Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

A Study on Model of Vocal Tract in Mandarin

北京大学中文系语言学语言学实验室
北京大学汉语语言学研究中心
Dept. of Chinese and Research Center of Chinese Linguistics
Peking University, China

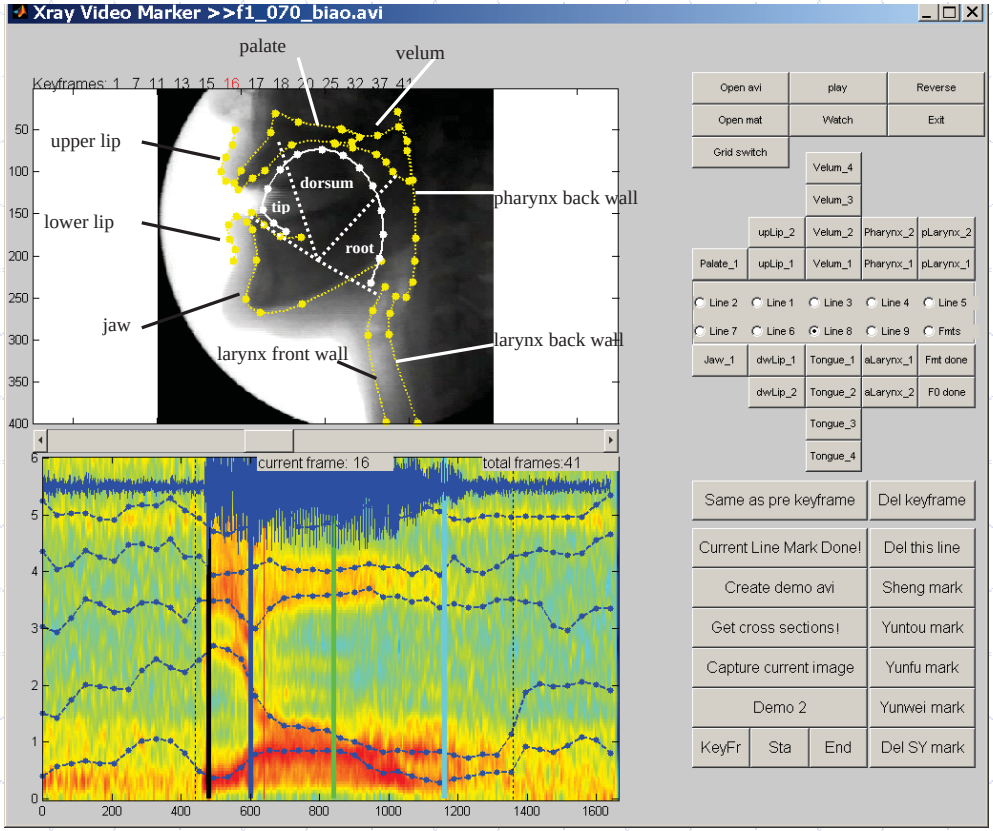
Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

X-ray Video of Mandarin

Gaowu Wang, Jiangping Kong, Huaiqiao Bao (2008). “Formant Estimation from Vocal Tract
with Reference to Mandarin Vowels”, Journal of Chinese Phonetics, Vol.1,pp. 164-170, 2008/04

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

X-ray Image Processing Platform



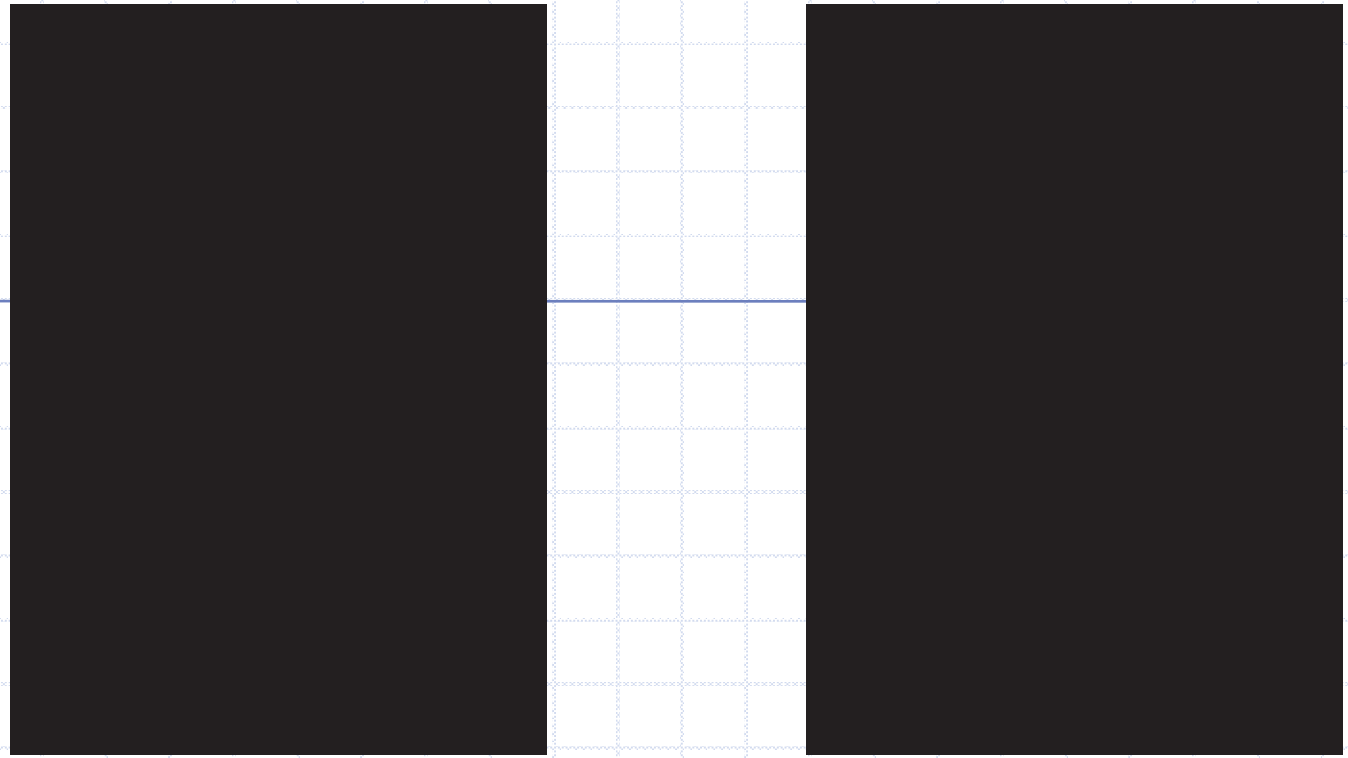
X光_单音节

Gaowu Wang, Jiangping Kong (2010) “The Relation between Larynx Height and F0 during the Four Tones of Mandarin in X-ray Movie”, ISCSLP2010 Taiwan

X光_儿化音□

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

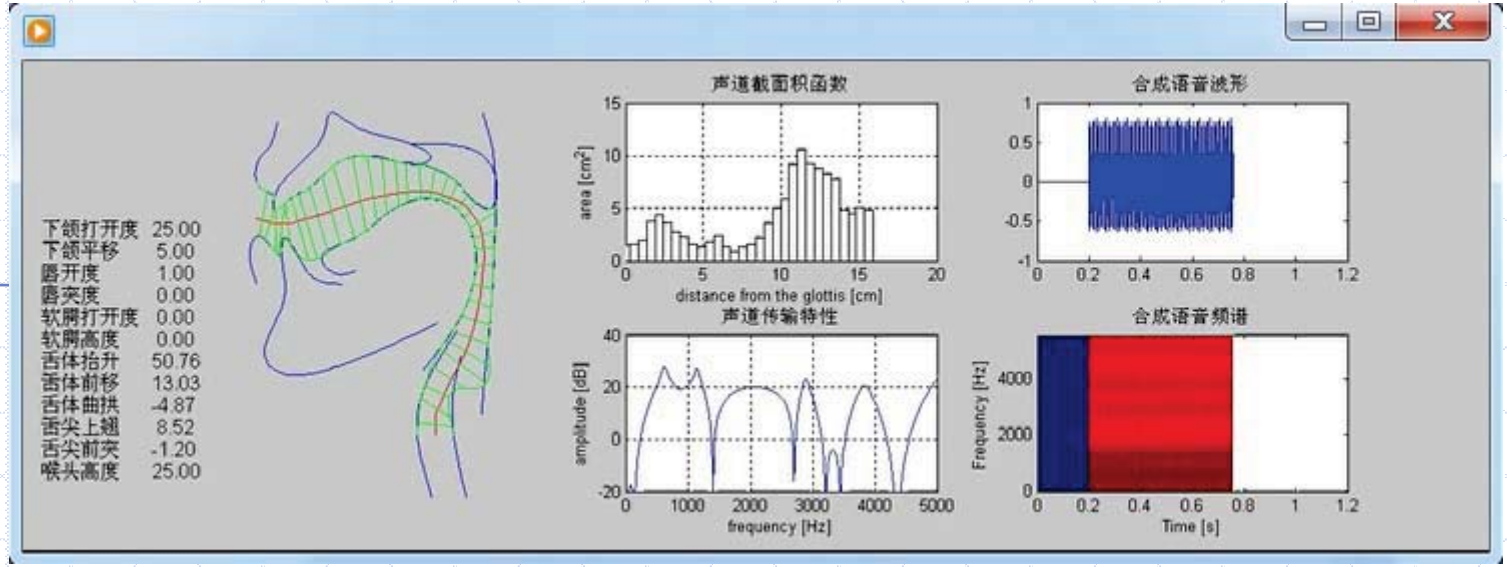
X-ray Video of Mandarin



Play video samples

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

An 2D Articulatory Model of Mandarin by X-ray Video



AVI

5_demo_01a_LFmodel_addsound.avi

Play the learning system

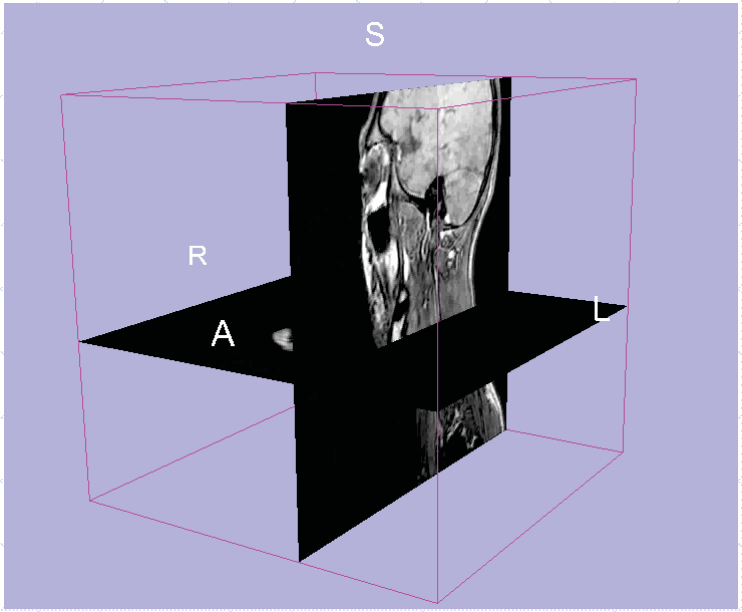
Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

An Putonghua Learning System Based X-ray Video

Play the learning system

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

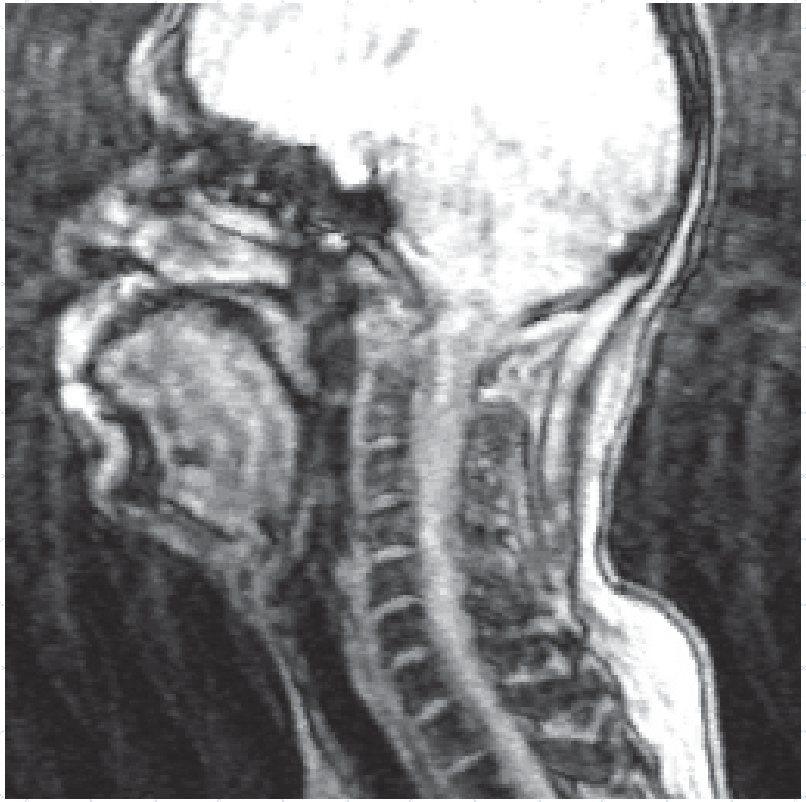
Study on Mandarin by X-ray Video



MRI sampling at ATR in Japan

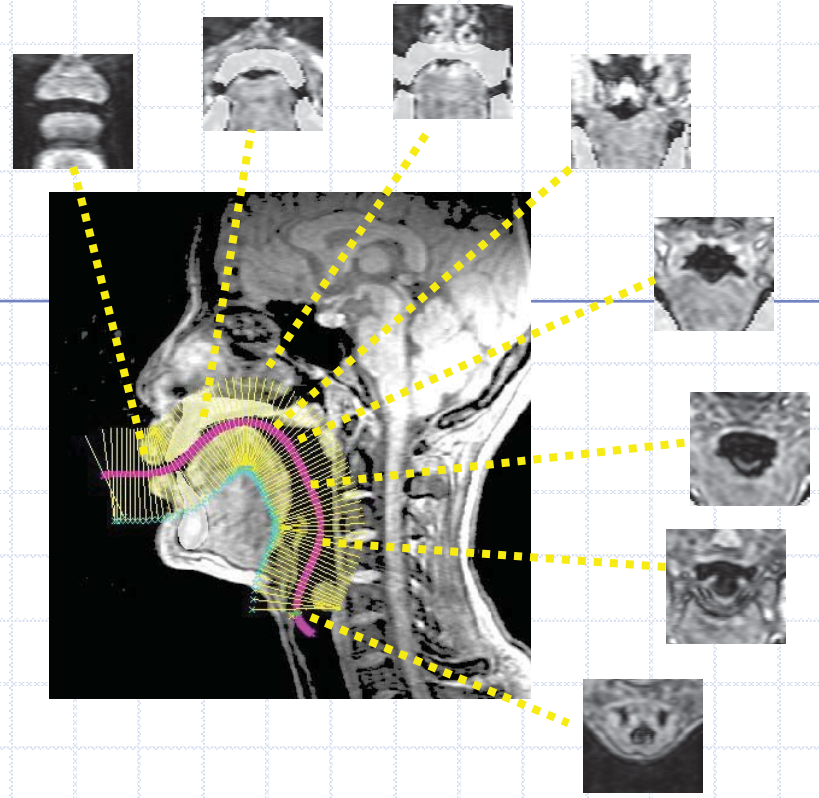
Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

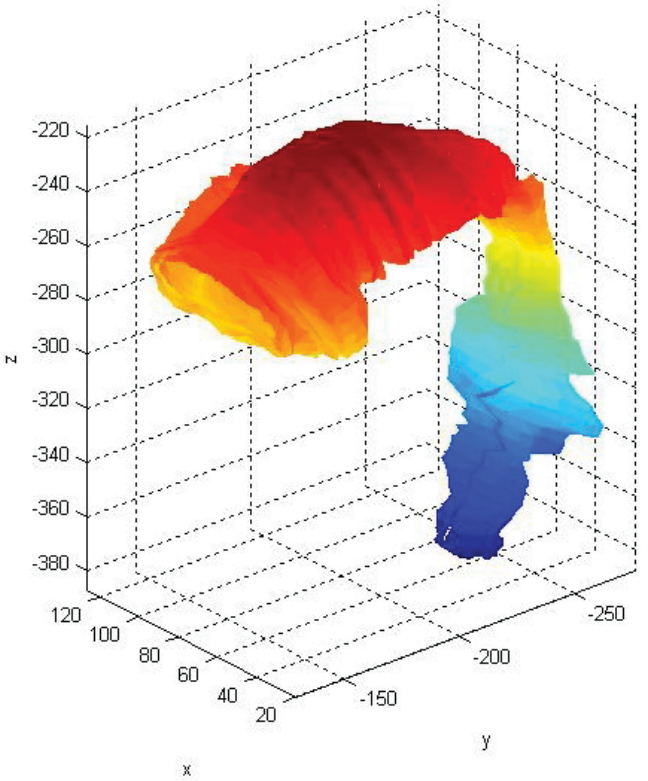
Two Samples by MRI



MRI sampling at ATR in Japan

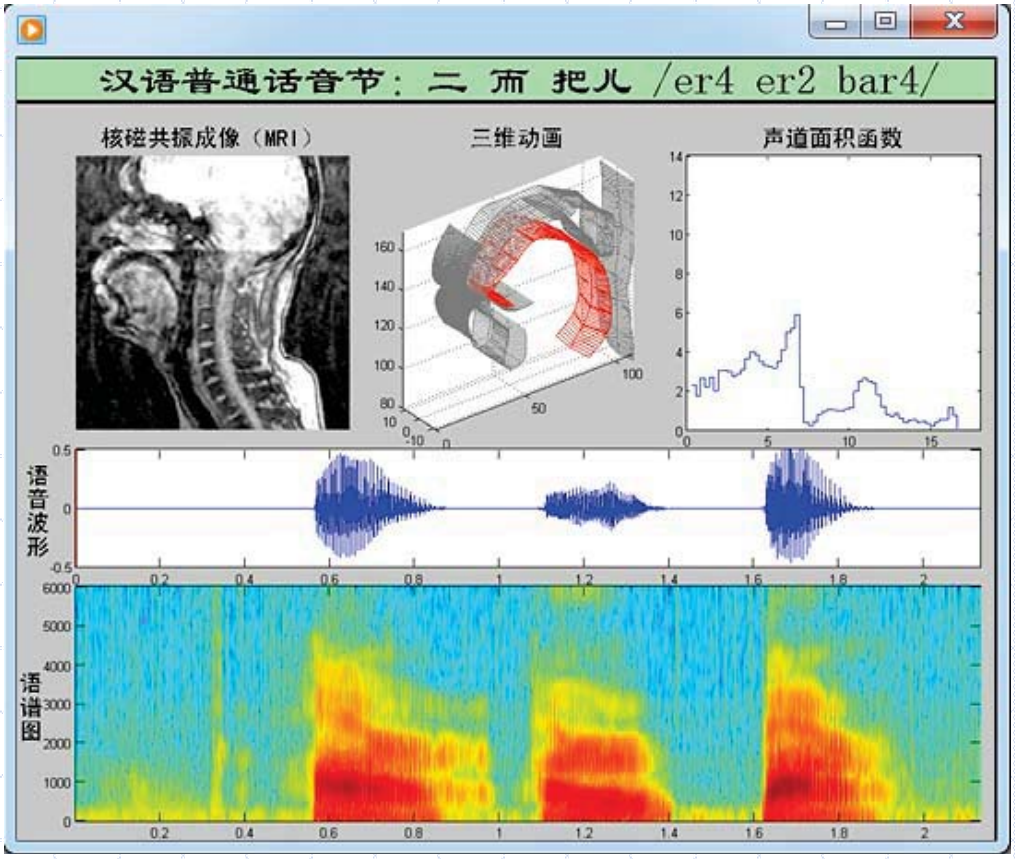
MRI 数据处理




-220
-240
-260
-280
-300
-320
-340
-360
-380
z
120
100
80
60
40
20
x
-250
-200
-150
y

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

Study on Mandarin by X-ray Video

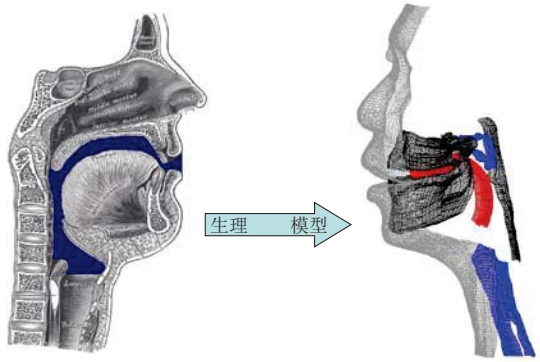


MRI_arerbari.avi

Model of Speech

1. Acoustical speech model
2. Geometric speech model
3. Physiological speech model
4. Neuro-speech model

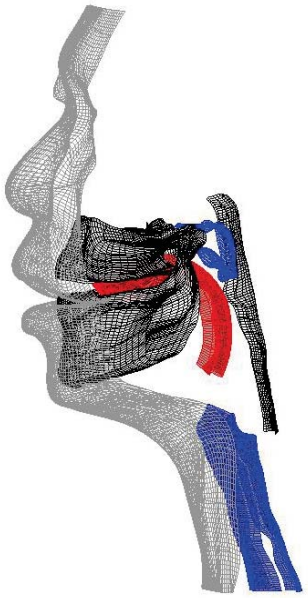
汉语普通话声道几何调音模型研究

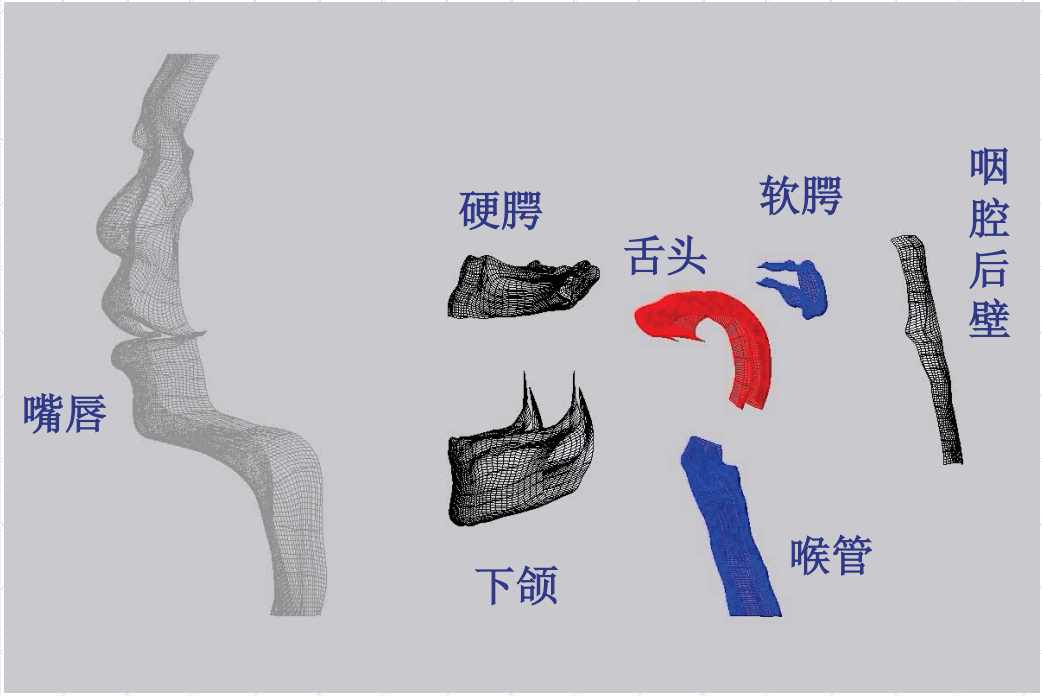


面积参数模型[Fant1960],[Flanagan1972],[Stevens1989],[Story1996,2005],……
几何调音模型[Coker1966],[Liljencrants1971],[Mermelstein1973],[Maeda1990],[Dang2002],……
生理调音模型[Perkell1974],[Payan1997],[Sanguineti1998],[Badin2002],[Dang2004],……

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

Future study





Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

A Study on Lip Model in Mandarin

北京大学中文系语言学语言学实验室
北京大学汉语语言学研究中心
Dept. of Chinese and Research Center of Chinese Linguistics
Peking University, China

1. A Study on 2D Lip Model in Mandarin

001.avi

002.avi

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

2. Definition of outer lips

Parameter Extracting

from video.

Liew's Model

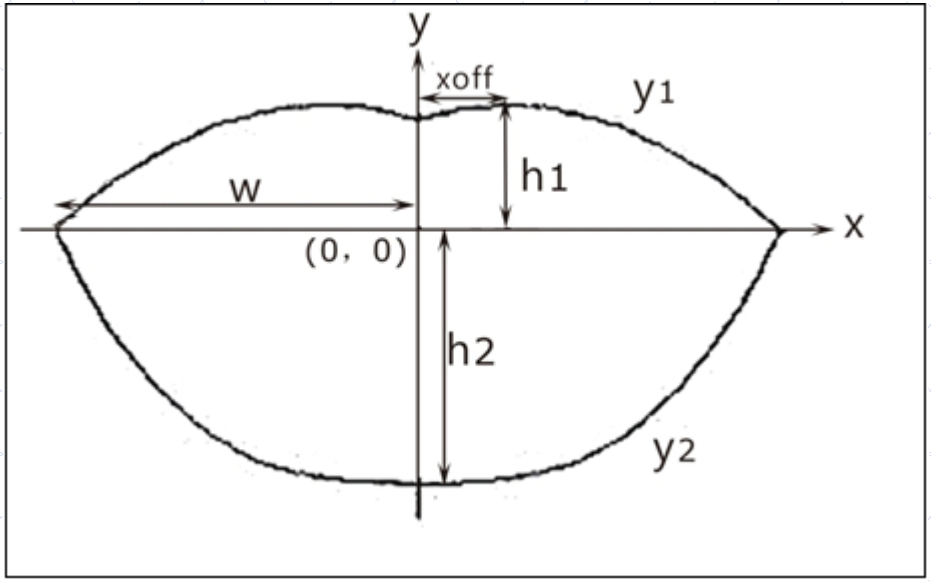
Para.:

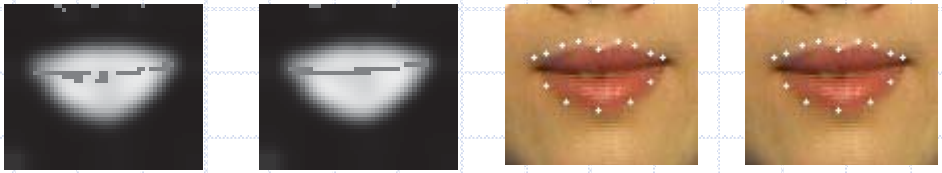
h1：height of upper lip

H2：height of lower lip

w：half width of lip

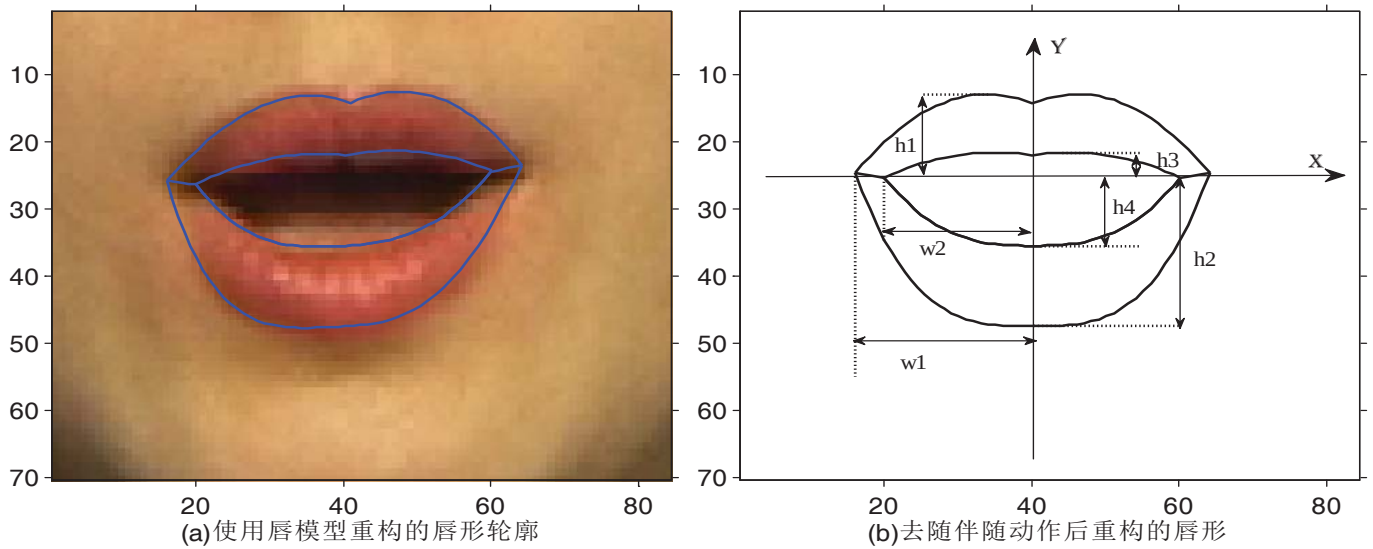
xoff：sunk part of upper lip





Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

3. Definition of inner and outer lips



(a)使用唇模型重构的唇形轮廓　(b)去随伴随动作后重构的唇形

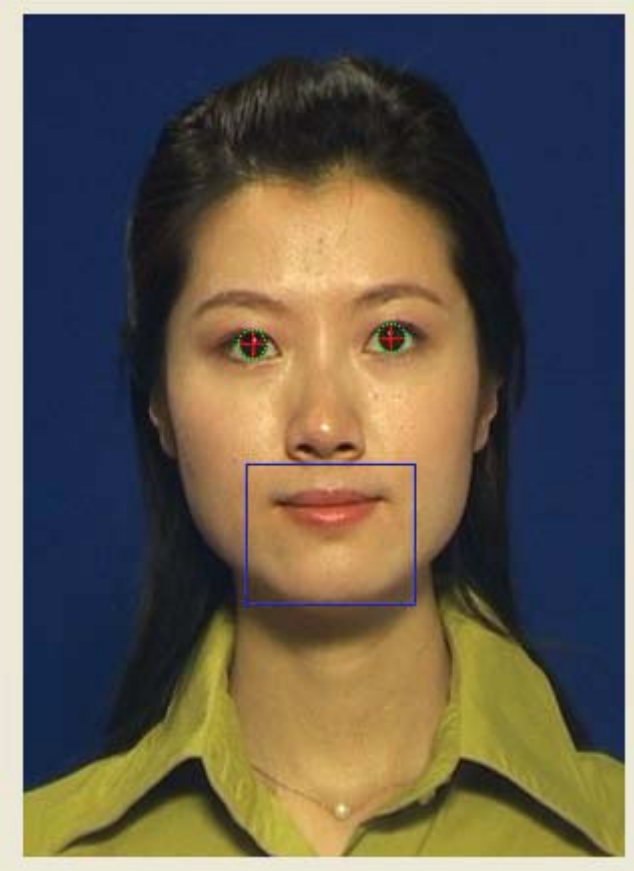
h1:外上唇高　h2:外下唇高　h3:内上唇高　h4:内下唇高

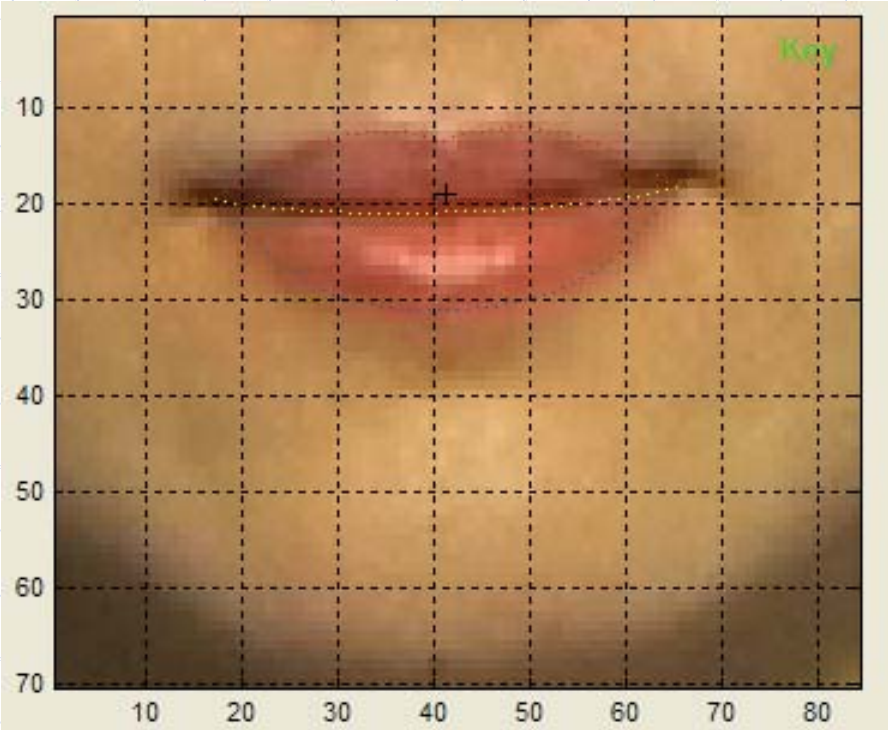
w1:外唇宽度　w2:内唇宽度　yh:下唇圆弧度　ax:人中凹陷度

q1:合口处曲率　qx:头部倾斜度　wz:歪嘴程度

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

4. 2D Lip Parameter Extracting

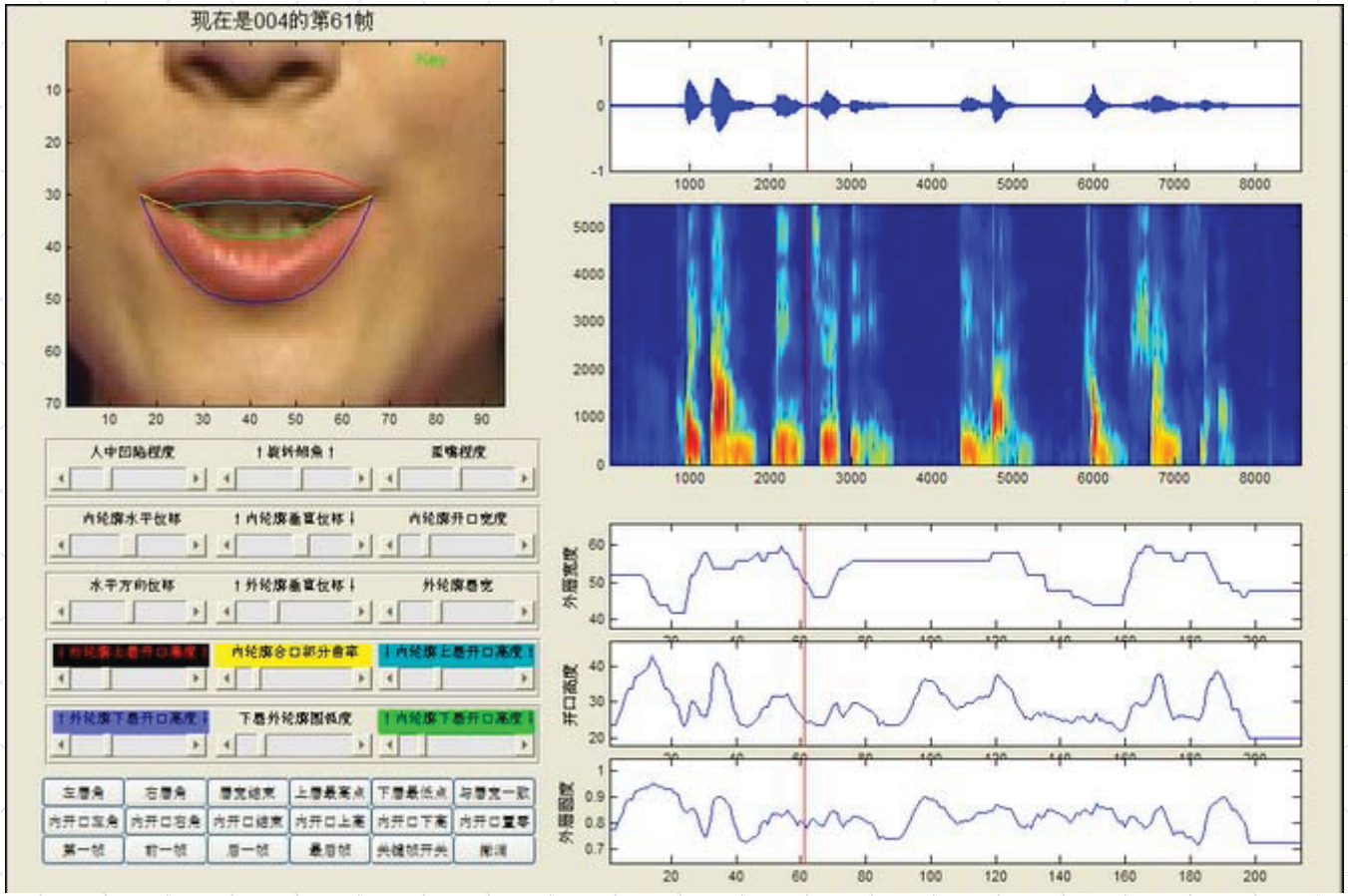




408 syllables are labeled.

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

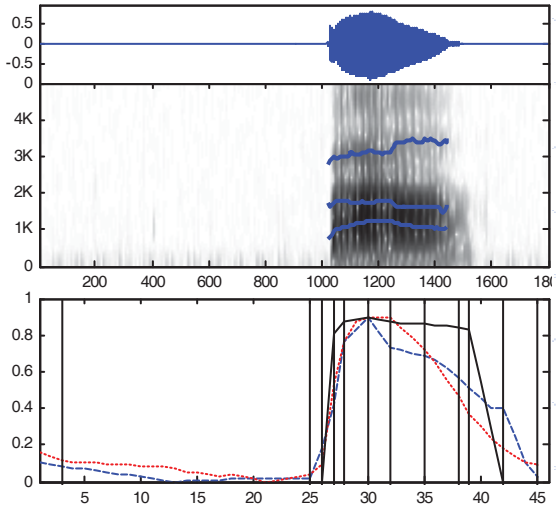
5. Analyzing Platform



wy.avi

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

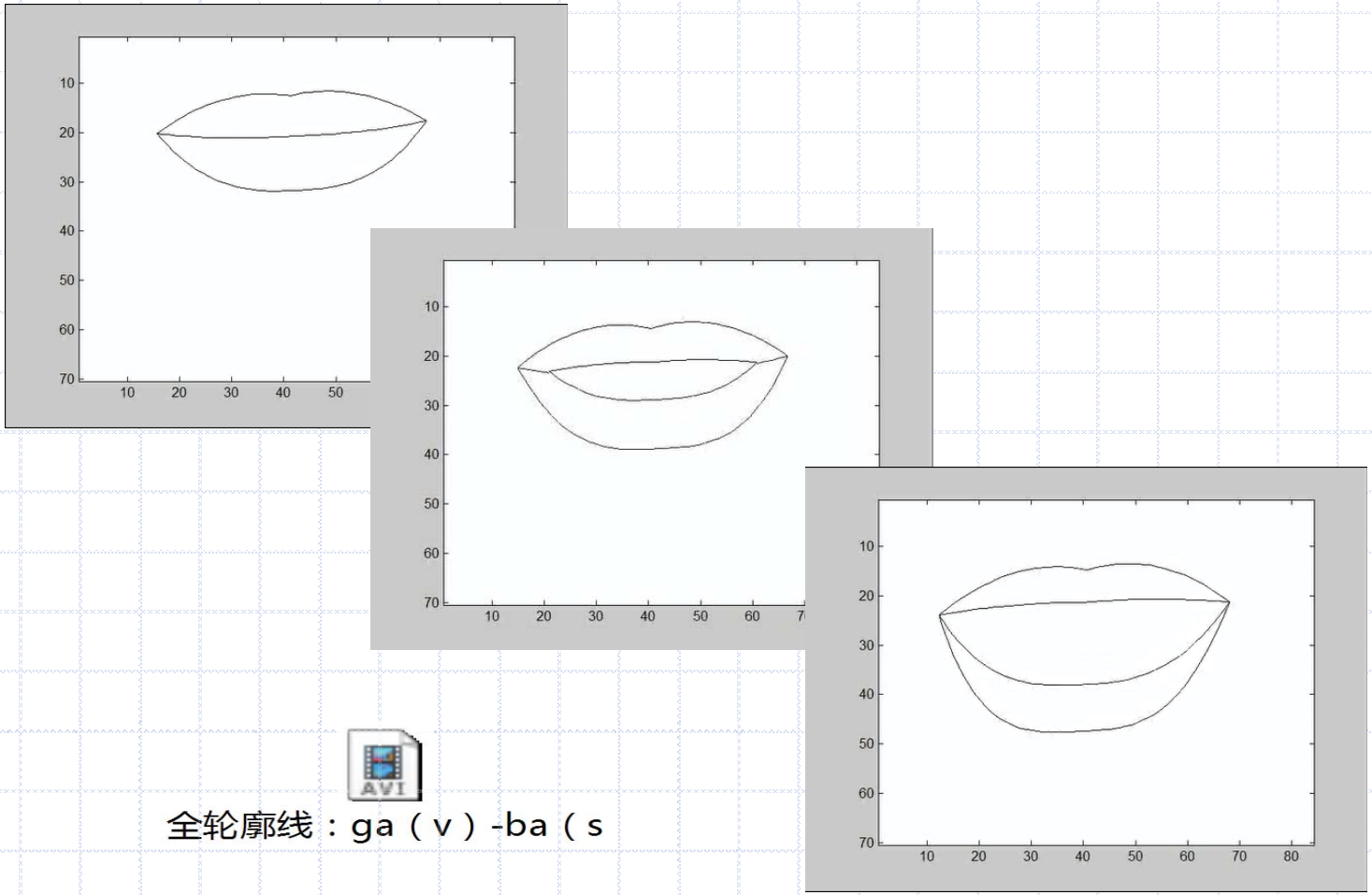
6. Analysis of one sample



	Fr3	Fr25	Fr26	Fr27	Fr28	Fr30	Fr32	Fr35	Fr38	Fr39	Fr42	Fr45
F1	-	-	-	870	1068	1213	1222	1049	-	-	-	-
F2	-	-	-	1675	1759	1733	1762	1605	-	-	-	-
F3	-	-	-	1675	1759	1733	1762	1605	-	-	-	-
P1	25	24	25	27	29	30	29	28	28	27	27	24
P2	18	15	17	31	39	43	43	37	29	26	20	17
P3	0	0	0	27	29	30	29	28	28	27	0	0

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

7. Synthesizing by a 2D model



全轮廓线：ga（v）-ba（s

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

8. McGurk Effect Study with 2D Model

da(v)-ba(s)

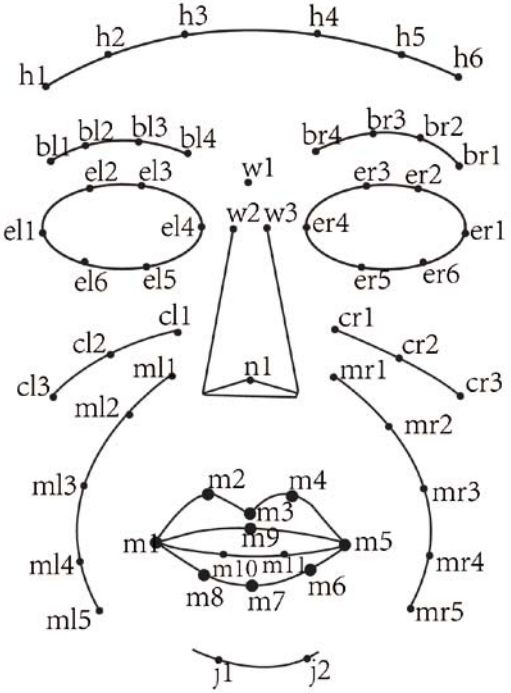
ga(v)-ba(s)

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

9. A 3D Study on Lip Model

Motion Capture

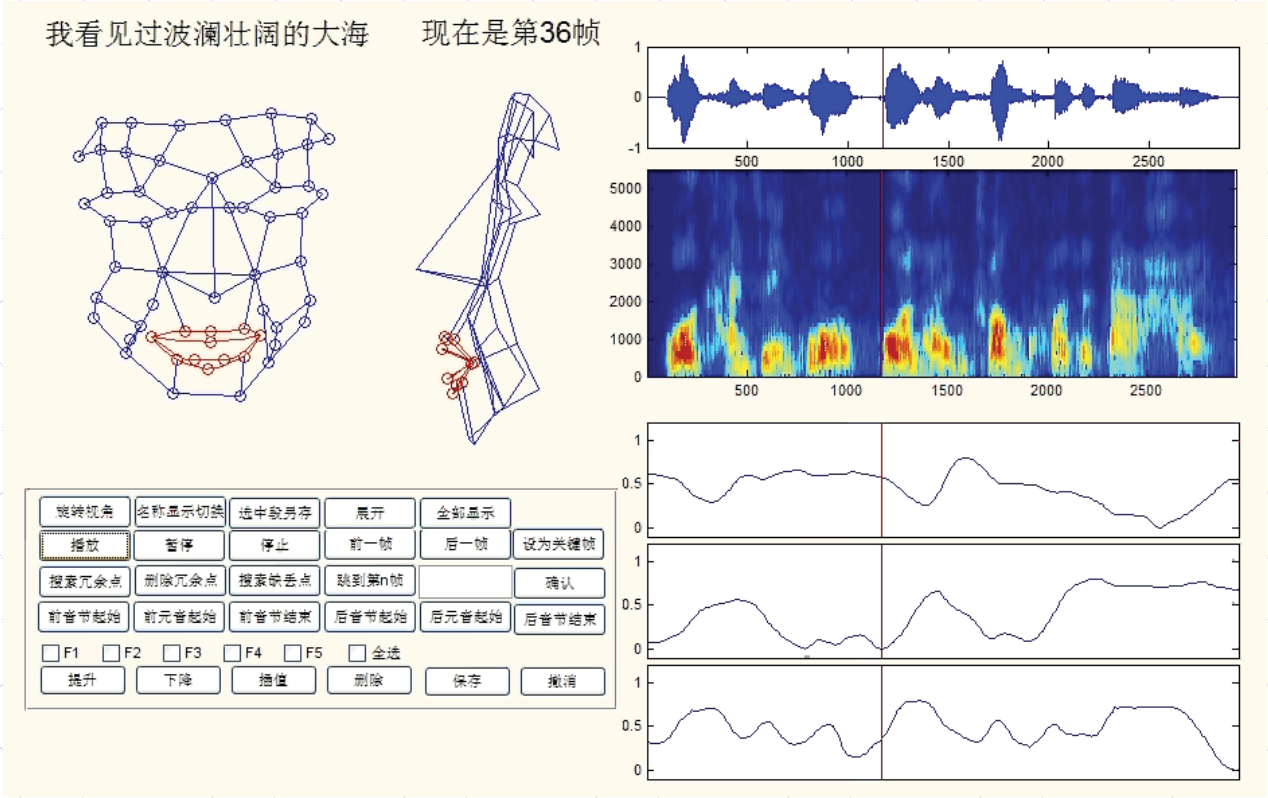
10. The Definition of Lip Sampling



ban.avi

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

11. The Definition of Lip Sampling



三维唇形_诗

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

A Study on Voice in Mandarin through High-speed Imaging

Dept. of Chinese Language and Literature
Center for Chinese Linguistics
A joint Lab for Language and Human Complexity
Peking University, P.R. China

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

A Voice Study in Ancient China

A story of artificial larynx

in ancient Chinese documents

Devices made by people from materials, such as bamboo, wood, ivory and bone, were called sound generators. A sound generator, which could be put into the throat and produce speech sound by whistling, was called a voice generator. A dumb person, who suffered from injustice, could not argue in court for himself. The judge let people put a voice generator in his throat, and asked him to speak. The speech articulated like puppet talk, but could roughly make sense. His injustice was finally redressed. The case is worthy of being documented.

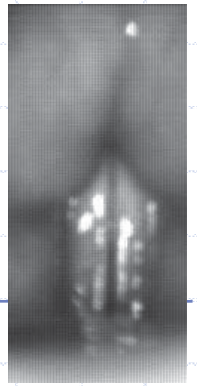
From an ancient Chinese book titled "Mengxi Bi Tan" (writing in the Mengxi Garden) by Shen Kuo (1031-1095). (translated by Kong Jiangping.)


夢溪筆談卷十三 一
世人以竹木牙骨之類爲叫子置人喉中吹之能作人言謂之類叫子嘗有病瘖者爲人所苦煩冤無以自言聽訟者試取叫子令類之作聲如傀儡子粗能辨其一二其冤獲伸此亦可記也

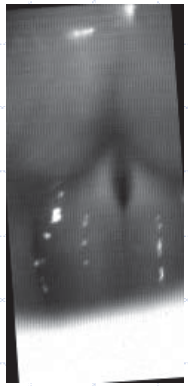
Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

1. A Study on Dynamic Glottis by HSI

In thisstudy，the system with endoscope is used for capturing samples. The sampling rate of image is 2000 frames/s and the size of image is 120×256 pixels in gray scale. The system is produced by Kay.



Tone 1



Tone 2

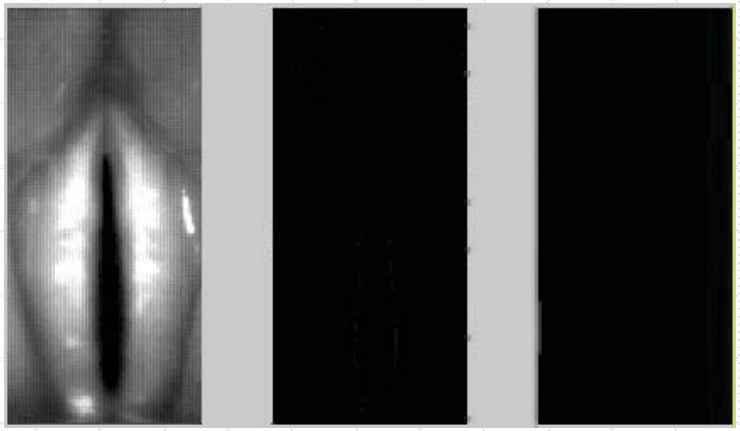
Tone 3

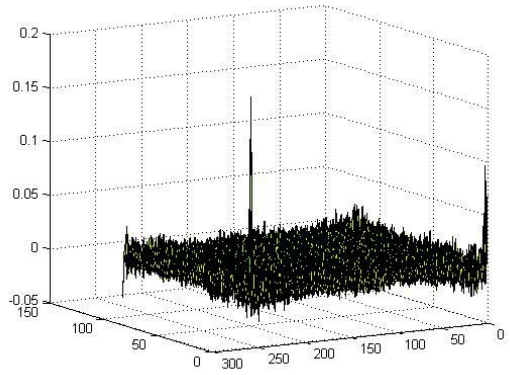
Tone 4

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

Movie Compensation

One of the adverse factors affecting the accuracy and validity of high-speed video (HSV) quantitative assessment is the motion of the endoscope's lens relative to the larynx. Endoscopic motion makes it difficult to track the dynamic characteristics of the laryngeal anatomic structures. The method of FFT-based cross-power spectrum published by D. Deliyski (2005) are mainly used.



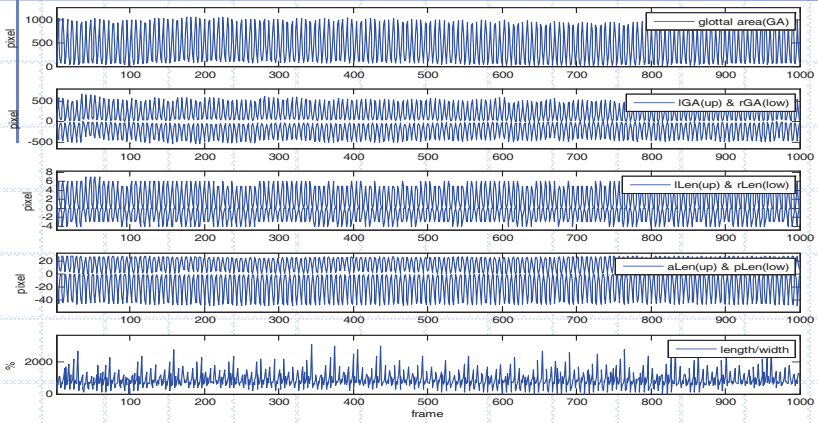


The videos show the computation of the time differentials of HSV sequence. The window includes 100 frames of images and the computing step is one frame. The figure shows the peak for cross-power spectrum similarity. The location of the peak reflects the shift between the 2 images.

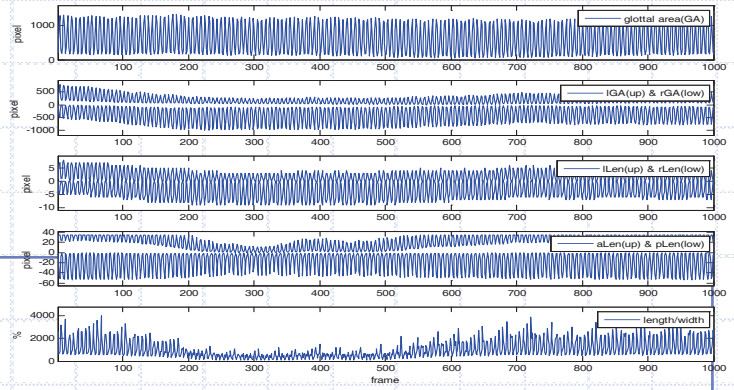
Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

Movie Compensation

The following figures show the result of MC. One is the parameters before the MC and the other shows the parameters after MC.



The glottal parameters after MC.



The glottal parameters before MC.

From up to bottom: 1) the glottal area (GA); 2) the left and right GA; 3) the left and right glottal width; 4. the anterior and posterior glottal length; 5) the ratio of glottal length to glottal width.

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

Contrast Adjusting

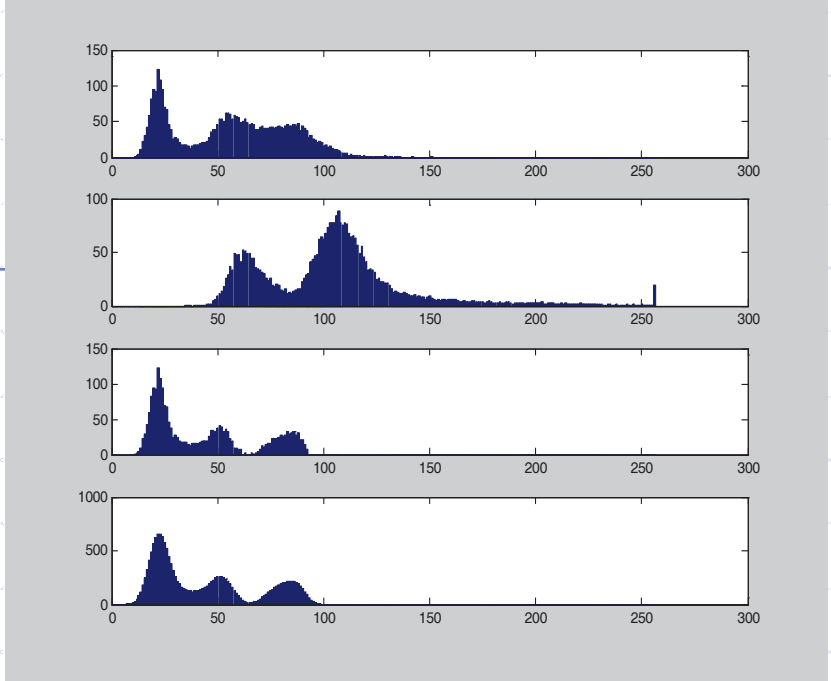
The method of contrast adjusting:

1. The accumulated histogram when the glottis is fully opened in the first 100 frames;

2. The accumulated histogram when the glottis is fully closed in the first 100 frames;

3. The subtraction of them gives us the histogram and the peak in the low gray region mainly reflects the gray value of the glottal region.

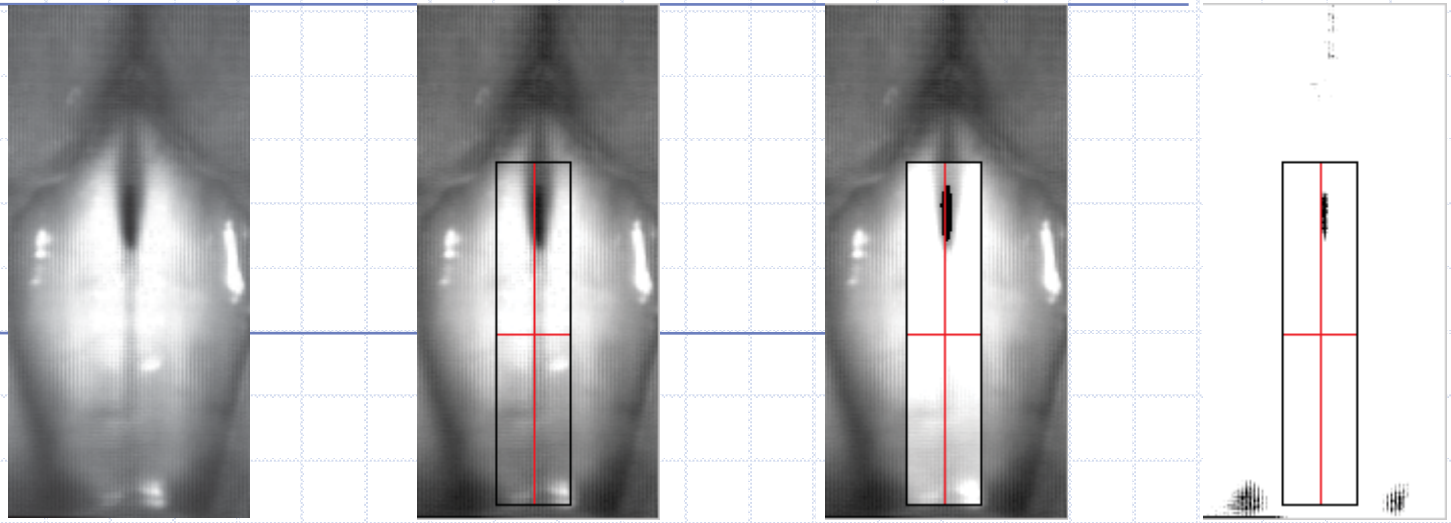
4. Smooth this histogram, and use the gray values at the left and right sides of the first peak to automatically adjust the contrast.



The ‘x’ axis is the value of gray scale (0 to 255) and the ‘y’ axis is the sum of value with the same gray scale.

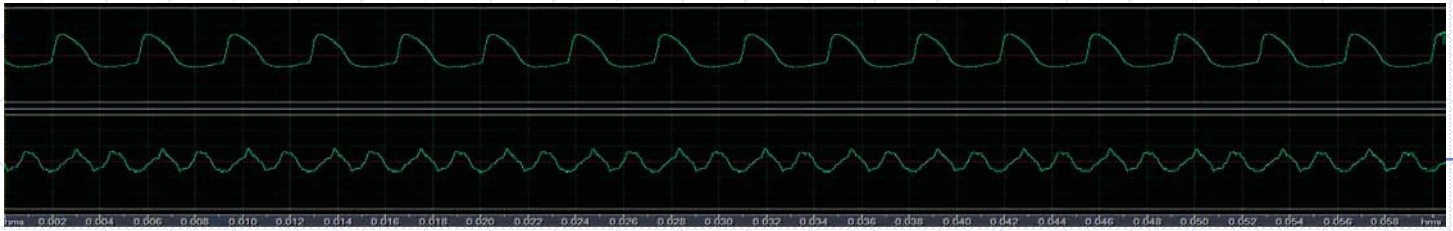
Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

Contrast Adjusting



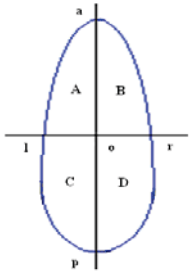
a b c d

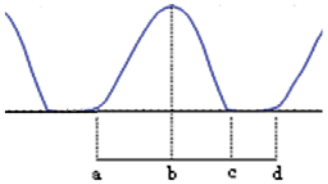
'a' is an original video; 'b' is a video with a window; 'c' is a video with the detected glottis; 'd' is a video with glottal area.



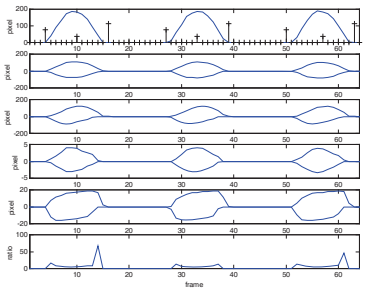
The figure show the EGG signal (up) and speech signal (bottom).

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University





The figure shows the basic definition of glottis and the F0, OQ and SQ.



The definitions:

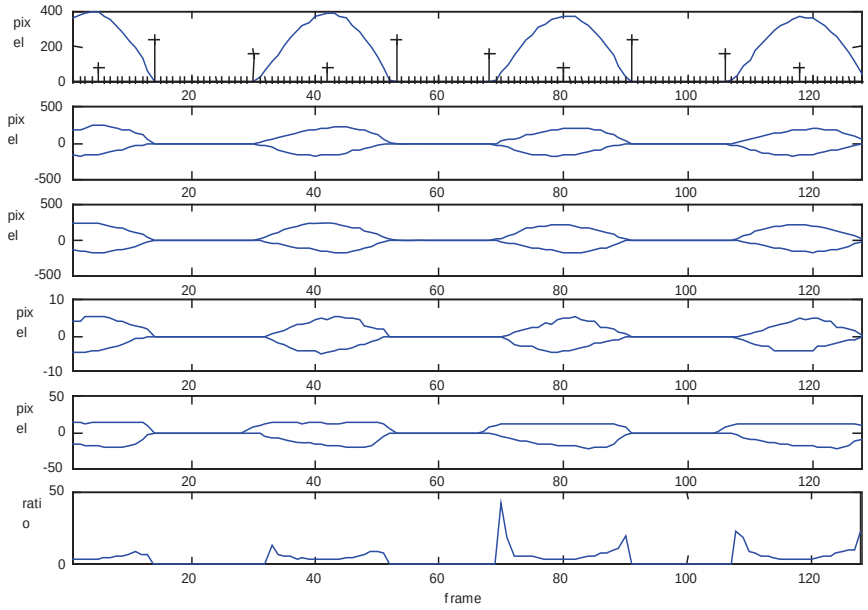
1)The fundamental frequency is defined as 1/'ad' (Hz).

2)The open quotient of glottal period is defined as the ratio of 'ac' over 'ad'.

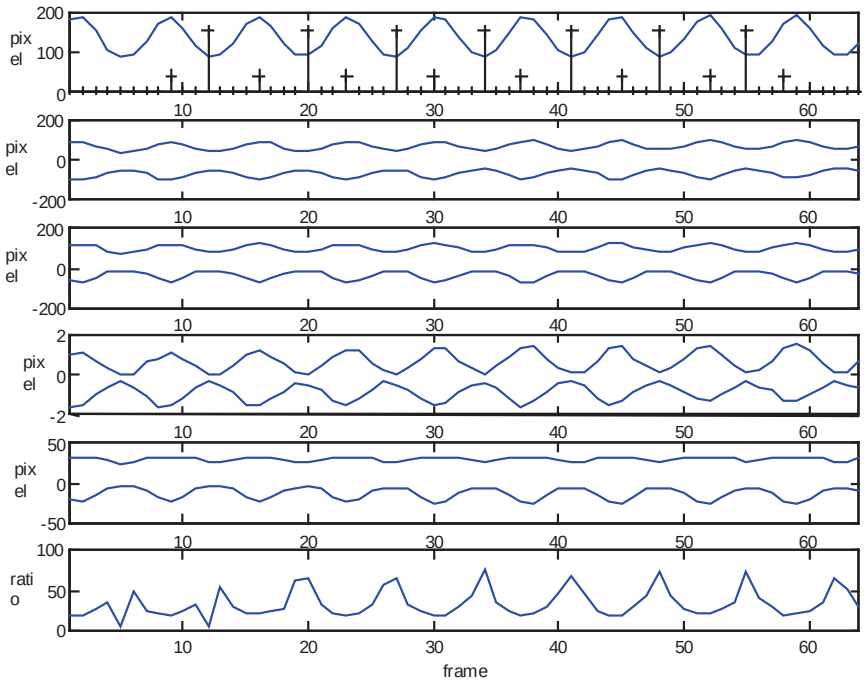
3)The speed quotient of glottal period is defined as the ratio of 'ab' over 'bc'.

1: Left width of glottal area
2: Right width of glottal area
3: Anterior length of glottal area
4: Posterior length of glottal area
5: Left glottal area
6: Right glottal area
7: Anterior glottal area
8: Posterior glottal area
9: Ratio of length\width
10: Glottal area
11: F0 of glottal area function
12: OQ of glottal area function
13: SQ of glottal area function

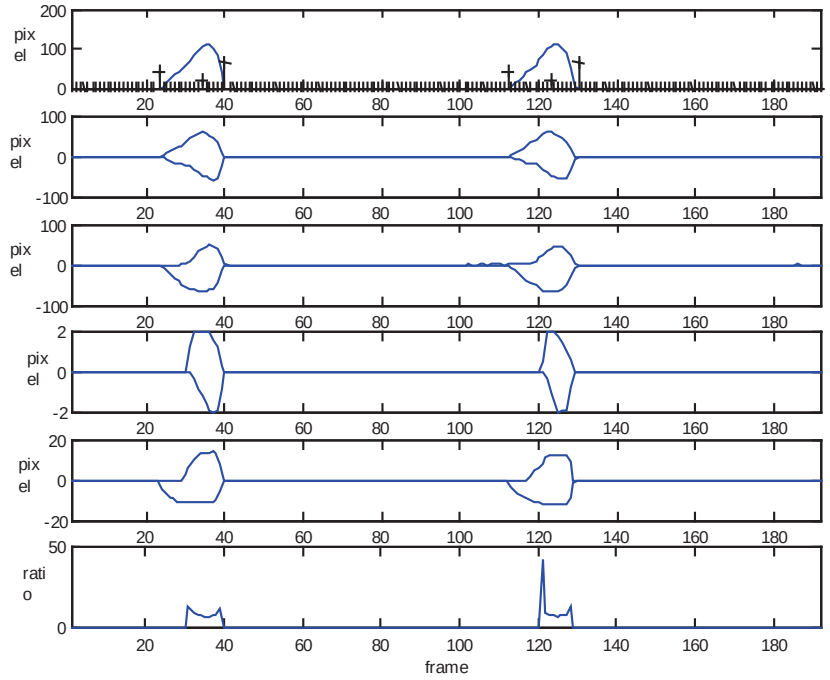
正常嗓音 (modal voice)



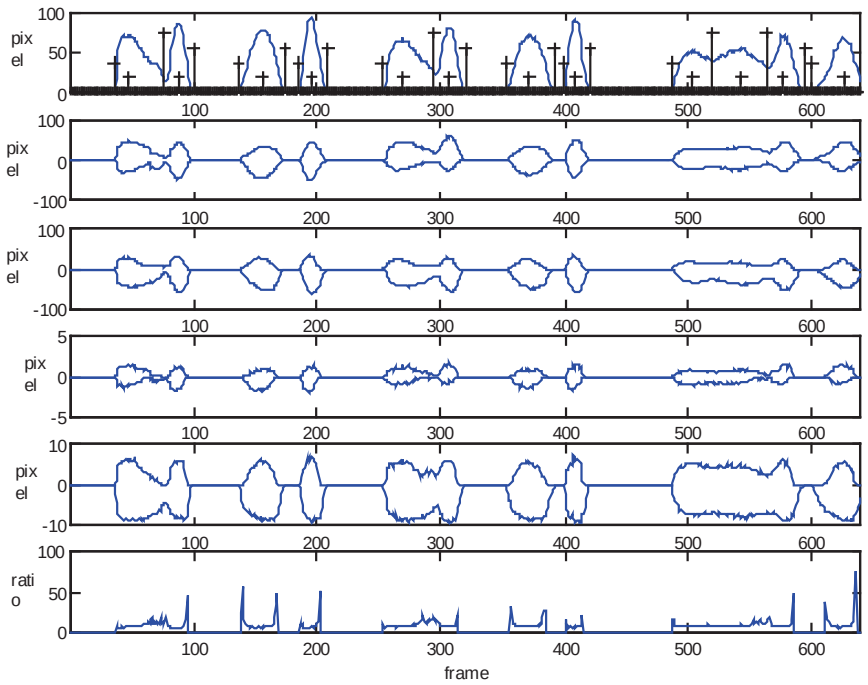
假声 (falsetto)


pix el
200
100
0
10 20 30 40 50 60
pix el
200
0
-200
10 20 30 40 50 60
pix el
200
0
-200
10 20 30 40 50 60
pix el
2
0
-2
10 20 30 40 50 60
pix el
50
0
-50
10 20 30 40 50 60
rati o
100
50
0
10 20 30 40 50 60
frame

气泡音（typical vocal fry）

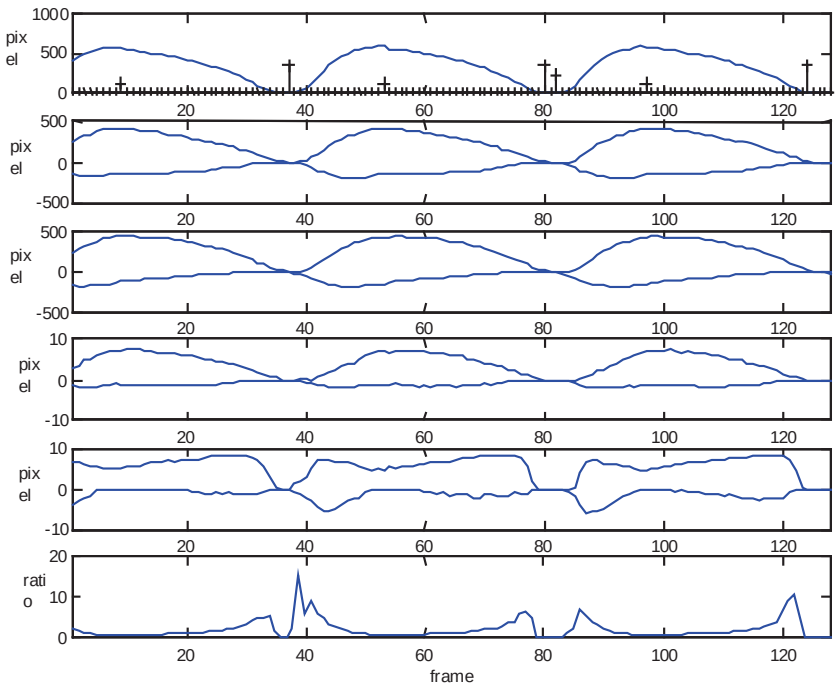


特殊气泡音（special vocal fry）

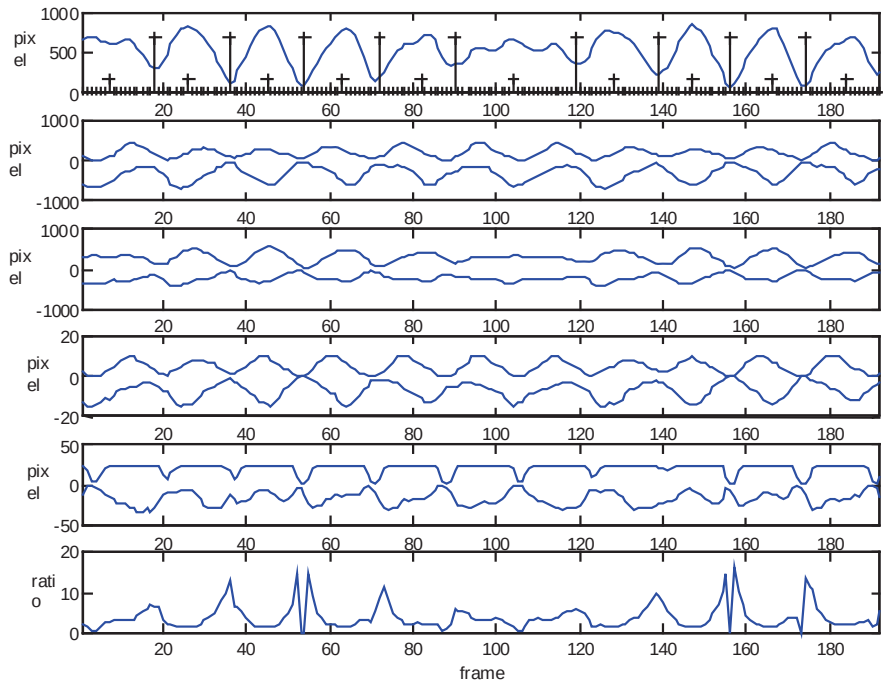


Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

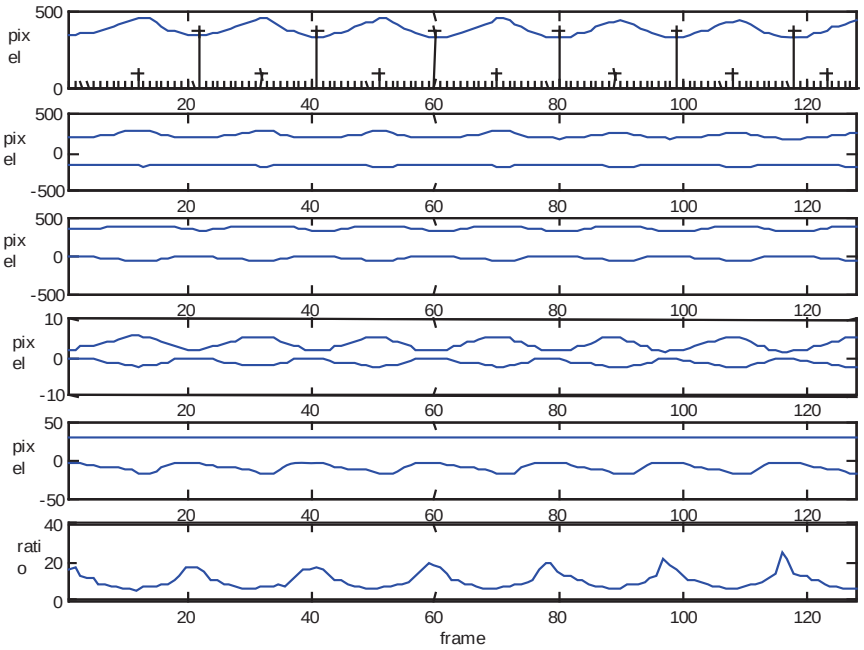
气嘎音 (breathy voice)



双音调噪音（diplophonia）



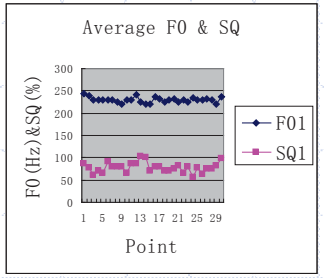
吸气音（inspiration）

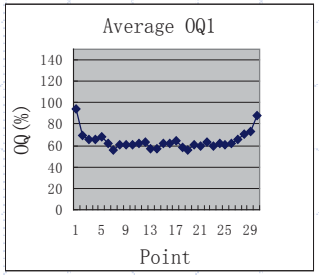


Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

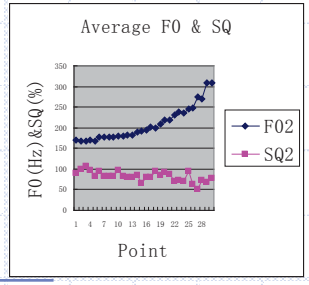
Properties in Tones

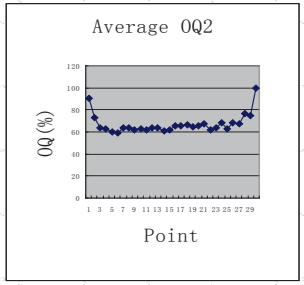
	F0	SQ	OQ
Tone 1	H	Level	level
Tone 2	R	F	level
Tone 3	FR	RF	level
Tone 4	F	F	level



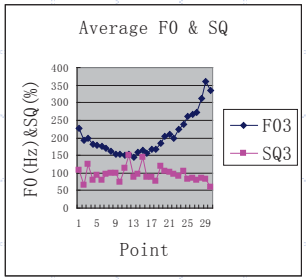


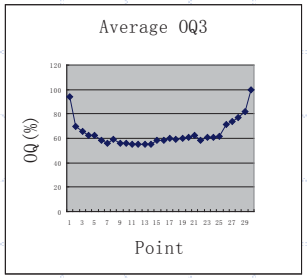
This figure shows the phonation parameters of F0,OQ and SQ of tone 1.



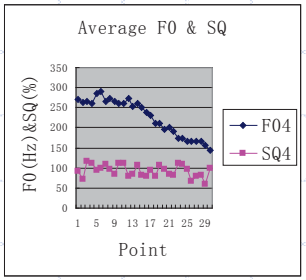


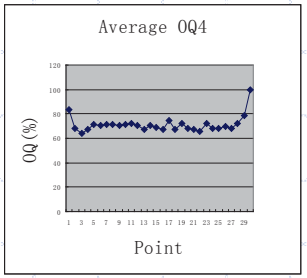
This figure shows the phonation parameters of F0,OQ and SQ of tone 2.





This figure shows the phonation parameters of F0,OQ and SQ of tone 3.

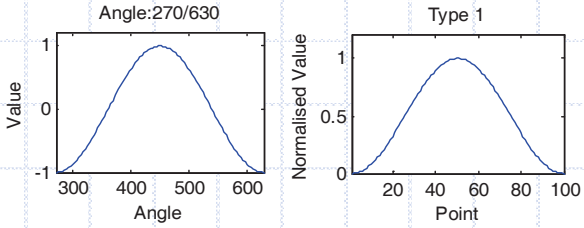




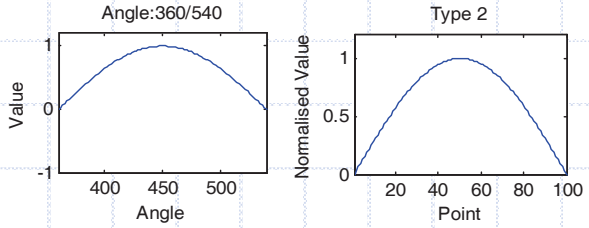
This figure shows the phonation parameters of F0,OQ and SQ of tone 4.

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

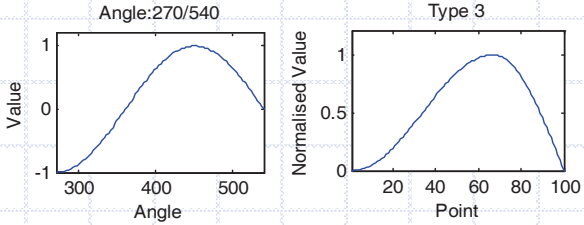
Modeling on Glottis



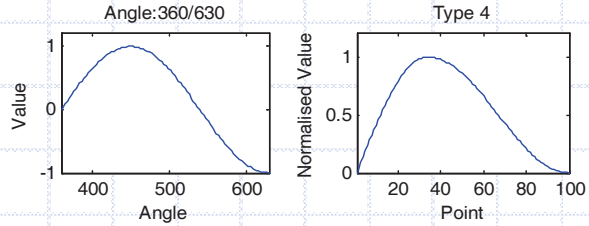
The left image shows a pulse obtained from 270 to 270 of sinusoid and the right image is shows the parameters of the left image which is normalized from 0 to 1.



The left image shows a pulse obtained from 360 to 180 of sinusoid and the right image is shows the parameters of the left image which is normalized from 0 to 1.



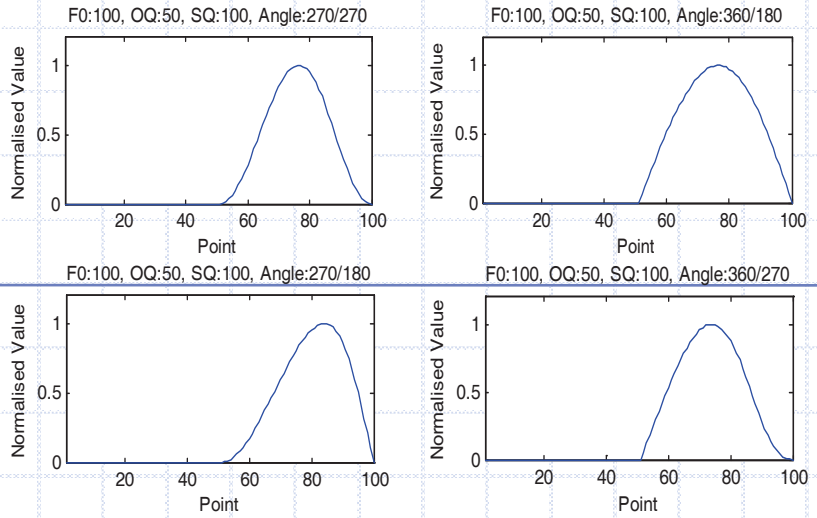
The left image shows a pulse obtained from 270 to 180 of sinusoid and the right image is shows the parameters of the left image which is normalized from 0 to 1.



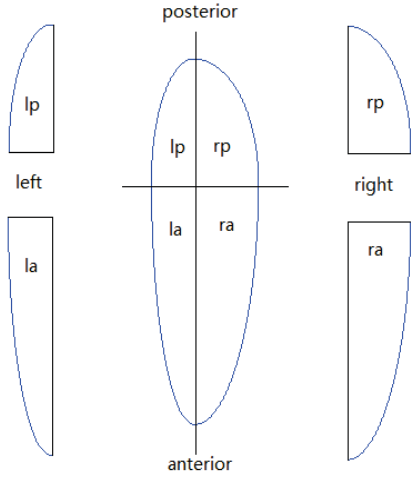
The left image shows a pulse obtained from 270 to 180 of sinusoid and the right image is shows the parameters of the left image which is normalized from 0 to 1.

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

Modeling on Glottis



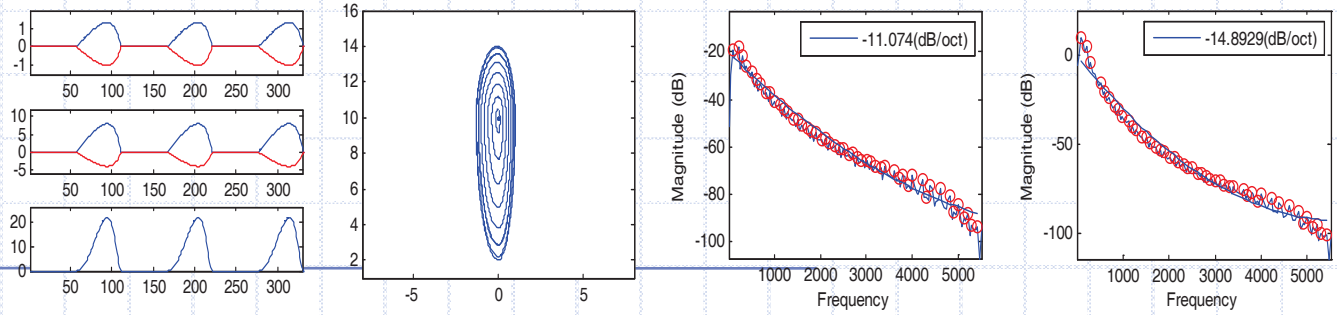
There are four plots which show the four typical types of dynamic glottal function.



The static glottis was modeled by four quarters of ellipses.

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

Synthesized tones



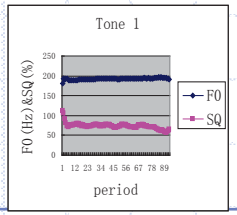
	F0	OQ	SQ	Angle 1	Angle 2	width/length
Left	100	50	300	360	180	1.3
Right	100	50	300	360	180	1
Anterior	100	50	300	360	180	8
Posterior	100	50	300	360	180	4

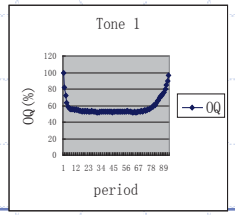
a211 i35

Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

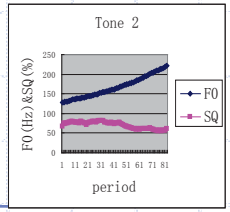
Synthesized tones

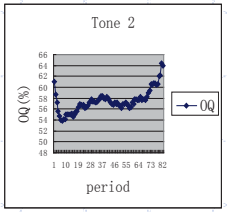
Tone 1



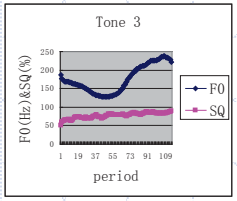


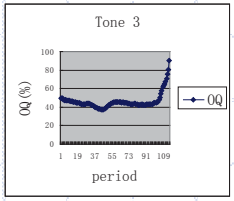
Tone 2



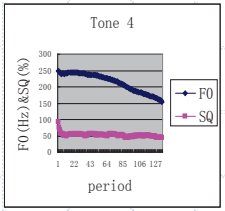


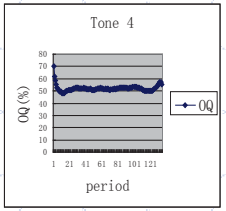
Tone 3





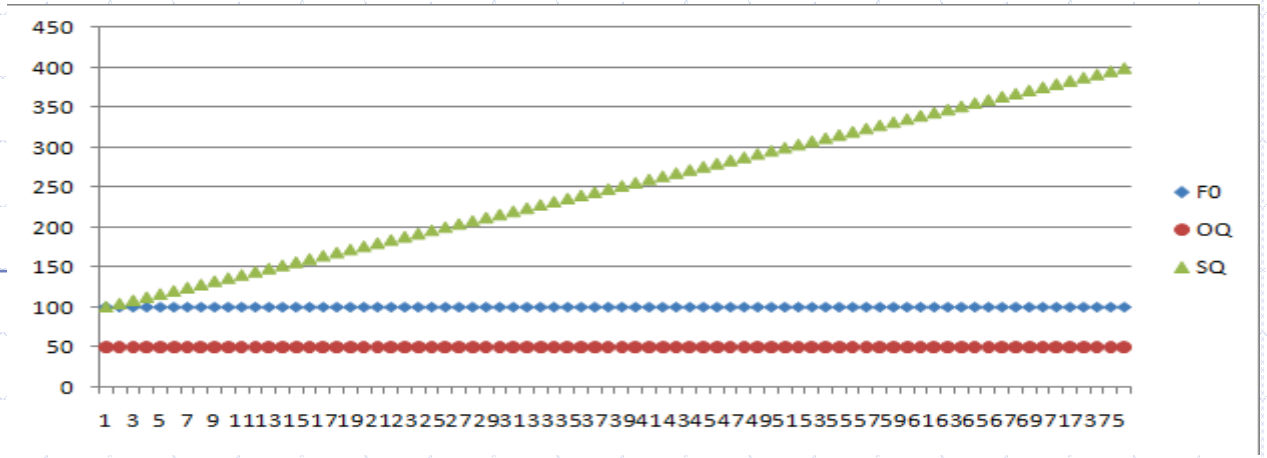
Tone 4





Linguistics Lab
Dept. of Chinese Language and Literature
Peking University

Sample with gliding SQ

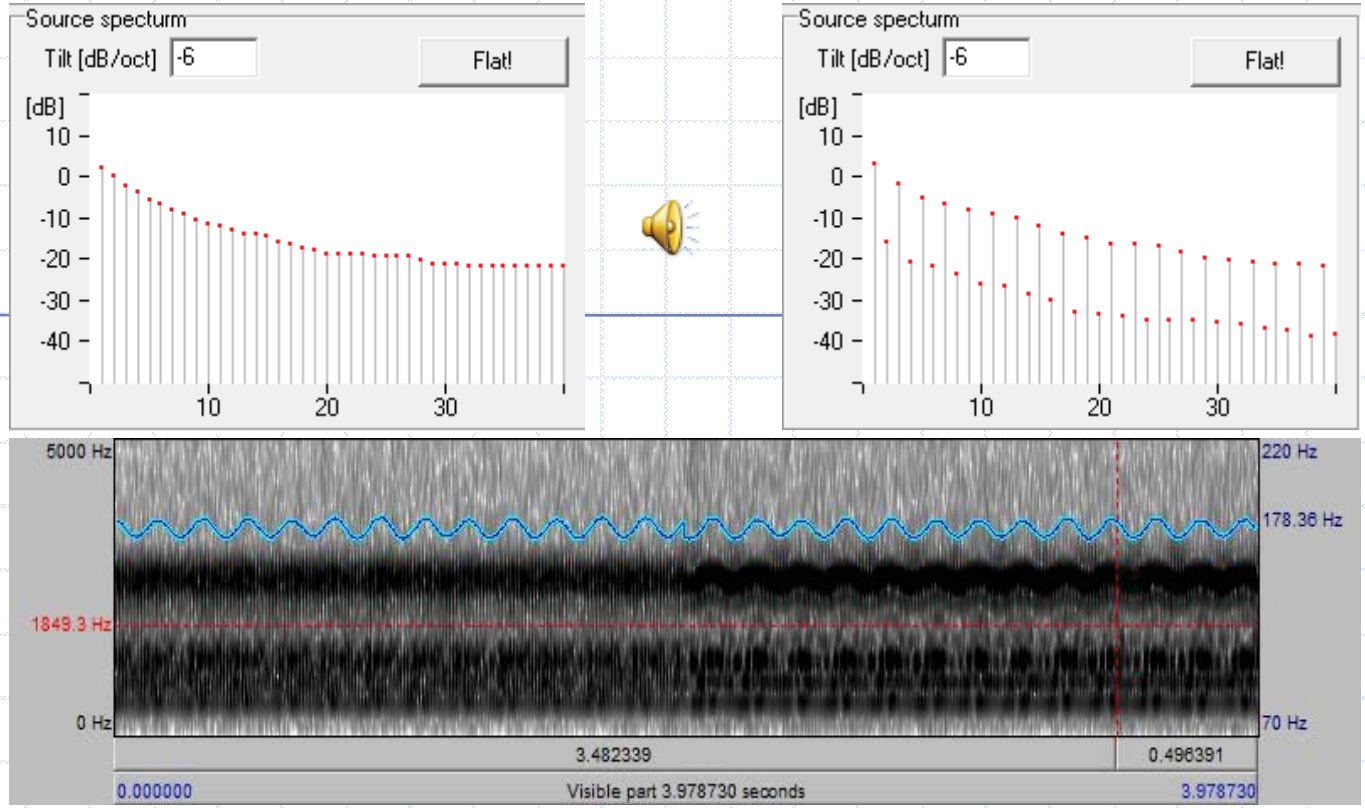


F0=100 Hz

OQ=50 %

SQ= 100 to 400%

Sample with different harmonics



Vibrato；harmonics；sub-harmonics；spectrum tilt

Thank you for your attention

谢　　谢!

jpkong@pku.edu.cn