Computational Linguistics and Chinese Language Processing
Vol. 12, No. 1, March 2007, pp. 33-48

Affective Intonation-Modeling for Mandarin Based on PCA

Zhuangluan Su*, and Zengfu Wang*

Abstract

The speech fundamental frequency (henceforth F0) contour plays an important role in expressing the affective information of an utterance. The most popular F0 modeling approaches mainly use the concept of separating the F0 contour into a global trend and local variation. For Mandarin, the global trend of the F0 contour is caused by the speaker's mood and emotion. In this paper, the authors address the problem of affective intonation. For modeling affective intonation, an affective corpus has been designed and established, and all intonations are extracted with an iterative algorithm. Then, the concept of eigen-intonation is proposed based on the technique of Principal Component Analysis on the affective corpus and all the intonations are transformed to the lower-dimensional eigen sub-space spanned by eigen-intonations. A model of affective intonations is established in the sub-space. As a result, the corresponding emotion (maybe a mixed emotion) can be expressed by speech whose intonation is modified according to the above model. The experiments are performed with the affective Mandarin corpus, and the experimental results show that the intonation modeling approach proposed in this paper is efficient for both intonation representation and speech synthesis.

Keywords: Eigen-Intonation, Affective Speech, Mixed Emotion, F0 Contour, Speech Synthesis

1. Introduction

Speech can convey not only literal meanings, but also the mood and emotion of a speaker. Some researchers have proven that the contour of the speech fundamental frequency (henceforth F0 contour) plays an important role in expressing the affective information of an utterance. It is concluded that some statistical characteristics of F0 play the most important roles in emotion perception [Tao and Kang 2005]. Especially, F0 contours differ from each

* Department of Automation, University of Science and Technology of China, Hefei 230027, China
E-mail: zfwang@ustc.edu.cn
The author for correspondence is Zengfu Wang.

[Received September 16, 2006; Revised October 10, 2006; Accepted November 1, 2006]

other because of the speaker's different emotion in Mandarin [Yuan *et al.* 2002]. Due to significance of F0, the F0 contour modeling is one of the key issues that should be addressed.

The most popular F0 modeling approaches mainly use the concept of separating the F0 contour into a global trend and local variation [Abe and Sato 1992; Bellegarda *et al,* 2001]. Mandarin is a tonal language including four basic tone types and a so-called 'light' tone. The F0 contour is composed of three elements [Zhao 1980]: the tone of the syllable, the variety of tone in continuous utterance, and the movement influenced by mood. How to extract tones and intonations from speech is a difficult problem. Tian and Nurminen have proposed a data-driven tone modeling approach to describe the tonal element [Tian and Nurminen 2004]. In previous work [Su and Wang 2005], the authors of this paper also proposed an affective-tone modeling approach for Mandarin to separate F0 contour into two elements: variational tones based on syllables and intonations for prosody phrases.

In this paper, the authors propose a data-driven intonation modeling approach based on Principal Component Analysis (henceforth, PCA [Fukunaga 2000]). For modeling affective intonations, an affective corpus of Mandarin has been designed and the corresponding intonations are extracted with an iterative algorithm from the original speech. The eigen-intonation concept is proposed based on the principal components of the above intonations obtained from the affective corpus, and all the intonations are then transformed into the sub-space spanned by the eigen-intonations. The distribution of affective intonations corresponding to an emotion in the above sub-space is a help to establish the corresponding affective intonation model. As a result, speech whose intonation is modified according to the model can express the corresponding emotion, even mixed emotions. In addition, the authors will also show emotion perception results using the proposed modeling approach.

The remainder of the paper is organized as follows. The speech corpus and some statistic results of F0 based on the database are described first. Then, the algorithm of eigen-intonation extraction is described, and some of the basic properties of the eigen-intonation representation are concluded. Next, how to model the affective intonation is discussed. Last, the performance of the proposed modeling approach is given by experimental results.

2. Speech Corpus and Statistic Results of F0

Carrying on the affective speech research, a reasonable classification of the emotion is needed first, and then the speech features with different emotions can be analyzed effectively. In emotional psychology, Robert Plutchik proposed a four pair emotional ring constructed of eight pure emotions, including anger, joy, acceptance, surprise, fear, sadness, hatred and expectation. In the affective speech research for Mandarin, four emotions are generally selected, either including anger, joy, fear, sadness [Yuan *et al.* 2002; Tao and Kang 2005], or including anger, joy, surprise and sadness [Zhao *et al.* 2004]. In contrast, five emotions are selected for this paper, and they are anger, joy, surprise, fear and sadness.

What is discussed in this paper is the global variety of the F0 contour, so a reasonable duration of the target needs to be considered. Due to the multi-level structure of prosody [Abney 1995; Li *et al.* 2000], a complicated sentence with many syllables can be divided into several simple prosody units with fewer syllables at prosody boundaries. So, studying intonation based on prosody units can transform this complicated problem into several simple ones. Moreover, it is known that prosodic phrases can keep a relatively stable intonation pattern. Therefore, the authors model intonation based on prosodic phrases in the paper.

It is known that F0 contour is influenced by several factors, including syntax, stress, speaker's emotion and his or her individual character. This paper focuses on the movement of intonation caused by emotion, and the influence of other factors such as syntax, stress, and the individual characters will not be considered. Currently, there are no effective methods that can eliminate the influence of these factors from the original speech signals directly, so the corpus used in the paper are obtained in such a way as to avoid these interferential factors' influence.

To avoid unwanted factors' influence and to simplify the following processing, the corpus is designed with some limitations. The authors have designed 40 sentences with different literal contents for the following test, and each sentence only consists of three components: subject, verb, and object. Furthermore, the subject, verb, and the object are all designed to be disyllabic words. So, each sentence only has 6 syllables in this case, and all of these sentences have the same syntax. As the length of a prosodic phrase is approximately six syllables [Zhao *et al*. 2002], each sentence consists of only one prosodic phrase. An example of such a sentence is given by "北京召开奥运". This design can be advantageous to the following experiments, and the model will be established directly based on one sentence. Each sentence is then performed by a female actor with all six emotions, including fear, sadness, neutral, anger, joy and surprise. In the end, the corpus used for analysis contains 240 total sentences, consisting of 1,440 syllables from a single speaker, with same syntax and the same individual characters. The speech signals are digitized at 16 kHz with 16-bit precision.

To evaluate the representational ability of the corpus, some experiments about the distributions of F0 are performed. Here, the F0 of a speech is extracted by using a modified autocorrelation algorithm. The results are demonstrated in Figure 1.

Figure 1 shows that "surprise", "happy" and "angry" make a very high F0, while "sad" generates lower value than the neutral state. It can also be found that the varying range of "sad" is smaller than the others. F0 parameters of "fear" make quite similar behaviors as "sad". "Angry", "happy", and "surprise" also behave similarly. All of the results accord with the conclusions given by other researches [Yuan *et al.* 2002; Zhao *et al.* 2004; Tao and Kang 2005]. So the speech corpus is representational and effective for the following analysis.

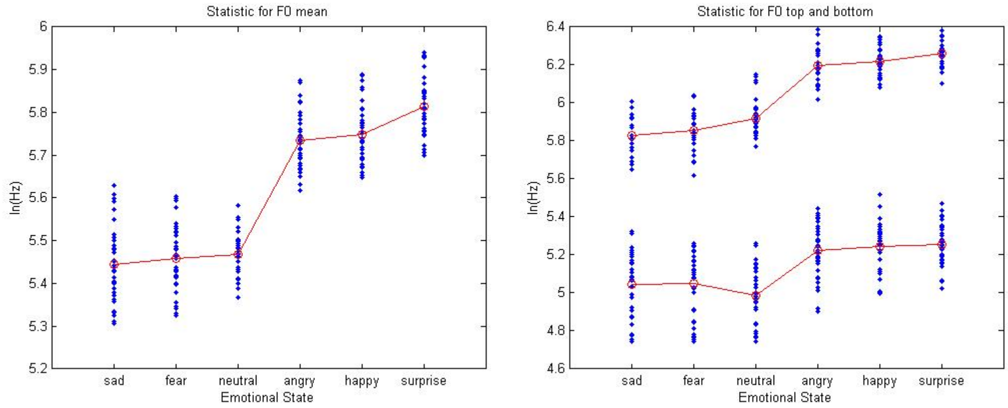


Figure 1. Statistic results for F0 with different emotional states

3. Concept of Eigen-Intonation

The affective intonation will be modeled with a concept called "eigen-intonation". The concept of eigen-intonation is derived through the use of the PCA technique. PCA [Fukunaga 2000] is a multivariate analysis method that carries out a compact description of a data set. In a PCA process, a set of correlated variables is transformed into a set of uncorrelated variables that are ordered by reducing variability, and these new uncorrelated variables are linear combinations of the original variables. It can be concluded that the first new variable contains the greatest amount of variation; the second contains the next greatest residual variance and orthogonal to the first, and so on. Thus, the last of these variables can be removed with a minimal loss of real data.

With the affective corpus in the paper, the speech intonations for sentences should be very similar in all configurations, and they should be able to be described by some "basic intonations". From the previous description, one knows that one of the main functions of PCA is that it can be used to extract new uncorrelated features from original data. According to these ideas, one can find the "basic intonations" that best account for distribution of speech intonations within the entire intonation space using the principal components analysis. The "basic intonations" are called "eigen-intonations".

With eigen-intonation, original intonations can be transformed to corresponding representations with lower dimensions. Some rules can also be possibly given out in the low-dimensional space. Moreover, the resultant rules with low dimensions have simpler expression, and it is advantageous to control the rules for the goal of this study.

4. Analysis for Eigen-Intonation

The concept of eigen-intonation is proposed based on PCA technique. Mathematically, the principal component analysis involves an eigen analysis on a covariance matrix. A good low-dimensional representation in the space of possible speech intonations can be achieved by considering only a few principal components or eigenvectors, corresponding to the first largest eigenvalues.

4.1 Extraction of Intonation

In order to obtain the intonation of a speech, the F0 contour of the speech should be extracted first. After that, the F0 contour will be separated into a global variety, which is regarded as intonation, and rapidly-varying components corresponding to local changes based on syllables. The details of intonation extraction are described in the following.

The entire intonation extracting algorithm can be divided into five main steps:

1) Estimating initial F0 values based the modified normalized autocorrelation from voiced regions of the original speech.
2) Cubic Hermite interpolating for unvoiced regions and obtaining a continuous F0 curve.
3) Filtering the continuous F0 contour with two serial modified smoothing processes.
4) Applying piecewise three-order polynomial iterative fitting to the entire F0 contour, the n-th iterative processing step is as:
    (a) Fitting the entire F0 contour with n pieces of cubic polynomial.
    (b) Calculating the fitting error E_n.
    (c) If $E_n < E_t$, ending the iterative algorithm and taking n pieces of cubic polynomial fitting as final resultant F0 contour. Else, $n = n + 1$, go to (a). Where E_t is a given threshold of maximal fitting error.
5) The ln(F0) contour is passed through a high-pass filter with a stop frequency at 0.5Hz, and the residual low frequency contour after filtering is denoted as L_F contour.

From the authors' previous work [Su and Wang 2005], The L_F contour can be regarded as the F0 global variety of a speech. As all sentences have the same syntax and each sentence consists of only one prosodic phrase in this corpus, the model can be established directly based on one sentence. It is to say that the resultant L_F contour of the algorithm for each sentence in the corpus is the modeling target, intonation based prosodic phrase (henceforth intonation). Finally, each intonation is normalized into an N-dimensional vector (N = 100 in the paper).

4.2 PCA for Intonation

Let the data set of intonations be $I_1, I_2, \ldots I_M$, where I_i is an N-dimensional intonation sample, and M is the number of intonations (M = 240 in the paper). Then the intonation covariance matrix $C_{N\times N}$ is computed by (1).

$$C = \frac{1}{M}\sum_{i=1}^{M}(I_i - m)(I_i - m)^T \tag{1}$$

Where, m is the average intonation calculated by (2).

$$m = \frac{1}{M}\sum_{i=1}^{M} I_i \tag{2}$$

The differential intonations matrix A is defined as (3).

$$A = \frac{1}{\sqrt{M}}\left[I_1 - m, I_2 - m, \ldots, I_M - m\right] \tag{3}$$

Then, $C = AA^T$ is an $N\times N$ covariance matrix. The eigen analysis on the covariance matrix $C_{N\times N}$ yields a set of positive eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ in descending order and the corresponding eigenvectors, $\{V_1, V_2, \ldots, V_N)$. The first L ($L < N$) eigenvectors, denoted as $U = \{V_i, i = 1, 2, \ldots, L\}$, are selected as principal components, and the intonations corresponding to these L vectors are so-called eigen-intonations, denoted as U_o.

The eigen sub-space spanned by the principal components U is called sub-space of intonation, denoted as P, and the original space of intonation is denoted as O. All intonations in O can be projected to be the corresponding representations in P. It is known that the dimension of P is lower than that of O, and one can establish the rules of intonation in P and then restore the resultant intonations in O. Obviously, rules with lower dimension are easily controlled. Next, restoration of intonation will be discussed.

4.3 Restoration Based on PCA

According to the principal component analysis, the original intonations in O are projected into the sub-space P as (4).

$$\Omega_k = U^T(I_k - m), \quad k = 1, 2, \ldots, M \tag{4}$$

Where, Ω_k is coordinate vector of the k-th intonation. With Ω, the intonation samples are restored as (5), and the final approximation of the original intonations I is given out as (6), denoted as J.

$$B = U\,\Omega \tag{5}$$

$$J_k = B_k + m \quad k = 1, 2, \ldots, M \tag{6}$$

Especially, let $B = U$ in (6), intonations corresponding to U can be given out, and that are eigen-intonations U_o. It can be concluded that although U_o is higher than U, the configuration of U_o is same as U. So the authors do not distinguish them when their configurations are discussed.

To evaluate the ability of restoration, the restoring rate for k-th intonation is defined as (7).

$$R_k = 1 - \frac{\|I_k - J_k\|}{\|I_k\|}, \quad k = 1, 2, ..., M \tag{7}$$

The final restoring rate of the entire algorithm is defined as (8).

$$r = \left\{ \sum_{k=1}^{M} R_k / M \right\} \times 100\% \tag{8}$$

5. Affective Intonation

5.1 About Affective Intonation

Affective intonation is the concept that a speech with a certain affective intonation can express a corresponding emotion. Some works of speech prosody have proposed much qualitative analysis for affective intonations, and this paper will try to give quantitative affective intonation rules. At last, speech whose intonation is modified according to a certain affective intonation obtained in the paper can express the corresponding emotion.

In order to research affective problems, emotion can be classified. Robert Plutchik [Plutchik 1960] considered that the emotions felt in normal human life were complicated and mixed, and considered some intensity of the eight pure emotions constructing a mixed emotion. So, in a similar way to him, all the mixed-emotional intonations are supposed to be defined by some vectors in the form of linear combination of the coefficients in the paper, where the vectors are the principal components U and the coefficient is the coordinate vector Ω_k in (4). Based on this assumption, one can easily change the coefficient corresponding to a certain eigen-intonation to control some configuration of final affective intonation for the goal. How to perform the assumption is discussed in the following.

5.2 Modeling Affective Intonation

Let the set of emotions be a, $a = 1, 2, \ldots, 6$ representing anger, joy, surprise, fear, sadness and neutral emotional state. Intonations extracted from the speeches with emotion a are denoted as N-dimensional vector I^a in original space O. Let $I = I^a$ in (4), and I^a be projected into the sub-space P, denoted as Ω^a. Ω^a is distributed in different regions in P for the different emotions a, and the mass kernel vectors $\bar{\Omega}^{\alpha}$ are computed as (9).

$$\overline{\Omega}^{\alpha} = \sum_{k=1}^{K_{\alpha}} \Omega_k^{\alpha} / K_{\alpha}, \quad \alpha = 1, 2, ..., 6 \tag{9}$$

Where, Ω_k^{α} is the projecting representation vector (henceforth projecting vector) in *P* of the k-th intonation with emotion *a*. K_a is the total number of all intonation samples with emotion *a*.

{$\overline{\Omega}^{\alpha}$, *a = 1, 2, …, 6*} are the resultant affective intonations with low dimension basing eigen-intonation. They are restored in the original intonation space *O* as (10).

$$T_{\alpha} = U\,\overline{\Omega}^{\alpha} + m, \quad \alpha = 1, 2, ..., 6 \tag{10}$$

Where T_a are the final affective rule-intonations (henceforth rule-intonations) and they can be applied directly to modify the target intonation for synthesizing affective speech, which will be performed in the following experiments.

6. Experimental Results and Discussion

6.1 Analysis on Eigen-Intonation

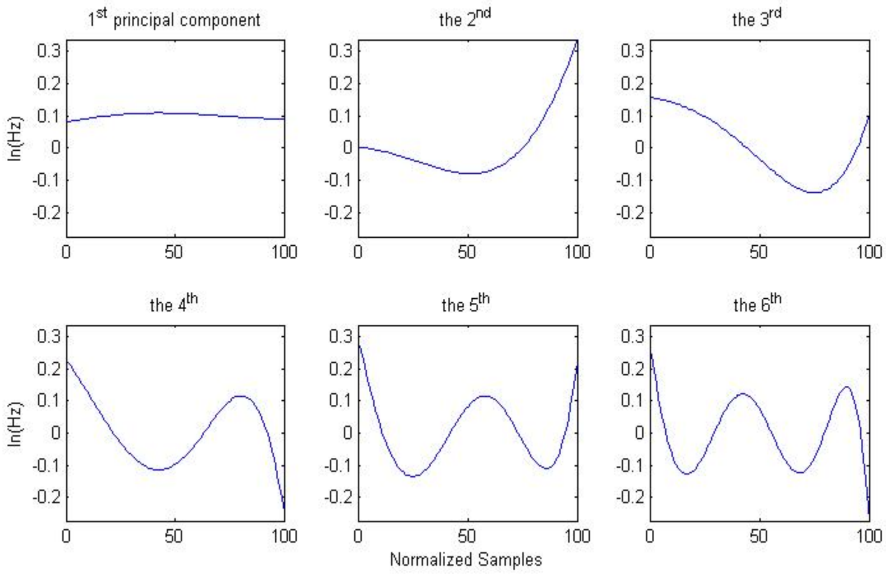


Figure 2. Eigen-intonation of the affective speech

To demonstrate the eigen-intonations, a PCA experiment using the affective speech corpus was performed. The first six principal components *U* are shown in Figure 2 and the authors do not distinguish the principal components selected and eigen-intonations here. It can be seen that the varying range of the first component is the smallest, and it is also the highest. So the first eigen-intonation represents the flat and positive pitch. The second eigen-intonation contributes a big rising component, and the third matches a falling intonation with a little rising at the end. The fourth can be viewed as adding a falling part to the end of the third. The varying ranges are same between the fifth and the sixth, and their global trends are flat with big rising and falling varying. These two can be viewed as adding a rising or a falling part to the end of the previous component. It will be known that the sixth component contains a very small contribution of energy or variance to the intonation contour in the following analysis.

Based on the previous resultant eigen-intonations, the authors carry out the restoring experiment using L components selected, respectively considering L be 3, 4, 5 and 6. The results are shown in Table 1.

Table 1. The restoring rate r with L components selected

L – component number	3	4	5	6
r – restoring rate	81.61%	95.71%	99.46%	99.89%

From Table 1, it can be concluded that selecting five components is acceptable, but with six principal components, the restoring rate is 99.89% and the approximation error is almost equal to zero. The approximating examples are shown in Figure 3. That means a good six-dimensional representation for the space of all speech intonations is achieved, and these eigen-intonations are very efficient for intonation representation.

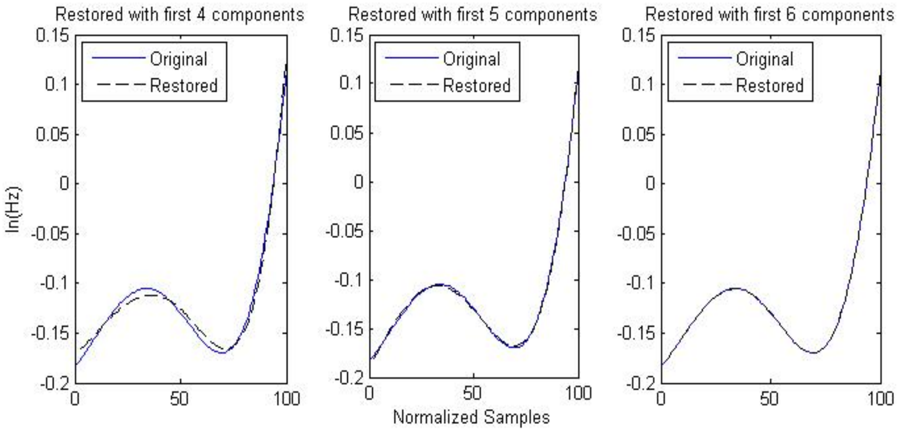


Figure 3. Illustration for restoring with eigen-intonations

6.2 Modeling Affective Intonation

The emotional state expressed by intonation of each affective speech in the corpus is known, and there are six categories of emotions, including the neutral state. And there are 40 speeches within each emotional state. According to Section 5.2, all affective intonations labeled with different emotions are projected into six-dimensional sub-space P spanned by eigen-intonations. The distribution of first three weights of the projecting vector Ω^a is shown as Figure 4, and the mass kernel of each emotional state is indicated by red color in the figure.

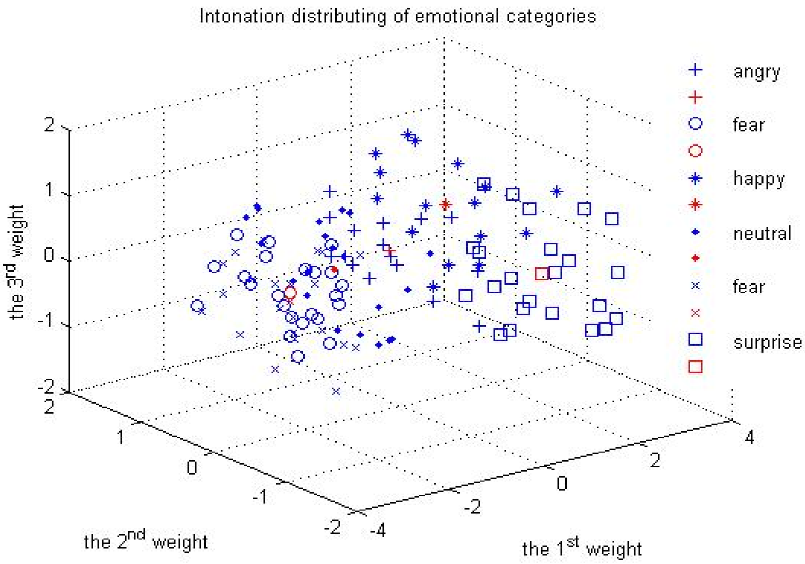


Figure 4. Distribution of first 3 weights of affective intonations in eigen sub-space

From Figure 4, one can see that the kernel of surprise, job and anger is far from that of neutral, where the “surprise” is farthest and then “angry” is next. However, the “fear” almost distributes in the same region with “sad”, and they can be distinguished from the neutral emotional region. In addition, it can be known from analysis on eigen-intonations that the last several weights corresponding to these three weights in the figure contain a very small contribution of energy or variance, so the difference of their distribution is not as clear as in Figure 4.

Now the projecting vectors Ω^a in P of original intonations labeled with emotion are given out as well as the corresponding kernel vector $\bar{\Omega}^{\alpha}$ for each emotional state. By restoring with eigen-intonations, the kernel vectors are transformed as (10) into the original space, there they are regarded as rule-intonations. The rule-intonations representing emotion states are illustrated in Figure 5. From the figure, one can see that the intonations of anger, job and surprise are high, where the variety of surprise is greatest. However, the “fear” is flat and low, similar to that of the “sad”. All these qualitative results are in line with the previous works of other researchers. So the resultant rule-intonations are efficient for expressing emotions in theory.

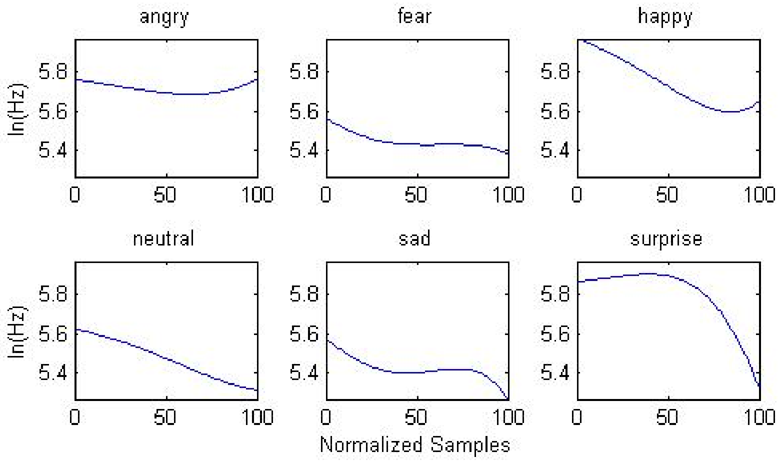


Figure 5. Affective rule-intonations T_α

6.3 The Mixed-Emotional Intonation

When the affective rule-intonation was modeled with eigen-intonation in the previous sub-section, the emotion labeled in the corpus and expressed by resultant intonation was supposed to be pure. It is known that the emotions of humans felt in normal life are not always so simple, and they are usually mixed with several so-called pure emotions, whose intensities differ corresponding to constructing the different emotions. The experiment is performed as the following to explain that the modeling approach proposed with eigen-intonation is also effective for representing the mixed-emotional intonation.

All affective intonations labeled with emotions have been projected into sub-space P and the distribution of first three weights of the projecting vector Ω^a in P has been shown in Figure 4. Now only the mass kernel of each emotional state, which is corresponding to the resultant rule-intonation, is represented in Figure 6.

Nine equal space points in line between the neutral kernel and the surprise kernel are selected and indicated in the figure. If the kernel explains pure emotions, then what the points selected explain are the mixed emotions. Along the arrow in Figure 6, points at the starting vertex explain more neutral and those at the ending vertex explain more surprise. So the emotions expressed by the intonations correspond to these points transfer from neutral to surprise along the arrow and they are mixed. The mixed-emotional intonations corresponding to the selected-points are restored in original space and shown in the left of Figure 7. It can be concluded from the figure that, along the arrow, the first rule-intonations can express more neutral and the last ones express more surprise and all of them express the mixed emotions.

Another nine equal space points between the happy kernel and the surprise kernel are also selected and the same experiment is performed. The illustrations of the experiment are shown in Figure 6 and the right of Figure 7.

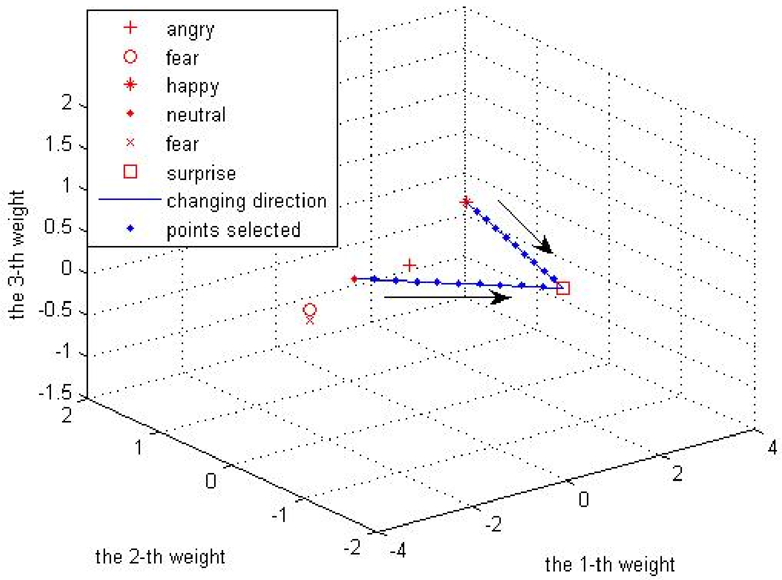
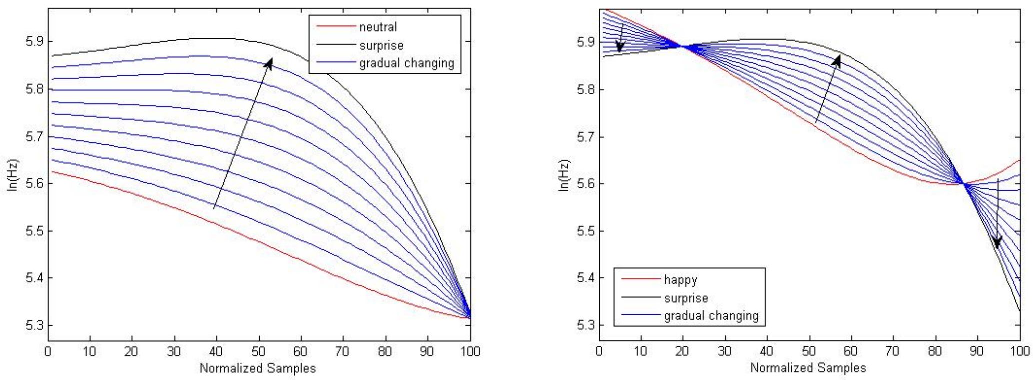


Figure 6. Transferring illustration of affective intonation in sub-space



Note:

The arrows in the figure indicate the gradual varying direction corresponding to that in sub-space showed in Figure 6 and each gradual changing curve is corresponding to one point selected in Figure 6.

Figure 7. Intonations transferring corresponding to that in sub-space

Figure 6 and Figure 7 show that the mixed-emotional intonation can be represented with eigen-intonation, so one can control the relative position of intonation-representation in the sub-space to explain the certain mixed-emotion felt in the usual human life. To sum up, the modeling approach proposed with eigen-intonation is effective for representing not only the simple emotional intonation but also the mixed-emotional intonation.

6.4 Synthesis with Affective Intonation

Based on the linear predictive coding technology [Quatieri 2004], the authors analyzed neutral speeches, modified their intonations with the six rule-intonations, respectively, and re-synthesized them. For example, the intonation of a neutral speech is modified to the surprise intonation, and the demonstration is shown as Figure 8. In the figure, the top is the waveform of the neutral speech, and the bottom includes the original F0 contour, the original intonation, the modified intonation, and the resultant F0 contour of the neutral speech. Moreover, the intonation of an original surprise speech is also plotted in the bottom figure for contrast. Figure 8 shows that the modified intonation is similar to the original intonation of the surprise speech, and the resultant F0 contour is higher than expressing surprise.

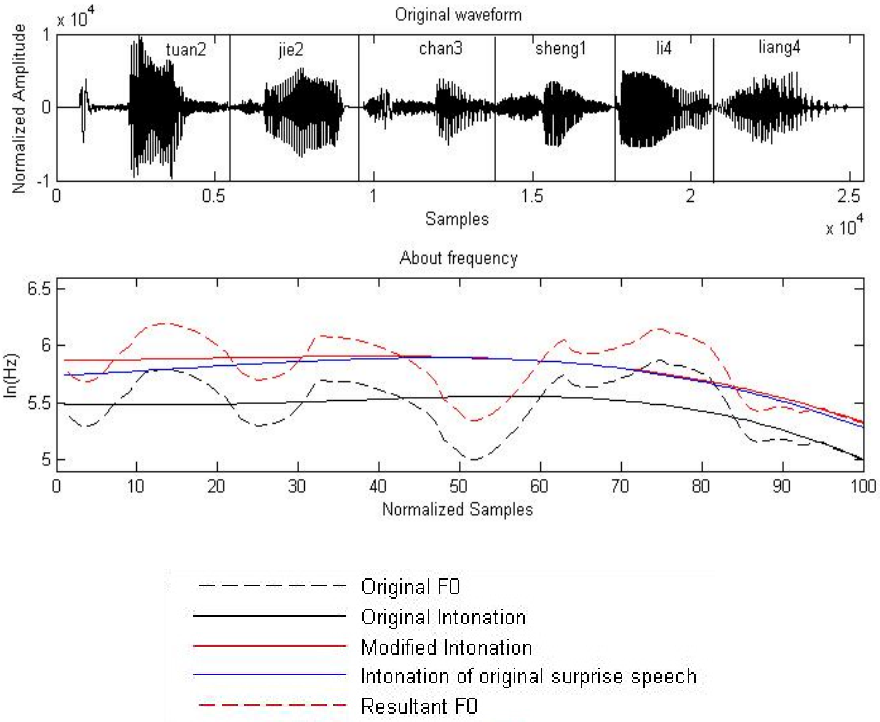


Figure 8. Illustration for modifying intonation with surprise rule-intonation

In the perception experiment, the listener was asked to judge the emotional state of the speech sound. The results show that, though it is difficult to distinguish anger from happy, and also can not point out whether the speech sounded closer to fear or sadness, it is easy to tell the emotional states such as joy, surprise, and fear of one speech. So one can conclude that the rule-intonations are almost corresponding to the emotional state and the eigen-intonation modeling method is efficient.

7. Conclusion

The F0 contour plays an important role in expressing the affective information of an utterance, and the most popular F0 modeling approaches are mainly using the concept of separating the F0 contour into a global trend and local variation. Mandarin is a tonal language, and the global trend of F0 contour is caused by speaker's mood and emotion, which is focused on in this paper, and that is called affective intonation. Affective intonation is the concept that a speech with a certain affective intonation can express a corresponding emotion. Some works of speech prosody have proposed much qualitative analysis for affective intonations, and the paper has given out quantitative rule-intonation.

In order to establish the model of affective intonation, an affective corpus of Mandarin was obtained with some limitation for affective research goal and all intonations were extracted from the original speeches. Then the eigen-intonation concept was proposed basing PCA on the affective corpus and all the intonations were transformed to lower-dimensional representations in the eigen sub-space spanned by eigen-intonations. A model of affective intonations was established in the sub-space and then was restored in the original space of intonation to form the rule-intonations. As a result, speech whose intonation is modified according to a certain rule-intonation can express the corresponding emotion, even the mixed emotion.

The authors have performed experiments with the affective Mandarin corpus. And the experimental results are in line with the theoretical analysis and the intonation modeling approach proposed is proved to be efficient for representing the simple emotional and mixed-emotional intonation. Future work will focus on how to accurately give out the boundaries of the pure emotional regions in sub-space with eigen-intonation.

Acknowledgements

This work is supported by Open Foundation of National Laboratory of Pattern Recognition, China. The authors would like to thank Dr. Tieniu Tan and Dr. Jianhua Tao for their help.

References

Abe, M., and H. Sato, "Two-stage F0 control model using syllable based F0 units," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, San Francisco, USA, 1992，pp.53-56.

Abney, S., "Chunks and dependencies: Bringing processing evidence to bear on syntax," in *Jennifer Cole and Georgia Green and Jerry Morgan(Eds.): Computational Linguistics and the Foundations of Linguistic Theory*, pp. 145-164, CSLI, 1995.

Bellegarda, J., K. Silverman, K. Lenzo, and V. Anderson, "Statistical prosodic modeling: from corpus design to parameter estimation," *IEEE Trans. Speech and Audio Processing*, 9(1), 2001, pp. 52-66.

Fukunaga, K., *Introduction to statistical pattern recognition*, Academic Press, Dordrecht, 2000.

Li, A., M. Lin, X. Chen, Y. Zu, G. Sun, W. Hua, Z. Yin, and J. Yan, "Speech corpus of Chinese discourse and the phonetic research," in *Proceedings of Sixth International Conference on Spoken Language Processing, 2000, Beijing, China,* pp. 13-18.

Plutchik, R. "The multifactor-analytic theory of emotion," *Journal of Psychology*, 50, 1960, pp. 153-171.

Quatieri, T. F., *Discrete-Time Speech Signal Processing: Principles and Practice*, House of Electronics Industry, Beijing, 2004.

Su, Z., and Z. Wang, "An Approach to Affective-Tone Modeling for Mandarin," *Lecture Notes in Computer Science 3784*，ed. By J. Tao, T. Tan, and R.W. Picard, Springer, 2005, pp. 390-396.

Tao, J., and Y. Kang, "Features Importance Analysis for Emotional Speech Classification," *Lecture Notes in Computer Science 3784*，ed. By J. Tao, T. Tan, and R.W. Picard, Springer, 2005, pp. 449-457.

Tian, J., and J. Nurminen, "On analysis of eigenpitch in Mandarin Chinese," in *Proceedings of 2004 International Symposium on Chinese Spoken Language Processing*, 2004, Beijing, China, pp. 89-92.

Yuan, J., L. Shen, and F. Chen, "The acoustic realization of anger, fear, joy and sadness in Chinese," in *Proceedings of seventh International Conference on Spoken Language Processing,* 2002, Denver, Colorado, USA, pp. 2025–2028.

Zhao, L., C. Jiang, C. Zou, and Z. Wu, "A study on Emotional Feature Analysis and Recognition in Speech," *Acta Electronica Sinica*, 32(4), 2004, pp. 606-609.

Zhao, S., J. Tao, and H. Cai, "Rule-learning Based Prosodic Structure Prediction," *Journal of Chinese Information Processing*, 16(5) , 2002, pp. 30-37.

Zhao, Y., *Problems of Language*, Commercial Press of China, Beijing, 1980.