INTERACTIVE DYNAMIC SOUNDFIELD RENDERING WITH VISUAL FEEDBACK

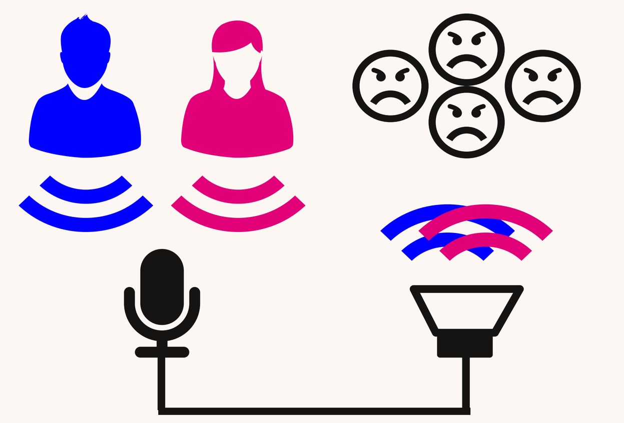
F. Antonacci†, L. Bianchi†, D. Marković†, A. Canclini†, R. Magalotti*, A. Sarti†, S. Tubaro†
† Image and Sound Processing Group, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano
* B&C Speakers, Bagno a Ripoli (FI), Italia

Abstract

In this demonstration we show a system for the rendering of directional sources based on a loudspeaker array. The rendering methodology is based on the idea of the decomposition of the sound field in plane waves. The loudspeaker array is divided into overlapping sub-arrays, each generating a plane-wave component. Individual plane waves are weighed by the desired directivity pattern. We demonstrate the capability of the rendering system in an acoustic scene consisting of two sources (a male and a female speaker) partially occluded by walls. A tracking and visualizazion system based on a Kinect camera and a 2D display is used to provide a visual feedback of the scene from the viewpoint of the listener. The Kinect device is also used as a Computer-Human Interface to modify the position of the walls with gestures of the hands. As a result, the listener can freely move in the scene to appreciate the spatial audio effect.

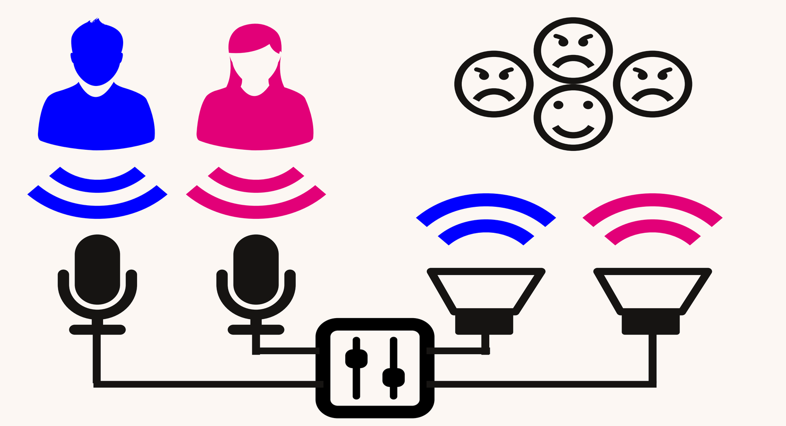
1. Motivation

Conventional mono



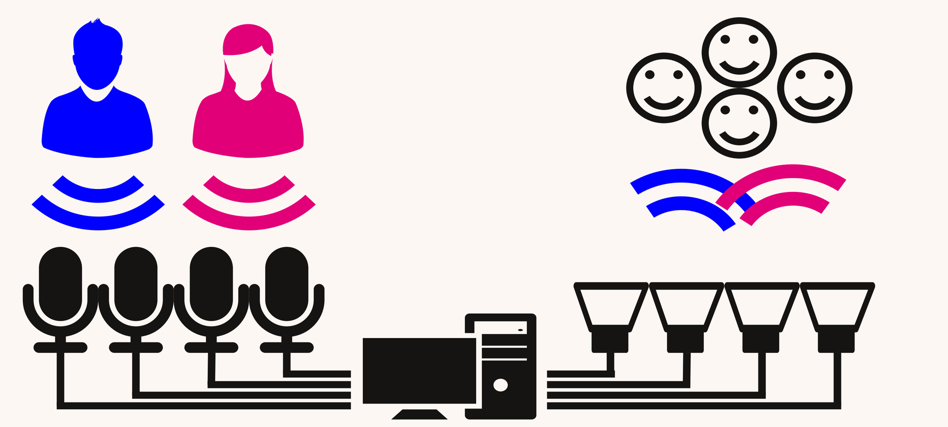
- Listeners can not understand who says what

Conventional stereo panning



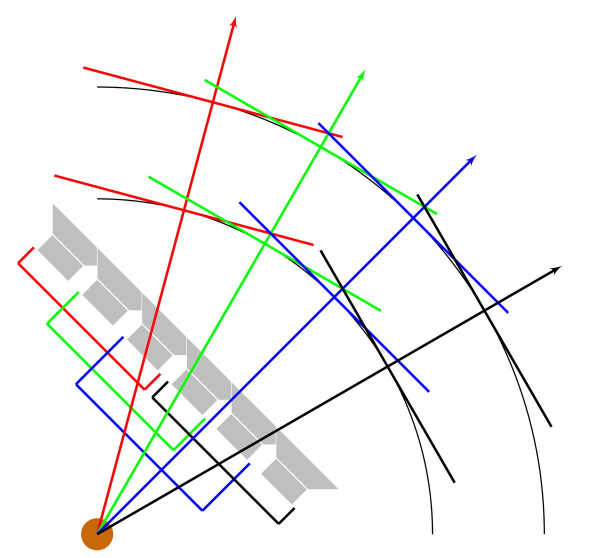
- Only the listener in the focal point can understand who says what

Soundfield rendering



- All listeners can localize each source and hence understand who says what

2. Soundfield Rendering based on Plane-Wave Decomposition

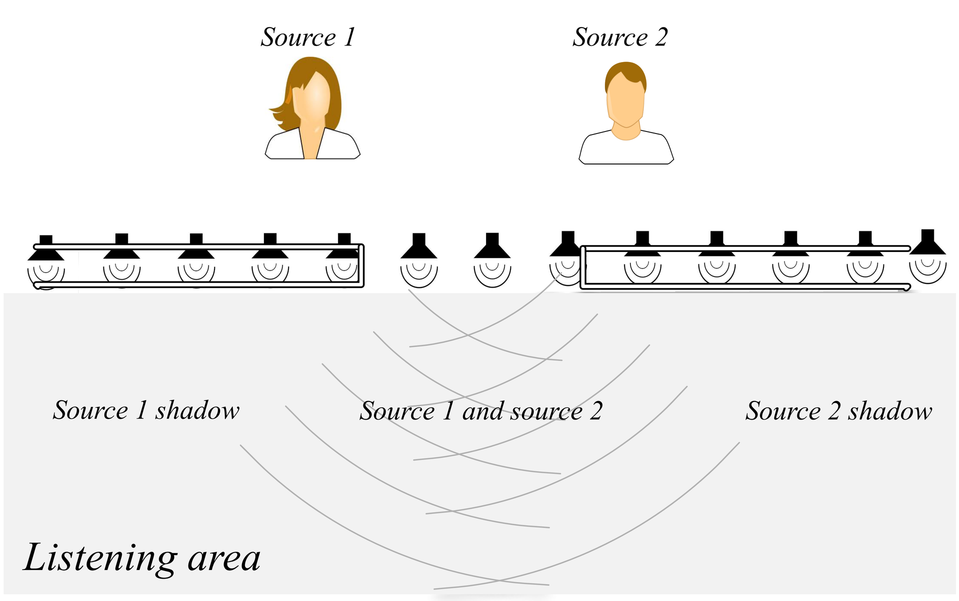


- Arbitrary sound field can be synthesized as the superposition of a finite number of plane waves components, each generated by a loudspeaker array

$$p(\mathbf{x}, \omega) \approx \sum_{i=1}^{M-W+1} d_i(\omega) e^{-j\frac{\omega}{c}\hat{\mathbf{k}}_i^T \mathbf{x}}$$

→ Uniform linear array of $M = 28$ elements subdivided into $M - W + 1 = 22$ sub-arrays of $W = 7$ elements
→ A beamformer is used to generate plane wave components
→ Directivity $d_i(\omega)$ weighs each plane wave component according to visibility and directivity pattern

3. Acoustic Scene

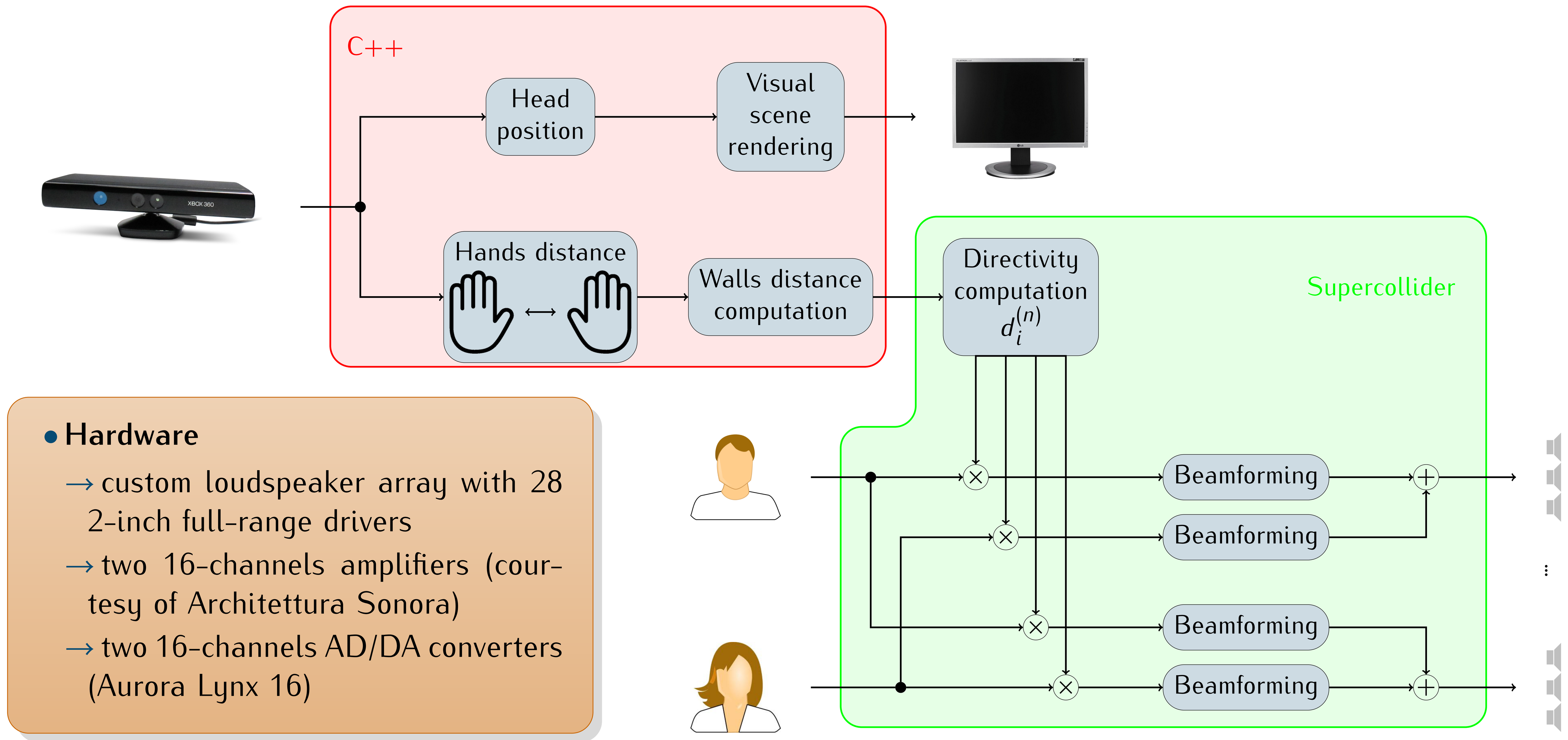


- Two sources behind two parallel walls that form an aperture in front of the listening area
- Three zones in the near field listening area
  → a region where only source 2 is audible
  → a region where both sources are audible
  → a region where only source 1 is audible
- Sources characterized by a cardioid directivity pattern, oriented towards the listening area
- Directivity functions of the nth source are related to the cardioid pattern $c(\omega)$ though

$$d_i^{(n)}(\omega) = \begin{cases} c(\omega) & \text{in the audible area} \\ \frac{1}{1000} c(\omega) & \text{in the shadow area} \end{cases}$$

→ actual average attenuation of sources in the shadow area has been measured to be approximately 25 dB

4. Soundfield display, visual interaction and feedback



- **Hardware**
  → custom loudspeaker array with 28 2-inch full-range drivers
  → two 16-channels amplifiers (courtesy of Architettura Sonora)
  → two 16-channels AD/DA converters (Aurora Lynx 16)

C++ application: capture Kinect signals and render the visual scene

- tracks head position of the listener and shows the visual scene from the listener's viewpoint
- captures hands gestures to control walls distance
- walls distance is sent to the audio rendering application through OSC messages

Supercollider application: perform soundfield rendering

- reads walls distance from OSC messages
- compute the directivity function $d_i^{(n)}$
- compute beamformer filters for each sub-array and weighs its contribution by the directivity
- synchronizes the wavefronts and merges all the contributions coming from all sub-arrays that insist on each loudspeaker
- output signals are sent to the loudspeaker array

References

[1] J. Ahrens and S. Spors, "Rendering of virtual sound sources with arbitrary directivity in higher order ambisonics," in *Audio Engineering Society Convention 123*, 2007.
[2] L. Bianchi, F. Antonacci, A. Canclini, A. Sarti, and S. Tubaro, "Localization of Virtual Acoustic Sources Based on the Hough Transform for Sound Field Rendering Applications," in *Proc. of ICASSP 2013, Int. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, CA, May 2013.
[3] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
[4] E. Corteel, "Synthesis of directional sources using wave field synthesis, possibilities, and limitations," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Feb. 2007.
[5] K. Kumatani, L. Lu, J. McDonough, A. Ghoshal, and D. Klakow, "Maximum negentropy beamforming with superdirectivity," in *European Signal Processing Conference (EUSIPCO)*, Aalborg, DK, Aug. 2010.
[6] E. Verheijen, *Sound Reproduction by Wave Field Synthesis*, Ph.D. thesis, Delft University of Technology, 1998.

ICASSP 2014
IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING