THE EVALUATION OF CHINESE ASPIRATION SOUNDS UTTERD BY JAPANESE STUDENTS USING VOT AND POWER

*Akemi Hoshino and Akio Yasuda**

Toyama National College of Maritime Technology
1-2 Neriya, Ebie, Shinminato City, Toyama 933-0293, Japan
hoshino@toyama-cmt.ac.jp
*Tokyo University of Mercantile Marine

ABSTRACT

A Chinese aspiration is generally considered to be very difficult to be reproduced by Japanese students. The difficulty is caused by lack of aspiration in Japanese pronunciation. Voice onset time (VOT) is one of the important measures to evaluate the quality of pronunciation. In this paper, VOT for Chinese aspiration sounds pa[p'a], pi[p'i], po[p'o] and pu[p'u] were measured, as they were pronounced by 40 Japanese students, who have been studying Chinese 3 hours per weak for one year. The uttered sounds were also evaluated by native Chinese speakers. The paper demonstrates that good pronunciation has generally long VOT. However, some exceptions were observed. Furthermore analysis of temporal variations of breathing power during the VOT period demonstrates that the power is low in the lower grade sound. The paper shows that the VOT is not only one measure of clear pronunciation of the Chinese aspirations. Other measures, such as the power and the energy, have been introduced and their importance as a means to evaluate the quality of pronunciation have been confirmed.

1. INTRODUCTION

There are many kinds of sounds in Chinese pronunciation. Most of them are peculiar in Japanese sounds. Uttering aspirated syllables requires exhaling. As Japanese has no aspirated sound, the Japanese students utter aspirated sounds after the native Chinese teacher, but many of them cannot pronounce the correct sounds. Recognition of aspirated sounds is difficult for them too.

In order to develop the instruction device for the pronunciation of aspirated syllables, we tried to establish evaluation measures.

In this paper, we analyze the labial sounds of [p'a], [p'i], [p'o] and [p'u] among Chinese aspiration sounds uttered by 9 native Chinese speakers and 40 Japanese students and show that the quality of the pronunciation depends not only on VOT but also on the power during VOT.

2. DIFFERENCE BETWEEN ASPRIRATED AND UNASPRIRATED SOUNDS

Figure.1 shows spectrograms of the unaspirated syllable ba[pa] (left) and aspirated syllable pa[p'a] (right). The aspiration appears in a brief interval, on the right hand side spectrogram, between the stop burst and the onset of vocal fold vibrations followed by a vowel [1]. This time interval is called the voice onset time (VOT). The onset of vocal fold vibration is so close to burst that the interval of aspiration does not appear on the left hand side spectrogram.

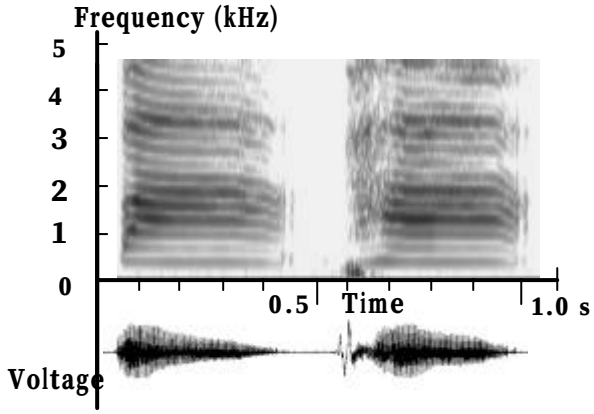


Figure 1: Spectrograms of unaspirated syllable ba[pa] (left) and aspirated syllable pa[p'a] (right).

3. MEASUREMENT OF VOT

3.1. Comparison Between VOTs of Native Chinese Speakers and Japanese Students

Figure 2 shows the spectrograms of the aspirated syllable pa[p'a] of the native Chinese speaker (left) and the Japanese student (right). The VOT is 60.3 ms on the left hand side for native Chinese. The VOT is only 11.9 ms on the right hand side for the Japanese student. It is quite brief compared with that of the native speakers.

Table 1 is the measured VOT of aspirated syllables of labial sounds, pa[p'a], pi[p'i], po[p'o] and pu[p'u] pronounced by 9 native speakers. Table 2 is those by 40

Japanese students who have been studying Chinese 3 hours per week for one year. We chose 44 pieces of data out of all their pronunciations based on the following conditions described in 3.2.

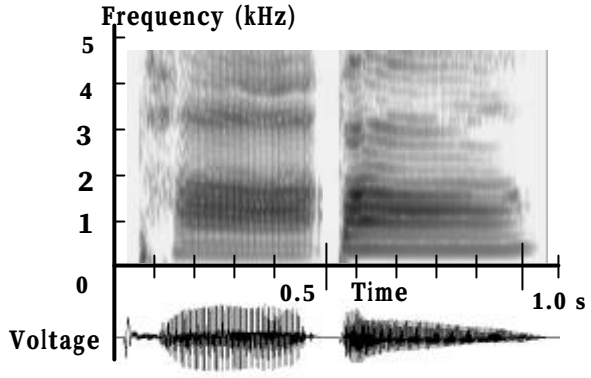


Figure 2: Spectrograms of aspirated syllable pa[p'a] of native Chinese speaker (left) and Japanese student (right).

Table 1: Measured VOT of aspirated syllable of labial sounds, pa, pi, po and pu pronounced by 9 Chinese native speakers

Syllable	pa[p 'a]	po[p 'o]	pi[p 'i]	pu[p 'u]
	(ms)	(ms)	(ms)	(ms)
Chinese 1	89.87	72.25	164.90	75.01
Chinese 2	77.51	53.25	89.32	81.42
Chinese 3	60.76	69.93	70.00	50.50
Chinese 4	76.46	85.00	90.07	40.36
Chinese 5	94.96	111.21	60.95	132.20
Chinese 6	75.01	64.94	89.98	67.57
Chinese 7	44.98	37.56	37.50	40.10
Chinese 8	65.04	34.92	40.00	75.37
Chinese 9	51.49	75.01	60.05	51.07
Average	70.68	67.12	78.09	68.18

The average lengths of VOT of the sounds pa[p'a], pi[p'i], po[p'o] and pu[p'u] of the Japanese students are 33.9 ms, 43.9 ms, 37.2 ms and 61.3 ms respectively and are a half of those by native speakers except for the sound pu[p'u]. This is considered to be a major reason why the aspirated pronunciation by the Japanese students sounds like an unaspirated one.

Table 2 : Measured VOT of aspirated syllable of labial sounds, pa[p'a], pi[p'i], po[p'o] and pu[p'u] pronounced by 40 students. (D:Data, G:Grade)

Pa	VOT		Po	VOT		Pi	VOT		Pu	VOT	
D	(ms)	G	D	(ms)	G	D	(ms)	G	D	T(ms)	G
D1	56.87	1.9	D12	35.01	2.6	D23	35.02	3.0	D34	39.12	2.5
D2	9.97	2.1	D13	35.02	2.5	D24	34.94	2.5	D35	95.98	2.7
D3	30.2	3.0	D14	49.98	3.0	D25	54.38	3.0	D36	70.22	3.0
D4	60.05	3.0	D15	64.97	2.9	D26	54.45	2.6	D37	64.24	2.6
D5	70.03	3.0	D16	19.95	2.4	D27	19.95	1.8	D38	18.74	1.3
D6	24.94	2.9	D17	14.97	1.0	D28	9.97	1.0	D39	70.02	2.6
D7	25.04	2.9	D18	49.98	3.0	D29	40.00	3.0	D40	60.05	2.3
D8	14.73	2.4	D19	50.00	2.7	D30	43.27	3.0	D41	35.60	2.9
D9	31.84	3.0	D20	42.00	2.7	D31	35.00	3.0	D42	78.70	2.9
D10	21.69	1.3	D21	80.00	3.0	D32	50.00	2.7	D43	54.19	2.9
D11	27.27	3.0	D22	40.82	2.9	D33	32.11	3.0	D44	87.62	3.0
Av.	33.88			43.88			37.19			61.32	

3.2 Grade Dependency of Pronunciation on VOT

We examine here the dependency of the grade of the pronunciation on VOT. 8 native Chinese speakers joined the hearing test as examiners for the 40 students' pronunciations. They put the mark 3 in the pronunciation which sounds aspirated, mark 1 in the unaspirated sounds and mark 2 in the unclear sounds. We collected data for table 2 by excluding the pronunciations of which evaluation splits largely and standard deviation is larger than 0.64 (one quarter among the total data), broken ones uttered very close to the microphone and ones with low S/N uttered very far from the microphone (one quarter) and ones of which VOT is long enough to get a good mark (one quarter). Table 2 shows the VOT and the average evaluation of the students. The average mark of the good pronunciation is larger than 2.6 : 5 of the examiners gave a 3 mark and 3 examiners gave a 2 mark.

It is generally said that the pronunciation of an aspirated syllable with brief VOT sounds unaspirated and that with a long VOT sounds aspirated [2]. We can find, however, some exceptions in table 2. In the pronunciation of the aspirated syllable pa, the average grade of D1 is 1.9, although VOT is 56.9 ms. The grades are 2.9 and 3.0 of D3, D6, D7, D9 and D11, although their VOTs are briefer than that of D1. In the case of po, the grades of D16 and D17 are 2.4 and 1.0 respectively and differ from each other largely, although the VOT difference is just 5.0 ms between them.

In the case of pi, the grades of D23 and D24 are 3.0 and 2.5. The VOTs are almost same. In the case of pu, the grade of D43 is better than that of D40, although VOT is a little bit less than that of D40.

These exceptions show that VOT is not a sole measure to evaluate the pronunciation of the aspirated syllables, although it is closely related to the grade.

4. AVERAGE POWER AND ENERGY DURING VOT AND EVALUATION

In the former chapter, we showed the exceptions in which the grade of the pronunciation of aspirated syllables does not depend on VOT. In order to find the reason, we measured the average power during VOT and examined the dependency of the evaluation.

4.1 Deduction of Relative Average Power

The tool used for speech analysis converted the voltage signal picked up by a microphone to the digital signal with a sampling frequency of 11kHz and a dynamic rang of 16 bits. It deduces the power by taking the average value V_a of the continually sampled values V(t) as a standard. Then the power P(0.005m) at every 5ms is given by

$$P(0.005m) = 20\log_{10}\left(\frac{V_a(0.005m)}{300}\right). \quad (1)$$

Where m and n are integers, $V_a(t)$ is

$$V_a(0.005m) = \frac{\sum_{n=55m+1}^{n=55m+55} |V(n) - V_A|}{55}, \quad (2)$$

and set 300 of digital value as 0 dB[3].

We deduced the total energy in VOT, dividing the P(t) in Eq.(1) by 10, converting reversely to the real value by anti-logarithmic calculation and multiplying 5ms. The energy deduced here however is arbitrary. Thus we name it W_{nom}.

Then we normalized it by the following procedure. We deduce the ratio R of P_1, maximum power inside VOT, against P_2, maximum power inside the voiced period.

$$r\ (dB) = P_1 - P_2\ (dB) \quad (3)$$

$$R = 10^{r/10} \quad (4)$$

Multiplying R and W_{nom}, we normalized the energy in VOT against that during the voiced period.

$$W_{rel} = W_{nom} \times R \quad . \quad (5)$$

Dividing W_{rel} by VOT, we obtain the relative average power P_{rel}.

$$P_{rel} = W_{rel} / VOT \quad . \quad (6)$$

4.2 The Consideration on Relation Between Grade and Average Power

Next we deduced the relative average power of equation (6) for the students' data in table 2. Figures 3 to 6 show the data distribution on the surface. The abscissa represents VOT and the ordinate relative average power. The average grade was added to the points of students. The points are also plotted for the native speakers as the reference.

4.2.1 Case of Grade Depending on Relative Average Power

D10, located the lowermost part in figure 3 of the aspirated syllable of pa, which has a longer VOT than D8 and almost same VOT to D6 and D7. But it has a lower power and it is less than one tenth of others'. The grade is as low as 1.3.

D22, locates the uppermost and left part in figure 4 of the aspirated syllable of po, has VOT of 40.8 ms and the grade of 2.9. D13, located just below D22, has VOT of 35.0 ms and grade of 2.5. Although the difference of VOT is just 5.8ms, the power difference of 41.9 and 0.5 is

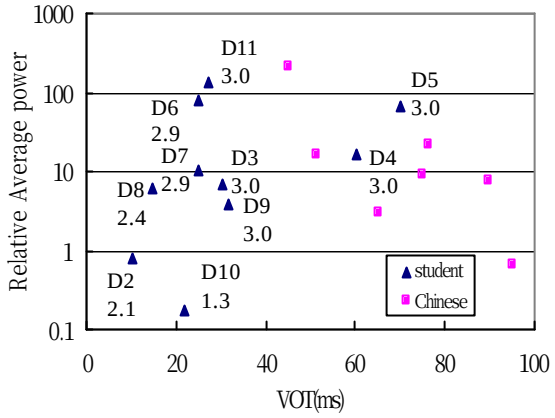
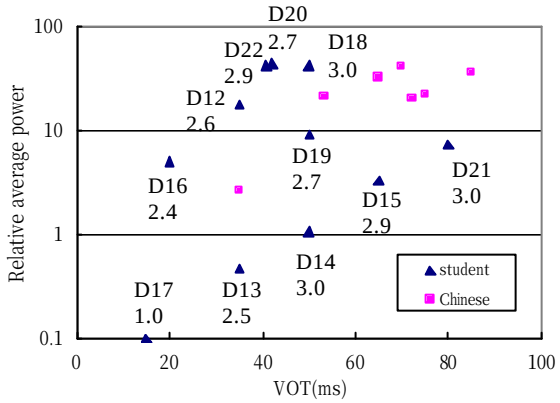
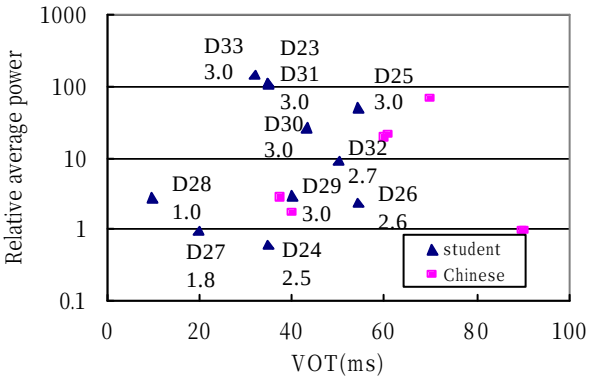


Figure 3: Data distribution of the aspirated syllable pa[p'a] on the surface of VOT in abscissa and relative average power in ordinate



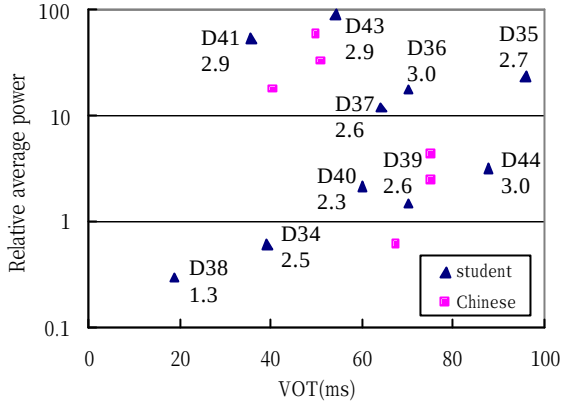
Figures 4: Data distribution of the aspirated syllable po[p'o] on the surface of VOT in abscissa and relative average power in ordinate.



Figures 5: Data distribution of the aspirated syllable pi[p'i] on the surface of VOT in abscissa and relative average power in ordinate.

large. The pronunciation with a higher power gets a better grade.

D33, located the uppermost part in figure 5 of the aspirated syllable pi, has a briefer VOT of 32.1 ms than that of 34.9 ms of D24, located at the lower part. But it has a fairly higher power and gets a better grade of 3.0 than the grade of 2.5 in D24. Here it is demonstrated



Figures 6: Data distribution of the aspirated syllable pu[p'u] on the surface of VOT in abscissa and relative average power in ordinate.

that the pronunciation with a higher power gets a better grade. Overlapped points of D23 and D31 have the VOT of 35.02 and 35.01 ms respectively. Their VOT is just longer than that of D24. But the powers are larger than 100 times that of the power of D24. And their grades are 3.0.

D41, located at the upper and left part in figure 6 of the aspirated syllable pu, has a VOT of 35.6 ms and the power of 54.44. D34, is located far below D41, has a longer VOT of 39.12 ms than that of D41. But the power is as low as 0.61. Their grades split between 2.9 and 2.5.

The upper examples show that the pronunciation with higher power gets a better grade even though their VOT is nearly equal or shorter than those with lower power.

The grade of pronunciations with VOT between 9 and 30 ms of the aspirated syllable pa, that between19 and 30 ms of po, that between 32 and 60 ms of pi and that between 35 and 70 ms of pu does not depend so much on the length of VOT rather depends on the average power used to breathe during VOT.

4.2.2 Case of Very Short VOT

The dependency of the grade on the power is not always true. Some examples show that dependency is not found, if the VOT is very short. The lowest grade of aspirated syllable pi in figure 5 is 1.0 of D28, whose VOT is 10.0 ms. The power of this datum is slightly higher than those of D27 and D24. Then the grades are 1.8 and 2.5, respectively.

4.2.3 Case of VOT with Enough Length

The other examples show that the grade does not depend so much on the breathing power. Although the grades in figure 3 with longer VOT than 30 ms are all 3, their powers varies largely ranging 0.66 and 212.5. D18 and D14 in figure 4 both have grade 3.0 with the same length of VOT. But their powers are 42 and 1 respectively. The grade is high in the pronunciation of the students if VOT is longer than 50 ms. A quarter of all the data of the students have long enough VOT to get a good grade, although they are not cited in table 2.

We show above the accepted theory that the VOT is a sole measure to evaluate the pronunciation of labial aspiration is not always true. But it is shown that the grade is not much correlated with the power, if the length of VOT is longer than some values.

4.3 Correlation between Grade and Evaluation Variables

We show in section 4.2.1 that the grade of the pronunciation of labial aspiration does not always depend on the length of the VOT in some specific ranges of VOT. We deduced the correlations between the grade and evaluation variables, the length of VOT, energy and average power during VOT. We summarized the result in table 3. The correlation with the power and the energy are superior to that with VOT in the above ranges. As for the syllable po, it is shown that the correlation with the power and energy is better than that with VOT, even though the sample is just 3.

Table 3: Correlations between the grade and evaluation variables, the length of VOT, energy and average power.

	VOT Range	No.of Data	Cor.vs VOT	Cor.vs Power	Cor.vs Energy
pa	9～32ms	8	0.596	0.819	0.842
po	19～35ms	3	0.770	0.993	0.983
pi	32～60ms	10	0.462	0.736	0.793
pu	35～70ms	7	-0.029	0.786	0.779

5. CONCLUSION

We examined the VOT and average power during VOT of labial sounds of [p'a], [p'i], [p'o] and [p'u] among Chinese aspiration sounds uttered by 9 native Chinese speakers and 40 Japanese students. The pronunciation grade of each sound was determined as the average value of the 3 evaluation marks by the hearing test of 8 native Chinese speakers. The result shows that the quality of the pronunciation depends not only on VOT but also on the power during VOT.

We continue examining other aspirated syllables in Chinese to establish reliable measures to evaluate the pronunciation.

The authors appreciate very much the suggestion by Dr. R. Tachita (Matsushita Communication Industry) in carrying on this research.

REFERENCES

[1]Ray. D. Kent and Charles Read, 'The Acoustic Analysis of Speech, ' Singular Publishing Group, Inc, San Diego and London, p107, 1992

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

[3]H. Imagawa and S. Kiritani, 'A real-time pitch and formant extraction system using a DSP, and its applications for speech pronunciation training,' IEICE Technical Report, SP89-36, pp.17--24, July 1989.