A Study on Emotional Feature Recognition in Speech

Li Zhao    Wei Lu    Ye Jiang   Zhenyang Wu
Department of Radio Engineering Southeast University
Nanjing, Jiangsu 210096, People’s Republic of China
Email: zhaoli@seu.edu.cn

ABSTRACT
This paper analysis the feature of the time, amplitude, pitch and formant construction involved such four emotions as happiness, anger, surprise and sorrow. Through comparison with non-emotional speech signal, we sum up the distribution law of emotional feature including different emotional speech. Nine emotional features were extracted from emotional speech for recognizing emotion. We introduce three emotional recognition methods based on principal component analysis and the results show that these method can provide an effective solution to emotional recognition.

1. INTRODUCTION
Language is a very important tool for us to communicate. Our speech includes not only symbols and words but also our emotion and mood. For example, listener will feel different if speaker in different mood speaks the same sentence. The traditional information science only deals with the “non-neural” knowledge world, such as the accuracy of the deliberation of information, but it totally neglects the emotional element. So it only reflect one aspect of information. The emotional science world is corresponding to the knowledge science world. It’s an important component of information process. So the artificial process of emotional feature are of very important meaning in the field of signal process and artificial intelligence[1, 2].

This paper analyzes the feature of the time, amplitude, pitch and formant construction of different emotional speech, then compare them to the neutral speech to find the different emotion feature’s distribution. At the same time this paper extracts 9 emotional feature parameters, applying principal element analysis method to identify different emotions. We got a satisfaction result.

2. SELECTION OF EMOTIONAL SENTENCE
This paper takes two aspects into account: First, the sentences don’t have any emotional tendency. Second, the sentences can involve all kinds of emotion to analysis and compare. According to the two rules, we use four sentences as the speech material. The classification of emotion is a very important part of the research of emotional analysis. There’re different methods to classify emotion. But in view of engineering no subject has been proposed yet. This paper sorts emotion into happiness, anger, surprise and sadness. To obtain the original data of the speech, we use five male speakers who are good at acting to speak the sentences three times in emotion of happiness, anger, surprise and sadness. At the same time we make the speakers speak each sentence three times in a neutral way. In this way, we get 300 sentences for experiment. Among them, 200 are used for training and 100 are used for test.

The digital signal was stored in a PC computer. We played all the emotional sentences randomly, the listeners (none of them are speakers) would decide which type of emotion each sentence involve. The listeners do this all by their feeling subjectively. After repeated comparison, we selected the fittest sentences.

3. ANALYSIS AND EXTRACTION OF EMOTIONAL FEATURE
General speaking, the emotional feature of speech signal are always represented as the change of prosodic[3, 4]. For example, when a man is in rage, his rate of speech, volume and tone will all get higher. Because the emotional information of speech signal is more or less affected by the meaning of the sentence, we get the distribution and construction characteristic by analyzing the relationship between emotional speech and neutral speech to remove the effect and to recognize different emotion. In our study, we compare the feature of the time, amplitude, pitch and formant construction of four types of...
emotional speech to neutral ones to discover the distribution and tectonic characteristic of different emotional speech. For convenience’s sake, the ratios of the emotional feature parameter are shown below. They are the ratios between emotional feature parameter and neutral ones.

3.1 Analysis of time construction

With an eye on the difference of different emotional speech’s time construction, analysis of the time construction of emotional speech is to analyze and compare the difference of the change of duration aroused by emotion. In this paper, we calculate the duration of each emotional sentence. They include the part of silence, because these parts contribute to the emotion. Then we compare and analyze the relationship between the average duration of emotional sentence (T), the average rate of speaking and emotion.

From the results, we can see that the duration of anger and surprise is shorter than that of neutral and happiness and on the contrary the duration of sad speech are shorter. The duration of anger is shorter than that of surprise. The duration of neutral speech is much shorter than the one of sadness and little shorter than the one of happiness. Through farther observing, we can know these phenomena are caused by some faintly pronouncing, prolonged, omitted phoneme (compared to the neutral speech) in emotional speech. Based on the results above, we can recognize happiness, sadness and other emotion easily by comparing their duration to the neutral one. We can also set some thresholds to recognize happiness and sadness. But we can’t distinguish anger and surprise efficiently only by duration.

3.2 Analysis of amplitude construction

In general, the amplitude feature of signal closely relates to all kinds of emotional information. We can feel in our real life that when a man is in a rage or in surprise, his volume is very high, however, when he is in sorrow, his volume is very low. So in some study of emotional analysis, amplitude construction is regarded as the most important feature. In our paper, we mainly take the average amplitude energy and the range into account to analyze and compare. We calculate the shot term energy of each frame of the signal, and analyze their characteristics that vary with time. To remove the effect of the silent and noisy part of the speech, we only take the average value of the absolute value of the amplitude into account. All the absolute values are bigger than a threshold.

We can see from the results that the amplitudes of signals which involve happiness, anger, surprise are bigger than the one of neutral one, however the signal that involves sadness shows contrary characteristic. We can also learn from the listening experiment that the bigger the average of amplitude of happiness, anger, and surprise are or the smaller the one of sadness is, the more emotional effect the sentence will show.

3.3 Analysis of pitch construction

Pitch construction is also an important feature that reflects emotional information. To analysis the characteristic of pitch construction in emotional speech signal, we calculate the smooth pitch envelops of the emotional speech signals, then analyze the characteristic of different pitch envelopes to find out the characteristic of the fundamental frequency constructions of different emotional speech signals. This paper analyzed the range, the average fundamental frequency and its rate of change of the envelopes of different emotional speech signals. Here \( F_0 \) refers to the mean absolute value of the difference between each frame of speech signal’s fundamental frequencies.

Compared to the neutral speech signal, the range and \( F_0 \) of happiness, anger and surprise are bigger while the ones of sadness are smaller. In view of happiness, anger and surprise the parameters of surprise is the biggest, then are happiness and anger. In addition, we can see that the envelop of surprising speech signal has the characteristic of raising at the end of the sentence. This characteristic is very useful for us to distinguish surprise from other emotion.

3.4 Analysis of formant construction

Formant is an important parameter that reflects the characteristic of track. We can predict the different location of formant of different emotional speech signals because different emotional speech changes track differently. We get formant in two steps: 1) Apply LPC method to calculate the PSD’s envelop of track; 2) apply the Peak Picking method to calculate the frequency of the formant. This paper only researches the average, range and rate of change of \( F_1 \) of the first formant.

We can see that compared to neutral, the frequencies of
first formant of happiness and anger are little higher while the one of sadness is much lower. We can also learn by farther observation that when we express the emotion of happiness or anger, our mouths always open bigger than usual and when we express emotion of sadness, our mouths open smaller and our speech always go with faint snuffle. The $F_{1_{\text{way}}}$ of four types of emotion are bigger than the one of neutral. Among them, surprise is the biggest. However, the $F_{1_{\text{way}}}$ of all types of emotion are smaller than the one of neutral. Among them, sadness is the smallest.

We compared the four type of emotional speech to the neutral. We can get Tab 1

<table>
<thead>
<tr>
<th>Tab 1. The variability of all the parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
</tr>
<tr>
<td>Happiness</td>
</tr>
<tr>
<td>Anger</td>
</tr>
<tr>
<td>Surprise</td>
</tr>
<tr>
<td>Sadness</td>
</tr>
</tbody>
</table>

( + : increase ++ : increase more _ : decrease _ _ : decrease more - : no change )

4 EXPERIEMNT AND RESULTS OF EMOTIONAL RECOGNITION

In our experiment, we extract nine emotional feature parameters said above and apply principal element analysis to carry out the experiment. So any emotional sentence can be characterized as a 9-dimensional origin eigenvector ($X_i$).

Before analyzed, each element of the vector will be transfer into the parameters that conform to the standard gaussian distribution, because the units of each elements aren’t consistent with each other.

4.1 Method of recognition

4.1.1 Method 1

First we apply principal component analysis\(^7\) to transfer a D-dimensional vector $\{x_{ij}, x_{i2}, \ldots, x_{id}\}$ into a p-dimensional one $\{\bar{x}_{ip}, x_{i2}, \ldots, x_{ip}\}$ ($p \leq D$). Then we computer the mean value and the corresponding variance. Then accord formula (1) we computer the Mahalanobis distance measure to do recognition directly.

$$D_i = (\bar{y} - \mu_1) \Sigma^{-1} (\bar{y} - \mu_1)$$

(1)

4.1.2 Method 2

We apply gaussian mixture model (GMM) in method 2. For example, the accumulated probability of a original vector can be get by formula (2)

$$P_k = w_1 f_1(\bar{y}) + w_2 f_2(\bar{y}) + w_3 f_3(\bar{y})$$

(2)

where $f_i(\bullet) (i = 1, 2, 3)$ obey the gaussian distribute law, and $w_1 + w_2 + w_3 = 1$ is weight coefficient. In training step, we first get the codebook of the original vector by applying vector quantization(VQ). Then we computer the corresponding variance of each codevector. At last we computer the probability of the principal component $Y_{ip}$ of each emotion type by formula (2). The biggest probability corresponds to the result of emotion recognition.

4.1.3 Method 3

To choose more reasonable virtual principal element, we firstly transform principal element of emotional feature parameters based on maximum separability rule, then according to formula (3) we mark $Z_{ij}$ of the $j$th main element of the $i$th sentence.

$$Z_{ij} = \sum_{i=1}^{n} \frac{a_{ij}(x_{ij} - \bar{x}_i)}{\sqrt{\Sigma_{ii}}}$$

$$x_{ij} = \frac{1}{N} \sum_{i=1}^{n} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)$$

(3)

Here $a_{ij}$ is the element of the $j$th basic vector of main...
element—— $A = (a_1, \ldots, a_d, \ldots, a_M)'$, and $X_{id}$ is the element of the 9-dimensional original eigenvector of the $i$th sentence.

$$P'_{ij}(Z_{ij}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(Z_{ij} - \mu_{ij})^2}{2\sigma^2}\right)$$  (4)

$$P'_{ij} = \prod_{j=1}^{L} P'_{ij}(Z_{ij})$$  (5)

$$\max_{i=1}^{P} \{P'_{ij}\}$$  (6)

We can mark the virtual main element of the $X_i$ of the emotional sentence to be recognized, then we can use the value of $\mu_{ij}$ and $\sigma_{ij}$ to calculate the output probability of the mark $Z_{ij}$.

The joint probability of the virtual main element of each type of emotion can be calculated according formula (5). Here $P$ is the number of virtual main element. The result is the type $k$ that accord with formula (6).

### 4.2 Results

As was said above, we use 100 sentence to carry out the emotional feature recognition experiment. The results are shown in tab.2.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>happiness</th>
<th>anger</th>
<th>surprise</th>
<th>sorrow</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>method 1</td>
<td>80</td>
<td>85</td>
<td>70</td>
<td>100</td>
<td>83.75</td>
</tr>
<tr>
<td>method 2</td>
<td>80</td>
<td>90</td>
<td>80</td>
<td>100</td>
<td>87.5</td>
</tr>
<tr>
<td>method 3</td>
<td>85</td>
<td>95</td>
<td>80</td>
<td>100</td>
<td>90</td>
</tr>
</tbody>
</table>

### 5 CONCLUSION

This paper analyzes and compares four aspects of four types of emotion and finds out the distribution of different emotional speech. Based on these, we propose a emotional feature recognition method. We use MATLAB to do the emulation experiment by testing 100 emotional speeches. We get a the results that are close to the one s of human. On the other hand the parameter is not fit for distinguishing happiness and surprise. We will devote ourselves to finding more efficient parameters to do farther analysis and experiment in a wide field.

### 6. REFERENCE