



**HAL**  
open science

# Discrimination of Chinese pronunciations of aspirated dental and retroflex syllables according to breathing power and its frequency dependency during VOT

Akemi Hoshino, Akio Yasuda

► **To cite this version:**

Akemi Hoshino, Akio Yasuda. Discrimination of Chinese pronunciations of aspirated dental and retroflex syllables according to breathing power and its frequency dependency during VOT. *Acoustics* 2012, Apr 2012, Nantes, France. hal-00810643

**HAL Id: hal-00810643**

**<https://hal.science/hal-00810643>**

Submitted on 23 Apr 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# ACOUSTICS 2012

## Discrimination of Chinese pronunciations of aspirated dental and retroflex syllables according to breathing power and its frequency dependency during VOT

A. Hoshino and A. Yasuda

Toyama National College of Technology, 1-2 Ebie-neriya, Imizu city, Toyama, 933-0293, JAP,  
933-0293 Toyama, Japan  
hoshino@nc-toyama.ac.jp

Retroflexing of aspirates in Chinese is generally difficult for Japanese students learning pronunciation, because the Japanese language has no such sounds. In particular, discriminating between utterances with aspirated dental and retroflex syllables is the most difficult thing to learn. As we observed in our study, many students could not produce the correct sounds. Warping the tongue was not enough to produce the correct articulation, because there is no retroflex sounds amongst the Japanese syllables. In the present paper, the authors extracted the features of the correct pronunciation of the aspirated dental syllables  $ca[t\text{ʰ}a]$ ,  $ci[t\text{ʰ}i]$ , and  $ce[t\text{ʰ}ɿ]$  and the aspirated retroflex ones  $cha[t\text{ʂ}a]$ ,  $chi[t\text{ʂ}i]$ , and  $che[t\text{ʂ}ɿ]$  by analyzing the spectrum of breathed power during voice onset time (VOT) of sounds uttered by nine Chinese native speakers. We found the main difference between aspirated dental and retroflex syllables appeared in the spectrogram of the breathed power during VOT. Then, we examined the relationship between the scores of 20 students' pronunciations and deduced evaluation parameters. We will develop an automatic discrimination system by using the deduced evaluation standards.

## 1 Introduction

Retroflexing of aspirates in Chinese is generally difficult for Japanese students learning pronunciation, because the Japanese language has no such sounds. In particular, discriminating between utterances with aspirated dental and retroflex syllables is the most difficult thing to learn. In our study, we observed a classroom of Japanese students of Chinese uttering aspirated retroflex sounds modeled after examples uttered by a native Chinese instructor. However, the utterances sounded like dental syllables to the instructor, and many students could not produce the correct sounds. They cannot warp their tongues enough to articulate correctly, because there is no retroflex sounds amongst the Japanese syllables.

We previously [1,2,3,4,5] showed that the breathing power,  $P_{rel}$ , during voice onset time (VOT) is a useful measure for evaluating the correct pronunciation of Chinese aspirates. We also developed an automatic evaluation system [6,7] for the students' pronunciation of the Chinese aspirated syllables in accordance with the two parameters of length of VOT and the breathing power during VOT, which the present authors proposed.

In this study, the authors extracted the features of the correct pronunciation of the aspirated dental affricate syllables  $ca[t\text{ʰ}a]$ ,  $ci[t\text{ʰ}i]$ , and  $ce[t\text{ʰ}ɿ]$  and the aspirated retroflex ones  $cha[t\text{ʂ}a]$ ,  $chi[t\text{ʂ}i]$ , and  $che[t\text{ʂ}ɿ]$  by analyzing the spectrum of breathed power during VOT of sounds uttered by nine Chinese native speakers. Then, we deduced these parameters for the pronunciations of 20 Japanese students and examined the relationship between the quality of the pronunciation and these parameters. For the next step, we are preparing to develop an automatic discrimination system by using the deduced evaluation standards.

## 2 Difference between aspirated dental affricate syllables and aspirated retroflex ones

There is an affricate in the pronunciation. The affricate is the complex sound that is generated by simultaneously articulating explosive and fricative sounds as one sound in the same point of articulation.

The dental sound  $[t\text{ʰ}]$  of the aspirate of Chinese is called affricate of alveolar and is formed by articulating an explosive and a fricative sound at the point between the tip of the tongue and the tooth.

The Chinese aspirated retroflex sound  $[t\text{ʂ}ʰ]$  is also called the sublamino-postalveolar affricate, and the point of

articulation is the sublamino-postalveolar. In articulating it, one warps the tongue firmly and articulates explosive and fricative sounds simultaneously.

In this chapter, we define the feature that discriminates between the dental affricate  $[t\text{ʰ}]$  and retroflex one  $[t\text{ʂ}ʰ]$  by examining the spectrogram of the pairs of  $ca[t\text{ʰ}a]$  -  $cha[t\text{ʂ}a]$ ,  $ci[t\text{ʰ}i]$  -  $chi[t\text{ʂ}i]$ , and  $ce[t\text{ʰ}ɿ]$  -  $che[t\text{ʂ}ɿ]$  uttered by a native Chinese speaker.

### 2.1 Difference between aspirated dental affricate syllable $ca[t\text{ʰ}a]$ and aspirated retroflex affricate syllable $cha[t\text{ʂ}a]$

Figure 1 shows temporal evolution of spectrograms of the aspirated retroflex sound  $cha[t\text{ʂ}a]$  (left) and the aspirated dental sound  $ca[t\text{ʰ}a]$  (right) uttered by a Chinese speaker. The lower part of the figure shows the waveform of the voltage evolution picked up by a microphone. The ordinate extended upward shows the frequency component and the darkness of the stripes implies the approximate power level at the corresponding time and frequency. The aspirate appears in the brief interval in the right spectrogram of  $ca[t\text{ʰ}a]$ , indicated by light and thin vertical stripes during VOT, between the stop burst and the onset of vocal fold vibrations followed by a vowel sound. This time interval is called the voice onset time, VOT [8]. It is long, 160 ms. Although the slightly darker stripes appear between 2500 and 5000 Hz in frequency and 70 and 150 ms in VOT, temporal variation of the breathing power during VOT is hardly seen.

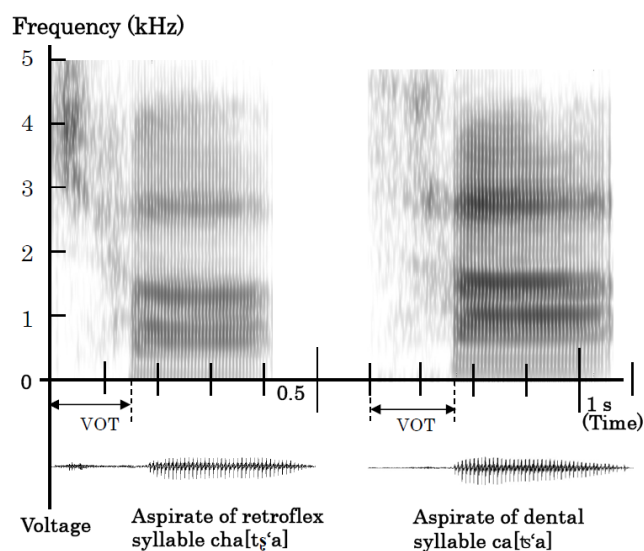


Figure 1: Spectrograms of retroflex aspirate syllable

cha[tʂ'a] (left) and dental aspirated syllable ca[tʂ'a] (right) pronounced by a Chinese speaker.

The left spectrogram in figure 1 is for aspirated retroflex sound cha[tʂ'a] uttered by a Chinese speaker. The VOT was long, 150 ms. Dark vertical stripes are observed between 2500 and 5000 Hz in frequency, upper left during 0~70 ms in VOT. This is a friction of the breath caused by the breath explosion, which arises at a spot between the warped tongue and posterior alveolar. A large energy in the mouth dissipates at the early stage of VOT and generates the high breathing power there.

## 2.2 Difference between aspirated dental affricate syllable ci[tʂ'i] and aspirated retroflex affricate syllable chi[tʂ'i]

Figure 2 shows temporal variation of spectrograms of the aspirated retroflex sound chi[tʂ'i] (left) and the aspirated dental sound aspirated ci[tʂ'i] (right) uttered by a Chinese speaker. The lower part of the figure shows the voltage evolution picked up by a microphone. The VOT of the aspirated dental ci[tʂ'i] was long, 225 ms, in the right hand side spectrogram. The unvarying darkness of the vertical stripes shows the breathing power is rather steady during VOT.

The left spectrogram in figure 2 is for the aspirated retroflex sound chi[tʂ'i]. The VOT was long, 250 ms. During almost the whole VOT, the dark vertical stripes are observed in the frequency band of 2000~5000 Hz. This is a friction of the breath caused by the breath explosion, which arises at a spot between the warped tongue and posterior alveolar. A large energy in the mouth dissipates during the whole VOT, and the strong breathing power is generated. Moreover, the weaker power was consumed at the frequency lower than 1500 Hz in the VOT, as the light vertical stripes imply.

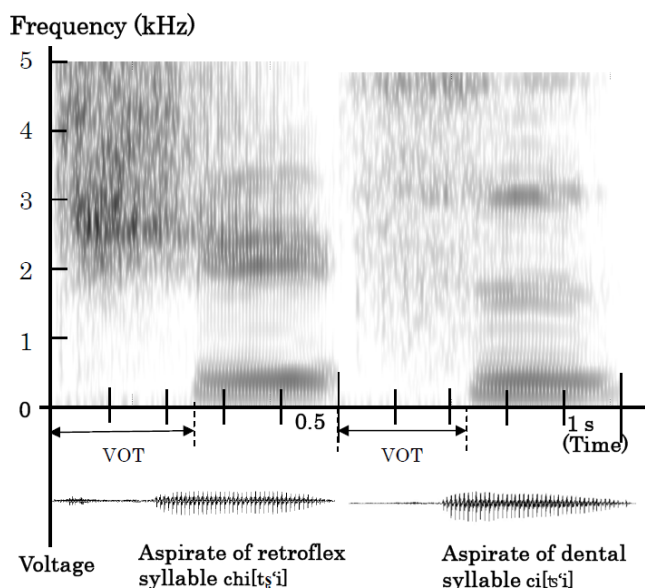


Figure 2: Spectrograms of retroflex aspirate syllable chi[tʂ'i] (left) and dental aspirated syllable ci[tʂ'i] (right) pronounced by a Chinese speaker.

## 2.3 Difference between aspirated dental affricate syllable ce[tʂ'ɿ] and aspirated retroflex affricate syllable che[tʂ'ɿ]

Figure 3 shows temporal variation of spectrograms of the aspirated retroflex sound che[tʂ'ɿ] (left) and the aspirated dental sound ce[tʂ'ɿ] (right) uttered by a Chinese speaker. The VOT of the aspirated dental sound ce[tʂ'ɿ] was long, 180 ms. The stripes around 2000 Hz are somewhat darker, and slightly stronger power is observed there.

The left of figure 3 shows a spectrogram of the aspirated retroflex syllable che[tʂ'ɿ]. The VOT was 125 ms. In the first half portion of VOT, dark vertical stripes in the frequency of 2000 Hz~5000 Hz are observed. This is also generated by a friction of the breath caused by the breath explosion, which arises at a spot between the warped tongue and posterior alveolar with a strong power, generated in this portion.

Moreover, for the frequency lower than 1200 Hz in VOT, the darkness of the vertical stripes is light in accordance with the weak power.

The distinctive feature of retroflex aspirated syllables is that they have non-uniform spectrum in frequency and/or time during VOT, whereas aspirated dental ones have rather uniform spectrum.

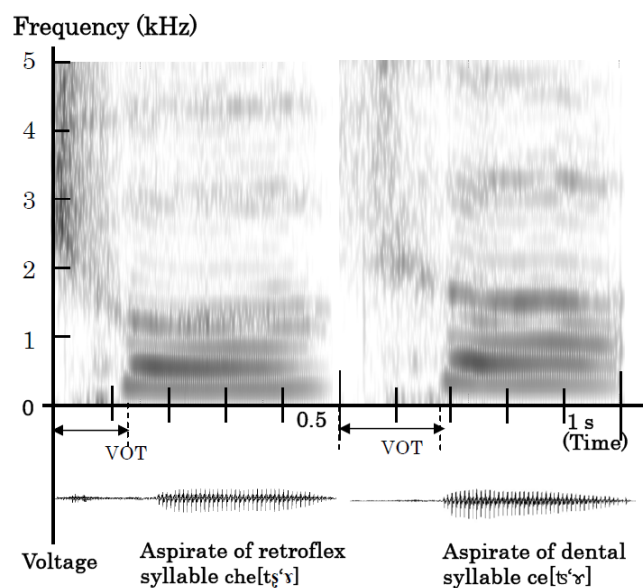


Figure 3: Spectrograms of retroflex aspirate syllable che[tʂ'ɿ] (left) and dental aspirated syllable ce[tʂ'ɿ] (right) pronounced by a Chinese speaker.

## 3 Relationship between power and its frequency dependency during VOT and the quality of pronunciations

Although several reports of research [9,10] on voiced retroflex have been published before, there are still few reports on research of aspirated retroflex. In this chapter, we define the discrimination standards of the aspirated dental affricate syllables and aspirated retroflex ones by examining the VOT and the power spectrum during VOT of the pronunciations of the pairs of ca[tʂ'a] - cha[tʂ'a], ci[tʂ'i] - chi[tʂ'i], and ce[tʂ'ɿ] - che[tʂ'ɿ] uttered by 20 Japanese students.







## References

- [1] A. Hoshino, A. Yasuda, "Evaluation of Chinese aspiration sounds uttered by Japanese students using VOT and power (in Japanese), *Acoust. Soc. Jpn.*, **58**, No. 11, pp.689-695,(2002).
- [2] A. Hoshino and A. Yasuda, "The evaluation of Chinese aspiration sounds uttered by Japanese student using VOT and power," 2003 International Conference on Acoustics, Speech, and Signal Processing IEEE Proceedings, Hong Kong, pp. 472-475, 2003.
- [3] A. Hoshino and A. Yasuda, "Dependence of correct pronunciation of Chinese aspirated sounds on power during voice onset time," Proceeding of ISCSLP 2004, Hong Kong, pp. 121-124, 2004
- [4] A. Hoshino and A. Yasuda, "Effect of Japanese articulation of stops on pronunciation of Chinese aspirated sounds by Japanese students," Proceeding of ISCSLP 2004, Hong Kong, pp. 125-128, 2004
- [5] A. Hoshino and A. Yasuda, "Evaluation of aspiration sound of Chinese labial and alveolar diphthong uttered by Japanese students using voice onset time and breathing power," Proceeding of ISCSLP 2006, Singapore, pp. 13-24, 2006.
- [6] A. Hoshino and A. Yasuda, "Pronunciation Training System for Japanese Students Learning Chinese Aspiration," The 2nd International Conference on Society and Information Technologies(ICSIT), Orlando, Florida, USA,pp.288-293, 2011
- [7] A. Hoshino and A. Yasuda, "Pronunciation Training System of Chinese Aspiration for Japanese Students," Acoustical Science and Technology, Japan, Vol.32, No4, pp.154-157, July, 2011
- [8] Ray. D. Kent, Charles Read, "The Acoustic Analysis of Speech," Singular Publishing Group, Inc., San Diego and London, pp.105-109, 1992
- [9] C. Zhu, "Studying Method of the Pronunciation of Chinese Speech for Foreign Students (in Chinese)," Yu Wu Publishing Co. China, pp. 63-71, 1997
- [10]L. Zhou, H. Segi and K. Kido, "The investigation of Chinese retroflex sounds with time-frequency analysis," The Acoustical Society of Japan, Vol54, No.8, pp.561-567, 1998.