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

Spring 2008

# Outline

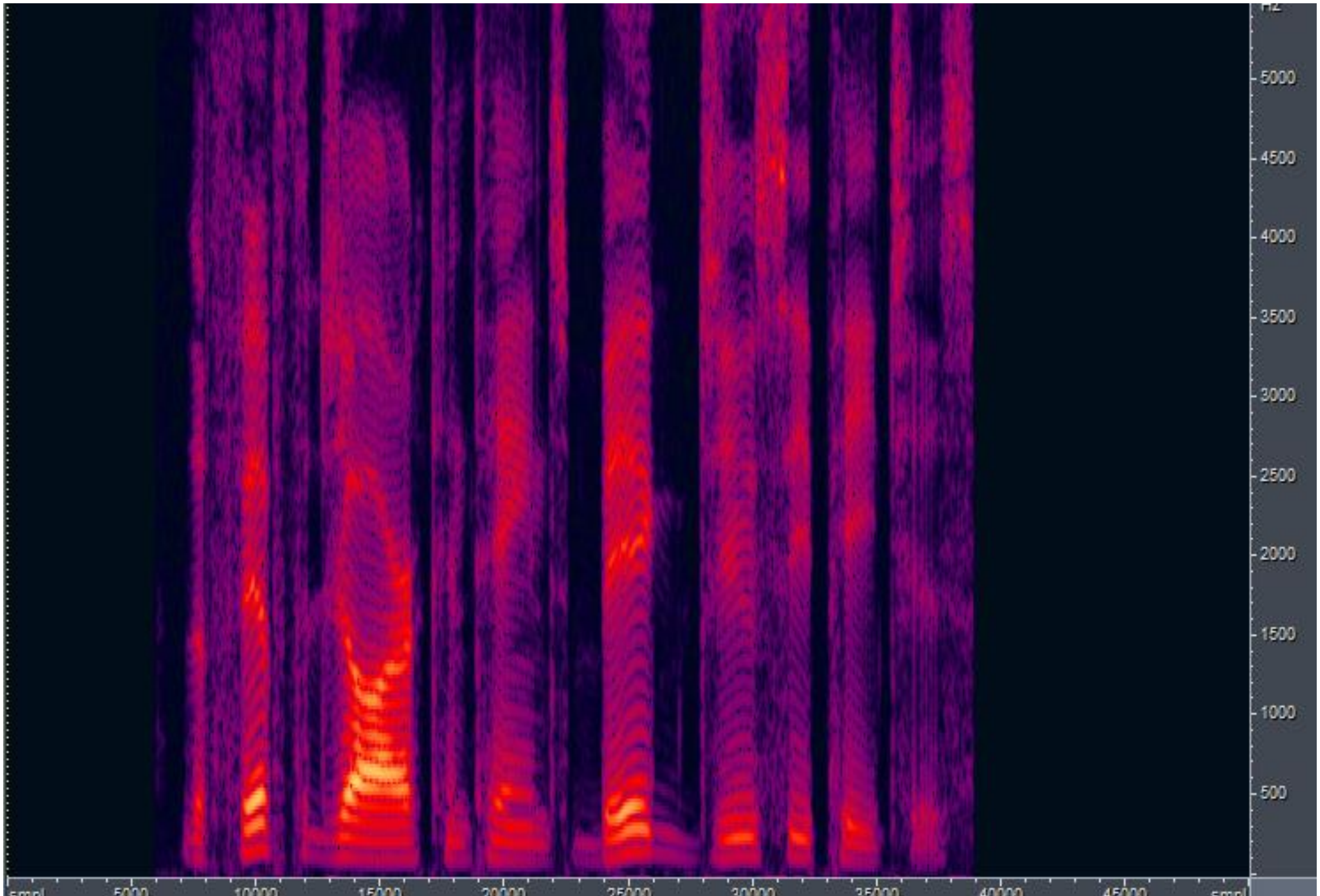
- Introduction
- Problem Formulation
- Possible Solutions
- Proposed Algorithm
- Experimental Results
- Conclusions

# 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.

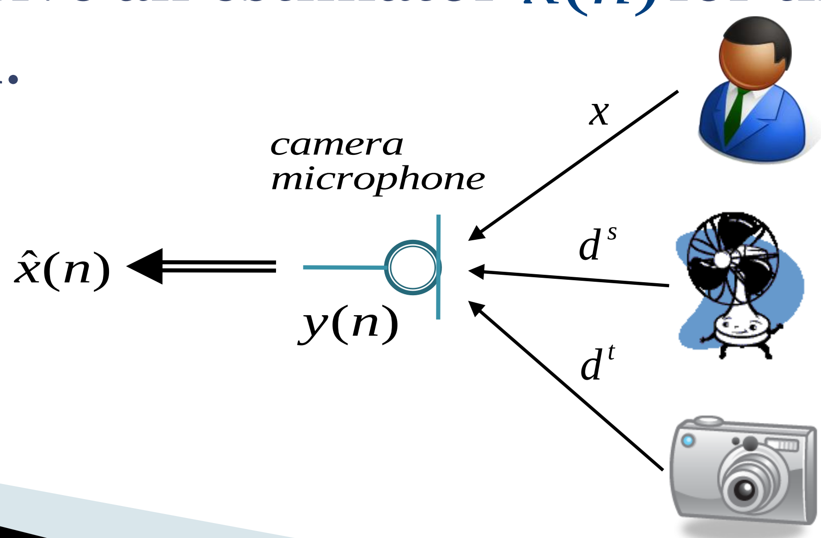
# Introduction – cont.

## Speech + Motor noise 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.

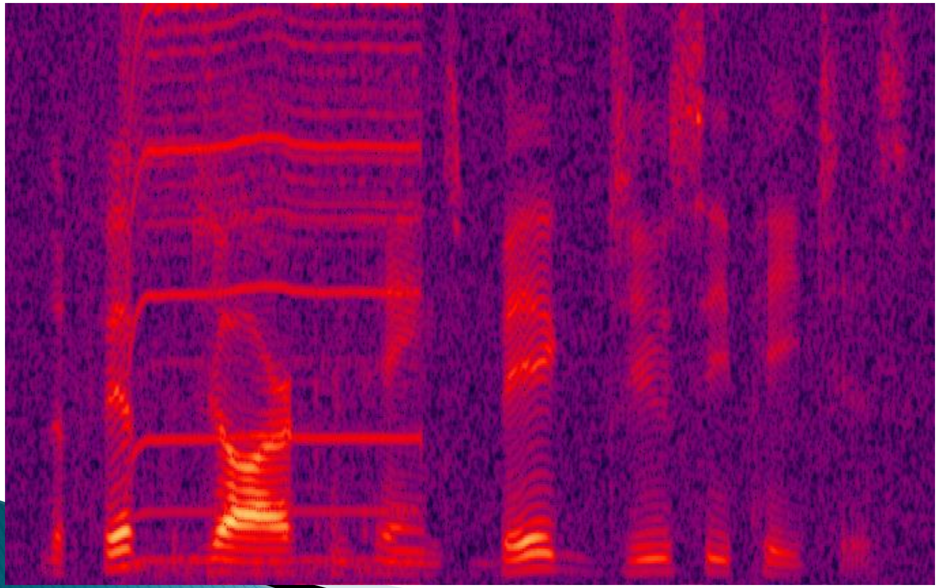
- The desired estimate of $\hat{X}_{lk}$ is : $\hat{X}_{lk} = G_{lk} \cdot Y_{lk}$ where the gain function $G_{lk}$ is achieved by minimizing a cost-function: $\arg\min\limits_{G_{lk}} E\left\{ d\left( X_{lk}, \hat{X}_{lk} \right) \right\}$
- There are different ways to measure the distortion function.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

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

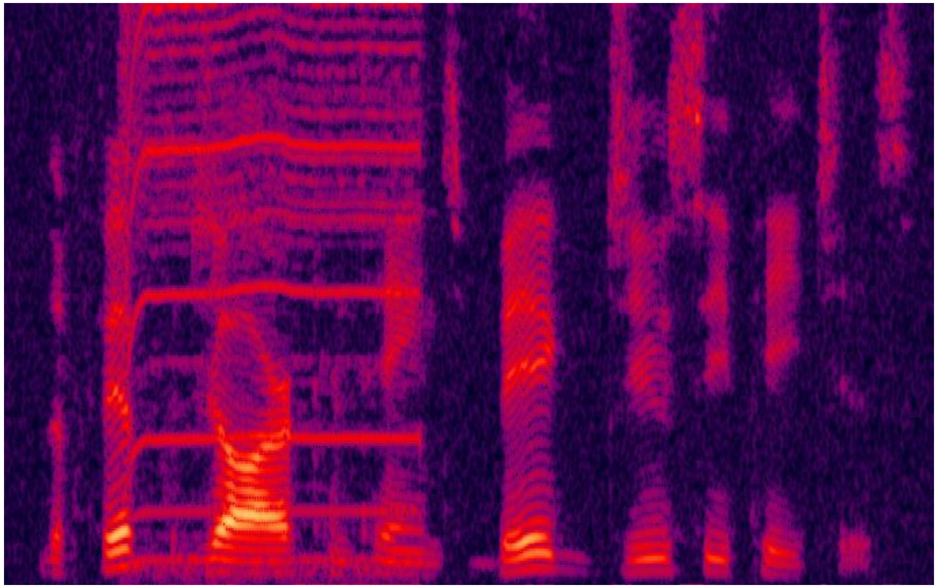
# Spectral Enhancement Techniques – cont.

- 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:

$$H_{1s}^{lk}: Y_{lk} = X_{lk} + D_{lk}^{s}$$
$$H_{1t}^{lk}: Y_{lk} = X_{lk} + D_{lk}^{s} + D_{lk}^{t}$$
$$H_{0s}^{lk}: Y_{lk} = D_{lk}^{s}$$
$$H_{0t}^{lk}: Y_{lk} = D_{lk}^{s} + D_{lk}^{t}$$

$\Downarrow$

$H_{1}^{lk}$ : 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 $(l,k)$:

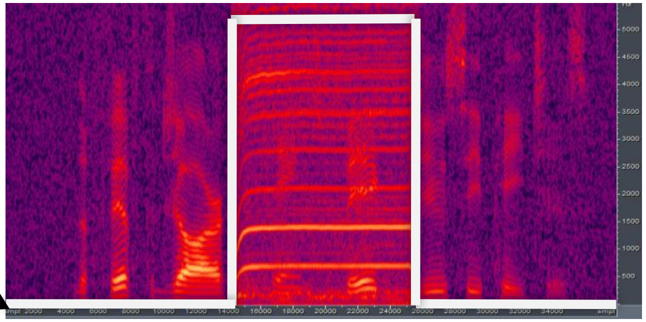
$\eta_0^{lk}$ *– transient is a noise component*

$\eta_1^{lk}$ *– 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 $\eta_j^{lk}$ :

$$\hat{A}_{lk} = \arg\min_{\hat{A}} \Big\{ C_{1j}\, p\left(H_{1s}^{lk} \cup H_{1t}^{lk} \mid \eta_j^{lk}, Y_{lk}\right) \times E\left[ d\left(X_{lk}, \hat{A}\right) \mid Y_{lk}, H_{1s}^{lk} \cup H_{1t}^{lk} \right] + C_{0j}\, p\left(H_{0s}^{lk} \cup H_{0t}^{lk} \mid \eta_j^{lk}, Y_{lk}\right) d\left(G_{\min} R_{lk}, \hat{A}\right) \Big\}$$

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

# Proposed Gain Function – cont.

- Based on above definitions, the gain function is defined : $\hat{A}_{lk} = G_{\eta_j}(\xi_{lk}, \gamma_{lk}) Y_{lk}$

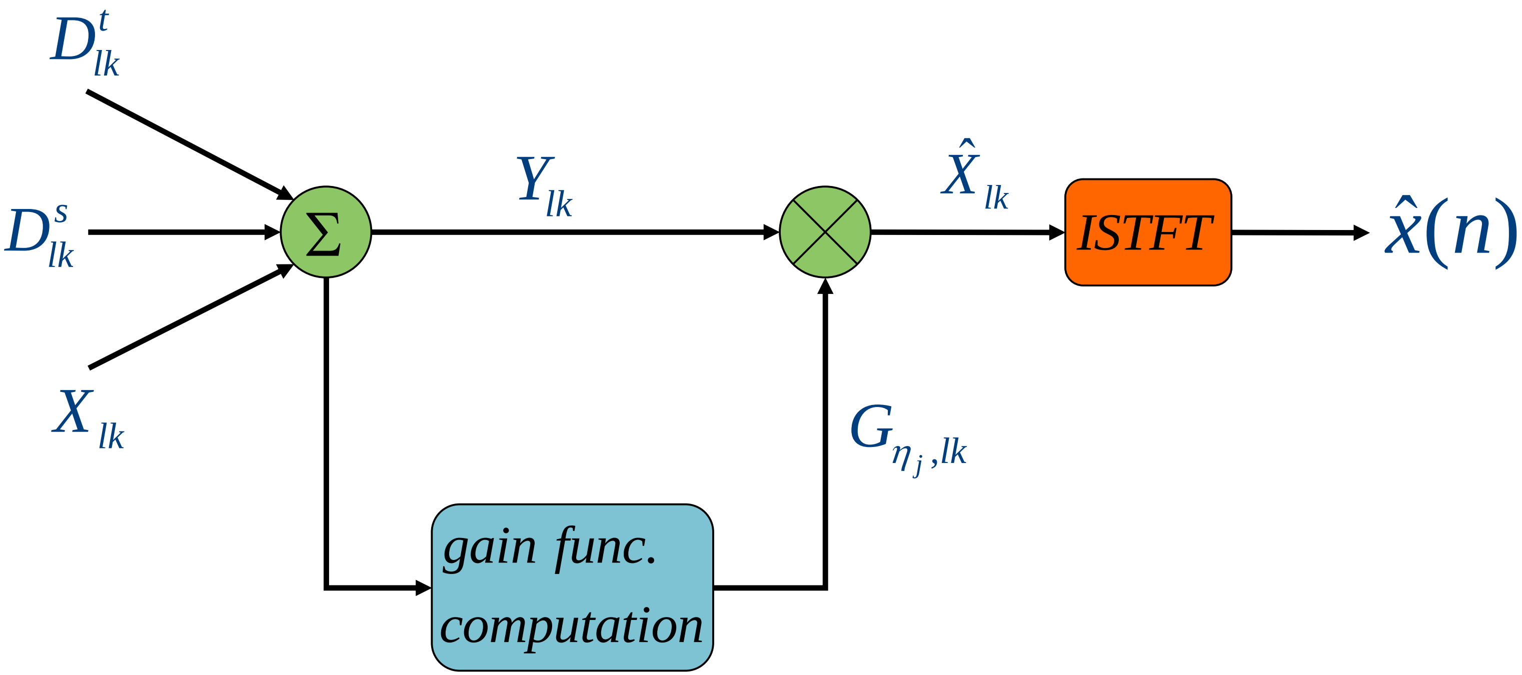
$$\textit{where } G_{\eta_j}(\xi_{lk}, \gamma_{lk}) = G_{\min}^{1-a} G_{LSA}(\xi_{lk}, \gamma_{lk})^a$$

$$\gamma_{lk} = \frac{|Y_{lk}|^2}{\lambda_{s,lk} + \lambda_{t,lk}} \text{ :a-posteriori SNR}$$

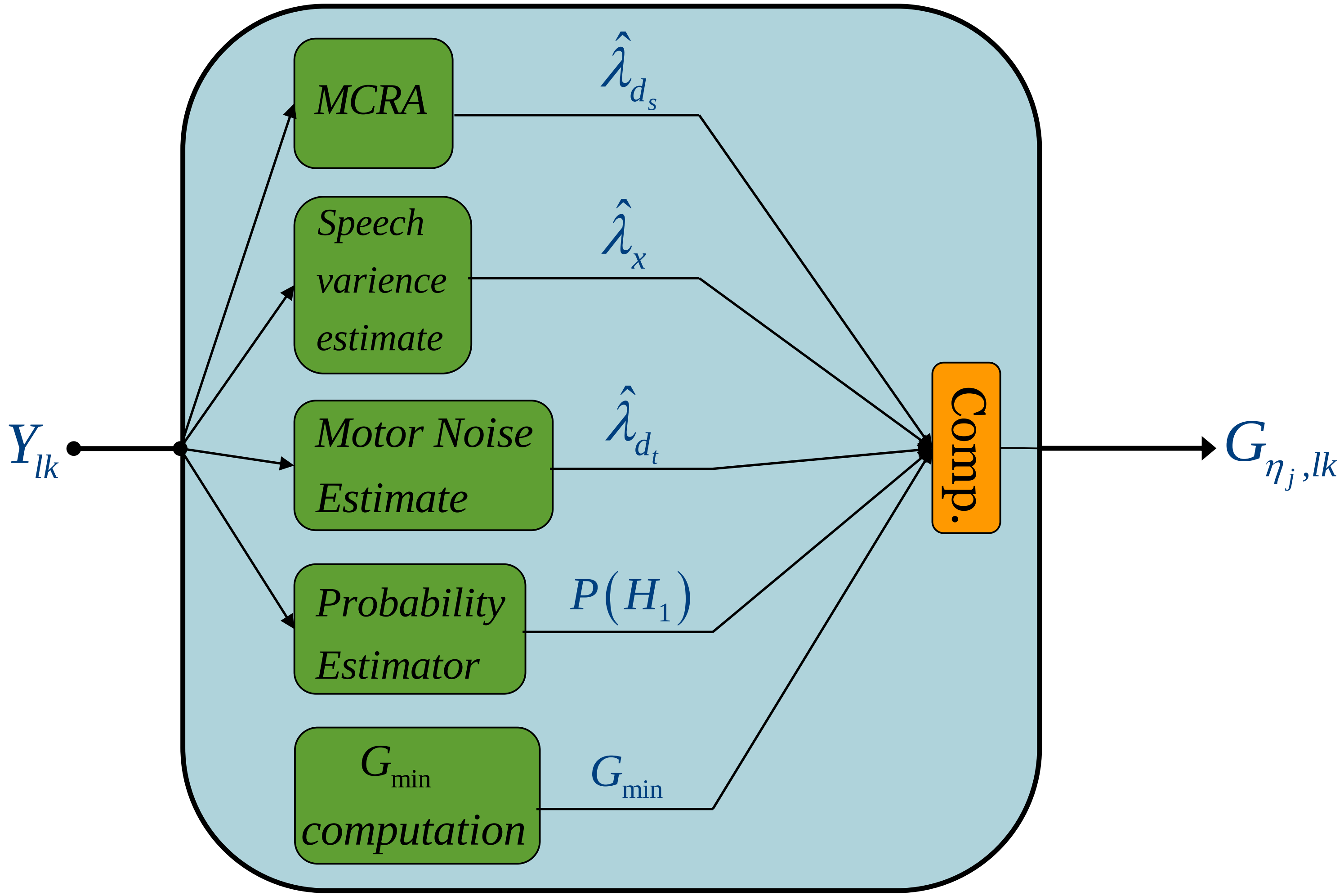
$$\xi_{lk} = \frac{\lambda_{x,lk}}{\lambda_{s,lk} + \lambda_{t,lk}} \text{ :a-priori SNR}$$

- When no motor noise exists (indicator=0 ), we will use the conventional OMLSA: $a = P(H_1^{lk})$.

# Block Scheme



# Block Scheme

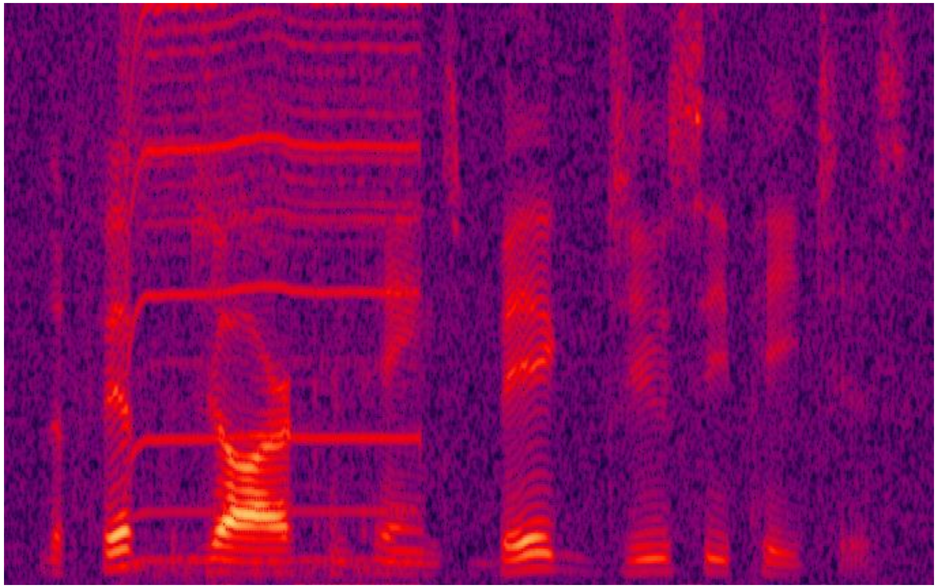


# Experimental Results

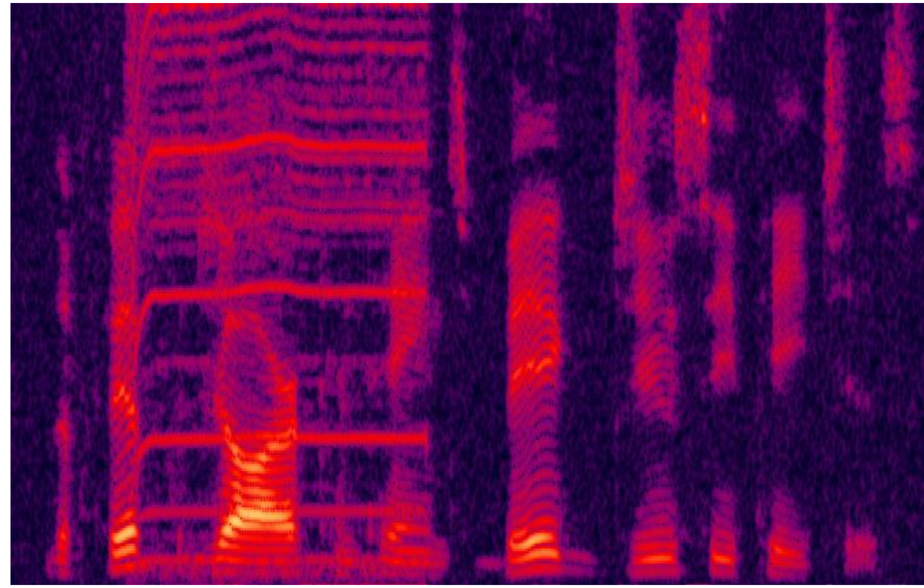
## Parameters Setup:

- 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.

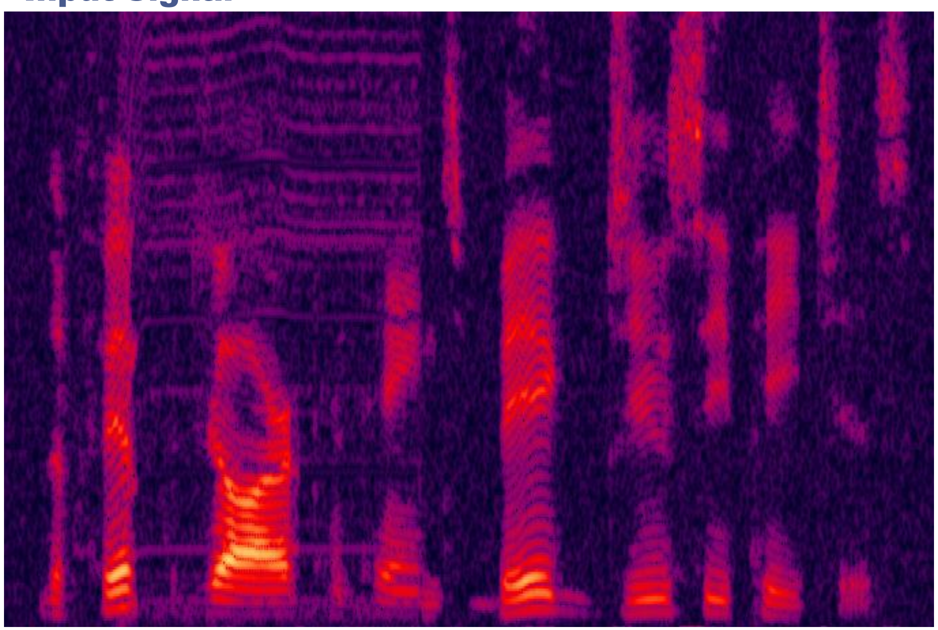
# Full Zoom SNR=8dB, Male



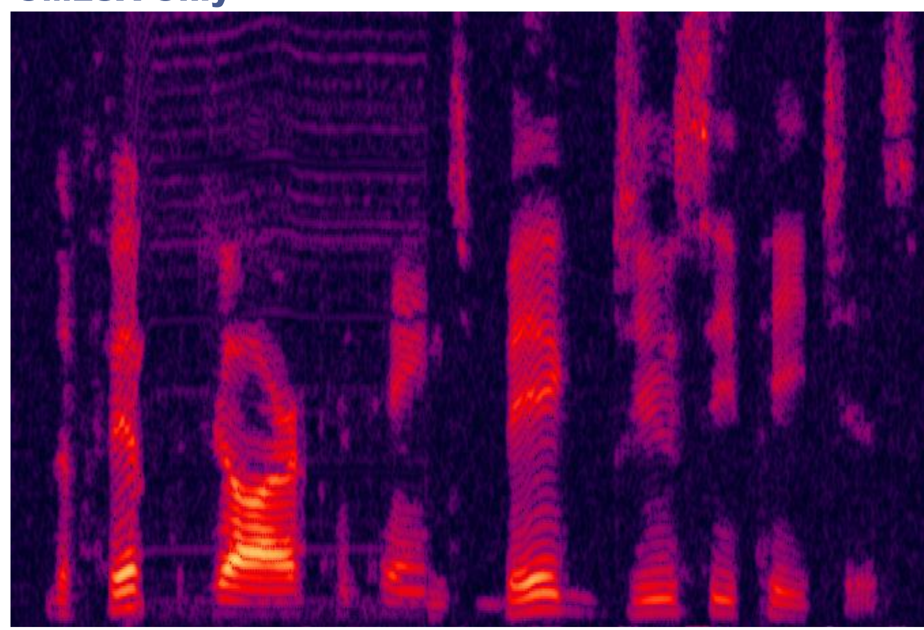
**Input Signal**



**OMLSA Only**

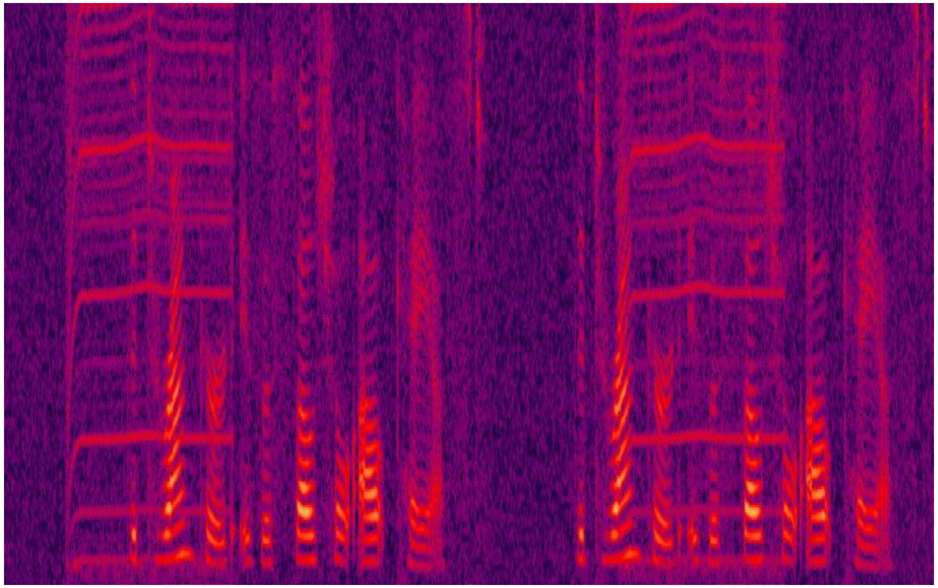


**Gf=-15dB**

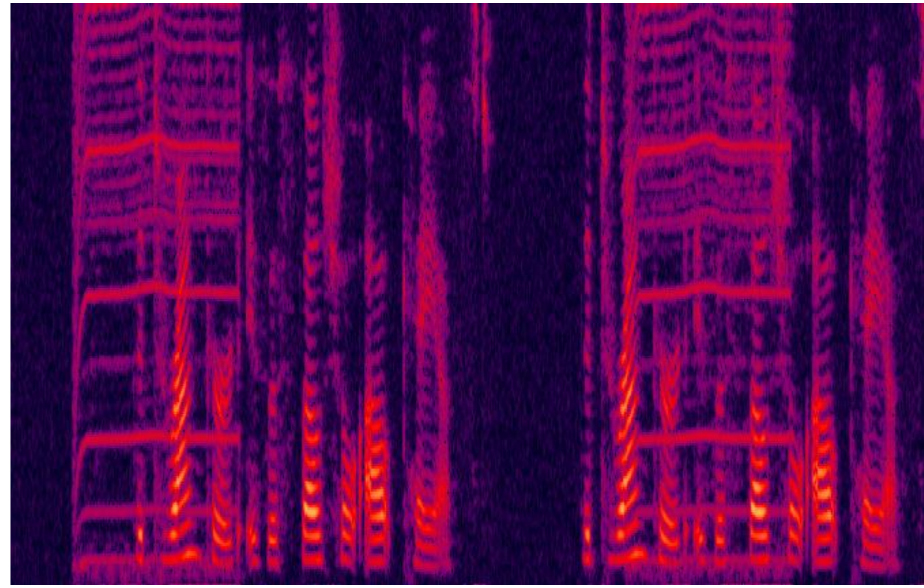


**Gf=-20dB**

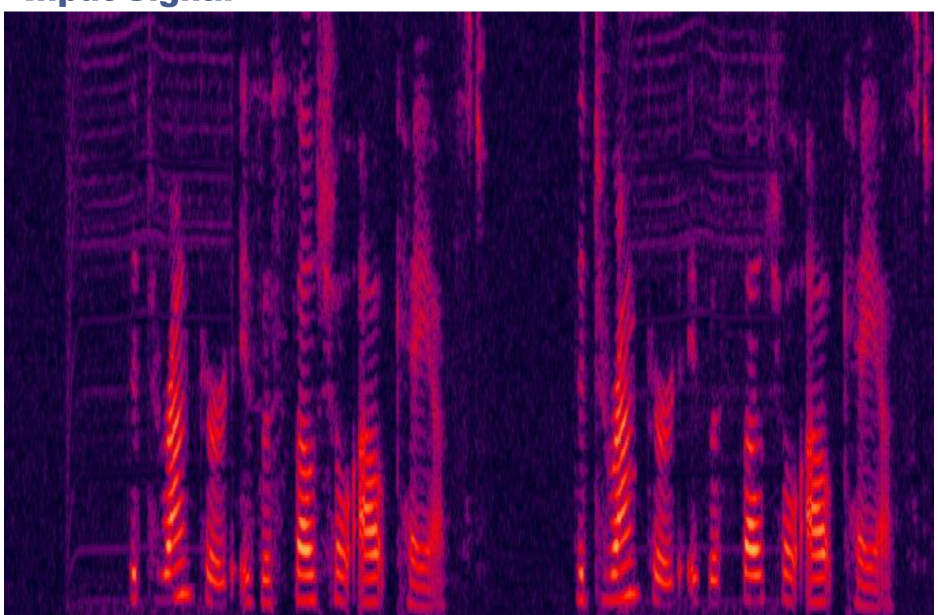
# Full Zoom SNR=10dB, Female



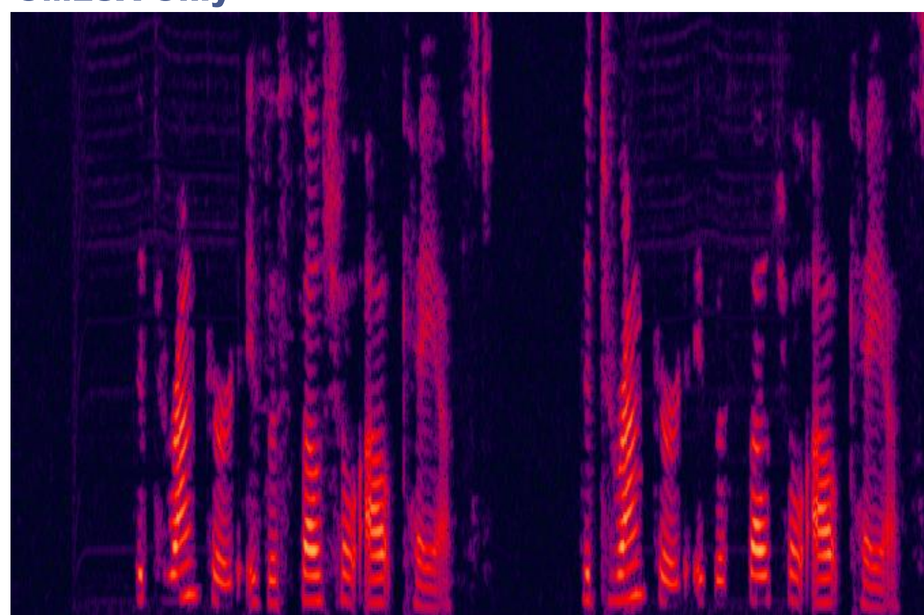
Input Signal



OMLSA Only

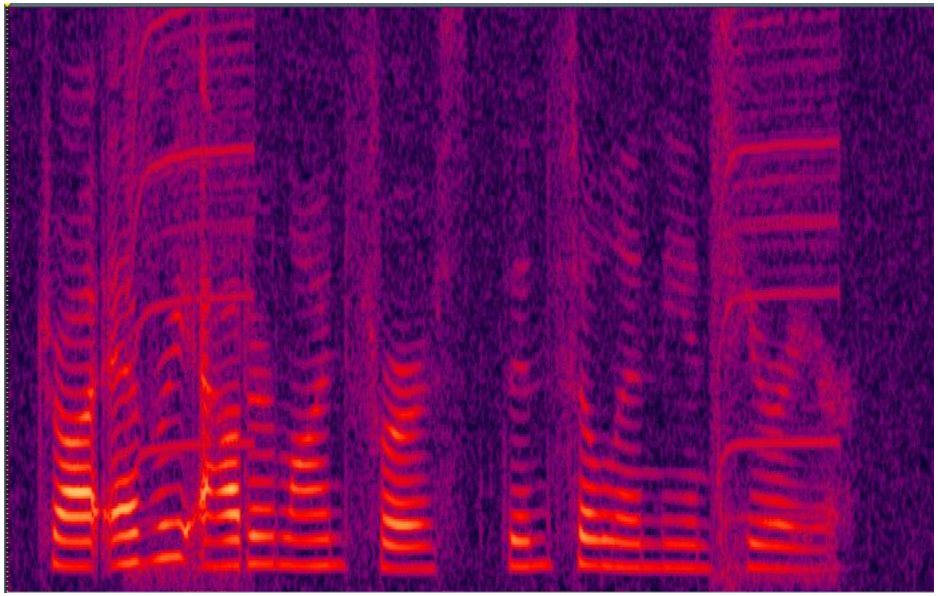


Gf=-15dB

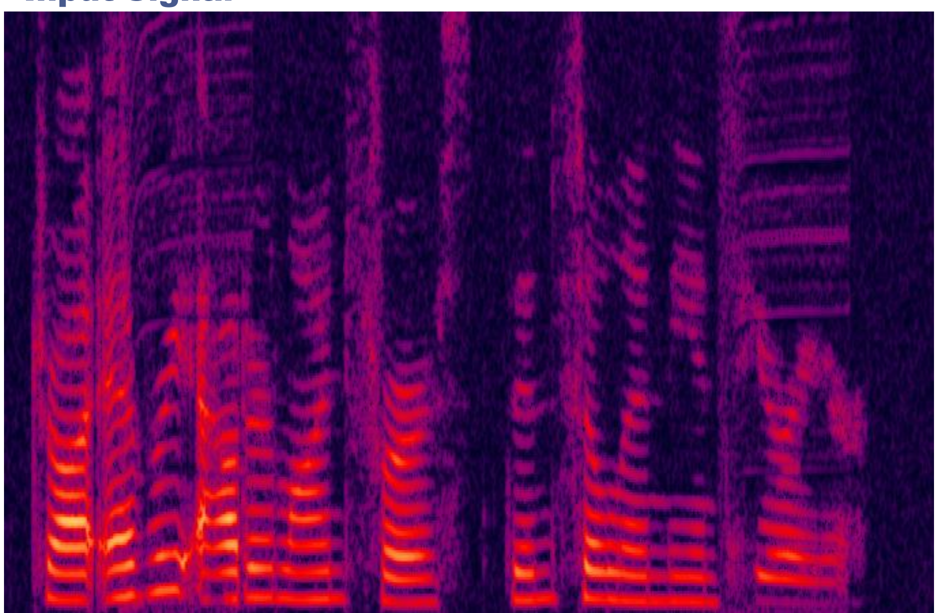


Gf=-25dB

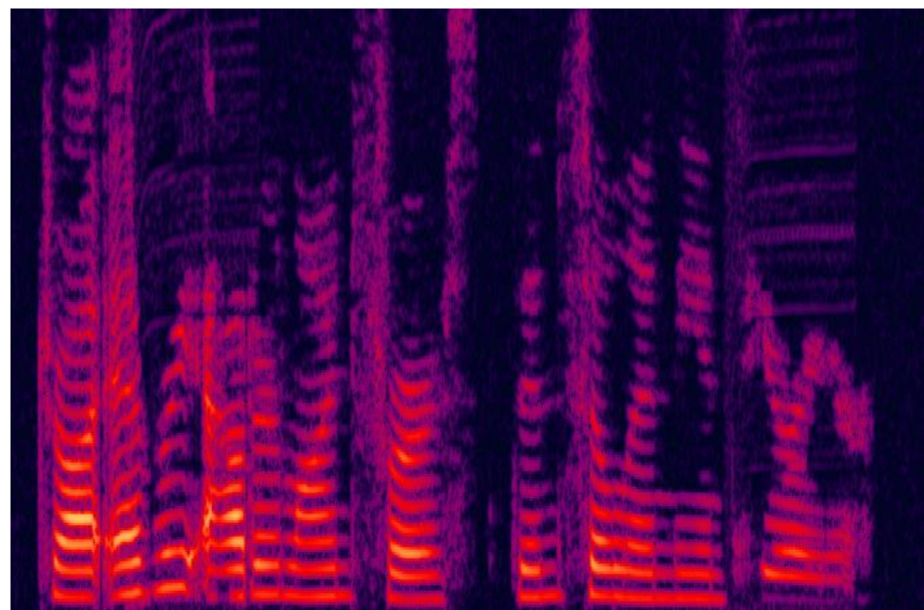
# 2 parts Zoom SNR=15dB, Female



Input Signal

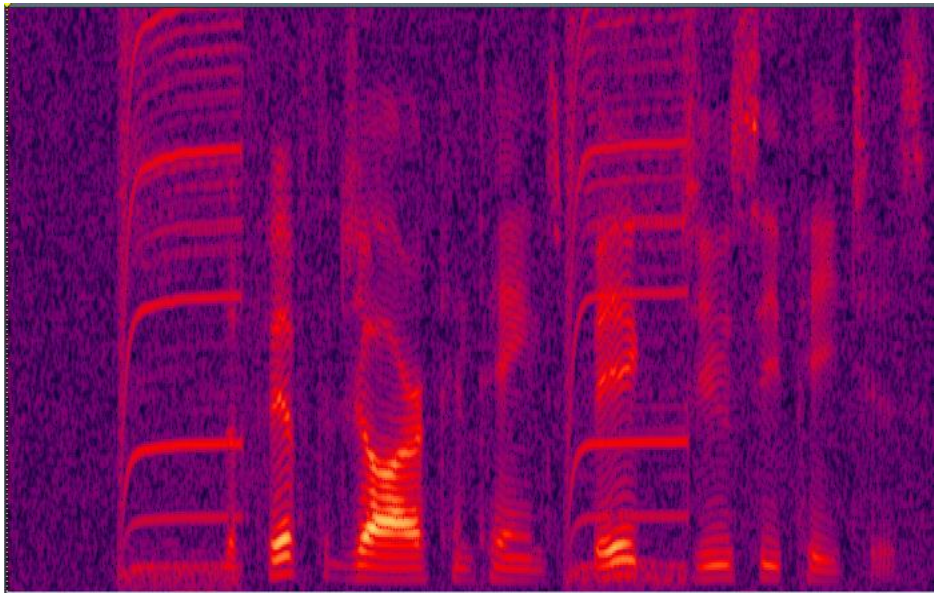


Gf=-12dB

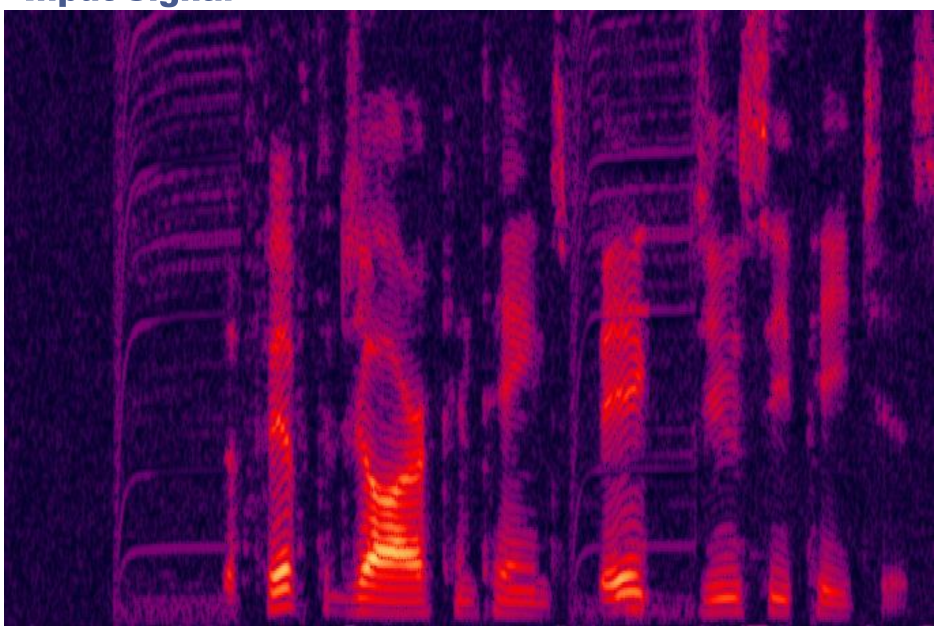


Gf=-20dB

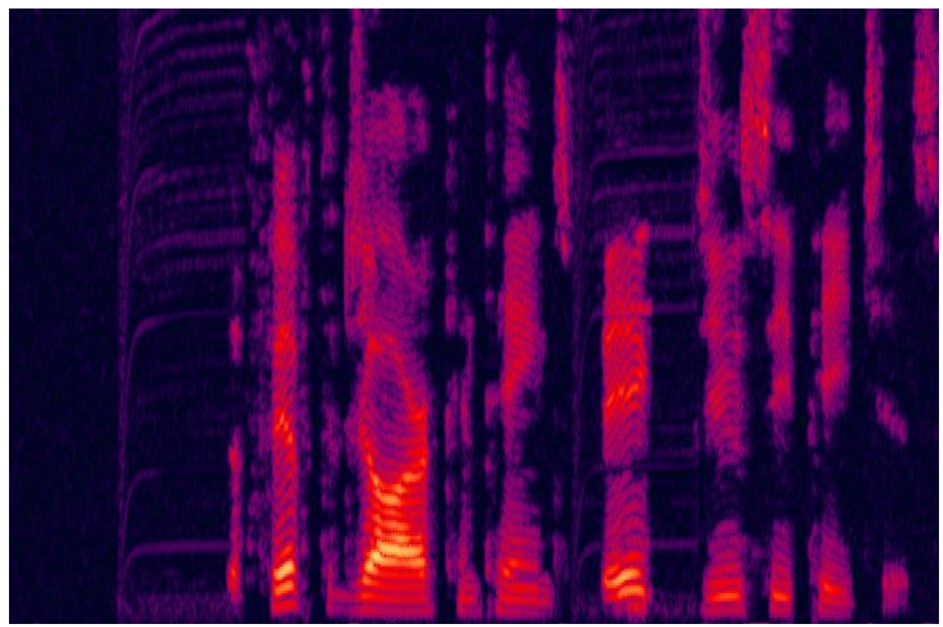
# 2 parts Zoom SNR=10dB, Male



Input Signal

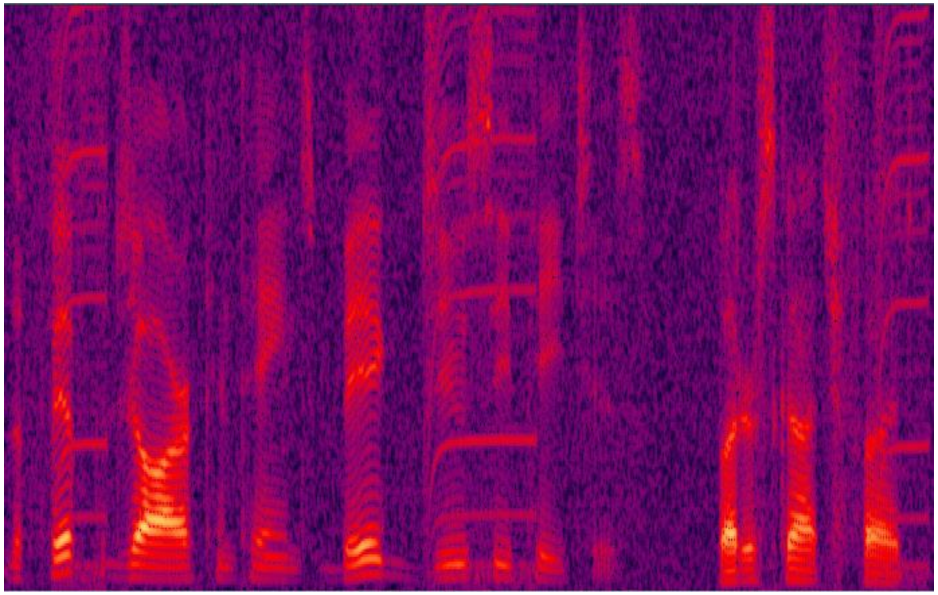


Gf=-15dB

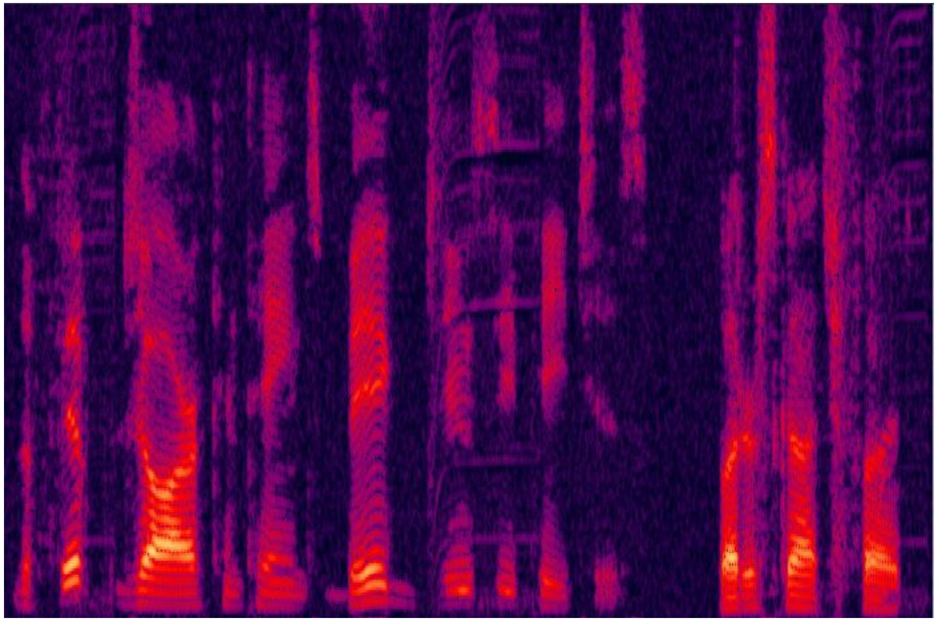


Gf=-25dB

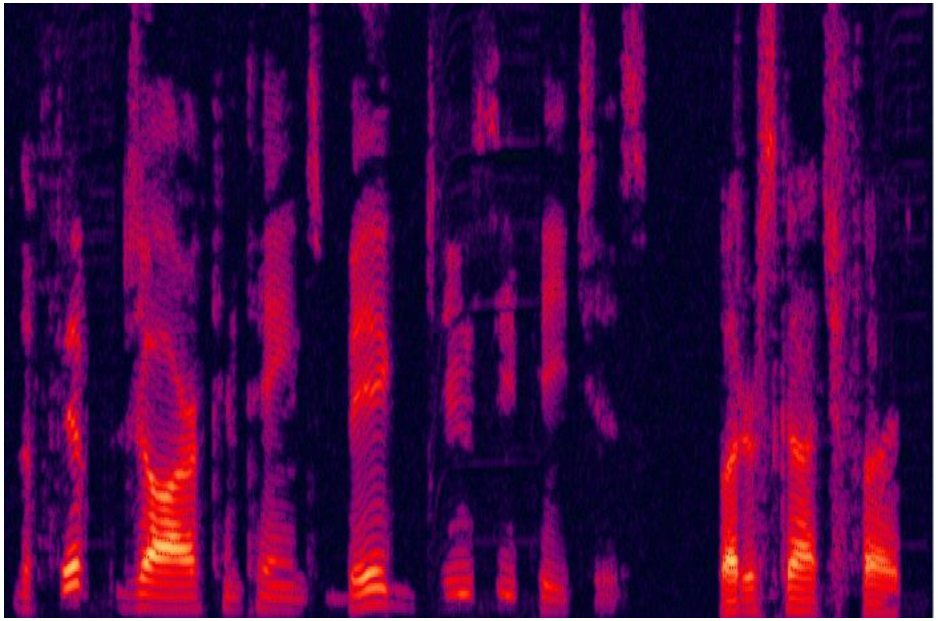
# 3 parts Zoom SNR=15dB, Male



Input Signal

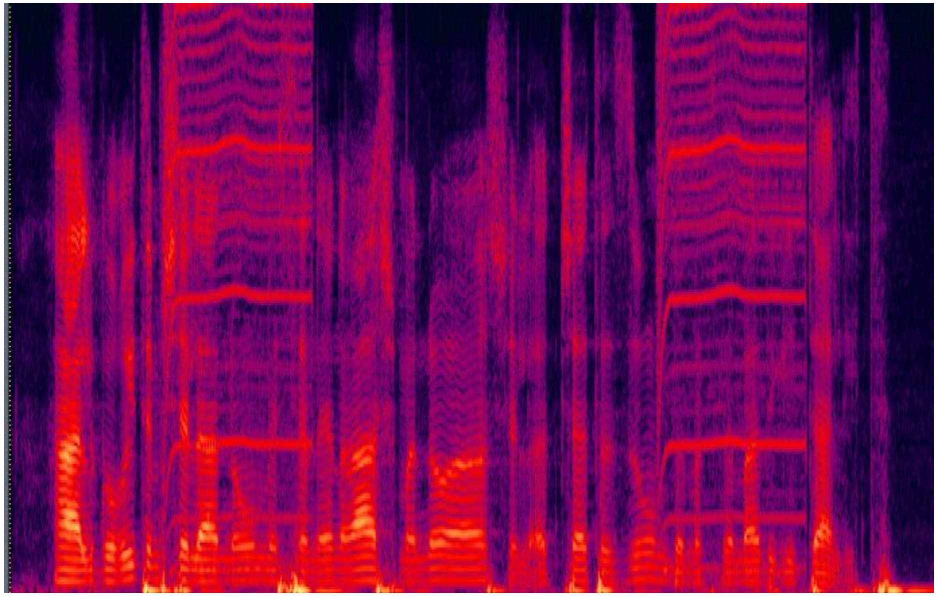


Gf=-15dB

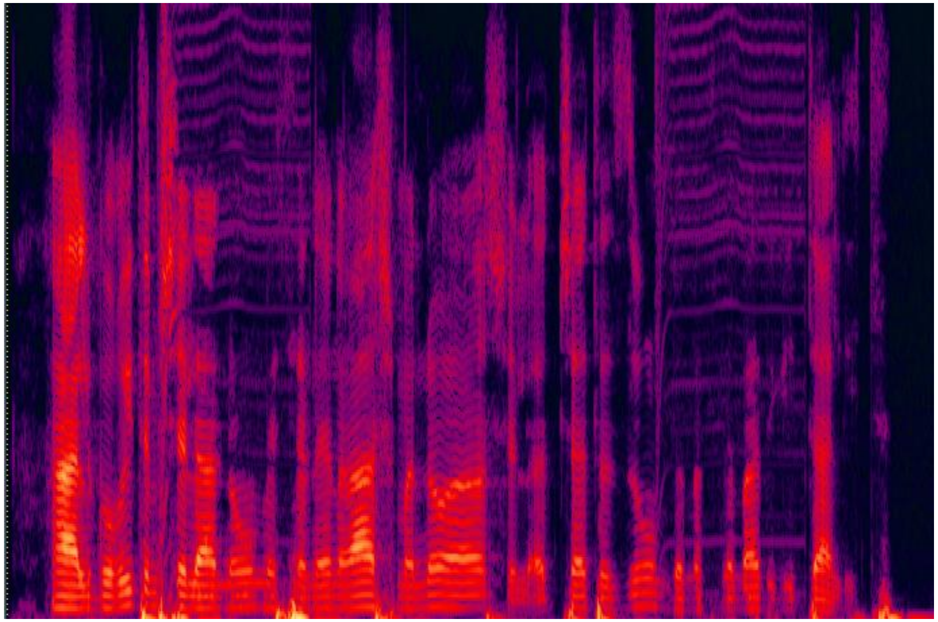


Gf=-25dB

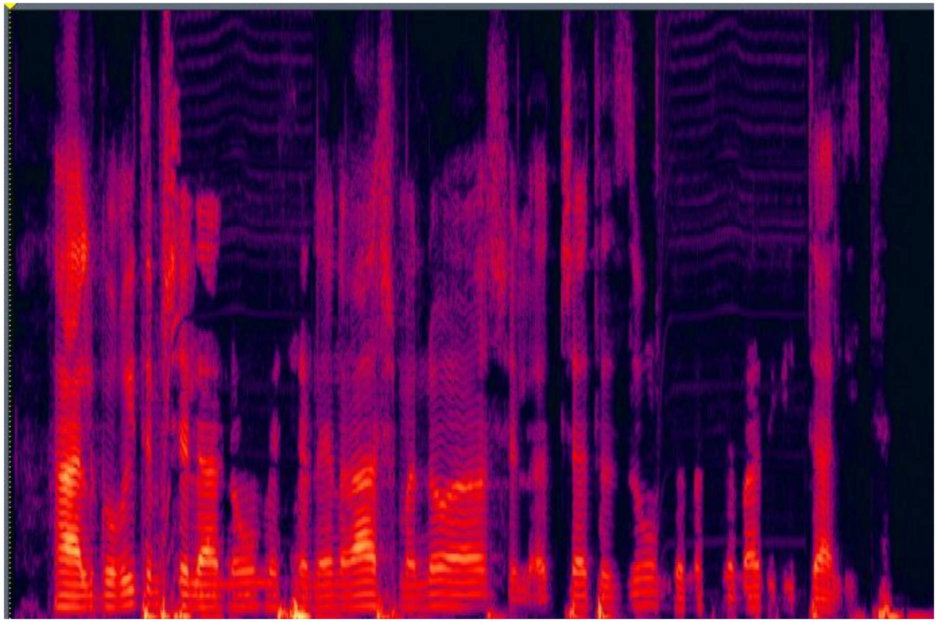
# Full Zoom Real Recording, Male



Input Signal

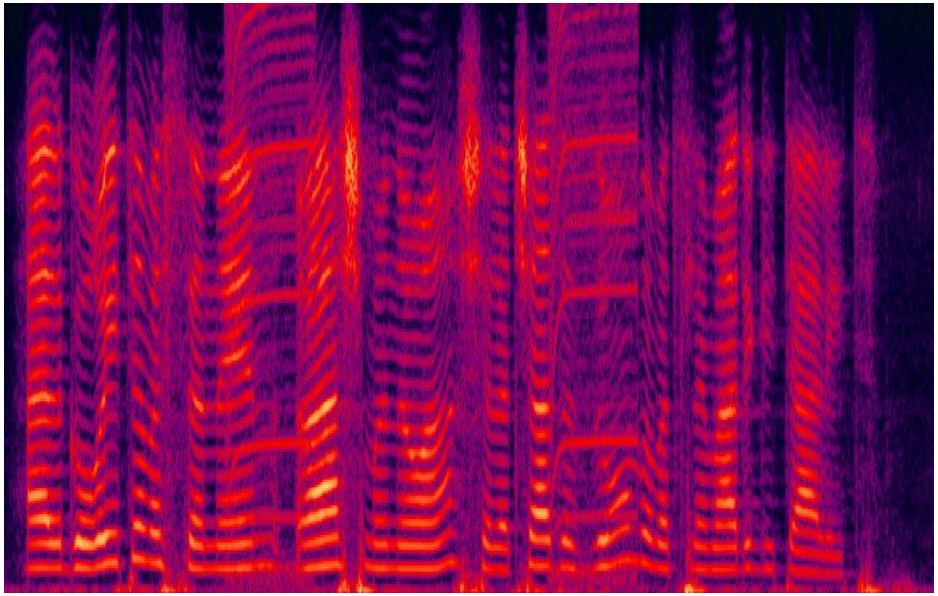


Gf=-15dB

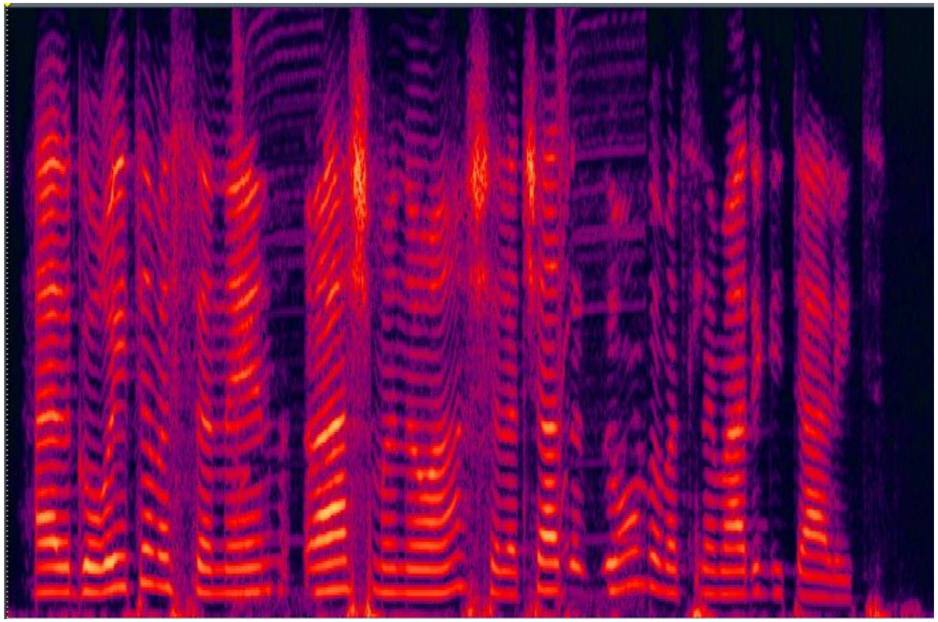


Gf=-25dB

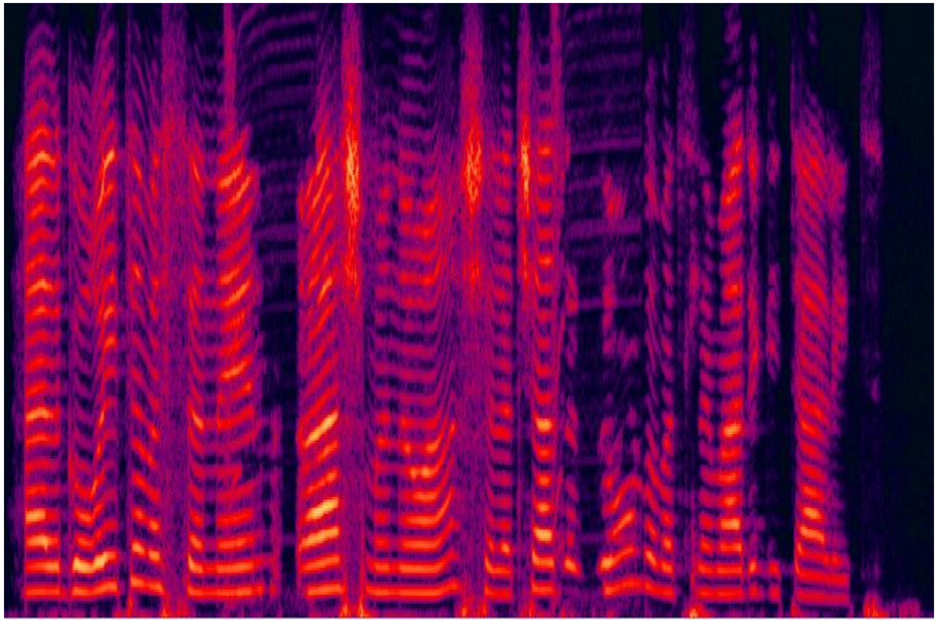
# 2 parts Zoom Real Recording, Female



Input Signal



Gf=-15dB



Gf=-20dB

# Summary

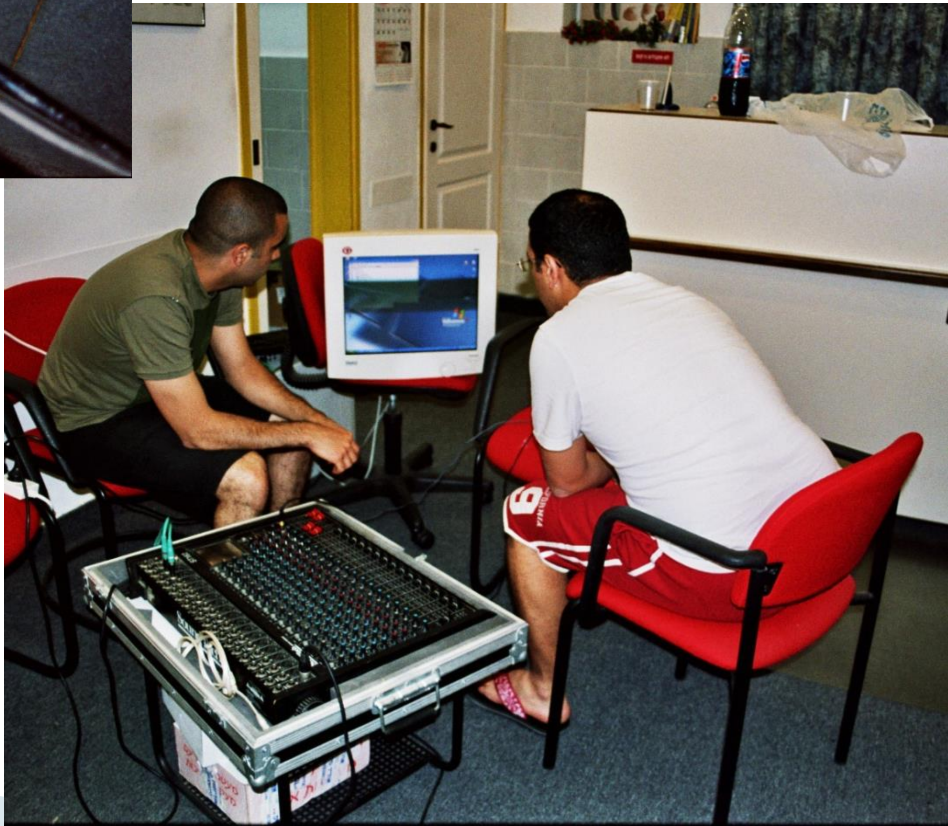
- An algorithm for **suppressing lens motor noise** has been introduced.
- **An optimal estimator**, is derived, while assuming some indicator for the motor-noise presence in the time domain.
- A-priori motor noise spectrum estimate is acquired .
- **A substantial suppression** of the motor noise is 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.



# Motor Noise Estimation

- The a-priori estimation for the motor noise is achieved using an average of early acquired recordings $\lambda_0$.
- The algorithm updates the initial estimation according to pre-determined regions.
  The result is the desired $\hat{\lambda}_t$ :

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

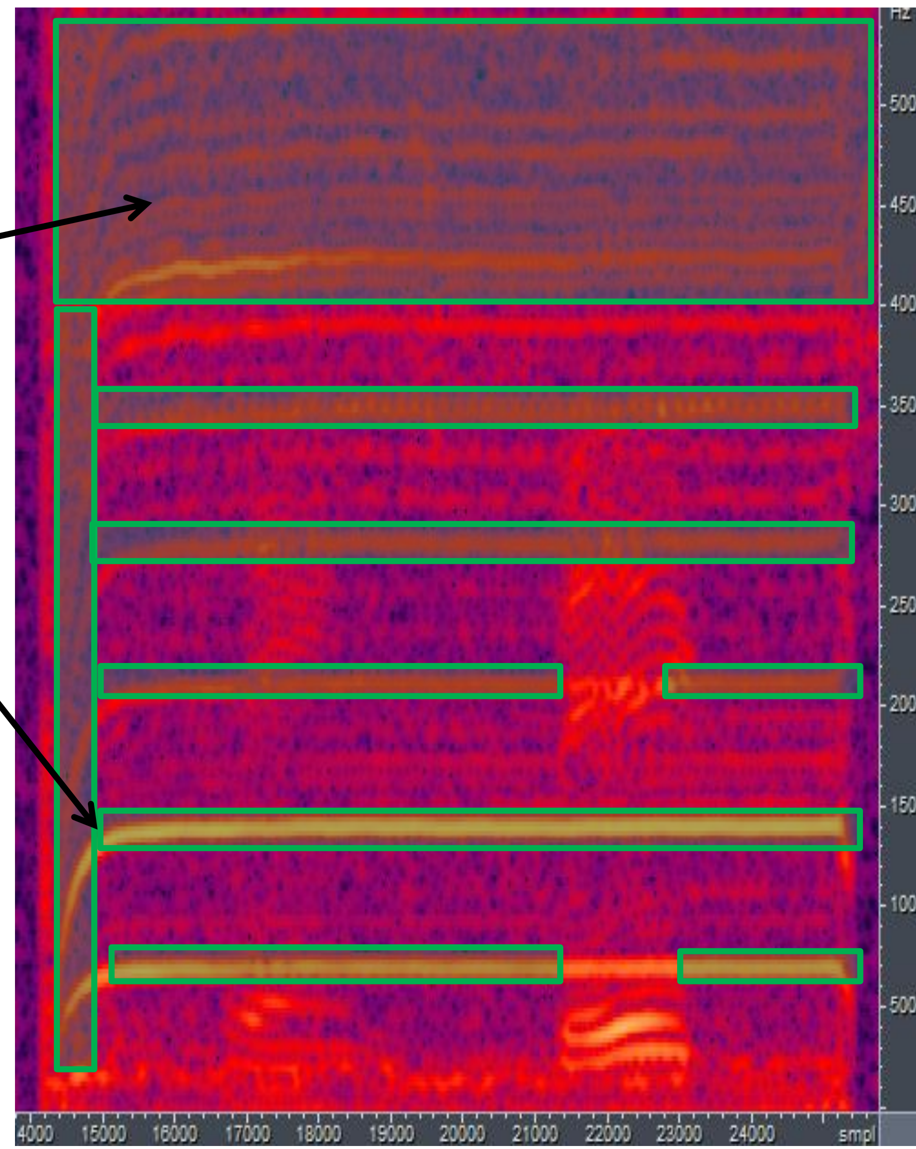
$$\tilde{H}_1 : \hat{\lambda}_t\left(l,k\right) = \alpha\lambda_0(l,k) + \left(1-\alpha\right)\hat{\lambda}_t(l-1,k)$$

- 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 $\tilde{H}_1$.



# Speech Spectral Variance

- In general the speech spectral estimation is calculated by subtracting the motor noise estimation and the background noise estimation from the observed signal.

$$\hat{\lambda}_{x,lk} = \max\left\{ \underbrace{\alpha G_{LSA}{}^{2}\left(\hat{\xi}_{l-1,k}, \gamma_{l-1,k}\right)\left|Y_{l-1,k}\right|^{2}}_{\text{Previous frame estimate}} + \underbrace{(1-\alpha)\left(\left|Y_{l,k}\right|^{2} - \hat{\lambda}_{s} - \hat{\lambda}_{t}\right)}_{\text{Current frame estimate}},\ \lambda_{\min} \right\}$$

# Noise Spectral Estimation

- Using the MCRA algorithm the noise spectrum is estimated. Let $\hat{\lambda}_{s,lk}$ be the noise spectrum estimation.
- Let $p'_{lk}$ denote the conditional speech presence probability, therefore the update equation for $\hat{\lambda}_{s,lk}$ is :

$$\hat{\lambda}_s(l+1,k)=\tilde{\alpha}_d(l,k)\hat{\lambda}_s(l,k)+\left[1-\tilde{\alpha}_d(l,k)\right]\left|Y(l,k)\right|^2$$

where $\tilde{\alpha}_d(l,k)=\alpha_d+(1-\alpha_d)\,p'(l,k)$.

- Let $S_r(l,k)=S(l,k)/S_{\min}(l,k)$ denote the ratio between the local energy of the noisy signal and its derived minimum.
- **The decision rule is:** $S_r(l,k)\underset{\tilde{H}_1}{\overset{\tilde{H}_0}{\gtrless}}\delta$ , $\delta$ threshold value.

# Constant Attenuation

- In order to suppress the noise (stat. & transients) when speech is absence, minimizing the next equation yields the solution above:

$$\arg\min_{G_{\min}} \left\{ E\left[ G_{\min}\left(\lambda_{s,lk} + \lambda_{t,lk}\right) - G_f \lambda_{s,lk} \right] \right\}$$

- Let $G_{\min}$ denote the constant attenuation under speech absence:

$$G_{\min} = G_f \frac{\lambda_{s,lk}}{\lambda_{s,lk} + \lambda_{t,lk}}$$

# Speech Presence Prob.

- Let $P(H_1) = \left\{ 1 + \frac{\hat{q}_{lk}}{1-\hat{q}_{lk}} \left(1+\xi_{lk} \exp(-\upsilon_{lk})\right) \right\}^{-1}$

$$\hat{q}(l,k) = 1 - P_{local}(l,k) P_{global}(l,k) P_{frame}(l)$$

- Where $\hat{q}_{lk}$ is the estimator for the a-priori signal absence probability.
- $\hat{q}_{lk}$ is larger if either previous frames or recent neighboring frequency bins do not contain speech.