Special Session on Forensic Voice Comparison and Forensic Acoustics @ 2nd Pan-American/Iberian Meeting on Acoustics, Cancún, México, 15–19 November, 2010
http://cancun2010.forensic-voice-comparison.net

3pSC4

Accuracy and Precision of Forensic Voice Comparison Using the Chinese /iao/ Triphthong

Cuiling Zhang
Department of Forensic Science and Technology
China Criminal Police University

Geoffrey Stewart Morrison
Forensic Voice Comparison Laboratory
School of Electrical Engineering & Telecommunications
University of New South Wales

中国刑事警察学院
China Criminal
Police University

Introduction

- The formants of vowels reflect the dimensions of the vocal tract and the articulation patterns of speakers.
- Are mean formant frequencies enough?
- Dynamic spectral properties, such as formant trajectories, may be useful for forensic voice comparison.
- But how valid are they?

Introduction

- The traditional approach to the use of formants in forensics voice comparison has been via visual inspection of spectrographs – the so called “voiceprinting” approach.
- This approach lacks objectivity and demonstrated validity.

Introduction

- The current study evaluates the validity of measurements of static and dynamic formant features for forensic voice comparison using acoustic measurements and statistical calculation of likelihood ratios.

Likelihood ratio framework

- The rational evaluation of the strength of evidence with respect to the competing prosecution and defense hypotheses.
- It has been applied in forensic comparison science including DNA and voices.

Likelihood ratio framework

$$LR = \frac{p(E_{fsr} \mid H_{ss})}{p(E_{fsr} \mid H_{a})}$$

P： probability

H_{SS}： prosecution hypothesis, assuming same speaker

H_{a}： defense hypothesis, assuming different speakers

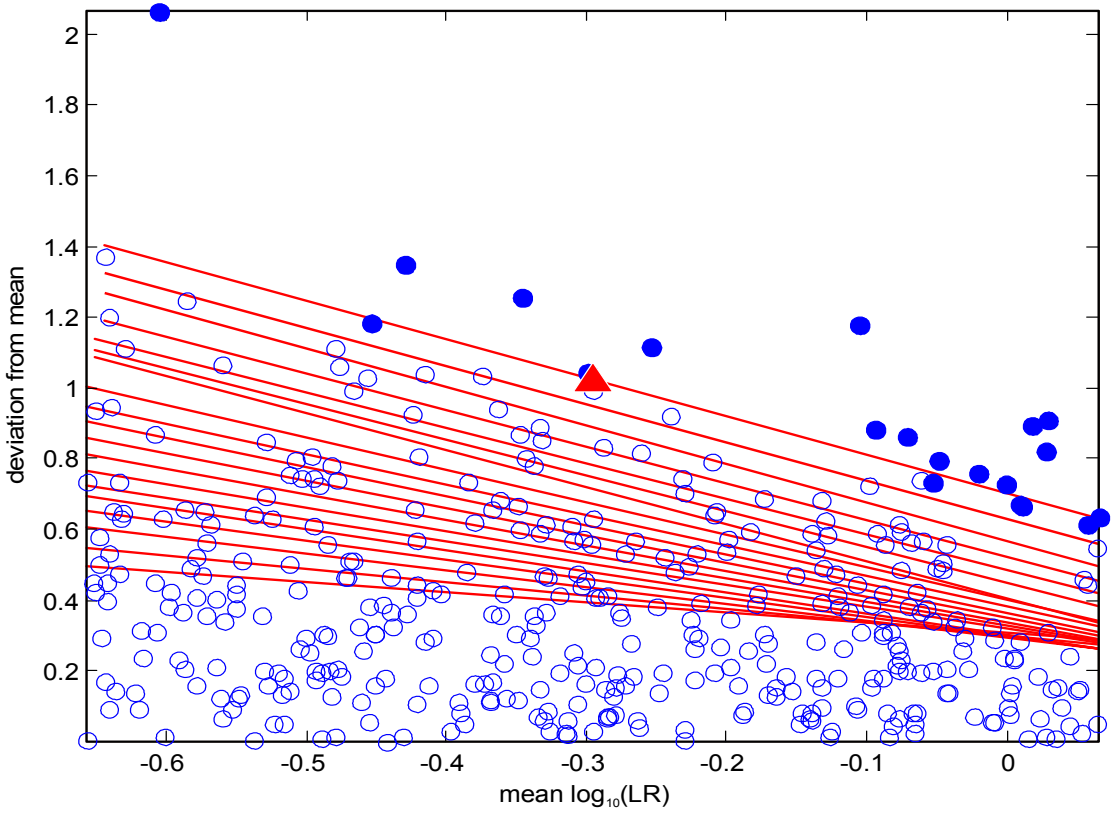
E_{fsr}： forensic voice comparison evidence--difference between suspect & offender speech sample

Validity testing

- Collect lots of pairs of samples from the same speaker speaking on different occasions (SS)
- Collect lots of pairs of samples from different speakers (DS)
- Estimate LRs for SS and DS comparisons
- Good performance is the extent to which LR < 1 for DS and LR > 1 for SS
- Calculate log-likelihood-ratio cost, C_{llr}
  (Brümmer & du Preez, 2006)

Reliability testing

- 95% Credible Interval calculated using non-parametric procedure (Morrison, Thiruvaran, & Epps, 2010)
- Calculated using different-speaker pairs only



Materials and methodology

Monophthong Data

- ~10 tokens of each of Chinese /i/, /e/, and /a/ vowels.
- Answers to interview questions recorded over a landline telephone system (speaker in quiet room).
- 64 male speakers, students from China Criminal Police University, 19~23 years old, spoke Standard Chinese, without apparent accent or speech impediments.
- Each speaker was recorded on two separate occasions separated by approximately two weeks.
- Sampling rate of 16KHz

Materials and methodology

Triphthong Data

- 20 tokesn of Chinese /iao/ vowel.
- Read-speech (phrases) studio recordings
- 20 female speakers, students from China Criminal Police University, 21~24 years old, spoke Standard Chinese, without apparent accent or speech impediments.
- Each speaker was recorded on two separate occasions separated by approximately two weeks.
- Sampling rate of 16KHz

Materials and methodology

Acoustic and Statistical Analysis

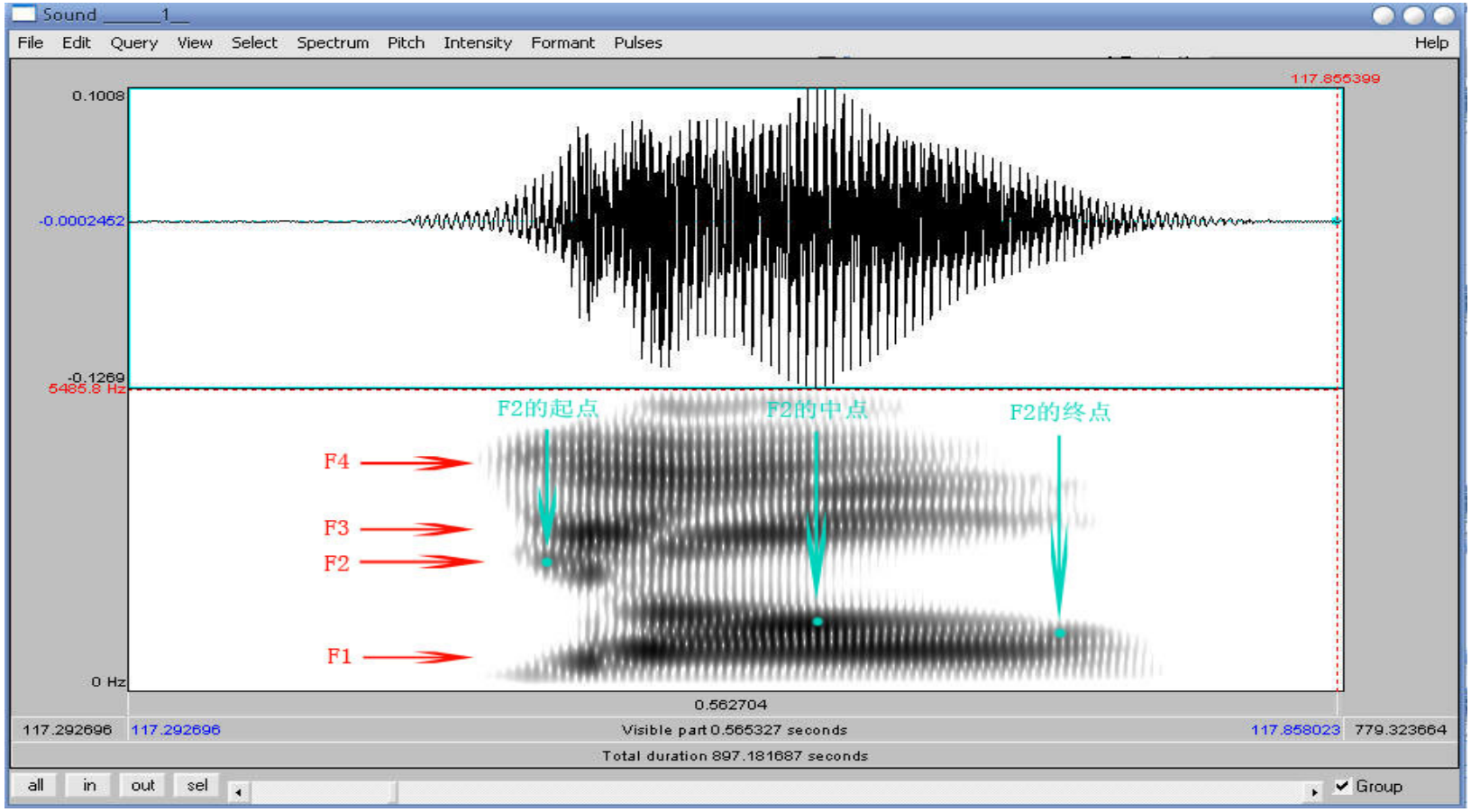
- Vowel tokens acoustically analyzed using *Praat*.
- F1,F2, F3 measured using LPC
  - mean value for monophthongs
  - at three points (start, middle and end) for tripthong
- Likelihood ratios calculated using the multivariate-kernel-density method (MVKD) (Aitken & Lucy, 2004)
- Likelihood ratios calculated for all possible same-speaker and different-speaker pairs in the data set

Materials and methodology

Acoustic and Statistical Analysis

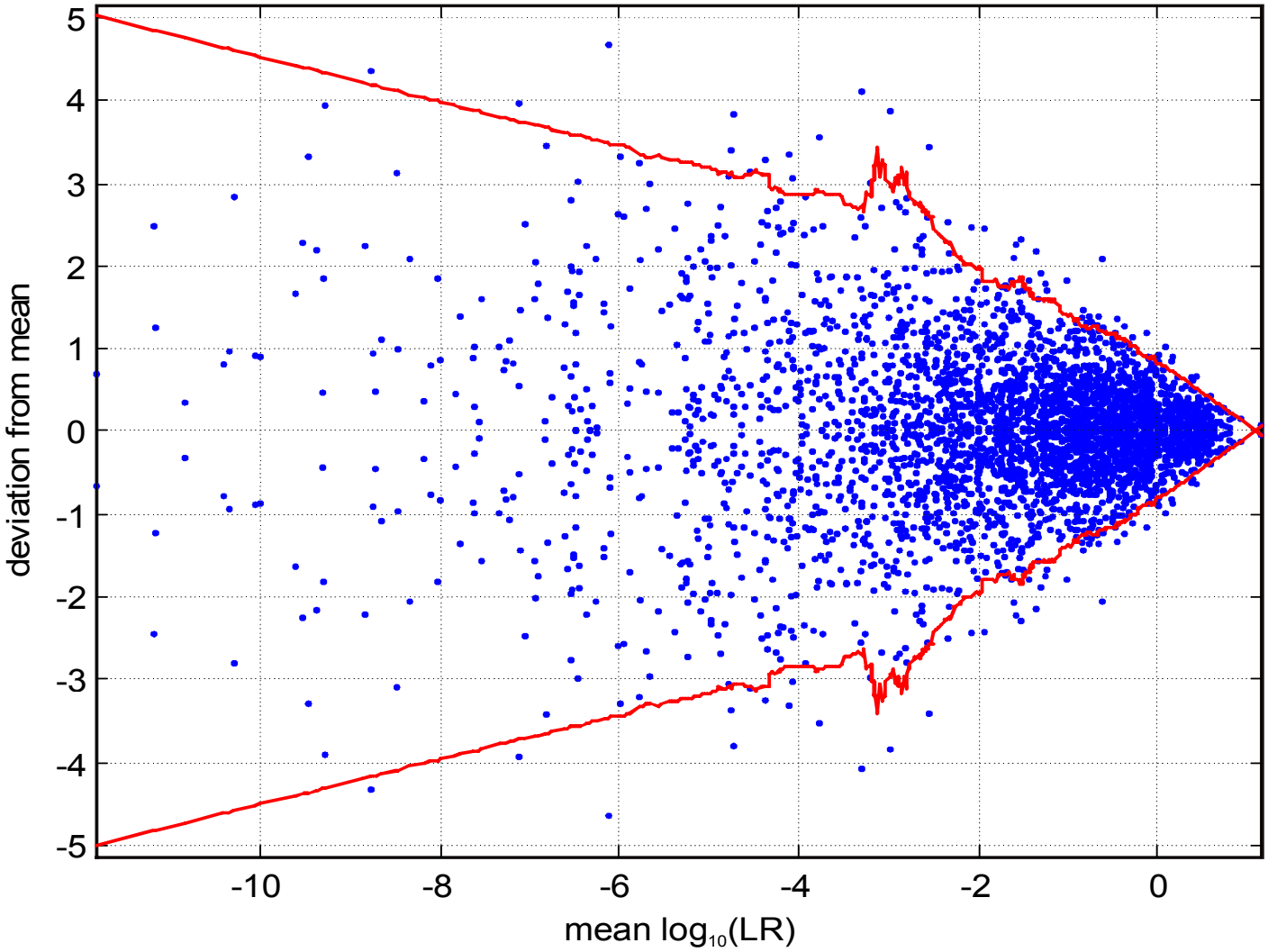
- Results from three monophthongs fused using logistic regression
- Validity of fused-monophthong and triphthong results assessed using the log-likelihood-ratio cost (C_{llr}).
- Reliability of monophthong results assessed using estimate of 95% Credible Interval

Measurements



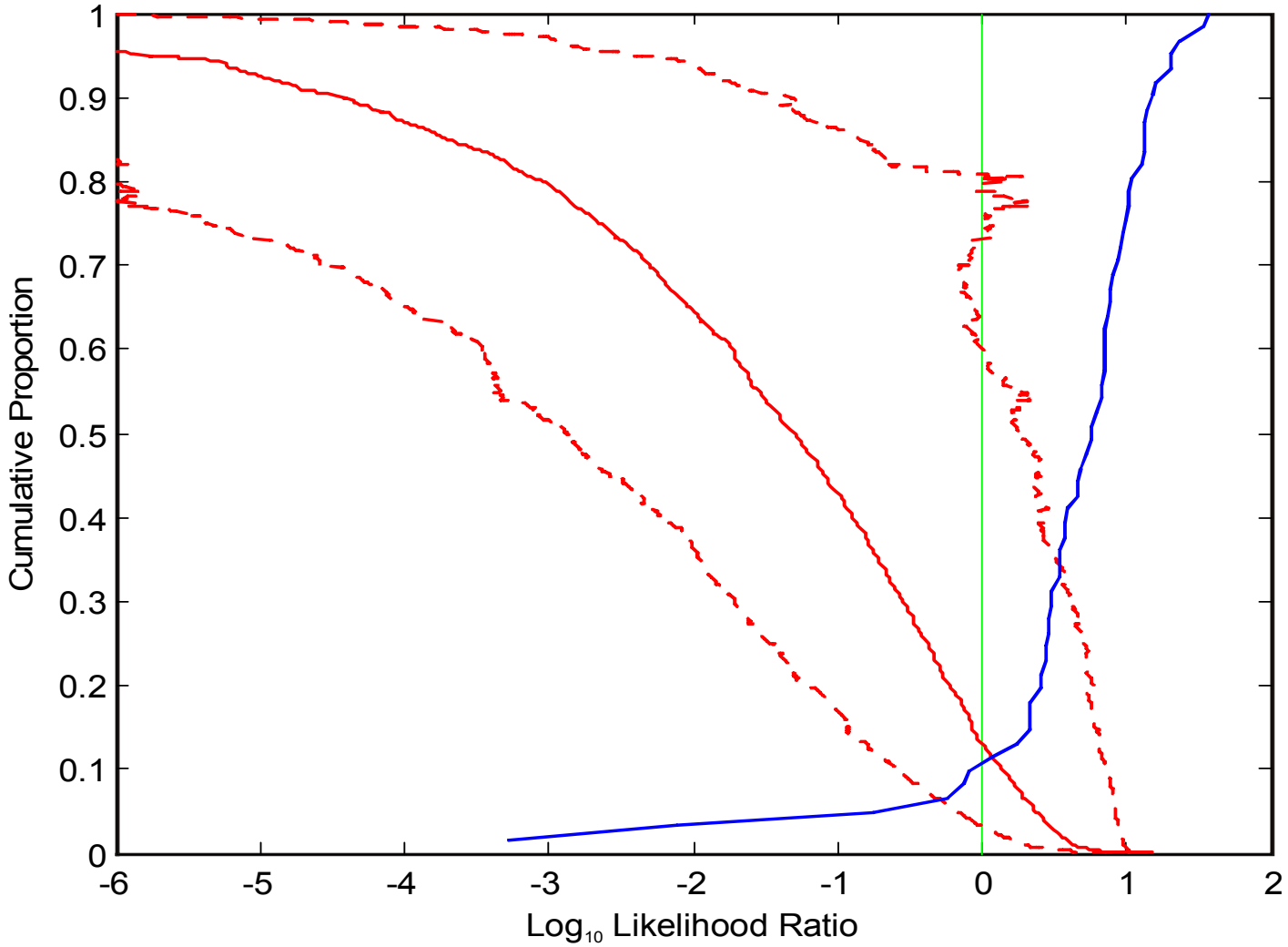
First three formant frequencies at three points (start, middle and end) of /iao/

Results: fused /i/, /e/, /a/



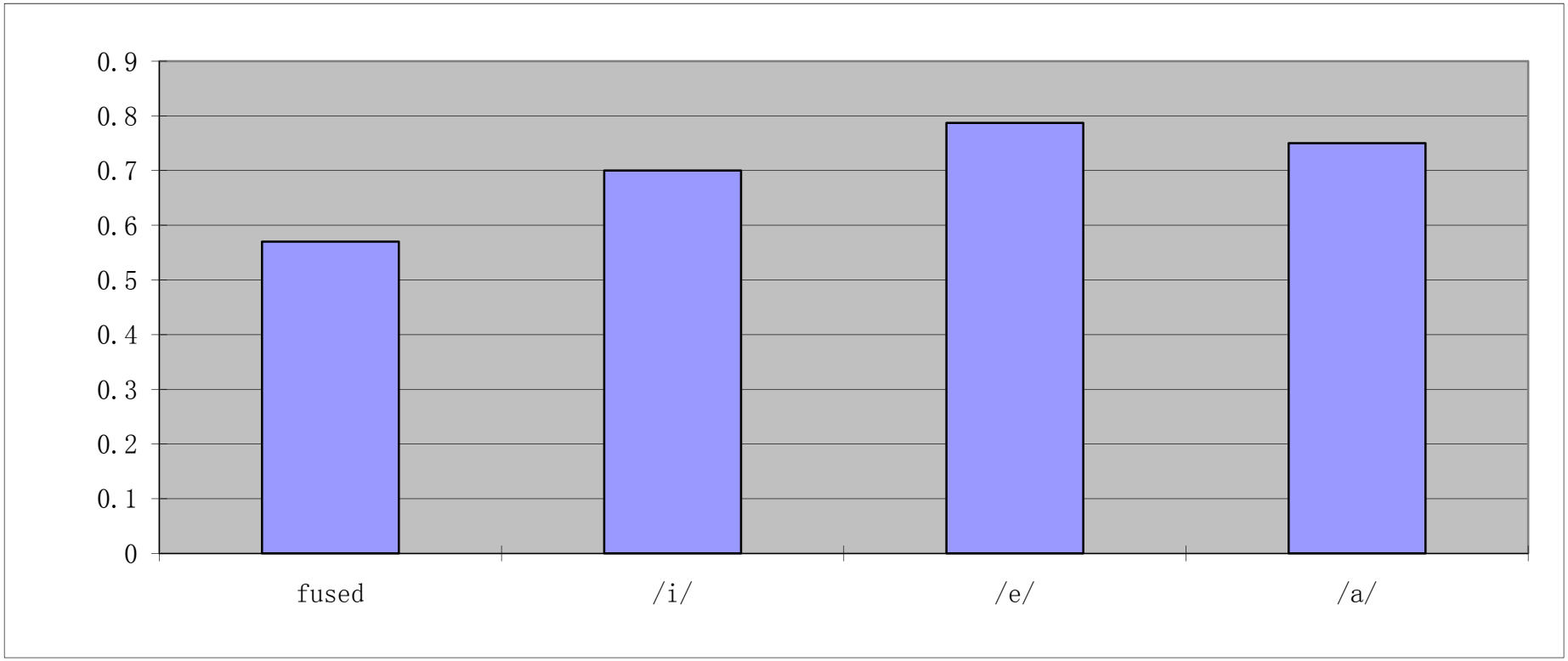
Results: fused /i/, /e/, /a/

C_{llr} = 0.570



Comparison between individual monophthongs and fused system

C_{llr} results for /i/, /e/, /a/

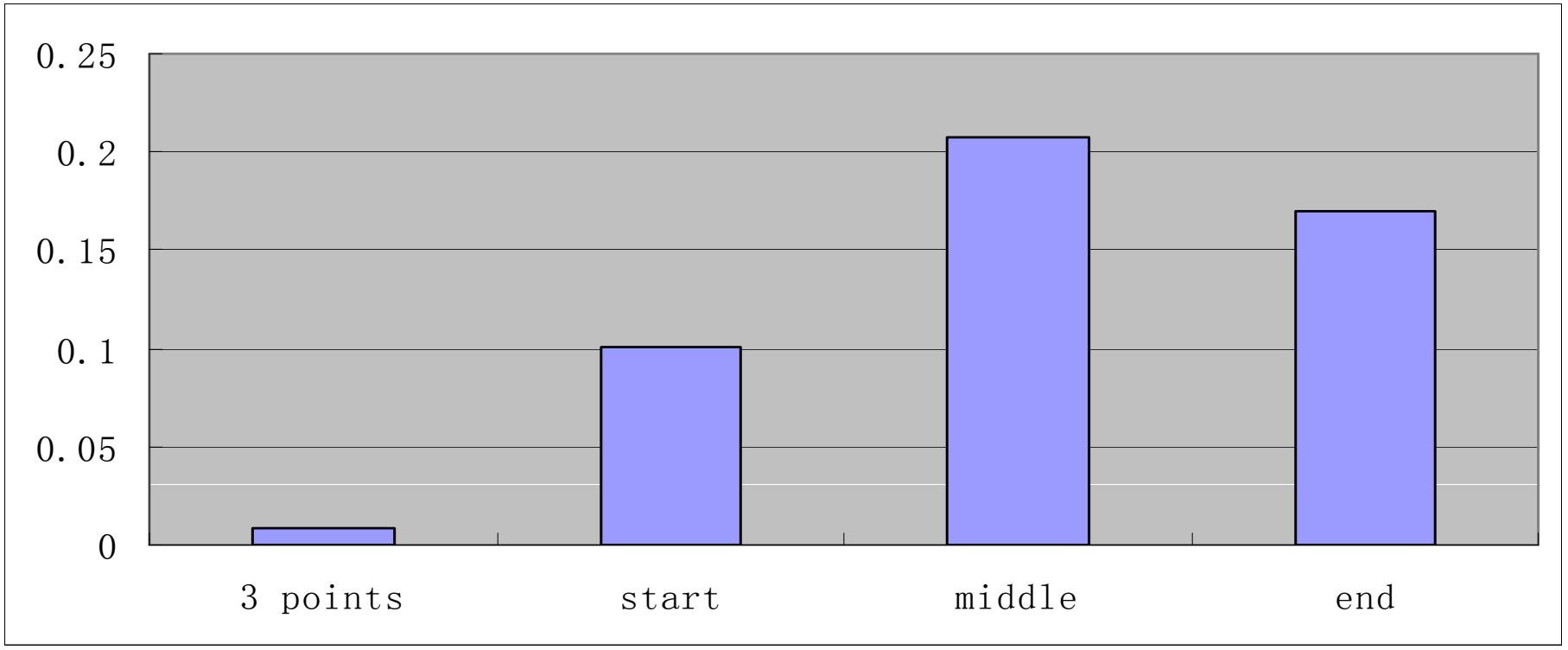


Results: /iao/

C_{llr} = 0.008

Comparison between 3-points fused and each individual point

C_{llr} results for /iao/



Conclusions

- Results from static formant measurements of monophthongs were poor
- Preliminary results from dynamic formant measurements of triphthong are promising
- This preliminary results for /iao/ were based on read speech recorded under laboratory conditions. Different speech styles, recording conditions, and techniques for measuring formant dynamic should also be tested.

Further study

- Bigger size of database (IAFPA project，60 female speakers)
- More vowels (diphthongs and triphthongs)
- More parameters (formants, F0 of tones)
- More realistic data
- More effective formant measurement method (formant tracker）
- Curve fitting representation of formant trajectories
- Comparison of statistical models for LR calculation (MVKD v GMM-UBM)