MANDARIN ACCENT ANALYSIS BASED ON FORMANT FREQUENCIES

Kun Liu[1], Zhiwei Shuang[2], Yong Qin[2], Jianping Zhang[1], Yonghong Yan[1]

ThinkIT Speech Lab, Institute of Acoustics, Chinese Academy of Sciences[1]
IBM China Research Lab[2]
{kliu, jzhang, yyan}@hccl.ioa.ac.cn, {shuangzw,qinyong}@cn.ibm.com

ABSTRACT

Accent analysis for Mandarin Chinese based on formant frequencies is presented in this paper. Five monophthongs [a, o, e, i, u] of 430 speakers across eight accents were analyzed with univariance analysis of variance (UNIANOVA). The results show that accent has significant influence on the second formant frequency of monophthongs [o, i, u] and has no obvious influence on the formant frequencies of monophthong [a]. In addition, accent has no obvious influence on the first and the third formant frequencies of all five monophthongs.

Index Terms— Accent, formant, UNIANOVA

1. INTRODUCTION

The speaker variability caused by accent is one of the most critical issues for speech signal processing, especially for automatic speech recognition (ASR) [1]. ASR systems trained on Standard Mandarin Chinese often experience a dramatic accuracy loss when tested by speakers with strong accents. Accent can be defined as the patterns of pronunciation features which characterize an individual speech belonging to a particular native language group. When learning a second Language, the speaker will carry these accent patterns of his native language into the new language spontaneously. Therefore, many accent traits will persist in his speech [2].

Many researches on accent in English and Mandarin Chinese have been reported. For English, accent analysis and classification were explored with Hidden Markov Model (HMM) codebooks among three types of accented English by non-native speakers and standard American English [2]. Support Vector Machine (SVM) and HMM were proposed for accent classification by Tang [1]. Also accent detection had been studied with spectral emphasis [3]. Recently, perceptual assessment of accent variation in US native English was given by Lin [4]. While for accented Mandarin Chinese, which are strongly influenced by the speakers' regional dialects of Chinese, Shanghai accented Mandarin is always the one having been explored. Yu and Li compared nine diphthongs between standard Mandarin and Shanghai accented Mandarin with medium accent [3]. After the phonetic research on three dialectal regions: Shanghai, Wuhan and Xiamen, Li suggested that supra-segmental features played an important role in rating accent degree [5]. And the accuracy of Shanghai accented speech recognition had been improved through accent detection, accent discriminative acoustic features and acoustic adaptation techniques [6]. The acoustic model auxiliary reconstruction with accent decision trees introduced about 4% absolute WER reduction evaluated on Cantonese and Wu accented recognition [7]. Although certain researches on accented Mandarin have been done, no statistical conclusions about the influence of accent on formant frequencies are given.

In this paper, we investigate the influences of accent on formant frequencies. Eight types of Mandarin accents are studied based on the formant frequencies. The results show that accent plays an important role in the second formant for monophthongs [o, i, u]. This paper is arranged as follows. In section II, the corpus for analysis and the criteria for data selection are described. Automatic parameter selection procedure is introduced in Section III. Analysis and discussion about accent are described in Section IV. The conclusion is made in Section V.

2. CORPUS AND DATA SELECTION

2.1. Corpus

Mandarin-5000 database purchased from Siemens AG was used in this study. This data contains the recordings of 4,752 speakers across 32 accented Mandarins. All speech was recorded over the telephone channel as PCM format, 8-bit, 8 kHz sampling rate. And the accent of the speech was judged as medium by subjective estimation of dialectologists.

2.2. Data selection

2.2.1. Vowel Selection

Consider the coarticulatory effects in the diphthongs and triphthongs, only monophthongs are chosen to be

analyzed here. In this paper, five monophthongs [a, o, e, i, u] were selected for accent analysis.

2.2.2. Syllable selection for each monophthong

To alleviate the effects of different neighboring consonants on each monophthong, we select seven syllables 'fa1', 'ba1', 'po1', 'bo1', 'che1', 'chu1', 'qi1' including these five monophthongs for analysis. And tone 1 is used to here avoid the potential effects of tones.

After the syllable selection procedure, more than ten thousands syllables in Mandarin from 430 speakers across the following eight accents: Beijing, Sichuan, Shanghai, Jiangsu, Henan, Shandong, Guangdong and Guangxi are used in this paper.

3. PARAMETER EXTRACTION

3.1. Formant Frequencies Extraction

Ideally, the hand-track formants should be used to avoid errors, but it is very time consuming in practice. State-of-the-art tool Praat (http: //www. praat. org) is applied here to extract the formant frequencies of all the data. The maximum formant (Hz) is set to 4000 Hz for all the speakers and the maximum number of formants is set to 3 and 4 for female and male respectively.

3.2. Automatic Formant Selection

3.2.1. Automatic Formant Selection Method

Instead of using all the formants of a monophthong, we choose to use the formant at the stable frame in the middle of the monophthong to exclude the formants in the transition part. When the variances of formant frequencies for several continuous frames remain in predefined ranges, the formant frequencies at these frames can be regarded as relatively stable formant frequencies. The means of the formant frequencies at these stable frames are used as the formant frequencies for this monophthong. Later, the dispersed formant frequencies are eliminated using mean and variance to compare the differences with other formant frequencies for the same monophthong.

This procedure is named as an Automatic Formant Selection (AFS) procedure, which is proposed for obtaining the stable formant frequencies automatically. The predefined maximal variation ranges for the first three formant frequencies are set to 30 Hz, 50 Hz and 100 Hz respectively and the minimum length of the stable segment is set to 40 ms.

3.2.2. Performance of AFS

The results from AFS are compared with those from manual formant selection (MFS) on a small database of seven speakers to test the performance of AFS procedure. MFS procedure is done with similar criteria except that the stable frames are judged by persons manually. This small database consists of seven speakers: two females and five males. 93 syllables containing 6 monophthongs [a, o, e, i, u, y] were recorded for each speaker. All speech was sampled at 8 kHz, 16 bit. Because the comparison results for all the 6 monophthongs are very similar, we just list the comparison results of monophthong [a] from the seven speakers in Table 1 for short.

Table 1: Comparison between AFS and MFS

	F_1(Hz)		F_2(Hz)		F_3(Hz)		Diff
	MFS	AFS	MFS	AFS	MFS	AFS	(%)
F1	1066.2	1041.6	1577.5	1615.4	3132.1	3118.4	-0.11
F2	1079.6	1101.7	1713.0	1755.1	3042.1	2981.5	0.84
M1	843.0	821.0	1266.4	1302.4	2480.0	2378.7	1.28
M2	831.5	846.3	1241.5	1309.7	2831.6	2675.2	0.58
M3	834.9	874.0	1409.0	1438.8	2896.8	2853.9	1.77
M4	769.5	763.1	1192.1	1203.4	2560.5	2578.3	0.27
M5	812.8	801.4	1238.1	1242.1	2386.4	2404.8	-0.10

Where, the symbols in the first column represent the ID of speakers. The six columns in the middle show the results of AFS and MFS for the first three formant frequencies. The difference in the last column is calculated by the mean of the differences between AFS and MFS for the first three formant frequencies. For all monophthongs including [a], the difference between AFS and MFS is less than 2%, and such difference is acceptable for formant analysis purpose. So AFS can be used to substitute manual selection.

3.2. Results of formant frequencies

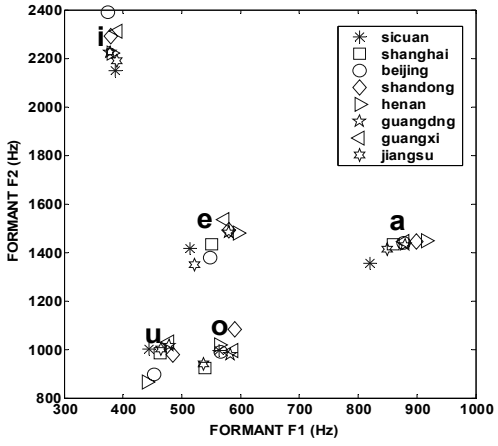

Figure 1: F_1 vs. F_2 across accented Mandarin

The scatter plot of all the mean frequency pairs (F_1, F_2) for each monophthong across the eight accents is shown in Figure 1. The mean frequencies with a particular accent for each monophthong are obtained by averaging the formant frequencies of all the speakers in this accent

group. Eight different symbols are used to represent different accents shown in the legend of Figure 1. For example, the point marked by symbol star (*) shows the frequency pairs (F_1, F_2) with Sichuan accent. And the name of each monophthong is shown near the corresponding symbols.

In this figure, the mean frequency pairs of different accents are clearly distinguished.

4. ANALYSIS AND DISCUSSION

Because formant frequencies are not just influenced by accent, other factors, such as gender and age, are also very important. UNIANOVA in SPSS (Statistical Package for the Social Science) is used to analyze the influences of accents and other factors on the formant frequencies here. UNIANOVA is widely used to analyze the main and interaction effects of independent variables on a single dependent variable. And General Linear Model (GLM) is often used to implement this UNIVNOVA procedure. This procedure assumes:

- Independence - the observations in each of the groups are independent.
- Normality - the distributions in each of the groups are normal.
- Homogeneity of variances – the variance of data in groups should be the same (Levene's test is used for homogeneity of variances).

4.1. Analysis model

The same model is built for each formant frequency of every monophthong. Variables in the model are defined as follows:

F_i (i=1, 2, 3) is the dependent variable representing the first, the second and the third formant frequency respectively.

'Accent' and 'Gender' are both fixed factors, which are categorical predictors. 'Accent' has eight levels representing eight different accents. And '0' and '1' are used to represent female and male separately.

'Age' is the random factor.

We use 'Accent + Gender + Age' to represent this design of this model. With the model built, we can find the main effects of 'Accent', 'Gender' and 'Age' on each F_i.

4.2. Analysis results

4.2.1. Normality test

The normality of the data should be tested first because UNIANOVA assumes the normality of the observations. One-Sample Kolmogovov-Smirnov procedure [8] is used to test normality of F_i for each monophthong. If the significance value is above the significance level 0.05 and the number of observations for testing is enough (the minimum number is 25 or 30), it is safe to determine that the data is adequately normal to perform UNIANOVA. And the normality test results for F_i are given in Table 2.

Table 2: Results of One-Sample Kolmogorov-Smirnov Test

Sig.	a	o	e	i	u
F1	**0.019**	0.112	0.179	**0.004**	**0.036**
F2	**0.000**	0.084	0.771	0.194	0.292
F3	0.082	**0.000**	0.280	0.125	**0.000**
N	403	316	139	356	224

Where, N means the number of syllables analyzed for each monophthong. From this table, we can see that F_1 of monophthong [a, i, u], F_2 of monophthong [a] and F_3 of monophthong [o, u] do not satisfy normality because sig. of these F_i is less than 0.05. If these F_i are analyzed with UNIANOVA model, the results may be doubtful.

4.2.2. Analysis results of F_i for these five monophthongs

The significance values of influences from different factors on F_i for each monophthong are given the middle three columns of Table 3. If the significance value is less than 0.05, the factor has a significant effect on F_i. The smaller the sig., the stronger the effect will be.

Table 3: Analysis results of F_i for each monophthong

Mono	Formant	Sig. of Effect Size			Sig. of Homogeneity Test
		Accent	Gender	Age	
a	F_1*	0.115	**0**	0.689	0.716
	F_2*	0.472	**0**	0.502	0.263
	F_3	0.369	**0**	0.515	0.277
o	F_1	0.146	**0**	0.580	0.106
	F_2	**0.001**	**0**	0.187	0.538
	F_3*	0.667	**0**	0.235	0.427
e	F_1	**0.023**	**0**	0.183	0.503
	F_2	0.122	**0**	0.603	0.188
	F_3	**0.049**	**0**	0.058	0.251
i	F_1*	0.767	**0**	0.006	**0.001**
	F_2	**0.001**	**0**	0.499	0.556
	F_3	**0.047**	**0**	0.962	0.489
u	F_1	0.544	**0**	0.009	0.582
	F_2	**0.019**	**0**	0.639	0.994
	F_3*	0.526	**0**	0.036	0.851

Meanwhile the results of homogeneity test are listed in the last column. If the significance value of homogeneity test is larger than 0.05, the null hypothesis that each accent group has equal variance is accepted. Where, the asterisk (*) is used to indicate that the corresponding result may not be sure because the normality or the homogeneity of this variable are not satisfied.

4.4. Discussion

4.4.1. Influence of accent on formant frequencies

From Table 3, we can draw two conclusions:

- Accent has a great influence on F_2 for monophthongs [o, i, u] because their significant values of 'Accent' on F_2 are less than 0.05 and close to 0.

As commented by Wu in [9], formant frequencies are related to the articulating organs and F_2 is inversely proportional to the front-back of the tongue and is influenced by the shape of the lip. Therefore this conclusion suggests that accents may affect the position of the articulating organs relative to F_2, i.e., the front-back of the tongue and the shape of the lip, for accented Mandarin speakers.

- In addition, accent has no significant effect on the formant frequencies of monophthong [a] since all the significant values of 'Accent' on F_1, F_2, and F_3 are much larger than 0.05.

This result suggests that monophthong [a] is very stable for each accent. And this can be easily understood since syllable 'ma1' exists in almost every language and it is always the first syllable that infants are able to speak.

The results also show that accent contributes less to the F_1 and F_3 of the five monophthongs although the significant value of F_1, F_3 for the monophthong [e] and F_3 of monophthong [i] are slightly smaller than 0.05.

4.4.2. Influence of gender and age on formant frequencies

Zero for each significant value of 'Gender' on F_i shows that gender contributes greatly to formant frequencies.

The effects from the age of the speaker is not significant in total because most significance values are much larger than 0.05 and we can hardly find any rule from the occasional small significance values.

5. CONCLUSION

Five monophthongs of 430 speakers across eight accents in China are used to analyze the influences of accent on formant frequencies in this paper. Our results suggest that accent has a significant influence on the second formant frequencies for monophthongs [o, i, u] and no obvious influence on monophthong [a]. In addition, no obvious influence of accent is found on the first and the third formant frequencies of all the five monophthongs. These results are valuable for further accented Mandarin analysis, voice morphing and robust Mandarin ASR system.

6. ACKNOWLEGEMENTS

This research was conducted during the internship in IBM China Research Lab. The author is thankful to Yi Liu and all the other team members in speech group who gave great help to this research. And this work is supported in part by Chinese 973 program under Grant No. 2004CB318106 and National Natural Science Foundation of China under Grant No. 10574140 and 60535030.

7. REFERENCES

[1] Hong Tang, Ali A. Ghorbani, "Accent Classification Using Support Vector Machine and Hidden Markov Model," *Canadian Conference on artificial intelligence*, Halifax, pp.629-631, June 2003.

[2] Levent M. Arslan and John H.L. Hansen, "Language Accent Classification in American English," *Speech Communication*, Vol. 18, pp. 353-367, June 1996.

[3] Mattias Heldner, "Spectral Emphasis as an Additional Source of Information in Accent Detection," *In Proc. of ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*, Red Bank, NJ, pp. 57-60, 2001.

[4] Xiaofan Lin and Steven Simske, "Role of Prosody in the perception of US Native English Accents," *Proceedings of INTERSPEECH 2006,* Pittsburgh, Pennsylvania, pp. 437-440, September 2006.

[5] Aijun Li, Qiang Fang, Ziyu Xiong, "Phonetic Research on Accented Chinese in Three Dialectal Regions: Shanghai, Wuhan and Xiamen," *Proceedings of INTERSPEECH 2006,* Pittsburgh, Pennsylvania, pp. 705-708, September 2006.

[6] Yanli Zheng, Richard Sproat, Liang Gu, "Accent Detection and Speech Recognition for Shanghai-Accented Mandarin," *Proceedings of INTERSPEECH 2005,* Lisbon, Portugal, pp. 217-220 September 2005.

[7] Liu Yi, Pascale Fung, "Multi-Accent Chinese Speech Recognition," *Proceedings of INTERSPEECH 2006,* Pittsburgh, Pennsylvania, pp. 133-136, September 2006.

[8] Http://staff.harrisonburg.k12.va.us/~gcorder/test_normality_Kolmogovov_Smirnov.html

[9] Zongji Wu, Maocan Lin, "Summary of experimental phonetics," *Higher education press*, 1989.

[10] Jue Yu, Aijun Li, Xia Wang, "A Contrastive Investigation of Diphthongs between Standard Mandarin and Shanghai Accented Mandarin," *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, Beijing, China, pp. 113-117, March, 2004.

[11] Zhiwei Shuang, Raimo Bakis, etc, "Frequency Warping Based on Mapping Formant Parameters," *Proceedings of INTERSPEECH 2006,* Pittsburgh, Pennsylvania, pp. 2290-2293, September 2006.