Investigation on pleasure related acoustic features of affective speech*

Dandan Cui, Lianhong Cai, Yongxin Wang, Xiaozhou Zhang

Key Laboratory of Pervasive Computing (Tsinghua University), Ministry of Education
Beijing 100084, P.R.China

{cuidd02}@mails.tsinghua.edu.cn
{clh-dcs}@tsinghua.edu.cn

Abstract. This paper presents our recent work on the investigation of affective speech along the pleasure-displeasure (P) dimension in the PAD emotional space. First, 76 conventional features are taken as candidates. And a novel feature, F0 Dominant Ratio, which is designed to reflect the dominant pitch, is also introduced. Correlative coefficients are calculated and a preliminary selection is done. Factor analysis and co-clustering are then employed. After another two rounds of selection, a set of 5 correlative features is selected. It includes Spectral Roll off, Spectral Low-high Ratio, Average F0, Min. F0, and the F0 Dominant Ratio, which can illustrate the acoustic differences between arousal-equaled emotion categories. Then a conversion experiment is carried out. Our model is built through stepwise linear regressions, and modifications are implemented carefully. Perceptual test shows that the conversion result of pleasure intensity is acceptable. Among the 5 features, the spectral ones show the highest favor. The dominant pitch appears relative, yet it requires further investigation. Meanwhile, intensity of arousal (A) in these sentences is agreed to be unchanged, i.e. the pleasure-oriented features are separated nicely from the arousal-oriented ones.

1 Introduction

Affective computing of speech is extensively attractive in human-computer interaction [1]. It is very appealing if the computer can recognize emotions in speech signal or synthesize speech with affect. Particularly, in recent years, with the overwhelming development of Internet and mobile applications, the conversion of emotion in speech has become more and more desirable.

Emotion can be described either in discrete categories or in continuous dimensions. Conventional approaches of affective speech analysis and processing are mostly based on the former, e.g., most commonly, “the Big Six” (fear, anger, sadness,

* The work in this paper is supported by National Science Foundation of China (No. 60433030, No. 60418012) and the Special Funds for Major State Basic Research Program of China (973 Program) (No. 2006CB303101).

happiness, disgust, and surprise) [1]. In fact, continuous dimensions apparently exceed because:

- They are more similar with the nature of human affect. When people feel or express some emotion, we don't just choose one from happy, sad, angry or surprise, etc., but a little bit active, a little bit happy, a little bit dominant, and so on.
- They are more suitable for computing. Emotional states are denoted by points in the multi-dimension space, and can be computed just using their ordinates.
- They can express much more. In the continuous multi-dimensional space, many that can not be illustrated by categories are clearly illustrated, such as the relative positions between emotions, subtle or mixed emotions.

Thus, we use the PAD space to represent emotions in our approach. PAD Emotion Model is a 3-D emotion model getting increasing applications in human-computer interaction. It has three nearly independent continuous dimensions: Pleasure-Displeasure (P), Arousal-Nonarousal (A), and Dominance-Submissiveness (D) as stated in literature [7]. Emotion states can be denoted as points in the space.

Unfortunately, in spite of their advantages, the analysis and processing of speech on continuous dimensions are truly much more difficult. Here follow the major barriers:

- Modeling in continuous and multi-dimensional space is much more complex than merely a classification problem.
- Corpus building is really difficult. The collection of effective and sufficient data can hardly be covered. And the authority of rating for emotional ordinates still requires improving.
- Moreover, what are the acoustic correlates of emotional dimensions in speech is still a problem. For the widely agreed two basic emotional dimensions, "arousal" ("activity") and "valence" ("evaluation", "pleasure"), most acoustic features relate to the former. Although human can judge the pleasure or displeasure in speech easily, their acoustic correlates remain confusing as literature [6][10] have stated, and even so for neuroscience. The situation of D dimension is still more blurry.

Yet, its superiority is so attractive that there are still consecutive efforts made on the multi-dimensional space. Particularly, novel approaches and findings on pleasure-related speech analysis are published recently, e.g. literature [2] and [3], inspired by musical and biological sciences separately.

Meanwhile, for modification, algorithms used to be barely satisfied, but they have been improved so that the modification are feasible now as stated in literature [4] [5].

Thanks to all the efforts above, we challenge the conversion of affective speech with respect to the continuous pleasure dimension. The approach is a three stage process:

1. Feature selection. In most cases of affective speech analysis, conventional acoustic features are used without paying much attention to their selection. Whereas in our approach we take it of great importance because finding the acoustic correlates of pleasure intensity in speech itself is a challenging task, and is crucial for the later stages as well.
2. Modeling. A model is built based on our data, through which the value of the acoustic features for target speech can be predicted according to its pleasure ordinate.

3. Modification. The speech is modified according to the values of its original target features. Modifications of timing, pitch, and spectrum are carried out separately, where different algorithms are employed.

This paper focuses on the first stage of the process: feature selection. First, our corpus and the statistical, data-mining methods employed in our approach are introduced in section 2. In section 3, numerous features are investigated and selected using those methods, then discussed. Features in the selected set are then modeled, and modified. An experiment is carried out in section 4, and the result of subjective evaluation is presented. Finally, in section 5, we state the conclusion and future work.

2 Corpus and methods

Before the process, it is necessary to introduce the corpus on which the study is carried out, and the methods that are employed. They are in the following subsections.

2.1 Corpus

First, ten emotional categories that are typical and can cover all the octants of PAD space are selected. They are: exuberant, relaxed, docile, disdainful, disgusted, angry, fearful, anxious, surprised, and sad. Along with the neutral, we have 11 emotional categories.

Then 10 passages under certain situation are designed for each category, and each passage contains 100 syllables or so. In every passage, we embedded a sentence that is emotionally unbiased and common for all the 11 emotions. These passages are read or recited by 20 college students, including 10 boys and 10 girls. Thus, a data set of 2200 passages and accordingly 2200 emotionally unbiased sentence are obtained.

Boundaries of prosodic constituents are annotated, together with the F0 contours. Besides the emotional categories, PAD values are also scored using a reliable and valid method introduced in literature [14]. More details can be found in a previous paper [9].

Moreover, in order to find the subtle acoustic cues for pleasure changes in speech, a lot of purification and normalization need to be done before analysis: to delete the wild samples, to eliminate the effect of different text content, different recording days, etc. Finally, the study is carried out on the emotionally unbiased sentences from one boy. Allowing for the phenomena in literature [11], each sentence is an intonation phrase and the real total number is 124. Based on these sentences, features are exacted and the 3 methods below are implemented.

2.2 Correlative coefficient

The most common method to investigate the linear correlation between variables, is employed in our approach as a general metric of correlations ($R(i,j)$) between acoustic features (i) and PAD ordinates (j) as can be illustrated by (1).

$$R(i,j) = \frac{C(i,j)}{\sqrt{C(i,i)C(j,j)}} \tag{1}$$

2.3 Factor analysis

The statistical method initially used in psychological researches, and now is quite popular in statistical analysis. It can simplify the complex interrelations among the numerous observed variables (x), by finding a relatively small number of common factors (f), as illustrated in literature [12].

$$x = \mu + \Lambda f + \varepsilon \tag{2}$$

This method is employed to find the emotion-dimension-related common factors in speech. For that purpose, the number of common factors is set to be 3 with respect to our 3-dimension emotional space. The factor loadings matrix (Λ) is computed through the maximum likelihood estimate (MLE), which presents the contribution of the features to the common factors.

2.4 Co-clustering

It is a fairly new clustering method. Unlike the conventional methods, it can group the features ($E = \{e_1,...,e_M\}$) and samples ($S = \{s_1,...,s_N\}$) simultaneously according to the mutual information entropies ($I(S;E)$) between them [13]. Thus, not only the sample clustering is optimized, but the correlations between features are also mined out.

$$I(S;E) = \sum_s \sum_e p(s,e) \log_2 \frac{p(s,e)}{p(s)p(e)} \tag{3}$$

Allowing for the intrinsic 11 emotional categories in our corpus, which cover the octant varieties of PAD space as section 2.1 has introduced, this method will help us find the grouping influences of acoustic features during the clustering of emotional categories. The number of sample clusters is specified as 11, while 3 for the features.

3 Feature selection

Then, with the help of those 3 methods, a feature selection procedure is carried out on the corpus, from initial candidate feature set, through several intermediate statuses, to a final selected set. As Fig. 1 shows, first, correlative coefficients between the initial candidate features and the utterances' PAD values are calculated, and a 37-feature set is selected preliminarily. After that, co-clustering and factor analysis are employed. A crucial set with 15 features turns out. Then a second round of factor analysis and co-clustering is implemented on the 15-feature set, by which we get a final set with 5 features.

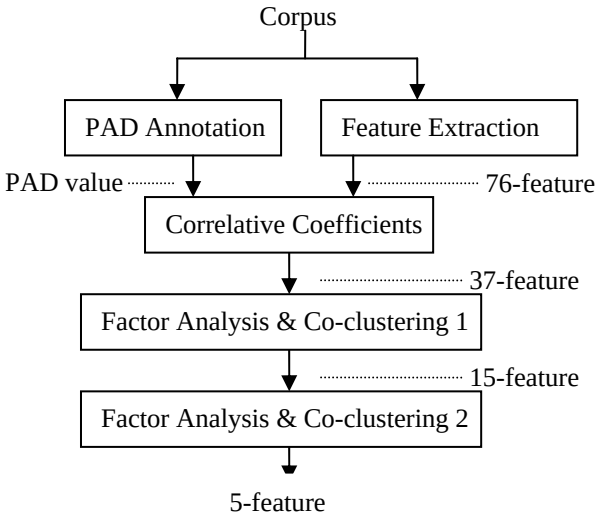

Fig. 1. Process for feature selection

Additionally, in the work presented in this paper, we focus on the universal features of utterance (intonation phrase) for the consideration of:

- Emotion is a more global and long-termed variation rather than the local micro-prosody employed in a common-sensed TTS.
- Local modification is dangerous especially for Mandarin as a tonal language. And
- Context may matter, yet limited by the current scale of data set, it can hardly be modeled [5]. We will investigate it afterwards.

3.1 Initial set of candidate features

In researches on affective speech, numerous features have been involved, which can be classified into prosodic and spectral ones [8]. Prosodic features such as F0, duration and energy are used most often. Several spectral ones have also been employed, which have been found to be distinctive for anger and joy in literature [2]. In our initial set of candidate features, most of the conventional features and their possible transformations are included. They are seen in Table 1.

Among the 76 features, novel ones are also introduced. As it is found in literature [3] that the dominant distribution of pitches may reflect the valence in speech, to simplify, we define a feature named F0 Dominant Ratio that refers to:

$$DominantRatio(F0) = \frac{Average(F0) - Min.(F0)}{Range(F0)} \tag{4}$$

For the consideration of robust stability, ⑮ and ⑯ are also introduced. And similar features for syllable and pause durations are introduced analogically.

Table 1. Initial set of candidate features

Spectral features:	
① Average of Spectral Centroid	② Average of Spectral Roll off
③ Average of Spectral Slope	④ Average of Spectral Low-high Ratio
⑤ Average of Spectral Flux	⑥ Average of Band Periodicity
⑦ Average of Band Periodicity (2000Hz~4000Hz)	
⑧… ⑳ Average of the 1-13 order MFCC features	
㉑… ㊵ Average of the first differences of feature①… ⑳ (absolute)	
Energy:	
㊶ Average of Short Term Energy	㊷ Standard deviation of Short Term Energy
F0:	
㊸ Average of F0	㊹ Standard derivation of F0
㊺ Max. F0	㊻ Min. F0
㊼ F0 range	㊽ F0 Dominant Ratio
㊾ Average of the first order differences of F0 (absolute)	
㊿ Standard deviation of the first order differences of F0	
⑱, ⑲ Slope and intercept of F0	
㉔ Median F0	㉕ Average F0/Median F0
㉖ (Average F0-Min. F0)/ (Median F0-Min. F0)	
Syllable duration:	
㉑ Average of syllable durations	㉒ Standard derivation of syllable durations
㉓ Max. syllable duration	㉔ Min. syllable duration
㉕ Syllable duration range	㉖ Syllable duration Dominant Ratio
㉗ Average of the first order differences of syllable durations (absolute)	
㉘ Standard deviation of the first order differences of syllable durations	
⑳, ㉑ Slope and intercept of syllable duration	
Pause duration:	
㉙ Average of pause durations	㉚ Standard derivation of pause durations
㉛ Max. pause duration	㉜ Min. pause duration
㉝ Pause duration range	㉞ Pause duration Dominant Ratio
㉟ Average of the first order differences of pause durations (absolute)	
㊱ Standard deviation of the first order differences of pause durations	
㉒, ㉓ Slope and intercept of pause duration	
㉗ Ratio of syllable duration to total utterance length	

Then all the features are calculated and normalized. The following computing is implemented on the z-scores for which segmental dependence is eliminated.

3.2 Feature selection round by round

As we see, the number of candidate feature is large which may wrack the clustering or factor analysis. Therefore, the selection is carried out round by round as Fig. 1 shows.

First, selection 1 calculates the correlative coefficients between the initial candidate features and the utterances' PAD values. Thus, features with higher correlative coefficients with P than other features and those with the highest for P among P, A, D are preferred. And a 37-feature set is selected preliminarily.

Then, factor analysis and co-clustering are employed in selection 2, from the point of view of emotion-dimension-related common factors and taking advantage of the intrinsic octant-covering categories separately. Comprehending their results, the number of features is reduced to 15.

15 is still a bit large. Therefore a second round of factor analysis and co-clustering is implemented on the 15-feature set as selection 3, by which we get a final set with 5 features. The result (loading matrix) of final factor analysis is showed in Table 2. For the co-clustering result, P, A, and D are in separate groups, while feature ④ and ⑨ are in the same group with P.

Table 2. The result of final factor analysis on the 15-feature set

Factor 1	Factor 2	Factor 3	Features
0.89543	-0.04424	0.43671	①
0.50814	-0.09254	**0.78997**	②
0.96255	-0.06584	0.21523	③
-0.078078	0.031425	**-0.78947**	④
0.93921	0.070619	0.12572	⑤
-0.92755	-0.00374	-0.33066	⑨
0.36317	-0.22326	0.12811	⑩
0.86138	-0.02188	0.018139	⑪
0.9715	0.015062	0.061754	⑬
0.8564	-0.01767	-0.09182	⑯
-0.10915	-0.0013	**0.2813**	⑱.
0.045499	**0.98223**	-0.04645	⑳
-0.006219	**-0.99079**	0.00662	㉑
0.19521	-0.0567	-0.02901	㉓
0.96265	0.021872	0.061001	㉔
-0.26857	0.20011	**-0.4571**	P
0.8336	0.19723	-0.04077	A
0.0076053	0.032609	**0.25862**	D

3.3 The result of selection

Finally, the 3 features that are in the same factor with P are selected, they are

② Average of Spectral Roll off (Rolloff)

④ Average of Spectral Low-high Ratio (LhRatio)

⑱ F0 Dominant Ratio (DominantRatio)

Allowing for the difficulty of modification and correlations between the selected features, feature ⑨ is not included in the final set.

Although they are not in the same factor with P, the following two features are also selected because of their common attendance in speech analysis.

⑬ Average of F0 (MeanF0)

㊻ Min. F0 (MinF0)

Thus a final 5-feature set is selected.

As it presents, the F0 Dominant Ratio does show correlation with P as we desired. Specially, in every round of co-clustering and factor analysis, A and P are in different groups or factors, which proves the independent function of A and P. Meanwhile, P and D are in the same group or factor in some cases, e.g. table 2.

For the correlative coefficients with P, most spectral features have higher values than the prosodic ones, just in line with what is found in literature [2]. Additionally, we eliminated some features with relatively significant correlations yet with no clear physical meanings, e.g. the 4th and 9th order MFCC features.

3.4 Discussions

Now back to the 11 categories recorded in our corpus as introduced in section 2.1, relaxed, docile, surprised, exuberant, disdainful, disgusted, fearful, sad, anxious, and angry, let's reinvestigate the relativities between them with respect to acoustic features and emotional ordinates.

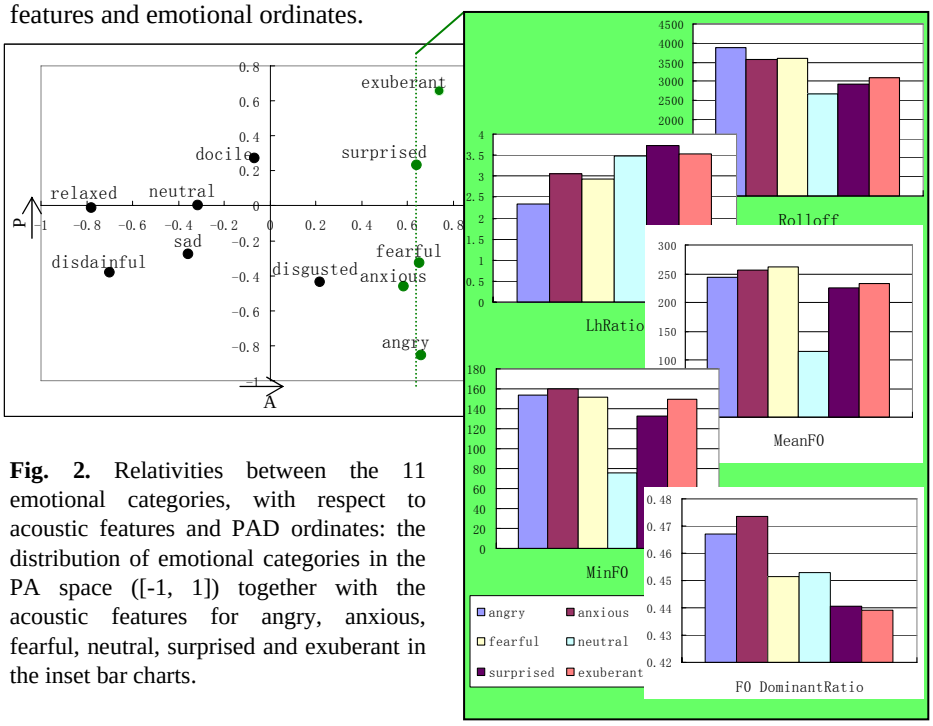

Fig. 2. Relativities between the 11 emotional categories, with respect to acoustic features and PAD ordinates: the distribution of emotional categories in the PA space ([-1, 1]) together with the acoustic features for angry, anxious, fearful, neutral, surprised and exuberant in the inset bar charts.

The scatter plot in Fig. 2 shows the distribution of the 11 emotions in our corpus in a PA space normalized to [-1, 1] according to the average of real annotated PAD values. It is seen that the there are several emotions around the vertical dot line at the A value about 0.6 to 0.8. They are angry, anxious, fearful, surprised and exuberant, which have similar positive A value, while the P values vary over a wide range. Then

their acoustic features are compared in the inset bar charts together with those of the neutral emotion.

For the two spectral features, consistent with the above conclusions, both show obvious P effect, especially the LhRatio. Maybe the Rolloff is influenced a little by the A.

Inversely, meanF0 and minF0 shows little cue of P, but apparently influenced by the A. It implies that they mainly contributes to the A intensity of speech. While they may also function in the P intensity, but the operation must be different from what they do with A.

The DominantRatio is somewhat strange. It shows the clearest difference between opposite, negative and neutral P emotions, and is beautifully independent of A, which is coincident with its performance in Table 2. But, the trend is contrary to what we have desired according to the statement in literature [3]. It may stem from the creaky voice that has been observed at the end of several sad sentences. Is it caused by our unsatisfactory quality of speech data? If not, is it unique for male? Or is it decided by some other dominant factor for Mandarin as a tonal language rather than what is found in the pitch accent one, Japanese, in the literature? A lot of further investigation is needed, and it worths while.

4 Experiment on speech conversion

Given the feature set, preliminary modeling and modification are tried. According to the target values predicted by the model, a neutral sentence is modified, and a perceptual test is conducted.

4.1 Modeling and modification

Since the correlations between P and the features may not be just linear and the concrete modality is not known, 6 elementary functions of P is chosen as the variables and a stepwise linear regression is implemented for each feature. Thereby a predict model is obtained. The 6 functions are as below:

$$P,\ P^2,\ exp(P),\ exp(-P),\ atan(P),\ atanh(P/2.5)$$

Modifications are carried out in two groups. Prosodic features are modified by TD-PSOLA algorithm. Then spectral features are modified by LP-PSOLA algorithm developed in our lab for simplification. Particularly, the features are integrated in advance, and the sequence of modifications is carefully arranged so that the conflict and artifact could be reduced as much as possible.

4.2 The experiment

The sentence is selected from the utterances those did not appear in the training set and is narrated by the same speaker. Context of the sentence is emotionally unbiased and the utterance is dominated by a single intonation contour.

Then it is modified by 3 degree towards displeasure (z-score -1), and 3 degree towards pleasure (z-score 1), thus 7 samples from displeasure to pleasure gradually. In order to compare the function of different features, the sentence is modified in 5 ways: with the spectral features, with the average and the min. F0, with spectral features and the average and the min. F0, with the average, the min., and the dominant ratio of F0, and finally with all the five features.

Thus, together with the original neutral utterance, there are 35 utterances. The perceptual test is then performed by 8 untrained listeners. The utterances are played in the 5 groups according to the different choices of modified features as mentioned above, and within each group the 7 samples are aligned with the displeasure-to-pleasure sequence. Subjects are asked to rate the difference between each current utterance and its previous one in term of pleasure-displeasure with a 5 degree ([-2, 2]) form. For the first sentence of each group, the absolute perception value is demanded.

At the end of this test, each subject will choose one group as his/her favorite.

4.3 Test results

Statistics of the result are shown in Table 3 and Table 4.

Table 3. The result of perceptual test, mean scores

Sample	Group1	Group 2	Group 3	Group 4	Group 5
1*	-1.25	-0.75	-1.75	-0.375	-1.5
2	0.125	-0.625	1	0	0.125
3	0.75	0.125	0.875	-0.375	0.375
4	1	-0.375	0.375	0.25	1
5	0.75	-0.5	0.875	-1.375	-1.125
6	1.125	-0.375	0.875	-0.125	0.125
7	1.375	0.25	0.75	0.625	1

Table 4. The result of perceptual test, confusion matrix in person number (pleasure by displeasure)

Sample	Group1	Group 2	Group 3	Group 4	Group 5
1*	0	0	0	0	0
2	2	4	0	0	0
3	0	2	0	4	1
4	0	5	2	1	0
5	0	4	0	7	7
6	0	2	0	1	0
7	0	0	0	0	0

*The score for sample 1 are according to absolute perception while others are relative with previous one.

For all the five groups, no confusion is found at the poles, especially with the negative pole. The features really influence. Yet, group 1 and group 3 are apparently superior, while reversely the other three groups are barely satisfied. Unfortunately, F0 Dominant Ratio shows very little help in the comparison between group 2 and group4. The F0 features perform badly, and even the group 5 is implicated. This may

arise from the disturbance in the model stem from wild point in the data set. Refining is needed in future work both for the data, and the modeling methods.

In line with the MOS and confusion matrix results, all the subjects are in favor of group 3 or group 1. All of them imply the importance of spectral features for pleasure in speech, which coincides with what literature [2] has found.

Additionally, it is reported informally that no subject feels the variety of arousal, i.e. the acoustic features correlative with the P dimension are separated successfully from the ones correlative with the A dimension.

5 Conclusions

As stated above, our recent work is the conversion of affective speech along the pleasure-displeasure (P) dimension in the PAD emotional space. The process is by 3 stages, and this paper focuses on the first: feature selection.

First, numerous conventional features are investigated. And a novel feature that presents the dominant pitch is also introduced. Correlative coefficients are calculated and a preliminary selection is done. Factor analysis and co-clustering are then employed. Finally, a set of 5 correlate features is selected which includes: Spectral Roll off, Spectral LhRatio, Average F0, Min. F0, and the F0 Dominant Ratio. These features can explain the differences between the speech arousal-equivalent emotion categories in our corpus. Modeling and modifications are also implemented. Therefore, perceptual experiment is able to be carried out.

The task is challenging, yet promising. Listening test shows that the conversion result of pleasure intensity is acceptable. Meanwhile, the arousal intensity of these sentences are agreed to be unchanged, i.e. the pleasure-oriented features are separated nicely from the arousal-oriented ones.

Among all the candidate features, most spectral ones show relatively significant correlates with P value. Among the 5 final features, the spectral ones also get highest favor.

The dominant pitch appears relevant to P in factor analysis, yet, unfortunately, it doesn’t perform as well in the modeling stage and the modification. A lot of further investigations are still required.

In all, the conversion of pleasure intensity and the multi-dimension emotion ordinates in speech is a long way to go, while our current model is just the first step. Corpus building is still a tremendous barrier. Richer and more efficient data are needed. The intra-sentence variance will be studied; tonal and intonational effects are to be concerned, which may take more responsibility for the harmony and melody in Mandarin than the non-tonal languages.

Besides the major conclusions, some interesting details are also noticed, and also worth further investigation in future work. E.g. some acoustic features of surprise speech are out of the trends; the duration features show relative to a separate factor in Table 2, which may have relations with findings of breath in a previous research [9].

Limited by the length of this paper, only feature selection is focused on. We will represent other details in a later paper, with successive improvements.

6 Acknowledgements

We would like to thank Professor Yi Xu for his kind advice. Thank Ling WEI from Institute of Psychology, CAS for her hard work on the PAD annotation of the corpus.

References

1. Rosalind W. Picard: Affective Computing. Cambridge, Mass.: MIT Press. 1997
2. Suthathip Chuenwattanapranithi, Yi Xu, Bundit Thipakorn, et al: Expressing anger and joy with the size code. In: Proc. of Speech Prosody, 2006: pp. 487-490
3. Norman D. Cook, Takashi X. Fujisawa, Kazuaki Takami: Evaluation of the affective valence of speech using pitch substructure. IEEE Trans. on Audio, Speech, and Language Processing, Vol. 14, No.1, Jan. 2006
4. Hideki Kawahara, Hisami Matsui. Auditory Morphing Based on an Elastic Perceptual Distance Metric in an Interference-free Time-frequency Representation. In: Proc. of ICASSP, 2003, vol. I: pp 256-259
5. Yongguo Kang, Jianhua Tao, Bo Xu: Applying pitch target model to convert F0 contour for expressive mandarin speech synthesis. In: Proc. of ICASSP 2006, vol. I: pp 733- 736
6. Ladd S, Silverman K, Bergmann G, et al: Evidence for independent function of intonation contour type, voice quality, and F0 in signaling speaker affect. J Acoust Soc Am, 1985, 78(2): pp. 435-444
7. Mehrabian, A.: Framework for a comprehensive description and measurement of emotional states. Genetic, Social, and General Psychology Monographs. 1995, 121: pp. 339-361
8. Cowie R, Douglas-Cowie E, Tsapatsoulis N, et al: Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine, 2001, 18(1): pp. 32-80.
9. Dandan Cui, Lianhong Cai: Acoustic and physiologic analysis of affective speech. In: Proc. of ICIC, 2006: pp. 912-917
10. Schröder, M., Cowie, R., Douglas-Cowie, E., et al: Acoustic correlates of emotion dimensions in view of speech synthesis. In: Proc. of Eurospeech, 2001: pp. 87–90
11. YANG Yufang, WANG Bei: Acoustic correlates of hierarchical prosodic boundary in Mandarin. In: Pro. of Speech Prosody, 2002: pp. 707-710
12. Harman, H. H., Modern Factor Analysis, 3rd Ed., University of Chicago Press, Chicago, 1976
13. I.S. Dhillon, S. Mallela, and D.S. Modha: Information Theoretic Co-clustering. In: Proc. of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington. DC, USA ,2003: pp. 89-98
14. Xiaoming Li, Haotian Zhou: The Reliability and Validity of the Chinese Version of Abbreviated PAD Emotion Scales. In: International Conference on Affective Computing and Intelligent Interaction (ACII), 2005: 513-518