Time-of-Recording Estimation for Audio Recordings

Lilei Zheng, Ying Zhang, Chien Eao Lee and Vrizlynn L. L. Thing

Cyber Security Cluster, Institute for Infocomm Research (I^2R), Singapore

Outline

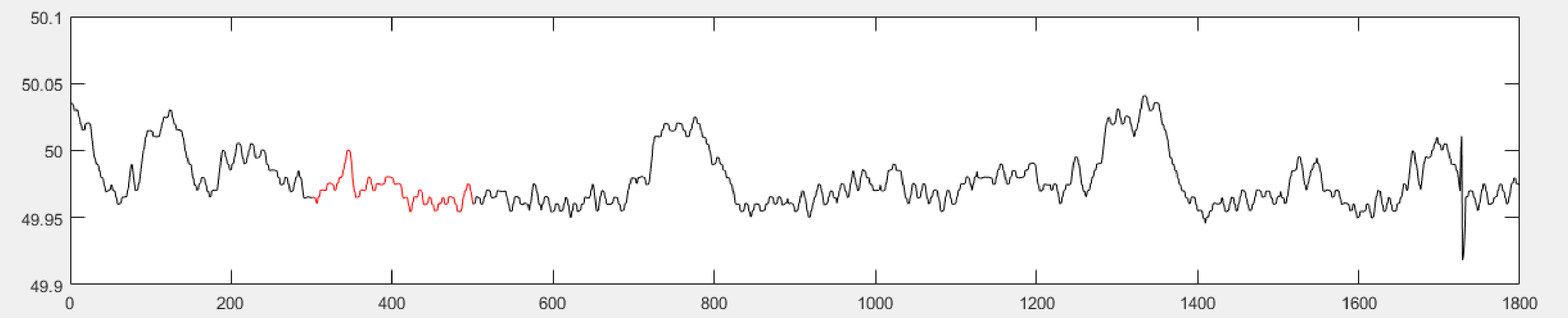
- Introduction: Electric network frequency (ENF)
- Method: Distance/similarity between two ENF sequences
- Application: Time-of-recording estimation (ToRE)
  - Demo video show

Outline

- Introduction: Electric network frequency (ENF)
- Method: Distance/similarity between two ENF sequences
- Application: Time-of-recording estimation (ToRE)
  - Demo video show

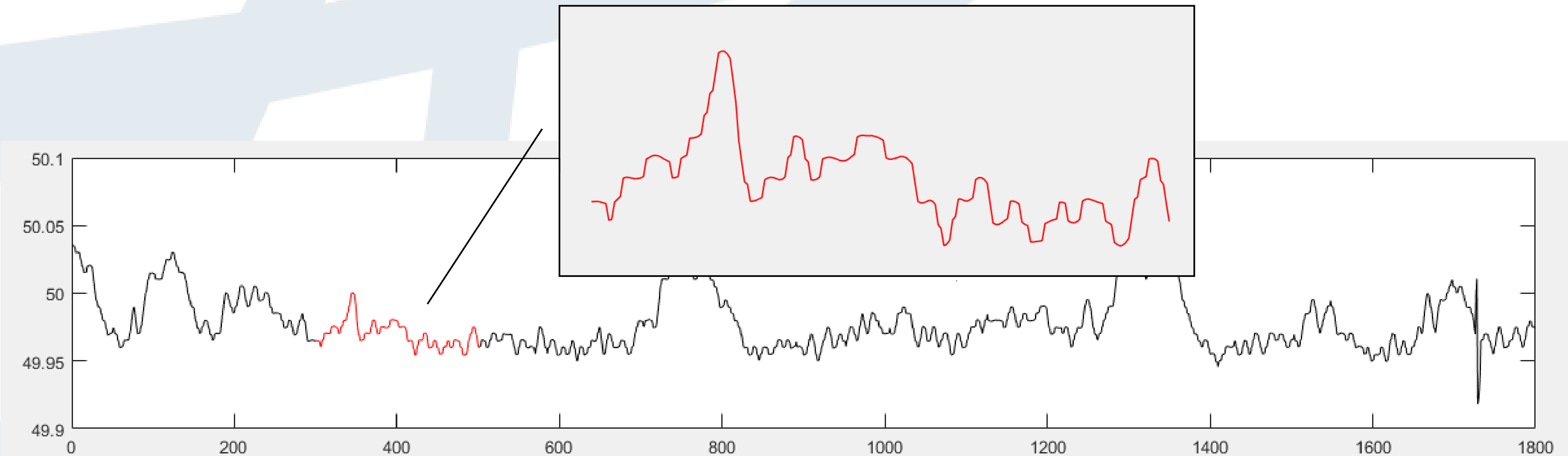
Electric Network Frequency

- A electric power grid has a nominal frequency of either 50 Hz (e.g., Singapore) or 60 Hz (e.g., United States).
- ENF pattern: random fluctuations over time around the nominal value, e.g., [49.9, 50.1] Hz.



Electric Network Frequency

- An ENF pattern is a sequence of values. The longer it is, the larger probability it is unique from other ENF sequences.



Electric Network Frequency

- All equipment connected to the electric power operate in accordance with the ENF.
  - The ENF pattern is consistent across different places within the same power grid.
  - Recording device plugged into power mains can pick up the ENF signal.
  - Portable recording device near other electric-powered devices can capture the ENF signal, e.g., through acoustic hum and mechanical vibrations.

- ENF is a good timestamp.

Use Case

- An audio recording is used as evidence, but the claimed recording time is doubted.

- Conditions
  - The recording was taken in an environment surrounded by electric devices such as lights and TV, i.e., we can extract the ENF pattern from the recording.
  - We have the historical ENF data (from long time ago to today) covering the possible recording time.
- It is possible to know the actual recording time by pattern matching.

ENF Timestamp

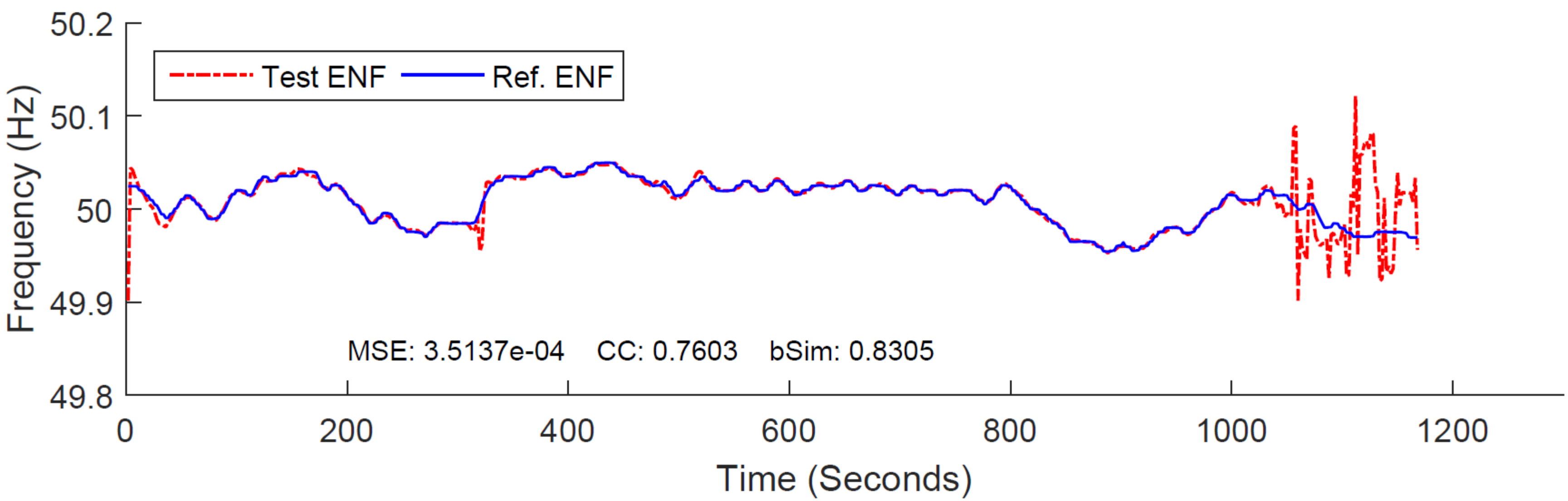
- Recording device plugged into power mains is used to pick up the ENF signal as reference ENF.
  - Non-interrupted recording through years.

ENF Timestamp

- The ENF signal, unintentionally captured by electric-powered or portable recording devices, is called test ENF.
  - Many recorders are able to make audios.

ENF Timestamp

- Different from the reference ENF, the test ENF may be unavailable or partially “noisy” due to, e.g.,
  - Far away from any electric-powered device.
  - Quick move of the audio recorder.



Time-of-recording Estimation

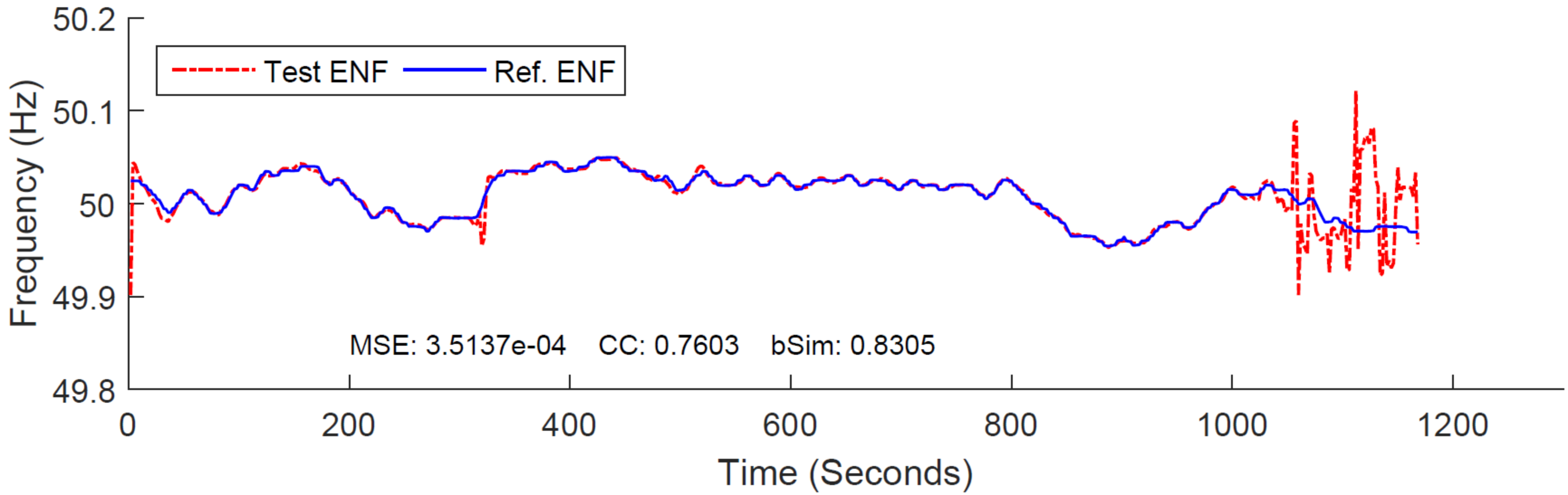
- It is hard to find the exact match within the reference ENF for the test ENF.
- We search for the best match and take its timestamp as the estimated recording time for the test audio.
  - A proper similarity to measure the matching.
  - A fast search algorithm.

Outline

- Introduction: Electric network frequency (ENF)
- Method: Distance/similarity between two ENF sequences
- Application: Time-of-recording estimation (ToRE)
  - Demo video show

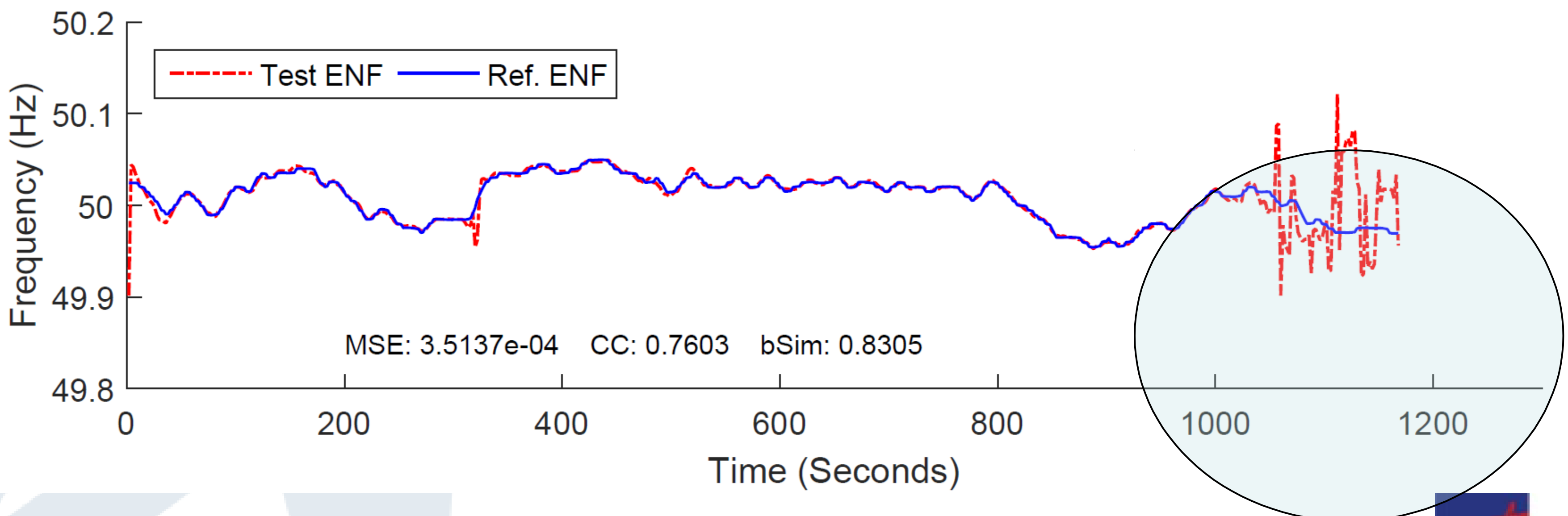
Visual Comparison

- Visual comparison
  - Easy to know how and where is the match.
  - Inefficient and subjective (human biased).



MSE and CC

- Mean squared error (MSE) and correlation coefficient (CC)
  - Traditional measures used in prior arts.
  - Problem: larger local mismatches will contribute more to the final similarity/distance score.



Bitwise Similarity

- Bitwise Similarity (bSim)

$$bSim(\mathbf{t}, \mathbf{r}) = \frac{1}{N}\sum_{i=1}^{N} s_i, \quad s_i = \begin{cases} 1, & \|t_i - r_i\| < \theta \\ 0, & \|t_i - r_i\| \geq \theta \end{cases}$$

- Using a threshold to truncate the local mismatch
  - All the local mismatches larger than the threshold are treated the same.
  - A binary version of MSE.

Outline

- Introduction: Electric network frequency (ENF)
- Method: Distance/similarity between two ENF sequences
- Application: Time-of-recording estimation (ToRE)
  - Demo video show

Fast Search

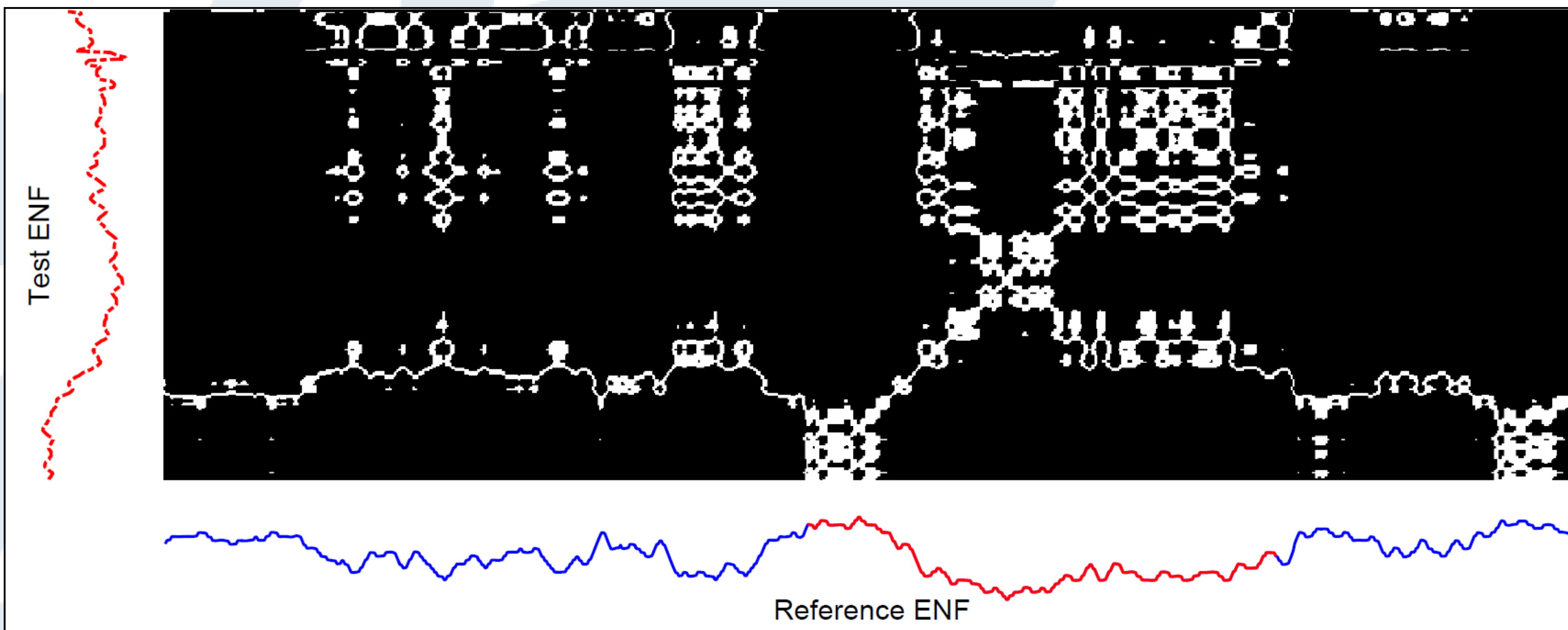
- Bitwise Similarity (bSim)

$$bSim(\mathbf{t}, \mathbf{r}) = \frac{1}{N}\sum_{i=1}^{N} s_i, \quad s_i = \begin{cases} 1, & \|t_i - r_i\| < \theta \\ 0, & \|t_i - r_i\| \geq \theta \end{cases}$$

- The distance map between the test ENF and the reference ENF is binary.
  - 1 for local match, 0 for local mismatch.
  - Operating on binaries is beneficial for fast computation.

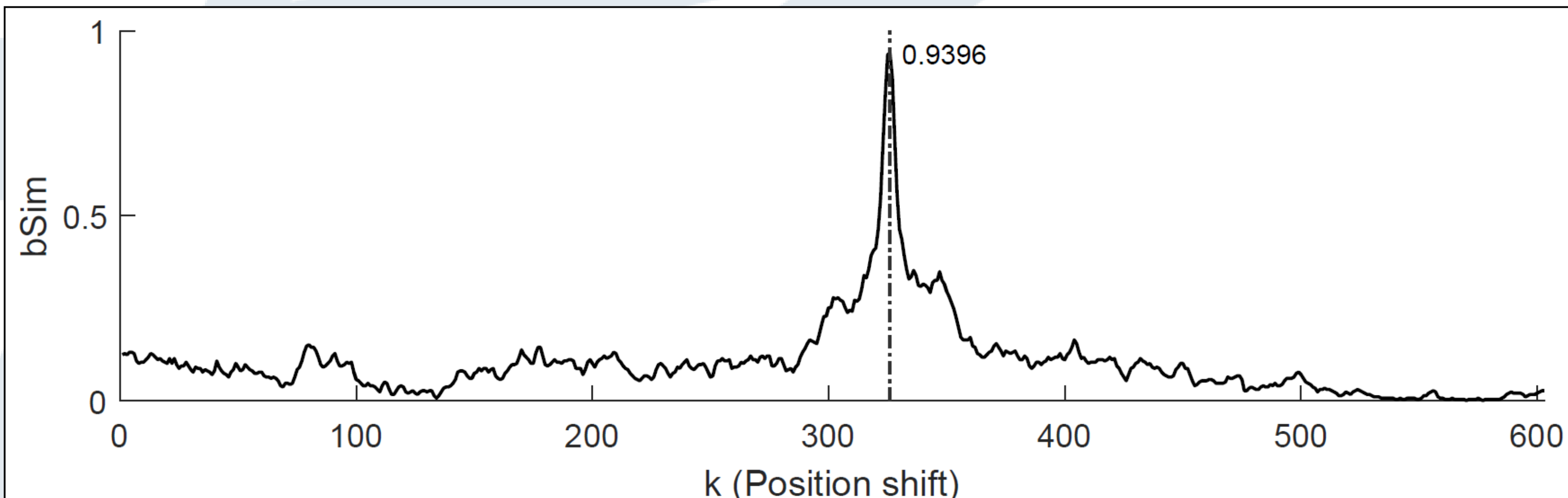
Fast Search

- Binary distance map
    - White for 1, black for 0.
    - White diagonal indicates the position of the matched segment on the reference ENF.



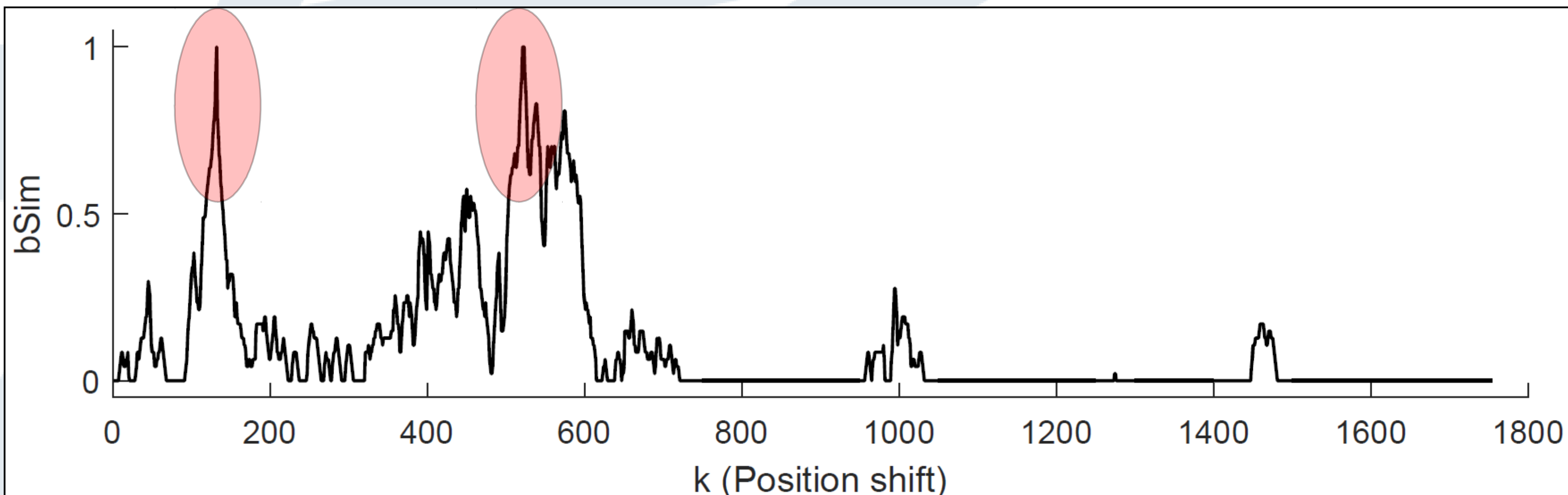
Fast Search

- Curve of bSim values
  - Similarity values of all the diagonal lines.



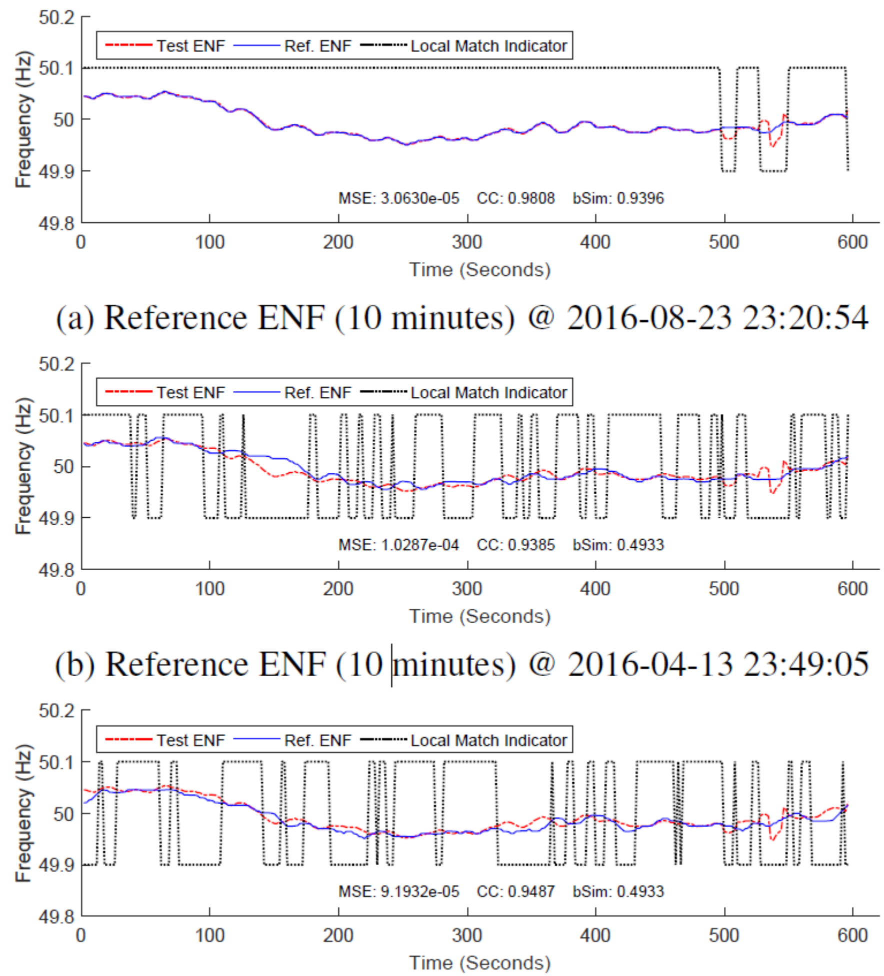
Conditional Uniqueness

- A short test ENF may have multiple matches on the reference ENF.



Top-3 Retrieval

- Retrieving the Top-3 best matches to avoid wrong decision due to the probable problem of multiple matches.
- The Top-1 similarity should be significantly larger than the other two similarities.



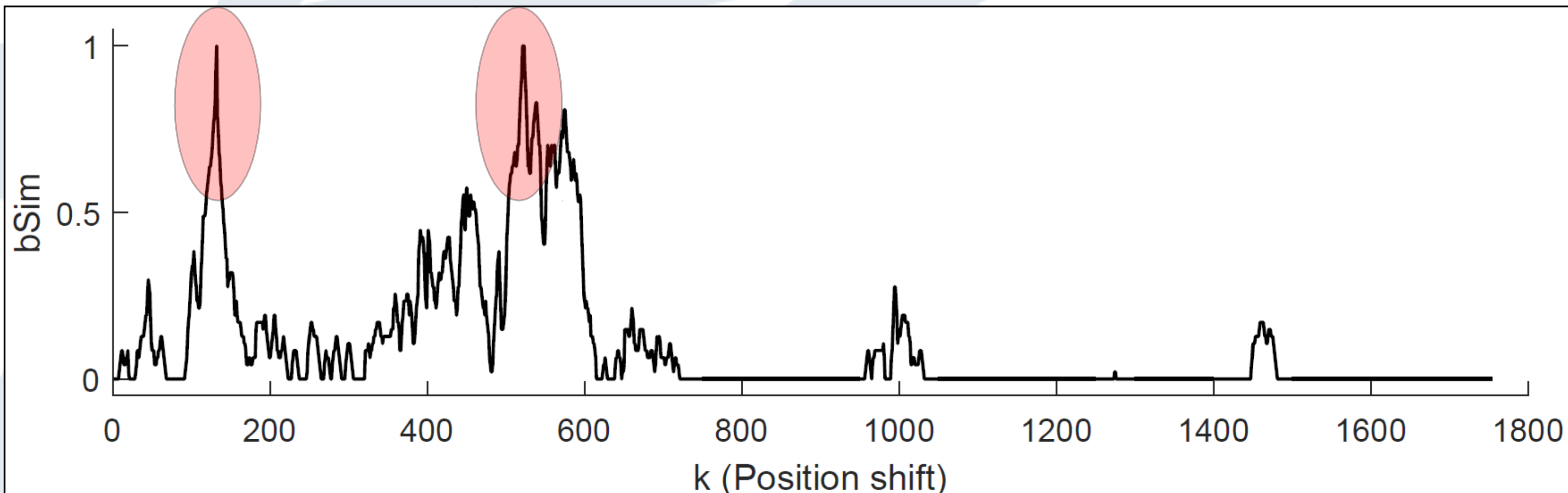
(a) Reference ENF (10 minutes) @ 2016-08-23 23:20:54

(b) Reference ENF (10 minutes) @ 2016-04-13 23:49:05

(c) Reference ENF (10 minutes) @ 2016-04-22 18:46:24

Top-3 Retrieval

- Exceptions are given up with “no decision”.
  - To ensure high precision of the final decisions.
  - High precision is essential for supporting evidence in forensics.



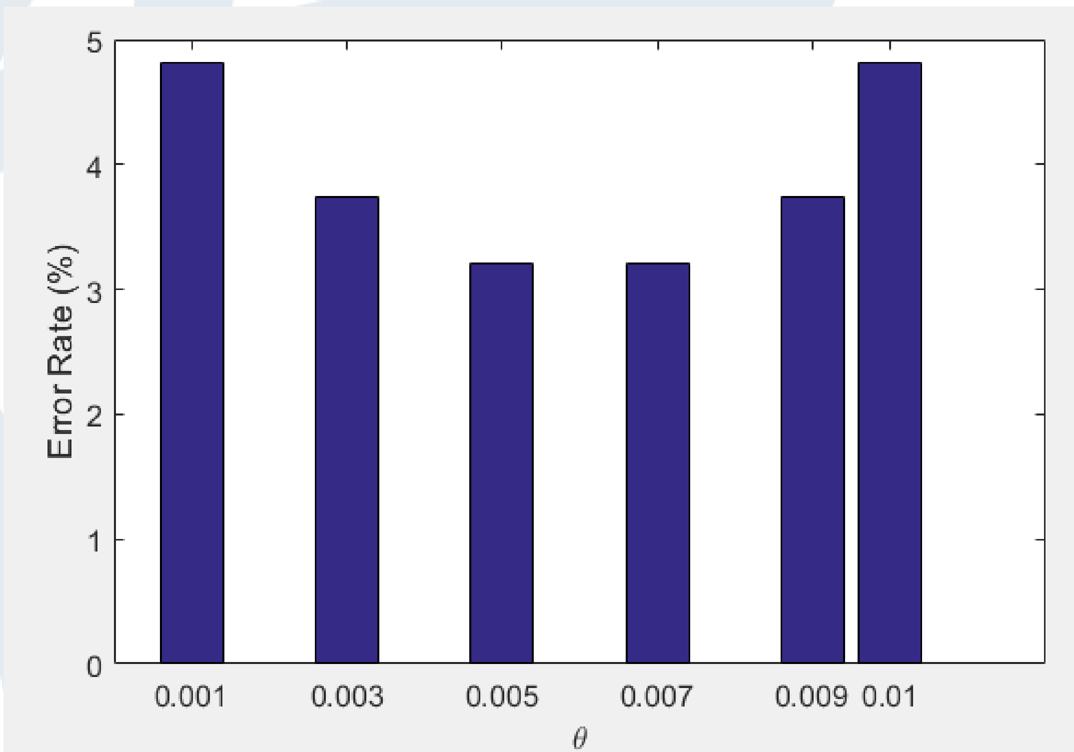
Experiment Setup

- Reference ENF data of Singapore since 2013.
- Test ENF data collected in August, 2016, including 187 audio recordings.

- For each test audio, the estimated time within a shift of 1 minute to the ground truth is considered as correct.
  - People usually note time up to minutes, e.g., 16:50.
  - The estimation can be accurate up to seconds.
- Error rate; precision & recall.

Experiment (1)

- Error rates with respect to the threshold θ
  - The best θ is between 0.005 and 0.007.



Experiment (2)

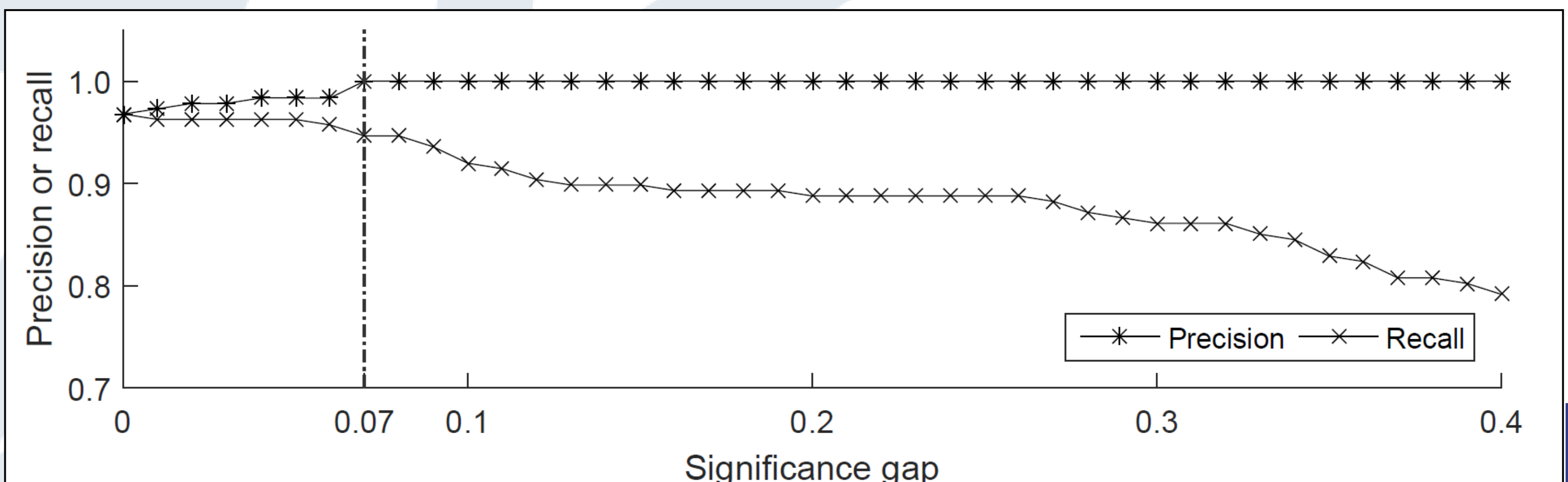
- Comparison to prior arts
  - Lowest error rate
  - Fastest search

Approaches	Min. MSE (baseline 1)	Max. CC (baseline 2)	DMA [1]	Max. bSim (this work)
Top-1 error	22.46%	22.46%	20.32%	**2.67%**
Searching time	1.4649	1.9698	41.0444	**0.8973**

- Binarization is beneficial for this task
  - Robust to local mismatch (e.g., noise)
  - Efficient computation on binaries

Experiment (3)

- Top-3 retrieval can ensure high precision
  - Significance gap: the similarity gap between the Top-1 result and the others
  - Precision reaches 100% when the significance gap is larger or equal to 0.07, with the recall as high as 94.65%.



Conclusions

- Bitwise similarity (bSim) for accurate and fast ENF matching.
- Top-3 retrieval for making confident decisions of timestamp estimation.

- A beta-version application implemented in MATLAB.
  - Video demo show in next slide.

Demo Video