Demonstrator for Controllable Focused Sound Source Reproduction

Alois Sontacchi, Markus Guldenschuh, Thomas Musil, and Franz Zotter

Abstract—**The implementation of a prototype to establish a controllable sound field within a restricted area utilizing distributed loudspeakers is presented. Based on the near field beam-forming approach a demonstrator setup has been developed and implemented. The proposed solution should be a primary step towards providing a convincing alternative instead of wearing headphones which are inconvenient over long time periods for instance at a controller working position (CWP) in airport traffic control towers. Therefore this invention might improve the working process and therefore positively influence the safety conditions at all.**

The proposed setup consists of an off-the-shelf computer, a multi-channel audio converter, a flexible array of tiny loudspeakers and a tracking device. On the computer the sound field rendering process is realized with the programming language Pure Data (pd). Due to tracking data the sound field rendering process is able to trace the position of the listener in space and to calculate the required loudspeaker signals.

Within this article the principle of function is briefly revisited and the required setup components as well as the implementation strategies are treaded.

Index Terms— **Acoustic beam focusing, sound field, focus**

I. Introduction

COMMUNICATION between aircraft pilots and controllers play one of the major roles within the air traffic management. Therefore reliability and security of the communication line are of greatest importance. Beside limited and degraded communication channels both communication participants have to handle with their tasks in rather loud environments.

With the proposed demonstrator we try to pay attention to the current controller working conditions. It provides an alternative communication interface instead of wearing headphones. Headphones exhibit demanding properties in long term use. Unpleasant effects like heating-up and compression of the user’s ears show up eminently in the case of circumaural headphones, but even in the case of supra-aural ones, too. Other types like earbuds (also called earphones) and in-ear monitors that do not cause the above mentioned drawbacks suffer form booming effects within the head.

Manuscript received September 15, 2008. This work was supported in part by the Eurocontrol under Research Grant Scheme – Graz, (08-120918-C).

All authors are with the Institute of Electronic Music and Acoustics, University for Music and Dramatic Arts Graz, 8010 AUSTRIA (corresponding author to provide phone: A. Sontacchi; ++43/316-389-3548; fax: ++43/316-389-3171; e-mail: sontacchi@iem.at).

This article explores the implementation of an innovative communication interface based on loudspeaker reproduction. Details concerning the treaded components of the various approaches can be found in an accompanying paper (see [1]). Within the next section the motivation and the aims of the proposed demonstrator will be presented. In section 3 the general concept of the proposed communication device is introduced. The set-up of the demonstrator as well as the different aspects concerning the implementation will be treated in section 4.

II. Motivation & Aims

In order to provide the controllers at the CWP an appropriate freedom and naturalness, a sophisticated sound reproduction system is requested. This novel audio system should claim for an isolated sound reproduction at the different working positions, whereby it is able to track and follow the position of a defined controller at the CWP. Furthermore controller and co-controller should have the possibility to work within there own isolated communication zone but also should have access to a shared area where information exchange with the partner is possible without losing the contact to the own zone. The vision of two crossing so called “sound bubbles” is depicted in figure 1.

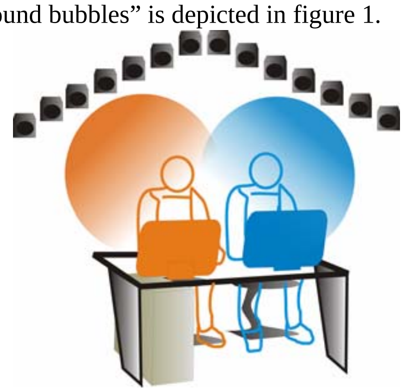

Fig. 1. Schematic depiction of so called “Sound bubbles” - Acoustically isolated/transparent working positions.

Based on first studies in [1], where four different methods have been examined concerning the beam-steering purpose, the most promising method has been selected. A near field beam-forming approach [2] including delay and gain differences is combined with a harmonic nested windowing technique. In addition to that, different shapes of the loudspeaker arrays have been examined (for details see [3]) and to benefit from spatial information crosstalk cancellation and binaural pre-processing have been added to the system. To provide users/listeners more freedom and to increase the naturalism and immersion within the virtual audio scene head tracking is considered, too.

III. Concept of the Demonstrator

Based on the real-time tracking data (i.e. head position and head orientation) the audio reproduction system will produce time varying sound pressure signals in the vicinity of the tracked user-ears. Hence, caused by the geometrical arrangement of the loudspeaker array and e.g. the video based tracking system, as well as the optical information display at the CWP the target user space will be limited. The resulting scenario for a specific configuration is depicted in figure 2. In the presented configuration the curved loudspeaker array line is placed in front of the user and the video camera is mounted on top of the CWP Display.

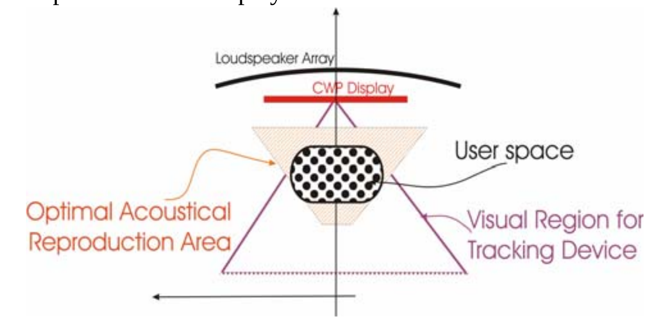

Fig. 2. Schematic representation of the overall structure.

A. Tracking Device

Different commercial tracking devices are available on the market to provide head tracking. The mechanical approach, where the tracking system is mounted on the body of the tracked head and each head movement causes an alteration of a resistor (potentiometer) is not very practical due to the reduced flexibility of the user. In addition to the head orientation the position of the head related to a defined point in space is required, too. Therefore electromagnetic based systems like „Isotrak“ and „Fastrack“[1], or “Flock of Birds”[2] might be better suited. Beyond that, acoustical systems based on ultra sound waves (e.g. Logitech 3D Head Tracker[3]), optical systems, indoor GPS tracking, or tracker devices utilizing the inertia system can be used. However, sensors mounted on the users head are required. This necessity can be realized by combining the required target sensor with the head mounted microphone. Alternatively, if neither a head mounted microphone nor any sensor should be worn by the user, video based tracking (e.g. eye-ball tracking) in combination with a microphone array can be used.

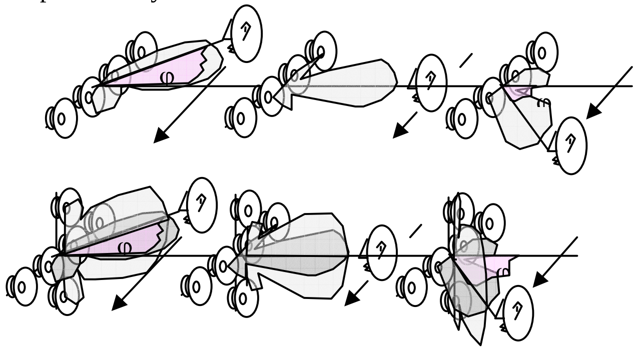

Fig. 3. Achievable schematic acoustic beams with horizontal line array (upper row), and with additional vertical arranged loudspeakers (lower row).

B. Acoustic Focusing

Based on near field beam-forming focused sound spots can be achieved. If loudspeakers with the same orientation are arranged along one line, the produced sound field will exhibit a beaming within the plane where the loudspeakers are oriented and a spreading within the orthogonal plane. To prevent the excitation of the diffuse sound field additional loudspeakers orthogonal to the line array (see Fig. 3) or even such with higher directivity in this direction can be used. Isolated sound babbles can be realized by taking these aspects into account.

To improve the beam-forming procedure described in [3] we add spatial information at the focus spots in the vicinity of the user’s ear entrance (see Fig. 4). Therefore the mono signal from the radio communication channel is converted into a binaural signal, whereby the perceived target direction corresponds with the optical information display. In order to provide dynamic listening cues the binaural cues are adapted to the different head orientations. This procedure can be realized very efficiently by the usage of the Ambisonics approach (cf. [4]).

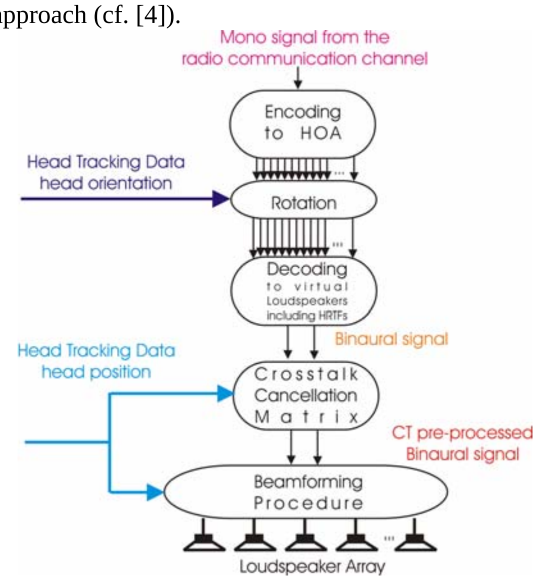

Fig. 4. Schematic representation of the overall structure

[1] http://www.polhemus.com; copyright 2005 Polhemus. 40 Hercules Drive, P.O. Box 560, Colchester, VT 05446, USA

[2] http://www.ascension-tech.com; copyright 2005 Ascension Technology Corporation. P.O. Box 527, Burlington, VT 05402, USA.

[3] www.logitech.com

Furthermore, caused by the fact, that two foci are realized in closed vicinity the crosstalk between these two spots has to be reduced (for details see [3], too).

IV. Test Set-Up

In the following section the set-up of the demonstrator will be treated. A basic overview of the involved hardware equipment is presented in Fig. 5. The demonstrator consists of the following components:

- Sixteen separate driven loudspeakers combined to a specific arrangement (loudspeaker array).
- One 16-channel amplifier
- A PC with installed pure data (pd)
- A PC multi-channel audio interface (AD- & DA-converter)
- Head Tracking Device (e.g. webcam)

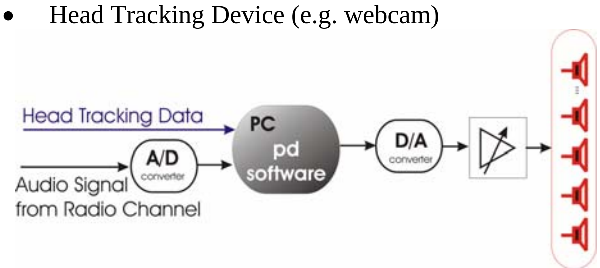

Fig. 5. Block diagram of the hardware setup.

A. Arrangement of the Loudspeakers

In order to get the most suitable shape for the speaker arrangement a simulation tool was developed in MATLAB. This tool calculates direct sound pressure level of the loudspeaker array driven by beam-steering methods and windowing techniques applied to harmonic input signals. The directivities of the simulated speakers are assumed to be omni-directional. The simulation shows the magnitude of the sound pressure field and directivity patterns (screen-shot, see Fig. 6).

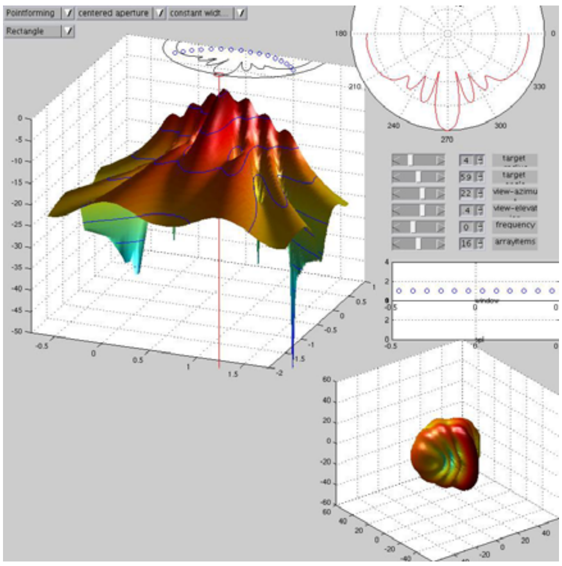

Fig. 6. Near field beam-forming with applied windowing technique for a circularly array segment consisting of 16 speakers at an interval of 0.09 [m]. (Frequency used for depiction: 1000 [Hz]).

B. Software – control parameters

For the implementation of the required audio algorithms the software Pure Data ("pd", Miller S. Puckette[4]) is used. The developed software provides the possibility to adjust different array shapes with a fixed number of loudspeakers. Different window-shapes and the window-sliding can be applied to the weighting process (harmonic nesting, see [3]). Furthermore it provides the inclusion of the head tracking data and realizes the binaural pre-processing as well as crosstalk-cancellation.

In addition to that, the overall amplification gain, and equalizing the non-idealized frequency responses of the loudspeakers to avoid timbre colorations is realized. The maximum number of speakers in the current presented implementation is restricted to 16, so far.

C. Signal processing

A full description of the virtual Ambisonics approach for the binaural pre-processing can be found in [4]. This approach provides the possibility to realize the calculation of time variant binaural cues by using time invariant filters (head related transfer functions, HRTFs) in combination with easily rotatable weighted basis functions of the decomposed sound field.

The crosstalk cancellation matrix can be easily determined by visiting the sketch in Fig.7. Regarding the received signal at the user's left ear *L(k)* it can be described in the frequency domain by multiplying the sum of transfer functions from F_1 to F_N with the transmitted signal spectrum S_{Left}, (cf. Eq.1). Whereby the individual transfer functions F_i are determined by calculating the complex gains with the near field beam-forming approach. These gains are based on the head (i.e. the user's ears) position data, obtained by the head tracking system.

$$L(k) = S_{Left}(k) \cdot \sum_{i=1}^{N} F_i(k) = S_{Left}(k) \cdot H_1(k) \quad (1)$$

However, applying a specific signal for the left ear with the given loudspeaker arrangement the right ear *R(k)* will obtain some modified copy of it (cf. Eq.2).

$$R(k) = S_{Left}(k) \cdot \sum_{i=1}^{N} \hat{F}_i(k) = S_{Left}(k) \cdot H_2(k) \quad (2)$$

To prevent the observed crosstalk a cancellation matrix (**CC**, see Eq.3) has to be applied.

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} H_1 & H_2 \\ H_2 & H_1 \end{bmatrix} \cdot \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} L_{Binaural} \\ R_{Binaural} \end{bmatrix} = \mathbf{H} \cdot \mathbf{CC} \cdot \begin{bmatrix} L_{Binaural} \\ R_{Binaural} \end{bmatrix} \quad (3)$$

The crosstalk is removed, if the product of the matrix H and **CC** in Equation 3 yields the identity matrix I. After some calculations we obtain the elements of the crosstalk-cancellation matrix **CC** (see Eq.4).

$$\mathbf{CC} = \frac{1}{H_1^2 - H_2^2} \begin{bmatrix} H_1 & -H_2 \\ -H_2 & H_1 \end{bmatrix} \quad (4)$$

[4] Details about pd and download links can be found at: http://crca.ucsd.edu/~msp/

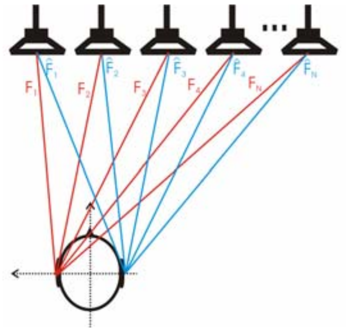

Fig. 7. Sketch of the various acoustical transmission paths (F_i and $\hat{F}_i$) from the individual loudspeakers to the user's ears.

The beam-forming processing consists simply of a delay and a FIR filter for each speaker channel (see Fig. 8). In Figure 9a and 9b the determination of the gain and delay distribution is shown for an example of a line array consisting of 5 loudspeakers.

The AM radio channel used in airplane radio has a very small bandwidth of only 3.4 kHz, thus filtering can be done at low sampling rates. Our FIR filters are FFT based (see Fig. 11) and use filter lengths of 64 samples at a sampling rate of 11025 Hz.

Filter coefficients are determined by the window function applied to the speaker array. Due to harmonic nesting and a window center that slides along the array axis, window weights need to be calculated for each steering position and frequency bin. The (high pass) frequency response compensation is multiplied to the filter responses, so we need only one filter per loudspeaker. The windows are determined by their off-center shift and the desired beam width (see Fig. 10).

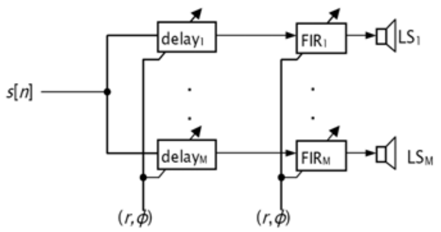

Fig. 8. Block diagram of the beam-forming procedure for one sound spot.

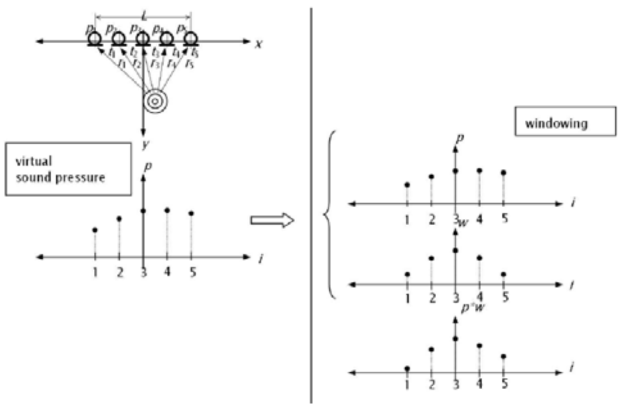

Fig. 9a. Determination schema of loudspeaker feeds in case of a linear array consisting of five speakers. Gain calculation.

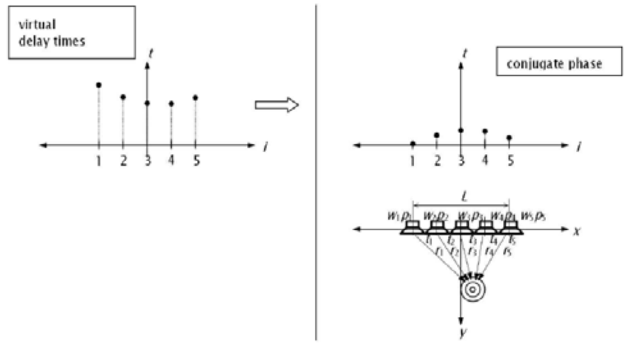

Fig. 9b. Determination schema of loudspeaker feeds in case of a linear array consisting of five speakers. Delay time estimation.

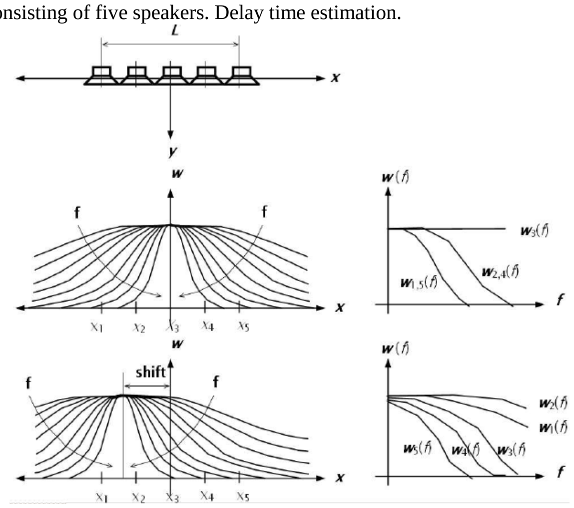

Fig. 10. Depiction of applied harmonic nesting (second row) and the sliding-window effect (at the bottom).

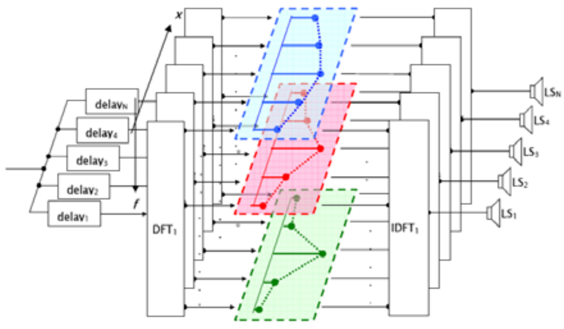

Fig. 11. Schema of the FFT based block signal processing for each beam.

V. Conclusion

Within this article the implementation of a demonstrator for reproduction of controllable focused sound sources with loudspeakers in a defined area has been presented. Based on the near field beam-forming approach additional spatial information is introduced. Besides tracking the head position, the head orientation is tracked, too. Therefore more naturalism and flexibility can be achieved by the adapted binaural cues. In order to stabilize and improve the reproduced cues crosstalk cancellation is considered.

Future work will be directed to handle two interactive "sound bubbles", to test various proper tracking devices and to improve the sound focusing quality by extending the acoustical path i.e. considering real reflections, too.

REFERENCES

[1] F. Zotter, “Studies on control of spatial sound field concentration”, IEM-Report, November 2003.
[2] E. Williams, “Fourier acoustics“, London, Academic Press 1990.
[3] M. Guldenschuh et al., “Prinicples and Considerations to Controllable Focused Sound Source Reproduction”, submitted to the 7th EUROCONTROL Innovative Research Workshops and Exhibition, Eurocontrol Experimental Centre, Bretigny sur Orge, December 2-4, 2008.
[4] M. Noisternig, T. Musil, A. Sontacchi, and R. Hoeldrich, *"3D Binaural Sound Reproduction using a Virtual Ambisonics Approach"*, Proc. Int. Symposium on Virtual Environments, Human-Computer Interface, and Meas. Sys. (VECIMS), Lugano, Switzerland, July, 2003.