Acoustic investigation of /tʰ/ lenition in Brunei Mandarin

Shufang Xu
Faculty of Arts and Social Science, University of Brunei Darussalam, Brunei Darussalam
12h8151@ubd.edu.bn

Abstract
This study investigates the acoustic characteristics of /tʰ/ lenition in conversational speech of Brunei Mandarin, a variety of Mandarin Chinese. Based on data from 20 Chinese Bruneians, /tʰ/ lenition was found in the third-person pronoun tā /tʰə/, which is frequently pronounced as hā [ha]. Perceptual judgments, spectrographic analysis and acoustic measurements were conducted to examine the features of this sound change. In comparison with the perceptual judgments, it was found that the spectrographic inspection yielded 83.6% correct classification of [tʰ] and [h] for female speakers and 77.2% for male speakers, indicating there is reasonably high reliability in identification in terms of spectral properties. Results of the acoustic measurements showed that there is an increase in high frequency intensity after the release of the closure for [tʰ] while there is little change in intensity during the frication for [h]. The results showed that the lack of burst and little increase in intensity are reasonably reliable cues for stop lenition.

Index Terms: Brunei Mandarin, /tʰ/ lenition, auditory judgment, spectrographic observation, intensity

1. Introduction
Lenition is a common phonological phenomenon in which the consonant becomes weaker, especially during fast speech [1, 2, 3]. Accounting for a series of segmental alternations in a variety of languages in the world, it has a firm place in phonological theory, and many scholars focus on its description and analysis. It is a phonetic process described as consonant reduction, normally found in spontaneous speech [4], with fast speech rate as a possible trigger, and sometimes vice versa. /tʰ/ lenition, in which /tʰ/ at the start of a third-person pronoun tā /tʰə/ is pronounced as [h], is another distinctive feature.

Lenition of /t/ to [h] is not uncommon in the literature on English. Hickey [1] describes this /t/ to [h] phenomenon especially in word-final position by Dublin speakers, such as Saturday [sæhəd]. In addition, Liverpool English is well known for its syllable-final /t/ to [h] lenition [3, 10, 11]. There are some restrictions on the environments in which lenition of /t/ to [h] tends to occur, including monosyllabic function words with short vowels and /t/ in syllable-final position, such as but [bɑːθ], that [θæt], and it [ɪt] [3, 12].

Though Duanmu [9] briefly mentions that a similar phenomenon for the tā pronoun exists in Taiwan accented Standard Chinese, this sound change from /tʰə/ to [ha] has not previously been reported in other varieties of Chinese, and it has never been investigated acoustically or systematically.

In spectrograms, one fundamental cue for plosives is the burst, described as a spike, a vertical bar immediately after the period of closure. The absence of this burst is a reliable cue for lenition, indicating the loss of strength for plosives [13, 14]. Intensity measurement is another reliable correlate of consonant weakening. It is reported that there is less increase in energy during the release of the burst for a stop that has undergone lenition than for one that has no lenition [15].

The current study provides an acoustic analysis of /tʰ/ lenition found in Brunei Mandarin. Its objectives are twofold: first, it examines the extent of the reliability of spectrographic analysis by comparing with the results of perceptual judgment; second, it investigates the extent to which we can discriminate between [tʰ] and [h] by acoustic measurements of intensity.

2. Method

2.1. Subjects
Twenty Chinese Bruneians were recruited in this study, ten females and ten males. At the time of the study, they had a mean age of 21 years, ranging from 19 to 23 years, and all were tertiary students. Of the 20 students, 12 were from Universiti Brunei Darussalam (UBD), seven were from Institut Teknologi Brunei (ITB), and one was from Kolej IGS Brunei Darussalam (IGS), a private tertiary institution in Brunei. Subjects will be referred to using letter-number identifiers. For example, F1 refers to female number one. Though one male subject (M5) did not want to reveal his age, he mentioned that he was doing his degree in ITB at the time of recording, so his age can be assumed to be between 19 and 23. Native or near native Chinese was an essential requirement for inclusion in this study, as we are looking at the acoustic characteristics of /tʰ/ lenition in proficiently spoken Brunei Mandarin.
2.2. Experimental task and stimuli
A face-to-face interview was conducted by the researcher with each participant. The participants were asked to talk freely about their recent travelling experiences or their daily activities. Each interview lasted about six minutes (average 6.3 minutes). The participants spoke fast and tended to use a fairly informal style of speech, such as laughing, using slang, and occasionally switching into English, and the speech frequently involves sound changes such as reduction and elision.

2.3. Recording procedure and data extraction
The interviews were recorded in a quiet office at UBD and a quiet room at ITB. The speech was recorded directly onto a Sony laptop computer at a sampling frequency of 44,100 Hz, with a high-quality University Sound US692 dynamic microphone positioned a few inches from the mouths of the speakers.

The data were saved as .wav files and analyzed using Praat [16]. The interviews were transcribed into text format for analysis. Sentences containing the target pronoun /t ha/ were extracted and they yielded a total of 217 tokens of the pronouns from the 20 speakers.

Auditory judgments were performed first to identify the tokens of [tʰ] and [h] from the data. The researcher listened to all the tokens and identified them as starting with [tʰ] or [h], and then her supervisor did the same to obtain an estimate of inter-rater auditory reliability.

Acoustic measurement in this study consists of two parts. Firstly, spectrographic identification was conducted to see if visual examination can indicate [tʰ] and [h]. Secondly, acoustic measurement was made to describe the intensity differences between [tʰ] and [h].

2.4. Visual inspection
When looking at the spectrograms, segmentation can be rather problematic as some segments may not have a clear start or end [17]. However, in this case it was fairly straightforward to identify the start of the vowel in /tʰa/. A Praat script was then used to obtain a one second segment with the vowel in /tʰa/ at the mid-point and the spectrograms were then saved as .jpg files. Thus, 217 spectrographic images were extracted, and in nearly all cases, the token of /tʰa/ can be identified easily.

The criteria used in classification were as follows:

For the aspirated plosive [tʰ]:
- There is a visible sudden sharp spike/burst in the spectrogram indicated by the arrow in Figure 1.
- There is lots of energy in the higher frequency part of the spectrogram immediately after the sharp spike.

For the voiceless glottal fricative [h]:
- Energy is distributed over a wide range of frequencies as shown by the arrows above the spectrogram in Figure 2.
- The energy is not so intense as for [tʰ].
- There is sometimes a clear formant pattern during the fricative.

![Figure 2: [ha] as spoken by F5 (187.24s)](image)

However, there are instances in which what was auditorily perceived to be [ha] was judged to be [tʰa] from the spectrographic examination, because there seems to be a sudden burst. Figure 3 shows an instance of an auditory [ha] that was thought to be [tʰa] because of the apparent spike indicated by the arrow.

![Figure 3: [ha] as spoken by F2 (313.44s)](image)

Meanwhile, an auditory [tʰa] was at times judged as [ha] when such a burst is non-existent, as in Figure 4. We can see that the region below the arrows has a lot of random noise and some faint formant bands. However, it was auditorily judged to be [tʰa], even though it looks like a [h] in the spectrogram.

![Figure 4: [tʰa] as spoken by F5 (303.3s)](image)

2.5. Acoustic measurement
As already discussed, aspiration for [tʰ] is expected to show up on a spectrogram as a burst of high-frequency energy. We can try to measure whether there is a sudden increase in high-frequency energy using the following methodology.

Because we are only concerned with high-frequency energy, the speech was filtered to remove all energy below 2000 Hz, using a Hann band stop filter. Figure 5 shows a token of [tʰa] after filtering. Location C is the start of the vowel. The intensity at a position 80 msec from the start of the vowel (location A) was measured and compared with the intensity at a position 40 msec from the start of the vowel (Location B). These times were chosen because we expect the aspiration for [tʰ] to be a little over 60 msec [7]. For [tʰ] we expect Location B to have more intensity than Location A as

![Figure 5: [tʰa] as spoken by F5 (303.3s)](image)
in Figure 5, but for [h], we expect there to be little difference between the intensity at Locations A and B as in Figure 6.

![Figure 5: \( \text{[\text{th}]} \) as spoken by F7 (78.21s)]

![Figure 6: \( \text{[h]} \) as spoken by F7 (108.15s)]

There is a red line on the spectrograms to show intensity. As shown in Figure 5, the red line has an upward trend from A to B, indicating an increase in intensity. However, it fluctuates evenly from A to B in Figure 6, indicating that the intensity does not change much during the frication for [h].

## 3. Results

For the auditory judgment, the results of the researcher and her supervisor were compared and it was found that they differed in nine tokens out of the total of 217, making the percentage of agreement nearly 96%. All nine of these tokens involved the researcher hearing \( \text{[\text{th}]} \) while her supervisor heard [h]. The results of the researcher are reported in this study, so we can conclude that the incidence of [h] reported here is a conservative estimate. The results for the auditory perception are shown in Table 1. Overall, 40.6% of the tokens had lenition.

<table>
<thead>
<tr>
<th>Table 1. Auditory perception</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>[h]</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

It can be seen that the incidence of \( \text{[\text{th}]} \) lenition demonstrated by women is substantially higher than by men, as men have a greater tendency to maintain the standard pronunciation, \( \text{[\text{th}]} \). The difference between women and men is highly significant (\( \chi^2=7.619, \text{df}=1, \text{p}=0.006 \)). Women tend to be the linguistic trendsetters in many societies [18], so this suggests that [h] may become the norm in Brunei Mandarin in the future.

The results for the spectrographic observation are shown in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Spectrographic observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectrographic observation</td>
</tr>
<tr>
<td>[h]</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Comparing Table 2 with Table 1, there are more cases in which perceptual \( \text{[\text{th}]} \) is judged as [h] from the examination of the spectrograms than the other way round, particularly for the male speakers.

Table 3 shows the differences between the spectrographic observation and the auditory judgments.

<table>
<thead>
<tr>
<th>Table 3. Differences from spectrographic observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditory [h] observed as ( \text{[\text{th}]} )</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Clearly, there is a tendency for auditory \( \text{[\text{th}]} \) to be judged as [h] when examining the spectrograms. Thus, the accuracy of spectrographic analysis reaches 83.6% for female speakers and 77.2% for male speakers based on perceptual judgments. In other words, the absence of the burst release is a reasonably reliable cue for the detection of \( \text{[\text{th}]} \) lenition, because it indicates that there is less of an energy increase when this sound change takes place, but there are quite a number of exceptions.

A summary of the results for intensity differences is shown in Table 4. As expected, the difference in intensity is larger for \( \text{[\text{th}]} \) than for [h].

<table>
<thead>
<tr>
<th>Table 4. Average intensity differences between locations A and B (in dB) for ( \text{[\text{th}]} ) and [h]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average differences between A and B</td>
</tr>
<tr>
<td>( \text{[\text{th}]} )</td>
</tr>
<tr>
<td>[h]</td>
</tr>
</tbody>
</table>

This difference is highly significant (\( t=6.82, \text{df}=215, \text{two-tailed, paired-sample, p}<0.0001 \)).

However, there is also a substantial overlap. Let us consider a threshold for the \( \text{[\text{th}]} \)/[h] separation. If we make an increase in energy of 3.25 dB to be the threshold, we find that 37 tokens of \( \text{[\text{th}]} \) have an increase in energy of less than 3.25dB, and 25 tokens of [h] have an increase in energy of more than 3.25 dB. In other words, 155 tokens were classified correctly using this method, representing 71.4%.
4. Discussion and Conclusions

A high-frequency burst is considered a distinct indicator for [tʰ], but this study did not find consistent burst characteristics for [tʰ] in connected speech. However, the relatively reliable acoustic cue for [tʰ] is the widespread energy that is concentrated at the high-frequency region of the spectrograms. [h], on the other hand, can be usually recognized on spectrograms.

Strictly speaking, [h] is not a real voiceless fricative as the source of the noise is not air being forced through a narrow gap [19]. As a result, there is not much energy in the upper frequency range on the spectrogram. Although sometimes there is a faint trace of a burst at the beginning of [h], it has consistent acoustic cues most of the time.

It is widely agreed that spectrographic analysis works well for vowels, especially for monophthongs, but it often becomes more difficult when analyzing consonants [20]. In addition, the acoustic cues presented in the spectrograms are not always consistent due to many factors, such as background noise and speech rate, so perceptual judgment is also essential in this study.

In conclusion, the absence of the release of burst and little increase in intensity during the release of the burst can be reliable cues for identifying lenition from [tʰ] to /h/. However, we cannot solely depend on the acoustic results. A better way to describe the phonetic properties is to combine the results of perceptual and acoustic analysis, though it is also likely that improved acoustic techniques can be developed.

5. Acknowledgements

I am indebted to my supervisor, Prof. David Deterding, who provided invaluable help throughout the writing of this paper, including the recruit of subjects, the guidance on recording, and the feedback and suggestion on my writing.

6. References