Discrimination of Chinese Aspirated and Un-aspirated Sounds
Uttered by Japanese Students

Yasukazu Kanamori[1] , Teruaki Tokoro[2]

[1,2] Aichi Prefectural University, 1522-3, Kumabari, Nagakute-cho, Aichi-gun, Aichi, Japan,

[1] **kanamori@ist.aichi-pu.ac.jp**

A Chinese aspiration is generally considered to be very difficult to pronounce by Japanese student. The difficulty is caused by lack of the aspiration in Japanese pronunciation. Voice onset time (VOT) is one of important parameter to discriminate aspirated and unaspirated sounds. However, some exceptions of long VOT existed in the aspirated sounds. Moreover, the researches of the aspirated sounds only to explosive sounds [p'], [p], [b'] and [b] were often made. Furthermore, affrication is carried out rarely in past papers. Not only the explosive unaspirated sounds [g] and aspirated sounds [k], but affrications are also the object of this study. In this paper, it proposes distinguishing using spectral tilts of some partial bands and the state of time change of energy. The distinguishing rates of [g] and [k] are 74.8% and 84.2% for the conventional method (both used VOT and Dynamic low-passed power). It is become to 88.7% and 94.7% for the proposing method respectively. Further analysis was made for six affrications of unaspirated sounds [z], [zh], [j] and aspirated sounds [c], [ch], [q]. When the proposing method is used, the rate is improved from 83.9% to 93.5% for [c/z] pair, and improved from 88.7% to 93.7% for all affricated unaspirated / aspirated sounds. This paper showed the validity of the proposing method.

1 Introduction

Many people around the world want to learn Chinese. Each Chinese syllable has three basic components: consonant, vowel, and tone. Some common Chinese words contain syllables that contain aspirated consonants, it differs from Japanese. The appearance probability of aspirated consonants is 13.4% when zero consonants are included in total consonants. The utterance is difficult although this has only the number of 1/3 of unaspirated consonants which appearance probability is 39.8% [1-2]. The importance of pronouncing these consonants properly is obvious.

When we try to listen to or to speak in a foreign language, there is a typical problem which is the native Japanese speakers' difficulty with the English sounds /r/ and /l/. Because the /r/-/l/ distinction does not exist in Japanese, many Japanese find it extremely hard to hear the difference or to pronounce the two sounds correctly, particularly in real-world situations [3-4].

A similar problem arises here when Japanese learn Chinese, since Chinese has many consonant types, for example, aspirated sounds, while Japanese does not (excepting some individual cases). Some Chinese teachers have commented on the difficulty of the Japanese students in pronouncing aspirated sounds correctly. So that they are well distinguished, aspirated affrications are especially difficult.

Kobozono provided an overview of the research on the acquisition of phonology and language universals, in particular the acquisition of vowels, consonants, syllable structure and prosody, mainly for Japanese learner [5].

It was reported that it is difficult to utter the vowel /an/ by Japanese students as compared with the Chinese vowel /ang/ [6]. On the other hand, some reports have made about Chinese aspirated consonants [6-8]. However, these papers are only discussed to aspirated plosive consonants (/p/, /b/) pair or / and (/t/, /d/) pair. The pair of (/k/, /g/) is discussed rarely. The aspirated affrication is more difficult to pronounce when we got a pronounce experiment for Japanese students [9].

Our goal is to provide feedback to learners as they learn how to improve their pronunciation skills. In first, this paper conducts an auditory evaluation. Then we analyzed the differences in VOT and energy of aspirated sounds and unaspirated sounds, mainly affrications, to determine their characteristics using some proposed parameters. Finally, a discrimination experiment is also performed.

2 Speech Materials

Four native Japanese university students participated in this study. They were first year under-graduates; time spent learning Chinese phonology was only three months. The speech material included 72 two-syllable significant words. The number of words with plosive consonant is 38, and it has 46 words with affrication consonant. The sequence of the consonant in a word is not taken into consideration. One native Chinese teacher also uttered the same material. The detailed numbers of plosive and affrication consonants of aspirated and unaspirated sounds are shown in Table 1 and Table 2 with their perception results in next

subsection. Three of them appear in study of affrications.

The materials were recorded in a sound proof room inside the university. The equipment included a 24 track digital recorder D24224LV (FOSTEX) and a microphone MD421-II (SENNHEISER). Speech samples recorded at 48 kHz sampling rate with 24 bit precision and later down-sampled to 16 bits at 16 kHz.

3   Perception Results

First, we checked pronunciation quality. The native Chinese teacher evaluated each learner's utterance. Each syllable was assessed as being right, vague, or wrong. The criteria are based on those consonants.

The evaluation results are shown in Table 1 and Table 2, corresponding to plosive sounds and affrications, respectively. In here, "Clear" in the tables is corresponding to evaluation of "right". On the other hand, the term of "Unclear" is corresponding to evaluation of "vague" and "wrong". Table 1 is shown that these students pronounced clearly the unaspirated plosive sounds with 99.4%, but it is 85.5% for the aspirated plosive sounds. This means that pronunciation of the aspirated plosive sounds is more difficult. It is same tendency as that in table 2, but articulation fell greatly for affrications. The articulation of unaspirated affrications is 84.8%, it was not able to utter correctly about 15%. Aspirated affrications cannot be uttered correctly 40%. It has appeared well for utterance of aspirated affrications to be difficult.

Table 1. Number of plosive sounds and its perception results (Four speakers)

	Aspirated plosive sounds				Unaspirated plosive sounds				Total
	/p/	/t/	/k/	Subtotal	/b/	/d/	/g/	Subtotal	
Clear[%]	90.0	74.1	94.1	85.5	100.0	98.2	100.0	99.4	92.8
Unclear[%]	10.0	25.9	5.9	14.5	0.0	1.8	0.0	0.6	7.2
Number	40	54	51	145	48	55	59	162	307

Table 2. Number of affrication sounds and its perception results (Three speakers)

	Aspirated affrications				Unaspirated affrications				Total
	/c/	/ch/	/q/	Subtotal	/z/	/zh/	/j/	Subtotal	
Clear[%]	45.8	47.6	72.7	59.6	83.3	72.7	93.8	84.8	73.2
Unclear[%]	54.2	52.4	27.3	40.4	16.7	27.3	6.3	15.2	26.8
Number	24	21	44	89	24	33	48	105	194

In next sections, those syllables with clear pronunciation were used for the study of analysis and discrimination. This means that the number of consonants differs for each speaker.

4   Acoustical Analysis

4.1   Speech Characteristics

Figures 1 (a) and (b) show, respectively, the examples of spectrum and waveform of the syllable /cao/ and /zong/ with affrications which spoken by the teacher. From top to bottom, these figures show the semi-syllable label, and spectrum and speech waveform. In order to distinguish the unaspirated or aspirated consonant, the parameter VOT (Voice Onset Time) is used usually. This is the time length from a burst point in time to a start time of vibration of the vocal cord. The length of VOT becomes long in order that a "breath" may follow an aspirated sound on utterance. On the other hand, since unaspirated sound does not have the interval accompanied by a breath, the length of VOT becomes short. Therefore, the existence of the interval by which a breath is accompanied can be checked by investigating the length of VOT. This can be checked from Fig. 1. Although VOT is a good parameter, it is still imperfect to the purpose achievement.
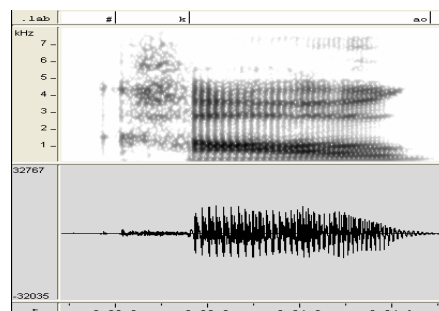
(a) Aspirated affrication /k/
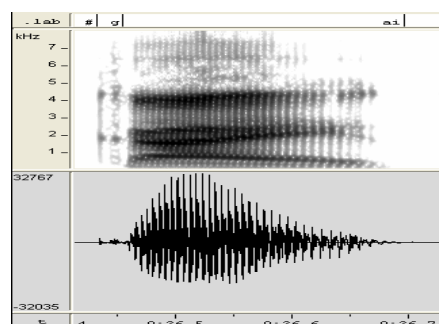
(b) Unaspirated affrication /g/

Fig.1 Example of affrications uttered by teacher

If the spectral are seen in detail, it turns out that distribution of energy differs on both a time axis and a frequency axis. Some parameters which approximate these properties in time domain and frequency domain could be devised.

4.2 VOT

Figure 2 and figure 3 are shown the average length of VOT of the plosive and affricative consonants, respectively. These words are uttered by teacher and four students or three students which are indicated above. From figure 2, average VOT is different between aspirated plosive consonant /g/ and unaspirated plosive consonant /k/. The difference in VOT is 2.8 to 5.6 times. It is also large than other plosive consonant pairs. On the other hand, the difference in VOT is 1.4 to 2.2 times for affrications of /c/ and /z/. It becomes small. Considering the variation of VOT, distinguishing them may be difficult in this case.
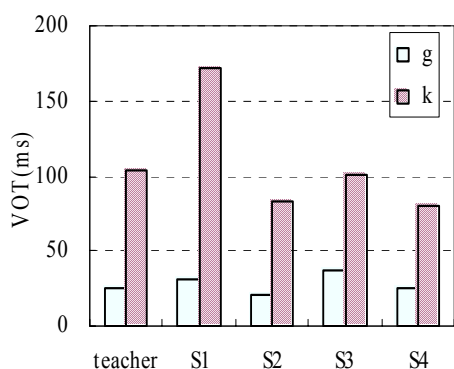

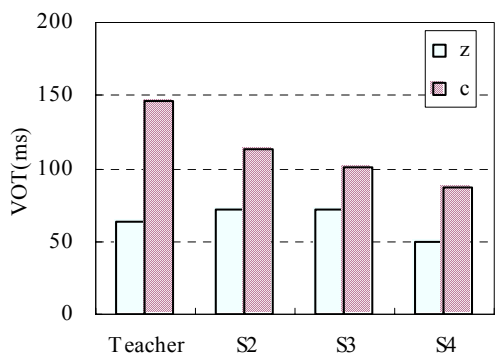**Fig.2** Average VOT of plosive consonants (/g/ and /k/)

Fig.3 Average VOT of affrications (/z/ and /c/)

4.3 Number of the Energy Peaks (NEP)

If the spectrum of aspirated and unaspirated consonant is observed in detail, the difference in the strength of energy will be seen by each frequency band. Because the energy is variable in each frequency area for different consonants, the band is set up separately. Here, we select six bands within 8000 Hz frequency area. For example, the band of /c/and /z/ is divided into [0, 600, 1600, 4000, 5000, 6000, 8000] Hz, and it is

called band 1, band 2, to band 6 from a low frequency band, respectively.

The example of the energy variation in band 1 of /g/ and /k/ uttered by teacher is shown in Fig.4. It was difficult to discriminate from /k/ and /g/ by exploiting only the average energy of each band. If we pay attention to fluctuation, it is clear that there were some peaks appeared. It is also shown that the number of peaks differs. In order to calculate correctly the Number of the Energy Peak (NEP), the variation curvilinear was smoothed in Fig. 4. This NEP is used for the further study as a parameter.
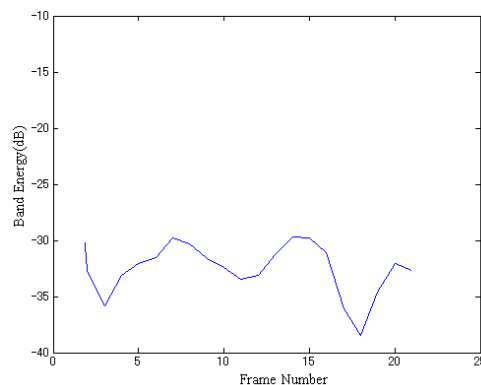
(a) Unaspirated consonant /g/
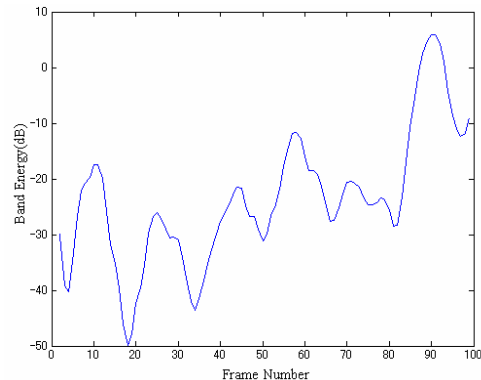
(b) aspirated consonant /k/

Fig 4. An example of the varied energy of band 1 (uttered by teacher)

Figure 5 is shown the difference of NEP for band 2 in case of affrication /c/ and /z/. Aspirated sound /c/ has larger value than unaspirated sound /z/.

4.4 Position of Maximum Energy Peak (PMEP)

The position which has maximum energy of peak (PMEP) is also different for aspirated and unaspirated consonants. In figure 4(b), corresponding to aspirated /k/, the position is in the right side, and PMEP is large

when it is compared with unaspirated /g/ in figure 4(a). This is seen also by affrications.

We also use the energy difference from 64 bands to indicate the characteristics of aspirated and unaspirated sounds.
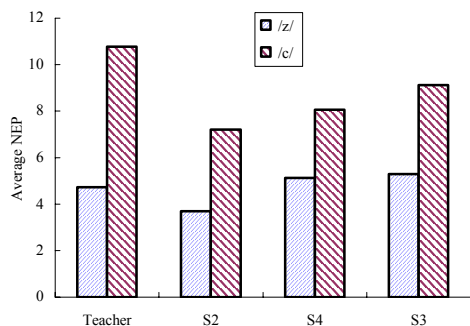

Fig. 5 The result of NEP in band2

5 Discrimination experiments

In order to see the validity of the proposed parameter, the discrimination experiment was conducted. The DLP (Dynamic low-passed power) method [7] in the conventional method was also examined. The rates of discrimination of 6 explosive sounds ([p'], [p], [k] and [b'], [b], [g]) will be increased from 92.4% (using the conventional method: VOT and DLP) to 94.7% (only add the parameter NEP). The rate is improved from 89.4% to 93.7% for all affricated unaspirated / aspirated sounds ([z], [zh], [j] and [c], [ch], [q]). The maximum rate of 95.1% is obtained when 5 parameters are all used.

6 Summary

This paper examined the problem of how to assist Japanese students when learning to speak Chinese. A key difficulty is correctly pronouncing the aspirated / unaspirated affrications. We conducted auditory tests (subjective) and found that three Japanese subjects found aspirated affrications /c/, /ch/ and /q/ to be more difficult to pronounce than unaspirated affrications and plosive sounds.

The results of acoustic analysis showed the utility of using the difference in the energy of some frequency bands to distinguish the pronunciation of aspirated sounds and unaspirated sounds. We proposed two parameters to present the difference between aspirated and unaspirated sounds, NEP (Number of the Energy Peak) and PMEP (Position of Maximum Energy Peak). The rate of discrimination of 6 explosive sounds was increased from 92.4% to 94.7% (when using the parameter NEP). The rate is improved from 89.4% to

93.7% for all affricated unaspirated / aspirated sounds when proposed parameter is used.

References

[1] Yongbing Chen, Renhua Wang, "Speech signal processing", Chinese technological university publishing company, 1990 (in Chinese)

[2] Rijia Wang, "A study of the word education", Beijing university publishing company and 1992 (in Chinese)

[3] Susan G. Guion, James E. Flege, Reiko Akahane-Yamada and Jesica C. Downs-Pruitt, "Acquisition of english phonetic categories by Japanese speakers: Evidence from categorical discrimination tests," Proc. Cong. Acous., 135th MASA, Vol.4, pp.2973-2974, (1998)

[4] Reiko Akahane-Yamada, "Learning non-native speech contrasts: What laboratory training studies tell us," Proc. ASA/ASJ Third Joint Meeting, 3pSC21, pp.953-958 (1996.12)

[5] Reiko A. Yamada, Winifred Strange, James S. Magnuson, John S. Pruitt, William D. Clarke III., "The Intelligibility of Japanese Speakers' Productions of American English /r/, /l/ and /w/, as Evaluated by Native Speakers of American English," Proceedings of the 1994 International Conference on Spoken Language Processing, Yokohama, pp.2023-2026, (1994.9).

[6] Haruo Kubozono, "Acquisition of phonology and language universals," Journal of the Phonetic Society of Japan, Vol.7, No.2, pp.5-17 (2003) (in Japanese)

[7] J. Miwa, "The distinction of some mind sound," Electronic information communication academic meeting thesis journal, Vol.J85-D-II and pp.842-850.2002 (in Japanese)

[8] J. Tian, J. Miwa, "A discrimination of Chinese aspirated/unaspirated sounds for speech learning system," IEICE, Vol.J85-D-II, No.5 pp.459-468(1992) (in Japanese)

[9] A. Hosino, A. Yasuda, "Evaluation of Chinese aspiration sounds uttered by Japanese students using VOT and power," ASJ, Vol.58 pp.689-695(2002) (in Japanese)

[10] T. Tokoro, Y. Kanamori, "The feature extraction and discriminant to aspirated and unaspirated affrication sounds uttered by Chinese learners," General conf. of IEICE, D1-4-9(2005) (in Japanese)