



An Empirical Study of Tone Mispronunciation
Detection Methods for Mandarin CAPT Applications

實證探究聲調辨識技術於電腦輔助
華語發音訓練之應用

¹ Hsiao-Tsung Hung (洪孝宗)

² Yuwen Hsiung (熊玉雯)

¹ Berlin Chen (陳柏琳)

¹ Dept. of Computer Science & Information Engineering &

² Center of Learning Technology for Chinese,
National Taiwan Normal University

Outline

- Introduction
- Problem Formulation
- Pitch Contour (Fo) Normalization
- Tonal Feature Extraction
- Empirical Experiments
- Conclusion and Future Work



Introduction:

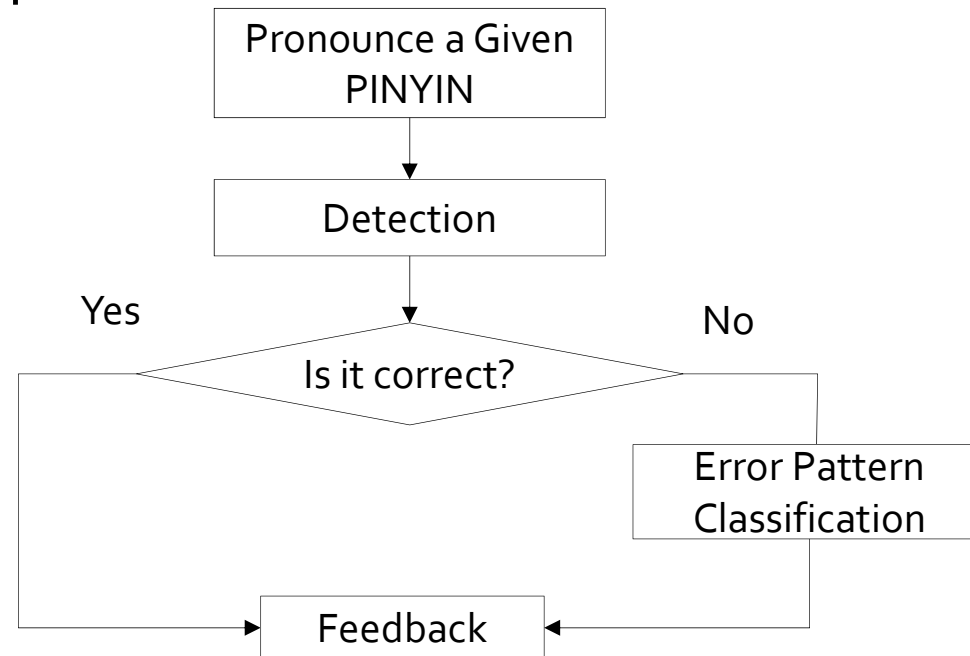
Applying ASR Techniques for CAPT

- CAPT: Computer-Assisted Pronunciation Training
 - Task of applying ASR for CAPT : automatically detect pronunciation errors and evaluate pronunciation quality
 - Facets for CAPT in Mandarin Chinese
 - Lexical Tones
 - Pronunciation Scoring for Sub-word Units (syllable, INITIAL/FINAL)
 - Fluency/Proficiency (Duration/Speaking Rate)
 - Overall Scoring (word-, phrase- and sentence-levels)
 - In this paper we focus exclusively on developing robust methods for automatic detection of lexical tone pronunciation errors
-

Introduction:

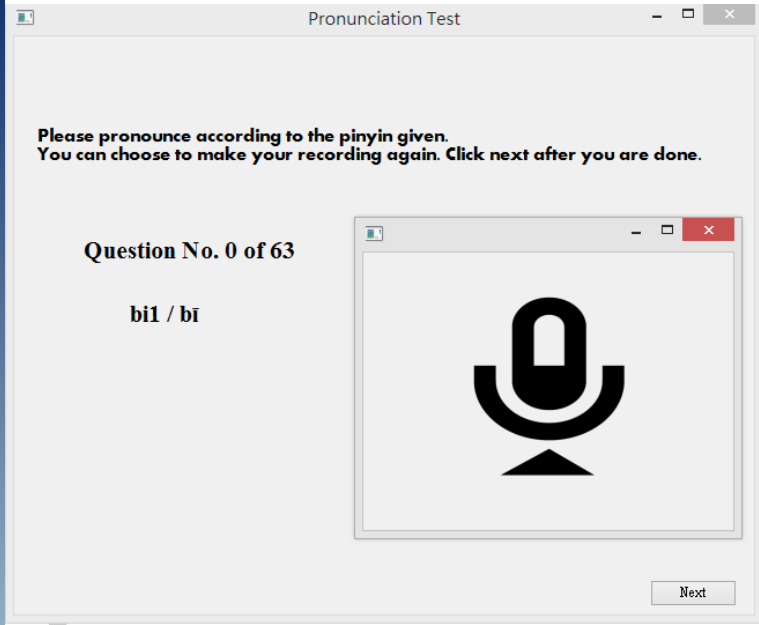
Detection and Assessment of Pronunciations

- Automatic detection and assessment of lexical tone and syllable pronunciations
 - This paper focuses exclusively on the detection and assessment of 4 lexical tones (Tone 1- Tone 4)
- Schematic Depiction:



Introduction: Recently Developed Prototype System

- For example, in the test phase, the system can detect possible mispronunciations and corresponding error patterns of the user



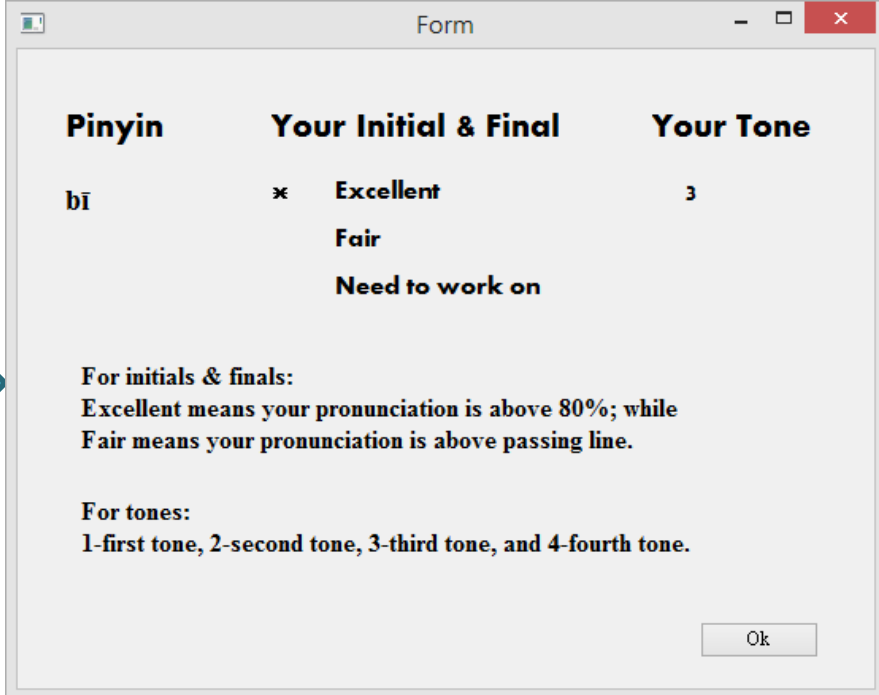
Pronunciation Test

Please pronounce according to the pinyin given.
You can choose to make your recording again. Click next after you are done.

Question No. 0 of 63

bi1 / bi

Next



Form

Pinyin	Your Initial & Final	Your Tone
bi	✘ Excellent Fair Need to work on	3

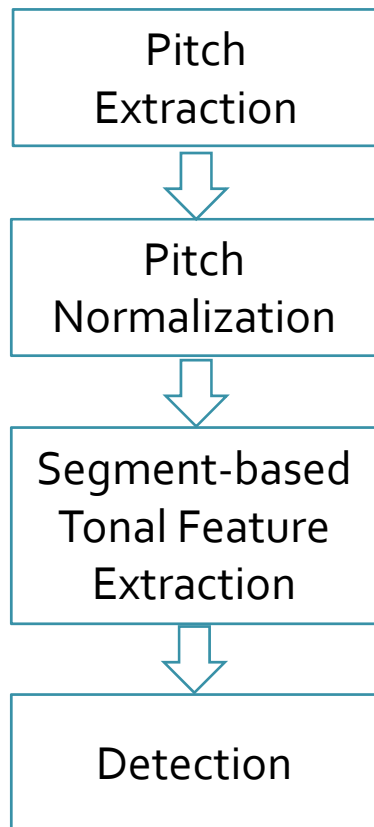
For initials & finals:
Excellent means your pronunciation is above 80%; while
Fair means your pronunciation is above passing line.

For tones:
1-first tone, 2-second tone, 3-third tone, and 4-fourth tone.

Ok

Problem Formulation of Lexical Tone Detection

- Typical Steps for Lexical Tone Detection



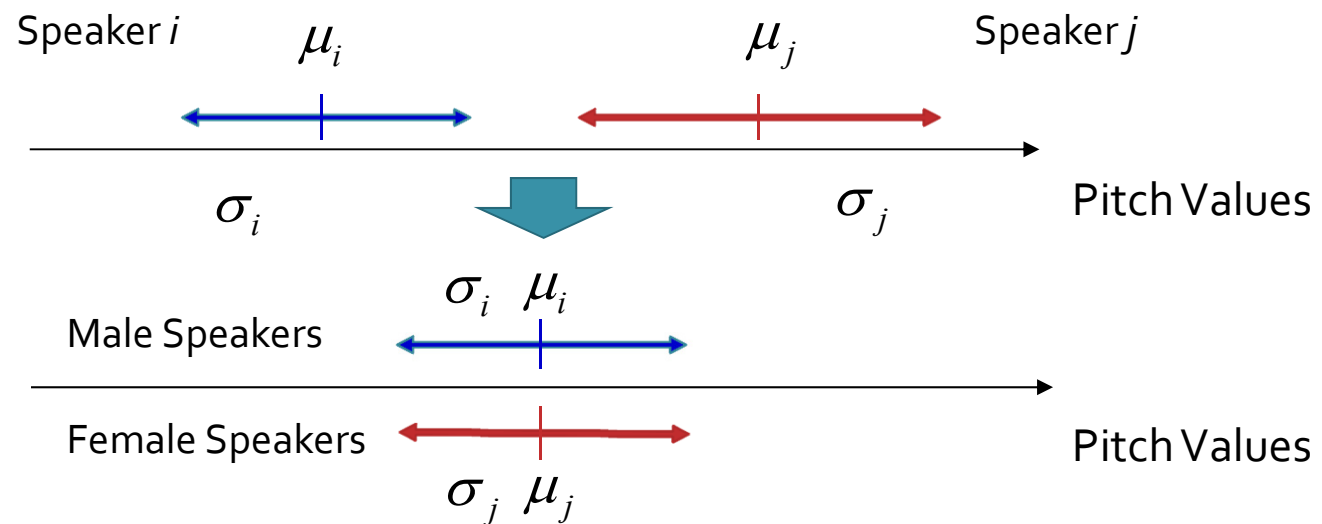
How to mitigate the negative effects caused by speaker and environmental variations?

How does the subtleness (granularity) of tonal features affect mispronunciation detection?

MVN-Based Pitch Contour Normalization

- Mean Variance Normalization (MVN) in a speaker-wise manner
- Can deal with monotonic and linear distortions/variations
- Assume a linear transformation existing between the original and normalized pitch contours

$$y = T[x] = \alpha x + \beta$$



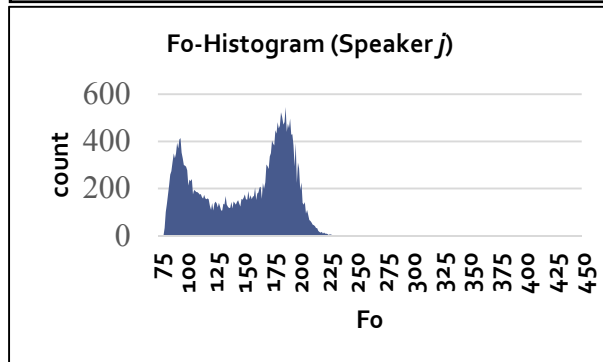
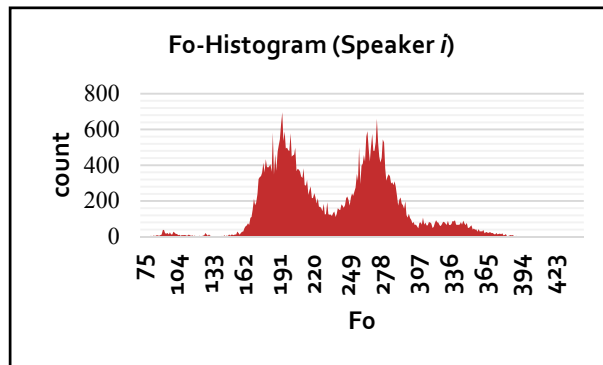
CDF-based Pitch Contour Normalization

- Can deal with monotonic and non-linear distortions/variatioins
- CDF-Matching (or Histogram Equalization)

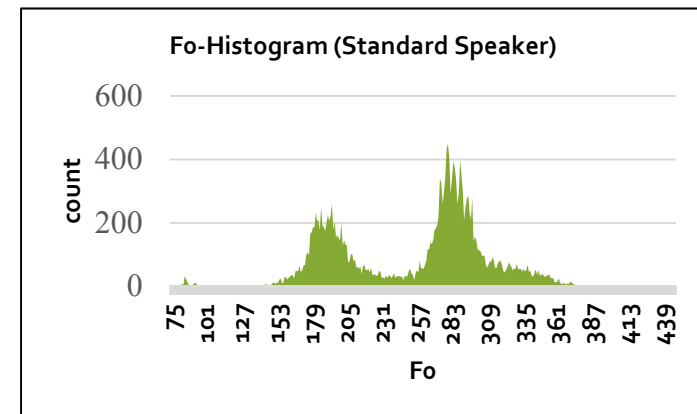
$$y = T[x] = C_Y^{-1}(C_X(x)),$$

where C is a cumulative distribution function

- Schematic Illustration



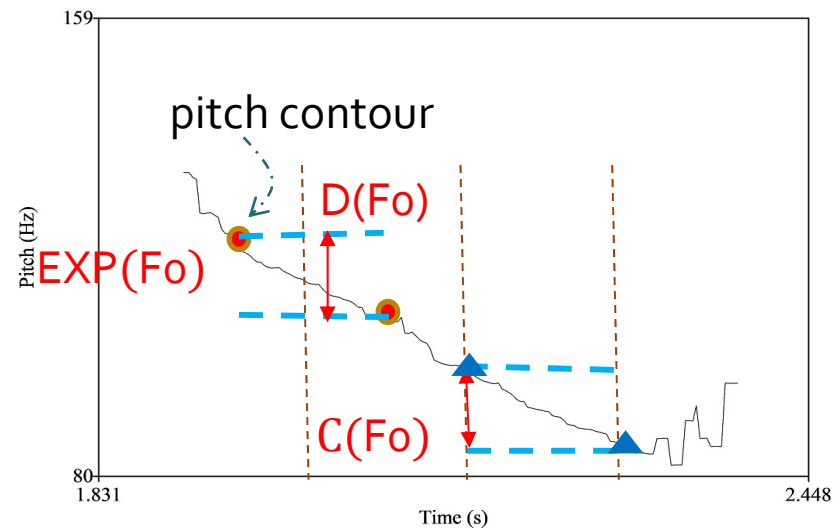
CDF
Normalization



Segment-based Tonal Features

Types of Features

EXP[Fo]	Mean Fo in each segment
C(Fo) (Within-Segment Δ Fo)	Difference of beginning and ending Fo values within each segment
D(Fo) (Between-Segment Δ Fo)	Difference of EXP[Fo] values between any pair of segments



Experimental Setup

- Some Statistics of Training and Test Sets

	Training Set (2.2 hours)	Test Set (2.3 hours)
Speakers	3 male + 5 female	2 male + 5 female
Tone 1	2,641 / 6,248 syllables	2,137 / 4,184 syllables
Tone 2	1,799 / 7,103 syllables	1,077 / 5,553 syllables
Tone 3	1,533 / 7,747 syllables	1,058 / 5,643 syllables
Tone 4	2,819 / 6,089 syllables	1,986 / 5,599 syllables
Total	8,792 / 27,097 syllables	6,258 / 20,979 syllables

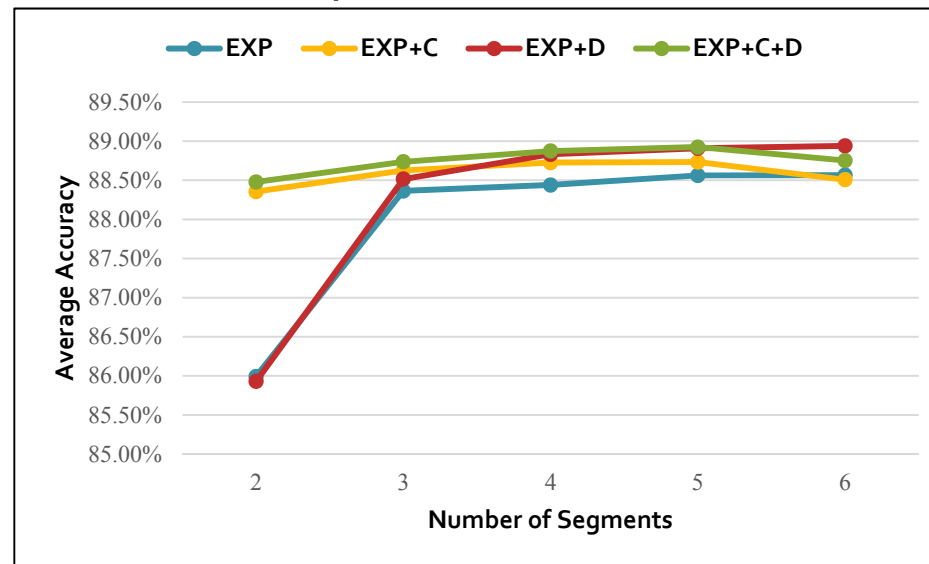
correct pronunciations

mispronunciations

Experimental Results:

1. Different Numbers of Segments for Tonal Feature Extraction

- MVN-based pitch value normalization works in concert with SVM-RBF back-end classifier
- Average accuracy on detecting correct pronunciations and mispronunciations of 4 lexical tones

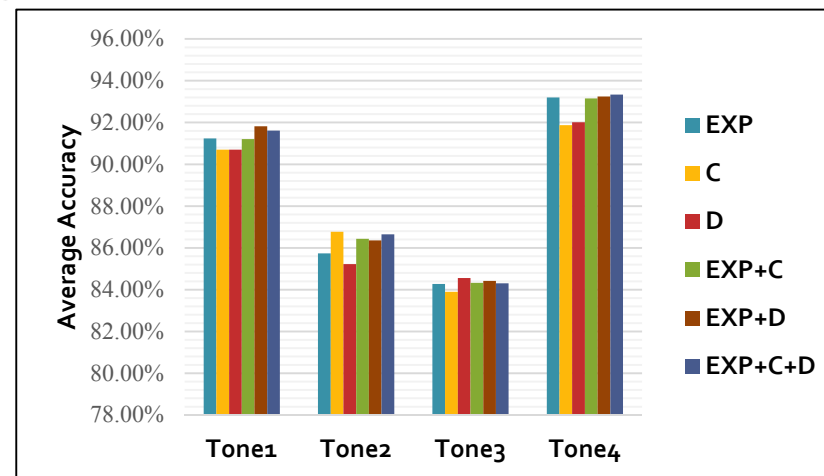


- The performance is improved when the number of the segments becomes larger; the improvements, however, seem to soon reach a plateau when the number of the segments is set to 5

Experimental Results

2. Detailed Performance for 4 Lexical Tones

- MVN-based pitch value normalization in concert with SVM-RBF back-end classifier
- Average accuracy on detecting correct pronunciations and mispronunciations of 4 lexical tones
 - Using 5 Segments for Tonal Feature Extraction

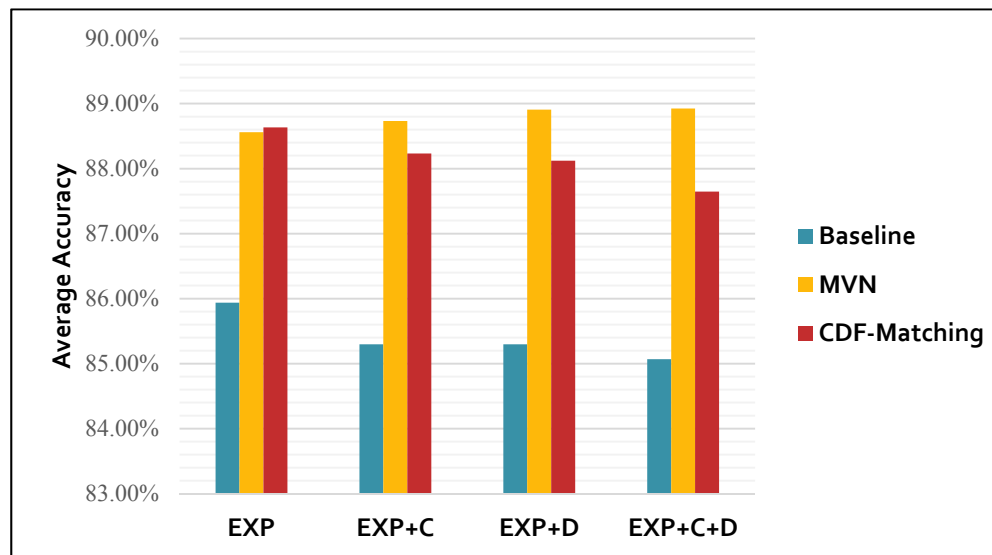


- Mean values "EXP" (i.e., information about pitch levels) are important for lexical tone detection
- On the other hand, the differential statistics (either "C" or "D") are especially beneficial for identifying Tone-1 and Tone-2

Experimental Results:

3. Comparisons Among Two Different Normalization Methods

- Average accuracy on detecting correct pronunciations and mispronunciations of 4 lexical tones
 - Using 5 Segments for Tonal Feature Extraction



- Both methods (MVN and CDF-matching) can offer significant performance boots compared to the baseline (without normalization)

Conclusion & Future Work (1/2)

- How to robustly normalize pitch contours and to determine the subtleness of tonal features are critical to the success of automatic detection of lexical tone pronunciation errors
- As to future work, we would like to apply and extend our methods to automatic pronunciation scoring for sub-word (syllable, INITIAL/FINAL) units and overall pronunciation quality evaluation
- In addition, we are planning to leverage more state-of-the-art machine learning techniques for CAPT in Mandarin Chinese

Conclusion & Future Work (2/2)

- Building a Chinese Learning and Assessment System

結合語音處理技術之華語診斷教學平台
Chinese Learning and Assessment Systems Integrated with Speech Processing Technologies

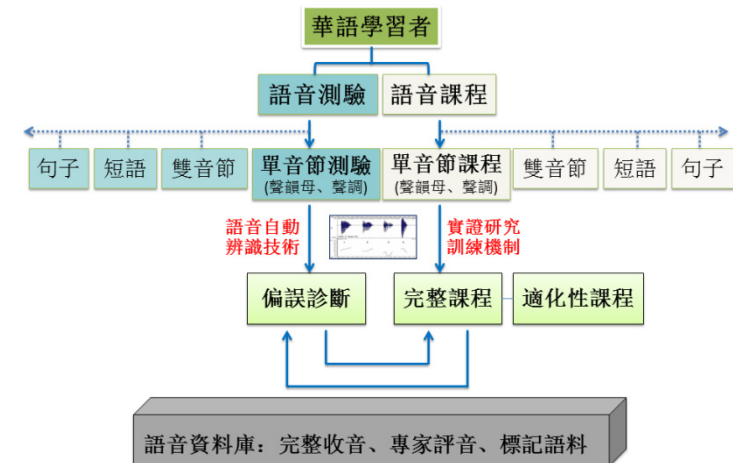
NEWS
系統改版中 101.10.15

INTRODUCTION
This research project aims to build CALL systems that provide Chinese pronunciation assessment and learning programs integrated with the advancing automatic speech processing technologies. We set out to develop several core ingredients: the automatic pronunciation assessment technology that can analyze a L2 Chinese learner's pinyin sounds and tones. [more](#)

Microphone Testing

Chinese Learning and Assessment Systems Integrated with Speech Processing Technologies
Automatic Assessment and Interactive Chinese Listening Classroom

TEL : 886-2-7734-6672 | E-mail : berlin@csie.ntnu.edu.tw | ADD : No.88, Sec. 4, Tingzhou Rd., Wenshan Dist., Taipei City 116, Taiwan (R.O.C.)
COPYRIGHT 2012 NTNU. ALL RIGHTS RESERVED.



Thank You!