Original Article

Language *and* Speech

Acoustic Cues in the Production and Perception of Cantonese Sarcasm

Language *and* Speech
1–34

DOI: 10.1177/00238309251333766
journals.sagepub.com/home/las
Sage

Chen Lan and Peggy Mok
The Chinese University of Hong Kong, Hong Kong

Abstract

Sarcasm has been associated with several acoustic cues, but exploration of such cues in Cantonese has been limited. The present study revisited the production and perception of sarcasm in Cantonese, investigating how prosodic features and voice quality measures signal sarcastic speech and how well they can be recognized. Eighteen native Hong Kong Cantonese speakers produced colloquial sentences with three attitudes: sarcasm, neutrality, and sincerity. Five prosodic parameters and three voice quality parameters were analyzed and compared between attitudes, genders, and individual speakers. In addition, 42 native listeners rated the degree of sarcasm and sincerity of the target utterances. Average rating scores were compared between attitudes. Results show that Cantonese sarcasm is characterized by a slower speech rate, lower mean F0, narrower F0 range, lower mean amplitude, greater amplitude range, higher harmonic-to-noise ratio (HNR), lower jitter, and lower shimmer than sincere speech. Speakers utilized different combinations of acoustic cues to express sarcastic feeling. Listeners were able to distinguish sarcasm from sincerity according to the acoustic cues alone in the absence of verbal context. The more acoustic cues were utilized in a sarcastic utterance, the easier it would be for the listeners to understand the implied sarcastic meaning. Moreover, the insertion of an intensifier "*zan55hai22 (really)*" enhanced the sarcastic intonation, increasing listeners' accuracy at recognizing the speaker's sarcastic intention. The present study contributes to a more comprehensive understanding of the relationship between prosody and sarcastic speech by using an improved method and providing evidence of production and perception in native Cantonese speakers.

Keywords

Sarcasm, speech production, speech perception, Cantonese, prosody

Introduction

Verbal irony has generally been described as a rhetorical device for either implying the opposite of what the literal content would be, or for expressing a different meaning from what is said (Brown & Levinson, 1978; Myers, 1990). Ironic criticism (using positive content to deliver a negative

Corresponding author:
Peggy Mok, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, Hong Kong.
Email: peggymok@cuhk.edu.hk

meaning) and ironic compliments (making use of negative content to give positive comments) are two types of irony (Dews et al., 1995). Sarcasm is generally associated with the former (Long & Graesser, 1988). Even though "irony" or "ironic" has been used to generally describe "sarcasm" or "sarcastic" situation, irony and sarcasm should be considered as two different phenomena. Sarcasm mostly occurs among people who are very familiar with each other, and thus usually has a particular target, conveying the most harmful type of negative attitude, and are more explicit and direct with clearer cues (Garmendia, 2018; Kreuz, 1996). Irony is with a more general target and usually associated with the phenomena that are funny or strange; for example, something happened out of your expectation, which is not the case of sarcasm. The present study focuses on sarcasm and the sarcastic intention delivered by the speakers or perceived by the listeners.

It has been reported that several cues convey sarcastic meanings. Facial expressions are non-verbal cues visually marking sarcasm; for example, raised eyebrows, mouth movement, blank face, pointing, or laughing (e.g., Attardo et al., 2003; Rockwell, 2001; Tabacaru, 2020; Utsumi, 2000). Contextual or lexical cues, prosodic cues, and semantic information are verbal cues of sarcasm. For instance, interjections, hyperbolic words such as extreme constructions of adverbs and adjectives, metaphors, or echoic mentions are reported to implicitly convey sarcastic intentions (e.g., Kreuz & Caucci, 2007; Kreuz & Roberts, 1995; Utsumi, 2000). Prosody shifts, such as heavy stress and slow speaking rate, are important paralinguistic cues displaying sarcasm (e.g., Utsumi, 2000). The present study mainly explores the prosodic cues of sarcasm in Cantonese, focusing on the acoustic parameters such as duration, fundamental frequency (F0), amplitude, and a combination of these parameters and voice quality parameters.

Previous studies on sarcasm have suggested that F0, duration, and amplitude were essential phonetic cues to distinguish sarcasm from non-sarcasm, although patterns varied across languages. For example, English sarcastic utterances were marked by a lower mean F0 and a slower speech rate (e.g., Bryant & Fox Tree, 2002; Cheang & Pell, 2008; Chen & Boves, 2018; Rockwell, 2000), while sarcasm in Italian was produced with a higher mean F0, a slower speech rate, and greater amplitude (e.g., Anolli et al., 2002). Separate patterns also existed for the same language. For example, a lower mean intensity for sarcasm than for sincerity in English was reported in Cheang and Pell (2009), whereas a reverse pattern was indicated in Rockwell (2000). In addition, voice quality is found to be modulated by speakers in affective communication. English, Mandarin, and Korean speakers have been reported to change voice quality while expressing a sarcastic attitude (Cheang & Pell, 2008; S. Li et al., 2020; Yang, 2021).

The sarcastic tone is commonly used in Cantonese, but the acoustic cues of Cantonese sarcastic speech have not been well examined. The present study continues the research on sarcastic speech by investigating the prosodic parameters (i.e., speech rate, F0 measures, and amplitude measures) and voice quality parameters (i.e., harmonic-to-noise ratio [HNR], jitter and shimmer) of sarcasm in Cantonese, aiming to identify how prosodic and voice quality features signal the sarcastic tone produced by native Cantonese speakers and how well native listeners can recognize it.

1.1 Acoustic features of sarcasm

There have been some studies investigating the prosodic properties of sarcasm, most of which focused on sarcasm in English. Earlier studies usually provided definitions of all the target attitudes to the speakers and used written materials to elicit different attitudes; for example, displaying the contexts, the attitude, and the target sentences with a card (e.g., Cheang & Pell, 2008; Rockwell, 2000; Rockwell, 2007). These studies analyzed the sarcastic utterances as a whole acoustically or perceptually, agreeing that conversational sarcastic utterances in English are characterized by a slower speech rate, lower mean F0, and narrower F0 range. A recent study asked the participants to

imagine themselves having a phone talk with a good friend and say the provided sentences as a response in sarcastic and sincere attitudes in British English and analyzed the acoustic measures of some key words, suggesting that sarcasm is distinguished from sincerity in key words with a longer duration and a flatter fall of F0 (Chen & Boves, 2018). Based on the acoustic analysis on some key words, Chen and Boves suggested that sarcasm is distinguished from sincerity in key words with a longer duration and a flatter fall of F0. Perception studies on English sarcasm usually asked the participants to identify the attitudes or rate the degree of sarcastic intentions by the sound of voice they heard (e.g., Bryant & Fox Tree, 2005; Cheang & Pell, 2011). These studies indicated that listeners were able to recognize the sarcastic meaning of a sentence or discriminate between sarcastic and non-sarcastic utterances in their native language according to prosodic cues, although individual differences were found. The speaker, the information that was presented beforehand, and the immediate context of the utterances affected how the sarcastic tone of voice was produced (Bryant & Fox Tree, 2005). For instance, speakers may produce one sarcastic utterance differently from another, and different prosodic strategies might be used to accommodate the diversity of contexts and the contents being conveyed. Native listeners may rely on multiple sources of information to infer a speaker's sarcastic intent.

In addition to the studies on English sarcasm, a few studies have examined sarcastic speech in other languages (e.g., Anolli et al., 2002 on Italian; Cheang & Pell, 2009 on Cantonese; Loevenbruck et al., 2013 on French; Rao, 2013 on Mexican Spanish; Niebuhr, 2014 on German; Tang & Gu, 2015 and S. Li et al., 2020 on Mandarin). These studies supported that duration, F0, and amplitude were significant markers of sarcasm, although the patterns varied across languages. Rao (2013) provided contextual information and the conditions (i.e., positive or negative) in written form and asked the speakers to produce the given responses with sarcastic or sincere attitudes, examining sarcastic speech in Mexican Spanish in terms of sentence- and word-level factors. In comparison to sincerity, sarcasm was produced by decreasing the speech rate and mean F0 of the sentence as a whole and by lengthening the stressed syllable in attitudinally relevant words. Niebuhr (2014) extended the investigation to voice quality, asking native German speakers to sarcastically and neutrally utter the target sentences to their friends without providing any contextual information. Niebuhr's study revealed a breathier voice quality and a higher degree of segmental reduction for German sarcasm. Considering the results found in English and other languages, speech rate was the most consistent prosodic marker of sarcasm with similar pattern across languages, and F0 was also a prominent cue, but the actual patterns varied across languages.

Gender differences have been found in Mexican Spanish, English, and Mandarin, but the results were mixed. For example, male speakers of British English relied more on lengthening the duration while female speakers relied more on lowering the F0 (Chen & Boves, 2018), whereas the effect of F0 measures on attitude was stronger for males than females in Mexican Spanish (Rao, 2013). A very recent study examining voice quality in Mandarin sarcastic speech also revealed that males' sarcastic production was charactered by a lower H1-H2, lower jitter, and lower shimmer, while females' sarcastic production had a higher H1-H2 and a higher H1-A1 (S. Li et al., 2020). The present study further explored the gender effect in production and perception of sarcasm by providing evidence of Cantonese.

Considering the research methods, early studies on sarcastic speech analyzed speech materials recorded by experienced actors hoping to ensure that the sentences were clearly produced with a sarcastic tone (e.g., Anolli et al., 2002; Rockwell, 2000). For example, Rockwell (2000) recruited professional radio announcers who were experienced in producing sarcastic speech. Contextual information was provided in written form for elicitation by most of the studies (e.g., Anolli et al., 2002; Cheang & Pell, 2009; S. Li et al., 2020; Niebuhr, 2014; Rockwell, 2000), while Loevenbruck and colleagues' (2013) study gave the contexts in the form of audio recording. There were also

some studies equipping the participants with a real person, asking them to produce the target sentences as a response to their friends in different attitudes without providing contextual information (e.g., Chen & Boves, 2018; Rao, 2013). It is possible that the different acoustic patterns of sarcasm found in these previous studies are due to the different elicitation methods.

Furthermore, most studies did not control the semantic cues in the experiment. Some semantic cues are used very frequently by speakers that they become markers of sarcasm (such as *thanks a lot* in English), regardless of the prosodic pattern of the utterance (Haiman, 1998; Kreuz & Caucci, 2007). It is difficult to conclude whether semantic cues or prosodic cues signal sarcasm more if the materials are not controlled for semantic cues. In addition, in some studies, a perceptual validation test was conducted before the acoustic analysis to check whether the utterances produced by the participants delivered the intended attitudes (e.g., Cheang & Pell, 2009, 2011; S. Li et al., 2020). However, not being perceived as expressing a sarcastic meaning clearly does not necessarily mean that the speaker did not produce the sentence with a sarcastic attitude. There might be multiple combinations of acoustic parameters signaling the colloquial production of sarcasm, and only some combinations are more easily noticed by the listeners. Therefore, a study of more natural data using a rigorous method is warranted.

Most previous studies on sarcasm were on English or European languages, with few being done on typologically different tone languages. Tang and Gu (2015) conducted a perceptual and acoustic study of six attitude pairs in Mandarin Chinese. The speakers were provided a role-play scenario for attitude elicitation, and the listeners were asked to listen to the target sentences and identify the attitude. Their study suggested that Mandarin sarcasm could be perceptually recognized with higher accuracy compared to other attitudes, but it might not be well distinguished from praising by prosodic cues (e.g., F0 measures). A recent study expanded the understanding of the acoustic features of Mandarin sarcasm by investigating the role of voice quality in expressing sarcastic attitude (S. Li et al., 2020). They applied a similar elicitation method as in Tang and Gu (2015) but compared only sarcastic and sincere attitudes. In comparison to sincere speech, sarcastic speech had a creakier voice, a lower F0, a greater degree of vocal fold adduction, and less noise. These two studies also reported that sarcastic Mandarin was produced with a lower mean F0 (S. Li et al., 2020; Tang & Gu, 2015). It will be interesting to further examine how F0 is related to sarcasm in another tone language with a more complicated tone system (e.g., Cantonese). Also, in addition to F0, what and how other acoustic cues characterize Cantonese sarcasm can also be investigated for a more comprehensive understanding of the association between acoustic properties and sarcasm.

1.2 Sarcasm in Cantonese

Previous research on Cantonese sarcasm or other ironic forms in Cantonese mostly focused on syntactic structures (M. K. M. Chan, 2002; J. P. W. Li et al., 2013; Matthews & Yip, 1994). For instance, the Cantonese sentence-final particle (SFP) /tsɛk55/ is commonly used to mark a sense of irony in a positive literal utterance according to the speech context (K. K. L. Chan & To, 2016). Although some studies mentioned the role of prosody in the comprehension of irony (e.g., J. P. W. Li et al., 2013), only one study has investigated the prosodic features. Cheang and Pell (2009) measured the acoustic parameters of the utterances produced by six native Cantonese speakers in Canada with four attitudes (sarcasm, humor, sincerity, and neutrality), indicating that a higher mean F0, a narrower F0 range, a slower speech rate, and a more restricted amplitude range distinguished sarcastic utterances from non-sarcastic ones. A higher HNR value also differentiated sarcasm from sincere utterances, which aligned with the results for Mandarin (S. Li et al., 2020). Cheang and Pell (2011), which is the only published study focusing on the perception of sarcasm

in Cantonese, further revealed that Cantonese and English listeners were able to recognize sarcasm and distinguish sarcasm from sincerity in their native languages; however, listeners in both groups had little ability to recognize sarcasm in a non-native language. Based on their finding that Cantonese and English speakers employed opposite patterns of mean F0 to convey sarcasm versus sincerity, with a higher mean F0 for sarcasm in Cantonese and a relatively lower mean F0 for sarcasm in English, Cantonese and English listeners might have different expectations about the alteration of mean F0 while detecting the sarcastic and sincere intentions, which may account for the incorrect recognition of the speaker's intentions in the non-native language.

Previous studies on the production and perception of Cantonese sarcasm have provided an important first step in the investigation of sarcastic cues in Cantonese using controlled materials; however, there are some limitations which may render their findings tentative. Regarding the materials, some target utterances in Cheang and Pell (2009) appeared to be unidiomatic for native Cantonese speakers. For example, it is uncommon for Cantonese speakers to use key phrases such as "係啩 *hai22gwa33* (*I suppose.*)," "係咩 *hai22me55* (*Is that so?*)," or "嘩哎 *wa55aai55* (*Oh boy.*)" to start expressions of different attitudes. Even though the key phrases and the combined sentences used in their study are syntactically acceptable, it is not usual to express sarcasm using these structures, which may affect the naturalness of the utterances. Also, SFPs play an essential role in conveying different attitudes and emotions in Cantonese (Fung, 2000; Law, 2002; J. P. W. Li et al., 2013; Matthews & Yip, 1994). The absence of the SFPs in Cheang and Pell (2009) rendered the target sentences less colloquial. The procedure of the production experiment posed a second limitation. Sarcasm is most naturally expressed during conversation with close friends (Rockwell, 2000, 2007); thus, a better way to elicit sarcastic speech is to put the participants in a conversational context. However, the materials in Cheang and Pell (2009) were presented in a written format which might reduce the naturalness of the participants' responses. Also, even though the participants and the encoders in Cheang and Pell (2009, 2011) were born, raised, and educated in a Cantonese environment (Hong Kong or Guangzhou), they emigrated to Canada as young adults, meaning that they had been in an English immersion environment for at least a few years at the time of their participation. Their Cantonese speech patterns might have been affected by the exposure to English prosodic patterns. Therefore, a more comprehensive study is needed to corroborate the findings in previous studies.

1.3 The present study

Although the prosodic correlates of sarcastic speech in non-tone languages have been studied, the acoustic cues of sarcasm in tone languages have not been systematically investigated. Cantonese has a different prosodic system compared to non-tone languages and a more complex tone system compared to Mandarin. The communication styles of Cantonese speakers also differ from the speakers of other languages, for example, using SFPs. Considering the limitations of previous studies, the present study revisited the acoustic cues of sarcasm in Cantonese with a more rigorous method and a greater number of participants, investigating both the production and perception of sarcastic speech in Cantonese.

All the target sentences were first judged by two native Hong Kong Cantonese speakers to ensure that they were commonly used among speakers in daily communication for expressing sarcastic or sincere attitudes. The SFPs "呀 *aa33*" or "喎 *wo33*" were chosen for more natural utterances in our study. A scenario approach (Scherer et al., 2001) was used to elicit expressions naturally, providing the participants with short scenarios with positive or negative situations.

Considering that sarcasm is most naturally expressed with close friends, the contexts were presented using audio recorded by native speakers to put the participants in a conversational context.

In addition, as most of the previous studies analyzed the attitudinal utterance as one whole unit, little is known about the acoustic patterns of smaller units, such as constituents (e.g., adjectival phrase [AP]). The present study conducted both sentence-level and phrase-level analyses, aiming to investigate whether the acoustic cues associated with sarcasm change throughout the utterance; for example, whether the cue difference is consistent in each part of an utterance and as a whole; or only appears in some parts of the sentence. A degree modifier (DM), an AP, and an SFP were examined in all the sentences. The DMs " 好 *hou25*" or "幾 *gei25* (*very*)" locating before the adjectives serve as an important component for a colloquial sentence in Cantonese and are acceptable for expressing both positive and negative feelings, which are literally neutral terms for native speakers and listeners. AP is the most relevant element for attitudinal communication since it usually carries a positive or negative feeling of the speakers. In the current study, all the APs were literally positive in Cantonese. The SFPs "呀 *aa33*" and "喎 *wo33*" are acceptable for ending positive or negative utterances by native Cantonese speakers and listeners, and thus could be regarded as literally neutral terms.

Some comparative expressions such as "*so*" in English are considered as typical elements to indicate sarcasm (Camp, 2011; Lebedeva, 2021), and such words tend to make more sarcastic tonal effect. The intensifier "真係 *zan55hai22* (*really*)" is frequently used to express criticism and for assuring truth or sincerity in Cantonese (Fung, 2000). Since adding this intensifier may reinforce the intention in the utterances, it raises a question whether including this intensifier would induce a focusing effect manifesting itself in the form of an amplitude or F0 peak. Therefore, in the present study, the target intensifier was included as the fourth key phrase, aiming to examine whether the presence and absence of the intensifier influence the production and perception of sarcasm. If this intensifier was the focused element, it would be further explored whether its acoustic features typically signal Cantonese sarcasm and whether there would be interaction between this target intensifier and the other key phrases. Finally, the investigation of Cantonese sarcasm was extended by exploring variability between genders and among speakers.

In summary, the present study asked the following research questions: (1) What are the acoustic cues of Cantonese sarcastic speech? (2) How well can native Cantonese speakers perceive sarcasm according to the acoustic cues? and (3) Does having an intensifier enhance sarcastic intonation?

2 Method

2.1 Participants

Eighteen native Hong Kong Cantonese speakers (nine females and nine males) aged between 18 years and 5 months and 23 years and 8 months participated in the production task, and 42 native Hong Kong Cantonese listeners (23 females and 19 males) aged between 17 years and 11 months and 23 years and 5 months took part in the perception experiment. According to their language background questionnaires, all the participants were born and grew up in Hong Kong, having at least one of their parents being a native Hong Kong Cantonese speaker. They went to local primary and secondary schools and currently were undergraduate students at a university in Hong Kong. Cantonese was the most used language in their daily communication taking up around 85.8% of their time in comparison to the percentage of using other languages (e.g., English, Mandarin). There was no overlap between the production and the perception participants. All the participants reported no speech or hearing problems or learning difficulties.

Table 1. Example of the target utterances with English translations.

		Intensifier	Degree modifier	Adjectival phrase	Sentence-final particle
a	你 (You)		好 (very)	醒 (smart)	呀
b	你 (You)	真係 (really)	好 (very)	醒 (smart)	呀

In the table, (a) is a sentence without an intensifier, and (b) is a sentence with the target intensifier.

Figure 1. Example of the negative and positive scenarios and the corresponding pictures.

2.2 Production materials

Two sets of sentences were designed for the present study. The first set contained the target sentences with a DM, an AP, and an SFP. The intensifier "真係 *zan55hai22* (*really*)" was inserted before the DM in the first set to create the second set of target sentences. In total, 12 pairs of sentences as exemplified in Table 1 were produced with three attitudes (neutral, sincere, sarcastic) and with three repetitions (see Appendix Table A1 for all the target sentences and the scenarios). Each participant produced 216 target utterances (12 groups of utterances × 2 sets × 3 attitudes × 3 repetitions). Positive and negative scenarios recorded by two native Cantonese speakers and visual aids in the form of relevant pictures were provided (see Figure 1). For neutral speech, no biasing context was provided. Instead, the participants received an instruction to read out the sentences displayed on the screen.

2.3 Perception materials

A subset of sarcastic and sincere production data was selected to form the perception stimuli. To investigate how well the native listeners could detect the sarcastic meanings in colloquial speech, all the stimuli were not manipulated, namely that the original recordings produced by the native speakers were heard and judged by the listeners. An acoustic measure significantly distinguishing a sarcastic sentence from its sincere counterpart following the general pattern found in the

production experiment was determined as an acoustic cue utilized by the speaker to signal sarcasm. Each stimulus contained at least one acoustic cue, and all the stimuli displayed different combinations of acoustic cues used by the speakers. In total, audio recordings of 100 utterances, consisting of 50 pairs of sentences produced with two attitudes (sarcastic and sincere) and two sentence sets (with and without the intensifier), were included.

2.4 Procedure

Two groups of participants were paid to attend the production or perception experiments. Before the experiment, the participants were briefed about the study. With their consent, a language background questionnaire asking for some personal information (e.g., age, languages used in different situations) was filled. The speakers participated in the production experiment individually in a sound-treated room. A ZOOM H2n solid-state recorder with a sampling rate of 44,100 Hz was used for recording. The recorder was placed in front of the speaker 20 cm away from the speaker's mouth. During the experiment, no definition of sarcasm and the purpose of study was provided, and the target sentences were not provided to the participants in advance for practicing. In the first part, the target sentences were presented using PowerPoint, and each slide contained only one sentence. No biasing context was given. The participants were instructed to read the displayed sentences one by one neutrally and naturally. There were practice trials which contained a sentence with the target intensifier and a sentence without the target intensifier to familiarize the participants with the experiment procedure. In the second part, a picture and a target sentence were presented on each slide, and the audio scenario was played automatically. The speakers listened to the audio and produced the target sentence according to the context provided by the audio and the picture. The sentences were randomized and shown on the screen in different orders in each repetition. The participants went through the practice trials containing a sentence with the target intensifier and a sentence without the target intensifier with both positive and negative scenarios. There was no overlap between the materials for the practice trials and for the formal experiment.

For the perception experiment, the participants completed an online perception task individually. During the experiment, the audio recordings of the stimuli were randomized and presented to the listeners without providing them with the sentences in written form nor the verbal contexts. The listeners were instructed to click the link of the sound file, listen to the stimuli, and rate each target utterance on a 6-point Likert-type scale from 1 to 6 to indicate whether they perceived the sentence as being produced with a very sincere (1) or very sarcastic (6) tone of voice, or somewhere in between. The listeners did not attend professional training on perceptual analysis of acoustic features prior to the task. They were not expected to point out which acoustic features were used by the speaker; instead, they only needed to judge and rate the degree of the attitudes delivered by the speakers as native listeners.

2.5 Data analysis

Both acoustic and statistical analyses were applied to the production data. A total of 3,888 utterances (12 pairs of target utterances $\times$ 2 sentence sets $\times$ 3 attitudes $\times$ 3 repetitions $\times$ 18 speakers) were measured in Praat (Boersma & Weenink, 2024) using ProsodyPro (Xu, 2013). Eight acoustic parameters, including the speech rate, mean F0, F0 range, mean amplitude, amplitude range, the HNR, jitter, and shimmer were measured for each utterance as a whole and for the four key phrases including the intensifier *zan55hai22* "*really*," the degree modifiers *hou25* or *gei25* "*very*," the APs, and the SFPs *aa33* or *wo33*. The number of syllables and the total duration of each utterance or phrase were extracted to compute the speech rate. Speech rate was calculated by dividing the

number of syllables by the duration (in second) of each utterance or phrase to capture how fast or how slowly a speaker delivered different attitudes. For the F0 parameters, mean F0, minimum F0, and maximum F0 were measured in Hertz (Hz). Mean F0 was measured to evaluate the pitch level characteristic of a speaker delivering different attitudes. The F0 range was obtained by subtracting the minimum F0 from the maximum F0 to evaluate how F0 was distributed within an utterance or a phrase and how the variation differed according to attitudes. Regarding the amplitude variables, mean amplitude, minimum amplitude, and maximum amplitude were measured in decibels (dB). Mean amplitude was measured to quantify the acoustic energy in a speaker's sound while conveying different attitudes; simply, it described the loudness of the utterances or phrases. The amplitude range was obtained by subtracting the minimum intensity from the maximum intensity to capture how the acoustic energy was distributed within an utterance or a phrase and how the variation differed according to attitudes. In addition, the HNR value, jitter, and shimmer were measured to capture the voice characteristics of the speakers and how their voice quality changed with different attitudes. The HNR refers to the degree of acoustic periodicity, assessing the ratio between periodic and non-periodic components of voiced speech. A higher HNR value represents less noise in the speech signal. The HNR value in the present study was measured in dB between the frequency range of 0 and 5,000 Hz using the algorithm with cross-correlation method in Praat (Boersma & Weenink, 2024; Xu, 2013). Jitter indicates the cycle-to-cycle variation in frequency, calculated by dividing the mean absolute difference between consecutive intervals by the mean interval. Shimmer refers to the variation of amplitude, measured by dividing the mean absolute difference between the amplitudes of consecutive intervals by the mean amplitude. The jitter and shimmer are associated with the breathiness or roughness of a speaker's voice.

All data were converted into z-scores using each person's mean before statistical analysis. A linear mixed-effects model was conducted to analyze each acoustic parameter in the R program (R Core Team, 2022) using the lmerTest package (Kuznetsova et al., 2017). The maximal random effect structure was applied for each model for hypothesis testing (Barr et al., 2013). For sentence-level analysis, in each model, Attitude, sentence Set, and Gender were entered as a fixed effect (Attitude × Set × Gender), and the random intercepts for Speaker and Sentence were included. The random slopes for attitude and set by speaker and by sentence were excluded for the convergence of the models. For phrase-level analysis, linear mixed-effects models were conducted to analyze the eight parameters of the four key phrases in the sentences with the target intensifier, and of the three key phrases (DM, AP, and SFP) in the sentences without the target intensifier. In each model, Attitude and Phrase (Attitude × Phrase) were entered as a fixed effect while Speaker and Sentence were set as the random effects.

For the perception data, the rating scores from 4,200 responses (50 target utterances × 2 attitudes × 42 participants) were analyzed, considering Attitude, sentence Set, and Gender (Attitude × Set × Gender) as fixed effect, and Listener as a random effect. Wilcoxon signed rank test was conducted to compare the rating scores between the two Attitudes and two sentence Sets. Simple linear regressions were conducted to explore the relationship between the acoustic cues used by the speakers and the degree of sarcastic attitude perceived by the listeners. To observe the inter-rater reliability, the Intraclass Correlation Coefficient (ICC) was calculated using the "icc" function in the irr package (Gamer & Lemon, 2019) based on an average rating (k = 42), absolute agreement, and a two-way model. The 95% confident interval of the ICC estimate for the perceptual rating in the present study is 0.974–0.985 ($p < .001$), indicating an excellent absolute agreement between raters according to Koo and Li's (2016) guideline.

For each model, the "anova" function in the lmerTest package (with Satterthwaite's method) was used to calculate the F value and p value of the main and interaction effects and determine degree of freedom. Pairwise comparisons were further conducted when a significant interaction effect was revealed.

Table 2. Mean values (*z-scores*) of the target acoustic variables.

	Gender	Speech rate	Mean F0	F0 range	Mean amplitude	Amplitude range	HNR	Jitter	Shimmer
Sarcasm		−0.76 (0.87)	−0.33 (0.93)	−0.04 (0.97)	0.11 (0.92)	0.30 (1.01)	0.42 (1.05)	−0.16 (0.94)	−0.22 (0.96)
	Female	−0.78	−0.50	−0.06	0.08	0.26	0.42	−0.07	−0.22
	Male	−0.74	−0.15	−0.03	0.13	0.34	0.41	−0.25	−0.22
Neutrality		0.46 (0.73)	−0.16 (0.79)	−0.18 (0.90)	−0.61 (0.82)	−0.28 (0.94)	−0.47 (0.85)	0.30 (0.99)	0.29 (0.97)
	Female	0.53	−0.16	−0.19	−0.62	−0.30	−0.53	0.28	0.39
	Male	0.40	−0.16	−0.16	−0.60	−0.26	−0.41	0.32	0.19
Sincerity		0.29 (0.85)	0.46 (0.96)	0.18 (0.91)	0.50 (0.90)	−0.02 (0.95)	0.05 (0.86)	−0.16 (0.88)	−0.09 (0.90)
	Female	0.25	0.65	0.21	0.54	0.05	0.10	−0.23	−0.20
	Male	0.33	0.27	0.14	0.47	−0.08	0.00	−0.10	0.01

Standard deviations are listed in the brackets.

3 Results

3.1 Production

Table 2 summarizes the average z-transformed values of the target prosodic parameters and the voice quality measurements, including the speech rate, mean F0, F0 range, mean amplitude, amplitude range, the HNR value, jitter, and shimmer.

3.1.1 Speech rate. The analysis of **speech rate** revealed a significant main effect of Attitude, $F(2, 3865) = 1571.6691$, $p < .001$, and of Set, $F(1, 3865) = 1045.2718$, $p < .001$. Sarcasm was expressed significantly more slowly than sincerity (Est. = 1.095, $SE = 0.047$, $t(3865) = 23.197$, $p < .001$) and neutrality (Est. = 1.376, $SE = 0.047$, $t(3865) = 29.161$, $p < .001$). As shown in Figure 2, two sets of sentences shared the overall pattern, but the speaking rate of the utterances with the intensifier was faster than that of the utterances without the intensifier.

3.1.2 F0 measures. The analysis of **mean F0** revealed that Attitude was a significant predictor, $F(2, 3865) = 302.2286$, $p < .001$, together with a significant main effect of Set, $F(1, 3865) = 19.3781$, $p < .001$. There was a significantly lower mean F0 for the sarcastic speech than for the sincere speech (Est. = 1.122, $SE = 0.067$, $t(3865) = 16.641$, $p < .001$) and for neutral speech (Est. = 0.257, $SE = 0.067$, $t(3865) = 3.804$, $p = .0001$). Both sets shared the same pattern as the overall pattern, but the utterances with the intensifier had a higher mean F0. The **F0 range** analysis revealed a significant main effect of Attitude, $F(2, 3865) = 52.5686$, $p < .001$, with a significantly narrower and higher F0 range for sarcasm than for sincerity (Est. = 0.246, $SE = 0.069$, $t(3865) = 3.802$, $p = .0001$) and for neutrality (Est. = −0.178, $SE = 0.069$, $t(3865) = -2.566$, $p = .010$), respectively. Set was also a significant predictor, $F(1, 3865) = 45.3828$, $p < .001$, with a higher F0 range for the utterances with the intensifier.

3.1.3 Amplitude measures. The analysis of **mean amplitude** yielded a significant main effect of Attitude, $F(2, 3865) = 555.3574$, $p < .001$, with a significantly lower mean amplitude for sarcastic speech than for sincere speech (Est. = 0.428, $SE = 0.068$, $t(3865) = 7.138$, $p < .001$) and a significantly higher mean amplitude for sarcasm than for neutrality (Est. = −0.565, $SE = 0.068$, $t(3865) = -8.363$, $p < .001$). Set was also reported as a significant main effect, $F(1, 3865) = 7.3921$, $p = .007$, together with a significant interaction between Attitude and Set, $F(2, 3865) = 7.2853$, $p = .0007$. Both sets shared the same pattern as the overall pattern, but the utterances with the intensifier had a lower mean amplitude and a larger distinction between sarcasm and neutrality. The analysis of the **amplitude range** revealed a significant main effect of Attitude, $F(2, 3865) = 172.3087$, $p < .001$, and of Set, $F(1, 3865) = 63.4851$, $p < .001$. Sarcastic speech had a significantly greater amplitude range than sincere speech (Est. = −0.156, $SE = 0.063$, $t(3865) = -2.477$, $p = .013$) and neutral speech (Est. = −0.479, $SE = 0.063$, $t(3865) = -7.601$, $p < .001$). The two sentence sets shared the same pattern as the overall pattern, with a greater amplitude range for the utterances with the intensifier.

3.1.4 Voice quality. The analysis of the **HNR** revealed a significant main effect of Attitude, $F(2, 3865) = 464.2376$, $p < .001$, and of Set, $F(1, 3865) = 169.1013$, $p < .001$, together with a significant interaction between Set and Attitude, $F(2, 3865) = 10.8731$, $p < .001$. Sarcastic utterances had a significantly higher HNR value (i.e., less breathy voice) than sincere utterances (Est. = −1.056, $SE = 0.059$, $t(3865) = -7.074$, $p < .001$) and the neutral utterances (Est. = −0.414, $SE = 0.059$, $t(3865) = -18.036$, $p < .001$), coinciding with the patterns of both sets. Compared to the sentences

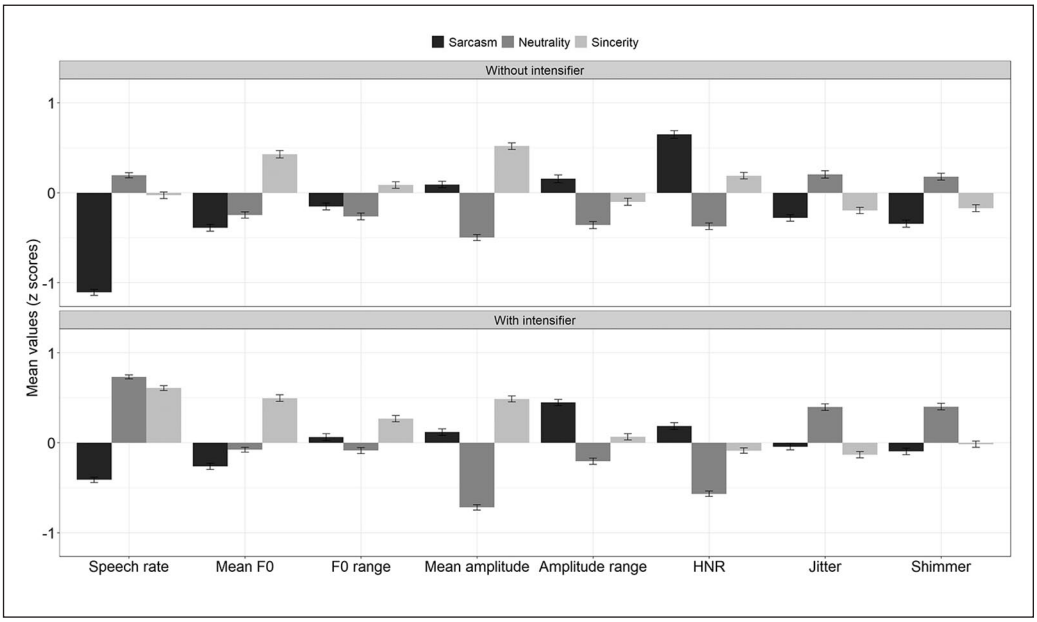

Figure 2. Mean values (z-scores) of the acoustic parameters across three attitudes in the utterances without (top) and with (bottom) the target intensifier.
Note. Error bars indicate the standard errors.

without the intensifier, those with the intensifier had a lower HNR value and a smaller distinction between sarcasm and non-sarcasm. The **jitter** analysis revealed a significant main effect of Attitude, $F(2, 3865) = 124.2660$, $p < .001$, and of Set, $F(1, 3865) = 34.7596$, $p < .001$, together with a significant interaction between Attitude and Set, $F(2, 3865) = 3.3197$, $p = .030$. The jitter value differentiated sarcasm from neutrality with a significantly lower jitter (Est. = 0.373, $SE = 0.068$, $t(3865) = 5.491$, $p < .001$), but no significant change was found between sarcasm and sincerity. The pattern of both sentence sets aligned with the overall pattern, displaying a higher value for those with the intensifier. The analysis of **shimmer** showed a significant main effect of Attitude, $F(2, 3865) = 118.4812$, $p < .001$, and of Set, $F(1, 3865) = 52.5659$, $p < .001$. A lower shimmer significantly distinguished sarcasm from neutrality (Est. = 0.597, $SE = 0.069$, $t(3865) = 8.642$, $p < .001$), which appeared in both sentence sets. A lower shimmer significantly differentiated sarcastic from sincere utterances without the intensifier. Overall, Cantonese sarcasm is not breathier than neutrality or sincerity.

3.1.5 Phrase-level analyses. Figure 3 shows the normalized mean values of eight acoustic parameters for the four key phrases, including the target intensifier "*zan55hai22 (really)*," the DM, the AP, and the SFP, in the utterances with the intensifier and for the three key phrases in the utterances without the intensifier across attitudes. For all the parameters in both sentence sets, there were significant main effects of Attitude and significant interaction between Attitude and Phrase (see Table 3). Phrase had a significant main effect for all the parameters except for the F0 range, mean amplitude, jitter, and shimmer in the utterances without the intensifier. The following paragraphs report the patterns according to the key phrases, and Table B1 in Appendix B provides detailed statistical information.

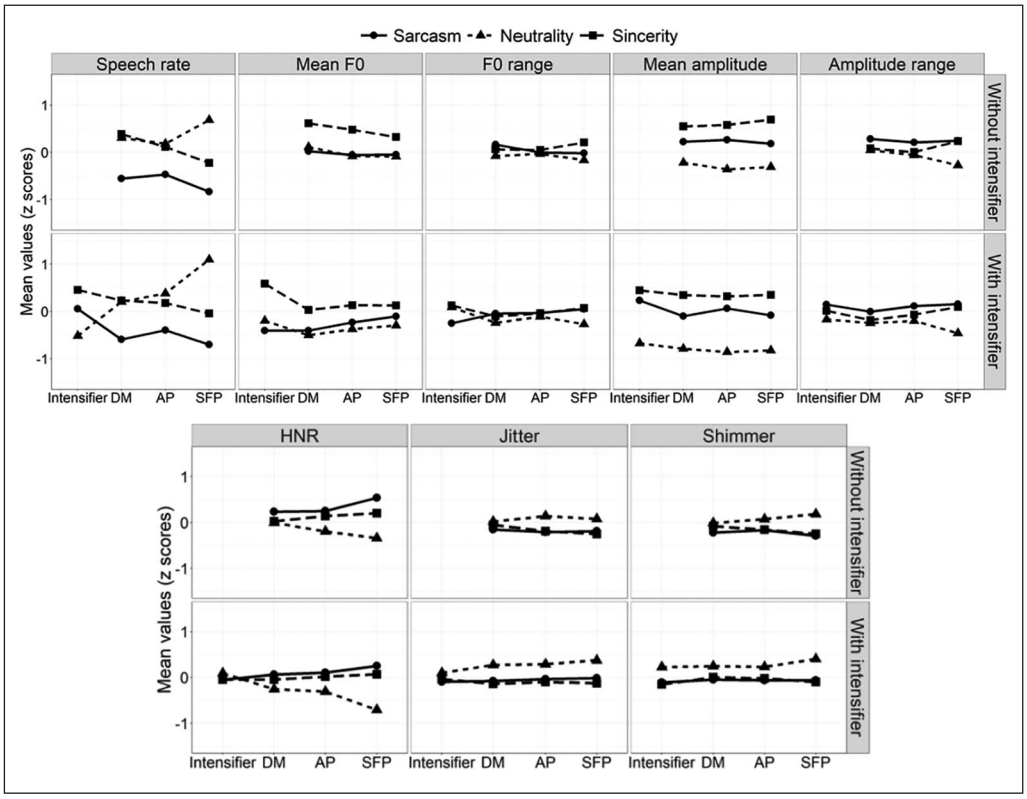

Figure 3. Mean values (z-scores) of the acoustic parameters of the three key phrases in the utterances without the target intensifier (upper panels) and the four key phrases in the utterances with the intensifier (lower panels) across the three attitudes.

First, compared to the sincere utterances, the **target intensifier** in the sarcastic utterances was produced significantly more slowly, with a declination of mean F0, F0 range, and mean amplitude as well as an increase in amplitude range, coinciding with the sentence-level pattern. The patterns comparing sarcasm and neutrality aligned with the overall sentence pattern of the eight parameters except for speech rate and HNR. The sarcastic intensifier was produced faster and with a lower HNR than the neutral one. Larger mean F0 and F0 range distance between sarcasm and sincerity were found for the intensifier than for other key phrases, while no distinctive distance was found for other acoustic parameters, suggesting that the target intensifier mainly carried the alteration of F0 measures.

Second, Cantonese speakers significantly reduced their speaking rate, mean F0, as well as the mean amplitude and increased the amplitude range while producing the sarcastic **DMs** in comparison to the sincere ones, aligning with the sentential pattern. A higher HNR and a lower shimmer were only found in the sentences without the intensifier. Coinciding with the sentence-level patterns, the sarcastic DMs were spoken more slowly, with a greater F0 range, higher mean amplitude, higher HNR, lower jitter, and lower shimmer compared to the neutral speech. Significant change in mean F0 was only found in the sentences with the intensifier, showing an increase in the mean F0, which was opposite to the overall pattern.

Table 3. *F* values for the main effects of Attitude and of Phrase, and the interaction between Attitude and Phrase in the utterances without and with the target intensifier.

Parameters	Utterances without the target intensifier			Utterances with the target intensifier		
	Attitude *df* = 2, 5,823	Phrase *df* = 2, 5,823	Attitude × Phrase *df* = 4, 5,823	Attitude *df* = 2, 7,760	Phrase *df* = 3, 7,760	Attitude × Phrase *df* = 6, 7,760
Speech rate	820.567***	22.690***	89.903***	573.097***	15.489***	281.852***
Mean F0	227.204***	26.205***	3.685**	364.968***	41.617***	26.984***
F0 range	29.940***	1.930	9.863***	19.553***	6.730***	20.997***
Mean amplitude	544.214***	0.621	5.302***	1221.919***	20.795***	4.060***
Amplitude range	68.054***	4.406*	13.512***	108.498***	8.075***	10.162***
HNR	167.147***	3.104*	24.684***	124.335***	6.754***	45.400***
Jitter	65.826***	2.557	5.951***	114.418***	3.287*	4.539***
Shimmer	75.019***	0.828	9.369***	115.867***	3.731*	2.991**

Significant level (p value): "***" < .001; "**" < .01; "*" < .05; " " not significant.

Third, being consistent with the overall sentence pattern, the speech rate, mean F0, and mean amplitude of the sarcastic **APs** significantly decreased while amplitude range was significantly increased, and in addition, there was a higher HNR value as well as a lower jitter and shimmer relative to the neutral APs. A significantly higher mean F0 for sarcastic APs only appeared in the sentences with the target intensifier, which differed from the general sentence pattern.

Fourth, compared to the sincere attitude, the **SFPs** with sarcastic attitude had a significantly slower speaking rate, lower mean F0, and lower mean amplitude, following the sentence-level pattern. The change of the other parameters varied according to the sentence sets. There was a significantly narrower F0 range, a higher HNR, and a lower shimmer for sarcasm than for sincerity in the sentences without the intensifier, while these three parameters did not significantly change while the SFPs were produced with the intensifier. A decrease in the speaking rate, mean amplitude, jitter, and shimmer, together with a greater F0 and amplitude range and a higher HNR significantly discriminated sarcastic SFPs from the neutral ones in both sentence sets, being consistent with the sentential pattern. A significantly higher mean F0 only appeared in the sentences with the intensifier, but the pattern differed from the sentential patten.

To sum up, a slower speech rate, lower mean F0, lower mean amplitude, and higher HNR distinguished sarcasm from sincerity throughout the whole utterance. The SFP did not carry the change in amplitude range, and the decrease in jitter and shimmer only appeared on the DM. AP in the utterance without the intensifier as well as DM and AP in the utterance with the intensifier did not carry the alteration in F0 range. A higher mean amplitude, greater amplitude range, higher HNR value, and lower jitter and shimmer distinguished sarcasm from neutrality throughout the whole utterance. A faster speech rate appeared on the intensifier, whereas slower speech rate was distributed for DM, AP, and SFP. Without the intensifier, the declination on mean F0 happened for DM only, and higher F0 range appeared on DM and SFP. The change of mean F0 was distributed throughout the utterance with the intensifier, with a lower mean F0 for the intensifier and higher mean F0 for DM, AP, and SFP.

3.1.6 Gender differences. The analysis on **speech rate** revealed a significant interaction between Attitude and Gender, $F(2, 3865) = 11.45$, $p < .001$. Both male (Sincerity: Est. = 1.063, $SE = 0.038$, $t(1930) = 28.025$, $p < .001$; Neutrality: Est. = 1.133, $SE = 0.038$, $t(1930) = 29.861$, $p < .001$) and female speakers (Sincerity: Est. = 1.035, $SE = 0.037$, $t(1930) = 27.788$, $p < .001$; Neutrality: Est. = 1.313, $SE = 0.037$, $t(1930) = 35.249$, $p < .001$) produced sarcasm with a significantly slower speech rate compared to other attitudes, aligning with the general sentence pattern; however, as shown in Table 2, the durational difference between sarcasm and non-sarcasm was larger in males' speech. **Mean F0** also had a significant interaction effect, $F(2, 3865) = 59.055$, $p < .001$. The mean F0 was significantly reduced to discriminate sarcasm from sincerity by both males (Est. = 0.510, $SE = 0.070$, $t(1930) = 7.293$, $p < .001$) and females (Est. = 1.122, $SE = 0.064$, $t(1930) = 17.488$, $p < .001$), aligning with the overall sentence pattern, with a larger difference for females. Significant difference between sarcasm and neutrality only appeared among female speakers (Est. = 0.257, $SE = 0.064$, $t(1930) = 3.998$, $p < .001$). There was a significant interaction between Attitude and Gender regarding **jitter**, $F(2, 3865) = 10.5982$, $p < .001$, and **shimmer**, $F(2, 3865) = 17.6818$, $p < .001$. A lower jitter and shimmer significantly distinguished sarcasm from neutrality by male speakers (jitter: Est. = 0.595, $SE = 0.067$, $t(1930) = 8.881$ $p < .001$; shimmer: Est. = 0.449, $SE = 0.069$, $t(1930) = 6.501$, $p < .001$) and female speakers (jitter: Est. = 0.373, $SE = 0.069$, $t(1930) = 3.440$ $p < .001$; shimmer: Est. = 0.597, $SE = 0.068$, $t(1930) = 8.745$, $p < .001$), coinciding with the general sentence pattern. However, significant difference between sarcasm and sincerity only occurred in males' production (jitter: Est. = 0.200, $SE = 0.067$, $t(1930) = 2.978$, $p = .003$; shimmer: Est. = 0.275, $SE = 0.069$, $t(1930) = 3.976$ $p < .001$).

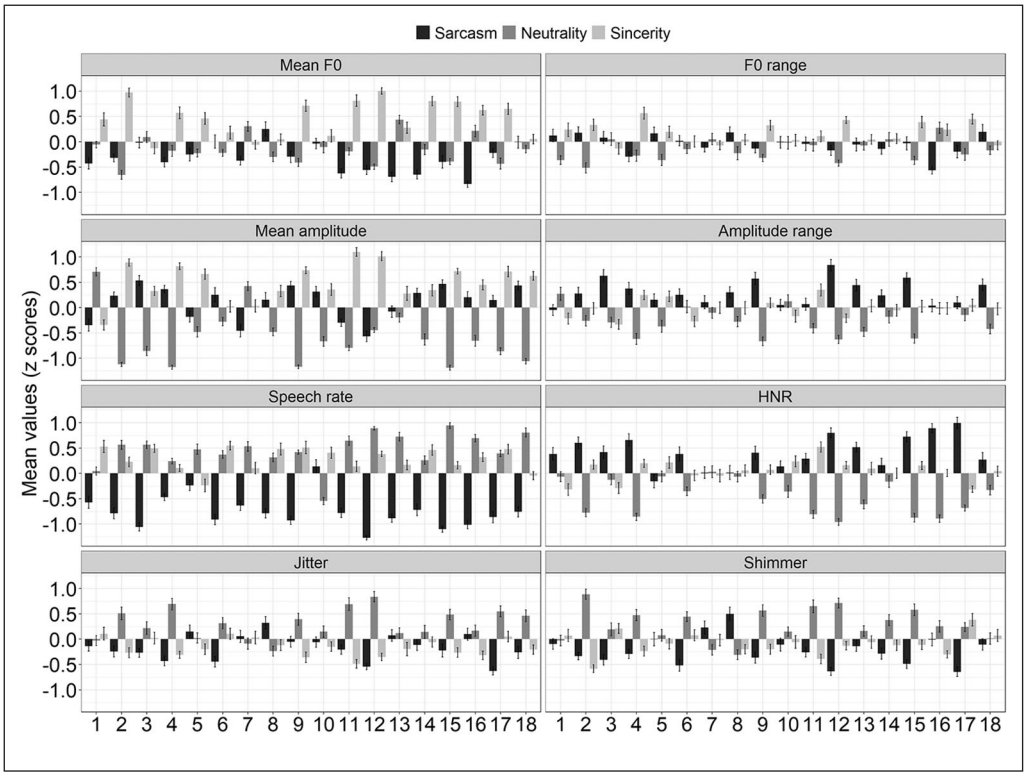

Figure 4. Mean values (z-scores) of the eight acoustic parameters across the three attitudes by speakers.

Table 4. Number of the speakers regarding the use of acoustic cues in sarcastic utterances compared to sincere and neutral utterances (Total number of speakers: 18).

Sarcasm vs. Sincerity		Sarcasm vs. neutrality	
Acoustic cues	N	Acoustic cues	N
Slower speech rate	17	Slower speech rate	17
Lower mean F0	13	Lower mean F0	6
Narrower F0 range	6	Larger F0 range	3
Lower mean amplitude	10	Higher mean amplitude	14
Larger amplitude range	9	Larger amplitude range	12
Higher HNR	12	Higher HNR	14
Lower jitter	3	Lower jitter	11
Lower shimmer	5	Lower shimmer	9

To sum up, gender difference existed particularly for mean F0 and voice quality parameters. Female speakers favored the reduction of mean F0, while male speakers usually reduced jitter and shimmer.

3.1.7 Individual patterns. Figure 4 shows the normalized mean values of the acoustic parameters in the production of each speaker (see Table S1 in the Supplementary Material for the statistical results). An acoustic parameter which was significantly different between a speaker's sarcastic and non-sarcastic production following the overall sentential pattern was considered as being utilized by this speaker as an acoustic cue to distinguish sarcasm from non-sarcasm. For instance, when an individual speaker delivered sarcastic attitude with a significantly slower speech rate and lower mean F0 compared to sincere attitude with no significant difference on other parameters, this speaker was considered to have used a combination of a slower speech rate and a lower mean F0 as the acoustic cues to convey sarcasm.

Table 4 presents how many speakers used a particular cue to differentiate sarcasm from other attitudes. Results showed that the speakers utilized different combinations of acoustic cues to convey Cantonese sarcasm. None of the speakers applied all the eight cues together. Over half of the speakers used more than four acoustic cues. A slower speech rate was utilized by most of the speakers (94%) to distinguish sarcasm from sincerity, followed by a lower mean F0 (72%), a higher HNR (67%), and a lower mean amplitude (56%). For the distinction between sarcasm and neutrality, similarly, 94% and 78% of the speakers used a slower speech rate and a higher HNR. Amplitude variables were also favored by the speakers, with fourteen speakers increased the mean amplitude and twelve speakers enlarged the amplitude range for sarcasm. In addition, speakers relied more on the modulation of their voice quality for distinguishing sarcasm from neutrality than from sincerity. Over 61% of the speakers decreased jitter and 50% of the speakers decreased shimmer for the former distinction, while less than 28% of the speakers applied these two cues for the latter distinction.

Variations also existed between the group pattern and the individual pattern for some of the acoustic parameters. For example, while the group pattern showed that Cantonese speakers usually increased the mean amplitude and decreased the shimmer to distinguish sarcasm from neutrality, Speaker 7 produced sarcasm with a significantly lower mean amplitude and higher shimmer value. While the group pattern indicated that sarcastic speech usually had a greater amplitude range and higher HNR value than sincere speech, Speaker 11 produced sarcasm with a significantly narrower amplitude range and lower HNR. Instances could also be found in other speakers' production in terms of all the eight parameters (see Figure 4).

To sum up, in comparison to sincere utterances, Cantonese sarcasm was produced with a slower speech rate, lower mean F0, lower mean amplitude, narrower F0 range, greater amplitude range, and less breathy voice (i.e., higher HNR and lower shimmer). Speech rate, mean F0, mean amplitude, amplitude range, and HNR significantly distinguished sarcasm from other attitudes, regardless of whether the utterances contained the target intensifier. F0 measures were not consistently used for the distinction between sarcasm and neutrality. Also, the acoustic cues may not be all jointly used by the speakers to deliver sarcasm.

3.2 Perception

Figure 5 summarizes the mean scores of the Cantonese listeners' perceptual ratings for the two attitudes across sentence sets (with intensifier vs. without intensifier). Wilcoxon signed rank test comparing the rating scores between the two attitudes (sarcasm vs. sincerity) revealed a significant difference (Wilcoxon $Z = 903$, $p < .001$) with a higher rating score for the recognition of sarcastic utterances (Mean score = 4.02, $SD = 0.34$) than for sincere utterances (Mean score = 2.46, $SD = 0.45$). There was a significant main effect of Attitude, $F(1, 120) = 1235.0626$, $p < .001$, sentence Sets, $F(1, 120) = 101.0122$, $p < .001$, and Gender, $F(1, 40) = 6.1952$, $p = .017$. The rating score of

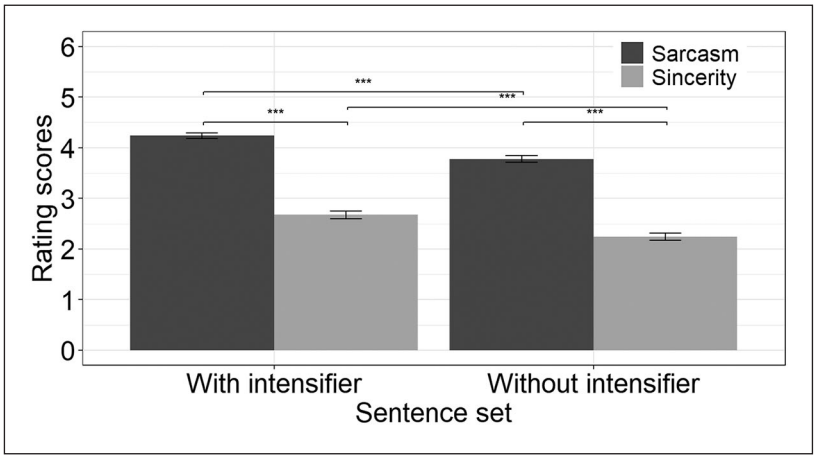

Figure 5. Mean rating scores of the two attitudes across sentence sets.
Note. Error bars indicate the standard errors.

Cantonese sarcasm was significantly higher than that of sincerity (Est.$=-1.649$, $SE=0.084$, $t(120)=-19.675$, $p<.001$). Utterances with the target intensifier were rated with a significantly higher score compared to those without the intensifier (Est.$=-0.540$, $SE=0.084$, $t(120)=-6.447$, $p<.001$), suggesting that the sarcastic and the sincere utterances with the intensifier were respectively perceived as less sincere and more sarcastic than those without the intensifier. Both male and female listeners shared the abovementioned patterns, but the males' ratings were significantly higher than the females' ratings (Est.$=0.149$, $SE=0.129$, $t(40)=1.153$, $p<.001$).

3.2.1 Correlation between the acoustic cues and the perceptual rating. The relationship between acoustic cues and the perceptual rating on sarcastic utterances was examined by investigating (1) the correlation between the number of acoustic cues used by the speaker to utter a sarcastic sentence and the rating by the listeners, and (2) the correlation between the acoustic value of each individual cue and the rating by the listeners.

In terms of the first correlation, whether a speaker utilized an acoustic cue to deliver sarcasm was determined following the method used for the analysis of individual patterns. A simple linear regression model indicated that the number of acoustic cues used by the native speakers significantly predicted how Cantonese listeners would rate the sarcastic utterances, $R^2=0.11$, $F(1, 48)=6.945$, $p=.011$. The rating scores increased as a function of the number of the eight prosodic cues used (Est.$=0.254$, $SE=0.10$, $t(48)=2.635$, $p=.011$), displaying a significantly moderate positive correlation ($r=0.33$, $p=.019$).

Several regression models further explored how individual acoustic measure was associated with the degree of sarcasm perceived by the listeners. Speech rate did not significantly predict the rating score on sarcasm, $F(1, 2098)=0.336$, $p=.562$, while F0 measures were significant predictors for the rating scores (Mean F0: $R^2=0.048$, $F(1, 2098)=106.5$, $p<.001$; F0 range: $R^2=0.032$, $F(1, 2098)=71.28$, $p<.001$). The sarcastic utterances were rated significantly more sarcastic (i.e., higher score) as the mean F0 decreased (Est.$=-0.372$, $SE=0.04$, $t(2098)=-10.32$, $p<.001$) or as the F0 range became narrower (Est.$=-0.286$, $SE=0.03$, $t(2098)=-8.443$, $p<.001$). Besides, amplitude variables also significantly predicted the rating scores (Mean amplitude: $R^2=0.022$, $F(1, 2098)=48$, $p<.001$; Amplitude range: $R^2=0.023$, $F(1, 2098)=49.97$, $p<.001$). The sarcastic

Table 5. Mean rating scores and the number of instances for six types of the combinations of acoustic cues used in the 50 sentences for the perception stimuli (1 = very sincere; 6 = very sarcastic).

Combination of acoustic cues used	Mean rating score (*SE*)	Count
Type 1: Speech rate	2.81 (0.14)	2 (4%)
Type 2: Fundamental frequency + Amplitude + Voice quality	3.41 (0.12)	3 (6%)
Type 3: Speech rate + Amplitude + Voice quality	3.89 (0.09)	7 (14%)
Type 4: Speech rate + Fundamental frequency + Amplitude	4.08 (0.07)	10 (20%)
Type 5: Speech rate + Fundamental frequency + Voice quality	4.65 (0.13)	3 (6%)
Type 6: Speech rate + Fundamental frequency + Amplitude + Voice quality	4.17 (0.05)	25 (50%)

utterances were rated significantly more sarcastic as the mean amplitude decreased (Est. = −0.262, $SE = 0.04$, $t(2098) = -6.928$, $p < .001$) or as the amplitude range enlarged (Est. = 0.224, $SE = 0.03$, $t(2098) = 7.069$, $p < .001$). With regard to the voice quality parameter, HNR, $R^2 = 0.004$, $F(1, 2098) = 9.601$, $p = .002$, significantly predicted the perceptual rating, with an increasing score for sarcastic utterances as the amount of noise reduced (Est. = 0.148, $SE = 0.05$, $t(2098) = 3.099$, $p = .002$). Jitter, $R^2 = 0.007$, $F(1, 2098) = 15.34$, $p < .001$, and shimmer, $R^2 = 0.004$, $F(1, 2098) = 10.02$, $p = .002$, were also significant predictors, displaying that the decreasing jitter (Est. = −0.165, $SE = 0.04$, $t(2098) = -3.916$, $p < .001$) and shimmer values (Est. = −0.097, $SE = 0.03$, $t(2098) = -3.165$, $p = .002$) contributed to an increasing rating score on sarcastic utterances. Model comparisons suggested that the effect of a lower mean F0 for the prediction of perceptual rating on sarcasm was better than other acoustic cues (see Table C1 in Appendix C for the statistical information). Across perceptual correlates, pitch, being index with a lower mean F0 and narrower F0 range, played a more important role in predicting the degree of sarcasm compared to other variables. However, the low adjusted R-squared value of all the individual cues indicated that each individual cue could not explain much variance of the perceptual rating on the degree of sarcasm, suggesting that it may not be easy for the native listeners to clearly understand the sarcastic meanings with only one acoustic cue.

Table 5 summarizes the acoustic cues used by the speakers in the stimuli. The analysis of listener's rating, with the Acoustic cues as fixed effect and Listeners as a random intercept, revealed a significant main effect of Acoustic cues, $F(5, 2053) = 21.813$, $p < .001$. It was difficult for the listeners to identify the sarcastic meaning based on a slower speech rate alone. Also, without either a slower speech rate (i.e., Type 2) or a change in F0 (i.e., Type 3), the rating scores for the sarcastic utterances were lower than the overall mean scores and the rating for the sarcastic sentences with either one of these two cues (i.e., Type 4, Type 5, and Type 6). The combination of speech rate, F0 measures, and voice quality (Type 5) were rated significantly more sarcastic than each of the other types (see Table C2 in Appendix C).[1]

To sum up, a slower speech rate was an important cue differentiating sarcasm from sincerity by Cantonese speakers, but it may not serve as a sufficient cue of sarcasm for the native listeners. Pitch was an important correlate for perceiving Cantonese sarcasm, considering that a lower mean F0 or narrower F0 range could predict a higher degree of sarcastic detection, followed by the change in amplitude variables and the voice quality shifts. Compared to individual cues, it would be easier for the native listeners to identify the implied sarcastic meaning when more acoustic cues were jointly utilized in a sarcastic utterance by native Cantonese speaker.

4 Discussion

The present study examined three prosodic variables, including speech rate, F0 variables, and amplitude variables, as well as voice quality measures, including HNR, jitter, and shimmer, for utterances with sarcastic, sincere, or neutral attitudes in Cantonese. Measurements of each parameter on the key phrases (i.e., the intensifier, the DM, the AP, and the SFP) were done to evaluate how Cantonese speakers change their prosodic features and voice quality when expressing different attitudes. Our findings indicate that Cantonese speakers usually change at least one of the acoustic variables to differentiate their expressions of sarcasm from non-sarcasm. Cantonese listeners can discriminate sarcasm from sincerity based on the acoustic cues alone without semantic contexts. The more acoustic cues were utilized in a sarcastic speech, the easier it would be for the listeners to understand the implied sarcastic meaning.

4.1 Acoustic cues of sarcasm

Compared to sincere utterances, Cantonese sarcasm is conveyed by a slower speech rate, decreases in mean F0 and mean amplitude, a narrower F0 range, a greater amplitude range, a higher HNR, and lower jitter and shimmer.

Speech rate has been reported to be a strong prosodic marker of sarcasm by studies on different languages, and this is also confirmed by our study on Cantonese sarcasm. Most of the Cantonese speakers decreased their speaking rate for different parts across the sarcastic sentence, and such a declination also helped enhance the accuracy rate together with other acoustic cues for Cantonese listeners to recognize sarcastic meanings. Considering that a slower speech rate has been reported to significantly distinguish sarcasm and sincerity in Cantonese in terms of production in a previous study (Cheang & Pell, 2009) as well as the production and perception in our study, this cue can be regarded as a stable and prominent acoustic marker for Cantonese sarcasm. A slower speech rate for sarcastic speech than for sincere speech has also been found in other languages, such as English (i.e., Bryant & Fox Tree, 2002; Cheang & Pell, 2008; Chen & Boves, 2018; Rockwell, 2007), Italian (i.e., Anolli et al., 2002), French (i.e., Loevenbruck et al., 2013), and Mexican Spanish (i.e., Rao, 2013). The purpose of communication contributes to variation in speech rate, and speech rate influences global perception of speaker confidence on persuasion; for example, a slower speech rate is perceived to be more persuasive (Fujihara, 1986; Guyer et al., 2019; Haiman, 1998; Kreuz & Roberts, 1995). When expressing sarcasm, speakers tend to use acoustic cues to deliver intentions which do not align with the literal meanings of the content, and slowing down the speed can help attract listeners' attention to the intended sarcastic meaning. Therefore, it is not surprising that a slower speech rate marks sarcasm across many languages which have been investigated.

Our findings further suggested that amplitude variables can be regarded as a consistent cue of Cantonese sarcasm. For example, over half of the speakers modified the mean amplitude and amplitude range to distinguish sarcasm from other attitudes. Also, the patterns of how the speakers changed their mean amplitude to produce sarcastic sentences were aligned in our study and in Cheang and Pell (2009). The consistency of the amplitude cue in Cantonese sarcasm diverged from its role in English sarcasm in which amplitude was an inconsistent cue based on varied results across studies (e.g., Cheang & Pell, 2008; Rockwell, 2007).

F0 is another variable being regarded as a prominent cue of sarcasm in different languages, although the actual patterns vary; for instance, a lower mean F0 and more restricted F0 range are associated with English sarcasm (i.e., Bryant & Fox Tree, 2002; Cheang & Pell, 2008; Chen & Boves, 2018; Rockwell, 2007) while a higher mean F0 and greater F0 range are associated with

Italian sarcasm (i.e., Anolli et al., 2002). Our findings suggest that a lower mean F0 functions as the secondary cue stably distinguishing sarcasm from non-sarcasm by native Cantonese speakers and serves as a prominent cue enhancing the degree of sarcasm perceived by the native listeners. It is reasonable that F0 measures play a useful role because a change of F0 functions as an important strategy for Cantonese speakers to convey pragmatic and affective states (Chan & To, 2016). Most importantly, our findings contrast with those in Cheang and Pell (2009), which reported that Cantonese sarcastic speech was marked by an increase in mean F0. A possible explanation for the difference may be the improved elicitation method in our study, which lets the speakers respond in a more natural way. Also, our study recruited native Hong Kong Cantonese speakers who were less influenced by the exposure to English intonation compared to the participants in Cheang and Pell (2009) who emigrated to Canada as young adults. Considering the more consistent results for both production and perception and the larger sample size of participants and responses in our study, we believe that a lower mean F0 found in our study is reliable. Unlike the studies on non-tone languages suggesting that speakers were sensitive to F0 measures in distinguishing sarcastic speech from normal speech (i.e., neutrality) and from sincere speech, this cue was mostly used by Cantonese speakers to distinguish sarcasm from sincerity. For example, over 72% of the speakers lowered their mean F0 in sarcastic speech compared to sincere speech while only 33% of the speakers modified the mean F0 for the distinction between sarcasm and neutrality. Furthermore, Cantonese speakers did not use F0 range as a major cue to express sarcasm; for instance, less than 33% of the speakers applied this cue to distinguish sarcasm from non-sarcasm. It has been a research question as to whether or not local pitch movement may affect the application of F0 for expressing or perceiving the global meaning of an attitude in tone languages (Ross et al., 1986). Some early studies reported that pitch in tone languages signals tones directly, and the intonation of an utterance may be restricted by the need to preserve the citation tones within the utterance, which may further reduce the freedom of using F0 for "affective-prosodic signaling" (Connell et al., 1983; Ross et al., 1986). The inconsistency of changing the F0 for distinguishing sarcasm from non-sarcasm in our study may be such an example. Therefore, F0 variable, specifically F0 range, may not be as stable as other prosodic cues (i.e., speech rate and amplitude) in tone languages in the distinction of sarcasm from other attitudes (e.g., neutral speech).

In addition, our findings provide more evidence for the notion that speakers may modulate their voice quality while expressing sarcasm, as reported in previous studies (e.g., Cheang & Pell, 2009; Gobl & Chasaide, 2003; S. Li et al., 2020; Yang, 2021). A higher HNR value stably marks Cantonese sarcasm, distinguishing sarcasm and non-sarcasm by native speakers and enhancing the recognition of the sarcastic meanings by the listeners. Our findings also align with the pattern reported in Cheang and Pell's (2009) study. However, unlike the HNR value, Cantonese speakers did not usually modulate jitter and shimmer to distinguish sarcasm from sincerity. For example, over 50% of the speakers in our study lowered jitter and shimmer to discriminate sarcastic speech from neutral speech, whereas less than 28% of them applied these two cues for the distinction between sarcasm and sincerity. The combination of a higher HNR and a lower jitter found in Cantonese sarcasm also marks Mandarin sarcastic speech (S. Li et al., 2020), displaying a different pattern from English (Cheang & Pell, 2008, 2009; Yang, 2021) and Korean (Yang, 2021), where higher jitter and lower HNR value conveyed sarcastic utterances. Further examination of voice quality, for example, auditory analysis by listeners' ratings, could be conducted for a more comprehensive understanding of how voice quality characterizes sarcasm.

Given the above comparisons, sarcasm is delivered by a combination of different acoustic cues, and the patterns are language specific. The slower speech rate and the decrease in the amount of noise prominently characterize Cantonese sarcasm. The lower mean F0 and lower mean amplitude are the stable cues for the distinction between Cantonese sarcasm and sincerity, while the higher

mean amplitude, higher amplitude range, and lower jitter and shimmer distinguished sarcasm from neutrality. For the universality of acoustic cues, the alteration of the speaking rate may be the most reliable cue for speakers of different languages to deliver sarcasm, and the manipulation of voice quality may also robustly signify the expression of sarcasm, even though the patterns varied across languages.

In addition, our findings demonstrate an interaction between context and prosody. A previous discussion regarding the interplay between context and prosody argued that whether any tone of voice works together with a positive context should be perceived as sincere (Woodland & Voyer, 2011). Our finding suggests that the positive context produced with a sarcastic tone can be perceived as sarcastic instead of sincere, even though the rating for the sarcastic utterances was closer to the mid-range compared to that of sincerity. This may be explained by the opinion in Woodland and Voyer's (2011) study indicating that it is more likely for the listeners to rate the sentences with an incongruent match of context and prosody (e.g., positive context with sarcastic tone) as "neutral" compared to the congruent context and prosody pairing (e.g., positive context with sincere tone). This finding further suggests a significant effect of prosody on the perception of sarcasm. Listeners were able to recognize a negative meaning while hearing a positive content with a sarcastic tone of voice.

Our findings also provide some preliminary insights about the relationship between production and perception of sarcastic speech. First, native Cantonese speakers are used to decrease their speaking rate, mean F0, mean amplitude, and the amount of noise to deliver sarcastic meanings, but these cues may not equally enhance the degree of sarcastic tone perceived by the native listeners. For example, the decreased speaking rate was the prominent acoustic feature marking the production of Cantonese sarcasm, but this cue did not guarantee a higher degree of sarcastic tone perceived by the listeners. The presence of a slower speech rate played an important role in the identification of sarcasm, but it was difficult for native listeners to identify the sarcastic meaning with this cue alone. The effect of the lower mean F0 and mean amplitude aligned between production and perception of Cantonese sarcasm, marking the production of sarcasm and enhancing the degree of sarcastic tone perceived by the listeners. Second, differing from the discrepancy between production and perception with regard to the individual cue, native speakers and listeners consistently expressed or detected sarcastic meanings with a combination of several acoustic cues. The more acoustic cues applied by the speakers to deliver sarcastic feeling, the easier it would be for the listeners to receive the sarcastic meaning. However, since the production and perception data in the present study come from the different participants, it is unknown how the same group of participants would produce and perceive Cantonese sarcasm. More research can be conducted for a deeper understanding about the relationship between production and perception of sarcastic speech.

4.2 Gender difference and individual variability

In addition to the language-specific features, gender differences and individual variation also characterized the acoustic features of sarcasm. Male and female Cantonese speakers in our study rely on different cues to distinguish sarcasm from other attitudes. It is shown that in addition to the slower speech rate, higher HNR value, and larger amplitude range, male speakers relied on the reduction of jitter and shimmer, while female speakers usually lower the mean F0. Such gender difference can be expected since the normal F0 range of an adult female voice is wider than that of an adult male voice, providing more room for female speakers to shift their F0 values to express different attitudes. For male speakers, there might be other cues functioning as compensation for the narrower scope of F0 variation, such as duration reported in British English (Chen & Boves, 2018). The lower jitter and shimmer may also have similar compensation function for male

speakers producing Cantonese sarcasm. Similar gender differences regarding the modification of voice quality were reported by S. Li et al. (2020), in that the HNR was considered as one of the strong predictors of Mandarin sarcasm for female speakers and the jitter predicted sarcasm by male speakers. These findings further support the notion that gender affects how sarcastic cues change within an utterance (Rao, 2013).

Besides, our findings provide evidence for answering the question on whether or not the presentation of sarcasm is regular among speakers (Bryant & Fox Tree, 2005; Cheang & Pell, 2008, 2009, 2011). The speakers in our study freely used different combination of acoustic cues to express sarcasm, and the pattern of the acoustic parameter produced by some speakers did not align with the overall group pattern, yielding a large individual variation. Different degrees of sarcasm rated by the listeners also suggest a free combination of the acoustic cues across utterances, and different cues influenced how well they detect the sarcastic tone. These findings indicate that the strategy for the expression of sarcasm is not regular but speaker-dependent.

4.3 Sentence-level patterns versus phrase-level patterns

The current study has taken a preliminary step to narrow the investigation on how the prosodic parameters change at the sentence level and at the phrase level for Cantonese sarcasm. There was a suggestion that the expression of sarcasm is a global prosodic phenomenon (Cheang & Pell, 2009, 2011); however, our study indicates that it is not necessarily the case in Cantonese. While the pattern of the cue changes between sarcasm and sincerity in terms of speech rate, mean F0, and mean amplitude were consistent for all the key phrases and across the sentence as a whole, the alteration of the F0 range, amplitude range, and the voice quality parameters only appeared in some of the key elements. For example, the SFP was the only phrase carrying the change in F0 range and without a change in amplitude range. Similarly, while speech rate, mean amplitude, amplitude range, jitter, and shimmer displayed a cue-distributed pattern for the distinction between sarcasm and neutrality, other parameters did not change in all the key phrases, or the phrasal pattern did not align with the sentential pattern. For example, mean F0 of all the four key phrases was significantly changed, but only the intensifier and the SFPs displayed the sentential pattern, with a lower mean F0 for sarcasm. The consistency and variation in the phrase-level patterns of amplitude variables and pitch variables, respectively, further explain why pitch may not be a stable cue for distinguishing sarcasm from neutrality in Cantonese, while amplitude is. It is possible that amplitude serves as compensation for the lack of global changes in pitch for sarcasm compared to neutral speech in tone languages.

In addition, regarding the global and local F0 features of attitudinal speech in tone languages, it has been suggested that the local F0 features in the words carrying sentential stress are more contrastive than the overall F0 characteristics in the entire utterance (Gu & Fujisaki, 2016), but this concept was not borne out in the current study. For example, the mean F0 of the sarcastic and neutral utterances without the intensifier were significantly more distinctive for the entire sentences than for the key phrases.

Considering that some of the acoustic cues distributed throughout the whole utterance whereas some cues mainly focused on particular parts of the sentence, Cantonese sarcasm is generally a global phenomenon, but not all cues are equally robust.

4.4 Influences of the insertion of an intensifier

The insertion of an intensifier "*zan55hai22 (really)*" affected the F0 shift in the production of Cantonese sarcasm. The mean F0 difference between sarcasm and non-sarcasm was more

distinctive for the utterances with the target intensifier than for those without the target intensifier, and the F0 range could not significantly distinguish sarcasm from neutrality without the target intensifier. The difference regarding the mean F0 and the F0 range between sarcasm and non-sarcasm was more distinctive for the intensifier than for other key phrases. In addition, the insertion of the target intensifier influenced the listeners' perception of Cantonese sarcasm, making them perceive the sentences as more sarcastic or less sincere. These findings suggest that Cantonese speakers may emphasize the target intensifier to facilitate the distinction between sarcasm and non-sarcasm, contributing to a more exaggerated change in F0, which may assist the listeners to perceive the sarcastic tone more clearly.

The intensifier may also provide a focusing effect with a syntactic cue for expressing criticism, making the sarcastic intentions more direct and indicative for the listeners. However, even though this syntactic cue resulted in a slight increase in the rate for sincere utterance, the ratings were still within the expected range (i.e., 1–3), suggesting that the prosodic cues still take the prominent role, and the syntactic cue may jointly work with the prosodic cues in sarcasm perception.

4.5 Limitations and future work

Data of the present study supports the essential role of prosodic parameters and voice quality parameters in distinguishing sarcasm and non-sarcasm. More research methods (e.g., recording electroglottographic signals, auditory analysis) and more parameters (e.g., contact quotient, peak increase in closure, H1-H2) can be included for further investigations on how voice quality cues signal the production and perception of Cantonese sarcasm. Furthermore, the present study collected data from college-aged Cantonese-speaking participants in Hong Kong. Future investigations can include larger group of participants diverging in age (e.g., children vs. young adults vs. older-aged speakers) or with different language backgrounds (e.g., heritage Cantonese speakers or bilingual speakers, L2-Cantonese or L3-Cantonese speakers). The connection between social factors and how speakers produce sarcastic tone is also worthy of investigation. For example, it is interesting to know whether a speaker's attitude toward sarcasm and whether the speakers having or growing up with different social status may influence how they produce or perceive a sarcastic tone of voice.

In conclusion, sarcastic speech is marked by a combination of acoustic correlates, but the patterns are language-specific and speaker-dependent. Cantonese listeners were able to recognize the sarcastic tone based on the prosodic cues alone. The more prosodic cues being applied by the speakers, the easier it would be for the listeners to understand the implied sarcastic meaning. The insertion of the target intensifier "*zan55hai22 (really)*" functions as an enhancement of the sarcastic intonation in Cantonese. Our study contributes to a more comprehensive understanding about the relationship between prosody and sarcastic speech by using an improved method and extends the investigation on Cantonese sarcasm by including more voice quality parameters. The sentence-level and phrase-level findings provide evidence for the discussion on whether sarcasm is a global or local phenomena, and more studies can be conducted to corroborate our findings.

Acknowledgements

We would like to thank Casper Pak Long Hui for his contribution in the early stages of this project.

Data accessibility statement

The production and perception data analyzed in the current study are publicly available online: https://github.com/sapphirelch/Supplementary-Material-LAS-23-0005.git.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project was supported by the funding of the authors' department.

ORCID iDs

Chen Lan https://orcid.org/0000-0001-8822-3919
Peggy Mok https://orcid.org/0000-0002-9284-6083

Supplemental material

Supplemental material for this article is available online.

Note

1. A possible explanation for the higher score for Type 5 than for Type 6 is that one of the stimuli in Type 5 was rated with a score of 5–6 by all the listeners, resulting in an increase in the total rating score for this type. Several acoustic parameters in this utterance had a larger degree of modification than utterances in Type 6, which might contribute to a higher rating score.

References

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. https://doi.org/10.1080/00207590244000106

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. https://doi.org/10.1515/humr.2003.012

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* [Computer program] (Version 6.4.06). https://www.praat.org

Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In E. N. Goody (Ed.), *Questions and politeness: strategies in social interaction* (pp. 56–311). Cambridge University Press.

Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, *17*(2), 99–119. https://doi.org/10.1207/s15327868ms1702_2

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, *48*(Pt. 3), 257–277. https://doi.org/10.1177/00238309050480030101

Camp, E. (2011). Sarcasm, pretense, and the semantics / pragmatics distinction. *Noûs*, *46*(4), 587–634. https://doi.org/10.1111/j.1468-0068.2010.00822.x

Chan, K. K. L., & To, C. K. S. (2016). Do individuals with high-functioning autism who speak a tone language show intonation deficits? *Journal of Autism and Developmental Disorders*, *46*(5), 1784–1792. https://doi.org/10.1007/s10803-016-2709-5

Chan, M. K. M. (2002). Gender-related use of sentence-final particles in Cantonese. In *Gender across languages: The linguistic representation of women and men*. https://doi.org/10.1075/impact.10.08cha

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, *50*(5), 366–381. https://doi.org/10.1016/j.specom.2007.11.003

Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America*, *126*(3), 1394–1405. https://doi.org/10.1121/1.3177275

Cheang, H. S., & Pell, M. D. (2011). Recognizing sarcasm without language. *Pragmatics & Cognition*, *19*(2), 203–223. https://doi.org/10.1075/pc.19.2.02che

Chen, A., & Boves, L. (2018). What's in a word: Sounding sarcastic in British English. *Journal of the International Phonetic Association*, *48*(1), 57–76. https://doi.org/10.1017/S0025100318000038

Connell, B. A., Hogan, J. T., & Rozsypal, A. J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics*, *11*, 337–351.

Dews, S., Kaplan, J., & Winner, E. (1995). Why not say it directly? The social functions of irony. *Discourse Processes*, *19*(3), 347–367. https://doi-org.easyaccess1.lib.cuhk.edu.hk/10.1080/01638539509544922

Fujihara, T. (1986). Effects of speech rate and hand gesture on attitude change and impression formation. *The Japanese Journal of Psychology*, *57*(4), 200–206. https://doi.org/10.4992/jjpsy.57.200

Fung, R. S.-Y. (2000). *Final particles in standard Cantonese: Semantic extension and pragmatic inference* [Doctoral dissertation, Ohio State University].

Gamer, M., & Lemon, J. (2019). *irr: Various coefficients of interrater reliability and agreement* (R Package Version 0.84.1). https://CRAN.R-project.org/package=irr

Garmendia, J. (2018). Sarcasm and Humour. In *Irony* (pp. 126–146). https://doi.org/10.1017/9781316136218.007

Gobl, C., & Chasaide, A. N. I. I. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, *40*(1–2), 189–212.

Gu, W., & Fujisaki, H. (2016). Data acquisition and prosodic analysis for Mandarin attitudinal speech. In G. Peng & F. Shi (Eds.), *East flows the great river: Festschrift*. (pp. 483–500) City University of Hong Kong Press.

Guyer, J. J., Fabrigar, L. R., & Vaughan-Johnston, T. I. (2019). Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion. *Personality and Social Psychology Bulletin*, *45*(3), 389–405. https://doi.org/10.1177/0146167218787805

Haiman, J. (1998). *Talk is cheap: Sarcasm, alienation, and the evolution of language*. Oxford University Press.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kreuz, R. J. (1996). The use of verbal irony: Cues and constraints. In J. S. Mio & A. N. Katz (Eds.), *Metaphor: Implications and applications* (pp. 23–38). Lawrence Erlbaum.

Kreutz, R. J., & Caucci, G. M. (2007). Lexical influences on the perception of sarcasm. In *Proceedings of the Workshop on Computational Approaches to Figurative Language* (pp. 1–4). Association for Computational Linguistics.

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. https://doi.org/10.1207/s15327868ms1001_3

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). ImerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/JSS.V082.I13

Law, A. (2002). Cantonese sentence-final particles and the CP domain. *UCL Working Papers in Linguistics*, *14*, 375–398.

Lebedeva, E. A. (2021). *The pragmatic role of intensifiers in American media political discourse*. https://doi.org/10.15405/epsbs.2021.12.03.65

Li, J. P. W., Law, T., Lam, G. Y. H., & To, C. K. S. (2013). Role of sentence-final particles and prosody in irony comprehension in Cantonese-speaking children with and without autism spectrum disorders. *Clinical Linguistics and Phonetics*, *27*(1), 18–32. https://doi.org/10.3109/02699206.2012.734893

Li, S., Gu, W., Liu, L., & Tang, P. (2020). The role of voice quality in Mandarin sarcastic speech: An acoustic and electroglottographic study. *Journal of Speech, Language, and Hearing Research*, *63*, 2578–2588. https://doi.org/10.1044/2020_JSLHR-19-00166

Loevenbruck, H., Ameur, M., Jannet, B., D'imperio, M., Spini, M., & Champagne-Lavau, M. (2013). Prosodic cues of sarcastic speech in French: Slower, higher, wider. In *Proceedings of 14th Annual Conference of the International Speech Communication Association* (pp. 3537–3541). International Speech Communication Association (ISCA).

Long, D. L., & Graesser, A. C. (1988). Wit and humor in discourse processing. *Discourse Processes*, *11*(1), 35–60. https://doi-org.easyaccess2.lib.cuhk.edu.hk/10.1080/01638538809544690

Matthews, S., & Yip, V. (1994). *Cantonese: A comprehensive grammar*. Routledge. https://doi.org/10.4324/9780203835012

Myers, G. (1990). The rhetoric of irony in academic writing. *Written Communication*, *7*(4), 419–455. https://doi.org/10.1177/0741088390007004001

Niebuhr, O. (2014). "A little more ironic"—Voice quality and segmental reduction differences between sarcastic and neutral utterances. In *Proceedings of 7th International Conference of Speech Prosody* (pp. 608–612). https://doi.org/10.13140/2.1.5088.0006

Rao, R. (2013). Prosodic consequences of sarcasm versus sincerity. *Concentric: Studies in Linguistics*, *2*, 33–59.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, *29*, 483–495. https://doi.org/10.1023/A:1005120109296

Rockwell, P. (2001). Facial expression and sarcasm. *Perceptual and Motor Skills*, *93*(1), 47–50. https://doi.org/10.2466/pms.2001.93.1.47

Rockwell, P. (2007). Vocal features of conversational sarcasm: A comparison of methods. *Journal of Psycholinguistic Research*, *36*(5), 361–369. https://doi.org/10.1007/s10936-006-9049-0

Ross, E. D., Edmondson, J. A., & 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. *Journal of Phonetics*, *14*(2), 283–302. https://doi.org/10.1016/s0095-4470(19)30669-2

Scherer, K. R., Banse, R., Wallbott, H. G., Rimé, B., Kappas, A., Manstead, A., & Ricci-Bitti, P. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Vocal Emotion Expression Journal of Cross-Cultural Psychology*, *32*(1), 76–92.

Tabacaru, S. (2020). Faces of sarcasm. Exploring raised eyebrows with sarcasm in French political debates. In A. Athanasiadou & H. Colston (Ed.), *The diversity of irony* (pp. 256–277). De Gruyter Mouton. https://doi.org/10.1515/9783110652246-012

Tang, P., & Gu, W. (2015). Perceptual experiment and acoustic analysis of Chinese attitudes: A preliminary study. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–5). The University of Glasgow.

Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony. *Journal of Pragmatics*, *32*(12), 1777–1806. https://doi.org/10.1016/S0378-2166(99)00116-2

Woodland, J., & Voyer, D. (2011). Context and intonation in the perception of sarcasm. *Metaphor and Symbol*, *26*, 227–239.

Xu, Y. (2013). ProsodyPro — A tool for large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* (pp. 7–10). Aix-en-Provene.

Yang, S. Y. (2021). Listener's ratings and acoustic analyses of voice qualities associated with English and Korean sarcastic utterances. *Speech Communication*, *129*, 1–6. https://doi.org/10.1016/j.specom.2021.02.002

Appendix A

Table A1. The target sentences and the scenarios with negative and positive situations.

Sentence	Set	Target sentences	Negative scenarios	Positive scenarios
1	a	你煮野食好叻呀。You cook very well.	我不嬲都冇煮開嘢食，琴日試下煮咗碟中式牛柳俾我呀媽食，點知啲牛肉煮得太耐，我同佢都咬唔開。I seldom cooked before. I cooked a beef fillet for my mother yesterday, but it was too chewy.	我琴日煮咗碟中式牛柳俾我呀媽食，佢話好好食！I cooked a beef fillet for my mother yesterday, and she said it was tasty.
	b	你煮野食真係好叻呀。You cook really (very) well.		
2	a	你好醒呀！You are very smart.	下？乜琴日reg科咩？我以為係今日reg添！What? Was yesterday the deadline for course registration? I thought it would be due today.	出面落緊雨，以你性格一定冇帶遮，所以我帶多咗一把比你用。It's raining. I know you have not taken you umbrella with you, so I bring one for you.
	b	你真係好醒呀！You are really (very) smart.		
3	a	呢套戲聽落幾精彩喎。This movie sounds very amazing.	呢套戲呀？都唔知個故事係到講乜，淨係見到有兩個人係咁結婚離婚咗十萬次。I don't know what the story of this movie was telling about. I only found two people getting married and divorced for one hundred thousand times.	呢套戲好睇呀！個故事又好，啲演員又正，一定要去戲院睇囉。This movie was great! The story was marvelous, and the actors performed well. You must go to the theater to watch this movie.
	b	呢套戲聽落真係幾精彩喎。This movie sounds really (very) amazing.		
4	a	場波幾好睇呀。This game of football was very good.	琴日果場波呀？車路士對利物浦囉，兩邊淨係係度猜來猜去，射門都無個，咪0比0囉。The game of football yesterday was Chelsea versus Liverpool. Two teams only passed the ball without attempting on goal, so it ended zero-zero.	琴日果場波，車路士大戰利物浦，雖然最尾一比一，但係兩邊有攻有守，質素好高。Even though the game of football yesterday ended one-one, both Chelsea and Liverpool football teams had high-quality offense and defense.
	b	場波真係幾好睇呀。This game of football was really (very) good.		
5	a	你睇書睇得幾快喎。You read books very quickly.	呢本書我買咗一年，淨係睇咗一個 chapter 咁多。I've bought this book for one year, but I've only read one chapter.	呢本書我用三日就睇晒啦。It took me only three days to finish reading this book.
	b	你睇書真係睇得幾快喎。You read books really (very) quickly.		

(continued)

Table A1. (continued)

Sentence	Set	Target sentences	Negative scenarios	Positive scenarios
6	a	你都幾大食喎。You have a very big appetite.	我通常食半碗飯就飽架喇。I'm usually full after I've eaten half bowl of rice.	我係屋企食飯座底都三碗飯啦！I usually eat three bowls of rice at home.
	b	你都真係幾大食喎。You have a really (very) big appetite.		
7	a	幾香喎。It smells very nice.	嘩！隔離街爆屎渠，隔咗幾條街都聞到！Waa! There is a toilet pipe on the nearby street bursting. The smell is spreading across streets!	你聞下我用個款香水香唔香？Does my perfume smell nice?
	b	真係幾香喎。It smells really (very) nice.		
8	a	好平喎！It's very cheap.	呢本書你咪睇佢薄 citcit 咁呀，成兩百蚊架！Even though this book is very thin, it sells for 200 HKD.	呢本書我用 20 蚊就買到啦！I spent only 20 HKD to buy this book.
	b	真係好平喎！It's really (very) cheap.		
9	a	你都幾有耐性喎。You have a lot of patience.	呢間餐廳有無搞錯呀，成五分鐘都未上菜，五分鐘夠我食晏啦，真係嘥晒我哋時間。What's wrong with this restaurant? They've still not served the dishes while five minutes have passed away. It really wastes my time.	頭先教阿妹一題數教咗幾次，講左成粒鐘先明，都好，總算冇白費心機。I spent one hour teaching my younger sister a math question for several times. My effort was not in vain since she finally made it.
	b	你都真係幾有耐性喎。You really have a lot of patience.		
10	a	佢好大力喎。He is very strong.	你知唔知呀，佢拎個半磅嘅蛋糕都拎到手震呀！Carrying a half-pound cake made his hand trembled.	今日我細佬幫阿媽抬咗包五十公斤嘅米返屋企呀。Today my younger brother helped our mother to carry 50 kilograms of rice back home.
	b	佢真係好大力喎。He is really (very) strong.		
11	a	佢好乖喎。He behaves very well.	頭先搭地鐵個陣有個細路係咁係度周圍走呀，猛咁撞到其他人。When I took the subway, I saw a boy walking around and running into other people.	頭先搭地鐵個陣有個五歲嘅細路讓座比個婆婆坐呀。When I took the subway, I saw a 5 year-old boy offered his seat to a granny.
	b	佢真係好乖喎。He behaves really (very) well.		
12	a	佢跑得好快喎。He runs very quickly.	佢用咗成分鐘都跑唔晒一百米呀！He can't run across 100 meters in 1 minute.	佢跑一百米十秒都唔洗呀！He spent less than 10 seconds running across 100 meters!
	b	佢真係跑得好快喎。He runs really (very) quickly.		

Appendix B

Statistical information for the phrase-level analysis

Table B1. Results of the linear mixed effects models on eight prosodic parameters for the target intensifier, degree modifier (DM), adjectival phrase (AP), and the sentence-final particle (SFP) in the utterances without and with the intensifier.

Parameters	Phrase	Predictor	Utterances without the intensifier				Utterances with the intensifier			
			Estimate	*SE*	*t* value	Pr(>|*t*|)	Estimate	*SE*	*t* value	Pr(>|*t*|)
Speech rate	Intensifier	(Intercept)					0.055	0.085	0.641	.532
		Neutrality					−0.572	0.048	−11.909	**<.001**
		Sincerity					0.399	0.048	8.305	**<.001**
	DM	(Intercept)	−0.555	0.151	−3.687	.003	−0.597	0.131	−4.565	.0005
		Neutrality	0.858	0.046	18.711	**<.001**	0.800	0.039	20.445	**<.001**
		Sincerity	0.942	0.046	20.539	**<.001**	0.830	0.039	21.194	**<.001**
	AP	(Intercept)	−0.470	0.247	−1.90	.084	−0.398	0.243	−1.636	.13
		Neutrality	0.650	0.026	25.12	**<.001**	0.781	0.026	30.202	**<.001**
		Sincerity	0.584	0.026	22.55	**<.001**	0.575	0.026	22.237	**<.001**
	SFP	(Intercept)	−0.833	0.075	−11.11	<.001	−0.705	0.075	−9.394	<.001
		Neutrality	1.516	0.035	43.33	**<.001**	1.801	0.037	48.878	**<.001**
		Sincerity	0.608	0.035	17.39	**<.001**	0.662	0.037	17.969	**<.001**
Mean F0	Intensifier	(Intercept)					−0.408	0.131	−3.103	.009
		Neutrality					0.211	0.043	4.936	**<.001**
		Sincerity					0.996	0.043	23.350	**<.001**
	DM	(Intercept)	0.022	0.157	0.140	.890	−0.407	0.090	−4.509	.0005
		Neutrality	0.094	0.040	2.353	**.019**	−0.096	0.035	−2.694	**.007**
		Sincerity	0.594	0.040	14.809	**<.001**	0.440	0.035	12.396	**<.001**
	AP	(Intercept)	−0.056	0.126	−0.445	.664	−0.231	0.095	−2.434	.030
		Neutrality	−0.024	0.041	−0.575	.566	−0.143	0.045	−3.187	**.001**
		Sincerity	0.536	0.041	13.042	**<.001**	0.362	0.045	8.088	**<.001**
	SFP	(Intercept)	−0.050	0.060	−0.827	.420	−0.105	0.050	−2.124	.044

(Continued)

Table B1. (continued)

Parameters	Phrase	Predictor	Utterances without the intensifier				Utterances with the intensifier			
			Estimate	*SE*	*t* value	Pr(>|*t*|)	Estimate	*SE*	*t* value	Pr(>|*t*|)
		Neutrality	−0.037	0.047	−0.780	.436	−0.195	0.050	−3.887	**.0001**
		Sincerity	0.374	0.047	7.960	<**.001**	0.230	0.050	4.591	<**.001**
F0 range	Intensifier	(Intercept)					−0.253	0.128	−1.981	.070
		Neutrality					0.342	0.047	7.325	<**.001**
		Sincerity					0.380	0.047	8.137	<**.001**
	DM	(Intercept)	0.164	0.084	1.967	.069	−0.046	0.066	−0.690	.501
		Neutrality	−0.241	0.043	−5.657	<**.001**	−0.196	0.038	−5.171	<**.001**
		Sincerity	−0.099	0.043	−2.317	**.021**	−0.060	0.038	−1.568	.117
	AP	(Intercept)	−0.003	0.097	−0.030	.977	−0.033	0.101	−0.321	.753
		Neutrality	−0.031	0.044	−0.704	.482	−0.078	0.046	−1.684	.092
		Sincerity	0.052	0.044	1.189	.235	−0.005	0.046	−0.105	.916
	SFP	(Intercept)	−0.020	0.079	−0.259	.799	0.051	0.076	0.679	.508
		Neutrality	−0.144	0.047	−3.038	**.002**	−0.319	0.048	−6.668	<**.001**
		Sincerity	0.231	0.047	4.877	<**.001**	0.021	0.048	0.444	.657
Mean amplitude	Intensifier	(Intercept)					0.229	0.088	2.613	.021
		Neutrality					−0.903	0.046	−19.645	<**.001**
		Sincerity					0.215	0.046	4.686	<**.001**
	DM	(Intercept)	0.224	0.139	1.611	.133	−0.101	0.110	−0.918	.376
		Neutrality	−0.443	0.045	−9.813	<**.001**	−0.690	0.043	−15.945	<**.001**
		Sincerity	0.325	0.045	7.192	<**.001**	0.444	0.043	10.256	<**.001**
	AP	(Intercept)	0.263	0.145	1.816	.095	0.063	0.144	0.440	.668
		Neutrality	−0.625	0.040	−15.543	<**.001**	−0.920	0.041	−21.961	<**.001**
		Sincerity	0.315	0.040	7.839	<**.001**	0.252	0.041	6.019	<**.001**

(Continued)

Table B1. (continued)

Parameters	Phrase	Predictor	Utterances without the intensifier				Utterances with the intensifier			
			Estimate	*SE*	*t* value	Pr(>|*t*|)	Estimate	*SE*	*t* value	Pr(>|*t*|)
	SFP	(Intercept)	0.180	0.097	1.862	.085	−0.084	0.071	−1.179	.257
		Neutrality	−0.490	0.044	−11.006	<**.001**	−0.743	0.048	−15.618	<**.001**
		Sincerity	0.513	0.044	11.535	<**.001**	0.432	0.048	9.071	<**.001**
Amplitude range	Intensifier	(Intercept)					0.145	0.101	1.433	.175
		Neutrality					−0.315	0.051	−6.162	<**.001**
		Sincerity					−0.133	0.051	−2.594	**.009**
	DM	(Intercept)	0.281	0.190	1.483	.165	−0.009	0.121	−0.071	.945
		Neutrality	−0.283	0.046	−5.110	<**.001**	−0.238	0.044	−5.421	<**.001**
		Sincerity	−0.199	0.046	−4.276	<**.001**	−0.176	0.044	−3.996	<**.001**
	AP	(Intercept)	0.211	0.171	1.235	.241	0.110	0.173	0.635	.538
		Neutrality	−0.270	0.046	−5.925	<**.001**	−0.313	0.045	−7.018	<**.001**
		Sincerity	−0.206	0.046	−4.514	<**.001**	−0.179	0.045	−4.010	<**.001**
	SFP	(Intercept)	0.248	0.097	2.547	.024	0.153	0.087	1.759	.100
		Neutrality	−0.520	0.051	−10.286	<**.001**	−0.614	0.051	−12.046	<**.001**
		Sincerity	−0.011	0.051	−0.208	.835	−0.060	0.051	−1.182	.237
HNR	Intensifier	(Intercept)					−0.050	0.103	−0.487	.635
		Neutrality					0.147	0.052	2.824	**.005**
		Sincerity					0.004	0.052	0.085	.932
	DM	(Intercept)	0.233	0.210	1.110	.29	0.061	0.190	0.322	.753
		Neutrality	−0.242	0.042	−5.763	<**.001**	−0.319	0.040	−8.048	<**.001**
		Sincerity	−0.207	0.042	−4.943	<**.001**	−0.110	0.040	−2.770	**.006**
	AP	(Intercept)	0.248	0.244	1.016	.331	0.106	0.238	0.443	.666
		Neutrality	−0.445	0.031	−14.540	<**.001**	−0.414	0.033	−12.687	<**.001**
		Sincerity	−0.109	0.031	−3.558	**.0004**	−0.093	0.033	−2.839	**.005**

(Continued)

Table B1. (continued)

Parameters	Phrase	Predictor	Utterances without the intensifier				Utterances with the intensifier			
			Estimate	*SE*	*t* value	Pr(>|*t*|)	Estimate	*SE*	*t* value	Pr(>|*t*|)
	SFP	(Intercept)	0.534	0.099	5.352	.0001	0.247	0.101	2.446	.029
		Neutrality	−0.873	0.046	−18.910	**<.001**	−0.953	0.048	−19.702	**<.001**
		Sincerity	−0.333	0.046	−7.212	**<.001**	−0.182	0.048	−3.756	**.0002**
Jitter	Intensifier	(Intercept)					−0.096	0.090	−1.070	.303
		Neutrality					0.197	0.051	3.887	**.0001**
		Sincerity					0.059	0.051	1.166	.244
	DM	(Intercept)	−0.155	0.082	−1.899	.075	−0.077	0.086	−0.901	.382
		Neutrality	0.175	0.045	3.877	**.0001**	0.348	0.048	7.230	**<.001**
		Sincerity	0.099	0.045	2.190	**.029**	−0.062	0.048	−1.295	.196
	AP	(Intercept)	−0.212	0.113	−1.880	.084	−0.035	0.132	−0.264	.796
		Neutrality	0.350	0.042	8.312	**<.001**	0.322	0.049	6.618	**<.001**
		Sincerity	0.020	0.042	0.479	.632	−0.064	0.049	−1.312	.190
	SFP	(Intercept)	−0.190	0.065	−2.881	.012	−0.016	0.074	−0.211	.835
		Neutrality	0.267	0.042	6.368	**<.001**	0.389	0.052	7.541	**<.001**
		Sincerity	−0.061	0.042	−1.444	.149	−0.112	0.052	−2.173	**.029**
Shimmer	Intensifier	(Intercept)					−0.108	0.051	−2.137	.042
		Neutrality					0.332	0.051	6.438	**<.001**
		Sincerity					−0.045	0.051	−0.878	.380
	DM	(Intercept)	−0.221	0.062	−3.549	.002	−0.048	0.061	−0.778	.445
		Neutrality	0.208	0.048	4.368	**<.001**	0.291	0.052	5.630	**<.001**
		Sincerity	0.140	0.048	2.946	**.003**	0.053	0.052	1.033	.302
	AP	(Intercept)	−0.167	0.126	−1.330	.209	−0.064	0.139	−0.459	.654
		Neutrality	0.239	0.041	5.882	**<.001**	0.294	0.048	6.098	**<.001**
		Sincerity	0.007	0.041	0.162	.871	0.045	0.048	0.930	.352
	SFP	(Intercept)	−0.291	0.071	−4.102	.001	−0.062	0.067	−0.934	.363
		Neutrality	0.473	0.043	10.885	**<.001**	0.462	0.051	9.058	**<.001**
		Sincerity	0.042	0.043	0.962	.336	−0.035	0.051	−0.691	.490

The intercept refers to *sarcasm* produced by the speakers.

Appendix C

Statistical information for the analysis on the perception data

Table C1. Summary of the adjusted R-squared, sigma, AIC, BIC, and the *p* value of the regression models on perceptual rating scores on sarcastic utterances as a function of each acoustic parameter.

	adjusted R^2	sigma	AIC	BIC	*p* value
Speech rate	−0.0003	1.54	7,763	7,780	.562
Mean F0	0.048	1.50	7,660	7,677	<.001
F0 range	0.032	1.51	7,694	7,711	<.001
Mean amplitude	0.022	1.52	7,716	7,733	<.001
Amplitude range	0.023	1.52	7,714	7,731	<.001
HNR	0.004	1.53	7,754	7,771	.002
Jitter	0.007	1.53	7,749	7,765	<.001
Shimmer	0.004	1.53	7,754	7,771	.002

Table C2. Summary of the results for the comparison of the perceptual rating scores on the utterances with different combinations of acoustic cues (*df* = 2053).

	Estimate	*SE*	t value	Pr (>|*t*|)
(intercept)	4.651	0.14	33.735	<.001
Type: Type 1 Speech rate	−1.841	0.21	−8.860	<.001
Type: Type 2 F0 + Amplitude + Voice quality	−1.238	0.19	−6.661	<.001
Type: Type 3 Speech rate + Amplitude + Voice quality	−0.766	0.16	−4.879	<.001
Type: Type 4 Speech rate + F0 + Amplitude	−0.574	0.15	−3.783	<.001
Type: Type 6 Speech rate + F0 + Amplitude + Voice quality	−0.530	0.14	−3.818	<.001

The intercept refers to *Type 5: Speech rate + Fundamental frequency + Voice quality*.