ON THE USE OF BINAURAL RECORDINGS FOR DYNAMIC BINAURAL REPRODUCTION

Reference PACS: 43.60.Dh

Hoffmann, Pablo F.; Christensen, Flemming

Acoustics, Department of Electronic Systems, Aalborg University
Fredrik Bajers Vej 7, DK-9220 Aalborg East, DENMARK
pfh@es.aau.dk, fc@es.aau.dk

ABSTRACT

Binaural recordings are considered applicable only for static binaural reproduction. That is, playback of binaural recordings can only reproduce the sound field captured for the fixed position and orientation of the recording head. However, given some conditions it is possible to use binaural recordings for the reproduction of binaural signals that change according to the listener actions, i.e. dynamic binaural reproduction. Here we examine the conditions that allow for such dynamic recording/playback configuration and discuss advantages and disadvantages. Analysis and discussion focus on two case studies of reproduction of background sound in a car interior and of the rowing sound captured from elite athletes.

RESUMEN

Las grabaciones binaurales se consideran sólo aplicables a la reproducción binaural en condiciones estáticas. Esto se debe a que la reproducción de grabaciones binaurales sólo es válida para la posición y orientación de la cabeza usada durante las grabaciones. Sin embargo, dada ciertas condiciones es posible reproducir grabaciones binaurales que cambien de acuerdo a las acciones del oyente. Este tipo de reproducción lo llamamos reproducción binaural dinámica, y ofrece ciertas ventajas en la reproducción de audio tridimensional de alta fidelidad. En el presente artículo examinamos las condiciones que permiten el uso de reproducción binaural dinámica, y discutimos cuáles son las ventajas y desventajas de esta forma de reproducción de audio tridimensional. Análisis y discusión se dan en el marco de dos casos específicos, el primero correspondiente a la reproducción de ruido de fondo en el interior de un auto, y el segundo correspondiente a la reproducción del sonido de remo producido por remadores profesionales.

1. INTRODUCTION

Binaural technique implementations can be divided into two main categories: Binaural recording and binaural synthesis. Typically, binaural recordings are associated with the capturing of sound events that can be played back and thus accurately reproduce the recording scenario, whereas binaural synthesis is relying on simulations and eventually allowing the listener to move in the simulated scenario. This paper investigates a new idea of a hybrid method allowing listeners to move in recorded binaural scenarios.

1.1 Binaural Recording and Playback

Binaural recordings can be made with microphones mounted in the ear canal of an artificial head, or if the microphones are small enough, in the ear canals of humans. In both cases all sounds contained in a given scenario are captured in the recording process. The positioning of the recording head (human or artificial) defines the perspective of the sound scenario at the time of recording. When the recording is played back over headphones, what was recorded in the ears of the recording head is fed to the ears of the listener in a one to one transmission. In a real life listening situation, movements by the listener will result in a different perspective relative to the individual sound sources i.e. the listener can move in the scenario. When listening to binaural recordings, however, the listener will hear the sound scenarios from listening positions defined by the recording heads position and orientation at the time of recording, and as such eventual movements done by the listener are not reflected in the sound scenario.

1.2 Binaural Synthesis

Binaural synthesis is usually implemented as mathematical convolution of impulse responses with anechoic sound recordings. The impulse responses are measured (or calculated) as either:

- Head Related Impulse Responses (HRIR) i.e. impulse responses measured in the ears of a human or artificial head. To make these responses describe sound from a direction rather than a specific loudspeaker, the impulse response from the sound source to a position in the centre of the head (head absent) are de-convolved.
- or Binaural Room Impulse Responses (BRIR) i.e. impulse responses measured from a sound source in a room to the ears of a human or artificial head.

If HRIRs are collected from many directions, BRIRs can be modelled by a calculated Room Impulse Response where the individual parts (direct sound and individual reflections) are convolved with the HRIRs from the corresponding directions. This allows for the BRIRs to be changed according to listener movements at the time of sound reproduction.

Real sound scenarios are often very complex, and re-creating sound scenarios in this manner is complicated since one BRIR is needed for every different sound source in a sound scenario. Simplifications are often introduced in order to overcome the task e.g. assumption of point sources, and simulation of space distributed sound sources by a limited number of point sources.

Thus, the binaural synthesis allows for the creation of interactive sound scenarios, but the complexity of the sound scenarios i.e. number of individual sound sources and shape of the sound sources are restricted by the computation power among other factors.

1.3 A hybrid method

The idea of the hybrid method is to be able to utilize the advantages of binaural recording in capturing all parts of a sound scenario in one recording process and making the playback system responsive to listener movements to the same degree as in binaural synthesis systems. This idea is further pursued in the following.

2. METHODS

In the following, two different practical cases are described with the aim of studying the hybrid method of Dynamic Binaural Reproduction.

2.1 Case A: Car Audio Simulator

In the development of car audio equipment there is a need to optimize systems by comparing e.g. different loudspeaker types or signal processing algorithms in listening tests. Such tests can be very hard to perform in the real car environment since e.g. different loudspeaker sets would have to be installed in the same car simultaneously. Similarly it is difficult to make a direct subjective comparison of an audio system installed in different cars, due to the unreliability of human auditory memory required when moving from car to car. For this reason a binaural simulation system for car audio equipment was constructed. For an extended description of the system and the related investigations see [1].

The system should simulate automotive audio through headphones with the purpose of making listening experiments in the laboratory. It consists of two main components, one for reproducing the sound of the audio system itself and another for reproducing the background noise in the cabin, which could e.g. mask parts of the music from the audio system and therefore influence subjective judgements.

The use of binaural room synthesis i.e. measurement and application of Binaural Room Impulse Responses are straightforward for the audio system part, and comparable to methods used in e.g. binaural room scanning as described by [2]. A measurement signal is fed into the audio system and recorded with an artificial head at the listeners' position, and the system impulse response is calculated. However, the sound from the audio system is by far not the only one reaching the listeners ears. A car environment has – depending on the type of car – a substantial amount of background noise coming from e.g. the engine, exhaust pipe, tyres, suspension, side mirrors (turbulent air flow) and external noise sources such as sound from other cars, reflections of the cars own emitted noise by nearby objects, rain on the wind shield, etc. The source signal and the transmission characteristics can, for these parts of the sound, not be separated easily. For some parts of the background sounds, the source and transmission path interacts, so the system is non-linear, and the transmission can as such not be described by an impulse response [3]. So modelling such a system would require a complicated mathematical model including both a sound or vibration source and transmission structure.

A way of getting around the modelling problem would be to make a binaural recording of a real system i.e. inside a driving car. A normal binaural recording would facilitate only a fixed head position in the recording and listening situation. This is inadequate since e.g. rotation of a listeners head (in a real life listening situation) could result in one or both ears moving in or out of mode-positions in the sound field and result in different degrees of masking of the music from the audio system. To solve this issue, the binaural reproduction is implemented by recording the cabin noises with the artificial head rotated to various positions and reproducing the recordings through headphones. During playback the listener's head rotations are tracked and both the binaural synthesis (of the car Hi Fi) and the playback of binaural recordings are updated accordingly in real time.

The binaural audio system and cabin synthesis would typically be handled by simultaneous convolution of the sound source signal with the impulse responses of the relevant directions followed by a sort of cross fading to handle transitions between measured directions as the head is turned (illustrated in upper part of figure 1). The same strategy is used for the dynamic reproduction of binaural recordings (lower part of figure 1). However, in order for this to work we should be able to cross-fade between recordings from adjacent directions without a clearly audible transition from one to the other caused by differences in the source signals. Such differences could appear for many types of live signals (e.g. music or speech signals). A transition from one recording to the next can happen at arbitrary periods in time depending on the movement of the listener. The only situation where one can be sure that a live signal will result in the right cross-faded signal is if the signal can be considered stationary (for the time of the recording session). Such signals could for instance be random or quasi-random signals with unchanging timbre as for instance the noise in a car cabin. In the car background noise situation, it should thus be possible to use this strategy.

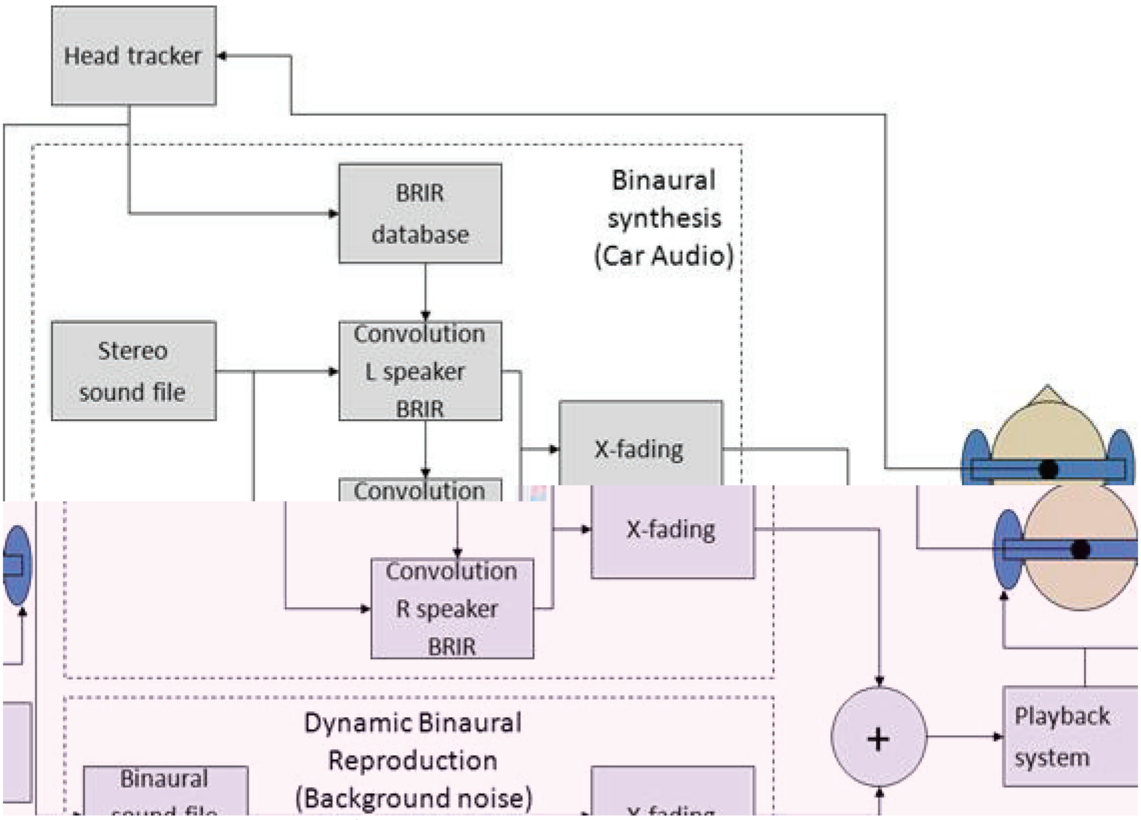

Figure 1 Car audio and background noise simulation system

2.2 Case B: Binaural Sound of Rowing

As part of the implementation of a virtual training environment for rowing [4], a sound rendering module has been included that enables dynamic binaural reproduction of real-life sounds recorded at the ears of experienced athletes.

In addition to the natural sounds that are present in a typical rowing setting, the sound of rowing is mostly dominated by the sound of the rowing stroke. Among the sounds that constitute the sound of the rowing stroke we can identified the sounds from the athlete's movements, splash sounds from the oar blades entering and exiting the water, the propelling sound of the boat as it cuts the water, the mechanical sounds produced by the interaction between the oars and oar locks, and the general transmission of sound through the structure of the boat. Same as for the sound of the car environment, the complex sound of rowing poses a challenge to its rendering if a high level of authenticity is desired. Conventional binaural synthesis and mixing of the individual sources require the rendering of a prohibitive large number of sources. Instead, we analysed binaural recordings that were obtained from experienced rowers. Specifically, binaural recordings were segmented according to velocities, in strokes per minute (SPMs), that span a range between 18 SPMs and 40 SPMs, and stroke phase corresponding to entry, drive, recovery and finish.

The sound rendering system for rowing was implemented as follows. For each velocity, given in SPM, and each stroke phase, three sound excerpts were manually selected from the binaural recordings of the most experienced rower. A database of audio files was constructed using these sound excerpts. Audio files were stored in WAV format (16-bit and 48-kHz sampling rate) and had a fixed duration of 2 seconds. The duration of the longer stroke phase was approximately about 1.2 seconds, and thus each audio file in the database started with the sound of its corresponding phase and the remaining part was filled with environmental sound, i.e. sound from the boat when not in motion or receding in speed, birds, trees moving with the wind, etc. Figure 3 provides a schematic description of the sound rendering system. The sound renderer was synchronized with velocities and rowing phases reported by the rowing simulator. Every 10 ms (100-Hz update) the system received velocity and phase information that used to index and retrieve at random one of the three audio files corresponding to that particular combination of velocity and phase (as an example, the selected sound is represented by the yellow box in the figure). The idea of having more than one sound per velocity and phase was borrowed from audio rendering techniques for games, and the reason was an attempt to avoid making the sound rendering too boring or predictable [6].

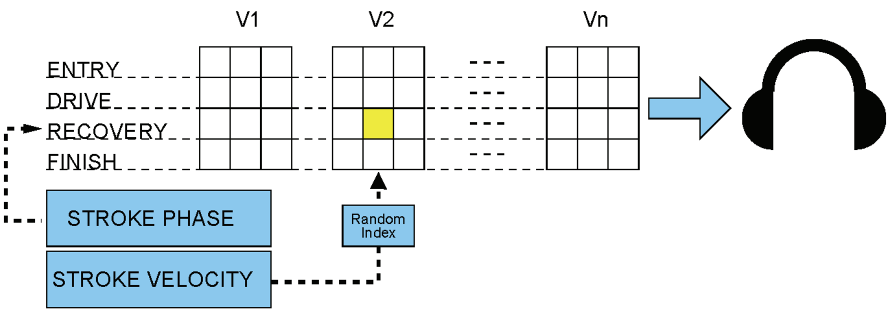

Figure 2 Schematic representation of binaural dynamic reproduction of rowing sounds

The aspect of the rendering system we want to emphasize is that the implementation combines binaural recordings and concatenative synthesis [5] in the sense that different recordings are selected, concatenated and reproduced in response to the user's inputs. It seems that provided the sound to be rendered has a strong cyclic component, as in this case, it is possible to render a realistic simulation of that sound based on binaural recordings. This appears to be particularly advantageous considering that for complex sound fields having numerous sound sources that cannot all be accounted for, or recorded individually, binaural synthesis cannot achieve the same degree of perceived quality as that of binaural recording and reproduction.

3. DISCUSSION

We have presented two cases in which we propose the use of dynamic binaural reproduction for interactive virtual environments. This approach is feasible provided that certain conditions are met. What seems common to the two cases described is that for dynamic binaural reproduction to work one requires that one or some physical dimension in the recording is independent of sound direction. That is, in the first case the main assumption is that timbre in the background noise of a car interior is constant with respect to spatial position. In the second case we assumed that when the sound field is strongly influenced by rhythmic or cyclic components, the binaural recordings can be parsed into segments representing specific static auditory spatial scenes whose sequential concatenation renders the desired sound field, e.g. the phases in rowing stroke.

Because the rendering principle behind dynamic binaural reproduction is based on the concatenation of binaural recordings, dynamic binaural reproduction can reconstruct all the spatial, temporal and qualitative characteristics of the captured sound field in the same way as conventional binaural reproduction. Although we do emphasize the binaural aspect of reproduction in terms of the level of authenticity that can be achieved without much processing on the reproduced signals, it is important to mention that probably the most critical aspect of dynamic reproduction is the procedure used to concatenate the different binaural audio segments. During playback of a single event all the spatial information can be authentically reproduced except at the time the concatenation or cross-fade takes place. That is, it is critical to evaluate whether artefacts produced by the concatenation and looping of the audio segments are not audible, or whether they can be masked by the sounds being concatenated.

A limitation of dynamic binaural reproduction is that it is necessary to impose constraints on the user movements within the virtual auditory environment. That is, user's actions need to be bounded. For example, sound rendering on the rowing simulator works well as long as the simulator provides a valid rowing phase and the user is moving the head backwards and forwards according to the motion required by the correct rowing technique. If the user rotates the head around, the system is not capable to respond to these changes and thus the sound does not update accordingly. Allowing this would require an even bigger data base of recordings, where the expert rower turned the head in several directions at which the recordings were repeated.

We hope to further explore the possibilities that binaural recordings may offer for their implementation in dynamic interactive virtual environments.

4. ACKNOWLEGMENTS

Pablo Hoffmann acknowledges support from the SKILLS Integrated Project (IST-FP6 #035005, http://www.skills-ip.eu) funded by the European Commission.

5. REFERENCES

[1] Christensen, F., Martin, G., Minnaar, P., Song, W., Pedersen, B., & Lydolf, M. (2005). A listening test system for automotive audio-part 1: System description. Proceedings of *AES 118th Convention, Barcelona, Spain. Paper Number 6358.*
[2] Karamustafaoglu A., Horbach U., Pelegrini R., Mackensen P., Theile G. (1999): "Design and Applications of a Data-based Auralisation System for Surround Sound", *Proceedings of 106th Audio Engineering Society Convention, Munich, Germany, May 8-11, 1999, preprint 4976*.
[3] Oppenheim A. V., Schafer, R. W. (1989): "Discrete-time Signal Processing", *Prentice-Hall, Inc. 1989. ISBN 0-13-216771-9*
[4] Ruffaldi E., Filippeschi. A., Avizzano, C. A., Bergamasco, M. (2010) Skill modeling and feedback design for training rowing with virtual environments, *in Proceedings of 3rd International Conference on Applied Human Factors and Ergonomics*, ed. David B. Kaber, Edward P. Fitts, CRC Press / Taylor & Francis, June 2010
[5] Schwarz, D. (2005) Current research in concatenative sound synthesis, *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, Spain, September 5-9.
[6] Vachon, J-F. (2009) Avoiding tedium – Fighting repetition in game audio*, Proceedings of the 35th AES International Conference: Audio for Games*, London, UK, February 11-13.