<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Acoustic cues for the perception of aspiration in Cantonese initial stops</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Other Contributor(s)</strong></td>
<td>University of Hong Kong</td>
</tr>
<tr>
<td><strong>Author(s)</strong></td>
<td>Poon, Man-wai, May; 潘文瑋</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2000</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/56229">http://hdl.handle.net/10722/56229</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; The author retains all proprietary rights, such as patent rights and the right to use in future works.</td>
</tr>
</tbody>
</table>
Acoustic cues for the perception of aspiration

in Cantonese initial stops

Poon Man Wai, May

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, May 10, 2000.
ABSTRACT

This study investigated the significant acoustic cues used in the perception of aspiration contrast in Cantonese initial stops by normal hearing adults. The effects of vowel characteristics, aspiration noise level and voice-onset time (VOT) duration on aspiration perception in Cantonese initial stops were studied. The stimuli were consonant-vowel (CV) real words recorded by a male Cantonese native speaker. Initial stops /p, pʰ, t, tʰ, k, kʰ/ combined with the same vowel /a/ to give the stimuli. A total of twenty-four stimuli conditions were edited from the natural stimuli by varying the vowel ('Aspirated vowel' or 'Unaspirated vowel'), aspiration noise level ('✓ Aspiration noise' or '✗ Aspiration noise') and VOT duration ('Long VOT' or 'Short VOT') independently. Ten subjects listened to the stimuli and selected one from the six real words (/pa/ 'father', /pbʰa/ 'on all fours', /tʰa/ 'dozen', /ta/ 'he', /ka/ 'home', /kʰa/ 'compartment') which he or she might find the same or the most similar to the presented stimulus. The results showed that the vowel characteristics of the stimuli were the main cues used on aspiration perception by the normal hearing adults. The cueing effect of 'Aspirated vowel' was independent from the other two acoustic parameters. Although both aspiration noise level and VOT duration showed significant cueing effects on aspiration perception in Cantonese initial stops, only the co-existing of '✓ aspiration noise' and 'Long VOT' was sufficient to cue for the aspirated perceptual response in the absence of 'Aspirated vowel'.
INTRODUCTION

This study investigated the significant acoustic cues used in the perception of aspiration contrast in Cantonese initial stops by normal hearing adults. Syllable-initial stop is characterized by an articulatory occlusion (closure), followed by a release of a burst of energy in form of air escapes and then an articulatory stop to vowel transition. Acoustically, the closure-release-transition phases are represented by a stop gap, a noise burst and the formant shifts respectively on the spectrogram or digital waveform displays (Kent & Read, 1992).

Previously, the perception of English initial stops have been widely studied (Lisker & Abramson, 1964; Stevens & Klatt, 1974; Summerfield & Haggard, 1977; Massaro & Oden, 1980; Revoile et al., 1987). Some were concentrated on the perception of place of articulation and some were focused on the voicing perception. The acoustic characteristics of English initial stops were also explained thoroughly in Kent and Read (1992). However, other than Lisker and Abramson (1974), very few studies attempted to investigate the perception of Cantonese initial stops.

Tsui and Ciocca (2000) studied the differences in the perception of aspiration and place of articulation of Cantonese initial stops among normal and sensorineural hearing impaired people. They suggested that in the perception of Cantonese initial stops, aspiration noise level and formant transitions was the main cue used in aspiration perception by the normal and the sensorineural hearing impaired listeners respectively. Voice-onset time (VOT) was only a weak cue for both group of listeners.
Clinically, a clear picture on the independent cueing effects of different acoustic parameters of stops is helpful in planning and providing speech and aural rehabilitation programmes for the Cantonese-speaking society.

The characteristics of Cantonese and English initial stops

1. English.

In English, different places of articulation of initial stops are categorized by forming the articulatory closure in the bilabial, alveolar or velar position. There is voiced and voiceless contrast among English initial stops determined by the presence or absence of glottal buzz during the interval of closure (Lisker & Abramson, 1964). In releasing the stop-burst, it is further differentiated into aspirated or unaspirated. “Aspiration is a breathy noise generated as air passes through the partially closed vocal folds and into the pharynx.” (Kent & Read, 1992).

Between the articulatory release and the onset of vocal fold vibrations, there is a time interval called the voice onset time (VOT). In the spectrogram, it is marked as the duration of interval between the onset of the stop-burst release and the starting of the glottal signal (Liser, 1975; Lisker & Abramson, 1970, as cited in Revoie, et al., 1987).

In English, an aspirated stop is characterized by having long VOT (greater than about 50 ms) and containing breathy noise in the release burst. In contrast, an unaspirated stop is characterized by having short VOT (less than 20-30 ms) with no breathy noise in the release burst (Stevens & Klatt, 1974; Kent & Read, 1992).

The vocal fold vibrations start off the stop to vowel transition. It involves the changes
in vocal tract shape from stop occlusion to vowel configuration. It is associated with the rapid spectral change in the form of shifting formants (Kent & Read, 1992). According to Stevens and Klatt (1974), the English voiced stops normally have significant formant transitions following vowel onset whereas these formant transitions are essentially completed prior to vowel onset for the English voiceless stops.

2. Cantonese.

Cantonese initial stops share most of the features with the English initial stops. They have the same categorization of places of articulation. Stops are classified as either aspirated or unaspirated also. However, the aspiration in Cantonese stops is phonemic but it is not in English stops (Kent & Read, 1992). For example, in Cantonese, /pau/ and /pʰau/ are representing two different words ‘包’ (bread) and ‘抛’ (throw) respectively. The aspiration contrast between /p/ and /pʰ/ is used to distinguish two phonemes. In English, although the ‘p’ in pie (/pʰai/) is produced differently from the ‘p’ in spy (/spai/) in terms of aspiration, they are still representing the same phoneme.

As stated, English initial stops with the same place of articulation are contrasted in both voicing and aspiration. There are voiced unaspirated stops /b, d, g/ and voiceless aspirated stops /p, t, k/. In Cantonese, however, there is no phonetic voicing contrast among syllable-initial stops. All stops are voiceless. They are classified into voiceless unaspirated stops /p, t, k/ and voiceless aspirated stops /pʰ, tʰ, kʰ/ according to their aspiration difference only (Lisker & Abrahamson, 1964).
Previous study

Previously, it was found that aspiration noise level and voice onset time (VOT) duration of the release burst were two important acoustic cues used in the perception of voicing in English initial stops (Revoile et al., 1987). In Tsui and Ciocca (2000), the level of aspiration noise was also claimed to be the dominant cue on aspiration perception in Cantonese initial stops by the normal hearing listeners.

Formant transitions (vowel-onset transitions) was another parameter suggested to be important on voicing perception in English initial stops (Stevens & Klatt, 1974; Revoile et al., 1987). Their cueing effects on aspiration perception in Cantonese initial stops, especially for the sensorineural hearing impaired listeners, were proved in Tsui & Ciocca (2000).

However, most of these studies were limited by the covariation effect among acoustic parameters in editing the stimuli. For example, the VOT duration and aspiration noise level were not varied independently. The shortening of VOT duration would be accompanied by the reduction in aspiration noise level, and vice versa. The consequence was that without examining the ‘short VOT with high aspiration noise level’ and the ‘long VOT with low aspiration noise level’ conditions, it was hard or even impossible to investigate the independent effects of these two acoustic parameters on voicing or aspiration perception.

Revoile et al. (1987) did attempt to test for the independent effect of VOT duration on voicing perception by including the ‘Silence-for-aspiration’ condition in their study. This method was also used in the study of Tsui and Ciocca (2000). They inserted a silent phase
into the unaspirated stops so as to form the 'long VOT with low aspiration noise level' (Long VOT) condition. However, the cueing effect of 'short VOT with high aspiration noise level' condition on voicing or aspiration perception is still unknown.

Study goals

The goals of the present study were to investigate both the independent and interacting effects of (1) vowel characteristics, (2) aspiration noise level and (3) VOT duration on aspiration perception in Cantonese initial stops by normal hearing adults. The term 'vowel characteristics' was used instead of 'formant transitions' in this study because other than the formant transitions during vowel-onset, the onset fundamental frequency was also suggested to be possible in responsible for the voicing perception in English initial stops (Kent & Read, 1992). Therefore, this study used 'vowel characteristics' to represent all the possible vowel-related cues used in the perception of aspiration in Cantonese initial stops.

The independent effects of these three acoustic parameters were investigated by varying them independently in editing the stimuli. It is hypothesized that if any one of these three acoustic parameters is the dominant cue that responsible for the perception of aspiration in Cantonese initial stops, this acoustic parameter will have an absolute determining effect on aspiration perception. This absolute effect will not be affected by the conditions of other parameters of the stops.

The perceptual values corresponding to the changes in these three acoustic parameters would also be evaluated.
METHOD

Participants

Ten participants with no known hearing problem participated in this study. They aged from twenty to twenty-three years old (Mean age = 22.03). All of them were native speakers of Cantonese. None of them has ever received previous training on acoustic perception.

Stimuli

Six monosyllabic real words with aspirated or unaspirated initial stops were used to be the natural stimuli. They were: 爸 /pa/ (father), 扑/pʰa/ (on all fours), 打 /ta/ (dozen), 他 /tʰa/ (he), 家 /ka/ (home), 角 /kʰa/ (compartment).

Cantonese, one of the many Chinese dialects, is a tone language. Matthews and Yip (1994) stated that the pitch or pitch pattern with which a syllable (or word) is pronounced in a tone language is crucial to the identification of the syllable (or word). There are six commonly distinguished tones in Cantonese (Tone 1 to 6). They are characterized in terms of their starting and ending pitch levels (1 to 5 in order of ascending pitch): High level /55/, High rising /35/, Mid level /33/, Low rising /23/, Low level /22/ & Low falling /21/. In order to eliminate the possible context effects caused by tonal difference, all the stimuli used in this study had a high level tone (Tone 1).

The possible vowel context effects were also eliminated by combining the six stops with the same vowel /a/ to form the CV structured stimuli. Stops from all the bilabial (/p/ & /pʰ/), alveolar (/t/ & /tʰ/) and velar (/k/ & /kʰ/) positions were included to make the study
result independent from the place of articulation of the stops. Thus, six monosyllabic real words were used to avoid the difficulty in identifying meaningless syllables.

Recording and editing

A male Cantonese native speaker (20 years old) recorded the six unmodified stimuli. The speaker was unknown by all subjects so as to avoid the practice effect due to familiarity of the speaker's production. His articulation was judged to be clear and intelligible by a student speech therapist.

The stimuli were recorded in a ICAC single-walled sound-proof room using a low noise omnidirectional Microphone (Brüel and Kjær, Type 4003). The speech samples were stored on the hard disk of a Macintosh HD computer equipped with a GW instruments sound scope. Digital waveform displays for editing were obtained using the Sound Scope software.

During the recording, the six real words (爸, 扒, 打, 他, 家, 卡) were presented one by one to the speaker in a randomized order. The speaker was asked to read aloud the presented word in a normal pitch and a normal volume. A total of ten trials were recorded from the speaker. The best trial for each word in terms of clarity and amplitude was chosen to be the stimulus.

The recorded stimuli were edited along three dimensions: (1) Vowel characteristics ('Aspirated vowel' or 'Unaspirated vowel'), (2) Aspiration noise level ('✓ Aspiration noise' or '✗ Aspiration noise') and (3) VOT duration ('Long VOT' or 'Short VOT').
<table>
<thead>
<tr>
<th>Vowel characteristic</th>
<th>Aspiration noise</th>
<th>VOT duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level</td>
<td></td>
</tr>
<tr>
<td>Aspirated vowel</td>
<td>√ Aspiration noise</td>
<td>Long VOT (Stimulus 1)*</td>
</tr>
<tr>
<td>Aspirated vowel</td>
<td>× Aspiration noise</td>
<td>Long VOT (Stimulus 3)</td>
</tr>
<tr>
<td>CV stimuli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(labial/alveolar/velar)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unaspirated vowel</td>
<td>√ Aspiration noise</td>
<td>Short VOT (Stimulus 6)</td>
</tr>
<tr>
<td>Unaspirated vowel</td>
<td>× Aspiration noise</td>
<td>Long VOT (Stimulus 7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Short VOT (Stimulus 8)</td>
</tr>
</tbody>
</table>

(Remark: the asterisk is placed next to the unmodified stimulus)

There were all together twenty-four stimuli with eight created from each places of articulation of stops (labial/ alveolar/ velar). For each set of eight stimuli, only the one with 'Aspirated vowel - √ Aspiration noise - Long VOT' condition (Stimulus 1) was remained unmodified (natural aspirated stimulus). The original noise level of the natural unaspirated stimulus was set to zero to become the 'Unaspirated vowel - × Aspiration noise - Short VOT' condition (Stimulus 8).

1. Vowel characteristics.

The 'Aspirated vowel' was the vowel portion of the waveform edited from the natural aspirated stimuli started from the onset of voicing (or in the other words, the beginning of the vowel-onset formant transitions). For the 'Unaspirated vowel', it came from the vowel portion of the natural unaspirated stimulus.
2. Aspiration noise level.

The two conditions of aspiration noise level were ‘✓ Aspiration noise’ (original noise level of the natural aspirated stimulus) and ‘✗ Aspiration noise’ (set to zero level). Fant (1973, as cited in Revoile et al., 1987) suggested that there were three successive phases in voiceless-aspirated stops named (1) the transient phase, (2) the frication phase and (3) the aspiration phase. Different phases correspond to different types and locations of sound source changed during the production of stops. The natural aspirated and unaspirated stops differ in the aspiration phase, which is the main part of the stop-burst.

In order to increase the aspiration noise level of an unaspirated stimulus, its aspiration phase was replaced by the one cut from the aspirated stimulus. In the ‘✗ Aspiration noise’ conditions, the noise level of their aspiration phase was set to zero. For the ‘✓ Aspiration noise - Short VOT’ conditions, the short aspiration phase was filled up by the beginning and the ending segments of the aspiration phase from the natural aspirated stimulus. (Appendix A) These two segments were chosen so as not to miss any possible transitional information contained in the aspiration phase of the natural aspirated stimulus.

3. VOT duration.

According to the transient-frication-aspiration phases suggested by Fant (1973, as cited in Revoile et al., 1987), the ‘Long VOT’ and ‘Short VOT’ conditions differed in the duration of aspiration phase only.

For the ‘Unaspirated vowel - ✓ Aspiration noise - Long VOT’ condition, the aspiration
phae of the unaspirated stimulus was elongated by the addition of a silent phase. Therefore, the VOT duration was increased independently (from ‘Short VOT’ to ‘Long VOT’) without changing the aspiration noise level.

Ignore the place errors

In this study, although stops from bilabial, alveolar and velar places of articulation were included, errors made on place of articulation in the perceptual response were ignored. It is because differences in formant transitions (second and third formants) and VOT duration are cues used in the perception of place of articulation of initial stops also (Kent & Read, 1992; Tsui & Ciocca, 2000). Lisker and Abramson (1964) so as Kent and Read (1992) suggested that the average VOT duration of English initial stops of different places of articulation were ranked in the following order: shortest for bilabial stops, intermediate for alveolar stops and longest for velar stops. Therefore, varying these two acoustic dimensions would affect the identification on aspiration contrast so as the difference in place of articulation.

For our participants, it was the most difficult for them to identify the alveolar stops (/t/ and /tʰ/) in terms of place of articulation. This was consistent with Tsui and Ciocca (2000) that in the perception of place of articulation in Cantonese initial stops, VOT modification affects the identification of alveolar stops the most. As only the effects made on aspiration perception in Cantonese initial stops were interested in this study, the experimenter ignored all the unwanted distracting results (including place errors) and only the aspiration-related results are recorded.
Procedure

The experiment was carried out in the same ICAC single-walled sound-proof room as the recording. The whole experiment required approximately thirty minutes to finish. Each participant sat in front of a computer and listened to the stimuli with a headphone individually. Before listening to the stimuli, the experimenter introduced and read aloud the six real words (爸, 扒, 打, 他, 家, 卡) to the participants. Then, the participants repeated the six words so as to ensure that they were familiar with the words. Both written and oral instructions were given to the subjects.

In the experiment, the stimuli were presented through a hypercard program using a sound card via Sennheiser headphones. A block of twenty-four stimuli was presented eleven times to each participant. A total of two hundreds and sixty-four trials of stimuli were presented. The stimuli within each block were ordered randomly. In each trial, the stimulus was presented twice. After the second presentation of the stimulus, six real words (爸, 扒, 打, 他, 家, 卡) were displayed on the screen. Six number boxes (1 to 6) were shown under the real words correspondingly. The participant was then asked to select one from six of those which he or she might find the same or the most similar to the presented trial by clicking on the corresponding number box. After each selection, the six real words were cleared up from the screen before going on to the next trial.

The firstly presented stimuli block (the first twenty-four trials) served as a practising block for the participants to get familiar with the experiment. The results obtained from these twenty-
four trials were not counted (this was unknown by the participants). The participant’s judgement on the aspiration condition (aspirated or unaspirated) of each stimulus was recorded with place errors ignored. The average response proportion for each stimulus condition among the ten participants was then summarized for further analysis.

RESULT

The twenty-four stimuli were categorized into the following eight conditions with the differences in place ignored. For example, all the unmodified /pʰaᵢ/, /tʰaᵢ/ and /kʰaᵢ/ stimuli (Stimulus 1) were grouped under c1 [Av-An-L].

Table 1. Conditions of stimuli with place errors ignored

<table>
<thead>
<tr>
<th>Condition</th>
<th>Combination of acoustic parameters</th>
<th>Short form</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>Aspirated vowel - ✓ Aspiration noise - Long VOT</td>
<td>[Av-An-L]</td>
</tr>
<tr>
<td>c2</td>
<td>Aspirated vowel - ✓ Aspiration noise - Short VOT</td>
<td>[Av-An-S]</td>
</tr>
<tr>
<td>c3</td>
<td>Aspirated vowel - ✗ Aspiration noise - Long VOT</td>
<td>[Av-Un-L]</td>
</tr>
<tr>
<td>c4</td>
<td>Aspirated vowel - ✗ Aspiration noise - Short VOT</td>
<td>[Av-Un-S]</td>
</tr>
<tr>
<td>c5</td>
<td>Unaspirated vowel - ✓ Aspiration noise - Long VOT</td>
<td>[Uv-An-L]</td>
</tr>
<tr>
<td>c6</td>
<td>Unaspirated vowel - ✓ Aspiration noise - Short VOT</td>
<td>[Uv-An-S]</td>
</tr>
<tr>
<td>c7</td>
<td>Unaspirated vowel - ✗ Aspiration noise - Long VOT</td>
<td>[Uv-Un-L]</td>
</tr>
<tr>
<td>c8</td>
<td>Unaspirated vowel - ✗ Aspiration noise - Short VOT</td>
<td>[Uv-Un-S]</td>
</tr>
</tbody>
</table>

In summarizing the aspiration responses made by the ten participants on all stimuli, all the place errors were ignored. For example, in perceiving the unmodified /pʰaᵢ/ stimulus, the choice of either ʰ/pʰaᵢ/, ʰ/tʰaᵢ/ or ʰ/kʰaᵢ/ was regarded as an aspirated response. Otherwise, it was recorded as an unaspirated response.
The following graph shows the average response proportions of the eight conditions [c1-c8] made by the ten participants.

![Average response proportion made by all participants](image)

Figure 1. The average response proportions of all conditions made by all participants

Nonparametric test was used in this study to calculate the significant level of differences between different conditions due to the presence of ceiling or bottom effect for most of the conditions. The response proportions made by the subjects did not fall into the normal distribution. Wilcoxon matched-pairs tests were chosen to be an alternative for parametric \( t \) test. It is distribution free and suitable for the measurement with skewed distribution. Also, it is popularly used in the studies of behavioral sciences to test for the significant level of differences between dependent samples (Siegel & Castellan, 1988; Maxwell & Satake, 1997).

**Independent effects**

Each stimulus condition is a combination of three acoustic parameters (vowel
characteristics, aspiration noise level and VOT duration). The 'independent effect' of an acoustic cue reflects its importance on aspiration perception in Cantonese initial stops under all combinations of parameters' conditions.

For vowel characteristics, its independent effect was calculated by comparing the average aspirated response proportions (ARP) of the conditions with ‘Aspirated vowel’ (c1 to c4) to that of the conditions with ‘Unaspirated vowel’ (c5 to c8). For aspiration noise level, the average ARP of the conditions with ‘\(\checkmark\) Aspiration noise’ (c1, c2, c5, c6) was compared to that of the conditions with ‘\(\times\) Aspiration noise’ (c3, c4, c7, c8). Then, the average ARP of the conditions with ‘Long VOT’ (c1, c3, c5, c7) was compared to that of the conditions with ‘Short VOT’ (c2, c4, c6, c8) to find the independent effect of VOT duration.

From the Wilcoxon matched-pairs tests, the average ARP decreased significantly from ‘Aspirated vowel’ conditions (92.1%) to ‘Unaspirated vowel’ conditions (28.3%), \(T = 0, N = 10, p = .005\). Significant dropping in average ARP was also found from ‘\(\checkmark\) Aspiration noise’ conditions (73.9%) to ‘\(\times\) Aspiration noise’ conditions (46.5%), \(T = 0, N = 10, p = .005\). and from ‘Long VOT’ conditions (73.1%) to ‘Short VOT’ conditions (47.4%), \(T = 0, N = 10, p = .005\). It was concluded that all three acoustic cues have significant independent effects on aspiration perception.

In order to rank for the independent effect of the three acoustic parameters, the average ARP of the ‘Aspirated vowel’ (92.1%), ‘\(\checkmark\) Aspiration noise’ (73.9%) and ‘Long VOT’ (73.1%) conditions were compared. It was found that vowel characteristics had the greatest
independent effect on aspiration perception in Cantonese initial stops where the independent effects of aspiration noise level and VOT duration were similar.

Vowel characteristics

With a near to ceiling average ARP (92.1%) for the presence of 'Aspirated vowel' (c1 to c4) and near to bottom average ARP (0.97%) for the presence of 'Unaspirated vowel' (c6 to c8) except for c5, vowel characteristics were claimed to be the dominant cue for aspiration perception in Cantonese initial stops. They had an absolute effect on aspiration perception over all stimuli conditions except for the co-existing of '✓ Aspiration noise' and 'Long VOT' (c5).

Aspiration noise level

Without 'Aspirated vowel' or 'Long VOT', the presence of '✓ Aspiration noise' alone did not have any significant effect on aspiration perception. For example, when comparing c6 [Uv-An-S] to c8 [Uv-Un-S], there was no significant different between their ARP, where c6 = 0.9%, c8 = 0.6%, T = 2, N = 10, p = .593.

When the '✓ Aspiration noise' was accompanied by the 'Aspirated vowel', it did have some significant effect on aspiration perception of Cantonese initial stops though it was not the determining factor. Significant different was found between the ARP of c2 [Av-An-S] (99.7%) and that of c4 [Av-Un-S] (88.3%), T = 0, N = 10, p = .018.

With the presence of 'Unaspirated vowel', aspiration noise level did become a determining factor in c5 [Uv-An-L] where '✓ Aspiration noise' and 'Long VOT' were co-
existing. Significant different was found between the ARP of c5 [Uv-An-L] (97.7%) and that of c7 [Uv-Un-L] (14.1%), T = 0, N = 10, p = .005.

VOT duration

Although the ARP for the presence of 'Long VOT' alone without 'Aspirated voice' or '✓ Aspiration noise' was low (14.2%) and was not sufficient to be the determining factor over the other two acoustic parameters, its effect in cueing for the aspirated response was still found to be significant {comparing the ARP of c7 [Uv-Un-L] (14.1%) to that of c8 [Uv-Un-S] (0.6%), T = 3, N = 10, p = .021}. However, its effect became insignificant when it was accompanied by 'Aspirated vowel' {comparing the ARP of c3 [Av-Un-L] (83.0%) to that of c4 [Av-Un-S] (88.3%), T = 10, N = 10, p = .263}.

As mentioned, with the co-existing of '✓ Aspiration noise' and 'Long VOT', VOT duration became one of the determining factors on aspiration perception in c5 [Uv-An-L]. Significant different was found between the ARP of c5 [Uv-An-L] (97.7%) to that of c6 [Uv-An-S] (0.9%), T = 0, N = 10, p = .005. When either 'Long VOT' or '✓ Aspiration noise' was absent in the 'Unaspirated vowel' conditions [c6 to c8], over 85% of the stimuli were perceived as unaspirated.

DISCUSSION

This study investigated the significant acoustic cues used in the perception of aspiration contrast in Cantonese initial stops by normal hearing adults.

The results of the present study showed that in perceiving Cantonese words with initial
stops, the characteristics of the vowel following the stops were the most determining acoustic
cue for aspiration perception. The presence of 'Aspirated vowel' was almost a guarantee for
the stop to be perceived as aspirated. This effect was independent from the conditions of
aspiration noise level and VOT duration.

For stops followed by 'Unaspirated vowel', they were only perceived as aspirated with
the co-existing of '✓ Aspiration noise' and 'Long VOT'. Either one of these two cues alone
was not sufficient to cue for the perception of aspiration in the presence of 'Unaspirated
vowel'.

VOT duration

The weak cueing effect of VOT duration on aspiration perception in this study
supported the findings in Tsui and Ciocca (2000). It was also consistent with the study of
Revoile et al. (1987). The insertion of Silence-for-aspiration phase (Short VOT to Long VOT)
in the voiced stops in Revoile's study was insufficient to cue for the voiceless response
(voiced to voiceless response proportion: 78% to 22%).

Aspiration noise level

In the present study, the aspiration noise level was found to be as ineffective as the VOT
duration in cueing for aspiration response when '✓ Aspiration noise' was present alone. Only
the co-existing of '✓ Aspiration noise' and 'Long VOT' (c5) was sufficient to cue for the
aspirated perceptual responses.

This contrasted to certain degree with the results of Tsui and Ciocca (2000) and Revoile
et al. (1987). Both of them claimed that people relied heavily on aspiration noise level in the perception of aspiration contrast in Cantonese initial stops or voicing contrast in English initial stops. They concluded that aspiration noise level was a stronger cue on aspiration or voicing perception of initial stops than VOT duration. Revoile et al. (1987) even stated that “the aspiration cue was dominant over the transition cue to voicing perception for the normal-hearing group”.

1. Language difference.

The difference between the results of the present study and that of Revoile et al. (1987) could be accounted by the fundamental different on aspiration perception in Cantonese initial stops and voicing perception in English initial stops.

2. Ambiguous stop-vowel boundary.

Other than the possible language difference, the contradicting results might be caused by the ambiguous stop-vowel boundary. This boundary lies across the transition from the end of the aspiration phase of the stop-burst to the voice-onset. The transition duration only lasts for about 50ms (Kent & Read, 1992). With the differences in voice-onset locations between aspirated and unaspirated stops in related to the transition interval, the significance of formant transitions remained after voice-onset was claimed to be a possible candidate responsible for aspiration perception (Stevens & Klatt, 1974).

It was found that significant formant transitions would only appear after the voice-onset (stop-vowel boundary) of the unaspirated stops. However, when the stop-vowel boundary
was not placed sharply on the voice-onset point in editing the stimuli but varies a bit towards the vowel portion, the segment of formant transitions might have been cut away. If this formant transitions segment was the only marker to discriminate the unaspirated stop from the aspirated vowel, the dominant cueing effect of the vowel portion would be undiscovered. Therefore, the significant independent effect of aspiration noise level, so as that of the VOT duration, on the aspiration perception might become prominent after eliminating the dominant role of the vowel characteristics.


In Tusi and Ciocca (2000), the perceptual response for the Short VOT condition (equivalent to c4 - [Av-Un-S]) showed that normal hearing people were sensitive to the reduction in VOT accompanied by the removal of aspiration noise from the aspirated stops. Significant shift of perceptual response from aspirated to unaspirated was resulted. For the Long VOT condition (equivalent to c7 - [Uv-Un-L]), the increase in VOT of the unaspirated stops alone without the addition of aspiration noise did not have any significant effect on aspiration perception. These come up with their conclusion that the aspiration noise level was a stronger cue on aspiration perception in Cantonese initial stops than VOT duration did.

However, they did not include the [Uv-An-S] condition from the present study, which reflected the effect of the addition of aspiration noise in the short VOT of the unaspirated stops. Without investigate the [Uv-An-S] condition, it was not sufficient to conclude that aspiration noise level is a stronger cue than VOT duration because the independent effect of
aspiration noise level on aspiration perception was still not found.

In the present study, the low ARP (14.1%) for c7 [UV-Un-L] was consistent with that of the Long VOT condition in Tsui and Ciocca (2000). However, the low ARP (0.9%) for c6 [UV-An-S] indicates that the addition of aspiration noise without increasing the VOT was still not enough to cue for the aspirated response. In the presence of ‘Unaspirated vowel’, only the co-existing of ‘✓ Aspiration noise’ and ‘Long VOT’ (c5) was sufficient to cue for the aspirated responses (ARP = 97.7%).

Therefore, the study results of Tsui and Ciocca (2000) could only used to conclude that the independent effect of VOT duration was insufficient to cue for aspiration perception in Cantonese initial stops. However, they were not enough to conclude for the independent effect of the aspiration noise level on aspiration perception. The same limitation was also found in Revoile’s study (1987).

Vowel characteristics

As the presence of ‘Aspiration vowel’ was almost a guarantee for aspirated perceptual response in this study, some of its characteristics, that were not present in the ‘Unaspirated vowel’, must be responsible to cue for the perception of aspiration. Although the underlining determining characteristic(s) of vowel had not been further investigated in this study, we could still hypothesize some possible characteristics from the study results together with the review in literatures.
1. Formant transitions.

In the previous studies, formant transitions were suggested to have heavy weighting in voicing perception (Stevens & Klatt, 1974; Revoile et al., 1987; Kent & Read, 1992). Well-defined first formant (F1) transitions can be found after the voicing-onset in the voiced stops. For the voiceless stops, these transitions are usually started prior to or even completed before the voicing-onset (Stevens & Klatt, 1974). 'F1 cutback' was introduced by Kent and Read (1992) to be the delay in the release of F1 energy for voiceless stops in compare with other higher formant energy.

These transitional-related phenomena might be the characteristics of vowel responsible for the aspirated perceptual responses in this study if they were applicable in Cantonese. Tsui and Ciocca (2000) also supported for the cueing effects of formant transitions on aspiration perception, which were especially important for the hearing-impaired listeners to rely on.

2. Fundamental frequency.

According to Kent and Read (1992), the onset fundamental frequency of the vowel following the English voiceless stops is higher than that following the English voiced stops. Although Revoile et al. (1987) concluded that the onset value of the fundamental frequency of the vowel was not an effective cue for voicing perception in related to other transitional information, its effect on aspiration perception in Cantonese stops was still unknown.

Again, this was a possible candidate for accounting for the results observed in the present study.
3. Carrying on of aspiration noise after vowel-onset.

The phenomenon of carrying on of aspiration noise after vowel-onset was observed from the digital waveform displays of the stimuli. (Appendix B) The first few cycles of the 'Aspirated vowel' are found to be more fluctuating and less sharp than the first few cycles of the 'Unaspirated vowel'. The breathy noise liked fluctuating waveform indicated that noise from the aspiration phase of the aspirated stop had entered into the first few cycles of the vowel. The noise was not cut out sharply by the stop-vowel boundary. A short period of fade-out phenomenon occurred after the vowel-onset. Therefore, all the conditions containing the 'Aspirated vowel' were accompanied by a small amount of aspiration noise in the first few cycles of the vowel. If this breathy noise liked period was perceived by the normal hearing listeners, it could become a marker for aspiration.

Trading relationship- F1 onset frequency and VOT

The trading relationship between the onset frequency of the first formant (F1) and the VOT on the perception of voicing was suggested by Summerfield and Haggard (1977). In their study, the perceived frequency of F1 at the onset of voicing, rather than the presence of significant F1 transitions, was claimed to be the major effect of F1 on voicing perception in English initial stops. Due to the trading relationship, the VOT boundary for voicing contrast varies with the F1 onset frequency of the following vowel. Longer VOT was required to cue for the perception of voicelessness with low F1 onset frequency, and vice versa. Therefore, the effect in voicing perception caused by varying the VOT duration maybe, actually, induced
by the variation in the F1 onset frequency.

In Cantonese initial stops, the voicing-onset also starts before the stop-vowel formant transitions (from low frequency to high frequency) for the unaspirated stops but during or even after the transitions for the aspirated stops. Therefore, the F1 frequency during voicing-onset of the aspirated stops should be higher than that of the unaspirated stops. Trading relationship was seemed to be applicable in aspiration perception of Cantonese initial stops.

In the present study, 'Aspirated vowel' was found to have a comparatively high F1 onset frequency (689Hz to 904Hz). According to Summerfield and Haggard (1977), only a short VOT duration was required to cue for aspirated perceptual response. Therefore, stimuli with 'Aspirated vowel' accompanied by 'Short VOT' were still enough to be perceived as aspirated (c2 and c4). On the other hand, as the F1 onset frequency of 'Unaspirated vowel' was found to be comparatively low (517Hz to 646Hz). A long VOT duration was suggested to be required to cue for the aspirated perceptual response. Together with the co-existing effect between 'Long VOT' and "Aspiration noise", this could accounted for the high ARP (97.7%) of c5 [Uv-An-L] from this study.

In order to prove for the hypothesis of trading relationship between F1 onset frequency and VOT on aspiration perception in Cantonese initial stops, it is recommended to carry out an experiment to investigate the effect in changing F1 onset frequency on the location of VOT boundary for the aspiration contrast in the perception of Cantonese initial stops.
Covariation and interaction effects

Till now, the whole study was mainly focusing on the discovery of particular acoustic parameter, which had an absolute determining effect on aspiration perception in Cantonese initial stops. However, if further study is going to be carried out on investigating the cueing effects of each characteristic of vowel (e.g. formant transitions, fundamental frequency, F1 onset frequency, etc) on aspiration perception, the possibility of having covariation and interaction effects among the characteristics should not be ignored.

The covariation of the F1 onset frequency, the duration of F1 transition and its extent suggested by Summerfield and Haggard (1977) was a typical example of covariation effect. Even the dimensions of aspiration noise level and VOT duration were proved to have interaction effects on aspiration perception in this study. Therefore, although the independent effect of each characteristics of vowel is clinically valuable, the possible effects brought by the covariation and interaction relationships among different vowel characteristics should also be carefully considered.

Perceptual values of varying the acoustic dimensions

In this study, the digital waveform display and the spectrogram were used to analysis for the acoustic parameters of the stimuli and carry out editing. Clinically, these devices are also helpful in providing speech intervention and aural rehabilitation training. Objective and quantitative visual feedback on the acoustic features of speech sound can be given to the client. In addition, they are particularly useful in setting concrete and measurable intervention
goal and monitoring treatment progress.

However, as it is found that not all changes in acoustic parameters showed on the digital waveform display or the spectrogram are perceptually valuable, focusing on the right acoustic parameters is the essential way to make good use of the visual information.

For example, the present study claimed that either the presence of ‘Aspirated vowel’ or the co-existing of ‘✓ Aspiration noise’ and ‘Long VOT’ was sufficient for a stop to be perceived as aspirated. Therefore, one of the targets of the intervention program for a speech patient with deaspiration process can be focus on increasing both the aspiration noise level and VOT duration for the production of aspirated stops. Digital waveform display or the spectrogram can be used during the intervention to give objective and quantitative feedback to the patient. However, it may be meaningless if only the VOT duration of the patient’s production is worked on. The results of this study suggested that even the patient can reach the target VOT duration in producing aspirated stops according to the measure in the digital waveform display or the spectrogram, his production will still be perceptually unaspirated without the co-existing of ‘✓ Aspiration noise’.

Therefore, the corresponding perceptual value of each acoustic parameter of the speech sound should be measured first before setting them as the treatment targets. Only working on those visually but not perceptually noticeable acoustic parameters is not clinically valuable.
CONCLUSION

On aspiration perception in Cantonese initial stops, the characteristics of the following vowel were the dominant cue relied on by the normal hearing adults. Initial stops with 'Aspirated vowel' and 'Unaspirated vowel' were perceived to be aspirated and unaspirated respectively. The only exception was the co-existing of high aspiration noise level and long VOT duration, which was perceived to be aspirated even with the presence of 'Unaspirated vowel'. However, without the co-existing relationship, the independent effect of either one of them was insufficient to cue for the aspirated perceptual response.
ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisors, Dr. Valter Ciocca and Dr. Alexander Francis, for their invaluable comments and supports. I would also like to give special thanks to my speaker, Mr. Jason Kan S. L. and all my dearest participants.
REFERENCES


Figures a, b and c indicating the formation of 'short VOT with high aspiration noise' condition (Figure c). The short VOT duration in Figure b (approximately from 7.5ms to 13ms) was filled up by the beginning segment (approximately 10ms to 15ms) and ending segment (approximately 62ms to 65ms) of the aspiration phase of Figure a. Figure c was the outcome containing high aspiration noise in short VOT.
Figure d. Digital waveform display of 'Unaspirated vowel'.

Figure e. Digital waveform display of 'Aspirated vowel'.

Figure d and e showed the first few cycles of 'Unaspirated vowel' and 'Aspirated vowel' respectively. Those waveform of 'Aspirated vowel' (figure e) was more fluctuating than that of 'Unaspirated vowel' (figure d). This indicated that aspiration noise had entered into the first few cycles of the 'Aspirated vowel', which might be responsible for the aspirated perceptual responses on the stimuli with 'Aspirated vowel'.