Reproducing 3D-Sound for Videoconferencing: a Comparison between Holophony and Ambisonic

Rozenn Nicol (rozenn.nicol@cnet.francetelecom.fr)
Marc Emerit (marc.emerit@cnet.francetelecom.fr)

France Telecom CNET Lannion

Abstract

Current research in videoconferencing is focussed on 3D-Sound reproduction over a wide listening area. This paper compares two methods: Holophony and Ambisonic. First, it will be shown that Ambisonic actually corresponds to a particular case of holophonic reconstruction, then the specificities of the two approaches will be analysed in terms of complexity and size of listening area. As a result, a third method, which shall combine the advantages of both Holophony and Ambisonic, will be suggested.

1 Introduction

In a videoconference, people communicate via an audiovisual interface. In order to enhance the transparency of this interface, current developments intend to introduce 3D-reproduction methods both for image and sound [1]. This paper deals only with the problem of *3D-Sound reproduction*. First, one point should be kept in mind: in a videoconference, the listening area must be wide, since it is addressed to several listeners. This requirement clearly excludes the stereophonic and binaural technologies, which are restricted to a single listener.

In fact, methods based on physical reconstruction of the acoustical field offer the most relevant solutions to create an extensive listening area. Two approaches have been considered: *Holophony* –acoustical equivalent to holography– and *Ambisonic*. From a theoretical point of view, Holophony is the only solution which ensures a perfect accurate reproduction, but its implementation is rather complicated. On the contrary, Ambisonic is remarkably simple and easy to use. However, the reconstruction is accurate over such a small area that it hardly encompasses one listener.

Holophony and Ambisonic thus achieve opposite compromises between system complexity and size of listening area. Consequently, by comparing their approach, it is intended to deduce a third method which combines their advantages. In other words, Ambisonic could be improved via Holophony or Holophony simplified via Ambisonic. Before linking them, the theoretical basis of each method is briefly recalled.

2 Holophony

Holophony is derived from the Huygens' Principle [2], which conveys the idea that an acoustical field within a volume Ω can be expressed as an integral over the boundary $\partial\Omega$ [3]:

$$p(\vec{r}) = \iint_{\partial\Omega} \left[\vec{\nabla} p_0 . \vec{n} - \frac{\vec{R}}{R} . \vec{n} \ (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4\pi R} \, dS_0 \quad \forall \vec{r} \in \Omega \quad (1)$$

where p refers to the pressure field inside Ω and p_0 to the pressure field over $\partial\Omega$ (cf. Fig.1).

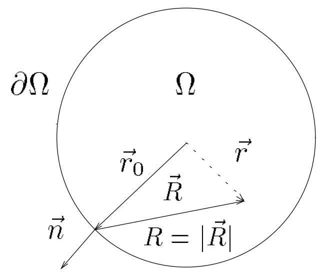

Fig.1 - *Notations*

This integral, known as the *Kirchhoff-Helmholtz Integral*, may be interpreted as a distribution of secondary sources along the boundary $\partial\Omega$ and therefore defines a method to reproduce a pressure field within an extensive area. First, the field must be recorded along the boundary both by pressure and velocity microphones, in order to pick up the pressure (p_0) and its gradient ($\vec{\nabla} p_0$). Secondly, microphones are replaced by loudspeakers, which are fed by the signals previously recorded, so that they perfectly reconstruct the original field. It should be noticed that, as for the microphones, 2 different types of loudspeakers must be used: monopole and dipole sources (cf. equ.1).

However, the holophonic reconstruction is perfectly accurate, provided that a continuous distribution of transducers is used both for recording and reproduction. In practice, only discontinuous array of microphones and loudspeakers are available, which raises the problem of *spatial sampling* [4]. The transducer spacing is determined by the minimal wave length of the recorded field, in order to avoid aliasing according to the Shannon criterion. The number of transducers is therefore dictated not only by the size of the area $\partial\Omega$, but also by the spectrum of the field. Unfortunately, it is obvious that the Shannon criterion can not be practically satisfied for high frequencies.

Nevertheless, investigations carried out at the D.U.T. (Delft University of Technology) have shown that spatial aliasing occuring at high frequencies does not severely degrade the result, as long as the aliasing frequency is greater than 1.5 kHz [4]. Additionnally, they point out further simplifications. First, if the reproduction is limited to a plane instead of a volume, the secondary source distribution can be restricted to a curve instead of a surface, by using the *Stationary Phase Approximation* [5]. Secondly, the contribution of the monopole and dipole are redundant, so that only monopoles fed by velocity microphones are sufficient.

3 Ambisonic

The theory of Ambisonic specifically deals with the problem of plane wave reconstruction, but this assumption is not restrictive, insofar as any acoustical field can be expressed as a sum of plane waves [3]. The reconstruction process is based on the expansion of a plane wave (amplitude a, wave number k, angle of incidence θ) into a *Fourier-Bessel series* [6]:

$$p(r,\phi) = aJ_0(kr) + 2a\sum_{n=1}^{+\infty} j^n J_n(kr)\cos(n\theta)\cos(n\phi) + 2a\sum_{n=1}^{+\infty} j^n J_n(kr)\sin(n\theta)\sin(n\phi) \quad (2)$$

where the pressure field p is described in cylindrical co-ordinates (r,ϕ) and the function J_n refers to the nth order Bessel function of the first kind. By introducing the vectorial notation:

$$\mathbf{c}^T = \sqrt{2}\,[\tfrac{1}{\sqrt{2}}\cos(\theta)\ \sin(\theta)\cdots\cos(n\theta)\ \sin(n\theta)\cdots]$$
$$\mathbf{h}^T = \sqrt{2}\,[\tfrac{J_0(kr)}{\sqrt{2}}\ j\cos(\phi)J_1(kr)\quad j\sin(\phi)J_1(kr)\cdots j^n\cos(n\phi)J_n(kr)\quad j^n\sin(n\phi)J_n(kr)\cdots]$$

the previous equation becomes:

$$p = a\ \mathbf{c}^T\ \mathbf{h}\ . \quad (3)$$

The information concerning the spatial distribution of the plane wave is fully described by the vector $\mathbf{c}$, which means that the sound pick-up consists in identifying the coefficients c_n. However, for an exhaustive characterization, these coefficients must be known for $n = 0$ to $n = +\infty$. In practice, the expansion will be inevitably truncated. Moreover, to measure the physical quantities c_n, microphones whose directivity follows a $\cos(n\theta)$ or $\sin(n\theta)$ law are required, and, except for $n = 0$ (omnidirectional microphone) and $n = 1$ (bidirectional microphone), such microphones do not exist. Therefore, until now, Ambisonic recordings, which use the *Soundfield* microphone designed par M. Gerzon [7], are truncated at the first order.

For the sound reproduction, the listener is surrounded by N loudspeakers equally distributed along a circle centered on him. Provided that the radius of the circle is greater than the wave length, the field induced by each loudspeaker at the position of the listener is assimilated to a plane wave:

$$p_i = a_i\ \mathbf{c}_i^T\ \mathbf{h}\ .$$

The superposition of the N waves gives:

$$p = \sum_{i=1}^{N} a_i\ \mathbf{c}_i^T\ \mathbf{h} = \mathbf{A}^T\ \mathbf{C}\ \mathbf{h}\ , \quad (4)$$

with: $\mathbf{A}^T = [a_1\ a_2 \cdots\ a_N]$, $\mathbf{C}^T = [\mathbf{c}_1\ \mathbf{c}_2 \cdots\ \mathbf{c}_N]$. By identifying the field reproduced (equ. 4) with the original field (equ. 3), the loudspeaker weightings $\mathbf{A}$ can be obtained as follows [8]:

$$\mathbf{A} = \frac{a}{N}\ \mathbf{C}\ \mathbf{c} \quad (5)$$

provided that $N \geq 2M+1$, where M is the truncature order of the Fourier-Bessel series (in practice $N = 2M+1$), so that $\mathbf{C}\mathbf{C}^T = \frac{1}{N}\ \mathbf{1}$ ($\mathbf{1}$: Identity Matrix). Under this condition, the reconstruction is perfectly accurate.

4 Linking Ambisonic and Holophony

To link Ambisonic and Holophony, the Kirchhoff-Helmholtz Integral is derived for the case of a plane wave reconstruction by a circular layout of loudspeakers, i.e:

- the 2D-integral is reduced to an 1D-integral by applying the Stationary Phase Approximation, so that $\partial\Omega$ becomes a circle of radius r_0,
- the pressure field p_0 is a plane wave described by the equation 2.

Finally:

$$p(r,\phi) = \frac{r_0}{2\sqrt{2\pi jk}}\int_0^{2\pi}\left[\frac{\partial p_0}{\partial r_0} - \cos\alpha\ (1+jkR)\frac{p_0}{R}\right]\frac{e^{-jkR}}{\sqrt{R}}d\phi_0 \quad (6)$$

where:

$$R = \sqrt{r^2 + r_0^2 - 2rr_0\cos(\phi - \phi_0)}$$
$$\cos\alpha = \frac{r\cos(\phi-\phi_0)-r_0}{R}$$

Since, for Ambisonic, the contribution of each loudspeaker is assimilated to a plane wave, the sources are set to infinity by replacing the Bessel functions by their asymptotic expansion in equ.6. Then the limit when r_0 tends to infinity is taken, so that:

$$p(r,\phi) = \frac{1}{2\pi}\int_0^{2\pi}\left[1 + 2\sum_{n=1}^{+\infty}\cos n(\phi_0 - \theta)\right] e^{jkr\cos(\phi-\phi_0)}d\phi_0 \quad (7)$$

If a discontinuous distribution of N sources is considered:

$$p(r,\phi) = \frac{1}{N}\sum_{i=1}^{N}\left[1 + 2\sum_{n=1}^{+\infty}\cos n(\phi_i - \theta)\right] e^{jkr\cos(\phi-\phi_i)} \quad (8)$$

and it is easy to verify that this result coincides with equ. 4, which points out that Ambisonic is a particular case of the holophonic approach.

5 Discussion

In this section, a parallel will be drawn between the holophonic and ambisonic methods concerning the sound pick-up, which will be analysed in terms of system complexity (number of signals recorded or microphones required), reconstruction performance (accuracy, size of listening area) and implementation feasibility.

5.1 Sound Pick-up Strategy

Whereas, for Holophony, the pressure field p_0 is recorded along a *curve* (a circle), for Ambisonic, the recording is made at a *point* (the centre of the circle) and consists in extracting the coefficients c_n of the Fourier-Bessel series (cf. equ.3). However, these two approaches are based on equivalent description of the spatial behaviour of the acoustical field. Indeed, for Holophony, the pressure $p_0(r_0, \phi_0)$ recorded along a circle of radius r_0 may be developed into a complex Fourier series:

$$p_0(r_0,\phi_0) \equiv p_0(\phi_0) = \sum_{n=-\infty}^{n=+\infty} \gamma_n\ e^{jn\phi_0} \quad (9)$$

The coefficients γ_n are deduced from the Fourier Transform of the signal p_0. For a plane wave, a comparison between equ.2 and 9 yields:

$$\begin{cases} \gamma_0 & = J_0(kr_0) = J_0(kr_0)\ c_0 \\ \gamma_n & = j^n J_n(kr_0)e^{-jn\theta} \qquad (n \geq 1) \\ & = \frac{j^n}{\sqrt{2}}\ J_n(kr_0)\ [c_{2n-1} - jc_{2n}] \\ \gamma_{-n} & = j^n J_n(kr_0)e^{jn\theta} \\ & = \frac{j^n}{\sqrt{2}}\ J_n(kr_0)\ [c_{2n-1} + jc_{2n}] \end{cases} \quad (10)$$

so that the coefficients c_n can be derived from the coefficients γ_n:

$$\begin{cases} c_0 = \frac{\gamma_0}{J_0(kr_0)} \\ c_{2n-1} = \frac{[\gamma_n + \gamma_{-n}]}{j^n\sqrt{2}J_n(kr_0)} \\ c_{2n} = j\ \frac{[\gamma_n - \gamma_{-n}]}{j^n\sqrt{2}J_n(kr_0)} \end{cases} \quad (11)$$

5.2 Encoding Error

Although both Holophony and Ambisonic sound pick-up fundamentally provide an equivalent and exhaustive characterization of the pressure field, their implementation raises problems of different nature: *spatial sampling* of the pressure signal p_0 for Holophony and *truncature* of the series c_n for Ambisonic. On one hand, spatial sampling of p_0 produces aliasing of its spectrum, i.e. of the coefficients γ_n. From equ.10, it can be seen that the size of this spectrum is theoretically infinite, but it can be considered as finite in a first approximation, because of the marked decline of Bessel functions. Nevertheless it increases with the frequency (k) and the radius r_0. On the other hand, the truncature of the Fourier-Bessel series (cf. equ.2), which recalls the Taylor's Formula, may influence the size of the area where an accurate reconstruction is expected.

Spectral aliasing and truncature induce an *encoding error*, which may be evaluated by:

$$e(r,\phi) = |p(r,\phi) - \hat{p}(r,\phi)| \quad (12)$$

where p is the original field and $\hat{p}$ the field reconstructed according to a Fourier-Bessel series, by considering either the coefficients γ_n computed from the sampled field p_0 or the $2M+1$ first coefficients c_n. The *mean error*, which is defined as:

$$\overline{e}(r) = \frac{1}{2\pi}\int_0^{2\pi} e(r,\phi)d\phi$$

is illustrated by the Fig.2. The parameter S refers to the number of recorded signals, i.e. for Holophony $S = N$ if the pressure field is recorded at N equally distributed points and for Ambisonic $S = 2M+1$ if M is the truncature order, so as to compare the two methods for equal cost.

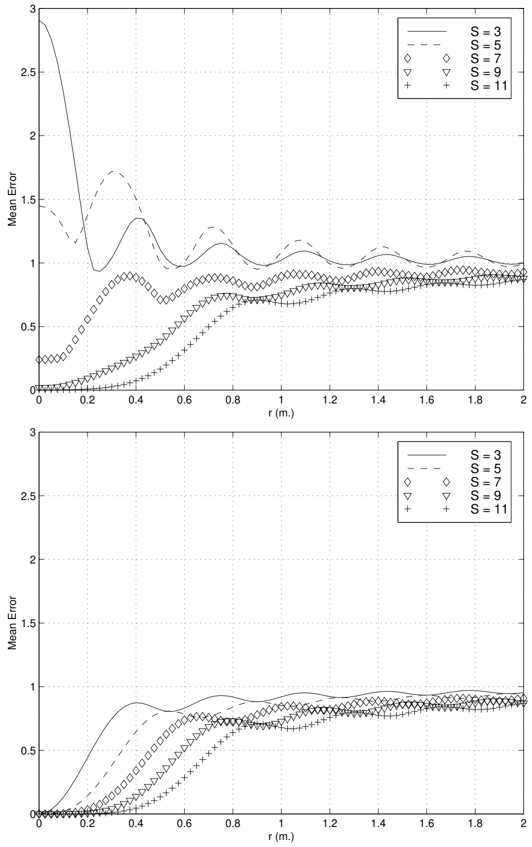

Fig. 2 - *Mean Encoding Error (f = 500 Hz) for Holophony (top) and Ambisonic (bottom)*

First, it should be noticed that, whatever the number S, the Ambisonic error is zero at the centre of the circle and keeps very low at its neighbourhood, but, as expected (cf. Taylor's Formula), it increases with radius r quite rapidly. However, the higher the truncature order, the wider the area of accurate reconstruction. On the contrary, the holophonic error for low values of S is high everywhere, and particularly near the centre of the circle. As S increases, the error noticeably declines, until the aliasing effect disappears. A truncature effect remains, which results from the finite number N of recorded signal, so that the holophonic error converges on the ambisonic error.

Therefore, for low values of S, the ambisonic encoding seems to be the more efficient and, for high values of S, Ambisonic and Holophony achieve equivalent error. However, it must be borne in mind that the Ambisonic sound pick-up for order greater than 1, i.e. recording of the $c_{2n-1} = \cos(n\theta)/c_{2n} = \sin(n\theta)$ signals, is –at least presently– unfeasible, whereas Holophony requires only pressure or velocity microphones. But the previous results (cf. equ.11) suggest that it is possible to derive the c_n signals from a holophonic sound pick-up based on a circular array of microphones, provided that a sufficient number of microphones is used. By including higher order signals, the accurate reconstruction area may be substantially extended (cf. Fig.2).

6 Conclusion

It has been shown that Holophony and Ambisonic are based on equivalent approaches of acoustical field reconstruction. In the light of this equivalence, it is intended to compare the two approaches, in terms of reconstruction performance relative to system complexity and implementation feasibility. This first study was focussed on sound pick-up and has pointed out that Ambisonic spatial information encoding is more efficient, but less practical. However, it should be noticed that, though directly unfeasible, ambisonic sound pick-up including high orders may be derived from a holophonic recording. Another consequence of the link drawn between the two methods is the possibility of separating sound pick-up and reproduction in order to derive a hybrid approach. Future research will complete this analysis, first by examining the sound reproduction step, secondly by studying the method robustness towards the actual behaviour of transducers.

7 Acknowledgements

The authors would like to thank Mr. J.-M. Jot for many helpful discussions and suggestions in connection with this paper.

References

[1] M. Miyoshi and N. Koizumi. NNT's research on acoustics for future telecommunication services. *Applied Acoustics*, 36:307–326, 1992.

[2] M. Jessel. *Acoustique théorique, propagation et holophonie*. Masson, Paris, 1973.

[3] M. Bruneau. *Introduction aux théories de l'acoustique*. Université du Maine, Le Mans, 1983.

[4] E. N. G. Verheijen. *Sound Reproduction by Wave Field Synthesis*. PhD thesis, Delft University of Technology, 1998.

[5] J. D. Murray. *Asymptotic Analysis*, chapter 4. Clarendon Press, Oxford, 1974.

[6] J. S. Bamford. *An Analysis of Ambisonic Sound Systems of First and Second Order*. PhD thesis, University of Waterloo, Ontario Canada, 1995.

[7] Ken Farrar. Soundfield microphone. *Wireless World*, pages 48–50, october 1979.

[8] J. Daniel, J.-B. Rault, and J.-D. Polack. Ambisonics encoding of other audio formats for multiple listening conditions. to be published in J.A.E.S.