3D sparse imaging in biosonar scene analysis

Bertrand Fontaine, Herbert Peremans, Jan Steckel

3D sparse imaging in biosonar scene analysis

Bertrand Fontaine, Herbert Peremans and Jan Steckel
Active Perception Lab, Universiteit Antwerpen
13, Prinsstraat, 2000 Antwerpen, Belgium
Contact email: bertrand.fontaine@ua.ac.be

***Abstract*—Bats can navigate through foliage thanks to their biosonar. To build an image of their environment they emit ultrasonic pulses and analyse the corresponding echo signals which consist of a complex superposition of a set of highly overlapping reflected pulses coming from the targets at their different positions. We propose to use a sparse representation of echoes in overcomplete dictionaries to perform 3D imaging using biosonar signals. We show that even if the echosignal is a superposition of highly overlapping echoes the algorithm can efficiently reconstruct the image only from a single measurement. Those results can be useful from a robot sensor perspective where the robot has to find a free path. It can also help to understand better how bats can efficiently navigate through dense foliage.**

I. Introduction

Navigation through a realistic outdoor environment, i.e one containing trees, bushes and other natural reflectors, while solely relying on sonar would seem to be an impossible task if it were not routinely solved by bats. A bat flying in a natural environment, e.g. a forest, is confronted with large numbers of (randomly) distributed scatterers of various sizes. Yet, it manages to extract the spatial information it needs in order to manoeuvre with great skill through its environment. To do so, the so-called FM-bat emits short Frequency Modulated (FM) calls (e.g. in Fig.1(a)) and analyse the corresponding echoes (e.g. in Fig.1(b)).

If we assume that the environment reflects towards the bat a number of glints (strong echoes) much smaller than the number of samples needed to represent the received sound, this 3D image reconstruction task can be formulated as a sparse problem. To solve it, a redundant dictionary is built using shifted version of a simple reflector echo signal as atoms. The shift in time, which corresponds to a certain distance traveled by the soundwaves, allows a reconstruction along the depth axis. It is similar to what is done for seismic signals analysis [1].

The atoms of the dictionary are further filtered as they would be by the bat's Head-Related Transfer Function (HRTF) providing spatial cues [2], [3] about the reflector's position. This spatial dependent filtering provides information about the other two degrees of freedom: azimuth and elevation. This approach is similar to the one proposed in [4] for separating sounds from different sources. The HRTF processing is modelled using a phase array which is optimised with respect to a sparse recovery criterion [5].

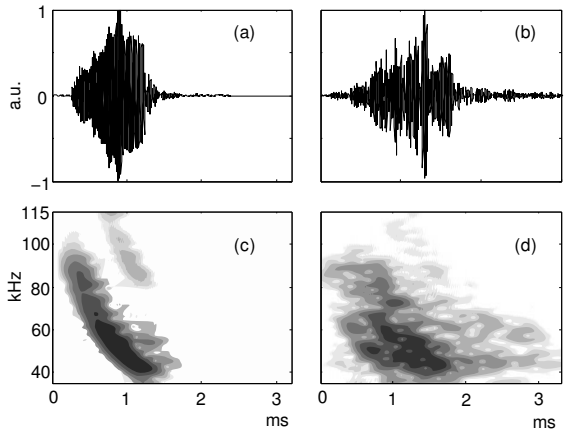

Fig. 1. Emitted call in its time-pressure representation (a) and spectrogram (c). Echo from a complex object in its time-pressure representation (b) and spectrogram (d)

Using a robotic setup, we show how this 3D imaging technique can localise multiple simple objects very efficiently based on a single measurement. We also apply the technique on more complex reflecting objects and show that, even if the results are still good, other problems arise that are not dependent on the algorithm self but e.g. the choice of the bases or the difference in energy between echoes coming from different directions.

II. Measurement setup

A polaroid emitter [6] is mounted on a pan/tilt platform. The transmit signal sent to the emitter is a downward sweep from 100 khz down to 40kHz with a duration of 1ms at a sampling frequency of 500kSamples/sec. Due to the non-linearities of the polaroid transducer, the first harmonic is also present in the emitted acoustic signal and the duration is a bit longer. The sent signal, which resembles a typical FM-bat call [7], can be seen in Fig.1.

The emitter, which has a piston-like directionality, concentrates the emitted energy mainly in the first lobe pointing forward. Therefore, the region in space where signals of sufficient energy can be reflected from is restricted to a truncated cone as shown in Fig.2. The maximum aperture considered is 25° in azimuth and elevation.

The receiver consists of a 16 channel (4x4), 2D array (s. Fig. 3) built using Knowles FG23329-PO7 microphones and

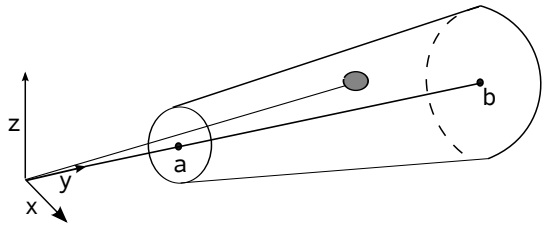

Fig. 2. Due to the directivity of the emitter the covered space is a truncated cone.

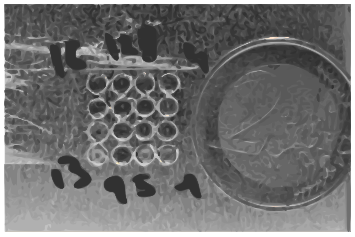

Fig. 3. The 2D array consisting of 16 microphones.

connected to a custom made 16 channel A/D system sampling the 16 microphone signals simultaneously at 500kSamples/sec. As explained further, the 16 received signals from the array are combined and this combined signal is bandpass filtered between 115kHz and 35kHz, and downsampled to 250kSamples/sec.

III. Depth axis

As a first approximation, the echo signal received by a bat from a small object can be considered as a linear superposition of echoes, each one a scaled and delayed version of the emitted call. Hence, if it is assumed that the phase of the emitted pulse is not changed upon reflection, a target image R can be entirely characterised by a sum of scaled and time-shifted Dirac impulses

$$R(t) = \sum A_i \delta(t - t_i) \tag{1}$$

where A_i denotes the amplitude of the echo reflected by reflector i and $t_i = 2 * r_i / v_s$ with r_i denoting the range along the depth axis of reflector i and v_s the speed of sound. This can be extended to the case where the depth axis is longer, i.e. when the glints might come from different objects.

The measured echo $S(t)$ can be written as the convolution of the target impulse response and the emitted call $C(t)$ which in our case is a chirp from 100kHz down to 40kHz. Hence, introducing sampled versions of the signals of interest i.e., the column vector $S = [S(0.T_s), S(1.T_s), \cdots, S(n.T_s)]^T$ denoting the received signal with n the total number of samples and T_s the sample period, and analogously for the vectors C and R denoting the emitted call and the target image respectively, yields

$$S = \mathbf{D}R \tag{2}$$

where $\mathbf{D}$ is a $n \times n$ matrix whose k-th column contains the sampled version of the emitted call C delayed by k samples.

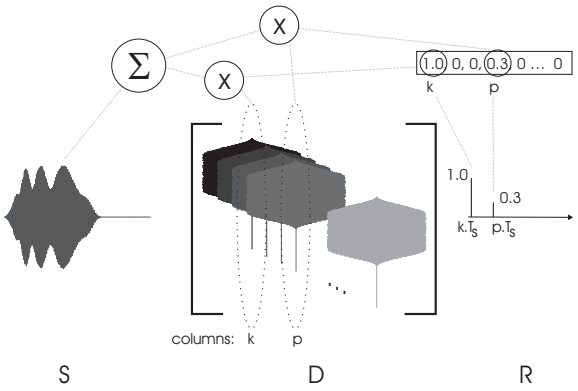

Fig. 4. The sparse representation R together with the dictionary $\mathbf{D}$ allows to reconstruct the received signal S.

$$\mathbf{D} = \begin{bmatrix} C(0.T_s) & 0 & \cdots & 0 \\ C(1.T_s) & C(0.T_s) & \cdots & 0 \\ C(2.T_s) & C(1.T_s) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ C(m.T_s) & C((m-1).T_s) & \cdots & 0 \\ 0 & C(m.T_s) & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & C(0.T_s) \end{bmatrix} \tag{3}$$

For example (see Fig. 4), if the received signal contains two echoes with amplitudes 1 and 0.3 respectively arriving at times $k \cdot T_s$ and $p \cdot T_s$, the 'simplest' way to build S from R using $\mathbf{D}$ is to set the k-th and the p-th samples of R equal to 1.0 and 0.3 and zero all the other ones.

The matrix $\mathbf{D}$ is the dictionary to build S from R. In the ideal case, the shifted signal in the base is a perfect copy of the emitted call. In reality, even a finite sphere filters the reflected echo. The question then arises as what should be the signal taken as a basis. Two choices are tried; the emitted call and an echo coming from a wooden sphere of diameter 6cm. In any case, an object has a finite number of facets reflecting towards the direction of emission. Therefore, the number of echoes in S and thus the number of non-zero entries in R will be small with respect to the number of samples n, R can be considered as sparse.

The problem of finding the sparsest solution R of Eq.2 is relaxed, a non negativity constraint is added and noise is taken into account. The final formulation is:

$$min_R \left[\frac{1}{2} \|S - \mathbf{D}R\|_2^2 + \gamma \|R\|_1 \right], \; R \geq 0 \tag{4}$$

with $\gamma > 0$. To solve this convex quadratic optimisation problem, an interior-point method is used [8]. The algorithm, whose code can be found at http://www.stanford.edu/~boyd/l1_ls/, is fed with D and its transpose as operators. The computational burden can be alleviated significantly by exploiting the fact that the columns of D are shifted versions of each other

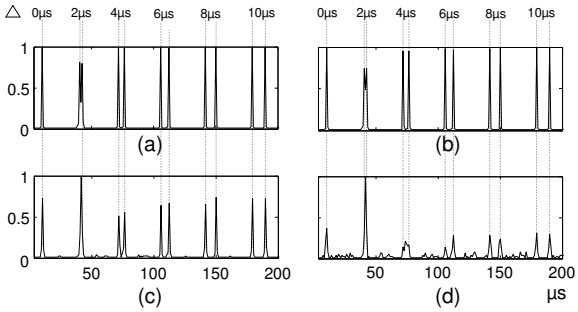

Fig. 5. Mean reconstructed target images (50 trials) at 4 SNR conditions: (a):SNR=45dB, (b): SNR=35dB, (c): SNR=25dB and (d): SNR=15dB. The true target image consists of 11 echoes (vertical lines) all having the same amplitude.

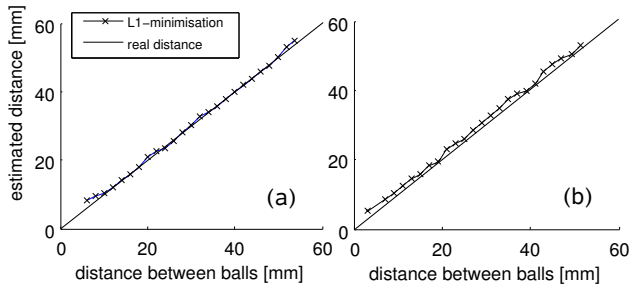

Fig. 6. Results for spheres of diameter 12mm (a) and 6mm (b) with the sparse representation (cross) compared with the real distance (solid line)

allowing to use cross correlations (in the Fourier domain) to compute the inner products. Indeed, every point of the cross correlation between two signals is the inner product of one signal with a time-reversed shifted version of the other signal.

To illustrate the approach we create a simulated echo composed of 11 equal amplitude echoes each delayed by respectively 20,80,84,144,152,212,242,284,300,360 and 380 μs. The simulated object is thus 6.1cm long and contains point-reflectors spaced by minimum 0.6mm. We can see in Fig.5 that for high Signal to Noise Ratio (SNR as defined in [9]) the reconstruction is perfect in both amplitude of the glints and their spacing. Even for lower SNR's the spacing estimation remains very good.

To validate the simulations, small wooden spheres with diameter of 12mm and 6mm were attached in pairs on wires mounted one after the other. They were moved apart along the z-axis so that the closest sphere would not shadow the other. The different bases tried were the reflections of single spheres at the position x=0. The results can be seen in Fig.6.

IV. Reflector localisation

From spatial hearing theory it is known that upon arrival at the listener's head the received acoustic signals are filtered by the Head Related Transfer Function (HRTF) [10]. In particular for bats this HRTF induced filtering is considered to provide important spatial cues that help the bat localise its target [2], [3]. We use the filtering induced by the HRTF to estimate the position of the reflector.

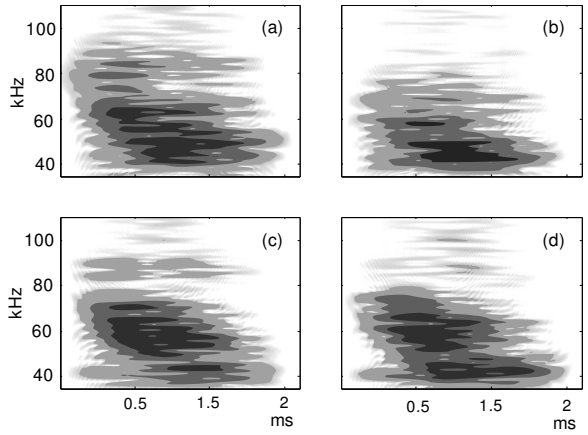

Fig. 7. Echoes from a sphere positioned at different angles. The phased array and the associated delay and sum beamformer filters the echoes in a spatially dependent way.

Assuming that the HRTF filtering is mostly due to the fact that, before reaching the tympanic membrane, sounds bounce against different parts of the outer ear we model this by delaying slightly the individual channels of the microphone array introduced in Section II and scaling each of them differently. Those 32 parameters (16 delays values and 16 scaling factors) are optimised using a genetic algorithm. Although the model could be optimised to match a real bat HRTF, we choose to use a criterion more specific to sparse recovery, i.e. the Fuchs criterion [5].

The spectrograms of four filtered echoes are shown in Fig.7. Due to the different delays and gains applied to the 16 microphone signals by the delay and sum beamformer, additional spectral cues i.e., notches and peaks, are introduced which depend on the spatial position of the reflector. Note that this approach has similarity with the one proposed in [4]. However, in [4], HRTF filtering is used to separate sound sources and not to localise them.

V. 3D imaging

The total dictionary $\mathbf{D}_t$ is built up by taking the union of the dictionaries $\mathbf{D}$ each containing shifted atoms filtered by an HRTF corresponding to position i, one dictionary for each possible target position.

$$\mathbf{D}_t = |\mathbf{D}_1|...|\mathbf{D}_i|\mathbf{D}_k| \tag{5}$$

To indicate how such an approach would indeed allow to simultaneously determine the target's biosonar image as well as the target's position, four sets of spheres are arranged as shown in Fig.8(b). The two first sets are horizontal in the frontal plane and the two other ones are vertical in the back plane. The two planes are separated by 20cm. The space between [-20°, 20°] degrees azimuth and elevation is divided in steps of 2.5°. The depth axis is divided in 200. The energy in the resulting image from each dictionary is taken as the probability to have a glint at this position. The 3D reconstruction is shown in Fig.8(b) in spherical coordinates

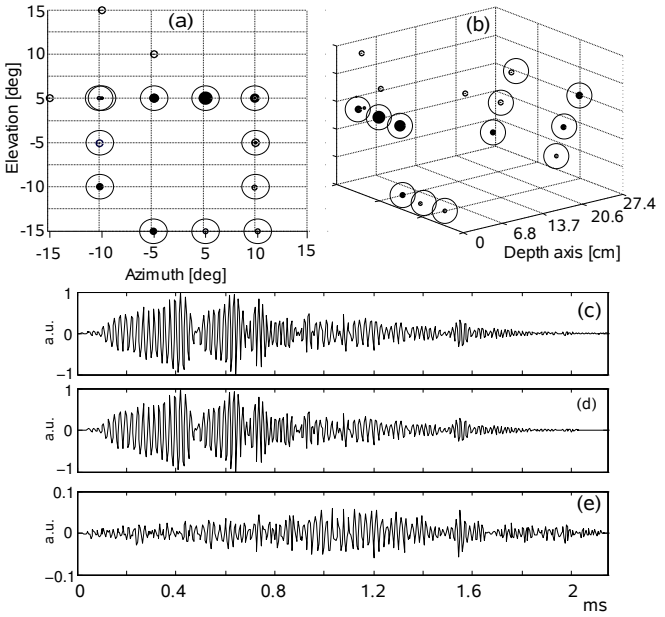

Fig. 8. 3D reconstruction of four objects, each made out of three wooden spheres. The big empty circles are centered on the true positions of the wooden spheres. The filled circles indicate the positions found with the sparse representation. Their radius is proportional to their value in the R vector. (a) Frontal view, (b) 3D view. (c) The measurement, (d) the reconstructed signal using the weights given in the sparse representation and (e) the error signal.

and a frontal view is shown in Fig.8(a). The measured signal (Fig.8(c)), the reconstructed signal (Fig.8(d)) and the error (Fig.8(e)) are also shown.

In the previous experiments, the signals from the ensonified objects consist of overlapping versions of signals used as bases. This assumption should be relaxed if one wants to estimate the position of a complex objects which might not be consider as a set of point reflectors. To do so, pieces of crumpled paper are fixed to a wire and mounted at different positions and configurations in front of the emitter. Most of the results only reconstruct parts of the objects, namely the parts reflecting most energy, as shown in Fig.9. Indeed, the directivity of the emitter biases the reconstruction towards the ensonified cone and in addition, all of the facets of the objects do not reflect directly at the emitter either. The technique can however be improved if the emitter is allowed to scan across space and/or move around the objects and partial reconstructions are combined. However, a few favorable cases allowed good reconstruction in 3D based on a single measurement only, as shown in Fig.10. Those measurements were often with objects elongated along the depth axis.

VI. Discussion

We have shown that sparse analysis could be used to estimate the positions of objects in 3D space. The algorithm performs very good as shown by the close similarity between the signal reconstructed from the sparse coefficients and the real measurement. We believe that the problems encountered are not inherent to the technique proposed but are due to some of the simplifying assumptions made. The scheme works

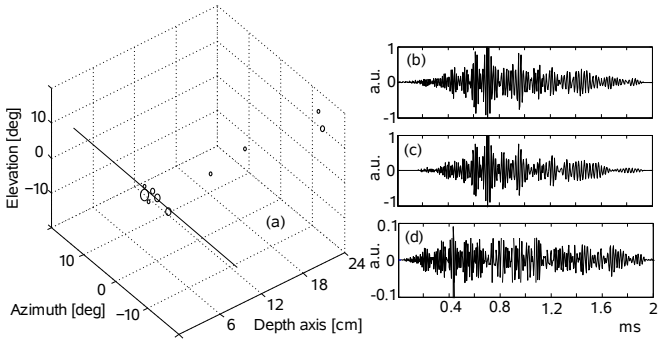

Fig. 9. (a) 3D reconstruction of a complex object. The solid line is the approximate position of the object. The circles indicate the positions found with the sparse representation. Their radius is proportional to their value in the R vector. (b) The measurement, (c) the reconstructed signal using the weights given in the sparse representation and (d) the error.

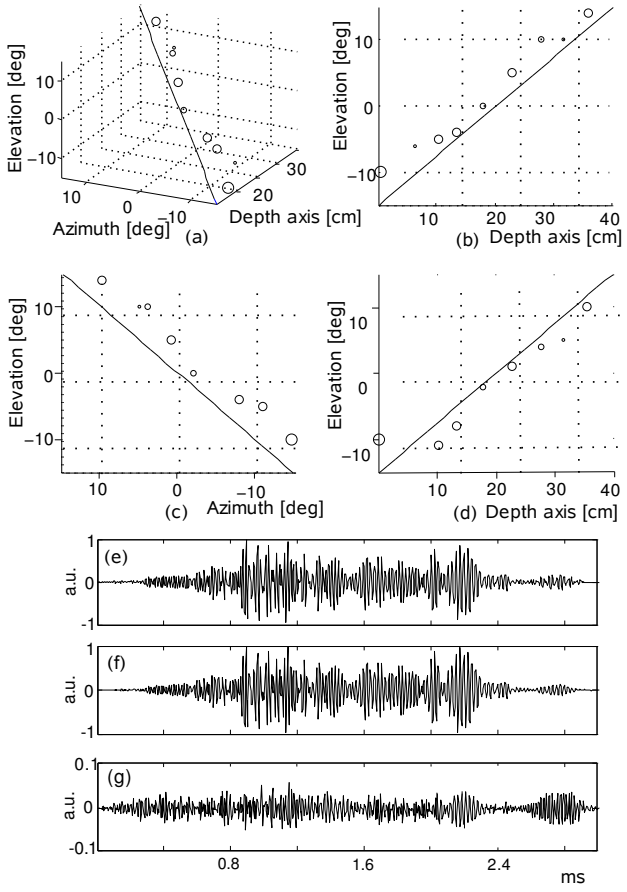

Fig. 10. (a) 3D reconstruction of a complex object. The solid line is the approximate position of the object. The circles indicate the positions found with the sparse representation. Their radius is proportional to their value in the R vector; (b) side view, (c) frontal view and (d) view from above. (e) The measurement, (f) the reconstructed signal using the weights given in the sparse representation and (g) the error.

very well when an object is composed of a superposition of base signals but performs less well when analysing complex object echoes. This is probably due to the fact that the facets of a complex object do not reflect exactly as spheres or perfect reflectors. This can be improved by choosing a more appropriate base. Possible solutions to address these problems could consist of averaging the features of various possible objects or learning the dictionary [11], [12].

The most important problem with the practical implementation of this scheme is the difference in returned energy between different directions. This can be due to the fact that not all facets of a reflecting object are pointing towards the emitter. However, the difference in emitted energy, due to the directionality of the emitter, limited the extent of the reconstructed space based on a single measurement most in our experiments. As we have seen, the sparse reconstruction will favor the strongest glints and will tend to disregard the weaker ones. Moreover, whereas the difference in energy is logarithmic, the sparse representation returns a linear scale solution. In an engineering context, this can be overcome by using a more omnidirectional emitter which would make sure that equal energy is radiated in all directions. Alternatively, moving around emitter and receiver [13], as the bat does by moving its pinnae and its head [14], would allow to combine partial reconstructions by focusing on different regions in space consecutively.

We propose that, apart from its engineering interest, studying echolocation in a sparse context could allow to understand better how the brain processes the information contained in the echoes. Indeed, sparse codes have been widely studied in neuroscience [15] and have proven to be especially efficient for vision [16]. A scheme similar to the approach proposed here has been successfully appied to neural networks [17]. Moreover, the very low number of spikes produced in the bat auditory system [18] could make sparse coding a very efficient alternative to rate coding.

REFERENCES

[1] S. Mallat, *A wavelet tour of signal processing: the sparse way*, third edition ed. Academic Press, 2009.

[2] M. Aytekin, E. Grassi, M. Sahota, and C. Moss, "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," *J. Acoust. Soc. Am.*, vol. 116, no. 6, pp. 3594–3605, 2004.

[3] J. M. Wotton, T. Haresign, and J. A. Simmons, "Spatially dependent acoustic cues generated by the external ear of the big brown bat, *Eptesicus fuscus*," *J. Acoust. Soc. Am.*, vol. 98, no. 3, pp. 1423–1445, 1995.

[4] B. A. Pearlmutter and A. M. Zador, "Monaural source separation using spectral cues," in *ICA*. Springer-Verlag, 2004, pp. 478–485.

[5] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Info. Theo.*, vol. 60, no. 6, pp. 1341–1344, 2004.

[6] C. Biber, S. Ellin, E. Shenk, and J. Stempeck, "The polaroid ultrasonic ranging system," in *67th Convention of the Audio Engineering Society*, New York, October 1980.

[7] A. Surlykke and C. F. Moss, "Echolocation behavior of big brown bats, eptesicus fuscus, in the field and the laboratory," *J. Acoust. Soc. Am.*, vol. 108, no. 5, pp. 2419–2429, 2000.

[8] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "A method for large-scale l1-regularized least squares," *IEEE J. on Sel. Topics in Sig. Proc.*, vol. 1, no. 4, pp. 606–617, 2007.

[9] F. M. Simmons, J.A. and C. Moss, "Echo-delay resolution in sonar images of the big brown bat, eptesicus fuscus," *Proc. Nat. Acad. Sci.*, vol. 95, pp. 2647–12 652, 1998.

[10] J. Blauert, *Spatial Hearing*. Cambridge: MIT Press, 1997.

[11] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, pp. 4311 – 4322, 2006.

[12] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Comp.*, vol. 12, no. 2, pp. 337–365, 200.

[13] V. A. Walker, H. Peremans, and J. C. T. Hallam, "One tone, two ears, three dimensions: A robotic investigation of pinnae movements used by rhinolophid and hipposiderid bats," *J. Acoust. Soc.Am.*, vol. 84, p. 16761679, 1998.

[14] J. Mogdans, J. Ostwald, and H. Schnitzler, "The role of pinna movement for the localization of vertical and horizontal wire obstacles in the greater horseshoe bat, rhinolophus ferrumequinum," *J. Acoust. Soc.Am.*, vol. 84, p. 16761679, 1988.

[15] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Curr. Opinion Neurobiol.*, vol. 14, pp. 481–487, 2004.

[16] ——, "Sparse coding with an overcomplete basis set: A strategy emploeyed by v1?" *Vision Reas.*, vol. 37, no. 23, pp. 3327–3338, 1997.

[17] H. Asari, B. Pearlmutter, and A. Zador, "Sparse representations for the cocktail party problem," *J. Neurosci.*, vol. 26, no. 28, pp. 7747–7490, 2006.

[18] M. Sanderson and J. Simmons, "Neural responses to overlapping fm sounds in the inferior colliculus of echolocating bats," *J. Neurophysiol.*, vol. 83, pp. 1840–1855, 2000.