Expressing the Opposite: Acoustic Cues of Thai Verbal Irony

Nimit Kumwapee
Department of Linguistics,
Faculty of Arts,
Chulalongkorn University
`nimit.kumwapee@gmail.com`

Sujinat Jitwiriyanont
Department of Linguistics and
Southeast Asian Linguistics Research Unit,
Faculty of Arts, Chulalongkorn University
`sujinat.j@chula.ac.th`

Abstract

The present study examined the acoustic cues of sarcasm in Thai. Four Thai speakers participated in two reading tasks: neutrality and sarcasm reading. Speech rate, F0 mean, F0 range, amplitude mean, amplitude range, F0 slope, and F0 intercept were measured and analyzed. The results indicated that sarcasm was produced at a faster speech rate and with a higher F0 mean, a wider F0 range, and a higher amplitude mean. Amplitude range reported no statistical significance. F0 slope alone was an insignificant cue, but F0 slope together with F0 intercept could distinguish between sarcasm and neutrality. Regarding gender differences, male speakers decreased their speech rate and increased their F0 mean while female speakers increased their F0 range when expressing sarcasm. Also, both male and female speakers increased their amplitude mean when producing sarcasm.

1 Introduction

Verbal irony is a linguistic device utilized to convey an opposite meaning from the literal meaning embedded in its linguistic form (Searle, 1991). To illustrate, the utterance "Your cooking is the best." means that the person to whom the utterance is addressed is literally the best cook while can also suggest that the person is the worst cook if the speaker intends to say otherwise. Additionally, irony is regarded as flouting the Maxim of Quality (Grice, 1989). In other words, the speaker of the utterance above was lying to the addressee that they are good at cooking.

Sperber and Wilson (1981) propose another framework that irony is an echoic expression of thoughts that conveys a dissenting attitude such as skepticism, mockery, or contempt. Thus, the speaker of the utterance above could be mocking the person they address. Despite having more definitions than presented here, irony can be viewed as a form of communication expressing a different/opposite facet of meaning with different attitudes.

Different meanings within a statement can be expressed via different methods employed whether it be different word choice, syntactic structure or means of communication. That is, some words convey a more formal meaning than another and some structures emphasize on different pieces of information. Different tones of voice are used to signify different meanings apart from the original statement. During communication, meaning is not just encoded in linguistic forms. It is both linguistic and non-linguistic cues that play a role in an effective communication among interlocutors. Linguistic cues are encoded within sounds, words, and structures as previously explained. Non-linguistic cues, on the other hand, involve gestures, facial expressions, and the situations/contexts in which the communication takes place. Hellbernd and Sammler (2016) found that extralinguistic cues such as speech prosody also function in conveying intentions. They found that prosodic cues such as F0 rise, mean F0, mean intensity, and duration provided a foundation for listeners to recognize the intention of speakers. This shows that acoustic cues also help convey information that is not encoded within the linguistic form—intentions—which in turn helps decode the

intended meaning of an utterance. Bryant and Fox Tree (2002) similarly examined the role of contextual and prosodic information in the recognition of verbal irony and found that participants rated a sentence as more sarcastic[1] when provided with acoustic contents or irony-biasing contexts. This shows that both acoustic cues and contextual information work hand in hand during the process of inferring or decoding ironic intents. Hence, prosodic cues do not only function as local linguistic cues, but also function as pragma-linguistic cues that help in the signaling and interpreting processes of intents in communication. Speakers undoubtedly employ such cues when expressing verbal irony.

1.1 Acoustic Cues and Sarcasm

There were a wide range of studies regarding the acoustic characteristics of sarcasm. Each study examined different acoustic parameters of a different language, but speech rate, F0 mean, and F0 range appeared as variables in all studies. Majority of studies found that sarcasm was produced with a slower speech rate when compared to neutral speech regardless of language. However, F0 mean and F0 range appeared to be language-dependent and varied across studies. For example, sarcasm was expressed with a lower F0 mean in English (Cheang and Pell, 2008; Bryant, 2010; Chen and Boves, 2018), Spanish (Rao,2013) and Cantonese (Lan et al., 2019), but was produced with a higher F0 mean in French (Lœvenbruck et al., 2013) and Italian (Anolli et al., 2002). Although there were different findings concerning F0 mean in English and Cantonese (Rockwell, 2007; Cheang and Pell, 2009), such difference could be a result of different methodologies used in each study. Apart from speech rate, F0 mean, and F0 range, amplitude mean and range –previously found to be insignificant cues in English –were also found to be another significant cue for sarcasm in various works (Anolli et al., 2002; Cheang and Pell, 2009; Lan et al., 2019). To recap, these studies support the previous point that speakers employ different acoustic cues when expressing verbal irony and these cues are found to contrast with neutral speech and vary

across language.

Apart from the differences in acoustic cues between speech types found, the studies by Rao (2013), Chen and Boves (2018), and Lan et al. (2019) also found the differences between sarcasm produced by male and female speakers. However, there were differences across language. To elaborate, Rao (2013) found that Mexican Spanish male speakers would significantly decrease their speech rate and suppressed their F0 range more than female speakers when producing sarcasm. Additionally, Chen and Boves (2018) discovered that British English male speakers showed larger durational difference than female speakers whereas female speakers would lower their mean pitch when expressing sarcasm. Lan et al. (2019) found a similar pattern to Chen and Boves (2018) in term of the durational difference; however, they also found that female speakers showed a larger F0 mean difference than male speakers and F0 range was found significantly smaller only in female speakers. Hence, speakers of different gender could employ acoustic cues differently when producing sarcasm and such differences are language dependent like the acoustic characteristics of sarcasm themselves.

1.2 Acoustic Study of Sarcasm in Thai

Majority of works that studied sarcasm/verbal irony in Thai concerned pragmatics (Panpothong, 1996; Kongchang, 2017; Bunnag, 2017) and stylistics (Anansapsuk, 2016). Nevertheless, no studies examined sarcasm in Thai from an acoustic perspective. An acoustic study of Thai most relevant to the study of verbal irony was carried out by Sonboonta (2010). Sonboonta (2010) studied the acoustic characteristics of the What-word /ʔarai/ in direct and indirect speech acts and found that despite the fact that no difference between speech acts could be established, F0 of a positive indirect speech act was found to be higher than that of a negative indirect speech, suggesting that there was difference in acoustic values of different speech acts to some extent. However, the study only examined the acoustic characteristics of a single word produced by female speakers. Consequently, an acoustic study of Thai sarcasm in the sentence level produced by both male and female speakers is needed so as to generalize the characteristics of Thai sarcasm and to provide a broader

---

[1]Majority of studies seem to use the term sarcasm in Searle's sense, and both are used interchangeably despite the distinctions. When sarcasm is used in this study, it refers to the general sense of verbal irony.

picture about acoustic cues of sarcasm from the perspective of the Thai language. It is also important to note that most of the languages studied are non-tonal with the only exception of Cantonese. An acoustic study of Thai sarcasm would also add to the existing literature the acoustic characteristics of sarcasm from the point of view of a tonal language such as Thai. The aims of this study are (1) to study the acoustic cues of Thai sarcasm in the sentence level and (2) to investigate whether there is a difference in acoustic cues of sarcasm between male and female speakers.

2 Methodology

2.1 Participants

Two male and two female Thai speakers participated in this study ($M = 24$ years old, $SD = 0.71$). All participants spoke Bangkok Thai fluently. They were picked from a self-selection process and voluntarily participated without being paid. The participants reported no speech hearing and production problems.

2.2 Materials

Thirty sentences and a biasing situation for each sentence were prepared for this study. Each sentence did not end with any final particles and was affirmative sentence with the number of words ranging from five to thirteen ($M = 7.27$ words/sentence, $SD = 2.1$). Two sets of materials were prepared: (1) Baseline Reading Set and (2) Biased Reading Set. The former was just the thirty baseline sentences without any context provided whereas the latter was a set of each baseline sentence preceded by its biasing context as presented in Table 1.

2.3 Data Collection

Two production tasks were conducted: baseline sentence reading and biased sentence reading.[2] The reading was self-recorded by each participant using a mobile application with a sampling rate between 41-48 kHz. The participants were instructed to record a sample of their voice and send back to check whether the voice level was not too loud or

---

[2]While some of the previous studies used a perception test to select only the productions that sound sarcastic and some did not, the current study did not use a perception test because it focused on production, the signalling of sarcasm through acoustic cues.

too low and whether they placed their cellphone too close or too far before the recording began. Afterwards, they received the Baseline Reading Set and they were instructed to read each sentence once with 2-3s pause between each. After they sent back their baseline recording, they received the Biased Reading Set and were instructed to read the situation context silently before producing the bolded and underlined target sentence. Total tokens (4 participants x 2 readings x 30 sentences) acquired were 240 tokens.

2.4 Acoustic Value Extraction and Calculation

Acoustic measures such as utterance duration, maximum pitch, minimum pitch, mean pitch, pitch listing, maximum intensity, minimum intensity/amplitude, and mean intensity/amplitude were extracted manually for each token using Praat. At this stage, mean F0 (Hz) and mean amplitude (dB) were automatically measured without any further calculation needed. Afterwards, speech rate (word/s) was calculated by dividing the syllable number of a sentence by the utterance duration. F0/amplitude range was calculated by subtracting the maximum value by the minimum value. F0 values gained from the pitch listing function in Praat were converted into a data point set. The data point set was then used to calculate a slope (Hz/ms) and its intercept (Hz) using the Least Square Regression method. The extraction and calculation proceeded as above for all 240 tokens.

2.5 Acoustic Analyses

An average of each acoustic cue was calculated for each participant, for each gender, and for all participants. All data were converted into z-scores. The acoustic characteristics of sarcasm and neutrality between participants, between gender, and for overall participants were then observed from the z-scores.

2.6 Statistical Analyses

Raw acoustic values stored as a data table with information about gender and speech type were imported into R to perform a statistical analysis. ANOVA test for each acoustic variable considering speech type as a factor was then carried out to test the difference in acoustic values between types of speech whereas Paired t-test for each acoustic cue of each speech

Biasing context	Baseline sentence
You went out with a group of friends and then they invited a person you did not like to join as well. After finishing, your friend asked you how it was. You ironically replied:	/cʰăn miː kʰwaːm sùk mâːk/ "I am very happy."
You assigned an employee to arrange a set of documents. Although there were a few documents, he took half a day to finish. You then said:	/kʰăw tʰam ŋaːn rew ciŋ ciŋ/ "He works really fast."

Table 1: Examples of Baseline Sentences together with Their Biasing Contexts.

type was conducted to test whether there was a difference between gender. The alpha level used in this study was 0.05.

3 Results

3.1 General Acoustic Cues between Neutrality and Sarcasm

Figure 1 showed that the values of speech rate, F0 mean, F0 range, amplitude mean, and amplitude range for sarcasm was higher than that of neutrality. However, the statistical significance was found only in the case of speech rate, F0 mean, F0 range, and amplitude mean. To elaborate, sarcasm is significantly produced at a faster speech rate than neutrality, $F(1,238) = 13.45$, $p = .0003$. When expressing sarcasm, speakers increased their F0 mean significantly, $F(1,238) = 5.308$, $p = .0221$, and exhibited a significant wider F0 range than when they produced neutral speech, $F(1, 238) = 6.176$, $p = .0136$. As for amplitude, sarcasm was significantly produced with a higher amplitude mean, $F(1,238) = 6.985$, $p = .0088$, but there is no significant difference found in amplitude range, $F(1,238) = 0.016$, $p = .901$. This suggested that when Thai speakers produced sarcasm, they did so by increasing the speech rate, F0 mean, F0 range, and amplitude mean. Although the values of amplitude range appeared to be higher in sarcastic speech, amplitude range might not be a reliable cue when it came to distinguish sarcasm and neutrality.

Figure 2 illustrated the relationship between F0 slope (x-axis) and F0 intercept (y-axis) obtained from the Least Square Regression method. Sarcasm seemed to stay in a higher region for both male and female speakers (See Figure 2). 65.83 % of F0 slope
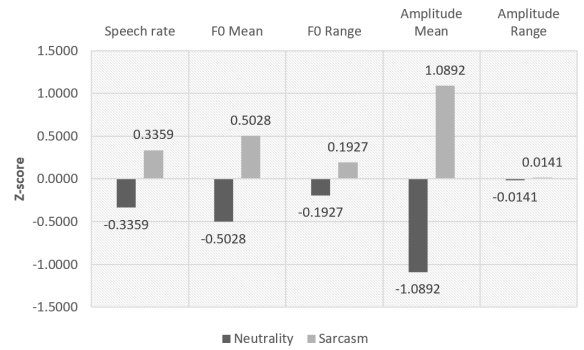

Figure 1: Five Acoustic Values (Z-Scores) of Two Speech Types.

in sarcastic speech were less than their original neutral slopes, suggesting that speakers employed different F0 contours for sarcasm. 83.33 % of F0 intercept in sarcasm were higher than their neutral counterparts, supporting that sarcasm was produced with a higher F0 mean than neutral speech. However, there was no significant difference in F0 slope between speech types. Still, there was a significant difference in F0 intercept between neutrality and sarcasm, $F(1,238) = 10.58$, $p = .0013$. Additionally, there was a significant difference in F0 slope together with F0 intercept between neutral and sarcastic speech, $F(1,238) = 10.55$, $p < .01$. The interaction between F0 slope and F0 intercept was also found to be statistically significant, $F(1,238) = 6.85$, $p < .01$. This signified that the two values taken together as a parameter could distinguish between sarcasm and neutrality. Moreover, sarcasm appeared to exhibit more variability than neutrality. To illustrate, F0 slope of sarcasm showed more variability ($M = -0.3122$, $SD = 0.4963$) than that of neutrality ($M = -$

0.2134, *SD*= 0.377). Likewise, F0 intercept of sarcasm also showed more variability (*M*= 207.6355, *SD*= 66.9692) than that of neutrality (*M*= 180.5076, *SD*= 61.568). It could be that speakers manipulated their F0 when they produced a sarcastic sentence more than when they made a neutral sentence.
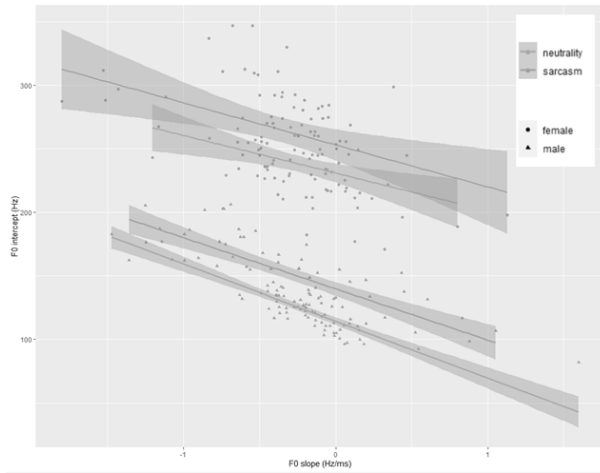

Figure 2: Relationship between F0 Slope and F0 Intercept by Speech Type and Gender with Linear Fitted Smooths (Upper lines of each pair represent sarcasm) and Confidence Bands.

3.2 Acoustic Cues between Sarcasm and Neutrality by Participants and Speech Types

As for speech rate (See Figure 3 Panel A), although M1 speaker seemed to significantly emphasize his sarcastic reading, resulting in a slower speech rate for sarcasm than neutrality, the overall speech rate of sarcasm was still faster than neutral speech. With regard to gender, male speakers produced a sarcastic speech with a significantly slower rate than neutral speech, $t(1,118) = 2.4092$, $p = .0175$. On the other hand, female speakers produced a sarcastic speech with a significantly faster rate than neutral speech, $t(1,118) = -8.9696$, $p < .0001$. This illustrated that there was gender difference despite the general characteristics of speech rate for overall participants.

Regarding F0 Mean, male speakers seemed to significantly increase their F0 mean in sarcastic speech, $t(1,118) = -9.8553$, $p < .0001$. However, there was no significant difference in F0 mean between sarcasm and neutrality within female speakers, $t(1,118) = -1.7472$, $p = .0832$. This was due to F2 speaker whose F0 mean did not differ much between sarcasm and neutrality (See Figure 3 Panel B).

For F0 range, there was no significant difference in F0 range between sarcasm and neutrality in male speakers, $t(1,118) = -0.6338$, $p = .5274$, possibly because M2 speaker produced sarcasm with a narrower F0 range (See Figure 3 Panel C). However, female speakers significantly exhibited a wider F0 range ($t(1,118) = -3.2322$, $p = .0016$) when producing sarcasm as can be seen in both F1 and F2 speakers.

For amplitude mean (See Figure 3 Panel D), both male and female speakers increased their amplitude mean when expressing sarcastic speech, $t(1,118) = -2.0723$, $p = .0404$, and $t(1,118) = -2.6534$, $p = .0091$, respectively. However, M2 speaker showed a different pattern from the rest of the participants. Despite this difference, the overall characteristic of amplitude means for overall speakers, male speakers, and female speakers was still statistically significant.

For amplitude range, the difference between male and female speakers could not be generalized because there were both cases that amplitude range was wider (M1, F1) and narrower (M2, F2) for sarcasm within both group (See Figure 3 Panel E).

4 Conclusion and Discussion

This study examined speech rate, F0 mean, F0 range, amplitude mean and amplitude range along with F0 slope and F0 intercept of sarcasm and neutrality. Speech rate, F0 mean, F0 range, and amplitude mean were found to be a significant cue that differentiated between speech types. Speech rate for sarcasm in Thai is different from other languages because it is expressed at a faster speech rate (See Table 1). However, one speaker appeared to emphasize and stress each word significantly, resulting in a slower speech rate. The difference in speech rate might be a result from different emotions expressed in different contexts such as in the study by Tumtavitikul and Thitikannara (2006) that showed that there was a difference in duration for speech produced with different emotions. The current work was in line with Yimngam et al. (2011) that neutral speech was produced at the slowest speech rate from other types of emotions. This study did not control situational or emotional contexts because it

A

SPEECH RATE

B

F0 MEAN

C

F0 RANGE
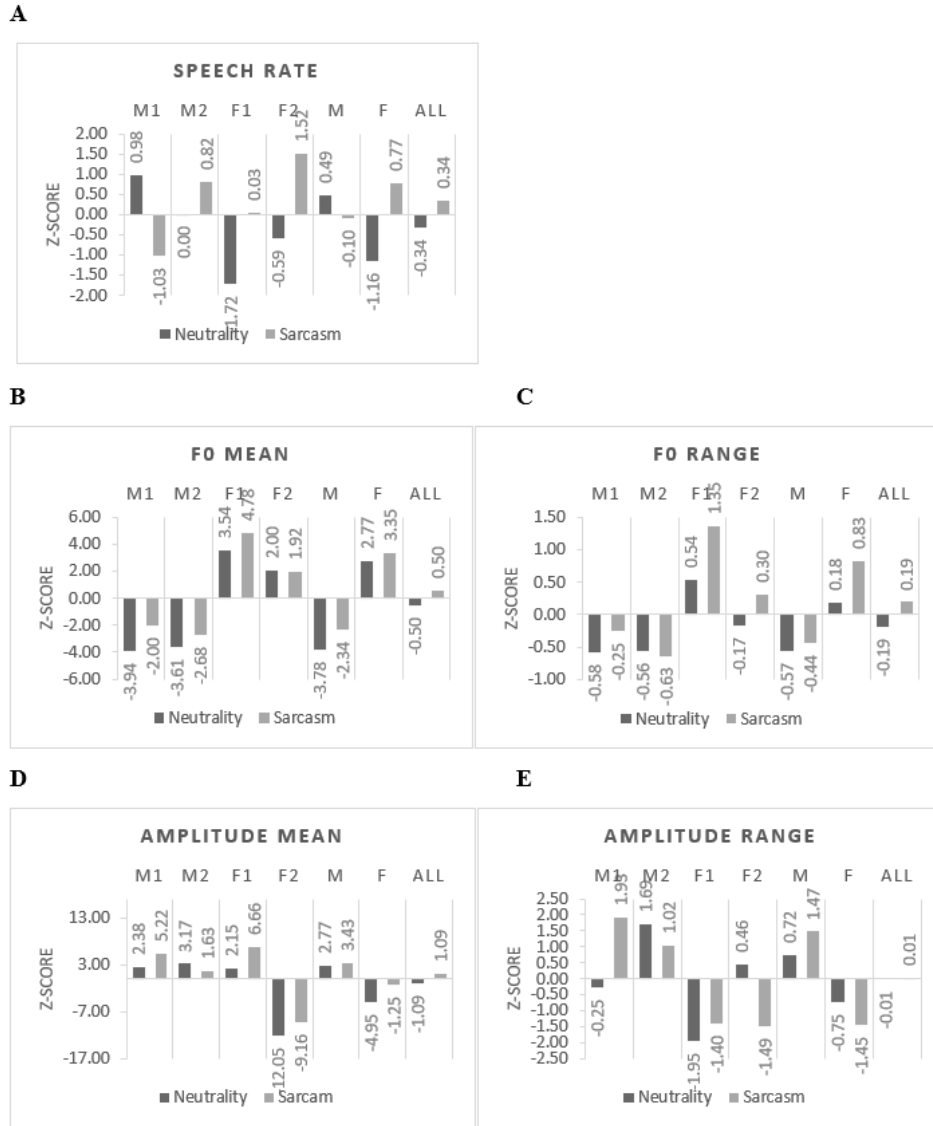
D

AMPLITUDE MEAN

E

AMPLITUDE RANGE

Figure 3: Acoustic values of Sarcasm and Neutrality across Participants and Genders and for Overall.

aimed to explore the general characteristics of sarcastic speech. The variation in sarcasm in term of types and emotions awaits further study.

For F0, the result in F0 mean that sarcasm is produced with a higher pitch corresponds with studies in French (Lœvenbruck et al., 2013) and Italian (Anolli et al., 2002). Surprisingly, the result of F0 mean in this study is different from Cantonese (Lan et al., 2019), which is also a tonal language. Regarding F0 range, that sarcasm is expressed with a wider F0 range in this study is accordance with studies of French (Lœvenbruck et al., 2013), Italian (Anolli et al., 2002), and Cantonese (Cheang and pell, 2009;

Lan et al., 2019). Regarding amplitude, that sarcasm is expressed with a higher amplitude is in line with the studies in Italian (Anolli et al., 2002) and Cantonese (Cheang and pell, 2009; Lan et al., 2019). Although sarcasm seems to be expressed with a wider amplitude range as in Italian (Anolli, et al., 2002) and Cantonese (Lan et al., 2019), there is no significant difference in amplitude range found in the current study.

As for F0 slope, there is no significant difference between speech types. However, F0 slope together with F0 intercept distinguishes between neutrality and sarcasm. Sarcasm seems to exhibit a larger vari-

ability in F0 slope than neutral speech. The variability might be due to different emotional contexts in which sarcasm is expressed. Study of F0 contour by Gu and Lee (2007) showed that different emotional speech displayed different sentential F0 declination. Additionally, different emotions affect the pattern of F0 contour or tone as in Li (2015). Hence, the variability could arise from the emotional differences which in turn result in different F0 contour patterns. Even though there was no significant difference found for F0 slope, F0 movement and contour are found to be specific to different emotions as in the works by Paeschke and Sendlmeier (2000) and by Paeschke (2004). Moreover, that there was no significance found is possibly due to the nature of a tonal language that F0 movement of a sentence could not be so different from its original neutral sentence as Wu (2019) found that unnaturalness in F0 affected perceivability. As some manipulations of F0 seem to exist, future works could explore whether there is a difference in intonation patterns of different speech types or whether sarcasm affects the intonation patterns in Thai.

This study also examined gender as a factor and found that within the general acoustic characteristics found for sarcasm, there are gender differences across different acoustic cues. That male speakers produce sarcasm with a slower speech rate is in line with the studies by Rao (2013), Chen and Boves (2018), and Lan et al. (2019), but female speakers shows more durational difference between sarcasm and neutrality than male speakers in this study. Additionally, this study found that male speakers increase their F0 mean significantly when expressing sarcasm whereas female speakers exhibit a wider F0 range. Likewise, both genders increase their amplitude mean significantly when making a sarcastic sentence. Nevertheless, this study found that male speakers rely not only on decreasing their speech rate as found in Rao (2013), Chen and Boves (2018) and Lan et al. (2019), but also on increasing their F0 mean. Whereas Lan et al. (2019) found that F0 range was significantly smaller only in female speakers, this study found a different pattern that F0 range was significantly wider only in female speakers. These similarities and differences support that male and female speakers use different acoustic cues for sarcasm in different languages and gender should

also be included as one of the variables so as to find out if there is to be difference within the overall acoustic characteristics.

In this study, sarcasm was found to be expressed with a faster speech rate, higher F0 mean, wider F0 range, and higher amplitude mean. The gender difference was also observed across acoustic cues of sarcasm. Also, this study found that there are variations within sarcasm in term of emotional and situation contexts that awaits future works. Future works may explore various sarcastic sentence patterns apart from an affirmative type and consider different types of sarcasm expressed with different emotions such as anger or joy or in different settings such as friendly or unfriendly so as to provide a more comprehensive picture about the variations and characteristics of sarcastic speech in Thai.

References

Anansapsuk, Ornkanya. (2016). Verbal irony in Udom Taepanich's 7-10 episode stand-up comedy. Master thesis, M.A (Thai). Bangkok: Graduate School, Chulalongkorn University.

Anolli, L., Ciceri, R., and 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.

Boersma, P., and Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved from https://www.praat.org/

Bryant, G. A. (2010). Prosodic contrasts in ironic speech. *Discourse Processes*, 47(7): 545–566.

Bryant, G. A., and Fox Tree, J. E. (2002). Recognizing Verbal Irony in Spontaneous Speech. *Metaphor and Symbol*, 99-117.

Bunnag, Orawee. (2017). The functions of verbal irony for commenting via Facebook of mobile operator in Thailand. *Vacana Journal*, 5(2): 24-40.

Cheang, H. S., and Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, 50(5): 366-381.

Cheang, H. S., and Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America*, 126(3): 1394-1405.

Grice, H. P. (1989). *Studies in the way of words.* Cambridge, Mass.: Harvard University Press.

Gu, W., and Lee, T. (2007). Quantitative analysis of F0 contours of emotional speech of Mandarin. In *SSW*, 228-233.

Hellbernd, N., and Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act

perception. *Journal of Memory and Language*, 88: 70-86.

Kongchang, Rumrada. (2017). Responding strategies to verbal irony in Thai: a case study of speakers of equal status. Master thesis, M.A (Thai). Bangkok: Graduate School, Chulalongkorn University.

Lan, Chen, Hui, Pak, Xu, Wenwei and Mok, Peggy. (2019). Revisiting Acoustic Markers of Sarcasm in Cantonese. In *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)*, 77-81.

Li, A. (2015). Emotional Intonation and Its Boundary Tones in Chinese. In *Encoding and Decoding of Emotional Speech*, 133-164. Springer, Berlin, Heidelberg.

Lœvenbruck, H., Jannet, M. A. B., d'Imperio, M., Spini, M., and Champagne-Lavau, M. (2013). Prosodic cues of sarcastic speech in French: slower, higher, wider. Institut Universitaire de France, (August), 3537-3541.

Paeschke, A. (2004). Global trend of fundamental frequency in emotional speech. In *Speech Prosody 2004*, International Conference.

Paeschke, A., and Sendlmeier, W. F. (2000). Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion.*

Panpothong, Natthaporn. (1996). A pragmatic study of verbal irony in Thai. Doctoral dissertation, University of Hawai'i.

R Development Core Team (2008). R: A language and environment for statistical computing. [Computer program]. Version 4.0.0, retrieved from http://www.R-project.org.

Rockwell, P. (2007). Vocal features of conversational sarcasm. *Journal of Psycholinguistic Research*, 29(5): 361-369.

Searle, John R. (1991). Metaphor. In Steven Davis (ed.), *Pragmatics: A reader*, 519–39. Oxford and New York: Oxford University Press.

Sonboonta, Wanlapapaporn. (2010). An Acoustic Analysis of Thai Wh-word [?arai] in direct speech acts and indirect speech acts. Master thesis, M.Ed (Educational Linguistics). Bangkok: Graduate School, Srinakharinwirot University

Sperber, Dan, and Wilson, Dierdre. (1981). Irony and the Use-Mention Distinction. In Peter Cole (ed.), *Radical Pragmatics*, 295-318.

Tumtavitikul, A., and Thitikannara, K. (2006). The intonation of Thai emotional speech. In *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, New Zealand, (December), 6-8.

Wu, M. (2019). Effect of F0 contour on perception of Mandarin Chinese speech against masking. PloS one, 14(1).

Yimngam, S., Premchaisawadi, W., and Kreesuradej, W. (2011). Prosody analysis of Thai emotion utterances. In *International Conference on Application of Natural Language to Information Systems*, (June), 177-184. Springer, Berlin, Heidelberg.