

A Spatial Sampling Approach to Wave Field Synthesis: PBAP and Huygens Arrays*

Julius O. Smith III

Center for Computer Research in Music and Acoustics (CCRMA)
Department of Music
Stanford, California 94305 USA

<jos at ccrma.stanford.edu>
<https://ccrma.stanford.edu/~jos/>

May 15, 2020

Abstract

A simple approach to microphone- and speaker-arrays is described in which the microphone array is regarded as a sampling grid for the acoustic field, and the corresponding speaker-array is treated as a “spatial digital to analog converter” that reconstructs the acoustic field from its spatial samples. Advantages of this approach include ease of understanding and teaching, ease of deployment, effective practical guidelines for deployment, and significant computational savings in special cases. In particular, in the far-field case (virtual acoustic sources many wavelengths away from a linear array of speakers) it is possible to quantize source angles slightly so that *no* processing per speaker is required beyond pure integer delay. Smoothly moving sources are obtained using well known delay-line interpolation techniques such as linear (cross-fading) and Lagrange (polynomial) interpolation between/among speakers. We call the far-field line-array case Planewave-Based Angle Panning (PBAP), in reference to the well-known Vector-Base Amplitude Panning (VBAP) family of techniques, some of which are derived here as special cases: When speakers undersample the acoustic field, the result may be considered a form of VBAP, and VBAP is also obtained as a limiting case of polygonal PBAP arrays truncated to the polygon perimeter. Spatial samples need not be on a linear array, leading to a simple spatial audio system we call Huygens Arrays (HA). HAs are quite general for sources located behind the speaker array, which no longer needs to be linear, and the sources are no longer restricted to the far field. Multiband and hybrid arrays employing VBAP (or stereo) and subwoofer(s) are discussed, using sampling theory to inform the choices of crossover frequencies. In summary, various sound spatialization techniques are discussed, spanning VBAP, PBAP, Huygens Arrays, and special cases of Wave Field Synthesis (WFS). All may be unified under the general topic of spatial acoustic field sampling, and all were suggested by attempts to derive WFS systems as properly bandlimited spatial interpolators.

*Published 2019-11-18: <http://arxiv.org/abs/1911.07575>

Contents

1	Introduction	3
2	Plane Wave Sampling Theory	8
2.1	Only Planar and Spherical Arrays Can Work Perfectly	10
2.2	Fractional Delay	10
2.3	A Synthesis Scenario	11
2.4	Quantized Angles of Arrival	11
2.5	Four-Quadrant PBAP	12
2.6	Polygonal PBAP (PPBAP)	13
2.7	Circular PBAP Truncates to VBAP	14
2.8	Wrapped Polygonal PBAP (WPPBAP)	15
2.9	Combining Line Arrays to make Polygons	15
2.10	Increasing the Number of Source Angles	15
2.11	Continuous Angles of Arrival	16
2.12	Covering the Spatial Hearing Frequency Range	16
2.13	Delaying High Frequencies to Suppress Aliasing via Precedence Effect	16
2.14	Finite-Array Correction	17
2.15	A Four-Inch Grid Implementation	17
2.16	Extension from a 2D Listening Plane to 3D	18
3	Huygens Arrays (HA)	19
3.1	Linear Huygens Arrays	20
3.2	Sampling Spreading Loss	20
3.3	Virtual Sources in Front of the Array	21
3.4	Undersampled Huygens Arrays become VBAP	21
3.5	More General Huygens Arrays	22
3.6	Interpolation Accuracy	23
3.7	Evolving Radiation Pattern (Sampling Kernel) due to Diffraction	24
3.8	Specific Sample Shapes (Speaker Radiation Patterns)	24
4	Multiresolution Spatial Sampling Arrays	25
4.1	Multiresolution Coaxial Drivers Array	26
4.2	Multiresolution Line Array	26
4.3	Multiresolution Sampling Considerations	27
4.4	Multiresolution Speaker Systems	27
4.5	Four-Way Huygens Arrays	28
4.6	Huygens Octave Panels (HOP)	29
4.7	Five-Band HOPs	30
4.7.1	HOP Stage Angle	31
4.8	Upper, Middle, and Lower HOPs	32
4.9	Nominal Design Guidelines	33
4.10	An Equilateral Triangle Design Example	33
4.10.1	Stereo As Two Spatial Samples	33

5	Conclusions and Future Work	34
6	Acknowledgments	35
A	Relevant Software	35
A.1	Soundfield Synthesis (SFS) in Matlab	35
A.2	PBAP in FAUST, Integer Delays	35
B	Speakers as Spatial Samples	36
B.1	Point-Source Speakers	36
B.2	Overlap-Add of Point-Source Frequency Responses	37

1 Introduction

There is an extensive literature on microphone- and speaker-arrays for audio measurement and reconstruction [Pulkki, 2017; Ahrens, 2012]. Let M denote the number of microphones, and N the number of speakers. When $M = N = 1$, we have monaural recording and playback, while $M = N = 2$ describes stereo, etc. There are many approaches to making larger microphone and/or speaker-arrays (M and/or N greater than 2). Since only a small number of speakers is affordable in typical practice, we are normally very concerned with human *perception* of spatial sound [Blauert, 1997], informing stereophonic, quadraphonic, and more generally *ambisonic* sound systems [Cooper, 1972; Gerzon, 1974, 1985]. Ambisonics extends stereo and quad with an expansion of the soundfield in terms of *spherical harmonic* basis functions centered on one listening point.¹ Such systems must deal with the psychoacoustics of direction and timbre perception in various frequency ranges and for various geometries.

Given a very large number of microphones and speakers, it is possible to approximate complete reconstruction of the soundfield in a given space. The best known approach to this problem is *Wave Field Synthesis* (WFS) [Berkhout et al., 1993],² also called “acoustic holography,” or “holophony” [Berkhout, 1988]. WFS reproduces (or synthesizes) a recorded soundfield *physically*, so that psychoacoustic questions can in principle be avoided.³ However, for best results at minimum expense, psychoacoustic considerations remain important.

The basic idea of an “acoustic curtain” for reconstructing soundfields in principle was described by Harvey Fletcher [1934], and at that time, two or three speakers was considered an adequate psychoacoustic approximation [Steinberg and Snow, 1934]. Generating wave propagation from spherical waves (“secondary sources”) emitted along the wavefront is the essence of Huygens’ Principle (1690).⁴ Both Huygens and Fletcher called for a *continuum* of wavefront samples. The theory of *bandlimited sampling* was introduced by Nyquist [1928], which, together with basic wave theory, can be considered the basis of this paper.

¹A nice overview of ambisonics references appears on the Web: http://www.york.ac.uk/inst/mustech/3d_audio/gerzonrf.htm

²https://en.wikipedia.org/wiki/Wave_field_synthesis

³Ambisonics can also approach physical completeness of the soundfield, as the order (number of spherical harmonics) increases, but most practical systems are fairly low order. (The highest order used at CCRMA is presently seven.) First-order ambisonics is essentially stereo (representable as a monopole plus one left-right dipole) augmented to include front-back and top-down dipoles.

⁴<https://www.britannica.com/science/Huygens-principle>

Deriving WFS begins with the Kirchhoff-Helmholtz integral (or in simplified form from the Rayleigh integral), which expresses any source-free acoustic field as a sum of contributions—called secondary sources—from the boundary of any enclosing surface [Firtha, 2018; Pierce, 1989; Berkhout et al., 1993; Ahrens, 2012]. The same basic approach is used by the well known *Boundary Element Method* (BEM) for numerically computing a wavefield from its values along a boundary surface [Kirkup, 2007]. The secondary sources in WFS aim to reconstruct (in the listening zone) the same soundfield produced by the original (primary) sources “on stage.” In practice, the secondary sources are simplified from an enclosing sphere down to (typically) a ring of loudspeakers in a line array around the listening space (which should not be reverberant and ideally anechoic—a major goal of Berkhout’s WFS formulation was to include the reverberant as well as the direct soundfield). There are many variations on the details of deriving a practical WFS system, and some of them get close to the sampling-based point-of-view taken here. However, there does not appear to be a WFS paper which formulates the problem as basic soundfield sampling and reconstruction from samples (spatial analog-to-digital and digital-to-analog conversion). As a result, differences in final implementation do emerge, as will be brought out below.

Far-Field WFS (FFWFS) is the limiting form of WFS in which the sources are many wavelengths away from the recording mics and listening audience. By adopting this simplifying assumption, which is not restrictive in many applications, we can derive FFWFS very easily and clearly from *sampling theory*, and this is where we begin below.

Overview

This paper defines several speaker-array systems based on a sampling-theory approach:

- Planewave-Based Angle Panning (PBAP) consists of at least one linear speaker array implementing an approximation to Far-Field Wave Field Synthesis (FFWFS).
- Polygonal PBAP (PPBAP) uses multiple PBAP arrays to define an interior acoustic space which is preferably polygonal in the 2D case. Thus, multiple line arrays are combined to form a polygonal listening area, such as a square, rectangle, octagon, etc.
- Truncated Polygonal PBAP is a variation on PPBAP which trims each line array down to the polygon side it includes. In the limit of many polygon sides, a circular array is obtained, giving the simplest form of Vector-Base Amplitude-Panning (VBAP) [Pulkki, 1997, 2001] in which only one speaker (the one closest to the desired angle) emits each source.
- Huygens Array (HA) is a generalized approach allowing sources and speakers/microphones to be placed anywhere on the listener-side of the source(s). In this case, sampling-based insights provide guidance for what to expect, and how to avoid artifacts by maintaining a valid (nonuniform) sampling grid.
- Various straightforward extensions to 3D are discussed.
- Multiband arrays and integrations with VBAP, stereo, and subwoofers are discussed.

When conditions for sample-based reconstruction are *not* met, we transition as gracefully as possible to some form of Vector Base Amplitude Panning (VBAP) [Pulkki, 2001].

A benefit of extending VBAP to PBAP is providing a larger “sweet spot,” since the quasi spherical waves emanating from VBAP speakers become upgraded to approximations of sampled

plane waves in PBAP, and plane waves are the same for all listeners to within a delay. In other words, it is well known that point sources, such as ideal speakers used for ordinary stereo, produce listening levels that are sensitive to listener proximity, while a soundfield composed of a plane wave from each far-field source gives a constant level versus position for each source.

Background

This paper evolved from 2011 to now as a back-burner interest by the author who has never published previously in the field of spatial audio, and who has limited experience in the field, but much experience with bandlimited sampling theory. This paper contains the author's accumulated reactions to drilling down on currently used spatial audio methods such as Wave Field Synthesis (WFS) and Vector-Base Amplitude Panning (VBAP), without the level of background research (*i.e.*, reading all known relevant papers) expected of a typical graduate student in the field. As a result, it is possible and even likely that some or many of the speaker arrays "introduced" here exist already in the literature and/or patent record in some form. The author welcomes citations that can be incorporated into an updated version of this paper.

Another result of the evolving nature of this paper is that it unfolds progressively from basic soundfield sampling to PBAP and its variations and finally to Huygens Arrays (HA). While it could make sense to write a fresh, shorter paper on one or more of these subtopics, the full story has greater tutorial value, and is less work to write, so here it all is in one relatively long paper.

Overview of Planewave-Based Angle Panning (PBAP)

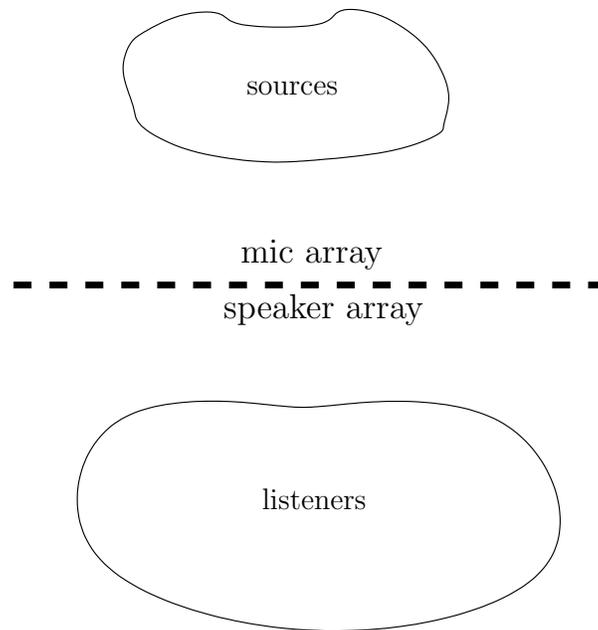


Figure 1: Assumed geometry of primary sources, mic- or speaker-array (or *secondary sources*), and listening space in PBAP. The primary sources and mic-array are separated by many wavelengths so that the microphones only need to record *pressure samples* of a superposition of *plane waves*. The speaker-array is then designed to try to reproduce this superposition of plane waves from its samples along the array.

Studying Wave Field Synthesis (WFS) led to the idea of the considerably simpler Planewave-Based Angle Panning (PBAP), which could just as well be called Far-Field Wave Field Synthesis (FFWFS). The simplifying assumptions of PBAP are

1. the microphone-array is positioned between the primary sources and the listeners, *e.g.*, between a “stage” and all the “seats” in an auditorium, for example, as shown in Fig. 1;
2. the microphone-array is at least several wavelengths away from the nearest source, so that the mic-array sees a superposition of *plane waves* to a good approximation;⁵ and
3. the speaker-array lies along a line parallel to or coincident with the mic array in the direction of the audience.

The simplest case of assumption 3 is when the speakers and microphones are *coincident*, *i.e.*, $N = M$ and the speaker array lies along the same line as the mic array, so that each speaker reproduces the signal recorded at its microphone. The mic array can be viewed as a spatial A/D converter, sampling the incident wavefront, and the coincident speaker array gives the corresponding spatial D/A converter for these microphone samples. When $N \neq M$ or when not coincident and not spaced in a compensating manner, we must compute the speaker signals from spatial-resampling of the mic signals.

The main point is that PBAP is obtained from WFS by moving all acoustic sources to “infinity”—*i.e.*, onto the “celestial sphere”.⁶ In practical acoustics, “infinity” means “many wavelengths,” and thus depends on frequency. Huygens Arrays, developed next, generalize the linear/planar sampling arrays to more geometries while keeping a separation plane between primary sources and listeners.

Figure 1 illustrates the basic geometry assumed for single-line-array PBAP. All sources are confined to a “stage area,” which can be a 3D distribution of sources when the microphone-array is planar, but such a distribution is effectively 2D due to assumption 2 (the microphones are in the “far field” of each source). All listeners are confined to the “audience area” which can also be a 3D distribution of “seats,” when the speaker-array is planar, but again the far-field assumption implies that each seat hears substantially the same planar distribution of sources far away. The microphone- and speaker-arrays thus form a *separation plane* between the stage and audience areas. For the line-array case, the sources are considered as being arranged on a flat stage (same height—only azimuth varying—the same case addressed by ordinary stereo).

In summary, our assumptions allow us to reconstruct a superposition of plane waves given pressure samples along any plane separating the source and audience area. Since the microphone-array is our pressure-sampling array, it is natural to choose *uniform spacing* of the microphones along each coordinate dimension. (It is reasonable to choose a smaller spacing for the horizontal direction since people are more sensitive to azimuth than to elevation of source angle-of-arrival.) Furthermore, we need to ensure that the incident wavefield is properly *band-limited*⁷ [Smith, 2010]. In normal signal sampling, an anti-aliasing lowpass-filter removes high-frequency components that

⁵Since every source-free soundfield can be constructed as a superposition of plane waves, it follows that solving the plane-wave synthesis problem is quite general.

⁶This analogy for PBAP sources as point-sources (stars) on the celestial sphere motivates the alternate name “Star Field Synthesis (SFS).” Viewing a starfield through a window is a vivid analogy for PBAP using a rectangular speaker array. However, diffraction effects are much more significant in SFS (especially at low frequencies) than when viewing light through a rectangular aperture, so the analogy can be misleading.

⁷<https://ccrma.stanford.edu/~jos/resample/>

would alias. In the spatial sampling array, we can simply require that the incoming plane waves have a known maximum angle that does not “alias” spatially. This limits the angular “stage width.”

2 Plane Wave Sampling Theory

This section develops a special case of Wave Field Synthesis (WFS) by spatially *sampling* simple plane waves. Sampling plane waves is much simpler than the traditional WFS formulation which begins with the classical Kirchhoff-Helmholtz integral [Pierce, 1989; Firtha, 2018]. In return for this simplicity, we are restricted to virtual primary sources that are many wavelengths away from the speaker array, and on the other side of it from the listener. As we shall see, we can relax these restrictions in various ways, and the remaining sampling conditions are generally equally binding for WFS systems. In other words, spatial sampling theory is fundamental to all spatial audio systems using discrete drivers arranged in linear, planar, or even more general array geometries. What does not seem to be generally known, however, is that a sampling-based approach is *sufficient* (and much more to the point) for deriving and optimizing the system, as pursued in this paper.

Figure 2 shows a monochromatic plane wave traveling down and to the right at a 45 degree angle. The solid black line across the middle represents the microphone-array, ideally a uniformly spaced grid of tiny omnidirectional pressure microphones having no “acoustic shadow” at all; these microphones serve to *sample* the plane wave along the line. In the 3D case, the line represents one cut along a planar microphone-array. The sinusoid drawn along the microphone line indicates the pressure seen by each microphone. By the sampling theorem (applied now to spatial sampling using a microphone-array), we must have *more than two microphones per wavelength* λ_x along the line array. Thus, the required microphone density is determined by the minimum incident wavelength λ_{\min} and the maximum angle of incidence θ_{\max} , as derived below.

Figure 3 illustrates the geometry of the wavelengths involved. The wavelength of the incident sinusoidal plane wave is denoted λ , and λ_x denotes the wavelength of the sinusoidal pressure fluctuation seen by the microphone line array. As Fig. 3 makes clear, from the angle of incidence θ and incident wavelength λ , we have

$$\sin(\theta) = \frac{\lambda}{\lambda_x}. \quad (1)$$

Let Δ_x denote the microphone spacing along the x axis. Then the sampling theorem requires

$$\Delta_x < \frac{\min\{\lambda_x\}}{2} = \frac{1}{2} \frac{\lambda_{\min}}{\sin(\theta_{\max})} \quad (2)$$

$$= \frac{c}{2 \cdot f_{\max} \cdot \sin(\theta_{\max})} \quad (3)$$

where c is the speed of sound (m/s), f_{\max} is the maximum temporal frequency in Hz (typically 20 kHz for audio), and θ_{\max} (radians) is the maximum plane-wave angle allowed.

For example, choosing $f_{\max} = 19.5$ kHz and $\theta_{\max} = \pi/4$ (stage angle 90 degrees), and using $c = 343$ m/s for sound speed, we obtain $\Delta_x < 12.5$ mm, or about half-inch spacing for the microphones. (The coincident speaker-array has the same sampling-density requirement as the microphone-array.)

Reducing either f_{\max} or θ_{\max} relaxes the spatial sampling density requirement. For example, if the “stage width” is reduced from 90 degrees ($\theta_{\max} = \pi/4 \approx 0.8$) down to 40 degrees ($\theta_{\max} =$

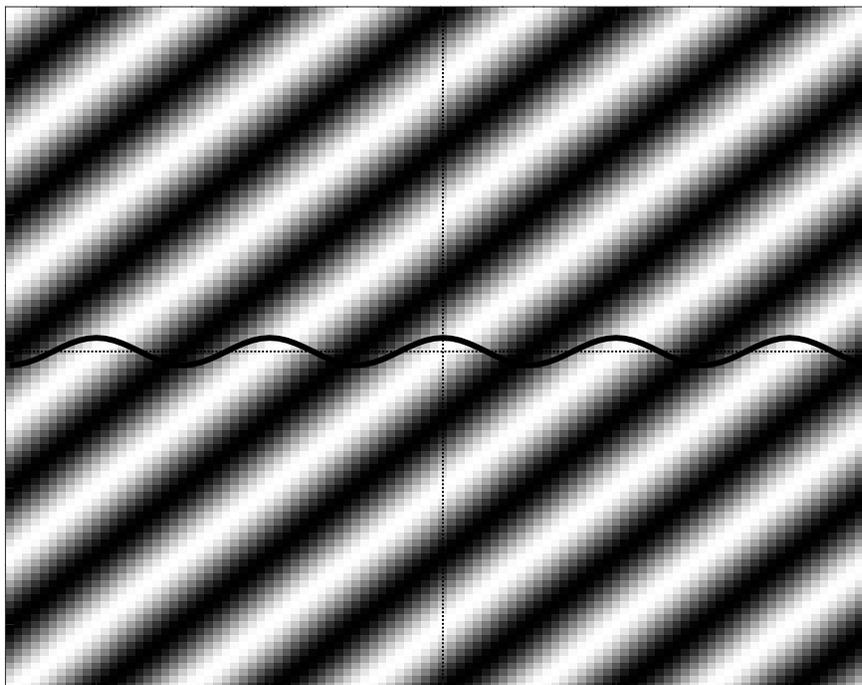


Figure 2: Cross-section of a single-frequency plane wave traveling down and to the right into a line-array of microphones.

$\pi \cdot 20/180 \approx 0.35$), then one-inch spacing of the microphones (and speakers) is allowed. If we band-limit our reconstruction bandwidth to 5 kHz, then we get by with four-inch spacing, as pursued below in a practical PBAP design (§2.15).

If we don't band-limit to below the spatial Nyquist limit, then we obtain “spatial angle aliasing” at very high frequencies for sources near the left or right edge of the “stage viewing window”. That is, for sources near the left or right edges of the “stage”, the highest-frequency components may not appear to come from the same direction as components below the cutoff frequency of 5 kHz. On the other hand, perception is such that the apparent angle-of-arrival typically may not alias at high frequencies because the desired angle remains a psychoacoustic choice and keeps the whole source spectrum in one place. Sources near the center of the stage are spatially oversampled by the microphone-array at all frequencies, so they are never a problem. In lowpassed-wideband-noise tests (see Appendix A), high-frequency spatial aliasing has been observed to break up a formerly coherent wideband virtual noise source, but not in a manner indicating “folding” as in aliasing of tones due to temporal sampling, but instead as the sound of a spurious new noise source somewhere along the array. The psychoacoustics of spatial aliasing perception is a fascinating topic for ongoing research.

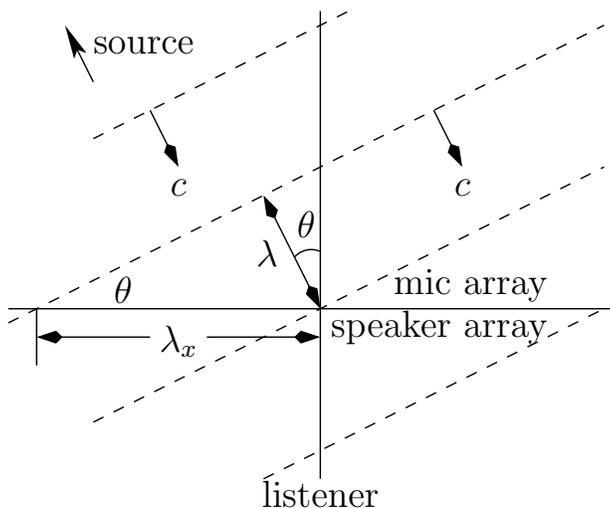


Figure 3: Wavelength geometry.

2.1 Only Planar and Spherical Arrays Can Work Perfectly

It is well known that only spherical waves propagate in 3D without a wake, which includes plane waves as a limiting case. A line array produces a cylindrical wave, which has a wake [Morse and Ingard, 1968, p. 243]. This takes the form of temporal dispersion behind the wavefront in the soundfield reconstruction from linear arrays. Interestingly, nobody seems to complain about it, so it can't be a particularly audible distortion. However, for best results, planar (or spherical) sample distributions are preferred.

Another disadvantage of cylindrical waves relative to plane waves is amplitude decay by $1/\sqrt{r}$ (consider energy conservation), where r is distance from the cylinder axis. This is less of a problem, since it preserves audio fidelity and spatial locations, and is potentially even desirable since it gives the listener a way to alter listening level by merely moving closer to or farther from the array (-3dB per distance doubling).

We will continue to consider primarily linear arrays, in which soundfield reconstruction from samples happens only along the left-right dimension, and VBAP or stereo is used for the vertical dimension, if anything. There will be other approximation errors as well, and the key question as always is the *audibility* of those errors.

2.2 Fractional Delay

A nice property of ideal plane waves is that they do not decay, *i.e.*, the amplitude of a given source is the same at every spatial sampling point (microphone or speaker). The only change from one spatial sample to the next is a *relative delay* corresponding to the different propagation distance. Thus, the only processing needed for each source for each speaker is the *fractional delay* [Franck, 2008, 2011; Välimäki, 1995] corresponding to the source-speaker distance. Furthermore, in a line or plane array, the change in delay from one spatial-sample to the next is *constant* along any straight line, and this can be used to save overall computation [Franck, 2011].

2.3 A Synthesis Scenario

Let N_s denote the number of virtual acoustic sources and let N be the number of speakers. Then in PBAP, the j th speaker signal is given by a sum of delayed source signals:

$$s_j(n) = \sum_{i=1}^{N_s} x_i(n - D_{ij}), \quad j = 1, 2, \dots, N \quad (4)$$

where $x_i(n)$ is the i th source signal, and D_{ij} is the time-delay, in samples, from virtual source i to speaker j . The signal $x_i(n - D_{ij})$ is obtained as the output of a fractional-delay filter driven by $x_i(n)$. The same delay line can be used for source i to all N speakers, using “taps” to interpolate [Smith et al., 2002].

2.4 Quantized Angles of Arrival

Since high-quality fractional-delay filtering is expensive, it is worth considering restriction to angles-of-arrival corresponding to *integer* delays (in samples). If the speaker-to-speaker spacing along a line array is Δ_x , then the speaker-to-speaker delay for a plane wave at angle-of-incidence θ is $\Delta_x \cdot \sin(\theta)/c$, where c denotes sound speed. Thus, an angle-of-arrival θ_n corresponds to an integer speaker-to-speaker delay n (in samples) when

$$\frac{\Delta_x}{cT} \sin(\theta_n) = n, \quad n = 0, \pm 1, \pm 2, \dots, \pm N_a \quad (5)$$

where T denotes the temporal sampling interval in seconds, and

$$N_a = \left\lfloor \frac{\Delta_x}{cT} \right\rfloor \triangleq \text{floor} \left(\frac{\Delta_x}{cT} \right).$$

Note that increasing the speaker spacing Δ_x for a given temporal sampling rate $f_s = 1/T$ gives more integer-delay angles θ_n . However, doing this also decreases the stage-width (or supported bandwidth) by the same factor.

It is clearly inaudible to shift the location of each virtual source x_i so that the time delay to the nearest speaker is an integer number of samples. Then having an integer number of samples for each inter-speaker delay makes all the delays integer. Finally, this can all be implemented as a single delay line with a *tap* (non-interpolating) for each speaker signal. For moving sources, to avoid clicks, moving taps should be cross-faded from one integer delay to the next in the usual way [Smith, 2010].⁸

Solving Eq. (5), the collection of angles θ_n corresponding to integer inter-speaker delays n (in samples) is

$$\theta_n = \sin^{-1} \left(\frac{cT}{\Delta_x} n \right), \quad n = 0, \pm 1, \pm 2, \dots \quad (6)$$

For example, with $\Delta_x = 0.1$ m ($\approx 4''$), $c = 343$ m/s, and $f_s = 48$ kHz, the first 11 available angles are

$$\theta_n^\circ \in \pm[0, 4.1, 8.1, 12, 16, 21, 25, 30, 35, 40, 45] \quad (7)$$

⁸https://ccrma.stanford.edu/~jos/pasp/Delay_Line_Signal_Interpolation.html

degrees, to two digits precision. Azimuth perception is accurate to approximately 1 degree at center-front.⁹

Figure 4 depicts the available geometric rays of plane-wave propagation for this example. Thicker rays are drawn for 0 degrees (directly in front) and ± 90 degrees (full left and right).

29 angles, fs=48 kHz, Speaker Spacing = 4.0 in (Meyer MM-4XP)

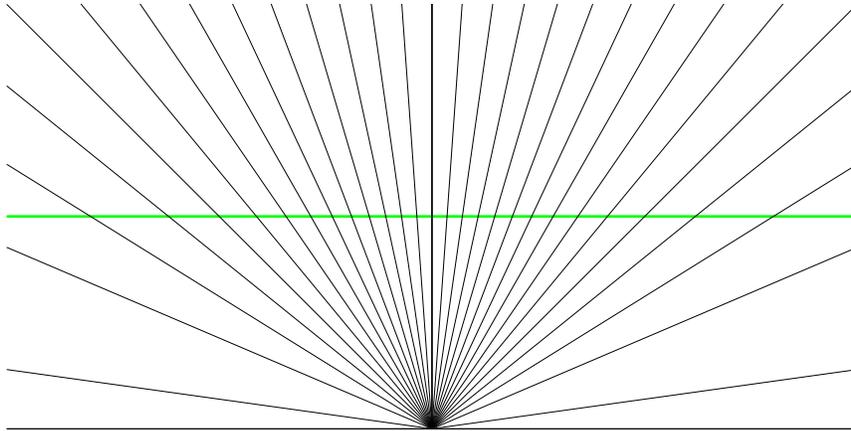


Figure 4: Rays of propagation toward a listener in the center for the available plane-wave angles from a line-array (four-inch spacing) at a sampling rate of 48 kHz. These are the source angles requiring no interpolation—just pure integer delay from one speaker to the next.

If the 21 angles-of-arrival across a 90° stage listed in Eq. (7) are deemed sufficient, then PBAP is essentially free: just provide the appropriate integer adjacent-speaker delays for each source in the sum for each speaker. As is well known, an integer delay is an $\mathcal{O}(1)$ computation, requiring only a single read, write, and circular-buffer pointer-increment each sampling instant [Smith, 2010].¹⁰

2.5 Four-Quadrant PBAP

If *four* line arrays are arranged in a “tic tac toe board” configuration (or perhaps just a square) enclosing the listener in its central square, as shown in Fig. 5, then each line array need only cover a ± 45 degree range, which is more uniformly sampled in angle.

Four-Quadrant PBAP uses only a *fourth* of the speakers for each plane wave. Infinitely long line arrays emit cylindrical waves, which are equivalent (ignoring the wake) to plane waves in one

⁹<http://acousticlab.org/psychoacoustics/PMFiles/Module07a.htm>

¹⁰https://ccrma.stanford.edu/~jos/pasp/Software_Delay_Line.html

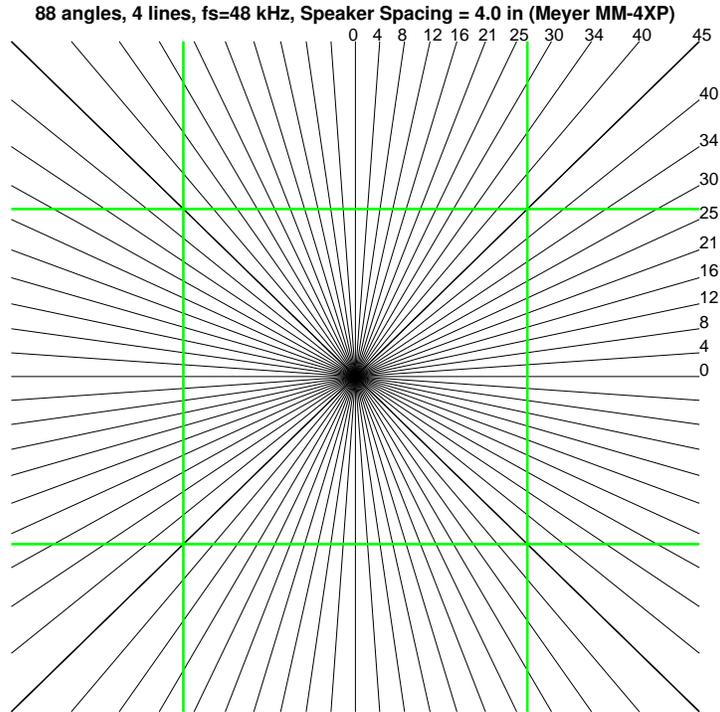


Figure 5: Ray coverage for the same example considered in Fig. 4 using *four* line arrays enclosing the listener at the center of a “tic tac toe board” configuration. The rays at ± 45 are drawn thicker.

listening plane passing through the cylindrical axis. However, more practical *truncated* line arrays can benefit from using more than a fourth of the speakers. It is intuitively obvious that at least *half* of the speakers could help construct a desired plane wave at the listener—all speakers having a radial component in the desired direction.¹¹ At a single listening point, as in ambisonics, *all* of the speakers can be put to work toward approximating the desired soundfield pressure and velocity versus frequency at that point.

2.6 Polygonal PBAP (PPBAP)

The Four-Quadrant PBAP of the previous section readily extends to N -sided polygons. Figure 6 shows a progression from $N = 4$ (left), to $N = 8$ (middle), and $N = 16$ (right). Along the top of the figure is the result of using line arrays (as long as available, yielding the best plane-wave slices in the listening plane). Along the bottom is the compromise obtained by *truncating* each line array to the portion that borders the interior polygonal listening area (Truncated PPBAP).

¹¹The analogous condition for Wave Field Synthesis is to use only secondary sources that are “illuminated” by the virtual source being rendered [Ahrens, 2012].

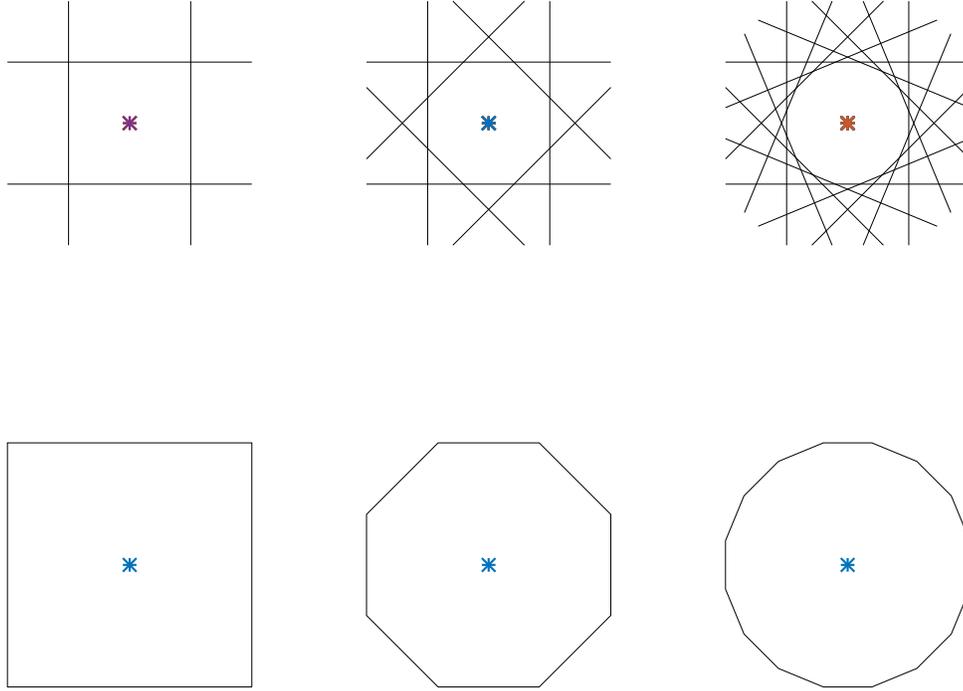


Figure 6: (Top) PPBAP using N line arrays to create an N -sided polygon for the interior listening area. (Bottom) Truncated PPBAP.

2.7 Circular PBAP Truncates to VBAP

In the limit as the number of polygon sides N becomes large, we obtain a *circular* array, having only one speaker representing each line array in the truncated PPBAP case. Furthermore, the available angles (when avoiding interpolation) are simply the speaker angles. This coincides with zeroth-order VBAP (Vector Based Amplitude Panning) [Pulkki, 2001, 1997].

In practical VBAP, sources are typically enlarged to more than a single speaker, both to make a sonically larger source, and so that they pan more smoothly from one location to the next around the ring (or dome, etc.). This of course gets us into interpolation, and can be viewed as such. The multispeaker interpolation strategies of VBAP can be applied to PBAP, both to the interior polygon/circle as in VBAP, and more generally to the line arrays creating an N -sided polygon. In the latter case, PBAP can be viewed as a sweet-spot enlarging strategy for VBAP. The longer the tangential line arrays, the straighter the plane waves emitted, and the less sensitivity to listener position downstream from the plane waves.

2.8 Wrapped Polygonal PBAP (WPPBAP)

A problem with VBAP’s sweet-spot size is that each speaker is approximately a spherical wave source. Therefore, Truncated Polygonal PBAP produces quasi spherical waves from each polygon side (when the wavelength is large compared with polygon side length). To address this, we can use nearly half of all speakers surviving the truncation to participate in the plane-wave generation from each original line array.

In particular, starting with Polygonal PBAP, the process of truncating to the interior polygon can include summing the signal “seen” from each truncated speaker arriving at each surviving speaker. This is essentially just applying the sampling principles used to derive PBAP in the first place. Thus, each speaker signal will include the signal from its original line array, plus a delayed and attenuated signal from each truncated speaker that is “behind it” relative to the listener, and reasonably close. The truncated speakers are now treated as point sources, so the attenuation is proportional to inverse-distance as usual for spherical waves. A maximum distance is set beyond which speakers from the (imagined) extended line array are not heard. This leaves less than half of the surviving speakers receiving contributions from the truncated speakers on any given line array.

2.9 Combining Line Arrays to make Polygons

When avoiding delay-line interpolation and accepting the angles given by integer interspeaker delays, we should choose the sampling rate f_s and speaker line-array spacing X so that the angles available from each line array include the angles to the polygon vertices.

For an N -sided polygon, the two needed angles are $\pm\pi/N$. The set of all vertex angles is $\arcsin(ncT/X)$, $n = 0, \pm 1, \pm 2, \dots$, where c is sound speed and $T = 1/f_s$ is the sampling interval. Thus, we need some integer n to give $\sin(\pi/N) \approx ncT/X$, or

$$f_s(n) = n \frac{\csc\left(\frac{\pi}{N}\right)}{X/c}$$

for some $n = 1, 2, \dots$.

In the four-quadrant case (four line arrays defining a square) with speaker spacing $X = 4$ in (example from §2.15 below), the sampling rate wants to be a multiple of 4802.2, and 48 kHz happens to be 9.9954×4802.2 , so the 10th angle is very close to 45 degrees. For $N = 8$ polygon sides and four-inch speaker spacing, we need f_s to be a multiple of 8873.3, and it so happens that 44.1 kHz is close to five times that (4.97).

2.10 Increasing the Number of Source Angles

The number of available ray angles in PBAP can be increased by various means.

Oversampling in time to $f_s = 96$ kHz doubles the number of integer-delay angles over the same range. The example of Eq. (7) on page 11 goes from 21 to 41 angles-of-arrival across a 90° frontal stage.

Another avenue for doubling the number of source angles is to implement a half-sample fractional-delay filter. In that case, half of the available angles require use of the filter while the other half require no interpolation filter.

Generalizing, given $L - 1$ fractional-delay filters, with the l th filter providing fractional delay l/L samples, $l = 1, 2, \dots, L - 1$, the number of available arrival angles is increased by the factor L .

A benefit of a fixed grid of available angles is that each fractional-delay filter can be individually designed and optimized for the delay it provides. Three such filters ($L = 4$) suffice to provide approximately complete perceptual resolution in the example of Eq. (7).

2.11 Continuous Angles of Arrival

Quantized angles-of-arrival do not suffice when sources move over time. Also, continuous angles-of-arrival allow a less quantized “source width” parameter for each source—as if the source were coming from a distant solid-angle region (like a nebula) instead of one point (star).

There are various methods for continuously variable fractional-delay filtering [Smith, 2010]. Perhaps the simplest is by means of *Lagrange interpolation* [Välimäki, 1995; Franck, 2008, 2011; Smith, 2010].

The case of first-order Lagrange interpolation is especially simple, being simple linear interpolation. Thus, one can linearly “cross-fade” in amplitude from one source angle to the next to implement a moving source. Distributed sources can be formed as a linear combination of adjacent source angles.

2.12 Covering the Spatial Hearing Frequency Range

Spatial hearing [Blauert, 1997] is accomplished by two ears sampling the acoustic field through small apertures (ear canals) having diameters smaller than the wavelength across almost the entire audio band. As a result, the directionality of a sound is inferred primarily from the relative intensity (ΔI) and time-of-arrival (ΔT) at the two ears. There is also directionality information impressed on the signal by pinnae filtering and shoulder reflections, etc., that are especially important for elevation perception.¹² For azimuth perception, ΔT is the dominant cue below about 800 Hz, and ΔI dominates above 1600 Hz or so.¹³ In the octave between these limits, both ΔT and ΔI are used. Also, ΔT is picked up as *phase delay* for low frequencies, and *group delay* at high frequencies (ibid.). Perceptual accuracy is on the order of 1° for azimuth in front of the listener. The lower limit of azimuth perception based on ΔT is approximately 80 Hz, below which phase differences become imperceptible. Thus, our spec is to synthesize correct ΔT s down to 80 Hz. Note, however, that since the acoustic wavelength at 80 Hz is over 4 meters long, we could get by with reduced spatial resolution in this frequency range, such as simple stereo. Multiresolution speaker arrays are discussed starting in §4 below.

2.13 Delaying High Frequencies to Suppress Aliasing via Precedence Effect

As mentioned in §2, spatial aliasing limits the highest frequency and/or the widest source angle supported by a uniformly sampled line array. As also mentioned there, spatial aliasing may not be perceived because the spectrum as a whole may lock in perceptually at the correct angle. In other words, the ambiguity of the spatial angle of the highest-frequency components may be resolved perceptually by the brain’s natural desire to “make sense” of an auditory scene. This effect can be enhanced by slightly *delaying* the high-frequency components relative to the low-frequency

¹²http://en.wikipedia.org/wiki/Head-related_transfer_function

¹³http://en.wikipedia.org/wiki/Sound_localization#ITD_and_ILD.

In [Gerzon, 1974], citing Rayleigh from 1907, the low-frequency crossover is given as 700 Hz. It is also noted in [Gerzon, 1974] that pinnae filtering is thought to be important above 5 kHz.

components that have no aliased interpretation. The idea is to force perception to hear the desired angle before the ambiguous spectral components are heard, so that they will all fuse at correct angle of arrival. This is of course altering the timbre of the sound, and may be considered off limits for that reason.

2.14 Finite-Array Correction

In practice, it is necessary to truncate an array to finite bounds. This causes reconstruction error analogous to the error obtained when restoring a continuous-time signal from a finite segment of its samples. Thus, most of the error occurs for sources near the edges of the array (*i.e.*, arriving from the maximum angles-of-arrival supported). This error can be reduced by compensating for the missing contributions from the truncated “sampling kernels”. From this point of view, the error is equivalent mathematically to the “Gibbs phenomenon,” and many forms of “windowing” and “apodization” have been advanced to address the issue [Smith, 2011].¹⁴ One can also formulate a customized optimization that maximizes perceptual criteria; for this problem, linear programming formulations may suffice for correcting amplitude error.¹⁵ A Hann window is used for array windowing in the Sound Field Synthesis Toolbox [Wierstorf and Spors, 2012].¹⁶

2.15 A Four-Inch Grid Implementation

The Meyer Sound MM-4XP Miniature Loudspeaker (Fig. 7) is a self-powered speaker that provides an approximate four-inch by four-inch cell. Thus, with this speaker we can make either a line array or a 2D grid with four-inch spacing. While this speaker is relatively expensive, it exhibits excellent power and linearity for its size.



Figure 7: The Meyer Sound MM-4XP Miniature Loudspeaker

¹⁴https://ccrma.stanford.edu/~jos/sasp/Spectrum_Analysis_Windows.html

¹⁵https://ccrma.stanford.edu/~jos/sasp/Window_Design_Linear_Programming.html

¹⁶<https://github.com/sfstoolbox/sfs>



Figure 8: Eight-channel array of Meyer Sound MM-4XP miniature loudspeakers

The MM-4XP power output is 113 dB at 120 Hz. Extending the low end beyond the spatial-hearing range down one octave requires a 6 to 12 dB sacrifice in power for the same peak diaphragm excursion.¹⁷ Thus, the extended-low-end array is plenty loud—and we may cross over to a subwoofer below the spatial hearing range.

2.16 Extension from a 2D Listening Plane to 3D

Since PBAP converges to VBAP when using many line arrays truncated to the enclosed polygon, a simple VBAP-style extension to 3D is to place a new speaker directly overhead and a second new speaker directly below. Then, elevation cues can be imparted by mixing in the above or below speaker according to a psychoacoustically measured *panning law* between the array and an out-of-plane speaker above or below.

The proper extension of PBAP to 3D is of course obtained using *sampled plane waves* arriving at the correct 3D angles, instead of cylindrical waves sampled in the listening plane for 2D PBAP from a line array. That means our line arrays must be replaced by planar speaker arrays, and the polygonal listening space becomes a polyhedron, or sphere in the limit. This of course also reduces to non-interpolating spherical VBAP when the planar arrays are truncated down to one point in each plane, and normal VBAP interpolation can be used as described above here as well. As in the 2D case (§2.7 on page 14), 3D PBAP can be used as a sweet-spot enlarger for 3D VBAP.

¹⁷To see this, consider that the speaker is much smaller than the wavelength produced at its low end, so it can be regarded as an approximate “point source” in that frequency range. From the theory of a point source [Morse and Ingard, 1968, p. 310], the peak pressure-amplitude from a sinusoidally oscillating point source is proportional to the peak *volume acceleration* from the source, which is in turn proportional to the *radial acceleration* (second time-derivative of spherical radius) for any small sphere used to model the simple source. (Any sphere much smaller than a wavelength in diameter will do.) Let the spherical radius acceleration be denoted $a(t) = A \cos(\omega t)$, which is proportional to the far-field pressure a fixed distance away. Then the peak radial excursion of the spherical surface is given by A/ω^2 , and keeping the excursion fixed while decreasing ω by one octave reduces the far-field pressure by a factor of four, or -12 dB. When operating as a cell of a planar array, on the other hand, far-field pressure is proportional to the driver surface *velocity* instead of acceleration. In that case, only 6 dB per octave is lost integrating velocity to get displacement.

3 Huygens Arrays (HA)

We have so far considered only PBAP and its variants, which can be considered Far-Field Wave Field Synthesis (FFWFS), in which each source contributes a plane wave to each listening point. This is pretty general, in principle, because, as is well known, every source-free soundfield can be expressed as a sum of plane waves at various amplitudes, phases, and directions of arrival. In fact, Fourier transform methods can be used for this purpose.¹⁸ Therefore, a straightforward path from FFWFS to full-fledged WFS is to decompose any desired soundfield into a sum of plane waves, and then generate those plane waves using FFWFS/PBAP. There are many known methods for so-called Plane-wave Decomposition (PWD) [Pulkki, 2017].

A more direct extension of PBAP toward WFS is based again on simple *sampling* of the acoustic source wave, but now allowing *spherical waves* instead of only plane waves in the soundfield reconstruction. We could call this *Sphere-Based Range and Angle Panning* (SBRAP). However, reconstructing a wavefront as a superposition of spherical waves is essentially the idea of *Huygens' Principle*. We therefore choose the name Huygens Array (HA) for the extension of PBAP to include spherical as well as planar wavefronts.

For constructing a Huygens Array, each virtual source \underline{s}_i is at a known location in 3D space:

$$\underline{s}_i^T = (x_i, y_i, z_i), \quad i = 1, 2, \dots, N$$

According to the basic sampling assumption in PBAP, each speaker location is also a microphone location, so we can denote the i th speaker/mic location by \underline{m}_i , $i = 1, 2, \dots, M$. Different mic distributions are obtainable via spatial resampling as before.

In PBAP, each virtual source \underline{s}_i was characterized by an angle of arrival θ_i , which determined the inter-speaker delay

$$\tau_i = \frac{X}{c} \sin(\theta_i)$$

in seconds along the line array, where X denotes the center-to-center speaker spacing. To generalize from plane waves to spherical waves, we need both a delay d_{ij} and a gain g_{ij} describing the acoustic ray from source i to speaker j .

Let A_i denote the amplitude of source i at a distance one meter from its center. (Each source is assumed to be a point source for now; distributed sources can be modeled as weighted sums of point sources.) Then for the delays we have

$$d_{ij} = \frac{\|\underline{s}_i - \underline{m}_j\|}{cT}$$

where $\|\underline{x}\|$ denotes the Euclidean norm of \underline{x} , c denotes sound speed, and T is the digital audio sampling interval, as before. For the gains we have

$$g_{ij} = \frac{A_i}{d_{ij}},$$

and, if desired, lowpass-filtering due to air absorption can be included:

$$G_{ij}(z) = g_{ij}L_{d_{ij}}(z)$$

where $L_d(z)$ is a standard air absorption filter corresponding to propagating d meters through air at some assumed standard conditions (humidity level being the most important) [Smith, 2010].

¹⁸https://ccrma.stanford.edu/~jos/pasp/Vector_Wavenumber.html

3.1 Linear Huygens Arrays

We have so far not used any assumptions regarding the microphone/speaker array to be used.

The sampling analysis of §2 on page 8 made use of the far-field assumption in obtaining a spatial aliasing limit that depended only on the source angle θ and spatial frequency k . Generalizing to the *near field* (arbitrary source *distances*) means that the sampling analysis is applicable only *locally* along the array. That is, the wavelength seen by the line array depends on both the source angle and the *distance* of the source to the array (or equivalently, the relative distance of the source to the array and to the listening point). For example, the line from the source normal to a horizontal array (see Fig. 3 on page 10) is at angle $\theta = 0$, which is always oversampled by the array. Points on the array far away from the normal line, however, see an angle approaching 90 degrees to the right and -90 degrees to the left. If a source touches the array at $x = y = 0$, then all of array points other than the point at $x = 0$ see a right angle (± 90 degrees). This behavior means we cannot set a limited stage-angle to avoid spatial aliasing like we did in the far-field case (§2). We can now accept a 180-degree stage, or limit the closeness and layout-width of the sources to obtain a worst-case angle limit (maximally close to the array at the edge of the allowed stage), and treat that as before.

3.2 Sampling Spreading Loss

In addition to spatial audio oscillations from a point source, there is amplitude change due to “spreading loss” away from the source. To look at this, consider that a unit pressure point-source at the origin can be expressed as the real part of [Morse and Ingard, 1968]

$$p(r, t) = \frac{e^{j(\omega t - kr)}}{r}$$

where $r = \sqrt{x^2 + y^2 + z^2}$ is the radial coordinate axis. There is no change in the complex amplitude along directions with constant radius r . Along r we observe the maximum amplitude change-rate. This rate of change is approached asymptotically along the array in both directions. The pressure gradient is given by

$$\nabla p(r, t) = \frac{\partial}{\partial r} p(r, t) = -\left(\frac{1}{r} + jk\right) p(r, t)$$

Intuitively, a sampling grid that is adequate for sampling spatial frequencies k should be adequate for sampling spherical spreading (decay by $1/r$) when

$$\frac{1}{r} \ll k_{\max} = \frac{2\pi}{\lambda_{\min}}$$

or

$$r \gg \lambda_{\min}. \tag{8}$$

That is, to keep the rate of amplitude-change due to spreading loss much less than that due to acoustic vibration, we can keep all sources a few minimum-wavelengths or more away from the line array. Note that this strategy only provides approximately valid sampling of the $1/r$ spreading-loss decay, because $1/r$ is not a bandlimited function and therefore cannot be sampled without some error in the reconstruction.¹⁹ Fortunately, the error can be made zero *psychoacoustically* at

¹⁹The 2D Fourier transform of $1/r$ can be shown to be $1/k_r = \lambda_r/(2\pi)$:
http://sepwww.stanford.edu/public/docs/sep103/jon3/paper_html/node3.html

a reasonable sampling density. Amplitude error perception generally requires at least a quarter-dB difference, and that’s in the most demanding case of comparing alternating amplitude levels.

Since the speaker spacing X must be smaller than half the minimum wavelength λ_{\min} , we can stipulate that all sources should stay at least several speaker-spacings away from the array.

An alternative strategy to Eq. (8) is to double the linear sampling density of the array and allow amplitude-change due to spreading loss become comparable to that due to vibration. In this case, the minimum approach distance becomes ($r \geq 1/k_{\max} = X/\pi$), allowing sources get to within $\approx \lambda_{\min}/6.28$ of the minimum wavelength from the array, or about a third of the center-to-center speaker spacing. Intuitively, thinking of the speaker array as a sampling grid, it makes sense to keep virtual point-sources on the order of a sample away or more.

We learn in a first course on digital signal processing that a signal must be bandlimited to less than half the sampling rate in order to avoid aliasing [Smith, 2007b]. Setting a minimum on how close a virtual source may approach the sampling array effectively *spatially bandlimits* the wavefront geometry. This enables soundfield sampling to work as intended. Analogous bandlimiting happens along the time dimension when we apply an A/D lowpass filter prior to sampling in time.

3.3 Virtual Sources in Front of the Array

In all cases considered so far, the virtual sources (primary sources) have been restricted to be *behind* the speaker array by some minimum distance for valid sampling. We can now extend as in WFS to allow sources *between* the array and the listener, but we still must maintain the same minimum distance, but now from the other side of the array. There are differences, however, to keep in mind relative to the behind-the-array case. For simplicity, consider a line or plane array as a starting point.

- Unlike the primary sources behind the array, those in front of the array have to first propagate the “wrong way” to the array to provide recorded signal components (spatial samples) that can be played back from the array to create a converging wavefront back to the virtual source and then on to the listener. This means there is an added delay between the source and the listener, as if the listener can only hear the first-order *reflection* of the source bouncing off the array, with no direct signal from the virtual source. The listener hears the reflection from the array after it passes through the point of convergence at the virtual source and then propagates to the listener.
- The listener receives a *mirrored* reflection of the *rear* of the virtual source. This is no problem for an isotropic source, like any monopole, but it can be an error for distributed sources trying to achieve a specific natural radiation pattern. As a result, primary sources should face the array instead of the listener, and be flipped as needed.
- In offline applications the extra delay for interior sources is easily removed in post-processing.

3.4 Undersampled Huygens Arrays become VBAP

A practical issue that arises when an array is allowed to undersample the soundfield is *level normalization*. An extreme example is the amplitude (∞) of a point source when it touches a microphone/speaker point on the array, which is well out of bounds for any reasonable practical system.

Level regularization is another reason to keep sources a few wavelengths or more away from the microphone/speaker array.

When the array becomes undersampled, say because a wideband source is approaching an extreme stage angle, the array can be regarded as transitioning from soundfield reconstruction by summing interpolation kernels to simple *panning* between/among available speakers (*i.e.*, what we’re normally always doing). Note that the high-frequencies must go into “panning mode” first, while lower frequencies may remain adequately sampled. We can implement an adaptive spectral partition between sample-based reconstruction below and panning above.²⁰

3.5 More General Huygens Arrays

We can also drop the restriction to microphone/speaker line/plane arrays and allow the speakers to be more generally laid out, such as on a sphere (a typically available layout for ambisonics systems). The basic Huygens principle remains the same: when a wavefront reaches a speaker, it fires out a spherical secondary wave (or hemispherical wave aimed toward the sphere’s interior). This configuration was simulated with a plane wave excitation, and a decent plane-wave reconstruction was obtained inside the spherical space. For a spherical speaker array, only half of the speakers become activated by a source outside the sphere.

More generally, when virtual sources are kept away from uniformly laid out speakers, Huygens’ Principle tends to hold up pretty well, according to simulation results to date.

It remains preferable to have a reasonable sampling grid for most frequencies, and we continue to prefer a separation plane between the sources and listeners (Fig. 1 on page 6).

One argument in favor of a separation plane has to do with the fundamental limitation of Huygens’ Principle, which only considers pressure (a scalar) and not velocity (a 3D vector). Without velocity matching, a line array of point sources creates a cylindrically symmetric output. A plane array of point sources similarly emits identical wavefronts in both directions away from the plane. If all listeners are on one side of the speaker array, then we don’t care what happens on the other side. However, this argument can be overcome.

Since audio loudspeakers are normally baffled, we get an approximate hemispherical source from them, which is ideal in the limit of a continuous distribution of point sources. Also, matching pressure across both time and space implies velocity matching, since ultimately the two are tied together (in a source-free region) by the wave equation, as discussed further in the next subsection.

From our sampling point of view, what we want are speakers having a radiation pattern that serves well as an *interpolation kernel* (the spatial shape of one sample) for reconstructing a soundfield from its samples in one direction leaving the speaker array. Like Huygens, we want to neglect velocity and work only with pressure samples, but generate the correct velocities indirectly using pressure differentials along the array and across time, when the speakers are close enough for this to work, as they are in a valid sampling grid.

Another argument in favor of a source-listener separation plane is that there can be no *standing waves* when sound generally propagates *from* a set of sources *to* a set of listeners. Standing waves could pose problems if/when our microphone/speaker array happens to line up with a node line. A soundfield is *uncontrollable* and *unobservable* at nodes of vibration, leading to possible degeneracies in implementation. A progressive wavefield cannot have these problems.

²⁰This spectral partition issue is related to the classic “panning problem” in which low frequencies see a 3dB boost relative to high frequencies, due to coherent versus noncoherent summation from stereo speakers for an off-axis listener: https://ccrma.stanford.edu/~jos/sasp/Panning_Problem.html

It can also be taken as a simple *design decision* that we want our wavefronts to cross a separation plane from sources to listeners. This implies we are not trying to synthesize the reverberant field like WFS, and we set up our arrays in normal acoustic environments, as opposed to the anechoic environments called for by WFS. WFS solves for signals at all speakers enclosing the listening space to produce the desired interior field, even when it contains sources, and including any reverberation. We are less ambitious with Huygens Arrays, and we can expect more robustness and intuition-guided design freedom as a result.

3.6 Interpolation Accuracy

In a uniform speaker array (line or plane) that can be considered a “spatial digital to analog converter (D/A)”, the speaker’s radiation pattern $g(\theta)$ plays the role of “sampling kernel,” or “reconstruction lowpass-filter impulse-response” used in D/A conversion.

A speaker’s radiation pattern is naturally given as a function of *angle* relative to the speaker’s central axis. It may therefore appear to be a problem that our acoustic “samples” are changing (diffracting) as they propagate away from the speaker. However, this is ok as long as there is *any* distance from the speaker array at which the soundfield has been reconstructed (both pressure and velocity versus position). Beyond that, the wavefront “takes care of itself” (consider the more complete Huygens-Fresnel principle [Miller, 1991]).

In the far-field case (PBAP), we observe only traveling plane waves, where pressure and velocity are proportional to each other. In the case of a plane wave impinging on the array at angle $\theta = 0$, the velocity is obviously in the correct direction by symmetry. For an angled plane wave, it is easy to show that the velocity angle is correct far downstream from the array, since the contribution of each point-source along the array becomes a plane wave, and the lines of constant phase are along the desired angle due to the relative timing of the point sources. Having the pressure of a plane wave propagating in the desired angle means that its velocity points in that direction as well.

It is more difficult to show velocity reconstruction in the near-field case (general HA), where diverging wavefronts are sampled and re-emitted by the array. In this case, we need to show that the pressure samples add up to give the both the correct pressure and velocity for continuing the spherical wavefront expansion from the source. We have found simulation results to be helpful in the absence of analytical results (see Appendix A). There is no error in the timing of the secondary wavefronts, making the simulation results look great (a reconstructed plane wave looks very planar), but there is in principle amplitude error along the synthesized wavefront caused by the individualized $1/R$ factors. This error declines in relative terms downstream from the array.

While the instantaneous pressure $p(\underline{x})$ along a line or plane does not determine the corresponding velocity, unless we can assume a traveling plane wave, etc., the wavefront pressure $p(\underline{x}, t)$ over a nonzero time interval *does* determine the particle velocity $u(\underline{x}, t)$. In fact, the wave equation itself can be integrated to compute it. The time interval creates an interval of pressure history which allows the pressure gradient to be calculated (in a progressive wave), and the pressure gradient drives the velocity in the absence of a coincident source. Specifically, the sound velocity is given by the time-integral of the pressure-gradient divided by the air’s mass-density ρ (Newton’s second law of motion $f = ma$, which, in the wave equation, appears as $\partial p(x, t)/\partial x = -\rho \partial u(x, t)/\partial t$, neglecting high-order terms [Morse and Ingard, 1968, p. 243]).

It thus suffices to reconstruct sound pressure as a function of time and position along any parallel line (or plane) in front of the array. Since we assume no sources along that line or plane, the velocity is determined by the pressure in any spatial neighborhood.

In the case of the line array, we are only concerned about the velocity vector lying in the plane determined by the line array and any vector from the line array to the sources or listeners, all of which are assumed to lie in one plane.

For a planar microphone/speaker array, the velocity vector can point anywhere in the half-space from the array toward the listeners.

3.7 Evolving Radiation Pattern (Sampling Kernel) due to Diffraction

It is well known that the far-field radiation pattern of an acoustic source is proportional to the spatial Fourier transform of the source’s radiating amplitude distribution. In optics, this is called Fraunhofer diffraction theory [Goodman, 2005]. By linearity of the Fourier transform, it follows that if the speaker array emits valid sampling kernels near the array (*i.e.*, the radiation patterns overlap and add to a constant overall gain for a plane wave), then the far-field patterns will similarly overlap and add to a constant gain. All points in between must be valid as well, being progressive diffractions of the source distribution, but a rigorous proof with quantified approximation error would be nice to see (there are always terms to neglect, and it is good to be mindful of them).

3.8 Specific Sample Shapes (Speaker Radiation Patterns)

Perhaps the most obviously ideal sampling kernel $g_y(x)$ for a speaker centered at $x = 0$ within a line-array with cell-width X is the rectangular pulse:

$$g_y(x) = \begin{cases} \frac{1}{X}, & |x| < \frac{X}{2} \\ 0, & \text{otherwise} \end{cases}$$

That is, each speaker in the array pushes a unit volume of air in one second (unit area). Such a speaker array can be regarded as a series of contiguous pistons, each of width X .

The unit-area pulse function is not bandlimited. A more graceful choice is the “spatial sinc function,” which is the ideal sampling kernel used in bandlimited audio sampling:

$$g_y(x) = \frac{1}{X} \text{sinc}\left(\frac{x}{X}\right)$$

where

$$\text{sinc}(x) \triangleq \frac{\sin(\pi x)}{\pi x}.$$

It is well known that the sinc function is the Fourier transform of a rectangular pulse, and vice versa. Therefore, rectangular-pulse samples at the line array propagate and diffract to become overlapping sinc functions in the far field, while sinc-shaped radiation patterns diffract to become adjacent or overlapping rectangular spatial bands in the far field. Of course, a true sinc function can never be implemented precisely because it is infinite in spatial extent. Still, it is interesting to imagine disjoint spatial samples in the far field, and consider whether approaching that might be desirable in some situations, such as delivering different program material to different listening positions (a current problem in automotive sound).

The most typical example in practice is a *circular* driver—an ordinary circular speaker cone. In this case, the far field shape is the so-called “Airy function” involving Bessel function J_1 .²¹

²¹<https://adriftjustoffthecoast.wordpress.com/2013/06/06/2d-fourier-transform-of-the-unit-disk/>, <http://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf>

In addition to the rectangular piston and sinc function, or circular driver and Airy pattern, any spectrum-analysis window $w(x)$, or its Fourier transform $W(\omega)$, having the constant-overlap-add (COLA) property

$$\sum_{n=-\infty}^{\infty} w(x - nX) = \text{constant}$$

can be used as a spatial sampling kernel, with the added stipulation that w should have an effective width on the order of X (one spatial sample) in order that spatial resolution be maximized. In other words, it suffices for the speaker radiation gains at a particular distance away from the array to overlap-and-add to a constant gain-versus-position at the speaker spacing X used. In the limit of infinite sampling density, *all* windows w have the COLA property, including hemispherical, which is in part why Huygens’ Principle works as a spatial D/A converter with its spherical/circular wavefronts. A review of COLA windows and their design is given in [Smith, 2011].

Spatial oversampling gives more flexibility in the choice of radiation pattern. For example, doubling the spatial sampling rate allows the radiation patterns to overlap by an additional factor of two with no loss of spatial resolution along the line array. On the other hand, that factor of two could be given to the aliasing cutoff frequency, which is probably a better use of it.

As in audio interpolation, *windowed sinc* interpolation can be used in practice [Smith and Gossett, 1984], if a speaker driver can be devised to generate that radiation pattern at some distance from the speaker.

The spatial sampling kernel should ideally be frequency *independent*, but this is never the case for typical speaker systems. Instead, typical speakers look like point sources at low frequencies, radiate efficiently and widely at wavelengths comparable to the speaker diameter, and begin to “spotlight” increasingly at higher frequencies (where the diameter is multiple wavelengths).

Due to the naturally narrowing spatial beam-width with increasing frequency for typical speaker drivers, a sufficient sampling density for high frequencies corresponds to heavy oversampling (spatially) at low frequencies. PBAP and its extensions therefore should either be implemented in separate frequency bands (multiband PBAP is discussed below), or using a new kind of speaker having a frequency-independent radiation pattern that sums to a constant when the speaker outputs are all added together at any point of the listening region. One solution is to approximate a point source (see Appendix B), for which the sampling kernel is a substantially identical sphere for all speaker drivers much smaller than a wavelength. Such speakers radiate inefficiently, but they are already widely used at the low-frequency end in practice (subwoofers are smaller than most of the wavelengths they must produce). The main problem with a single set of point-source-approximation drivers is that they must be packed very densely for the high frequencies and also have long-throw excursion for low frequencies—expensive. Furthermore, there is always intermodulation distortion in any wideband driver (*e.g.*, Doppler shift of high-frequency components by low-frequency excursion, which nobody apparently compensates).

4 Multiresolution Spatial Sampling Arrays

Since typical speakers have a frequency-dependent radiation pattern, it makes sense to take a multiband approach and combine an array of “woofers” with a denser grid of midrange speakers and a yet denser array of “tweeters,” for example. In principle, each speaker is assigned a frequency band containing wavelengths up to its diameter or so, and larger wavelengths can be pushed using

additional volume-velocity drive. Of course, one large 17-meter (56') diameter diaphragm can handle the entire audio band, but then we have poor spatial resolution at high frequencies.

4.1 Multiresolution Coaxial Drivers Array

One approach suitable for 2D arrays is coaxial heterogeneous drivers, as depicted in Fig. 9. The Kenwood KFC-1695PS, for example, provides three concentric drivers with diameters 6 1/2, 1 9/16, and 1/2 inches (75 Watts RMS). This could be packed into a rectangular panel and treated as a multiresolution Huygens Array or PBAP/VBAP system.

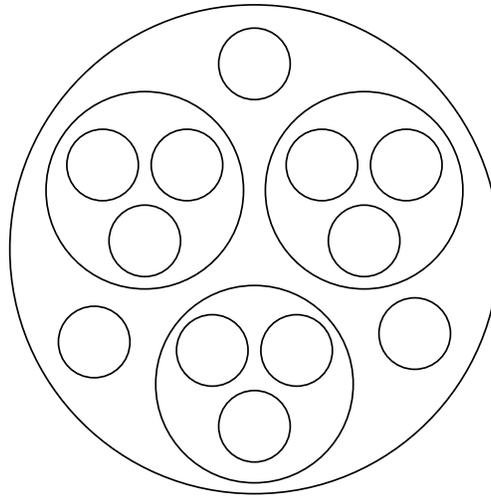


Figure 9: Example of a coaxial multiresolution driver geometry.

One can also imagine this type of array implemented using pipes of various diameters, with the pistons driving the larger pipes operating behind the pistons of the smaller pipes, where each piston is driven at the end of a narrow rod that passes through a small hole in the piston(s) behind.

4.2 Multiresolution Line Array

A simpler geometry for the line-array case is parallel *strips* covering different bands (Fig. 10).

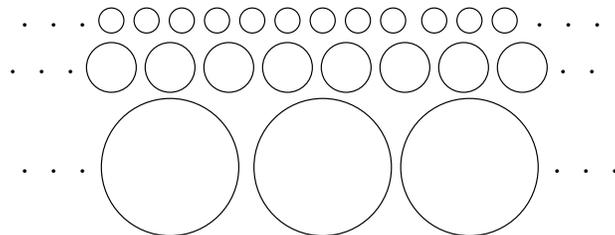


Figure 10: Example multiresolution line array implemented using strips of drivers at different sizes.

Of course, the drivers do not have to be circular, and some tweeter geometries are non-circular.

Also, circular drivers can drive the base of horns having rectangular exit apertures that tessellate a linear or planar region.

4.3 Multiresolution Sampling Considerations

Intuitively, spatial sample reconstruction requires that each driver be capable of pressurizing a fraction of one wavelength (in its band) to the desired pressure level. Thus, while the speakers are ideally very small, on the order of drinking-straw diameters at high frequencies, the drivers need a long excursion in order to push enough air down the straw to achieve the desired pressure within the subwavelength zone being served. It is straightforward to calculate the maximum piston excursion needed for a given sound pressure level and lowest sinusoidal frequency. Dividing up the spectrum into frequency bands makes this easier.

The exact shape of the drivers is not important when they are smaller than a wavelength, only that they can pressurize their subwavelength zone as needed. Perhaps the easiest solution conceptually is a grid of contiguous square pistons. In that case it is easy to see that it must work very well, because the pistons can generate the wave propagation leaving the surface in great detail.

In classical “critical sampling,” there would two pistons per wavelength, one to push while the adjacent piston pulls. In practice, critical sampling may cause undesired noise due to turbulence, since there is no guarantee that laminar flow is maintained. Obtaining silent pressurization of half a wavelength may prove difficult at high sound pressure levels, so spatial oversampling helps.

4.4 Multiresolution Speaker Systems

We are all familiar with speaker cabinets containing woofers, midrange, and tweeters, etc. Each cabinet can be considered a monaural multiresolution speaker system, typically three-way. Additionally there is often a subwoofer somewhere putting out the deep bass.

The Kenwood JL-840W speaker systems are four-way:²² They use four circular drivers having diameters 30, 12, 6, and 3 cm.²³ The crossover frequencies are at 2, 5, and 10 kHz, which is at $ka = 5.5$ for the speaker driving below crossover, where k is wave number in radians per meter, as usual, and a is speaker radius.²⁴ The nominal total frequency range of the system is 20–20 kHz, but amplitude drops off in the woofer (30 cm) for wavelengths much greater than the diameter, which must be compensated by extra drive. The super-tweeter, tweeter, and midrange drivers have diameters on the order of one to two wavelengths. The woofer high-end is near that range, but must handle all lower frequencies as well.

We can extend existing N -way speaker systems to multiresolution line arrays in a straightforward manner. We use the term Huygens Array (HA) to refer to any collection of drivers used as a spatial imaging array, and in particular, Huygens Octave Panels (HOP) will refer to multiresolution

²²<https://www.acs.psu.edu/drussell/Demos/BaffledPiston/BaffledPiston.html>

²³These are “octave spaced” from midrange to super-tweeter, with an extra large woofer. To adhere to octave spacing, the woofer diameter could be changed to 24 cm and the lower frequencies could be taken over by a subwoofer where the woofer leaves off.

²⁴This implies the lower speaker diameter is 1.75 wavelengths at crossover while the upper speaker diameter is 0.7 (midrange) or 0.88 (tweeter and super-tweeter) wavelengths at crossover. The geometric means of these two diameters in wavelengths are 1.1 (woofer-midrange) and 1.24 (other two crossovers) wavelengths. The general tradeoff is that driving with diameter less than a wavelength is inefficient (below cutoff), but yields nicely omnidirectional radiation, while driving with diameter much larger than a wavelength becomes highly directional.

line arrays having *octave* divisions.²⁵

4.5 Four-Way Huygens Arrays

A four-way line array inspired by the Kenwood JL-840W four-way speaker system can be made using four parallel rows of contiguous speaker drivers, packed together as closely as possible, as indicated in Fig. 11.

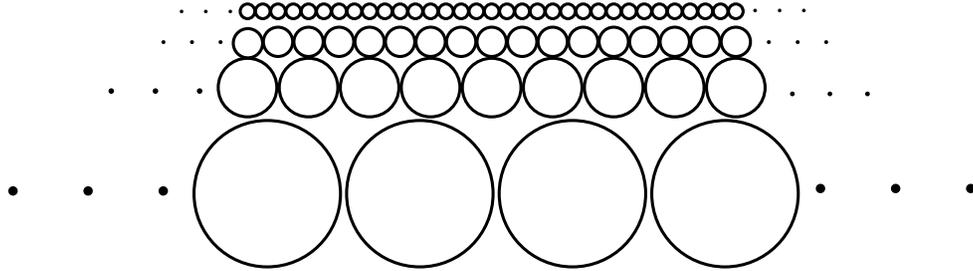


Figure 11: Example of a four-way multiresolution Huygens array implemented using strips of drivers at four sizes, scaling by a factor of two from one strip to the next (octave strips), with extra large woofer row.

The extra large woofer breaks the pattern, suggesting using two of them as a stereo pair. Doing this makes room for a center channel, as shown in Fig. 12.

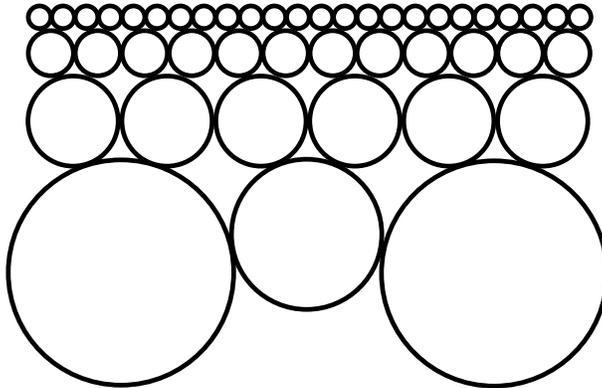


Figure 12: Four-way multiresolution Huygens array using the same driver diameters of Fig. 11, offering only stereo for the woofer, plus a 20 cm center-channel. Forty-five channels = $24+12+6+3$. Nominal width 0.72 meters = $6 \times 12 = 12 \times 6 = 24 \times 3$ cm.

²⁵We avoid the term “Huygens Octave Array” (HOA), which would be nice to use for planar arrays as in Fig. 9, due to the common use of HOA as “Higher Order Ambisonics”. It is convenient to organize multiresolution line arrays into panels, so not much is lost.

4.6 Huygens Octave Panels (HOP)

In audio, octave-based resolution is very common. If the upper limit of the audio bandwidth is taken to be 20 kHz, then the top octave covers 10–20 kHz, the next octave down is 5–10 kHz, and so on, giving crossover frequencies at 10, 5, 2.5, and 1.25 kHz, and 612, 312, 156, 78, and 39 Hz, which can be taken as the crossover to the lowest octave including 20 Hz. Thus, the complete audio spectrum spans 10 octaves, and the 9 crossover frequencies are given in kHz by 20×2^{-n} , where $n = 1, 2, \dots, 9$ is the octave number counting down from the top.

It is common to use a subwoofer to take the low end (spanning 20–80 Hz for THX, 20–100 Hz for “pro”, or 40–200 for “consumer” quality level, etc.²⁶), so that only 8 (starting from 78 Hz) or 7 (from 156 Hz) octave bands are needed from a multiresolution array.

It is instructive to look at the wavelengths in each band, since each speaker-driver diameter needs to be on the order of a wavelength. Let’s define the center-frequency of each octave as the geometric mean of its limits, so $f_c(n) = 20 \times 2^{-(n+1/2)}$, $n = 1, 2, \dots, 9$, which gives

$$\underline{f}_c = [7071, 3536, 1768, 884, 442, 221, 110, 55, 28]^T$$

for the center frequencies in Hz. Then for a speed of sound $c = 343$ m/s, using $k = \omega/c = 2\pi f/c = 2\pi/\lambda$, we obtain the center-frequency wavelengths $\lambda_c(n) = c/f_c(n)$ to be

$$100\underline{\lambda}_c = [4.85, 9.70, 19.40, 38.81, 77.61, 155.22, 310.45, 620.89, 1241.79]^T \quad \text{cm}$$

or

$$\underline{\lambda}_c'' = [1.91, 3.82, 7.64, 15.28, 30.56, 61.11, 122.22, 244.45, 488.89]^T \quad \text{in}$$

or

$$\underline{\lambda}_c' = [0.16, 0.32, 0.64, 1.27, 2.55, 5.09, 10.19, 20.37, 40.74]^T \quad \text{ft.}$$

We see that even consumer quality level wants a lowest-octave driver diameter on the order of 10 feet, and THX and pro quality want a 20-foot cone! Practical systems rarely go for such large drivers. Instead, we settle for the top five or six octaves, and drive the lowest octave with additional gain to get the desired power level. In other words, the low-end speaker(s) operate in a “rolling off” region, driving only a fraction of a wavelength, and so they require a 6 dB boost for each halving of frequency in that zone. We can make up for driving less than a wavelength by using extra power, to the extent no audible turbulence is generated.

Another point to consider is that sound localization is only meaningful down to a fraction of a wavelength, and wavelengths much larger than our heads cannot be localized based on the steady-state soundfield, because our ears get nearly identical signals from the field.²⁷ We generally localize a sound based on its higher frequency components, such as above 500 Hz where azimuth changes cause noticeable Interaural Intensity Differences (IID), i.e., audible “head shadowing”. The localization determined during a sound’s onset, which normally has the most high-frequency content, tends to persist psychologically even after the high-frequencies that localized it have faded away.²⁸

²⁶2019-09-30: <https://en.wikipedia.org/wiki/Subwoofer>

²⁷<http://acousticlab.org/psychoacoustics/PMFiles/Module07a.htm>

²⁸A great demonstration of this effect in stereo, is to turn on a gated sinusoid in, say, the right speaker, cross-fade the signal over to the left speaker, then pull out the right speaker cable and hand it to the befuddled listener who still hears the tone coming out of the right speaker. Thanks to Bill Putnam for showing me this one.

4.7 Five-Band HOPs

Extending the four-way Huygens Array of §4.5 to five bands, and enforcing octave-band design all the way down yields a five-band Huygens Octave Panel for our consideration.

Figure 13 shows the driver outlines for two five-octave HOPs side by side. Each panel is about half a meter wide, so a 4m wall needs 8 HOPs.

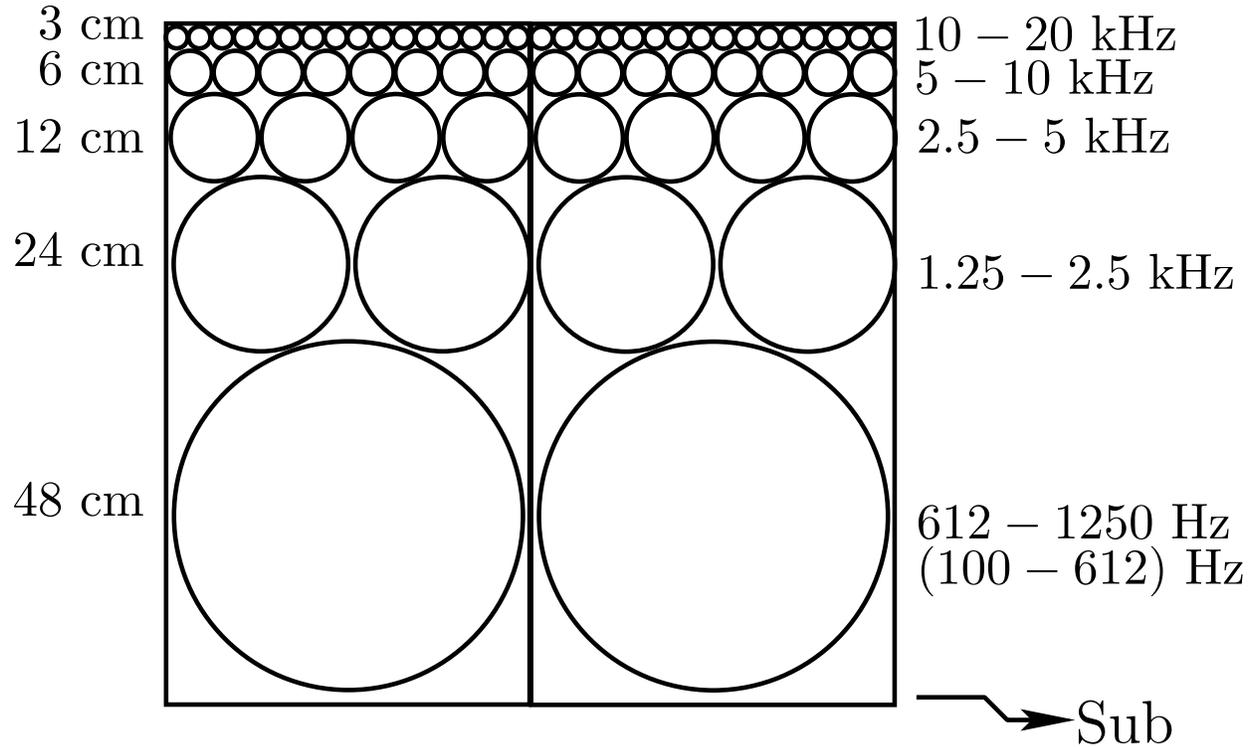


Figure 13: Two modular five-octave Huygens arrays using the smallest three driver diameters of Fig. 11, plus two more octave-band extensions downward. Speaker diameters 3, 6, 12, 24, and 48 cm. Thirty-one channels per panel = 1+2+4+8+16. Channel 32 can be used for a subwoofer. Nominal panel width is 1/2 meter, so 1 meter for the pair.

The top three rows use drivers of the same sizes used in the Kenwood JL-840W four-way speakers (3, 6, and 12 cm), while the fourth row is reduced to 24 cm (down from 30 cm) in order to continue the octave divisions, and a fifth row is added employing 48 cm (19 in) drivers. Fifteen inch speakers are very common, and could be used as a compromise for the bottom row, and subwoofer drivers of diameter 18 and 20 inches appear to be available, but probably quite expensive to use in an array system.²⁹

²⁹<https://www.svsound.com/blogs/svs/strengths-and-pitfalls-of-big-subwoofer-drivers>

4.7.1 HOP Stage Angle

An immediate question that arises is how large is the maximum angle-of-arrival for a PBAP array made using such a HOP? From §2 on page 8, we have

$$\Delta_x \sin(\theta_{\max}) < \frac{c}{2f_{\max}}$$

Assuming it is possible to pack the speakers contiguously, and setting $c = 343$ m/s, we obtain for a top row using driver diameters 0.03 m as in Fig. 13

$$\theta_{\max} < \sin^{-1} \left(\frac{343}{2 \cdot 20,000 \cdot 0.03} \right) \approx 0.290 \text{ rad}$$

or 16.6 degrees, giving a 33.2-degree stage at f_{\max} . The same result is obtained for the highest frequency in each band of this example HOP, due to its strict octave scaling. The alias-free angle-of-arrival range is unfortunately narrow, but it is calculated at the high edge of each band. At the bottom of each band, we obtain a much nicer 70 degree sound-stage. At the center-frequency of each band, defined as the geometric mean of the band edges, we find that a 47.7-degree sound-stage is supported without spatial aliasing, which is not too bad. Again, when pushing the angle-of-arrival limits, only certain frequency bands near the top of each octave band start to spatially alias, and the spectrum as a whole is likely to keep the brain fusing everything together at the intended angle of arrival.

To reduce spatial aliasing or increase the stage width of this HOP, it is tempting to shift the operating range of each speaker down by some fraction of an octave, thereby increasing the stage-width while making each speaker a less efficient radiator, a better approximation to a point source, and a finer spatial sample within the array. Decreasing driver diameter is of course equivalent to decreasing the frequency range over which it operates. However, since most people cannot hear frequencies near 20 kHz, it makes sense to push first on downward frequency scaling. This point is pursued further in §4.9 below.

The need to downscale the frequency band served by each speaker driver is a fundamental problem with circular drivers laid out in a row. Alternatives driver geometries include (1) overlapping driver cones (perhaps using a shared membrane across multiple drivers instead of individual cones), (2) staggered packing of two or more identical rows, as in the hexagonal mesh, or (3) square or rectangular horn drivers with contiguous exit-apertures driven by long-throw pistons.

To summarize the situation from an elementary spatial-sampling perspective, the maximum stage width determines the minimum wavelength seen by the array, and we need two or more drivers per wavelength for proper spatial sampling. The maximum width (diameter) of each driver must be less than half the minimum spatial wavelength seen by the array (which can be made arbitrarily large by restricting the stage angle). If wide stage angles are to be supported, then every driver must be smaller than spatial wavelengths it is emitting. They're all approaching "simple sources" emitting omnidirectional radiation patterns.

The crossover frequency 612 Hz for the low end of the 48-cm woofer in Fig. 13 is based on the preferred driving frequencies and dispersion pattern for a conventional loudspeaker of that size, as chosen in the Kenwood JL-840W four-way speaker system. We see that stage-width and spatial-aliasing considerations argue for a smaller crossover frequency, perhaps even an octave lower, if adequate undistorted power can be obtained covering that range.

The nominal lowest frequency in the fifth row (612 Hz) is well above the typical subwoofer crossover range of 80–200 Hz. This means we have almost three missing octaves between this particular five-band HOP and the ideal subwoofer taking over at 80 Hz. Even downscaling the frequency bands by an octave (discussed above), two missing octaves remain. Raising the subwoofer cutoff from 80 to 160 gets one more, leaving almost a one-octave gap, which can be reduced further from either side. A sixth row in the HOP would call for a 96 cm speaker diameter (38 inches or 3.15 feet), which does not appear to be practical using conventional driver technology. We therefore presently allow the bottom row to go undersampled, and cross-fade over to VBAP signals, or even simple stereo using the left $N/2$ speakers as a left channel and the right $N/2$ speakers as a right channel (panning each virtual source accordingly in the stereo field). Fewer than $N/2$ speakers can be used for wider separation but less power. Experiments are needed to determine best practices in this (undersampled) frequency range. At frequencies immune to spatial aliasing, we expect to be limited only by driver quality and noise.

4.8 Upper, Middle, and Lower HOPs

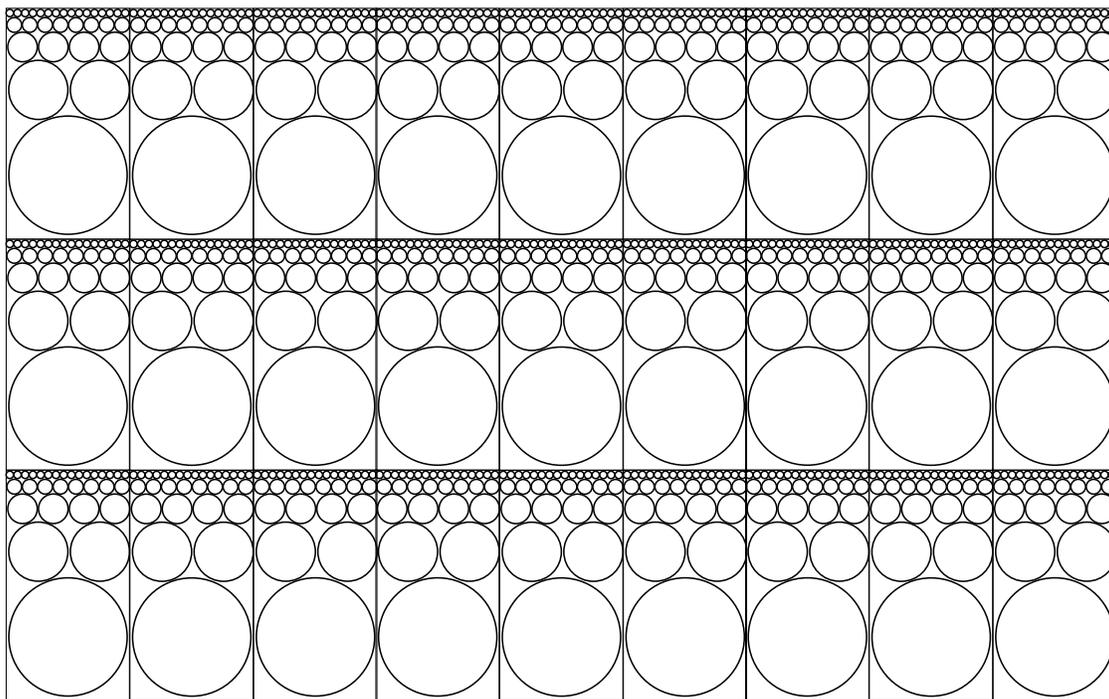


Figure 14: Three rows of nine HOPs \approx 3 m high and 4.5 m wide

Figure 14 illustrates a simple extension from one to three multiband line-arrays for the purpose of providing three elevation levels. In the vertical dimension, it can be driven as three-channel VBAP, or stereo with a center channel. While azimuth perception is accurate to \approx 1 degree straight ahead, elevation perception is an order of magnitude less accurate, so a cruder sampling of the vertical axis is appropriate.

4.9 Nominal Design Guidelines

In general, we want

1. Subwoofer(s) up to f_1 Hz
2. VBAP from f_1 to f_2 Hz, or some simple stereo reduction
3. HOPs from f_2 to f_3 , where f_3 depends on listener age, etc.

Crossover frequency f_1 is normally set near the lower limit of stereo perception, which we are taking to be around 80 Hz (§2.12 on page 16).

Crossover frequency f_2 is chosen to be the lowest frequency at which the lowest octave of the HOP is adequately sampled, following the analysis of §2 on page 8. For the five-band HOP example, it is $f_2 = 612$ Hz (§4.7 on page 30). From a spatial sampling point of view, f_2 depends on listening geometry, particularly the stage-width, minimum-source-distance, and listener-distance from the array.

The setting for the upper limit f_3 can be based on how much high-frequency spatialization is desired. For older listeners,³⁰ the top row is normally inaudible in arrays such as the previous example and in the five-band HOP examples above, such as Fig. 13 on page 30. Therefore, the top row can either be simply omitted for older listeners, or, as a compromise preserving at least the presence of the high end for the occasional younger listener, replaced by conventional stereo tweeters (two-channel VBAP), or any number of tweeters in an undersampled VBAP row, or coaxial mounts along any lower row. Also, at such high frequencies, wavelengths are so small that left-right processing is noncoherent, so that left-or-right “presence” is the only available cue for perception. Thus, an undersampled top row should work quite well for those who can hear it at all. Finally, much auditory direction perception research ignores frequencies above 10 kHz.

4.10 An Equilateral Triangle Design Example

Suppose we build a linear array forming an equilateral triangle with the centered listener, as in typical stereo speaker placements. Denote the side-length by S . The listener distance is then $S\sqrt{3}/2 \approx 0.87S$ from the array. Suppose we want to sit back three meters from the array for normal listening. Then $S \approx 3.5$, which is about 7 HOPs across, following the example of Fig. 13. To reduce edge effects, we can add one more HOP on each side to obtain nine HOPs, as shown in Fig. 14. As discussed in §4.6 on page 29, the five-band HOP places the crossover to VBAP/stereo at $f_2 = 612$ Hz, or lower as discussed in §4.7 above, and the subwoofer nominally crosses over at 80 Hz, or wherever the VBAP/stereo band lower limit may be.

4.10.1 Stereo As Two Spatial Samples

It is interesting to consider the woofer upper frequency f_1 in this example from a spatial sampling point of view. Suppose for simplicity that stereo will be used immediately above f_1 . The array width is S , and for simplicity, consider the effective stereo speaker separation to be $\Delta = S/2$ when half the speakers comprise the left channel and the other half comprise the right, or $\Delta \approx S$ when using only the extreme left and right drivers, or various choices in between. Considering the stereo

³⁰See International Standard ISO-7029:2017 (3rd edition): <https://www.iso.org/standard/42916.html>

pair as two soundfield samples, stereo should take over for the subwoofer before the wavelength shrinks to $\lambda_1 = 2\Delta$ (critical spatial sampling). The formula for f_1 is then

$$f_1 = \frac{c}{2\Delta}, \quad \Delta \in \left[\frac{S}{2}, S \right].$$

In the above example with $S \approx 3.5$, we obtain $f_1 = 343/(2\Delta) \in [49, 98]$ Hz, which includes the desirable 80 Hz setting. In the much smaller four-way Huygens Array of Fig. 11 on page 28, the stereo speaker separation (center to center) is close to $S = 1/2$ meter, where we obtain the range [343, 686] Hz, which is quite high compared to normal subwoofer crossovers. It is apparent that in typical listening geometries the subwoofer cutoff due to perceptual limits is comparable to that obtained by considering the stereo speakers as a pair of spatial samples needing to be outside each other’s “high-correlation zone.”³¹

5 Conclusions and Future Work

We have approached the problem of soundfield synthesis from the point of view of spatial sampling theory. Starting with the simplest case of far-field sources, we derived Planewave-Based Angle Panning (PBAP) which can be considered both a special-case of wave field synthesis (WFS) for distant sources without reverberation, and a path for enlarging the “sweet-spot” in Vector-Base Amplitude Panning (VBAP). Linear, planar, and more generally distributed arrays were considered, with the most general case termed Huygens Arrays (HA). Considerations of basic sampling theory led to straightforward practical guidelines on how to use the arrays, such as determining speaker spacing and size, speaker radiation pattern (interpolation kernel), maximum source angle, and how closely a source can approach the array. Finally, various multiband systems were considered, allowing each driver to focus on a particular band, such as an octave band. Multiband linear arrays, for use with conventional stereo and/or subwoofer(s) at low frequencies, were found to be particularly attractive and relatively practical.

Interesting next steps include studying the overlap-add properties of the near- and far-field radiation patterns for circular and other commonly used drivers. The SFS software (linked in Appendix A) could be adapted for this purpose. The simple case of point-source drivers is given a start in Appendix B.

The shape of a spatial sample (speaker radiation pattern) can be optimized for various purposes. One is to make it easy to produce using conventional drivers, such as the simple matrix of square pistons considered in §3.8. Another is to optimize far-field considerations, such as minimizing cross-talk between angular directions (also considered in §3.8). Other optimization criteria can be imagined. For example, if the center of the array at a particular listening distance is given special status, then the array can be optimized for that region in various ways, including minimizing the spatial width of the radiation pattern and thence array truncation effects.

³¹A single-frequency soundfield is always highly correlated at distances less than a quarter wavelength or so, even when it is randomly constructed as a sum of plane waves from all directions with random phases (a “diffuse field” [Pierce, 1989; Beranek, 1986; Smith, 2010]). Even in a richly reverberant environment, one can imagine “correlation bubbles” on the order of the wavelength at each frequency.

6 Acknowledgments

Thanks to Hyung-Suk Kim for fruitful discussions on the topic of this paper, and to Madeline Huberth for reporting errata and providing helpful suggestions.

A Relevant Software

A.1 Soundfield Synthesis (SFS) in Matlab

The Soundfield Synthesis (SFS) Toolbox for Matlab [Wierstorf and Spors, 2012; Ahrens, 2012]: <https://github.com/sfstoolbox/sfs>

Note that there are many useful pointers to references in the software documentation.³²

A.2 PBAP in Faust, Integer Delays

Below is the initial interactive test program in the FAUST language for integer-delay PBAP on the eight-channel array described in §2.15 on page 17. Integer delays are extremely efficient computationally, but assume fixed source positions. The lowpass filter is provided to facilitate experiments with the perception of spatial aliasing at various cutoffs.

```
N = 8; // number of channels (speakers)
MAXDELAY = 1024; // maximum delay-line length for each channel
lpf = lowpass(3); // Butterworth lowpass filter to use (filter.lib)

declare name "Far Field Wave Field Synthesis (FFWFS) Simple Tests";
declare author "Julius O. Smith (jos at ccrma.stanford.edu)";

import("oscillator.lib"); // saw2, pink_noise
import("filter.lib"); // lowpass, smooth, ...

// GUI:
level = hslider("v:FFWFS/[0]Level (dB)", -10, -70, 10, 0.1);
del = hslider("v:FFWFS/[1]Interspeaker Delay (samples)",0,-misd,misd,1)
      with { misd = int(0.5*MAXDELAY/N);}; // max inter-speaker delay
freq = hslider("v:FFWFS/[2]Frequency (Hz)",440,20,10000,1);
cutoff = hslider("v:FFWFS/[3]Lowpass Cutoff (Hz)",1000,20,20000,1);
nsw = checkbox("v:FFWFS/[4]Sawtooth (instead of Pink Noise)");

// Signal Processing:
amp = level : db2linear : smooth(0.999);
signal = _ + select2(nsw,pink_noise,saw2(freq)) : lpf(cutoff);
ffwfs(i) = delay(MAXDELAY,int(MAXDELAY/2)-i*del);
process = signal * amp <: par(i,N,ffwfs(i));
```

³²<https://sfs.readthedocs.io/en/3.2/problem/>

B Speakers as Spatial Samples

The *polar pattern* for a microphone or loudspeaker is its gain along a circle of constant radius away from the diaphragm/driver. For a spherical-wave “point-source”, the polar pattern is simply a constant at each radius, *e.g.*, $p(\theta) = p_1/r$, where p_1 denotes the pressure-scaling at $r = 1$ and r denotes the distance from the center of the source.

Since our speaker arrays are typically flat, we need to calculate a *slice* through the polar pattern along a *listening line* (or plane) which we will take to be parallel to the array and to the x axis, as shown in Fig. 15. The polar-pattern slice is then be considered as one *sample* (interpolation kernel) used to reconstruct the soundfield at a distance z_l from the array.

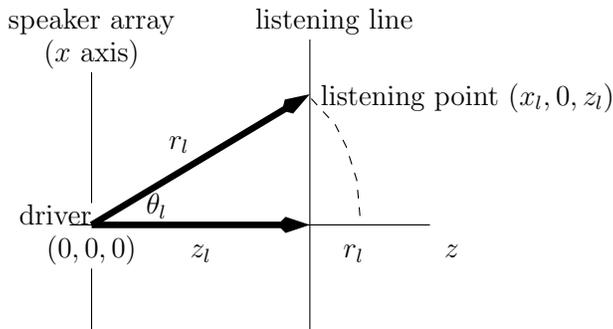


Figure 15: Geometry of x -slice through polar pattern to evaluate the effective sampling kernel along a line parallel to the speaker array.

B.1 Point-Source Speakers

Consider a *spherical* secondary source (“point source”) located at $\underline{x} = \underline{0}$ as shown in Fig. 15. Then the pressure amplitude along the listening line $(x, 0, z_l)^T$ for all x is given by

$$p_l(x) = p_l(0) \frac{z_l}{\sqrt{z_l^2 + x^2}} \triangleq \frac{p_l(0)}{\sqrt{1 + \left(\frac{x}{z_l}\right)^2}} \triangleq \pi_l(\theta) = \frac{p_l(0)}{\sqrt{1 + \tan^2(\theta)}} \quad (9)$$

where $p_l(0)$ denotes the pressure amplitude observed from the point source at $\underline{x} = (0, 0, z_l)^T$, and $\theta \triangleq \tan^{-1}(x/z_l)$ denotes the angle of the line from the source center to the line-array point at x , *i.e.*, $\underline{x} = (x, 0, z_l)^T$. This polar-pattern slice is plotted in Fig. 16 for the triangular array+listener geometry such as described in §4.10 on page 33, with $p_l(0) \triangleq 1$. The top plot shows $p_l(x)$ over a 12 m width centered about a 6 m wide array, and the bottom plot shows $\pi_l(\theta)$ for $\theta \in [-\pi/2, \pi/2]$, thereby covering the entire axis containing array.

In addition to the gain variation $p_l(x)$ in Eq. (9), we also have *phase effects* that we don’t have when keeping radius r constant and only varying θ . The envelope $p_l(x)$ in Fig. 16 is due only to spherical spreading loss according to $1/r$. The difference in propagation distance between $p_l(0)$ and $p_l(x)$ is

$$r_d(x) \triangleq r_l(x) - r_l(0) = r_l(x) - z_l = \sqrt{z_l^2 + x^2} - z_l = z_l \left[\sqrt{1 + \tan^2(\theta)} - 1 \right]. \quad (10)$$

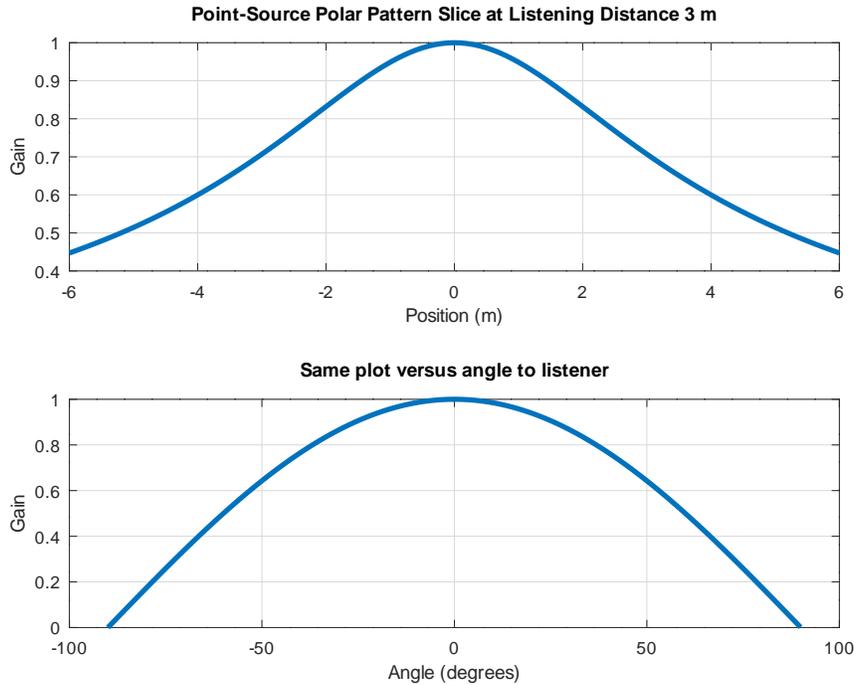


Figure 16: Point-source polar-pattern slice for a listening-line $z_l = \lambda$ away from the source. Top: Gain p_l versus position x . Bottom: Gain π_l versus angle $\theta \triangleq \tan^{-1}(x/z_l)$.

The resulting frequency-response $\text{re} \{ p_l(x) e^{jk r_d(x)} \}$ is plotted in Fig. 17 for 1 kHz (wavelength about a foot (1.126 ft)) and $p_l(0) = 1$, where $k = \omega/c = 2\pi \cdot 1000 / 343$ is the wavenumber (spatial radian frequency), which has made its appearance for the first time in these formulas.³³

For small $|\theta|$, we can use the approximation

$$r_d(x) \approx \frac{1}{2} z_l \tan^2(\theta) = \frac{x^2}{2z_l}.$$

B.2 Overlap-Add of Point-Source Frequency Responses

The listening point receives a sum of the radiating point-sources along the array. For an incident plane-wave traveling toward the listener, the sources are all “in phase”, emitting identical signals. Let’s focus on two adjacent sources for this case.

Figure 18 shows the magnitude of the sum of two identical point-sources separated by half a wavelength, and observed along a listening line one wavelength away from the two-source line array. Also shown is an overlay of the real parts of the component point-sources being summed, as in Fig. 17. Notice how the pressure sums coherently near the center, but largely cancels far away from the center, especially to the far left and right where the half-wavelength spacing of the sources leads to almost complete cancellation.

³³For a definition of “frequency response” and related terms, see, *e.g.*, [Smith, 2007a].

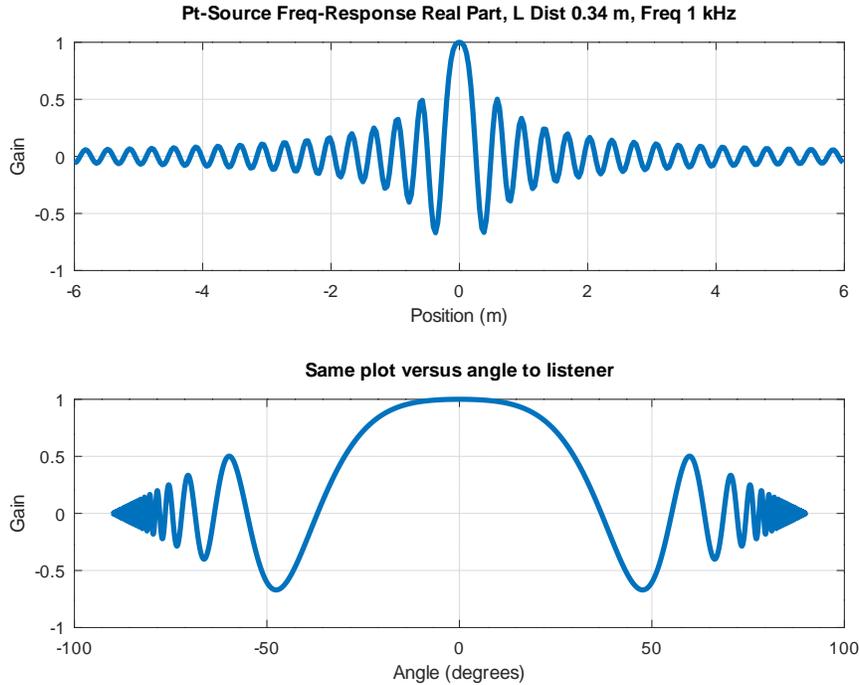


Figure 17: Real part of point-source frequency-response for a listening-line one wavelength away ($z_l = \lambda$). Top: Gain p_l versus position x . Bottom: Gain π_l versus angle $\theta \triangleq \tan^{-1}(x/z_l)$.

Consider now a larger array spanning 6 meters centered within a 12 m listening line, again one wavelength away. Figure 19(a) shows the OverLap-Add (OLA) result for the point-source frequency-response of Fig. 17, again using source-separation $\lambda/2 \approx 17.2 \text{ cm} \approx 6.76 \text{ in}$, corresponding to “critical sampling” for extreme stage angles $\pm\pi/2$ at 1 kHz. The Gibbs phenomenon is highly visible, strongly suggesting a tapering window to reduce array truncation effects (§2.14). Tapering sinusoidally to zero (Hann half-windows) over 3 wavelengths (6 source points) produces the result shown in Fig. 19(b). The normalized driving amplitude from each source is indicated by a small black circle.³⁴

While spherical waves do not give perfect-reconstruction at half-wavelength overlap-add, one wavelength out, we see from the dB plots in Fig. 19 that the error is quite small in audio terms. The “ripple” stays small down to well under wavelength $z_l < \lambda$ before source proximity effects increase the ripple significantly. Note, incidentally, that we are ignoring the reactive “mass” component of the point-source field near its center, which is non-propagating [Morse and Ingard, 1968].

The ripple reduces as we get farther out from the array ($z_l > \lambda$) because the interpolation kernels expand, giving more overlap. (Increase the number of sources by the same factor to see this more clearly without additional array-truncation error.) Since we normally listen to arrays at some distance away, we see that a larger issue than sampling density is array extent; looking

³⁴Matlab for these figures is available at <https://ccrma.stanford.edu/~jos/huygens/matlab.tgz>

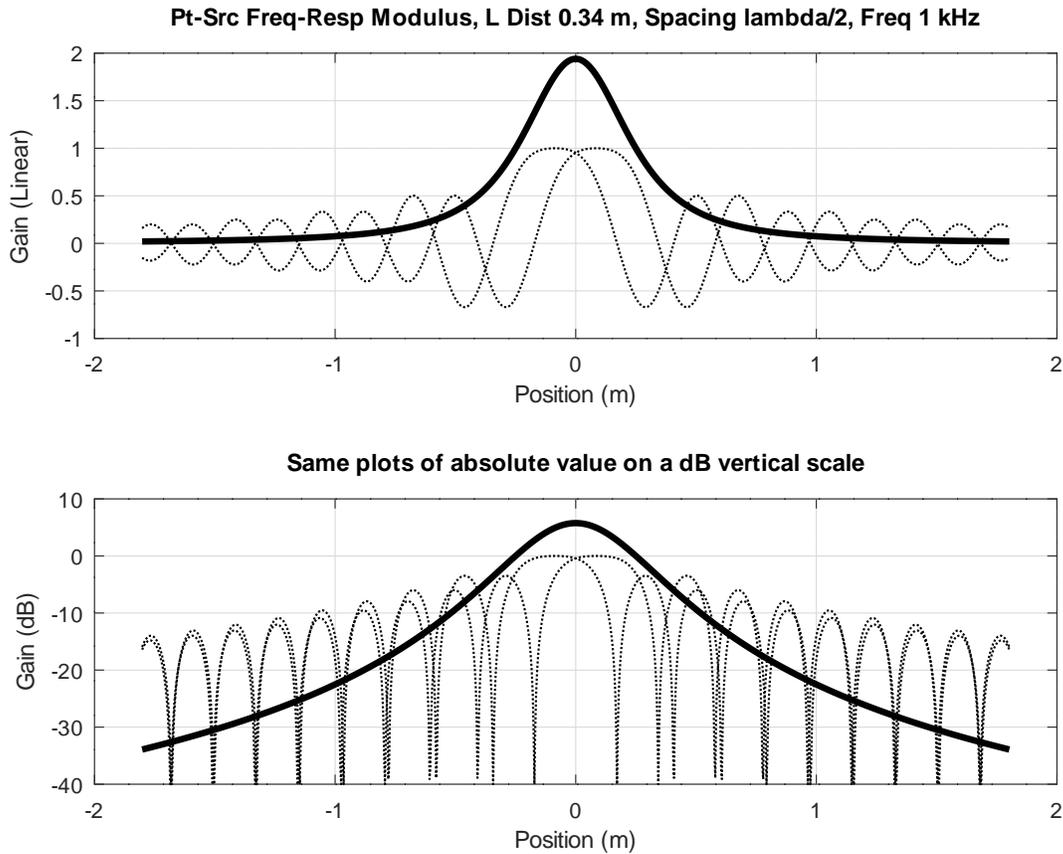
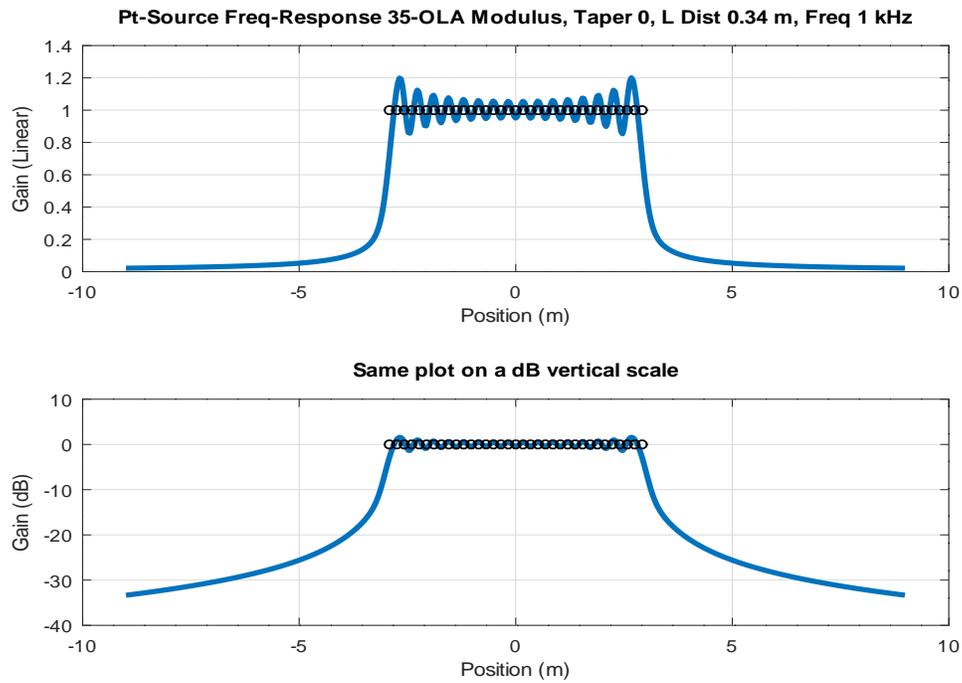
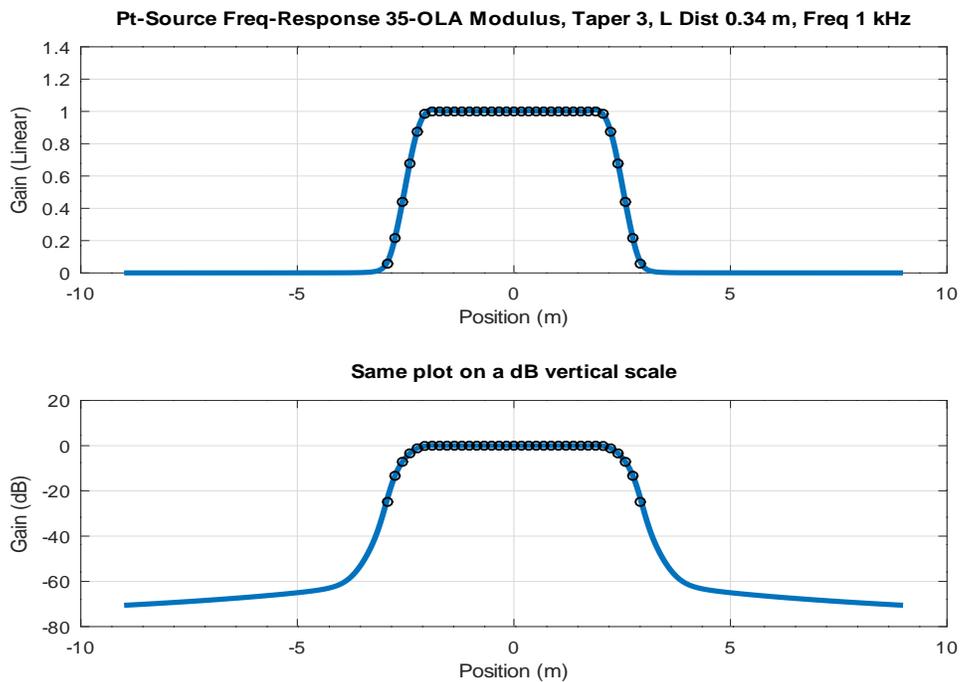


Figure 18: Point-source frequency-response real parts and overlap-add magnitude, for two sources separated by half a wavelength (≈ 17 cm), with the listening-line one wavelength away.

in the direction a wave is coming from, we should see plenty of samples surrounding the point that our “look direction” intersects along the array. The commonly used surrounding ring/sphere architectures (often set up for ambisonics or VBAP) are excellent in this respect: Every look direction has samples uniformly about/around it. Linear/planar arrays are at a disadvantage because they must be truncated or windowed, limiting the virtual stage view.



(a) No array taper, showing much Gibbs oscillation due to array truncation.



(b) Sinusoidally tapered array edges over three wavelengths on each side.

Figure 19: Point-source frequency-response overlap-add magnitude for a 6 m array, with half-wavelength source spacing (≈ 17 cm), observed from a 12 m listening-line one wavelength away.

References

- J. Ahrens. *Analytic Methods of Sound Field Synthesis*. Springer, New York, 2012. ISBN 978-3-642-25743-8. doi: 10.1007/978-3-642-25743-8. URL <https://www.springer.com/gp/book/9783642257421>.
- L. L. Beranek. *Acoustics*. American Institute of Physics, for the Acoustical Society of America, <http://asa.aip.org/publications.html>, 1986. 1st ed. 1954.
- A. J. Berkhout. A holographic approach to acoustic control. *Journal of the Audio Engineering Society*, 36(12):977–995, 1988.
- A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5):2764–2778, 1993. doi: 10.1121/1.405852. URL <https://doi.org/10.1121/1.405852>.
- J. Blauert. *Spatial Hearing. The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA, USA, 1997.
- T. Cooper, Duane H.; Shiga. Discrete-matrix multichannel stereo. *J. Audio Eng. Soc*, 20(5): 346–360, 1972. URL <http://www.aes.org/e-lib/browse.cfm?elib=2070>.
- G. Firtha. *Unified Wave Field Synthesis Framework with Application for Moving Virtual Sources - DRAFT*. Budapest University of Technology and Economics, Hungary, February 7, 2018. Accessed October 19, 2019 from http://last.hit.bme.hu/download/firtha/WFS_system/Firtha_Dissertation_DRAFT.pdf.
- H. Fletcher. Auditory perspective - basic requirements. *Electrical Engineering*, 53(1):9–11, Jan. 1934. doi: 10.1109/EE.1934.6540356.
- A. Franck. Efficient algorithms and structures for fractional delay filtering based on lagrange interpolation. *Journal of the Audio Engineering Society*, 56(12):1036–1056, Dec. 2008. URL <http://www.aes.org/e-lib/browse.cfm?elib=14647>.
- A. Franck. *Efficient Algorithms for Arbitrary Sample Rate Conversion with Application to Wave Field Synthesis*. PhD thesis, Technischen Universität Ilmeneau, Nov. 2011.
- M. A. Gerzon. Surround-sound psychoacoustics. *Wireless World*, 80:483–486, Dec. 1974. available online from The Gerzon Archive at <http://www.audiosignal.co.uk>.
- M. A. Gerzon. Ambisonics in multichannel broadcasting and video. *Journal of the Audio Engineering Society*, 33(11):859–871, 1985.
- J. W. Goodman. *Introduction to Fourier Optics, 3rd Edition*. Roberts & Company, Englewood, CO, 2005. URL https://books.google.com/books?id=ow5xs_Rtt9AC.
- S. Kirkup. *The Boundary Element Method in Acoustics*, volume 8. World Scientific Publishing, 01 2007. ISBN 0953403106. doi: 10.1016/S0218-396X(00)00016-9. URL <http://clok.uclan.ac.uk/7333/1/tbemia07.pdf>.

- D. A. B. Miller. Huygens's wave propagation principle corrected. *Optics Letters*, 16:1370–2, 1991. doi:10.1364/OL.16.001370.
- P. M. Morse and K. U. Ingard. *Theoretical Acoustics*. McGraw-Hill, New York, 1968.
- H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, April 1928. doi: 10.1109/T-AIEE.1928.5055024.
- A. D. Pierce. *Acoustics*. American Institute of Physics, for the Acoustical Society of America, 1989. <http://asa.aip.org/publications.html>.
- V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–, June 1997.
- V. Pulkki. Tech. rep. 62. spatial sound generation and perception by amplitude panning techniques. Technical report, Lab. of Acoustics and Audio Signal Processing, Helsinki University of Technology, 2001. <http://lib.tkk.fi/Diss/2001/isbn9512255324/isbn9512255324.pdf>.
- V. Pulkki, editor. *Parametric Time-Frequency Domain Spatial Audio*. John Wiley and Sons, Inc., New York, 2017. URL <https://onlinelibrary.wiley.com/doi/book/10.1002/9781119252634>.
- J. O. Smith. *Introduction to Digital Filters with Audio Applications*. <https://ccrma.stanford.edu/~jos/filters/>, Sept. 2007a. online book.
- J. O. Smith. *Mathematics of the Discrete Fourier Transform (DFT), with Audio Applications, Second Edition*. <https://ccrma.stanford.edu/~jos/mdft/>, Apr. 2007b. online book.
- J. O. Smith. *Physical Audio Signal Processing*. <https://ccrma.stanford.edu/~jos/pasp/>, Dec. 2010. ISBN 978-0-9745607-2-4. online book.
- J. O. Smith. *Spectral Audio Signal Processing*. <https://ccrma.stanford.edu/~jos/sasp/>, Dec. 2011. ISBN 978-0-9745607-3-1. online book.
- J. O. Smith and P. Gossett. A flexible sampling-rate conversion method. In *Proc. 1984 Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP-84), San Diego*, volume 2, pages 19.4.1–19.4.2, New York, Mar. 1984. IEEE Press. expanded tutorial and associated free software available at the Digital Audio Resampling Home Page: <https://ccrma.stanford.edu/~jos/resample/>.
- J. O. Smith, S. Serafin, J. Abel, and D. Berners. Doppler simulation and the leslie. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx-02), Hamburg, Germany*, pages 13–20, September 26 2002. <https://ccrma.stanford.edu/~jos/doppler/>.
- J. C. Steinberg and W. B. Snow. Auditory perspective - physical factors. *Electrical Engineering*, 13(2):12–17, Jan. 1934. doi: 10.1002/j.1538-7305.1934.tb00661.x.
- V. Välimäki. *Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters*. PhD thesis, Report no. 37, Helsinki University of Technology, Faculty of Electrical Engineering, Laboratory of Acoustic and Audio Signal Processing, Espoo, Finland, Dec. 1995. http://www.acoustics.hut.fi/~vpv/publications/vesa_phd.html.

H. Wierstorf and S. Spors. Sound Field Synthesis Toolbox. In *132nd Convention of the Audio Engineering Society*, 2012.