Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Circumaural transducer arrays for binaural synthesis

Raphaël Greff[a] and Brian Katz[b]

[a]A-Volute, 4120 route de Tournai, 59500 Douai, France
[b]LIMSI-CNRS, B.P. 133, 91403 Orsay, France
raphael.greff@a-volute.com

Binaural cues such as the interaural time and level differences are the primary cues for estimation of the lateral position of a sound source, but are not sufficient to determine elevation or the exact position on the "cone of confusion". Spectral content of the head-related transfer function (HRTF) provides cues that permit this discrimination, notably high frequency peaks and notches created by diffraction effects within the pinnae. For high quality binaural synthesis, HRTFs need to be individualized, matching the morphology of the listener. Typical means for this are to measure or calculate the HRTF of the listener, but these lengthy and costly methods are not feasible for general public applications. This paper presents a novel approach for HRTF individualization, separating the head and torso effect from that of the pinnae. The head/torso component is numerically modeled while the pinnae component is created using a multiple transducer array placed around each pinna. The philosophy of this method consists in trying to excite the correct localization cues provided by the diffraction of the reconstructed wave front on the listener's own pinnae. Simulations of HRTF reconstruction with various array sizes and preliminary auditory localization tests are presented.

1 Introduction

High quality binaural synthesis requires individualized HRTF, *i.e.* matching the morphology of the listener. Measuring or calculating the HRTF of a listener are common but lengthy and costly methods that are not feasible for general public applications. In order to simplify the estimation procedure of individualized HRTF, a great deal of research efforts aim to model the relation between morphology and the HRTF and provide a fast individualization based on anthropometric measurements [1,2]. However, accurate estimations of morphological parameters may only be guaranteed using accurate measurement tools, essentially for estimating pinnae dimensions.

In this paper, a novel approach for HRTF individualization is proposed, by separating the head and torso effect from that of the pinnae using a multiple transducer array placed around each pinna. The goal is to reduce the individualization of the HRTF to the modeling of the head/torso component, and to try to excite the correct localization cues provided by the diffraction of the reconstructed wave front on listener's own pinnae.

2 Binaural holophony

The concept of binaural holophony consists in using holophonic techniques to reconstruct the sound fields in each pinna area. In practice, this should be done by using circum-aural transducer arrays mounted on a headset and by feeding them correctly in order to reconstruct the desired binaural signal at listener's eardrums.

To characterize array signals, one needs first to understand the theory of general holophony and see how this is used for reproducing sound fields.

2.1 Kirchhoff-Helmholtz integral

The basic principle of holophony is based on the Huygens's principle. Huygens stated that a wave front of a propagating wave at any instant conforms to the envelopes of spherical waves emanating from every point on the wavefront at the prior instant. The mathematical foundation of this illustrative description is given by the Kirchhoff-Helmholtz integral [4] which expresses that the acoustical pressure p within a source-free volume V is derived from the knowledge of the acoustical pressure p_0 and its gradient over the boundary S enclosing the considered volume. $\forall \vec{r} \in \Omega$:

$$p(\vec{r}) = \iint_S \left[\vec{\nabla} p_0 \cdot \vec{n} - \frac{\vec{R}}{R} \cdot \vec{n}(1+jkR)\frac{p_0}{R} \right] \frac{e^{-jkR}}{4\pi R} dS_0 \quad (1)$$

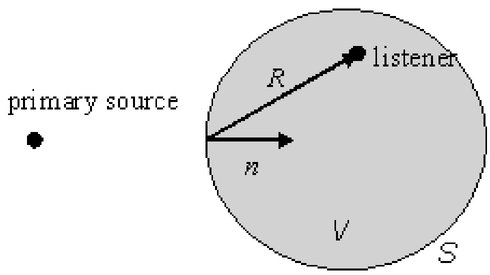


Fig.1 Parameters used in the expression of the Kirchhoff-Helmholtz integral.

2.2 Spatial sampling

Though the Kirchhoff-Helmholtz integral provides a very attractive solution for sound field reproduction, in practice only discontinuous array of secondary sources are available, which raises the problem of spatial sampling. Discrete arrays can not correctly sample incident waves whose wavelength is small compared to the transducer spacing Δ. Such spatial aliasing occurs above the so-called "spatial aliasing frequency" (c is the speed of sound):

$$f_a = \frac{c}{2\Delta} \quad (2)$$

In practical situations, the integral in Eq.(1) should be replaced by a summation. This gives a straightforward way of reproducing a sound field. At the recording stage, the listening area is surrounded by a microphone array composed by pressure and velocity microphones which record the primary sound field due to external sources. For the reproduction stage, loudspeakers are substituted for the microphones by replacing pressure microphones by dipole sources and velocity microphones by monopole sources. Each loudspeaker is fed with the signal recorded by its associated microphones (assuming the use of some weighting factors due to the spatial distribution of the secondary sources).

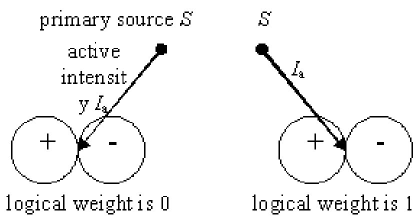
2.3 Single directivity array

It has been remarked that the two elementary transducers of the secondary source are highly redundant, so that only one of them is necessary [3,4]. In practice, only monopole loudspeakers are used and fed by figure-of-eight or even cardioid microphones. Figure-of-eight solution is an exact solution if one is able to record the contribution of each primary source separately. The recorded contribution of a primary source is taken into account only if this source is viewed by the positive pole of the microphone. For planar array, this selection is done by construction [4], but for complex array geometries, the logical weighting can only be performed if the exact location of each primary source is known. For this reason cardioid microphones are often preferred when recording real complex sound fields using complex array geometry.

Nevertheless, the location of a sound source relative to an arbitrary field point is given by the active intensity at this point. Therefore, we may try to generalize the logical weighting - when using figure-of-eight microphones - to any arbitrary primary sound field, *i.e.* created by unknown sources.

In the next section, three kinds of holophonic recording solutions are employed for investigating the concept of binaural holophony:

- *CARD*: cardioid recording,
- *FIG-8 S-LOG*: figure-of-eight recording, considering the contribution of each primary sources separately,
- *FIG-8 T-LOG*: figure-of-eight recording, considering the total sound field.



.Fig.2 Illustration of logical weighting.

3 Simulations

The array used here has a hemispherical geometry. Based on Listen and CIPIC anthropometric databases, we evaluated that a radius of 50mm is sufficient to surround all of the pinnae, and 90% of them when centring the array on ear canal entrance. Simulations are only performed with centred positioning. The primary array is composed of 640 equally distributed secondary sources to avoid spatial aliasing in the considered frequency range (3.5 to 16 kHz).

The goal here is to compare the original HRTF to a reconstructed HRTF using holophonic techniques for different array sizes. By introducing two kind of filters, the transducer driving filters (CAHF for Circum-Aural Holophonic Filters) and the transfer functions representing the path from a considered transducer to the entrance of the ear canal (PRTF for near-field Pinnae-Related Transfer Functions), the reconstructed *HRTF* can be expressed as follows (where N is the total number of transducers and i the index of the considered transducer):

$$HRTF_r = \sum_{i=1}^{N} CAHF[i].PRTF[i] \qquad (3)$$

For estimating the quality of reconstruction, two metrics are introduced, correlation and distance. Both are calculated on a logarithmic scale and based on decibel spectral values of the HRTF.

3.1 Baffled pinnae

Simulations have been performed on 6 different baffled pinnae meshes (Cortex, DB60, DB65, DB90, DB95, Yuvi), modelled by Yuvi Kahana [5]. In this case, the three kinds of holophonic recording solutions previously mentioned can be used, because all sources are known (the primary source and its image). CAHF are calculated analytically due to the simplicity of the problem. Original HRTF and PRTF are estimated via BEM techniques, respecting the sampling rule of 6 nodes per wavelength [5,6].

For all pinnae models, results indicate a good match between original HRTF and reconstructed HRTF using CARD and FIG-8 S-LOG solutions (cf. Fig.3). FIG-8 T-LOG is dependant on lateral angle: quality of reconstruction decreases with lateralization of the primary source (cf. Fig.4), because contributions of the primary source and those of the image source tends to oppose.

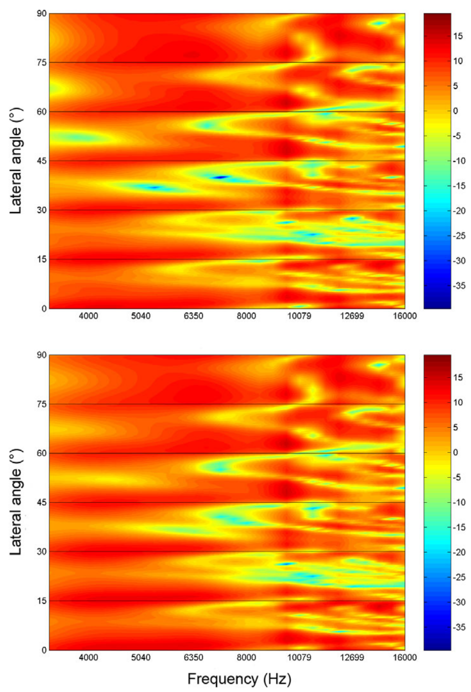


Fig.3 Cortex pinna - original HRTF (upper) and reconstructed HRTF using an array of 640 transducers with FIG-8 S-LOG recording method (lower).

If the number of secondary sources is too small, spatial aliasing occurs which degrades the quality of reconstruction. By decreasing the number of transducers from 640 to 8 (keeping a homogeneous repartition), the correlation decreases from more than 0.9 to less than 0.2

and spectral distances increases from approximately 1 dB to more than 5 dB (FIG-8 S-LOG method).

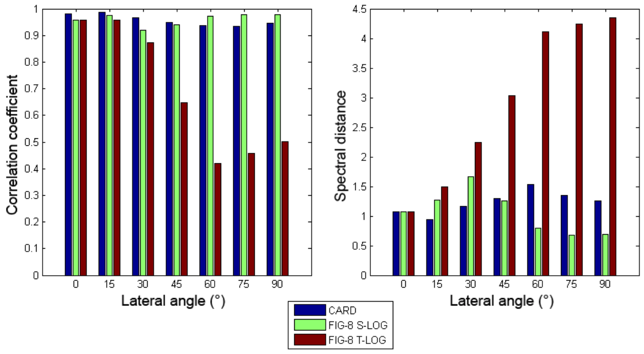


Fig.4 Cortex pinna – correlation coefficient (left) and spectral distance (right) per sagittal plane.

3.2 Head model

Simulations have been done on a model of the Neumann KU-100 dummy head, obtained via a MRI scan [7]. In this case, only CARD and FIG-8 T-LOG solutions have been calculated, because of the complexity of the problem. Moreover, all transfer functions have been estimated via BEM calculations. Because of the size of the model, a sampling of 6 nodes per wavelength is only respected up to 10 kHz, but simulation are still conducted up to 16 kHz.

For the ipsilateral HRTF, results are equivalent to those of the baffled pinnae simulations. For the contralateral HRTF, quality of reconstruction decreases with respect to lateralization for both recording methods, but CARD solution seems to be more robust.

Decreasing the number of secondary sources degrades the reconstruction for both recording method, wherever the primary source is located. For small array sizes (under 20 secondary sources), FIG-8 T-LOG method offers better results than CARD for median plane HRTF, but can not reconstruct the contralateral HRTF properly.

3.3 Equivalent spherical head

BEM calculations dedicated to CAHF estimation are very time-consuming (the current calculation required more than three weeks on a cluster of three dual core processors with a total of 9 GB RAM). So it is interesting to evaluate the degradation of HRTF reconstruction using an equivalent spherical head model which permits an analytical resolution of the problem. A similar approach from that Algazi proposed [8] has been used to find the radius of the iso-ITD sphere relative to the KU-100 dummy head. CAHF have then been calculated for this sphere.

No relevant difference has been noticed compared to KU-100 results. The spherical head approximation seems to offer an easy way to design a binaural holophonic system.

4 Optimizations

It was desired, for practical reason, to arrive at a solution for a binaural holophonic system using a maximum of 16 channels, *i.e.* 8 transducers per pinna. Precedent simulations have shown that raw applications of holophony principles provide bad results when spatial aliasing is occurring. Transducer driving filters and transducer positioning need to be optimized.

4.1 Non pinnae-related optimization

When the number of secondary sources decreases, the area associated to each microphone (or transducer) increases. The pressure (and pressure gradient) recorded at a single point may not be representative of what is happening in the entire region. A simple solution for obtaining a more relevant recording is to average the whole area information.

Applying this approach to the precedent simulations provides a significant enhancement of the reconstruction quality. For example, correlation coefficient increases from less than 0.2 to over 0.5.

4.2 Pinnae-related optimization

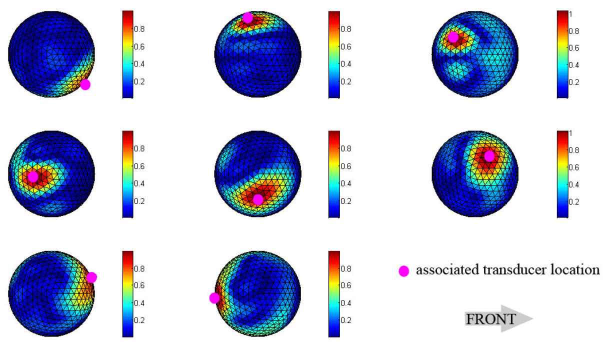


Fig.5 Optimized directivities given by "subset selection" analysis.

A way of optimizing both the location and the associated CAHF of a transducer is to apply statistical analysis on a database of PRTF (corresponding to the 640 secondary source locations of the primary array). Three types of optimizations have been investigated: PCA, ICA and "subset selection" [9]. The PRTF database is composed from that of 7 pinnae (Cortex, DB60, DB65, DB90, DB95, Yuvi and KU-100). The considered number of transducer is 8 per pinnae.

The optimization procedure consists first in separating the PRTF database into a reduced number of filters and directivities. Optimized locations of the transducers are derived from the directivity patterns (directions pointed by each maximum of directivity). As associated filters may not correspond to the PRTF at those locations - for PCA and ICA - the directivity patterns can not be used directly to calculate the optimized CAHF. Projecting whole PRTF database on the PRTF associated to the optimized locations permits one to obtain the desired optimized associated directivities. These directivities are then used as masks to derive the optimized CAHF associated with the optimized array. Naming $CAHF_{prim}$ the driving filters of the primary array of 640 secondary sources and D_{opt} the optimized directivities, the optimized driving filters $CAHF_{opt}$ of the optimized array (of 8 transducers in this case) are obtained as follows:

$$CAHF_{opt}[k] = \sum_{i=1}^{N} D_{opt}[i,k] \cdot CAHF_{prim}[i] \qquad (4)$$

where N is the total number of the primary array, i the index of the secondary source related to the primary array and k the index of the transducer related to the optimized array.

Results indicate that all three methods highly increase the quality of HRTF reconstruction. However, "subset selection" seems to be the best solution (correlation coefficient and spectral distance reach values that exceed 0.9 and fall under 1 dB respectively). Moreover, it is the only optimization that produces directly the optimized directivities (cf. Fig.5): the filters issued from the "subset selection" analysis correspond exactly to the PRTF pointed by each directivity pattern.

5 Subjective evaluations

A modular hand-made headset prototype has been constructed to perform subjective evaluations of different array configurations. A preliminary listening test has been conducted to choose the transducer dimensions. The protocol of this test is not presented here, but results indicated that it is preferable to use transducers as small as possible.

5.1 Configuration comparison

Three array configurations have been tested, composed of 8 transducers per pinnae. In configuration ***A*** transducers are equally distributed and CAHF are calculated using the first optimization (average per area). For configuration ***B***, the associated filters are obtained through the "subset selection" optimization method. Configuration ***C*** is based on psychoacoustic hypothesis: as pinnae spectral cues are predominant cues in the median plane [10,11], transducers are placed at grazing incidence where recorded energy is maximal when primary sources are located on median plane.

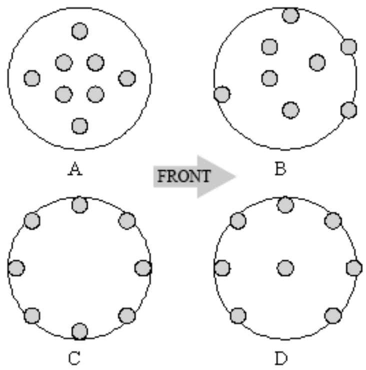


Fig.6 Tested array configurations (schematic representation).

Localization tests were performed only for virtual sources in the median plane. Perceived direction was reported on a visual interface. The stimulus used was a 200ms pink noise burst modulated at 25Hz. 4 subjects participated to the test. Each tested location was repeated 5 times.

Results, reported in Table 1, show that configuration ***C*** offers the best performance. "Subset selection" optimized configuration ***B*** has good simulation results, where an exact positioning of the headset is guaranteed, but compared to configuration ***C*** it seems not to be the best way for exciting pinnae cues when headset placement is approximate. Further simulations must be conducted to evaluate the degradation of reconstruction relative to off-centred positioning of the array. However, localization accuracy is better using ***B*** compared to ***A***, which is coherent with simulations results.

(min./max)	***A***	***B***	***C***
Mean error E (°)	28 (19/39)	23 (17/27)	**14** (11/19)
Standard deviation Std (°)	20 (16/24)	**17** (14/21)	19 (14/25)
Front-back confusions FB (%)	25 (7/38)	34 (11/60)	**20** (13/27)

Table 1 Localization test results for all configurations (***A***, ***B***, ***C***) - median plane.

E (°)	*Std (°)*	*FB (%)*
11 (7/13)	12 (10/13)	20 (11/29)

Table 2 Localization test results for configuration ***C*** – horizontal plane.

Horizontal plane localization tests have then been conducted on retained configuration C. Even if results reported in Table 2 are good, informal listening tests and discussions with tested subjects indicate a strange sensation for lateral source. It is considered that this may be due to the fact that all transducers are placed at grazing incidence so that a lateral wavefront can not be recreated. We decided to test a fourth configuration (***D***) obtained by relocating one transducer in order to improve the coherence of recreated lateral wavefronts. Results (cf. Table 3) are comparable to those of the configuration ***C*** and informal discussions indicate that the strange sensation has disappeared.

(min./max)	*E (°)*	*Std (°)*	*FB (%)*
Median plane	18 (16/21)	17 (14/19)	9 (2/18)
Horizontal plane	11 (7/13)	10 (8/12)	21 (7/31)

Table 3 Localization test results for configuration ***D*** – median and horizontal plane.

5.2 CAHF vs. individualized HRTF

Localization tests have been conducted on 4 other subjects, using the binaural holophonic system (configuration ***D***) and also individualized HRTF. Performances are compared per listener.

Results show that both spatialization systems are comparable (cf. Table 4 and Table 5). In our experiment, individualized binaural accuracy is worth than what is reported in the literature [12], but generally the rate of tested median plane localisation is less important, median plane localization task being the most difficult task.

		Median plane		*Horizontal plane*	
		Binaural holophony	*Personal HRTF*	*Binaural holophony*	*Personal HRTF*
1	*E*	24	20	9	11
	Std	**25**	**44**	10	10
	FB	18	24	13	18
2	*E*	16	15	7	9
	Std	18	16	9	16
	FB	**2**	**16**	**7**	**20**
3	*E*	18	16	13	10
	Std	14	16	10	6
	FB	**38**	**24**	**31**	**13**
4	*E*	30	27	8	13
	Std	**20**	**30**	11	11
	FB	16	20	**22**	**36**
Mean	*E*	22	20	9	11
	Std	19	27	10	11
	FB	19	21	18	22

Table 4 Comparison of binaural holophony and personal HRTF results.

6 Conclusion

Simulations and subjective evaluations show that it is possible to excite pinnae cues using multiple transducers well-placed around listener's pinnae, and using transducer driving filters related only to a simple spherical head model. Comparison of the proposed system to individualized HRTF don't permit to conclude directly that the proposed binaural holophonic system offers localization performances as accurate as individualized HRTF system do, because statistics on 4 subjects may not be representative of the entire population. However, it justifies the use of binaural holophony for general public application as a promising alternative to common HRTF-based binaural systems.

Moreover, the holophonic approach offers a way to decompose the HRTF into head/torso component - CAHF - and pinnae component - PRTF (exact physical decomposition in comparison to Algazi's heuristic signal approach [13]). This allows us to study separately the link between pinnae morphology and associated spectral modifications. Pinnae-related optimizations presented above offer also another point of view for characterizing this relation and open the way of fast HRTF composing.

Acknowledgments

Special thanks to Yuvi Kahana for access to his pinnae models. Funding was provided by the ANRT CIFRE program in collaboration with A-Volute.

References

[1] C. Jin, P. Leong, J. Leung, A. Corderoy and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements", *International Symposium on Multimedia Information Processing, Sydney, Australia* (2000)

[2] D. N. Zotkin, R. Duraiswami and L. S. Davis, "Customizable auditory displays", *International Conference on Auditory Display, Kyoto, Japan* (2002)

[3] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis", *Journal of the Acoustical Society of America* 93, 2764-2778 (1993)

[4] R. Nicol, "Restitution sonore spatialisée sur une zone étendue : application à la téléprésence", *PhD thesis, Université du Maine* (1999)

[5] Y. Kahana, "Numerical modelling of the head-related transfer function", *PhD thesis, University of Southampton* (2000)

[6] B. F.G. Katz, "Measurement and calculation of individual head-related transfer functions using boundary element model including the measurement and effect of skin and hair impedance", *PhD thesis, Pennsylvania State University* (1998)

[7] R. Greff, B. F.G. Katz, "Round robin comparison of HRTF simulation results: preliminary results", *Audio Engineering Society Convention, New York, USA* (2007)

[8] V. R. Algazi, C. Avendano and R. O. Duda, "Estimation of a spherical-head model from anthropometry", *Journal of the Audio Engineering Society* 49(6), 472-478 (2001)

[9] V. Larcher, "Techniques de spatialisation des sons pour la réalité virtuelle", *PhD thesis, Université Paris VI* (2001)

[10] J. Hebrank, D. Wright, "Spectral cues used in the localisation of sound sources on the median plane", *Journal of the Acoustical Society of America* 56, 1829-1834 (1974)

[11] F. Asano, Y. Suzuki and T. Sone, "Role of spectral cues in median plane localization", *Journal of the Acoustical Society of America* 88, 159-168 (1990)

[12] J.-M. Pernaux, "Spatialisation du son par les techniques binaurales : application aux services de telecommunications", *PhD thesis, Institut National Polytechnique de Grenoble* (2003)

[13] V. R. Algazi, R. O. Duda, R. P. Morrison, D. M. Thompson, "Structural composition and decomposition of HRTF", *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, USA* (2001).