

The Correlation Between Signal Distance and Consonant Pronunciation in Mandarin Words

Huijun DING, Chenxi XIE, Lei ZENG, Yang XU, Guo DAN

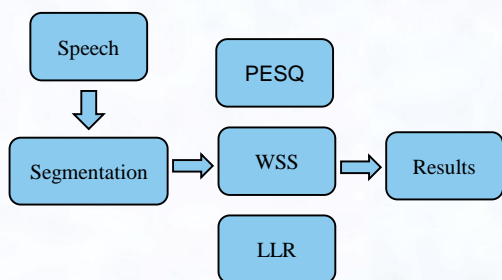
Introduction

In Mandarin language speaking, some consonant and vowel pairs are hard to be distinguished and pronounced clearly even for some native speakers. This study investigates the signal distance between consonants compared in pairs from the signal processing point of view to reveal the correlation of signal distance and consonant pronunciation.

Materials:

SPEECHOCEAN's Chinese Mandarin Speech Recognition Database (serial no.: King-ASR-028). It contains the voices of 98 native speakers from China. Each speaker utters 1267 monosyllables. They are sampled at 44.1 kHz, 16 bits, 3 channels. This study selects the voice uttered by a female from Peking with the age of 25.

Methods



Syllable analysis

A total of 397 meaningful syllables, the permutations of all the 23 consonants and 35 vowels in Mandarin (exclusive lexical tones), are analyzed.

Consonant-vowel segmentation

The audio file of each syllable is truncated into consonant segment and vowel segment by manual cutting.

Signal distance calculation

Three different objective measures, namely weight ed-slope spectral distance (WSS), perceptual evaluation of speech quality (PESQ), log-likelihood ratio (LLR) are used to calculate the signal distance for predicting the consonant and vowel pairs (CVP) confusion.

CVP Confusion

Some easy-confusing consonant and vowel pairs (CVP) and the areas in China with high incidence of such confusion are listed in this table.

No.	CVP	Possible Region of China
1	/l/ - /n/	Centre
2	/y/ - /r/	Northeast
3	/f/ - /h/	Fujian Province
4	/z/ - /zh/ /c/ - /ch/ /s/ - /sh/	South
5	/in/ - /ing/ /en/ - /eng/ /an/ - /ang/	Northwest
6	...	

Results

CVP \ Algorithm	LLR	PESQ	WSS
/l/ - /n/	0.47	1.42	23.48
/l/ - /b/	0.71	1.17	36.13
/l/ - /c/	0.95	1.20	31.94
/l/ - /ch/	1.92	1.07	36.92
/l/ - /d/	0.36	1.06	31.47
/l/ - /f/	1.54	1.39	33.77
/l/ - /g/	1.42	1.45	29.10
/l/ - /h/	1.27	1.47	32.77
/l/ - /k/	1.76	0.91	32.78
/l/ - /m/	0.26	1.73	27.77
/l/ - /r/	1.02	1.32	26.33
/l/ - /s/	3.52	0.79	40.34
/l/ - /z/	0.67	0.98	22.60
/l/ - /zh/	0.38	1.23	21.93
Ranking	4	4	3
Probability	78.57%	78.57%	85.71%

The signal distances of CVPs followed with vowel /an/ in Tone 3.

Algorithm	LLR	PESQ	WSS
Average Ranking	2.8	4.4	1.4
Probability	82.38%	56.38%	97.14%

The average ranking and probability of similarity between initial /l/ and /n/ with different finals

Conclusion

The shorter the distance is, the higher probability that the CVP causes confusion in pronunciation is. It can be best predicted by WSS, followed by LLR and PESQ.

*This research was granted by the national innovation and entrepreneurship student training program (201610590046). It is also supported in part by the Shenzhen Science and Technology Innovation council (JCYJ20140418095735605, JCYJ20140617104848154, KQTD2015033016104926), Natural Science Foundation of SZU (827/000088) and Science and Technology Planning Project of Guangdong Province (2015B020214007).