Statistical analysis of acoustic characteristics of Tibetan Lhasa dialect speech emotion

Dandan Guo*, Hongzhi Yu, Axu Hu & Yanbing Ding
Key Lab of China's National Linguistic Information Technology Northwest University for Nationalities, Lanzhou, Gansu, China

ABSTRACT: The paper makes a quantitative analysis and comparison on the continuous speech emotion of Lhasa Tibetan in the four basic emotional patterns (happy, surprise, sad, neutral) pitch, energy and time length by experimental phonetics and the linear statistical research methods, found that there is a positive correlation between the Lhasa Tibetan emotional speech and pitch, energy and duration ,etc. And the pitch, energy and duration of negative emotion acoustic parameters are bigger than positive emotion, on this basis, drawing the Lhasa Tibetan speech emotion acoustic feature patterns. Compared with the Chinese language and the Tibetan, even though both have the tone prosodic features, they also have significant differences in the acoustic characteristics of the speech emotion.

Keywords: Tibetan Lhasa Dialect; Speech Emotion; Acoustic Characteristics

1 INTRODUCTION

Language is the most important tool for human communication, and it is also the main medium of emotion expression. In human language, in addition to the information of the text symbols, it also contains information on peoples' emotion and attitude change, and these changes are mainly reflected by voice. The same speech may have different meaning and communicative function due to the change of the emotional attitude and so on. On the contrary, people can communicate their emotions through the change of mood, intonation and so on. A lot of research has found that the prosodic of the speech signal features (fundamental frequency, energy, speed, etc.) is the most important factor to effect the speech emotion expression. Therefore, it is of great significance and value to study the emotion information of speech through the analysis of the acoustic features.

In the international world, Arnott and Murray have obtained the conclusion, shown in Table 1, on the corresponding relationship between different emotion and speech features [1]. Jiahong Yuan and Jianhua Tao in China have obtained some qualitative conclusions about the emotional acoustic features of the Chinese language through experiments [2].

Table 1. Corresponding relationships between different emotion and speech features

Characteristic	happy	fear	angry	sad	hate
Speed	fast or slow	soon	Slightly faster	Slightly lower	Especially fast
Average fundamental frequency	very high	especially high	especially high	slightly lower	especially low
Pitch range	very wide	very wide	very wide	slightly wide	slightly wide
Intensity	high	normal	high	low	low
Pitch change	upward bending, smooth	normal	Stress point mutation	downward bending	wide, tail down

The emotional information is social, regional and cultural, that is different cultures from different regions of different countries in different ways to express feelings. Therefore, there are many problems: Are the acoustic features of the emotional speech in Mandarin Chinese also present in the minority language? Is the language of emotion recognition need to be combined with the actual cultural background? The paper regard the Lhasa Tibetan as the research object, using linear statistical method by extracting the acous-

*Corresponding author: 920233231@qq.com

DOI: 10.1051/shsconf/20162501017

tic characteristics of the language of speech emotion to solve the above problems and provides the basis for general significance. The same as Mandarin, Tibetan Lhasa dialect also belongs to a tonal language [3][4]. It also has segmental and suprasegmental features, which both play important role in the expression of emotional speech. Therefore the paper mainly study Lhasa Tibetan emotional speech features through investigating pitch, energy and duration time of sound.

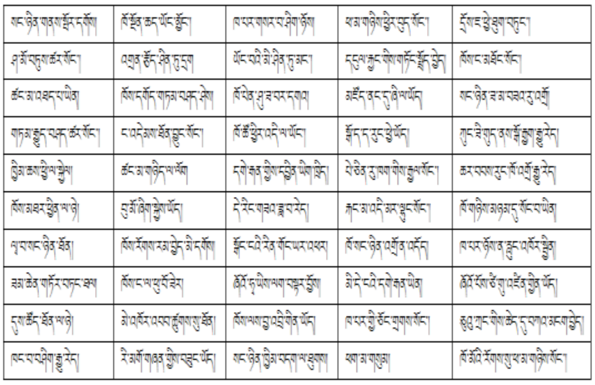
2 EMOTIONAL SPEECH FEATURE COLLECTION

2.1 *Emotional speech Collecting method*

Emotional speech can be divided into three kinds of imitation speech in accordance with the different ways of the acquisition, mimic speech, induced speech and natural voice. The experiment adopts the way of setting up the scene induced speech to guarantee the authenticity of the emotion expression. In particular, the recording text has a high degree of emotional freedom, which can be directly related to different emotions. Specifically, every sentence in the text literal are neutral (leave the context cannot certain the intrinsic emotion), but associated with different context naturally arouse different feelings. Human emotion is complex and changeable, so there is no certain theory to classify the emotion yet. This paper mainly investigates the acoustic characteristics of Tibetan Lhasa dialect under the happy, surprised, sad and neutral emotion.

2.2 *Recording text design*

Table 2. Recording text



གང་ཞིན་གནས་སྐོར་དགོས།	ཁོ་སྟོན་ཆང་ཡོང་རྒྱུང་།	ཁ་པར་གསར་པ་ཞིག་ཉོས།	ཕ་མ་གཉིས་ཕྱིར་ཕྱུར་སོང་།	རྫོགས་ཟ་ཕྱེ་ཐུག་བཏུང་།
ཉ་མོ་བཏུག་ཚར་སོང་།	འགྲན་རྩོད་ཤིན་ཏུ་དྲག	ཡོང་བའི་མི་ཤིན་ཏུ་མང་།	དངུལ་རྒྱུང་ཡིས་གཏོང་སྐྱོང་ཕྱེར།	ཁོས་ང་མཐོང་སོང་།
ཚང་མ་འཐད་པ་ཡིན།	ཁོས་དགོད་གཏམ་བཤད་ཤེས།	ཁོ་པེན་ཤུ་ཟ་བར་དགའ།	མཛོད་ནང་དུ་ཞི་ལ་ཡོད།	གང་ཞིན་ཟ་མ་བཟའ་རུ་འགྲོ།
གཏམ་རྒྱུད་བཤད་ཚར་སོང་།	ང་འདེམས་ཐོན་བྱུང་སོང་།	ཁོ་ཚོ་ཕྱིར་འདི་ལ་ཡོང་།	སྦོད་ད་ཙུང་ཕྱེ་ཡོད།	ཀུང་ཟེ་ཤུད་ནས་སྒོ་རྒྱག་རྒྱུ་རེད།
ཁྱིམ་ཆས་ཕྱིལ་སྐྱེས།	ཚང་མ་གཉིད་ལ་སོང།	དགེ་རྒན་གྱིས་དབྱིན་ཡིག་ཕྲེད།	པེ་ཅིན་ཏུ་ཡག་གིས་རྒྱལ་སོང་།	ཆར་པས་ཙུང་པོ་འགྲོ་རྒྱུ་རེད།
ཁོས་མཐར་ཕྱིན་ལ་ཞེ།	བུ་མོ་ཞིག་སྐྱེས་ཡོད།	དེ་རིང་གཟའ་ཟླ་བ་རེད།	ཀང་མ་འདི་མར་ཕྱུང་སོང་།	ཁོ་གཉིས་མཉམ་དུ་སོང་བ་ཡིན།
ལྷུ་པ་གང་ཞིན་ཐོན།	ཁོས་རོགས་རམ་བྱེད་མི་དགོས།	སྐྲང་ངའི་རིན་གོང་ཡར་འཕར།	ཁོ་གང་ཞིན་འགྲོན་འདོད།	ཁ་པར་ཉོས་ན་རྒྱང་འཁོར་སྤྱིན།
ཟམ་ཆེན་གཏོར་བཏང་ཟས།	ཁོས་ང་ལ་ཕུ་བོ་ཟེར།	ཁོའི་ཏྲུ་ཡིས་ལག་བསྐྱར་རྫོགས།	མི་དེ་ངའི་དགེ་རྒན་ཡིན།	ཁོའི་རོགས་ཚོ་ཀུ་འཛིན་གྱིན་ཡོད།
དུས་ཆེན་ཐོན་ལ་ཞེ།	མི་འཁོར་འབབ་རྟུགས་སུ་ཐོན།	ཁོས་ལས་བྱ་འདྲི་གིན་ཡོད།	ཁ་པར་གྱི་ཚོང་གྲགས་སོང་།	ཧྲུའུ་ཀྲང་ཡིས་ཆེད་དུ་བཀའ་མང་གཏེར།
ཁང་པ་བཤིག་རྒྱུ་རེད།	རི་མགོ་གཞན་གྱིས་བཟུང་ཡོད།	གང་ཞིན་ཁྱིམ་བདག་ལ་ཟུགས།	ལག་མ་གསུམ།	ཁོ་མིའི་རོགས་སུ་ཕ་མ་གཉིས་སོང་།

Choosing the proper recording text is the primary premise to record. This study select the recording text according to the reference [5]: (1) Each recording scripts has a higher degree of freedom, and is suitable for joy, surprise, sadness and neutral emotional expression; (2) Recording script with semantic neutral and does not contain a kind of obvious emotional tendencies; (3) Recording script is oral declarative sentence, the content involves the daily life, covering a wide range; (4) Recording script speech collection can include main vowels and consonants of language speech; (5) Recording scripts are 5-8 word phrases. According to the above criteria, selecting 50 phrases as recording text, as shown in Table 2.

2.3 *Recording*

The recording is done in the professional voice laboratory with good airtight by Lenovo notebook computer with XP windows system and headphones. Recording software with Cool Edit Pro2.0, sampling rate is 22050hz, the sampling rate of Cool, single channel, 16 bit sampling accuracy, audio files are stored in wav format. And the recording is to save in the Cool Edit with taking a sample as a unit, while using Praat voice analysis software to analyze the recorded voice. In the experiment, four students of Lhasa Tibetan College students who are good at emotional expression were selected, two male and two female, aged about 20 years old.

Before recording, a designed context is provided according to the recording text to help stimulate the needed emotion. In corpus design, 4 kinds of emotions are considered, which are happy, surprise, sadness and neutral. Each emotion has a 50 sentence corpus which are read by these four subjects using the above 4 different emotions, a total of 800 statements. Each of the emotional states of the statement repeated for three times, and the recorders are required to maintain the same intensity of the emotion as much as possible. In order to maintain the stability of the emotion, the sentence of the same kind of emotion is arranged together. Finally let the recorder to confirm the adequacy of the emotional expression.

3 DATA ANALYSIS

At present, the research of emotion speech is mainly focused on the basic frequency, energy, duration and vowel resonance peak and so on. [6]. Because of the difference in the study, the acoustic characteristics extracted are different. The mainly acoustic parameters we analyzed including the following several aspects, and analyze statistically these acoustic parameters of the data.

3.1 *Fundamental frequency parameters analysis*

The pitch frequency is the frequency of vibration of the vocal cords when voiced sound，which is one of the important parameters of speech signal. In different emotional states, the same words with different characteristics of the fundamental frequency. Since the Tibetan Lhasa dialect is also a tonal language, the change of tone can convey a lot of information and has a great role on discourse expression. And the tone feature is mainly through the change mode of fundamental frequency. The research of emotional speech at home and abroad shows that the fundamental frequency (the maximum, minimum, mean, fundamental frequency curve, etc.) plays an important role in the study of emotional speech. In this paper, the analysis of the fundamental frequency is also a key point.

In consideration of the difference between male and female's vocal organs, we make a statistical analysis of the fundamental frequency of different emotion by averaging.

Table 3. Tibetan Lhasa dialect boy's frequency value of different emotion (Hz)

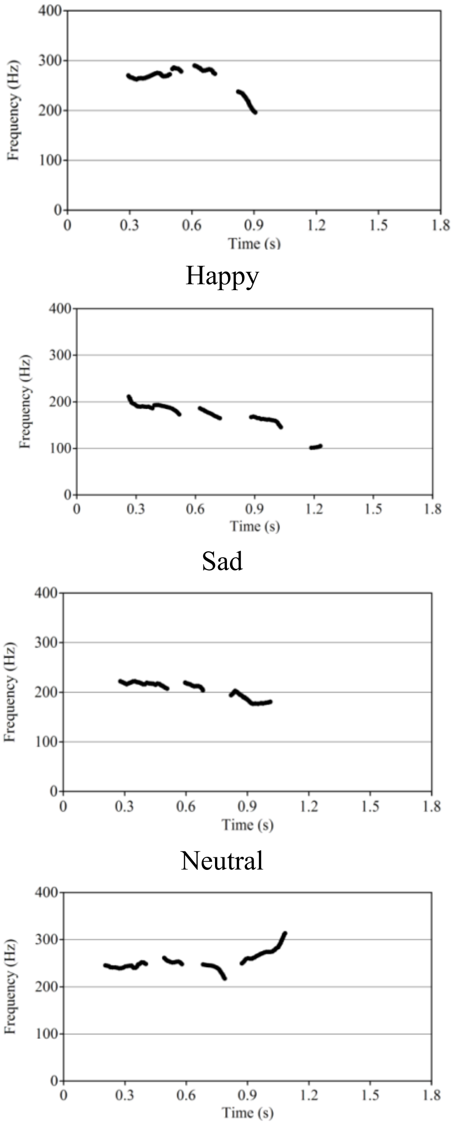
Emotion Type	Male（Hz）				
	Sentence-initial fundamental frequency	Sentence end fundamental frequency	Maximum fundamental frequency	Minimal fundamental frequency	Average fundamental frequency
happy	265.6	220.7	297.6	169.7	256.7
surprise	243.3	278.8	324.9	207.3	255.1
sad	191.1	126.0	211.2	99	171.3
neutral	217.3	180.5	227.6	100.3	194.2
average	229.3	201.5	265.3	144.1	219.3

Table 4. Tibetan Lhasa dialect girls' frequency value of different emotion (Hz)

Emotion Type	Female（Hz）				
	Sentence-initial fundamental frequency	Sentence end fundamental frequency	Maximum fundamental frequency	Minimal Fundamental frequency	Average Fundamental frequency
happy	543.6	481.6	635.9	263.7	480.2
surprise	504.9	572.1	535.9	457.3	492.6
sad	463.2	383.6	526.9	367.8	441.9
neutral	370.1	321.6	460.8	283.7	350.4
average	470.5	418.2	539.9	343.1	441.3

As can be seen from Table 3 and 4, the girl's fundamental frequency is obviously higher than the boys, and the girls' average fundamental frequency range of four emotions is 343.1Hz to 539.9Hz, while the boys' is 144.1Hz to 265.3Hz. From the physiological point of view, this is mainly because the girl's vocal cords compared to the boy's fine and long, so the sound frequency is higher than the boys. In addition, whether it is a boy or a girl, the basic fundamental frequency value of surprise is higher than other emotional types, happy emotion the second.

On the basis of the data parameters, the basic frequency curves of the sentences were extracted with Praat software, to better observe and compare the fundamental frequency of different emotions and their performance in the fundamental frequency curve. The fundamental frequency curve can reflect the change of the fundamental frequency of the whole sentence. The expression of different emotions is mapped by the sudden elevation or decrease of the fundamental frequency. Through the analysis of the fundamental frequency curve, we can find the difference between different emotions. Figure 1 is the same sentence in different emotional state of the fundamental frequency curve.



Happy

Sad

Neutral

Surprise

Figure 1. The same statement in different emotional state of the fundamental frequency curve

Figure 1 shows that the same sentence in different emotional state, the envelope of the fundamental frequency curve is different. The most significant change appeared in the beginning and end of a statement. Accordingly seen happy overall statement presents the trend of increasing at first and then decreasing; sad statement has been in a downward trend; at the end of the sentence of surprise statement is clear upward trend, as distinguished from the other emotional states sign characteristic; and the fundamental frequency curve of neutral statement, compared with other three emotional states, is slower.

3.2 *Energy parameter analysis*

Energy is another important feature of speech, which contains a wealth of emotional information; it will also change with the speaker's emotional state. The statistical information of the energy of the speech can be used as a significant basis for judging the emotional changes. For example, people speak louder when people are in high spirits. When the mood is low, the energy is lower, and the energy curve is relatively flat. Therefore, in the case of the same time domain, the energy and range of the four different emotions are extracted, which are shown in Table 5.

Table 5. Tibetan Lhasa dialect different emotional energy value

Emotion Type	happy		surprise		sad		neutral	
	male	female	male	female	male	female	male	female
Average energy	72.3	67.6	75.5	74.1	68.7	66.5	65.2	64.3
Maximum energy	82.6	78.2	84.1	80.3	77.2	74.8	76.7	72.8
minimal energy	33.7	30.8	31.3	20.8	28.5	27.7	30.2	30.6
Range value	49.9	47.3	52.8	49.6	48.7	47.0	46.6	42.3

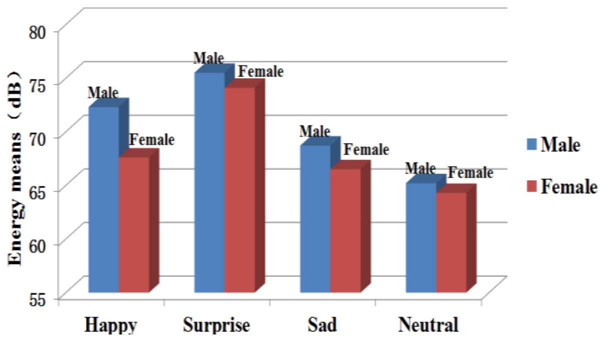


Figure 2. Average energy distribution of different emotions

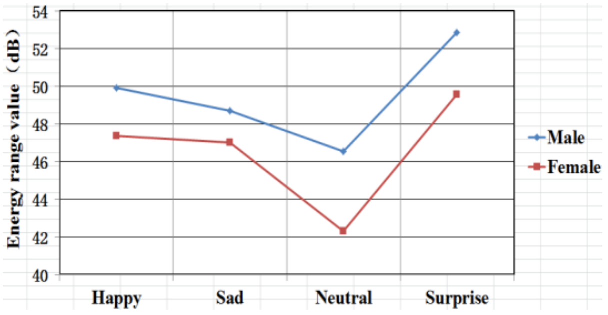


Figure 3. Dynamic ranges of different emotions

Figure 2 is the average energy distribution about the same text statement in different emotional state, from which can be seen, the lower the level of sadness is lower than neutral emotional state, that is, the energy level is low, but the energy of joy and surprise is higher, which is related to the intuitive feelings of people on the loudness of the sounds. According to the maximum and minimum values of the energy of different emotions in the same time domain, the energy dynamic range of different emotions is summarized, that is variation range of energy, which is shown in Figure 3. From this, we can see that the energy fluctuation in the excited state is larger, which reflects the fluctuation range of the energy curve. Followed by that of happy emotion, sad emotion is smaller, which is also the most easy to identify.

From a gender perspective, it is found that male and female students show obvious gender differences in the expression of the same kind of emotion, and it can be very good to explain that the male' voice is more stable than the female.

3.3 *Time parameter analysis*

The related features of the speech duration also contain the emotional prosodic information. In different emotional states, the speed is different, and presents the difference of time in acoustics. Therefore in the time parameters of the traditional emotional speech analysis, the speed was often regarded as an effective feature to distinct emotional state; the unit is bytes / sec. The paper counts the temporal structure characteristics of the sentence with the same text in different emotional states, including the length of the duration and the corresponding ratio of the length on quiet sound. As shown in Table 6. In addition, Figure 4 also gives the corresponding relationship between the duration and speed of different emotional states.

Table 6. Time structure characteristics of different emotional states

Emotion Type	happy		surprise		sad		neutral	
	male	female	male	female	male	female	male	female
Average time(s)	1.33	1.41	1.22	1.27	1.62	1.73	1.42	1.47
Speed	3.76	3.54	4.1	3.93	3.08	2.89	3.52	3.4
Speed ratio	1.07	1.04	1.16	1.15	0.88	0.85	1	1

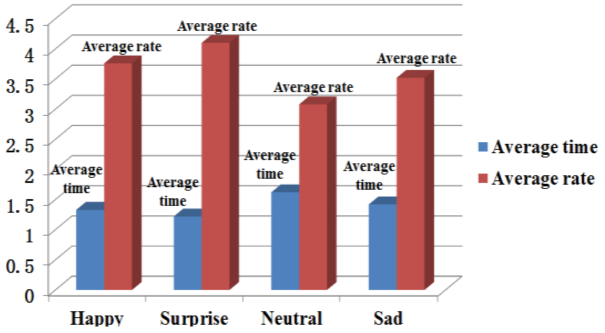


Figure 4. The corresponding relationship between different emotional state of long time and speed

By Table 6 and Figure 4, the sad state is the longest, followed by a happy, neutral, and surprise. And then the negative emotion of the time is significantly greater than the positive emotion. At the same time, the quantitative results verify the auditory perception of sad slow, happy light.

In the emotional state of surprise and happy, the speaker's speed will be faster when the pronunciation statement text word number is equal, therefore, people will get the feeling of euphoria of listening sense. In the sad emotional state, speed is slow, and people will feel deep and slow. And that due to the influence of the emotional state, duration and rate of speech showed negative correlation relationship. In addition, on the whole, in the expression of the same emotion, boys' speed is significantly faster than the girls.

4 DISCUSSION

4.1 *The corresponding relationship between acoustic parameters and emotion*

There is a positive correlation between the acoustic parameters of Tibetan Lhasa dialect and frequency, energy, time and so. Pitch, energy and duration of negative emotion are greater than positive emotion. From a cognitive point of view, this is because the negative emotion is more sensitive than the positive emotion in the human cognitive system.

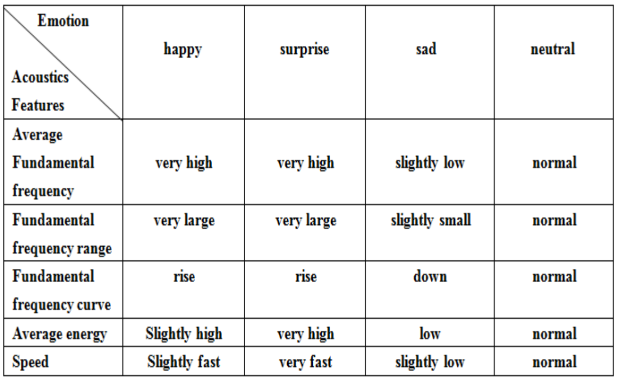
4.2 *Gender difference*

The research on the relation of Tibetan Lhasa and pitch, energy and duration show significant gender differences, that is girls' fundamental value was significantly higher than boys; and boys' energy values were significantly higher than those girls, speed also than girls' faster. This is because men and women' emotional expression is different, the girls are relatively more delicate, and the boys are more rough. In addition to the physiological mechanism of the expression of emotion, the female vocal cords were shorter and thinner, so the fundamental frequency of the students was higher than that of boys.

The results also show that the expression of emotion is not a single parameter at work, because of the difference of gender and language style, and there are also differences in the performance of the acoustic model.

4.3 *The acoustic mode of Tibetan speech*

Table 7. The Tibetan emotional acoustic features



Emotion / Acoustics Features	happy	surprise	sad	neutral
Average Fundamental frequency	very high	very high	slightly low	normal
Fundamental frequency range	very large	very large	slightly small	normal
Fundamental frequency curve	rise	rise	down	normal
Average energy	Slightly high	very high	low	normal
Speed	Slightly fast	very fast	slightly low	normal

4.4 *Comparison of the emotional speech acoustic characteristics of Tibetan Lhasa dialect and Mandarin*

Although the Tibetan and Chinese language both belong to Sino Tibetan Languages and have the tone, but compared with the Mandarin emotional speech patterns [6], when expressing the same emotional state, the energy value and fundamental value of Lhasa Tibetan are greater than Chinese language; speed than that of Chinese is faster, showing the obvious national character. This result is closely related to the unique culture and living background of each nation. Of course, this conclusion remains to be further expanded the number of corpus to verify.

5 CONCLUSION

The paper summarizes the pitch, energy and duration characteristics of Tibetan Lhasa dialect speech emotion by statistical analysis. Explaining many of the daily speech perception with experimental methods, and fully verify the relevant theoretical issues of traditional linguistics. But also for minority language phonetics study provides the acoustic model, provides the voice parameters for the recognition and synthesis of speech engineering, for cognitive affective study provides more precise classification basis of emotion, and expands the emotion cognition research fields from the perspective of cross culture.

6 ACKNOWLEDGEMENT

This project is subsidized by Nation-al Natural Science Foundation of China (61462075) - the Research of the Lhasa Tibetan Rhythm Cognitive Theory Based on EEG.

7 REFERENCES

[1] Murray I.R. & Arnott J.L. 1993. Toward a simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *Journal of the Acoustical Society of American*, 93(2): 1097-1108.

[2] Tao J. 2003. *Emotional Control of Chinese Speech Synthesis in Natural Environment*. Euro Speech.

[3] Kong Jiangping. 1995. Tibetan Lhasa dialect tone perception study. *The National Language*, 3.

[4] Zheng Yuling. 1998. Tibetan dialect quantitative analysis. *The national language*, 5.

[5] Boersma, P., Weenink, D. 2001. PRAAT, a system for doing phonetics by computer. *Glot International*, 5(9/10): 341-345.

[6] Zhao Li, Qian Xiangmin. 2000. Research on extracting the emotional characteristics from the speech signal, data acquisition and processing, 15(1): 3.