YAMAHA
Make Waves

Immersive Sound Technology

A Comparative Study of DBAP, VBAP, and WFS

Hiraku Okumura

Research & Development Division
Yamaha Corporation

October 28th, 2025

Immersive Audio

Spatial Audio Technology

YAMAHA

Sound Sources

Channel-based Audio

Traditional audio format using fixed speaker channels (e.g., stereo, 5.1 surround).

Object-based Audio

Audio elements are treated as independent objects with metadata.

Scene-based Audio

Captures and reproduces the entire sound field.

Playback

Psychoacoustics-Based Methods

Based on human auditory perception.

- Conventional 2D Panning
- 3D Panning:
  Vector Based Amplitude Panning (VBAP)
  Distance Based Amplitude Panning (DBAP)

Physical Acoustics-Based Method

Based on physical modeling of sound propagation.

- Wave Field Synthesis (WFS)
- Pressure Matching Method
- HOA(Mode Matching Method)
- Binaural (for headphone/earphone listening)

Spatial Audio Technology

YAMAHA

Sound Sources

Channel-based Audio

Traditional audio format using fixed speaker channels (e.g., stereo, 5.1 surround).

Object-based Audio

Audio elements are treated as independent objects with metadata.

Scene-based Audio

Captures and reproduces the entire sound field.

Playback

Psychoacoustics-Based Methods

Based on human auditory perception.

- Conventional 2D Panning
- 3D Panning:
  Vector Based Amplitude Panning (VBAP)
  Distance Based Amplitude Panning (DBAP)

Physical Acoustics-Based Method

Based on physical modeling of sound propagation.

- Wave Field Synthesis (WFS)
- Pressure Matching Method
- HOA(Mode Matching Method) *for reference
- Binaural (for headphone/earphone listening)

Psycho-acoustically based method

YAMAHA

- **Vector Based Amplitude Panning (VBAP)**
  - Based on the **direction vector from the listener to the virtual source**, VBAP assigns gain weights to two (in 2D) or three (in 3D) loudspeakers.

  - Solve for gain vector $\mathbf{g} = [g_1 \quad g_2 \quad g_3]$ such that

$$\mathbf{p} = g_1\mathbf{l}_1 + g_2\mathbf{l}_2 + g_3\mathbf{l}_3$$

Subject to:

$$g_i \geq 0$$
$$\|\mathbf{g}\| = 1$$

  - This can be solved as:

$$\hat{\mathbf{g}} = \mathbf{p}\begin{bmatrix}\mathbf{l}_1\\ \mathbf{l}_2\\ \mathbf{l}_3\end{bmatrix}^{-1}$$

$$\mathbf{g} = \frac{\hat{\mathbf{g}}}{\|\hat{\mathbf{g}}\|}$$

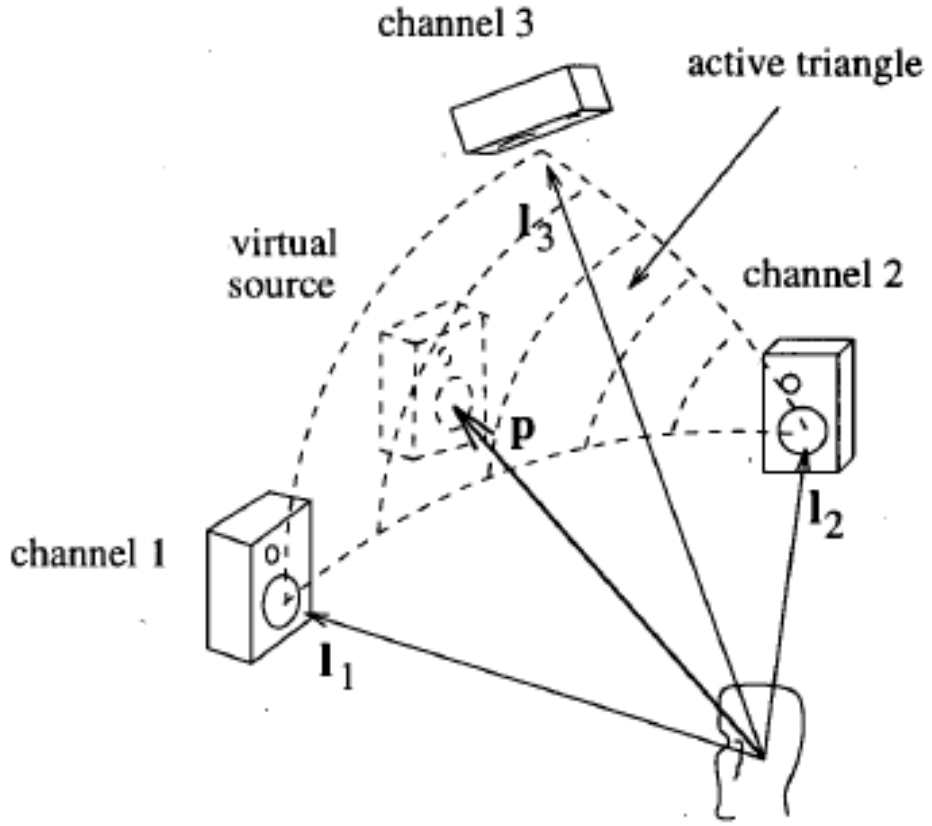


Fig. 5. Sample configuration for three-dimensional amplitude panning. Loudspeakers form a triangle into which the virtual source can be placed.

(Pulkki, 1997)

Psycho-acoustically based method

- **Vector Based Amplitude Panning (VBAP)**
  - The **sweet spot is relatively wide**; however, for listeners positioned farther from the intended position, localization errors increase.
  - VBAP **does not simulate depth**, causing virtual sources to appear confined to the speaker surface.
    To create a sense of distance, additional processing—such as adding reverberation—is required.
  - The **perceived size of the phantom source** varies depending on the virtual source direction and speaker positions, which may result in unnatural impressions during source movement.
  - Collision detection between the vector from the listener's position to the virtual source and the triangles formed by speaker triplets is required to determine which speakers to use.
    Due to numerical errors, the vector may fall between triangles, so additional handling is needed to prevent selection failures.
    The collision detection algorithm differs between 2D and 3D, as it involves checking against line segments in 2D and triangular surfaces in 3D.

Psycho-acoustically based method

- **Distance Based Amplitude Panning (DBAP)**
  - DBAP assigns gain weights to each speaker based on the **distance between the virtual source and the speakers.**

    Unlike VBAP, DBAP does not rely on fixed speaker pairs or triplets; instead, it considers all speakers simultaneously.

  - With N speakers, the distance d_i from the virtual source to the i-th speaker is

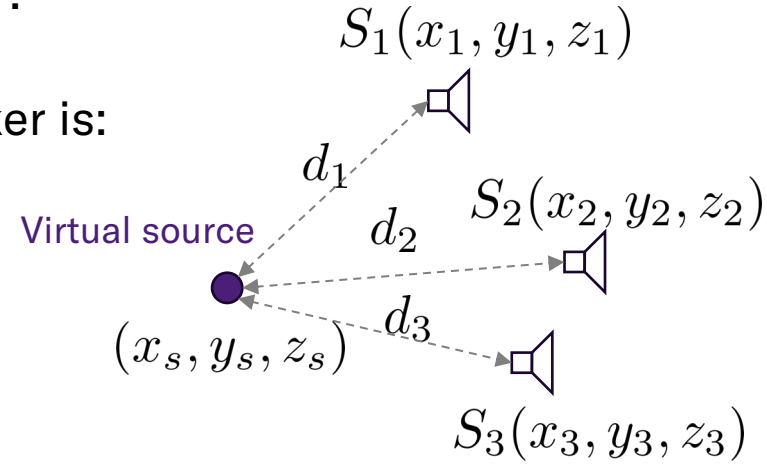
$$d_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2 + (z_i - z_s)^2 + r_s^2}$$

    r_s is a spatial blur factor to prevent gain divergence when $d_i \simeq 0$.

  - Using an attenuation coefficient $a = \dfrac{R}{20 \log_{10} 2}$, gain for each speaker is:

$$A = \sum_{j=1}^{N} \frac{1}{d_j^{2a}}$$

$$g_i = \frac{1}{\sqrt{A}} \frac{1}{d_i^a}$$



  - In a free-field 3D space, $R = 6$ dB corresponds to natural attenuation. R is usually set to $3 - 5$dB in real environments due to room reverberation.

Psycho-acoustically based method

- **Distance Based Amplitude Panning (DBAP)**

  - The **sweet spot is relatively wide**, and since there is no predefined listening position, spatial effects are preserved even for the listeners positioned away from the center of the listening area.

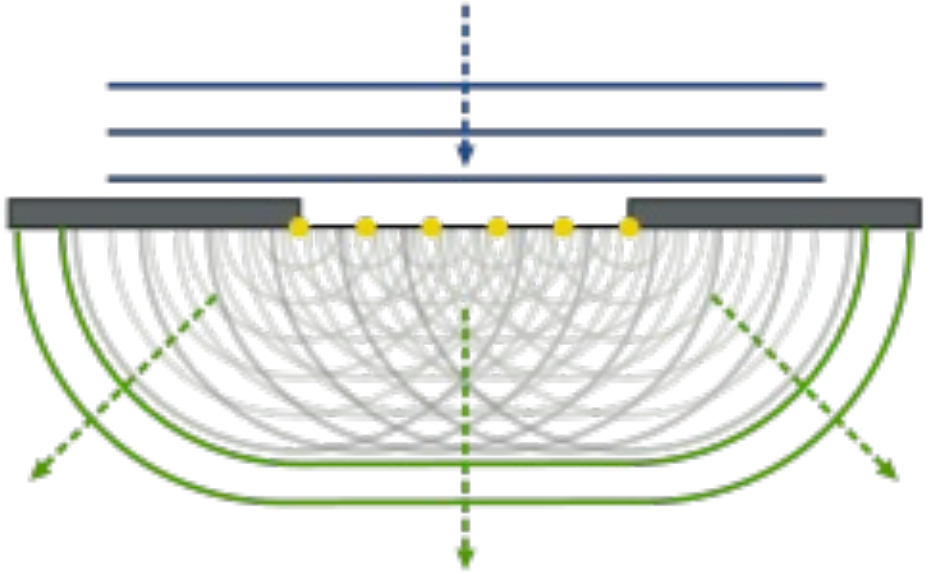
  - DBAP **does not simulate depth directly**; however, virtual sources tend to become **blurred** when they are positioned far from the speaker surface.
    To enhance depth perception, additional effects such as reverberation can be applied.

  - The **perceived size of the phantom source varies** depending on the virtual source and speaker positions, but this can be **mitigated more easily** than with VBAP.

Physical Acoustics-Based Method

- **Wavefront Propagation Principle**
  Wavefront propagation, such as sound waves, can be explained using the **Huygens–Fresnel principle**. At each moment, every point on a wavefront emits spherical wavelets. These wavelets combine to form the next wavefront, and repeating this process results in wave propagation.

  >> This concept forms the basis of **Wave Field Synthesis (WFS)**.



https://en.wikipedia.org/wiki/Huygens%E2%80%93Fresnel_principle

Physical Acoustics-Based Method

- **Mathematical Foundation**

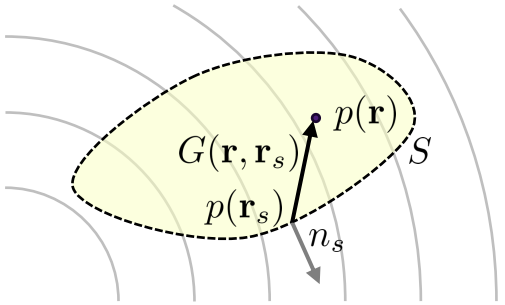
  This process is mathematically described by the **Kirchhoff–Helmholtz integral**, which shows that the sound field inside a boundary is determined by the sound pressure and particle velocity on its surface.

$$p(\mathbf{r}) = \int_S \left[p(\mathbf{r}_s) \frac{\partial G(\mathbf{r}, \mathbf{r}_s)}{\partial n_s} - G(\mathbf{r}, \mathbf{r}_s) \frac{\partial p(\mathbf{r}_s)}{\partial n_s} \right] dS$$

Sound pressure on the surface S.

Particle velocity on the surface S.

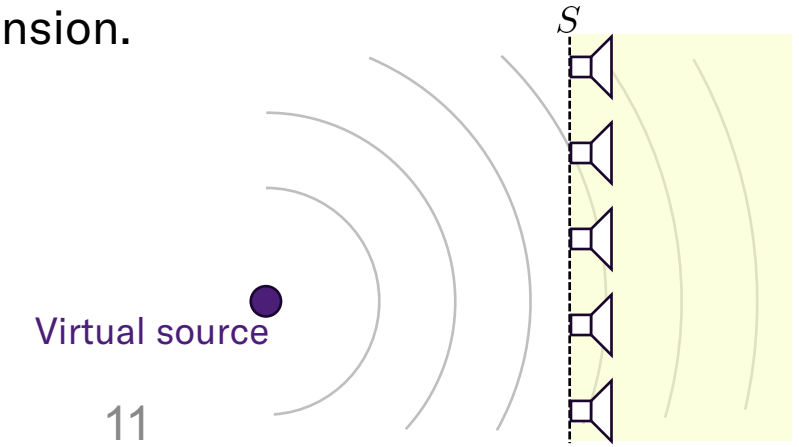
$$G(\mathbf{r}, \mathbf{r}_s) = \frac{e^{ik|\mathbf{r}-\mathbf{r}_s|}}{4\pi|\mathbf{r}-\mathbf{r}_s|}$$



  >> This is the foundation of the **Pressure Matching Method**.

- **Application to Linear Arrays**

  When the boundary is treated as an infinite plane, this leads to a **linear array WFS** model — a planar array simplified by ignoring the vertical dimension.



Physical Acoustics-Based Method

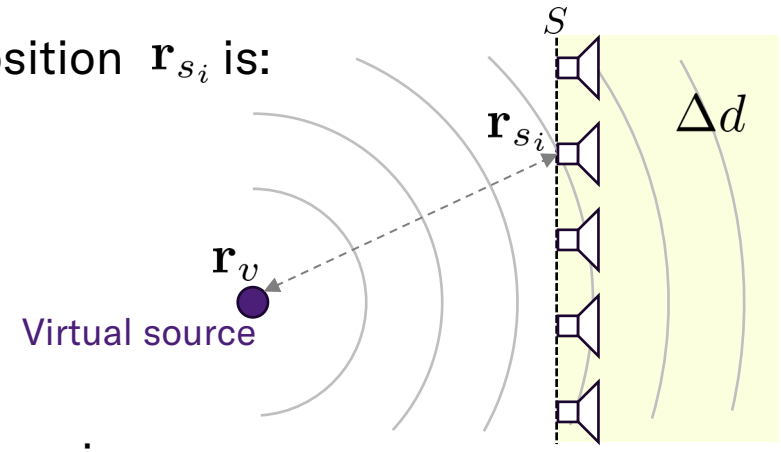
- **Wave Field Synthesis (WFS) for Linear Array**
  - For a virtual source at position $\mathbf{r}_v$, the gain for the i-th speaker at position $\mathbf{r}_{s_i}$ is:

$$g_i = \frac{e^{-jk|\mathbf{r}_{s_i} - \mathbf{r}_v|}}{\sqrt{|\mathbf{r}_{s_i} - \mathbf{r}_v|}}$$

Delay

Amplitude

where, $k = \frac{2\pi}{\lambda}$: wavenumber



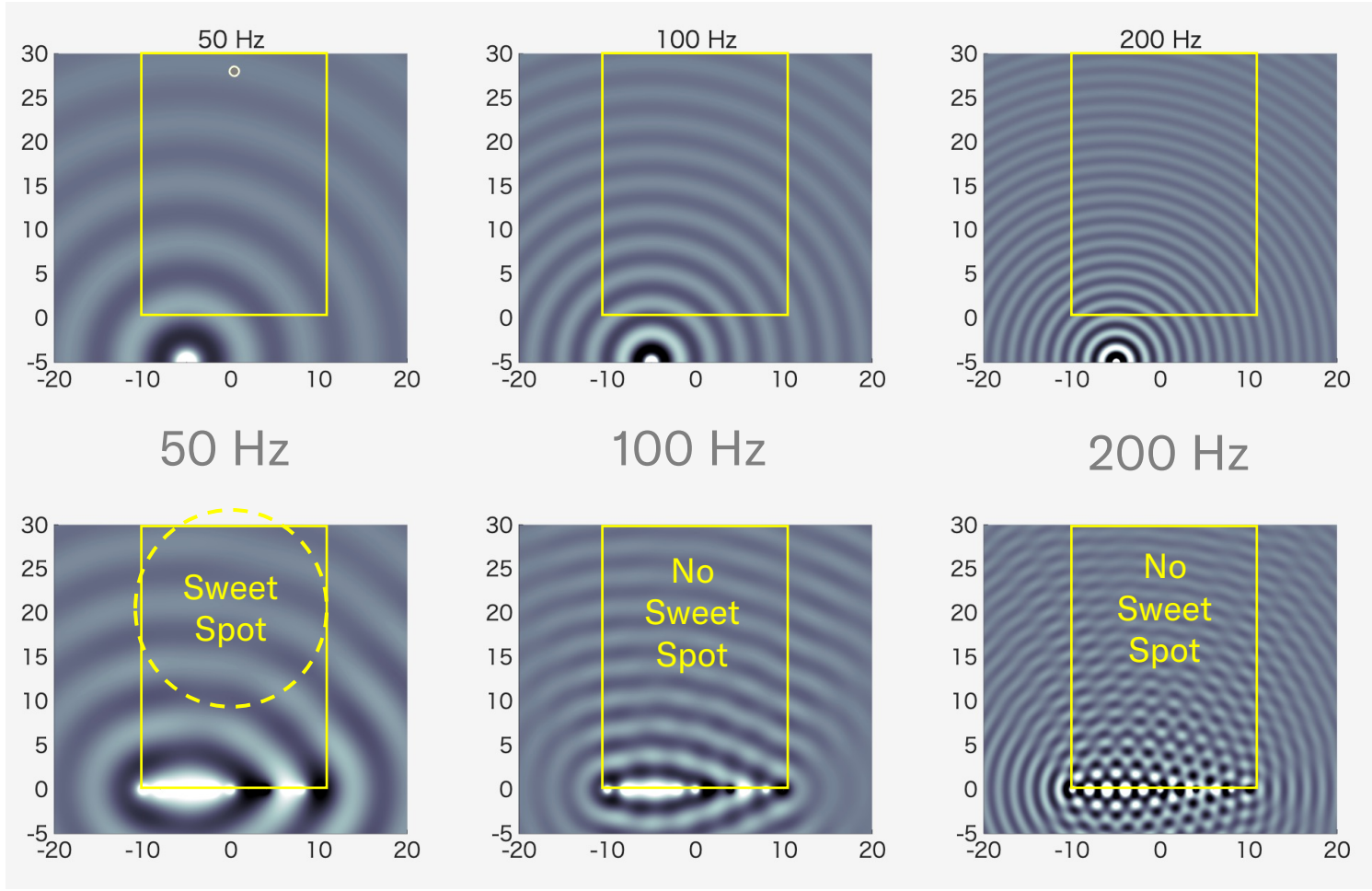
  - In WFS, the wavefront can only be accurately reconstructed for frequencies whose half-wavelength is larger than the speaker spacing Δd:

$$\lambda/2 \geq \Delta d$$

$$f \leq \frac{c}{2\Delta d}$$

Physical Acoustics-Based Method

- **Wave Field Synthesis (WFS) for Linear Array**
  - Case Example: Speaker Spacing $\Delta d = 2.0\ \mathrm{m}\ (f \leq 85\ \mathrm{Hz})$



Physical Acoustics-Based Method

- **Wave Field Synthesis (WFS)**
  - **Wide Sweet Spot**: Offers stable spatial perception across a broad listening area.
  - Accurately **reproduces depth** by simulating wavefront curvature.
  - **Maintains consistent phantom source** size during movement.
  - **Phase Interference**: Source movement causes delay changes, leading to phase artifacts; compensation is required.
  - **Effective only below a certain frequency**, determined by speaker spacing.
  - Performance can **degrade due to reflections and reverberation** in real environments.

Physical Acoustics-Based Method

- **Mode Matching Method (Higher Order Ambisonics: HOA)**
  - The sound pressure $p(\mathbf{r}, k)$ at position $\mathbf{r} = (r, \theta, \phi)$ is expressed as:

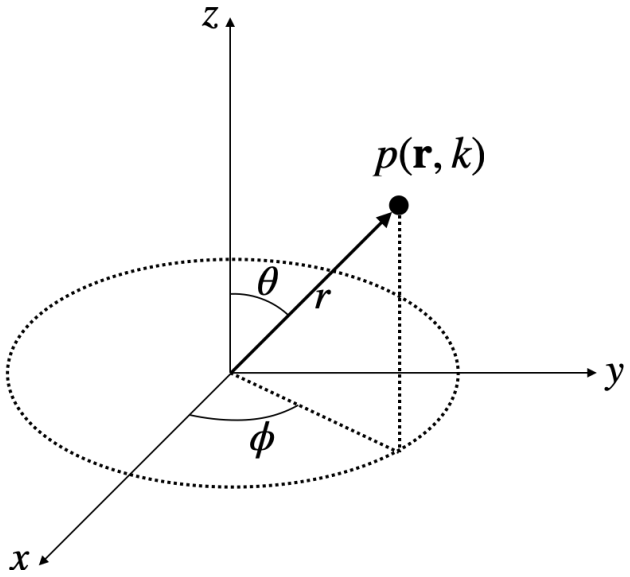
$$p(\mathbf{r}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} A_n^m(k) j_n(kr) Y_n^m(\theta, \phi)$$

Wave number $k = \frac{2\pi}{\lambda}$

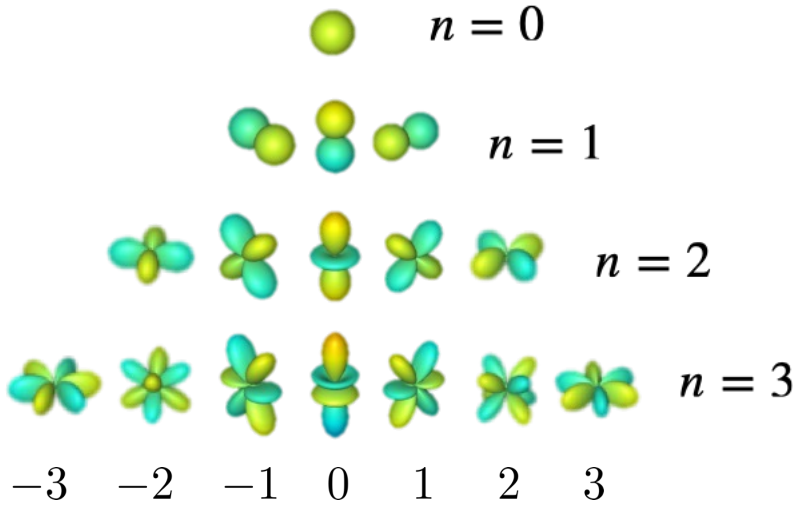
Expansion Coefficient

Spherical Bessel function of the first kind

Spherical harmonics of degree m and order n



Spherical harmonics



Physical Acoustics-Based Method

YAMAHA

- **Mode Matching Method (Higher Order Ambisonics: HOA)**
  - Assuming uniformly distributed points $\mathbf{r}_i = (r_0, \theta_i, \phi_i)$ on a spherical surface of radius r_0, this can be written in matrix form as:

$$\mathbf{p} = \mathbf{YJA}$$

where,

$$\mathbf{p} = \begin{bmatrix} p(\mathbf{r}_1, k) & p(\mathbf{r}_2, k) & \cdots \end{bmatrix}^T$$

$$\mathbf{Y} = \begin{bmatrix} Y_0^0(\theta_1, \phi_1) & Y_1^{-1}(\theta_1, \phi_1) & Y_1^0(\theta_1, \phi_1) & Y_1^1(\theta_1, \phi_1) & \cdots \\ Y_0^0(\theta_2, \phi_2) & Y_1^{-1}(\theta_2, \phi_2) & Y_1^0(\theta_2, \phi_2) & Y_1^1(\theta_2, \phi_2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$\mathbf{J} = \mathrm{diag}\begin{pmatrix} j_0(kr_0) & j_1(kr_0) & j_1(kr_0) & j_1(kr_0) & \cdots \end{pmatrix}$$

$$\mathbf{A} = \begin{bmatrix} A_0^0(k) & A_1^{-1}(k) & A_1^0(k) & A_1^1(k) & \cdots \end{bmatrix}^T$$

the expansion coefficients can be solved as: $\mathbf{A} = \mathbf{J}^{-1}\mathbf{Y}^H\mathbf{p}$ **(Encode)**

- When applying HOA to object-based audio, one practical approach is to assume that a **plane wave arrives from the direction of each object**.

Physical Acoustics-Based Method

YAMAHA

- **Mode Matching Method (Higher Order Ambisonics: HOA)**

$$\mathbf{r} = (r, \theta, \phi)$$

  - Assuming that **point sources are uniformly distributed** on a spherical surface of radius R, the sound field they generate can be expressed as:

$$\hat{p}(\mathbf{r}, k) = ik \sum_{n=0}^{\infty} \sum_{m=-n}^{n} j_n(kr) h_n(kR) Y_n^m(\theta, \phi) \sum_{l=1}^{\infty} w_l(k) Y_n^m(\theta_l, \phi_l)^*$$

Spherical Hankel function of the first kind

Weight factor

  - To reconstruct the original sound field $p(\mathbf{r}, k)$, we solve for the weight factor $w_l(k)$ such that $\hat{\mathbf{p}} = \mathbf{p}$.
  - Then, the weight vector $\mathbf{w}$ can be obtained by: $\mathbf{w} = \Upsilon \mathbf{H}^{-1} \mathbf{A}$ **(Decode)**

  - Standard HOA **assumes a spherical speaker array**, which enables accurate mode matching and spatial reconstruction.
  - **Applying HOA to linear arrays is challenging** due to insufficient angular coverage and poor mode representation.
  - When **the number of speakers is limited** or they are **not uniformly distributed**, the reconstruction **accuracy of the sound field significantly decreases**.

Comparison of VBAP, DBAP, and WFS

Feature	VBAP	DBAP – AFC Image	WFS
Speaker Configuration Flexibility	**Moderate** – requires structured layout (pairs/triangles)	**High** – works with arbitrary layouts	**Low** – requires dense, and approximately uniform
Sweet Spot Size	**Moderate** – limited to central area	**Wide** – no fixed listening point	**Wide** – consistent across large area
Processing Load	**Moderate**	**Low**	**High** – requires real-time delay and filtering
Localization Sharpness	**High** near center, decreases off-center	**Moderate**	**High** – accurate wavefront reconstruction in lower frequencies
Depth Perception	**Poor** – requires added reverb	**Limited** – can be enhanced with reverb	**Good** – simulates wavefront curvature
Smoothness of Sound Movement	**Moderate** – may cause artifacts at transitions	**Smooth** – continuous gain changes	**Smooth**, but require phase compensation
Timbral Stability during Sound Movement	**May vary** due to phantom source size changes	**Stable**	**May be affected by phase shifts** from delay changes, causing artifacts

Demo

Conclusion

- Provided a brief overview of immersive sound technologies.
- Formulated and compared rendering methods for object-based audio, including VBAP, DBAP, WFS, and HOA.
- Demonstrated simple playback examples using a linear speaker array with VBAP, DBAP, and WFS.
- In the AFC Image system, DBAP is employed for its minimal timbral coloration and ease of implementation.

YAMAHA

Make Waves