

Revisiting Acoustic Markers of Sarcasm in Cantonese

Chen Lan, Pak Long Hui, Wenwei Xu, Peggy Mok

The Chinese University of Hong Kong, Hong Kong

lchsapphire@gmail.com, casphui@gmail.com, 1155061931@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

ABSTRACT

The present study investigated the acoustic cues of sarcasm in Cantonese. Ten native Hong Kong Cantonese speakers were elicited by pictures and audios to produce target utterances with three attitudes: neutrality, sincerity, and sarcasm. Six prosodic features were measured (speech rate, mean F0, F0 range, mean amplitude, amplitude range, and harmonics-to-noise ratio HNR). Results of the acoustic analysis indicated that a reduction in the speech rate, an enlargement of the amplitude-range, and an increase in the HNR distinguished sarcasm from the other attitudes. Relative to sincerity in particular, sarcasm in Cantonese was also marked by lower mean F0, smaller F0-range, and lower mean amplitude. These findings of mean F0 and amplitude-range differed from the results of previous studies. Gender differences were observed in this study as well. Between sarcasm and sincerity, females had smaller speech rate differences and larger mean F0 differences than males.

Keywords: sarcasm, acoustic cues, prosodic features, Cantonese

1. INTRODUCTION

Verbal irony has generally been described as a rhetorical device for either implying the opposite of what the content is literally [3], or expressing a different meaning from what is said [13]. Ironic criticisms, using positive contents to deliver negative meanings, and ironic compliments, which making use of negative contents to give positive comments, were two types of irony [12]. In this study, we investigated and discussed the acoustic features of the former one, which was generally referred to as sarcasm. Previous studies on sarcasm suggested that prosodic properties such as pitch, duration, and amplitude were essential verbal cues to distinguish sarcasm and non-sarcasm. However, the patterns of these cues were varied across languages. For example, English sarcastic utterances were marked by a lower pitch and a slower speech rate [4, 6, 8, 15], while sarcasm in Italian was produced with a higher pitch level, a slower speech rate, and greater amplitude [1]. Cantonese is spoken in the community with different “communication style” compared to the aforementioned languages [7].

Previous research on Cantonese sarcasm often focused on the syntactic structures, and little of them studied the prosody system. This study investigates the acoustic markers of sarcasm in Cantonese, aiming to identify the prosodic features of Cantonese sarcasm and compare the findings with the patterns of other languages.

1.1. Sarcasm in Cantonese

A previous study on Cantonese sarcasm [7] investigated six prosodic variables, displaying a pattern involving the combination of a higher mean fundamental frequency (F0), a narrower F0 range, a slower speech rate, and a more restricted amplitude range. In addition, Harmonic to Noise Ratio (HNR) was found as a significant cue in differentiating sarcasm from humorous utterances, as the average HNR values of sarcastic phrases were lower than that of humorous phrases. However, their non-colloquial materials (including both the target utterances and biasing sentences) posed a limitation for the above study, rendering their findings as tentative. This study revisits this issue with more rigorous methods.

1.2. Gender differences

Gender differences on sarcasm were examined in Mexican Spanish and English, 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 their pitch levels [8], but for Mexican Spanish, the effect of pitch variables on attitude was stronger for males than females [14]. Gender differences in sarcasm had not been studied in Cantonese. Hence, this study also investigated this factor.

2. METHOD

2.1. Participants

Ten native Hong Kong Cantonese speakers (five females and five males) who were undergraduate students at a university of Hong Kong participated in the production task (with a mean age of 21 years old). According to their language background questionnaires, all of the participants were born in Hong Kong with parents being native Hong Kong Cantonese speakers, and went to local primary and

secondary school. Cantonese was the most often used language in their daily communication with a mean percentage of 85.8%. The participants were paid to attend the experiment and reported with no speech, hearing, or study problems.

2.2. Materials

Two sets of simple sentences commonly used in Cantonese were designed for this study. The first set contained the target utterances with a degree modifier, an adjectival phrase, and a sentence final particle. Cantonese sentence final particles were frequently used to convey different attitudes in speech [10]. The final particle /a˩/ or /wo˩/ was chosen for more natural utterances. Furthermore, in Cantonese, the intensifier /tsɛn˩hɛi˩/ ‘really’ was used frequently for expressing criticism as well as for assuring sincerity [9], working naturally for both sarcasm and sincerity. As a result, this intensifier was used to create the second set of target utterances, aiming to examine whether results varied with this intensifier. To summarize, 12 sets of target sentences as exemplified in Table 1 were produced in three attitudes (neutral, sincere, sarcastic) with three repetitions. In total, each participant produced 216 target utterances.

To naturally elicit the emotional expressions, this study applied scenario approach [17], providing the participants short scenarios with positive or negative situations commonly happened in daily life (see Table 1). These contexts were presented using audios recorded by two native Cantonese speakers and pictures describing the contexts. For neutral speech, no biasing sentences were provided. Instead, the participants received an instruction that they only need to read out the sentences displayed on the screen.

Table 1: Example of the contexts (1. negative; 2. positive) and the target utterances with English translations (a. sentence without intensifier; b. sentence with a target intensifier).

Biasing Utterances	
1	What? Was yesterday the deadline for course registration? I thought it would be due today.
2	It’s raining. I know you have not taken your umbrella with you, so I bring one for you.
Target Utterances	
a	你好醒呀！ You are so smart!
b	你真係好醒呀！ You are really (so) smart!

2.3. Procedure

All participants individually completed a production task in a soundproof recording room. A solid-state recorder with the sampling rate of 44100 Hz was used for the recording. During the experiment, no

definition of sarcasm was provided. In the first part of the experiment, target utterances were presented using PowerPoint, and each slide contained one sentence. No biasing contexts were given. The participants were instructed to read the displayed sentences one by one neutrally. In the second part, a picture and a target utterance were presented on each slide, and a biasing sentence was played automatically. The participants were required to listen to the audio first, and then produce the target utterance according to the context provided by the audio and the picture. Target utterances were randomized and shown up on the screen in different orders in each repetition.

2.4. Acoustic analyses

2160 utterances (12 target utterances × 3 attitudes × 2 sentence sets × 3 repetitions × 10 participants) were measured in Praat [2] using ProsodyPro [18]. Following [7], speech rate, mean F0, F0 range, mean amplitude, amplitude range, and the HNR were measured for each utterance as a whole. The number of syllables and the total duration of each utterance were measured, and the speech rate was calculated by dividing the number of syllables by the length of each utterance. For the pitch variables, mean F0, minimum F0, and maximum F0 were measured in Hertz (Hz), and F0 range was provided by subtracting the minimum F0 from the maximum F0. The F0 contours were time-normalized, and each syllable was equally divided into 10 points. F0 value at each point was extracted. Regarding the amplitude variables, mean intensity, minimum intensity, maximum intensity, and the HNR were measured in decibel (dB), and amplitude range was provided by subtracting the minimum intensity from the maximum intensity.

2.5. Statistical analyses

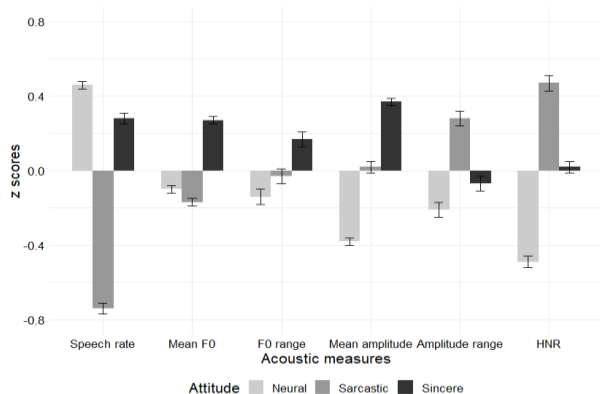
All data were converted into z-scores. Two-way ANOVAs with repeated measures were conducted for each variable considering two factors: Attitude (sarcasm, sincerity, and neutrality) and the sentence Set (utterances with and without the target intensifier). Furthermore, for the analyses on gender differences, one-way ANOVAs with repeated measures were conducted to compare the acoustic measures. Additionally, paired-sample t tests further compared the variables of sarcasm to that of neutrality and sincerity within each gender. Sentence sets were omitted in these comparisons.

3. RESULTS

Fig. 1 summarises the average normalized values of the six acoustic variables, including the speech rate,

mean F0, F0 range, mean amplitude, amplitude range, and the HNR.

Figure 1: Mean values (z-scores) of the six acoustic variables across three attitudes. Error bars indicate the standard errors.

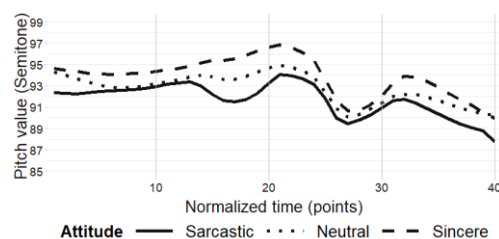


3.1. Speech rate and F0 measures

A significant interaction between Attitude and Set was found ($F(2,700) = 18.741, p < .001$), together with main effects for the two factors: Attitude ($F(2,700) = 1899.921, p < .001$); Set ($F(1,350) = 715.135, p < .001$). Results indicated that sarcasm was expressed significantly more slowly than sincerity and neutrality, and neutrality was produced with a significantly faster speech rate than the other attitudes.

Fig. 2 provides an example of the averaged time-normalized pitch contours of a target utterance produced by the female participants in three attitudes, revealing a flattened pitch in the sarcastic production. Compared with the other attitudes, sarcasm produced an F0 valley in a particular syllable, as shown in the second syllable – the degree modifier, in the exemplified sentence. Similar pattern was found regarding the degree modifiers in other target utterances produced by all the participants. The analysis of the mean F0 revealed a significant interaction between Attitude and Set ($F(2,700) = 4.674, p = .010$), together with a main effect for Attitude ($F(2,700) = 172.511, p < .001$). The post hoc comparison suggested that Cantonese participants produced sarcasm with a significantly lower mean F0 than neutrality and sincerity, and neutrality with a significantly lower mean F0 than sincerity. Regarding the F0 range, significant main effects were found for Attitude ($F(2,700) = 26.416, p < .001$) and Set ($F(1,350) = 8.512, p = .004$). Sarcasm was produced with a significantly smaller F0 range than sincerity but with a significantly greater F0 range than neutrality. In addition, the F0 range of sincerity was significantly greater than that of sarcasm and neutrality.

Figure 2: Average time-normalized pitch contours of 你好醒呀 ('You are so smart') in three attitudes produced by the female speakers. Vertical lines indicate syllable boundaries.



3.2. Amplitude measures

The analysis of the mean amplitude provided a significant interaction between Attitude and Set ($F(2,700) = 4.241, p = .015$), together with significant main effects for Attitude ($F(2,700) = 450.587, p < .001$) and Set ($F(1,350) = 22.173, p < .001$). Results further illustrated that the mean amplitude of neutrality was significantly lower than that of sarcasm and sincerity. Also, sarcasm was produced with less energy compared to sincerity. With respect to the amplitude range, significant main effects for Attitude ($F(2,700) = 106.764, p < .001$) and Set ($F(1,350) = 44.660, p < .001$) were found. Post hoc tests suggested that the participants performed a significantly greater amplitude range in sarcastic speech than in the speech of other attitudes. In addition, sincerity was expressed with a significantly narrower amplitude range than neutrality. The analysis of the HNR revealed a significant interaction between Attitude and Set ($F(2,700) = 14.499, p < .001$), and main effects for Attitude ($F(2,700) = 395.206, p < .001$) and Set ($F(1,350) = 111.430, p < .001$). The post hoc comparison displayed that the HNR value of sarcastic utterances was significantly higher than that of sincerity and neutrality. Neutrality was conveyed with a significantly lower HNR than sincerity.

3.3. Gender differences

Analyses of the speech rate, mean F0, amplitude range, and the HNR found significant interactions between Attitude and Gender. Results of the paired-sample t tests suggested that sarcastic utterances were produced with a significantly slower speech rate than the sincere and neutral utterances by both male and female speakers, but the durational difference was significantly larger in male's speech ($t(1,359) = -7.88, p < .001$, see Fig. 3). Regarding the pitch variables (see Fig. 4), mean F0 was found to significantly discriminate sarcasm from sincerity for both males and females. However, between sarcasm and neutrality, a significant difference was found only for female speakers ($t(1,359) = -7.33, p < .001$). In

addition, the mean F0 difference between sarcasm and sincerity by female speakers was significantly larger than those by males ($t(1,359) = 5.17, p < .001$). F0 range of sarcastic speech was significantly narrower than that of sincere speech by both males and females. However, sarcasm was found with significantly smaller F0 range than neutrality by female speakers only ($t(1,359) = 2.83, p = .005$). The amplitude variables were found to significantly distinguish sarcasm from the other attitudes for both male and female speakers, but the difference between the HNR of sarcastic utterances and that of sincere utterances was significantly greater in female's production ($t(1,359) = 2.91, p = .004$).

Figure 3: Average speech rate (z-scores) of the target utterances across three attitudes by female and male participants.

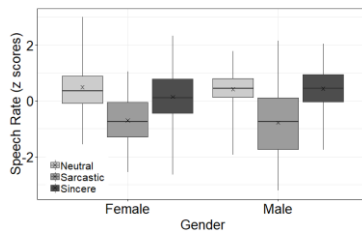
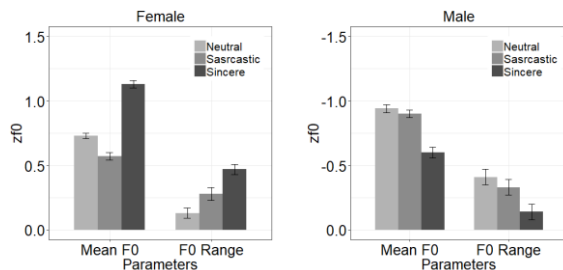


Figure 4: Average z-transformed F0 and F0 range of the target utterances across three attitudes produced by female (left) and male (right) participants. Error bars indicate the standard errors. Mirror image (right) is used for the comparison purpose.



4. DISCUSSION

This study analysed six acoustic parameters for utterances with sarcastic, sincere, or neutral attitude in Cantonese, and compared each of the parameters across attitudes by gender.

Overall, a slower speech rate, a greater amplitude range, and a higher HNR were reported as the most consistent markers of sarcasm in Cantonese, since they significantly distinguished sarcasm with all the other attitudes within each sentence set. For speech rate of sarcasm, our findings shared the same pattern with that of English [4, 6, 8, 15], Italian [8], French [11], and Mexican Spanish [14]. However, sincerity and neutrality was found to have the fastest speech rate respectively in the previous study on Cantonese [7] and in the current study. In terms of amplitude

range, [7] indicated that sarcasm was produced with a narrower amplitude range, which was contrary to our findings.

In addition, a reduction of mean F0 and mean amplitude as well as a restriction of F0 range also differentiated sarcasm from sincerity. It is reasonable that pitch variables are the significant cues since a change of pitch functions as an important strategy for Cantonese speakers to convey pragmatic and affective states [5]. Across languages, our findings of F0 measures concur with those of English and Mexican Spanish [4, 6, 8, 14, 15] but differ from those of Italian and French [1, 11]. Most importantly, our findings are in contrast to those in [7] saying that Cantonese sarcasm was marked by a rise in pitch. Possible explanations for the difference can be the improved elicitation method in our study which let the participants respond in a more natural way. Considering the larger size of the responses and also more participants in our study than [7], we believe that the patterns found in the current study are reliable. Despite the contrary finding on mean F0, the two studies agreed on the pattern of F0 range.

Furthermore, amplitude variables were reported as an inconsistent cue of sarcasm in English since results varied in different studies [6, 16]. However, for sarcasm in Cantonese, our findings about amplitude were consistent with [7], showing the lowest mean amplitude for neutrality and a lower mean amplitude of sarcasm compared to that of sincerity. It seems reasonable to conclude that amplitude is an acoustic cue distinguishing sarcasm and sincerity in Cantonese.

Gender differences were examined across attitudes in this study. Male and female speakers used different cues to distinguish sarcasm from the other attitudes. For instance, male speakers relied more on decreasing their speech rates, while female speakers mostly relied on changing their pitch variables. This finding suggested that the compensation function of duration used by male speakers of British English [8] was also applied by Cantonese male speakers.

With respect to the target intensifier /tsɛn˧˥hei˧˥/, our findings showed that the acoustic values were differed in the two sentence sets. For example, utterances with the intensifier had a lower mean amplitude and HNR as well as a faster speech rate. However, it was unclear whether these changes were attributed to the intensifier itself or the change on values of the other parts of the sentence. Future analysis of the data is underway.

To conclude, sarcastic intonation in Cantonese was investigated using six acoustic markers. Additionally, male and female speakers also relied on different cues to signal sarcasm. Further study examining the perception of these various sarcastic cues is currently underway.

5. REFERENCES

- [1] 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.
- [2] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved from <http://www.praat.org/>.
- [3] Brown, P., Levinson, S. C. 1978. Universals in language usage: Politeness phenomena. In *Questions and politeness strategies in social interaction*.
- [4] Bryant, G. A., Fox Tree, J. E. 2002. Recognizing Verbal Irony in Spontaneous Speech. *Metaphor and Symbol*.
- [5] 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.
- [6] Cheang, H. S., Pell, M. D. 2008. The sound of sarcasm. *Speech Communication*, 50(5), 366–381.
- [7] 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.
- [8] 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.
- [9] Fung, R. S.-Y. 2000. *Final Particles in Standard Cantonese: Semantic Extension and Pragmatic Inference*. PhD dissertation. Ohio State University.
- [10] Law, A. 2002. Cantonese sentence-final particles and the CP domain. *UCL Working Papers in Linguistics* 14: 375–398.
- [11] Løevenbruck, H., Jannet, M. A. B., Imperio, M. D., Spini, M., Champagne-lavau, M. 2013. Prosodic cues of sarcastic speech in French: slower, higher, wider. *Institut Universitaire de France*, (August), 3537–3541.
- [12] Mauchand, M., Vergis, N., Pell, M. 2018. Ironic tones of voices. *9th International Conference on Speech Prosody 2018*, (June), 443–447.
- [13] Myers, R. A. 1978. *Irony in Conversation*. Ann Arbor: University of Microfilms International.
- [14] Rao, R. 2013. Prosodic Consequences of Sarcasm Versus Sincerity. *Concentric: Studies in Linguistics*, 2(November), 33–59.
- [15] Rockwell, P. 2000. Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, 29(5), 483–495.
- [16] Rockwell, P. 2007. Vocal features of conversational sarcasm: A comparison of methods. *Journal of Psycholinguistic Research*, 36(5), 361–369.
- [17] Scherer, K. R., Banse, R., Wallbott, H. G. 2001. Emotional Inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural Psychology*, 32(1), 76-92.
- [18] 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)*.