Wave field synthesis: The next dimension

François Germain
iZotope, Inc.

March 9, 2021
What is this about?

We are here to talk about rendering soundfields.

Wait... Didn’t we do that already?

Well, sort of, Ambisonics is a way to render soundfields. But there are also other ways!
What is this about?

Errrrr… What?

Yes! There are different ways(/camps) on how to render soundfields

- Ambisonics
- Plane-wave decomposition
- Wave field synthesis
- ...

francois@ccrma.stanford.edu (iZotope)
A brief history

- Pioneered by A.J. Berkhout (a geophysicist!) out of TU Delft in 1988
- Many theoretical bricks brought along the years out of TU Delft, IRCAM and TU Berlin
- Today, research is showing up all around with large dense speaker array installations in more and more institutions, as well as traveling installations
- Sometimes also referred to as holophon (yes, holo- with the same meaning as in holography!)
Examples of installations

Royal Conservatoire, The Hague, The Netherlands
TU Berlin, Berlin, Germany
RPI, Troy, NY
U. Hamburg, Hamburg, Germany
NTT, Yokosuka, Japan
Game of Life installation

francois@ccrma.stanford.edu (iZotope)
Wave equation - Remember?...

- Wave equation:

\[ \nabla^2 p(r, t) - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}(r, t) = 0 \]

- Helmholtz equation:

\[ \nabla^2 P(r, \omega) + \frac{\omega^2}{c^2} P(r, \omega) = 0 \]

---

\( p \): acoustic pressure, \( P \): Fourier transform of acoustic pressure

\( r \): position
Huygens-Fresnel principle

Any physical system that follows the wave equation obeys the Huygens-Fresnel principle.

Huygens-Fresnel principle: A propagating wave can be recreated as the sum of elementary wave sources placed on its wavefront.
Kirchhoff-Helmholtz integral

*Kirchhoff-Helmholtz integral*: If you know the sound pressure and velocity in any point on the surface of a source-free volume, you have complete knowledge of the sound field inside.

*Rayleigh integral*: If you know the sound pressure and velocity in any point on an (infinite) plane bordering a source-free half-space, you have complete knowledge of the sound field inside.
The WFS promise

- “The only true multiuser solution” (kind of...)
- Simple formulation (kind of...): record on the surface, 1 microphone to 1 speaker, very flexible on arbitrary speaker setup

What does that look like mathematically?...
Green’s theorem

- Green’s theorem:

\[
\iint_V (\Phi \nabla^2 \Psi - \Psi \nabla^2 \Phi) \, dV = \iint_S \left( \Phi \frac{\partial \Psi}{\partial n} - \Psi \frac{\partial \Phi}{\partial n} \right) \, dS
\]

- Helmholtz equation and Green’s theorem

\[
\begin{align*}
\nabla^2 \Phi &= -k^2 \Phi \\
\nabla^2 \Psi &= -k^2 \Psi
\end{align*}
\Rightarrow \iint_V (\Phi \nabla^2 \Psi - \Psi \nabla^2 \Phi) \, dV = 0
\]

(The acoustic pressure field follows the Helmholtz equation, remember?...)

---

S: Surface enclosing volume \( V \)

\( n \): unit vector normal to \( S \) pointing inwards into \( V \)
Kirchhoff-Helmholtz integral equation

\[ G(r|r') \text{ the } Green's \text{ function} \text{ solution of :} \]

\[ \nabla^2 G(r|r') + k^2 G(r|r') = \delta(r - r') \]

Working through the math comes the Kirchhoff-Helmholtz integral equation:

\[ P(r') = \iint_{r \in S} \left( P(r) \frac{\partial G}{\partial n}(r|r') - G(r|r') \frac{\partial P}{\partial n}(r) \right) dS \]
Pressure and velocity field

\[ P(r') = \int\int_{r \in S} \left( P(r) \frac{\partial G}{\partial n}(r|r') - G(r|r') \frac{\partial P}{\partial n}(r) \right) dS \]

- \( P(r) \) is the pressure on the surface.
- From linear acoustics theory:

\[ \nabla P(r) = -j\omega \rho_0 U(r) \]

\[ \Rightarrow \frac{\partial P}{\partial n}(r) (= \nabla P(r).n) \] is proportional to the acoustic velocity projected on \( n \).

\( U \): Fourier transform of the acoustic velocity \( u \)
Pressure and velocity field

\[ P(r') = \int \int_{r \in S} \left( P(r) \frac{\partial G(r \mid r')}{\partial n} - G(r \mid r') \frac{\partial P(r)}{\partial n} \right) dS \]

- \( P(r) \) is the pressure on the surface.
- From linear acoustics theory:

\[ \nabla P(r) = -j \omega \rho_0 U(r) \]

\[ \Rightarrow \frac{\partial P}{\partial n}(r) (= \nabla P(r).n) \text{ is proportional to the acoustic velocity projected on } n. \]

\textbf{U}: Fourier transform of the acoustic velocity \( u \)
Wave field synthesis (Finally!!)

\[
P(r') = \iint_{r \in S} \left( P(r) \frac{\partial G}{\partial n}(r | r') - G(r | r') \frac{\partial P}{\partial n}(r') \right) dS
\]

⇒ We can reconstruct the sound field inside a closed volume free of acoustics sources through the superposition of:

- a **first source distribution** on \( S \) driven by the **pressure field** on \( S \);
- a **second source distribution** on \( S \) driven by the **normal velocity field** on \( S \).
Wave field synthesis (Finally!!)

\[ P(r') = \int \int_{r \in S} \left( P(r) \frac{\partial G}{\partial n}(r|r') - G(r|r') \frac{\partial P}{\partial n}(r) \right) dS \]

⇒ We can reconstruct the sound field inside a closed volume free of acoustics sources through the superposition of:

- a first source distribution on \( S \) driven by the pressure field on \( S \);
- a second source distribution on \( S \) driven by the normal velocity field on \( S \).
Free-field Green’s function

Free-field ("Anechoic") Helmholtz equation:

\[ \nabla^2 G + k^2 G = \delta(r - r') \quad \text{with} \quad G(r| r') \xrightarrow{|r| \to \infty} 0 \]

- \[ G(r| r') = \frac{e^{jk|r-r'|}}{4\pi|r-r'|} \Rightarrow \text{This is a monopole source expression.} \]
- \[ \frac{\partial G}{\partial n}(r| r') = \frac{\partial}{\partial n} \left[ \frac{e^{jk|r-r'|}}{4\pi|r-r'|} \right] \Rightarrow \text{This is a dipole source expression.} \]
Free-field wave field synthesis

\[ P(r') = \iint_{r \in S} \left( P(r) \frac{\partial}{\partial n} \left[ \frac{e^{ijk|r-r'|}}{4\pi|r-r'|} \right] - \frac{e^{ijk|r-r'|}}{4\pi|r-r'|} \frac{\partial P}{\partial n}(r) \right) dS \]

We can reconstruct the sound field inside a closed volume free of acoustics sources through the superposition of:

- a **monopole source distribution** driven by the pressure field;  
- a **dipole source distribution** driven by the normal velocity field.

⇒ This is the mathematical foundation of wave field synthesis.
Free-field wave field synthesis

\[ P(r') = \int\int_{r \in S} \left( P(r) \frac{\partial}{\partial n} \left[ \frac{e^{jk|\mathbf{r} - \mathbf{r}'|}}{4\pi |\mathbf{r} - \mathbf{r}'|} \right] - \frac{e^{jk|\mathbf{r} - \mathbf{r}'|}}{4\pi |\mathbf{r} - \mathbf{r}'|} \frac{\partial P}{\partial n}(r) \right) dS \]

We can reconstruct the sound field inside a closed volume free of acoustics sources through the superposition of:

- a monopole source distribution driven by the pressure field;
- a dipole source distribution driven by the normal velocity field.

⇒ This is the mathematical foundation of wave field synthesis
Free-field wave field synthesis

\[ P(r') = \iint_{r \in S} \left( P(r) \frac{\partial}{\partial n} \left[ \frac{e^{jk|r-r'|}}{4\pi|r-r'|} \right] - \frac{e^{jk|r-r'|}}{4\pi|r-r'|} \frac{\partial P}{\partial n}(r) \right) dS \]

We can reconstruct the sound field inside a closed volume free of acoustics sources through the superposition of:

- a monopole source distribution driven by the pressure field;
- a dipole source distribution driven by the normal velocity field.

⇒ This is the mathematical foundation of wave field synthesis.
Wave field synthesis in practice

\[ P(r') = \iint_{r \in S} \left( P(r) \frac{\partial}{\partial n} \left[ \frac{e^{jk|r-r'|}}{4\pi |r-r'|} \right] - \frac{e^{jk|r-r'|}}{4\pi |r-r'|} \frac{\partial P}{\partial n}(r) \right) dS \]

Problems?
Sources and Neumann Green’s function

Do I really need *two source types*??

Neumann’s Green function:

\[
\nabla^2 G_N(r|r') + k^2 G_N(r|r') = \delta(r - r') \quad \text{with} \quad \frac{\partial G_N}{\partial n}(r|r') = 0
\]

For a *planar* source distribution:

\[
G_N(r|r') = 2G(r|r')
\]

\[
\Rightarrow P(r') = -2 \int\int_{r \in S} G(r|r') \frac{\partial P}{\partial n}(r) dS
\]

(\Rightarrow \text{we only need to know/measure the acoustic velocity})

This source expression is also used for non-planar source distribution
(as a high-frequency approximation with *directional* truncation to limit artifacts)
Typical driving functions

- **Plane waves:**
  \[
  U(r) = j\omega n_w \frac{n_w}{c} e^{-j\omega \cdot \frac{n_w}{c} (r-r_s)} \quad \text{and} \quad u(r) = \frac{n_w}{c} \frac{ds}{dt} \left( t - \frac{n_w \cdot (r-r_s)}{c} \right)
  \]

- **Spherical waves:**
  \[
  U(r) = \left( \frac{1}{|r-r_s|} + \frac{j\omega}{c} \right) \frac{r-r_s}{|r-r_s|^2} e^{-j\omega \frac{r-r_s}{c}}
  \]
  \[
  u(r) = \frac{r-r_s}{|r-r_s|^2} \left[ \frac{1}{|r-r_s|} + \frac{1}{c} \frac{d}{dt} \right] s \left( t - \frac{|r-r_s|}{c} \right)
  \]

  https://www.youtube.com/watch?v=3ALsjtMNsiE&list=PL-CY-hgT9swxoRAf6dElilWsPEIghRz_2

- **Focused waves** (using **time reversal**):
  \[
  U(r) = \left( \frac{1}{|r-r_s|} + \frac{j\omega}{c} \right) \frac{r-r_s}{|r-r_s|^2} e^{+j\omega \frac{r-r_s}{c}}
  \]
  \[
  u(r) = \frac{r-r_s}{|r-r_s|^2} \left[ \frac{1}{|r-r_s|} + \frac{1}{c} \frac{d}{dt} \right] s \left( t + \frac{|r-r_s|}{c} \right)
  \]

  https://www.youtube.com/watch?v=7BSSn6zJGok&list=PL-CY-hgT9swzeM9x8wgE6PbtPEEE2eWf5
Sampling and spatial aliasing

Wait... *Did I see an integral in there?!*
We need to deal with *sampling* (Same as for ambisonics)

- Practical source distribution are *discrete*
- For audio, sampling creates *aliasing*: high frequencies get “folded” down
- Spatial sampling creates *spatial aliasing*: spurious wavefronts appear

Typical aliasing analysis is done for plane waves and linear arrays:

- Worst-case conditions: \( f_{alias} = \frac{c}{2\Delta x} \) for \( \Delta x \) speaker spacing
- Typical example: \( \Delta x \approx 15\text{cm} \approx 6\text{in} \Rightarrow f_{alias} \approx 1\text{kHz} \)
- Direction-dependent conditions (!!): \( f_{alias} = \frac{c}{\Delta x|1-\sin \theta|} \)

with \( \theta \) incidence angle of the wave

*Trick:* no “backward” wave

https://www.youtube.com/watch?v=3ALsjtMNsiE&list=PL-CY-hgT9swwgLLmk4jC_oQKaBVzKAYXo
Truncation

Do I really need to enclose the space with loudspeakers?!

Many loudspeaker configuration use truncated arrays.

Truncation creates “Gibbs-like” phenomenon with wave front ripples.

Truncation creates a sweet-spot (oops): The visibility area

To some extent, truncation mitigates spatial aliasing issues(!!).

Mitigation: Tapering window (at the expense of sweet-spot size)

https://www.youtube.com/watch?v=U81al_45LxQ&list=PL-CY-hgT9swzDPKBWQ8wuTiHsr2AAKH7T
Wait... Have we been talking about 2D source distributions all this time?!

Planar arrays are... expensive!!

Can we use linear arrays instead?

Yes... with some limitations. But how?!

We can interpret the linear array as the projection of the planar array, i.e., the horizontal plane as a projection of the 3D space.
In projected 2D (3D with 1D sources), The Kirchhoff integral equation is still valid but the free-field Green’s function changes:

$$G(r|r') = \frac{j}{4} H_0^{(2)} \left(\frac{\omega}{c} |r' - r|\right)$$

Approximation: $G(r|r') \approx \sqrt{d} \times \Gamma(\omega) \times \frac{e^{jk|r-r'|}}{4\pi|r - r'|}$

The source energy decays in $1/r$ instead of $1/r^2$

$$\Gamma(\omega) \propto \sqrt{\omega} \approx +3\text{dB/oct} \text{ (The array efficiency decreases with frequency)}$$

In summary: We can approximate things by adding a scaling and filter

Designing the scaling and filter is its own thread of research...

$H_0^{(2)}$: Hankel function of the second kind of order 0

$d$: Arbitrary reference distance from the array
Real sources and spaces (1)

Real sources are *not monopoles*:
- Closed-cabinet speakers are nearly *monopoles* at low frequency
- Closed-cabinet speakers are nearly *dipoles* at high frequency

Mitigation is similar to 2.5D WFS approach: Find an approximate compensating equalization and integrate that contribution into the driving function.

Real spaces are *reverberant*  
⇒ The Green’s function solution is incorrect: Aliased wave fronts are reflected too!...

Practical solution: (Realtime?) Equalization of WFS (requires microphone array or simulation)
My speakers are not well aligned, does that matter?
Yes and no, the errors get bigger as the frequency increases, but there’s some tolerance.

I don’t like fractional delays, can I do without?
Yes and no, there is evidence that non-fractional delays don’t degrade the sound that much.

And this spatial aliasing thing, what can we do then?
- Not much besides having more speakers.
- People have proposed hybrid methods to help.
- It also depends on the target sweet-spot (same as for Ambisonics and other methods)
- There is evidence it doesn’t matter too much, but it adds coloration
Current research

- 2.5D WFS: driving function design
- Complex acoustic sources
- Moving sources (Doppler effect,...)
- Spatial aliasing mitigation (hybrid methods...)
- Perceptual evaluation
The WFS promise?

- Large sweet-spot by design
- “Simple” basic formulation, easy to implement (delay and gain), 1 microphone to 1 speaker, very flexible on arbitrary speaker setup
- Focused sources!! (possible and easy to make)

**but...**

- A *little* math needed for precise use in practice (same as Ambisonics)
- Spread-out, somewhat complex error (aliasing—unlike Ambisonics—, reverberation—same as Ambisonics—,...)
- Requires object-based format, especially to scale up/down (unlike Ambisonics)
- Heavy hardware needs

https://www.youtube.com/watch?v=3ALsjtMNsiE&list=PL-CY-hgT9swvvZ_bB1TZubYQnraRz30CF
Some tools

- For research: Sound Field Synthesis toolbox (Matlab, Python, Julia)
  https://github.com/sfstoolbox

- For playing: SoundScape Renderer
  http://spatialaudio.net/ssr/
Parting thoughts

- Elevation reproduction?
  ⇒ Most proposed solution come down to hybrid systems

- Sound reinforcement

- Wave field interpolation/extrapolation

- WFS = general formulation of sound propagation theory

- At the end, *it all converges!* (WFS, HOA,...)
Questions?