BINAURAL IN-EAR MONITORING OF ACOUSTIC INSTRUMENTS IN LIVE MUSIC PERFORMANCE

Elías Zea[*]

Sound and Music Computing Group
KTH Royal Institute of Technology
Stockholm, Sweden
ezea.audio@ieee.org

ABSTRACT

A method for Binaural In-Ear Monitoring (Binaural IEM) of acoustic instruments in live music is presented. Spatial rendering is based on four considerations: the directional radiation patterns of musical instruments, room acoustics, binaural synthesis with Head-Related Transfer Functions (HRTF), and the movements of both the musician's head and instrument. The concepts of static and dynamic sound mixes are presented and discussed according to the emotional involvement and musical instruments of the performers, as well as the use of motion capture technology. Pilot experiments of BIEM with dynamic mixing were done with amateur musicians performing with wireless headphones and a motion capture system in a small room. Listening tests with professional musicians evaluating recordings under conditions of dynamic sound mixing were carried out, attempting to find an initial reaction to BIEM. Ideas for further research in static sound mixing, individualized HRTFs, tracking techniques, as well as wedge-monitoring schemes are suggested.

1. INTRODUCTION

Nowadays the amplified playback provided for the musicians on-stage consists in either IEM or floor wedge-monitoring situations. Standard techniques for monitoring live music have traditionally neglected certain aspects of the music-making experience which are relevant and valuable to the performer and audience alike. A good musical interpretation is indeed a complex process in which musical expression, communication, creativity and imagination interact in unfathomable ways and determine the artistic value of the performance. If technology is to be introduced at all in the process, it should be done for the sake of enhancing the enjoyment of the musical experience itself, specially in the case of musicians and/or audiences who are not naturally attracted to the use of technology.

It is worth to remind ourselves of the seemingly trivial -yet extraordinary- fact that the human ear has the natural ability to perceive with impressive accuracy the spatial location and motion of a sound source. In non-amplified situations, this is a natural component of the music-making process. On the other hand, the movements of the musicians and their instruments on stage are normally the spontaneous bodily response to their emotional involvement with the music (e.g. when a player moves its instrument around in various directions, towards or away from other musicians). It is reasonable to think that musicians would like to hear what they are playing as coming from their own instrument (rather than coming from earphones or speakers), and also that the amplified sound conveys their spatial location and movements on stage [1]. So the need arises for monitoring technology to provide an immersive music environment with binaural sound in which the spatialization of the instruments is clearly conveyed. Such spatial information is an essential part of the musical experience which must have an enhancing effect on the artistic value of a live performance.

Nowadays, in the particular case of binaural synthesis, measurements of HRTF [2], [3], [4], among others, provide a method for adding spatial information to a sound source heard by a listener using headphones. It is worth to mention that every individual will have his/her own head-related impulse response. The latter is referred to as individualized response when the HRTF corresponds to measurements of a certain subject. In the present study, the author will refer to such databases as individualized HRTF measurements from human subjects (but not the ones under scrutiny for the tests performed).

The directivity function of an acoustic instrument depends on two parameters: the listener's orientation respect to the frontal axis of a sound source, and the frequency content of the sound heard. Extensive research has been carried out on the directional properties of musical instruments [5], [6], [7] and [8]; yielding reliable models with databases of discrete-sampled radiation patterns [9].

As regards monitoring schemes, Dugan [10] and [11] was the first to do research on automatic mixing models in live music. Work on the musician's sweet-spot in live monitoring has been going on at Queen Mary, University of London [12] and [13]. Among all these studies, gains for each audio channel are presented as a conjunction of matrices for optimum playback from each monitor on stage; since it is of great importance for avoiding any undesirable feedback as well as providing an adequate playback for the musicians.

In 1999, Savioja et al. [14] introduced a framework for an audiovisual immersive and interactive environment with MIDI-based orchestral instruments (DIVA). Savioja's study dealt with the directional characteristics of the musical instruments, the room acoustic response and binaural, transaural and/or multichannel synthesis. Movements of the immersed listener were tracked in order to achieve full spatial exploration of the acoustic environment.

The present paper describes a framework for BIEM of acoustic instruments under live performance conditions. The outline of the paper is described as follows. Section 2 consists of a brief description and motivation of IEM technology, presenting BIEM and backwards compatibility. Section 3 presents the concepts of static and dynamic sound mixes, according to the use of tracking techniques, as well as a broader range of musical instruments, i.e.

[*] IEEE and AES Student Member

non acoustic ones. Directional modeling of acoustic instruments is presented in Section 4, followed by a bilinear interpolation method within a radiation database of orchestral instruments. In Section 5, Pure Data is presented as the main development tool, overviewing the relevant aspects of the code and useful externals for the accomplishment of the framework. Section 6 describes the tests attempted to identify errors in the method, the results obtained as well as possible solutions and improvements. An assessment of possible implications of this technology for further development is presented in Section 7, leading to the conclusions in Section 8.

2. B + IEM = BIEM

In-Ear Monitoring is an alternative method to conventional wedge-monitors, which provides the feedback to the musicians onstage. As the name suggests, IEM involves the use of earplugs. Despite the disadvantage of isolating the musicians from onstage sounds others than those coming from the mix, IEM has the great benefit of making the sound synthesis independent of room acoustics. Actually, IEM is becoming quite popular for reasons of sound quality and hearing protection, and is further justified by the equipment requirements of the musicians [15]. Furthermore, a comparative study of IEM and Wedge monitoring in regard to the musician's perception of latency effect has yielded an encouraging conclusion: latency is weakly perceived with the use of IEM when compared to floor wedges [16].

On the other hand, recording and sound quality reinforcement techniques for live music have improved considerably, producing Hi-Fi models that provide a good basis for backwards compatibility with future technologies. Since one of the goals in audio design is to create software and hardware that is compatible with previous technologies, BIEM (Binaural IEM) presents itself as a good option in this attempt to push forward without loosing backwards compatibility. In the following section, the definitions of static and dynamic sound mixes are presented, pointing out the use of tracking technologies into monitoring schemes and the range of musical instruments.

3. STATIC VS. DYNAMIC SOUND MIXING

Two concepts for sound mix were found by the author in [17]: static and dynamic. The following subsections will overview each of these concepts. The author uses the notation of M for the number of musicians involved in the performance.

3.1. Static Sound Mix (SSM)

In such cases where the musicians do not tend to move on stage (e.g. academic music), the author's hypothesis is that tracking techniques are not necessary to implement [17]. In addition, models for the directional radiation of instruments can be absent in the framework as well, thus the system becomes easier to design and evaluate in real-life situations.

The author presents the spatial matrix, shown in Figure 1, and referred to as $\boldsymbol{S}$ in the study. Let us consider the rider the musicians provide to the sound engineer prior to the live performance. A two-dimensional grid can then be considered (commonly referred to as stage plot), thus an M x 2 matrix is defined in equation 1 with the coordinate pairs of each musician on the stage (x_i, y_i), $i = 1, 2, ..., M$

$$\boldsymbol{S} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_M & y_M \end{bmatrix} \tag{1}$$

where (x_i, y_i) corresponds to the (static) spatial localization of the i-th musician in the stage plot.

The spatial matrix is used to compute the corresponding angles for spatial rendering with HRTFs. The azimuth angle θ_{ij} of musician i relative to musician j is computed in equation 2. Azimuth angle and elevation angle of each musician relative to himself/herself may be set both to zero radians to simplify the problem, though it could be modified according to the musician's preferences.

$$\theta_{ij} = \tan^{-1}\frac{\boldsymbol{S_{i2}} - \boldsymbol{S_{j2}}}{\boldsymbol{S_{i1}} - \boldsymbol{S_{j1}}}, \quad i \neq j \tag{2}$$

The advantages of SSM are that there is no need for tracking the instruments/musicians, and the directional models can be bypassed in the signal processing chain. The latter consideration may incur in a simpler model for binaural monitoring of any instruments/voices, e.g. electronic instruments. In the present document the author will not present the implementation of SSM, but that of binaural monitoring under dynamic conditions.

3.2. Dynamic Sound Mix (DSM)

The author has developed a framework for a monitoring system with motion capture of the musicians' head and instruments [1]. Such a scheme is based on a preliminary work done by Savioja et al. in [14]. The concept of DSM refers to the situations in which musicians are likely to move on the stage (e.g. rock or pop music) and a dynamic mix is designed in order to compensate the rotational motion of the musicians' head and instruments. The main goal is to recreate the acoustic events that occur under non-amplified conditions, and bring them into real-life scenarios.

The range of musical sounds that can be monitored with DSM may be somewhat restricted due to the directional radiation models (mostly acoustic instruments and voices). Whilst SSM is based on a fixed spatial matrix $\boldsymbol{S}$, DSM comprises the use of directional radiation of acoustic instruments, motion capture techniques and the automation of the spatial matrix. The latter is defined via two matrices M x 5, referred to as $\boldsymbol{P}$ and $\boldsymbol{Q}$, containing five degrees of freedom of the performers and the instruments, respectively, necessary to compute the relative orientation for both directional radiation and spatial rendering.

$$\boldsymbol{P} = \begin{bmatrix} x_1^p & y_1^p & z_1^p & \theta_1^p & \psi_1^p \\ x_2^p & y_2^p & z_2^p & \theta_2^p & \psi_2^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_M^p & y_M^p & z_M^p & \theta_M^p & \psi_M^p \end{bmatrix} \tag{3}$$

where (x_i^p, y_i^p, z_i^p) is the tridimensional position of the head of musician i. Variables θ_i^p and ψ_i^p correspond to the azimuth and the elevation angles in radians, respectively, relative to the median plane and the horizontal plane of the head of performer i.

$$\boldsymbol{Q} = \begin{bmatrix} x_1^q & y_1^q & z_1^q & \theta_1^q & \psi_1^q \\ x_2^q & y_2^q & z_2^q & \theta_2^q & \psi_2^q \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_M^q & y_M^q & z_M^q & \theta_M^q & \psi_M^q \end{bmatrix} \tag{4}$$

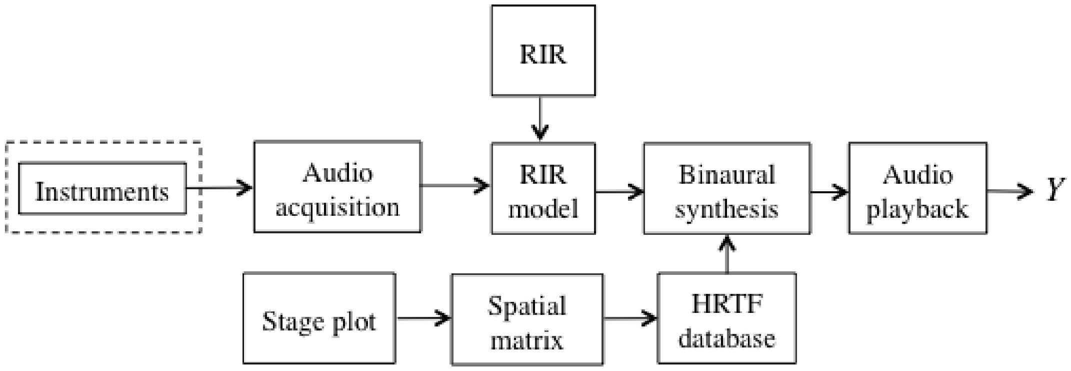

Figure 1: *Block diagram of BIEM with static sound mixing.*

where (x_i^q, y_i^q, z_i^q) is the tridimensional position of instrument i. Variables θ_i^q and ψ_i^q correspond to the azimuth and the elevation angles in radians, respectively, relative to the median plane and the horizontal plane of the instrument i.

In addition to the problem of SSM, azimuth and elevation angles are subdivided into two categories: orientation for directional radiation and orientation for spatial rendering. Equations 5 and 6 define the relative orientation $(\theta_{d_{ij}}, \psi_{d_{ij}})$ for directional radiation of instrument j relative to musician i.

$$\theta_{d_{ij}} = \cos^{-1} \frac{(\boldsymbol{P_{i1}} - \boldsymbol{Q_{j1}}) \cdot \cos \boldsymbol{Q_{j4}} + (\boldsymbol{P_{i2}} - \boldsymbol{Q_{j2}}) \cdot \sin \boldsymbol{Q_{j4}}}{\sqrt{(\boldsymbol{P_{i1}} - \boldsymbol{Q_{j1}})^2 + (\boldsymbol{P_{i2}} - \boldsymbol{Q_{j2}})^2}} \tag{5}$$

$$\psi_{d_{ij}} = \cos^{-1} \frac{(\boldsymbol{P_{i1}} - \boldsymbol{Q_{j1}}) \cdot \cos \boldsymbol{Q_{j5}} + (\boldsymbol{P_{i3}} - \boldsymbol{Q_{j3}}) \cdot \sin \boldsymbol{Q_{j5}}}{\sqrt{(\boldsymbol{P_{i1}} - \boldsymbol{Q_{j1}})^2 + (\boldsymbol{P_{i3}} - \boldsymbol{Q_{j3}})^2}} \tag{6}$$

Likewise, equations 7 and 8 define the orientation $(\theta_{b_{ij}}, \psi_{b_{ij}})$ for spatial rendering of instrument j relative to musician i.

$$\theta_{b_{ij}} = \tan^{-1} \left[\frac{\boldsymbol{P_{i2}} - \boldsymbol{Q_{j2}}}{\boldsymbol{P_{i1}} - \boldsymbol{Q_{j1}}} \right] - \boldsymbol{P_{i4}} \tag{7}$$

$$\psi_{b_{ij}} = \tan^{-1} \left[\frac{\boldsymbol{P_{i3}} - \boldsymbol{Q_{j3}}}{\sqrt{(\boldsymbol{P_{i1}} - \boldsymbol{Q_{j1}})^2 + (\boldsymbol{P_{i2}} - \boldsymbol{Q_{j2}})^2}} \right] - \boldsymbol{P_{i5}} \tag{8}$$

Figure 2 depicts the block diagram of BIEM under conditions of dynamic sound mixing. The framework as well as the tests carried out are described later in greater detail in the present document.

4. DIRECTIONAL FUNCTIONS OF ACOUSTIC INSTRUMENTS

Acoustic musical instruments have a frequency response that is determined by the energy modes resonating in particular regions of their body. This response is obviously not flat, and is in fact a time-varying function that depends also on the orientation and motion of the instrument relative to a listening point. For example, listening to a double bass when facing its front and when facing its back are totally different experiences.

Pätynen constructed a database of radiation patterns of acoustic instruments with 22 microphones positioned in a tetrahedral configuration surrounding the musician; thus responses were found for one-third octave-band frequencies of typical woodwind, brass and string instruments [9]. Averaged data for different tones was computed in order to provide a generalized directivity function of the instruments, depending of three variables: azimuth angle, elevation angle and frequency. The results obtained by Pätynen correspond to the SPLs obtained from each microphone in the array, at 28 discrete frequency bins. In the following section, a computational model of these directivity functions is presented.

4.1. Directional Transfer Functions (DTF)

Sound radiation of acoustic instruments is modeled as an N-length discrete function of frequency, and azimuth and elevation angles: $H[k, \theta, \psi]$, with $k = 0, 1, ...N - 1$. Let us first approximate the instruments as monopole sound sources. As shown in Figure 3, for a certain listening point in space relative to an instrument, there will be a frequency-dependent directional function characterizing the response of the sound radiator. Once orientation is interpolated within the radiation database [9], these frequency representations are referred to as Directional Transfer Functions (DTFs). In this study, the latter consist of discrete values of sound level at $N = 28$ frequency bins, which provide directional radiation to instrument j relative to musician i. The author will refer to such a DTF as an N x 1 vector $\boldsymbol{D}^{(i,j)}$.

$$\boldsymbol{D}^{(i,j)} = H_{ij}[k, \theta_{d_{ij}}, \psi_{d_{ij}}] = \begin{bmatrix} H_{ij}[0, \theta_{d_{ij}}, \psi_{d_{ij}}] \\ H_{ij}[1, \theta_{d_{ij}}, \psi_{d_{ij}}] \\ \vdots \\ H_{ij}[N-1, \theta_{d_{ij}}, \psi_{d_{ij}}] \end{bmatrix} \tag{9}$$

4.2. Bilinear Interpolation of DTFs

Bilinear interpolation is applied to $(\theta_{d_{ij}}, \psi_{d_{ij}})$ at each row of vector $\boldsymbol{D}^{(i,j)}$, in order to find the nearest transfer function from the measurements in [9] for instrument j relative to musician i. The problem can be seen as a two-dimensional grid, shown in Figure 4, with four lattices representing the known radiation functions of

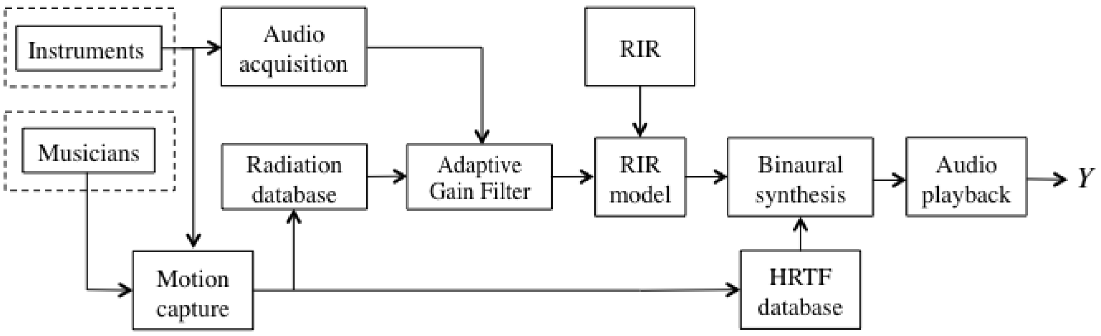

Figure 2: *Block diagram of BIEM with dynamic sound mixing.*

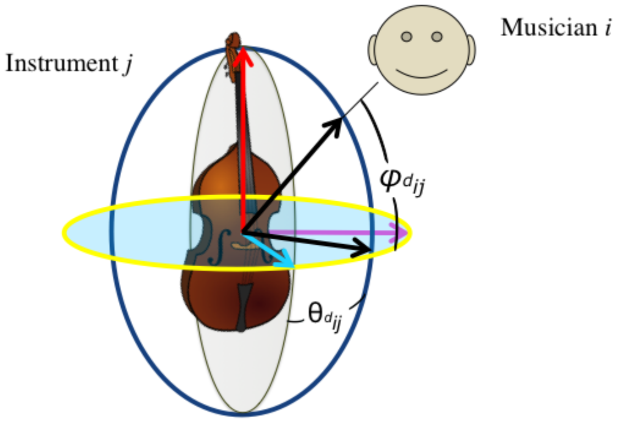

Figure 3: *Diagram depicting directional orientation ($\theta_{d_{ij}}$, $\psi_{d_{ij}}$) of the head of musician i relative to instrument j. The grey circle and the yellow circle are the median and horizontal planes of the instrument, respectively.*

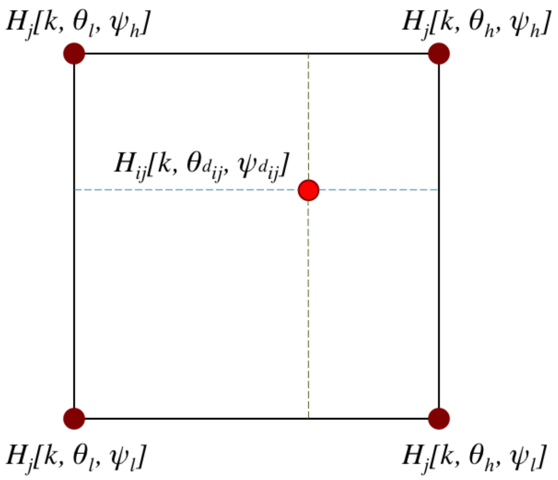
$H_j[k, \theta_l, \psi_h]$
$H_j[k, \theta_h, \psi_h]$
$H_{ij}[k, \theta_{d_{ij}}, \psi_{d_{ij}}]$
$H_j[k, \theta_l, \psi_l]$
$H_j[k, \theta_h, \psi_l]$

Figure 4: *Two-dimensional grid depicting four known DTFs from the database and an unknown function $H_{ij}[k, \theta_{d_{ij}}, \psi_{d_{ij}}]$.*

instrument j from the database at higher- and lower-edges for azimuth and elevation angles

$$\begin{aligned} \theta_l < \theta_{d_{ij}} < \theta_h \\ \psi_l < \psi_{d_{ij}} < \psi_h \end{aligned} \tag{10}$$

Indexes l and h in equation 10 correspond to lower- and higher-edge variables, respectively.

The manipulation of the variables and the microphone arrays done by the author with the database realized by Pätynen is presented in greater detail in [17]. We are interested in computing an unknown DTF, $H_{ij}[k, \theta_{d_{ij}}, \psi_{d_{ij}}]$, of instrument j relative to musician i at a non-sampled angular orientation $(\theta_{d_{ij}}, \psi_{d_{ij}})$. Thus, the interpolation is applied to the vector $\boldsymbol{D}^{(i,j)}$ via equation 11

$$\begin{aligned} \boldsymbol{D}_{\boldsymbol{I}}^{(i,j)} = & H_j[k, \theta_l, \psi_l](\theta_h - \theta_{d_{ij}})(\psi_h - \psi_{d_{ij}}) + \\ & H_j[k, \theta_l, \psi_h](\theta_h - \theta_{d_{ij}})\psi_{d_{ij}} + \\ & H_j[k, \theta_h, \psi_l]\theta_{d_{ij}}(\psi_h - \psi_{d_{ij}}) + \\ & H_j[k, \theta_h, \psi_h]\theta_{d_{ij}}\psi_{d_{ij}} \end{aligned} \tag{11}$$

where index I denotes interpolated vector of $\boldsymbol{D}^{(i,j)}$, and $k = 0, 1, ..., 27$.

5. IMPLEMENTATION OF BIEM WITH DSM

A model of BIEM under conditions of dynamic sound mixing (DSM) was computed in Pure Data (Pd)[1] in order to make pilot tests and listening experiments. Pd is based on graphical language and is an open-source programming tool for real-time signal processing. The following sections present the Pd implementation (hardware and software) of the audio retrieval, playback and the four components of BIEM with DSM. Given the limitations of space and audio channels for playback, $M = 2$ in the present study.

5.1. Audio In: Multi-channel Acquisition

Three microphones were used: one omnidirectional (Behringer ECM 8000) and one violin pickup (Brüel and Kjær 4021). A multi-

[1] See http://puredata.info/

channel audio interface[2] was used to retrieve the signals coming from the microphones. The interface was connected via FireWire to a personal computer. An **adc~** object with 2 channels is used to sample the signals at 44.1 kHz.

5.2. Motion Capture Scheme

A motion capture system with eight infrared cameras was used for the purpose of dynamic tracking [3]. Rigid bodies were constructed for every headphones used for playback, as well as for every instruments involved in the pilot test. These rigid bodies consisted of a set of 3-5 infrared markers asymmetrically positioned in a rigid object. Four rigid bodies were built in total, providing the 5 degrees of freedom for matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, corresponding to performers and instruments, respectively.

Data from the cameras was retrieved with a motion capture software (ARENA [4]), with frame rate of 120 FPS, latency of 8.33 ms and spatial resolution of 1 mm; thus satisfying the requirements for adequate auralization obtained by Sandvad on his study on dynamic tracking for virtual acoustic environments [18].

The data from ARENA was sent to a Pd network patch via Open Sound Control (OSC) by means of a network client [5]. Data sent from the network was read by a receiver in Pd, and unpacked for every rigid body. Scaling is applied to x-, y- and z-axis components, according to the size of the room. Yaw and pitch were transformed into azimuth and elevation angles, respectively.

5.3. Directional Radiation in an Audio Channel

The computation of the directivity functions was done for every instrument j relative to musician i. The latter will have the adequate playback with the directional radiation of the former. Let us denote $x_j[n]$ the signal of instrument j acquired with the audio interface.

The values from matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ were used to compute the orientation for directional radiation $(\theta_{d_{ij}}, \psi_{d_{ij}})$ of instrument j relative to the head of musician i. The orientation was interpolated in order to obtain the DTF that convolves with $x_j[n]$ and outputs the signal $x_{d_{ij}}[n]$. The convolution process is described in the following section.

5.3.1. Adaptive Gain Filter

Once the matrix $\boldsymbol{D}^{(i,j)}$ is computed for instrument j relative to head of musician i, convolution with the signal $x_j[n]$ can be performed in the frequency domain. An adaptive gain filter was coded in Pd with arrays of 28 elements, corresponding to the root mean square sound levels at the frequency bins presented in equation 9 that multiply the audio signal.

Four overlapping Hann window functions were used to create the periodic discrete sequences of $x_j[n]$ for FFT (fast Fourier Transform) convolution. Signal $x_j[n]$ was then converted into the frequency domain via Discrete Fourier Transform (DFT), referred to as $X_j[k']$ with $k' = 0, 1, ..., 4095$.

The value in each row of $\boldsymbol{D}^{(i,j)}$ is raised to four and multiplied by a normalization factor according to the FFT block size and the amount of overlapping windows. Equation 12 defines the filter $\boldsymbol{G}^{(i,j)}$ that multiplies real and imaginary parts of $X_j[k']$, where k' is obtained with a rounding algorithm via equation 13

$$\boldsymbol{G}^{(i,j)} = \frac{1}{6144}\begin{bmatrix} ({\boldsymbol{D}_{\boldsymbol{I}_0}}^{(i,j)})^4 \\ \vdots \\ ({\boldsymbol{D}_{\boldsymbol{I}_{\boldsymbol{N'}-1}}}^{(i,j)})^4 \end{bmatrix} \tag{12}$$

where ${\boldsymbol{D}_{\boldsymbol{I}_{k'}}}^{(i,j)}$ is the k'-th component of the vector $\boldsymbol{D}_{\boldsymbol{I}}^{(ij)}$, with $k' = 0, 1, ..., N-1$ and $N' = 4096$.

$$k' = round\left\{\frac{2N' f_k}{f_s}\right\} \tag{13}$$

We must observe that the size of $\boldsymbol{D}_{\boldsymbol{I}}^{(i,j)}$ is different from that of the vector $\boldsymbol{G}^{(i,j)}$, thus the magnitude of the FFT bins, except the ones corresponding to one-third octave bands, were set to unity. This filter convolved with every channel acquired with the audio interface relative to every listening position.

5.4. Room Impulse Response Model

One of the reasons of choosing binaural synthesis is the fact that the room acoustic response of the place where the live performance is carried out does not significantly affect the signal processing chain; given that near-field microphones are used. Thus, any room impulse response (RIR) can be used for modeling a virtual environment where the musicians are performing.

On the other hand, sound source externalization is one of the biggest challenges in binaural technology. Therefore, an RIR model is introduced to enhance lateral externalization of the sounds played back through headphones. A Pd external object (**partconv~**) and an RIR[6] were used for modeling the virtual acoustic environment.

The convolution was done with an FFT block size of 1024 elements. In the present work, the pilot tests were done in a small room and an RIR of 2 x 3 m^2 is used, given the tracking volume of the motion capture system ($V \simeq 10$ m^3).

5.5. Binaural Synthesis

A set of individualized HRTFs measured from human subjects in [3] was used along with a PD external: **cw_binaural~** [19] to provide the adequate interaural time (ITD) and level differences (ILD) for spatial rendering. It is worth pointing out that the HRTFs of the subjects under test were not measured.

An interpolation in the external is performed to find the HRFT with the relative orientation $(\theta_{b_{ij}}, \psi_{b_{ij}})$ (obtained in equations 7 and 8) which characterizes the spatial rendering of instrument j relative to the head of musician i. The external decomposes the HRTFs into all-pass and minimum-phase components [19].

The incoming signal from the RIR model was passed as input to **cw_binaural~**, which outputs left and right binaural channels for instrument j relative to the i-th musician: $x_{L_{ij}}[n]$ and $x_{R_{ij}}[n]$.

5.6. Audio Out: Multi-channel Monitors

Two arrays, $\boldsymbol{L}$ and $\boldsymbol{R}$, were coded respectively for left and right channels of the contributions of the M instruments relative to the M performers.

[2] See http://www.motu.com/products/motuaudio/traveler-M,k3

[3] See http://www.naturalpoint.com/optitrack/

[4] See http://www.naturalpoint.com/optitrack/products/arena/

[5] See http://old.code.zhdk.ch/projects/gesture/browser/branches/active?rev=2

[6] See http://www.openairlib.net/

$$\boldsymbol{L} = \begin{bmatrix} x_{L_{11}}[n] & x_{L_{12}}[n] & \cdots & x_{L_{1M}}[n] \\ x_{L_{21}}[n] & x_{L_{22}}[n] & \cdots & x_{L_{2M}}[n] \\ \vdots & \vdots & \ddots & \vdots \\ x_{L_{M1}}[n] & x_{L_{M2}}[n] & \cdots & x_{L_{MM}}[n] \end{bmatrix} \qquad (14)$$

$$\boldsymbol{R} = \begin{bmatrix} x_{R_{11}}[n] & x_{R_{12}}[n] & \cdots & x_{R_{1M}}[n] \\ x_{R_{21}}[n] & x_{R_{22}}[n] & \cdots & x_{R_{2M}}[n] \\ \vdots & \vdots & \ddots & \vdots \\ x_{R_{M1}}[n] & x_{R_{M2}}[n] & \cdots & x_{R_{MM}}[n] \end{bmatrix} \qquad (15)$$

An M x 2 playback matrix, $\boldsymbol{Y}$, was coded and each row was passed to $2M$ stereo channels of a **dac~** object, sending them to the interface for every musician. Multi-channel wireless headphones [7] were used for playback.

$$\boldsymbol{Y} = \begin{bmatrix} \sum_{j=1}^{M} \boldsymbol{L_{1j}} & \sum_{j=1}^{M} \boldsymbol{R_{1j}} \\ \sum_{j=1}^{M} \boldsymbol{L_{2j}} & \sum_{j=1}^{M} \boldsymbol{R_{2j}} \\ \vdots & \vdots \\ \sum_{j=1}^{M} \boldsymbol{L_{Mj}} & \sum_{j=1}^{M} \boldsymbol{R_{Mj}} \end{bmatrix} \qquad (16)$$

6. EVALUATION AND ANALYSIS OF THE METHOD

Pilot experiments and listening tests were carried out in order to evaluate the computational model of BIEM with DSM, identify errors and tentative solutions, and an assessment of the initial response of professional musicians to such a monitoring system presented in this study [17]. The following sections overview the design of the experiments, the subjects used and the surveys provided to evaluate the perceived quality of BIEM.

6.1. Pilot Experiments

Four subjects were involved in the pilot tests, grouped in pairs and performing a musical score using two monitoring techniques, referred to as System A and System B. The former technique fed the audio signals of the instruments convolved with the RIR, whereas the latter fed the playback matrix $\boldsymbol{Y}$.

Four amateur musicians (3M, 1F) with ages between 20 and 47 years-old and with a musical experience ranging from 13 to 40 years. Two played the violin, one the oboe and one the bassoon. All the subjects have performed with monitoring systems before.

The score is a trio for oboe, violin and violoncello, composed by Luis Zea. It is a short, tonal piece, easy to read and perform at first sight.

The subjects were asked to perform the following tasks for each of the monitoring techniques used:

- Task 1: Perform the musical score while sitting
- Task 2: Perform the musical score while walking slowly around the sitting position in Task 1

Five questions were asked to the musicians in order to evaluate the pilot experiments. The participants were also asked to include any additional comments regarding the tests. The scale for perceived quality corresponds to the range from "Very little" to "Very much" as from 1 to 7.

1. To what extent each system reproduces the spatial localization of the instruments through your headphones?
2. To what extent each system reproduces your motion and that of the other musician through your headphones?
3. Rate how much attention did you pay to your musical interpretation while playing and using each monitoring system
4. To what extent did each system enhance your musical experience while playing?
5. If you needed to monitor your musical performance, rate how much would you prefer to use systems A and B

The results obtained with the survey provided to the musicians in the pilot experiments are presented in Figure 5. None of the questions, except for number 5, provide a significant result given the high standard deviations that overlap among Systems A and B. We clearly see that musicians would rather to use System A for their live performance. However, all of the participants referred to System B as presenting terrible glitches while performing the second task. The author attributes these results and comments to flaws in the tracking system.

On the other hand, an hypothesis was established by the author in [17], related to the musicians' awareness of tridimensional sound. One of the biggest challenges of using binaural technology with music is that listeners and/or performers are not used to the experience. In this way, we can observe a tendency of System B in Figure 5, from Question 1 to Question 2, that the musicians perceived more the motion of the sounds than the localization. This slightly 14% difference, which reciprocally decreases in 10% for System A, is attributed to the fact that musicians did not fully understand what the playback of System B was about before Question 2. The author's reasoning is that performing experiments with musicians that are not aware of binaural technology was not the correct approach to validate the BIEM scheme. It is likely that musicians would have a better opportunity to evaluate the usefulness of BIEM if they had known before hand that System B simulated a spatial rendering of the instruments. The author also considers the survey could have been designed in a different way.

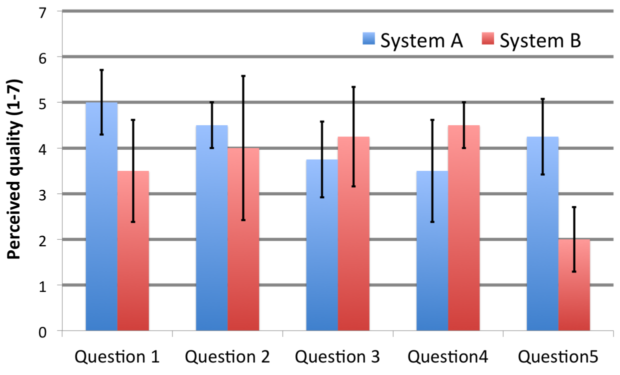

Figure 5: *Results for perceived quality obtained in the questions of the pilot tests for System A and System B.*

6.2. Listening Tests

Two recordings were performed during the pilot experiments, referred to as Version A and Version B, corresponding to the musicians performing with systems A and B, respectively. A violinist

[7] See http://www.sennheiserusa.com/wireless-audiophile-headphones-rs220-502029

and a bassoonist were recorded while performing an alternative version of the tasks in the pilot experiments. The playback sent to the violin player was the recorded signal, and the violin player was moving whilst the bassoonist remained seated. An audio file was published in Vimeo [8], putting together Version A and Version B.

Twenty three (23) professional musicians (15M, 8F), with averages of 29 years of musical experience and age of 46, were asked to participate in the listening test with the audio file and a survey. All of the subjects have performed live music with monitoring systems.

Likewise, a survey was designed for the listening tests. However in this case, the questions were geared towards an assessment of the initial response of professional musicians to BIEM, and they were encouraged to provide additional comments on the applicability of this technology. The participants were told to rate their perceived quality in a scale from "Very little" to "Very much" ranging from 1 to 7. The following questions were asked for Versions A and B:

1. Rate to what extent the monitoring system in each version reproduces (through your headphones) the location of the instruments in space
2. Rate to what extent the monitoring system in each version reproduces (through your headphones) the movements of the instruments in space
3. Rate to what extent did the monitoring system in each version enhance your musical experience while listening
4. If you needed to use amplification for your musical performance, rate how much would you prefer to use each monitoring system

The results of the survey provided to the professional musicians are shown in Figure 6. We might mention the survey was not correctly designed given the large error bars. Only Question 2 provides a significant result, which refers to a perceived quality of motion of the instruments about 70% higher in Version B than in Version A. Nevertheless, the additional comments provided by the subjects strengthen the author's opinion that musicians do not know about 3-D sound. Some musicians confused spatial rendering with panning, mono with stereo, among many other comments. It is hard to answer questions concerning the potential of BIEM unless a minimum of knowledge about it exists. Therefore a better explanation of the System B was needed.

An interesting comment left by two professional singers was the applicability of BIEM for singing choirs. These subjects suggested the use of spatial rendering of voices under monitored conditions. In addition, one musician mentioned that musical creativity may be stimulated with the use of spatialized sounds; even though the subject did not recognize the use of binaural audio.

7. GENERAL DISCUSSION

The author believes that the implementation and validation of Binaural IEM under conditions of dynamic mixing becomes difficult in real-life situation (e.g. a concert). In this way, the first consideration for future research is the evaluation of BIEM with static sound mixing (SSM). The latter, as overviewed in the present paper, is easier to implement and evaluate in live performances than with DSM. This could take us to an assessment of binaural technology with a broader range of musical instruments than in the DSM scheme.

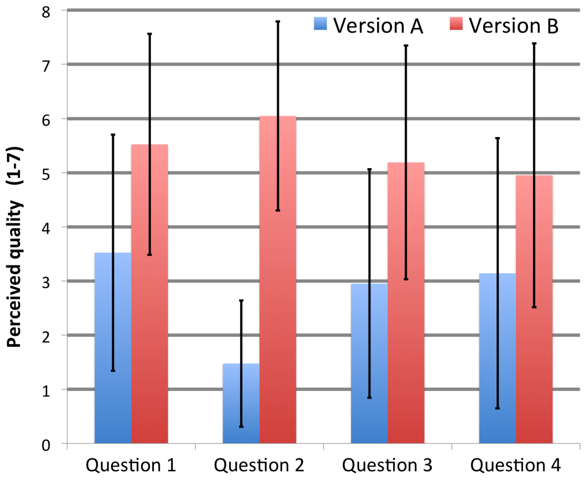

Figure 6: *Results for perceived quality obtained in the questions of the listening tests for Version A and Version B.*

On the other hand, the author suggests the incorporation of machine learning with the anthropomorphic data of the performers under test. The individualized HRTF databases (e.g. CIPIC and/or Listen), which provide anthropomorphic data of the subjects whose response was measured, can be trained with an unsupervised machine learning algorithm. Multiple features can be handled, and the most suitable HRTF for a given musician can be found from the database via Support Vector Machine Classification (SVM). Thus, aural perception and sound externalization in the playback may be greatly enhanced.

In addition, another alternatives of tracking mechanisms can be taken into consideration. The author encourages future studies with accelerometers and gyroscopes, given the flexibility, accuracy and low-cost of such components.

We can also start thinking of transaural audio with crosstalk cancellation. With this approach, the room acoustic model is different from the one presented in the paper, due to the fact that the wedge-floor monitors have a directivity response, and complex reflection patterns are likely to occur with the surfaces of the room.

8. CONCLUSIONS

A methodology for Binaural In-Ear Monitoring (BIEM) of acoustic instruments was presented. The importance of including spatial information in live music performance was discussed, since it has been neglected in traditional monitoring technologies and may enhance the music-making experience and the artistic value of live performances. A brief description and justification of In-Ear Monitoring technology was included, as well as a discussion of backwards compatibility of BIEM. The definitions and block diagrams of static and dynamic sound mixing were presented in detail, pointing out the applicability of binaural audio according to emotional involvement and instruments, as well as the use of motion capture techniques. By means of Pätynen's database [9], a computational model of directional transfer functions was presented, followed by

[8] See https://vimeo.com/43886864

a bilinear interpolation algorithm to provide radiation directivity to a given acoustic instrument relative to a musician in space.

A computational method of BIEM under conditions of DSM with Pure Data was presented, describing the signal processing chain from audio acquisition to playback. Evaluation of the method was done via pilot experiments with amateur players and listening tests with professional musicians. From such tests, the most remarkable results point towards unsatisfactory design of the questions in the surveys and complete unawareness of the musicians of binaural sound. The participants of the pilot tests would prefer to use a conventional monitoring technique than BIEM. Some musicians that performed the listening tests confused spatial rendering with panning, although a higher perceived quality of the movements of the instruments was found with BIEM than with conventional monitoring. Flaws in the motion capture system were significantly affecting the quality of the playback in the pilot tests. The author suggests that the musicians must know about binaural sound prior to evaluate BIEM.

The author left further development of BIEM under conditions of static mixing. A machine learning algorithm might be useful to find an adequate individualized HRTF for the performers using BIEM. Other tracking techniques are suggested for BIEM with DSM, as well as future insights on transaural audio with crosstalk cancellation techniques.

9. ACKNOWLEDGMENTS

Thanks to Damian Murphy, Roberto Bresin, Anders Askenfelt, Sten Ternström, Anders Friberg and Kjetil Hansen for the motivation, knowledge and support.

Thanks to Jukka Pätynen for his support in providing the directivity functions of orchestral instruments.

Special thanks to my wonderful family, whose unconditional support, artistic professionalism and passionate commitment to their art have encouraged me to think of such fascinating questions in live music performance.

Acknowledgments to the Regents of the University of California for providing permission to make use of the CIPIC HRTF database in the present study.

10. REFERENCES

[1] E. Zea, “A framework for spatial rendering of amplified musical instruments,” in *Proc. Special issue of the Speech, Music and Hearing*, Sweden, April, 2012, pp. 46–47.

[2] B. Gardner and K.D. Martin, “HRTF measurements of a KEMAR,” *J. of the Acoustical Soc. of America*, vol. 97, no. 6, pp. 3907–3908, June, 1995.

[3] V. R. Algazi et al, “The CIPIC HRTF database,” in *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, Mohonk Mountain House, New Paltz, NY, Oct., 2001, pp. 99–102.

[4] AKG and IRCAM, “Listen HRTF database,” Available at Room Acoustics Group IRCAM, Paris, France, accessed January 20, 2012.

[5] R. Caussé, J. Bresciani, and O. Warusfel, “Radiation of musical instruments and control of reproduction with loudspeakers,” in *Proc. of the International Symposium on Musical Acoustics*, Tokyo, 1992.

[6] G. Weinreich, “Radiativity revisited: theory and experiment ten years later,” in *Proc. of the Stockholm Music Acoustics*, Sweden, 1994, p. 436.

[7] G. Weinreich, “Directional tone color,” *J. of the Acoustical Soc. of America*, vol. 101, no. 4, pp. 2338–2346, April, 1997.

[8] J. Pätynen, V. Pulkki, and T. Lokki, “Anechoic recording system for symphony orchestra,” *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 856–865, Dec., 2008.

[9] J. Pätynen and T. Lokki, “Directivities of symphony orchestra instruments,” *Acta Acustica united with Acustica*, vol. 96, no. 1, pp. 138–176, Feb., 2010.

[10] D. Dugan, “Automatic microphone mixing,” *J. Audio Eng. Soc*, vol. 23, no. 6, pp. 442–449, 1975.

[11] D. Dugan, “Application of automatic mixing techniques to audio consoles,” in *AES Convention 87*, Oct., 1989.

[12] J.D. Reiss and E. Perez Gonzalez, “An automatic maximum gain normalization technique with applications to audio mixing,” in *AES Convention 124*, May, 2008.

[13] M.J. Terrell and J.D. Reiss, “Automatic monitor mixing for live musical performance,” *J. Audio Eng. Soc*, vol. 57, no. 11, pp. 927–936, 2009.

[14] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, “Creating interactive virtual acoustic environments,” *J. Audio Eng. Soc*, vol. 47, no. 9, pp. 675–705, 1999.

[15] S. Armstrong, K. Gordon, and B. Rule, “Assessing the suitability of digital signal processing as applied to performance audio such as in-ear monitoring systems,” in *AES Convention 119*, Oct., 2005.

[16] M. Lester and J. Boley, “The effects of latency on live sound monitoring,” in *AES Convention 123*, Oct., 2007.

[17] E. Zea, “Binaural monitoring for live music performances,” M.S. thesis, KTH Royal Institute of Technology, 2012.

[18] J. Sandvad, “Dynamic aspects of auditory virtual environments,” in *AES 100*, May, 1996.

[19] D. Doukhan and Sédès, “CW_binaural∼: A binaural synthesis external for Pure Data,” in *3rd. PD International Convention in Säo Paulo*, 2009.

[20] Wikipedia, “Bilinear interpolation,” Available at Bilinear Interpolation, accessed February 23, 2012.