INTRODUCTION

Normal phonation (EP) involves an outward airflow generated by the respiratory system, which drives the adducted vocal folds into vibration. Inspiratory phonation (IP), however, requires a reverse flow of air which brings air into the lungs through the glottis. The unique voicing mechanism of IP has been examined by researchers. For example, Orlikoff et al. (1997) examined sustained vowels produced using IP by using a combination of acoustic, electrogastrographic (EGG), and stroboscopic measurements. They found that IP was associated with an increased airflow, decreased amount of vocal fold contact, and a higher fundamental frequency (F0) when compared with EP. Kelly and Fisher (1999) examined the acoustic and stroboscopic data obtained from sustained /i/ using IP and found similar results. They also noted that intensity level of IP vowels failed to vary consistently with increase of F0. Robb et al. (2001) further examined the interaction between articulation and phonation during IP and suggested that the articulatory-phonatory coordination during IP was not precisely the same as that during EP. However, they did not explain the details of such unique laryngeal maneuver. All of these studies mainly focused on IP productions in isolated vowel context. The mechanism of IP production of vowels in a CV context is not investigated. It is not known if and how IP affects the coordination between phonation and articulation.

The present study examined the timing coordination between the phonatory and articulatory systems during EP and IP. VOT values obtained from Cantonese CV syllables were correlated with listeners’ perceptual identification of stops produced using EP and IP. In Cantonese, all stops are voiceless and contrasted by aspiration and three places of articulation (bilabial, alveolar and velar), yielding a total of six different stops. By comparing the VOT characteristics of these six Cantonese stops produced using EP and IP, the effect of phonation (EP vs. IP) on interarticulator timing control could be revealed.

METHOD

Participants. Thirty (15 male &15 female) healthy adults who were 19 – 28 years of age (mean age = 22.53 years) served as speakers in the study. All of them were capable of sustaining the vowel /a/ steadily using IP for at least 2 seconds. Five male and five female adult Cantonese speakers with ages ranging from 19 to 24 years (mean age = 20.40 years) served as listeners. Both participants and listeners were native speakers of Hong Kong Cantonese with no known history of respiratory, vocal and/or auditory problems.

Speech Tasks. Three aspirated (/pʰ, tʰ, kʰ/) and three unaspirated (/p, t, k/) voiceless stops followed by vowel /a/ were used in the study. This provided a complete set of contrast of voiceless unaspirated and voiceless aspirated stops at three places of articulation (bilabial, alveolar and velar). As lexical tone is beyond the focus of the study, all CV syllables were produced at the high-level tone. This yielded six meaningful Cantonese monosyllabic words: /pʰa/ “to grovel”, /tʰa/ “to help”, /kʰa/ “to stop”, /pa/ “father”, /ta/ “toad”, /ka/ “to add”. All Cantonese words were embedded in a carrier phrase: /tʃ/k/ta/ meaning “Read word”.

Procedure. Speech samples were recorded by using a high-quality microphone (Shure, SM 58) with a 10-cm mouth-to-microphone distance and a preamplification unit (M-Audio, MobilePre USB). Acoustic signals were digitized at 20 kHz and 16 bits/sample by using Praat. Of the three trials recorded, only the last (a total of 720 productions) two were chosen for VOT measurement and listening experiment. VOT values were measured from the waveform of speech samples, with the help of a wideband spectrogram (window size = 25 ms). The speech samples were also used in the perceptual experiment during which the listeners were asked to identify the different stops perceived.

Statistical analysis. A 2 (phonation type) x 2 (aspiration) x 3 (place of articulation) three-way Analysis of Variance (ANOVA) was used to determine if significant differences were present in VOT between EP and IP, between aspirated and unaspirated stops, and among the bilabial, alveolar, and velar stops.

RESULTS AND DISCUSSION

Perception. Identification of Cantonese stops produced using EP achieved a level of accuracy of over 90% (see Table 1). Misperception of place of articulation appeared to be the predominant error, with few errors in aspiration. However, all Cantonese stops produced using IP were identified at a significantly lower level of accuracy (< 65%) (see Table 2). Aspiration confusion appears to be the common error pattern, with relatively higher accuracy in identification of place of articulation. Listeners demonstrated difficulties in distinguishing between aspirated and unaspirated stops produced at same place of articulation using IP. They also misidentified place of articulation of Cantonese stops produced using IP. This may be related to the small VOT contrast between aspirated and unaspirated stops in IP.

Aspiration.

<table>
<thead>
<tr>
<th>Aspirated</th>
<th>Unaspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP 96.86 ms</td>
<td>15.66 ms</td>
</tr>
<tr>
<td>IP 37.30 ms</td>
<td>28.24 ms</td>
</tr>
</tbody>
</table>

Place of articulation.

For EP: /kʰ, k/ > /tʰ, t/ > /pʰ, p/.

For IP: /kʰ, k/ > /tʰ, t/, /pʰ, p/.

Due to the smaller VOT difference between aspirated and unaspirated IP stops.

- Matches with perceptual results.
- VOT of EP stops significantly increases as the place of articulation changes from bilabial to velar. Suggested reasons:
  1. The introral pressure behind the occlusion,
  2. Speed of articulator movement, and
  3. The extent of articulatory contact area.

Phonation (EP vs. IP). VOT of EP > VOT of IP for aspirated, but VOT of EP < VOT of IP for unaspirated stops. VOT difference between EP and IP is due to the physiological difference of the two phonatory mechanisms:

1. Difference of air columns and air reserve capacity between EP and IP in the production of aspirated stops: Aspirated EP stops are produced using two air columns due to limited natural lung capacity: the first air column is used to open the oral occlusion (conditions: oral occlusion + vocal folds abducted), and the second air column is used to set vocal folds into vibration (conditions: sufficient subglottal pressure + vocal folds adducted). However, it is likely that aspirated IP stops are produced with only ONE air column due to infinite air reserve capacity of atmosphere. It could be that the ONE air column is employed to release oral occlusion and to render vocal fold vibration with the facilitation of a negative transglottal pressure differential (TPD) (Orlikoff et al., 1997).

2. Difference of mechanisms in establishing adequate TPD, regardless positive or negative, in EP and IP productions.