SHAZAM

Marco Gazzoli
Michele Svanera

Audio Fingerprinting and Recognition System

Università Degli Studi Di Brescia

A.A. 2012-2013

Index

- Aim – Targets
- Fingerprinting Algorithm Overview
- Matching/Scoring
- Implementation Keypoints
- Live Recording Results
- Future Works
- Conclusions
- Bibliography

Aims - Targets

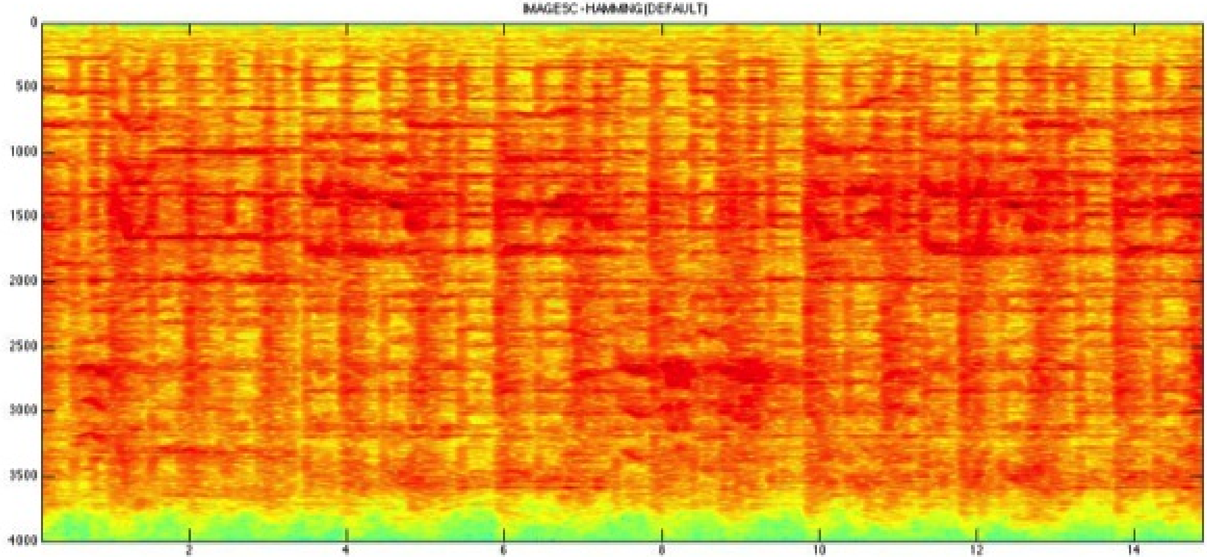
- Audio Recognition (excluding live performance and/or cover)
- Robustness against:
  - Noise (environment)
  - Interference (additive speech)
- Wide DB – Narrow Signature
  - Fast Computation
  - Fast Search

Fingerprinting Algorithm Overview

- Fingerprint = Representative Information Signature
- How to:
  - Spectrogram
  - Peaks Detection (most relevant in half second slot) – Anchor Point (AP)
  - For each AP:
    - Peaks Detection (most relevant in half second slot next to the AP) - Nearby Peak (NP)
    - HASH Generation (delay invariant)

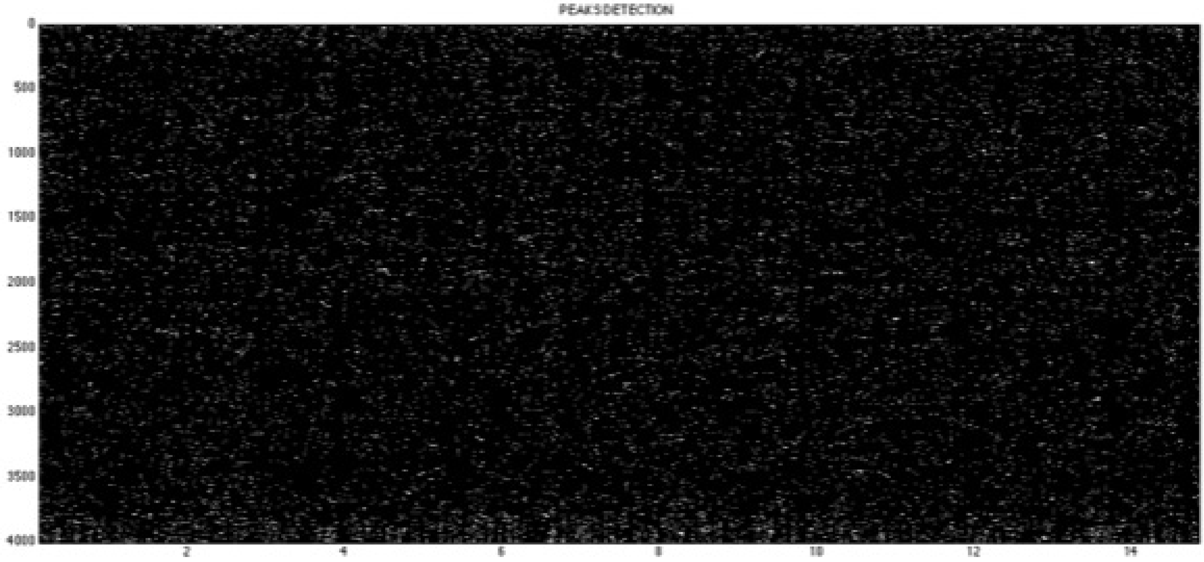
Fingerprinting Algorithm Overview

- Spectrogram



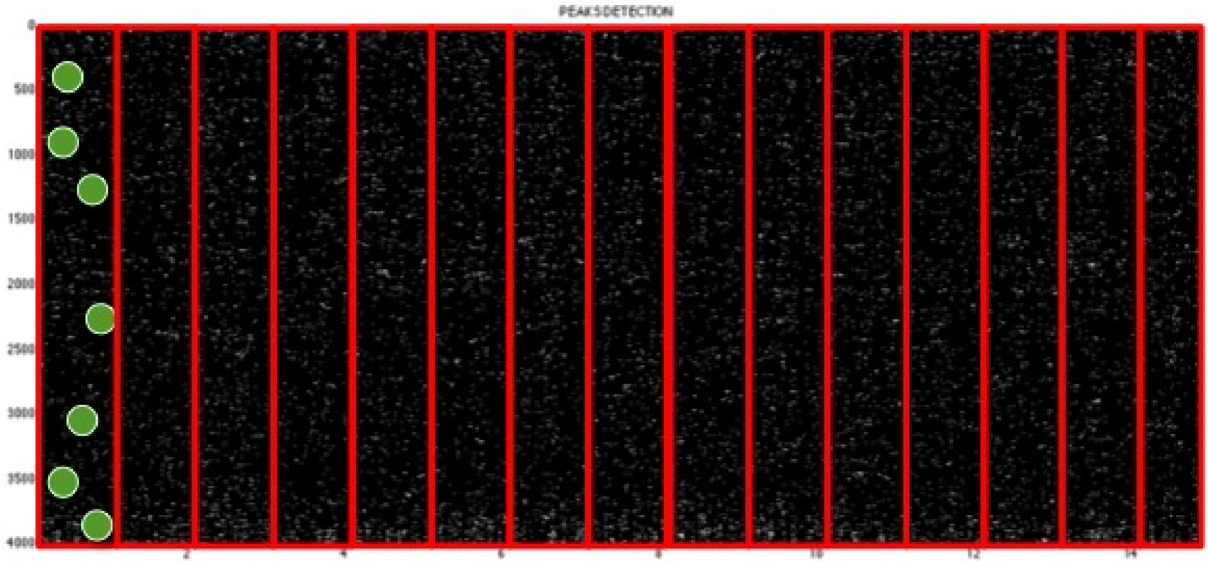
Fingerprinting Algorithm Overview

- Peaks Detection (most relevant in half second slot) – Anchor Point (AP)



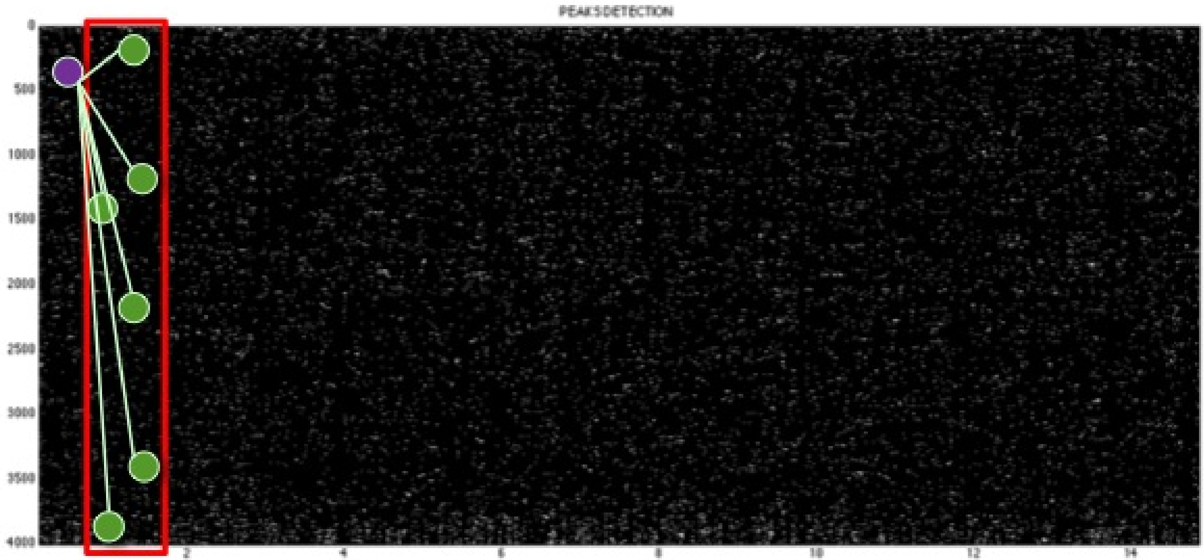
Fingerprinting Algorithm Overview

- HALF SECOND + PEAKS SELECTION (HIGHEST)



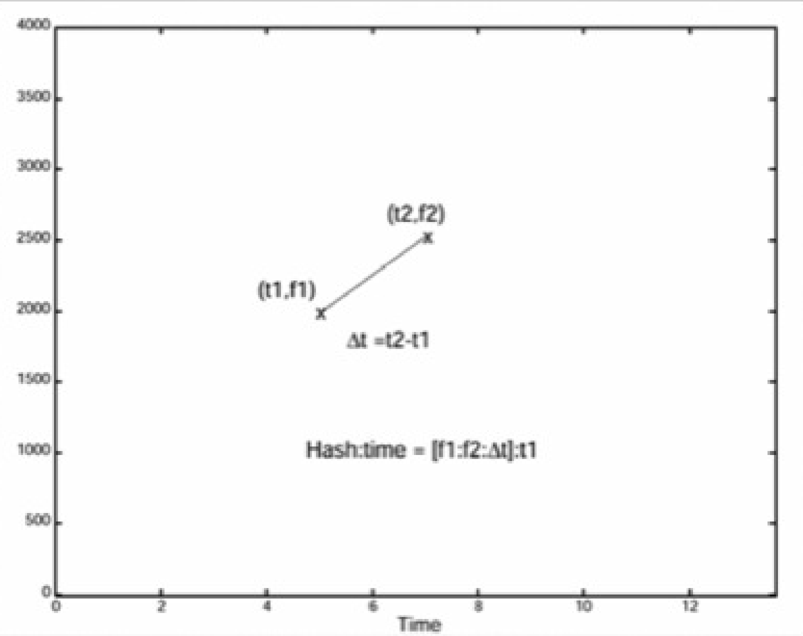
Fingerprinting Algorithm Overview

- PEAKS ASSOCIATION (NP – Nearby Peaks)



Fingerprinting Algorithm Overview

- HASH Generation



Matching

- HASH STRUCTURE:
  - [f1:f2: Δ t]:t1 (for each couple of AP + NP)
  - Delay Invariant Information + Offset Information
- SERVER SIDE:
  - Each DB-file generates its HASH signature
- CLIENT SIDE
  - Sample HASH
- MATCH:
  - [… : … : …] content must be the same in Sample HASH and in DB-file HASH

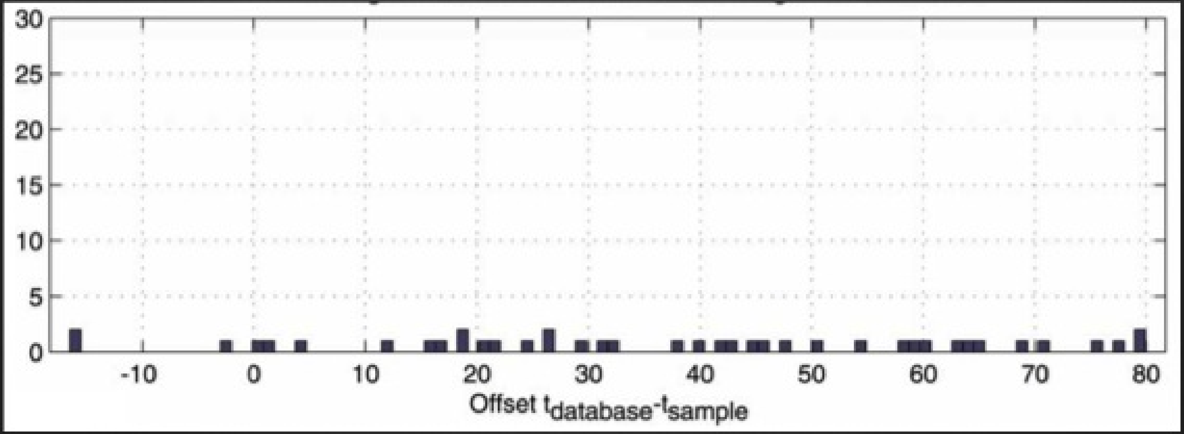
Scoring

- Scatter Plot Histogram
  - For each match, keep track of $\Delta t = t_{sample} - t_{db}$
  - Probability Function

- Most probable Song = Highest Peak among all Scatter Plot Histograms

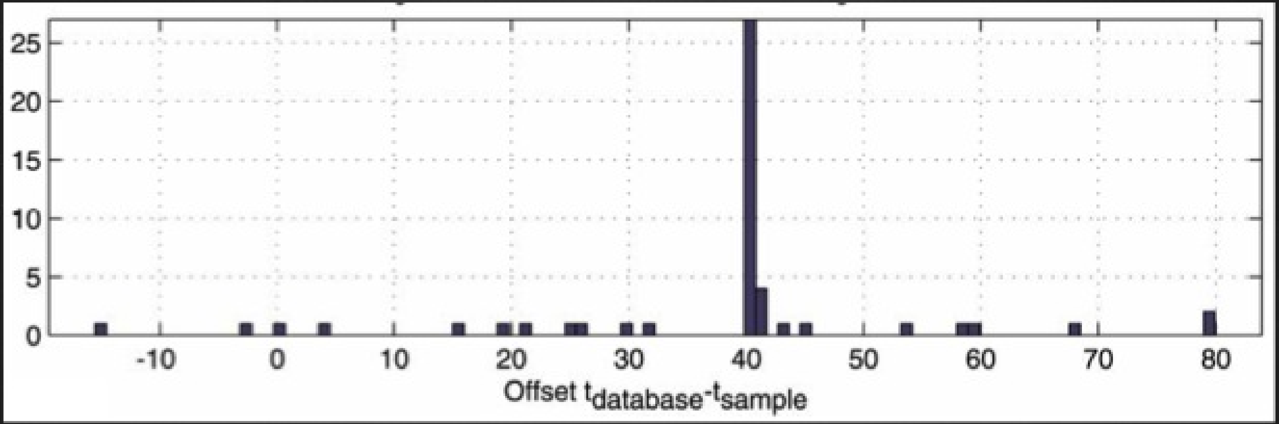
Scoring - Example

- Match Scoring – Scatterplot Histogram



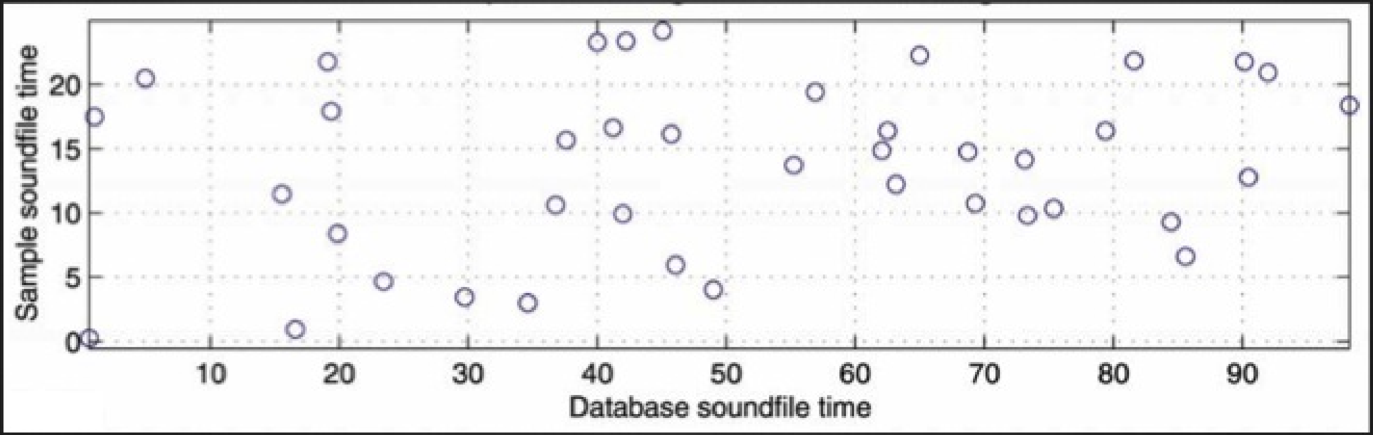
Scoring - Example

- Match Scoring – Scatterplot Histogram



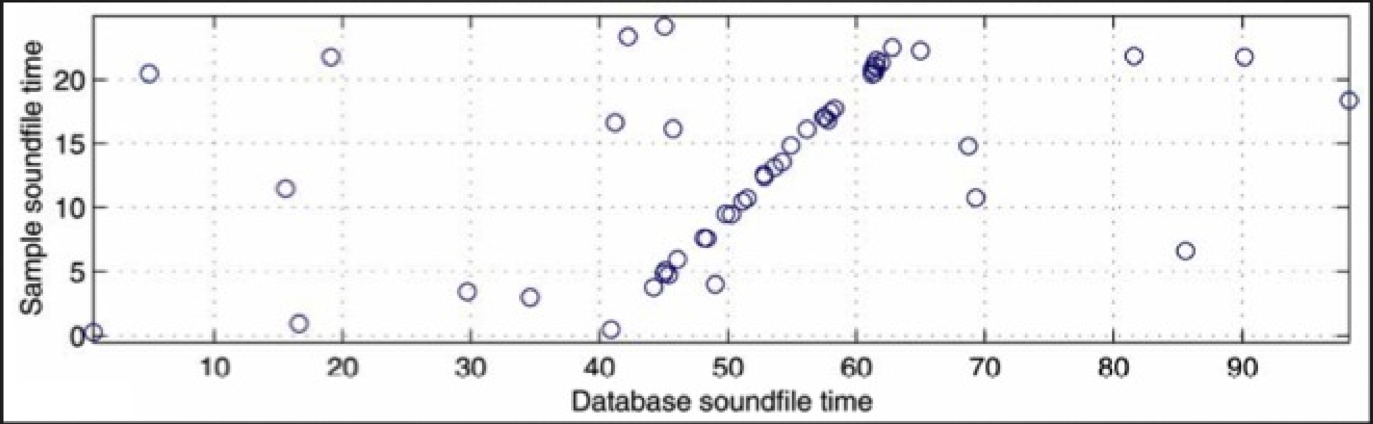
Scoring - Example

- Match interpretation



Scoring - Example

- Match interpretation



Implementation Keypoints

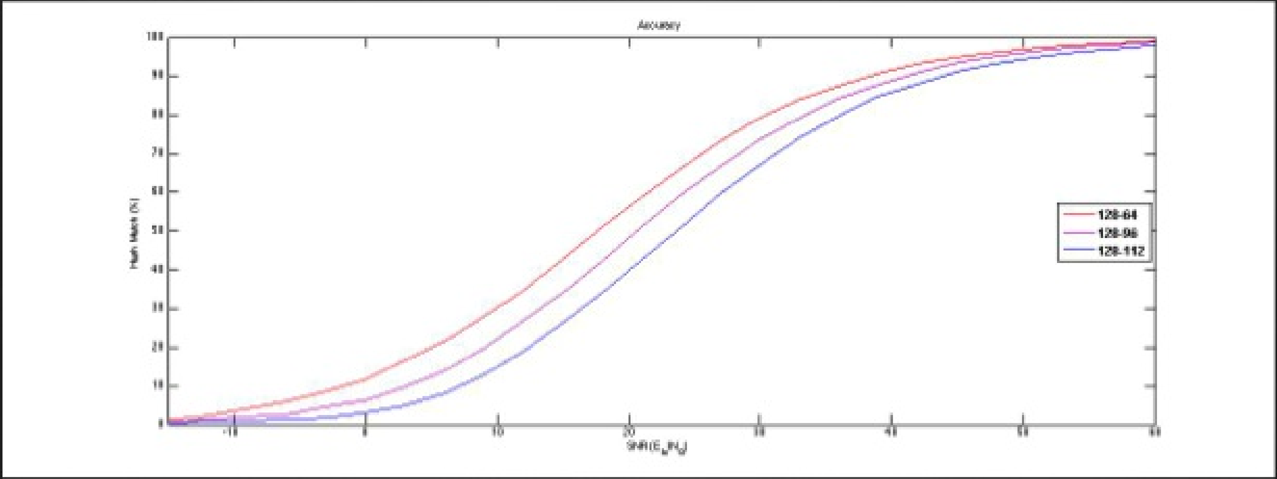
Accuracy → Window Size

Discrimination on wide DB → N FFT

Time peaks localization → Overlap

Window Size

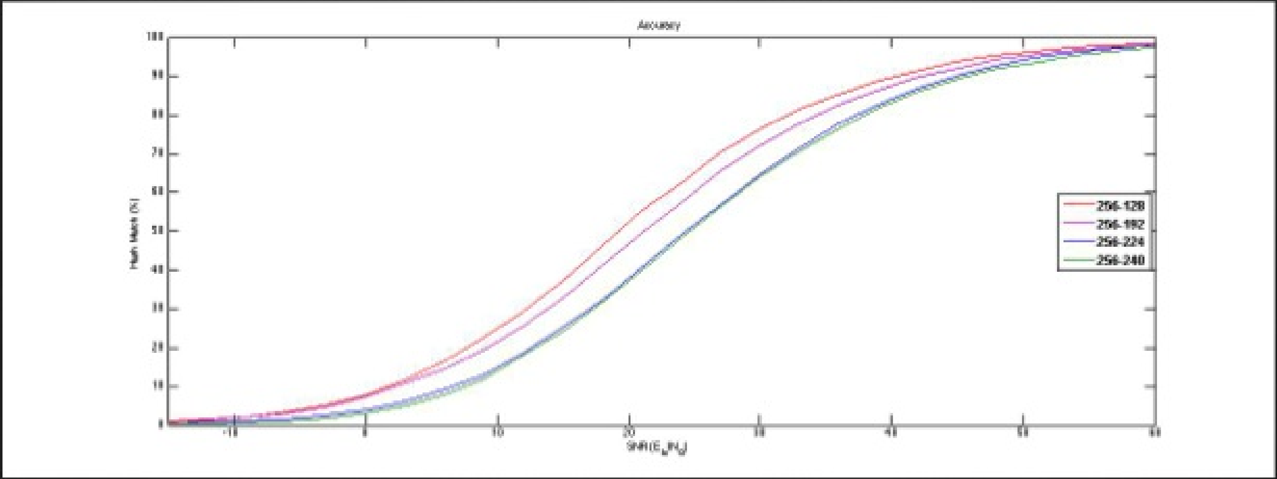
- Best choice?



- First Attempt: Fix Window – Variable Overlap

Window Size

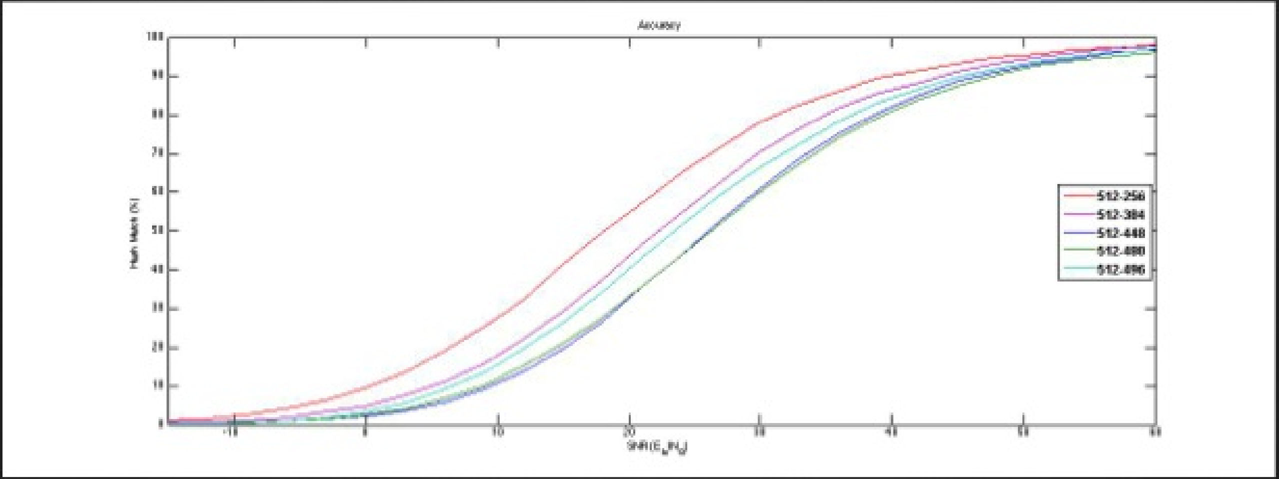
- Best choice?



- First Attempt: Fixed Window – Variable Overlap

Window Size

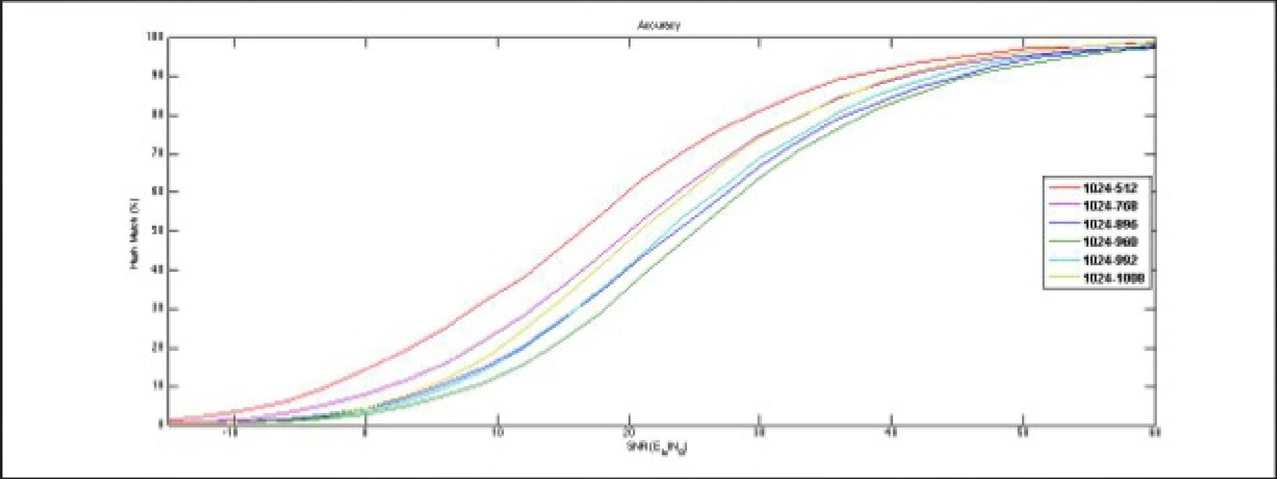
- Best choice?



- First Attempt: Fixed Window – Variable Overlap

Window Size

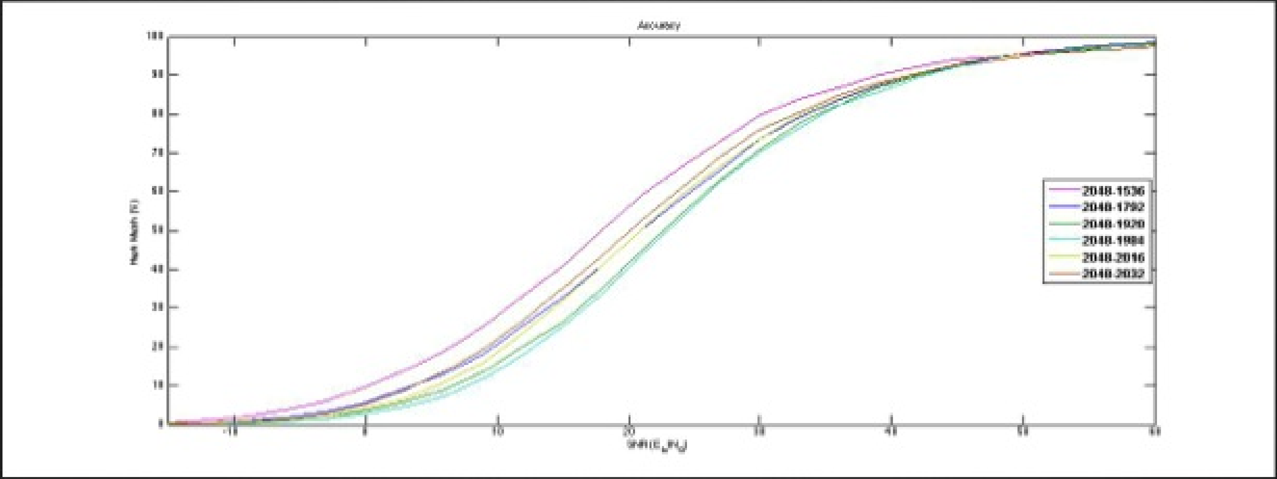
- Best choice?



- First Attempt: Fixed Window – Variable Overlap

Window Size

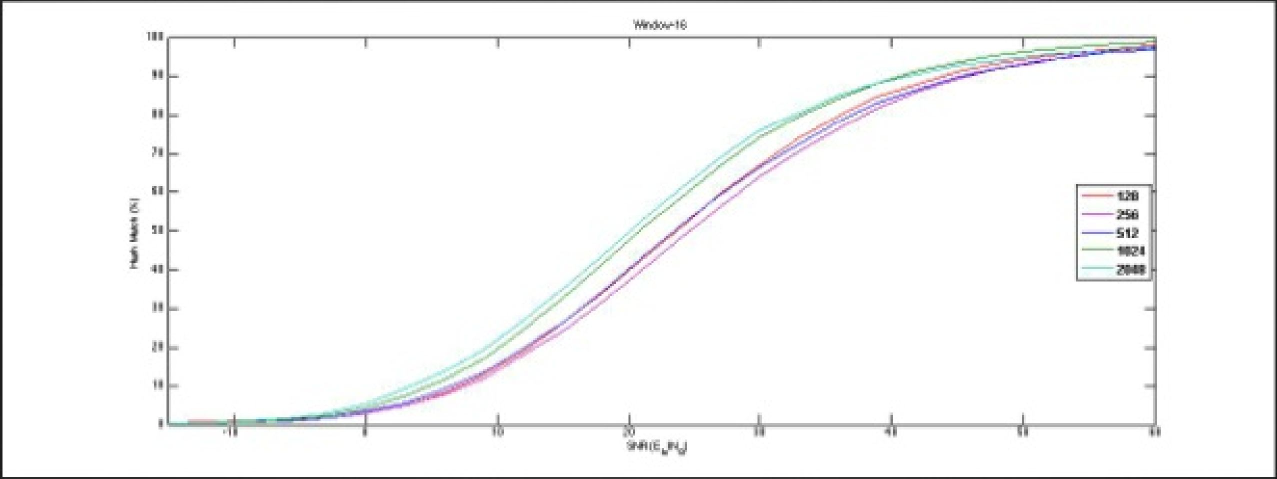
- Best choice?



Not Comparable Information Results

Window Size

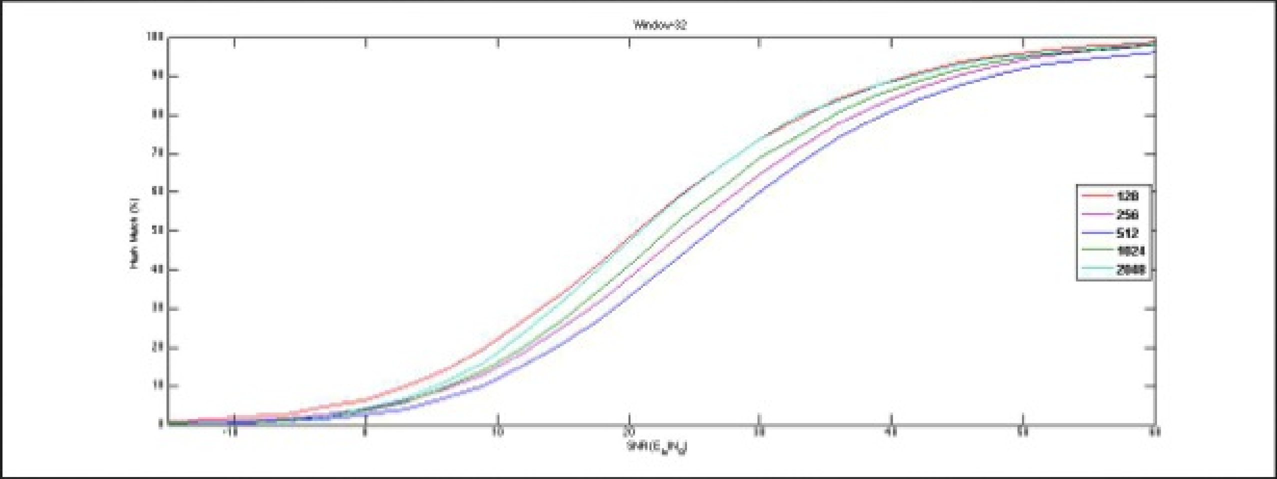
- Best choice?



- Second Attempt: Fixed Hop Size – Variable Window

Window Size

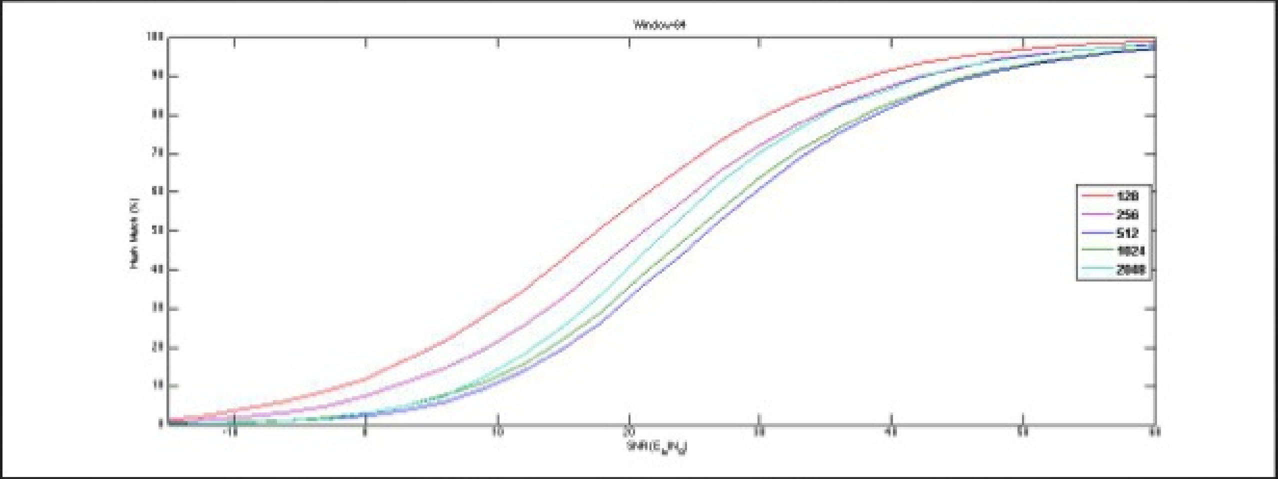
- Best choice?



- Second Attempt: Fixed Hop Size – Variable Window

Window Size

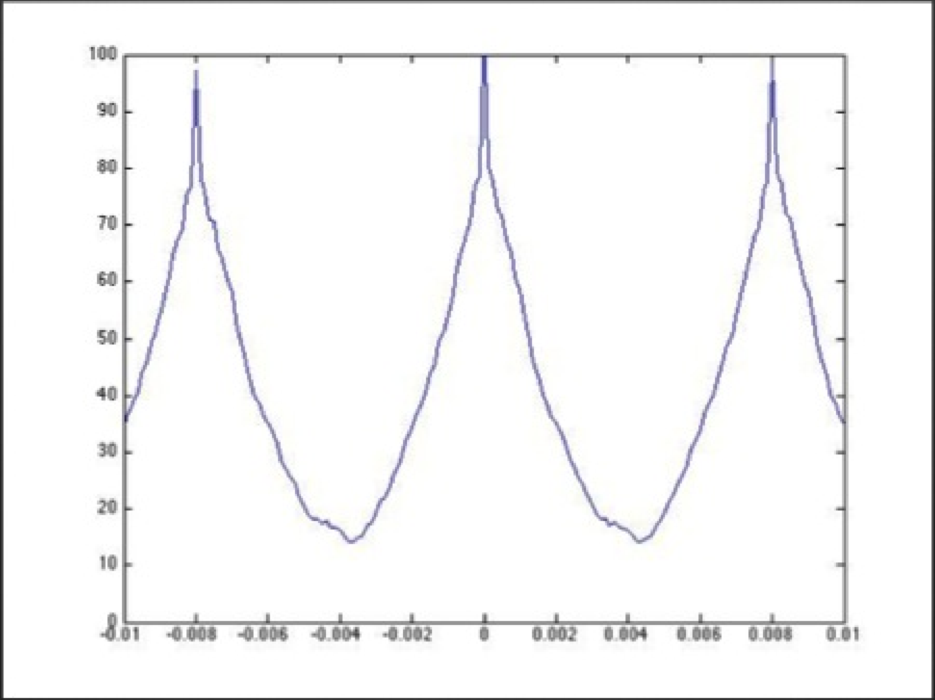
- Best choice?



Overlap = Window – 64

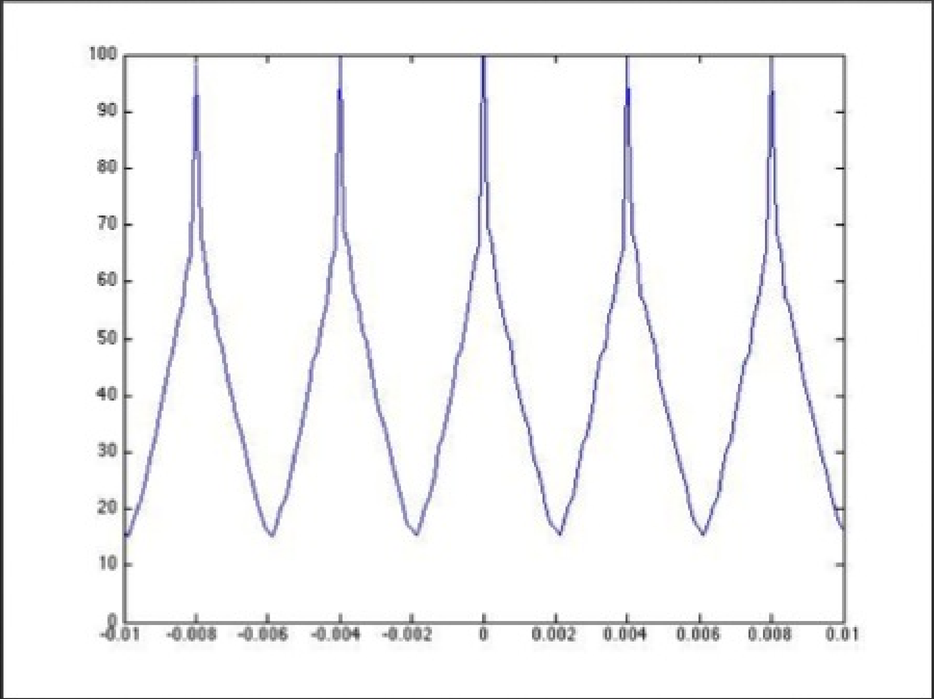
Is the best compromise between Accuracy and Computational Cost

Offset Analysis

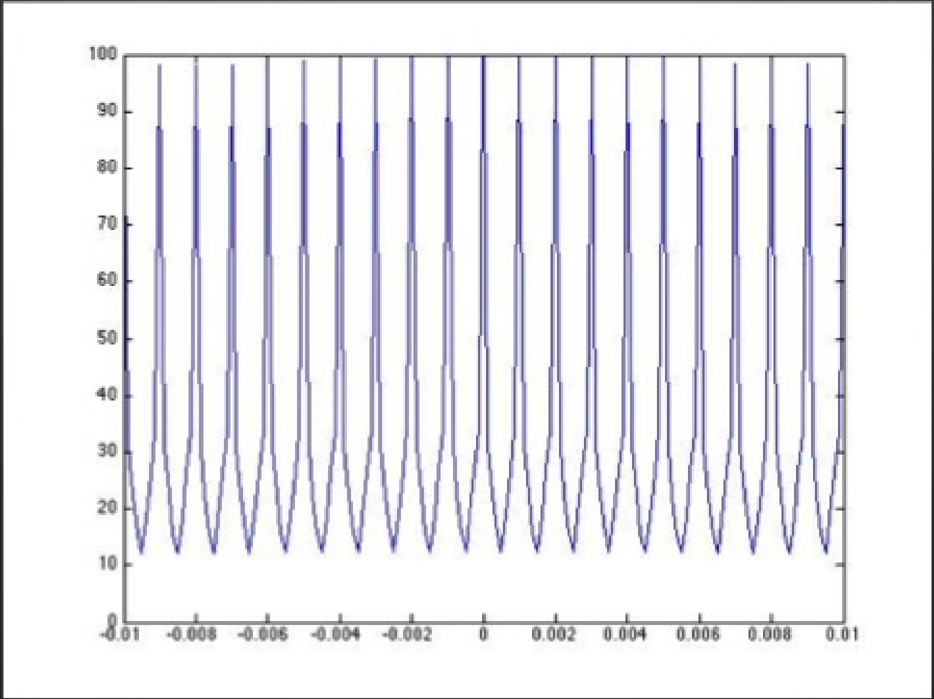


NO AWGN

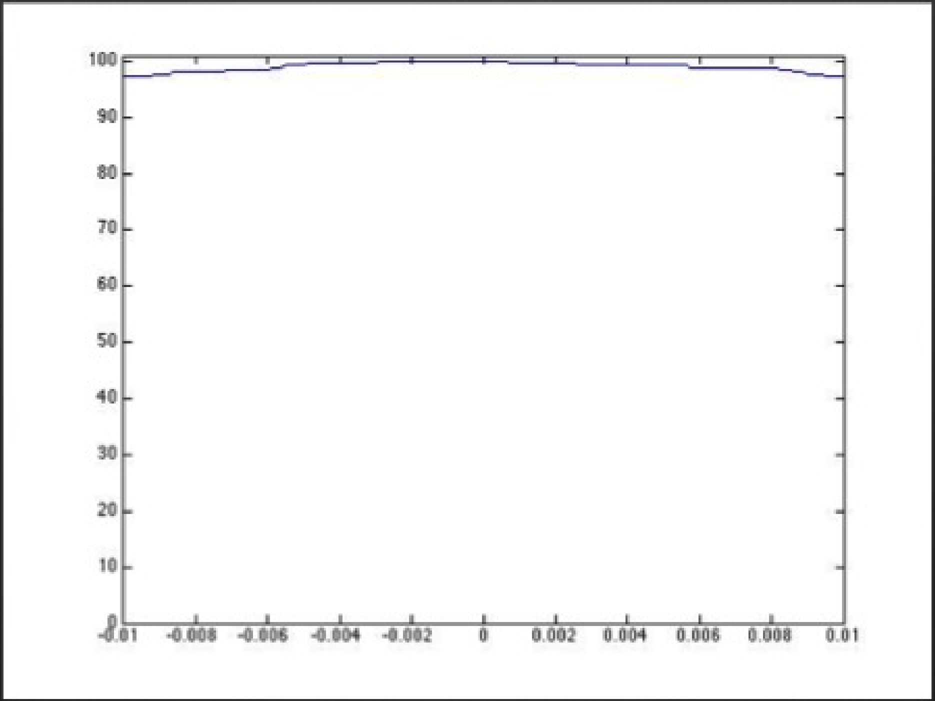
Offset Analysis



Offset Analysis

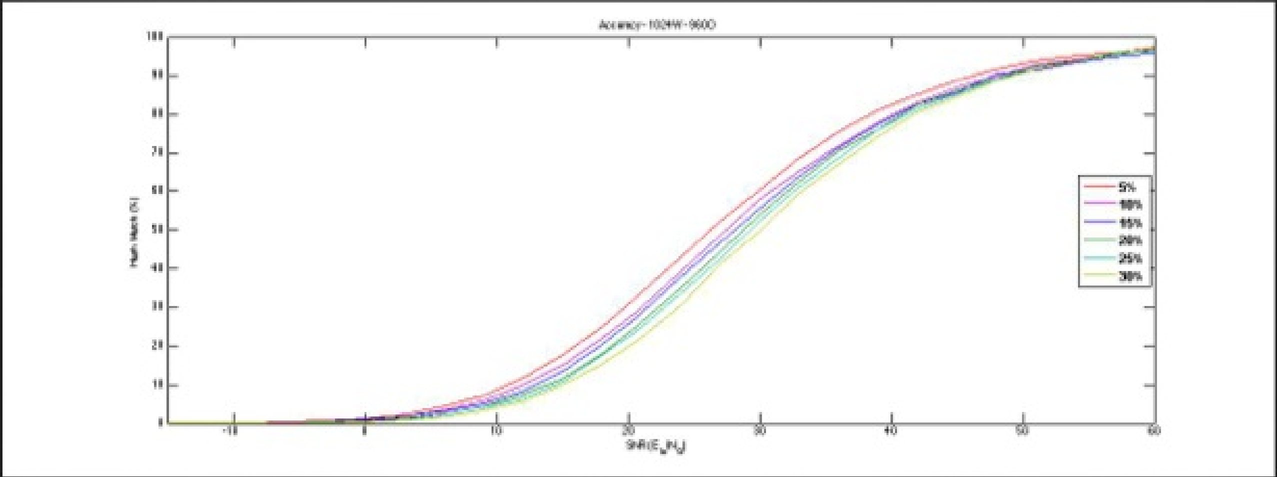


Offset Analysis



AWGN Analysis

- Two unknowns: Speaker TF + Microphone TF



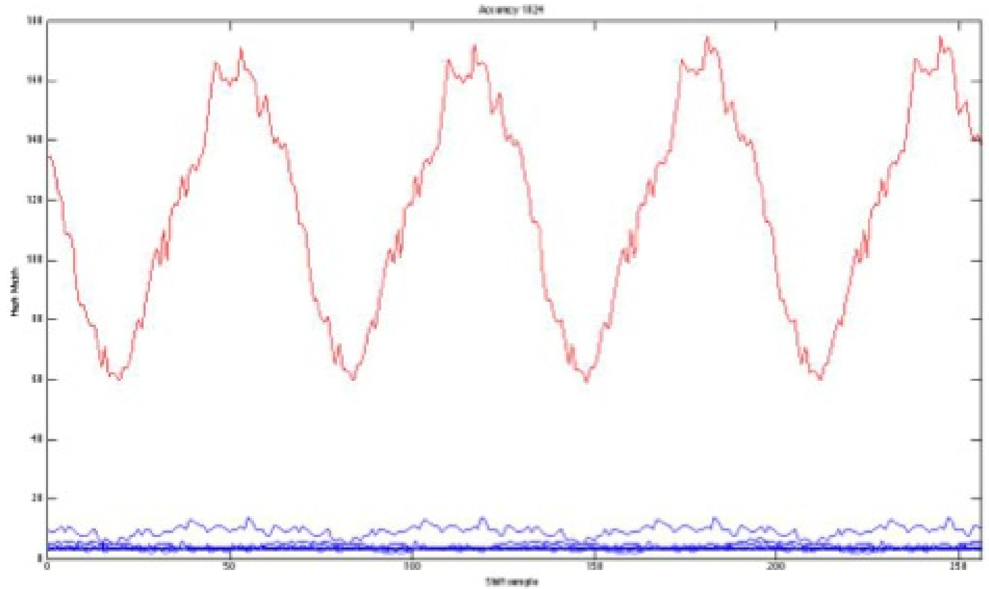
- Solution: NOT ALL F-Domain needs to be considered!

Live Recording Results

- TEST SCENARIO:
  - 1 or 2 Mac Speaker System/s
  - Cellphone Microphone
  - Air conditioning system working

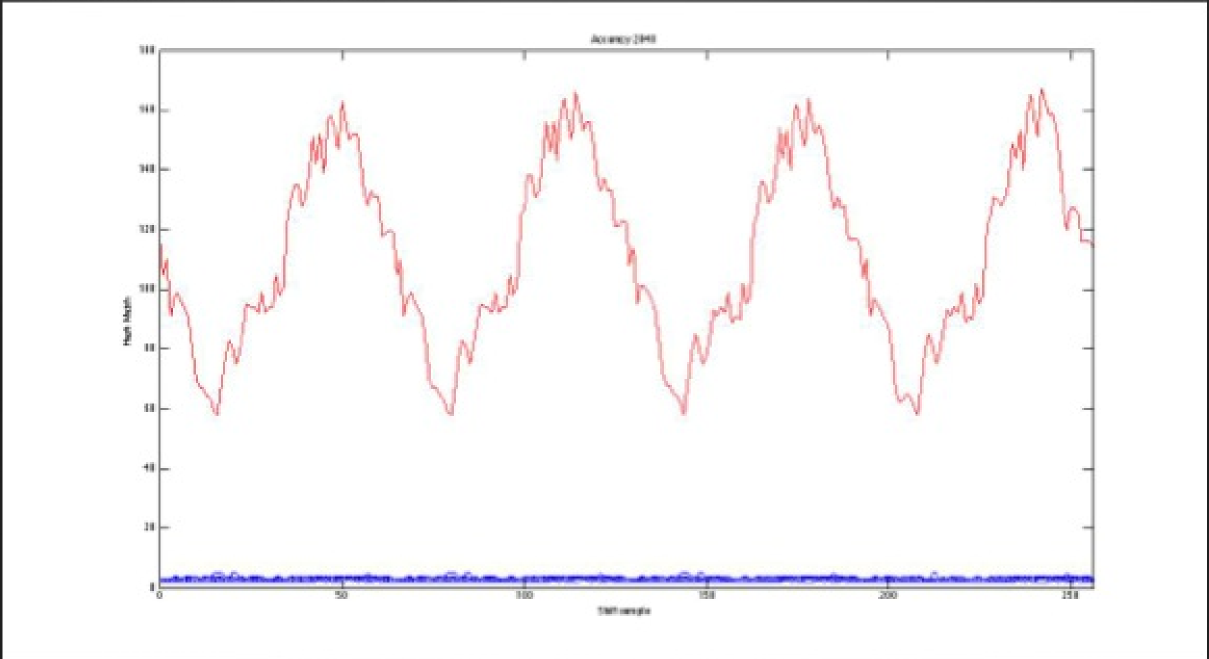
Live Recording Results

- CASE #1: Song Recorded Without External Interference
- 1024 window – 960 overlap



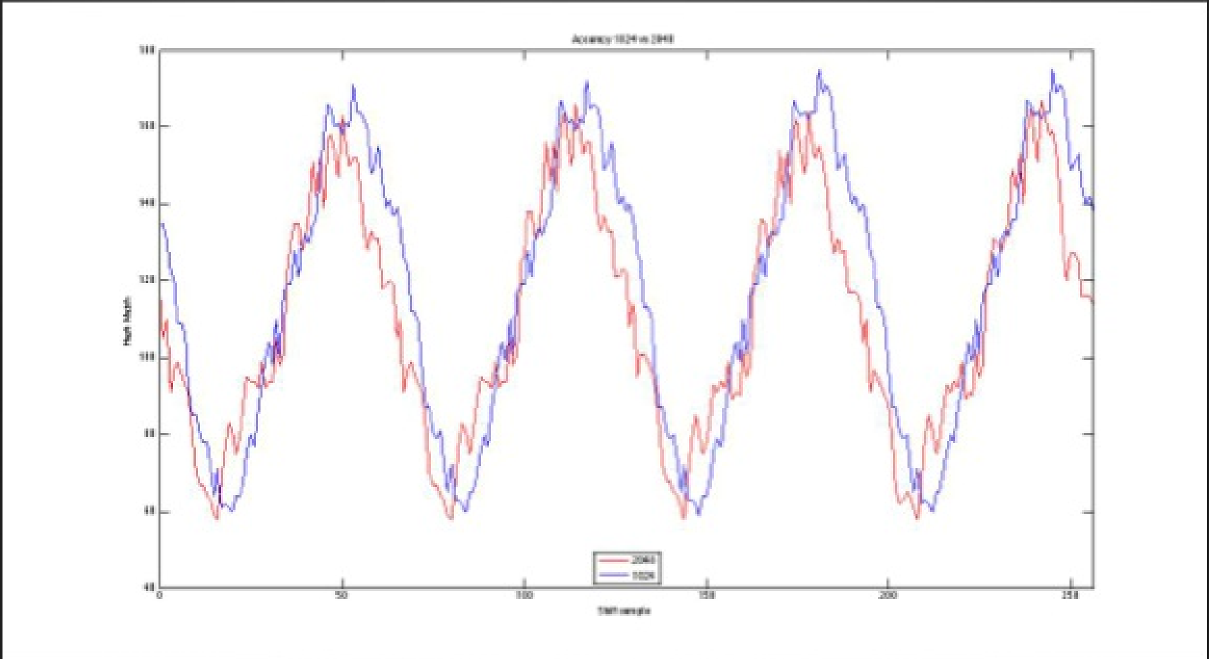
Live Recording Results

- Song Recorded Without External Interference
- 2048 window – 1984 overlap



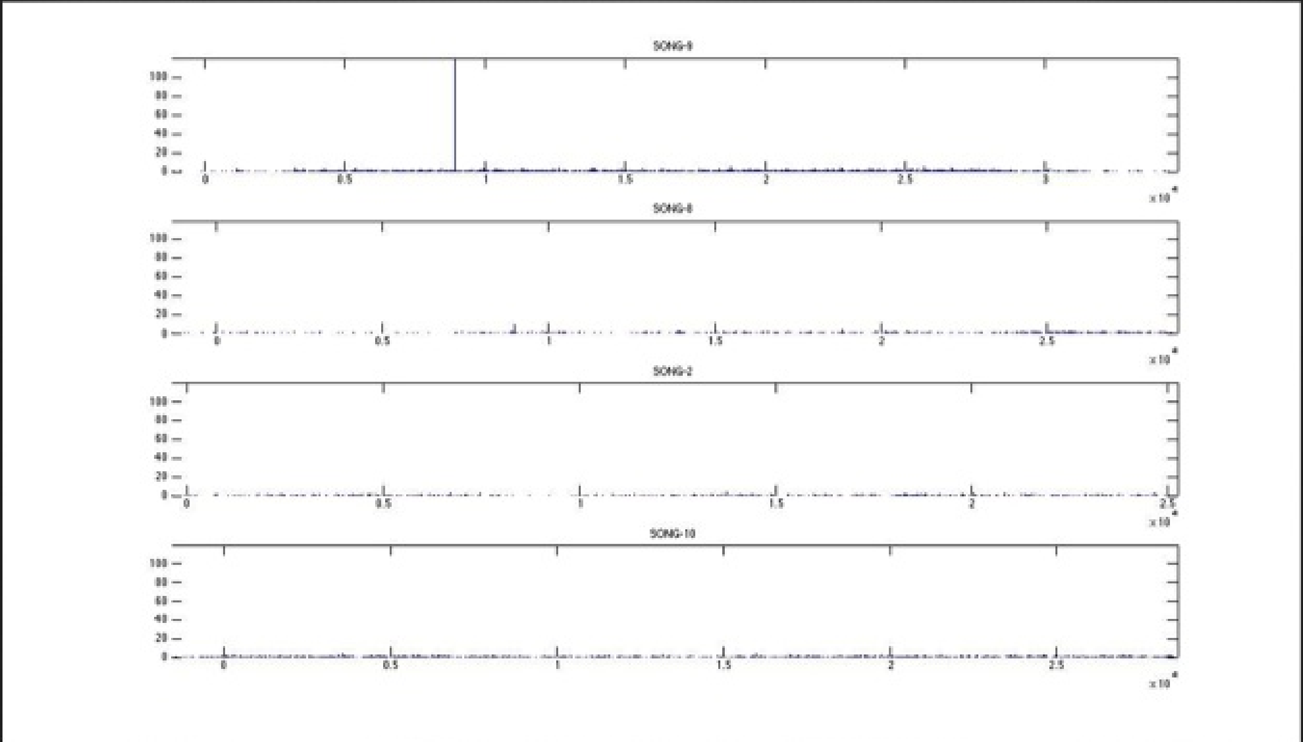
Live Recording Results

- 2048 vs 1024 window



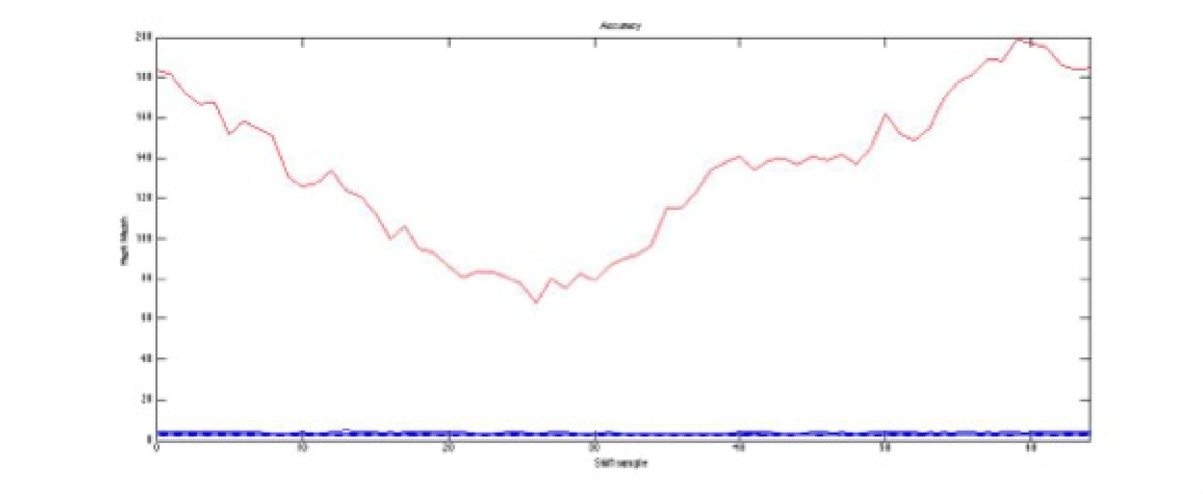
Live Recording Results

- Discrimination in DB



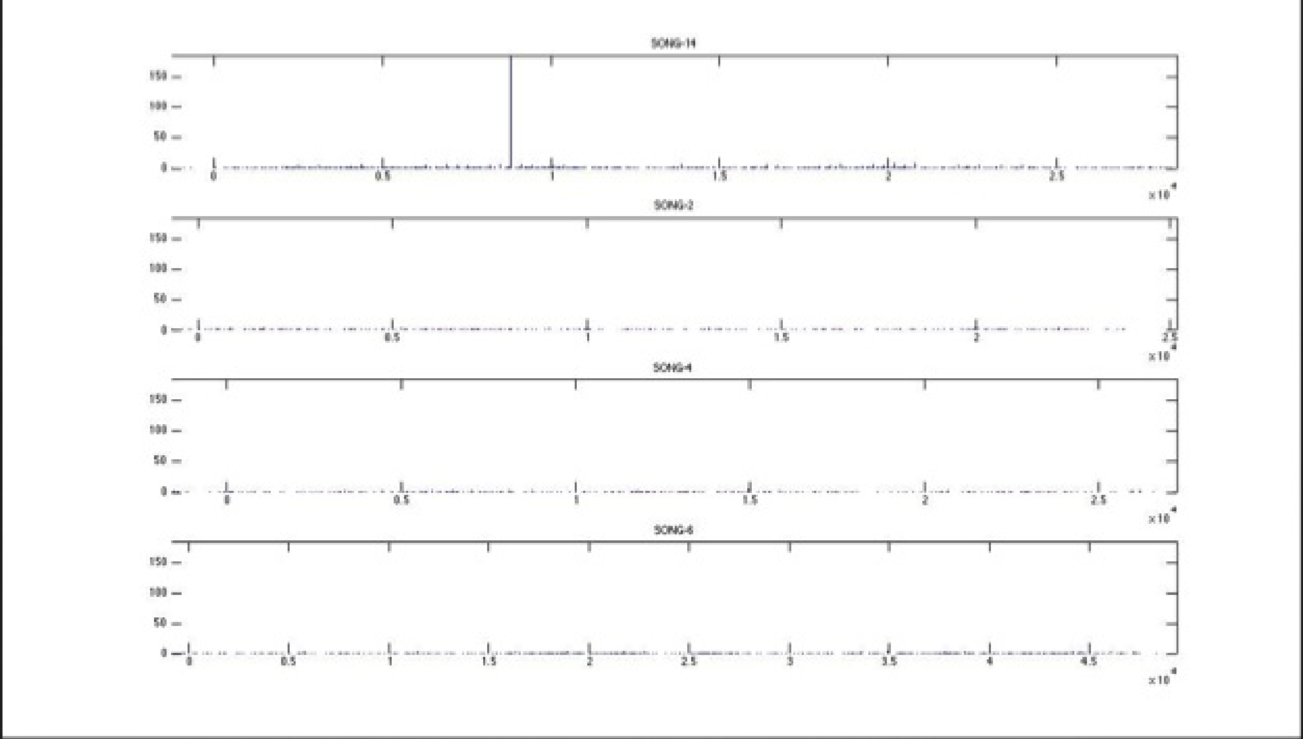
Live Recording Results

- CASE #2: Song Recorded With External Interference
- 2048 window – 1984 overlap



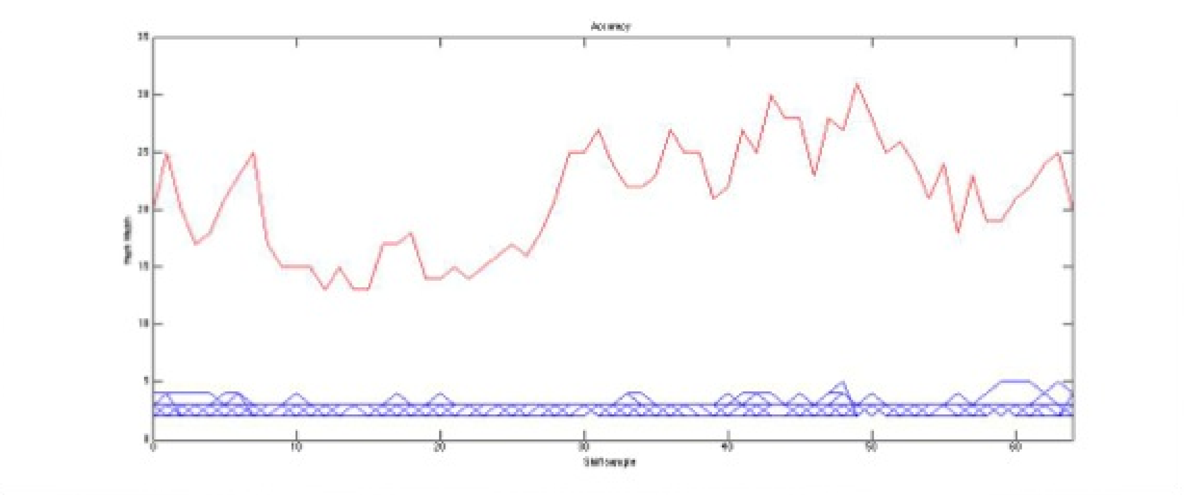
Live Recording Results

- Discrimination in DB



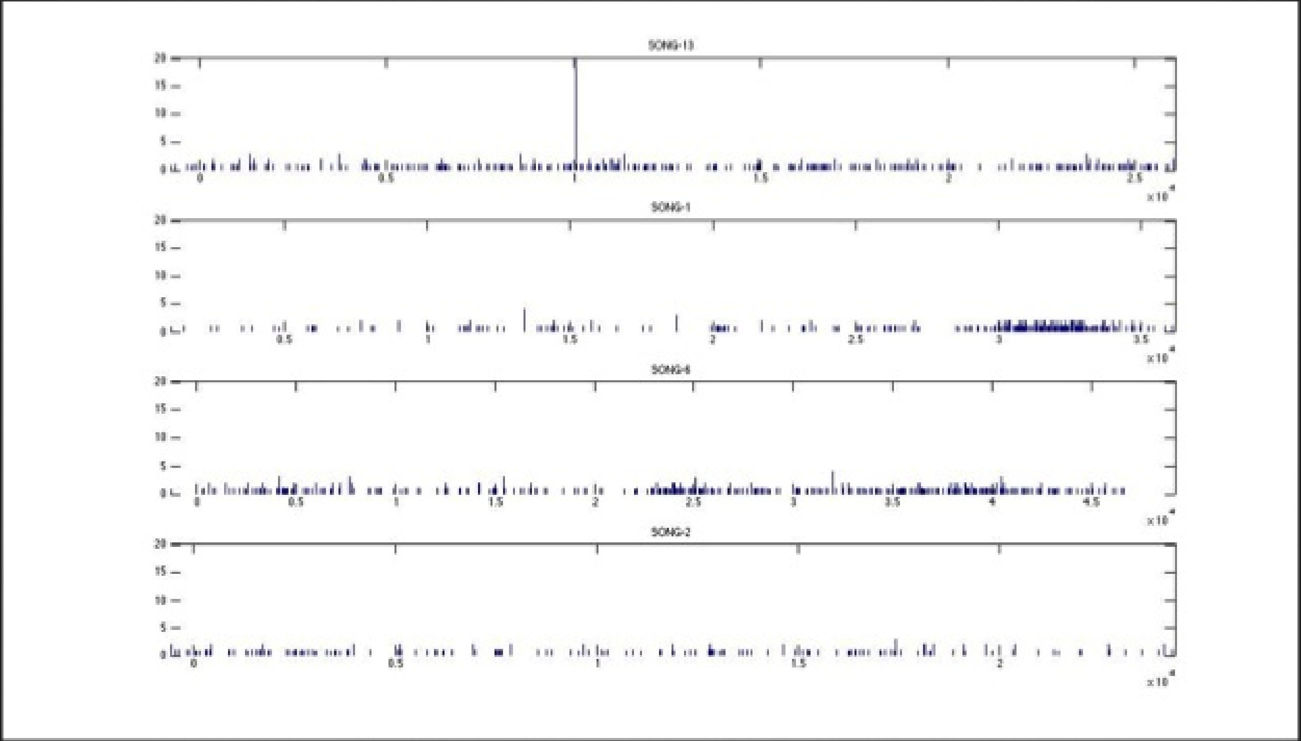
Live Recording Results

- CASE #3: Speaker 1 playing a DB song – Speaker 2 playing a NOT DB song
- 2048 window – 1984 overlap



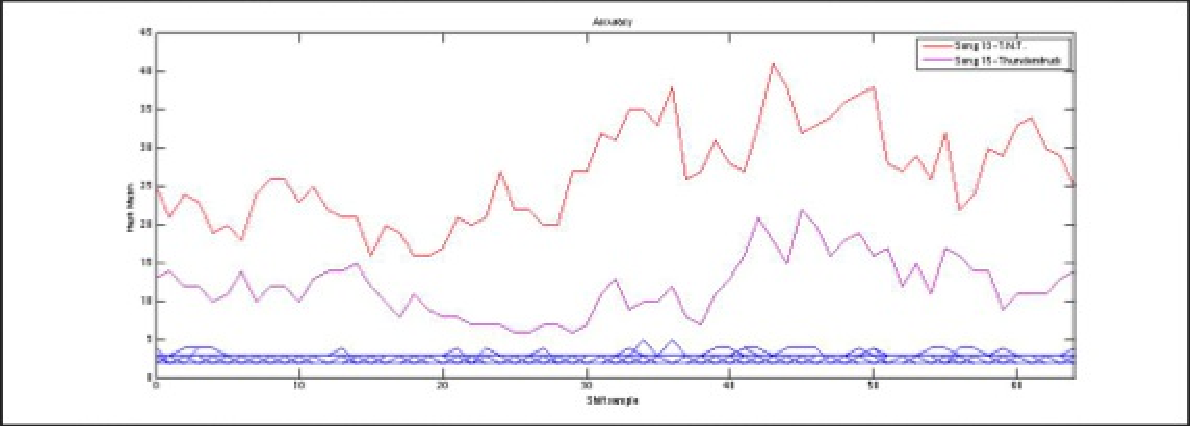
Live Recording Results

- Discrimination in DB



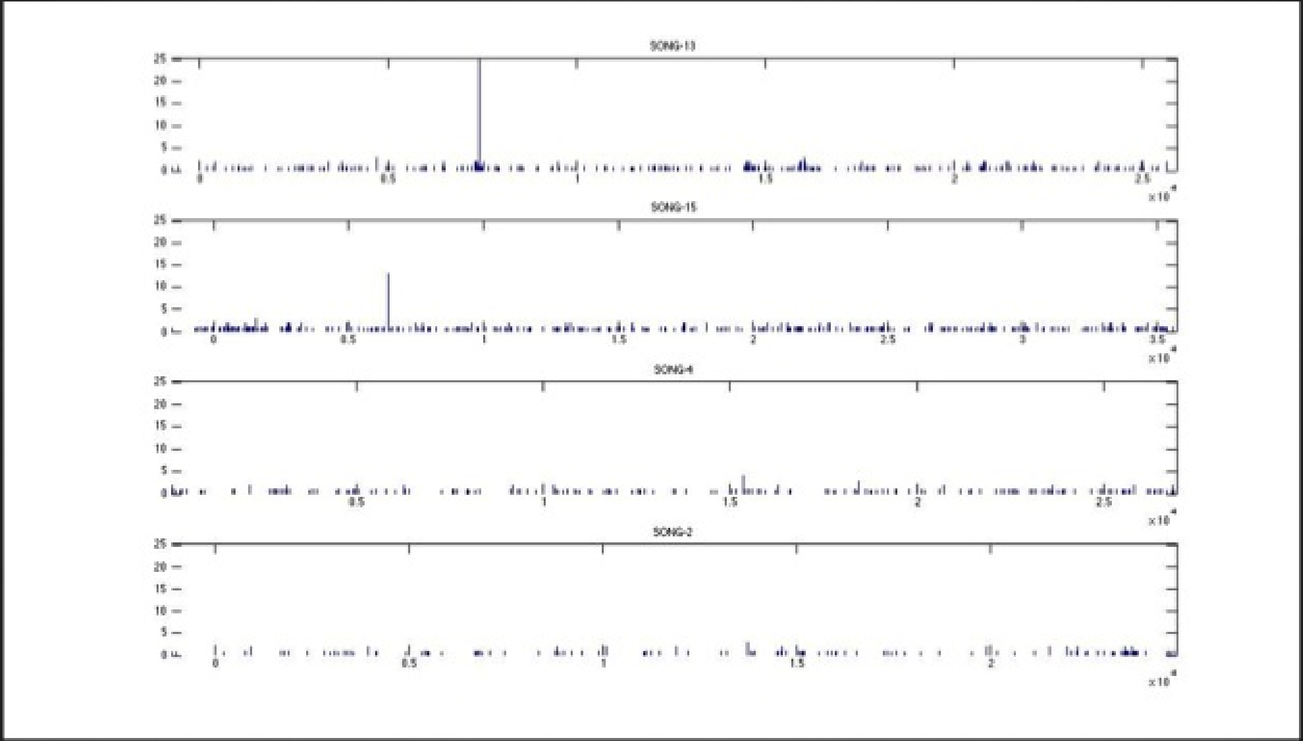
Live Recording Results

- CASE #4: Both Speaker playing 2 DIFFERENT DB SONGS
- 2048 window – 1984 overlap



Live Recording Results

- Discrimination in DB



Conclusions

- PROS:
  - Simplicity
  - Effectiveness/Reliability

- CONS:
  - Needs optimization with large DB
  - Could be useful try different peak localization techniques

Future Work

- Integration/Expansion to Cover-Song Recognition Systems
- C/C++ Implementation
- Search Technique Improvement (SQL, metric distance, ecc…)
- Client/Server Implementation
- Mobile Application Implementation

- Commercialization…

SGHAZAM

vanera

azzoli

Marco Gazzoli
Michele Svanera

Audio Fingerprinting and Recognition System

Università Degli Studi Di Brescia
A.A. 2012-2013

Bibliography

- "An Industrial-Strength Audio Search Algorithm", *Avery Li-Chun Wang, Shazam Entertainment, Ltd.*

- "A Review of Algorithms for Audio Fingerprinting", *Pedro Cano and Eioi Batlle, Ton Kalker and Jaap Haitsma*

- "Fingerprinting to identify repeated sound events in long-duration personal audio recordings", *James P. Ogle and Daniel P.W. Ellis, Columbia University*

- "A Highly Robust Audio Fingerprinting System", *Jaap Haitsma and Ton Kalker*