International Symposium on Chinese Spoken Language Processing (ISCSLP 2004)
Hong Kong
December 15-18, 2004

DEPENDENCE OF CORRECT PRONUNCIATION OF CHINESE ASPIRATED SOUNDS ON POWER DURING VOICE ONSET TIME

*Akemi Hoshino and Akio Yasuda**

Toyama National College of Maritime Technology, Toyama 933-0293, Japan
hoshino@toyama-cmt.ac.jp
*Tokyo University of Marine Science and Technology, Tokyo 135-8533, Japan

ABSTRACT

The length of voice onset time (VOT) in uttering Chinese aspirated sounds, which are difficult for Japanese to pronounce, is an important factor in evaluating the quality of pronunciation. In this paper, both the length of the VOT and the power used during the VOT for 21 single-vowel syllables of six different Chinese aspirates were measured for 40 Japanese students and nine native speakers of Chinese. The quality of the students' pronunciation was evaluated using a hearing test judged by eight native Chinese. The results indicated that the correlation between the quality of the students' pronunciation and the power used in uttering a sound was greater than to the VOT within a certain range of VOT which varied for different syllables. Thus, we conclude that power is also an important factor in evaluating the quality of pronunciation.

1. INTRODUCTION

The number of students learning Chinese in Japan has increased in recent years along with the development of the Chinese economy. There are many different sounds in Chinese pronunciation. As most of them are quite different from Japanese sounds, many Japanese students have difficulty in pronouncing them, especially aspirated sounds. Uttering aspirated syllables requires the speaker to exhale. As Japanese has no aspirated sounds, Japanese students attempt to imitate the sounds made by their native Chinese teachers, but many of them are unable to pronounce them correctly. Recognizing aspirated sounds is also difficult for them.

There are 21 aspirated Chinese sounds of single-vowel syllables of six different aspiration sounds: bilabial (p[p‘]), alveolar (t[t‘]), velar (k[k‘]), palatal (q[tɕ‘]), retroflex (ch[tʂ‘]) and dental (c[ʦ‘])[1], as shown in table 1. We analyzed the VOT and the power used during the VOT for four bilabial syllables in our previous study[2][3] and showed that the quality of pronunciation depended not only on the VOT but also on the power used during the VOT. In this study, we analyzed another 17 single-vowel syllables of five different aspirated sounds pronounced by nine native Chinese speakers and 40 Japanese students, who had studied Chinese for 3 hours per week for one year, and again found that the quality of pronunciation depended not only on the VOT but also on the power used during the VOT as reported for bilabial sounds in the previous papers.

2. DIFFERENCES BETWEEN ASPIRATED AND UNASPIRATED SOUNDS

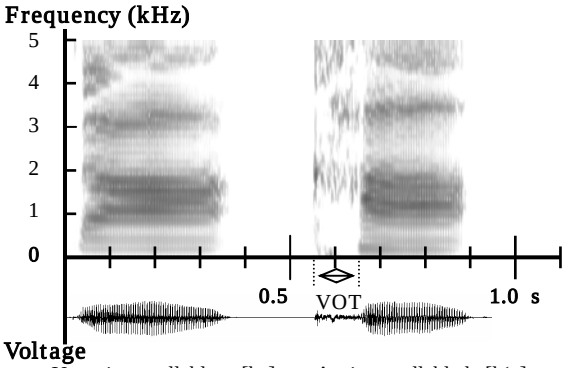

Figure 1: Spectrograms of unaspirated syllable ga[ka] (left), and aspirated syllable ka[k'a] (right) pronounced by Chinese speakers.

Figure 1 shows the air vibrations of uttered sounds (lower) and spectrograms of the unaspirated syllable, ga[ka], (left) and the aspirated syllable, ka[k'a], (right). The darker the horizontal bands, the higher the power of the frequency components. The aspirate appears in a brief interval in the right spectrogram, indicated by vertical stripes, between the stop burst and the onset of vocal fold vibrations followed by a vowel. This time interval is called the voice onset time (VOT)[4]. The onset of the vocal fold vibration is so close to the burst in the left spectrogram that no aspiration interval appears. These data were acquired and analyzed using a tool of Multi-Speech (Model 3700, Kay Elemetrics Corp., USA).

3. METHOD USED TO EVALUATE PRONUNCIATION

We calculated the VOT and the relative average power during VOT from spectrograms using the procedure reported previously[2][3]. The sounds uttered by 40 Japanese students were ranked in a hearing test of the reproduced sounds conducted by eight native Chinese speakers. The grades were as follows: 3 = pronunciation in which the aspirate was sounded; 2 = unclear sounds; and 1 = unaspirated sounds. The examiners checked with each other that their pronunciations were perfectly aspirated. Data were excluded in cases of split evaluations and a standard deviation larger than 0.64, broken sounds uttered very close to the microphone, and sounds with a low S/N uttered away from the microphone. An average grade of more than 2.6 was defined as a good pronunciation, five of the examiners awarded a 3 and three examiners awarded a 2.

4. COMPARION BETWEEN VOTs of CHINESE SPEAKERS AND JAPANESE STUDENTS

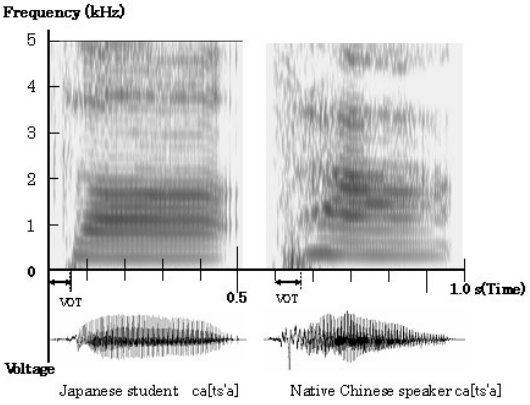

Figure 2: Spectrograms of dental aspirate syllable ca[ts'a] pronounced by native Chinese speaker (right) and Japanese student (left).

It is generally agreed that when an aspirated syllable is pronounced with a brief VOT it sounds unaspirated, and when it is pronounced with a long VOT it sounds aspirated[5]. However, an investigation of all single-vowel aspirated syllables in Chinese shows that there are some exceptions. Figure 2 shows spectrograms of the dental aspirated syllable, ca[ts'a], pronounced by a Japanese student (left) with a VOT of 69 ms and by a Chinese speaker with a VOT of 78 ms. Although the difference was just 9 ms, the student's pronunciation received a low grade of 1.2.

The example, also found for other aspirated syllables, shows that the VOT is not the only relevant factor in evaluating the pronunciation of aspirated syllables, although it is closely related to the grade received.

5. DEPENDENCE OF EVALUATION ON AVERAGE POWER

Table 1: Averaged VOT and power of aspirated syllable of pronunciation by Japanese student and Chinese native speakers. Grade is just for the students.

	Syllable		VOT(ms)	Power	Grade
Bilabial	pa[p'a]	Student	32.24	32.24	2.7
		Chinese	71.12	38.7	
	po[p'o]	Student	43.88	15.47	2.6
		Chinese	65.04	24.78	
	pi[p'i]	Student	37.19	42.57	2.6
		Chinese	64.09	16.28	
	pu[p'u]	Student	59.58	20.7	2.5
		Chinese	59.91	47.56	
Alveolar	ta[t'a]	Student	46.63	3.60	2.6
		Chinese	76.56	5.12	
	te[t'ɤ]	Student	48.78	1.24	2.2
		Chinese	61.23	4.86	
	ti[t'i]	Student	54.66	1.22	2.7
		Chinese	87.25	0.87	
	tu[t'u]	Student	51.62	2.18	2.3
		Chinese	77.58	4.70	
Velar	ka[k'a]	Student	71.37	0.23	2.8
		Chinese	81.9	0.27	
	ke[k'ɤ]	Student	77.6	0.08	2.4
		Chinese	99.51	0.14	
	ku[k'u]	Student	85.53	0.07	2.5
		Chinese	90.0	0.82	
Palatal	qi[tɕ'i]	Student	109.23	0.21	2.6
		Chinese	122.9	0.61	
	qu[tɕ'u]	Student	116.96	0.21	2.3
		Chinese	147.04	0.8	
Retroflex	cha[tʂ'a]	Student	64.83	0.42	2.3
		Chinese	122.83	1.21	
	che[tʂ'ɤ]	Student	82.13	0.18	2.0
		Chinese	155.64	0.19	
	chi[tʂ'i]	Student	118.0	0.44	2.2
		Chinese	132.47	0.18	
	chu[tʂ'u]	Student	106.36	0.26	2.0
		Chinese	181.38	1.10	
Dental	ca[ts'a]	Student	95.92	0.1	2.1
		Chinese	106.92	0.37	
	ce[ts'ɤ]	Student	71.2	0.013	2.0
		Chinese	119.08	0.029	
	ci[ts'i]	Student	101.58	0.014	2.2
		Chinese	146.76	0.103	
	cu[ts'u]	Student	86.71	0.35	2.1
		Chinese	151.93	4.74	

In the present study, we found some cases in which student pronunciation of aspirated sounds received a low grade even when the VOT was almost the same as, or longer than that of a Chinese speaker. To find the reason for this, we calculated the relative average power during the VOT[2][3] of the sounds of all single-vowel aspirated syllables in Chinese uttered by 40 Japanese students and nine native Chinese speakers and examined the dependence of the

evaluation on them.

Table 1 summarizes the average VOT and relative average power for all the single-vowel aspirated syllables. The average student VOT for the syllable, pu[p‘u], had the least difference from that of the Chinese speakers for the four bilabial syllables. However, these syllables received a relatively low average grade of 2.5. The average power was low, and there was a large difference between the power of the Chinese speakers and that of the students. In the case of the alveolar aspirated syllable, te[t‘ɤ], the average power of the students showed the largest difference from that of the Chinese speakers for the four syllables, and the pronunciation received a lower grade of 2.2. The students received a high average grade of 2.8 for ka[k‘a] of the three velar syllables and their average power was similar to that of the Chinese speakers for this sound. The students’ pronunciation of the other aspirated sounds, which had a similar average power to that of the Chinese speakers, also received good grades.

These examples indicate that the VOT is not the only factor to consider in evaluating the pronunciation of aspirated syllables, although it is closely related to the grade received.

5.1 Relationship Between Grade and Relative Average Power

Figures 3 and 4 show the distributions of data for two Chinese aspirated sounds, tu[t‘u] and cu[ʦ‘u]; the abscissa represents the VOT, and the ordinate represents the relative average power. The average grade was added to the students’ individual marks. Points were also plotted for the Chinese speakers to provide a reference.

5.1.1 Dependence of Grade on Relative Average Power

Figure 3 shows the data for the alveolar aspirate, tu[t‘u]. D1, located at the top left, has a VOT of 30 ms and a relatively high power of 30.7; the grade was 2.7. D2 and D3, below D1, had a longer or almost the same VOT as D1; however, they had a lower power of 1.2 and 0.77, and received low grades of 2.6 and 1.9, respectively. The second data sample from the left at the top, D8, had a briefer VOT of 25 ms compared to 36 ms for D7, located at the bottom. However, it had 127 times more power than D7 and received a much higher grade of 2.6 compared to 1.0 for D7. These data show that syllables pronounced with more power received a higher grade. In this assessment, the student data is grouped on the lower left of the figure, while the Chinese data is grouped on the upper right. This clearly shows that the power used by the students was lower than that used by the Chinese speakers, and this was responsible for the low average grade of 2.3 in Table 2. In this case, if the VOT was between 21 and 55 ms, pronunciation with a higher power received a higher grade.

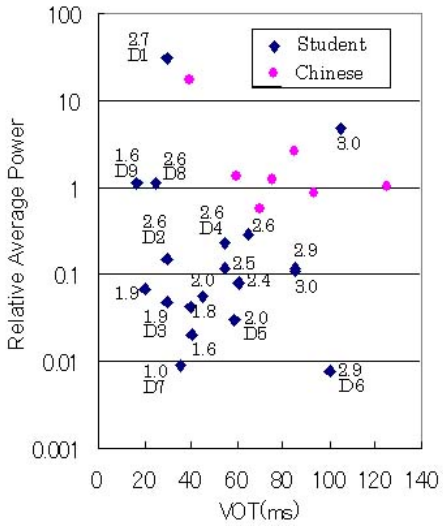

Figure 3: Distribution of data for alveolar aspirated syllable tu[t‘u]; on the surface of VOT on abscissa and relative average power on ordinate.

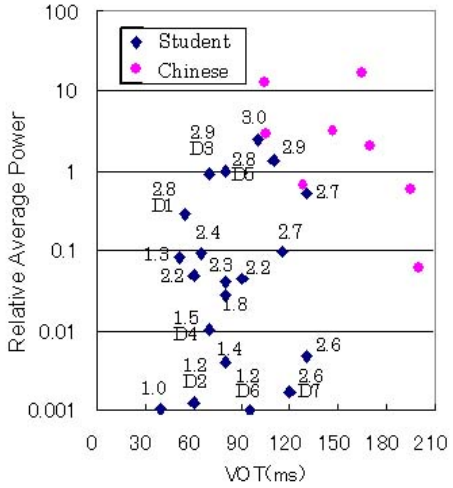

Figure 4: Distribution of data for dental aspirated syllable cu[ʦ‘u] on the surface of VOT on abscissa and relative average power on ordinate.

Figure 4 shows the data for the pronunciation of the dental aspirated syllable, cu[ʦ‘u]. D1, located in the upper left of Figure 4, had a VOT of 55 ms and a power of 0.29. Although D2, located second from left, had a 5-ms longer VOT than D1, it received a grade of 1.2 because the power was about 1/290 that of D1. Although D5, located third from the top, had a shorter VOT than D6, located at the bottom, it

received a higher grade (2.8 compared to 1.2) than D6 because of its higher power. The Chinese data are grouped in the upper right of the figure. The students' pronunciation had a shorter VOT and lower power than that of the Chinese speakers.

The above examples show that pronunciation with higher power received higher grades, even when the VOT was nearly equal to, or shorter than that for sounds uttered with lower power. The results of assessing the pronunciation of 21 single-vowel syllables of six different aspirated syllables showed that the grades for pronunciation with a range of VOT, as shown in Table 2, depended less on the length of the VOT and more on the average power used for breathing during the VOT.

5. 1.2 VOTs of Very Short and of Adequate Length

Some examples showed that there was no dependent relationship if the VOT was very short. The grade for the pronunciation of the alveolar aspirated syllable, tu[t'u], at the far left of Figure 3, was 1.6 for D9, with a VOT of 17 ms. Although D8, next on the right, had the same power as D9, it received a higher grade (2.6) than D9.

In other examples with a longer VOT, the grade did not depend as much on breathing power. As Figure 3 shows for the alveolar aspirated syllable, tu[t'u], D6, located at the bottom right, had a similarly low power to D7 at the bottom left, but it received a good grade for pronunciation (2.9) with a long VOT of 100 ms.

For the dental aspiration, cu[ʦ'u], D7, located at the bottom right of Figure 4, had a slightly higher power than that of D6. But D7, which had a longer VOT, received a higher grade. As this figure shows, pronunciation with a VOT longer than 110 ms received grades above 2.6, which is considered a good grade.

As shown above, grades did not correlate closely with power if the length of the VOT was shorter or longer than certain values. The same tendency was also observed in other syllables.

5.2 Correlation between Grade and Evaluation Variables

As we showed in Section 5.1.1, grades for the pronunciation of aspirated sounds did not always depend on the length of the VOT in some specific ranges of the VOT. Coefficients for the grades and the variables (length of VOT, and average relative power during the VOT) are summarized in Table 2. The result of bilabial syllables is in the last paper[3]. These show that the correlation for power is superior to for VOT in the ranges denoted in the table.

Table 2: Correlations between grades and the evaluation variables.

	Syllable	VOT Range	Data#	Cor. To VOT	Cor. To Power
Alveola	ta[t'a]	25–48ms	11	0.19	0.60
	te[t'ɤ]	29–44ms	8	0.40	0.77
	ti[t'i]	33–50ms	9	0.23	0.44
	tu[t'u]	21–55ms	11	0.12	0.49
Velar	ka[k'a]	30–55ms	8	0.52	0.86
	ku[k'u]	35–60ms	5	0.56	0.77
	ke[k'ɤ]	35–60ms	8	0.16	0.65
Palatal	qi[ʨ'i]	75–105ms	11	0.45	0.64
	qu[ʨ'u]	75–115ms	9	0.38	0.73
Retroflex	cha[ʧ'a]	45–75ms	14	0.12	0.65
	che[ʧ'ɤ]	75–120ms	16	0.23	0.62
	chi[ʧ'i]	66–120ms	14	0.01	0.72
	chu[ʧ'u]	60–145ms	21	0.32	0.67
Dental	ca[ʦ'a]	60–85ms	10	−0.03	0.55
	ce[ʦ'ɤ]	70–90ms	7	−0.13	0.95
	ci[ʦ'i]	60–110ms	12	−0.42	0.74
	cu[ʦ'u]	60–100ms	13	−0.02	0.65

6. CONCLUSION

To develop methods for teaching students how to pronounce Chinese aspirated syllables correctly, we attempted to establish some evaluation measures. We examined the VOT and average power used during the VOT for 21 single-vowel syllables of six aspirated sounds uttered by nine native Chinese speakers and 40 Japanese students. Grades for the pronunciation of each sound were determined by a hearing test of eight native Chinese speakers. When the length of the VOT for the Chinese aspiration was within a certain range, pronunciation made using higher power received a higher grade. The results indicate that the quality of the pronunciation of aspirated sounds depends not only on the VOT but also on the power used during the VOT.

7. REFERENCES

[1] T. Inuishi, 'Point Learning-Chinese for Beginners' (in Japanese), Toho Publishing Corp, Tokyo, 1997.

[2] A. Hoshino and A. Yasuda, 'Evaluation of Chinese aspiration sounds uttered by Japanese students using VOT and power (in Japanese),' The Journal of the Acoustical Society of Japan, Vol. 58, No. 11, pp. 689-695, Nov. 2002.

[3] A. Hoshino and A. Yasuda, 'Evaluation of Chinese aspiration sounds uttered by Japanese students based on VOT and power,' Proceeding of ICASSP 2003, pp.472-475, Hong Kong, 2003.

[4] Ray D. Kent and Charles Read, 'The Acoustic Analysis of Speech,' Singular Publishing Group, Inc., San Diego and London, p. 107, 1992

[5] Zhu Chuan, 'Studying Method of the Pronunciation of Chinese Speech for Foreign Students (in Chinese)', Yu Wu Publishing Co., China, pp. 63-71, 1997