EUSIPCO 2018

Lossy Audio Compression Identification

Bongjun Kim
Northwestern University, USA
bongjun@u.northwestern.edu

Zafar Rafii
Gracenote, USA
zafar.rafii@nielsen.com

1. Introduction

Lossy Audio Compression

- Reducing the size of audio data by removing perceptually less significant information.
- Popular formats: MP3, AAC, AC-3, Vorbis, and WMA.

Audio Compression Identification

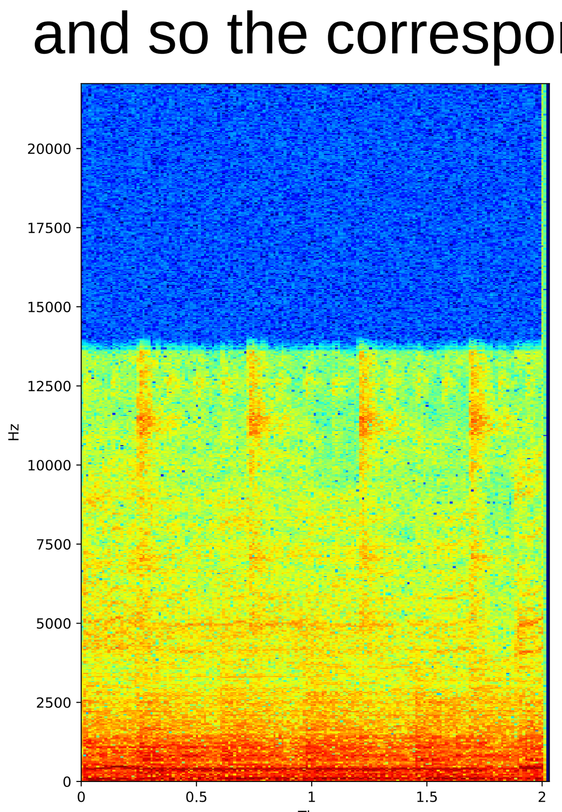
- Identifying information regarding the data compression that an audio signal has undergone, regardless of the content.
- Applications include detection of alterations in audio data, authentication of the audio quality, and identification of the original source of an audio signal (e.g., a TV channel or streaming service).

Contributions of this Work

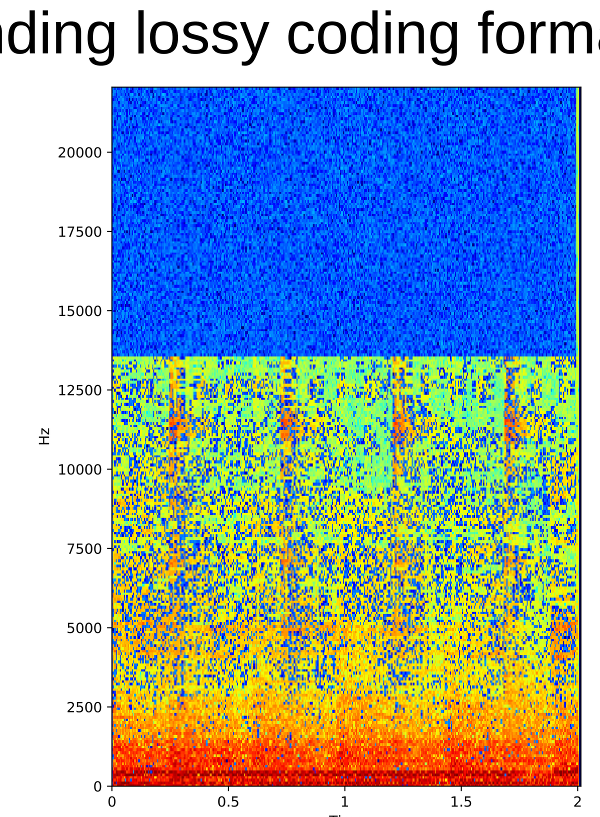
- A new metric for measuring traces of compression, robust to content variations.
- A new method for combining the estimates from multiple audio blocks to refine the results.
- A full evaluation with all the popular formats, and various compression and conditions.

2. Idea

- Different coding formats typically depend on different parameters for encoding the audio data (e.g., window functions and lengths).
- Traces of compression become visible in the time-frequency representation only when the compression parameters and framing conditions match those used for the encoding.
- By measuring the energy in the time-frequency representations of the audio signal for different framing conditions, we can then identify the correct compression parameters and so the corresponding lossy coding format.

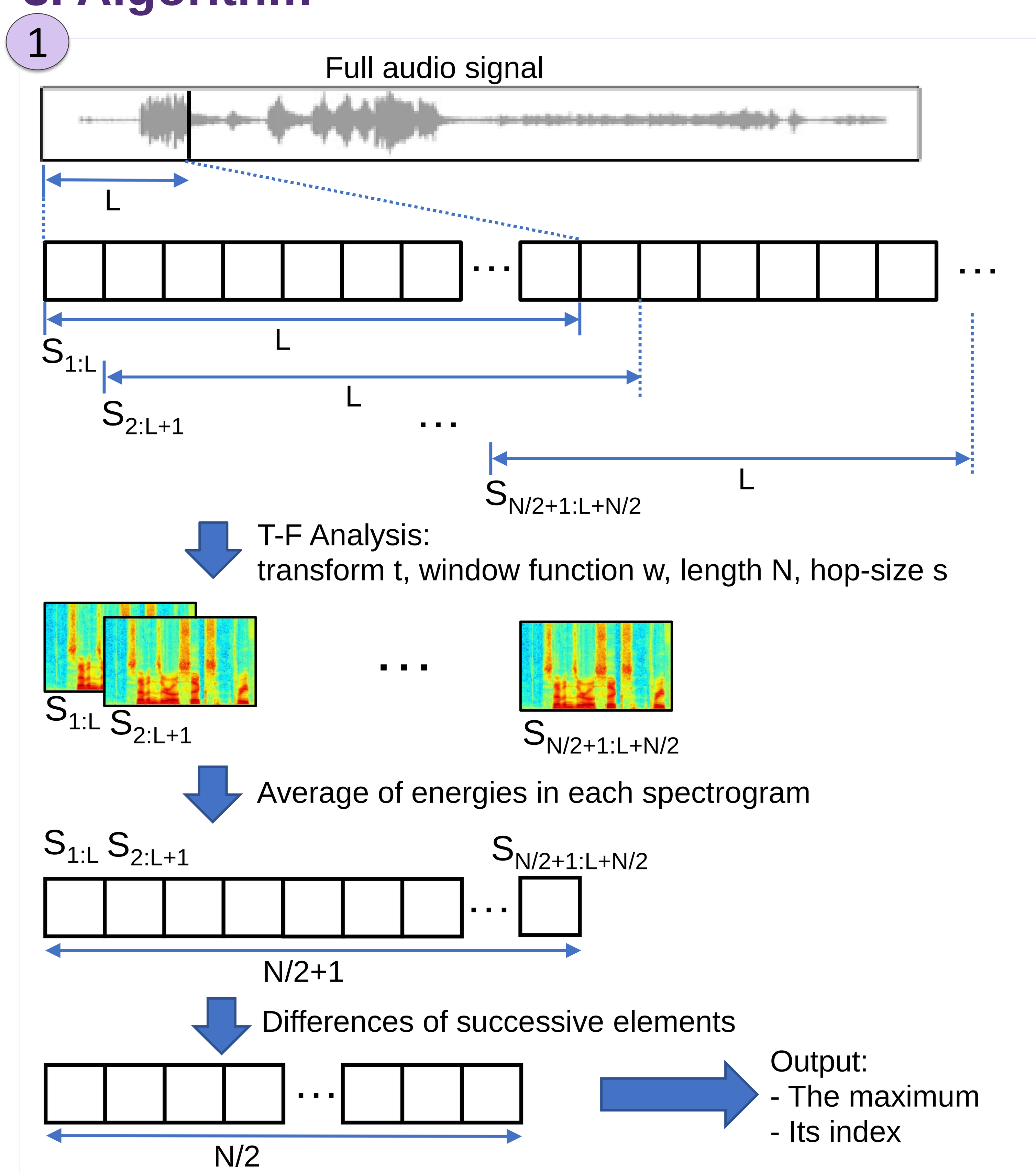


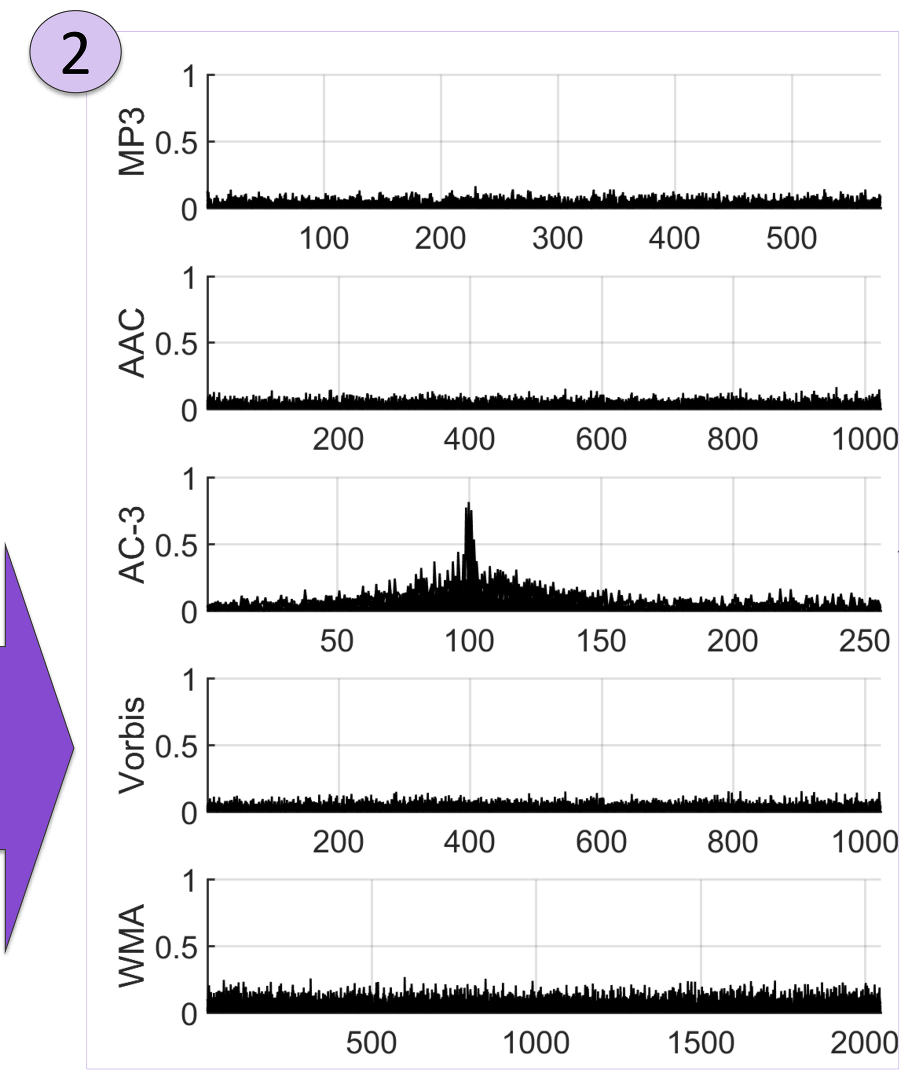
(a) Parameters not matched

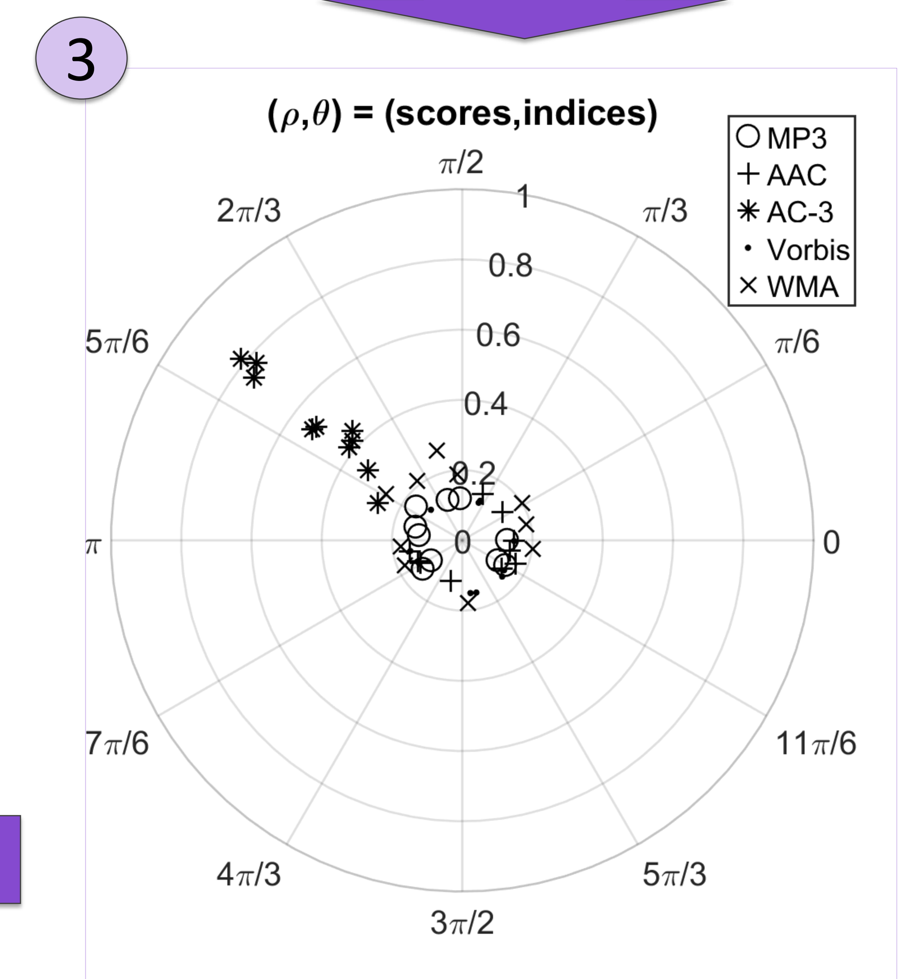


(b) Parameters matched

3. Algorithm







A group of points with the highest circular mean
➔ AC-3

4. Evaluation

A. Dataset

- 20 one-minute audio excerpts from 10 different songs (CDs, 44.1 kHz) and 10 different movies (Blue-rays, 48 kHz).
- Each excerpt was compressed using the 5 popular lossy coding formats and at 5 different bit rates (including high bit rates), and then converted back to a WAV file.
- We used both digital and analog transfer, the latter one which introduces sample desynchronization and electronic noise.

➔ 1,000 one-minute audio examples

B. Settings

Coding format	Window function	Window length
MP3	Sine	1152
AAC	KBD (α=4)	2048
AC-3	KBD (α=5)	512
Vorbis	Slope	2048
WMA	Sine	4096

* KBD: Kaiser-Bessel-Derived window

- We used the modified discrete cosine transform (MDCT) for the time-frequency transform and half the window length for the hop size.

C. Results

Digital	96k	128k	192k	256k	320k	All
MP3	1.0	1.0	1.0	1.0	0.9	0.98
AAC	1.0	1.0	1.0	1.0	1.0	1.0
AC-3	1.0	1.0	1.0	1.0	1.0	1.0
Vorbis	1.0	1.0	1.0	1.0	1.0	1.0
WMA	1.0	1.0	1.0	1.0	1.0	1.0
All	1.0	1.0	1.0	1.0	0.98	**0.996**

Analog	96k	128k	192k	256k	320k	All
MP3	1.0	1.0	0.95	0.75	0.7	0.88
AAC	1.0	1.0	1.0	1.0	0.8	0.96
AC-3	0.95	1.0	1.0	0.95	0.75	0.93
Vorbis	1.0	1.0	1.0	1.0	0.25	0.85
WMA	1.0	1.0	1.0	1.0	1.0	1.0
All	0.99	1.0	0.99	0.94	0.7	**0.924**

Conclusion: Our system can identify the correct coding format in almost all cases, even at high bit rates and with distorted audio, with an overall accuracy of 96%.