Research on Lhasa Tibetan Prosodic Model of Journalese Based on Respiratory Signal

Shuni Guo

Inner Mongolia University for Nationalities, Tongliao, China

Email: guo_shuni@126.com

Lu Gao, and Hongzhi Yu

Key Lab of China's National Linguistic Information Technology, Northwest University for Nationalities

Lanzhou, China

Email: Rabbit5835@163.com

***Abstract*—In accordance with the actual development for Tibetan speech synthesis, the paper has taken news text as training corpora, analyzed the speech and prosodic features of Tibetan Lhasa dialect and confirmed the respiratory signal parameters with prosodic features. It has confirmed 6 classes of 39 dimensions context feature parameters in terms of previous prosodic structure analysis results. It uses RBF neural network to establish prosodic model and output 10 dimensions prosodic control parameters, and testing to know the predictable nature of the established model.**

Keywords—Tibetan Lhasa dialect; Prosodic model; RBF; respiratory signal

I. INTRODUCTION

In order to eliminate the synthesized speech and human nature language differences between flow, and the synthesis with high degree of natural voice, it must establish a rhythm of the model of high quality. The rhythm model of the middle of the last century has many research results, such as: ⊖the United States established the German University of Idaho Natural Language Flow Changes Synthesis Model (2004); ⊜ the ATR Information Science Research Institute of Japan, and Nick Campbell studied the characteristics of the rhythm reflect Japanese conversation model (2004). Domestically, voice rhythm of the research object model mainly concentrated in the Chinese language, and also gained plenty of research results. Because of the recent technology of Tibetan speech, rhythm, the study of the model is deficient. Particularly there is no research report of high quality in rhythm modeling.

In addition, your chest and abdominal breathing signals produce two important physiological signals. Kong Jiangping Beijing University takes the lead in using breathing belt acquisition of bust or waist circumference values changes to study the rhythm of speech characteristics [1]. This article on the human respirator physical process analysis, designed a set of measurements applicable to research the Lhasa words of Tibetan breathing signal parameters, and these parameters will be used as the training data model established rhythm, and link the speech multimodal with computer words link project.

II. CORPUS

The corpus used in this article was chosen from 2007 in the *Tibet Daily*, and reflects the importance of the syntax, the Tibetan intonation and rhythm, especially intonation information structure. At the same time, it covers the language syntax structure and context information and may be an important role for prosodic feature of factors. With reference to the sentence pattern classification, consider Tibetan sentence length, both before and after the sound league match syllables, syllable combination of syllables and position in natural language flow appeared in frequency, we chose sentences containing a combination of continuous tone sentences – 2000 words in total.

III. WORDS BREATHING SIGNALS

At present, the focus of the speech discipline research has produced by the voice of the acoustic and physiological speech production to biological mechanism. The multidisciplinary research as a whole is the embodiment of "multi-modal of speech". Words are a breath of motive mechanism, basic words are one of the physiological processes. The signals in the chest and abdomen when breathing are two important physiological signals that are produced in the process of pronunciation.

A. Breathing Signal Acquisition

Collection is through the use of MLT1132 breath when the main sensor to breathe with chest expansion of change. Contract the two root respiratory respectively with sensors in the pronunciation of the chest and abdomen, breathing with sensors will bust

978-1-61284-881-5/11/$26.00 ©2011 IEEE

and waist circumference contraction caused by the expansion with length change by breathing equipment into electrical signals piezoelectric chest, to get the breathing rhythm signal.

B. Breathing Signal Parameters Settings

Through the analysis of the human respiratory physiological processes, this will be a complete breathing process definition for Breath Cell. In a breath element, there are two parameters defined: Inspiration Cell and Expiration Cell, corresponding to a breathing process consists of two sub-processes: breathe in and out.

For breathing signals, key research has two variables: duration, said a breath on the duration of the process, and amplitude signals, the extent of change. For every process, we will study six parameters: breath cell duration, breath cell amplitude, expiration duration, expiration amplitude, inspiration duration, and inspiration amplitude [2], as shown in Figure 1 below.

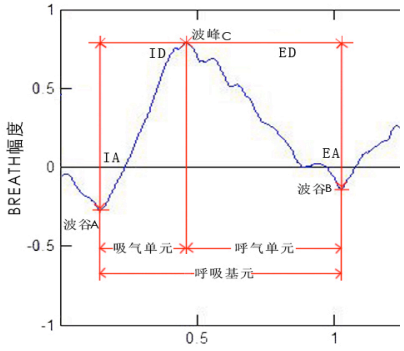

Figure 1. Breathing signal parameter diagram

Note: ID is inspiration duration; ED is inspiration duration; IA is expiration amplitude; EA is inspiration amplitude.

Through the above, we can get the following: breath cell duration (BD) = ID + ED; breath cell amplitude (BA) = IA + EA.

C. Breathing Signals and Prosodic Feature Words of Contact

1) Level 3 Breathe the Definition of the Unit

The 2000 words of corpus news collection of breathing sounds chest signal. With each place breathe in place for breathing reset, and through the definition of automatic tagging analysis platform, we get 560 breathing reset places. 560 breathing reset duration and reset amplitude in SPSS16.0. We draw the histogram, as shown in Figure 2 and 3.

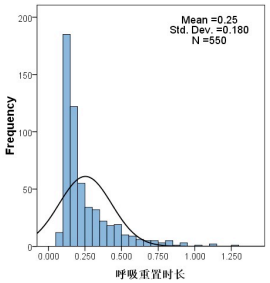

Figure 2. Reset Duration Distribution

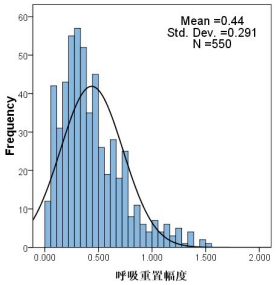

Figure 3. Reset Amplitude Distribution

By resetting the size of the breathing unit by different increments, we can be divided into three types of language flow breathing units: big breathing unit, breathing unit and small breathing unit. Three levels of breathing units based on different increments respectively and the size of the corresponding semantic unit are shown in Figure 4.

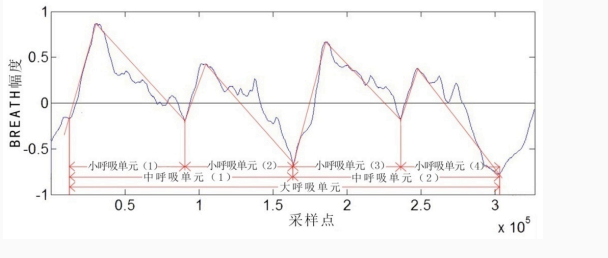

Figure 4. Breathing unit schemes

2) The Relationship of Breathing Unit and Rhythm Unit

This paper chose the corpus from the newspaper, and the structure and format of the news are relatively fixed, can form the fixed breathing pattern, and show the rhythm of the special features. Before the 560 mark for a breathing reset in analysis, respectively in each syllable, statistical reset a Chinese word and phrase number and its duration, rhythm that unit out data, breathe in rhythm as is shown in Table 1.

TABLE 1. UNIT OUT OF RHYTHM. BREATHE IN DATA

Breath Cell		**Level 1**	**Level 2**	**Level 3**
Breathing reset duration (second)	Average	0.78	0.33	0.19
	sd	0.18	0.17	0.09
Breathing reset amplitude	Average	0.75	0.68	0.39
	sd	0.18	0.17	0.09
Rhythm unit number (average)	Syllables	80	20	7
	Words	31	7	3
	Phrases	15	4	1
	Sentences	3	1	0
	Section	1	0	0

The data in Table 1 demonstrates the following conclusion:

a. When breathing unit higher levels, this unit contains the longer duration and the greater amplitude. When the breath of the unit is at a relatively low level, the various data is relatively stable.

b. In this level 3 breathing unit, unit at all levels contained in the duration and amplitude breathe in unit of change are positively correlated.

3) The Relationship Expiratory Unit and Rhythm Unit

Expiratory unit in the corresponding waveform is speech section, and its level is the same as in the level of the Inspiration unit before. We get the data of extraction unit and statistics, as shown in Figure 5.

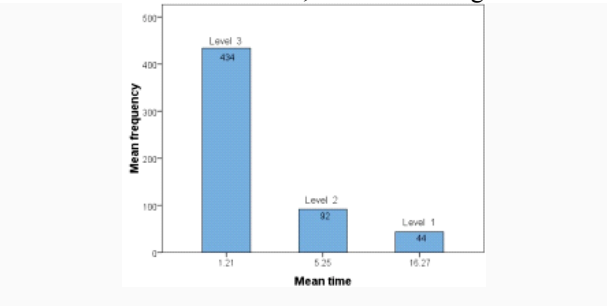

Figure 5. Expiratory unit duration data statistics

The data analysis of the available information:

a. The higher level of breath cell, it appears the less number, reflecting many even less citation characteristics in read pause.

b. The higher level of breath cell, the longer time unit, the corresponding voice of the duration for long.

4) Breathing and the Relationship Between the Pause

In human nature language, a section of the flow of the pronunciation of physiological needs is a pause. Since most of the language is an expiration sound, so humans exhale as you exhale only for pronunciation. Because lung capacity are limited and not able to reel off a long words, especially in the speech process to take a breath, so need always to breathe in pause, can produce the reflected in waveform is silent section.

IV. THE RBF NEUTRAL NETWORK MODEL RHYTHM

Radial basis function RBF neural network (hereinafter referred to as RBF network) first is in the real changeful interpolation problem, by introducing theory that J. Moody and C. Darken put forward in the 1980s. It is a single hidden layer of three layers, network feed forward with arbitrary precision can approximate any continuous function. The RBF network input and output are two vector data sets. The initial input vector is the influence of prosodic feature parameters of contextual information, and the output target is prosodic feature vector control parameters.

A. Context Information Parameters

Prosodic features in different context will have a lot of reflection, context information parameters and the prosodic features are closely linked. Tibetan Lhasa words, according to the context information, the context of the text information and the Tibetan prosodic feature different levels of influence, will the prosodic word syllable–according to the rhythm phrases-statements from–to the level of the unit cell. In this rhythm model, the parameters by the current context information (C), before adjacent syllables information (L), after the adjacent syllables information (N), the information of the syllable in the prosodic word (W), the information of the syllable in prosodic phrases (P) and syllable information in the statement (S).

B. Speech Signal Context Information Parameters

Tibetan syllables duration, how pronunciation weight factors such as express and pauses with rhythm the role of information. Pitch and duration, etc. should be able to reflect the acoustic parameters prosodic feature; at the same time the syllables of Tibetan between the strong co-articulation phenomenon, language flow of adjacent syllable these acoustic variation of parameters is the effect of synthetic speech nature degree of important factors [3].

C. Breathing Signal Contextual Information Parameters

Breathing signals to breathe in the amplitude of the unit in different reflect the corresponding language flow subordinate followed by the different level rhythm; Breathe out the slope of the unit to show the rate of breathing, which can affect a pitch value. In the original context information parameters in join the current unit slope and exhale syllables in range, before unit adjacent syllables expiratory unit slope and inspiration cell amplitude in unit of 11 D breathing signal amplitude of the parameters of the Tibetan, will speech rhythm contextual information model parameters set extension for six groups 39 d parameters, as shown in Table 2.

TABLE 2. CONTEXT INFORMATION PARAMETER SET

The current syllable information C	Syllable duration	C1
	Initials type	C2
	Finals type	C3
	Tone type	C4
	Stress type	C5
	Position in prosodic word	C6
	Coupling with former syllable	C7
	Coupling with after syllable	C8
	Pitch reset with former syllable	C9
	Pitch reset with after syllable	C10
	The distance with former stress syllable	C11
	The distance with after stress syllable	C12
	Expiration cell signal slope	C13
	Inspiration cell signal amplitude	C14
Before the adjacent information L	Syllable duration	L1
	Finals type	L2
	Tone type	L3
	Stress type	L4
	Expiration cell signal slope	L5
	Inspiration cell signal amplitude	L6
After the adjacent syllable information N	Syllable duration	N1
	Initials type	N2
	Tone type	N3
	Stress type	N4
	Expiration cell signal slope	N5
	Inspiration cell signal amplitude	N6
The information of the syllable in the prosodic word W	Syllables	W1
	Position in prosodic word	W2
	Expiration cell signal slope	W3
	Inspiration cell signal amplitude	W4
The information of the syllable in prosodic phrases P	Syllables	P1
	The prosodic word	P2
	Position in sentence	P3
	Expiration cell signal slope	P4
	Inspiration cell signal amplitude	P5
Syllable information in the statement S	Syllables	S1
	The prosodic words	S2
	The prosodic phrases	S3
	Breath cell	S4

D. Rhythm Control Parameters

Rhythm control parameters should be able to reflect the corresponding to the text in the rhythm of your properties. Natural language the rhythm of the main cash flow characteristics for the tone of voice, duration and stress, because the pitch and duration are two of the most important characteristics of synthetic speech rhythms, the perception of the nature degree has the important meaning, so choose the pitch and duration syllables as prosodic feature control parameters.

Rhythm control parameters mainly include the following parameters: the position of syllables starting A1, the position of syllables ending A2, the position of pitch starting N1, the position of pitch ending N2, maximum pitch position N3, minimum pitch position N4, the initial pitch value B, ending pitch value F, maximum pitch value H and minimum pitch value L f, as shown in Figure 6.

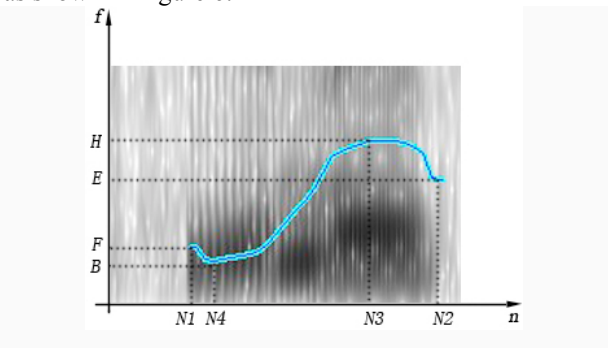

Figure 6. Prosodic feature control parameters schemes

E. Model Implementation

Due to the rhythm of the Tibetan Lhasa words characteristics, the traditional is administrative neural network model rhythm isn't a good rhythm reflect character, and in training network convergence and mapping capability will get great restrictions. According to the characteristics of rhythm control parameters influence degree, will context information parameter is divided into three groups, are shown below.

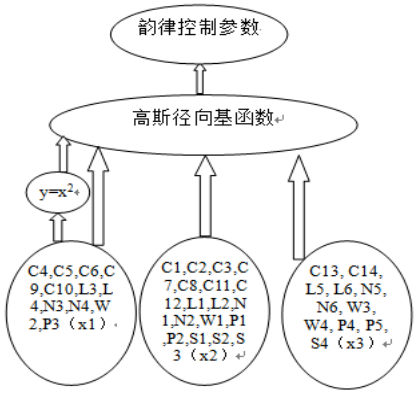

Figure 7. Neural network model rhythm schemes

Research shows that, the influence of the X1 group parameters on the rhythm of the control parameters is bigger. The Chinese in reference to deal with the problems of the existing achievement, the paper in input layer X1 group parameters and hidden between

neurons function is to join the $y=x^2$ weighted hidden layer, used to increase the weight of these parameters, and highlighted its in generating prosodic feature control parameters, and the effect of eventually to build a model [4].

V. The Experimental Results and Analysis

A. *Experiment*

Will the corpus of corpus 80% for training corpus, 20% of the corpus model for testing. Test randomly from corpus for a "ཨོན་ཀྱང་གནད་དོན་ཁག་གསུམ་གནས་པར་བསམ་གཞིག་སྐྱེད་དགོས་ཞེས་གསུངས་པ་རེད།" as a case, the rhythm control parameters were analyzed.

B. *The Result Analysis*

First to rhythm with test corpus training model, when model after training, the statement of the input context information model parameters, the sentence will output of the rhythm control information parameters. The statements of artificial extraction and model output the rhythm of the syllable duration control parameters, a frequency, maximum and minimum this model is compared, and evaluation.

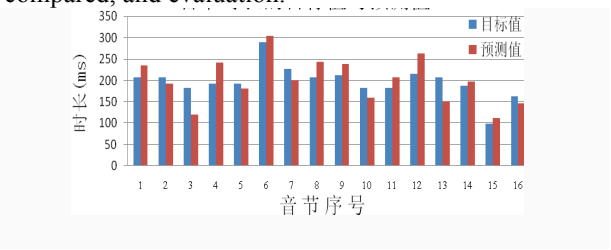

Figure 8. Syllable duration target and test values

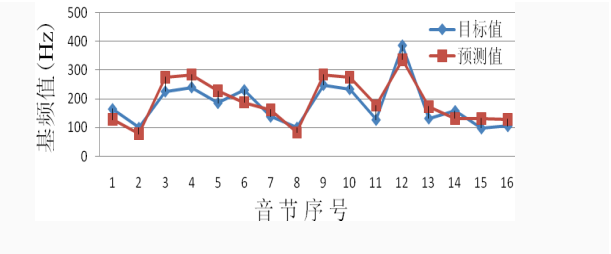

Figure 9. The pitch of maximum target and test values

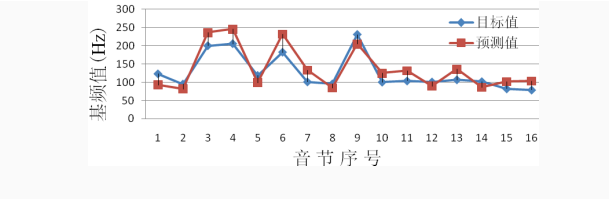

Figure 10. The pitch of the minimum target and test values

Figure 8 is long syllables and predicted the contrast of the target, we can see that the target and predicted the error between the time the average within 60 ms, showed a good prediction. In a frequency, the comparison of the maximum and minimum by the Figure 9 and Figure 10 can be seen, the model will be predicted error control in about 40 Hz ac.

VI. Summary

In this paper, the Lhasa words and Tibetan rhythm structure based on the research of the signal and breathing, the corresponding relationship between prosodic feature, and the influence of prosodic feature breathing signal parameters. Using RBF neural network established the Tibetan Lhasa words news body rhythm model. Through a model for training evaluation get results in output, after the syllable duration and frequency value are better reflect the characteristics of itself corpus and good results are obtained.

But in the process also to have the question, speech, breathing signal is interpersonal interaction field a new object of study, in the future to further study the signal to the speech signal breathing corresponding relation, breathing signals and prosodic feature the link between affecting prosodic feature of signal parameter, in addition, breathing in using RBF neural network model to establish rhythm, the input context information model parameters and the rhythm of the output control parameters optimized, choose more parameters, can reflect the parameters of prosodic feature, the performance of the optimization model.

Acknowledgment

The success of this project was subsidized by the National Support Project (2009BH41B03).

References

[1] J. Tan, J. Kong, and Y. Li "Chinese putonghua different style when you read breathing reset study,". *Journal of Tsinghua University (Natural Science Edition)*, 09, 2007.

[2] Q. Chen, H. Yu, and Y. Li. "Words breathing signal analysis system design and realization," *Computer Engineering and Applications*, (s20-40), 2009.

[3] J. Tao, L. Cai, and Z. Wu. "Based on the statistical model, the rhythm of the modeling method," in *The Sixth National Human Voice Communication Academic Conference Proceedings*, 2001

[4] Q. Chen, H. Yu, C. Chen, and J. Shi. "Take a show: Study of Tibetan Prosodic Model on the site have respiratory signals," in *Proc. Of the ICIA 2010*, 2010.