Audio Engineering Society
Convention Paper
Presented at the 120th Convention
2006 May 20–23 Paris, France

Room impulse responses measurement using a moving microphone

Thibaut Ajdler[1], Luciano Sbaiz[1], Martin Vetterli[1,2]

[1] *Ecole Polytechnique Fédérale de Lausanne (EPFL), Laboratory of Audiovisual Communications, 1015 Lausanne, Switzerland*

[2] *Department of EECS, University of California at Berkeley, Berkeley CA 94720, USA*

Correspondence should be addressed to Thibaut Ajdler (thibaut.ajdler@epfl.ch)

ABSTRACT

In this paper, we present a technique to record a large set of room impulse responses using a microphone moving along a trajectory. The technique processes the signal recorded by the microphone to reconstruct the signals that would have been recorded at all possible spatial positions along the array. The speed of movement of the microphone is shown to be the key factor for the reconstruction. This fast method of recording spatial impulse responses can also be applied for the recording of head-related transfer functions.

1. INTRODUCTION

For different spatial audio applications (e.g. Wave Field Synthesis[1, 2], beamforming) one would like to measure room impulse responses (RIRs) at a large number of positions in space. In most of the applications, precision and performance of the algorithms increase when using a larger number of RIRs. 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, even the array of microphones would have to be displaced and the measurement could not happen fast. Furthermore, 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. In this paper we introduce a novel technique to easily achieve the recording of a large number of RIRs. We consider a setup with a fixed loudspeaker and a moving microphone. The microphone is moving along a trajectory (e.g. circular trajectory) with constant speed. Also, the acquisition of the data is not done position per position but happens without interruption at any specific spatial position. The movement of the microphone is uniform and does not stop during the acquisition. Thanks to this, one avoid problems linked with abrupt stops of the microphone leading to oscillations of the position and waiting time for the microphone to have obtained its position. From the one dimensional signal gathered by the moving microphone, the two dimensional dataset (spatial and temporal) containing the RIRs at all the different spatial positions along the trajectory is reconstructed. In this paper, an analysis is performed to study the influence of the different parameters to be chosen to achieve the reconstruction of the RIRs (e.g. size 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 paper presents the trade off existing between the speed of movement of the microphone and the spacing between the frequencies contained in the excitation signal. The presented theory is shown together with simulations and real measurements. These real measurements are obtained using a precision moving camera holder that achieves a precision of a few hundredths of a degree when rotating in the median plane. An interesting application of this setup can be found in the measurement of Head-Related Transfer Functions (HRTFs). These measurements are typically done in anechoic chambers to measure the influence of our body on the sound perceived by our brain. HRTFs are filters that model the shape of our head, pinnae and shoulders. In that case, the impulse responses to be measured are typically very short (on the order of a few milliseconds) since no room reflections needs to be captured. This allows to record the whole dataset of azimuthal angles in a very fast manner. The outline of the paper is as follows. In Section 2, the sound field is studied along different geometries (line and circle array). The 2-dimensional spectrum of the sound field is studied for those geometries. In Section 3, we consider that a microphone is moving along the studied trajectories at a certain speed. In this scenario the Doppler effect needs to be considered. This Doppler effect is briefly reviewed in Section 3.1 and it is shown in Section 3.2 that using the 2-dimensional spectrum representation of the sound field along a line, the Doppler effect can be put in evidence. The signal recorded by a moving microphone along a circle is then studied in Section 3.3. Using the knowledge presented in Section 2 and 3, the main novelty of the paper is presented in Section 4. It is shown how from the gathered signal by the moving microphone along a circle, it is possible to reconstruct the sound field or RIRs at all possible angular positions. The technique is presented in Section 4.1. Some further remarks are given in Section 4.2. Finally the theory is compared with real measurements in Section 5.

2. SPATIAL IMPULSE RESPONSES

In this section, we recall results obtained on the study of the 2-dimensional Fourier transform (2D-FT) of the sound field along different geometries. First, the sound field along a line in the room will be presented in Section 2.1 followed by the sound field along a circle in Section 2.2.

2.1. 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 ω_0

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

with $k_0 = \frac{\omega_0}{c}$ and c the speed of sound propagation as shown in Fig. 1. The 2D-FT of (1) is [1]

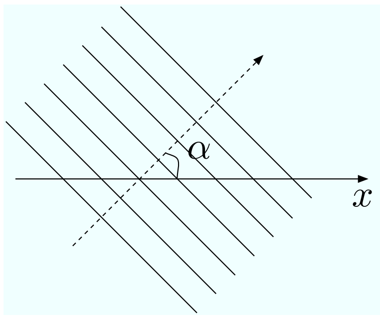


Fig. 1: Projection of the sound field.

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

with ϕ and ω the spatial and temporal frequency respectively.

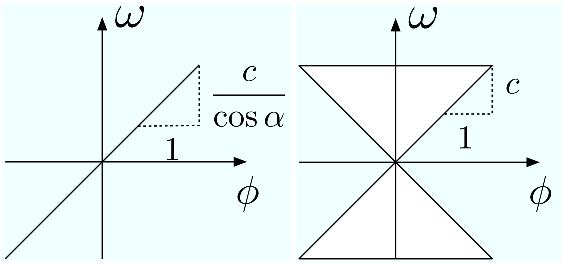
Considering a plane wave emitting all possible frequencies, (1) becomes

$$p(x,t) = \delta\left(t + \frac{x\cos\alpha}{c}\right). \tag{3}$$

The 2D-FT of (3) is then

$$p(\phi,\omega) = 2\pi\delta\left(\phi - \frac{\omega\cos\alpha}{c}\right). \tag{4}$$

This 2-dimensional spectrum is shown in Fig. 2(a). When plane waves arrive from all possible angles, the support of the spectrum is represented in Fig. 2(b).



(a) **(b)**

Fig. 2: 2-dimensional spectrum of the sound field. (a) In case of a single plane wave. (b) In case of plane waves arriving from all possible angles.

Further work 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 [3, 4].

2.2. Sound field along a circle

For the study of the sound field along a circle, consider the scheme presented in Fig. 3(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. We assume the sound source to have coordinates (s_x, s_y, s_z). The time of arrival from the source to the receiver is given by the following expression

$$h(\theta) = \frac{\sqrt{(s_x - r\cos\theta)^2 + (s_y - r\sin\theta)^2 + (s_z - m_z)^2}}{c}. \quad (5)$$

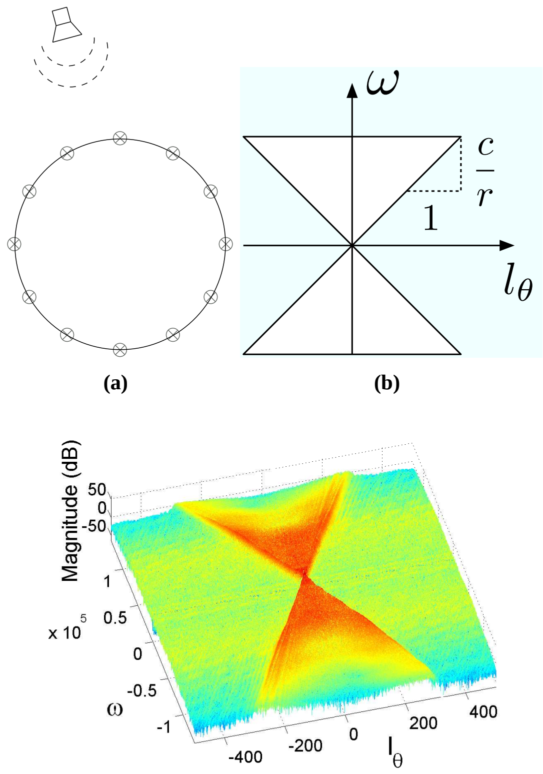
The free field pressure recorded on the circle of microphones due to the emission of a Dirac by the source is given by [5]

$$p(\theta, t) = \frac{\delta(t - h(\theta))}{4\pi c h(\theta)}. \quad (6)$$

Remark that $p(\theta, t)$ is 2π-periodic with respect to θ. The 2D-FT of (6) denoted as $p(l_\theta, \omega)$ has been studied in [6, 7]. Note that l_θ is the index of the Fourier series with respect to the θ-axis. As shown in Fig. 3(b), most of the energy of the spectrum is located into a bow-tie region spectrum satisfying

$$|l_\theta| \leq |\omega| \frac{r}{c}. \quad (7)$$

Fig. 3(c) shows an example of the 2D-FT of RIRs measured on a circle. These measurements were taken every .36 deg along a circle of radius .55 m.



(a) **(b)**

(c)

Fig. 3: Sound field along a circle. (a) Schematic view of the setup. (b) Support of the 2D-FT of the sound field analyzed on a circle. (c) 2D-FT of room impulse responses measured on a circle.

3. MOVING MICROPHONE SIGNAL

In Section 2, the sound field has been studied at different spatial position. Using this knowledge, we would like to study the sound that is recorded by a microphone when it is moving along a trajectory. First, the Doppler effect will be reviewed in Section 3.1. Further, the moving microphone signal will be studied when the microphone is moved along a line in Section 3.2 and along a circle in Section 3.3.

3.1. Doppler effect

.

When considering a moving source or microphone, a frequency shift is observed on the recorded signal. This is known as the Doppler effect and can be expressed by the following formula:

$$\omega = \omega_0 \left(\frac{c \pm v}{c \pm v_s} \right), \qquad (8)$$

with ω the observed frequency, ω_0 the emitted frequency, v_s the speed of movement of the source and v the speed of movement of the receiver. In the sequel, we will only consider the case of microphone movement. The sign of the speed of the receiver is considered as positive when the source and the receiver are coming towards each other. The sign is negative when source and receiver are moving apart.

3.2. Microphone moving along a line

The pressure recorded at all positions along a line trajectory is denoted as $p(x,t)$ as discussed in Section 2.1. The 2D-FT of this signal is $p(\phi,\omega)$. Consider a setup with a fixed source and a moving microphone with constant speed v. The sound recorded by the moving microphone is $r(t) = p(vt,t)$. The spectrum of this recorded sound can be calculated as follows:

$$r(\gamma) \;=\; \int_{-\infty}^{\infty} p(vt,t) e^{-j\gamma t} \mathrm{d}t. \qquad (9)$$

Also remark that

$$p(vt,t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\phi,\omega) e^{j(\omega t + \phi v t)} \mathrm{d}\phi \mathrm{d}\omega. \qquad (10)$$

Using (10), Expression (9) can be rewritten as:

$$\begin{aligned} r(\gamma) &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\phi,\omega) e^{-jt(\gamma - v\phi - \omega)} \mathrm{d}t\mathrm{d}\phi\mathrm{d}\omega \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\phi,\omega)\delta(\gamma - v\phi - \omega)\mathrm{d}\phi\mathrm{d}\omega \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} p(\phi, -v\phi + \gamma)\mathrm{d}\phi. \end{aligned} \qquad (11)$$

For each frequency γ, the value of the spectrum of the recorded signal is obtained by projection of the 2-dimensional spectrum $p(\phi,\omega)$ following the direction $\vec{\mathbf{v}} = \frac{(1,-v)}{\sqrt{v^2+1}}$ on the ω axis. This construction is presented in Fig. 4 and 5 where the Doppler effect is also put in evidence. Consider a plane source emitting a plane wave arriving on one microphone line with angle $\alpha = 0^\circ$ as shown in Fig. 4(a). Consider further a moving microphone along an infinite line. The microphone is moving with a constant speed v away from the plane source in the positive x direction. As seen in Section 3.1, the movement leads to a frequency shift that lowers the perceived frequency at the microphone. The receiver signal can be obtained by projection of $p(\phi,\omega)$ along the direction $\vec{\mathbf{v}}$. The component of $p(\phi,\omega)$ at frequency ω_0 is simply the point $\left(-\frac{\omega_0}{c}, \omega_0\right)$. The projection of this point on the ω axis happens at

$$\omega = \omega_0 \left(1 - \frac{v}{c}\right). \qquad (12)$$

This is exactly the result obtained when considering the Doppler effect (8) for a receiver moving away from the source at speed v.

Similarly to the previous case, Fig. 5(a) presents the situation where the receiver is moving towards the source along the negative x direction. As can be seen in Fig. 5(b), the projection of the point $\left(-\frac{\omega_0}{c}, \omega_0\right)$ on the ω axis is now

$$\omega = \omega_0 \left(1 + \frac{v}{c}\right). \qquad (13)$$

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

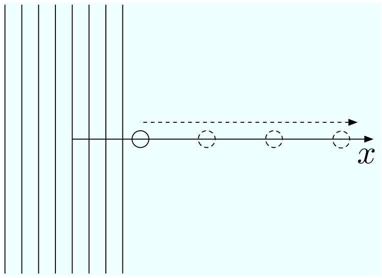
3.3. Microphone moving along a circle

We consider a setup similar to the one presented in Section 2.2. 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 of v rad/s is $r(t) = p(vt,t)$. Similarily to (11), it can be derived that

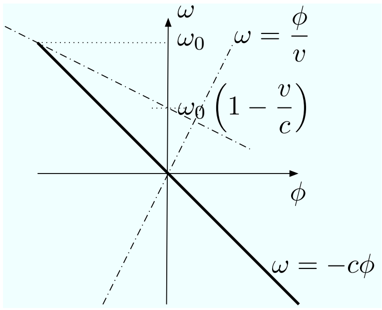
$$r(\gamma) = \sum_{l_\theta=-\infty}^{\infty} p(l_\theta, -vl_\theta + \gamma). \qquad (14)$$

4. DYNAMIC RECONSTRUCTION OF ROOM IMPULSE RESPONSES

This section presents the main result of this paper. From the recording of a moving microphone signal, the purpose is to recover the different RIRs at any position along the microphone trajectory. The different aspects related to the presented technique are presented in Section 4.1. The relation between the speed of movement of the microphone and the spacing between the frequency components of the excitation signal is discussed. Section 4.2

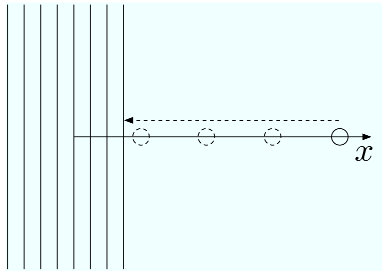


(a)

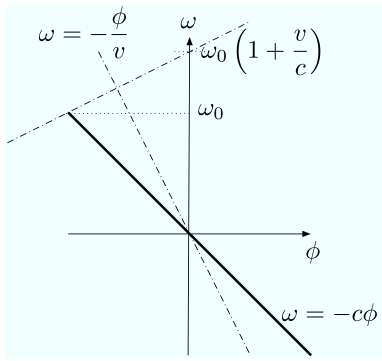


(b)

Fig. 4: 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.



(a)



(b)

Fig. 5: 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.

discusses different remarks related to the presented techniques. Application of the technique to HRTFs measurements is mentioned. Note that the technique is presented for the case of the microphone moving along a circle but can easily be done for the line case.

4.1. Spatio-temporal reconstruction of room impulse responses

With the knowledge of the emitted sound, the purpose of this paper 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 2-dimensional Fourier representation will again be used. Fig. 6 shows the spectrum $p(l_\theta, \omega)$ representing the spectrum of the different signals gathered at the different angular position along the circle. To apply our algorithm of reconstruction, we need 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 $p(l_\theta, \omega)$. The signal that is recorded by the microphone is given by (14). The 2-dimensional spectrum is projected following the direction $\vec{\mathbf{v}}$ on the ω 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 $p(l_\theta, \omega)$ will not overlap once projected on the ω axis. Consider that the maximal frequency emitted by the source is ω_1. In the 2-dimensional spectrum this frequency component corresponds to the line $\omega = \omega_1$ or to the segment $|p_1p_2|$. When projected on the ω axis, new frequency components appear in the range $[\omega_1 - 1/2\Delta\omega_p, \omega_1 + 1/2\Delta\omega_p]$ 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. We have to study where the projection of the point p_1 intersects the bow-tie spectrum. The intersection point is denoted as i. To obtain the ordinate of i, denoted as ω_2, recall that the minimal slope of the triangular spectrum is given by (15):

$$\omega = \frac{c}{r} l_\theta. \tag{15}$$

ω_2 is obtained by solving a system representing the in-

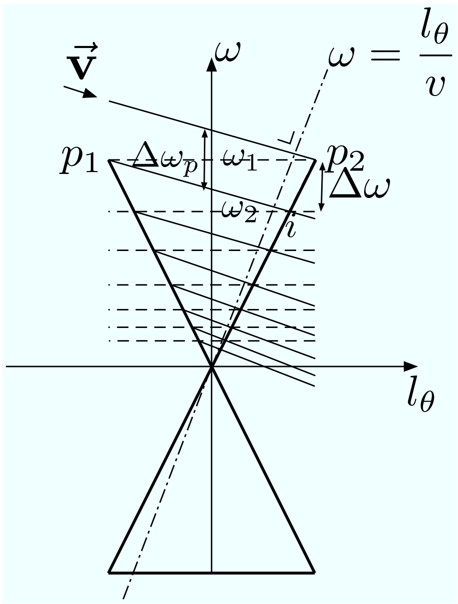


Fig. 6: Projection of the sound field.

tersection between the segment $|p_1 i|$ and one side of the bow-tie:

$$\begin{cases} \omega = -vl_\theta + \omega_1(1 - \frac{rv}{c}) \\ \omega = \frac{c}{r} l_\theta. \end{cases} \tag{16}$$

The solution of this sytem is $\omega_2 = \omega_1 \left(\frac{c-rv}{c+rv}\right)$. Defining $\Delta\omega$ as the frequency spacing between the two consecutive temporal frequency components ω_1 and ω_2 emitted by the source, we find that

$$\Delta\omega = \omega_1 \frac{2vr}{c + rv}. \tag{17}$$

As can be seen in Fig. 6, the spacing allowed between successive temporal frequencies is diminishing for smaller temporal frequencies. Therefore, the excitation signal could be very dense at the low frequencies. It can be shown that the frequencies allowed to be emitted by the sources satisfy

$$\omega_{i+1} = \omega_1 \left(\frac{c - rv}{c + rv}\right)^i. \tag{18}$$

Nevertheless, to ease the calculation and make the processing simpler, we choose to work with a constant spacing $\Delta\omega$ between the temporal frequencies of the excitation signal. It means that the excitation signal is a periodic signal of period $T_S = \frac{2\pi}{\Delta\omega}$. Therefore, to avoid any temporal aliasing, the duration of the RIR to be recorded, denoted as T, has to be smaller than the sampling period T_S, i.e.

$$T < \frac{\pi(c + rv)}{rv\omega_1}. \tag{19}$$

To record RIRs of length T, the maximal speed at which the rotation can be done is obtained using (19):

$$v_{max} = \frac{\pi c}{r(\omega_1 T - \pi)}. \tag{20}$$

4.2. 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 Section 4.1. This signal is recorded using a microphone moving at a certain speed satisfying (20). After recording, the Fourier transform of the microphone signal is taken. It corresponds to $r(\gamma)$ in (14). As discussed in Section 4.1, this signal corresponds to a projection of $p(l_\theta, \omega)$ on the ω axis. This projection needs to be undone to recreate the 2-dimensional spectrum. This can be done at the condition that no overlapping was present in Fig. 6. This is assured when (20) is satisfied. To obtain the different RIRs, one simply needs to divide this 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 locations. By interpolation in time and angular domain [6], the RIRs at any angular position can finally be obtained.

A numerical example is given. Using (20), to record RIRs of 100 ms on a circular array of radius of 1m 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 of approximatively 74 s. 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.

As was shown in Section 4.1, 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.

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 HRTFs measurement is of the order of 10 ms. Furthermore, as was discussed in [6], the support of the 2-dimensional spectrum has most of its energy in the region satisfying

$$|l_\theta| \leq |\omega| \frac{d}{2c}, \tag{21}$$

with d being the diameter of the head (typically of the order of 18 cm). With this modification, (20) becomes

$$v_{max} = \frac{2\pi c}{d(\omega_1 T - \pi)}. \tag{22}$$

The maximal speed to reconstruct the HRTFs up to 20 kHz is therefore of 542 deg/s. In only $.66$ s, the measurement of HRTFs for all angles in the median plane can be achieved.

5. 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 to rotate the microphone with a precision of $.03$ 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 (20). We chose to apply a rotation of 1.8 deg/s. Therefore, the full rotation was achieved in 200 s. The Fourier transform of the recorded signal (in blue full lines) and of the emitted signal (in black dotted lines) are shown for the low frequencies in Fig. 7(a) and for the high frequencies in Fig. 7(b). As discussed in Section 4.1, 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 2-dimensional spectrum needs to be reconstructed. This is achieved by undoing the effect of the projection appearing because of the movement of the microphone.

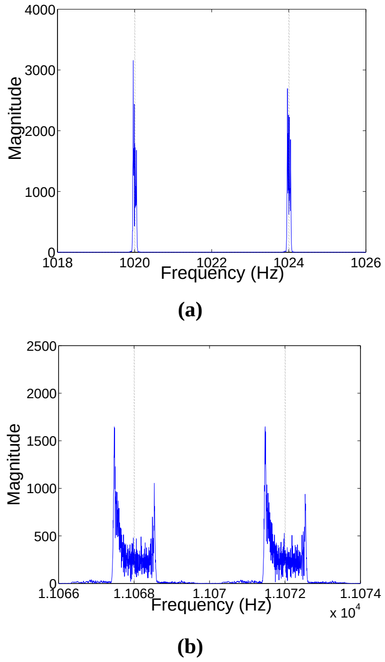


Fig. 7: Spectrum of the moving microphone signal(a) at low frequencies (between 1018 and 1026 Hz).(b) at high frequencies (between 11066 and 11074 Hz).

The obtained reconstructed 2-dimensional spectrum is shown in Fig. 8. This spectrum is then further divided by the spectrum of one period of the excitation signal. Taking the 2-dimensional inverse Fourier transform of the result leads to the RIRs at the different angular positions. A typical room impulse response obtained by the algorithm is shown in Fig. 9. The RIR is not exactly zero at its beginning because of time aliasing due to the fact that the RIR is not exactly zero after 250 ms but still contains little 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 due to the 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

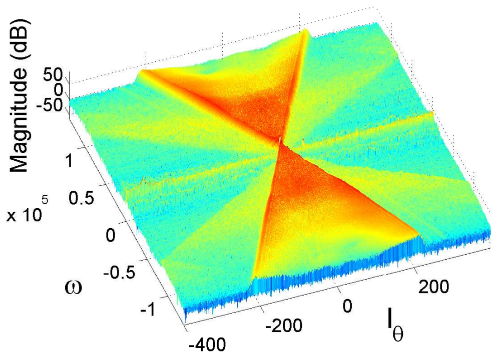


Fig. 8: 2D-FT transform of the recorded data obtained after undoing the effect of the projection due to the recording by the moving microphone.

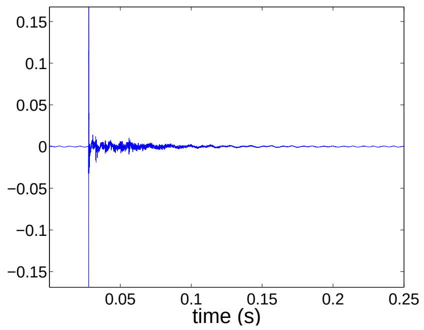


Fig. 9: Reconstructed room impulse response.

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 well as the intrusion of a person in the room leads to variations in the speed of the sound as was shown in [8, 3]. It is shown in [8] that a variation of $.1^\circ$ can create a misalignment between RIRs of more than 25 dB. These different aspects make the comparison in MSE sense difficult. Nevertheless, it can be observed that the relative time difference and attenuation between the reconstructed RIRs is very coherent.

This paper has presented a new technique to record RIRs in a very fast and easy manner and measurements have shown that the goals presented can be achieved. Some further work still needs to be addressed to solve the minor issues discussed above.

6. CONCLUSION

In this paper, a technique was presented to record large sets of room impulse responses using a microphone moving along a trajectory. The paper first studied the 2-dimensional spectra of the recordings obtained by a microphone moving along a line or a circular trajectory. Using these results, the technique to reconstruct the room impulse responses at any angular position along a circular array was presented. 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.

7. REFERENCES

[1] A.J. Berkhout, *Applied Seismic Wave Theory*, Elsevier Science, 1987.

[2] E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved microphone array configurations for auralization of sound fields by wave field synthesis," in *110th Conv. of the AES*, 2001.

[3] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *accepted in IEEE transactions on Signal Processing*, 2005.

[4] T. Ajdler and M. Vetterli, "The plenacoustic function, sampling and reconstruction," in *IEEE ICASSP*, 2003.

[5] P.M. Morse and K.U. Ingard, *Theoretical Acoustics*, McGraw-Hill, 1968.

[6] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Head-related transfer functions interpolation considering acoustics," *118th Conv. of the AES*, 2005.

[7] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function on the circle with application to HRTF interpolation," *In proceedings of IEEE ICASSP*, 2005.

[8] 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, 2003, pp. 67–70.