Audio Engineering Society

Convention e-Brief

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

Sharper spectrograms with Fast Local Sharpening

Robin Lobel

Divide Frame, 20 rue Raymond Aron, 75013 Paris, France

contact@divideframe.com

ABSTRACT

Spectrograms have to make compromises between time and frequency resolution because of the limitations of the short-time Fourier transform (Gabor, 1946). Wavelets have the same issue. As a result spectrograms often appear blurry, either in time, frequency, or both.
A method called Reassignment was introduced in 1978 (Kodera et al.) to make spectrograms look sharper. Unfortunately it also adds visual noise, and its algorithm does not make it suitable for realtime scenarios.
Fast Local Sharpening is a new method which attempt to overcome both theses drawbacks.

1 Introduction

The reassignment method was introduced in 1978 [1] to limit the uncertainty principle of Short Time Fourier Transforms [2]. It takes a spectrogram as an input, as well as its derivates in time & frequency (which can be computed from the same data set [3]), and estimate new locations for the spectrum pixels. As a result the spectrum looks much sharper, but major drawbacks come with this approach :

- While the overall spectrum power is kept constant, individual power peaks aren't accurate anymore when reassignment is made by adding reassigned pixels (standard reassignment method [1][3]). Noise is also introduced due to the high dispersion created by colliding frequencies and fragmented reassignment.
- Power peaks can be kept accurate when reassignment is made by maxing reassigned pixels instead of adding them, but more noise is then visible.
- It takes too long to reassign pixels dynamically in a realtime scenario with standard computer hardware.

The Fast Local Sharpening method whose results are shown here is only based on local, in-place informations from the derivates and therefore has none of theses drawbacks.

2 Peak Power Comparison

Linear Power and Frequency Increase (2D)

This example showing a linear increase in power (from -120dB to 0dB) and frequency (from 0Hz to 600Hz) over 250ms is designed to show the limitations of the Adding Reassignment method.

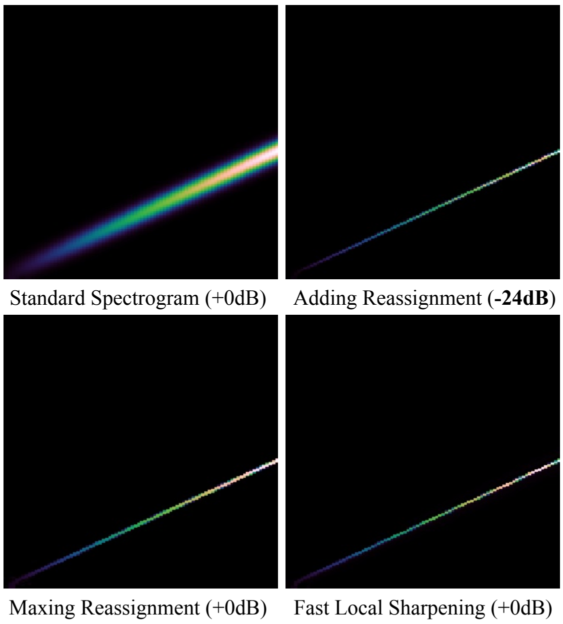

Fig. 1a. STFT: 2048 samples, Blackman-Harris
Time Overlap: x16 - Frequency Overlap: x4

As shown in figure 1a, Adding Reassignment peak power doesn't match the real peak power values: the power scale of the Adding Reassignment spectrogram had to be adjusted by -24dB to match the real values.

Linear Power and Frequency Increase (3D)

The same example is then shown with a vertical 3D power projection to compare the shape of the peak power over time.

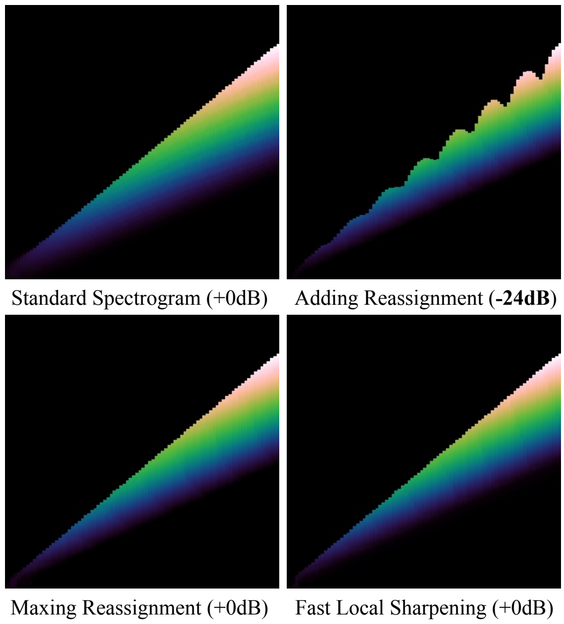

Fig. 1b. STFT: 2048 samples, Blackman-Harris
Time Overlap: x16 - Frequency Overlap: x4
Vertical 3D Power Projection

As shown in figure 1b, Adding Reassignment does not linearly increase power as opposed to the other methods. There's no strict peak power shape correlation with the real values.
Adding Reassignment is therefore discarded from the following comparisons with Standard Spectrogram, Maxing Reassignment and Fast Local Sharpening.

3 Ground Truth Comparison

Cross

The spectral cross is a worst case scenario where a continuous frequency (a pure horizontal spectral line) meets a click (a pure vertical spectral line) of equal power. The two distort each other's spectrogram in time and frequency at the crossing point.

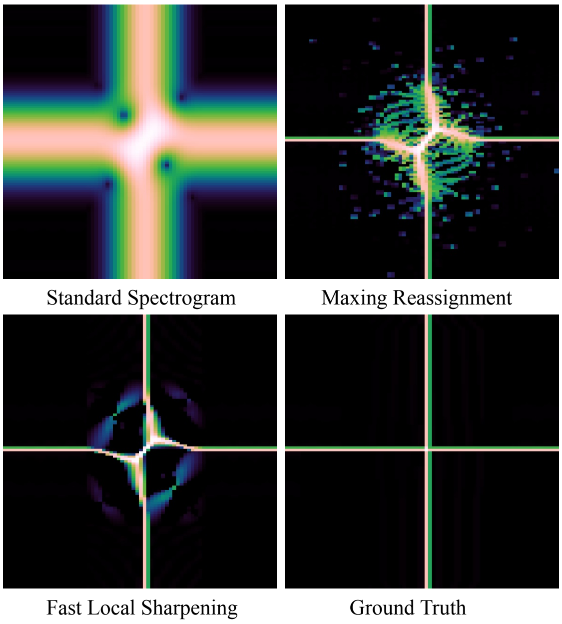

Fig. 2. STFT: 2048 samples, Blackman-Harris
Time Overlap: x32 - Frequency Overlap: x8

As shown in figure 2, MR and FLS both increase accuracy, but FLS also limits the spectral dispersion, getting closer to the ground truth spectrogram.

Frequency Modulation

The modulated frequency example is another difficult case where a frequency can distorts its own spectrogram by oscillating close to the time and frequency uncertainty of the STFT.
Here a frequency oscillating at 440Hz is modulated at 20Hz by ±44Hz.

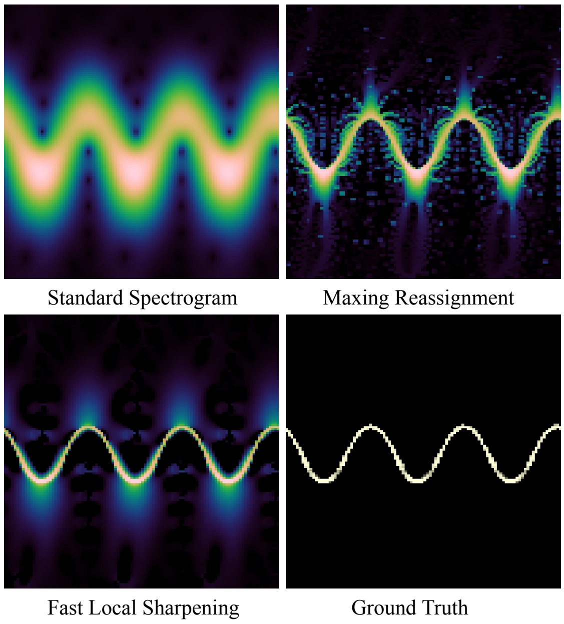

Fig. 3. STFT: 2048 samples, Blackman-Harris
Time Overlap: x32 - Frequency Overlap: x8

Figure 3 shows that FLS increases accuracy and limits spectral dispersion when compared to MR, getting closer to the ground truth spectrogram.

4 Resolution Comparison

White Noise

A pure white noise signal is compared at different resolutions, by increasing the time and frequency overlap of the STFT.

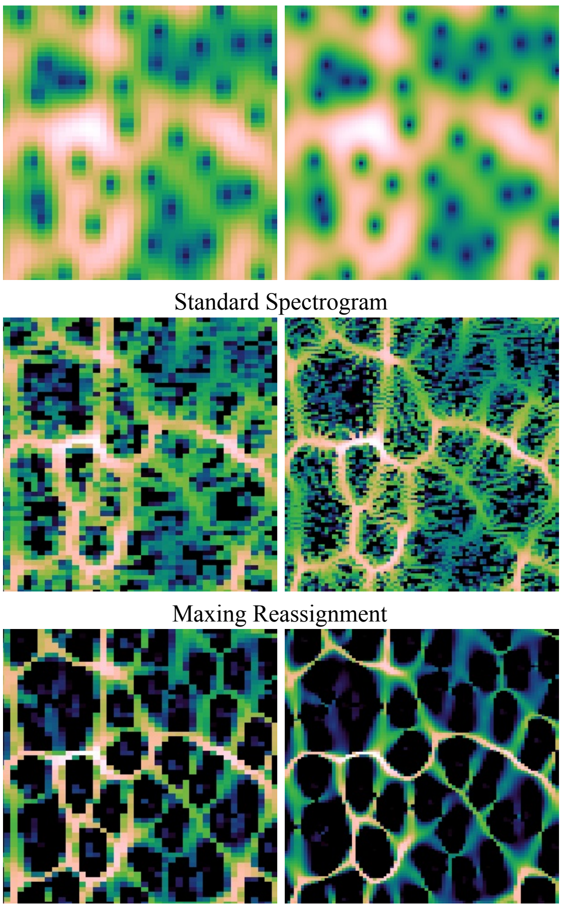

Fig. 4. STFT: 2048 samples, Blackman-Harris
Left: Time Overlap: x16 - Frequency Overlap: x4
Right: Time Overlap: x32 - Frequency Overlap: x8

While there's no ground truth spectrogram for white noise signals by definition, we can observe how MR and FLS converge at different resolutions towards an underlying pattern.

While MR remains mostly fuzzy with the first overlap settings, FLS already show clear patterns. Increasing resolution reveals more MR details but lot of areas are still undetermined, while FLS refines what was already apparent.

Vocal and Piano

This example is a very short extract from the song *City of Stars*, where vocals and piano mix over a little background noise.

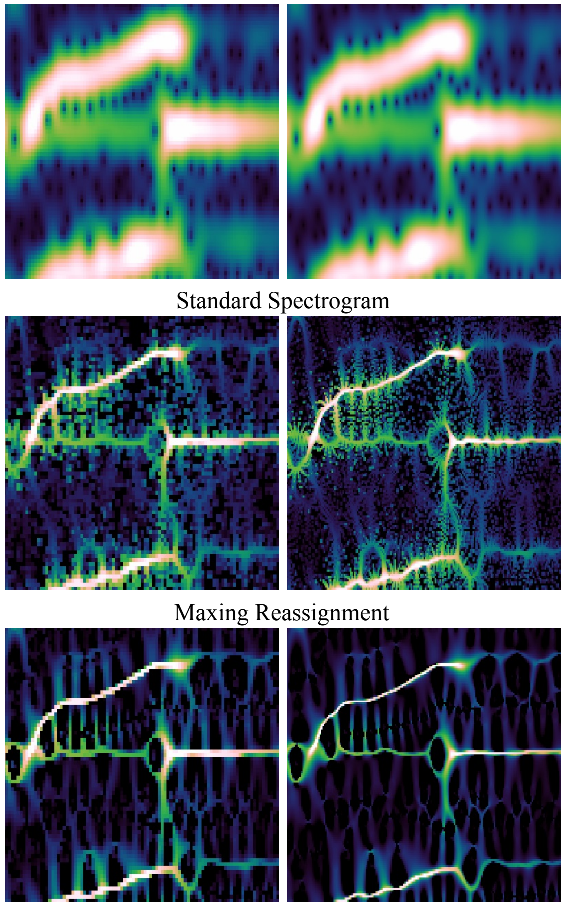

Fig. 5. STFT: 4096 samples, Blackman-Harris
Left: Time Overlap: x16 - Frequency Overlap: x4
Right: Time Overlap: x32 - Frequency Overlap: x8

Same remarks as with the white noise example.

5 Speed Comparison

Another issue with MR is speed. Despite being very simple operations, it requires pseudo-random memory access on non-aligned memory blocks, a sub-optimal scenario for standard computer architectures.

Therefore it cannot be vectorized (make use of SIMD instructions) and cannot be properly multi-threaded because several thread sync would be required, which would kill the performances.

As opposed to MR, FLS is designed to be highly vectorizable and multi-threadable. Despite requiring more calculus, it performs optimally on standard computer architectures.

	Proc. Time	Speed Boost
Initial Computation	18ms	
MR	52ms	
FLS (scalar, 1 core)	34ms	x1.5
FLS (SSE, 1 core)	5ms	x10
FLS (SSE, 2 cores)	2.9ms	x18

Table 1. STFT: 2048 samples, Blackman-Harris
Time Overlap: x16 - Frequency Overlap: x4
310 Transforms - Intel Core i5-4200U CPU

	Proc. Time	Speed Boost
Initial Computation	26ms	
MR	216ms	
FLS (scalar, 1 core)	125ms	x1.7
FLS (SSE, 1 core)	16ms	x13
FLS (SSE, 2 cores)	9.7ms	x22

Table 2. STFT: 2048 samples, Blackman-Harris
Time Overlap: x32 - Frequency Overlap: x8
620 Transforms - Intel Core i5-4200U CPU

SSE instructions and dual-cores processors were introduced as early as 2006 (Intel Core 2, AMD Athlon 64). The same data set was used in both benchmarks.

If the Initial Computation needs to be recalculated (due to a change of data for instance), IC+MR takes 70ms to compute in the first scenario, and 242ms in the second scenario, while IC+FLS takes 21ms and 36ms, resulting in an overall x3.3 and x6.7 speed boost. The further resolution increases, the faster FLS is over MR.

6 Conclusion

The results of Fast Local Sharpening shown here outperform the Reassignment methods both in accuracy and speed, making spectrogram sharpening a viable option both for peak power analysis and spectral morphology.

Speed could further be increased by using AVX instructions, more cores, or GPU processing.

References

[1] K. Kodera, C. de Villedary, and R. Gendrin, *A new method for the numerical analysis of time-varying signals with small BT values* Phys. Earth Planet. Interiors, no. 12, pp. 142-150, 1976.

[2] D. Gabor, *Theory of communication*, J. Inst. Electron. Eng., vol. 93, no. 11, pp. 429457, Nov. 1946.

[3] F. Auger and P. Flandrin, *Improving the readability of time-frequency and time-scale representations by the reassignment method*, IEEE Trans. Signal Process., SP-43(5), 1068–1089, 1995.