The application of compressive sampling to the analysis and synthesis of spatial sound fields

Nicolas Epain¹, Craig Jin¹, André van Schaik¹

¹Computing and Audio Research Laboratory (CARLab), School of Electrical and Information Engineering, The University of Sydney, NSW 2006, Australia

Correspondence should be addressed to Craig Jin (craig@ee.usyd.edu.au)

ABSTRACT

Compressive sampling provides a new and interesting tool to optimise measurements of physical phenomena with a small number of sensors. The essential idea is that close to perfect reconstruction of the observed phenomenon may be possible when it can be described by a sparse set of basis functions. In this paper, we show how to apply compressive sampling techniques to the recording, analysis and synthesis of spatially extended sound fields. Numerical simulations demonstrate that our proposed method can dramatically improve the playback of spatialised sound fields as compared, for example, with High Order Ambisonics.

1. INTRODUCTION

Compressive sampling describes an ensemble of mathematical methods and theorems that enables faithful recovery of signals from sub-Nyquist-rate sampling. The key idea behind compressive sampling is that if the observed phenomenon can be described by a sparse set of basis functions, then the number of observations or sensors required to perfectly reconstruct this phenomenon may be much smaller than indicated by classic sampling criteria. For instance, a time-domain signal resulting from the sum of a few pure tones could be reconstructed perfectly even if the sampling rate is dramatically less than that dictated by the Nyquist-Shannon theorem. In this paper, we propose a technique to apply the mathematical tools and concepts of compressive sampling to the recording, analysis, and synthesis of spatial sound fields.

In a first part, we briefly review the mathematical concepts and tools behind compressive sampling. If the observed physical phenomenon is sparse in a certain domain, then in some sense it is optimal to mea-
sure the signal in a domain whose basis functions or waveforms are incoherent with the basis functions describing the sparse domain. Independent of the domain in which the signals are acquired, there are advantages for signal reconstruction in the sparse domain. In short, with compressive sampling we find the sparsest description of the measurement data by solving for a vector of coefficients that describe the measurement data in the sparse domain. This is typically achieved by applying linear programming methods to find the set of coefficients with smallest L1 norm.

In a second part, we show how the concepts of compressive sampling can be applied to the problem of analysing and reconstructing a sound field using an array of microphones. While various sparse domains can be chosen, it is often the case that a sound field originates from a small number of sound sources. Indeed, an interesting choice of a sparse basis consists simply of the sound fields emitted by the loudspeakers used for the audio playback. The intuition behind the selection of this sparse basis is that it is frequently preferable that a small number of loudspeakers are used for the playback of each sound source. A more complete and improved sparse basis is a plane-wave basis with high spatial resolution that does not depend on the playback setup. We also discuss the influence of the observation domain on the sound field analysis.

In the third part, we present primarily results of frequency-domain numerical simulations and also one time-domain simulation. We show that our proposed method can dramatically improve the process of spatial sound field analysis and playback over a set of loudspeakers, as compared, for instance, with Higher Order Ambisonics. In particular, results show that it is possible to analyse the spatial properties of the sound field very efficiently, even when the frequency range is restricted by the size of the microphone array, e.g., above the so-called spatial aliasing frequency. Interestingly, we also demonstrate that the precise knowledge of sound source locations enables the use of panning methods, such as vector-based amplitude panning (VBAP). In summary, the application of compressed sensing to the reconstruction of spatial sound fields ensures that a minimal number of loudspeakers are used for every sound source and thus considerably enlarges the sweet spot of the loudspeaker array.

2. SPATIAL SOUND FIELD ANALYSIS AND SYNTHESIS BASED ON SPHERICAL HARMONIC REPRESENTATIONS

In the following, we briefly survey how the spherical harmonic formalism is used to spatially analyse and synthesise sound fields recorded by sensor arrays. The reader familiar with the notions of spherical harmonic expansion as used in the Higher Order Ambisonics method can skim this section as much of it will be a review.

Spherical harmonic methods are based on the idea that a sound field can be expressed as a sum of elementary functions of the spherical coordinates. Solving the wave equation in this coordinate system shows that, in the frequency-domain, any sound field formed of incoming waves can be written as follows [1]:

$$ p(r, \theta, \varphi, \omega) = \sum_{m=0}^{\infty} \sum_{n=-m}^{m} b_{mn}(\omega) j_m(kr) Y^m_n(\theta, \varphi), $$

where $\omega$ is the frequency, $k$ is the wave number, $(r, \theta, \varphi)$ are the spherical coordinates, $j_m$ is the spherical Bessel function of degree $m$ and the $Y^m_n$ functions are the so-called spherical harmonics. Equation 1, known as the Fourier-Bessel representation of the sound field, establishes that the sound field can be entirely described by a set of complex numbers $b_{mn}$. Describing the sound field throughout the entire space requires an infinite number of coefficients. However, as shown in [2], truncating the series in Equation 1 up to order $M$ provides a good approximation of the acoustic pressure in the vicinity of the origin:

$$ p(r, \theta, \varphi, \omega) \approx \sum_{m=0}^{M} \sum_{n=-m}^{m} b_{mn}(\omega) j_m(kr) Y^m_n(\theta, \varphi) $$

with $kr \leq M$. (2)

As there are $(M+1)^2$ spherical harmonic functions up to order $M$, the same number of complex coefficients are required to describe the sound field at every frequency value.

Spherical harmonic analysis, as used in HOA, consists in finding the spherical harmonic expansion of
the sound field based on the information provided by a microphone array. Assuming the sensors behave linearly, such an array acts as a linear operator on the vector of spherical harmonic coefficients, as follows
\[
m(\omega) = T(\omega)b(\omega),
\]
(3)
where \(m(\omega)\) is the vector of the complex microphone gains, \(T(\omega)\) is the matrix defining the effect of the array, and \(b(\omega)\) is the vector of the \(b_{mn}\) coefficients. Matrix \(T\) depends on the sensors’ location, directivity and pointing direction. In the case of omnidirectional microphones in an open array the coefficients are found using the following equation
\[
T_{lmn}(\omega) = \delta_{lj}Y_{mn}^l(\theta_l,\phi_l),
\]
(4)
where \((r_l, \theta_l, \phi_l)\) are the spherical coordinates of the \(l\)th microphone. Formulae establishing these coefficients for other types of microphones can be found in [3], for instance.

Retrieving the spherical harmonic coefficients from the sensor signals basically consists in solving Equation 3 for \(b\) at every frequency value. This is usually implemented using finite impulse response filters whose frequency response determines the pseudo-inverse of matrix \(T\) for every frequency bin, as follows
\[
E(\omega) = \text{pinv}(T(\omega)),
\]
(5)
where \(E(\omega)\) is the matrix whose coefficients are given by the frequency response of the encoding filters.

Once the spherical harmonic expansion has been determined, it can be used for beamforming with the microphone array or to play back a recorded sound field over a set of loudspeakers. Both tasks are usually implemented using linear combinations of the spherical harmonic domain signals:
\[
g(\omega) = Db(\omega),
\]
(6)
where \(g(\omega)\) is the vector of loudspeaker complex gains and \(D\) is known as a decoding matrix. In the play-back case, \(D\) is frequently calculated based on the assumption that the loudspeakers act as plane-wave sources. Mathematically, this assumption leads to the following simple formula:
\[
b(\omega) = Y_{spk}g(\omega),
\]
(7)
where the columns of matrix \(Y_{spk}\) are the \(Y_{mn}^l(\theta_l,\phi_l)\) function values for each loudspeaker direction \((\theta_l, \phi_l)\). It follows that the decoding matrix is usually calculated as the pseudo-inverse of matrix \(Y_{spk}\), thus ensuring a perfect reconstruction of the sound field to a given order of spherical harmonic expansion.

Spherical-harmonic based methods possess a number of advantages which include the ability to easily reconstruct the sound field using various and arbitrary loudspeaker configurations. However, it also suffers from limitations related to both the encoding and decoding process. Firstly, as a finite number of sensors is used to observe the sound field, the encoding suffers from spatial aliasing at high frequencies [4]. Secondly, when the number of loudspeakers that are used for playback is larger than the number of spherical harmonic components used in the sound field description, one generally finds a deterioration in the fidelity of the constructed sound field [5]. Interestingly, in both cases, the limitations are related to the fact that an under-determined problem is solved using the pseudo-inverse method. In the following sections we show that, under certain assumptions, these limitations are circumvented by using compressive sampling methods.

3. COMPRESSIVE SAMPLING METHODS

Compressive sampling denotes an ensemble of mathematical concepts and tools that can be used to solve inverse problems arising in data acquisition. In the following, we briefly describe compressive sampling’s main results and tools. More information about the method can be found in [6].

When sampling a physical phenomenon in time or space, it is well known that perfect reconstruction of the signal is, in the general case, possible only if the sample rate satisfies Shannon’s sampling theorem. The main idea behind compressive sampling is that Shannon’s theorem is only a sufficient condition for the success of signal reconstruction: under certain assumptions, the observations provided by a sub-Nyquist sampling rate process contain enough information to describe the observed phenomenon perfectly.

According to compressive sampling theory, the main condition for a phenomenon to be efficiently observed through a small number of samples is that
it must be inherently simple. Mathematically, this means that the phenomenon can be described as a sum of contributions from a small number of elementary functions. If \( y \) denotes the vector of the observed (sampled) phenomenon, it can then be written as the following matrix-vector product:

\[
y = \Psi x,
\]

(8)

where \( \Psi \) is a basis of elementary functions and nearly all coefficients in \( x \) are null. If \( S \) coefficients are non-null, the phenomenon is then said \( S \)-sparse in the sparsity domain \( \Psi \). A good example of a sparsity domain is the Fourier domain for sinusoidal time signals: although such signals seem “full” in the time domain, they are perfectly described by a vector with only one non-null coefficient in the frequency domain. Compressive sampling’s main result may then be summarised as follows: given a physical phenomenon, if it is \( S \)-sparse in a particular domain, then the number of observations required to describe it perfectly is proportional to \( S \).

Another important concept in the compressive sampling theory is incoherence. Assuming the observed signal is sparse in a certain domain, then the number of observations required to describe the phenomenon perfectly will be lower if the observation domain is incoherent with the sparsity domain. In other words, if the effect of sampling the phenomenon is described by the matrix \( \Phi \) and the signal is sparse in a sparsity domain defined by matrix \( \Psi \), then the sampling will be more efficient if the correlation between \( \Phi \)’s rows and \( \Psi \)’s columns is small.

Perfect recovery of a sparse phenomenon from a small number of observations requires one to solve Equation 8 for \( x \). As the number of observations is small compared to the number of elementary functions in \( \Psi \), the system is under-determined and there are an infinite number of solutions. The compressive sampling approach to solving this system and resolving the ambiguity is to look for the sparsest of these solutions, which translates into the following optimisation problem:

\[
\text{minimise } ||x||_1 \\
\text{subject to } y = \Psi x.
\]

(9)

Note that the L1-norm is used to measure the sparsity of vector \( x \) instead of the L0-norm. This is because, unlike the L0-norm, L1-norm minimisation leads to a roughly equivalent optimisation problem than can be solved by linear programming methods.

4. APPLYING COMPRESSIVE SAMPLING METHODS TO SPATIAL SOUND FIELD ANALYSIS AND SYNTHESIS

As stated in the previous section, the application of compressive sampling methods requires the prior knowledge of a sparsity domain. Although the spherical harmonic domain presents interesting properties, it is not a good candidate in terms of sparsity. There is no reason for the spherical harmonic expansion to be sparse unless all of the sound sources are located in very particular directions such as \((0,0)\). On the other hand, if the sound field results from a few sound sources, it is likely to be described by a small number of coefficients in a source domain. In the following, we show how a discrete plane-wave basis can be used as a sparsity domain for the compressive sampling approach of sound field analysis/synthesis.

Let us return to the spherical harmonic decoding problem: assuming that the spherical harmonic expansion is specified up to an order \( M \), a gain must be found for the loudspeakers such that it solves Equation 7. The classical manner in which this linear problem is solved is to calculate the inverse or pseudo-inverse of \( Y_{\text{spk}} \). This method provides a
Fig. 2: Speaker gains obtained when decoding an order-4 encoded plane-wave over 48 loudspeakers by: 1- the pseudo-inverse method; 2- solving the optimisation problem 10 with \( \epsilon = 10^{-3} \)

unique and exact solution when the system is underdetermined, i.e., when the number of plane-waves is greater than the number of spherical harmonic components. This solution is referred to as the least-square solution because of its characteristic property of having the lowest energy (and L2-norm) among all of the exact solutions.

However, the least-square solution tends to distribute the energy evenly between the loudspeakers, as illustrated in Figure 1 for the case where only one source is being encoded/decoded. This becomes a problem when the number of loudspeakers is greater than the number of spherical harmonic components being used for the expansion of the sound field. In this case, we frequently find many loudspeakers turned on and driving similar signals leading to spectral distortion that reduces the size of the sweet spot. In this sense, we obtain a larger sweet spot when a minimal number of loudspeakers are used to recreate a given source. This implies there is a benefit when the vector \( \mathbf{g} \) of loudspeaker gains is made appropriately sparse.

A first and intuitive application of compressive sampling methods is therefore to solve the decoding problem by finding a solution to the following optimisation problem:

\[
\begin{align*}
\text{minimise} & \quad ||\mathbf{g}||_1 \\
\text{subject to} & \quad ||Y_{\text{spk}} \mathbf{g} - \mathbf{b}||_2 \leq \epsilon,
\end{align*}
\]

which means searching for the sparsest gain vector that explains the observed spherical harmonic expansion with a specified precision \( \epsilon \). Figure 2 shows the result obtained when solving this problem using CVX, a package for specifying and solving convex programs [7, 8], as compared to the gains obtained from using the pseudo-inverse of \( Y_{\text{spk}} \) in the case where a unique plane-wave source is present. The solution obtained by the compressive sampling approach clearly locates the plane wave more accurately, while achieving a relative error of 0.1%.

In the compressive sampling approach to HOA decoding described above, we have used the basis formed by the speaker contributions to the spherical expansion as the sparsity domain, which can be seen as performing a discrete plane-wave analysis on this basis. Although this idea seems promising, the spatial resolution of the analysis is still imposed by the loudspeaker set-up. Furthermore, the sparsity of our solution depends on the dimension of the basis: the more plane-waves, the sparser the solution.
Therefore, better results may be obtained by performing the analysis using a plane-wave basis that has greater angular resolution. This is illustrated in Figure 3 for the case in which there is a two degree resolution between different plane-wave directions.

Once the plane-wave analysis is complete, the results can be easily played back over loudspeakers or headphones and also used for beamforming. In both cases, the most straightforward method is to amplitude-pan the different plane waves to the different speakers/beams. This is achieved by forming a linear combination of the plane-wave signals, as follows:

\[ \mathbf{b}(\omega) = \mathbf{A} \mathbf{s}(\omega), \]  

where \( \mathbf{s}(\omega) \) is the vector of plane wave gains, \( \mathbf{b} \) is the vector of speaker/beam gains and \( \mathbf{A} \) is the matrix defining the panning law between the plane-wave directions and the speaker directions. \( \mathbf{A} \) can be chosen as the VBAP panning law, for instance. Alternatively, the plane-wave signals may be re-encoded in the HOA domain at some high order \( M' \) as shown by the following equation:

\[ \mathbf{b}'(\omega) = \mathbf{Y}_{\text{plw}}^T \mathbf{s}(\omega), \]

where \( \mathbf{Y}_{\text{plw}} \) is the matrix defining the contribution of the different plane-waves to the spherical harmonic expansion \( \mathbf{b}'(\omega) \) up to order \( M' \).

5. NUMERICAL SIMULATIONS

We now present simulation results demonstrating the advantage of using the compressive sampling approach versus the HOA method in the case where the sound field can be modelled as sparse or nearly-sparse in the plane-wave domain. This section is divided into three parts: we first describe the original sound field; we then describe the HOA reconstructed sound field; finally we describe the sound field reconstructed using compressive sampling. Please note that for these simulations, the loudspeakers are modelled as spherical sources.

5.1. Frequency-domain simulations

5.1.1. Original Sound Field

Given a set of spherical sources located at different positions in space, the resulting complex pressure field is given by

\[ \mathbf{p}_{\text{orig}} = \mathbf{T}_{\text{sou/pts}} \mathbf{g}_{\text{sou}}, \]  

where \( \mathbf{p}_{\text{orig}} \) is the vector of pressure values for some specified set of field points, \( \mathbf{g}_{\text{sou}} \) is the vector of the source gains and \( \mathbf{T}_{\text{sou/pts}} \) is the transfer matrix between the sources and the field points, whose coefficients are given by:

\[ \mathbf{T}_{\text{sou/pts}}(m,n) = \frac{e^{-ikr_{mn}}}{r_{mn}}, \]  

where \( k \) is the wave number and \( r_{mn} \) is the distance between the \( mn \)th field point and the \( mn \)th source. In the ensuing discussion this sound field is denoted as the original sound field, which we then observe using a microphone array and try to re-synthesise.

5.1.2. HOA Sound Field

Microphone signals are recorded by a microphone array positioned in the original sound field. These signals are given by:

\[ \mathbf{s}_{\text{mic}} = \mathbf{T}_{\text{sou/mic}} \mathbf{g}_{\text{sou}}, \]

where \( \mathbf{T}_{\text{sou/mic}} \) is the transfer matrix between the sources and the microphones. For the general case, this transfer matrix can be modelled using a spherical harmonic expansion, as follows

\[ \mathbf{T}_{\text{sou/mic}} = \mathbf{T}_{\text{sph/mic}} \mathbf{T}_{\text{sou/sph}}, \]

where \( \mathbf{T}_{\text{sph/mic}} \) represents the transfer matrix between the spherical harmonic domain and the microphone signals and \( \mathbf{T}_{\text{sou/sph}} \) is the transfer matrix between the source signals and the spherical harmonic domain. Note that for the simulations presented below the order of the expansion is chosen sufficiently high so that the acoustic behaviour of the microphone array is modelled accurately.

For our simulations, the microphone array consists of omnidirectional microphones located at the surface of a rigid sphere, therefore \( \mathbf{T}_{\text{sph/mic}} \) is given by

\[ \mathbf{T}_{\text{sph/mic}} = \mathbf{Y}_{\text{mic}}^T \mathbf{W}_{\text{mic}}, \]

where \( \mathbf{Y}_{\text{mic}} \) is the matrix whose columns are the values of the spherical harmonic functions for each microphone direction and \( \mathbf{W}_{\text{mic}} \) is the diagonal matrix whose coefficients are defined by

\[ w_{\text{mic}}(m) = i^m \left( j_m(kR) - h_m^{(2)}(kR) + j_m^{(2)}(kR) \right), \]
where $R$ is the radius of the sphere, and $h^{(2)}$ is the spherical Hankel function of the second kind. Note that $m$ denotes the order of the spherical harmonic expansion.

The coefficients of the transfer matrix between the sources and the spherical harmonic expansion are given by the following formula:

$$ t_{\text{s commun}}(m, n, s) = i^{-m} \frac{h^{(2)}_m(kr_s)}{h^{(2)}_0(kr_s)} Y^m_n(\theta_s, \varphi_s). \quad (19) $$

Our overall objective is to compare the sound fields re-synthesised using HOA and our compressive sampling method. For the case in which HOA is used, the spherical harmonic coefficients up to order $M$ are estimated using:

$$ b_{\text{HOA}} = \text{pinv} \left( \hat{T}_{\text{sph/mic}} \right) s_{\text{mic}}, \quad (20) $$

where $\text{pinv} \left( \hat{T}_{\text{sph/mic}} \right)$ denotes the regularised pseudo-inverse of the truncated-to-order-$M$ matrix $T_{\text{sph/mic}}$. Then, the loudspeaker gains are computed using the basic HOA decoding given by:

$$ g_{\text{HOA}}^{\text{spk}} = \frac{1}{N_{\text{spk}}} Y^T_{\text{spk}} b_{\text{HOA}}, \quad (21) $$

where $N_{\text{spk}}$ is the number of loudspeakers and $Y_{\text{spk}}$ is the matrix whose columns are the values of the spherical harmonic functions for each loudspeaker direction (up to order $M$). Finally, the re-synthesised sound field in the HOA case is given by:

$$ p_{\text{HOA}}^{\text{pts}} = T_{\text{sph/pts}} g_{\text{HOA}}^{\text{pts}}, \quad (22) $$

where $T_{\text{sph/pts}}$ is the transfer matrix between the loudspeakers and the field points, whose coefficients are computed similarly to those of $T_{\text{s commun}}$.  

### 5.1.3. Compressive Sampling Sound Field

In the case where the compressive sampling approach is used, the first step is a plane-wave analysis of the sound field. This is implemented by solving the following optimisation problem:

minimise $||g_{\text{plw}}||_1$

subject to $||T_{\text{plw/mic}} g_{\text{plw}} - s_{\text{mic}}||_2 \leq \epsilon_1$

and $||\text{pinv} \left( T_{\text{plw/hoa}} \right) b_{\text{HOA}}||_2 \leq \epsilon_2, \quad (23)$

where $T_{\text{plw/mic}}$ is the transfer matrix between the plane waves and the microphones, computed in a similar way as $T_{\text{s commun/mic}}$. With regard to the second constraint, $T_{\text{plw/hoa}}$ is the transfer matrix between the plane waves and the HOA-estimated spherical harmonic expansion. The reason for the second constraint will be made clear in Section 7 and essentially provides a means to control the distance between the HOA solution and the compressive sampling solution. In a first reading, this constraint can be ignored. In any case, one should note that the HOA-estimated spherical harmonic expansion is also derived via the microphone array. Thus, $T_{\text{plw/hoa}}$ is given by:

$$ T_{\text{plw/hoa}} = \text{pinv} \left( T_{\text{sph/mic}} \right) T_{\text{sph/mic}} Y_{\text{plw}}, \quad (24) $$

where $Y_{\text{plw}}$ is the matrix whose columns are the values of the spherical harmonic functions for each plane wave direction. The loudspeaker gains are then found by amplitude-panning the plane-wave gains, as follows

$$ g_{\text{spk}}^{\text{cs}} = P_{\text{plw/spk}} g_{\text{plw}}, \quad (25) $$

where $P_{\text{plw/spk}}$ components are found using the VBAP method. Finally, the re-synthesised sound field in the compressive sampling approach case is given by

$$ p_{\text{pts}}^{\text{cs}} = T_{\text{spk/pts}} g_{\text{spk}}^{\text{cs}}. \quad (26) $$

### 6. RESULTS

In our simulations, the microphone array is a 4 cm radius rigid sphere with 32 omnidirectional microphones evenly distributed on the surface. The sound fields are re-synthesised using a ring of 48 loudspeakers with a radius of 1 m. In the HOA case, the microphone gains are HOA-encoded up to order 4. The compressive sampling plane-wave analysis is performed using a basis of 360 plane waves evenly distributed in the horizontal plane. The values of the $\epsilon_1$ and $\epsilon_2$ coefficients have been fixed to $10^{-3}$ and 2, respectively. In every case, the directions of the sound sources that define the sound field have been randomly chosen in the horizontal plane.

Figure 4 shows the result obtained for a sound field defined by four spherical sources at 2kHz. In the compressive sampling approach case, only the loudspeakers that are the closest to the different source
Fig. 4: Frequency-domain simulation of sound field spatial analysis/synthesis using a 32-sensor spherical microphone array and a circular 48-loudspeaker playback setup. Frequency is 2kHz. Top-left: original sound field ; Top-right: HOA encoded/decoded sound field ; Bottom-left: compressive sampling approach ; Bottom-right: speaker gains comparison. Note: the circle has been plotted to help visualisation.

directions are used for playing the sound field back. In contrast, many loudspeakers are used for the HOA decoding. This explains why the area in which the sound field is accurately reproduced is much larger for the compressive sampling case.

The compressive sampling algorithm is designed to work when the observed phenomenon can be described as sparse in a certain domain. This is clearly the case for the previous simulation where the sound field consists of only four sources. But the application of the algorithm should still provide reasonable results when the sound field is nearly sparse, i.e., it should demonstrate robustness to sparsity. This is important, for example, when there are a few primary sources and several secondary ones due to reflections.

Figure 5 shows the results obtained at 2kHz, when the sound field consists of 4 primary sources and 8 secondary ones. As the figure indicates, the compressive sampling approach accurately detects not only the primary sources, but also many of the secondary ones. Note that the two secondary sources with small gain located around 300° arise because the sources have opposite phase. We find again that the overall quality of the sound field reproduction improves with compressive sampling when compared
Fig. 5: Frequency-domain simulation of sound field spatial analysis/synthesis using a 32-sensor spherical microphone array and a circular 48-loudspeaker playback setup. Frequency is 2kHz. Top-left: original sound field; Top-right: HOA encoded/decoded sound field; Bottom-left: compressive sampling approach; Bottom-right: speaker gains comparison.

to HOA.

One of the known advantages of the compressive sampling approach is that it allows one to sample at sub-Nyquist rates. This is important because HOA generally suffers from spatial aliasing when the wavelength is small compared to the distance between the microphones. In our simulations, the spherical harmonic expansion is found using the classic HOA-encoding scheme, which means there is spatial-aliasing. However, when the source direction is known, we actually know its complete influence on the estimated expansion, including the aliasing phenomenon. Thus, the compressive sampling approach may still give good results at frequencies where HOA cannot.

Figure 6 shows the results obtained at 16kHz, in the case where, again, the sound field results from 4 primary sources and 8 secondary ones. Although the frequency is about twice the spatial aliasing frequency of the microphone array, the compressive sampling approach offers a very accurate spatial analysis, detecting every primary source and a large portion of the secondary ones. This results in a very accurate reproduction of the sound field when compared to the one obtained with HOA.
Fig. 6: Frequency-domain simulation of sound field spatial analysis/synthesis using a 32-sensor spherical microphone array and a circular 48-loudspeaker playback setup. Frequency is 16kHz. Top-left: original sound field; Top-right: HOA encoded/decoded sound field; Bottom-left: compressive sampling approach; Bottom-right: speaker gains comparison.

6.1. Time-domain simulation

Frequency-domain simulations of the compressive sampling approach show promising results. However, these results are only representative for what happens in a purely sinusoidal sound field, which does not represent a realistic recording situation. The time-domain implementation of the compressed-sensing approach is still a work in progress as will be discussed in the next section. However, we provide some initial results from a time-domain implementation where only one source is present and these are shown in Figure 7. In the case where HOA is used, a large number of loudspeakers are used for the re-synthesis and the wave is focused in a very small area surrounding the centre of the reproduction area. Note that the curvature of the wave is at first inverted when compared with the original sound field. By contrast, the compressive sampling approach ensures that only two loudspeakers strongly contribute to re-creating the sound field. The resulting sound field therefore more accurately matches the original one, with a sweet spot several time larger than the one observed with HOA.

7. CONCLUDING REMARKS AND PERSPECTIVES

In this paper we have shown how compressive sampling methods can be applied to spatial sound field analysis/synthesis. In the case where only a few
Fig. 7: Time-domain simulation of a spatial sound field analysis and re-synthesis using a 32-microphone spherical array and a circular 48-loudspeaker playback system. From left to right: order-4 HOA encoding and decoding; original soundfield; compressive sampling plane-wave analysis and amplitude panning.
sound sources are present, a plane wave basis can be chosen as the sparsity domain. L1-norm minimisation algorithms then provide a high-resolution spatial analysis of the sound field, which can be used for beamforming or spatial synthesis over loudspeakers. The reconstructed sound field then shows a much larger sweet spot as compared to the one obtained using classic HOA methods.

While we have not emphasised it so far, with Equation 23, we have shown how to smoothly control the variation of the compressive sampling solution from the HOA solution. This is the purpose of the second constraint in Equation 23. It is an open research question to understand from the acoustic measurement data when it is beneficial to treat the sound field as sparse. Often with research related to compressive sampling, sparsity is simply taken for granted, but this is frequently not the situation with an acoustic sound field and therefore we have highlighted the ability to smoothly vary the HOA solution using compressive sampling methods. The HOA and compressive sampling methods differ in that the HOA reconstruction filters do not depend on the sound field. Because the compressive sampling reconstruction filters vary with the instantaneous sound field, we have the chance to improve the reconstruction of the sound field when it is sparse. In this sense, we are really taking advantage of the fact that the sound field reconstruction problem can be viewed as an acoustical inverse problem for which compressive sampling methods can yield better results.

8. REFERENCES


