

# CMDMC: A Diachronic Digital Museum of Chinese Mandarin

Hou Min<sup>1</sup>, Zou Yu<sup>1</sup>, Teng Yonglin<sup>1</sup>, He Wei<sup>1</sup>  
Wang Yan<sup>1,2</sup>, Liu Jun<sup>1,2</sup>, and Wu Jiyuan<sup>1,2</sup>

<sup>1</sup>Broadcast Media Language Branch, National Language Resources  
Monitoring and Research Center at Communication University of China

<sup>2</sup>School of Literature, Communication University of China  
Beijing 100024, China

{houmin, zouiy, tengyonglin, hewei}@cuc.edu.cn,  
forget1812@sina.com, {aaa\_0119, wjy\_\_00}@163.com

## Abstract

Modern Chinese Mandarin has gone through near a hundred years, it is very important to store its representative sample in digital form permanently. In this paper, we propose a Chinese Mandarin Digital Multi-modal Corpus (CMDMC), which is a digital speech museum with diachronic, opened, cross-media and sharable features. It has over 3460 hours video and audio files with metadata tagging. The materials, which were generated by the authoritative speakers (e.g. announcers at TV or radio station) with normality, are required samples if we can get them. Based on this resource, we also intend to analyze the syntactic correlations of prosodic phrase in broadcasting news speech, and compare the phonetic and prosodic features in movie dialogues among several same-name movies in different historical eras.

## 1 Introduction

Modern Chinese Mandarin has gone through near a hundred years. As language changes as society develops, Mandarin must be periodically marked with the different features of different historical eras. It is very important to design and construct a Chinese Mandarin Digital Multi-modal Corpus (CMDMC), and store its representative sample in digital form permanently.

It's international trend to establish large-scale natural language corpus, and many countries

pay more attention to research and preserve their national language. For instance, the Linguistic Data Consortium (LDC) is an open consortium of universities, companies and government research laboratories. It creates, collects and distributes speech and text databases, lexicons, and other resources for research and development purposes<sup>1</sup>. Since its foundation, the LDC has delivered data to 197 member institutions and 458 non-member institutions. Moreover, European Language Resources Association (ELRA)<sup>2</sup> is the driving force to make available the language resources for language engineering and to evaluate language engineering technologies. In order to achieve this goal, ELRA is active in identification, distribution, collection, validation, standardization, improvement, in promoting the production of language resources, in supporting the infrastructure to perform evaluation campaigns and in developing a scientific field of language resources and evaluation. In this paper, we intend to establish the CMDMC with the goal of showing the history of the development of Chinese Mandarin, and representation the real character in different historical eras.

The paper is organized as follows: Section 2 describes the resources and data processing of our CMDMC. The experiment and evaluation is designed and carried out in section 3. Section 4 is dedicated to analyze the syntactic correlations of prosodic phrase in broadcast news speech on CNR (China National Radio), and compare the

---

<sup>1</sup> The Linguistic Data Consortium (LDC),  
<http://www ldc upenn edu>.

<sup>2</sup> European Language Resources Association (ELRA), <http://www elra info/>.

phonetic and prosodic features in movie dialogues. Finally, some conclusions and outlines of our future work are given in section 5.

## 2 General Description of CMDMC

In order to show the history of the development of Chinese Mandarin, and representation the real character in different historical periods, the CMDMC, which is a dynamic miniature model (or speech museum) with diachronic, opened, cross-media and sharable features, is designed and constructed by Broadcast Media Language Branch of National Language Resources Monitor & Research Center at Communication University of China.

In China, announcers in Radio & TV stations, as well as movie or stage actors, are the authority of the national language standardization. Therefore, the speech in radio, television and movie can be taken as the paradigm and representative of Mandarin. They can reflect the phonetic situation of that era. All of these are the source of the sample data for CMDMC.

### 2.1 Description of Resources

In order to fully demonstrate the development of Chinese Mandarin by the past 100 years, we try to collect all the video or audio materials in different periods. Therefore, a state-of-the-art classification is defined based on the corpora that we got.

**Language styles:** According to characteristic speaking styles of different media, there are three categories was defined, such as broadcast media language, movie or drama dialogue, and the dialogue in folk art (e.g. *xiangsheng*, *pingshu* etc.) and so on. To sum up, the three speaking styles accounted for about 64.9%, 27.2% and 7.9% of total corpora, respectively.

**Mediums:** The materials can be divided into audio, video, text and image/picture. The audio or video files are the main materials in our corpus, and the aligned texts are transcribed based on the audio or video. The documents of image are subsidiary corpora.

**Historical eras:** Based on the characteristics of social and language changes, we also define six historical stages of Chinese Mandarin: 1)

Before1949 (or 1919-1949), it is a theoretical stage for corpora collection. In fact, the earliest speech materials, which we can collect, is released in 1932; 2) 1949-1965; 3) 1966-1977; 4) 1978-1989; 5) 1990-1999; 6) 2000 to today. Table 1 shows the distribution of detailed data in different eras.

Eras	Broadcast media (hours)	Movie /drama (hours)	Folk art (hours)	Percent of total (%)
1932-49		39.3		1.1
1949-65	5.2	191.4	20	6.2
1966-77	17.5	93.0		3.2
1978-89	52.4	145.9	75.5	7.9
1990-99	43.5	137.5	11.5	5.6
2000--	2131.5	337.0	167.1	76.0
Total	2250.1	944.1	274.1	

Table 1: The distribution of video and audio materials in different eras.

### 2.2 Data Processing

The data processing includes metadata tagging, text transcription and aligning, phonetic and prosodic annotation, POS and syntactic tagging and so on.

As for labeling prosodic phrase boundaries, we strictly dependent on the prosodic criteria and perception by using the wave files and their transcriptions, which use many prosodic features such as F0 contour, energy contour etc. At the same time, some spoken phenomena are considered.

## 3 Experiment and Evaluation

Firstly, in order to investigate the correlations between prosody and syntax, about 13 hours speech materials were selected to segment and label, including break index, stress index and summary of emotional tendentiousness etc. Before the real annotation, six transcribers have been trained in accordance with the prosodic labeling conventions, until a high consistency of prosodic annotation can be carried out.

According to above experiment and annotation, the number of occurrences of the various boundaries was calculated in table 2.

Secondly, we also designed a perception experiment to determine phonetic diversification for elimination as much as possible the subjectivity which could be caused by the different personal intuition of language. Ten people at-

tended the perception experiment of this study: 3 men and 7 women. The average age is 25 years. Nearly all of them were graduates majoring in linguistics. During the experiment, the participants were asked to discriminate 12 paragraphs of random materials and judge the naturalness, pitch, and speech rate of the sentences produced in each paragraph. These 12 paragraphs consisted of 4 from 21 paragraphs of the 1995 version, 4 from 21 paragraphs of the 1975 version and 4 from modern materials.

Boundaries			Frequency
Types	Index	Marker	
PW	1	/1, /1+	55237
PP	2	/2	28867
C-PP	2	/2*	5976
IP	3	/3	7147
IG	4	/4	2781
MEC	5	/5	1770

Table 2: Distribution of all boundaries. The PW, PP, C-PP, IP, IG and MEC are the abbreviation of prosodic word, normal prosodic phrase, complex prosodic phrase, intonational phrase, intonational group and meaning expression cluster respectively.

In the perceptive procedure, we disordered all these materials for experiment, and three choices were given to these ten people: 1) natural, in conformity with the standard of modern Mandarin; 2) fairly natural, close to the standard of modern Mandarin; 3) unnatural, a little stogy. Every paragraph was released twice with an interval of 10 seconds. After one hour of continuous work, a 10-minute break was given.

Only the results with at least a 90% agreement rate were considered for analysis.

## 4 Related Works

Based on this resource, we intend to analyze the syntactic correlations of prosodic phrase in broadcasting news speech on CNR, and compare the phonetic and prosodic features in movie dialogues among several same-name movies in different historical eras.

### 4.1 Correlation between Syntax & Prosody

In English, there is a strong correlation between prosodic phrase boundaries and syntactic phrase boundaries (Price *et al.* 1991). That is to say,

prosodic phrase boundaries can play an important role in understanding utterance as punctuation marks do in written language. An investigation propose that boundary strength according to the measure, which the boundary strength is applied to syntactic structures and the phrase structure is viewed as an immediate constituency tree exclusively, corresponds much more closely to empirical prosodic boundary strength than does syntactic boundary strength according to a standard measure (Abney, 1992). In Greek, some study indicated that prosodic phrasing has a 95% identification rate, and a major effect on final tonal boundaries (Botinis *et al.* 2004).

In Chinese, some researchers also proposed a statistical model to predict prosodic words from lexical words. In their model, both length of the word and the tagging from POS are two essential features to predict prosodic words, and the results showed approximately 90% of prediction for prosodic words (Chen *et al.* 2004).

What the correlation between syntax and prosody is in Chinese broadcasting news speech? In order to investigate the syntactic correlations of prosodic phrase in real read speech on radio, we chose the representative speech materials from *Xinwen he Baozhi Zhaiyao (News and Newspapers Summary)* from CMDMC, which is a very famous broadcast news program of CNR.

This news program contains more syntactic, semantic and prosodic information, speaking styles and high quality voice in real context. Therefore, 908 programs, which contain 454 hours speech data from January 2006 to June 2008, were selected for pre-processing. After the pre-processing step, we selected two female’s 13 hours speech materials (one female announcer’s material forms the main data, and another one’s is supplemented for comparable data) as a core database, which segmentation, transcription and prosodic annotation (including break index, stress index and summary of emotional tendentiousness etc) was made by six transcribers.

According to the characteristic of broadcasting news speech, a new prosodic hierarchical structure (Zou *et al.* 2009) and two different types of prosodic phrase (i.e. the normal prosodic phrase and the complex prosodic phrase) boundaries were defined and used in our data labeling.

Categories	Location	Top pitch value			Bottom pitch value		
		N	SD	Mean	N	SD	Mean
PW	Left	3478	3.917	16.1	3253	4.761	8.5
	Right	3701	4.894	14.7	3165	5.457	9.9
PP	Left	1741	3.891	14.7	1718	4.302	6.2
	Right	627	3.481	16.5	492	5.077	9.3
C-PP	Left	314	4.085	13.5	317	4.135	4.8
	Right	361	3.616	17.9	285	5.092	10.0
IP	Left	536	4.817	12.9	456	5.575	3.9
	Right	531	3.019	18.8	473	3.720	13.8
IG	Left	211	4.363	11.4	203	6.055	4.7
	Right	229	2.377	19.4	185	2.927	15.0
MEC	Left	104	4.238	8.1	95	4.937	2.6
	Right	22	2.178	18.7	12	2.893	16.2

Table 3: The distribution of pitch on different boundaries. The phonetic acoustic data of each syllable was extracted by Praat script, and the foundational frequency was normalized by semitones, the normalization formula is  $ST=12*\log(F_0/F_{ref})/\log 2$  (the female’s reference frequency is 100Hz). (“top” is the mean of the highest pitch value at the first tone and the fourth tone; “bottom” is the mean of the lowest pitch value at the third tone and the fourth tone; “N” refers the number of samples; “SD” is the abbreviation of standard deviation)

In the further step, we selected 100 minutes speech materials from core annotated data, and investigated its features of pitch and duration at boundary (Zou *et al.* 2010). The detailed data are shown in table 3 and 4 respectively.

Boundaries		N	Mean	SD
Types	Marker			
PW	/1 or/1+	118	65.2	61.714
PP	/2	659	97.6	84.140
C-PP	/2*	193	108.7	82.483
IP	/3	877	343.2	138.906
IG	/4	375	699.2	254.287
MEC	/5	31	771.0	208.580

Table 4: The mean of silent pause duration at boundaries.

There are two ways of representation to pitch feature at prosodic boundary: Firstly, the pitch contour is un-continuity; secondly, the pitch resetting of the declination contour (de Pijper *et al* 1994). According to Table 3, we can find that there is a few resetting of bottom pitch value at PW boundary, that is to say, the bottom of the PW boundary right is 1.4 semitones higher than that of its left. At other boundaries, the bottom pitch values at right side are much higher than that at left side, for instance, there is 3.1, 5.2, 9.9, 11.3 and 13.6 semitones resetting from PP

to MEC boundary successively. Especially, at the IP boundary its resetting has about two times than that of C-PP boundary. This shows that there are very obvious prosodic feature at various boundaries in broadcasting news speech.

Generally, we know that 90ms is the floor of threshold for perceiving the silent pause. From Table 4, the mean of silent pause duration from long to short followed by MEC > IG > IP > C-PP > PP > PW. Except there is no perceived silent pause at PW boundary, the other boundaries have obvious silent pause that can be perceived. The length of silent pause at PP and C-PP are 97.6ms and 108.7ms respectively, and the length at IP has over three times longer than that at C-PP. According to this, we propose that the PP and C-PP lie in the same position at the prosodic hierarchical structure, and the C-PP is a special prosodic phrase.

From our core data we got 6728 C-PPs. According to the C-PP that contains the number of PW, we divided them into four categories, such as three-PW, four-PW, five-PW and six-PW. The distribution of them is shown in Table 5.

After preliminary analysis we found that the C-PP, which contains three PWs, has a simple syntactic structure although it is absolute majority in the number, and that is compose of four PWs should be done for correlations of prosody and syntax. There are about 6 types of prosodic

structure if the C-PP contains four PWs. The detail data of this type C-PP followed in table 6.

From the data, we know that the fourth type, which is  $(A+B) + (C+D)$ , is the most, and that is composed by  $(A+B) + C + D$  is the least in all of the six types. Although there are just six types of prosodic structure that can be found, there are more than 985 syntactic categories in this 1835 C-PPs. There are 23 types which occur more than 10 times, and most of them occur only one time. To some extent, it can explain that the syntactic structure is more complex than the prosodic one.

An example of prosodic and syntactic structures in the utterance, which is *ou1 yang2 yu3 hang2 yi4 zhi1 shou3 jin3 jin0 bao4 zhu4 lou2*

*ti1 de0 lan2 gan1* (Ouyang Yuhang held fast to the staircase railing with one hand), is given in figure 1. The left side of figure is the prosodic structure, and the syntactic one lies at the right side.

In figure 1, there is a little difference of *jin3 jin0* *jin0 bao4 zhu4 lou2 ti1 de0 lan2 gan1* (紧紧抱住楼梯的栏杆) between its prosodic structure “ $A+(B+C+D)$ ” and its syntactic structure “[VP [VP *jin3jin0/adv bao4zhu4/v*] [NP [AP *lou2ti1/n de0/u*] [NP *lan2gan1/n*]]]”, but the differences between its prosodic and syntactic structure are obvious because the *jin3 jin0* is stressed in speech for semantic expression.

Categories	Example	Num.
Three-PW	开展/1+ 互利/1 合作/2*	4433
Four-PW	第九次/1+ 全国/1 代表/1 大会/2*	1835
Five-PW	国际/1 市场/2 原油/1 期货/1 价格/2*	414
Six-PW	遭/1+ 不明/1 身份/2 武装/1 人员/2 袭击/2*	46
Total		6728

Table 5: The distribution of four kinds of C-PP

Types	Example	Num.	Percent (%)
A+(B+C)+D	与/1+ 不利/1 因素/2 并存/2*	441	24.03
A+(B+C)+D	分获/1+ 一/1 二/1 三等奖/2*	495	26.98
A+B+(C+D)	国家/1+ 著名/1+ 一级/1 演员/2*	97	5.29
(A+B)+(C+D)	将/1 上涨/2 一成/1 左右/2*	529	28.83
(A+B+C)+D	中国/1 民间/1 国宝/2 称号/2*	259	14.11
(A+B)+C+D	受/1 美股/2 大跌/2 拖累/2*	14	0.76
Total		1835	100

Table 6: The distribution of prosodic type in C-PP of four-PW

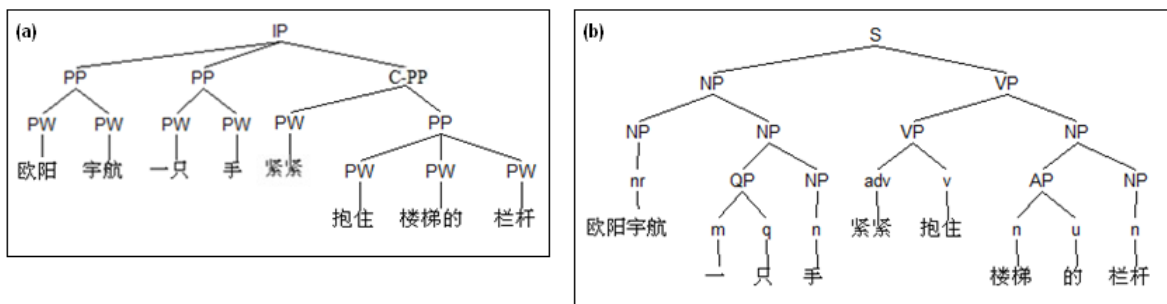


Figure 1: An example of (a) prosodic structure vs. (b) syntactic structure in an utterance: *ou1 yang2 yu3 hang2 yi4 zhi1 shou3 jin3 jin0 bao4 zhu4 lou2 ti1 de0 lan2 gan1* (Ouyang Yuhang held fast to the staircase railing with one hand).

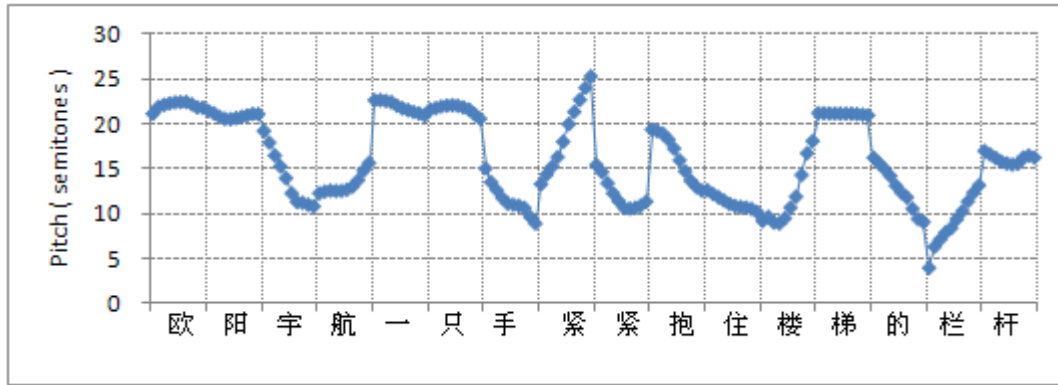


Figure 2: The pitch contour of the same utterance.

Figure 2 shows the pitch contour of the same utterance. In this utterance, there is a nesting structure at *jin3 jin0 bao4 zhu4 lou2 ti1 de0 lan2 gan1* (held fast to the staircase railing) based on the length of perceived silent pause. Furthermore, the pitch declination trend within the C-PP is obvious despite small resetting between *zhu4* and *lou2*. So we suggest that there is a stable prosodic pattern within a C-PP in broadcasting news speech.

Conversely, what is the correlation between the prosody and syntax? From above analysis, we know that the conjunction and particle, such as 的(*de0*), 等(*deng3*), 和(*he2*), 但(*dan4*) and so on, more likely attached to the end of left structure or the beginning of right one and form a prosodic word. If it has just four lexical words including the conjunction or particle they form a prosodic word by itself. That is to say, it has very great flexibility in prosodic structures for conjunctions and particles, such as “占(*zhan4*)/1+ 全国(*quan2guo2*)/1 湿地(*shi1di4*)/1 面积的(*mian4ji1 de0*)/2\* (*occupy/1+ country-wide/1everglade/1 acreage/2\**)”, “和(*he2*)/1 社会(*she4hui4*)/2 救助(*jiu4zhu4*)/1 制度(*zhi4du4*)/2\* (*and/1 social/2 assistance/1 system/2\**)” and so on.

#### 4.2 Diachronic Comparative Phonetic and Prosodic Analysis in Movie Dialogues

Which diachronic phonetic changes happened in Mandarin by the past 100 years? We also analyze and compare the phonetic features of Chinese Mandarin among several same-name movies in different historical eras from CMDMC (Wang *et al.* 2010). In order to minimize the

divergence of the variables and maximize the reliability of conclusions, we chose two pairs of same-name movies screened in different historical periods. These movies are: *Pingyuan Youjidi* (*The Plains Guerrillas*) shot in 1955 and 1975, *Dujiang Zhencha Ji* (*Reconnaissance across the Yangtze River*) shot in 1954 and 1974 respectively.

**Pitch Feature:** In the analysis of pitch, we put aside the stresses and the neutral tone syllables, and make the statistical investigations on the top pitch value and the bottom pitch value of the syllables.

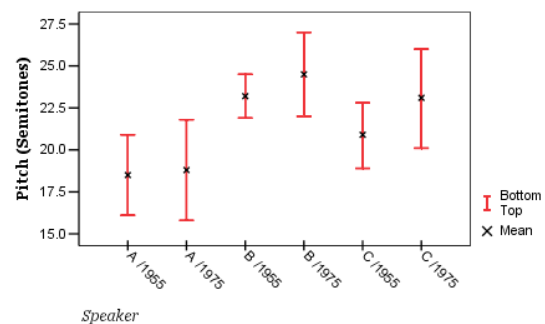


Figure 3: The pitch data of 1955 and 1975 version in *the Plains Guerrillas*. The fundamental frequency also was normalized by semitones; the male's reference frequency is 50Hz.

Figure 3 shows that the mean of the top pitch value in the 1950s' materials is lower than that of 1970s'. In the 1955 version, the leading character, Speaker A, possesses a mean value of the top pitch value which is 20.9 semitones. This value is lower than that of 1975s' by a difference of 0.9 semitones. The negative character, Speaker B, has a mean value of the top pitch

value which is 24.5 semitones in the 1955 version. The value in the 1975 version is 27 semitones, with a difference of 2.5 semitones left, also showing that the value in the 1975 version is comparatively high. Comparing the data of the bottom pitch value in the 1955 version with that in the 1975 version, we know that these data seem closer than the top pitch value, but still the higher ones belong to the 1975 version. That the bottom pitch value is higher tells us that the whole pitch register is raised.

Furthermore, we can easily see from Figure 3 that the pitch range of the same character in the 1975 version is wider. Speaker A of the 1955 version has a pitch range of 4.8 semitones. In contrast, the same character in the 1975 version has a pitch range of 6 semitones. Speaker C of the 1955 version has 4 semitones pitch range, but in the 1975 version, he has 5.9 semitones pitch range. The gap between them is 1.9 semitones. Through this comparison, we find that the pitch range in the 1975 version is wider than that in the 1955 version in the whole.

To some extent, the speaking, both the top pitch value and the bottom pitch value in the 1975 version are higher. This proves that, on the whole, the pitch of the 1970s' materials is higher and more unnatural than that of 50s' because of the effect by the Cultural Revolution era. And this also proves the feeling of the participants in the perceptual experiment at section 3 about the 1970s' materials, that is, the 1970s' Mandarin has a loud and sonorous voice; the characters pronounce harder; the general pitch is higher.

**Duration feature:** In the respect of duration, we also compared and analyzed the presenters' speech on TV in 2005<sup>3</sup> with the materials extracted from the movie dialogues the 1955 and the 1975. Table 7 is the relevant data.

According to table 7, there is a little difference of the durations mean among them (following four tones), especially it's very closely between the 1975 and the 2005, and those of the 1975 version are a few longer than those of the 1955 version. But, except the first tone (Sig. =.077), the differences of the duration means between the others, which is in the 1955, the

1975 and the 2005, are significant (Sig. =.000, .000, .002 < .05 respectively).

		mean	SD	N
Movie:1955	T1	153.6	69.5	243
	T2	136.8	58.1	242
	T3	132.8	58.7	321
	T4	133.5	52.0	539
Movie:1975	T1	177.8	72.1	258
	T2	155.5	52.0	263
	T3	152.5	57.6	289
	T4	156.7	59.9	505
TV: 2005	T1	163.1	65.7	1471
	T2	156.0	66.5	1743
	T3	156.8	67.6	1054
	T4	145.9	62.3	2652

Table 7: The duration mean of four tones in movie dialogues (1955 and 1975) vs. that of presenters' spoken language on TV in 2005(ms).

**Demonstrations of the four-syllable prosodic words:** The comparative pitch contour of two four-syllable prosodic words, which are “*bu2 yao4 lu4 mian4*” (don't appear) and “*gan4 shen2 me0 de0*” (What are you doing?), are shown in Figure 4 and 5, respectively.

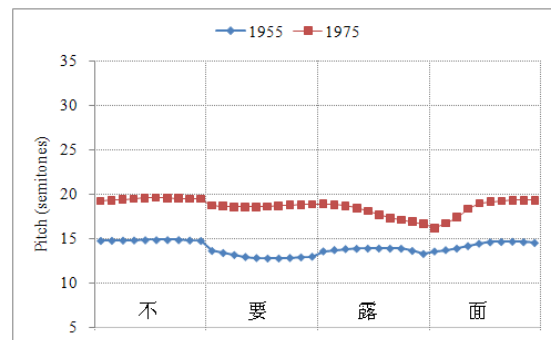


Figure 4: The pitch contour of “*bu2 yao4 lu4 mian4*” (don't appear)

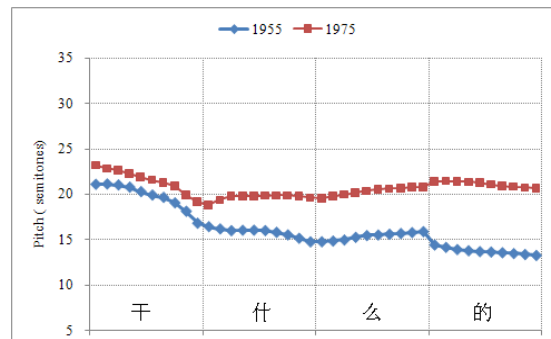


Figure 5: The pitch contour of “*gan4 shen2 me0 de0*” (What are you doing?)

<sup>3</sup> In this work, we just chose the male's speech data from Zou (2007).

By observing the above two figures, we find that the pitch contour of the 1975 and that of the 1955 are almost identical except the latter is always lower than the former. This may explain that although the Mandarin has gone through a hundred years, the pitch pattern is relatively stable.

## 5 Conclusions and Future Work

This paper proposes to design a Chinese Mandarin Digital Multi-modal Corpus (CMDMC). Through this corpus, the historical trace of Mandarin development can be followed; the fresh and alive data and material resources can be drawn up for the modern researchers and successors. We also intend to analyze the syntactic correlations of prosodic phrase in broadcasting news speech, and compare the phonetic and prosodic features in movie dialogues among several same-name movies in different historical eras. The contributions are as follows.

Firstly, the syntactic structure is more complex than the prosodic structure, some conjunction and particle, such as *de0*, *deng3*, *he2*, *dan4* and so on, more likely attached to the end of left structure or the beginning of right one and form a prosodic word, if the number of lexical words mismatch the prosodic words. Otherwise, they have almost similar structure.

Secondly, the speech of 1970s in last century is greatly influenced by the special era. People usually use exaggerated voice, pronounce hard and raise the pitch unnaturally, giving others a taste of lecturing and ordering. In contrast, the speech of Mandarin in 1950s is more natural and close to the daily life pronunciation and intonation. Even so, the pitch patterns have no big changes, and this may explain that the pitch patterns are comparatively stable in Chinese Mandarin.

Future research will include treatment of correlation between syntax and prosody within IP or IG, ideally comparing the diachronic phonetic or prosodic changes in Mandarin by the past 100 years. Additionally, we would like to tackle the problem of data management, update and periodical increasing as time passes.

## 6 Acknowledgements

This work was supported by the Department of Science and Technology at Ministry of Educa-

tion (No. 107118), and “211” Key Projects of Communication University of China (No. 21103010105, 21103010106). We would like to thank the anonymous reviewers for their insightful comments.

## References

- Abney, S. 1992. Prosodic Structure, Performance Structure and Phrase Structure. *Proceedings of 5th Darpa Workshop on Speech & Natural Language*.
- Botinis, A., Ganetsou, S., Griva, M., and Bizani, H. 2004. Prosodic Phrasing and Syntactic Structure in Greek. *Proceedings of FONETIK 2004*, Dept. of Linguistics, Stockholm University.
- Chen, Keh-jiann, Tseng, Chiu-yu, Peng, Hua-jiu and Chen, Chi-ching. 2004. Predicting Prosodic Words from Lexical Words -- A First Step towards Predicting Prosody from Text. *Proceedings of the 4th International Symposium on Chinese Spoken Language Processing (ISCSLP 2004)*. Hong Kong, 173-176.
- de Pijper, J. R., and Sanderman, A. A. 1994. On the Perceptual Strength of Prosodic Boundaries and its Relation to Suprasegmental Cues. *Journal of the Acoustical Society of America*, 96(4), 2037-2047.
- Price, P., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. 1991. The Use of Prosody in Syntactic Disambiguation. *Journal of the Acoustic Society of American*, 90, 2956-2970.
- Wang Yan, Liu Jun, Kan Minggang, Hou Min, Zou Yu. 2010. Phonetic Diachronic Diversification in Mandarin: A Case of the Same Movie's Dialogue in 1950s and 1970s. *Proc. of YWCL 2010*, Wuhan, Hubei, Oct 10-13. (Accepted)
- Zou Yu. 2007. *A Formal Study on Prosody of Presenter's Spoken Language Based on Broadcast Speech Corpus*. PhD thesis, Communication University of China.
- Zou Yu, He Wei, Zhang Yuqiang, Hou Min and Zhu Weibin. 2009. A Special Prosodic Phrasing in Broadcasting News Programs. *Computational Sciences and Optimization: Theory, Simulation and Experiment (Vol. 2)*, Sanya, Hainan, China, 24-26 April, 406-408.
- Zou Yu, Wu Jiyuan, He Wei, Hou Min, Teng Yonglin. 2010. Syntactic Correlations of Prosodic Phrase in Broadcasting News Speech. *The 6th IEEE International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE 2010)*, Beijing, China, Aug. 21-23.