Acoustic Prosodic Features In Sarcastic Utterances

Ahmed Abu-Shnein
University of Kufa

Introduction:

The main goal of this study is to determine if sarcasm can be detected through the analysis of prosodic cues or acoustic features automatically. If we obtain an acceptable accuracy, then the machine can be used instead of human detectors. This study lies in the fields of Psycholinguistics, Computational linguistics, and Natural Languages Processing.

Sarcasm can be defined as the use of expressions with a completely different meaning that is far from its literal meaning. The most common definition in dictionaries implies irony to intend to offend or mock other peoples' behavior or comments. Nevertheless, sarcasm in spoken language can be used with the purpose to simply avoid answering questions directly. To understand when an expression is used for this purpose, we usually rely on the context where the expression takes place, and on the cultural knowledge shared by the speakers. Also, the acoustic features (physical features of sounds) and the prosodic cues (suprasegmental features of utterances) used in the expressions are considered important factors in the understanding of the real meaning of the utterance. In this sense, Voyer (2008) describes sarcasm as a special situation in speech perception as it reflects a case where literal and prosodic contents are incongruent. The recognition of sarcastic speech, from this point of view, necessarily implies the study of prosody as a critical component of the understanding.

Since sarcasm is used very often in daily conversations, several studies in natural languages processing have been conducted to propose which linguistic cues can be used for systematic detection of sarcasm the thing that could help in using the machine in sentiment analysis in a broader sense or even on a larger scale. That is the case with Tepperman et al. (2006). In their study they propose some prosodic cues to detect sarcasm in spoken language. More precisely, they studied the expression “Yeah right” which is widely used in American English with both sarcastic and literal meaning which is the neutral meaning in this context. In this project, we propose to study the same expression “yeah right” in spoken language to detect and compare when the expression implies sarcasm or when it is merely used to follow the conversation, which can be said as being a neutral utterance. The previous study tried to train an automatic sarcasm recognizer; meanwhile our goal is to use a computer program that is called *Praat* (designed by Paul Boersma and David Weenink from the University of Amsterdam) to identify the prosodic cues proposed by different studies in the field and cues presented in the following studies.

Related Work

Tapperman et al. (2006) present some experiments on sarcasm recognition using prosodic, spectral, and contextual cues. They tend to train an automatic sarcasm recognizer using all the cues proposed. They claim that prosody alone is not sufficient to discern whether an utterance is sarcastic. In the sense of prosodic features, they used 19 different cues. The data was 131 uninterrupted occurrences of the phrase "yeah right" found in the Switchboard and Fisher Corpora. In experiment one, they used just the prosodic features to recognize sarcasm which resulted in an accuracy of 0.69, and this was considered very low for the recognizer but still better than human annotators listening without context. In the second and third experiments, was adapted in Tepperman's and Cheang's studies. It can be used to differentiate the semantic meaning of the expression Yeah right. Other studies in the field can be used as supporting material for our project since they have used the computer program *Praat* to analyze the quality and prosodic content of different expressions and also analyze other acoustic features from those proposed by Tepperman.

In this sense, Tepperman proposed 19 prosodic features to characterize the quality of the words *Yeah* and *right* in order to understand their meaning in the utterance. In our study, we are just parameterizing the tone of voice in terms of pitch, intensity, energy and duration to analyze if these features can give enough information to determine the orientation of the expression. Our project for detecting sarcasm was designed following the prosodic important to our project, since our goal is to confirm whether this affirmation is viable or not in all cases. We use the prosodic features Tepperman et al. describe as more significant in detecting sarcasm when analyzing our data with *Praat*. As they stated in their paper, even when prosodic cues did not perform the best accuracy, they can be more accurate than a human annotator without context, and that's exactly what we try to figure out in our project.

Voyer and Techentin (2010) focused on a specific component of prosody, that is, the subjective auditory cues conveying sarcasm to determine which prosodic cues are more relevant in the detection of sarcasm. They claim that sarcasm is perceived through the integration of multiple subjective auditory features, similar to the interpretation of any other type they used spectral and contextual cues respectively, with 0.77 and 0.84 of accuracy separately. But in experiment four, they used both at the same time and reached 0.87 accuracy which exceeded the inter-human agreement reached by the human annotators. Adding contextual, spectral and prosodic cues, in experiment five, the accuracy reached 0.86, what demonstrated their claim that prosodic cues are not necessary when paired with contextual or spectral cues. The prosodic cues that contributed the most in the classifier were the rising pitch frame in “yeah” and “right” separately and the energy over right. The authors concluded with the idea that a sarcasm detector can ignore prosody and focus on contextual and spectral features.

The aforementioned study is very

terms of sarcasm perception.

The results obtained in this study shed some light on the detection of sarcasm. The prosodic cues proposed for this research can be used in the analyses with a computer program. Besides, their interpretation of prosody is a critical aspect in the understanding of sarcasm, even though no contextual cues are present. It supports the goal of our project and contradicts what is proposed by Tepperman. The authors went further in their statement providing that prosodic cues alone can resolve ambiguity in any statement.

Another important research was carried out by Cheang and Pell (2008). In their study, they studied four different attitudes in English utterances as: sarcasm, humor, sincerity, and neutrality. They focused on analyzing sarcasm through the following prosodic of prosody. For this, they carried out two critical data analyses of 12 sentences obtained from TV comedies, web pages, and every day conversations. The subjects were 151 undergraduate students. In the first one, ratings were compared across tone of voice to identify specific subjectivity aspects of prosody that might count for sarcasm perception. The second one, they analyzed the data basing on the subjective perceptual rate of the subjects involved in the study in order to determine whether specific groupings of dimensions or tones of voice would emerge. The results suggested that sarcastic speech can be characterized by having less pitch and intensity variations, less resonance, and less clarity than sincere speech. As conclusion, the authors believed that their study supports the role of these factors in the interpretation of prosody in features: fundamental frequency (F0), F0 range, mean amplitude, amplitude range, speech rate, harmonic to noise ratio (HNR), and spectral values. The data was 96 English non-spontaneous utterances recorded by six native speakers of English; 24 representing each attitude. The subjects were trained with a definition of each attitude, and the data obtained was analyzed using *Praat* speech analyzer also. Each utterance was analyzed and measured in terms of the prosodic features exposed before. For this study, a reduction in mean F0 was the most consistent feature observed correlated with sarcasm; amplitude was not important for sarcasm recognition in this data since it did not show any predictable pattern. The study concluded suggesting that the distinct pattern associated with sarcasm in speech is the reduction of F0 and HNR and decreased F0 standard deviation. For our project, the measure of F0 as prosodic cue in the recognition of sarcasm is part of the acoustic features that can help to detect sarcasm because of the significance the authors concluded in the detection of sarcasm. Based on the findings of this research, we use a similar procedure to collect our data.

Finally, Gonzalez-Ibañez et al. (2011 reported on an empirical study on the use of lexical and pragmatic factors to distinguish sarcasm from positive to negative sentiment expressed in Twitter messages. The main purpose of the study is to build a corpus that includes only sarcastic utterances that have been identifying the composer of the message, and to report on the difficulty of distinguishing sarcastic tweets from tweets that are straightforwardly report the difficulties of detecting sarcasm. This research used different factors to study sarcasm, and still the performance was not very high. Of course, in writing, prosodic features cannot be used; nevertheless, it shows the complexity of determining and classifying this kind of utterances.

Method

Our present study is a combined replica of the previous work conducted by Tapperman et al. (2006) and Henry S. Cheang & Marc D. Pell (2008). We extracted the most significant features from both articles and undertook our empirical study. Our only concern is the prosodic cues in the spoken language as they mostly, according to the previous research done, reflect the speaker's sentiment. We didn't search for any spectral or contextual cues as the researchers above did. The other features which positive or negative. The data were 900 tweets in each of the categories: sarcastic, positive and negative, analyzed in terms of lexical and pragmatic features. In order to reflect on the difficulty concerning the task they conducted experiments with human judges, where the subjects had to identify 10% of the total data into the three sentiment categories. As a result, the agreements between the subject`s judgment was 50%. They trained the rest of the data with their classifier obtaining 57.6% accuracy. They found that the pragmatic features were the most useful in classifying feelings and the low performance of the humans suggests that gold standards built by using labels given by human coders other than human authors are not reliable. In this sense, even when our project tries to detect sarcasm in speech, we study sentences out of context. Also, we are correlated with other majors are out of the scope of our study.

In this study, we randomly chose nine English native speakers (5 males and 4 females) because the native speakers are completely aware of how they use their language. Their age ranges (16–45 years). They are mostly high school to graduate students. They were asked to produce 'Yeah right!' on two different occasions; neutrally and sarcastically according to their personal attitude toward the statements posed by the interviewer. Their responses were either neutral or sarcastic in order to express negative verbal irony toward the biased statements. Responses were recorded as naturally as possible as a reflection to two kinds of triggers: neutral and sarcastic.

The objective was to figure out how the acoustic/prosodic cues in his/her voice are to be modulated so as to express sarcasm. The 'yeah right' utterance was recorded twice for each speaker. Recording all the times occurred with the elicitation of neutral utterance, then the sarcastic one. Before being recorded, speakers were informed that the recording aims at eliciting neutral and sarcastic same utterance. To familiarize speakers with the recording procedure, a few similar trials were completed with each speaker before starting with actual recordings. We neither led speakers to produce 'proper' rendition of the required attitude nor did we coach them.

All utterances were acoustically analyzed by using Praat speech analyses software (Boersma and Weenink, 2007). A number of acoustic parameter was selected for the purpose of figuring out acoustic

b. Sarcastic utterance of yeah right

One can clearly see the difference in the two spectrographs; however, the closer look into the cues and reading the nuances will be discussed in the results and discussion part.

Prosodic Features

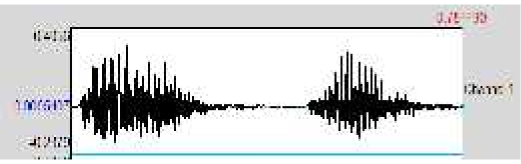
We picked the following features among the other features which could be found through the analysis of Praat. They are: the standard deviation, pitch range for the whole utterance, the number of rising and falling frames, duration of each word, average energy in each word, the number of inter-frames, the pitch range of each word, and other parameters. Though the following are of great relevance to our research.

Results and Discussion

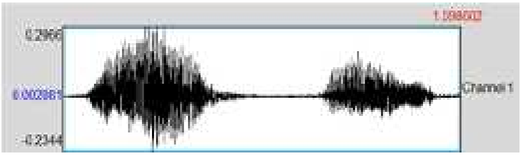
Based on the findings of previous research and our present study the relevant cues would be listed as follows: Mean F0 cues of sarcasm: mean F0, mean pitch, maximum pitch, minimum and maximum intensity, energy, and duration.

When we elicited the intended utterances, it was very difficult to know the reference of each utterance without referring to the tag we attached to it. In other words, the results rendered by *Praat* were very sophisticated and very detailed to the extent that it exceeds the ability of many of us in determining the exact orientation of the separate utterances. Let's look at the results and try to extract a pattern related to sarcasm.

Spectrographs



a. Neutral utterance of yeah right



Subjects	Neutral utterance	Sarcastic utterance
1	25.8535 Hz	113.2177 Hz
2	116.3896	104.7436
3	**75.44412**	**76.14486**
4	**78.26759**	**133.6111**
5	**77.47681**	**80.0063**
6	135.9757	120.4343
7	**101.0981**	**110.0847**
8	**176.0454**	**188.2164**
9	189.7460	140.9667

The F0 stands for the fundamental frequency in a series of sounds; i.e., the lowest frequency of the utterance. The results collected from the analyzed data were inconsistent and peoples' responses varied a lot between lower and higher F0. This inconsistency stands against formulating a hypothesis or a pattern.

Mean Pitch

Subjects	Neutral utterance	Sarcastic utterance
1	47.6513 Hz	24.3043 Hz
2	171.6849	129.5256
3	186.3183	171.3530
4	175.5071	171.2168
5	92.59608	75.0438
6	174.2772	147.6343
7	115.2076	101.6611
8	190.8087	178.0649
9	164.0588	152.8801

The pitch is the relative highness or lowness of a sound. The mean pitch results were very encouraging to us. The sarcastic utterances tend to be produced by lower mean pitch than that of the neutral utterances. From the table above a pattern can be seen and a hypothesis is due. The first acoustic cue relevant to sarcasm and found to be consistent is the 'mean pitch'. Sarcasm can be all the elicited responses. Minimum Intensity

Sub jects	Neutral utterance	Sarcastic utterance
1	35.6803 dB	53.0096 dB
2	45.4068	46.1688
3	33.5902	34.1823
4	**37.8039**	**34.6945**
5	**42.3495**	**33.6532**
6	28.6744	33.7506
7	37.5477	39.1481
8	20.2187	23.0203
9	**37.2807**	**36.9922**

Which is the energy carried by sound waves. The minimum intensity recorded about 70% consistency with our data. The majority of utterances recorded higher minimum intensity than the neutral utterances. This relative consistency does not formulate a pattern.

characterized by lower mean pitch.

Maximum Pitch

Subj ects	Neutral utterance	Sarcastic utterance
1	428.2789 Hz	31.7786 Hz
2	482.8828	399.5204
3	**505.2518**	**507.2908**
4	288.2602	234.9009
5	**115.9350**	**165.5293**
6	328.3972	313.7723
7	**131.1031**	**134.2509**
8	**203.7436**	**230.7824**
9	**189.7460**	**245.8205**

The maximum pitch is relatively inconsistent. The sarcastic responses recorded lower and higher maximum pitch; however, these findings cannot be considered a pattern that might be generalized to when looking for acoustic features correlated with sarcasm.

Energy

Sub jects	Neutral utterance	Sarcastic utterance
1	0.00876 Pascal^2 sec	0.00630 Pascal^2 sec
2	0.00659	0.00209
3	0.00077	0.00076
4	0.00175	0.00148
5	0.00805	0.00100
6	**0.00223**	**0.00273**
7	0.00757	0.00502
8	0.00094	0.00075
9	**0.00124**	**0.00126**

The second prosodic cue relevant to the sarcastic utterance is the energy. We found about 82% consistency of

Maximum Intensity

Sub jects	Neutral utterance	Sarcastic utterance
1	75.0563 dB	77.9639 dB
2	**77.4803**	**74.6441**
3	**69.8441**	**69.8300**
4	74.4733	74.8278
5	78.9558	81.2200
6	76.8450	77.1911
7	**81.5786**	**79.5343**
8	**70.4197**	**69.9812**
9	**72.3498**	**71.8733**

The results of maximum intensity seem to vary a little between the neutral and sarcastic utterances. They are totally inconsistent and cannot be taken into consideration findings that sarcastic utterances are mostly uttered with lower energy.

Duration

Subjects	Neutral utterance	Sarcastic utterance
1	0.585214 seconds	0.726897 seconds
2	0.599112	0.623825
3	0.861405	0.889546
4	0.603980	0.675981
5	0.798925	1.056320
6	0.635515	0.861359
7	0.947386	0.962597
8	0.654653	0.796255
9	0.855235	1.192131

The third completely reliable prosodic feature of sarcasm is the time duration of the sarcastic utterance. Our data recorded 100% consistency that the sarcastic utterance relatively takes longer time to be produced in comparison to the neutral utterance. The table above shows the time duration it takes to produce 'yeah right'. Hence, a pattern can be concluded that the sarcastic utterance takes longer duration than the neutral utterance.

Conclusion

In this project we have examined different acoustic/prosodic cues to detect sarcasm, all adopted from previous studies. We used *Praat* to analyze our data which is a very useful and sophisticated tool in the study of 7 acoustic and prosodic features which is the same software used by many other previous studies concerning the study of acoustic features of humans speech. Our main goal was to determine if sarcasm can be detected through the analysis of prosodic cues or acoustic features. Our results showed that, from all the cues covered in our study, mean pitch, energy and duration can characterize sarcasm in the sarcastic utterances. The rest of features did not present a predictable difference compared to the neutral utterances.

Based on our data and the findings that the present study yielded, we can conclude that sarcastic expressions present lower energy than neutral expressions. They require lower effort from the speaker to produce. And also they take relatively longer duration. Nevertheless, detection of sarcasm is a difficult task for human detection and also for the studies of prosody. Even though we were able to find a pattern in the prosodic features, an in-depth study should include other cues to detect sarcasm in order to achieve higher accuracy.

The scope of our research did not cover the cognitive or neurological factors that influence the production of sarcasm, or even the psychological mechanisms that lead to producing sarcastic utterances with lower energy or longer duration. What exactly makes speakers think in a specific way, or reflect their own attitudes by manipulating their sounds? What we are sure of is that speakers tend to modulate their sounds, unconsciously, to present sarcasm even though they are not aware of that modulation.

In terms of the data, it was not obtained as excerpts from natural conversations. The circumstances were not qualitatively standard as we didn't record in sound-proof booths. There might be background noise. We also tried our best to obtain natural-like expressions, but our data were not void of exaggerated expressions. Moreover, the number of our subjects was not large enough to support strongly our claim that prosodic cues could be indicators of sarcasm. Further work is needed to prove if in natural conversations other patterns can be found.

References :

- Boersma, P., Weenik, D. (2007). Doing phonetics by computer (version 5.3.11) [computer program] (Retrieved 31.03.2012) http://www.praat.org/
- Cheang, H.S., Pell, M.D., (2008). The sound of sarcasm. *Speech communication 50*, 366-381
- González-Ibáñez, R., Muresan, S., and Wacholder, N. (2011). Identifying sarcasm in Twitter: a closer look. in proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011). Portland, Oregon: June 19-24, 2011. *shortpapers*, 581–586.
- Reyes, A., Rosso, P., & Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering, 74*, 1-12.
- Tapperman, Traum, Nayaranan (2006) “yeah” right: sarcasm recognition for spoken dialogues systems, Conference on spoken. Iska-speech.org
- Voyer, D., & Techentin, C. (2010). Subjective Acoustic Features of Sarcasm. Metaphor and Symbol, 25, 227-242.
- Voyer, D., Bowes, A., & Techentin, C. (2008). On the perception of sarcasm in dichotic listening. Neuropsychology, 22, 390-399.