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

Examining melodiousness in sarcasm: Wiggliness, spaciousness, and contour clustering

Csilla Tatár[1], *Jonathan R. Brennan*[1], *Jelena Krivokapić*[1], *Ezra Keshet*[1]

[1]University of Michigan, US
cstatar@umich.edu

Abstract

This paper compares ways of quantifying the phonetic correlates of sarcasm focusing on two novel measures, *wiggliness* and *spaciousness*, neither of which has been examined in affective prosody before. We compare these to further F0 measures and evaluate their role in distinguishing sarcasm via contour clustering.

In a production study, American English speakers (N=12) were recorded producing identically worded utterance pairs presented in contexts conducive to sarcasm and sincerity. Utterances were analyzed for wiggliness, spaciousness, F0 range, and the SD of F0 mean. The measures were entered into logistic regression models as predictors for sarcastic affect (wiggliness and spaciousness jointly, the others separately); model fit was evaluated with pseudo-R statistics. Results show that wiggliness and spaciousness together are comparable to F0 mean SD and F0 range in that reduced wiggliness and spaciousness distinguish sarcasm from sincerity for many of the speakers (N=8). By-speaker contour clustering was performed to examine the categorization of affect based solely on acoustic properties. The quality of speakers' clusters in capturing sarcasm vs. sincerity varied, indicating a variable match between speaker intent and acoustic properties. Further analyses showed that wiggliness and spaciousness capture the differences between sincere and sarcastic contours for some speakers.

Index Terms: wiggliness, spaciousness, F0 contour clustering, sarcasm, affect

1. Introduction

Sarcasm may be defined as a critical message expressed in positive language. Speakers mark sarcasm in several ways, among them manual gestures, facial expressions, lexical choices, and prosody (see [1-8]). The present paper focuses on the prosodic marking of sarcasm, in particular it examines the phonetic correlates of sarcastic speech, specifically F0.

Previous work found that sarcasm can be marked by differences in F0 variability, quantified as F0 range ([9-22]) and standard deviation around mean F0 ([9], [19], [21], [22], [23]). The present study explores two novel measures of F0 variability in marking sarcasm, *wiggliness* (the number of turns in the F0 contour per second) and *spaciousness* (the average of the two greatest F0 movements within a contour), both introduced in [24]. Wiggliness allows for the direct measurement of how much the contour changes (instead of relying on averaged values), while spaciousness measures range in a more fine-grained manner than overall F0 range does.

With respect to differences in intonational contours in sincere and sarcastic speech, [4] and [15] found specific contours to mark sarcasm in Catalan and French, respectively, and [14] found a flatter pitch contour in British English. As studies report smaller F0 range or SD in sarcasm for various dialects of English ([10], [14], [19], [21]), a less variable F0 contour may be characteristic of sarcasm in American English. As an initial exploration of sarcastic contours, the present study examines the clustering patterns of sincere and sarcastic utterances in terms of their F0 contours independently of speaker intent with the clustering method developed in [25], with the goal of examining if speakers produce contours characteristic of sarcasm, and if contours differ on the wiggliness and spaciousness dimensions.

2. Methods

2.1 Participants

Students (17f, 1m) at the University of Michigan participated in the study. Participants were in their early 20s. The gender and age distribution is due to the availability of participants. Participants' first language was English, and they reported no speech impediments or hearing loss. Participants were compensated for their time and effort ($20.00/h, prorated).

2.2 Materials

Production materials consisted of short context - target utterance pairs. Contextual information was given to aid speakers in producing the utterances with the appropriate attitude. For each target utterance, three context-setting situations were written: one encouraging a sarcastic reading of the target utterance, one encouraging sincerity, and one disbelief (the latter one was included to avoid binary readings). Structurally, the target utterances were similar to those in [13]: "[*X*] *is really quite a* [*positive adjective*] [*noun*]". To reliably track F0, an effort was made to include adjectives and nouns with predominantly voiced segments. There were 32 utterances per affect condition, with two repetitions (1536 sentences total). The first repetition of sarcastic and sincere utterances was analyzed, unless there was disfluency, in which case the second repetition was used, for a total of 768 utterances (12 speakers x 2 affect conditions x 32 sentences).

2.3 Procedure

Speakers were recorded individually with the first author present in a sound attenuated booth, where two laptops were set up: one for recording and one for stimuli presentation. Recording took place with an AKG C4000 B microphone connected to a Focusrite Scarlett Solo. To gain familiarity with the task and the equipment, each participant completed a training session. Following this, they were presented with the

10.21437/SpeechProsody.2024-137

speech stimulus (contexts) and the target sentences. Recording was self-paced, and participants advanced by right-clicking on the touchpad of the laptop that was set up for stimuli presentation. Stimuli were shown in random blocks of ten, between which participants could take a break if they so wished. They were encouraged to take breaks as needed. Recording took 45 minutes per speaker on average.

Participants were instructed to read the written contextual information (white text on black background) silently. Below each context was the target sentence (light orange color on black background) that they were asked to produce aloud. Speakers were instructed to produce the target sentence in a way that felt appropriate to them given the contextual information provided; regarding prosody, participants did not receive instruction. The contexts and the corresponding target sentences were presented in pseudo-random order two times. Participants had the option of repeating an utterance if they were discontented with their first attempt. Additionally, they were asked to repeat an utterance if there was noticeable disfluency in their production.

2.4 Phonetic analysis

Phonetic analysis was completed on twelve speakers' data, after participants were screened for potential issues (giggling (N=1), experimental error (N=1)). Participants were further excluded if there was considerable creakiness in their production (N=2), as the phonetic feature association models all rely on F0-related measures, and creaky voice prevents F0 from being reliably tracked. Tokens with such potential issues were removed from the remaining participants' sets of tokens as well. Given the imbalance in the gender of the participants who responded to the recruitment flier, the single male participant was excluded as well. Participants whose production was included in the analysis are henceforth coded as P1-P12. In total, 768 utterances were analyzed in Praat ([26]) and R (v4.2.1 [27]).

F0-variability was quantified in three ways for each speaker individually: (1) in terms of F0 range, (2) in terms of the standard deviation around mean F0, and (3) in terms of wiggliness and spaciousness (novel measures introduced in [24]). Acoustic features were measured in Praat. F0 range was calculated from F0 minimum and maximum and is reported in Hz. F0 mean was extracted from Praat, from which the standard deviation was calculated for each speaker; it is reported in Hz as well. Wiggliness and spaciousness are defined in [24] as follows: wiggliness is the number of turns in the F0 contour per second, and spaciousness is the average of the two greatest movements in F0. These were measured using the method laid out in [28] with the Praat and R scripts provided. First inter-pausal units (IPUs) were identified and labeled on TextGrids in Praat; note that for the purposes of the project, each individual target utterance was determined to be an IPU with their duration labeled in milliseconds. Then, the *mausmooth* software [29] was used in Praat to inspect and correct the F0 contours for any tracking errors and creakiness. Next, with *smoosplit* (developed by Wehrle and Cangemi (see [28]) the pitch points were interpolated, and the contour was smoothed and stylized to a two-semitone (ST) resolution following the suggestion of [28]. The F0 measures and their time points of the original contour and the stylized contour were extracted into text files with *smoosplit*. Finally, wiggliness and spaciousness values were calculated in R (v4.2.1 [27]) with the *Wiggliness and Spaciousness Calculation* script developed by Wehrle and Gauser (see [28]).

2.5 Contour cluster analysis

By-speaker contour clustering was performed to examine if sarcasm and sincerity cluster separately based on their F0, without using speaker intent in the clustering algorithm (thus independently of speaker intent). Clustering analysis was done on the same data as in 2.4, but the tokens belonging to the two affect conditions were conflated and examined together. The cluster analysis methods and scripts provided in [25] were used to conduct by-speaker analysis on the F0 contours. As each speaker was examined individually for their sarcastic and sincere productions, no speaker normalization was applied. To obtain F0 measurements, the sound files and TextGrids were run through the Praat script provided in [25] with the default settings. In the R tool, Euclidean (L2 norm) distance was set as the distance measure, and linkage criterion was set to *complete* (default settings, and as suggested in [25]). Clustering was performed on data converted to semitones to facilitate the comparison with wiggliness and spaciousness. Rows marked for error by the script and NAs were removed (this process removed a larger portion of the utterances of speakers *P1, P2,* and *P11*). Two clusters were created. If a sole token constituted a cluster, the token was removed from the analysis as a likely outlier, and cluster analysis was performed again (n=1). The distribution of sarcastic and sincere tokens per cluster for each speaker is given in Table 1.

2.5 Statistical analysis

2.5.1 Prosodic feature associations

Prosodic feature associations were analyzed via by-speaker logistic regression analyses carried out in R (v4.2.1 [27]); The following models were tested:

$$Sarcasm = \beta_0 + \beta_1(\text{Wig}) + \beta_2(\text{Spac}) \quad (1)$$

$$Sarcasm = \beta_0 + \beta_1(\text{F0 range}) \quad (2)$$

$$Sarcasm = \beta_0 + \beta_1(\text{F0 mean SD}) \quad (3)$$

In all models, the dependent variable was sarcastic intent (i.e. the intended affect of the speaker). In (1), Wiggliness and Spaciousness measures were entered as predictor variables, in (2), F0 range was the predictor, and in (3), the standard deviation around the F0 mean was tested as predictor. Model fit by speaker was estimated with pseudo-R2 statistics (McFadden's R2 [31]). R packages used were *tidyverse* [32], *hqmisc* [33], *pscl* [34], *ggplot2* [35], *dplyr* [36].

2.5.2 Clustering

As an exploratory analysis to evaluate if wiggliness or spaciousness distinguish between the clusters, by-speaker logistic regression analyses were performed on the data of speakers who had a minimum of ten tokens in each cluster, where cluster group was the dependent variable (cluster 1 = 0, cluster 2 = 1) and measured wiggliness and spaciousness values were predictors in separate regression models.

$$Cluster\ 2 = \beta_0 + \beta_1(\text{Wiggliness}) \quad (1)$$

$$Cluster\ 2 = \beta_0 + \beta_1(\text{Spaciousness}) \quad (2)$$

3. Results

3.1 F0 measures

Raw values of spaciousness (ST) and wiggliness (ST) are given in Figure 1, and raw values of F0 range (Hz) and the standard deviation around F0 mean (Hz) are given in Figure 2. The figures illustrate the by-speaker differences between the sarcastic and sincere renderings of the utterances.

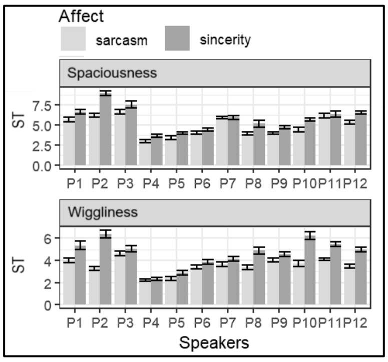

Figure 1: *By-speaker spaciousness (ST) and wiggliness (ST) in sincerity and sarcasm.*

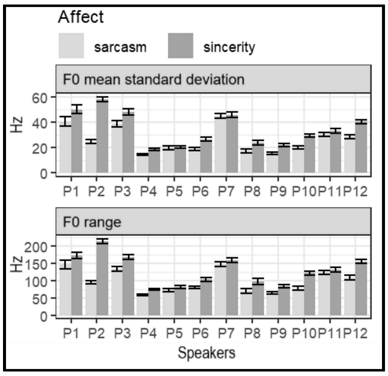

Figure 2: *By-speaker F0-mean SD (Hz) and F0 range (Hz) in sincerity and sarcasm.*

3.2 Prosodic feature associations

This subsection reports the results of the by-speaker logistic regression analyses. All significant results are at $p<0.05$, and model fit is estimated with McFadden's Pseudo-R^2.

In the model of wiggliness and spaciousness predicting sarcasm, wiggliness emerged as a significant predictor for six speakers (*P1* [β=-0.55], *P2* [β=-1.3], *P8* [β= -0.87], *P10* [β=-1.2], *P11* [β=-2.01], *P12* [β=-1.03]) and spaciousness for three speakers (P2 [β=-1.0], P4 [β=-0.64], P9 [β=-0.94]). Where significance was reached, reduced wiggliness and reduced spaciousness were characteristic of sarcasm. For these speakers, apart from *P4*, model fit reaches the threshold of $R^2 \geq 0.2$. For a further three speakers (*P2, P10, P11*), model fit was exceptionally good at $R^2 \geq 0.5$.

In the model of F0 range predicting sarcasm, F0 range was a significant predictor for eight speakers (*P2* [β=-0.08], *P3* [β=-0.02], *P4* [β=-0.07], *P6* [β=-0.04], *P8* [β=-0.02], *P9* [β= -0.06], *P10* [β=-0.06], *P12* [β=-0.04]), and in all cases, reduced F0 range predicted sarcasm. The model reached the $R^2 \geq 0.2$ threshold of good fit in all cases where the predictor was significant. For a further two speakers (*P2, P10*), model fit was exceptionally good at $R^2 \geq 0.5$.

In the model of the standard deviation around the F0 mean predicting sarcasm, the predictor was significant for eight speakers (*P2* [β=-0.31], *P3* [β=-0.06], *P4*[β=-0.23], *P6* [β= -0.18], *P8* [β=-0.08], *P9* [β=-0.26], *P10* [β=-0.19], *P12* [β= -0.14]), with smaller SD predicting sarcasm. For all of them, the model reached the $R^2 \geq 0.2$ threshold of good fit. For one further speaker (*P2*), model fit was exceptionally good at $R^2 \geq 0.5$.

3.3. Contour clustering

The results of the by-speaker clustering analysis with two assumed clusters are given in Table 1. The quality of the clusters is first analyzed by calculating what percentage of *sincere* tokens group together in a speaker's cluster 1 versus their cluster 2, and what percentage of *sarcastic* tokens group into their cluster one versus cluster 2. Then, the quality of clusters is examined in terms of cluster composition, calculating what percentage of *all* tokens in a given cluster of a speaker are sincere vs. sarcastic. 70% was chosen as an arbitrary threshold in the evaluation of the clusters.

3.3.1 By-speaker sincere and sarcastic token distribution across clusters

In terms of the first calculation, there appears to be a strong separation of sincere and sarcastic contours for speakers *P4* and *P5*, and a somewhat weaker separation for speakers *P10, P3,* and *P11*. Speaker *P4* appears to keep sincere and sarcastic contours separate with 87% of sincere tokens grouping together in cluster 2 and 70% of sarcastic tokens in cluster 1. Similarly, speaker *P5*'s sincere and sarcastic contours appear to separate with 81% of sincere tokens in cluster 2 and 90% of sarcastic ones in cluster 1. Speaker *P10* is again similar with all their sarcastic tokens in grouping into cluster 2 and 65% of all sincere tokens into cluster 1. Likewise, for speaker *P3*, 91% of their sarcastic tokens are in cluster 1 and there is a tendency for sincere tokens to group into cluster 2 (at 64%). Speaker *P11* is alike as well with 96% of the sincere tokens in cluster 1 and 63% of sarcastic tokens in cluster 2.

There appears to be some separation, although much weaker, between sincerity and sarcasm for speakers *P2, P7*, and *P9*. They all show a near equal distribution of their sincere contours across their clusters, but sarcastic contours show a strong tendency to group into one cluster (100%, 97%, and 96%, respectively). For these speakers, it appears that sarcasm is less variable in terms of F0 contours.

Conversely, for speaker *P1*, it is the sarcastic tokens that show a near equal distribution into two clusters, while sincere contours group predominantly (85%) into one. In this case, sarcasm appears to be more variable.

Finally, in the case of speakers *P6*, *P8*, and *P12*, the majority of their sincere tokens cluster together (70%, 93%, 95%, respectively) and most of their sarcastic tokens group together as well (96%, 90%, 87%, respectively). The two affect groups do not show a separation, however, with most sincere and sarcastic tokens clustering together.

In sum, it appears that speakers distinguish between sarcastic and sincere speech to varying degrees in their intonational contours, with only five speakers (P3, P4, P5, P10, P11) making clear distinctions.

3.3.2 By-speaker cluster composition

In terms of the second calculation, i.e. by-speaker cluster composition, *P2*, *P3*, *P4*, *P5*, *P10*, and *P11* are similar in that one of their clusters is composed predominantly of sincere tokens (100%, 82%, 75%, 89%, 100%, 77%, respectively), while the other contour cluster consists mostly of sarcastic tokens (81%, 79%, 84%, 81%, 73%, 92%, respectively).

Speakers *P1*, *P6*, *P7*, *P9* and *P12* are comparable in that only one of their clusters show an affect-based distinction in their distribution. For speakers *P1* and *P12*, one cluster consists mostly of sarcastic tokens (86% and 80%, respectively), while the other cluster is composed of both sincere and sarcastic tokens to a near-equal degree. Conversely, for speakers *P6*, *P7*, and *P9*, the cluster that shows an affect-based distinction consists predominantly of sincere tokens (90%, 93%, and 92%, respectively), while their other cluster is composed of both sincere and sarcastic tokens.

In the case of *P8*, neither cluster shows a preponderance of tokens from one versus the other affect condition.

In sum, half the speakers show a relatively strong separation of sincerity and sarcasm in the composition of their two clusters, while the other half shows more variability.

Table 1: *Clustering with two clusters assumed. (Note: the clusters are to be interpreted within-speaker only, i.e. cluster 1 for P1 is not the same as cluster 1 for P2, etc.)*

Speaker	***Affect***	***Tokens in cluster 1***	***Tokens in cluster 2***
P1	*sincere*	11	2
	sarcastic	11	12
P2	*sincere*	6	5
	sarcastic	26	0
P3	*sincere*	8	18
	sarcastic	30	3
P4	*sincere*	4	27
	sarcastic	21	9
P5	*sincere*	6	25
	sarcastic	26	3
P6	*sincere*	21	9
	sarcastic	27	1
P7	*sincere*	16	14
	sarcastic	29	1
P8	*sincere*	28	2
	sarcastic	26	3
P9	*sincere*	15	11
	sarcastic	22	1
P10	*sincere*	17	9
	sarcastic	0	24
P11	*sincere*	23	1
	sarcastic	7	12
P12	*sincere*	21	1
	sarcastic	27	4

3.4 Contour clusters, wiggliness, and spaciousness

With respect to how the measured features of wiggliness and spaciousness relate to the contour clusters, logistic regression analyses were performed on the data of speakers *P1*, *P3*, *P4*, *P5*, *P6*, *P7*, *P9*, *P10*, and *P11*. Results are reported at p<0.05 significance level.

For *P1*, *P3*, *P4*, *P5* and *P11*, results show that neither of the measures are significant predictors of their cluster 2.

For speakers *P6* and *P9*, spaciousness is positively correlated with cluster 2 (*P6*: β=1.62; *P9*: β=0.76), indicating that the contours constituting cluster 2 are relatively more spacious than their cluster 1 contours. For both speakers, subsection 3.3.2 reported cluster 2 to consist predominantly of sincere tokens; it appears therefore, that a more spacious contour is likely to be sincere rather than sarcastic.

For speakers *P6* and *P7*, wiggliness is positively correlated with cluster 2 (*P6*: β=0.7, *P7*: β=1.1). For these speakers, cluster 2 consisted of sincere tokens predominantly, suggesting that sarcastic utterances are less likely to be produced with increased wiggliness. For speaker *P10*, wiggliness is negatively correlated with cluster 2 (β=-0.44), indicating that the contours constituting cluster 2 have reduced wiggliness relative to cluster 1. As reported in 3.3.2, sarcastic utterances predominantly contributed to this speaker's cluster 2, suggesting that sarcastic utterances are produced with a less wiggly contour.

In sum, wiggliness and spaciousness may contribute to the by-speaker grouping of some contours, and the differences may relate to sarcastic affect. Figure 3 shows the contour clusters for *P6*, *P7*, *P9* and *P10*, for whom wiggliness or spaciousness was a significant predictor of cluster group.

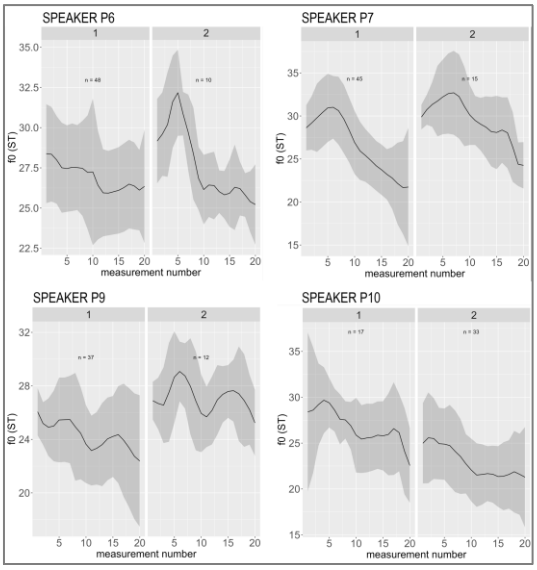

Figure 3: *Contour clusters for speakers P6, P7, P9 and P10.*

4. Discussion

Similarly to prior studies, the present study found many speakers to mark sarcasm with reduced F0 variability relative to sincerity. Reduced wiggliness and spaciousness *together* were found to be significant predictors for over half of the speakers (spaciousness to a lesser degree), producing comparable results to the averaged measures of F0 range and mean SD, and demonstrating the ability of the novel measures to capture differences in affective prosody. Contour clustering demonstrated the tendency of roughly half the speakers to produce unique contours more characteristic either of sincerity or of sarcasm. Further analyses of the clusters demonstrated that relative to sincerity, some speakers mark sarcasm by less wiggly and/or less spacious contours. Further distinguishing factors may include differences in accentuation and accent scaling (see P6). Notable, however, is the mismatch between many speakers' intention of producing sincere versus sarcastic utterances and how their F0 contours cluster. This, in part, could contribute to the difficulty of many in perceiving sarcasm (see [37] for a review). It remains to be examined whether the results reported here generalize to multiple genders.

5. References

[1] Aguert, M. (2022). Paraverbal expression of verbal irony: vocal cues matter and facial cues even more. Journal of Nonverbal Behavior, 46(1), 45-70.mmm

[2] Kreuz, R. (2020). *Irony and sarcasm*. MIT Press.

[3] Garmendia, J. (2018). *Irony*. Cambridge University Press.

[4] González-Fuente, S., Escandell-Vidal, V., & Prieto, P. (2015). Gestural codas pave the way to the understanding of verbal irony. Journal of Pragmatics, 90, 26–47. https://doi.org/10.1016/j.pragma.2015.10.002

[5] Caucci, G. M., & Kreuz, R. J. (2012). Social and paralinguistic cues to sarcasm. Humor, 25(1), 1-22.

[6] Pexman, P. M. (2008). It's Fascinating Research: The Cognition of Verbal Irony. Current Directions in Psychological Science, 17(4), 286–290. https://doi.org/10.1111/j.1467-8721.2008.00591.x

[7] Rockwell, P. (2000). Lower, Slower, Louder: Vocal Cues of Sarcasm. Journal of Psycholinguistic Research, 29(5), 13.

[8] Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. Metaphor and symbol, 10(1), 21-31.

[9] Leykum, H. (2020). Voice quality of ironic utterances in Standard Austrian German: preliminary results. In Akten der Konferenz" Phonetik und Phonologie im deutschsprachigen Raum (P&P14)"/Proceedings of the Conference, Phonetics and Phonology in the German Language Area (P&P 14) (pp. 94-98).

[10] Mauchand, M., Vergis, N., & Pell, M. D. (2020). Irony, prosody, and social impressions of affective stance. Discourse Processes, 57(2), 141-157.

[11] Jansen, N., & Chen, A. (2020). Prosodic encoding of sarcasm at the sentence level in Dutch. 10th International Conference on Speech Prosody 2020, 409–413. https://doi.org/10.21437/SpeechProsody.2020-84

[12] Braun, A., & Schmiedel, A. (2018). The phonetics of ambiguity: A study on verbal irony. Cultures and Traditions of Wordplay and wordplay research, 6, 111.

[13] Mauchand, M., Vergis, N., & Pell, M. D. (2018, June). Ironic tones of voices. In 9th International Conference on Speech Prosody (pp. 443-447).

[14] Chen, A., & Boves, L. (2018). What's in a word: Sounding sarcastic in British English. Journal of the International Phonetic Association, 48(1), 57–76. https://doi.org/10.1017/S0025100318000038

[15] González-Fuente, S., Prieto, P., & Noveck, I. (2016). A fine-grained analysis of the acoustic cues involved in verbal irony recognition in French. 902–906. https://doi.org/10.21437/SpeechProsody.2016-185

[16] Niebuhr, O. (2014). "A little more ironic" Voice quality and segmental reduction differences between sarcastic and neutral utterances. 7th International Conference on Speech Prosody 2014, 608–612. https://doi.org/10.21437/SpeechProsody.2014-109

[17] Rakov, R., & Rosenberg, A. (2013, August). " sure, I did the right thing": a system for sarcasm detection in speech. In Interspeech (pp. 842-846).

[18] Lœvenbruck, H., Jannet, M. B., d'Imperio, M., Spini, M., & Champagne-Lavau, M. (2013, August). Prosodic cues of sarcastic speech in French: slower, higher, wider. In Interspeech 2013-14th Annual Conference of the International Speech Communication Association (pp. 3537-3541).

[19] Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. Speech Communication, 50(5), 366–381. https://doi.org/10.1016/j.specom.2007.11.003

[20] Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. The Journal of the Acoustical Society of America, 126(3), 1394–1405. https://doi.org/10.1121/1.3177275

[21] Bryant, G. A., & Fox Tree, J. E. (2005). Is there an Ironic Tone of Voice? Language and Speech, 48(3), 257–277. https://doi.org/10.1177/00238309050480030101

[22] 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. https://doi.org/10.1080/00207590244000106

[23] 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.

[24] Wehrle, S., Cangemi, F., Krüger, M., & Grice, M. (2018). Somewhere over the spectrum: Between robotic and singsongy intonation. Il Parlato Nel Contesto Naturale. Speech in the Natural Context, 4, 179-194.

[25] Kaland, Constantijn. (2021). Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association*. doi:10.1017/S0025100321000049 [pdf]

[26] Boersma, Paul & Weenink, David (2023). Praat: doing phonetics by computer [Computer program]. Version 6.1.54, retrieved 2 September 2020 from http://www.praat.org/

[27] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[28] Wehrle, Simon (2022). A brief tutorial for using Wiggliness and Spaciousness to measure intonation styles. OSF; retrievable at osf.io/5e7fd.

[29] Cangemi, Francesco (2015). *mausmooth*. Retrievable online at http://phonetik.phil-fak.uni-koeln.de/fcangemi.html

[30] Kaland, C; Peck, N; Ellison, T. M. & Reinöhl, U. (2021). An initial exploration of the interaction of tone and intonation in Kera'a. *Proceedings of the 1st International Conference on Tone and Intonation (TAI)* (pp. 132-136). Sønderborg, Denmark. doi:10.21437/TAI.2021-27 [pdf]

[31] McFadden, D. (1987). Regression-based specification tests for the multinomial logit model. Journal of econometrics, 34(1-2), 63-82.

[32] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, LD., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, TL., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, DP., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H. (2019). "Welcome to the tidyverse." _Journal of Open Source Software_, *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

[33] Quene H (2022). _hqmisc: Miscellaneous Convenience Functions and Dataset_. R package version 0.2-1, https://CRAN.R-project.org/package=hqmisc.

[34] Jackman, S. (2020). pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory. United States Studies Centre, University of Sydney. Sydney, New South Wales, Australia. R package version 1.5.5. URL: https://github.com/atahk/pscl/

[35] Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

[36] Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.2, https://CRAN.R-project.org/package=dplyr.

[37] Pexman, P., Reggin, L., & Lee, K. (2019). Addressing the challenge of verbal irony: Getting serious about sarcasm training. *Languages*, *4*(2), 23.