



Review

A study of auditory localization mechanism based on thought experiments

Yi-de Zhang^{*}, Wei Liu

Department of Mathematics and Physics, North China Electric Power University, China

Available online 28 January 2019

Communicated by Felix Schoeller

Abstract

The focus of this study is auditory localization. Based on wave mechanics and structural mechanics, we analyze the sound field encircling external ear and the function of the soft tissues in middle ear respectively. And then, with the help of two rules, some details of the generation of spatial hearing are revealed. For auditory direction perception, three semicircular canals work as the reference coordinate system, and the cues are the direction of the concentrated force acting on tympanic membrane and the synchronous stress state. For the distance perception, because the distance information of the sound source is converted into some comparisons, the accuracy is closely related to the amount of useable sound sources. Therefore, the bad accuracy of the distance perception is inevitable in most cases. When it is necessary, many cues based on experience will be utilized to enhance the accuracy of the distance perception, which brings the diversity to auditory localization. Additionally, our results also can be used to explain some well known acoustic phenomena, such as auditory localization with one ear and the cocktail party effect.

© 2019 Elsevier B.V. All rights reserved.

Keywords: Auditory localization; Direction perception; Distance perception; Thought experiment

1. Introduction

Position detection is one of basic survival skills of living things, which plays an important role in predation, escape, migration and other behaviors. Also, position detecting techniques are significant over the course of human history, with which the footprint of mankind has been all over the world and is extending to outer space.

Even today, it is still in dispute how auditory system locates a sound source. One popular opinion was presented by Rayleigh and Strutt [1], which is called Duplex theory. According to Duplex theory, just because auditory system consists of two ears, the interaural time difference and interaural intensity difference exist, which are two important cues of auditory localization. However, the findings of two acoustic phenomena, i.e., auditory localization is not influenced by the azimuth ambiguity [2,3] and one ear also can orient a sound source [4], make this theory deep in crisis.

^{*} Corresponding author at: 619 North Yonghua Street, Baoding, 071003, China.
E-mail addresses: 51752062@ncepu.edu.cn, zhy___zhy@sina.com.cn (Y. Zhang).

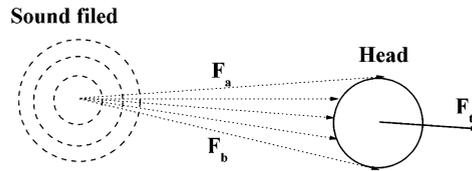


Fig. 1. The symmetry of the sound field. This sound field is excited by a point source. The dash line denotes the wave front of the sound field, the dot line with an arrow denotes the force acting on the surface of the head with a certain incident angle such as F_a and F_b , and the solid line with an arrow denotes the concentrated force acting on the head.

To explain such acoustic phenomena, kinds of attempts are made. Some researchers think that the Pinna Effect [5,6] is helpful, because lots of useful information is carried by the reflection and diffraction in the journey of sound waves from the source to the ear. With more and more cues are found, people regard auditory localization as a complex procedure influenced by many factors, and describe it by the head related transfer function [7–9] (HRTF). Additionally, as auditory localization is a physiological function rather than an acoustic phenomenon alone, the methods or theories rooting in other subjects also are utilized to analyze the generation of spatial hearing, such as neurology [10], psychology [11], statistics [12], etc.

Another difficulty in the study of auditory localization is to verify whether a cue is really used in auditory localization or not. And the reasons are not only in techniques but also in the poor known about the conversion between sound signals and nervous impulses. Therefore, two rules of the signal conversion are introduced in this article, whose reasonability is based on thought experiments. In the discussion section, several sound phenomena are explained with our findings, which also can be seen as the verification of our results.

2. Theory

2.1. Acoustical field encircling external ear

As shown in Fig. 1, middle ear is separated from external ear by tympanic membrane and connected to inner ear via ossicular chain. External ear, middle ear and inner ear are in charge of sound signal receiving, transmitting and processing respectively.

It is troublesome to show the details of the sound field encircling external ear, because many influence factors exist, such as the obstacles between the sound source and the head, the shape of head and the physiological structure of external ear. Fortunately, as the elasticity modulus of tympanic membrane is much smaller than that of bones, the stress concentration will appear in the region of tympanic membrane, which means that the force acting on tympanic membrane F_t can be computed by a surface integral

$$F_t = \int_S F dS \tag{1}$$

where the outside surface of head S is spherical approximately for human beings and the vector function F denotes the force acting on each point in S .

As all sources can be seen as the superposition of point sources, a point source is used here. On the other hand, due to the diffraction of sounds, most obstacles cannot change the propagation direction of sound waves. In other words, the influence of obstacles can be ignored. And then, with the assumption of spherical head, it is true that the direction of the concentrated force F_t in formula (1) is in the line from the source to the center of the head, because of the symmetry of the sound field, referring to Fig. 1.

In most cases, the restriction on the shape of head can be abolished, because the maximal angle between two forces, such as F_a and F_b in Fig. 1, acting on the surface S is very small. For example, according to some geometrical relationships, when the size of head is 0.1 m and the distance of the sound source is 2 m, the maximal angle is no more than 6° . Furthermore, this angle will decrease as distance of the sound source increases.

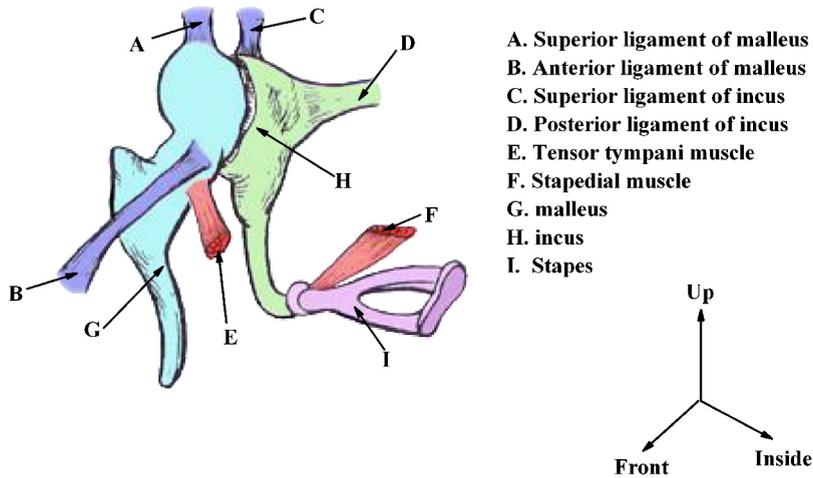


Fig. 2. Ossicular chain and some soft tissues in middle ear.

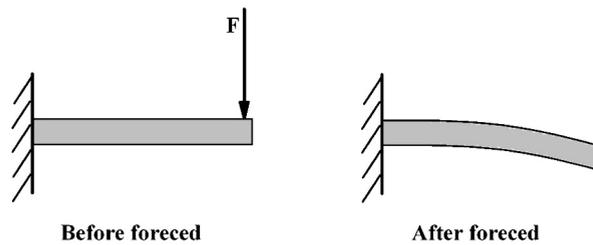


Fig. 3. The deformation of a cantilever beam when forced. The soft tissue in middle ear is regarded as a cantilever beam, because one end of it attaches to the ear bone, which is free, and the other end of it attaches to the inner wall of the cranial cavity, which is fixed.

2.2. Sound energy transmission in middle ear

The sound-conducting system in middle ear is comprised of ossicular chain and some soft tissues, referring to Fig. 2. At two ends of ossicular chain, malleus and stapes are attached to tympanic membrane and oval window via annular ligaments respectively. One function of the ligaments is to make ossicular chain suspended in tympanic cavity.

If all components of the sound-conducting system including two joints of ossicular chain are rigid, it is true that the direction of the concentrated force caused by the outside sound field and acting on tympanic membrane will be maintained until this force is transferred onto oval window, and so does the sound energy flow.

Unfortunately, the ligaments and muscle tendons in middle ear are not rigid enough. Some researches [13] indicate that stapes footplate just has one mode of vibration, i.e., translational motion, if the frequency of the sound wave is lower than 700 Hz, and the vibration of stapes footplate contains both translational motion and rotational motion if the frequency is higher than 5 kHz. Actually, the rotational motion is caused by the flexibility of the soft tissues in middle ear. In the following, we will analyze how the flexibility of the sound-conducting components influences the motion of stapes footplate.

Firstly, the ligaments not including annular ligaments and muscle tendons are considered. As shown in Fig. 3, this kind of soft tissue can be seen as cantilever beams forced at the free end. Based on the theory in mechanics of materials, some results are shown below.

The axial component and the radial component of a force are named as F_x and F_r respectively. The axial displacement of the free end u_x is

$$u_x = 4lF_x / (\pi E d^2), \tag{2}$$

where l , E , and d respectively denotes the length, Young modulus, and cross sectional diameter of the cantilever beam. The radial displacement of the free end u_r is

$$u_r = 64l^3 F_r / (3\pi E d^4), \quad (3)$$

and then, the ratio of the axial displacement and the radial displacement is

$$u_x/u_r = 3d^2 F_x / (16l^2 F_r). \quad (4)$$

For the ligaments and muscle tendons in Fig. 2, because of $d \ll l$, their capability of compressive deformation is much lower than that of the bending deformation.

Moreover, since these ligaments and muscle tendons attached to ossicular chain constitute a tridimensional system, when the force acting on ossicular chain goes beyond the limit, the exceeding part will be collectively borne by the soft tissues in different directions. Thus, under the influence of these soft tissues, the redundant sound energy flow will be preferentially converted into shearing strain energies, which leads to the rotational motion of three auditory ossicles; the rest of the sound energy flow is transferred into inner ear while the direction is remained, which leads to the translational motion of auditory ossicles.

Secondly, the annular ligaments on tympanic membrane and stapes footplate are considered. This kind of soft tissue can be seen as a cantilever beam with an annular cross section, and the results are as follows

$$u_x = 4l F_x / [\pi E (b^2 - a^2)], \quad (2a)$$

$$u_r = 64l^3 F_r / [3\pi E (b^4 - a^4)], \quad (3a)$$

$$u_x/u_r = 3(b^2 + a^2) F_x / (16l^2 F_r), \quad (4a)$$

where b and a respectively denotes the outer diameter and the inner diameter of the cross section. For two annular ligaments, because of $a, b \gg l$, their capability of compressive deformation is much higher than that of the bending deformation, which means that the flexibility of the annular ligaments mainly bring translational motions. In fact, such translational motion can be neglected, because the length l is very short.

Thirdly, two joints of ossicular chain are considered. Here, the hinge joint is used to describe the flexibility feature of the joint, which means that the moment of force cannot be transmitted. As the rotational motion of auditory ossicle can be seen as the effect of the moment of couple, the role of the joint is to reduce the amplitude of the rotational motion. Specifically, the sound energies carried by the rotational motion are partially converted into strain energies of the joint.

Actually, as cochlea is filled with lymph whose shear modulus approximately is zero, the sound energies carried by rotational motions will be restricted in a small region and cannot excite auditory neurons. And then, the helpful information in auditory localization is carried by translational motions.

From the above, the rotational motion relates to the flexibility of the sound-conducting system and the translational motion relates to the rigidity of the sound-conducting system. Additionally, the rigidity of the sound-conducting system maintains the directivity of the sound energy flow or the concentrated force. Consequently, we have two results as shown below.

The rigidity of the sound-conducting system brings translational motions to auditory ossicles. Via such kind of translational motion, the direction information of the sound source is transmitted into inner ear.

The flexibility of the sound-conducting system brings rotational motions to auditory ossicles, and then the redundant sound energies are translated into the strain energies of some soft tissues. In other words, the main function of the soft tissue is to reduce the damage to inner ear when the intensity of the sound energy is too high.

2.3. The generation of direction perception

Since the sound wave in the air is the longitudinal wave, the direction of the concentrated force caused by the sound field is in the line from the source to the head rather than always towards the head or the source. And then, the direction information extracting cannot be realized only with the direction measurement of the concentrated force acting on tympanic membrane. Fortunately, when acoustic nerves are stimulated, the synchronous stress state of them also is known. Actually, once the direction of the force and the synchronous stress state are recorded, the sound source will be oriented. Specifically, when a tensile force works, the direction of the force points to the sound source, otherwise the source is in the opposite direction.

In wave theories, the direction of the force coincides with that of the particle velocity and the stress state can be described by the pressure. And then, the principle of signal processing procedure also can be expressed by the formulas below. Readers unfamiliar with wave theory can skip these formulas.

$$P_n = -v_n \cdot p, \quad n = x, y, z \quad (5)$$

$$\alpha = \arctg(P_y/P_x), \quad \text{if } P_x < 0 \quad \text{and} \quad P_y < 0 \quad (6a)$$

$$\alpha = \arctg(P_y/P_x) + \pi/2, \quad \text{if } P_x > 0 \quad \text{and} \quad P_y < 0 \quad (6b)$$

$$\alpha = \arctg(P_y/P_x) + \pi, \quad \text{if } P_x > 0 \quad \text{and} \quad P_y > 0 \quad (6c)$$

$$\alpha = \arctg(P_y/P_x) - \pi/2, \quad \text{if } P_x < 0 \quad \text{and} \quad P_y > 0 \quad (7a)$$

$$\beta = \arctg \left| P_z / \sqrt{P_x^2 + P_y^2} \right|, \quad \text{if } P_z < 0 \quad (7b)$$

$$\beta = \arctg \left| P_z / \sqrt{P_x^2 + P_y^2} \right| - \pi/2, \quad \text{if } P_z > 0 \quad (7c)$$

In formula (5), (6) and (7), the subscript denotes the direction, p is pressure, v_n is the component of particle velocity, P_n is the component of acoustic Poynting vector [14], α is the azimuth angle of the source and β is the dip angle of the source.

Before showing the generation of auditory distance perception, one rule about the generation of perception is introduced. The expression of information is composed of two parts. One is used for denoting the quantity, which usually appears in the form of numbers. The other is used for denoting the reference object, which usually appears in the form of units. Therefore, the reference object is necessary in the generation of perception.

Similarly, there are two key points in the generation of auditory direction perception, i.e., how those two angles are accepted by nerve system and who plays the role of the reference object. For the former, the size of the angles can be converted into the amplitude ratios of nervous impulses. For the latter, three semicircular canals are suitable, since they are mutually perpendicular. In details, each component of the force and the synchronous stress state will lead to one response of the nerve in certain semicircular canal.

According to the above, the generation of auditory direction perception is summarized as follows.

When stapes vibrates under the influence of the outside sound field, the auditory nerve is excited by the translational motion and then generates some nervous impulses carrying two kinds of information. The direction of the translational motion is translated into the amplitude ratio of the nervous impulses, and the synchronous stress state of the auditory nerve is expressed as the polarity of another kind of nervous impulse. Finally, the direction perception generates, with the help of three semicircular canals.

2.4. Auditory distance perception

In 1999, Bronkhorst and Houtgast [15] pointed out that auditory distance perception in a room relates to the energy ratio of direct and reflected sound. Except for their work, very poor researches are focused on auditory distance perception. Actually, the generation of auditory distance perception is quite different from that of auditory direction perception, especially in the conversion between sound signals and nervous impulses. In this part, we will show why the difference exists and what it brings.

One important difference between the direction information and the distance information is their value ranges, i.e., the value range of the former is limited and that of the latter is infinite. As mentioned in section 2.3, the expression of information generally includes two parts. One denotes the quantity and the other denotes the reference object. Because of the difference in value range, these two kinds of information cannot be converted into nervous impulses in the same way. In details, for the direction information, it is feasible to convert the quantity into the amplitudes of nervous impulses directly, and for the distance information, such kind of conversion will be rejected.

Actually, when the outside information with an infinite value range is converted into nerve impulses, it is avoided as far as possible that the amplitude of the nerve impulse is directly related to the part denoting the quantity, which is named as the rule of amplitude avoiding in this paper. Although this rule is violated sometimes, such as the generation of the pain, the violation contains the risk. For example, without the treatment, the sharp pain for a long time will bring death to the patient. Since the distance perception is generated constantly, we think that the rule of amplitude avoiding must be obeyed in its generation procedure.

According to the rule of amplitude avoiding, the reference objects used in distance perception are also different from those in direction perception essentially. For the direction perception, a fixed reference object, such as the semi-circular canal, is acceptable, while it is not so for the distance perception. If a fixed reference object for the distance perception really exists, the rule of amplitude avoiding will be violated inevitably, especially when the source is very far. Therefore, another kind of expression of the distance information is employed in the signal processing of auditory distance perception, i.e., the size of the distance is converted into some comparisons, such as $C < A < B$ denotes that B is further than A and A is further than C.

Due to the application of comparison, the accuracy of distance perception is sensitive to the amount of reference objects. Because auditory distance perception possesses a bad accuracy, we verify this inference by a visual phenomenon. Compared with the visual distance perception in the daytime, there is a sharp decline in the accuracy of that at night. If the distance information is expressed in the form of numbers and units, the decline in accuracy would not appear. Thus, the decline in accuracy is caused by the application of comparison in distance perception. Plenty of natural light makes lots of objects become reflectors, which also can be seen as light sources, in the daytime, while only a small number of objects emit or reflect light at night. In other words, the amount of reference objects in the daytime is much more than that at night. And then, the sharp decline in accuracy appears.

With the analysis above, it is true that the amount of the available sound sources is closely related to the accuracy of auditory distance perception. In a natural environment, although many things can make sounds, the duration of the sound usually is short. Moreover, because of the attenuation [16], a sound source cannot be detected by auditory system if it is very far. Additionally, auditory masking effects [17] make weak sound signals difficult to be detected. Consequently, the effect of auditory perception is similar to that of visual perception at night, i.e., the inherent auditory distance perception has a very bad accuracy.

It must be pointed out that the bad accuracy of auditory distance perception is caused by the signal processing procedure of auditory nerves and is independent of the cues of sound localization. In other words, it is hard to enhance the accuracy of auditory distance perception generally. And this is why normal persons are accustomed to acquire the distance information via visual system rather than via auditory system.

3. Discussion

Firstly, the cues of auditory localization in this article are based on a monaural model, which means that our study reveals the working mechanism of the orientation via one ear. Furthermore, the one to one correspondence is established between the direction information of a sound source and the features of the sound signals, which can be used to explain why the azimuth ambiguity not exists in auditory localization.

Secondly, auditory direction perception has the excellent noise immunity, which is reflected in the cocktail party effect [18]. This phenomenon relates to the generation of auditory direction perception. As deduced in section 2.3, the cues of the direction perception are not the real amplitudes of sound signals but the amplitude ratios of them. Moreover, since each sound wave has its own crest and trough, a weak sound signal also has the opportunity of being highlighted. Additionally, the isotropous absorption of sound energy cannot influence the orientation of a sound source, according to formula (5), (6) and (7). Therefore, in some cases, although listeners cannot know what the speaker says, they can know where he/she is. Actually, because the same detection principle is used, the more discussions on the noise immunity of auditory direction perception can be found in the researches on vector sound field technique [19,20].

Thirdly, based on the rule of amplitude avoiding, an interesting inference is gained, that the cues provided by binaural effect are difficult to be used by auditory system. The basis of binaural effect is that the distance exists between two ears. And then, any kind of cue derived from binaural effect is connected with this distance. If such distance is not utilized, both the direction information and the distance information of the source can hardly be described, which conflicts with the good accuracy of auditory direction perception. On the other hand, once the distance is used, it would act as a fixed reference object, which conflicts with the rule of amplitude avoiding. Consequently, although these cues really contain the position information of the sound source, they do not play the important roles in auditory localization as people expected.

Finally, some features in the evolution of auditory system accord with the inferences of this study. It must be pointed out that the evolution here mainly relates to auditory distance perception, since the accuracy of the direction perception is good enough in most cases. Theoretically, because of the application of comparison, the most direct way to improve the distance perception is to increase the amount of reference objects. As natural sound sources cannot emit

sounds continuously, the auditory system of some animals, such as bats and dolphins, evolves into biosonar system. One important function of biosonar system is to convert the objects not emitting sound waves into acoustic reflectors so that the amount of reference objects is increased.

However, most living things do not possess biosonar system and they choose another way to improve their distance perception, i.e., the physiological function of auditory distance perception is replaced by the logical deduction. If a certain kind of logical deduction is frequently used, it will evolve into conditioned reflex. And this is why the persons with visual impairment have the good auditory distance perception in their homes. Unfortunately, for these persons, the distance perception will become bad if the sound scene is changed, which indicates that the memory trace plays the role of the reference object in their distance perception.

Actually, besides of biosonar system, other ways to improve auditory distance perception generally root in experience. As mentioned in section 2.3, the reference object is necessary in the generation of distance perception. Without biosonar system, the reference object outside bodies cannot meet the requirement, because the amount is not enough. And then, the only way to acquire suitable reference objects is to search the memory trace.

Due to the individual difference, the improvement rooting in experience will make auditory distance perception present diversity and developmental plasticity, which are partially verified by the study of Keating, Dahmen and King [10]. Furthermore, just because of these two features, it is nearly impossible to find a commonly used cue of auditory distance perception.

4. Conclusion

Our results are summarized as follows.

(1) The direction of the concentrated force acting on tympanic membrane is in the line from the sound source to the center of the head. Via the translational motion of auditory ossicles, this kind of directivity is transmitted into inner ear, which is the cue of auditory direction perception.

(2) The direction of the concentrated force and the synchronous stress state are converted into nervous impulses, and then auditory direction perception is generated. In this signal processing, semicircular canals work as the reference coordinate system of the direction perception.

(3) To avoid the damage caused by the stimulation with high intensity, the distance information of sound sources is expressed in the form of comparison in nerve conduction. Therefore, the accuracy of the distance perception relates to the amount of the available sound sources, which leads to the low accuracy of auditory distance perception. In practice, to enhance the accuracy, the cues rooting in experience usually participate in the generation of the distance perception. And this is why it is difficult to find a widely used cue of the distance perception.

Acknowledgements

We thank the Guest editor Felix Schoeller and the anonymous reviewers for their suggestions to revise the manuscript. We also thank Prof. Hengshan Hu and Genshan Jiang for their advice and help to polish up the manuscript. This work is supported by the National Natural Science Foundation of China (No. 11674093 and No. 11474091). This work is also supported by State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences (SKLA201808).

References

- [1] Rayleigh L, Strutt J. XII. On our perception of sound direction. *Philos Mag Ser 6* 1907;13(74):214–32.
- [2] Wallach H. On sound localization. *J Acoust Soc Am* 1939;10(4).
- [3] Zhang Y-d, Hu H. A technique to eliminate the azimuth ambiguity in single-well imaging. *Geophysics* 2014;79(6):D409–16.
- [4] Slattery WH, Middlebrooks JC. Monaural sound localization: acute versus chronic unilateral impairment. *Hear Res* 1994;75(1):38–46.
- [5] Batteau DW. The role of the pinna in human localization. *Proc R Soc Lond B, Biol Sci* 1967;168(1011):158–80.
- [6] Musicant AD, Butler RA. The influence of pinnae-based spectral cues on sound localization. *J Acoust Soc Am* 1984;75(4).
- [7] Krysl P, Cranford TW. Directional hearing and head-related transfer function in odontocete cetaceans. In: *Basic aspects of hearing: physiology and perception*, vol. 875. 2016. p. 583–7.
- [8] Gamper H. Head-related transfer function interpolation in azimuth, elevation, and distance. *J Acoust Soc Am* 2013;134(6):EL547.
- [9] Kulkarni A, Colburn HS. Role of spectral detail in sound-source localization. *Nature* 1998;396(6713):747–9.

- [10] Keating P, Dahmen JC, King AJ. Complementary adaptive processes contribute to the developmental plasticity of spatial hearing. *Nat Neurosci* 2015;18(2):185–7.
- [11] Kuk F, et al. Effects of training on the use of a manual microphone shutoff on a BiCROS device. *J Am Acad Audiol* 2015;26(5):478–93.
- [12] Shinn-Cunningham BG, Kopco N, Martin TJ. Localizing nearby sound sources in a classroom: binaural room impulse responses. *J Acoust Soc Am* 2005;117(5):3100–15.
- [13] Blauert J. *Communication acoustics*. Berlin, Heidelberg: Springer; 2005.
- [14] Auld BA. *Acoustic fields and waves in solids*. John Wiley; 1973.
- [15] Bronkhorst AW, Houtgast T. Auditory distance perception in rooms. *Nature* 1999;397(6719):517–20.
- [16] Aki K, Richards PG. *Quantitative seismology: theory and method*. W. H. Freeman & Co.; 1980.
- [17] Ma D. *Basic theory of modern acoustics*. Science Press; 2004.
- [18] Hawley ML, Litovsky RY, Culling JF. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *J Acoust Soc Am* 2004;115(2):833–43.
- [19] Hawkes M, Nehorai A. Acoustic vector-sensor correlations in ambient noise. *IEEE J Ocean Eng* 2001;26(3):337–47.
- [20] Hawkes M, Nehorai A. Acoustic vector-sensor beamforming and Capon direction estimation. *IEEE Trans Signal Process* 1998;46(9):2291–304.