

# THE PRODUCTION AND PERCEPTION OF IRONY IN SHORT CONTEXT-FREE UTTERANCES

*Astrid Nauke & Angelika Braun*

University of Trier, Germany

nauke@uni-trier.de; brauna@uni-trier.de

## ABSTRACT

The present study examines single-word-utterances and focuses on (1) production patterns of irony as opposed to sincerity and (2) listeners' ability to distinguish between the two without any contextual information. The results of the acoustic measurements show a lower average F0, less F0 variability, a smaller F0 range, lower intensity and longer durations for the ironic stimuli. The perception experiment resulted in overall recognition rates of more than 80%, the sincere utterances being identified significantly better than the ironic ones. Female listeners tend to be better at recognizing irony in female speakers than are male listeners.

**Keywords:** irony, sarcasm, sincerity, vocal cues

## 1. INTRODUCTION

Irony is a highly complex phenomenon which can hardly be explained or defined satisfactorily and exhaustively [1, 8, 9, 10]. However, what is generally considered to be an important element of irony is a certain discrepancy between meaning and wording [5, 9] as well as the expression of a (mostly negative) emotion or attitude [1, 4, 8, 11, 12] in being ironic.

Another problem in analyzing ironic speech is the nature of irony. The two terms 'irony' and 'sarcasm' have been used synonymously in some previous studies because subjects were found to have only vague ideas of the meaning of 'irony', whereas the notion of 'sarcasm' was more familiar and intuitive to them [2]. In other research, however, explicit mention is made that 'sarcasm' as one form of 'irony' is examined [4, 11]. The results of [1] point to the fact that there are substantial differences in the vocal features between 'kind irony' ("praise by blame") and 'sarcastic irony' ("blame by praise"). The focus of the present study is on sarcastic irony.

Previous studies have demonstrated that the best way for listeners to distinguish between irony and sincerity is co(n)text [6, 8, 9, 12]. The term

co(n)text refers to the linguistic setting, while context refers to the situational setting.

There are numerous ways to signal irony, i.e., gestures, facial expression, as well as vocal cues [6, 8, 13]; but a classification or some kind of typology of 'irony signals' is impossible [14]. Any form of contrast or incongruity of a remark with its co(n)text will trigger the listener to challenge the meaning of an utterance. As long as the listener is aware of that discrepancy between reality and expectation [5], anything can serve as an indicator [5, 10].

In the forensic setting, there is often no co(n)text available and some decisions must be made on a single word like "yes" or "no". Certain studies examining acoustic features show that recipients rely mostly on pitch and duration when facing the decision whether an utterance is sincere or not [1, 3, 4, 11, 13]. Ironic utterances are generally longer [1, 4, 11, 13] and the decreased tempo might be an even more distinct cue if the ironic stimulus is very short [4].

Regarding F0 in sarcastic irony researchers report both a lower mean, less variability and a smaller range [1, 4, 7], and a higher mean F0 and more variability [3, 13]. Intensity measurements showed either inconsistent differences or no differences at all [1, 3, 4, 11] between the two. Voice quality as a vocal cue has been largely neglected, but there are observations reporting creak or vocal fry [7, 13] and larger amounts of noise [4] in ironic speech. Those contradictory findings may arise from the varied research parameters in regards to speakers, speech material, language examined and irony type.

The present study attempted to address the following research questions:

1. How is irony coded in one-word-utterances?
2. Is irony detectable on the basis of context-free one-word-utterances?

## 2. MATERIALS AND METHODS

### 2.1. Subjects

10 female speakers aged 21–25, with a mean of 22.3 years, volunteered to produce the utterances for the study. A group of 28 listeners (22 female, 6 male) took part in the perception test. They ranged in age from 20–29 years with a mean of 23.7 years. All subjects were native Germans and had no known speech-language-hearing pathologies.

### 2.2. Stimuli

The 10 female speakers were asked to produce 10 different one-word-utterances both in a sincere and ironic manner. The target words have no more than two syllables and occur literally as well as ironically in general language use. They were presented in brief scenarios either evoking a sincere or an ironic reaction (Table 1).

**Table 1:** Sample scenario: ‘Lecker’ (‘Tasty’) in a sincere and an ironic context.

Sincere	Ironic
Paul and Paula are going out to eat. They are looking at the menu. Paul remarks: “Oh look, they’ve got your favourite pasta!”	Paul and Paula are going out to eat. They are looking at the menu. Paul remarks: “Oh look, they’ve got snails and frog’s legs!”
Paula replies: “Tasty!”	Paula replies: “Tasty!”

In addition to *Lecker* (‘tasty’) the words *Danke* (‘thank you’), *Klar* (‘sure’), *Klasse* (‘great’), *Nett* (‘nice’), *Schön* (‘fine’), *Spitze* (‘great’), *Super* (‘super’), *Toll* (‘wonderful’) and *Wahnsinn* (‘awesome’) were used.

Speakers were recorded individually directly onto a PC (stimuli were sampled at 44.1 kHz, 16 bit, mono). Speakers were instructed to produce natural-sounding speech, not a caricature. Each target word was repeated until the speakers were satisfied with the result.

A total of 200 stimuli (100 sincere, 100 ironic) were thus produced by the speakers.

### 2.3. Procedure

The stimuli were subjected to acoustic analysis using Praat 5.0.30. Various measurements regarding fundamental frequency, intensity, duration and voice quality (jitter, shimmer) were carried out. For the listening test a recording was prepared which contained all 200 stimuli in random order. Each stimulus was presented twice with a 0.5 second pause in between, followed by a

2 second pause between stimuli. After every 10th stimulus pair the 2 second pause also contained a 440 Hz tone in order to provide some orientation for the listeners. The recording was played to the listeners over loudspeakers in a quiet room. Their task was to decide while listening whether an utterance was sincere or ironic.

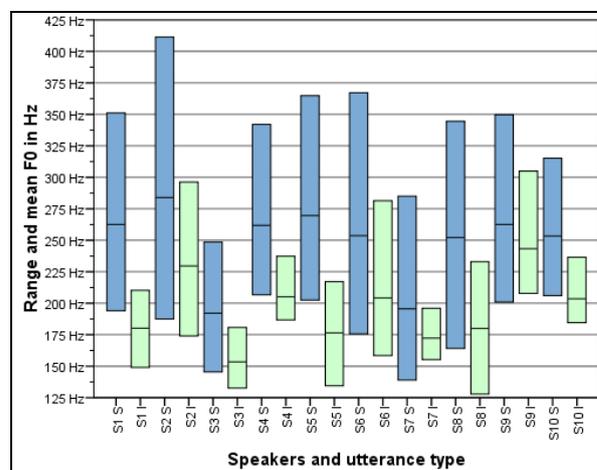
## 3. RESULTS

### 3.1. Acoustic analysis

#### 3.1.1 Fundamental frequency

The means of the F0 parameters, i.e. mean, SD, min, max and range – show that almost all values of the ironic stimuli are lower than those of the sincere ones. Sole exceptions are the minimum F0 for speakers 7 and 9. Figure 1 illustrates the individual differences between speakers and speech mode and demonstrates the tendency to produce the ironic stimuli with a lower F0, less modulation and a smaller F0 range.

**Figure 1:** Range and mean F0 (in Hz) for each speaker. The range is represented as a bar between min and max values. The horizontal lines dividing the bars indicate the mean. Dark colour and capital ‘S’ behind speaker numbers represent sincere utterances, accordingly the ironic ones are marked by light colour and capital ‘I’.



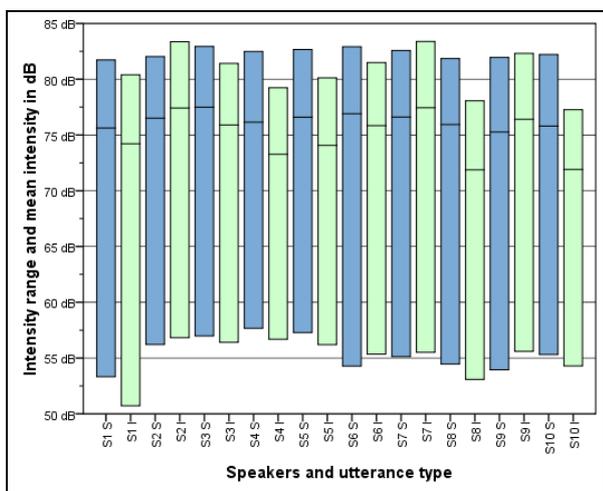
The statistical analysis confirms this impression. Even in this small sample the differences between the means of the F0 parameters for ironic and sincere stimuli are highly significant ( $p = 0.000$  for mean F0, SD, max, range;  $p = 0.023$  for min).

#### 3.1.2 Intensity

Regarding the intensity it can be seen (Figure 2) that the differences between sincere and ironic

means are not as distinct as for fundamental frequency. Still, the mean results for the ironic stimuli are mostly lower or almost equal to those of the sincere utterances. The paired t-tests showed significant values ( $p < 0.05$ ) for mean, SD, max and range but not for the intensity minimum ( $p = 0.364$ ).

**Figure 2:** Intensity range and mean intensity (in dB) for each speaker. The range is represented as a bar between min and max values. The horizontal lines dividing the bars indicate the mean. Dark colour and capital ‘S’ behind speaker numbers represent sincere utterances, accordingly the ironic ones are marked by light colour and capital ‘I’.



3.1.3 Duration and voice quality

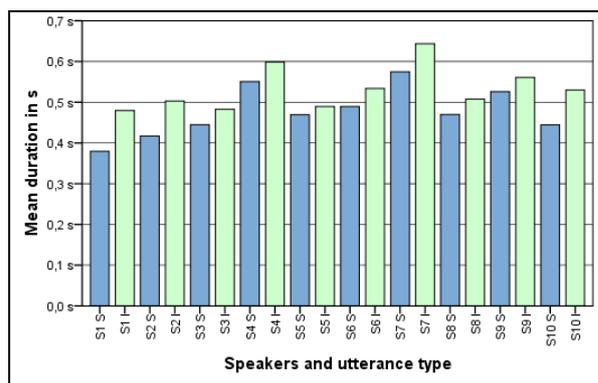
The results of the individual jitter and shimmer values with respect to the voice quality are rather varied (Table 2). Overall the means of both parameters in the ironic utterances are lower than those in the sincere ones, but the differences are not significant (jitter:  $p = 0.171$ , shimmer:  $p = 0.200$ ).

**Table 2:** Duration, jitter and shimmer: means and associated standard deviations in parentheses of sincere and ironic utterances for each speaker (S1 to S10).

		S1	S2	S3	S4	S5
Duration	sincere	0,38 (0,1)	0,42 (0,2)	0,45 (0,1)	0,55 (0,1)	0,47 (0,1)
	ironic	0,48 (0,1)	0,50 (0,1)	0,48 (0,1)	0,60 (0,1)	0,49 (0,1)
Jitter rap	sincere	1,35 (0,6)	0,69 (0,4)	1,23 (0,7)	0,57 (0,2)	0,79 (0,4)
	ironic	0,89 (0,7)	0,57 (0,2)	0,90 (0,5)	0,69 (0,3)	0,73 (0,3)
Shimmer apq11	sincere	8,54 (3,3)	11,50 (2,6)	8,28 (3,7)	7,86 (1,9)	8,57 (2,8)
	ironic	9,16 (3,9)	9,96 (2,4)	8,50 (2,5)	7,57 (1,6)	6,59 (3,1)
		S6	S7	S8	S9	S10
Duration	sincere	0,49 (0,1)	0,57 (0,1)	0,46 (0,1)	0,53 (0,1)	0,44 (0,1)
	ironic	0,53 (0,1)	0,64 (0,2)	0,50 (0,1)	0,56 (0,1)	0,53 (0,1)
Jitter rap	sincere	0,64 (0,2)	2,12 (1,5)	0,97 (0,6)	0,79 (0,4)	0,77 (0,3)
	ironic	0,65 (0,2)	0,86 (0,8)	1,35 (0,8)	0,58 (0,2)	0,63 (0,2)
Shimmer apq11	sincere	7,94 (2,8)	8,28 (2,8)	8,56 (2,8)	8,67 (1,6)	7,94 (1,4)
	ironic	7,96 (2,6)	5,99 (2,6)	9,61 (3,0)	8,15 (1,6)	7,81 (2,1)

The mean duration of the one-word-utterances ranges between 0.38 s and 0.57 s for sincerity and 0.48 s and 0.64 s for irony (Table 2). Each speaker produced the ironic stimuli with longer durations than the sincere ones (Figure 3). Speaker 5 has the smallest difference (0.02 s) while speaker 1 has the largest (0.1 s). The significance level for this parameter is again very high ( $p = 0.000$ ).

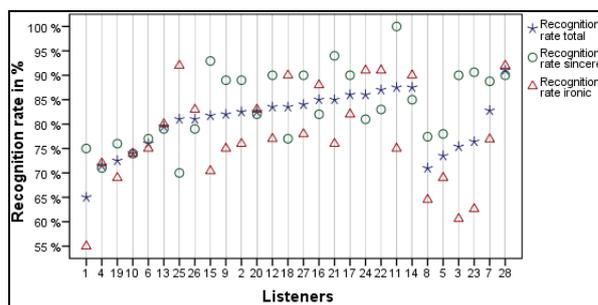
**Figure 3:** Mean duration (in s) for each speaker. Dark colour and capital ‘S’ behind speaker numbers represent sincere utterances, accordingly the ironic ones are marked by light colour and capital ‘I’.



3.2. Perception experiment

Of the 5600 stimuli rated in total (28 listeners x 200 stimuli) 5553 could be subjected to statistical analysis. The absent 47 stimuli were either rated not clearly or not at all. The individual recognition rates for all stimuli and subdivided into sincere and ironic ones are illustrated in Figure 4.

**Figure 4:** Listener recognition rates in ascending order of their means: total (N = 200), sincere (N = 100) and ironic (N = 100). The male listeners (No. 3, 5, 7, 8, 23, 28) are grouped together on the right.



As can be seen, listener performance ranges from 65% (listener 1) to 91% (listener 28) correct. The diagram also shows obvious differences in the identification rates for irony and sincerity in each listener. Overall, the sincere utterances were identified significantly better than the ironic ones ( $p = 0.018$ ). Furthermore there was no correlation

between listeners' ability to identify sincere and ironic utterances, i.e. those who were good at one task did not necessarily do well at the other.

The average total recognition rates and the same separated for the two speech modes (Table 3) show that 80.5% of all stimuli were identified correctly. Regarding a potential gender difference in the distinction ability it can be seen that the female listeners were slightly, but not significantly better than the males ( $p = 0.341$ ).

**Table 3:** Average recognition rates total, sincere and ironic for all listeners and grouped by gender (N female = 22, N male = 6).

	Recognition rate total	Recognition rate sincere	Recognition rate ironic
Total	80,5 %	83,6 %	77,4 %
Female	81,1 %	83,0 %	79,2 %
Male	78,3 %	85,8 %	70,9 %

Looking at the correct identification of sincere utterances only, the performance of the male listeners is slightly better than that of the females, but that difference does not reach significance either ( $p = 0.437$ ). The gender difference between the recognition rates of the ironic stimuli, however, is more distinct. The percentage of 79.2% for the females as compared to 70.9% for the males reaches a significance level of  $p = 0.074$ .

#### 4. DISCUSSION

Owing to the fact that previous studies focusing on vocal cues of irony differ in material (sentences, utterances, single words), method (posed vs. spontaneous), examined language and subjects (female and/or male), the discussion is not easy. The results of the present study confirm previous findings that ironic speech differs from literal or sincere speech in pitch, intensity and duration. It is widely agreed [1, 4, 11, 13] that ironic utterances have a longer duration and this study again provides support. The present findings of significantly lower values in all F0 parameters for ironic utterances are in line with previous research [1, 4, 7]. On the other hand, our finding that ironic utterances contain a significantly lower amplitude level, standard deviation and range than sincere ones contradict previous findings [1, 3, 4, 11]. A possible explanation for this is that subjects actually succeeded in producing very natural-sounding stimuli which may have resulted in a very subtle coding of irony.

The high overall recognition rates of 80.5% which are actually very similar to those reported in [13], indicate that listeners were nonetheless

provided with quite distinct cues on which to base the identification task. The observed gender difference in recognition rates will have to be tested against speech material produced by male speakers. As shown in [1] it seems appropriate to look at other types of irony in more detail and to distinguish between an emphatic and a more subtle coding of irony.

#### 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] Bryant, G., Fox Tree, J. 2002. Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol* 17(2), 99-117.
- [3] Bryant, G., Fox Tree, J. 2005. Is there an ironic tone of voice? *Language and Speech* 48(3), 257-277.
- [4] Cheang, H., Pell, M. 2008. The sound of sarcasm. *Speech Communication* 50(5), 366-381.
- [5] Colston, H., O'Brien, J. 2000. Contrast and pragmatics in figurative language: Anything understatement can do, irony can do better. *Journal of Pragmatics* 32, 1557-1583.
- [6] Cutler, A. 1974. On saying what you mean without meaning what you say. *Chicago Linguistic Society: Papers from the Tenth Regional Meeting*. Chicago: University of Chicago, 10, 117-127.
- [7] Fónagy, I. 1971. Synthèse de l'ironie. *Phonetica* 23, 42-51.
- [8] Hartung, M. 2002. *Ironie in der Alltagssprache – Eine gesprächsanalytische Untersuchung*. Radolfzell: Verlag für Gesprächsforschung.
- [9] Haverkate, H. 1990. A speech act analysis of irony. *Journal of Pragmatics* 14, 77-109.
- [10] Lapp, E. 1992. *Linguistik der Ironie*. Tübingen: Narr.
- [11] Rockwell, P. 2000. Lower, slower, louder: Vocal cues for sarcasm. *Journal of Psycholinguistic Research* 29(5), 483-495.
- [12] Rosengren, I. 1986. Ironie als sprachliche Handlung. In Sitta, H. (ed.), *Sprachnormen in der Diskussion. Beiträge vorgelegt von Sprachfreunden*. Berlin/New York: de Gruyter, 41-71.
- [13] Schaffer, R. 1981. Are there consistent vocal cues for irony? In Masek, C.S., Hendrick, R.A., Miller, M.F. (eds.), *Papers from the Parasession in Language and Behaviour*. Chicago: Chicago University Press, 204-210.
- [14] Warning, R. 1976. Ironiesignale und ironische Solidarisierung. In Preisendanz, W., Warning, R. (eds.), *Das Komische*. München: Fink, 416-423.