Conception and implementation of a test system for automated tests of audio outputs with the method of acoustic fingerprinting

by

Prof. Dr.-Ing. Klaus Solbach
University of Duisburg-Essen
Department of Microwave and RF-Technology

Ali Sen
University of Duisburg-Essen
Department of Microwave and RF-Technology

Dipl.-Ing. Michael Reiner
Nokia Automotive Business Line
R&D Center Bochum
Contents

• Introduction
• Overview of the Test System
• Development Process
• Audio Fingerprinting Theory
• Function of Audio Fingerprinting
• Implementation in Nokia Test Tool
• Result of Tests with a Car kit
• Conclusion
Introduction

What is an audio fingerprint?

- Small distinguishing feature of an audio signal
- Independent from artist or song name (Metadata)
- Consists of the main perceptual properties of the audio signal

What are the requirements for audio fingerprint comparison:

- Only the plain audio is needed
- Reliability
- Real-time audio comparison
- Support for short audio samples
- Support for both music and speech commands
- Limited implementation effort
Overview of the Test System

- **Passed / Failed**
- **Test PC**
- **Soundcard**

**Windows XP**
- CAN
- Power
- Simple Thought
- Audio Fingerprint
  - Generate
  - Compare

**CANoe**
Car simulation environment

**Com-port**

**CAN Bus**
Wiring loom for car simulation

**Audio output**

**Bluetooth**
Development Process

First stage of development:
- Used existing fingerprint theory and create fingerprint program in “MATLAB “

Second stage of development:
- Use “MATLAB” automatic C++ code generation and implement the fingerprint program in “Visual Studio”

Last stage of development:
- Implementation of the fingerprint program in “Nokia test tool”
Contents

• Introduction
• Overview of the Test System
• Development Process
• **Audio Fingerprinting Theory**
  • Music Identification using Audio Fingerprinting
  • Overview of the Fingerprint Algorithm
  • The Pro and Cons
• Function of Audio Fingerprinting
• Implementation in Nokia Test Tool
• Result of Tests with a Car kit
• Conclusion
Music Identification using Audio Fingerprinting

**Music identification process:**

1. Create audio fingerprint from audio file

2. Read second audio signal from soundcard and extract fingerprint

3. Compare both fingerprints
Overview of the Fingerprinting Algorithm

Front-End from chapter 3.2

1. Converting to mono (8bit)
2. Resampling to 5000KHz
3. Windowing by „Hanning Window“
4. Fast Fourier Transformation (FFT)
5. Absolute value (ABS) (Amplitude)
6. Divide the generated spectrum into Bark bands
7. Energy computation and energy differences
The Pros and Cons

Disadvantages:
• Reacting to single sinusoidal sounds and DTMF tones

Advantages of this fingerprinting scheme:
• Adapted for all types of audio signals (speech, music, etc.)
• High accuracy regardless of small disturbance in recordings
• Independent from amplitude of the audio signal
• Easy “MATLAB”-implementation
• Good descriptive documents
Contents

• Introduction
• Overview of the Test System
• Development Process
• Audio Fingerprinting Theory
  • Function of Audio Fingerprinting
    • Optimisation of Audio Fingerprint Comparison
    • Results of Comparison with different Settings
    • Audio Fingerprint Comparison I
    • Real-time Audio Matching
    • Possible Applications
    • DTMF Tones
      • DTMF Tone Detection
        • Implementation in Nokia Test Tool
        • Result of Tests with a Car kit
  • Conclusion
Optimisation of Audio Fingerprint Comparison
## Results of Comparison with Different Settings

### Final result:

<table>
<thead>
<tr>
<th>Settings</th>
<th>Audio samples</th>
<th>Speech commands</th>
<th>Music’s like Pop or Techno</th>
<th>Classics</th>
<th>Sinusoidal tones</th>
<th>DTMF tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td></td>
<td>92.13%</td>
<td>95.83%</td>
<td>98.50%</td>
<td>1.37%</td>
<td>3.65%</td>
</tr>
</tbody>
</table>

### With different setup:

<table>
<thead>
<tr>
<th>Settings</th>
<th>Audio samples</th>
<th>Speech commands</th>
<th>Music’s like Pop or Techno</th>
<th>Classics</th>
<th>Sinusoidal tones</th>
<th>DTMF tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduce fingerprint size (N = 128)</td>
<td></td>
<td>92.13%</td>
<td>95.96%</td>
<td>98.96%</td>
<td>2.15%</td>
<td>2.47%</td>
</tr>
<tr>
<td>Increase fingerprint size (N = 384)</td>
<td></td>
<td>92.13%</td>
<td>96.05%</td>
<td>98.00%</td>
<td>1.82%</td>
<td>4.95%</td>
</tr>
<tr>
<td>Sample rate 3000 Hz</td>
<td></td>
<td>83.81%</td>
<td>84.62%</td>
<td>86.02%</td>
<td>1.52%</td>
<td>1.89%</td>
</tr>
<tr>
<td>Sample rate 4000 Hz</td>
<td></td>
<td>94.53%</td>
<td>94.01%</td>
<td>97.27%</td>
<td>2.91%</td>
<td>2.99%</td>
</tr>
<tr>
<td>Sample rate 7500 Hz</td>
<td></td>
<td>89.49%</td>
<td>1.80%</td>
<td>92.90%</td>
<td>3.15%</td>
<td>3.65%</td>
</tr>
<tr>
<td>Window size 1024</td>
<td></td>
<td>89.68%</td>
<td>92.71%</td>
<td>96.88%</td>
<td>2.65%</td>
<td>3.65%</td>
</tr>
<tr>
<td>Window size 4096</td>
<td></td>
<td>95.91%</td>
<td>95.96%</td>
<td>98.44%</td>
<td>3.56%</td>
<td>1.56%</td>
</tr>
<tr>
<td>Overlapping factor 32</td>
<td></td>
<td>92.42%</td>
<td>96.35%</td>
<td>98.63%</td>
<td>2.55%</td>
<td>2.15%</td>
</tr>
<tr>
<td>Overlapping factor 128</td>
<td></td>
<td>76.52%</td>
<td>79.21%</td>
<td>80.15%</td>
<td>5.54%</td>
<td>5.94%</td>
</tr>
<tr>
<td>Hamming window</td>
<td></td>
<td>92.70%</td>
<td>95.57%</td>
<td>98.57%</td>
<td>2.15%</td>
<td>1.56%</td>
</tr>
<tr>
<td>Blackman window</td>
<td></td>
<td>89.33%</td>
<td>95.31%</td>
<td>98.44%</td>
<td>1.22%</td>
<td>1.82%</td>
</tr>
<tr>
<td>Triangle window</td>
<td></td>
<td>91.76%</td>
<td>91.08%</td>
<td>97.85%</td>
<td>1.78%</td>
<td>3.78%</td>
</tr>
<tr>
<td>Only 10 bands</td>
<td></td>
<td>91.01%</td>
<td>95.75%</td>
<td>98.70%</td>
<td>0.95%</td>
<td>1.82%</td>
</tr>
<tr>
<td>Only 5 bands</td>
<td></td>
<td>85.96%</td>
<td>92.19%</td>
<td>93.75%</td>
<td>2.99%</td>
<td>3.32%</td>
</tr>
<tr>
<td>Vector elements 1 &amp; -1</td>
<td></td>
<td>96.05%</td>
<td>96.86%</td>
<td>98.98%</td>
<td>3.29%</td>
<td>3.68%</td>
</tr>
</tbody>
</table>
Audio Fingerprint Comparison I

Preparations of the audio file to improve the comparison results:

Audio files with silent ending:
Delete the silent part at the beginning and end of audio file

Extend to short audio files:
After resampling the window size of a short audio file is under 2048 of window size

Extract fingerprint at each part:
For tests with forwarding
Real-time Audio Matching

Three different settings for real-time audio matching:

1. **Best setting:**
   Fingerprint size = 256
   (Need higher CPU speed)

2. **Optimal settings:**
   Fingerprint size = 128

3. **Fastest settings:**
   Fingerprint size = 128
   (Only 5 bands for reduce the buffer)

---

**Buffer for fingerprint size 128 (1.85 sec)**

128 (Fingerprint size) x 64 (overlap) = 8192 (Samples)

8192 (Samples) x (8000 Hz / 4000 Hz) = 16384 (Samples)

16384 (Samples) / 3 = 5461 (Samples) Buffer size

1,2,3,...,128 Fingerprint size

---

**Buffer for fingerprint size 256 (3.6 sec) = 8192 (Samples) Buffer size**
Possible Applications:

- Validation of voice recordings (no voice recognition!)
- Comparison of voice commands like phone numbers (0454418156)
- Multiple fingerprints for one audio file (different positions) to check different sound playback features (e.g. fast forward)
- Search audio in database
DTMF Tones

697 Hz Sine Wave + 1209 Hz Sine Wave = DTMF Tone "1"

<table>
<thead>
<tr>
<th>DTMF keypad frequencies (with sound clips)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>697 Hz</td>
</tr>
<tr>
<td>770 Hz</td>
</tr>
<tr>
<td>852 Hz</td>
</tr>
<tr>
<td>941 Hz</td>
</tr>
</tbody>
</table>
DTMF Tone Detection

Band division for defined tone

Energy computation

Compute proportions

Tone A Detection

Tone A Detection

Framing → FFT → ABS

Tone A +86 Hz

Tone A +60 Hz

Tone A +20 Hz

Tone A -60 Hz

Tone A -80 Hz

Tone B +86 Hz

Tone B +60 Hz

Tone B +20 Hz

Tone B -60 Hz

Tone B -80 Hz

Tone B Detection

Defined Tone A 697 Hz

Defined Tone B 1209 Hz

Tone A

Tone B

>10

>10

>10

>10

>10

>10

>10

>10

>10

>10
Contents

- Introduction
- Overview of the Test System
- Development Process
- Audio Fingerprinting Theory
- Function of Audio Fingerprinting
- Implementation in Nokia Test Tool
  - Fingerprint Example in Manual Mode
  - Test Script
- Result of Tests with a Car kit
- Conclusion
Fingerprint example in Manual Mode

- Fingerprint file
- Quality settings
- Duration time in sec
# Test script

## Test script:

<table>
<thead>
<tr>
<th>#</th>
<th>Test script</th>
</tr>
</thead>
<tbody>
<tr>
<td># Filename:</td>
<td><strong>Fingerprint.csv</strong></td>
</tr>
<tr>
<td># ID:</td>
<td><strong>Fingerprint</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Label</th>
<th>Interface</th>
<th>Command</th>
<th>Parameter</th>
<th>Condition</th>
<th>Timeout</th>
<th>Appearance</th>
<th>EventOnError</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td># Example</td>
<td>AudioCmp</td>
<td>Preload</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Power</td>
<td>Setvoltage</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AudioCmp</td>
<td>Generate Fingerprint</td>
<td>C:\Music\Music12.wav,C:\Fingerprint\Music12.mat2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AudioCmp</td>
<td>Compare</td>
<td>C:\Fingerprints\Music12.mat</td>
<td>50 %</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**END** System NOP

---

**End of the Fingerprint script with no operation**
Contents

• Introduction
• Overview of the Test System
• Development Process
• Audio Fingerprinting Theory
• Function of Audio Fingerprinting
• Implementation in Nokia Test Tool
• Result of Tests with a Car kit
• Conclusion
Result of Tests with a Car kit

Tests:
- Play, Stop, Pause
- Repeat
- Intro scan
- Skip track
- Shuffle
- Fast forward and fast backward

<table>
<thead>
<tr>
<th>Recorded audio samples</th>
<th>Original audio samples</th>
<th>Techno</th>
<th>Pop</th>
<th>Classic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Techno</td>
<td></td>
<td>93.13%</td>
<td>8.37%</td>
<td>9.55%</td>
</tr>
<tr>
<td>Pop</td>
<td></td>
<td>11.89%</td>
<td>95.03%</td>
<td>10.36%</td>
</tr>
<tr>
<td>Classic</td>
<td></td>
<td>13.53%</td>
<td>7.46%</td>
<td>98.58%</td>
</tr>
</tbody>
</table>

Test results
- 90% Passed
- 10% Failed
Contents

• Introduction
• Overview of the Test System
• Development Process
• Audio Fingerprinting Theory
• Function of Audio Fingerprinting
• Implementation in Nokia Test Tool
• Tests with the Audio Identification System
• Conclusion
Conclusion

- Fingerprint can be extracted from audio file
- Audio signal can be identified
- Similarity rate of almost 90%
- Identification rate of exact 100%.
- Stability of the developed test system is given
- To advance a search and record function can be added in future
Thank you for your attention!