Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

KA210-VET - Small-scale partnerships in vocational education and training

AURALIZE

Innovative Course in Room Acoustics and Auralization for Sound Professionals

Project No. 2024-1-IT01-KA210-VET-000245142

Virtual acoustics, production workflow, and software

Assoc. Prof. Marko Horvat

05/05/2025 – University Zagreb, Academy of Music

Event organizer: PP – UNIVERSITY OF ZAGREB
Faculty of Electrical Engineering and Computing, HR

Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

AURALIZE

Erasmus+ Project

www.auralize.eu

CONSORZIO FUTURO IN RICERCA, IT – LEAD PARTNER

SVEUČILIŠTE U ZAGREBU, Faculty of Electrical Engineering and Computing (FER), CRO – PROJECT PARTNER

Project duration: 12 months 01/11/2024 – 31/10/2025

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

AURALIZE

Erasmus+ Project

www.auralize.eu

Objectives

- Develop and implement an **innovative vocational training course** in **room acoustics and auralization**.
- Provide hands-on experience in acoustic measurements and creative labs for students and professionals.
- Enhance digital listening skills and foster sonic thinking in professional and academic fields.
- Strengthen cross-national cooperation in vocational training between Italy and Croatia.

CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

AURALIZE

Erasmus+ Project

www.auralize.eu

WP1 - Project Management and Communication

WP2 - Development of the Course in Room Acoustics an Auralization

WP3 - Course Implementation, Acoustic Measurements and Creative Labs

Kick-off meeting: February 28th, 2025 (online)

Intermediate meeting: May 6th, 2025, Zagreb (CRO)

Final meeting: September 12th, 2025, Bologna (ITA)

CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

AURALIZE - Workplan

1. Lecture A1 - Historical overview and importance of acoustics as intangible heritage (CFR - **March 21, 2025** - online);
2. Lecture A2 - Fundamentals of room acoustics. (FER, **April 14, 2025** - in person - Zagreb);
3. Lecture A3 - Equipment and methods of recording acoustic measurements, analysis, and interpretation - theoretical and practical part (CFR in Zagreb, **April 14, 2025** - in person - Zagreb);
4. Acoustic measurements in Zagreb (CRO): Vatroslav Lisinski Concert Hall and National Theatre of Zagreb - (CFR & FER, **April 15, 2025** - in person);
5. Acoustic measurements in Faenza (ITA): Teatro Masini in Faenza (CFR & FER, **April 29-30, 2025** - in person);
6. Lecture A4 - Virtual acoustics, production workflow, and software (FER, **May 5, 2025** - in person - Zagreb);
7. Lecture A5 - Multi-disciplinary uses of virtual acoustics (cultural heritage, sonic archaeology, games, media, etc.) (CFR, **May 5, 2025** - in person, Zagreb).

CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

A4 Virtual acoustics, production workflow, and software

Contents

1. How can we (re)create a sound environment?
2. Mono systems
3. Stereo systems
4. Surround systems
5. VBAP systems
6. Binaural systems
7. Cross-talk cancellation
8. Ambisonics
9. Wave-field synthesis

CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

1. How can we (re)create a sound environment?

- Live recording
  - documenting an environment at a particular moment in time (e.g. a recording of a classical concert in a concert hall)
- Synthesis
  - creating an environment in „simulated" condition (e.g. mixing and mastering of a song, panning/placing the sound sources in a virtual environment)
- Measurements and convolution
  - obtaining the impulse response(s) of a space through measurements for specific source/receiver combinations and convolving them with „dry" sound
- Room acoustics simulation
  - obtaining the impulse responses by calculating them from 3D models of rooms for specific source/receiver combinations, with the surfaces and objects assigned with absorption and diffusion properties, and convolving them with „dry" sound

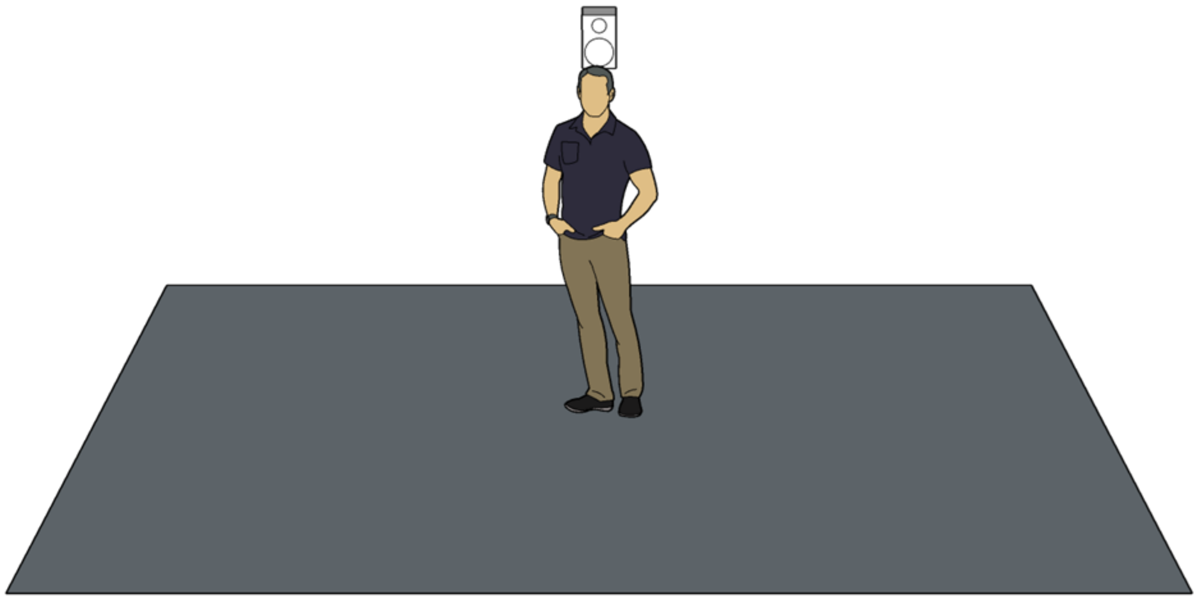
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

2. Mono sound (re)production

- Simplest and oldest format – one channel for recording, one channel for reproduction
- No perception of 3D space – the sound is arriving from a single source
- One microphone required for recording (or measurement)
    - Omnidirectional microphone – all directions equally favored
    - Directional microphone – certain directions favored at the expense of others
- The recording (or measurement) of a sound environment will contain all the desired sonic content, but no information on the spatial distribution of sound sources
- One loudspeaker required for reproduction



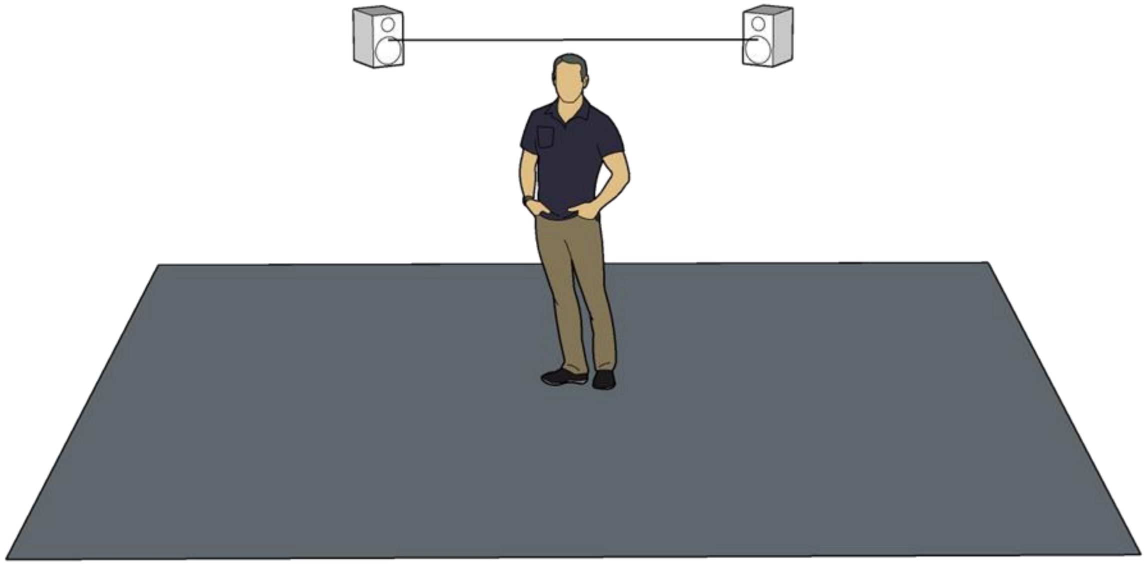
CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

3. Stereo sound (re)production

- Extension of mono – two channels for recording, two channels for reproduction
- Greatly improved perception of 3D space (although still limited) – the sound is arriving from a multitude of sources spanning between two loudspeakers (stereo image)
- Two microphones required for recording (or measurement)
  - Intensity stereophony – relies on the level difference between the two channels
  - Time-of-arrival stereophony – relies on the time difference between the two channels



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

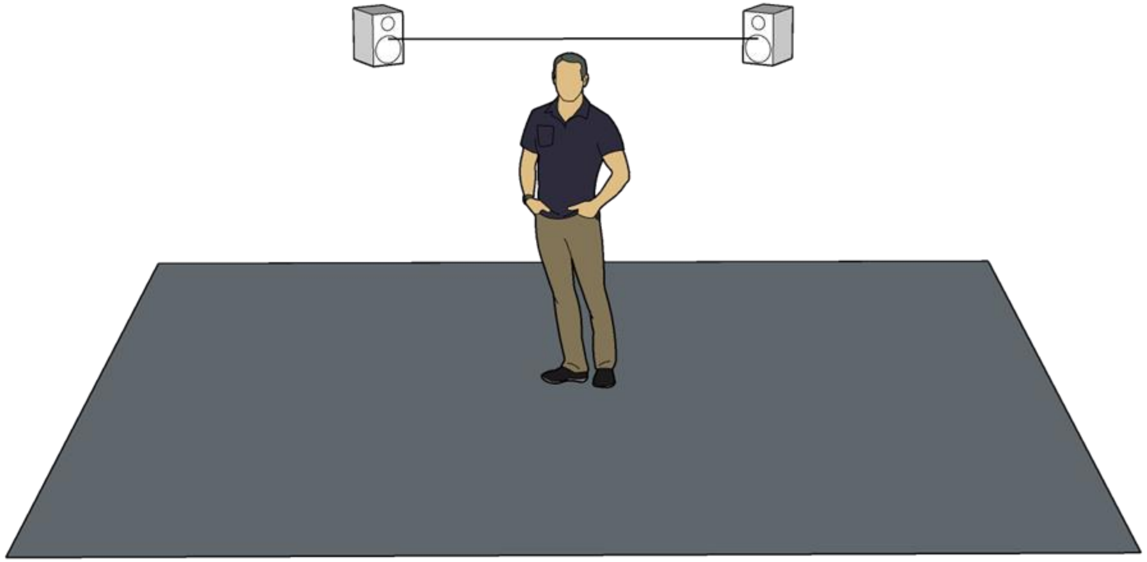
Co-funded by the European Union

3. Stereo sound (re)production

- Synthesis (creation of stereo image by positioning a mono source with sonic content *I*) –

$$L = aI, \quad R = bI$$

  - *a* and *b* are gains assigned to the L and R channels for the individual source (level/gain panning) – the gain will be higher for the loudspeaker closer to the sound source
  - requirements:
    - the source should maintain constant loudness, no matter where it is panned (i.e. in energy terms, the sum of gains should ideally remain constant)
    - panning direction depends on the difference of gains
- The recording (or measurement) of a sound environment will contain all the sonic content, and basic information on the spatial distribution of sound sources in the stereo image
- Two loudspeakers required for reproduction, located at ±30° left and right of the listener (standard stereo listening setup)



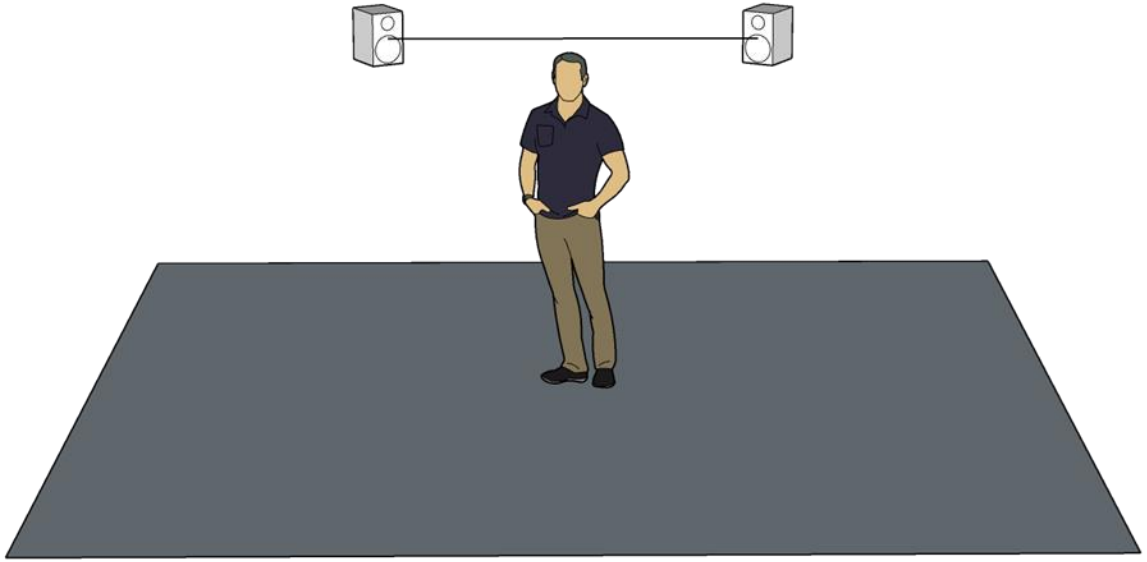
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

3. Stereo sound (re)production

- Two loudspeakers required for reproduction, located at ±30° left and right of the listener (equilateral triangle - standard stereo listening setup)
- Listening over headphones is common
  - different experience compared to loudspeaker-based listening
  - each of the two channels reproduced only to its respective ear
  - with loudspeaker-based listening, each channel/loudspeaker is heard by both ears



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

3. Stereo sound (re)production

- Most stereo recording techniques are straightforward
  - The L and R microphones record
  - The corresponding L and R channels that are directly reproduced by
  - L and R loudspeakers (each loudspeaker reproduces only its own dedicated channel)
  - The stereo image is created by the human hearing system from the recorded level and/or time differences contained in the L and R channel
- Types of stereo recording techniques
  - Coincident stereo pair
  - Spaced stereo pair
  - Near-coincident stereo pair
  - Stereo pair with a baffle

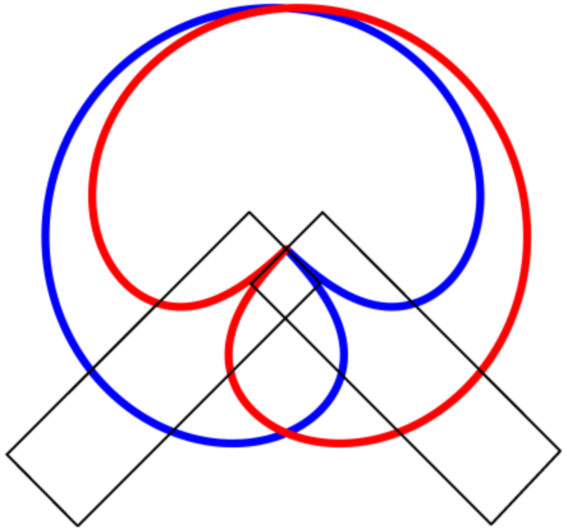
CFR CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

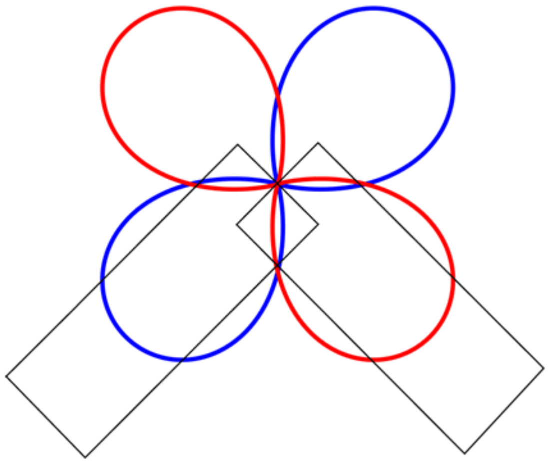
Co-funded by
the European Union

3. Stereo sound (re)production

- Coincident stereo pairs
  - Directional microphones (figure-of-eight or cardioid) with capsules at the same position in space
  - Only intensity stereophony (level differences between two channels)
  - Examples: XY, Blumlein



XY – two cardioids angled at 90°



Blumlein – two figure-of-eights angled at 90°

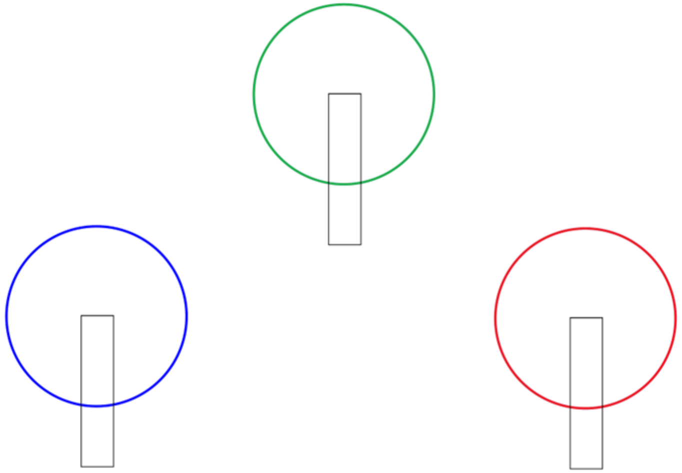
CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

3. Stereo sound (re)production

- Spaced stereo pairs
  - Omnidirectional microphones (or directional, if oriented the same way) set at a fixed distance between them (30 cm or more)
  - Only time-of-arrival stereophony (time differences between two channels; the level is essentially the same)
  - Examples: AB, Decca tree

AB – two omnis spaced at 30 cm or more



Decca tree – two omnis cca 2 m apart with a centre omni cca 1.5 m in front

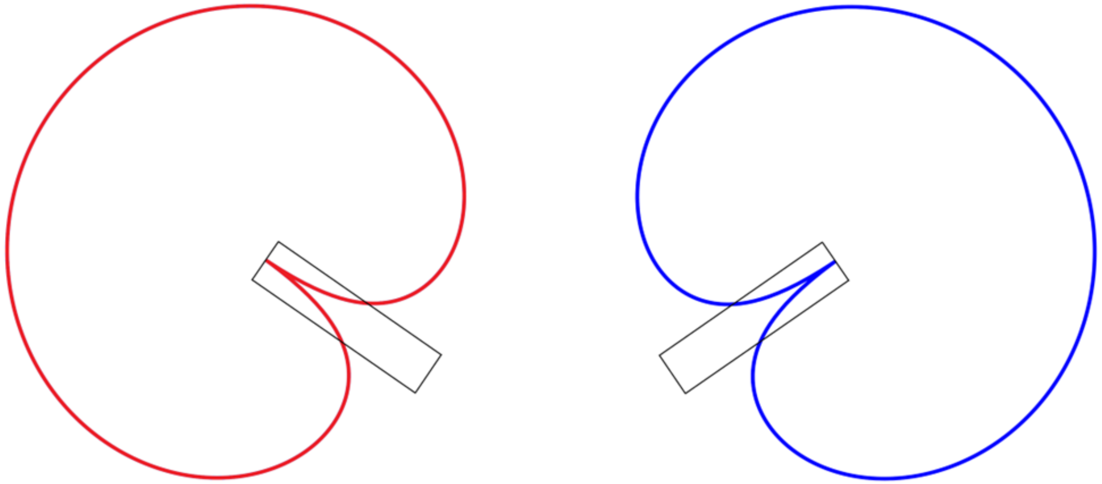
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

3. Stereo sound (re)production

- Near-coincident stereo pairs
  - Directional microphones (cardioids) set at a fixed distance between them (17 to 30 cm)
  - Emulation of the behaviour of the human hearing
  - Both intensity stereophony and time-of-arrival stereophony (level and time differences between two channels)
  - Examples: ORTF, NOS, RAI

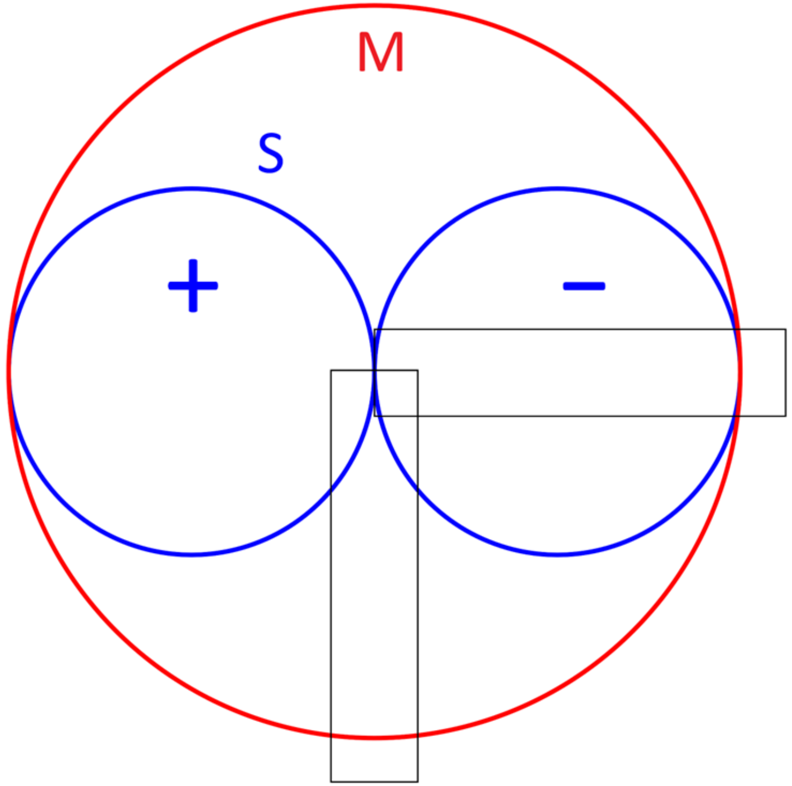


ORTF – two cardioids spaced at 17 cm and angled at 110°

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

3. Stereo sound (re)production

- Specific stereo recording technique – the Mid-Side Stereo (MS)
- Coincident stereo pair
  - Omnidirectional M (Mid) microphone
    - entire sonic content, but no information on the direction of the sources = mono signal
  - Figure-of-eight S (Side) microphone
    - supplement to the M microphone that carries the basic information on the direction of sound sources in the stereo image



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

3. Stereo sound (re)production

- the Mid-Side Stereo (MS)
  - Not straightforward as the other techniques in terms of stereo reproduction
    - The stereo image (layout of sound sources) is **encoded** in the M and S signals
    - For reproduction, i.e. the feeds for the L and R loudspeakers, the information needs to be **decoded** from M and S as

$$L = M + S, \quad R = M - S$$

  - Important! The final L and R feeds for the left and the right loudspeaker are created from both M and S

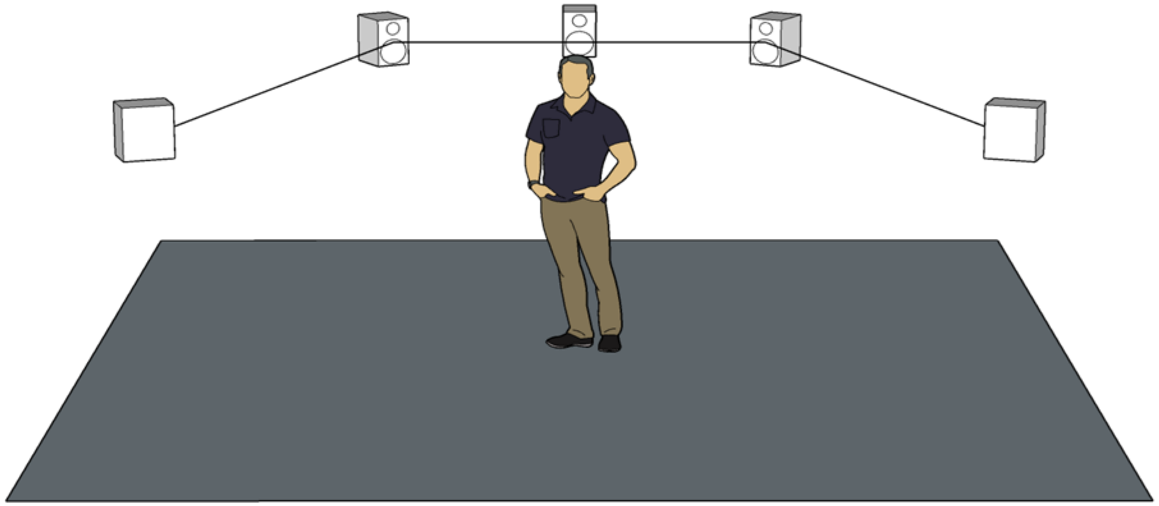
CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

4. Surround sound

- Extension of classic stereo to the horizontal plane (full or a part), with or without height coverage
- Height coverage provided by
  - Physical loudspeakers located on the ceiling
  - Virtual coverage that employs reflections from the ceiling of sound that originates from loudspeakers positioned elsewhere in the room
- Numerous configurations available
  - 5.1, 7.1,…
  - 3.1.2, 4.1.4,…



CFR
CONSORZIO FUTURO IN RICERCA

Co-funded by the European Union

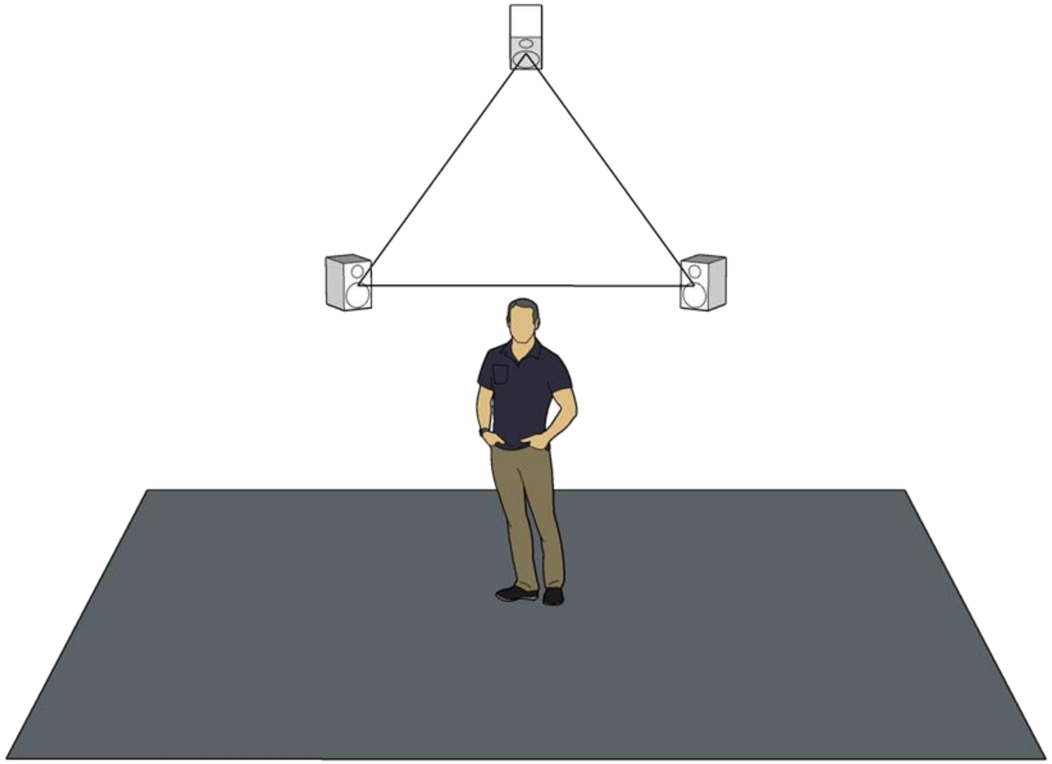
Erasmus+
Enriching lives, opening minds.

5. Vector-Base Amplitude Panning (VBAP)

- Extension of conventional intensity-based stereo to full 3D space
- Adding a loudspeaker above the stereo pair (L and R) creates an elementary triangle
- A sound source can be placed anywhere inside that triangle using these three loudspeakers
- Possible cases for sound source location
  - In a vertex -> only the corresponding loudspeaker is active
  - On an edge -> two loudspeakers are active (the ones on the end of that edge)
  - Inside the triangle -> all three loudspeakers are active
- Synthesis (positioning a mono source I) –

$$L1 = aI, \quad L2 = bI, \quad L3 = cI$$

  - *a, b,* and *c* are gains assigned to the three loudspeaker feeds (level/gain panning) – the gain will be higher for the loudspeaker closer to the sound source (it will be louder)

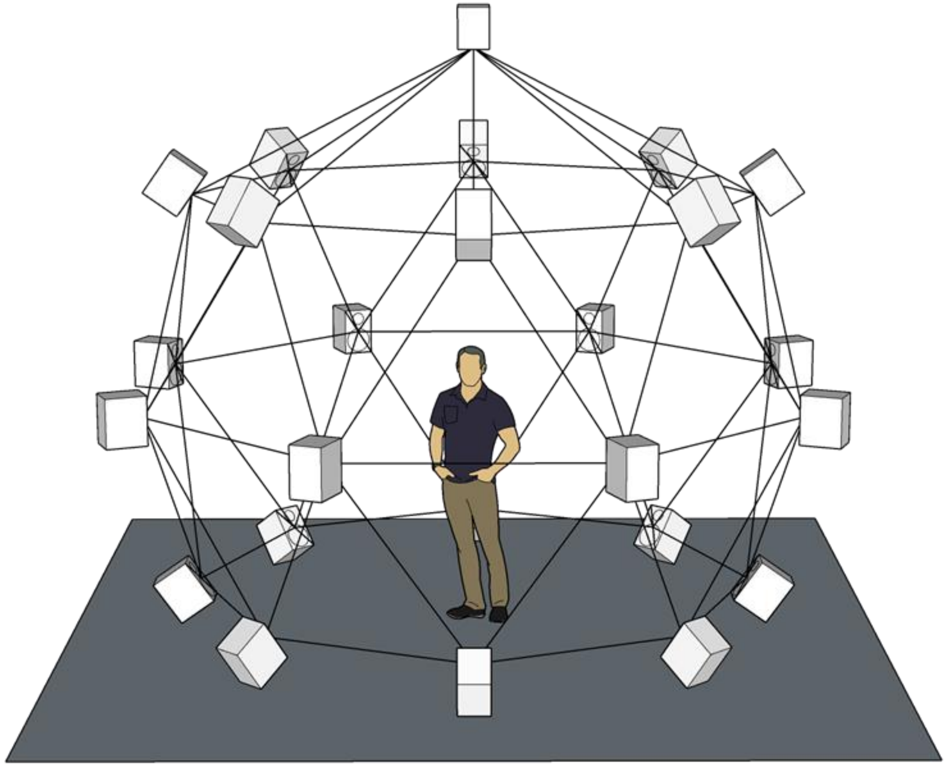


CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

5. Vector-Base Amplitude Panning (VBAP)

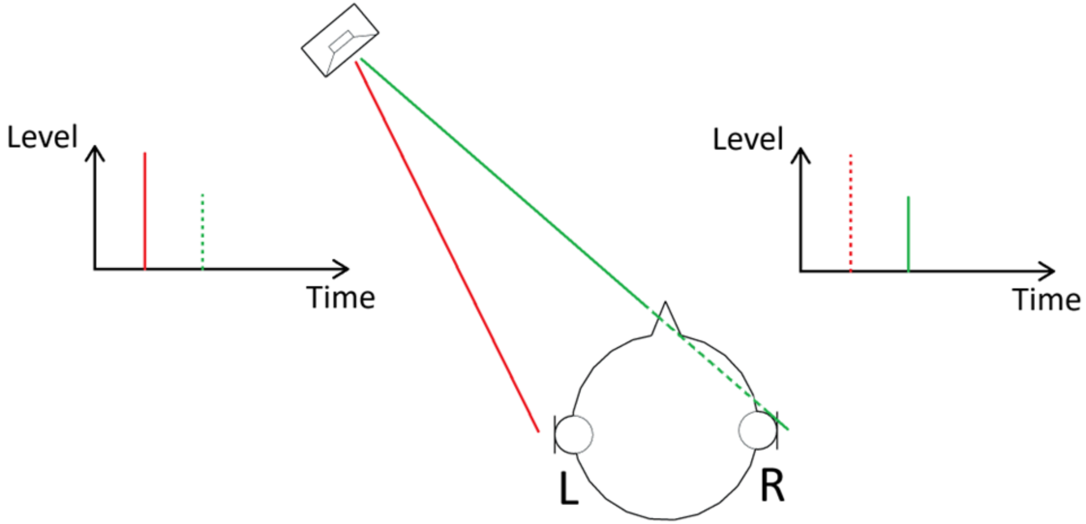
- How to create a full 3D sound reproduction system?
  - Place multiple loudspeakers around the listener position in a sphere-like configuration
  - The surface of the sphere is divided into triangles
  - Each triangle is covered by three loudspeakers
  - If a virtual sound source is placed within a triangle, only the three loudspeakers that form that triangle will participate in the reproduction of that source



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

6. Binaural systems

- Binaural hearing and localization - two basic mechanisms
  - ILD (Interaural Level Difference) = the level difference at the two eardrums
    - Dominant mechanism at high frequencies – the head shadows the sound for the far ear (seen from the source)
  - ITD (Interaural Time Difference) = the difference in time the sound needs to reach the two eardrums
    - Dominant at low frequencies, as the head does not represent an obstacle to sound



- Example
  - The sound coming from the source located in front and to the left of the listener will reach their left ear first and then the right ear, and will have higher amplitude/level at the left ear

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

6. Binaural systems

- Is this enough to describe binaural localization mechanisms? No!
- Very similar or identical ITD and ILD can be obtained for different directions of arrival, and yet, humans can localize the source at its correct position
- Example:
  - For the source located directly in front of the listener, the ITD and the ILD will be zero
  - The same occurs if the source is directly above or directly behind the listener
  - How can we distinguish between these three positions?
  - Solution? **HRIRs and HRTFs!**

CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

6. Binaural systems

- Sound is changed on its way from the source to the eardrums by the presence of the human body on that way (pinna, head, torso)
  - This change is unique for every direction!
  - It is described with HRIRs (Head-Related Impulse Responses) and HRTFs (Head-Related Transfer Functions)
  - A pair of HRIRs or HRTFs (one for the left, and one for the right ear) is defined/measured for many directions with reference to the listener and calculated for the rest -> HRTF set for all directions
- Individual HRIRs/HRTFs -> HRIRs/HRTFs are unique for every person!
  - need to be measured for every person – complicated and not feasible!
- Generic HRIRs/HRTFs
  - measured on a dummy head as the representation of the average human head (and sometimes torso)

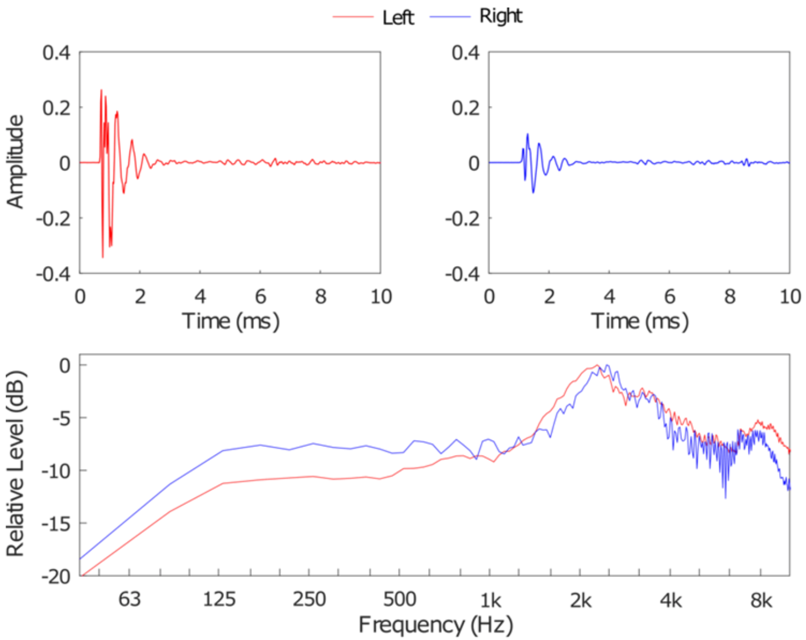


Image source: https://vocal.com/audio/head-related-transfer-functions/

Photo source: https://www.neumann.com/en-us/products/microphones/ku-100

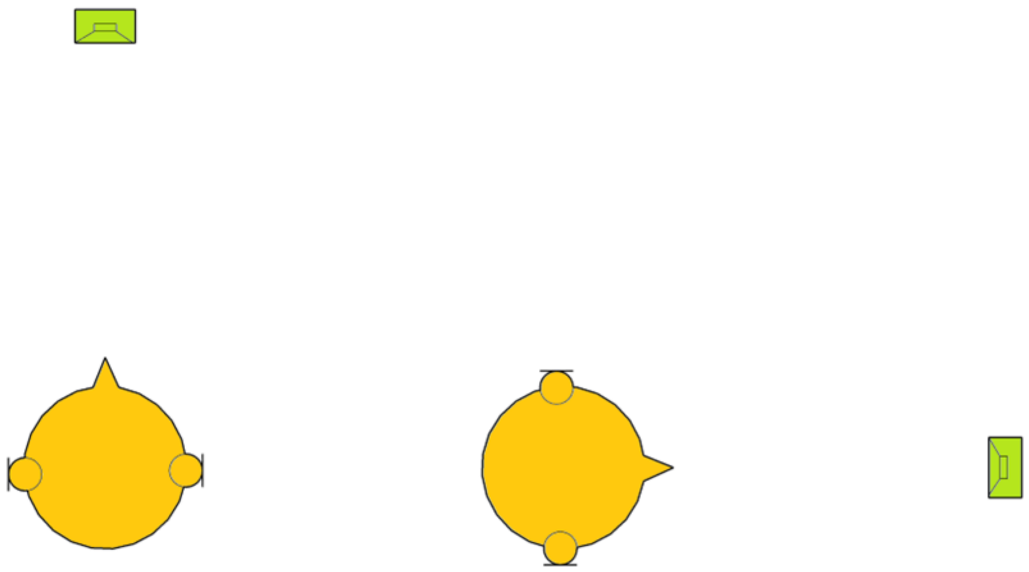
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

6. Binaural systems

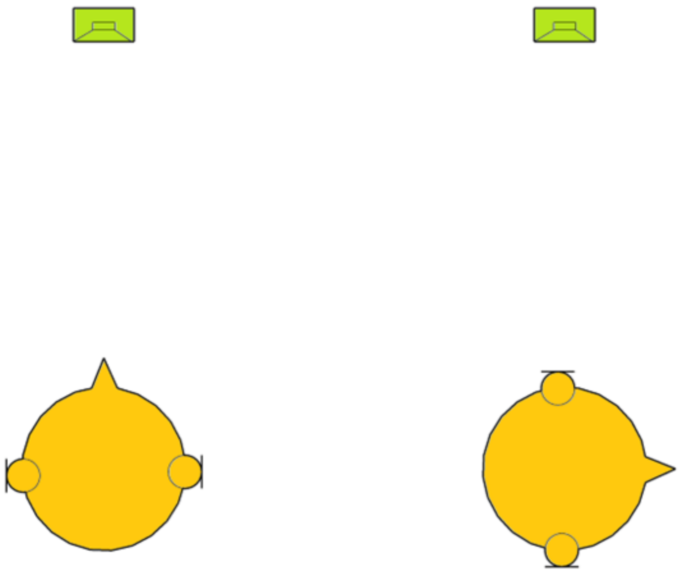
- Static binaural systems
  - Example:
    - The listener is facing to the front and perceives a virtual sound source in a binaural recording to appear in front of them
    - When the listener turns their head 90° to the right, they still perceive the virtual sound source to be in front of them
  - With head movement, the sound image „follows" the head
  - To maintain a stable 3D binaural image of the sound environment, the listener must not move their head while listening through headphones (much like standard stereo)
  - Is this so in real life? No!
  - In this example, the sound source should appear to be to the left of the listener when they turn their head as described



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

6. Binaural systems

- Dynamic binaural systems
  - Example:
    - The listener is facing to the front and perceives a virtual sound source in a binaural recording to appear right in front of them
    - When the listener turns their head 90° to the right, the virtual sound source is now perceived to be to the left
  - With head movement, the sound image remains stable
  - Is this so in real life? Yes!
  - To maintain a stable 3D binaural image of the sound environment, the orientation of the listener’s head must be known -> head tracking!
  - The head tracking data is used to dynamically update the binaural 3D sound image in real time
  - Example:
    - For the listener to perceive a sound source in front of them, a pair of HRIRs/HRTFs for the frontal direction must be used
    - For the listener to perceive a sound source to the left after turning their head to the right, a pair of HRIRs/HRTFs for the left direction (90° to the left) must be used
    - For the transition to appear smooth (so that the source does not „jump” from the front to the left), pairs of HRIRs/HRTFs must be used for the in-between directions (in this example, one might use such pairs for directions 10°, 20°, 30°,…,70° and 80° to the left)



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

6. Binaural systems

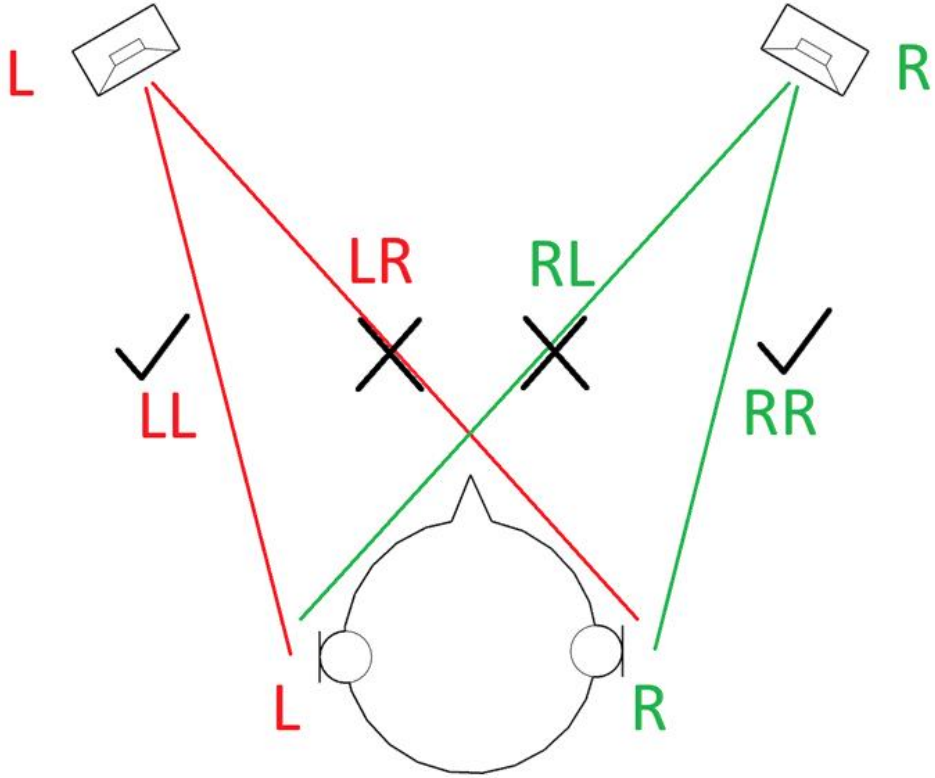
- How to go about (re)creating a sound environment?
  - Recording
    - A binaural recording of a particular event can be made using a dummy head at a specific position
    - Alternatively, in-ear microphones can be used on a person who then acts as a dummy head
  - Measurement
    - Binaural Room Impulse Response (BRIR) using a dummy head for a specific source-receiver combination
    - An extension of an ordinary Room Impulse Response – the direction of arrival of every reflection is recorded and known thanks to the changes to sound made by the dummy head (as described with HRIRs/HRTFs)
    - To be convolved with a dry signal – the reproduced result sounds as if the source of the dry signal is positioned at the location of the actual source during measurements
  - Production
    - On top of the usual procedures employed in sound production (mixing, adding effects, etc.), binaural panning allows the source to be placed anywhere in 3D space – its position is determined by using a specific pair of HRIRs/HRTFs on dry signal
  - Room acoustic simulations
    - The source and receiver position can be chosen at will
    - The BRIR is calculated using a generic HRTF set (measured on a dummy head), and then convolved with a dry signal

  - By definition, binaural signal is to be reproduced over headphones – it is crucial to have the sound contained in the L and R channels reproduced only to the respective ears – no crosstalk!

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

7. Cross-talk cancellation systems

- A way of reproducing binaural audio using loudspeakers
- Binaural audio can be reproduced over loudspeakers if and only if each ear hears the sound only from its respective loudspeaker
  - LL path = from the left loudspeaker to the left ear – desired and needed ✓
  - RR path = from the right loudspeaker to the right ear – desired and needed ✓
  - LR and RL paths = from loudspeakers to opposite ears – crosstalk! – undesired, must be eliminated/cancelled!



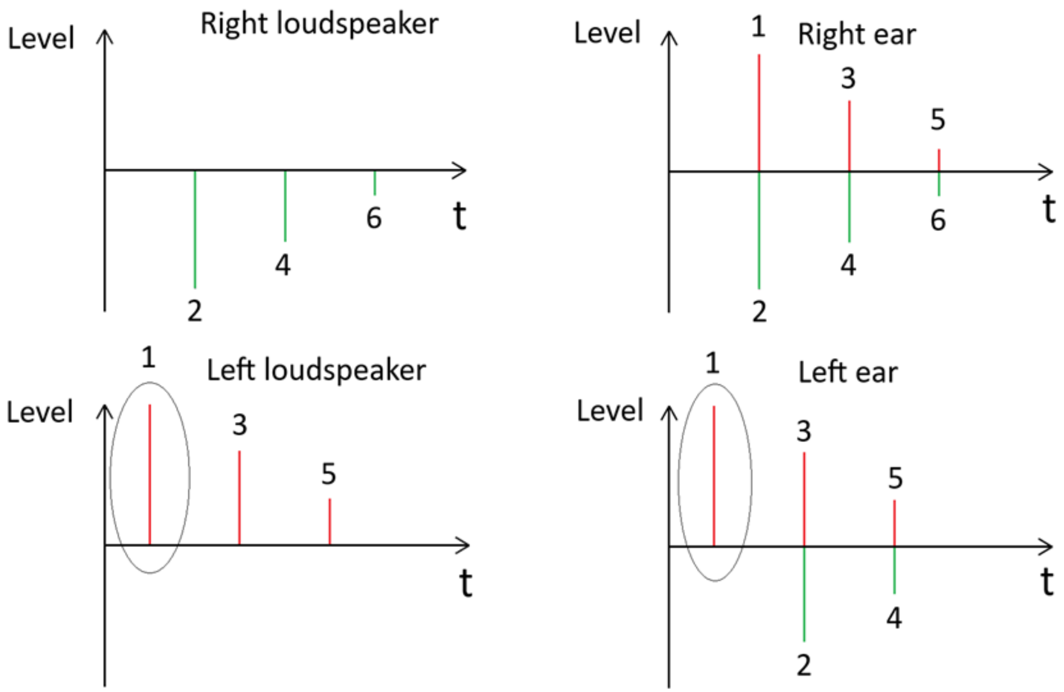
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

7. Cross-talk cancellation systems

- Basic principle of CTC – example
  - Step 1
    - **Goal:** a single pulse/click 1 is emitted by the left loudspeaker (circled) and it needs to be heard only by the left ear
    - pulse 1 arrives at the right ear a bit later and a bit weaker - to be cancelled
  - Step 2
    - right loudspeaker emits pulse 2 to cancel pulse 1 at the right ear
    - pulse 2 arrives at the left ear a bit later and a bit weaker – to be cancelled
  - Step 3
    - left loudspeaker emits pulse 3 to cancel pulse 2 at the left ear
    - pulse 3 arrives at the right ear a bit later and a bit weaker – to be cancelled
  - …
  - Step 6
    - right loudspeaker emits pulse 6 to cancel pulse 5 at the right ear
    - pulse 6 is not heard anymore by the left ear – cancellation is not required anymore!



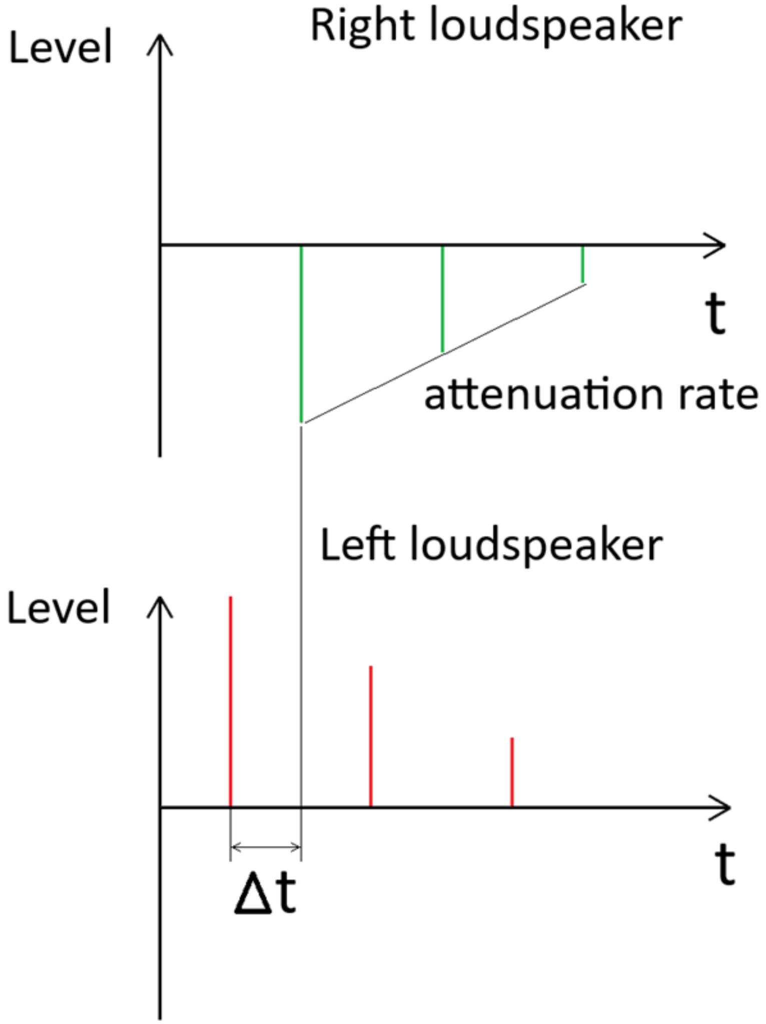
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

7. Cross-talk cancellation systems

- Most important parameters
  - Δt = the difference of arrival times to the ears
  - Attenuation rate = the reduction of level of each cancellation component referenced to the previous one
    - e.g. if set to 3 dB, the initial pulse will have the level of 0 dB, the second one -3 dB, the third one -6 dB, etc…
  - Number of cancellation components
  - Used to defined the CTC filters that need to be applied to both binaural channels
- Static CTC systems
  - The filters are fixed = calculated for a fixed listener position that does not change, and the listener must not move
- Dynamic CTC systems
  - The listener is allowed to move
  - The position and orientation of their head with respect to loudspeakers needs to be known -> head tracking!
  - The filters are updated in real time based on head tracking data

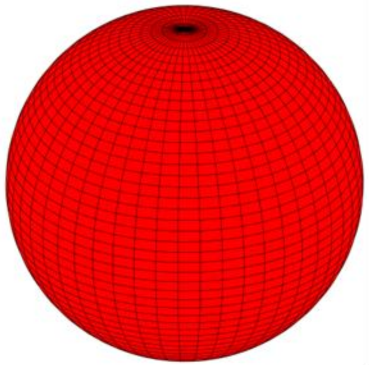


Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

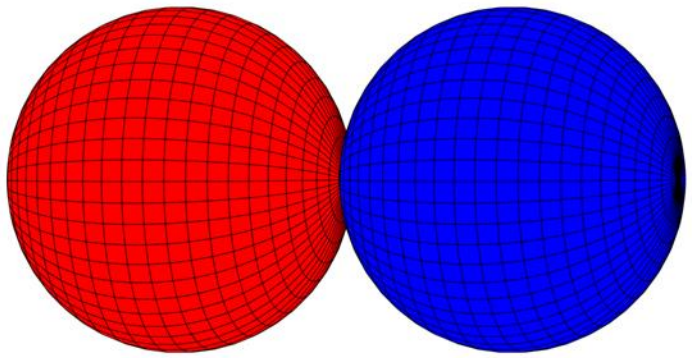
Co-funded by the European Union

8. Ambisonics

- A way to describe a sound field on a sphere around a single point in space
- In essence, an elaborate and versatile variant of coincident microphone recording
- The M/S stereo = the most primitive Ambisonics system!
  - The M channel is recorded with an omni microphone and contains all the sound content, but no information on the position of the sound sources
  - The S channel is recorded with a figure-of-eight microphone oriented on the left-right axis, perpendicular to the sound source
  - The information on the content and the position of the sound sources is **encoded** in the M and S signals
  - **Result?** The sound sources can be placed in the stereo image (left-right)
  - For reproduction, the information needs to be **decoded** for the L and R loudspeakers from M and S signals



M = omni



S = left-right figure-of eight

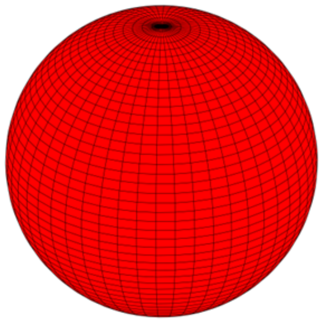
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

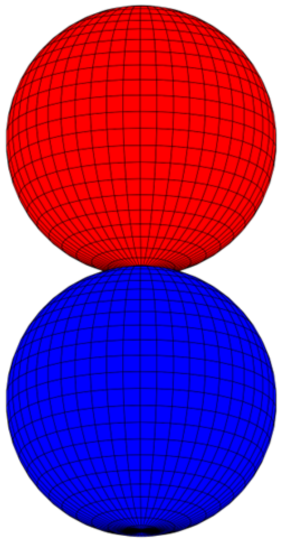
Erasmus+
Enriching lives, opening minds.

8. Ambisonics

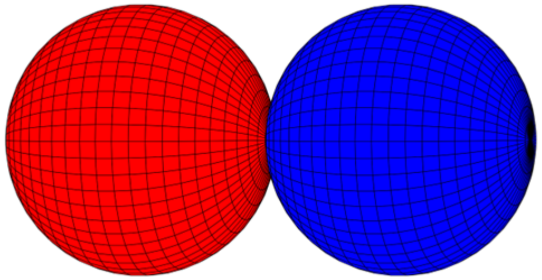
- **Double M/S = 1st order Ambisonics (2D)**
  - An extension of the M/S technique
  - A second figure-of-eight microphone added to capture the front-back directional information – oriented to the front
  - The information on the content and the position of the sound sources is **encoded** in these signals:
    - W (= M, omni),
    - X (front-back figure-of-eight),
    - Y (= S, left-right figure-of-eight)
  - **Result?** The sound sources can be placed anywhere in the horizontal plane around the listener/microphones



W = omni



X = front-back figure-of eight



Y = left-right figure-of eight

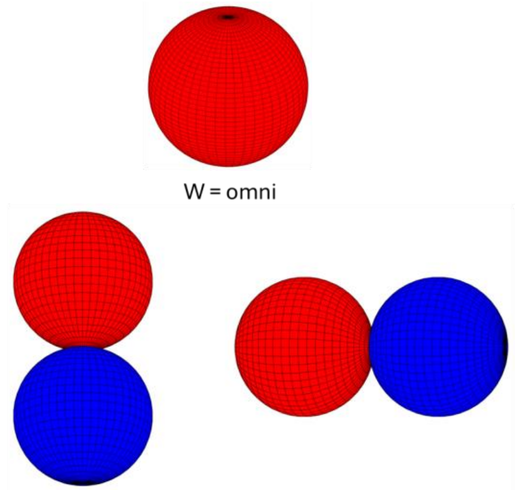
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

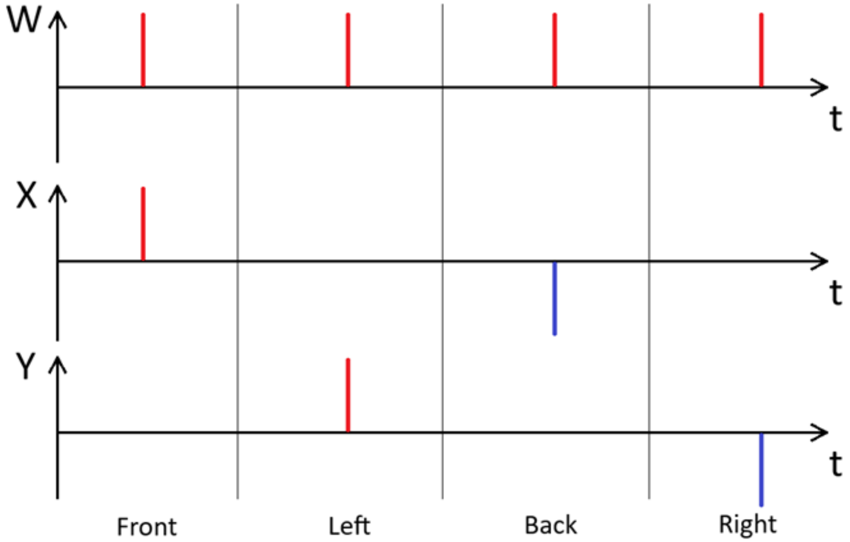
8. Ambisonics

- **Double M/S = 1st order Ambisonics (2D)**
  - Examples
    - Source in front of the microphones -> W signal at full level, X signal at full level positive, no Y signal
    - Source left of the microphones -> W signal at full level, no X signal, Y signal at full level positive
    - Source behind the microphones -> W signal at full level, X signal at full level negative, no Y signal
    - Source left of the microphones -> W signal at full level, no X signal, Y signal at full level negative



W = omni

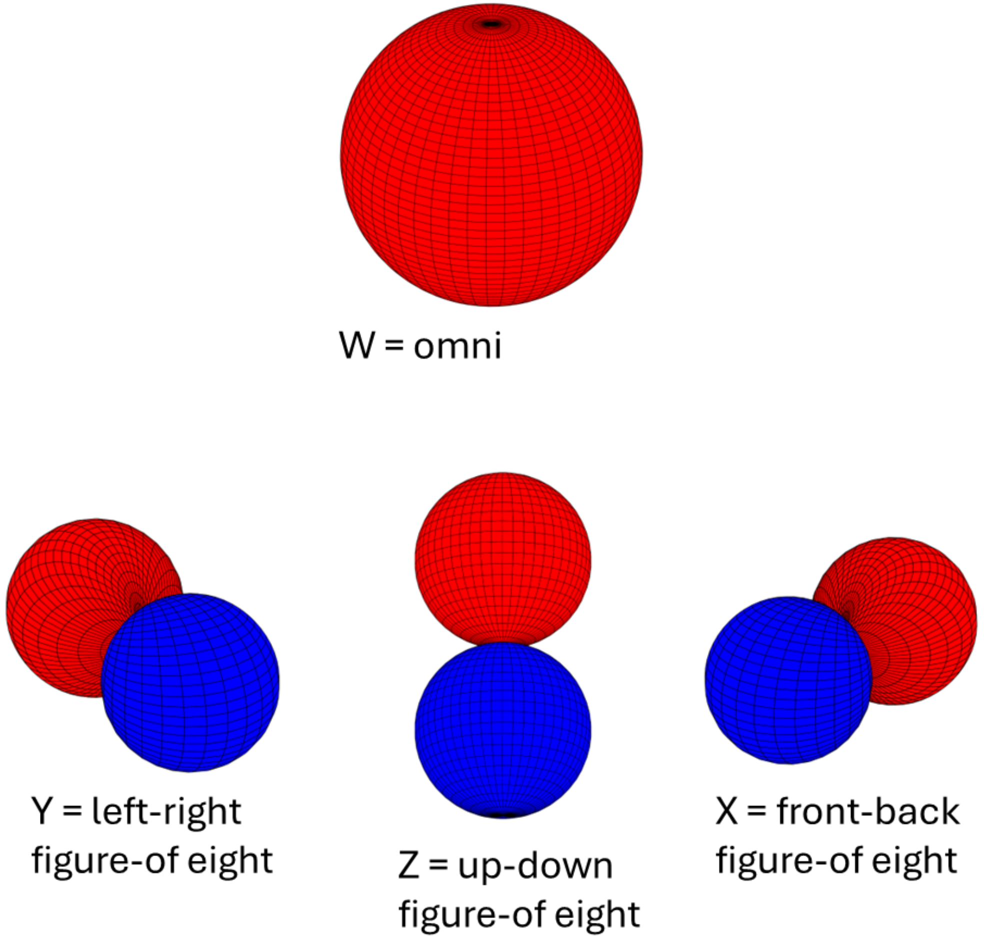
X = front-back figure-of eight   Y = left-right figure-of eight



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

8. Ambisonics

- **„Triple" M/S = 1st order Ambisonics (3D)**
  - An extension of the double M/S technique
  - A third figure-of-eight microphone added to capture the up-down directional information – oriented upwards
  - The information on the content and the position of the sound sources is **encoded** in these signals:
    - W (omni),
    - X (front-back figure-of-eight),
    - Y (left-right figure-of-eight)
    - Z (up-down figure-of-eight)
  - **Result?** The sound sources can be placed anywhere in 3D space around the listener/microphones



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

8. Ambisonics

- **„Triple" M/S = 1st order Ambisonics (3D)**
  - The most basic 3D Ambisonics audio format
  - In theory, recordings and measurements can be made with „ordinary" microphones – 1 omni (W) and 3 figure-of-eights (X, Y, Z)
  - **Result:** B-format Ambisonics
  - Practical issue:
    - All four microphones should be in the same point in space, or at least as close to each other as possible – very difficult to achieve
  - Solution:
    - Use a microphone with four cardioid capsules mounted in a tetrahedral layout – not ideal, but much more compact
    - **Result:** raw A-format Ambisonics

Photo source: https://rode.com/en/microphones/360-ambisonic/nt-sf1

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

8. Ambisonics

- **1st order Ambisonics (3D)**
  - Simple conversion from A-format to B-format
  - If the microphone capsules (raw signals recorded with four cardioids) are named as:
    - FLU (Front Left Up)
    - FRD (Front Right Down)
    - BLD (Back Left down)
    - BRU (Back Right Up)
  - the B-format (as if it was recorded with an omni and three figure-of-eight microphones) can be obtained as:
    - W = FLU + FRD + BLD + BRU
    - X = FLU + FRD - BLD - BRU
    - Y = FLU - FRD + BLD - BRU
    - Z = FLU - FRD - BLD + BRU

Photo source: https://rode.com/en/microphones/360-ambisonic/nt-sf1

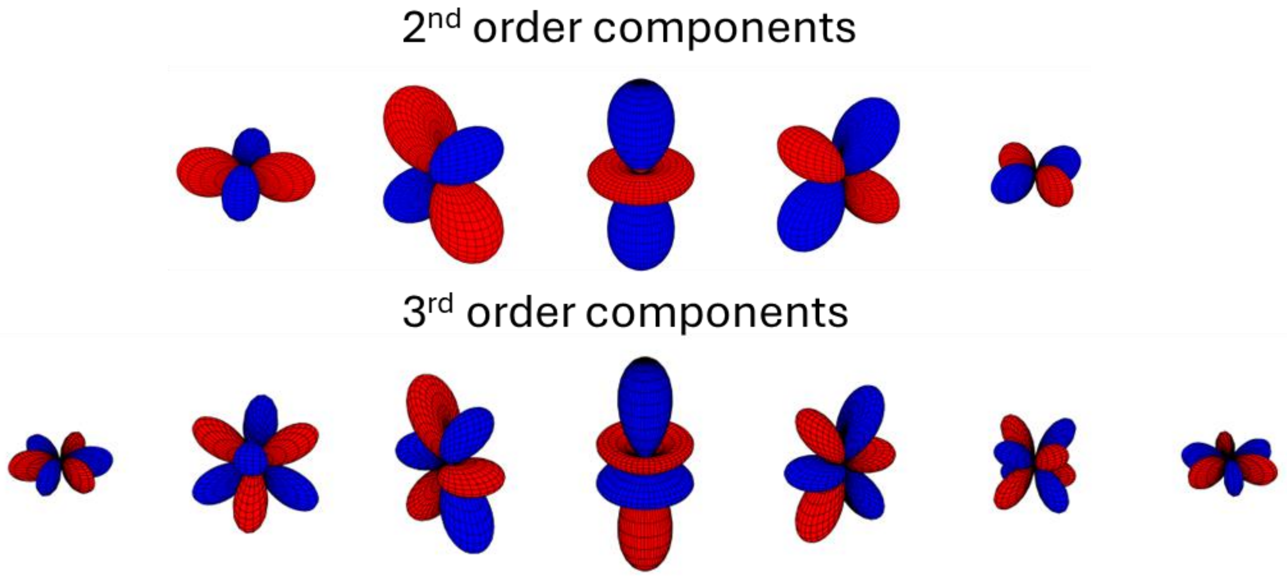
CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

8. Ambisonics

- **Higher-order Ambisonics**
  - On top of the omni (W) and the basic three figure-of-eight (X, Y, Z) components, more complex components are added
  - The shape of these components is not chosen at random = spherical harmonics!
  - **Advantage:** source position in 2D/3D space with regard to the listening/recording point can be described with greater accuracy
  - **Practical issue:**
    - Can these higher-order components be directly recorded? No! There are no real microphones with such directional characteristics.

- **Solution:**
  - Use complex microphone configrations for recording and obtain Ambisonics components (B-format) from raw signals (A-format) in post-processing
  - By taking just the right amount of each raw signal and combining these together, higher-order components can be obtained as if they were indeed recorded with real microphones with the displayed directional characteristics (the same as for the 1st order Ambisonics components X, Y, Z)



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

8. Ambisonics

- **Higher-order Ambisonics – examples of microphones for recording**
  - Core Sound Octomic
    - 8 cardioid capsules in an octahedral configuration
    - up to 2nd order Ambisonics recording
  - ZYLIA ZYLIA ZM-1
    - 19 omnidirectional capsules on a sphere
    - up to 3rd order Ambisonics recording
  - mh acoustics Eigenmike 64
    - 64 omnidirectional capsules on a sphere
    - up to 6th order Ambisonics recording

Photo source: https://www.zylia.co/zylia-zm-1-microphone.html

Photo source: https://www.core-sound.com/products/octomic

Photo source: https://eigenmike.com/eigenmike-64

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

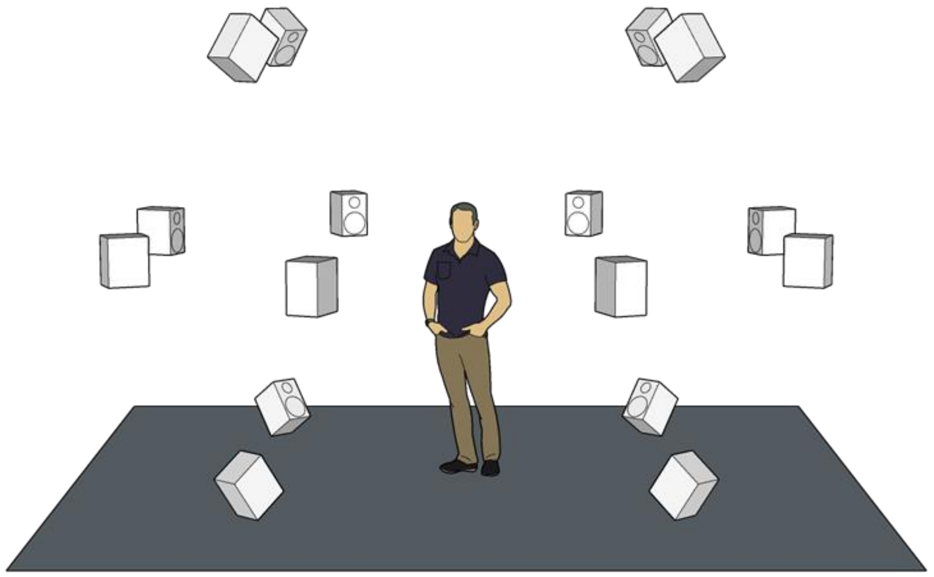
8. Ambisonics

- **Ambisonics – recording microphones**
  - Flexibility – almost any microphone setup can be emulated by processing raw signals
  - An Ambisonics microphone can be used for:
    - Classic mono recording (omni, figure-of-eight, cardioid, etc.)
    - Classic stereo recording - any coincident pair technique (XY, Blumlein, M/S) – both „microphones" in the pair can be emulated simultaneously
    - Surround recording – more than two microphones, all emulated simultaneously
    - …

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

8. Ambisonics



- **Decoding the Ambisonics format for reproduction**
  - Flexibility – Ambisonics format can be decoded for any reproduction setup
    - Mono
    - Stereo
    - Surround formats
    - Binaural reproduction
    - Ambisonics setups (circular for 2D in the horizontal plane, spherical for 3D in full space)
  - E.g. a 4-8-4 configuration
    - The 8 loudspeakers in the ring around the listener can be used for 2D reproduction
    - The 4 loudspeakers below the ring and the 4 above provide additional sonic information for full 3D reproduction

- Important!
  - **All** Ambisonics signals (W, X, Y, Z,…) are used to create feeds for **all** loudspeakers (remember M/S stereo!)
  - How much of each signal will be taken to form a feed for an individual loudspeaker depends on the position of that loudspeaker –> when setting up an Ambisonics decoder, it is crucial to have accurate data on the positions of loudspeakers

CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

8. Ambisonics

- How to go about (re)creating a sound environment?
  - Recording
    - A full 3D recording of a particular event using an Ambisonics microphone for a specific microphone position
    - B-format Ambisonics - to be decoded for reproduction
  - Measurement
    - Ambisonics Room Impulse Response (ARIR) using an Ambisonics microphone for a specific source-microphone combination
    - An extension of an ordinary Room Impulse Response – the direction of arrival of every reflection is recorded and known
    - To be convolved with a dry signal – the resulting signal is decoded for reproduction
  - Production
    - On top of the usual procedures employed in sound production (mixing, adding effects, etc.), Ambisonics panning allows the source to be placed anywhere in 3D space – its position is encoded in the Ambisonics components
  - Room acoustic simulations
    - The source and receiver position can be chosen at will
    - The ARIR is calculated for the chosen source-receiver combination and then convolved with a dry signal

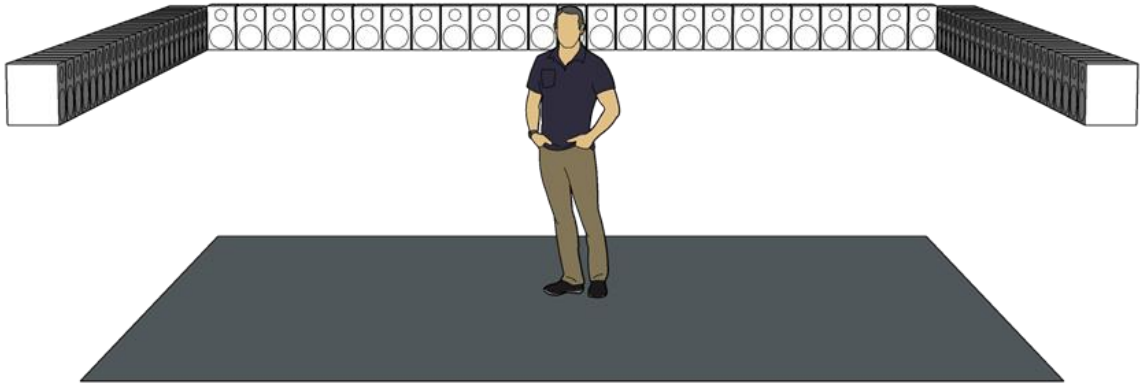
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by
the European Union

Erasmus+
Enriching lives, opening minds.

9. Wave-Field Synthesis

- A way to create the appearance of a virtual sound source by producing artificial wavefronts of sound as if emitted by that source
- How is it done?
- The Huygens principle
    - a wavefront can be viewed/constructed as a sum of spherical waves emitted by a large (infinite) number of elementary sources
- Practical application
    - Elementary sources = loudspeakers; a large number of them set as close to each other as possible
    - Appropriate time delay and gain/level are applied to each loudspeaker signal
- Full 3D would require too large a number of loudspeakers
- Application restricted to the horizontal plane

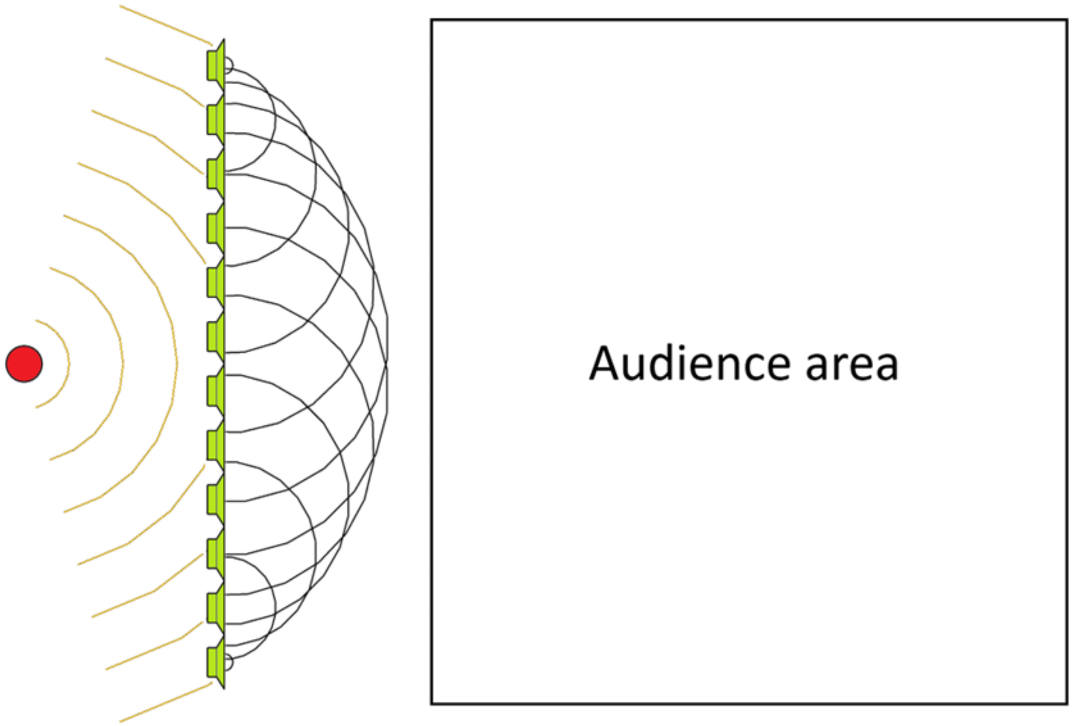


CFR
CONSORZIO FUTURO IN RICERCA

Co-funded by the European Union

9. Wave-Field Synthesis

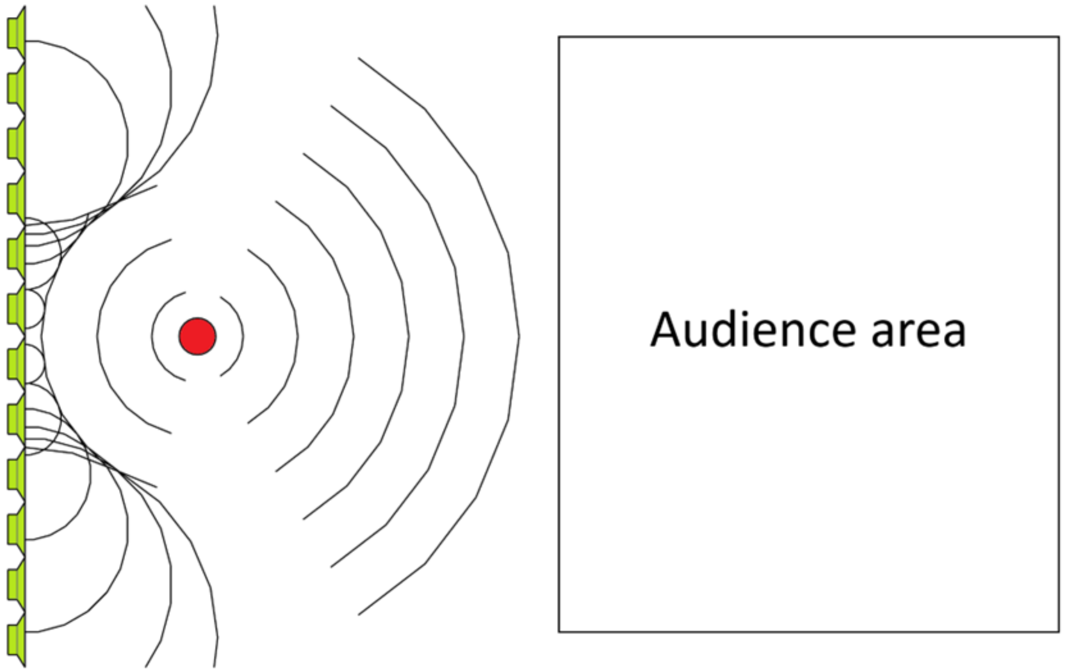
- Creation of a virtual source behind the loudspeaker array (possibly outside the room)
  - If the source (red dot) was at the marked location, the sound would reach central loudspeakers first, so they start emitting sound first
  - The farther a loudspeaker is from the virtual source, the later it starts to emit sound (needs time delay), and the lower its level is (needs level correction)
  - Working all together, the create a convex wavefront very similar to the one a real source would create if it was at that position
  - The audience (located to the right of the loudspeaker array) perceives the created wavefronts as if they were emitted by a source located at the marked position



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

9. Wave-Field Synthesis

- Creation of a virtual source in front of the loudspeaker array (inside the room)
  - To create a virtual source (red dot) at the marked location, the loudspeaker array needs to create concave wavefronts
  - These wavefronts converge to a focal point where the virtual source should be and continue propagating into the audience area as convex wavefronts (similar to a magnifying glass or concave mirror!)
  - Creation of a concave wavefront: outermost loudspeakers emit sound first; the closer a loudspeaker is to the center of the array, the later it emits sound, and the lower its level should be (time delay and level correction)
  - The audience (located to the right of the loudspeaker array) perceives the created wavefronts as if they were emitted by a source located at the marked position
  - The space between the loudspeakers and the virtual source must not be disturbed!



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

9. Wave-Field Synthesis

- Often used in research and (experimental) music production and performances
  - E.g. a concert by solo horn player with WFS recreation of a virtual source (himself) on the stage
- Virtual enviroment created by synthesis
  - Positioning the virtual sources (arbitrary number of them)
  - Adding early reflections to the reproduced sound
- Recreation of a sound environment by measuring (numerous) impulse responses is possible, but complex and time-consuming
- Advantage over other spatial audio techniques!
  - Correct re/creation of the sound field is not limited to a rather small sweet spot!
  - It extends to a sweet area (much larger than a sweet spot)!

CFR
CONSORZIO FUTURO IN RICERCA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

KA210-VET - Small-scale partnerships in vocational education and training

AURALIZE

Innovative Course in Room Acoustics and Auralization for Sound Professionals

Project No. 2024-1-IT01-KA210-VET-000245142

Virtual acoustics, production workflow, and software

Assoc. Prof. Marko Horvat

Thank you for your participation

CFR
CONSORZIO FUTURO IN RICERCA

UNIVERSITAS STUDIORUM ZAGRABIENSIS MDCLXIX