

# Dynamic measurement of room impulse responses using a moving microphone

Thibaut Ajdler,<sup>a)</sup> Luciano Sbaiz, and Martin Vetterli<sup>b)</sup>

*Audiovisual Communications Laboratory, Ecole Polytechnique Fédérale de Lausanne (EPFL),  
1015 Lausanne, Switzerland*

(Received 10 October 2006; revised 24 March 2007; accepted 3 July 2007)

A technique for the recording of large sets of room impulse responses or head-related transfer functions is presented. The technique uses a microphone moving with constant speed. Given a setup (e.g., length of the room impulse response), a careful choice of the recording parameters (excitation signal, speed of movement) leads to the reconstruction of all impulse responses along the trajectory. In the case of a moving microphone along a circle, the maximal angular speed is given as a function of the length of the impulse response, its maximal temporal frequency, the speed of sound propagation, and the radius of the circle. As a result of the presented algorithm, head-related transfer functions sampled at 44.1 kHz can be measured at all angular positions along the horizontal plane in less than 1 s. The presented theory is compared with a real system implementation using a precision moving microphone holder. The practical setup is discussed together with its limitations. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2766776]

PACS number(s): 43.55.Mc, 43.20.Ye, 43.20.Bi, 43.60.Ac [WMC]

Pages: 1636–1645

## I. INTRODUCTION

One of the key elements in the success of multimedia communication systems is the capability of reproducing virtual environments to be perceived by the listener as naturally as possible. To achieve realism, either real measurements need to be introduced in the system, or the environment needs to be modeled and synthetically generated.<sup>1,2</sup> In both cases, room impulse responses (RIRs) are used to give listeners the impression of being in a real environment and enveloped by the sound. In the case of headphone playback, head-related transfer functions (HRTFs) are usually added to achieve externalization and more realistic impressions.<sup>3</sup> The RIRs and HRTFs to be used need to be either measured at a large number of positions in the considered space or modeled using different techniques such as the ones described in Ref. 4 or 5. The RIRs are then provided to the playback system. This is typically the case in wave field synthesis (WFS) systems where measured RIRs at a large number of positions<sup>6,7</sup> are used to increase the realism of the reproduced field. Therefore, one would like to find a way to easily and rapidly measure large sets of RIRs. The usual technique is to use a single microphone or a microphone array. Nevertheless, to capture hundreds of RIRs, the array of microphones needs to be displaced to several positions. The intrusion of a person to modify the setup (e.g., displace the array) changes greatly the characteristics of the room and the temperature field inside of the room<sup>8</sup> and makes the measurement very time consuming. In this paper, a technique is introduced that achieves a fast recording of a large number of RIRs. It considers a fixed loudspeaker with a moving microphone. The moving element follows a trajectory (e.g., circular trajectory) with con-

stant speed. Also, the acquisition of the data is not done position after position but happens continuously along the trajectory. The movement is uniform and does not stop during the acquisition. Thanks to this setup, one avoids problems linked to abrupt stops leading to oscillations and waiting time for the microphone to get to its next position. From the one-dimensional signal gathered by the microphone, the two-dimensional (2D) dataset (spatial and temporal) containing the RIRs at all the different spatial positions along the trajectory is reconstructed. The algorithm takes into account the Doppler shift inherent to the moving element and cancels its effect in the reconstructed RIRs. An analysis is performed to study the influence of the different parameters to be chosen so as to achieve the reconstruction of the RIRs (e.g., length and period of the excitation signal, frequencies contained in the excitation signal, speed of movement of the microphone, temporal and spatial frequencies of the reconstructed dataset). The trade-off existing between the speed of the moving element and the spacing between the frequencies contained in the excitation signal is discussed.<sup>9</sup> The presented theory is shown together with real measurements. These measurements are obtained using a moving microphone holder that achieves a precision of a few hundredths of a degree when rotating in the horizontal plane. An interesting application of this setup can be found in the measurement of HRTFs. These measurements are typically done in anechoic chambers to describe the influence of our body on the sound measured at the entrance of our ears.<sup>3</sup> For these specific filters, the impulse responses to be measured are very short (on the order of a few milliseconds) since no room reflections need to be captured. The whole dataset of azimuthal angles can then be recorded in a very fast manner. In this paper, it is shown that with the presented technique, measurement of all HRTFs in the horizontal plane can be achieved in less than 1 s.

<sup>a)</sup>Electronic mail: [ajdler@gmail.com](mailto:ajdler@gmail.com)

<sup>b)</sup>Also at Department of EECS, University of California at Berkeley, Berkeley, CA 94720.

## A. Related work

Different methods exist for the measurement of RIRs and HRTFs. To measure a RIR, the typical setup considers a calibrated loudspeaker and microphone. By measuring the signal recorded by the microphone, it is possible to obtain the room impulse response. Different excitation signals can be provided to the loudspeaker such as logarithmic sweeps<sup>10</sup> or maximum length sequences.<sup>11</sup> The simplest and most common technique for HRTF measurement consists in putting a microphone at the entrance of the ear canal and recording the sound emitted by a source located at a certain azimuth and elevation angle. By comparing the recorded signal and the emitted one, it is possible to extract the influence of the listener. This procedure needs to be repeated for every desired direction as shown in Fig. 1. Different stimuli can be used for the purpose of measuring HRTFs. White Gaussian noises, MLS, or Golay codes<sup>12</sup> are often used. Comparison of these signals can be found in Ref. 13. To reduce the time needed for the acquisition of large datasets of HRTFs, an interesting method has been presented by Zotkin *et al.*<sup>14</sup> By putting a loudspeaker emitting sound near the ear entrance, the sound pressure field is recorded with receivers located on a circular or on a spherical array of microphones. Like this, all HRTFs can be measured simultaneously. A third technique is described in this paper where a fixed source and a moving person are considered. In the proposed algorithm, it can be shown that the measurement of all possible HRTFs along the horizontal plane can be achieved in less than 1 s. Note that a similar approach was followed in Refs. 7 and 15 where room impulse response measurements were carried out using a microphone moving along a circular trajectory. Nevertheless, the assumption was made that at low angular speed, the RIRs can be obtained without taking the Doppler effect into account. This is not the case in our study where the RIRs are theoretically perfectly reconstructed. Other methods using moving microphones have been presented for improving wave field extrapolation techniques. As is described in Ref. 16, the first step is to measure the field along a certain trajectory, and, from the gathered information, extrapolation can be performed using the Kirchhoff-Helmholtz integral.<sup>17</sup> Considering the field to be extrapolated as stationary, the different measurements do not need to be performed simultaneously. In Refs. 18 and 19 it is explained that using

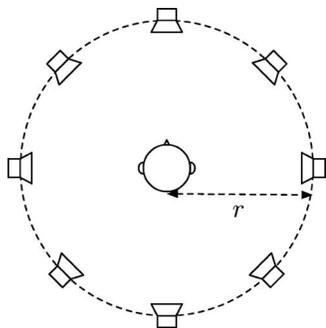


FIG. 1. Setup for the recording of HRTFs.

a moving microphone allows for measurement of the field in more positions and therefore better extrapolation can be achieved.

## B. Contributions

The contributions of this work are the following:

- (i) The Doppler effect is put in evidence in the two-dimensional spectrum representation of the sound pressure field when a moving receiver is used in the setup.
- (ii) An algorithm for the reconstruction of the static RIRs or HRTFs is presented by removing the Doppler effect present in the recording.
- (iii) A formula describing the maximal rotation speed is given as a function of the different parameters of the system.
- (iv) The measurement of HRTFs at every azimuthal angle along the horizontal plane can be achieved in less than 1 s at a sampling frequency of 44.1 kHz.
- (v) Experimental measurements using a moving microphone demonstrate the applicability of the algorithm and confirm the presented theory.

## C. Outline

The outline of the paper is as follows. Section II describes some existing results on the study of the sound pressure field gathered along a line and a circle. The two-dimensional Fourier transform (2D-FT) of the considered sound pressure field is described for the two setups in Secs. II A and II B. Further, Sec. II C briefly recalls the Doppler effect. Section III considers the setup with a microphone moving along a trajectory with a uniform speed. In this scenario, the Doppler effect needs to be considered. It is shown in Sec. III A that, using the 2D spectrum representation of the sound pressure field along a line, the Doppler effect can be put in evidence. The signal recorded by a microphone moving along a circle is then described in Sec. III B. Section IV presents the main result of this work. It develops an algorithm to reconstruct the RIRs at every possible position, from the gathered signal by the microphone. The technique to achieve this reconstruction is presented in Sec. IV A. Some further remarks are given in Sec. IV B. A performance analysis comparing the signal to noise ratio of both the dynamic and the static measurement setups is described in Sec. V. The theory is then compared with experimental measurements in Sec. VI. The conclusions are drawn in Sec. VII.

## II. SPATIO-TEMPORAL IMPULSE RESPONSES

In this section, existing results are recalled on the study of the 2D-FT of the sound pressure field along different geometries. First, the spectrum of the sound pressure field along a line in the room is presented in Sec. II A followed by the sound pressure field along a circle in Sec. II B. Finally, Sec. II C briefly reviews the Doppler effect widely used in the rest of the paper.

## A. Sound field along a line

Consider the sound pressure field studied along a line in a room. As a first example, consider a plane wave of temporal frequency  $\omega_0$  arriving on the line with an angle  $\alpha$ ,

$$p(x,t) = e^{j(\omega_0 t + k_0 x \cos \alpha)}, \quad (1)$$

with  $k_0 = \omega_0/c$  and  $c$  the speed of sound propagation as shown in Fig. 2. The 2D-FT of a function  $p(x,t)$  is defined as follows

$$\hat{p}(\phi, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x,t) e^{-j(\omega t + \phi x)} dt dx, \quad (2)$$

with  $\phi$  and  $\omega$  the spatial and temporal frequencies respectively.

In the case of Eq. (1), its 2D-FT is<sup>17</sup>

$$\hat{p}(\phi, \omega) = 4\pi^2 \delta(\omega - \omega_0) \delta\left(\phi - \frac{\omega_0 \cos \alpha}{c}\right). \quad (3)$$

We therefore see that the 2D-FT of a plane wave of temporal frequency  $\omega_0$  only has energy in one single point in the spectral domain.

Considering a plane wave emitting all possible frequencies, Eq. (1) becomes

$$p(x,t) = \delta\left(t + \frac{x \cos \alpha}{c}\right). \quad (4)$$

The 2D-FT of Eq. (4) is then

$$\hat{p}(\phi, \omega) = 2\pi \delta\left(\phi - \frac{\omega \cos \alpha}{c}\right). \quad (5)$$

This 2D spectrum is shown in Fig. 3(a). When plane waves arrive from all possible angles, the support of the spectrum is represented in Fig. 3(b) and satisfies the rule

$$|\phi| \leq \frac{|\omega|}{c}. \quad (6)$$

As can be seen, the support has a butterfly spectrum, which can intuitively be explained as follows. At low temporal frequencies, the sound wavelength is very large and therefore, spatially, the wave is varying slowly, which explains the small spatial support. For higher temporal frequencies, the sound wavelength is smaller, which makes the spatial variation larger. The spatial support for higher temporal frequencies is therefore larger. As a result, when sampling the sound

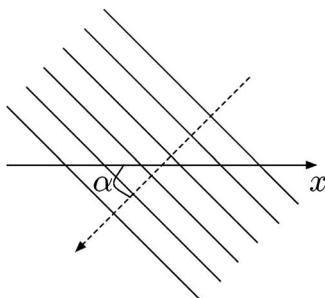


FIG. 2. A plane wave arrives on a line of microphone along the  $x$  axis with an angle of arrival of  $\alpha$ .

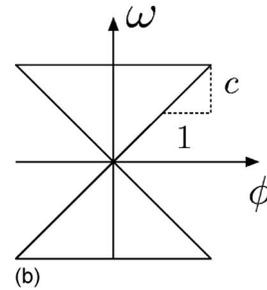
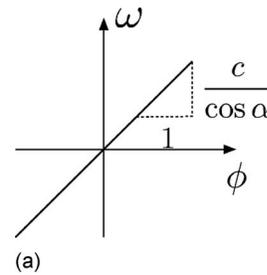


FIG. 3. Two-dimensional spectrum of the sound pressure field, (a) In case of a single plane wave, (b) In case of plane waves arriving from all possible angles.

field up to a temporal frequency of  $\omega_{\max}$ , the necessary spacing between the microphones needed to interpolate the field at any position along a line is of  $D$ , with

$$D < \frac{\pi c}{\omega_{\max}}. \quad (7)$$

Further work on the sampling and interpolation of the sound field has been done when considering not only the plane wave assumption but the real solution of the wave equation. More details can be found in Refs. 20 and 21.

## B. Sound field along a circle

For the study of the sound pressure field along a circle, consider the scheme presented in Fig. 4(a). Consider a circular microphone array of radius  $r$ . The coordinates of the different microphones are  $(m_x, m_y, m_z)$ , with  $m_x = r \cos \theta$ ,  $m_y = r \sin \theta$ , and  $m_z$  constant. The sound source has coordinates  $(s_x, s_y, s_z)$ . The time of arrival from the source to the receiver is given by the following expression:

$$h(\theta) = \frac{\sqrt{(s_x - r \cos \theta)^2 + (s_y - r \sin \theta)^2 + (s_z - m_z)^2}}{c}. \quad (8)$$

The free field pressure recorded on the circle of microphones due to the emission of a Dirac by the source is given by<sup>22</sup>

$$p(\theta, t) = \frac{\delta(t - h(\theta))}{4\pi c h(\theta)}. \quad (9)$$

Remark that  $p(\theta, t)$  is  $2\pi$  periodic with respect to  $\theta$ . The 2D-FT of Eq. (9) is denoted as  $\hat{p}(l_\theta, \omega)$ . Note that  $l_\theta$  is the index of the Fourier series with respect to the  $\theta$  axis. In Refs. 23 and 24, it has been shown that most of the energy of the spectrum is located into a butterfly region spectrum satisfying

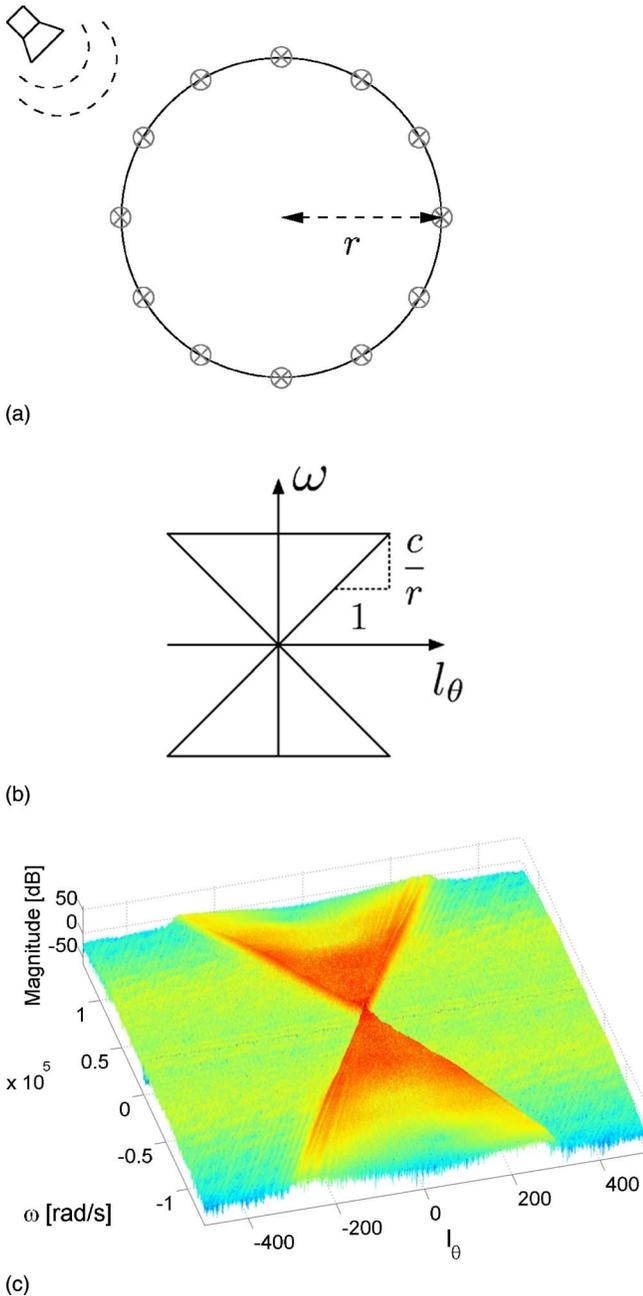


FIG. 4. (Color online) Sound field along a circle, (a) Schematic view of the setup, (b) Support of the 2D-FT of the sound pressure field analyzed on a circle. (c) 2D-FT of room impulse responses measured on a circle.

$$|l_\theta| \leq |\omega| \frac{r}{c}. \quad (10)$$

The support of the spectrum is schematized in Fig. 4(b). The intuition behind the butterfly spectrum is exactly the same as the one discussed in Sec. II A. Considering that no energy is present outside of the butterfly region, the angular spacing  $\Delta\theta$  necessary to interpolate the sound field up to  $\omega_{\max}$  at any azimuthal angular position is given by

$$\Delta\theta < \frac{\pi c}{r\omega_{\max}}. \quad (11)$$

Figure 4(c) shows an example of the 2D-FT of RIRs measured on a circle. These measurements were taken every

0.36 deg along a circle of radius 0.55 m. The details of the measurements as well as the database are available online.<sup>25</sup>

### C. Doppler effect

In Secs. II A and II B the sound pressure field is considered to be measured at every positions along a line or a circle. In this paper, these measurements are obtained by using a moving microphone. When considering a moving element, a frequency shift is observed on the recorded signal. This is known as the Doppler effect and can be expressed by the following formula:<sup>26</sup>

$$\omega = \omega_0 \left( \frac{c + v_r}{c + v_s} \right), \quad (12)$$

with  $c$  the speed of sound propagation,  $\omega$  the observed frequency,  $\omega_0$  the emitted frequency,  $v_s$  the speed of movement of the source, and  $v_r$  the speed of movement of the receiver. A convention for the sign of the speeds in Eq. (12) needs to be adopted. The positive direction is considered as the direction “listener towards source.” The setup considering a moving microphone is discussed in Sec. III. There, using this convention, the sign of the speed of the receiver is considered as positive when the source and the receiver are coming towards each other as shown in Fig. 5(a), The sign of the received speed is negative when source and receiver are moving apart as in Fig. 5(b).

### III. RECORDED SIGNAL FOR A MOVING MICROPHONE

In Sec. II, the sound pressure field has been described along a line and a circle. For both cases, the sound pressure field has been shown to have a butterfly spectrum. This specific shape of the spectrum will be used to study the sound field gathered by a microphone in the presence of a moving microphone.

Note that in the rest of this paper, we consider the loudspeaker and microphone to be omnidirectional devices and to present flat frequency responses. In a more realistic setup, the loudspeaker and microphone should be calibrated to be able to remove their influence on the impulse responses.

In this section, the moving microphone signal is studied when the microphone is moved along a line in Sec. III A and along a circle in Sec. III B.

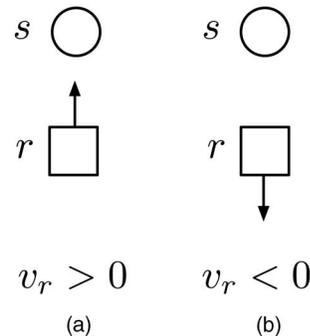


FIG. 5. Different setups for the study of the Doppler effect.

## A. Microphone moving along a line

Consider a setup with a fixed loudspeaker and a microphone moving with a constant speed of  $v$  m/s along a line. Due to the constant speed of the microphone movement, the position  $x$  along the line is proportional to the time  $t$  as

$$x = vt. \quad (13)$$

The sound pressure at all positions along a line trajectory is denoted as  $p(x, t)$ . The 2D-FT of this signal is  $\hat{p}(\phi, \omega)$ . At each position the sound recorded by the microphone is the convolution of the source signal denoted as  $s(t)$  with the corresponding RIR to be found in the RIRs database denoted as  $h(x, t)$ . Therefore,

$$p(x, t) = \int_{-\infty}^{\infty} s(i)h(x, t - i) di. \quad (14)$$

The sound recorded by the moving microphone is denoted as  $r(t)$  with

$$r(t) = p(vt, t). \quad (15)$$

Note that Eq. (15) corresponds to a slice of the function  $p(x, t)$  along the line given by Eq. (13). The spectrum of this recorded sound can be calculated as follows

$$\hat{r}(\gamma) = \int_{-\infty}^{\infty} p(vt, t)e^{-j\gamma t} dt. \quad (16)$$

Also remark that

$$p(vt, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{p}(\phi, \omega) e^{j(\omega t + \phi vt)} d\phi d\omega. \quad (17)$$

Using Eq. (17), Eq. (16) can be rewritten as

$$\begin{aligned} \hat{r}(\gamma) &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{p}(\phi, \omega) e^{-jt(\gamma - v\phi - \omega)} dt d\phi d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{p}(\phi, \omega) \delta(\gamma - v\phi - \omega) d\phi d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{p}(\phi, -v\phi + \gamma) d\phi. \end{aligned} \quad (18)$$

For each frequency  $\gamma$ , the value of the spectrum of the recorded signal is obtained by projection of the 2D spectrum  $\hat{p}(\phi, \omega)$  following the direction  $\mathbf{v}=(1, -v)$  onto the  $\omega$  axis. This result is known as the *projection-slice theorem*: a slice in the time domain corresponds to a projection in the frequency domain.<sup>27</sup> This theorem can be used in this scenario since the moving microphone signal corresponds to a slice of the spatio-temporal sound field as shown in Eq. (15). With an appropriate choice of the excitation signal emitted by the loudspeaker, the projection will be shown to be possibly inverted and the spatio-temporal signal reconstructed. The projection-slice theorem construction is presented in Figs. 6 and 7 where the Doppler effect is also put in evidence. Consider a plane source emitting a plane wave arriving on one microphone line with angle  $\alpha=180$  deg as shown in Fig. 6(a). Consider further a moving microphone along the infinite line. The microphone is moving with a constant speed

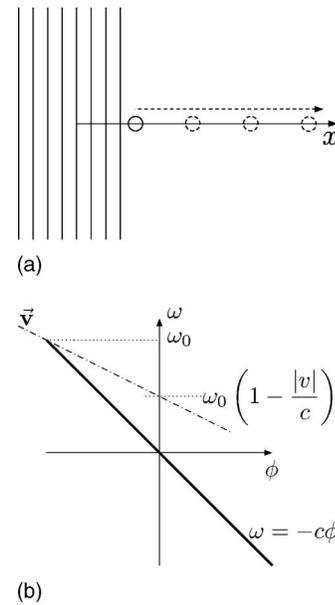


FIG. 6. Doppler effect with a receiver moving away from the source, (a) Schematic view of the situation, (b) Analysis of the situation in the 2D-FT domain.

$|v|$  away from the plane source in the positive  $x$  direction. As seen in Sec. II C, the movement leads to a frequency shift that lowers the perceived frequency at the microphone. The receiver signal can be obtained by projection of  $\hat{p}(\phi, \omega)$  along the direction  $\mathbf{v}$ . As the sound pressure field is a plane wave with an arriving angle of  $\alpha=180$  deg, it has been shown in Sec. II that its 2D spectrum lies on the line of equation

$$\omega = -c\phi. \quad (19)$$

This spectrum is shown in bold in Fig. 6(b). The component of  $\hat{p}(\phi, \omega)$  at frequency  $\omega_0$  is simply the point  $(-\omega_0/c, \omega_0)$  in

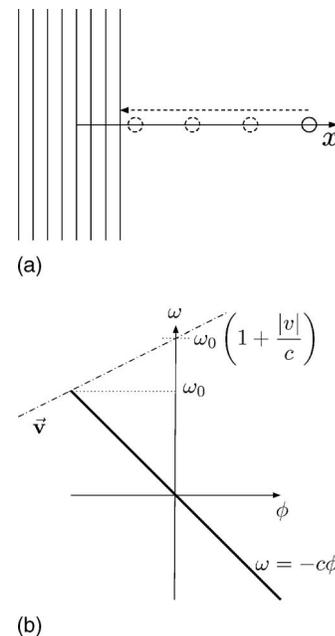


FIG. 7. Doppler effect with a receiver moving towards the source. (a) Schematic view of the situation. (b) Analysis of the situation in the 2D-FT domain.

Fig. 6(b). The projection of this point on the  $\omega$  axis happens at

$$\omega = \omega_0 \left( 1 - \frac{|v|}{c} \right). \quad (20)$$

This is exactly the result obtained when considering the Doppler effect for a receiver moving away from the source as previously described in Eq. (12). The speed of the receiver is  $v_r = -|v|$  since the case of Fig. 5(b) is considered.

Similarly, Fig. 7(a) presents the situation where the receiver is moving towards the source along the negative  $x$  direction. As can be seen in Fig. 7(b), the projection of the point  $(-\omega_0/c, \omega_0)$  on the  $\omega$  axis is now

$$\omega = \omega_0 \left( 1 + \frac{|v|}{c} \right). \quad (21)$$

This corresponds to the Doppler effect for a receiver moving towards the source.

## B. Microphone moving along a circle

In this section, a source emits sound and a receiver is moving along a circle. The pressure measured at the different positions along the circle is given by  $p(\theta, t)$ . The sound recorded by the receiver moving with an angular speed<sup>28</sup> of  $v$  rad/s is  $r(t) = p(vt, t)$ . Similarly to Eq. (18), it can be derived that

$$\hat{r}(\gamma) = \sum_{l_\theta=-\infty}^{\infty} \hat{p}(l_\theta, -vl_\theta + \gamma). \quad (22)$$

For completeness, note that when the source emits a periodic signal  $s(t)$  with period  $2\pi/v$ ,  $r(t)$  becomes also periodic with the same period. In that case,  $r(l_\gamma)$  is the Fourier series of  $r(t)$  with

$$\hat{r}(l_\gamma) = \sum_{l_\theta=-\infty}^{\infty} \hat{p}(l_\theta, -vl_\theta + l_\gamma). \quad (23)$$

The results presented in this section will further be used in Sec. IV.

## IV. SPATIO-TEMPORAL RECONSTRUCTION OF ROOM IMPULSE RESPONSES

This section presents the main result of this work. Considering a moving microphone in the setup, the purpose is to recover the different RIRs at any position along the trajectory from the recording of the microphone. The different aspects of the technique are explored in Sec. IV A for a moving microphone. The relation between the speed of movement and the spacing between the frequency components of the excitation signal is specified. Section IV B discusses different remarks related to the presented techniques and compares both setups. Application of the technique to HRTF measurements is described. Note that in the present paper, the technique is presented for the case of a movement along a circle but can easily be applied for the linear trajectory or other trajectories.

## A. Reconstruction algorithm for a moving microphone

With the knowledge of the emitted sound, the purpose of this technique is to reconstruct the different RIRs at any angle from the recording of the moving microphone. The speed of the moving microphone will be shown to be the key factor in the possible reconstruction of the RIRs at any possible angle. For this purpose, the 2D Fourier representation is used. Figure 8 shows the magnitude of  $\hat{p}(l_\theta, \omega)$  representing the spectrum of the different static signals gathered at the different angular positions along the circle. To apply our algorithm of reconstruction, one needs to impose that not all temporal frequencies are present in the emitted signal. Energy is present only for the temporal frequencies shown as dashed lines in the spectrum  $\hat{p}(l_\theta, \omega)$ . The signal that is recorded by the microphone is given by Eq. (22). As described in Sec. III, the 2D spectrum is projected following the direction  $\mathbf{v}$  on the  $\omega$  axis. To be able to reconstruct the different RIRs, the emitted signal by the source has to be such that all the lines containing energy in the spectrum  $\hat{p}(l_\theta, \omega)$  do not overlap once projected on the  $\omega$  axis. Consider the maximal frequency emitted by the source to be  $\omega_1$ . In the 2D spectrum, this frequency component corresponds to the segment  $[p_1 p_2]$ . When projected on the  $\omega$  axis, new frequency components appear in the range  $\omega \in [\omega_1 - \Delta\omega_p/2, \omega_1 + \Delta\omega_p/2]$  due to the Doppler shift. To avoid any overlapping in the projections, the next frequency component emitted by the source has to be chosen carefully. Denote by  $i$  the intersection point of the projection of the point  $p_1$  with the butterfly spectrum. To obtain the ordinate of  $i$ , denoted as  $\omega_2$ , recall that the minimal slope of the triangular spectrum is given by

$$\omega = \frac{c}{r} l_\theta. \quad (24)$$

The frequency  $\omega_2$  is obtained by solving a system representing the intersection between the segment  $[p_1 i]$  and one side of the butterfly spectrum:

$$\begin{cases} \omega = -vl_\theta + \omega_1 \left( 1 - \frac{rv}{c} \right), \\ \omega = \frac{c}{r} l_\theta. \end{cases} \quad (25)$$

The solution of this system leads to the value  $\omega_2 = \omega_1((c-rv)/(c+rv))$ . Defining  $\Delta\omega$  as the frequency spacing between the two consecutive temporal frequency components  $\omega_1$  and  $\omega_2$  emitted by the source, it can be shown

$$\Delta\omega = \omega_1 \frac{2vr}{c+rv}. \quad (26)$$

As can be seen in Fig. 8, the spacing allowed between successive temporal frequencies is diminishing for smaller temporal frequencies. Therefore, the excitation signal could be very dense at low frequencies. It can be shown that the frequencies emitted by the sources need to satisfy the following relation:

$$\omega_{i+1} = \omega_1 \left( \frac{c-rv}{c+rv} \right)^i. \quad (27)$$

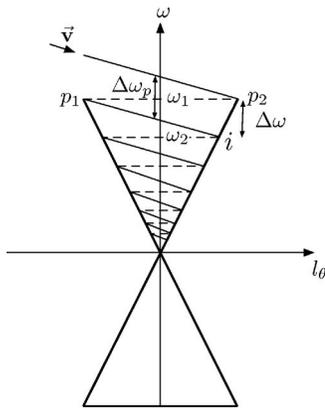


FIG. 8. Projection of the sound pressure field for an emitting sound containing frequencies spaced following Eq. (27).

Nevertheless, to ease the calculation and make the processing simpler, a constant spacing  $\Delta\omega$  between the temporal frequencies of the excitation signal is chosen as in Eq. (26). This new setup is presented in Fig. 9. The excitation signal is therefore a periodic signal with period  $T_S = 2\pi/\Delta\omega$ . Remark that although the excitation signal only contains discrete frequencies, it is possible to determine the RIRs at every frequency thanks to the sampling theorem in the frequency domain.<sup>29</sup> Reconstruction at all frequencies is possible when considering the RIR to be of finite length. The duration of the RIR to be recorded, denoted as  $T$ , has to be smaller than the sampling period  $T_S$  to avoid any temporal aliasing, i.e.,

$$T < \frac{\pi(c + rv)}{rv\omega_1}. \quad (28)$$

To record RIRs of length  $T$ , the maximal speed at which the rotation can be applied is obtained using Eq. (28):

$$v_{\max} = \frac{\pi c}{r(\omega_1 T - \pi)}. \quad (29)$$

## B. Discussion

To summarize, a description of the presented technique is given here. An excitation signal is transmitted to a loudspeaker. The emitted signal cannot contain every frequency but has to be designed carefully as discussed in Sec. IV A. The maximal speed of the moving element follows the rule given in Eq. (29). Note that, in order to keep the periodic character of the excitation signal, one will need to adapt the speed such that the time to make a full rotation corresponds to a multiple of the period of the excitation signal. After recording, the Fourier transform of the microphone signal is taken. In the case of the moving microphone this spectrum corresponds to  $\hat{r}(l_\gamma)$  in Eq. (23). As discussed in Sec. IV A, it corresponds to a projection of  $\hat{p}(l_\theta, \omega)$  on the  $\omega$  axis. This projection needs to be undone to recreate the 2D spectrum. This can be achieved at the condition that no overlapping was present in Fig. 8. This is assured when Eq. (29) is satisfied. To obtain the different RIRs, one simply needs to divide the obtained spectrum by the spectrum of one period of the excitation signal and take the inverse Fourier transform of the dataset to obtain the RIRs at the different spatial

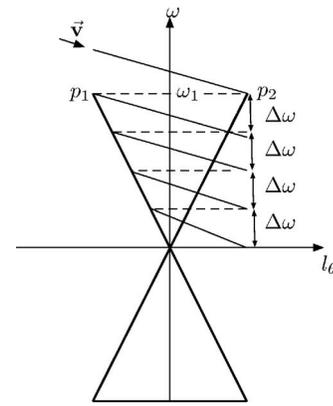


FIG. 9. Projection of the sound pressure field with an emitting signal containing frequencies spaced every  $\Delta\omega$ .

locations. As the obtained 2D spectrum is almost bandlimited along the angular frequency dimension as was described in Sec. II, interpolation of the RIRs at any angle can be achieved.

A numerical example is given.

Using Eq. (29), to record RIRs of 100 ms on a circular array of radius of 1 m up to a frequency of 20 kHz, the rotation needs to be achieved at a maximal speed of  $v = 4.9$  deg/s. The duration of one full rotation is approximately 74 s.

As was shown in Sec. IV A, due to the Doppler effect, the recorded sound can contain higher frequencies than the original emitted sound. Therefore, to avoid aliasing due to the creation of these new frequencies, one has to choose a temporal sampling frequency slightly higher than twice the largest emitted frequency component.

## C. Individualized HRTF measurement

The presented technique is very suitable for the recording of very short impulse responses as is the case for HRTFs. A typical length of interest for HRTF measurement is of the order of 10 ms. Also, as was shown in Ref. 23, the support of the 2D spectrum of HRTFs along the horizontal plane has most of its energy in the region satisfying

$$|l_\theta| \leq |\omega| \frac{d}{2c}, \quad (30)$$

with  $d$  the diameter of the head (typically of the order of 18 cm). With this modification Eq. (29) becomes

$$v_{\max} = \frac{2\pi c}{d(\omega_1 T - \pi)}. \quad (31)$$

The maximal speed to reconstruct all HRTFs up to 20 kHz is therefore of 542 deg/s. In only 0.66 s, the measurement of HRTFs for all angles in the horizontal plane can be achieved. This interesting result finds application in the fast measurement of HRTFs. This technique could be further developed to allow every person to measure his own set of HRTFs and be able to use them when listening during headphone playback. Obviously, the content of the audio playback has to be adapted to support the use of HRTFs. This is precisely the goal of the new MPEG standard<sup>30</sup> that will allow the use of

individualized HRTFs for optimal audio rendering using headphones.

## V. SIGNAL TO NOISE RATIO ANALYSIS

In this section, the performance of the proposed method with respect to a set of static measurements is studied. The goal is to determine, for the two methods, the signal to noise ratio (SNR) as a function of the measurement time and compare the results.

To simplify the analysis, it is assumed that for both the static and the dynamic measurements, a periodic signal of period  $T_S$  and duration  $T_P$  is used: moreover, the measurement noise is considered to be white on the bandwidth of interest and with spectral density  $W$ .

For the static case, a measure has to be performed at each sampling position on the circle. The minimum number of measurements  $N$  is obtained by imposing that there is no superposition on the spectral replicas of the spatio-temporal room impulse response. From Eq. (10), the value

$$N = \frac{2r}{c} \omega_1 \quad (32)$$

is obtained, where  $\omega_1$  corresponds to the maximal frequency present in the excitation signal. Each measurement consists of a repetition of the same period. In the frequency domain, the spectrum consists in a set of lines in the range  $[-\omega_1, \omega_1]$  spaced at angular frequency  $2\pi/T_S$  and approximate width  $2\pi/T_P$ . The measurement noise influences the result only for the contributions corresponding to the lines. As in the considered interval,  $\omega_1 T_S/\pi$  lines are present; the noise variance  $\sigma_s^2$  is given by

$$\sigma_s^2 = \frac{\omega_1 T_S}{\pi} \frac{W}{T_P}. \quad (33)$$

If  $e_S$  is the energy associated to one period of the recorded signal, the SNR for the static measurement, denoted as  $\text{SNR}_s$ , can be defined as the ratio between the power of the recorded signal  $e_S/T_S$  and the noise variance  $\sigma_s^2$ , namely

$$\text{SNR}_s = \frac{e_S}{W} \frac{\pi}{\omega_1 T_S^2} T_P. \quad (34)$$

As one could have expected, the SNR increases linearly with the time taken by each measure. The total measurement time  $T_m$  is given by

$$T_m = N(T_P + T_t), \quad (35)$$

where the additional term  $T_t$  is added to take into account the transient between consecutive measurements. By combining Eqs. (32), (34), and (35), the relation between  $\text{SNR}_s$  and the measurement time is

$$\text{SNR}_s = \frac{e_S}{W} \left( \frac{\pi c}{2r\omega_1^2 T_S^2} T_m - \frac{\pi T_t}{\omega_1 T_S^2} \right). \quad (36)$$

In the case of dynamic measurements, a single measurement is sufficient to recover the information needed to reconstruct any impulse response along the circle. In this case the variable  $T_P$  represents the duration of the full measurement rather than the duration of a single measurement. However,

the Doppler shift will increase the width of the spectral lines of the recorded signal, which will imply a higher noise variance on the reconstructed room impulse responses. The  $i$ th line of the recorded signal will present an angular frequency  $2\pi i/T_S$  and a width of<sup>31</sup>

$$|i| \frac{2\pi 2rv}{T_S c}, \quad (37)$$

where the index  $i$  is in the range  $\pm\omega_1 T_S/2\pi$  (for simplicity, the quantity  $\omega_1 T_S/\pi$  is assumed to be an even integer). To compute the noise variance,  $\sigma_d^2$ , the spectral density has to be integrated on the support, of the lines; this gives approximately

$$\sigma_d^2 = W \frac{r\omega_1^2 v T_S}{2\pi^2 c}. \quad (38)$$

The SNR for the dynamic measurement,  $\text{SNR}_d$ , is obtained by dividing the power of one period of the recorded signal (this is assumed to be approximately constant for the different positions of the microphone) by the noise variance  $\sigma_d^2$ . By using the relation  $v = 2\pi/T_P$ , the resulting expression for  $\text{SNR}_d$  is

$$\text{SNR}_d = \frac{e_S}{W} \frac{\pi c}{r\omega_1^2 T_S^2} T_P. \quad (39)$$

In the case of the dynamic measurement, the measurement time,  $T_m$ , is given by

$$T_m = T_P + T_t, \quad (40)$$

where the additive term  $T_t$  takes into account the time needed to reach the steady state at the beginning of the measurement. Combining Eqs. (39) and (40), the relation between  $\text{SNR}_d$  and measurement time is determined by

$$\text{SNR}_d = \frac{e_S}{W} \frac{\pi c}{r\omega_1^2 T_S^2} (T_m - T_t). \quad (41)$$

By comparing this last relation with Eq. (36), it can be observed that the two methods deliver similar SNR. The main difference is given by the additional time  $T_t$ , which is  $N$  times larger for the static case. This extra time can be used in the dynamic case to obtain a larger SNR. Furthermore, the component of the SNR related to the measurement time  $T_m$  is essentially twice smaller for the static method than for the dynamic one. In the dynamic case, one takes advantage of the butterfly shaped spectrum, which allows an average reduction of the width of the spectral lines by a factor of 2.

The increase of SNR obtained using the dynamic compared to the static measurement setup can be quantified as follows. From Eqs. (36) and (41), it can be obtained that

$$\frac{\text{SNR}_d}{\text{SNR}_s} = \frac{2(T_m - T_t)}{T_m - NT_t}. \quad (42)$$

Therefore, for large RIRs ( $T_m \gg NT_t$ ), the dynamic SNR will be about 3 dB higher than the static SNR. Nevertheless, for very short RIR measurements ( $T_m \approx NT_t$ ), the dynamic SNR increases indefinitely compared to the static one. It can therefore be concluded that the dynamic setup leads to the highest SNR gain when small RIRs need to be measured.

## VI. EXPERIMENTS

Experiments have been carried out in a sound insulated room. A loudspeaker (Genelec 1029A) was used to generate an excitation signal, recorded by a microphone Beyerdynamic MC-740. Rotation of the microphone was performed using a Pan/Tilt Unit PTU-D46-70. This device allows the rotation of the microphone with a precision of 0.03 deg and with a maximal rotation speed of 60 deg/s. The room, where the measurements were performed, was so that the RIRs have very small energy after 250 ms. The excitation signal was made of the sum of sinusoids spaced with 4 Hz. Each sinusoid component was given a random phase. The frequencies covered by the excitation signal ranged from 20 Hz to 22 kHz. The period of the excitation signal was of length 250 ms. To reconstruct the signal up to 22 kHz on a circular array of radius 60 cm, the maximal rotation speed is of 3 deg/s, as follows from Eq. (29). We chose to apply a rotation of 1.8 deg/s. Therefore, the full rotation was achieved in 200 s. Considering the analysis described in Sec. V, the difference between the two SNRs using either the static or dynamic method can be calculated. In the current scenario, the transient time satisfies  $T_t \geq 250$  ms. Then, using Eq. (42), the dynamic SNR is shown to gain 7.8 dB compared to the static SNR. The Fourier transform of the recorded signal (in blue full lines) and of the emitted signal (in black dotted lines) are shown for low frequencies in Fig. 10(a) and for high frequencies in Fig. 10(b). As discussed in Sec. IV A, it can be observed that for the low frequencies, the projection of the triangular spectrum is very narrow while it becomes wider for larger frequencies. From the spectrum of the recorded signal, the original 2D spectrum needs to be reconstructed. This is achieved by undoing the effect of the projection appearing because of the movement of the microphone. The obtained reconstructed 2D spectrum is shown in Fig. 11. This spectrum is then further divided by the spectrum of one period of the excitation signal. Taking the 2D inverse Fourier transform of the result leads to the RIRs at the different angular positions. A typical RIR obtained by the algorithm is shown in Fig. 12. The RIR is not exactly zero at its beginning because of time aliasing. Aliasing happens since the RIR is not exactly zero after 250 ms but still contains some energy.

Estimating the correctness of the reconstructed RIRs is not a straightforward task. This is mainly due to different limitations related to the setup of the experiment. First of all, the hardware used did not allow a perfect synchronization between the moving motor and the emitted sound. Another limitation of the material was the presence of noise emitted by the motor when the Pan/Tilt unit moves. Adaptively removing the noise or putting around the motor some sound absorbing material could be considered to attenuate this noise source. Nevertheless, the reconstructed RIRs were compared with statically measured RIRs, and the mean squared error (MSE) between them was of the order of  $-10$  to  $-15$  dB. Note that the measurement conditions were not exactly similar for the measurement of the two sets of RIRs (one set obtained with the moving microphone and the other set using a static microphone). The temperature variation as

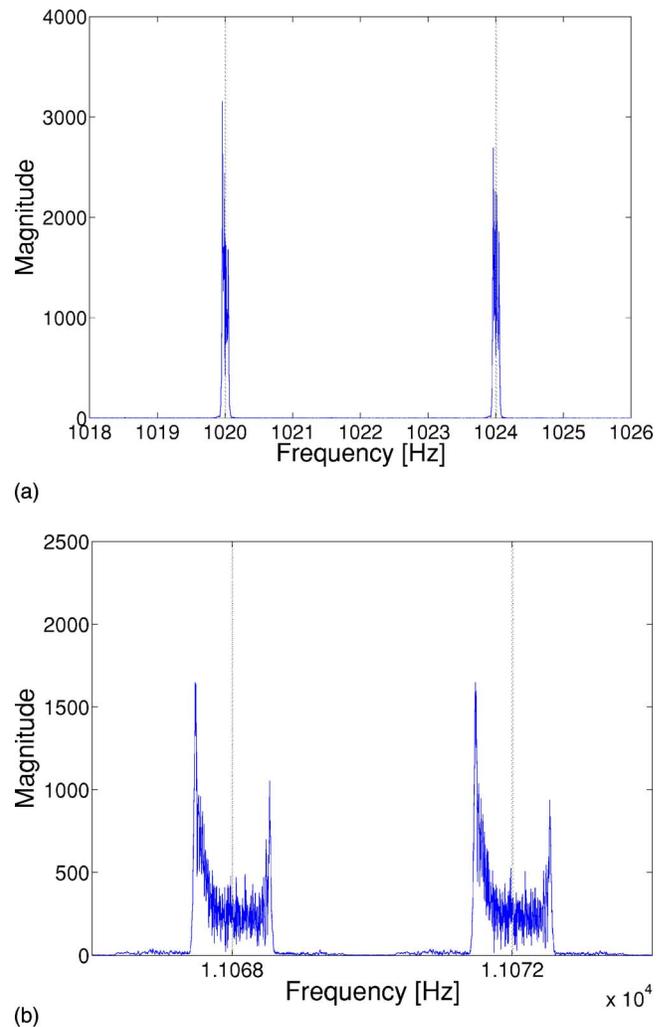


FIG. 10. (Color online) Spectrum of the moving microphone signal (a) at low frequencies (between 1018 and 1026 Hz) and (b) at high frequencies (between 11 066 and 11 074 Hz).

well as the intrusion of a person in the room can lead to variations in the speed of the sound as was shown in Refs. 8 and 20. It is shown in Ref. 8 that a variation of 0.1 deg can create a misalignment between RIRs of more than 25 dB. These different aspects make the comparison in the MSE

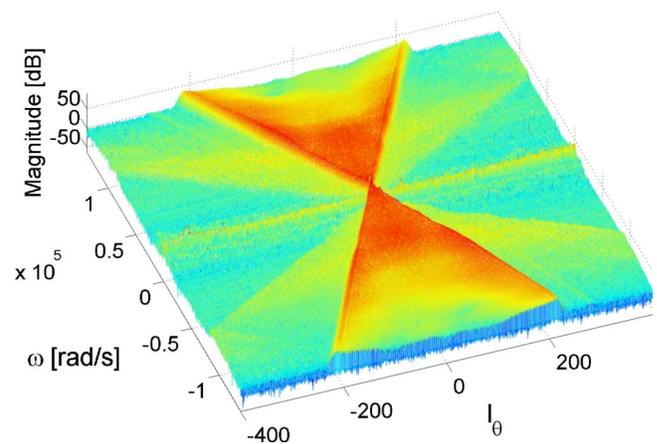


FIG. 11. (Color online) 2D-FT of the recorded data obtained after undoing the effect of the projection due to the recording by the moving microphone.

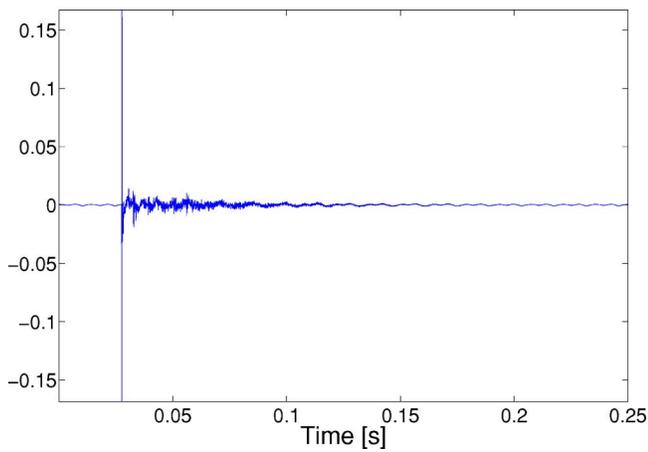


FIG. 12. (Color online) Example of a reconstructed RIR using the presented algorithm.

sense difficult. Nevertheless, the relative time difference and attenuation between the reconstructed RIRs is coherent and similar to the theoretical results obtained in Sec. II or found using a standard static technique.

This section has presented experimental results obtained using hardware showing certain limitations. The purpose of the experiments was to show the possible applicability of the presented algorithm. Further work involving the design of a system using dedicated hardware could be considered. Better attenuation of the motor noise and a better synchronization of the system would lead to better MSE between the static and dynamically reconstructed room impulse responses. The engineering effort to devise a better system is beyond the scope of this paper, since the aim of the experiments was mainly to validate the theory. The obtained MSE is sufficient for this purpose.

## VII. CONCLUSION

In this paper, a technique was presented to record large sets of room impulse responses using a microphone moving along a trajectory. The technique reconstructs the room impulse responses at any position along a linear or a circular array. The speed of movement of the microphone has been shown to be the key factor for the reconstruction. The theoretical aspects of the presented algorithm have been compared with experimental results.

<sup>1</sup>D. Zotkin, R. Duraiswami, and L. Davis, "Rendering localized spatial audio in a virtual auditory space," *IEEE Trans. Multimedia* **6**, 553–564 (2004).

<sup>2</sup>J. Miller, "SLAB: A software-based real-time virtual acoustic environment rendering system," in *International Conference on Auditory Display*, Espoo, Finland (2001).

<sup>3</sup>J. Blauert, *Spatial Hearing* (MIT, Cambridge, MA, 2001).

<sup>4</sup>L. Savioja, "Modeling Techniques of Virtual Acoustics," Doctoral thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, Report TML-A3, 1999.

<sup>5</sup>T. Ajdler, "The plenacoustic function and its applications," Ph.D. thesis, Ecole Polytechnique Fédérale de Lausanne, 2006.

<sup>6</sup>A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.* **93**, 2764–2778 (1993).

<sup>7</sup>D. de Vries, S. Brix, and E. Hulsebos, "Measurements of church impulse responses using a circular microphone array for natural spatial reproduction of a choir concert recording," in *21th Audio Engineering Society Conference* (2002).

<sup>8</sup>G. Elko, E. Diethorn, and T. Gänsler, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *IWAENC*, Kyoto, Japan (2003), pp. 67–70.

<sup>9</sup>Remark that in this work, the turbulence of air due to the moving element is not taken into account in the analysis.

<sup>10</sup>S. Müller and P. Massarani, "Transfer function measurement with sweeps," *J. Audio Eng. Soc.* **49**, 443–471 (2001).

<sup>11</sup>J. Borish and J. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," *J. Audio Eng. Soc.* **31**, 478–488 (1983).

<sup>12</sup>B. Zhou, D. M. Green, and J. C. Middlebrooks, "Characterization of external ear impulse responses using golay codes," *J. Acoust. Soc. Am.* **92**, 1169–1171 (1992).

<sup>13</sup>C. Cheng and G. H. Wakefield, "Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in time, frequency and space," *J. Audio Eng. Soc.*, **49**, 231–249 (2001).

<sup>14</sup>D. Zotkin, R. Duraiswami, E. Grassi, and N. Gumerov, "Fast head-related transfer function measurement via reciprocity," *J. Acoust. Soc. Am.* **120**, 2202–2215 (2006).

<sup>15</sup>E. Hulsebos and D. de Vries, "Parametrization and reproduction of concert hall acoustics measured with a circular microphone array," in *112th Audio Engineering Society Convention* (2002).

<sup>16</sup>E. G. Williams, *Fourier Acoustics* (Academic New York, 1999).

<sup>17</sup>A. Berkhout, *Applied Seismic Wave Theory* (Elsevier Science, Amsterdam, 1987).

<sup>18</sup>Y.-H. Kim, "Can we hear the shape of a noise source?" in *The 18th International Congress on Acoustics* (2004).

<sup>19</sup>H.-S. Kwon and Y.-H. Kim, "Moving frame technique for planar acoustic holography," *J. Acoust. Soc. Am.* **103**, 1734–1741 (1998).

<sup>20</sup>T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its samplings," *IEEE Trans. Signal Process.* **54**, 3790–3804 (2006).

<sup>21</sup>T. Ajdler and M. Vetterli, "The plenacoustic function, sampling and reconstruction," in *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, (2003), Vol. **5**, pp. 616–619.

<sup>22</sup>P. Morse and K. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

<sup>23</sup>T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Head-related transfer functions interpolation considering acoustics," 118th Conv. of the Audio Engineering Society (2005).

<sup>24</sup>T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function on the circle with application to HRTF interpolation," in *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)* (2005), pp. 273–276.

<sup>25</sup><http://lca.vwww.epfl.ch/~thibaut/Measurements/Last> viewed on 23 March 2007.

<sup>26</sup>The formula is valid for the case of a movement along the same direction as the source-receiver axis.

<sup>27</sup>R. N. Bracewell, *The Fourier Transform and its Applications* (McGraw-Hill, New York, 2000).

<sup>28</sup>Note that in the case of a circular trajectory the speed is an angular speed and has units rad/s while along a line the speed "v" has units m/s. We chose to keep the same letter for simplicity of notation.

<sup>29</sup>A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1999).

<sup>30</sup>J. Breebaart, J. Herre, L. Villemoes, C. Jin, K. Kjöring, and J. Plogsties, "Multichannel goes mobile: Mpeg surround binaural rendering," in *29th Audio Engineering Society Conference* (2006).

<sup>31</sup>In Eq. (37), the width of  $2\pi/T_p$  due to the finite duration of the emitted signal is not considered.