Assessment of Musical Noise using Localization of Isolated Peaks in Time-Frequency Domain

Ronan Hamon[1] Lucas Rencker[2] Valentin Emiya[1] Wenwu Wang[2] Mark Plumbley[2]

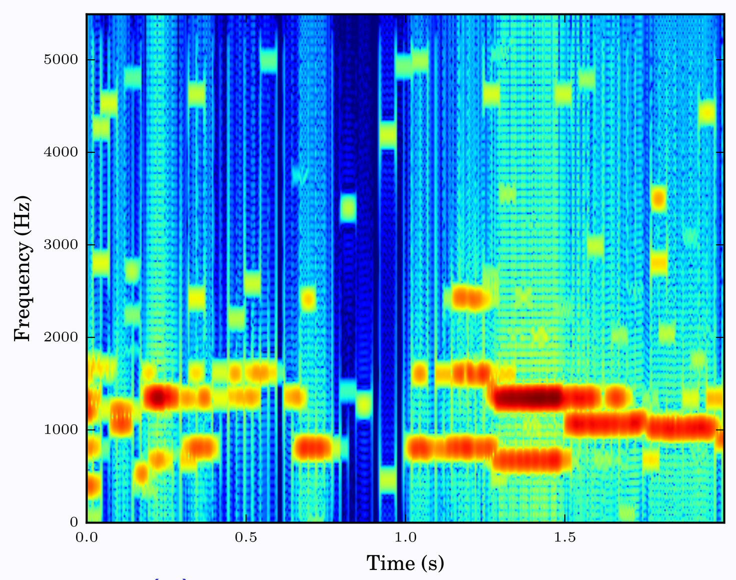
[1]Aix Marseille Univ, CNRS, LIF, Marseille, France {firstname.lastname}@lif.univ-mrs.fr

[2]CVSSP, University of Surrey, Guildford, UK {firstname.lastname}@surrey.ac.uk

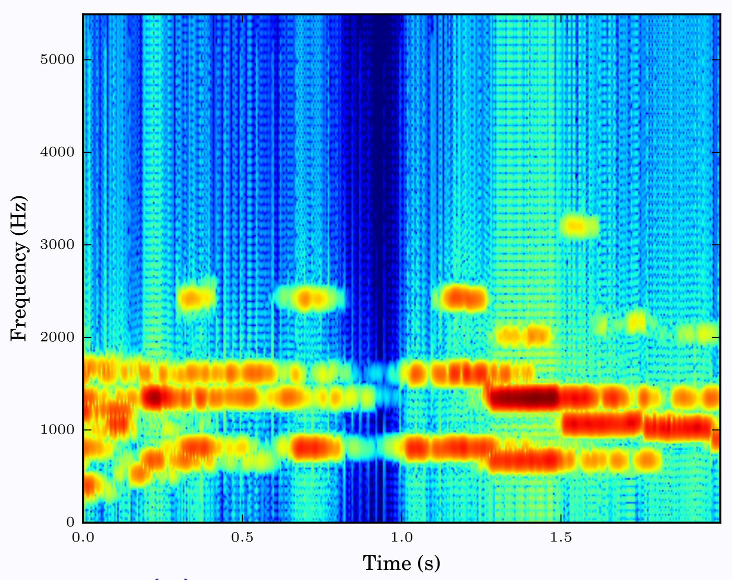
- Novel approach to assess the amount of musical noise in an audio signal based on the minima of the STFT
- Better understanding of the impact of artificial isolated peaks in the time-frequency domain

1. Musical noise

- Artificial noise appearing in processed spectrograms when using spectral techniques such that denoising or blind source separation
- Induced by isolated time-frequency coefficients
- Sound examples: `http://lif.univ-mrs.fr/ronan.hamon/icassp2017/`



(a) High level of musical noise



(b) Low level of musical noise

2. Time-Frequency characterisation of musical noise

Short-time Fourier Transform of a real signal

From a real signal x of length N:

- STFT:

$$\mathbf{F}(n,m) = \mathrm{STFT}(x) = \sum_{k=0}^{N-1} x(k+Rm)w(k)e^{-i2\pi\frac{kn}{N}}$$

- Spectrogram: $\mathbf{S}(n,m) = |\mathbf{F}(n,m)|^2$
- Inverse STFT:

$$\tilde{x}(l) = \sum_m s(l-mR)\sum_n \mathbf{F}(n,m)e^{i2\pi n\frac{l-mR}{N}}$$

- $\mathbf{F}$ is consistent if $\mathbf{F} = \mathrm{STFT}[\mathrm{STFT}^{-1}(\mathbf{F})]$

Consistent representation of an isolated peak

- Single isolated peak in Time-Frequency plane:

$$\mathbf{F}_0(n,m) = \begin{cases} 1+0i \text{ if } n=\alpha \text{ and } m=\beta \\ 0+0i \text{ otherwise.} \end{cases}$$

- Resulting synthesized signal $\tilde{\mathbf{x}}[l]$:

$$\tilde{\mathbf{x}}[l] = s[l-\beta R]e^{i2\pi\alpha\frac{l-\beta R}{N}}.$$

- Time-frequency representation $\hat{\mathbf{F}}_0$ of $\tilde{\mathbf{x}}$:

$$\hat{\mathbf{F}}_0(n,m) = \left|\sum_{k=0}^{N-1} e^{-i2\pi\frac{k(n-\alpha)-R(m-\beta)}{N}}w(k)s[k-R(\beta-m)]\right|^2.$$

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

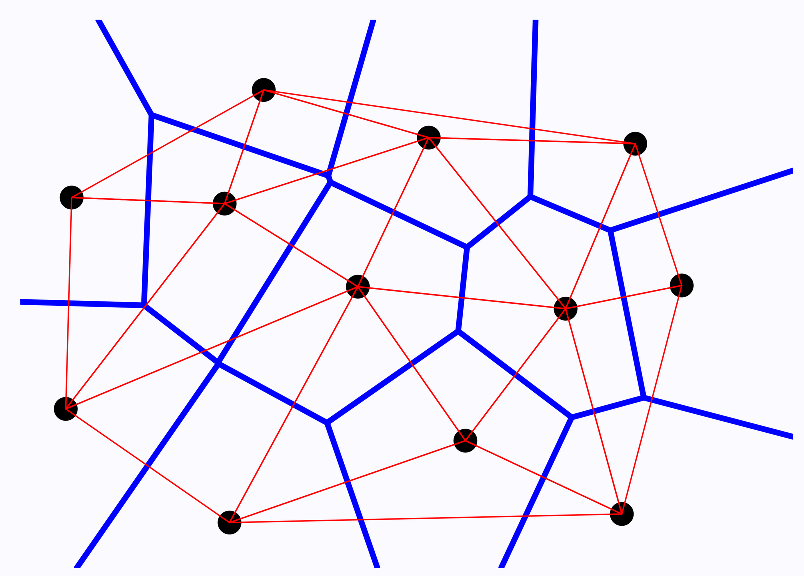
3. Domain localisation in audio spectrograms

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
- Choice of a threshold to avoid selecting minima in high-energy regions

Delaunay triangulation over minima



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 $\mathbf{P_k}$ 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

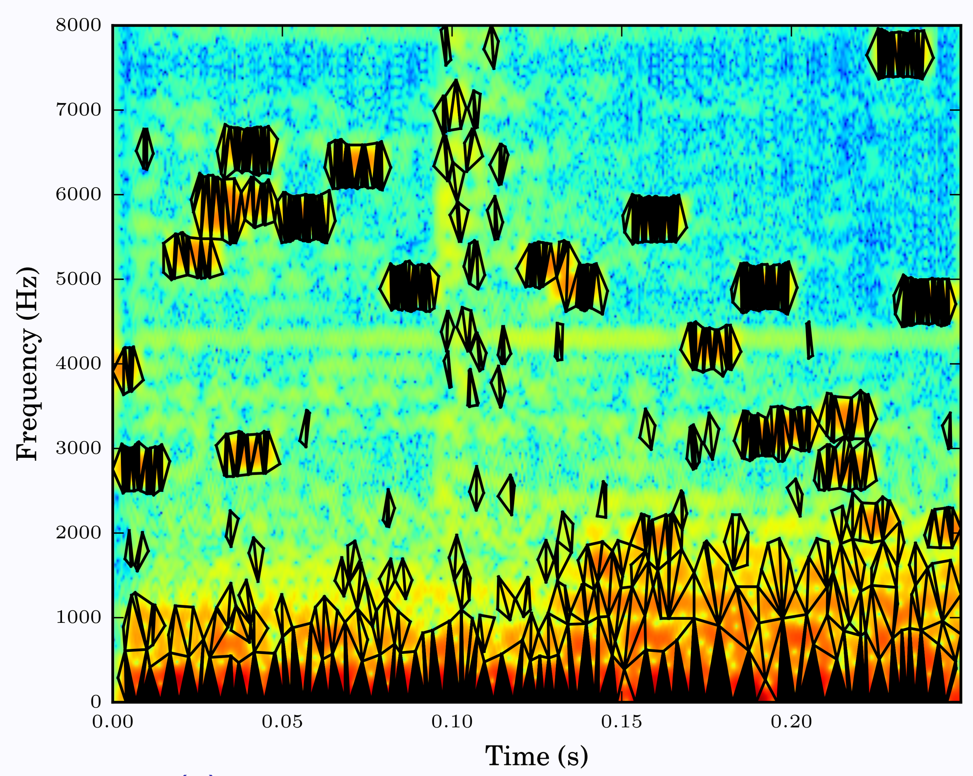
- Triangles are retained according to the length of their edges
- The frequency contribution is used to catch high-energy regions

Grouping of triangles in domains

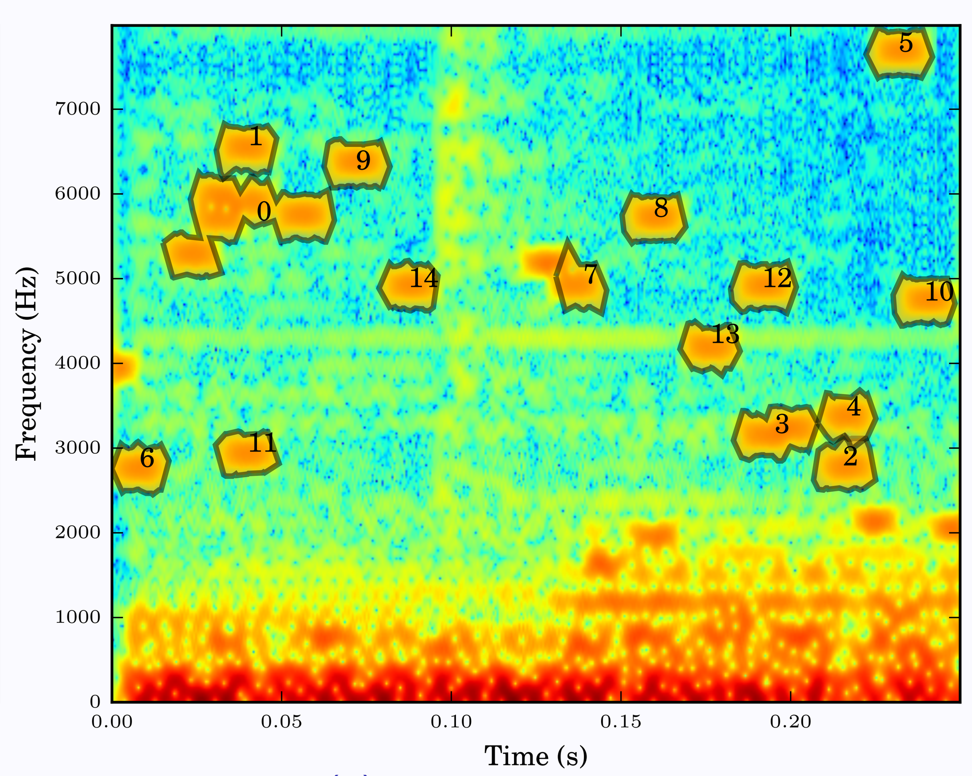
- Merging of triangles to construct high-energy regions in the time-frequency plane
- Selection of domains according to the expected duration of a spot

Evaluation of musical noise

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



(c) Triangulation after selection of triangles



(d) Retained domains

4. Experiments

Generation of musical noise

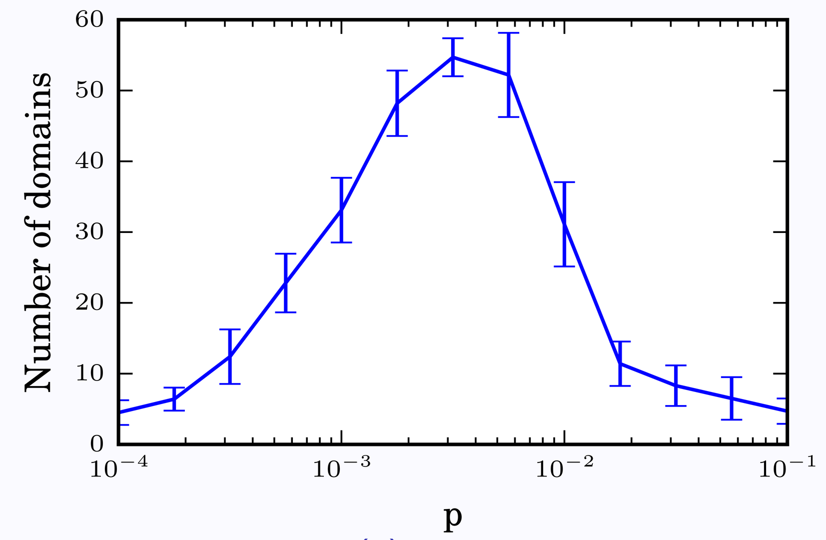
AIP: Adding artificial isolated peaks in the spectrogram: $\tilde{\mathbf{S}} = \mathbf{S} + \mu\mathbf{M}$, where $\mathbf{M}$ is a Bernoulli matrix of parameter p.

- If p is too low, there is no isolated peaks
- If p is too high, isolated peaks merge together and produces white noise

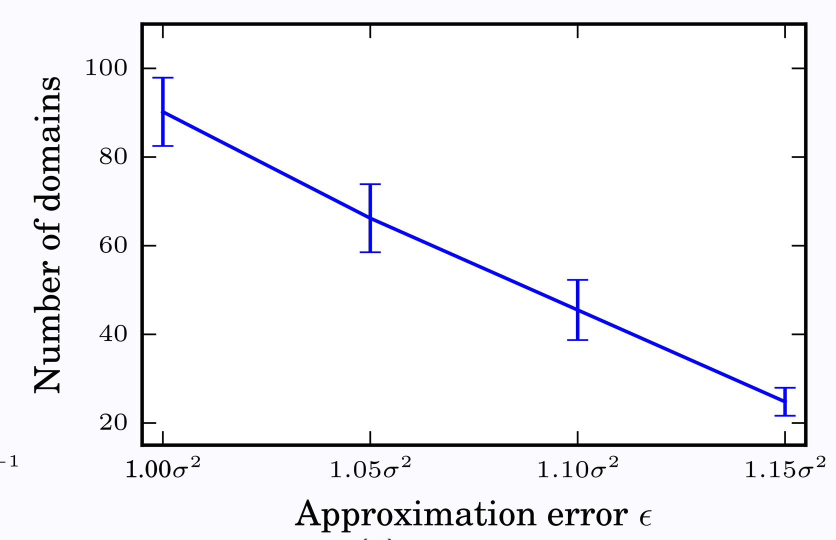
OMP: Orthogonal Matching Pursuit denoising: estimation of a sparse approximation of a noisy signal, controlled by the approximation error ε.

- A low value of ε removes noise to the cost of low quality
- A high value of ε preserves the original signal but tends to produce musical noise

Results



(e) AIP



(f) OMP

⇒ 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

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 threholding*, 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
6 **Flandrin** *Time-Frequency filtering based on spectrograms*, 2015.

Acknowledgements

- ANR JCJC program MAD (ANR-14-CE27-0002)
- European Union’s H2020 Framework Programme (H2020-MSCA-ITN-2014) - Grant agreement no 642685 MacSeNet.
- Engineering and Physical Sciences Research Council (EPSRC) of the UK grant number EP/L027119/2.