



Tampering Detection On Digital Audio Using Gabor Filterbank

Fajri Kurniawan¹, Mohammed S. Khalil¹, Muhammad Khurram Khan¹, Hafiz Malik²

¹Center of Excellence Information Assurance
King Saud University
Riyadh, Saudi Arabia
fkurniawan.c@ksu.edu.sa
sayimkhalil@gmail.com
mkhurram@ksu.edu.sa

²Electrical and Computer Engineering Department
University of Michigan
Dearborn, Dearborn, MI 48128, USA
hafiz@umich.edu

CoEIA. King Saud University, 20 October 2015



Background

- Digital multimedia are widely used in many website or online media to deliver information more attractively.
- The raw content must contain authentic information as it capture the real-time situation. However, after post-processing the information might altered or twisted when someone intentionally tamper the raw content.
- The audio forgery detection are very useful to combat digital forgery problems such as copyright issue, blurring court evidence, modify voice recording of public figure or politician for black campaign, authenticate correctness of financial recording and other sensitive issue.

Introduction

- In sophisticated forgery, they would attempt to hide audible trace (due to forgery) using some techniques such as by applying some noises or filters at post-processing.
- Audio signal that have unstable fingerprint can be considered as tampered audio signal.
- Related works in forgery detection for audio signal can be found in [1-16, 24]. However, it is obvious detecting various of noises and filters at a time will time consuming and superfluous.
- Hence, this study proposed a novel approach to detect the forgery by considering microphone's fingerprint.

Related Works

- Rui Yang, "**Additive noise detection and its application to audio forensics**," Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA) , vol., no., pp.1,5, 9-12 Dec. **2014**.
- Zhao, H., Malik, H., **Audio Recording Location Identification using Acoustic Environment Signature**. IEEE Transactions on Information Forensics and Security, vol. 8(11), pp. 1746 – 1759, Nov. **2013**.
- S. Hicsonmez, E. Uzun, and H.T. Sencar, "**Methods for identifying traces of compression in audio**," in Proc. of the 1st International Conference on Communications, Signal Processing, and their Applications, Sharjah, **2013**, pp. 1–6.
- R. Yang, Z. Qu, and J. Huang, "**Exposing mp3 audio forgeries using frame offsets**," ACM Transactions on Multimedia Computing, Communications and Application, vol. 8, no. S2, pp. 35:1–35:20, **2012**.
- R. Yang, Q.Y. Shi, and J. Huang, "**Detecting Double Compression of Audio Signal**," Proc. SPIE 7541, SPIE Press, **2010**; doi:10.1117/12.838695.

Objectives

- To conduct tampered detection of audio signal based on the digital traces.
- To analyze the inconsistency digital traces of tampered audio.
- To evaluate performance of three fingerprints feature (Gabor, MFCC & PLP).

Proposed Method (1 of 3)

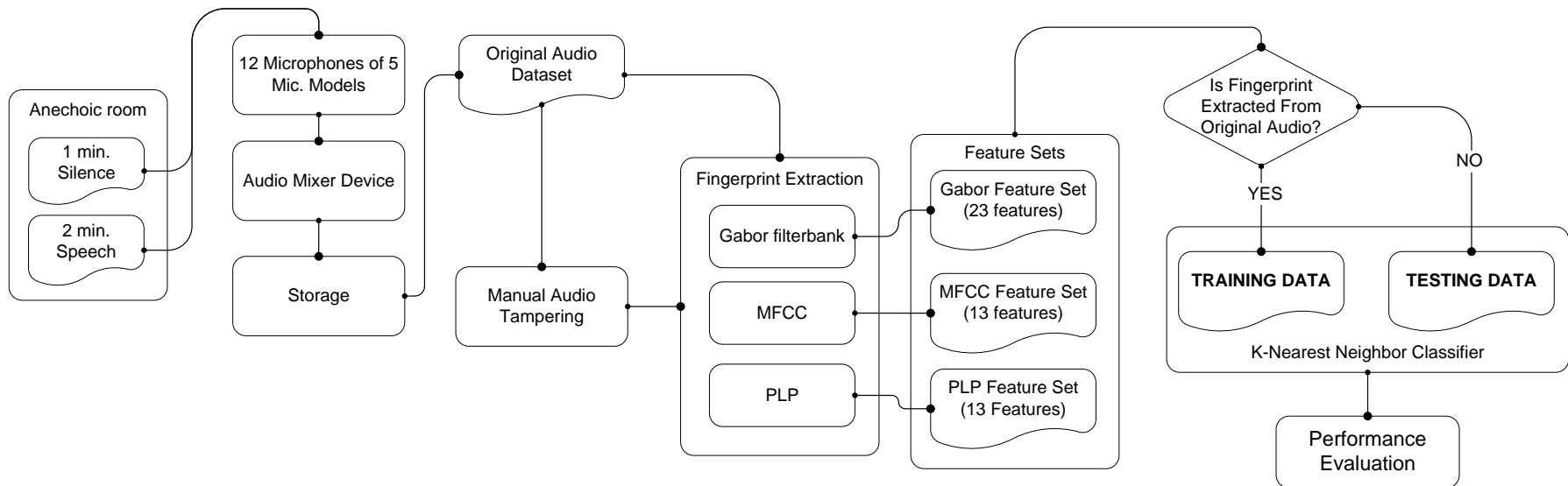


Diagram of Proposed Feature Extraction for Microphone's Fingerprint

Proposed Method (2 of 3)

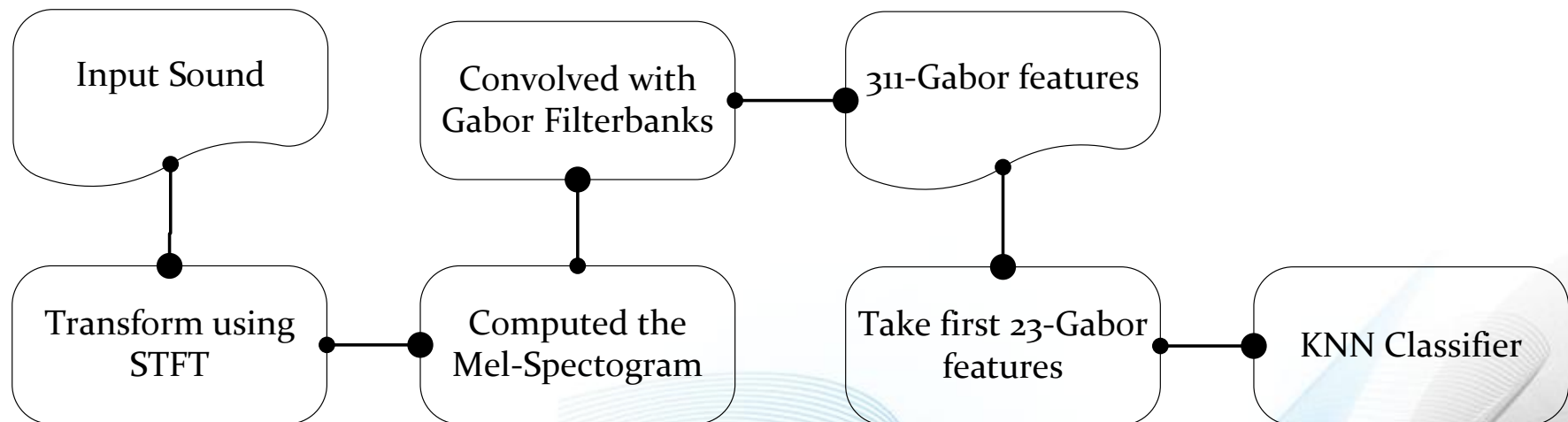
MFCC & PLP

- MFCC and PLP can be considered as baseline on the comparison because those features have been widely used and proven robust in speech recognition [22, 23].
- In general, PLP and MFCC shared several analogous steps.
- First, both features applied hamming window and DFT on input signal. Then, a set of filterbank are employed to generate the power spectrum.
- The main difference between PLP and MFCC are on filterbank that utilized and how processing the power spectrum to produce 13 features set.
- For the detailed MFCC and PLP features can be read in reference [22, 23].

Proposed Method (3 of 3)

Gabor filterbank Auditory Spectrum

- Schadler et al. [20] proposed GFAS features for speech recognition.
- The Gabor function is considered as proven able to represent speech signal in compact spectro-temporal structure, allow constant overlap, compress the output excess and has flexible parameters.



Dataset (1 of 2)

- 14 audio recording are collected prior to generate tampered audio files.
- 5 different mic. models utilized, where each of them at least has two identical model.
- It recorded simultaneously in anechoic room by organizing the microphone using a mic's stand such that it well-organized.
- The recording session has 3 minutes recording, 1st minute is silence and remaining is speech.
- There is fixed 30 cm distance between the person's lips and the mic.
- The tampered audio are generated by replacing "destination file" at particular time with audio signal taken from "source file" as described in Table III.

Description	Value
Format	Wave
Audio Format	PCM
Codec ID	1
Bit rate	705.6 Kbps
Channel(s)	1 channel
Sampling rate	44.1 KHz
Bit depth	16 bits
File size	~16.9 MB
Overall bit rate mode	Constant
Bit rate mode	Constant
Format settings, Endianness	Little
Format settings, Sign	Signed

Dataset (2 of 2)

TABLE I: Microphone Features And Specifications






<u>Shure SM-58</u> (Mic ₁ : 3 units)	Electro-Voice RE-20 (Mic ₂ : 2 units)	Coles 4038 Ribbon Microphone (Mic ₃ : 2 units)	<u>Sennheiser MD421II</u> <u>Dynamic Cardioid</u> Microphone (Mic ₄ : 3 units)	AKG C 451 B Condenser Microphone (Mic ₂ : 2 units)
				
SHU_0058	ELE_0020	COL_4038	SEN_0421	AKG_0451

TABLE III: Tampered Description

No.	File Name	Short Name	Source File	Destination File	Start Time	End Time
1	T_AKG_0451_m1_m2	T_AKG	AKG_0451_m2	AKG_0451_m1	0:01:12	0:01:19
2	T_COL_4038_m2_m1	T_COL	COL_4038_m1	COL_4038_m2	0:02:26	0:02:45
3	T_ELE_0020_m1_m2	T_ELE	ELE_0020_m2	ELE_0020_m1	0:01:21	0:01:38
4	T_SEN_0421_m1_m2	T_SEN_A	SEN_0421_m2	SEN_0421_m1	0:02:36	0:02:53
5	T_SEN_0421_m3_m1	T_SEN_B	SEN_0421_m1	SEN_0421_m3	0:01:03	0:01:19
6	T_SHU_0058_m1_m3	T_SHU_A	SHU_0058_m3	SHU_0058_m1	0:02:09	0:02:24
7	T_SHU_0058_m2_m3	T_SHU_B	SHU_0058_m3	SHU_0058_m2	0:01:38	0:01:56

Experimental Settings

- The experiment is conducted by **grouping same models** into two or three classes depend on corresponding number of microphone of the model.
- **Intra-class problem is the main concern** in this experiment. Hence, the classifier will be burdened with less number of classes such that can reveal the robustness of each features under identical model.
- All three features called **Gabor filterbank, MFCC and PLP will be compared**. Those feature extraction methods are applied on both original and tampered dataset.
- Afterward, **train-data and test-data are fairly constructed through 10-fold cross validation**.
- **The K-NN classifier is utilized** to classify the microphone model definitely after it trained with the train-dataset.

Experimental Result (1 of 3)

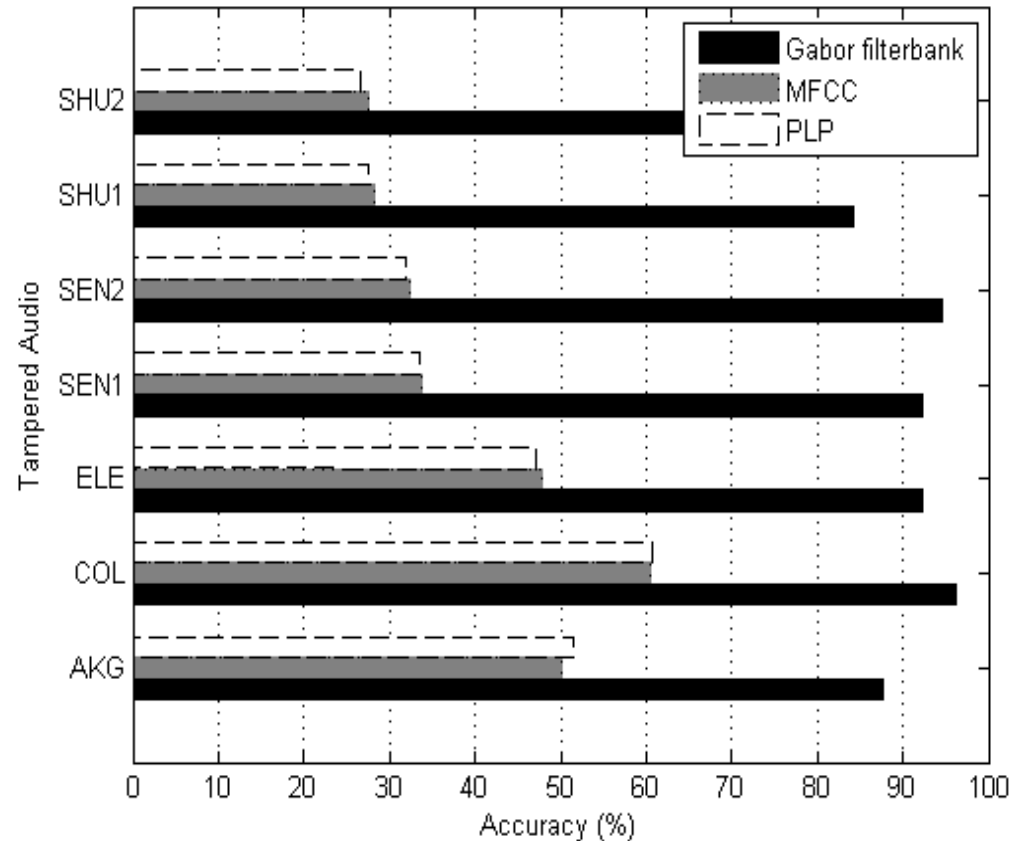


Fig. 3: Accuracy of K-NN Classifier Identify the Tampered Audio.

Experimental Result (2 of 3)

TABLE IV: Accuracy (in %) of K-NN Classifier

Description	T_AKG	T_COL	T_ELE	T_SEN_A
<u>gabor</u> : 23f_2970d	87.73	96.21	92.38	92.39
<u>mfcc</u> : 13f_2980d	50.13	60.55	47.81	33.86
<u>plp</u> : 13f_2980d	51.48	60.75	47.19	33.49

T_SEN_B	T_SHU_A	T_SHU_B
94.59	84.3	85.78
32.38	28.2	27.46
31.86	27.64	26.58

Experimental Result (3 of 3)

- It is clearly shows that 23 features of reduced-Gabor filterbank obtained highest accuracy followed by 13 features of MFCC and last is PLP feature.
- From table IV, the highest rates using reduced-Gabor feature can obtained 96.21% on tampered audio of COL_4038 model.
- Accuracy of reduced-Gabor on other models are also promising with at least 84% correct rates.
- MFCC and PLP achieved maximum correct rates not more than 61%. In addition, both features even give very low accuracies less than 34% correct rates for model SEN_0421 and SHU_0058.

Conclusions

- This study compared three features that exploited as microphone's fingerprint to identify the microphone model.
- Experimental result shows the tampered audio can be identified with high correct rates using Gabor filterbank features.
- Inconsistency in audio recording can be detected based on the digital traces even though it tampered using identical model.
- The Gabor filterbank feature outperform with accuracy of 96.21%.

Future Works

- This study can be extended to study **blind forgery detection**. In such case, **no prior knowledge is required** to locate the tampered region.
- More study can be carried out on **various places that covered indoor and outdoor** environment.
- It is interesting to know more **how echo, reverberant or any noise** can affect the digital traces.

Acknowledgment

This Project was funded by the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award Number **12-INF2634-02**



Thank You

