Technion - Israel Institute of Technology
Department of Electrical Engineering

Signal and Image Processing Laboratory

Robust Detection Of Watermarks in Audio Signals

Shay Mizrachi

Supervisor: Prof. David Malah

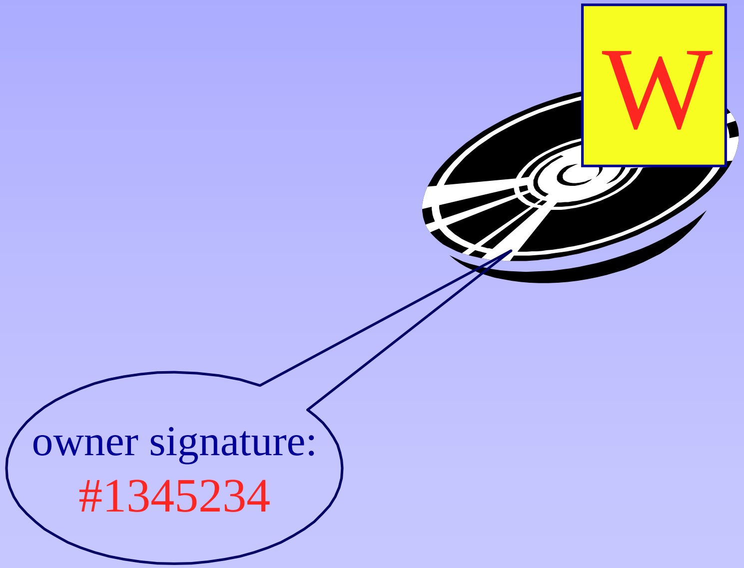
14/8/2000

Outline

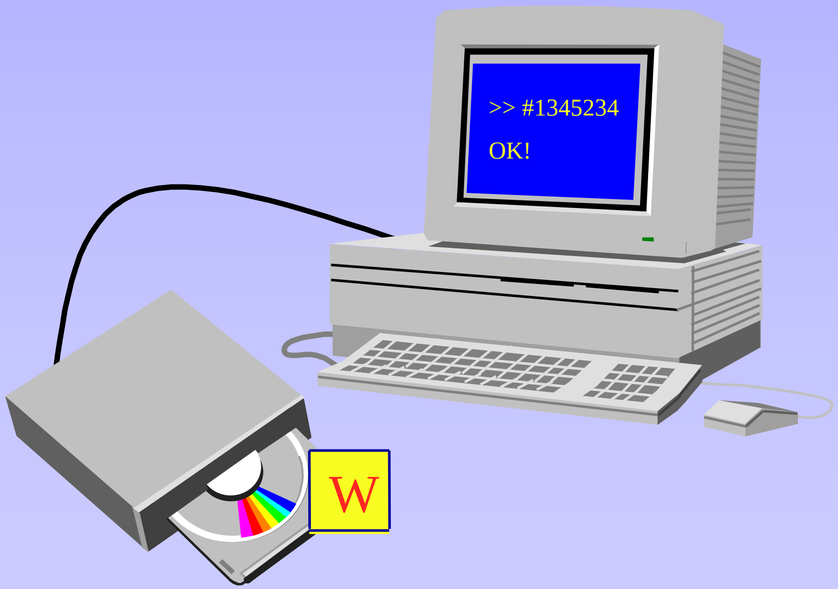
- What is digital watermarking?
- The main requirements from watermarking system.
- How is signature is embedded?
  - The psycho-acoustic model.
- The detection mechanism.
- Possible attacks and the modification needed in the detection mechanism.
- Comparison of different solutions.
- Results and Conclusions.

Digital Watermarking
For Copyright Protection

Signature embedding



Signature detection

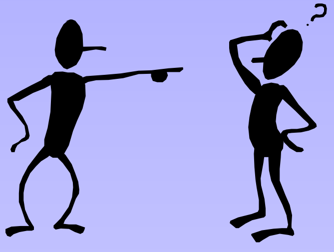


Signature Requirements

- Must be embedded within the data itself.
- Inaudible to the human ear.
- Knowledge of algorithm doesn't allow signature removal.
- Any damage to the signature will cause a damage to the signal itself.
- False alarms are much more acute than misses and must be prevented as much as possible.

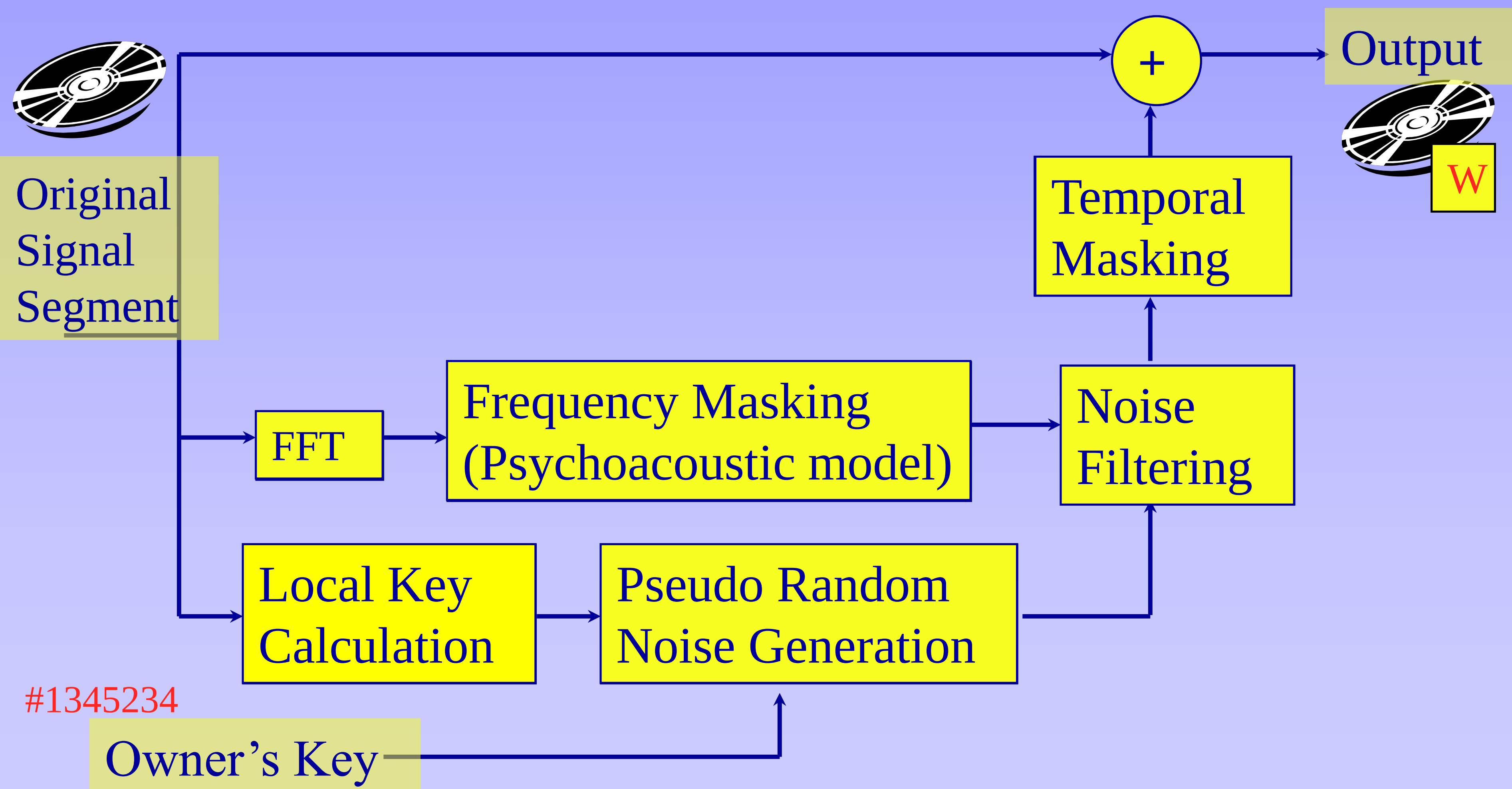
Signature Requirements (cont’d)

- The Deadlock problem, i.e., multiple ownership claims, must be solved.



- A Solution is to keep the owner’s original file, or parts of it.

Signature Embedding Mechanism



Frequency Masking

- Finding the spectral threshold using a psycho-acoustic model.

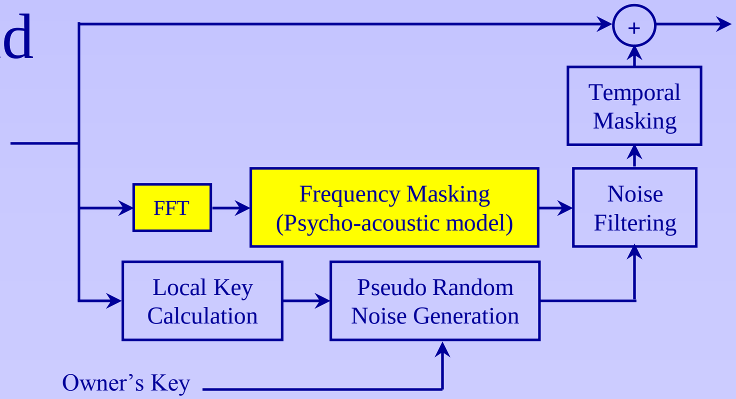
Spectrum calculation → Find Tonal and Non-tonal components

↓

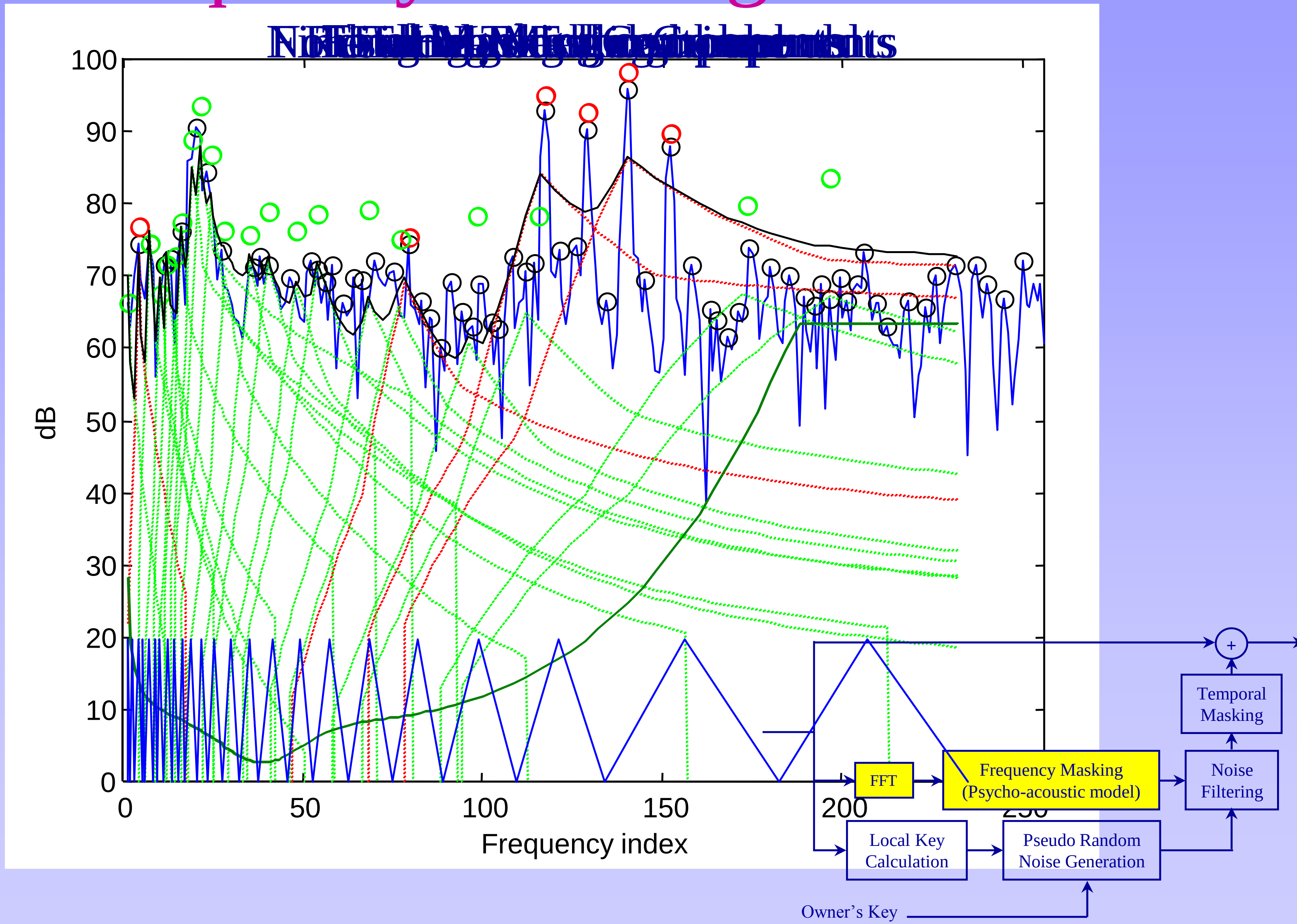
Use hearing threshold

Use individual masking threshold

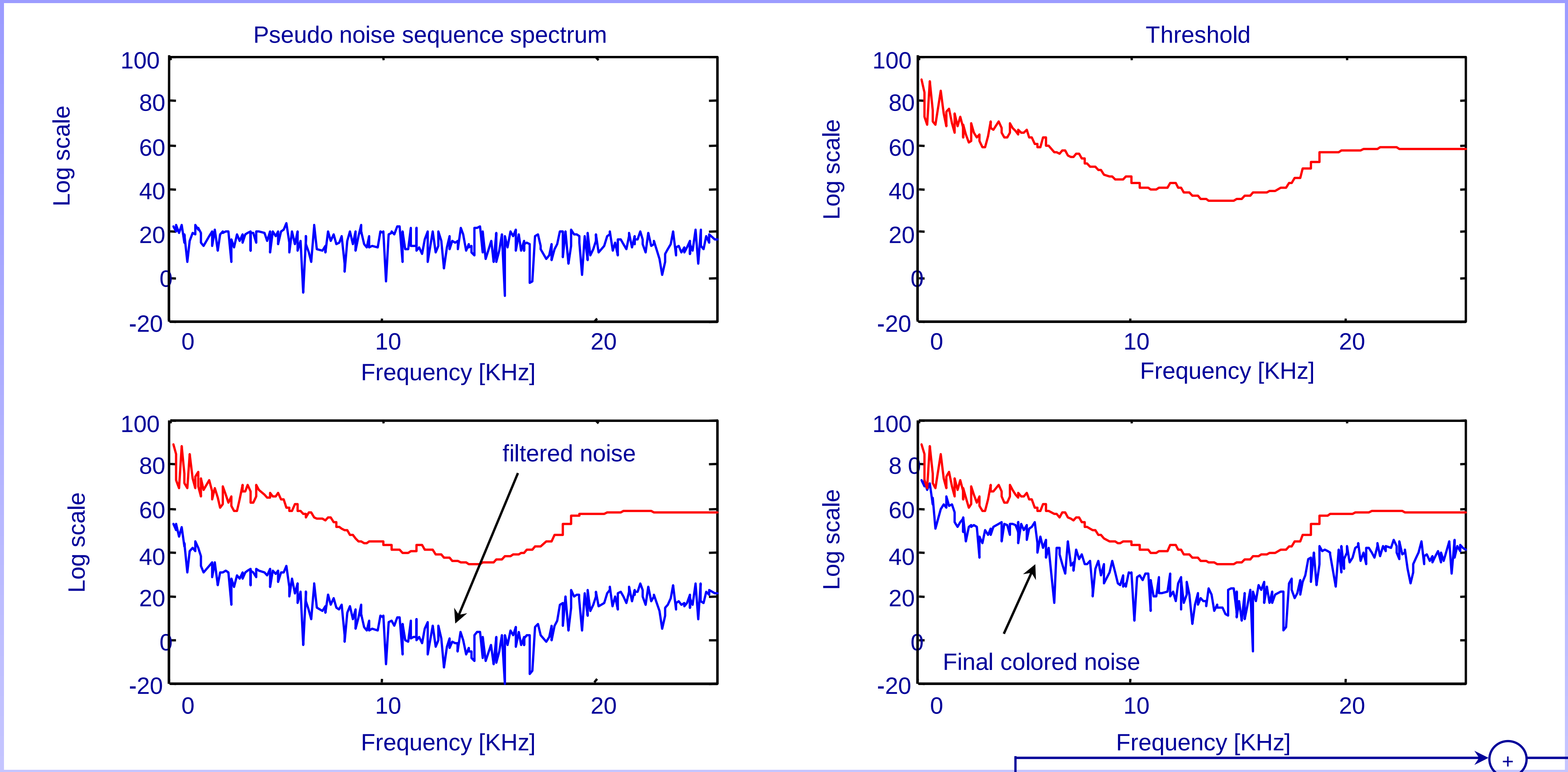
Generate global masking threshold

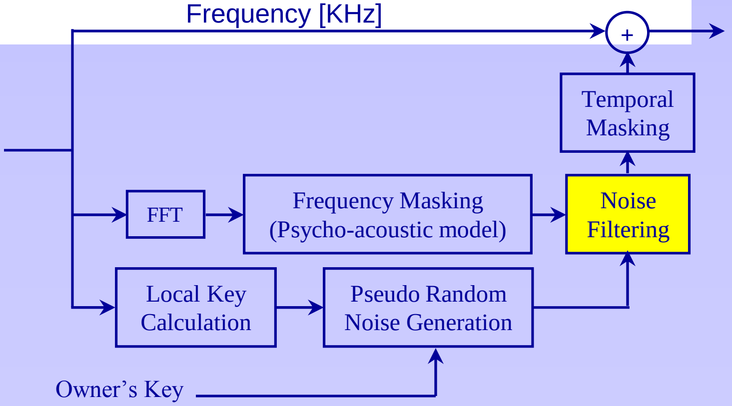


Frequency Masking - continue

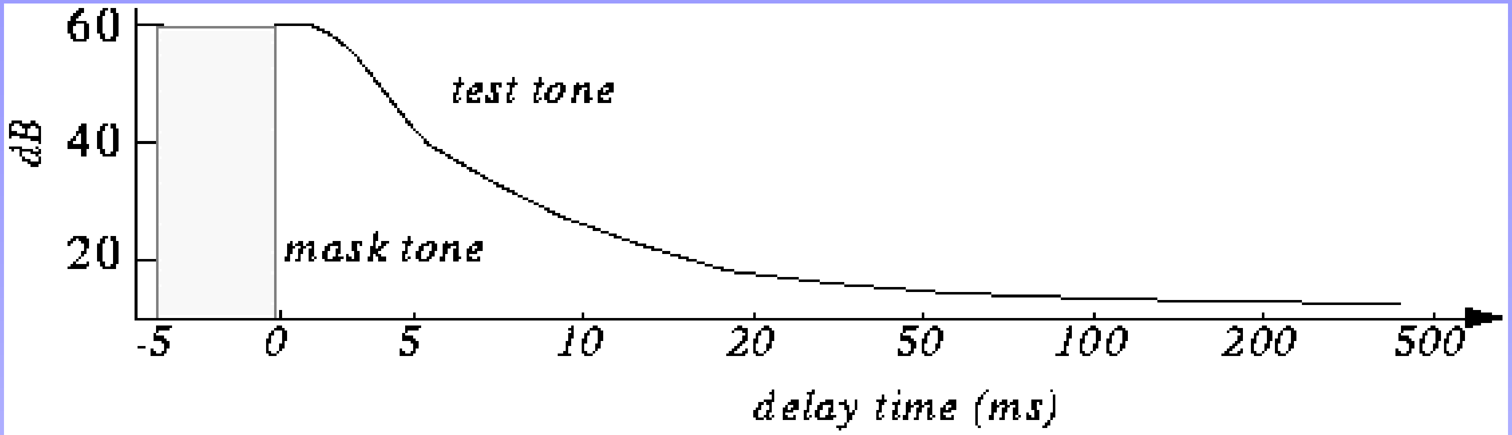


Noise Filtering



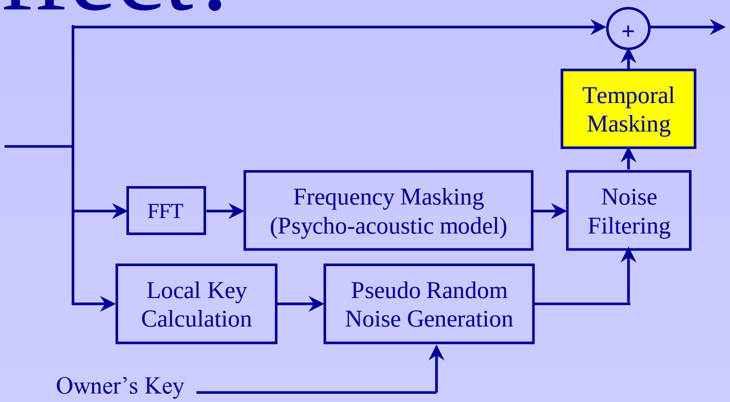


Temporal Masking



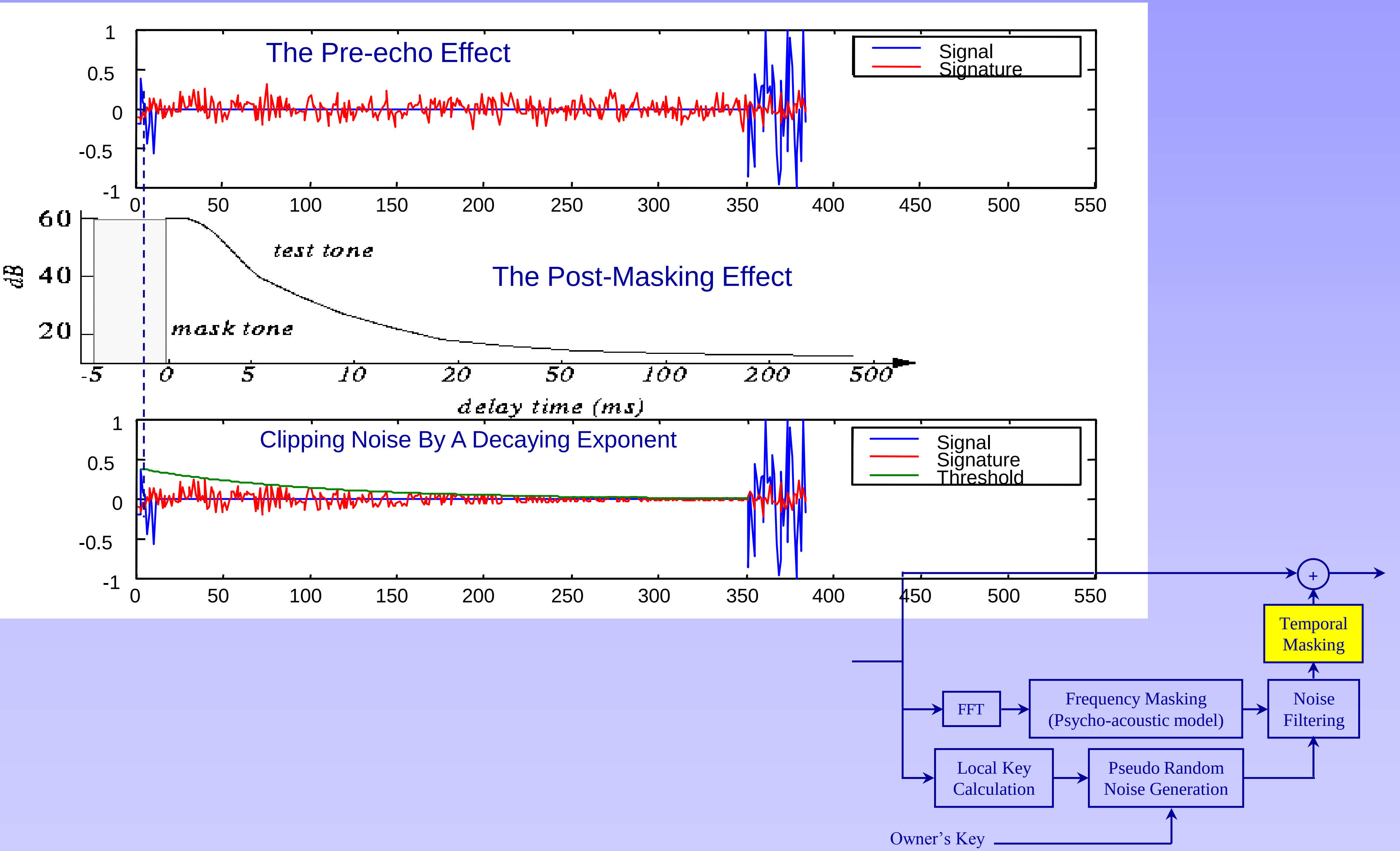
Temporal masking is used in order to prevent the pre-echo effect!

What is the Pre-Echo effect?

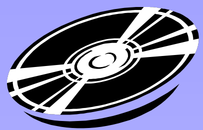


Temporal Masking

– using the post masking effect to reduce pre-echo



Demos

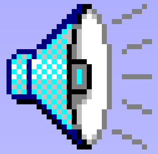


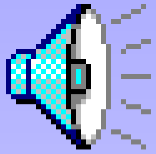
W1

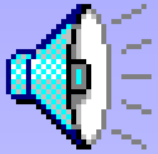
original

Watermarked

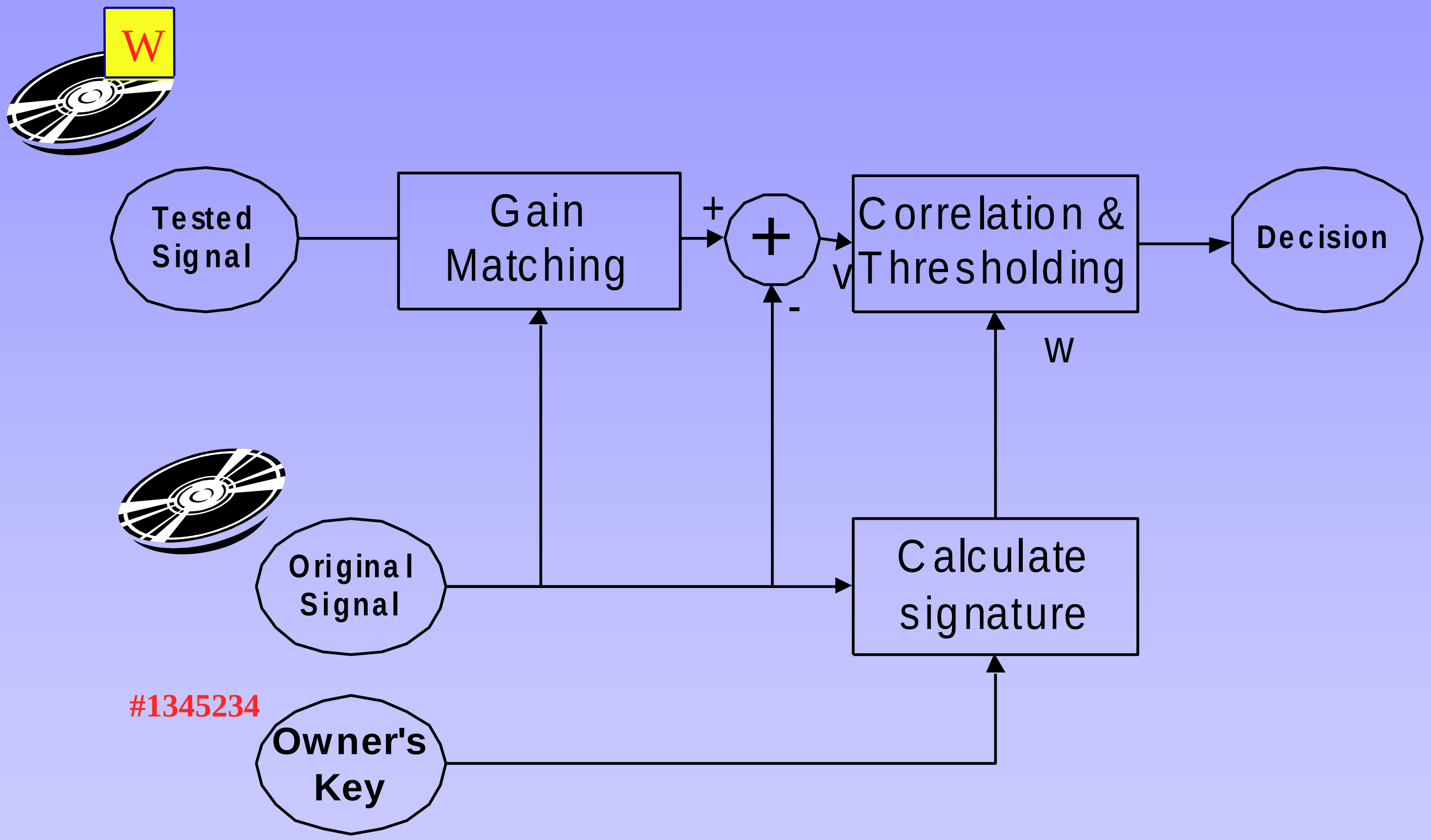
Watermark





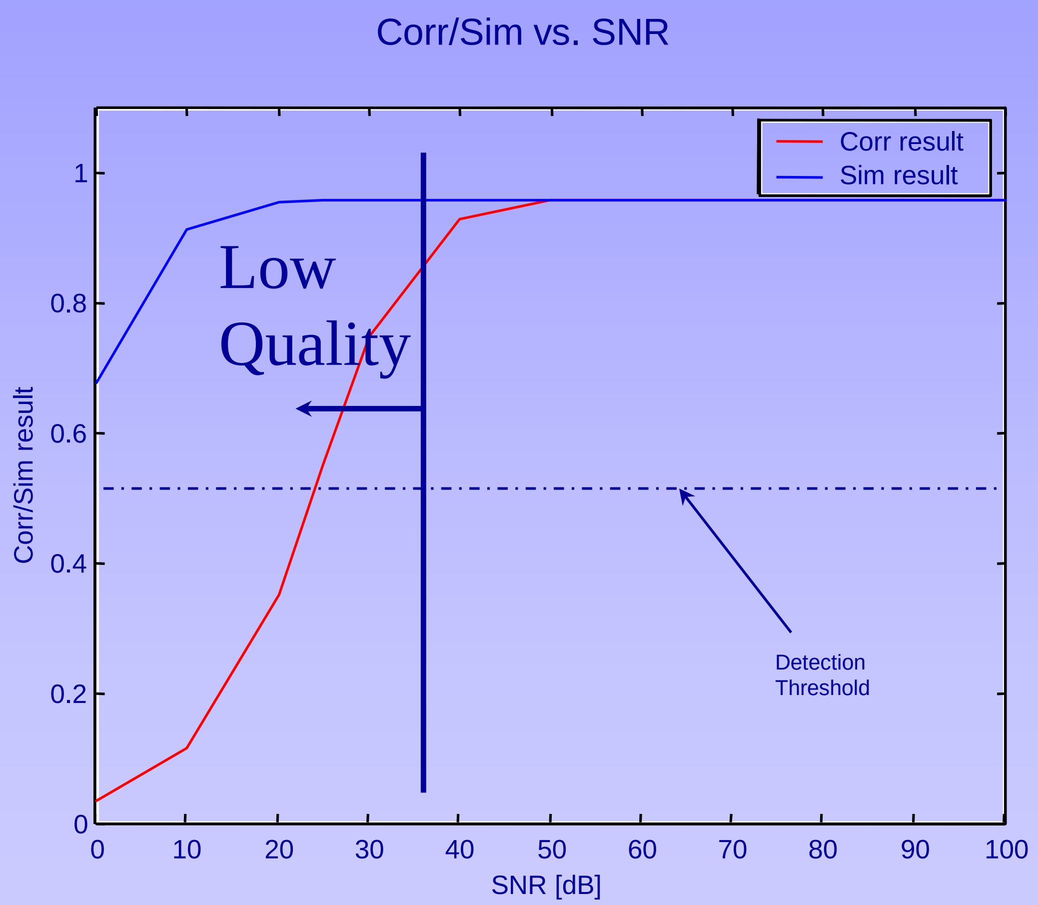


Detection Mechanism



Correlation vs. SNR

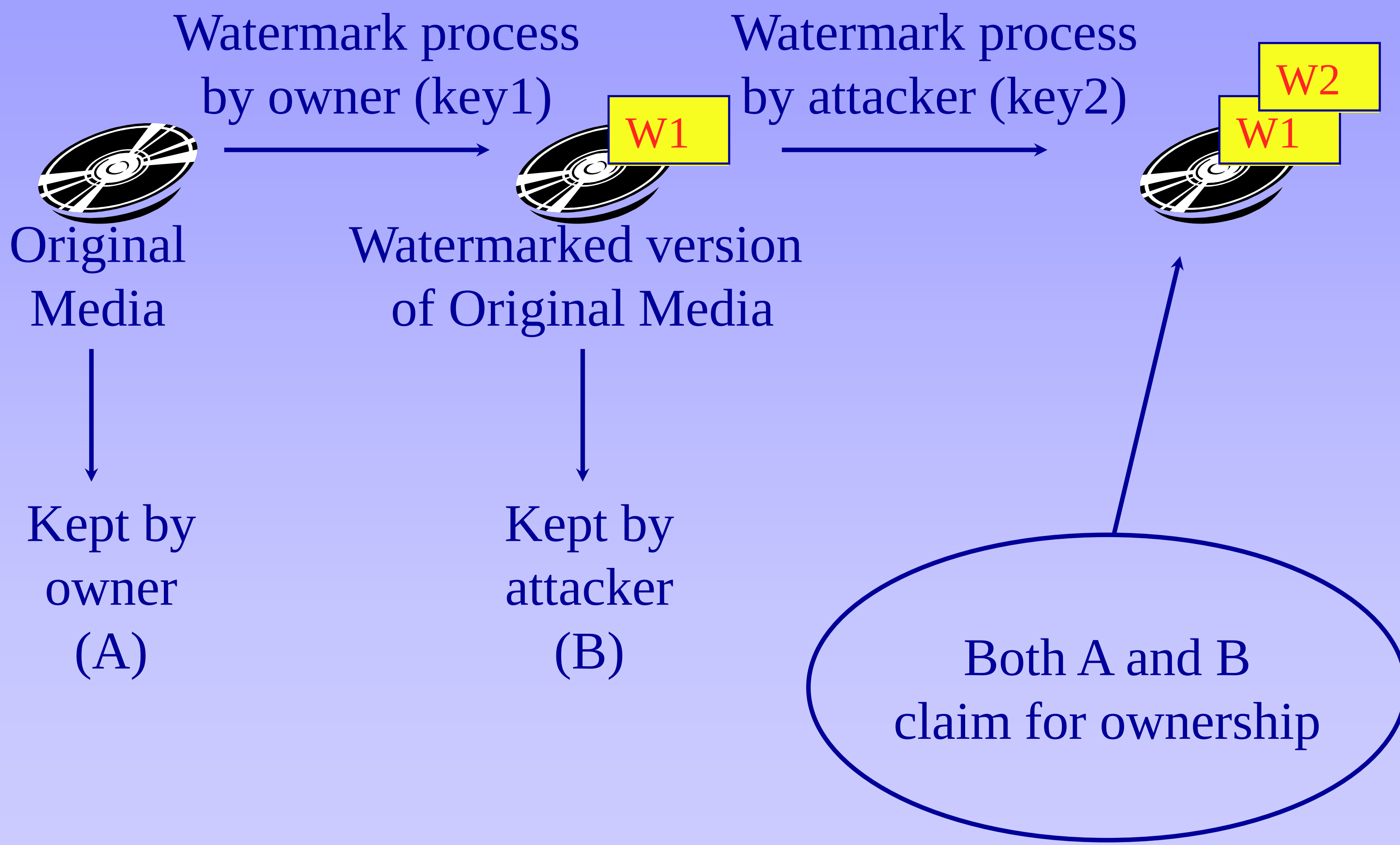
- White Gaussian noise was added in attempt to <u>destroy</u> the signature.
- The signature is still detected with high correlation/similarity.



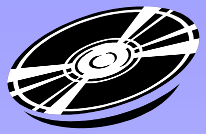
$$corr(w,v) \equiv \frac{\sum_{i=1}^{N} w_i v_i}{\sqrt{\sum_{i=1}^{N} w_i^{\,2}} \times \sqrt{\sum_{i=1}^{N} v_i^{\,2}}}$$

$$sim(w,v) \equiv \frac{\sum_{i=1}^{N} w_i v_i}{\sum_{i=1}^{N} w_i^{\,2}}$$

How is the dead lock problem solved?

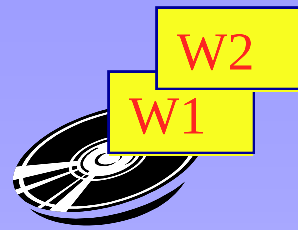


How is the dead lock problem solved? (Cont'd)

A has  and together with key1 can create W1

B has  W1 and together with key2 can create W2

How is the dead lock problem solved? (Cont'd)

<u>First phase:</u> check if 

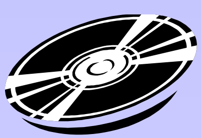
includes W1 and W2

In our case both signatures will be detected in the tested media.

<u>Second phase:</u> check if original of B includes W1

is the original of B and it does include W1

<u>Third phase:</u> check if original of A includes W2


is the original of A and it does not include W2

→ <u>A is the owner</u>

Work Goals:

Dealing with ownership claims of attackers.

To do so:

- Find the characteristics of the attacker's system.
- Modify the detection system in order to increase the signature detection probability for these attacks.

Fundamental assumption:

The attack is limited in the sense of preservation audio quality.

Attacks

Naive attacks:

- Gain
- Coping part of a file
- Equalization
- Compression
- …

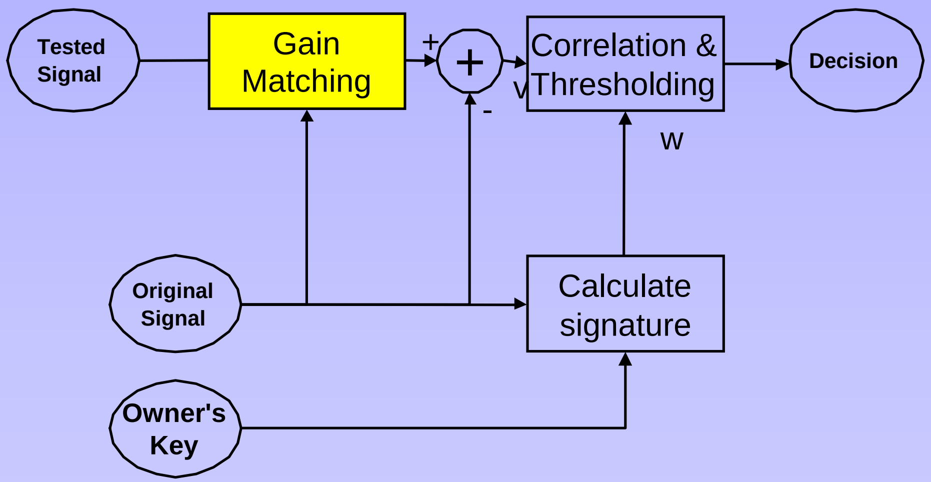
Sophisticated attacks:

- Non-linear transforms
- All-pass filters
  - Fixed and time varying
- Equalization
- Noise
  (White or Colored – perceptually based)
- Time scale modification
- Echo
- …

These attacks should not decrease the audio quality!

Gain Problem

Solved by finding the gain value using Tested and Original signals. Using this value we can correct the tested signal.



Copy part of a file / Offset

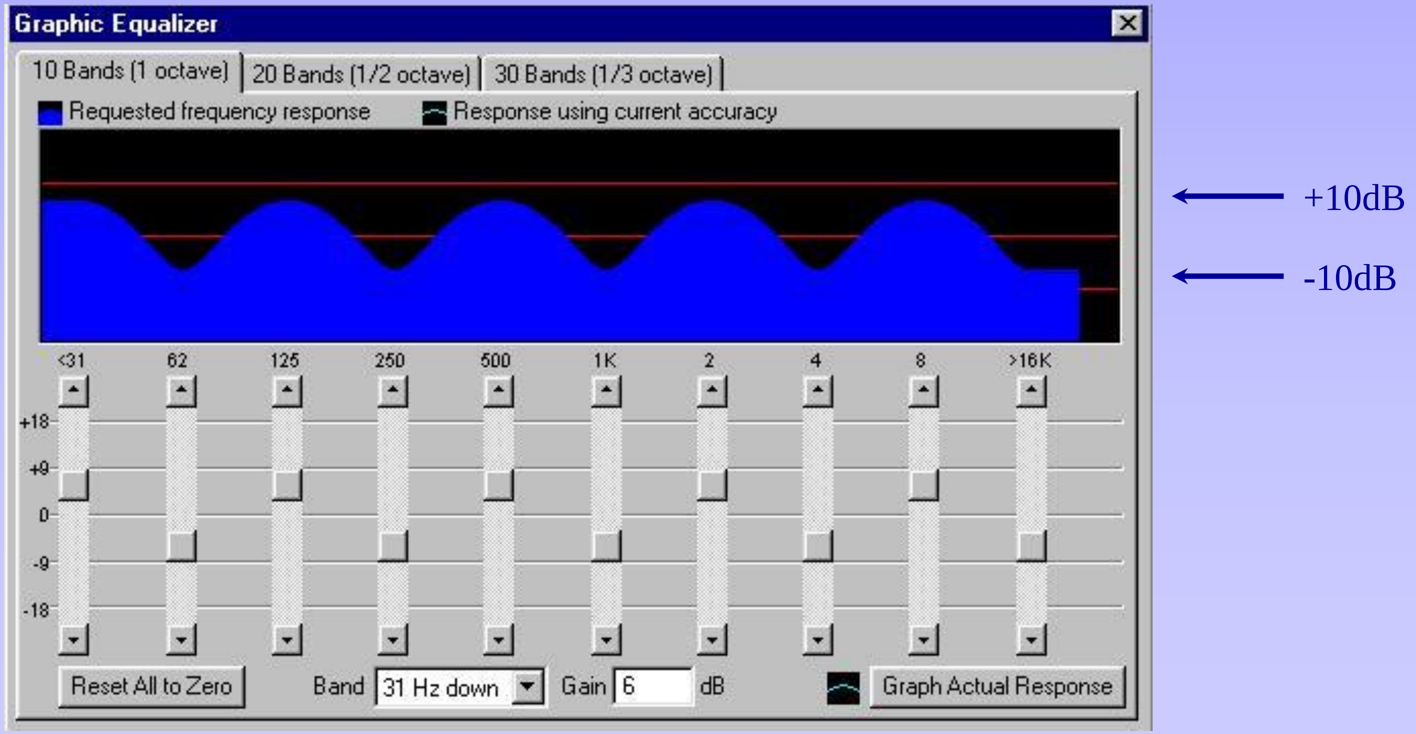
Solved by finding the number of samples offset between the Tested and Original signals.

This could be done using cross-correlation (searching for maximum value of cross-correlation as a function of the samples index.)

Equalization

Soft equalization does not cause a significant reduce in the correlation value.

Example of equalization using Cool-Edit Pro utility:



Compression

We examined the effects of MP3 compression.

The equivalent white noise SNR appears below.

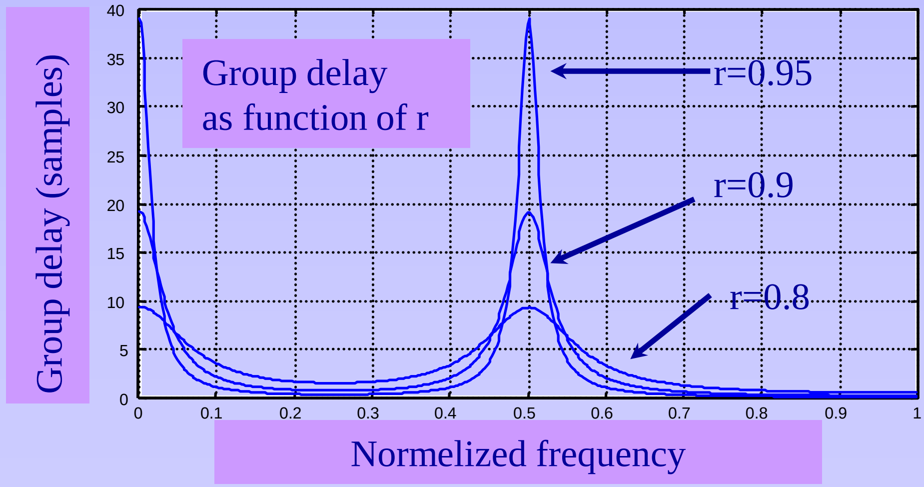
Compression rate	Similarity result	Correlation result	Equivalent SNR
128K	0.92	0.9	~30dB
96K	0.88	0.76	~20dB
64K	0.76	0.49	~15dB

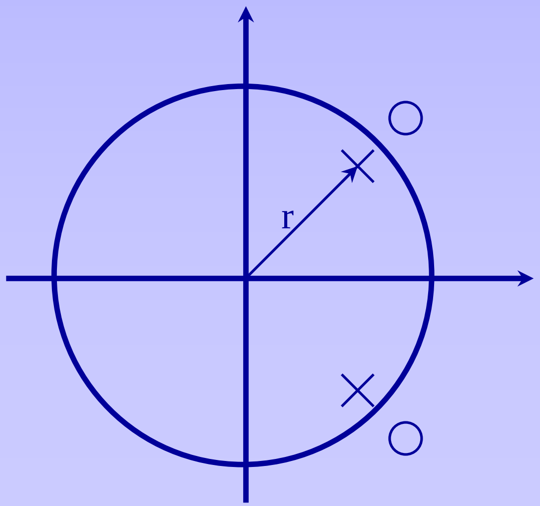
Handling Sophisticated Attacks

- All-pass filter
- Time varying all-pass filter
- Non-linear process
- Combined time varying all-pass filter and non-linear process

Fixed All-Pass Filter

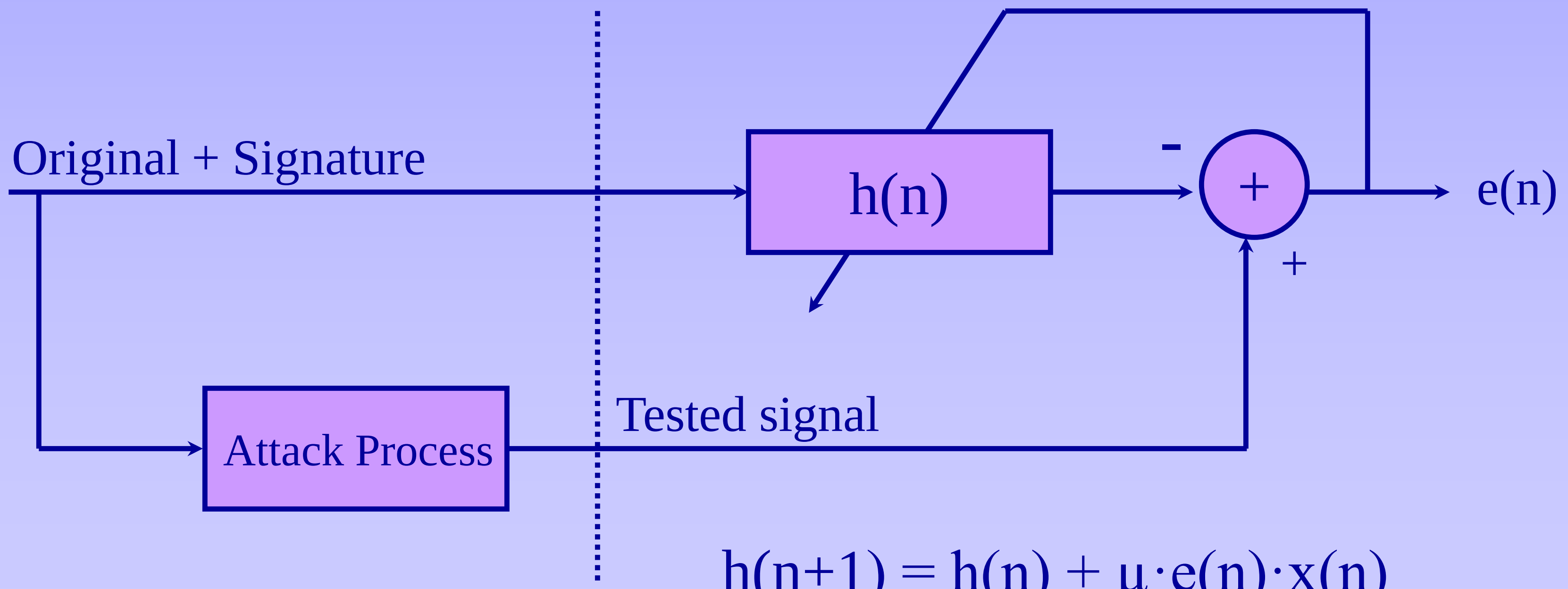
- Does not reduce the audio quality.
- The closer r to 1 the bigger the group delay.
- Example: All-pass filter with poles at: 0.9, 0.9i, -0.9i, Reduce to correlation metrics from 1 to about 0.5 .





All-Pass Filter - Solution

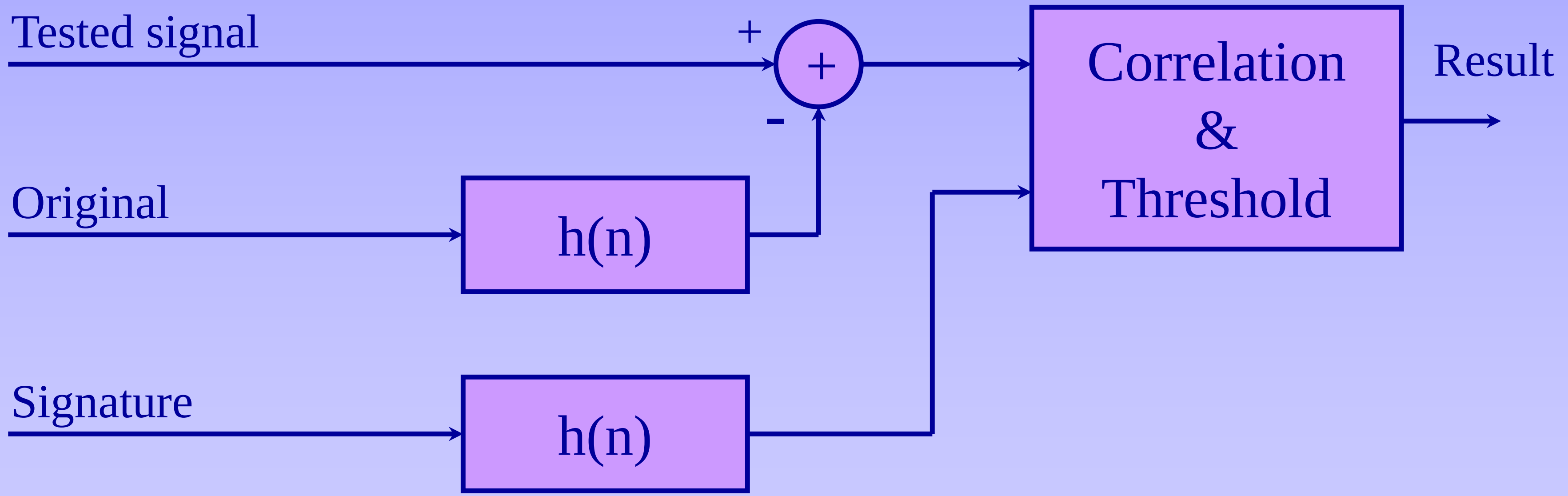
- Using the known watermarks signal (original + signature) find an FIR filter that matches the filter response.
- Could be done by using the LMS algorithm.



$$h(n+1) = h(n) + \mu \cdot e(n) \cdot x(n)$$

All-Pass Filter – Solution (Cont’d)

- Now, when the attack process is estimated using h(n) use the next system for finding the correlation value.



All-Pass Filter – NLMS Solution:

How to set μ?

Instead using LMS we are using Normalized LMS (NLMS)!

$$h(n+1) = h(n) + alpha \cdot e(n) \cdot x(n) / (|x(n)|^2)$$

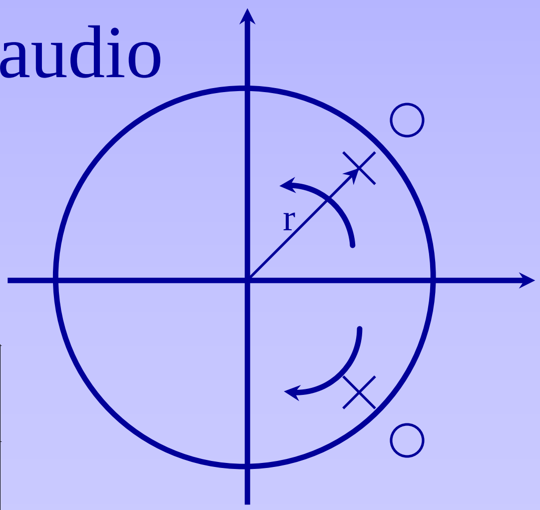
$$0.1 < alpha < 2$$

For the related example:
Using the NLMS system we improved the correlation value from ~0.5 to ~0.9 !

Time Varying All-Pass Filter

- Same as before but now the poles and zeros locations vary in time.

- The rate of change was determine by listening to the audio signal.
  The maximal rate depends on r and the audio signal characteristics.



Typical minimum cycle duration (samples):

	r=0.8	r=0.9	r=0.95
Classica1 music	10000	25000	40000
Vocal music	2000	3000	8000

Time Varying All-Pass Filter - Solution

Can apply NLMS process but this must be done on short segments.

Segment length should be short relative to the variable filter cycle duration but should be longer than NLMS convergence time.

The Correlation should be computed on short segments as well.

Example: Time varying all-pass filter (r=0.9):
(Using repetitive process on 500 samples segment)

Cycle duration (samples):	40000	30000	25000	20000
Correlation value (NLMS):	0.91	0.95	0.65	0.58
Correlation value (RLS):	0.99	0.99	0.7	0.63

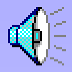
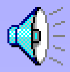
Non -Linear Distortions

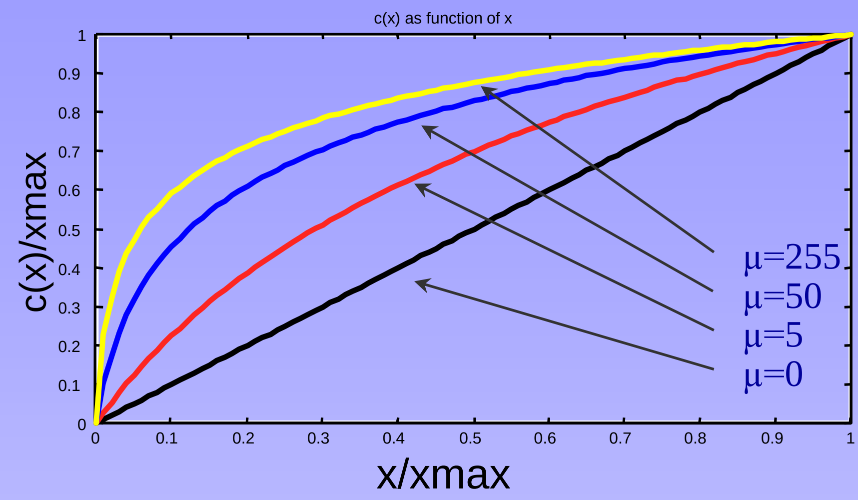
- Clipping or Central Clipping (adaptive and non-adaptive type) distortions effects were examined for both media quality reduction and Correlation value reduction.
- Correlation value at the point where the distortion is not heard is not reduced significantly.

Distortion	Hearing point	Correlation value
Clipping	0.9	0.99
Adaptive Clipping	0.8	0.99
Center Clipping	0.038	0.99
Adaptive Center Clipping	0.1	0.98

μ-Law Non -Linear Distortion

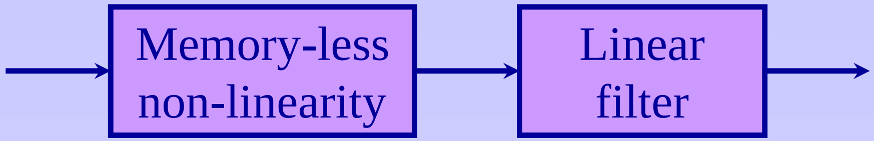
We used μLaw distortion with μ=5.
(the hearing threshold is μ=~3.)

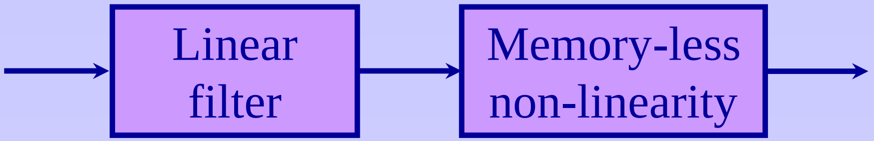
Original:  μ=5: 



- This distortion reduces the correlation measure to ~0.95
- Adding this non-linear distortion before/after all-pass filtering reduces the correlation to 0.36/0.38 (μ=5.)
- The solution: insert an non-linearity into the distortion model.

The new distortion models:

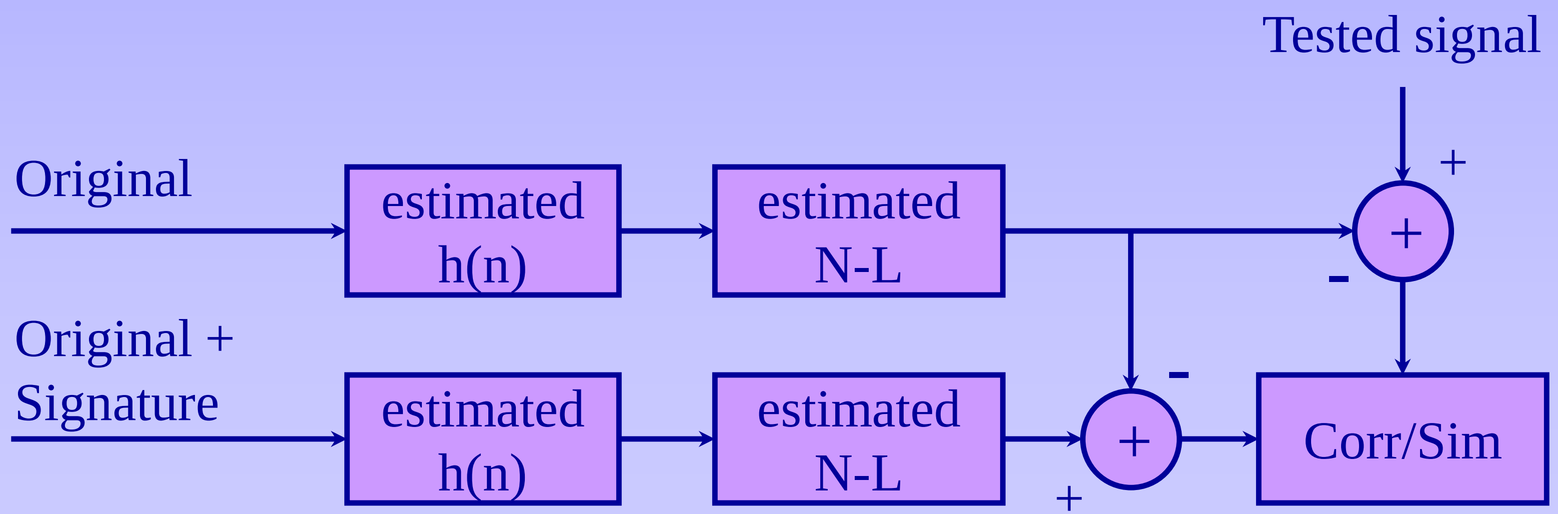




Handling Linear Filtering Followed By Non-Linearity - Detection system

Assumption:

The N-L is almost linear for short segments



Possible Model Estimation

- Volterra series method:

  Instead using x(1),…,x(N) in the standard LMS process, estimate the linear filter with inputs:

  $x(1),\ldots x(N), x(1)\cdot x(1), x(1)\cdot x(2),\ldots, x(1)\cdot x(N), x(2)\cdot x(1), x(2)\cdot x(2),\ldots x(N)\cdot x(N)$

  - For N linear filter coefficients in normal LMS we get $N+N^2$ coefficients in the Volterra method.

  - Convergence rate in much slower for the Volterra method compared to LMS !

  - Special problem when dealing with a time varying all-pass filter.

Proposed solution

Assumptions:

- The non-linearity is not high
- Anti-symmetric

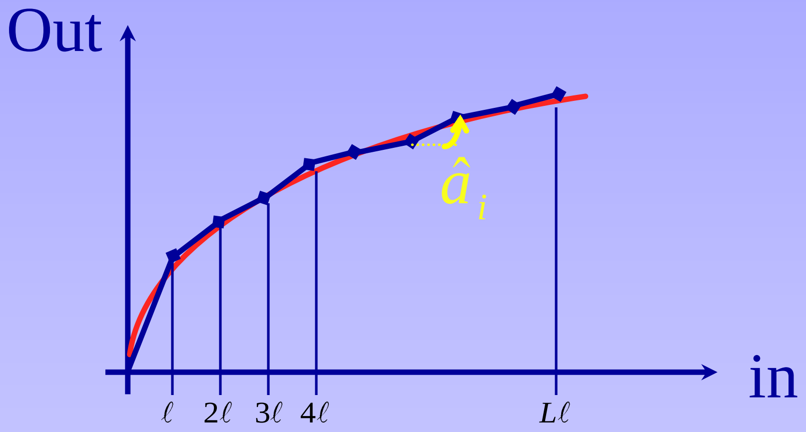
Model: piece-wise linear approximation.

Non-Linearity Piece Wise Approximation

$$out = \hat{f}\left(in\right)$$

Where:

$$\hat{f}\left(x\right)=\begin{cases}\hat{f}_1\left(x\right) & 0\le x\le \ell \\ \hat{f}_2\left(x\right) & \ell\le x\le 2\ell \\ \vdots & \\ \hat{f}_L\left(x\right) & \left(L-1\right)\cdot\ell\le x\le L\cdot\ell\end{cases}$$



$$\hat{f}_i\left(x\right)=\hat{a}_i\cdot\left(x-\left(i-1\right)\cdot\ell\right)+\ell\cdot\sum_{k=0}^{i-1}\hat{a}_k \qquad \left(\hat{a}_0=0\right)$$

Non-Linear Estimation

Two methods:

- Using LS criterion and (x,y) pairs to estimate the slopes coefficients.
- Adaptive system for estimating the line slope using sample by sample adaptation.

Non-Linearity Estimation – LS Criteria

For each segment calculate $\hat{a}_i$ using the *i-th* segment samples: $(x, y) \in \{(x, y) : (i-1)\cdot\ell \le x \le i\cdot\ell\}$

Begin with $\hat{a}_1$, continue with $\hat{a}_2, \hat{a}_3, \ldots, \hat{a}_L$.

Note: D_i depends on $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_i$

Distortion for the i-th segment:

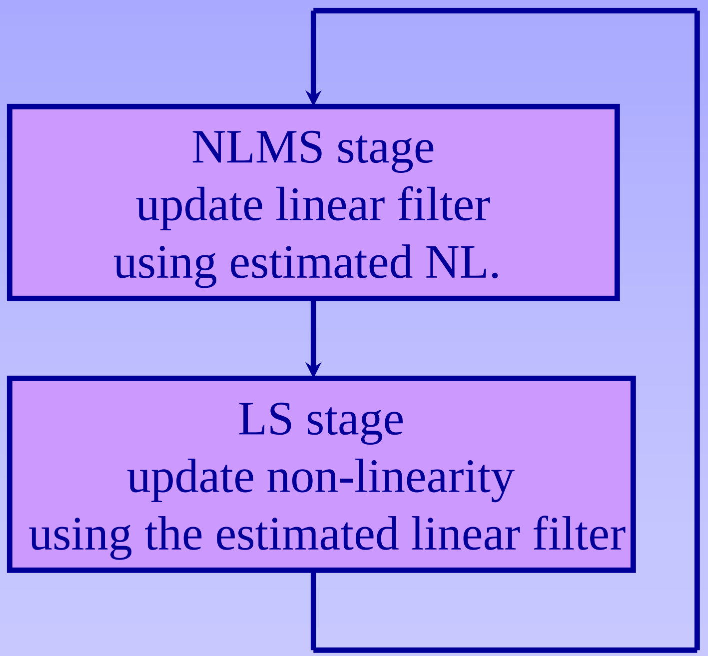
$$D_i = \sum_{(x,y)\in\{(x,y)\}_i} \left(\hat{a}_i \cdot \left(x - (i-1)\cdot\ell \right) + \ell \cdot \sum_{k=1}^{i-1} \hat{a}_k - y \right)^2$$

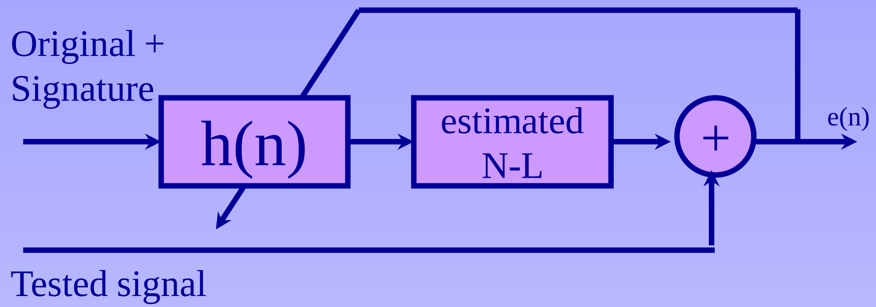
gives:

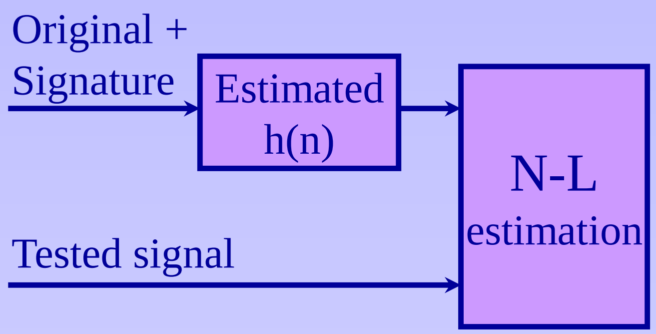
$$\hat{a}_i = \frac{\sum_{(x,y)\in\{(x,y)\}_i} \left(y - \ell \cdot \sum_{k=1}^{i-1} \hat{a}_k \right) \cdot \left(x - (i-1)\cdot\ell \right)}{\sum_{(x,y)\in\{(x,y)\}_i} \left(x - (i-1)\cdot\ell \right)^2} \qquad i = 1, \ldots, L$$

Estimating The Complete Distortion Model (LS Criterion)

- Two stage process:

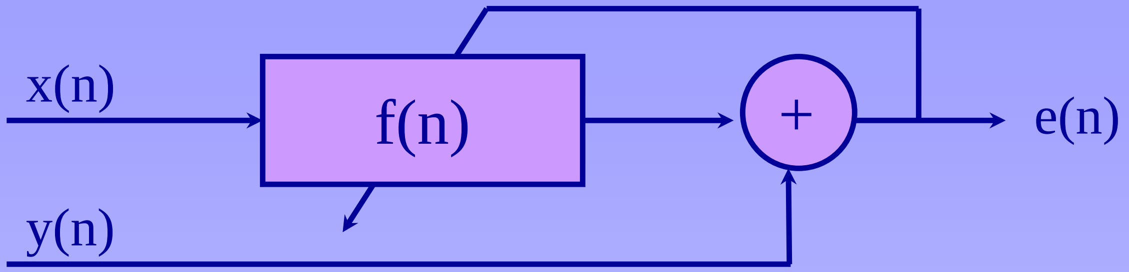






Typically, convergence is achieved in 2-3 iterations

Non-Linearity Estimation – Adaptive Method



For each sample pair *x(n)*,*y(n)* update $\left\{\hat{a}_i\right\}_{i=1}^{L}$ as follows:

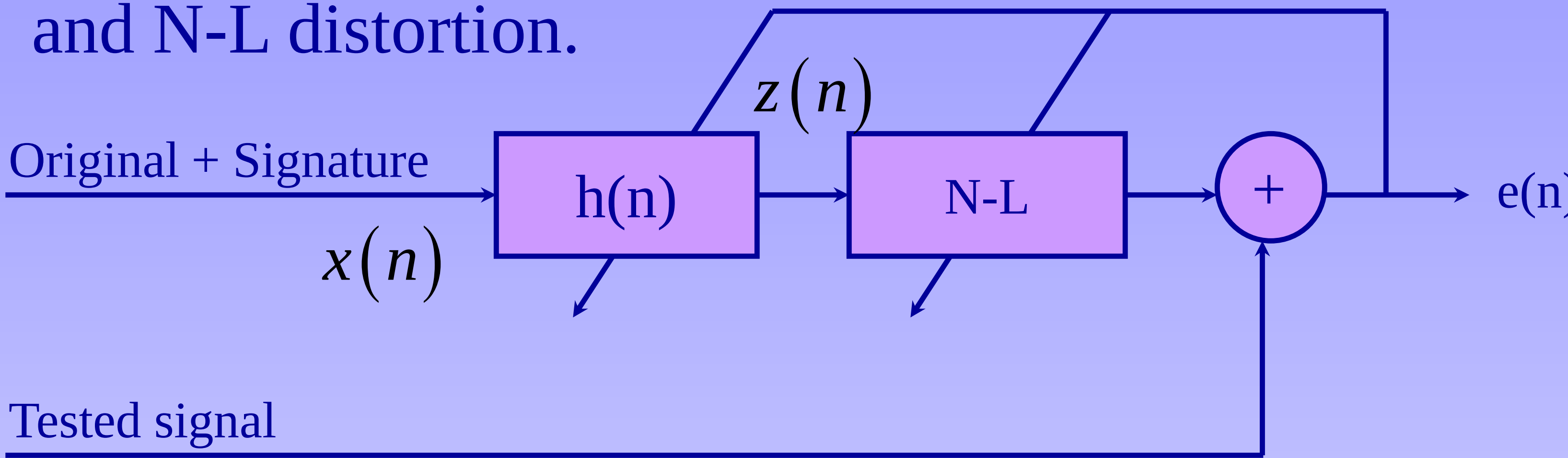
$$\hat{a}_j^{m+1} = \hat{a}_j^{m} + \begin{cases} \mu \cdot e(n) \cdot \left(x(n) - \left(i\left(x(n)\right) - 1 \right) \cdot \ell \right) & \text{if } j = i\left(x(n)\right) \\ \mu \cdot e(n) \cdot \ell & \text{if } j < i\left(x(n)\right) \\ 0 & \text{elsewhere} \end{cases}$$

$$j = 1, \ldots, L$$

When *i(x)* denotes the segment that includes *x*.

Estimating The Complete Distortion Model (Adaptive Method)

- Use NLMS method to estimate both linear filter and N-L distortion.

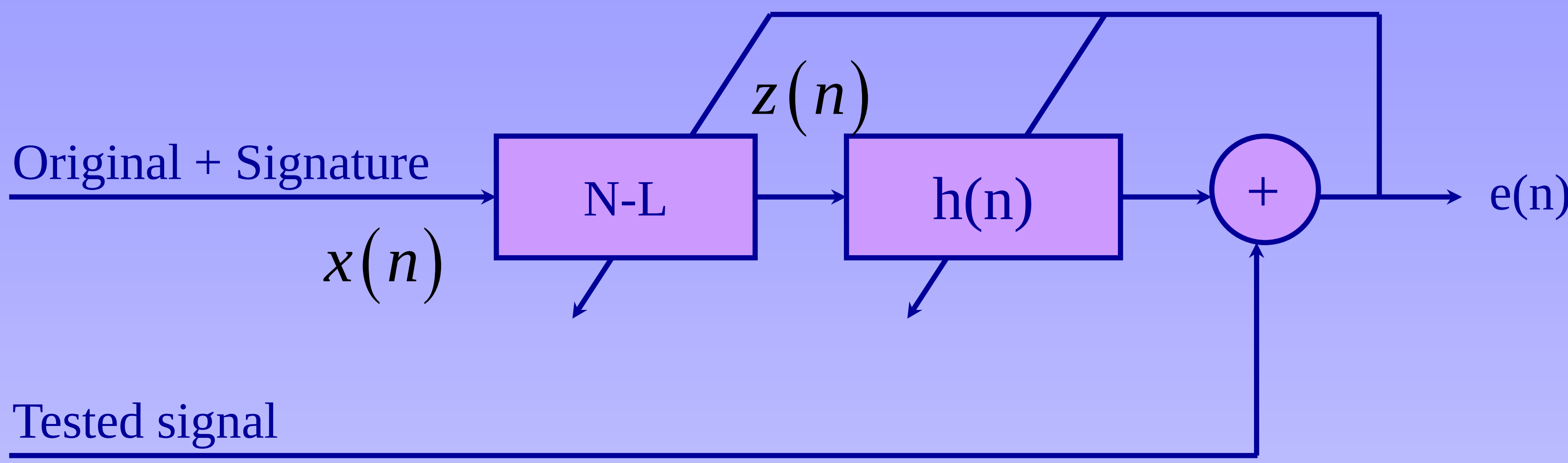


$$\hat{h}_j^{m+1} = \hat{h}_j^{m} + \mu \cdot e(n) \cdot x(n-j) \cdot \hat{a}^{m}_{i(z(n))} \qquad j = 1, \ldots, N$$

$$\hat{a}_j^{m+1} = \hat{a}_j^{m} + \mu \cdot e(n) \cdot \begin{cases} \left(z(n) - \left(i(z(n)) - 1\right) \cdot \ell\right) & \text{if } j = i(z(n)) \\ \ell & \text{if } j < i(z(n)) \\ 0 & \text{elsewhere} \end{cases}$$

$$j = 1, \ldots, L$$

Estimating The Complete Distortion Model (Adaptive Method)



$$\hat{h}_j^{m+1} = \hat{h}_j^{m} + \mu \cdot e(n) \cdot z(n-j) \qquad j = 1,\ldots,N$$

$$\hat{a}_j^{m+1} = \hat{a}_j^{m} + \mu \cdot e(n) \cdot \left[\sum_{k=0}^{N-1} \hat{h}_k^{m} \cdot \begin{cases} \left(x(n-k) - \left(i\left(x(n-k)\right) - 1 \right) \cdot \ell \right) & \text{if } j = i\left(x(n-k)\right) \\ \ell & \text{if } j < i\left(x(n-k)\right) \\ 0 & \text{elsewhere} \end{cases} \right]$$

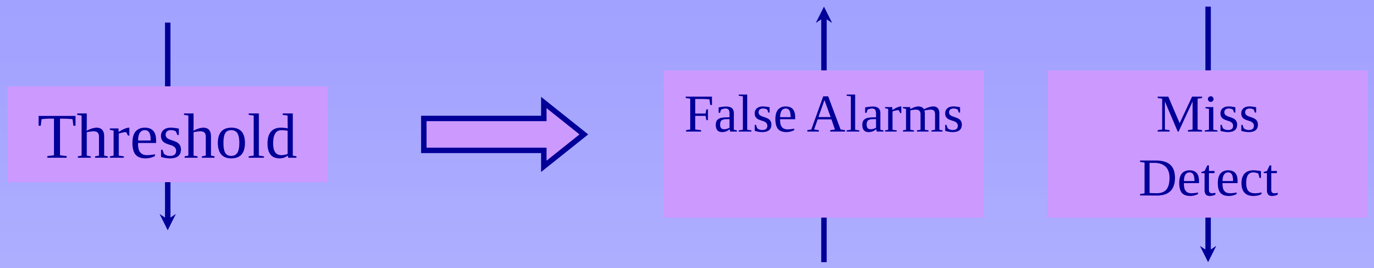
$$j = 1,\ldots,L$$

Comparison between LS method and Adaptive Method

- Both methods give about the same results for linear filtering followed by NL.
- The adaptive method can handle NL followed by linear filtering while LS is expected to have problems approximating an inverse of the linear filter.

Finding Detection Threshold Value

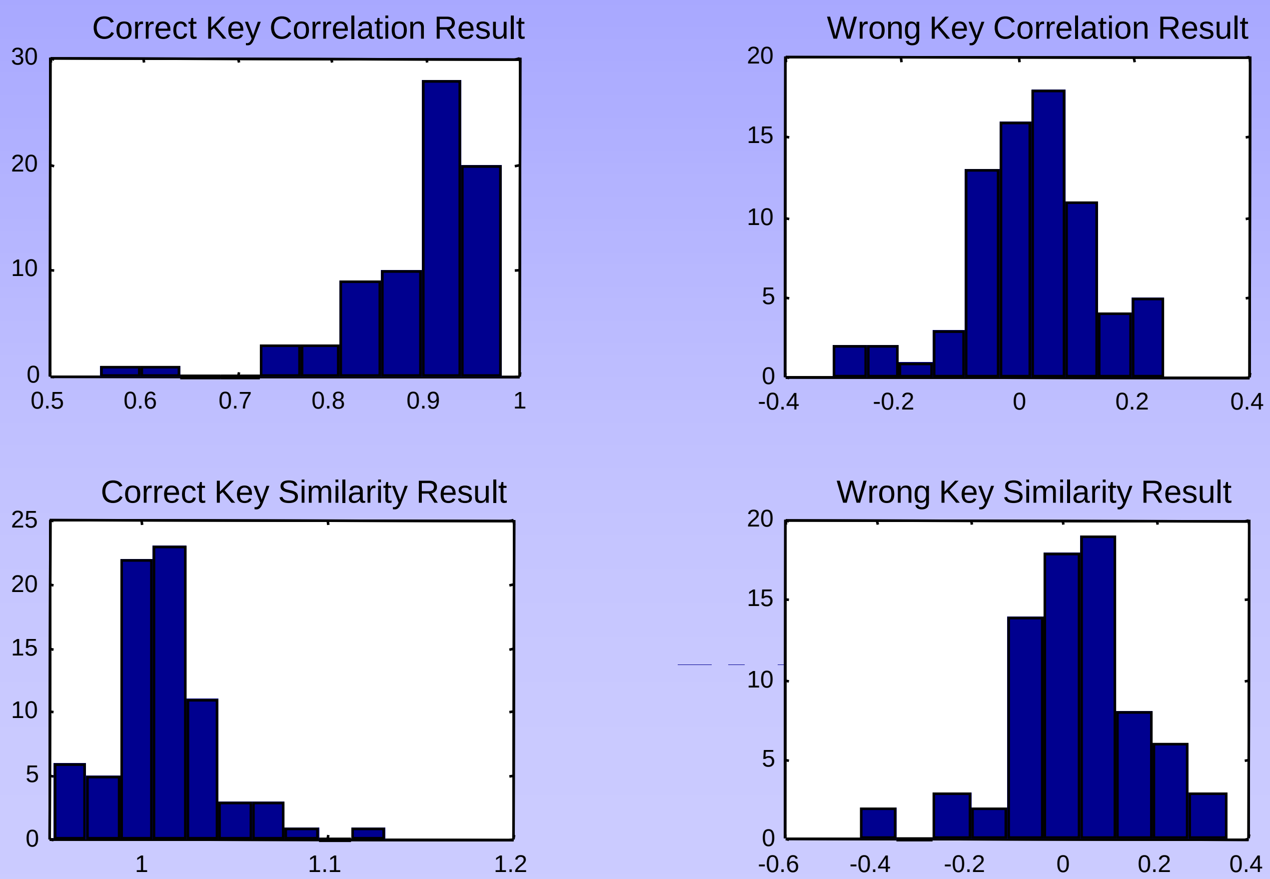
- False Alarms Vs. Miss Detection



- Purpose: find the detection threshold for a given false alarm rate.
- Influenced by Correlation segment length.
- Can be done using histogram analysis for both cases, False Alarms and Miss Detection.
- Threshold for Correlation found to be 0.5 for segment length of above 2000 samples. It gets 0 False Alarms in all of our tests and Correct Detection in more than 95% of segments.

Finding Detection Threshold Value (Cont'd)

Example: Histograms result for segment length of 2000 samples



Summary and Conclusion

- We implemented a watermarking embedding and detection system and found the detection performance under common types of attacks.
- We searched for attacks limited in the sense of preserving the audio quality and focused in these that reduce detection performance.
- We defined the attacker's global model that combines NL and linear filtering, estimated its parameters using system identification methods (including LS, Normalized-LMS and RLS) that use both the tested signal and the reference signal.
- We determined detection threshold to meet false alarms rate requirement for a give correlation segment length.

Further Work

- Handling linear speed change and time scale modification:
  May be handled by doing time warping calculation using dynamic programming.
- Handling echo addition:
  May be handled by defining an echo model and estimating its parameters.
- Low quality compression.