Nasal spectra for Forensic Voice Comparison

Ewald Enzinger, Cuiling Zhang

FORENSIC VOICE COMPARISON LABORATORY
SCHOOL OF ELECTRICAL ENGINEERING & TELECOMMUNICATIONS
THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY • AUSTRALIA

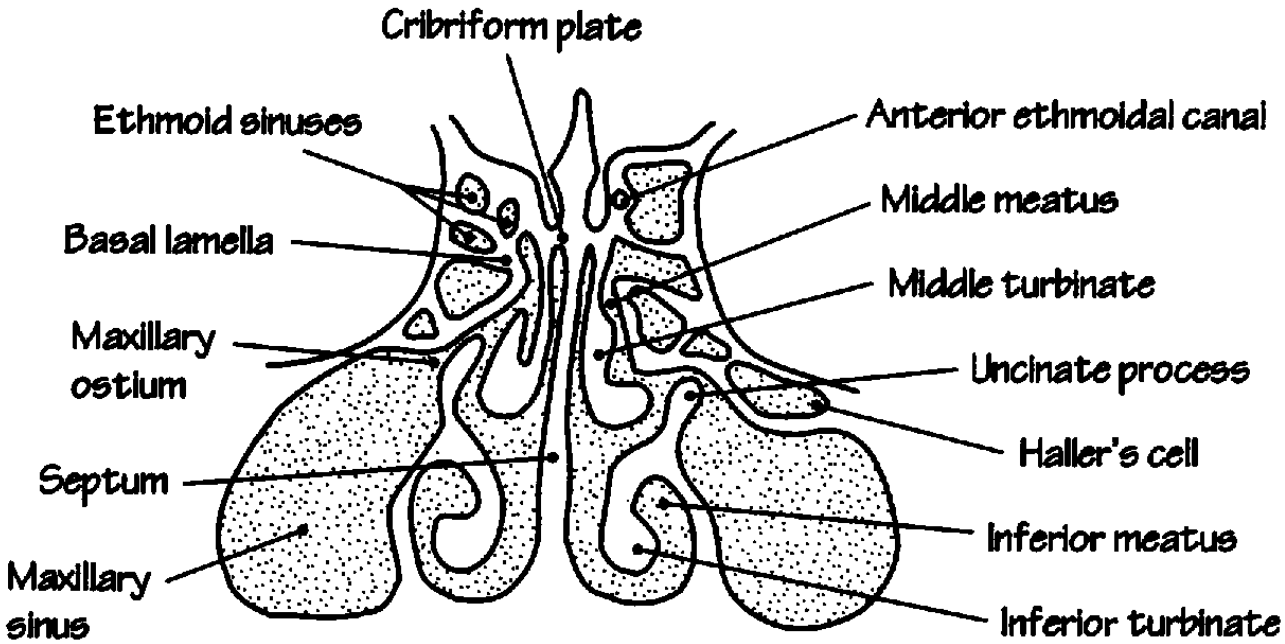
162nd ASA Meeting, San Diego, California
4aSCa5

Acknowledgement of Funding

- Data collection was funded by an International Association of Forensic Phonetics and Acoustics (IAFPA) Research Grant.
- Analysis was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL). All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U.S. Government.
- Presentation supported by the Australian Research Council, Australian Federal Police, New South Wales Police, Queensland Police, National Institute of Forensic Science, Australasian Speech Science and Technology Association, and the Guardia Civil through Linkage Project LP100200142. Unless otherwise explicitly attributed, the opinions expressed are those of the authors and do not necessarily represent the policies or opinions of any of the above mentioned organizations.

Why nasal consonants for FVC?

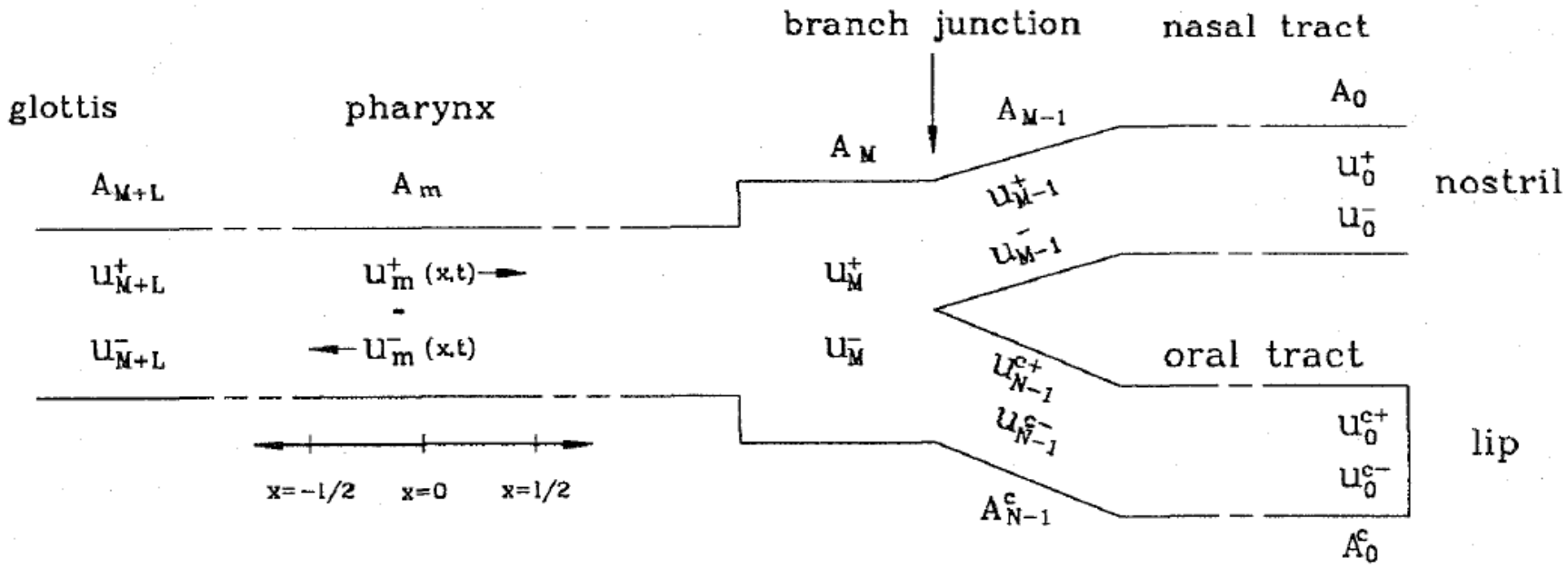
- Relatively fixed nasal and paranasal cavities
  ➡ potentially low within-speaker variability
- Complicated structure of nasal cavity
- Asymmetries in paranasal cavities (sinuses)
  ➡ potentially high between-speaker variability



(Stevens, 1999, p.19)

Representations of nasals

- Branch in oral and nasal cavity as well as sinuses cause zeros

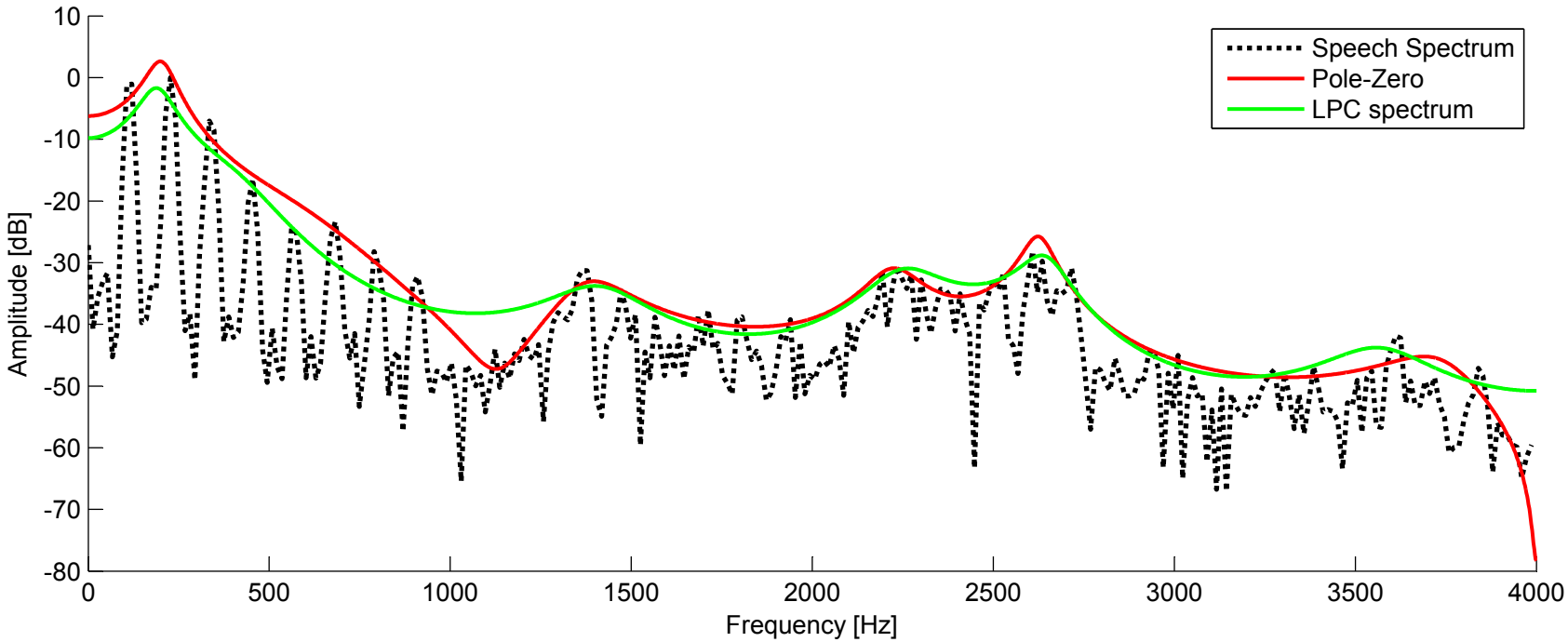


(Lim & Lee, 1996)

Representations of nasals

- Transfer function is adequately represented by a Pole-Zero model

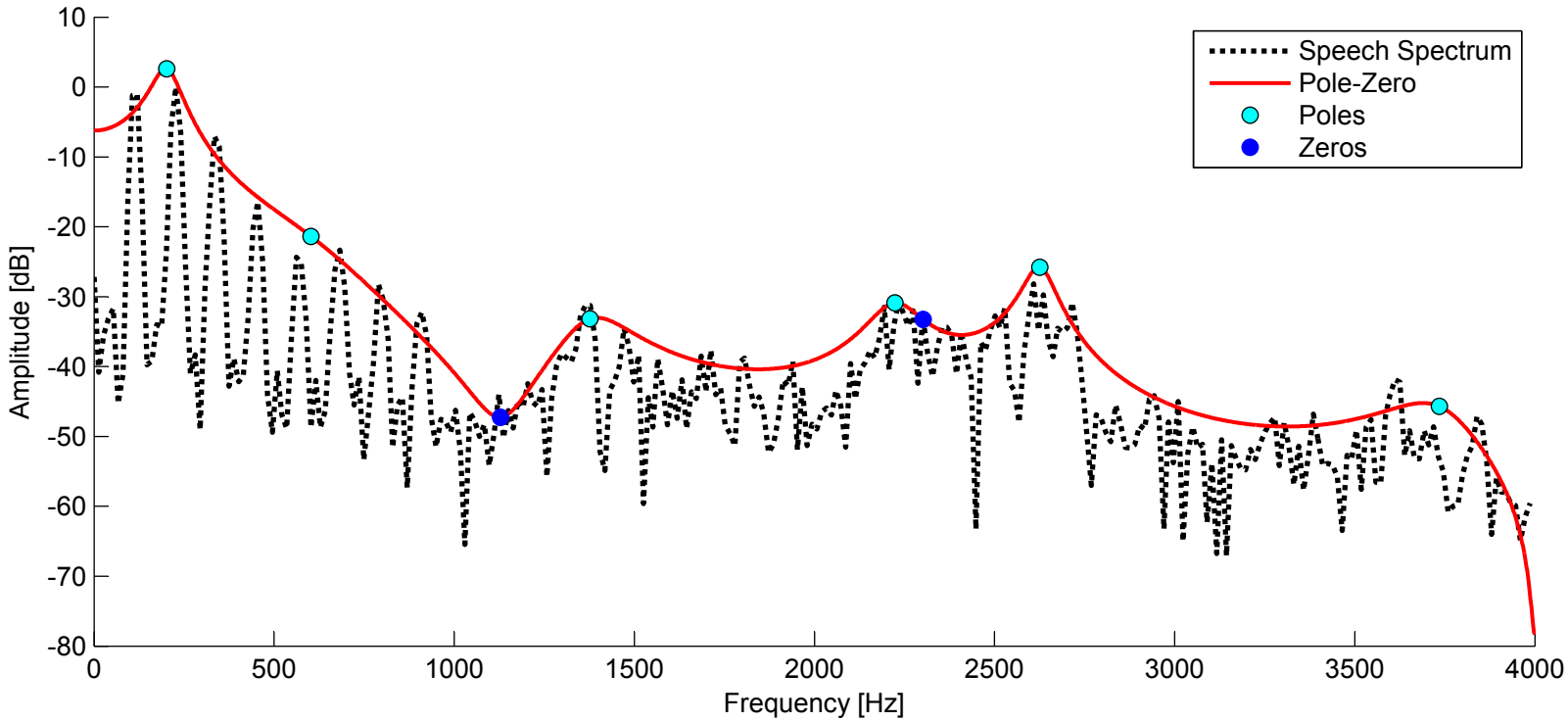
$$G(z,\theta) = \frac{B(z,\theta)}{A(z,\theta)} = \frac{\sum_{l=0}^{n} b_l z^{-l}}{\sum_{l=0}^{m} a_l z^{-l}}$$



Representations of nasals

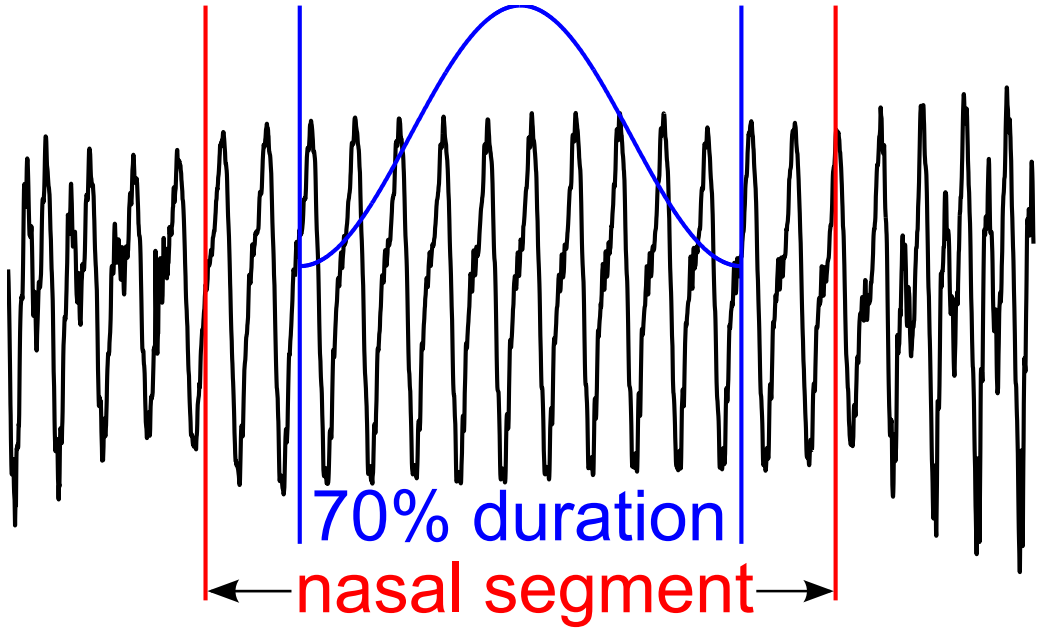
- Cepstral features
  - Mel-frequency cepstral coefficients (MFCCs)
  - Linear Prediction cepstral coefficients (LPCC)
  - Pole/Zero cepstral coefficients (PZCC)
- Angular positions of roots of numerator/denominator polynomials

(Enzinger et al., 2011, Enzinger & Balazs, 2011)



Representations of nasals

- Unsupervised (automatic) feature extraction
- Features obtained from window of 70% segment duration



Likelihood ratio calculation

- Multi-variate kernel density (Aitken & Lucy, 2004)
- Logistic-regression calibration and fusion
- MFCC-based GMM-UBM system
  ➡ Entire speech-active portion of recording

Data

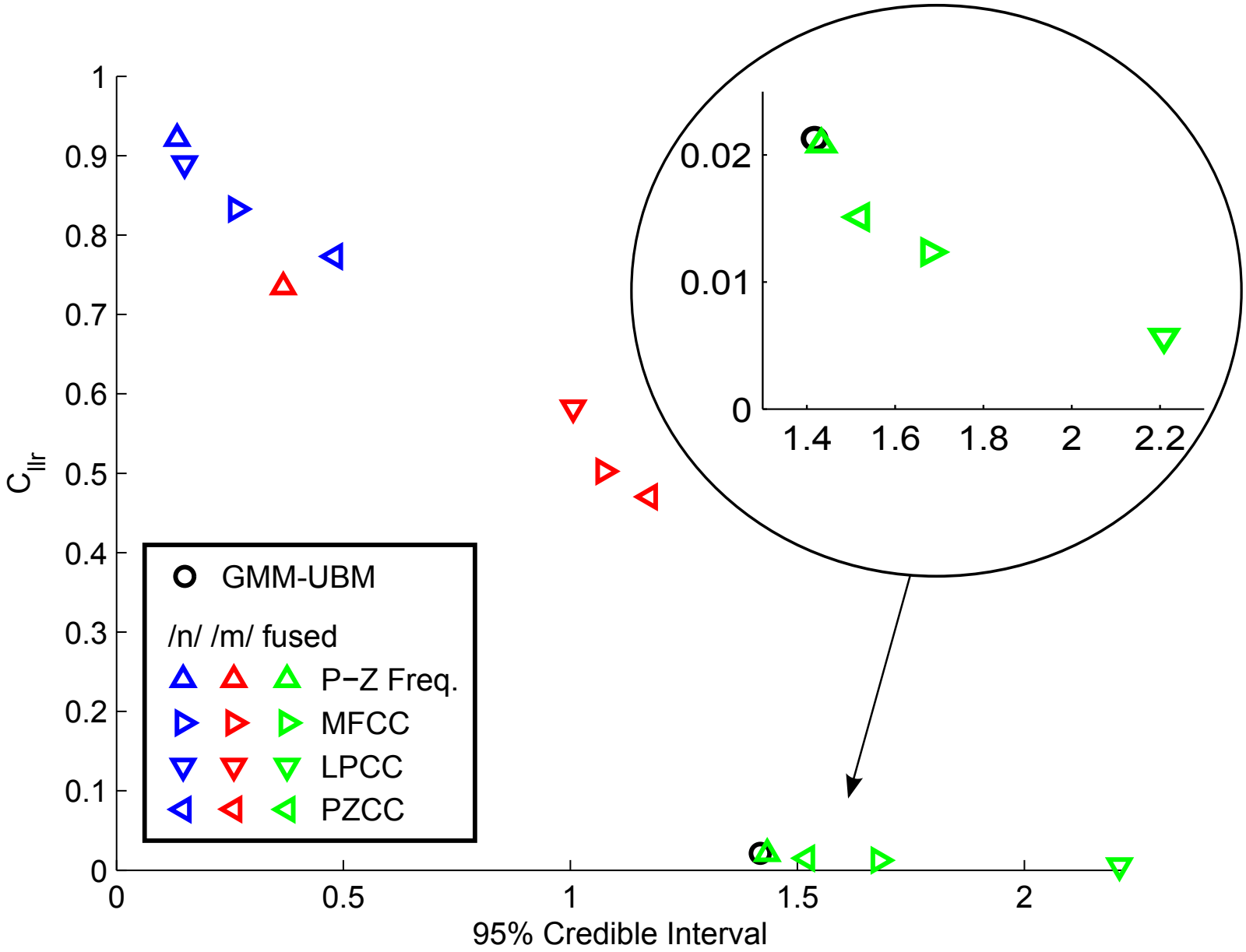
- 60 female Standard Chinese speakers
- Split into 3 groups of 20 speakers
  - background database
  - development set
  - evaluation set
- Information-exchange task over the telephone
- Two recording sessions separated by 2–3 weeks
- High quality data
- /n/ and /m/ tokens analyzed

`http://databases.forensic-voice-comparison.net/`

Evaluation measures

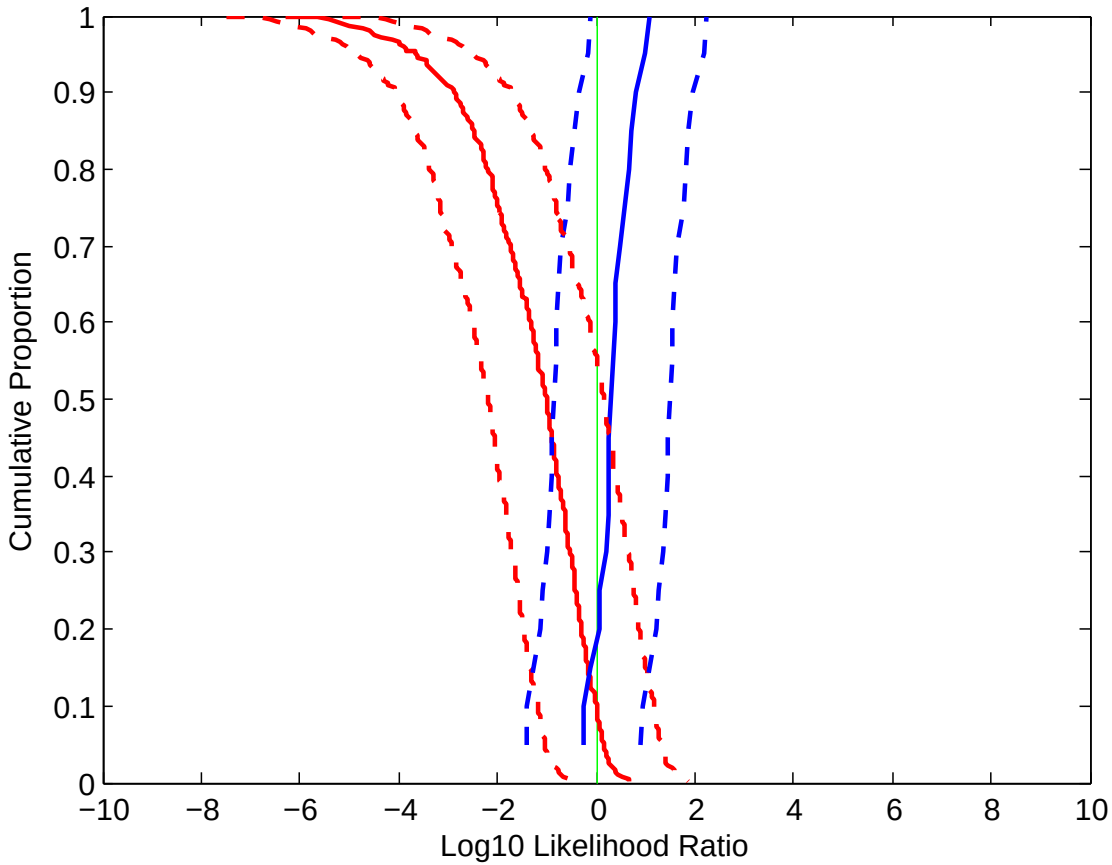
- Validity / Accuracy
  - Log-likelihood ratio cost (C_{llr}) metric
- Reliability / Precision
  - 95% credible interval (Morrison, 2011)
  - Parametric estimation method

Results


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
C_{llr}
0
0.5
1
1.5
2
95% Credible Interval
GMM-UBM
/n/ /m/ fused
P−Z Freq.
MFCC
LPCC
PZCC
0.02
0.01
0
1.4
1.6
1.8
2
2.2

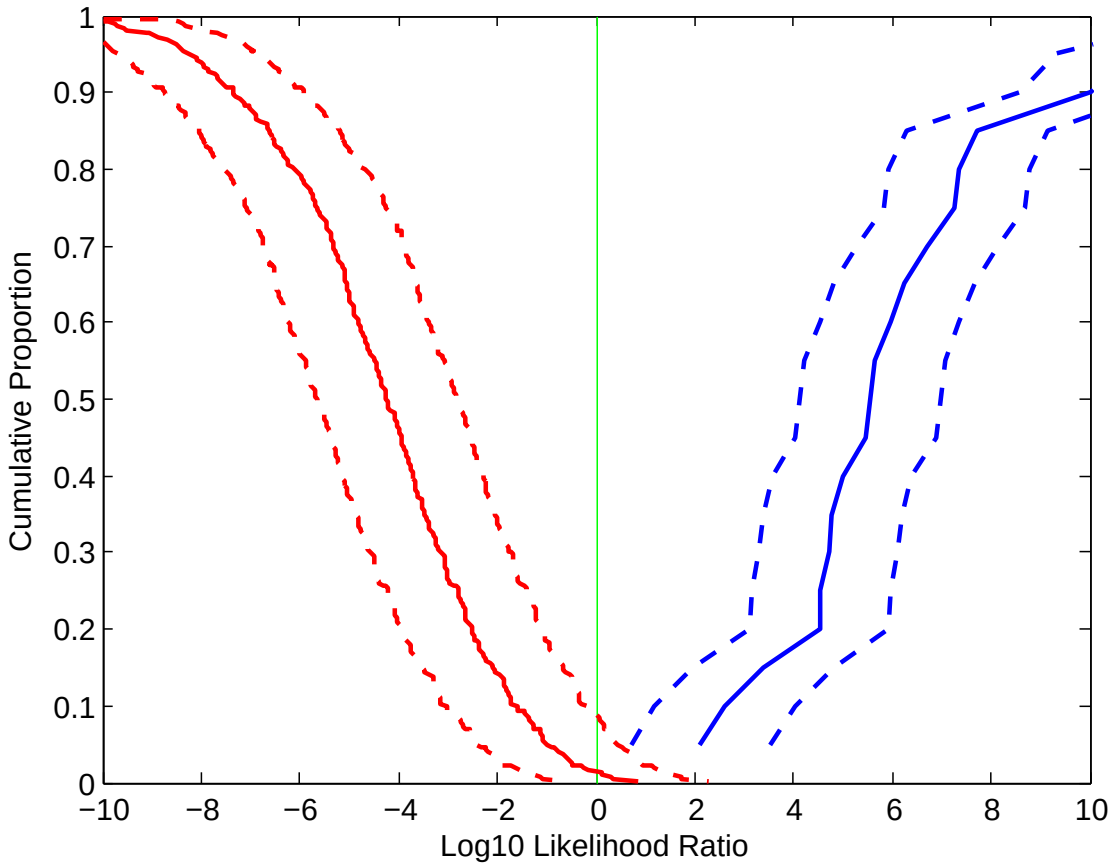
Results

PZCC /m/



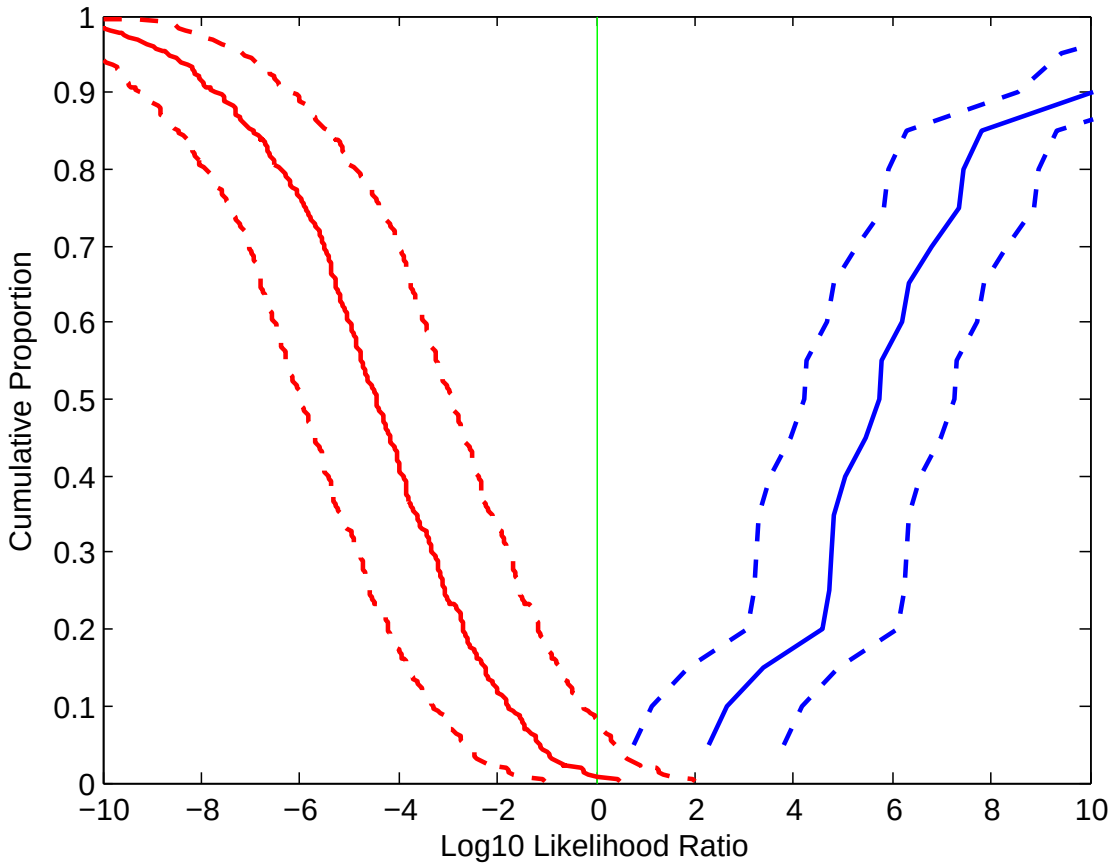
Results

Baseline GMM-UBM



Results

Baseline GMM-UBM + PZCC /m/



Conclusion

Pole/Zero features for nasals

- Theoretically well motivated
- Poor performance of /n/ compared to /m/
  - Low number of tokens per speaker
- Fusion with baseline increases accuracy with a loss in precision

Open issue: Channel mismatch

- GSM: LPC (all-pole) in Adaptive Multi-Rate codec

Thanks!

References

Aitken, C. G. G., & Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data. *Applied Statistics*, *53*(1), 109–122.

Enzinger, E., & Balazs, P. (2011). Speaker Verification using Pole/Zero Estimates of Nasals. In *Proceedings of the Multi-Conference on Systems & Structures (SysStruc '11)*. Reşiţa, Romania.

Enzinger, E., Balazs, P., Marelli, D., & Becker, T. (2011). A logarithmic based pole-zero vocal tract model estimation for speaker verification. In *Proceedings of the International Conference on Audio, Speech and Signal Processing (ICASSP)*, (pp. 4820–4823).

Lim, I.-T., & Lee, B. G. (1996). Lossy pole-zero modeling for speech signals. *IEEE Transactions on Speech and Audio Processing*, *4*(2), 81–88.

Morrison, G. S. (2011). Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice*.

Stevens, K. N. (1999). *Acoustic Phonetics*. Cambridge, MA: MIT Press.