BINAURAL FOR POPULAR MUSIC: A CASE OF STUDY

Simone Fontana[1,2], Angelo Farina[1], and Yves Grenier[2]

[1]Ecole Nationale Supérieure des Télécommunications, TSI Paris, France
[2]Università di Parma, Parma, Italia
fontana@laegroup.org

ABSTRACT

The goal of this study is to retrieve useful information about the reactions of listeners to different recording techniques, namely binaural and stereo. This has been done comparing different mixes of the same song. Each mix is obtained from stereophonic and binaural recordings, or processing proximity recordings with stereo panning, binaural synthesis or a hybrid approach. The comparison is made through listening tests with headphones and an analysis of subjects' reactions and meaningful subjective parameters ratings.

[Keywords: Spatialization, Binaural, Listening Tests]

1. INTRODUCTION

Binaural technology can be approached from different perspectives and in different domain of application. Scientific research dedicated a lot of study to this technology, both to binaural recording and synthesis issues. Augmented virtual reality massively uses binaural technology to improve realism and immersion. Binaural technology is employed for acoustic measurements in hearing and audiology, sound quality tests in telecommunications and automotive systems, room acoustics, psychoacoustics. Field recording employs binaural techniques for 'virtually taking the listener there': it is used for event and travel reporting, sometimes in an anthropological, ethnographical, historical or musicological context. Environmental sound recording and bioacoustics, so as alternative medicine show interests for the 3D rendering of ambient sounds with binaural microphones or synthesis.

Binaural techniques are also used for artistic creations, namely to design soundscapes or to enhance the spatial dimension in music recordings. Classical binaural recordings have obtained good comments in the audiophile community, so as binaural audio dramas and electronic/ambient music, but the market part of this kind of products is limited compared to the popular music market.

Popular music is only marginally touched by binaural technologies. Examples can be found in Tchad Blake producer discography (Pearl Jam, Latin playboys, Tom Waits) or in albums like Pink Floyd "The final cut", Michael Jackson "History", or in some Stevie Wonder and Lionel Ritchie albums, that used the holophonic technique (by Hugo Zuccarelli). Nevertheless the application of binaural techniques in popular music stays anecdotic. This lack of success seems to be due both to technical reasons and more general considerations. As a matter of fact, binaural recordings through dummy heads presented in the past major problems, mostly linked to the possibility of listening to binaural mixes on stereo loudspeaker. The mix sounded thin, distant, and without low frequencies [1]. Today this kind of problems has been solved by marks such as Neumann, through diffuse field equalization of the dummy head microphones [2], even if, in our opinion, a certain voice presence loss can still be perceived. The loss of the 'you are here' effect with loudspeaker listening can be avoided using crosstalk cancellation solutions currently available [3], [4]. More general compatibility problems are object of current researches for binaural upmix/downmix to and from different multichannel technologies, such as Ambisonics, or 5.1, [5]. We remark that the only 'you are here' effect is lost in normal loudspeaker reproduction, but the binaural recording has its own sound on loudspeaker, just as any other stereophonic technique.

Another problem often related is the common studio use of proximity/multimic recording (namely for enhancing the voice presence), which seems to represent an obstacle to all-binaural recording in studio. Binaural synthesis is candidate solution to face the multimic constraint, but other related difficulties appear, namely concerning HRTF modelling, individualization, and measurement [6].

To this kind of technical problems, we have to add the individual sound engineers' studio practice, their habits, and their (relative) inertia to major technological evolutions, which involve a more or less longer time of adaptation in order to develop field experience. The final success is, anyway, up to audience acceptance of the new techniques, and is then related to how listeners react faced to binaural audio products. The current study presents some issues concerning these matters, namely how listeners react to binaural products compared to traditional stereo ones.

In order to identify the potentials of binaural technologies compared to more traditional stereophonic approaches, five preliminary mixes have been carried out from the recording session of a professional guitar-voice duo, for an original composition of theirs. AB recording and stereo panning, dummy head binaural recording and HRTF synthesis, so as a hybrid technique have been considered.

A first final mix has been carried out: it contains the full song, seen from different 'perspectives', that is the different recording and synthesis techniques. The listening perspective changes dur ing the song, so that a listener can fully enjoy, in a fluid way and without abrupt changes, the rendering differences for the five considered strategies. This could be seen as a 'marketing' approach of comparing recording techniques.

A second approach, based on listening of selected extracts of the 5 mixes, numerical rating and statistical processing has been followed in order to try a more scientific analysis of the problem.

2. RECORDING AND MIXING DETAILS

The recording has been carried out in a little studio (6x5x4 m) acoustically treated at Ecole Nationale Supérieure of Telecommunications in Paris. The singer was positioned in front of the dummy head, at 1 metre, and an AB couple which were nearly coincident. The guitarist sat on her left, one meter further.

The binaural recording has been made with a Neumann Ku-100 dummy head; the AB couple is composed by two MK2 Schoeps. The proximity recording is made with a Soundelux Ifet7 for the voice and a MK2 Schoeps for the guitar. HRTFs for synthesis have been chosen within the IRCAM LISTEN database [8], by listening to several sets and choosing the best one according to the author and the sound engineer advice. All the 5 mixes that are contained in the first 'multiperspective' listening sample have been done on a analogue Trident Series75 using hardware equipment (D.W Fearn VT-2..), as well as in the box, in Logic Pro7 on an Apple G5 1,8Ghz.

In a first mixing step, complete freedom has been accorded to the sound engineer, in order to let him use all the potential of a professional studio to achieve a product matching discography standards. For example, in the hybrid mix he used:

- regular stereo mix on the 2 mono-proximity sources, panned in stereo, with equalization, compression, a touch of widening (M/S Matrix), effects (3 different convolution reverbs panned in various directions).
- binaural recording compressed through an SSL Talk Back compressor (plug-in version), a bit of stereo reverb applied, and submixed 10dB lower than the stereo mix. The final result has been down mixed in one audio file.

A similar processing has been performed for the other mixes; no particular sound engineering problems have been encountered during the production process. The first final mix presented to the listeners contained the whole song, obtained by mixing sequentially some extracts of each one of the mixes, so that the song changed continuously from one technique to the other, without abrupt changes, making the listening easy and fluid. Each change was announced by a superposed voice. The multiperspective sample has been presented in an unblind fashion to the listeners, which knew which extracts belonged to which strategy. This is a common marketing approach used for presenting comparative listening, [7].

One remark can be made: both the unblindness and the slightly different processing of the 5 mixes can introduce a bias in a statistical analysis of the data. However, the goal of this first listening test was to record the advices of the listeners on a 'final product'. The judgement were more qualitative than quantitative and were aimed to provide a very general idea on the sensitivity of the audience to a true sound engineered production of the song, as it could be found on a normal CD. The listening sample has been made available on audio-dedicated and academic research Internet sites and more than 20 feedbacks have been received from audio professionals. These feedbacks are available contacting the author.

In a second phase, some unprocessed extracts of the recordings and synthesis mixes have been used for quantitative rating. This time three extracts has been obtained from each one of the 5 mixes, that is

- BR1,BR2, BR3 from the binaural recording mix
- AB1, AB2, AB3 from the AB recording mix
- SP1,SP2, SP3 from the stereo panned mix
- BS1,BS2, BS3 from the binaural synthesis mix
- HY1,HY2, HY3 from the hybrid mix

The BR and AB series are the recorded files without processing; the SP series has been obtained panning the voice proximity recording in the centre and the guitar on the left (45 degrees); the BS series has been obtained using the LISTEN IRC002 set raw HRTFs corresponding to azimuth 0 and elevation 0 for the voice, and azimuth 45 degrees on the left, 0 elevation for the guitar; the hybrid mix has been obtained mixing SP and BS (this one 3dB lower).

We considered two approaches: a 'live recording approach', as the one which is often used in classical music context, and a 'studio recording approach', as the one used in pop music. The first one consists in recording with two microphones set in the performing area, in a well-studied position: the recording contains the contributions of both musicians and hall, with weak possibilities for post processing (which is not always welcomed in classical context). The mix is up to the performing musicians, the recording engineer has to be transparent.

The second one is based on 'proximity recording', that is with microphones set in proximity of each musician. This type of recording is useful because it allows for further processing in order to achieve accurate mixing of the single tracks, effects insertion and space organization: the mix is up to the sound engineer, which is directly involved in the sound creation.

The first category test sample has been obtained mixing sequentially in randomized order the three extracts of the BR and AB series; the same for the second category test sample with the three extracts of the SP, BS and HY series. A reference signal REF has been put at the beginning of the sample: it is composed by a mono mix of the proximity recordings for the voice and the guitar. Six sequences (and the reference) for the live approach and nine sequences (and the reference) for the studio approached were then considered. All the extracts have been normalized. The test was blind and performed using Sennheizer HD600 headphones and a B&K ZE0769 headphones amplifier.

3. LISTENING TESTS

We asked the listeners to rate some parameters related to the spatial and timbre dimensions of the mixes, and also to provide a general advice about the pleasantness of each mix, according to their personal expectations and taste.

We choose to define the spatial dimension of the 'live mixes' with three parameters: **localization, sound relief, spaciousness**. Localization is related to the possibility for the listener to associate one sound to one position in space, without fluctuations or 'diffuse' sound effects. If there is a plan organization (background, foreground) in the performance, the 'sound relief' parameter is related to the correct reproduction of it. The spaciousness is related to the listener envelopment and the source apparent width, LEV and ASW respectively.

Given the fact that in the 'studio mixes' we did not synthesize a recording space, or a plan organization, we chose to characterize the spatial dimension in these mixes with **out-of-head localization and source width.** Localization 'out of the head' with headphone reproduction is believed to happen when all the hearing system localization cues are provided by the audio chain. This is theoretically guaranteed by binaural

synthesis. The 'source width' parameter characterizes the difference existing between a point source (issued from synthesis) and an extended source perception issued from recording).

The 'timber quality' and 'pleasantness' parameters are common for the two mixes categories.

All the parameters have been described with care to the listeners. The listeners were asked to rate each sequence (6 in the first test, 9 in the second) in a range for 1 to 5. The semantic value of these rates is reported in table 1.

Localization:	5: stable and unique 1: fluctuating and diffuse
Spaciousness:	5: enveloping and large 1: isolated and thin
Sound relief:	5: separated and 'in perspective' 1: confuse and flat
Timber:	5: natural 1: with artifacts
Pleasantness:	5: very pleasant 1: very unpleasant
Out-of-head localization:	5: well externalized 1: totally inside the head
Source Width:	5: large source 1: point source

Table 1: *semantic of the proposed parameters for rating*

15 subjects with supposed experience in audio (Sound and Audio engineers and Musicologists) performed the test in the listening room of the Casa della Musica in Parma. They could listen to the mixes as many time as they want. Each test lasted at least 30 minutes, at most 45 minutes.

4. RESULTS

4.1. Reliability of the test subjects

A first common remark made by the majority of the subjects was that the test was quite difficult. The subjects did differentiate the sequences, but the focusing on the different parameters was something they were not used to do. Focusing on localization, for example, without considering spaciousness could be very difficult for a person without a deep analytical and more precisely, 'spatial', listening. We could suppose that the difference between techniques is not so flagrant or that the performance used for recording is not the most suitable for such differentiation, but the fact that every listener perceived a difference in the sequences, made us think that it was just a matter of going deeper to try to find out 'why' they were different.

A second remark is related to the fact that three different extracts of the song were presented to the subjects, each one with one of the techniques under study. We expected that the same technique should provide similar results for all the extracts. We observed that the assumption on the coherence of rating is namely verified for a group of listeners, while for the others, strong discrepancies can subsist between same technique but different sequences. This group was composed by listeners that have a particular experience in spatial listening, because they work, as scientist or musicologists, on sound spatialisation.

The sources of such incoherence can be further discussed. First, the semantic of the proposed attributes should be considered. For example, it happened that, even between the experienced 'spatial' listeners, there were misunderstanding on the sense of the 'source width' parameter, which provided exactly opposite answers on this parameters while identical answers for other parameters. The mapping between perceptive attributes and listening cues should be defined on the basis of a common agreement between listeners, in order to setup, if possible, a stable listening framework.

Second, the assumption that the same technique should provide coherent results even if rated over different sound sequences can be criticized. Parameters such as 'timber' and 'pleasantness' can easily introduce bias when considered on different sequences, where the natural dynamic of the performance or the artistic intention can radically change. The nature of the considered sound or the emotion belonging to a specified sequence can significantly influence the perceptual parameter.

Third, we observed two different behaviors in test subjects. One part of the subjects first tried to identify the techniques used in each sequence, then rated the technique, more than the sequence, simply copying the results of one sequence, let us say the binaural recorded, in all the sequences supposed to have been recorded in binaural. This, of course, provides statistic consistency. The only problem is when listeners make mistakes in correctly identifying the pair sequence-technique. A bias can be introduced by the process of pre-evaluation. Another part of the listener subjects simply rated each sequence independently from the other, providing so, in some cases, contradicting results. This can be due to the fact that, in most cases, their judgment was based on comparison between the present sequence and the previous one, loosing then a long term memory that could have preserved global coherence.

These remarks being done, we decided to accept some incongruence between the different sequences ratings (averaging the result of a single technique over all the sequences), but to discard excessive variations. In order to evaluate the frequency of rating incongruence, we choose to compare statistical quartiles as provided by the Matlab 'notched boxplot' function. In a notched box plot the notches represent a robust estimate of the uncertainty about the medians for box-to-box comparison. Boxes whose notches do not overlap indicate that the medians of the two groups differ at the 5% significance level. We fixed a threshold of acceptance discarding boxplots whose notches do not overlap. The little residual variations can be seen as normal fluctuations due to the different sequence audio and emotional content, semantic misunderstanding and listening approach.

In figure 1 we reported an example of discarded subject. On the x axis, the three sequences used in the first 'live recording' test (for binaural recording), on the y axis, the three quartiles computed on an average of the 5 parameters rating. It is possible to observe that boxplots notches do not overlap and then the median can be significantly (as long as a statistic made on 5 samples can be considered as significant) considered as belonging to different groups, that is non unique recording technique. Note that in this way 'nearly constant' answers are

accepted. In figure 2 we plotted an accepted listener (AB recording).

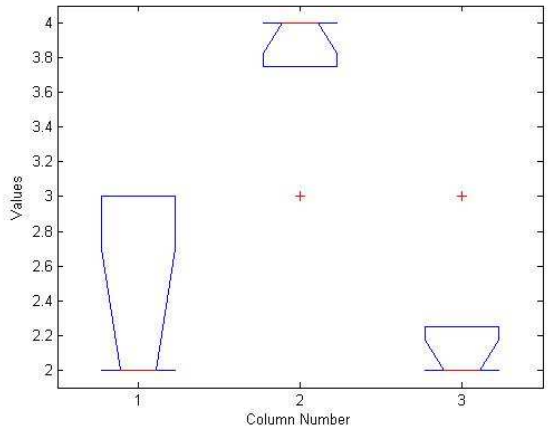

Figure 1: *Discarded subject: on the x axis, the three binaural sequences used in the first 'live recording' test; on the y axis, the three quartiles computed on the average on the 5 parameters rating*

We discarded 2 subjects: the following analysis is then made on the basis of 13 subjects' answering grid, considering the mean ratings computed on the three sequences.

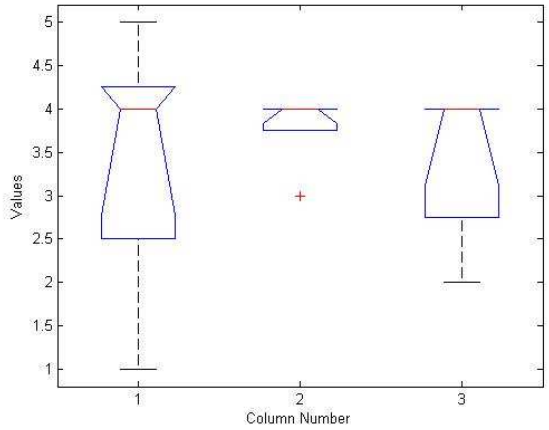

Figure 2: *Accepted subject: on the x axis, the three sequences used in the first 'live recording' test (AB recording); on the y axis, the three quartiles computed on the 5 parameters rating*.

4.2. Analysis of useful results

In figure 3,4,5,6 we report the results of the tests.

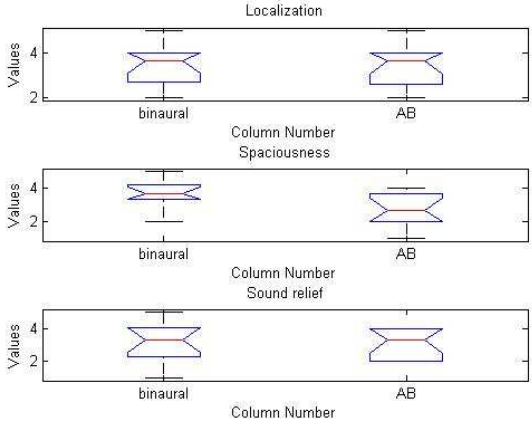

Figure 3: *Localization, spaciousness and sound relief results for the 'live recording' test.*

In figure 3 we plot the results for the 'live recordings' test parameters: localization, spaciousness, sound relief. We remark:

- The spaciousness is, as expected, better for binaural recording, while the other parameters are equivalent. This parameter seems to be statistically consistent, due to the separation of boxplots.
- The sound relief does not seem to be affected by the recording technique: this effect can be due to the simple sound scene that we have recorded which certainly does not present evident sound plans (the guitarist was less than 1 meter behind the singer).
- Localization has been rated in an equivalent way for the two techniques. This could confirm that the spectral cues introduced by dummy head recording (and not present in AB recording, due to the absence of head diffraction) are not enough to enhance localization as individual HRTFs do [9].

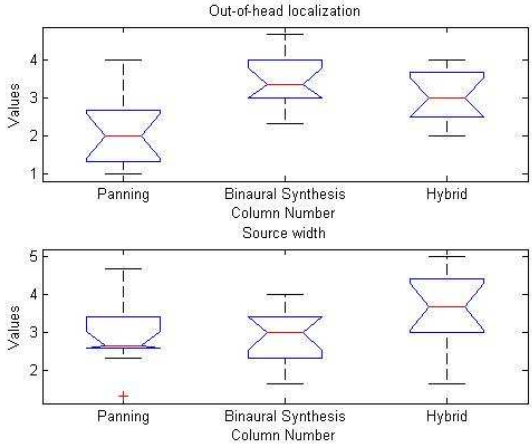

Figure 4: *Out-of-head localization and source width result or 'studio mixes'*

In figure 4 we plot the results for the 'studio mixes' test parameters: out-of-head localization and source width.

We note that :

- Out-of-head localization is enhanced by binaural synthesis and, in a more moderate way, in the hybrid mix. This parameter seems to be statistically consistent, due to the separation of boxplots.
- Source width is better for hybrid mix (but no statistical consistency can be assumed). The similar results for panning and binaural synthesis can be due to misunderstanding or different interpretation of the parameter.

In figure 5 and 6 we plotted the global parameters referring to the five considered techniques. We remark that:

- The timber is correctly reproduced in almost all the techniques. Binaural synthesis obtained worst results: listeners reported a strong comb filtering-like effect at high frequencies. This can be due to problems due to HRTF individualization and equalization. Using diffuse-field equalized HRTFs can be an interesting solution: the inter-differences between HRTFs coming from different subjects are mostly sensitive to high frequencies, for which the ear shape starts to plays a significant role. The attenuation of high frequencies due to diffuse field equalization should at least reduce energy in problematic zones of the spectrum.

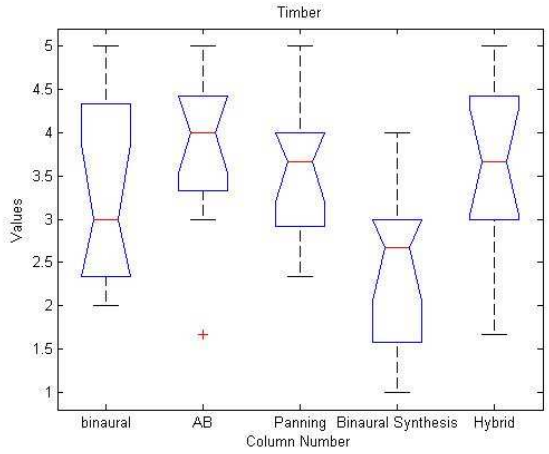

Figure 5*: Timber quality results*

- The listeners found the hybrid mix and the AB recording as the most pleasant techniques, but this is only true for median value analysis: the information is not statistically consistent, because the notches in figure 6 strongly overlap. It seems that spatial enhancement due to binaural technologies and timber deterioration (for binaural synthesis) compensate, 'aligning' in some way the pleasantness for the techniques under study. Alternatively, it can be said that spatial cues enhancement (which is statistically consistent) is not significant in pleasantness enhancement.

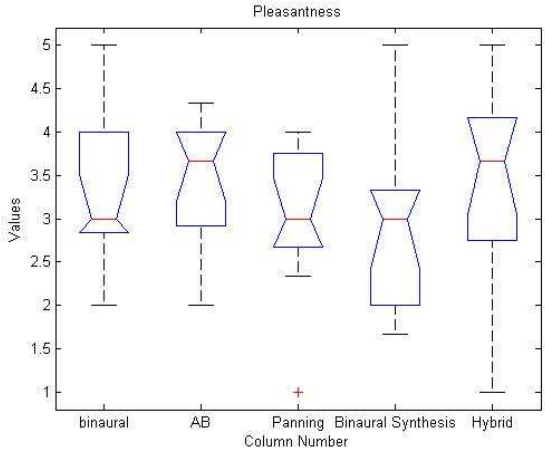

Figure 6: *Pleasantness results*

5. FURTHER STUDIES

A stereo dipole ([3]) system is being installed at the Casa della Musica of Parma: this system is a transaural system that allows, in theory, to perform cross-talk cancellation and then preserve all the characteristics of a binaural recording over loudspeakers. The present test will be re-performed using normal stereo and stereo dipole systems, to identify the eventual deteriorations in reproducing binaural through loudspeakers. Some pioneer tests have been made in this direction.

6. CONCLUSIONS

In this study we aimed to evaluate reactions of experienced audio people to popular binaural music.

The results of perceptual tests on sound-caring listeners show that nowadays they do not seem sensitive to the benefits of binaural technologies: even if binaural techniques were recognized to provide increased spaciousness and out-of-head localization, this has not been a sufficient reason for preferring binaural technologies to stereo ones.

This problem can be due to technological limitations: pop music is basically studio-oriented, due to the possibility of multimic recording and mixing. The studio process is then 'binaural synthesis' oriented and demand for timber precision that is not completely provided by the present HRTF technology. Diffuse-field HRTFs are available [8] and can be tested, but HRTF individualization is a current research problem.

Apart for these technical issues, the feeling is that listeners were just not used to a new approach in listening, that is spatial, and namely, binaural listening. They rated in the same way all the technologies but they recognized, if solicited, the spatial dimension enhancement provided by binaural. Binaural recording and hybrid techniques do not seem to be affected by timber deterioration, and enhance the spatial dimension in hearing. If these techniques have not been chosen as the best

ones, it can be reasonable to think that it is because the spatial dimension that they introduce has not been considered as important by the listeners.

Will popular music listeners become aware of this kind of listening perspectives, maybe through a massive use of their iPods, and some marketing or educational campaign? Maybe the major labels could just produce binaural music (not more expensive than producing normal stereo music) that will be fully enjoyed by 'spatial listeners', keeping intact the listening pleasure of the common listener. The 'spatial listener' can be the added-value market target, but this 'species' should first be created.

Or, as someone said, binaural will remain "just a cool effect with which you can impress people every few years (when it gets rediscovered by someone) - but not make best-selling recording?". Maybe a gradual introduction of binaural cues in recording, through hybrid mixes, could represent a painless intermediary step.

7. ACKNOWLEDGEMENTS

We would really like to thanks Carole Masseport and Stephane Rombi for their patience and their capability of keeping a deep artistic touch in a scientific context. Thanks also to Thomas Van Den Heuvel of VDH recording studios for its essential contribution in the recording and mixing process. Finally, thanks to all the people who gave their feedback.

8. REFERENCES

[1] http://www.binaural.com

[2] G. Theile, "The importance of diffuse-field equalization for stereophonic recording and reproduction", 13 Tonmesteirtagung, 1984.

[3] O. Kirkeby, P. A. Nelson, H. Hamada, "The Stereo Dipole A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers", J. Audio Eng. Soc, 46(5), 1998

[4] Cooper, D. H., and J. L. Bauck, "Prospects for Transaural Recording", J. Audio Eng. Soc., 37(1/2), 1989

[5] J.Jakka, "Binaural to multichannel Audio Upmix", Master Thesis, HUT, 2005

[6] J Huopaniemi, M Karjalainen, "Review of Digital Filter Design and Implementation Methods for 3-D Sound ", AES 102nd Convention, Munich, Germany, Mar, 1997

[7] http://www.qsound.com

[8] http://recherche.ircam.fr/equipes/salles/listen/

[9] P. Minaar, S. Olesen, F. Christensen, H. Moller, "Localization with Binaural Recording from Artificial and Human Heads", J. Audio Eng. Soc., 49(5), 2001