Localization of Reflections in Auditoriums using Time Delay Estimation

5168 (T - 2)

Van Lancker Eric
EPFL DE LEMA, Lausanne, Switzerland

Presented at
the 108th Convention
2000 February 19-22
Paris, France

AES

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd St., New York, New York 10165-2520, USA.

AN AUDIO ENGINEERING SOCIETY PREPRINT

Localization of Reflections in Auditoriums using Time Delay Estimation

Eric Van Lancker
Swiss Federal Institute of Technology of Lausanne
Ecublens CH-1015 Lausanne, Switzerland
eric.vanlancker@epfl.ch

Abstract

The impulse response or echogram of a hall enables its acoustic properties to be assessed. The system being studied adds the spatial dimension to this time-energy representation. The method is based on the time delay estimation between microphones of a cubic array with 25 cm edges. The paper presents the results of tests carried out in the Stravinsky Auditorium (Montreux).

Introduction

In the field of concert hall or auditorium acoustics, practitioners are led to make a judgement on the listening comfort for the audience. The impulse response or echogram represents the amplitudes and arrival times of the sound rays. The time distribution of the direct ray, the indirect rays and the diffuse sound field characterise the qualities of a hall. Unfortunately, no information on the spatial distribution of these contributions is available. The system being presented makes up for this lacuna by localizing the direct and indirect rays. The method presented in section 2 is based on Time Delay Estimation (TDE) between microphones of a cubic shaped array. The processing tools used draw on the Cross-Power Spectrum Phase (CPSP). Once the components of the various contributions have been calculated, a sorting algorithm recombines the cross-correlation peaks in order to reconstruct the wave vectors. This stage, optimised from the point of view of computation time, introduces selection criteria in relation to the costs and to constraints linked to the propagation and reception conditions (geometry, antenna size). Section 3 explains how echogram (temporal module) and localization (spatial module) interconnect to provide the spatial echogram.

The performance of the system is expressed in terms of localization precision, probability of detection and in terms of the number of contributions detected.

1 Propagation time and wave vector

In a multipath environment, such as a concert hall, the output of an array transducer, $x_m(t)$, associated with a controlled source signal, $s(t)$, can be expressed as:

$$x_m(t) = \sum_{i=0}^{k} \alpha_i s(t - \tau_i) + n(t) \tag{1}$$

where k is the number of reflections considered, α_i, the attenuation of the i^{th} reflection, τ_i is the corresponding path delay, and $n(t)$ is the ambient noise. The direct path is associated with $i = 0$. The time delay between transducers δ is equal to the difference of time propagation. For a microphone pair 1-2 (Fig 1):

$$\delta_{12} = (\tau_1 - \tau_2) \tag{2}$$

Each delay corresponds to the wave vector projection on the selected sensors-pair. Since M transducers distributed in 3-D space yields $M - 1$ independent relative delays, four microphones are sufficient to provide a Direction of Arrival (DOA) estimation without ambiguities. The wave vector can be derived from the delay vector $\vec{\tau}$, the relative microphone positions matrix D, and sound velocity c :

$$\vec{n} = c \cdot D^{-1} \cdot \vec{\delta} \tag{3}$$

To minimize the estimation error, it is profitable to increase the number of sensors. In this case, the system becomes overspecified and the inversion of the non-square matrix D requires least-squares techniques as the the Singular Value Decomposition (SVD) method.

2 Antenna

The antenna is a cubic-shaped array with 25 cm edges made up of 8 omnidirectional electret microphones situated at the apexes (Fig 2). The sensors in fact belong to a sphere and with such a geometry the system can operate without any a priori information on the source location.

This cubic-shaped array enables us to calculate simultaneously each of the wave vector components of four TDE, corresponding to the four parallel edges of the cube , and theorically increase the SNR about 6dB. Furthermore, the algorithm checks that the plane wave hypothesis is verified. This coherent processing technique has demonstrated a good behavior in terms of detection performance [1]. The threshold SNR at which the correlator performance deviates from the Cramer-Rao Lower Bound (CRLB) is rejected lower. The detection performs better in low SNR conditions (e.g. reverberant environment).

The captured signals are post-processed on a PC. The acquisition chain made up of the array, the preamplifiers and A/D converters has been carefully designed to avoid any phase differences between channels.

3 Active Localization

3.1 Transmitted signal

Quazi [2] calculated the Cramer-Rao lower bound for the variance of the TDE in a case of an active system. The CRLB appears to be inversely proportional to the bandwidth cubed. This optimal bound, calculated for baseband signal, degrades when a bandpass signal is considered. Actually, low frequencies, i.e. wavelengths more than two times greater than antenna aperture, allow true peaks to be found without any ambiguities and high frequencies sharpen the peaks and give accurate estimations. Both large-band Chirp and MLS sequences were tested. The results present similar performances as long as their bandwidths are equal. At the opposite of the Chirp, the length of which is fixed for the experiment, the MLS is more flexible in regard to the segmentation and averaging steps.

3.2 Time Delay Estimation

The calculation of the $M-1$ relative delays in 3 is performed by TDE techniques. Classified in the Generalized Cross Correlator, the Phase correlator (also called CPSP method) emphasizes the phase difference information since the cross-correlation is normalised by its magnitude. This corresponds in the Fourier Domain to the Cross Power Spectrum normalised by the Spectral Density. A modified version of the CPSP Time Delay Estimation is given by [4]:

$$\widehat{C_{12}} = \mathcal{F}^{-1}\left[\frac{X_1(\omega)X_2(\omega)^*}{(|X_1(\omega)||X_2(\omega)|)^\rho}\right] \qquad 0 \leq \rho \leq 1 \tag{4}$$

where $X_m(\omega)$ is the Fourier transform of the $x_m(t)$, $\mathcal{F}^{-1}$ represents the inverse Fourier transform, and * denotes the complex conjugate. Setting ρ to zero produces the unnormalized cross-correlation, while setting it to one produces the CPSP. Optimally set by experimentation, ρ mainly varies with room caracteristics (e.g. RT60) and noise level. The correlator resolution is then improved through interpolation. The three-point method, performing a parabolic interpolation around the peaks, is a very low cost (CPU) solution, as opposed to the SINC interpolation which introduces much more calculations but denotes better performances. A truncated SINC appears to be a good compromise.

3.3 Detection

After a peak-piking step expressed as:

$$\delta_{12} = \tau : \max_{\tau} \quad \widehat{C_{12}}(\tau) \tag{5}$$

a sorting algorithm deals with delays combination to reconstruct the different contributions (direct component and reflections). The best combinations are determined according to a selection criteria including a weighted constraint on the consistence of the delays, the norm of the wave vectors and the energy of the peaks. The detection step can be enhanced with averaging of the CPSP over several successive time segments (non-coherent process) and tracking algorithms.

4 Spatial Echogram

In parallel to the localization module, the classical echogram is calculated. A peak-piking algorithm selects the peaks with most energy for each response. The peak sets for one sensors-pair are mixed in order to find relative delays verifiing the physical travel time assumption. As shown in (Fig 3) for the sensors-pair $1-2$, the algorithm deals with time differences of propagation sets issued from the echogram module and relative delays sets issued from the localization module. The combination of "Echogram" sets Δ and "Localization" sets δ is obtained by minimizing the Euclidian distance criterium expressed by:

$$(I,J) = \min_{i,j} \sum_{n=1}^{N} (\Delta^{i}_{n_1 n_2} - \delta^{j}_{n_1 n_2})^2 \quad i,j = 1,\ldots,k \tag{6}$$

where N is the number of sensor-pairs considered (2 typically), $\Delta^{i}_{n_1 n_2}$ is the relative delay between sensors 1 and 2 of the sensor-pair n issued from the Echogram module.

5 Experimentation

The test presented took place in the Stravinsky Auditorium in Montreux. The goal was to justify the presence of hall sound reflectors for the orchestra (Fig 4) by localizing early reflections (first 30 ms) coming from the reflectors. The trial involved a source piloted by PC (signal generation) and the cubic-shaped array. The source (5" loudspeaker) is situated backstage on a chair $50cm$ high and the antenna at the conductor's location $9m$ away from the source and $1m$ up from the floor. The antenna is facing the source in such a way that the direct contribution is expected at roughly Azimuth 0°. The source - antenna pair is aligned to the central reflector below it. Two extra vertical panels, disposed between the source and the antenna allow localization attemps with no direct contribution, without disturbing other reflected contributions.

The first test session, without vertical panels, allowed validation and localization precision estimation to be made. A classical Echogram based on Hadamar Transform (Fig 5), clearly shows the direct contribution followed by three major reflections. The localisation module results are shown in Figure 6. The total observation time was set to $5s$, the acquistion length of each segment was fixed to 1024 samples and ρ coefficent to 0.8. The results enables us to attribuate the:

- direct contribution to the source direction with an estimated precision of 1 or 2 degrees (Azimuth 1°, Elevation -4°)
- reflection $k = 1$ to the central reflector (Azimuth 1°, Elevation 47°)
- reflection $k = 2$ to both adjacents reflectors (Azimuth 329°+ 32°, Elevation 52°)
- reflection $k = 3$ to a second order reflection central reflector-floor (Azimuth 1°, Elevation -48°)
- reflection observed $1ms$ after the direct contribution, to a floor reflection (Azimuth 1°, Elevation -13°)

The second test session, with vertical panels hidding the source, enable more reflections to be localized. Contributions from lateral rear wall panels (Azimuth 34°, Elevation -4°), as well as third order reflections source - central reflector - floor - central reflector - antenna with (Azimuth 1°, Elevation 82°) could be identified.

6 Conclusion

The modified CPSP has demontrated its potential to deal with a reverberant environment. Actually, the ρ coeffficent allows an optimal adaptation of the localisation engine to the acoustical characteristic of the Auditorium. The spatial echogram, concatenation of the localisation module brought about by the antenna and the echogram module, corresponds to expectations. The direct contribution and five reflections have been identified. In particular, the localisation enable us to attribuate the second reflection peaks ($k = 2$) of the echogram to two separate contributions arriving simultaneously at the antenna due to the symetric disposition of the reflectors. Extra reflections have been observed by hidding the source behing panels, increasing the total observed reflections to seven contributions.

References

[1] Jan, E., Flanagan, J.L. 1996. Sound Source Localization in Reverberant Environments Using a Outlier Elimination Algorithm. In ICSLP 1996, Philadelphia, 3.

[2] Quazi, A.H. 1981. An Overview on Time Delay Estimate in Active and Passive Systems for Target Localization. In IASSP, 29(3), Pt 2 pp. 527-533.

[3] Scarbrough, K., Carter, G.C., Tremblay, R.J. 1984. Implications of Threshold Effects for Coherent and Incoherent Processing Technique of Time Delay Estimation. In ICASSP 1984, 1:15.1/1-4.

[4] Rabinkin, D.V., Renomeron, R.J., Dahl, A., French, J.C., Flanagan, J.L., Bianchi, M.H. 1986. A DSP Implementation of Source Location Using Microphone Arrays. In JASA, 99(4), Pt.2, pp. 2503

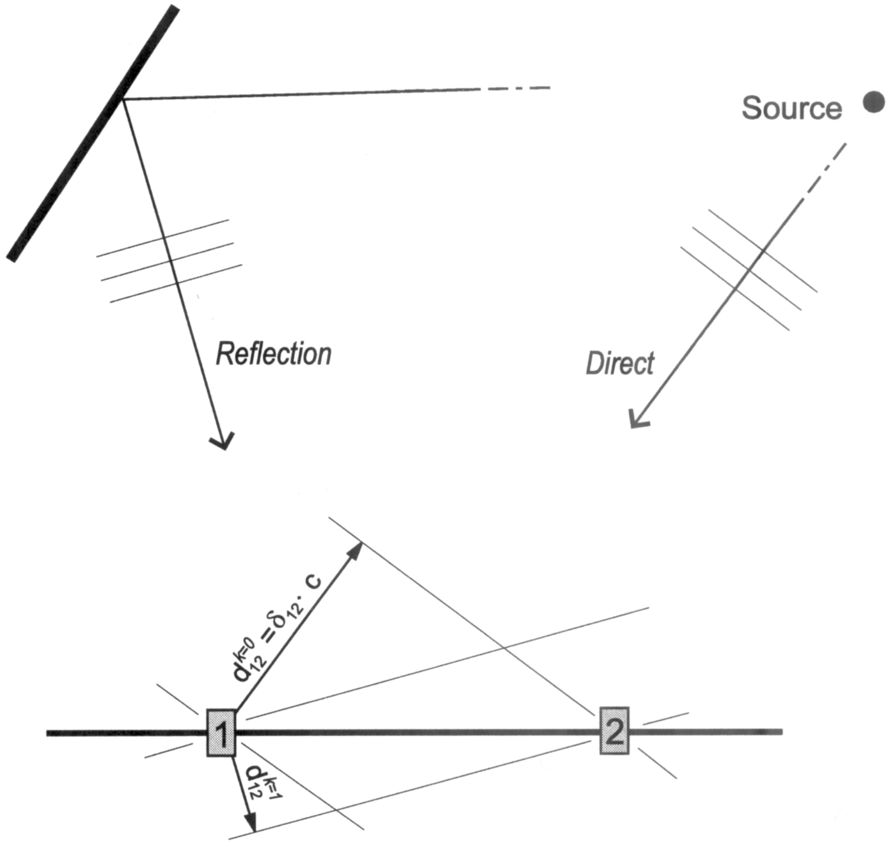

Figure 1: Direct wave, Reflections

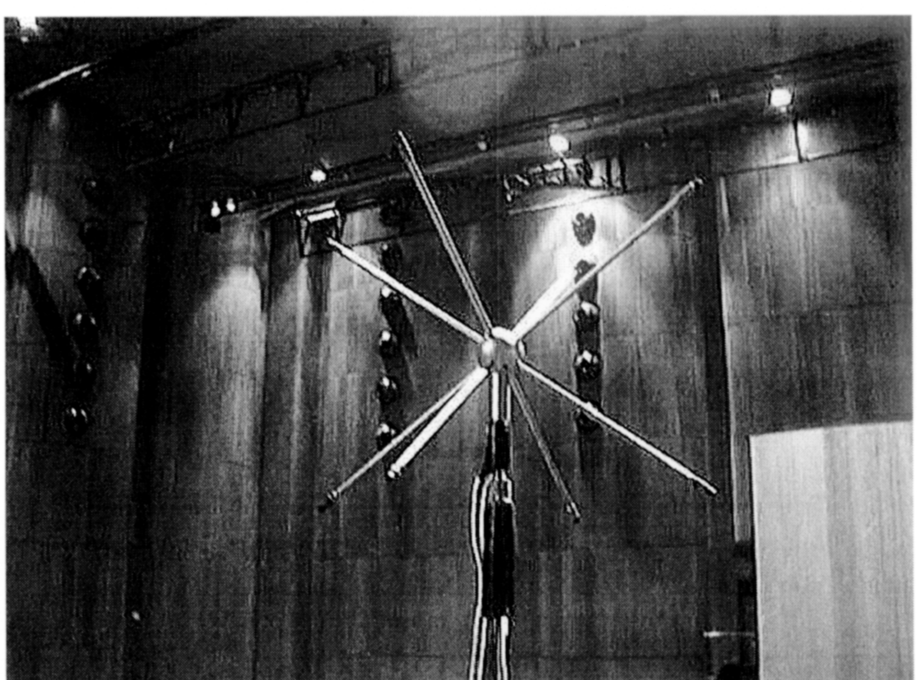

Figure 2: Antenna in Stravinski Auditorium

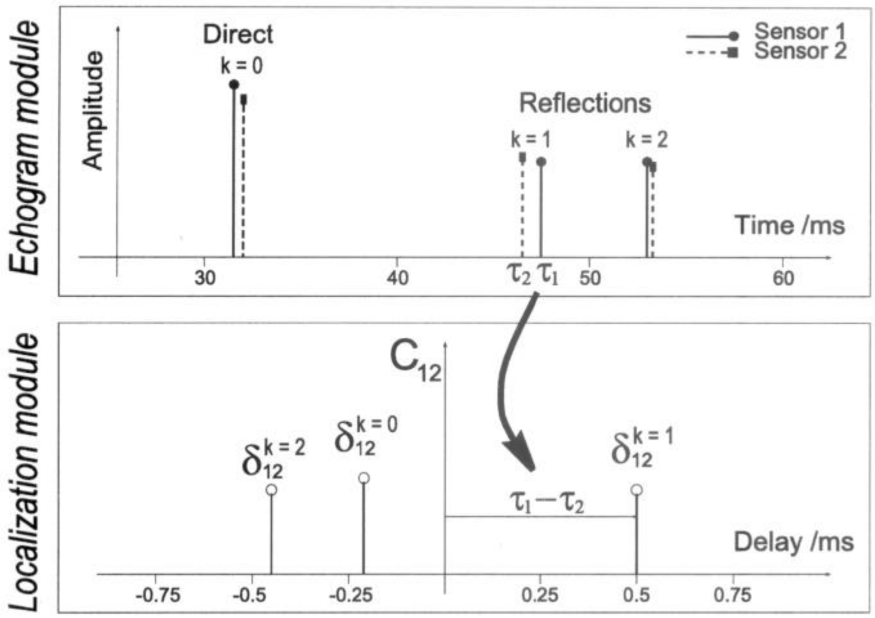

Figure 3: Contributions identification

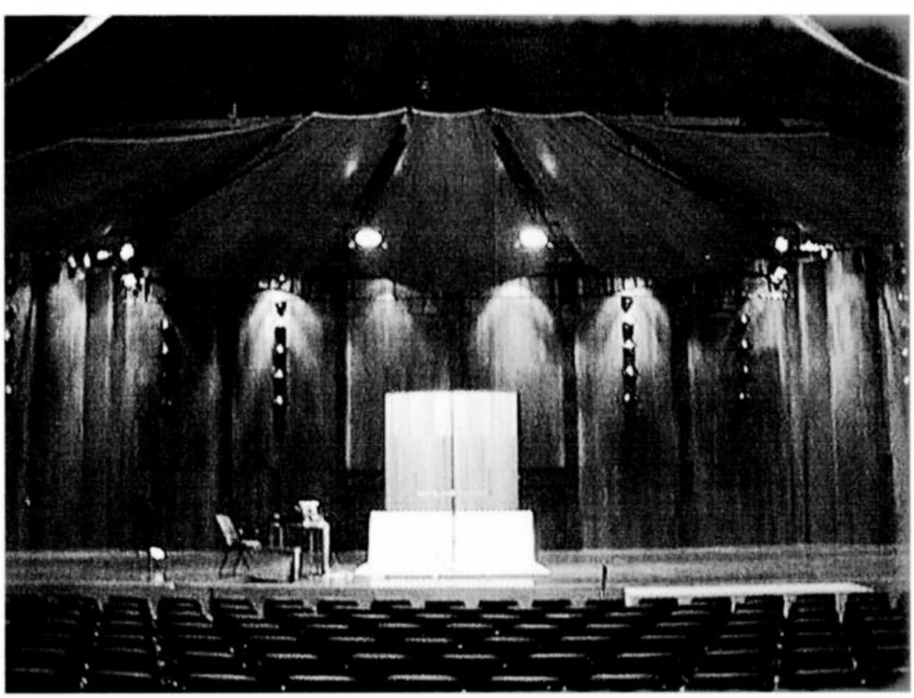

Figure 4: Orchestra sound reflectors

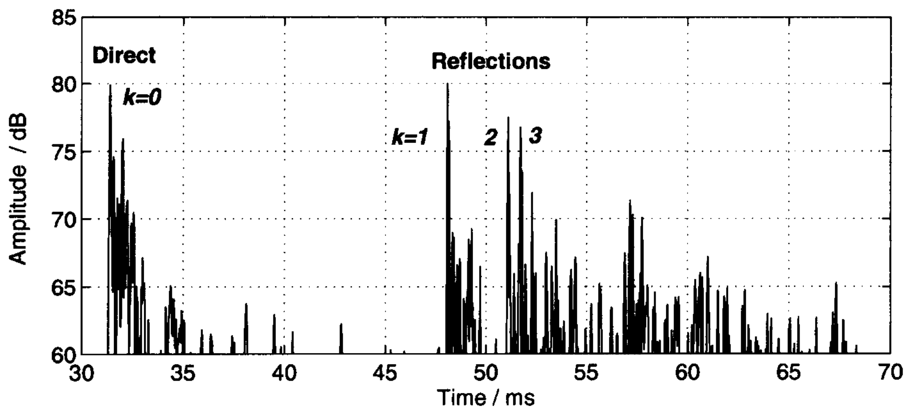

Figure 5: Echogram

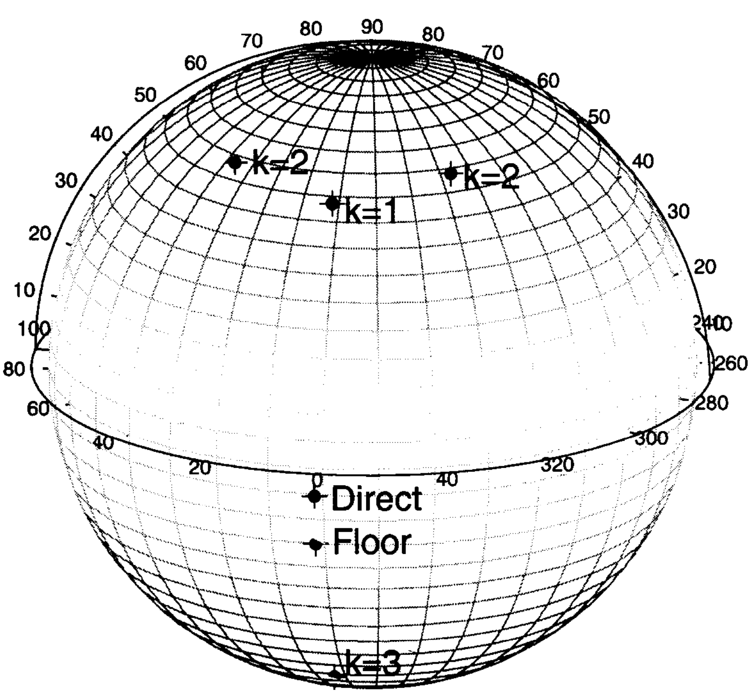

Figure 6: Localisation