Effects of telephone transmission on the validity of formant-trajectory-based forensic-voice-comparison systems

Cuiling Zhang
Ewald Enzinger
Geoffrey Stewart Morrison
Felipe Ochoa

FORENSIC VOICE COMPARISON LABORATORY
SCHOOL OF ELECTRICAL ENGINEERING & TELECOMMUNICATIONS

UNSW
THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY • AUSTRALIA

Acknowledgment of Funding

- International Association of Forensic Phonetics and Acoustics (IAFPA) Research Grant.
- 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.
- The China Scholarship Council State-Sponsored Scholarship Program for Visiting Scholars.
- The Ministry of Education of the People's Republic of China "Program for New Century Excellent Talents in University" (NCET-11-0836).
- Other
- 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.

Research Questions

- Formant measurement is very common in acoustic-phonetic forensic voice comparison.

- The speech on offender recordings has often been transmitted through a telephone system.

- What is the effect of telephone transmission on the validity of forensic-voice-comparison systems based on:
  - human-supervised formant-trajectory measurement?
  - fully-automatic formant-trajectory measurement?

- Improvement over baseline fully-automatic MFCC system?

Data

- 60 female speakers of Standard Chinese
  - 20 for background
  - 20 for development
  - 20 for test
- Information-exchange task over the telephone
- Two recording sessions separated by 2–3 weeks
  - 4–5 minutes of speech per speaker per session
- Chinese /iau/ tokens
  - 15–30 tokens per speaker per recording

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

Formant measurement

- Manual marking of /iau/ tokens (CZ)
  - SOUNDLABELLER
- Human-supervised formant tracking (CZ)
  - FORMANTMEASURER
- Fully-automatic formant tracking
  - WAVESURFER
  - PRAAT
  - Nearey, Assmann, Hillenbrand (2002) [NAH2002]
  - Mustafa, Bruce (2006) [MB2006]
  - Rudoy, Spendley, Wolfe (2007) [RSW2007]

Forensic-voice-comparison systems

- Formant-trajectory systems
  - F2 and F3 trajectories of /iau/ tokens
  - discrete cosine transform (DCT)
  - zeroth through fourth coefficients
  - multivariate kernel density (MVKD) formula

- Baseline MFCC + Δ, GMM-UBM
  - entire speech-active portion of recording

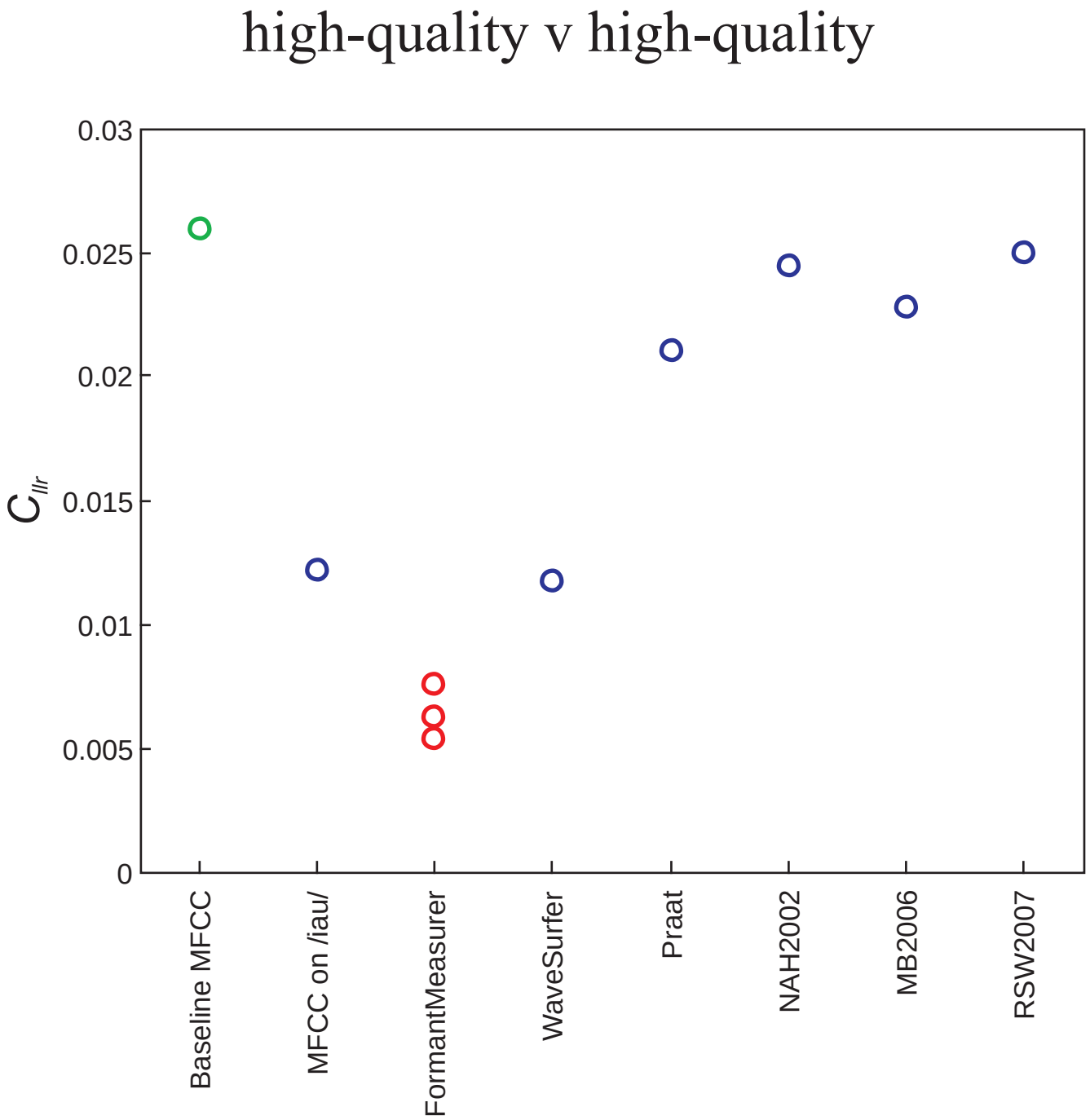
- MFCC + Δ, GMM-UBM
  - on /iau/ tokens only

- Logistic-regression fusion

Testing

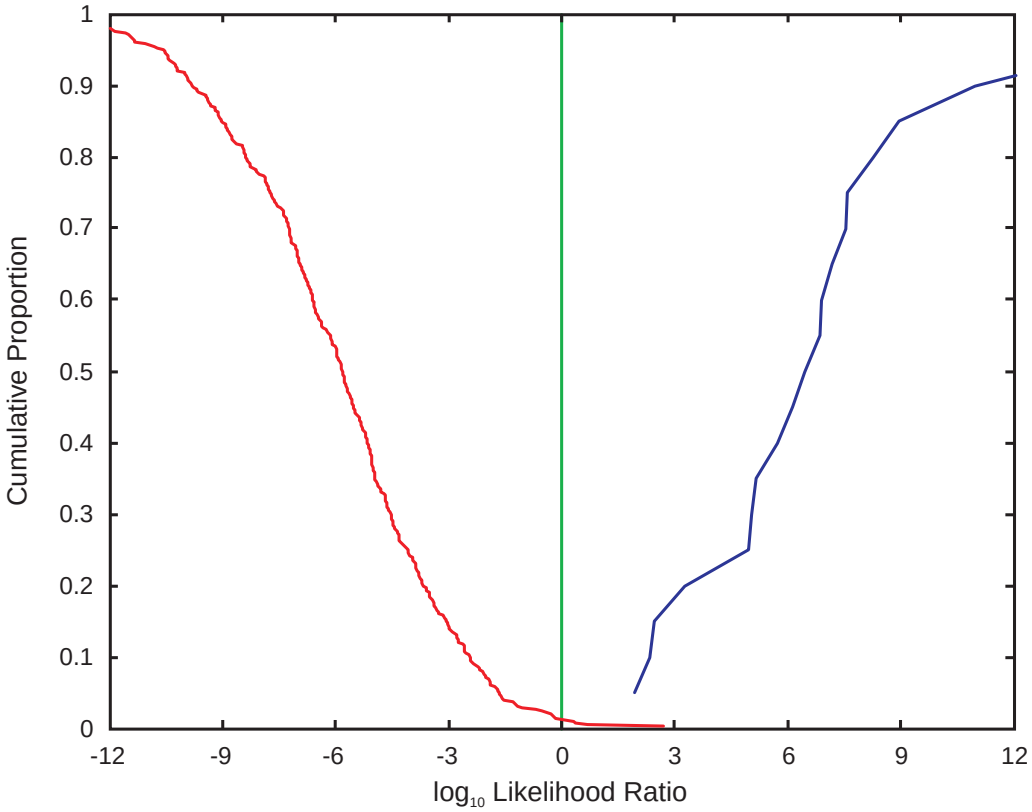
- Background
  - high-quality audio
- Suspect
  - high-quality audio
- Offender
  - high-quality audio
  - landline-to-landline
  - mobile-to-mobile
  - mobile-to-landline

Results

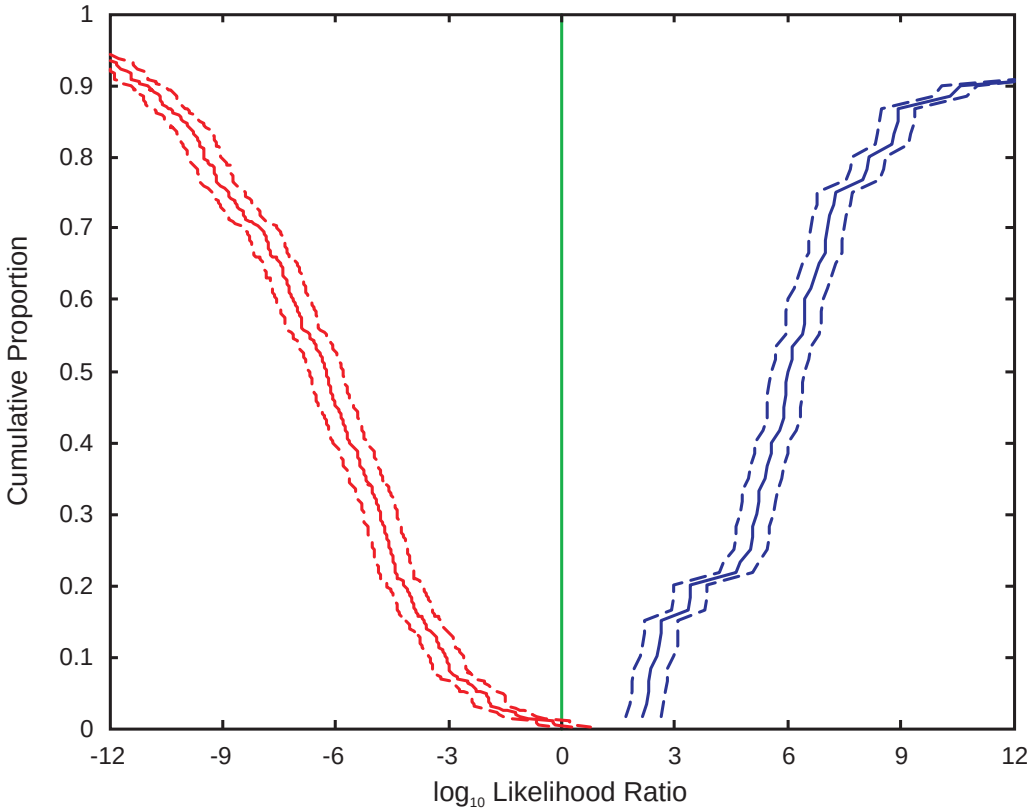

high-quality v high-quality
C_{llr}
0.03
0.025
0.02
0.015
0.01
0.005
0
Baseline MFCC
MFCC on /iau/
FormantMeasurer
WaveSurfer
Praat
NAH2002
MB2006
RSW2007

Results

high-quality v high-quality



Baseline MFCC



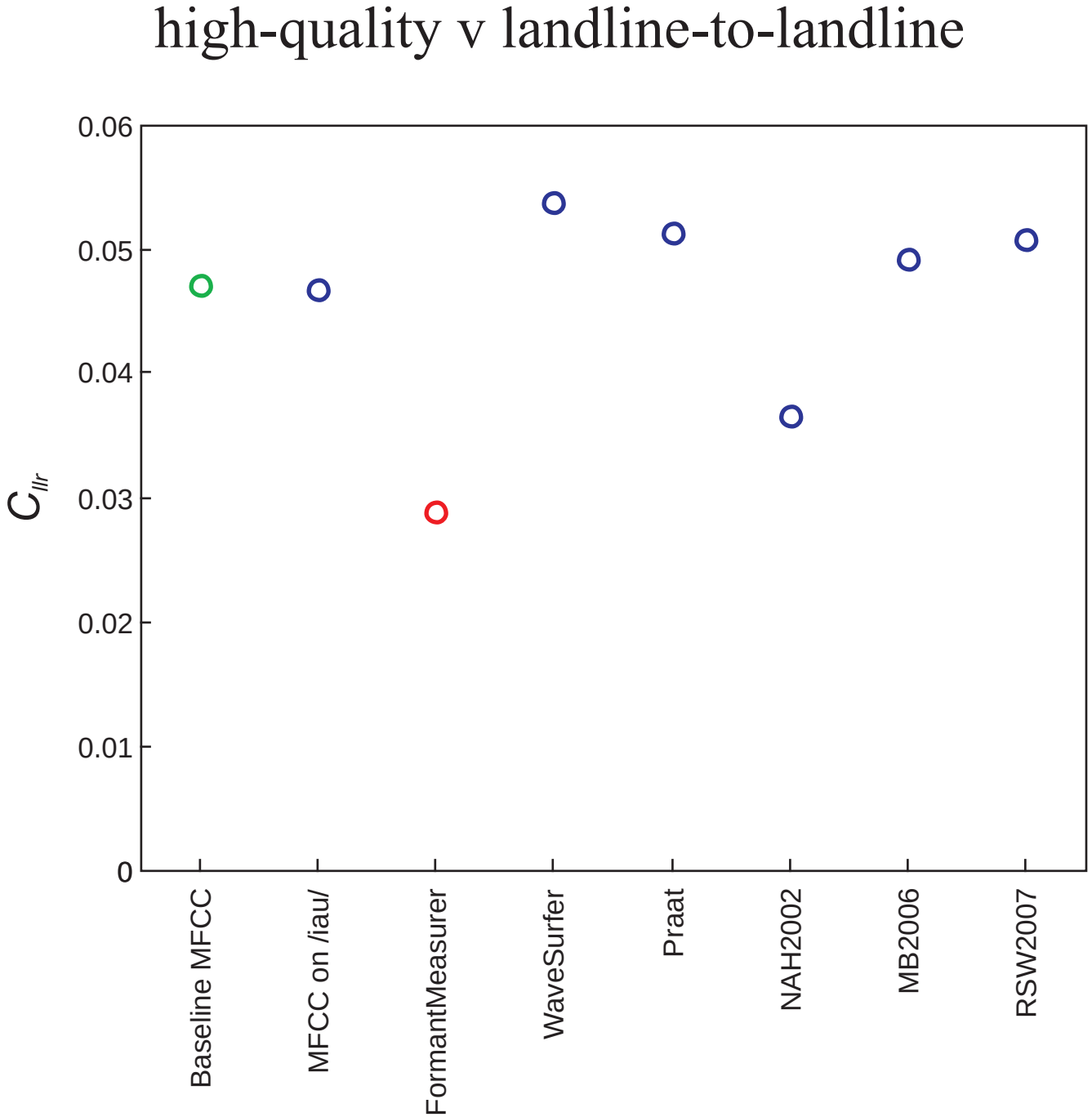
FormantMeasurer

Conclusions

high-quality v high-quality

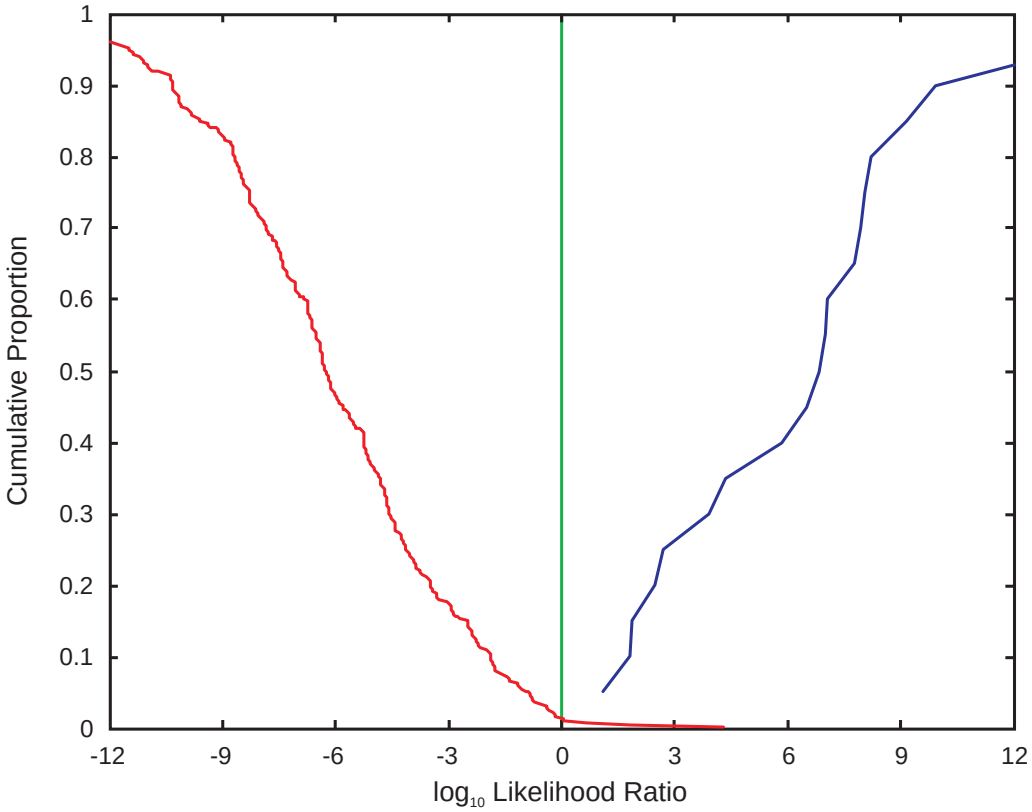
- Human-supervised formant measurement worthwhile?
  - maybe
    - MFCC on /iau/ and WAVESURFER almost as good
- Fully-automatic formant measurement worthwhile?
  - yes for WAVESURFER
    - but MFCC on /iau/ about as good
- Manual segment selection worthwhile?
  - yes

Results

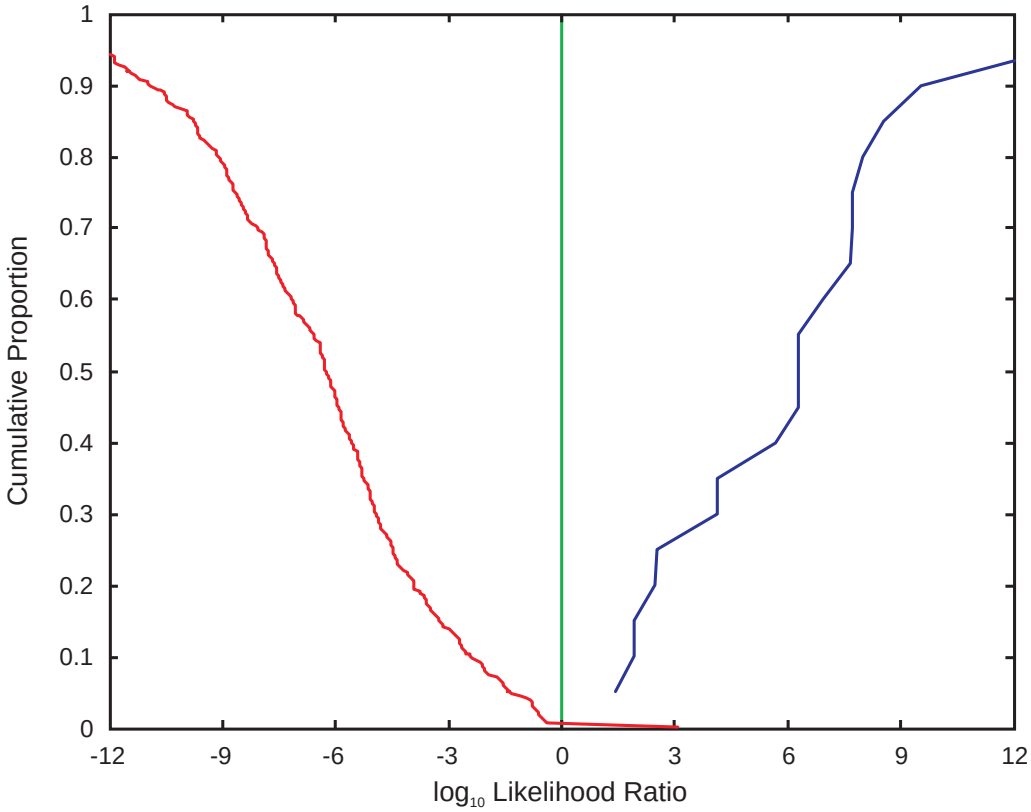

high-quality v landline-to-landline
C_{llr}
0.06
0.05
0.04
0.03
0.02
0.01
0
Baseline MFCC
MFCC on /iau/
FormantMeasurer
WaveSurfer
Praat
NAH2002
MB2006
RSW2007

Results

high-quality v landline-to-landline



Baseline MFCC



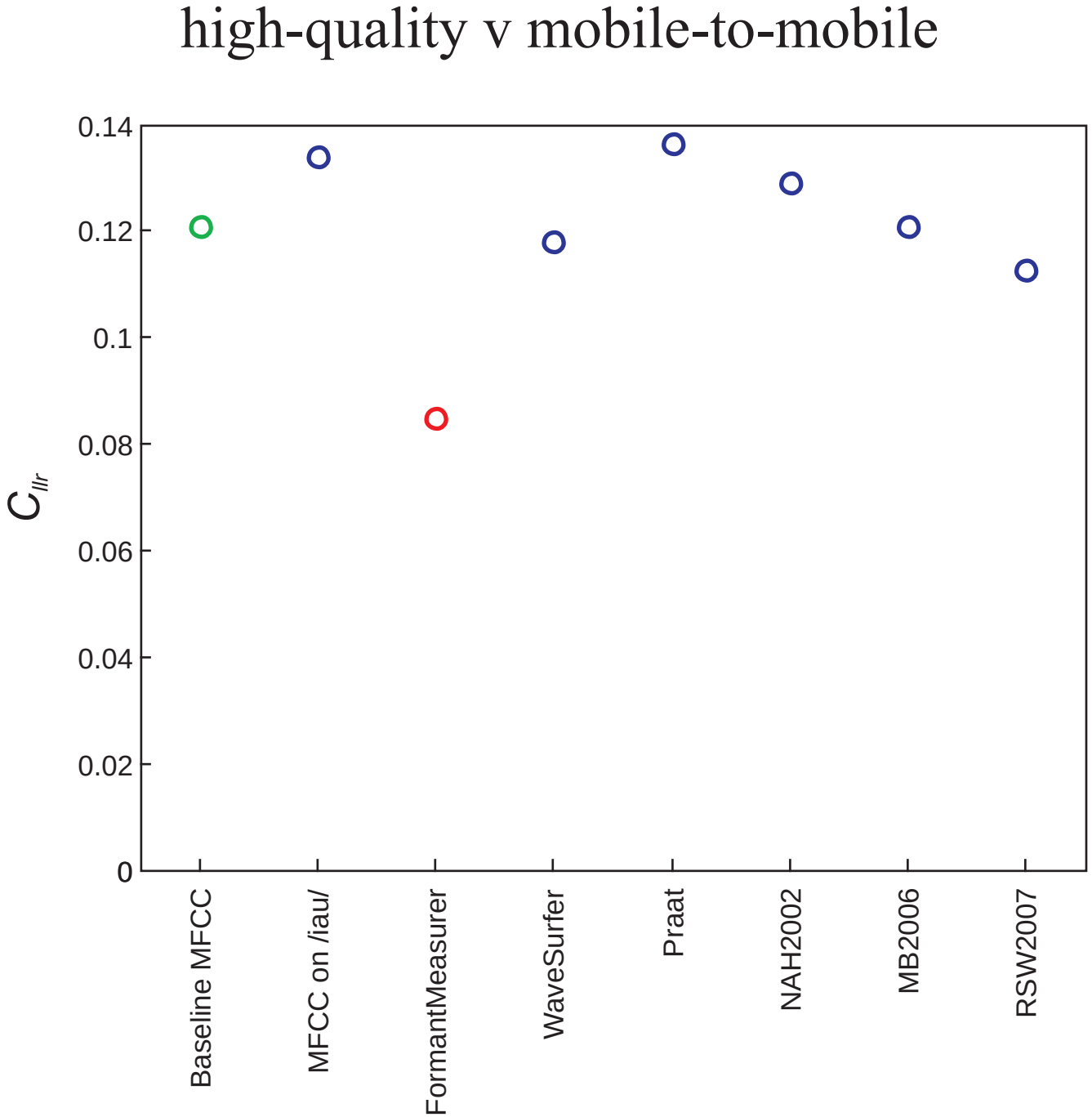
FormantMeasurer

Conclusions

high-quality v landline-to-landline

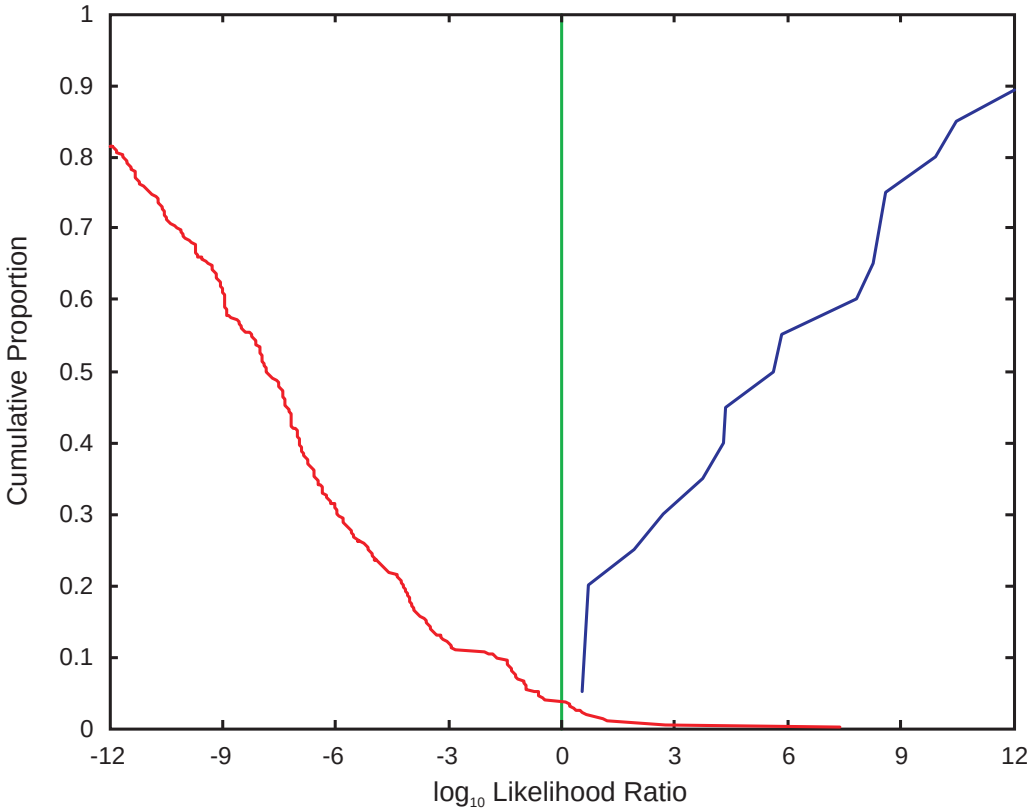
- Human-supervised formant measurement worthwhile?
  - maybe
    - some improvement over baseline
- Fully-automatic formant measurement worthwhile?
  - maybe
    - NAH2002 best candidate
- Manual segment selection worthwhile?
  - yes, contingent on formant measurement

Results

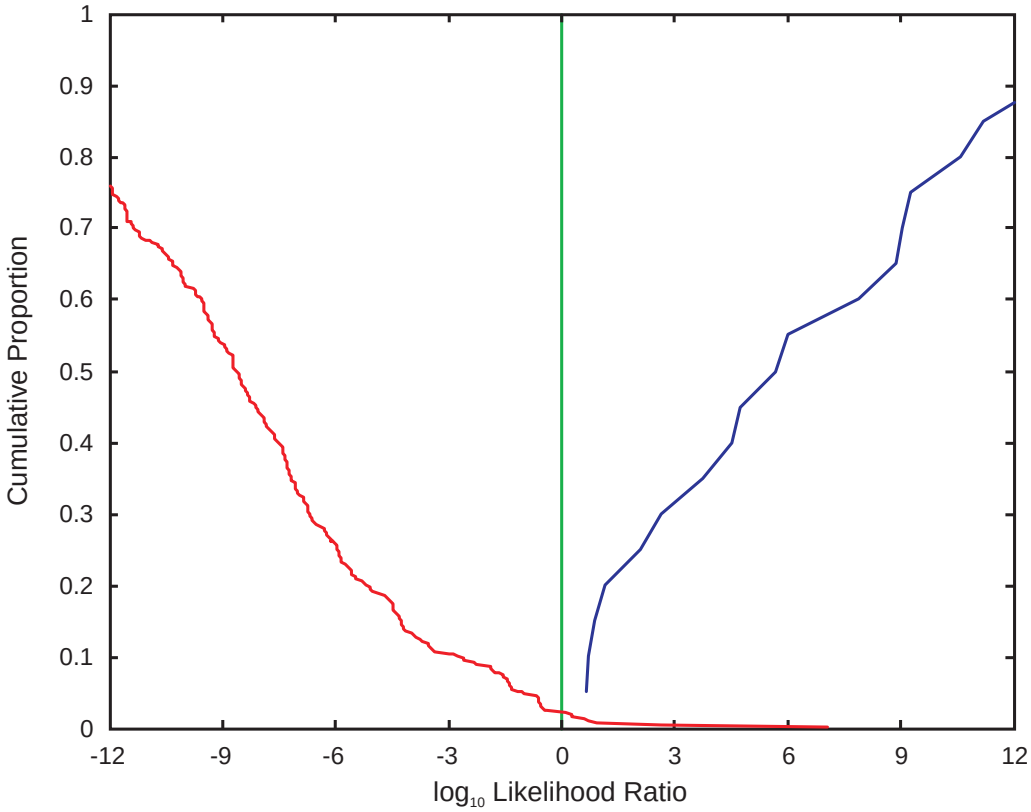

high-quality v mobile-to-mobile
C_{llr}
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Baseline MFCC
MFCC on /iau/
FormantMeasurer
WaveSurfer
Praat
NAH2002
MB2006
RSW2007

Results

high-quality v mobile-to-mobile



Baseline MFCC



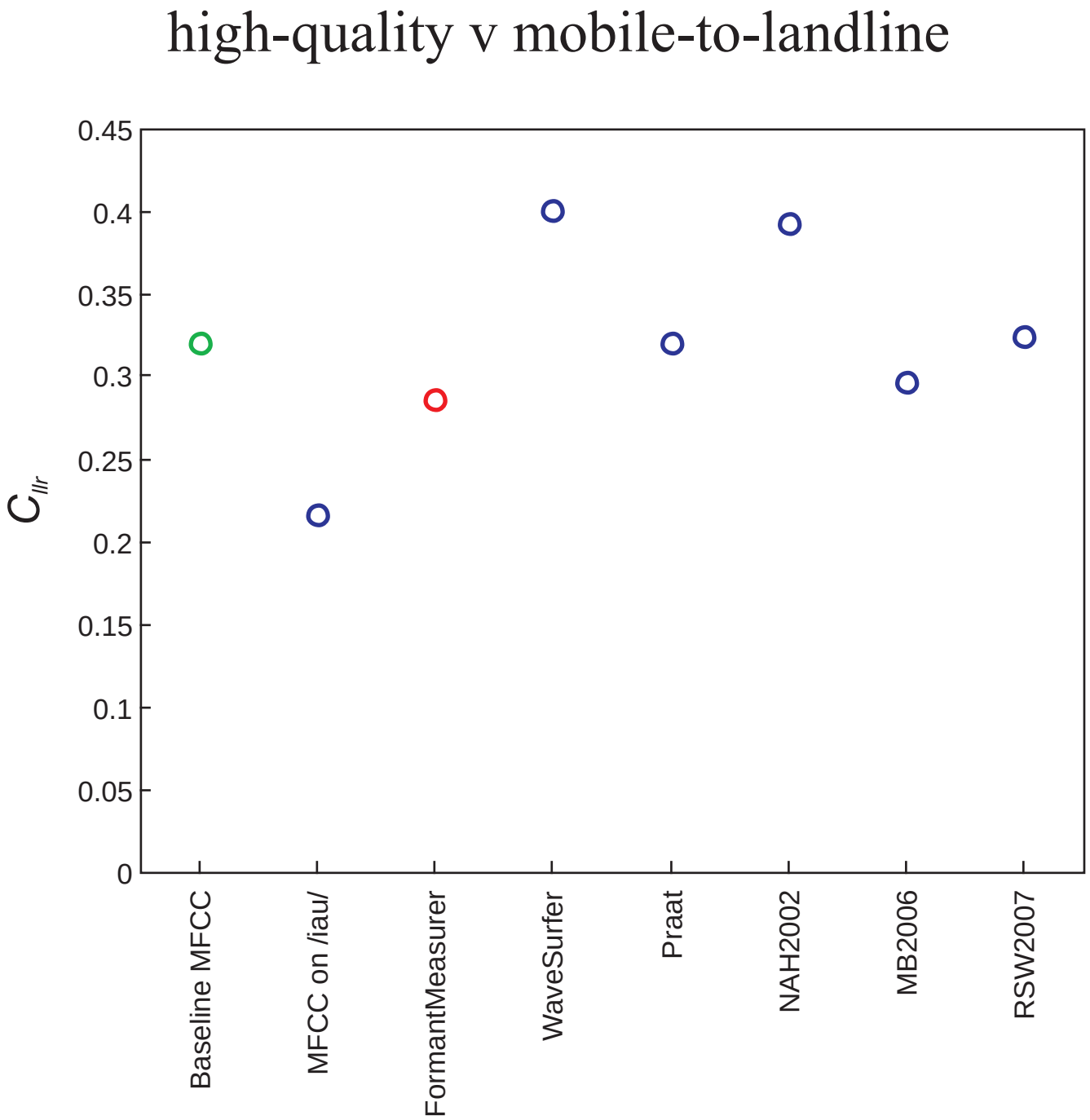
FormantMeasurer

Conclusions

high-quality v mobile-to-mobile

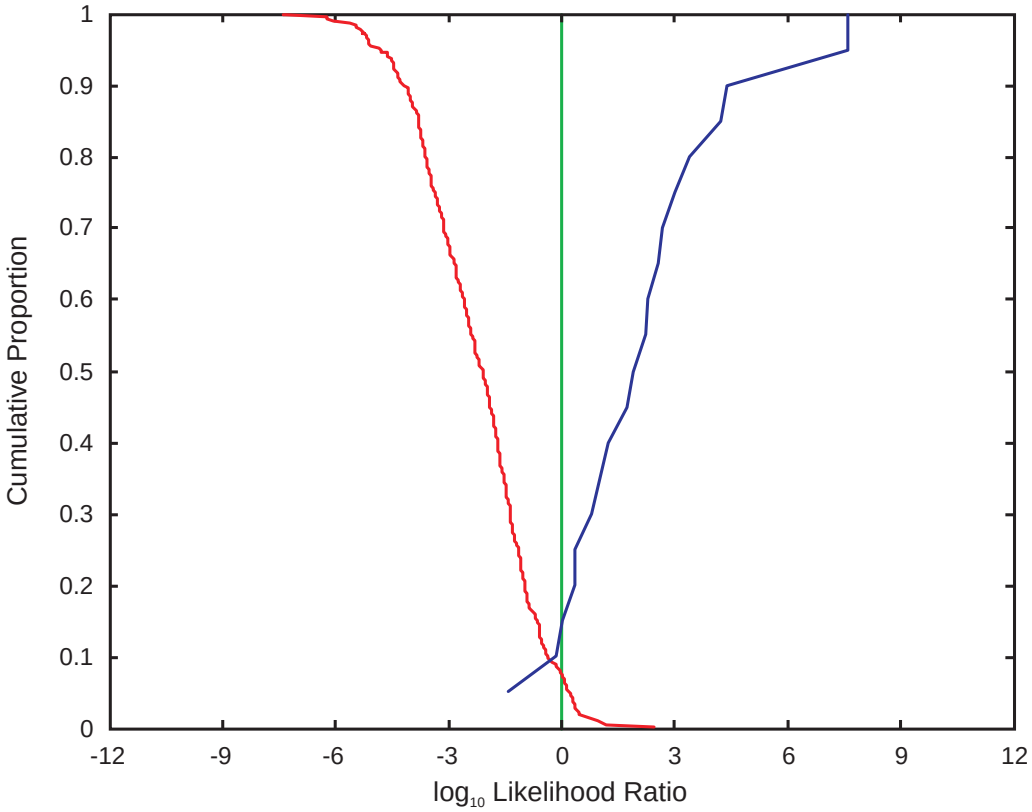
- Human-supervised formant measurement worthwhile?
  - no
- Fully-automatic formant measurement worthwhile?
  - no
- Manual segment selection worthwhile?
  - no

Results

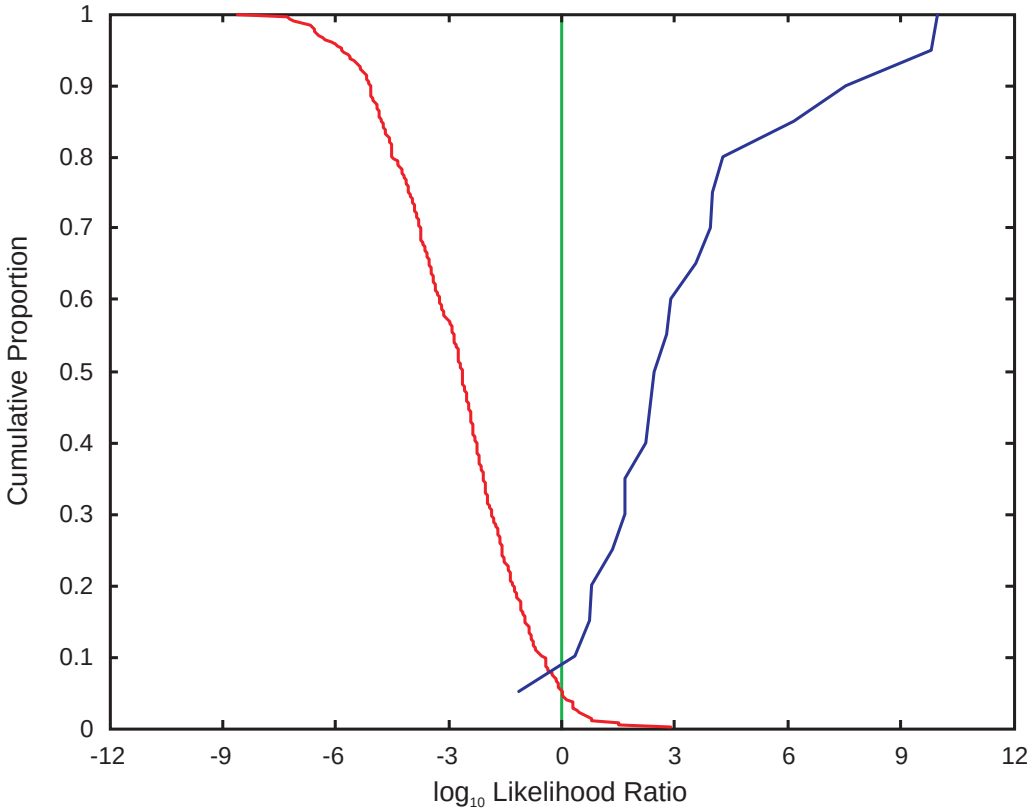

high-quality v mobile-to-landline
C_{llr}
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Baseline MFCC
MFCC on /iau/
FormantMeasurer
WaveSurfer
Praat
NAH2002
MB2006
RSW2007

Results

high-quality v mobile-to-landline



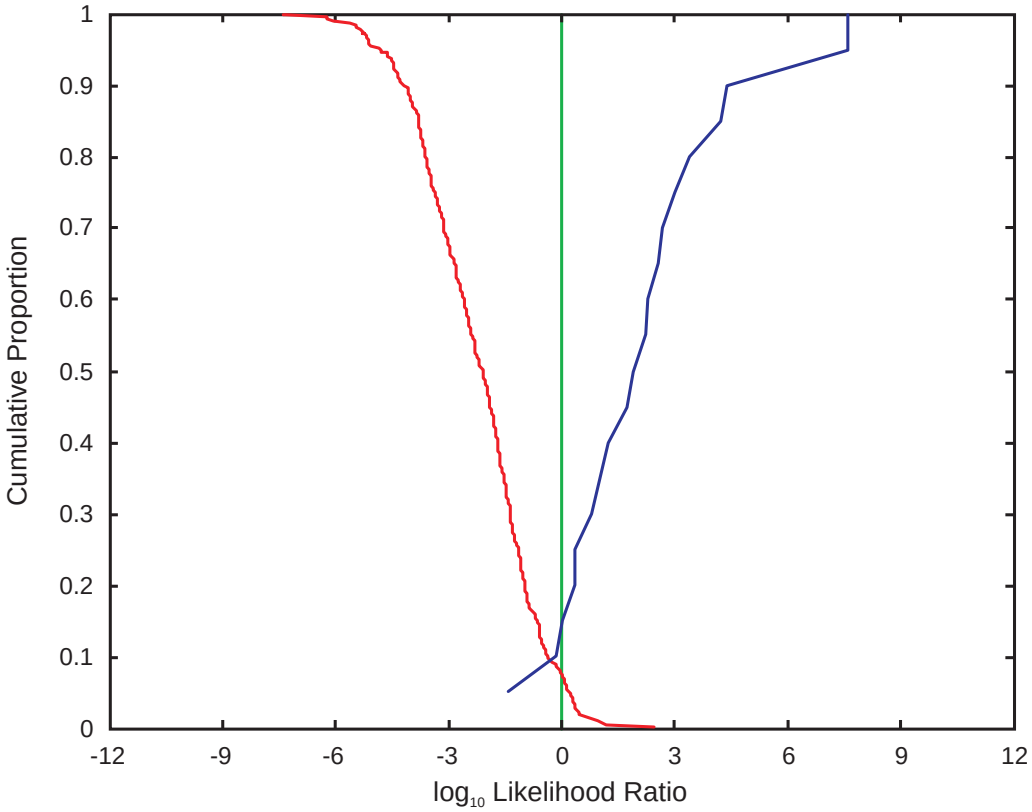
Baseline MFCC



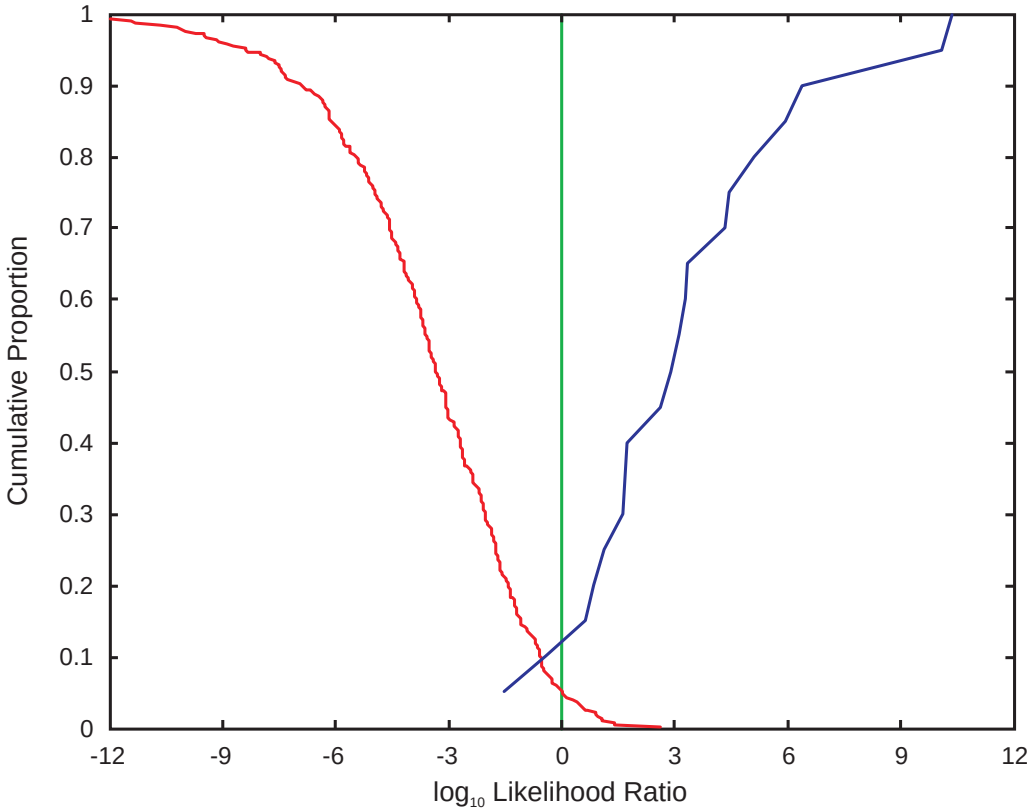
FormantMeasurer

Results

high-quality v mobile-to-landline



Baseline MFCC



MFCC on /iau/

Conclusions

high-quality v mobile-to-landline

- Human-supervised formant measurement worthwhile?
  - no
- Fully-automatic formant measurement worthwhile?
  - no
- Manual segment selection worthwhile?
  - probably no
    - some improvement for MFCC on /iau/

Conclusions

- Human-supervised formant measurement in the landline-to-landline condition may be a worthwhile use of resources.

- Fully-automatic formant measurement in the landline-to-landline condition may be a worthwhile use of resources.

- Neither human-supervised nor fully-automatic formant measurement is a worthwhile use of resources in any condition involving a mobile telephone.

Thank You