Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Can Cantonese listeners identify the prosodic cues of sarcasm?

Chen Lan, Peggy Mok

The Chinese University of Hong Kong
lchsapphire@gmail.com, peggymok@cuhk.edu.hk

Abstract

This study investigated how prosodic features characterize Cantonese sarcasm and how these features help native listeners understand sarcastic meanings. 28 native Hong Kong Cantonese listeners listened to 50 sentences naturally produced by 6 native Cantonese speakers with a sarcastic attitude and with a sincere attitude. For each sentence the listeners rated whether they perceived the sentence as being produced with a very sincere (1) or very sarcastic (6) tone on a 6-point Likert scale. Acoustic analysis of the stimuli revealed that a slower speech rate, a lower mean F0, a lower mean amplitude, a narrower F0 range, a greater amplitude range, and a higher HNR value are all significant prosodic cues for Cantonese sarcasm although these cues may not be jointly present to deliver a sarcastic meaning. Listeners' ratings indicated that Cantonese listeners were able to discriminate sarcasm from sincerity based on the prosodic features. The combination of the prosodic cues used by the speakers influenced how well the listeners could perceive sarcasm. Listeners could not recognize sarcasm if only one of the cues was applied. The more prosodic cues were utilized in a sarcastic speech, the easier it would be for the listeners to understand the implied sarcastic meaning.

Index Terms: prosodic features, perception, Cantonese sarcasm

1. Introduction

Verbal irony has generally been described as a rhetorical device for either implying the opposite of what the content is literally [1] or expressing a different meaning from what is said [2]. Ironic criticisms, which use positive content to deliver negative meanings, and ironic compliments, which make use of negative content to give positive comments, are two types of irony [3]. In this study, we investigated and discussed the prosodic features of the former one, which was generally referred to as sarcasm.

1.1. Production and perception of sarcasm

Previous studies on the prosodic features of sarcasm mainly focused on the production of sarcasm. Duration, pitch, and intensity have been reported as the important cues distinguishing sarcasm from non-sarcasm even though patterns varied across languages. For example, English sarcasm was delivered with a lower pitch and a slower speech rate [4, 5, 6], while French produced sarcasm with a higher pitch level, a slower speech rate, and a greater amplitude range [7]. Voice quality can also be modulated by speakers in affective communication. English, Mandarin, and Korean speakers have been reported to change voice quality while expressing a sarcastic attitude [4, 8, 9].

Previous studies on the perception of sarcasm argued whether prosodic cues alone could be used to detect sarcasm. [10] suggested that sarcastic utterances could not be detected by prosody alone. Contextual features should be paired with prosodic features to mark sarcasm more accurately. Other researchers argued that listeners were able to figure out sarcasm via the change in some prosodic features (e.g., lower F0 in English) [11, 12]. However, the non-colloquial materials and the presented method of the stimuli (e.g., providing the written form of the stimuli) posed a limitation for the above studies on sarcasm perception. This study explores how prosodic cues work for detecting sarcasm with more rigorous methods.

1.2. Cantonese sarcasm

A sarcastic tone of voice is commonly used in Cantonese, but most of the previous research on Cantonese sarcasm or other ironic forms in Cantonese focused on syntactic structures. For instance, the Cantonese sentence-final particle (SFP) /ʦɛk55/ is commonly used to mark a sense of irony in a positive literal utterance according to the speech context [13, 14, 15]. To our knowledge, there are only two studies that investigated the prosodic cues of Cantonese sarcastic speech [4, 16]. [4] 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 sarcasm from non-sarcastic utterances. Furthermore, a higher HNR value was also reported to mark Cantonese sarcasm, i.e., sarcastic phrases were less breathy. [16] applied a more rigorous method (e.g., more participants and more colloquial stimuli) to revisit the prosodic markers of Cantonese sarcasm, revealing a contrary finding regarding the mean F0, and the amplitude range, that is, Cantonese sarcasm was marked by a lower mean F0 and a greater amplitude.

In addition to the production of Cantonese sarcasm, the authors of [4] also investigated how Cantonese listeners identify sarcasm in their native language and a non-native language (English), which is the only published study focusing on the perception of sarcasm in Cantonese [11]. In their study, the listeners were required to make a forced choice to detect the attitude expressed by the speakers. Both Cantonese and English listeners were able to recognize sarcasm and distinguish sarcasm from sincerity in their native languages. In addition, the prosodic features such as F0 played an essential role in the perception of sarcasm for listeners. However, the stimuli used in [11] may not be very colloquial for native Hong Kong Cantonese speakers according to the judgements by two native speakers. For example, SFPs play an important role in conveying different attitudes in Cantonese, but all the target utterances in [11] were without an SFP, rendering them less colloquial. Regarding the research method, in addition to

10.21437/SpeechProsody.2024-138

asking the listeners to identify whether a sentence is delivered with a sarcastic attitude, it will also be useful to ask them to rate how sarcastic the speech is, which can help connect the prosodic cues used by the speakers and the degree of sarcastic attitude perceived by the listeners.

The current study investigated the perception of Cantonese sarcasm by native Hong Kong Cantonese listeners, aiming to understand whether prosodic features alone were able to distinguish sarcasm from non-sarcasm. Also, this study explored the correlation between the use of prosodic cues in sarcastic speech produced by the speakers and the degree of recognition of sarcastic meaning by the listeners.

2. Method

2.1. Participants

Six native Hong Kong Cantonese speakers (3F), who were undergraduate students aged between 18;5 and 23;8 at a university in Hong Kong, were recruited to record the stimuli. Twenty-eight native Hong Kong Cantonese speakers (14F), who were undergraduate students aged between 17;11 and 24;0 at a university in Hong Kong participated in the perceptual rating task. According to their language background questionnaires, all the speakers and listeners were born and grew up in Hong Kong, having at least one of their parents being a native Hong Kong Cantonese speaker. Also, they went to local primary and secondary schools, and 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 speakers and the listeners. All the participants reported no speech or hearing problems or learning difficulties.

2.2. Stimuli

The materials contained two sets of simple sentences which are commonly used in colloquial Cantonese. As exemplified in Table 1, the first set contained the target utterances with a degree modifier, an adjectival phrase, and an SFP. An intensifier zan55hai22 *'really'* was inserted before the degree modifier to create the second set of utterances, aiming to examine whether results varied with the insertion of this intensifier. This intensifier was used frequently for expressing criticism as well as for assuring sincerity in Cantonese, working naturally for both sarcasm and sincerity [17]. To elicit the sincere and sarcastic attitudes, scenarios with positive or negative situations were presented using audios recorded by two native Hong Kong Cantonese speakers, and visual aids in the form of relevant pictures were also provided. A picture and a target utterance were presented on each slide, and the audio scenario was played automatically. The speakers listened to the audio and produced the target utterance according to the context provided by the audio and the picture. The target utterances were randomized and shown on the screen in different orders in each repetition. All the recordings were conducted in a sound-treated room with a solid-state recorder with a sampling rate of 44100 Hz. Acoustic analyses were conducted to compare the prosodic features of sarcastic utterances and sincere utterances. Utterances with different combinations of prosodic cues were selected for the perception experiment. In total, 100 target utterances consisting of 50 pairs of sentences produced with two attitudes and two sentence sets were used as the stimuli.

2.3. Procedure

All the participants were paid to attend the online perception experiment. With their consent, the participants were required to fill in a language background questionnaire asking for some personal information about them. 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. The participants were instructed to click the link of the sound file, listen to the stimuli, and rate each target utterance on a 6-point Likert 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.

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

Scenarios
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 sentences
a. 你好醒呀 [You are so smart]
b. 你真係好醒呀 [You are really (so) smart]

2.4. Data analysis

The stimuli were acoustically analyzed in Praat [18] using ProsodyPro [19]. 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 the F0 range was provided by subtracting the minimum F0 from the maximum F0. Regarding the amplitude variables, mean intensity, minimum intensity, and maximum intensity were measured in decibels (dB), and the amplitude range was provided by subtracting the minimum intensity from the maximum intensity. For the voice quality, the HNR value was measured. All acoustic data were converted into z-scores using each person's mean before statistical analysis. How these prosodic features were manipulated by the speakers in each sentence was examined. When the change of a prosodic feature is aligned with that of the overall pattern, it can be considered as a prosodic cue utilized by the speaker to deliver sarcastic meaning. The number of prosodic cues utilized in each sentence was calculated and classified into different types of combinations of prosodic cues.

For the perception experiment, the rating scores from 2800 responses (50 target utterances × 2 attitudes × 28 participants) were analyzed with a linear mixed model in the R program [20] using the lmerTest package [21]. In the model, the interaction between Attitude and sentence Set (Attitude * Set) was set as a fixed effect, and Participant was entered as a random effect. Independent-sample t-tests were used to further compare the listeners' ratings for two attitudes and two sentence sets. Simple linear regressions were conducted to explore the relationship between the prosodic cues used by the speakers and the degree of sarcastic attitude perceived by the listeners.

3. Results

3.1. Prosodic cues in production

Figure 1 summarizes the average normalized values of the six prosodic parameters, including the speech rate, mean F0, F0 range, mean amplitude, amplitude range, and the HNR. The statistical analysis revealed a significant main effect for Attitude (F (1, 588) = 12.25, p = .001). Significant differences were found between the two attitudes in terms of each prosodic feature. Compared to sincere sentences, sarcastic sentences in Cantonese were produced with a significantly slower speech rate, a lower mean F0, a lower mean amplitude, a narrower F0 range, an enlargement of amplitude range, and an increase in the HNR. These prosodic cues characterized the sarcastic tone of voice in Cantonese. The examination of how these prosodic cues were used showed speaker variability. These cues may not be jointly present to express a sarcastic attitude by Cantonese speakers. Seven combinations of the prosodic cues were found in the stimuli as listed in Table 2. Among the 50 sarcastic utterances, 8% of them were produced with less than three prosodic cues (i.e., Type 1 and Type 2). 40% of the sarcastic sentences were observed with three prosodic cues. The speakers produced most of the sarcastic utterances (52%) by changing all four prosodic variables (i.e., speech rate, pitch, amplitude, and voice quality).

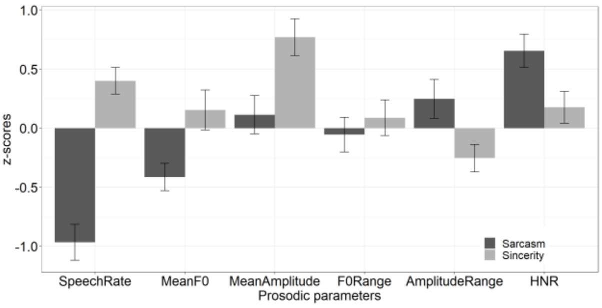

Figure 1: *Mean values (z-scores) of the six acoustic variables across two attitudes. Error bars indicate the standard errors.*

3.2. Overall perceptual patterns

Wilcoxon signed rank test comparing the rating scores between two attitudes (sarcasm vs. sincerity) revealed a significant difference (Wilcoxon Z = 406, p < .001) with a higher rating score for the recognition of sarcastic utterances (Mean score = 4.00, SD = 0.43) than for the sincere utterances (Mean score = 2.40, SD = 0.47). Figure 2 summarizes the mean scores of the Cantonese listeners' perceptual ratings for the two attitudes across sentence sets.

A linear mixed model revealed a significant main effect for Attitude (|t| = 21.900, p < .001) and for sentence Sets (|t| = 5.173, p < .001). Independent-sample t-tests further indicated that the rating score of Cantonese sarcasm was significantly higher than that of sincerity in the sentences with the intensifier (|t| = 21.923, p < .001) and without the intensifier (|t| = 22.695, p < .001). Sarcastic utterances with the intensifier were rated significantly more sarcastic than those without the intensifier (|t| = 5.981, p < .001), while sincere utterances with the intensifier were rated significantly less sincere than those without the intensifier (|t| = 6.131, p < .001).

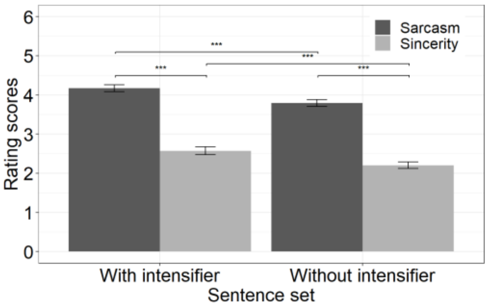

Figure 2: *Mean rating scores of two attitudes across sentence sets. Error bars indicate the standard errors.*

3.3. Correlation between the prosodic cues and the sarcasm perception

Two criteria were applied to analyze the relationship between the use of prosodic cues in sarcastic speech and the rating for the degree of sarcasm. Under criterion 1, six prosodic cues were included separately for comparison. For example, if a sarcastic sentence has a slower speech rate, a lower mean F0, a lower mean amplitude, and a higher amplitude range compared to its sincere counterpart, four prosodic cues were calculated. In total, six prosodic cues were considered under criterion 1. Criterion 2 aligned with the types of the combination of prosodic cues used as listed in Table 2. Mean F0 and F0 range were classified as pitch variables, while mean amplitude and amplitude range were classified as amplitude variables. Speech rate, pitch variables, amplitude variables, and voice quality (the HNR value) were included for comparison.

Table 2: *Mean rating scores for seven types of combinations of prosodic cues used.*

Combinations of prosodic cues used	**Mean rating score (*SE*)**
Type 1: Speech rate	2.68 (*0.02*)
Type 2: Speech rate + Amplitude	3.38 (*0.02*)
Type 3: Pitch + Amplitude +HNR	3.20 (*0.02*)
Type 4: Speech rate + Pitch + Amplitude	3.97 (*0.01*)
Type 5: Speech rate + Amplitude + HNR	4.11 (*0.01*)
Type 6: Speech rate + Pitch + HNR	4.69 (*0.02*)
Type 7: Speech rate + Pitch + Amplitude + HNR	4.63 (*0.002*)

Figure 3 shows the number of prosodic cues used by the speakers and the average rating scores under two criteria, and the results of simple linear regressions are also presented. The number of prosodic cues used significantly predicted how Cantonese listeners rate the sarcastic utterances under both criteria. The rating scores increased as a function of the number of six prosodic cues used (b = 0.344, SE = 0.13, t = 2.619, p = 0.012) and as a function of the number of four prosodic variables used (b = 0.518, SE = 0.20, t = 2.589, p = .013). There is a significant positive correlation between the rating scores and the number of prosodic cues used by the speakers to deliver a sarcastic attitude, that is, the more prosodic cues were utilized in a sarcastic speech, the easier it would be for the listeners to understand the implied sarcastic meaning.

Further examination was conducted regarding the relationship between the rating scores and different combinations of prosodic cues used by Cantonese speakers. Table 2 shows the mean rating score for each combination of prosodic cues. The mean rating score for the sarcastic utterances with only one prosodic cue (i.e., speech rate) was within the range of 1-3, suggesting that it is difficult for the listeners to identify sarcastic meaning based on only one prosodic cue. In addition, without either a slower speech rate (i.e., Type 3) or a higher HNR value (i.e., Type 2 and Type 4), the rating scores for the sarcastic utterances were significantly lower than the overall mean scores, and they were also significantly lower than the rating for the sarcastic sentences with these two prosodic cues (i.e., Type 5, Type 6, and Type 7). To sum up, compared to the pitch variables and the amplitude variables, speech rate and voice quality (HNR) are the more important cues for the listeners to perceive Cantonese sarcasm.

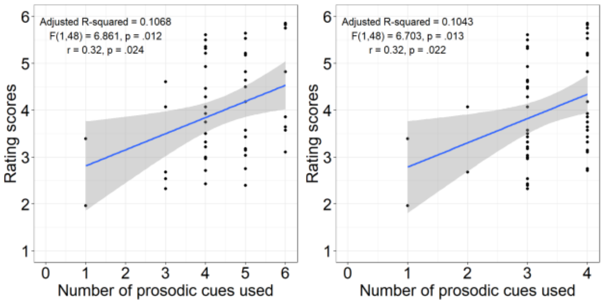

Figure 3: *The correlation between the number of six prosodic cues used and the rating scores (left) and between the number of four prosodic variables and the rating scores (right).*

4. Discussion

The present study investigated how prosodic features distinguish sarcasm from sincerity in Cantonese and how these features help native listeners understand sarcastic meanings. In general, a slower speech rate, a declination in mean F0 and mean amplitude, a narrower F0 range, and an increase in amplitude range and the HNR characterize Cantonese sarcasm. Native speakers utilized different combinations of these prosodic cues to convey sarcastic meanings in daily communication.

In terms of the perception of Cantonese sarcasm, our finding indicates that prosody alone can help listeners recognize the sarcastic intention of the speakers. With the same context, Cantonese listeners were able to discriminate sarcasm from sincerity according to the alteration of several prosodic parameters. How well the listeners detect the sarcastic tone may rely on the number of prosodic cues used by the speakers. It is difficult for the listeners to successfully identify the sarcastic tone with only one or two prosodic cues. It would be more helpful for delivering a sarcastic attitude to the listeners if more prosodic cues were utilized. Furthermore, unlike a previous study on the perception of Cantonese sarcasm which said that pitch was the prominent cue for sarcasm identification [4], this study found that speech rate and voice quality also play an essential role in perceiving Cantonese sarcasm. The alteration of the speaking rate is regarded as an important prosodic cue of sarcasm which has been reported in different languages (e.g., English [4, 5, 6], Italian [22], French [7], Mexican Spanish [23], Mandarin [8], and Cantonese [4, 16]). Our finding supports the significant role of this cue in signaling sarcastic attitude from the perceptual perspective. Considering that the slower speech rate has been reported to significantly distinguish sarcasm and sincerity in Cantonese in terms of production in the previous studies [4, 16] and perception in this study, this cue can be regarded as a stable prosodic marker for Cantonese sarcasm. The significant influence of the HNR on perceiving sarcasm supports the existence of voice quality modulation in affective communication, and the listeners were able to recognize a decrease in the amount of noise in the speaker's voice.

Our finding also exposes 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 [24]. Our finding suggests that the positive context produced with a sarcastic tone of voice can be perceived as sarcastic instead of sincere, even though the rating for the sarcastic utterances was closer to the mid-range compared to the rating of sincerity. This may be explained by the opinion in [24] 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 are able to recognize a negative meaning while hearing a positive content with a sarcastic tone of voice.

In addition, the insertion of the intensifier zan55hai33 *'really'* significantly influenced the listeners' perception of Cantonese sarcasm, making them perceive the sentences as more sarcastic or less sincere. There are two possible explanations. From the perspective of speech, the speakers may emphasize the target intensifier while producing the utterances, which may lead to a more exaggerated change in several prosodic parameters. The listeners may perceive the sarcastic tone more clearly with these modulations. The second explanation is that the insertion of the target intensifier may trigger a syntactic cue in addition to the prosodic cues. Since this intensifier can be used for expressing criticism in Cantonese, the listeners may correlate this word with negative intention, considering it as a cue to sarcasm. However, prosody still plays a significant role in the rating for sarcastic and sincere sentences. Listeners were able to distinguish sarcasm and sincerity without the intensifier, and with the intensifier, sarcastic and sincere utterances were rated within the expected range (e.g., 1-3 for sincere). This finding suggests that the syntactic cue and the prosodic cues may jointly work in sarcasm perception.

In conclusion, prosodic features can distinguish sarcasm and non-sarcasm not only in production but also in perception. How several prosodic cues are utilized by the speakers influences how well the listeners perceive the sarcastic meanings. More data will be collected to have a more comprehensive understanding of the prosodic cues and sarcasm.

5. Acknowledgements

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

6. References

[1] P. Brown and S. C. Levinson, Universals in language usage: Politeness phenomena. In *Questions and politeness strategies in social interaction*. 1978.

[2] R. A. Myers, *Irony in Conversation*. Ann Arbor: University of Microfilms International. 1978.

[3] M. Mauchand, N. Vergis and M. D. Pell, Ironic tones of voices. *Proceedings of the International Conference on Speech Prosody*, Poznan, Poland. 443–447, 2018.

[4] H. S. Cheang and M. D. Pell, The sound of sarcasm. *Speech Communication, 50*(5), 366–381, 2008.

[5] A. Chen and L. Boves, What's in a word: Sounding sarcastic in British English. *Journal of the International Phonetic Association, 48*(1), 57–76, 2018.

[6] P. Rockwell, Vocal features of conversational sarcasm: A comparison of methods. *Journal of Psycholinguistic Research, 36*(5), 361–369, 2007.

[7] H. Loevenbruck, M. Ameur, B. Jannet, M. D'imperio, M. Spini and M. Champagne-Lavau, Prosodic cues of sarcastic speech in French: Slower, higher, wider. *Proceedings of 14th Annual Conference of the International Speech Communication Association*, Lyon, France. 3537-3541, 2013.

[8] S. Li, W. Gu, L. Liu and P. Tang, The Role of Voice Quality in Mandarin Sarcastic Speech: An Acoustic and Electroglottographic Study. *Journal of Speech, Language, and Hearing Research, 63*, 2578–2588, 2020.

[9] S. Y. Yang, Listener's ratings and acoustic analyses of voice qualities associated with English and Korean sarcastic utterances. *Speech Communication, 129*, 1–6, 2021.

[10] T. Joseph, D. Traum and N. Shrikanth, Yeah right: Sarcasm recognition for spoken dialogue systems. *Proceeding of INTERSPEECH 2006 ICSLP*, 1838-1841, 2006.

[11] H. S. Cheang and M. D. Pell, Recognizing sarcasm without language. *Pragmatics & Cognition, 19*(2), 203–223, 2011.

[12] P. Rockwell, Lower, Slower, Louder: Vocal Cues of Sarcasm. *J Psycholinguist Res, 29*, 483–495, 2000.

[13] S. Matthews and V. Yip, *Cantonese: A Comprehensive Grammar*. Routledge. 1994.

[14] M. K. M. Chan, Gender-related use of sentence-final particles in Cantonese. *Gender Across Languages: The linguistic representation of women and men* (2). 2002.

[15] J. P. W. Li, T. Law, G. Y. H. Lam, and C. K. S. To, 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, 2013.

[16] C. Lan, P. L. Hui, W. W. Xu, and P. Mok, Revisiting acoustic markers of sarcasm in Cantonese. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)*. Melbourne. 2019.

[17] R. S.-Y. Fung, *Final Particles in Standard Cantonese: Semantic Extension and Pragmatic Inference*. PhD dissertation. Ohio State University. 2000.

[18] P. Boersma and D. Weenink, Praat: doing phonetics by computer [Computer program]. Version 6.4.04, retrieved from http://www.praat.org/. 2024.

[19] Y. Xu, ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France. 7-10, 2013.

[20] R Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2022.

[21] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26, 2017.

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

[23] R. Rao, Prosodic Consequences of Sarcasm Versus Sincerity. *Concentric: Studies in Linguistics, 2*(November), 33–59, 2013.

[24] J. Woodland and D. Voyer, Context and intonation in the perception of sarcasm. *Metaphor and Symbol, 26*: 227-239, 2011.