Microphone Identification
using
Higher-Order Statistics

Hafiz Malik, John Miller

Information Systems, Security, and Forensics (ISSF) Lab
Department of Electrical and Computer Engineering
University of Michigan – Dearborn
Outline

• Motivation
• Microphone Nonlinearity Modeling
• Higher-Order Spectral Analysis (HOSA)
• Nonlinear Identification using HOSA
  – Invariant Moments
• Experimental Results
• Conclusion
The goal of microphone forensics is to find answers to the following questions:

- Does a microphone leave a fingerprint in the recording; if so, how can we extract it?
- Does the microphone signature depend on the underlying transducer technology?
- Is a unique mapping of a recording to the source possible?

This work focuses on finding answers to these questions.
Microphone Forensics: State-of-the-Art

- Buchholtz et. al. (2010)
- Kreatzer et. al. (2009, 2007)
Microphone Distortion Modeling

- Microphone distortions can be classified into
  
  1. Harmonic distortion,
  2. Intermodulation distortion, and
  3. Difference-frequency distortion.

- The intermodulation effect produces an output signal made of sums and differences of the input signals fundamental frequencies and their harmonics, that is,

\[ \omega_2 \pm \omega_1, \omega_2 \pm 2\omega_1, \omega_2 \pm 3\omega_1, \text{ etc.} \]
The microphone response can be approximated using the following discrete time-invariant Hammerstein series model,

\[ y[n] = \sum_{k} g_1[k] x[n-k] + \sum_{k} g_2[k] x^2[n-k] + \sum_{k} g_3[k] x^3[n-k] + \cdots \]

where \( g_i[k] \): \( i=1, 2, 3 \) characterize linear, quadratic, and cubic response of the microphone.
Higher-Order Spectral Analysis to Capture Mic. Artifacts

• The HOSA can be used to analyze the nonlinearity of a system operating under a random input.

• Why HOSA?

• The HOSA not only reveals the amplitude information, but also the phase information about the underlying process.
  – The correlation based methods are phase blind.
  – The cumulant-based methods are blind to any kind of a Gaussian process, whereas, correlation is not.

  ➔cumulant-based methods are capable of handling nonlinear and non-Gaussian processes more effectively than the correlation-based methods (Nikias & Petropulu, 1993).
HOSA: Example

- Consider input signal
  \[ x[n] = c_1 \cos[\omega_1 n + \phi_1] + c_2 \cos[\omega_2 n + \phi_2] \]

and second-order nonlinearity, i.e.,

\[ y[n] = x[n] + \alpha x^2[n] \quad 1 \geq \alpha > 0 \]

This nonlinearity introduces harmonics with strongly correlated phases and magnitudes, i.e.,

\[(2\omega_1, 2\phi_1), (2\omega_2, 2\phi_2), (\omega_1 + \omega_2, \phi_1 + \phi_2) \text{ and } (\omega_1 - \omega_2, \phi_1 - \phi_2)\]

**Note:**
The quadratic nonlinearity results in quadratic phase coupling (QPC).
The QPC an interaction between two harmonic components causing contribution to the power at their sum and/or difference frequencies.
Bicoherence

• **The bicoherence**, a normalized bispectrum (the Fourier transform of the third-order cumulant or bicorrelation) is used to capture for nonlinearities due to microphone.

• The **bicoherence**, \( b_y^3(\omega_1, \omega_2) \), is defined as,

\[
b_y^3(\omega_1, \omega_2) = \frac{C_y^3(\omega_1, \omega_2)}{\sqrt{Y(\omega_1)Y(\omega_2)} \sqrt{Y(\omega_1 + \omega_2)}}
\]

where

\[
C_y^3(\omega_1, \omega_2) = E\{Y(\omega_1)Y(\omega_2)Y^*(\omega_1 + \omega_2)\}
\]

\[Y(\omega) = F(y(t))\]

here
\( F(.) \) is the Fourier Transform operator, \( E\{.\} \) is the expected value operator, and * denotes complex conjugate.
Bicoherence Mag. Spec. of Ambient Noise

$M_5$: Behringer EMC8000 - 1  
Bicoherence Magnitude Plots  
$M_6$: Behringer EMC8000 - 2

$\omega_2$  
$\omega_1$

1$^{st}$ Recording  
2$^{nd}$ Recording  
3$^{rd}$ Recording

46th AES Conference
Audio Forensics
Recording, Recovery, Analysis, and Interpretation
14 - 16 June 2012  
Denver, CO, USA  
College of Engineering  
& Computer Science  
Dearborn
Bicoherence Ph. Spec. of Ambient Noise

M_5: Behringer EMC8000 - 1

Bicoherence Phase Plots

M_5: Behringer EMC8000 - 2

1^{st} Recording

2^{nd} Recording

3^{rd} Recording

\omega_1
Feature Extraction

- It can be observed that bicoherence magnitude plots exhibit 12 regions of symmetry.
- Image moments are commonly used to characterize images based on scale, centroid, and orientation.
- The *scale invariant Hu moments* are used to detect nonlinearities due to microphone.
Feature Extraction:

*Scale Invariant Hu Moments*

- The scale invariant Hu moments \( \eta_{i,j} \) where \( i + j \geq 2 \) is computed as,

\[
\eta_{i,j} = \frac{\mu_{i,j}}{\mu_{0,0}^{1 + \frac{i+j}{2}}}
\]

here, the central moment \( \mu_{i,j} \) can be computed as,

\[
\mu_{i,j} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j b(x, y)
\]

where \( \bar{x} \) and \( \bar{y} \) are the components of the centroid of the bicoherence magnitude spectrum, \( b(.,.,.) \)
Data Set

- Three sets of recordings were made using eight microphones and an eight-channel USB multitrack recorder.

- The multitrack recorder, a Zoom R16, was powered via the USB connection and recordings were saved on a PC running Windows 7.

- A 12-inch fan approximately 0.7 meter from the microphones operating at maximum speed was used as the sound source.

- The microphones were directed towards the back of the fan so as to minimize any effects from wind noise.

- Recordings were made at a level of approximately -45 dB for each channel with a duration of about one minute and about ten minutes between recordings.
## List of the Microphones Used

<table>
<thead>
<tr>
<th>Mic. ID</th>
<th>Microphone Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1$</td>
<td>Samson R19 Dynamic Mic.-1</td>
</tr>
<tr>
<td>$M_2$</td>
<td>Samson R19 Dynamic Mic.-2</td>
</tr>
<tr>
<td>$M_3$</td>
<td>Radio Shack Electret Mic.-1</td>
</tr>
<tr>
<td>$M_4$</td>
<td>Radio Shack Electret Mic.-2</td>
</tr>
<tr>
<td>$M_5$</td>
<td>Behringer EMC8000 Measurement Mic.-1</td>
</tr>
<tr>
<td>$M_6$</td>
<td>Behringer EMC8000 Measurement Mic.-2</td>
</tr>
<tr>
<td>$M_7$</td>
<td>Zoom R16 Recorder Built-in Mic.-Left</td>
</tr>
<tr>
<td>$M_8$</td>
<td>Zoom R16 Recorder Built-in Mic.-Right</td>
</tr>
</tbody>
</table>
Parameter Settings

- Each recording is segmented into frames of four seconds duration with a 50% overlapping factor.

- Bicoherence is estimated from each audio segment using the direct (fft-based) approach (Nikias & Petropulu, 1993).

- The bicoherence is estimated with the following parameter settings:
  1. 128-point segment length,
  2. 256-point FFT length,
  3. no overlap, and
  4. Rao-Gabr optimal window for frequency domain smoothing.
Experiment 1: Classification using Phase Similarity
Inter-Class Phase Variability

XCorr. between Ph. Spectra of 1st Recording of M_6 and 1st Recording of M_1

XCorr. between Ph. Spectra of 1st Recording of M_6 and 1st Recording of M_2

XCorr. between Ph. Spectra of 1st Recording of M_6 and 1st Recording of M_3

XCorr. between Ph. Spectra of 1st Recording of M_6 and 1st Recording of M_4

XCorr. between Ph. Spectra of 1st Recording of M_6 and 1st Recording of M_7

XCorr. between Ph. Spectra of 1st Recording of M_6 and 1st Recording of M_8
Inter-Class Phase Variability

XCorr. between Ph. Spectra of 1\textsuperscript{st} Recording of $M_6$ and 2\textsuperscript{nd} Recording of $M_5$

XCorr. between Ph. Spectra 1\textsuperscript{st} and 2\textsuperscript{nd} Recordings of $M_6$

XCorr. between Ph. Spectra of 1\textsuperscript{st} Recording of $M_6$ and 3\textsuperscript{rd} Recording of $M_5$

XCorr. between Ph. Spectra of 1\textsuperscript{st} and 3\textsuperscript{rd} Recordings of $M_6$
Experiment 2: Classification using Magnitude Similarity
Scatter Plot of Scale Invariant Hu Moments: M1 & M2

$M_1$: Samson R19 - 1

$M_2$: Samson R19 - 2

$\mu_{3,0}$

$\mu_{1,1}$

$\mu_{2,0}$

$\mu_{1,1}$
Scatter Plot of Scale Invariant Hu Moments:

M3 & M4

M₃: Radio Shack Electret - 1

1ˢᵗ Recording

M₄: Radio Shack Electret - 2

2ⁿᵈ Recording

3ʳᵈ Recording

μ₃,₀

μ₂,₀

μ₁,₁
Scatter Plot of Scale Invariant Hu Moments: M5 & M6

M₅: Behringer EMC8000 - 1

M₆: Behringer EMC8000 - 2

1ˢᵗ Recording

2ⁿᵈ Recording

3ʳᵈ Recording
Scatter Plot of Scale Invariant Hu Moments:
M7 & M8
Limitations

• Not applicable to speech data
• Audio artifacts can be modeled and estimated.

• Microphone artifacts are useful in forensics

• Combine with reverberation, ambient (background) noise, transcoding artifacts, ENF, etc.
Future Directions

• Investigate robustness against lossy compression, transcoding, additive noise attacks

• Investigate the HOS of invariant moments of the magnitude spectrum

• Extend this framework for speech

• Investigate uniqueness of microphone fingerprints
Questions/Comments?