Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2016) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

2016 IEEE International Symposium on Robotics and Intelligent Sensors, IRIS 2016, 17-20 December 2016, Tokyo, Japan

Predictive Acoustic Tracking with an Adaptive Neural Mechanism

Danish Shaikh[a,*], Poramate Manoonpong[a]

[a] *Embodied AI and Neurorobotics Laboratory, Centre for BioRobotics, Maersk Mc-Kinney Moeller Institute, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark*

Abstract

Tracking an acoustic signal in motion is pertinent in several domains such as human-robot interaction and search-and-rescue robotics. Conventional approaches to acoustic tracking acquire time-of-arrival-difference signals from multi-microphone arrays and localise the acoustic signal using Kalman or particle filtering, generalised cross-correlation or steered response power techniques. The authors have previously developed a biologically-inspired mechanism that utilises two microphones to reactively track an acoustic signal in motion. The mechanism leverages the directional response of an mathematical model of the lizard peripheral auditory system to extract information regarding sound direction. This information is utilised by a neural machinery to learn the acoustic signal's velocity through fast and unsupervised correlation-based learning adapted from differential Hebbian learning. This approach has previously been validated in simulation and via robotic trials to track a continuous pure tone acoustic signal with a semi-circular motion trajectory and a constant but unknown angular velocity. The neural machinery has been shown to be able to learn different target angular velocities in independent trials. Here we extend our previous work by demonstrating that an identical instance of the mechanism can be used to successfully predict the future spatial location of an acoustic signal with an identical semi-circular motion trajectory and a constant but unknown angular velocity. We evaluate the prediction performance of the simulated mechanism in independent trials for three different angular velocities.

Peer-review under responsibility of organizing committee of the 2016 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2016).

Keywords: motion extrapolation, predictive acoustic tracking, correlation-based learning, sound localisation, lizard peripheral auditory model

1. Introduction

Predicting the future spatial location of an acoustic signal can be relevant in several applications. Verbal human-robot interaction in social robots is deemed richer if the robot's auditory focus is oriented and maintained towards the speaker [1,2] throughout their spatial trajectory via appropriate acoustomotor responses, for example a human walking around in a room while addressing the robot via speech commands. In audio-visual teleconferencing systems, automatically steering microphone systems that follow a speaker as they move about in a room could dynamically

* Corresponding author. Tel.: +45-65509526 ; fax: +45-66157697.
E-mail address: danish@mmmi.sdu.dk

1877-0509
Peer-review under responsibility of organizing committee of the 2016 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2016).

maximise the power of the incoming audio signal or orient the camera towards the current speaker[3,4]. In robot phonotaxis, the robot could localise acoustic sources and navigate towards them[5,6].

Tracking an acoustic signal along an unknown but fixed motion trajectory with an unknown but constant velocity entails knowledge of its instantaneous spatial location. Depending on the sound frequency, either the sound level difference (interaural intensity difference or IID), the time-of-arrival-difference of sound (the interaural time difference or ITD) or both can be used for instantaneous localisation. This necessitates an arrangement with at least two microphones with a fixed displacement between them. Successful repetition of this localisation at a sufficiently fast rate can then minimise the tracking error. Here we attempt to predictively track an acoustic signal in motion using only ITD information. An acoustic signal in motion in a given direction and a constant velocity with respect to the microphones generates dynamically varying ITD cues. The microphone separation and the relative instantaneous position of the acoustic signal with respect to the median plane determine the instantaneous values of these cues. The rate of variation of these cues is dependent on the relative velocity of the acoustic signal. Tracking an acoustic signal in motion therefore entails the transformation of such velocity- and relative position-dependent cues into some desired behaviour. For orientation behaviour that requires a motor response, any processing delays in the sensorimotor pathway must be compensated for and this is possible by extrapolating the target's motion and thus predicting its future position.

Common techniques for reactive acoustic target tracking extract ITD information via multi-microphone arrays[7,8,9] with at least 4 microphones that are arranged as either linear, square or circular arrays or are spatially distributed. These techniques use particle filtering algorithms[10,11] to compute the relative sound source location from raw ITD data; conventional approaches[12,13,14,15,16] are based on generalised cross-correlation[17] or steered response power[18,19]. A greater number of microphones can improve localisation accuracy, but this requires greater computational complexity and expensive hardware to synchronise and process multi-channel acoustic signals. However, to the best of our knowledge predictive robotic tracking of moving sound has not been reported in the literature.

We have previously reported a reactive tracking mechanism[20] with two microphones, which couples a model of the lizard peripheral auditory system[21] with a neural learning machinery. The lizard peripheral auditory system provides relative directional information about the acoustic signal. It has been extensively studied via bio-faithful mathematical modelling utilising biophysical data *in vivo* to determine the parameters of the model, as well as via various robotic implementations[22]. The mechanism has been validated in simulation and in robotic trials for reactive acoustic tracking where it learned various target angular velocities in separate trials[20]. We extend our earlier work in the following manner. We implement an identical instance of the previously proposed neural mechanism in a robotic agent in simulation to first learn the constant but unknown angular velocity of a virtual and continuous pure tone acoustic signal in motion following a semi-circular trajectory. We then demonstrate that, by removing the criterion that stops the learning in the learning algorithm, the learning continues stably after the correct target angular velocity has been learned. The mechanism subsequently learns a new angular velocity that allows the robotic agent to predict and orient towards the future spatial location of the target.

The remainder of this article is structured as follows. The lizard peripheral auditory model and its directional response is described in Sect. 2. Section 3 describes the neural mechanism and the experimental setup. Prediction performance of the proposed approach is reported in Sect. 4. The research is summarised in Sect. 5 and further research is highlighted.

2. Background

The remarkable directionality, i.e. the ability to extract the relative position of a relevant acoustic signal, of the lizard peripheral auditory system[23,24] found in *Gekko gecko* (commonly known as the tokay gecko, Fig. 1**A**) can be attributed to the internal acoustical connection between the animal's two eardrums established by efficient sound transmission via passages in the head's interior (Fig. 1**B**). Although the eardrum separation for most lizard species is between 10–20 mm there is strong response to sound wavelengths within 340–85 mm (corresponding to sound frequencies within 1.0–4.0 kHz), where the sound diffracting over the animal's head results in insignificant sound pressure difference between the eardrums and thus almost negligible (1–2 dB) IID information[24]. The system essentially converts μs-scale interaural phase differences between sound at the two eardrums (which correspond to ITDs) into relatively greater (up to 40 dB) interaural vibrational amplitude differences[23] that encode sound direction information. The superposition of two acoustic components determines each eardrum's vibrations–an external sound pressure at the

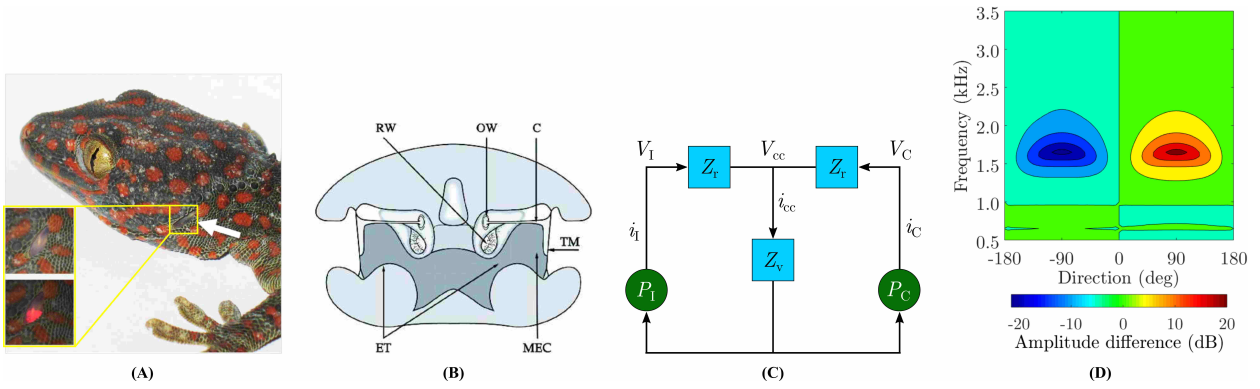

Fig. 1. **A.** A tokay gecko eardrum (modified from [24]). **B.** Cross-section of the *Sceloporus* lizard peripheral auditory system (borrowed from [23]). **C.** Electrical circuit equivalent of the peripheral auditory system (based on [25,26] and redrawn from [27]). **D.** Directionality as binaural subtraction of the two eardrum responses [Eqn. (1)]. The response is good within a 1.0–2.2 kHz range, with a peak response at approx. 1.6 kHz.

eardrum's periphery and an internal equivalent sound pressure at its interior generated due to acoustic interference in the internal passages. Therefore the ear nearer to the relevant acoustic signal exhibits stronger vibrations as compared to the ear more distant from it, and the vibration intensities depend on the sound frequency and the interaural phase difference. An equivalent electrical circuit modelled after the peripheral auditory system [25,26] (Fig. 1**C**) allows the directionality to be visualised (Fig. 1**D**) as the difference between the vibrational amplitudes i_I and i_C of the ipsilateral (towards the acoustic signal) and contralateral (opposite to the acoustic signal) eardrums respectively, as given by

$$\left|\frac{i_I}{i_C}\right| = \left|\frac{G_I \cdot V_I + G_C \cdot V_C}{G_C \cdot V_I + G_I \cdot V_C}\right| \equiv 20\,(\log|i_I| - \log|i_C|)\ \text{dB}\ . \tag{1}$$

Frequency-dependent gains G_I and G_C respectively reflect the effect of sound pressure on the ipsilateral and contralateral eardrum motion. They are determined experimentally from eardrum vibration measurements via laser vibrometry [23] and are digitally implemented as 4th-order infinite impulse response bandpass filters. The ratio $\frac{i_I}{i_C}$ is positive for $|i_I| > |i_C|$ and negative for $|i_C| > |i_I|$. The model's symmetry implies that $\left|\frac{i_I}{i_C}\right|$ is identical with respect to the median at $\theta = 0°$ as well as locally symmetrical within the pertinent range of sound direction $[-90°, +90°]$. Equation (1) defines a differential signal whose sign specifies sound direction as arriving from the ipsilateral (positive sign) side or from the contralateral (negative sign) side. Its magnitude relates non-linearly to the relative angular position of the acoustic signal with respect to the midpoint of the head.

3. Materials and methods

We define the task of acoustic motion extrapolation as follows–a robotic agent must learn an appropriate angular velocity that aligns the agent sufficiently quickly with the extrapolated future spatial location of an acoustic signal in motion. The signal moves with an unknown but constant angular velocity in a given direction along a pre-defined arc-shaped semi-circular trajectory. To solve this task we devise an adaptive closed-loop learning mechanism [20] embedded in the task environment (Fig. 2**A**). The mechanism combines the auditory processing of the lizard peripheral auditory model, which provides sound direction information and the Input Correlation (ICO) learning algorithm [28], which learns appropriate synaptic weights that determine the agent's angular velocity. The synaptic weights encode the temporal relation between the sound direction perceived by the lizard peripheral auditory model *preceding* and *succeeding* the agent's spatial rotations. Since this relation is inversely proportional to the angular velocity of the acoustic signal, a fixed set of synaptic weights can only represent a fixed angular velocity and these must be re-learned for a new angular velocity.

The experimental setup in simulation (Fig. 2**B**) comprises a two-dimensional virtual loudspeaker array, with 37 loudspeakers arranged in a semi-circle, which generates relevant tones. Consecutive loudspeakers are separated by a 5° angular displacement. Acoustic motion is simulated by sequential playback, one loudspeaker at a time, beginning

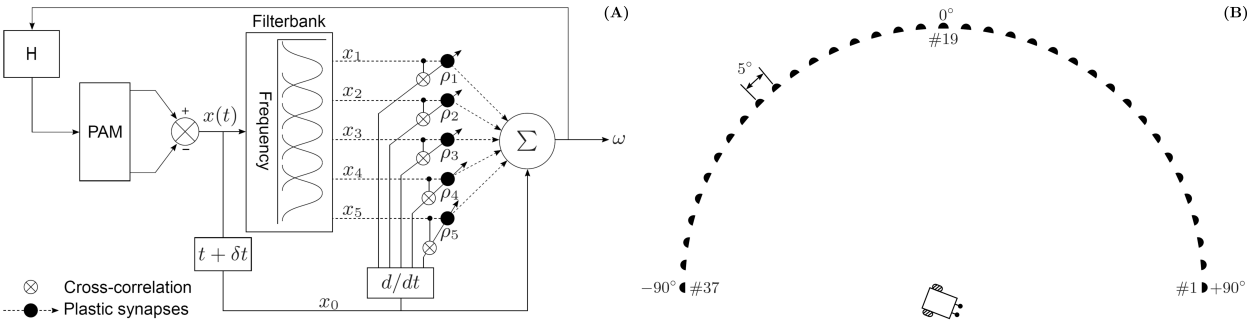

Fig. 2. **A.** The closed-loop learning mechanism. It computes angular velocity ω, defined as the angular deviation per time step, required to align the robotic agent with the extrapolated future spatial location of the acoustic signal in one time step. During learning, ω is converted into corresponding ITD information via the environmental transfer function H. The peripheral auditory model PAM converts these cues into a differential signal $x(t)$ [Eqn. (1)] that encodes information about sound direction. Decomposition of $x(t)$ into its frequency components $x_k(t)$, where $k = 1, \ldots, N$, is done by a filter bank to extract frequency information. The filter bank comprises five bandpass filters, each implemented digitally as a second-order infinite impulse response filter. The centre frequencies of these filters lie at every 200 Hz between 1.2–2.0 kHz. Each filter has a 3 dB cut-off frequency of 200 Hz. Due to the non-linear response of the peripheral auditory model to sound frequency, in the absence of sound frequency information it provides ambiguous sound direction information, thus necessitating a filter bank. The magnitude responses of the filters represent spectro-temporal receptive fields [29] of individual auditory neurons, i.e. the range of sound frequencies that provide optimal neuronal stimulation. $x_k(t)$ are correlated with the derivative of $x_0(t)$ to update the synaptic weights ρ_k. $x_k(t)$ are "predictive" signals that determine the future spatial location of the acoustic signal before turning, while $x_0(t)$ is the "retrospective" signal generated after turning. **B.** The experimental set-up in simulation (taken from [20]).

from either end of the array. Sound continuity is maintained by instantaneous playback of consecutive loudspeakers, thus simulating an acoustic signal in continuous motion, albeit in discrete steps. The angular velocity, defined as the angular displacement in radians every 10 time steps, is set by playing any given loudspeaker for 10 time steps. This process is repeated until the end of the array is reached, marking the end of one learning iteration. The direction of acoustic motion is from loudspeaker #1 to the left to loudspeaker #37 to the right of the array. The robotic agent rotates on a fixed axis perpendicular to and passing through the centre point of the diameter of the semi-circle. To predict the future spatial location of the acoustic signal, the robotic agent must first determine the signal's angular velocity and then extrapolate its motion by turning with a adequately greater angular velocity to align with the *next* spatial location of the acoustic signal along its trajectory in *one* time step.

The learning takes place as follows. Initially the robotic agent is oriented towards a randomly chosen spatial location (97°). Loudspeaker #1 then plays a 2.2 kHz tone signal, a frequency chosen to ensure that the peripheral auditory model can extract adequate directional cues, and the robotic agent uses this information to compute $x_k(t)$ and estimate the angular velocity [Eqn. (2)] with which to turn towards the said loudspeaker. After completing the turn, the robotic agent again uses the extracted sound direction information to compute $x_0(t + \delta t)$ and updates the synaptic weights ρ_k appropriately [Eqn. (3)]. This process is repeated for each loudspeaker.

$$\omega = \rho_0 x_0 + \sum_{k=1}^{N} \rho_k x_k, \text{ where } N = 5 \tag{2}$$

$$\frac{d\rho_k(t)}{dt} = \mu x_k(t) \frac{dx_0(t)}{dt}, \text{ where } k = 1, \ldots, N \tag{3}$$

We evaluate the motion extrapolation performance for three separate target angular velocities – 5°/ 10 time steps, 10°/ 10 time steps and 15°/ 10 time steps. For all trials, we set the learning rate μ to 0.0001 and the synaptic weight ρ_0 to 0.00001. We initialise all plastic synaptic weights ρ_k to zero. In our earlier work a stopping criterion halted the learning when the robotic agent oriented to within 0.5° of the currently playing loudspeaker [20]. Here we remove this constraint and allow the learning to progress until the synaptic weights stabilise, but the maximum number of iterations is limited to 150 to reduce the simulation run time. After the learning stops, the learned synaptic weights are frozen and serve to compute the angular velocity required to extrapolate and align with the future spatial location of the acoustic signal.

4. Results and discussion

The learning can be divided into two phases – a *reactive* phase where the robotic agent learns the target's angular velocity and is able to orient towards the currently playing loudspeaker within one time step and a *predictive* phase where the robotic agent exceeds the target's angular velocity and learns the correct angular velocity that allows it to orient itself towards the next loudspeaker along the trajectory within one time step. Figure 3**A** shows the tracking error θ_e during the reactive phase for a target angular velocity of 15° / 10 time steps as an example. The spikes in θ_e visible in the inset represent a mismatch between the last orientation of the robotic agent and the current spatial location of the acoustic signal. This produces finite ITD information that is used by the lizard peripheral auditory model to extract directional information about the acoustic signal. The robotic agent then rotates towards the acoustic signal with the last learned angular velocity, thereby decreasing θ_e. At each time step, this process is repeated, thereby reducing the tracking error exponentially. In the predictive phase (Fig. 3**B**) the robotic agent's learned angular velocity exceeds that of the target and it overshoots the target in one time step by a progressively larger amount at every iteration. The overshoot at any given time step is compensated for by a small amount in the next time step due to sign reversal of $x_k(t)$ in Eqn. 1, decreasing the synaptic weights and causing a progressively greater overshoot in the opposite direction. This leads to exponentially growing oscillations of the robotic agent around the currently playing loudspeaker. The oscillations stabilise when the synaptic weight updates in either direction are matched (Fig. 3**C**). Table 1 lists the predicted angular displacements and the corresponding prediction errors $\Delta\theta_e$, which are relatively small in all trials.

Table 1. Prediction tracking performance for the three target angular velocities.

Target angular velocity (angular displacement / 10 time steps)	Predicted angular displacement	Prediction error $\Delta\theta_e$
5° / 10 time steps	10.02°	0.02°
10° / 10 time steps	20.75°	0.75°
15° / 10 time steps	31.75°	1.75°

5. Conclusions

We have reported on a neural closed-loop learning mechanism for acoustic motion extrapolation. It allows a robotic agent in simulation to predict the future spatial location of a virtual acoustic signal in motion along a fixed semi-circular trajectory with a constant but unknown angular velocity. The mechanism successfully computes the agent's angular velocity that aligns the agent towards the future spatial location of the acoustic signal by correlating

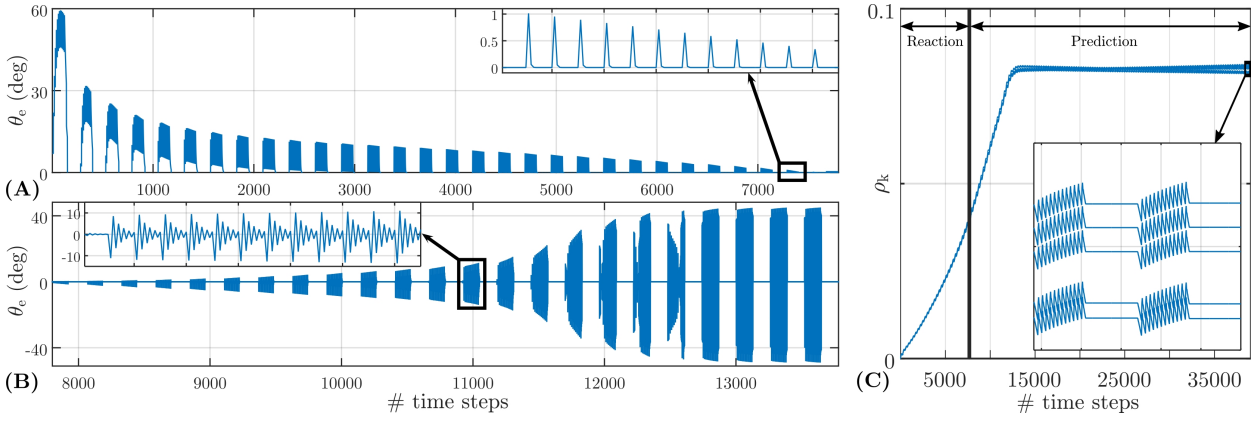

Fig. 3. Predictive tracking performance for a target angular velocity of 15° / 10 time steps. The insets show snapshots for an iteration. **A.** Tracking error θ_e for the reactive phase where the robotic agent matches the target's angular velocity. **B.** Tracking error θ_e for the predictive phase where the synaptic weights stabilise and the robotic agent learns the correct angular velocity that predicts the target's future position. **C.** Evolution of the synaptic weights ρ_k. The solid vertical line marks the instant at which the angular velocity ω of the robotic agent matches the target's angular velocity. The inset shows the evolution of the synaptic weights ρ_k for the last two iterations after they have stabilised.

the sound direction, perceived by a lizard peripheral auditory model, preceding and succeeding the agent's spatial rotations. The next step is to validate these results in an identical experimental setup in the real world by realising the neural learning mechanism on a mobile robot. The proposed approach can also be applied to spatial tracking of speech by tuning the lizard peripheral auditory model parameters to respond to human speech. The neural machinery itself is agnostic to the sensory modality used and can also be applied to visual motion extrapolation.

References

1. Nakadai, K., Lourens, T., Okuno, H., Kitano, H.. Active audition for humanoid. In: *In Proceedings of 17th National Conference on Artificial Intelligence (AAAI-2000) (2000), AAAI*. AAAI; 2000, p. 832–839.
2. Okuno, H., Nakadai, K., Hidai, K.I., Mizoguchi, H., Kitano, H.. Humanrobot non-verbal interaction empowered by real-time auditory and visual multiple-talker tracking. *Advanced Robotics* 2003;**17**(2):115–130.
3. Wang, H., Chu, P.. Voice Source Localization for Automatic Camera Pointing System in Videoconferencing. In: *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97) -Volume 1 - Volume 1*; ICASSP '97. Washington, DC, USA: IEEE Computer Society; 1997, p. 187–190.
4. Brandstein, M., Ward, D., editors. *Microphone Arrays: Signal Processing Techniques and Applications*. Digital Signal Processing. Springer Berlin Heidelberg; 1 ed.; 2001.
5. Reeve, R., Webb, B.. New neural circuits for robot phonotaxis. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 2003;**361**(1811):2245–2266.
6. Oh, Y., Yoon, J., Park, J., Kim, M., Kim, H.. A name recognition based call-and-come service for home robots. *IEEE Transactions on Consumer Electronics* 2008;**54**(2):247–253.
7. Ju, T., Shao, H., Peng, Q.. Tracking the moving sound target based on distributed microphone pairs. In: *Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2013 10th International Computer Conference on*. 2013, p. 330–334.
8. Ju, T., Shao, H., Peng, Q., Zhang, M.. Tracking the moving sound target based on double arrays. In: *Computational Problem-Solving (ICCP), 2012 International Conference on*. 2012, p. 315–319.
9. Kwak, K.. Sound source tracking of moving speaker using multi-channel microphones in robot environments. In: *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*. 2011, p. 3017–3020.
10. Ward, D., Lehmann, E., Williamson, R.. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. 2003.
11. Lehmann, E.. *Particle filtering methods for acoustic source localisation and tracking*. Ph.D. thesis; The Australian National University; 2004.
12. Cai, W., Wang, S., Wu, Z.. Accelerated steered response power method for sound source localization using orthogonal linear array. *Applied Acoustics* 2010;**71**(2):134–139.
13. Wan, X., Wu, Z.. Improved steered response power method for sound source localization based on principal eigenvector. *Applied Acoustics* 2010;**71**(12):1126–1131.
14. Marti, A., Cobos, M., Lopez, J., Escolano, J.. A steered response power iterative method for high-accuracy acoustic source localization. *The Journal of the Acoustical Society of America* 2013;**134**(4):2627–2630.
15. Zhao, X., Tang, J., Zhou, L., Wu, Z.. Accelerated steered response power method for sound source localization via clustering search. *Science China Physics, Mechanics and Astronomy* 2013;**56**(7):1329–1338.
16. Lima, M., Martins, W., Nunes, L., Biscainho, L., Ferreira, T., Costa, M., et al. A volumetric srp with refinement step for sound source localization. *IEEE Signal Processing Letters* 2015;**22**(8):1098–1102.
17. Knapp, C., Carter, G.. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1976;**24**(4):320–327.
18. DiBiase, J.. *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*. Ph.D. thesis; Brown University; 2000.
19. DiBiase, J., Silverman, H., Brandstein, M.. *Robust Localization in Reverberant Rooms*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2001, p. 157–180.
20. Shaikh, D., Manoonpong, P.. *An Adaptive Neural Mechanism with a Lizard Ear Model for Binaural Acoustic Tracking*. Springer International Publishing; 2016, p. 79–90.
21. Wever, E.. *The Reptile Ear: Its Structure and Function*. Princeton University Press; 1978.
22. Shaikh, D., Hallam, J., Christensen-Dalsgaard, J.. From "ear" to there: a review of biorobotic models of auditory processing in lizards. *Biological Cybernetics* 2016;doi:10.1007/s00422-016-0701-y.
23. Christensen-Dalsgaard, J., Manley, G.. Directionality of the Lizard Ear. *Journal of Experimental Biology* 2005;**208**(6):1209–1217.
24. Christensen-Dalsgaard, J., Tang, Y., Carr, C.. Binaural processing by the gecko auditory periphery. *Journal of Neurophysiology* 2011;**105**(5):1992–2004.
25. Fletcher, N., Thwaites, S.. Physical Models for the Analysis of Acoustical Systems in Biology. *Quarterly Reviews of Biophysics* 1979;**12**(1):25–65.
26. Fletcher, N.. *Acoustic Systems in Biology*. Oxford University Press, USA; 1992.
27. Zhang, L.. *Modelling Directional Hearing in Lizards*. Ph.D. thesis; Maersk Mc-Kinney Moller Institute, Faculty of Engineering, University of Southern Denmark; 2009.
28. Porr, B., Wörgötter, F.. Strongly improved stability and faster convergence of temporal sequence learning by utilising input correlations only. *Neural Computation* 2006;**18**(6):1380–1412.
29. Aertsen, A., Johannesma, P., Hermes, D.. Spectro-temporal receptive fields of auditory neurons in the grassfrog. *Biological Cybernetics* 1980;**38**(4):235–248.