Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: .http://hdl.handle.net/10985/8229

To cite this version :

Marc REBILLAT, Victor BENICHOUX, Romain BRETTE, Makoto OTANI, Renaud KERIVEN - Estimation of the low-frequency components of the head-related transfer functions of animals from photographs - Journal of the Acoustical Society of America - Vol. 135, n°5, p.2534–2544 - 2014

Any correspondence concerning this service should be sent to the repository Administrator : scienceouverte@ensam.eu

Arts et Métiers Institute of Technology

Estimation of the low-frequency components of the head-related transfer functions of animals from photographs

Marc Rébillat[a) b)]
PIMM, Arts et Métiers Paristech, Paris, France.

Victor Benichoux and Romain Brette
Département d'Études Cognitives, École Normale Supérieure, & Laboratoire Psychologie de la Perception, CNRS and Université Paris Descartes, Paris, France.

Makoto Otani
Faculty of Engineering, Shinshu University, Nagano, Japan.

Renaud Keriven
Acute3D, Centre International de Communication Avancée, Sophia-Antipolis, France.

Reliable animal head-related transfer function (HRTF) estimation procedures are needed for several practical applications. For example, to investigate the neuronal mechanisms of sound localization using virtual acoustic spaces, or to have a quantitative description of the different localization cues available to a given animal species. Here two established techniques are combined to estimate an animal's HRTF from photographs by taking into account as much morphological detail as possible. The first step of the method consists in building a 3D-model of the animal from pictures taken with a standard camera. The HRTFs are then estimated by means of a rapid boundary-element-method implementation. This combined method is validated on a taxidermist model of a cat by comparing binaural and monaural localization cues extracted from estimated and measured HRTFs. It is shown that it provides a reliable way to estimate low-frequency HRTF, which are difficult to obtain with standard acoustical measurements procedures because of reflections.
Keywords: Head-Related Transfer Function, Boundary Element Method, Inter-aural Time and Level Differences, Monaural and binaural Localization Cues

PACS numbers: 43.20.Rz, 43.66.Pn

I. INTRODUCTION

Animals can localize a sound source in space by analyzing the sound signals arriving at their two ears. The acoustical transformation occurring between a point source and a receiving ear is usually termed *"Head Related Transfer Function"* (HRTF) (Wightman and Kistler, 2005). HRTFs convey all the acoustical cues available for the animal to localize a sound source in space: *binaural* cues, such as interaural-time differences (ITDs) and interaural level differences (ILDs), and *monaural* cues, such as spectral cues (Blauert, 1996). These acoustical cues however greatly depend on the morphology of the animal under study (Xu and Middlebrooks, 2000; Schnupp *et al.*, 2003; Tollin and Koka, 2009b,a; Jones *et al.*, 2011). Reliably estimating the HRTFs of a given animal is a challenging task that is necessary for various applications.

For example, to investigate the neuronal mechanisms of sound localization with physiological and behavioral experiments, HRTFs can be used to generate controlled binaural

[a)] Marc Rébillat and Victor Benichoux contributed equally to this work.
[b)] Electronic address: marc.rebillat@ensam.eu

stimuli presented through headphones using a virtual auditory environment (Jacobson *et al.*, 2001). Ideally, the use of HRTFs that are not specific to the animal under study should be avoided as it alters the virtual-spatial sensitivities of neurons in the central nervous system (Mrsic-Flogel *et al.*, 2001; Schnupp *et al.*, 2001; Sterbing *et al.*, 2003). Reliable HRTFs are also needed to quantitatively estimate the localization cues available for different animal species. For this purpose, animal HRTFs are commonly measured experimentally by placing miniature microphones at the ear canal entrances and by measuring their responses to controlled sounds played through a movable loudspeaker (Mehrgardt and Mellert, 1977; Roth *et al.*, 1980; Xu and Middlebrooks, 2000; Spezio *et al.*, 2000; Maki and Furukawa, 2005; Koka *et al.*, 2008, 2011; Tollin and Koka, 2009b,a; Kim *et al.*, 2010; Jones *et al.*, 2011). Such procedure nevertheless require dedicated hardware and facilities as well as a substantial amount of time. In particular, sound localization in the horizontal plane relies mainly on low frequency ITDs, but such measurements are often limited in the low frequency range by the quality of the measurement setup (which is mostly degraded by loudspeaker response) and the measurement room (anechoic or not) being used. Alternative HRTFs estimation methods are thus needed to address those issues and to obtain reliable localization cues in the low frequency range.

To avoid HRTF measurements for each animal, HRTFs of a given animal can be estimated using HRTFs measured on another animal of the same species on the basis of some morphological parameters. For cats (Xu and Middlebrooks, 2000), ferrets (Schnupp *et al.*, 2003), and gerbils (Maki and Furukawa, 2005), HRTF amplitudes can be quite accurately predicted by a scaling operation on the log-frequency axis and by a rotation of the HRTFs around the source coordinate sphere. The optimal scaling factor (OSF) and the optimal coordinate rotation (OCR) can furthermore be predicted on the basis of morphological parameters. However, this approach requires that for each species under study one reference measurement and the relations between morphological parameters and the OSF and OCR are available.

Simplified geometrical models such as spherical models (Brown and Duda, 1998; Duda *et al.*, 1999; Algazi *et al.*, 2001b), head and torso models (Algazi *et al.*, 2002; Pernaux, 2003; Zotkin *et al.*, 2003), or dummy head models (Dellepiane *et al.*, 2008) have also been proposed to estimate human listener's HRTFs. The morphological parameters needed to feed the geometrical model can be easily estimated, for example from photographs (Pernaux, 2003; Zotkin *et al.*, 2003; Dellepiane *et al.*, 2008). However, the resulting estimated HRTFs are only approximate and have generally been limited to humans.

Finally, animal HRTFs can be estimated using a full 3D-model of the animal's body obtained for example from a scanning laser (Quaranta, 2003; Muller, 2004; Grace *et al.*, 2008; De Mey *et al.*, 2008). As all the fine morphological details are taken into account in the full 3D-model, HRTFs estimated using this method are valid over a larger frequency range than when estimated using simplified geometrical models. However, this approach relies on expensive equipment that is not always available in practice.

This work thus focuses specifically on the low-frequency components of HRTFs, which convey mainly binaural cues (ITDs and ILDs). As mentioned above these low-frequency components are challenging to obtain with standard acoustical measurements procedures because reflections on walls impact their reliability unless a large anechoic room is used. Although this problem is often addressed in practice by approximating HRTFs as linear-phase filters, it is well known that phase-derived ITDs do in fact substantially depend on frequency (Kuhn, 1977; Roth *et al.*, 1980), and therefore that their accurate estimation require specific measurements procedures in the low frequency range. For example, in the cat Roth *et al.* (1980), Fig. 1.(a), have shown that these ITDs are significantly larger below 1 kHz than above. This general finding is also observed by the authors of the present paper in recordings of ITDs on a taxi-

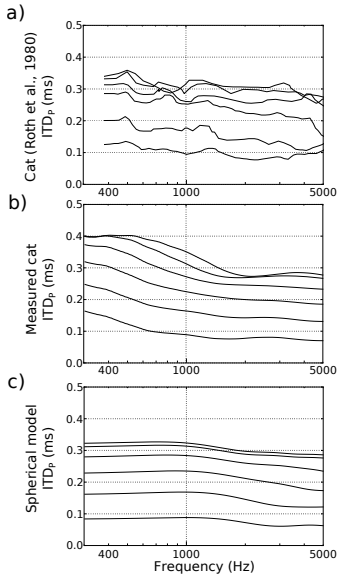

Figure 1. Frequency-dependent ITDs in low-frequency cat HRTFs for several azimuth (15°, 30°, 45°, 60°, 75°, 90°) in the horizontal plane. (a) ITD as reported in Roth *et al.* (1980) for a cat. (b) ITD measured on a taxidermized animal. (c) ITD simulated in a spherical model with the same interaural distance as the cat in (b).

dermized cat, Fig. 1.(b), and on a spherical model Fig. 1.(c). This frequency-dependent behavior of ITDs in the low-frequency range has also been highlighted in human HRTFs (Kuhn, 1977) and for other small mammals (Koka *et al.*, 2011). Moreover, panels (a) and (b) of Fig. 1, suggest that the variation of ITD across frequency exhibits regular patterns that cannot be attributed only to measurement noise. As an example, the bump in ITD around 500 Hz seen in Fig. 1.(a) is present at all positions, and therefore is not merely due to measurement noise. These fine features are furthermore not accounted for by a spherical model of the head as shown by Fig. 1.(c). A natural question to ask is whether these patterns seen in acoustical measurements are effectively due to measurement artifacts, such as reflections, or whether they results from an interaction of the incoming sound field with the animal body, and thus potentially constitute a localization cue. This question can be addressed by computing ITDs from HRTFs obtained from 3D models of the same animals which are by essence free of any measurement noise.

In this paper, we present and validate a general method to estimate HRTFs from raw photographs by taking into account as much morphological detail of the animal as possible. Then, we use this model to measure frequency-dependent ITD cues and assess the nature of the dependence of ITD on frequency for a given position. The first step of the method consists in building a 3D-model of the animal from photographs taken with a standard camera (Lafarge *et al.*, 2010). Based on this 3D-model, the HRTFs are then computed by means of a rapid boundary element method (BEM) implementation (Otani and Ise, 2006). As compared to other methods that use photographs for BEM (Pernaux, 2003; Zotkin *et al.*, 2003; Dellepiane *et al.*, 2008), and that adapt a human head mesh on the basis of morphological parameters estimated from photographs, the

presented method is more general as it builds the whole mesh on the sole basis of these photographs. The main advantages of this method are thus that it is generic, fast, and necessitates no direct acoustical measurements as well as no expensive equipment.

The proposed method is described in Sec. II and used to estimate the HRTFs of a taxidermized cat, which are also experimentally measured and compared to a simple spherical model. The method is then validated in Sec. III by comparing the binaural and monaural localization cues provided by the estimated, spherical, and measured HRTFs. The limitations and new possibilities offered by the method are finally discussed in Sec. IV.

II. METHODS

The HRTFs estimation pipeline presented here is based on two steps and is illustrated on the taxidermized cat (a female *felis domestica*) shown in Fig. 2(a). The HRTFs of the taxidermized cat are also measured on the actual taxidermized cat and compared to a simple spherical head model.

A. HRTF estimation procedure

1. 3D head and body model reconstruction

In order to build a 3D-model of a given object, some input data carrying geometrical information regarding the object to be modeled is needed. A classical way to acquire such data consists in using laser scanning equipments (Bernardini and Rushmeier, 2002). However, such equipment is particularly expensive and hard to use in practice. An alternative method to build 3D-models consists in using a set of raw photographs of the object as input data (Seitz *et al.*, 2006; Strecha *et al.*, 2008). This kind of method requires only a standard camera and is able to provide 3D-models that are qualitatively in good agreement with the ones obtained through laser scanners (Seitz *et al.*, 2006; Strecha *et al.*, 2008).

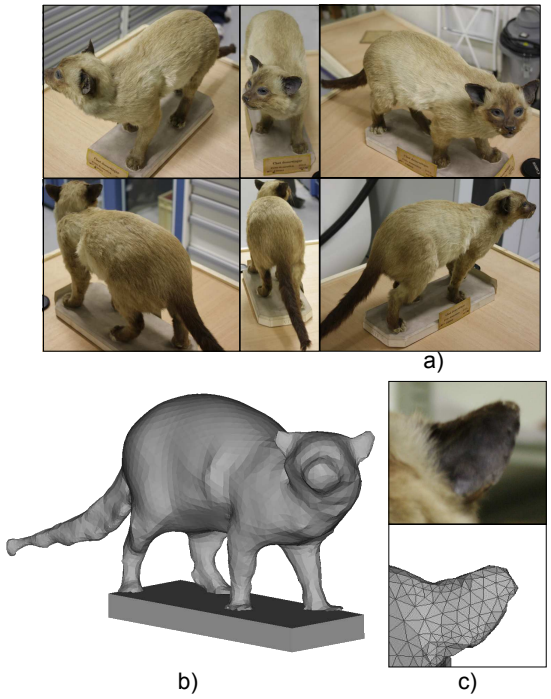

Figure 2. Taxidermized cat used to validate the HRTF computation procedure. (Color online) (a) Examples of photographs used for the 3D-model estimation. (b) Overview of the estimated 3D-model of the cat. (c) Detail of the left ear of the cat: real cat (top) and 3D-model (bottom).

For its practical ease of use, a method based on photogrammetry (Leberl and Thurgood, 2004; Mayer, 2008) and using raw photographs as input data has been retained here (Lafarge *et al.*, 2010). This method models an arbitrary 3D-object as a combination of meshes and of geometrical primitives. On the basis of raw photographs, a Jump-Diffusion process (Grenander and Miller, 1994) is designed to sample these two types of elements simultaneously. The 3D-models reconstructed by this method have been shown to be qualitatively comparable to models acquired through the use of laser scanners (Lafarge *et al.*, 2010). Only the camera focal distance is required for the method to output a scaled 3D-model.

We applied this 3D-model estimation procedure to the cat shown in Fig. 2(a). A 3D-model of the taxidermized animal was estimated based on 68 photographs taken with a standard 10 Mpixels camera under normal lighting conditions, see Fig. 2(b). This model is made up of 11526 triangles with sides of $\simeq 1$ cm, which is larger than the resolution of scanning lasers ($\simeq 1$ mm).

As is shown in Fig. 2(c) the fine geometrical details, such as the exact concave shape of the cat pinnae, are not well captured by this method. It should also be noted that the cat fur has been modeled here as a continuous surface. However, all the major body part shapes, *i.e.* the head, body, legs and tail are accurately modeled. The 3D-model obtained here is thus more precise than a simple geometrical model (spherical model or head and torso model) but not as precise as would be obtained using a precise laser scanner.

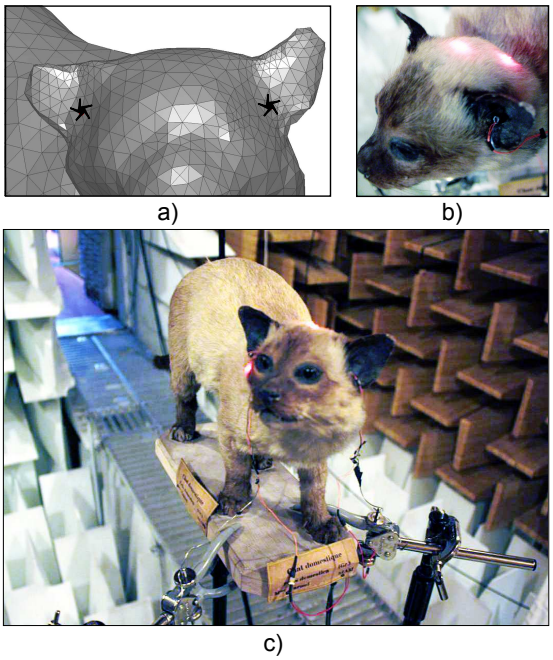

Figure 3. Overview of the measurement and BEM procedures. (Color online) (a) Black stars indicate the position of the acoustical sources in the BEM calculations. HRTF measurement setup for the cat: (b) zoom on the positioning laser and ear canal microphone and (c) overview.

2. Boundary element method formulation

The HRTFs are then computed from this 3D-model using the fast HRTF calculation algorithm proposed by Otani and Ise (2006) based on the boundary element method (BEM). In this BEM formulation, the reciprocity theorem is incorporated into the computational process in order to shorten the processing time. Moreover, all the factors independent of the source position are precomputed in advance. Using this algorithm, the HRTFs for any source position are obtained in a few seconds with a standard PC, once the pre-computation process has been achieved on the estimated 3D-model.

It is generally assumed that BEM performs well up to the frequency for which there are 6 to 10 triangular mesh elements per wavelength (Katz, 1998; Kahana, 2000). Given that the nodes are spaced by $\simeq 1$ cm (see Sec. II.A.1) and according to this criterion, the upper frequency limit below which the HRTFs are accurately computed is expected to be between 3 and 5 kHz.

Regarding boundary conditions, the acoustical properties of the cat fur are not known. It is however known that absorption coefficients are larger for a haired animal than for a hairless one (Ackerman *et al.*, 1957; Katz, 2000). More specifically, some studies focused on the effects of the hair and clothes on human HRTFs: Burkhard and Sachs (1975) found that the ear entrance sound pressure was relatively insensitive to the head impedance. Kuhn (1977) has shown that ITD and ILD differences between a bare and a clothed torso are relatively small below 2 kHz. Katz (2001); Treeby *et al.* (2007a,b) studied the effect of hair on HRTFs and concluded that they do not play an important role below 3 kHz. Katz (2001) furthermore emphasized the fact that modeling hair or fur as a normal reactive impedance is only a rough approximation of the physical reality. Thus, since 1) we

are interested mainly in the frequency range below 3 kHz due to the relatively coarse meshing of our model, 2) no data about the impedance of the cat fur were available, and 3) the BEM implementation of impedance still is not perfect, we chose to apply rigid boundary conditions over the whole cat surface.

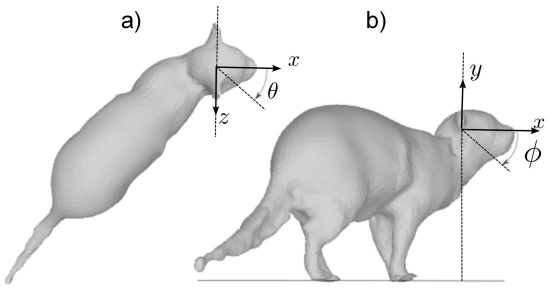

Figure 4. Coordinate system definition: a) top view ($\phi = 0°$) with azimuth θ and b) side view ($\theta = 0°$) with elevation ϕ. See (Warusfel, 2002) for more details.

We computed the HRTFs for the cat from the previously obtained 3D-model, see Fig. 2(b), with rigid boundary conditions. Acoustical sources were placed sequentially at two points situated a few millimeters in front of the 3D-reconstruction of the cat ears, as shown by the red dots in Fig. 3(a). The HRTFs were computed for 651 positions at a distance of 1.95 m, with a frequency resolution of 43 Hz and a sampling frequency of 44.1 kHz. The spatial resolution of the computed HRTFs is 5° in azimuth, from −175° to 180°, and 15° in elevation, from −40° to 90° following the LISTEN coordinates system (Warusfel, 2002) shown in Fig. 4. Precomputation took about 7 hours on a Red Hat Linux, Xeon 3.33 GHz, 8 cores, 48 Gbyte RAM workstation.

B. HRTF measurements for the cat

In order to validate the HRTF estimation pipeline described above, the actual HRTFs of the cat have also been measured experimentally. The HRTFs of the cat shown in Fig. 2(a) were measured by means of the IRCAM HRTF measurement system (Warusfel, 2002). Measurements were done in an anechoic chamber covered with glass wool wedges, see Fig. 3(c). HRTFs were obtained through the blocked ear canal measurement method, which has been shown to be sufficient to acquire directional information in humans (Hammershoi and Moller, 1996; Wightman and Kistler, 2005). A pair of miniature microphones have been placed at the entrance of the occluded ear canal of the taxidermized animal as shown in Fig. 3(b). The cat head was positioned so that the center of the interaural axis was located at the center of the sphere and the horizontal plane was parallel to the support plane. This was accomplished using two low-powered lasers to align the head correctly. The sound source was a speaker placed on a rotating crane. HRTFs were measured at the same 651 positions at which computations were performed, with a frequency resolution of 11.7 Hz and a sampling frequency of 192 kHz.

C. HRTF estimation using a spherical head model

In order to emphasize the benefits offered by the estimation procedure proposed here in comparison with procedures based on simple geometrical models, the HRTFs of the cat have also been estimated using a simple spherical model. The implementation of the spherical model proposed by Duda and Martens (1998) has been used here. The diameter of the sphere modeling the cat's head has been chosen to be equal to the interaural cat distance, *i.e.* 7.5 cm (see Fig. 3). Using this spherical model, the HRTFs were estimated for a distance of 1.95 m at the same 651 positions at which computations and measurements were performed, with a frequency resolution of 10.8 Hz and a sampling frequency of 44.1 kHz.

III. COMPARISON OF LOCALIZATION CUES

In this section the localization cues provided by the different HRTF sets obtained in the

previous section are compared. HRTFs estimated with the procedure proposed in the paper will be referred to as *estimated* HRTFs (see Sec. II.A), experimentally measured HRTFs as *experimental* HRTFs (see Sec. II.B), and HRTFs obtained using the spherical model as *spherical* HRTFs (see Sec. II.C).

A. Definitions

Before comparing the different HRTF sets, the monaural and binaural localization cues that will be used in the following are defined.

The monaural spectral cues for the left and right ears are defined as the left and right directional transfer functions (DTF) as done by Kistler and Wightman (1992). For each considered HRTF set, the mean across the 651 positions of the HRTF log-magnitude is computed. These mean functions include the direction-independent spectral features shared by all the HRTFs of a given set. To remove these features, the appropriate mean function is then subtracted from the log-magnitude of each HRTFs of a given HRTF set. With means removed, the resulting 651 log-magnitude functions, denoted $\mathrm{DTF_l}(f)$ and $\mathrm{DTF_r}(f)$ in the following, represent direction-dependent spectral effects.

Let us now consider a pair of HRTFs, $H_l(f)$ and $H_r(f)$, corresponding to a given source position. The interaural level differences, $\mathrm{ILD}(f)$ (in dB), and interaural phase differences, $\mathrm{IPD}(f)$ (in radians), are defined according to Eqs. (1) and (2), where $\angle(.)$ denotes the *unwrapped-phase* operator, *i.e.* the operator that changes absolute jumps greater than or equal to π to their 2π complement:

$$\mathrm{ILD}(f) = 20\log_{10}\left[\left|\frac{H_l(f)}{H_r(f)}\right|\right] \tag{1}$$

$$\mathrm{IPD}(f) = \angle\left[\frac{H_l(f)}{H_r(f)}\right] \tag{2}$$

Two binaural time localization cues are defined from the interaural phase differences by Eqs. (3) and (4) following Roth *et al.* (1980). $\mathrm{ITD_p}(f)$ is the interaural phase delay and $\mathrm{ITD_g}(f)$ is the interaural group delay. They correspond to the interaural delay of the temporal fine structure and of the envelope, respectively. These two quantities generally differ and depend on frequency because of sound diffraction by the head and body (Kuhn, 1977; Roth *et al.*, 1980).

$$\mathrm{ITD_p}(f) = -\frac{\mathrm{IPD}(f)}{2\pi f} \tag{3}$$

$$\mathrm{ITD_g}(f) = -\frac{1}{2\pi}\frac{d\mathrm{IPD}(f)}{df} \tag{4}$$

To quantify this difference, we additionally define in Eq. (5) a new index, the *Interaural Diffraction Index* $\mathrm{IDI}(f)$ (in radians). This cue denotes the phase-lag induced by diffraction effects between the envelope and the fine structure of the incoming sound. For example, an acoustically transparent head (*i.e.* a case without any diffraction effects), leads to equal and frequency-independent inter-aural group and phase delays, and thus would give a zero IDI at all considered frequencies.

$$\mathrm{IDI}(f) = 2\pi f\left[\mathrm{ITD_p}(f) - \mathrm{ITD_g}(f)\right] \tag{5}$$

Considering all the above-defined cues, there are thus six frequency-dependent localization cues (specifically $\mathrm{ITD_p}(f)$, $\mathrm{ITD_g}(f)$, $\mathrm{IDI}(f)$, $\mathrm{DTF_l}(f)$, $\mathrm{DTF_r}(f)$ and $\mathrm{ILD}(f)$) to be compared for the three HRTF sets obtained in Sec. II. All cues are smoothed in the frequency domain with a third octave sliding window, narrower than the cat auditory filters (Mc Laughlin *et al.*, 2008).

B. Qualitative comparison

In Fig. 5, amplitude-based localization cues $\mathrm{DTF_l}(f)$, $\mathrm{DTF_r}(f)$, and $\mathrm{ILD}(f)$, computed according to the procedure described in Sec. III.A

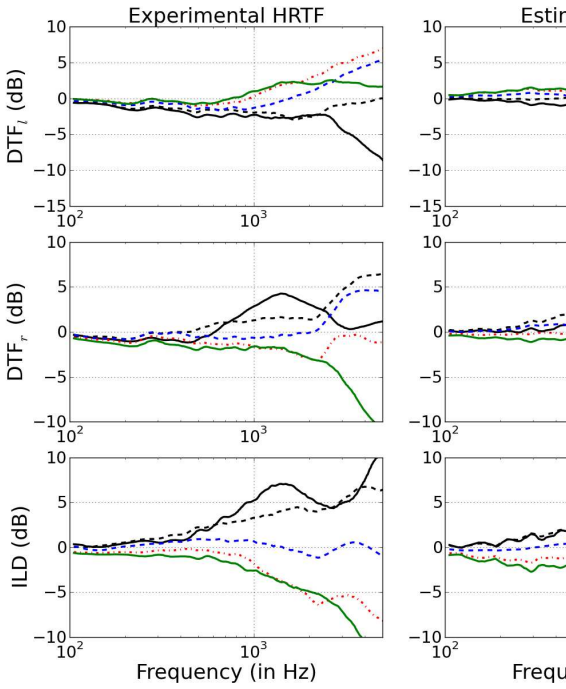

Figure 5. Comparison of amplitude-based localization cues $\mathrm{DTF_l}(f)$, $\mathrm{DTF_r}(f)$, and $\mathrm{ILD}(f)$ for the three HRTF sets for 5 positions in the frontal azimuth plane and between 100 Hz and 5 kHz. (Color online)

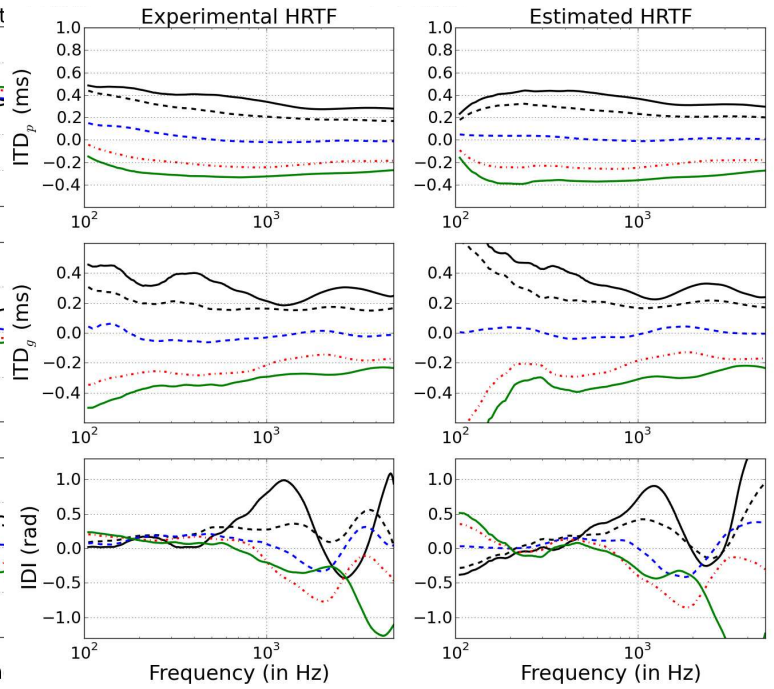

Figure 6. Comparison of time-based localization cues $\mathrm{ITD_p}(f)$, $\mathrm{ITD_g}(f)$ for the three HRTF sets for 5 positions in the frontal azimuth plane and between 100 Hz and 5 kHz. (Color online)

and to Eq. (1) are plotted for the *experimental*, *estimated* and *spherical* HRTF sets for 5 azimuths in the frontal horizontal plane. In Fig. 6, time-based localization cues $\mathrm{ITD_p}(f)$, $\mathrm{ITD_g}(f)$, and $\mathrm{IDI}(f)$, computed according to Eqs. (3), (4) and (5), are plotted for the three HRTF sets and for the same 5 azimuths.

It can be seen in Fig. 5 that there is very good qualitative agreement between the amplitude-based localization cues obtained through the proposed method and those experimentally measured. A much less accurate agreement is observed between those obtained using the simple spherical model and the experimental ones. It is also particularly striking to see that the proposed method is able to render fine changes of the various localization cues with frequency and that the simple spherical model cannot capture such details (see for example the curves for $\theta = -90°$). In Fig. 6, it can furthermore be observed that the same comments generally hold for time-based localization cues above 200 Hz. There is thus globally an excellent qualitative agreement between the *estimated* and *experimental* HRTF sets for the 5 positions tested here.

C. Quantitative comparison

In order to perform a global comparison between the *estimated* (or *spherical*) and the *experimental* HRTF sets over the whole sphere, the differences observed between the monaural and binaural localization cues over the 651 posi-

tions have been computed between 100 Hz and 5 kHz. Fig. 7 shows histograms of the differences between the *estimated* and the *experimental* HRTF sets and Tab. I provides the mean and standard deviation of the difference between the *estimated* (or *spherical*) and the *experimental* HRTF sets. In the following, we denote the differences associated with each cue as ΔDTF_l, ΔDTF_r, ΔILD, ΔITD_p, ΔITD_g and ΔIDI.

	Experimental versus Estimated		Experimental versus Spher	
Cue	Mean	STD	Mean	STD
ΔDTF_l	−0.42 dB	0.95 dB	0.05 dB	2.36 dB
ΔDTF_r	0.42 dB	1.24 dB	−0.05 dB	1.98 dB
ΔILD	0.84 dB	1.21 dB	−0.05 dB	3.05 dB
ΔITD_p	1.2 μs	23 μs	6.8 μs	170 μs
ΔITD_g	10.1 μs	38 μs	11.7 μs	159 μs
ΔIDI	0.07 rad	0.27 rad	0.009 rad	0.47 rad

Table I. Mean and standard deviation (STD) of the differences between the localization cues extracted from the *experimental*, *estimated*, and *spherical* HRTFs sets over the 651 positions between 300 Hz and 3 kHz.

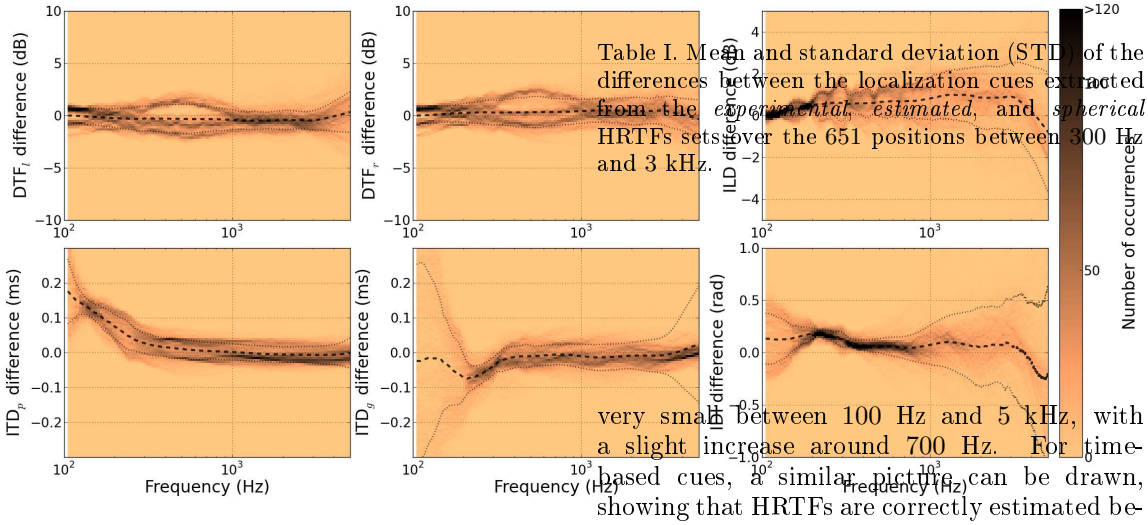

Figure 7. Histograms of the differences in localization cues between the *experimental* and *estimated* HRTF sets over the 651 positions and between 300 Hz and 3 kHz. For each localization cue, the dashed line (- - -) stands for the mean difference and the dotted line (·····) depicts two lines lying one standard deviation apart from the mean difference line. (Color online)

From Fig. 7 it can be observed that for amplitude-based cues, the differences between the *experimental* and *estimated* data sets remains very low between 300 Hz and 3 kHz highlighting the fact that there is quantitative agreement between the amplitude-based cues derived from the two HRTF sets in that frequency range. Regarding ΔILD only, it can be observed that the standard deviation of these differences has a global tendency to increase with the frequency. Focusing on ΔDTF_l and ΔDTF_r, it can be seen that the differences are very small between 100 Hz and 5 kHz, with a slight increase around 700 Hz. For time-based cues, a similar picture can be drawn, showing that HRTFs are correctly estimated between 300 Hz and 3 kHz. ΔITD_p exhibits a positive bias below 300 Hz and remains very low above. Standard deviations associated with ΔITD_g and ΔIDI also increase below 300 Hz and above 3 kHz. From Tab. I, it can also be observed that the differences between the *experimental* and *spherical* data sets have a higher standard deviation, and thus a lower precision, for all the localization cues that are considered here.

In summary, there is a global quantitative agreement between 300 Hz and 3 kHz between the *estimated* and *experimental* HRTF sets for the 651 positions tested around the whole sphere. The agreement between the *estimated* and *experimental* HRTF sets is furthermore better than the agreement obtained using the *spherical* HRTF set in the same frequency range.

IV. DISCUSSION

A. Validity range of the estimation procedure

Results from behavioral experiments in cats indicate that the just-noticeable difference is $\simeq$ 20 μs for $\mathrm{ITD_p}$ and $\simeq$ 1 dB for ILD (Wakeford and Robinson, 1974). This is the same order of magnitude as the differences obtained between *experimental* and *estimated* HRTF sets between 300 Hz and 3 kHz (see Sec. III.C). We can thus conclude that for the presented cat example, the HRTF estimation procedure is accurate up to behavioral precision in that frequency range.

For frequencies higher than 3 kHz, it can be seen in Fig. 7 that the differences between the *estimated* and *experimental* HRTFs suddenly increases for all the localization cues, except $\mathrm{ITD_p}(f)$. As measurements are thought to be reliable in that frequency range, this implies the estimation procedure produces unreliable results here. This upper frequency limit is nevertheless in good agreement with the upper frequency limit related to the BEM procedure (see Sec. II.A.2). We conclude that the 3 kHz upper bound obtained here is a direct consequence of the limited mesh resolution of the 3D-model used here. A solution to go beyond this upper frequency limit could thus be to increase the 3D-model mesh resolution by refining the mesh using smaller triangles or by taking photographs from more viewpoints. Nevertheless, as the influence of fur is supposed to increase with the frequency, the upper frequency limit may not be extended broadly using only mesh refinement.

For frequencies lower than 300 Hz, it can be seen in Fig. 7 that the differences between the *estimated* and *experimental* HRTFs also suddenly increase but only for the time-based localization cues. By looking closely at Fig. 6, it can be seen that there is a systematic increase of $\mathrm{ITD_p}(f)$ as the frequency decreases for the *experimental* HRTFs, which is not observed for the two other HRTF sets. As there is no obvious reason to explain this systematic shift, we postulate that the *experimental* HRTFs are poorly impacted by the measurement setup and room reflections below 300 Hz. By comparing the *estimated* and *spherical model* HRTFs in that frequency range, one can see that there is a good agreement above 200 Hz. Below that frequency, $\mathrm{ITD_p}(f)$ seems to converge to 0 and $\mathrm{ITD_g}(f)$ to diverge for the *estimated* HRTFs and not for the *spherical model*. This may be a side effect of the BEM procedure which produces only one point every 43 Hz in the frequency domain (see Sec. II.A.2) and thus does not provide enough reliable information in the low-frequency range. A solution to go below this lower frequency limit could thus simply be to increase the frequency resolution associated with the BEM procedure.

Measurements of HRTF have been used mainly to study high frequency spectral cues, and therefore a method valid in low frequency may seem to have little practical value. This method is indeed appropriate for the study of temporal cues, rather than spectral cues. Neurophysiologists working on the processing of temporal cues in the auditory brainstem of mammals generally record neurons with characteristic frequency below 3 kHz (Joris and Yin, 2007; Grothe *et al.*, 2010), that is, below the limit of phase locking. Secondly, ITD processing in humans is perceptually dominated by the low frequency range ($<$ 1.5 kHz) (Wightman and Kistler, 1992). Thirdly, it is shown in Figs. 5 and 6 that a simple spherical model cannot account for all the detailed ITD and ILD variations that are observed below 3 kHz. Therefore, accurate HRTF measurements in low frequency are required for this type of study. However, due to the limitations explained above, the present model is unable to render high frequency spectral cues, which are cues to elevation (Algazi *et al.*, 2001a).

Finally, our measurements and estimations were done on a taxidermized animal, which may differ from a live animal by its acoustical impedance. Nevertheless, the acoustical impedance differences between the taxidermized and living animal are expected to be small in comparison with the acoustical impedance differences between the taxidermized animal and a rigid model. As the latter differences are expected to have a very small influence on the binaural cues below 3 kHz (see

Sec. II.A.2), results obtained here for a taxidermized cat can thus confidently be extended to a living cat.

In summary, the estimation procedure proposed here produces HRTFs that can be used for a live animal and that are accurate up to behavioral precision between 300 Hz and 3 kHz. These frequency bounds can furthermore be easily extended as they are directly related to technical implementation details and not to a fundamental limitation of the method itself.

B. Deviations from the spherical head model and the effect of posture

The method proposed here allows estimating the HRTFs of a given animal on the basis of a 3D-model built from photographs. One interesting point regarding this method is that once the 3D-model has been acquired, it can easily be modified or incorporated in complex environments. Indeed, a 3D-model is a computational object that can be modified and integrated with other 3D-models in many ways. This computational object is easier to manipulate than the original physical object (i.e., the taxidermized cat). For example, such manipulations can be useful to assess in a systematic manner the impact of posture or of the environment on localization cues.

To illustrate the potential applications offered by this method, $\mathrm{ITD}_p(f)$ is plotted in Fig. 8 for the cat model with its head turned by 50°, which was the original posture, or modified to be straight. The front horizontal plane corresponds to θ varying between $-90°$ and 90° by steps of 5°. The back horizontal plane corresponds to θ varying between 90° and 180° by steps of 5°. This figure highlights first the fact that $\mathrm{ITD}_p(f)$ varies in a complex, but organized, manner depending both on frequency and head position. This cannot be accounted for by a spherical model. By comparing front and back curves, this figure also emphasizes the fact that the organization of the $\mathrm{ITD}_p(f)$ curves is greatly influenced by the presence of the cat's body. Building on the pioneering work of Algazi *et al.* (2002) in that direction, this method thus potentially allows studying on a systematic manner the effect of body posture on binaural cues.

Acknowledgments

The authors would like to thank the *Museum d'Histoire Naturelle de Paris* for lending the taxidermized cat. The authors would also like to thank IRCAM, and especially O. Warusfel, for the anechoic chamber and the LISTEN HRTF measurement setup. The authors would finally like to thank Brian F.G. Katz and Philip Joris for providing valuable comments regarding former versions of the manuscript. This work was supported by the European Research Council (ERC StG 240132).

Bibliography

Ackerman, E., Adam, A., Berger, R. L., Campanella, A. J., Danner, P. A., Farwell, R. W., Frings, H. W., Oda, F., and Tu, L. (**1957**). "Sound absorption at the surfaces of small laboratory animals. technical report 57-461.", Technical Report, Wright Air Development Center.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001**a). "Elevation localization and head-related transfer function analysis at low frequencies", Journal of the Acoustical Society of America **109**, 1110–1122.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001**b). "Estimation of a spherical-head model from anthropometry", Journal of the Audio Engineering Society **49**, 472–479.

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (**2002**). "Approximating the head-related transfer function using simple geometric models of the head and torso", The Journal of the Acoustical Society of America **112**, 2053–2064.

Bernardini, F. and Rushmeier, H. (**2002**). "The 3D model acquisition pipeline", Computer Graphics Forum **21**, 149–172.

Blauert, J. (**1996**). *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Local-*

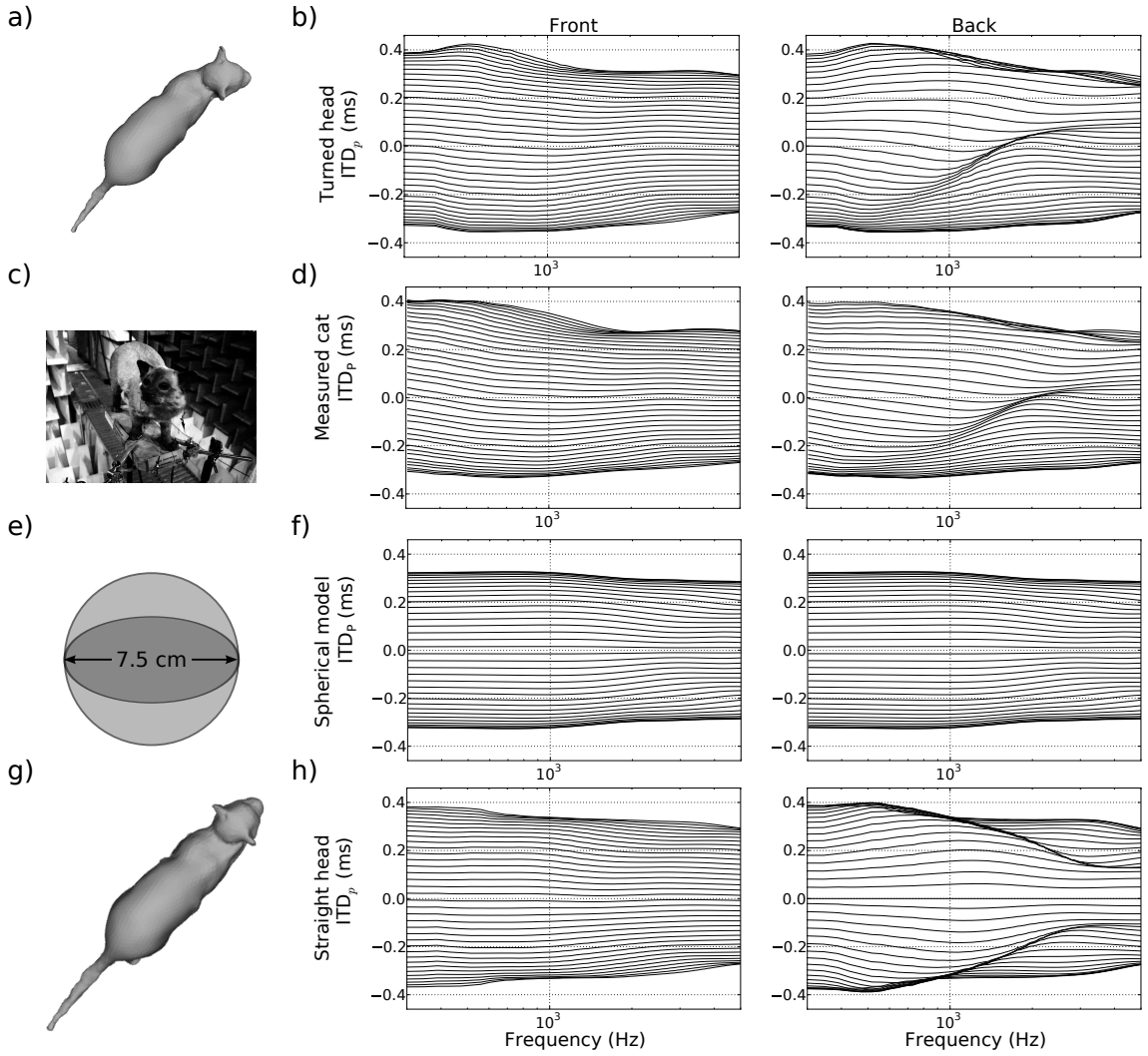

Figure 8. Influence of posture on $\mathrm{ITD}_p(f)$ in the horizontal plane. Each row shows a set of $\mathrm{ITD}_p(f)$ vs. frequency curves for sources with azimuth varying between $-90°$ and $90°$ (front), and between $180°$ by steps of $5°$ (back). (a) 3D model of the cat (head is turned). (b) $\mathrm{ITD}_p(f)$ extracted from computed HRTF of the model. (c) Photo of the taxidermized cat in the anechoic room. (d) $\mathrm{ITD}_p(f)$ measured in the acoustical recordings. (e) Spherical model with the same interaural distance as the cat (7.5 cm). (f) $\mathrm{ITD}_p(f)$ in the spherical model. (g) Modified 3D-model of the cat with the head straight. (h) $\mathrm{ITD}_p(f)$ extracted from computed HRTF of the modified model.

ization (494 pages) (The Massachusett Institute of Technology Press, Cambridge, MA, USA).

Brown, C. and Duda, R. (**1998**). "A structural model for binaural sound synthesis", IEEE Transactions on Speech and Audio Processing **6**, 476–488.

Burkhard, M. D. and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research", Journal of the Acoustical Society of America **58**, 214–222.

De Mey, F., Reijniers, J., Peremans, H., Otani, M., and Firzlaff, U. (**2008**). "Simulated head related transfer function of the phyllostomid bat Phyllostomus discolor", Journal of the Acoustical Society of America **124**, 2123–2132.

Dellepiane, M., Pietroni, N., Tsingos, N., Asselot, M., and Scopigno, R. (**2008**). "Reconstructing head models from photographs for individualized 3D-audio processing", Computer Graphics Forum **27**, 1719–1727.

Duda, R. and Martens, W. (**1998**). "Range dependence of the response of a spherical head model", Journal of the Acoustical Society of America **104**, 3048–3058.

Duda, R. O., Avendano, C., and Algazi, V. R. (**1999**). "An adaptable ellipsoidal head model for the interaural time difference", in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, 965–968.

Grace, S. M., Quaranta, E., Shinn-Cunningham, B. G., and Voigt, H. F. (**2008**). "Simulation of the binaural environmental transfer function for gerbils using a boundary element method", Acta Acustica United With Acustica **94**, 310–320.

Grenander, U. and Miller, M. I. (**1994**). "Representations of knowledge in complex systems", Journal of the Royal Statistical Society. Series B (Methodological) **56**, pp. 549–603.

Grothe, B., Pecka, M., and McAlpine, D. (**2010**). "Mechanisms of sound localization in mammals", Physiological Reviews **90**, 983–1012.

Hammershoi, D. and Moller, H. (**1996**). "Sound transmission to and within the human ear canal", Journal of the Acoustical Society of America **100**, 408–427.

Jacobson, G., Poganiatz, I., and Nelken, I. (**2001**). "Synthesizing spatially complex sound in virtual space: an accurate offline algorithm", Journal of Neuroscience Methods **106**, 29–38.

Jones, H., Koka, K., Thornton, J., and Tollin, D. (**2011**). "Concurrent development of the head and pinnae and the acoustical cues to sound location in a precocious species, the chinchilla (chinchilla lanigera)", Journal of the Association for Research in Otolaryngology **12**, 127–140.

Joris, P. and Yin, T. C. (**2007**). "A matter of time: internal delays in binaural processing", Trends Neuroscience **30**, 70–78.

Kahana, Y. (**2000**). "Numerical modelling of the head-related transfer function", Ph.D. thesis, University of Southampton.

Katz, B. F. G. (**1998**). "Measurement and calculation of individual head-related transfer function using a boundary element model incuding the measurement and effect of skin and hair impedance.", Ph.D. thesis, Pennsylvania State University.

Katz, B. F. G. (**2000**). "Acoustic absorption measurement of human hair and skin within the audible frequency range", Journal of the Acoustical Society of America **108**, 2238–2242.

Katz, B. F. G. (**2001**). "Boundary element method calculation of individual head-related transfer function. ii. impedance effects and comparisons to real measurements", Journal of the Acoustical Society of America **110**, 2449–2455.

Kim, D. O., Bishop, B., and Kuwada, S. (**2010**). "Acoustic Cues for Sound Source Distance and Azimuth in Rabbits, a Racquetball and a Rigid Spherical Model", Journal of the Association for Research in Otolaryngology **11**, 541–557.

Kistler, D. J. and Wightman, F. L. (**1992**). "A model of head-related transfer functions based on principal component analysis and minimum-phase reconstruction", Journal of the Acoustical Society of America **91**, 1637–1647.

Koka, K., Jones, H. G., Thornton, J. L., Lupo, J. E., and Tollin, D. J. (**2011**). "Sound pressure transformations by the head and pinnae of the adult Chinchilla (Chinchilla lanigera)", Hearing Research **272**, 135–147.

Koka, K., Read, H. L., and Tollin, D. J. (**2008**). "The acoustical cues to sound location in the rat: Measurements of directional transfer functions", Journal of the Acoustical Society of America **123**, 4297–4309.

Kuhn, G. (**1977**). "Model for interaural time difference in azimuthal plane", Journal of the Acoustical Society of America **62**, 157–167.

Lafarge, F., Keriven, R., Brédif, M., and Vu, H.-H. (**2010**). "Hybrid multi-view reconstruction by jump-diffusion", in *Conference on Computer Vision and Pattern Recognition (CVPR)* (San Fransisco).

Leberl, F. and Thurgood, J. (**2004**). "The promise of softcopy photogrammetry revisited", International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences **35**, 759–763.

Maki, K. and Furukawa, S. (**2005**). "Acoustical cues for sound localization by the Mongolian gerbil, Meriones unguiculatus", Journal of the Acoustical Society of America **118**, 872–886.

Maki, K. and Furukawa, S. (**2005**). "Reducing individual differences in the external-ear transfer functions of the mongolian gerbil", Journal of the Acoustical Society of America **118**, 2392–2404.

Mayer, H. (**2008**). "Object extraction in photogrammetric computer vision", ISPRS Journal of

Photogrammetry and Remote Sensing **63**, 213–222.
Mc Laughlin, M., Chabwine, J. N., van der Heijden, M., and Joris, P. X. (**2008**). "Comparison of bandwidths in the inferior colliculus and the auditory nerve. ii: Measurement using a temporally manipulated stimulus", Journal of Neurophysiology **100**, 2312–2327.
Mehrgardt, S. and Mellert, V. (**1977**). "Transformation characteristics of external human ear", Journal of the Acoustical Society of America **61**, 1567–1576.
Mrsic-Flogel, T. D., King, A. J., Jenison, R. L., and Schnupp, J. W. H. (**2001**). "Listening through different ears alters spatial response fields in ferret primary auditory cortex", Journal of Neurophysiology **86**, 1043–1046.
Muller, R. (**2004**). "A numerical study of the role of the tragus in the big brown bat", Journal of the Acoustical Society of America **116**, 3701–3712.
Otani, M. and Ise, S. (**2006**). "Fast calculation system specialized for head-related transfer function based on boundary element method", Journal of the Acoustical Society of America **119**, 2589–2598.
Pernaux, J. M. (**2003**). "Spatialisation du son par les techniques binaurales: application aux services de télécommunications. (sound spatialization though binaural techniques: application to telecommunication)", Ph.D. thesis, Institut National Polytechnique de Grenoble.
Quaranta, E. (**2003**). "Application of the boundary element method to computation of the head-related transfer function for gerbils", Master's thesis, Boston University, College of Engineering.
Roth, G., Kochhaar, R., and Hind, J. (**1980**). "Inter-aural time differences - implications regarding the neurophysiology of sound localization", Journal of the Acoustical Society of America **68**, 1643–1651.
Schnupp, J., Mrsic-Flogel, T., and King, A. (**2001**). "Linear processing of spatial cues in primary auditory cortex", Nature **414**, 200–204.
Schnupp, J. W. H., Booth, J., and King, A. J. (**2003**). "Modeling individual differences in ferret external ear transfer functions", Journal of the Acoustical Society of America **113**, 2021–2030.
Seitz, S., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (**2006**). "A comparison and evaluation of multi-view stereo reconstruction algorithms", in *International Conference on Computer Vision and Pattern Recognition*, volume 1, 519 – 528.
Spezio, M., Keller, C., Marrocco, R., and Takahashi, T. (**2000**). "Head-related transfer functions of the Rhesus monkey", Hearing Research **144**, 73–88.
Sterbing, S. J., Hartung, K., and Hoffmann, K.-P. (**2003**). "Spatial tuning to virtual sounds in the inferior colliculus of the guinea pig", Journal of Neurophysiology **90**, 2648–2659.
Strecha, C., von Hansen, W., Van Gool, L., Fua, P., and Thoennessen, U. (**2008**). "On benchmarking camera calibration and multi-view stereo for high resolution imagery", IEEE Conference on Computer Vision and Pattern Recognition, Vols 1-12 2838–2845.
Tollin, D. J. and Koka, K. (**2009**a). "Postnatal development of sound pressure transformations by the head and pinnae of the cat: Binaural characteristics", The Journal of the Acoustical Society of America **126**, 3125–3136.
Tollin, D. J. and Koka, K. (**2009**b). "Postnatal development of sound pressure transformations by the head and pinnae of the cat: Monaural characteristics", The Journal of the Acoustical Society of America **125**, 980–994.
Treeby, B. E., Pan, J., and Paurobally, R. M. (**2007**a). "The effect of hair on auditory localization cues", The Journal of the Acoustical Society of America **122**, 3586–3597.
Treeby, B. E., Paurobally, R. M., and Pan, J. (**2007**b). "The effect of impedance on interaural azimuth cues derived from a spherical head model", Journal of the Acoustical Society of America **121**, 2217–2226.
Wakeford, O. and Robinson, D. (**1974**). "Lateralization of tonal stimuli by cat", Journal of the Acoustical Society of America **55**, 649–652.
Warusfel, O. (**2002**). "The LISTEN database: http://recherche.ircam.fr/equipes/salles/listen/ (date last viewed 07/22/2013)", .
Wightman, F. and Kistler, D. (**2005**). "Measurement and validation of human HRTFs for use in hearing research", Acta Acustica united with Acustica **91**, 429–439.
Wightman, F. and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization", Journal of the Acoustical Society of America **91**, 1648–1661.
Xu, L. and Middlebrooks, J. C. (**2000**). "Individual differences in external-ear transfer functions of cats", The Journal of the Acoustical Society of America **107**, 1451–1459.
Zotkin, D. N., Hwang, J., Duraiswami, R., and Davis, L. S. (**2003**). "HRTF personalization using anthropometric measurements", in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 157–160.