Multisensory guidance of goal-oriented behaviour of legged robots

D. SHAIKH* and P. MANOONPONG†

Embodied AI and Neurorobotics Laboratory, Centre for BioRobotics, The Maersk Mc-Kinney Moller Institute, University of Southern Denmark
Campusvej 55, 5230 Odense M, Denmark
*E-mail: *danish@mmmi.sdu.dk, †poma@mmmi.sdu.dk*
http://ens-lab.sdu.dk

G. TUXWORTH

School of Information and Communication Technology, Griffith University
170 Kessels Road QLD 4111, Australia
E-mail: g.tuxworth@griffith.edu.au

L. BODENHAGEN

SDU Robotics, The Maersk Mc-Kinney Moller Institute, University of Southern Denmark
Campusvej 55, 5230 Odense M, Denmark
E-mail: lebo@mmmi.sdu.dk

Biological systems often combine cues from two different sensory modalities to execute goal-oriented sensorimotor tasks, which otherwise cannot be accurately executed with either sensory stream in isolation. When auditory cues alone are not sufficient to accurately localise an audio-visual target by orienting towards it, visual cues can complement their auditory counterparts and improve localisation accuracy. We present a multisensory goal-oriented locomotion control architecture that uses visual feedback to adaptively improve acoustomotor orientation response of the hexapod robot AMOS II. The robot is tasked with localising an audio-visual target by turning towards it. The architecture extracts sound direction information with a model of the peripheral auditory system of lizards to modulate locomotion control parameters driving the turning behaviour. The visual information adaptively changes the strength of the acoustomotor coupling to adjust turning speed of the robot. Our experiments demonstrate improved orientation towards the audio-visual target emitting a tone of frequency 2.2 kHz located at an angular offset of 45° from the robot.

Keywords: lizard peripheral auditory system, audio-visual localisation, multisensory integration, hexapedal locomotion

2

1. Introduction

Vision and audition are the two most dominant senses in humans as well as higher-order animals for directing attention towards relevant objects and events in the real world. Vision outperforms audition in spatial perception. However, research has shown that these two senses do not operate in isolation, but rather share neural resources[1]. Auditory attention may orient vision towards an audio-visual target that is initially not in the field of view or is not aligned with the spatial orientation of visual attention. Audition is thought to provide a rough initial estimate of target location, orienting gaze towards it, after which vision guides and provides accurate localisation. This may allow for better detection of unexpected visual targets. For example, we are able to localise the source that calls out our name from a location which is either outside our field of view or at a location other than where our eyes are currently focussed. In this way multisensory spatial cues guide orientation behaviour[2]. Audio-visual multisensory integration has been studied from the perspective of Bayesian inference[3]. To the best of our knowledge, closed-loop multisensory guidance of legged robots has not been fully investigated. Research in legged robots focusses mainly on the generation of animal-like gaits[4,5]. The contribution of this work is therefore to develop mechanisms for multisensory guidance of legged locomotion. Potential applications for legged robots are as search-and-rescue robots in natural disaster scenarios such as searching for earthquake survivors amongst rubble and large debris.

Here we present closed-loop multisensory goal-oriented locomotion control that uses visual feedback to reduce auditory orientation errors. The mechanism uses a model of the lizard peripheral auditory system which provides sound direction information for auditory localisation. The peripheral auditory model has been thoroughly studied via several robotic implementations[6]. Visual feedback is provided by a standard digital video camera module. The mechanism is implemented on the hexapod robot AMOS II[7], which must execute a turning behaviour to orient itself towards an out-of-view audio-visual target located at a relative angular offset of 45°. We compare the preliminary performance of the mechanism to two control conditions—(a) auditory orientation and (b) auditory orientation to bring the target into the field of view, followed by visual orientation.

This article is organised in the following manner. Section 2 describes the multisensory goal-oriented locomotion control mechanism and the experimental setup. Section 3 presents the experimental results from robotic trials. Section 4 summarises the work and discusses future directions.

2. Materials and methods

We use the hexapod robot AMOS II (inset in Fig. 1) that is inspired by cockroach morphology. The walking behaviours of AMOS II are driven by neural control[7] (Fig. 1). The robot must orient towards a loudspeaker emitting a 2.2 kHz tone and with a 5 mm green LED (Light Emitting Diode) mounted atop its housing. It is located $\sim$2 m away at an angular offset of 45° to the left of the robot's initial orientation. The target is initially out of view so the robot must always initiate orientation via audition.

Sound direction is extracted by an electrical model[8] (Fig. 2**A**, bottom) of the lizard peripheral auditory system (Fig. 2**A**, top). This model is chosen due to its computational simplicity. Voltages V_I and V_C respectively model sound pressures at the two ears. Currents i_I and i_C model eardrum vibrations. Impedances Z_r model the total acoustic filtering due to the ET and eardrum. Impedance Z_v models the acoustic filtering by the mouth cavity. The equivalent sound pressure in this cavity, modelled by voltage V_cc, due to the interaction of internal sound pressures from either side induces current i_cc. i_cc models the propagation of sound waves inside the mouth cavity. Sound direction is modelled as the binaural subtraction (Fig. 2**B**) of the perceived amplitude of eardrum vibrations, given by
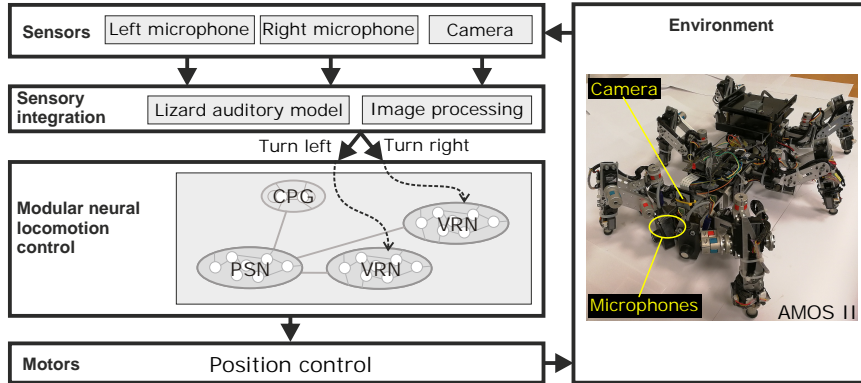

Fig. 1.   The multisensory goal-oriented locomotion control architecture for AMOS II. The locomotion controller comprises a central pattern generator (CPG), a Phase Switching Network (PSN) and two Velocity Regulation Networks (VRNs). Each VRN controls the three ipsilateral Thoraco-Coxal (TC) joints that enable forward, backward and turning movements. The TC joints are located closest to the robot's body. The VRNs control the walking speed and turning direction of the robot based on their inputs, and scaling the inputs changes the walking speed. Auditory and visual signals are used to control the turning behaviour of the robot.
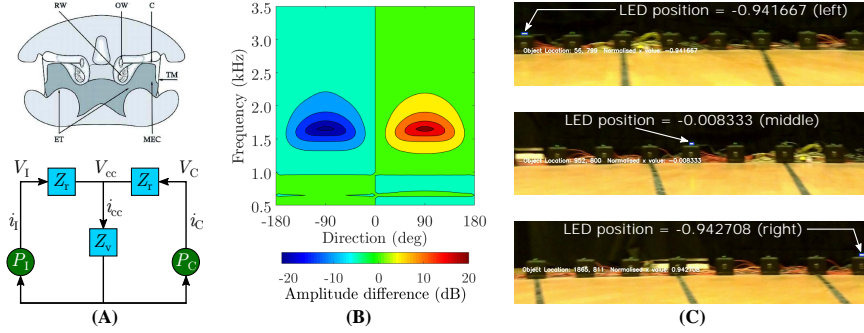
4

Fig. 2.   Auditory and visual information driving orientation behaviour. **A** Cross-sectional diagram[9] (top) of the lizard (genus *Sceloporus*) peripheral auditory system and ideal lumped-parameter electrical model[8] (bottom). The system possesses two ears (TM) connected via air-filled Eustachian tubes (ET) passing through the mouth cavity. This design permits sound transmission between the ears. This translates small phase differences (proportional to time differences) between sound waves arriving at either ear into relatively larger differences in perceived sound amplitude at either ear. The amplitude of the phase difference and thus the perceived amplitude difference depends on the relative sound direction. **B** Contour plot modelling binaural subtraction of the ipsilateral and contralateral responses as defined by Eq. (1). **C** Example snapshots of the robot's camera image showing the detection of a green LED mounted atop the loudspeaker.

$$\left|\frac{i_{\mathrm{I}}}{i_{\mathrm{C}}}\right| = \left|\frac{G_{\mathrm{I}} \cdot V_{\mathrm{I}} + G_{\mathrm{C}} \cdot V_{\mathrm{C}}}{G_{\mathrm{C}} \cdot V_{\mathrm{I}} + G_{\mathrm{I}} \cdot V_{\mathrm{C}}}\right| \equiv 20 \left(\log|i_{\mathrm{I}}| - \log|i_{\mathrm{C}}|\right) \ \mathrm{dB} \ , \qquad (1)$$

where G_{I} and G_{C} are ipsilateral (towards the acoustic signal) and contralateral (away from the acoustic signal) frequency-specific (1 kHz-2.2 kHz) gains. These are determined experimentally via laser vibrometry[9] measurements of eardrum vibrations and are implemented digitally as 4th-order infinite impulse response bandpass filters. The symmetry of the model implies that its response $\left|\frac{i_{\mathrm{I}}}{i_{\mathrm{C}}}\right|$ is zero for a sound signal arriving directly from the front or back in the azimuth plane ($|i_{\mathrm{I}}| = |i_{\mathrm{C}}|$), positive for $|i_{\mathrm{I}}| > |i_{\mathrm{C}}|$ and negative for $|i_{\mathrm{C}}| > |i_{\mathrm{I}}|$. Thus a positive sign means the sound signal is on the ipsilateral side while a negative sign means it is on the contralateral side. Furthermore, the magnitude of the response varies symmetrically but non-linearly within the relevant range of $[-90°, +90°]$ of sound direction.

Visual information is obtained by detecting the green LED via the OpenCV software image processing library. The detection algorithm localises a group of pixels with luminance greater than a fixed threshold in the green channel of a captured RGB image. It returns the relative position of the LED in the image, normalised within the range $[-1, +1]$, where

negative and positive values imply that the LED is respectively detected to the left and right of the median in the image (Fig. 2). Thus negative values correspond to the LED located on the left side of the robot and positive values correspond to the LED located on the right side of the robot.

The lizard peripheral auditory model, the image processing routines and the neural control are all implemented on a Raspberry Pi 2 (Model B+) controller connected to a FPGA board (model LOGI Pi from ValentFX). An analogue-to-digital converter (ADC) driver is implemented on the FPGA using the VHDL (VHSIC Hardware Description Language) programming language (VHSIC stands for Very-High-Speed Integrated Circuits). The driver reads raw digitised audio signals from a 2-channel simultaneous ADC connected to two omnidirectional microphones (model FG-23329-P07 from Knowles Electronics) mounted 13 mm apart at the front of the robot. The microphone separation matches the eardrum separation of the lizard from which the peripheral auditory model's parameters have been derived. An otherwise mismatched separation creates a mismatch between the extracted ITD cues and the ITD cues to which the peripheral auditory model is tuned. The controller is powered by a 12 Ah lithium polymer battery (model Xtorm AL450 from A-solar). The robot's servo motors are powered by a tethered external power supply. A 5-megapixel RGB digital video camera with a measured effective horizontal field of view of $\pm 19°$ (Raspberry Pi camera module), is mounted between the microphones on the robot.

The sensorimotor couplings are defined such that the robot must turn either left or right, with a given angular velocity, based on whether the sensory signal is positive or negative. For auditory orientation, the robot turns left when $\left|\frac{i_\mathrm{I}}{i_\mathrm{C}}\right|$ is positive and right when it is negative. For visual orientation, the robot turns to the side on which the LED is detected in the camera image as explained earlier. In both cases the robot's angular velocity is fixed at $\sim 2.426°$/sec. Such turning manoeuvres aim to equalise the outputs i_I and i_C of the peripheral auditory model during auditory orientation and nullify the detected relative location of the LED during visual orientation. The binaural subtraction, i.e. the right-hand term in Eq. (1), results in a zero value when the two microphones are equidistant from the target. This occurs when the robot is directly facing the target. In this condition there is no phase difference between the sound signals arriving at the two microphones and thus the peripheral auditory model's outputs are identical. Any displacement from this position results in a non-zero response from the peripheral auditory model and the robot takes corrective action by turning in the appropriate direction. Similarly, when the robot is pointing directly

6

towards the target, the LED is detected to be horizontally centred in the camera image, resulting in the relative position of the LED being reported as zero. Any displacement of the robot from this position results in a shift in the detected location of the LED in the camera image. The robot consequently takes corrective action by turning in the appropriate direction to centre the target in the field of view.

Inter-limb coordination during turning in hexapod robots is typically coarse, implying that fine turning movements are difficult to achieve in real world conditions. Therefore the robot may overshoot the target as it attempts to centre itself on it. For unimodal sensory signals, i.e. using only audition or only vision, the defined sensorimotor couplings can cause the robot to consequently oscillate about the relative location of the target if its turning velocity is kept constant. However adaptive scaling, using multimodal sensory information, of the drive signal that determines the turning velocity may reduce the amplitude of these oscillations. We test this hypothesis by determining the robot's orientation performance in terms of the amplitude of the oscillations and their relative offset from the location of the target. The orientation performance is compared for three separate trials. Firstly, the robot must orient towards the target using audition alone. Secondly, the robot must orient towards the target using audition until it is in the field of view, and then switch to vision to continue orientation. Finally, the robot must orient towards the target using audition but its angular velocity is modulated by visual feedback when the target is in its field of view. This is done by multiplying the input to the VRNs (Fig. 1), which lies within the range $[0, 1]$, by the absolute value of the relative position of the LED in the camera image.

3. Results and discussion

A single trial was conducted for each orientation strategy. Figure 3 shows the orientation errors during localisation. All orientation strategies exhibit similar initial movement towards the audio-visual target as evident by an almost linear decrease in orientation errors, followed by oscillatory movements. Auditory orientation (Fig. 3**A**) generates relatively large oscillations as compared to auditory-then-visual orientation (Fig. 3**B**) and vision-modulated auditory orientation (Fig. 3**C**). This is because the peripheral auditory model's outputs exhibit errors, due to acoustic noise from sound reflections in the environment as well as from the robot's movements, introduced into the sound signals recorded by the microphone. These oscillations are relatively symmetrical about the target. Auditory-then-visual orienta-
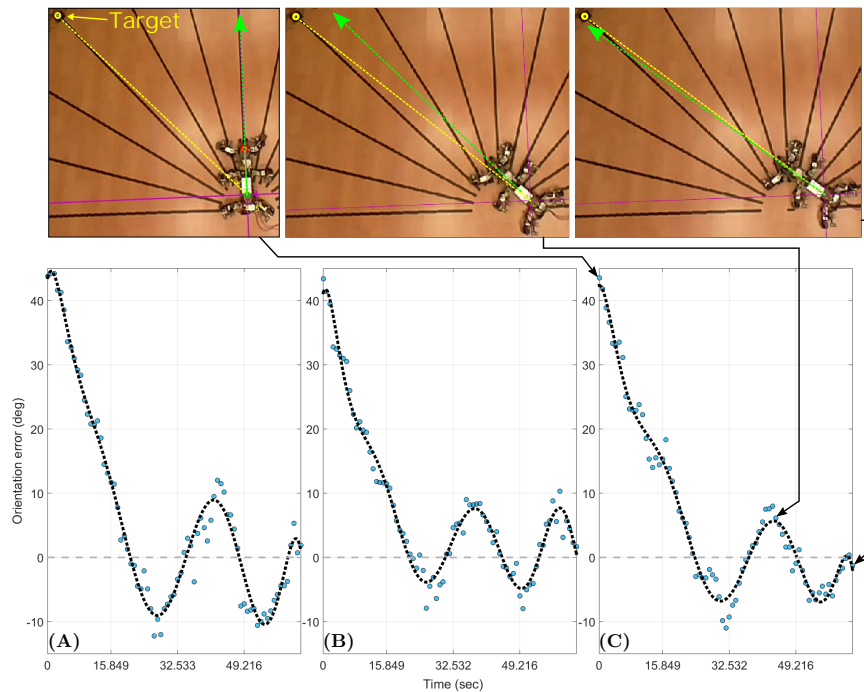
Fig. 3.   Orientation errors during localisation (circles) and fitted 9-degree polynomial curves (dashed lines). **A** Auditory orientation. **B** Auditory orientation followed by visual orientation. **C** Visual scaling of the acoustomotor coupling. Snapshots show robot orientation (green dashed line) relative to the target (top left corner of image).

tion generates relatively smaller oscillations that are offset from the target. The smaller amplitude is a result of relatively low error in the detection of the LED due to relatively noise-free images recorded by the camera. Visual modulation of auditory orientation also generates relatively smaller oscillations that are symmetrical about the target as well. This is a consequence of the adaptive visual scaling of the motor drive signals that determine angular velocity in the acoustomotor coupling. Thus, a multi-modal integration strategy produces relatively symmetrical oscillations with relatively lower amplitude. These observations are quantified in Table 1.

4. Conclusions and future directions

We presented a mechanism for multisensory guidance of orientation behaviour in a hexapod robot. The mechanism uses visual feedback to suppress oscillations during auditory localisation of an audio-visual target. We

8

Table 1.   Mean and standard deviation of the oscillations in the orientation error.

Orientation strategy	Mean (deg)	Standard deviation (deg)
Auditory orientation	-0.1731	± 6.3862
Auditory-then-visual orientation	1.8903	± 4.0663
Visual modulation of auditory orientation	-0.5951	± 4.4313

compared the orientation performance to that of two control conditions—auditory-only orientation and auditory-then-visual orientation. Results in robotic trials confirm that visual feedback generates relatively small oscillations centred on the target. However, we aim to perform multiple trials to draw stronger conclusions.

In our experiments the scaling factor is the visual feedback signal itself, which is explicitly designed. In order to minimise the oscillations, a gain must be introduced to scale the visual feedback signal and decelerate the robot during multisensory orientation behaviour. The optimal value of this gain could be learned by correlating the auditory and visual sensory information in time such that optimal deceleration can occur. In this way the robot can learn to optimally orient towards an audio-visual target.

5. Acknowledgements

References

1. D. Smith, B. Davis, K. Niu, E. Healy, L. Bonilha, J. Fridriksson, P. Morgan and C. Rorden, *Journal of Cognitive Neuroscience* **22**, 347 (2010).
2. J. Maier and J. Groh, *Hearing Research* **258**, 106 (2009).
3. B. David and A. David, Chapter 14 combining visual and auditory information, in *Visual Perception—Fundamentals of Awareness: Multi-Sensory Integration and High-Order Perception*, eds. S. Martinez-Conde, S. Macknik, L. Martinez, J.-M. Alonso and P. Tse, Progress in Brain Research, Vol. 155, Part B (Elsevier, 2006) pp. 243–258.
4. A. Schneider, J. Paskarbeit, M. Schilling and J. Schmitz, HECTOR, a bio-inspired and compliant hexapod robot, in *Proceedings of the 3rd Conference on Biomimetics and Biohybrid Systems*, 2014.
5. K. Karakasiliotis, R. Thandiackal, K. Melo, T. Horvat, N. K. Mahabadi, S. Tsitkov, J. M. Cabelguen and A. J. Ijspeert, *Journal of The Royal Society Interface* **13** (2016).
6. D. Shaikh, J. Hallam and J. Christensen-Dalsgaard, *Biological Cybernetics* **110**, 303 (2016).

7. P. Manoonpong, U. Parlitz and F. Wörgötter, *Frontiers in Neural Circuits* **7**, p. 12 (2013).
8. N. Fletcher and S. Thwaites, *Quarterly Reviews of Biophysics* **12**, 25 (1979).
9. J. Christensen-Dalsgaard and G. Manley, *Journal of Experimental Biology* **208**, 1209 (2005).