Keynote

Parametric Spatial Room Impulse Response Analysis and Synthesis: A High-Resolution Approach

*Sakari Tervo*, Post-doctoral researcher
Department of Computer Science
Aalto University School of Science, Finland
Introduction to parametric estimation
Parametric estimation

Consider a note played with a bowed violin

The player played the note A3
The model

Parametric estimation requires a model of the physical world.

What defines a note?
Is it the fundamental frequency?
Which model describes the fundamental frequency?
Parametric estimation

F0 \sim 440 \text{ Hz}

Model 0
"The strongest mode"

Probability

Frequency

Magnitude [dB]
Parametric estimation

Refining the model

Model 1
“First strong mode”

Probability

F0 ~ 220 Hz
Parametric estimation

Using a more accurate signal model

F0 ~ 221 Hz

Model 2
”The average difference between strong modes”

Probability

Frequency

"The average difference between strong modes"
Parametric estimation

Using a probability distribution
F0 ~ 221.5 Hz

Model 3
"The average difference between the mean of strong modes"

Probability

Frequency

Aalto University
School of Science

Sakari.Tervo@aalto.fi
8th of September 2016, DAFx 2016
Parametric estimation

Using a probability distribution with a noise

F0 ~ 221.7 Hz

Model 4

"The average difference between the mean of strong modes, given the background noise"

Probability

Frequency

F0 ~ 221.7 Hz
Parametric estimation

https://www.youtube.com/watch?v=6JeyiM0YN04
Parametric estimation: Conclusions

Requires a model for the *signal* and the *noise*

The model is defined by a set of parameters, which are to be estimated.

The selected model depends on the application.

In reality, after George Box

> All models are wrong, some are less wrong

Typically the trade-off is

> Complexity vs accuracy
Parameterization of room impulse responses
Time-frequency dependency

Room impulse responses are time and frequency dependent.
Given a room impulse response $x$, what is the highest time-frequency resolution the analysis can have?

**Time windowing**

Inspect the signal in one time and frequency instant $x(t_0, \omega_0)$

The time window of length $L$ should include at least three wavelengths of the inspected frequency $L > 3 \times 2 \pi/\omega$
High-resolution analysis

The signal at \((t_0, \omega_0)\) in the continuous domains is therefore
\(x(t) = a \sin(\omega_0 t)\) in the time-domain and
\(x(\omega) = a \exp(-i\omega_0 + \phi_0) \delta(\omega - \omega_0)\) in the frequency domain.
Parameterization of a room impulse response

What kind of model can we assume for a time instant at one frequency? Total response is expressed as the superposition of several sinusoids

\[ x(\omega_0) = \sum_i c_i(\omega_0), \text{ where } x(t, \omega_0) \]

\( i \) is the index of the acoustic event, e.g. \( i = 0 \) is the direct sound, and \( c \) is the complex response of the event.

Since the room impulse response is linear we can write:

\[ x(\omega) = \sum_i c_i(\omega), \]

which describes the whole impulse response as a sum of sinusoidal signals.
Parameterization of spatial room impulse responses
Spatial room impulse response

A room impulse measured with a microphone array. At \( t_0 \) and in \( \omega_0 \) we have

\[
x(t, \omega) = \begin{pmatrix}
x_1(t_0, \omega_0) \\
\vdots \\
x_M(t_0, \omega_0)
\end{pmatrix}
\]

Where \( M \) is the number of microphones.

Image from mhacoustics.com
Spatial room impulse response

Again, in the time-domain, the signals are sinusoids, but with different scaling and delay.

\[ x_1(t, \omega_0) \]

\[ \vdots \]

\[ x_M(t, \omega_0) \]

time, \( t \)
Direction of arrival (DOA) $\Omega$

Since the microphones are in different positions we can obtain the direction of arrival of a sound wave.

In general we require that $M > 3$, in order to obtain the Direction of Arrival $\Omega$. 
Direction of Arrival

How many DOAs $D$ can we find from a single measurement?

In general, $D < M$, but depends on the array and frequency.
Spatial room impulse response

The measurement is affected by

\[ l(\omega) \quad h(\omega) \quad a(\Omega, \omega) \quad n(\omega) \]

Probability

\[ \Omega_s \]

DOA \( \Omega \)

(Angular frequency)

Images from genelec.com and mhacoustics.com

Sakari.Tervo@aalto.fi
8th of September 2016, DAFx 2016
Source response

Source response is different in each direction and at each distance $l(\Omega, r, \omega)$

---

Figure 5. The upper curve group shows the horizontal directivity characteristics of the 8010A measured at 1 m. The lower curve shows the system's power response.

Images from genelec.com and mhacoustics.com
Air absorption

Depends on the frequency and travelled distance
Affected by the composition of air, temperature, humidity, etc.

Images from Wikipedia
The acoustic path

The sound in the acoustic path is altered by other phenomena besides air absorption, for example.

- **Diffraction**
- **Reflections on the boundaries**

Images from Wikipedia
Array response, i.e., steering vector or array manifold

The transfer function $a(\Omega, \omega)$, i.e., response of each microphone w.r.t. frequency and DOA
Can be measured or modelled. Measurement is recommended in the general case.
Noise

Spatially white, independently and identically distributed.
Cause by the thermal noise in the electronic devices, A/D conversion etc.

Image from Wikipedia
When the source is in the far-field, the sound field is a plane wave from the array’s perspective.
Spatial room impulse response

What is simplest way of parameterizing this problem?

\[ l(\omega) \]

\[ h(\omega) \]

\[ a(\Omega, \omega) \]

Probability

\[ \Omega_s \]

DOA \( \Omega \)
A simple parameterization

A source signal $s(\omega)$ in direction $\Omega_s$

$\Omega_s$

Probability

DOA $\Omega$

$l(\Omega_s, \omega) \times h(\omega) = s(\omega)$
Parametric model

Noise is additive, the responses are convolved with each other

\[
\begin{bmatrix}
    x_1(\omega) \\
    \vdots \\
    x_M(\omega)
\end{bmatrix} =
\begin{bmatrix}
    a_1(\Omega_S, \omega)s(\omega) + n_1(\omega) \\
    \vdots \\
    a_M(\Omega_S, \omega)s(\omega) + n_M(\omega)
\end{bmatrix}
\]

In vector format

\[
x(\omega) = s(\omega)a(\Omega_S, \omega) + n(\omega)
\]
More than one “source”

\[ x = sA(\Omega) + n, \]
where
\[ \Omega = [\Omega_1, \Omega_2, \Omega_3] \]
\[ A(\Omega) = [a(\Omega_1), a(\Omega_2), a(\Omega_3)] \]
\[ s = [s_1, s_2, s_3] \]

For simplicity \( \omega \) is omitted from the above notation.
How general is the model?

The impulse response will always follow the deterministic model at a time instant and in a single frequency given that the array is in the far-field and that the array response is accurate.

Very General!
Parametric estimation
Unknown parameters

Unknown parameters in the model are DOAs $\Omega$, source signals $s$, and noise variance $\sigma^2$. 

$$x(\omega) = s(\omega)A(\Omega, \omega) + n(\omega)$$
Assumptions on the covariances

The noise is uncorrelated and identically distributed with a variance $\sigma^2$

$$E[n(\omega)n^\dagger(\omega)] = \sigma^2 I.$$ 

The source signal is deterministic

$$E[x(\omega)] = A(\Omega, \omega)s(\omega),$$

since

$$x(\omega) = A(\Omega, \omega)s(\omega) + n(\omega).$$

Thus

$$E[(x(\omega)-E[x(\omega)])(x(\omega)-E[x(\omega)])^\dagger] = \sigma^2 I,$$

and

$$E[x(\omega)x^\dagger(\omega)] = A(\Omega, \omega)s(\omega)s^\dagger(\omega)A^\dagger(\Omega, \omega) + \sigma^2 I$$
Maximum Likelihood (ML) Estimation

If we assume that the noise is of Gaussian shape we can write a likelihood function, for simplicity $\omega$ is omitted

$$p(\Omega, \Sigma, s) = |\pi \Sigma|^{\frac{1}{2}} \exp\left(-\left(x - A(\Omega)s\right) \Sigma^{-1} (x - A(\Omega)s)^{H}\right).$$

Since we have $\Sigma = \sigma^2 I$ the likelihood is presented as

$$p(\Omega, \sigma^2, s) = |\pi \sigma^2 I|^{\frac{1}{2}} \exp\left(-\left[(x - A(\Omega)s) \sigma^2 (x - A(\Omega)s)^{H}\right]\right)$$

The minimum argument of the negative log-likelihood gives the ML estimates

$$\hat{\Omega}, \hat{\sigma}^2, \hat{s} = \arg\min_{\Omega, \sigma^2, s} \left(\log(|\pi \sigma^2 I|) + \log\left([\left(x - A(\Omega)s\right) \sigma^2 (x - A(\Omega)s)^{H}\right]\right)$$

This function is in general non-linear multidimensional and has multiple minima.
Concentrated ML estimation

One can only find the estimates via non-convex optimization. The source signal \( s \) is difficult to obtain via optimization because of the circularity of the phase. When we assume that \( \Omega \) and \( s \) are fixed, the noise is given as

\[
\hat{\sigma}^2(\Omega) = \frac{1}{M} \text{Tr}(P_A^\perp(\Omega)\hat{R}), \quad (1)
\]

where

\[
P_A^\perp(\Omega) = I - A(\Omega)A^\dagger(\Omega), \text{ is the orthogonal projector to the null-space and}
\]

\[
\hat{R} = x x^H \quad \text{is the estimated covariance matrix of the microphones signals. Inserting the (1) into the likelihood gives a quadratic function}
\]

\[
\hat{\Omega}, \hat{s} = \arg \min_{\Omega, s} \left( (x-A(\Omega)s)(x-A(\Omega)s^H) \right)
\]
Concentrated ML estimation

When we minimize the quadratic function w.r.t. to $s$ we obtain

$$\hat{s}(\Omega) = A^\dagger(\Omega)x. \quad (2)$$

$$\hat{\sigma}^2(\Omega) = 1/M \text{Tr}(P_A^\perp(\Omega)\hat{R}), \quad (1)$$

Inserting (1) and (2) to the negative log-likelihood

$$\hat{\Omega}, \hat{\sigma}, \hat{s} = \arg\min_{\Omega, \sigma^2, s} \left( \log(|\pi \sigma^2 I|) + \log([((x-A(\Omega)s) \sigma^2 (x-A(\Omega)s)^H]) \right)$$

returns

$$\hat{\Omega} = \arg\min_{\Omega} \left( \text{Tr}(P_A^\perp(\Omega)\hat{R}) \right). \quad (3)$$
Concentrated ML in Matlab

1. function [nLogL, sigma2, s] = CML(Omega, M, x)
2. R = x*x'; % Covariance matrix of the measurements
3. A = getSteeringVector(Omega); % Array Response
4. invA = pinv(A); % Pseudo-inverse of A
5. PIA = eye(M)-A*invA; % Orthogonal projector to the null-space
6. nLogL = trace(PIA*R); % Equation (3), negative log-likelihood
7. sigma2 = 1/M*nLogL % Equation (1), variance
8. s = invA*y; % Equation (2), source signal
Example, two-way loudspeaker

A two-way loudspeaker has two sources, different DOAs and different source signals

\[ \Omega = [\Omega_1, \Omega_2] \]

\[ A(\Omega) = [a(\Omega_1), a(\Omega_2)] \]

\[ s = [s_1, s_2] \]
Example, two-way loudspeaker

A two-way loudspeaker has two sources, different DOAs and different source signals

\[ \Omega = [\Omega_1, \Omega_2] \]
\[ A(\Omega) = [a(\Omega_1), a(\Omega_2)] \]
\[ s = [s_1, s_2] \]
Microphone array

Instead of commercial one we use a DIY array.
1. Buy a bottle of champagne
2. Share it with our colleagues
3. Attach microphones on the surface
Example, measurement of array response

The steering vectors are measured in the far-field. Measurement is implemented in an office space, and windowing is applied to avoid reflections.

Figure 1: Diagram of the measurement setup.
Example, measurement of steering vectors

Steering vector $a(\Omega, \omega)$ magnitude response for a microphone
Example, measurement setup

Near-field measurements of the loudspeaker
Example, log-likelihood

Log-likelihood (DML) with $D = 1$ source
Example, log-likelihood

Log-likelihood (DML) with $D = 2$ sources
Example, two-way loudspeaker

Simulated two-way loudspeaker
Example, different two-way loudspeakers

Transfer functions at different distance and angles with 3 loudspeakers types
Example, different two-way loudspeakers

Transfer functions at different distance and angles with 3 loudspeakers types

Varying distance
\( \phi = 0^\circ \)
\( \theta = 0^\circ \)
\( d = \{ 0.5\text{m}, 1.0\text{m}, 1.5\text{m} \} \)

Vertical off-axis
\( d = 1\text{m} \)
\( \theta = 0^\circ \)
\( \phi = \{ -10^\circ, 0^\circ, 10^\circ \} \)

Horizontal off-axis
\( d = 1\text{m} \)
\( \phi = 0^\circ \)
\( \theta = \{ -5^\circ, 0^\circ, 5^\circ \} \)
Example, different two-way loudspeakers

Transfer functions at different distance and angles with 3 loudspeakers types
Detection of reflections
Detection of reflections

Parametric estimation requires the knowledge of the number of reflections/sources.

The number cannot be assumed to be known a priori in the general case, although, so far, it is always assumed in the literature.
Detection of reflections

Simultaneous detection and estimation detection methods are the only approaches that can be applied to the deterministic model in the coherent case.

Simultaneous estimation and detection algorithm:
Initialize \( D = 0 \)
1. \( D = D + 1 \)
2. Estimate the parameters
3. Calculate the test statistic for the parameters
4. Decide based on a priori distribution if the model fits the data
   1. If the zero hypothesis cannot be accepted Goto 1
   2. Else stop iteration
Likelihood based detection

Cochran’s Theorem states that a normalized ML estimate of variance follows Chi-squared distribution

\[ R = 2M \frac{\hat{\sigma}^2}{\hat{\sigma}^2} = 2\text{Tr}(P_A^{-1}(\hat{\Omega})\hat{R})/\hat{\sigma}^2 \sim \chi^2(\nu), \]

where \( \hat{\sigma}^2 \) is a consistent noise estimate and \( \nu \) is the degrees of freedom. The decision \( H_0 \) is accepted if

\[ D_e(H_0) = 1 \text{ if } R < \text{Chi-Inv-}\chi^2(\nu, \gamma), \]

where \( \gamma \) is a pre-defined significance level, e.g., 99%.
function D = LBD(x, sigma2c, M)
D = 0;
while 1
    D = D + 1;
    v = 2*(M-2*D); % Degrees of freedom
    fun = @(Omega) CML(Omega, M, x); % Find DOAs
    Omega = findMinimum(fun, [zeros(2*d,1)]);
    nLogL = CML(Omega, M, x); % Likelihood
    if nLogL/sigma2c < chi2inv(.99, v)
        break; % Stop iteration if criteria met
    end
end
Example: likelihood based estimation and detection

Measurement of a corner in a semi-anechoic room.

Fig. 5: The measurement setup in the semi-anechoic room. After [12].
Example: Obtaining a reference

The reference is obtained by analyzing the impulse response with short time windows 64 samples at 48 kHz and assuming $D = 1$.

Fig. 6: Real part of SH coefficients $a_{nm}$ in the time domain, and the windowing (- - - ) applied to derive the reference values. Note that the $(m + 1)^2 = 16$ SH coefficients overlap in the visualization heavily. After [12].
Example: likelihood based estimation and detection

In the evaluation we use a rectangular window length of 256 samples with an overlap of 255 samples at 48 kHz, and analyze the response at 4 kHz.

Fig. 6: Real part of SH coefficients $a_{nm}$ in the time domain, and the windowing (---) applied to derive the reference values. Note that the $(m+1)^2 = 16$ SH coefficients overlap in the visualization heavily. After [12].
Example: likelihood based estimation and detection

The reference for the detection in a window of length $L = 256$
Example: Likelihood based estimation and detection

Detection rate 92.0 %
Estimation results

Large errors in the reflection signal estimates
Reproduction of spatial sound via room impulse responses
Spatial sound reproduction

Assign the estimated source signals $\hat{s}$ to the estimated direction $\hat{\Omega}$. Playback the source signal with vector based amplitude panning, wavefield synthesis, nearest neighbor, ambisonics (whatever order).

Convolve each loudspeaker with a source signal for spatial sound.
Example: Nearest neighbour playback
Thank you for your attention!

Sakari.Tervo@aalto.fi
https://mediatech.aalto.fi/en/research/virtual-acoustics

The research leading to these results has received funding from
- the Academy of Finland, project nos. [257099]
B. Ottersten *et al.*: “Exact and Large Sample ML Techniques for Parameter Estimation and Detection in Array Processing”, in *Radar Array Processing*, Simon Haykin (ed.), Springer-Verlag, Germany, 1993
