

Is there an Ironic Tone of Voice?

Gregory A. Bryant and Jean E. Fox Tree

University of California, Santa Cruz

Key words

implied language

prosody

sarcasm

*spontaneous
speech*

verbal irony

Abstract

Research on nonverbal vocal cues and verbal irony has often relied on the concept of an *ironic tone of voice*. Here we provide acoustic analysis and experimental evidence that this notion is oversimplified and misguided. Acoustic analyses of spontaneous ironic speech extracted from talk radio shows, both ambiguous and unambiguous in written form, revealed only a difference in amplitude variability compared to matched nonironic speech from the same sources, and that was only among the most clear-cut items. In a series of experiments, participants rated content-filtered versions of the same ironic and nonironic utterances on a range of affective and linguistic dimensions. Listeners did not rely on any set of vocal cues to identify verbal irony that was separate from other emotional and linguistic judgments. We conclude that there is no particular ironic tone of voice and that listeners interpret verbal irony by combining a variety of cues, including information outside of the linguistic context.

1 Introduction

Verbal irony is a form of nonliteral language in which speakers communicate implied propositions that are intentionally contradictory to the propositions contained in the words themselves. In studies of how people use and understand verbal irony, many researchers rely on the concept of an ironic tone of voice (also called sarcastic intonation). We define an ironic tone of voice as some particular consistent prosodic (i.e., pitch, loudness, and duration) pattern with a distinct perceptual correlate that is systematically associated with verbal irony. Developmental studies in particular have relied on a fairly limited (and often vague) category of intonation features that are thought to represent irony such as prolonged articulation and exaggerated pitch

Acknowledgments: This project was made possible by a Special Research Grant from the University of California, Santa Cruz. We thank the many research assistants who helped prepare and run the experiments described here including Michael Andric, Greg Cohen, Sarah Jeffries, Jackie Morck, Jessica Petersen, and Harmony Reppond. Thank you to Don Brenneis, Bruce Bridgeman, Martine Grice, Sylvie Mozziconacci, and Patricia Rockwell for their helpful comments, and special thanks to Ray Gibbs for many excellent ideas during the course of this project.

Address for correspondence. Gregory A. Bryant, Center for Culture, Brain, and Development, UCLA, 1285 Franz Hall, 405 Hilgard Ave., Los Angeles, CA 90095-1563; phone: (310) 825-5326; e-mail: <gabryant@ucla.edu>.

(e.g., Ackerman, 1983; Capelli, Nakagawa, & Madden, 1990; deGroot, Kaplan, Rosenblatt, Dews, & Winner, 1995) though there are notable exceptions (Nakassis & Snedeker, 2002). The studies presented here suggest that there is no particular ironic tone of voice.

1.1

Background

When exploring the nonverbal aspects of verbal irony, researchers typically describe speech produced by actors instructed to speak sarcastically (e.g., Anolli, Ciceri, & Infantino, 2000; Milosky & Ford, 1997; Rockwell, 2000). Actors often indicate sarcasm by lowering their pitch, raising their amplitude, and speaking more slowly (Rockwell, 2000), which is similar to what television actors do (Bryant & Fox Tree, 2001), although results are mixed (Anolli et al., 2000; Attardo, Eisterhold, Hay, & Poggi, 2003). These characteristics have previously been described together as an ironic tone of voice.

But is this what people normally do when speaking ironically? Rockwell (2000) found that without explicit instructions to speak sarcastically, actors' sarcastic utterances were no different from nonsarcastic ones. Spontaneously produced ironic speech might not contain signature acoustic features as has been generally assumed, but instead might comprise of any combination of prosodic features resulting from, for example, word emphasis or emotional expression (Bryant & Fox Tree, 2002).

Prosodic features are often characterized as *local* or *global*. Local features typically make linguistic distinctions and act on segmental and suprasegmental information. These cues often manifest as localized variations in fundamental frequency (F0) and amplitude that affect suprasegmental contours. For example, emphasized syllables are often followed by an abrupt fall (Bolinger, 1989). Semantic and contrastive focus are signaled by local prominences in the F0 and amplitude contours, which may also be followed by low F0 up to the end of the phrase (Cooper, Eady, & Mueller, 1985). There is, in fact, considerable debate about how local pitch movements relate to focus. What some researchers consider to be local prosodic cues might in fact be components of larger (i.e., global) prosodic units (Ladd, 1996).

Global features act on whole utterances, or groups of utterances, and are used to convey emotional and motivational information. This global, or *affective*, prosody generally contains greater F0 variability and F0 range than local prosody as well as more dynamic amplitude features (Frick, 1985). Overall F0 contours are used for affective expression and reliably correlate with basic emotional categories across various languages (Murray & Arnott, 1993). In addition, many animal species exhibit similar global prosodic features in affective communication such as the vocal expression of fear (Hauser, 1996). Shared global contours are also found when expressing emotion in music. For example, music is typically judged to be happy when it is in higher frequencies, a rising contour, and staccato articulation and is generally regarded to be sad when in lower frequencies, a falling contour, and legato, or smooth, articulation (Kratz, 1993). Parallels also exist between angry vocal contours, as expressed by descending F0 contour, staccato articulation, and rising amplitude contour (Johnson,

Emde, Scherer, & Klinnert, 1986), and angry musical passages (Dolgin & Adelson, 1990).

The two types of cues can interact. McRoberts, Studdert-Kennedy, and Shankweiler (1995) found that when contrastive stress increased (manifested as F0 prominence), less F0 final-rise was used to signal a question. But this is only when the F0 peak was associated with contrastive stress. No such trade-off occurred when the F0 peak conveyed affect. This suggests a dissociation between local and global prosody such that linguistic functions are separated from affective functions. One way to explain these phenomena is by means of articulatory constraints (McRoberts, et al., 1995). F0 changes can come about either by controlling vocal fold tension with cricothyroid (CT) laryngeal muscles or with subglottal pressure. The CT contributes to contrastive stress. Because the questions tested were short (3 syllables), CT activity for contrastive stress reduced the ability of the CT to produce the signature final-rise that indicates questions. Affect, however, can be conveyed by subglottal pressure, and so would not necessarily affect the fall-rise. Another possibility is that multiple CT muscles are at work and vocal physiology is functionally organized to perform multiple pitch production tasks (Bryant, 2004a).

Pell (2001) described the trade-offs in what people choose to convey locally and globally as a *prosodic load* problem (p. 1678). When told to place F0 stress on a focus word and produce a question, which requires a terminal F0 rise, people no longer produced local emotional distinctions on key words. Overall F0 differences still varied across emotion types, and other prosodic cues such as speech rate still provided disambiguating affective information, however. Trade-offs in prosodic production reflect strategies used by speakers when production needs outnumber production mechanisms. Speakers allocate prosodic production resources in a manner that maximizes effect and minimizes effort while sparing listeners' perceptual and cognitive resources as much as possible. In implied language such as verbal irony the prosodic load can become quite heavy. This is not to say that verbal irony always demands more prosodic work than other forms of literal or non-literal speech. But in general, any language use that reduces the number of surface propositions but maintains or increases the number of implied propositions must find a means for disambiguation other than the words themselves. The voice is a natural source for this as are the face and other body movements (Attardo et al., 2003).

Intonational phonology also provides a framework from which local and global prosodic features can be understood (e.g., Fujisaki, 1983; Ladd, 1996). In the auto-segmental-metrical (AM) approach, local cues roughly map onto what are called events, and these discrete events are strung together into structured contours. By this account, global prosody consists of discrete components (i.e., events) and the transitions between them. The most important events in a tonal string are pitch accents (associated with prominent syllables within segments) and edge tones (events on the edge of segments) and they manifest as speech melody that can be transcribed according to their movement associated with phrasal structure.

Fujisaki (1983) developed a model of intonation that distinguishes between accent commands and phrase commands. These commands are implemented by functionally distinct mechanisms that send their outputs to a glottal oscillation mechanism that

ultimately controls a final F0 contour. Accent and phrase components can be thought of roughly as linguistic and affective F0 movements respectively.

Approaches such as these resonate well with findings like those described above (i.e., McRoberts, et al., 1995) concerning the dissociation between linguistic and affective pitch production. A proper characterization of how prosodic features relate to language use must decompose the functionally separable aspects of prosodic production. Multiple theoretical perspectives are converging on some similar principles in describing prosody. Upon close examination, however, oversimplified prosodic notions such as an ironic tone of voice are often ruled out. Nevertheless, many people studying verbal irony find the idea compelling, and tend to resort to it in experimental work.

1.2

Summary of current studies

In an earlier paper, we presented preliminary evidence that listeners use local prosodic cues in identifying sarcastic utterances (Bryant & Fox Tree, 2002). The sarcastic and nonsarcastic utterances we tested could not be differentiated when participants read silently to themselves, but could easily be sorted when participants heard them. As the data in this paper will show, these utterances did not differ on several global parameters, and listeners could not differentiate the utterances when presented with only global prosodic information via low and high band-pass filtering. Because they could separate sarcastic from nonsarcastic items when they heard the words, and there were no global acoustic differences between the utterance types, we concluded that local prosodic information must have driven sarcasm identification in these items. That is, participants' judgments of sarcasm were driven by local prosodic cues that interacted with individual lexical items. There is a possibility that judgments were driven by global prosodic interactions with propositional information, but we feel this is less likely given not only the esoteric nature of the utterances (isolated pitch register information is not easily interpreted along with novel, isolated utterances), but also the lack of important baseline speech information.

It is, however, possible that listeners use global cues with other types of sarcastic utterances. Because the items tested above were ambiguous in their written forms, they constituted a special subset of the type of sarcastic utterances generally found on talk radio. To our ears, they could be described as *dry* sarcasm. In the current paper, we selected a new set of talk radio utterances that were, in pilot testing, clearly prosodically unambiguous, which we will label *dripping* sarcasm (see appendix for complete list of utterances and speaker information). The dry sarcastic utterances were also prosodically unambiguous, so the issue here is a matter of degree. Notably, many of the dripping utterances were obviously sarcastic in written form (e.g., "maybe they should call the mental health hotline"). In our observations, people generally produce sarcastic utterances that are both textually and prosodically unambiguous.

We compared dry, dripping, and nonironic utterances across five acoustic dimensions associated with global prosody, or, more specifically, an ironic tone of voice. These are (1) mean F0, (2) F0 range (Max-Min), (3) F0 variability (F0 *SD*), (4) amplitude variability (dB *SD*), and (5) speech rate, as measured by mean syllabic duration. Overall amplitude measurements were not analyzed. Because utterances

were obtained from talk radio, several recording level issues make this measurement useless. Additionally, between-speaker comparisons of amplitude without baseline data are meaningless. Nonetheless, amplitude variability can be compared.

We then measured how low and high band-pass filtered utterances affected listeners' judgments of the original speakers' affective and linguistic intentions. Because listeners could not identify which filtered dry utterances were sarcastic, we stopped testing listeners' affective or linguistic judgments of these items. We continued testing with the filtered versions of the dripping utterances because listeners could identify sarcasm in them. This allowed us to examine whether the perception of verbal irony in filtered utterances would be independent of other affective and linguistic judgments of the speech information. One assumption of the ironic tone of voice hypothesis is that participants' judgments of sarcasm should not be significantly confused with other (nonironic) categories of communicative intention.

All researchers studying verbal irony must inevitably deal with the issue regarding the difference between verbal irony and sarcasm, and research has addressed the issue directly. For example, ridicule seems to play an important role in sarcasm, but not verbal irony in general (Lee & Katz, 1998). By this account, sarcasm is a particular kind of personal criticism leveled against a person or group of persons that incorporates verbal irony. Gibbs (2000) suggested that sarcasm, hyperbole, understatement, rhetorical questions, and jocularity should all be considered kinds of verbal irony. However, people quite often consider instances of verbal irony to be sarcastic (Gibbs, 2000). In previous research, we found that people are much better at defining sarcasm than verbal irony (Bryant & Fox Tree, 2002). Because people are more familiar with the term sarcasm, this word is more appropriate when asking people to make verbal irony judgments.

Low band-pass (LBP) and high band-pass (HBP) filtering involves removing selected bands of acoustic information resulting in an impoverished speech signal. LBP filtered utterances often sound like muffled speech, but the F0 contour is fairly intact. This process removes most acoustic information *above* a designated cut-off point (e.g., ~500Hz) and thus eliminates most features needed for identifying words. HBP filtering removes most acoustic information *below* a designated cut off (e.g., ~2000Hz) and often results in a sound that does not resemble speech at all, but instead as high frequency chirps. One interesting phenomenon of HBP speech is that despite the removal of information needed to calculate F0, listeners do perceive a missing fundamental (Houtsma & Smurzynski, 1990). Research suggests that at least two pitch extraction processes are used for solving this perceptual problem (Winkler, Tervaniemi, & Naatanen, 1997). Speech production events with a concentration of energy in higher frequencies, such as stop consonants and fricatives, are particularly salient in HBP speech. Because of this, certain global phenomena such as stress patterns are easier to identify than in LBP or normal speech (Grant & Walden, 1996). Warren, Riener, Bashford, and Brubaker (1995) found that speech intelligibility was largely retained even in quite narrow bandwidths (e.g., $\frac{1}{20}$ -octave) if the center frequency was around 1500Hz. Moreover, Healy and Warren (2003) found that contrasting patterns of amplitude fluctuations in narrow speech bands contribute significantly to intelligibility.

First we tested whether listeners associated the global prosody remaining in the filtered utterances with affective information other than sarcasm. If speakers use different prosodic patterns to represent different emotions and different tones of voice, then what is identified as an ironic tone of voice should not also be identified as some other emotion. We selected anger from among the emotional possibilities because a lot of sarcastic talk has angry dimensions to it, and so presented the most definitive test of the ironic tone of voice hypothesis. If there is an ironic tone of voice that is truly distinct from an angry tone of voice, then we should not find an effect of anger because we did not select items on that dimension. But even if people are indicating sarcasm in part by acting angry, the relationship between our items and ratings of sarcasm should be stronger than between the items and ratings of anger.

As a further test of the independence of an ironic tone of voice, we tested a non-emotional but nonetheless prosodically dynamic dimension, inquisitiveness. Inquisitiveness is not integrally related to sarcasm and contains particular prosodic properties (Pell, 2001). Although one of our sarcastic items was phrased as a question, the ironic tone of voice hypothesis would predict that listeners would not differentially judge all items based on this orthogonal dimension. The global information retained in our filtered utterances may lead listeners to differentiate items based on inquisitiveness along the same lines as sarcasm and anger: listeners may associate features of sarcasm with the dynamic features of inquisitiveness. People might also detect overacting and label it inquisitiveness, especially if the utterance categories vary greatly in amplitude, suggesting the dynamics of exaggerated questions.

As a control, we tested a linguistic distinction that is generally not made with global prosodic cues, given/new structure. New information is often indicated by altering the pitch on a key word or phrase, a local prosodic cue (Ladd, 1996). Because listeners generally need both a prosodic baseline and the distinctively-pitched words to judge what is new, we expected listeners' ratings of whether speakers were introducing new information to be similar across ironic and nonironic utterances. Our utterances were short, and so unlikely to provide enough material for a baseline judgment. Also, there was no lexical information in the filtered stimuli so listeners would be unable to make tune-text associations (Ladd, 1996). Methodologically, this rating will also show whether or not listeners were merely using the tactic of responding differentially on a rating scale for any systematic acoustic difference they might have identified across the utterance types, regardless of question asked.

As a further control, we tested a sociolinguistic distinction that has no reliable local or global prosodic correlates, the degree of authority over a topic, expressed in our experiment as, "How likely does the speaker know what they are talking about?" Although there is some evidence that people use their voice to be credible and persuasive (e.g., Bugental, 1974; de Groot & Motowidlo, 1999), there are no known precise acoustic correlates of these dimensions. We expected listeners' judgments of authority to not vary across our utterance types.

This rating further tested whether listeners were answering the rating questions according to prosodic features that were conceptually related to the questions. Although in all the filtering experiments people reported that they felt like they were guessing, it is possible that their guesses were largely based on some arbitrary acoustic feature that systematically varied. For example, louder utterances could receive higher

ratings regardless of the question. Similar authority ratings across utterance types would rule out this possibility.

Nonetheless, despite our predictions for authority, the possibility remained that listeners would identify exaggerated prosody in the ironic targets, assume speakers were acting, and conclude that speakers were pretending to be authorities when they were not. Filtering studies offer a way of assessing what people are willing or not willing to infer from global prosodic cues even when there are no reliable acoustic correlates underlying their inferences.

2 Acoustic analyses

Utterances were digitized at 44.1kHz with 16-bit resolution and resampled to 11.025kHz to diminish aliasing (i.e., reduce the chance that extraneous frequency information not in the voices is introduced in the samples). All acoustic analyses were done with Multi-Speech, a Windows-based version of the Computerized Speech Lab (Kay Elemetrics Corp.). Fundamental frequency was determined using voiced period marks (peak impulse locations). All F0 values were converted to semitones (relative to 70 Hz) in order to correct for between-speaker variability issues. Mean syllabic duration was calculated by dividing the total time of the utterance by number of produced syllables (as distinguished from implied syllabic structure) thereby including internal pauses in the measurement. Statistical analyses were performed with SPSS using *t*-tests for independent means.

2.1

Dry sarcasm

In previous work, we selected 24 spontaneously produced textually ambiguous utterances (12 ironic and 12 nonironic) from talk radio programs. We verified experimentally that items were indistinguishable across ironic and nonironic categories when read, but easily distinguished when heard (Bryant & Fox Tree, 2002). Traditional acoustic analysis parameters revealed no differences across utterance categories. Table 1 summarizes these results.

Table 1

Acoustic analysis of textually ambiguous utterances

<i>Variable</i>	<i>Utterance Type</i>	
	<i>Ironic</i>	<i>Nonironic</i>
Mean F0Hz / (semitones)	155.1 (12.8)	142.9 (11.6)
F0 Range Hz / (semitones)	110.1 (32.6)	95.3 (26.5)
F0 Variability Hz / (semitones)	23.3 (3.37)	22.8 (4.03)
Amplitude variability (dB <i>SD</i>)	11.36	11.93
Mean syllabic duration (ms)	187	201

Note. All F0 comparisons done on semitone values. All differences nonsignificant.

Textually ambiguous items did not differ on any global acoustic parameters, suggesting that local cues were responsible for the auditory distinctions people made (Bryant & Fox Tree, 2002). These local prosodic cues must interact with lexical information in order to disambiguate speaker intentions. That is, in order to detect sarcasm, it is necessary to detect the contradiction between the literal meaning and the intended meaning. As we will show below, however, people can assign intentional meaning without hearing words (i.e., without access to a literal meaning).

2.2

Dripping sarcasm

Eleven new sarcastic targets were selected from the same radio talk show sources as the textually ambiguous materials. They were chosen based on exaggerated prosody. The twelfth sarcastic target was held over from the textually ambiguous set because of its high overall ratings for sarcasm and for its perceptually salient prosodic characteristics. The nonironic textually ambiguous targets were used as the comparison group for the current analysis, with three exceptions. These three utterances received higher sarcasm ratings than other nonironic targets ($M = 3.49$) and were replaced with utterances that received lower ratings ($M = 2.14$). If listeners can differentiate targets on ironic content even when filtered beyond lexical identification, then the associated acoustic variables that allow for these distinctions could be informative for future research. These sets were created based on prosodic salience in order to maximize the likelihood of finding an ironic tone of voice. Traditional acoustic analysis parameters revealed moderate differences across categories. Table 2 summarizes the results.

Table 2

Acoustic analysis of textually unambiguous materials

<i>Variable</i>	<i>Utterance Type</i>	
	<i>Ironic</i>	<i>Nonironic</i>
Mean F0Hz / (semitones)	176.4 (15.3)	142.7 (11.2)*
F0 Range Hz / (semitones)	122 (25.7)	92.6 (32.7)
F0 Variability Hz / (semitones)	27.4 (2.64)	23 (2.69)
Amplitude variability (dB <i>SD</i>)	14.5	16.1**
Mean syllabic duration (ms)	233	218

Note. All F0 comparisons done on semitone values * $p < .10$ ** $p < .05$

Although there were no reliable global differences between dry ironic and nonironic utterances, there were differences between dripping ironic and nonironic utterances. In these materials, dripping utterances tended to be higher pitched. This finding has been reported in research using actors, where those acting *sarcastic irony* produce higher pitched utterances than those acting *kind irony* (Anolli et al., 2000), but other studies have found the opposite effect (Rockwell, 2000). The utterances also

differed in amplitude variability with less variability in ironic utterances. This has not been reported before with actors. One surprising result is that the utterances did not differ with regard to duration. The typical finding that tends to corroborate with stereotypical notions of sarcastic speech is that sarcasm is spoken slower, and this has been found in spontaneous speech as well (Bryant, 2004b). While mean syllabic duration was slightly longer in ironic targets, this difference was not reliable.

3 Low and high band-pass filtered speech: Dry ironic versus nonironic

We assessed the degree to which listeners use low and high frequency information in making judgments of ironic intent by filtering words out of utterances and asking listeners to rate sarcasm. Given the lack of any reliable global acoustic differences between dry and nonironic utterances before filtering, we did not expect listeners to be able to distinguish filtered versions of the same utterances.

3.1

Participants

In this and all of the following studies, participants were native English speaking University of California, Santa Cruz students who participated for course credit. Fifty students participated in the low band-pass dry versus nonironic experiment, and 30 in the high.

3.2

Materials

3.2.1

Low band-pass filtered

Content filtering has been used widely in experimental research for the purpose of removing lexical information from the speech signal while leaving basic prosodic information intact (Rogers, Scherer, & Rosenthal, 1971). The 24 utterances were low band-pass filtered using the FFT filter function in Cool Edit Pro. All utterances were filtered initially at 500 Hz (all frequencies above 500 Hz were reduced with a 60 dB per octave roll-off). Utterances varied in intelligibility when filtered at 500 Hz, so in problem areas where clear articulation remained on single words or syllables, lower cut off values (all > 375 Hz) were used on small segments never exceeding 1000 ms. The primary objective of this filtering procedure was to eliminate lexical identification across all utterances with the minimal amount of signal loss, so slightly different settings were used to accommodate the variability in the different speech samples. Utterances were then normed for intelligibility with a group of participants from the same pool as the main experimental participants. Three rounds of 10 participants each were conducted until utterances were filtered just below the threshold of lexical identification by naïve listeners. One-word identifications were allowed for some utterances ($N = 6$), but no identifications were permitted beyond that. Utterances did not differ across types in overall amplitude or length.

Due to any number of technical audio issues (e.g., recording levels on original tape), overall amplitude differences not attributable to the speaker could confound the rating data (e.g., louder recordings could be judged as more angry).

3.2.2

High band-pass filtered

The 24 utterances were high band-pass filtered using the FFT filter function in Cool Edit Pro. All utterances were initially filtered at 1750 Hz (all frequencies below 1750 Hz were reduced with a 60 dB per octave roll-off). Utterances varied in intelligibility, so higher cut-off frequencies were used never exceeding 3250 Hz or on segments longer than 500 ms. Utterances were then normed for intelligibility with participants from the same pool as the main experimental participants. Two rounds of 10 participants each were conducted until utterances were filtered just below the threshold of any lexical identification by naïve listeners. One-word identifications were allowed for some utterances ($N=4$), but no identifications were permitted beyond that. Utterances did not differ across types in overall amplitude or length.

Utterances were transferred to cassette tape, with one tape for low band-pass filtered utterances and another for high.

3.3

Procedure

In this experiment and all of the following filtering experiments, participants were informed that they would be listening to filtered segments of spontaneous speech that would be quite difficult to understand, and they were asked to try to rate the utterances as best as they could. For each trial, participants heard a target utterance and rated the degree of sarcasm on a seven-point scale from (1) not sarcastic to (7) very sarcastic. A definition of sarcasm was not provided. Participants were also instructed to write down any words they thought they heard. Participants were told to not rewind the tape, but were allowed to stop it if necessary. Participants were tested individually. There was approximately five seconds from the offset of one trial to the announcement of the next trial.

4 Results and Discussion

4.1

Low band-pass filtered

As predicted, participants rated nonironic utterances ($M=3.3$, $SD=0.8$) and ironic utterances ($M=3.2$, $SD=0.9$) almost identically, $t1(49)=0.59$, $p=ns$, $t2(22)=.77$, $p=ns$. Some words were identified, but no systematic identifications occurred.

4.2

High band-pass filtered

Ironic utterances ($M=3.5$, $SD=0.9$) were rated slightly more sarcastic than nonironic utterances ($M=3.3$, $SD=0.5$) but this difference was not significant across participants,

$t1(29) = 1.46, p = ns$, or items, $t2(22) = .65, p = ns$. Some words were identified, but no systematic identifications occurred.

In previous work with the dry sarcastic materials, listeners did not judge ironic items as more ironic than nonironic items until information from the voice was available (Bryant & Fox Tree, 2002). This suggested that prosodic cues were informing their decisions. But listeners were unable to reliably identify originally ironic targets hearing only filtered speech. Because there were no differences in ratings of sarcasm, we did not ask any further questions on perceived affective or linguistic information for these items.

5 Low and high band-pass filtered speech: Dripping ironic versus nonironic

As with the textually ambiguous items, we assessed the degree to which listeners used low and high frequency information in making judgments of speaker intent by presenting content filtered sentences and asking participants to rate them on various dimensions. When reading text silently with no auditory or contextual information, participants ($N = 34$) were able to distinguish ironic from nonironic utterances. Ironic utterances ($M = 4.54, SD = 0.69$) were rated as more sarcastic than nonironic utterances ($M = 3.46, SD = 0.63$) and this difference was significant across participants and items, $t1(33) = 10.45, p < .001$; $t2(22) = 2.47, p < .05$.

5.1

Participants, materials, and procedure

The materials were low (LBP) and high band-pass (HBP) filtered as described above. Some words were identified, but no systematic identifications occurred. Two rounds of 10 participants each were conducted until utterances were filtered just below the threshold of any lexical identification by naïve listeners. One-word identifications were allowed for some utterances ($N = 3$), but no identifications were permitted beyond that. Utterances did not differ across types in overall amplitude or length.

Thirty-six students participated in the LBP *sarcasm* experiment, and 30 participated in the HBP. They rated the degree of sarcasm on a seven-point scale from (1) not sarcastic to (7) very sarcastic. Thirty-five students participated in the LBP *anger* experiment, and 30 in the HBP. They rated the degree of anger on a seven-point scale from (1) no anger to (7) very angry. Thirty-three students participated in the LBP *inquisitiveness* experiment, and 31 in the HBP. They rated the degree of inquisitiveness on a seven-point scale from (1) not inquisitive to (7) very inquisitive. Thirty-one students participated in the LBP *given/new* experiment, and 32 in the HBP. They rated the likelihood of the speaker's providing new information on a seven-point scale from (1) not likely to (7) very likely. Twenty-nine students participated in the LBP *authority* experiment, and 31 in the HBP. They rated the likelihood of a speaker's knowing what he/she was talking about on a seven-point scale from (1) not likely to (7) very likely. Other aspects of the procedure followed those of the prior experiments.

6 Results

In both LBP and HBP experiments, ironic utterances (Low: $M=3.7$, $SD=0.9$; High: $M=3.6$, $SD=1.2$) were rated as more sarcastic than nonironic utterances (Low: $M=2.8$, $SD=0.8$; High: $M=3.2$, $SD=0.9$). This difference was significant across participants and across items for LBP materials, $t1(35)=9.24$, $p<.001$, and $t2(22)=2.51$, $p<.05$, but only across participants for HBP materials, $t1(29)=2.9$, $p<.01$, and $t2(22)=1.38$, $p=ns$.

In both LBP and HBP experiments, ironic utterances (Low: $M=3.8$, $SD=0.8$; High: $M=3.8$, $SD=.7$) were rated as more angry than nonironic utterances ($M=3$, $SD=.9$; High: $M=2.9$, $SD=1$). This difference was significant across participants and marginally significant across items for both LBP and HBP materials, $t1(34)=9.2$, $p<.001$, and $t2(22)=1.9$, $p=.07$ for low, and $t1(29)=5.4$, $p<.001$, and $t2(22)=1.9$, $p=.07$, for high.

In both LBP and HBP experiments, ironic utterances (Low: $M=3.7$, $SD=0.7$; High: $M=3.5$, $SD=1$) were rated as more inquisitive than nonironic utterances (Low: $M=3.4$, $SD=0.7$; High: $M=2.9$, $SD=1$). This difference was significant across participants but not across materials for LBP materials, $t1(32)=2.50$, $p<.05$, and $t2(22)=1.19$, $p=ns$. For HBP materials, both were significant, $t1(32)=4.93$, $p<.001$, and $t2(22)=2.37$, $p<.05$.

Ratings for new information differed dramatically across LBP and HBP experiments. LBP ironic utterances ($M=4.01$, $SD=0.83$) were rated just as likely to contain new information as nonironic utterances ($M=4.04$, $SD=0.79$), $t1(30)=0.14$, $p=ns$, and $t2(22)=.09$, $p=ns$. But HBP ironic utterances ($M=4.0$, $SD=0.81$) were rated as more likely to contain new information than nonironic utterances ($M=3.5$, $SD=0.77$) across both participants, $t1(33)=3.51$, $p<.001$, and items, $t2(22)=2.2$, $p<.05$.

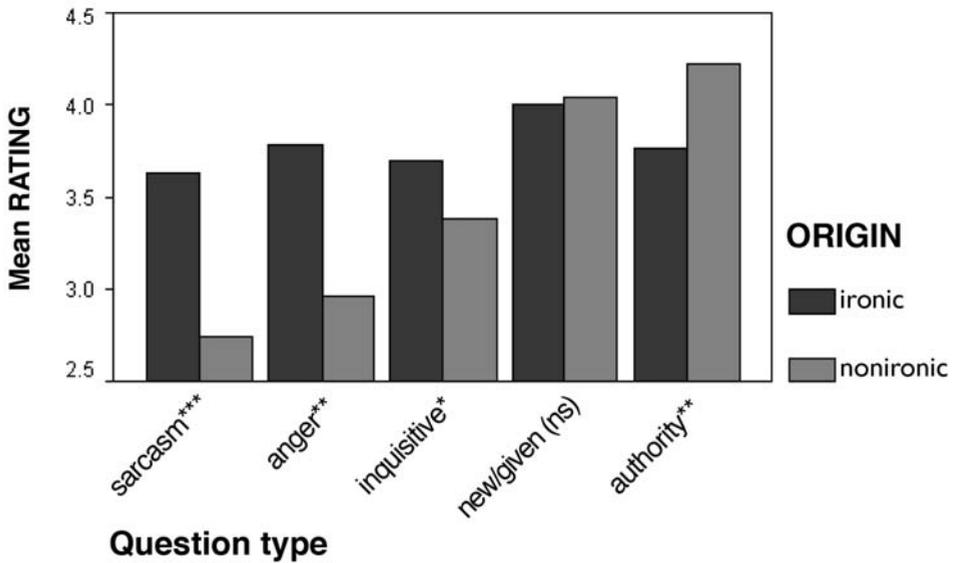
Ratings for authority also differed dramatically across LBP and HBP experiments. LBP ironic utterances ($M=3.8$, $SD=0.80$) were rated as less likely to be produced by a speaker that knew what they were talking about than nonironic utterances ($M=4.2$, $SD=0.89$). This difference was significant across participants, $t1(28)=4.21$, $p<.001$, and marginally significant across items, $t2(22)=1.96$, $p=.06$. But HBP ironic utterances ($M=3.71$, $SD=0.88$) were not rated as more likely to be produced by a speaker that knew what they were talking about than nonironic utterances ($M=3.69$, $SD=0.86$), $t1(30)=0.26$, $p=ns$, and $t2(22)=.08$, $p=ns$. See Figures 1 and 2 for summaries of results across filtering experiments.

7 Discussion

Listeners clearly distinguished between dripping ironic and nonironic utterances when provided either just LBP or HBP information. At first glance, these data suggest that there possibly is such a thing as an ironic tone of voice. Of the five dimensions we had listeners rate the utterances on, sarcasm was the only one in which the targets actually differed categorically. But listeners differentiated between the two types of utterances on a few dimensions, and with varying acoustic information.

Figure 1

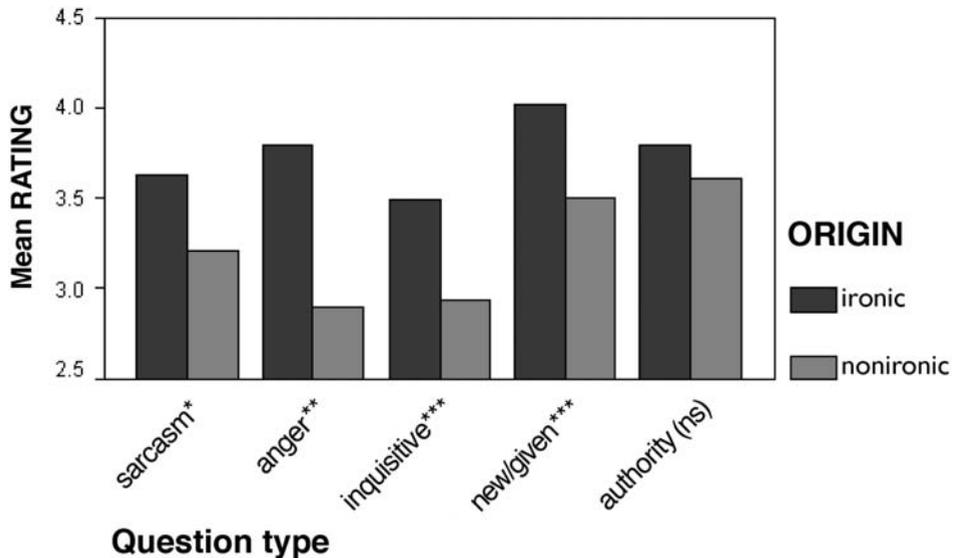
Rating differences across question types for low band-pass filtered utterances



***= t1 and t2 sig. **= t2 marg. sig. *= t2 not sig. ns = not sig.

Figure 2

Rating differences across question types for high band-pass filtered utterances



***= t1 and t2 sig. **= t2 marg. sig. *= t2 not sig. ns = not sig.

Because the items analyses were not significant in five of the 10 experiments where the participant analyses did reach significance, it is possible that a limited number of items was driving the effect across participants in those experiments. Due to the impoverished nature of the stimuli, perhaps only particularly unambiguous items were rated systematically. We found only one item that deviated from the norm for others (Item 9, see the Appendix). This nonironic item received exceptionally low ratings for sarcasm in the HBP condition ($M = 1.6$; All nonironic items: $M = 3.2$, $SD = 0.9$) and somewhat low ratings for inquisitiveness in the LBP condition ($M = 2.5$; All nonironic items: $M = 3.4$, $SD = 0.7$). When this item is removed from these analyses, means become only marginally different across participants in both experiments. The pitch, loudness, and speech rate characteristics of this utterance were not remarkable in comparison to the other nonironic items, but one notable feature was a non-fluctuating, staccato rhythmic quality that is recognizable in all states (normal, LBP, and HBP). On a postexperiment questionnaire, participants were asked how they made their judgments, and they often reported that they tried to identify speed and tone interactions for both sarcasm and inquisitiveness (e.g., “when the voice changed from low to high to low and got louder and faster”). This suggests rhythmic properties independent of global pitch or amplitude attributes were informing participants’ decisions at least about sarcasm and inquisitiveness. F0 excursions and speed changes have also been observed to affect other speech judgments such as liveliness (Traunmuller & Eriksson, 1995). In our studies, there were no systematic comments for other judgments. These data raise the possibility of what might be fair to call an ironic *rhythm* of voice, a topic we defer to future investigations.

Although the targets actually only differed on their levels of sarcasm, listeners also distinguished utterances based on anger, with dripping sarcastic utterances rated as more angry. So, for these materials, an ironic tone of voice could also be described as an angry tone of voice. In other materials, sarcasm and anger might not have overlapped so much. Listeners might have used amplitude variability to make anger judgments given that amplitude is an important cue for anger (Frick, 1985), and one dripping target with noticeably higher anger ratings had a marked rising amplitude contour.

Listeners’ categorizations were not restricted to affective judgments. Dripping items were not only rated as more angry, but more inquisitive as well. Listeners’ systematic categorization based on inquisitiveness surprised us. But as the given/new and authority ratings show, the systematic categorization is not likely to be a methodological artifact. That is, if it is a methodological artifact, the ratings should differ regardless of the test question. Instead, there is something about the global prosodic features that listeners tap into when making inquisitiveness judgments. For example, many participants commented that they used vocal activity at the end of utterances as a cue to sarcasm, but vocal activity at the end of utterances is also a characteristic of questions. However, all but one item in our materials were actually statements and the items did not differ across conditions with regard to high-rise terminals, decreasing the likelihood that activity at the end of the utterances was driving this effect. It is also unlikely that listeners were using amplitude variability, as there is no evidence, to our knowledge, that amplitude variability systematically varies across declaratives and interrogatives.

Listeners might have used pitch differences. The item that received the second highest rating for inquisitiveness was the one interrogative, and it contained the signature final pitch rise that is the most salient prosodic feature of questions. Pitch cannot entirely account for the categorizations, however. When reliable F0 information was available (LBP condition), listeners did not distinguish between ironic and ironic utterances for inquisitiveness. It is possible that other prosodic features affected pitch extraction processes that infer the missing fundamental. In addition, judgments could have been swayed by participants' expectations of interrogatives given the task demands. No other utterances contained high rise terminals, a phenomenon relatively common in declaratives in some contemporary English dialects.

Unlike the other categorizations, listeners' categorizations of given/new and authority were not systematic. Listeners did not use LBP information in making given/new distinctions, and they did not use HBP information in making authority distinctions. They used HBP information to categorize dripping items as more likely to be new, they used LBP information to categorize nonironic items as more likely to be spoken by an authority.

Although there are prosodic features that help distinguish between given and new information (e.g., amplitude and pitch help identify what is in focus), without baseline prosodic information or, more importantly, lexical information, one might expect that listeners would not make any systematic given/new distinctions. This expectation is also consistent with a metrical phonology approach where a listener requires a tune-text association in order to correctly use pitch accent information (Ladd, 1996). This was the case for the LBP results but not for the HBP results. One reason listeners might have categorized the HBP stimuli differently is that, as mentioned earlier, stress patterns are easier to identify in HBP speech. Listeners might have used amplitude stress patterns (i.e., amplitude prominences that reveal rhythmic properties) to categorize HBP speech as new, and the more prominently stressed items might have coincided with the ironic utterances. If participants relied on pitch information to make their judgments then they should have distinguished the utterance types in the LBP condition, but they did not. Sarcasm is often reported to be associated with pronounced stress (Murray & Arnott, 1993) and this could be confused with new information when verbal content is filtered and amplitude stress patterns are salient. Verbal irony can be seen as an instance of providing new information: either the expression of a new, attributed attitude toward an old proposition (Sperber & Wilson, 1986, 1995), the recognition of a speaker pretending to be a person that would say the ironic utterance literally (Clark, 1996), or a combination of the two (e.g., Gibbs, 2000).

There are no known reliable prosodic correlates of authority, so listeners should not make any systematic authority distinctions especially on utterances that do not actually differ on that dimension. This was the case for the HBP condition, but not for the LBP. While the utterances only differed acoustically in amplitude variability, some distinguishing dynamic in the prosody could have been relevant for judgments of authority and obscured in the HBP condition. If listeners detect pretense in the ironic items and interpret that as insincere language, we might expect the nonironic items to be more associated with speakers who believe what they are saying, or take themselves as an authority. This concurs with research on credibility and the voice.

Speakers who are identified as experts and speak in a more spontaneous, less scripted style are rated as more credible (Bugental, 1974).

8 General Discussion

From the standpoint of numerous theoretical approaches to verbal irony (e.g., Clark, 1996; Sperber & Wilson, 1986, 1995), it is not surprising that the prosody of ironic speech should vary, even with actors. Speakers are communicating multiple messages by layering propositional and non-propositional information, and providing prosodic cues in contextually dependent ways that map differentially onto the simultaneously presented information (e.g., a pretend anger pitch contour superimposed onto laughing to assist a listener with an ironic implicature). The affective prosody serves to communicate metarepresentational and direct affective information. We should expect a variety of prosodic strategies to accommodate this incredible diversity of nested emotional and propositional messages. Attempts to make one-to-one mappings between acoustic parameters and ironic content seem futile from this angle.

These data speak to the high degree of similarity and overlap in which varying emotional and linguistic messages can manifest prosodically and conceptually. An utterance can be sarcastic, angry, inquisitive, provide new information, and be spoken with authority all at the same time (e.g., upon learning that their child has wrecked the family car, a parent could exclaim, “Were you like me at your age, and thought that you were immune to driving accidents...because...‘oh, I’m special!?’”). Understanding the prosodic production behind such a sentence requires an analysis that breaks down the component propositions and considers the varying impact attempts at disambiguation may have at different levels. What seems like an ironic tone of voice is likely an emergent product of interpretations informed by multiple sources of information, many not acoustic. The folk notion of sarcasm as a fairly uniform category of language use could contribute to the illusion of prosodic consistency that an ironic tone of voice implies.

It is unclear how local and global prosody manifest in filtered speech. Listeners can clearly infer many properties of a speech signal from a fairly impoverished stimulus. As mentioned earlier, listeners are able to make fairly accurate estimations of the missing fundamental with only high-order harmonics of a complex tone (Houtsma & Smurzynski, 1990). Many of the similarities in our results between the high and low band-pass conditions could be due to listeners inferring F0 based on the limited high frequency information. This could account for the sarcasm, anger, and inquisitive responses being similar across filtering conditions. Perhaps listeners were relying more on pitch contours to answer these questions. By contrast, in the two conditions where global pitch information independent of lexical information is not useful, listeners might have relied on amplitude variability information (amplitude variability was reliably different between utterance types, and mean F0 was marginally different). This information was salient in the HBP condition, but obscured in the LBP condition. By this reasoning, we should expect that the given/new distinction would be difficult with LBP stimuli, and easier with the HBP stimuli, which is exactly what occurred.

This account does not explain the responses to the question regarding authority of the speaker. In the HBP condition, listeners apparently did not make use of either the missing fundamental, or amplitude variability in making judgments regarding authority over a topic. This question was partially included as a control as there is no empirical evidence to suggest that listeners should be able to perform such a task. As mentioned earlier, in the LBP condition, participants may have inferred pretense in the ironic items, and thus judged nonironic utterances to be more authoritative. Another less likely possibility, though not mutually exclusive, is that the marginally significant overall F0 difference may have lead listeners to make sex stereotypical judgments related to gender and authority. That is, the ironic utterances were higher pitched, and so were possibly identified as more likely to be female or a child (Traunmuller & Eriksson, 1995), and therefore possessing less authority. Another (non-mutually exclusive) possibility is that lower voices, on average, are perceived as more dominant, and thus, more authoritative (e.g., Putz, 2004).

Other acoustic variables need to be examined in ironic speech. For example, some researchers have speculated that nasality is indicative of a sarcastic intention (Cutler, 1974; Haiman, 1998). One might even suggest an embodied relationship between the emotion of disgust and voicing that represents removing air from the nasal passage (Don Brenneis, personal communication). The stimuli obtained from the radio examined here do not afford the type of audio quality needed for reliable long-term spectral mean analyses, but the percept of nasality has spectral slope correlates, and this could be a fruitful approach (Tsang & Trainor, 2001). It would be an acoustic dimension such as this that may actually provide some rationale for the folk notion of an ironic tone of voice. If this were the case, one could still argue that nasality is not specific to verbal irony, but merely more closely associated with it than other global properties such as those examined in the present study.

Another time when spectral slope may shift in verbal irony is when a speaker engages in pretense. By acting out some scenario where conversationalists take on other personas, vocal changes not only signal that an “act” is happening, but provide participants a medium with which to carry out many kinds of communicative goals (Clark, 1996). Vocal imitation allows speakers to comment on a whole variety of aspects related to personalities, behaviors, and attitudes of others. Prosodic shifts from baseline speech are used by natural conversationalists when speaking ironically (Bryant, 2004b). Speakers contrast pitch, loudness, and speech rate dimensions of verbal irony utterances with baseline speech presumably in an effort to signal ironic intentions, but it is not yet known exactly how these shifts affect listeners’ interpretations. Prosodic contrasts are not special to verbal irony, but instead apply to any number of speech phenomena.

The acoustic analyses and perceptual studies of spontaneous ironic speech presented here provide very little support for the notion of an ironic tone of voice; that is, prosodic consistency across verbal irony utterances. The perception of any such tone appears to be a result of the integration of multiple sources of information (including, we believe, non-acoustic) and thus likely more an illusion than an actual speech production phenomenon. This is not to say that particular vocal characteristics do not accompany categories of language in systematic ways, but rather, extensive discussion and description at that level of analysis may be an exercise in futility.

Researchers should examine how language use interacts with prosodic production and perception by breaking up the exploration into more specific problems such as those related to sentence focus, reference, and emotional communication.

manuscript received: 09. 14. 2004
manuscript accepted: 08. 04. 2005

References

- ACKERMAN, B. P. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology*, **35**(3), 487–508.
- 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.
- ATTARDO, S., EISTERHOLD, J., HAY, J., & POGGI, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, **16**(2), 243–260.
- BOLINGER, D. (1989). *Intonation and its uses*. Stanford, CA: Stanford University Press.
- BRYANT, G. A. (2004a). *The functional organization of vocal physiology for pitch production*. Paper presented at the 16th Annual Meeting of the Human Behavior and Evolution Society, Berlin, Germany, July, 2004.
- BRYANT, G. A. (2004b). *Prosodic features of verbal irony in spontaneous speech*. Doctoral dissertation, University of California, Santa Cruz, Department of Psychology.
- BRYANT, G. A., & FOX TREE, J. E. (2001). *Is there an ironic tone of voice? Yeah, right*. Paper presented at the annual meeting of the Western Psychological Association, Maui, HI, May, 2001.
- BRYANT, G. A., & FOX TREE, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, **17**(2), 99–117.
- BUGENTAL, D. E. (1974). Interpretations of naturally occurring discrepancies between words and intonation: Modes of inconsistency resolution. *Journal of Personality and Social Psychology*, **30**(1), 125–133.
- CAPELLI, C. A., NAKAGAWA, N., & MADDEN, C. M. (1990). How children understand sarcasm: The role of context and intonation. *Child Development*, **61**, 1824–1841.
- CLARK, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- COOPER, W. J., EADY, S. J., & MUELLER, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, **77**, 2142–2156.
- CUTLER, A. (1974). *On saying what you mean without meaning what you say*. Papers from the Tenth Regional Meeting of the Chicago Linguistics Society (pp. 117–127). Chicago: Department of Linguistics, University of Chicago.
- De GROOT, A., KAPLAN, J., ROSENBLATT, E., DEWS, S., & WINNER, E. (1995). Understanding versus discriminating nonliteral utterances: Evidence for a dissociation. *Metaphor and Symbol*, **10**(4), 255–273.
- De GROOT, T., & MOTOWIDLO, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, **84**(6), 986–993.
- DOLGIN, K. J., & ADELSON, E. H. (1990). Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music*, **18**, 87–98.
- FRICK, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, **97**(3), 412–429.

- FUJISAKI, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter MacNeilage (Ed.), *The production of speech* (pp.39–55). Heidelberg: Springer-Verlag.
- GIBBS, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*, **15**(1 & 2), 5–27.
- GRANT, K. W., & WALDEN, B. A. (1996). Spectral distribution of prosodic information. *Journal of Speech and Hearing Research*, **39**(2), 228–238.
- HAIMAN, J. (1998). *Talk is cheap: Sarcasm, alienation, and the evolution of language*. New York: Oxford University Press.
- HAUSER, M. (1996). *The evolution of communication*. Cambridge: MIT Press.
- HEALY, E. W., & WARREN, R. M. (2003). The role of contrasting temporal amplitude patterns in the perception of speech. *Journal of the Acoustical Society of America*, **113**(3), 1676–1688.
- HOUTSMA, A. J. M., & SMURZYNSKI, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America*, **87**(1), 304–310.
- JOHNSON, W. F., EMDE, R. N., SCHERER, K. R., & KLINNERT, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry*, **43**, 280–283.
- KRATUS, J. (1993). A developmental study of children's interpretation of emotion in music. *Psychology of Music*, **21**, 3–19.
- LADD, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- LEE, C. J., & KATZ, A. N. (1998). The differential role of ridicule in sarcasm and irony. *Metaphor and Symbol*, **13**(1), 1–15.
- McROBERTS, G. W., STUDDERT-KENNEDY, M., & SHANKWEILER, D. P. (1995). The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation. *Perception and Psychophysics*, **57**(2), 159–174.
- MILOSKY, L. M., & FORD, J. A. (1997). The role of prosody in children's inferences of ironic intent. *Discourse Processes*, **23**, 47–61.
- MURRAY, I. R., & ARNOTT, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, **93**(2), 1097–1108.
- NAKASSIS, C., & SNEDEKER, J. (2002). Beyond sarcasm: Intonation and context as relational cues in children's recognition of irony. In A. Greenhill, M. Hughs, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the Twenty-sixth Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
- PELL, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *Journal of the Acoustical Society of America*, **109**(4), 1668–1680.
- PUTZ, D. A. (2004). *Female choice, male dominance, and the evolution of low voice pitch in men*. Doctoral dissertation, University of Pittsburgh, Department of Anthropology.
- ROCKWELL, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, **29**(5), 483–495.
- ROGERS, P. L., SCHERER, K. R., & ROSENTHAL, R. (1971). Content filtering human speech: A simple electronic system. *Behavior Research Methods and Instrumentation*, **3**(1), 16–18.
- SPERBER, D., & WILSON, D. (1986, 1995). *Relevance: Communication and cognition*. Cambridge: Harvard University Press.
- TRAUNMULLER, H., & ERIKSSON, A. (1995). The perceptual evaluation of F0 excursions in speech as evidenced in liveliness estimations. *Journal of the Acoustical Society of America*, **97**(3), 1905–1915.
- TSANG, C. D., & TRAINOR, L. J. (2001). *Nasality and spectral slope: Infants are tuned to the timbres of speech and music*. Presented at the American Psychological Society, Toronto, Canada, June, 2001.

- WARREN, R. M., RIENER, K. R., BASHFORD, J. A., & BRUBAKER, B. S. (1995). Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits. *Perception and Psychophysics*, **57**, 175–182.
- WINKLER, I., TERVANIEMI, M., & NAATANEN, R. (1997). Two separate codes for missing-fundamental pitch in the human auditory cortex. *Journal of the Acoustical Society of America*, **102**(2), 1072–1082.
-

Appendix

Dripping Ironic and Nonironic Utterances

The following 24 utterances were extracted from various talk radio programs between January, 1999 and March 2000. The programs were of varied types but mostly political. All utterances were spontaneously produced in conversation by either hosts or people being interviewed. Sex of speakers in parentheses, and asterisks indicate duplicate speakers.

Originally Ironic Items

1. yeah that shocked me too. (M)
2. obviously they were all crooks. (M)
3. hey, you know what? give him his kid back too. (M) **
4. that's tough to spell. (F)
5. she didn't say anything to Clinton about a pardon. (M) *
6. never would've thought that one out, huh? (M)
7. maybe they should call the mental health hotline. (F)
8. he'll win a lot of people over to his way of thinking that way. (M) **
9. sure you do. (F)
10. wow, there's a concept. (M)
11. thanks a lot for redeeming me. (F)
12. he's a hip guy, he's funny, and he's smart. (M) *

Originally Nonironic Items

1. this is a good thing. (M)
2. it violates the laws of physics on a very basic level. (M)
3. to me that is not worth the price. (M)
4. they really make life a lot of fun. (F)
5. oh, they are ingenious. (F)
6. of course you wouldn't know that by looking at me now. (M)
7. that sounds like fun. (F)
8. you know, Katie's been trying so hard. (M) *
9. because obviously you don't stop there. (M)
10. I do it for the money and nothing else. (M)
11. imagine that. (M) **
12. they are really some really smart people. (M)]

note: * same speaker; ** same speaker