Aalto University
School of Electrical
Engineering

Communication acoustics
Ch 14: Sound reproduction

Ville Pulkki

Department of Signal Processing and Acoustics
Aalto University, Finland

September 29, 2022

Sound reproduction applications

- Public address
- Full-duplex speech communication over technical channels
- Audio content for music and cinema industries
- Broadcasting of sound in radio or of audiovisual content in TV
- Computer games and virtual reality
- Accurate reproduction of sound
- Enhancement of acoustics and active noise cancellation
- Aided hearing

A loudspeaker is always involved, and often also a microphone. Required technical specifications are very different in different applications

This chapter

- Audio content production
- Listening set-ups
- Recording techniques
- Virtual source positioning
- Binaural techniques
- Digital audio effects
- Reverberators

Audio content production

- Audio content: sound signals produced that have meaning or value to a listener.
- Audio engineering: production of audio content
- Audio engineer: recording, manipulation, mixing, mastering, and reproduction of sound
- Recording: process of capturing sound
- Mixing: process of adding different recorded tracks together
- Mastering: preparing and transferring the mixed audio track to media
- Live sound: on-line mixing and mastering during live concerts

Listening set-ups

- Headphone (monotic), headphones (diotic-dichotic)
- Loudspear setups
  - Mono, 1 loudspeaker
  - Stereo, 2 loudspeakers
- Surround setups
  - Number of loudspeakers around the listener [dot] number of subwoofers [dot] number of elevated loudspeakers
  - 5.1, 6.1, 7.2, 12.2, 22.2, 5.1.2, 7.1.4
- What more loudspeakers, that larger listening area
- More complete coverage of directions in reproduction is achieved with wider and denser loudspeaker setups

Listening room acoustics

- Rooms have different acoustic conditions
- Room acoustics has vast effect on frequency spectrum of ear canal signals
- Listeners actively adapt to rooms, and thus audio content is perceived very similar in different rooms
- Potential problems in audio content production and listening due to different acoustics in studios and domestic conditions are also mitigated by adaptation
- Standardized room acoustics exist, a few parameters are defined in certain limits

Audio-visual reproduction systems

- Loudspeaker set-up + video display
- Cross-modal effects
  - Better audio quality can make video degradation less annoying, but good video quality was not found to improve the perceived audio quality.
  - Synchronization: lead of audio in the recommendation is 20 ms, and correspondingly the maximum tolerated lag is 40 ms (ITU-T)
  - Color affects loudness
  - Audio affects direction of gaze
  - Ventriloquism

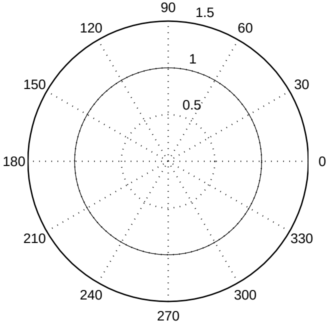
Audio-tactile systems

- Sound is also perceived through the sense of touch
- Tactile presentation of low frequencies increases the loudness with about 1 phon
- Headphones + vibrating chair, higher audio reproduction quality with music [demo by Merchel]
- Perception of asynchronicity about 10–24 ms
- Interesting future applications: haptic mixing desk knobs reproduce the track signal etc

Recording techniques

- How to position microphone(s) in relation to the sources and to the room
- Monophonic
- Spot microphone techniques
- Microphone placement techniques dedicated to certain loudspeaker listening set-ups
- Coincident techniques
- Non-linear perceptual reproduction methods

Microphone polar patterns



Omnidirectional

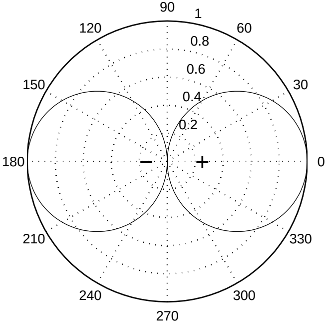
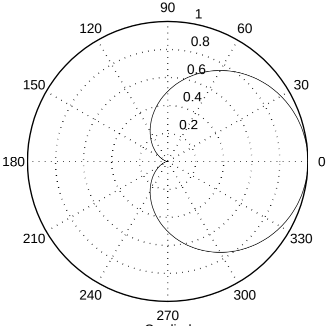


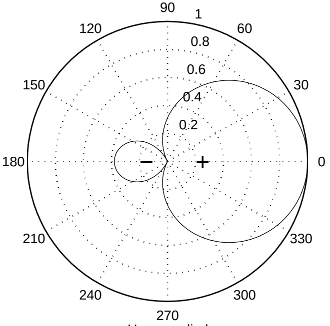
Figure of eight

Multiple coincindent microphones, polar patterns are additive.

Microphone polar patterns



Cardiod



Hypercardiod

When two signals from directive microphones are summed, the resulting (virtual microphone) signal has a directional pattern also

$x_{\mathrm{cardioid}}(t) = 0.5\left(x_{\mathrm{omni}}(t) + x_{\mathrm{dipole}}(t)\right)$

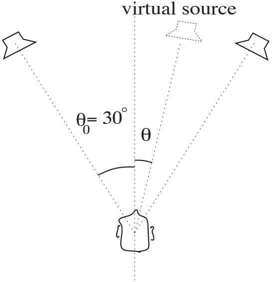
$x_{\mathrm{cardioid}}(t) = 1/3\left(x_{\mathrm{omni}}(t) + 2x_{\mathrm{dipole}}(t)\right)$

Monophonic recording

- Single microphone close to source, signal $x(t)$
  - Very probably most often used recording technique
  - Phones – walkie-talkies — etc
  - Captures only one sound signal, room effect not present
- Single far-away microphone
  - Captures all sources present
  - Reverberant sounds appear colored when reproduced with mono

Two-channel stereophony

- Two loudspeakers
- How to position two microphones to record for this layout?



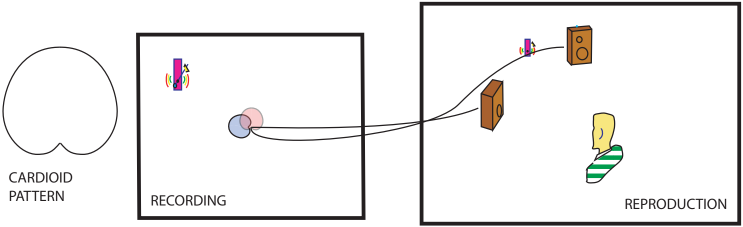
Model-based analysis of recording techniques

- Binaural auditory model
- Estimate the dependency of loudness of a source rotating around the microphone array (loudness plot)
- Estimate ITD and ILD cues for real sources
- Map ITD and ILD cues measured from recording techniques to ITD angles (ITDA) and ILD angles (ILDA)
- With perfect reproduction system, a virtual source at x° should reproduce ITDA and ILDA values of x°

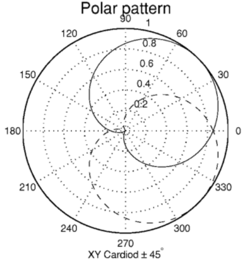
Pulkki, Ville. "Microphone techniques and directional quality of sound reproduction." Audio Engineering Society Convention 112. Audio Engineering Society, 2002.

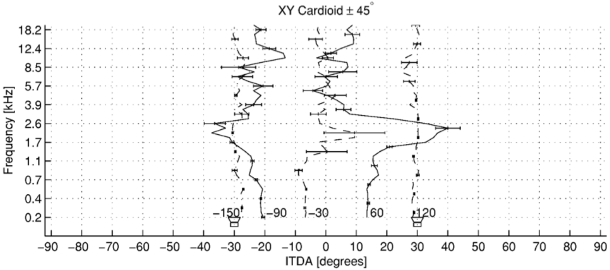
Coincident techniques for stereophony

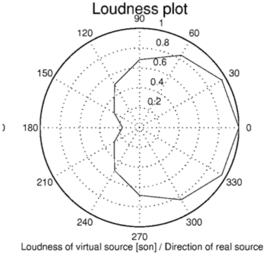
- Two directive microphones in coincident positioning
- XY (cardioids or similar), Blumlein (Dipoles)
- Virtual sources relatively point-like
- May suppress reverberation

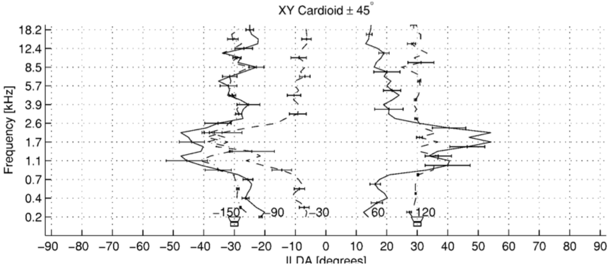


XY cardioids towards $\pm 45^\circ$

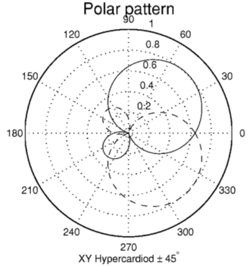


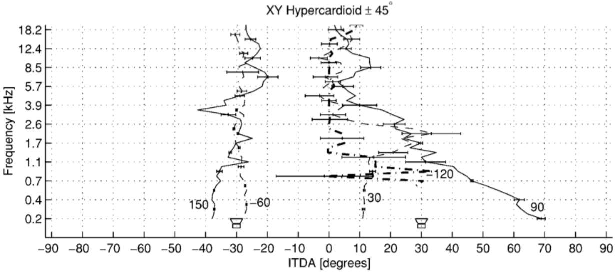


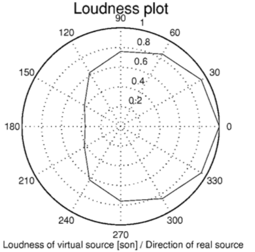


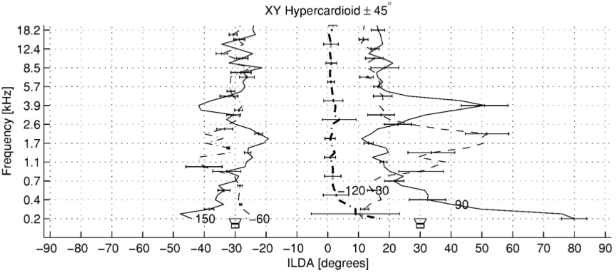


XY hypercardioids towards $\pm45^\circ$

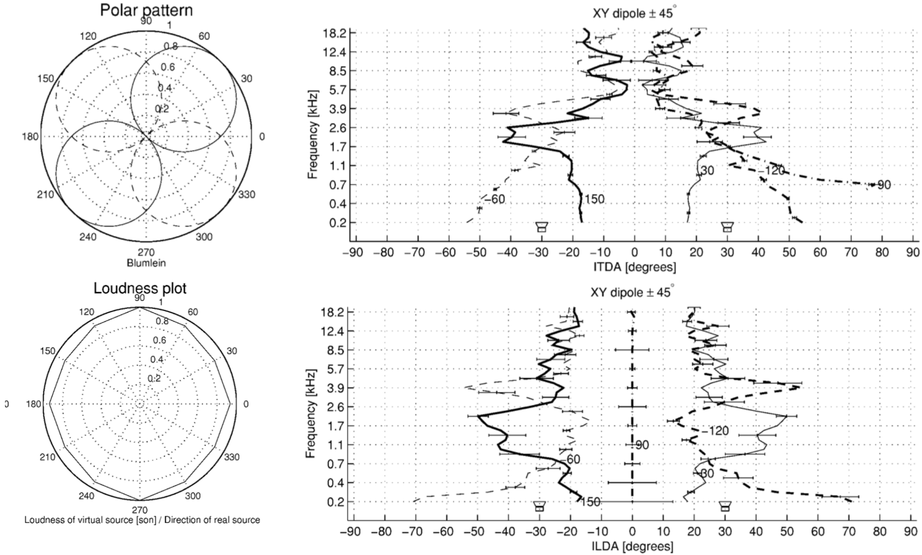






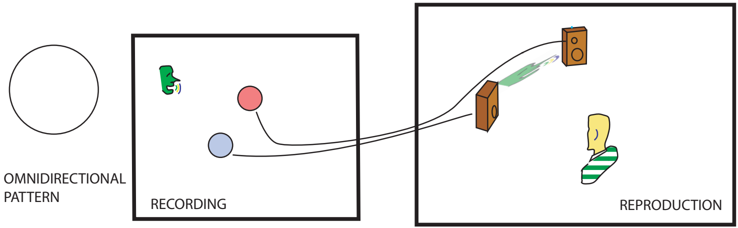


XY dipoles towards $\pm 45^\circ$ (Blumlein pair)

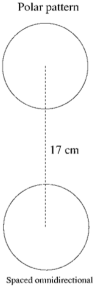


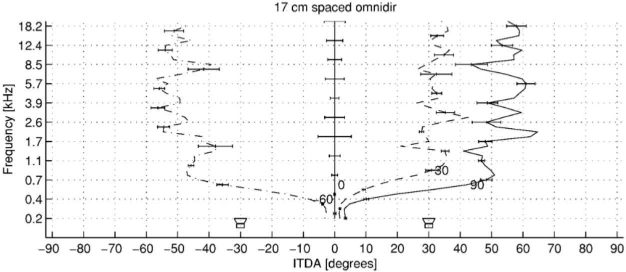
Spaced techniques for stereophony

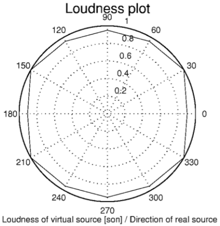
- Two directive or omnidirectional microphones spaced by 20cm – few meters
- AB technique
- Virtual sources relatively broad, and localization depends on frequency
- Reverberation perceived "airy", "open", not suppressed

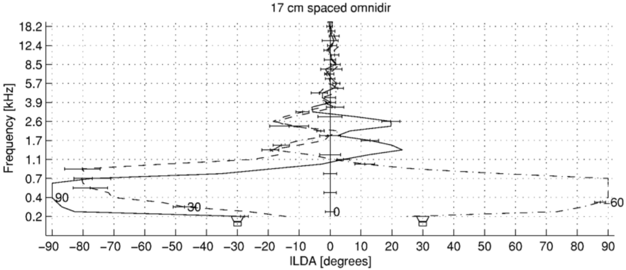


Spaced omnidirectional microphones (AB technique)



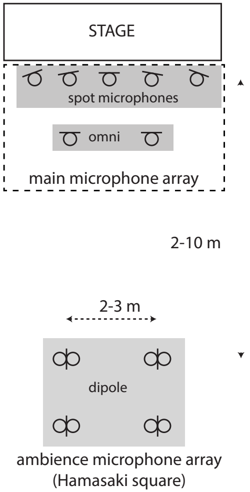




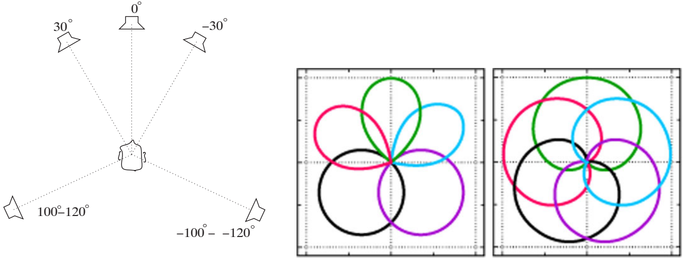


Spot microphone recording

- Multiple sources, e.g., an orchestra on stage
- A "spot" microphone near each source, optimally capturing only single source signal
- Spot microphones are mixed together
- Often far-away "ambience" signals are also recorded with far-away microphones, and mixed with spot microphone signals



Microphone techniques for multichannel

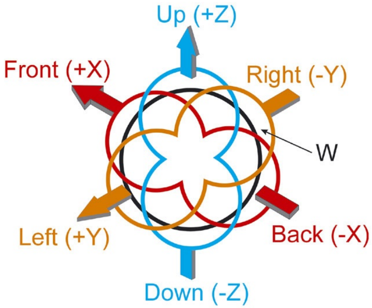


- Center: Ideal microphone patterns for 5.1 loudspeaker setup
- Right: First-order directional patterns
- Too broad patterns cause loudspeaker signals to be coherent
- Comb-filter effects, "muffled" sound, stereo image blurred

B-format recording

- B-format microphones
- Omni + 3 dipoles on Cartesian axis
- Steerable first-order microphone
- Cardioid or hypercardioid for each loudspeaker

B-format recording



www.soundfield.com

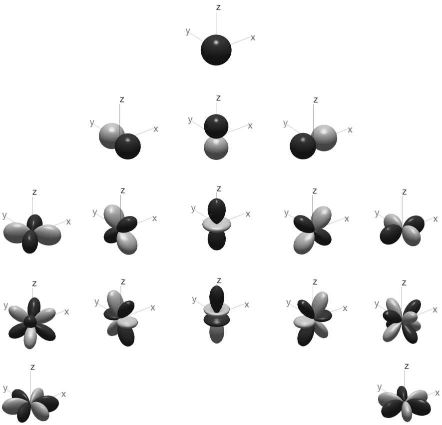
First-order Ambisonics

[Gerzon 70's]

- A signal for each loudspeaker is decoded from B-format
- Loudspeaker channels are relatively coherent
- Coloring
- OK quality in best listening position, and in good listening room
- Nearmost loudspeaker dominates outside best listening position

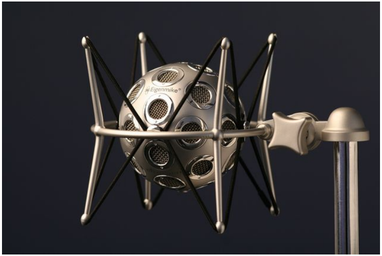
Microphone techniques for multichannel

- Higher-order directional patterns would potentially solve the problem
- Narrow patterns could be composed by combining higher-order patterns
- Higher-order Ambisonics



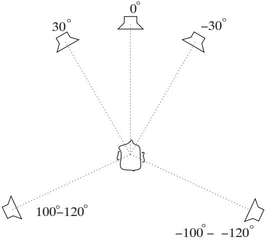
Higher-order microphones

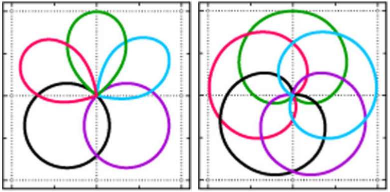
- Requires tens of microphones
- Serious noise problems at low frequencies in decoded spherical harmonics
- Serious problems at frequencies above spatial aliasing frequency



http://www.mhacoustics.com

Microphone techniques for multichannel

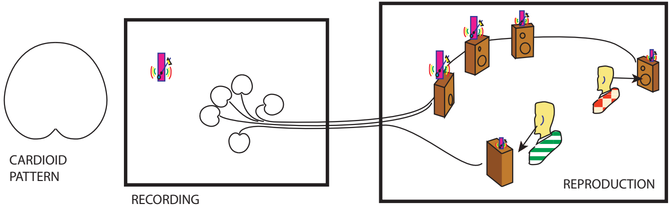




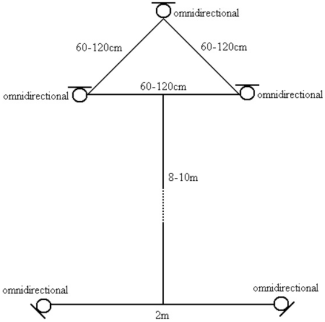
- Center: Ideal microphone patterns for 5.1 loudspeaker setup
- Right: First-order directional patterns
- Too broad patterns cause loudspeaker signals to be coherent
- Comb-filter effects, "muffled" sound, stereo image blurred

Spaced microphone techniques for multichannel

- A set of [usually first-order] directive microphones in some layout
- Large enough spacing to avoid too high coherence btw loudspeaker channels
- Directional patterns provide some kind of reproduction of source directions
- Trade-offs, no generic solution

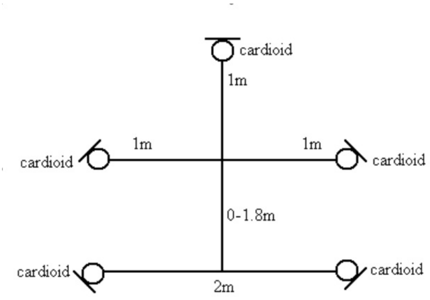


Spaced microphone arrays for multichannel



Decca tree

Spaced microphone arrays for multichannel

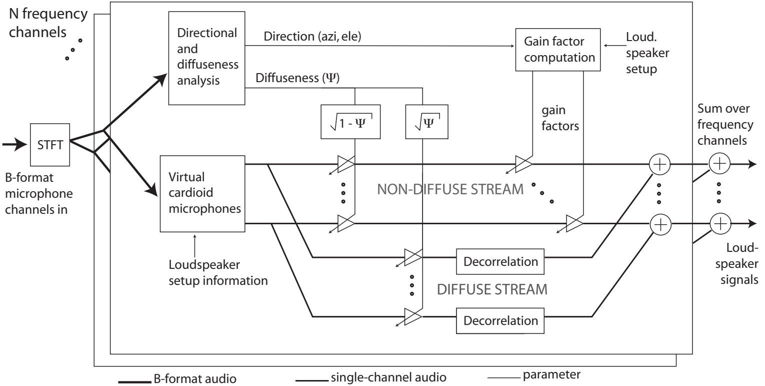


Fukada tree

Non-linear time-frequency-domain reproduction

- Assumption: human perceives only one direction at one time at one frequency channel
- Build system that analyzes sound direction from coincident recording in time-frequency domain, and
- utilizes the analyzed direction to route sound to correct directions
- Directional audio coding, Harpex
- Non-linear signal-dependent spatial-sound-field-dependent techniques
- Enhance quality in most acoustic situations
- Very challenging acoustic conditions cause artifacts

Non-linear time-frequency-domain reproduction

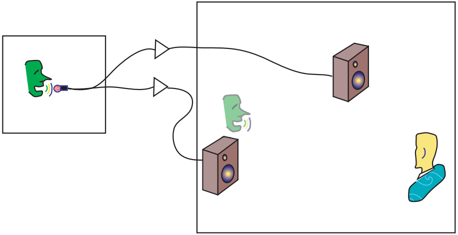


Virtual source positioning

- Input: N monophonic signals
- Output: loudspeaker or headphone signals
- Process each monophonic signal in such a way, that the desired direction is perceived for corresponding virtual source

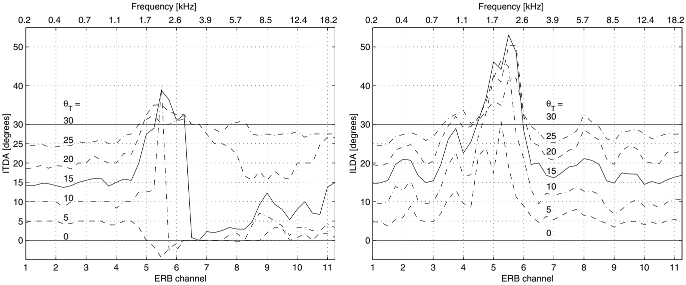
Amplitude panning

- Panpot in mixers: most used virtual source positioning technique
- Equivalent to coincident microphone techniques



Amplitude panning

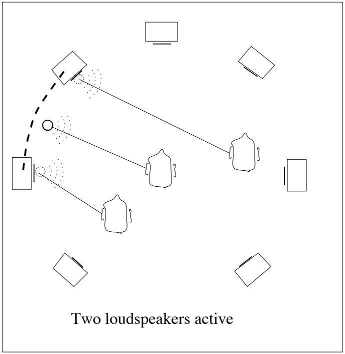
- loudspeaker amplitude difference changes to interaural time difference at low frequencies
- loudspeaker amplitude difference changes to interaural level difference at high frequencies
- does not color sound in any position, although directional effect may be lost

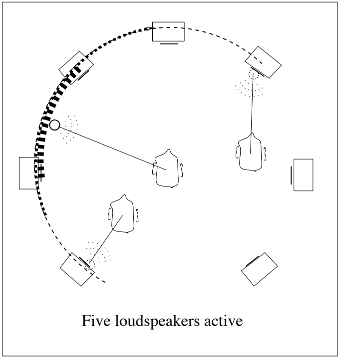


2D panning

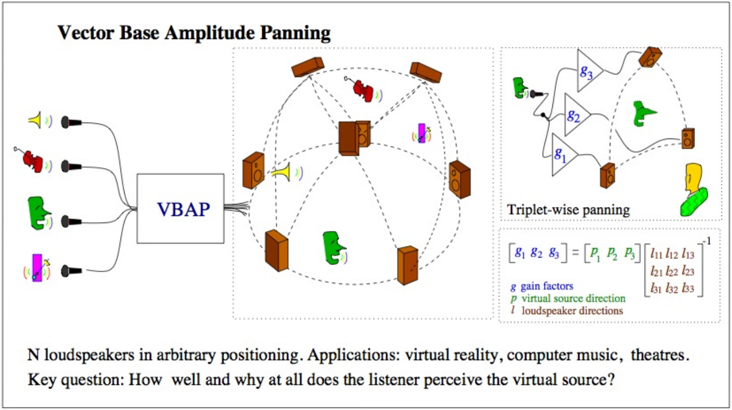
- Pairwise panning

Matrixing





3D Amplitude panning



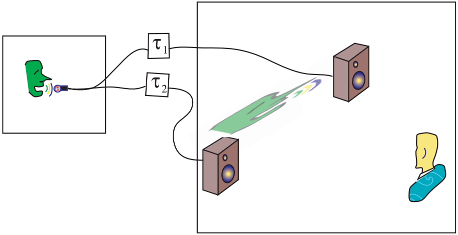
PhD of Ville Pulkki (1995-2001)

Products with "VBAP inside"

- ITU MPEG-H audio standard (broadcast)
- DTS:X audio format (cinema + blueray)
- Sony Playstation VR (gaming)
- Dedicated audio programming softwares

Time delay panning

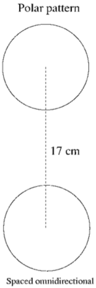
- Used mostly as an effect; creates a spatially spread virtual source
- Equivalent to stereophonic spaced-microphone techniques

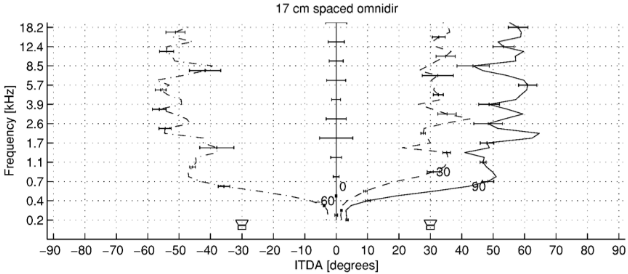


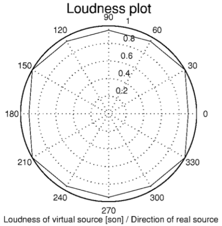
Time delay panning

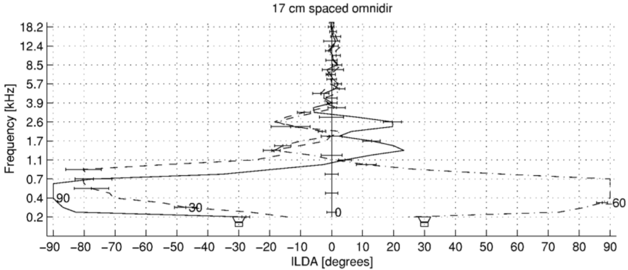
- loudspeaker time delay changes to frequency-dependent interaural level difference at low frequencies
- loudspeaker time delay changes to interaural phase difference at high frequencies
- virtual sources with harmonic spectrum are localized to different directions depending on frequency

Spaced omnidirectional microphones (AB technique)



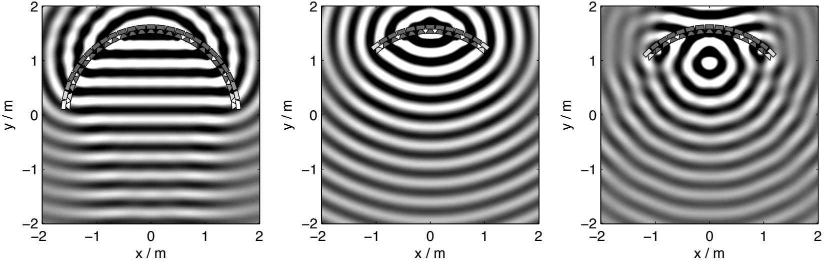






Wave field synthesis

- Try to control the complete wave field
- Helmholtz-Kirchhoff integral
- Can position virtual sources also closer than the loudspeakers are



Wave field synthesis

- Hundreds of loudspeakers needed for 2D loudspeaker setups
- Hundreds of thousands of loudspeakers would be needed for 3D setups
- Not practical as recording technique, possible as virtual source positioning technique
- Spatial aliasing occurs typically near 1kHz, depending on spacing between loudspeakers
- Applications: large venues and installations
- Sound field control, silent and loud zones, noise suppression

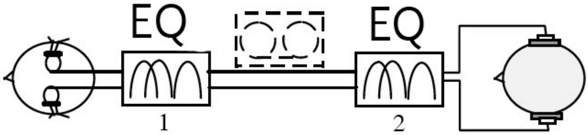
Binaural techniques

- Ear canal signals are the main input to hearing
- Why not replicate only them?
- Recording/reproduction/synthesis of ear canal signals
- Challenges: dynamic cues (head movements), tactile perception

Binaural recording, headphone playback

- careful microphone and headphone equalization
- binaural cues and auditory spectrum reproduced as were in recording
- in some cases this is appealing solution

Applications: personalized recording, academic use, noise measurements, augmented reality audio

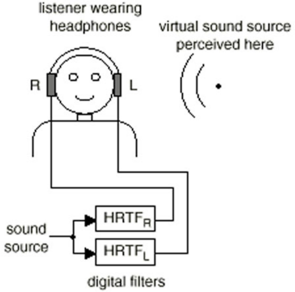


Binaural recording

Challenges

- headphone equalization is problematic
- listener head movements does change binaural cues inside-head localization
- front-to-back confusions
- vision conflicts with audition
- works best only with recordings made with your own head

Binaural synthesis, headphones

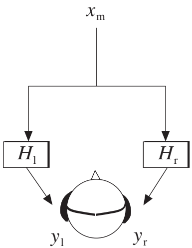


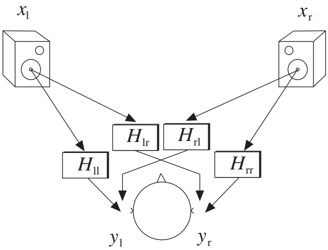
©Bill Gardner

- convolve monophonic sound tracks with measured [individual] HRTFs
- auditory objects can be positioned in 3D virtual space
- inside-head localization, front-back confusions
- need of individual HRTFs
- head tracking may be used to resolve this
- virtual reality, gaming, aviation
- playback of surround audio content over multiple virtual loudspeakers

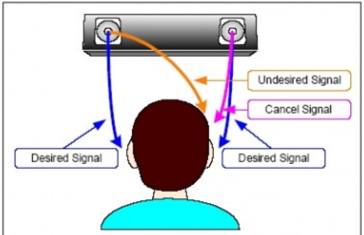
Binaural recording, loudspeaker playback

- Left loudspeaker sound signal reaches also right ear, and vice versa
- "Cross-talk" is a problem
- Could cross-talk be avoided?





Binaural recording, cross-talk cancelled playback



Concept image of Transaural Processing

©JVC

- head has to be placed with about 1cm accuracy
- reflections should not exist
- applicable in some special cases
- back-to-front confusions

References

These slides follow corresponding chapter in: Pulkki, V. and Karjalainen, M. Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics. John Wiley & Sons, 2015, *where also a more complete list of references can be found.*