

SSSP- LAVAGNATE

MUSIC ENGINEERING,
POLITECNICO DI MILANO



ANGELO ANTONA

Audio equalizers

Audio equalizers $\rightarrow$ audio equalizer $\rightarrow$ equal means to filter it things $\rightarrow$ beam filter $\rightarrow$ also the multiply also for response then on top of this eq

SHELVING FILTERS

Then we have category of shelving filter $\rightarrow$ HF shelving filter $\rightarrow$ multiply by shelf. They multiply the freq response sound $\rightarrow$ extend high freq

PEAK FILTERS

C-CONTINUOUS TRANS FUNCTION

LF shelving

$$H_L(s) = \frac{1 + \frac{s}{\omega_0}}{1 + \frac{s}{Q\omega_0}}$$

HF shelving

$$H_H(s) = \frac{1 + \frac{s}{Q\omega_0}}{1 + \frac{s}{\omega_0}}$$

when $\omega_0 = 2\pi f_0$
 $\omega = j\omega$

T-CONTINUOUS TRANS FUNCTION

$$H_L(s) = \frac{s}{1 + \frac{s}{Q\omega_0}}$$

when $\omega_0 = 2\pi f_0$
 $\omega = j\omega$

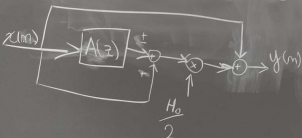
Z-DOM TRANS FUNC

We use a first order all pass filter:

$$H_L(z) = 1 + \frac{H_0}{2} \left(1 \pm A(z) \right) \text{ where } A(z) = \frac{z^{-1} + 1}{1 + z^{-1}}$$

$\rightarrow$ LF
 $\rightarrow$ HF

In the previous Z-domain filter, we catch the pattern of us



Z-DOM TRANS FUNC

We use a second order all pass filter:

$$H_L(z) = 1 + \frac{H_0}{2} \left(1 \pm A(z) \right) \text{ where } A(z) = \frac{-z^{-1} + 1 + \alpha z^{-2}}{1 + \alpha z^{-1} + z^{-2}}$$

In the previous Z-domain filter, we catch the pattern of us. $\rightarrow$ catch the pattern of us. $\rightarrow$ catch the BW of the peaking filter

Integer and fractional delay

INTEGER AND FRACTIONAL DELAY

We mean that a simple continuous time delay filter is given by $y(t) = x(t - \tau)$

When we go on discrete time domain, the equation becomes: $y(n) = x(n - D) \rightarrow$ there are two cases: $\rightarrow$ integer $\rightarrow$ integer delay $\rightarrow$ $D = \lfloor D \rfloor + \delta$ $\rightarrow$ $D = \lfloor D \rfloor + \delta$ $\rightarrow$ the complete integer time delay is not an integer multiple of T_s

If we are dealing with a fractional delay, we must perform the following steps:

① We interpolate the signal, extending the continuous time axis

$$\hat{x}(t) = \sum_{m=-\infty}^{\infty} x(nT_s) h_c(t - nT_s)$$

some domain filtering = convolution

② We apply the delay $D T_s$ to the continuous time signal, and then we re-sample it

We can see the delay filter as $y(n) = x(n - D) \rightarrow Y(z) = X(z) z^{-D} \rightarrow Y(f) = X(f) e^{-j2\pi f D} \rightarrow H(f) = \frac{Y(f)}{X(f)} = e^{-j2\pi f D}$

If we apply the IFT to the impulse response of the delay filter, we get: $h(t) = \int_{-\infty}^{\infty} e^{-j2\pi f D} e^{j2\pi f t} df = \delta(t - D)$

There are two problems: $\rightarrow$ integer delay $\rightarrow$ the sampled $\rightarrow$ c is 0 pulse $\rightarrow$ FTR $\rightarrow$ fractional delay $\rightarrow$ the sampled $\rightarrow$ c is not a pulse $\rightarrow$ IIR



• **Analog equalization** → correct in filtering a signal using analog filters

• The basic categories of analog equalization filters are:

Shelving filter
They are based on HPF and LPF

$$H(s) = \frac{1 + \frac{s}{\omega_c}}{1 + \frac{s}{\omega_c}}$$

We can implement a shelving filter using a 1st order all-pass FIR filter

$$H(z) = 1 + \frac{1}{2} \left(\frac{1+z}{1+z^{-1}} \right) \rightarrow \lambda(z) = \frac{z+z^{-1}}{1+z^{-1}}$$

Peaking filter
They are based on BPF

$$H(s) = \frac{1 + \frac{s}{\omega_c}}{1 + \frac{s}{\omega_c}}$$

We can implement a peaking filter using a second order all-pass FIR filter

$$H(z) = 1 + \frac{1}{2} \left(\frac{1+z}{1+z^{-1}} \right)^2 \rightarrow \lambda(z) = \frac{z^2 + 2z + 1}{1+z^{-1}}$$

If we perform the inverse fourier transform of the transfer function we get a sinc function: $\lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega n} d\omega = \delta[n]$

• **Approximate a delay filter with a FIR filter** to approximate a Delay filter with a FIR filter of the type $H(z) = \sum_{k=0}^{N-1} h_k z^{-k}$

$$\|E\|^2 = \|H_d(\omega) - H(\omega)\|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_d(\omega) - H(\omega)|^2 d\omega \rightarrow \sum_{n=0}^{N-1} |h_n - h_d(n)|^2 \rightarrow \sum_{n=0}^{N-1} (h_n - h_d(n))^2 = 0 \rightarrow \sum_{n=0}^{N-1} h_n = \sum_{n=0}^{N-1} h_d(n)$$

Good measure: they are unwanted ripples of ± 1 due to the FIR filter approximation. We have three ways to reduce the error:

① Less pass the error → it reduces ripple but we lose for details

② Less pass the error of the error (Good least squares approximation): instead of less pass the error we reduce the error by calculating the error of the error: $\|E\|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_d(\omega) - H(\omega)|^2 d\omega$

③ Lagrange interpolation → we want to make the error zero at N points, we set the first N samples of the error = 0:

$$\frac{dE}{d\omega} = 0 \text{ with } l=1, N \Rightarrow E(\omega) = H_d(\omega) - H(\omega)$$

$$H_d(\omega) = e^{-j\omega D} \quad H(\omega) = \sum_{k=0}^{N-1} h_k e^{-j\omega k}$$

$$\frac{dE}{d\omega} = (-jD) e^{-j\omega D} - \sum_{k=0}^{N-1} h_k (-j k) e^{-j\omega k} = 0 \Rightarrow D e^{-j\omega D} = \sum_{k=0}^{N-1} h_k e^{-j\omega k}$$

$$\frac{dE}{d\omega} = (-jD) e^{-j\omega D} - \sum_{k=0}^{N-1} h_k (-j k) e^{-j\omega k} = 0 \Rightarrow D e^{-j\omega D} = \sum_{k=0}^{N-1} h_k e^{-j\omega k} \Rightarrow D = \sum_{k=0}^{N-1} h_k k$$

The solution to the equation can be given by:

$$h_k = \frac{1}{N} \frac{D!}{k! (N-k)!}$$

This is a binomial expansion with N terms. We can solve it using the binomial theorem:

$$h_k = \frac{1}{N} \frac{D!}{k! (N-k)!} \Rightarrow \sum_{k=0}^N h_k = 1$$

• **Approximate delay filter with a IIR filter**: we can also implement the delay filter using an all-pass IIR filter:

$$H(z) = \frac{z^{-N} A(z^{-1})}{A(z)} \text{ where } A(z) = a_0 + a_1 z^{-1} + \dots + a_N z^{-N}$$

In an ideal delay filter, $|H(e^{j\omega})| = 1$ and $\angle H(e^{j\omega}) = -D\omega$ → a solution is to use the Thiran all-pass design: $H(z) = \frac{1}{(1+z^{-1})^N}$ where $a_k = \frac{1}{2^N} \binom{N}{k}$

Feed forward delay → mix between input signal and delayed input signal: $y(n) = x(n) + g x(n-D) \Rightarrow Y(z) = X(z) + g X(z) z^{-D}$

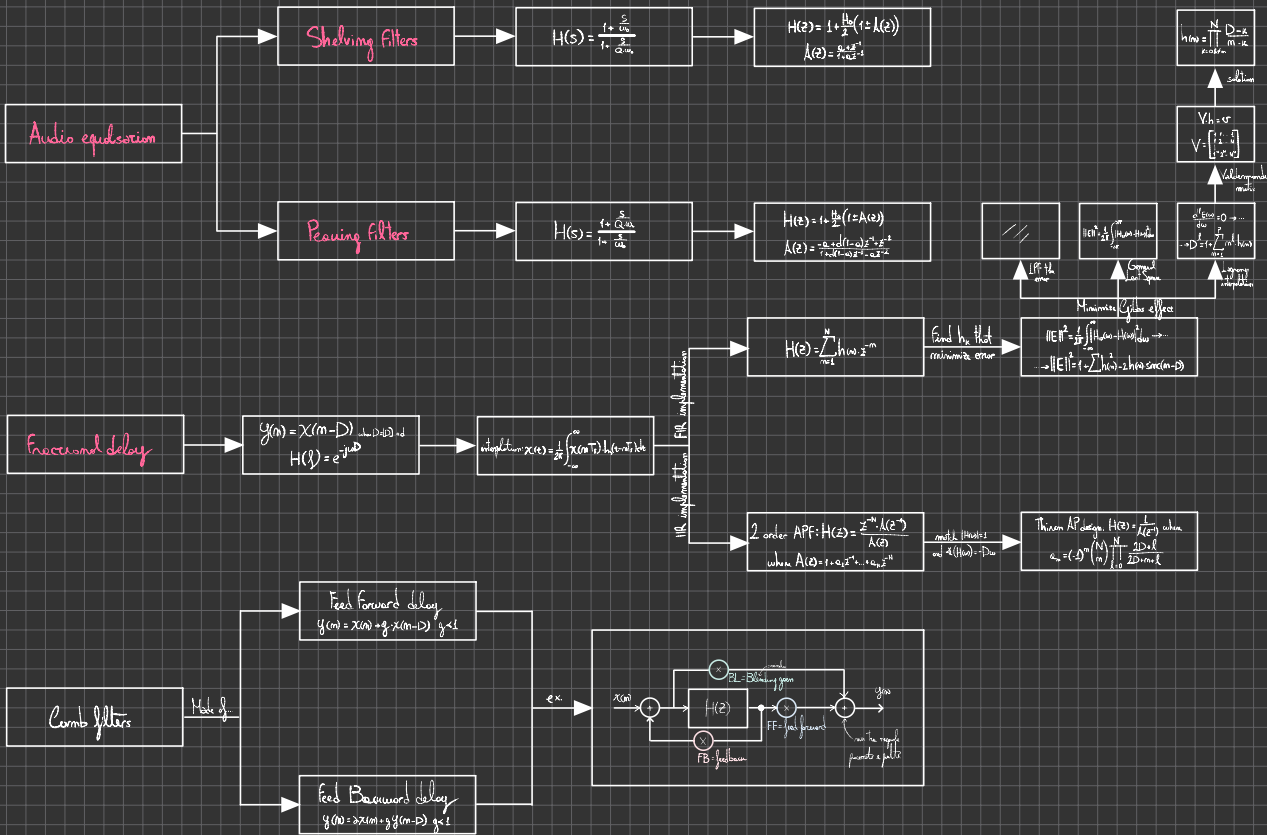
$$H(z) = 1 + g z^{-D}$$

• **Feed back delay** → mix between input signal and delayed output signal: $y(n) = x(n) + g y(n-D) \Rightarrow Y(z) = X(z) + g Y(z) z^{-D}$

$$H(z) = \frac{1}{1 - g z^{-D}}$$

• **Universal comb filter** → combination of feed-forward and feed-back delay:





Fourier representation of different types of signals

	Periodic Signal	Finite Duration	Non-periodic Signal
Data	$X(t) = \sum_{n=-\infty}^{\infty} x_n e^{j n \omega_0 t}$	$X(t) = \int_{-\infty}^{\infty} x(\omega) e^{j \omega t} d\omega$	$X(t) = \int_{-\infty}^{\infty} x(\omega) e^{j \omega t} d\omega$
Inverse	$x_n = \frac{1}{T} \int_0^T X(t) e^{-j n \omega_0 t} dt$	$x(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(t) e^{-j \omega t} dt$	$x(\omega) = \left(\frac{1}{2\pi} \right) \int_{-\infty}^{\infty} X(t) e^{-j \omega t} dt$



Acoustic field Homogeneous Equations

An acoustic field is a function $p(x, y, z, t)$ that varies on time and space. If we have a homogeneous equation, we can find the homogeneous solution.

- Time domain homogeneous equation: $\nabla^2(p(x, y, z, t)) = \frac{1}{c^2} \frac{\partial^2 p(x, y, z, t)}{\partial t^2}$
- Frequency domain homogeneous equation: $\nabla^2(p(x, y, z, \omega)) = -\frac{\omega^2}{c^2} p(x, y, z, \omega)$ — Helmholtz homogeneous equation

The homogeneous solution of the field is the free field, $p(x, y, z, t) = P(x, y, z) e^{j \omega t}$

Solutions of Homogeneous equations

The acoustic field is a homogeneous field, which means that it is a solution to the homogeneous equation. We can find the homogeneous solution by using the separation of variables method.

Separation of Variables

Let $p(x, y, z, t) = X(x)Y(y)Z(z)T(t)$. Substituting into the homogeneous equation, we get:

$$\frac{1}{X} \frac{d^2 X}{dx^2} + \frac{1}{Y} \frac{d^2 Y}{dy^2} + \frac{1}{Z} \frac{d^2 Z}{dz^2} + \frac{1}{T} \frac{d^2 T}{dt^2} = -\frac{\omega^2}{c^2}$$

Each term is a function of a single variable, so each term must be equal to a constant. Let $k_x^2, k_y^2, k_z^2, k_t^2$ be the separation constants. Then:

$$\frac{1}{X} \frac{d^2 X}{dx^2} = -k_x^2, \quad \frac{1}{Y} \frac{d^2 Y}{dy^2} = -k_y^2, \quad \frac{1}{Z} \frac{d^2 Z}{dz^2} = -k_z^2, \quad \frac{1}{T} \frac{d^2 T}{dt^2} = -k_t^2$$

The general solution is:

$$p(x, y, z, t) = \sum_{k_x, k_y, k_z, k_t} A_{k_x, k_y, k_z, k_t} X_{k_x}(x) Y_{k_y}(y) Z_{k_z}(z) T_{k_t}(t)$$

Separation of Variables

The general solution is the sum of the homogeneous solutions. For example, for a rectangular domain, the solution is:

$$p(x, y, z, t) = \sum_{k_x, k_y, k_z, k_t} A_{k_x, k_y, k_z, k_t} \cos(k_x x) \cos(k_y y) \cos(k_z z) \cos(k_t t)$$

Solutions of Inhomogeneous equations

Acoustic field Inhomogeneous Equations

INHOMOGENEOUS ACOUSTIC WAVE EQUATION

If we consider a source of sound, the inhomogeneous equation is:

$$\nabla^2(p(x, y, z, t)) = \frac{1}{c^2} \frac{\partial^2 p(x, y, z, t)}{\partial t^2} + S(x, y, z, t)$$

The general solution is the sum of the homogeneous solution and the particular solution.

3D SOLUTION

The solution of the inhomogeneous equation is the sum of the homogeneous solution and the particular solution. The particular solution is found by using the method of Green's functions.

2D SOLUTION

The solution of the inhomogeneous equation is the sum of the homogeneous solution and the particular solution. The particular solution is found by using the method of Green's functions.

• Fourier representations for continuous signals:

Periodic in time $\longrightarrow$ Discrete in frequency $X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$ $x(t) = \sum_{-\infty}^{\infty} X(\omega) e^{j\omega t}$	Non periodic in time $\longrightarrow$ Continuous in frequency $X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$ $x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega$	Multidim non periodic in time $\longrightarrow$ Multidim non periodic in freq $X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$ $x(t) = \left(\frac{1}{2\pi}\right)^N \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega$
--	---	---

• We can describe a field on a medium using the following equations

Medium with uniform density: homogeneous wave equation

- Trans domain: $\nabla^2 P(r, \omega) = \frac{\partial^2 Q(r, \omega)}{\partial t^2}$
 - Freq domain: $\nabla^2 P(r, \omega) = -(\omega)^2 P(r, \omega) = \text{Helmholtz eq.}$
- $\longrightarrow$ These equations are independent from the coordinate system, so we can solve them in different coordinate systems.

Cartesian coordinates

Spherical domain

The solution to the homogeneous Helmholtz eq. in cartesian domain is:

$$P(r, \omega) = e^{j(k_x x + k_y y + k_z z)}$$

If we consider plane waves, we have that k is $\perp$ to the plane. In that case, we can say that $k_x^2 + k_y^2 + k_z^2 = (\omega/c)^2$. Given this, there are three possibilities:

- ① $(\omega/c)^2 > k_x^2 + k_y^2 \rightarrow k_z \in \mathbb{R}$
- ② $(\omega/c)^2 < k_x^2 + k_y^2 \rightarrow k_z \in \mathbb{I}$

The solution to the Helmholtz homogeneous equation in spherical domain is expressed in the form $P(r, \omega) = \frac{Q(r, \omega)}{r} \Phi(\theta) \Theta(\phi)$, where Φ, Θ, Q are functions of r, θ, ϕ respectively.

$$\begin{aligned} \bullet \Phi(\theta) &= e^{j m \theta} & \bullet \Theta(\phi) &= P_l^m(\cos(\phi)) \\ \bullet R(r) &= R_1 Y_l(\frac{\omega}{c} r) + R_2 Y_l(\frac{\omega}{c} r) \\ &= R_3 h_l^{(1)}(\frac{\omega}{c} r) + R_4 h_l^{(1)}(\frac{\omega}{c} r) \end{aligned}$$

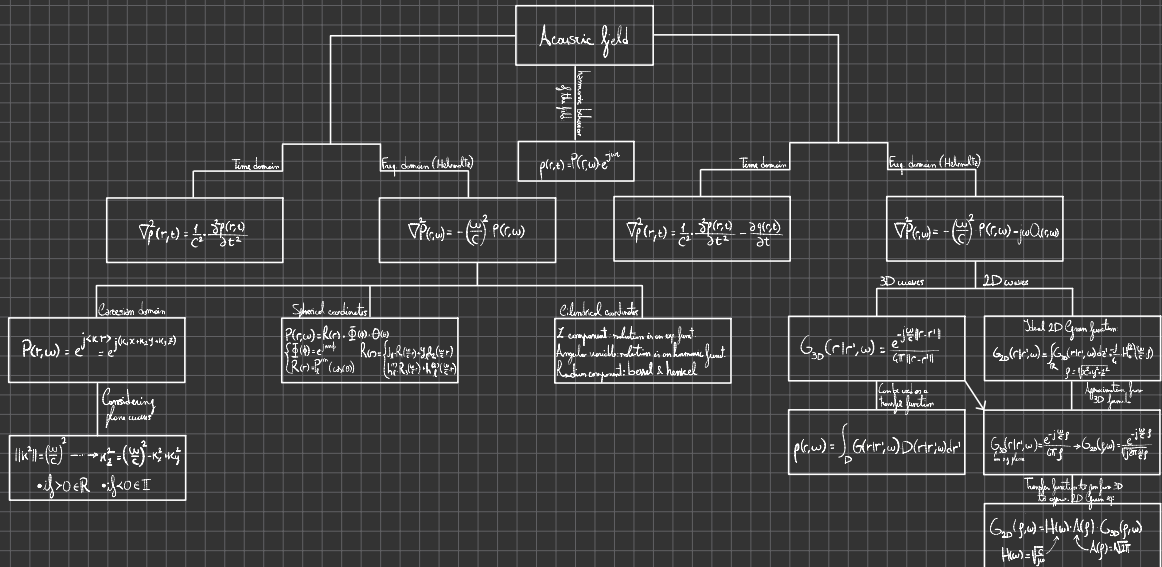
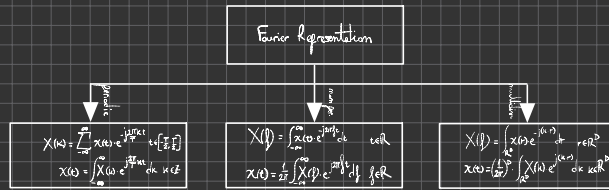
Medium with non-uniform density: inhomogeneous wave equation

- Trans domain: $\nabla^2 P(r, \omega) = \frac{\partial^2 Q(r, \omega)}{\partial t^2}$
 - Freq domain: $\nabla^2 P(r, \omega) = -(\omega)^2 P(r, \omega) - j\omega Q(r, \omega)$
- $\longrightarrow$ The solution of this inhomogeneous field equation has a particular form: **Green function**: $G_{\text{sp}}(r(r'), \omega) = \frac{e^{j\omega(r-r')}}{4\pi|r-r'|}$

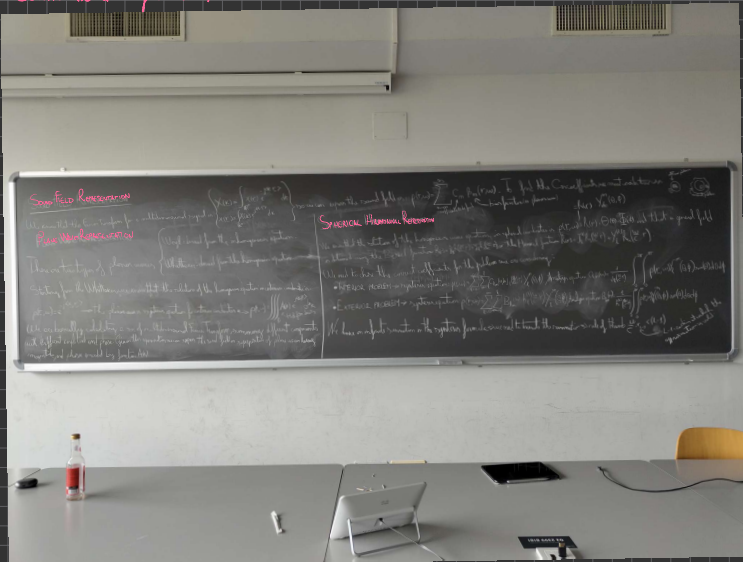
Considering an infinite line source, one idea could be to derive the 2D Green function integrating the 3D one along the source direction:

$$G_{2D} = \int_{-\infty}^{\infty} G_{3D} dz \longrightarrow \text{the problem here is that we cannot build an infinite long line source, so we must derive the 2D Green function approximating the 2D one: } G_{2D}(r(r'), \omega) = \frac{e^{j\omega(r-r')}}{4\pi|r-r'|} \xrightarrow{\text{one-dim plane}} G_{2D}(r(r'), \omega) = \frac{e^{j\omega(r-r')}}{2\pi\sqrt{r^2 - z^2}} \xrightarrow{\text{long time operation (imag. part)}} G_{2D}(r(r'), \omega) = \frac{e^{j\omega(r-r')}}{\sqrt{r^2 - z^2}}$$

It's also possible to get the transfer function to go from the G_{2D} to the G_{3D} : $G_{3D}(r(r'), \omega) = H(\omega) G_{2D}(r(r'), \omega)$



Sound Field Representation



- **Sound field representation:** sound field is representable as $p(r, \omega) = \sum_{n=0}^{\infty} C_n B_n(r, \omega)$. We need to use the correct coefficients for every situation. There are two main cases: coefficients to be found.

INTERIOR PROBLEM



the sources located inside a closed region, we are analyzing the effect of the field in an external region Ω_e

EXTERIOR PROBLEM



the source is located outside of an area, which corresponds to the region of space Ω_i that we are analyzing

There are two types of **plane wave representations**

Wave based on non homogeneous wave equations

Whitaker based on homogeneous wave equations $\rightarrow$ the solution in cartesian plane is $p(r, \omega) = e^{i \omega t}$

Taking in consideration the Whitaker equations we can use it to represent the sound field as the summation of infinite plane waves having different amplitudes and phases:

$$p(r, \omega) = \iiint_{\mathbb{R}^3} A(\mathbf{k}) e^{i \mathbf{k} \cdot \mathbf{r}} d\mathbf{k}$$

$A(\mathbf{k})$ encodes amplitude and phase of each plane wave

$$\Phi(\mathbf{r}) = e^{i \mathbf{k} \cdot \mathbf{r}}$$

hyperbolic polynomial of order k and l

$$\Theta(\mathbf{r}) = P_l^m(\cos \theta)$$

$$= R(r) Y_l^m(\theta, \phi)$$

- We know that the **spherical harmonic representation** of a field has a solution of the type $p(r, \omega) = R(r) \Phi(\theta) \Theta(\phi)$ and that a general spherical wave can be expressed as $R(r) Y_l^m(\theta, \phi)$

Similarly to what we do with the plane wave representation, we can represent the sound field as the superposition of an infinite number of spherical waves. This is done differently for interior and exterior problem:

INTERIOR PROBLEM SPHERICAL REPRESENTATION

• Synthesis equation: $p(r, \omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(\omega) j_l(kr) Y_l^m(\theta, \phi)$

• Analysis equation: $C_{lm}(\omega) = \frac{1}{j_l(kr)} \int_0^{\pi} \int_0^{2\pi} p(r, \omega) Y_l^{*m}(\theta, \phi) \sin \theta d\theta d\phi$

EXTERIOR PROBLEM SPHERICAL REPRESENTATION

• Synthesis equation: $p(r, \omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^l B_{lm}(\omega) h_l^{(2)}(kr) Y_l^m(\theta, \phi)$

• Analysis equation: $B_{lm}(\omega) = \frac{1}{h_l^{(2)}(kr)} \int_0^{\pi} \int_0^{2\pi} p(r, \omega) Y_l^{*m}(\theta, \phi) \sin \theta d\theta d\phi$

● Sound field Representation (Later schematization of concepts)

- Sound fields can be represented through two main eqs:
 - Wave equation: derived from inhomogeneous field equations.
 - Helmholtz representation: derived from homogeneous equations.

We analyse only this Helmholtz representation. In the type of representation we want to express the field as result of the superposition of elementary waves. Let's see how we can do that using planar and spherical waves.

Planar waves

In this case we consider the field as formed by infinite planar waves, each one having its amplitude and phase. Mathematically this is expressed as:

$$P(x, y) = \iint_{\mathbb{R}^2} A(\omega) e^{i\omega \cdot x} d\omega$$

↗ partially multidimensional Fourier transform
↘ A(ω) under this amplitude and the phase for each plane wave

Spherical waves

This method involves representing the wave field as the superposition of its spherical components. We know that the solution of the field equation in spherical coordinates is:

$$P(r, \theta) = R(r) \Phi(\theta) \Theta(\phi)$$

↗ radial physical
↘ azimuthal physical

$$R(r) = R_1(r) \left(\frac{e^{-ir}}{r} \right) + R_2(r) \left(\frac{e^{ir}}{r} \right)$$

↗ inward propagating
↘ outward propagating

$$= R_1(r) \sqrt{\frac{2}{\pi}} \left(\frac{e^{-ir}}{r} \right) + R_2(r) \sqrt{\frac{2}{\pi}} \left(\frac{e^{ir}}{r} \right)$$

Knowing that we can express the sound field as the superposition of its harmonic spherical components, but we need to use this correct formulation based on the case we are dealing with:

INTERIOR PROBLEM CASE EXTERIOR PROBLEM CASE

- This is the case in which there is a source inside a volume and we want to express the sound field inside a closed area Ω .



- The spherical waves summation for an exterior problem case is:

we need these coefficients

$$P(r, \omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm}(\omega) j_l(kr) Y_l^m(\theta, \phi)$$

$$C_{lm}(\omega) = \frac{1}{4\pi} \iint_{\Omega} P(r, \omega) \cdot Y_l^m(\theta, \phi) \cdot \sin\theta d\theta d\phi$$

- This is the case of a sound wave/plane outside a closed control area Ω , in which we want to express the sound field.



- In that case, the spherical harmonic summation is:

we need these coefficients

$$P(r, \omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^l B_{lm}(\omega) h_l^{(2)}(kr) Y_l^m(\theta, \phi)$$

$$B_{lm}(\omega) = \frac{1}{4\pi} \iint_{\Omega} P(r, \omega) \cdot Y_l^m(\theta, \phi) \cdot \sin\theta d\theta d\phi$$

- **HRTF (Head Related Transfer Function)**: HRTF are a pair of transfer functions (one per ear) that describe how sound as filtered by listener's ear head and body before reaching the ear canal. It is defined as $HRTF = \frac{\text{sound pressure at the ear}}{\text{sound pressure reference}}$. The HRTF varies among different individuals, but there are some common characteristics: filter approximately 15kHz, a drop in amplitude is observed due to the effect of the head which creates a shadowing effect.
 Current state: there is a ramp increase up due to the fact that the form of the binaural tends to significantly influence the sound, adding direction-dependent filtering effect.
- **Different methods to measure HRTF** are:
 - Pseudo-random noise: a broad-spectrum noise used to analyze the frequency response.
 - Gating sequences: a pair of complementary sequences used to obtain the impulse response efficiently.
 - Sweep: frequency sweep (chirp) that cover the entire audible spectrum, providing detailed frequency response data.

● **HRTF (Head Related Impulse Response)**: it is derived from the HRTF and it represents the time domain impulse response of the listener's anatomy.

- **Distance perception**:
 - in free field we cannot determine distance perception because we have not acoustic cues (Note: a free field is an environment without any reflection, mirror or an acoustic chamber).
 - in a reverberant environment we can achieve distance perception through the following principles:
 - **Energy ratio comparison**: when a sound is closer, the direct sound is more dominant compared to its reverberations. As the distance increases, the proportion of reverberant sound energy increases relative to the direct sound.
 - **Uncorrelated ear signals**: reverberation uncorrelates the signals from the two ears. In fact, when sound waves bounce off various surfaces they arrive at the ears at slightly different times and intensities, creating a decorrelation that can be used by our auditory system to infer distance.
 - **Spectral coloration**: the reverberant component of the signal that occurs to our ears has a different spectral coloration compared to the main wave. Normally, higher frequencies are absorbed more than lower ones. This is an additional cue to estimate the distance of the sound source.

● **Reflections**: in a reverberant environment we have two different types of reflections:

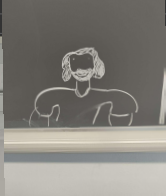
- **Early reflections** → on the initial reflections that occur shortly after the direct sound (50ms). The **precedence effect** is a phenomenon where the localization of a sound source is dominated by the direct sound instead of the early reflections.
- **Late reflections** → reflections that occur after the early reflections.

 We can also define a **precedence region**, which is the time in which reflections are not perceived as different signals to instead of the direct wave. Beyond this precedence region, the brain starts to perceive distinct echoes.
 The precedence effect allows to maintain clear and stable sound localization in reverberant environments by prioritizing the direct sound over reflections.

- **Primary auditory cues**: one ear used to handle the sound direction
 - **ILD (Interaural Level Difference)**: related for $f > 7000\text{Hz}$. It is the difference in the sound level reaching each ear due to the shadow effect. The ILD varies with the position of the head (and θ).
 - **ITD (Interaural Time Difference)**: related for $f < 1500\text{Hz}$. It is the difference in time of arrival of the sound between the two ears.

Using these two effects together (duplex theory) we can determine the direction of sound sources in the horizontal plane. However, this method does not distinguish a sound that comes from angle 0 and one coming from angle 180°.

Oscillators



Oscillators are used to produce a periodic waveform.

SINE WAVE

- Ideal sine wave is a smooth, continuous waveform.
- Sine wave is a periodic waveform.
- Sine wave is a periodic waveform.
- Sine wave is a periodic waveform.

GRANULAR SYNTHESIS

SALTOOTHY WAVES

SQUARE WAVES

TRIANGLE WAVES

SIGNAL BASED SYNTHESIS

- In signal based synthesis, we produce sounds directly, focusing on its characteristics.
- **Oscillators**: an oscillator is a circuit to produce a periodic waveform. We will study 4 types of oscillators: sine, triangle, square, and sawtooth.

Sine wave

It is the simplest type of oscillator. Its mathematical representation is: $\cos(\theta - \phi) = 2\cos(\theta)\cos(\phi) - \cos(\theta + \phi)$. From this equation, using trigonometric identities, we can get the following system representation: $\begin{bmatrix} \cos(\theta - \phi) \\ \cos(\theta + \phi) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix}$. We can get the next state of the system: $y = R \cdot x$. In order to ensure stability, $\det(R) = 1$. We can implement efficiently the rotational matrix using the Givens matrix: $\begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$.

Sometimes, the output of the system can diverge due to the approximation of the coefficients. To manage this problem, we can use the ASC (Automatic gain control), which consists of:

- 1) Calculate the scaling factor (if response grows).
- 2) Scale the system parameters by the scaling factor.

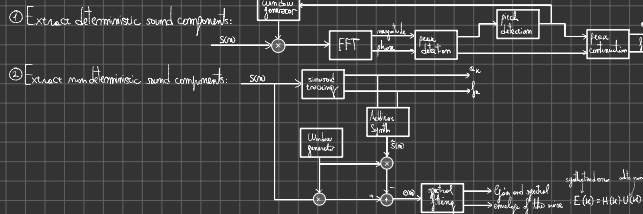
Another way (more efficient from a point of view of performance and memory) is to use the wavetable oscillator. It works as follows:

- 1) We precompute and store in the wavetable a single period of the wave we want to generate.
- 2) When we want to generate the waveform, we read the single period of the desired function. We can also change the pitch of the sound, reading the samples of the period at interval Δ . $\Delta < 1$ if we want a higher frequency, $0 < \Delta < 1$ if we want a lower frequency. The step Δ can be determined using the formula: $\Delta = \frac{f_s}{f}$, where f_s is the sample frequency.

Additive synthesis: in additive synthesis we build the waveform summing its individual components: $x(n) = \sum_{k=0}^{m-1} a_k \cos(k\omega)$. We have two main ways to implement this theoretical concept.

- **Waveform concatenation:** in this method we generate a single period of freq f_0 and then we concatenate it. When we copy this one period in time domain we are also concatenating a "sine" for a pulse of f_0 in the freq domain, actually adding frequency components.

- **Synthesis by endless computation:** in this technique we extract some characteristics of the sound (analysis) in order to be able to re-synthesize it.



• Sine wave oscillators

A sine wave oscillator produces a waveform that ramps upward and then sharply drops. Its mathematical representation is:

$$\sin(\omega) = 2 \cdot \left[\frac{m}{\pi} \right]_{\text{mod } 1} - 1$$

In a real-time system the roll-off factor in -6dB/octave is the closing problem (frequency components that roll-back in audible frequency range) critical to manage this problem we can use one of these techniques:

- ① Band-limited algorithm: significant only components below Nyquist.
- ② Over-band: limits algorithm has fast frequency domain.
- ③ Spectral tilt modifications: reshape the closing into non audible range.

We can generate a real-time system with limited closing using the Differential Feedback Wave method (DFW):

$$p(n) = \int_{-\pi}^{\pi} x \, dx \quad \sin(\omega) = \frac{d}{dt} \cos(\omega)$$

The roll-off factor for a real-time system generated with the DFW technique is -12dB/sec (Based on spectral tilt modifications).

Additive synthesis formulation: $S(m) = -\sum_{k=1}^{K-1} \frac{1}{k} \sin(2\pi k \frac{m}{n})$

• Square wave oscillators

A square wave oscillator produces a waveform that alternates from a high and a low states. Its mathematical representation is:

$$\sin(\omega) = 2 \cdot \text{sgn}(\sin(2\pi \frac{m}{n})) - 1$$

In a square wave the roll-off factor in -6dB/octave, so the closing problem is critical. The way to limit this problem is the use of the real-time wave.

• Triangle wave oscillators

A triangle wave oscillator produces a waveform that linearly goes up and down. Its mathematical representation is:

$$\sin(\omega) = \frac{2}{\pi} \cdot \arcsin(\sin(2\pi \frac{m}{n})) - 1$$

For triangle waves the roll-off factor is of -12dB/sec, so the closing problem is less prominent.

- **Granular synthesis:** it is a sound synthesis technique that involves manipulating short pieces of sound, called grains. The mathematical representation of this process is: $x(t) = \sum_{n=1}^N \sum_{m=1}^M g_m(t) \cdot \sin(2\pi f_m t) \cdot e^{j2\pi f_m t}$. There are two main types of granular synthesis: ① Synchronous Granular Synthesis: grains are distributed with a constant density over time (generator: harmonic relations). ② Asynchronous Granular Synthesis: the density of grains distribution varies over time (generator: more natural behavior).

● **Source filter models:** in source filter models, the generation of sound is modelled as follows:



Shaping filter = resonant order LP filter $\rightarrow H(f)$
 $\left(\frac{1}{(1 + j\omega/\omega_0)^2} \right) \left(\frac{1}{(1 + j\omega/\omega_1)^2} \right)$
 in actual form of the filter
 2nd order resonance
 2nd order resonance

● **Non-linear models:** until now we've seen linear Time Invariant models, that do not add new freq component to the sound. To achieve this even we must use non-linear models which introduce harmonic distortion. This quantity is measured through the total harmonic distortion factor $\rightarrow$ THD = $\sqrt{\frac{\sum_{k=2}^N A_k^2}{A_1^2}}$
 2nd order of the non-linear
 2nd order of the non-linear
 We can implement non linear models in three main ways:

• **Wave shaping**

Are memoryless models where the output depends directly from the input. They can be expressed using a truncated Taylor expansion:

$$y(m) = F(x(m)) = \sum_{k=0}^N a_k x^k(m)$$

Wave shaping can introduce both even and odd harmonics.

- Symmetric distortion $\rightarrow$ odd harmonics
- Asymmetric distortion $\rightarrow$ both even and odd harmonics

• **Multiplicative synthesis (ring modulation)**

This technique involves multiply two signals together. Typically we have $c(m)$ = carrier signal
 $m(m)$ = modulating signal.
 The resulting signal is $y(m) = c(m) \cdot m(m)$

• **Frequency modulation (FM)**

The technique involves modulating the frequency of a signal with another signal, often a sinusoidal carrier. For example: $S(m) = A \sin(\omega_c m + \phi(m))$
 modulating signal

NON LINEAR MODELS

• Another way to shape the sound of an oscillator is to use non-linear models $\Rightarrow$ introduce harmonic distortion, measured through

Total Harmonic Distortion: $THD = \sqrt{\frac{\sum_{k=2}^N A_k^2}{A_1^2}}$

• There are three main approaches to implement non linear models in oscillators:

• **Wave shaping / memoryless model** (the output depends only on the current input). They can be expressed through a truncated Taylor expansion:

$$y(m) = F(x(m)) = \sum_{k=1}^N a_k x^k(m)$$

- Symmetric distortion introduces odd harmonics
- Asymmetric distortion introduces both odd and even harmonics

• **Multiplicative synthesis:** multiply two signals together. Typically the signals are: $f_c(m)$ = carrier signal
 $f_m(m)$ = modulating signal.

The resulting signal is $y(m) = c(m) \cdot m(m)$

• **Frequency Modulation Synthesis:** modulate the frequency of the carrier signal with a modulating signal:

$$S(m) = A \cdot \cos(\omega_c m + \phi(m))$$

FM

AM - MODULATOR

modulating signal

MODEL BASED SYNTHESIS

• In model based synthesis we change the sound "invariant": we focus on how to model the instrument that produces the same instead of directly model the sound itself.
We can use three main approaches to model this process: the instrument that produces the sound itself.

• **Lumped models:** treat the system as a collection of discrete components, each one modelled by ORDINARY DIFFERENTIAL EQUATIONS.
Each component represent a physical property such as mass, spring or damper.
This approach is indicated for model systems where the interactions are concentrated at specific points, like circuits or mechanical linkages.

Note: DWF belong to this category. In fact they model lumped model converting their continuous relations into scattering relations. They will be explained in the next chapter.

• **Distributed models:** consider the system as a continuous entity. The spatial and temporal variations in the system are described by PARTIAL DIFFERENTIAL EQUATIONS.
(ex. wave equation for a vibrating string) They are described by Partial Diff. Eq:

Partial Differential Equations

- 1) We start from the D'Alembert equation
$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 y}{\partial t^2} \rightarrow \frac{\partial^2 y}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 y}{\partial t^2} = 0$$
- 2) We decompose the square of difference into a product:
$$\left(\frac{\partial y}{\partial x} - \frac{1}{c} \frac{\partial y}{\partial t} \right) \left(\frac{\partial y}{\partial x} + \frac{1}{c} \frac{\partial y}{\partial t} \right) = 0$$

← we need to characterize them
- 3) The general solution is: $y(x,t) = y^+(c t - x) + y^-(c t + x)$

• **Cellular models:** models the system breaking it into many interacting particles (cells). Each cell interacts with its neighbours through simple rules.
(ex. granular synthesis).

• **Carver's crumple:** is a cellular model technique used to model mechanical systems such as strings, membranes and other vibrating objects. As we could suppose given that it is a cellular model, the carver's crumple approach decompose the model into many interacting particles.

It considers forces and displacement particles on a single axis (the fact that the simulation is only one dimensional is also the algorithm's main limit).

PHS

• **Physical Model Synthesis:** using one or more of the techniques described above, we can model the entire instrument that emits the sound. To do so, we must conceptually divide the instrument in interacting blocks and model each one of those using the techniques described above. However, non linearity is crucial to maintain a stable oscillation we need to manage the non-linearity.

Let's analyse the technique that we can use to model the different parts of a musical instrument:

Frequency of resonance: a model vibration of a cord or a cone structure.
• **Modal synthesis** (lumped model): it is a technique used to analyse the behavior of complex vibrating systems. This is done by decompose these systems into independent oscillators, called normal modes. Each normal mode is characterized by its specific resonance frequency.

Once the string propagation length is corded, it's more vibrant.
• **D'Alembert equation** (distributed model): the D'Alembert equation describes the propagation of a wave in an elastic medium. The equation is:
• general equation (1D): $\frac{\partial^2 y}{\partial x^2}(x,t) = \frac{1}{c^2} \frac{\partial^2 y}{\partial t^2}(x,t)$ where $c = \sqrt{\frac{T}{\mu}}$ (transverse wave velocity)
• general solution: $y(x,t) = y_1(c t - x) + y_2(c t + x)$ (linear wave theory)

We can use these results in two ways:
(finite difference approximation) (DWF)
① **Discretization of the general equation**
In the case we want to directly numerically resolve the continuous PDE using difference we can approximate partial derivatives.
The discretized equation is:
$$y(n+1, t) = y(n, t) + c \cdot [y(n, t) - y(n-1, t)]$$

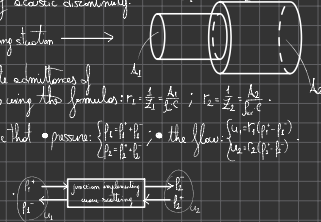
② **Discretization of the general solution**
Let's sample the solution:
$$y(x, t) = y_1(c t - x) + y_2(c t + x)$$

The discretized equation is:
$$y(n+1, t) = y_1(n - m) + y_2(n + m)$$

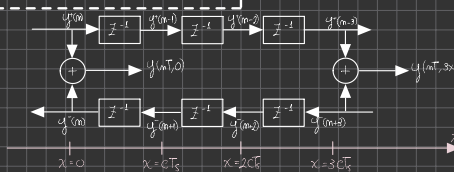
It is necessary to respect the five orders, every condition on sampling steps in time on space: $\Delta t = \frac{\Delta x}{c} \leq \frac{1}{c \cdot f_s}$
The objective is to map the continuous relation of the D'Alembert equation into a digital representation or discrete solution.

Scattering junction for 1D (passive)
• **Kelly-Lockhart junction:** this junction models the behavior of sound waves in joints of acoustic discontinuity.

Consider the following situation:
We can calculate admittances of the two surfaces using the formulas: $Z_1 = \frac{p_1}{u_1}$; $Z_2 = \frac{p_2}{u_2}$
We can also write that: pressure: $p_1 = p_2 = p$; the flow: $u_1 = u_2 = u$



Given these informations we want to get the scattering matrices:
① We must apply the continuity conditions:
• $u_1 + u_2 = 0$
• $p_1 = p_2 = p$
(the latter is due to the fact that the pressure is the same in all the two pipes)



Substituting each term with its expression we get:

$$\begin{cases} u_1, u_2 \Rightarrow r_1(p_1 - p_1^-) + r_2(p_1^- - p_2^-) = 0 \\ p_1^- - p_2^- \Rightarrow p_2 = p_1 + p_1^- - p_2^- \Rightarrow \begin{cases} p_1 = p_2 - p_1^- \\ p_2 = p_2 - p_1^- \end{cases} \end{cases}$$

$$\rightarrow r_1(2p_1^- - p_2) + r_2(2p_2 - p_1) = 0 \Rightarrow -p_2(r_1 + r_2) = -2r_1p_1^- - 2r_2p_2^+$$

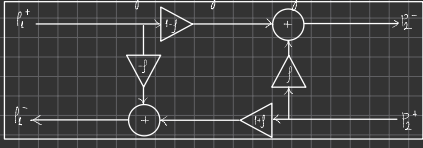
$$\rightarrow p_2 = \frac{2r_1p_1^- + 2r_2p_2^+}{r_1 + r_2}$$

[Note: p_1^- and p_2^+ are constants]

② Using this last equation we can obtain the scattering equations:

$$\begin{aligned} p_2 &= \frac{2r_1p_1^- + 2r_2p_2^+}{r_1 + r_2} \rightarrow \begin{cases} p_1^- = p_1 + p_1^- + p_2^- = \frac{2r_1p_1^- + 2r_2p_2^+}{r_1 + r_2} \rightarrow p_1^- = \frac{2r_1p_1^- + 2r_2p_2^+ - r_1p_1^- - r_2p_2^-}{r_1 + r_2} = \frac{r_1p_1^- + 2r_2p_2^+ - r_2p_2^-}{r_1 + r_2} \\ p_2^- = p_2 + p_2^- + p_1^- = \frac{2r_1p_1^- + 2r_2p_2^+}{r_1 + r_2} \rightarrow p_2^- = \frac{2r_1p_1^- + 2r_2p_2^+ - r_1p_1^- - r_2p_2^-}{r_1 + r_2} = \frac{r_1p_1^- + 2r_2p_2^+ - r_1p_1^- - r_2p_2^-}{r_1 + r_2} \end{cases} \end{aligned}$$

• The block scheme of the Kalou-Lobbaum algorithm is:



Kalou algorithm: it is a synthesis technique used to simulate plucked strings. It is a very fast algorithm, running at suitable for real time use. The fundamental component of this algorithm is the 1st order NR comb filter: $H_{NR} = \frac{1}{1 - gZ^{-m}}$.

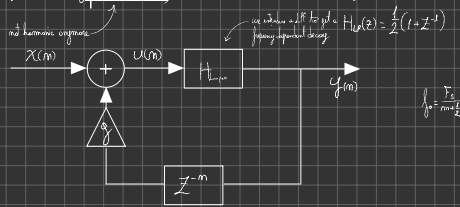
$$y(n) = x(n) + g \cdot y(n-m) \rightarrow Y(z) = X(z) + g \cdot Y(z)Z^{-m} \rightarrow X(z)(1 - gZ^{-m}) = Y(z) \rightarrow \frac{Y(z)}{X(z)} = \frac{1}{1 - gZ^{-m}}$$

Note: for fine-tuning (accurately adjust the frequency) we must use fractional delay

In this way, we obtain an harmonic frequency similar to with a fundamental frequency that is $f_0 = \frac{F_s}{m}$ if the sign of the waves is constant

• $f_0 = \frac{F_s}{m}$ if the sign is reversed at each reflection (frequency odd multiples of f_0)

• In order to obtain a more realistic model of the harmonic decay of the string we can add a LPF in the feedback of the algorithm:



$$f_0 = \frac{F_s}{m \cdot 2}$$

We have seen that a physical model is generally specified by $\left\{ \begin{array}{l} \bullet \text{ a set of CONTINUOUS EQUATIONS (ODE/PDE), each describing a functional element.} \\ \bullet \text{ a set of CONTINUITY CONSTRAINTS that specify how such elements interact with each other.} \end{array} \right\}$. In the analogy world, this is represented by a set of equations that need to be solved. We convert to turn this set of equations into a computable signal flow.

Definition of Voltage Waves: consider a n -port circuit element. The Kiroff variables for that object are $\left\{ \begin{array}{l} \bullet \text{ the port voltages } v_m \\ \bullet \text{ the port currents } i_m \end{array} \right\}$. Given that the voltage waves are defined as $\left\{ \begin{array}{l} \bullet a_m = v_m + Z_0 i_m \text{ (incident wave)} \\ \bullet b_m = v_m - Z_0 i_m \text{ (reflected wave)} \end{array} \right\}$ where $i_m = \frac{a_m - b_m}{2Z_0}$. We can also write the Kiroff-Volter and Wave-Kiroff Transformations in matrix form:

- Knopf $\rightarrow$ Welle: $\begin{bmatrix} a_m \\ b_m \end{bmatrix} =$

- Ume $\rightarrow$ Kineff: $\begin{bmatrix} G_m \\ i_m \end{bmatrix} = \frac{1}{2} \cdot \begin{bmatrix} 1 & 1 \\ \frac{1}{Z_n} & \frac{1}{Z_n} \end{bmatrix} \cdot \begin{bmatrix} u_m \\ b_m \end{bmatrix}$

● In Wave Digital Filters we model the individual modules and the topology separately. Let's analyze each one of these aspects.

Constitutive equations of one-pore elements: for one-pore elements, in continuum, we have that $\tilde{h}(\psi(x), \alpha) = 0$. In discrete terms the equation becomes $\tilde{h}(\psi(x), \alpha) = 0$. The majority of linear one-pore elements is characterized by a constitutive equation in the form: $\psi(x) = P_0(x) + \alpha(x) \cdot \psi(x)$. In this equation, we put the definitions of incident and reflected waves, we get:

$\begin{aligned} \vec{v}[k] &= R_k^{-1} \cdot \vec{r}_k[k] = V_k[k] \longrightarrow \frac{\vec{r}_k[k] - b[k]}{2} = R_k[k] \cdot \frac{\vec{a}[k] - b[k]}{2Z[k]} + V_k[k] \longrightarrow a[k] Z[k] + b[k] Z[k] = R_k[k] \cdot \vec{a}[k] + b[k] \cdot 2Z[k] V_k[k] \longrightarrow b[k] Z[k] + R_k[k] = a[k] R_k[k] Z[k] + 2Z[k] V_k[k] \longrightarrow b[k] = \frac{R_k[k] Z[k]}{Z[k] + R_k[k]} \cdot a[k] + \frac{2Z[k]}{Z[k] + R_k[k]} V_k[k] \end{aligned}$

$\vec{r}[k] = \frac{\vec{a}[k] \cdot b[k]}{2}$, $b[k] = \frac{a[k] \cdot b[k]}{2Z[k]}$

$\bullet$ a vector of port voltages: $V = [V_1, \dots, V_N]^T$ form a subset of independent port voltages we can set this vector as: $V = V_0 \cdot V_{\text{base}}$ where selecting the input voltage.

generalized out at resistors

Topological Functions: an N-port topological junction is an open interconnection node characterized by:

- a vector of port voltages: $V = [V_1, \dots, V_N]$ → from a triplet of independent port voltages we can get this vector as: $V = Q \cdot V_1$ where Q is the incidence matrix
- a vector of port currents: $I = [I_1, \dots, I_N]$ → from a triplet of independent port currents we can get this vector as: $I = B \cdot I_1$ where B is the fundamental loop matrix

- **Topological Junctions for Wave Digital Structures:** in the WD domain a junction is modelled as a SCATTERING MATRIX (ADAPTOR). Given:
 - $\mathbf{a} = [a_1, \dots, a_N]^T$ vector of incident waves to be split
 - $\mathbf{b} = [b_1, \dots, b_N]^T$ vector of reflected waves to be split
 - $\mathbf{Z} = \text{diag}(Z_1, \dots, Z_N)$ diagonal matrix of port parameters
 The scattering relation is $\mathbf{b} = \mathbf{S} \mathbf{a}$
 - $\mathbf{S}$ = scattering matrix (N x N)

One pair of a topological WD junction can be made reflection-free (optimal). The M -th pair of a WD junction is made reflection-free if the M -th diagonal entry S_{MM} is composed of 0. This condition can be obtained by properly setting the free parameter Z_M (root of the nonlinear element).

Wave Digital Filters will only non-linearly: in case of presence of one non linearly, the WDF is modeled as a CONNECTION TREE (where NON are the adaptors). In case the topology is made only by series and parallel connections, the WDF can be modeled as a BINARY CONNECTION TREE (BCT) $\Rightarrow$ makes one 3-ports parallel or series adaptors.

Parallel 3-port junctioning $\rightarrow$ consider the following circuit connection:

We know from the Digital Wave Definition that $\begin{cases} a_n = \frac{1}{2}(u_n + b_n) \\ b_n = \frac{1}{2}(u_n - a_n) \end{cases}$

• If we apply the continuity equation we get $\begin{cases} \psi_1 = \psi_2 = \psi_3 = \psi_1 + \psi_2 = \psi_1 + \psi_3 = \psi_2 + \psi_3 \\ i_1 + i_2 + i_3 \rightarrow G_1(\psi_1^* \psi_1) + G_2(\psi_2^* \psi_2) + G_3(\psi_3^* \psi_3) \end{cases}$ and writing the reflected waves as a function of the incident waves we get $\begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} = \underbrace{\begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix}}_{\text{SCATTERING MATRIX}} \cdot \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix}$ where $\alpha_i = \frac{2G_i}{G_1 + G_2 + G_3}$ and $\beta, \gamma = \frac{\text{reflection}}{\text{sum of all}}$

Calculating V_1 , V_2 and V_3 from the previous material formulae, we discover that all the ports exhibit a LOCAL NONRECURSIVE REFLECTION. To turn this "generic" parallel junction into an adaptor, we need to add a constraint on the three reference port voltages to make it REFLECTION FREE. In fact, adaptors are special type of junctions that have one adaptor port, in that they can form circuits being connected with each other without creating computability problems.

Computational flow in connection trees (BCT):

- ① FORWARD PASS: waves reflected by the linear elements (LEs) are computed and propagated through the junctions up to the non-linear element.
- ② LOCAL NON-LINEAR SCATTERING STAGE: given the incident waves the waves reflected by the non-linear element is computed.
- ③ BACKWARD PASS: waves are propagated through the junctions down to the linear elements.

Differences Between DWS and DWF:

Angela Ancona

Digital Wave Slices

Digital Wave Forms

Optimal context of use	Ideal to model concern in which waves propagate in a continuous medium	Useful to model physical electrical circuit and, in general, systems where port-to-port interconnections are predominant
Fundamental principles	Based on the derivation of waves equation's relation (D'Alambert)	Based on the conversion from Kneiff "domain" (current and tension) to Digital Wave domain (incident and reflected wave).
Key components	Digital delay lines: use model the travelling of the waves in the continuous medium with a series of delay lines; each delay line represents a discrete spatial step and time step. Scattering Junctions: 2x2 blocks that manage reflection and transmittance of waves when they go through an impedance discontinuity. Apply the Kneiff continuity law for position and velocity. D'Alambert equation: provides the relation of the equation describing travelling waves on a continuous medium & in the mathematical case of DWS. Characteristic impedance: determines the relation between Kneiff variables and waves variables.	Kneiff to Wave mapping: convert current and tension into incident and reflected waves. Scattering Junctions: they are 2x2 junctions that manage the port-to-port connections between different modules of a lumped model circuit. They implement the Kneiff continuity laws for tension and current. Discretized circuit elements: resistor, inductor and capacitor are converted in discrete representations, computable in wave domain. Free parameters: suitable parameters to eliminate not computable logs. Impedance R: used in K$\leftrightarrow$W mapping.
Stability	Stability is assured by the fact that all elements are passive elements.	To ensure stability we eliminate non computable logs using free parameters.

3D AUDIO

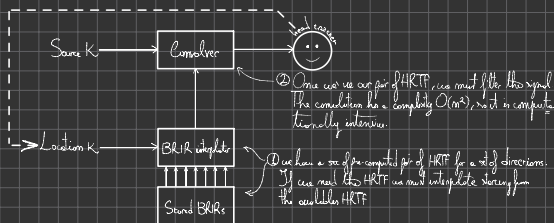
The objective of 3D audio is to create a sense of immersion when one is listening to a sound.

- **Spatial sound with headphones: binaural rendering** → the reproduction of spatial sound through headphones allows to experience a sense of immersion. There are two ways to implement 3D audio through headphones, each with its recording and rendering techniques.

HRTF FILTERS METHOD

- **Simple binaural recordings:** for binaural recordings are placed two microphones positioned in a dummy head in order to imitate the information about primary cues and reverberation. The problem of this implementation is the front-back confusion and the fact that the sound does not move when one rotates the head.

HRTF Binaural space rendering: this method consists in the use of a pair of HRTF filters for each direction. These transfer functions are valid for anechoic environment, which is unobstructed and does not correct distance perception.



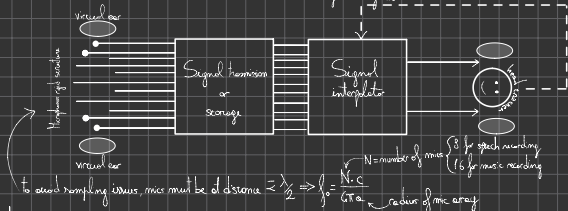
Pro: requires conventional recording practices.

Cons: computationally inefficient, not effective for low sound since it simulates an anechoic chamber.

MOTION TRACKED BINAURAL METHOD

- **Motion tracked binaural method:** in this method one uses a microphone 'array' structure having dimension on shape of a head instead of the HRTF filters:

Head tracking is used to control interpolation so that outputs are interpolated to match the orientation of the listener's ears, ensuring that the location of sound sources are perceived to be independent of the head.



Pro: computationally simpler effective for low sound.

Cons: listener not able to move freely (only head rotation), doesn't require conventional recording practices.

- **Spatial audio with Loudspeakers:** use want to reproduce spatial audio using loudspeakers. As we have seen in the headphones case, we must firstly record the "spatial audio signal" properly. There are 3 main type of audio recording/reproduction techniques:

Channel-Based Approach

- In channel based approach, utilize un numero fisso di canali audio per rappresentare il campo sonoro. Ogni canale è destinato a una specifica posizione nel lo spazio d'ascolto.
- **Fase di registrazione:** microfones are positioned to record sound for each specific channel (ex. central channel, lateral channels and presence channels in surround 5.1).
- **Fase di riproduzione:** speakers are positioned exactly in the same position of the microphones in the recording phase.

Transform-Domain Approach

- In Transform Domain Approach we represent the sound field using mathematical transformations like harmonic decomposition. Sound is captured and reproduced as a combination of coefficients that describe sound field in terms of spherical waves.
- **Fase di registrazione:** in this type of techniques specialized microphones are used to capture the sound from all the directions and decompose it into a set of coefficients.
- **Fase di riproduzione:** speakers can be configured in various positions because we can adjust the coefficients to reproduce the field in that configuration.

Object-Based Approach

- In this category of techniques we represent the sound field as a set of independent "sound objects", each one with its characteristics (position, density, ...).
- **Fase di registrazione:** sounds are recorded as independent object with spatial information (volume, metadata on position).
- **Fase di riproduzione:** "sound objects" are dynamically placed into the listen area using this metadata. The reproduction system use a spatial rendering engine to calculate how each "sound object" must be reproduced by the speakers.

Continue in the next page.

• We analyze these reproduction techniques more in detail

Stereophony panning

- Two-channel stereophony is the simplest and more diffuse technology to present virtual sound events. It exploits ILD (Inter-aural Level Difference) and ITD (Inter-aural Time Difference):

SO = input of the virtual source

$$\begin{cases} \text{for 1st signal: } d_1(t) = g_1(t - \tau_1) \\ \text{for 2nd signal: } d_2(t) = g_2(t - \tau_2) \end{cases} \text{ where } \begin{cases} \tau_1 - \tau_2 = \text{ITD} \\ g_2 - g_1 = \text{ILD} \end{cases}$$

We want now to get the PANING MATHEMATICAL LAWS:

(assumptions: • loudspeakers in far field of the source, • $\theta \in [0, \pi]$ (between x axis and))

- 1) We consider the image on the right, the sound field generated from each one of the two loudspeakers is $P_L(x, y, z) = g_L e^{i k \cdot r_L}$ and $P_R(x, y, z) = g_R e^{i k \cdot r_R}$ where $\mathbf{k} = k \hat{\mathbf{x}}$ and $\mathbf{r}_L = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + z \hat{\mathbf{z}}$ and $\mathbf{r}_R = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + z \hat{\mathbf{z}}$. The total field is: $P(x, y, z) = g_L e^{i k \cdot r_L} + g_R e^{i k \cdot r_R}$

Applying the Euler's Identity assuming $kx \ll 1$ we get that the phase is same for $\cos$ in $P(x, y, z) = \frac{g_L}{2} e^{i k \cdot r_L} + \frac{g_R}{2} e^{i k \cdot r_R}$

- 2) We know that the phase of a plane wave that comes from a target direction θ is $\phi(\theta) = k \cdot \mathbf{r} = k r \cos(\theta)$.

- 3) We want to get the gain values to reproduce the effect of the plane wave coming from direction θ :

$$\frac{g_L}{2} e^{i k \cdot \mathbf{r}_L} + \frac{g_R}{2} e^{i k \cdot \mathbf{r}_R} = \cos(\theta) x \rightarrow \cos(\theta) = \frac{g_L}{2} \frac{\partial}{\partial x} - \frac{g_R}{2} \frac{\partial}{\partial x}$$

we change the gain values in order to change the position of θ (the loudspeakers are fixed)

Multi-channel stereophony: VBAP (Vector Based Amplitude Panning)

VBAP is a panning technique that extends panning to multi-channel configurations (sources are binaural):

- 1) For each virtual sound source, the gain value for the two loudspeakers that will represent that source.
- 2) The amplitude of the signal sent to the loudspeakers is calculated so that there are no sound gain peaks in the desired point of the virtual source.
- 3) The vector of the direction of the sound is represented as a linear combination of the position of the two loudspeakers. We know that the vector of the direction of the sound is $\mathbf{d}$ and the position of the loudspeakers is $\mathbf{L}_1$ and $\mathbf{L}_2$. To create immersion perception, VBAP exploits ILD and ITD.

Practically, the panning coefficients are calculated using the formula:

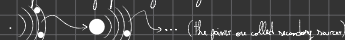
$$\mathbf{g} = \sum_{i=1}^N g_i \mathbf{L}_i$$

where $\mathbf{L}_i$ are the position vectors of the loudspeakers and $\mathbf{g}$ is the vector of the panning coefficients.

- Wave Field Synthesis (WFS) aims to create an immersive sound field on an extended area. It uses a series of secondary sources whose reproduction recreates the original field.

Physical principles:

- 1) Huygens' Principle: any point on a propagating wavefront can be seen as a point source for producing secondary waves.



- 2) Kirchhoff-Helmholtz integral: we can measure the pressure field $P(\mathbf{r}, \omega)$ inside a volume V summing up all the infinite contributions of the secondary sources placed on the bounding surface ∂V .

Implementation:

- Loudspeakers forming a high number of loudspeakers are positioned on the boundary of the listen area. Each loudspeaker works as a secondary source.
- Driving functions calculation: this needs to be sent to the loudspeakers, so calculated using the Kirchhoff-Helmholtz integral.

$$P(\mathbf{r}, \omega) = -\frac{1}{4\pi} \int_{\partial V} \left[\frac{\partial}{\partial n} P(\mathbf{r}', \omega) G(\mathbf{r}, \mathbf{r}') - P(\mathbf{r}', \omega) \frac{\partial}{\partial n} G(\mathbf{r}, \mathbf{r}') \right] dA(\mathbf{r}')$$

gain function: transfer function of a sample

- Ambisonics is a technique for recording/reproduction of binaural and sound field based on spherical harmonics.

Basic principles:

- In recording phase, sound is captured using an array of microphones typically settled in a tetrahedral or spherical configuration at the center of the sphere than in the area that has to be listened. The microphone samples the sound field in its harmonic components (3D space).
- In reproduction phase, speakers can be positioned in a linear configuration (2D plane). We can basically use whatever speaker configuration we want. This signal to be reproduced by the speakers are calculated using this spherical harmonic components of the field and the information about loudspeakers positions.

2D Ambisonics (Linear and Loudspeakers on the same plane)

Assume loudspeakers placed on a circle with radius R . We want to create a listening area inside the circle ($r < R$). The internal sound field will be represented in terms of circular harmonics, which must be approximated with a limited number of harmonics. The truncation leads to an error ϵ . We can calculate it and use it to get the acceptable number of harmonics M :

$$M = \left\lceil \frac{1}{\epsilon} \left(e^{-\frac{r}{R}} - \frac{r}{R} \right) \right\rceil$$

order of the signal

- 3D Ambisonics: in this case coefficients are estimated from the microphone signals in the enclosed phase. The idea is the same of the 2D case but now we consider a sphere instead of a circumference.
- High Order Ambisonics: increase the sound quality increasing the number of channels, allowing a better reproduction and a higher variability of the sound signal.

* Note clarification on loudspeakers configurations:

Channel based configurations: each source in the sound field is represented by a loudspeaker (source technique). This category includes the two channel stereophony.

Model based configurations: sources are not associated to physical sources, instead they are considered the virtual sound sources. This category includes VBAP, WFS, AMBISONICS.

Pros: supports conventional reading, information
long computation is sufficient due to convolution and interpolation

- Pros: supports conventional reading, no memorization
- Cons: computationally inefficient due to convolution and integration

[illegible][illegible]

EXTENDED AREA SPATIAL SOUND

- [illegible]

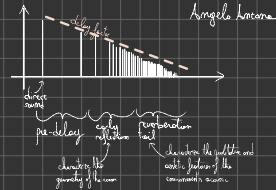
MESSAGE MULTICHANNEL

- [illegible]

REVERB

Angela Brown

Reverberation is an effect that prolongs the sound giving the characteristic of an environment. Basically, it is a very fast delayed repetition of the signal. There are three different stages in reverb:



- **Reverb descriptors:**
 - **Energy Decay Curve (EDC):** measures the energy left on the Reverb Impulse Response after a time t : $\text{EDC}(t) = \int_0^{\infty} h^2(\tau) d\tau$. The time that the function takes to reach the threshold of -60dB is called T_{60} . Note: if the noise floor is higher than -60dB threshold, we should use a higher threshold.
 - **Center Time:** is the centroid of the squared impulse response: $T_c = \frac{\int_0^{\infty} \tau h^2(\tau) d\tau}{\int_0^{\infty} h^2(\tau) d\tau}$. (Note: h is the impulse response of the room)
 - **Clarity Index:** is an indicator for speech intelligibility: $C_{50} = 10 \log_{10} \left(\frac{\int_0^{0.5} h^2(\tau) d\tau}{\int_{0.5}^{\infty} h^2(\tau) d\tau} \right) \text{ (dB)}$ (Note: h is the impulse response of the room)
 - **Definition Index:** similar to Clarity Index but we compare the first 50ms of reverberation with the entire reverberation.

There are three main ways to implement reverberation:

DELAY NETWORKS

In this type of implementation, the signal is passed through a series of delay lines. The delayed signal is fed back into the system to create a series of reflections that mimic the reverberation found in natural environment.

- **Components:**
 - delay lines: units that delay the signal by a specified amount of time.
 - filter: used to simulate the natural frequency spectral distribution.
 - feedback path.

- We need two model:
 - **early reflections:** are modeled using a tapped delay line with a FIR filter.
 - **late reflections:** add a reverb block to control late reverberations. This block can be implemented as:

Comb filter (IR filter)

$$H(z) = \frac{z^{-m}}{1 - g z^{-m}}$$

All pass filter:

$$H(z) = \frac{1 - g z^{-m}}{1 - g z^m}$$

• **Room's Tapped Delay Line** is a structure in which the late reverb is driven by the output of an early echo tapped delay line (TDL) to increase the echo density of the late reverberations.

CONVOLUTIONAL REVERBERATION

Impulse convolving the dry audio signal with an impulse response IR recorded or estimated from a real or modeled space.

- **Components:**
 - impulse response (IR): captures acoustical behavior of an environment.
 - convolution engine: the filter.

Deepen look into the various propagation algorithms:

FINITE ELEMENT METHOD

The environment is once it is divided in small finite elements (usually tetrahedrons or hexahedrons). For each element, the propagation of the wave is described by the discretization of the PDE.

BOUNDARY ELEMENT METHOD

In this approach, the PDE are converted in integral equations on the boundaries of the domain. The solution is calculated only on the borders of the domain, reducing problem's complexity.

IMAGE SOURCE METHOD

In the method reflections are modeled as secondary sources placed under the boundaries of the environment. It is assumed that sound propagates only straight and that reflections follow Snell's law (perfectly specular reflection).

BEAM TRACING

It is a more optimized version of the ray tracing. In this technique we trace beams of rays. The algorithm is the following:

- 1) Initialization: we create the virtual model for the environment and we define the source position of the surfaces.
- 2) Division of the area into small regions.
- 3) Generation and tracing of the beams.
- 4) Calculation of the reflection and refractions.
- 5) Build the beam tree: record all the interactions and reflections in the beam tree.
- 6) Sound field calculation: use on the beam tree to calculate the energy that reaches the listener.

Reverb

Delay is an effect that changes the location of the sound going out the boundaries of an enclosed room.

There are different indicators to describe the quality of a reverb effect.

There are three main ways to implement reverb:

A) DELAY NETWORKS: in this case we implement reverb by a series of delay modules.

- early reflections are modeled using a tapped delay line (TDL)
- late reflections can be modeled in two ways:
 - comb filter
 - All pass filter
- also can simulate the frequency dependent reflection characteristics

B) Convolutional Delay

In this technique we numerically reflect the IR of a environment. In the reverb process just consider the delay imposed by the IR.

- Early Decay Time is measure the decay of the IR
- $E_{DT} = T_{60} \rightarrow EDC(0) = \int_0^{T_{60}} IR dt$
- Decay time is control of the reverb IR: $DT = \frac{\int_0^{T_{60}} IR dt}{\int_0^{T_{60}} IR dt}$
- Clarity Index is an indicator that measures speech intelligibility $CI = \log_{10} \frac{\int_0^{T_{60}} IR dt}{\int_0^{T_{60}} IR dt}$
- Definition index is number for the clarity index but it is smaller.



C) Computational Acoustic Simulation with statistical physical modelling to simulate the sound propagation in a modeled geometry. The main components of this model are:

- 1) geometric model of the environment
- 2) wave propagation algorithm. There are 4 main methods:
 - ① Ray Tracing Method
 - ② Image Source Method
 - ③ Beam Tracing Method
 - ④ Finite Elements Method

FINITE ELEMENTS METHOD
This numerical way is used in real world and it is not easy to use. The sound waves are modeled through the behavior of the PDE.

BOUNDARY ELEMENTS METHOD
In this method the sound waves are modeled on the boundaries of the environment. The way we can calculate the reflection is by the boundary impedance of the surface.

IMAGE SOURCE METHOD
The image source method is based on the idea of image sources. The sound waves are modeled as if they were coming from a virtual source. The sound waves are modeled as if they were coming from a virtual source.

RAY TRACING METHOD
The ray tracing method is based on the idea of rays. The sound waves are modeled as if they were traveling in straight lines. The sound waves are modeled as if they were traveling in straight lines.

③ Ray Tracing Method
④ Beam Tracing Method
⑤ Finite Elements Method

③ Ray Tracing Method
④ Beam Tracing Method
⑤ Finite Elements Method

