Characteristics of authentic anger in Hebrew speech

Noam Amir, Shirley Ziv, Rachel Cohen

Department of Communication Disorders
Sackler Faculty of Medicine, Tel Aviv University, Israel
noama@post.tau.ac.il

Abstract

In this study we examine a number of characteristics of angry Hebrew speech. Whereas such studies are frequently carried out on acted speech, in this study we used recordings of participants in broadcasted, politically oriented talk shows. The recordings were audited and rated for anger content by 11 listeners. 12 utterances judged to contain angry speech were then analyzed along with 12 utterances from the same speakers that were judged to contain neutral speech. Various statistics of the F0 curve and spectral tilt were calculated and correlated with the degree of anger, giving a number of interesting results: for example, though pitch range was significantly correlated to anger in general, pitch range was significantly negative-correlated to the degree of anger. A separate test was conducted, judging only the textual content of the utterances, to examine the degree to which it influenced the listening tests. After neutralizing for the textual content, some of the acoustic measures became weaker predictors of anger, whereas mean F0 remained the strongest indicator of anger. Spectral tilt also showed a significant decrease in angry speech.

1. Introduction

Numerous studies have been carried out in recent years, which examine the manifestations of emotion in the speech signal. One of the main problems in this type of study is obtaining a corpus of emotional speech. One approach that has been adopted widely is to obtain acted emotions, from professional or non professional actors, as done by Banse and Scherer and others [1,2,3]. Using this type of data always raises the question whether the manner in which emotion is expressed by actors is the same as that in which it is expressed in spontaneous speech or in dialog. This point has been discussed in various papers [4,5,6], though no sweeping consensus has been arrived at.

In recent years more studies have appeared in which naturally occurring speech has been used, in various settings, such as television talk shows [4] or dialogs recorded during the performance of a given task [6]. It is no coincidence that together with the trend toward more natural settings, the definition of the emotions to be studied becomes more troublesome. Whereas actors can be requested to produce speech simulating such basic emotions such as anger, sadness, fear, etc., the emotions found in more naturally occurring speech can be more subtle and also more complex. For this very reason we find that these same studies propose more flexible means to classify emotions, such as the method proposed by Cowie [7], and different classifications than the basic emotions often found when using acted speech [6].

The objective of the current study was to bridge this gap to a certain extent, which is more easily performed for one of the more basic and widespread emotions, namely anger. Though it may prove difficult and even unethical to provoke anger in an experimental setting, it can often be found in political talk shows, which exist on a number of channels of Israeli television. Due to the complicated political situation in the Middle East, inhabitants of this region tend to have a high degree of political awareness. Politics receive a large amount of television broadcast time, and the participants tend to have a strong emotional involvement in the subjects being discussed.

In the present study, a number of utterances were recorded from such political talk shows. These were analyzed for pitch and voice quality on one hand, and judged by human listeners for the presence and degree of anger. The correlation between various feautures extracted from the raw data and the human judgement was examined in detail.

2. Data collection

2.1 Raw speech material

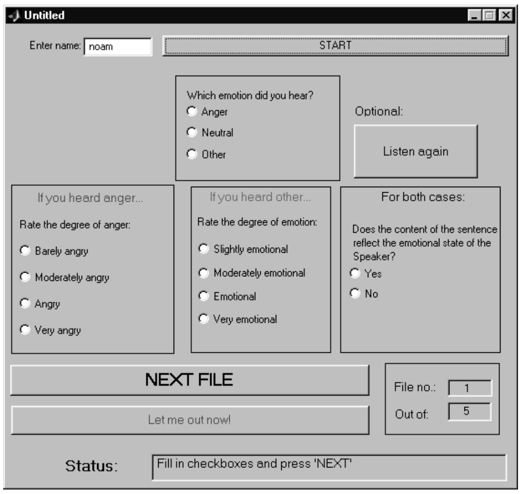
Extensive recordings were carried out, of televised political talk shows. One of the problems often encountered in such programs it that the participants tend to interrupt each other and speak simultaneously, which renders such recordings useless for the type of analyses carried out here. Eventually a sufficient amount of data was collected, from 8 adult native Hebrew speakers, 6 men and 2 women. The data was separated into phrases consisting of one breath unit. Any phrases that contained questions were removed, in order not to confuse final rises with rises due to emotion.

2.2 Data validation listening tests

The phrases collected were subjected to a listening test, using 11 adult native Hebrew speaking judges, 3 men and 8 women. The phrases were presented at random order, and the judges were asked to fill the automated questionnaire shown in figure 1.

For phrases that were judged to contain anger, listeners were asked to rate the degree of anger on a scale from 1 to 4. 12 phrases (uttered by 5 speakers) that received an overall score of 22 or more (out of a maximum of 44) were retained as representing anger. 12 phrases (from 8 speakers) that were scored by 7 or more judges as neutral were retained to represent neutral speech.

In order to examine the influence of the textual content of the utterances on the judgment of the utterances, a different set of 11 judges was asked to score only the texts of these utterances, on a simpler scale of: anger, neutral or other emotion. Only three texts were judged to represent anger by more than 6 judges.



Untitled

Enter name: noam

START

Which emotion did you hear?
○ Anger
○ Neutral
○ Other

Optional:

Listen again

If you heard anger...

Rate the degree of anger:
○ Barely angry
○ Moderately angry
○ Angry
○ Very angry

If you heard other...

Rate the degree of emotion:
○ Slightly emotional
○ Moderately emotional
○ Emotional
○ Very emotional

For both cases:

Does the content of the sentence reflect the emotional state of the Speaker?
○ Yes
○ No

NEXT FILE

Let me out now!

File no.: 1

Out of: 5

Status: Fill in checkboxes and press 'NEXT'

Figure 1: A screen shot of the listening test questionnaire

3. Feature extraction and normalization

The raw F0 features that were extracted from the utterances were mean pitch, minimum pitch, maximum pitch and pitch range. Since these are speaker dependant to a certain degree, they must be normalized before a comparison is carried out across speakers. Normalization is carried out with respect to values obtained from neutral utterances, though the exact normalization to be used can vary. Several possible normalizations were examined in this study, giving the following F0-based feature set:

- R1-range(anger)/mean(anger)
- R2-range(anger)/mean(neutral)
- R3-range(anger)/range(neutral).
- A1- mean(anger)/mean(neutral)
- M1- minimum(anger)/mean(neutral)
- M2- minimum(anger)/minimum(neutral)
- X1- maximum(anger)/mean(neutral)
- X2- maximum(anger)/maximum(neutral)

Spectral tilt was also analyzed. In order to get an idea of the variability of this parameter, it was computed separately for each word in the recorded utterances. The LTAS (long term average spectrum) was computed for each word, and spectral tilt was obtained from the ratio between energy in the 0-1 kHz band to the energy in the 2-5 kHz band.

4. Results

4.1 Angry vs. neutral speech - overall

Figure 1 shows the unnormalized results for the 5 speakers, in neutral and angry speech. The differences in mean pitch across speakers are obvious from the figure. The differences between angry and neutral speech are clearly obvious in this figure: all 4 of the basic features tend to be higher for angry speech, for all speakers. There is some variation in detail, on the other hand. For example, for speaker 2, pitch range is nearly the same for neutral and angry pitch, but pitch minimum and maximum are clearly higher for anger than for neutral. For speakers 4 and 5 on the other hand, there is only a slight rise in minimum pitch during anger, and a much larger rise in maximum pitch.

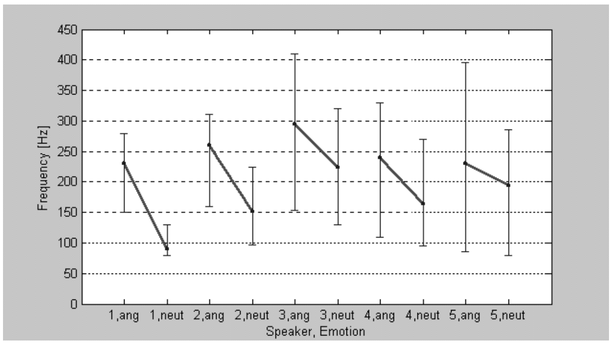


Figure 2: absolute F0 mean, max and min for neutral and angry speech, for 5 speakers

In order to compare these results across speakers, Table 1 shows the above parameters for anger, normalized by their corresponding values for neutral speech, i.e. features R3, A1, M2, X2

Table 1: Normalized features

Speaker	Mean (A1)	Range (R3)	Min (M2)	Max (X2)
1	2.39	2.33	2.005	2.14
2	1.69	1.06	1.78	1.37
3	1.33	1.33	1.19	1.27
4	1.39	1.23	1.19	1.22
5	1.19	1.54	1.07	1.42

All the values in this table are larger than 1, indicating that the corresponding feature is larger in anger than in neutral speech. Large differences in absolute values between speaker 3 and 4 for example, as in figure 1, translate to very similar normalized values as found in table 1. On the other hand we can find different means of expressing anger in different speakers. For example, for speakers 1-4, The normalized mean is higher than all other features. For speaker 5, on the other hand, mean F0 during anger is not much higher than neutral, nor is minimum F0, but the maximum F0 and therefore the range of F0 show a considerable increase during anger.

4.2 Analysis of degree of anger

An additional aspect of the present study is the fact that listeners were asked to rate the degree of anger in the angry utterances. A graphical summary of these ratings for the 12 angry utterances is presented in figure 3:

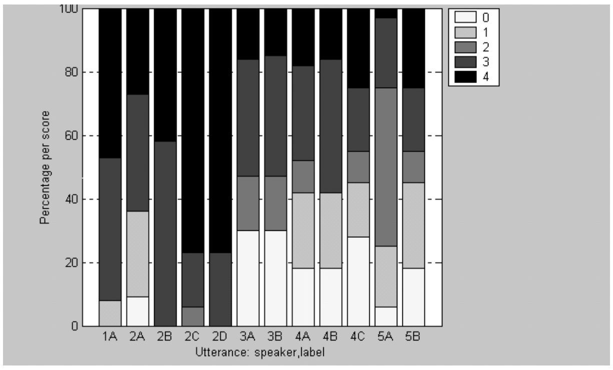


Figure 3: distribution of rating values for angry utterances

The utterances that received the most uniform ratings were those that were judged to be angriest, such as 1 and 2D. some speakers, such as speaker 2, received relatively uniform ratings, whereas ratings for speaker 5 for example had a larger degree of variability.

To learn more from the ratings received by each utterance, we examined their correlations with the various features listed above. The results are summarized in the following few paragraphs:

Normalized mean F0 (A1) during anger was found to have a strong and significant positive correlation with the degree of anger (P=0.0023, r=0.7893).

Normalized minimum F0 (M1) during anger was found to have a strong and significant positive correlation with the degree of anger (P=0.0007, r=0.8373).

Normalized maximum F0 (X1) during anger was found to have a weak and insignificant positive correlation with the degree of anger (P=0.2369, r=0.3697).

Normalized F0 range (R2) during anger was found to have a moderate and insignificant negative correlation with the degree of anger (P=-0.43, r=0.1629).

The above features were examined on an unnormalized scale also. This type of comparison is problematic for absolute values such as mean, minimum and maximum. On the other hand it can be considered valid for the measure of pitch range, since there is no absolute criteria for normalizing this parameter. It is therefore interesting to note that absolute F0 range (as opposed to normalized pitch range) exhibited a strong and significant negative correlation to the degree of anger (P=-0.853, r=0.0004).

4.3 The effect of textual content

One of the problems with analyzing anger in spontaneous speech is that there is no control over the textual content, as there is in acted speech. We therefore attempted to examine the degree to which the judges in the listening tests were influenced by the textual content of the utterances.

3 of the angry utterances were judged to have angry content by 7 or more of judges in the second test (text only), whereas 9 utterances were not. These 3 utterances were found to have received a significantly higher rating of anger in the listening tests, with an average score of 38.1 vs. an average score of 24.9 for the other 9 angry utterances.

The acoustic content of these two sets of utterances were also compared, appearing in Figure 4. F0 minimum was found to be significantly higher and F0 range was found to be significantly lower for the utterances whose textual content was judged to be angry.

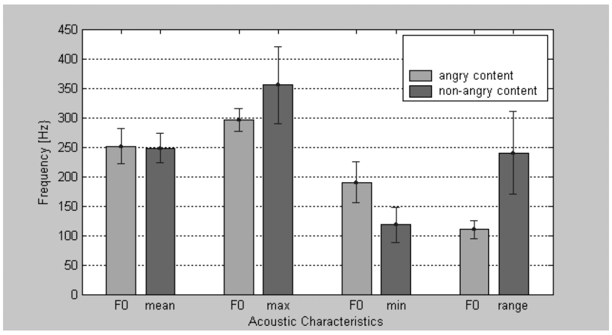


Figure 4: F0 features for angry utterances judged to have non-angry content

Correlation between the normalized acoustic variables and human ratings was carried out once more, for the utterances whose content was not judged to express anger. The general trends found in the previous subsection remained, but the only one that remained statistically significant was the positive correlation between F0 mean and degree of anger.

4.4 Spectral tilt

Spectral tilt was computed separately on the angry and neutral utterances, after the files were separated into discrete words. All speakers shows a significant decrease in spectral tilt for angry speech as compared to neutral speech. The results are presented graphically in figure 5. A similar comparison was carried out on a different measure of spectral tilt the slope obtained from the linear regression over the spectrum from 0 to 5 kHz. Once more this gave a significant decrease in slope of angry vs. neutral utterances.

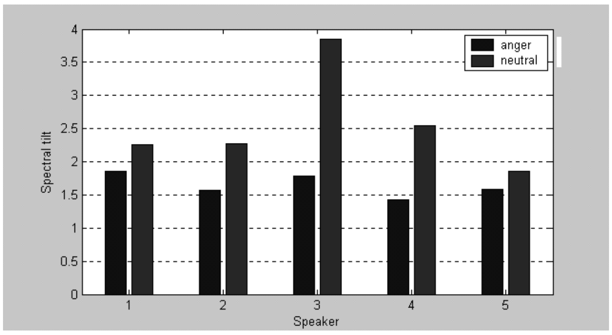


Figure 5: Average spectral tilt for each speaker in angry and neutral speech

5. Discussion

Despite the fact that the corpus examined in this study is rather small, a number of conclusions can clearly be drawn from the results presented above. All of the features used in this study showed a tendency to rise during anger, as compared to neutral speech, though they do not rise to the same degree for different speakers. As can be expected, the details of emotional expression vary somewhat from speaker to speaker. This can be expected to be a more marked effect in natural speech as opposed to acted speech.

The degree to which these features change was not necessarily significantly correlated to the degree of anger. Most markedly, mean F0 was found to be significantly and positively correlated to the degree of anger, yet it is not necessarily clear that this is the most important perceptual cue that listeners use to identify anger. For example, the utterance that consistently received the highest score in this study shows a marked increase in pitch. Yet the voice quality of the speaker of this utterance was so different from his voice quality in normal speech, that the listeners had no real baseline for comparison. In fact, this points to the fact that voice quality might indeed be an important perceptual cue to anger. This is borne out by the fact that a reduction in spectral tilt was found to be significantly correlated to anger.

It is interesting to note that though pitch range increased in angry speech as compared to neutral speech, it actually showed a significant negative correlation to the *degree* of anger. One possible explanation is that in mild anger the speaker may have momentary rises in pitch, but the pitch minimum is probably similar to that of neutral speech. In states of high emotional excitation, on the other hand, there is a rise in minimum pitch also, causing the pitch range to become in fact smaller.

Some further informal tests were carried out, in attempt to edit the pitch contour of angry utterances to observe whether they could be rendered emotionless. This was carried out using Praat software, by editing the pitch contour with the mouse (the pitch shifting itself is performed by Praat using a PSOLA algorithm). Preliminary results showed that performing a pitch shift on an entire angry utterance could reduce the impression of anger considerably. This is somewhat surprising, considering that the intonation curve in itself was not changed. The resultant spectral tilt was not examined, but listening to the resultant synthesis gave an impression of a more normal voice quality this may be due to the manner in which the PSOLA algorithm operates.

In summary, this study suggests that study of emotions as they occur in natural speech can give results that prove significant, both in listening tests and in acoustic analysis. This type of study can provide a better idea of the variability in expression of emotions as they are manifested in natural speech, as opposed to acted speech. Anger is an emotion that can be found relatively easily in the kind of setting examined here, but probably other emotions can be obtained also such as happiness, surprise, frustration, disgust, and maybe even fear. Kehrein [6] has shown that some of these emotions can also be elicited using a straightforward experimental setup.

It seems that one course of study not found widely in the literature is to change the properties of recorded utterances in an attempt to modify their emotional content using prevalent and easy to use speech analysis and synthesis software such as Praat. Such a methodology, combined with listening tests, can be used to verify general hypotheses as to the perceptual cues of emotion in speech. Such a further study is currently being carried out by the authors.

6. Conclusions

In this study we collected a corpus of angry and neutral speech from televised political talk shows. 12 angry utterances and 12 neutral utterances, as determined by listening tests, were analyzed for both F0 and spectral tilt. All of the F0 features examined were higher in angry speech as compared to neutral speech, and some of these features showed a significant correlation to the *degree* of anger: a rise in pitch mean, and a decrease in pitch range. A decrease in spectral tilt was also significantly correlated to the presence of anger.

This study shows that it is possible to obtain natural emotional speech from public sources, and that studying this type of data can be fruitful. It would probably be useful to obtain a larger corpus involving more emotions, using the same means. The resultant conclusions as to the manifestation of emotion in speech could then be further verified by manipulating the F0 contours of neutral and emotional speech, and conducting listening tests on the resultant synthetic speech.

7. References

[1] Banse, R. and Scherer, K., Acoustic profiles in emotion expression, *Journal of Personality and Social Psychology*, 70(3), 614-636, 1996

[2] Yang, L. and Yunxin, Z., Recognizing emotions in speech using short tem and long term features, *Proceedings of ICSLP 98*, Sydney

[3] Dellaert, F., Polzin, T, Waibel, A., Recognizing emotion in speech, *Proceedings of ICSLP 96*

[4] Douglas-Cowie, E., Cowie, R., Schroder, M., A new emotion database: consideration, sources and scope, *ISCA workshop on speech and emotion*, Belfast 2000

[5] Amir, N., and Ron, S. Towards an automatic classification of emotion in speech, *Proceedings of ICSLP 98*, Sydney

[6] Kehrein, R., Prosodie und Emotionen, Tuebingen Neimeyer 2002

[7] Cowie, R.,Describing the emotional states expressed in speech, *ISCA workshop on speech and emotion*, Belfast 2002