24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

MICROPHONE SETS, PANNING FUNCTIONS AND CHANNEL OBJECTS

Dylan Menzies, Filippo Maria Fazi
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, UK

A common approach for sound field reproduction is to encode the spatial audio scene with a fixed set of channel signals using panning functions or microphones. The channels are then used to drive loudspeakers either directly or via a linear decoding process. A combination of physical and psychoacoustic effects enable the approximate reproduction of the spatial percepts of the original scene. Older examples include Stereo and Ambisonics. With the arrival of object-based audio systems the problem has naturally arisen of how to include channel based material. Here we begin by presenting a framework for discussing channel encodings. This is applied to the general problem of converting between channel encodings, and specifically the decoding of channel based objects to arbitrary arrays. We then address the problem of representing a channel encoding using fewer channels, in order to reduce storage or transmission cost, and use weights to focus the encoding effort.

Keywords: spatial audio, 3D audio, ambisonics, loudspeaker array

1. Introduction

Channel-based audio reproduction systems use a fixed number of signals to encode a spatial sound scene. An encoding is a convenient representation of the scene. The encoding may be decoded to produce signals that are fed to an array of loudspeakers, in order to reproduce a recorded or simulated scene. In some cases the decoding process is trivial, the decoded signals are the same as the encoded signals.

A common type of encoding consists of microphone signals, or composed by applying panning functions to source signals. A microphone is described by a directivity function, which provides a transfer function from each incident plane wave signal to the microphone output. A panning function produces loudspeaker gains depending on the intended image direction, like the microphone directivity function. The gains are applied to the image signal. A scene may be composed by summing multiple such image signals.

For a given array, a set of microphones, decoding functions or panning functions are designed so that the original audio scene is reproduced faithfully, which is to say perceived images match the original recorded sources, or intended images. If a set of panning functions were replaced by microphones with the same directivities then the array would reproduce the sound field captured by the microphones. In this case the microphone decoding is trivial. Precise definitions for sound fields, microphone directivities and panning functions are given in Section 2.

The stereo system [1, 2] provides the simplest and oldest example of channel reproduction: Two channels are derived from a crossed pair of cardioid microphones, or else using a mixer with a panning function applied to each mono source signal. In reproduction these channels are fed directly to loudspeakers.

In the Ambisonic system, the channels are generated either from microphone arrays or with multichannel panning functions, using spherical harmonic directivity functions, [3, 4]. A variety of Ambisonic decoders have been designed according to different criteria [5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Each of these can be defined using loudspeaker panning functions. Channel-based encoding enables the compact representation of a sound field, which may include reverberation and complex objects. However it is not possible to directly manipulate the component sounds in a channel-encoding independently from one another.

Vector base amplitude panning (VBAP) [15] is a widely used method for panning over a 3-dimensional loudspeaker array. VBAP is not usually considered in the context of channel-based reproduction because the loudspeaker feeds are usually not transmitted, but rather generated at the point of reproduction. However VBAP is relevant for this article because it can be defined using a panning function for each loudspeaker [11].

In contrast with channel encoding, an *object-based audio* encoding consists of a variable number of audio objects, each representing a single source of some kind. Each object includes signals and other metadata information, for example about source position and size. This permits a great deal of freedom in manipulating the audio objects independently from one another, right up to the point of reproduction. Object encoding has been used for many years within the computer game industry, and has more recently been developed in standards for cinema and interactive broadcasting [16]. As transmission bandwidth increases and reproduction hardware becomes more sophisticated, object-based audio has become more attractive. In the context of spatial audio the possibility exists for optimising the reproduction according to each user's reproduction system and room. A channel encoding can be useful as an audio object because it can be used to efficiently capture an element of the sound field that is spatially complex, and for which detailed internal manipulation is not required. Already channel beds are used to provide a whole background scene. Here we consider more general channel objects that cover some part of the whole scene. Mixing of stereo signals to multichannel formats has been common practice in the recording industry, although there has been little research about this.

This investigation is motivated by the desire to better understand channel-based encodings, the information they contain and how they can be used, particularly within the context of object-based audio. In Section 2 we begin by describing a framework for the representation of the sound field and microphone directivity functions. This lays the foundation for the following work. In Section 3 this is applied to the conversion from one set of channel signals to another, in particular the case where the derived *decoded* signals directly drive a loudspeaker array reproduction. Comparison is made with some existing decoding schemes. The source signal set is considered as a channel object within an object encoding, and some examples are given for illustration. Finally in Section 4 a method is presented for reducing the channel count of encodings while retaining as much spatial accuracy as possible. This may be useful for compressing broadcast signals, or downmixing high resolution recordings. Weightings are introduced in order to emphasise the importance either of certain microphones or directions. Given the limited space here, many details are omitted. A complete account with examples will appear in a forthcoming extended article.

1.1 Notation

Signals and filters are represented in the frequency domain. For simplicity the frequency dependence is usually omitted, and frequency dependent directivity functions are not considered. Vectors and matrices are in bold type. Either may also be represented in component form with normal type, for example the element of matrix $\boldsymbol{A}$ in the i-th row and j-th column is written A_{ij}. j is also occasionally used for $\sqrt{-1}$, but not simultaneously as an index, so its meaning is always clear. A hat is used to denote a spatial vector of unit length, for example $\hat{\boldsymbol{x}} = \boldsymbol{x}/|\boldsymbol{x}|$. Operators or matrices, vector spaces, and functions are all capitalised. The complex conjugate is represented with a bar, for example $\bar{p}$. The transpose conjugate is written $\boldsymbol{A}^H$. A dual basis or space is represented with an asterix, like S^*.

2. Representing Sound Fields And Microphones

2.1 Sound Fields

There are several ways to represent a region of sound field in a 2D or 3D that is free of sources, and which so satisfies the homogeneous Helmholtz equation. The Herglotz expansion (HE), is built from a continuous set of plane wave basis functions. The pressure field as a function of position $\boldsymbol{x}$ and wave number k is

$$p(\boldsymbol{x}, k) = \int_{\hat{\boldsymbol{k}} \in \Omega} e^{-j\boldsymbol{k}\cdot\boldsymbol{x}} s(\boldsymbol{k}) d\Omega \tag{1}$$

where the integration variable is $\hat{\boldsymbol{k}}$ ranging over a surface Ω of radius 1, a circle in 2 spatial dimensions and a sphere in 3 dimensions. The positive frequency convention, with time dependence $e^{j\omega t}$ is used here for wave representation. The wave vector of each plane wave component is $\boldsymbol{k} = k\hat{\boldsymbol{k}}$, where the direction of travel of the wave is $\hat{\boldsymbol{k}}$. The Herglotz density function $s(\boldsymbol{k})$ contains the information that uniquely represents and encodes the sound field, and can be thought of as a signal density function for the direction $\hat{\boldsymbol{k}}$ and wave number k.

For numerical calculation the HE integral cannot be used directly. It can be approximated by sampling over a set of uniformly distributed directions, $\{\hat{\boldsymbol{k}}_i\}$. The encoding is then represented by a function $s(\boldsymbol{k}_i) = s(\hat{\boldsymbol{k}}_i, k)$. For brevity we hide the frequency dependence from k and write the encoding as a vector $\boldsymbol{s}$ with components $s_i = s(\hat{\boldsymbol{k}}_i) = s(\hat{\boldsymbol{k}}_i, k)$. $\boldsymbol{s}$ and $s(\boldsymbol{k})$ will be referred to *as the sound field*, since it contains the information content of the field, although the actual pressure field is given by (1). We also refer to the *space of sound fields* S, which is the vector space of all possible sound fields $\boldsymbol{s} \in S$. The continuous encoding $s(\boldsymbol{k})$ exists in an infinite dimensional space, however in this article we stick to the finite dimensional case and notation. The sound field pressure is now a sum

$$p(\boldsymbol{x}) = \sum_i e^{-j\boldsymbol{k}_i \cdot \boldsymbol{x}} s(\boldsymbol{k}_i) \Delta\Omega_i \ . \tag{2}$$

In the 2D case the directions $\{\hat{\boldsymbol{k}}_i\}$ can be spaced equally, with uniform $\Delta\Omega_j$. In 3D the optimal choice of $\{\hat{\boldsymbol{k}}_i\}$ and $\{\Omega_j\}$ is not trivial in general. There exist direction sets for which uniform $\{\Omega_j\}$ gives exact reconstruction for sound fields of finite polynomial complexity, the *spherical t-designs* [17, 14]. Generally for large uniform direction sets the variation in optimal $\Delta\Omega_j$ becomes small.

The 2D or 3D sourceless sound field region may also be expanded in terms of a countable set of localised regular harmonic basis functions $\{R_i\}$,

$$p(\boldsymbol{x}, k) = \sum_{i=1}^{\infty} b_i(k) R_i(\boldsymbol{x}, k) \ , \tag{3}$$

where $b_i(k)$ are the coefficients or signals encoding the sound field. The basis functions can be made orthonormal,

$$R_i \cdot R_j = \delta_{ij} \ , \tag{4}$$

with the dot product defined as an integral over space,

$$p_1 \cdot p_2 = C \int_{\boldsymbol{x} \in V} \bar{p}_1(\boldsymbol{x}) p_2(\boldsymbol{x}) dV \ , \tag{5}$$

for a normalisation constant C. V can be either the whole of 2D or 3D space.

The Fourier Bessel Expansion (FBE) provides an example in 3D, and has several variations including the *N3D* form [18], which has real valued functions. The basis functions are each the product of a spherical harmonic function of direction and spherical Bessel function of distance. Analogous functions exist in 2D based on sinusoidal functions of azimuth and Bessel functions, a common form is the *N2D* basis, [18]. The term FBE will be used to refer expansions of this type in either 3D or 2D.

An FBE can be approximated by truncation,

$$p(\boldsymbol{x}, k) \approx \sum_{i=1}^{N} b_i(k) R_i(\boldsymbol{x}, k) \ , \tag{6}$$

For typical sound fields the approximation is very good within a radius r depending on N and k. This property has been used in the Ambisonic reproduction method for sound field encoding. Variable truncation allows for variable resolution and compatibility of encodings with different resolution. The truncated encoding coefficients form a finite vector, $\boldsymbol{b} = \{b_i\}$, which is a representation of the sound field, like $\boldsymbol{s}$, but with a different basis.

The HE is a natural representation for the sound field in the sense that the signal content in each direction is represented directly. In the FBE the signal components are complex linear combinations of the plane wave signal components. However there are fundamental and practical reasons, not elaborated here, why the FBE is sometimes preferable. In the next section the directivity of microphones is represented using the natural counterparts to these sound field descriptions.

2.2 Directivity Functions

An ideal microphone is characterised by its directivity function. Specifically, the output q of a perfectly linear microphone is the bilinear function of a complex-valued directivity function (DF) $Q(\hat{\boldsymbol{k}})$, and the sound field encoding $s(\hat{\boldsymbol{k}})$,

$$q = \int_{\hat{\boldsymbol{k}}\in\Omega} Q(\hat{\boldsymbol{k}})s(\hat{\boldsymbol{k}})d\Omega \tag{7}$$

The normalised wavevector $\hat{\boldsymbol{k}}$ provides the direction of travel of the wave. Note that microphone directivity is often stated as a function of the reverse direction $\hat{\boldsymbol{\theta}} = -\hat{\boldsymbol{k}}$, towards the incoming wave. In the following formula the wave directions are indexed, so we don't have to choose between variables, and no confusion arises.

A panning function provides a loudspeaker gain as a function of the desired image direction. A common goal is to find a set of panning functions for a loudspeaker array such that the output gains produce a perceived image close to the desired image, for a range of desired images. A complex scene can be produced by separately panning multiple signals and summing. If a set of such panning functions were replaced by microphones with the same directivities then the array would reproduce the captured sound field, since each plane wave component is effectively panned by each microphone. A panning function can be viewed as a form of virtual microphone directivity. DF will be used to refer to both microphones and panning functions.

Using a discretised sound field with components $s_j = s(\hat{\boldsymbol{k}}_j)$ and discretised DF with components $Q_j = Q(\hat{\boldsymbol{k}}_j)$ a microphone signal given by (7) can be approximated as

$$q \approx \sum_{j=1}^{N} Q_j s_j \Delta\Omega_j \tag{8}$$

A set of DFs, representing a set of microphones or panning functions, can then be written as a matrix $\boldsymbol{Q}$ with components Q_{ij} where i indexes the DFs, and j indexes the plane wave directions. $\boldsymbol{Q}_i$ will denote the ith DF as a vector. For the sake of convenience we redefine s_j by absorbing the product with $\Delta\Omega_j$ into it. This keeps the signals and DFs, which are of most interest, constant whatever discretisation is used. Each DF $\boldsymbol{Q}_i$ produces a signal $\boldsymbol{q}_i$ given by (8). The set of signals can be written as a vector $\boldsymbol{q}$, and the corresponding set of equations is

$$\boldsymbol{q} = \boldsymbol{Q}\boldsymbol{s} \tag{9}$$

$\boldsymbol{q}$ can be viewed as a *lossy encoding* of the sound field represented by $\boldsymbol{s}$. To represent the sound field accurately the number of elements in $\boldsymbol{s}$ is much greater than the number of elements in $\boldsymbol{q}$. $\boldsymbol{q}$ then contains less information than $\boldsymbol{s}$ and the original sound field, and $\boldsymbol{Q}$ has *full row rank*.

$\boldsymbol{Q}$ could be defined instead by substituting its transpose $\boldsymbol{Q}^T$ or transpose conjugate $\boldsymbol{Q}^H$. This would perhaps be a more symmetric presentation since DFs are then column vectors like $\boldsymbol{s}$. The choice does not affect the discussion here.

According to (7) DF can be viewed as a linear map from the space of sound fields S to the complex numbers $\mathbb{C}$, so the space of all possible directivity functions is the *dual space* S^*, of S. S and S^* have the same internal structure but represent different types of object. The subspace of S^* spanned by DFs in $\boldsymbol{Q}$ will be written

$$S^*_{\boldsymbol{Q}} \subseteq S^* \ . \tag{10}$$

For any sensible choice of DFs they are linearly independent, and form a basis of a subspace in S^*. We assume this case unless stated otherwise. A linearly dependent set is called a *frame* [19], and may arise for example when two different basis sets are joined together.

Equation (9) can also be viewed as an equation with an operator $\boldsymbol{Q}$ that acts on the sound field according to the original definition in (7). All subsequent expressions have analogs with this interpretation. We focus on the discrete case to give a presentation in terms of familiar matrix operations. For background on the linear algebra and matrix results employed here refer to [20].

3. Conversion Between Signal Sets

A range of problems can be expressed as the task of converting one set of signals associated with a set of DFs to another set of signals associated with another set of DFs. The DFs are known, but the associated sound field is not. One specific example is finding signals for a standard multichannel microphone set given a non-standard set of microphone signals. Another example is finding loudspeaker feeds for an array specified with panning functions from signals of an unrelated microphone set. This conversion is an example of decoding, since it is the process of reproducing the sound field from a set of encoding signals.

More precisely, given signals $\boldsymbol{q}$ and DFs $\boldsymbol{Q}$ such that $\boldsymbol{q} = \boldsymbol{Q}\boldsymbol{s}$ for an unknown sound field $\boldsymbol{s}$, what is the best estimate for signals $\boldsymbol{r}$ such that $\boldsymbol{r} = \boldsymbol{R}\boldsymbol{s}$ for known DFs $\boldsymbol{R}$? Generally there will be many sound fields satisfying $\boldsymbol{q} = \boldsymbol{Q}\boldsymbol{s}$, since $\boldsymbol{Q}$ has full row rank as noted previously. A natural estimate is the sound field $\tilde{\boldsymbol{s}}$ for which the L^2 norm $\|\boldsymbol{s}\|$ is a minimum, since this will have the least total energy. Nearly all possible sound fields $\boldsymbol{s}$ have excessively high energy, being distant from $\tilde{\boldsymbol{s}}$. Sparsity is another possible criteria useful for estimating $\boldsymbol{s}$, since natural sound fields are sometimes sparse. The L^1 norm is one way to select for sparsity, and can be used in combination with energy criteria, but this is not explored here. The kind of sound field that is usefully represented by channel encodings is dense and complex rather than sparse.

If $\boldsymbol{Q}$ has full row rank, as discussed in the previous section, then the least power estimate, written $\tilde{\boldsymbol{s}}_{\boldsymbol{q}}$ to indicate the dependence on known signals $\boldsymbol{q}$, can be calculated using the Moore-Penrose pseudo inverse $\boldsymbol{Q}^+$, equivalent in this case to $\boldsymbol{Q}^H(\boldsymbol{Q}\boldsymbol{Q}^H)^{-1}$,

$$\tilde{\boldsymbol{s}}_{\boldsymbol{q}} = \boldsymbol{Q}^+\boldsymbol{q} \tag{11}$$

The pseudo inverse is well conditioned provided the DFs $\boldsymbol{Q}$ are chosen reasonably, with no DFs being close to linear dependence. Equation (11) can also be viewed as an expansion of $\tilde{\boldsymbol{s}}_q$ with a set of vectors we call $\{\boldsymbol{Q}^*_j\}$ where $\boldsymbol{Q}^*_j = \{Q^+_{ij}\}$, the j^{th} column vector of $\boldsymbol{Q}^+$,

$$\tilde{\boldsymbol{s}}_{\boldsymbol{q}} = \sum_j Q^+_{ij} q_j = \sum_i q_i \boldsymbol{Q}^*_i \tag{12}$$

The inner product of vectors $\boldsymbol{Q}_i$ with vectors $\boldsymbol{Q}^*_j$ defines a matrix

$$\boldsymbol{Q}_i \cdot \boldsymbol{Q}^*_j = \sum_k Q_{ik} Q^+_{kj} \tag{13}$$

$$= \boldsymbol{Q}\boldsymbol{Q}^+ = \boldsymbol{I} = \delta_{ij} \quad . \tag{14}$$

The equivalence to the identity in the second line follows because $\boldsymbol{Q}$ has full row rank. $\{\boldsymbol{Q}^*_i\}$ is therefore a *dual basis* for $\{\boldsymbol{Q}_i\}$. The * superscript is chosen to indicate this. The space spanned by the dual basis will be called

$$S_{\boldsymbol{Q}^*} \subseteq S \quad . \tag{15}$$

The subspaces $S_{\boldsymbol{Q}^*}$ and $S^*_{\boldsymbol{Q}}$ are dual to one another. The space of sound fields S can be written as an orthogonal inner direct sum of $S_{\boldsymbol{Q}^*}$ and the null space $S_{\boldsymbol{Q}_0} = \{\boldsymbol{s} : \boldsymbol{Q}\boldsymbol{s} = 0\}$:

$$S = S_{\boldsymbol{Q}^*} \oplus S_{\boldsymbol{Q}_0} \quad , \tag{16}$$

since any sound field can be written as $\boldsymbol{s} = \tilde{\boldsymbol{s}_q} + (\boldsymbol{s} - \tilde{\boldsymbol{s}_q})$, where $\boldsymbol{q} = \boldsymbol{Q}\boldsymbol{s}$, $\tilde{\boldsymbol{s}_q} \in S_{\boldsymbol{Q}^*}$ and $(\boldsymbol{s} - \tilde{\boldsymbol{s}_q}) \in S_{\boldsymbol{Q}_0}$ (since $\boldsymbol{Q}(\boldsymbol{s} - \tilde{\boldsymbol{s}_q}) = \boldsymbol{q} - \boldsymbol{q} = 0$).

Fig. 1 illustrates this relationship and other quantities discussed below.

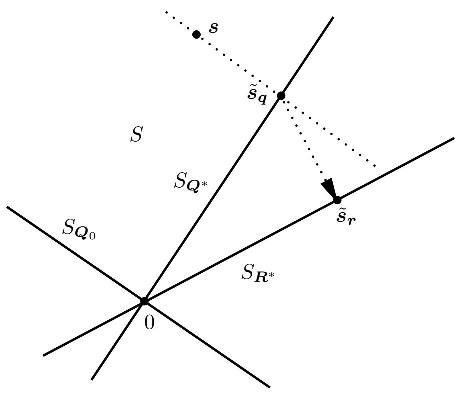

Figure 1: Illustration of relationships between quantities in the sound field space S. The view is normal to the plane of $\tilde{\boldsymbol{s}}_{\boldsymbol{q}}$ and $\tilde{\boldsymbol{s}}_{\boldsymbol{r}}$. $\boldsymbol{s}$ may be out of the plane.

From the estimate $\tilde{\boldsymbol{s}}_{\boldsymbol{q}}$ the corresponding estimate for $\boldsymbol{r}$ is

$$\tilde{\boldsymbol{r}} = \boldsymbol{R}\tilde{\boldsymbol{s}}_{\boldsymbol{q}} = \boldsymbol{R}\boldsymbol{Q}^{+}\boldsymbol{q} \tag{17}$$

$\boldsymbol{R}\boldsymbol{Q}^{+}$ is then the decoding/transcoding matrix from signals $\boldsymbol{q}$ to $\tilde{\boldsymbol{r}}$. If there exists a DF in $\boldsymbol{R}$ that is proportional with a DF in $\boldsymbol{Q}$, $\boldsymbol{R}_k = \alpha \boldsymbol{Q}_i$, then signal r_k can be recovered exactly, as we would hope: From (17) $r_k = \alpha \boldsymbol{Q}_i \boldsymbol{Q}^{+}\boldsymbol{q} = \alpha q_i$, using $\boldsymbol{Q}\boldsymbol{Q}^{+} = \boldsymbol{I}$.

4. Finding An Optimal Reduced Encoding

Given a set of DFs, it may be useful to find a smaller set of DFs from which the original DFs can be reconstructed. This would allow us to reduce the number of signals to be stored or transmitted. The reduction may be necessary because transmission is required over a lower bandwidth channel, or when several sets are combined, possibly with some redundancy. The goal then is to find the reduced set for which the original signals can be reconstructed as well as possible.

The reduced DF set will be chosen to minimise the difference between the signals from the original DFs and their estimates derived from the reduced DF signals, averaging over all plane waves. So, given N DFs $\boldsymbol{Q}$, how should M DFs $\boldsymbol{B}$ with $M < N$ be chosen to minimise the total estimated signal error over all plane waves, $\|\tilde{\boldsymbol{q}} - \boldsymbol{q}\|$, where $\tilde{\boldsymbol{q}} = \boldsymbol{Q}\boldsymbol{B}^{+}\boldsymbol{b}$, $\boldsymbol{b} = \boldsymbol{B}\boldsymbol{s}$? From here on $\boldsymbol{b}$ does not necessarily refer to an FBE encoding, although there is a connection that will become apparent. Likewise $\boldsymbol{B}$ does not necessarily refer to FBE DFs .

It can be shown that estimate for DFs $\boldsymbol{Q}$ constructed from DFs $\boldsymbol{B}$ that gives the lowest signal estimate error $\|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|$, averaged over all sound fields, is $\tilde{\boldsymbol{Q}}_{\boldsymbol{B}} = \boldsymbol{Q}\boldsymbol{B}^{+}\boldsymbol{B}$, with error $\|\tilde{\boldsymbol{Q}}_{\boldsymbol{B}} - \boldsymbol{Q}\|$. Hence the problem is equivalent to finding the value $\boldsymbol{B}$ giving the least error for estimates $\tilde{\boldsymbol{Q}}_{\boldsymbol{B}}$,

$$\boldsymbol{B}_{\boldsymbol{Q}} = \arg\min_{\boldsymbol{B}}(\|\tilde{\boldsymbol{Q}}_{\boldsymbol{B}} - \boldsymbol{Q}\|) \tag{18}$$

which is the DF set, spanning $S^{*}_{\boldsymbol{B}} \subseteq S^{*}$, that minimises the overall distance from $S^{*}_{\boldsymbol{B}}$ to the vectors $\{\boldsymbol{Q_i}\}$, illustrated in Fig. 2. The meaning of $\boldsymbol{B}_{\boldsymbol{Q}}$ is distinct from $\tilde{\boldsymbol{Q}}_{\boldsymbol{B}}$ and made clear by the presence of $\tilde{}$. The optimum set is not unique since any basis for $S^{*}_{\boldsymbol{B}}$ provides an alternative set.

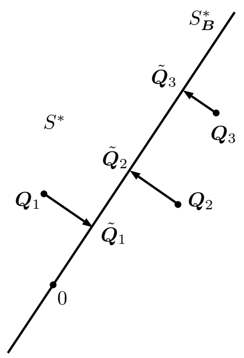

Figure 2: Choice of $S^{*}_{\boldsymbol{B}}$ in S^{*} minimising distance from mics $\boldsymbol{Q}$.

Singular Value Decomposition (SVD) solves this problem directly [20]: Given mics $\boldsymbol{Q}$ SVD provides unitary $\boldsymbol{U}$,$\boldsymbol{V}$ and diagonal $\boldsymbol{\Sigma}$ such that

$$\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{H} \tag{19}$$

and the diagonal entries of $\boldsymbol{\Sigma}$, the singular values, are real valued, non-negative, and ordered by decreasing size, $\boldsymbol{V}^{H}$ the complex conjugate of $\boldsymbol{V}$. Then a solution to (18) is given by

$$\boldsymbol{B}_{\boldsymbol{Q}} = \left[\boldsymbol{V}^{H}\right]_{M} \tag{20}$$

denoting the restriction of $\boldsymbol{V}^{H}$ to the first M rows, which form an orthonormal basis. The projection of $\boldsymbol{Q}$ into the space spanned by basis DFs $\boldsymbol{B}_{\boldsymbol{Q}}$ is

$$\tilde{\boldsymbol{Q}} = \boldsymbol{Q}\boldsymbol{B}^{H}_{\boldsymbol{Q}}\boldsymbol{B}_{\boldsymbol{Q}} \tag{21}$$

where $B_Q^H B_Q$ is a projection operator, since $B_Q B_Q^H = I$ by orthonormality of B_Q. Furthermore $B_Q^H = B_Q^+$ as B_Q has full row rank. The projection can also be evaluated as

$$\tilde{Q} = [U\Sigma]^M B_Q \tag{22}$$

where $[U\Sigma]^M$ is the restriction of $U\Sigma$ to the first M columns.

Finally, the reduced signals are given by

$$b = B_Q Q^+ q \ , \tag{23}$$

and to reproduce estimates of the original signals from the encoded signals,

$$\tilde{q} = Q B_Q^H b \ . \tag{24}$$

5. Conclusions

A framework was presented for representing sound fields, microphone directivity functions and panning functions, and the resulting signals. A method was found for converting signals from one directivity set to another, based on intermediate estimation of the sound field. This is compatible with conventional decoding methods including stereo and Ambisonics, and allows the decoding of general microphone signals to arbitrary loudspeaker arrays in a rational manner. An important general feature is that the psychoacoustical content of the loudspeaker panning functions is separated from the process of mapping the encoding functions on to the panning functions.

The overlap between encoding microphones allows channel signals to be compressed into fewer channels and restored approximately. While this causes some loss of spatial accuracy, only linear mixing is used, so the non-spatial fine structure of the signals is preserved. Weighting can be used to distribute spatial encoding accuracy non-uniformly in the compressed signals.

Acknowledgment

This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership. No new data were created during this study.

References

1. Bernfeld, B. Attempts for better understanding of the directional stereophonic listening mechanism, *Audio Engineering Society Convention 44*, March, no. C-4, (1973).

2. Blumlein, A. D. British patent specification 394,325 (improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems), *Journal of the Audio Engineering Society*, **6** (2), 91–130, (1958).

3. Craven, P. G. and Gerzon, M. A., (1977), *Coincident microphone simulation covering three dimensional space and yielding various directional outputs*. US Patent 4,042,779.

4. Gerzon, M. A. General metatheory of auditory localisation, *92nd Audio Engineering Society Convention, Vienna*, no. 3306, (1992).

5. Poletti, M. Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts, *Journal of the Audio Engineering Society*, **55** (7/8), 598–610, (2007).

6. Arteaga, D. An ambisonics decoder for irregular 3-d loudspeaker arrays, *Audio Engineering Society Convention 134*, May, (2013).

7. Benjamin, E., Heller, A. and Lee, R. Design of ambisonic decoders for irregular arrays of loudspeakers by non-linear optimization, *Audio Engineering Society Convention 129*, Nov, (2010).

8. Gerzon, M. A. NRDC, Design of ambisonic decoders for multispeaker surround sound, (1977).
9. Moore, D. and Wakefield, J. The design and analysis of first order ambisonic decoders for the itu layout, *Audio Engineering Society Convention 122*, May, (2007).
10. Scaini, D. and Arteaga, D. Decoding of higher order ambisonics to irregular periphonic loudspeaker arrays, *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Audio Engineering Society, (2014).
11. Batke, J.-M. and Keiler, F. Using vbap-derived panning functions for 3d ambisonics decoding, *2nd Ambisonics Symposium, Paris*, (2010).
12. Epain, N., Jin, C. and Zotter, F. Ambisonic decoding with constant angular spread, *Acta Acustica united with Acustica*, **100** (5), 928–936, (2014).
13. Wiggins, B. The generation of panning laws for irregular speaker arrays using heuristic methods, *Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio*, Audio Engineering Society, (2007).
14. Zotter, F. and Frank, M. All-round ambisonic panning and decoding, *J. Audio Eng. Soc*, **60** (10), 807–820, (2012).
15. Pulkki, V. Virtual sound source positioning using vector base amplitude panning, *J. Audio Eng. Soc*, **45** (6), 456–466, (1997).
16. Breebaart, J., et al. Spatial audio object coding (saoc)-the upcoming mpeg standard on parametric object based audio coding, *Audio Engineering Society Convention 124*, Audio Engineering Society, (2008).
17. Hardin, R. H. and Sloane, N. J. Mclaren's improved snub cube and other new spherical designs in three dimensions, *Discrete & Computational Geometry*, **15** (4), 429–441, (1996).
18. Daniel, J. Spatial sound encoding including near field effect, *Proc. AES 23nd International Conference, Helsinger, Denmark*, (2003).
19. Mallat, S., *A wavelet tour of signal processing*, Academic press (1999).
20. Golub, G. H. and Van Loan, C. F., *Matrix computations*, vol. 3, JHU Press (2012).
21. Fazi, F. M., Noisternig, M. and Warusfel, O. Representation of sound fields for audio recording and reproduction, *Acoustics 2012 Nantes*, (2012).
22. Simon, M., Menzies, D., Fazi, F. M., de Campos, T. and Hilton, A. A listener position adaptive stereo system for object based reproduction, *Proc. Audio Engineering Society 138th Convention, Warsaw*, May, (2015).
23. Kronlachner, M. and Zotter, F. Spatial transformations for the enhancement of ambisonic recordings, *2nd International Conference on Spatial Audio, Erlangen*, (2014).