

Humor: Prosody Analysis and Automatic Recognition for F * R * I * E * N * D * S *

Amruta Purandare and Diane Litman

Intelligent Systems Program

University of Pittsburgh

{amruta,litman}@cs.pitt.edu

Abstract

We analyze humorous spoken conversations from a classic comedy television show, FRIENDS, by examining acoustic-prosodic and linguistic features and their utility in automatic humor recognition. Using a simple annotation scheme, we automatically label speaker turns in our corpus that are followed by *laughs* as humorous and the rest as non-humorous. Our humor-prosody analysis reveals significant differences in prosodic characteristics (such as pitch, tempo, energy etc.) of humorous and non-humorous speech, even when accounted for the gender and speaker differences. Humor recognition was carried out using standard supervised learning classifiers, and shows promising results significantly above the baseline.

1 Introduction

As conversational systems are becoming prevalent in our lives, we notice an increasing need for adding social intelligence in computers. There has been a considerable amount of research on incorporating affect (Litman and Forbes-Riley, 2004) (Alm et al., 2005) (D'Mello et al., 2005) (Shroder and Cowie, 2005) (Klein et al., 2002) and personality (Gebhard et al., 2004) in computer interfaces, so that, for instance, user frustrations can be recognized and addressed in a graceful manner. As (Binsted, 1995) correctly pointed out, one way to alleviate user frustrations, and to make human-computer interaction more natural, personal and interesting for the users, is to model HUMOR.

Research in computational humor is still in very early stages, partially because humorous lan-

guage often uses complex, ambiguous and incongruous syntactic and semantic expressions (Attardo, 1994) (Mulder and Nijholt, 2002) which require deep semantic interpretation. Nonetheless, recent studies have shown a feasibility of automatically recognizing (Mihalcea and Strapparava, 2005) (Taylor and Mazlack, 2004) and generating (Binsted and Ritchie, 1997) (Stock and Strapparava, 2005) humor in computer systems. The state of the art research in computational humor (Binsted et al., 2006) is, however, limited to text (such as humorous *one-liners*, *acronyms* or *wordplays*), and to our knowledge, there has been no work to date on automatic humor recognition in spoken conversations.

Before we can model humor in real application systems, we must first analyze features that characterize humor. Computational approaches to humor recognition so far primarily rely on lexical and stylistic cues such as alliteration, antonyms, adult slang (Mihalcea and Strapparava, 2005). The focus of our study is, on the other hand, on analyzing acoustic-prosodic cues (such as pitch, intensity, tempo etc.) in humorous conversations and testing if these cues can help us to automatically distinguish between humorous and non-humorous (normal) utterances in speech. We hypothesize that not only the lexical content but also the prosody (or how the content is expressed) makes humorous expressions *humorous*.

The following sections describe our data collection and pre-processing, followed by the discussion of various acoustic-prosodic as well as other types of features used in our humorous-speech analysis and classification experiments. We then present our experiments, results, and finally end with conclusions and future work.

2 FRIENDS Corpus

(Scherer, 2003) discuss a number of pros and cons of using *real* versus *acted* data, in the context of emotional speech analysis. His main argument is that while real data offers natural expressions of emotions, it is not only hard to collect (due to ethical issues) but also very challenging to annotate and analyze, as there are very few instances of strong expressions and the rest are often very subtle. Acted data (also referred to as *portrayed* or *simulated*), on the other hand, offers ample of prototypical examples, although these are criticized for not being *natural* at times. To achieve some balance between naturalness and strength/number of humorous expressions, we decided to use dialogs from a comedy television show FRIENDS, which provides classical examples of casual, humorous conversations between friends who often discuss very real-life issues, such as job, career, relationships etc.

We collected a total of 75 dialogs (scenes) from six episodes of FRIENDS, four from Season I (Monica Gets a New Roommate, The One with Two Parts: Part 1 and 2, All the Poker) and two from Season II (Ross Finds Out, The Prom Video), all available on *The Best of Friends Volume I* DVD. This gave us approximately 2 hrs of audio. Text transcripts of these episodes were obtained from: <http://www.friendscafe.org/scripts.shtml>, and were used to extract lexical features (used later in classification).

Figure 1 shows an excerpt from one of the dialogs in our corpus.

3 Audio Segmentation and Annotation

We segmented each audio file (manually) by marking speaker turn boundaries, using Wavesurfer (<http://www.speech.kth.se/wavesurfer>). We apply a fairly straightforward annotation scheme to automatically identify humorous and non-humorous turns in our corpus. Speaker turns that are followed by artificial laughs are labeled as *Humorous*, and all the rest as *Non-Humorous*. For example, in the dialog excerpt shown in figure 1, turns 3, 7, 9, 11 and 16 are marked as *humorous*, whereas turns 1, 2, 5, 6, 13, 14, 15 are marked as *non-humorous*. Artificial laughs, silences longer than 1 second and segments of audio that contain purely non-verbal sounds (such as phone rings, door bells, music etc.) were excluded from the analysis. By considering only

-
- [1] Rachel: Guess what?
 - [2] Ross: You got a job?
 - [3] Rachel: Are you kidding? I am trained for nothing!
 - [4] <Laughter>
 - [5] Rachel: I was laughed out of twelve interviews today.
 - [6] Chandler: And yet you're surprisingly upbeat.
 - [7] Rachel: You would be too if you found John and David boots on sale, fifty percent off!
 - [8] <Laughter>
 - [9] Chandler: Oh, how well you know me...
 - [10] <Laughter>
 - [11] Rachel: They are my new, I don't need a job, I don't need my parents, I got great boots, boots!
 - [12] <Laughter>
 - [13] Monica: How'd you pay for them?
 - [14] Rachel: Uh, credit card.
 - [15] Monica: And who pays for that?
 - [16] Rachel: Um... my... father.
 - [17] <Laughter>
-

Figure 1: Dialog Excerpt

speaker turns that are followed by laughs as humorous, we also automatically eliminate cases of pure visual comedy where humor is expressed using only gestures or facial expressions. In short, non-verbal sounds or silences followed by laughs are not treated as humorous. Henceforth, by *turn*, we mean proper speaker turns (and not non-verbal turns). We currently do not apply any special filters to remove non-verbal sounds or background noise (other than laughs) that overlap with speaker turns. However, if artificial laughs overlap with a speaker turn (there were only few such instances), the speaker turn is chopped by marking a turn boundary exactly before/after the laughs begin/end. This is to ensure that our prosody analysis is fair and does not catch any cues from the laughs. In other words, we make sure that our speaker turns are *clean* and not garbled by laughs.

After segmentation, we got a total of 1629 speaker turns, of which 714 (43.8%) are humorous, and 915 (56.2%) are non-humorous. We also made sure that there is a 1-to-1 correspondence between speaker turns in text transcripts that were obtained online and our audio segments, and corrected few cases where there was a mis-match (due to turn-chopping or errors in online transcripts).

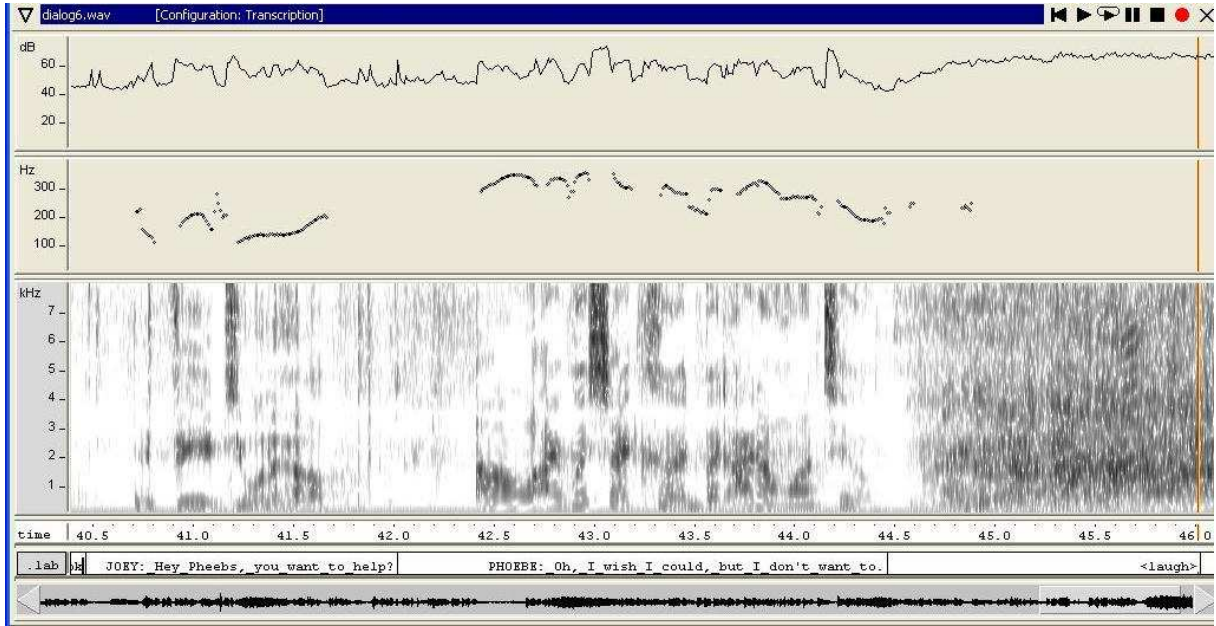


Figure 2: Audio Segmentation, Transcription and Feature Extraction using Wavesurfer

4 Speaker Distributions

There are 6 *main* actors/speakers (3 male and 3 female) in this show, along with a number of (in our data 26) guest actors who appear briefly and rarely in some of our dialogs. As the number of guest actors is quite large, and their individual contribution is less than 5% of the turns in our data, we decided to group all the guest actors together in one *GUEST* class.

As these are *acted* (not real) conversations, there were only few instances of speaker turn-overlaps, where multiple speakers speak together. These turns were given a speaker label *MULTI*. Table 1 shows the total number of turns and humorous turns for each speaker, along with their percentages in braces. Percentages for the Humor column show, out of the total (714) humorous turns, how many are by each speaker. As one can notice, the distribution of turns is fairly balanced among the six main speakers. We also notice that even though each guest actors' individual contribution is less than 5% in our data, their combined contribution is fairly large, almost 16% of the total turns.

Table 2 shows that the six main actors together form a total of 83% of our data. Also, of the total 714 humorous turns, 615 (86%) turns are by the main actors. To study if prosody of humor differs across males and females, we also grouped the main actors into two gender classes. Table 2 shows that the gender distribution is fairly bal-

Speaker	#Turns(%)	#Humor (%)
Chandler (M)	244 (15)	163 (22.8)
Joey (M)	153 (9.4)	57 (8)
Monica (F)	219 (13.4)	74 (10.4)
Phoebe (F)	180 (11.1)	104 (14.6)
Rachel (F)	273 (16.8)	90 (12.6)
Ross (M)	288 (17.7)	127 (17.8)
GUEST (26)	263 (16.1)	95 (13.3)
MULTI	9 (0.6)	4 (0.6)

Table 1: Speaker Distribution

anced among the main actors, with 50.5% male and 49.5% female turns. We also see that of the 685 male turns, 347 turns (almost 50%) are humorous, and of the 672 female turns, 268 (approximately 40%) are humorous. Guest actors and multi-speaker turns are not considered in the gender analysis.

Speaker	#Turns	#Humor
Male	685 (50.5% of Main)	347 (50.6% of Male)
Female	672 (49.5% Of Main)	268 (39.9% of Female)
Total Main	1357 (83.3% of Total)	615 (86.1% of Humor)

Table 2: Gender Distribution for Main Actors

5 Features

Literature in emotional speech analysis (Liscombe et al., 2003)(Litman and Forbes-Riley, 2004) (Scherer, 2003)(Ang et al., 2002) has shown that prosodic features such as pitch, energy, speaking rate (tempo) are useful indicators of emotional states, such as joy, anger, fear, boredom etc. While humor is not necessarily considered as an emotional state, we noticed that most humorous utterances in our corpus (and also in general) often make use of hyper-articulations, similar to those found in emotional speech.

For this study, we use a number of acoustic-prosodic as well as some non acoustic-prosodic features as listed below:

Acoustic-Prosodic Features:

- Pitch (F0): Mean, Max, Min, Range, Standard Deviation
- Energy (RMS): Mean, Max, Min, Range, Standard Deviation
- Temporal: Duration, Internal Silence, Tempo

Non Acoustic-Prosodic Features:

- Lexical
- Turn Length (#Words)
- Speaker

Our acoustic-prosodic features make use of the pitch, energy and temporal information in the speech signal, and are computed using Wavesurfer. Figure 2 shows Wavesurfer’s energy (dB), pitch (Hz), and transcription (.lab) panes. The transcription interface shows text corresponding to the dialog turns, along with the turn boundaries. All features are computed at the turn level, and essentially measure the mean, maximum, minimum, range (maximum-minimum) and standard deviation of the feature value (F0 or RMS) over the entire turn (ignoring zeroes). Duration is measured in terms of time in seconds, from the beginning to the end of the turn including pauses (if any) in between. Internal silence is measured as the percentage of zero F0 frames, and essentially account for the amount of silence in the turn. Tempo is computed as the total number of syllables divided by the duration of the turn. For computing the number of syllables per word, we used the General Inquirer database (Stone et al., 1966).

Our lexical features are simply all words (alphabetic strings including apostrophes and stopwords) in the turn. The value of these features is integral and essentially counts the number of times a word is repeated in the turn. Although this indirectly accounts for alliterations, in the future studies, we plan to use more *stylistic* lexical features like (Mihalcea and Strapparava, 2005).

Turn length is measured as the number of words in the turn. For our classification study, we consider eight speaker classes (6 Main actors, 1 for Guest and Multi) as shown in table 1, whereas for the gender study, we consider only two speaker categories (male and female) as shown in table 2.

6 Humor-Prosody Analysis

Feature	Humor	Non-Humor
Mean-F0	206.9	208.9
Max-F0*	299.8	293.5
Min-F0*	121.1	128.6
Range-F0*	178.7	164.9
StdDev-F0	41.5	41.1
Mean-RMS*	58.3	57.2
Max-RMS*	76.4	75
Min-RMS*	44.2	44.6
Range-RMS*	32.16	30.4
StdDev-RMS*	7.8	7.5
Duration*	3.18	2.66
Int-Sil*	0.452	0.503
Tempo*	3.21	3.03
Length*	10.28	7.97

Table 3: Humor Prosody: Mean feature values for Humor and Non-Humor groups

Table 3 shows mean values of various acoustic-prosodic features over all speaker turns in our data, across humor and non-humor groups. Features that have statistically ($p \leq 0.05$ as per independent samples t-test) different values across the two groups are marked with asterisks. As one can see, all features except Mean-F0 and StdDev-F0 show significant differences across humorous and non-humorous speech. Table 3 shows that humorous turns in our data are longer, both in terms of the time duration and the number of words, than non-humorous turns. We also notice that humorous turns have smaller internal silence, and hence rapid tempo. Pitch (F0) and energy (RMS) features have higher maximum, but lower minimum

values, for humorous turns. This in turn gives higher values for range and standard deviation for humor compared to the non-humor group. This result is somewhat consistent with previous findings of (Liscombe et al., 2003) who found that most of these features are largely associated with positive and active emotional states such as happy, encouraging, confident etc. which are likely to appear in our humorous turns.

7 Gender Effect on Humor-Prosody

To analyze prosody of humor across two genders, we conducted a 2-way ANOVA test, using speaker gender (male/female) and humor (yes/no) as our fixed factors, and each of the above acoustic-prosodic features as a dependent variable. The test tells us the effect of humor on prosody adjusted for gender, the effect of gender on prosody adjusted for humor and also the effect of interaction between gender and humor on prosody (i.e. if the effect of humor on prosody differs according to gender). Table 4 shows results of 2-way ANOVA, where Y shows significant effects, and N shows non-significant effects. For example, the result for tempo shows that tempo differs significantly only across humor and non-humor groups, but not across the two gender groups, and that there is no effect of interaction between humor and gender on tempo. As before, all features except Mean-F0 and StdDev-F0 show significant differences across humor and no-humor conditions, even when adjusted for gender differences. The table also shows that all features except internal silence and tempo show significant differences across two genders, although only pitch features (Max-F0, Min-F0, and StdDev-F0) show the effect of interaction between gender and humor. In other words, the effect of humor on these pitch features is dependent on gender. For instance, if male speakers raise their pitch while expressing humor, female speakers might lower. To confirm this, we computed means values of various features for males and females separately (See Tables 5 and 6). These tables indeed suggest that male speakers show higher values for pitch features (Mean-F0, Min-F0, StdDev-F0), while expressing humor, whereas females show lower. Also for male speakers, differences in Min-F0 and Min-RMS values are not statistically significant across humor and non-humor groups, whereas for female speakers, features Mean-F0, StdDev-F0 and tempo do not

show significant differences across the two groups. One can also notice that the differences in the mean pitch feature values (specifically Mean-F0, Max-F0 and Range-F0) between humor and non-humor groups are much higher for males than for females.

In summary, our gender analysis shows that although most acoustic-prosodic features are different for males and females, the prosodic style of expressing humor by male and female speakers differs only along some pitch-features (both in magnitude and direction).

Feature	Humor	Gender	Humor x Gender
Mean-F0	N	Y	N
Max-F0	Y	Y	Y
Min-F0	Y	Y	Y
Range-F0	Y	Y	N
StdDev-F0	N	Y	Y
Mean-RMS	Y	Y	N
Max-RMS	Y	Y	N
Min-RMS	Y	Y	N
Range-RMS	Y	Y	N
StdDev-RMS	Y	Y	N
Duration	Y	Y	N
Int-Sil	Y	N	N
Tempo	Y	N	N
Length	Y	Y	N

Table 4: Gender Effect on Humor Prosody: 2-Way ANOVA Results

8 Speaker Effect on Humor-Prosody

We then conducted similar ANOVA test to account for the speaker differences, i.e. by considering humor (yes/no) and speaker (8 groups as shown in table 1) as our fixed factors and each of the acoustic-prosodic features as a dependent variable for a 2-Way ANOVA. Table 7 shows results of this analysis. As before, the table shows the effect of humor adjusted for speaker, the effect of speaker adjusted for humor and also the effect of interaction between humor and speaker, on each of the acoustic-prosodic features. According to table 7, we no longer see the effect of humor on features Min-F0, Mean-RMS and Tempo (in addition to Mean-F0 and StdDev-F0), in presence of the speaker variable. Speaker, on the other hand, shows significant effect on prosody for all features. But

Feature	Humor	Non-Humor
Mean-F0*	188.14	176.43
Max-F0*	276.94	251.7
Min-F0	114.54	113.56
Range-F0*	162.4	138.14
StdDev-F0*	37.83	34.27
Mean-RMS*	57.86	56.4
Max-RMS*	75.5	74.21
Min-RMS	44.04	44.12
Range-RMS*	31.46	30.09
StdDev-RMS*	7.64	7.31
Duration*	3.1	2.57
Int-Sil*	0.44	0.5
Tempo*	3.33	3.1
Length*	10.27	8.1

Table 5: Humor Prosody for Male Speakers

Feature	Humor	Non-Humor
Mean-F0	235.79	238.75
Max-F0*	336.15	331.14
Min-F0*	133.63	143.14
Range-F0*	202.5	188
StdDev-F0	46.33	46.6
Mean-RMS*	58.44	57.64
Max-RMS*	77.33	75.57
Min-RMS*	44.08	44.74
Range-RMS*	33.24	30.83
StdDev-RMS*	8.18	7.59
Duration*	3.35	2.8
Int-Sil*	0.47	0.51
Tempo	3.1	3.1
Length*	10.66	8.25

Table 6: Humor Prosody for Female Speakers

surprisingly, again only pitch features Mean-F0, Max-F0 and Min-F0 show the interaction effect, suggesting that the effect of humor on these pitch features differs from speaker to speaker. In other words, different speakers use different pitch variations while expressing humor.

9 Humor Recognition by Supervised Learning

We formulate our humor-recognition experiment as a classical supervised learning problem, by automatically classifying spoken turns into humor and non-humor groups, using standard machine learning classifiers. We used the decision tree algorithm ADTree from Weka, and ran a 10-fold cross validation experiment on all 1629 turns in our data¹. The baseline for these experiments is 56.2% for the majority class (non-humorous). Table 8 reports classification results for six feature categories: lexical alone, lexical + speaker, prosody alone, prosody + speaker, lexical + prosody and lexical + prosody + speaker (all). Numbers in braces show the number of features in each category. There are total 2025 features which include 2011 lexical (all word types plus turn length), 13 acoustic-prosodic and 1 for the speaker information. Feature *Length* was included in the lexical feature group, as it counts the number of lexical items (words) in the turn.

¹We also tried other classifiers like Naive Bayes and AdaBoost, although since the results were equivalent to ADTree, we do not report those here.

All results are significantly above the baseline (as measured by a pair-wise t-test) with the best accuracy of 64% (8% over the baseline) obtained using all features. We notice that the classification accuracy improves on adding speaker information to both lexical and prosodic features. Although these results do not show a strong evidence that prosodic features are better than lexical, it is interesting to note that the performance of just a few (13) prosodic features is comparable to that of 2011 lexical features. Figure 3 shows the decision tree produced by the classifier in 10 iterations. Numbers indicate the order in which the nodes are created, and indentations mark parent-child relations. We notice that the classifier primarily selected speaker and prosodic features in the first 10 iterations, whereas lexical features were selected only in the later iterations (not shown here). This seems consistent with our original hypothesis that speech features are better at discriminating between humorous and non-humorous utterances in speech than lexical content.

Although (Mihalcea and Strapparava, 2005) obtained much higher accuracies using lexical features alone, it might be due to the fact that our data is *homogeneous* in the sense that both humorous and non-humorous turns are extracted from the same source, and involve same speakers, which makes the two groups highly alike and hence challenging to distinguish. To make sure that the lower accuracy we get is not simply due to using smaller data compared to (Mihalcea and Strappar-

Feature	Humor	Speaker	Humor x Speaker
Mean-F0	N	Y	Y
Max-F0	Y	Y	Y
Min-F0	N	Y	Y
Range-F0	Y	Y	N
StdDev-F0	N	Y	N
Mean-RMS	N	Y	N
Max-RMS	Y	Y	N
Min-RMS	Y	Y	N
Range-RMS	Y	Y	N
StdDev-RMS	Y	Y	N
Duration	Y	Y	N
Int-Sil	Y	Y	N
Tempo	N	Y	N
Length	Y	Y	N

Table 7: Speaker Effect on Humor Prosody: 2-Way ANOVA Results

Feature	-Speaker	+Speaker
Lex	61.14 (2011)	63.5 (2012)
Prosody	60 (13)	63.8 (14)
Lex + Prosody	62.6 (2024)	64 (2025)

Table 8: Humor Recognition Results (% Correct)

ava, 2005), we looked at the learning curve for the classifier (see figure 4) and found that the classifier performance is not sensitive to the amount of data.

Table 9 shows classification results by gender, using all features. For the male group, the baseline is 50.6%, as the majority class *humor* is 50.6% (See Table 2). For females, the baseline is 60% (for non-humorous) as only 40% of the female turns are humorous.

Gender	Baseline	Classifier
Male	50.6	64.63
Female	60.1	64.8

Table 9: Humor Recognition Results by Gender

As Table 9 shows, the performance of the classifier is somewhat consistent cross-gender, although for male speakers, the relative improvement is much higher (14% above the baseline), than for females (only 5% above the baseline). Our earlier observation (from tables 5 and 6) that differences in pitch features between humor and non-humor

```

(1)SPEAKER = chandler: 0.469
(1)SPEAKER != chandler: -0.083
|   (4)SPEAKER = phoebe: 0.373
|   (4)SPEAKER != phoebe: -0.064
(2)DURATION < 1.515: -0.262
|   (5)SILENCE < 0.659: 0.115
|   (5)SILENCE >= 0.659: -0.465
|   (8)SD_F0 < 9.919: -1.11
|   (8)SD_F0 >= 9.919: 0.039
(2)DURATION >= 1.515: 0.1
|   (3)MEAN_RMS < 56.117: -0.274
|   (3)MEAN_RMS >= 56.117: 0.147
|   |   (7)come < 0.5: -0.056
|   |   (7)come >= 0.5: 0.417
|   (6)SD_F0 < 57.333: 0.076
|   (6)SD_F0 >= 57.333: -0.285
|   (9)MAX_RMS < 86.186: 0.011
|   |   (10)MIN_F0 < 166.293: 0.047
|   |   (10)MIN_F0 >= 166.293: -0.351
|   (9)MAX_RMS >= 86.186: -0.972

```

Legend: +ve = humor, -ve = non-humor

Figure 3: Decision Tree (only the first 10 iterations are shown)

groups are quite higher for males than for females, may explain why we see higher improvement for male speakers.

10 Conclusions

In this paper, we presented our experiments on humor-prosody analysis and humor recognition in spoken conversations, collected from a classic television comedy, FRIENDS. Using a simple automated annotation scheme, we labeled speaker turns in our corpus that are followed by artificial laughs as humorous, and the rest as non-humorous. We then examined a number of acoustic-prosodic features based on pitch, energy and temporal information in the speech signal, that have been found useful by previous studies in emotion recognition.

Our prosody analysis revealed that humorous and non-humorous turns indeed show significant differences in most of these features, even when accounted for the speaker and gender differences. Specifically, we found that humorous turns tend to have higher tempo, smaller internal silence, and higher peak, range and standard deviation for pitch and energy, compared to non-humorous turns.

On the humor recognition task, our classifier

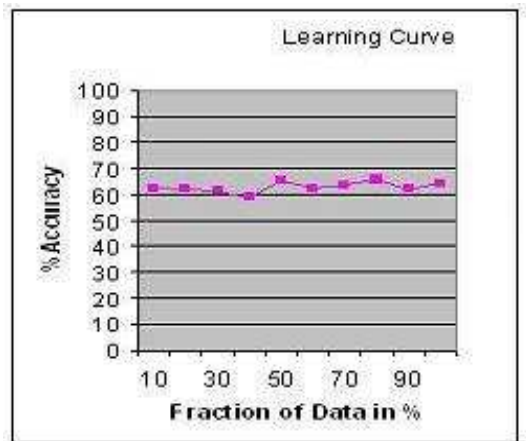


Figure 4: Learning Curve: %Accuracy versus %Fraction of Data

achieved the best performance when acoustic-prosodic features were used in conjunction with lexical and other types of features, and in all experiments attained the accuracy statistically significant over the baseline. While prosody of humor shows some differences due to gender, the performance on the humor recognition task is equivalent for males and females, although the relative improvement over the baseline is much higher for males than for females.

Our current study focuses only on lexical and speech features, primarily because these features can be computed automatically. In the future, we plan to explore more sophisticated semantic and pragmatic features such as incongruity, ambiguity, expectation-violation etc. We also like to investigate if our findings generalize to other types of corpora besides TV-show dialogs.

References

- C. Alm, D. Roth, and R. Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of HLT/EMNLP*, Vancouver, CA.
- J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke. 2002. Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In *Proceedings of ICSLP*.
- S. Attardo. 1994. *Linguistic Theory of Humor*. Mouton de Gruyter, Berlin.
- K. Binsted and G. Ritchie. 1997. Computational rules for punning riddles. *Humor*, 10(1).
- K. Binsted, B. Bergen, S. Coulson, A. Nijholt, O. Stock, C. Strapparava, G. Ritchie, R. Manurung, H. Pain, A. Waller, and D. O'Mara. 2006. Computational humor. *IEEE Intelligent Systems*, March-April.
- K. Binsted. 1995. Using humour to make natural language interfaces more friendly. In *Proceedings of the AI, ALife and Entertainment Workshop*, Montreal, CA.
- S. D'Mello, S. Craig, G. Gholson, S. Franklin, R. Picard, and A. Graesser. 2005. Integrating affect sensors in an intelligent tutoring system. In *Proceedings of Affective Interactions: The Computer in the Affective Loop Workshop*.
- P. Gebhard, M. Klesen, and T. Rist. 2004. Coloring multi-character conversations through the expression of emotions. In *Proceedings of Affective Dialog Systems*.
- J. Klein, Y. Moon, and R. Picard. 2002. This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14.
- J. Liscombe, J. Venditti, and J. Hirschberg. 2003. Classifying subject ratings of emotional speech using acoustic features. In *Proceedings of Eurospeech*, Geneva, Switzerland.
- D. Litman and K. Forbes-Riley. 2004. Predicting student emotions in computer-human tutoring dialogues. In *Proceedings of ACL*, Barcelona, Spain.
- R. Mihalcea and C. Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of HLT/EMNLP*, Vancouver, CA.
- M. Mulder and A. Nijholt. 2002. Humor research: State of the art. Technical Report 34, CTIT Technical Report Series.
- Scherer. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1-2):227-256.
- M. Shroder and R. Cowie. 2005. Toward emotion-sensitive multimodal interfaces: the challenge of the european network of excellence humaine. In *Proceedings of User Modeling Workshop on Adapting the Interaction Style to Affective Factors*.
- O. Stock and C. Strapparava. 2005. Hahaacronym: A computational humor system. In *Proceedings of ACL Interactive Poster and Demonstration Session*, pages 113-116, Ann Arbor, MI.
- P. Stone, D. Dunphy, M. Smith, and D. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press, Cambridge, MA.
- J. Taylor and L. Mazlack. 2004. Computationally recognizing wordplay in jokes. In *Proceedings of the CogSci 2004*, Chicago, IL.