The effective duration of the autocorrelation function of a sound signal: calculation methods, relationship with cognitive models and relevance on the subjective preference theory

Doctoral Thesis of Dario D'Orazio

Introduction

In this work various aspects, concerning the effective duration of the autocorrelation function of a sound signal, are treated. The effective duration (τ_e) of the autocorrelation function (ACF) has been proposed by Ando as one of the fundamental factor to evaluate the listening process in concert halls. The Ando model is a complex synthesis of different theories confirmed by a rigorous experimental approach. One of the aims of this work is to present with the necessary rigor the different studies which have been combined in the Ando system. A history of the autocorrelation-based models of auditory perception is proposed in chapter 1. When it is possible in compact form, the analytical formulations are presented, to show the assumption and simplifications made in the progressive system formulation. The recent models that used an autocorrelation-based schema are resumed in the last paragraph of chapter 1, to confirm the plausibility of the use of this model at the present time. The monaural criteria used in room acoustics evaluation are resumed in chapter 2, highlighting the progressive role of the envelope of ACF as a fundamental descriptor. The room criteria based on τ_e analysis are proposed in paragraph XX and the extension of the role of the temporal factors to other fields of acoustics is discussed in the last paragraph of chapter 2. The envelope extraction methods that can be found in current literature are discussed in chapter 3. The use of the decay parameter estimation for ACF, as proposed in /citeAndo89, is formalized in par. XX, comparing it to the decay parameter of the room impulse response. An alternative approach to the envelope detection, based on the statistical decision theory, is discussed in par. XX. Some notes about the reference of this theory and about the neurophysiologic characteristics of the auditory process (assumed by a real

observer) are resumed in chap. XX. The role of the envelope in some auditory models and the differences with the Ando's one are discussed in par. XX. An extended numerical evaluation of the decay parameters of the ACF is presented in chap. 4, with reference to the previous literature. A critical analysis of these methods, based on numerical results and mathematical considerations obtained in the frame of this work, is presented. A novel extraction method of τ_e is proposed in chapter XX. The algorithm proposed is discussed with reference to its neurophysiologic plausibility and the assumed statistical knowledge of the sound signal. In order to extend the use of τ_e analysis to other fields of acoustics a good performance in presence of low SNRs is required. The use of the d-SACF instead of the A-weighting ACF is discussed in chap. XX. A software ad hoc is then presented in its beta version. Finally a case study is presented: the 'Torri dell'acqua' auditorium, in Budrio. The acoustics of the auditorium has been designed in 2008 by the author of this work, following the criteria of subjective preference. The normal factors, in relation to the reference motif, have been calculated and are shown here. In particular, a new method to evaluate the Δt_1 is used for the first time. Being this auditorium a general purpose one, the use of dynamical DSP for loudspeaker systems may have sense inside it. Some consideration about a potential real-time optimization of the subjective preference are proposed in the last paragraph of chapter XX.

Chapter 1

Autocorrelation-based perception models

1.1 Background influences

1.1.1 Helmoltz and von Bekesy

Two dominant influences inspire the the birth of perception models in the 40's years: the Helmoltz's resonance-place theory and the von Bekesy's work. Helmoltz allocates each elementary sinusoidal oscillation, in which every stimulus can be decomposed, to its proper position along the length of the basilar membrane. Allocation determines the subjective pitch, the pitch being in one-to-one correspondence with the allocated place. We should note that Helmoltz's formulation [**?**] suggests a system in which a single compound signal is broken up into elementary parts and these elementary parts are transmitted in separate channels, spatially distinct from one to another. This conception was firstly proposed by Du Vernay [**?**] who called *tonotopical* this approach.

Von Bekesy [**?**] shows that the hydrodynamic action of the cochlea does indeed distribute different frequencies to different locations along the basilar membrane and that this effect is not very sharp. A sinusoidal oscillation of any given frequency, however, excites the vibration of a large part of the basilar membrane. The distribution of vibration amplitude has a maximum

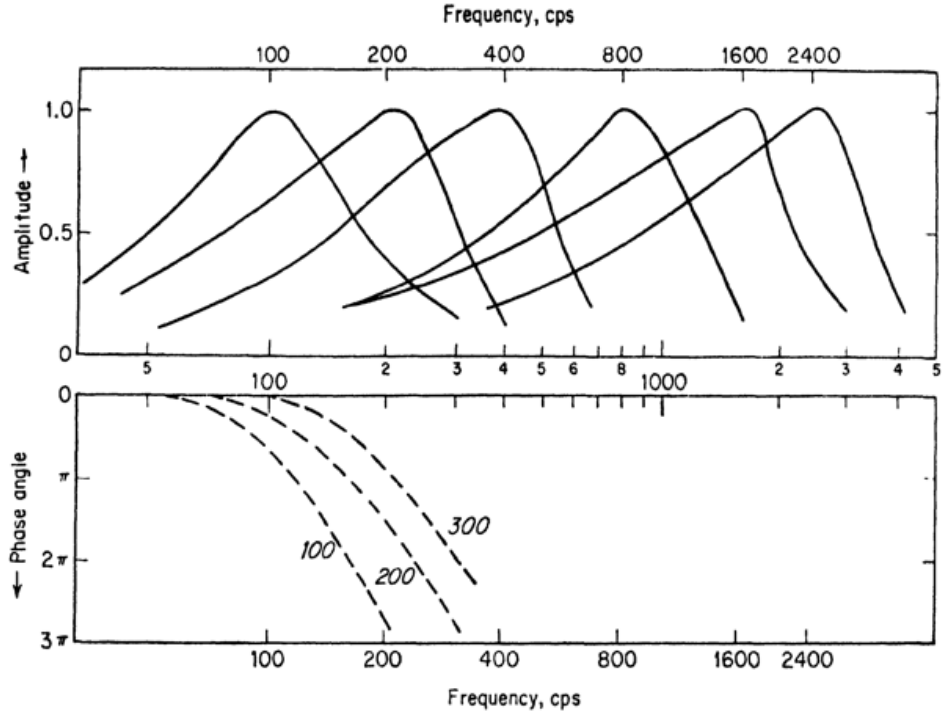


Figure 1.1: Amplitude and phase of the hydrodynamic frequency analysis performed by the cochlea determined by von Bekesy. from [33]

but, instead of Helmoltz's hypothesis (one single resonant string vibrating and its neighbors quiescent), the distribution is broader. This dullness of the mechanical frequency analysis (an equivalent of blurring in vision) changes in the '30s the conception of the auditory system: instead of discrete and distinct channels the von Bekesy's theory proposes a continuum of channels, broad and overlapped.

1.1.2 The statistical decision theory

The von Bekesy's cochlear analysis is essentially hydrodynamic and the relative formulations involves a combination of physics, anatomy, physiology, acoustic, optics and electronics. In the later 40's the rising autocorrelation-based perception model needs borrows from it and another contributions taken from mathematical statistics: the terms and the ideas of *ideal observer*, *real observer*, *threshold of perception*, *false alarm probability* and *payoff ma-*

Response	Stimulus	
	s	$\bar{s}$
S	V_{sS}	$V_{\bar{s}S}$
$\bar{S}$	$V_{s\bar{S}}$	$V_{\bar{s}\bar{S}}$

Figure 1.2: Payoff Matrix (from [33])

trix are defined. The ideas of statistical theory of decision became necessary in the second world war in target detection by radar systems. In acoustics these concepts has been developed by Smith and Wilson [**?**] at the M.I.T. Lincoln Laboratory, by Tanner, Swets and Green [63] at the University of Michigan, by Marill [**?**] at the M.I.T. Research Laboratory of Electronics. These models are very closely related: axiomatic the one proposed by Tanner, Swets and Green (TSG), less formal the one by Smith and Wilson (SW) and more tutorial the one by Marill (M). The TSG model is based on detection trials in which a signal, full known by the listener, is either presented (s, with probability $P(s)$) or not presented ($\bar{s}$ with probability $P(\bar{s})$) and a listener gives a response (s or $\bar{s}$) signifying whether or not, according to his judgement, the signal was presented. In the TSG model it is assumed that the listener knows the a priori probabilities $P(s)$ and $P(\bar{s})$ governing signal presentation and that the listener make his decision in relation to a *payoff matrix* (fig. 1.2). The signal presented to the listener (a single narrow frequency band with fixed or swept center frequency) is superposed to the background noise. The listener's decision is governed, according to TSG, by a decision process (a mathematical function of a priori probabilities, payoff matrix and sensory data). In order to connect the mathematical structure of the statistical hypothesis and the 'samples' of signals and noises, a signal space is defined: a n-dimensional space in which a signal is represented by n coordinates. The signal space for signals Δ_t long in time and Δ_f wide in frequency is a complex space and, consequently, there are $n = 2\,\Delta_t\,\Delta_f$ numbers to specify uniquely a periodic wave. In the TSG theory the background noise is approximately a gaussian-process noise, distributed in this space as

an $n-$dimensional gaussian density sphere, centered at the origin. The signal (and its related noise) is another $n-$dimensional gaussian density sphere, not centered at the origin. A waveform presented to the listener corresponds to a particular point P_j in the space. The listener's perception can be described in terms of probability function, as shown in fig. 1.3. The dimension n of the perception space varies, but with some hypotheses related to the Shannon information theory a dimension $n = 4$ may be fixed [**?**].

1.2 The Licklider's work

1.2.1 The duplex theory of pitch perception

In 1930 Wiener shows [**?**] the relation between the Fourier transform of power spectra and the autocorrelation function. The operations involved in carrying out the autocorrelational analysis are different from those involved in making the frequency analysis. The interest of acousticians, using until that time only the frequency analysis, is focused on the neurophysiological studies (temporal and statistical analysis of neuronal activity).

In 1942 Licklider writes his doctoral thesis on frequency localization in the auditory cortex of the cat [31], then he iworks at the Acoustics Laboratory of M.I.T. on speech perception in telecommunications. Following the Helmoltz's resonance-place theory [**?**] and the Fletcher's space-time pattern theory [], Licklider proposes the duplex theory of pitch perception in the 1951 [32]. The essence of the duplex theory is that the auditory system employs both frequency analysis and autocorrelation-based analysis. The frequency analysis is performed by the cochlea, the autocorrelation-based analysis by the neural part of the system. The latter is an analysis of trains of nerve impulses: an highly nonlinear process of neural excitation intervenes between the two analyses. If the lengthwise dimension of the uncoiled cochlea is designated as the x-dimension, the cochlea transforms the stimulus time function $f(t)$ into a running spectrum $F(t, x)$, position x being the neural correlate of stimulus frequency. The running spectrum, a spatial array of of time functions, is transmitted by the neurons of the auditory nerve. Each $F(t, \xi)$ is

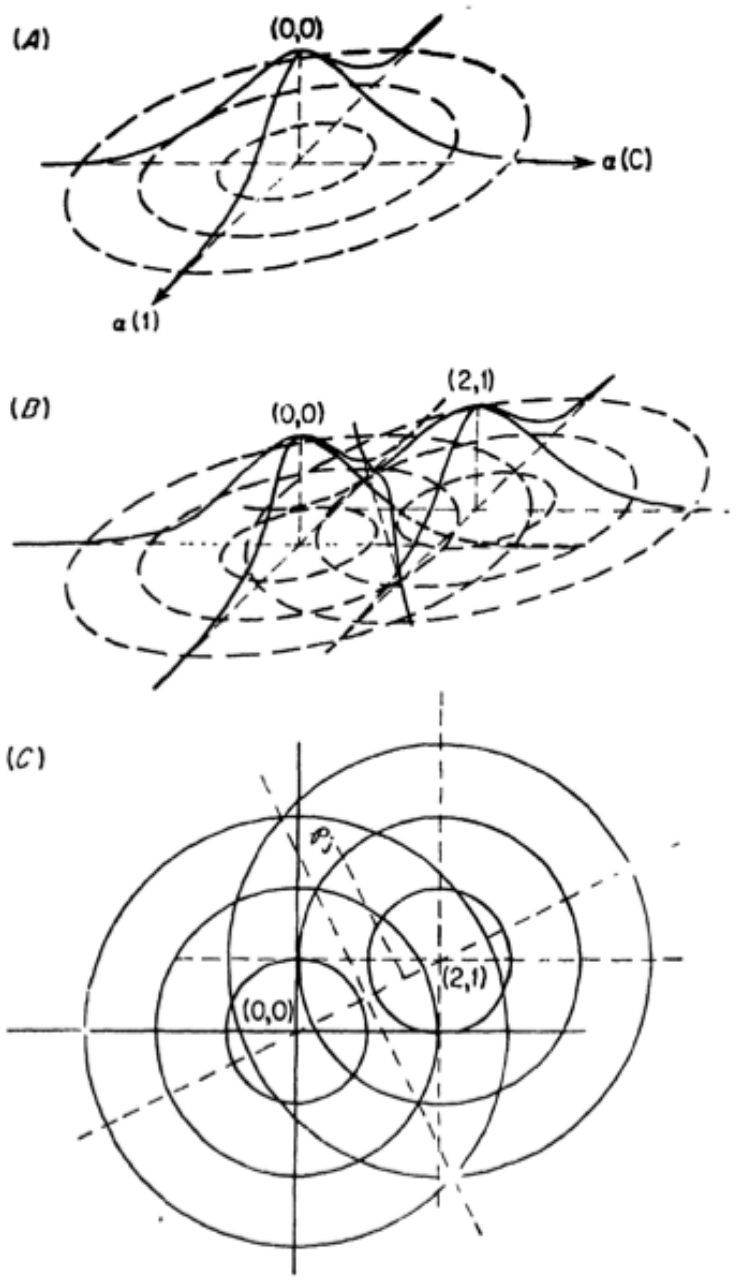


Figure 1.3: Representation in signal space. (A) Probability density function of the noise alone (0,0). (B) Probability density in case of noise (0,0) and signal plus noise (2,1). (C) Iso-probability curves, with an observator represented by the point P_j. Only two dimensions for the sake of simplicity are plotted. (from [33])

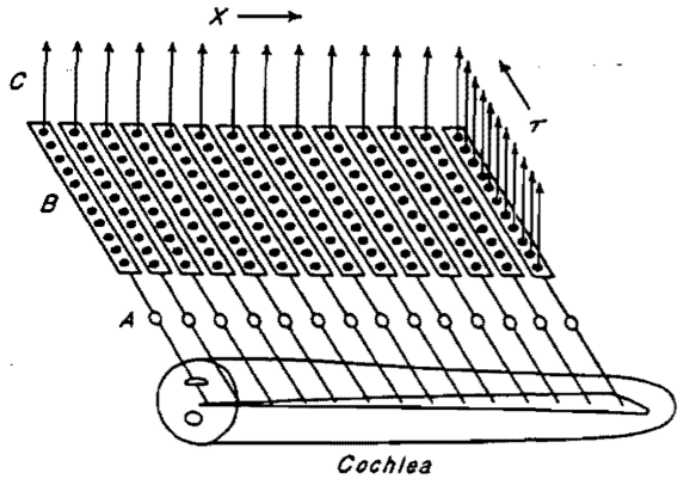


Figure 1.4: Schematics of Licklider's triplex theory of pitch perception. From [32]

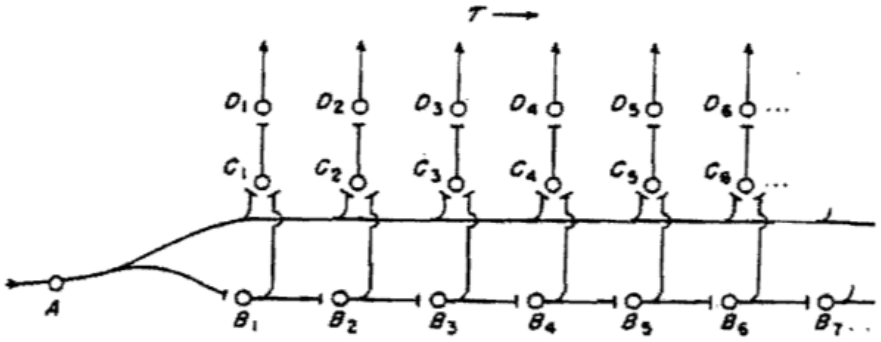


Figure 1.5: Basic schema of neuronal autocorrelator. From [32]

an integral over the behaviors of many neurons (fig. 1.4) A basic schema of neuronal autocorrelator is proposed in (fig. 1.5) A is the input neuron, B_i is the delay chain. The original and the delayed signal are multiplied with A and B_k feed C_k. A running integral of the product is obtained at the synapse between C_k and D_k, where excitation accumulates whenever C_k discharges and dissipates itself at a rate proportional to the amount accumulated. The excitatory states D_i provide a display of autocorrelation function of the input time function, the temporal course of discharge of A.

If $N_{ij}(t)$ is the formalization of the contribution an individual neuron (the $j-$th neuron in the group i) then $N_{ij} = 0$ if the neuron is quiescent or $N_{ij} = 1$

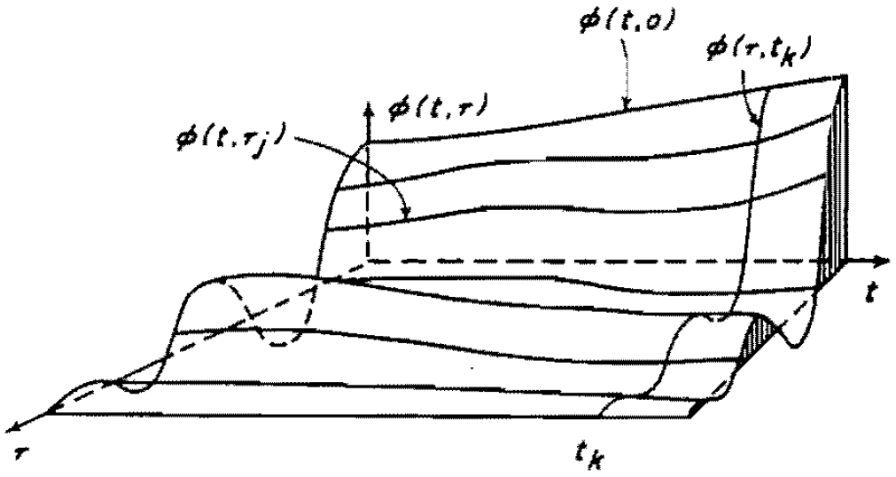


Figure 1.6: Running autocorrelation function (r-ACF): surface $\phi(t, \tau)$. From [32]

if it fires. The input functions $N_{ij}(t)$ are delayed by a variable τ $N_{ij}(t-\tau)$ and the running accumulation $\Phi_{ij}(\tau)$ is the product of the input function and the delayed one. By integration over a time interval $2T$ the running integral of the latter accumulation has the form of an autocorrelation of the signal in the $i-$th channel:

$$\Phi_i(\tau, x_i) = \lim_{2T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t, x_i) f(t+\tau, x_i)\, dt \tag{1.1}$$

Let's note that the Licklider's model is multi-channel (Φ is dependent from the place x) and that the neurophysiological plausibility of the model is confirmed by a Licklider's note: "The autocorrelational analysis must operate only for frequencies that can be represented by volleys in the first order neurons" [32]. This fact restricts the autocorrelational analysis to $1KHz$ or less.

Neverthless, the use of Licklider's formulation in the past literature refers to a mono-channel form. For the purpose of the duplex theory (pitch perception) a running integration of the ACF is proposed.

$$\Phi(\tau, t) = \lim_{2T\to\infty} \frac{1}{2T} \int_{t-T}^{t+T} W(\xi)\, f(\xi)\, f(\xi+\tau)\, d\xi \tag{1.2}$$

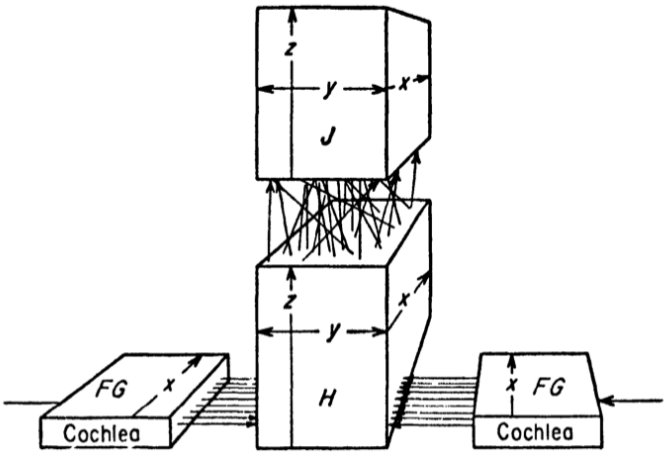


Figure 1.7: Schematics of Licklider's triplex theory of pitch perception (da Licklider 1956)

Where $W(\xi)$ is a window function. An exponential window with a time constant of 2-3 ms is proposed in [32], but this value could be reported to the very low memory capacity of the signal processors in the early '50.

1.2.2 The triplex theory of pitch perception

In 1956, when he is employed at Bolt Beranek and Newman, Licklider publishes an extension of his theory (Triplex theory of pitch perception [33]) in which the analysis proposed in the earlier works are widely discussed and integrated with the new literature.

In the paper [33] the author proposes also an improved version of autocorrelation mechanism. This work of 1956 in some manner anticipates the future auditory models like SACF [**?**]. With reference to (fig. 1.8) the function $\theta_1(x,t)$ is given by the original signal $\theta_0(t)$ filtered through a bandpass filter, according to von Bekesy's curves. At each point along x, $\theta_1(x,t)$ is subjected to rectification. The half rectified function $\hat{\theta}_1(x,t)$ is then smoothed by a low pass filter. The smoothed wave $\theta_2(x,t)$ is the function that controls the excitations of the auditory neurons at x. The successive recoveries of the hypothetical neuron are represented in (fig. 1.8(4)). The timing of the impulses is dependent from the interaction of $\theta_2(x,t)$ (gray line) and

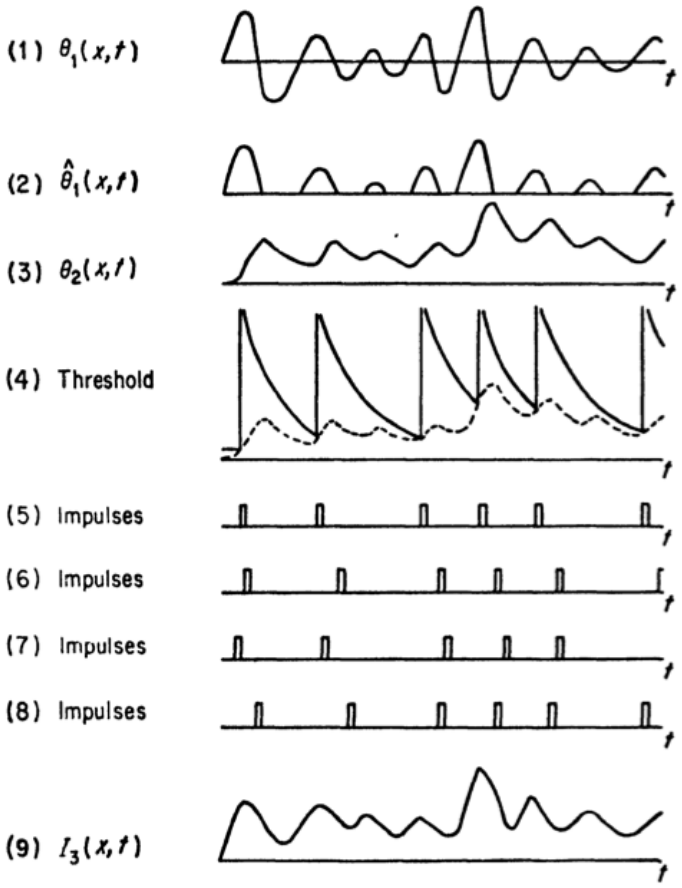


Figure 1.8: Proposal of multi-channel AC. After [33]

the recovery curve (solid line) of the neuron. Each time $\theta_2(x,t)$ exceeds the threshold curve, the neuron discharges (fig. 1.8(5)) and the recovery cycle begins again. In the (fig. 1.8(6-8)) are represented the discharges of the neighboring neurons. $I_3(x,t)$ represents the sum of the impulse curve of many neighboring neurons .

A similar mode, developed from quite different sets of experimental observation, is proposed by Sayers and Cherry [**?**].

1.3 Schroeder's work

1.3.1 An integrable model for the basilar membrane

In [53] Schroeder proposes an equivalent electrical circuit for the basilar membrane (fig. 1.10). The natural logarithm g of the ratio output/input of this circuit (when the output is terminated in its characteristics impedance) is given by:

$$g = -arcsinh\sqrt{\frac{j\omega\, L_1}{R + j\omega\, L + \frac{1}{j\omega\, C}}} \tag{1.3}$$

For $\Delta\, x \to 0$:

$$\lim_{\Delta\, x \to 0} g = -\sqrt{\frac{j\omega\, L_1}{R + j\omega\, L + \frac{1}{j\omega\, C}}} \tag{1.4}$$

Introducing the resonance frequency $\omega_r = \frac{1}{\sqrt{LC}}$, the loss factor at the resonance frequency $\delta(x) = \omega_r RC$ and the phase velocity at low frequencies $c(x) = \frac{\Delta\, x}{\sqrt{L_1 C}}$, eq. 1.4 may be rewritten and integrated from a place x to the basal input ($x = 0$), obtaining:

$$G(\omega, x) = \int_0^x \frac{g(\omega, \xi)}{\Delta\, x}\, d\xi = -j\omega(x) \int_0^x \frac{1}{c(x)\sqrt{1 - (\frac{\omega(x)}{\omega_r})^2 + j\delta\frac{\omega(x)}{\omega_r}}}\, d\xi \tag{1.5}$$

Schroeder needs to postulate some assumptions to integrate eq. 1.5:

- a logarithmic relation between place ξ and the resonance frequency ω_r:

$$\xi = d\ ln(\frac{\omega_m}{\omega_r(\xi)}) \tag{1.6}$$

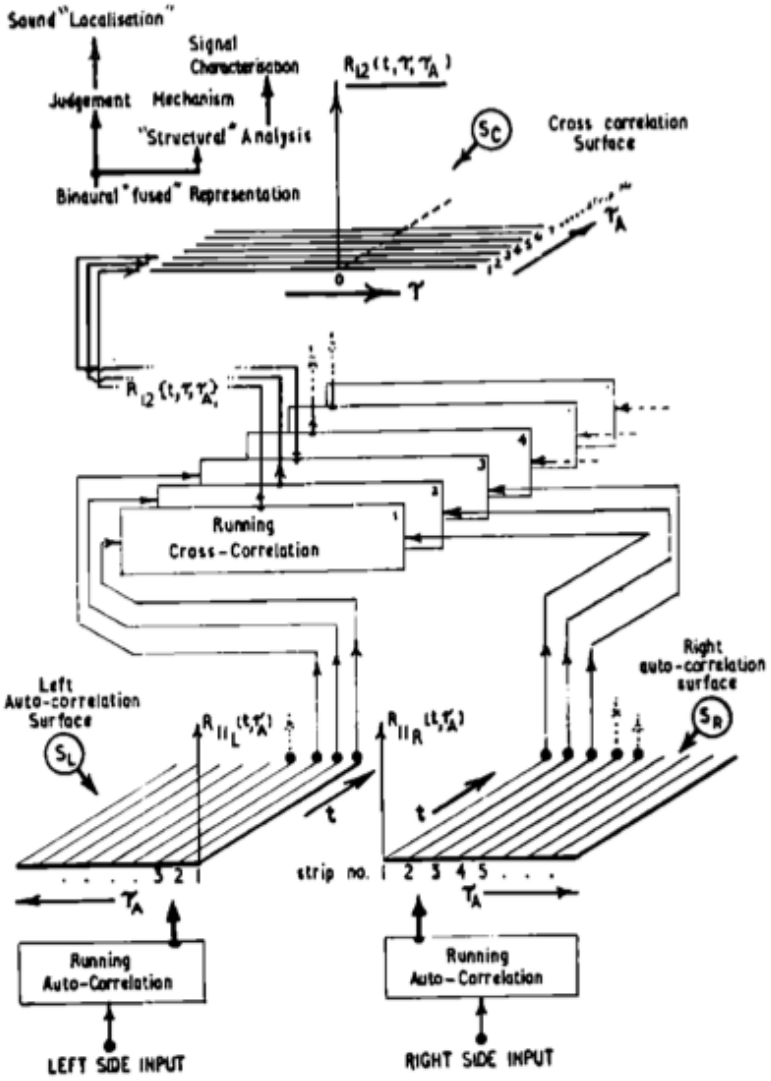


Figure 1.9: Mechanism of Formation of Monaural and Binaural *Gestalten*. After [?]

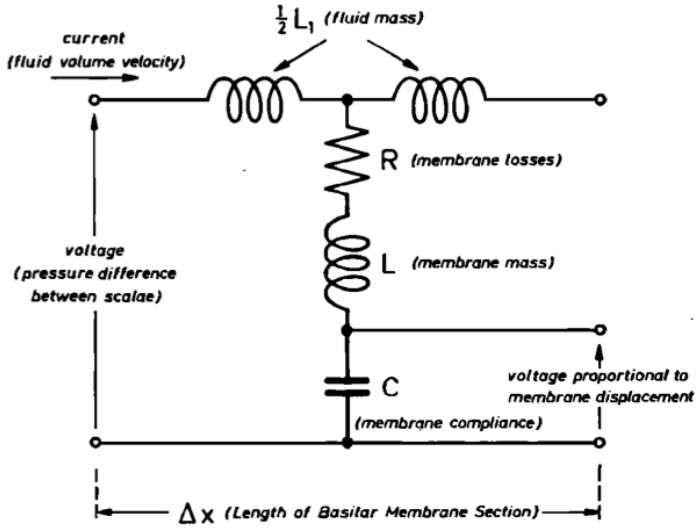


Figure 1.10: Equivalent electrical circuit of the basilar membrane. After [53]

- a proportionality between low-frequency phase velocity $c(\xi)$ and its resonance frequency $\omega_r(\xi)$ for each place ξ:

$$c(\xi) = \frac{\omega_r(\xi)}{k} \tag{1.7}$$

- the independence from place of the loss factor δ:

$$\delta(\xi) = \delta \tag{1.8}$$

With these assumptions the analytic integration of eq. 1.4 is possible and the formal result is:

$$G(\omega, x) \simeq -j\,k\,d(arcsin(\frac{\omega}{\omega_r} - \frac{\delta}{2}) + arcsin(\frac{\omega}{\omega_m} - j\frac{\delta}{2}) \tag{1.9}$$

Let's note that $\frac{k\,d}{2\,\pi}$ is approximately the number of wavelengths on the basilar membrane for radian frequencies small enough compared to ω_m. The solution 1.9 in accordance with the experimental results concerning the phase at low frequency, the amplitude and the phase above resonance, the group delay, the phase velocity and the group velocity. The solution 1.9 permits also to

determine the impulse response of the basilar membrane. For $\omega \ll \omega_r$ the arcsin function can be replaced by its argument. The impulse response of the basilar membrane, with this hypothesis, results as follows:

$$p(t,x) = k\, d\, t^{-1} e^{\frac{-\delta\omega_r}{2}} J_k(\omega_r\, t) \tag{1.10}$$

where J_k is a first type Bessel function of the $k-$th order.

1.3.2 Model for mechanical to neural transduction in the auditory receptor

Schroeder and Hall present in [54] a model for the transduction of the motion of the basilar membrane into spikes in the auditory nerve. The model is based on the process of generation/depletion of electrochemical quanta with three rules:

- quanta are generated in the hair cells at a fixed rate r;
- quanta disappear and cause the firing in the afferent nerve fiber with a probability related to the their number $n(t)$ and a permeability $perm(t)$ related to the instantaneous amplitude to the input signal.
- quanta disappear independently from the stimulation with a probability per unit time equal to $g\, n(t)$, without causing firings in these cases.

In the steady state the rate of generation r of quanta is equal to:

$$r = \frac{1}{T}\int_0^T n(t)\, perm(t)\, dt + \frac{1}{T}\int_0^T n(t)\, g\, dt \tag{1.11}$$

where: r is the number of quanta, $perm$ is the permeability function and $n(t)\, g$ is the probability (per quantum and unit time) of disappearance of a quantum.

The firing probability per unit time $f(t)$ is given by

$$f(t) = n(t)\, p(t) \tag{1.12}$$

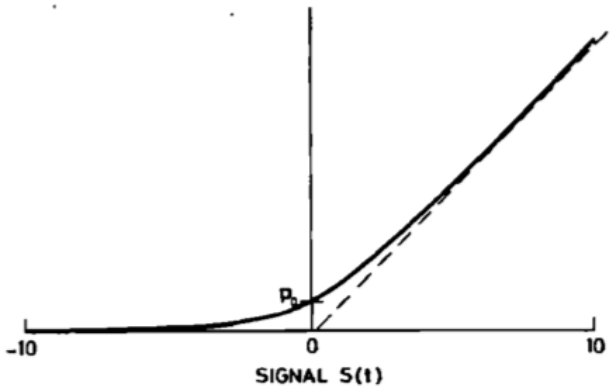


Figure 1.11: Permeability vs Input signal amplitude. From [53]

the permeability function is assumed as a soft half-wave rectifier (see fig. 1.11):

$$perm(t) = perm_0(\frac{1}{2} + \sqrt{1 + \frac{1}{4}s^2(t)}) \tag{1.13}$$

where $s(t)$ is the mechanical stimulation signal from the basilar membrane.

For small signals $perm(t) - perm_0$ is a nearly linear function of $s(t)$, for larger signals $\frac{p(t)}{p_0}$ approaches the half-wave rectified version of $s(t)$. For large signal amplitudes ($(\int p(t)\,dt \gg g)$ and for sufficiently high fundamental frequency of a stimulating periodic signal, the authors show that the firing probability is asymptotically independent of the signal amplitude:

$$f(t) \rightarrow \frac{r\,p(t)}{\frac{1}{T}\int_0^T p(t)\,dt} \tag{1.14}$$

The average of firing can be represented by an equivalent of an RC circuit approximation **??** in which the current through $perm(t)$ is proportional to the instantaneous firing rate, the source current r corresponds to the rate of generation of quanta, the current $f(t)$ through the variable resistance represents the firing probability of the nerve, the charge on the capacitor C corresponds to the average number of quanta in the hair cells. The expected

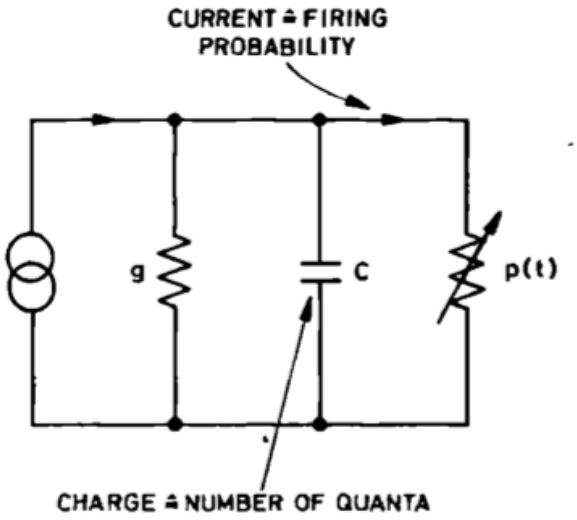


Figure 1.12: Equivalent electrical circuit of firing generation in hair cell. After [54]

number of quanta results from the differential equation:

$$\frac{dn}{dt} = r - n(t)(p(t) + g) \tag{1.15}$$

that gives:

$$n(t) = \frac{[r \int_0^T I(\tau)\, d\tau + n(0)]}{I(\tau)} \tag{1.16}$$

where

$$I(\tau) = e^{g\tau + \int_0^\tau p(\theta)\, d\theta} \tag{1.17}$$

1.4 Recent development of the ACF-based models

The physiological model of perception based on the autocorrelation function (ACF), firstly proposed by Licklider [32] considering both time and frequency domains. After this work, the model of monaural perception based on the

autocorrelation function has been developed by several authors. Patterson [**?**] introduces the pulse-ribbon and strobed auditory integration (SAI), as a crosscorrelation between the neural response within each channel with a strobe of single pulse per stimulus. The SAI is quite similar to autocorrelation.

Meddis [37, 38] sums the ACF across channels in the frequency domain and obtains a summary autocorrelation function (SACF) based on the probability of all-order interspike intervals. With the hair cell transduction model Meddis [**?**] introduces the pitch dominance region and permits to compute the relative weight of low-versus-high frequency channels and the breakdown of neural firing synchrony at high frequency. Cariani and Delgutte [10] show the relation between the shape of the autocorrelation histograms (ACH) (*all-order interspike interval* histograms for each channel) and the ACF. The raw waveform proposed by Yost [74] is also similar to SACF. Despite the recent improvement to the original Licklider's model at the present time (2011) the words of de Cheveigne are still valid: "autocorrelation model remains a good first-order model, attractive in terms of simplicity, explanatory power and physiological plausibility" [**?**]. The Licklider's triplex theory model [33] moreover reveals analogies with recent models based on multi-channel analysis and neuronal spiking.

Chapter 2

Monaural criteria in Concert Hall Acoustics

2.1 Earlier works

The optimum reverberation time is treated in others earlier works by several authors [**?**, 70] obtaining the optimum value in relation to the room volume or the sound intensity. MacNair at Bell Laboratories [**?**] proposes the conditions:

$$\int_{t_0}^{t_1} L_t \, dt = -K \tag{2.1}$$

in which t_0 is the time a sustained source of sound, E is cut off, T_1 the time the sound becomes inaudible, L_t the loudness of the sound at any istant t, K is a constant.

Lifschitz [34] states optimal conditions as:

$$\int_{0}^{t_1} log(\rho) \, dt = -a \, E \tag{2.2}$$

where ρ is the instantaneous value of the energy density of sound during decay, t_1 is the time of decay to minimum audibility, E is the rate of supply of energy during the steady state condition. The optimal conditions proposed by Lifschitz, in other words, are verified where ρ decays as a regular negative exponential.

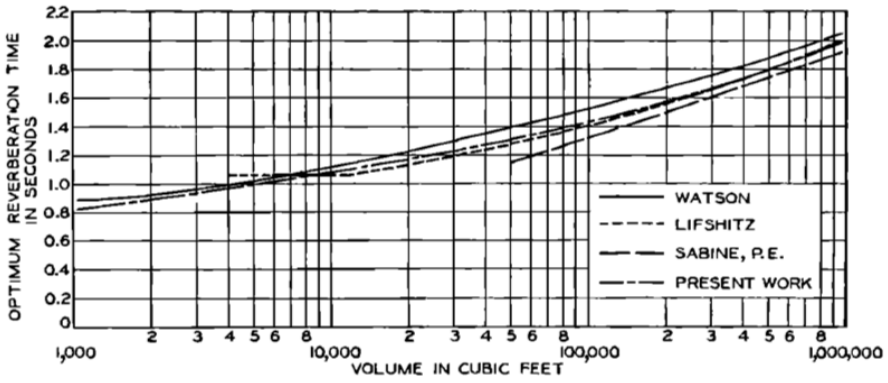


Figure 2.1: Optimum reverberation time at 512 Hz in relation to the volume of the hall. Comparison between the results of [34, **?**, **?**]. After [**?**]

Maxfield [**?**] in agreement with the empirical data and with the equation derived by Maxfield and Albersheim [**?**] proposes the conditions:

$$\int_0^{t_1} \rho \, dt = -a \, const, \quad E = G \, V^n \tag{2.3}$$

where G is a constant, V is the volume, n is a constant which should be less than unity ($n = \frac{2}{3}$ can be assumed). Measurements of MacNair, Lifschitz, Sabine are reported in fig. 2.1.

2.2 An isolated Russian proposal

Fourdouiev in the 1965 [**?**] proposes an optimal reverberation time that depends from the autocorrelation of the musical signal or speech signal. The degree of coherence between the signal and the same delayed signal is comparable with the correlation between the direct sound and each reflection. The exponential weighted running autocorrelation function (called *fonction d'autocorrelation evolutive* in the original paper) is defined by Fourdouiev as:

$$R(\tau) = \frac{1}{T} \int_{-\infty}^{T} e^{-\frac{t-x}{T}} \phi(x)\phi(x-\tau)) \, dt \tag{2.4}$$

where the time constant $T = 30\,ms$ is assumed long enough to approximate the form of $R(\tau)$ to a short-time autocorrelation. In case of non sta-

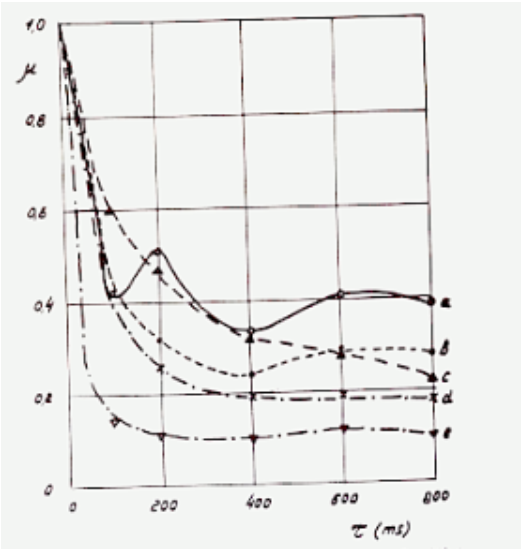


Figure 2.2: Degree of coherence μ of different motifs. α) Bach: Preludio and corale (organ) $\theta = 1000\,ms$ $\mu_0 = 0.3$ β) Reger: Suite, part I (alto) $\theta = 600\,ms$ $\mu_0 = 0.25$ γ) Glasounov: Valzer (orchestra) $\theta = 600\,ms$ $\mu_0 = 0.2$ δ) Frescobaldi: Passacaglia (clavicembalo) $\theta = 150\,ms$ $\mu_0 = 0.15$ ϵ) Grieg: Lieder (piano and voice)
$\theta = 140\,ms$ $\mu_0 = 0.15$. After [**?**]

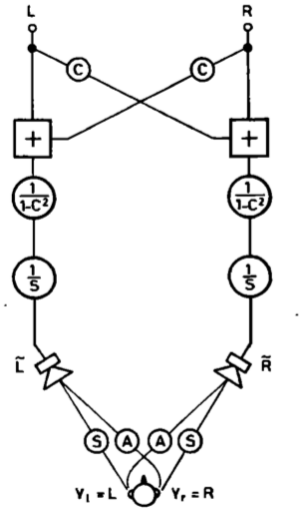


Figure 2.3: Schematics of listener's system with cross cancellation filtering. After [55]

tionary signals the coherence degree μ can be evaluated by the average of the normalized squared running autocorrelation $R(\tau)$:

$$\mu = \frac{\int_0^T R^2(\tau)\, d\tau}{\int_0^T R^2(0)\, d\tau} \tag{2.5}$$

The coherence degree of several motifs, so calculated, are showed in fig. 2.2. Different lower values of the decay $\mu_0 are chosen for evaluating the different mothifs. The$ values is chosen in the interval 0.15 - 0.3 and corresponds to the inflection point of the decay. θ is the delay τ at which $\mu(\tau) = \mu_0$.

The Fourdouiev's paper has been presented at the ICA conference of 1965 and remains an isolated proposal to the western scientific community.

2.3 The preferred space

A change of point of view from the 'optimum' to 'preferred' reverberation time is due to the group of acousticians at the Drittes Physikalisches Institut of the University of Gottingen (Schroeder, Gottlob, Siebrasse, Atal et Al.). In a comparative study of European concert halls [**?**] the cited authors propose

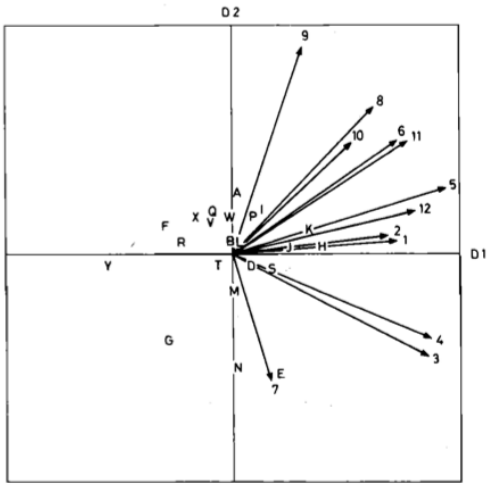


Figure 2.4: Preferred space of several european concert hall. After [55]

new methods for the evaluation of subjective perception. On one hand this study introduces a new method to evaluate the subjective experience in a concert hall: an anechoic registration is convolved with the binaural impulse response of the concert hall and the resultant two signals are proposed to the listener after a crosstalk cancellation (fig. 2.3).

On the other hand, the authors propose a new formalism: a preference space to evaluate the judgements of the listeners. In the preference space each vector represents the judgement of the listeners on a articular hall: the projection on the abscissa represents the mean consensus on the hall, the projection of the ordinate reflects individual preference disparities.

The preference space is used also to represent the correlation of each objective parameter with its subjective perception. Two geometric parameters obtained from drawings of the hall (volume V, width W of the hall or the first sidewall reflection) and four acoustic parameters obtained from measured impulse response ($ITDG$, T_{15}, definition D, $IACC$) are evaluated 2.5. The results show that the greater is reverberation time T_{15}, the greater the consensus preferences for these halls (high abscissa value, low ordinate value). The definition D has a highly negative correlation with the preference: lower values of definition corresponds to higher preference values (high negative abscissa value, low absolute value of the ordinate). $IACC$ is also

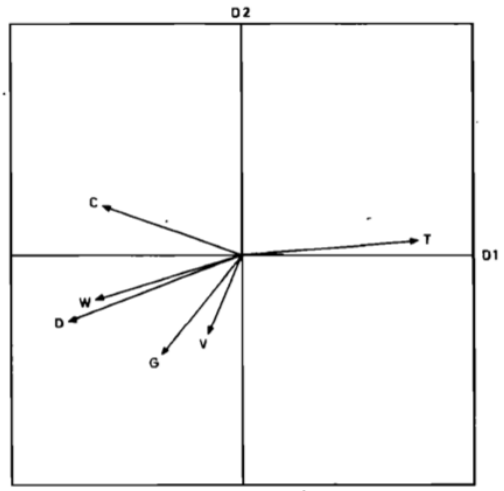


Figure 2.5: Consensus vs. Preference disparities for different room criteria. After [55]

strongly correlated to preference: a low $IACC$ value is expected to have a better judgement. The authors do not consider significant the correlations concerning the value of width, $ITDG$ and volume and the preference of listeners.

2.4 The work of Yoichi Ando

2.4.1 The preferred delay of the first reflection Δt_1

The previously cited work [?] a non significant relation between $ITDG$ and subjective preference was founded by the authors. Fig. 2.5 shows that, when the preference is good, a lower value of ITDG may be generally expected (the abscissa value is negative) but not always. In other words, not in all cases a lower ITDG corresponds to a listener's better judgement.

In a Ando's paper of 1977 [1] a listener's test setup similar to fig. 2.3 is used. An anechoic motif is proposed frontally to the listener in addition to a delayed version of the same motif, performed by lateral loudspeakers. A fixed direction of 36° is chosen for the lateral loudspeaker, simulating a typical reflection angle. Changing the delay time of the delayed signal and

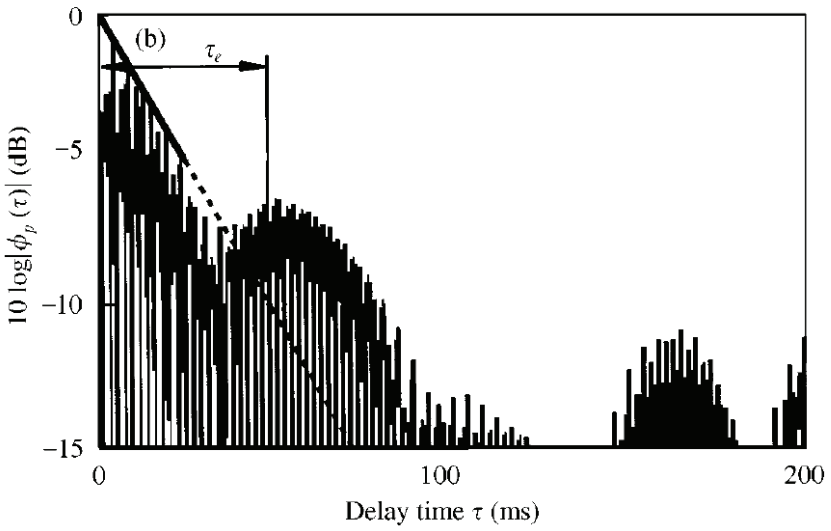


Figure 2.6: Definition of effective duration of ACF. From [3]

using Thurstone's law [65, 66], different levels of listener's preference are evaluated (fig. ando197701) . Following Licklider's formulation (eq. 1.1) the normalized autocorrelation function used by Ando is given by

$$\phi(\tau, t) = \frac{\Phi(\tau, t)}{\Phi(0, t)} \tag{2.6}$$

$$\Phi(\tau, t) = \frac{1}{2T} \int_{t-T}^{t+T} W(\xi)\, f_A(\xi)\, f_A(\xi + \tau)\, d\xi \tag{2.7}$$

where $W(\xi)$ is a rectangular window function, $f(\xi)$ is the original signal and $2T$ in this case is equal to $33s$ and $f_A(\xi)$ is

$$f_A(\xi) = f(\xi) * h_A(\xi) \tag{2.8}$$

where $h(\xi)$ is the impulse response of an A-weighting filter. The *effective duration* of the ACF (τ_e) is defined as *the first ten-percentile of envelope decay of the normalized running autocorrelation function* expressed in dB (fig. 2.6). In fig. 2.7 the relationship between the preferred Δt_1 and the effective duration of the ACF τ_e is shown: $\Delta\, t_1 = \tau_e$ (with the condition $A = 1$). Changing the angle of the first reflection and evaluating the relative

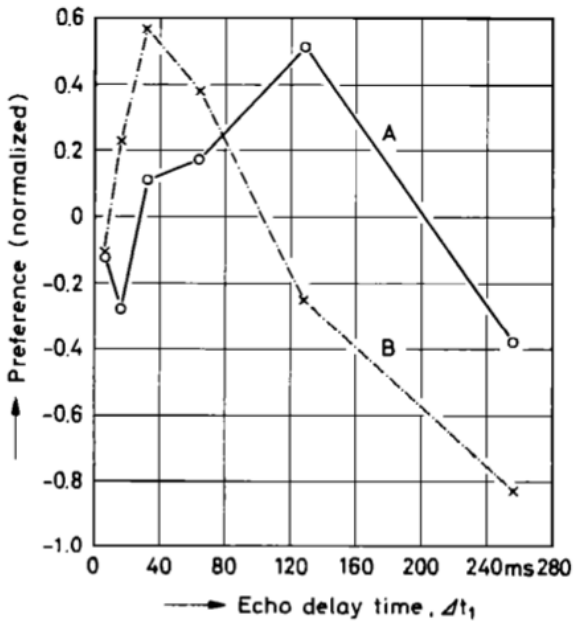


Figure 2.7: Relationship between subjective preference and Δt_1. Curve (A): Motif 1; Curve (B): Motif 2. After [1]

$IACC$, the listener's preferences are as shown in fig. 2.9. The maxima of listener's preference corresponds to the ACF effective duration of each motif (see fig. **??**:

$$\Delta t_1 = \tau_e \tag{2.9}$$

On one hand, the preference decreases in case $0 < \Delta t_1 < \tau_e$ for two reason:

- there are some interference phenomena in the coherent time region ($\tau < \tau_e$) due to the coloration effects of the reflections.
- the IACC increases when Δt_1 tends to zero: this fact lowers the total listener's preference.

On the other hand, the preference decreases when $\tau_e < \Delta t_1$ because echo effects disturb the listener's perception and the $IACC$ increases.

In a later paper [**?**] Ando and Alrutz study the coloration of the first reflections in terms of factors extracted from ACF: an analytical exposition of ACF envelope is proposed. According to Licklider [32][33], the signal

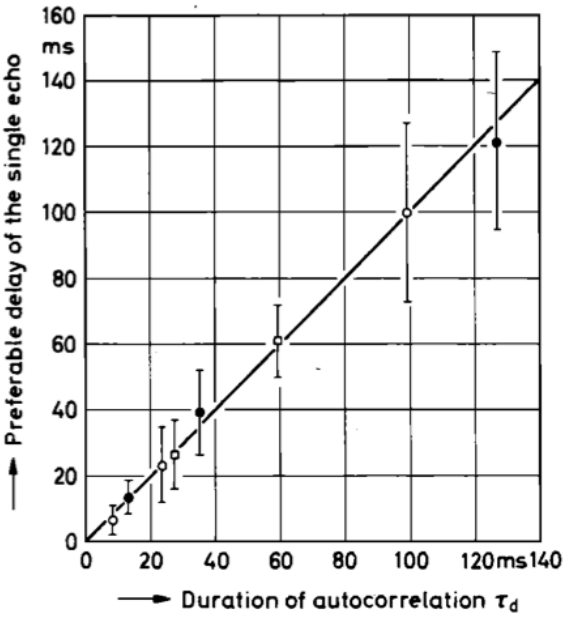


Figure 2.8: Relationship between preferred Δt_1 and τ_e value extracted from a long time ACF for different motif and regression straight line. After [1]

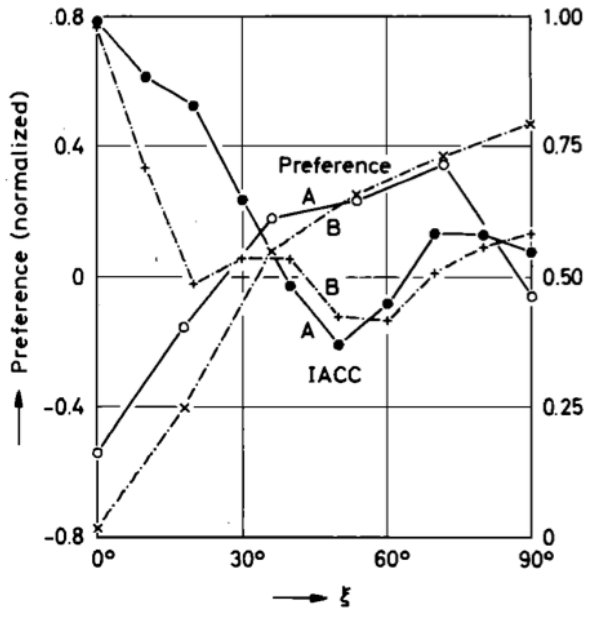


Figure 2.9: Relationship between preferred Δt_1, IACC value and lateralization angle ξ. Curve (A): Motif 1; Curve (B): Motif 2. After [1]

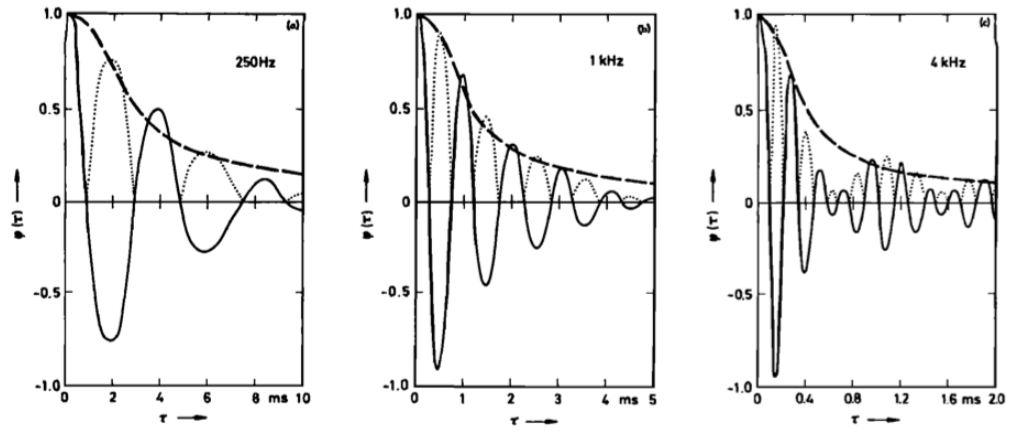


Figure 2.10: Autocorrelation of a Gaussian noise after passing through an ideal bandpass filter. From [**?**]

that represents the internal auditory excitation is supposed as an amplitude modulated AM signal. The autocorrelation of a Gaussian noise after passing through an ideal bandpass filter (with a flat response between the frequencies f_1 and f_2) is given by:

$$\phi(\tau) = \frac{2}{\Delta\omega\tau} sin(\frac{\Delta\omega\tau}{2}) cos(\frac{\Delta\omega_c\tau}{2}) \tag{2.10}$$

where $\Delta\omega = 2\pi(f_2 - f_1), \quad f_2 > f_1$, $\Delta\omega_c = 2\pi(f_2 + f_1)$. The envelope of the ACF is given by:

$$env_\phi(\tau) = \begin{cases} \frac{2}{\Delta\omega\tau} sin(\frac{\Delta\omega\tau}{2}) & 0 \leq \Delta\omega\tau \leq \pi \\ \frac{2}{\Delta\omega\tau} & \Delta\omega\tau > \pi \end{cases} \tag{2.11}$$

Atagi et Al. [**?**] study the the relation between the subjective preference and Δt_1 in case of a time variant sound field. The variant sound field is given *e.g.* by fluctuations due to the air conditioner. The Δt_1 in this case is 'modulated' with the air fluctuations. The study shows that the 'stationary' value of Δt_1 is related also to $(\tau_e)_{min}$ measured in non-stationary fields and that the sound fields with the modulated delay are more preferred than the ones with the fixed delay time of reflections.

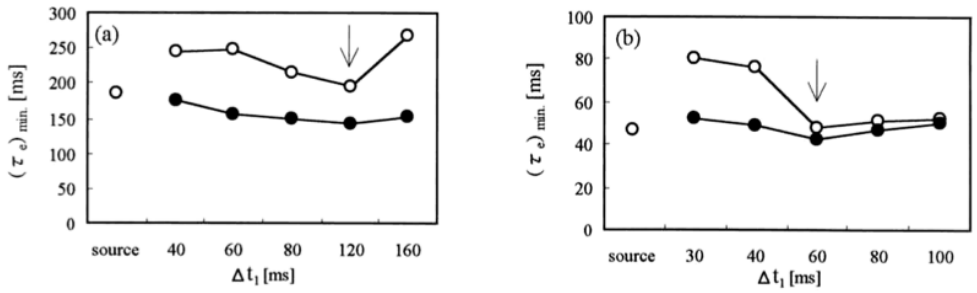


Figure 2.11: Relations between $(\tau_e)_{min}$ value and Δt_1 in case of time variant sound field. From Atagi??

Table 2.1: Motifs 1 and 2 with the relative values of τ_e calculated by Ando [4].

2.4.2 The preferred reverberation time T_{sub}

The binaural impulse response at the two ears is given by:

$$f_{l,r}(t) = p(t) * h_{l,r}(t) \tag{2.12}$$

where $h_{l,r}(t)$ is the binaural impulse measured at listener's place. Ando identifies two independent objective parameters in listener's preference as follows:

- the effective duration of the long time normalized autocorrelation function $\phi(\tau)$ in the sense of 2.6, 2.7
- the room impulse response represented by the structure of the early reflections and by the subsequent reverberation. In addition to these, another independent objective parameter is the $IACC$ (the maximum of long time inter-aural cross correlation function): the parameter in this investigation is fixed in a preferred condition and is not investigated as an independent parameter. Several anechoic motifs are proposed to the listener, with a method similar to that described in [1]. These motifs are represented by the relative τ_e values (see tab. 2.1)

The Thurstone's law of comparative judgements [65][66] is used for the evaluation of listener's preference, as in the statistical decision theory. The iso-

preference curves relative to the different motifs are shown in [fig 5a]. Fixed one parameter (the motif) the listener's preference is function of other two independent parameters: the delay of the first reflections Δt_1 and the T_{15} T_{sub}. The dashed lines represent the preferred values of Δt_1 founded in previous study [**?**]. The solid lines represents the preferred values of T_{sub} and they are founded by regression as:

$$[T_{sub}]_p \approx 23\,\tau_e \tag{2.13}$$

Fig 5a, 5b SCANSIONARE

The relation between the preferred T_{sub} and τ_e is shown in Fig. XX. Let's note that:

- in the case of the Motif 1 (Pavane), in which the effective duration of ACF has an absolute high value, the variance of preference judgement is higher than the one of the other's motifs.
- a long time ACF is used.

When the ACF is strongly damped (*e.g.* in case of speech signals) the relation **??** becomes critical. It's not clear if this result is due to the linearity faulit of relation 11 or to problems when a short value of τ_e(less than 10 ms) is calculated. The temporal window used in [2] contains all the anechoic motifs: the τ_e in this study is a constant and not a function of the running time, from which the minimum value is donsidere, as showed in later papers [4]. The preferred value of T_{sub} is given by:

$$[T_{sub}]_p \approx 23\,[\tau_e]_{min} \tag{2.14}$$

where $[\tau_e]_{min}$ is the minimum value of the $running\tau_e$

Fig 6 SCANSIONARE

2.4.3 Theory of acoustic design

The cited researches on preferred delay of first reflection and preferred subsequent reverberation time show the role of the τ_e as objective parameter for

characterize the audio signal that 'excites' a sound field. Moreover the more important role of the minimum of the running τ_e needs adequate methods of measurement. The literature and the theory that concerns these methods will show in detail in later part of this thesis. Figg. 5-6 (paragrafo precedente) propose some cross related analysis to evaluate the listener's judgement preference with some analogy to the preferred space of Schroeder et Al. Later studies found an high level of total subjective preference in case of preferred value of each factors, and an high of independence of these factor that authorize to postulate orthonormality of each factor respect to the others. In the statistical theory, as resumed (par.XX), a wave can be specify in a four dimension signal space: these coordinates are orthonormal and may represent uniquely the listener's perception.In a possible analogy with the hypothesis of statistical decision theory, Ando propose in a well know book an organic theory in which, by the principle of Thurstone's comparative judgements [65, 66], a scale value of subjective preference may be expressed. A recent form of this theory is the Theory of Acoustic Design [**?**], in which the total scale value of the subjective preference may be described in form of four orthonormal preference functions S_i.

$$S = (S_2 + S_3)_{left} + (S_1, S_4)_{right} \tag{2.15}$$

where $(S_2 + S_3)_{left}$ are related to temporal impression and $(S_1, S_4)_{right}$ are related to spatial impression. Each subjective preference function may be expressed in form of temporal and spatial factors, by a regression fitting from listener's evaluation (see fig. 2.12)

The expressions of normalized factors are:

$$S_i = \alpha(x_i)^{\frac{3}{2}}, \quad i = 1, 2, 3, 4 \tag{2.16}$$

Each normalized factor x_i may be expressed in form of temporal and spatial factor

$$x_2 = log\frac{\Delta t_1}{[\Delta t_1]_p} \tag{2.17}$$

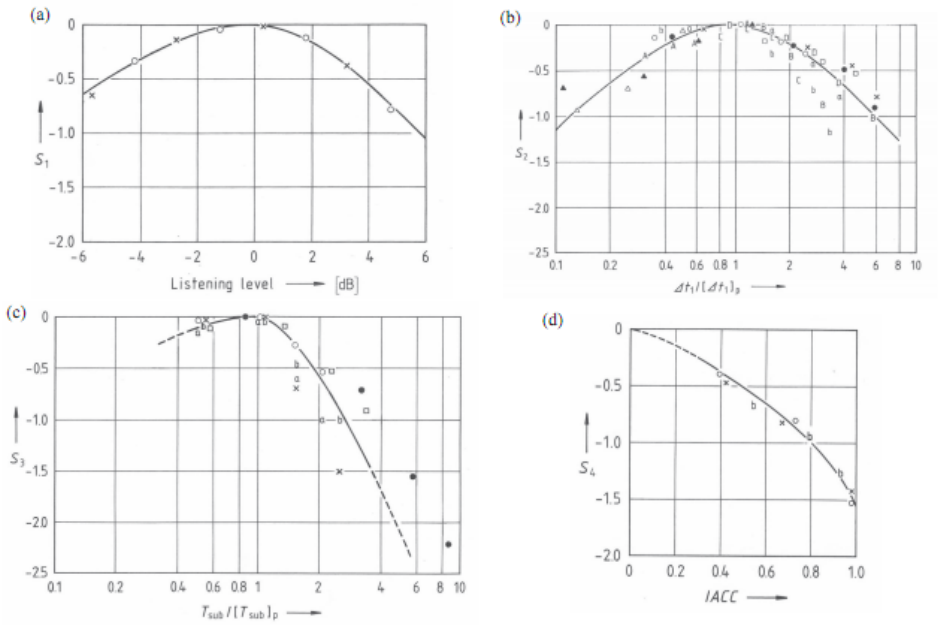


Figure 2.12: Normal factors of subjective preference. From [**?**]

$$x_e = log \frac{T_{sub}}{[T_{sub}]_p} \tag{2.18}$$

and

$$[\Delta t_1]_p = (1 - log_{10} A)(\tau_e)_{min} \tag{2.19}$$

$$[T_{sub}]_p = 23(\tau_e)_{min} \tag{2.20}$$

The normalized factors relative to spatial impression are:

$$x_1 = LL - [LL]_p = 20\, log \frac{p}{[p]_p} \tag{2.21}$$

$$x_4 = IACC \tag{2.22}$$

where LL is the pressure level $[p]_p$ is the preferred pressure level.

As noted in [**?**] the scale value have little change in the neighborhood of the most preferred conditions. Using the optimal value of subjective preference value a design criterion of concert hall is possible, as resumed in fig.

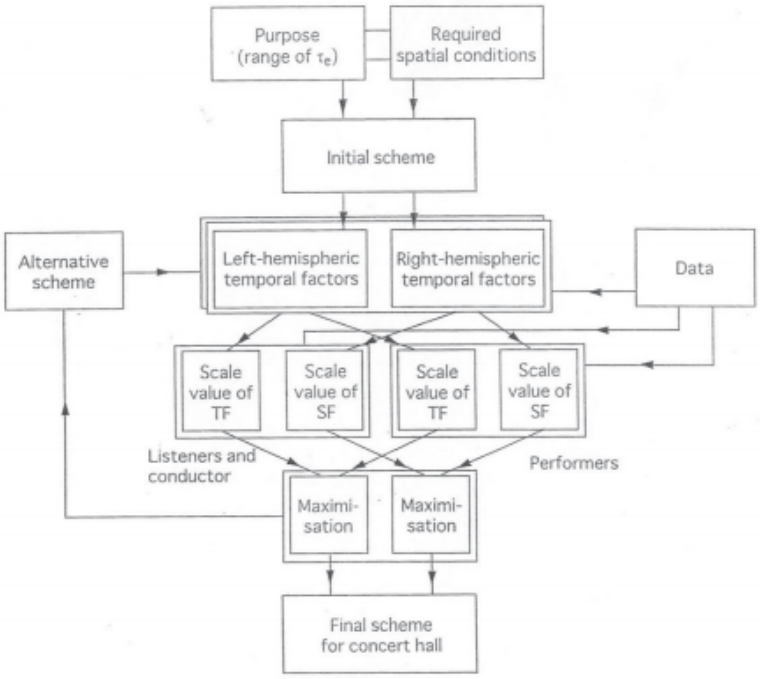


Figure 2.13: Design criterion of concert hall following the theory of acoustic design. From [**?**]

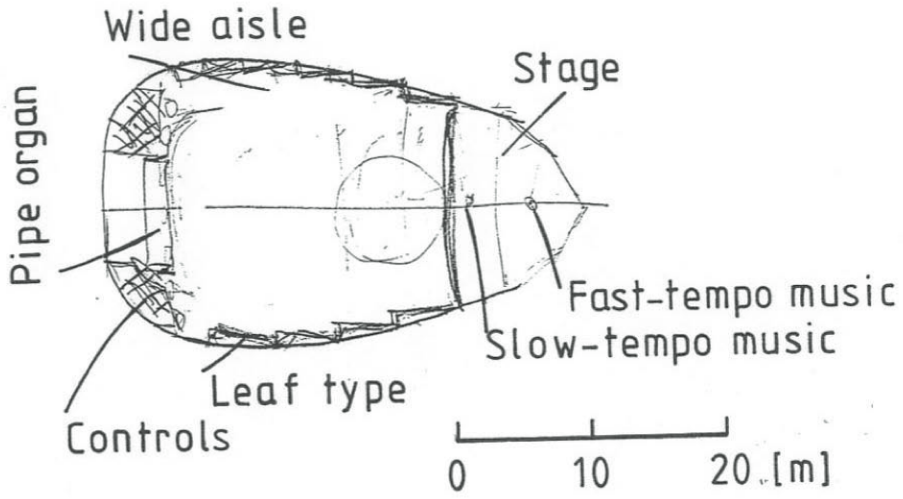


Figure 2.14: Sketch of Kirishima International Concert Hall. From [**?**]

2.13. The temporal and spatial criteria are exposed later (par. XX).

Two applications to this design theory are the Kirishima International Concert Hall and the Tsuyama Music Cultural Hall. For which concerns the monaural criteria different solutions are proposed in the two concert halls. In the Kirishima International Concert Hall the geometry of the reflectors above the stage permits to accord the $Deltat_1$ to $(\tau_e)_{min}$ of the execution by different positioning of the performers. In the Tsuyama Music Hall the variable height of the reflectors above the stage permits a variable $Deltat_1$, according to the τ_e of the execution. Fixing the reverberation time at 1.7s the minimum of tau_e that optimize this concert hall is located in the range 50-90 ms.

The theory of acoustic design authorize the use of algorithmic criteria of concert hall project or optimization. Genetics algorithm for the design of concert hall based on subjective preference theory are proposed by Sato [CERCARE][l'altro cinese dell'ISTD].

2.4.4 Theory of auditory temporal and spatial primary sensations

The results of the last year of investigations of Ando's group (Andolab) are resumed in the theory of auditory temporal and spatial sensations [**?**]. Brain activities corresponding to subjective preference are observed. Auditory brainstem responses (), slow vertex response (), Magnetoencephalogram () and Electroencephalogram () from right and left hemisphere are recorded. This studies appear that temporal factors extracted from are associated with the left hemisphere and the spatial factor extracted from are associated with the right one. In relation to the field of interest of this thesis the major results due to the cited papers are:

- Δt_1 and T_{sub} indicate a left hemisphere dominance [**?**, **?**, **?**]
- LL and IACC indicate right hemisphere dominance with an high degree of independence between the left and right factors. It may confirm some orthogonality of factors in the subjective preference theory.
- The scale of subjective preference is related to the τ_e value extracted from the ACF of the $\alpha-$wave in EEG when one of the temporal and spatial factors field is varied [**?**].
- The scale of subjective preference corresponds to the τ_e value extracted from the ACF of the $\alpha-$wave of MEG in the left hemisphere [**?**].

The paper cites the Cariani's interpretation of ACF mechanism of the spike activity in the auditory processing. The auditory model used in the paper has some modifications respect to the previous works, concerning the ABR waves (fig. 2.15). The "ear sensitivity" is described with the impulse response of an $A-$weighted networks, as the previous work, in reason of a general use. Thus, using the author's words, *the individual sensitivity is rigorously different*. Some limits of the proposed auditory model can be derived from the identification of fundamental frequency using the ACF first peak τ_1 [21]. The method fails in case of fundamental frequency above 1200 Hz (fig. 2.16). In the past section [riferimento] it will discussed if this limits concerning the

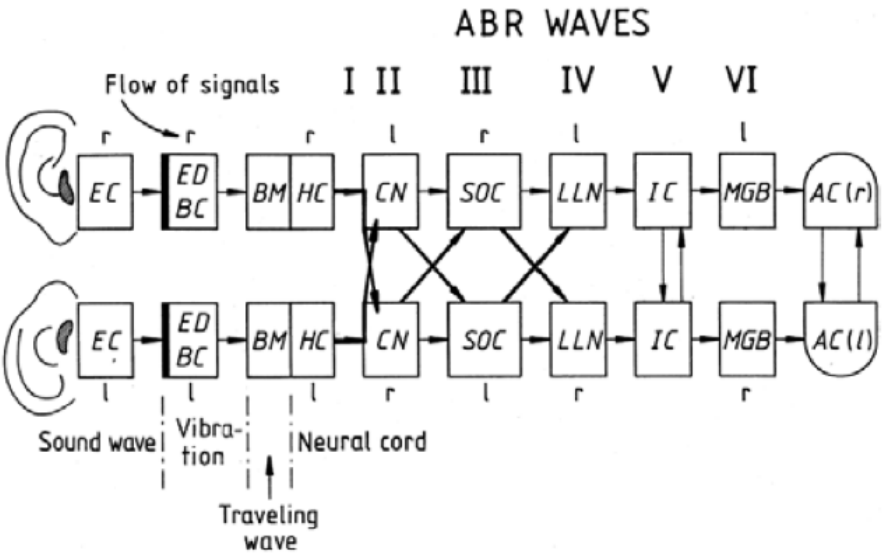


Figure 2.15: Auditory brain model with localization of ABR waves. *EC* is external canal, *ED* ear drum, *BC* ??, *BM* is the basilar membrane, *HC* hair cell, *CN* cochlear nuclei, *SOC* ??, *LLN* ??, *IC* inferior colliculus, *MGB* ??, *AC* are the autocorretional and crosscorretional process (from [**?**])

fine structure analysis may be supposed also in the envelope analysis (as the τ_e extraction).

2.4.5 Application of τ_e analysis in other fields of acoustics

τ_e and other factors extracted from ACF have been widely used in different fields of acoustics. In architectural acoustics, τ_e is related to the subjective preference [3, 2, 1, 47, 6]. In musical acoustics, τ_e is related to the analysis of performance [28]. τ_e has also been used to quantify the annoyance [5, 51, 61, 62]. In environmental acoustics, τ_e is used to analyze environmental noise [5], soundscape quality and aircraft noise [17, 50]. In speech analysis, τ_e is used for pitch detection [21] and speech recognition in a noisy environment [29]. Finally, τ_e has been used in building acoustics [25].

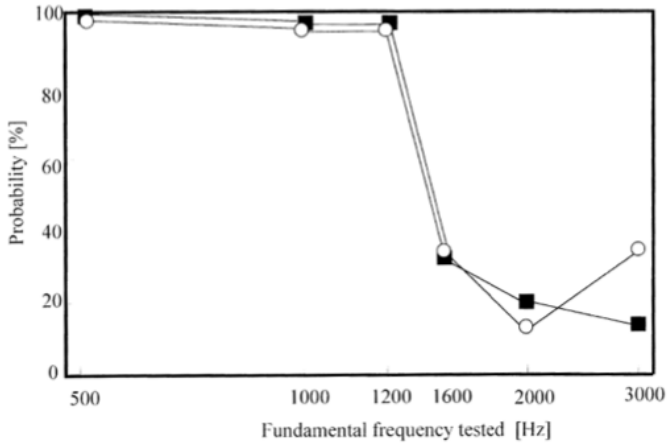


Figure 2.16: Probability of pitch detection vs. τ_1 identification. From [?]

2.5 Feedback process in musical execution

Winckel [?] studies the dynamic range and the tempo of several performed music in different halls. The Winckel's studies shows that dynamic range may depend on the noise level of the room and the state of sound diffusion. The tempo of music therefore is at maximum in concert halls which have optimum reverberation time and diffusion. For an optimum reverberation time the impedance matching of the radiating pressure to the surrounding room is better than for rooms with too small a reverberation time: there is a certain feedback mechanism which varies the impedance of the instrument (or the voice) to get the best adaptation.

Many studies concern the preferred first reflection of musical performer [?, ?, 28, ?]. The expression of preferred Δt_1 is formulated in a more general forms when the listener is a performer.

$$[\Delta t_1]_p = (log_{10}\frac{1}{k} - c\, log_{10}A)(\tau_e)_{min} \tag{2.23}$$

where k, c are constants, for cellist: $k = \frac{1}{2}$ anc $c = 1$ (Sato, 2000), for alto recorder soloist: $k = \frac{2}{3}$ anc $c = \frac{1}{4}$ [41]. For a listener (k=1, c=1) the eq. 2.23 coincides to eq. 2.17. A non-preferable echo disturbance may occur if the correlation between the direct sound and delayed reflections is smaller than a

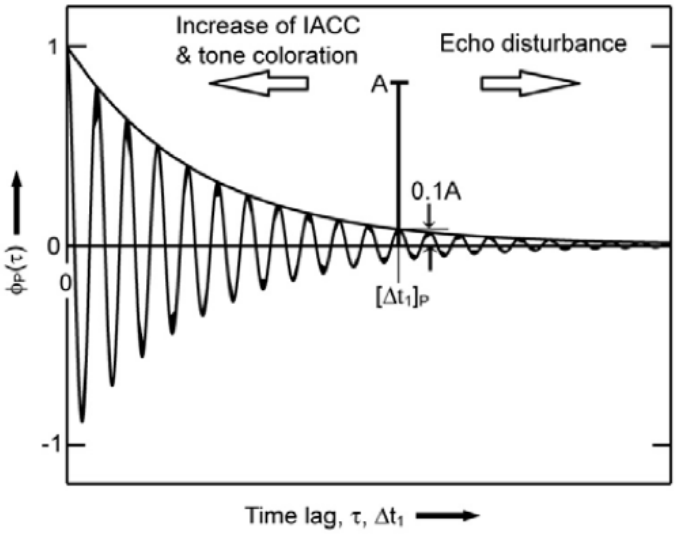


Figure 2.17: Preferred Δt_1 for performers. From [28]

certain value, then a very long initial time delay $Delta\, t_1$ causes the echo. In other words may assume that the performer needs to perceive his instruments in a preferred acoustic space, with a certain sustain. The dimensions of this acoustic space e the sustain needed varies with the characteristics of the instrument.

Chapter 3

Theory and methods of the envelope extraction

3.1 Decay Parameter Estimation

The impulse response of a resonant acoustic system can be expressed in the form: RICONTROLLARE FORMULA

$$h(t) = \sum_{i=1}^{N} (A_i \, e^{-\frac{t}{t-t_i}} \cos(\omega_i \, t + \phi_{0i}) \, u(t - t_i)) + n_{measurement}(t) \qquad (3.1)$$

where A_i, t_i, ϕ_{0i} are the amplitude, the arrival time and the relative phase of the $i-$th reflection of the acoustic system (*e.g.* a room), $u(t)$ is the step function, $n_{measurement}(t)$ is the noise in the measurement.

The normalized autocorrelation function ϕ can be expressed in the form:

$$\phi(\tau) = \sum_{i=1}^{N} (\phi_i \, e^{-\frac{\tau}{\tau-\tau_i}} \cos(\omega_i \, \tau + \phi_{0i}) \, u(\tau - \tau_i)) + n_{auditory}(\tau) \qquad (3.2)$$

where ϕ_i, τ_i, ϕ_{0i} are the amplitude, the delay and the relative phase of the $i-$th peak, $u(\tau)$ is the step function, $n_{auditory}(tau)$ is the noise in the auditory path. With these general hypothesis the ACF envelope may have the same form of an impulse response of a resonant acoustic system. That authorize

to evaluate the envelope of an ACF with the methods of an impulse response decay estimation. The decay of the impulse response and the decay of the ACF envelope in fact can be assimilated in first approximation to a straight-line fit when are considered in logarithmic forms.

Let's notes that some difference exists between the logarithmic form of a normalized ACF $\phi(t)$ and the logarithmic form of a impulse response $h(t)$. When are used the same methods for the decay estimation the two functions have to be expressed as:

$$h_{log}(t) = 10\, log_{10}(h(t)) \tag{3.3}$$

$$\phi_{log}(t) = 5\, log_{10}(\phi(t)) \tag{3.4}$$

The parallel between the decay of an impulse response and the envelope of the ACF is supported by a Ando's paper of 1989 [4]. (The analytical IPOTESI sono le ho scritte io e non ando. come lo dico?)

3.1.1 Least-squares line fitting

Least-squares line fitting (linear regression) is done by finding the optimal decay rate k and the initial level a that minimize the function:

$$\int_{t_1}^{t_2} (\phi_{log}(\tau) - (kt + a))^2\, dt \tag{3.5}$$

The precision of this method depends the choice of the interval of integration in which the decay of functions should be inherently nonlinear.

A particular discrete case of least-squares line fitting may be considered in the Ando's peak detection method. The best line fitting is done by finding the decay rate k and the initial level a that minimize the function

$$\sum_{i=1}^{N} (\phi_i - (k\,\tau_i + a))^2 \tag{3.6}$$

where ϕ_i, τ_i are the amplitude and the delay of the $i-$th peak of the logarithmic normalized ACF. Basing the method on neurophysiological mecha-

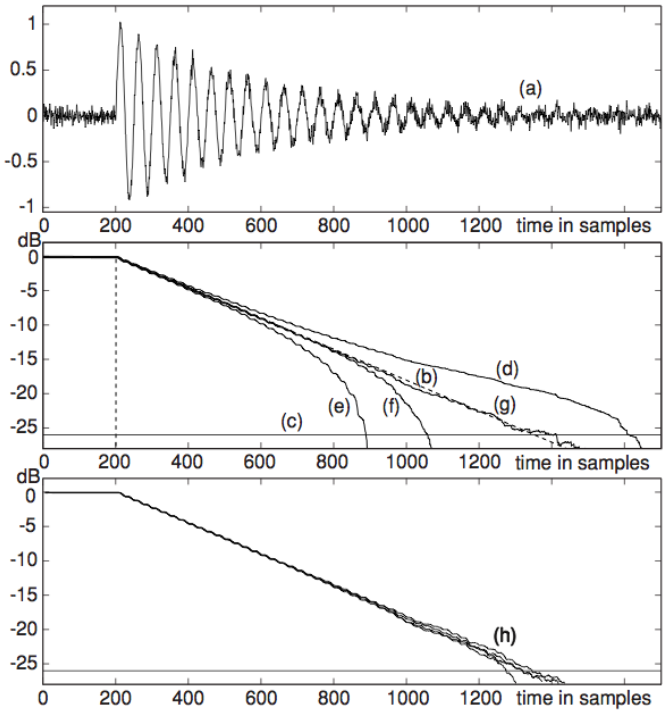


Figure 3.1: (A) Normalized impulse response; (B) Decay function for different ULI values; (C) Fitting of decay function. From [27]

nism, Ando proposes that only the N peaks located in the first 5 dB of the decay must be fixed.

3.1.2 Schroeder's backward integral

A monotonic and smoothed decay curve can be produced from an impulse response by a backward integration [**?**]. The decay function of an ACF $\phi(\tau)$ may be given by:

$$L(\tau) = \int_{\tau}^{ULI} \phi_{log}^2(\tau), d\tau \tag{3.7}$$

where ULI is the Upper Limit of Integration time. A known problem is the choice of ULI in relation to the background noise floor or generally in relation to the slope of the decay. In the first case if the background noise floor is included within the integration interval the decay function can have a raised ramps that biases upward the late part (or a significant part) of the

decay. In the second case a low slope of decay produces the same effects: the integration of the late part of the $\phi(\tau)$ can biases upward the decay function (see fig. 3.1). In order to reduce this problem many author proposes limits of integration or control techniques based on the knowledge of the noise floor level (or the slope of impulse response decay). As previous specified the autocorrelation function $ACF(\tau)$ and the impulse response can be treated in the same way: this fact permits to use in the ACF envelope analysis the same methods founded in $h(t)$ analysis.

The study of Kurer and Kurze [30], accepted in the ISO 3382 [**?**], suggests that the noise floor must be 10 dB below the lowest level used in the decay evaluation. In other words the ULI must be chosen at a level minor of -70 dB in case of evaluation of the T_{60}, minor of -45 dB in case of the T_{30}, etc... Xiang proposes [73] a non linear regression approach: in case of initial over estimated ULI the iterated regression permits a good agreement between the real decay and the so calculated decay. On the other hand in case of initial under estimated ULI the method cannot guarantee the expected convergence. Morgan proposes [**?**] an analytical evaluation of the error due to Schroeder's Backward Integration method. Basing his method on statistical hypothesis, Morgan postulates the squared impulse response equal to:

$$h^2(t) = (e^{-t/2}r(t) + \sqrt{a}n(t))^2 \tag{3.8}$$

where $r(t)$ is a Rayleigh fading process (citer Jeffress?) a is the noise floor power $n(t)$ is a Gaussian noise.

The backward integration integral is given to: RICONTROLLARE LE FORMULE QUI ED ESPRIME IN TERMINI DI phi

$$F(t) = \int_{t_0}^{ULI} h^2(t)dt \tag{3.9}$$

Let's note that others methods proposed for the right choice of ULI [11, 36] can be reformulate in the form of eq. 3.8 fixing an $ULI = ln\frac{1}{a}$. These methods may be viewed as a truncation at the knee of the curve.

Following the eq. 3.9 the statistical mean of the squared impulse response

is given by:

$$E\{h^2(t)\} = e^{-t} + a \tag{3.10}$$

The backward integration of this mean gives:

$$F(t) = \int_t^{ULI} E\{h^2(\xi)\}d\xi = e^{-t} - e^{-ULI} + a(ULI - t) \tag{3.11}$$

The straight line fit of this decay function can be written as:

$$\hat{F}(t) = (1 - e^{-ULI} + ULI\, a)\, e^{-t/\hat{\tau}} \tag{3.12}$$

where:

$$\hat{\tau} = \frac{ULI}{ln\frac{e^{-t_0}+e^{-ULI}+ULI\, a}{1-e^{-ULI}+ULI\, a}} \tag{3.13}$$

The error ratio between the right decay and decay curve obtained from Backward Integration is given by:

$$\epsilon(t) = \frac{e^{-t} - e^{-ULI} + a(ULI - t)}{(1 - e^{-ULI} + ULI\, a)e^{-t}} \tag{3.14}$$

The correct choice of the ULI is noted also by Kato [28]. In the same article it is proposed to apply the integration to the initial part of ACF only: the initial part is fixed in $50ms$ according to evaluation of a vocal musical signal. For a general usage this choice can limit the applicability of the technique, e.g. in case of 'slow' musical motifs or stationary tonal signals when high τ_e values are expected [62].

3.1.3 Hilbert's transform

The Hilbert transform of a normalized ACF $\phi(\tau)$ may be expressed as a convolution with the Hilbert's kernel $\frac{1}{\pi\tau}$:

$$H[\phi(\tau)] = \phi(\tau) * \frac{1}{\pi\tau} \tag{3.15}$$

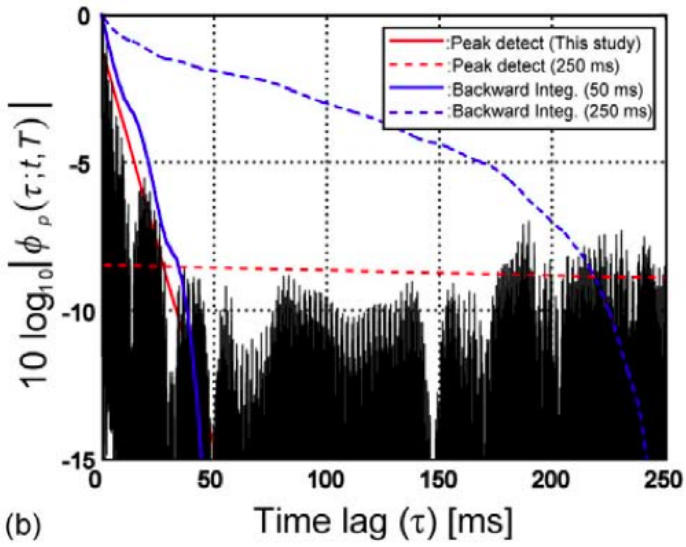


Figure 3.2: Schroeder's decay functions for different upper limit intergration (ULI) values. FromKato2006

where $*$ is the convolution operator. The discrete form of the Hilbert transform (HDT) is:

$$HDT\left\{\phi_n\right\} = h_n * \phi_n \tag{3.16}$$

where: ϕ_n is the sampled version of $\phi(\tau)$; h_n is the discrete Hilbert filter, defined as:

$$h_n = \begin{cases} 0 & n = \text{even} \\ \frac{2}{\pi n} & n = \text{odd} \end{cases} \tag{3.17}$$

The magnitude of HDT is the envelope of the normalized ACF.

verificare se mettere anche la parte scritta fin qui o solo quella da qui in poi.

The eq 3.2 in first approximation can be decomposed as:

$$\phi(\tau) = \sum_{i=1}^{M} \phi_i \, e^{j\omega_i \tau} \tag{3.18}$$

Assume that $\Delta\omega_i = \omega_i - \omega_1$ and that

$$\|\phi_1\| > \|\sum_{i=2}^{M} \frac{\phi_i}{\phi_1} e^{(j\Delta\omega_i \tau)}\| \tag{3.19}$$

It is possible to show that, with some conditions

$$\phi(\tau) = \sum_{i=1}^{M} \phi_i e^{j\omega_i \tau} = \phi_1 e^{-\widehat{\phi}(t)+j(\omega_1 t+\phi(t))} \tag{3.20}$$

where $\widehat{m}(t)$ is the Hilbert transform of the ACF $\phi(\tau)$. Sufficiente (?) condition for the eq. 3.20 is that $\phi(\tau)$ is a minimum phase signal. A multicomponent sinusoidal signal, with the assumptions of eq XX 3.19 is a frequency translated minimum phase signal. However if some component in the middle of the spectrum is large, then the signal is mixed phase and cannot be expressed in forms of eq. 3.20.

3.2 Notes about the signal modulation

3.2.1 Auditory frequency resolution

The role of the limited auditory frequency resolution, the envelope models and the hypothesis on the modulation of signals may be related. At the end of the chapter of this thesis that concerns the literature are necessary some simplified notes on the signal modulations proposed the auditory analysis. An history of the early studies is resumed in [19]. The Zwicker's experiment shows to a listener two signals having the same envelope but different fine structure (the one is amplitude modulated, the other one is frequency modulated). The expression of AM signals:

$$s_{AM} = \frac{m}{2}\cos(\omega_1 - \omega_2 t) + \cos(\omega_1 t) + \frac{m}{2}\cos(\omega_1 + \omega_2 t) \tag{3.21}$$

where ω_1 is the carrier frequency, ω_2 is the modulation frequency, m is the modulation depth.

The expression of signal FM can be assumed in detection literature as special case of QFM signal:

$$s_{QFM} = \frac{\beta}{2}\cos(\omega_1 - \omega_2 t) + \cos(\omega_1 - \frac{\pi}{2}t) + \frac{\beta}{2}\cos(\omega_1 + \omega_2 t) \tag{3.22}$$

where ω_1 is the carrier frequency, ω_2 is the modulation frequency, β is the modulation depth.

The QFM approximates a sine FM tone with a small modulation index β

$$s_{AM} \approx \sin(\omega_1 t + \cos(\omega_2 t)), \quad \beta^2 \ll 1 \tag{3.23}$$

Zwicker shows that for modulation frequency less or equal to a fraction the carrier frequency the signals are distinguishable, for modulation frequency over the fraction of the carrier frequency the two signal are indistinguishable (fig. 3.3). The fraction varies from 15% to 40%, depending on the sound pressure level of the stimulus. Goldstein extend this experiment to a QFM (frequency modulated tone)(eq. 3.22) and obtain the same results. Shoerer [**?**] and Sek [**?**], using a carrier of 1Khz, shows that the thresholds for detecting AM or FM signals are different under a frequency (Critical Modulation Frequency) from which the thresholds are the same for the two signals (fig. 3.4). The authors propose as descriptor of the CFM the logarithmical ratio $log_{10}\frac{\beta}{m}$ (where β is the modulation index of AM and m is the modulation index of FM). As shown in fig. 3.4 the frequency band of CMF could be equal to the band in which the ratio $\frac{\beta}{m}$ approaches unity.

The results of these study confirms that the ear is insensitive in the fine structure of signals at high frequency and that in these regions of frequencies differences in the envelopes of signals becomes important (see [**?**] for a complete bibliography of monaural and binaural experiments).

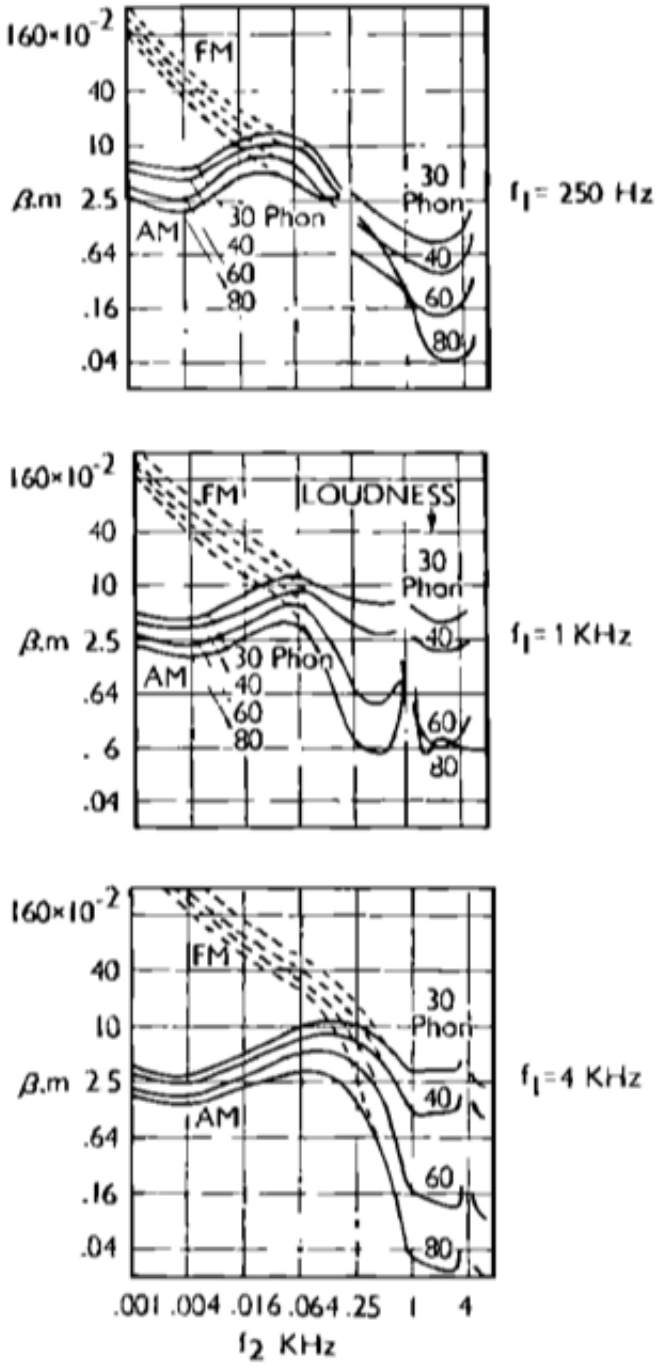


Figure 3.3: Modulation thresholds for sinusoidally modulated AM and FM for different carrier frequencies ($250Hz, 1KHz, 4KHz$). from [**?**]

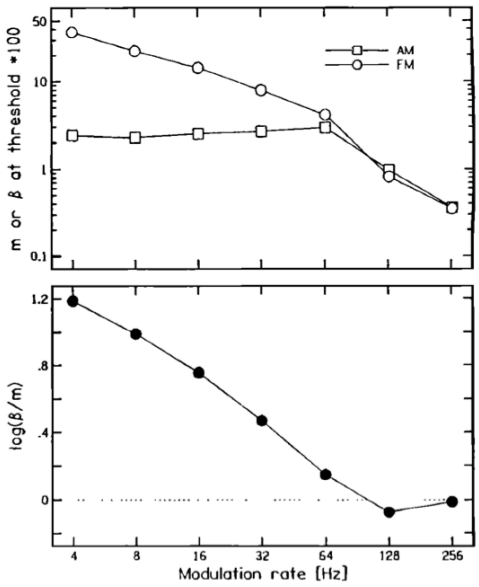


Figure 3.4: (A) Critical Modulation frequency (CMF) for AM and FM signals in function of modulation depths m and β. (B) Critical Modulation frequency (CMF) in function of log $\frac{\beta}{m}$. From [**?**]

3.2.2 Ideal envelope of minimum phase signals

A generic modulated signal (AM, QFM) may be formulated in terms of complex tone:

$$s(t) = a\cos(\omega_1 - \omega_2)t + \alpha + b\cos\omega_1 t + \beta + c\cos(\omega_1 + \omega_2)t + \gamma \qquad (3.24)$$

where for an AM signal $a = \frac{1}{2}, \alpha = 0, b = 1, \beta = 0, c = \frac{1}{2}, \gamma = 0$, for a QFM signal $a = \frac{1}{2}, \alpha = 0, b = 1, \beta = \frac{\pi}{2}, c = \frac{1}{2}, \gamma = 0$.

The ideal envelope of a complex tone $s(t)$ is given to magnitude of the positive frequency phasors:

$$env(t) = |a\,e^{j[(\omega_1-\omega_2)t+\alpha]} + b\,e^{j[\omega_1 t+\beta]} + c\,e^{j[(\omega_1+\omega_2)t+\gamma]}| = |a\,e^{j(\alpha-\omega_2 t)} + b\,e^{j\beta} + c\,e^{j(\gamma+\omega_2 t)}| \qquad (3.25)$$

It's useful to represent this envelope geometrically in therm of phasors (fig. 3.5). The ellipse is the vector sum of two sidetone phasors and the envelope is the magnitude of the phasor resultant. The normalized major

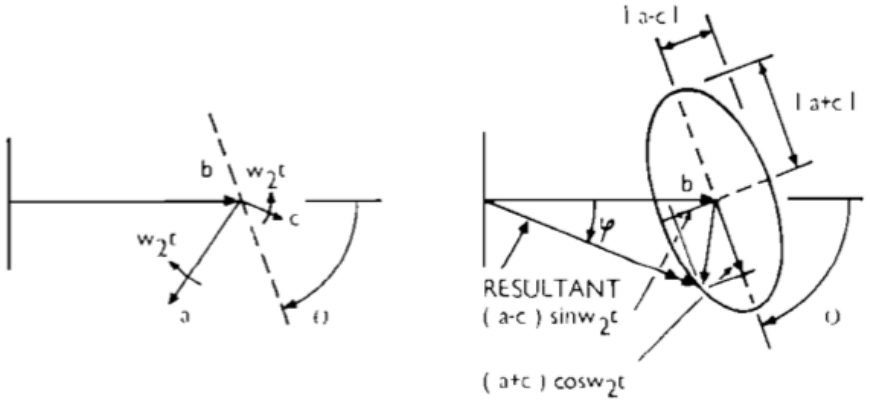


Figure 3.5: Magnitude of the positive frequency factors. From [?]

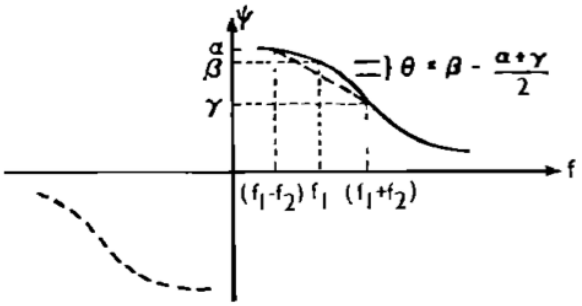


Figure 3.6: Approximation of hyperbolic sin function. From [?]

and minor axes of the ellipse are $A = \frac{a+c}{b}$ and $B = \frac{|a-c|}{b}$. The effective phase angle of the envelope response is $\theta = \beta - \frac{\alpha+\gamma}{2}$. Applying the analytic geometry to the phasor diagram a formulation of maxima and minima of the envelope is possible. For an AM signal (max when $\theta = 0$) results:

$$env_{AM,max} = \pm(1 + A) \tag{3.26}$$

$$env_{AM,min} = \begin{cases} \sqrt{\frac{B^2(A^2-B^2-1)}{(A^2-B^2)}} & \frac{A-B^2}{A} \geq 1 \\ \pm(1 - A) & \frac{A-B^2}{A} \leq 1 \end{cases} \tag{3.27}$$

For an amplitude modulated signal (max when $\theta = \frac{\pi}{2}$) results:

$$env_{QFM,max} = \begin{cases} \sqrt{\frac{A^2(A^2-B^2+1)}{(A^2-B^2)}} & \frac{A^2}{B} - B \geq 1 \\ \pm 1 + B & \frac{A^2}{B} - B \leq 1 \end{cases} \tag{3.28}$$

$$env_{QFM,max} = \pm(1 - B) \tag{3.29}$$

Let's note that neglecting the internal phase angle is equivalent to neglecting only the deviation of the internal phase characteristic from constant plus linear phase (fig. 3.6). This application is equivalent to assuming that the internal effective phase angle is much smaller than $\frac{\pi}{2}$ (minimum phase characteristics).

3.3 Theory of Signal Known Statistically (SKS)

In the par. the statistical decision theory has been briefly resumed. If the listener not knows a priori the signal a different theory (SKS) has been proposed. Some results of the theory SKS may are useful to define an envelope extractor for the ACF, as will showed in chapter XX.

3.3.1 Envelope detection

In case of signal known except for the phase the two distribution are non normal and the SN function have the larger variance. Rice [**?**] approach (?) the problem of the monaural detection by assuming a variable

$$V(t) = X \cos \omega t + Y \sin \omega t \tag{3.30}$$

where X and Y are normal variates of mean zero and variance σ^2. The quantity σ is the voltage of the band noise under consideration. The statistical analysis of $V(t)$ permits the analysis of the monaural envelope. In case of narrow band signal [24] $\sqrt{(X^2 + Y^2)}$ may be interpreted as the envelope of $V(t)$. In statistical terms the hypothesis of narrow band signal corresponds to the slowly varying of X and Y.

In case of presence of signal the term $A\cos\omega t$ is added (in phase for convenience) with $X\cos\omega t$, so:

$$V(t) = (A + X)\cos\omega t + Y\sin\omega t \tag{3.31}$$

A new pair of variables can be defined, $x = A + X$ and $y = Y$ with the probability-density functions:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-A)^2}{2\sigma^2}} \tag{3.32}$$

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-y^2}{2\sigma^2}} \tag{3.33}$$

The two functions are ortogonal and the joint density function is given by:

$$f(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2+A^2-2xA}{2\sigma^2}} \tag{3.34}$$

If the statistics of $a = \sqrt{x^2 + y^2}$ is studied (the distance from the origin) the probability density of a is given by:

$$f(a) = \frac{a}{2\pi\sigma^2} \int_0^{2\pi} e^{-\frac{a^2+A^2-2A\,a\,\cos\theta}{2\sigma^2}}\, d\theta \tag{3.35}$$

when $x = a\cos\theta$ in polar cordinates. The meaning of the distance a is the same proposed in the TSD theory (section XX).

The distribution can be writed as:

$$f(a) = \frac{a}{\sigma^2} e^{-\frac{a^2+A^2}{2\sigma^2}}\, I_0(\frac{aA}{\sigma^2}) \tag{3.36}$$

where I_0 is a the zero-order modified Bessel function of the first kind. The distribution of eq. 3.36 is called Rice's distribution. If no signal is added ($A = 0$), and since $I_0(0) = 1$ the Rice's distribution has the form knowed as Rayleigh's distribution:

$$f_0(a) = \frac{a}{\sigma^2} e^{-\frac{a^2}{2\sigma^2}} \tag{3.37}$$

(da qui in poi magari in appendice?) The mean of the Rice's distribution

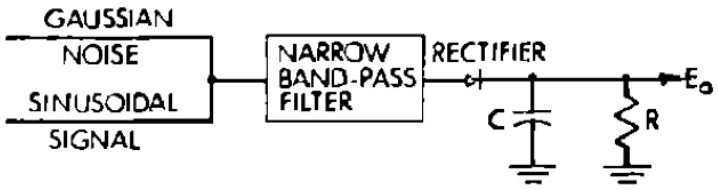


Figure 3.7: Block diagram of the envelope detector proposed in [**?**]

($\sigma = 1$ is assumed) is given by:

$$\mu_{Rice,\sigma=1} = \sqrt{\frac{\pi}{2}}[(1+A^2)I_0(\frac{A^2}{2})e^{-\frac{A^2}{2}} + A^2\, I_1(\frac{A^2}{2})e^{-\frac{A^2}{2}}] \tag{3.38}$$

The mean of the Rayleigh's distribution is given by:

$$\mu_{Rayleigh,\sigma=1} = \sqrt{\frac{\pi}{2}} \tag{3.39}$$

The variance of the Rice's distribution is given by:

$$var_{Rice,\sigma=1} = A^2 + 2 + \mu^2_{Rice,\sigma=1} \tag{3.40}$$

The variance of the Rayleigh's distribution is given by:

$$var_{Rayleigh,\sigma=1} = 2 - \frac{\pi}{2} \tag{3.41}$$

Rice computes also the expected number of maxima of the envelope for an ideal rectangular bandpass filter and it proves to be 0.6411 times the bandwidth. The same problem is analyzed in mathematical form in [**?**].

Jeffress [**?**, 24] using an envelope detector obtain the electrical analogue of the Rayleigh and Rice distributions. The envelope detector takes the form of a rectifier (either half wave or full wave) followed by a post detection filter (a capacitor shunted by a resistor) (see fig. 3.7). He studies the statistical value when the input signal is gaussian noise and when the input signal is deterministic. The results confirms that the the Rayleigh's distribution is a statistical model for the gaussian noise and the Rice's distribution is a statis-

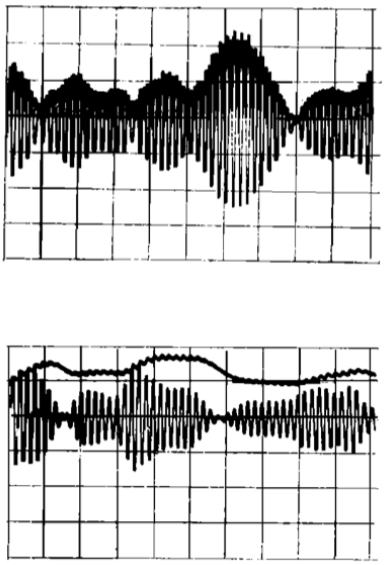


Figure 3.8: Performance of the post detection filter: (A) correct choice of constant time, (B) wrong choice of the constant time. From [**?**]

tical model for the deterministic signal. More in detail in [**?**] is showed that if the envelope is sampled at the instant when it has reached the maximum, a Rice distribution for envelope maxima is obtained.

In Jeffress's paper some observations about the optimal value of time constant RC in the post detector filter are proposed. He notes that in case of longer time constant the filter fails ti follow even the main individual fluctuations of the waveform (fig. 3.8) but the mathematical proof of this is not proposed.

Peterson et Al. [46] study the Rice's density function (eq. 3.36) using the signal to noise ratio $\frac{E}{N_0}$ (where E is the signal energy and N_0 is the power of noise) in form of 2TW samples (where W is the bandwidth and T is the observation time). Their resulting expression ($\sigma = 1$ is assumed):

$$f(a) = a\, e^{a^2+2\,T\,W}\, I_0(a\,\sqrt{2\,T\,W}) \tag{3.42}$$

For the ideal detector of the SKS case, Peterson et Al. show that the bandwidth of the filter being employed should be the reciprocal of the duration

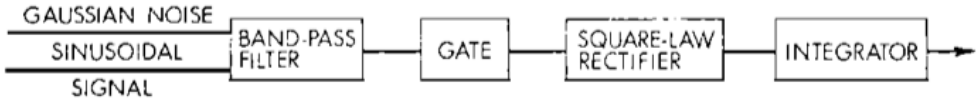


Figure 3.9: Block diagram of an energy detector. From [**?**]

of the sample of N or SN

3.3.2 Energy detection

Marrill [**?**] using a psychometric functions obtained in two-alternative forced choice [65, 66] derived the probability of being correct for the SKS detector:

$$P(c) = 1 - \frac{1}{2} e^{\frac{E}{2\,N_0}} \tag{3.43}$$

In 1966 McGill [**?**] shows (?) that the result of Marrill can be derived from the statistics of an energy detector (fig. 3.9), obtaining the expression for the probability density.

$$f(x) = \frac{1}{N_0} e^{-(\frac{E}{N_0} + \frac{x}{N_0}} \, I_0(2\frac{\sqrt{E\,x}}{N_0}) \tag{3.44}$$

which is a special case of noncentral χ^2 distribution with $\nu = 2$ degree of freedom (see later). The generic noncentral χ^2 distribution with ν degree of freedom can be expressed as:

$$f(a) = \sum_{i=0}^{\infty} \frac{\lambda\, e^{-\lambda}}{i!} \frac{\nu^{\nu+1} a^{2\nu+2i-1}}{2^{\nu+i-1}(v+i-1)!} e^{\frac{\nu\, a^2}{2}} \tag{3.45}$$

If $\nu = 1$ and non centrality parameter $\lambda = 1$ the eq. 3.45 corresponds to a Rayleigh's distribution, if $\nu = 1$ and $\lambda = \frac{A^2}{2}$ the eq. 3.45 corresponds to a Rice's distribution.

In fig. 3.9 the energy detector used by McGill is shown. Note the gate at the input of the filter and the integrator takes a true integral of the input. In this experiment for the first time a gate ($T = 1\,s$) is employed. McGill then derive Marill's equation **??**marrill:distribution using as ideal detector

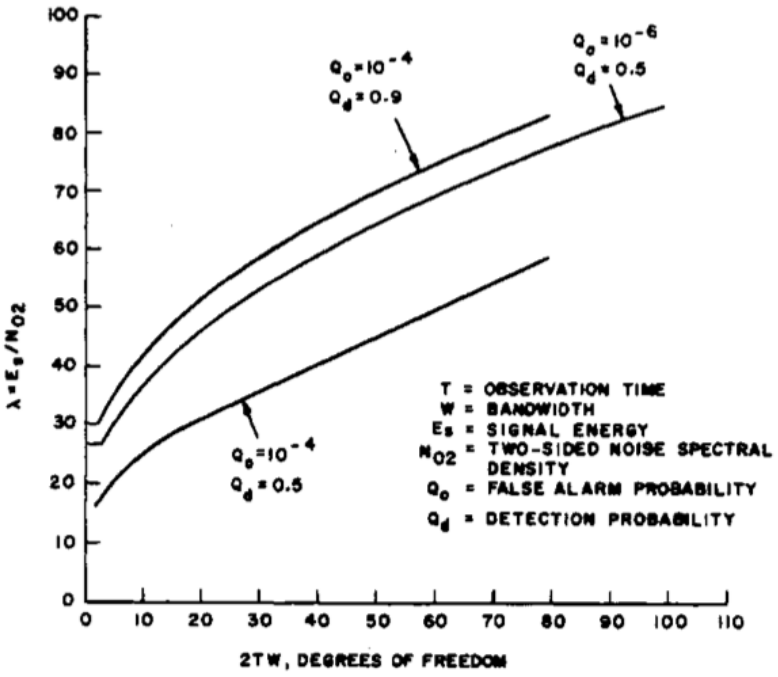


Figure 3.10: Detection probability curve in case of deterministic signal. The product 2TW of the observation is related with the degree of freedom ν of a non-central χ^2 distribution and with the non-centrality parameter λ. The curves represent different detection probability. From [67]

both energy detector and envelope detector, obtaining same results. Jeffress [?] discusses the choice of the time constant, basing on Zwislocki's work of temporal integration [?]. They find(s) that a 50ms integration time well approximates signals ranging from about 20 to about 100 ms. Shorter or longer integration time militate against longer or shorter signals, respectively [24]. It's the first evidence of an adaptive working (in terms of time integration and effective duration) of the auditory process. When duration of the signals exceeds $200ms$ the postdetection filter in fig. 3.9 saturates and further increases of duration have little effect to envelope results. Jeffress uses degree of freedom $\nu > 2$ ($\nu = 5, 7, 19$) to adapting the results of Marrill's experiments to statistical formulations. Urkowitz [67] discussed the same problem in more rigorous mathematical form. He show(s) that, in presence of deterministic signal, the distribution of the decision statistic at the output of

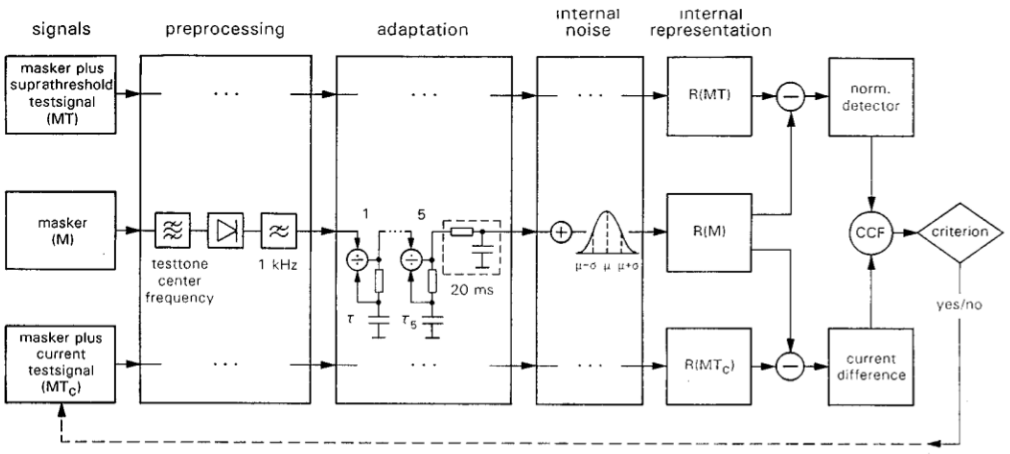


Figure 3.11: Dau's model of the internal representation of the acoustic perception. From [13]

the energy detector is a non-central χ^2 distribution with $\nu = 2TW$ degree of freedom and a non-centrality parameter given by ratio $\lambda = \frac{E}{2\,N_0}$ (see fig. 3.10). He show also that 2TW terms are sufficient to approximate the energy in a finite duration sample of a band limited process [67].

3.4 Interaural coherence and envelope coherence

At the high frequencies (above about $1.5KHz$) the phase locking in the responses of the inner hair cells decreases and the coding of the stimulus fine structure is lost [**?**, **?**]. At high frequencies only envelopes are represented in binaural and monaural [**?**] neural activity. The correlation process (autocorrelation and inter aural cross correlation) are based on envelope function, that in many auditory models is presented after the envelope extraction (e.g. in the Dau's model (fig. 3.11)). In this case, assuming that the internal representation (fig. 3.11) is given by autocorrelation and cross correlation function. The interaural envelope correlation (or *envelope coherence*) is given by:

$$cc_E(\tau) = \frac{1}{2T\sqrt{\Phi_{LL}(0)\Phi_{RR}(0)}} \int_{-T}^{2T} p_R(t)\, p_L(t+\tau)\, dt \qquad (3.46)$$

The autocorrelation may viewed as a particular case of eq. 3.11 ($p_R(t) = p_L(t) = p(t)$). The envelope coherence may be more interesting binaural measure than the coherence of the waveforms of those signals for high-frequency bands [**?**]. Some relation are founded between the envelope coherence and the cross correlation functions ($\phi_{LR}(\tau)$) [**?**, **?**]. Several analytical expressions for the envelope correlation are exposed in [**?**].

In the optics of this thesis some important differences should are considered: in the Ando's model the envelope concerns a monaural criterion of the sound, the analytical (statistical) expression proposed refer to to monaural and binaural criteria of the sound (with some differences). These relations may be COMUNQUE useful to relate the envelope meaning in the Ando's auditory model and envelope meaning in other auditory models.

Chapter 4

Some consideration about the τ_e extraction

4.1 References from literature

In the field of room acoustics the ACF is usually defined as [28]:

$$\Phi(\tau, t) = \frac{1}{2T} \int_{t-T}^{t+T} p(\xi) p(\xi + \tau) d\xi \tag{4.1}$$

where p(ξ) is the A-weighted sound pressure signal, 2T is the width of the temporal window, τ is the time delay of the ACF. The normalized ACF has the form:

$$\phi(\tau, t) = \frac{\Phi(\tau, t)}{\Phi(0, t)} \tag{4.2}$$

Where $\Phi(0, t)$ can also be seen as the energy of the signal inside the temporal window. The 'effective duration' of the ACF (τ_e) is defined here, following Ando [3], as 'the first ten-percentile of envelope decay of the normalized running autocorrelation function' expressed in dB.

4.1.1 Linear part

Many physiological models include an additive noise term in the internal representation of a sound signal[60]; this is equivalent taking into account,

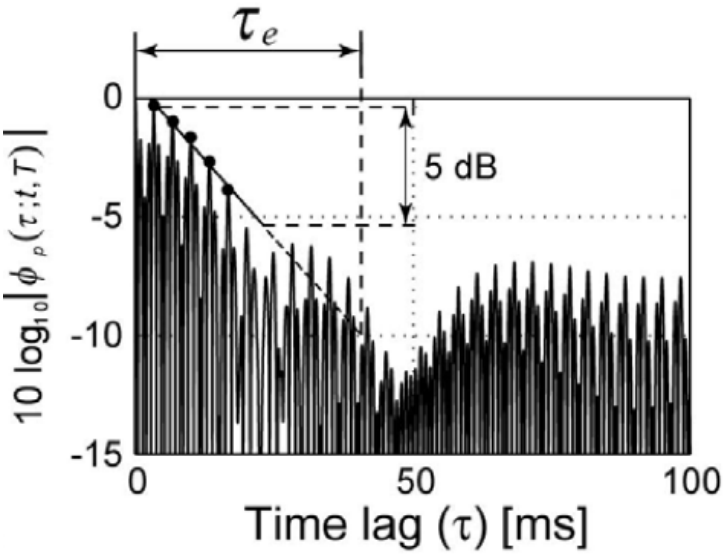


Figure 4.1: Dynamical and temporal definition of ACF initial part. From [28]

in the analysis of correlation, a threshold under which the decay is no more meaningful [60]. In the case of τ_e this threshold is set to -10 dB by definition [3]. Furthermore the ACF envelope is linearized over the first 5 dB.

differenza tra regione lineare e interpretazione del rumore di fondo.

4.1.2 Temporal window of integration

The discussion on the integration time of the running ACF should be concern both envelope and fine structure analysis. Different values are proposed in the two fields.

On one hand, the fine structure analysis of r-ACF requires low values of integration time. Licklider [32] originally suggests using a running ACF with a narrow temporal window shifted along time axis. This windows is represented by a low-pass filter (exponential window) having a time constant of $2 - 3ms$. [formulazione ACF di Licklider, magari richiamarla soltanto] Wiegrebe [71] proves evidence that multiple time constant are required, each proportional to the autocorrelation lag being evaluated. A value of twice time the the lag may be appropriated in case of IRN stimuli. Following

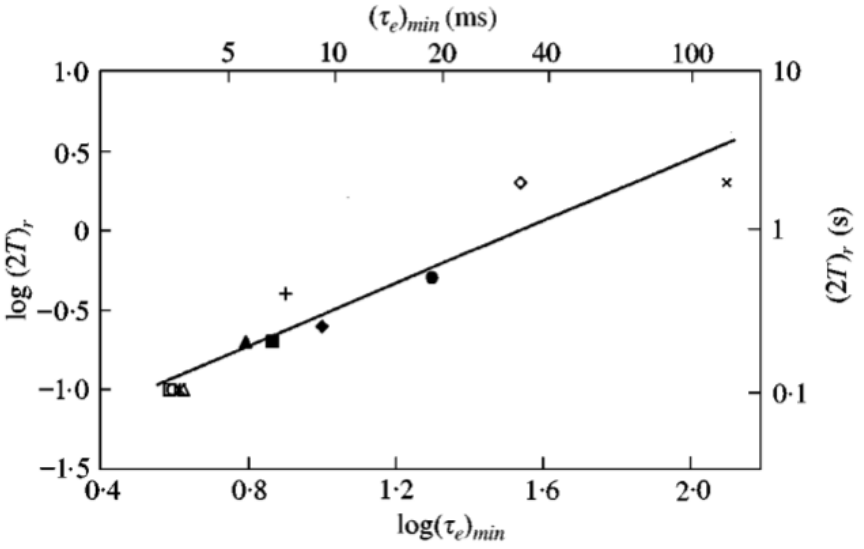


Figure 4.2: Logarithmical relationship between the width of the temporal window 2T and the mininum of τ_e. From [40]

the same rule, time constant between 40 and 4 ms would be appropriated for pitches in range $50 - 500Hz$. Some stimuli requires longer integration times and the author proposes to extend the model by passing the SACF through a low pass filter ($LP - SACF$). Some relations between lag window and physiological mechanism are resumed and discussed in [7], in which a variable 'lag window' integration of ACF is proposed.When the stimuli (e.g. aperiodic click stimuli) contains fluctuations whose durations are longer than the time constant other models can be used, with longer integration window which stabilize the output function [**?**].

On the other hand, the envelope analysis of r-ACF requires high values of integration time. Ando, using temporal window without low-pass filter, proposes initally [1] the use of long time temporal window ($2T = 30s$). Varying the amplitude of temporal window, he notes [4] that for $2T < 500ms$ too many fine structures appear and the τ_e cannot correspond to the instantaneous subjective response. The results with $2T = 2s$ may assumed as applicable are closely correspondent to those of long time ACF ($2T = 30 \div 35s$). Following the Fraisse's definition of "temporal present" [16], Ando suggests

a value of about 2s. All the signals presented to listeners in [4] contains music. Mouri et Al [40], studying different type of signals (voice, music, noise) find(s)? a relation between width of the temporal window and $(\tau_e)_{min}$. The regression time of the fig. 4.2 can be expressed approximately by the equation:

$$(2T)_r \approx 30\,(\tau_e)_{min} \tag{4.3}$$

(ANALOGIA con il feedback proposto da Jeffress)

The relation 4.3 is commonly used in the running analysis of the ACF. A confirm of that is proposed in figg. **??**, **??**. The minimum of τ_e of two motif from literature [9] is evaluated. Peak detection method (par.) is used with different width of the temporal window.

4.2 Peak detection method

In the so called *peak detection method* the ACF peaks located over the threshold are linearly interpolated and the abscissa of the -10 dB point on the interpolation straight line estimates τ_e [3]. In the peak detection method it is not clear whether all peaks should be taken into account or if only a subset of them are meaningful for the envelope analysis [4, 28]. In figure **??** three τ_e values are calculated using the peak detection method with different selection criteria of the 'relevant' peaks. Only the third selection criterion (fig. **??**) may return the correct τ_e because it considers only the local maxima which contribute to the envelope of the ACF. In fact, a correct peak selection is necessary for a good approximation of the envelope in the sense of Ando's definition of τ_e, but the selection criterion is not well defined in literature. Some methods, developed for pitch extraction from ACF [56], introduce the idea of the 'relevant peaks' but these methods are not focused on the envelope extraction but only on the first maximum of the ACF.

Furthermore, the lowest evaluable value of τ_e is another issue which should be considered. The methods for the evaluation of τ_e reported in literature are mainly related to musical motifs. In case of musical signals the expected value of τ_e ranges from $30ms$ to $300ms$ (and more). For environmental noise

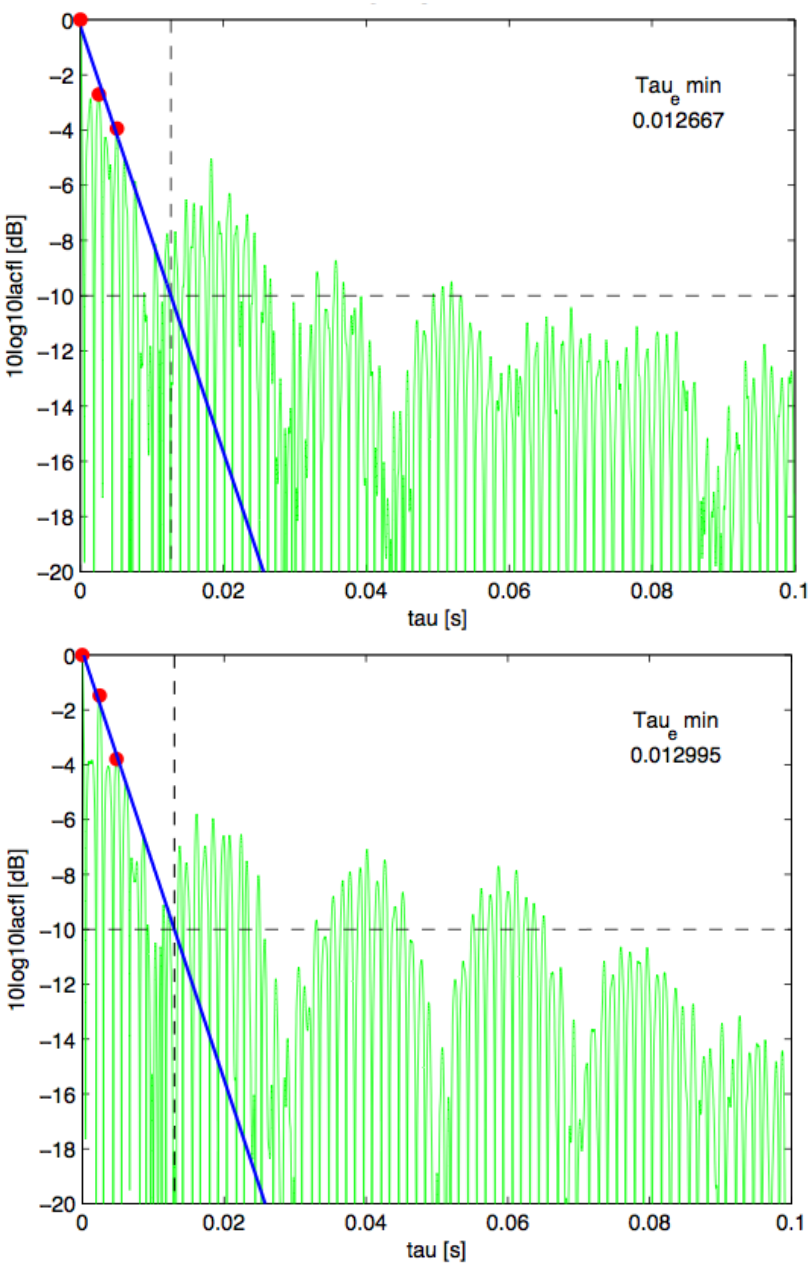


Figure 4.3: $(\tau_e)_{min}$ of mothif 1 with different width of temporal window. (A) 2T=250 ms (B) 2T=2000 ms

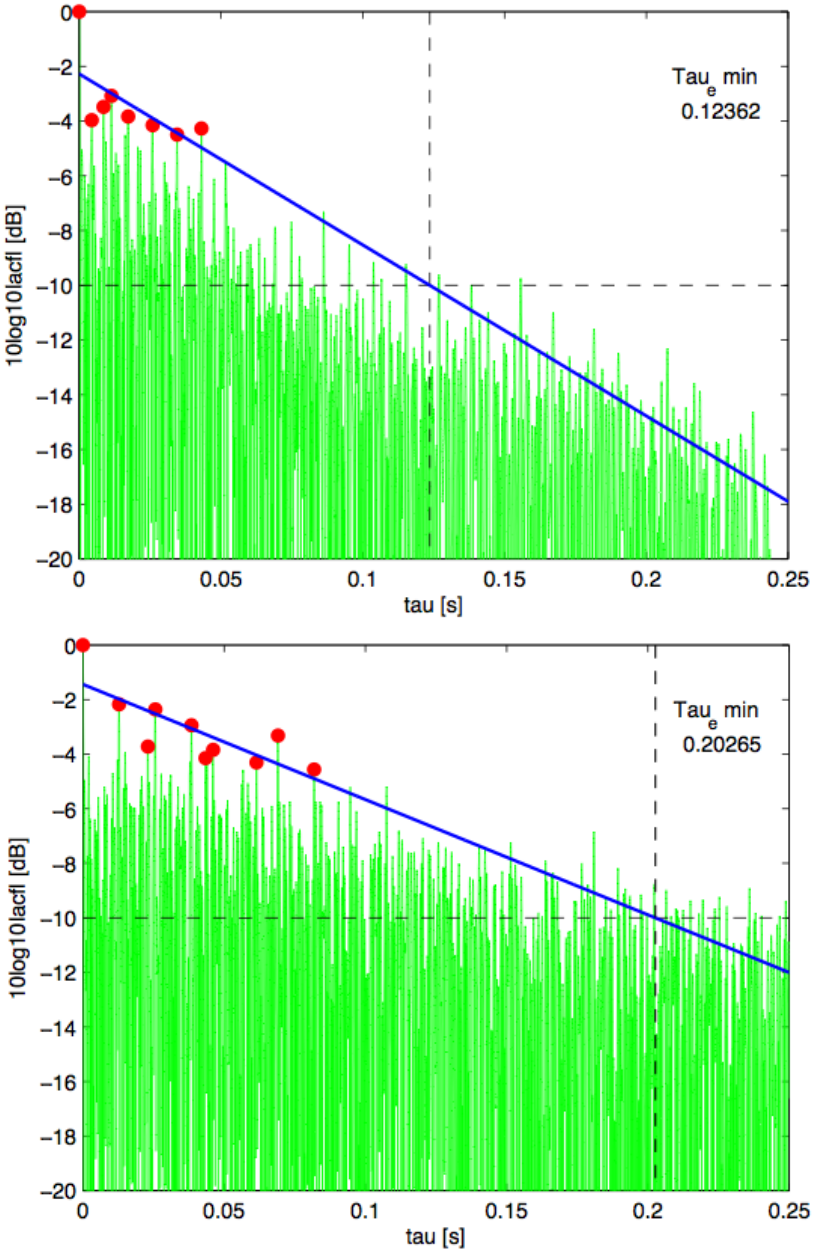


Figure 4.4: $(tau_e)_{min}$ of mothif 2 with different width of temporal window. (A) 2T=250 ms (B) 2T=2000 ms

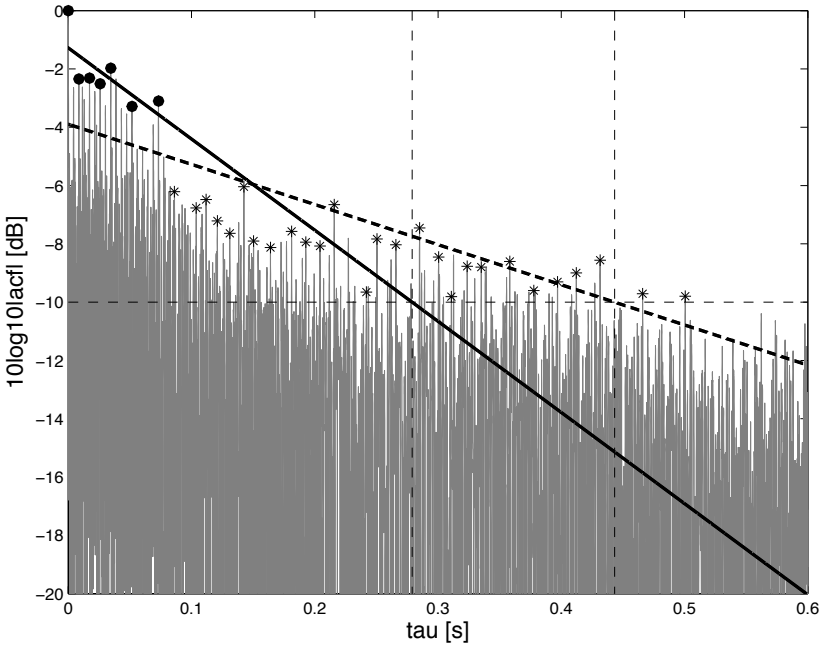


Figure 4.5: Calculation of $(\tau_e)_{min}$ with different threshold of linearization

analysis or impulsive signals the value of τ_e can be 15 ms or less [62]. In the peak detection method the -5 dB threshold may produce meaningless results in case, e.g., of impulsive signals or signals with inharmonic features [61]: only one peak located in the origin (at $\tau \approx 0$) may be detected and the relative interpolation is not possible. In these cases other methods should be considered if the scope is to generalize the use of τ_e analysis (*e.g.* in soundscape analysis, ...).

4.3 Schroeder's backward integration

A backward integration method, analogous to Schroeder's one [52], has ben proposed for the evaluation of ACF envelope [4, 28]. The decay function $D(t)$ can be obtained with the eq. XX (vedere cap prcedente). The decay curve, linearly fitted from 0 dB to -5 dB, extrapolated to -10 dB, permits to extract the τ_e value (fig. 4.11). Let's note that in this case the ACF should be expressed in non normalized form (Φ); in all other methods presented in

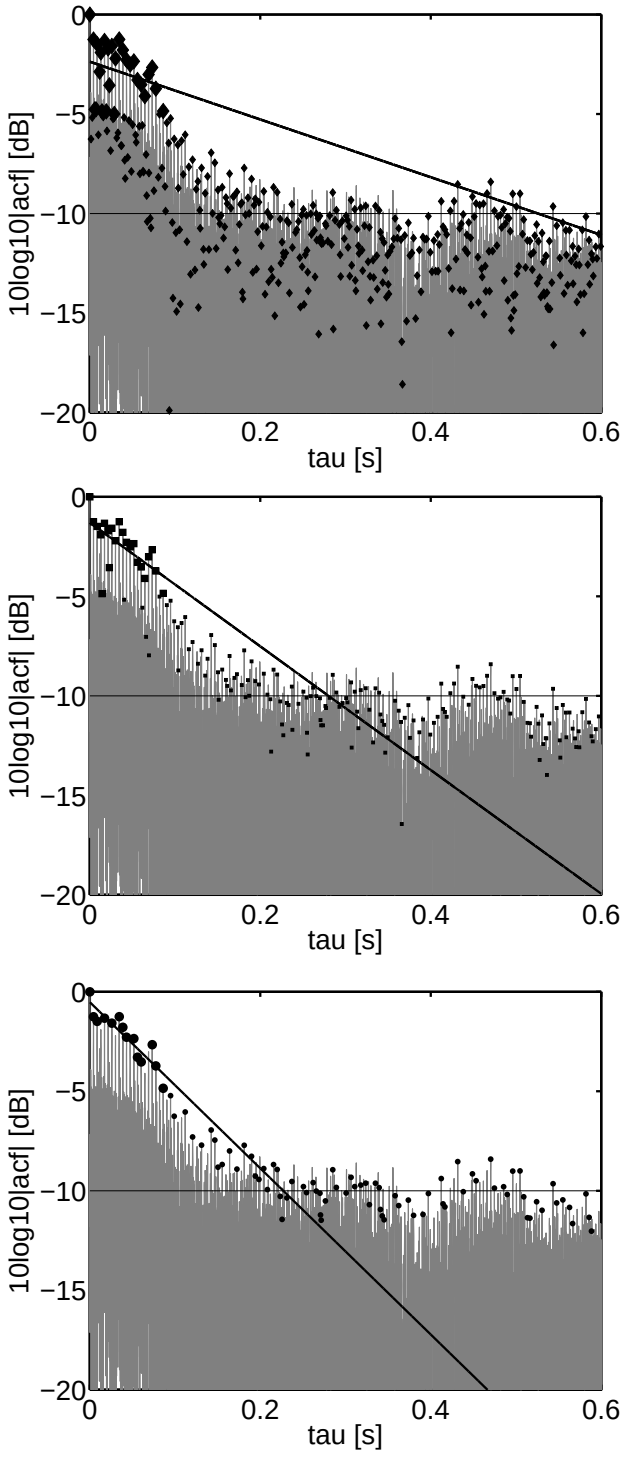


Figure 4.6: Decay of the ACF with three arbitrary criteria of selection, arbitrary motif. Criterion (A): τ_e=526. Criterion (B): τ_e=280. Criterion (C): τ_e=227 ms.

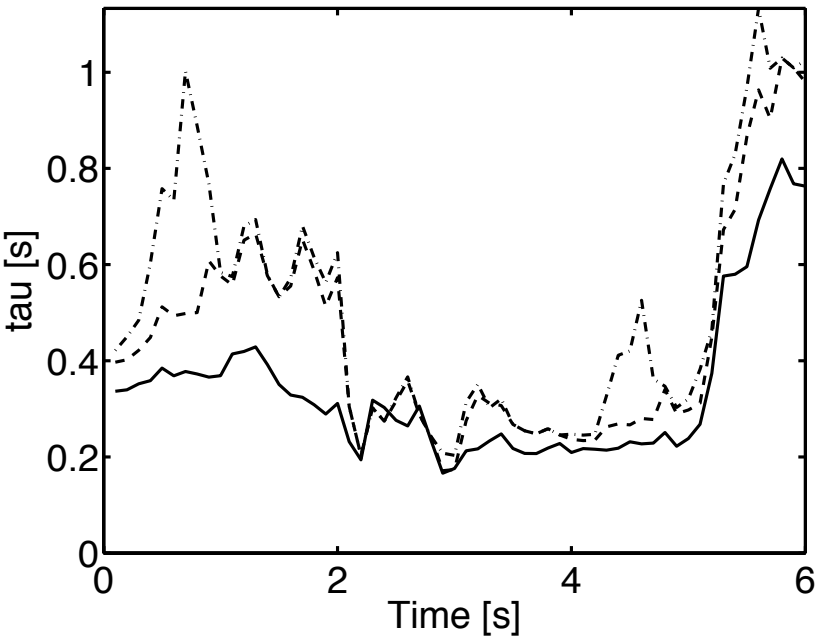


Figure 4.7: $\tau_e(t)$ curves relative to different criteria of peak selection

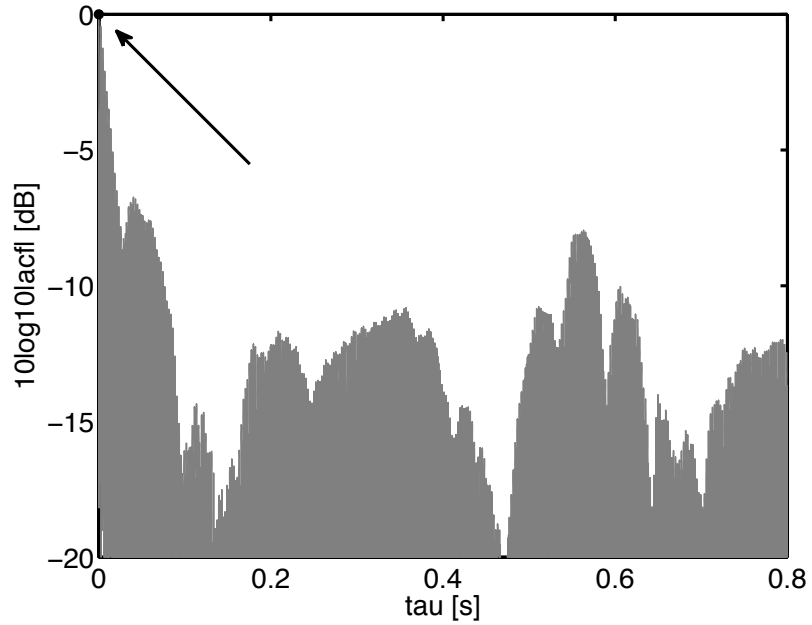


Figure 4.8: Peak detection in case of low SNR or impulsive signals. Only one peak is evaluated and the peak detection method may be unusable

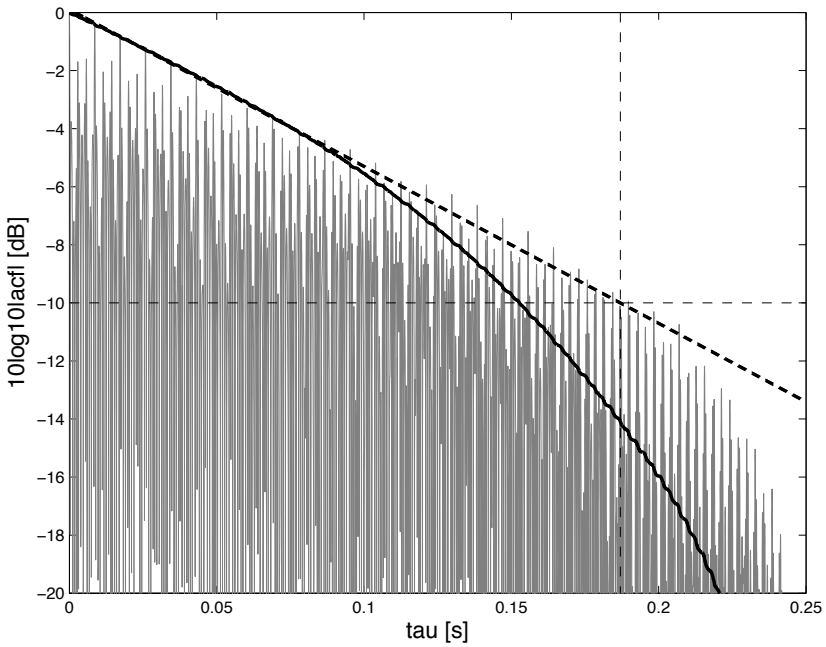


Figure 4.9: τ_e extraction from Schoeder's decay. Schroeder's Backward integration curve (black curve), straight line fitted over the first 5 dB of the Schroeder's decay (dashed line). 2T = 2000ms, ULI = 250 ms. The method in this case performs a correct τ_e extraction.

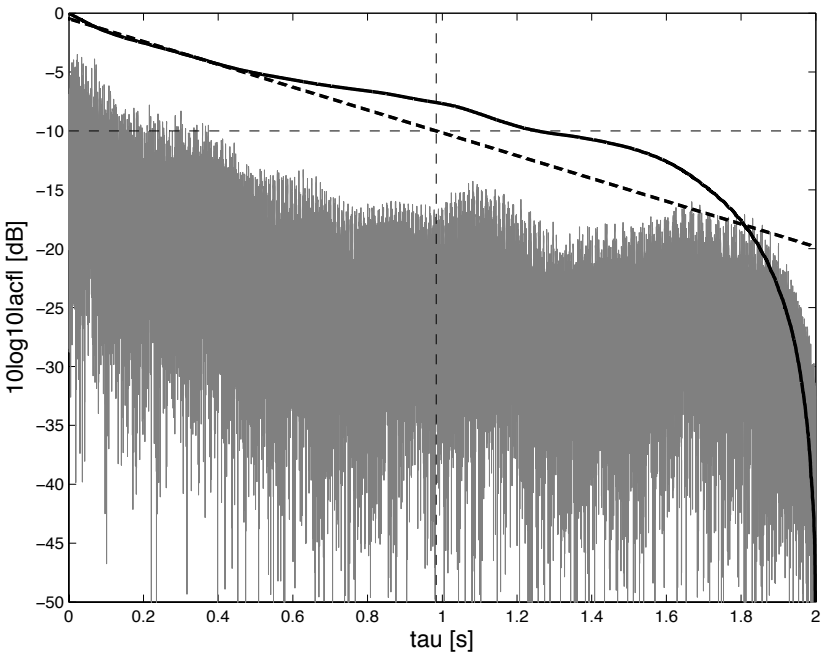


Figure 4.10: τ_e extraction from Schoeder's decay. Schroeder's Backward integration curve (black curve), straight line fitted over the first 5 dB of the Schroeder's decay (dashed line). 2T = 2000ms, ULI = 250 ms. The method in this case performs a wrong τ_e extraction.

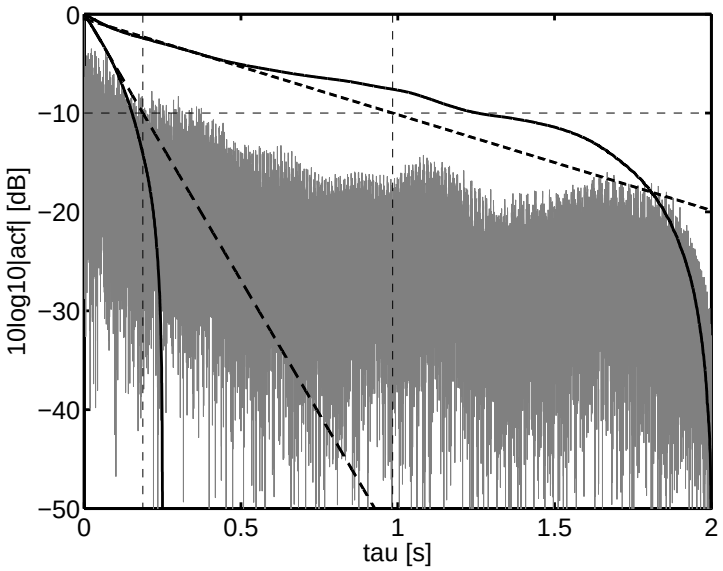


Figure 4.11: Comparison of the two precedent cases (figg. [**?**],[**?**]). Schroeder's backward integration method with two different initial times. Gray curve: normalized ACF; left solid curve: decay function with $ULI = 250ms$; left dashed line: interpolation of the decay function ($ULI = 250ms$) over the first 5 dB; right solid curve: decay function with $ULI = 2s$; right dashed line: interpolation of the decay function ($ULI = 2s$) over the first $5dB$.

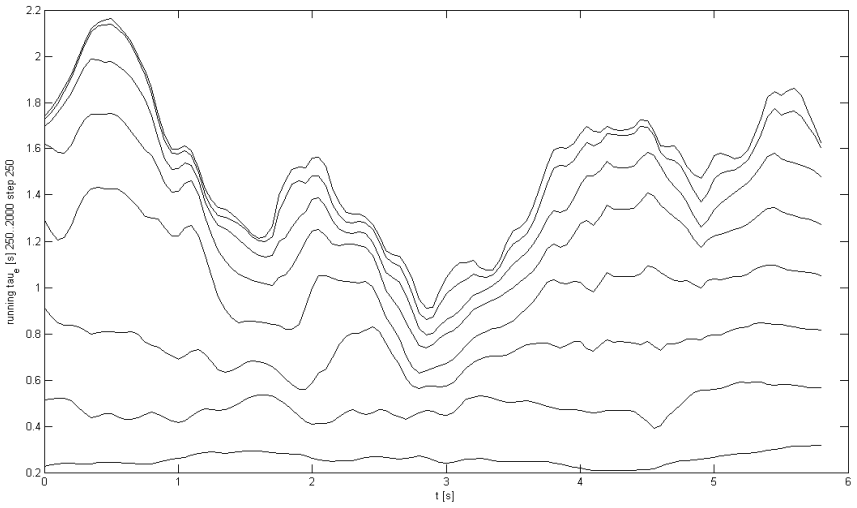


Figure 4.12: Schroeder's backward integration method with stepped values of ULI (from 250 ms to 2000 ms with step of 250ms). The curve are quite dependent from the ULI choice and not return any information of the correct $\tau_e(t)$ curve

this work the ACF is used in its normalized form (ϕ) (in figure 4.11 the ACF is presented already normalized, after the calculation of the backward integration curve, for the sake of uniformity with the other methods).

The decay function (the approximation of the ACF envelope) doesn't have a slope similar to that of the envelope of a room impulse response. So, there is a difference between the accuracy of the backward integration method when applied to a room impulse response and to the detection of the ACF envelope. A critical point is the choice of the initial time of the backward integration: the evaluation of the ACF envelope can be wrong if this initial time is not properly selected (see right curve in figure 4.11).

The phenomenon is noted also by Kato et Al. [28]. In the same article it is proposed to apply the integration to the 'initial part of ACF' only: the initial part is fixed in 50 ms according to evaluation of a vocal musical signal. For a general usage this choice can limit the applicability of the technique, e.g. in case of 'slow' musical motifs or stationary tonal signals when high τ_e values are expected [62].

Some considerations may be proposed, basing on the evaluation criteria for the decay estimation proposed in par XX and on statistical property of signals proposed in par XX

Some criteria for a correct decay estimation based knowledge of noise floor lever are proposed in literature [30, 69]. These methods suggest that the lower level of the decay in which the Schroeder's backward integral is calculated should be 10 dB above the noise floor level. In the case of integration of an autocorrelation function the background noise may be related to a threshold of $-10dB$, as noted in section **??** [60]. In case of ACF the noise floor level is located at the same level of the lower level of the decay: that may not permit a good evaluation of the decay [69, 73].

Some criteria for the choice of the initial time of backward integration have been suggested in literature (par XX). In these methods the choice of the initial time of the integration of a room impulse response may be related to knowledge of noise floor level and statistical hypothesis on the signal [**?**].

The Morgan's statistical analysis of the impulse response decay could be not applicable in case of autocorrelation function. In case of impulse response thus, the statistical process that can describe the decay is a Rayleigh's distribution, as proposed in [**?**]. In case of deterministic signal, as resumed in par XX, the statistical distribution that can describe the decay is a Rice's distribution (par XX). The two distributions may be generalized as two special cases of non central χ^2 with degree of freedom $\nu = 1$ and non centrality parameter $\lambda = \frac{SNR}{2}$ in case of autocorrelation decay (Rice's distribution), and $\lambda = 1$ in case of impulse response (Rayleigh's distribution). Let's note that in case $SNR = 0.5$ (or negative value in case of logarithmic expression of SNR)the two distribution could correspond and the Morgan's method may be useful for an evaluation of a correct ULI. This case is not possible in case of musical or speech signal but may possible in case of noise analysis. In the latter case, moreover, the expected value of τ_e should be so low and the shape of the expected decay should be so high that the problem in analysis (the correct choice of the ULI) falls in meaningful. (controllare grammatica)

In Schroeder's original paper of backward integration [52] a gaussian like signal is supposed. When apply the backward integration to deterministic

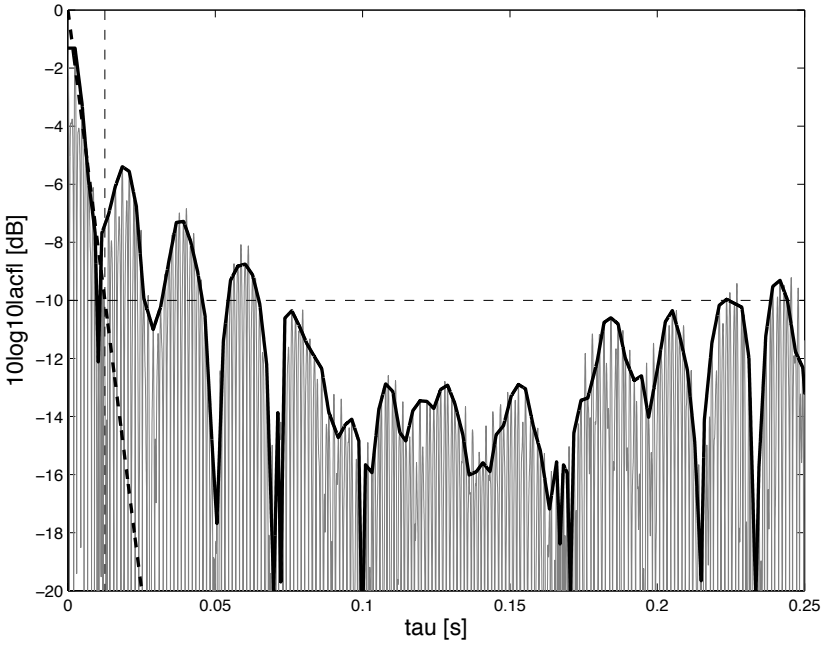


Figure 4.13: τ_e extraction based on Hilbert's transform (motif 2)

signal the proof of Schroeder's method falls and the decay function so calculated could not be assumed as convergent to the decay of the signal.

The Kato's proposal [28] of the integration of the 'initial part of ACF' only may not be useful for a general usage. The choice of to integrate only the first 50ms of the ACF can limit the applicability of the τ_e extraction technique, e.g. in case of 'slow' musical motifs or stationary tonal signals when high τ_e values are expected [62].

4.4 On the use of Hilbert's transform in extraction

Ando [4] proposed the use of the magnitude of HDT for the evaluation of the ACF envelope. Although the HDT is an accurate method for envelope extraction in signal processing, this method has been only proposed by Ando [4], but not used for the ACF analysis, due to its heavy computational load. An algorithm for the τ_e extraction from the Hilbert's envelope is implemented

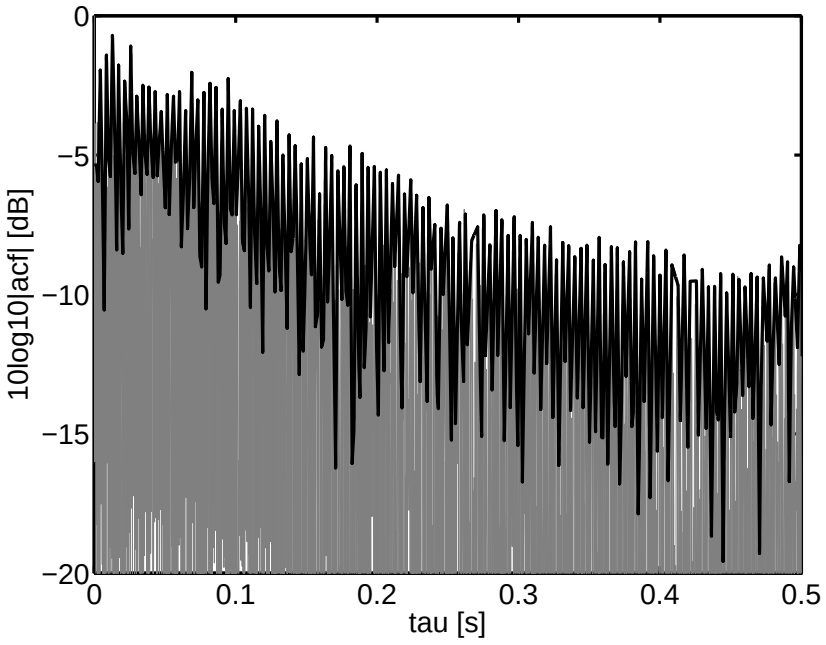


Figure 4.14: τ_e extraction based on Hilbert's transform (motif 1).

by the author of this thesis [**?**]. In some cases the Hilbert's envelope produces good results (fig. 4.13). In other cases the same method not produces useful results and a τ_e value is not evaluable (fig. 4.14).

Some considerations may be proposed. As shown in par XX a sufficient condition for the envelope extraction with Hilbert's transform is the minimal phase of the signal. The Hilbert's envelope extraction is commonly used on modulated signals, *e.g.* the single side band amplitude modulated signals ($SSB - AM$) (paragrafo kumaresan). It's possible to show that the autocorrelation of an AM signal is still an AM signal. The autocorrelation of a QFM signal is still not a QFM signal. The signal in input to the autocorrelation process is the output signal of the cochlea. Some authors suppose that the signal at the output of the cochlea is a minimum phase signal [**?**]. Other authors [32, **?**] suppose an AM signal at the input of auto correlational process. As shown in par.XX with some some hypothesis an AM signal may be supposed a minimal phase signal but these hypothesis depend from the frequency composition of the audio signal and are not generally verified (?). If the signal is mixed phase the eq. () may not be verified (?) and the

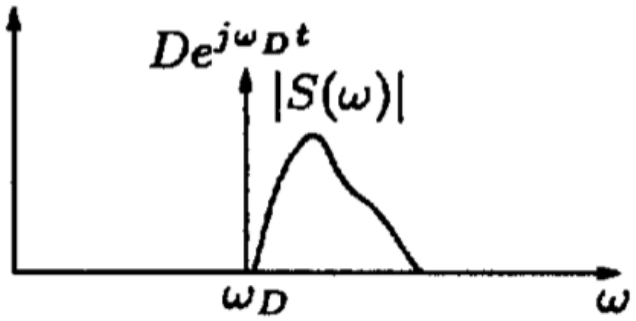


Figure 4.15: Sine wave added at the lower band-edge of a mixed phase signal: the signal sum of the two ones is a minimum phase signal. From Kumaresan95

Hilbert's envelope extraction falls. Moreover: the characteristic of minimum phase, present in the original signal, may be lost with the A-weighting or other filtering operation realized by an IIR filter: the filtered signal may be a mixed phase signal also when the original signal is a minimum phase signal. This fact may explain the ambiguity in Hilbert's envelope running, shown in fig **??**.

Some authors proposed mathematical method to transform a generic signal in a minimum phase signal [**?**]. Given a mixed phase signal $s(t)$ whose envelope may not calculated with Hilbert transform, consider a signal $s_{MP}(t)$ obtained by adding a sine wave a the lower band-edge of s(t):

$$s_{MP}(t) = s(t) + D\, e^{j\omega_D\, t} \tag{4.4}$$

It's possible to proof that $s_{MP}(t)$ is a minimum phase signal whose envelope may be described by Hilbert Transform. The eq. 4.4 could authorize the envelope detector Hilbert's transform-based shown in fig. 4.16. Moreover this method don't have a sufficient physiological support and should not be used.

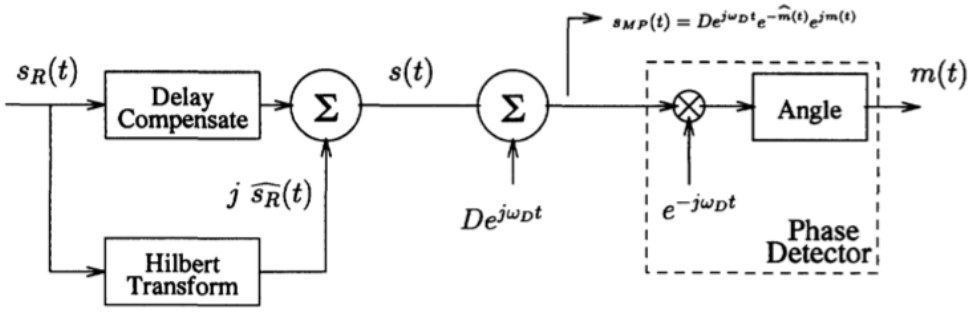


Figure 4.16: Algorithm to transform a mixed phase signal in a minimum phase signal. From Kumaresan95

Chapter 5

A new proposal for the τ_e extraction

All the previous methods can be interpreted as envelope extractors: in case of Schroeder's backward integration, the decay function is the envelope; in case of the Hilbert's transform method, the magnitude of HDT is the envelope; in case of the peak detection method, the interpolation straight line is a particular form of envelope due to the hypothesis of linearity of the decay. The linearization of the envelope returns the τ_e value.

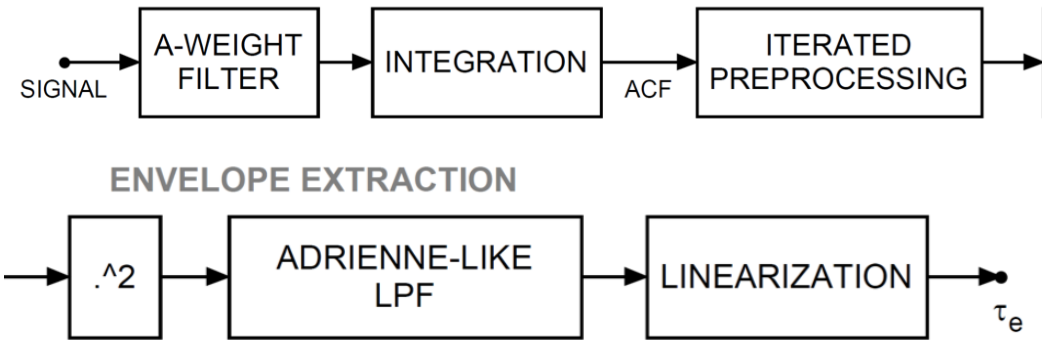


Figure 5.1: General scheme of the new proposed method

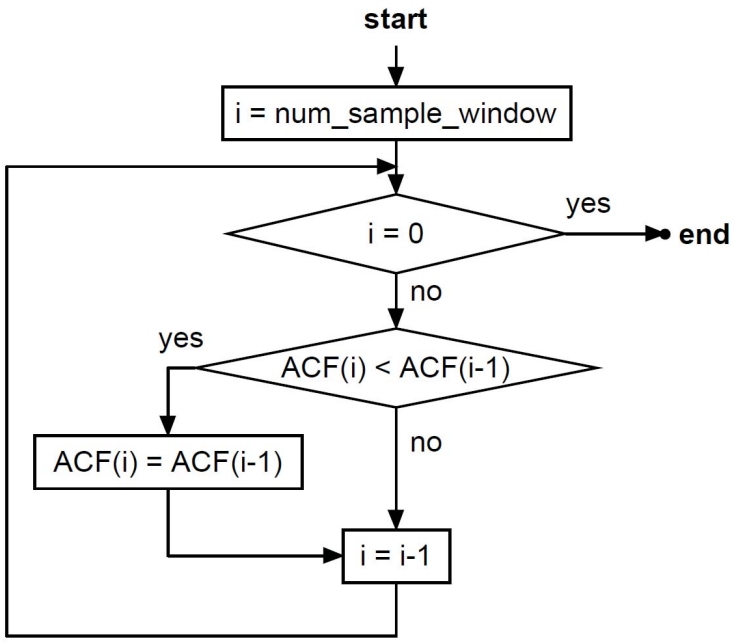


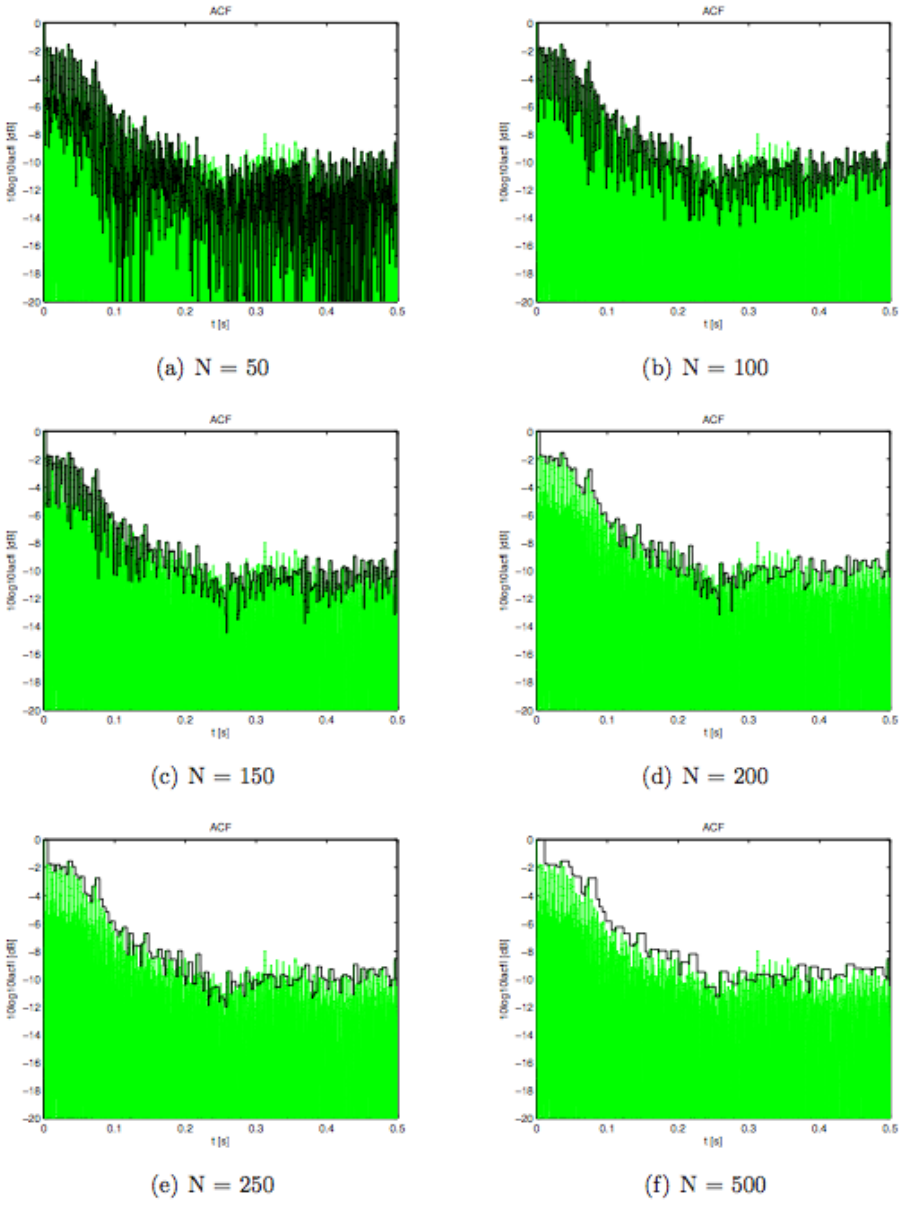
Figure 5.2: Flow diagram of the iterated pre proccesing sharpening

5.1 Energy detecion

5.1.1 Iterated sharpening

The flow diagram of the new proposed method (fig. 5.1) shows the differences with the previous ones. The ACF of the A-weighted signal is preprocessed by an iterated sharpening process. The basic algorithm of this iterated sharpening is as follows (see figure 5.2): at each iteration, the samples are compared two by two, from right to left, and for each comparison the smaller value is replaced by the higher. Since the autocorrelation function decreases with time, the algorithm proceeds from right to left on the temporal axis. In the case of musical signals, with a sample rate of 44 100 Hz, the number of iterations which permits to obtain an almost correct envelope is 200 (see fig. 5.3).

analogie con il preprocessing in [57, 8]



(a) N = 50 (b) N = 100

(c) N = 150 (d) N = 200

(e) N = 250 (f) N = 500

Figure 5.3: Envelope of ACF in function of sharpening iteration number

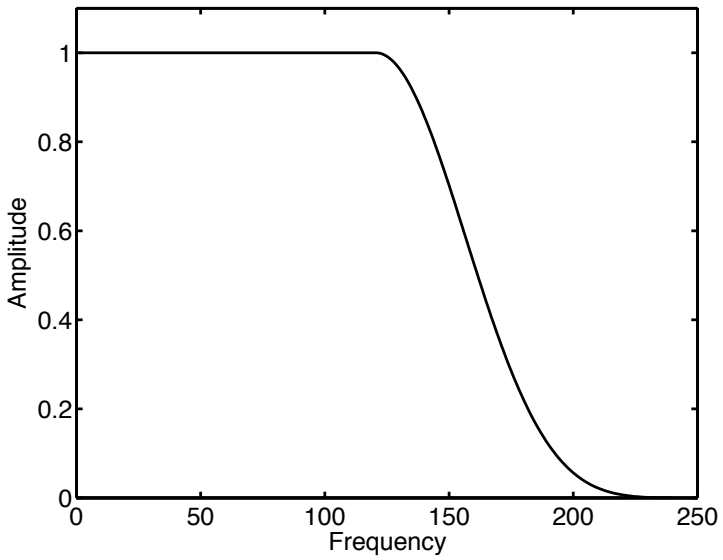


Figure 5.4: Shape of the postdetection filter (modified Blackmann-Harris frequency window)

5.1.2 Post detection filtering

After the preprocessing, the envelope extraction in this new method is based on a 'energy detection' method: a quadratic operator followed by a low pass filter (fig. 5.1). The 'energy detectors' in the early literature have been realized in the analog domain [67, 24] and the shape of the filter depended on *RC* cells. In this work a digital filter with an emphAdrienne-like frequency window is used [18]: the amplitude of the filter is unitary in the range $0 \div 120$ Hz and decays like a Blackman-Harris filter [20] for the next 120 Hz (range $120 \div 240$ Hz). MOTIVAZIONI SULL'ORDINE DEL FILTRO

MOTIVAZIONI SULLA CAPACITA' DI SCENDERE IN DINAMICA (mettere immagini dalla tesi di simona, su progressivo adeguamento del numero di iterazioni).

NOTE SUL FILTRO RETTANGOLARE

The shape of this filter permits a good envelope of the local maxima located in the first part of the ACF decay and permits also to have a useful range of decay of 25 dB [20]. The calculated envelope is linearized, like in the previous methods, over the first 5 dB of the decay and the τ_e value is

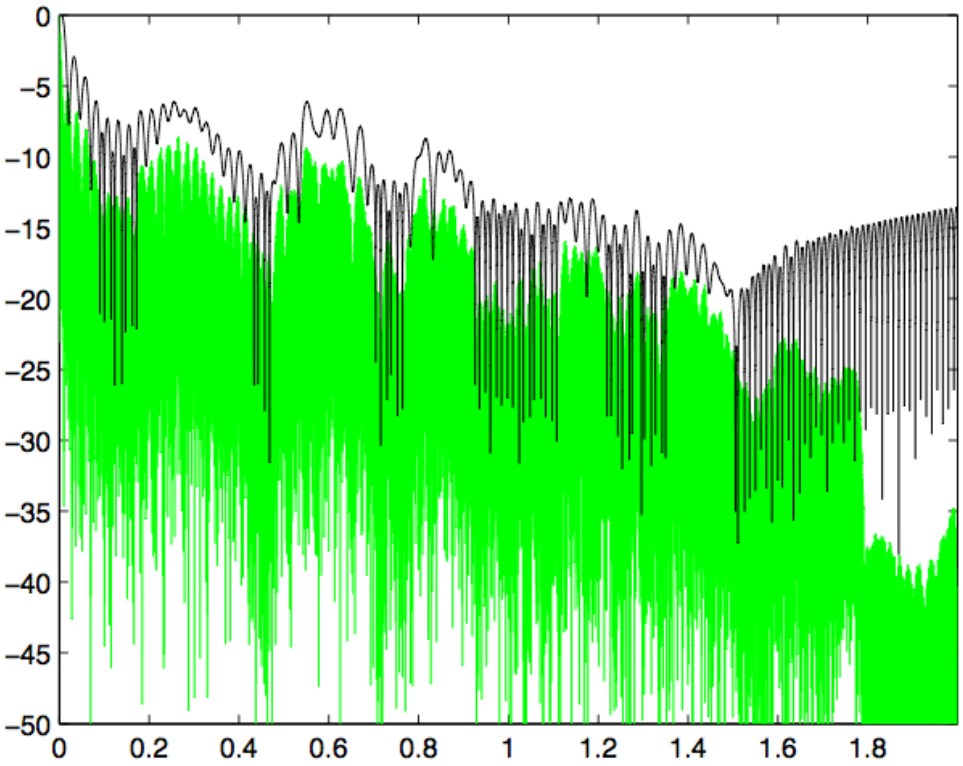
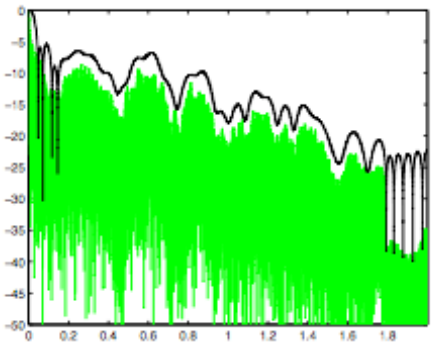
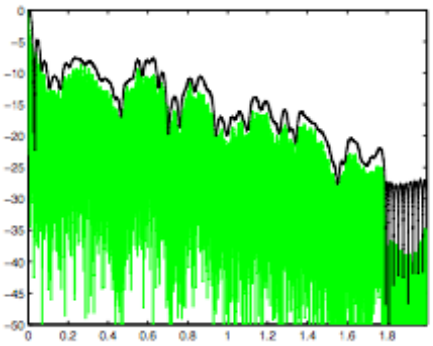


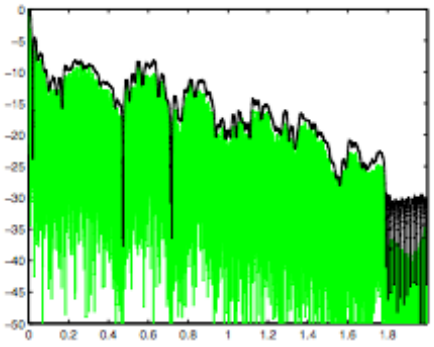
Figure 5.5: Envelope of ACF in case of a rectangular shaped post detection filter



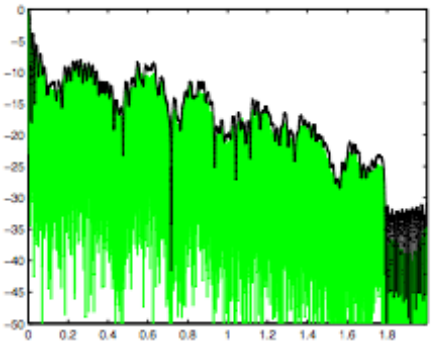
(a) N = 60



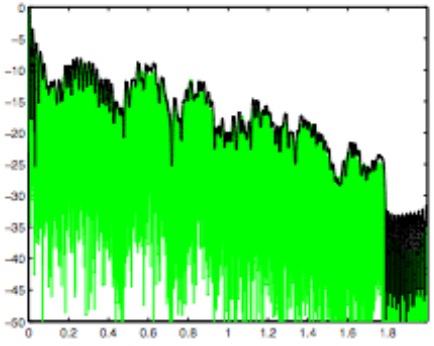
(b) N = 120



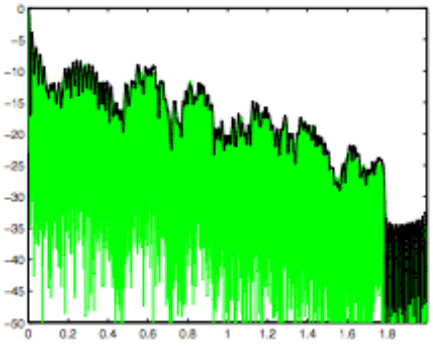
(c) N = 180



(d) N = 240



(e) N = 300



(f) N = 360

Figure 5.6: Envelope of ACF in function of sharpening iteration number

extracted.

5.2 Experimental results

Two reference motifs (tab. 2.1) from literature have been used for a comparison between the new proposed method (ACF version) and the previous ones [4, 28, 9]. As noted in section **??** and also by Kato [28], the peak detection method is the only method from literature for which numerical results have been published, so it is used here as reference. In figure **??** the running τ_e of motif 1, calculated with the two methods (peak detection vs new method based on ACF), is presented. As can be noted, the two curves of the running τ_e are quite similar and the two minimum values are comparable and correspond to the same temporal window. In figure **??** the relative decay of ACF generating the minimum value of τ_e ($(\tau_e)_{min}$) calculated with the new method based on ACF is presented. The decay of ACF generating the minimum value of τ_e ($(\tau_e)_{min}$) evaluated with the peak detection method actually returns a value different from that given in the literature ($15ms$ instead of $35ms$ [4], see fig. 5.8).

5.3 Discussion

Some considerations may validate physiological plausibility of this method. The energy detection describes in general sense the action of the hair cell in that the vibratory pattern of the basilar membrane produces neural impulse (squaring, but also the half wave rectification may used, with different results) and the membrane properties limit the ability of the neurons to phase lock to high frequency vibrations (low pass filter) [**?**].

Preprocessing the data with the sharpening algorithm before the envelope extraction improves the correct evaluation of the envelope, which otherwise could be overestimated. The sharpening block and the particular shape of the low pass filter realize the 'proper' envelope, in the sense discussed by Jeffress [24] (see fig. 3.8). The sharpening proposed in this method must

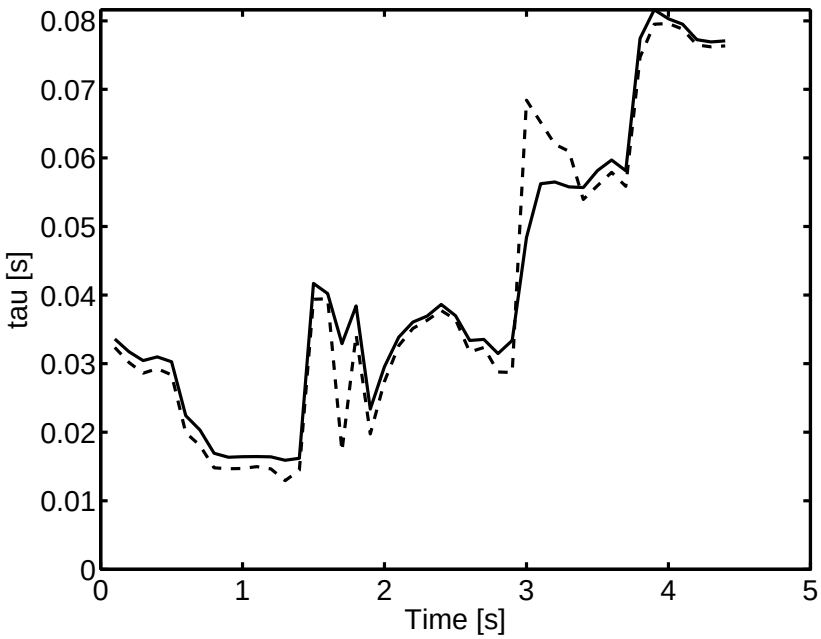


Figure 5.7: $\tau_e(t)$ curves: comparison between peak detection (dashed line) and new proposed method (black line). Motif 1. $2T = 2s$

not be confused with the sharpening block present in auditory model [14, 3]. The sharpening of the present model may be considered as a preprocessing to maximize the performance of energy detector. In this meaning may be associated to other non linear preprocessing-based algorithm that concerns the ACF [8, 57]: the last ones concerns the fine structure, the presented one concerns the envelope. The nonlinear processing of iterated sharpening changes both spectrum and statistical characteristics of the acf. The second ones may be analyzed with the theory SKS (signal known statistically) shown in par. XX. A deterministic signal (as the ACF function) may be described as a Rice's process. The envelope extraction of a Rice's process needs an energy detector with an high order of degree of freedom ν. The degree of freedom ν of the postdetection filter in the energy detector must be proportional to the product $2WT$ of the Rice's process (where T is the temporal window and W is the bandwidth of the signal) (eq. 3.42). The complexity of the filter depends QUINDI to the temporal window e to the bandwidth. The iterated sharpening, in this optics, may change the statistical characteristics

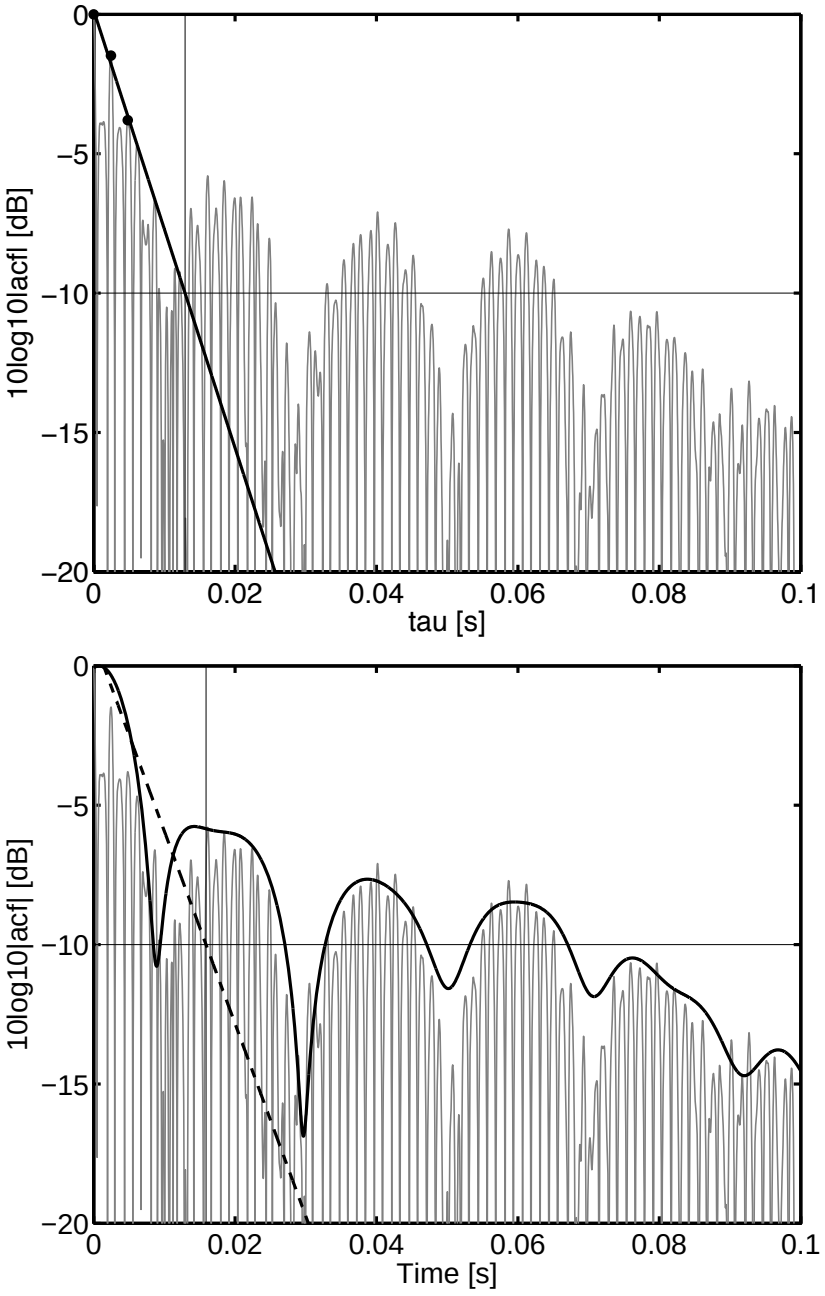


Figure 5.8: Minimum value of τ_e evaluated with peak detection (A) and new proposed method (B). Motif 1. $2T = 2s$

of the ACF 'centering' the χ^2 distribution (as a Rayleigh statistics) and pulls down the needed complexity of the filter. In the proposed system a relative simplex shape of the filter (chosen for his good performance with secondary lobes) permits to extract the correct envelope. A confirm of that may be represented to the different performance of the algorithm to different signal, different temporal window and (see later) the better performance in case of bandpass filtered ACF. In optics should be interpreted the directions proposed in chapter XX: 200 iterations may be indicated in case of music with a $2T = 2s$. Another limit of the system is represented by the minimum evaluable value of τ_e, that depends on the choice of the cut-off frequency of the filter, with a reciprocity principle: increasing the cut-off frequency of the filter the main lobe of the envelope decreases and permits the evaluation of lower values of τ_e. On the other hand an higher cut-off frequency of the filter may not permits a *proper* envelope detection. The *optimal* number of iterations of the sharpening algorithm depends also on the shape of the filter, so an adaptive variant of the choice of this number has been tested: in this case the cost function of the methods has been related to the distance between the peaks. The resulting envelope doesn't differ significantly from the other one (with a fixed number of iterations) and the processing time is higher: so the algorithm with a fixed number of iterations has been retained in the present work.

Chapter 6

On the use of the decimated Summary Auto Correlation Function

6.1 Definitions

Further modifications to the ACF model proposed by Licklider have been suggested in the literature. The same Licklider in his triplex theory extends the autocorrelational process in a binaual model, that concerns the binaural mechanisms, the neural firing process, the refractary (?) periods. The multichannels model resumed in par. XX have been discussed and extended in more complex models of perception. The normalization of SACF has been discussed by Bernstein [7]; it is also mentioned by Viemeister [68], and in the Ando's ACF formulation [3]. Pressnitzer[48] and Bernstein [7] proposed the usage of weighting functions for different channels. An exponential weighted ACF, based on Meddis' model [37], has been proposed by Bernstein [7] and Wiegrebe [71]; some notes on exponential weighting of ACF can also been found in the original Licklider's formulation [32]. Some authors [35, 22] proposed a non linear filter bank (instead to linear gammatone filters [45]) for modeling the mechanical compression in the middle ear.

In this chapter an improvement of the ACF-based model is presented,

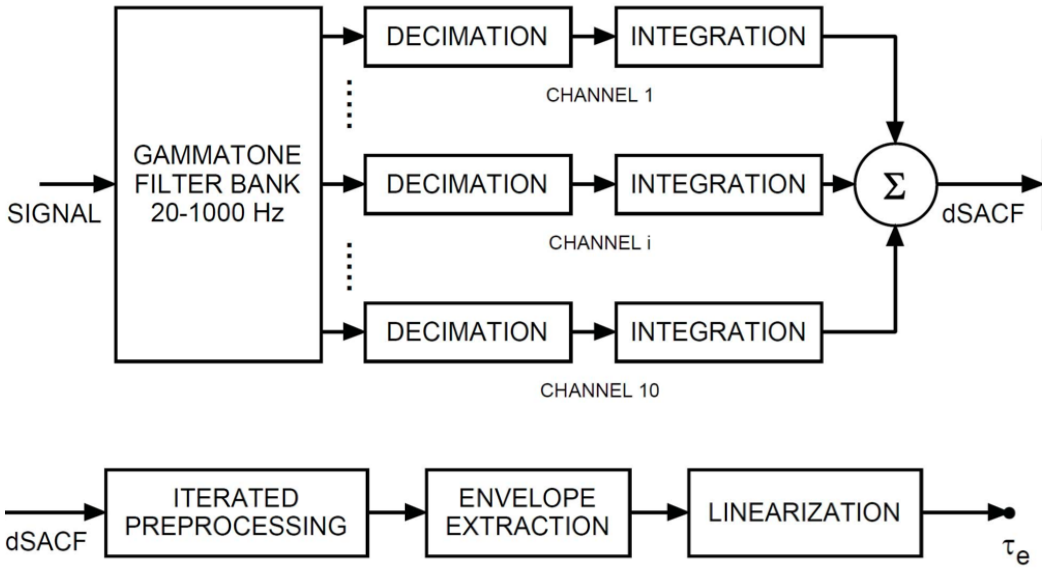


Figure 6.1: General scheme of the new method based on d-SACF.

which can maximize the performance of the new method of τ_e extraction proposed in the last chapter. 5. The use of a decimated summary autocorrelation function (dSACF) instead of an A-weigthed ACF is proposed. [**?**].

6.1.1 Gammatone filterbank

In the previous methods the blocks related to the basilar membrane and cochlear nuclei are represented by an A-filter applied to the signal. In the dSACF a bandpass filter block based on gammatone filters [44] is applied to the windowed signal. A filter bank of ten gammatone filters have been used, whose central frequency range from 20 Hz to 1 kHz. A gammatone function [43] is an equation that may used to give an analytic expression for the reverse correlation procedure [**?**]. A Gammatone filter is the Discrete Fourier Transform of the impulse response:

$$h_n(t) \approx t^{n-1} e^{-2\pi\, b\, t} \cos\left(2\pi\, f_0\, t + \psi\right) \tag{6.1}$$

where n is the order, b is a bandwidth parameter, f_0 is the filter centre frequency and ψ is the phase of the fine structure of the impulse response. The frequency of the higher gammatone used ($1KHz$) may be viewed as a prefiltering to envelope extraction. According to many algorithms that concern the envelope extraction [13, 49], the frequency components above $1kHz$ may be not considered.

6.1.2 Decimation and channel-based autocorrelation

The signal of every channel is decimated. Basing on a cognitive hypothesis of the hair cells refractory period [**?**, 58] one sample per millisecond is retained. A threshold of 'activation' of the channel is represented by the noise floor. This decimation decreases the computational load preserving the significativity of τ_e (as shown in section 5.2). The decimated signal is the autocorrelated and the autocorrelation function of each channel are summated in a *decimated Summary Autocorrelation Function* ($dSACF$). The model so realized is very similar to which proposed by Licklider in his triplex theory (fig. 1.7): same multichannel characteristics, same considerations of refractory period of mechanism of impulse generation. The Licklider's model of triplex theory suggest relations between the firing of one neuron and the firing of his neighbour neurons in similar mode to the [**?**]. The method proposed simplify this process and uses a summary of autocorrelation process. From the first formulation of ACF (duplextheory) the mechanism of autocorrelation is related to a place/channel. The gammatone filterbank is a common way to realize a multichannel model of basilar membrane. Other methods are proposed in literature, as the Gammachirp of Irino and Patterson [22], that weights the gammatone filter with exponential function, to compute the different weight of the different bands. The $d-SACF$ so calculated is then processed by an energy detector to extract the envelope.

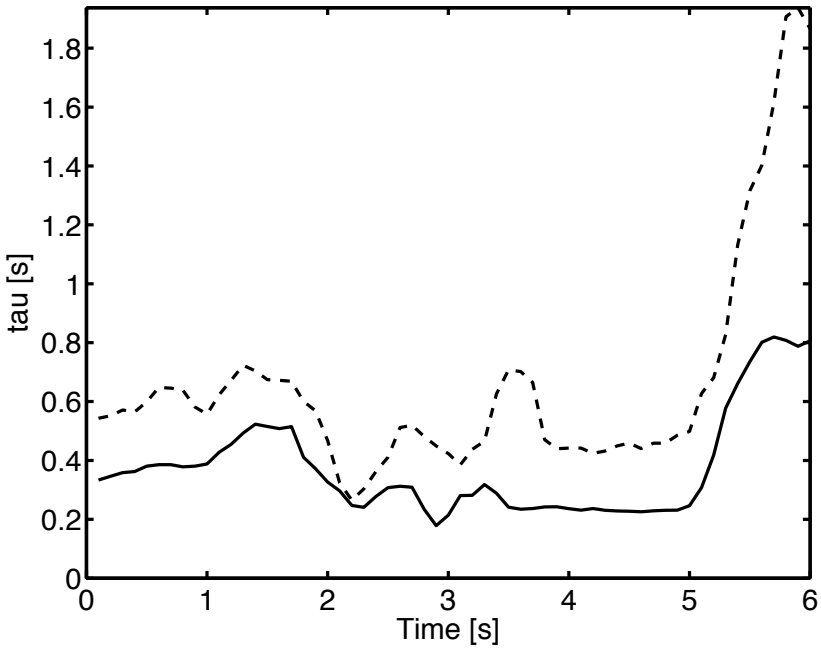


Figure 6.2: $\tau_e(t)$ curves: comparison between new proposed method based on A-weighted ACF (black line) and new proposed method based on d-SACF. Motif 1. $2T = 2s$

6.2 Experimental results

A redefinition of the proportionality coefficients between $(\tau_e)_{min}$ could be necessary after the calculation of new running τ_e in the motifs used as bases in the subjective preference theory. This may autorize also a possible use of the dSACF based new proposed method. In figure 6.5 the running τ_e of motif 2, calculated with the new method using the A-weighted ACF and the dSACF, is presented. Although the two curves return different values of τ_e, they have the same trend.

The use of the dSACF version of the new method permits the calculation of the of a signal in case of low SNR. The envelope extraction based on dSACF seems to be less dependent from the SNR. An analysis of running of motif 2 with a $SNR = 3$ and with SNR is presented in figure . The running τ_e analysis based on the A-weighted ACF in case of $SNR = 3$ (solid bold curve) is meaningless, because the minimum value of τ_e is lower than the minimum evaluable using this method within the given frequency window (about 10 ms). The running τ_e analysis based on dSACF returns two comparable results: the dashed curve (SNR $\approx \infty$) and the dashdot curve ($SNR = 3$) have the same trend and return a similar value of $(\tau_e)_{min}$ (fig. 6.5).

6.3 A.G.A.T.A. (A Gui for Ando's Temporal Analysis)

A graphic user interface has been developed to collecting all the methods proposed in this thesis. The functions that concerns the temporal analysis are coded with a common semantic in a Toolbox structure, based on a similar Toolbox developed for binaural analysis [**?**]. The visual aspect of a beta version of A.G.A.T.A. is shown in fig. 6.7.

SCRIVERE QUALCOSA SULLE FUTURE APPLICAZIONI DI AGATA

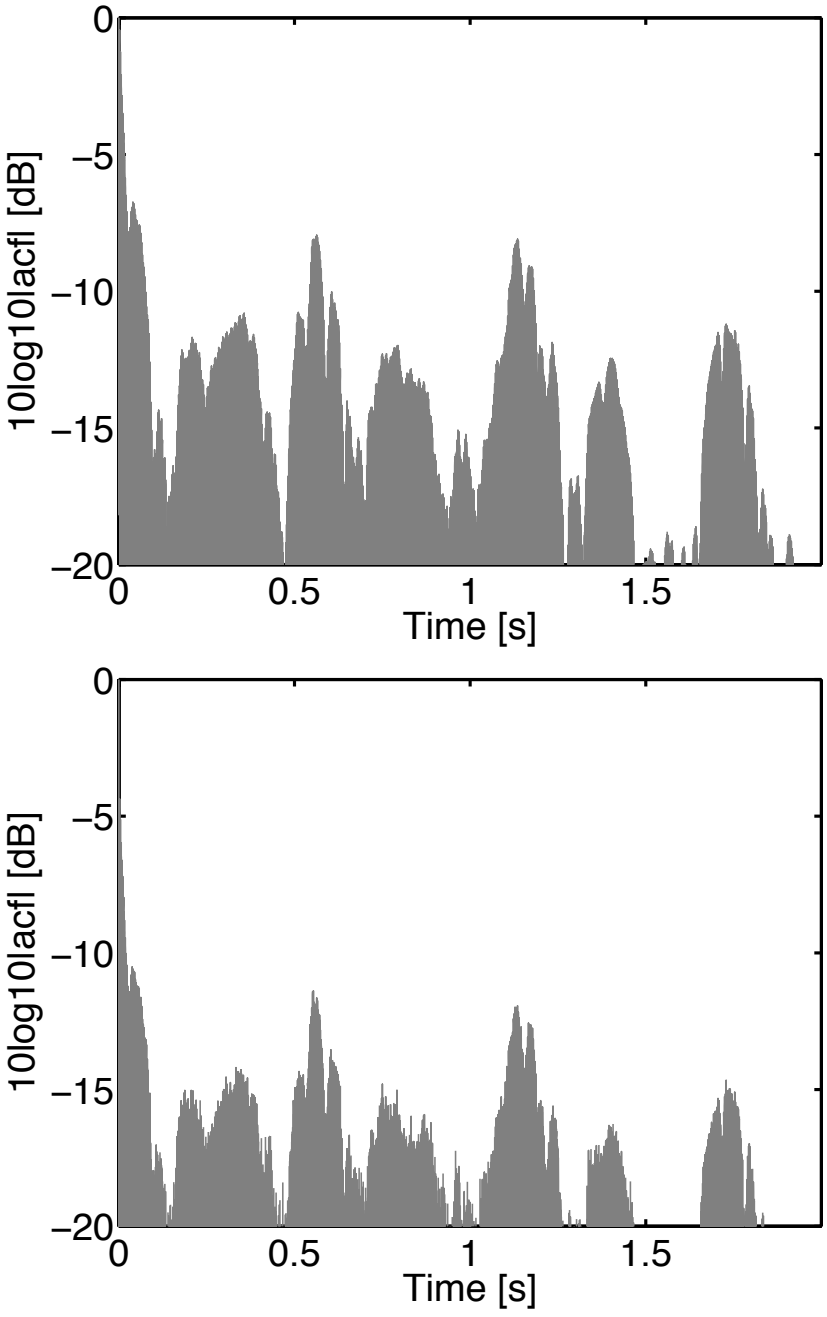


Figure 6.3: ACF with $SNR = \infty$ (A) and $SNR = 3$ (B)

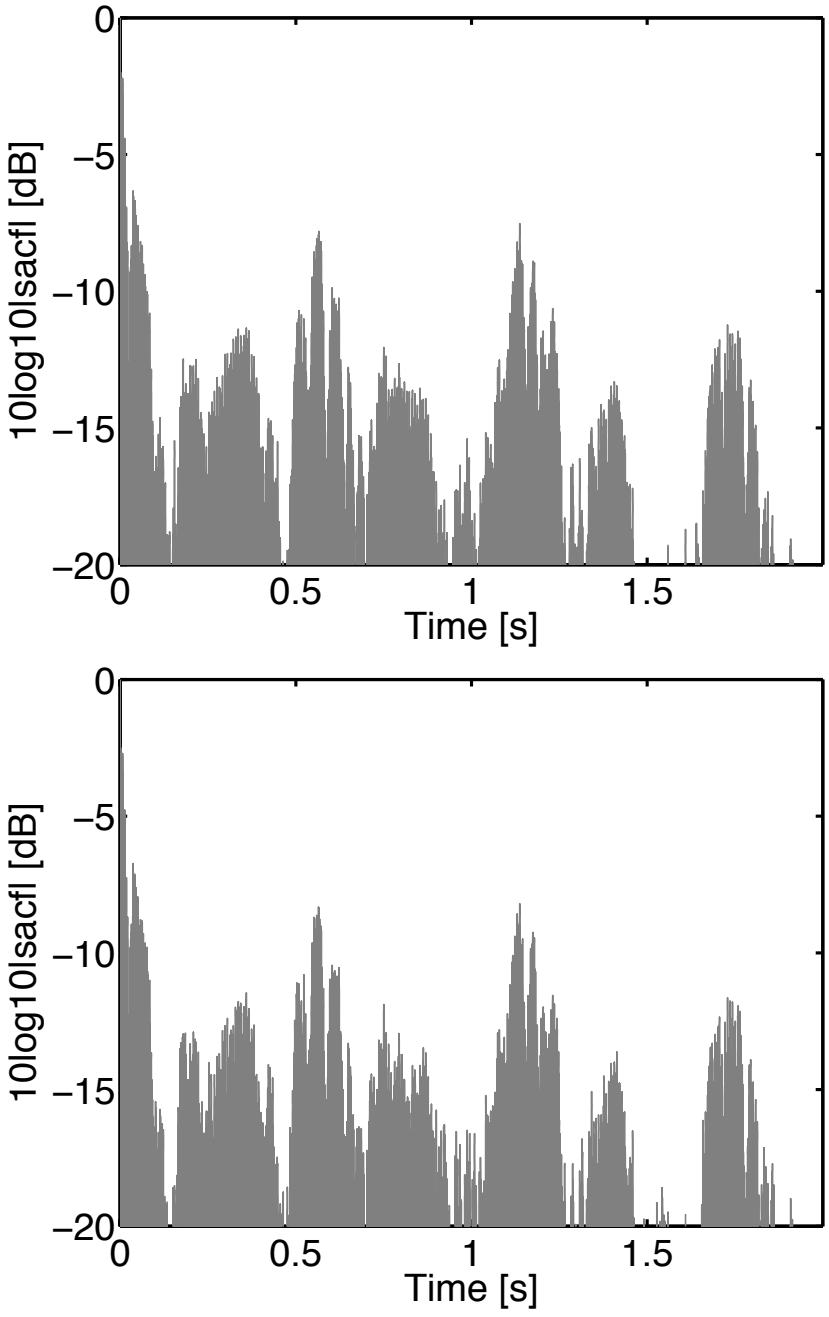


Figure 6.4: d-SACF with $SNR = \infty$ (A) and $SNR = 3$ (B)

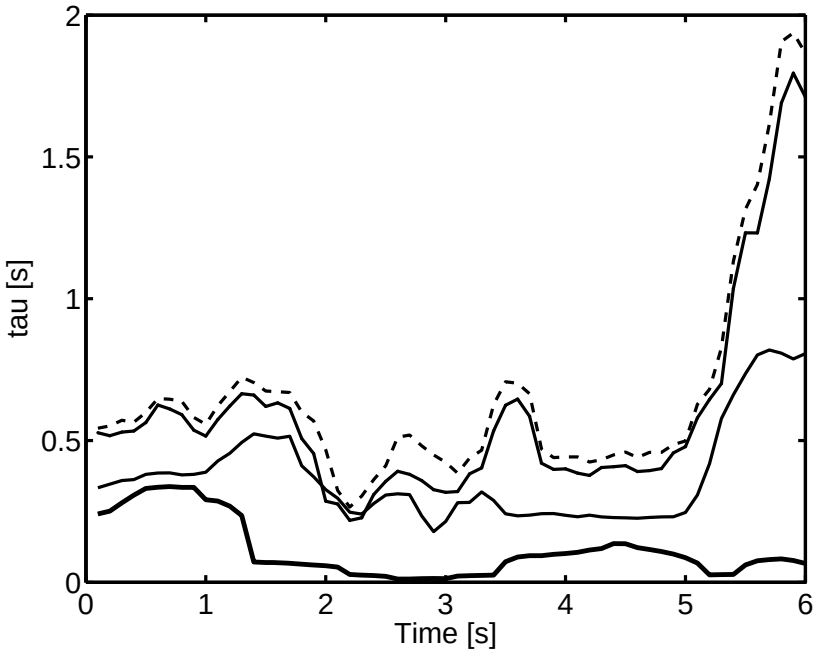


Figure 6.5: Comparison between the new proposed method ACF-based and dSACF-based. Motif 2 with SNR=3 and SNR $\approx \infty$. Running τ_e of new method based on A-weighted ACF (solid line (SNR $\approx \infty$) and solid bold line (SNR=3)) and new method based on dSACF (dotted line (SNR $\approx \infty$) and dashdot line (SNR=3)).

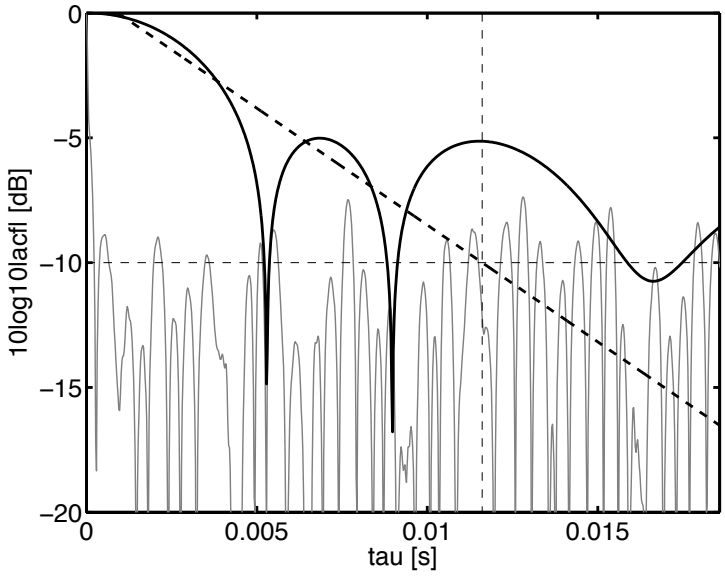


Figure 6.6: Minimum of τ_e extractd from an A-weighted based ACF in case of high SNR

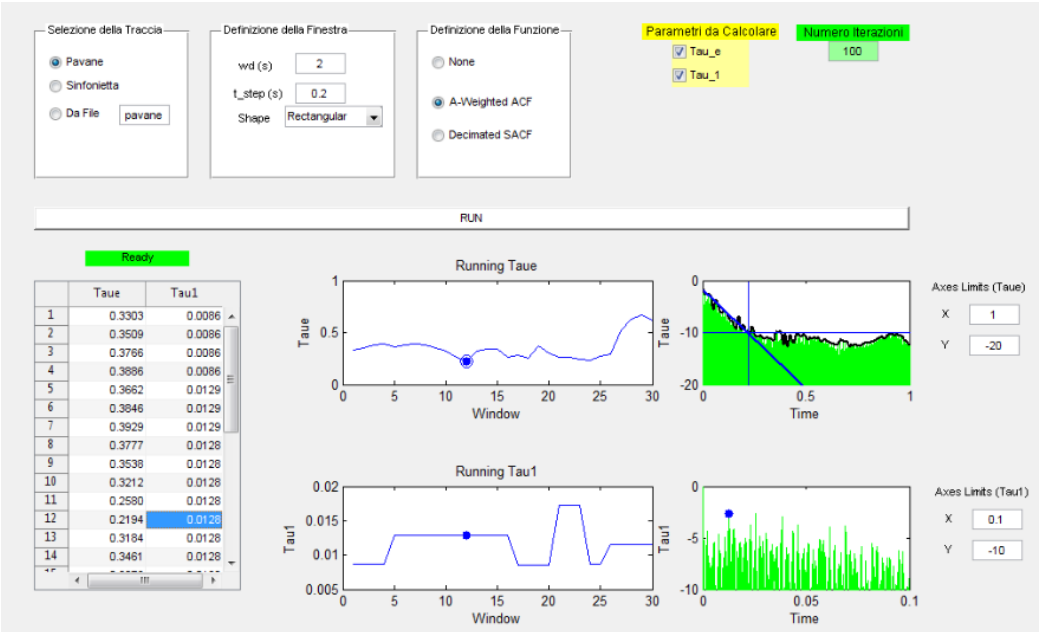


Figure 6.7: Screen shot of A.G.A.T.A. (beta version)

6.4 Discussion

vale quanto detto nel paragrafo precedente, solo che qui la banda di ciascun filtro inferiore e la perfonance pi smoothed.

confronto gammatone/schroeder. non c' molta differenza in fondo. ando conosceva schroeder ma ha usato filtro A. in fondo non c' troppa differenza. la differenza in questo metodo la fa invece la decimazione.

The new proposed method may solve some of the limitations of the methods from literature: this may permit some possible extensions of τ_e-based analysis.

minimum phase gammatone

Chapter 7

A little auditorium

7.1 The *Torri dell'Acqua* Auditorium in Budrio: project criteria

7.2 Measurement of normal factors

The values of Δt_1 are extracted from the impulse response with an algorithm based on iterated sharpening. The absolute value of the impulse response is preprocessed with an iterated sharpening. The sharpening permits to identify the local maxima of the impulse response. The local maxima becomes a nodes of a cubic spline. Another cubic spline, based on the last spline, permits to identify the reflections in which is present the maximum energy contribution.

The Δt_1 is assumed as the delay time (respect to the arrival instant of the direct wave) of the reflection with a maximum energy contribution. The use of the spline permits the non identification of the 'spurious' peaks near the direct wave and near the main reflections. This approach is quite different to that which consider the reflection characterized by the maximum of the amplitude. Fuji et Al [**?**] proposes a similar interpretation of the Δt_1, based on different extraction process. A running integration of the squared impulse response with a temporal window of 5ms is proposed: the Δt_1 is the delay of the maximum of the decay function so calculated.

The identification process of the measured IR in the auditorium are shown

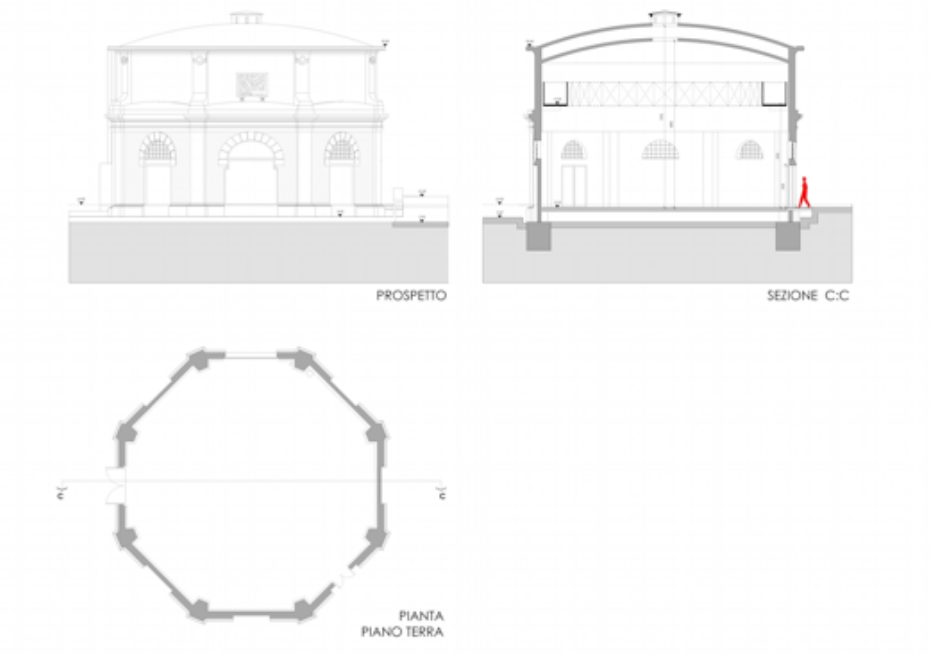


Figure 7.1: Prospect, plant and section of the *Torri dell'Acqua* Auditorium

in figg. 7.7, 7.8, 7.9, 7.10.

7.3 Evaluation of subjective preference

7.4 Active control of the running τ_e

When the music (or the speech) is played through an electroacoustic system more considerations on the τ_e value may be proposed. In case of acoustic instrument, as proposed in (Tsub chapter) the sound field at the entrance of the listener's ear is supposed as a convolution of the monaural source and the binaural impulse response of the hall. In case of electroacoustic system the eq. [**?**] may be written as

$$f_{dsp,l,r}(t) = p_{dsp}(t) * h_{l,r}(t;r) \tag{7.1}$$

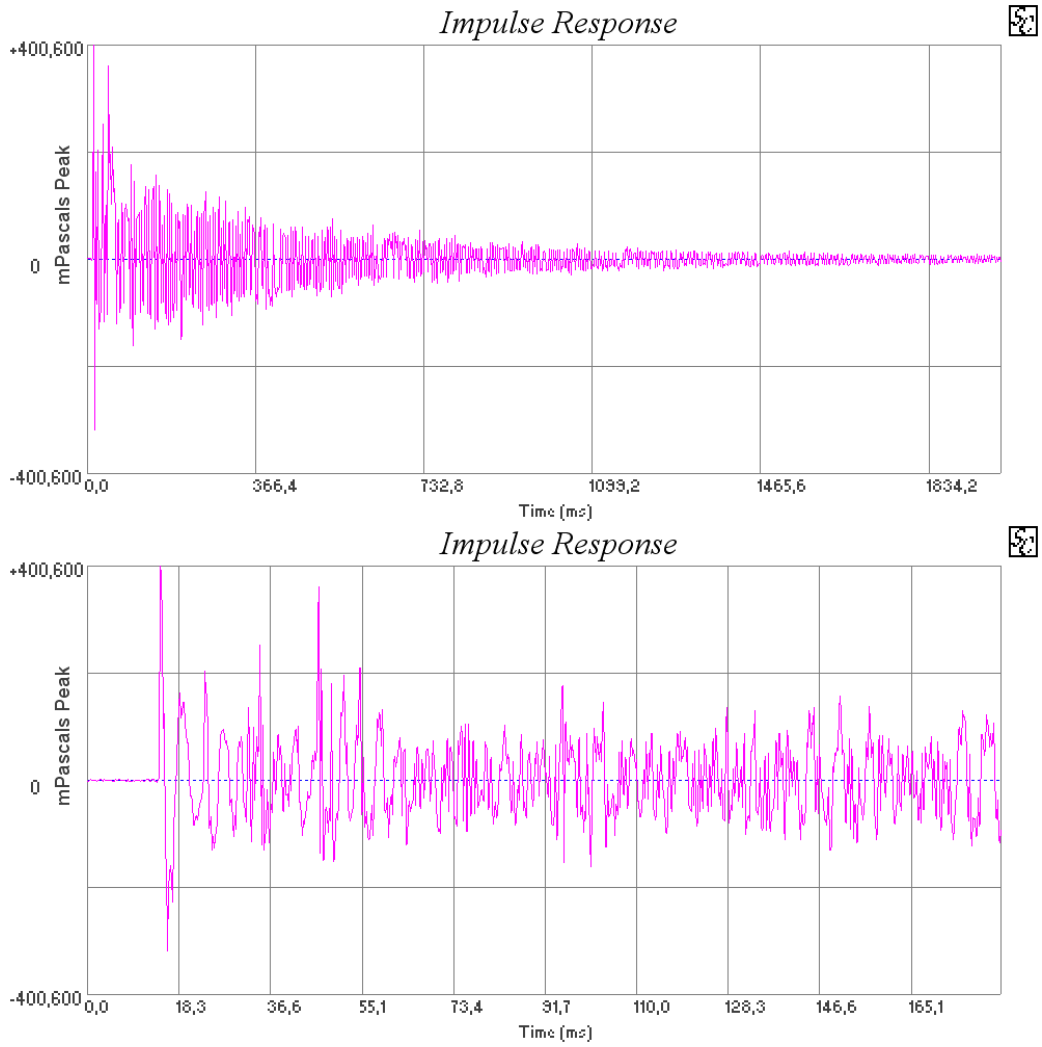


Figure 7.2: Measurements *ante operam*. (A) Decay of the impulse response. (B) Detail of the Early Reflections.

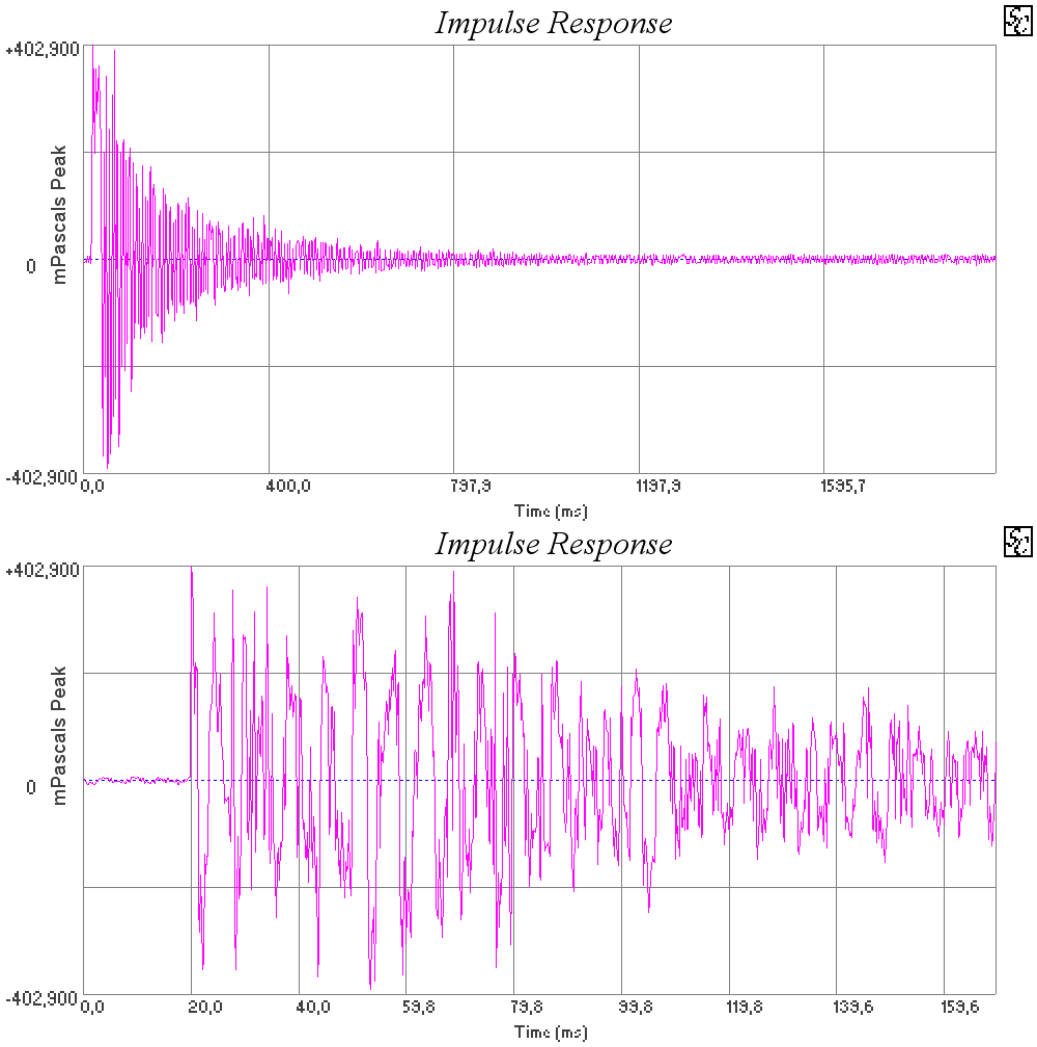


Figure 7.3: Measurements *post operam*. (A) Decay of the impulse response. (B) Detail of the Early Reflections.

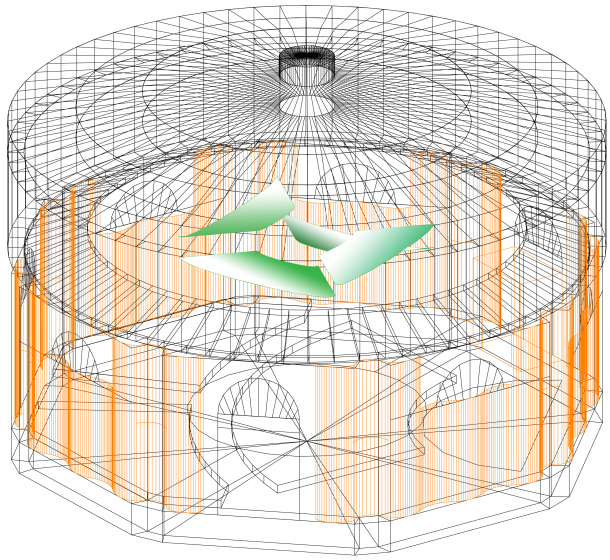


Figure 7.4: CAD of *Odeon*© simulation. Orange: lateral diffusion surfaces to controlling the IACC. Green: asymmetrical reflectors to control the Δt_1, the modal components in low frequency and the IACC

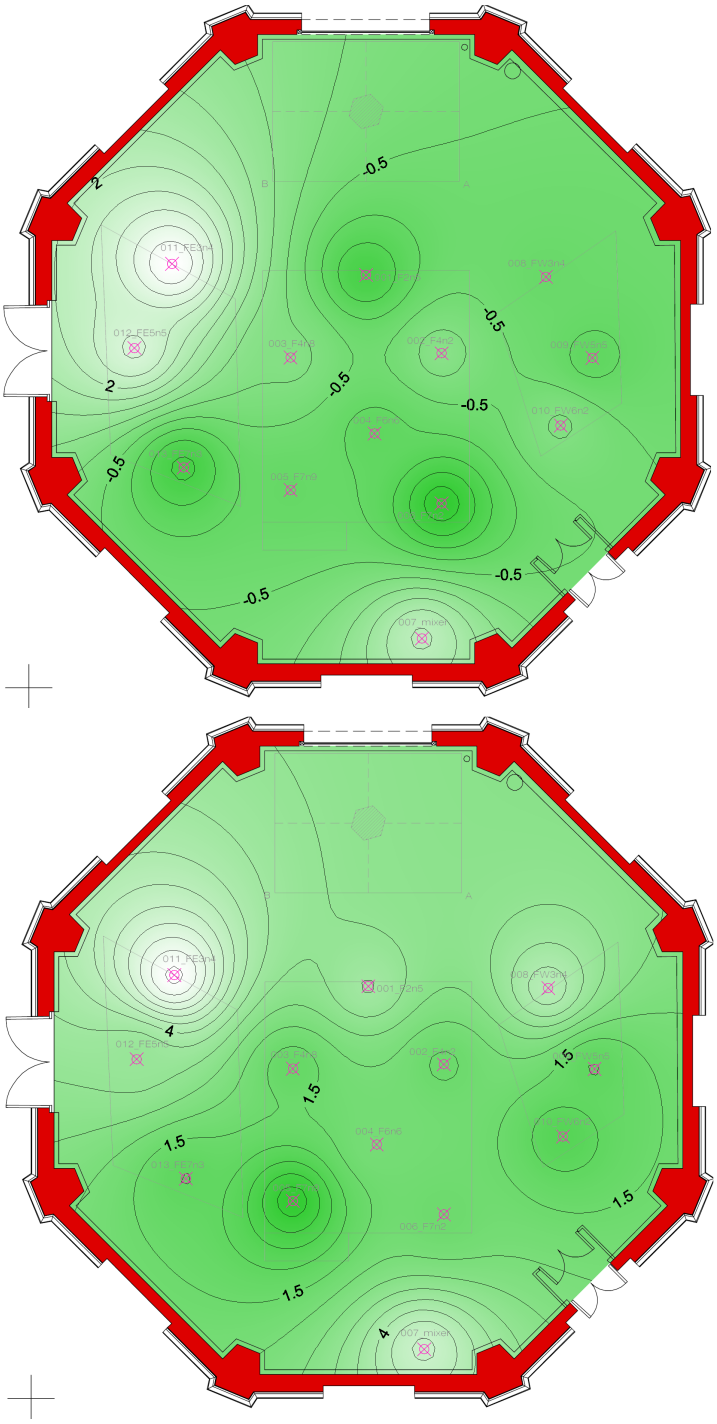


Figure 7.5: Measured values of C_{80}. Map @125 Hz (A), @500 Hz (B)

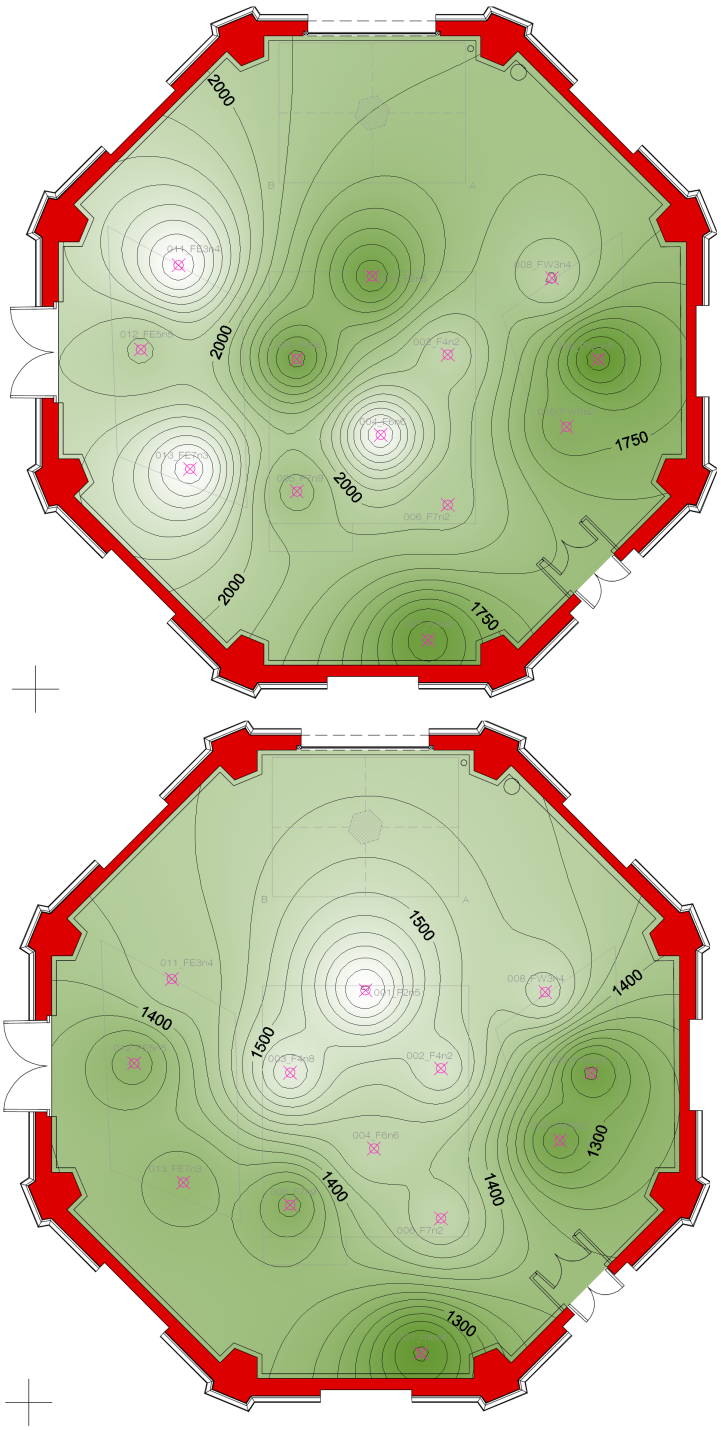


Figure 7.6: Measured values of EDT. Map @125 Hz (A), @500 Hz (B)

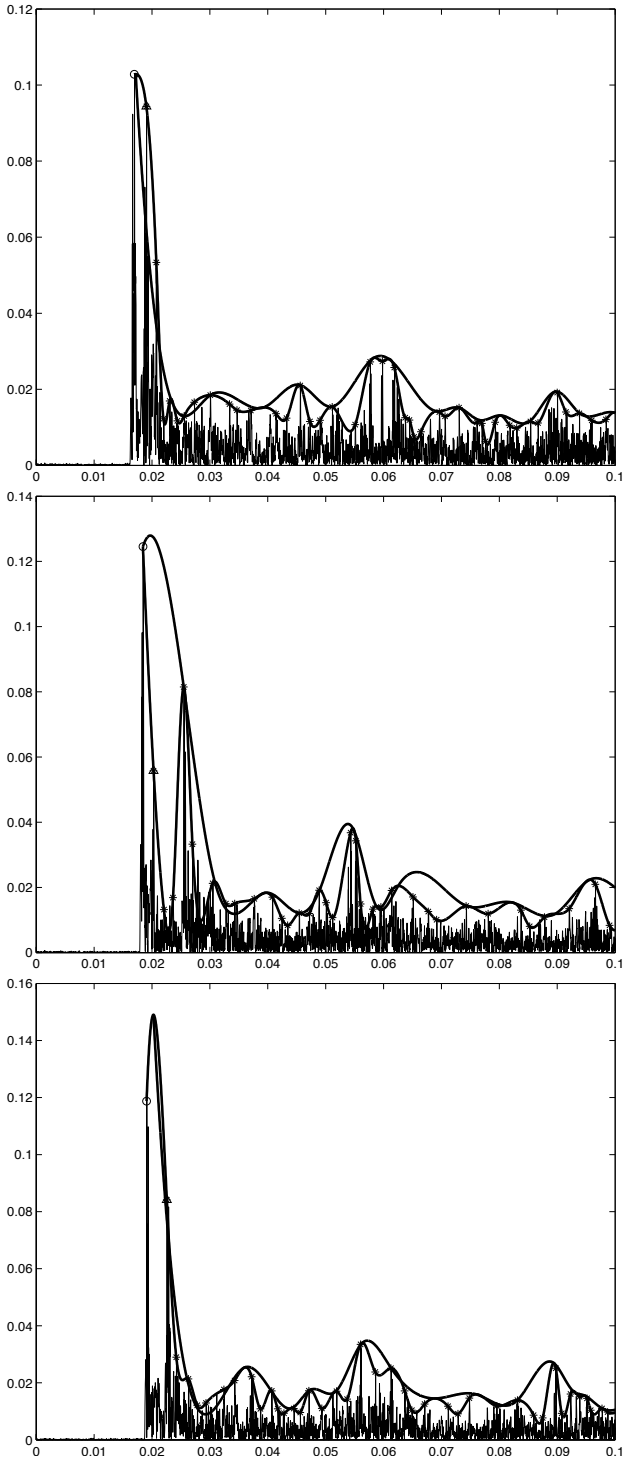


Figure 7.7: Measured values of Δt_1 in the *Torri dell'Acqua* auditorium. Positions 01, 02, 03

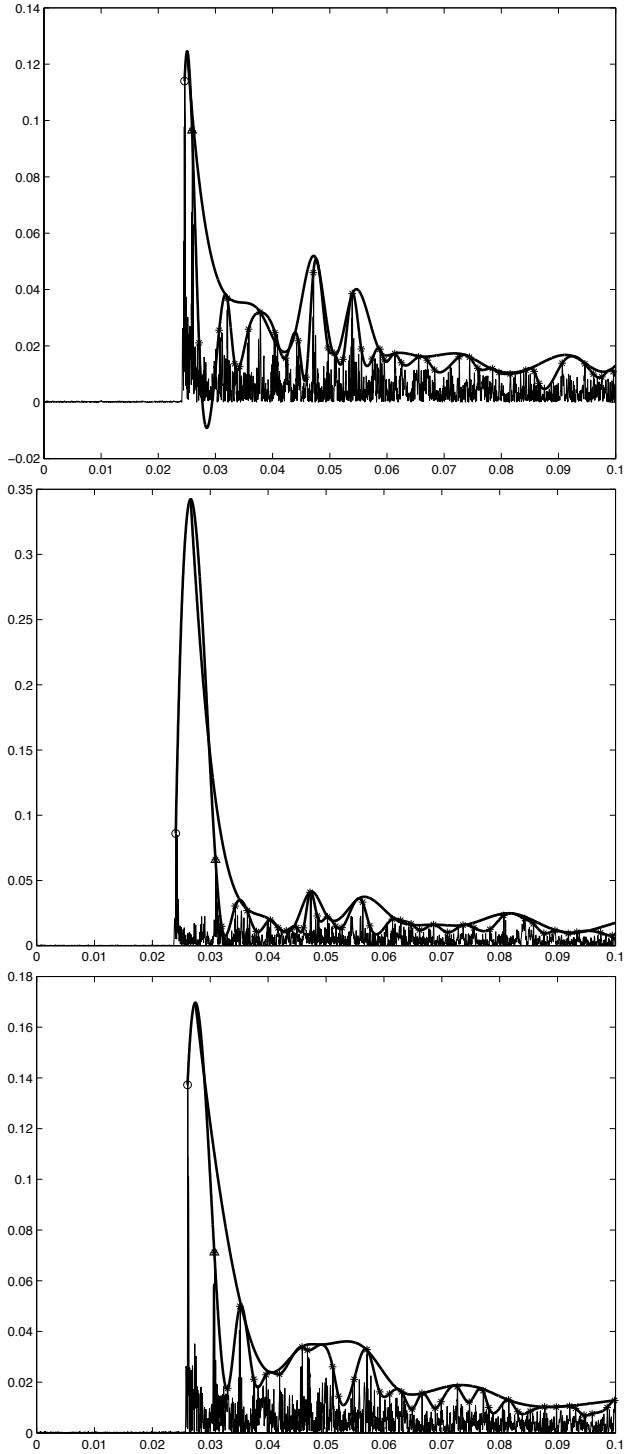


Figure 7.8: Measured values of Δt_1 in the *Torri dell'Acqua* auditorium. Positions 04, 05, 06

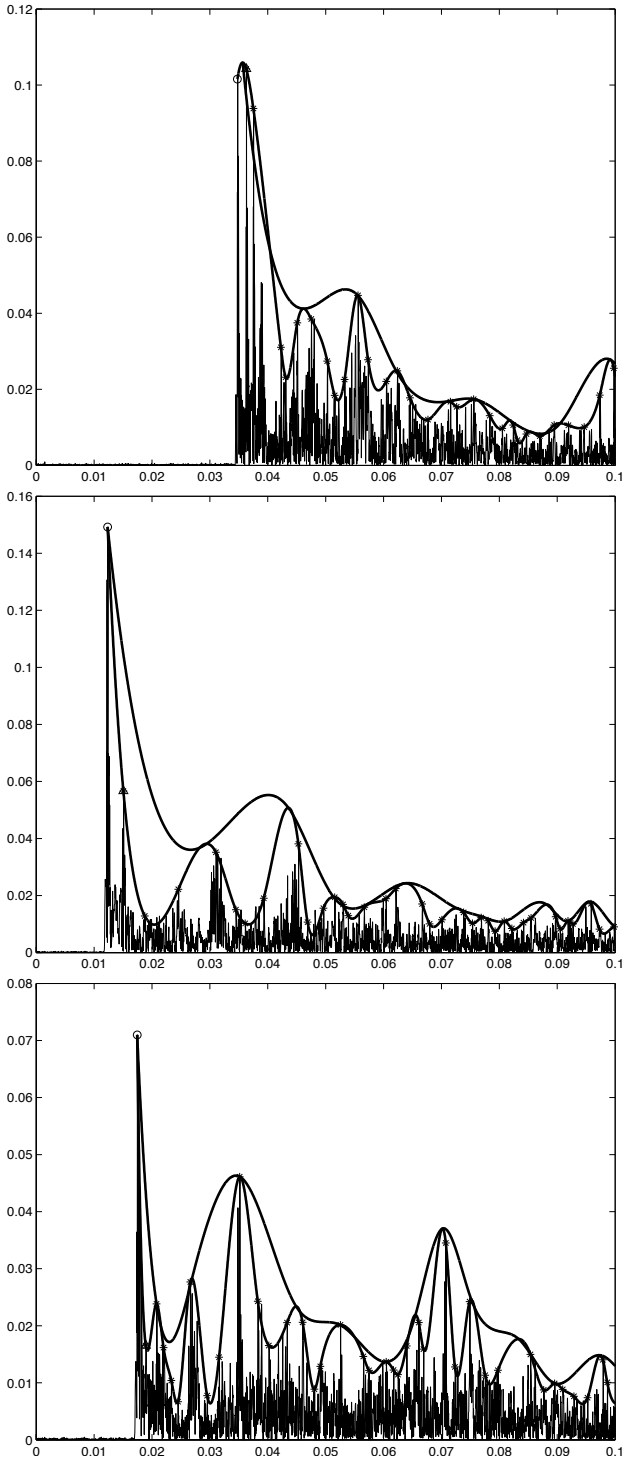


Figure 7.9: Measured values of Δt_1 in the *Torri dell'Acqua* auditorium. Positions 07, 08, 09

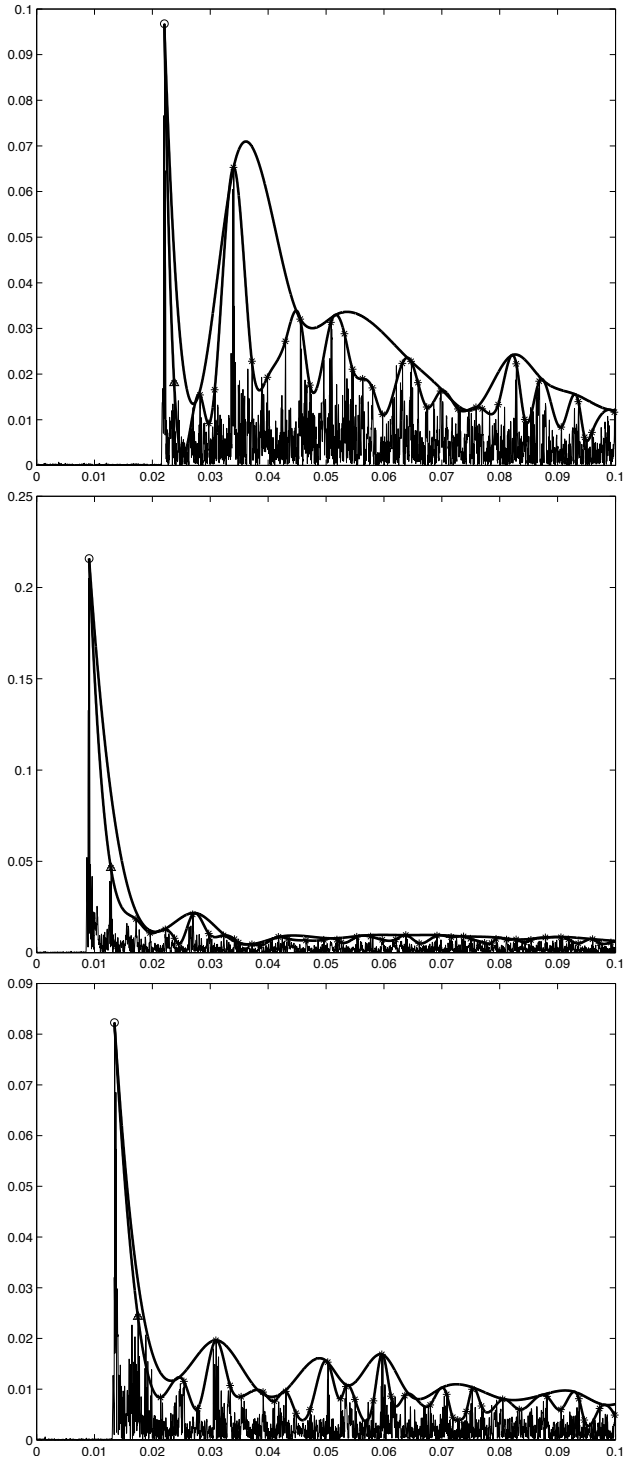


Figure 7.10: Measured values of Δt_1 in the *Torri dell'Acqua* auditorium. Positions 10, 11, 12

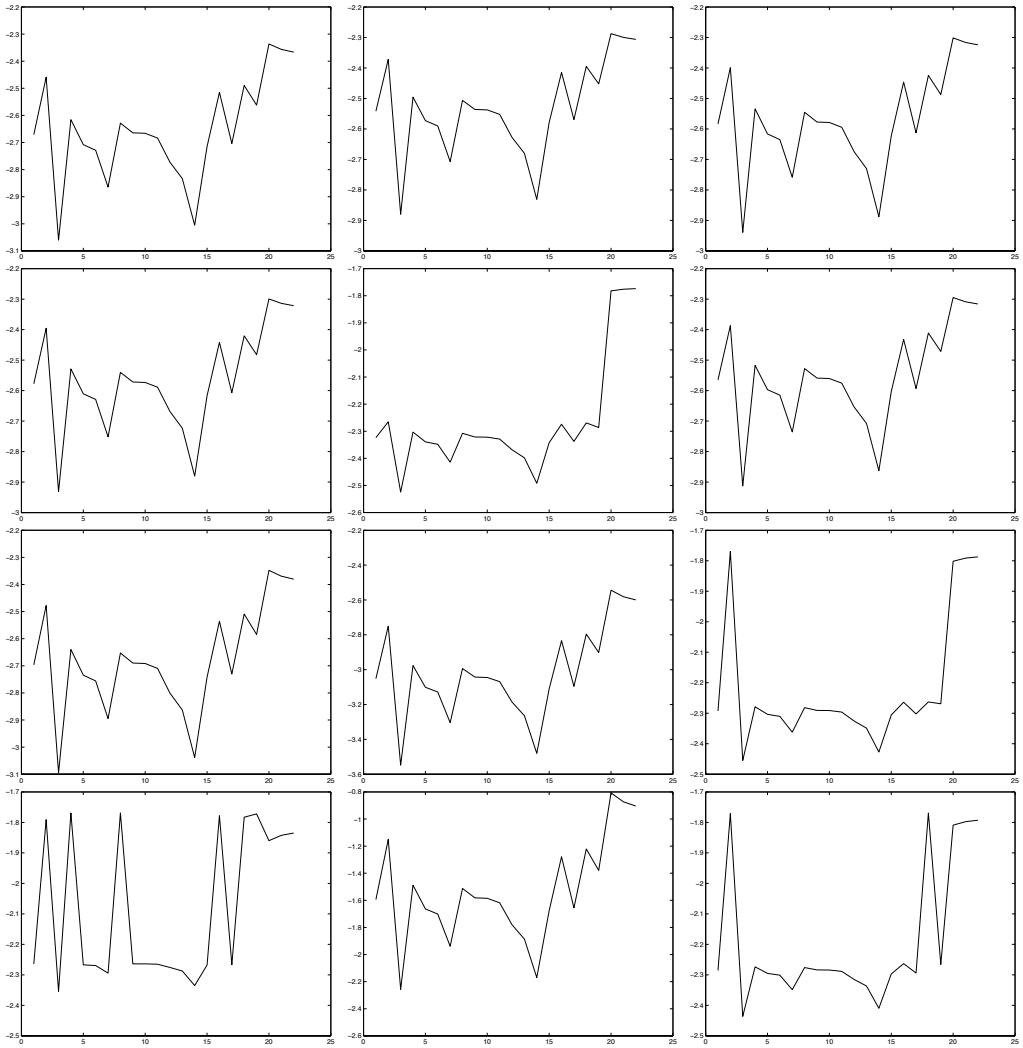


Figure 7.11: Measured values of running-$S2$. Mothif: 1 (sinfonietta). Positions 1-12

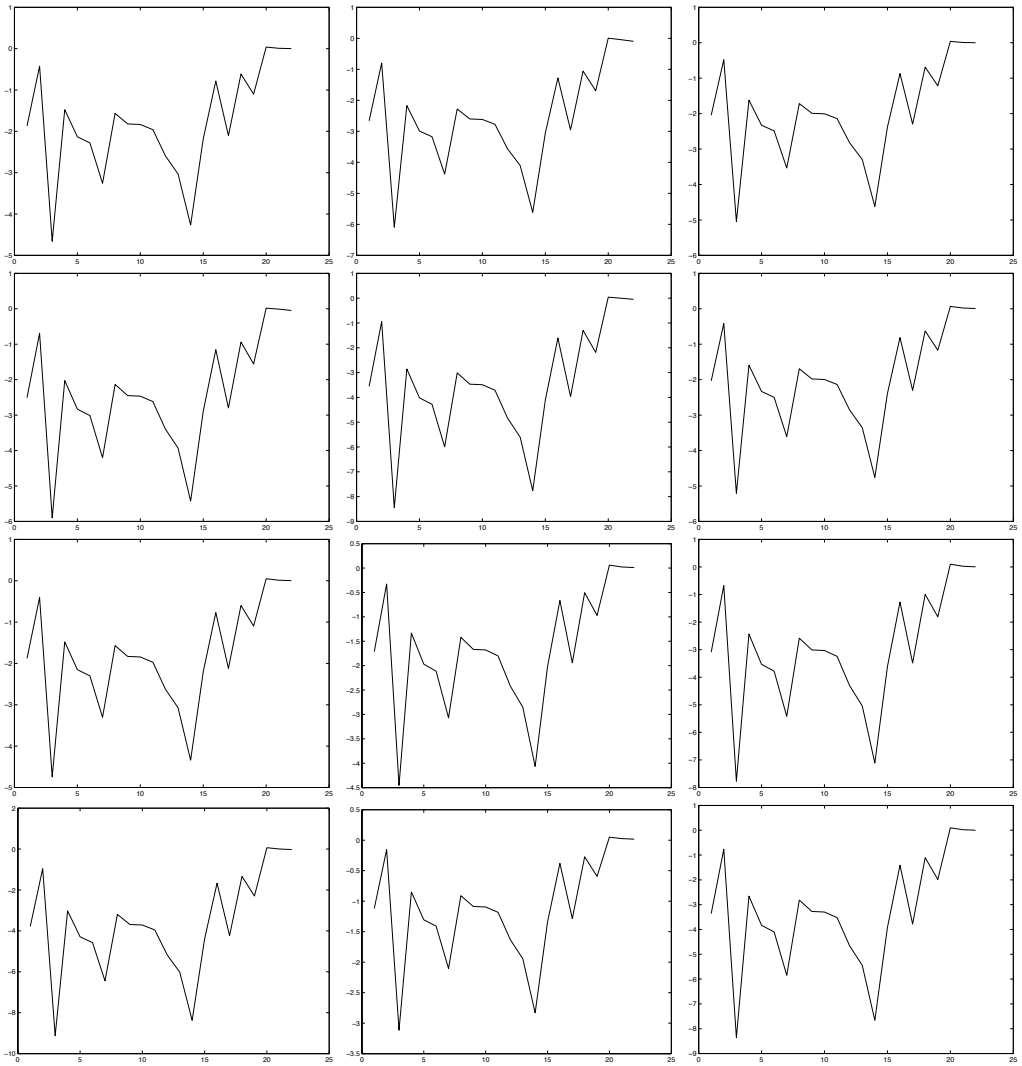


Figure 7.12: Measured values of running-$S3$. Mothif: 1 (sinfonietta). Positions 1-12

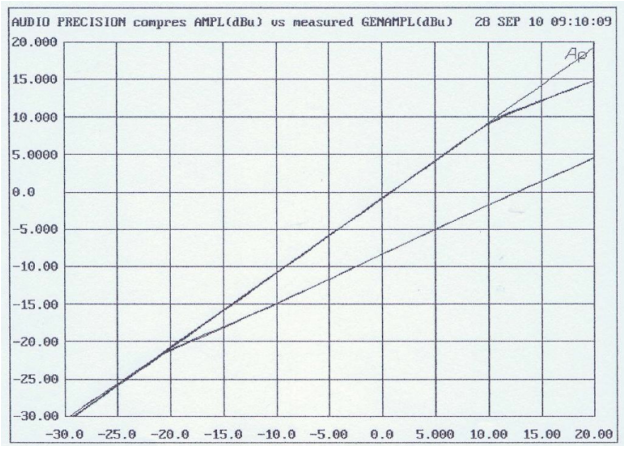


Figure 7.13: Definition of dynamical compression parameters

where $h_{l,r}(t;r)$ is the binaural impulse measured at listener's place (r), $p_{dsp}(t)$ is the signal processed in nonlinear mode. For simplicity the electroacoustic source (the loudspeaker) is supposed monaural. The processing of the signal may concern both the spectrum (equalizers) and the dynamic (limiters/compressors). The last ones are the object of this paragraph. A compressor is a non linear processor that modify(s) the amplitude of the signal, without effects to the spectrum of the signal. Above a selected threshold level (fig audio precision) the compressor multiply the relative level of the signal for a factor $\frac{1}{ratio}$, the relative level being the positive difference between the instant level and the threshold level. The signals that have a level minor of threshold have no modifications in dynamic. The ratio is general positive (compressors) but in some cases may have a negative value (distressors). In general use, the compression occurs some milliseconds after the instant when the signal exceed the threshold. This interval is called attack time and may range from 0 to $300ms$. The attack time 'preserves' the transitory of the signal, when it's needed, e.g. in case of voices or string instrument. Different uses of compressor must have different attack time and the attack time may considered a third parameter of the compression, the first being the threshold and the second being the ratio. The compressor are used in almost all the electroacoustic system. Compression is also used in all the music played by radio, television and in almost all the musical recordings. In first

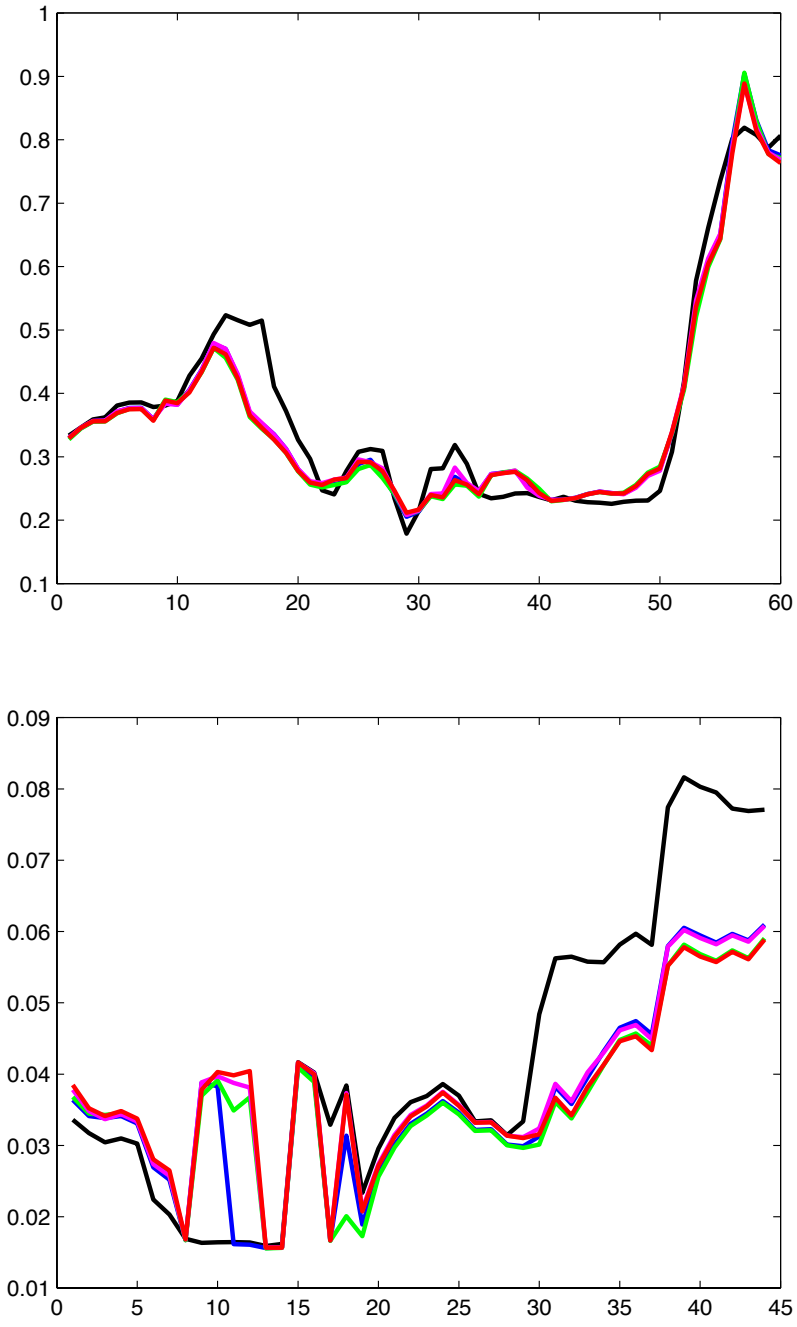


Figure 7.14: $\tau_e(t)$ in case of dynamical compression. Motif: Pavane (a), Sinfonietta (b). $2T = 2s, T_{step} = 100ms$. Curves: black (no compression), blue (Threshold level=-15dB, Ratio=5, Attack time=100ms), green (Threshold level=-15dB, Ratio=10, Attack time=100ms), magenta (Threshold level=-15dB, Ratio=5, Attack time=200ms), red (Threshold level=-15dB, Ratio=10, Attack time=200ms).

approximation it's assumed that, in case of compressor, only the envelope characteristics of the ACF may change, the fine structure being the same. The $\tau_e(t)$ is then evaluated for two motifs compressed with different values of ratio and attack time (fig. 7.14). As shown, the minimum value of $\tau_e(t)$ has no variations in case of dynamical compressions. Also in case of smaller attack (¡50ms) time values the minima values of τ_e not could change. More useful are the curves of compressed signals in case of Sinfonietta (fig. 7.14 (b)). The compression seems to limits also the τ_e values when these values increase over a certain value. The same limitations of the curves is shown in the first part of Pavane(fig. 7.14 (b)). In the second part the compressor not limits the τ_e: this may happen because the release time of the compressor is not considered in this study. This first considerations on nonlinear processing may suggests a possible interpretation of compression as a 'limiter' of τ_e. In other words, if some conditions of τ_e are expected to maximize the subjective preference over a long time exposition, a dynamical compression may maximize this effect. For a long time exposition it's assumed an interval in which the perceived value of $(\tau_e)_{min}$ may change. This feature of the auditory system, which may be supposed as adaptive in perceptive meaning (Ando, personal communication), is not yet included in the Ando's measurement system.

Conclusions

The main reasons to give relevance to the τ_e analysis, documented in a vast literature, has been synthesized in this work. Initially proposed as descriptor of preferred monaural parameters in architectural acoustics, the τ_e is the temporal factor that permits to relate the listener's preferred values of monaural room descriptor to the proposed sound signal. In the last years the use of τ_e has been proposed in other fields of acoustics related to the subjective perception (building acoustics, annoyance evaluation, soundscape analysis, etc). Both the classical domain of concert hall acoustics and the new fields of application need an accurate and robust algorithm for the estimation of τ_e.

More in detail, three decay extraction methods have been proposed to extract the τ_e from the ACF in literature [4]: the line fitting based on a least-square approximation for (some) initial peaks, the evaluation of the ACF decay function calculated with a Schroeder-like backward integration and the extraction of the envelope using the Hilbert's transform. The mathematical and technical considerations stemming from them may be resumed as follows.

The least-square fitting, also know as peak detection method, is the most used method fot the extraction of τ_e in current literature and has also a good neurophysiologic plausibility. As noted in par. XX, the criterion of selection of the peaks to be fitted is not well described in the literature. Also, some ambiguities may occur in case of impulsive sounds or signals with low SNRs. A reference version of this method, with an optimized criterion of peak selection, has been developed in the frame of the present work to have a reference against which to compare all other methods.

The evaluation of the ACF decay based on the Schroeder's integral re-

veals a dependence from the correct choice of the upper limit of integration (ULI). The decay (and also the τ_e value) are well characterized only in case of a correct choice of the ULI. Several methods have been developed to perform a correct choice of ULI for room impulse responses, but these methods cannot be used to select a correct ULI as far as the decay of the autocorrelation function is concerned. Statistical considerations suggest that the room impulse response may be associated to a Rayleigh's distribution, while the autocorrelation function may be associated to a Rice's distribution (par. XX). Moreover, if a limit of -10 dB is fixed to quantify the noise in the normalized ACF, other technical considerations confirm the inapplicability of these methods (par. XX).

The Hilbert's transform is a well know method to identify the envelope of a signal, provided some hypotheses are assumed. If the signal is a minimum phase one these hypotheses are satisfied. In case of a mixed phase signal these hypotheses are not verified and therefore the Hilbert's transform doesnt return the correct envelope. In par. XX a discussion of the minimum phase characteristics of the autocorrelation function is proposed. In general case this characteristic is not verified.

Let's note that, in case of an impulsive sound signal, the statistics of the ACF converges to the statistic of the gaussian noise (as the room impulse response) and both the decay evaluation method and the Hilbert's transform identify the correct envelope and return the correct τ_e value, as noted in some cases.

A new method has been proposed, based on considerations from neurophysiology and signal theory. Following some observations related to the SKS theory (par. XX) an improved version of the Energy Detector is proposed to identify the τ_e value. The new method is based on an iterative non linear preprocessing (sharpening) follow by an optimized post detection filter. In the chap. XX the new method is described in detail and some comparison with the reference peak detection method are proposed.

Bibliography

[1] Y. Ando, "Subjective preference in relation to objective parameters of music sound fields with a single echo", J. Acoust. Soc. Am. **62**, 1436–1441 (1977).

[2] Y. Ando, "On the preferred reverberation time in auditoriums", Acustica **50**, 134–141 (1982).

[3] Y. Ando, *Concert Hall Acoustics* (Springer-Verlag, New York) (1985).

[4] Y. Ando, T. Okano, and Y. Takezoe, "The running autocorrelation function of different music signals relating to preferred temporal parameters of sound fields", J. Acoust. Soc. Am. **86**, 644–649 (1989).

[5] Y. Ando, "A theory of primary sensations and spatial sensations measuring environmental noise", J. Sound and Vib. **241**, 3–18 (2001).

[6] Y. Ando and R. Pompoli, "Concert hall acoustics and opera house acoustics", Special issue of J. Sound and Vib. **258** (2002).

[7] J. Bernstein and A. Oxenham, "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination", J. Acoust. Soc. Am. **117**, 3816–3831 (2005).

[8] J. Brown and M. Puckette, "Calculation of a 'narrowed' autocorrelation function", J. Acoust. Soc. Am. **85**, 1595–1601 (1989).

[9] A. Burd, "Nachfaller musik fur akustische modelluntersuchungen", Rundfunktechn. Mitteilungen **13**, 200–201 (1969).

[10] P. Cariani and B. Delgutte, "Neural correlates of pitch of complex tone. i. pitch and pitch salience", J. Neurophysiol. **76**, 1698–1716 (1996).

[11] W.T. Chu, "Comparison of Reverberation Measurements Using Schroeders Impulse Method and Decay-Curve Averaging Method", J. Acoust. Soc. Am. **63**, 1444–1450 (1978).

[12] A. de Cheveignè, "Cancellation model of pitch perception", J. Acoust. Soc. Am. **103**, 1261–1271 (1998).

[13] T. Dau, D. Puschel, and A. Kohlrausch, "A quantitative model of the effective signal processing in the auditory system: I. model structure", J. Acoust. Soc. Am. **99**, 3615–3622 (1996).

[14] E. deBoer, "No sharpening? A challenge for cochlear mechanics", J. Acoust. Soc. Am. **73**, 567–573 (1983).

[15] V. Fourdouiev, "Evaluation objective de l'acoustique des salles", Proceedings of 5th International Congress on Acoustics, 41–54 (1965).

[16] P. Fraisse, *The Psychology of Music*, chapter Rhythm and tempo, 149–180 New York: Academic Press (1982).

[17] K. Fujii, Y. Soeta, and Y. Ando, "Acoustical properties of aircraft noise measured by temporal and spatial factors", J. Sound and Vib. **241**, 69–78 (2001).

[18] M.Garai and P.Guidorzi, "European methodology for testing the airborne sound insulation characteristics of noise barriers in situ: experimental verification and comparison with laboratory data", J. Acoust. Soc. Am. **108**, 1054–1067 (2000).

[19] J. Goldstein, "Auditory spectral filtering and monoaural phase perception", J. Acoust. Soc. Am. **41**, 458–479 (1966).

[20] F. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform", Proceedings of the IEEE **66**, 51–83 (1978).

[21] M. Inoue, Y. Ando, and T. Taguti, "The frequency range applicable to pitch identification based upon the auto-correlation function model", J. Sound and Vib. **241**, 105–116 (2001).

[22] T. Irino and R. D. Patterson, "A dynamic compressive gammachirp auditory filterbank", IEEE Trans. Audio, Speech, and Language Process **14**, 2222–2232 (2006).

[23] L. Jeffress, "Stimulus-oriented approach to detection", J. Acoust. Soc. Am. **36**, 766–774 (1964).

[24] L. Jeffress, "Mathematical and electrical models of auditory detection", J. Acoust. Soc. Am. **44**, 187–203 (1968).

[25] J. Y. Jeon, "Subjective evaluation of floor impact noise based on the model of acf / iacf", J. Sound and Vib. **241**, 147–155 (2001).

[26] P.I.M. Johannesma, "The pre-response stimulus ensemble of neurons in the cochlear nucleus", Proceedings of the Symposium on Hearing Theory, IPO, Eindhoven, 58–69 (1972).

[27] M. Karjalainen, P. Antsalo, A. Makivirta, T. Peltonen, and V. Valimaki, "Estimation of modal decay parameters from noisy response measurements", J. Audio Eng. Soc, **50**, 867–879 (2002).

[28] K. Kato, T. Hirawa, K. Kawai, T. Yano, and Y. Ando, "Investigation of the relation between $(\tau_e)_{min}$ and operatic singing with different vibrato styles", J. Temporal Des. Arch. Environ. **6**, 35–48 (2006).

[29] M. Kazama and M. Tohyama, "Estimation of speech components by acf analysis in a noisy environment", J. Sound and Vib. **241**, 41–52 (2001).

[30] R. Kurer and U. Kurze, "lntegrationsverfahren zur nachhallauswertung", Acustica **19**, 314–322 (1967).

[31] J. Licklider, "An electrical investigation of frequency localization in the auditory cortex of the cat", Ph.D. thesis, Univer. of Rochester (1942).

[32] J. Licklider, "A duplex theory of pitch perception", Experientia **7**, 128–134 (1951).

[33] J. Licklider, "Three auditory theories", 3^{rd} London Symposium on Information Theory, edited by C.Cherry, Butterworth's London, 41–144 (1956).

[34] Lifschitz, Phys. Rev. **27**, 618 (1926).

[35] E. Lopez-Poveda and R. Meddis, "A human nonlinear cochlear filterbank", J. Acoust. Soc. Am. **110**, 3107–3118 (2001).

[36] A. Lundeby, T.E. Vigran, H. Bietz, and M. Vorlander, "Uncertainties of Measurements in Room Acoustics", Acustica **81**, 344–355 (1995).

[37] R. Meddis and M. Hewitt, "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. i: pitch identification", J. Acoust. Soc. Am. **89**, 2866–2894 (1991).

[38] R. Meddis and L. OMard, "A unitary model of pitch perception", J. Acoust. Soc. Am. **102**, 1811–1820 (1997).

[39] B. Moore, *An introduction to the Psychology of Hearing*, 2nd edition (Academic, London) (1982).

[40] K. Mouri and K. Akiyama, "Preliminary study on raccomended time duration of source signals to be analyzed, in relation to its effective duration of the auto-correlation function", J. Sound and Vib. **241**, 87–95 (2001).

[41] I. Nakayama, "Preferred time delay of a single reection for performers", Acustica **54**, 217–221 (1984).

[42] I. Nakayama, "Preferred direction of a single reection for a performer", Acustica **65**, 205–208 (1988).

[43] R.D. Patterson, and B.C.J. Moore, Auditory filters and excitation patterns as representations of frequency resolution, in *Frequency Selectivity in Hearing*, 123–177. London: Academic (1986).

[44] R. D. Patterson, M. H. Allerhand, and C. Giguère, "Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform.", J. Acoust. Soc. Am. **98**, 1890–1894 (1995).

[45] R. Patterson and J.Holdsworth, "A functional model of neural activity", Advanced in Speech, Hearing and Language Processing **3**, 547–563 (1996).

[46] W.W. Peterson, T.G. Birdsall, and W.C. Fox, "The Theory of Signal Detectability", IRE Trans. Inform. Theory **4**, 172–212 (1954).

[47] R. Pompoli and Y. Ando, "Opera house acoustics", Special issue of J. Sound and Vib **232** (2000).

[48] D. Pressnitzer, R. Patterson, and K. Krumbholz, "The lower limit of melodic pitch", J. Acoust. Soc. Am. **109**, 20742084 (2001).

[49] L. Rabiner, M. Cheng, A. Rosenberg, and C. McGonegal, "A comparative performance study of several pitch detection algorithms", IEEE Trans. Acoust., Speech, Signal Processing **ASSP-24**, 399–417 (1976).

[50] H. Sakai, S. Sato, N. Prodi, , and R. Pompoli, "Measurement of regional environmental noise by use of a pc-based system. an application to the noise near airport 'g. marconi' in bologna", J. Sound and Vib. **241**, 57–68 (2001).

[51] S. Sato, T. Kitamura, and Y. Ando, "Loudness of sharply (2068 db/octave) filtered noises in relation to the factors extracted from the autocorrelation function", J. Sound and Vib. **250**, 47–52 (2002).

[52] M. Schroeder, "New method of measuring reverberation time", J. Acoust. Soc. Am. **37**, 407–412 (1965).

[53] M. Schroeder, "An integrable model for the basilar membrane", J. Acoust. Soc. Am. **53**, 429–439 (1965).

[54] M. Schroeder, "Model for mechanical to neural transduction in the auditory receptor", J. Acoust. Soc. Am. **74**, 1055–1060 (1974).

[55] M. Schroeder, D. Gottlob and F. Siebrasse, "Comparative study of Europen concert halls: correlation of subjective preference with geometric and acoustic parameters", J. Acoust. Soc. Am. **57**, 1195–1201 (1974).

[56] Z. Sen, "Visual approach for automatic pitch period extimation", ?? **??**, 1339–1342 (2000).

[57] C. Shahnaz, W. Zhu, and M. Ahmad, "A robust pitch estimation approach for colored noise-corrupted speech", Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on **4**, 3143–3146 (2005).

[58] W. Siebert, "Frequency discrimination in the auditory system: place or periodicity mechanisms?", Proceedings of the IEEE **58**, 723–730 (1970).

[59] M. Smith and E.A. Wilson "A model of the auditory threshold and its application to the problem of the multiple observer", Psychol. Monogr **67**, (1953).

[60] D. Soderquist and J. Lindsey, "Physiological noise as a masker of low frequencies: the cardiac cycle", J. Acoust. Soc. Am. **52**, 1216–1220 (1972).

[61] Y. Soeta, K. Ohtori, K. Fujii, and Y. Ando, "Measurement of sound transmission by hall doors", J. Sound and Vib. **241**, 79–86 (2001).

[62] Y. Soeta, "Annoyance of bandpass-filtered noises in relation to the factor extracted from autocorrelation function", J. Acoust. Soc. Am. **116**, 3275–3278 (2004).

[63] W.P.Jr. Tanner and J.A. Swets, "The human use of information, I : signal detection for the case of the signal known exactly", Inst. Radio Engrs Transactions on Information Theory PGIT-4 (1954).

[64] W.P.Jr. Tanner, "Theory of recognition", J. Acoust. Soc. Am. **28**, 882–888 (1956).

[65] L.L. Thurstone, "Psychophysical Analysis", AM. J. Psychol. **38**, 368–389 (1927).

[66] L.L. Thurstone, "A Law of Comparative Judgment", Psychol. Rev. **34**, 273–286 (1927).

[67] H. Urkowitz, "Energy detection of unknown deterministic signals", Proceedings of the IEEE **55**, 523–531 (1967).

[68] N. Viemeister, "Temporal modulation transfer functions based upon modulation thresholds", J. Acoust. Soc. Am. **66**, 1364–1380 (1979).

[69] M. Vorländer and H. Bietz, "Comparison of methods for measuring reverberation time", Acustica **80**, 205215 (1994).

[70] Watson, Architecture (1927).

[71] L. Wiegrebe, "Searching for the time constant of neural pitch extraction", J. Acoust. Soc. Am. **109**, 1082–1091 (2001).

[72] N. Wiener, "??", Acta Math. **117**, ?? (1930).

[73] N. Xiang, "Evaluation of reverberation times using a nonlinear regression approach", J. Acoust. Soc. Am. **98**, 2112–2121 (1995).

[74] W. Yost, R. Patterson, and S. Sheft, "A time domain description for the pitch strength of iterated rippled noise", J. Acoust. Soc. Am. **99**, 1066–1078 (1996).

[75] W. Yost, R. Patterson, and S. Sheft, "The role of the envelope in processing iterated ripple noise", J. Acoust. Soc. Am. **104**, 2349–2361 (1998).

List of Figures

List of Tables

Contents