A Multi-modal Interlanguage Speech Corpus of Chinese as a Second Language

Jinsong Zhang
SAIT lab
Beijing Language and Culture University
Jinsong.zhang@blcu.edu.cn

梧桐创新

语言资源高精尖创新中心
Beijing Advanced Innovation Center for Language Resources

Outline

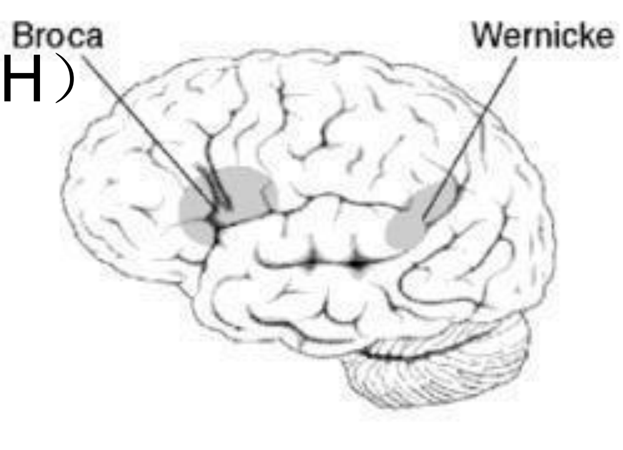
- □ The purpose of the database
- □ Feature descriptions
- □ Current status
- □ Conclusion

Background

- □ Pronunciation teaching is the 1st step in 2nd language learning.
- □ The results are unsatisfactory.
  - ■ Ex. 1: English spoken by Chinese - “Chinglish”
  - ■ Ex.2: Chinese spoken by foreigner- “洋腔洋调”
- □ Many explanations.

Reasons for Difficulties in 2nd Language Learning

- Critical Period Hypothesis（CPH）
  - CAH by R. Lado
  - SLM by J. Flege
  - PAM by C. Best
  - Etc.

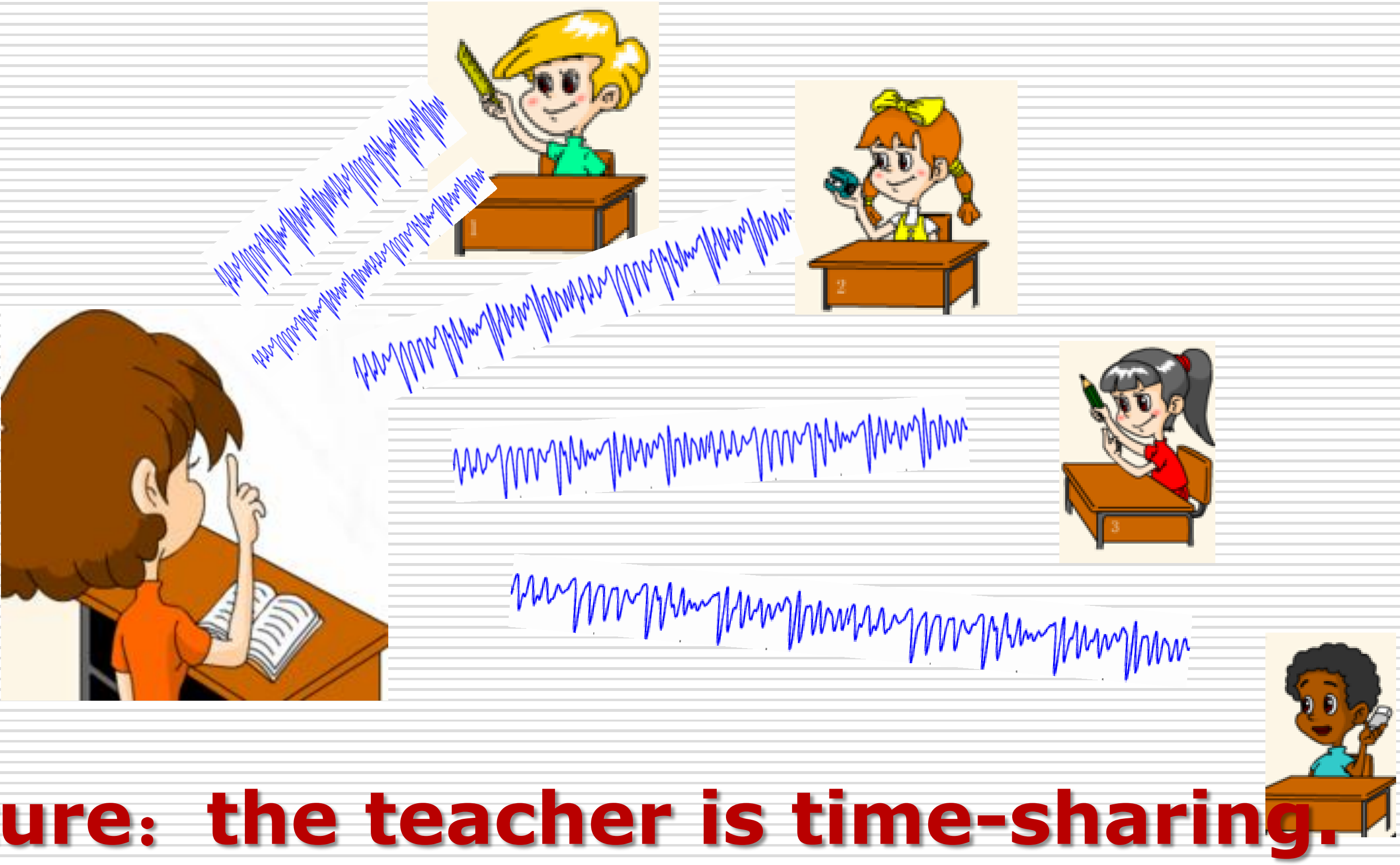


- Theories of 2nd Language Learning
  - UG-Language Acquisition Device by Chomsky
  - Input hypothesis by S. Krashen（explicit, implicit ）
  - Skill Acquisition theory by R.Dekeyser
  - Etc.

Optimal Requirements

- Large amounts of meaningful practices
- Instant and pertinent feedbacks

Language Teaching in Classrooms



Feature: the teacher is time-sharing.

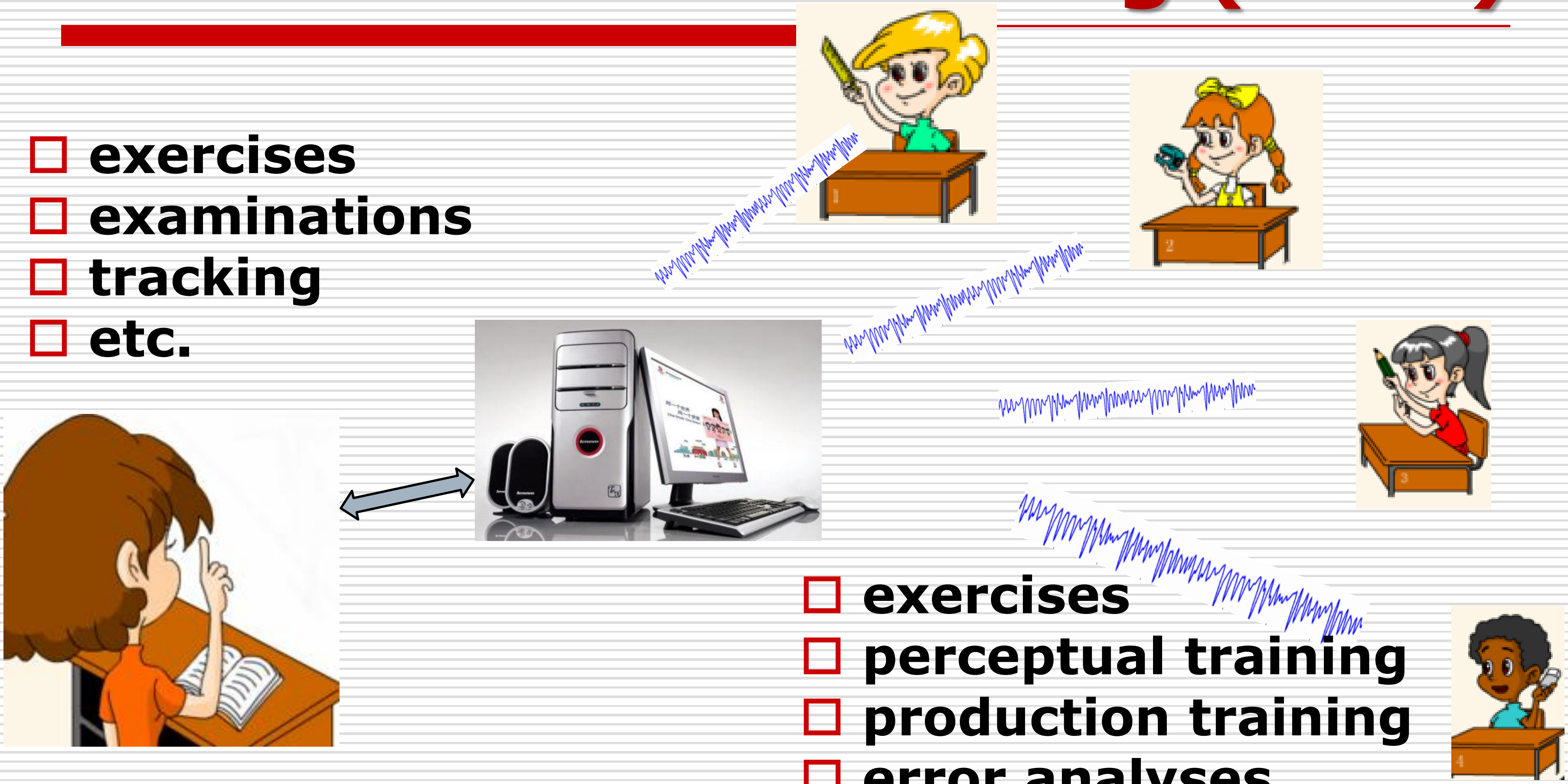
Limitations of Classroom Teaching

- □ Time-sharing classroom cannot fulfill the requirements！
- □ Solution：

Time-sharing teaching

Intelligent Technology For Pronunciation Teaching/Training

Intelligent Technology for Pronunciation Teaching (ITPT)



- □ exercises
- □ examinations
- □ tracking
- □ etc.

- □ exercises
- □ perceptual training
- □ production training
- □ error analyses
- □ etc.

Characteristics of ITPT

- Convenient for unlimited practicing；
- Individual training courses；
- Easy for teachers to know the pupils；
- Continuously keep tracking；
- Perceptual examination；
- Improve the objectiveness of examinations；
- Etc.

Outline

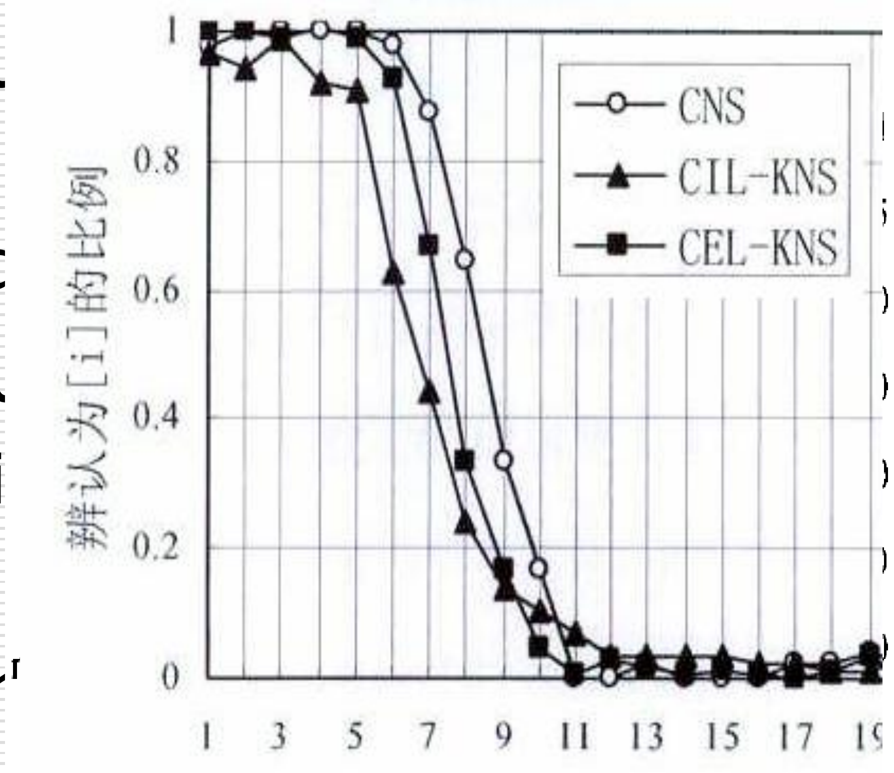
- □ The purpose of the database
- □ Feature descriptions
- □ Current status
- □ Conclusion

What to do at BLCU-SAIT?

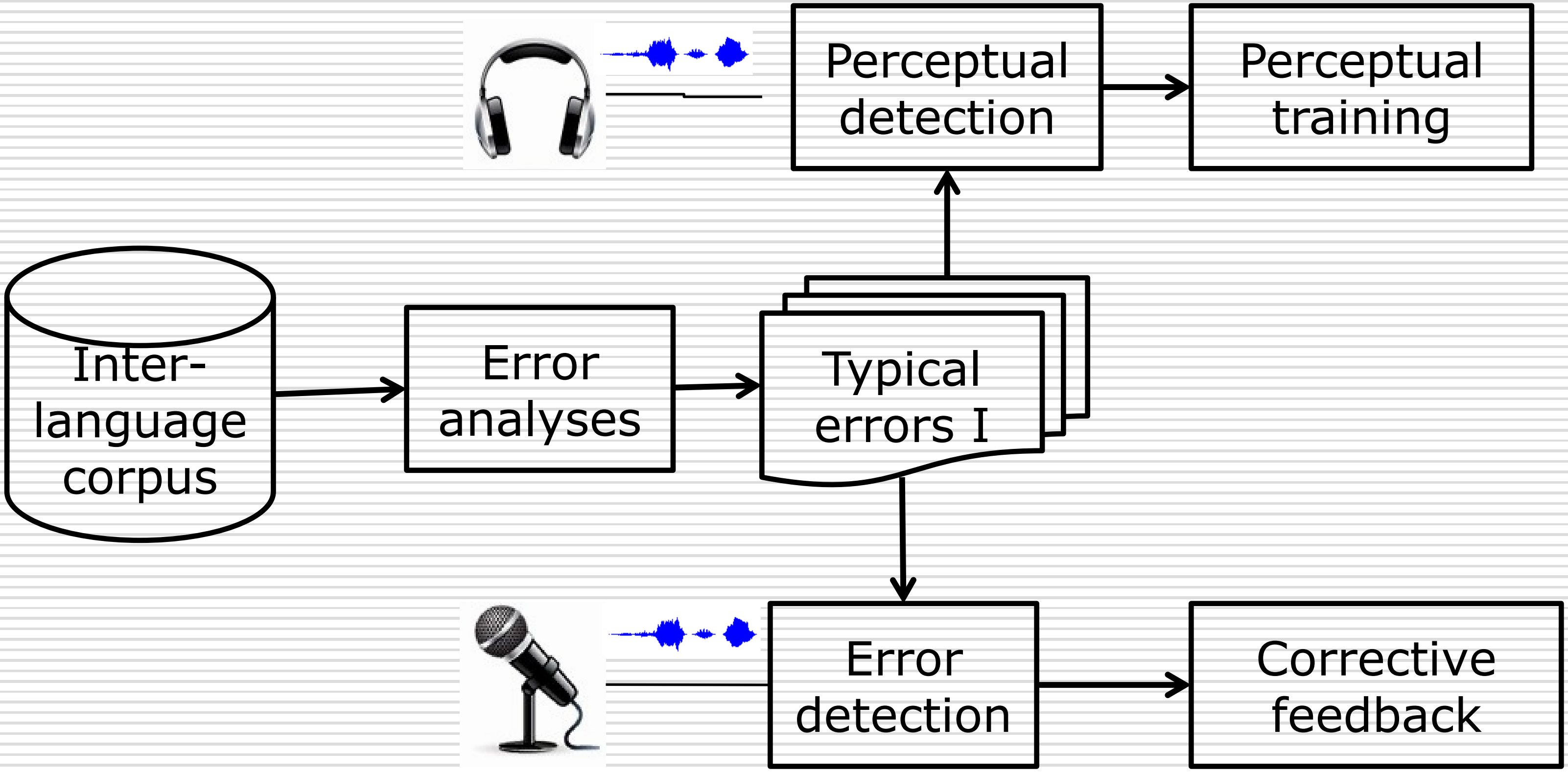
- □ Speech Acquisition and Intelligent Technology Lab (SAIT).
- □ The key problem dealt: **non-native accentedness** in Chinese spoken by 2nd language learners.
  - Perceptual ambiguity.
  - Production ambiguity.
  - Phonetic redundancy in speech communication.

Objective Reasons for Non-native accentedness

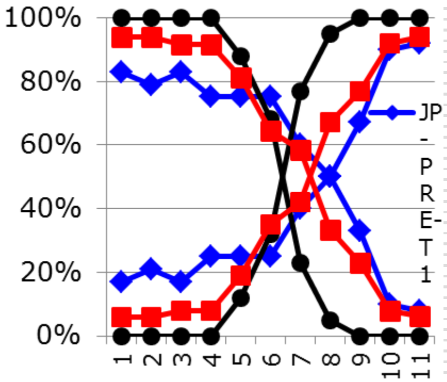
- □ Perceptual ambigu
  - Phonetic category
  - Perceptual depend
- □ Production ambigu
  - Acoustic overlappi
  - Negative transfer
- □ Communication ar
  - Silent feedback by native speakers.

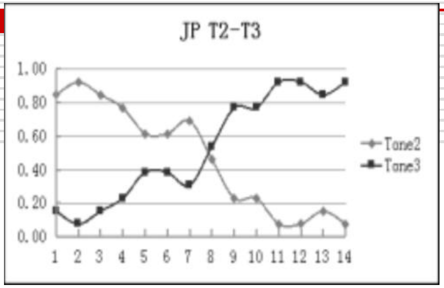


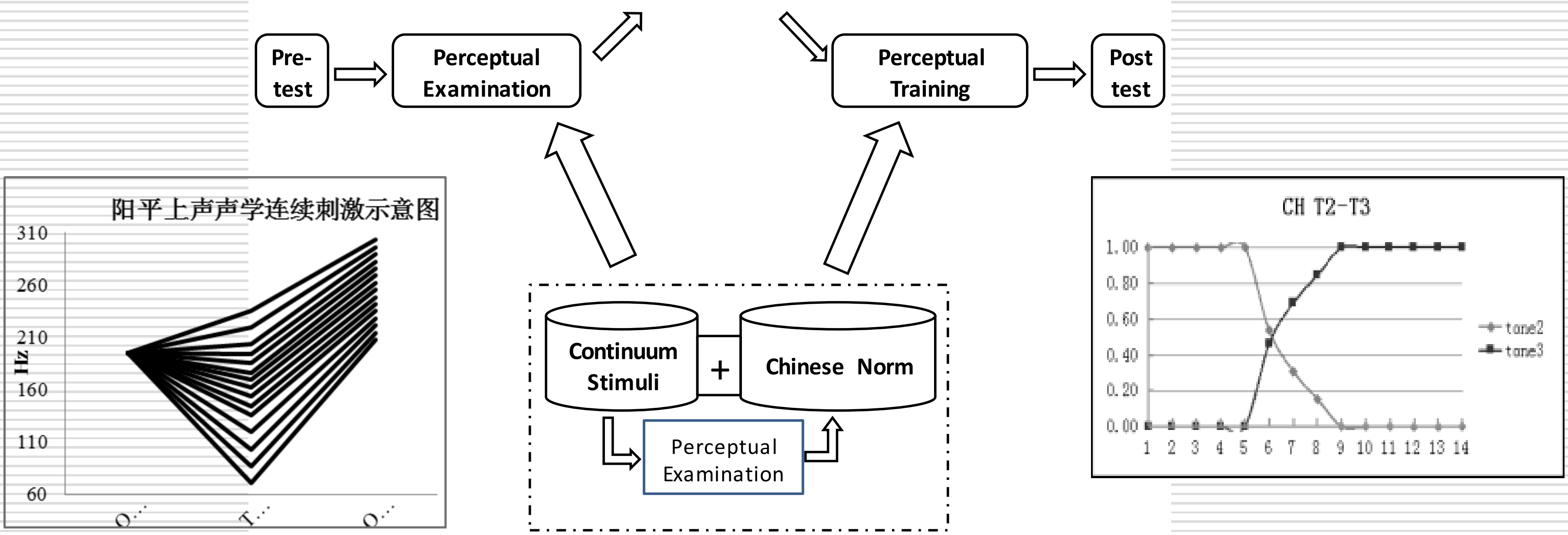
Block diagram of ITPT at SAIT



Perceptual Training of Tone






Pre-test

Perceptual Examination

Perceptual Training

Post test

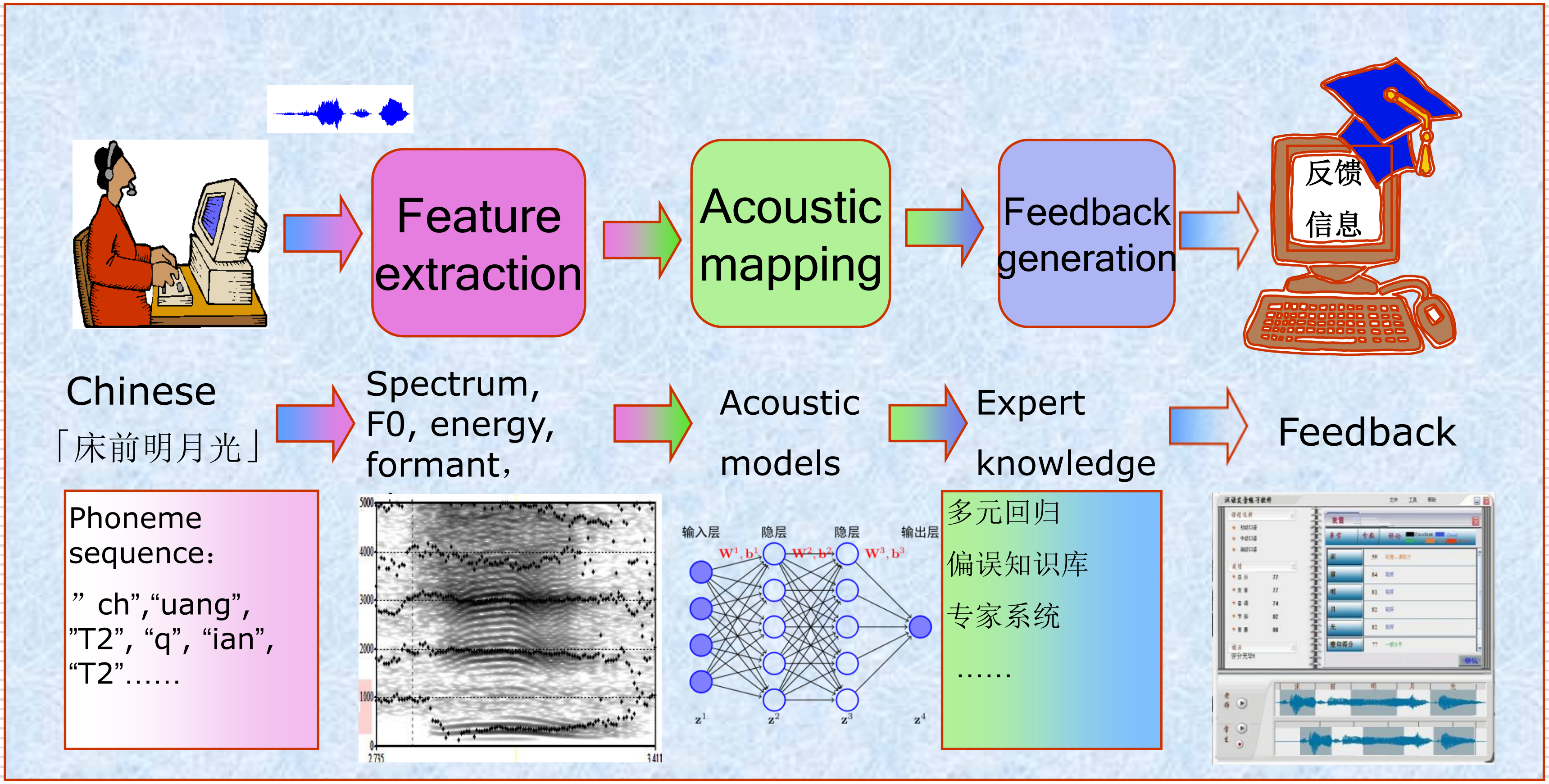
Continuum Stimuli

+

Chinese Norm

Perceptual Examination

Automatic Speech Assessment (ASA)



Feature extraction

Acoustic mapping

Feedback generation

反馈

信息

Chinese

「床前明月光」

Spectrum, F0, energy, formant,

Acoustic models

Expert knowledge

Feedback

Phoneme sequence:

” ch”,“uang”, ”T2”, “q”, “ian”, “T2”……

多元回归

偏误知识库

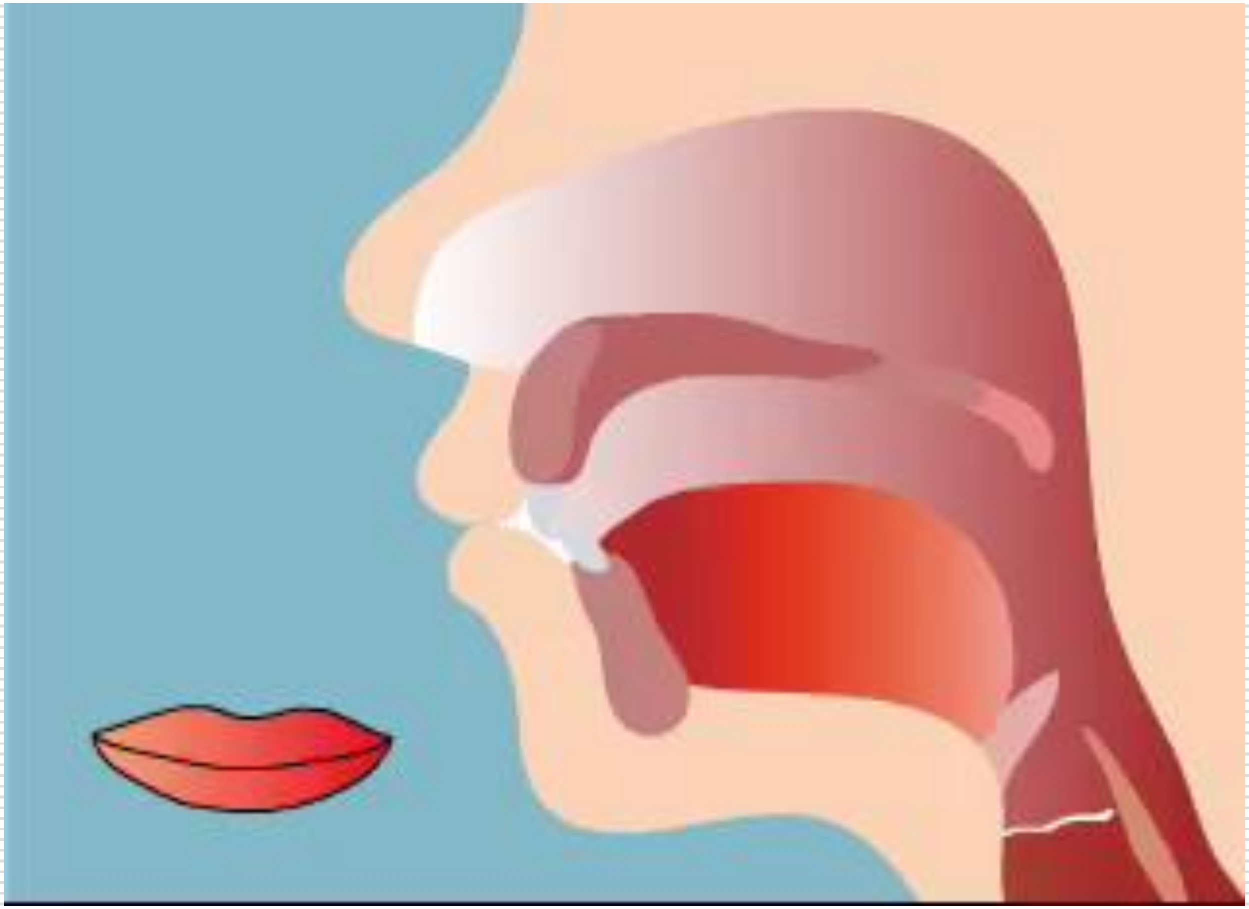
专家系统

……

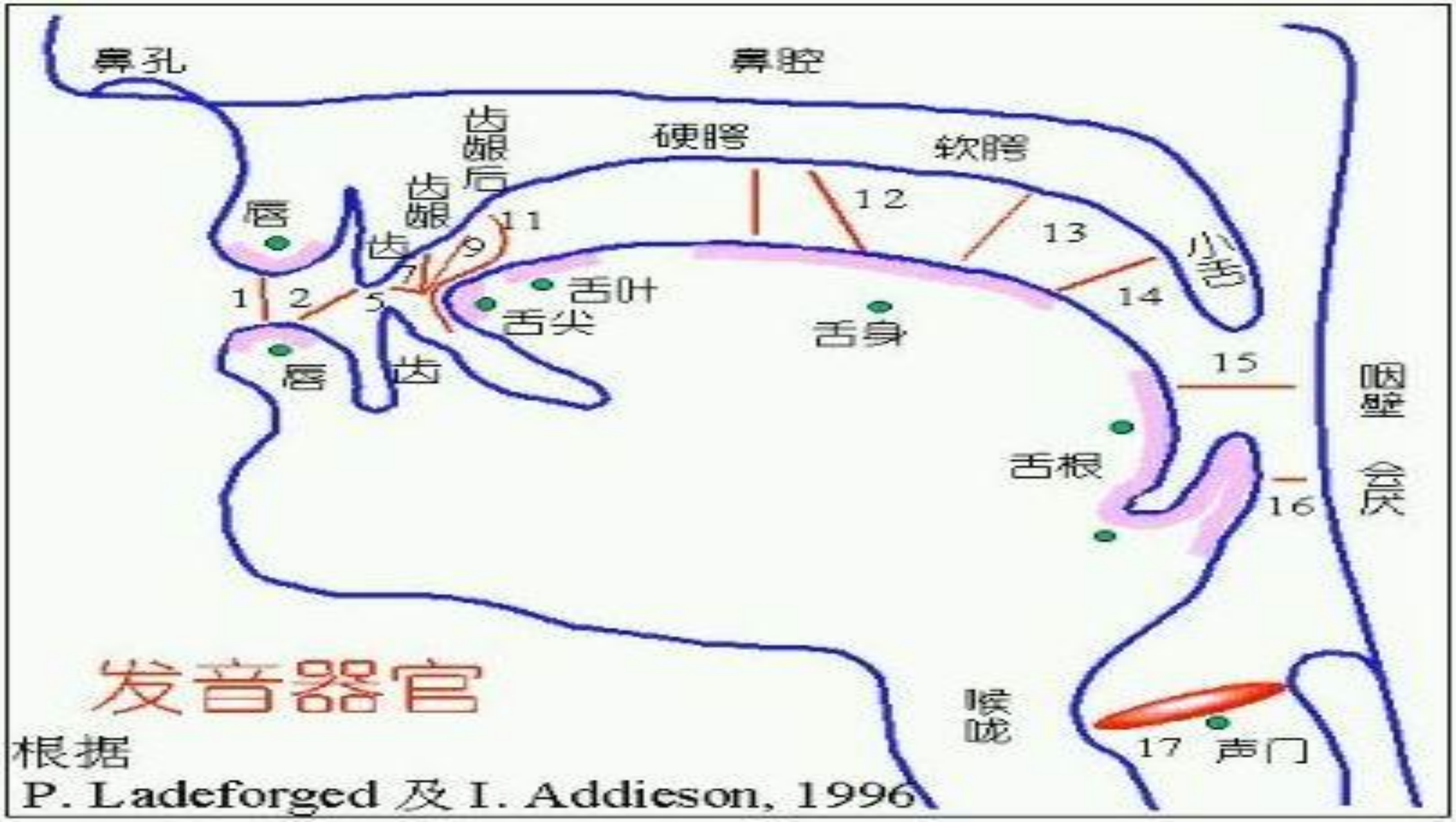
Feedbacks

- Scoring
- Corrective feedbacks:
  - Descriptions in speech or text
  - Picture
  - Audio Visual
  - 3-D animation
  - Etc.

Audio Visual Feedback



Real time 3d Articulatory Construction

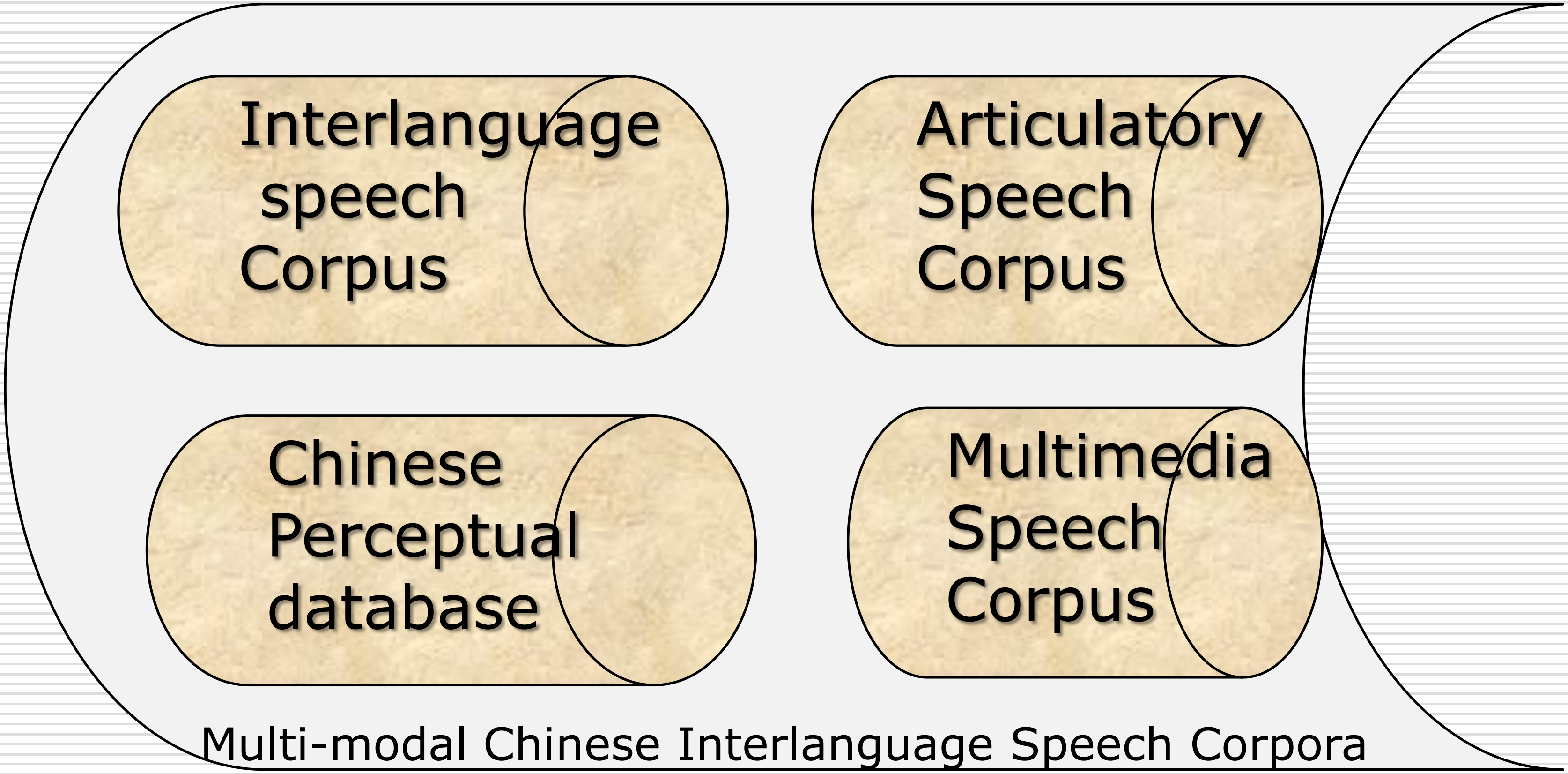


发音器官

根据

P. Ladeforged 及 I. Addieson, 1996

Overview of the Proposed Multi-modal Chinese Interlanguage Corpora



Interlanguage speech Corpus

Articulatory Speech Corpus

Chinese Perceptual database

Multimedia Speech Corpus

Multi-modal Chinese Interlanguage Speech Corpora

Outline

- The purpose of the database
- Feature descriptions
- Current status
- Conclusion

Modal I: Interlanguage Speech

- Design of the Chinese Interlanguage Speech corpus
  - Recording content
  - Speaker
- Representation of the corpus
  - Annotation considerations
  - Semi-automatic trial

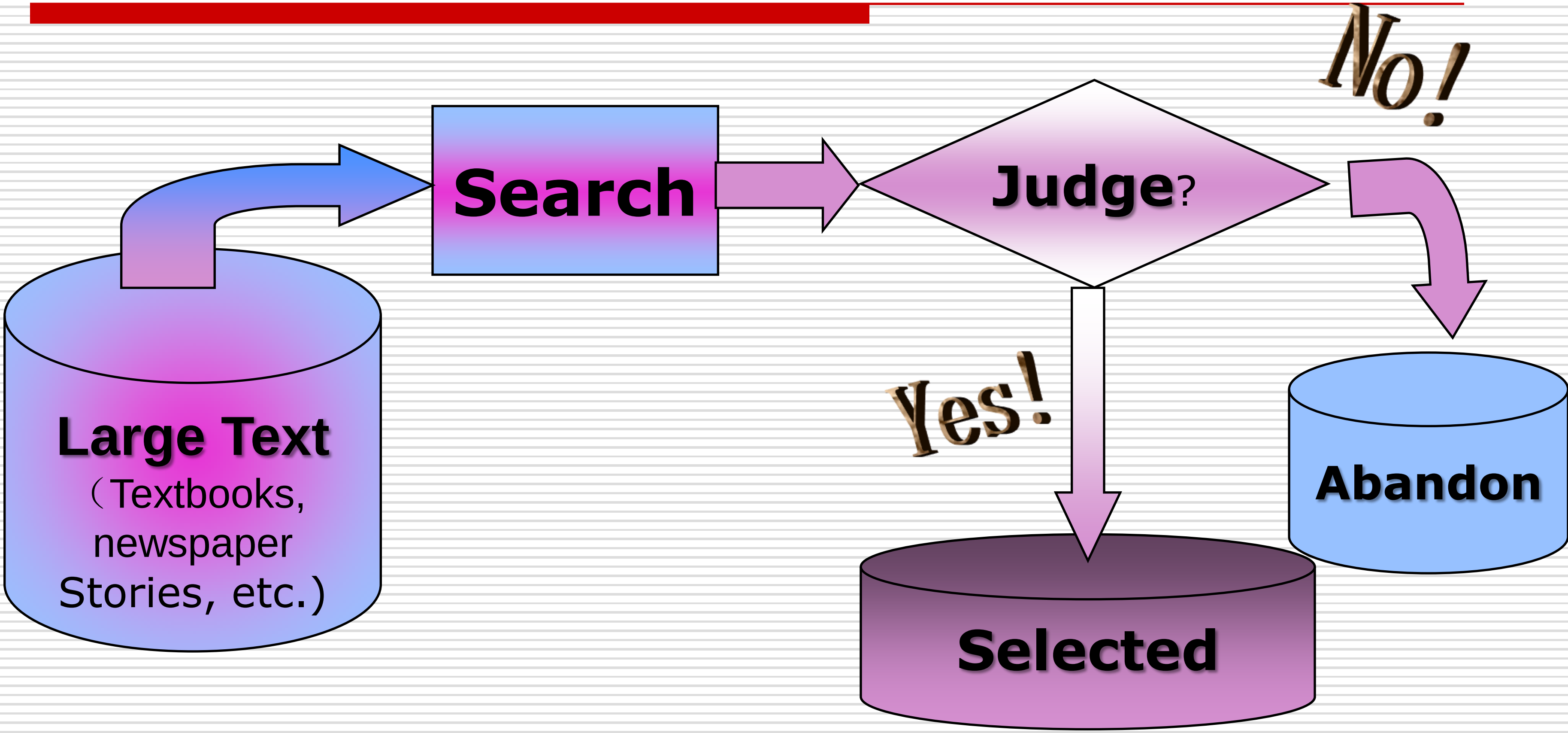
Content of the Chinese Interlanguage Speech corpus

- Two main applications:
  - 2 Language acquisition
  - Intelligent technology for pronunciation teaching/training
- Practical requirements:
  - Wide coverage of Chinese phonetic events
  - Small text size to facilitate collection of more speakers
- Phonetic events:
  - Basic: phonemes, tones
  - Co-articulations of: phones, tones
  - Prosody: phrasing, focus, intonation, etc.

Recording Script

- □ Mono-syllables
- □ Bi-syllables
- □ Minimum sentence set
- □ Short paragraph

Text Design: Search



Minimum Sentence Set

□ Tri-tone units

T3T4T1（打印机）

$$5 \times 5 \times 5 = 125$$

□ Boundary effect

T3 T4 ¦ T3 T4 T2
打 印 ¦ 机 械 图

T3T4T3、T4T3T4、T3T4T2
打印机　印机械　机械图

T3T4、T3T4 T3T4T2
打印　机械 机械图

$$125+5 \times 5 + 5 \times 5 + 5 = 180$$

Minimum Sentence Set：103

- Size
  - 103 sentences
  - 610 word tokens
  - 1340 characters

Speech duration < 10 Minutes.

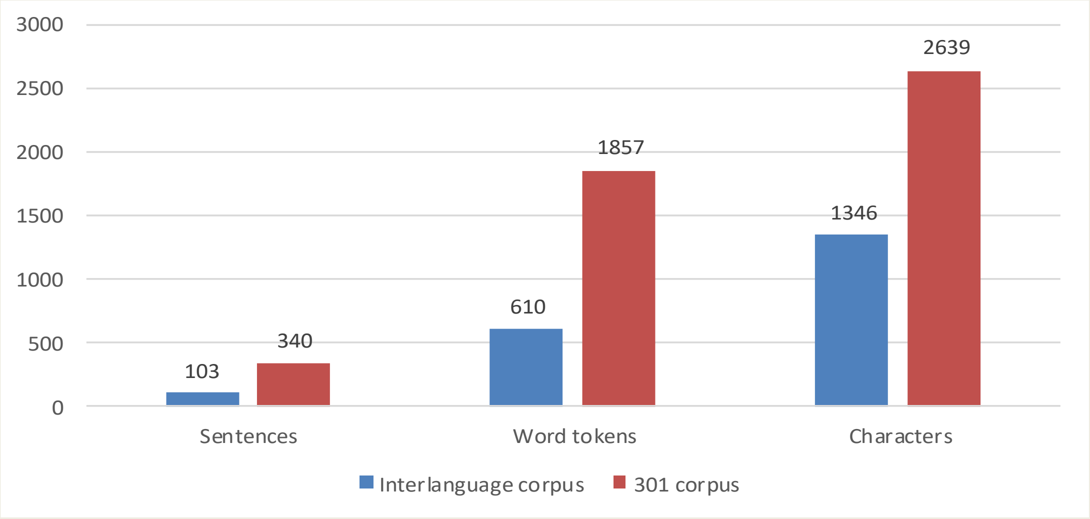
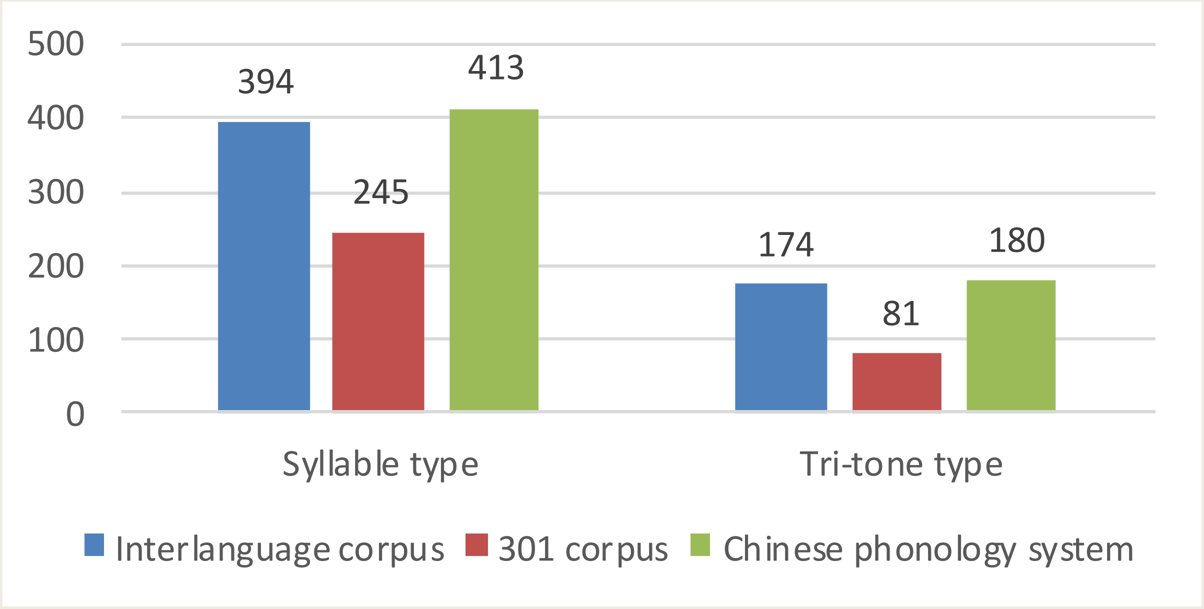
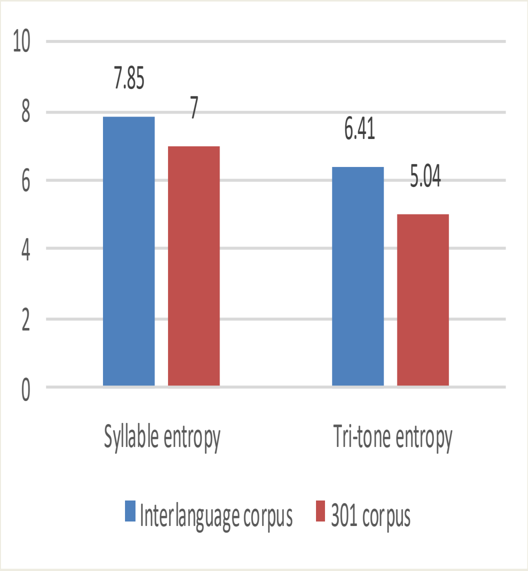


Fig 1: phonetic units of two corpora

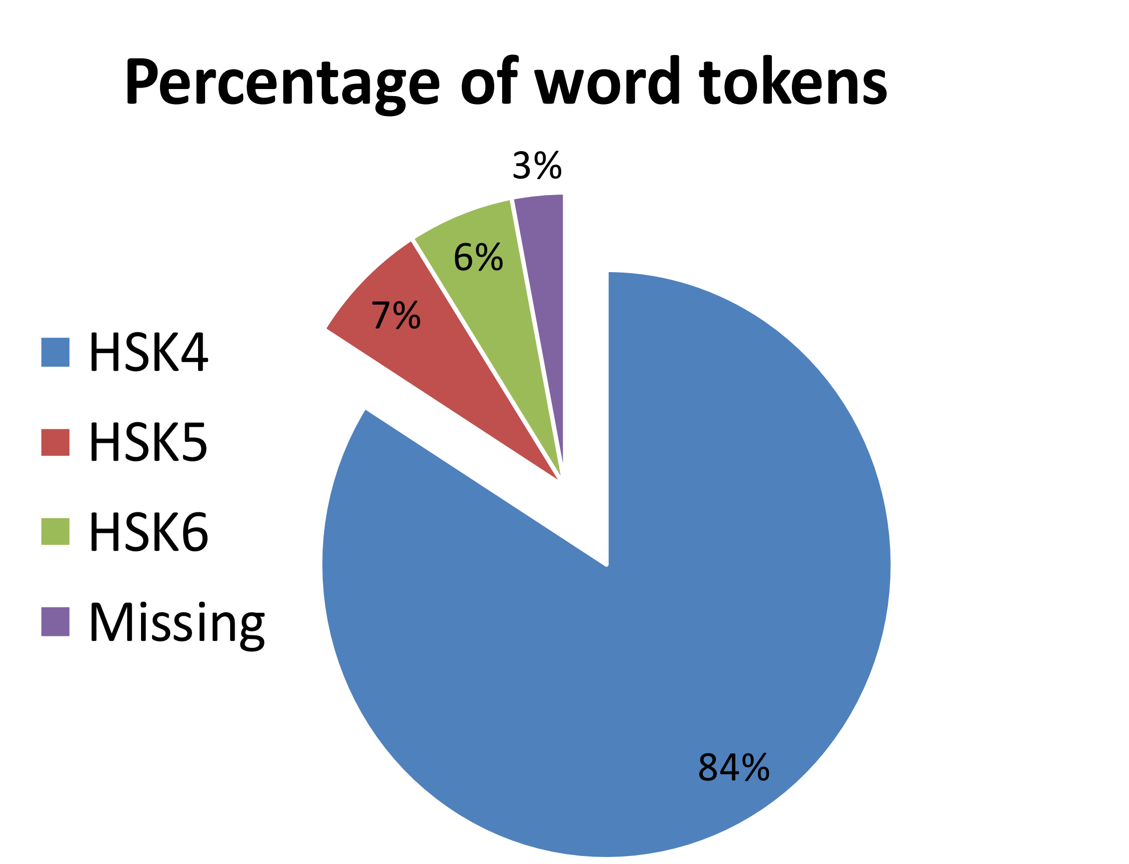
Statistics: 103 vs. 301

□ Phonetic types and its distribution





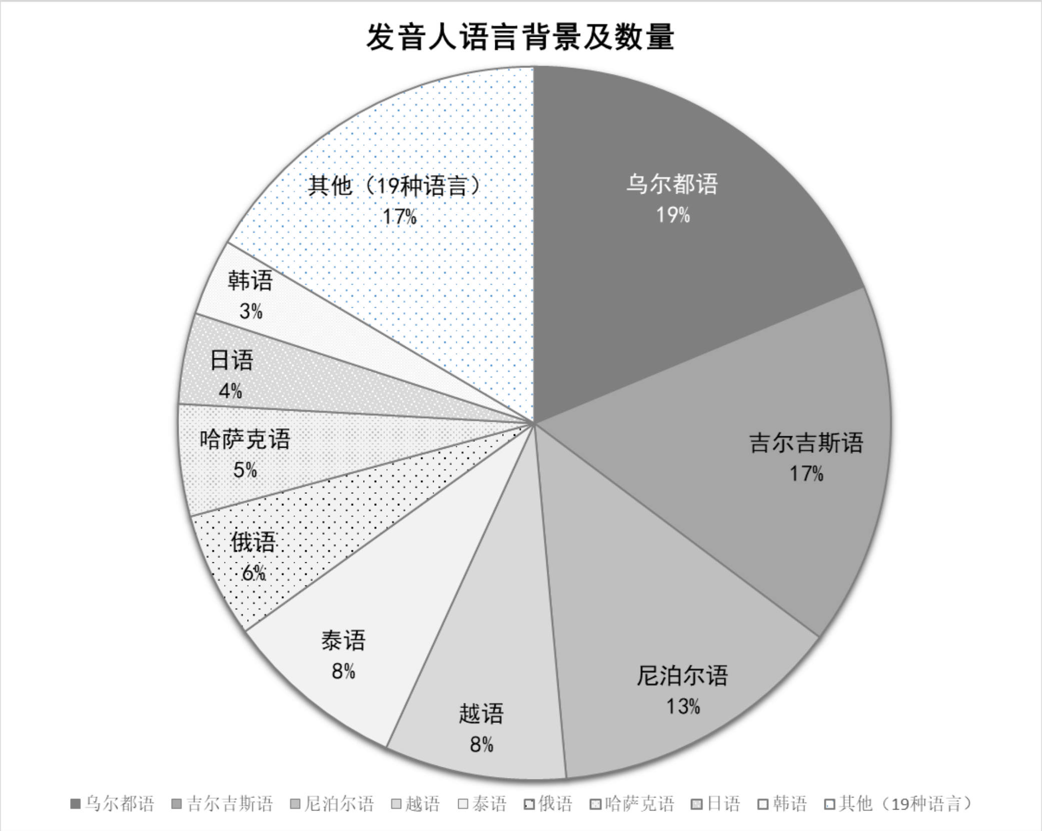
Difficulty of Vocabulary



Number of Speakers Collected by Now

- □ Total #: 314人
- □ Country #: 31
- □ Mother tongue: 27

国家	语言背景	人数
巴基斯坦	乌尔都语	59
吉尔吉斯	吉尔吉斯语	52
尼泊尔	尼泊尔语	42
越南	越语	26
泰国	泰语	26
俄罗斯	俄语	18
哈萨克斯坦	哈萨克语	16
日本	日语	13
韩国	韩语	11



Speakers’ Information

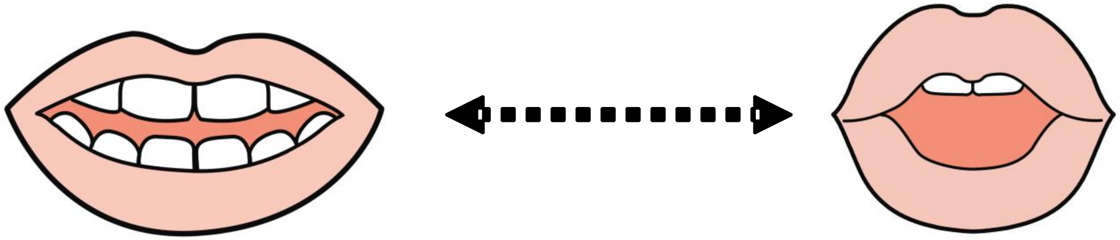
国家	语言背景	人数
马来西亚	马来语	9
印度尼西亚	印尼语	7
塔吉克	塔吉克语	6
缅甸	缅甸语	5
埃及、苏丹	阿拉伯语	4
英、加、美	英语	4
西班牙	西班牙语	3
乌兹别克	乌兹别克语	2
蒙古	蒙古语	2
印度	印地语	1
土库曼	土库曼语	1
科摩罗	斯瓦希里语	1
斯里兰卡	僧伽罗语	1
孟加拉	孟加拉语	1
卢旺达	卢旺达语	1
柬埔寨	高棉语	1
法国	法语	1
伊朗	波斯语	1
阿塞拜疆	阿塞拜疆语	1

Speakers’ information

国家	人数	语言	语族	语系
巴基斯坦	59	乌尔都语	印度语族	印欧语系（137人）
尼泊尔	42	尼泊尔语		
印度	1	印地语		
斯里兰卡	1	僧伽罗语		
孟加拉	1	孟加拉语		
英国	2	英语	日耳曼语族	
加拿大	1	英语		
美国	1	英语		
塔吉克	6	塔吉克语	伊朗语族	
伊朗	1	波斯语		
俄罗斯	18	俄语	斯拉夫语族	
法国	1	法语	罗曼语族	
西班牙	3	西班牙语	拉丁语族	
吉尔吉斯	52	吉尔吉斯语	突厥语族	阿尔泰语系（74人）
哈萨克斯坦	16	哈萨克语		
乌兹别克	2	乌兹别克语		
土库曼	1	土库曼语		
阿塞拜疆	1	阿塞拜疆语		
蒙古	2	蒙古语	蒙古语族	

Annotation: Pronunciation Erroneous Tendency

- Tendency instead of identification



Lip: spread rounding

Pinyin: “e” “o”

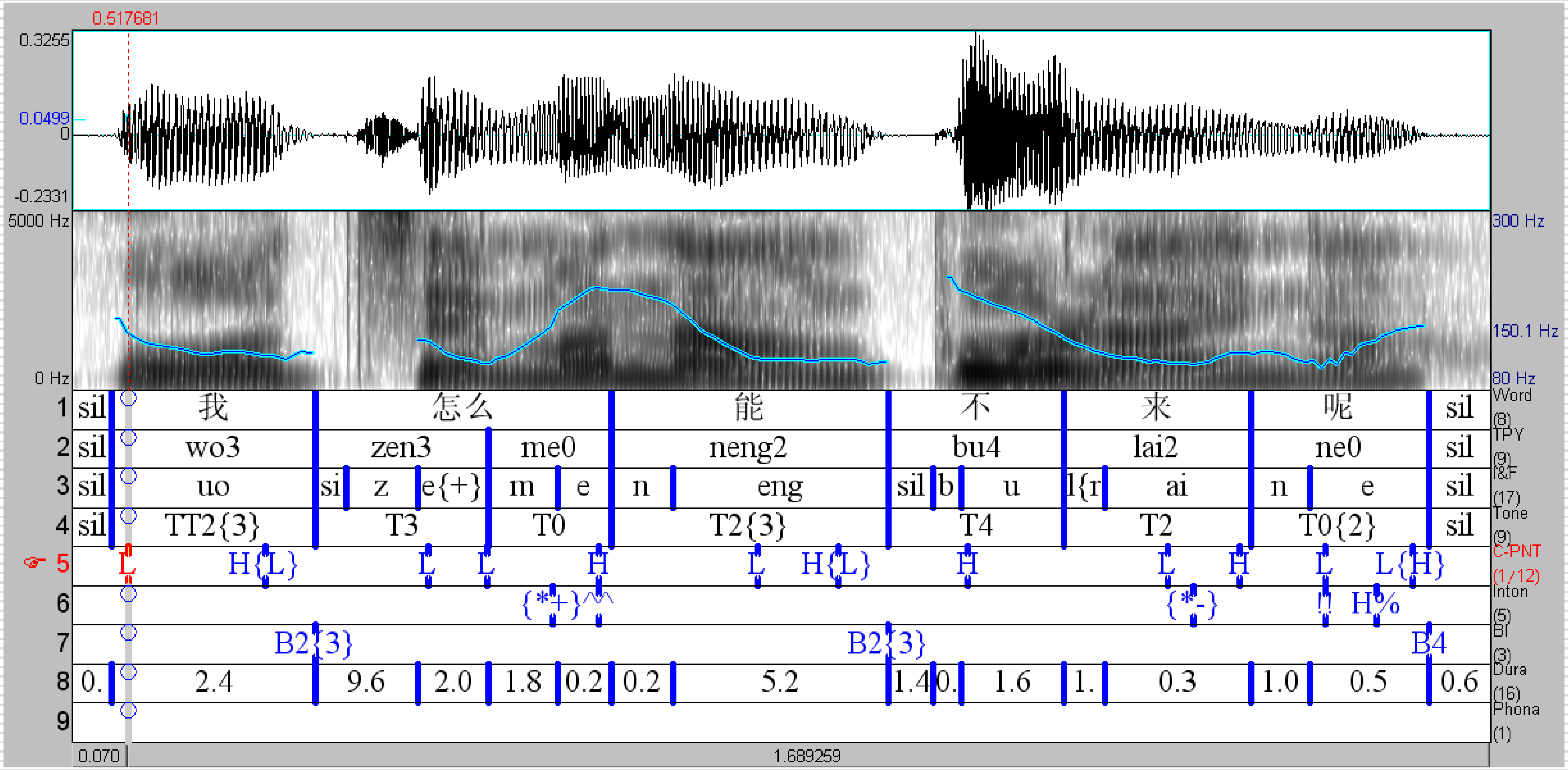
Rounded “e” sound: “e{o}”

Spreading “o” sound: “o{w}”

Examples

类型	标注符号	偏误举例	备注/说明
Higher	^	a{^}	a的舌位与标准音相比不够低，发音近似[©]
Lower	!	u{!}	u与标准音相比舌位过低，发音近似[¦]
Fronter	+	e{+}n	e的舌位靠前，en发音近似[”n]
Backer	-	n{-}	前鼻音发音近似后鼻音
Longer	:	z{:}	z[ts]（的擦音段）发音太长
Shorter	;	p{;}	p[p^h]（的送气段）时长不够
Central	”	uo{” }	uo中的o的舌位同时低化、前化，uo近似[u§]
Rounding	o	e{o}	e似被发成了圆唇音
Spreading	w	f{w}, u{w}	f被发成双唇擦音，u被发成了不圆唇音
Linguolabial	f	u{f}	u被发成[v]
Laminal	sh	sh{sh}	普通话的sh被发成[□]

Annotation Example



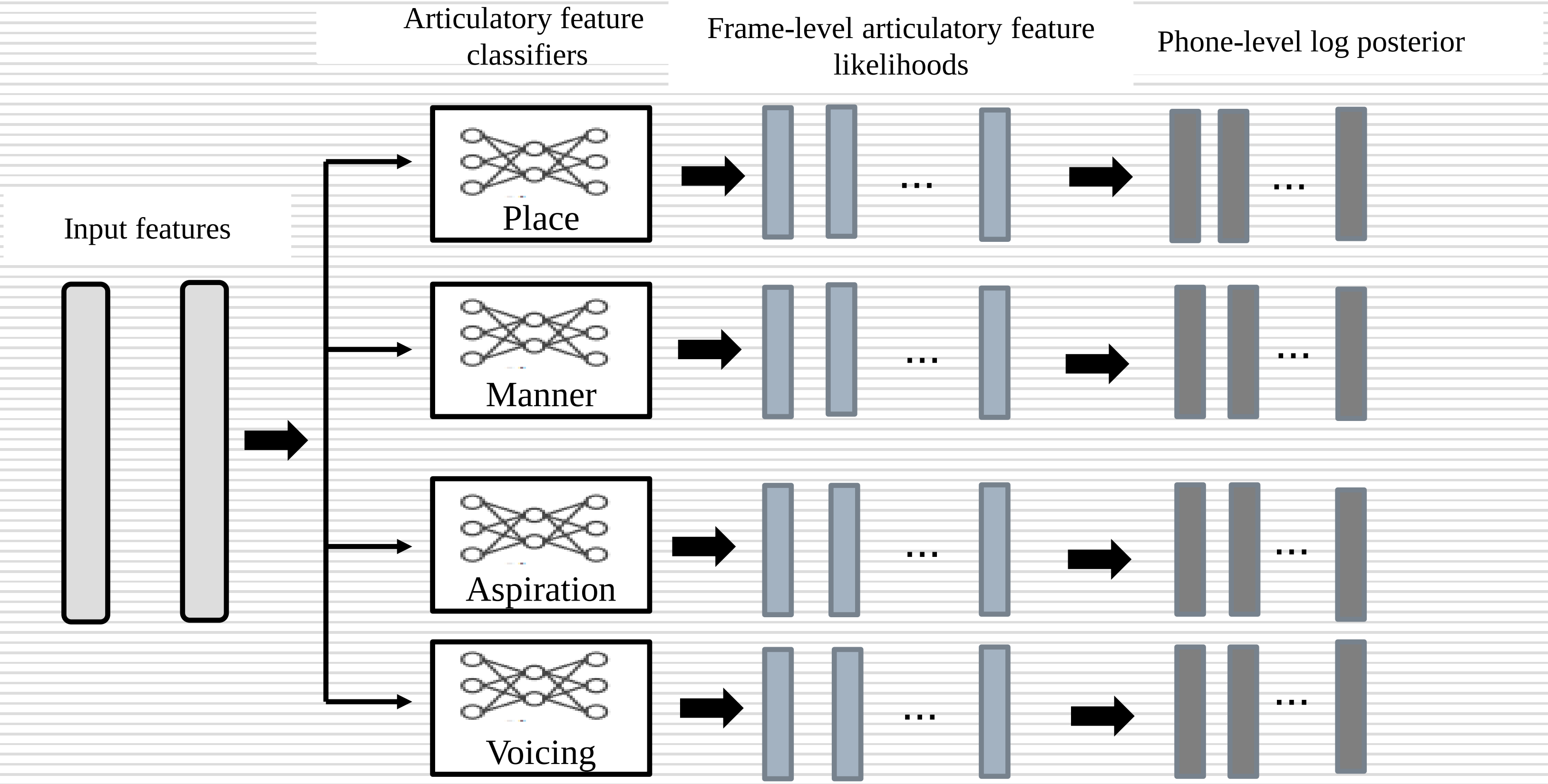
Semi-automatic Annotation

- Problem: manual annotation is of low efficiency.
- Solution: semi-automatic annotation
  - Automatic attribute detection
  - Attribute based PET prediction
  - Manual check

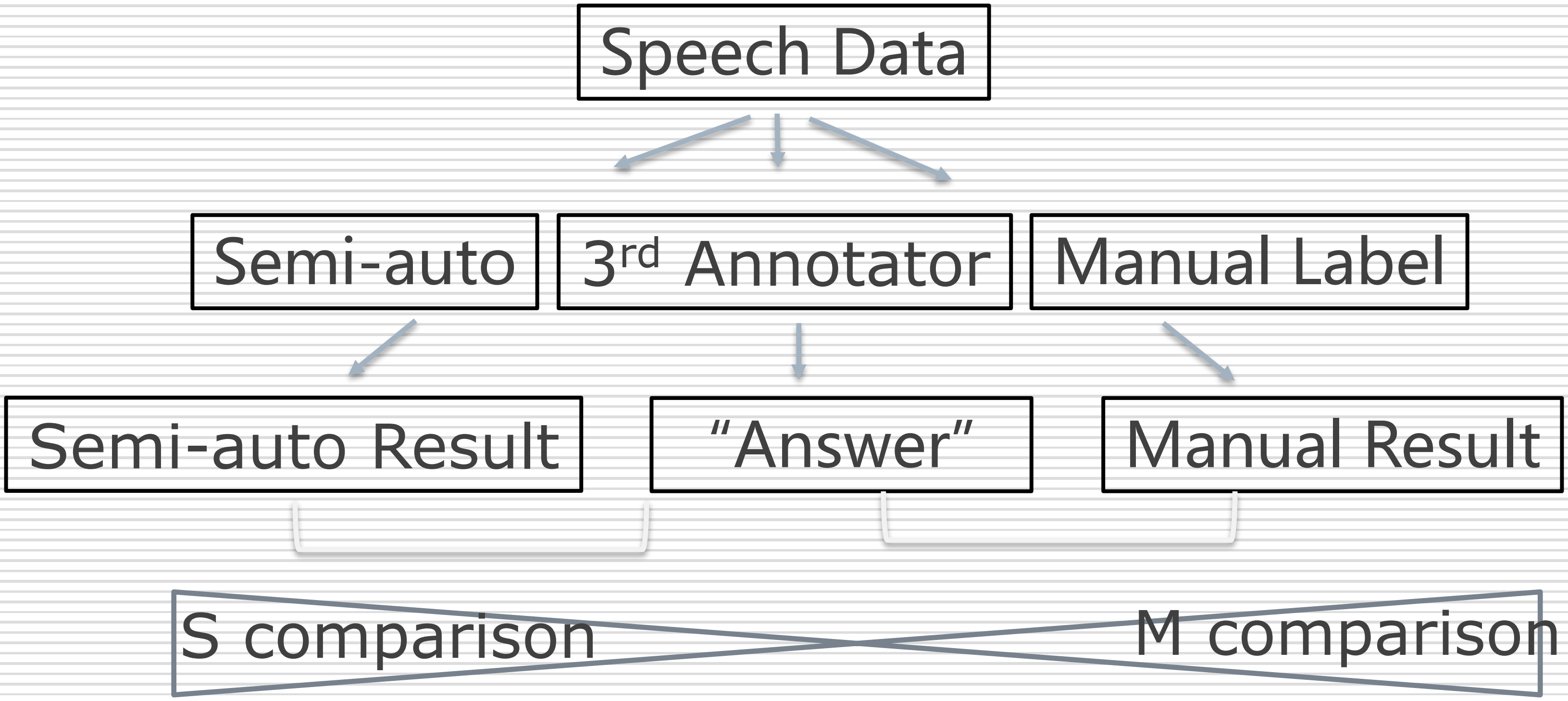
Attributes of Consonants

manner	Stops		Affricative		Fricative		Nasal Vcd	Lateral Vcd
Place	Unasprt	asprt	unasprt	asprt	unvcd	vcd		
Bilabial	b	p					m	
Labio-dental					f			
Dental			z	c	s			
Alveolar	d	t					n	l
Retroflex			zh	ch	sh	r		
Palatal			j	q	x			
Velar	g	k			h		ng	

DNN based Attribute Detection


Articulatory feature classifiers
Frame-level articulatory feature likelihoods
Phone-level log posterior
Input features
Place
Manner
Aspiration
Voicing
…

Pilot Study



Speech Data

Semi-auto

3rd Annotator

Manual Label

Semi-auto Result

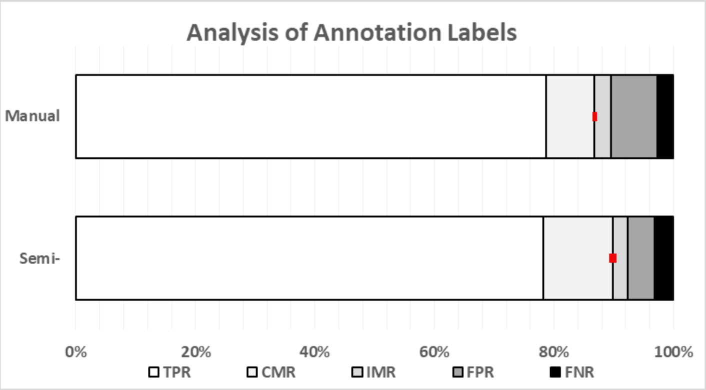
“Answer”

Manual Result

~~S comparison M comparison~~

Results

	Semi-Auto	Manual
consistency :	83.6%	86.6%
deletion :	4.5%	7.8%
insertion :	3.3%	2.7%
Correct hit :	89.8%	86.8%



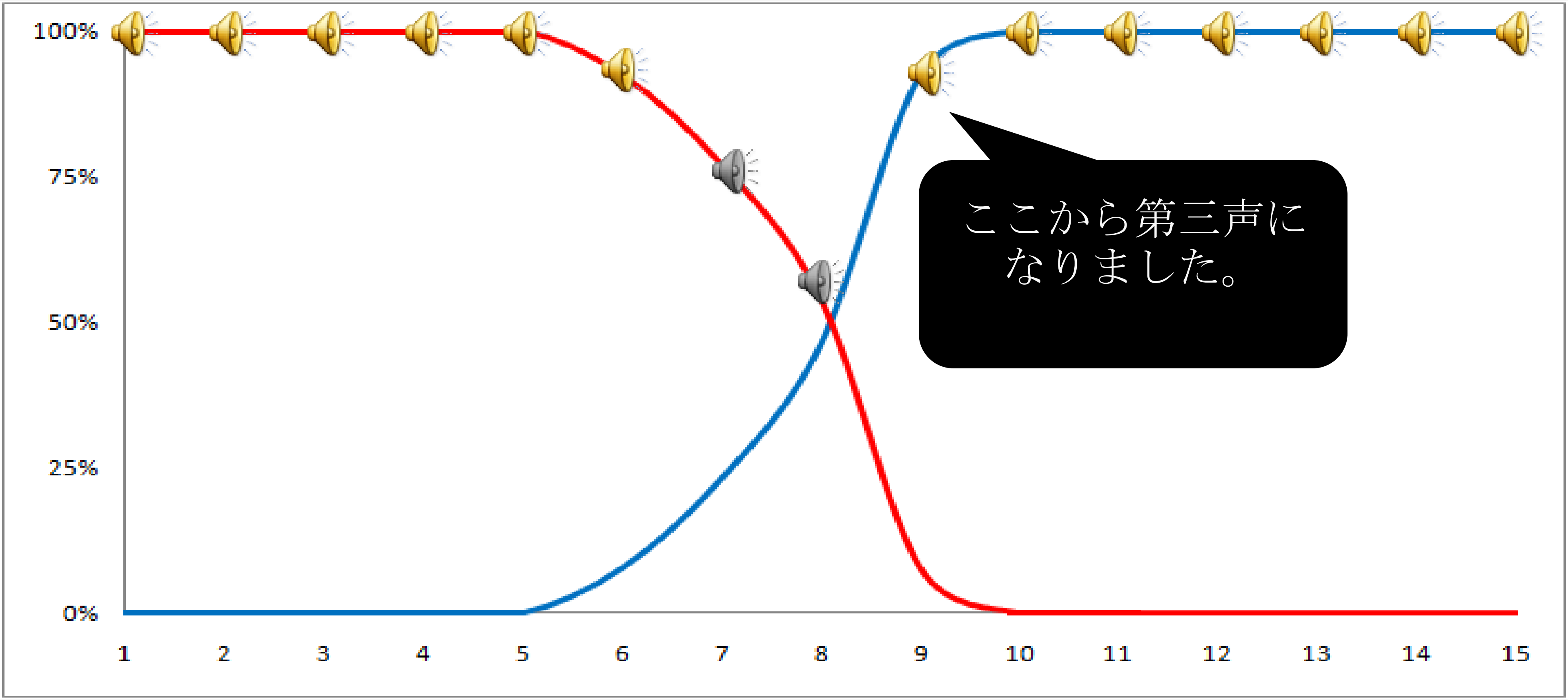
Modal 2: Perception Corpus

- Purpose:
  - Phonetic categories ←→ multi-dimensional acoustic cues
  - Influences of different mother tongues
  - Contrast analyses
  - Perceptual training
- Specifications:
  - Tones
  - Segments

Perceptual Studies

- □ Tonal perceptions of various kinds of mother tongues
- □ Influences of different formants on tone perception
- □ Aspiration's effects on tone perception of syllables with affricate Initials
- □ Formant's effects on perception of velar/alveolar Finals
- □ Influences from erroneous segments on tone perception
- □ Key acoustic cue to perception of "l/r", etc.

- 横軸は前図の番号で、縦軸は一般中国人判断の結果です。第二声は赤線で、第三声は青線で表示しています。



ここから第三声になりました。

Perceptual Results of T2_T3

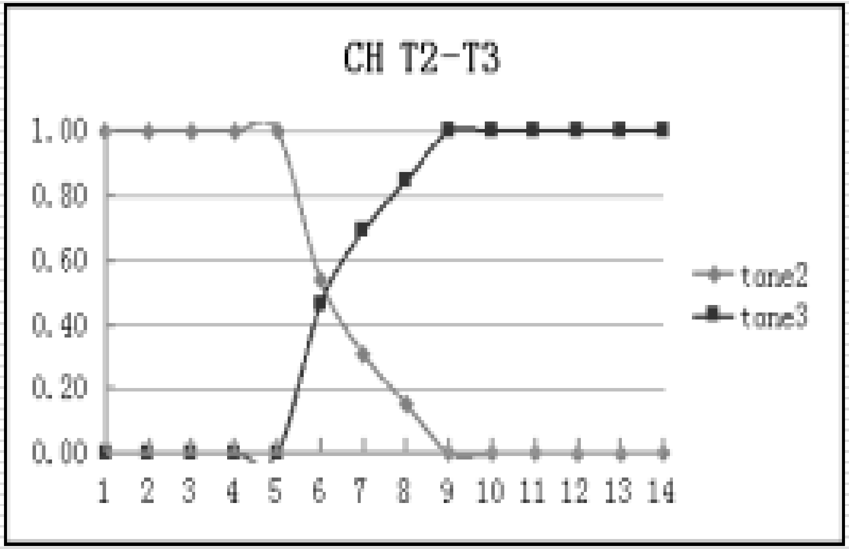


Fig. Identification curves of Chinese.

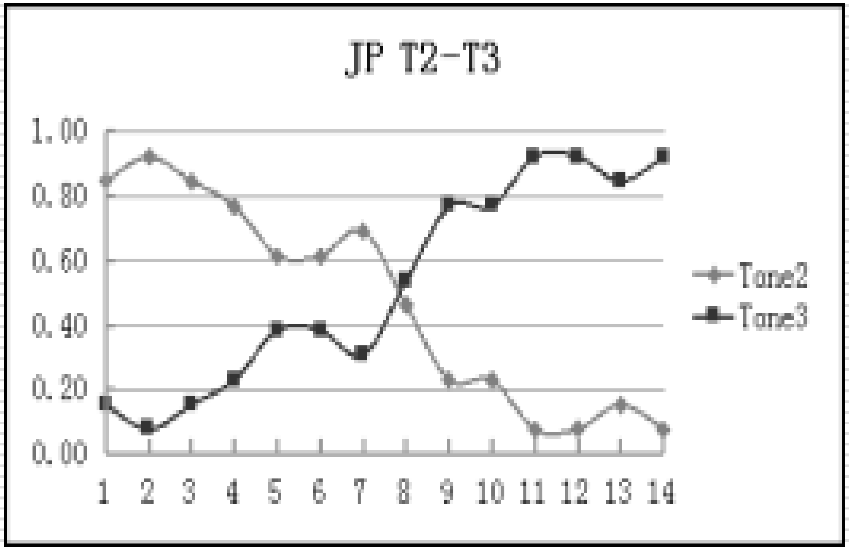
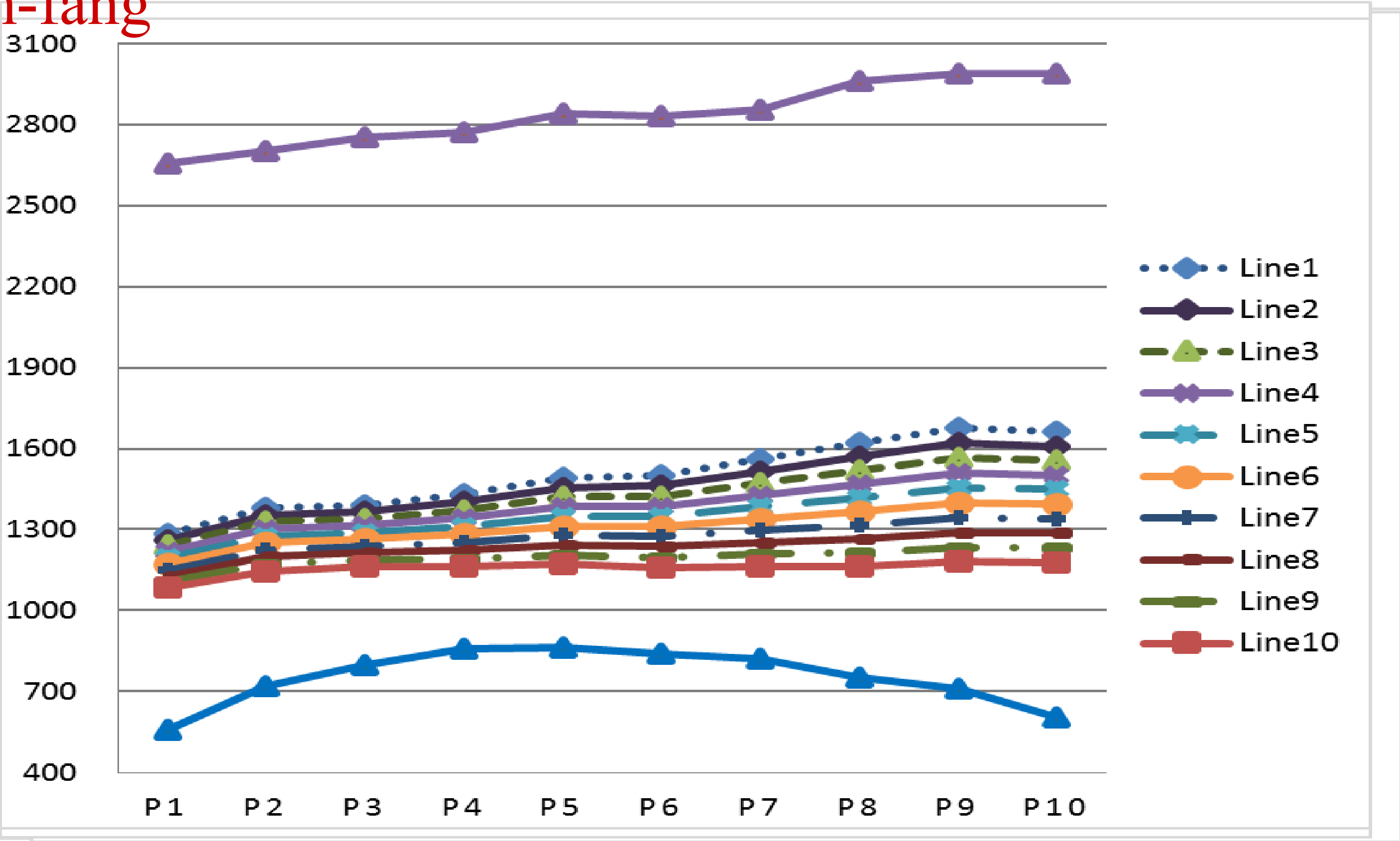


Fig. Identification curves of Japanese.

Illustration of Continua of Velar-Alveolar Nasals

Ex: fān-fāng

/f/ ✚



✚ /n/

Line1：fān　Line2：fāng

Perceptual Results of “fan_fang”

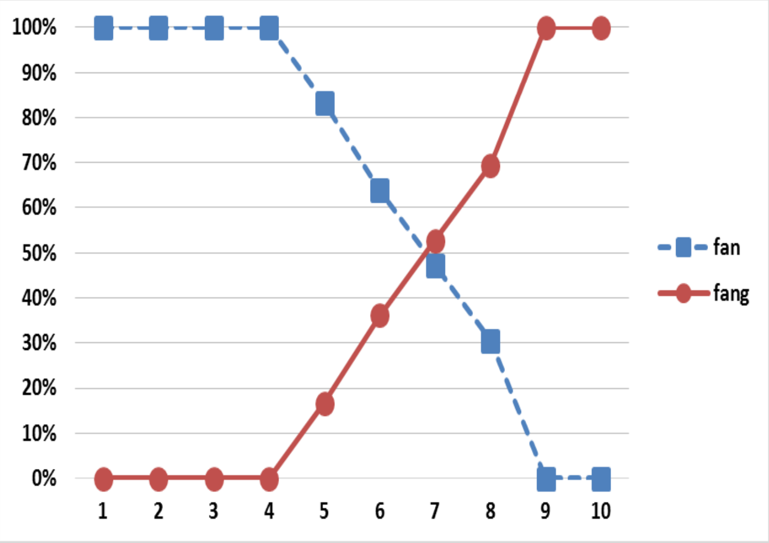


Fig. Chinese natives.

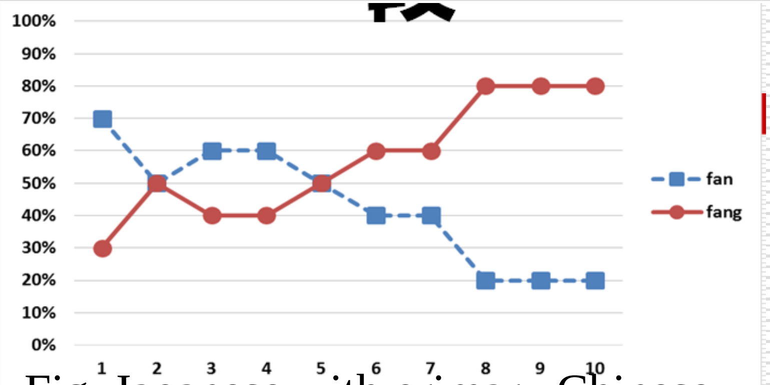


Fig. Japanese with primary Chinese language skills.

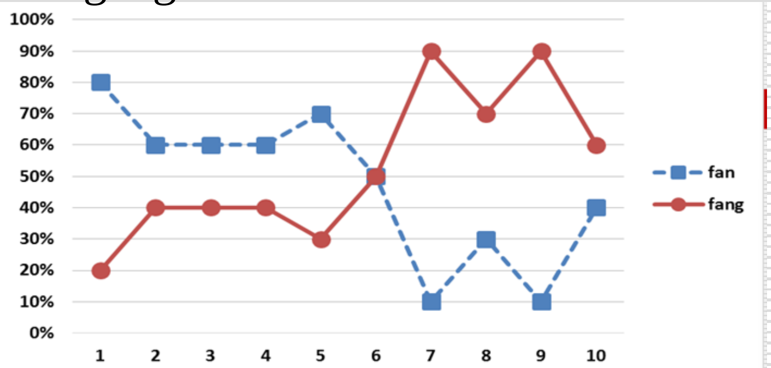
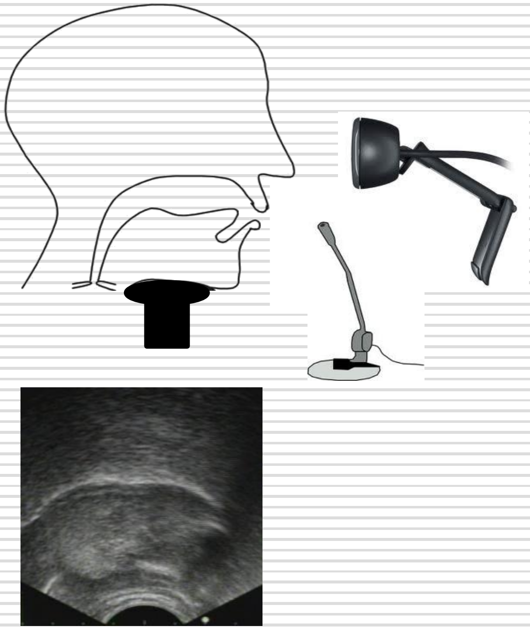


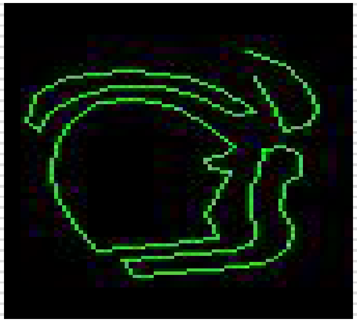
Fig. Japanese with mid-high level of Chinese language skills.

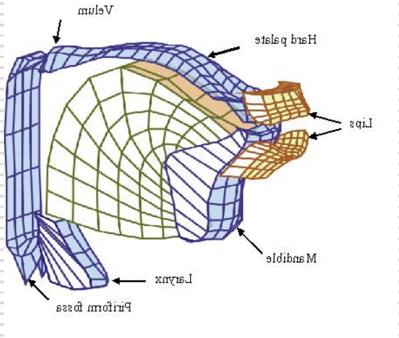
Modal III: Visual and Articulatory data

- Contents
  - Speech signal
  - Facial visual motion signal
  - Ultrasound articulatory data
  - EMA
  - MRI
- Speech materials:
  - Mono-syllables
  - Short sentences

Purpose







Acquisition

Including Speech, facial motion Articulator’s movements

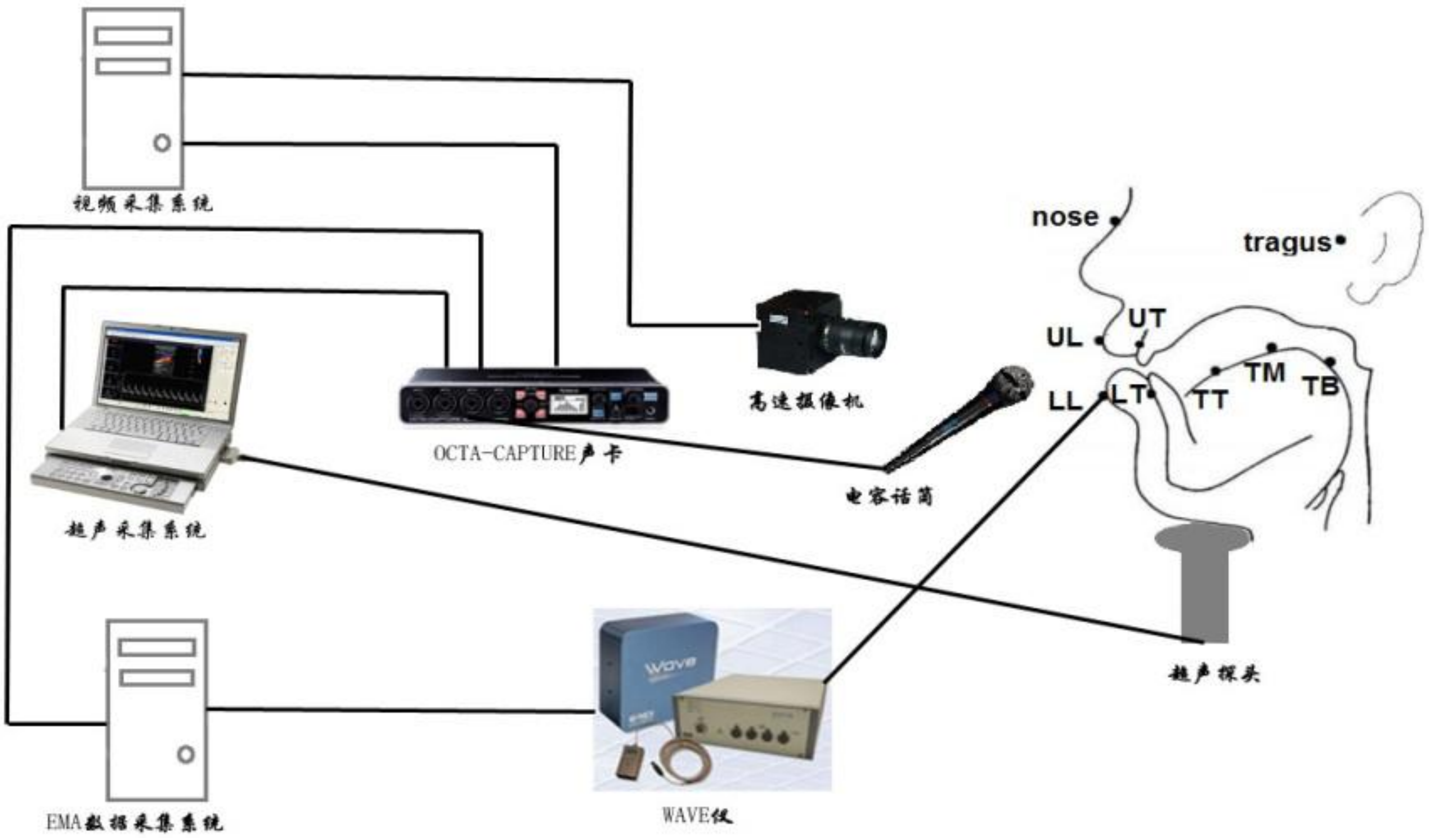
Image

Contour extraction of articulators

Animation

Generating the movement of vocal tract

Collection system

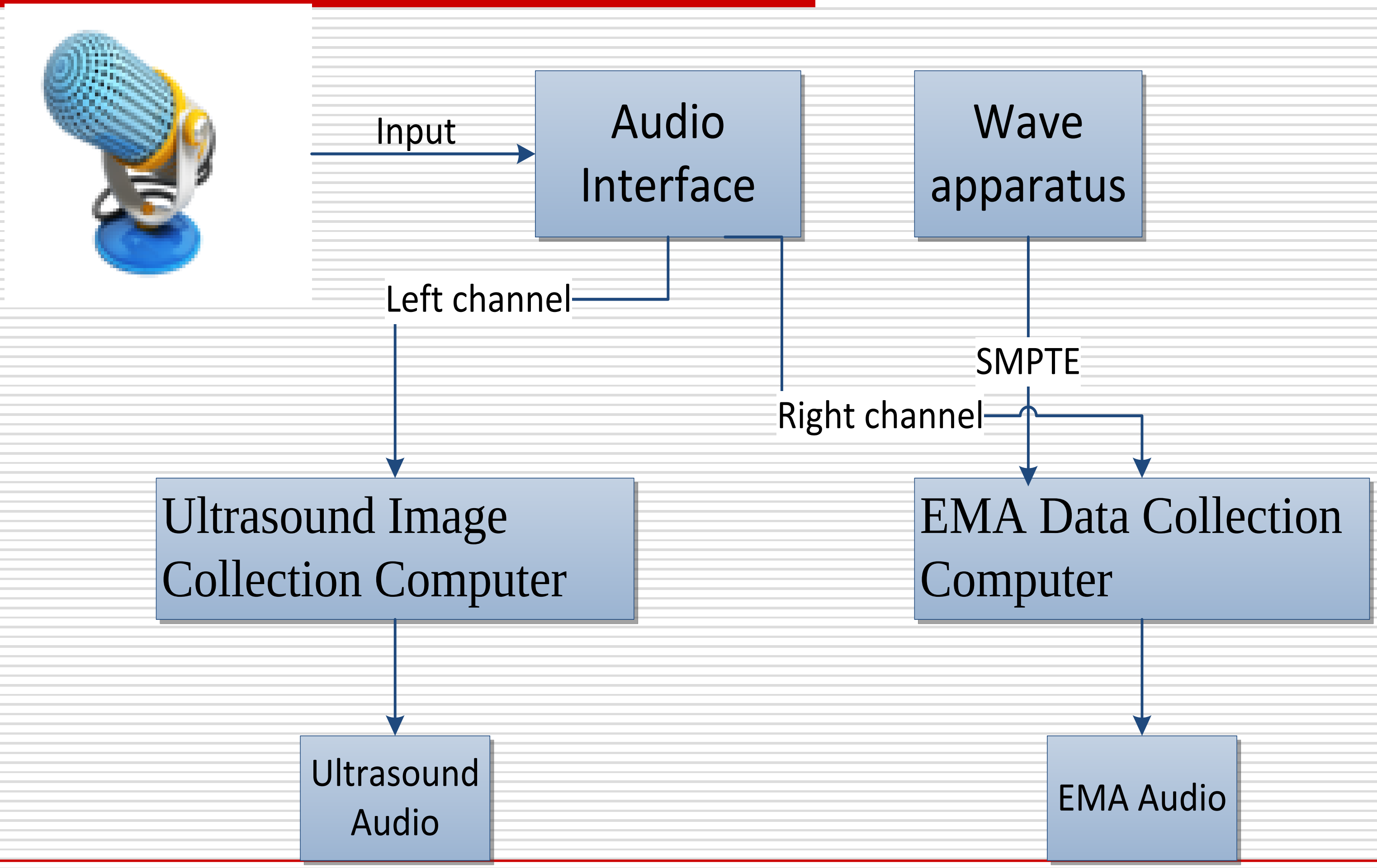


Introduction to the data and the equipment

- ➢Ultrsound Image Equipment
  - ➢ Terason, T3000 Ultrasound apparatus
  - ➢ Probe: 8MC3

- ➢EMA Data Collection Equipment
  - ➢ Canada, NDI, WAVE apparatus

Method to synchronize and device attachment



Input

Audio Interface

Wave apparatus

Left channel

SMPTE

Right channel

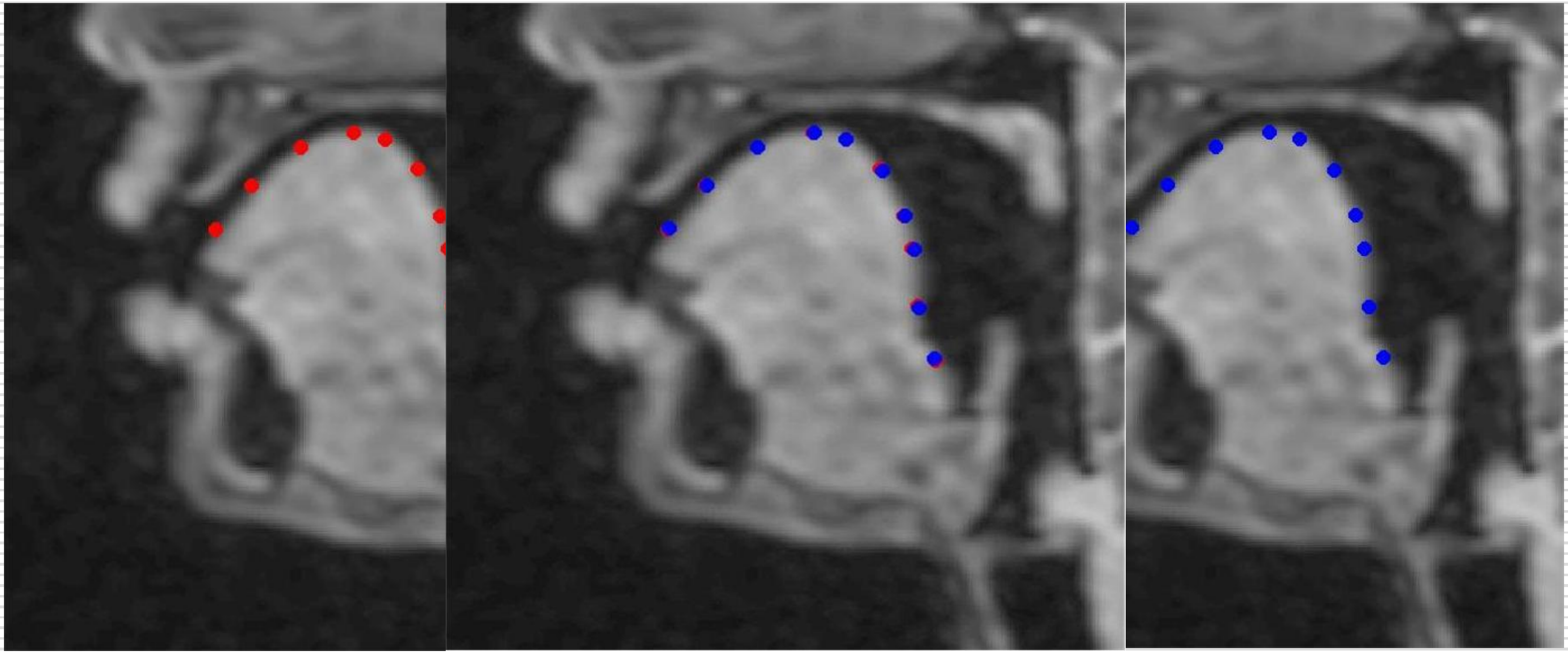
Ultrasound Image Collection Computer

EMA Data Collection Computer

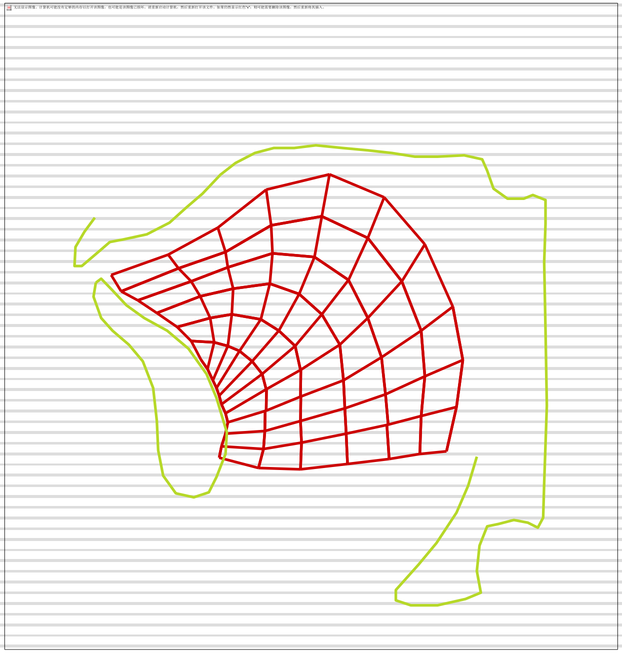
Ultrasound Audio

EMA Audio

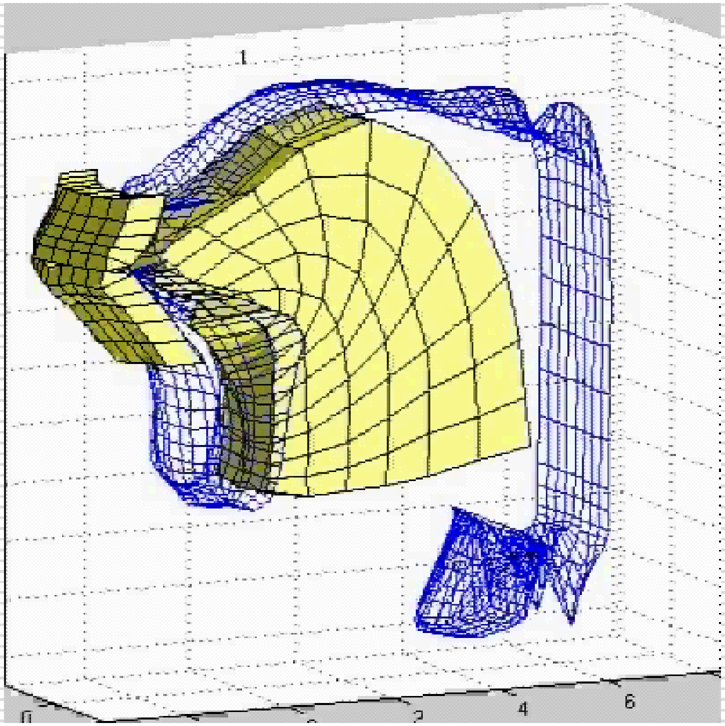
Data Synchronization



From Articulatory Data to 3-D Animation

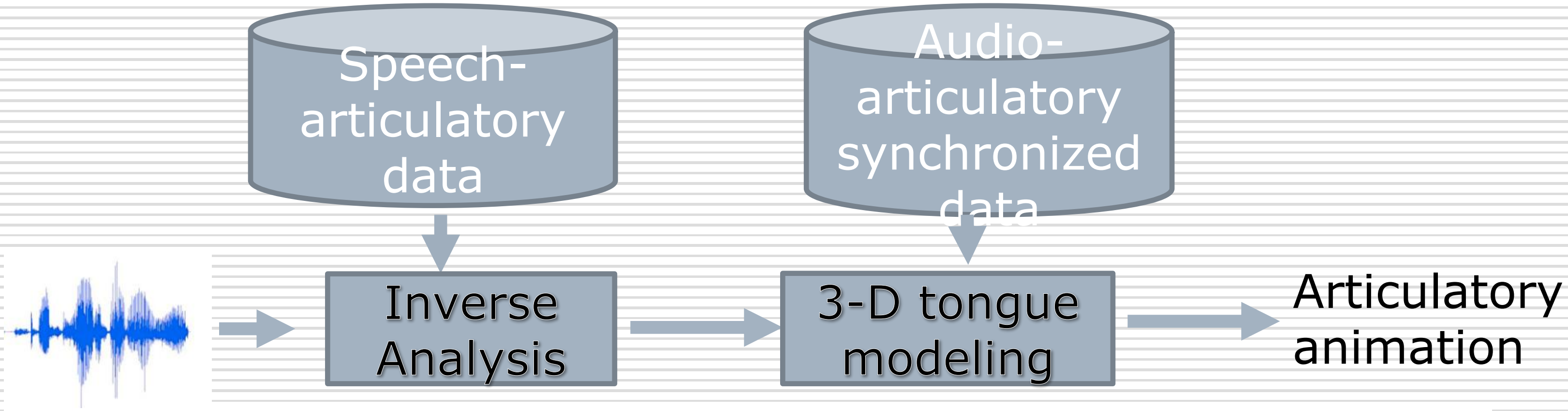


Tongue contour extraction



3-D modeling

DNN based Articulatory Synthesis



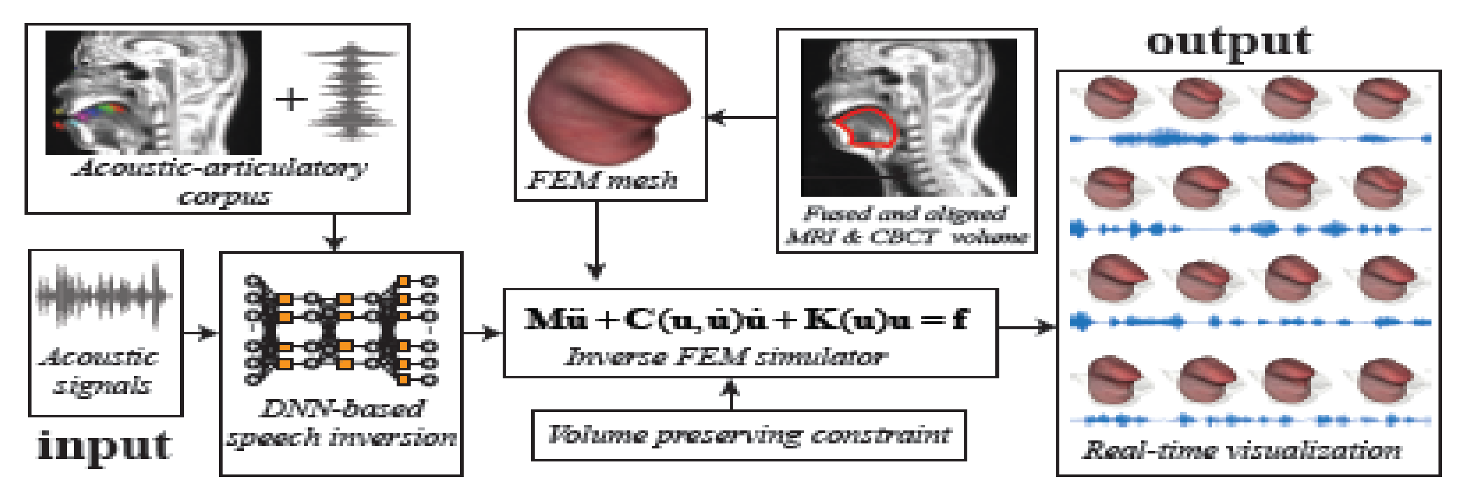
Speech-articulatory data

Audio-articulatory synchronized data

Inverse Analysis

3-D tongue modeling

Articulatory animation



Acoustic-articulatory corpus
FEM mesh
Fused and aligned MRI & CBCT volume
output
Acoustic signals
input
DNN-based speech inversion
$\mathbf{M\ddot{u}+C(u,\dot{u})\dot{u}+K(u)u=f}$
Inverse FEM simulator
Volume preserving constraint
Real-time visualization

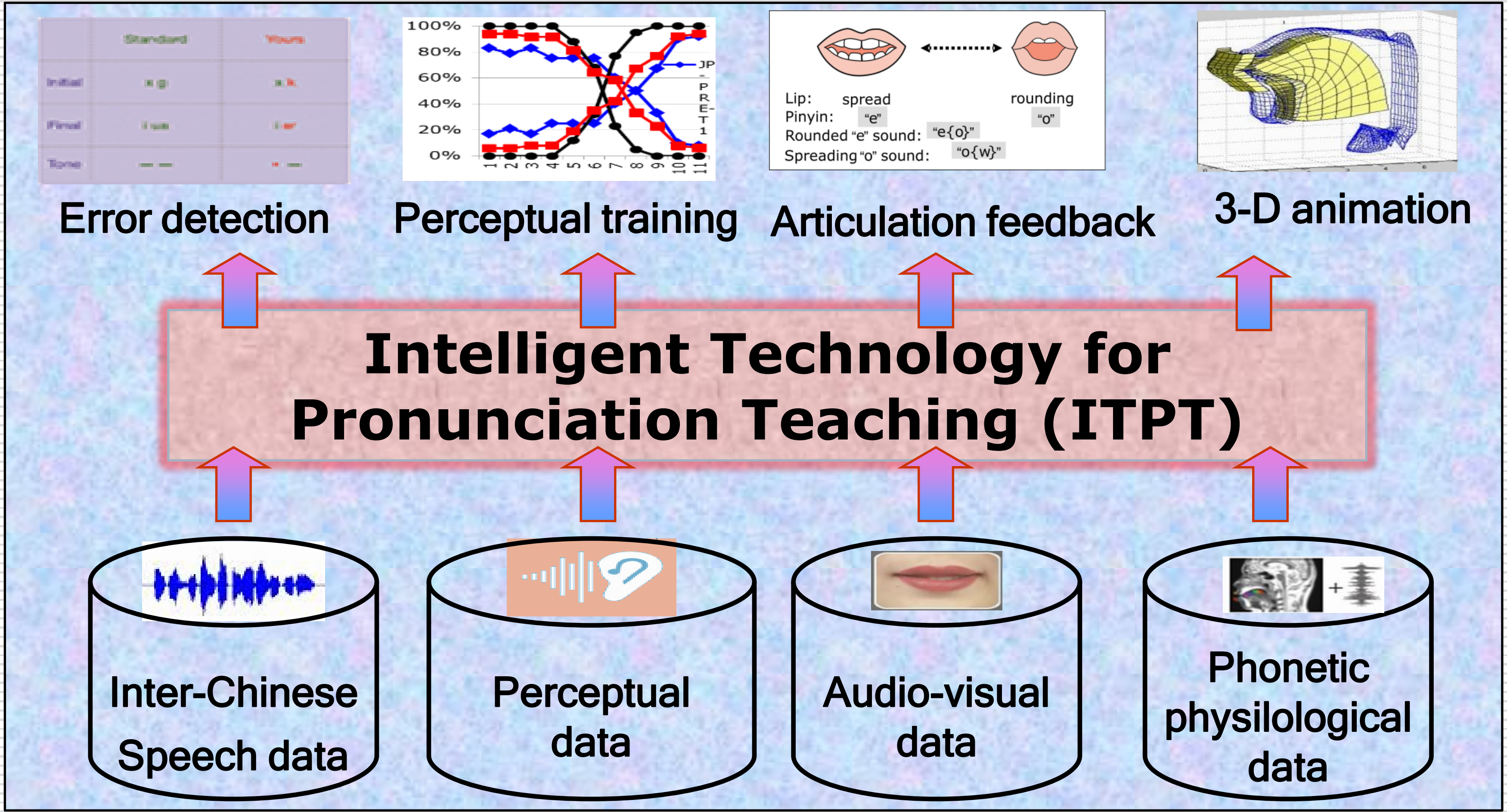
Outline

- □ The purpose of the database
- □ Feature descriptions
- □ Current status
- □ Conclusion

Conclusion

- □ The Key problem of 2nd language teaching is lack of “unlimited practices and feedbacks”.
- □ ITPT can be a possible solution.
- □ Our proposal is a combination of individual technologies.
- □ Multi-modal interlanguage Chinese Speech database is the basis.
- □ We are still on the way.

An Overview



Error detection

Perceptual training

Articulation feedback

3-D animation

Intelligent Technology for Pronunciation Teaching (ITPT)

Inter-Chinese Speech data

Perceptual data

Audio-visual data

Phonetic physilological data