ICA 2019 AACHEN

PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

HRTF and panning evaluations for binaural audio guidance

Sylvain Ferrand[1], François Alouges[2], Matthieu Aussal[3]

[1]CMAP, Ecole Polytechnique, Institut Polytechnique de Paris, France, sylvain.ferrand@polytechnique.edu
[2]CMAP, Ecole Polytechnique, Institut Polytechnique de Paris, France, francois.alouges@polytechnique.edu
[3]CMAP, Ecole Polytechnique, Institut Polytechnique de Paris, France, matthieu.aussal@polytechnique.edu

Abstract

We develop a device to guide blind people using binaural sound obtained by HRTF convolutions and reproduced by headphones. We have obtained good results in terms of user experience, but for guidance precision, the contribution of HRTFs compared to panning remained to be demonstrated.

In this study, we design different binaural filters and we ask the subjects to orient themselves in the direction of a sound source. We compare their performances with two HRTFs and two panning filters, both for static and continuously moving sound sources.

We show that HRTFs filtering allows the user to orient him/herself more precisely towards a sound source compared to a panning both in the static and the dynamic cases.

Keywords: Binaural, HRTF, blind people navigation, subjective evaluation

1 INTRODUCTION

Blind people use their ability to locate sounds on a daily basis. This is the case, for example, when they try to follow a person by following the sound of his/her footsteps. We already have used these capabilities to guide people with spatialized sounds played through headphones[1]. With the device that we have developed, the sound source is placed virtually in front of the person which allows him/her to be directed along a predefined path.

Human sound localization performances have been studied since a long time, either by *relative localization* tests (which consists in assessing the minimum audible angle between two identical sound sources [2]) or *absolute localization* (which aims to assess the subject's ability to designate the position of a sound source in space [3, 4]). Comparable studies have also been conducted for the localization of sounds spatialized by binaural filtering methods [5, 6, 7]. These studies usually focus on the localization of fixed sound sources in space and the question of the ability to quickly track changes in the position of moving sound sources has been little studied in the literature. We focus on the specific case of audio guidance which by nature concerns the azimuthal plane localization. In this application, the user has to follow a moving source, and continuously orient his/her head in the direction of sound. Therefore we specifically have to evaluate the capability of the subject to keep the sound in the frontal zone in a dynamic context.

The indices used for sound localization have also been extensively studied. Interaural Time Difference (ITD) and Interaural Level Difference (ILD) indices are known to play a major role in azimuthal location, while spectral indices are known to be mainly useful for median location [8, 9, 10, 11, 12]. However, these spectral indices are useful both for sound externalization and for the stability of sound image perception.

Therefore, we seek to evaluate the useful clues in horizontal localization in the context of moving sound sources. In order to do this, we propose to compare the azimuthal localization of moving sounds capabilities for different HRTFs with ITD+ILD panning filters derivated from these HRTF (which can be consider as HRTF without spectral cues).

2 FILTERS

2.1 HRTF

The experiments are performed with two HTRFs. We choose the HRTFs from two different projects with very distinct characteristics :

- An HRTF from the IRCAM Listen project [13]. We selected the subjet 1040 (human subject), well known in the binaural community for its externalization qualities. The HRTFs of the Listen IRCAM set are measured with a spatial resolution of 15°. Since we need a higher resolution, we have performed an interpolation at 2°.
- The HRTF of the Fabian project at TU-Berlin[14]. These HRTFs are measured with a head and artificial torso and have a native spatial resolution of 2° , a good phase linearity and good bandwidth qualities. We do not perform any additional processing on this HRTF.

2.2 Panning filters

We seek to compare the performance of these HRTFs with panning filters made up of ITDs and ILDs extracted from them, which can be considered as HRTFs from which the spectral component has been removed.

The *Interaural Time Difference* (ITD), is the difference in arrival time of a sound between the subjet's ears. It is well known to be an important cue for the localization of sounds in the azimuthal plane.

Several methods have been proposed to extract the ITDs. We can mention : the cross-correlation inter-aural [15], threshold detection and the group delay evaluation technique [16, 17].

It should be pointed out that that the notion of ITD is sometimes ambiguous: when the wave follows a multi-path travel around the head (typically in the contralateral case) several peaks are observed, of comparable energy levels, and corresponding to different arrival times.

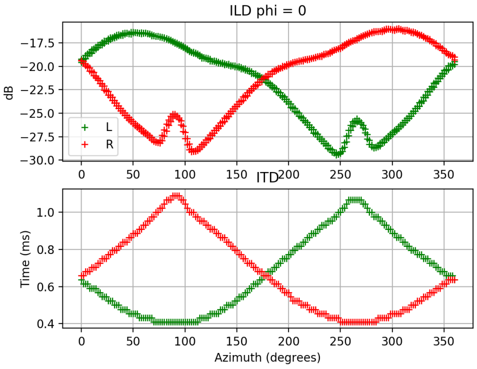

Figure 1. Fabian ITD-ILD at ϕ=0. Top: L/R levels for various azimuths, Down : L/R time of arrival of peaks in smoothed RIR

From a practical point of view, threshold detection methods are simple and generally effective. It is either possible to identify a threshold in the HRIR to distinguish the incoming signal from the initial noise, or to use the envelope of the signal (smoothing). We use this latter method in this study.

On the other hand, the *Interaural Level Difference* is the difference of sound level reaching the two ears due to the shadowing effect of the head. By nature, ILD are strongly frequency dependent. Because we want to discard the spectral aspect of the HRTF, we try to evaluate ILD using average sound level of the HRIR. To do this, we can consider the RMS value of the RIR but this approach leads to overestimate the ILD. To overcome this problem, we have decided to use the peak value of the smoothed RIR (fig. 1).

Additionally, extracted ITD and ILD can be easily smoothed or interpolated. Finally, using these data, we can construct a panning filter derived from each HRTF (fig 2). These panning filters typically have a very clean phase profile.

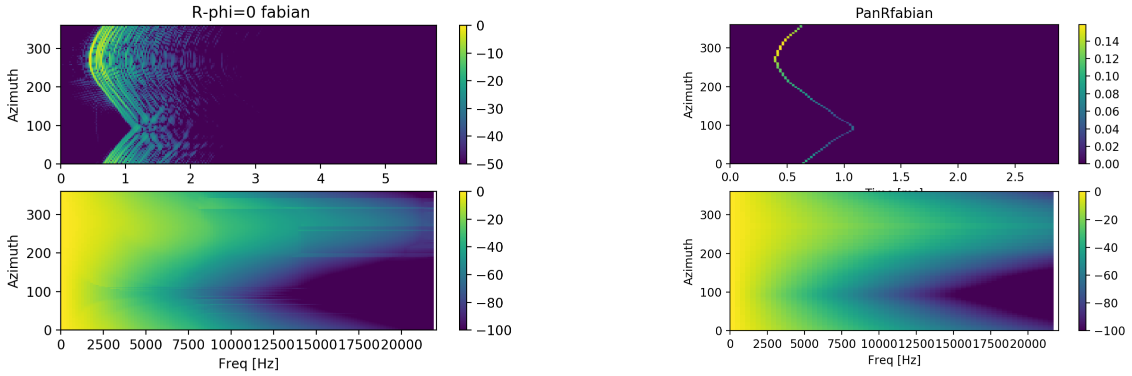

Figure 2. FABIAN HRTF and Panning RIR and phase

3 EXPERIMENTAL PROTOCOL

3.1 Method

The aim of these experiments is to measure subject's ability to locate fixed or moving sound sources using different spatialization filters. More precisely, we want to estimate the capability of the user to point and orient him/herself to a sound source. Thus, in these experiments, the subjects indicate the orientation of the sound source by orienting his/her head to the source (*nose pointing* technique). Hence, in all the experiments the position of the head is recorded using a head tracker, i.e. a compact and lightweight electronic device that measures the orientation of the head using gyro-magnetic sensors. This head-tracker is attached to the headphones. The measures from the head-tracker are used to compute azimuth (and elevation) of the sound source when the user changes his/her orientation, then the audio renderer take in account the orientation of the user in real time.

We are using the Bosch BNO-055 sensor which, well calibrated, performs with a typical 1-2° precision in the yaw axis (in the experimental conditions) with a latency shorter than 10ms.

Because the user orients him/herself to the sound using his/her head and body, we are mostly measuring the ability of the user to point and center the sound which is more relevant than other techniques (finger pointing) to evaluate the possibility to guide a person since it corresponds to the natural way of following a sound.

In all our experiments, subjects are placed in a rotating chair in a room isolated from outside noise and blindfolded (fig 3). Finally, they are equipped with a semi-open headset (BeyerDynamic DT990 pro) and the head-tracker. The user will listen a spatialized sound simulating a sound source on a self-centered circle and located horizontally at the level of the subject's ears. The user is asked to orient him/herself towards the sound source in order to position it in front of him/her.

In all series of experiments, we use white noise burst stimuli (100ms noise and 20ms silence repeated three times and 100ms silence). Participants are fully informed of the system and the objectives sought and are briefly trained to use the system.

We use the same positioning sequence (a random walk) for all users and experiments but the sequence is divided into sub-sequences distributed differently for each group of participants. These sub-sequences will be reorganized in each experiment and for each group to avoid attributing differences to training. Several transfer functions (HRTFs) are tested (i.e. the experiment is repeated several times). The order of passage of participants

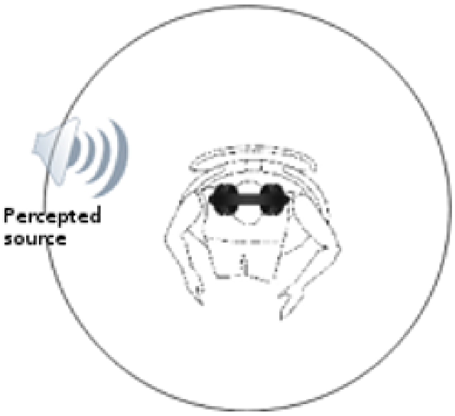

Figure 3. Experimental setup, the user is seated on a rotating chair and equipped with a headphone and a head-tracker

is randomized in order to avoid bias due to learning or tiredness.

3.2 Experiment 1 : stationary source

- The user listens to a spatialized sound simulating a sound source on a self-centered circle located horizontally at the level of his ears. The stimulus is emitted for a period of 10 seconds.
- The subject orients him/herself towards the sound.
- At t=10s, the orientation of the subject (i.e. the position of the head) is measured with the head tracker and the positioning error is calculated.
- The sequence continues.

The experiment lasts for 3 minutes, which corresponds to about 18 measurements.

3.3 Experiment 2 : continuously mobile source

This experiment is almost identical to the previous one, except that the user must continuously orient him/herself towards the sound source which is moving all the time. Ten times per second the orientation of the user's head and the position of the virtual sound source are recorded and compared.

The experiment lasts for 3 minutes. At the end of the experiment, the cumulative error is calculated and all measurement data are kept. For all subjects, this experiment is performed after the experience 1.

4 RESULTS

The tests have been performed by 13 subjects and we give here some preliminaries results, most of them should be statically confirmed with a larger experimental group.

4.1 Static tests

In this test we are comparing azimuthal sound source positioning for HRTF vs panning. The figure 4 gives the results, subject per subject, for IRCAM and FABIAN test. In both case, the Root Mean Square orientation Error (RMSE) is better for HRTF. Subject per subject comparison also shows that most participants perform better with HRTF compare to panning. Compared together FABIAN and IRCAM give very similar precision results.

Table 1. Comparison of HRTFs with panning in static case

Filter	IRCAM	IRCAM Panning	Fabian	Fabian Panning
OverAll RMSE (degrees)	5.42	8.72	5.32	7.01
Number of subjects who perform better with this filter	9	4	8	5

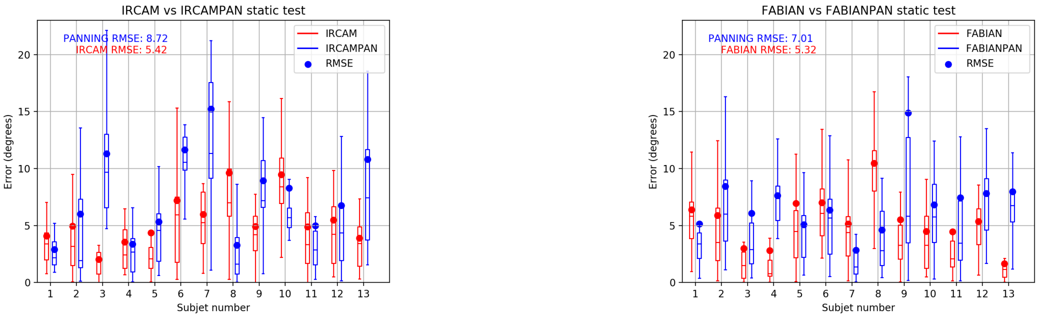

Figure 4. Static source pointing HRTF vs Panning for each subject

4.2 Dynamic tests

The dynamic test also shows better localization results for HRTF compared to panning, either for overall (RMSE) precision and if we compared the results subject per subject (fig 5, table 2).

Table 2. Comparison of HRTFs with panning in dynamic case

Filter	IRCAM	IRCAM Panning	Fabian	Fabian Panning
OverAll RMSE (degrees)	9.62	13.56	9.93	11.07
Number of subjects who perform better with this filter	11	2	9	4

The trajectory of the sound source is generated with a random walk on a circle around the listener. In order to go further it is possible to separate periods of slow and fast movements. The comparisons of RMSE for each subject during slow and fast source movement can be seen on figure 6. Predictably, the error increases with the speed of the source. In all cases, HRTFs perform again better than panning, but it seems that the gap between HRTF and panning widens when the sound source moves faster.

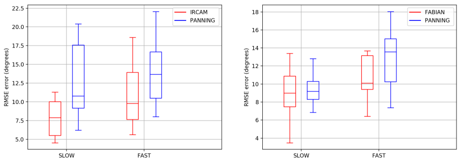

Figure 6. Slow/Fast source movement

4.3 Discussion

On every situation HRTF performs better than panning which seems to suggest that spectral cues are also important for this task.

If we compare together the two HRTFs (as a reminder, one is from a human subject with 15° resolution interpolated at 2°, the other one is from a mannequin measured at 2° native resolution), the results in localization precision are very similar.

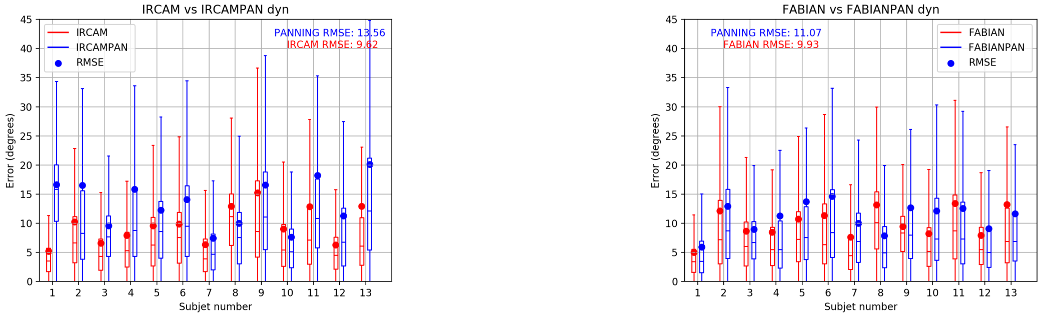

Figure 5. Overall dynamic source pointing HRTF vs Panning

If the results with HRTFs are extremely similar, we can observe some differences with panning. Indeed, panning computed from the Fabian database seems to perform slightly better than the one from Ircam. A possible explanation for this result could the overestimation of ILD for Fabian with our ILD extraction technique. The figure 7 shows our estimated ILD ($Level_{right} - Level_{left}$) for FABIAN and IRCAM that reveals a difference of 4dB. This overestimation of ITD could lead to a better capability to precisely *center* the sound for the user.

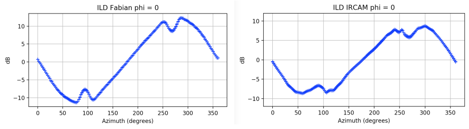

Figure 7. Comparison of ILD Fabian/Ircam. Value of extracted levels differences $Level_{right} - Level_{left}$ for all azimuth at $\phi = 0$

5 CONCLUSIONS

These tests broadly confirm the efficiency of HRTF compare to panning for orientation of people in the context of guidance (i.e. in the azimuthal plane and when user mostly keep the sound in the frontal area).

In the case of the azimuthal plane sound localization, HRTFs filtering is well known to allow the front-back discrimination, but spectral cues also seem important to perform sound source localization and centering which are very important for audio guidance.

If HRTFs seem more efficient, panning techniques also give reasonable results for guidance with the advantage of an easy and efficient implementation even on very low performance processors.

REFERENCES

[1] S. Ferrand, F. Alouges, and M. Aussal, "An augmented reality audio device helping blind people navigation," in *International Conference on Computers Helping People with Special Needs*, pp. 28–35, Springer, 2018.

[2] W. M. Hartmann and B. Rakerd, "On the minimum audible angle—a decision theory approach," *The Journal of the Acoustical Society of America*, vol. 85, no. 5, pp. 2031–2041, 1989.

[3] J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners," *The journal of the Acoustical Society of America*, vol. 87, no. 5, pp. 2188–2200, 1990.

[4] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz, "Auditory localization of nearby sources. ii. localization of a broadband source," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1956–1968, 1999.

[5] J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1493–1510, 1999.

[6] P. Zahorik, P. Bangayan, V. Sundareswaran, K. Wang, and C. Tam, "Perceptual recalibration in human sound localization: Learning to remediate front-back reversals," *The Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 343–359, 2006.

[7] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[8] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1829–1834, 1974.

[9] R. A. Butler and K. Belendiuk, "Spectral cues utilized in the localization of sound in the median sagittal plane," *The Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1264–1269, 1977.

[10] P. J. Bloom, "Determination of monaural sensitivity changes due to the pinna by use of minimum-audible-field measurements in the lateral vertical plane," *The Journal of the Acoustical Society of America*, vol. 61, no. 3, pp. 820–828, 1977.

[11] A. Kulkarni, *Sound localization in real and virtual acoustical environments*. PhD thesis, Boston University, 1997.

[12] H. Han, "Measuring a dummy head in search of pinna cues," *Journal of the Audio Engineering Society*, vol. 42, no. 1/2, pp. 15–37, 1994.

[13] O. Warusfel, "Listen hrtf database," *online, IRCAM and AK, Available: http://recherche. ircam. fr/equipes/salles/listen/index. html*, 2003.

[14] F. Brinkmann, A. Lindau, S. Weinzierl, G. Geissler, and S. van de Par, "A high resolution head-related transfer function database including different orientations of head above the torso," in *Proceedings of the AIA-DAGA 2013 Conference on Acoustics*, Citeseer, 2013.

[15] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992.

[16] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, and H. Møller, "The interaural time difference in binaural synthesis," in *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[17] B. F. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. 3530–3540, 2014.