Tongue Shape Variation Model for Simulating Mandarin Chinese Articulation

Jinguang Zhang, Xiyu Wu, Jiangping Kong
Department of Chinese Language and Literature, Peking University

We studied tongue shapes extracted from X-ray films which were taken during the process of mandarin Chinese articulation. Through factor analysis, we built an eight-parameter-driven tongue articulation model. This study reveals that the front of the tongue has large horizontal movement; the blade of the tongue has large vertical movement; whereas the back, as well as the root, of the tongue has small movement both horizontally and vertically. This model can be used to drive a 3D tongue model to control its articulatory behavior.

METHOD

Our mandarin Chinese X-ray database contains AVI format videos which were converted from PAL standard X-ray films. Contours of speech organs have been drawn manually. We made Visual C++ programs to process them and found out patterns of tongue shape variation. In order to separate shape variation from position change, we took tongue shape as an inscribed graphics of a rectangle. So we got five major feature points and eleven minor feature points, as shown in Figure 1.

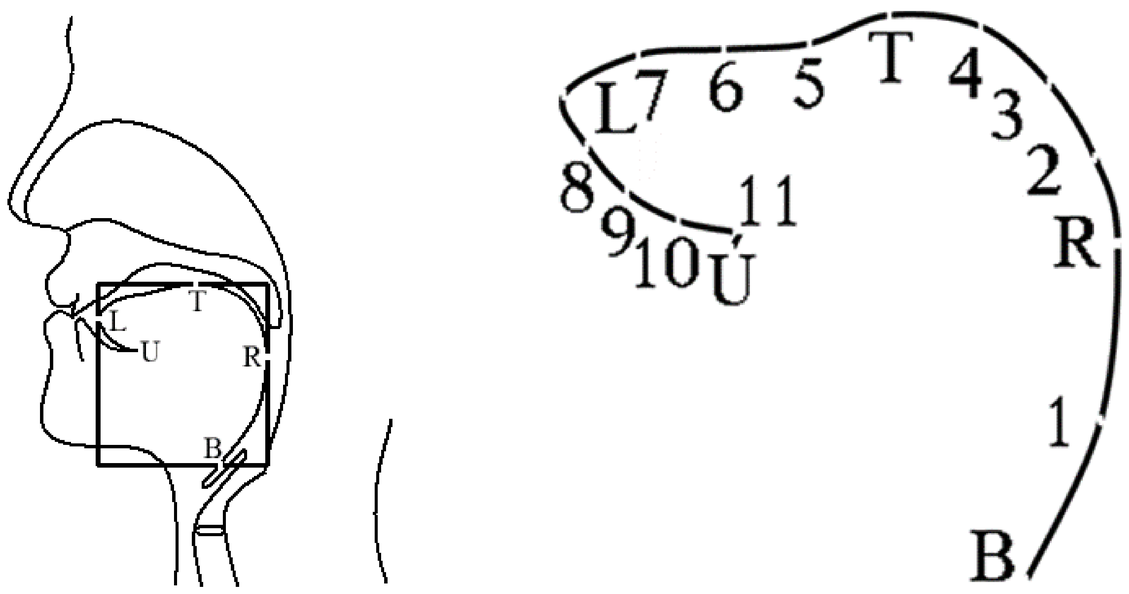
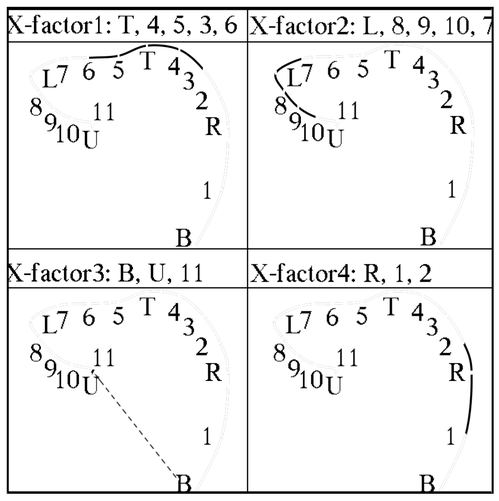
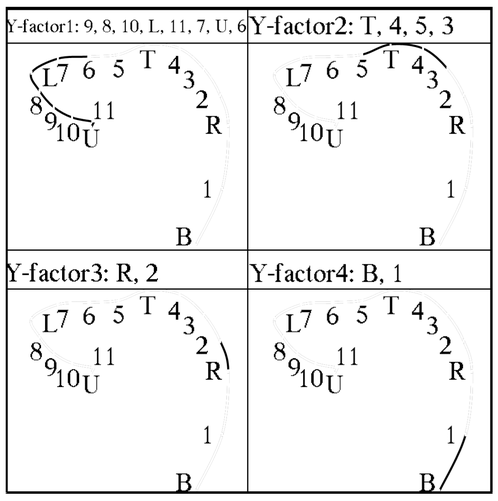


Figure 1: The choice of key feature points.

In our research, we used Cartesian coordinates. The coordinate origin is at the top-left corner on X-ray images. The X axis is positive along the right, and the Y axis is positive directly down. The extracted x, y coordinates were written into an Excel sheet, each frame as a record. There are totally 12657 records which are from 276 AVI videos. Each record contains 32 fields for coordinates of 16 key feature points, and a Name field for syllable name and frame number. Considering there are correlation among parameters, we need to do Principal Components Analysis and Factor Analysis to discover the underlying factors and to reduce dimension.

RESULTS





In the following figure 2, the model-simulated graphics (the outer curves) are compared with the original bitmap graphics (the inner curves) for the mandarin Chinese syllable [t^hu].

Figure 2: Frame 20, 25 and 42 of syllable [t^hu].

Our model matches well with speech spectrogram. We choose the data of syllable [piau] to illustrate the relationship between factor scores and speech formants in figure 3 to 5.

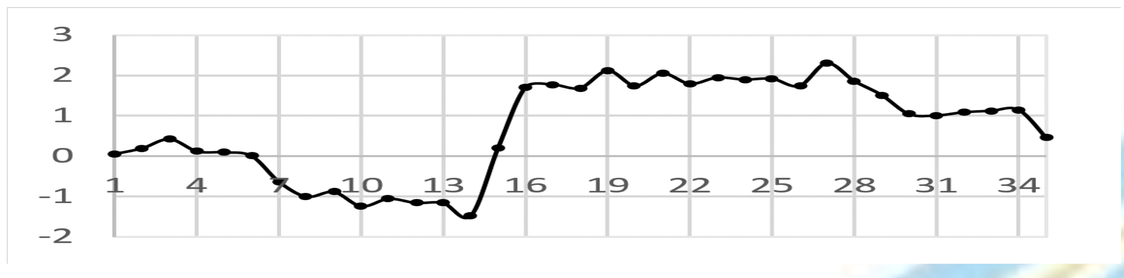


Figure 3: X-Factor1 scores of [piau].

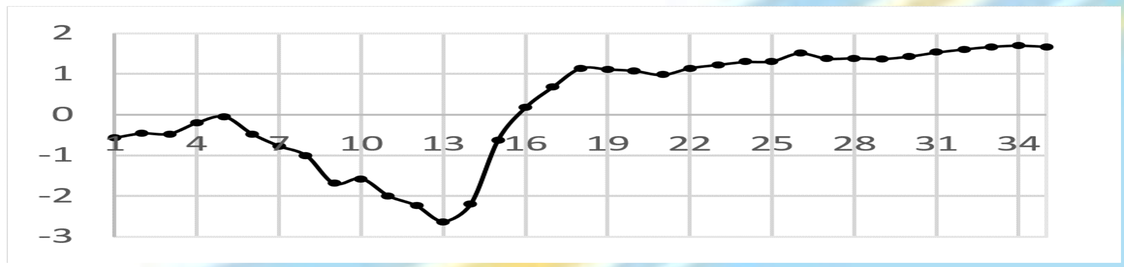


Figure 4: Y-Factor2 scores of [piau].

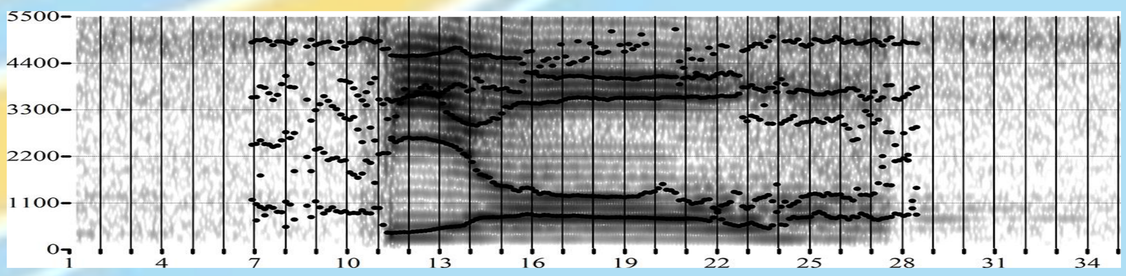


Figure 5: Spectrogram and formants of [piau].