F0 AS A CUE FOR IRONY IN SPONTANEOUS SPEECH

Helen Gent

University of Illinois at Urbana Champaign
hmgent2@illinois.edu

ABSTRACT

This study suggests that the sought-after universal prosodic pattern for irony is a fallacy, and instead irony prosody is individually determined - explaining past studies' conflicting results. In this naturalistic speech study, time-normalized F0 values were fitted to a generalized additive model (GAM) for each speaker (n = 5), and the full group. The results showed a significant difference in f0 over time between ironic and non-ironic speech – adding support to the prevailing opinion that prosody plays an important role in cuing verbal irony – as well as between speakers, providing a potential alternate path for research into irony prosody.

Rather than attempting to isolate a universal prosodic pattern for irony, future research can explore individual modifications to prosody in different domains and sociolinguistic contexts. The resulting expansion of understanding has potential applications in voice processing and artificial intelligence, first and second language acquisition, and social disorder research.

Keywords: prosody, irony, phonology, f0, naturalistic

1. INTRODUCTION

Past research into irony prosody has largely taken one of two approaches: either assuming a prosodic pattern for ironic speech and studying its effects on irony recognition and processing, or examining prosodic data itself. While the former path has found prosody to have an important role alongside discourse context [4], research into actual prosodic content has been sparse and has failed to produce a clear trend – particularly in the domain of f0. A wide variety of prosodic cues for irony have been identified, but the directionality of their differences from non-ironic speech has not been reliable. For example, some studies have found both higher and lower mean f0 in ironic speech than in non-ironic speech [1][3].

While some of the between-studies variation can be attributed to differing experiment design and data processing methods, the within-studies variation suggests that the issue may lie deeper – in the assumptions behind the research. This study suggests that the sought-after universal prosodic pattern for irony – the "ironic tone of voice"– is a fallacy. Though this is not the first study to suggest this [5], it presents a novel path forward, suggesting the reason for the conflicting results of past research is that irony prosody is individually determined, and therefore each individual will make different prosodic modifications to communicate irony.

2. RESEARCH QUESTIONS AND HYPOTHESES

This study is guided by the following research questions:

RQ1: Are there statistically significant differences between ironic and non-ironic prosody?

Hypothesis: There will be statistically significant differences between ironic and non-ironic prosody.

Prediction: These differences will not take a consistent form between speakers, but will nonetheless be distinguishable.

RQ2: If these differences exist, do they function on the level of individual speaker, speech community, or both?

Hypothesis: Different individual speakers will realize irony prosody differently, though speech community may have an effect.

3. METHODOLOGY

This study breaks from past research in three ways. First, it examines naturalistic speech, whereas much of the existing literature has used scripted irony. Secondly, it analyses and compares individual speakers' ironic prosody, to examine if irony prosody truly functions as an individual trait. Finally, rather than using a universal measure like mean, standard deviation, or range to study f0, this study utilizes a generalized additive model (GAM) in order to capture f0 behaviour relative to time.

3.1. The Corpus

The corpus was constructed using samples from the Stoat Party Podcast – a Dungeons and Dragons campaign. As the podcast is recorded casually and has a small listenership, the recorded speech is naturalistic and largely unaffected by the recording setting. The one-hour recording of a single episode yielded 781 utterances suitable for analysis between the five speakers.

3.1.1. Irony Labelling

To avoid subjectivity, irony in the corpus was labelled using the example of Brown [2] – meaning that an utterance was identified as ironic based on the response that followed it. Conversational responses to irony were based on the results of Gibbs [6]. Such that utterances were labelled as ironic if followed by any of the following: laughter, a literal response to the ironic meaning of the utterance, an ironic response (resulting in an irony chain that would need to be identified backward), or an explanation by the original speaker that irony was intended.

3.2. Selecting Samples for Analysis

After irony was labelled, samples were sorted by sentence type, and only statements were retained for analysis. Then, of the ironic statements, the median 80% by length was retained per speaker, and an equal number of non-ironic statements were randomly selected within this same length range for comparison. This yielded 15-23 ironic and non-ironic utterances per speaker.

3.3. Data Processing and Analysis

F0 information was extracted from the sound files using the Robust Epoch and Pitch EstimatoR (REAPER)[8]. F0 values were taken at the default interval of 5 ms and saved to plaintext files. Ironic and non-ironic f0 sample sets for each speaker were then fitted to a generalized additive model (GAM) in R [7], using the mgcv package [9]. GAMs were chosen in order to examine the complete f0 line over time, rather than relying on single measures like mean, range, or standard deviation.

The files were time-normalized to 75 time points per file, and ironic and non-ironic f0 lines were compared for each speaker. Raw f0 values were normalized within the GAM. An additional GAM was fitted with ironic and non-ironic data from all speakers combined, with speaker as a random variable. For this model, only 15 samples per speaker were included.

4. RESULTS

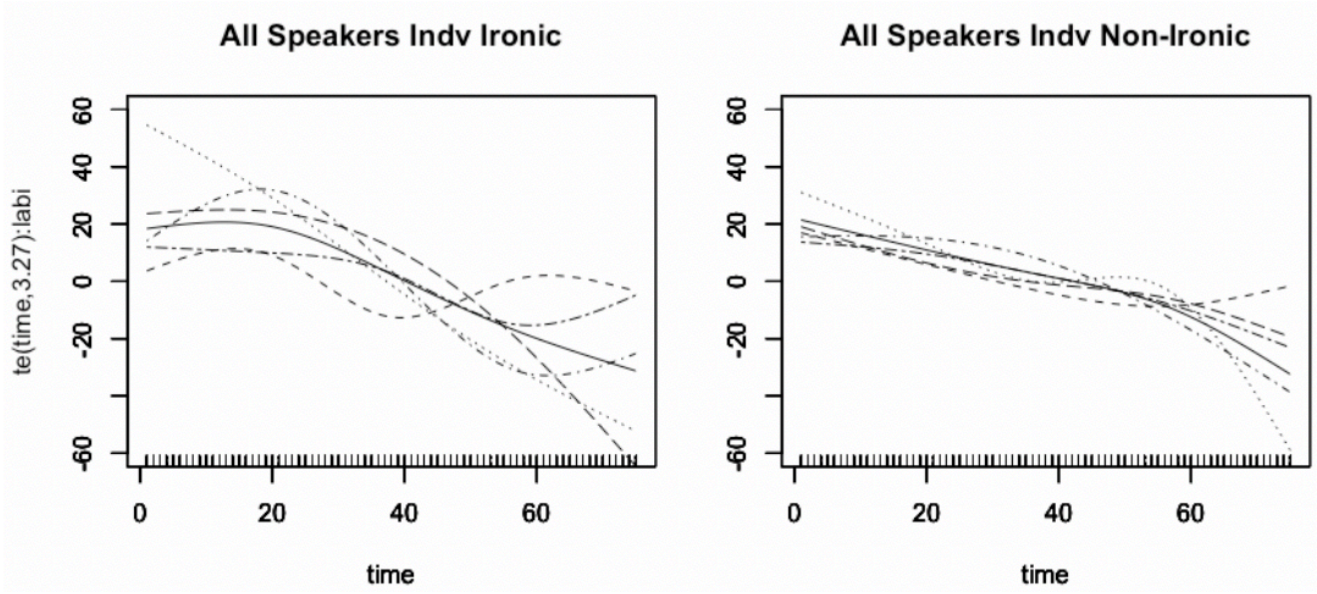
Significant effects were found for the presence or absence of irony over time in all models except individual speaker model for Speaker R. In the All Speakers model, speaker was also a significant effect.

Table 1: GAM results for all models

Model	P value	Deviance Explained
All Speakers	< 0.001 (irony) < 0.001 (speaker)	53.60%
Speaker B	< 0.001 (irony)	9.23%
Speaker G	< 0.001 (irony)	14.80%
Speaker P	< 0.001 (irony)	13.50%
Speaker R	0.084 (irony)	16.70%
Speaker Y	< 0.001 (irony)	6.69%

The significant effect of irony on the f0 over time supports the hypothesis that there would be a statistically significant difference between ironic and non-ironic prosody.

Figure 1: GAM results for all models. Solid line: all speakers; two dash: Speaker B; long dash: Speaker G; dot dash: Speaker P; dotted: Speaker R; dashed: Speaker Y).



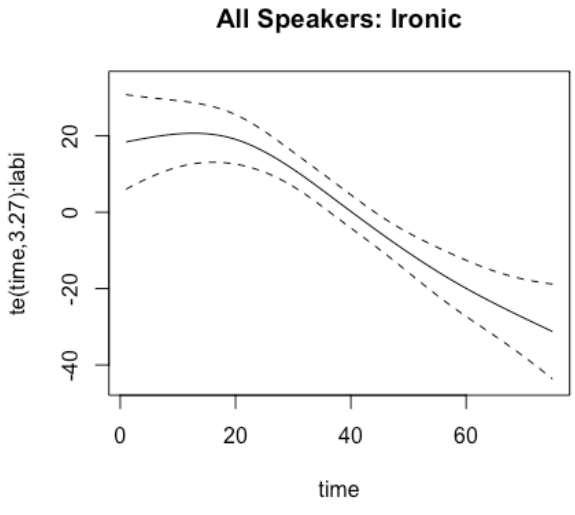
5. DISCUSSION

The significant effect of speaker on f0 over time for the all speakers model supports the hypothesis that each speaker alters prosodic material differently in order to express irony. This offers an explanation for the discrepancies in literature descriptions of irony prosody. Past research, when it has focused on the prosodic content of ironic speech, has largely focused on differences between ironic speech and non-ironic speech across speakers. These results indicate that the difference lies not in some universal cue for irony, but in the changes that individuals make to their speech in order to communicate their meaning. This is illustrated in Figure 1, wherein the f0 lines for all models are overlaid for comparison.

Comparing these two images, non-ironic speech for all speakers and for the group had a very similar f0 line for the majority of the length of an utterance, with the lines diverging dramatically only in the latter fifth or so of the time sample points. The ironic lines are never particularly close, or even parallel. Rather, each individual follows their own pattern separate from the group pattern and the patterns of all other speakers.

The ironic f0 pattern for this group, (illustrated in Figure 2) is characterized primarily by an overall fall, without the late rise-fall one might expect in a language like English which generally places phrase stress toward the end of sentences.

Figure 2: GAM results for the All Speakers model. Dashed lines indicate confidence intervals.



Different speakers utilize this pattern in different ways and to different extents. Speakers B and P have slight rising tails on their ironic f0 lines (Figures 3 and 4). Speakers R and G do not have an initial rise before the fall (Figures 5 and 6), and Speaker G uses a wider range than other speakers.

Figure 3: GAM results for Speaker B (ironic). Dashed lines are confidence intervals.

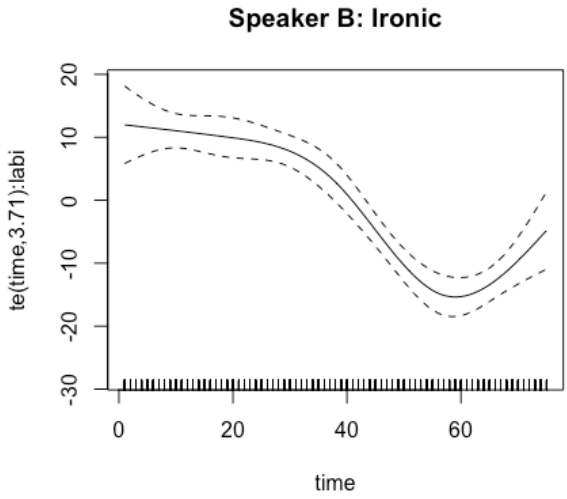


Figure 4: GAM results for Speaker P (ironic). Dashed lines are confidence intervals.

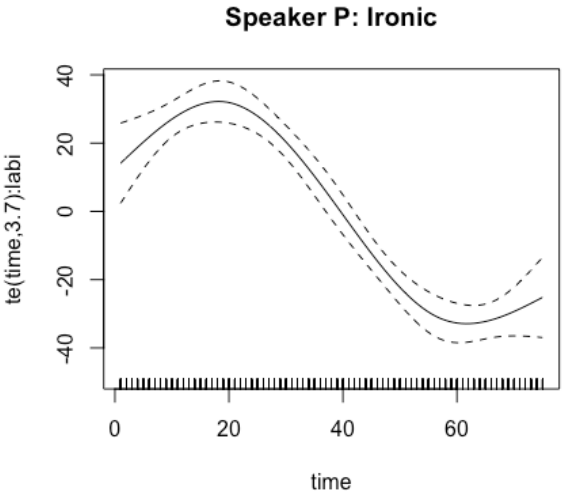


Figure 5: GAM results for Speaker R (ironic). Dashed lines are confidence intervals.

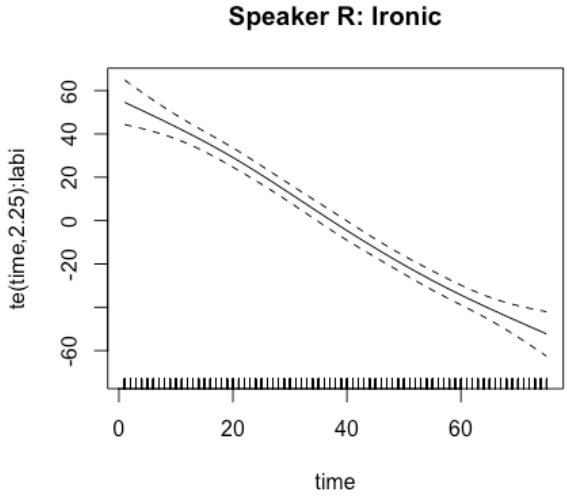
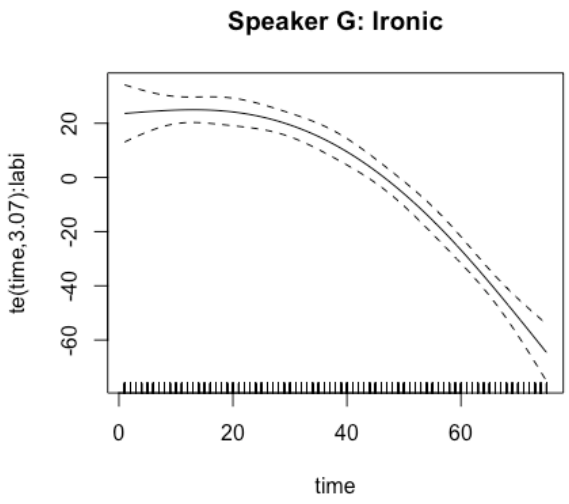
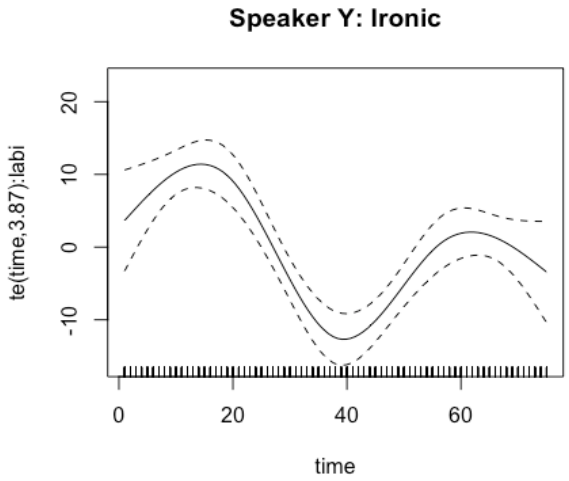


Figure 6: GAM results for Speaker G (ironic). Dashed lines are confidence intervals.



Finally, Speaker Y rarely or never uses the group pattern, such that the smooth curve for his model took an entirely different shape (Figure 6).

Figure 7: GAM results for Speaker Y (ironic). Dashed lines are confidence intervals.



What is the function of the group irony pattern, and why does Speaker Y not conform? The pattern itself is notable, as it is not in keeping with even the most general of literature accounts of irony prosody. Schaffer [8] described irony prosody in terms of "extremes," and yet this group as a whole seems to favour a more muted f0 line, and indeed a seeming suppression of expected phrasal stress patterns. One potential explanation is that the group pattern is a method of in-group signalling – that the muted shape is a deliberate signal that the joke should be obvious from the group's familiarity with the speaker. If this is the case, Speaker Y may not follow the group pattern due to less exposure to the other speakers. Though all five speakers have known each other for years and are close friends, Speaker Y has known the others for the shortest amount of time, and therefore may not have had as much time as the others to absorb this pattern of irony prosody.

If this is the case, it would indicate that irony prosody is not merely individual, but influenced enormously by the intended audience of the ironic utterance. One would use a different version of one's irony prosody when speaking to one friend group or another, to children or adults, in casual or formal settings, and when addressing a wide and varied audience, such as that of an actor on a television show. This would explain why speakers of North American English can summon to mind an 'ironic tone of voice,' but may not always use it in their everyday lives. The 'ironic tone of voice' is the expected prosody for the widest possible audience.

6. CONCLUSION

The presence of both an in-group irony prosody and individual variation indicates that attempting to isolate some universal feature set for ironic prosody may not be a worthwhile pursuit. Rather, it may be more productive for future research to focus on the question of whether or not it is possible to create a predictive model that – given a large enough sample set of an individual's speech in various social contexts and the social context of the utterance in question – could reliably identify irony by prosody for a given utterance. Such a model should include information about stress, duration, and amplitude, as well as f0. Directions for future study include the examination of ironic and non-ironic speech in terms of these domains as well, in order to move closer to a complete picture of what an individual might modify in order to express irony through prosody. Other directions for future research include expanding this method of analysis to more groups of people, studying the speech of people in different social contexts to see how their expression of irony changes, and truly interrogating the content of the ironic prosody seen on television, as this appears to be the closest thing there is to a universal ironic prosody pattern.

The results of this study open the door to a new method of examining the prosody of ironic speech and the potential revitalization of research in this area. While past literature has indicated that the search for a universal irony prosody is not a realistic goal, this study suggests that there is a way forward with the new goal of identifying a method of predicting whether an individual is being ironic based on the prosody of the utterance and past knowledge of the individual. In a world where individuals' speech is increasingly recorded, publicized, and archived, there is a wealth of data readily available for future research in pursuit of this goal.

7. REFERENCES

[1] Anolli, L., Ciceri, R., Infantino, 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] Brown, G., Currie, K., Kenworthy, J. 1980. *Questions of intonation*. London: Croom Helm.

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

[4] Bryant, G., Fox Tree, J. 2002. Recognizing Verbal Irony in Spontaneous Speech. *Metaphor and Symbol* 17(2), 99-119.

[5] Bryant, G., Fox Tree, J. 2005. Is there an ironic tone of voice? *Language and Speech* 48(3), 257-277.

[6] Gibbs, R. 2000. Irony in Talk Among Friends. *Metaphor and Symbol* 15(1-2), 5-27.

[7] R Core Team. 2016. R: A language and environment for statistical computing. Austria: R Foundation for Statistical Computing Vienna. URL: https://www.R-project.org/.

[8] Talkin, D. 2015. REAPER: Robust Epoch And Pitch EstimatoR [GitHub repository]. https://github.com/google/REAPER.

[9] Wood, S., Wood, M.S. 2017. mgcv.