



## Irony tones of voices

Maël Mauchand<sup>1</sup>, Nikolaos Vergis<sup>1</sup> and Marc D. Pell<sup>1</sup>

<sup>1</sup>McGill University, School of Communication Sciences and Disorders, Montreal, Canada

mael.mauchand@mail.mcgill.ca

### Abstract

While prosody is thought to play a major role in the production and comprehension of irony, the manner in which prosody is used to signal ironic intentions is still poorly understood. The complexity and variety of ironic interactions create divergences in the observations of irony production and interpretation, making the theoretical “ironic tone of voice” a challenging concept to define.

To examine the possibility of such a concept, acoustic and perceptual measurements were performed on literal or ironic criticisms and compliments. Our goal was to isolate cues specific to different attitudes conveyed and to relate these cues to the recognition and interpretation of particular attitudes.

The very accurate discrimination between literal and ironic utterances in the perceptual judgements contrasted with the diversity in prosodic strategies between and within each attitude. We found that ironic criticisms (sarcasm) could often be distinguished from literal compliments based on increased utterance duration and reduced pitch variability. However, none of the acoustic measures significantly predicted the distinction between ironic compliments (teasing) and literal criticisms. This asymmetry in the prosodic strategies, when related to the asymmetries in production and interpretation of ironies, highlighted the interdependence between prosodic consistency and functional interpersonal interactions in ironic speech.

**Index Terms:** speech prosody, non-literal communication, acoustics, sarcasm, teasing

### 1. Introduction

In everyday conversations, speakers are expected to be truthful and sincere in order to maintain functioning interactions [1]. For example, a sentence such as “You are a fantastic dancer” would convey nothing more than its explicit, literal meaning. Yet, speakers often intend to mean more, less, or even something else than what the content of their utterances suggest: “You are a fantastic dancer” could be used to mean that the person is actually terrible at dancing. In this case, the content of this sentence is not meant to be interpreted literally; the utterance can be defined as ironic. Irony is an indirect speech act in which what is meant is different from what is said; it is characterized by an incongruence between the context and the content of the utterance, from which the speaker is distancing himself [2]. To ensure that this incongruence is clear to listeners, speakers rely on a variety of strategies with more or less success. Assessing these strategies, such as the use of prosodic cues, can give insight into the production and interpretation of irony, and why interactions that center on ironic meanings sometimes fail.

#### 1.1. Ironic criticisms and compliments

In this study, we investigated the characteristics of two main categories of verbal irony: ironic criticisms and ironic compliments.

Ironic criticisms, generally referred to as *sarcasm*, are utterances that make use of a verbal compliment (i.e. positive content) in order to indirectly give a negative evaluation of the addressee (i.e. negative intent). It is accompanied by a context and/or non-contextual cues incongruent with the content, i.e. negative. For example, the sentence “You are a fantastic dancer” would be sarcastic if the dancer in question was undeniably terrible, and/or if the speaker displayed a facial expression or tone of voice that shows their negative intent to the interlocutor.

Ironic compliments, often called *teasing* (or banter, jocularly) work in a somewhat opposite manner: a verbal criticism (negative content) is used in order to indirectly and playfully praise the addressee, making use of positive contextual and/or non-contextual cues. This would be the case with a sentence like “You are a terrible dancer” uttered after a remarkable dance move, accompanied by friendly or playful facial and vocal cues.

#### 1.2. Asymmetry in production and interpretation of irony

The indirectness of irony leaves most of the meaning of an utterance up to the listener, making it a particularly challenging speech act to interpret, as ironic interactions are not as straightforward as theories suggest. Strategies of production and interpretation differ across ironic utterances, and some appear more successful than others. While sarcasm is clearly defined as negative [2]–[4], is often substituted for literal criticism [5], [6] and is accurately recognized and interpreted as unfriendly by listeners [4], [6]–[8], teasing is an much more delicate matter. Many speech acts with a structure similar to our definition of ironic compliments (verbal criticism with playful non-verbal cues), often also referred to as teasing, are used with opposite goals such as negative evaluation, social exclusion, provocation, etc. [9], [10]. Additionally, the production of ironic compliments are rarely preferred to literal compliments, with speakers even often choosing to remain silent when given the choice [5]. Finally, teasing is rarely interpreted as a compliment: listeners typically rate ironic compliments as very unfriendly, especially in contexts of non-solidary relationships [4], [7], [8]. This flagrant asymmetry between the relative simplicity of ironic criticism and the difficulties met with ironic compliments calls for further investigation into the structure of these speech acts.

#### 1.3. Ironic prosody

For conversations to proceed smoothly when using irony, the incongruence between content and context must be made very clear to listeners to allow them to understand the ironic attitude of the speaker and appraise their intent correctly. In many cases,

knowledge of the context is sufficient for this appraisal [3], [4], [7], [8], but speakers may also rely on additional non-contextual cues; in particular, an ironic tone of voice [2].

The variety of ironic utterances and their dependence on content make it difficult to investigate the possibility of a set of prosodic cues specific to irony. In fact, acoustic analyses of spontaneous ironic speech do not initially suggest any prosodic consistency across ironic utterances, at least when contextual cues are sufficiently obvious [11]. However, as context becomes more ambiguous, more distinct uses of prosody are made: greater prosodic contrast has been observed between non-ironic and ironic utterances when the former immediately preceded the latter [12], and in the case of sarcasm, lower pitch, slower speech rate and increased voice noise have been observed in acted productions of ironic speech [13]. Teasing prosody has been studied less and remains unclear. It is suggested to feature mostly exaggerating markers, which can take on many forms, such as elongated vowels, loud and rapid speech, laughter, singing voice; most changes seem to occur at the local level and to be highly context-dependent [10], [14], [15].

Rather than directly signaling the ironic attitude of the speaker, prosody is thought to be perceived as an additional incongruence with the content, an implicature suggesting a possible non-literal meaning [1], [16]. Although the concept of an ironic tone of voice is oversimplified, it is possible that some prosodic cues are used preferentially to signal a particular type of ironic incongruence (e.g., sarcasm). A recent study observed that listeners were extremely accurate in distinguishing ironic prosody from literal prosody, for both compliments and criticisms, without any other contextual cues [17], suggesting that there are indeed prosodic cues sufficient to identify irony. However, whether these cues vary across utterances or follow similar patterns when differentiating ironic/literal compliments from ironic/literal criticisms requires further analysis, which the present study aims to initiate through the observation of a few basic, global acoustic features.

## 2. Methods

### 2.1. Stimuli

The stimuli described below are the same as in a previous experiment by the authors; more extensive description of the methods can be found in this paper [17].

#### 2.1.1. Construction

Forty-eight sentences conveying judgements were constructed, all with identical grammatical form: “You are such a(n) *-adjective- -noun-*”. Nouns referring to jobs, activities or qualities (e.g. *guitarist*) each appeared once in two conditions (compliment/criticism). The distinction between compliment and criticism conditions was made by modulating the adjective (e.g. *amazing/horrible*), with 12 pairs of adjectives, each repeated in two sentences. The result was 24 unique compliments and 24 corresponding unique criticisms (*You are such a(n) amazing/horrible guitarist*).

#### 2.1.2. Recordings

Four speakers (2 males, 2 females, aged 25 to 50) with acting or public speaking experience were recorded uttering each sentence with two distinct intentions consecutively: first in a sincere, literal manner (literal compliments and criticisms), then in an insincere manner (sarcasm/ironic criticisms and

teasing/ironic compliments). This created four types of attitudes, referred to as *nice*, *sarcastic*, *mean* and *teasing*. No definition of sarcasm or teasing was given, speakers were only told what attitude should be conveyed by each utterance type (i.e. friendly for nice and teasing, unfriendly for mean and sarcastic); no models were provided to ensure more natural speech, but the verbal context was given during practice and in rare cases of difficult utterance production. Speakers were free to repeat utterances, and the most suitable repetitions were selected by the examiner (in the majority of cases the last one was kept). This yielded 96 productions for each attitude, and 384 utterances total. Stimuli were precisely segmented at onset/offset using Praat [18] and each .wav audio file was normalized to a peak intensity of 70dB to control for differences in sound recording levels.

#### 2.1.3. Validation and selection

Stimuli were perceptually validated by 20 online participants (10 males, 10 females) in order to select utterances that represented the best exemplars of target intentions. Participants were asked to judge the literality (“Does the speaker mean what they say?”) as well as the positivity (“Is the attitude of the speaker positive?”) of the speaker for each utterance, on a 5-point scale from *Not at all* to *Very much*. These ratings were then used to determine the utterances for which the combination of ratings of literality and positivity corresponded best to the prediction of the intent (e.g. for sarcasm, low literality and positivity). For each speaker, the five best token sets (all 4 attitudes for a given sentence model) that maximized the difference in ratings between literal and ironic counterparts were selected (4 speakers x 4 intentions x 5 items = 80 utterances).

## 2.2. Data collection

### 2.2.1. Acoustic measures

For each utterance, the following acoustic measures were computed:

- Duration (s): since every sentence was repeated in both literal and ironic manner, a simple measure of duration instead of speech rate was deemed to be sufficient to distinguish attitudes.
- Mean F0 (Hz, converted to z-scores): the conversion to z-scores was performed in order to control for natural differences in pitch between speakers (in particular, between males and females).
- Standard deviation of F0 (Hz): computed around the mean F0 to assess pitch variability.
- Intensity variation (dB): computed around the mean amplitude to determine intensity variability. Note that because the sounds were normalized, mean intensity was not considered in the measurements.
- Mean harmonics-to-noise ratio (HNR, in dB): computed on the basis of a forward cross-correlation analysis with time steps of .01s and 4.5 periods per window.

These measures are based on the methods of a previous experiment on acted sarcasm recordings [13] and constitute a preliminary exploration of possible global prosodic features of irony.

### 2.2.2. Perceptual measures

Forty participants (20 males, 20 females, aged 18 to 35), all native English-speaking Canadians, were recruited online via Prolific (www.prolific.ac) [19] to judge the selected stimuli. For each of the 80 utterances, subjects were asked to rate the friendliness of the speaker (“How friendly is the speaker?”) on a 5-point scale from *Not at all* to *Very much*. Ratings of friendliness were chosen in order to obtain natural and relevant impressions of speaker’s intent without directly pointing out the literal/ironic dimension of each utterance.

## 3. Results

Stepwise discriminant analyses were conducted for each type of content, positive (*You are such an amazing guitarist*) and negative (*You are such a horrible guitarist*), to predict whether an utterance with such content was literal or ironic. Predictors were the five acoustic measures listed above.

### 3.1. Positive content

When content was positive, perceptual ratings showed that listeners accurately discriminated irony from literality, rating sarcastic utterances (-.299, SD = .565) significantly less friendly than nice utterances (.843, SD = .342) [17]. This distinction was accurate and significant for all sentence tokens (see Figure 1).

The stepwise function added duration as a predictor in the first step ( $F(1, 38) = 18.904, p < .001$ ), and pitch variation in the second step ( $F(2,37) = 25.679, p < .001$ ). No additional predictors were added. The resulting function was significant (Wilk’s  $\lambda = .419, X^2 = 32.208, df = 2$ , canonical correlation = .762,  $p < .001$ ) and correctly predicted 82.5% of cross-validated group cases, with 75% of nice utterances predicted as nice and 90% of sarcastic utterances predicted as sarcastic (see Table 1).

Table 1: Classification results of the discriminant function for positive content

		Predicted group membership		Total
Attitude		Nice	Sarcastic	
Count	Nice	15	5	20
	Sarcastic	2	18	20
%	Nice	75	25	100
	Sarcastic	10	90	100

Table 2: Correlations of the predictors with the discriminant function and coefficients of used predictors (Scores at group centroids: Nice -1.148, Sarcastic 1.148)

Predictor	Structure matrix	Standardized canonical coefficients
Duration	.599	.955
F0(z-score) <sup>1</sup>	-.254	
F0 variation	-.489	-.877
Amplitude variation <sup>1</sup>	.271	
HNR <sup>1</sup>	-.029	

1. Variable not used in the analysis

The canonical coefficients and the structure matrix presenting the correlation between the predictors and the function, shown in Table 2, suggest that sarcastic utterances were predicted by a longer duration and reduced pitch variation.

When individual speaker patterns were qualitatively inspected, an interesting finding not represented in the discriminant analysis was the use of pitch by the two female speakers. For both of them, pitch z-scores were consistently positive for nice utterances and negative for sarcastic utterances, suggesting that reduction of pitch was a crucial strategy for their sarcastic speech, in contrast to male speakers. This possibility could be further tested in a larger group of male and female speakers.

### 3.2. Negative content

When content was negative, perceptual ratings showed that listeners still accurately discriminated irony from literality, with teasing utterances (-.387, SD = .622) rated more friendly than mean utterances (-1.07, SD = .295) [17]. Moreover, Figure 1 shows that both attitudes were correctly distinguished for all sentence tokens. This suggest that prosodic strategies produced efficient distinction of the attitude used. Some strategies appeared to be more efficient than others: it has been noticed that about a fifth of the teasing utterances used laughter to convey a playful attitude. These utterances were all rated significantly friendlier than the rest of the teasing stimuli.

However, no significant function was extracted by the stepwise discriminant analyses, for none of the potential predictors significantly improved the default function ( $F < 3.84$ ). This means that none of the prosodic features measured in the stimuli consistently changed with the literality or irony of the speakers.

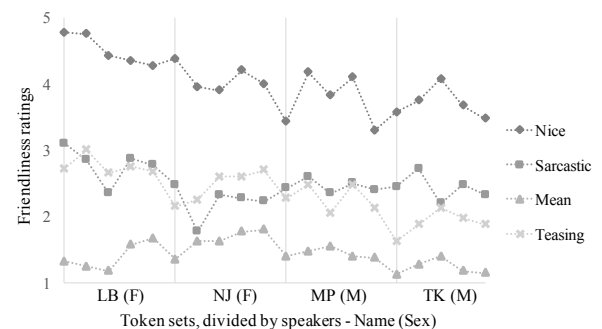


Figure 1: Average friendliness ratings for each utterance of each token set (How friendly is the speaker? 1=Not at all, 5=Very much). Each line groups all the utterances corresponding to one attitude; vertically aligned utterances belong to the same token set.

## 4. Discussion

### 4.1. Prosodic ironies

As expected, prosodic strategies largely differed between ironic compliments and ironic criticisms. The opposite nature of their intent (the former friendly, the latter unfriendly) already suggested that irony could not be characterized by one “tone of voice”, but even intent did not correlate with the use of prosodic cues: friendly intent did not differ from unfriendly intent in the same manner, regardless of whether the content was positive or negative. While ironic criticism seemed to be distinguished from literal compliments by longer utterances and a reduced

pitch variation, ironic compliments were not differentiated from literal criticisms using the same cues. These findings suggest that prosody may not work as a direct unidimensional indicator of the speaker's attitude or intent, but as a more complex, content- and context-dependent implicature from which the listeners are expected to infer the appropriate attitude [16], [20]. However, the use of syntactically consistent and semantically similar stimuli allowed to investigate the possibility of a corresponding consistency in prosody for each attitude.

#### 4.2. Ironic prosodies

Prosodic strategies did not only differ between each type of irony, but also across ironic utterances. Although the model described by the discriminant analysis for positive content was quite accurate, it still failed to predict the attitude of almost a fifth of the stimuli. Additionally, individual differences were noted, with for example a preferential use of pitch by females to distinguish sarcasm. This preliminary distinction could be explored in future works to assess more widely and precisely the dependence of prosodic strategies on sex and other individual differences. However, these differences did not significantly affect the recognition of intent, since listeners consistently distinguished irony from literality, which may suggest that conveying irony does not necessarily rely on the recurrent use of a specific set of the cues we examined; more extensive analysis of the stimuli and their features is needed to confirm this idea. These results show some overlap with previous findings: slower voice and lower pitch had been previously identified as sarcastic features [13]. However, divergences exist: lower pitch was reported to occur in sarcastic male speech, and voice noise has been shown to be increased in sarcasm [13]. As such, the findings presented herein are among the first to elucidate the role of pitch variation in ironic speech. Previous studies of irony primarily used stimuli produced in very diverse contexts, far from our controlled utterances [11]–[13]. As such, the resulting acoustic differences may have contributed to the context-dependent observations reported in prior studies of ironic prosody.

The variety of prosodic strategies in irony becomes even more obvious when looking at ironic compliments. None of the prosodic features measured successfully distinguished teasing from mean utterances; yet, in all cases, listeners accurately perceived the difference between the two attitudes across utterances with identical content. Recurrent global variations of the tone of voice may not be the main signals of teasing; the discrepancies across stimuli, as well as the occasional use of laughter, could indicate preferences in more local variations, as hypothesized and reported in prior studies [11], [12]. The restricted set of observed features limits our capacity to assert this strongly: additional investigations will be required.

The nature of irony makes it difficult to define a prototypical prosodic pattern used in the conveyance of ironic attitude. Since the key is to produce an implicature through incongruent production of the utterance, the strategies can vary depending on the speaker, the speaker's intent, the context, the type of sentence, etc. However, the degree to which irony production varies is not the same for all types of irony: sarcasm, for example, is shown to be easier to define with global prosodic cues, while teasing strategies are harder to assess with the same measurements.

#### 4.3. Ironic asymmetries

Although ironic utterances were all correctly distinguished from their literal counterpart, an asymmetry similar to previous

findings was observed [4], [7]. While ratings of ironic criticisms were concordant with the intended attitude (i.e. unfriendly), ironic compliments were rated much less friendly than could have been expected by the speaker; in fact, the perceived friendliness of teasing did not differ significantly from sarcasm. These results, frequently observed in prior studies [4], [7], [8], cannot be explained by a misunderstanding of the prosodic cues, since directed attention to these cues yield very accurate recognition of their friendly nature [17]; even the explicit incongruent context is not sufficient to facilitate a more friendly interpretation of teasing [4]. Recent evidence suggests that the source of the asymmetry actually arises from the nature of the content: negative content appears to be given more weight in the mental representation of a speaker's attitude, thus diminishing the importance of prosodic cues for the interpretation of intent [17]. The high risks to face value that a verbal criticism represents may lead the listener towards a more negative, but safer judgement.

This bias towards negative content may also explain the differences observed in the production of ironic speech. It is possible that ironic criticisms, using non-threatening, positive content, can be used more freely and frequently without worrying that the content will have an impact on the interpretation of intent. This can help explain why sarcasm can be used in more interpersonal contexts with a lower risk of relationship damage [5]. Hence, speakers can use more stereotypical, global sets of cues that depends less on the content, whereby the frequent use of these cues may facilitate the general concept of a sarcastic tone of voice and the interpretation of sarcasm in everyday conversation. However, ironic compliments may require much more careful manipulations of prosody to directly attenuate the hostility of a given content. Exaggerations on certain words, syllables, local modulations of pitch, as well as laughter can be teasing cues [10], but cannot be used in a consistent manner across syntactically and semantically different utterances. Additionally, the infrequent use of ironic compliments prevents the development of a popular consensus on what should teasing sound like. Even the use of laughter (or other apparently friendly cues) could actually be thought of as mocking instead of friendly, depending on the context and the speaker [9], [21]. Such misunderstandings can be avoided if the speaker and listeners are well acquainted: thorough knowledge of the speaker's general attitude and prosodic habits, coupled with a strong solidary relationship, could attenuate the negative effect of content and allow accurate recognition of teasing intent [8].

### 5. Conclusion

Irony can take on various forms, whether at the intentional, contextual or prosodic level. This variability, as well as the high risk associated with the use of certain forms or irony, may prevent proper agreement on what defines an utterance as ironic, thus leading recipients of ironic speech acts to adopt a defensive stance and focus on the threatening aspects of it. This defensive strategy in turn may lead to infrequent use of these forms of irony and prevent them to be universally defined with global, easily measurable paralinguistic cues.

### 6. Acknowledgements

We wish to thank Dr. Xiaoming Jiang and Dr. Jonathan Caballero for helpful advice on the methods of the experiment, and Deirdre Michael Truesdale for reviewing drafts of this paper.

## 7. References

- [1] H. P. Grice, "Logic and Conversation," in *Studies in the Way of Words*, 1989, pp. 1–13.
- [2] D. Wilson, "Irony, Hyperbole, Jokes and Banter," *Form. Model. Study Lang.*, no. January, pp. 1–8, 2017.
- [3] R. W. Gibbs and H. Colston, *Irony in language and thought: A cognitive science reader*. 2007.
- [4] P. M. Pexman and K. M. Olineck, "Does Sarcasm Always Sting? Investigating the Impact of Ironic Insults and Ironic Compliments Penny," *Discourse Process.*, vol. 33, no. 3, pp. 199–217, 2002.
- [5] J. K. Matthews, J. T. Hancock, and P. J. Dunham, "The Roles of Politeness and Humor in the Asymmetry of Affect in Verbal Irony," *Discourse Process.*, vol. 41, no. 1, pp. 3–24, 2006.
- [6] R. W. Gibbs, "Irony in Talk Among Friends," *Metaphor Symb.*, vol. 15, no. 1, pp. 5–27, 2000.
- [7] S. Dews, J. Kaplan, and E. Winner, "Why Not Say It Directly? The Social Functions of Irony," *Discourse Process.*, vol. 19, no. August, pp. 347–367, 1995.
- [8] P. M. Pexman and M. Zvaigzne, "Does irony go better with friends?," *Metaphor Symb.*, vol. 19, no. 2, pp. 143–163, 2004.
- [9] V. Sinkeviciute, "What makes teasing impolite in Australian and British English? 'step[ping] over those lines [...] you shouldn't be crossing,'" *J. Politeness Res.*, vol. 13, no. 2, pp. 175–207, 2017.
- [10] D. Keltner, L. Capps, A. M. Kring, R. C. Young, and E. A. Heerey, "Just Teasing: A Conceptual Analysis and Empirical Review," *Psychol. Bull.*, vol. 127, no. 2, pp. 229–248, 2001.
- [11] G. A. Bryant and J. E. Fox Tree, "Is there an ironic tone of voice?," *Lang. Speech*, vol. 48, no. 3, pp. 257–277, 2005.
- [12] G. A. Bryant, "Prosodic Contrasts in Ironic Speech," *Discourse Process.*, vol. 47, no. 7, pp. 545–566, 2010.
- [13] H. S. Cheang and M. D. Pell, "The sound of sarcasm," *Speech Commun.*, vol. 50, no. 5, pp. 366–381, 2008.
- [14] J. K. Alberts, Y. Kellar-Guenther, and S. R. Corman, "That's not funny: Understanding recipients' responses to teasing," *West. J. Commun.*, vol. 60, no. 4, pp. 337–357, 1996.
- [15] M. Haugh, "Jocular Mockery as Interactional Practice in Everyday Anglo-Australian Conversation," *Aust. J. Linguist. ISSN*, vol. 34, no. 1, pp. 76–99, 2014.
- [16] A. Wichmann, "Attitudinal intonation and the inferential process," *Speech Prosody 2002. Proc. 1st Int. Conf. Speech Prosody*, pp. 11–16, 2002.
- [17] M. Mauchand, N. Vergis, and M. D. Pell, "Irony, prosody, and social impressions of affective stance," *Rev.*
- [18] P. Boersma and V. van Heuven, "Speak and unSpeak with Praat," *Glott Int.*, vol. 5, no. 9–10, pp. 341–347, 2001.
- [19] E. Peer, L. Brandimarte, S. Samat, and A. Acquisti, "Beyond the Turk: Alternative platforms for crowdsourcing behavioral research," *J. Exp. Soc. Psychol.*, vol. 70, pp. 153–163, 2017.
- [20] A. Wichmann, "The attitudinal effects of prosody, and how they relate to emotion," *ISCA ITRW Speech Emot.*, pp. 143–148, 2000.
- [21] J. M. Bollmer, M. J. Harris, R. Milich, and J. C. Georgesen, "Taking Offense: Effects of Personality and Teasing History on Behavioral and Emotional Reactions to Teasing," *J. Pers.*, vol. 71, no. 4, pp. 557–603, 2003.