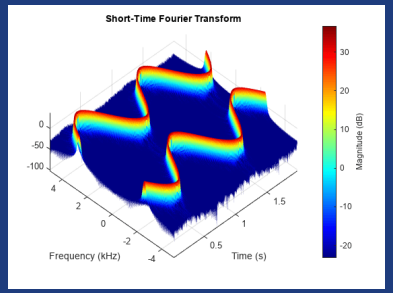
Automatic Tuning of STFT

Differentiable short-time Fourier transform



Maxime Leiber, Yosra Marnissi (Safran), Mohammed El-Badaoui (Safran), Laurent Massoulié (Inria)

08/11/2023

SAFRAN

Agenda

01
Context and motivation

02
Differentiable short-time Fourier transform

03
Optimization for best representation

04
Optimization for task performance

05
Conclusion

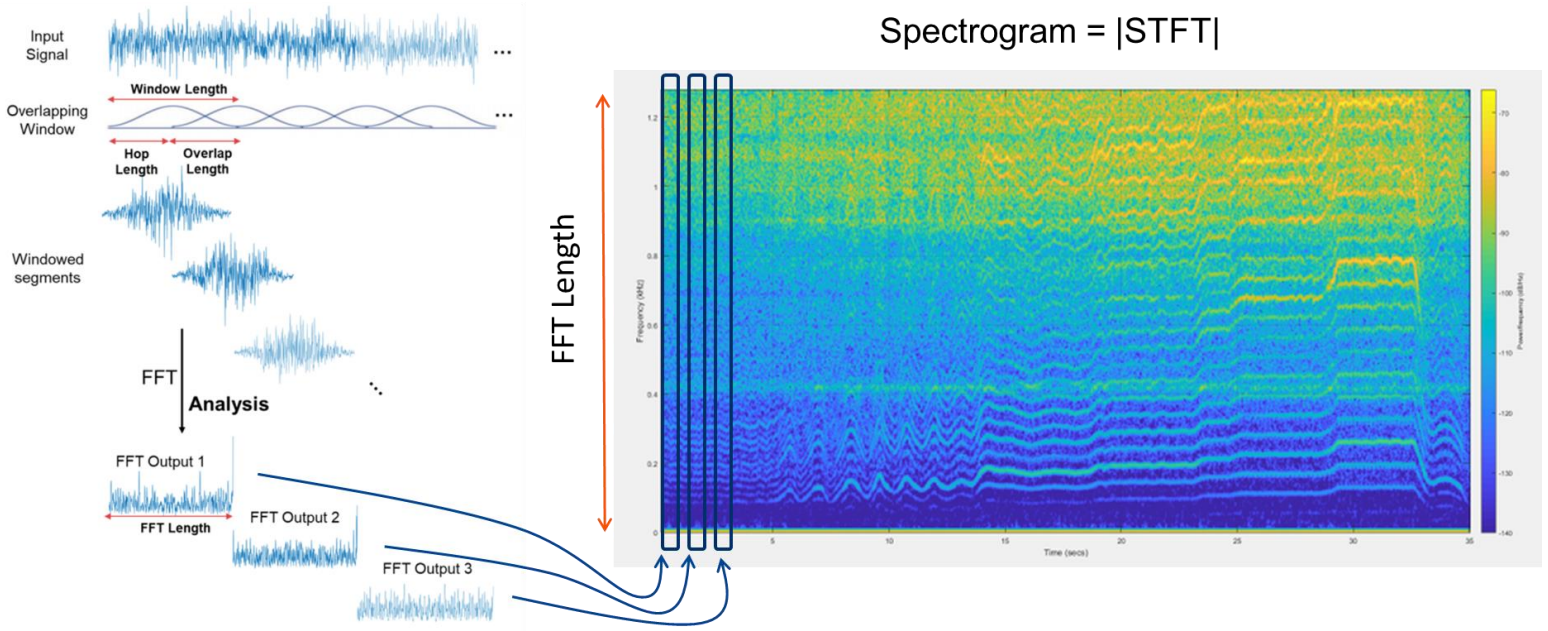
Part 1

Context and motivation

SAFRAN

Context

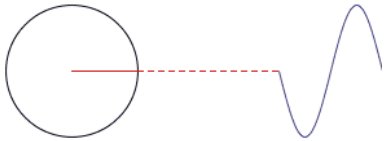
- **Short-Time Fourier Transform:** evolution of the frequency content over time
- **Parameters:** window shape, window length, hop length



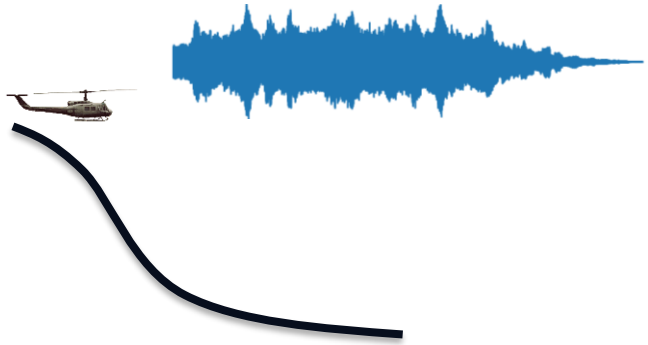
STFT in aeronautics

- **STFT remains one of the most widely used TF representations, especially in aeronautics**
  - Simplicity and low computational requirements
  - Sinusoidal basis functions are adapted to **rotating machines**
  - Data collected from aircraft sensors, such as vibration exhibit **non-stationary** behavior especially under **time-varying operational conditions**

Rotating machines generate quasi-periodic signals



Non-stationary signals



SAFRAN

STFT in aeronautics

- A **graphical display**, well adapted to non-stationarity that eases signal interpretation, in particular in exploratory data analysis
- Basis **representation to perform several tasks** such as instantaneous frequency estimation [1] and anomaly detection [2] in non-stationary regimes
- Input in many **data-based solutions** such as Non Negative Matrix Factorization [3] and Neural Networks [4]
- Used to **implement low-computationaly signal-processing tools** such as spectral kurtosis [5]

[1] Leclère, Q., et al. (2016). A multi-order probabilistic approach for Instantaneous Angular Speed tracking debriefing of the CMMNO 14 ‘diagnosis contest. Mechanical Systems and Signal Processing, 81, 375-386.

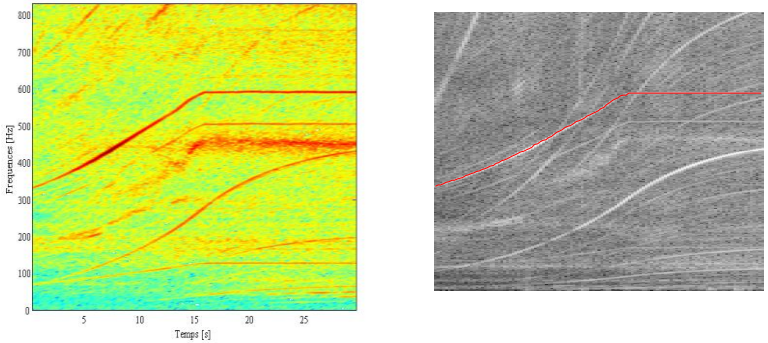
[2] Sayed, M. A. (2018). Représentations pour la détection d’anomalies: Application aux données vibratoires des moteurs d’avions (Doctoral dissertation, Université Paris Saclay (COmUE)).

[3] Lacaille, J., et al. (2020). Automatic Detection of Vibration Patterns During Production Test of Aircraft Engines. In Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 2 1 (pp. 81-93). Springer International Publishing.

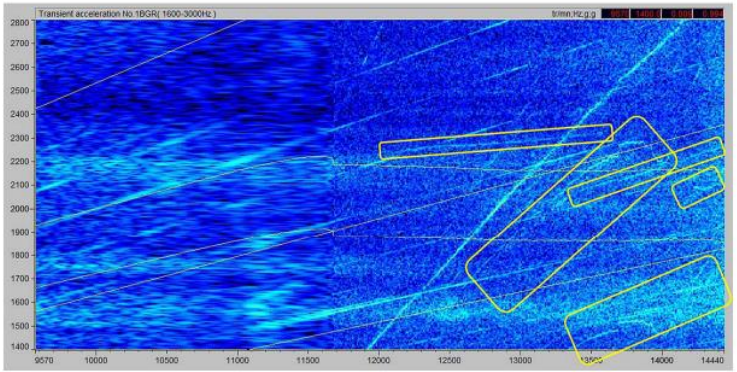
[4] Kulagin, V. et al (2021). Identification of temporal anomalies of spectrograms of vibration measurements of a turbine generator rotor using a recurrent neural network autoencoder. Russian Technological Journal, 9(2), 78-87.

[5] Antoni, J. et al, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, Mechanical Systems and Signal Processing 20 (2) (2006) 308–331.

STFT in aeronautics



Instantaneous frequency tracking from spectrogram of a helicopter vibration signal

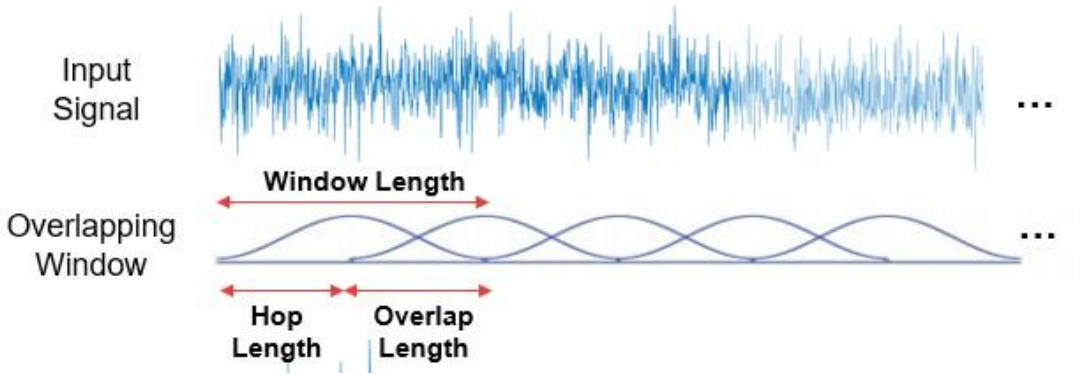


Manual annotation of suspicious signature by Safran Expert

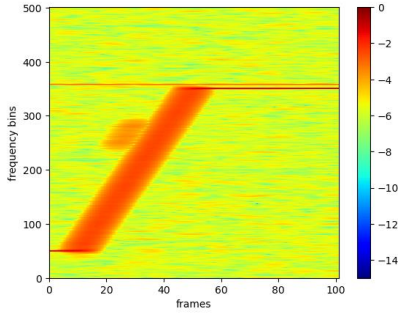
Limitation

STFT is highly sensitive to its parameters

- Window length: determines the trade-off between temporal and frequency resolution
- Temporal position of the windows (hop length or overlap length): allows the windows to be aligned with the signal components

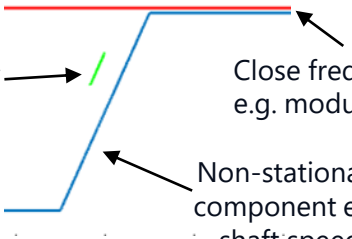


Impact of the window length



Large window length

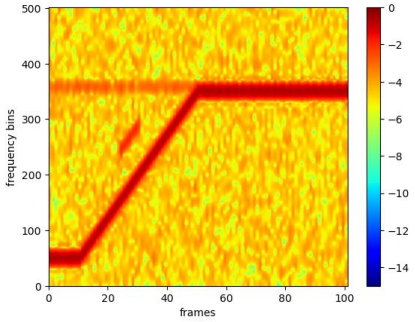
Good frequency resolution



Transient frequency
e.g. fan flutter

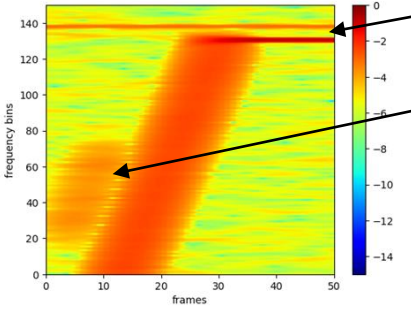
Close frequencies
e.g. modulations

Non-stationary
component e.g.
shaft speed



Small window length

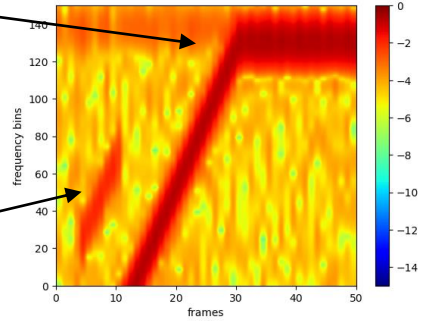
Good temporal resolution



Distinguish close
frequencies

Coarse temporal
resolution

Does not
distinguish close
frequencies

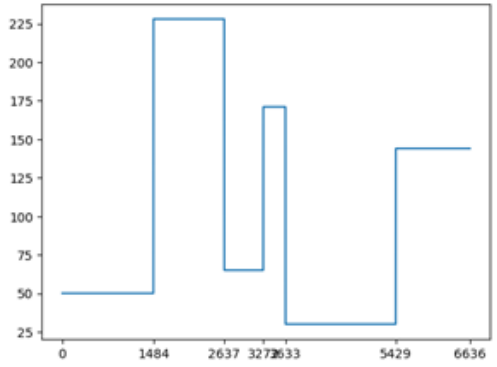


Better localize
transient events

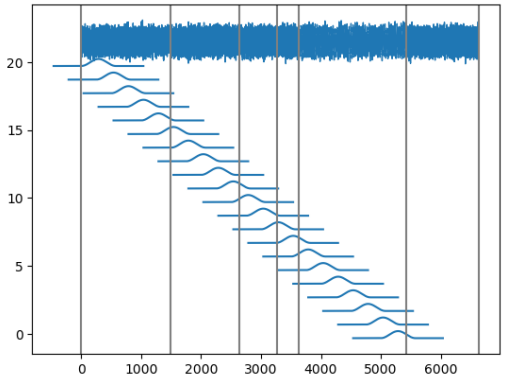
SAFRAN

Impact of the hop length

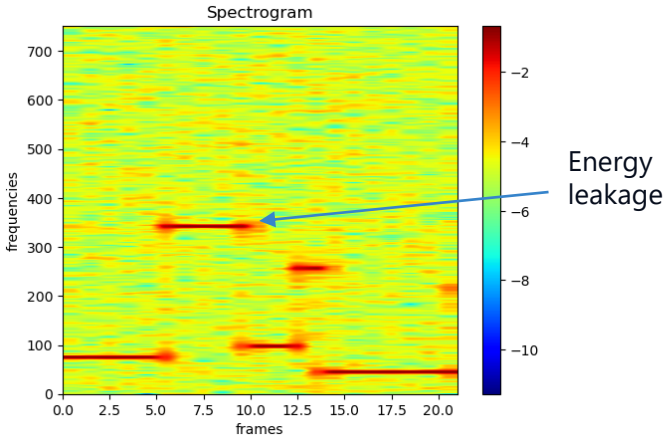
Frequencies over time



Windows are uniformly spread over signals



Spectrogram



Uniformly distributed windows along the signal

- Misalignment with the spectral content of the signal
- Energy leakage
- Errors in indicators and predictions

SAFRAN

Some proposed solutions in the literature

- **Pre-processing by searching optimal parameters with grid-search**
  - Define a representation metric to promote some target characteristics **[1]**
    - E.g. to promote energy concentration use some parsimonious losses such as kurtosis, entropy
  - Define a grid of parameters
  - Evaluate the metric on STFT representations generated from parameters of the grid
  - Choose the one that optimize the metric
  - Compute the STFT with these optimal parameters

- **Post-processing the STFT [2]**
  - Compute the STFT with given parameters
  - Transformation to sharpen the ridges in the TF plane e.g. reassignement, synhcrosqueezing

[1] Meignen, S. et al (2020, May). On the use of Rényi entropy for optimal window size computation in the short-time Fourier transform. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5830-5834). IEEE.

[2] Auger, F. et al (2013). Time-frequency reassignment and synchrosqueezing: An overview. IEEE Signal Processing Magazine, 30(6), 32-41.

Part 2

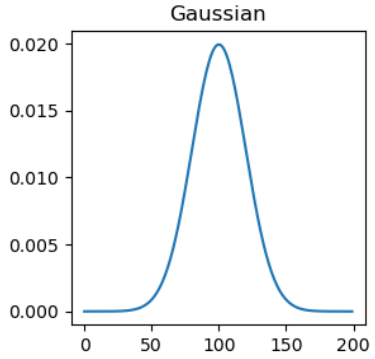
Differentiable Short Time Fourier Transform

SAFRAN

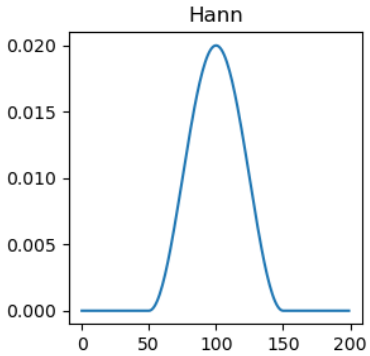
Objectives

STFT

$$S_L[i,f] = \sum_{k=t_i}^{t_i+L-1} w_L(k-t_i)\, s[k]\; e^{-\frac{2j\pi kf}{L}}$$



$$w_L(k) = \frac{1}{\sqrt{2\pi}L}\exp\left(-\frac{k^2}{2L^2}\right)$$



$$w_L(k) = \frac{1}{2} - \frac{1}{2}\cos\frac{2k\pi}{L}$$

L : integer window length
t_i: integer temporal position of frames

Objectives

- Make these parameters continuous
- Make STFT differentiable w.r.t. these parameters
- In order to optimize them using gradient descent

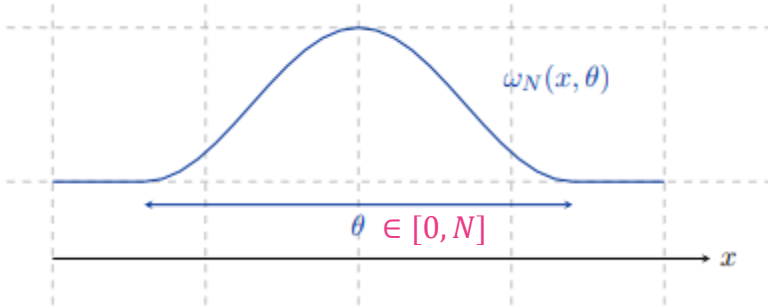
Proposed formulation

DSTFT

$$S_{\Omega}[i,f] = \sum_{k \in Z} \omega_N\left(k - t_i, \theta_{i,f}\right) s[k]\ e^{-\frac{2j\pi kf}{N}}$$

$$\Omega = \left\{\mathrm{N}, \theta_{i,f}, \mathrm{t_i}\right\}$$

N : window support (integer)
t_i: temporal position of frames (continuous)
$\theta_{i,f}$: window length (continuous)



- $\omega_N(x,\theta)$ is a two variable function
  - x is the domain of definition
  - θ determines the length of ω_N
- Both variables are **continuous**
- N is an **upper bound** of θ, characteristic of the maximum frequency resolution of the spectrogram

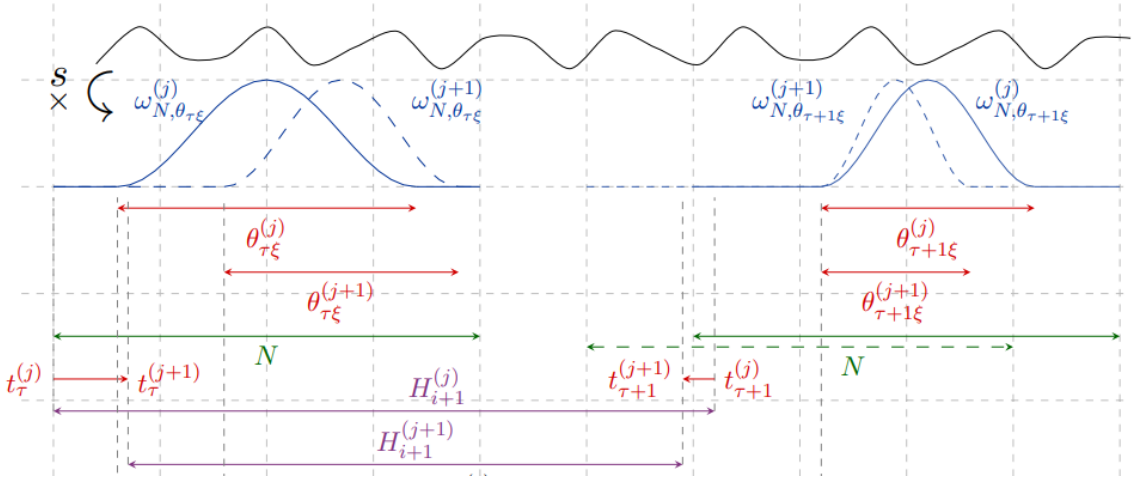
Example: for a Hann function $\omega(x,\theta) = \left(\frac{1}{2} - \frac{1}{2}\cos\frac{2\pi.x}{\theta}\right)1_{0\le x\le\theta}$
The translated function is

$$\omega_N(x,\theta) = \omega\left(x + \frac{\theta - N + 1}{2}, \theta\right)$$

Results

- **Proposition**

The modified STFT is differentiable with respect to both window length $\theta_{i,f}$ and temporal positions t_i (or equivalently the hop length $h_i = t_i - t_{i-1}$)



Interpretation

DSTFT

$$S_{\Omega}[i,f] = \sum_{k \in Z} \omega_N\left(k - t_i, \theta_{i,f}\right) s[k]\ e^{-\frac{2j\pi kf}{N}}$$

$$\Omega = \left\{\mathrm{N}, \theta_{i,f}, \mathrm{t_i}\right\}$$

- STFT is differentiable with respect to its parameters i.e., $\frac{\partial S_{\Omega}[i,f]}{\partial \theta_{i,f}}$ and $\frac{\partial S_{\Omega}[i,f]}{\partial t_i}$ exist and are finite
- Given a loss $L(S_{\Omega})$, then we can minimize $L(S_{\Omega})$ with respect to Ω using gradient based optimization with

$$\frac{\partial L(S_{\Omega})}{\partial \boldsymbol{\Omega}} = \sum_{i,f} \frac{\partial L(S_{\Omega})}{\partial \boldsymbol{S_{\Omega}}[i,f]} \frac{\partial \boldsymbol{S_{\Omega}}[i,f]}{\partial \boldsymbol{\Omega}}$$

Exists because DSTFT is differentiable

- We only need to adjust a higher bound N for the window length which is the support of the tapering window
- The temporal and frequency resolution is determined by $\theta_{i,f}$
- The window length $\theta_{i,f}$ **can take different values in the time-frequency plane** and be **optimized automatically**
- The hop length $h_i = t_i - t_{i-1}$ **can change over time** and be **optimized automatically**

Gradient expressions

- **Gradient computation are straightforward**
  - **Computationally equivalent to forward pass**

$$\frac{\partial \mathcal{S}[\tau, f]}{\partial \theta_{\tau,\xi}} = \mathcal{S}_{\bar{\omega}_N}[\tau, \xi]\mathbf{1}_\tau(\tau)\mathbf{1}_\xi(\xi) \quad \text{with} \quad \bar{\omega}_N(k, \theta_{\tau,\xi}) = \partial_\theta \omega_N(k, \theta_{\tau,\xi}).$$

$$\frac{\partial \mathcal{S}_{\omega_N}[\eta, \xi]}{\partial t_\tau} = \mathcal{S}_{\tilde{\omega}_N}[\tau, \xi]\mathbf{1}_\tau(\eta). \quad \text{with} \quad \tilde{\omega}_N(k, \theta_{\tau,\xi}) = \partial_x \omega_N(k, \theta_{\tau,\xi})$$

Variable parameters

$$\frac{\partial \mathcal{L}}{\partial \theta_{\tau,\xi}} = \frac{\partial \mathcal{L}}{\partial \mathcal{S}_{\omega_N}[\tau, \xi]} \frac{\partial \mathcal{S}_{\omega_N}[\tau, \xi]}{\partial \theta_{\tau,\xi}} = \frac{\partial \mathcal{L}}{\partial \mathcal{S}_{\omega_N}[\tau, \xi]} \mathcal{S}_{\bar{\omega}_N}[\tau, \xi]$$

$$\frac{\partial \mathcal{L}}{\partial t_\tau} = \sum_{\xi=0}^{N-1} \frac{\partial \mathcal{L}}{\partial \mathcal{S}_{\omega_N}[\tau, \xi]} \frac{\partial \mathcal{S}_{\omega_N}[\tau, \xi]}{\partial t_\tau} = \sum_{\xi=0}^{N-1} \frac{\partial \mathcal{L}}{\partial \mathcal{S}_{\omega_N}[\tau, \xi]} \mathcal{S}_{\tilde{\omega}_N}[\tau, \xi]$$

Shared parameters

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{\tau=0}^{T-1} \sum_{\xi=0}^{N-1} \frac{\partial \mathcal{L}}{\partial \mathcal{S}_{\omega_N}[\tau, \xi]} \frac{\partial \mathcal{S}_{\omega_N}[\tau, \xi]}{\partial \theta} = \left\langle \frac{\partial \mathcal{L}}{\partial \mathcal{S}_{\omega_N}}, \mathcal{S}_{\bar{\omega}} \right\rangle,$$

$$\frac{\partial \mathcal{L}}{\partial H} = \sum_{\tau=0}^{T-1} \sum_{\xi=0}^{N-1} \frac{\partial \mathcal{L}}{\partial \mathcal{S}_{\omega_N}[\tau, \xi]} \frac{\partial \mathcal{S}_{\omega_N}[\tau, \xi]}{\partial H} = \left\langle \frac{\partial \mathcal{L}}{\partial \mathcal{S}_{\omega_N}}, \mathcal{S}_{\tilde{\omega}} \right\rangle$$

DSTFT for best representation

- **Define a loss $L(S_\Omega)$ that is differentiable almost everywhere w.r.t to S_Ω**
  - Kurtosis of spectrogram
  - Entropy of spectrogram
  - Norms ℓ^1, ℓ^2, $\frac{\ell^1}{\ell^2}$ etc. of spectrogram
- **Compute gradient of the loss w.r.t toS_Ω**
  - E.g., $\frac{\partial \ell^1(S)}{\partial S_\Omega[i,f]} = \frac{S_\Omega[i,f]}{|S_\Omega[i,f]|}$ $if\ S_\Omega[i,f] \neq 0$ else 0
- **Compute gradient of the loss w.r.t to parameters**

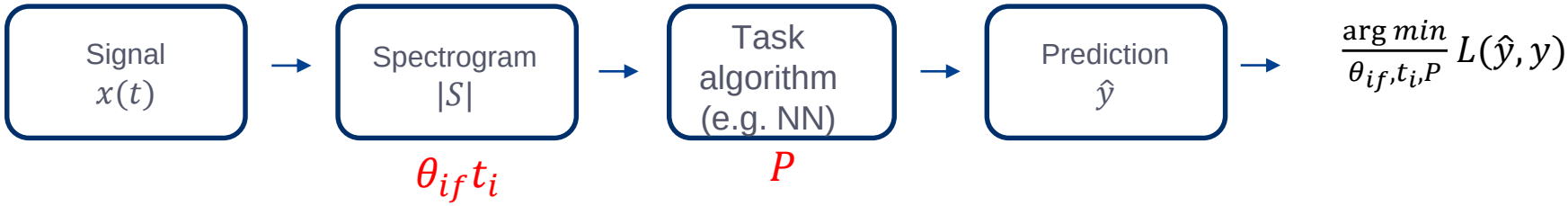
$$\frac{\partial L(S_\Omega)}{\partial \boldsymbol{\Omega}} = \sum_{i,f} \frac{\partial L(S_\Omega)}{\boldsymbol{\partial} S_{\boldsymbol{\Omega}}[i,f]} \frac{\boldsymbol{\partial} S_{\boldsymbol{\Omega}}[i,f]}{\partial \boldsymbol{\Omega}}$$

- **A gradient optimization step**

$$\boldsymbol{\Omega_{t+1}} = \boldsymbol{\Omega_t} - \boldsymbol{\alpha}.\frac{\partial L(S_\Omega)}{\partial \boldsymbol{\Omega_t}}$$

DSTFT for task performance

- **DSTFT can be seen as a neural network layer where the weights are the window lengths and positions**
  - Weights of NN and STFT parameters can be optimized jointly during the training of the NN
  - Can be also used for any task algorithm for regression/classification/anomaly detection e.g., SVM, Linear regression having as input STFT
  - Thus, parameters of STFT will be optimized from a whole dataset to achieve the optimal task metrics

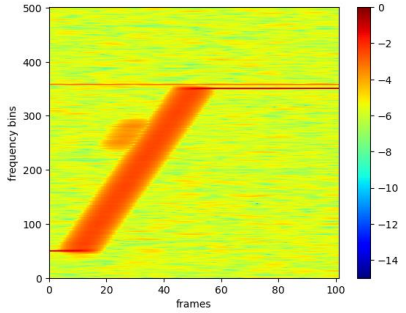


Part 3

Optimization for best representation

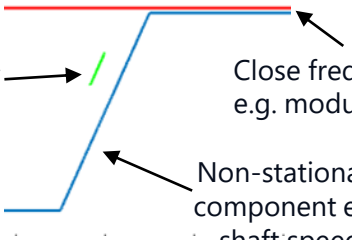
SAFRAN

Experiment



Large window length

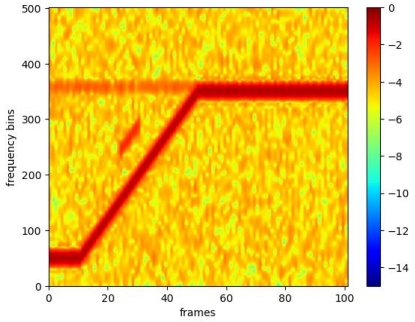
Good frequency resolution



Transient frequency
e.g. fan flutter

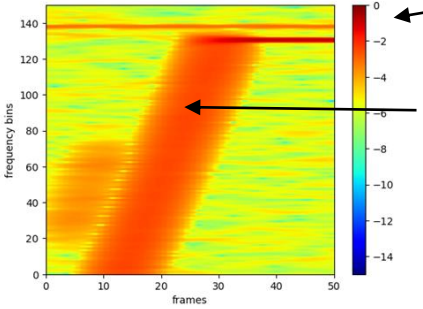
Close frequencies
e.g. modulations

Non-stationary
component e.g.
shaft speed



Small window length

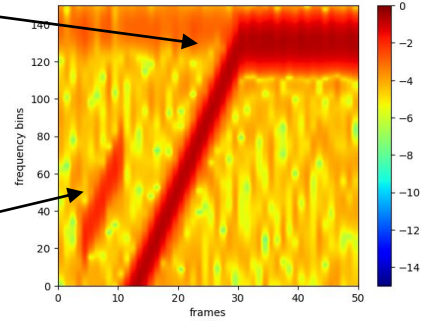
Good temporal resolution



Distinguish close
frequencies

Coarse temporal
resolution

Does not
distinguish close
frequencies



Better localize
transient events

SAFRAN

Considered criterion

- Energy concentration loss:
  - Entropy of the representation :

$$\mathcal{H}(\mathcal{S}_{\omega_N};\Omega) = -\sum_{\tau,\xi} p_{\tau\xi}\log(p_{\tau\xi}) \qquad \text{where } p_{\tau\xi} = \frac{|\mathcal{S}_{\omega_N}[\tau,\xi]|}{\sum_{k,l}|\mathcal{S}_{\omega_N}[k,l]|}.$$

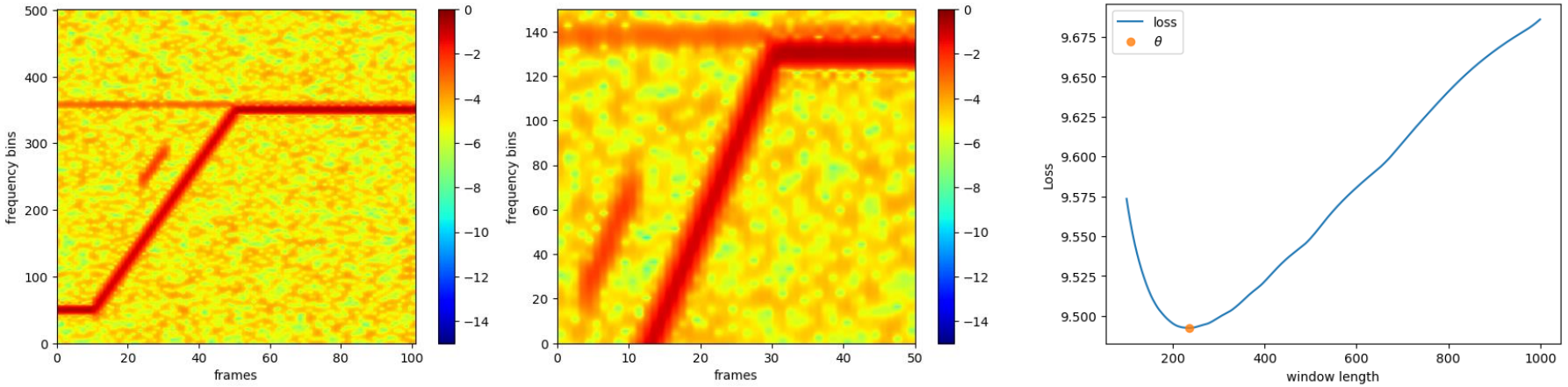
  - Kurtosis of the frame representation:

$$\mathcal{K}(\mathcal{S}_{\omega_N};\Omega) = \frac{1}{\sum_{\tau=0}^{T-1}\lambda_\tau}\sum_{\tau=0}^{T-1}\lambda_\tau\frac{\mathbb{E}_\xi[\mathcal{S}_{\omega_N}[\tau,\xi]^4]}{\mathbb{E}_\xi[\mathcal{S}_{\omega_N}[\tau,\xi]^2]^2}$$

- Regularisation for robustness to noise (Non-local total variation):

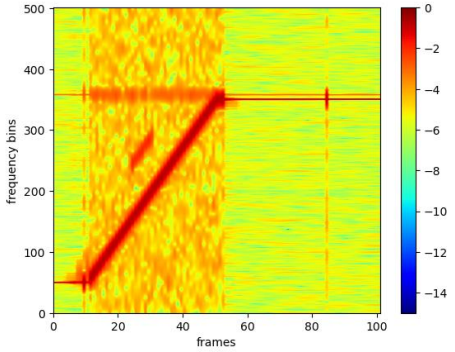
$$\mathcal{R}(\Theta) = \sum_{i,f}\sqrt{\sum_{(j,k)\in V_{i,f}}\gamma_{(j,k);(i,f)}(\theta_{i,f}-\theta_{j,k})^2}$$

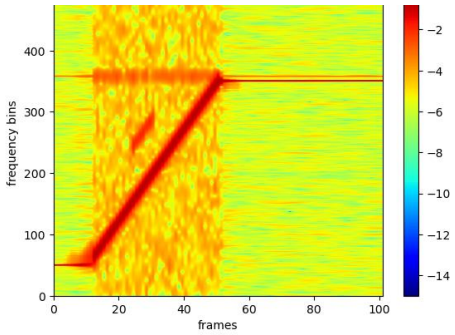
DSTFT with single window length θ



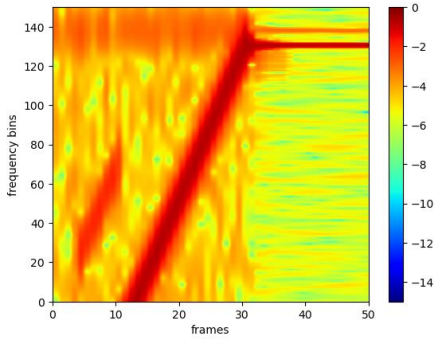
- Trade-off between time and frequency resolution
- Best representation using single window according to the criterion
- Does not distinguish well close frequencies neither localize well transient event

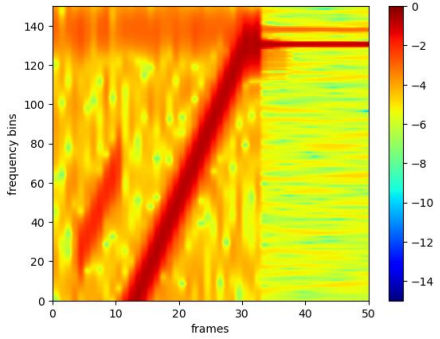
DSTFT with time-varying window length θ_t



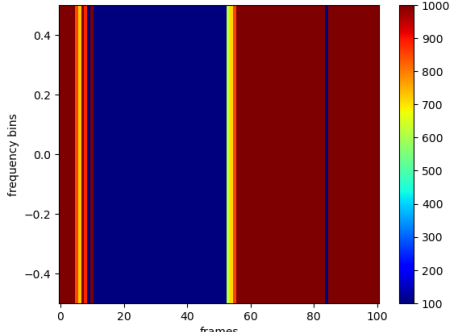


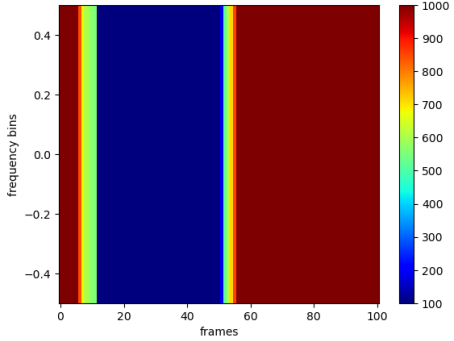
Spectrogram





Zoomed-in spectrogram





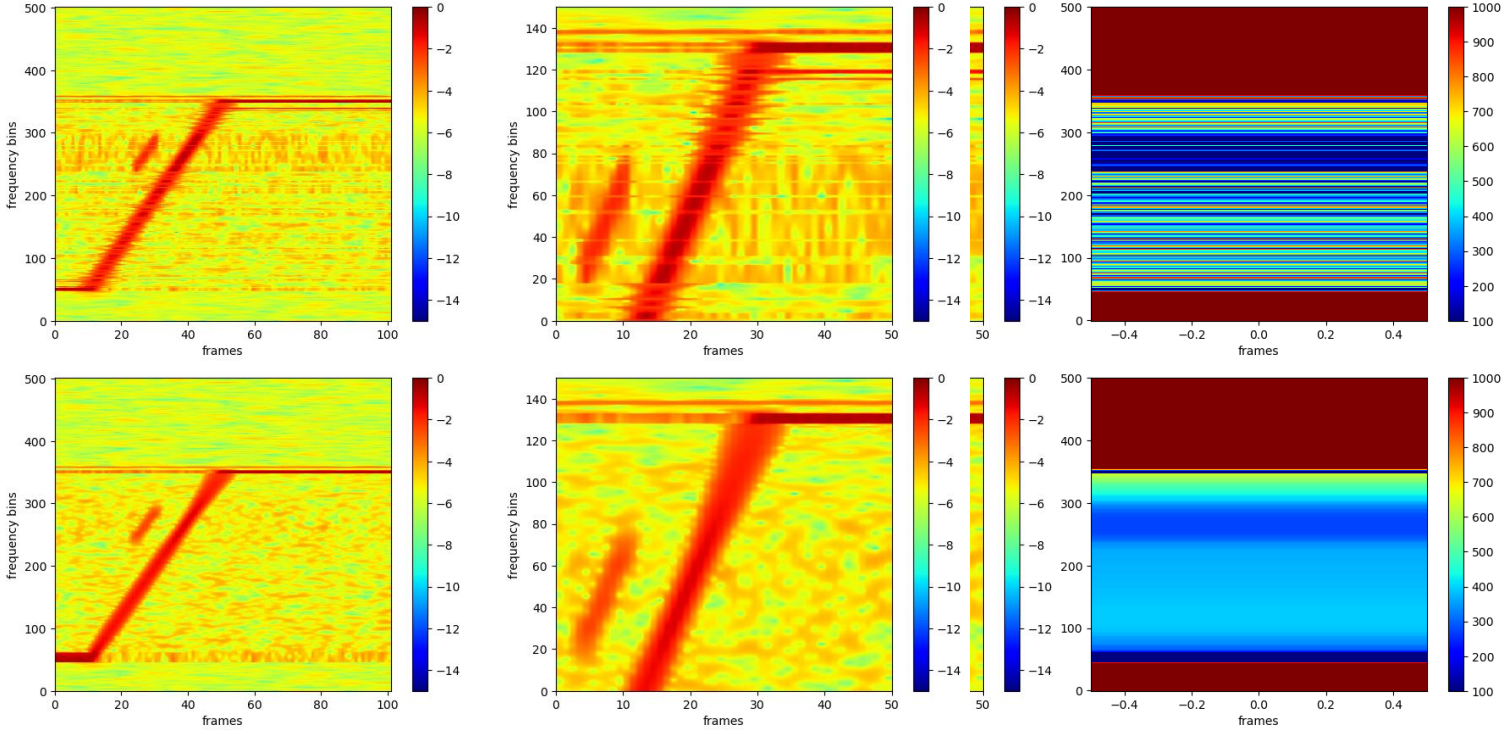
Distribution of window lengths

Without regularization

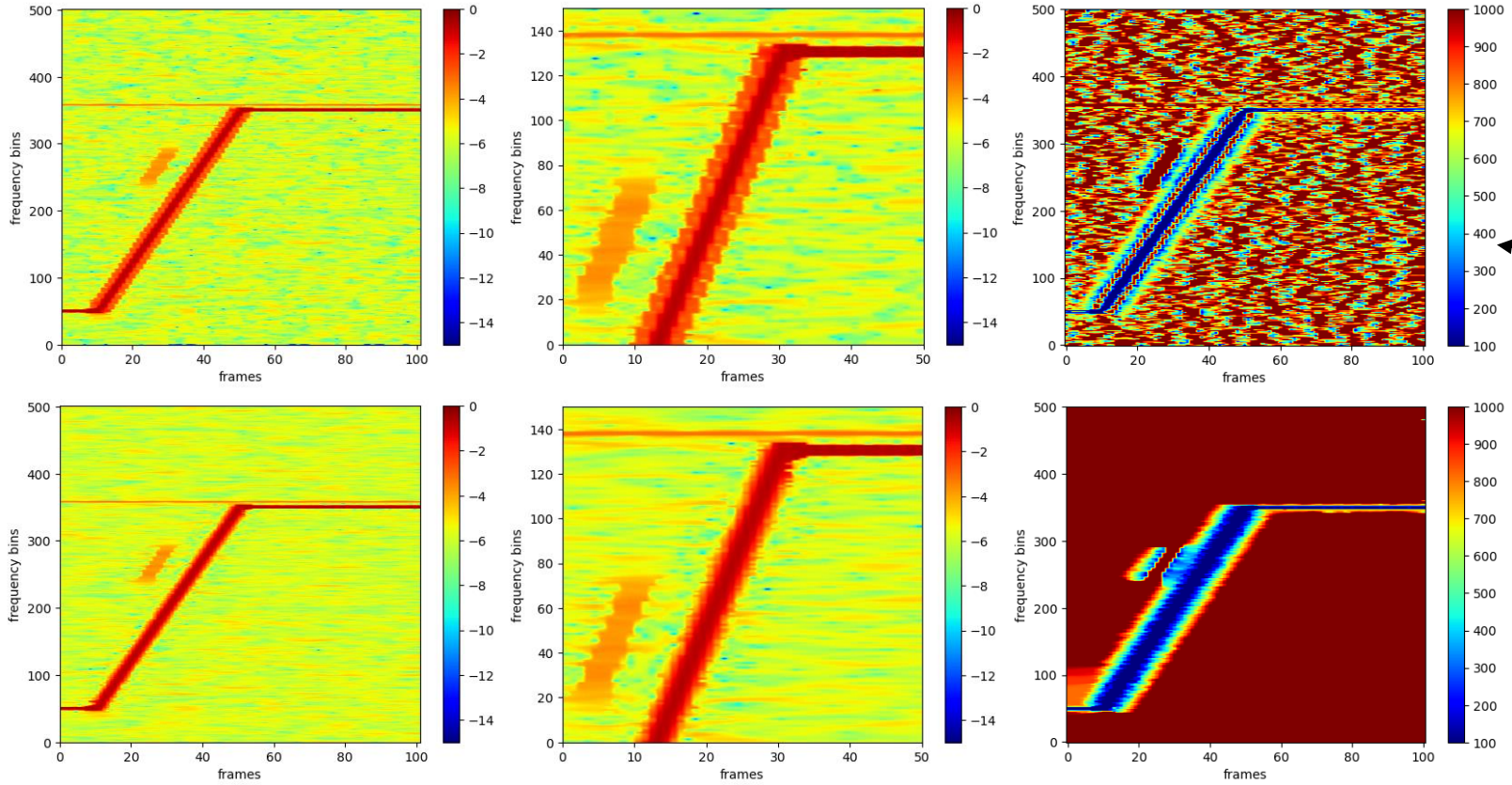
With regularization

SAFRAN

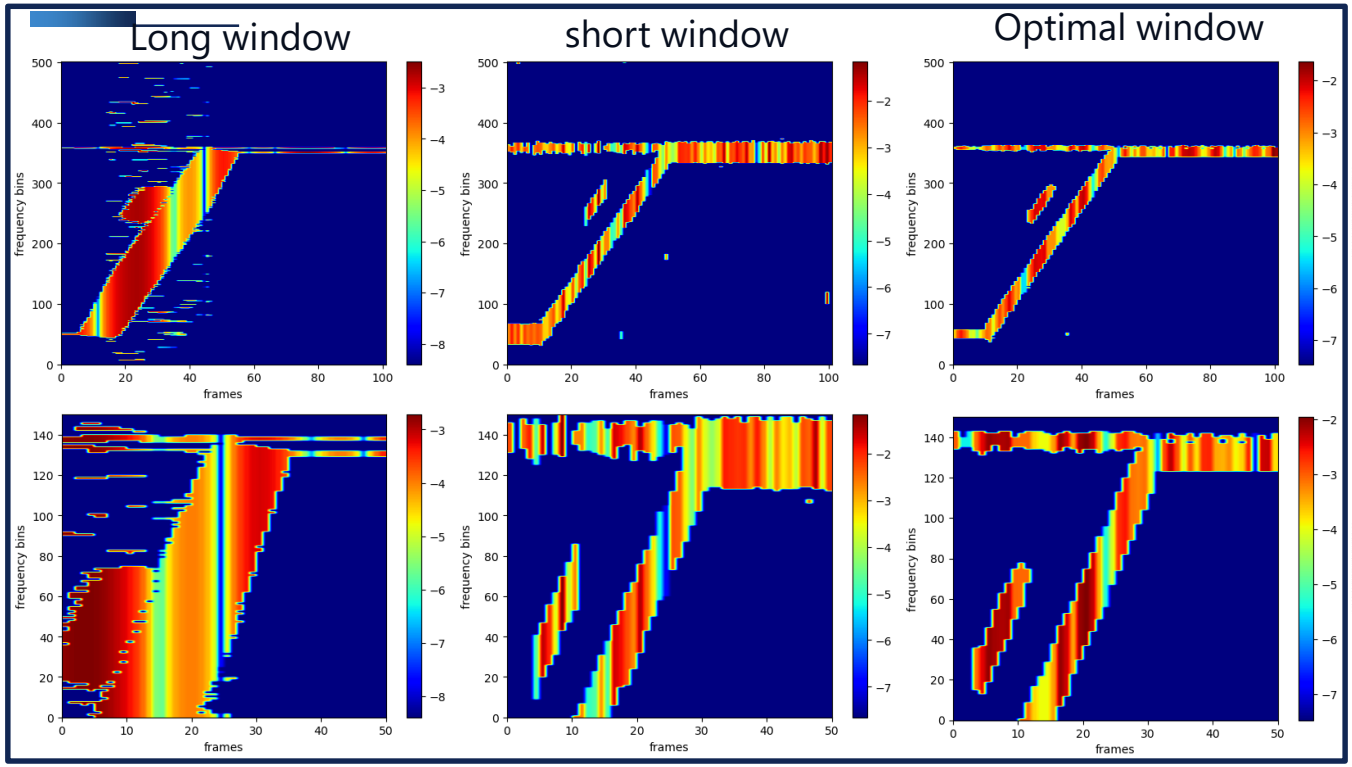
DSTFT with frequency-varying window length θ_f

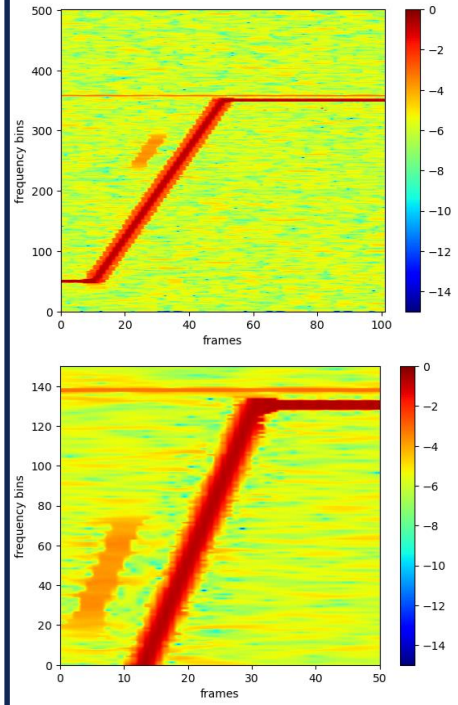


DSTFT with time- and frequency-varying window length



Comparison with synchrosqueezing





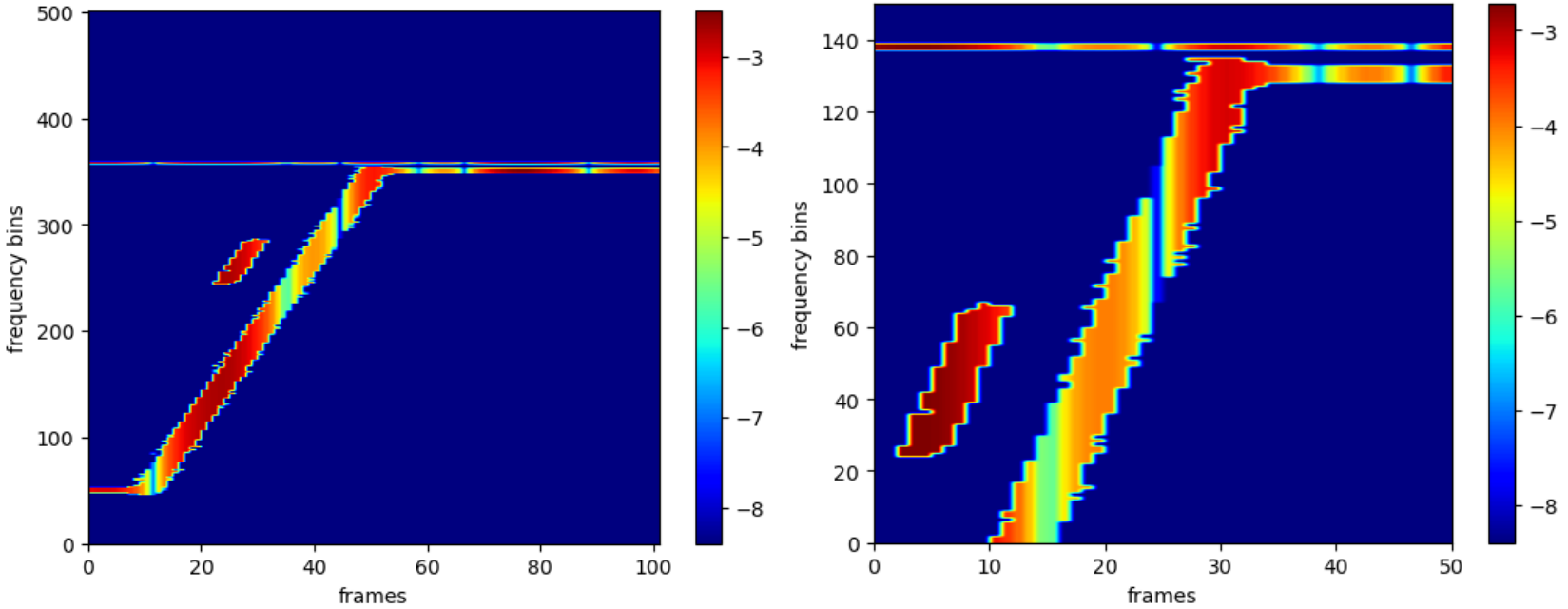
Shared (single) window + Synchrosqueezing

TF vaying window

- Regardless of the window length used in synchrosqueezing, it remains impractical to precisely locate all components in both time and frequency.

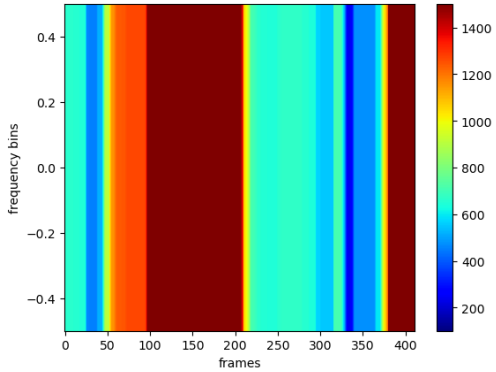
SAFRAN

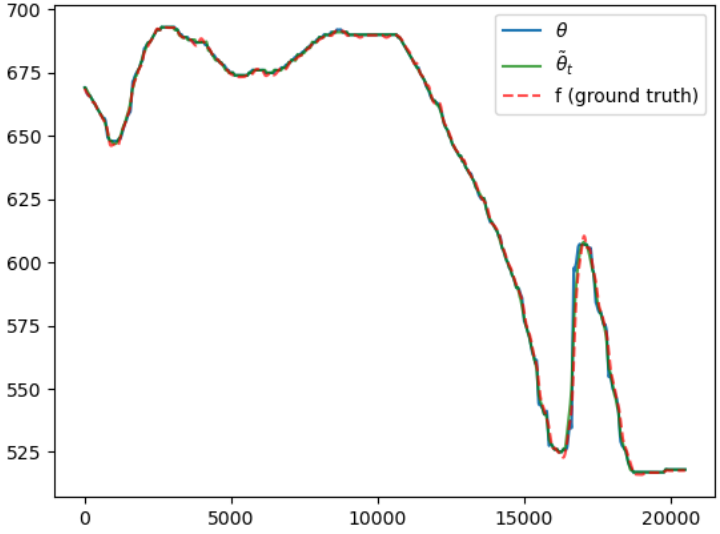
DSTFT wih time and frequency varying window+ synchrosqueezing

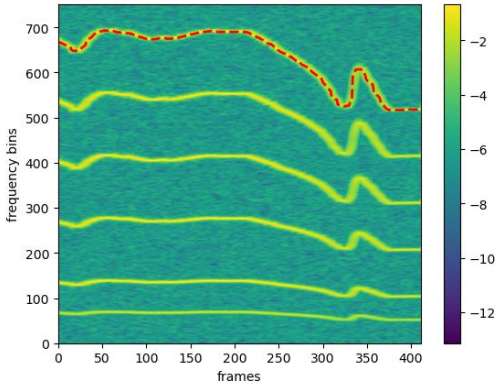


- Complementary methods: DSTFT and synchrosqueezing can be combined

Frequency tracking

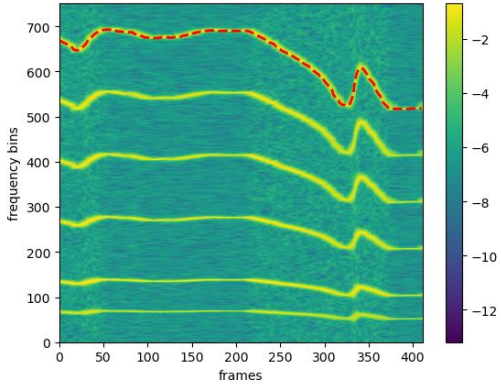






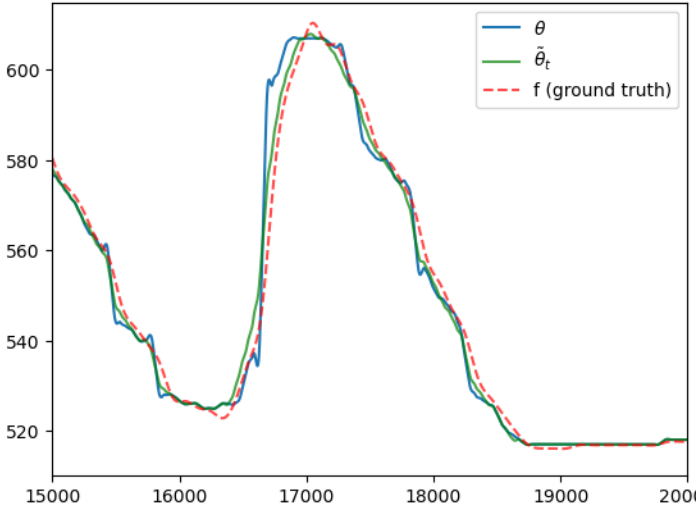
single window

$MSE(S_\theta) = 7.1$



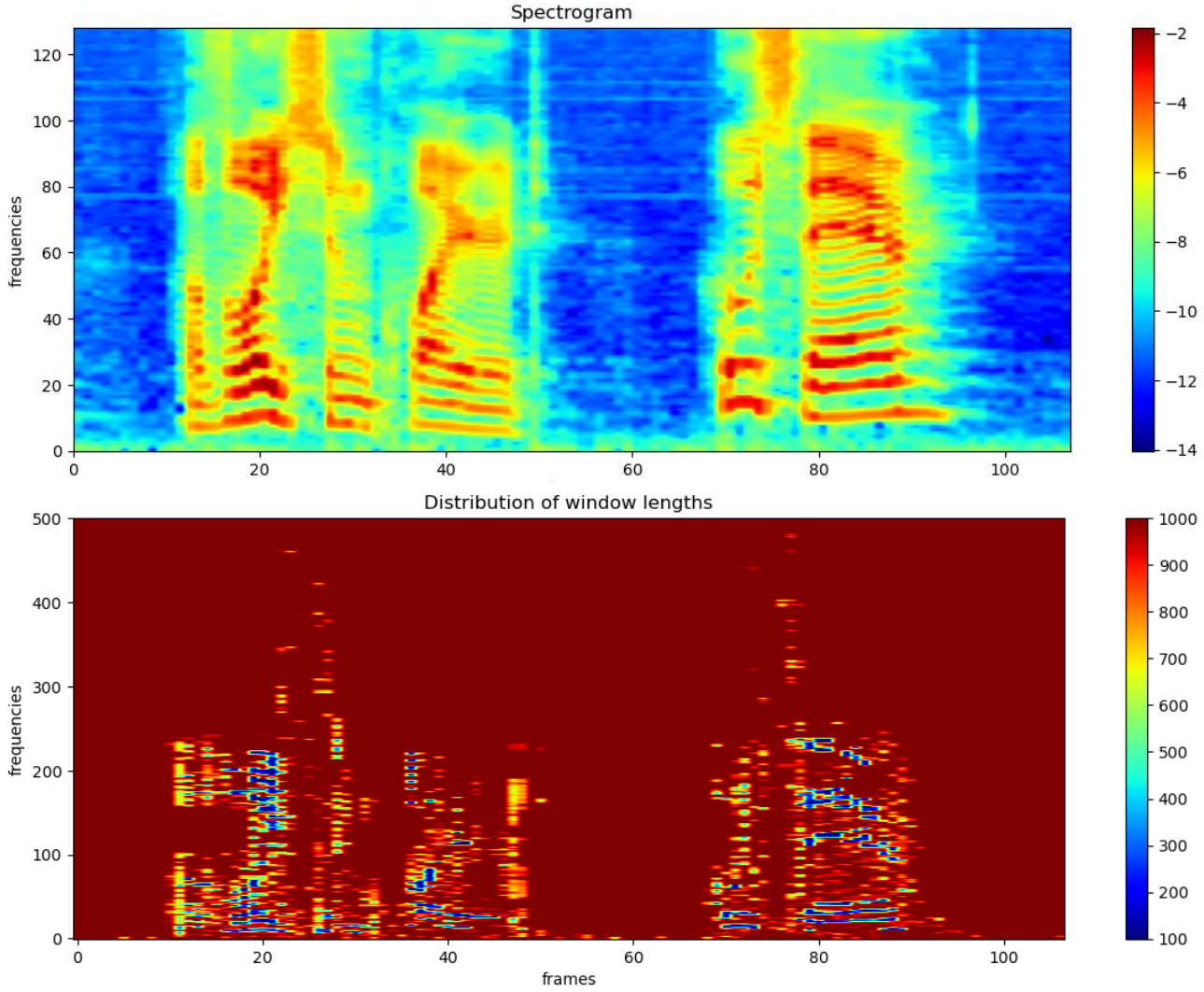
time-varying window

$MSE(S_{\theta_t}) = 2.9$

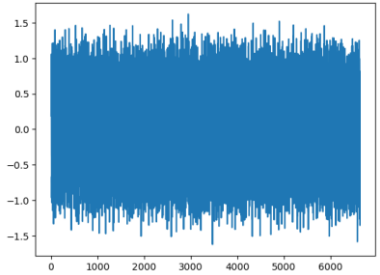


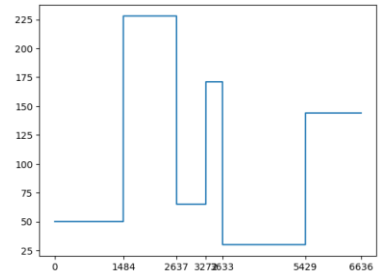
Speech signals

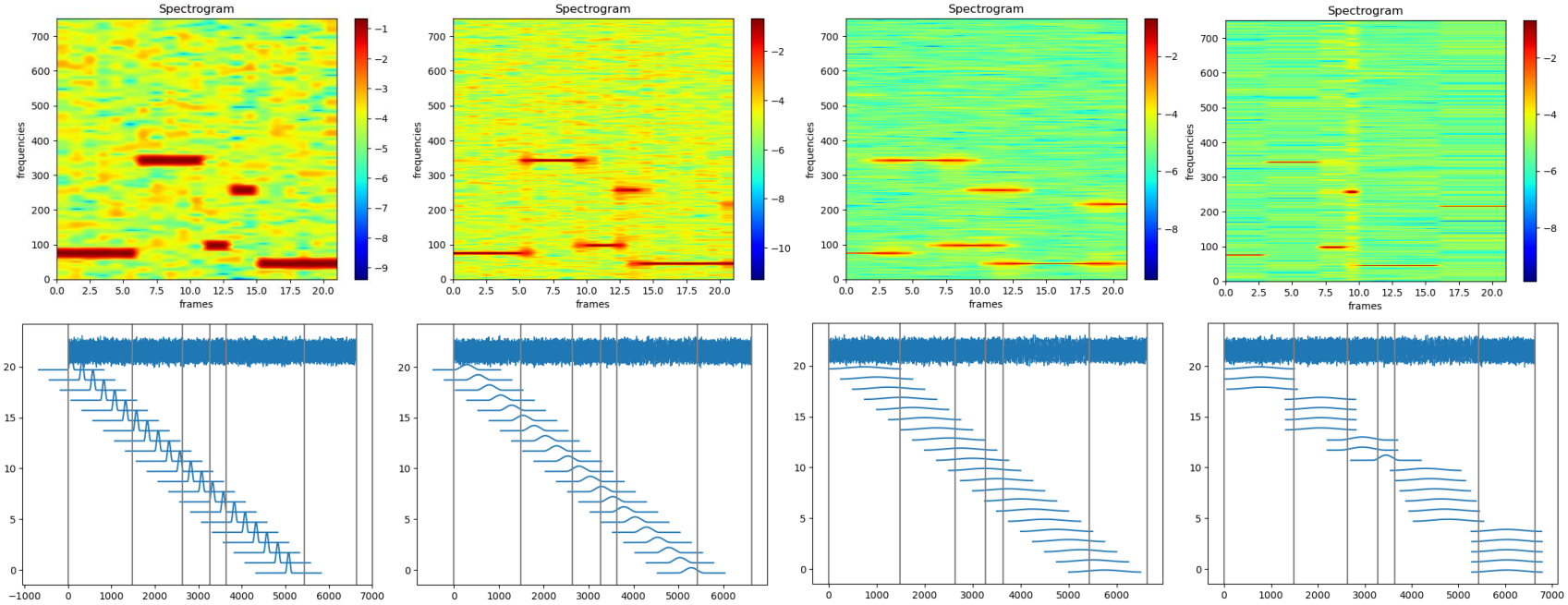
- DSTFT on speech signals
- Minimizing the entropy of the **mel-spectrogram**
- Enhanced resolution



Time-varying window and hop length







Part 4

Optimization for task performance

SAFRAN

Experiment : joint optimisation with a convolutional neural network

> Spoken digit classification task : Free Spoken Digit Dataset

> Objective is to find the optimal window length and network weights for the whole dataset.

> Optimization by minimizing the cross-entropy between the network prediction and the ground truth:

$$\mathcal{L}^{\theta,\omega}(\hat{y}_c, \hat{y}_{gt}) = -\frac{1}{N}\sum_{n=1}^{N} \log \frac{e^{\hat{y}_{gt}^n}}{\sum_{c=1}^{C} e^{\hat{y}_c^n}}$$

> θ: window length

> ω : network weights

> C: number of classes

> N: number of samples

> $\hat{y}_c^n$: network prediction of th n th sample associated with the class c

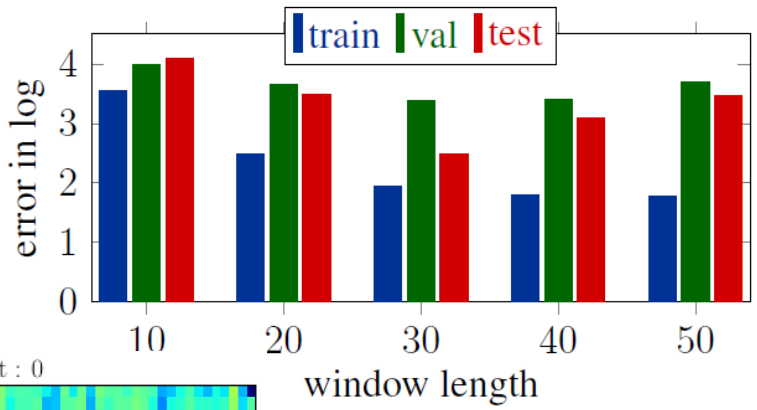
> $\hat{y}_{gt}^n$: network prediction of th n th sample for the clas corresponding to the ground truth

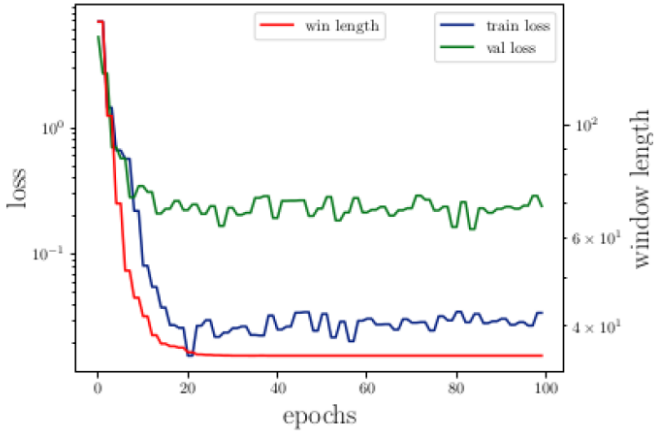
Experiment : joint optimisation with a convolutional neural network

> Losses decrease by jointly optimizing the window length and the weights of the neural network.

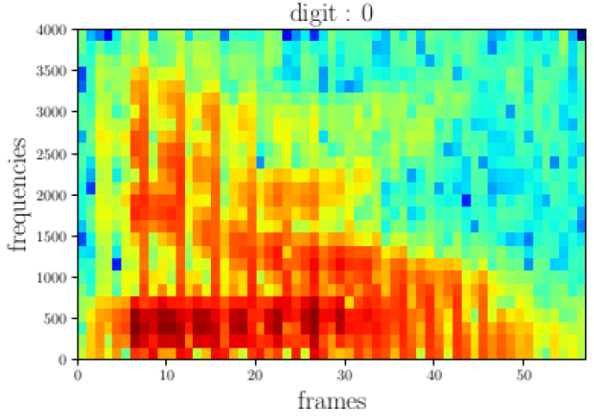
> During training, the time resolution continuous parameter θ converged to an optimal value.

Training, validation and testing losses





Training loss, validation loss and window length per epoch



Spectrogram of a sample obtained at the end of the training

SAFRAN

Part 5

Conclusion

SAFRAN

Conclusion

> STFT is now **adaptive** and **differentiable.**

> Allowing to adapt to the **time-varying spectral structure** of the signal.

> Can be of interest for **visualization** but also to **neural networks** and any **STFT-based signal processing algorithm**.

> Optimization of the parameters according to a **representation** or **task criteria**.

> extensions are also proposed:

> Fields of application: signal processing, speech recognition, audio, vibration diagnostics …

References and code

Leiber, M., Barrau, A., Marnissi, Y. and Abboud, D. (2022). **A differentiable short-time Fourier transform with respect to the window length**. In 2022 *IEEE European Signal Processing Conference (EUSIPCO).*

Leiber, M., Marnissi, Y., Barrau, A., and El Badaoui, M. (2023). **Differentiable adaptive short-time Fourier transform with respect to the window length**. In 2023 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*

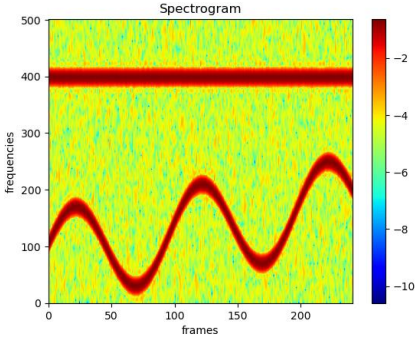
Leiber, M., Marnissi, Y., Barrau, A., and El Badaoui, M. (2023). **Differentiable short-time Fourier transform with respect to the hop length**. In 2023 *IEEE Statistical Signal Processing (SSP).*

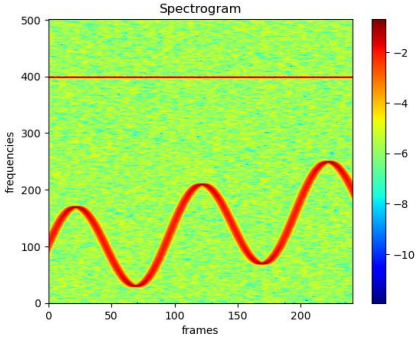
https://github.com/maxime-leiber/dstft

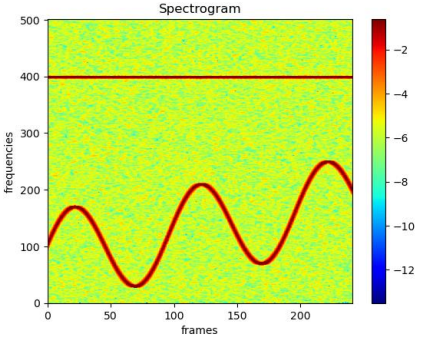
SAFRAN

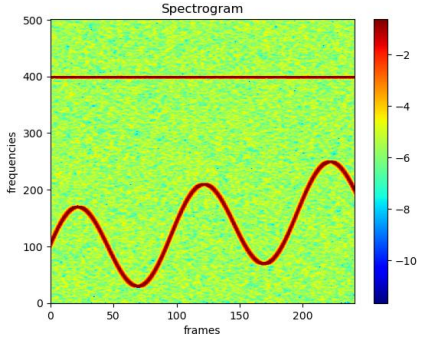
POWERED BY TRUST

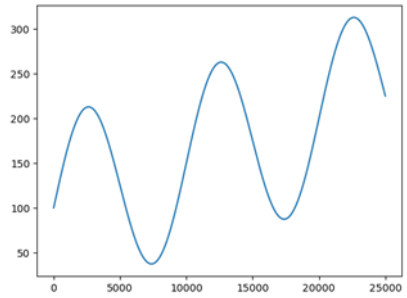
Experiments

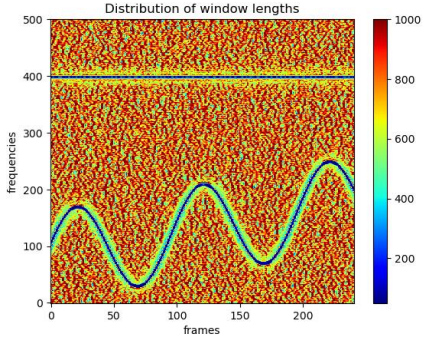


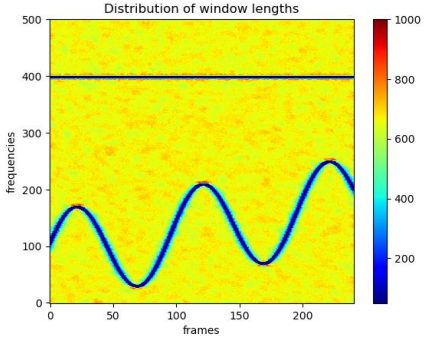












Time-varying window and hop lengths