ACOUSTIC CHARACTERISTICS OF VERBAL IRONY IN STANDARD AUSTRIAN GERMAN

Hannah Leykum[1]

[1]Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria
hannah.leykum@oeaw.ac.at

ABSTRACT

The present study investigates the use of acoustic irony markers by speakers of Standard Austrian German (SAG), a variety of German which has not yet been investigated with regard to irony marking. Short utterances were elicited in an ironic manner (ironic criticism) as well as expressing the positive literal meaning. Acoustic analyses of recordings of eight speakers of SAG revealed that ironic utterances are mainly characterised by a lower F0 with less variability (standard deviation and range of F0), a lower intensity and longer durations compared to the literal counterparts. Moreover, F1 and F2 frequencies of the stressed vowel are lower in ironic utterances compared to literal realisations of the same utterances leading to the assumption that speakers smile less when being ironic. In addition, some gender, task, and utterance specific differences occurred.

Keywords: verbal irony, sarcasm, ironic criticism, Standard Austrian German

1. INTRODUCTION

Irony is frequently used in interpersonal communication. Most speakers apply disambiguating cues highlighting the ironic intent of utterances to avoid misunderstandings. These cues can be verbal, non-verbal (e.g. visual) and / or paraverbal [21]. Paraverbal cues are frequently used mark irony [4], especially when speaker and listener are not very familiar with each other or when non-verbal cues are not available. The present study investigates the paraverbal features of ironic utterances (ironic criticism) in Standard Austrian German.

Acoustic characteristics of verbal irony have been investigated in some languages and language varieties. The studies lead to the conclusion that more or less the same parameters (especially F0, intensity, and duration) are used to highlight irony, however, the specific use of the cues differ between different languages. Especially concerning the use of F0 cues, language specific differences exist: In Italian [2], French [8, 12], Cantonese [6], and Japanese [1] the mean F0 is higher in ironic realisations of utterances and most of the studies found a larger F0 range and / or SD of F0 in ironic utterances. In addition, in French [8], ironic utterances are characterised by a rising F0 contour. In contrast to the aforementioned results, in several varieties of English [5, 7, 20], and in German in Germany [17, 21, 22] mean F0 values are lower in ironic utterances. Results on F0 variation and F0 contour differences were inconsistent. Some studies [5, 7, 17, 22] found a lower F0 variation or a flatter F0 contour in ironic utterances and some [20, 21] found no significant differences between ironic and literal utterances.

As some studies suggest there might be language or language type specific differences in the use of prosodic cues to mark irony. If the language type determines how F0 cues are used to highlight irony, the investigation of Standard Austrian German (SAG) is of special interest, since SAG is classified as a mixed-type language with word language characteristics as well as characteristics of quantifying languages [14]. In contrast, German in Germany is classified as a true word language [25].

Based on results on German in Germany [16, 17, 21, 22] and results of a pilot study with one speaker of SAG [10], the present study hypothesises that ironic utterances of speakers of SAG are characterised by longer durations, a lower mean F0, a smaller F0 variation + range, a lower intensity, a higher Harmonics-to-Noise Ratio (HNR), larger amplitude differences between the first and second harmonic (H1-H2), and hyperarticulated vowels. However, deviations from the expected results concerning the F0 measurements will not be surprising, since language variety specific differences are likely to exist.

2. METHODS

2.1. Recordings

All stimuli were either disyllabic one-word utterances or utterances consisting of two monosyllabic words with a positive meaning when realised literally (e.g.“Super!” ‘*super*’; “Sehr gut!” ‘*very good*’). All utterances are stressed on the first syllable / word. In a pre-test [10] the items were rated concerning their predominant manner of use and concerning the frequency of use by Austrians. These

ratings were conducted to have criteria to choose the items and to exclude non-familiar utterances and / or utterances which do not need to be marked acoustically as being ironic or literal, because the item itself is inherently ironic or literal.

For the recording session, the utterances were embedded in short scenarios evoking either a literal or an ironic realisation of the target word (following [16, 22]). All scenarios eliciting literal target utterance had a positive connotation, since neutral realisation of most of the target words are quite uncommon. An example scenario is given in Table 1. In the first part of the recordings, the speakers were instructed to put themselves in the position of person B and answer to the scenario accordingly (henceforth called scenario-condition). In the second part, the speakers were explicitly instructed to respond in an ironic or literal manner to ensure that they realise the utterances in the intended manner (henceforth called explicit-condition).

Recordings of eight speakers (23-32 years, balanced for gender) of SAG were conducted in a sound booth (IAC-1202A) with a cardioid microphone (AKG C451 EB). All SAG speakers were, as defined by [13, 15], born and raised in Vienna, had a high education level, and had at least one parent fulfilling the same criteria.

In addition to the material analysed in the present study, electroglottographic (EGG) and video recordings were conducted for further analyses.

Table 1: Example scenarios eliciting "Danke!" ('*thanks*') in an ironic and a literal manner.

Ironic	Literal
Person A gibt Person B einen dreckigen Putzfetzen und sagt „Ich habe dir was mitgebracht." - **„Danke!"**	Person A kommt lächelnd zur Tür rein und sagt: „Ich habe dir Blumen mitgebracht!" - **„Danke!"**
Person A gives person B a dirty cleaning rag and says: "I brought you something." - "Thanks!"	Person A smiles while entering the room and says: "I brought you some flowers!" - "Thanks!"

2.2. Analyses

For the present study ten utterances were realised by all speakers in both manners (ironic and literal) and in both tasks (scenario-condition and explicit-condition), resulting in a total of 320 utterances.

The recordings were semi-automatically segmented and analysed with STx [18]. The following parameters were measured, manually corrected, and extracted: total utterance duration, duration of the first (stressed) syllable, duration of the stressed vowel / diphthong; mean F0, standard deviation (SD) of F0, minimal F0 value ($F0_{min}$), maximal F0 value ($F0_{max}$), F0 range; mean intensity, SD of the intensity, formant frequencies (F1-F3) of the first vowel in each utterance, normalised F0 contour (= up to 20 equally distributed F0 values for each utterance). In addition, the stressed vowels of all utterances were analysed with VoiceSauce [23, 24] to extract HNR and corrected H1-H2 values. The frequency band for the HNR analyses was 0-3500 Hz.

The data was analysed statistically with R [19] mainly by fitting mixed effects models [3] with utterance and speaker as random factors. When necessary, Tukey post-hoc tests were carried out using the lsmeans package [9]. F0 contours were analysed by fitting a generalised additive mixed model (GAMM) [26]. In the results section, only effects of or interactions with manner of realisation are reported. Especially, obvious effects of speakers' sex or vowel specific differences are not mentioned.

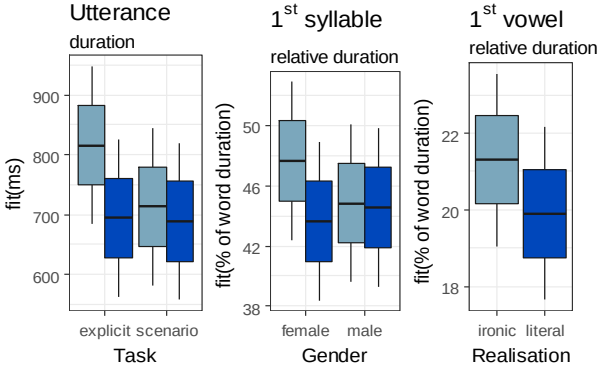
3. RESULTS

3.1. Duration

With respect to the total utterance duration a fitted mixed effects model revealed a significant interaction between manner of realisation and task ($F(1,303)=9.361$, $p=0.002$). Post-hoc tests showed that the duration of ironic and literal utterances differs only in the explicit-condition ($t(306)=5.394$, $p<0.001$) but not in the scenario-condition ($t(306)=1.089$, $p=0.697$) (see Fig. 1).

In order to see which part of the utterance is lengthened in ironic utterances, the relative duration (in % of utterance duration) of the first syllable / word was measured. The statistical model showed a significant gender*realisation interaction ($F(1,303)=10.636$, $p=0.001$), revealing in post-hoc tests that the first syllable of ironic utterances was longer than in literal utterances only for female speakers ($t(305)=4.929$, $p<0.001$) and not for male speakers ($t(305)=0.332$, $p=0.987$) (see Fig.1).

Figure 1: Duration of the utterance; relative duration of the stressed syllable and the stressed vowel (light blue = ironic; dark blue=literal).



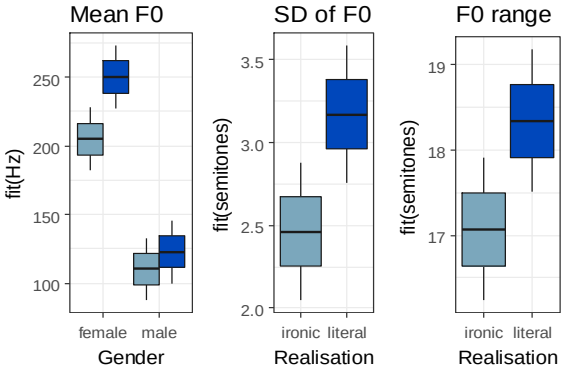
Concerning the relative duration of the stressed vowel (in % of utterance duration), a significant effect of realisation showed longer vowels (t(303)= -3.124, p=0.018) in ironic utterances (see Fig. 1).

3.2. Fundamental frequency

The mean F0 is significantly influenced by an interaction of manner of realisation and gender (F(1,303)=19.306, p<0.001). In post-hoc tests significant differences emerged for female speakers (t(305)=-8.571, p<0.001), and a tendency for a smaller effect was found for male speakers (t(305)=-2.377, p=0.084). For all speakers, the mean F0 was lower in ironic utterances (s. Fig. 2).

Concerning the standard deviation (SD) of F0 (converted to semitones), a significant main effect of manner of realisation (t(303)=4.294, p<0.001) showed a lower SD of F0 for ironic utterances compared to literal utterances (see Fig. 2).

Figure 2: Mean F0, SD of F0 and F0 range (light blue = ironic; dark blue=literal).



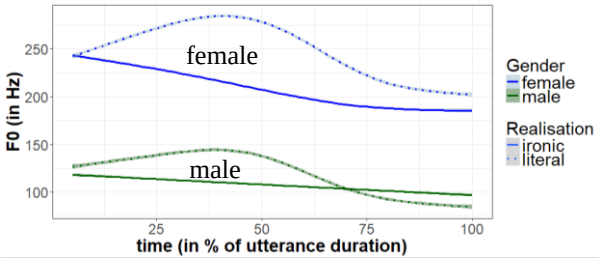
For $F0_{min}$ (in Hz) the model revealed an interaction between manner of realisation and task (F(1,302)=5.650, p=0.018) and a tendency occurred when $F0_{min}$ was converted to semitones (F(1,303)= 3.658, p=0.057). The effect exists for items in the explicit-condition (Hz: t(313)=3.715, p=0.001; ST: t(312)=3.510, p=0.003) but not in the scenario-condition (Hz: t(315)=0.481, p=0.963; ST: t(313)= 0.906, p=0.801). In the explicit-condition, ironic utterances were characterised by higher $F0_{min}$ values compared to literal utterances.

The $F0_{max}$ value (in Hz) showed no effect of manner of realisation (t(312)=1.523, p=0.129). When converted to ST, a significant interaction between manner of realisation and gender emerged (F(1,306)=10.160, p=0.002). Post-hoc analyses revealed lower $F0_{max}$ values in ironic utterances compared to literal utterance only for male speakers (t(307)=-3.659, p=0.002) but not for female speakers (t(315)=0.830, p=0.840).

The difference between $F0_{max}$ and $F0_{min}$ results in the F0 range, which was converted to semitones. The mixed effects model showed a main effect of manner of realisation (t(313)=3.141, p=0.002) with a lower F0 range in ironic utterances (see Fig. 2).

Since not only the total F0 range but also the shape of the F0 contour could differ between the two manners of realisations, the F0 contour of the utterances was analysed by fitting a GAMM [26]. The model revealed a significant interaction between manner of realisation and gender (t=-9.050, p<0.001). Predictions of the different F0 contours are shown in Fig. 3.

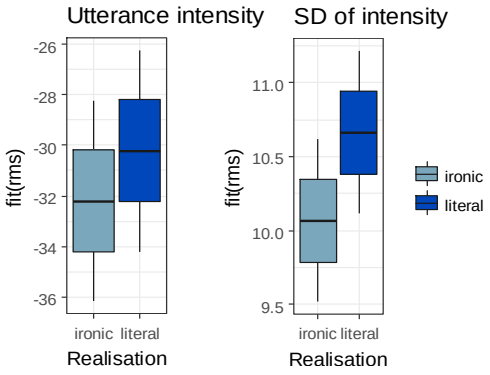
Figure 3: F0 contours (male and female speakers; ironic and literal utterances).



3.3. Intensity

The mean intensity of the utterances is significantly lower in ironic realisations of the utterances (t(303)=5.546, p<0.001). Moreover, the standard deviation of the intensity was measured. The statistical analyses showed a slightly but significantly lower SD of the intensity for ironic utterances (t(303)=-2.993, p=0.003) (see Fig. 4). The video recordings were checked for speaker movements towards or away from the microphone to ensure a comparability of the intensity measurements.

Figure 4: Mean utterance intensity and SD of utterance intensity.



3.4. Formant frequencies

The mean formant frequency (F1-F3) values of the first vowel of each utterance were measured. Concerning the F1 a significant main effect of manner of realisation (t(302)=3.074, p=0.002) revealed lower F1 values in ironic utterances (for all vowels). Equally, for F2 an effect of manner of realisation

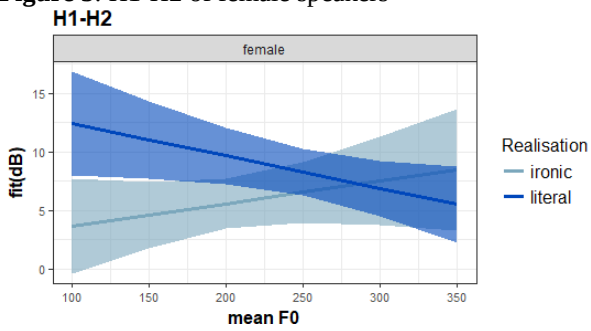
(t(303)=4.426, p<0.001) showed lower F2 values in the stressed vowels of ironic utterances, independent of the vowel quality. With respect to the F3 values, no differences between the two manners of realisation emerged (t(300)=1.216, p=0.225).

3.5. Voice quality

With respect to the voice quality of the stressed vowels, HNR and H1-H2 measurements were analysed. The analyses revealed a realisation*task interaction (F(1,395)=14.34, p<0.001) showing in post-hoc tests that in the explicit-condition, the HNR is higher in ironic utterances compared to literal utterances (t(306)=6.565, p<0.001), but not in the scenario-condition (t(306)=1.454, p=0.466).

Concerning the H1-H2 measurements a significant three-way interaction between manner of realisation, mean F0 and gender occurred (F(1,296)= 6.131, p=0.014). Post-hoc analyses showed a significant effect for female speakers (t(305)=-3.593, p=0.002) but not for male speakers (t(303)=-2.016, p=0.184). The realisation*meanF0 interaction of female speakers is shown in Fig. 5.

Figure 5: H1-H2 of female speakers



4. SUMMARY AND DISCUSSION

The present findings show that speakers of Standard Austrian German mark ironic utterances acoustically by using frequency (F0 and formants), intensity, and durational features. The use of some of the characteristics was limited to the condition in which the speakers were explicitly asked to produce ironic utterances compared to utterances which were only elicited via scenarios. Moreover, some features were used differently by male and female speakers.

With respect to the durational features, all speakers lengthened the ironic realisations in the explicit-condition. Independent of the task, the relative duration of the stressed vowel is longer in ironic utterances. For the relative duration of the first syllable, the gender*realisation interaction can be explained by the fact that female speakers mainly lengthen the stressed syllable to highlight an ironic intent while male speakers lengthen the whole utterance.

Having a closer look on the results of $F0_{min}$, $F0_{max}$, and F0 range, the smaller F0 range in ironic utterances is realised by female speakers by raising the $F0_{min}$ and by male speaker by raising $F0_{min}$ and lowering $F0_{max}$. Both SD of F0 and SD of intensity are lower in ironic utterances. Since both cues are used to generate speech melody, this points towards a more monotonous realisation of ironic utterances, which is supported further by the different shapes of the F0 contours.

Concerning the voice quality measurements, the higher HNR in ironic utterances (=less noise) and the lower values of H1-H2 point to breathier realisation of literal utterances. In addition, the lower H1-H2 values of female speakers in ironic utterances with a lower mean F0 could occur from more creaky realisations. These results will be verified by analysing the additionally recorded EGG signals (as already done for a small subset of the data [11]).

In contrast to [21], the analyses of F1 and F2 did not point to a larger vowel space (=hyperarticulation) since both formants were higher in literal realisations of the utterances independent of the vowel. A higher F2 in literal realisations could be explained by vocal tract shortening due to smiling. This assumption is consistent with finding of the pilot study [10] and will be verified further in a following study by analysing the facial expressions of the speakers, for which video recordings have been done.

To sum up, in SAG, ironic utterances as compared to literal utterances are characterised by a lower mean F0, a lower SD of F0, a smaller F0 range, lower mean intensity and lower intensity variation, longer durations (utterance, stressed vowel, stressed syllable), lower F1 and F2 frequencies of the first stressed vowel, differences in voice quality and a flatter F0 contour. Most of the results are consistent with results on German in Germany and English. However, in several aspects, the present results differ from studies on verbal irony in German in Germany [16, 17, 21, 22]: The present study did not show hyperarticulation of vowels but in general lower F1 and F2 values in ironic utterances; H1-H2 values were lower in ironic utterances and unlike [21] F0 contours differed between the two manners of realisation. There seems to be no major differences due to language type specific characteristics of SAG. Moreover, the present study extends the research on verbal irony on results of an additional language variety and shows the direction in which the investigation of irony in SAG will continue: The EGG and video recordings of all speakers will be analysed. Furthermore, a perception experiment with normal-hearing and cochlear-implant listeners is planned to see which acoustic and visual cues are used by the two groups to identify irony.

5. REFERENCES

[1] Adachi, T. 1996. Sarcasm in Japanese. *Studies in Language*. 20(1), 1–36.

[2] Anolli, L., Ciceri, R., Infantino, M. G. 2000. Irony as a Game of Implicitness: Acoustic Profiles of Ironic Communication. *Journal of Psycholinguistic Research*. 29(3), 275–311.

[3] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 67(1), 1–48.

[4] Bryant, G. A. 2010. Prosodic Contrasts in Ironic Speech. *Discourse Processes*. 47(7), 545–566.

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

[6] 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.

[7] Chen, A. and Boves, L. 2018. What's in a word: Sounding sarcastic in British English. *Journal of the International Phonetic Association*. 48(01), 57–76.

[8] Laval, V. and Bert-Erboul, A. 2005. French-Speaking Children's Understanding of Sarcasm: The Role of Intonation and Context. *Journal of Speech, Language, and Hearing Research*. 48, 610–620.

[9] Lenth, R. V. 2016. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*. 69(1).

[10] Leykum, H., in press. A pilot study on the diversity in irony production and irony perception. In: Colston, H. L., Athanasiadou, A. (eds), *The Diversity of Irony*

[11] Leykum, H. in press. Voice quality of ironic utterances in Standard Austrian German: preliminary results. *Proceedings of P&P14*

[12] Lœvenbruck, H., Ameur Ben Jannet, M., D'Imperio, M., Spini, M., Champagne-Lavau, M. 2013. Prosodic cues of sarcastic speech in French: Slower, higher, wider. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* Lyon, 3537–3541.

[13] Moosmüller, S. 1991. *Hochsprache und Dialekt in Österreich: Soziophonologische Untersuchungen zu ihrer Abgrenzung in Wien, Graz, Salzburg und Innsbruck*. Wien: Böhlau.

[14] Moosmüller, S. and Brandstätter, J. 2014. Phonotactic information in the temporal organization of Standard Austrian German and the Viennese dialect. *Language Sciences*. 46, 84–95.

[15] Moosmüller, S., Schmid, C., Brandstätter, J. 2015. Standard Austrian German. *Journal of the International Phonetic Association*. 45(03), 339–348.

[16] Nauke, A. and Braun, A. 2011. The production and perception of irony in short context-free utterances. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)* Hong Kong, 1450–1453.

[17] Niebuhr, O. 2014. "A little more ironic" – Voice quality and segmental reduction differences between sarcastic and neutral utterances. *7th International Conference of Speech Prosody* Dublin, 608–612.

[18] Noll, A., White, J., Balazs, P., Deutsch, W. 2007. *STx - Intelligent Sound Processing, Programmer's Reference.*

[19] R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. http://www.R-project.org/.

[20] Rockwell, P. 2000. Lower, Slower, Louder: Vocal Cues of Sarcasm. *Journal of Psycholinguistic Research*. 29(5), 483–495.

[21] Scharrer, L. and Christmann, U. 2011. Voice Modulations in German Ironic Speech. *Language and Speech*. 54(4), 435–465.

[22] Schmiedel, A. 2017. *Phonetik ironischer Sprechweise: Produktion und Perzeption sarkastisch ironischer und freundlich ironischer Äußerungen*. Berlin: Frank & Timme.

[23] Shue, Y.-L. 2010. *The voice source in speech production: Data, analysis and models*, UCLA dissertation.

[24] Shue, Y.-L., Keating, P., Vicenik, C., Yu, K. 2011. Voicesauce: a program for voice analysis. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)* Hong Kong, 1846–1849.

[25] Szczepaniak, R. 2007. *Der phonologisch-typologische Wandel des Deutschen von einer Silben- zu einer Wortsprache*. Berlin: Walter de Gruyter.

[26] Wood, S. and Scheipl, F. 2017. *gamm4: Generalized Additive Mixed Models using 'mgcv' and 'lme4'.* R package version 0.2-5. http://CRAN.R-project.org/package=gamm4.