The Audio Degradation Toolbox

http://code.soundsoftware.ac.uk/projects/audio-degradation-toolbox/

and its Application to Robustness Evaluation

Sebastian Ewert and Matthias Mauch
lossy compression

photo by dan taylor
bad analog-to-digital conversion

photo by emilio di fabio
low quality microphone

photo by JeffaCubed
Environmental noise
...and many other things degrade audio.

- irregular tape playback
- dynamic range compression in radio and tv broadcasts
- audio speedup on the radio
- noise
- clipping and other distortion
- ... and yet more.
Audio Collection Quality

- most audio collections
- contain some audio of low quality
- contain recordings of different qualities
- contain recording of *unknown* quality
Audio Collection Quality

- most audio collections
- contain some audio of low quality
- contain recordings of different qualities
- contain recording of *unknown* quality
Impact on Music Informatics

- methods are usually tested only on one (or few) audio collections, hence:
  - feature extractors (etc.) might fail in the real world
  - affects MIR researchers’ work
- if feature extractors work, it is not clear if they correlate with content or audio quality
  - affects ‘digital musicologists’ and industry
Audio Degradation Toolbox

- most comprehensive collection of Matlab code for audio degradation
- designed to make it easy to degrade audio in many different ways
- aim: encourage MIR researchers to test their algorithms under many different conditions

GPL open source on SoundSoftware
## Degradation Units

<table>
<thead>
<tr>
<th>Add Noise</th>
<th>Apply Impulse Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add Sound</td>
<td>High-pass filter</td>
</tr>
<tr>
<td>Attenuation</td>
<td>Low-pass filter</td>
</tr>
<tr>
<td>Aliasing</td>
<td>MP3 Compression</td>
</tr>
<tr>
<td>Clipping</td>
<td>Saturation</td>
</tr>
<tr>
<td>Delay</td>
<td>Speedup</td>
</tr>
<tr>
<td>Dynamic Range Compr.</td>
<td>Wow Resampling</td>
</tr>
</tbody>
</table>
Degradation Units

- Add Noise
- Add Sound
- Attenuation
- Dynamic Range Compr.

- Apply Impulse Response
- High-pass filter
- Low-pass filter
- MP3 Compression
- Saturation
- Speedup
- Wow Resampling

sounds included:
pub sound env.,
vinyl crackle
Degradation Units

- Add Noise
- Add Sound
- Attenuation
- Dynamic Range Compr.
- Apply Impulse Response
- High-pass filter
- Low-pass filter
- MP3 Compression
- Speedup
- Wow Resampling

Sounds included:
- Pub sound env.
- Vinyl crackle
- Room, microphone, speaker and vinyl player IRs
Degradation Unit Example

```matlab
parameter.noiseColor = 'brown';
[audio_out, timestamps_out] =
    degradationUnit_addNoise(audio, samplingFreq,
                          timestamps, parameter)
```

- example sound before / after
- why “timestamps” — we’ll see later.
Degradations

- To make complex “Degradations” we can make chains from degradation units
- ...like audio effects!

- Example: Radio Broadcast Degradation

Dynamic Range Compr.

Speedup
Degradiations

- to make complex “Degradiations” we can make chains from degradation units
- ... like audio effects!
- Example: *Radio Broadcast* Degradation

![Diagram]

Radio Broadcast → Dynamic Range Compr. → Speedup
Degradations — examples


- Examples with spectrogram:
  - Wow resampling on cello (file6)
  - Live Recording on file1
Comparing to Ground Truth

- one main purpose: evaluate methods under different degradations
- problem — we have time-distorting degradations
- solution: every degradation can also transform ground truth to the time line of the degraded audio
  - example: beat tracking ground truth after “Speedup” degradation
Comparing to Ground Truth

- one main purpose: evaluate methods under different degradations
- **problem** — we have time-distorting degradations
- **solution**: every degradation can also transform ground truth to the time line of the degraded audio

- example: beat tracking ground truth after “Speedup” degradation

```
original ground truth          time
                              |   |   |   |   |
```
Comparing to Ground Truth

- one main purpose: evaluate methods under different degradations
- **problem** — we have time-distorting degradations
- **solution**: every degradation can also transform ground truth to the time line of the degraded audio

- example: beat tracking ground truth after “Speedup” degradation

```
| original ground truth | | | | | | | |
| transformed           | | | | | | | |
```
Revisit Example

Radio Broadcast

Dynamic Range Compr.

Speedup

degradation unit

degradation

audio

ground truth

timestamps

transformed

audio

transformed

ground truth

Friday, 1 November 13
Experiments on ‘Real-World’ Degradations

- Live Recording
- Radio Broadcast
- Smartphone Playback
- Smartphone Recording
- Strong MP3 Compression
- Vinyl Recording
Results I — Audio ID

- audio ID fails for most “Real-World” degradations, not for mp3
- robustness to pink noise is ok

<table>
<thead>
<tr>
<th></th>
<th>correct</th>
<th>incorrect</th>
<th>not identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Live</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Radio</td>
<td>3</td>
<td>3</td>
<td>94</td>
</tr>
<tr>
<td>PhonePlay</td>
<td>0</td>
<td>1</td>
<td>99</td>
</tr>
<tr>
<td>PhoneRec</td>
<td>5</td>
<td>7</td>
<td>88</td>
</tr>
<tr>
<td>MP3</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vinyl</td>
<td>4</td>
<td>0</td>
<td>96</td>
</tr>
</tbody>
</table>
Results I — Audio ID

- audio ID fails for most “Real-World” degradations, not for mp3
- robustness to pink noise is ok

![Graph showing the relationship between correct audio identification and dB SNR](image.png)
Results II — Score-to-audio alignment

- pretty much falls over for "Live" and "Phone Playback" degradations
- explanations: onset duplication; bass harmony missing

![Box plot comparing score-to-audio alignment accuracy under different real-world degradations.](image-url)
Results III — Beat-tracking

- compare two methods: BeatRoot, Davies
- very similar, but Davies more robust to “Live” degradation

![Box plot comparison of beat-tracking performance under various conditions.](image)

- Original
- Live
- Radio
- PhonePlay
- PhoneRec
- MP3
- Vinyl

**F measure**

- BeatRoot
- Davies
Results IV — Chord recognition

- compare two methods: Chordino, HPA
- HPA usually better, Chordino more robust on “Phone Play”

![Box plot diagram showing relative correct overlap for different degradation units and methods.

- Y-axis: relative correct overlap
- X-axis: Original, Live, Radio, PhonePlay, PhoneRec, MP3, Vinyl
- Two categories: Chordino, HPA

The diagram illustrates the performance of Chordino and HPA across various degradation units and playback settings. Chordino tends to maintain a higher relative correct overlap compared to HPA, indicating its robustness in certain conditions.
Results IV — Chord recognition

- compare two methods: Chordino, HPA
- HPA usually better, Chordino more robust on “Phone Play”

![Box-and-whisker plots of the song-wise relative correct overlap for degradation, with medians dropping less than 10 percentage points: Chordino, HPA.](image)
Results IV — Chord recognition

- compare two methods: Chordino, HPA
- HPA usually better, Chordino more robust on “Phone Play”

![Chord detection performance: HPA chroma](image)

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Original</th>
<th>200Hz</th>
<th>400Hz</th>
<th>800Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.2</td>
<td>0.60</td>
<td>0.29</td>
<td>0.06</td>
</tr>
<tr>
<td>D</td>
<td>0.8</td>
<td>0.3</td>
<td>0.1</td>
<td>0.04</td>
</tr>
<tr>
<td>E</td>
<td>1.0</td>
<td>0.4</td>
<td>0.1</td>
<td>0.03</td>
</tr>
<tr>
<td>F</td>
<td>0.6</td>
<td>0.2</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>G</td>
<td>0.7</td>
<td>0.3</td>
<td>0.16</td>
<td>0.05</td>
</tr>
</tbody>
</table>

For every song against human annotated ground-truth [14]. We prepared 180 songs by the Beatles by degrading those five further degradations using the Radio HP 400, 800, 1000, 2000, and 3000, resulting in 1260 songs.

We used a 400Hz stop band edge. The high-pass filter is maintained for the 50Hz filter, but increasingly obfuscating the clear C major and A minor patterns.

To locate the reason for the strong drop-off, we studied the results by degradation and method. For the original audio file and a low-pass filter, the HPA’s advantage over the original is substantial. HPA’s advantage over Chordino, for the high-pass filter, is substantially lower. The obvious exception is the Chordino’s 0.77 (original: 0.94). As explained in the case of Chordino, Version 1.0 — PhoneRec, HP 400 Hz High-pass, HP 50 Hz High-pass, HP 100 Hz High-pass, HP 400 Hz High-pass, HP 800 Hz High-pass, HP 1000 Hz High-pass, HP 2000 Hz High-pass, and HP 3000 Hz High-pass.
Summary

- Audio Degradation Toolbox offers
  - easy-to-use degradations
  - more comprehensive than other existing toolboxes
  - ground truth time-line transform to evaluate on time-warping degradations

- Results show: ADT is useful to detect strengths and weaknesses of MIR methods

- For paper, audio examples, source code:
  http://code.soundsoftware.ac.uk/projects/audio-degradation-toolbox
What’s up next?

- convince everyone to use the ADT :)
- work with it ourselves...
  - degraded audio as additional training data
  - affect of degradation on human ground truth labelling