

CW_binaural~: A binaural synthesis external for Pure Data

David Doukhan
IRCAM/EPITA
1 place Igor Stravinsky
75004 Paris
France
david.doukhan@gmail.com

Anne Sédès
CICM - Université Paris VIII
MSH Paris Nord
93200 Saint Denis
France
sedes.anne@gmail.com

ABSTRACT

In this paper, we present a new binaural synthesis external for Pure Data, allowing a real-time rendering of 3D sound scenes through headphones. While the synthesis method is based on well-known paradigms, the work that has been done should be considered as a technological transfer aiming at the PD community. It provides a user-friendly tool, that will be appreciated by musicians, artists, or sonor designer. Furthermore, it allows to control and to add many synthesis parameters, making possible its use as a baseline implementation used by researchers.

Keywords

3D sound, Binaural synthesis, HRTF interpolation, Minimum Phase filter, Pure Data.

1. INTRODUCTION

Sound spatialization methods allow to position sound sources in space, and give the listener the ability to localize them. Those methods improve immersion of the listener, and have been used in a wide variety of domains: such as Teleconferences Applications, Virtual Environment [27], Interfaces for visually impaired users [11], video games, cinema and music [7, 24].

The most widely spread standards for sound spatialization allow to position sound sources in a 2D space (Stereo, Dolby setups). Other methods allow to obtain a better spatial resolution, or to position sound sources in a 3D space, such as Vector Based Amplitude Panning (VBAP) [23], Wave Field Synthesis (WFS) [4], or Ambisonic [9].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

PdCon09, July 19-26, , 2009, São Paulo, SP, Brazil.

Copyright remains with the author(s).

All those methods have an inherent cost, and are characterized by the size of their *sweet spot* (ie the quality of spatial perception outside the center of room), the frequency band for which the spatial informations will be kept, and constraints on the loudspeakers number and position for obtaining a given spatial resolution.

3D restitution methods based on the use of headphones, such as binaural synthesis, are of particular interest, because they are cheap, and provide an optimal spatial resolution. As a consequence, the use of real-time binaural synthesis components can be considered as an appropriate tool for the composition process, as well as for the monitoring of a live performance. It can also be considered, under certain assumptions that will be developed later, as a great technology for distributing spatialized music, that could be experienced with minimalistic setups.

2. CUES OF SPATIAL SOUND PERCEPTION

Psycho-acoustic basis are necessary to understand and evaluate our ability to localize sound sources, and have been widely investigated in the literature (see [5]). Those considerations should be taken into account for efficient 3D sound synthesis systems conception.

2.1 Interaural Time and Level Differences

The primary cues influencing our ability to localize sounds are the Interaural Level Differences (ILD), and the Interaural Time Differences (ITD). ILD is a consequences of the sound having to pass through the head to reach the furthest ear, and may results in sound level differences of the order of 9 dB. ITD represent the difference of arrival time of a sound between the two ears: it is null for front or back signal, and may reach 0.63 ms for a sound source located at one's far right or left. While those cues may be sufficient to explain how we localize sound on the front horizontal plane, they do not provide enough informations to appreciate the elevation of a source, nor to know is the sound sources are located behind or in front of the listener.

2.2 Body acts as a Filter

Depending on the localization of a given sound source, the head, torso, and the external ear will act as a linear time invariant filter that will transform the sound signal. In other word, a filtering operation will be applied by the body to the sound signal depending on its position in space. For each possible spatial position, a filter could be described through a corresponding impulse response called Head related impulse response (HRIR). The frequency domain representa-

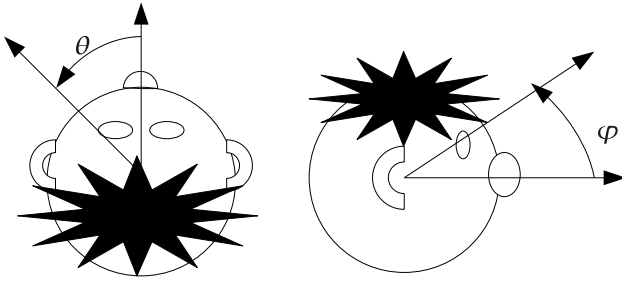


Figure 1: azimuth θ and elevation φ

tion of the HRIR is called the Head related transfer function (HRTF).

2.3 Localization Blur

Our ability to distinguish a sound source localization depends on several factors, such as its position in the horizontal plane (azimuth), in the median plane (elevation), and its distance (See Figure 1). [5] presented a comprehensive description of this ability, and provided estimations of the *localization blur*, defined as the amount of displacement of the position of a sound source perceived by subjects as a change in the position of the auditory event. On the horizontal plane, it lies between 3° (front sources), 10° for lateral sources, and 5° back sources. On the median plane, the localization may reach values up to 22° .

2.4 Miscellaneous cues

The nature of a sound source is another factor that may influence localization: sounds having energy for all of their frequency bands, such as white noise, are easier to localize than sounds having a sparse spectral content (sine tone). Familiar sounds are also easier to localize, since we have internal representation of the timbre of those particular sounds, for different spatial positions. Head motion helps in improving localization, mostly for distinguishing front versus back. Since vision has a predominant role in localization, it improves sound source localization as long as the visual stimulus is aligned with the sound stimulus: otherwise we may observe what is known as ventriloquism effect [12]. Other cues that may influence localization are early echo response, and reverberation, and will be discussed in the next section.

3. BINAURAL SYNTHESIS

3.1 Binaural Recording

Binaural recordings allow realistic restitution of the spatial cues through headphones. For a convincing demonstration of that technology, readers are invited to listen the amazing *Virtual Barber Shop*, which has been widespread on the Web.

Two main methods allow to produce binaural recording. The first one consists in mounting two high fidelity microphones in the ears of an individual, or in a dummy head (see the Neumann KU 100 Dummy Head Binaural Stereo Microphone). While this kind of approach is quite intuitive, and produce high quality results, the price of the corresponding hardware is prohibitive, and it imply moving the sound sources for real, which is not very practical when considering a huge number of sources having complex trajectories.

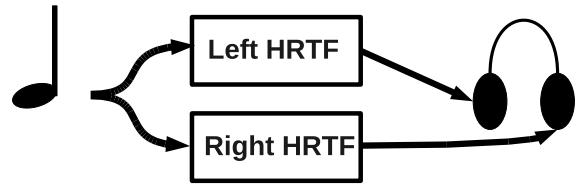


Figure 2: Monaural signal is processed by the left and right HRTF and restituted through headphones

The other approach, called binaural synthesis, consists in filtering the sound sources with the HRTFs corresponding to their requested spatial position defined by an azimuth, and an elevation (See Figure 2).

3.2 HRTFs

HRTFs should be measured in anechoic chambers, in order to constitute a database that would not take room dependent reflections into account. Since they model the way a given body filters the signal, they depend on the morphology of the body and are individuals. Using non individualized HRTFs may increase front-back confusion, as well as non externalization of the sound sources.

One solution that may help to tackle that problem would be to choose an HRTF set that bests reconstitute one's spatial cues. Other approaches [6] consists in using anthropometric manikins having an average morphology, or to consider that the HRTFs of a good localizer [28] can be considered as universal HRTFs.

However, one's can get used to someone else HRTFs, and [3] showed that using individualized HRTFs does not affect azimuth localization, mostly based on ITD and ILD indices, rather than on the spectral content of the signal.

Several HRTF databases measured on subjects having different morphologies are available to the public: such as Listen Project database [1], sampled every 15 degrees, or the CIPIC database [2] sampled every 5 degrees. Other database have been measured using the KEMAR (Knowles Electronics Manikin for Acoustic Research) Dummy-Head Microphone [10] and have been sampled every 5 degrees.

3.3 Improving localization

Several methods have been proposed to tackle localization problems inherent to binaural synthesis, such as distance estimation, front-back confusion, and minimization of the localization blur on the horizontal and median plane.

The most relevant methods consist in combining the binaural synthesis with a head-tracking system, that would select the best matching HRTF according to the orientation of the head of the listener, and to include the room effect (early reflections and reverberation) corresponding a given virtual room.

[3] did a comprehensive analysis of the impact of head tracking, reverberation and individualized HRTF for the spatial perception of speech source and showed that head tracking help to reduce the front back confusion, as long as the latency implied by the use of the head tracking system is acceptable. They also showed that room effect helps to lower the localization blur in the horizontal plane, as well as to improve the perception of distance, however, it also increase the localization blur in the median plane, and result in worst

estimation of the elevation of the sources. Moreover, while adding room effect may increase the realism of a virtual scene, its use in a musical context may lead to unwished changes in the timbre of the sources.

Consequently, binaural system users should be aware of those properties, and do the trade-off appropriate to their needs.

3.4 HRTF interpolation

Since azimuth and elevation are continuous values, sampling the whole set of spatial positions to get the corresponding HRTFs is infeasible, and binaural synthesis systems should be able to interpolate missing HRTFs from the available database. While considering moving sources, a brutal transition from one HRTF to another leads to artefacts: consequently, HRTF interpolation is a step that can not be skipped while designing binaural systems.

Several strategies have been implemented in existing PD externals in order to tackle that problem:

`earplug~`[29] did a linear interpolation based on the use of the nearest HRTFs of the requested position. While the amplitude of the obtained HRTF spectrum is the same as the ideally interpolated HRTF, the resulting ITD is biased, leading to localization artefacts.

`IEM bin_ambi` [20] use the virtual loudspeaker paradigm, consisting in encoding the sources in a M order ambisonic representation, and decode it using $(M + 1)^2$ virtual loudspeakers, associated with time invariant HRTFs. While that approach allow to make feasible the spatialization of a huge number of sources associated with a room effect, it does not take advantage of the whole HRTF database available, and may result in a lower localization precision than direct interpolation.

`Spat~`[13, 15] (available only for Max/MSP) used a method based on the decomposition of the HRTF into a minimum phase and a all pass component, and on an approximation of the HRTF consisting to represent the all pass component as a pure delay. The resulting structure consists in using a fractional delay corresponding to the ITD estimated from the HRIR, in cascade with a filter corresponding to the minimal phase component. [17] showed that this approximation leads to proper localization results for frequencies lower than 10 KHz, which are those used for localization. Once that decomposition has been done, the method consists in interpolating linearly the minimal phase part of the nearest HRTFs, and interpolating the ITD separately.

4. IMPLEMENTATION

The aim was of the binaural synthesis external that has been done was to fulfill two major requirement: being user-friendly, and providing an customizable synthesis method that would be as accurate as possible. It has three signal inlets: the source signal, the azimuth and the elevation, and two outlets corresponding to the signal that should be send to the left and right ears (See Figure 3).

It aims at providing an intuitive control interface similar to `earplug~`[29] that would overcome its major limitations, such as: using without efforts the HRTFs of a given subject, being able to use different set of HRTFs simultaneously, managing HRTFs database having a different number of measurements, choosing the size of the HRIR considered, and controlling the azimuth and elevation through signal inlets, instead of command inlets.

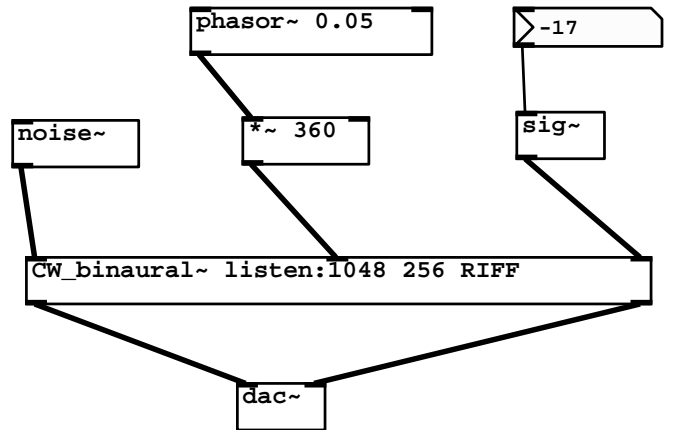


Figure 3: A white noise moving sound source doing rotations on the horizontal plane at elevation -17 degree. The HRTF set used comes from the Listen database, and corresponds to the subject 1048. The filtering method selected (RIFF) use 256 coefficients of the HRIR and do the convolution in the temporal domain.

To achieve that task, some points may be configured by the user through optional object arguments, such as

- the set of HRTF filter used (Listen[1] or CIPIC[2])
- the length of the considered HRIR (power of 2)
- the filtering method used (RIFF, FFT)
- the fractional delay method (Linear, Hermite,...)
- the ITD estimation method (Cross-Correlation, Over-sampled Cross-Correlation, ...)

In order to take advantage of all the HRTF data available, we kept the paradigm of real-time direct interpolation of the nearest HRTFs used in `earplug~`[29] and `Spat~`[13], and proceeded to the decomposition of the HRTFs into a pure delay and a minimal phase component, that would be interpolated separately to achieve optimal localization quality.

Since many steps of the synthesis are subject to improvements driven by research progresses (ITD estimation, filtering method, fractional delay, ...), we designed an external based on the use of the Object Oriented Paradigm in C++. While that choice has consequences on the speed of the implementation, the gain in modularity and maintainability provides a developer-friendly environment for scientists willing to improve a given part of the synthesis process.

4.1 HRTF decomposition

We proceed to the decomposition of the HRTF into a minimum phase (H_{min}) and a all pass (H_{ap}) component using

the following properties [21] used in Spat^{\sim} [13, 15]:

$$\begin{aligned} \text{HRTF} &= |\text{HRTF}| \cdot \exp(j \cdot \arg(\text{HRTF})) \\ &= |\text{HRTF}| \cdot \exp(j \cdot \text{mph}) \cdot \exp(j \cdot \text{eph}) \\ &= H_{\text{min}} \cdot H_{\text{ap}} \end{aligned}$$

with

$$\begin{aligned} H_{\text{min}} &= |\text{HRTF}| \cdot \exp(j \cdot \text{mph}) \\ H_{\text{ap}} &= \exp(j \cdot \text{eph}) \end{aligned}$$

with mph the phase of the minimal filter, obtained using the Hilbert Transform with the following relation:

$$\text{mph} = \text{Im}(\text{Hilbert}(-\log(|\text{HRTF}|)))$$

and eph the phase of the all-pass component called *excess phase*.

4.2 ITD estimation

Several methods have been proposed in the literature in order to estimate the ITD such as using spherical head models [16] or linear approximation [15], and have been analyzed in detail by [18]. For now, we provide the method proposed by [14] consisting in considering the maximum of the cross-correlation between the HRIRs.

Further development will include the use of methods based on the estimation of the group delay of the HRTF excess phase components evaluated at 0Hz recommended by [18].

4.3 Filtering

Two FIR (Finite Impulse Response) Filters have been implemented: a first one operating on the temporal representation of the signal, and another doing the computations on the spectral representation of the signal through the overlapp-add (OLAP) [21] method, providing exactly the same output for a given input signal. The difference between those filters rely in the fact that the filtering in the time domain is costly, but can be done with 0 latency, whereas the filtering in the spectral domain is fast, but require a latency corresponding to half of the HRIR considered. In other words, using the whole HRIR of the Listen Database (512 samples), we'll get a delay of 5.8 ms. Considering the first 128 samples of the HRIR allow to keep a localization accuracy considered as sufficient, and results in a delay of 1.5 ms.

While we do not consider that point as a priority, the latency could be lowered though the use of partitioned convolution algorithms that would result in bigger processing time.

4.4 Pure Delay

Several strategies may be used in order to implement the pure delay. The most intuitive consists in using the nearest sample of the requested value. Another method could be to use linear interpolation between the two nearest samples of the requested value. While those methods are computationally cheap, and sufficient when considering noisy sound sources (speech, saturated guitar, ...), they may cause clicks while considering sine tones, moreover they can lead to localization artefacts, since they modify the ITD.

PD [22] provides abstractions allowing to represent delays shorter than one sample such as vd^{\sim} based on cubic polynomial fit through 4 points, but resulted in a noticeable change in the timbre of the spatialized sources. Consequently, we investigated the use of other interpolation methods, and used a

Table 1: CPU usage (peak) using different filtering algorithms, and impulse response size

Filter	128	256	512
FFT	6.24	6.24	6.39
RIFF	20.2	31.2	54.6

Table 2: CPU usage (peak) of the different part of the algorithm, using the FFT filtering method with the 4-point, 3rd-order Hermite fractional Delay

HRTF size	HRTF Interpolation	Delay	Filtering
128	.93	3.12	3.12
512	1.09	3.12	3.27

4-point, 3rd-order Hermite x-form implementation described in [19], which gave us better results.

However, vd^{\sim} strategy, as well as ours, are interpolation methods based on the use of the current sample combined with two past samples and one future sample. Consequently, using those methods in real-time imply the addition of a 1 sample delay that could not be avoided, that could led to imprecisions when mixing binaurally synthesized sound sources with other ones.

For now, three kind of delays can be chosen by the users of the external: the non fractional, the delay based on linear interpolation, and the delay based on the 3rd-order Hermite x-form that is used by default.

Further developments will include the evaluation and integration of causal (based on the use of past values only) IIR and FIR variable fractional delay methods.

4.5 Performances

The benchmarking of the external have been done using a laptop equipped with a Intel Pentium T2330 CPU 1.6 GHz, which is considered as a Mid Range CPU (see <http://www.cpubenchmark.net>). Consequently, the results reported in this section are a baseline that could be easily surpassed using more recent hardware.

The results reported bellow correspond to the percentage of CPU usage for a given task. Since we aim to use the external in real time context, only the worst case performances (CPU peaks) have been reported.

In Table 1, we show the CPU usage for a fully operational binaural synthesis system, using the default 4-point, 3rd-order Hermite fractional delay, for different sets of filtering strategies and HRTF sizes. As expected, the FFT filtering method performs much better than the naive RIFF implementation, and do not show significative execution time difference for different HRTF sizes.

In Table 2, we investigate the computational cost of each task being performed, surprizingly, we saw that the fractional delay implementation, supposed to be less costly, was as expensive as the filtering.

Table 3: CPU usage (peak) for the Delay using different methods

Dummy	Linear	Hermite
1.24	1.56	3.12

In Table 3, we investigate the cost of all our delay implementation, sharing some common code.

The conclusions that could be drawn from those results are that in its current state of development, `CW_binaural` allows a high quality real-time rendering of more than 15 moving sound sources on a basic laptop, which may be enough for artistic purposes. Since the time spend in the delay is abnormally high (more than 1% of the CPU for a Dummy Delay), we may focus our efforts on C++ code optimization in order to avoid wasting too much time in that task, theoretically supposed to be less computationally intensive than the HRTF filtering. If we manage to do it, the next bottleneck will be the filtering, and will be hardly optimizable, since it already uses the highly optimized Mayer's real FFT provided in PD internals. [20] proposed a trick that could be used to reduce the filtering time by a factor of 2, consisting in mixing the spatialized sound sources in the frequency domain, and doing only 2 inverse FFT instead of 2 time the numbers of sources. Currently, we do not want to implement that method in `CW_binaural`, since it would mean not allowing the final user of the external to manipulate the output of each spatialized sound source separately, and would result in a lower component modularity.

5. ARTISTIC VALIDATION

The experimental art, non-standard, is a wonderful laboratory for testing the potential of musical expression and the rendering quality of new processes.

Anne Sedes will compose music for an electronic interactive sound installation, based on the use of the `CW_binaural` spatialization tool. One of the aims of this proposal is the investigation of music listening, unique to each individual. The sound space is a major dimension of music, especially electronic music [25]. It does not only necessarily concern the localization and simulation of "real world sounds" through room effects. Moreover, it helps composers to construct multidimensional dynamics variations carrying movement, timbre, rhythm, polyphony, musical sound, in order to create the consistency of music [26].

The proposal means: four headphones are connected to a cube that hides a computer with Pure Data, and a sound card with eight outputs; a multitouch interface is available to the listeners. The music is built based on "test" materials for binaural system: white noise, sinus, impulse, rich spectral variations, trajectories with variable speed, and natural recorded sound samples.

This box, equipped with a multitactile interface offers three listening options:

1. Hearing the piece of music as it was composed. Even if the piece is unique, it will be understood differently by each listener, in the frame of the oneness and own filtering of each human morphology. Listening is individual, and one can never know how another is listening, although the environment composed music itself is unique.
2. Real-time interactive manipulation of the position (azimuth, elevation) of various sound sources.
3. Controlling these parameters through physical models. A random part is incorporated into the composition.

This installation will be shown at the French exposure "Geste, Image Son" in the CFA Bagnole, Seine Saint-Denis (futur en seine, Cap Digital) in May-June 2009.

This artistic proposal aims a reflection about listening and about the singularity of each human being. Moreover, it helps us to test `CW_binaural` in a pragmatic and dynamic context of artistic production.

6. CONCLUSIONS AND FUTURE WORK

We presented a modular user-friendly binaural synthesis external allowing to perform an efficient real time interpolation of HRTFs, thus allowing a high quality perception of the position of the sources.

While the quality of the results obtained was satisfactory, they could be improved through the use of third-part head tracking components that could be plugged quite easily in our external.

Some points are still perfectible and may imply future developments, and the external design allows to integrate easily other binaural synthesis strategies, while preserving the compatibility with patches done before the improvements. The presented external did not include distance attenuation, neither does it include Doppler effect. That choice is deliberate since we consider those effects as independent of the spatialization method used.

The next improvement that will be done will consist in allowing the use of different signal sampling rates: for now it is limited to the sampling rate of the HRTF (44100 Hz for the Listen and CIPIC database). The last point where we may focus our efforts would be in the development of trajectory editors independent of the restitution device.

The external, as well as the conversion of the CIPIC database to a non proprietary format, will be available at http://cicm.mshparisnord.org/dl/index_en.html. We hope it will be enjoyed by the PD community.

7. ACKNOWLEDGMENTS

The authors would like to thank:

Khoa Van NGuyen, Thibaut Carpentier, Markus Noisternig, Grace Leslie and Olivier Warusfel from IRCAM's Room Acoustics team for their advices on the design of real-time binaural systems, as well as for their explanations of the mechanisms used in the `Spat` [13] and `bin_ambi` [20].

Ralph Algazi of the CIPIC Interface Laboratory for allowing us to convert and distribute the CIPIC HRTF database [2] to a non proprietary format that could be used with `CW_binaural`.

Charles Henry and Matt Barber for their explanations on efficient fractional delay implementation through the Pure Data mailing list.

Finally, we would like to thank `CyclikWeetos` [8] for their feed-back on the use of our external in their compositions, the name of the external: `CW_binaural` is a dedication to them.

8. REFERENCES

- [1] Listen HRTF database. <http://recherche.ircam.fr/equipes/salles/listen>.
- [2] V. Algazi, R. Duda, D. Thompson, and C. Avendano. The CIPIC HRTF database. In *Proc. 2001 IEEE*

- Workshop on Applications of Signal Processing to Audio and Electroacoustics*, pages 99–102, 2001.
- [3] D. Begault, E. Wenzel, and M. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.*, 49(10):904–916, 2001.
- [4] A. Berkhout, D. De Vries, and P. Vogel. Acoustic control by wave field synthesis. *The Journal of Acoustical Society of America*, 93(5):2764–2778, 1993.
- [5] J. Blauert. *Spatial Hearing*. 2nd ed. MIT Press, MA, 1997.
- [6] U. Burandt, C. Pösselt, S. Ambrozus, M. Hosenfeld, and V. Knauff. Anthropometric contribution to standardising manikins for artificial-head microphones and to measuring headphones and ear protectors. *Applied ergonomics*, 22(6):373, 1991.
- [7] J. Chowning. The simulation of moving sound sources. *Computer Music Journal*, pages 48–52, 1977.
- [8] CyclikWeetos. Electro-dancepunk-brutal-cannibal-zouk de pavillon. <http://www.myspace.com/cyclikweetos>.
- [9] J. Daniel, R. Nicol, and S. Moreau. Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging. *Preprints-Audio Engineering Society*, 2003.
- [10] B. Gardner and K. Martin. HRTF Measurements of a KEMAR Dummy-head Microphone. Technical report, MIT Media Lab, 1994.
- [11] S. Goose and C. Möller. A 3d audio only interactive web browser: using spatialization to convey hypermedia document structure. In *Proceedings of the seventh ACM international conference of Multimedia (Part 1)*, pages 363–371, 1999.
- [12] W. Hairston, M. Wallace, J. Vaughan, B. Stein, J. Norris, and J. Schirillo. Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, 15:20–29, 2003.
- [13] J. Jot and O. Warusfel. Spat : A Spatial Processor for Musicians and Sound Engineers. *CIARM: International Conference on Acoustics and Musical Research*, Ferrara, 1995.
- [14] D. Kistler and F. 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*, 91:1637, 1992.
- [15] V. Larcher. *Techniques de spatialisation des sons pour la réalité virtuelle*. PhD thesis, 2001.
- [16] V. Larcher and J.-M. Jot. Techniques d’interpolation de filtres audio-numériques, Application à la reproduction spatiale des sons sur écouteurs. In *Proc. CFA: Congrès Français d’Acoustique*, 1997.
- [17] S. Mehrgardt and V. Mellert. Transformation characteristics of the external human ear. *The Journal of the Acoustical Society of America*, 61:1567–1576, 1977.
- [18] P. Minnaar, J. Plogsties, S. Olesen, F. Christensen, and H. Mcler. The interaural time difference in binaural synthesis. *Preprints-Audio Engineering Society*, 2000.
- [19] O. Niemitalo. Polynomial interpolators for high-quality resampling of oversampled audio, 2001. <http://www.student.oulu.fi/~oniemita/dsp/deip.pdf>.
- [20] M. Noisternig, T. Musil, A. Sontacchi, and R. Holdrich. A 3D real time Rendering Engine for binaural Sound Reproduction. In *International Conference on Auditory Display (ICAD)*, volume 9, pages 107–110, 2003.
- [21] A. Oppenheim and R. Schaffer. *Digital signal processing*. 1975.
- [22] M. Puckette. *The Theory and Technique of Electronic Music*. World Scientific Publishing Co. Pte. Ltd., 2006.
- [23] V. Pulkki. *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. PhD thesis, Helsinki University of Technology, 2001.
- [24] J. C. Risset. Interaction and Spatialization: Three Recent Musical Works. *Sound & Music Computing*, 2006.
- [25] A. Sedes et al. *Espaces sonores: actes de recherches*. Editions musicales transatlantiques, 2004.
- [26] H. Vaggione. Some ontological remarks about music composition processes. *Computer Music Journal*, 25(1):54–61, 2001.
- [27] I. Viaud-Delmon, F. Znaïdi, N. Bonneel, D. Doukhan, C. Suied, O. Warusfel, K. N’Guyen, and G. Drettakis. Auditory-Visual Virtual Environments to treat Dog Phobia. In *International Conference Series On Disability, Virtual Reality And Associated Technologies (ICDVRAT)*, 2008.
- [28] E. Wenzel, F. Wightman, and D. Kistler. Localization with non-individualized virtual acoustic display cues. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, pages 351–359, 1991.
- [29] P. Xiang, D. Camargo, and M. Puckette. Experiments on spatial gestures in binaural sound display. In *International Conference on Auditory Display (ICAD)*, 2005.