SIPL SIGNAL & IMAGE PROCESSING LAB
Department of Electrical Engineering, Technion

Lens Motor Noise Reduction for Digital Cameras

Students: Avihay Barazany
Royi Levy
Supervisor: Kuti Avargel
In Association with:
Zoran, Haifa

ZORAN

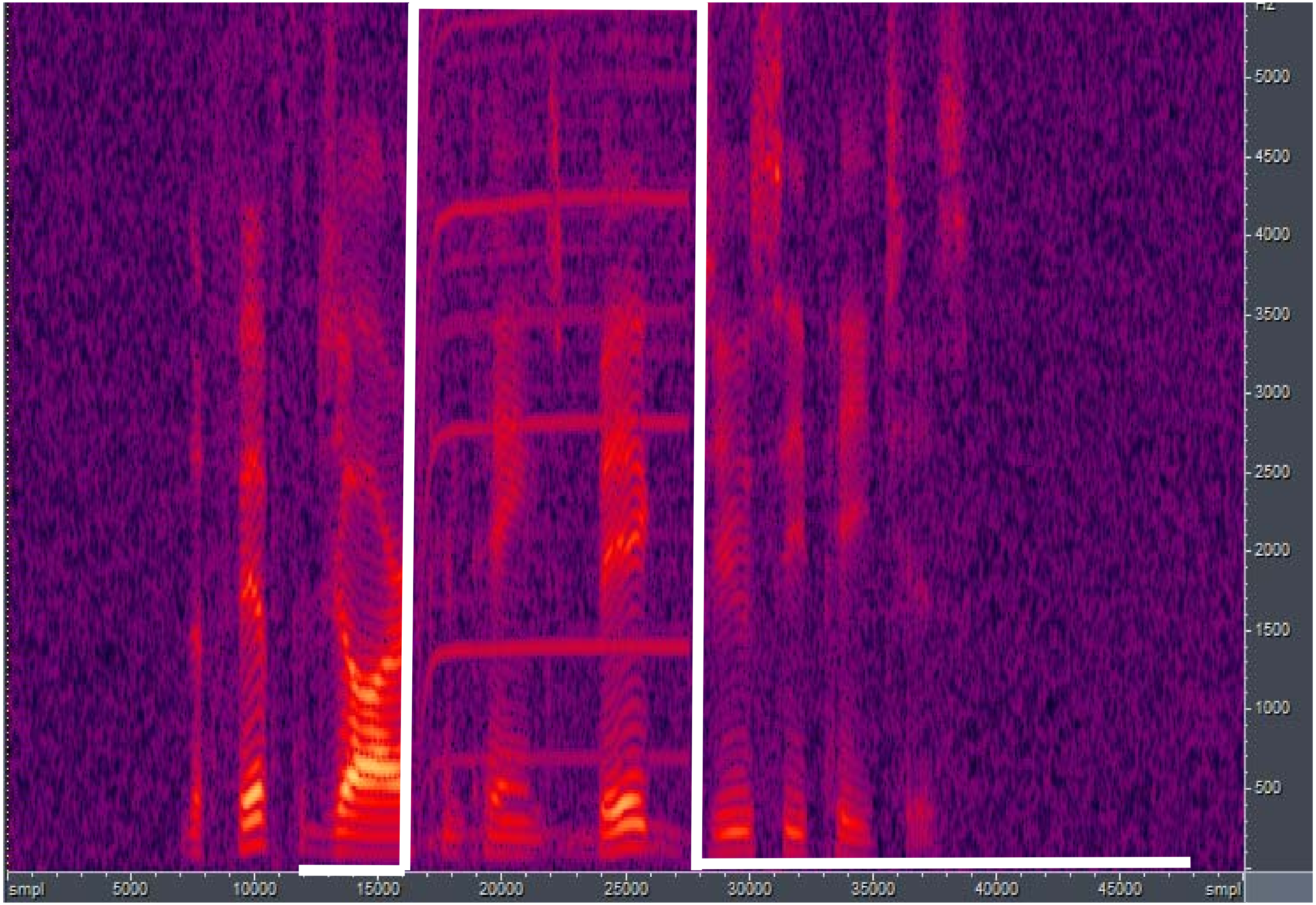
Spring 2007

Introduction

- **Digital still cameras** are widely used for video and audio recordings .
- When activating the **zoom lens-motor** during these recordings, the noise generated by the motor may be recorded by the camera's microphone.
- This noise may be extremely annoying and significantly **degrade the perceived quality** and intelligibility of the desired signal.

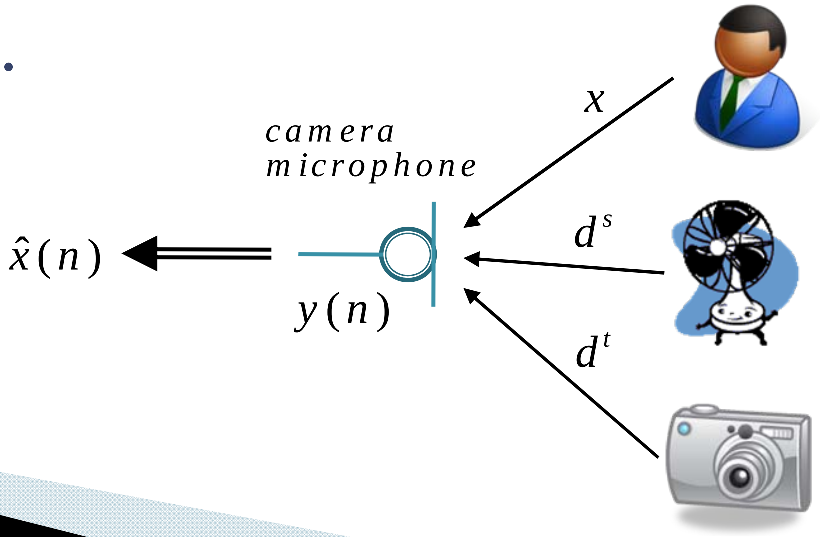
Introduction – cont.

Speech Signal Spectrum



Problem Formulation

- Let $x(n), d^s(n), d^t(n)$ denote the speech signal, background stationary noise, and zoom motor (non-stationary) noise, respectively.
- Let $y(n) = x(n) + d^s(n) + d^t(n)$ be the microphone signal.
- **Main goal:** to derive an estimator $\hat{x}(n)$ for the clean speech signal.



Possible Solutions

- ~~To solve this problem, many digital-cameras manufacturers disable the option of activating the lens motor during audio recordings.~~

- ~~**Adaptive solution** – Add a reference microphone and implement an **adaptive algorithm** for cancelling the motor noise in real-time.~~

- **Spectral enhancement** – Using spectral enhancement techniques for estimating the motor noise **spectrum** and enhancing the speech signal.

Spectral Enhancement Techniques

- The spectral enhancement approach is operated on the time-frequency domain.
- Let the observed signal be: $y(n) = x(n) + d(n)$
- The goal is to estimate the spectral coefficient of the speech signal.
- Let X_{lk} be the short time Fourier transform (STFT) of $x(n)$, i.e.,

$$X_{lk} = \sum_{m} w(lL - m)x(m)e^{-j\frac{2\pi}{N}km}$$

Spectral Enhancement Techniques – cont.

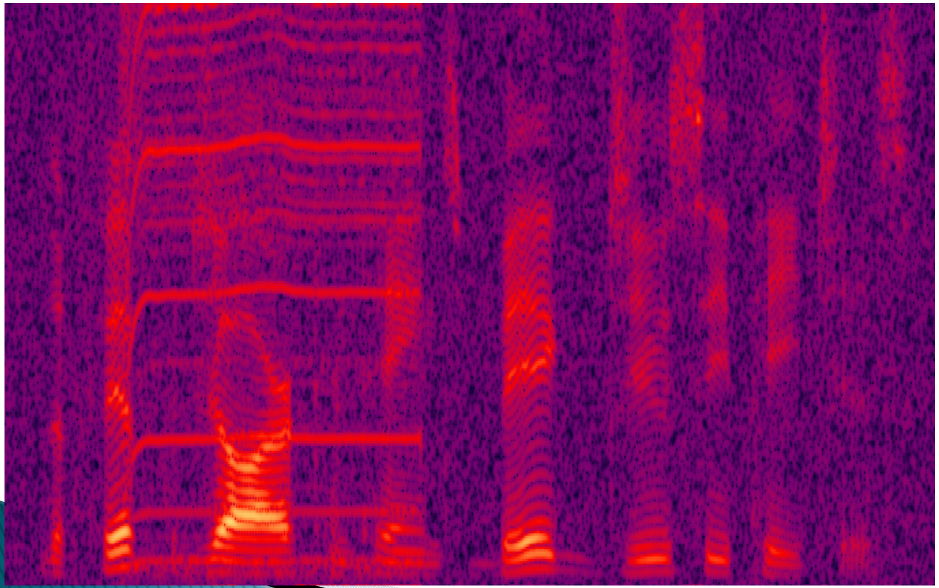
- The desired estimate of $\hat{X}_{lk}$ is : $\hat{X}_{lk} = G_{lk} \cdot Y_{lk}$ where the gain function G_{lk} is achieved by minimizing a cost-function: $\underset{G_{lk}}{\arg\min}\, E\left\{ d\left(X_{lk}, \hat{X}_{lk} \right) \right\}$
- There are different ways to measure the distortion function. The commonly used distortion functions are: $d\left(X_{lk}, \hat{X}_{lk} \right) = \left(\left| X_{lk} \right|^2 - \left| \hat{X}_{lk} \right|^2 \right)^2$ or

$$d\left(X_{lk}, \hat{X}_{lk} \right) = \left(\log\left| X_{lk} \right| - \log\left| \hat{X}_{lk} \right| \right)^2$$

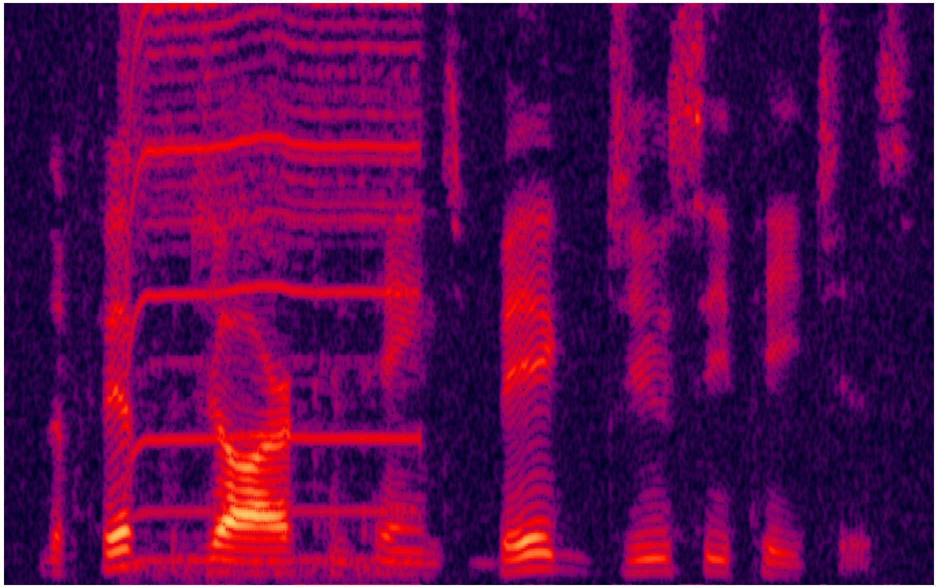
Spectral Enhancement Techniques – cont.

- The disadvantage of the above mentioned algorithms, is their difficulty to handle with highly non-stationary noises.

Input Signal



OMLSA Only



Proposed Algorithm

- The algorithm is based on paper:
  A. , Abramson, I. , Cohen, "Enhancement of Speech Signals Under Multiple Hypotheses using an Indicator for Transient Noise Presence", 2007

- Since the problem consists of 2 different types of noises, the definition of the observed signal is:

$$y(n) = x(n) + d^s(n) + d^t(n)$$

- And $X_{lk}, Y_{lk}, D_{lk}^s, D_{lk}^t$ are the STFT of $x(n), y(n),$ $d^s(n), d^t(n)$ accordingly.

Proposed Algorithm – cont.

- Since the motor noise not always present, we define the following 4 hypothesis:

$$H_{1s}^{lk}: Y_{lk} = X_{lk} + D_{lk}^{s}$$

$$H_{1t}^{lk}: Y_{lk} = X_{lk} + D_{lk}^{s} + D_{lk}^{t}$$

$$H_{0s}^{lk}: Y_{lk} = D_{lk}^{s}$$

$$H_{0t}^{lk}: Y_{lk} = D_{lk}^{s} + D_{lk}^{t}$$

H_1^{lk} : speech is more dominant than noise.

H_0^{lk} : noise is more dominant than speech.

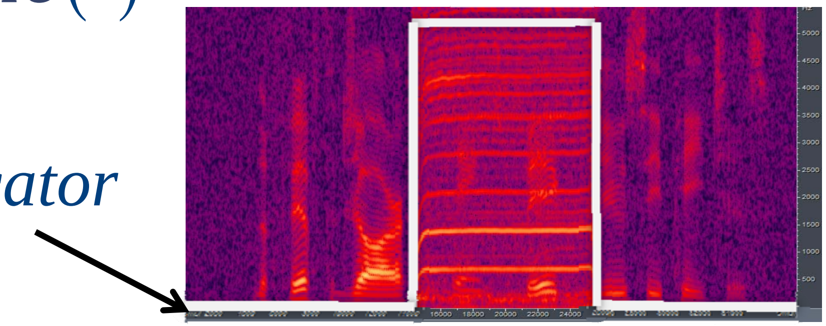
Proposed Algorithm – cont.

- Let $\eta_j^{lk}, j \in \{0,1\}$ denote the detector decision in the time-frequency bin (l,k):

$$\eta_0^{lk} - \textit{transient is a noise component}$$
$$\eta_1^{lk} - \textit{transient is a speech component}$$

- Let $C_{10,} C_{01}$ denote the cost of false-alarm / miss-detections, respectively.

- The algorithm assumes an indicator signal for the motor noise in the time frame (l).


Indicator

Estimation Criteria

- Let $A_{lk} = \left|X_{lk}\right|, \quad R_{lk} = \left|Y_{lk}\right|.$

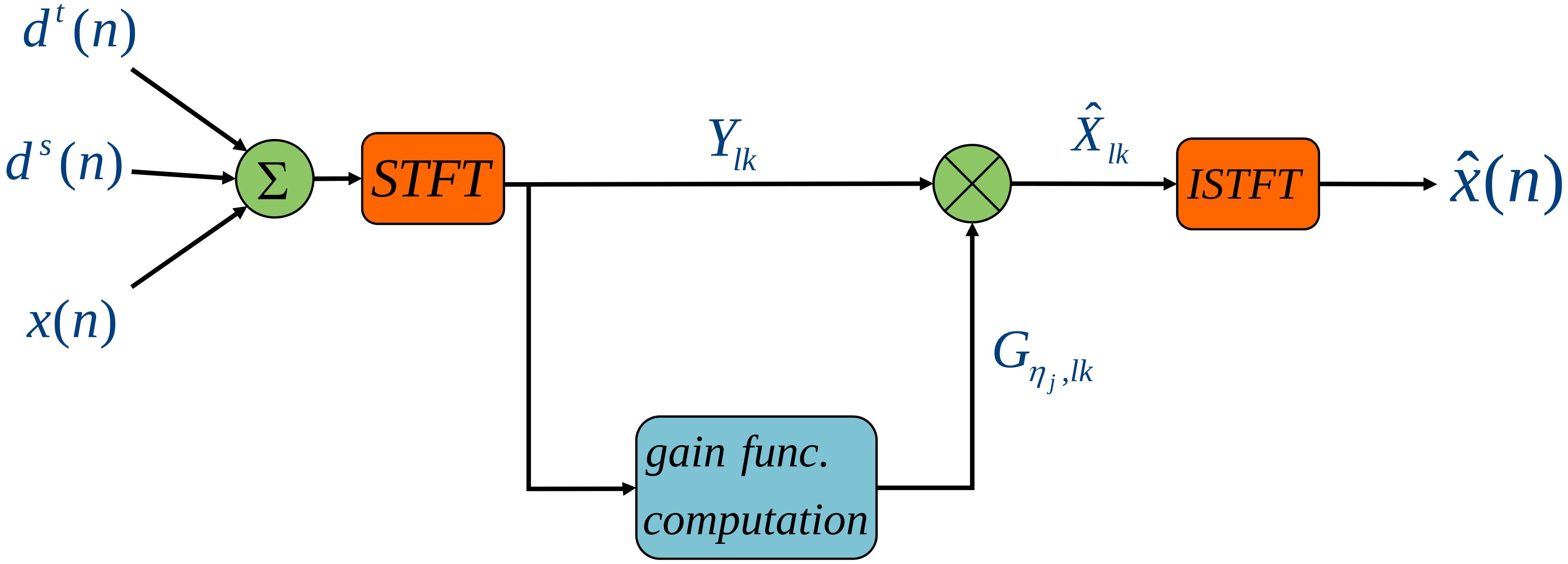
  The criterion for the estimation of the speech signal under the decision η_j^{lk} :

$$\hat{A}_{lk} = \arg\min_{\hat{A}} \Big\{ C_{1j}\, p\left(H_1^{lk} \mid \eta_j^{lk}, Y_{lk}\right) E\left[d\left(X_{lk}, \hat{A}\right) \mid Y_{lk}, H_1^{lk} \right]$$

$$+ C_{0j}\, p\left(H_0^{lk} \mid \eta_j^{lk}, Y_{lk}\right) d\left(G_{\min} R_{lk}, \hat{A}\right) \Big\}$$

  where $d(x, y) = \left(\log|x| - \log|y|\right)^2 .$

Block Scheme



Motor Noise Estimation

- The a-priori estimation for the motor noise is achieved using an average of early acquired recordings λ_0.
- The algorithm updates the initial estimation according to pre-determined regions.
  The result is the desired $\hat{\lambda}_t$:

$$\tilde{H}_0 : \hat{\lambda}_t(l,k) = \alpha\lambda_0(l,k) + (1-\alpha)\left\{\beta\hat{\lambda}_t(l-1,k) + (1-\beta)\left[\left|Y(l,k)\right|^2 - \hat{\lambda}_s(l,k)\right]\right\}$$

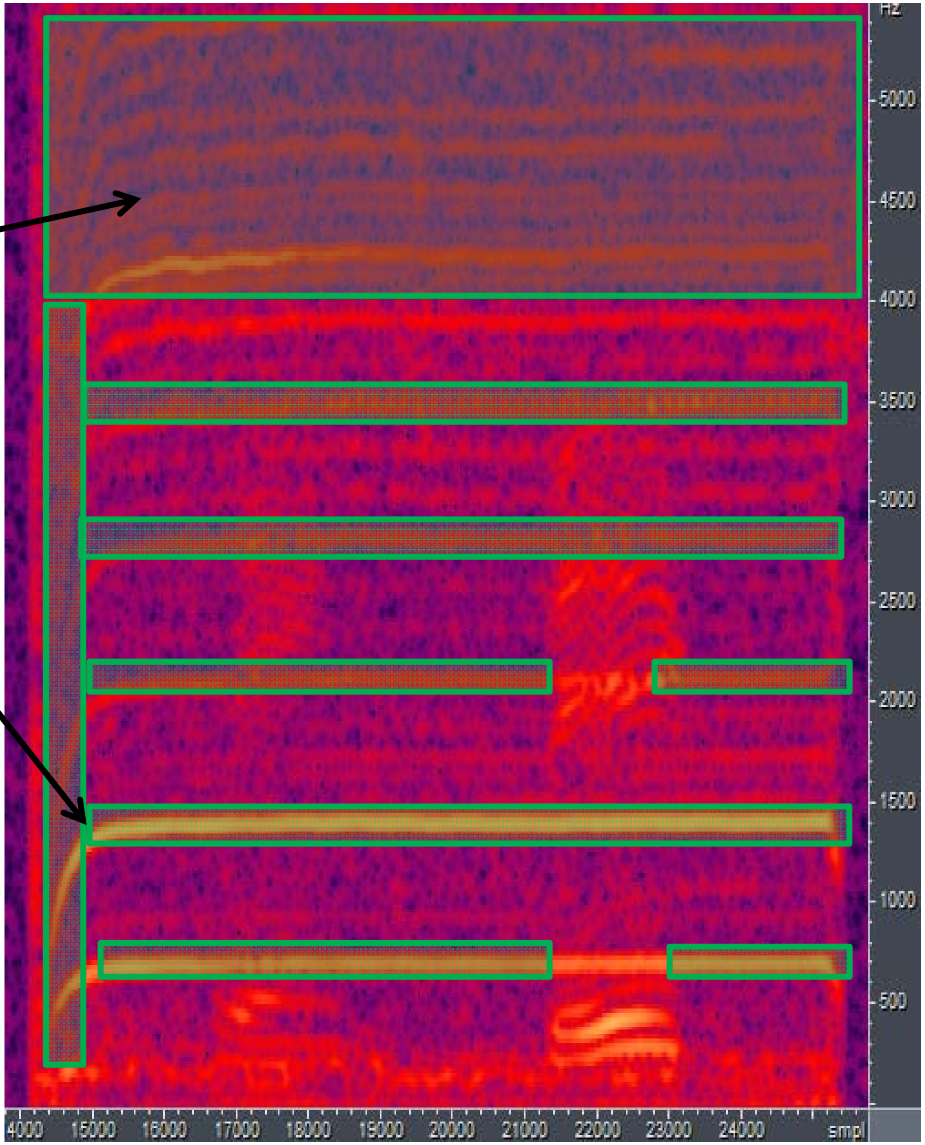
$$\tilde{H}_1 : \hat{\lambda}_t(l,k) = \alpha\lambda_0(l,k) + (1-\alpha)\hat{\lambda}_t(l-1,k)$$

- The noise is classified by the criteria:
  Motor noise level higher than speech level $\left(\tilde{H}_0\right)$.

Motor Noise Estimation – cont.

Region classification:

- Method of classification:
- Frequencies that are out of speech band [>4 KHz], are assumed to be in $\tilde{H}_0$.
- High amplitude harmonies in the motor noise estimation are classified as $\tilde{H}_0$ as well.
- High amplitude harmonies are determined by an empiric threshold.
- The rest of the spectrum is classified as $\tilde{H}_1$.

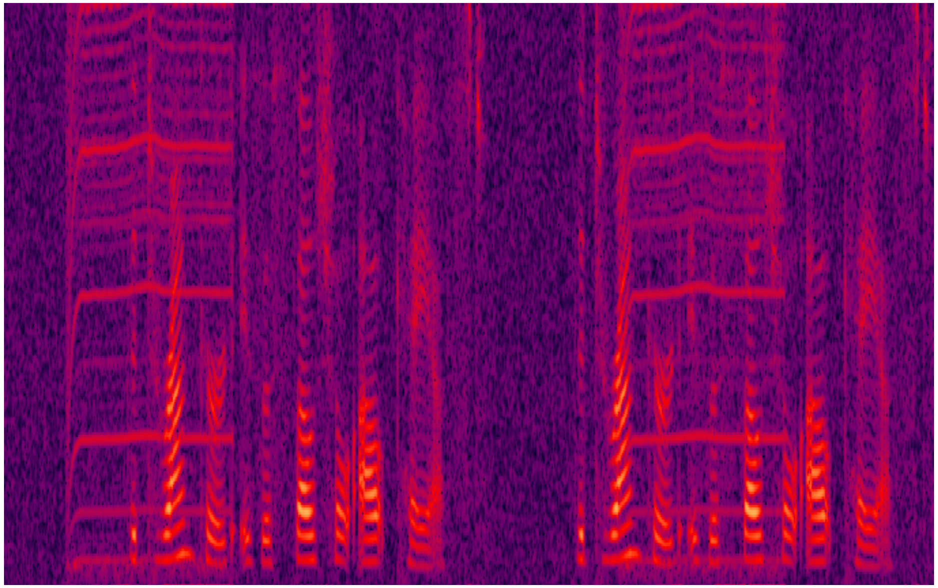


Experimental Results

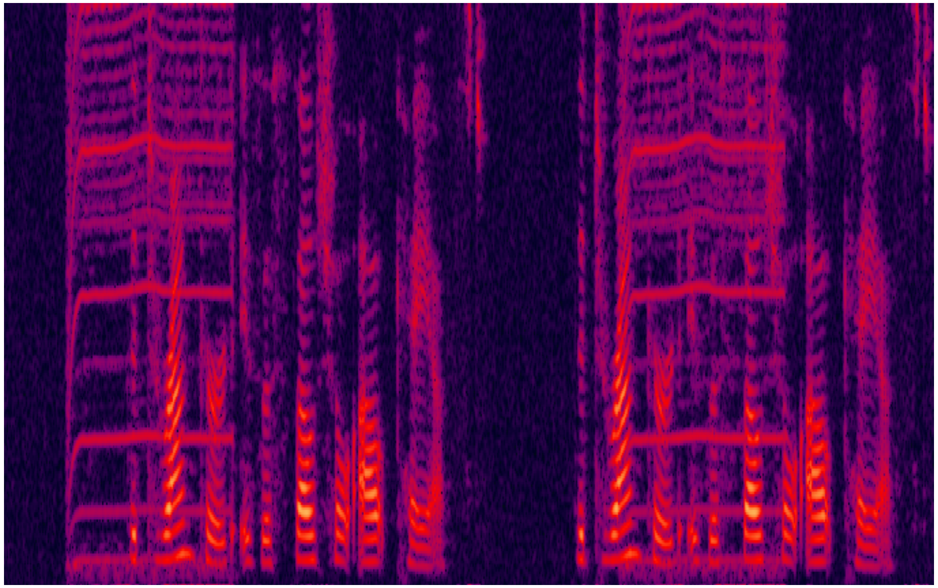
Parameters Setup:

- Several SNR’s of motor noise and speech were experimented.
- For each recording several G_f values were considered.
- Different parameter sets were tried out until the optimized ones were found.
- The performance of the proposed approach was compared to those of the conventional OMLSA.

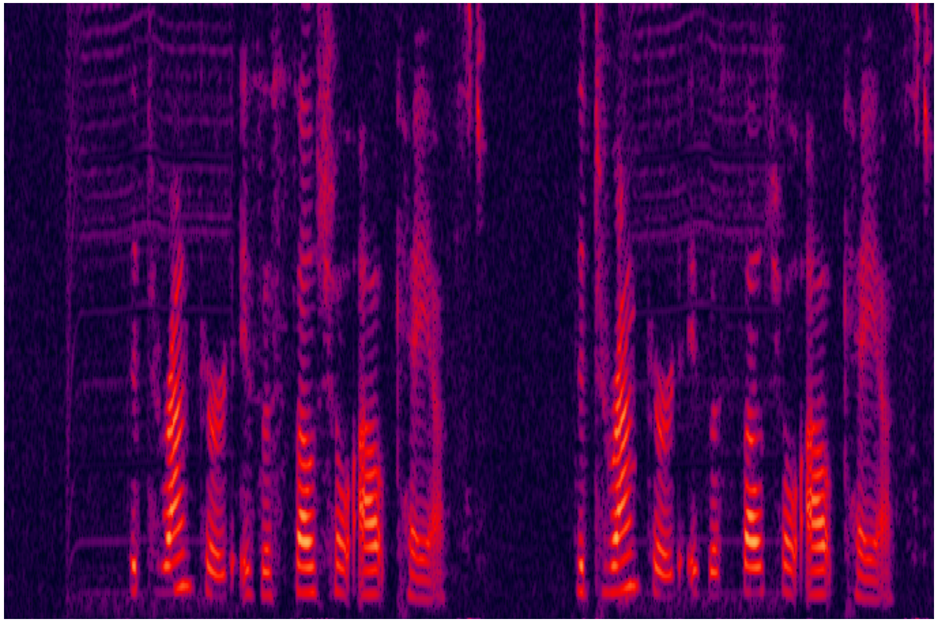
Full Zoom SNR=10dB, Female



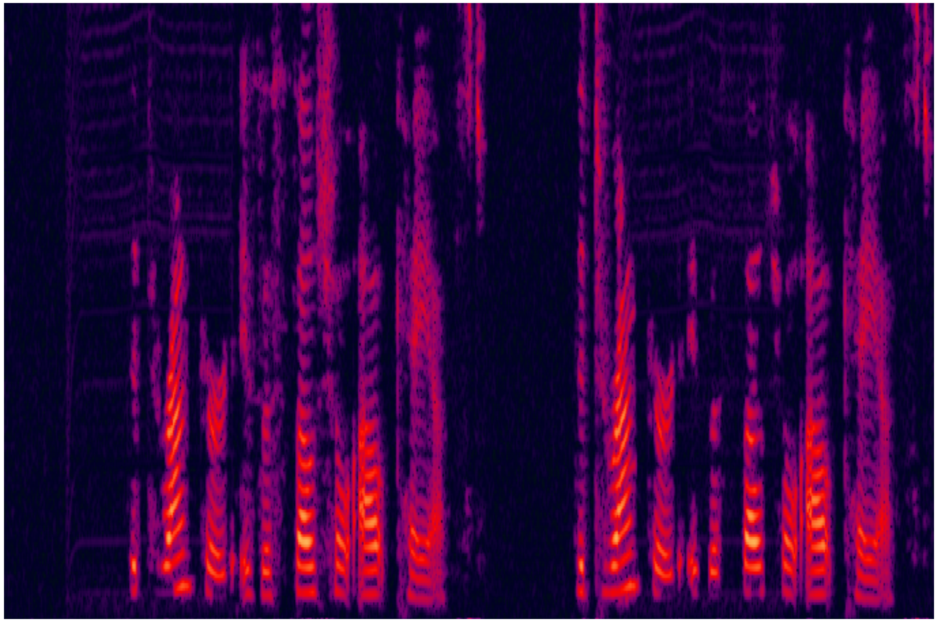
Input Signal



OMLSA Only

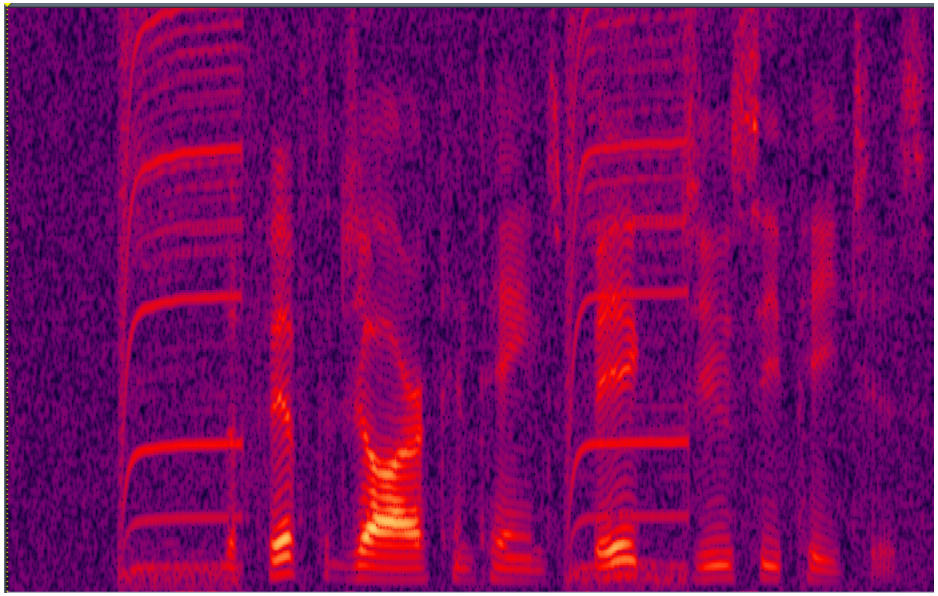


Gf=-15dB

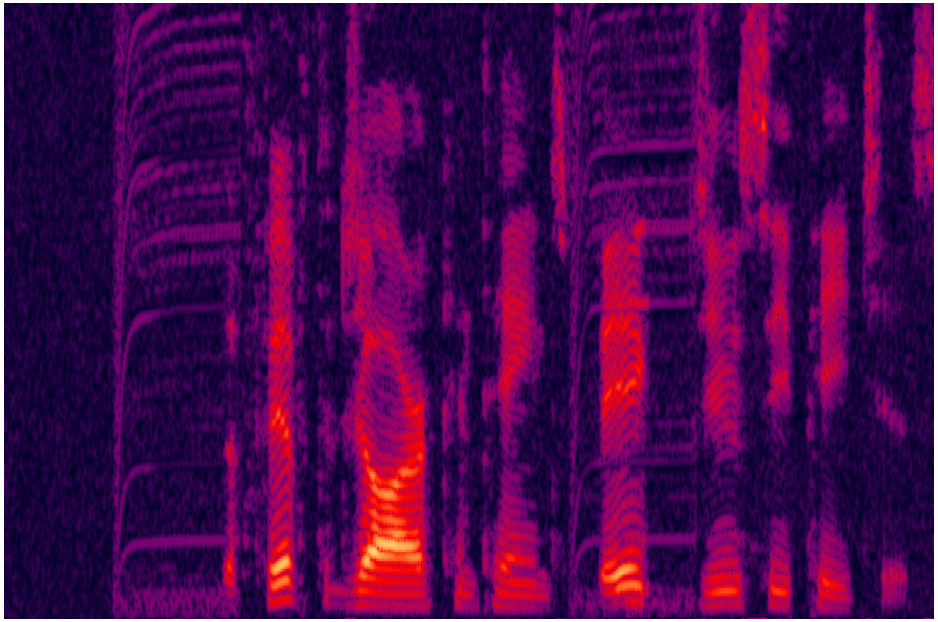


Gf=-25dB

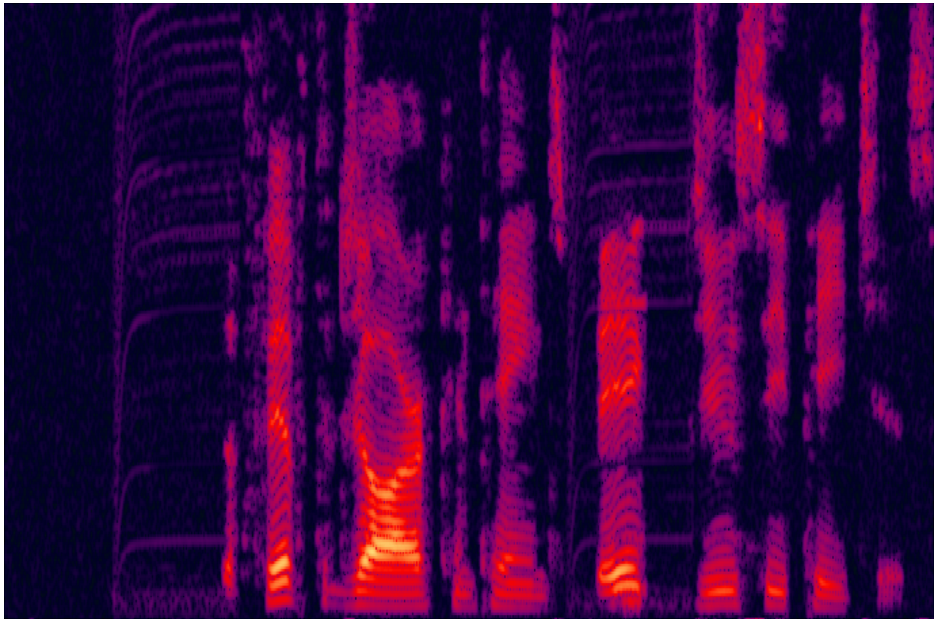
2 parts Zoom SNR=10dB, Male



Input Signal

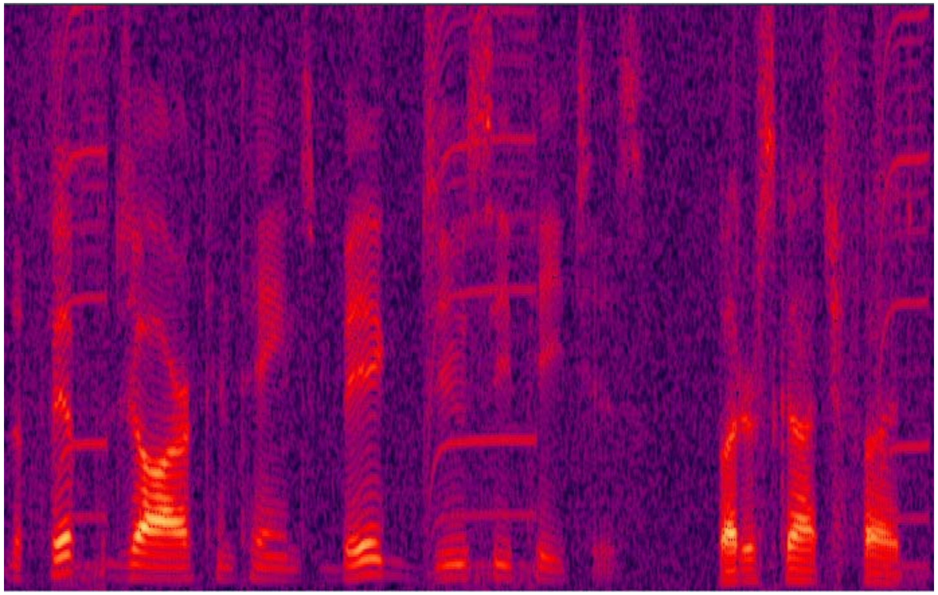


Gf=-15dB

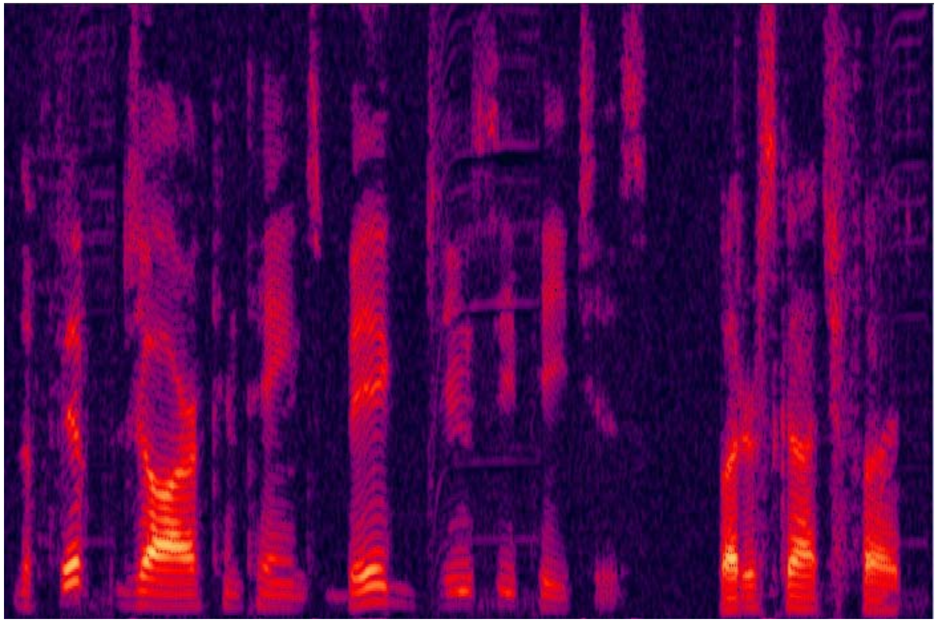


Gf=-25dB

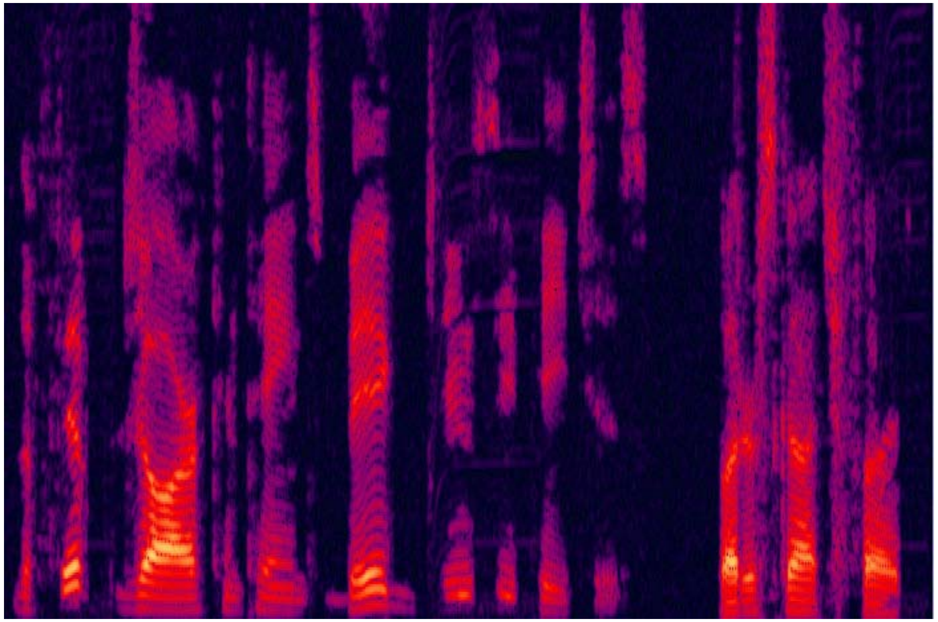
3 parts Zoom SNR=15dB, Male



Input Signal

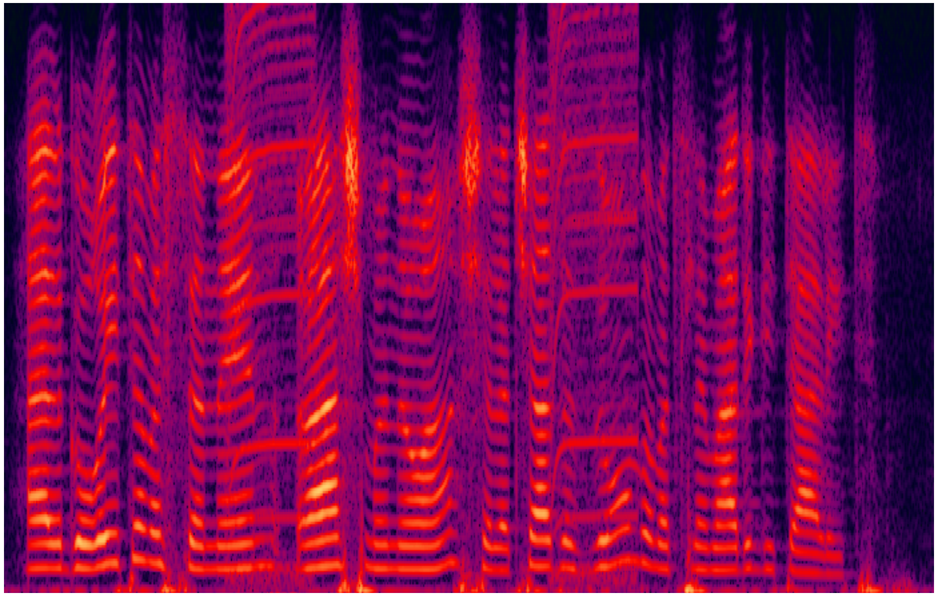


Gf=-15dB

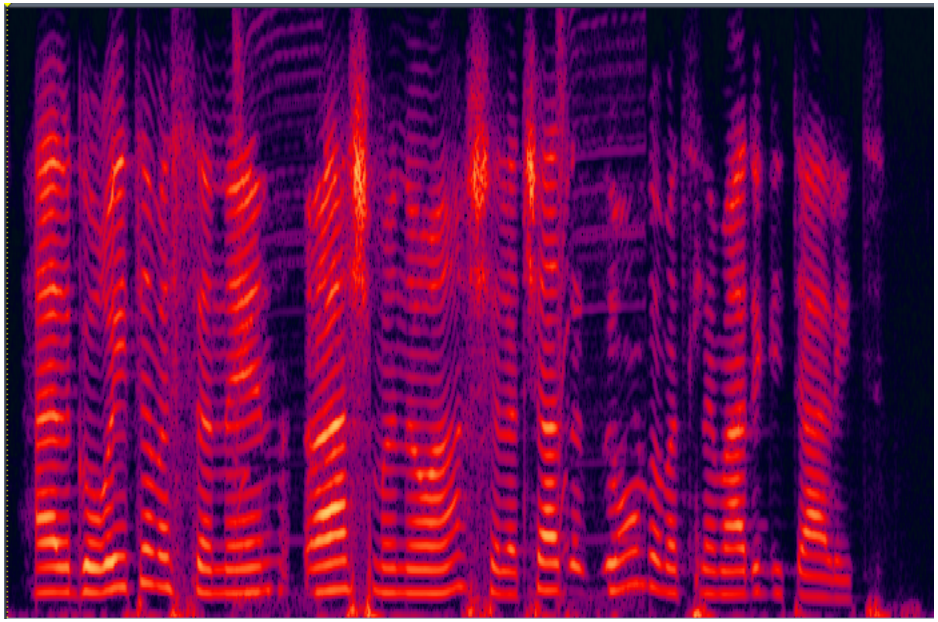


Gf=-25dB

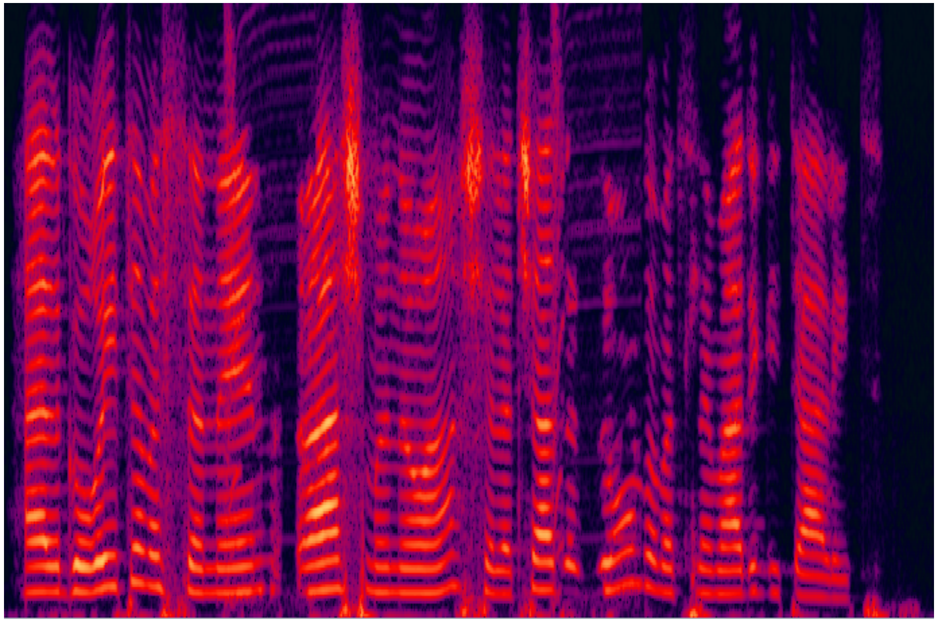
2 parts Zoom Real Recording, Female



Input Signal



Gf=-15dB



Gf=-20dB

Summary

- An algorithm for **suppressing lens motor noise** has been introduced.
- **An optimal estimator**, is derived, while assuming some indicator for the motor-noise presence in the time domain.
- A-priori motor noise spectrum estimate is acquired .
- **A substantial suppression** of the motor noise is achieved, **without degrading the perceived quality** of the desired signal.
- The proposed algorithm is **computationally efficient**.

Acknowledgments

- The Signal & Image processing lab for technical support during the entire work process.
- The Control & Robotics lab for assistance with assembling of the camera module together with an I/O control card.
- For all the guidance and academic support by Kuti Avargel.

References

- I. Cohen and B. Berdugo, “Speech Enhancement for Non-Stationary Noise Environments”, *Signal Processing*, Vol. 81, No. 11, pp. 2403-2418 , Nov. 2001.
- I. Cohen and B. Berdugo, “Noise estimation by minima controlled recursive averaging for robust speech enhancement”, *Signal* Processing, Vol. 9, Issue 1, pp. 12 – 15, Jan 2002.
- A. Abramson and I. Cohen," Enhancement of Speech Signals Under Multiple Hypotheses Using an Indicator for Transient Noise Presence " *Proc. 31th IEEE Internat.*
- A., Abramson, I. Cohen, “Simultaneous Detection and Estimation Approach for Speech Enhancement”, Audio, Speech, and Language Processing, IEEE Transactions on Vol. 15, Issue 8, pp. 2348 – 2359 , Nov. 2007.