Abstract—The relation between vertical laryngeal movement and vocal frequency (F0) change has attracted the attention of many researchers. This paper studies the vertical laryngeal position (larynx height) during the four tones of Mandarin, based on the X-ray movie data from one female speaker, with 36 monosyllables in tones. The location and movement of the larynx have been measured, and compared with the variation of the F0 during the four tones. Results show that, during the first tone, the larynx height is weakly correlative with F0, while during the other three tones the larynx height is positively correlative with F0. This suggests that the upward movement of larynx is partially (in tone 1) independent of F0, while the downward movement is accompanied by decreasing F0. And the quantitative mechanism of larynx height and F0 during the four tones can be integrated into articulatory model of vocal tract in Mandarin for better speech synthesis.

Keywords—laryngeal position; tones; Mandarin; X-ray

I. INTRODUCTION

It is well-founded that the vertical laryngeal position (larynx height) is related to the voice fundamental frequency (F0), which roughly means that the larynx moves up and down as F0 rises and falls[1]. This suggests that vertical movement of the larynx is a critical component of F0 control mechanisms, and therefore the relation between vertical larynx movement and F0 change has attracted the attention of many researchers [1-7].

On the other side, in articulatory model during F0 fluctuation, especially for those tonal languages, should the larynx height be adjusted accordingly, which will lead to the change of the vocal tract length? In other words, when the F0 of the synthesized speech rises and falls, should we adjust the larynx height? This will change the length of vocal tract, and accordingly influence the acoustic characteristics of synthesized speech. Specifically, what kind of quantitative mechanism of larynx height and F0 should be integrated into articulatory model for synthesizing tonal speech? This quantitative mechanism during the four tones will bring more precise vocal tract length when synthesizing tonal speech in Mandarin. From engineering point of view, all these also require us to explore the relation between larynx height and F0 [8-10].

However, the empirical observation is difficult because the larynx hides in human body. Previous measurements of larynx position have employed various techniques using optical instruments[2, 4], mechanical facilities[5], X-ray devices[6, 11], and Magnetic Resonance Imaging (MRI)[1,7,12]. Although MRI can give full volume vocal tract images, it is limited to a static, sustainable configuration. The X-ray permits real time tracking, so the full sagittal view of vocal tract including larynx during running speech provided by X-ray movie (cineradiography) remains unsurpassed, although the poor contrast of the cartilages in lateral cineradiography makes it a little technically difficult in measuring laryngeal movements

In the literature, there is scarce data of laryngeal movement in Mandarin, a tonal language with four tones: yin ping, yang ping, shang, qu. Therefore, in this study, the vertical laryngeal movement during the four tones of Mandarin has been observed and investigated, to explore the relation between the vertical movement and the F0 during Mandarin tones.

II. METHOD

A. X-ray movie

An X-ray movie database has been established from the original PAL videotape, which is the only cineradiography video of Mandarin available at present. Both sagittal cineradiography and speech sound are given simultaneously. A female speaker, who is a national broadcaster of Mandarin, pronounced 9 sets of monosyllables with different consonants and vowels in the 4 tones, totally 36 monosyllables in their corresponding characters. The speaker was required to articulate each syllable clearly and rest sufficiently between every two syllables, to avoid the interference of articulatory movements. The monosyllables and their corresponding characters in Mandarin are listed in table 1.

<table>
<thead>
<tr>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>yin ping</td>
<td>yang ping</td>
<td>shang</td>
<td>qu</td>
</tr>
<tr>
<td>ba</td>
<td>巴</td>
<td>ba1</td>
<td>把</td>
</tr>
<tr>
<td>bi</td>
<td>逼</td>
<td>bi1</td>
<td>骂</td>
</tr>
<tr>
<td>bu</td>
<td>逼</td>
<td>bu1</td>
<td>骂</td>
</tr>
<tr>
<td>lao</td>
<td>拽</td>
<td>lao1</td>
<td>拽</td>
</tr>
</tbody>
</table>

This study is sponsored by National Project of Scientific and Technical Supporting Programs Funded by Ministry of Science & Technology of China (NO. 2009BAH41B00).
This video has been transferred to avi files at 25 frames per second (fps), which is the same as original videotape. Next, the avi files are adjusted to make sure that the PAR (Pixel Aspect Ratio) equals 1, which means each pixel in the image is square, to avoid vertical or horizontal distortion. Thereafter, the avi files are cropped with a window of 400*400 pixels, to focus on the movements of articulators and to alleviate the processing burden. Finally, the avi files are segmented to single syllable.

The original speech sound is recorded simultaneously in the X-ray device room, where there exists some kind of noise. However, this will bring little influence to the detection of F0. The audio format is as follows: 44.1 kHz sampling rate, 16bits quantization, 1 channel, PCM wav file.

B. X-ray movie processing

A platform ‘VocalMarker’ in Matlab is programmed to trace the mid-sagittal articulatory movements, including the preprocessing of the images, the semi-automatic tracing of the articulatory movements, and the speech sound processing. Figure 1 is the demo of the ‘VocalMarker’, in which a syllable sample /biao1/ is being processed. The upper window shows the marking of articulatory movement, in which there are several lines with key points to mark the edges of articulators. The shape of larynx is marked by two lines ‘larynx front wall’ and ‘larynx back wall’. The articulators are marked in each frame of the X-ray movie, so finally the movement of the larynx can be obtained. The lower window shows the processing of synchronous speech sound, where the waveform and spectrogram of current syllable are shown. At the same time the acoustic parameters including formant frequency and F0 are calculated.

C. Larynx height measuring

Due to the poor contrast of the cartilages in lateral X-ray movie, we adjust the contrast of the movie, trying to make a precisely positioning of the larynx. Figure 2 shows the measuring of the larynx height. In the X-ray movie, the images of the cartilages are blurry, so the whole shape of the cartilages are not clearly shown. However, we can position the vocal fold as the maximum curvature along the larynx front wall, where is also the most protrudent place of the cricoid cartilage.

The original larynx height is defined as the vertical distance between the vocal fold (the third key point along the larynx front wall) and the lowest place of trachea in this video (the forth point along the larynx front wall). We set the original larynx height during speechless rest as the base larynx height. And finally the larynx height is defined as original larynx height minus the base larynx height. The higher value of larynx height indicates the higher position of vocal fold, and accordingly the shorter length of the vocal tract when other articulatory configurations remain the same. If the value of larynx height is minus, it means that the position of vocal fold is lower than its position during speechless rest.
III. RESULTS

The results are observed and discussed in two ways: the movement of the vertical laryngeal position and the relation between larynx height and F0.

A. The movement of the vertical laryngeal position

Figure 3 shows the movement of the vertical laryngeal position and fluctuation of the F0 during the four tones of the 9 sets of syllables. The x-axis is time in frames, and 1 frame is equivalent to 40 ms, since the fps is 25. The y-axis is F0 in Hz and larynx height in 0.1 mm at the same time for showing them both in the same figure. The red stars are F0 data, which show that the fluctuant F0 curves during the four tones: tone 1 (yin ping), tone 2 (yang ping), tone 3 (shang), tone 4 (qu), from left to right along the time axis. The starts and ends of every syllable are marked by eight vertical dashed lines. The blue points are larynx height data, which shows the rises and falls of the larynx during the four tones with 9 sets of syllables.

As shown in this figure, there usually is a lowering movement of larynx before each syllable with the decreasing larynx height. Moreover, some of the values of larynx height are minus, which means that the larynx goes lower than its position during speechless rest. This indicates the preparing action before the speaker pronounces each syllable. As we can notice, the 9 sets of syllables have different larynx height because of their different vowels, where the vowel /i/ has the maximum larynx height and the /u/ has the minimum one. However, for the moment, the intrinsic F0 and larynx height of different vowels are not considered.

As we can see that, during the tone 1, the larynx moves up while the F0 already starts at a high level around 250Hz and maintains flat, which means the speaker can produce a high tone without lifting up the larynx to a certain height necessarily. During the other 3 tones, basically, the larynx height moves up and down as F0 rises and falls.

B. The relation between the larynx height and F0

Having observed the fluctuation of larynx height and F0 in the time domain, we may explore whether it exists a quantitative correlation between them during the tone 2, 3, and 4, and what kind of mechanism between the larynx height and F0 we can adopt into the articulatory model in Mandarin.

Figure 4 shows the result. There are four sub-windows in which the results during four tones are shown respectively. The x-axis is F0, ranging from 100 to 300Hz, as a typical female speaker’s voice. The y-axis is larynx height, ranging from -6 mm to 6 mm. The different colors and symbols of the points represent the 9 sets of syllables with different consonants and vowels. Still the same, we do not consider the intrinsic F0 of vowels, so we just study the relation between the larynx height and F0.

In the first sub-window (from the left), during tone 1, when the F0 maintains at the same level around 250Hz, the larynx moves up. The correlation analysis result shows that the correlation factor between F0 and larynx height is 0.0274, which indicates they are weakly relevant, or nearly irrelevant.

During tone 2, when F0 rises, the larynx moves up, as indicated by the red line arrow. And the correlation factor between them is 0.5876. Therefore we use linear regression analysis to get the relation between larynx height (LH) and F0: LH = 0.6071*F0 - 135.9667.

During tone 4, when the F0 falls, the larynx moves down, as indicated by the red line arrow. And the correlation factor between them is 0.4639. Therefore we use linear regression analysis to get the relation between them: LH = 0.3055*F0 - 44.2634.

During tone 3, the relation is complicated. At first, the F0 falls from 200Hz to 150Hz, while larynx moves down, and we can use the equation during tone 2 to describe it. And then, the F0 remains low, but the larynx still moves down or up. Here the relation between the larynx height and F0 is still unclear since the phonation type of vocal fold may be different from modal voice. Finally, the F0
rises from around 150Hz to 250Hz, while the larynx moves up, and we can use the equation during tone 2 to describe it.

The changing F0 will bring a fluctuant range from -6 mm to +6 mm to larynx height, which would lead to the varying vocal tract length. The typical length of the whole vocal tract is about 150mm for an adult female speaker, so in articulatory model this will bring an 8% range of vocal tract adjustment during tonal speech synthesizing, thus may be beneficial to Mandarin TTS (Text to Speech).

Moreover, we here express our gratitude to the paper reviewers for their scrupulous work.

IV. CONCLUSION

As to the movement of the vertical laryngeal position and the relation between larynx and F0 during the four tones in Mandarin, there are mainly two different kinds of results: During tone 1, the vertical position of larynx is not correlative with F0. The F0 stays at a high level while the larynx moves up from its rest position, which means the speaker can produce a high tone without lifting up the larynx to a certain height necessarily. During the other 3 tones, the larynx position is positively correlative with F0, which roughly means that the larynx moves up and down as F0 rises and falls. Two equations can describe how the larynx height changes when F0 changing, and the larynx height changes at different ratio to F0 when moving up or down.

The results suggest that the upward movement of larynx is partially (in tone 1) independent of the F0, while the downward movement is accompanied by decreasing F0. This may be explained by the mechanism of the larynx structure, which still needs clarifying unfortunately. One drawback of the study is that the data size is scarce due to the radioactive harm of X-ray. Only a single speaker’s database containing 36 isolated syllables is used. This may need more empirical data and further study to explore the underlying mechanism controlling the movement of larynx and the change of F0. However, from the viewpoints of engineering, the quantitative mechanism of larynx height and F0 during the four tones can be integrated into articulatory model of vocal tract in Mandarin, which will bring more precise vocal tract length when synthesizing tonal speech in Mandarin.

ACKNOWLEDGMENT

We give our devout appreciation to Prof. Huaqiao BAO for providing the original videotape of the X-ray movie.

REFERENCES


Figure 4. The relation between larynx height and F0 during the four tones

<table>
<thead>
<tr>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0(Hz)</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Larynx height (0.1mm)</td>
<td>-100</td>
<td>-100</td>
<td>-100</td>
</tr>
</tbody>
</table>

Table 4. The relation between larynx height and F0 during the four tones.