45º CONGRESO ESPAÑOL DE ACÚSTICA
8º CONGRESO IBÉRICO DE ACÚSTICA
EUROPEAN SYMPOSIUM ON SMART CITIES AND ENVIRONMENTAL ACOUSTICS

MEASUREMENT AND VISUALIZATION OF HUMAN ACTIVE ECHOLOCATION CUES APPROACHING OBSTACLES

PACS: 43.55.MC

David Pelegrín García[1]; Marc Van Baelen[2]; Monika Rychtáriková[1]; Christ Glorieux[1]
[1] Lab. Acoustics and Thermal Physics, KU Leuven
Celestijnenlaan 200D, bus 2416, BE-3001 Heverlee, Belgium
Phone: +32 16 32 71 31
E-mail: david.pelegringarcia@fys.kuleuven.be, monika.rychtarikova@bwk.kuleuven.be, christ.glorieux@fys.kuleuven.be
[2] Dept. Engineering Management, University of Antwerp
Stadskampus, Koningstraat 8, S.Ko08.L15, BE-2000 Antwerp, Belgium
Phone: +32 32 65 46 70
E-mail: marc.vanbaelen@uantwerpen.be

ABSTRACT

Some blind people have the ability to use active echolocation, i.e. they manage to navigate in their environment and detect obstacles distantly by producing oral sounds and listening to their echoes. Using a dummy head with microphones at the ears, a directive loudspeaker near the mouth and a reference microphone, we measured the propagation of sound from the mouth to the ears, contained in the Oral-Binaural Room Impulse Responses (OBRIR), at different distances and orientations in front of obstacles made out of cardboard boxes. We analyse and graphically depict the OBRIRs so as to identify potential auditory cues used in human active echolocation.

RESUMEN

Ciertas personas invidentes utilizan la ecolocación activa: son capaces de desplazarse en entornos y detectar obstáculos a distancia mediante la producción de sonidos orales y la audición e interpretación de los ecos generados. Utilizando una cabeza artificial con micrófonos en los oídos, un altavoz directivo delante de la boca, y un micrófono de referencia, hemos medido la propagación aérea del sonido entre la boca y los oídos, caracterizada por la Respuesta Impulsional de Sala Oral-Binaural (OBRIR en inglés), a diferentes distancias y orientaciones en frente de obstáculos compuestos de cajas de cartón. En esta comunicación, analizamos y mostramos gráficamente las OBRIRs medidas para identificar las informaciones auditivas potencialmente utilizadas en la ecolocación activa.

1. INTRODUCTION

In the absence of sight, hearing plays a predominant role for interacting with the environment. In nature, some animals like bats use echolocation to navigate, avoid obstacles and hunt their prey. Echolocation consists in the emission of a sound signal, which reaches the ears following different paths, namely the direct one, and through reflections at boundaries. By recognizing audible features that result from the presence of both direct sound and echoes reflected by objects in the vicinity, the individual can determine properties about the boundary where the reflection was produced (i.e. size, distance, texture, orientation). Some blind people are also able to apply these principles to navigate in their environment, detect obstacles and landmarks, and effectively enhance their mobility skills [1]. Information about the distance from an object can be unambiguously extracted from the time delay of the reflection, and its attenuation. However, attenuation is also influenced by the size of the object and its texture/density. These properties are determined by the frequency balance of the reflection. Moreover, some authors suggest that the aperture of the reflection, or the solid angle from which the reflected sound returns to the echolocator, is linked to the size of the object [2].

Typical echolocation signals are a palatal click, which is typically used by experts as it is a very sharp sound with short duration, and a hissing sound, the latter used by beginners or during demonstration of this principle, because it produces audible coloration in the presence of nearby objects. Examples of the amplitude spectrum of anechoic recordings of such signals are shown in Figure 1. Whereas the click signal contains energy in a broad frequency band mainly between 1 kHz and 10 kHz, the hissing sound (in the figure, a sustained /s/ sound) has most of its energy between 4 and 11 kHz, with a clearly dominant peak at around 8 kHz.

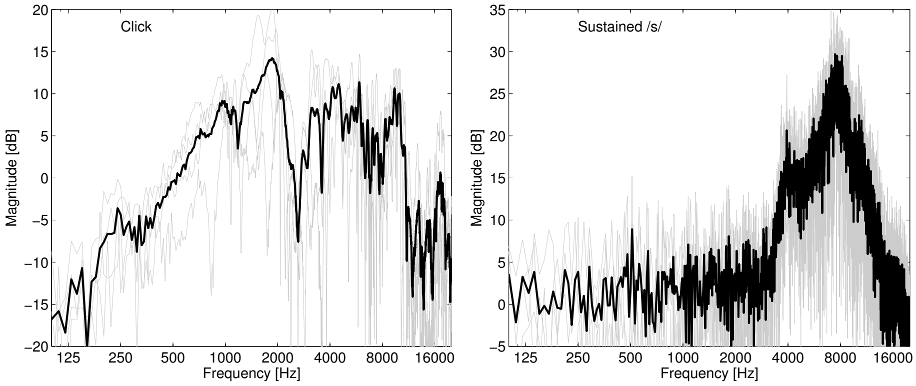

Figure 1. Amplitude spectrum of an oral click signal (left) and of a sustained /s/ sound. Individual repetitions are shown in grey and the average value is shown in black/bold.

Even though the perception of spatial properties of reflections in echoic environments is usually decreased due to the dominance of the direct sound, i.e. the "law of the first wavefront" or precedence effect [3], sudden changes in the echo structure lead to an enhanced perception of echoes (in the literature this has been coined as "breakdown of precedence phenomena" [3]). For this reason, head movements that allow comparisons of echoic patterns in different directions and positions are important for the correct identification of object properties [4].

In the context of human echolocation, a mixed-reality (MR) system for the auralization of one's own voice in virtual environments was recently proposed [5]. The system allows the user to hear himself or herself in an environment different from the actual one where the system is functioning. The system responds to changes in azimuthal orientation of the user. We intend to evaluate the system as a tool to train human echolocation in echolocation-naïve blind users, who would learn to recognize obstacles as they approach them using this system. After reaching a satisfactory performance with the MR system, our users will be evaluated in successfully detecting obstacles in real-life. For this particular experiment, we aim at evoking with the MR system the same auditory effects as those obtained in the actual environment in response to self-generated sounds. In order to obtain this effect, we need to determine the acoustic transfer function (or impulse response) corresponding with the propagation of airborne sound (both direct and reflected) between the mouth and the ears of the same person, at a certain position and orientation of the user. This impulse response is typically referred to as Oral-Binaural Room Impulse Response (OBRIR) [6]. By producing sequential OBRIR scans at different angles and positions, the MR system can react dynamically to changes in users' head orientation and displacements.

Given the number of variables that intervene in the development of the MR system and its use in training, this article has the limited scope of presenting the details of the method used for the acquisition of OBRIR scans. Potential cues in the OBRIRs that can lead to recognition of obstacles through echolocation are visualised and analysed.

2. ACOUSTIC PARAMETERS

The Oral-Binaural Room Impulse Response ($h_{me}(t)$), according to Cabrera et al. [6] can be measured with two microphones placed at the ears of a dummy head, and a microphone at the mouth reference point (MRP), situated approximately 3 cm in front of a loudspeaker at the mouth position. The electrical transfer functions (TFs), which characterize the transformation of the electrical signal sent to the loudspeaker and the electrical signal acquired at the microphones, are $H_{e,L}(f)$ and $H_{e,R}(f)$ at the ears and $H_m(f)$ at the MRP. Once the direct sound part of $H_m(f)$ has been determined by windowing $H_m(f)$ in order to remove room reflections (yielding $H_{m,dir}(f)$), the Oral-Binaural Room Transfer Function (OBRTF) $H_{me}(f)$ can be determined as the ratio between the cross-spectra of the direct sound TF at the MRP and the TF at the ears and the autospectrum of the direct sound at the MRP: The OBRIR can thus be obtained by inverse Fourier transform of the OBRTF. Hence,

$$h_{me,\{L,R\}}(t,\boldsymbol{r},\boldsymbol{\theta}) = \mathrm{F}^{-1}\left\{H_{me,\{L,R\}}\right\} = \mathrm{F}^{-1}\left\{\frac{H^{*}_{m,dir}H_{e,\{L,R\}}}{\left|H_{m,dir}\right|^{2}}\right\}. \quad (1)$$

In the above equation, it is made explicit that the OBRIR is unique for the left or the right ears ($h_{me,L}(t)$ for the left ear, or $h_{me,R}(t)$ for the right ear (the abbreviation $h_{me,\{L,R\}}$ is used, in order to express both ears). It depends on the position $\boldsymbol{r}$ and the orientation $\boldsymbol{\theta}$ of the user.

The signal received at the ears of the listener is the convolution of the echolocation signal generated at the MRP, and the OBRIR. As a consequence, the use of long echolocation signals (e.g. stationary sounds) may blur the temporal resolution of the OBRIR. The choice of echolocation signal, as shown in Figure 1, may emphasize different areas of the spectrum, so that only those range of frequencies contained in the echolocation signals are contained in the reflected signals. In order to avoid this, and to describe the echoes independently of the signal used, in the following, we assume an 'ideal' click, i.e. a Dirac delta signal which has a duration of one sample and equal energy at all frequencies with a 1s-equivalent SPL of 100 dB at the MRP. Note that the duration of real tongue clicks varies between 2 and 10 ms, corresponding with a bandwidth between 800 and 12000 Hz.

Interaural level difference

The spectral difference between the signals arriving at both ears are relevant to locate sounds proceeding from sources located off-axis in the horizontal plane.

$$\mathrm{ILD}(f,\boldsymbol{r},\boldsymbol{\theta}) = 10\log\left(\frac{\left|H_{me,L}(f,\boldsymbol{r},\boldsymbol{\theta})\right|^{2}}{\left|H_{me,R}(f,\boldsymbol{r},\boldsymbol{\theta})\right|^{2}}\right) \quad (2)$$

Average spectral level

For sources or reflections of sound located in the median plane, there are very little interaural differences, but localization in the vertical plane is still possible on a monaural basis due to head filtering. Thus, opposite to the interaural level differences, we consider the "monaural" average spectral level as the average of the signals at the two ears.

$$L_{me}(f,\boldsymbol{r},\boldsymbol{\theta}) = 0.5\times\left(10\log\frac{\left|H_{me,L}(f,\boldsymbol{r},\boldsymbol{\theta})\right|^2}{p_0^2}+10\log\frac{\left|H_{me,R}(f,\boldsymbol{r},\boldsymbol{\theta})\right|^2}{p_0^2}\right) \qquad (3)$$

with p_0=20 µPa rms the reference rms pressure. $H_{me,L}$ and $H_{me,R}$ have been calibrated based on a $H_{m,dir}$ with a 1s-equivalent SPL of 100 dB.

Spectral contrast

Perceiving the presence of an object is easier with comparisons of the OBRIR between one position and the next. For this reason, we define the spectral contrast as

$$\Delta L_{me}(f,\boldsymbol{r}=\boldsymbol{r_M},\boldsymbol{\theta}) = L_{me}(f,\boldsymbol{r}=\boldsymbol{r_M},\boldsymbol{\theta}) - L_{me}(f,\boldsymbol{r}=\boldsymbol{r_{M\text{-}1}},\boldsymbol{\theta}) \qquad (4)$$

Energy level vs time

When a reflection of sound follows the direct sound delayed by more than the “echo threshold”, which is about 2 and 6 ms [3], then the two arrivals are perceived as separate events. In view of this, one of the most obvious indicators of the presence of an echo appears to be the time structure of the OBRIR.

$$L_{me,\{L,R\}}(t,\boldsymbol{r},\boldsymbol{\theta}) = 10\log\frac{\left|h_{me,\{L,R\}}(t,\boldsymbol{r},\boldsymbol{\theta})\right|^2}{p_0^2} \qquad (5)$$

Time contrast

The time contrast is intended to enhance the visibility of changing patterns of reflections as the observer approaches an obstacle.

$$\Delta L_{me,\{L,R\}}(t,\boldsymbol{r}=\boldsymbol{r_M},\boldsymbol{\theta}) = L_{me,\{L,R\}}(t,\boldsymbol{r}=\boldsymbol{r_M},\boldsymbol{\theta}) - L_{me,\{L,R\}}(t,\boldsymbol{r}=\boldsymbol{r_{M\text{-}1}},\boldsymbol{\theta}) \qquad (6)$$

Interaural cross-correlation function

The interaural cross-correlation is a complementary measure of the similarity between the two ears. High values of cross-correlation at non-zero delays may arise from strong reflections from off-axis objects.

$$\rho(\tau,\boldsymbol{r},\boldsymbol{\theta}) = \frac{\int_{-\infty}^{+\infty} h_{me,L}(t,\boldsymbol{r},\boldsymbol{\theta})h_{me,R}(t+\tau,\boldsymbol{r},\boldsymbol{\theta})dt}{\sqrt{\int_{-\infty}^{+\infty} h_{me,L}^2(t,\boldsymbol{r},\boldsymbol{\theta})dt\cdot\int_{-\infty}^{+\infty} h_{me,R}^2(t,\boldsymbol{r},\boldsymbol{\theta})dt}} \quad \text{for -1 ms} \leq \tau \leq \text{1 ms} \qquad (7)$$

3. MATERIALS AND METHOD

The measurements were carried out in a mobility laboratory of dimensions (LxWxH) 16 m x 8.5 m x 3.2 m, located in the basement of Antwerp University Hospital. This is a long room, equipped with optical motion tracking devices, used for analysing the mobility of patients as they walk along it, though this is of no practical relevance in the present article. In our research, we plan to assess the performance of blind people avoiding obstacles by means of active echolocation after training sessions with virtual acoustics using the measured OBRIRs. The background noise level was approximately 45 dB(A). The reverberation time at mid-frequencies (average 500 – 1000 – 2000 Hz) was 0.55 s.

Series of measurements were made at 8 positions along the room (with a spacing of 1 m between them) with and without an obstacle, which was a "soft wall" made out of cardboard boxes covering an area of 1.85 m wide x 1.80 m high. The first position (1) was set at 0.85 m from the obstacle (the distance was measured between the centre of the head and the nearest part of the cardboard boxes), while the following positions were further away from the obstacle, along its normal. Thus, position number 8 was 7.85 m away from the obstacle.

The scheme of the measurement chain is shown in Figure 2a). The transducers were a pair of binaural microphones Sound Professionals mod. SP-TFB-2 mounted on the entrance of the blocked ear canal of a dummy head, and a Fostex 6301B loudspeaker near the position of the mouth. The ears of the dummy head were positioned at a height of 1.6 m. Both the dummy head and the loudspeaker were standing on a rotating table, with the centre of the dummy head aligned to its rotation axis. A PC with Matlab$^{®}$ and the ITA Toolbox$^{®}$ (RWTH, Aachen) was used for the acquisition of the OBRIRs with logarithmic sweep signals at a sampling rate of 44100 Hz. The loudspeaker and the microphones (via a microphone preamplifier) were connected to the PC via an audio interface RME Fireface UCX. The rotating table was controlled from the PC via RS-232 and rotated from -90º (looking left of the obstacle) to +90º (looking right of the obstacle) in steps of 15º, and at 0º the dummy head was looking towards the obstacle. A picture of the room and the measurement equipment at the closest position in front of the obstacle is shown in Figure 2b).

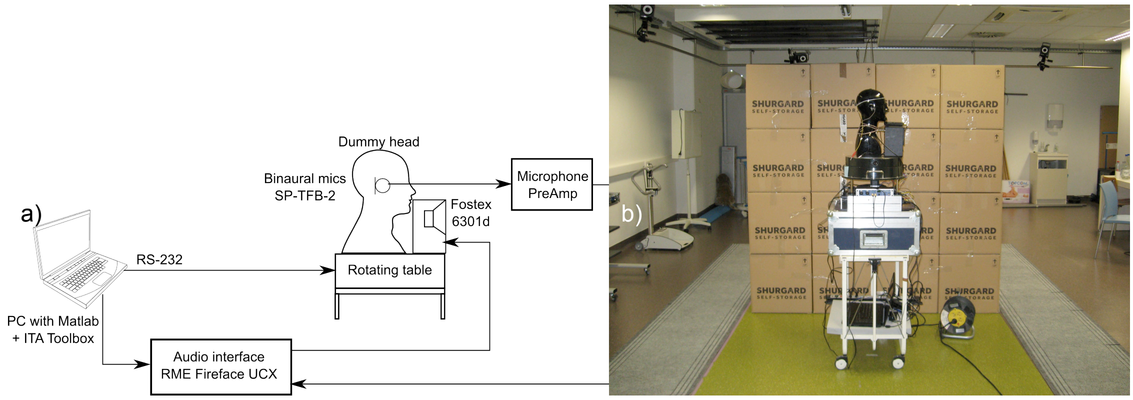

Figure 2. a) Measurement setup scheme. b) Measurement equipment in the first position, closest to the obstacle

With our setup, we measured only $H_{e,L}(f)$ and $H_{e,R}(f)$. The transfer function $H_{m,dir}(f)$ was measured later in an anechoic chamber with one of the binaural microphones positioned 3 cm in front of the loudspeaker. Before the OBRIRs were calculated by applying Eq. (1), all measured IRs were trimmed to the same length (0.5 s), in order to remove the noise floor to a maximum extent. All quantities shown in section 2 were derived, and those shown in the frequency domain are displayed after spectral analysis with a 1/12th octave-band filterbank.

4. RESULTS

The six parameters described in section 2 - namely the Interaural Level Difference (ILD), the Average Spectral Level $L_{me}(f)$, the Spectral contrast $\Delta L_{me}(f)$, the Energy Level vs Time $L_{me}(t)$, the Time contrast $\Delta L_{me}(t)$, and the Interaural Cross-Correlation function IACC or $\rho(\tau)$, have been calculated for angles between -90º and 90º in steps of 15º in the presence and the absence of an obstacle (cardboard wall). The first three parameters, which correspond to parameters in frequency domain, namely ILD(f), $L_{me}(f)$ and $\Delta L_{me}(f)$, are shown in Figure 3.

The ILD shows a particular pattern in the presence of the wall at the closest position (top left plot) in which turning towards left shows mostly a decrease in ILD (in blue) because the reflection at the right ear is stronger, while turning towards the right generates mostly an increase in ILD (in red) because the signal on the right ear becomes stronger. Spectral ripples (and thus audible spectral coloration) appear due to the frequency dependent destructive/constructive interference (alternating at $\Delta f=c/d=400$ Hz, with c the speed of sound and d=0.85m the difference in acoustic path length between the echo and the direct sound) of the reflection with the direct sound. The ripples are different between the left and right ear, as the path lengths of the corresponding reflections differ. The ILD is closer to zero at low frequencies, below 500 Hz, because the wavelength becomes significantly larger than the size of the head. But, as we observe from Figure 1, echolocation signals are marginally important in this frequency range. The patterns at larger distances have very little dependence on the angle and are largely dominated by the ILD of the direct sound (due to asymmetries in the dummy head), which has a marked peak at 7 kHz.

Similar observations can be made in terms of $L_{me}(f)$ (in the central columns in Figure 3). Only the plot with the wall at the first position shows more intense colors, indicating that the total (direct+reflected) level increases. This is better confirmed in the spectral contrast in the two right-most columns, indicating a clear increase when moving from position 2 to position 1 in the presence of the obstacle.

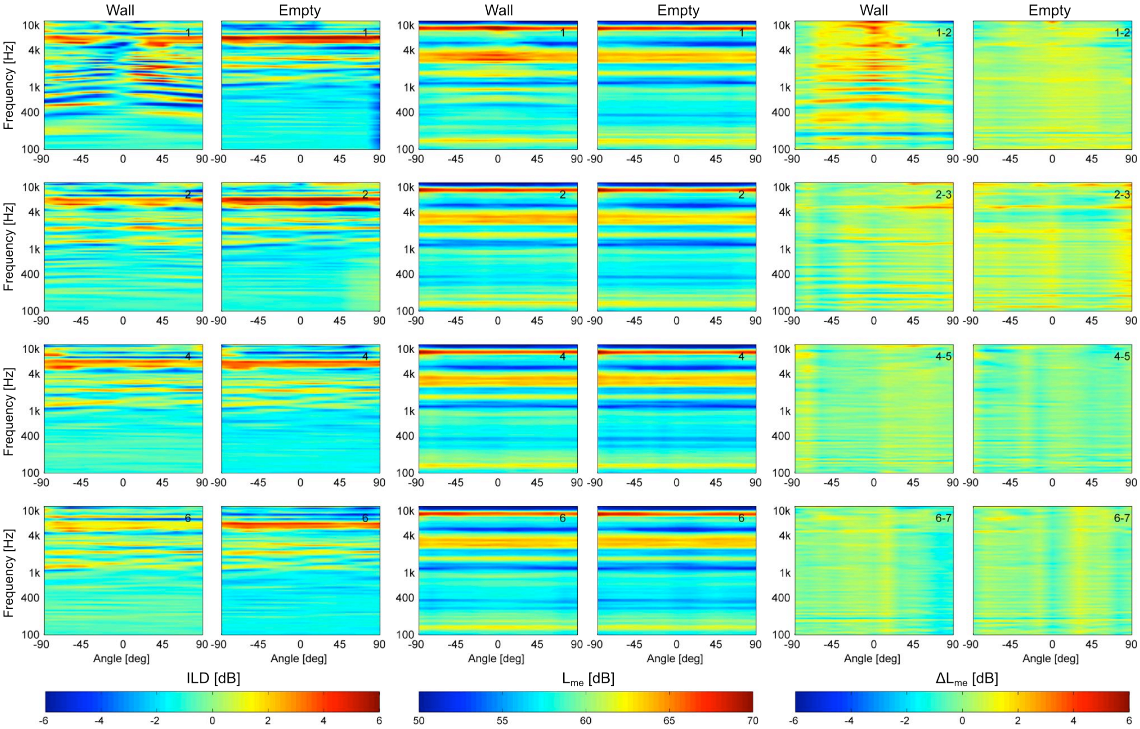

Figure 3. Interaural Level Difference (ILD) (1st and 2nd columns), Average Spectral Level $L_{me}(f)$ (3rd and 4th columns), Spectral contrast $\Delta L_{me}(f)$ (5th and 6th columns). Odd columns show parameters in the presence of the obstacle ("Wall"). Even columns show the parameters in its absence ("Empty"). 1st row: position 1 (0.85 m). 2nd row: position 2 (1.85 m). 3rd row: position 4 (3.85 m). 4th row: position 6 (5.85 m).

Time-related parameters, namely Energy Level vs Time $L_{me}(t)$, Time contrast $\Delta L_{me}(t)$, and Interaural Cross-Correlation function IACC, are shown in Figure 4. The two first parameters are shown only for the first 60 ms, i.e. for reflections coming from as far as 10 m. While the previous figure showed a remarkable dominance of the direct sound on the spectrum, $L_{me}(t)$ and $\Delta L_{me}(t)$ are only affected by the direct sound in the first 5 ms. Thereafter, room reflections from different walls introduce noticeable patterns that may be confounded with our obstacle of interest. Since the envelope of practical tongue clicks lasts for about 6 ms on average, it can be expected that echolocators are not able to exploit the timing information in the first 5ms. The images in the second and fourth column, which correspond to the absence of obstacles, display nevertheless the reflection from the back wall at around 0º by means of dark red clusters of energy, happening at approximately 38 ms (position 1), and happening later with increasing object distance (44 ms at position 2, 56 ms at position 4)[1]. In the presence of the obstacle (first column), it is more complicated to observe any effect in positions 1 or 2, due to the dominance of the direct sound at all angles. Nevertheless, at further distances, the energy clusters around 0º become noticeable (e.g. around 24 ms at position 4 or around 36 ms at position 6). In the first two columns, it is easy to identify the reflections from lateral walls because they are the ones that have maximum energy for angles of + or - 90º. Reflections from the ceiling and the floor have rather short delays. They are independent from the angle and appear within the first 10 ms.

The interaural cross correlation function of the signal (two last columns) as a whole does not show any clear hint on the location of the obstacle. The IACC is dominated by the direct sound - with much more energy than the reflected sound. This infers that one cannot rely on interaural time differences (ITD) for localizing an obstacle using a sustained /s/ sound. Nevertheless, at the closest distance from the obstacle, the IACC has a different pattern, as a result from the increased reflection intensity.

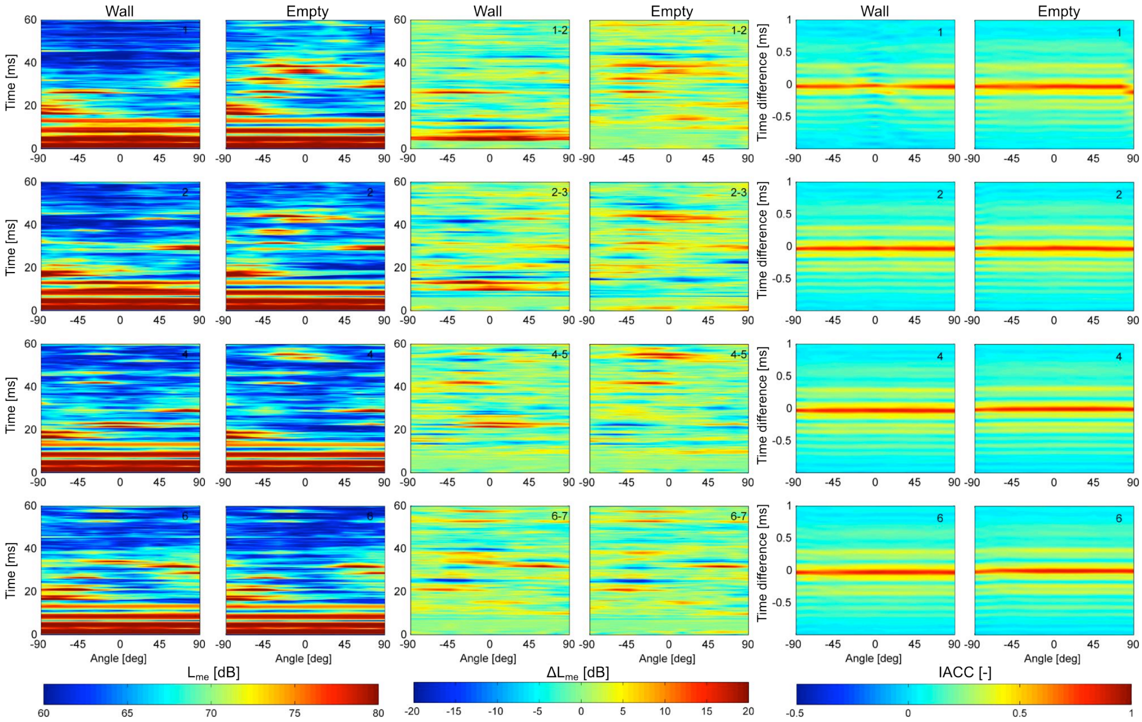

Figure 4. Energy Level vs Time $L_{me}(t)$ (1st and 2nd columns), Time contrast $\Delta L_{me}(t)$ (3rd and 4th columns), Interaural Cross-Correlation function IACC (5th and 6th columns). Odd columns show parameters in the presence of the obstacle. Even columns show the parameters in its absence. 1st row: position 1 (0.85 m). 2nd row: position 2 (1.85 m). 3rd row: position 4 (3.85 m). 4th row: position 6 (5.85 m).

5. DISCUSSION AND CONCLUDING REMARKS

In time domain, there is apparently sufficient information to extract the distance of a distant obstacle (further than 1 or 2 m, depending on the sharpness of the echolocation signal) from the OBRIR. However, it can be expected to be more difficult to retrieve information in time domain at short distances (less than 1 m), due to post-masking as a resulting of a too strong dominance

---

[1] Notice that the round trip for the reflection from an obstacle 1 m further is approximately 6 ms

of the direct sound. However, at these distances, information in frequency domain provides the clearest cues.

In the frequency-domain parameters, the direct sound is usually over-dominating the signal, masking the effect of the reflections, except at the closest position. While this can be a fair indication of auditory effects in the presence of stationary sounds, such as a sustained /s/, for transient sounds it is likely that the auditory salience of the direct sound decreases when the reflected sound arrives, as the events are non-simultaneous. Furthermore, we can hypothesize that performing systematically a repeatable click might lead to an increased suppression of the direct sound at some stage in the auditory pathway.

Spectrotemporal patterns are relevant for perception and a study of these patterns probably allow to elucidate cues that are more informative than purely spectral or purely temporal sound features. However, such analysis is difficult to illustrate in 2D graphs, since there are three independent variables (frequency, time and angle), plus one dependent variable. The same problem applies to a running cross-correlation function, which would reveal more accurate ITD information once the direct sound is left behind.

By pointing out the difficulty of visualizing actual human echolocation cues from actual measurements just by using frequency domain, we can conclude that stationary signals are not as effective as transient signals, which do benefit from time characteristics. At the same time, echolocation requires an intensive and exhaustive training before reaching a functional proficiency level, which would not happen if the information provided by the echoes was easily available. This fact should encourage us to continue the mission to identify the most salient cues (using auditory models with spectro-temporal information), and once identified, to try to explain them, in order to help training echolocators, or use assistive devices to enhance these cues.

ACKNOWLEDGEMENTS

IR measurement and post-processing has been realized by using the ITA-Toolbox for MATLAB developed at the Institute of Technical Acoustics at RWTH Aachen University.

This work is funded by FWO - Scientific Research Foundation in Flanders.

REFERENCES

[1] A. J. Kolarik, S. Cirstea, S. Pardhan, and B. C. J. Moore, “A summary of research investigating echolocation abilities of blind and sighted humans.,” *Hear. Res.*, vol. 310C, pp. 60–68, Apr. 2014.

[2] J. L. Milne, “Seeing with sound: investigating the behavioural applications and neural correlates of human echolocation,” The University of Western Ontario, 2014.

[3] R. Y. Litovsky, H. S. Colburn, W. a Yost, and S. J. Guzman, “The precedence effect,” *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 1633–1654, Oct. 1999.

[4] J. L. Milne, M. a Goodale, and L. Thaler, “The role of head movements in the discrimination of 2-D shape by blind echolocation experts.,” *Atten. Percept. Psychophys.*, May 2014.

[5] D. Pelegrin-Garcia, M. Rychtáriková, C. Glorieux, and B. F. G. Katz, “Interactive auralization of self-generated oral sounds in virtual acoustic environments for research in human echolocation,” in *Proceedings of Forum Acusticum 2014*, 2014, no. c.

[6] D. Cabrera, H. Sato, W. Martens, and D. Lee, “Binaural measurement and simulation of the room acoustical response from a person’s mouth to their ears,” *Acoust. Aust.*, vol. 37, no. 3, pp. 98–103, 2009.