2. Time-Frequency characterisation of musical noise

Short-time Fourier Transform of a real signal

From a real signal $x$ of length $N$:

- **STFT**:
  
  $$F(n, m) = \text{STFT}(x) \triangleq \sum_{k=0}^{N-1} (x \cdot Rm)w(k)e^{-j2\pi \frac{k}{N}}$$

- **Spectrogram**:
  
  $$S(n, m) = |F(n, m)|^2$$

- **Inverse STFT**:
  
  $$\hat{x}(l) = \sum_{n} s(l - m) \sum_{m} F(n, m)e^{j2\pi \frac{mn}{N}}$$

- $F$ is consistent if $F = \text{STFT}([F|^{T/2}]$)

Consistent representation of an isolated peak

- Single isolated peak in Time-Frequency plane:
  
  $$F_0(n, m) = \begin{cases} 1 + \alpha & \text{if } n = \alpha \text{ and } m = \beta \\ 0 + \beta & \text{otherwise.} \end{cases}$$

- Resulting synthesized signal $\hat{x}[l]$:
  
  $$\hat{x}[l] = [l - \beta R]e^{j2\pi \frac{\alpha l}{N}}$$

- Time-frequency representation $F_0$ of $\hat{x}$:
  
  $$F_0(n, m) = \sum_{k=0}^{N-1} e^{-j2\pi \frac{k}{N}(n-\alpha)} \sum_{k=0}^{N-1} w(k)(k-\beta R - m)^2.$$  

⇒ Consistent representation of isolated peaks appear as “spots”.

3. Domain localisation in audio spectrograms

Based on the work in [6].

Detection of local minima of the spectrogram

- Selection of time-frequency bins with an energy lower than adjacent bins

Grouping of triangles in domains

- Merging of triangles to construct high-energy regions in the time-frequency plane

Evaluation of musical noise

- Number of regions is used as indicator of the presence of musical noise

From a set of points $P_k \in \mathbb{R}^2$

- Voronoi cell of $P_k$: all $x \in \mathbb{R}^2$ such that $x$ is closer to $P_k$ than to any other point

- Delaunay triangulation: two points are connected if their Voronoi cells are adjacent

⇒ Nice set of triangles (avoids narrow triangles)

Selection of triangles

- Triangles are retained according to the length of their edges

- The frequency contribution is used to catch high-energy regions

4. Experiments

Generation of musical noise

- AIP: Adding artificial isolated peaks in the spectrogram: $\hat{S} = S + p\mathbf{M}$, where $\mathbf{M}$ is a Bernoulli matrix of parameter $p$.
  
  - If $p$ is too low, there is no isolated peaks
  
  - If $p$ is too high, isolated peaks merge together and produces white noise

- OMP: Orthogonal Matching Pursuit denoising: estimation of a sparse approximation of a noisy signal, controlled by the approximation error $\epsilon$.
  
  - A low value of $\epsilon$ removes noise to the cost of low quality

- A high value of $\epsilon$ preserves the original signal but tends to produce musical noise

Results

- High correlation between the number of detected domains and the expected level of musical noise according to the parameters of generation

Perspectives

- More advanced descriptors of the presence of musical noise using detected domains

- Better evaluation of the performance:

  - Wider range of spectral techniques to generate musical noise
  
  - Comparison with state-of-the-art techniques
  
  - Listening tests

- Build new strategies to reduce musical noise

Acknowledgements

- ANR (JLL program MADO) (ANR-14-CE27-0005)

- European Union’s H2020 Framework Programme (H2020-MSCA-ITN-2014) - Grant agreement no 642685 MacSeNet.

- Engineering and Physical Science Research Council (EPSRC) of the UK grant number EP/L027119/2.

References

1. Berouti et al., Enhancement of speech corrupted by acoustic noise, 1979

2. Ephraim and Malah, Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator, 1984

3. Yu et al., Audio denoising by time-frequency block thresholding, 2007

4. Uemera et al., Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics, 2008

5. Derakhshan et al., An objective measure for the musical noise assessment in noise reduction systems, 2009