Sound synthesis and spatial processing
Politecnico di Milano – Polo Regionale di Como
SUMMARY

- Wavefield Synthesis
  - strict foundation on the basic laws of acoustical wave propagation
- Ambisonics
  - represent the sound field by an expansion into three-dimensional basis functions
- Microphone Acquisition
Wavefield synthesis

Basic idea:

- reconstruct the soundfield within a volume by recreating the wavefront in a closed surface that contains it (through sampling)
Wavefield synthesis

- **Huygens principle:**
  “every points of a wavefront can be considered as the source of an elementary wave that interacts with the surrounding environment”

- The wavefield produced by a primary source $\Psi$ can be reconstructed using a distribution of secondary sources, for example obtained by placing the speakers $P_1, P_2, \ldots, P_M$
Wavefield synthesis
Wavefield synthesis

Let:

- \( Q(z, \omega) \) be the Fourier Transform of the emitting source;
- \( z \) be the vector of polar coordinates \((r, \theta, \phi)\) also referred as the listener position;
- \( P(z, \omega) \) be the Fourier Transform of the acoustic pressure in any point \( z \);
- \( V \) be the volume that encloses the listening area;
- \( S \) be the surface that encloses the volume \( V \).
Wavefield synthesis

The Huygens principle is mathematically described by the Kirchhoff-Helmholtz integral

\[ \iiint_{\mathbb{S}} \left[ P(z', \omega) \frac{\partial}{\partial n} G(z \mid z', \omega) - \frac{\partial}{\partial n} P(z', \omega) G(z \mid z', \omega) \right] dS = \begin{cases} P(z, \omega) & z \in \mathbb{V} \\ 0 & z \notin \mathbb{V} \end{cases} \]

where:

- \( G \) is the Green Function. The GF describes the contribution of a point source at position \( z' \) to the soundfield at position \( z \).
- Point source Green Function:

\[ G(z \mid z', \omega) = \frac{1}{4\pi} \frac{e^{-jk||z-z'||}}{||z-z'||} \]

- \( P(z', \omega) \) is the pressure along the enclosing surface \( \mathbb{S} \)
- \( \partial / \partial n P(z', \omega) \) is the pressure gradient along the surface normal
Wavefield synthesis

\[ \iiint_S \left[ P(z', \omega) \frac{\partial}{\partial n} G(z | z', \omega) - \frac{\partial}{\partial n} P(z', \omega) G(z | z', \omega) \right] dS = \begin{cases} P(z, \omega) & z \in V \\ 0 & z \notin V \end{cases} \]

The wavefield in a volume \( V \) is completely described by the pressure \( P(z', \omega) \) along the enclosing surface \( S \) and the pressure gradient \( \nabla P(z', \omega) \) along the surface normal \( n \).

According to the Huygens principle, we can interpret the above integral as a set of secondary sources along the surface \( S \).
Wavefield synthesis

- Choose $G$ as Green’s function of a point source (monopole)
  \[
  G(z | z', \omega) = \frac{1}{4\pi} \frac{e^{-jk\|z-z'\|/c}}{\|z-z'\|}
  \]
  \[
  \frac{\partial}{\partial n}(G) \quad \text{is the Green’s function of a dipole}
  \]
- Approximate the distribution of monopoles and dipoles on the surface $S$ by suitable types of loudspeakers.
- Excite the loudspeakers with suitable driving signals which are derived from $P(z', \omega)$ and $\nabla P(z', \omega)$
Acquisition system

- The holophonic acquisition system is derived from the Huygens principle and the Kirchhoff integral, and consists of the spatial sampling of
  - 1) pressure $P(z', \omega)$
  - 2) pressure gradient $\nabla P(z', \omega)$
- Pairs of microphones
  - omnidirectional for pressure
  - directional for gradient (figure eight)
- Pairs of speakers
  - Monopolar (closed speakers)
  - Dipolar (open speakers)
- As shown in the K-H integral, the omnidirectional microphone feeds the dipole while the il directional one feeds the non-dipolar (closed) speaker
acquisition system
acquisition system
Wavefield reconstruction

- Microphones are now replaced by speakers
  - driving signals are the same that were acquired by the microphones
- Speakers must be used in pairs with similar polar characteristics as the microphones used for the acquisition
- The perfect reconstruction is obtained when we can have a continuous distribution on $S$
- Distance btw microphones and btw speakers determines the minimum wavelength that can be recorded (Shannon)
# Wavefield reconstruction

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two kinds of sources at the boundary: monopoles and dipoles</td>
<td>Elimination of the dipoles</td>
</tr>
<tr>
<td>Generation of sound signals on the whole surface</td>
<td>Distribution of point sources only on a closed curve on the horizontal plane</td>
</tr>
<tr>
<td>Generation of a continuous source distribution</td>
<td>Discrete sources (loudspeakers)</td>
</tr>
</tbody>
</table>
Wavefield reconstruction

- Elimination of dipoles:
  - admit a certain sound pressure also for $z'$ outside $V$
  - use a modified Green’s function $G_1$ in order to verify the condition $\frac{\partial}{\partial n} G_1 = 0$ on $S$

$$- \iint_S \frac{\partial P(z', \omega)}{\partial n} G_1(z | z', \omega) dS = \begin{cases} P(z, \omega) & z \in V \\ \text{arbitrary} & z \notin V \end{cases}$$

only monopoles are now used

- cons:
  - speaker selection: windowing artifacts
  - soundfield generated outside $V$
Wavefield reconstruction

Reproduction of monochromatic plane wave …

…using all secondary sources

…using selected secondary sources
Wavefield reconstruction

• From 3D to 2D:
  • reconstruct the wavefield on a plane (listener head)

• Spatial Sampling:
  • Sampling the curve:
    • Trade-off: N. of speakers VS. Max sampled freq.

• We limit our bandwidth to frequencies below 1.5KHz
  (above 1.5 the damage is quite contained)
Wavefield reconstruction

Spatial sampling artifacts:

\[ f_{pw} = 1000 \text{ Hz} \]

\[ f_{pw} = 2000 \text{ Hz} \]
Example: 8-channel circular rig

- Wavefield synthesis requires an **anechoic** rendering environment
  - When the environment is not anechoic we can adopt a scheme with filters that are design to cancel (or minimize the impact of the environment)
Non-anechoic environments

- $G(z)$ can be determined through a calibration procedure
- Adapt a filter in such a way to minimize the error btw the acquired signal in the rendering room $y[n]$ (using microphones that are similar to those used during the recording) and the reference signal $d[n]$
  - Calibration signal is a pseudo-random noise
  - Filter lengths are proportional to reverberation time
Wavefield synthesis: applications

- Example of application: recreating the soundfield of a different environment
Wavefield synthesis: applications

- improving the acoustics of an auditorium by acquiring the sound and re-synthesizing it in the same environment
Wavefield synthesis

- using WFS together with beamforming it is possible to silence some areas
AMBISONICS
Ambisonics - overview

- Originally developed in the ’70s by Gerzon, Barton and Fellgett as a novel multi-channel recording-rendering technique with the aim of enhancing immersivity
  - With ambisonics the sound is encoded together with its directional components in a vector as a set of spherical harmonics
  - Similarly to holophonic systems, we can apply a transform matrix (decoder) to such vector in order to produce the signals that feed a speaker array that surrounds the environment
Ambisonics - overview

Recording

Soundfield Microphone

WXY(Z) coding

Decoding and conversion

2 ch. headphone

2 ch. LR

3/2

N ch.

Sweet-spot
Pros & cons

Main advantages
- Simple recording/coding method. We need multiple microphones (at least 4 for 3d – 3 for 2D) placed in the same point at the center of the acoustic scene;
- Independent coding/decoding methods
  - We can decode sound to produce a stereo signal, or a signal that is compatible with a standard 3/2 or with an N-channel system ($N=4, 6, 8, \ldots$)

Main disadvantages
- Sources can only produce planar wavefronts (fundamental assumption of Ambisonics)
  - Not restrictive as any wavefield can be modeled as the combination of planar wavefronts
- The rendered wavefield must be a combination of planar wavefronts (i.e. speakers must be sufficiently far from the listener)
  - Not a negligible problem for small environments such as cars
Ambisonics systems are based on transducers (mics) that produce a sampling of directional components of

- order zero (pressure microphones)
- order one (pressure gradient microphones)
- ...

Higher-order components correspond to a sampling of the acoustic field with many more (at least 9) super-directional microphones.
Ambisonics

- Directional characteristics of the wavefield are reconstructed by summing the spherical components of the field for orders $m=0,1,2$; each acquired with a microphone having similar directivity characteristics (from an omnidirectional microphone)
Coding formats

• A-format – used for coincident directional microphones (symmetrically oriented - soundfield)
• B-format – developed for studio equipment (most widespread)
  • Used also when microphones have similar directional characteristics as the spherical harmonics
  • Signal are coded into W,X,Y,Z
• C-format – for transmission
• UJT-format – for multi-channel coding (compatible with mono and stereo formats)
• D-format – for decoding and rendering
Coding formats

- When using a microphone for each harmonic component, signals coming from $W, X, Y, Z$ (B-format) can be encoded in any other format.
- $Z$ is omitted in 2D systems.
Theory

- **Hypothesis**
  - Sources are considered point source
  - Sources emit planar wavefronts

- **Notation**
  - We adopt cylindrical coord \((r, \phi, z)\)
  - Let us consider (for simplicity) the 2D case \((z=0)\)
  - Let \(k\) be the wave vector, of magnitude \(k=2\pi/\lambda\) (wave number) and same orientation as \(\Psi\)
  - \(k = [k_x \ k_y] = [\cos \Psi \ \sin \Psi]\)
  - Let \(r = (r, \phi)\) be the space vector

- The soundfield produced by a planar wave is

\[
P_{\Psi} e^{jk \cdot r}
\]

where \(P_{\Psi}\) is the pressure
Theory

- Pressure of reference field in a specific point \((r, \phi)\) is

\[ P_R(r) = P_\Psi e^{jkr} = P_\Psi e^{jkr\cos(\phi - \Psi)} \]

- Goal: to reproduce the planar wave relative to the listening reference pt by replacing the real source with a rendering system made of \(N\) speakers symmetrically arranged around the listening pt.
Theory

\[ P_R(r) = P_\psi e^{jk \cdot r} = P_\psi e^{jk r \cos(\phi - \Psi)} \]

**proof**

\[
\begin{align*}
  k &= [k_x \quad k_y] = [k \cos \psi \quad k \sin \psi] = k [\cos \psi \quad \sin \psi] \\
  r &= [r_x \quad r_y] = [r \cos \phi \quad r \sin \phi] = r [\cos \phi \quad \sin \phi] \\

  &\Rightarrow \\
  kr &= kr [\cos \psi \quad \sin \psi] \begin{bmatrix} \cos \phi \\ \sin \phi \end{bmatrix} \\
  kr &= kr (\cos \psi \cos \phi + \sin \psi \sin \phi) \\
  &\Rightarrow \\
  kr &= kr \cos(\psi - \phi) = kr \cos(\phi - \psi)
\end{align*}
\]
Rendering

- Each speaker generates a wavefield

\[ P_n(r, \phi_n) = P_n e^{jkr} = P_n e^{jkr\cos(\phi - \phi_n)} \]

- At the listening pt we have

\[ P_A(r) = \sum_{n=1}^{N} P_n(r, \phi_n) = \sum_{n=1}^{N} P_n e^{jkr\cos(\phi - \phi_n)} \]

- To reproduce the original field we need

\[ P_A(r) \equiv P_R(r) \]
The planar wave is characterized by a constant pressure and a phase component that can be expanded using a Fourier-Bessel series.

\[
P_R(r) = P_\Psi e^{jk r \cos(\phi - \Psi)} = \\
= P_\Psi J_0(kr) + 2P_\Psi \sum_{m=1}^{\infty} j^m J_m(kr) \cos[m(\phi - \Psi)] = \\
= P_\Psi \left( J_0(kr) + 2 \sum_{m=1}^{\infty} j^m J_m(kr) \cos(m\phi) \cos(m\Psi) + \\
+ 2 \sum_{m=1}^{\infty} j^m J_m(kr) \sin(m\phi) \sin(m\Psi) \right)
\]
Synthesis phase

- Truncate series as we have N speakers only

\[
P_A(\mathbf{r}) = \sum_{n=1}^{N} P_n(\mathbf{r}, \phi_n) = \sum_{n=1}^{N} P_n e^{jkr \cos(\phi - \phi_n)}
\]

\[
= \sum_{n=1}^{N} P_n \left( J_0(kr) + 2 \sum_{m=1}^{\infty} j^m J_m(kr) \cos(m\phi) \cos(m\phi_n) + \right.
\]

\[
+ 2 \sum_{m=1}^{\infty} j^m J_m(kr) \sin(m\phi) \sin(m\phi_n) \right)
\]

\[
P_A(\mathbf{r}) \equiv P_R(\mathbf{r})
\]

Matching conditions

\[
P_\Psi = \sum_{n=1}^{N} P_n;
\]

\[
P_\Psi \cos(m\Psi) = \sum_{n=1}^{N} P_n \cos(m\phi_n);
\]

\[
P_\Psi \sin(m\Psi) = \sum_{n=1}^{N} P_n \sin(m\phi_n).\]
Facts

- As N tends to infinity
  - the order-N approximation tends to an exact matching, so the synthesized wavefront is exactly the same as the reference one
  - the sweet spot expands
Vector-form expression for analysis

- $P_r(r)$ can be express in the vector-form

$$
\mathbf{c}^T = \begin{bmatrix}
1 & \sqrt{2} \cos \Psi & \sqrt{2} \sin \Psi & \cdots & \sqrt{2} \cos m\Psi & \sqrt{2} \sin m\Psi & \cdots
\end{bmatrix};
$$

$$
\mathbf{h}^T = \begin{bmatrix}
J_0(kr) & j\sqrt{2}J_1(kr) \cos \phi & j\sqrt{2}J_1(kr) \sin \phi & \cdots \\
j^m\sqrt{2}J_m(kr) \cos m\phi & j^m\sqrt{2}J_m(kr) \sin m\phi & \cdots
\end{bmatrix}.
$$

$$
\mathbf{P}_R(r, \phi) = \mathbf{P}_\Psi \mathbf{c}^T \mathbf{h}
$$

- We decoupled the direction information of the plane wavefront ($\Psi$) from information on the soundfield ($\phi$)
- The information on the spatial distribution of the plane wavefront is all in $c$
Vector-form expression for analysis

\[ P_R(r, \phi) = P_\Psi c^T h \]

\[
c^T h = J_0(kr) + \sqrt{2} \cos \psi \cdot j \sqrt{2} J_1(kr) \cos \phi + \sqrt{2} \sin \psi \cdot j \sqrt{2} J_1(kr) \sin \phi + \ldots + \\
+ \sqrt{2} \cos m \psi \cdot j^m \sqrt{2} J_m(kr) \cos m \phi + \sqrt{2} \sin m \psi \cdot j^m \sqrt{2} J_m(kr) \sin m \phi
\]

\[
c^T h = J_0(kr) + 2j J_1(kr) \cos \psi \cos \phi + 2j J_1(kr) \sin \psi \sin \phi + \ldots + \\
+ 2j^m J_m(kr) \cos m \psi \cos m \phi + j^m J_m(kr) \sin m \psi \sin m \phi = \\
= J_0(kr) + 2 \sum_{m=1}^{\infty} \left(j^m J_m(kr) \cos m \psi \cos m \phi + j^m J_m(kr) \sin m \psi \sin m \phi \right)
\]

\[
c^T h = J_0(kr) + 2 \sum_{m=1}^{\infty} \left(j^m J_m(kr) \cos[m(\phi - \psi)] \right)
\]
Vector-form expression for analysis

- First components of vector $c$ multiplied $P_\Psi$

\[
P_\Psi c_0 = W = P_\Psi \\
P_\Psi c_1 = X = P_\Psi \sqrt{2} \cos \Psi \\
P_\Psi c_2 = Y = P_\Psi \sqrt{2} \sin \Psi
\]

- Which are for $m=0$ (zero order) and $m=1$ (first order)
Vector-form expression for analysis

- In the acquisition phase (mics) we need to identify the coefficients $c_m$ of vector $c$
  - In principle we need $m=0,\ldots,\infty$
  - In practice, we need to truncate the expression
- In order to measure $c_m$ we would need mics whose directivity matches $\cos(m\Psi) e^{\sin(m\Psi)}$. Except for $m=0$ (omnidirectional mic) and $m=1$ (bi-directional mics), for $m\geq2$ there are no matching mics
  - In practice, we stop at order 1 or 2
  - For high order we need to mathematically determine coefficients $c_m$
Fig. SH: Cortical surface decomposition with 256, 48, 26, 14, 8 und 4 degrees.

© 2009 Ziegler@http://dbm.neuro.uni-jena.de
Vector-form expression for synthesis

In the synthesis phase

\[ P_A(r) = \sum_{n=1}^{\bar{N}} P_n (r, \phi_n) = \sum_{n=1}^{\bar{N}} P_n e^{jkr\cos(\phi-\phi_n)} \]

\[ = \sum_{n=1}^{\bar{N}} P_n \left( J_0(kr) + 2 \sum_{m=1}^{\infty} j^m J_m(kr) \cos(m\phi) \cos(m\phi_n) + + 2 \sum_{m=1}^{\infty} j^m J_m(kr) \sin(m\phi) \sin(m\phi_n) \right) \]

\[ c_n^T = \begin{bmatrix} 1 & \sqrt{2} \cos \phi_n & \sqrt{2} \sin \phi_n & \cdots & \sqrt{2} \cos m\phi_n & \sqrt{2} \sin m\phi_n & \cdots \end{bmatrix} \]
Vector-form expression for synthesis

- Let $P_n(r, \phi) = P_n c_n^T h$ be the soundfield produced by the $n$-th speaker.
- We have

$$P_A(r) = \sum_{n=1}^{N} P_n(r, f_n) = \sum_{n=1}^{N} P_n c_n^T h = A^T C^T h$$

$$C = [c_1 \quad c_2 \quad \ldots \quad c_N]$$

$$A^T = [P_1 \quad P_2 \quad \ldots \quad P_N]$$
Coding

- The recorded sound contains the whole information required for the coding
- So the *Ambisonic signal* contains the whole information (magnitude and direction) of the sources
  - E.g.: the 2D Ambisonic signal for a source $P_\Psi$ oriented as $\Psi$ is given by

\[
\begin{align*}
  m = 0 & \quad W = P_\Psi / \sqrt{2} \\
  m = 1 & \quad X = P_\Psi \cos \Psi \quad Y = P_\Psi \sin \Psi \\
  m = 2 & \quad U = P_\Psi \cos 2\Psi \quad V = P_\Psi \sin 2\Psi \\
  \vdots & \quad \vdots
\end{align*}
\]
• In ambisonics recordings the order $M$ is extremely limited
  • Typical choice is $M=1$ but sometimes $M=2$ or $M=3$ are used for experimental reasons
  • Truncation implies a limited listening region: as $M$ increases, the sweet spot grows
  • Truncation implies a limited localization
• The notation $(W,X,Y,U,V)$ is the typical one of the 2D case. The harmonics along $z$ are, in fact, zero
  • $m=1$ implies $Z=0$
  • $m=2$ implies $R=S=T=0$
Decoding

We start from the matching conditions

\[ P_\Psi = \sum_{n=1}^{N} P_n ; \]
\[ P_\Psi \cos(m\Psi) = \sum_{n=1}^{N} P_n \cos(m\phi_n) ; \]
\[ P_\Psi \sin(m\Psi) = \sum_{n=1}^{N} P_n \sin(m\phi_n) . \]

and we want

\[ P_A (r) = A^T C^T h \]
\[ P_R (r) = P_\psi c^T h \]

Where:

\[ A^T C^T h = P_\psi c^T h \]

we can define \( b = \begin{bmatrix} W & X & Y & (Z) & (R) & (S) & (T) & U & V & \ldots \end{bmatrix} \)

\[ b^T = A^T C^T \quad \Rightarrow \quad b = CA \]
Decoding

We thus have \( \mathbf{b} = \mathbf{C} \mathbf{A} \)
where
\[
\mathbf{b} = \begin{bmatrix} W & X & Y & (Z) & (R) & (S) & (T) & U & V & \ldots & \end{bmatrix}
\]
\[
\mathbf{A}^T = [P_1 \quad P_2 \quad \ldots \quad P_N]
\]

- \( \mathbf{C} \) is dependent on the speaker positions

\[
\mathbf{C} = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
\cos \phi_1 & \cos \phi_2 & \cdots & \cos \phi_N \\
\sin \phi_1 & \sin \phi_2 & \cdots & \sin \phi_N \\
\vdots & \vdots & \ddots & \vdots \\
\cos m\phi_1 & \cos m\phi_2 & \cdots & \cos m\phi_N \\
\sin m\phi_1 & \sin m\phi_2 & \cdots & \sin m\phi_N
\end{bmatrix}
\]

The only element we don’t know is \( \mathbf{A} \) which is the vector of the unknowns magnitudes that drive the speakers

- As \( \mathbf{C} \) is not square, we need to resort to pseudo-inversion (least mean square solution)

\[
\mathbf{A} = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{b}
\]
Speaker placement

• Speaker signals are determined by the Ambisonics weights according to their spatial location
• When speakers are placed on a circle (2D case) we have

\[ A = (C^T C)^{-1} C^T b \]

\[ P_n = \frac{1}{N} \left( W + 2X \cos \phi_n + 2Y \sin \phi_n + 2U \cos \phi_n + 2V \sin \phi_n + \cdots \right) \]

• Likewise, a closed-form expression can be found for speakers placed on a spherical surface (3D case)
Remarks

- Coding (recording) format \((W,X,Y,Z,\ldots)\) is independent on the No. and on the configuration of the rendering speakers
  - A large number of equidistant and uniformly distributed speakers guarantees a larger sweet spot and a more stable localization of sound
  - Many speakers guarantee better realism
- Asymmetric speaker placement requires a different weighting of components coming from
  \[
P_n = \frac{1}{N}(\alpha W + 2\beta X \cos\phi_n + 2\gamma Y \sin\phi_n + 2\delta \cos\phi_n + 2\varepsilon \sin\phi_n + \ldots)
\]
- Speaker characteristics must be as similar as possible
Remarks

- For a good 2D rendering we need at least 4 speakers
- For a correct localization we need 5 or more speakers
- For 3D Ambisonics (complete sphere) we need 6-8 speakers at least
- There are currently systems based on over 128 speakers
Comparative remarks

• Both holophonic systems (wavefield synthesis) and Ambisonics attempt a complete characterization of the soundfield but they profoundly differ in their approach.

• We consider a number of speakers \( N=S \) number of mics, and the order given \( 2M+1 \)
  • Holophony is based on spatial sampling of the wavefield
    • Spatial aliasing
    • For small \( S \), due to aliasing, the error is uniformly distributed over the volume
    • Increasing \( S \) the error has a fast decrease
  • Ambisonics truncate the spherical harmonic series
    • Approximation that narrows the sweet spot
    • For \( M \) low, the error is minimal at the center of the sweet spot and it increases as we move out of it
    • Increasing \( M \) the sweet-spot increase and the error decrease
Comparative remarks

- For small S Ambisonic work better than holophonic approach.
- For high order they are almost equivalent
- The difference is in mics acquisition:
  - In Ambisonic for high order we need specific directional mic (they could even not exist)
  - In Holophony we can increase the number of Omnidirectional and “otto” mics
References

Microphone acquisitions
Microphones

- The mic is an engine which produces an output voltage $v(t)$ proportional to some scalar or vector acoustic variable $g(t)$

- Microphones can be classified according to the law of proportionality btw acoustic and electric variable:
  - Microphones that are sensitive to the \textit{(scalar)} acoustic pressure
    - Non-directional (omni-directional) microphones
  - Microphones that are sensitive to the \textit{(vector)} pressure gradient
    - Directional microphones whose characteristics depend on the order $m$ of the gradient
Directional characteristics

- Radiation diagrams

a) Pressure microphone (omni-directional)
b) Low-order pressure gradient microphone (directional)
c) High-order pressure gradient microphone (very directional)

NB: radiation diagrams depend on the frequency and generally we have different diagrams for different frequencies
It is possible to construct alternative radiation patterns through:

- combining and connecting several microphone capsules (possibly with different characteristics) within the same device
- Adopting particular physical geometries around the capsule(s)
- E.g. Cardioid pattern can be obtained through a series connection of an omnidirectional capsule with a figure-of-eight combined in the same enclosure.
Acquisition techniques

- Three different strategies
  - Directional microphones in close proximity (coincident)
    - intensity depends on orientation only
    - referred to as *Intensity difference* $\Delta I$.
  - Omni-directional microphones placed at a distance of tens of centimeters from one another
    - referred to as *time difference* $\Delta T$
    - As microphones are omni-directional, intensity difference is negligible
  - Mixed microphones: mix directional microphones (cardioid or 8-figure) placed at a certain distance from one another
    - we have both a $\Delta T$ and a $\Delta I$
Intensity stereophony

- XY techniques (intensity stereophony)
  - two directional microphones at the same place, and typically pointing at an angle 90° or more to each other
  - A stereo effect is achieved through differences in sound pressure level between two microphones.
    - Due to the lack of differences in time-of-arrival / phase-ambiguities, the sonic characteristic of XY recordings is generally less spacy and has less depth compared to recordings employing an AB-setup
  - When the microphones are bidirectional and placed facing +/-45° with respect to the sound source the XY-setup is called a Blumlein pair.
    - The sonic image produced by this configuration is considered by many authorities to create a more realistic, almost holographic soundstage
Intensity stereophony

(a) $\Delta I$ pair (XY config.) with cardioid mics
(b) $\Delta I$ pair (XY config.) with figure 8 mics (Blumlein pair)
(c) $\Delta I$ pair with one cardioid and one 8-figure (Mid-Side configuration)
Time-of-Arrival Stereophony

- A-B techniques (time-of-arrival stereophony)
  - two parallel omnidirectional microphones some distance apart, so capturing time-of-arrival stereo information as well as some level (amplitude) difference information, especially if employed in close proximity to the sound source(s)

  - At a distance of about 50 cm the time delay for a signal reaching first one and then the other microphone from the side is approximately 1.5 msec (1 to 2 msec)
Acquisition techniques

(e) $\Delta T$ pair (A-B) with omni-directional mics
Near-coincident technique: mixed stereophony

- the ORTF stereo technique of the Office de Radiodiffusion Télévision Française = Radio France, calls for a pair of cardioid microphones placed 17 cm apart at a total angle between microphones of 110 degrees which, according to Eberhard Sengpiel, results in a stereophonic pickup-angle of 96°.
- Notice that the spacing of 17 cm has nothing to do with human ear distance. The recorded signals are generally intended for playback over stereo loudspeakers and not for ear phones.
Acquisition techniques

(f) $\Delta T - \Delta I$ pair (A-B) w/ cardioids (ORTF - Office de Radiodiffusion Television Française)
Examples of ORTF rigs
Acquisition techniques

For multi-channel application

• $\Delta T$ and mixed techniques use one mic for each reproduction channel (holophonic)

$$c_i(t) = g_i m_i(t); \quad i=1, 2, \ldots, N$$

where $g_i$ is a constant, $c_i$ is the signal to feed to the $i$th speaker and $m_i$ the signal captured from the $i$th mic.

• $\Delta I$ (coincident) techniques can use a smaller number of mics than the number of speakers – The signal fed into speakers are the obtained by linear combination of signal captured from mics, through a coding matrix

$$c_i(t) = \sum_{k=1}^{M} g_{ik} m_k(t); \quad i=1, 2, \ldots, N$$

where $M$ is the number of mics.
Surround Atmo-Mikrofon

- “Surround-Atmo-Mikrofon” – composed by 4 cardioid mics
- Much used to record concerts
Multi-channel Mid-Side

- Mid-Side (MS)
- It uses the two cardioid mics ($M_f$ and $M_b$) and one “otto” figure mic ($S$)
- It possible to build a 5 channel system (with C central – L Left – R Right – Ls Left back – Rs Right back)
Multi-channel Mid-Side

\[
\begin{bmatrix}
C \\
L \\
R \\
L_S \\
R_S
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & \alpha \\
1 & 0 & -\alpha \\
0 & 1 & \beta \\
0 & 1 & -\beta
\end{bmatrix} \begin{bmatrix}
M_F \\
M_B \\
S
\end{bmatrix}
\]

With $\alpha$ and $\beta$ constant
Microphone arrays

• Discrete arrays
Discrete microphone arrays

- Sets of discrete microphones with various geometries and appropriate directivity patterns
3D arrays

- Soundfield microphone
High-order arrays

- Trinnov’s 3D array with 24 microphones arranged in a pseudo-random fashion
3°-order directivity responses

<table>
<thead>
<tr>
<th>l=0</th>
<th></th>
<th></th>
<th>m_l=0</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>l=1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>l=2</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>l=3</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
<td><img src="image25.png" alt="Image" /></td>
</tr>
</tbody>
</table>
High Order Ambisonics (HOA)

- France Telecom also developed spherical arrays with 32 microphones (4th order)
Ambisonics and WFS acquisition and rendering
Ambisonics array

- The directional 3D diagram of the first-order Ambisonics is made of an omnidirectional figure and by three figures eight, one per axis
  - One pressure microphone (omnidirectional)
  - Three gradient pressure microphones
- Sounds coming from mics are those of the B format
  - Omnidirectional mic: $W$ signal
  - Directional mics: $X,Y$ and $Z$ signals
The ambisonics acquisition system is a coincident array. In order to avoid mic cluttering, we can use the so-called **Soundfield microphone** of type $\Delta l$

- First-order approximation, the microphone is made of 4 cardioid directional cartridges placed on a tetrahedron (one mic per direction)
- The soundfield mic picks up 4 signals: RF right-front, RB right-back, LF left-front and LB left-back
- This set of signals is called A-format
B-format can be derived by the A-format as

\[
\begin{bmatrix}
W \\
X \\
Y \\
Z
\end{bmatrix} =
\begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
L_F \\
L_B \\
R_F \\
R_B
\end{bmatrix}
\]
Limits and other formats

- Because of the discrete distribution of microphones, there is an upper bandwidth limit on the 4 components of the B-format.
- From [W X Y Z] is possible to derive other formats.
  - With the so called UHJ encoding matrix is possible to obtain various broadcast formats from the B-Format.

\[
\begin{align*}
\Sigma &= 0.9397W + 0.1856X \\
\Delta &= j(-0.342W + 0.5099X) + 0.655Y \\
T &= j(-0.1432W + 0.6512X) - 0.7071Y \\
Q &= 0.9772Z \\
R &= \frac{\Sigma - \Delta}{2} \\
L &= \frac{\Sigma + \Delta}{2}
\end{align*}
\]
Second-order ambisonics

- First-order ambisonics with coincident microphones determines a panning law that is quite far from ideality
- In order to better approximate ideal directivity patterns we can adopt a second-order approximation
  - The second order coincident array is made of 12 cardioid or hyper-cardioid microphones arranged on a dodecahedron
Problems and limits

- Coincident microphone arrays are used for stereo or surround sound recording but there are some factors that affect the correct reconstruction of the spatial acoustic image
  - Distance btw microphones is 3-10 cm instead of less than 5 mm
  - Microphone directivity diagrams are not ideal (especially at high frequencies)
Altering directivity

In the first-order case we can construct a tf function matrix \( G \) to perform beamforming so that the new signals \( X', Y', Z', W' \) will exhibit the desired polar diagram.
Altering directivity

- Correction of tetrahedron response

\[ G_1(z) \] for the omnidirectional signal \( W \)
\[ G_2(z) \] for the figure eight signals \( X, Y, Z \)