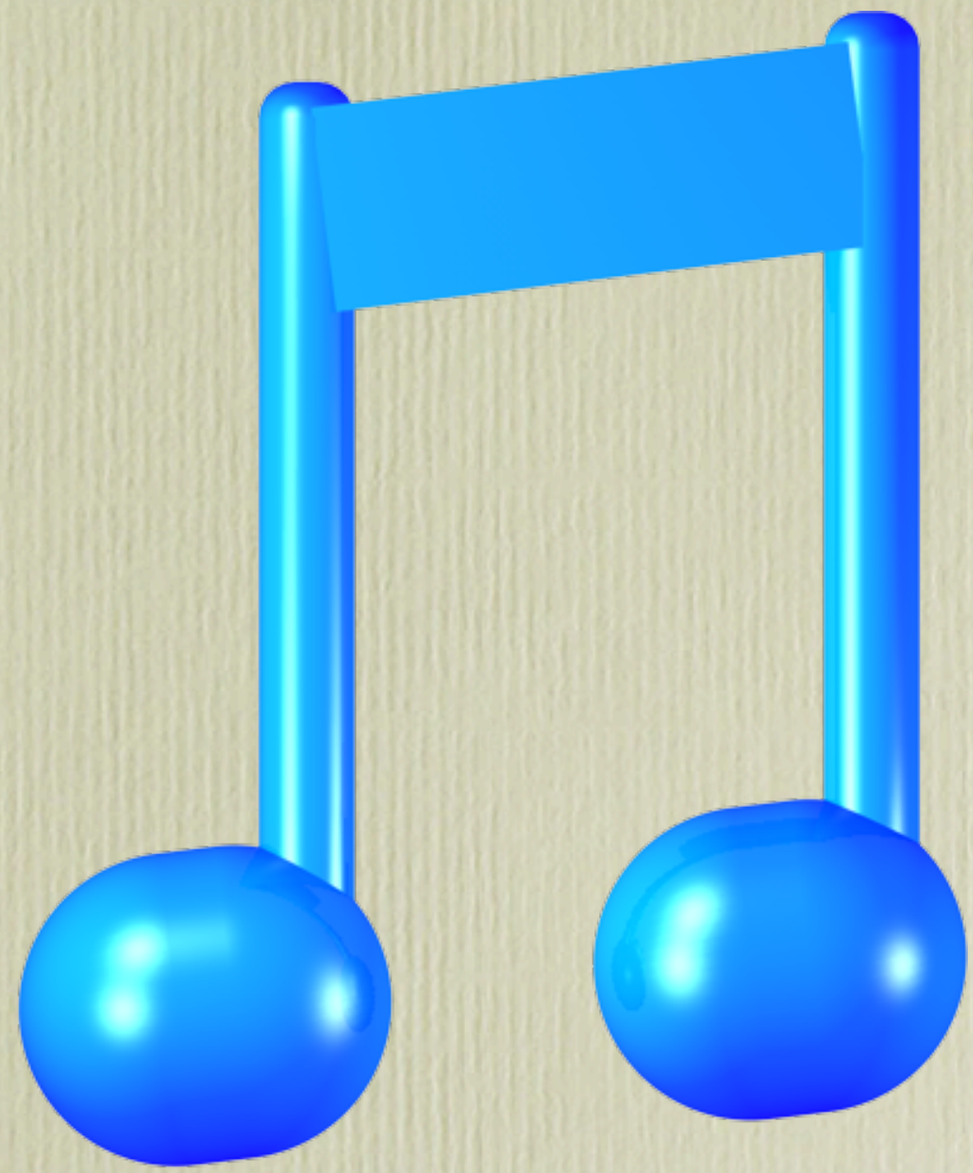
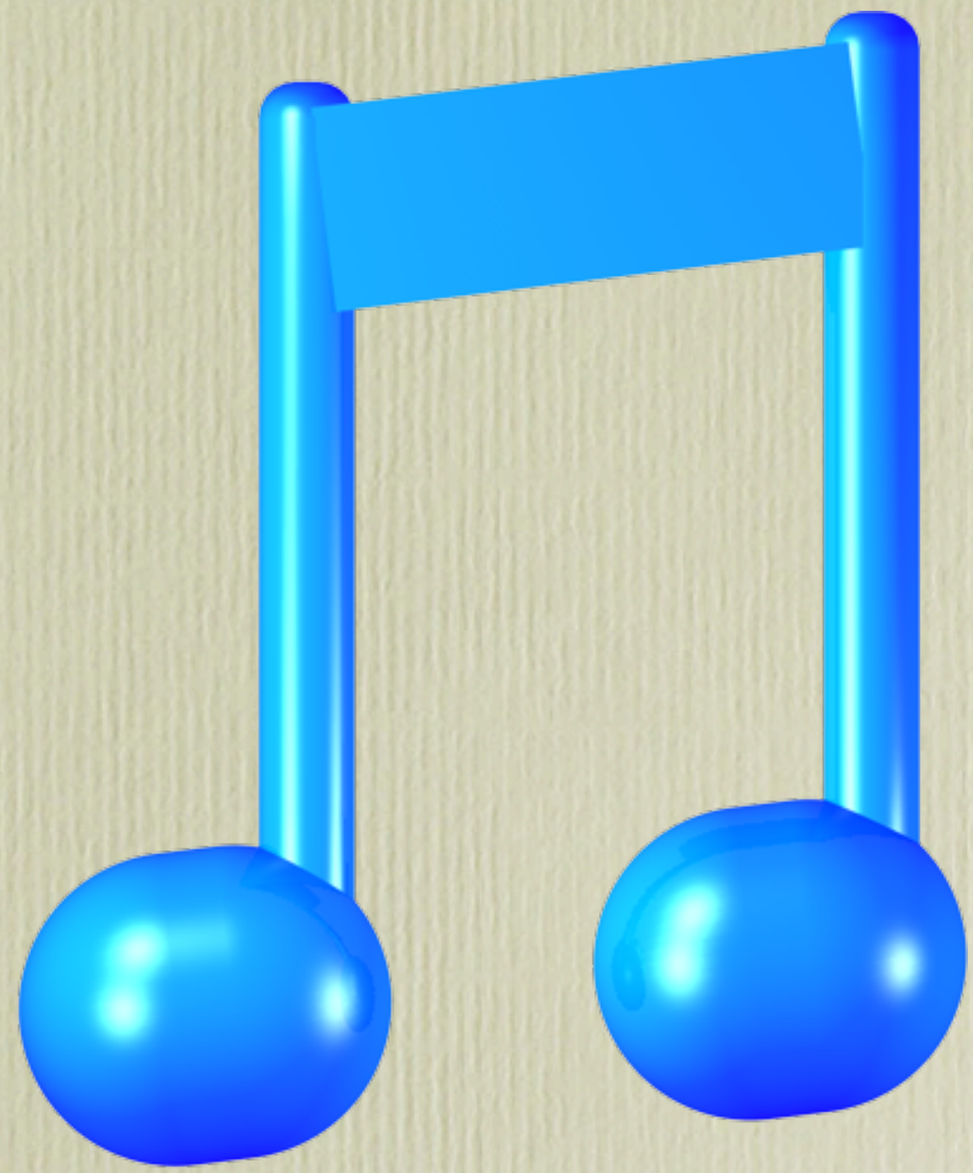
Bachelor's Thesis

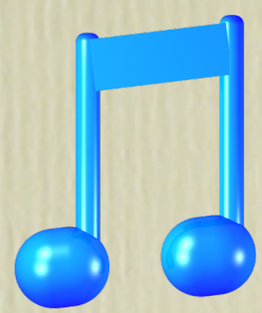
Audio Denoising with Non-Local Algorithms from Image Processing

Jan Hendrik Dithmar
dithmar@mia.uni-saarland.de





Motivation



- Similar tools: f. ex. the Fourier transform or wavelet-based approaches.
- Concept of noise also present in both approaches.
- We want insights in audio denoising from an image processing viewpoint.

Image source:
http://www.1clipart.com

Outline

I. Similarity Measurement (+ prerequisites)

II. Methods

- Non-local inpainting
- Averaging on a sample basis
- NL-means
- NL-means by Szlam

III. Experiments

I

Similarity Measurement

Similarity Measurement

- Prerequisites
  - Short-Time Fourier Transform
  - Hamming Window
- Measure
- Algorithm

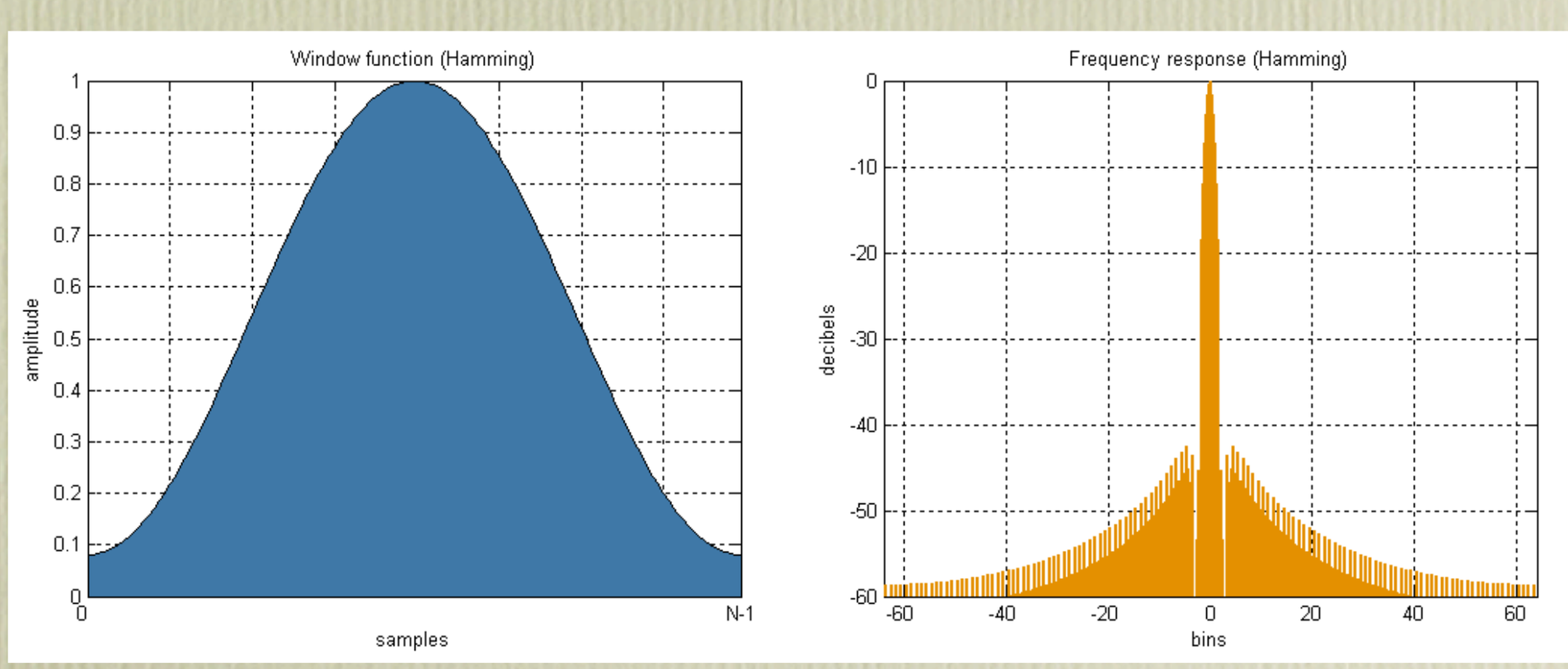
Short-Time Fourier Transform

- Basic idea:
  - Use a (sliding) window instead of the whole data.
    - Compute the Fourier transform for each window.
- Tells you something about the temporal appearance.

Hamming Window

The Hamming window is given by

$$w(n) = 0.53836 - 0.46164 \cos\left(\frac{2\pi n}{N-1}\right)$$



Hamming window. **Left:** Window function in the sample domain. **Right:** Frequency response. **Source:** Wikipedia.

Measure

- Power spectrum:

$$|\hat{f}|^2 = \left(\mathrm{Re}(\hat{f})\right)^2 + \left(\mathrm{Im}(\hat{f})\right)^2$$

- Measure:

$$\left\{ i \;\middle|\; \max_i \frac{\left\langle |\hat{f}_i|^2, |\hat{f}_{ref}|^2 \right\rangle}{\left\| |\hat{f}_i|^2 \right\| \cdot \left\| |\hat{f}_{ref}|^2 \right\|} \right\}$$

$|\hat{f}_i|^2$: power spectrum of current frame

$|\hat{f}_{ref}|^2$: power spectrum of reference frame

Similarity Measurement (Algorithm)

1. Take a part of the audio file of length N
2. Apply the Hamming window
3. Use the FFT on this window
4. Compute the power spectrum for this window
5. Go step size many steps forward and repeat until EOF

II

Methods

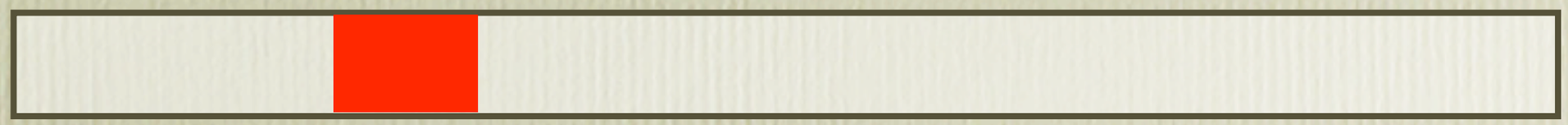
Non-Local Inpainting

- Find similar frames and choose one.
- Replace the noisy frame by simply copying the similar one.

Non-Local Inpainting

Noisy file

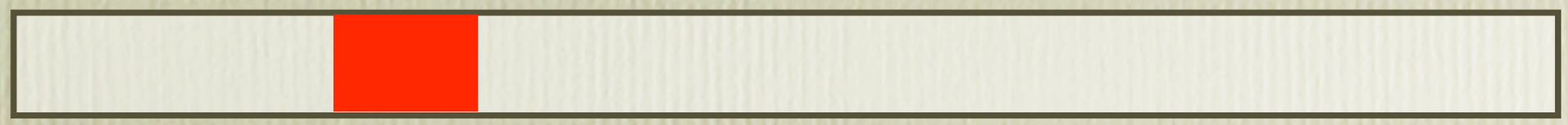
Denoised file



Non-Local Inpainting

Noisy file

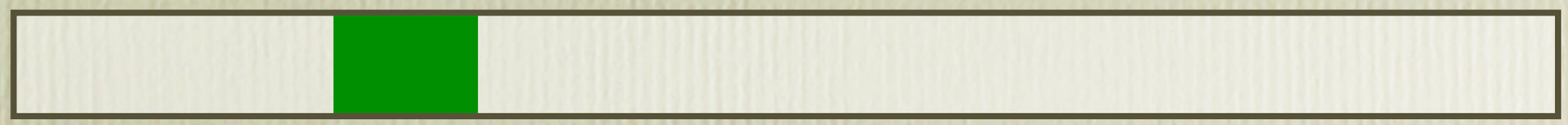
Denoised file



Non-Local Inpainting

Noisy file

Denoised file



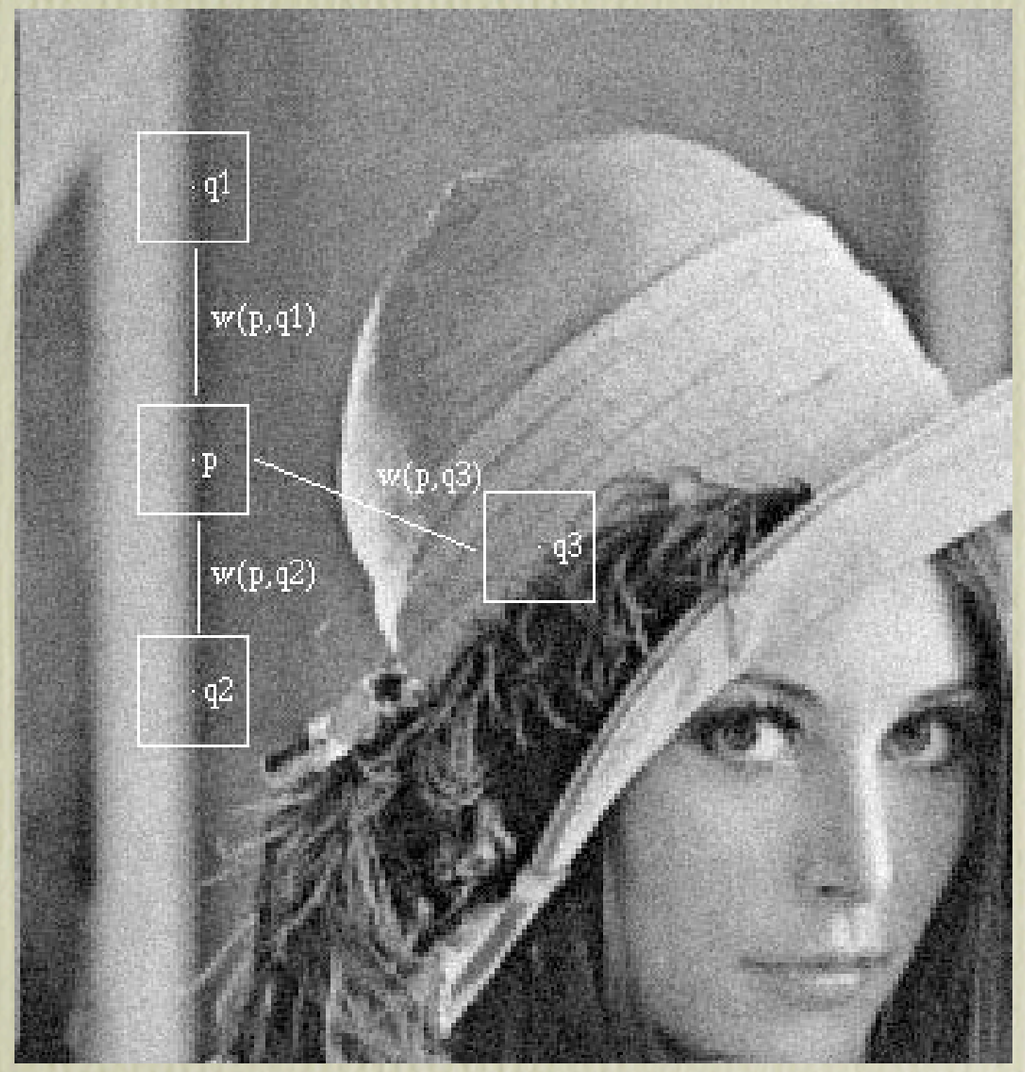
Averaging on a Sample Basis

- Instead of choosing one similar frame, average over all similar frames.
- Weighting function:

$$w = \frac{1}{frames}$$

- Do a non-local inpainting with the averaged result.

NL-means in Image Processing



NL-means strategy. Similar pixel neighborhoods give a large weight, while much different neighborhoods give a small weight. **Authors:** A. Buades, B. Coll, J.-M. Morel.

NL-means

- Adaptive weighting: similar structures are more important.
- Weighting function:

$$w(cur) = e^{-\lambda \cdot \frac{best - cur}{best}}$$

- Do a non-local inpainting with the averaged result.

Averaging / NL-means

Noisy file

Averaging on a Sample Basis

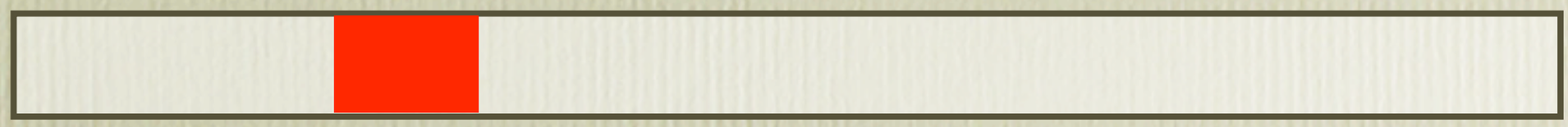
$$w = \frac{1}{frames}$$

weighting function

NL-means

$$w(cur) = e^{-\lambda \cdot \frac{best-cur}{best}}$$

Denoised file



Averaging / NL-means

Noisy file

Averaging on a Sample Basis

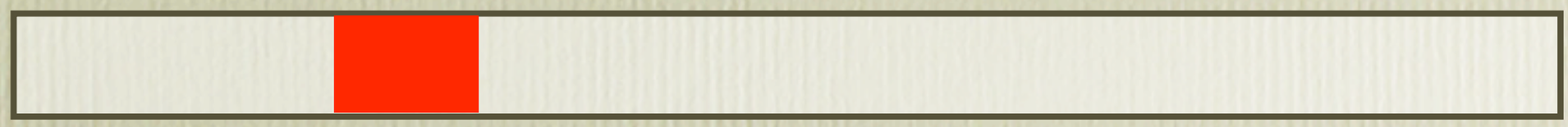
$$w = \frac{1}{frames}$$

weighting function

NL-means

$$w(cur) = e^{-\lambda \cdot \frac{best-cur}{best}}$$

Denoised file



Averaging / NL-means

Noisy file

Averaging on a Sample Basis

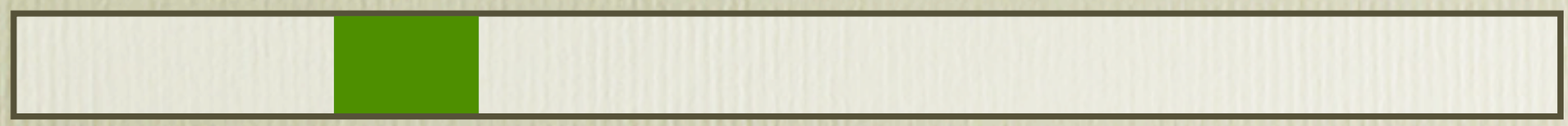
$$w = \frac{1}{frames}$$

weighting function

NL-means

$$w(cur) = e^{-\lambda \cdot \frac{best-cur}{best}}$$

Denoised file



NL-means by Szlam

- Entire signal is degraded by Gaussian noise.
- Only slight modifications of the code needed.
- Introduced clicking noise, but why?
- Further investigation was not possible within the time frame of this thesis.

Performance

- Memory requirements? Runtime?
- configuration 1:
  - step size = 256, window size = 1024
- configuration 2:
  - step size = 2048, window size = 4096

Memory Requirements and Runtime

Example: 21 MB audio file.
Searching the 10 best results, using non-local inpainting with the best result for denoising.

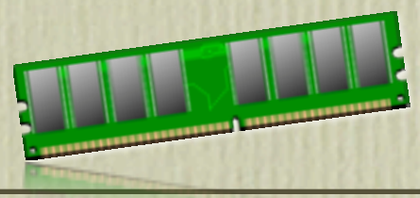
	configuration 1	configuration 2
		
		

Image sources:
http://www.1clipart.com
http://www.apple.com

Memory Requirements and Runtime

Example: 21 MB audio file.
Searching the 10 best results, using non-local inpainting with the best result for denoising.

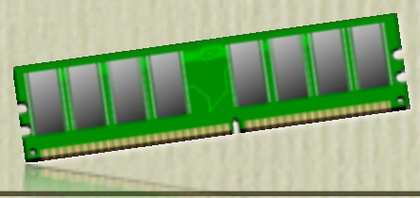
	configuration 1	configuration 2
	288 MB	165 MB
	28 seconds	7 seconds

Image sources:
http://www.1clipart.com
http://www.apple.com

III

Experiments

Averaging on a Sample Basis Example

- White noise with a duration of 1 second
  - Original file
  - Noisy version
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

Averaging on a Sample Basis Example

- White noise with a duration of 1 second
  - Original file ***now playing***
  - Noisy version
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

Averaging on a Sample Basis Example

- White noise with a duration of 1 second
  - Original file
  - Noisy version ***now playing***
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

Averaging on a Sample Basis Example

- White noise with a duration of 1 second
  - Original file
  - Noisy version
  - Denoised version ***now playing***
- Configuration
  - step size: 2048
  - window width: 4096

NL-means Example

- Silence with a duration of 0.5 seconds
  - Original file
  - Noisy version
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

NL-means Example

- Silence with a duration of 0.5 seconds
  - Original file ***now playing***
  - Noisy version
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

NL-means Example

- Silence with a duration of 0.5 seconds
  - Original file
  - Noisy version ***now playing***
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

NL-means Example

- Silence with a duration of 0.5 seconds
  - Original file
  - Noisy version
  - Denoised version ***now playing***
- Configuration
  - step size: 2048
  - window width: 4096

NL-means by Szlam Example

- Gaussian noise on the whole signal
    - Original file
    - Noisy version
    - First iteration
    - Third iteration
- Configuration
    - step size: 2048, window width: 4096

NL-means by Szlam Example

- Gaussian noise on the whole signal
  - Original file ***now playing***
  - Noisy version
  - First iteration
  - Third iteration
- Configuration
  - step size: 2048, window width: 4096

NL-means by Szlam Example

- Gaussian noise on the whole signal
  - Original file ***now playing***
  - Noisy version
  - First iteration
  - Third iteration
- Configuration
  - step size: 2048, window width: 4096

NL-means by Szlam Example

- Gaussian noise on the whole signal
    - Original file
    - Noisy version ***now playing***
    - First iteration
    - Third iteration
- Configuration
    - step size: 2048, window width: 4096

NL-means by Szlam Example

- Gaussian noise on the whole signal
    - Original file
    - Noisy version ***now playing***
    - First iteration
    - Third iteration
- Configuration
    - step size: 2048, window width: 4096

NL-means by Szlam Example

- Gaussian noise on the whole signal
    - Original file
    - Noisy version
    - First iteration ***now playing***
    - Third iteration
- Configuration
    - step size: 2048, window width: 4096

NL-means by Szlam Example

- Gaussian noise on the whole signal
    - Original file
    - Noisy version
    - First iteration ***now playing***
    - Third iteration
- Configuration
    - step size: 2048, window width: 4096

NL-means by Szlam Example

- Gaussian noise on the whole signal
    - Original file
    - Noisy version
    - First iteration
    - Third iteration ***now playing***
- Configuration
    - step size: 2048, window width: 4096

NL-means by Szlam Example

- Gaussian noise on the whole signal
    - Original file
    - Noisy version
    - First iteration
    - Third iteration ***now playing***
- Configuration
    - step size: 2048, window width: 4096

Non-Local Inpainting Example

- Sine noise with a duration of 5 seconds
  - Original file
  - Noisy version
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

Non-Local Inpainting Example

- Sine noise with a duration of 5 seconds
  - Original file ***now playing***
  - Noisy version
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

Non-Local Inpainting Example

- Sine noise with a duration of 5 seconds
  - Original file ***now playing***
  - Noisy version
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

Non-Local Inpainting Example

- Sine noise with a duration of 5 seconds
  - Original file
  - Noisy version ***now playing***
  - Denoised version
- Configuration
  - step size: 2048
  - window width: 4096

Non-Local Inpainting Example

- Sine noise with a duration of 5 seconds
  - Original file
  - Noisy version
  - Denoised version ***now playing***
- Configuration
  - step size: 2048
  - window width: 4096

One giant leap for me...

...but a small step for mankind

- Non-Local Inpainting
  - noisy structures (up to 5 seconds) can be almost fully reconstructed
- Averaging on Sample Basis
  - can fill in something plausible
- NL-means approach
  - works well with appropriate threshold λ

Summary

- Approach is simple, but it works.
- It is possible to use image processing methods for audio denoising.
- NL-means also seems to be well suited for audio data.

References

- Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 60-65. IEEE Computer Society, June 2005.

- Alexei A. Efros and Thomas K. Leung. Texture synthesis by non-parametric sampling. In *Proc. IEEE International Conference on Computer Vision*, pages 1033-1038. IEEE Computer Society, September 1999.

- Please check the thesis for further references.

Thank you!