

(51) International Patent Classification:
H04S 3/00 (2006.01)(21) International Application Number:
PCT/GB2009/051292(22) International Filing Date:
1 October 2009 (01.10.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0817950.9 1 October 2008 (01.10.2008) GB(71) Applicants (for all designated States except US): **UNIVERSITY OF SOUTHAMPTON** [GB/GB]; Highfield, Southampton, Hampshire SO17 1BJ (GB). **ELECTRONICS & TELECOMMUNICATIONS RESEARCH INSTITUTE OF KOREA** [KR/KR]; 161 Gaejeong-dong, Yusong-ku, Taejeon (KR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **NELSON, Philip** [GB/GB]; 3 Montfort Heights, Halterworth Lane, Romsey, Hampshire SO51 9LP (GB). **FAZI, Filippo, Maria** [IT/IT]; Via Sant'Ossvaldo 87, I-39100 Bolzano (IT). **SEO, Jeongil** [KR/KR]; 709-1401 Banseok-Maeul 7, Banseok-dong, Yuseong-gu, Daejeon 305-150 (KR). **KANG, Kyeongok** [KR/KR]; 101-605 Samsung-Puren Apartment, Jeonmin-dong, Yuseong-gu, Daejeon 305-390 (KR).(74) Agent: **BAKER BRETTELL LLP**; Medina Chambers, Town Quay, Southampton, Hampshire SO14 2AQ (GB).

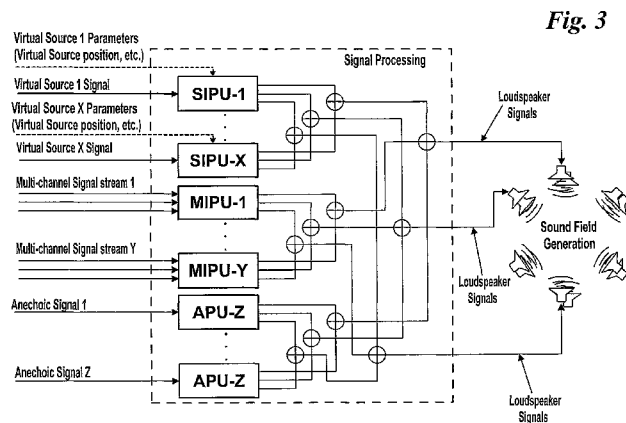
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: APPARATUS AND METHOD FOR SOUND REPRODUCTION



(57) Abstract: Method and, apparatus for implementing the method, the method comprising determining control signal data for an array of loudspeakers, the control signal data being such as to control the loudspeakers to produce a desired sound field associated with an audio signal, the method comprises determining control signal data for different frequency components of the desired sound field in respect of respective different positions in a listening volume of the loudspeaker array, wherein determination of the control signal data comprises sampling the desired sound field at the surface of a control volume (V).

APPARATUS AND METHOD FOR SOUND REPRODUCTION

Technical Field

The present invention relates to an apparatus and method for sound reproduction.

Background

- 5 In the prior art, systems are known for the recording and the reproduction of sound. Such systems provide, with an array of loudspeakers, the physical reconstruction of a desired sound field over a region of space. The sound field generated by the loudspeakers should give to the listeners located in the listening area the realistic perception of a desired virtual sound source or of a virtual sound scene.
- 10 One well known technology of this kind is Wave Field Synthesis disclosed in patent applications US 2006/0098830 A1, WO2007/101498 A1, US 2005/0175197 A1, US 2006/0109992 A1, for example. The technology uses the Kirchhoff Helmholtz equation which implies, in theory, the use of both dipole-like and monopole-like secondary sources (the loudspeakers), the strength of which (that
- 15 are proportional to the loudspeaker signals) is explicitly given by the value of the sound field on the integration contour and its normal derivative, respectively.

- Another known technology is Ambisonics, for example as disclosed in US 5,757,927 - 1998, High Order Ambisonics as disclosed in US 2006/0045275 A1 and another technology disclosed in US 2005/0141728 A1. The theoretical
- 20 formulation of these methods involve a large use of cylindrical or spherical harmonics and Legendre polynomials, in the same way that the use of sines and cosines or complex exponentials arise in the theoretical formulation of any technology based on the traditional Fourier transform. These prior art technologies involve the use of a matrix-based processing to encode the recorded
- 25 signals and generate an intermediate format, and a matrix-based system to decode this intermediate format and generate the driving signals for the loudspeakers.

We seek to provide an improved apparatus and method for reproduction of sound.

Summary

According to one aspect of the invention there is provided a method of determining control signal data for an array of loudspeakers, the control signal data
5 being such as to control the loudspeakers to produce a desired sound field associated with an audio signal, the method comprises determining control signal data for different frequency components of the desired sound field in respect of respective different positions in a listening volume of the loudspeaker array, wherein determination of the control signal data comprises sampling the desired
10 sound field at the surface of a control volume.

According to another aspect of the invention there is provided sound reproduction apparatus for processing an audio signal, the apparatus configured to output control signal data for an array of loudspeakers to produce a desired sound field associated with the audio signal, wherein the apparatus configured to determine
15 the control signal data for different frequency components of the desired sound field in respect of respective different positions in a listening volume of the loudspeaker array, wherein determination of the control signal data comprises sampling the desired sound field at the surface of a control volume.

A further aspect of the invention relates to a signal processor configured to process
20 the audio signal of the above aspects of the invention and output the control signal data. The signal processor may be configured by suitable machine-readable instructions, and the instructions may be realised in the form of a signal or a data carrier device.

Brief Description of the Drawings

25 Various embodiments of the invention will now be described by way of example only, with reference to the following drawings in which:

Figure 1 shows a co-ordinate system,

Figure 2 shows of different three-dimensional regions,

Figure 3 is a schematic diagram showing various functional components of a sound reproduction system.

Figure 4 is a block diagram of a single input processor arrangement,

5 **Figure 5** is a block diagram of a multiple input processor arrangement,

Figure 6 is a block diagram of auralisation processor arrangement,

Figure 7a and 7b are perspective views of a microphone array,

Figure 7c is a detailed view of part of the microphone array of Figures 7a and 7b, and

10 **Figure 7d** is a plan view of microphone array of Figures 7a and 7b

Detailed Description

The theoretical background on which the inventive apparatus is based is as follows. The signals driving the array of loudspeakers are designed to be such that the difference (more specifically the L^2 norm of the difference) between the desired or
15 target sound field and the sound field generated by the array of loudspeakers is minimized on the surface of a three dimensional region, herein called the control volume. The problem is formulated mathematically as an integral equation of the first kind, and its solution is used in order to suitably configure the signal processing apparatus embedded in the system. The solution of the integral
20 equation is often an approximated solution because of the mathematical ill-posedness of the problem. As described in detail below, the solution of the integral equation can be computed either with an analytical approach or with a numerical method. The two cases define two different approaches for the design of the signal processing apparatus. Both methods are described in this application.

An aspect of the below described embodiments is that different choices of some parameters and/or solution methods are chosen depending on the frequency of the sound to be reproduced.

When the integral equation is solved with a numerical method, the latter requires
5 that the target sound field is defined at a finite number of points on the boundary of the control volume, which is therefore sampled. The numerical solution of the integral equation is influenced by two sources of error: one due to ill-conditioning and one due to spatial aliasing. The former is more likely to affect the performance of the system at low frequencies, while the latter degrades the
10 performance at high frequencies. These undesired effects can be avoided or contained by a wise choice of the control volume and of the sampling scheme used on the boundary of the control volume. It can be shown that at low frequencies the effects of ill-conditioning can be limited by choosing a large control volume. On the other hand the effects of the spatial aliasing can be reduced by choosing a
15 regular sampling scheme of the surface of the control volume, for which the average distance between any two neighboring sampling points is less than half of the wavelength considered. For this reason, we define a control volume which is a function of the frequency of the sound to be reproduced: a large control volume at low frequencies and a gradually smaller control volume for higher frequencies.
20 Even though the size of the control volume varies with the frequency its shape does not vary, and the number and geometrical arrangement of the samples of the boundary of that region generally does not vary, but in some cases it may be advantageous to allow such a variation. This approach allows the definition, for each frequency, of a control volume which is large enough to avoid the problems
25 arising from ill-conditioning, keeping at the same time the distance between neighboring sampling points below the spatial aliasing limit.

A further aspect of the embodiments below relate to the solution of problems which can arise at certain frequencies that correspond to the so-called Dirichlet eigenvalues of the control volume. These problems arise when an attempt is made
30 to reconstruct a sound field which is defined, at those frequencies, only on the

boundary of the control volume. These critical frequencies are determined by the shape and size of the control volume. Our choice of a frequency dependent control volume has been applied to overcome these difficulties. In addition to that, we have chosen to define or measure the sound field at one or more locations in the interior of the control volume. In the design of the microphone array discussed below, when the microphones are arranged on spherical layers, we have chosen to include a microphone also in the centre of the microphone array. This choice has proven to overcome the problems due to the first critical frequency.

Another aspect of the embodiments below is related to the reproduction of a high frequency sound field due to a point source at a given location. In this case the approach used at low frequencies would lead all or almost all the loudspeakers of the array to contribute to the reproduced sound field. They would generate a large amount of acoustic energy, and the sound fields due to the different loudspeakers would generate a complex pattern of destructive and constructive interferences which would reconstruct the target field just over a limited region of the space. We have observed that the digital filters, which are part of the signal processing apparatus and correspond to those loudspeakers which are closer to the location of the virtual source, exhibit an asymptotic behavior at high frequencies. The amplitude of these filters tends to be constant and their phase tends to become linear. This fact has been proven mathematically in the special case when the loudspeakers are regularly arranged on the surface of a sphere, but similar conclusions can be inferred also for other geometric arrangements. For this reason, for the Single Input Mode described below, the processing applied to a high frequency component of the audio signal is different than that applied to the low frequency component of the signal. For the high frequency processing, only the loudspeakers which are the closest to the location of the virtual source are activated, and the corresponding digital filters have been substituted with simple gains, which resemble the asymptotic behavior of those filters. The Single Input Mode of the system may be viewed as a hybrid of a sound field reconstruction system and a three dimensional sound panning system. This approach has multiple advantages: in the first place, the amount of acoustic energy generated by the

system is reduced. As a second instance, the real time calculation of the gains is computationally considerably less expansive than the real time calculation of an equivalent number of digital filters. Finally, the fact that only the loudspeaker, or those loudspeakers, close to the virtual source location are activated provides
5 better cues at high frequencies for the human localization of a virtual sound source.

An especially innovative aspect of the embodiments described below is that the control volume can be chosen to be dependent on the frequency of the sound to be reproduced, and different processing steps are in general applied to the signals at different frequency bands. It is also worthwhile mentioning that the recording
10 devices and reproduction devices of this system are designed to operate as a unique system, and there is no intermediate audio format on which the information between the recording part and the reproduction part of the system is transmitted. Finally, a further innovative feature of this invention is constituted by the Auralisation Mode (and the Auralisation Processing Unit), which allows to apply
15 the theory of the reconstruction of a sound field to the design of a multi-channel reverberation simulator.

The principles described above have been applied to the design of the different components which constitute the signal processor apparatus of the sound reproduction system. The signal processor apparatus can be used to generate
20 sound fields of three different characteristics:

1. A sound field generated in the free field by a virtual point source, whose location in the space and whose radiation pattern are selected.
2. A sound field generated by an omnidirectional point source in a reverberant environment, whose reverberant field is described by a set of measured or
25 simulated impulse responses.
3. A generic sound field, described by a set of recorded sound signals acquired using a microphone array.

These three different characterizations define three different operating modes for the system, which are here defined as the Single Input Mode, the Auralisation Mode and the Multiple Input Mode, respectively. The general layout of the sound reproduction system comprises sound recording devices, modular signal processors and an array of loudspeakers (and their associated amplifiers). While the loudspeaker array is the same for the three different operational modes, the input devices and the signal processing are different for each of the modes.

In the Single Input Mode, the input signal is a single audio signal. The latter is obtained by capturing with a microphone the sound field generated by a sound source in an anechoic environment, or alternatively in a moderately reverberant environment, with the microphone located at a short distance from the source of sound. Additional data are provided, which describe the location of the virtual source (azimuth, elevation and distance) and preferably its radiation pattern as a function of two angles and of the frequency. The signal is digitally processed by the Single Channel Processing Unit, which is described in detail below. The location of the virtual source can be within or outside the loudspeaker array, but always outside of the control volume.

In the Multiple Input Mode, the input is a set of audio signals acquired with an array of microphones, designed for this purpose. The signals are processed by the Multiple Input Processing Unit, which is described in detail later and which is constituted by a matrix of digital filters. These are computed from a set of impulse responses describing the transfer function between each loudspeaker of the loudspeaker array and each microphone of the microphone array. These impulse responses can be either measured or computed from a theoretical model.

In the Auralisation mode the input signal is the same as for the Single Input Mode (a single anechoic audio signal), while the signal is digitally processed by the Auralisation Processing Unit described later. The latter is constituted by a set of digital filters, which are computed from two sets of impulse responses. The first set is constituted by the impulse responses of the considered reverberant

environment measured with a specially designed microphone array. The second set of impulse responses is the same as that described for the Multiple Input Mode (describing the transfer functions between loudspeakers and microphones of the two arrays).

- 5 There is no *a priori* constraint on the loudspeaker arrangement, apart from the mild constraint that they should be arranged on a three dimensional surface, which surrounds the listening area and their axes are pointed towards the geometrical centre of the array. However, the system gives the best performance when a relatively large number of loudspeakers (eight or more) is used, they exhibit a
- 10 preferably omnidirectional radiation pattern and they are regularly arranged on the surface of a sphere or of a hemisphere. Regular arrangement means here that the average distance between any two neighboring loudspeakers is constant or almost constant.

An example is now given to highlight the differences between the multiple input

15 mode and the auralisation mode. One desires to reproduce the effect of the sound field generated by a violin playing in a concert hall (which is a reverberant environment). One can either

1. Use multiple input mode by recording the violin playing in the hall with the microphone array and then use the output of the microphone array as the input to
- 20 the Multiple Input Processing Unit. The information on the sound of the violin and on the reverberant field of the hall are captured at the same time by the array and cannot, in principle, be divided, or

2. Use the auralisation mode by recording with the microphone array a set of impulse responses describing the reverberant field of the hall considered. These
- 25 measurements do not depend on the sound of the violin and are used to calculate the digital filters in the Auralisation Processing Unit. Then one records the violin playing in an anechoic environment (an artificial non-reverberant environment, without reflections) and then use that single audio signal as the input to the

Auralisation Processing Unit. The information on the sound of the violin and on the reverberant field of the hall are captured separately.

In summary, the input to the multiple input unit are the signals from the microphone array. No assumption regarding the form of the field is made and the
5 reverberant characteristic of the hall is not involved on the calculation of the digital filters. They are computed only from the knowledge of the characteristics of the microphone array and of the loudspeaker array geometry.

On the other hand, the input to the Auralisation Processing unit is a single audio signal (the sound of the violin playing in a non-reverberant environment). The
10 reverberant characteristics of the hall considered, represented by a set of impulse responses, are this time involved in the calculation of the digital filters of the Auralisation Processing Unit.

While the reverberant field and the sound of the violin are captured together for the Multiple Input Mode, they are in the Auralisation Mode captured with separate
15 procedures and are artificially merged by the auralisation processing unit.

The Multiple Input Mode has the advantage of capturing and reproducing a natural and real sound of an acoustic source in a given reverberant or non-reverberant environment.

The Auralisation mode has the advantage that the reverberant characteristic of the
20 room and the sound from the given acoustic source are acquired separately and merged together later. With the Auralisation mode, it is possible to use the same reverberant hall and to change the sound source (for example a violin, a piano etc playing in a given concert hall) or conversely to use the same source of direct sound (the violin) and change the reverberant environment (with the artificial
25 effect of having the same musician playing his/her violin in different concert halls).

Figures 7a to 7d show microphone array 100 comprising a framework for supporting a set of thirty microphones. The array is designed to be used for the auralisation mode, for measuring the impulse responses of the reverberant environment to be emulated. The framework comprises two substructures 110 and 120 of substantially arcuate form. The substructure 110 comprises two spaced apart arcuate members 112 and 113 connected by a bridging member 114. The substructure 120 comprises two spaced apart members 122 and 123, the spaced apart members connected by a bridging member 124. Each of the substructures is disposed at different radii 121 and 122 from a centre point. The substructures 110 and 120 are pivotably mounted about a pole 150. Apertures 140 are provided in the arcuate members and the bridging members and are dimensioned to locate the microphone devices (not illustrated). A removable rigid sphere 160 is located inside the array. It is to be noticed that the two layers of microphones do not extend over a entire sphere but extend only over sectors. For this reason, the measurement of the impulse responses must be repeated eight times. At each measurement, thirty signals are acquired by the thirty microphones. After each measurement step, the array is rotated of 45° around its vertical axis. The eight measurements allow one to obtain a set of two hundred and forty impulses, captured at locations corresponding to an approximately regular sampling of two concentric spherical surfaces. Because the two hundred and forty signals can not be captured at the same time, this microphone configuration can be used for the auralisation mode only and not for the multiple input mode, for which the microphone signals should be acquired simultaneously.

Theoretical Analysis

In what follows, vectors are represented by lower case bold letters. The convention for the spherical co-ordinates r_x , θ_x , ϕ_x of a given vector $\mathbf{x}$ is illustrated in Figure 1. Matrices are represented by capital bold letters. The imaginary unit is given by $i = \sqrt{-1}$. The symbol $[\cdot]^*$ represents the complex conjugate and the symbol $[\cdot]^H$ represents the Hermitian transpose (or conjugate transpose) of a matrix. δ_{nm} is the Kronecker delta ($\delta_{nm} = 1$ if $n = m$, $\delta_{nm} = 0$ if $n \neq m$).

Let $V \subseteq R^3$ be a bounded and simply connected volume of the three dimensional space, with boundary ∂V (of class C^2). A cross section of this volume is represented in Figure 2. This region is referred to as the control volume. The assumption is made that the target sound field $p(\mathbf{x}, t)$ is due to sources of sound
 5 that are not contained in V , and that no scattering object is contained within this region. $p(\mathbf{x}, t)$ satisfies the homogeneous wave equation

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} = 0 \quad (1)$$

in V , where c is the speed of sound, considered to be uniform in V . When a single frequency ω is considered and $p(\mathbf{x}, t) = \text{Re}\{p(\mathbf{x})e^{-i\omega t}\}$, equation (1) can be
 10 reformulated as the Helmholtz equation

$$\begin{aligned} \nabla^2 p(\mathbf{x}) + k^2 p(\mathbf{x}) &= 0 \\ \mathbf{x} &\in V \end{aligned} \quad (2)$$

where $k = \omega/c$ is the wave number and the time dependence $e^{-i\omega t}$ has been omitted. Let now $\Lambda \subseteq R^3$ be a bounded and simply connected region of the space, with boundary $\partial\Lambda$ of class C^2 , that fully encloses V . Assume now that a continuous
 15 distribution of an infinite number of secondary, monopole-like sources is arranged on the boundary $\partial\Lambda$. This continuous distribution of secondary sources is the ideal model of the loudspeaker array, which is useful for the mathematical formulation of the problem. Later on the assumption that the number of secondary sources is infinite will be removed.

20 The assumption is made that the sound field $p_y(\mathbf{x})$ generated by the secondary source located at $\mathbf{y} \in \partial\Lambda$ can be represented by a Green function $G(\mathbf{x}|\mathbf{y})$, thus satisfying the inhomogeneous Helmholtz equation

$$\begin{aligned} \nabla^2 p_y(\mathbf{x}) + k^2 p_y(\mathbf{x}) &= -a(\mathbf{y})\delta(\mathbf{x} - \mathbf{y}) \\ \mathbf{x} &\in \overline{V}, \quad \mathbf{y} \in \partial\Lambda \end{aligned} \quad (3)$$

where the function $a(\mathbf{y})$ represents the complex strength of the secondary sources.

In a free field, $p_y(\mathbf{x})$ can be represented by the free field Green function

$$p_y(\mathbf{x}) = a(\mathbf{y})g(\mathbf{x}|\mathbf{y}) = a(\mathbf{y}) \frac{e^{ik|\mathbf{y}-\mathbf{x}|}}{4\pi|\mathbf{y}-\mathbf{x}|} \quad (4)$$

$\mathbf{x} \in R^3, \mathbf{y} \in \partial\Lambda$

In a reverberant environment, the Green function has a more complex expression,
 5 which strongly depends on the shape of the reverberant enclosure and on its impedance boundary conditions.

The sound field $\hat{p}(\mathbf{x})$ generated by the infinite number of secondary sources uniformly arranged on $\partial\Lambda$ can be expressed as an integral, representing the linear superposition of the sound fields generated by the single sources:

$$10 \quad \hat{p}(\mathbf{x}) = (Sa)(\mathbf{x}) = \int_{\partial\Lambda} G(\mathbf{x}|\mathbf{y})a(\mathbf{y})dS(\mathbf{y}) \quad (5)$$

$\mathbf{x} \in V$

In what follows, $\hat{p}(\mathbf{x})$ is also referred to as the reconstructed or reproduced sound field, while the integral introduced is often referred to as a *single layer potential* and $a(\mathbf{y})$ is called the *density* of the potential.

Let now consider the following differential equation

$$15 \quad \begin{aligned} \nabla^2 p(\mathbf{x}) + k^2 p(\mathbf{x}) &= 0 & \mathbf{x} \in V \\ p(\mathbf{x}) &= f(\mathbf{x}) & \mathbf{x} \in \partial V \end{aligned} \quad (6)$$

where the function $f(\mathbf{x})$ describes the value field $p(\mathbf{x})$ on the boundary ∂V . This differential equation is known as the Dirichlet problem (and the related boundary condition is called after the same name). As the Kichhoff-Helmholtz integral equation suggests, the knowledge of both the sound field and its normal derivative
 20 on the boundary (the Cauchy boundary condition) uniquely defines a sound field

in the interior region V . However, this condition can be relaxed. In fact, under the above mentioned conditions and with the appropriate regularity assumption on the function $f(\mathbf{x})$, the Dirichlet problem (5) has a unique solution. This implies that the knowledge of the acoustic pressure on the boundary ∂V of the control
 5 volume is enough to define completely the sound field in the interior of volume V . This holds as long as the wave number k is not one of the Dirichlet eigenvalues k_n . The latter are defined as the infinite and countable wave numbers k_n such that the differential equation (5) with homogeneous boundary conditions $f(\mathbf{x})=0$ is satisfied.

10 It is worth mentioning that the sound field in V can be ideally extended by analytical continuation to the exterior of the control volume, provided that the considered region does not contain any source of sound. This implies that if the acoustic field is reconstructed perfectly on the boundary of the control volume V , then it is reconstructed perfectly also in its interior and partially in the neighboring
 15 exterior region.

The determination of the density $a(\mathbf{y})$ can be formulated as

$$p(\mathbf{x}) = (Sa)(\mathbf{x}) = \int_{\partial\Lambda} G(\mathbf{x}|\mathbf{y})a(\mathbf{y})dS(\mathbf{y}) \quad (7)$$

$\mathbf{x} \in \partial V$

where $p(\mathbf{x})$ is given and the density $a(\mathbf{y})$ is the unknown of the problem. Note that the integral operator $(Sa)(\mathbf{x})$ has been defined here as the restriction of the
 20 single layer potential (5) to the boundary ∂V . Equation (7) is an integral equation of the first kind and the determination of $a(\mathbf{y})$ from the knowledge of $p(\mathbf{x})$ on the boundary ∂V represents an inverse problem. It is important to highlight that equation (7) represents a problem that is, in general, ill-posed. This implies that a solution $a(\mathbf{y})$ might not exist, and even if it exists it might be non-unique or not
 25 continuously dependent on the data $p(\mathbf{x})$. The latter concept implies that small variations or errors on $p(\mathbf{x})$ can result in very large errors in the solution $a(\mathbf{y})$,

which is therefore said to be unstable. It is however always possible to compute an approximate but robust solution by applying a regularization scheme.

It is important to highlight that the integral equation (7) is different from the Kirchhoff-Helmholtz equation (often also called Green formula) on which the technology called Wave Field Synthesis is grounded. The first main difference is that the integrand in the Kirchhoff-Helmholtz equation involves both monopole-like and dipole-like secondary sources, and the expression of their strength is expressed explicitly. On the other hand, the proposed approach relies on the use of monopole-like secondary sources only, and the determination of their strength is determined by the solution of the integral equation. The second main difference is that the field is known on the boundary ∂V , which is not the same as the boundary $\partial \Lambda$ on which the secondary sources are arranged. This point describes also the main difference between the proposed approach and the Simple Source Formulation. It is possible to choose the control volume as a function of the frequency, while the secondary sources are arranged on the surface $\partial \Lambda$ which does not depend on the frequency.

It is now possible to seek a solution to equation (7) using two different methods, one based on an analytical approach and the other based on a numerical solution of discretised version of the integral. As described above, the solution of the integral is applied to the design the signal processing unit embedded in the proposed system. In order to do that, the continuous solution provided by the analytical approach needs to be slightly modified in order to be adapted to the finite number of loudspeakers. The number of loudspeakers is L and the position of the acoustic centre of the l -th loudspeaker is identified by the vector $\mathbf{y}_l$.

25 Analytical solution of the integral equation

The first method to calculate a solution to equation (7) is an analytical method based on the singular value decomposition of the integral operator and is described in what follows.

To begin with, the scalar product of two square integrable functions $f(\mathbf{y})$ and $g(\mathbf{y})$, with the same domain D , is here defined as

$$\langle f|g \rangle = \int_D f(\mathbf{x})^* g(\mathbf{x}) dS(\mathbf{x}) \quad (8)$$

The adjoint operator S^+ of S is defined to be such that

$$5 \quad \langle g|S a \rangle = \langle S^+ g|a \rangle \quad (9)$$

It can be easily verified that, for the operator S introduced in equation (7), the adjoint operator S^+ is given by

$$(S^+ g)(\mathbf{y}) = \int_{\partial V} G(\mathbf{y}|\mathbf{x})^* g(\mathbf{x}) dS(\mathbf{x}) \quad (10)$$

$\mathbf{y} \in \partial\Lambda$

The physical meaning of the single layer potential $(Sa)(\mathbf{x})$ is represented by the sound field generated by the continuous distribution of secondary sources on $\partial\Lambda$, evaluated at $\mathbf{x} \in \partial V$. Similarly, its adjoint operator $(S^+ g)(\mathbf{y})$ can be regarded as the time reversed version of the sound field generated by a continuous distribution of monopole-like secondary sources on ∂V , evaluated at $\mathbf{y} \in \partial\Lambda$. The time reversal is due to the fact that the kernel of the integral (10) is the complex conjugate of the Green function $G(\mathbf{y}|\mathbf{x})$. Considering, as an example, the case of the free field Green function $g(\mathbf{x}|\mathbf{y})$ introduced in equation (4), it can be easily verified that $g(\mathbf{y}|\mathbf{x})$ and $g(\mathbf{y}|\mathbf{x})^*$ differ because of the sign of the exponential. This implies that while $g(\cdot|\mathbf{x})$ can be considered as the representation of a sound field generated by a source of outgoing (diverging) spherical waves located at $\mathbf{x}$, $g(\cdot|\mathbf{x})^*$ could be understood as the case of a source of incoming (converging) spherical waves located at the same position $\mathbf{x}$. Alternatively, the sound field represented by $g(\cdot|\mathbf{x})^*$ could be regarded as the time reversed version of $g(\cdot|\mathbf{x})$, as

$$\operatorname{Re}\left\{e^{-i\omega t} g(\mathbf{x}|\mathbf{y})^*\right\} = \operatorname{Re}\left\{\frac{e^{-i(k|\mathbf{x}-\mathbf{y}|+\omega(-t))}}{4\pi|\mathbf{x}-\mathbf{y}|}\right\} = \operatorname{Re}\left\{e^{-i\omega(-t)} g(\mathbf{x}|\mathbf{y})\right\} \quad (11)$$

For both interpretations, the generated spherical wave fronts are converging towards $\mathbf{x}$. As a consequence of what has been said, the adjoint operator S^+ could be interpreted as a continuous distribution of "sources of incoming waves" on ∂V .

- 5 It can be shown that the operator S is compact and therefore its adjoint operator is compact too and the composite operator S^+S is compact and self-adjoint. It is therefore possible to apply the properties of compact self adjoint operators and to perform a spectral decomposition of S^+S . As a first step, the set of functions $a_n(\mathbf{y})$ is computed, these functions being solutions of the eigenvalue problem

$$10 \quad (S^+S a_n)(\mathbf{y}) = \lambda_n a_n(\mathbf{y}) \quad (12)$$

$$\mathbf{y} \in \partial\Lambda, \quad n = 1, 2, 3, \dots, \infty$$

- where the eigenvalues λ_n are real, positive numbers. Considering the interpretation of S and S^+ introduced above, the effect of the composite operator $(S^+Sa)(\mathbf{y})$ could be understood as follows: a sound field is generated by the continuous distribution of monopole-like sources on $\partial\Lambda$ with strength $a(\mathbf{y})$, and
- 15 this field, on the boundary ∂V , is described by the function $\hat{p}(\mathbf{x})$. Mathematically this implies that $\hat{p}(\mathbf{x}) = (Sa)(\mathbf{x})$. A different sound field is then generated by a continuous distribution of monopole-like sources on ∂V (the operator S^+), and the strength of the sources is determined by the function $\hat{p}(\mathbf{x})$. The sound field is then time reversed (or alternatively the secondary sources on ∂V can be replaced by
- 20 "sources of incoming waves"), and the function $\hat{a}(\mathbf{y})$ describes the value of this field on $\partial\Lambda$. In mathematical terms, this corresponds to $\hat{a}(\mathbf{y}) = (S^+\hat{p})(\mathbf{y})$. Summarizing, the effect of the operator S^+S can be understood as if the field generated by the secondary sources on $\partial\Lambda$ was propagated from $\partial\Lambda$ to ∂V and then propagated back to $\partial\Lambda$. Attention is now focused on the relation between
- 25 $\hat{a}(\mathbf{y})$ and $a(\mathbf{y})$: if these two functions are such that $\hat{a}(\mathbf{y}) = \lambda_n a(\mathbf{y})$, $\lambda_n \in R^+$, then the

function $a(\mathbf{y})$ is one of the eigenfunctions $a_n(\mathbf{y})$ of S^+S , and is a solution of (12). This means that the action of S^+S on $a_n(\mathbf{y})$ is simply an amplification or attenuation of the function, corresponding to positive real number λ_n .

The set of functions $a_n(\mathbf{y})$ constitutes an orthogonal set of functions for $\partial\Lambda$ meaning that any square integrable function $a(\mathbf{y})$ defined on $\partial\Lambda$ can be expressed as

$$a(\mathbf{y}) = \sum_{n=1}^N \langle a_n | a \rangle a_n(\mathbf{y}) + (Qa)(\mathbf{y}) \quad (13)$$

The integer N depends on the dimension of the range of S and might be infinite. $(Qa)(\mathbf{y})$ is the orthogonal projection of $a(\mathbf{y})$ on the null-space of S . The latter is defined as the set of functions $\tilde{a}(\mathbf{y})$ such that

$$N(S) = \{\tilde{a}(\mathbf{y}) : (S\tilde{a})(\mathbf{x}) = 0\} \quad (14)$$

If this set is empty, then the set of functions $a_n(\mathbf{y})$ is complete and equation (13) can be regarded as a generalized Fourier series. We can also generate a set of orthogonal functions $p_n(\mathbf{x})$ on ∂V by letting the operator S act on the functions $a_n(\mathbf{y})$ such that

$$\sigma_n p_n(\mathbf{x}) = (S a_n) \quad n = 1, 2, 3, \dots \infty \quad (15)$$

where the positive real numbers $\sigma_n = \sqrt{\lambda_n}$ are the singular values of S . It can be also proved that

$$\sigma_n a_n(\mathbf{y}) = (S^+ p_n) \quad n = 1, 2, 3, \dots \infty \quad (16)$$

It is possible to express any square integrable function $p(\mathbf{x})$ on ∂V as the infinite series

$$p(\mathbf{x}) = \sum_{n=1}^N \langle p_n | p \rangle p_n(\mathbf{x}) + (Rp)(\mathbf{x}) \quad (17)$$

(Rp)($\mathbf{x}$) being the orthogonal projection of $p(\mathbf{x})$ on the null-space of S^+ .
Combining equations (13) and (15) and keeping in mind that $(S(Qa))(\mathbf{x}) = 0$ because
of the definition of the null-space (14), it is possible to express the action of S on
5 $a(\mathbf{y})$ as

$$(Sa)(\mathbf{x}) = \sum_{n=1}^N \sigma_n \langle a_n | a \rangle p_n(\mathbf{x}) = \hat{p}(\mathbf{x}) \quad (18)$$

$\mathbf{x} \in \partial V$

It is now possible to calculate an approximate solution of the integral equation (7)
as

$$a(\mathbf{y}) = \sum_{n=1}^N \frac{1}{\sigma_n} \langle p_n | p \rangle a_n(\mathbf{y}) \quad (19)$$

- 10 This equation provides a very powerful method for solving the problem of sound field reproduction with a continuous layer of monopole-like secondary sources. In fact, considering equations (15), (17) and (19), the single layer potential with the density computed in equation (18) is given by

$$\begin{aligned} \int_{\partial\Lambda} G(\mathbf{x} | \mathbf{y}) a(\mathbf{y}) dS(\mathbf{y}) &= \int_{\partial\Lambda} G(\mathbf{x} | \mathbf{y}) \left(\sum_{n=1}^N \frac{1}{\sigma_n} \langle p_n | p \rangle a_n(\mathbf{y}) \right) dS(\mathbf{y}) = \\ &= \sum_{n=1}^N \frac{1}{\sigma_n} \langle p_n | p \rangle (Sa_n)(\mathbf{x}) = \sum_{n=1}^N \langle p_n | p \rangle p_n(\mathbf{x}) = p(\mathbf{x}) - (Rp)(\mathbf{x}) = \hat{p}(\mathbf{x}) \end{aligned} \quad (20)$$

- 15 This implies that the reconstructed sound field $\hat{p}(\mathbf{x})$ is the component of the target field that does not belong to the null-space of the adjoint operator S^+ , or equivalently that $\hat{p}(\mathbf{x})$ is the projection of $p(\mathbf{x})$ on the sub-space defined by the range of S . In other words, with the usual condition on the Dirichlet eigenvalues,

if the target field has a pressure profile $p(\mathbf{x})$ on ∂V that can be expressed as a linear combination of the orthogonal functions $p_n(\mathbf{x})$, then it is ideally possible to determine a density $a(\mathbf{y})$ such that $\hat{p}(\mathbf{x})=p(\mathbf{x})$ in V . In the case when $p(\mathbf{x})$ is not in the range of S , it can be shown that the reconstructed field $\hat{p}(\mathbf{x})$ in (20) is the

5 function that belongs to the range of S that minimizes the L^2 norm of the difference $\|p(\mathbf{x})-\hat{p}(\mathbf{x})\|_{L^2}$ on ∂V .

Considering equation (16), it can be noticed that even if some of the functions $p_n(\mathbf{x})$ do not rigorously belong to the null-space of S^+ , their related singular value σ_n can be so small that it is possible to consider these $p_n(\mathbf{x})$ as if $(S^+ p_n)(\mathbf{x}) \approx 0$.

10 This sheds some light on the ill-conditioning of the inverse problem represented by the integral equation (7). In fact, if for a given n the corresponding singular value σ_n is very small, then its reciprocal is very large and the norm of the density $a(\mathbf{y})$ computed with the series (19) might become unreasonably large. Furthermore, the factor $1/\sigma_n$ is related to the amplification of errors contained in the data $p(\mathbf{x})$ and

15 therefore to the stability of the system. In order to prevent this ill-conditioning problem, it is possible to regularize the solution (19), applying a cut-off to the spectrum of the operator, using the Tikhonov regularization or any other regularization technique. In the design of the digital filters used in the system described in this application, a combination of spectral cut-off and Tikhonov

20 regularization is used and the application of this is described later.

It has been shown that the possibility of calculating a density $a(\mathbf{y})$ for reconstructing the target sound field $p(\mathbf{x})$ with the single layer potential (5) from the knowledge of the boundary values of $p(\mathbf{x})$ on ∂V is strongly related to the null-space of the adjoint operator S^+ . It is important to emphasize that if it is not

25 possible to calculate exactly the density $a(\mathbf{y})$, this does not imply that the latter does not exist and that the target sound field can not be perfectly reconstructed by the continuous distribution of secondary sources. An analogous if not identical problem arises in the field of Acoustic Holography: suppose that an attempt is

made to determine the surface normal velocity of a radiating plate from the knowledge of the radiated sound field on a given region. It turns out that the vibration modes that generate evanescent waves cannot in practice be determined from far field measurements. However, this does not mean that these low-
 5 efficiency vibro-acoustic modes do not exist. This highlights the fact that the feasibility of the reconstruction of the target sound field not only depends on the arrangement $\partial\Lambda$ of the secondary sources, but also is very much related to the surface ∂V on which the sound field has been defined or measured.

This method of solution has the big advantage that the density $a(\mathbf{y})$ has an
 10 analytical expression and the design of the digital filters implemented in the system does not require any numerical matrix inversion. However, there are two disadvantages of this method. The first is that the eigenfunctions $a_n(\mathbf{y})$ strongly depend on the geometry of V and Λ and their explicit calculation is usually not trivial. Their formulation is known for a limited number of geometries. The
 15 second disadvantage is that, when the continuous distribution of secondary sources is substituted by an array of a finite number of loudspeakers, the performance of the system whose signal processing units have been designed with this method are more effective if the distribution of the secondary sources (the loudspeakers) is regular.

20 As an example, the special case is now considered, where the boundaries of the two volumes V and Λ are two concentric spheres, with radius R_V and R_Λ , respectively. The assumption of a free field is also made here. For the case under consideration the functions $a_n(\mathbf{y})$ and $p_n(\mathbf{x})$ and the singular values σ_n can be expressed analytically as

$$\begin{aligned}
 a_n(\mathbf{y}) &= \frac{1}{R_\Lambda} Y_n^m(\theta_y, \phi_y) \\
 25 \quad p_n(\mathbf{x}) &= \frac{\gamma_n}{R_V} Y_n^m(\theta_x, \phi_x) \\
 \sigma_n &= R_\Lambda R_V k \left| h_n^{(1)}(kR_\Lambda) j_n(kR_V) \right|
 \end{aligned} \tag{21}$$

$j_n(\cdot)$ is the spherical Bessel function of order n , $h_n^{(1)}(\cdot)$ is the spherical Hankel function of the first kind and order n . The spherical harmonics $Y_n^m(\theta, \phi)$ are defined as

$$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos\theta) e^{im\phi} \quad (22)$$

- 5 where $P_n^m(\cdot)$ are associated Legendre functions. The factor γ_n , having unitary norm, is given by

$$\gamma_n = i \frac{h_n^{(1)}(kR_\Lambda) j_n(kR_V)}{\left| h_n^{(1)}(kR_\Lambda) j_n(kR_V) \right|} \quad (23)$$

- and represents a phase shift applied to each spherical harmonic of order n due to the action of S . Equation (21) can be verified by substituting it into equation (12),
 10 (15) or (16) and applying the spherical harmonic expansion of the free field Green function (A1) together with the completeness and orthogonality relations of the spherical harmonics, equations (A5) and (A6) respectively.

- It can be noticed that the spherical harmonics $Y_n^m(\theta, \phi)$ have two indices, while the singular values σ_n have only one index. This is due to the degeneracy of the
 15 singular values, and it implies that one eigenspace of dimension $(2n+1)$ is associated with the singular values σ_n . Hence, for each order n , it is possible to generate a set of $(2n+1)$ orthogonal spherical harmonics which span that subspace. In other words, all the spherical harmonics of order n and degree m are associated with the same singular values σ_n . This degeneracy is typical of symmetrical
 20 geometries (such as the sphere), and arises in many other fields of physics (a well known example in quantum physics is the degeneracy of two electronic configurations, which have the same energy level).

As a result of the orthogonality of the spherical harmonics (A5), it is easy to verify the mutual orthogonality of the set of functions $p_n(\mathbf{x})$ and $a_n(\mathbf{y})$ defined by equation (21). Using the expansion of the free field Green function (A1), it is possible to show that the functions $a_n(\mathbf{y})$ and $p_n(\mathbf{x})$ and the singular values σ_n satisfy equations (12), (15) and (16).

The spherical wave spectrum $S_{nm}(r)$ of the target sound field $p(\mathbf{x})$, calculated at $r = R_V$, is defined as

$$S_{nm}(R_V) = \int_0^{2\pi} d\phi \int_0^\pi p(R_V, \theta_x, \phi_x) Y_n^m(\theta, \phi)^* \sin(\theta) d\theta \quad (24)$$

It is now possible to calculate analytically the density $a(\mathbf{y})$ as

$$a(\mathbf{y}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{Y_n^m(\theta_y, \phi_y)}{R_\Lambda^2 i k h_n^{(1)}(k R_\Lambda) j_n(k R_V)} S_{nm}(R_V) \quad (25)$$

It can be observed that the denominator of (22) equals zero for those wave numbers k_n such that $j_n(k_n R_V) = 0$. These wave numbers correspond to the Dirichlet eigenvalues introduced previously, and identify the frequencies of resonance of a spherical cavity with radius R_V and pressure release boundaries (sometimes called the Dirichlet sphere). Under these circumstances, it is not possible to calculate uniquely the density $a(\mathbf{y})$ with (25) because of the non uniqueness of the solution of the Dirichlet problem (6). Considering equation (16), it is easy to notice that the function $Y_n^m(\theta_x, \phi_x)$ belongs to the null-space of S^+ and therefore is not in the range of the single layer potential S . Once again, it is important to emphasize that this problem in the reconstruction is not due to the arrangement of the layer of secondary sources on $\partial\Lambda$, but it is due to the boundary ∂V where the target sound field has been defined. For that reason, the density $a(\mathbf{y})$ can be determined by changing the radius of the control volume V .

The determination of the spherical spectrum of the target sound field involves the calculation of the integral (24). For some special cases, this integral can be solved analytically. Considering the expansion of the free field Green function (A1) it is possible to derive the following relation for the spherical spectrum of an omnidirectional point source located at $\mathbf{z} > R_V$ (formally a monopole with volume velocity $q_{VOL} = -(ipck)^{-1}$):

$$S_{nm}^{ps}(R_V) = ikh_n^{(1)}(kr_z) j_n(kR_V) Y_n^m(\theta_z, \phi_z)^* \quad (26)$$

It is important to emphasize that the source location $\mathbf{z}$ should not be in V but can be within Λ . Considering the spherical harmonic summation formula (A3) and the trigonometric relations (A4), a simplified formula can be derived for the density for the reconstruction of the sound field due to a monopole source:

$$a^{ps}(\mathbf{y}) = \sum_{n=0}^{\infty} \frac{h_n^{(1)}(kr_z)}{R_{\Lambda}^2 h_n^{(1)}(kR_{\Lambda})} \frac{2n+1}{4\pi} P_n(\cos(\zeta)) \quad (27)$$

where $P_n(\cdot)$ is the Legendre polynomial of degree n and (see relation (A4))

$$\cos(\zeta) = \frac{\mathbf{y} \cdot \mathbf{z}}{R_{\Lambda} r_z} = \sin(\theta_y) \sin(\theta_z) \cos(\phi_y - \phi_z) + \cos(\theta_y) \cos(\theta_z) \quad (28)$$

and represents the cosine of the angle between the vectors $\mathbf{y}$ and $\mathbf{z}$. It can be noticed that the summation over the different degrees m has been reduced to the computation of a single Legendre polynomial.

The application of this solution to the design of the signal processing apparatus of the system requires the series (25) and (27) to be truncated to a given order N . Under this assumption, for high operating frequencies such that $kR_{\Lambda}, kr_z \gg N$, then the large argument limits or far field approximation (A7) of the spherical Hankel functions and the finite summation formula for Legendre polynomial (A2) can be used in (27), which can be rewritten as

$$a^{hf}(\mathbf{y}) = \frac{e^{ik(r_z - R_\Lambda)} (N+1) P_N(\cos(\zeta)) - P_{N+1}(\cos(\zeta))}{R_\Lambda r_z 4\pi} \frac{1}{1 - \cos(\zeta)} \quad (29)$$

This is a very powerful formula, which is extremely useful for the real time implementation of a panning function. In fact, the only frequency dependent part of (29) is the complex exponential $e^{ik(r_z - R_\Lambda)}$, and its Fourier transform corresponds in the time domain to a simple delay equal to the distance $r_z - R_\Lambda$. The denominator of (29) is a simple attenuation due to the distance of the point source and the term represented by the series of Legendre polynomials is a gain factor depending on the relative angle between $\mathbf{z}$ and $\mathbf{y}$.

In the very special case when the distance of the virtual source r_z equals the sphere radius R_Λ , equation (27) can be rewritten in the very simple formulation

$$a^{real}(\mathbf{y}) = \frac{(N+1) P_N(\cos(\zeta)) - P_{N+1}(\cos(\zeta))}{R_\Lambda^2 4\pi} \frac{1}{1 - \cos(\zeta)} \quad (30)$$

The continuous density function calculated with equations (19), (25), (27), (29) or (30) has to be transformed into a finite set of (possibly frequency dependent) coefficients corresponding to the each loudspeaker of the system. This can be done by applying a quadrature of the integral (7). The coefficient a_l corresponding to the l -th loudspeaker is therefore computed as

$$a_l = a(\mathbf{y}_l) \Delta S_l \quad (31)$$

where ΔS_l has the dimension of an area and depends on the loudspeaker arrangement. If these are arranged regularly, then

$$\Delta S_l = \frac{A_{\partial\Lambda}}{L} \quad (32)$$

where $A_{\partial\Lambda}$ is the area of the boundary $\partial\Lambda$ and L is the total number of loudspeakers of the system. If the boundary $\partial\Lambda$ is a sphere of radius R_Λ , then

$$\Delta S_l = \frac{4\pi R_\Lambda^2}{L} \quad (33)$$

Numerical solution of the integral equation

- 5 The second method to solve the integral equation (7) is numerical. As a first step, the boundaries $\partial\Lambda$ and ∂V must be sampled. The sampling scheme adopted for $\partial\Lambda$ is given by the loudspeaker array: the boundary $\partial\Lambda$ is divided into L surfaces, each of them corresponding to a loudspeaker. The surface ΔS_l corresponding to the l -th loudspeaker is the same as in equations (31), (32) or (33). The boundary
- 10 ∂V is divided into Q surfaces. The q -th surface is identified by its geometrical centre $\mathbf{x}_q$, hereafter called a sampling point. It is recommended that the number of sampling points is chosen to be such that $Q > L$. The sampling points should be chosen in such a way that the average distance δx between two neighboring points is constant or approximately constant. In order to avoid problems arising from
- 15 spatial aliasing, it is recommended that, for a given angular frequency ω ,

$$\delta x < \frac{c\pi}{\omega} \quad (34)$$

In what follows, this relation is referred to as the aliasing condition. As will be discussed later, it might be desirable to choose a control volume V which is a function of the frequency.

- 20 In order to avoid the problems arising from the Dirichlet eigenvalues explained above, it can be useful to add some additional sampling points in the interior of V , thus increasing the number Q of sampling points. In the case of ∂V being a sphere, it might be a wise choice to add an additional sampling point in the centre of the sphere, in order to avoid the problems arising from the first Dirichlet eigenvalue,
- 25 whose characteristic wave number is identified by the first zero of the spherical

Bessel function $j_0(kR_V)$. Another way of avoiding these problems is by choosing, for a given frequency $\hat{\omega}$, a control volume $V(\hat{\omega})$ for which the frequency considered does not correspond to one of its Dirichlet eigenvalues. Some possible strategies for choosing a frequency dependent control volume are described later.

- 5 The vector $\mathbf{p}$ is defined as the set of values of the target field $p(\mathbf{x})$ evaluated at the positions of the sampling points. In the case that the target field is due to an omnidirectional point source located at $\mathbf{z}$ in the free field, as in the case of equation (26), then the q -th element of the vector $\mathbf{p}$ is defined as

$$p_q = g(\mathbf{z} | \mathbf{x}_q) = \frac{e^{ikd_{qz}}}{4\pi d_{qz}} \quad (35)$$

- 10 where $d_{qz} = \|\mathbf{z} - \mathbf{x}_q\|$ is the distance between the q -th sampling point on ∂V and the virtual source, and it can depend on the frequency if the control volume V also depends on the frequency. The dimension of $\mathbf{p}$ is Q . The operator S can be transformed into matrix $\mathbf{S}$, which is defined as

$$S_{ql} = g(\mathbf{y}_l | \mathbf{x}_q) = \frac{e^{ikd_{ql}}}{4\pi d_{ql}} \quad (36)$$

- 15 where $d_{ql} = \|\mathbf{y}_l - \mathbf{x}_q\|$ is the distance between the q -th sampling point on ∂V and the position of the l -th loudspeaker. This distance can depend on the frequency if the control volume V also depends on the frequency. The dimension of $\mathbf{S}$ is therefore Q by L . The regularized pseudo inverse matrix $\mathbf{S}^+$ (not to be confused with the adjoint operator) is defined as

$$20 \quad \mathbf{S}^+ = (\mathbf{S}^H \mathbf{S} + \beta \mathbf{I})^{-1} \mathbf{S}^H \quad (37)$$

where β is a regularization parameter and $\mathbf{I}$ is the identity matrix of dimension L by L . The dimension of $\mathbf{S}^+$ is L by Q . It is important to emphasise that this matrix

depends only on the loudspeaker arrangement and on the sampling scheme on the boundary of the control volume, and does not depend on the position of the virtual source. It is now possible to compute the coefficient a_l corresponding to the l -th loudspeaker as

$$5 \quad a_l = \sum_{q=1}^Q S_{lq}^+ p_q \quad (38)$$

It is important to notice that both matrices $\mathbf{S}$ and $\mathbf{S}^+$ and the vector $\mathbf{p}$ depend on the wave vector k and hence on the operating frequency ω .

A major feature of the two methods which have been presented for solving the integral equation (7) and to derive the loudspeaker coefficients a_l , is that they give
10 identical results when the number of loudspeakers L tends to infinity.

It is now important to discuss how the control volume V can be modified depending on the operating frequency. For what has been said about the aliasing condition, it may seem to be wise to choose the control volume to be as small as possible. However, the study of the stability of the condition number of matrix $\mathbf{S}$
15 as a function of the operating frequency ω shows that if the control volume is too small, then the conditioning of the matrix $\mathbf{S}$ is poor.

In order to respect the sampling condition for all the considered frequencies and to have, at the same time, a well-conditioned matrix $\mathbf{S}$, it might be desirable to choose a frequency dependent control volume $V(\omega)$. Suppose that, for a given frequency
20 $\hat{\omega}$, a control volume $V(\hat{\omega})$ has been chosen which is a star-convex set and a set of sampling points $\mathbf{x}_l(\hat{\omega})$ have been chosen which respect the sampling condition, which grants good conditioning of the matrix $\mathbf{S}$ and $\hat{\omega}$ does not correspond to one of the Dirichlet eigenvalues for $V(\hat{\omega})$. In this case it is possible to define a frequency dependent control volume $V(\omega)$, which has the same shape of $V(\hat{\omega})$ and
25 which is identified by the sampling points

$$\mathbf{x}_l(\omega) = \frac{\omega}{\hat{\omega}} \mathbf{x}_l(\hat{\omega}) \quad (39)$$

Provided that $Q > L$, a suitable choice for a frequency dependent control volume $V(\omega)$ is a sphere centered in the origin with radius

$$R_V = \frac{c(\sqrt{L}-1)}{\omega} \quad (40)$$

- 5 The motivation of this choice is non trivial and is related to the behavior of the spherical Bessel functions. It has been shown that for spherical geometries the Bessel functions appear in the expression (21) of the singular values σ_n , and the radius R_V of the control volume appears in the argument of these functions. It can be inferred that, even in the discrete case, the L eigenvalues of the squared matrix
- 10 $\mathbf{S}^H \mathbf{S}$ are related to the spherical Bessel functions. The degeneracy of the eigenvalues σ_n in the case of spherical geometries has also been discussed above. This implies that if $\mathbf{S}^H \mathbf{S}$ has dimension L by L and always in the case of spherical geometry, then it has ideally $N = \sqrt{L} - 1$ independent eigenvalues. This approximation is exact when L tends to infinity and the boundaries ∂V and $\partial \Lambda$ are
- 15 sampled regularly. The condition number of $\mathbf{S}^H \mathbf{S}$ is related to the ratio between the largest and the smallest singular values, both related to the spherical Bessel functions. Using these arguments, it can be deduced that an optimal choice for the radius R_V is given by $R_V = N/k$, which leads to equation (40).

- In both cases of equation (39) and equation (40), the additional constraint must be
- 20 applied that all loudspeakers and the virtual source position must be located outside of the control volume.

- It is worth highlighting, once again, that the control volume does not correspond to the listening area, as an accurate reproduction of the target sound field can be achieved, because of analytical continuation, also in the exterior of the control
- 25 volume.

If the target sound field is due to a point source located at $\mathbf{z}$, it is useful to introduce the factor

$$\Delta\Phi_{lz} = e^{ik(r_l - r_z) + i\omega\delta t} \quad (41)$$

This represents a simple phase compensation factor, which compensates for the delay arising in the filter computation caused by the difference of the radial coordinate of the l -th loudspeaker r_l and of the virtual source r_z . The term δt is a small constant quantity, corresponding to a modeling delay, which has been introduced in order to guarantee the causality of the digital filters.

DESCRIPTION OF POSSIBLE IMPLEMENTATIONS

Figure 3 reports a diagram of the signal processing apparatus of the sound reproduction system. The apparatus comprises functional modules, called Single Input Processing Units (SIPU), Multiple Input Processing Units (MIPU) and Auralisation Processing Units (APU). They are the signal processing units corresponding to the three different operational modes of the system described above. It will be appreciated that the limit on the total number of SIPU, MIPU and APU modules of the system is not given *a priori* and depends on the computational power of the electronic hardware used for the implementation of the system. The different modules are composed by different components, described in detail in what follows. Digital filters are one of these components. They are defined in what follows with mathematical expressions, and from these expressions digital Finite Impulse Response filters (or possibly Infinite Impulse Response Filters) can be easily designed using standard techniques, such as for example the frequency sampling method. The use of Finite Impulse Response filters is suggested, as the filters described in what follows usually show a strong decay in the time domain. It is also relevant to highlight that most of the components of the different modules, such as digital delays and filters, are often described as separate elements for sake of clarity. However, it can be useful implementing a series of one or more of these elements as a single digital filter.

As an example, referring to Figure 4, it is possible to implement the Low Pass Filter (*LPF*) and each of the filters in the low frequencies bus (F_1, F_2 , etc.) as a single filter.

Single Input Processing Unit (SIPU)

- 5 Figure 4 shows a block diagram of a Single Input Processing Unit. This module is designed to generate the L loudspeaker signals which allow the reproduction of the sound field due to a virtual source at a given position in the free field, with a given radiation pattern and with a given orientation. The SIPU receives as an input a single audio signal and some additional data, describing the location (distance and
10 direction) of the virtual source, its radiation pattern and its orientation. With reference to Figure 1, the distance of the virtual source r_z is described by a scalar and positive number, while the direction of the virtual source is described by the two angles θ_z and ϕ_z . The orientation of the virtual source is described by two angles and its radiation pattern by a complex function of the frequency and of two
15 angles.

- The audio signal is first filtered by the digital filter F_H , which is defined depending on the orientation and distance of the virtual source. The signal is then divided into two busses, called a high frequency bus and a low frequency bus respectively. A high pass filter $HPF(\omega)$ and a low pass filter $LPF(\omega)$ are applied to the two
20 signals respectively. The two filters are such that

$$LPF(\omega) + HPF(\omega) = e^{i\omega\Delta t}$$

- This means that if one signal is filtered separately by the two filters and the two outputs are summed together, the result is a delayed version of the original signal. The cut-on frequency of the high pass filter (-6dB) and the cut-off frequency of the
25 low pass filter (-6dB) is the same and is called ω_c . The latter can be chosen to be the smallest frequency ω satisfying the condition

$$\frac{\left| h_N^{(1)}(r_{l\min} \omega/c) \right|}{\left| h_0^{(1)}(r_{l\min} \omega/c) \right|} \approx 1$$

where $r_{l\min}$ is the smallest of all loudspeaker radial co-ordinates r_l , that is to say the radial co-ordinate of the loudspeaker whose position is the closest to the origin.

The signal on the high frequency bus and the signal on the low frequency bus are
 5 processed in different ways, applying a set of operation called High Frequency Signal Processing and Low Frequency Signal Processing, respectively. They are described in detail below.

It should be noted that the existence of both the High Frequency Bus and the Low Frequency Bus is not essential. For example, a variant of the Single Input
 10 Processing Unit may comprise either the High Frequency Signal Processing or the Low Frequency Signal Processing. In these special cases, it is clear that the High Pass Filter $HPF(\omega)$ and the Low Pass Filter $LPF(\omega)$ would need to be removed.

The signal on the low frequency bus is filtered in parallel by a bank of L digital filters, labeled $F_1, F_2, \dots, F_L$ in Figure 4. Each filter corresponds to a different
 15 loudspeaker. These filters can be defined in two different ways, called numerical filter computation and analytical filter computation, both depending on the position of the virtual source and on the position of the loudspeaker considered.

1- For the numerical filter computation, a control volume V is chosen as explained above. Its geometrical centre coincides with the origin of the co-ordinate system.
 20 As described above, a set of Q regularly arranged sampling points is defined on the control surface. The q -th sampling point is identified by the vector $\mathbf{x}_q$. All loudspeakers and the location of the virtual source lie outside of the control volume. As has been discussed, the control volume and the sampling points can be chosen to be frequency dependent. A suitable choice for a frequency dependent
 25 control volume is a sphere with radius

$$R_V = \min \left[\frac{c(\sqrt{L}-1)}{\omega}, r_z, r_{l \min} \right]$$

As discussed above, it is beneficial to include some extra points in the interior of V in order to avoid problems arising from the Dirichlet eigenvalues. The frequency dependent vector $\mathbf{p}(\omega)$ is defined as in equation (35). The frequency dependent matrices $\mathbf{S}(\omega)$ and $\mathbf{S}^+(\omega)$ are defined as in equations (36) and (37). It is important to highlight that these matrices depend only on the loudspeaker arrangement and on the sampling scheme of the boundary of the reconstruction area, and do not depend on the position of the virtual source. For this reason, they can be computed off-line and not in real time. The digital filter corresponding to the l -th loudspeaker is computed from

$$F_l(\omega) = \Delta \Phi_{lz}(\omega) \sum_{q=1}^Q S^+_{lq}(\omega) p_q(\omega)$$

2- The digital filters in Figure 4, when they are computed following the analytical approach, have a frequency dependence of the form

$$F_l(\omega) = \Delta S_l(\omega) \Delta \Phi_{lz}(\omega) \sum_{n=1}^L \frac{a_n(\mathbf{y}_l, \omega)}{\sigma_n(\omega) + \beta(\omega)} \int_{\partial V(\omega)} g(\mathbf{z} | \mathbf{x}, \omega) p_n(\mathbf{x}, \omega) * dS(\mathbf{x}, \omega)$$

where the functions $a_n(\mathbf{y}, \omega)$ and $p_n(\mathbf{x}, \omega)$ and the singular values $\sigma_n(\omega)$, defined by equations (12) and (15), are computed analytically as described previously. $\partial V(\omega)$ is the boundary of the control volume and $\beta(\omega)$ is a regularization parameter; both can be chosen to be frequency dependent. It is recommended to choose the order N of truncation of the series to be equal to (or possibly smaller than) the number of loudspeakers L . In the case when the loudspeakers are regularly arranged on a sphere with radius R_Λ and the control volume is also a sphere with radius R_V , then following equation (27) the frequency response of the digital filter corresponding to the l -th loudspeaker is defined by

$$F_l(\omega) = \Delta \Phi_{lz}(\omega) \sum_{n=0}^M \frac{h_n^{(1)}(kr_z)}{h_n^{(1)}(kR_\Lambda)} \frac{2n+1}{L} P_n(\cos(\zeta))$$

where $\cos(\zeta)$ is defined by equation (28) and the order M of truncation of the series must be a natural number and is chosen to be

$$M \leq \sqrt{L} - 1$$

- 5 The signal on the high frequency bus is first delayed by an amount of time Δt that takes in consideration the length of the digital filters in the low frequencies bus and the quantity δt introduced in equation (41). The signal is then multiplied by a bank of parallel gains, each of them corresponding to a different loudspeaker. They are labeled $G_1, G_2, \dots, G_L$ in Figure 4. There are two possible ways of
- 10 defining these gains. The first method is derived from equation (29), corresponding to the high frequency approximation of equation (27). The gain corresponding to the l -th loudspeaker is defined as

$$G_l = \begin{cases} \frac{r_l(N+1)}{r_z} \frac{P_M(\cos(\zeta_l)) - P_{M+1}(\cos(\zeta_l))}{1 - \cos(\zeta_l)} \frac{\Delta S_l}{r_l^2} & \text{if } \zeta_l < \zeta_M \\ 0 & \text{if } \zeta_l \geq \zeta_M \end{cases}$$

- $\cos(\zeta)$ is defined by equation (28). M depends upon the average distance of the
- 15 neighboring loudspeakers and for a regular arrangement can be chose as $M \leq \sqrt{L} - 1$. In the case when the loudspeakers are regularly arranged on the surface of a sphere, then $\Delta S_l / r_l^2 = 1/L$. The angle ζ_M defines the semi-aperture of the main lobe of the function $(P_M(\cos(\zeta)) - P_{M+1}(\cos(\zeta))) / (1 - \cos(\zeta))$ and can be defined by the relation

$$20 \quad P_M(\cos(\zeta_M)) + P_{M+1}(\cos(\zeta_M)) = \min[P_M(\cos(\zeta)) + P_{M+1}(\cos(\zeta))]$$

This implies that only the loudspeakers which are closest to the location of the virtual source are active, and that they all operate in-phase.

The second method is derived from the assumption that the digital filters on the low frequency bus designed with the numerical approach and corresponding to the loudspeakers which are closer to the location of the virtual source show an asymptotic behavior at high frequencies. It is supposed that after the cut-off
 5 frequency ω_c the magnitude of these filters remains constant and the phase is very close to 0. The gain corresponding to the l -th loudspeaker is therefore defined as

$$G_l = \begin{cases} \left| \sum_{k=1}^K S_{lk}^+(\omega_c) p_k(\omega_c) \right| & \text{if } \zeta_l < \zeta_M \\ 0 & \text{if } \zeta_l \geq \zeta_M \end{cases}$$

where the angle ζ_M , the matrix $S^+(\omega)$ and the vector $\mathbf{p}(\omega)$ are defined as above.

Microphone array and Multiple Input Processing Unit (MIPU)

10 The Multiple Input Processing Unit is designed to generate the L loudspeaker signals which allow the reproduction of a sound field which has been captured using a specially designed array of microphones.

The array of microphones is designed in connection with the reproduction system, meaning that the microphone array is to be considered as a part of the whole
 15 system. The microphone array comprises a plurality of omnidirectional capsules regularly arranged on multiple surfaces. It will be appreciated that a microphone capsule relates to a portion where a microphone membrane is located. These surfaces define the boundaries of multiple, concentric control volumes. The choice of multiple control volumes arise from the fact that for a given number of
 20 sampling points the ideal size of the control volume, which respects the aliasing condition and allow a good conditioning of matrix $S(\omega)$, depends on the considered frequency. It is not practicable to choose a control volume which changes continuously as a function of the frequency. It is however possible to choose a finite number of control volumes $V_1, V_2 \dots V_F$, each of them dedicated to a
 25 given frequency range. A set of Q_f omnidirectional microphones are regularly

arranged on the boundary ∂V_f of the control volume V_f . The set of all the microphones arranged on the same control volume is referred to as a microphone layer. The total number of microphones Q is given by

$$Q = \sum_{f=1}^F Q_f$$

- 5 As explained above, problems can arise at those frequencies, which correspond to the Dirichlet eigenvalues of the control volume. The use of a multiple microphone layer can partially overcome this problem, but it can happen that one of the frequencies corresponding to one of the Dirichlet eigenvalues of the volume V_f belongs to the range of frequencies to which that control volume is dedicated.
- 10 This problem can be partially, when not totally, overcome by adding one or more microphones in the interior of the control volume. A wise choice for an additional microphone location is in the centre of the microphone array, especially when the control volumes are spherical. This is due to the fact that the first critical frequency for a given volume V_f is identified by the first zero of the spherical
- 15 Bessel function $j_0(R_f \omega / c)$. In physical terms this means that the microphone array can not detect the component of the sound field corresponding to the zero order spherical harmonic. The microphone in the centre of the array can, on the contrary, detect only that missing component, thus overcoming the problem. A higher number of additional microphones might be needed for higher critical
- 20 frequencies. It is also possible to use, as an additional sampling point for a given layer ∂V_f , one of the microphones arranged on a different layer $f' \neq f$.

A suitable choice for the different control volumes is given by a set of concentric spheres. As a guideline for the choice of the radius of these spheres, if the upper frequency for the operating frequency range of the control volume V_f is ω_f , the

25 radius R_f of that control volume can be chosen to be

$$R_f = \frac{c(\sqrt{\min_{LQ_f}} - 1)}{\omega_f}$$

where $\min_{LQ_f}$ is the smallest number between the number of loudspeakers L and the number of Q_f of microphones on that layer. This guideline arises from considerations of the spherical Bessel functions and on the conditioning of matrix $\mathbf{S}^H \mathbf{S}$, which have been discussed above (refer to equation (40)). This approach suggests that it is also possible to split a given frequency range corresponding to a control volume V_f with Q_f microphones into two sub-ranges by simply reducing the number of microphones used in the processing of the lower frequency range. This can be especially useful at very low frequencies, where the choice of radius discussed above would result in a very large value. As an example, consider a system composed by forty loudspeakers and a layer of thirty six microphones regularly arranged on the sphere ∂V_f dedicated to the audio frequencies below 500Hz. Following the above guideline, the radius R_f is approximately 0.5m. It is possible to choose a subset of eight microphones on that layer and define an additional frequency range with higher limit of approximately 200Hz.

A microphone array with just one layer can be considered as a special case of the microphone array described. Another variant is constituted by a microphone array having a scattering object (as for example a rigid sphere) in the region of the space contained inside the smallest control volume. The filter computation described in what follows remains the same. It is also straightforward to perform the analytical calculation of the digital filters described later for the case corresponding to a set of microphones arranged around or on the surface of a rigid sphere.

The output signals from the microphone array are processed by the Multiple Input Processing Unit, represented by Figure 5. This figure illustrates the example corresponding to a microphone array with only two layers, but this approach can be identically extended to configurations with more microphone layers. The output signals of each layer are processed separately, as shown in the figure. As a

first step, each signal is filtered through a Band Pass Filter ($BPF_f(\omega)$) whose cut-on and cut-off frequencies are defined by the frequency range corresponding to that microphone layer. For the layer f , the cut-on frequency is ω_{f-1} and the cut-off frequency is ω_f . These band pass filters must respect the condition

$$5 \quad \sum_{f=1}^F BPF_f(\omega) = e^{i\omega\Delta t}$$

analogous to the Low and High Pass Filters in the SIPU. The signals are then process by a matrix of digital filters, labeled $F_{1,1,1}$, $F_{L,1,1}$, ..., $F_{L,Q11}$, in Figure 5.

These filters can be defined in three different ways: analytically, numerically or with measurements. All these formulation are grounded on the same theoretical

10 background, which has been discussed above.

1-Using a purely numerical approach, the filters corresponds to the elements of the frequency dependent matrix $S^+(\omega)$ defined by equation (37). The filter corresponding to the l -th loudspeaker and the q -th microphone on the layer f is computed from

$$15 \quad F_{l,q,f}(\omega) = e^{i\omega\delta t} S_{lq}^+(\omega)$$

where the small δt has been introduced in order to ensure that the filter is causal (when this is needed).

2.-The filters can be also calculated after having measured, possibly in an anechoic environment, the impulse response between each loudspeaker and each
20 microphone on the given layer. The measurement can be carried out using standard techniques (swept sine, MLS, etc.) and is subject to the well-known sources of error that affect these kinds of measurements. The microphone array must be arranged in such a way that its geometrical centre corresponds to the origin of the co-ordinate system. It is preferable to exclude reflections generated

by the surrounding environment in the measurement. This can be done by carrying out the measurements in an anechoic environment or by windowing the measured impulse response in order to take into account only the initial part of the measured signal. The acquired impulse responses need to be transformed in the frequency domain by applying a Fourier transform. The set of acquired measurements constitutes the matrix $\mathbf{H}(\omega)$. Its element $H_{qlf}(\omega)$ represents the transfer function between the l -th loudspeaker and the q -th microphone on the layer f . Following a procedure analogous to equation (37), matrix $\mathbf{H}^+(\omega)$ is computed from

$$\mathbf{H}^+(\omega) = (\mathbf{H}^H(\omega)\mathbf{H}(\omega) + \beta(\omega)\mathbf{I})^{-1}\mathbf{H}^H(\omega)$$

where the elements $\beta(\omega)$ and $\mathbf{I}$ are defined as for equation (37). It is very important to apply a regularization scheme to the inversion of matrix $\mathbf{H}^H(\omega)\mathbf{H}(\omega)$, as the presence of measurement errors can result in the computation of unstable filters. In the proposed approach, Tikhonov regularization with frequency dependent regularization parameter $\beta(\omega)$ is suggested. The filter corresponding to the l -th loudspeaker and the q -th microphone on the layer f is computed from

$$F_{l,q,f}(\omega) = e^{i\omega\delta_l} H_{lqf}^+(\omega)$$

where δ_l is as defined above.

3- An analytical computation method for the filters can be derived from discretising, in equation (19), the integral $\langle p_n | p \rangle$ defined by equation (8). Equation

$$(19) \text{ can therefore be reformulated as } a_f(\mathbf{y}) = \sum_{q=1}^Q p(\mathbf{x}_q) \Delta S'_q \sum_{n=1}^M \frac{1}{\sigma_{n,f}} p_{n,f}(\mathbf{x}_q) a_{n,f}(\mathbf{y})$$

where $\Delta S'_q$, analogous to the coefficient where ΔS_l described above, has the dimension of an area and depends on the microphone arrangement on the given layer. The order of truncation $M = \min_{LQf}$ is the smallest number between the

number of loudspeakers L and the number of Q_f of microphones on the layer f . The subscript index $[\cdot]_f$ is due to the relevant fact that, in general cases, the eigenfunctions and eigenvalues $p_{n,f}(\mathbf{x})$, $a_{n,f}(\mathbf{y})$ and $\sigma_{n,f}$ can be different for different layers. Considering finally equation (31) the frequency response of the filter corresponding to the l -th loudspeaker and the q -th microphone on the layer considered is therefore defined by

$$F_{l,q,f}(\omega) = e^{i\omega\delta t} \Delta S_l \Delta S_q \sum_{n=1}^M \frac{p_{n,f}(\mathbf{x}_q, \omega) a_{n,f}(\mathbf{y}_l, \omega)}{\sigma_{n,f}(\omega) + \beta(\omega)}$$

In the special case when both the loudspeakers and the microphones on the considered layer are regularly arranged on two spheres of radius R_Λ and R_f , respectively, the filter is computed from

$$F_{l,q,f}(\omega) = e^{i\omega\delta t} \frac{4\pi}{LQ_f} \sum_{n=1}^{M'} \frac{(2n+1)P_n(\cos(\zeta_{lq}))}{jkj_n(kR_f)h_n^{(1)}(kR_\Lambda) + \beta(\omega)}$$

where equations (21), (33), (A2) and (A3) have been used and $\cos(\zeta_{lq})$ is the cosine of the angle between the vectors identifying the locations of the microphone and of the loudspeaker considered (refer to equation (28)). The order of truncation M' is chosen to be

$$M' \leq \sqrt{\min_{LQ_f}} - 1$$

The outputs of the digital filters are finally combined as shown in Figure 5.

Auralisation Processing Unit (APU)

Figure 6 reports a block diagram of the Auralisation Processing Unit. This module is designed to generate the L loudspeaker signals which allow the reproduction of the sound field due to a point source located at a given position in a given reverberant environment, such as a concert hall or an enclosure. The digital filters

labeled $G_1, G_2, \dots, G_L$ in Figure 6 are computed by combining the filters of the Multiple Input Processing Unit with a set of impulse responses describing the reverberant field of the reverberant environment considered. These impulse responses are the impulse responses measured between a reference source (for example a loudspeaker or an omnidirectional source) and the microphone array 100 described above. Both the reference source and the microphone array are located in the reverberant environment considered and the measurements can be carried out using one of the conventional standard techniques mentioned above. The set of Q measured impulse responses are transformed in the frequency domain by applying a Fourier transform, and are labeled $R_1, R_2, \dots, R_Q$. The filter corresponding to the l -th loudspeaker is computed from

$$G_l(\omega) = \sum_{q=1}^Q R_q(\omega) F_{l,q,f}(\omega) BPF_{f(q)}(\omega)$$

where the filters $F_{l,q,f}(\omega)$ are defined in the same way as for the MIPU. The Band Pass Filter $BPF_{f(q)}(\omega)$ depends on the layer on which the q -th microphone is arranged.

When designing a Finite Impulse Response filter from the formulation of $G_l(\omega)$ given above, it is important to consider that while the filters $F_{l,q,f}(\omega)$ and $BPF_{f(q)}(\omega)$ are in general short in the time domain, the impulse responses $R_q(\omega)$ are in general very long, their length depending on the reverberation time of the measured reverberant environment. This factor is vital when defining the filter length, which must be the same if the filters are defined in the frequency domain. In order to avoid this difficulty it is also possible to define the filters in the time domain as

$$G_l(t) = \sum_{q=1}^Q \mathfrak{S}^{-1}[R_q(\omega)] \otimes \mathfrak{S}^{-1}[F_{l,q}(\omega) BPF_{f(q)}(\omega)]$$

where the operator $\mathfrak{F}^{-1}[\cdot]$ represents the inverse Fourier transform and the symbol $\otimes$ represents a convolution in the time domain.

The Auralisation Processing Unit shares some strong conceptual similarities with the Multiple Input Processing Unit, but while the input to the latter is a stream of
5 Q audio channels which are processed by a matrix of Q by L by F digital filters, the input to the APU is a single audio signal, processed by a bank of L filters. The latter are computed from set of measurements, but their computation can be made off-line. This implies that the real time implementation of an MIPU is much more computationally expensive than that of an APU.

APPENDIX

Spherical harmonic expansion of the free field Green function

$$\frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|} = \sum_{n=0}^{\infty} ikh_n^{(1)}(k|\mathbf{y}|)j_n(k|\mathbf{x}|) \sum_{m=-n}^n Y_n^m(\theta_x, \phi_x) Y_n^m(\theta_y, \phi_y)^* \quad (\text{A1})$$

Finite summation of Legendre polynomials

$$5 \quad \sum_{n=0}^N (2n+1)P_n(x) = (N+1) \frac{P_N(x) - P_{N+1}(x)}{1-x} \quad (\text{A2})$$

Summation formula for the spherical harmonics

$$\sum_{m=-n}^n Y_n^m(\theta, \phi) Y_n^m(\theta', \phi')^* = \frac{2n+1}{4\pi} P_n(\cos(\zeta)) \quad (\text{A3})$$

where $P_n(\cdot)$ is the Legendre polynomial of degree n and ζ is the angle between the
10 directions identified by θ, ϕ and θ', ϕ' . It holds that

$$\begin{aligned} \cos(\zeta) &= \cos(\phi)\sin(\theta)\cos(\phi')\sin(\theta') + \sin(\phi)\sin(\theta)\sin(\phi')\sin(\theta') + \cos(\theta)\cos(\theta') = \\ &= \sin(\theta)\sin(\theta')\cos(\phi - \phi') + \cos(\theta)\cos(\theta') \end{aligned} \quad (\text{A4})$$

Orthogonality of the spherical harmonics

$$\int_0^{2\pi} d\phi \int_0^\pi Y_n^m(\theta, \phi) Y_n^{m'}(\theta, \phi)^* \sin(\theta) d\theta = \delta_{nn'} \delta_{mm'} \quad (\text{A5})$$

Completeness relation of the spherical harmonics

$$15 \quad \sum_{n=0}^{\infty} \sum_{m=-n}^n Y_n^m(\theta, \phi) Y_n^m(\theta', \phi')^* = \delta(\phi - \phi') \delta(\cos(\theta) - \cos(\theta')) \quad (\text{A6})$$

Large argument approximation of spherical Hankel functions ($x \rightarrow \infty$)

$$h_n^{(1)}(x) \approx (-i)^{n+1} \frac{e^{ix}}{x} \quad (\text{A7})$$

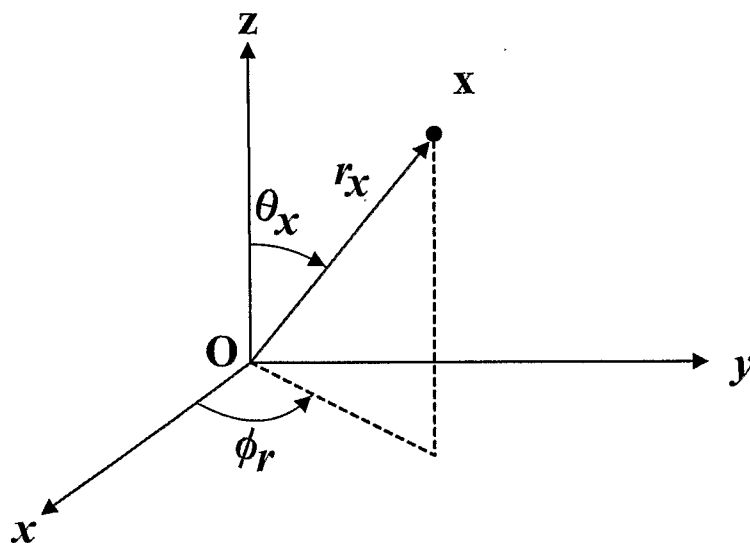
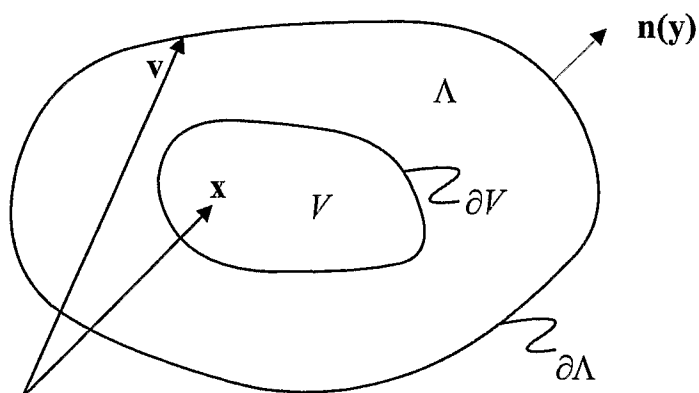
CLAIMS

1. Method of determining control signal data for an array of loudspeakers, the control signal data being such as to control the loudspeakers to produce a desired sound field associated with an audio signal, the method comprises determining
5 control signal data for different frequency components of the desired sound field in respect of respective different positions in a listening volume of the loudspeaker array, wherein determination of the control signal data comprises sampling the desired sound field at the surface of a control volume (V).
2. Method as claimed in claim 1 in which the size of the control volume is
10 dependent on the frequency component being sampled such that a larger control volume (V) is used for lower frequency components and a smaller volume is used for higher frequency components.
3. Method as claimed in claim 2 in which different sizes of control volume (V) are used for different frequency bands.
- 15 4. Method as claimed in claim 2 in which different sizes of control volume (V) of substantially the same shape.
5. Method as claimed in any preceding claim in which substantially the same number of sampling points is used for different sizes of the control volume (V).
6. Method as claimed in any preceding claim in which the average distance
20 between sampling points is less than half the wavelength considered.
7. Method as claimed in any preceding claim comprising sampling the desired sound field internally of the control volume (V).
8. Method as claimed in claim 7 in which sampling internally of the control volume (V) is effected at those frequencies which correspond to Dirichlet
25 eigenvalues of the control volume.

9. Method as claimed in claim 7 or 8 in which a sample internally of the control volume (V) is taken at the geometric centre of the control volume.
10. Method as claimed in any preceding claim in which the control signal data are such that only the or each loudspeaker of the loudspeaker array closest to a
5 virtual source generated by the loudspeaker array is/are activated.
11. Method as claimed in claim 10 which is used for high frequency component signal processing.
12. Method as claimed in any preceding claim in which in addition to the input audio signal, data indicative of location, radiation pattern and orientation of the
10 virtual source are used to determine the control signal data.
13. Method as claimed in any of claims 1 to 12 in which the audio signal comprises a signal representative of the sound of an acoustic source and a signal representative of characteristic data of an acoustic environment.
14. Method as claimed in claim 13 in which the sound of the acoustic source and
15 the data of an acoustic environment are included in a signal obtained by sampling the acoustic source in the acoustic environment.
15. Method as claimed in claim 13 in which the sound of the acoustic source and the data of the acoustic environment are separate signals.
16. Sound reproduction apparatus for processing an audio signal, the apparatus
20 configured to output control signal data for an array of loudspeakers to produce a desired sound field associated with the audio signal, wherein the apparatus configured to determine the control signal data for different frequency components of the desired sound field in respect of respective different positions in a listening volume of the loudspeaker array, wherein determination of the control signal data
25 comprises sampling the desired sound field at the surface of a control volume (V).

17. Apparatus as claimed in claim 16 in which the size of the control volume (V) is dependent on the frequency component being sampled such that a larger control volume is used for lower frequency components and a smaller volume is used for higher frequency components.
- 5 18. Apparatus as claimed in claim 16 or claim 17 in which the apparatus comprises a signal processor comprising one or more impulse response filters arranged to process the audio signal and output the control signal data.
19. Apparatus as claimed in any of claims 16, 17 or 18 which comprises a microphone array (100) for obtaining the audio signal, the array comprising a
10 microphone (140) at each of a different distance from the centre of the array.

1/6

*Fig. 1**Fig. 2*

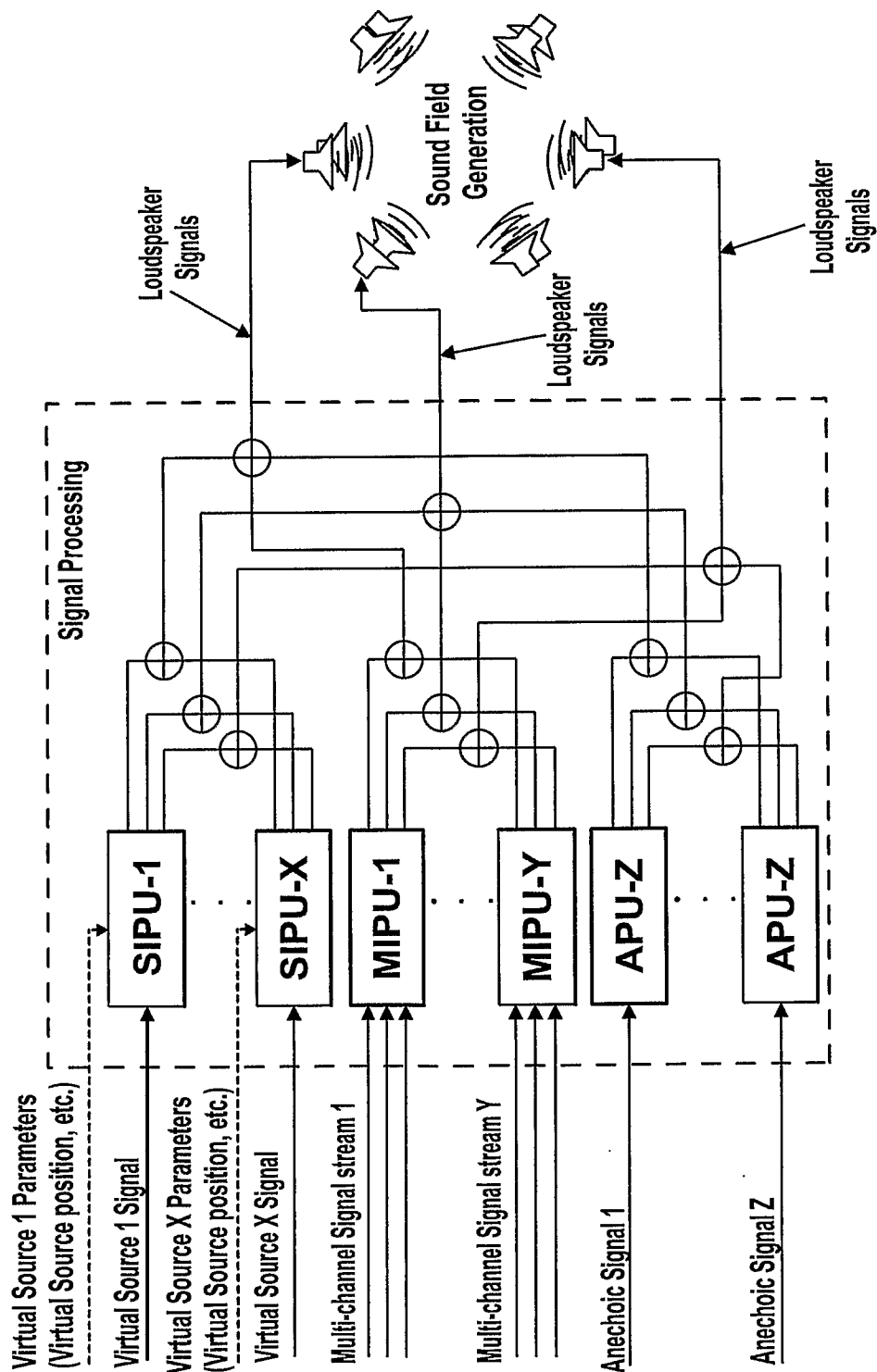


Fig. 3

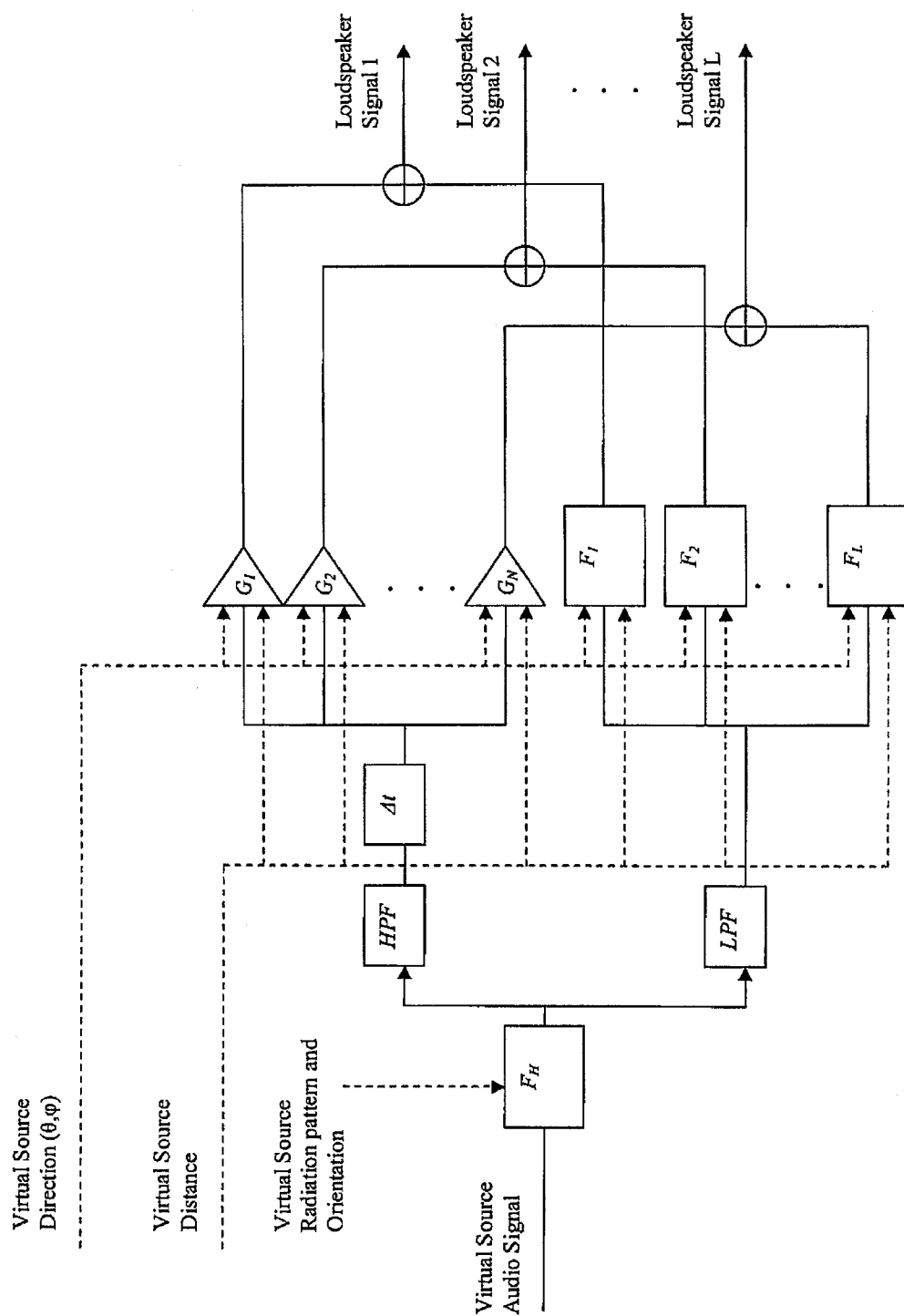


FIGURE 4

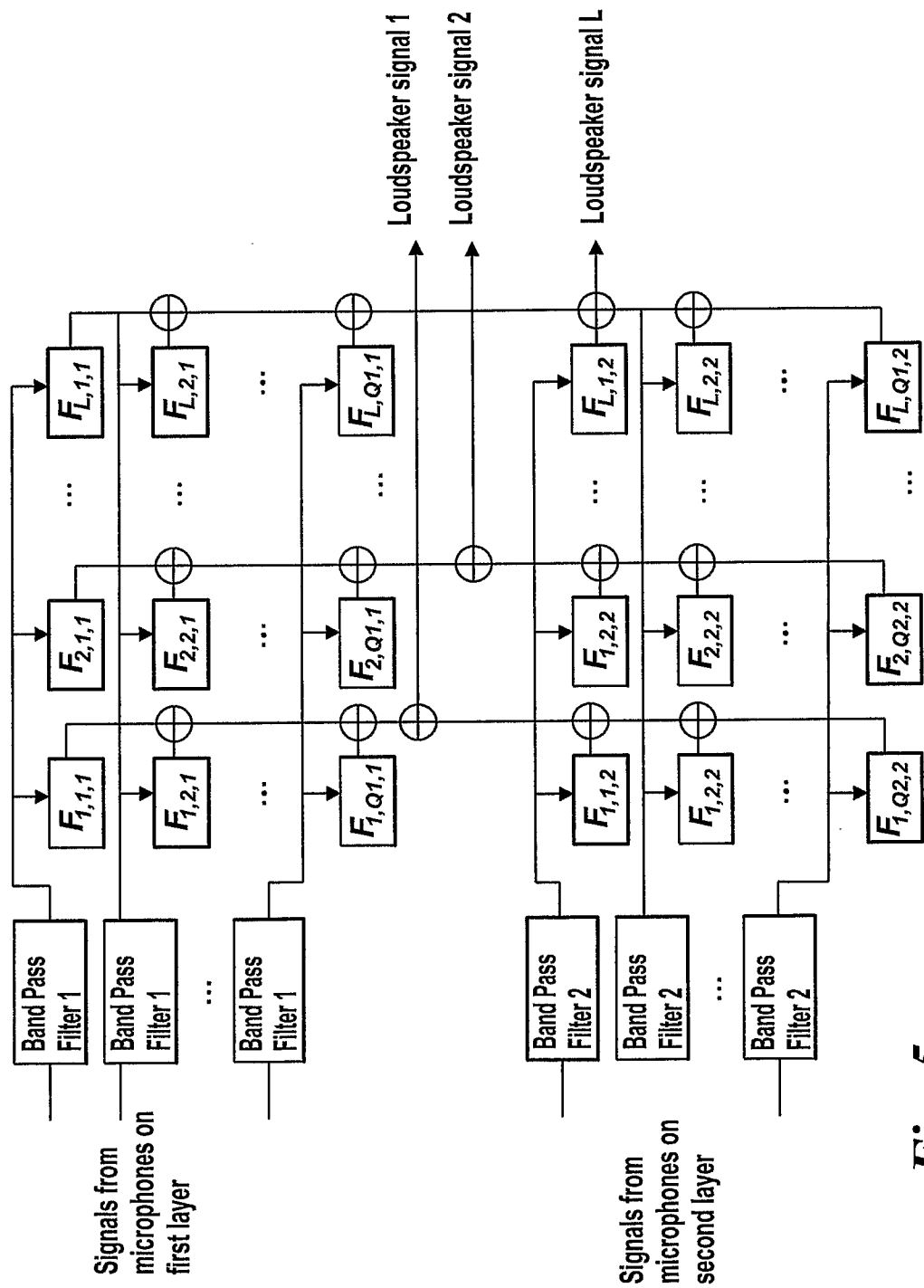


Fig. 5

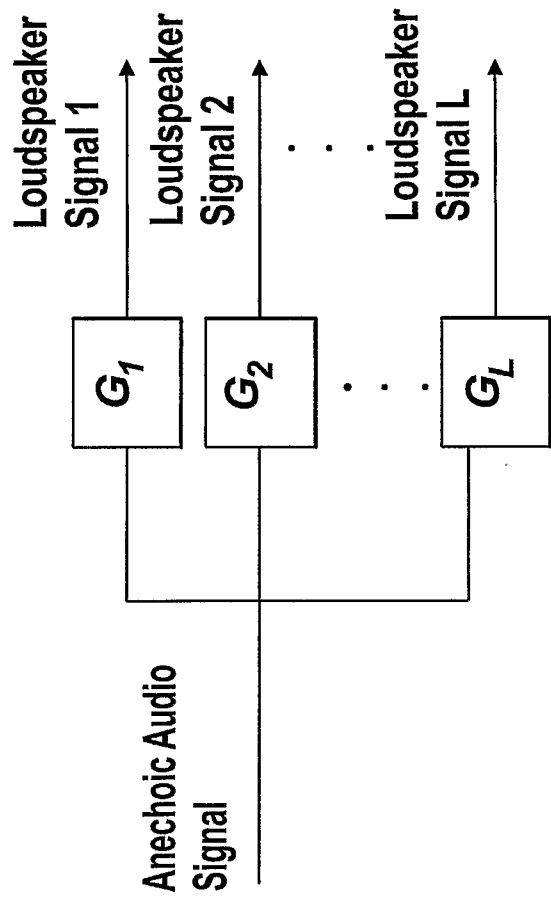


Fig. 6

