ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INDUSTRIALES Y DE TELECOMUNICACIÓN

Titulación :

INGENIERO TÉCNICO DE TELECOMUNICACIÓN,
ESPECIALIDAD EN SONIDO E IMAGEN

Título del proyecto:

COMPARISON OF DIFFERENT METHODS FOR TIME DELAY ESTIMATION.

Marta Abad Sorbet
Tutor: D. Luis Serrano

Pamplona, 7 mayo 2010

Resumen

A lo largo de los últimos años, la estimación del retardo entre señales recibidas por micrófonos separados espacialmente ha demostrado ser un método útil para múltiples aplicaciones. Algunos ejemplos de estas aplicaciones basadas en este método son la localización de hablantes, la localización de fuentes de sonido, las comunicaciones de radares, la sismología, ...

En este proyecto se estudian tres métodos distintos. Se va a presentar cada uno de estos métodos y se van a implementar con Matlab. Estos tres métodos son la correlación cruzada (CC), la correlación cruzada general (GCC) y por último un algoritmo adaptativo basado en la descomposición en valores propios (AED, adaptive eigenvalue decomposition algorithm).

Se ha discutido las ventajas e inconvenientes de cada método y se ha realizado una comparación entre los tres algoritmos. Para ello, utilizando las implementaciones de cada algoritmo, se han realizado diferentes simulaciones para ver como influye en la estimación del retardo según se varíe la reflexión de las paredes de la sala en la que se encuentra la fuente de sonido y los sensores, la distancia de la fuente a los sensores o las respuestas al impulso de la sala.

El objetivo es:

1. Realizar la mejor aproximación posible del retardo.
2. Comparar los diferentes métodos bajo distintas condiciones.

Los modelos matemáticos que describen el problema de la estimación del retardo entre señales recibidas por los sensores son: el modelo ideal y el modelo real. Los dos primeros algoritmos utilizados se basan en el modelo ideal es decir, no tiene en cuenta las reflexiones de el sonido, sino que unicamente tiene en cuenta la señal directa captada por los micrófonos.

El primer método utilizado es la correlación cruzada (CC). La correlación cruzada es una de las soluciones básicas para este problema y muchos otros métodos están basados en éste. La CC considera que el retardo entre las señales se corresponde con el pico máximo de la correlación cruzada.

Para mejorar la detección del pico y por consiguiente la estimación del retardo se usan filtros o funciones de ponderación despues de realizar la correlación cruzada. El retardo estimado es obtenido como el *time-lag* que maximiza la correlación cruzada entre las versiones filtradas de las señales recibidas en los sensores. Esta técnica se llama correlación cruzada general.

La mayoría de los métodos se basan en un modelo matemático ideal. En el método propuesto por Benesty [3] se presenta un nuevo algoritmo basado en un modelo real en el que se tiene en cuenta la reverberación. Este método usa la desomposición en valores propios para estimar el retardo en ambientes reverberantes. El vector propio correspondiente con el mínimo valor propio de la matriz de covarianza de las señales recibidas por los micrófonos contiene la respuesta al impulso entre la fuente de sonido y las señales captadas por los micrófonos.

En este proyecto se ha realizado un estudio comparativo entre estos tres métodos explicados. Para las simulaciones se han usado dos tipos de señales: señales aleatorias y se señales de hablantes.

El sonido en una sala esta compuesta por el sonido directo de la fuente y el sonido reflejado. En una sala, parte del sonido es reflejado, parte es absorbida por los materiales y otra parte se transmite a traves de estos materiales.

En el caso en el cual la fuente de sonido es un hablante se han realizado dos experimentos distintos. El primer experimento se realiza con señales obtenidas en una sala sin reverberación y el segundo experimento con señales obtenidas en una sala reverberante. En cada caso, con cada uno de los métodos, se ha estimado el retardo y se ha comparado con el real. En el primer caso (sala anecoica), la señal esta grabada con una frecuencia de muestreo de de 48000 Hz y el retardo entre las dos señales captadas por los dos micrófonos es de 200 muestras. En el caso de la sala reverberante (sala ecoica) el coeficiente de reflexión de las paredes es de 0.3.

Después de realizar la comparación entre el retardo obtenido teniendo en cuenta la reverberación de la sala o no, a continuación se ha modificado la respuesta al impulso de la sala. La reverberación de la sala puede ser simulada realizando la convolución de la señal de entrada con la respuesta al impulso de la sala. Se ha visto cuál es el efecto sobre la estimación del retardo.

Por último se han comparado los resultados cuando se varían los coeficientes de reflexión de las paredes y cuando se modifica el tamaño de la sala. En este ejemplo vamos a realizar la simulación en una sala de dimensiones ocho veces el volumen de la sala utilizada hasta ahora.

Las conclusiones a las cuales se han llegado han sido las siguientes:

Si comparamos el método de la correlación cruzada con el de la correlación cruzada general el resultado no varía de manera significante. El objetivo del filtro o de la función de ponderación en este último método es que el pico que se corresponde con el retardo sea más claro, pero bajo las condiciones que hemos estudiado la diferencia es muy pequeña. Los dos métodos son eficientes tanto en salas reverberantes como en salas no reverberantes, sin embargo, aunque la diferencia es pequeña, el pico que se obtiene con la correlación cruzada general es mas claro y mejor definido.

Si comparamos ahora estos dos métodos basados en la correlación con el algoritmo AED no se observa mucha mejora en salas con reverberación baja.

Cuando modificamos las respuestas al impulso de la sala se observa como con todos los algoritmos tienen soluciones muy cercanas a el retardo real, sin embargo el método más eficiente y con menor error es el AED.

Cuando se incrementa la reverberación de la sala los métodos pierden precisión en la estimación. Al aumentar el tamaño de la sala la estimación empeora, pero con coeficientes de reverberación bajos los resultados son aceptables. El problema es cuando se utilizan coeficientes de reflexión de las paredes elevados (mayor de 0.6). En ese caso los métodos basados en la correlación tienen más error que el método AED.

Declaration

I hereby declare that I have done the present work autonomously. I have used no other sources than indicated in bibliography.

Karlsruhe, March the 27th, 2010

Marta Abad Sorbet

TABLE OF CONTENTS

LIST OF FIGURES

TABLE INDEX

ABSTRACT

Time delay estimation (TDE) between signals received at different sensors has been proven to be a useful method for many applications. Speaker localization and meeting activity detection are some examples of applications based on TDE.

Three time delay estimation methods are described in this paper and implemented using MATLAB. These methods are cross-correlation (CC), General cross-correlation(GCC) and a method based on an adaptive eigenvalue decomposition algorithm (AED) [3]. We will discuss the pros and cons of each individual algorithm, and outline their inherent relationships. We also provide experimental simulations to illustrate the results.

The objetive is:

1. Estimate the best approximation time-delay
2. Compare different methods.

1. INTRODUCTION

During the last years, the problem of estimating the time delay between signals received at two spatially separated microphones has been considered for a variety of applications. Time delay estimation has been a research topic of significant practical importance in many fields:

- Radar Communications
- Microphone array processing systems
- Speech recognition
- Source localization
- seismology
- geo- physics
- ultrasonics
- hands-free communications, etc

This physical problem in two dimensions is shown in Figure 1.1:

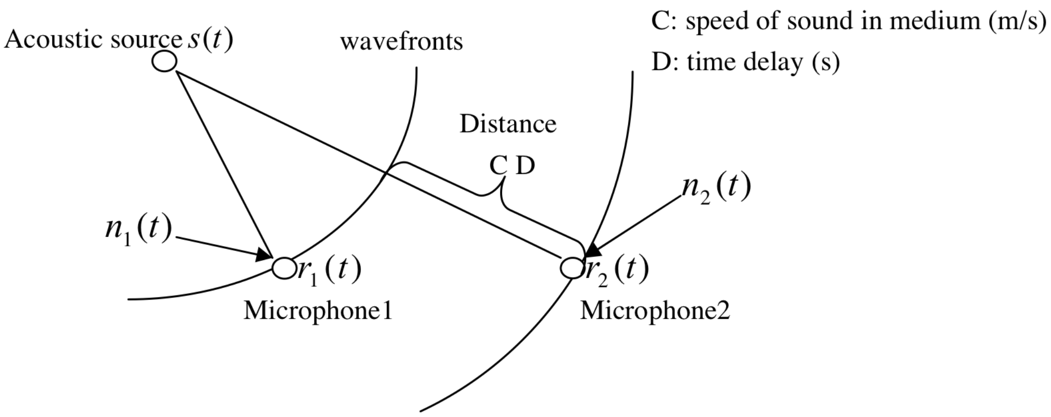

Figure 1.1 Time-delay associated with two microphones

The received signal at the two microphones can be modelled by:

$$x_1(n) = s(n) + b_1(n)$$
$$x_2(n) = s(n\text{-}\tau) + b_2(n) \qquad (1.1)$$

where x_1 (n) and x_2 (n) are the outputs of two spatially separated microphones, s(n) is the source signal, and b_1(n) and b_2(n) represent the additive noises and τ yields the time delay between the two received signals.

The objective is to estimate the time delay. One specific problem, common to all methods, is the severe degradation of performance at low signal-to-noise ratio (SNR) for TDE of narrowband signals.

Time delay estimation is difficult because of the nonstationary of speech and of room reverberation. Furthermore, signal-to-noise ratio (SNR) becomes also a problem if the SNR becomes smaller than 20 dB [3].

The estimation would be an easy task if the two received signals were merely a delayed and scaled version of each other. In reality, however, the source signal is generally in ambient noise since we are living in a natural environment where the existence of noise is inevitable. Furthermore, each observation signal may contain multiple attenuated and delayed replicas of the source signal due to reflections from boundaries and objects. This multipath propagation effect introduces echoes and spectral distortions into the observation signal, termed as reverberation, which severely deteriorates the source signal. All these factors make time delay estimation a complicated and challenging problem.

2. TIME DELAY ESTIMATION BASICS

2.1 INTRODUCTION

The mathematical models that describe an acoustic environment for the TDE problem will be presented. There are the ideal single-path propagation model, the multipath model, and the reverberation model. We are going to describe each model.

2.2 MODELS FOR TDE PROBLEM

2.2.1 Ideal model

If we take a signal s(n) propagating through a generic noisy free space, the signal acquired by the i-th (i= 1,2) microphone can be expressed as follows [3]

$$x_i(n) = \alpha_i\, s\,(n\text{-}\tau_i) + b_i(n) \tag{2.1}$$

where x_i (n) denote the i-th microphones signal, α_i is an attenuation factor due to propagation loss, τ_i is the propagation time from the unknown source s(n) to microphone i, and b_i (n) is additive noise. The time delay of arrival between the two microphones signals 1 and 2 is defined as

$$\tau_{12} = \tau_1\text{-}\tau_2 \tag{2.2}$$

Furthermore, we assume that s(n), b_1(n), and b_2(n) are zero-mean, uncorrelated, stationary Gaussian random process. In this case, a mathematically clear solution for τ_{12} can be obtained from the ideal model that is widely used for the classical TDE problem.

This model is ideal in the sense that the solution for determining τ_{12} is clear. Indeed, let's first write equation (2.1) in the frequency domain

$$x_i(f)=\alpha_i \mathrm{S}(f)e^{-j2\pi f\tau_1}+B(f) \tag{2.3}$$

and then take the (complex) sign of the cross-spectrum S_{x1x2} between $X_1(f)$ and $X_2(f)$,

$$sgn[S_{x1x2}(f)]=sgn\left[E\left\{X_1(f)X_2^*(f)\right\}\right]=e^{-j2\pi f\tau_{12}} \tag{2.4}$$

we can easily see the inverse Fourier transform of equation (2.4) will result in a sharp peak in the time domain corresponding to the delay τ_{12} [3].

2.2.2 Multipath model

The ideal propagation model takes only the direct-path signal into account. In many situations, however, each sensor receives multiple delayed and attenuated replicas of the source signal due to reflections of the wavefront from boundaries and objects in addition to the direct-path signal. In this case, the received signals are often described mathematically as [4]

$$x_i(n) = \sum_{j=1}^{J} \alpha_{ij}\, s\,[n\text{-}t\text{-}\tau_{ij}] + b_i(n) \tag{2.5}$$

where α_{ij} is the attenuation factor from the unknown source to the i-th sensor via the j-th path, t is the propagation time from the source to sensor 0 via direct path, τ_{ij} is the relative delay between sensor n and sensor 0 for path m with τ_{01} = 0, J is the number of different paths, and $b_i[n]$ is stationary Gaussian noise and assumed to be uncorrelated with both the source signal and the noise signals observed at other sensors.

The primary interest of the TDE problem for this model is to measure τ_{n1} , i=1,...,N−1, which is the TDOA between sensor i and sensor 0 via direct path [4].

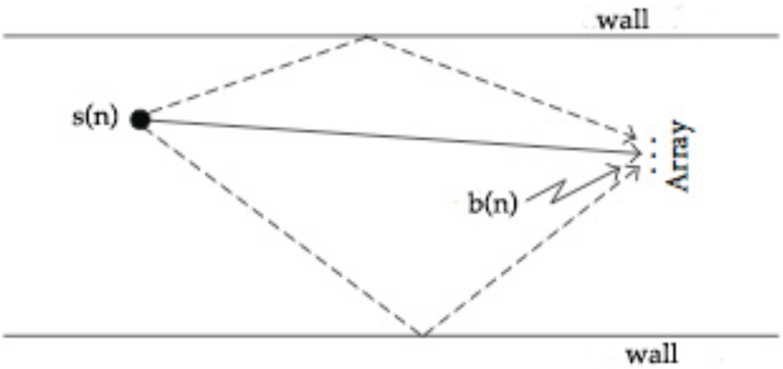

Figure 2.1. Ilustration of the signal model in a multipath environment

2.2.3 Real model(Real reverberant model)

The ideal free-field model model is simple and only few parameters need to be determined. But unfortunately, in a real acoustic environment we must take into account the reverberation of the room. Then, a more complicated but more complete model for the microphones signals $x_i(n)$ (i=1,2) can be expressed as follows[3]

$$x_i(n) = g_i * s(n) + b_i(n) \tag{2.6}$$

where * denotes convolution and g_i is the acoustic impulse response of the channel between the source and the i-th microphones. Moreover, $b_1(n)$ and $b_2(n)$ might be correlated, which is the case when the noise is direccional.

For the real reverberant model, we do not have an "ideal" solution to the TDE problem, as for the previous model, unless we can accurately (and blindly) determine the two impulse responses, which is a very challenging problem [3].

3. TDE IN REVERBERANT ENVIRONMENT

3.1 INTRODUCTION

There are many algorithms to estimate the time delay. Numerous algorithms have been developed, and can be categorized from the following points of view [5]:

i. the number of sources in the wavefield, that is, single-source TDE techniques and multiple-source TDE techniques.

ii. how the propagation condition is modeled, that is, the ideal single-path propagation model, the multipath propagation model, and the reverberation model.

iii. what analysis tools are employed, for example, generalized cross-correlation (GCC) method, higher-order-statistics-(HOS) based approaches , and blind channel identification based algorithms.

iv. how the delay estimate is updated, that is, non-adaptive and adaptive approaches.

The cross-correlation (CC) method is one of the basic solutions to the TDE problem. Many other TDE methods are based on this algorithm. The CC method considers the time argument that corresponds to the maximum peak in the output as the estimated time delay.

To improve the peak detection and time delay estimation, various filters, or weighting functions, have been suggested to be used after the cross correlation [2]. The estimated delay is obtained by finding the time-lag that maximizes the cross-correlation between the filtered versions of the two received signals. This technique is called generalized cross-correlation (GCC) [2].

The GCC method, proposed by Knapp and Carter in 1976, is the most commonly used method for TDE due to their accuracy and moderate computational complexity. The role of the filter or weighting function in GCC method is to ensure a large sharp peak in the obtained cross-correlation thus ensuring a high time delay resolution [1].

In this technique, time delay is obtained as the time-lag that maximizes the cross correlation between filtered versions of the received signals. There are many techniques used to select the weighting function; such as the Roth Processor, the Smoothed Coherence Transform (SCOT), the Phase Transform (PHAT), and the Maximum Likelihood (ML) estimator [1, 2]. They are based on maximizing some performance criteria. These correlation-based methods yield ambiguous results when the noises at the two sensors are correlated with the desired signals [1].

Most of the methods are based on an ideal model. In the paper written by Benesty [3], a new method based on a real signal model with reverberation is proposed. This method uses eigenvalue decomposition to estimate the time delay in reverberant environment. The eigenvector corresponding to the minimum eigenvalue of the covariance matrix of the microphone signals contains the impulse responses between source and the microphones signals [3].

Adaptive algorithms such as LMS can also be introduced into the TDE . In these algorithms, the delay estimation process is reduced to a filter delay that gives minimal error.

3.2 CROSS CORRELATION

3.2.1 BACKGROUND

The correlation analysis is a method for determining the degree of similarity between variations of two signals in time. The autocorrelation is the point-to-point multiplication of a waveform by a delayed version of itself, followed by a summation process or integration. This indicates the similarity of a signal with a time-delayed version. Mathematically, the autocorrelation function of a function x (t), denoted R_{xx} (τ) is given by:

$$R_{xx}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x(t)x(t+\tau)dt \tag{3.1}$$

where τ is the delay between the two versions of the function f(t).

In signal processing, cross-correlation is a measure of similarity of two waveforms as a function of a time-lag applied to one of them. The cross-correlation involves two time signals x (t) and y (t). It consists of the multiplication of a signal x (t) by a delayed version of y (t) followed by an addition or integration. The cross correlation function is given by:

$$R_{xy}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x(t)y(t+\tau)dt \tag{3.2}$$

and it indicates how much the signal resemble to a delayed version of the second.

A system, like the one in the figure 1.1 which represents a source in the presence of noise controlled by two separate sensors, it was shown:

$$\begin{aligned} x_1(n) &= s(n) + b_1(n) \\ x_2(n) &= \alpha\, s\,(n\text{-}\tau) + b_2\,(n) \end{aligned} \tag{3.3}$$

Assuming that propagation medium is homogeneous, where b_1 (n), b_2 (n) are random signals. x_1 (n) consists of the signal transmitted by the source s(n) plus noise b_1 (n) caused by a source close to her. x_2 (n) is composed of a delayed version τ seconds of the signal s(n) multiplied by an attenuation α and b_2 is the noise caused by another source. The random signals b_1 (n) and b_2 (n) are uncorrelated, i.e. there is no relationship between the frequency components that make up those signals.

3.2.2. CROSS CORRELATION (CC) ALGORITHM

One common method to estimate the time delay D, is to compute the cross correlation function between the received signals at two microphones. Then locate the maximum peak in the output which represents the estimated time delay [1].

The cross-correlation (CC) method is the most straightforward and the earliest developed TDE algorithm, which is formulated based on the single-path propagation model given in (2.1) with only two receivers [4].

The CC can be modelled by:

$$R_{x1x2}(\tau) = \mathrm{E}\,[x_1(n)\; x_2(n\text{-}\tau)] \tag{3.4}$$

$$D_{cc}\text{=arg max}\,[R_{x1x2}\,(\tau)] \tag{3.5}$$

A block diagram of a cross-correlation processor is shown in Figure 3.1.

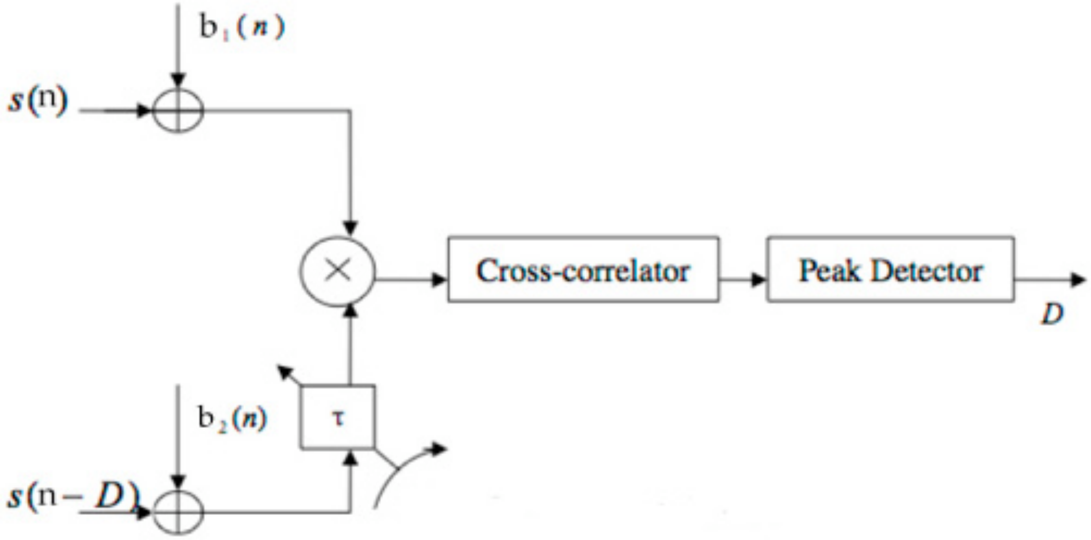

Figure 3.1. Cross-correlation processor

3.3 GENERAL CROSS CORRELATION (GCC)

3.3.1 BACKGROUND

The GCC is the Inverse Furier Transform (IFT) of the cross spectrum between $x_1(n)$ and $x_2(n)$, multiplied by a weighting function Φ(t). In this section we discuss the GCC function and the weighting functions.

The generalized cross-correlation (GCC) algorithm can be treated as an improved version of the CC method. Not only does it unify various correlation-based algorithms, but it also provides a mechanism to incorporate knowledge to improve the performance of TDE [4].

This method is based on the ideal model but is the most commonly used even in very reverberant environments. This method has gained its great popularity since the landmark paper was published by Knapp and Carter in 1976. The delay estimate is obtained as the value of τ that maximizes the general cross correlation function given by:

$$\begin{aligned}\psi_{x_1x_2}(\tau) &= \int_{-\infty}^{+\infty} \Phi(f) S_{x_1x_2}(f) e^{j2\pi f\tau} df \\ &= \int_{-\infty}^{+\infty} \Psi_{x_1x_2}(f) e^{j2\pi f\tau} df,\end{aligned} \tag{3.6}$$

where $\Phi(f)$ is a weighting function and

$$\Psi_{x_1x_2}(f) = \Phi(f) S_{x_1x_2}(f) \tag{3.7}$$

is the generalized cross-spectrum. Then, the GCC TDE may be expressed as:

$$T_d = \arg\max_\tau \psi_{x1x2}(\tau) \tag{3.8}$$

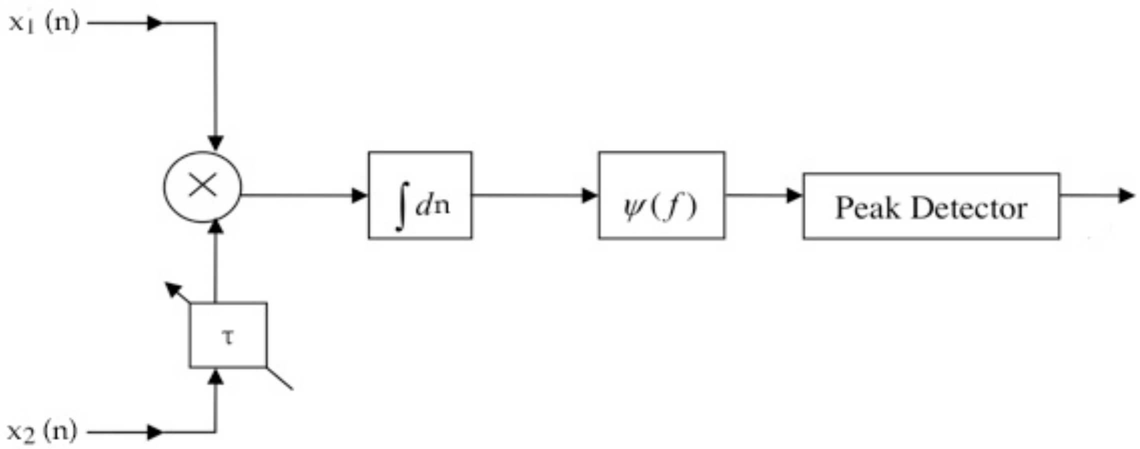

Figure 3.2. Generalized cross-correlation processor

S_{x1x1}(f) and S_{x2x2}(f) are the power spectra of input and output autocorrelation signals, respectively, and S_{x1x2} (f) is the power spectrum of cross-correlation. These are defined as follows:

$$S_{x1x1}(f)=\int_{0}^{\infty} R_{x1x1}(\tau)\, e^{-j2\pi ft\, dt} \tag{3.9}$$

$$S_{x2x2}(f)=\int_{0}^{\infty} R_{x2x2}(\tau)\, e^{-j2\pi ft\, dt} \tag{3.10}$$

$$S_{x1x2}(f)=\int_{0}^{\infty} R_{x1x2}(\tau)\, e^{-j2\pi ft\, dt} \tag{3.11}$$

The coherence function γ^2 (f) is used as a measure in each frequency component f to know the dependence between the output y(t) and the input x(t) and gives us an idea of the signal to noise ratio (SNR) between input and output.

$$\gamma^2(f) = \frac{|S_{x1x2}(f)|^2}{S_{x1x1}(f)S_{x2x2}(f)} \tag{3.12}$$

3.3.2 GENERAL CROSS CORRELATION ALGORITHM

Typically, for periodic signals containing high-power components, it can be difficult to estimate the time delay because the frequencies that don't correspond to the periodic signals are filtered when we make the correlation. Because of this we use prewhitening filters to have a better estimation [10].

A way to sharpen the cross correlation peak is to whiten the input signals by using weighting function, which leads to the so-called generalized cross-correlation (GCC).

In general, the prefilter enhances the frequency bands where the signal is strong and attenuates the bands where noise is strong.

The selection of Φ (f) depends on the method being used. Table 3.1 presents the different weighting functions.

Method	$F(f)=H_1(f)H^*_2(f)$
SCC	1
ROTH	$1/S_{x1x1}(f)$
SCOT	$1/\sqrt{S_{x1x1}(f)S_{x2x2}(f)}$
PHAT	$1/abs(S_{x1x2}(f))$
ML	$\gamma^2_{x1x2}(f)/[1-\gamma^2_{x1x2}(f)]abs(S_{x1x2}(f))$

Table 3.1. weighting functions

The cross-correlation between signals x_1 (n) and x_2 (n) is related to the function of power density spectrum crossed by Inverse Fourier Transform (IFT):

$$S_{x1x2}(f) = \alpha\, S_{x1x1}(f)\, e^{-j\,2\pi f\; Td} + S_{b1b2}(f) \qquad (3.13)$$

If there is no relationship between b_1 (n) and b_2 (n) then S_{b1b2} (t) = 0. If S_1 (t) is white noise, then S_{x1x1} (f) be a constant. The IFT of S_{x1x1} (f) is a delta. The IFT of S_{x1x2} (f) is a delta with a delay T_d.

The choice of Φ (f) is important in practice. We are going to describe the most important functions:

3.3.2.1 UNFILTERED

When the filters H1 (f) = H2 (f) = 1, the GCC function will be equivalent to the standard cross-correlation function (CC):

$$\psi_{x1x2}(\tau) = \int_{-\infty}^{+\infty} \mathrm{S}_{x1x2}(f)\, e^{j2\pi f\tau}\, df \tag{3.14}$$

The estimate delay is the abscissa's value where is localizated the highest peak of the expressed function.

3.3.2.2 ROTH

Roth weighting function is:

$$\Phi(f) = \frac{1}{\mathrm{S}_{x1x1}(f)} \tag{3.15}$$

When we evaluate the GCC we have:

$$\psi_{x1x2}(\tau) = \int_{-\infty}^{+\infty} \frac{\mathrm{S}_{x1x2}(f)}{\mathrm{S}_{x1x1}(f)}\, e^{j2\pi f\tau}\, df \tag{3.16}$$

The equation estimates the impulse response

$$H_m(f) = \frac{\mathrm{S}_{x1x2}(f)}{\mathrm{S}_{x1x1}(f)} \tag{3.17}$$

Which is the best aproximation of the mapping x_1 (n) x_2 (n).

3.3.2.3 SCOT

The reason for the division between the cross spectrum and autospectrum of x_1 (t) makes sense when you are in a linear system, as the case of the Roth processor.

This is one technique that does not give preference to S_{x1} (f) or S_{x2} (f). The weighting function is:

$$\Phi(f) = \frac{1}{\sqrt{S_{x1x1}(f)\, S_{x2x2}(f)}} \tag{3.18}$$

Generalized cross-correlation function gives:

$$\psi_{x1x2}(\tau) = \int_{-\infty}^{+\infty} \frac{S_{x1x2}(f)}{\sqrt{S_{x1x1}(f) S_{x2x2}(f)}}\, e^{j2\pi f\tau}\, df \tag{3.19}$$

Function SCOT is a very robust estimator for signals with low SNR. When $S_{x1x1}(f)$=$S_{x2x2}(f)$ SCOT is equivalent to Roth.

3.3.2.4 PHAT

The PHAT is a GCC procedure which has received considerable attention due to its ability to avoid the spreading of the peak of the correlation function [9, 2]. This can be expressed mathematically by

$$\Phi(f) = \frac{1}{|S_{x1x2}(f)|} \tag{3.20}$$

$$\psi_{x1x2}(\tau) = \int_{-\infty}^{+\infty} \frac{S_{x1x2}(f)}{|S_{x1x2}(f)|}\, e^{j2\pi f\tau}\, df \tag{3.21}$$

$$T_d\text{=arg max }[\mathrm{S}_{x1x2}(f)] \tag{3.22}$$

where $\mathrm{S}_{x1x2}(f)$ is the cross-spectrum of the received signal, $\psi_{x1x2}(f)$ is the PHAT weighting function and T_d is the time delay estimation. According to [10], only the phase information is preserved after the cross-spectrum is divided by its magnitude. Ideally (no additive noise), this processor approaches a delta function centered at the correct delay. In noiseless case, it depends only on the impulse responses and can perform well in moderately reverberant room.

3.3.2.5 HT

To use the model shown in equation described above

$$x_i(n) = \alpha_i\, s\,(n\text{-}\tau_i) + b_i(n) \tag{2.1}$$

it is necessary to assume that s(n) and b_i (n) are gaussians.

The weighting function is

$$\Phi(f) = \frac{1}{|\mathrm{S}_{x1x2}(f)|}\,\frac{|\gamma_{x1x2}(f)|^2}{[1\text{-}\,|\gamma_{x1x2}(f)|^2\,]} \tag{3.23}$$

The ML estimator is considered the optimal weighting function, it gives more weight where the coherence is close to unity and decreasing where coherence is near zero [1]. The ML estimator achieves minimum variance only if it has good SNR.

3.4 ADAPTIVE EIGENVALUE DECOMPOSITION

All the algorithms outlined in the previous sections achieve delay estimate by measuring the cross-correlation between two or multiple channels. A common assumption with these methods is that each sensor receives only the direct-path signal. In this section a completely different approach than GCC is proponed [3].

An adaptive eigenvalue decomposition (AED) algorithm was proposed to deal with TDE in room reverberant environment by Benesty [3].

This method focuses directly on the impulse responses between the source and the microphones in order to estimate the time-delay. Apparently, this algorithm takes fully into account the reverberation effect during time delay estimation [3].

3.4.1 BACKGROUND

We assume that the system (room) is linear and time invariant; therefore, we have the following relation:

$$x_1^T(n)\; g_2 = x_2^T(n)\; g_1 \;, \tag{3.24}$$

where

$$x_i(n) = [x_i(n) \quad x_i(n\text{-}1) \;\; \dots \;\; x_i\,(n\text{-}M\text{+}1)]^{\,T} \quad , \quad i\text{=}1,2 \tag{3.25}$$

are vectors of signal samples at the microphone outputs, *T* denotes the transpose of a vector or a matrix, and the impulse response vectors of length *M* are defined as

$$g_i = [g_{i,0} \quad g_{i,1} \;\; \cdots \;\; g_{i,M\text{-}1}\,]^{\,T} \quad , \quad i\text{=}1,2 \tag{3.26}$$

This linear relation follows from the fact that $x_i = s * g_i$, $i =$ 1,2, thus $x_1 {*} g_2 = s {*} g_1 {*} g_2 = x_2 {*} g_1$.

The covariance matrix of the two microphone signals is:

$$R=\begin{bmatrix}R_{x1x1} & R_{x1x2}\\ R_{x2x1} & R_{x2x2}\end{bmatrix} \tag{3.27}$$

where

$$Rx_ix_j=\mathrm{E}\left\{x_i(n)x_j^T(n)\right\},\quad i,j=1,2 \tag{3.28}$$

is the covariance matrix of the sensor signals.

Consider the 2 M x 1 vector

$$u=\begin{bmatrix}g_2\\ -g_1\end{bmatrix} \tag{3.29}$$

From Eqs. 3.22 and 3.25 , it can be seen that:

$$Ru=0 \tag{3.30}$$

This implies that vector u which consists of two impulse responses is in the null space of R. More specifically, u is the eigenvector of R corresponding to the eigenvalue 0.

Moreover, the covariance matrix R has one and only one eigenvalue equal to 0 if the following two conditions hold [4]:

(i) the polynomials formed from g_1 and g_2 are coprime, or they do not share any common zeros.
(ii) the autocorrelation matrix of the source signal s(n) is of full rank.

In practice, accurate estimation of the vector u is not trivial, because of the nature of speech, the length of the impulse responses, the background noise, etc. However, for this application we only need to find an efficient way to detect the direct paths of the two impulse responses [3].

3.4.2 ADAPTIVE ALGORITHM

In practice, it is simple to estimate iteratively the eigenvector (here u) corresponding to the minimum (or maximum) eigenvalue of R, by using an algorithm similar to the Frost algorithm which is a simple constrained Least-Mean- Square (LMS) [12].

In the following, we show how to apply these techniques to our problem. Minimizing the quantity u^T Ru with respect to u and subject to $\| u \|^2 = u^T u = 1$ will give us the optimum filter weights u_{opt}.

We are going to define the error signal ($\|\cdot\|$ denotes the l_2 norm of a vector or matrix):

$$e(n) = \frac{u^T(n)x(n)}{\|u(n)\|} \tag{3.31}$$

where $x(n) = [\ x_1^T(n) \quad x_2^T(n)\ ]^T$.

Note that minimizing the mean square value of $e(n)$ is equivalent to solving the above eigenvalue problem. Taking the gradient of $e(n)$ with respect to u(n) gives

$$\nabla e(n) = \frac{1}{\|u(n)\|}\left[x(n) - e(n)\ \frac{u(n)}{\|u(n)\|}\right] \tag{3.32}$$

and we obtain the gradient-descent constrained LMS algorithm:

$$u(n+1) = u(n) - \mu e(n) \nabla e(n) \tag{3.33}$$

where μ , the adaptation step, is a positive constant. Substituting Eqs. 3.29 and 3.30 into Eq. 3.31 gives:

$$u(n+1) = u(n) - \frac{\mu}{\|u(n)\|}\left[x(n)x^T(n)\ \frac{u(n)}{\|u(n)\|} - e^2(n)\frac{u(n)}{\|u(n)\|}\right] \tag{3.34}$$

and taking mathematical expectation after convergence, we get:

$$R\frac{u(\infty)}{\|u(\infty)\|} = \mathrm{E}\{e^2(n)\}\frac{u(\infty)}{\|u(\infty)\|} \tag{3.35}$$

the eigenvector u(∞) corresponding to the smallest eigenvalue $E[e^2(n)]$ of the covariance matrix R.

In practise, is advantageous to use the following adaptation scheme to avoid roundoff error propagation:

$$u(n+1) = \frac{u(n) - \mu e(n)\nabla e(n)}{\|u(n) - \mu e(n)\nabla e(n)\|} \tag{3.36}$$

Note that if this trick is used, then $\|u(n)\|$ (which appears in $e(n)$ and $\nabla e(n)$) can be removed, since we will always have $\|u(n)\| = 1$

A simplified algorithm

The algorithm Eq. 3.34 presented above is a little bit complicated and is very general to find the eigenvector corresponding to the smallest eigenvalue of any matrix R.

When an independent white noise signal is present on each sensor, it will regularize the covariance matrix; as a consequence, R does not have a zero eigenvalue anymore. In such a case, an estimate of the impulse responses can be achieved through the following algorithm, which is an adaptive way to find the eigenvector associated with the smallest eigenvalue of R [3].

In practice, if the smallest eigenvalue is equal to zero, which is the case here, the algorithm can be simplified as follows:

$$e(n) = u^T(n)x(n) \tag{3.37}$$

and

$$u(n+1) = \frac{u(n) - \mu e(n)x(n)}{\|u(n) - \mu e(n)x(n)\|} \tag{3.38}$$

with the constraint that $\|u(n)\| = 1$

With the identified impulse responses g_1 and g_2 , the time delay estimate is determined as the difference between two direct paths, that is [4]:

$$\tau = \arg\max_\tau |g_{1,\tau}| - \arg\max_\tau |g_{2,\tau}| \tag{3.39}$$

Note that this algorithm can be seen as an approximation of the previous one by neglecting the terms in $e^2(n)$ in Eq. 3.34, which is reasonable (since the smallest eigenvalue is equal to zero). In this application, the two algorithms Eq. 3.34 and 3.36 should have the same performance after convergence even with low SNRs. Moreover, in all experiments the unconstrained frequency-domain adaptive filter (UFLMS) is used to implement the impulse response estimation algorithm. Note that this algorithm is still efficient from a complexity point of view but it requires seven fast Fourier transform (FFT) operations per block (because we need to go back to the time-domain to apply the norm constraint), while the PHAT requires only three FFT operations per block [3].

4. SIMULATION AND RESULTS

4.1 INTRODUCCTION

In this chapter, a comparative study between the different methods is explained. For the simulations, we have used different test signals. We are going to compare the results depending on the kind of signal that is used. These signals are:

- Random signals
- Speech signals
- Speech signals using different room impulse responses.

The sound in a room is composed of the direct sound of the source and the reflected sound too. In a room, part of the sound will be reflected, while another part will be absorbed for the material, and another part will be transmitted through it. The first experiment involves a data obtained in nonreverberant (simulated by setting all the reflection coefficients to 0) and the second experiment involves a data obtained in reverberant environment.

In each case, we will estimate the time delay and compare it with the real time delay. Besides the study between real and simulated delay will also perform a comparison between different methods and under different conditions. The different approaches will be evaluated under different conditions considering accuray.

4.2 RESULTS

In this section, the obtained results after performing various tests with the three explained methods will be presented. The analysis will be divided into two parts depending on the data used to estimate the delay.

First of all, it is explained what occurs when is used a random signal with a delay fixed by us. In the second part we will use speech signals. Besides, these signals will be simulated varying the impulse response of the room.

As discussed earlier, the ideal propagation model takes only the direct-path signal (anechoic environment) into account. In many situations, each sensor receives multiple delayed and attenuated replicas of the source signal due to reflections of the wavefront from boundaries and objects in addition to the direct-path signal (echoic environment). For this reason, we will simulate scenes in two ways to compare the difference.

4.2.1 RANDOM SIGNALS

First of all, a random signal will be used to verify the accuracy of the three methods used in this work.

In this case the source is a random signal that is created with MATLAB with the *randn function* and a delay between the two sensors equal to 4.

We will examine the efficiency of the methods with this kind of signals. First of all, we will see the results using cross correlation method, then using GCC method and finally using the adaptive eigenvalue descomposition algorithm.

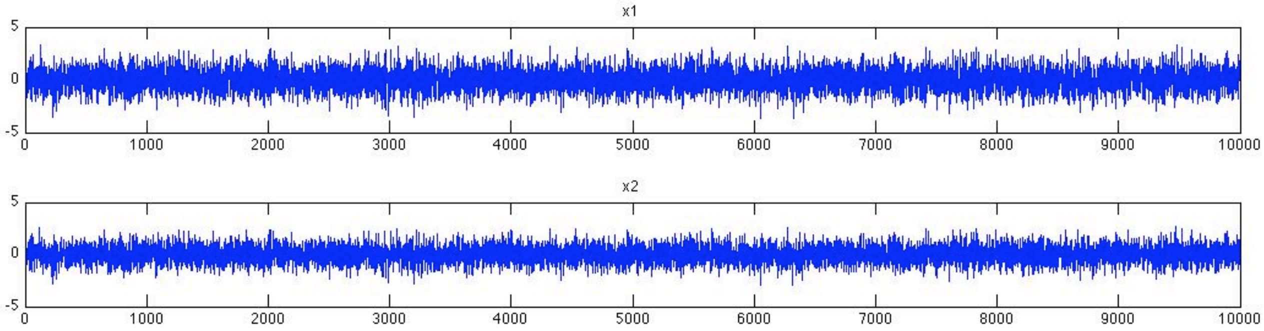

Fig. 4.1. Random signals

CROSS CORRELATION (CC):

We calculate the cross-correlation between two random signals with a delay of 4 samples. These signals can be consider as Gaussian noise. The result can be seen in the next figure:

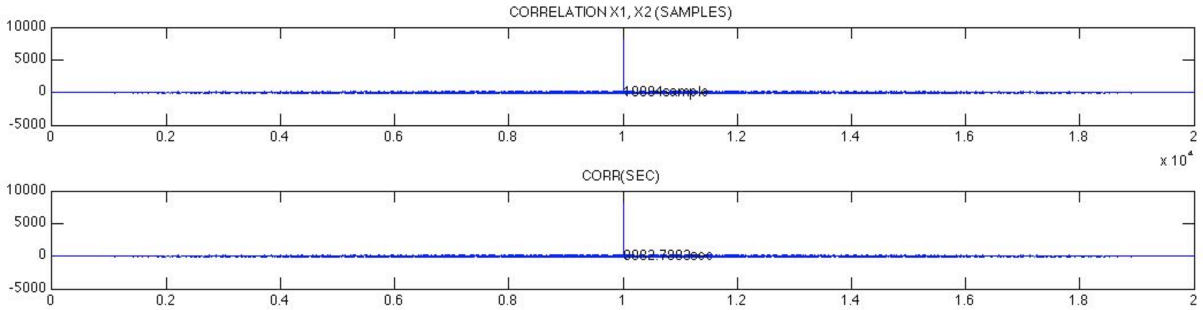

Fig. 4.2. Cross correlation when source is a random signal

The first and the second lines are the representation of the two random signals from which we estimate the delay. The third line shows the correlation between these two signals. The x-coordinate denotes the time, and the y-coordinate denotes the resulted cross-correlation.

As we said in theory, time delay estimation using CC algorithm is obtained as the lag time that maximizes the cross-correlation function between two received signals. The peak in this case is clearly seen in the graph and it is placed in sample 1004. From Figure 4.2, it can be seen that the peak occurs at the actual time delay.

This method works for random signals very efficiently. Then we will see the result of GCC method.

GENERAL CROSS CORRELATION (GCC):

Now, we are going to use the GCC method when the source is a random signal. The result can be seen in the following figure:

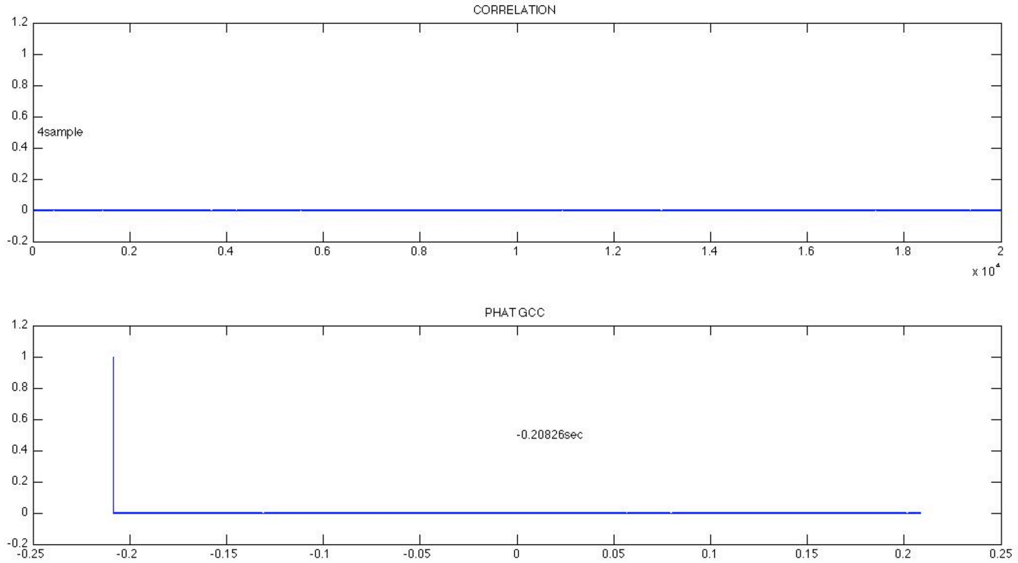

Fig. 4.3. GCC Path of random signals

The Figure 4.3 represents the General Cross Correlation (GCC) using Phat method. It is clear that the peak position corresponds to the actual time delay.

Figure 4.4 shows the results if we use the other weighting functions:

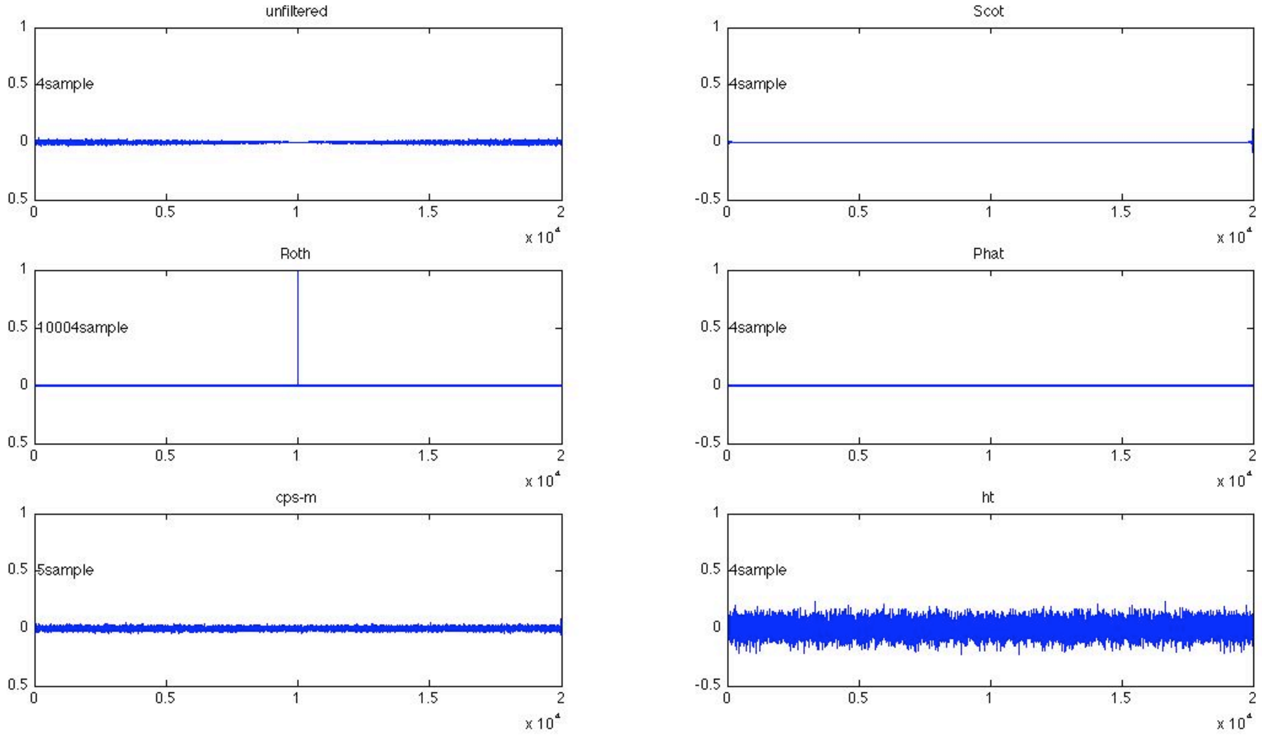

Fig. 4.4. Unfiltered, scot, Roth, Phat, cps-m, ht GCC with ramdom signals

We can see that in all cases, the peak correspondes with the real delay. Only cps-m case fails.

ADAPTIVE EIGENVALUE DECOMPOSITION ALGORITHM:

This is the last method we have to check when the source is a random signal. The results are:

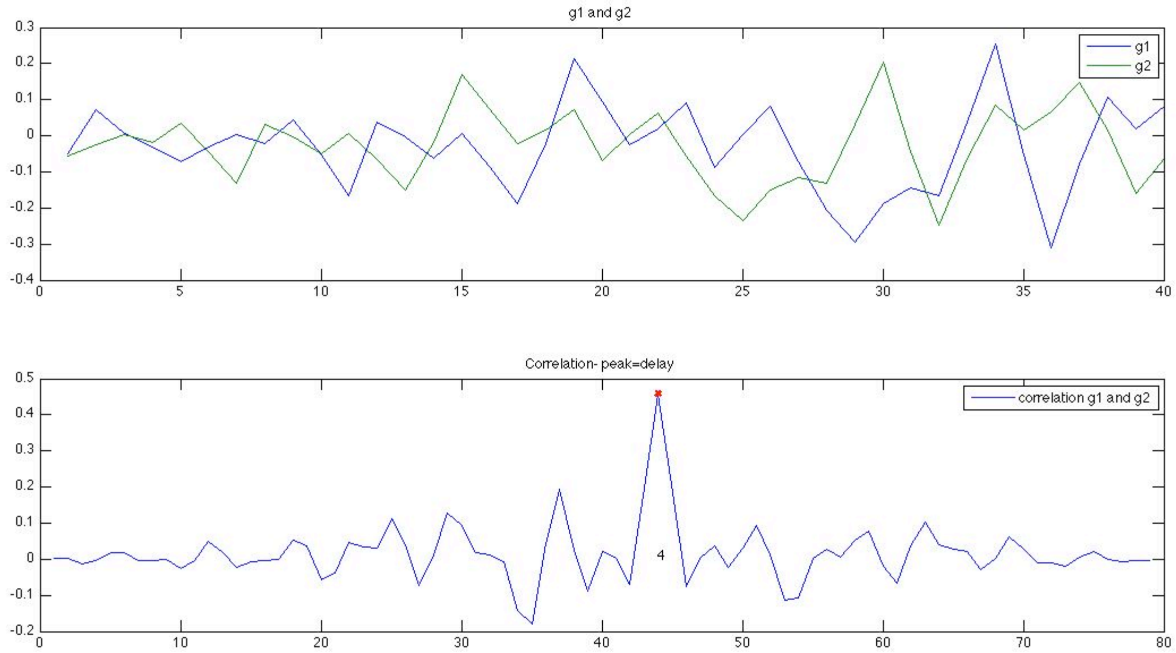

Fig. 4.5. AED applied to random signals

In figure 4.5 we can see in the first line the two impulses responses g_1 and g_2 and in the second line the correlation between these two signals. The peak position corresponds to the real delay.

From the results, we can conclude that the estimated delay with random signals with all algorithms is correct. The estimation of time delay is more accurate when the SNR is high, than when the SNR is low. The appearance of noise helps correct estimation of time delay.

The next table is a summary about the obtained results. It is a comparison between the three methods we used. It shows if the estimated delay is correct or not. True delays are compared with estimated delays.

Method	True delay	Estimate delay
CC	4	4
GCC	4	4
AED	4	4

Table 4.1 comparison between the three methods.

4.2.2 REAL SPEECH SIGNALS

The speech signal originated from the speaker's vocal cord contains a sequence of periodic correlation. It is considered that the voiced speech (with fundamental periodicity) signal is periodically correlated and the unvoiced signal is not [11].

In this section we will check the results when the source is a speech signal and it is received by two sensors between which there is a delay in the arrival of the signal.

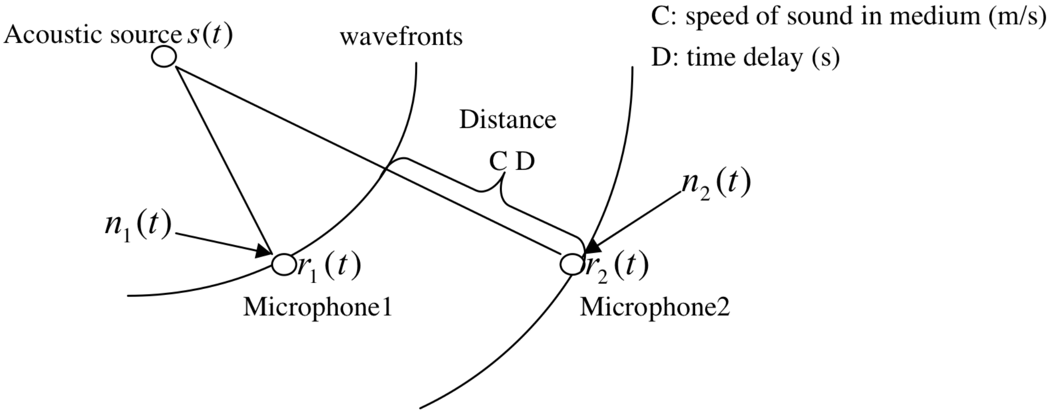

Fig. 4.6 Time-delay associated with two microphones

The analysis will be made first with a data obtained in nonreverberant environment (anechoic environment) and then in reverberant environment (anechoic environment).

4.2.2.1 ANECHOIC ENVIRONMENT

Under free-field conditions, only the direct sound which is radiated from the sound source is observed, without any obstacle for sound propagation.

In a free space, the acoustic waves are propagated from the source to infinity. In a room, the reflections of the sound on the walls produce a wave which is propagated in the opposite direction and comes back to the source. In anechoic rooms, the walls are very absorbent in order to eliminate these reflections. The sound seems to die down rapidly.

Anechoic rooms are especially suitable for making accurate acoustical measurements, such as source radiation patterns, microphone calibration, sound power emission of machines, headrelated transfer functions etc.

In this section we will use a speech signal recorded under anechoic conditions, that means, the only sound that reaches the microphone is the direct sound. This is not a real case but the direct sound has most energy of the total signal received.

An example will be use to explain the results when the source is a speech signal in an anechoic environment. In the next example the signal has a delay of 200 samples and it is recorded with a sample rate of 48000 Hz.

An analysis of the results obtained with the three different methods under consideration in this study (cross correlation, GCC and AED method) will be made.

CROSS CORRELATION (CC)

At first, the efficiency of the CC method when the source is a speech in anechoic conditions will be discussed.

The results are presented in the figure 4.7:

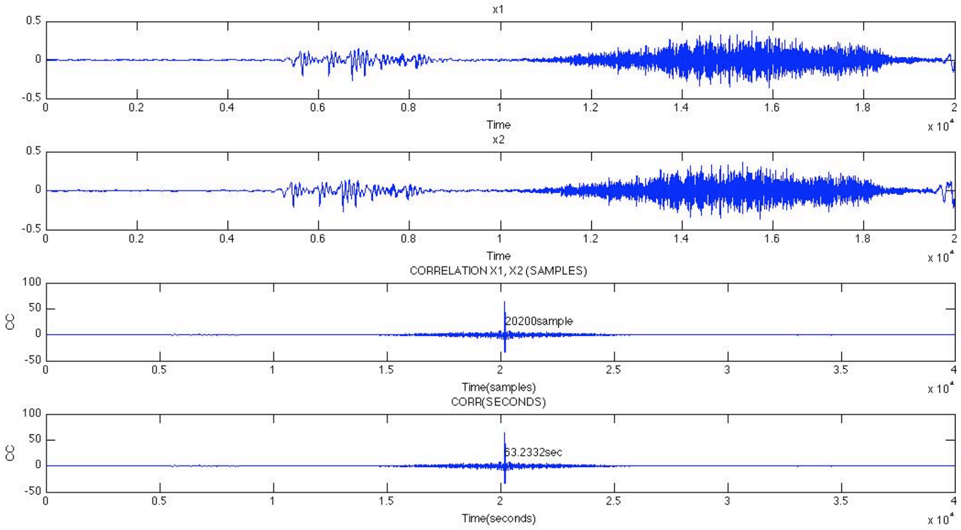

Fig. 4.7. Cross Correlation (delay=200)

In the example, the true time delay between the two sensors is equal to 200 (samples). The first two graphs represent the signals (x_1 and x_2) that we correlate in the third line. The x-corrdinate denotes the time in samples and y-coordinate denotes the resulted cross-correlation. It can be seen that the peak is in the sample 20200 and we can conclude that the peak position corresponds to the real delay (200) because if we make the correlation, we obtain a new signal (third line) that it is the double of the previous signal. If we have the peak in the sample 20000 (signal length/2), it means that the delay is zero so in this case the estimated delay is 200. The CC algorithm converges very fast to a good time delay estimate, it converges in less than 2 ms.

The following representation will be used in this text to present the results. In these figures cross-correlation values are calculated for a fixe frame-size. The frame is moved and delay is plotted over moved frames. The amplitude is encoded by color.

In this case, in all frames, the peak is located in the sample 200. The delay is constant throughout the entire signal.

From the results, one can see that for anechoic speech signals, the estimation of the delay using the cross correlation algorithm do not fail.

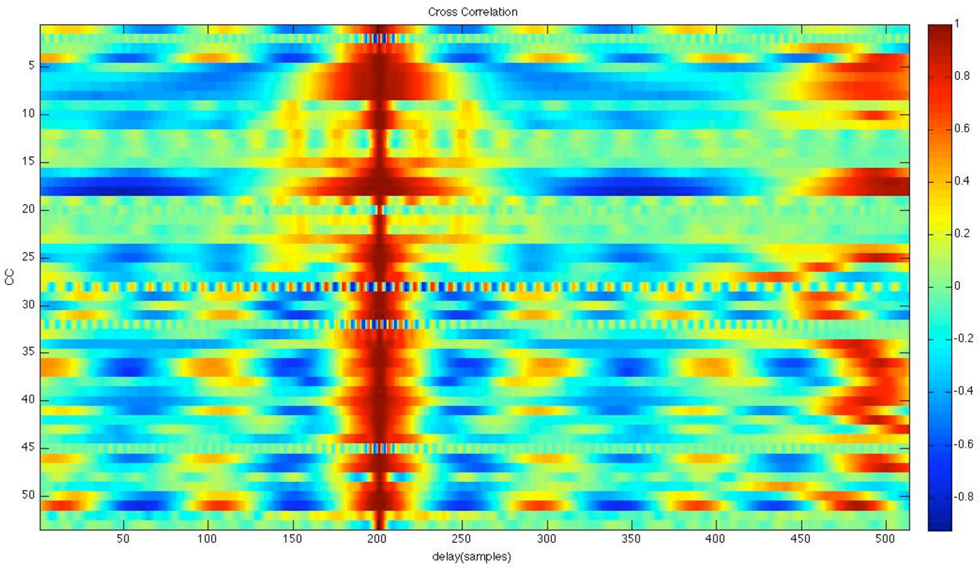

Fig. 4.8 Cross correlation with a delay of 200 samples

GENERAL CROSS CORRELATION

This example will be analyzed by all methods. The cross-correlation method was correct. Now, the results obtained using the GCC will be presented.

Using Matlab the results are shown below:

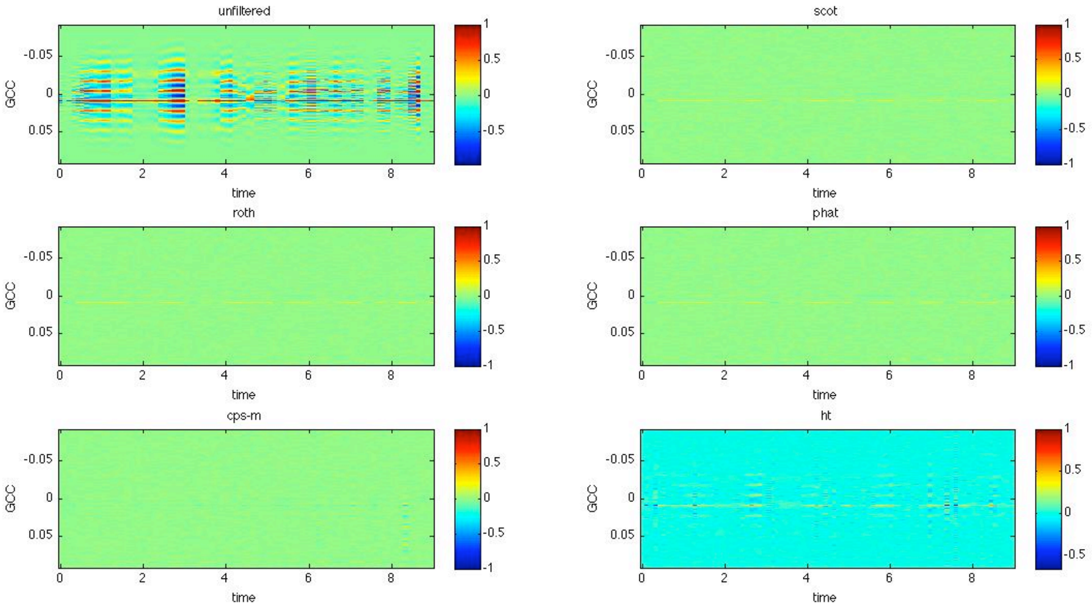

Fig. 4.9 Unfiltered, Scot, Roth, Phat, cps-m and ht GCC

Figure 4.10 shows the result of GCC method depending on the weighting function that is used. The x-coordinate denotes time and the y-coordinate denotes the GCC. It can be seen that in all cases the result is the same. In addition, the delay obtained is always constant.

Below, there is another graph that represents, as in the previous graph, the GCC results and it shows the estimated delay in samples.

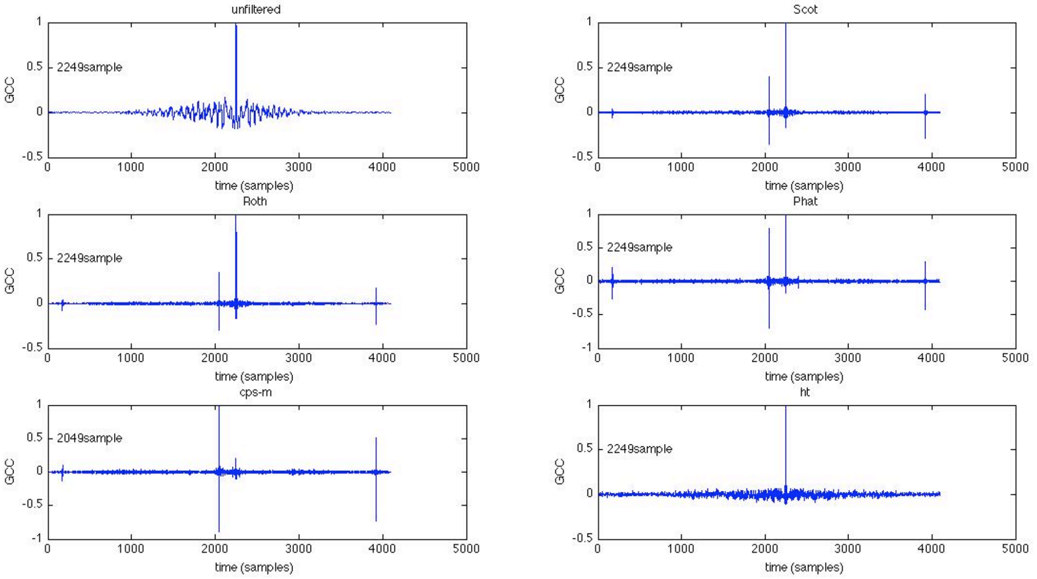

Fig. 4.10 Unfiltered, Scot, Roth, Phat, cps-m and ht GCC in samples

The peak is localizated in the same sample for all the algorithms. The real delay is 200 and we can see that the result obtained using this method is 201. When the correlation is calculated, the length of the signal is twice the length of the two previous signals. The delay is the difference from the length of the signal before being correlated, to the peak obtained by performing the correlation.

Thus, in this case, the length of the correlation between the two signals is 4096. The peak is shown in 2249, so the delay estimation has a value of 201.

From the results, one can conclude that this algorithm can adjust to the delay in only one sample. The results that we obtain with this method are also good, because the error is only a sample. If we represent the delay in time we can see that is the error or the delay equal to 0.0928 seconds (Figure 4.11).

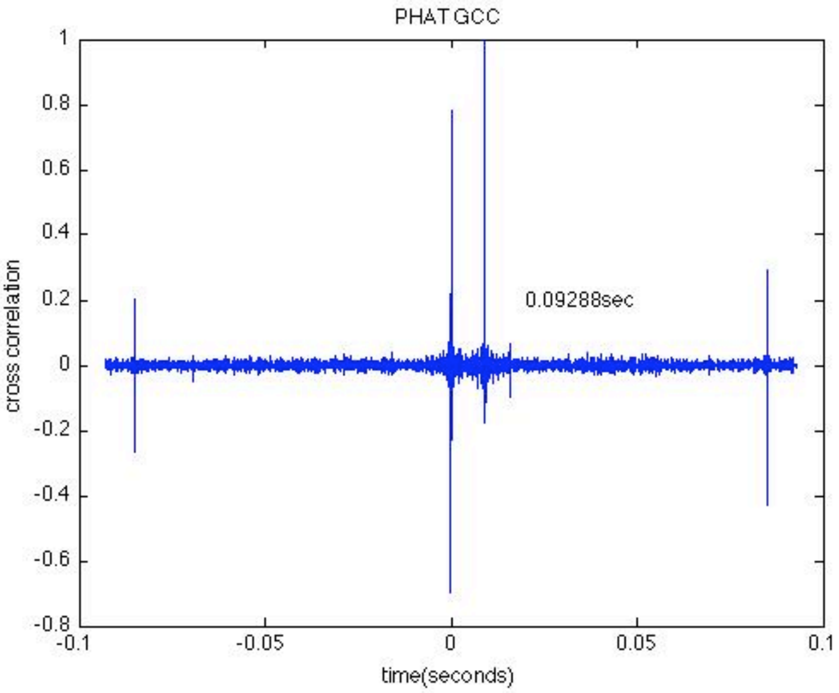

Fig. 4.11 TDE by Phat algorithm.

In theory, the GCC methods can give good results when the reverberation of the room is not very high, but when the reverberation becames important all of these techniques fail because they are based on a simple signal model that not represent reality [3].

ADAPTATIVE EIGENVALUE DECOMPOSITION:

So far, the results that have been obtained are good. The following result has been obtained using the AED algorithm:

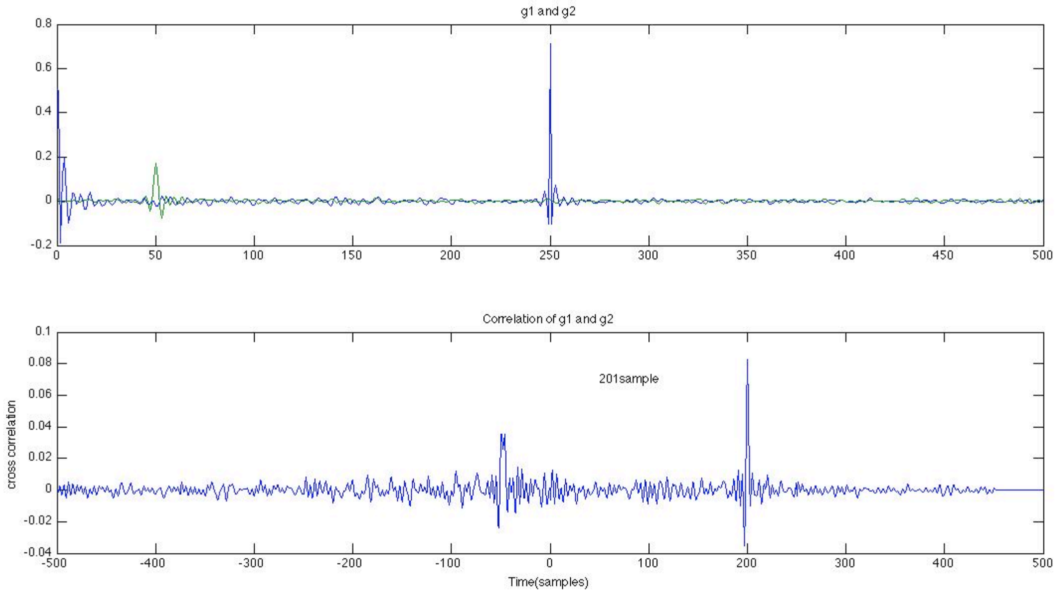

Fig. 4.12 AED method. First line represents the impulses responses g_1 and g_2 and second line represents the correlation about g_1 ang g_2. In the second line, the peak's position indicates the delay.

It was explained in theory that the relative delay is the difference between the indices corresponding to the two peaks of the impulses responses g_1 and g_2. The delay is the difference between the peaks of the impulse response g1 and the impulse response g_2.

The figure 4.12 shows each signal, first g_1 and g_2 and then, the correlation between these two signals. The x-coordinate denotes the time in samples and the y-coordinate denotes the correlation between g_1 and g_2. The peak indicates the value of the delay. It can be seen that de peak is in sample 201 and we can conclude that the estimated delay is very close to the real delay. Comparing the previous two algorithms with the AED algorithm, we did not see much improvement.

Next figure represents the comparison between the results obtained with the three different methods. In each case, it is represented the value of the estimated delay.

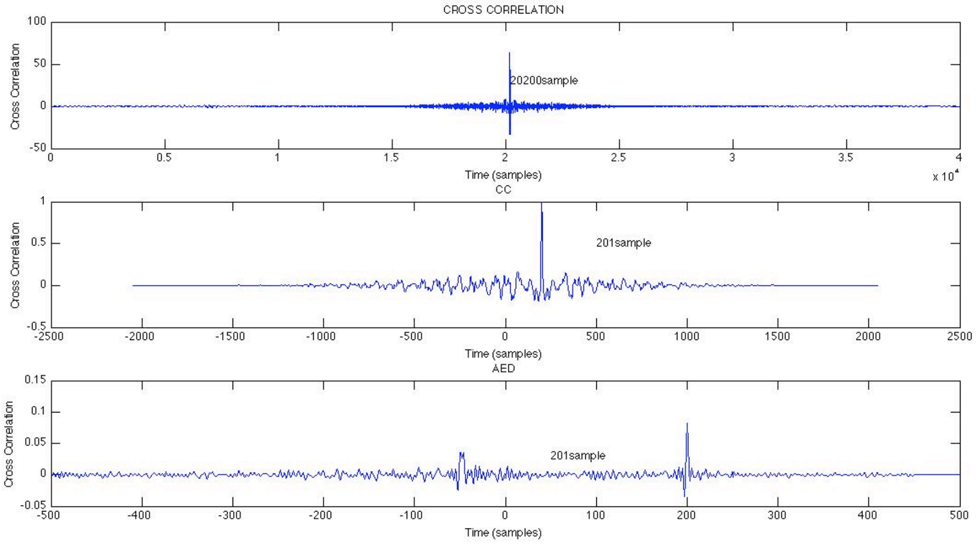

Fig. 4.13 CC, GCC and AED method comparison

It can be seen, as the error is very small in cases of GCC and AED method. The more accurate method is the cross-correlation method (CC). All the algorithms are close to the solution but the CC method is the most accurate.

4.2.2.2 ECHOIC ENVIRONMENT

In the open space, the sound emitted by a source will propagate away from the source and its intensity will decay quadratically.

In a room, the sound collide with obstacles (walls, surfaces, etc). Part of the sound will be reflected, while another part will be absorbed (dissipated as heat) for the material, and another part will be transmitted through it.

The signal received at the microphone is not only the direct sound as in the previous section. The received signal consists of the direct sound and the reflections due to walls and objects in the room.

The signal we are going to use in this section are composed of:

- Direct sound: in the line of sight, the direct sound is a peak corresponding to the shortest travel path
- Early reflections part
  - First reflection (usually the reflection from the ground)
  - Second and other reflections: more reflections still clearly distinguishable

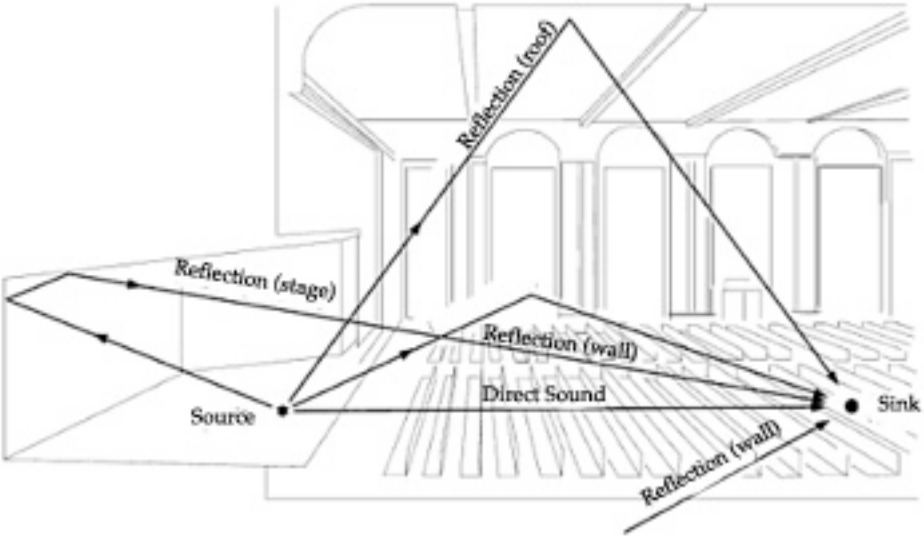

Fig. 4.14 Example of arrival of direct sound and early reflections to the receiver

The first reflections give us information about the position of the source and the reverberation give us information about the dimensions of the room. The reverberation, RT_{60}, is the time required for reflections of a direct sound to decay by 60 dB below the level of the direct sound.

As follows, simulation will be performed in an echoic environment. The estimation with the three methods will be described and we will see if there are differences between results in anechoic and echoic environments. In this example the true delay between the sensors is zero. Simulation was performed using a sampling rate of 48000 Hz and with a reflection coefficient of the walls of 0.3.

CROSS CORRELATION

As in the previous section, the same analysis but now with a speech signal in an echoic environment will be performed. Then, a comparison between anechoic and echoic environments will be done.

The first method is the cross correlation. The results obtained in this case can be seen in the following figure:

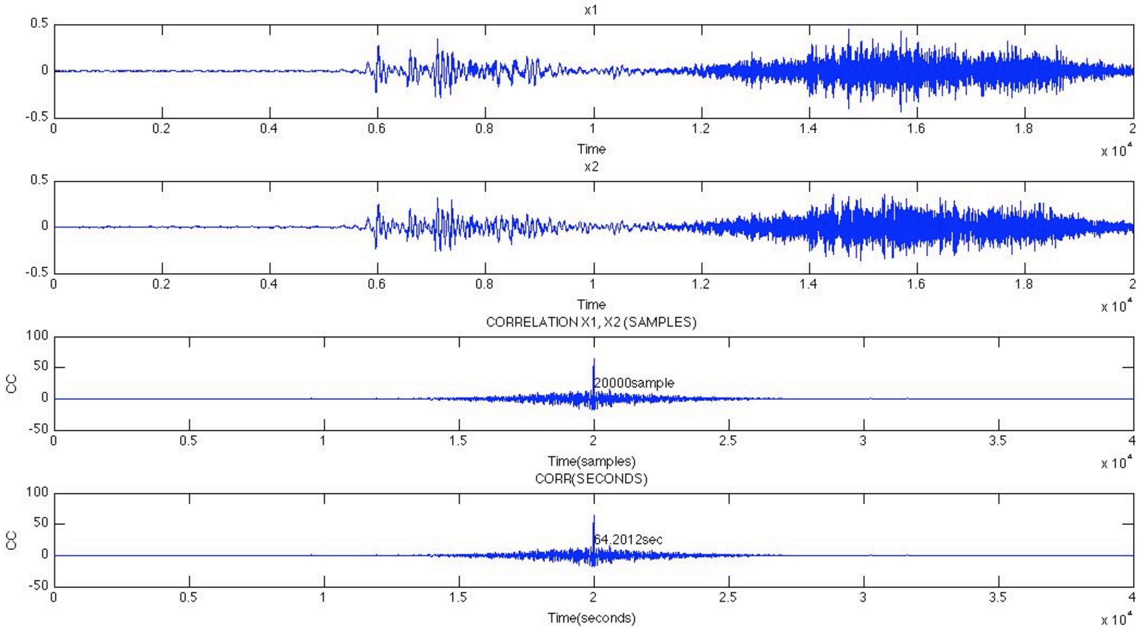

Fig. 4.15 Cross correlation. First and second lines are the signals captured by sensors. Third line represents the correlation between these two signals. The peak position indicates the delay.

In this example the true delay is zero. The CC results are shown in Fig.4.15 and it can be seen that this algorithm do not fail because the peak is located in sample 20000 (delay equal to 0).

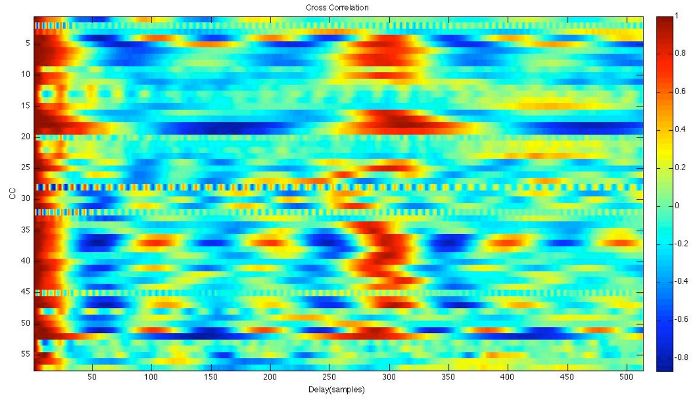

Fig. 4.16 Cross Correlation

It can be seen that the delay is constant and equal to 0 in the Fig.4.16.

GENERAL CROSS CORRELATION:

The results of the GCC algorithm are shown in the next figure:

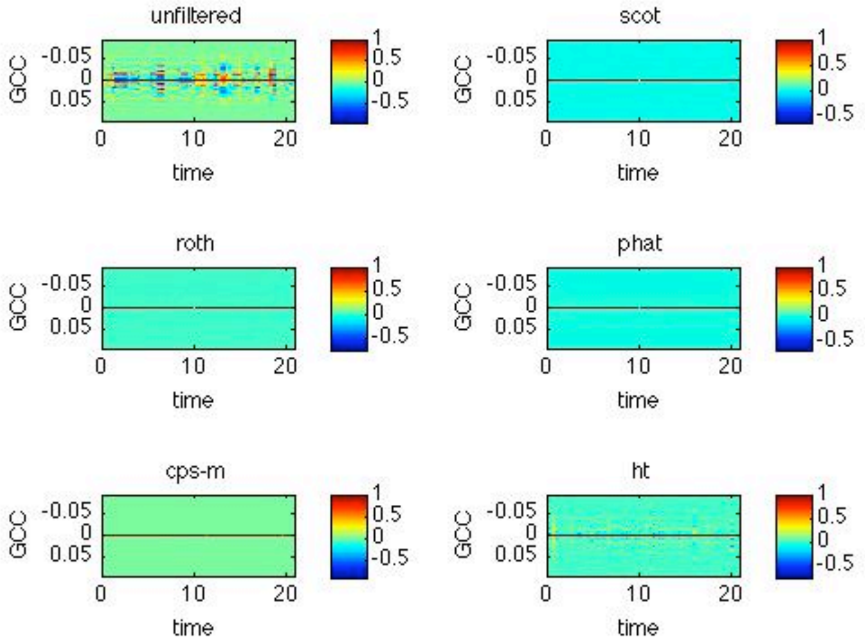

Fig. 4.17 Unfiltered, Scot, Roth, Phat, cps-, and ht GCC

This graph shows the results of the GCC algorithm depending on the weight function that is used. In all of them, there is a constant maximum located at zero. Comparing the CC algorithm with the AED algorithm, we can not see much improvement. We can see that in GCC method the peak is more clearly defined.

Next figure presents the Phat GCC algorithm, where it can be seen that the delay is constant and equal to 0.

.

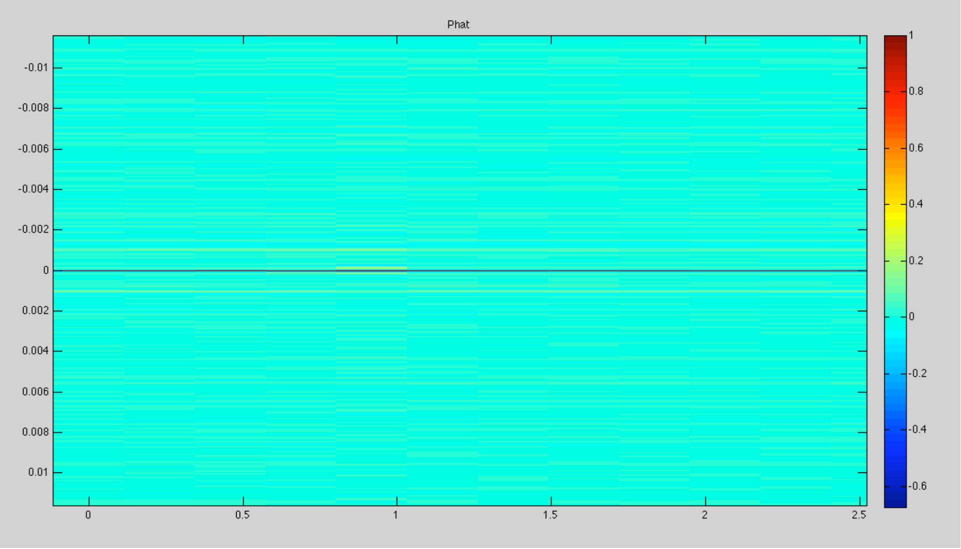

Fig. 4.18 Phat

Figure 4.19 shows the results in samples. Note that the peaks are better defined than that on figure 4.15 (CC method).

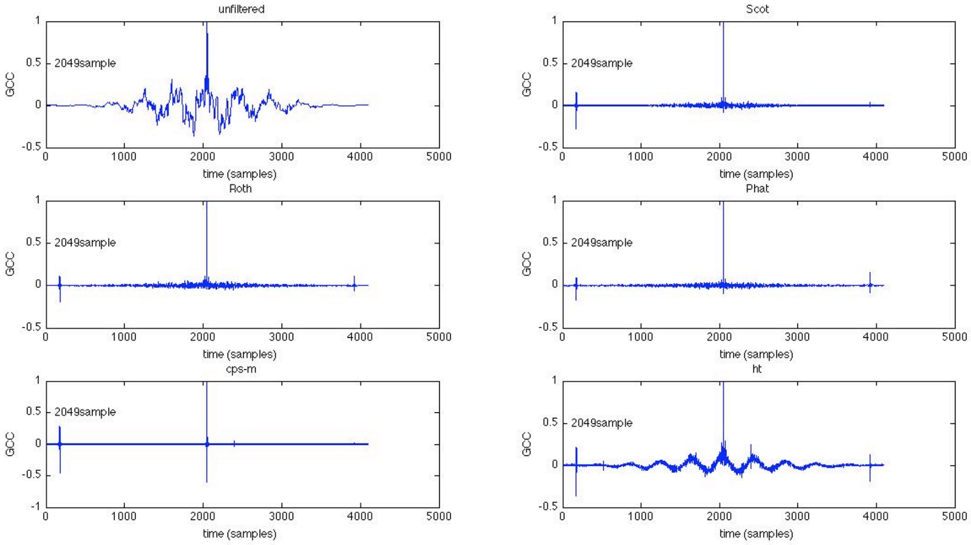

Fig. 4.19 GCC in samples

ADAPTIVE EIGENVALUE DECOMPOSITION

And finally, the results using the adaptative eigenvalue decomposition algorithm will be ilustrated. The estimated delay is shown in the following figure:

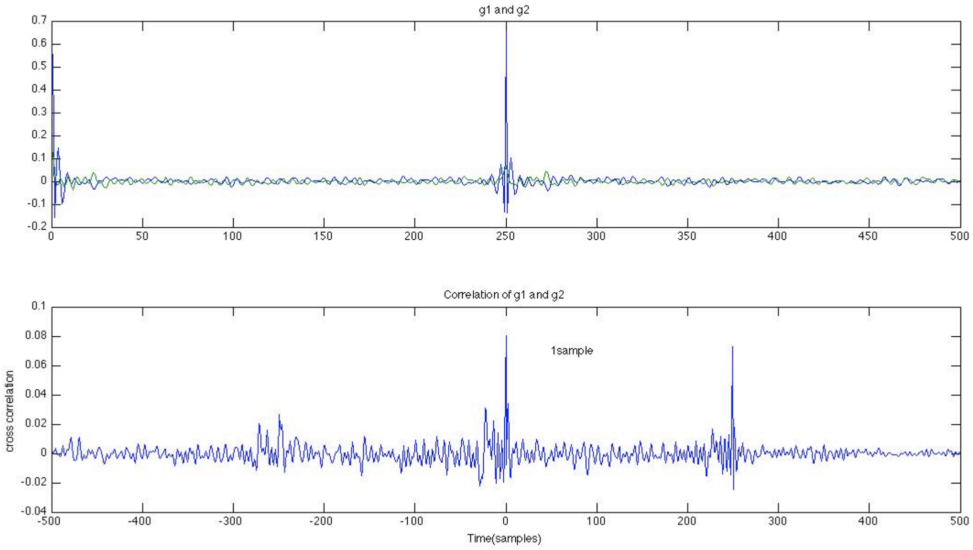

Fig. 4.20 AED algorithm

As in previous cases, the results we have obtained with this method are good due to the estimated delay is equal to 1 and the real delay is zero. The peak is good defined.

From the results, we can see that all the algorithms can adjust to the true delay in 1 sample. All the algorithms are very close to the solution. Besides, comparing the results between echoic and anechoic environment, we can not see significant differences.

Fig. 4.21 shows a comparative graph of the three methods and as in the previous section the results are good. From this analysis, it can be seen that the estimated delay is very close to the solution. The results show that if we have a low reverberation, we can have a good estimation with speech signals.

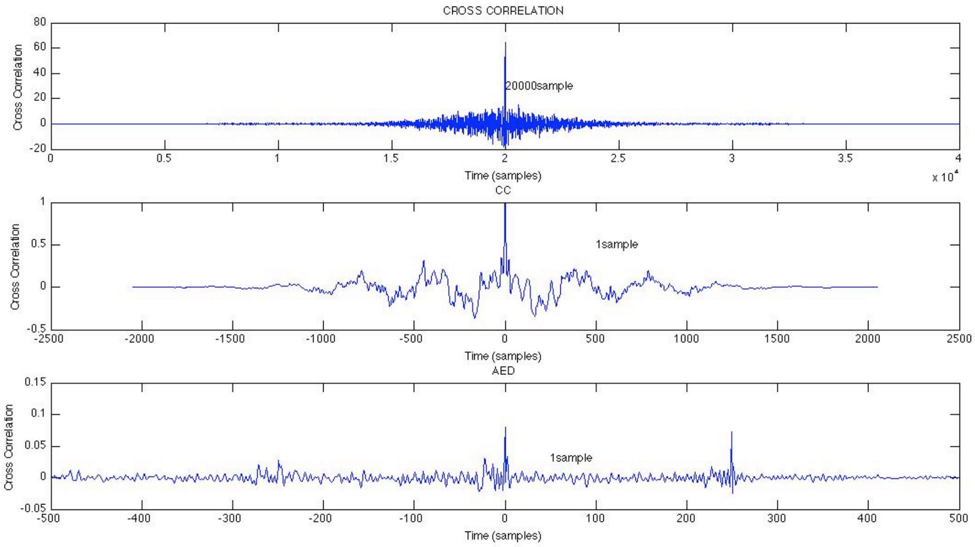

Fig. 4.21 CC, GCC and AED methods comparison

4.2.2.3 ROOM IMPULSE RESPONSES

INTRODUCCTION

Room impulse responses (abbreviated as RIRs) have the following components:

- Propagation delay: the length in time the sound travels from the source to the listener
- Direct sound: in the line of sight, the direct sound is a peak corresponding to the shortest travel path
- Early reflections part
  - First reflections (usually the reflection from the ground)
  - Second and other reflections: more reflections still clearly distinguishable
- Reverberation Tail part: this is the stochastic part of the reverberation where so many reflections are present that they cannot be separated any more.

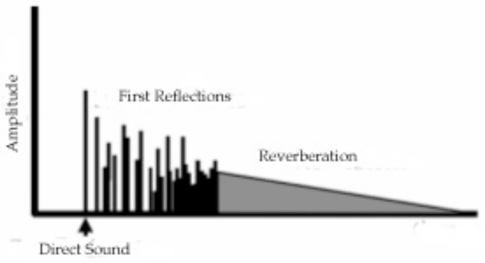

Fig. 4.22. Room impulse response

An impulse response is a transfer function between the input and the output of an LTI - linear, time-invariant-system, and contains all information about it. One of the most important usages of the room impulse responses is the ability to calculate the acoustical parameters. These are used for objective evaluation of the rooms.

In the open space, the sound emitted by a source will propagate away from the source and its intensity will decay quadratically. The sound level is reduced by 6 dB every time you double the distance.

In a room due to collide with obstacles (walls, surfaces, etc..,), part of the sound will be reflected, while another part will be absorbed (dissipated as heat) for the material, and another part will be transmitted through it (Figure 4.23).

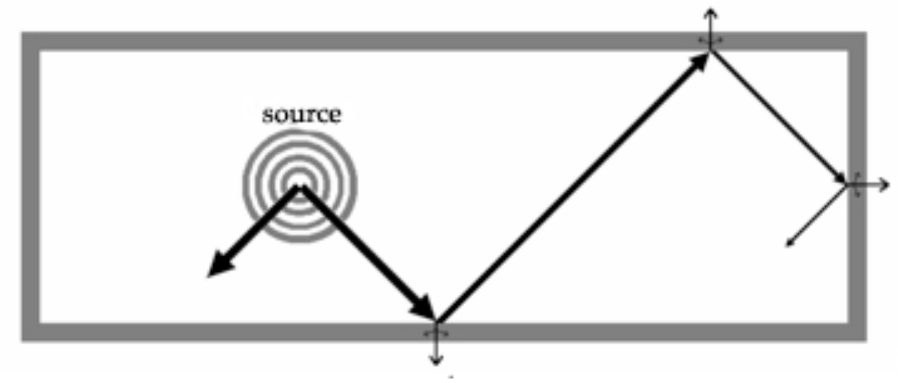

Fig. 4.23 Sound propagation in a room

The transfer function can be identified using a simple impulse as input. Thereafter, the charecteristics can be extracted easily.

The reverberation of a room can be simulated by the convolution of an input signal with the impulse response of the room. We will see the results when the impulse response of the room is considered. We will make the convolution between various impulse responses and one speech signal and we will see the effect in the time delay estimation.

For the example, we will use a speech signal ('germany4.wav') and the room impulse response when the room is an office.

When we make the convolution between the speech signal and the impulse response we obtained two signals called sensor 1 and sensor 2. Figure 4.24 shows these signals:

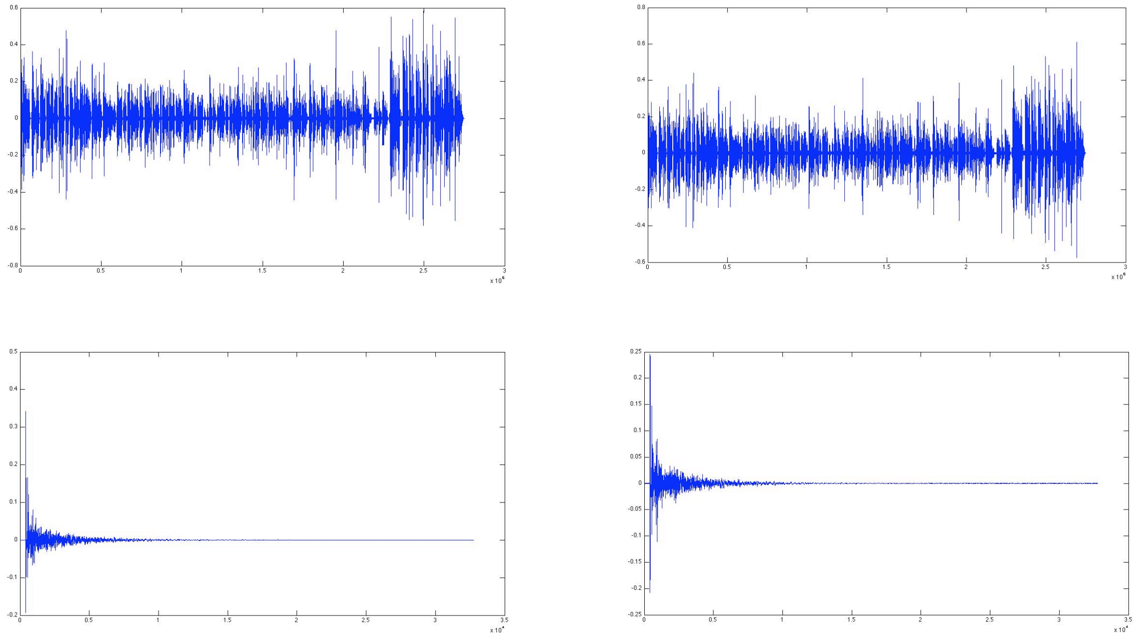

Fig. 4.24 First line: sensor 1 and sensor 2. The signals that have been convoluted with the room impulse response. Second line: room impulse response.

The delay between sensor 1 and sensor 2 is represented in Figure 4.25 The delay is near one sample. It can be seen in the following figures:

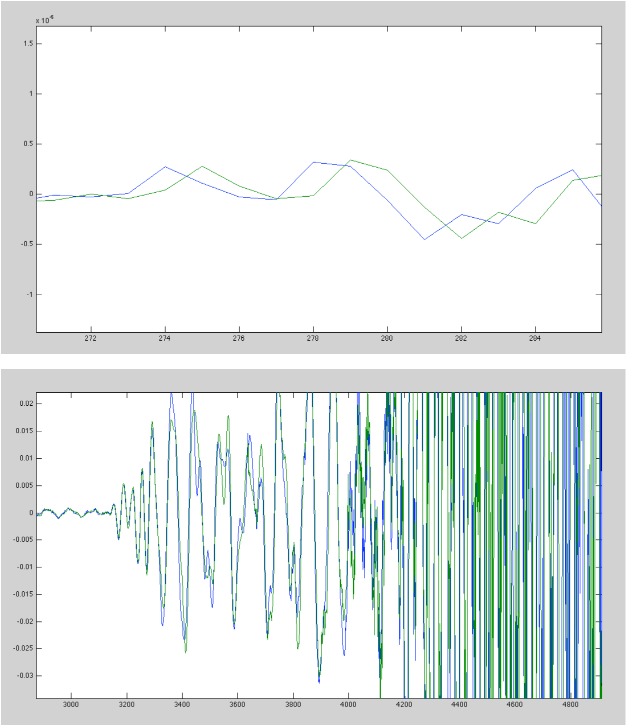

Fig. 4.25 Sensor 1 and sensor 2

The real delay a value near to zero. The effectiveness of delay estimation methods will be examined. The results obtained with each of the methods in this example will be presented comparing the estimate delay and real delay. Then, a comparison between the estimate results and real delay will be made.

CROSS CORRELATION:

In first time, the results obtained using the cross correlation algorithm is presented. In the next figure, it can seen that the peak corresponds to the real delay.

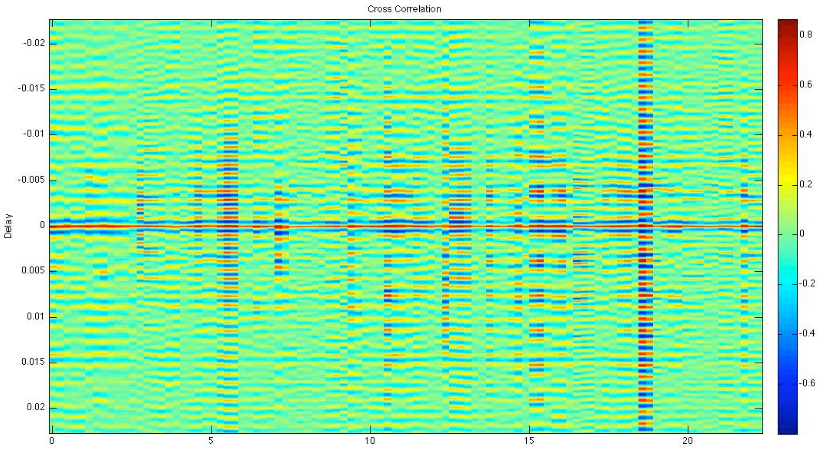

Fig. 4.26 Cross correlation

GENERAL CROSS CORRELATION:

The GCC results are presented in the figure 4.27 and 4.28

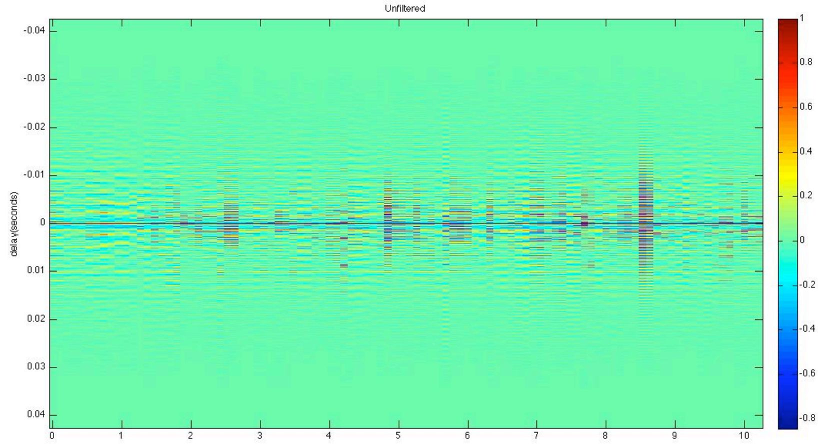

Fig. 4.27 GCC Unfiltered

As in CC method, in this case the estimate matches the actual delay. If we represent the results in terms of the weighting, all the algorithms can accurately identify the time delay.

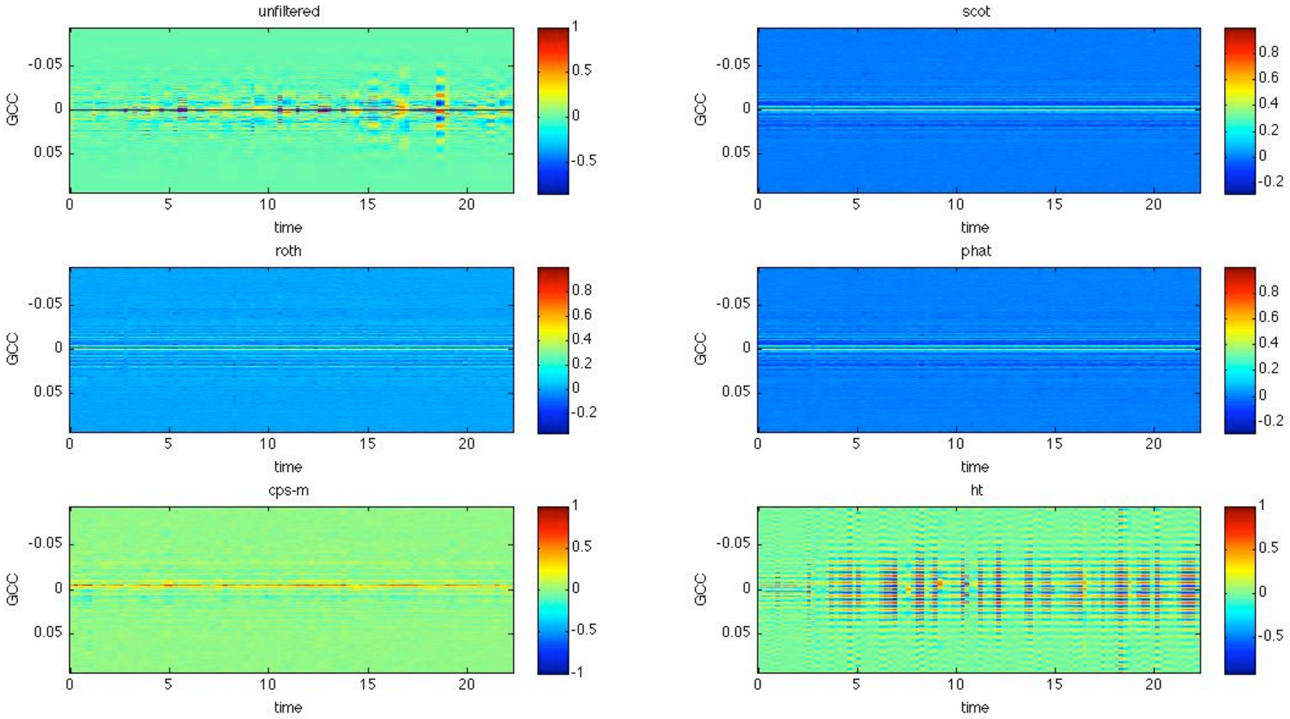

Fig. 4.28 Unfiltered, Scot, Roth, Phat, cps-m and ht GCC

In figure 4.28 it can be seen that the delay with all the methods is constant very close to 0. From the results, one can see that all the algorithms have a delay very close to the real delay. The error is less than 0,5 ms.

The results of GCC method in samples are represented in figure 4.29 :

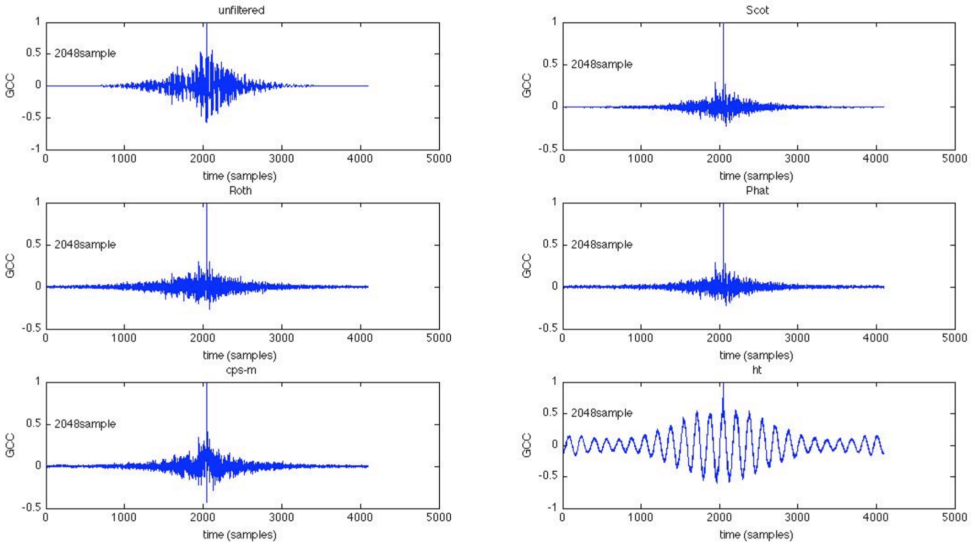

Fig. 4.29 Unfiltered, Scot, Roth, Phat, cps-m and ht GCC in samples

In figure 4.29 it can seen that the time delay is equal to 0 (samples).

ADAPTIVE EIGENVALUE DECOMPOSITION:

Finally , the result using the AED algorith, is represented in figure 4.30 :

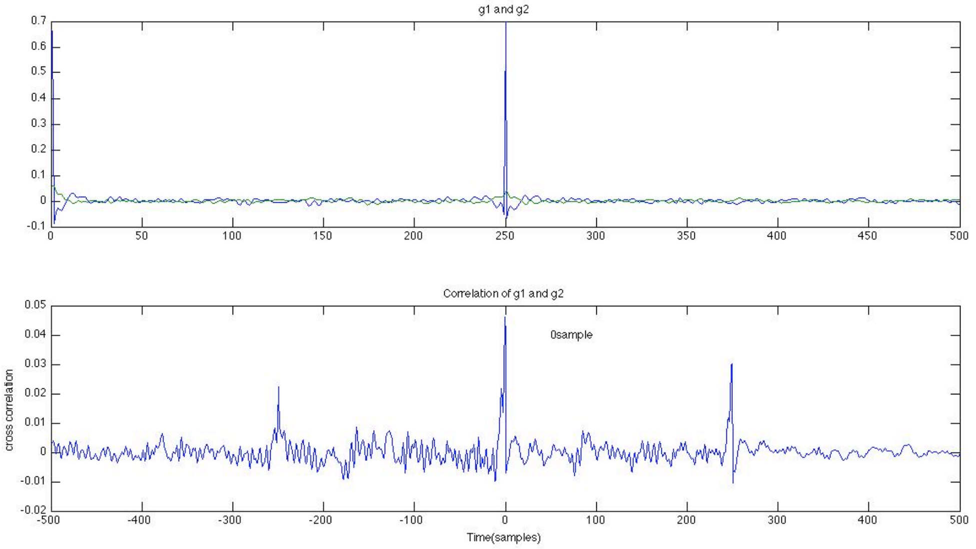

Fig. 4.30 AED Method. The peak is in zero

The peak position corresponds to the real delay .

Next figure represents the comparison between the results obtained for each method. It can be seen that the estimate delay corresponds to the real delay in the three cases.

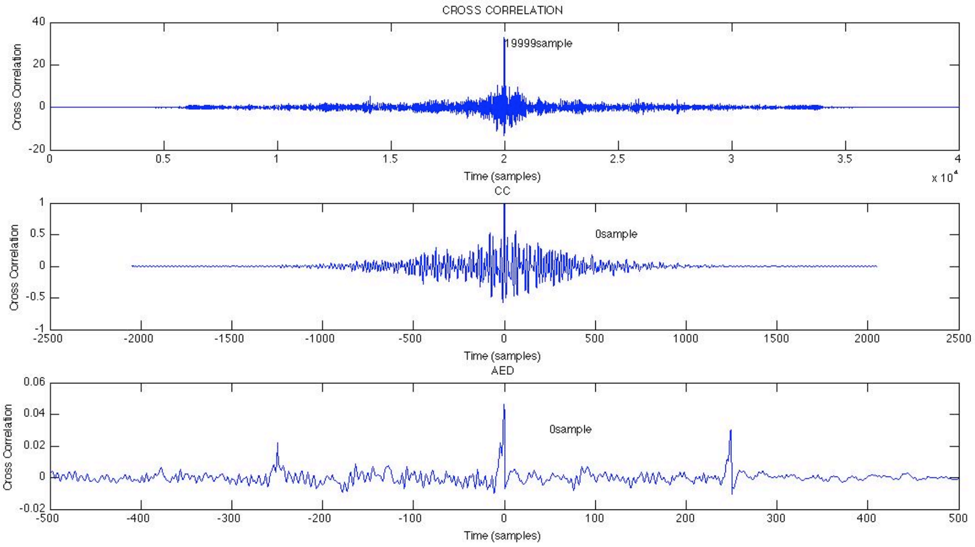

Fig. 4.31 CC, GCC and AED methods comparison

It is worth mentioning that in Figure 4.31, for CC algorithm, the peak is located at sample 19999, that is, the delay is -1. We have obtained a negative delay. We have assumed that sensor 1 arrives earlier than sensor 2, but for the results we can conlude that sensor 2 arrives earlier than sensor 1 due to the negative value has been obtained.

It can be seen from this example that GCC and AED methods perform better and are the more accurate than CC method.

Figure 4.32 shows TDE with differents room impulse responses. The source is the same speech signal. The first, second and third columns correspond, respectively, to a office room impulse, a booth room impluse and a lecture room impulse. The first, second and third lines correspond, respectively, to the TDE by the CC, GCC and AED algorithms. The true delay have a value aproximated of one sample in the three cases (Figure 4.35). It can be seen from this example that for office room impulse GCC and AED performs better than CC. With booth and lecture room impulses responses, GCC and AED preforms better but all the algorithms are very close to the solution. The peak using AED algorithm is clearer than the peaks that we obtained using the other two methods.

Figure 4.33 shows the room impulse responses. The first, second and third line correspond, respectively, to an office room impulse, booth room impulse and a lecture room impulse. It is important to know the room impulse because it give us information about the room. It can be seen, that the lecture room is more reverberant that the other two and the less reverberant room is the booth room.

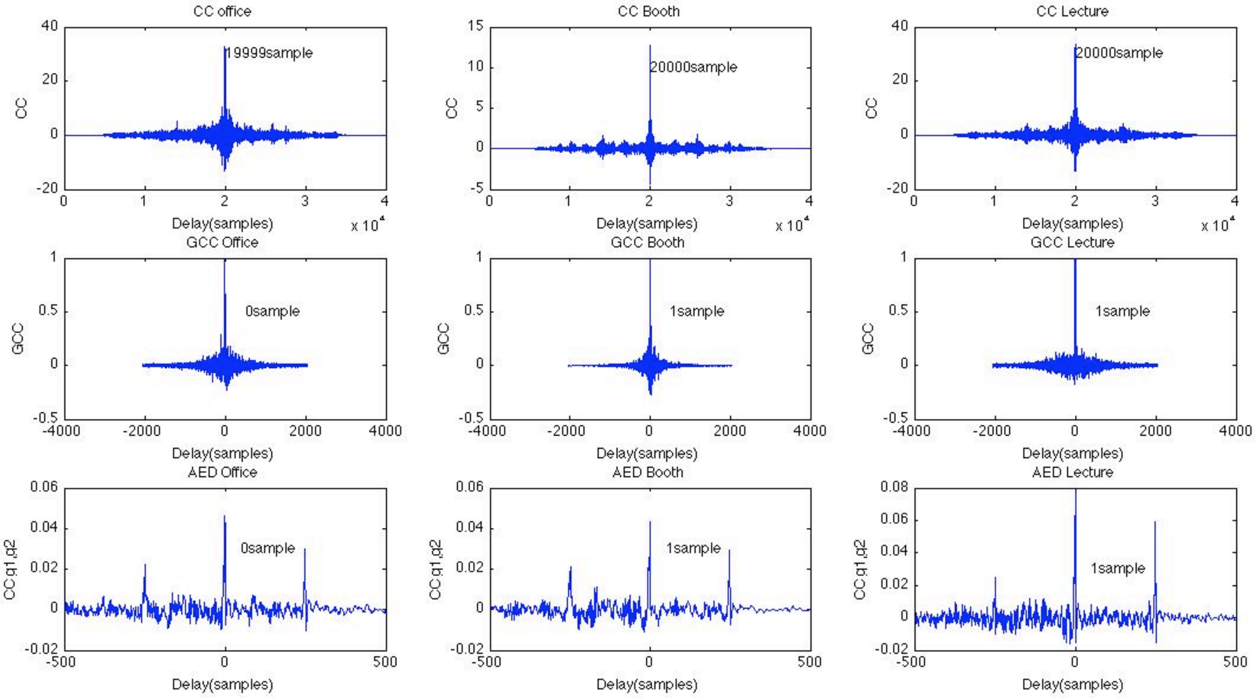

Fig. 4.32 TDE TDE with differents room impulse responses. The source is the same speech signal. The first, second and third columns correspond, respectively, to an office room impulse, a booth room impluse and a lecture room impulse. The first, second and third lines correspond, respectively, to the TDE by the CC, GCC and AED algorithms.

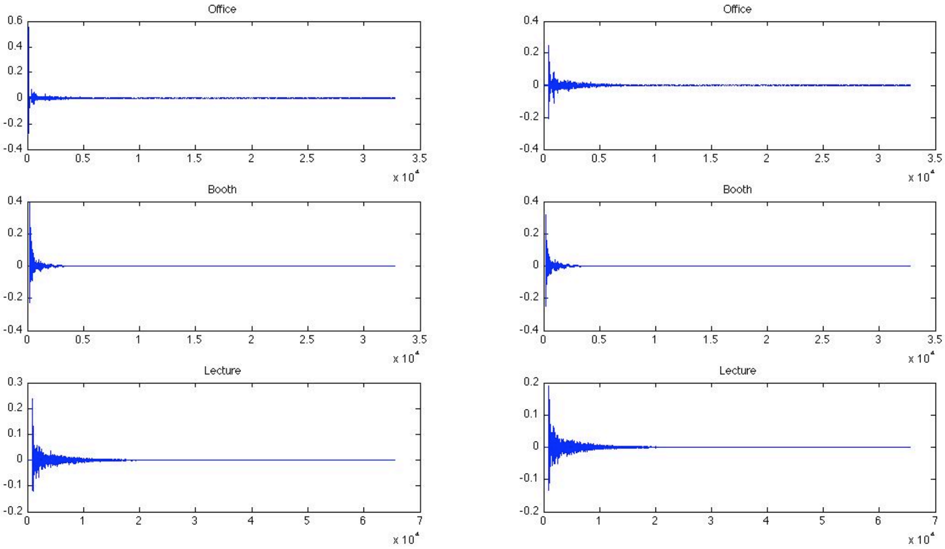

Fig. 4.33 Room impulse responses.

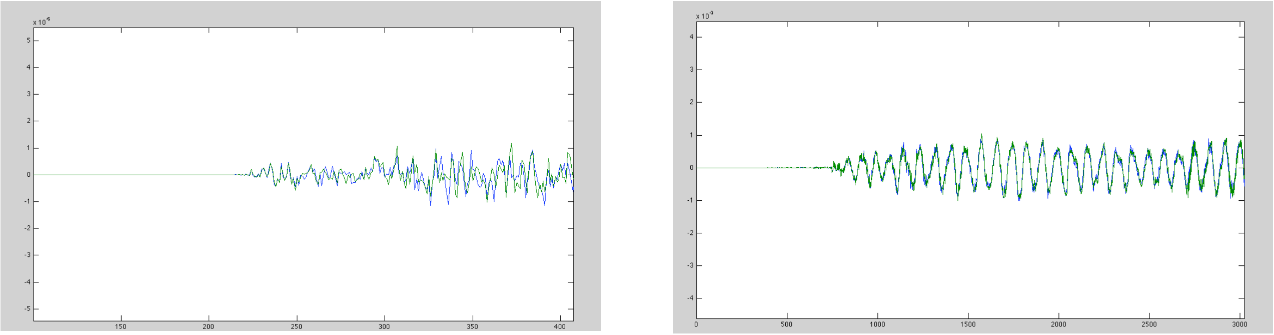

Fig. 4.34 Received signals in an booth room and in a lecture room respectively. True delay is near zero.

Now, the performance of the AED algorithm is compare to Phat and CC algorithm when we modify the reverberation. Two parameters that have relation with the reverberation are the reflection coefficient of the walls and the size of the room. In the examples before, the room was 10x10x3 metres, and in this example we are going to use a bigger room (20x40x3 metres). Four different reflection coefficients were selected: 0.1, 0.3, 0.6 and 0.8. The speech signal 'germany4.wav' was used. Two microphones and one source were simulated as we can see in the next figure:

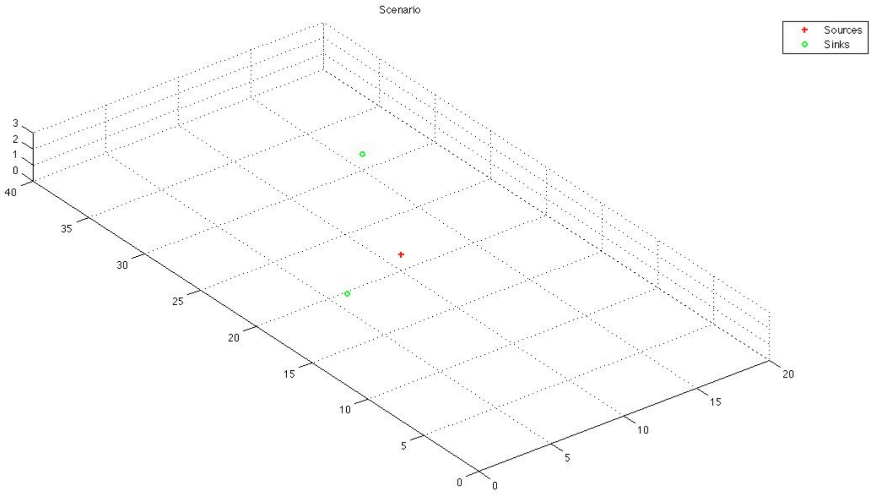

Fig. 4.35 Source and sinks localization

Figure 4.36 shows the obtained results. The source is a speech signal and the position is in the center. The first, second, third and fourth columns correspond, respectively, to a reflection coefficient of 0.1, 0.3, 0.6 and 0.8. The first, second and third lines correspond, respectively to CC, Phat and AED algorithm. The true delay is 283 samples and in each figure it is written the estimated delay.

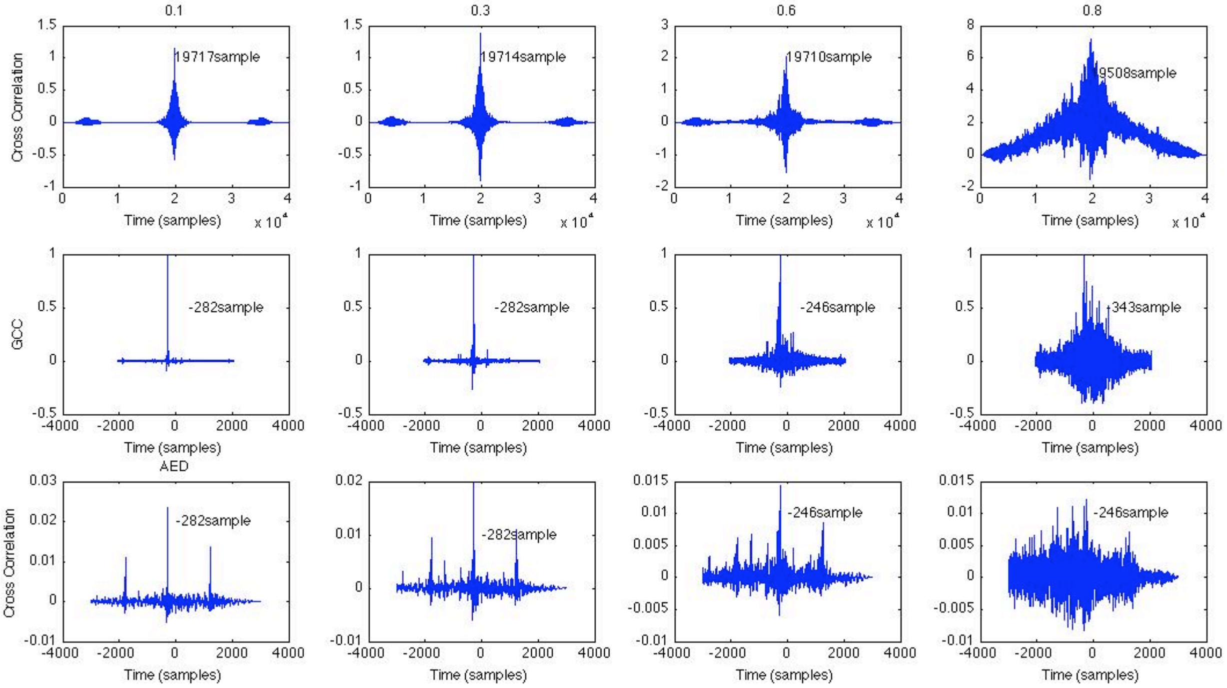

Fig. 4.36 Example varying the reflection coefficient of the room. The true delay is 283.

It can be seen from this example that the AED algorithm performs better and its more accurate when the reverberation is high. We have good results for low and moderately reverberation but all methods fail for high reverberation. The AED algorithm is the algorithm that have less error.

The next figure represents the error of the estimation with the different reflection coefficients. I can be seen that the error grow up when the reverberation is high. In this case all the methods fail but the AED method is the most accurate. When the reverberation is low, the estimation is good for all the methods.

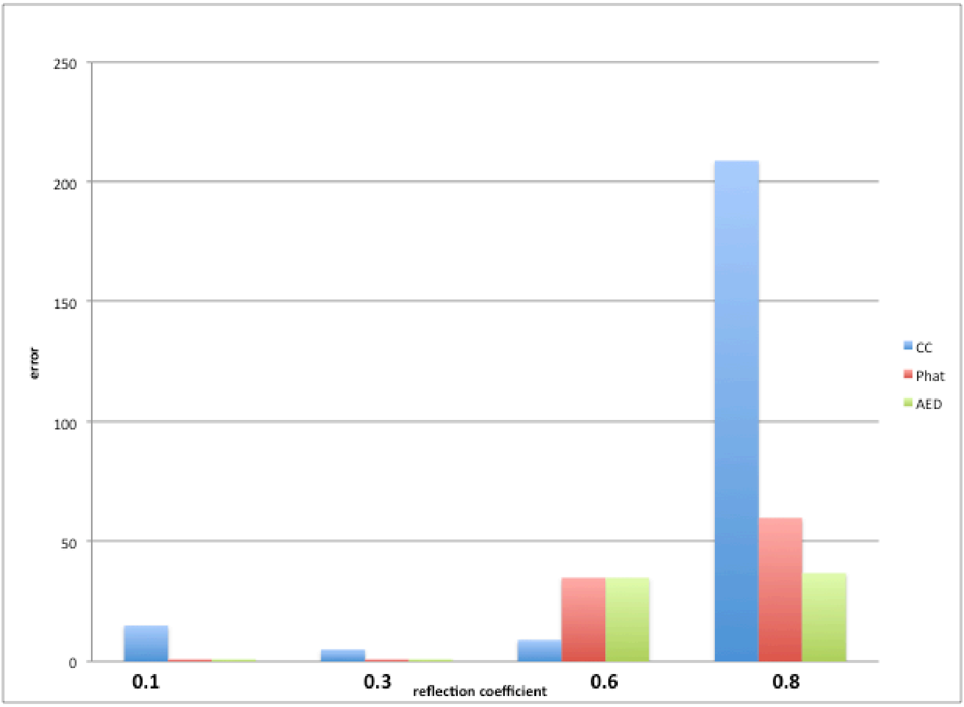

Fig. 4.37 Error of time delay estimation

		Error (samples)	
reflection	CC	Phat	AED
0.1	15	1	1
0.3	5	1	1
0.6	9	35	35
0.8	209	60	37

Table 4.2 Error of time delay estimation

5. CONCLUSIONS

In this paper, a comparison between three TDE methods has been presented. CC and GCC methods are based on an ideal model. Nevertheless, AED method first identifies the channel impulse responses from the source to the two sensors. This method focuses directly on the impulse responses between the source and the microphones in order to estimate the time-delay. Apparently, this algorithm takes fully the reverberation effect during time delay estimation into account [3].

Comparing CC algorithm with the GCC algorithm, we did not see much improvement. The role of the filter or weighting function in GCC method is to ensure a large sharp peak in the obtained cross-correlation thus ensuring a high time delay resolution [1]. In the conditions that we have used, the difference between these two methods is very small. CC and GCC methods are efficients and accurates in anechoic and in echoic environments. However we can see that the Peak is clearer and more definited in GCC method.

Comparing now CC and GCC algorithms with AED algorithm, we did not see much improvement either in a room with a low reverberation.

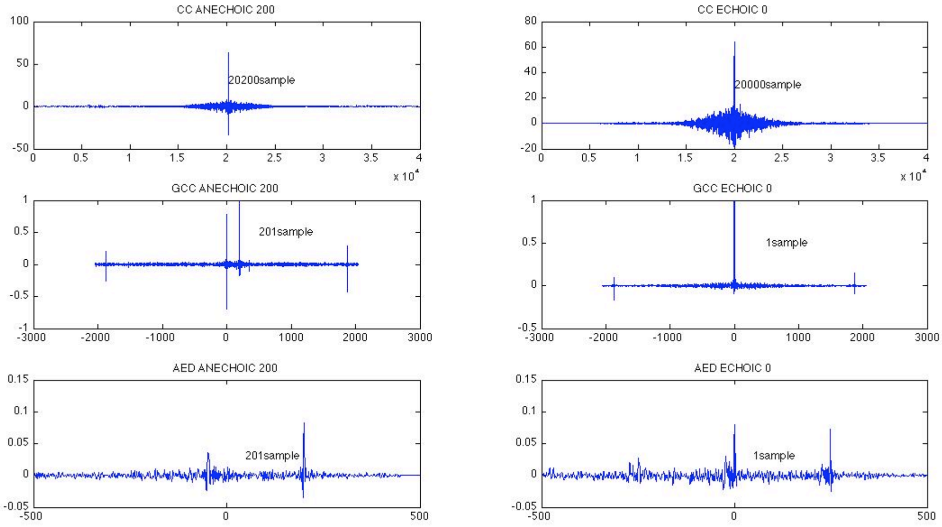

Fig. 5.1 TDE when the source is a speech signal. The first, second and third lines are, respectively, the TDE by the CC, GCC an AED algorithms. The first and second columns correspond, respectively, to a anechoic environment (delay 200) and a echoic environment(delay 0).

In figure 5.1 we can see that the differences between anechoic and echoic environments are almost nonexistent. However we can see that the peaks are clearer

If we compare now the results we have obtained changing the room impulse responses we can see that with all the room impulse responses, all the algorithms are close to the solution but the GCC and AED algorithm are the most accurate.

We have seen that the methods fail when we increase the reverberation of the room. When we used a bigger room with a low reflection coefficient the results are good for all the methods but when we make the estimation using a high reflection coefficient all the methods fail. However, the most accurate is the AED method (Fig. 4.38).

References:

[1] G. C. Carter: "Coherence and time delay estimation: an applied tutorial for research, development, test, and evaluation engineers", Piscataway, NJ: IEEE Press, 1993

[2] C. H. Knapp and C. G. Carter: "The generalized correlation method for estimation of time delay", IEEE Trans, Acoust, Speech, Signal Processing, vol. ASSP-21, pp. 320-327, August 1976

[3] J. Benesty. Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. 29 September 1999.

[4] Jingdong Chen, Jacob Benesty, and Yiteng (Arden) Time Delay Estimation in Room Acoustic Environments: An Overview Bell Laboratories, Lucent Technologies, Murray, 26 September 2005

[5] S. Björklund. Experimental Evaluation of some Cross Correlation Methods for Time Delay Estimation in Linear Systems. Technical report, Linkopings University, 15 April 2003.

[6] K. Astrom and T. Hagglund. PID Control lers: Theory, Design and Tunning. Instrument Society of America, 1995.

[7] L. Ljung. Identification for control: Simple process models. In Proceedings of the 41st IEEE Conference on Desicion and Control, December 2002.

[8] William D. Stanley. Digital signal processing. Reston Publishing Company, INC. 1975.

[9] H. Wang and P. Chu: "Voice Source Localization For Automatic Camera Pointing System In Videoconferencing", IEEE Acoustics, Speech, and Signal Processing, vol.1 pp.187-190, April 1997

[10] Z. CH. Liang, X. ZH. Liu and Y. T. Liu: "A modified time delay estimation algorithm based on higher order statistics for signal detection problems", IEEE Signal Processing, vol.1, pp.255-258, Aug 2002

[11] Md. Khademul Islam Molla, Keikichi Hirose and Nobuaki Minematsu: "Robust Determination of Periodic Correlation of Speech Signals using Empirical Mode Decomposition and Higher-Order Spectra"

[12] O. L. Frost: "An algorithm for linearly constrained adaptive array processing" Proc. IEEE 60, 926-935 (1972)