

# **Exploring a Robotic Model of the Lizard Peripheral Auditory System**

Danish Shaikh



A dissertation submitted to the University of Southern Denmark  
in partial fulfillment of the requirements of the degree of Doctor of Philosophy

January 8, 2012



*“To my wonderful daughter Olivia...”*



# Abstract

It is well known that lizard species such as the *Gekko gecko* and *Mabuya macularia* have evolved an amazingly directional peripheral auditory system that exhibits the strongest directionality of any terrestrial vertebrate's auditory system that has been investigated. The fascinating feature of the system is that it has rather small dimensions in comparison with the wavelengths of the frequencies to which it most sensitive. This implies that the binaural time and intensity cues, generated by the time-of-arrival differences and the sound shadow created by the animal's head, respectively, are rather small. In spite of this fact, the strong directionality has been experimentally determined to be a consequence of the acoustical coupling of the contralateral and ipsilateral tympanal membranes via a hollow Eustacian tube through the mouth cavity. Studies using laser vibrometry have shown that the strong directionality is manifested in the form of tympanal vibrations which are stronger on the side which is nearer to the source of the sound.

The work presented here builds upon recent research on modelling the peripheral auditory system of lizards and on validation of the obtained model coupled with a decision model in phonotaxis tasks through simulation and robotic experiments. Initially, the phonotactic performance of the auditory model is re-investigated through a robotic implementation inspired by Braitenberg vehicles with simple and direct sensorimotor cross-couplings between the auditory model and the actuators. In subsequent experiments, isolated and independent decision models are embedded into the individual sensorimotor cross-couplings, and the phonotactic performance is again investigated. The effect of parameter-scaling on the response of the model is then investigated and the directionality of the scaled models is experimentally determined via a robotic implementation.

In those experiments the robot model has the ability to move freely. However, sound localization based on the implemented strategies fails when movement is not allowed because the auditory cues cannot be directly mapped to the sound direction by such strategies. For a stationary system to localize sound using the auditory model, an internal representation of the sound direction is necessary. In order to automatically generate such a representation, the model is then coupled with a Cerebral Model Articulation Controller based neural network and the system's ability to learn the sound direction representation is investigated.

In all the above-mentioned experiments, the robot model is a wheeled robot. In order to extend the investigation to sensory guidance of bio-inspired locomotion patterns, the model is then coupled to a Central Pattern Generator locomotion model based on the spinal cord of salamanders, and the auditory cues extracted by the auditory model are employed to modulate the parameters of the locomotion model to generate turning motion in phonotaxis tasks. Finally, three dimensional sound localization is investigated via a pan-tilt assembly with three microphones employing two independent auditory models performing sound localization in the horizontal and vertical planes, and preliminary results are presented.

## **Declaration**

I hereby declare that the material presented in this dissertation is original and is the result of my own work, except where explicit reference is made to the work of others.

Danish Shaikh



# Acknowledgements

First and foremost, I would like to deeply thank my doctoral supervisors Prof. Dr. John Hallam from the Maersk Mc-Kinney Moller Institute and Prof. Dr. Jakob Christensen-Dalsgaard from the Institute of Biology, whose inexhaustible pools of knowledge and deep scientific insight into the fields of robotics and biology as well as numerous other relevant disciplines provided a enlightening environment for me during the past three years. Without their constant support and consistent supervision, this dissertation and the work presented herein would not be possible. I would also like to express my deep gratitude towards a number of people at the École Polytechnique Fédérale de Lausanne who were instrumental during my research visit, namely, Prof. Dr. Auke Jan Ijspeert, Dr. Alessandro Crespi, Jérémie Knüsel and Stéphane Bonardi from the Biorobotics Laboratory for their substantial help with my research efforts with the *Salamandra robotica II* and for making my stay enjoyable from both a research and social perspective, and towards Dr. Hervé Lissek, Cédric Monchâtre and Lionel Velut from the Laboratory of Electromagnetics and Acoustics for allowing me to avail of their anachocic chamber facilities for experimental purposes during my stay.

A special thanks goes to my previous colleagues and current friends, namely, Dr. Mirko Bordignon, Dr. Franco Ricardo Mendoza Garcia, Dr. Andreas Lyder, Dr. David Johan Christensen and Dr. David Brandt have all played an important role during my studies, distracting me just enough from my work to enjoy their company in social and informal settings. Furthermore, I would also like to thank Dr. Lei Zhang for his help in the initial stages of my work.

I would like to express my immense gratitude to my parents and my brother, who unconditionally accepted my at times almost eremitic lifestyle during my studies and never once let their incredible love and support for me diminish.

Last, but certainly not the least, there are two people who have been affected the most by my work schedule, and they are my wonderful wife Agnieszka and daughter Olivia. Words cannot express how indebted I am to them for their patience and thoughtful-

ness, and the unconditional love and care they have displayed during these testing times. Without their support, I would not be where I am today.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	The Peripheral Auditory System of a Lizard . . . . .	3
1.2	Key Contributions . . . . .	5
1.3	Thesis Organization . . . . .	6
<b>2</b>	<b>Literature Review</b>	<b>9</b>
2.1	Introduction . . . . .	9
2.2	Directional Hearing in Nature . . . . .	11
2.2.1	Barn Owl . . . . .	12
2.2.2	Crickets . . . . .	14
2.2.3	<i>Ormia ochracea</i> . . . . .	16
2.3	Synthetic Approaches to Sound Localization . . . . .	19
2.3.1	Traditional Approaches . . . . .	19
2.4	Biomimetic Robot Models . . . . .	22
<b>3</b>	<b>Braitenberg Models</b>	<b>29</b>
3.1	Introduction . . . . .	29
3.2	Theoretical Model of the Auditory System . . . . .	30
3.2.1	Auditory Model Implementation . . . . .	32
3.3	Braitenberg Vehicles . . . . .	33
3.4	A Basic Braitenberg Lizard . . . . .	35
3.4.1	The Robotic Platform . . . . .	35
3.4.2	Experimental Setup . . . . .	37
3.4.3	Results and Discussion . . . . .	40
3.5	A Braitenberg Lizard with Decision Models . . . . .	46
3.5.1	The Robotic Platform . . . . .	46
3.5.2	Experimental Setup . . . . .	47
3.5.3	Results and Discussion . . . . .	49
3.6	Summary and Conclusions . . . . .	55

<b>4</b>	<b>Modifying Directionality Through Auditory System Scaling</b>	<b>57</b>
4.1	Introduction . . . . .	57
4.2	Auditory System Scaling . . . . .	58
4.3	Design and Implementation . . . . .	60
4.3.1	Scaled Auditory Models . . . . .	60
4.3.2	Robotic Implementation . . . . .	60
4.4	Experimental Setup . . . . .	62
4.5	Results and Discussion . . . . .	64
4.5.1	Trajectory Directness . . . . .	66
4.6	Summary and Conclusions . . . . .	67
<b>5</b>	<b>Learning to Represent Sound Direction</b>	<b>71</b>
5.1	Introduction . . . . .	71
5.1.1	Cerebellar Model Articulation Controller . . . . .	72
5.2	Design and Implementation . . . . .	74
5.2.1	The Auditory Model Implementation . . . . .	74
5.2.2	The Learning Architecture . . . . .	76
5.2.3	The Robotic Platform . . . . .	80
5.3	Experimental Methods . . . . .	82
5.3.1	Simulation . . . . .	82
5.3.2	Robotic Trials . . . . .	90
5.4	Summary and Conclusions . . . . .	98
<b>6</b>	<b>Sensory Modulation of Motion</b>	<b>101</b>
6.1	Introduction . . . . .	101
6.2	<i>Salamandra Robotica II</i> . . . . .	102
6.2.1	Shortcomings in the Robot Model . . . . .	105
6.3	Design and Implementation . . . . .	107
6.4	Experiments and Results . . . . .	111
6.4.1	Phonotactic Performance in Simulation . . . . .	111
6.4.2	Robotic Experiments . . . . .	119
6.5	Summary and Conclusions . . . . .	126
<b>7</b>	<b>Three-dimensional Sound Source Localization</b>	<b>129</b>
7.1	Introduction . . . . .	129
7.2	Three-dimensional Sound Localization . . . . .	130
7.3	Design and Implementation . . . . .	131

7.4	Experimental Setup . . . . .	134
7.5	Results and Discussion . . . . .	135
7.6	Summary and Conclusions . . . . .	138
<b>8</b>	<b>Conclusions</b>	<b>141</b>
8.1	Summary and Contributions . . . . .	141
8.2	Future Directions . . . . .	144
<b>Bibliography</b>		<b>149</b>
<b>List of figures</b>		<b>163</b>
<b>List of tables</b>		<b>167</b>
<b>Appendix A</b>		<b>169</b>
<b>Appendix B</b>		<b>173</b>



*“Wer wartet mit Besonnenheit,  
der wird belohnt zur rechten Zeit...”*

— Till Lindemann



# Chapter 1

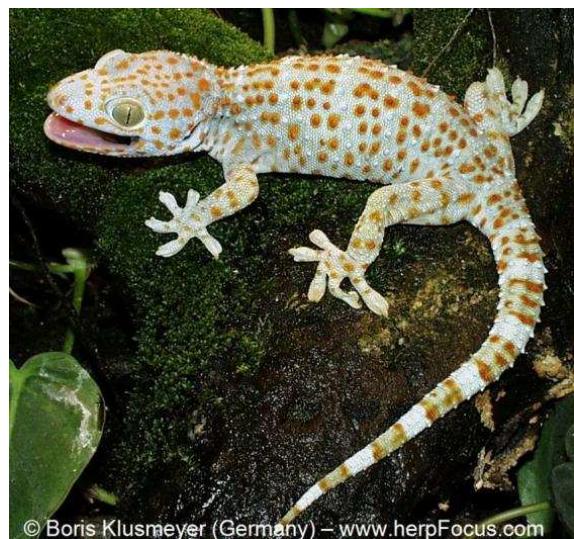
## Introduction

### 1.1 The Peripheral Auditory System of a Lizard

It has been known for quite some time that lizards such as the bronze grass skink, or *Mabuya macularia*, and the Tokay gecko, or *Gekko gecko* (refer to Fig. 1.1) possess a peripheral auditory system that is substantially directional [19, 135, 20]. The directionality is a consequence of an internal acoustical coupling of the two eardrums of the animal, formed by efficient transmission of sound through internal pathways in the head.



(a) The bronze grass skink or *Mabuya Macularia* (Image courtesy: Wikimedia Commons).



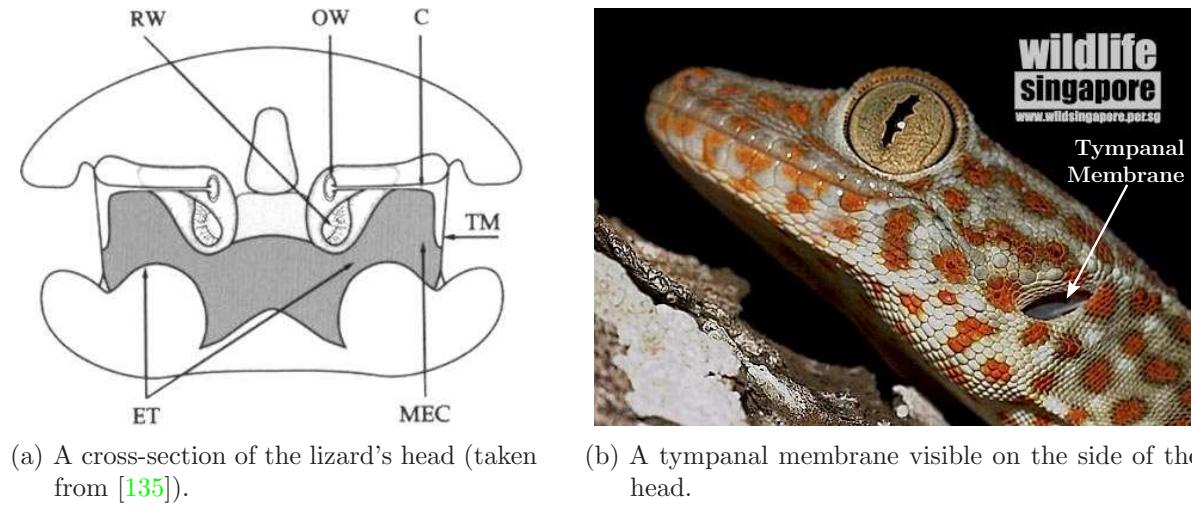
(b) The Tokay gecko (*Gekko gecko*)

© Boris Klusmeyer (Germany) – [www.herpFocus.com](http://www.herpFocus.com)

**Figure 1.1:** Two common lizard species that exhibit a highly directional peripheral auditory system.

The system exhibits symmetry with respect to the median plane and is relatively simple from a design point of view. What is interesting is that it is quite small in size with respect to the wavelength of the sounds that it responds to; most lizards have a tympanal separation (the distance between the eardrums) of approximately 10 mm–20 mm, and the sound wavelengths to which their ears are most sensitive are between 340 mm–85 mm, or 1000 Hz–4000 Hz respectively [20]. As a consequence, the incident sound waves tend to diffract around the animal’s head and body, resulting in the sound pressure being the same at both the ears. However, there is a small but finite phase difference between the sound waves arriving at either side, and its magnitude depends on the direction from which the sound appears to originate relative to the animal. The system essentially converts this small interaural phase difference cue into a relatively larger difference in the perceived amplitude of the sound on either side. The animal may employ this perceived difference in sound amplitudes to localize a relevant sound source, but this has not been verified through any behavioural studies.

As depicted in a cross-section of the lizard’s head in Fig. 1.2(a), the peripheral auditory system consists of the tympanal membranes (TM) or eardrums, the middle ear cavity (MEC), Eustachian tubes (ET), the cochlea (C), the round window (RW) and the oval window (OW) and the central mouth cavity in the middle. The tympani are located on either side of the head (refer to Fig. 1.2(b)), and are connected to each other via the Eustachian tubes through the central mouth cavity. Thus, for instance, when external sound waves impress upon the left ear, they cause the vibration of the left tympanum, and this consequently produces an internal sound pressure on inside of the left tympanum, in the left middle ear cavity. This sound pressure generates internal sound waves which travel via the Eustachian tubes, through the mouth cavity, to the right side and affect the sound pressure in the right middle ear cavity on the inside of the right tympanum. However, there are external sound waves impressing simultaneously upon the right tympanum as well and thus the motion of this tympanum is a result of the superposition of external and internal sound pressures. The resulting motion of the right tympanum is therefore due to the difference between instantaneous sound pressures on either side of it, which depends on the relative phases of the external sound pressure at the two ears. This process occurs simultaneously in the opposite direction as well, i.e. from right to left. These interactions create an acoustical coupling [22] between the two eardrums which transforms the lizard’s peripheral auditory system into what is referred to as a pressure difference receiver ear, which exhibits a directionality higher than that for any other known vertebrate [21]. Pressure difference receiver ears have been quite



**Figure 1.2:** The structure of the lizard ear.

widely studied and modelled both theoretically and experimentally. They are known to occur in lizards [21], frogs [19, 22], birds [67] and crickets [90, 86].

Much of the work presented in this dissertation follows that of Dr. Lei Zhang with a lumped-parameter model of the peripheral auditory system of as part of his doctoral research [137]. The engineering goal of this work is to develop a robotic prototype for sound localization, while the scientific goal is to explore a robotic model of the auditory system and its directionality. To this end, a series of independent experiments are performed, investigating the directionality of the auditory model in different scenarios. The outline of this dissertation is presented in the next section.

## 1.2 Key Contributions

The work presented in this dissertation can be characterised by the following achievements and contributions made in chronological order.

- A Lego mobile robot prototype inspired by Braitenberg vehicles was designed and constructed to investigate robotic phonotaxis.
- Simple decision models were independently embedded into the Braitenberg sensorimotor couplings of the mobile robot, and the phonotactic performance was evaluated.

- A second and more robust mobile robot was designed and constructed. The scaling properties of the lizard auditory model were investigated with the help of this robot.
- The lizard auditory model was coupled to a Cerebellar Model Articulation Controller and the ability of the system to learn a representation of the sound direction was investigated.
- The lizard auditory model was coupled to a Central Pattern Generator model of the salamander-like robot *Salamandra robotica II* and the sensory guidance of locomotion patterns was investigated.
- A pan-tilt sound localization system was constructed and the feasibility of three-dimensional sound localization with the lizard auditory model was investigated.

### 1.3 Thesis Organization

In Chap. 2, research related to directional hearing in animals, and their robotic implementations are discussed. Furthermore, a non-exhaustive survey of the typical as well as the biomimetic approaches to sound localization is presented.

In Chap. 3, the theoretical model of the peripheral auditory system of a lizard is presented, along with the digital implementation of the same. The frequency response of the theoretical model and the implementation model are presented as well. Next, Braitenberg vehicle-like robotic models, based on a combination of a digital signal processor and the Lego Mindstorms NXT brick as the robotic platform and implementing continuous control via direct sensorimotor cross-couplings, are introduced and their phonotactic performance is evaluated and compared to the step-control implementation using a decision model as in the robotic model of Zhang et al. [138]. In addition, two types of decision model are embedded into the sensorimotor couplings of the Braitenberg vehicle-like robotic model and their performances are compared for varying control loop frequencies.

Chapter 4 presents the scaling properties of the auditory model. Here, the hypothesis that consistent scaling of the auditory model and the ear separation results in a proportional scaling of the frequency response is formulated and subsequently verified via phonotaxis experiments. In addition, the effect of violating the scaling relationship between the auditory model and the ear separation is investigated. Finally, a new robotic model is introduced which utilizes programmable logic for parallel computation.

In Chap. 5, the auditory model is coupled with a Cerebral Model Articulation Con-

troller based neural network and the ability of the system to learn a representation of sound is investigated in simulation and in practice.

In Chap. 6, the auditory model is coupled with a bio-inspired Central Pattern Generator locomotor model of *Salamandra robotica II*, a salamander-like quadruped robot, essentially forming a Braitenberg vehicle-like quadruped robot. Sensorimotor couplings between the auditory model's outputs and the locomotor model's drive inputs are devised to elicit naturalistic locomotion patterns for phonotaxis behaviour.

Chapter 7 presents preliminary work towards three-dimensional sound source localization with the auditory model using three microphones. A pan-tilt electro-mechanical system is assembled, controlled by two identical auditory models, and the localization performance of the control strategy in three-dimensional space is evaluated.

Finally, the summary of the experimental work performed and the contributions thereof are presented in Chap. 8, along with potential future directions of research.



# Chapter 2

## Literature Review

### 2.1 Introduction

The sense of hearing is has evolved to be of utmost importance in humans and a number of animal species. The necessity of the hearing sense and the ability to locate a sound source is arguably more important to animals than to humans, since a number of species rely on it for hunting, avoid predators, and in general for survival as a whole. Fay and Popper [32] state that

*“...one of the fundamental and most important features of sound source detection is the ability of an animal to estimate source location as a first step in behaving appropriately in response to the sound. The need for sound source localization thus has become a fundamental feature of hearing in most hearing organisms, and one could argue that it is inconceivable that a sense of hearing could have evolved at all without the ability to locate and segregate sources. ”*

Sound source localization is the process of determining the relative location of a sound source in terms of its azimuth and elevation in space. There are three possible binaural (i.e., relating to two ears) cues which are available to an animal in order to localize sound — the Interaural Time Difference (ITD), the Interaural Level Difference (ILD) and the spectral cue. The ITD cue is the difference between the time-of-arrival of the incident sound waves at two spatially separate auditory receivers, and thus is also known as the Time-Difference-of-Arrival (TDOA). This cue provides information about the lateral position of the sound source, since it depends on the relative distances of the two receivers from the sound source. The ILD cue is the difference between the level of the sound pressure measured at two spatially separate auditory receivers, generated

due to the attenuation of the sound waves between the two receivers. This cue can provide information about the lateral position of the sound source as well, provided that the sound wavelength is less than the head size, a necessary condition for the sound to be blocked by the head (also referred to as sound shadowing). The spectral cue is generated by reflections from various body parts, for instance in the case of humans, from the pinnae, shoulders and upper torso.

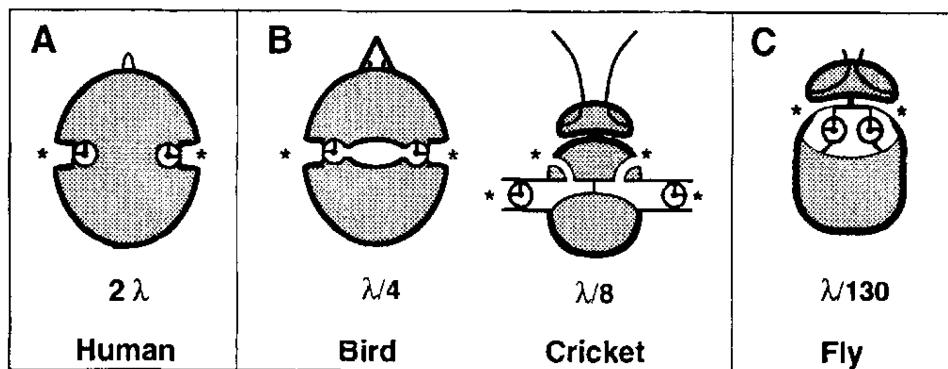
In 1907, Lord Rayleigh proposed the duplex theory, which stated that humans localize low frequency sounds (less than 2 kHz) through the ITD cues, while high frequency sounds (greater than 2 kHz) are localized through ILD cues. However, there is an overlap between the frequency ranges for ITD and ILD cues and therefore to localize natural sounds, for instance speech, which contains both low and high frequency components, the auditory system may use both cues to localize a sound source [111]. The duplex theory however does not provide an explanation for the ability of humans to distinguish between sounds coming the front and the back of the head. Furthermore, the theory does not explain localization in the vertical plane. For humans and animals with symmetrical ear placement, a combination of ILD and ITD cues can provide complete information about the azimuth [111], while the spectral cue can provide information about the elevation [39, 5].

Given the immense diversity in animal species these cues are tailored to individual species, depending upon the specific functional morphology of their acoustically relevant body parts such as the size and shape of the head and ears, as well as the anatomical position and orientation of the same, such as the separation between the ears and their orientation. A comparative survey of the correlation between sound localization ability and the anatomy of mammalian species has been compiled by Rickye S. Heffner [46]. She summarized that the hearing ranges of mammals is correlated with the functional head size i.e. the time taken by sound to travel between the two ears, which depends directly on the distance between the ears and indirectly on the size of the head. As the functional head size increases, the range of audible frequencies decreases, and vice versa. Naturally, the correlation between functional head size and physical head size is greater for species in which the outer ears or pinnae are located directly at the side of their heads, than for species in which the pinnae are located on the upper end of the sides of the head, such as the European Roe Deer (*Capreolus capreolus*), rabbits etc.. For the latter species, the ILD cues are minimal for elevated sounds since there is minimal sound shadowing effect. These animals must rely on ITD cues to localize sound in the vertical plane. However, nature has provided these creatures with movable pinnae, which can aid sound localization by generating ILD cues via movement. The tilting of

the head is also a common method used by a number of animals to localize sound coming from above or below. The red fox or *Vulpes vulpes*, are known to successively tilt their heads on either side while localizing terrestrial prey such as small rodents, which can be buried as deep as 6 feet below snow, in order to generate ITD and ILD cues for precise elevation determination. Isley and Gysel [61] determined the localization accuracy of red foxes under captivity to be greater than 90% for low frequencies, and 84% for high frequencies. The difference between mammals and smaller animals such as lizards and frogs is that the latter can extend their hearing to relatively higher frequencies, where sound shadowing is robust.

## 2.2 Directional Hearing in Nature

Directional hearing is a common trait visible among all hearing animals in nature. Depending on the selective pressure a particular species has been subjected to during its evolution, the auditory systems and sound localization abilities of different animals have evolved to different extents. There are three types of mechanisms for directional hearing that can be found in nature — pressure receivers, pressure difference receivers and mechanically coupled receivers. Figure 2.1 depicts these three types.



**Figure 2.1:** Different mechanisms of directional hearing observed in nature (taken from [104]). (A) Pressure receivers. (B) Pressure difference receivers. (C) Mechanically coupled receivers. The asterisks indicate the location of the sound inputs to the auditory system. The ratio between the interaural distance of the animal and the wavelength  $\lambda$  at 5 kHz is given as a fraction of  $\lambda$ .

In pressure receivers such as those in humans and most mammals, the auditory organs are disconnected from each other, acoustically as well as mechanically. Therefore, only the outer sides of the tympanal membranes experience the external sound pressure.

In birds, insects and reptiles, the auditory receivers are acoustically coupled together. This acoustical coupling allows sound to move from the ipsilateral tympanal membrane to the contralateral tympanal membrane, and this acoustical interaction results in the tympani experiencing sound pressures on both the internal and external sides, setting up a pressure gradient across either tympanum. Flies are unique in that their auditory periphery consists of a pair of tympani that are mechanically coupled to each other.

The discord between the head size and sound frequency is evident from the figure. For humans, the interaural distance, i.e. the distance between the ipsilateral and contralateral auditory organs is twice the sound wavelength of a 5 kHz tone (6.8 cm), taken as an example from [104]. For animals with pressure difference receivers, the interaural distance is one-fourth the wavelength for birds and one-eighth the wavelength for insects such as crickets. Among flies, the interaural distance is over a 100 times smaller than the sound wavelength. The smaller the interaural distance, the smaller are the ITD cues and thus greater is the challenge in localizing sound via time difference cues. This demonstrates that different species have evolved different mechanisms for sound localization. In the next sections a select few species are presented, which are known to have exceptional directional hearing.

### 2.2.1 Barn Owl

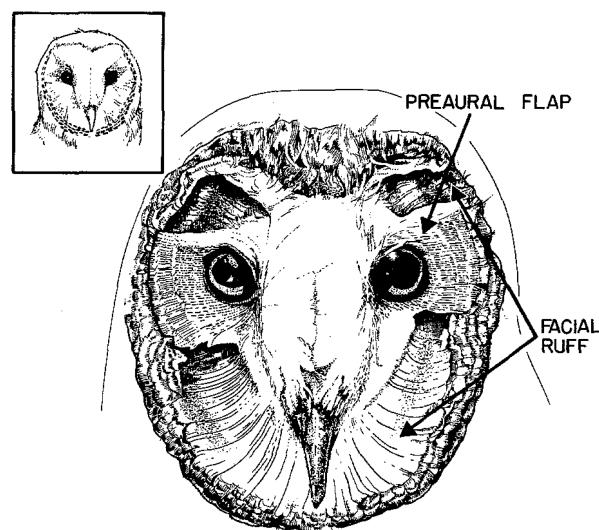
The barn owl or *Tyto alba* (refer to Fig. 2.2) is known for its strong directional hearing. It has been demonstrated by Roger S. Payne that the animal is sensitive to the sound, rather than the heat signature or the odour, generated by their prey [98]: understandably so, because the animal relies on its hearing to detect and localize its prey in complete darkness. In a typical prey capture scenario, the animal is perched on a tree branch and must localize a potential target such as a mouse on the ground. Since the prey is unpredictable and can be either stationary or in motion, the owl must estimate its azimuth and elevation as accurately as possible in order to be successful.

It has been reported by Knudsen et al. [69] that the animal can perform successful bi-coordinate sound localization without any head movements. Subsequent experiments with the barn owl by Knudsen and Konishi [70] demonstrated that the animal employed both the ITD and ILD cues to localize sounds in the azimuth plane, and only ILD cues to localize elevated sounds. The asymmetrical placement of the ear canal openings and the preaural flaps (refer to Fig. 2.3) which cover the openings are responsible for generating ILD cues needed for determining elevation [93, 124]. The ear canal openings

are surrounded by the facial ruff, consisting of densely packed sound reflecting feathers, and the left ear canal opening is located higher than the right one.



**Figure 2.2:** The barn owl or *Tyto alba* (Image Courtesy: [Wikimedia Commons](#)).



**Figure 2.3:** The barn owl ear placement (taken from [70]).

For an elevated sound source, at frequencies greater than 4 kHz the facial ruff reflects sound waves towards the ear canal opening, and due to the asymmetrical placement the left ear becomes more sensitive to sound waves impinging from below the head, while

the right ear becomes more sensitive to sound waves impressing from above the head, generating ILD cues which vary according to the relative elevation of the sound source. At frequencies lower than 4 kHz the corresponding wavelength ( $> 8.5$  cm) is greater than the dimensions of the facial ruff (6 cm–8 cm) and in this case the facial ruff has little effect, and the ILD cues corresponding to elevation are relatively small. It has been experimentally shown by Moiseff and Konishi [94] that the animal's interaural pathway attenuates the sound pressure significantly (by 13 dB on average for low frequencies and by 63 dB on average for high frequencies) and thus the ears do not function as pressure difference receivers. A Bayesian model of the sound localization by barn owls has been recently proposed by Brian Fischer [33].

## 2.2.2 Crickets

The cricket ear belongs to the pressure difference receiver family. The two most widely studied subfamilies of crickets are bush-crickets or *Tettigoniidae* and field crickets or *Gryllinae* (refer to Fig. 2.4). The males of the species emit chirping sounds or stridulation to attract females during the mating season. The female then has to localize the male cricket based on acoustical cues extracted from the calling songs via their directional hearing abilities.



(a) The bush-cricket *Tettigonia viridissima* (Image courtesy: [Wikimedia Commons](#)).



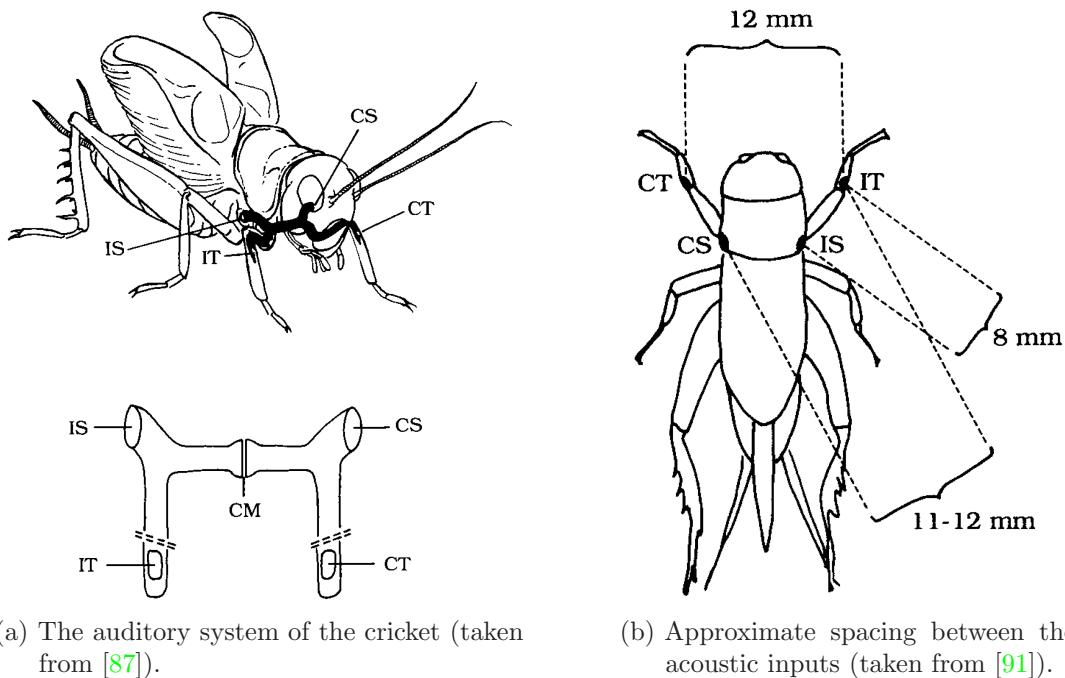
(b) The field cricket *Gryllus bimaculatus* (Image courtesy: [Shutterstock Images LLC](#)).

**Figure 2.4:** Two subfamilies of crickets.

The ears of crickets are located in the two front legs. Each tympanal membrane is connected to a spiracular opening on the ipsilateral side of the thorax through the acoustic trachea, which is basically an air-filled tube. These two acoustical tracheae are connected to each other via a transverse trachea, which contains a central membrane

(CM). Thus, there are four possible acoustic inputs into the auditory system of the cricket (refer to Fig. 2.5(a)) — the ipsilateral tympanum (IT), the contralateral tympanum (CT), the ipsilateral spiracle (IS) and the contralateral spiracle (CS). This implies that sound may propagate from the ipsilateral side to the contralateral side through the transverse trachea. Furthermore, due to the acoustical couplings, either tympanum can experience internal sound pressures similar to the lizard's tympani, apart from the external sound pressure.

The calling song frequency of the males is an approximately 4.8 kHz pure tone, with a corresponding wavelength of approximately 7 cm. At such wavelengths the diffraction effects are minimal due to the relatively small spacing between the acoustic inputs (refer to Fig. 2.5(b)), resulting in minimal differences in sound pressures between the ipsilateral and contralateral sides. This implies that the only cues available to crickets for sound localization are the ITD cues.



**Figure 2.5:** The cricket ear structure.

The directional hearing in crickets has been extensively studied. Huber et al. [54] have published a thorough review. It was originally proposed by Michelsen and Nocke [89] that the directionality was caused by ILD cues due to sound shadowing by the body. However, Hill and Boyan [48] proposed an alternate theory that the directionality was due to differences in the phases of the internal and external sound pressures on either tympanum due to internal sound transmission from the contralateral tympanum to the ipsilateral

tympanum through the transverse trachea. Hill and Boyan [49] performed behavioural experiments on crickets to support their theory. Larsen and Michelsen [88] suggested that the differences in the phases of the internal and external sound pressures on either tympanum were due to the sound input from the contralateral spiracle. However, Weber and Thorson [131] demonstrated that blocking the transverse trachea did not hinder sound localization. Finally, by measuring the change in amplitude and phase of the sound travelling from the contralateral tympanum, the contralateral spiracle and the ipsilateral spiracle, Michelsen et al. [91] demonstrated that both the contralateral and the ipsilateral spiracle sound inputs played a significant role, but the directionality was the direct consequence of phase shifting in the sound path from the contralateral spiracle to the ipsilateral eardrum.

Recently, Schöneich and Hedwig [112] have observed the phonotactic steering in the female field cricket tethered on a trackball to a male calling song. They discovered that females performed reliable phonotaxis for angular deviation of  $1^\circ$  in the frontal  $\pm 30^\circ$ . Furthermore, laser vibrometry revealed that the interaural amplitude differences increased linearly between  $0^\circ$  and  $30^\circ$ .

### 2.2.3 *Ormia ochracea*

Perhaps the most unique ear belongs to a nocturnal fly, the tiny *Ormia ochracea* as depicted in Fig. 2.6. The female of the species is an acoustic parasitoid of crickets, and thus their auditory system is tuned to the calling songs of crickets as reported by William H. Cade [14].



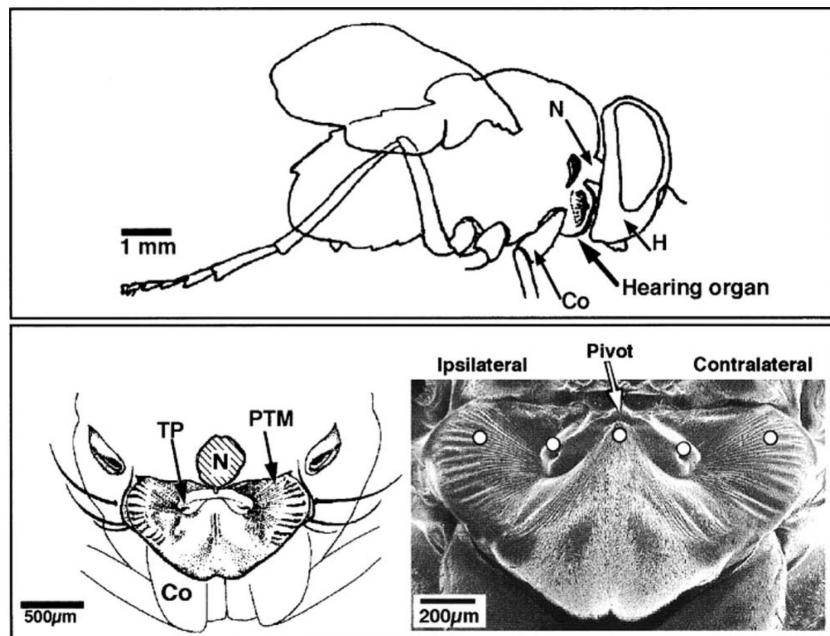
**Figure 2.6:** The parasitoid fly *Ormia ochracea* (Image courtesy: [Wikimedia Commons](#)).

Interestingly, the separation between the fly's ears is approximately 1.2 mm–1.5 mm, while the peak frequency of the cricket's mating call (which is relatively pure in frequency) is 4.8 kHz, which corresponds to a wavelength of approximately 70 mm. Furthermore, the auditory organs of the fly are located quite close to each other (approximately 520  $\mu$ m apart) in a common air-filled cavity [106, 29]. Such a minuscule ear separation renders the ITD cue between the ipsilateral and contralateral tympani significantly small, of the order of 1  $\mu$ s–2  $\mu$ s. Furthermore, the maximum length of the animal is less than 1 cm along any axis, and thus the 4.8 kHz sound waves do not diffract around the body, rendering the ILD cue almost non-existent. Acoustical measurements made with probe microphones at the the tympanal membranes by Robert et al. [104] revealed that the maximum ITD cues were of 1.45  $\mu$ s, and the maximum IID cues were less than 1 dB.

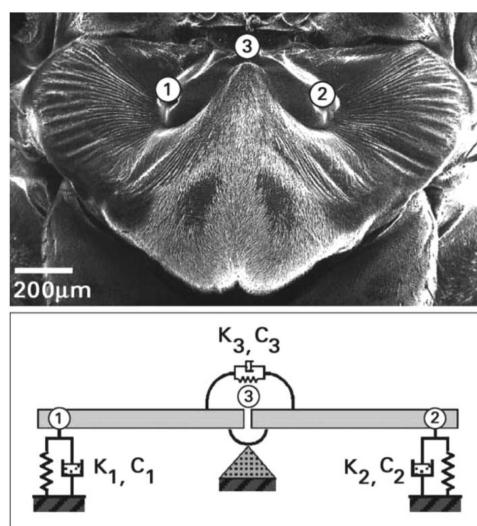
In spite of the relatively tiny amount of acoustic information available, it has been demonstrated by Thomas J. Walker [125] that the animal can unequivocally localize a cricket's calling song in the field. Later, Mason et al. [82] have demonstrated under laboratory conditions that the animal can localize the direction of sound sources in the azimuth plane with extraordinary precision, within 2°, during walking phonotaxis. A subsequent kinematic study on the directional acuity in the azimuth plane during walking phonotaxis by Mason at al. [81] has demonstrated that the rotational velocity of the fly, which translates to the directional orientation, varies systematically with sound source direction. This remarkable directional acuity is attributed to the unique structure of the ears (refer to Fig. 2.7). The ears are located inside the prothorax of the animal, just above the bases of their front legs, and behind the head.

The animal's two tympani are mechanically coupled to each other via a flexible inter-tympanal bridge (refer to Fig. 2.8) that pivots about a fulcrum. A mechanical spring-mass-damper model of the ear (refer to Fig. 2.8) proposed by Miles et al. [92] describes this mechanical coupling as a flexible mechanical lever, which causes the two tympanal membranes to deflect in opposite directions. When external sound pressure is impressing upon the ipsilateral side, the resultant vibrations of the ipsilateral tympanum are immediately coupled to the contralateral side, generating forces on the contralateral tympanum that effectively cancel out its vibrations due to the external sound pressure on the contralateral side. Further support for validity of the mechanical model was later provided by Robert et al. [105], who were able to elicit a corresponding mechanical response in one tympanum by mechanically stimulating the other tympanum. A more detailed functional explanation of the mechanical operation of the inter-tympanal coupling was subsequently provided by Daniel Robert [103]. The ear essentially amplifies

the tiny ITD cues into larger mechanical cues that can be processed by the nervous system.



**Figure 2.7:** The anatomy of the *Ormia ochracea* ear (taken from [1])



**Figure 2.8:** The inter-tympanal bridge coupling the tympanal membranes in *Ormia ochracea* and its equivalent mechanical model (taken from [1]).

## 2.3 Synthetic Approaches to Sound Localization

The problem of sound localization has been extensively studied from an engineering perspective. This section presents a non-exhaustive overview of some of the latest techniques reported in the literature. Paweł Strumillo [118] has compiled a comprehensive collection of the latest approaches in sound localization.

### 2.3.1 Traditional Approaches

Sound source localization is a central challenge in diverse fields, for instance mobile robotic audition, human-robot interaction and social robotics applications, hearing aids for the deaf. DiBiase et al. [28] have compiled a review of various localization strategies. The traditional methods for sound source localization are the steered beamformer based localization and the time-difference-of-arrival (TDOA) method. The former is based on the concept that localization can be achieved by filtering and summing up the contribution of a delayed version of each microphone's output in a multi-microphone array and determining the source direction for which this sum is maximised [16]. The principle behind the latter this approach is that given an array of microphones that are displaced from each other in space by a known distance, the sound waves arriving at any two consecutive microphones will be successively delayed in time by a fixed value proportional to the distance between the two microphones. By cross-correlating the signal obtained at two microphones and observing the location of the peaks in the cross-correlated signal, information about the relative location of the source can be extracted [68, 7, 15]. Knapp and Carter [68] have suggested that sound localization based on the TDOA methods present significant computational advantages over steered beamformer methods. Harry L. van Trees [122] has presented a through treatment of array signal processing. Brandstein and Ward [11] have compiled a detailed reference on microphone arrays and associated techniques.

The majority of the TDOA approaches employ multi-microphone arrays, typically in either a linear or grid configuration. The typical number of microphones utilized varies from as low as 4 to as high as 128. Kobayashi et al. [71] have evaluated a number of multi-microphone configurations in terms of their localization performance. However, grid arrays are advantageous in that they can accommodate a larger number of microphones in a relatively smaller and more convenient profile than linear arrays. It is however infeasible to employ large microphone arrays in robotic and hearing aid

applications due to constraints on size and available computational resources. Kagami et al. [62] developed a robotic sound localization system for two-dimensional sound localization with a 24-microphone circular array, and a localization accuracy of 5°. Tamai et al. [119] demonstrated a circular microphone array for two-dimensional sound localization with 32 microphones with a localization accuracy of 3°. Later, Tamai et al. [120] extended their previous work with a multi-ring microphone array for three-dimensional localization, with 32 microphones, in which 24 microphones are arranged in three rings, while a fourth ring of 8 microphones encompasses the three rings. They report the localization error to be within 5°. Valin et at. [121] developed a 8-microphone array three-dimensional localization system in the form of a rectangular cuboid. Coupled with weighting of regions in the spectrum with high signal-to-noise ratio (SNR), the system achieves a localization accuracy of 3° degrees. Liu and Shen [77] proposed a modified cross-correlation approach for sound localization with a 4-microphone array. Hwang et al. [55] employed sparse coding to decompose two-microphone signals into their frequency, time and amplitude components and then utilized a Self Organizing Map (SOM) [43] to estimate the azimuth. Murray et al. [96] proposed a combination of cross-correlation and a recurrent neural network (RNN) [43]. Kim et al. [66] proposed a 2-microphone azimuth localization system for humanoid robots. Mungamuru and Aarabi [95] proposed a new class of sound localization approaches designed to work with microphone arrays by exploiting a priori information regarding the directivities of the source and the microphones as well as the source-microphone distances, claiming higher localization accuracy than the existing approaches.

In general, the trend has been shifting towards minimizing the number of microphones required to perform sound localization, at the cost of increasing computational complexity. However, given the continuous rise in computational power and resources available for robots, computational complexity is not seen as a limiting factor.

Another localization approach is based on calculating the spectral cues generated by the filtering effects of the pinnae, head and torso in dummy heads and humans. Here the head-related transfer function (HRTF) [17], either individualized or generalized, is determined first and based on the HRTF response to incoming sound, the location of the sound can be determined. These approaches usually employ two microphones placed inside the ears of an artificial human or robotic head, since the focus here is using spectral cues which are dependent on the specific morphology of the head and pinnae. However, multi-microphone arrays are not uncommon in such approaches. Huang et al. [53] proposed a 4-microphone array-based robotic three-dimensional sound localization system where they determined the three-dimensional ITD cues by measuring the HRTFs of a

spherical head and cross-correlating the same. Kwon et al. [73] extended the cross-correlation approach by devising a mapping function from the measured HTRF of a spherical robotic head, which maps the cross-correlated output signals of each microphone pair of a 3-microphone array into the spatial domain around the head. Sounds are then localized by summing all the microphone signals in the spatial domain to obtain an estimate of the azimuth and elevation. Recently however, Rodemann et al. [107] have demonstrated that with a combination of two microphones and custom designed bionic ears, three-dimensional sound localization with a robotic head is possible without calculating the HRTF. They report an azimuth error of  $2.8^\circ$  with a 71% accuracy and an elevation error of  $12.3^\circ$  with a 54% accuracy.

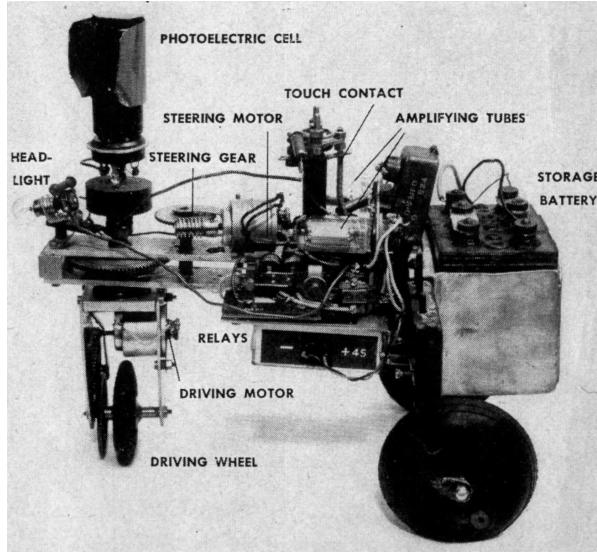
Continuing the trend for smaller microphone arrays, Eugeniusz Kornatowski [72] has very recently proposed the use of a single point surround microphone called TetraMic from Core Sound Inc. (refer to Fig. 2.9) to determine a direction of the sound source localization (SSL). This is a special professional quality microphone that utilizes four transducers oriented towards the vertices of a tetrahedron. Kornatowski reported average azimuth errors of  $3^\circ$ – $6^\circ$  and average elevation errors of  $4^\circ$ – $6^\circ$ .



**Figure 2.9:** The single point surround microphone TetraMic from Core Sound Inc. (Image courtesy: [Core Sound Inc.](#)).

## 2.4 Biomimetic Robot Models

Robotic models of biological systems have been in existence for at least the last 60 years. The first biologically inspired robots, robotic tortoises to be precise (refer to Fig. 2.10), were built by W. Grey Walter [127, 128] in the 1950s. His inspiration was the question whether the complexity of cerebral functions of the brain depended less on the number of units that made up the entire brain than on the richness of their interconnections. Over 30 years later, Valentino Braitenberg [10] described the functions of the brain in psychological terms, such as fear, aggression etc., as products of diverse forms of interconnections between sensors and actuators through his robotic vehicles.

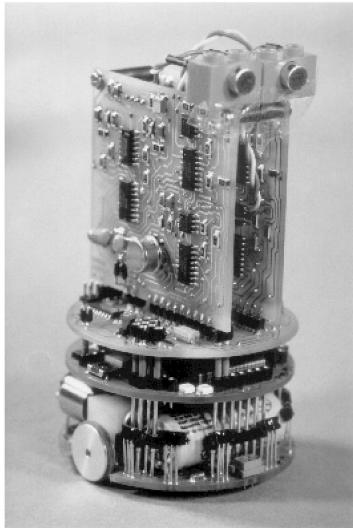


**Figure 2.10:** One of the robotic tortoises built by W. Grey Walter (taken from [50]).

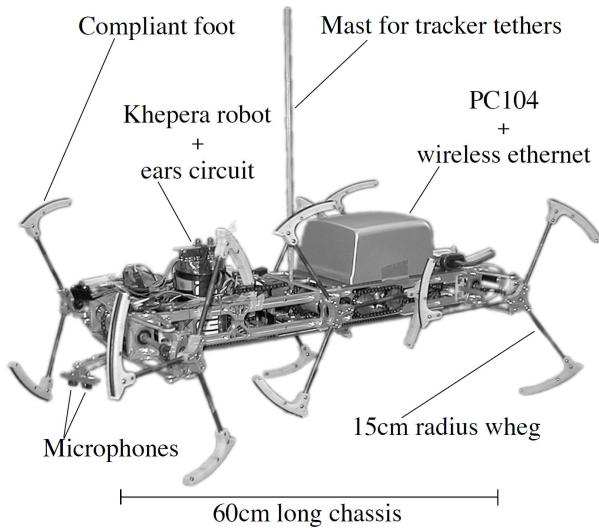
Even in the era when biology and robotics did not overlap, a wide variety of biological systems were studied by roboticists to draw inspiration for mobile machines. Walking creatures inspired a plethora of legged robots, with the reasoning that walking machines were decidedly better at traversing rough terrain than wheeled ones. The primary goal was to mimic the gait patterns, not to understand the biological reasons for that particular gait. Similarly, in recent times flying insects and birds have inspired a number of flapping wing robots, and interest in fish locomotion is on the rise. The number of biologically-inspired robots that have been built has been steadily increasing, and there are numerous such robot systems in existence today. These robots can be broadly classified into two types — those that have been built purely to mimic a particular animal species' ability to solve a relevant problem and those that have been built to mimic *and* understand the underlying biological principles of a particular trait of a particular ani-

mal species. The former group includes a large percentage of all the biologically-inspired walking robots, since there is a strong focus on legged locomotion in the robotics research community. The latter group however, comprises a much smaller percentage, mainly due to the relatively smaller number of researchers active in this area. However, employing robots as tools to study biological principles is steadily becoming the norm. Extensive literature can be found regarding the former group, but here the focus is maintained on the latter group. In the next few paragraphs, some of the robots from the latter group are introduced.

One of the most well studied biological systems that has been studied via biorobotic modelling is the cricket's auditory system. Lund et al. [78] implemented an auditory model of the cricket with only two sound inputs instead of four on a Khepera mobile robot (refer to Fig. 2.11(a)) and demonstrated successful phonotaxis towards a recorded version of a male cricket's calling song, as well as the frequency selectivity of the auditory model. Horchler et al. [52] implemented the cricket auditory model on the Whegs mobile robot (refer to Fig. 2.11(b)), which is an autonomous outdoor robot platform inspired by cockroach locomotion [74]. They demonstrated successful phonotaxis with the robot in an outdoor environment.



(a) The Khepera robot used by Lund et al. (taken from [78]).



(b) The Whegs robot used by Horchler et al. (taken from [52]).

**Figure 2.11:** Two of the different robotic platforms used in cricket hearing research.

Webb and Scutt [130] implemented the recognition behaviour of female crickets for the calling songs of the males, via a spiking-neuron model with only four neurons. They

were able to demonstrate that the directionality of the robot depended on the relative latencies in the firing of the neurons, and that calling song recognition implicitly emerged from the neurons' properties. Later, Webb and Harrison [129] combined the cricket auditory model with analogue VLSI optomotor circuitry and demonstrated that the addition of a visual sense can enhance phonotactic steering. Reeve and Webb [100] evaluated a number of different neural circuits for audition, vision and motor control for cricket phonotaxis on a Koala robotic platform (from the [K-Team Corporation](#)).

In the last decade, the salamander has emerged as a model organism for the study of neurobiological principles of vertebrate locomotion. The salamander is a amphibian tetrapod (refer to Fig. 2.12), and is considered to be a close link between aquatic and the first terrestrial vertebrates. Since it can exhibit both swimming and walking locomotion patterns, it is regarded as a key animal in the evolution of terrestrial locomotion from aquatic locomotion [24, 38]. The number of neurons in the central nervous system of the salamander is several orders of magnitude less than mammals [109, 110]. This makes it relatively simpler to understand and to model artificially.



**Figure 2.12:** A fire salamander or *Salamandra salamandra* (Image courtesy: [Wikimedia Commons](#)).

Electromyographic (EMG) muscle recordings performed by Frolich and Biewener [37] demonstrated that during swimming, the salamander exhibits travelling waves of motion that propagate from the head to the tail along the axis of the body, with the limbs folded in towards the body. During terrestrial locomotion, the salamander exhibits a single standing wave of motion, with the nodes of the wave centred on the anterior and posterior pelvic girdles. In synchrony with this standing wave, the movement of diagonally opposite limbs are in phase, while laterally opposite limbs are out of phase, in co-ordination with the movements of the body. Further EMG recordings performed by Delvolv   et al. [26] confirmed the same and proposed that a chain of coupled segmental oscillatory networks may generate both swimming and walking locomotion patterns. Ashley-Ross and Bechtel [6] performed kinematic analysis of the locomotion transition

from swimming to walking and arrived at the same conclusion.

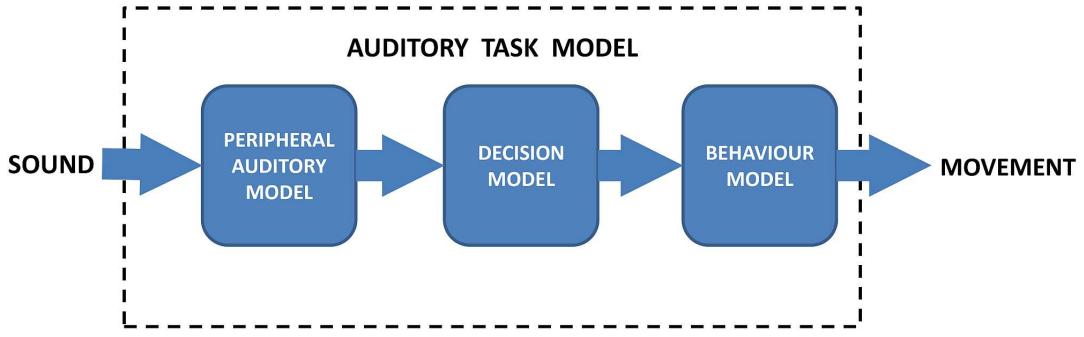
In subsequent in vitro experiments with an isolated spinal cord preparation of the adult newt *Pleurodeles waltl*, Delvolv  et al. [27] demonstrated that motor patterns were generated by central pattern generators (CPGs) [51] distributed in the spinal cord. Since the salamander's central nervous system is quite similar another thoroughly studied aquatic animal, the lamprey, studies of the lamprey's neurophysiological mechanism of motor pattern generation have aided the understanding of the salamander's motor pattern generation mechanisms. The lamprey is an eel-like fish that also exhibits the anguiliform swimming pattern as seen in salamanders. Therefore, CPGs have been extensively studied with the lamprey as an ideal test subject by Grillner et al. [40, 42, 41], Andrew D. McClellan [84] and Buchanan and Grillner [13]. Ekeberg et al. [31] proposed a computer-based biophysical model of the lamprey swimming circuitry. A connectionist model was subsequently proposed by  rjan Ekeberg [30]. Wall n et al. [126] also proposed a computer simulation of a biophysical model of the lamprey's spinal cord networks. James T. Buchanan [12] subsequently proposed a connectionist model as well.

Auke J. Ijspeert [56] developed a three-dimensional biomechanical model of the salamander in a realistic physics-based simulation, where the body and limbs were modelled as an interconnection of articulated segments, and the muscles of each segment were modelled as a spring-and-damper system. He subsequently developed a biologically plausible connectionist model of a central pattern generator capable of generating the aquatic and terrestrial gaits of the salamander and evaluated the same with a two-dimensional biomechanical model of the salamander in simulation [57]. Subsequently, Ijspeert et al. [59, 60] developed a robotic model, *Salamandra robotica* (refer to Fig. 2.13), of the salamander as well, which realistically reproduced the swimming and walking locomotion patterns of the salamander. Later, Chevallier et al. [18] described the walking and swimming gaits of the salamander from a biological and kinematic perspective. Karakasiliotis and Ijspeert [63] subsequently designed and built the second generation of the salamander robot, *Salamandra robotica II*.



**Figure 2.13:** The first version of the salamander robot, *Salamandra robotica*. Photograph by A. Crespi, courtesy Biorobotics Laboratory, EPFL.

Directional hearing in lizards has been extensively studied by Jakob Christensen-Dalsgaard [21, 22, 20, 23]. Lei Zhang performed a significant amount of research on robotic modelling of the peripheral auditory system of lizards. In his work, he questioned the nature of the coupling between the auditory signals of the two ears and its importance for directional hearing. He proposed an “auditory task model” [137] which incorporated the peripheral auditory system, a decision model and a behaviour model. The peripheral auditory model was utilized to model the effects of the physical structure of the lizard’s auditory periphery. The neuronal computations in the auditory nervous system to sense the tympanal vibrations, i.e. the output of the peripheral auditory model, and to generate motor commands were modelled by the decision model. The behaviour model was utilized to model a robotic actuation system which responded to the motor commands from the decision model. Figure 2.14 depicts the auditory task model.



**Figure 2.14:** The auditory task model proposed by Lei Zhang (redrawn from [137]).

Initially, Zhang et al. [138] designed and built a Lego based robotic model to evaluate the phonotactic performance of the auditory task model with a single sound source, and demonstrated the strong directionality of the system. Later, they investigated the directness of phonotaxis, i.e. the straightness of the robotic trajectories towards a sound source in the presence of a second sound source acting as a distractor [140]. They discovered that if the difference between the loudness of the sound source of interest and that of the distractor was less than 3 dB, then the robot had difficulty localizing the sound source of interest. Next, left-right asymmetry was introduced into the peripheral auditory model, and its effect on the localization error was investigated in simulation [139]. It was discovered that localization error was minimally degraded by asymmetries in the auditory task model for the peak directional frequency, i.e. the frequency for which the peripheral auditory model’s showed peak response, and the degradation increased for sound frequencies greater or lower than the peak directional frequency. In other words, the useful localization bandwidth, i.e. the band of frequencies for which the auditory task

model exhibited relatively low localization error was reduced when asymmetries were introduced into the peripheral auditory model. Subsequently, Lei Zhang [137] employed various neural networks as decision models in order to investigate whether it was possible to compensate for asymmetries in the peripheral auditory model by mechanisms in the decision model. It was discovered that neural networks were able to compensate for the asymmetries, and reduce the localization errors.



# Chapter 3

## Braitenberg Models

### 3.1 Introduction

The lizard’s peripheral auditory system has been extensively studied from a robotic perspective. It has been modelled and implemented as a step-controller in conjunction with a simple decision model (refer to Fig. 2.14) for a mobile robot by Zhang et al. [138], and successful phonotaxis behaviour has been experimentally demonstrated. The decision model was a set of *if–then* control rules — 1) *if* sound is coming from left *then* turn left with fixed speed, 2) *if* sound is coming from right *then* turn right with fixed speed, and 3) *if* sound is coming from front *then* go forward with fixed speed. Such a control strategy can be viewed as a sensor providing input to a robot controller, which makes motor decisions based on the sensor’s outputs. Although the desired robotic behaviour can be successfully generated by this decision model in terms of the robot actually localizing a sound source and the quality of the robot’s trajectory, it is rather the manner of interpretation of the auditory model’s outputs that is directly responsible for this behaviour and it reflects less upon the capabilities of the model in itself in performing successful phonotaxis. Furthermore, this does not imply that removing the decision model would necessarily result in failed phonotaxis behaviour. This warrants further investigation, by means of blurring the boundary between the sensory system and the controller such that there is a direct coupling between the sensory subsystem and the motor subsystem of the robot.

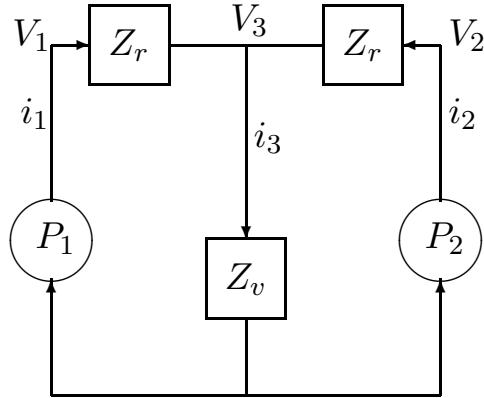
In this chapter, we investigate whether the auditory model can be directly employed to generate phonotactic behaviour on a mobile robot platform and evaluate its performance in a similar manner. Towards this end, simple Braitenberg vehicle-like sensori-motor couplings between the auditory model’s outputs and a mobile robot’s motor drive

signals are established, resulting in a control strategy where the sensor is regarded as the controller itself. Two independent experiments are performed. First, the experiments performed by Zhang et al. in [138] are repeated to compare the Braitenberg mobile robot's phonotactic performance in the absence of any explicit decision model to that when utilizing a decision model. The Braitenberg sensorimotor coupling also renders the control continuous, as opposed to step-control utilized by Zhang et al. Second, the individual sensorimotor couplings are independently modified by embedding various decision models indirectly into them, and the phonotactic performance is re-evaluated.

The work presented in this chapter covers and elaborates further upon the work reported by the author in [117] and in [114].

## 3.2 Theoretical Model of the Auditory System

Figure 3.1 shows the equivalent electrical circuit model [35, 36] of the pressure difference receiver ear model.



**Figure 3.1:** Lumped-parameter model of the lizard's peripheral auditory system.

Voltages  $V_1$  and  $V_2$  model sound pressures  $P_1$  and  $P_2$  at the left and right ear respectively. Currents  $i_1$  and  $i_2$  model the left and right tympanic motion in response to the sound pressures acting upon them. Impedance  $Z_r$  models the total effect of tympanal mass and stiffness and the Eustachian tubes connecting the tympani to the central cavity, while  $Z_v$  models the central cavity itself.  $Z_r$  is the same for both left and right sides since, for the sake of simplicity, the ear structure is assumed to have left-right symmetry. Voltage  $V_3$  represents the sound pressure generated in the central cavity due to the interaction of the sound pressures experienced from the left and right side. This causes

current  $i_3$  to flow through the impedance  $Z_v$ , modelling the movement of sound waves inside the central cavity as the pressure inside it varies. All impedances are complex numbers with frequency-dependent values.

The physical amplitude of the external sound waves is the same at both the tympani due to diffraction effects mentioned in Sect. 1.1, and thus  $V_1$  and  $V_2$  differ only in terms of their phase. On the other hand,  $i_1$  and  $i_2$  differ in both phase and amplitude, because of the interaction between the left and right internal sound pressures and this difference represents the proximity of either of the ears to the sound source. We can represent this difference mathematically [138] as

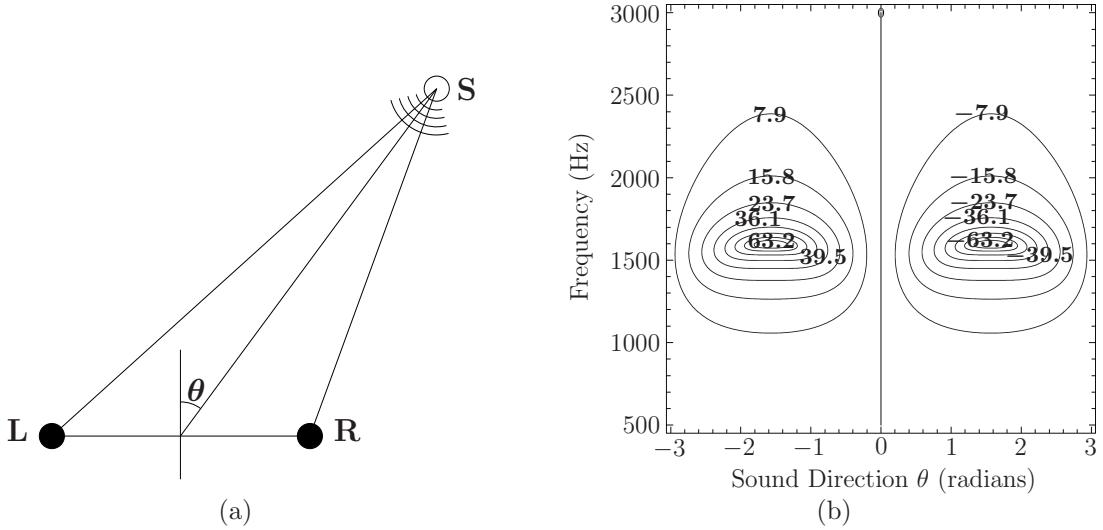
$$\left| \frac{i_1}{i_2} \right| = \left| \frac{G_I \cdot V_1 + G_C \cdot V_2}{G_C \cdot V_1 + G_I \cdot V_2} \right| = \left| \frac{G_I + G_C \cdot \frac{V_2}{V_1}}{G_C + G_I \cdot \frac{V_2}{V_1}} \right| , \quad (3.1)$$

where frequency-dependent gains  $G_I$  and  $G_C$  model the effect of sound pressure on the motion of the ipsilateral (on the same side as the sound source) and contralateral (on the opposite side as the sound source) tympani respectively. In signal processing terminology these gains are analogue filters and their coefficients have been experimentally determined, by taking measurements of the tympanic membrane vibrations via laser vibrometry, by Christensen-Dalsgaard and Manley [21]. Expressing  $i_1$  and  $i_2$  in dB, (3.1) can be rewritten as

$$i_{\text{ratio}} = 20 (\log |i_1| - \log |i_2|) \text{ dB} . \quad (3.2)$$

A plot of the current ratio given by (3.2) is shown in Fig. 3.2(b) for different frequencies and radial positions  $\theta$  of the sound source  $\mathbf{S}$  with respect to the left and right eardrums  $\mathbf{L}$  and  $\mathbf{R}$  respectively (refer to Fig. 3.2(a)).

It is evident that the model responds well to a wide range of frequencies (about 1000 Hz–2200 Hz), with a peak response around 1600 Hz. Henceforth, this range is considered as the frequency range of interest. Positive values of  $i_{\text{ratio}}$  indicate  $|i_1| > |i_2|$  and negative values indicate vice versa. Since the model is symmetrical, the values are identical in magnitude on either side of the center point  $\theta = 0$ . Also, the response is locally symmetrical about  $-90^\circ$  and  $+90^\circ$ , and henceforth this range  $[-90^\circ, +90^\circ]$  is considered as the range of interest of radial positions.

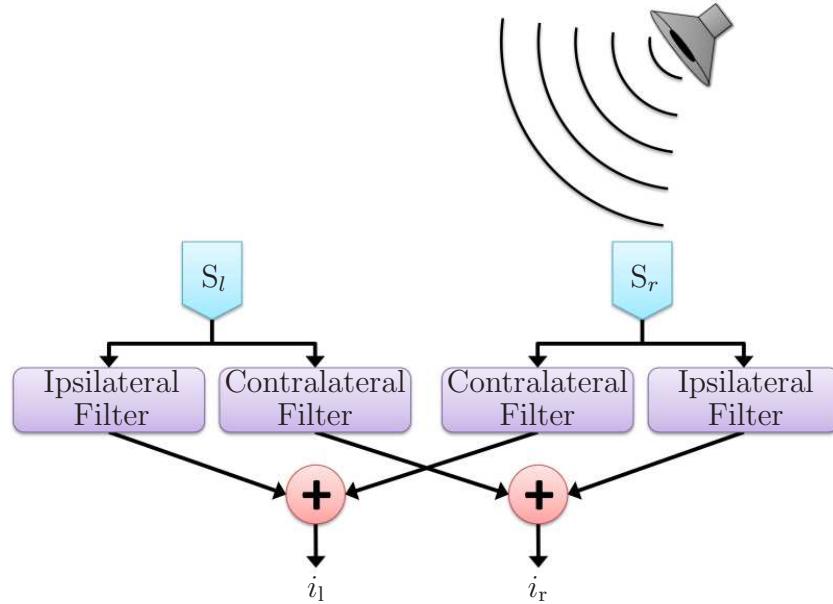


**Figure 3.2:** Directionality of the peripheral auditory system model. (a) Radial displacement of sound source **S** with respect to the left (**L**) and right (**R**) eardrums, taken and redrawn from [138]. (b) Current ratio  $i_{\text{ratio}}$  plot from (3.2) in dB.

### 3.2.1 Auditory Model Implementation

The auditory model is a set of coupled analogue filters with coefficients experimentally determined via laser vibrometry [21]. The electronic components, namely the inductors, required to implement these filters have extremely large values of the order of  $10^3$  H. An inductor, for example of the air core variety, with such large inductance would need to be approximately 30 m in diameter and 25 m in length. Although it is possible to scale the values of the other components down such that the required inductance is of realizable magnitude, it is quite simply a more complicated and time-consuming task and there is not enough justification for realizing the model using analogue components in terms of experimental gains. For this reason the auditory model was realized digitally. Therefore, digital counterparts of these filters were derived, for a sampling rate of 50 kHz which ensured the absence of anti-aliasing effects within the 1000 Hz–2200 Hz frequency range. Of course, a sampling rate of 5 kHz is more than sufficient to avoid aliasing in this range, but as a general rule of thumb oversampling is preferable. The digital filters thus obtained were 4<sup>th</sup> order bandpass filters with four coefficients per filter. Figure 3.3 depicts the block diagram of the implementation of the peripheral auditory model.

Consider for example the sound signal  $S_l$  on the left side, it is filtered by the ipsilateral filter, a process which is analogous to the motion of the left eardrum due to the external sound pressure, and by the contralateral filter, a process which is analogous to the

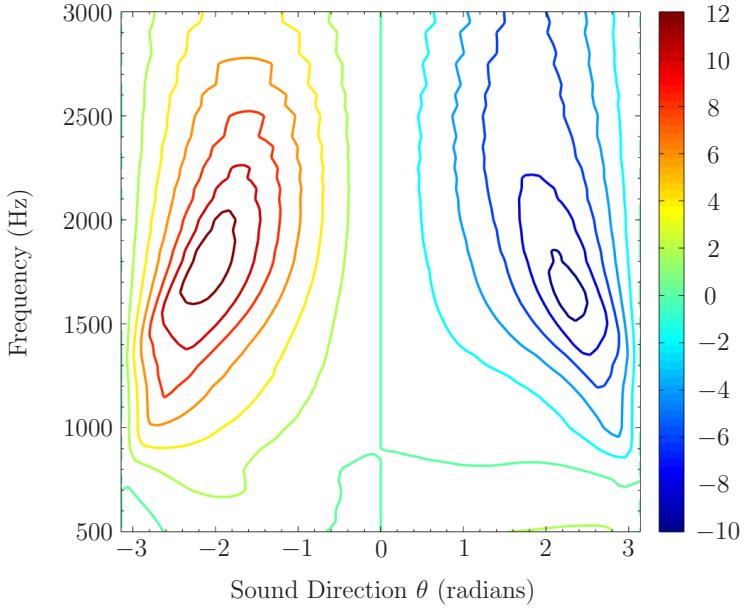


**Figure 3.3:** The digital implementation of the peripheral auditory system.

motion of the left eardrum due to the external sound pressure at the right eardrum. Similarly, for the sound signal  $S_r$  on the right side, the filtering by the ipsilateral and contralateral filters is analogous to the motion of the right eardrum due to the external and internal sound pressures respectively. The addition of the left ipsilateral filter output with the right contralateral filter output is analogous to the superposition of the external and internal sound pressures experienced by the left eardrum, while the addition of the right ipsilateral filter output with the left contralateral filter output is analogous to the superposition of the external and internal sound pressures experienced by the right eardrum. After addition we get the left and right outputs  $i_l$  and  $i_r$ , which are analogous to the resultant motion of the left and right eardrum respectively. Finally, these are converted into their corresponding decibel values. Figure 3.4 depicts the response of the digital implementation.

### 3.3 Braitenberg Vehicles

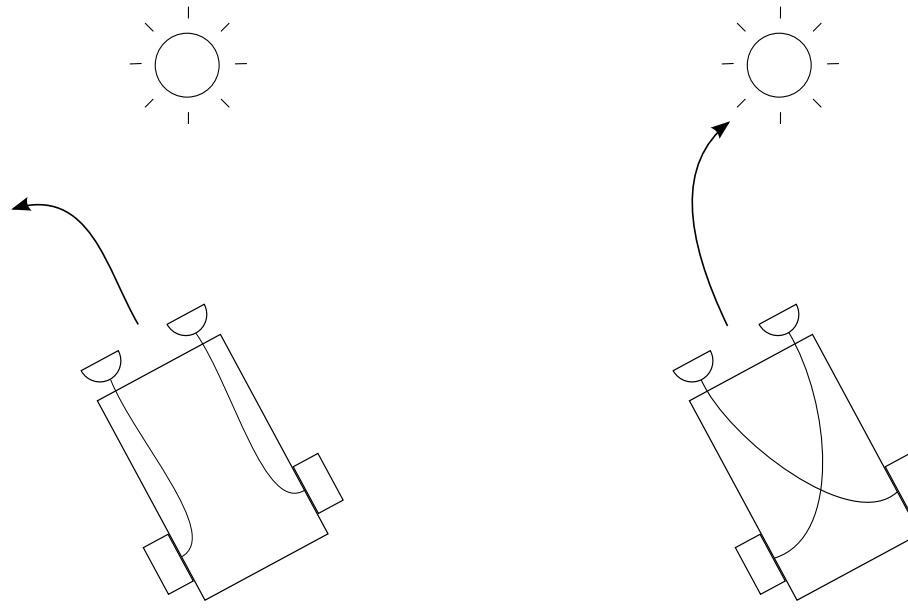
The conceptual Braitenberg vehicles were proposed by Valentino Braitenberg in an attempt to describe the evolution of different behaviours as a consequence of modifying the structure of the sensorimotor couplings [10]. A Braitenberg vehicle is an autonomous vehicle consisting of one or more simple sensor(s) (which could be any kind of sensor, for example a sound sensor such as a microphone in this case) and independent motor-



**Figure 3.4:** The relative response in dB of the digital implementation of the auditory model.

driven wheels as actuators. In the simplest instance of this vehicle, a sensor is directly connected to a single motor, which in turn is connected to a single wheel. The amplitude of the output of the sensor directly affects the speed of the motor to which it is connected. The greater the excitation of the sensor, the larger is the amplitude of its output, the greater is the speed of the corresponding motor, and the faster the connected wheel rotates.

These vehicles have no decision model in control, since the sensorimotor coupling is straightforward and direct. Depending on the nature and mechanism of the coupling between the sensors and motors, the vehicle exhibits different goal-oriented behaviours — illustrated through an example of a Braitenberg vehicle with two light sensors in Fig. 3.5. In the vehicle depicted in Fig. 3.5(a) the ipsilateral sensor, i.e. the one nearest to the light source, is excited to a greater degree than the contralateral sensor. This results in the ipsilateral motor having a greater speed than the contralateral one, turning the vehicle away from the light source. In the vehicle depicted in Fig. 3.5(b) once again the ipsilateral sensor is excited to a greater degree than the contralateral sensor. However, due to the cross-coupling, this results in the contralateral motor having a greater speed than the ipsilateral one, turning the vehicle towards from the light source. This simple cross-coupling serves as the underlying control architecture for evaluating the phonotactic performance of the auditory model.



- (a) The left and right sensors are connected to the respective motors. This vehicle always avoids the light source.
- (b) The sensorimotor couplings are crossed. This vehicle always moves towards the light source.

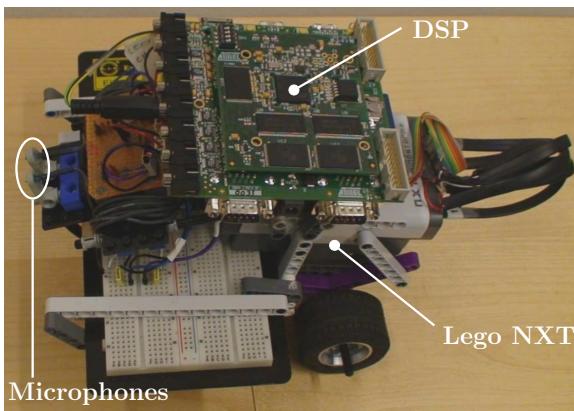
**Figure 3.5:** Simple Braitenberg vehicles with light sensors redrawn from [10].

## 3.4 A Basic Braitenberg Lizard

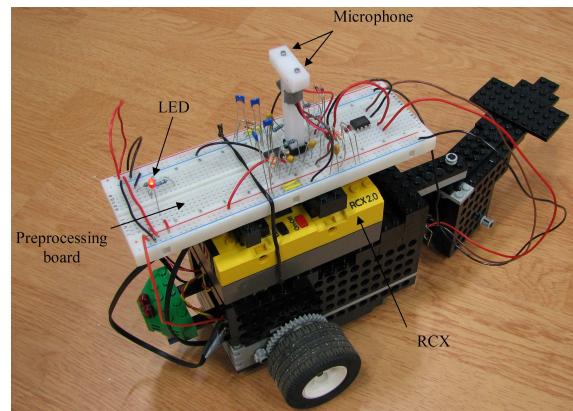
### 3.4.1 The Robotic Platform

The physical robot consisted of an Atmel DIOPSIS® digital signal processor (DSP) development board with the lizard ear model implemented on it in the form of digital filters, and a Lego Mindstorms NXT brick controlling two motor-driven wheels. The sole processing unit of the board was the Atmel DIOPSIS® 940HF, a dual-core DSP chip optimized for audio applications, integrating an Atmel ARM9 processor (specifically, the ARM926EJ-S) and a VLIW floating-point DSP on a single chip. The main reasons for choosing this particular platform were its ease of programmability and flexibility. Two omnidirectional microphones (model FG-23329-P07 from Knowles Electronics) were mounted at the front of the robot with a physical separation of 13 mm between them. This separation is similar to that found between the tympani of a typical lizard [21], because the auditory model's parameters have been derived based on measurements with the real animal and a different microphone separation would create a mismatch between the model and the external stimuli it expects. The voltage signals from these microphones were pre-amplified and fed into the DSP, where they were processed by

the lizard ear model and the left and right output powers were computed in dB as  $20 \log |i_l|$  and  $20 \log |i_r|$ . These values were then scaled to lie within the range of the argument to the motor drive speed command of the NXT brick (0–100), and serially transmitted as 8-bit values to it. The range 0–100 signified that the power applied to the motor lay between 0% (no power applied, motor is not running) and 100% (full power applied, motor is running at full speed). The scaling was done by first reducing the magnitude of both powers in fixed decrements of 1 until either of the values lay between 0 and 1. The resulting values were then multiplied by a scaling factor of +10, and an offset of +28 was added to both. With these choices, the forward and turning speeds of the Braitenberg vehicle implementation were roughly matched to those of the robot used by Zhang et al.. The resulting values were transmitted to the corresponding motors thus implementing the crossed sensorimotor coupling of a Braitenberg vehicle. In contrast, the step-control implementation employed by Zhang et al. used the aforementioned *if-then* decision model, with the robot driving straight or turning left or right with identical and constant speeds, as determined by the sign of current ratio in dB given by (3.2). Figure 3.6 depicts the NXT-based mobile robot employed and the RCX-based robot as employed by Zhang et al..



(a) The LEGO NXT based mobile robot.



(b) The LEGO RCX based mobile robot used by Zhang et al..

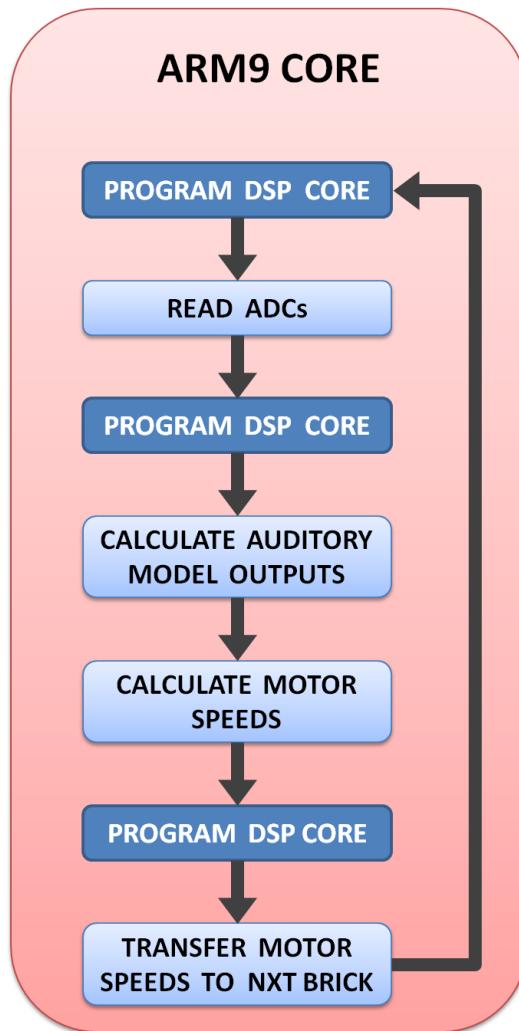
**Figure 3.6:** The mobile robot platforms.

The step-control strategy used by Zhang et al. consisted of calculating the auditory model outputs every 130 ms or at a frequency of approximately 7.7 Hz. After every calculation, the drive signals were presented to the motors, and the motors were driven until the next calculation was available. The speed of both the motors was always maintained such that the forward speed of the robot was 117.25 mm/s. In the continuous control strategy, the minimum and maximum forward speed of the robot were 0 mm/s

and approximately 345 mm/s respectively, and the average forward speed of the robot during the trials was measured to be approximately 118.6 mm/s. Furthermore, the auditory model outputs were calculated at the maximum possible frequency of approximately 166.67 Hz. This maximum possible control frequency was employed to ensure that the sensorimotor coupling remained as faithful as possible to the one depicted in Fig. 3.5(b). Since there are no delay elements embedded in the couplings, all attempts were taken to ensure that there were no delays in the implementation except those that were absolutely unavoidable, namely, the inherent processing delays which accumulated to approximately 6 ms. This delay was incurred due to the limited functionality offered by the software tools required for development, which at the time were released while still in their early stages of development. The compiler libraries for the ARM9 processor core did not support reading data from the analogue-to-digital converters (ADCs) or communicating over the digital I/O port used to transfer the motor speed data to the NXT brick. Rather, it was the core libraries for the DSP processor core that provided functions for performing DSP-related tasks and for communicating with the ADCs and over digital I/O ports. Therefore the three main tasks, namely, the reading of the ADCs, the auditory model calculations and data transfer to the NXT brick were implemented on the DSP core. However, the DSP core could only run when programmed through the ARM9 core, and furthermore could perform one task at a time, i.e. it could not perform communication with the ADCs and then the DSP calculations as one task for instance. Thus the DSP core had to be reprogrammed online through the ARM9 core for each task. A supervisory control loop on the ARM9 core was in charge of scheduling each task. At the start of every new control loop, the ARM9 core first programmed the DSP core to read the ADCs, then to perform the auditory model calculations, and finally to calculate and transfer the motor speed values to the NXT brick. These programming cycles introduced the aforementioned delays in the control loop. Figure 3.7 depicts the supervisory control loop running on the ARM9 core.

### 3.4.2 Experimental Setup

The experimental setup was similar to the one used by Zhang et al. in [138]. A sound dampening chamber was utilized as the experimental arena. A loudspeaker (model MD 42491 from Medion AG) served as a continuous monotone sound source. The robot started from a fixed initial location 2 m in front and 1.5 m to either the left or right of the loudspeaker, oriented along the  $x$  axis and allowed to move autonomously within the

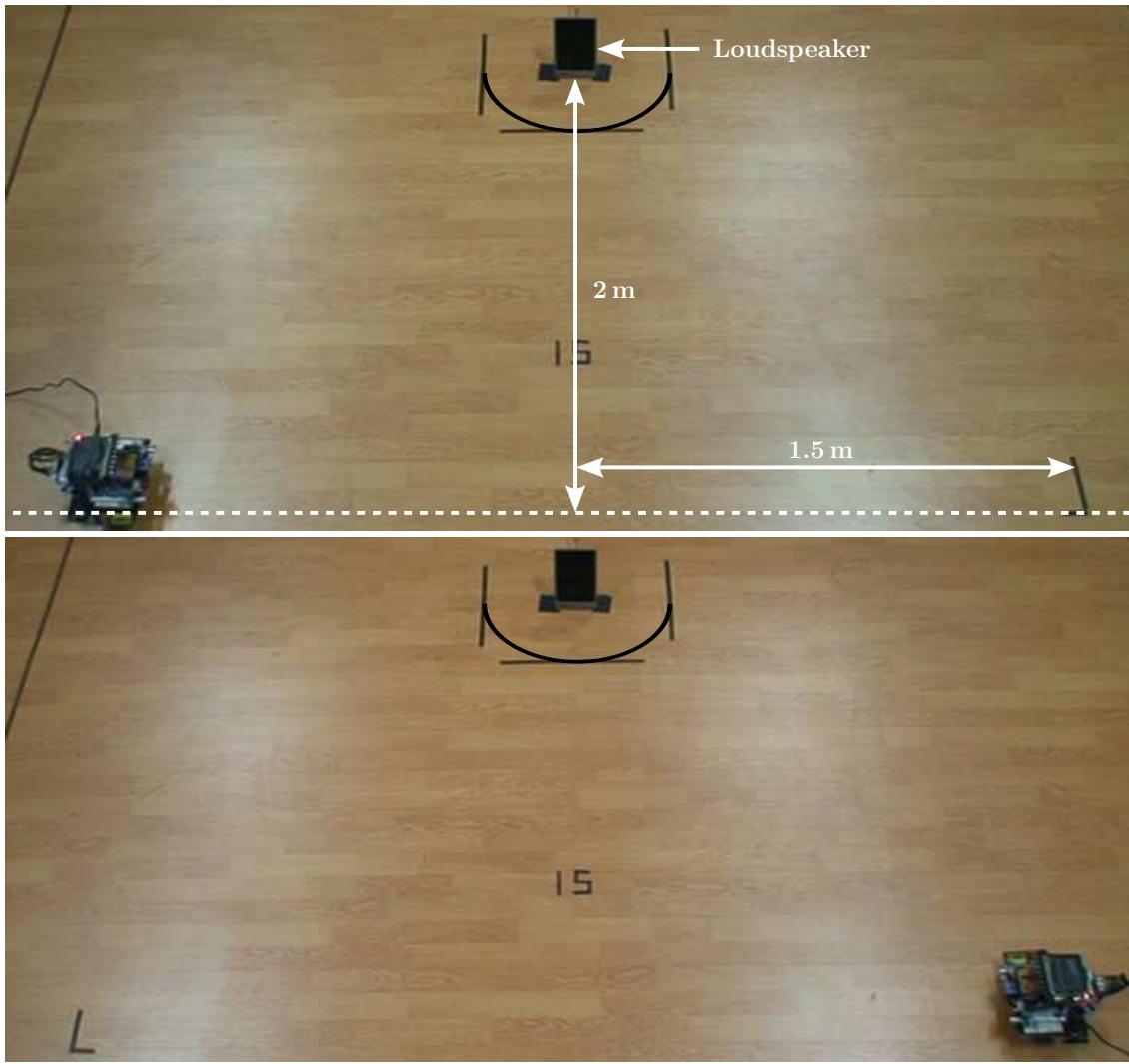


**Figure 3.7:** The supervisory control loop running on the ARM9 core. The dark blue boxes signify time delays introduced into the loop.

test arena boundaries until it either 1) hit the loudspeaker, 2) moved behind the loudspeaker or 3) moved outside the arena boundaries. During the motion of the robot, its position in terms of  $x$  and  $y$  coordinates was recorded via an overhead motion tracking system using an infrared camera tracking an infrared LED mounted on top of the robot. Figure 3.8 depicts the experimental arena and the robot's initial positions.

Multiple sets of experiments were performed in the relevant frequency range of 1000 Hz to 2200 Hz in steps of 50 Hz. In each set, the loudspeaker emitted a continuous tone of a given frequency, and the robot had to localize and move towards the loudspeaker. The trajectory of the robot was tracked from the initial position until the one of the three finishing conditions was met. This was done 20 times in total, divided into 10 trials with the robot starting from the left side of the loudspeaker and 10 trials

with the robot starting from the right side. The final outcome of each trial was classified as either a *hit* if the robot hit the loudspeaker, a *near hit* if the robot passed within a circular boundary of radius 26 cm around the loudspeaker or a *miss* if the robot either stayed outside the circular boundary or moved outside the arena boundaries. The total number of *hits*, *near hits* and *misses* were recorded for all the trials. As in the step-control experiments, successful localization was defined as achieving a minimum of 50 % *hits*.



**Figure 3.8:** The experimental arena and the initial positions of the robot.

### 3.4.2.1 Trajectory Analysis

In order to determine the performance of the robot in terms of its trajectories, a “directness” statistic [113] given by (3.3) was employed. It measured the average straightness or

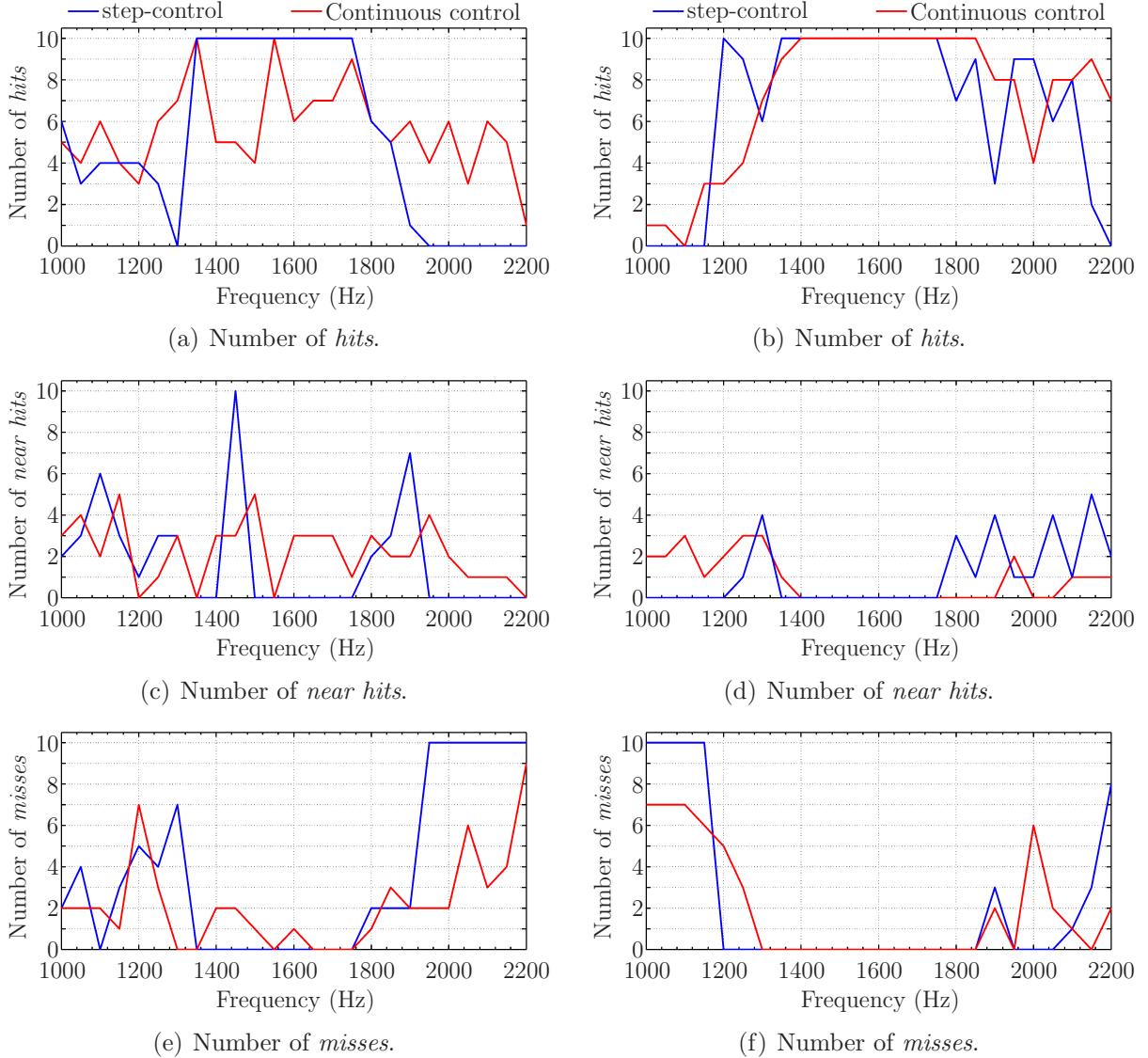
“directness” of a given robot trajectory vector from the starting point to the loudspeaker. A given trajectory was divided into  $n = 13$  vectors, each of length  $l$ . For each vector, the heading  $\theta$  relative to the position of the loudspeaker was calculated. Then these headings were averaged over the total number of vectors to obtain the average heading  $\vec{v}_{\text{avg}}$ . This procedure was repeated for trajectories corresponding to all frequencies. The polar co-ordinates (1,0) represent the ideal trajectory with average vector of length 1 and direction 0, i.e. the robot moves in a perfectly straight line from the initial position to the loudspeaker. The closer the data points for the 10 trials are to (1,0), the straighter or more “direct” are the trajectories on an average.

$$\vec{v}_{\text{avg}} = \frac{1}{\sum_{i=1}^n l_i} \left( \sum_{i=1}^n l_i \cos \theta_i, - \sum_{i=1}^n l_i \sin \theta_i \right) \quad (3.3)$$

### 3.4.3 Results and Discussion

Figure 3.9 depict the results of the step and continuous control implementations in terms of the number of *hits*, *near hits* and *misses*. It is evident that the physical implementation validates the theoretical model, showing strong directionality over the range of 1400 Hz–1900 Hz.

Viewing the loudspeaker from the robot’s perspective, when the initial position of the robot is to the right side of the loudspeaker both implementations exhibit quite similar performance statistics over the entire frequency range. It is evident from Fig. 3.9(b) that the robot exhibits successful localization over the continuous frequency range 1300 Hz–1950 Hz. In comparison, when employing step-control with a decision model, successful localization is observed over the continuous frequency range of 1350 Hz–1850 Hz. When the initial position of the robot is to the left side of the loudspeaker, there are distinct fluctuations in the performance statistics of the continuous control implementation compared to the step-control implementation, as evident in Fig. 3.9(a). However, the robot exhibits successful localization over the frequency range 1300 Hz–1950 Hz as well. In comparison, when employing step-control with a decision model, successful localization is observed over the continuous frequency range of 1350 Hz–1850 Hz. The overall performance trend over the relevant frequency range for the two implementations tends to be similar, but the presence of a decision model results in a better overall performance than the continuous control strategy. This is a result of both the inclusion of the decision



**Figure 3.9:** Comparison of step vs. continuous control in terms of the number of *hits*, *near hits* and *misses*. Figures (a), (c) and (e) depict the results for trials from the left initial position. Figures (b), (d) and (f) depict the results for trials from the right initial position.

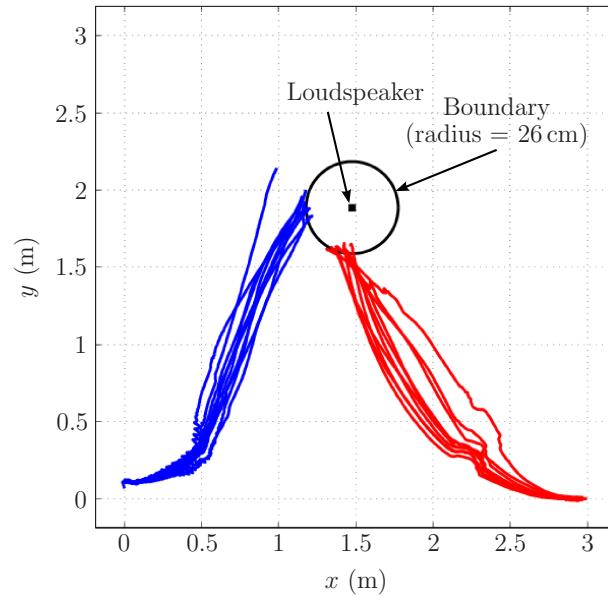
model and the kinematics of the robot model. The decision model implemented three simple behaviours based solely on the difference between the left and right output powers of the auditory model. If the difference was positive or negative it commanded the robot to respectively turn left or right with a fixed speed for a fixed amount of time irrespective of the magnitude of the difference, and if the difference was zero it commanded the robot to drive straight ahead with a fixed speed for a fixed amount of time. This strategy makes the steering quite insensitive to the individual magnitudes of the auditory model's

output powers. Furthermore, the relatively slow speed of 117.25 mm/s and a small but fixed turning radius of 1 body length or 142 mm also helped to limit the magnitude of movement error made by the robot. Since the motors are driven for 130 ms between two successive motor commands, the maximum forward distance covered by the robot between two successive motor commands is  $117.25 \cdot 130 \cdot 10^{-3} = 15.2425$  mm. In the continuous control strategy, the motor speeds are directly proportional to the respective outputs of the auditory model, and thus the steering is quite sensitive. Furthermore, since the turning radius of the robot is dynamically varying depending on its orientation relative to the loudspeaker, it easily surpassed the minimum value of 178 mm during the trials. Naturally, it is possible to tune the scaling parameters of the Braitenberg sensorimotor couplings to further reduce the motor speeds which would help in reducing the movement error, but this would create an unwanted mismatch between the speeds of the two robots.

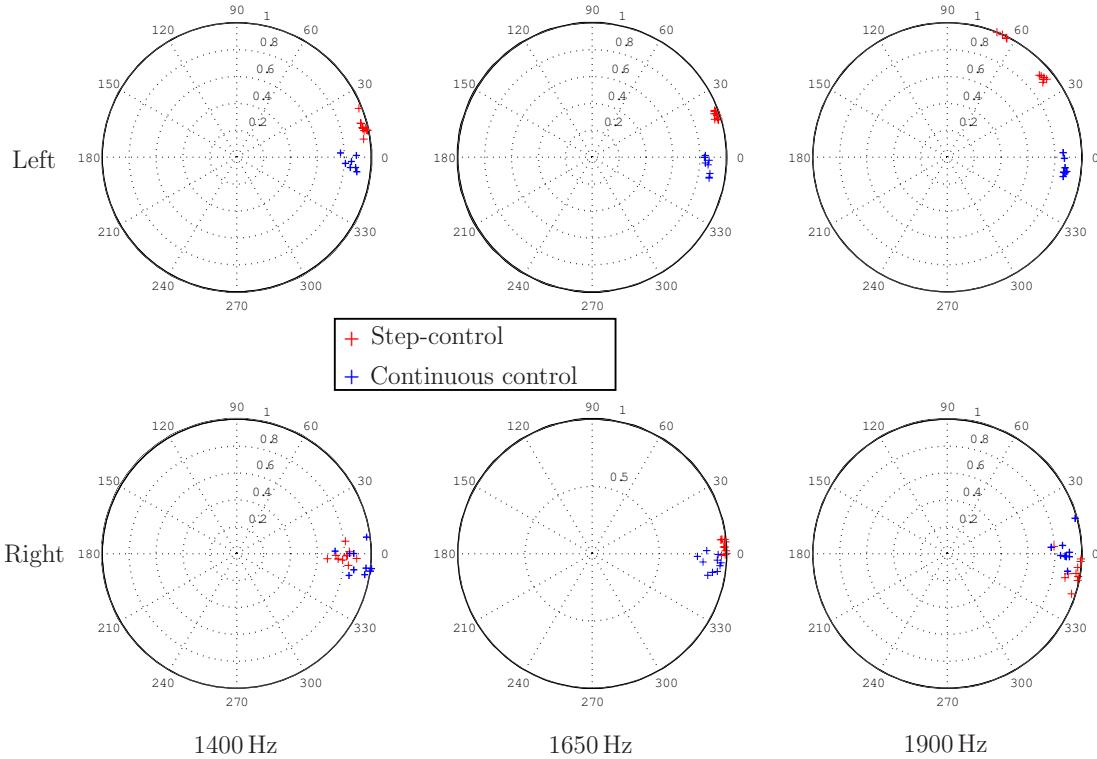
The performance statistics for the case when the initial position of the robot is to the left side of the loudspeaker appear to be worse than those when the initial position of the robot is to the right side. It was observed that the smooth wooden surface in the arena on which the trials were performed was more slippery in some areas than others and especially near the left initial position, and this resulted in the wheels slipping significantly during the first few seconds in each trial. When the robot was placed at the left initial position, the loudspeaker was to the left of the robot, and consequently the speed of the right motor was considerably greater than that of the left motor during the first few seconds in the trials. There was a relatively sharp bend observed in the trajectories on the left side when the robot subsequently hit a section of the floor with greater friction and turned sharply. As a consequence, during the first few seconds into a trial the orientation of the robot did not vary as smoothly as it did when the robot was on the right side of the loudspeaker. Figure 3.10 depicts this effect as an example in the trajectories for left and right sides at a sound frequency of 1600 Hz.

### 3.4.3.1 Trajectory Directness

Figure 3.11 depicts the trajectory directness polar plots for the step and continuous control strategies for three frequencies, namely 1400 Hz, 1650 Hz and 1900 Hz, chosen from the lower end, the middle and the upper end of the 1000 Hz–2200 Hz frequency range. The data for the step-control strategy was made available by Zhang et al.. The complete set of directness plots are presented in Appendix A.



**Figure 3.10:** Sample robot trajectories for 1600 Hz tone from the loudspeaker.



**Figure 3.11:** Trajectory directness in step-control vs. continuous control

It is evident that the average headings of the trajectories for both left and right sides lie in the vicinity of the ideal trajectory heading. The trajectories on the right side exhibit a slightly greater spread than those on the left side. This is a consequence of

the relatively sharp bend seen during the trials as mentioned above. The orientation of the robot with respect to the loudspeaker in the left initial position was such that the loudspeaker was to the left of the robot. The robot's sudden transition from a low friction surface to a high friction one early into the trials caused it to be oriented such that the loudspeaker was now on the right side of the robot. This caused the robot to take corrective action towards the right side. Due to this oscillatory effect, when computing the trajectory directness, the average heading thus obtained was quite close to the ideal heading. On the other hand, when the robot started from the right initial position, it was oriented such that the speaker was to its right. But since the arena surface was more uniform on that side, the robot's orientation varied smoothly during the trials, such that the speaker was almost always to the right of the robot. Thus the trajectories on the right side were always curving to the right. Therefore, there was more variation in the average headings computed for these trajectories.

Comparing the step-control strategy with the continuous control strategy from the trajectory directness' perspective, although it is evident that the continuous control strategy performs better than the step-control strategy for the left side, this is due to the sharp bend in the trajectories on the left side. For the right side, there is little difference between the performance of both the strategies.

### 3.4.3.2 Statistical Analysis

In order to determine whether there is any statistical difference between the left and right sets of trajectories, their directness was examined via the Mann-Whitney U test [79] to determine whether the sets belonged to populations with equal distributions. This is a non-parametric statistical hypothesis test to determine whether two samples are from two populations, say A and B with sizes  $n_A$  and  $n_B$  respectively, that have equal distributions. It involves the calculation of the U statistic and the probability  $\rho$  of observing a U value at least as extreme as the one calculated, given that the null hypothesis is true. If  $\rho < 0.05$ , then the null hypothesis is rejected at the 5% statistical significance level. Since the directness values as computed earlier do not have a fixed interval spacing but can be ordered in terms of their relative magnitude, they are ordinal data, and the Mann-Whitney U test is a suitable test for such data.

Here, the two populations being examined were the left and right cartesian distances between the directness of the ideal trajectory given by (1,0) and that of the individual trajectories as given by (3.3). For each initial position of the robot (left and right) the difference between the directness of the 10 trajectories and that of the ideal trajectory

given by the point (1,0) were calculated for each of the three frequencies considered — 1400 Hz, 1650 Hz and 1900 Hz. These three frequencies represent the start, middle and end respectively of the frequency range of 1400 Hz–1900 Hz where we see consistently good performance. This resulted in 6 populations, each with 10 samples. The null hypothesis was that for any given sound frequency, the population distributions for the left and right trajectories are equal, and the alternative hypothesis was that the two population distributions are not equal.

Table 3.1 lists the U statistics and the  $\rho$ -values calculated for the three sound frequencies. It can be seen that for 1400 Hz and 1650 Hz, the  $\rho$ -value is greater than 0.05. This implies that the probability of observing a U statistic at least as extreme as the calculated value is greater than 5%. Consulting the table for critical values of U which denote the upper and lower boundaries of U at the 5% significance level, i.e. the value of U which has a 5% probability of being observed in a non-directional test, it was noted that the upper and lower critical values of U at the 5% significance level were 77 and 23 respectively. Since the calculated U values for the left and right trajectories for the 1400 Hz and 1650 Hz frequencies were within the upper and lower critical bounds of U, it means that the probability of observing such U values is greater than 0.05 as confirmed by the  $\rho$ -values of 0.217562 and 0.970512 respectively, and thus the left and right trajectories are not significantly different for a non-directional test and the null hypothesis that the two populations have equal distributions cannot be rejected at the 5% significance level. For the 1900 Hz frequency, the calculated U values coincided with the upper and lower critical bounds of U and the  $\rho$ -value was 0.043257, which is just beyond the significance level of 0.05. This implied that the two populations were significantly different just beyond the 0.05 level, or that the difference was marginally significant for a non-directional test.

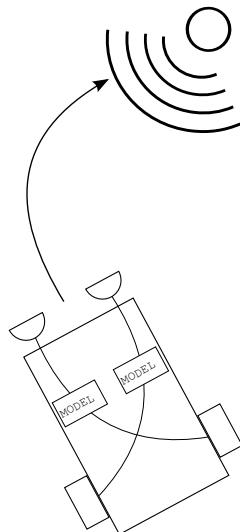
**Table 3.1:** Mann-Whitney U values.

Frequency (Hz)	$U_{left}$	$U_{right}$	$\rho$
1400	67	33	0.217562
1650	49	51	0.970512
1900	23	77	0.043257

## 3.5 A Braitenberg Lizard with Decision Models

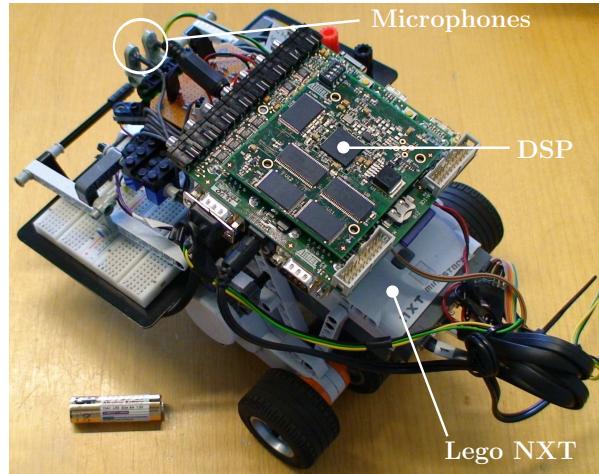
### 3.5.1 The Robotic Platform

Once again, the Braitenberg vehicle from Fig. 3.5(b) was implemented. Two kinds of decision models, ***MAX*** and ***AVG***, were individually inserted in turn into both the left and right sensorimotor couplings between the sensors and motors, while still maintaining their independence from each other. Figure 3.12 depicts the Braitenberg vehicles with the models embedded in the sensorimotor couplings. In this manner two different kinds of vehicles were designed, one with the ***MAX*** model and the other with the ***AVG*** model. The ***MAX*** model collected as input the signals produced by the corresponding auditory model output over a given period of time as determined by the control loop frequency and produced as output the *maximum* of these values, which was passed on to the corresponding motor. The ***AVG*** model also collects as input the signals produced by the corresponding auditory model output over a given period of time as determined by the control loop frequency and outputs the *average* of these values, which is passed on to the corresponding motor. This process is explained in more detail in the following paragraphs. These particular decision models are “continuous” in contrast to the discrete *if-then* rules used by [138], and they represent a lower level decision-making process as compared to higher level decision-making processes as represented by the discrete decision models.



**Figure 3.12:** Braitenberg vehicle with embedded decision models.

Figure 3.13 depicts the physical mobile robot. It was identical to the previous robot except for a few structural adjustments to provide additional structural rigidity.



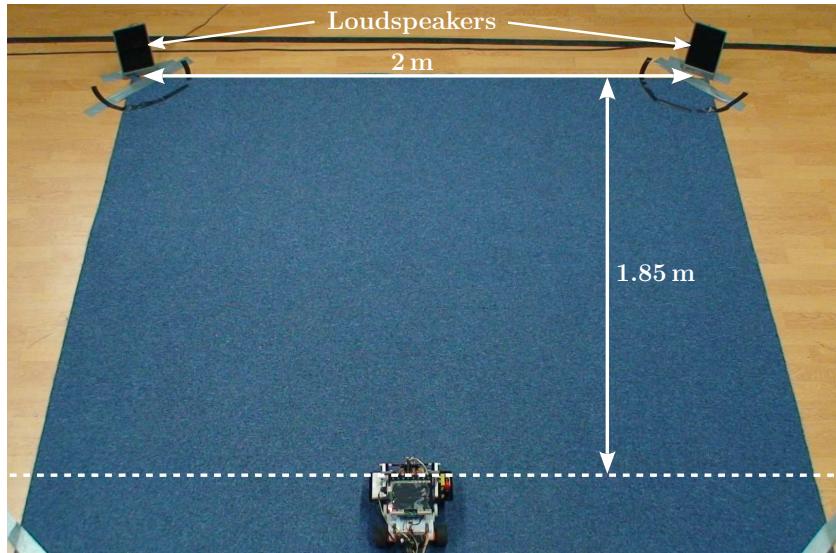
**Figure 3.13:** The Lego NXT based mobile robot.

As before, the voltage signals from the microphones were pre-amplified and fed into the DSP, where they were processed by the lizard ear model and left and right output powers were computed in dB. These values were then fed into the decision model as described above. The outputs of the decision models were then scaled to lie within the range of the argument to the motor drive speed command of the NXT brick (0–100), and serially transmitted to it identically to the previous implementation. Once again, the range 0–100 signified that the power applied to the motor lay between 0 % (no power applied, motor is not running) and 100 % (full power applied, motor is running at full speed). The scaling procedure was identical to that in the previous implementation.

### 3.5.2 Experimental Setup

The experimental setup consisted of two loudspeakers (model MD 42491 from Medion AG) placed 2 m apart, that served as monotone sound sources. The goal of the robot was to localize the active loudspeaker. The robot started from a fixed initial position 1.85 m in front of the mid-point between the loudspeakers, oriented straight ahead, and was allowed to move autonomously within the arena boundaries until it a) hit the loudspeaker, b) moved behind the loudspeaker or c) travelled outside the arena boundaries. During the movement of the robot, its position in terms of  $x$  and  $y$  coordinates was recorded via an infrared overhead camera system tracking an infrared LED (Light Emitting Diode)

mounted on top of the robot. At the same time, a digital video camera was also used to record the entire experimental proceedings. Figure 3.14 depicts the experimental arena.



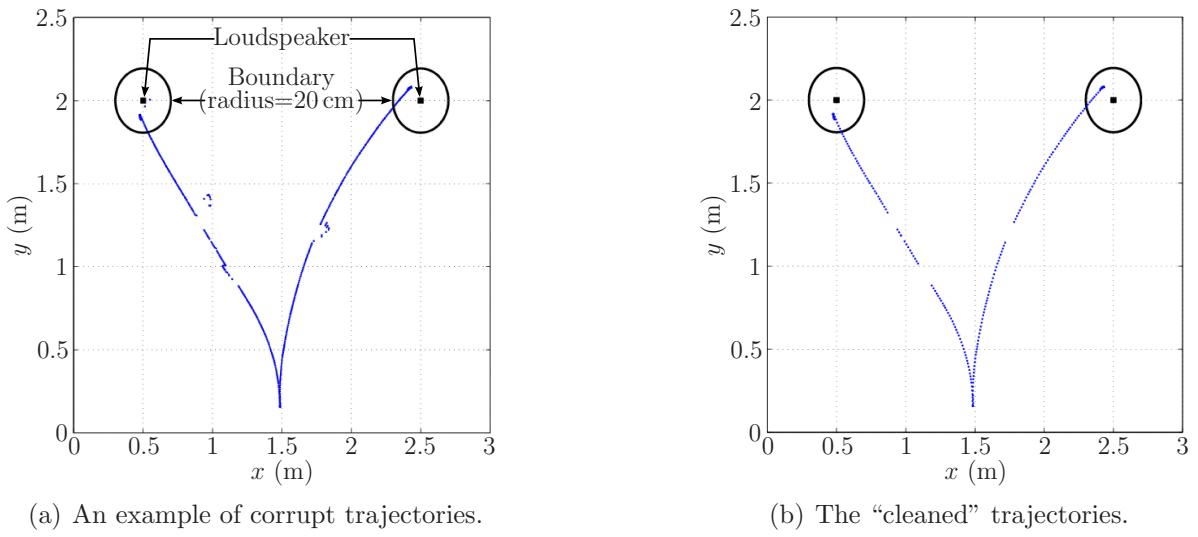
**Figure 3.14:** The experimental arena.

The experiments were performed for both the decision models individually. For each model, 4 different values, namely, 1 s, 333.333 ms, 100 ms and 33.333 ms were individually employed as the total delay between consecutive execution of the control loop, taking into account the approximately 6 ms delay incurred due to the reprogramming of the DSP core. Alternatively, four different control loop frequencies, namely,  $\frac{1}{1\text{s}} \approx 1\text{Hz}$ ,  $\frac{1}{333.333\text{ms}} \approx 3\text{Hz}$ ,  $\frac{1}{100\text{ms}} \approx 10\text{Hz}$  and  $\frac{1}{33.333\text{ms}} \approx 30\text{Hz}$  were employed. These particular values were chosen because they represented a linear increase in the robot's sensor-processing rate on a logarithmic scale. For each control loop frequency, 5 sets of experiments were performed in the frequency range of 1150 Hz–2150 Hz in steps of 250 Hz (i.e. one set per sound frequency), thereby covering the relevant frequency range of 1000 Hz–2200 Hz of the circuit model. In each set, one of the loudspeakers continuously emitted a tone of the selected frequency, and the path of the robot was tracked from the starting point until the one of the three terminating conditions mentioned above was met. This was done 10 times in total per set, with 5 trials with the left loudspeaker activated and 5 trials with the right loudspeaker activated. The loudspeakers were alternately activated for each set, starting with the left loudspeaker. The final outcome of each trial was classified as either (a) a *hit*, i.e. the robot hits the loudspeaker, (b) a *near hit*, i.e. the robot passes within a circle of radius 20 cm around the loudspeaker's mid-point or (c) a *miss*, i.e. the robot travels outside the arena boundaries without satisfying condition

(b). For each decision model and each control loop frequency, the total number of *hits*, *near hits* and *misses* were determined for the 10 trials, for all 5 sound frequencies used.

### 3.5.3 Results and Discussion

During the subsequent analysis of the logged trajectory data, it was discovered that there were points in the data which indicated that the robot had been suddenly displaced laterally from its current position. Comparing the individual video recordings of the trajectories followed by the robot during the actual trials with the trajectories recorded by the camera system, it was possible to determine whether the actual trajectory followed by the robot matched the recorded trajectory. It was discovered that the sudden deviations in the data points were typically greater in magnitude than the width of the body of the robot did not actually occur in the actual robot trials. Thus, video evidence of the robotic trials confirmed that the erroneous data points in the trajectories did not correspond to the actual trajectories followed by the robot. This led to the conclusion that these displaced data points were corrupt data. Such data points were eliminated from the trajectory data. This consequently created gaps in the trajectories that were not filled in with interpolated data because the length of the gaps did not typically exceed one body length of the robot. This elimination resulted in “clean” trajectories. Figure 3.15 depicts as an example the corrupt trajectory data and its “clean” version.

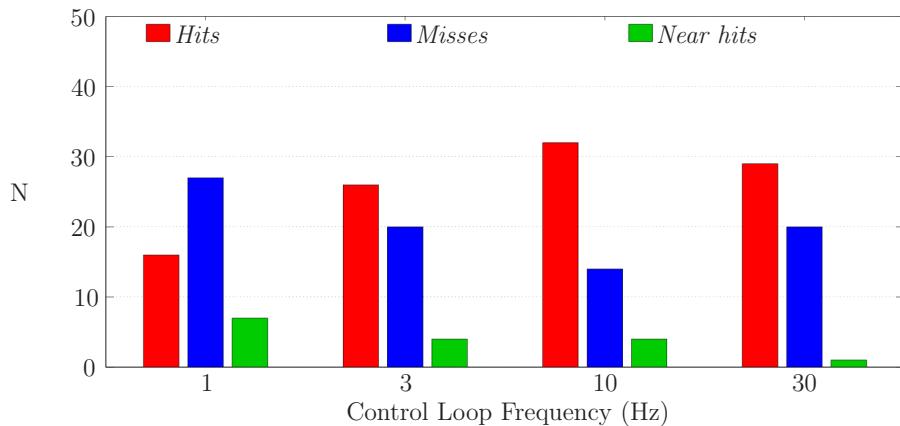


(a) An example of corrupt trajectories.

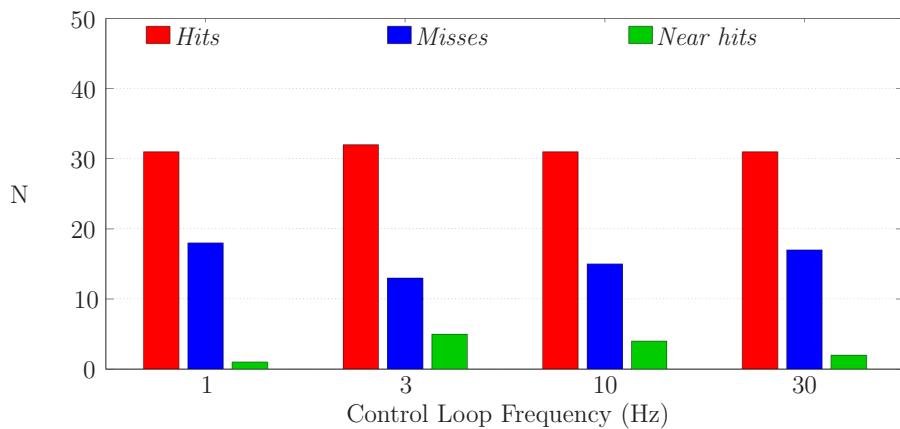
(b) The “cleaned” trajectories.

**Figure 3.15:** Example trajectories depicting the corrupt data and the “cleaned” version.

Figure 3.16 depicts the phonotactic performance in terms of the number of *hits*, *near hits* and *misses* for the **MAX** and **AVG** models. For the **MAX** decision model, it is evident from Fig. 3.16(a) that the performance of the robot, in terms of the number of *hits*, *near hits* and *misses*, improves as the frequency increases and reaches a peak at about 10 Hz. Since the **MAX** model's output is the maximum of the given set of inputs it processes in every execution of the control loop, the rate at which it produces an output is directly proportional to the control loop frequency. When the control loop frequency is varied, the reaction time of the robot to the external sound stimulus from the loudspeaker tends to vary. Higher control loop frequencies signify a lower reaction time and thus a faster response rate.



(a) Phonotactic performance of the **MAX** model.



(b) Phonotactic performance of the **AVG** model.

**Figure 3.16:** The phonotactic performance in terms of the number of *hits*, *near hits* and *misses* versus the control loop frequency for the two decision models.

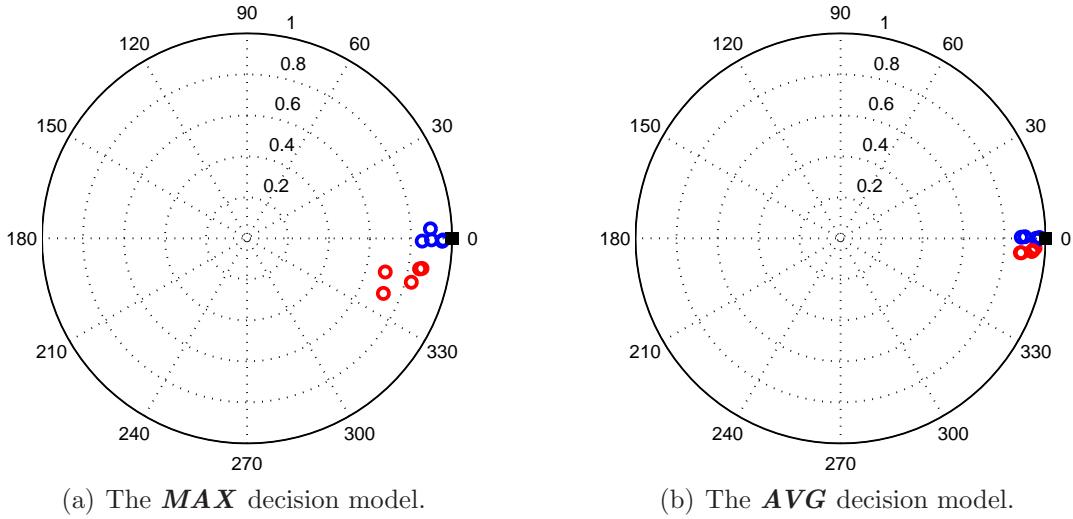
The **MAX** model is sensitive to variations in the control loop frequency. At low frequencies, once the model has generated an output, the robot behaves accordingly for a longer period of time. Therefore, if the magnitude of the model's output is large or small, the turning speed of the robot is respectively faster or slower, but in both cases the turn lasts for a relatively longer period of time than when the control loop frequency is high. In this case the robot's turns last for a relatively shorter period of time. This means that at high control loop frequencies, any overshoots in the robot's trajectory are quickly corrected for, while at low control loop frequencies the robot takes a longer time to take corrective action. Therefore, at high frequencies the robot's turns are tighter than those at low frequencies and it is able to localize the sound much better at high control loop frequencies than at low ones. However, at a control loop frequency of 30 Hz the performance drops slightly. Since there is no noise suppression at either the input or output of the ADCs, it is plausible that at relatively high control loop frequencies the noise component in the auditory signal has a relatively more pronounced effect, resulting in a relatively greater magnitude of error in the calculation of the auditory model's outputs. Over time, this may overcome the corrective actions taken by the robot and result in negative localization performance.

From Fig. 3.16(b), it is evident that for the **AVG** decision model, the performance of the robot is consistent across all control loop frequencies. Since the model's output is the average of the set of inputs it processes in every execution of the control loop, the magnitude of its output for any given frequency is not as large as that of the **MAX** model. This means that for any given frequency, the robot's turns are generally slower as compared to those with the **MAX** model. At lower control loop frequencies, even though the response rate of the robot is slower and the turns last for a longer period of time, the **AVG** model compensates for this with relatively smaller outputs. Consequently, the turns of the robot are relatively slower. This means that the overshoots in the robot's trajectory are smaller in magnitude. At higher control loop frequencies, the robot's response rate is faster and the turns last for a shorter period of time. This also results in overshoots of smaller magnitude, which are then corrected faster.

### 3.5.3.1 Trajectory Directness

In order to determine the performance of the robot in terms of its trajectory, the “directness” statistic given by (3.3) was computed. The trajectories were divided into  $n = 10$  vectors and the average heading was calculated as in Sect. 3.4.2. As before, the polar co-ordinates  $(1, 0)$  represent the ideal trajectory with average vector of length 1 and

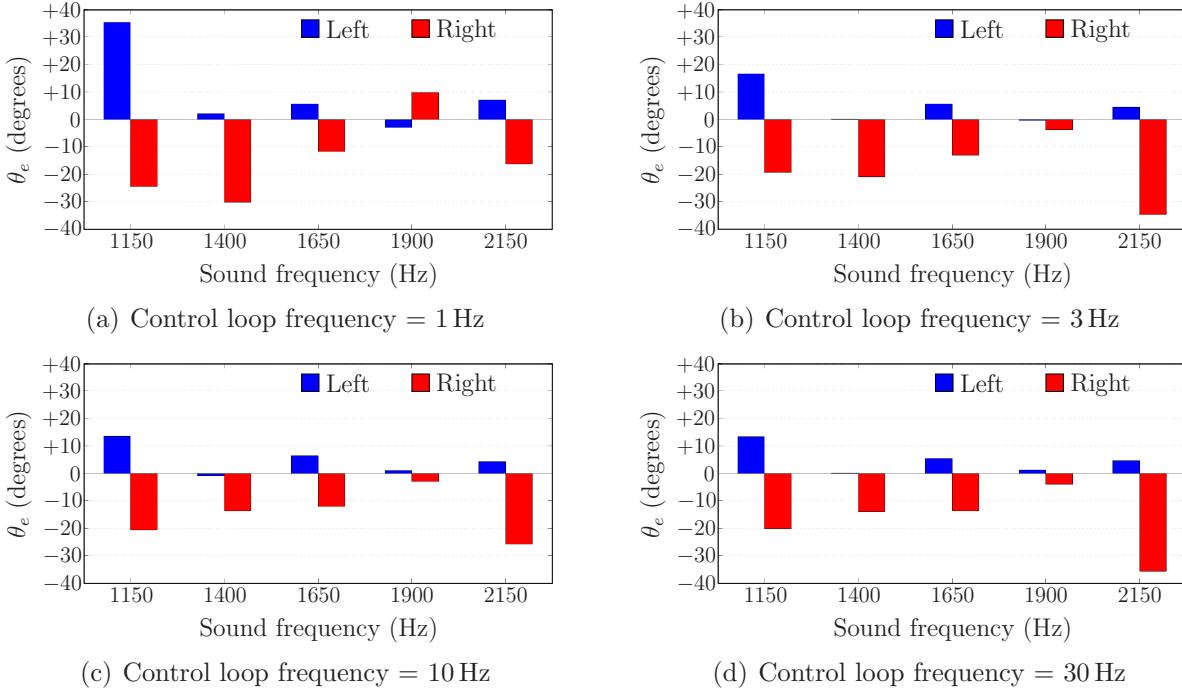
direction  $0^\circ$ , i.e. the robot moves in a perfectly straight line from the starting point to the loudspeaker. Figure 3.17 shows as an example the directness plots of the ideal and actual average trajectory headings for the **AVG** and **MAX** decision models at a 30 Hz control loop frequency in response to a 1400 Hz sound frequency. The complete set of directness plots are presented in Appendix B.



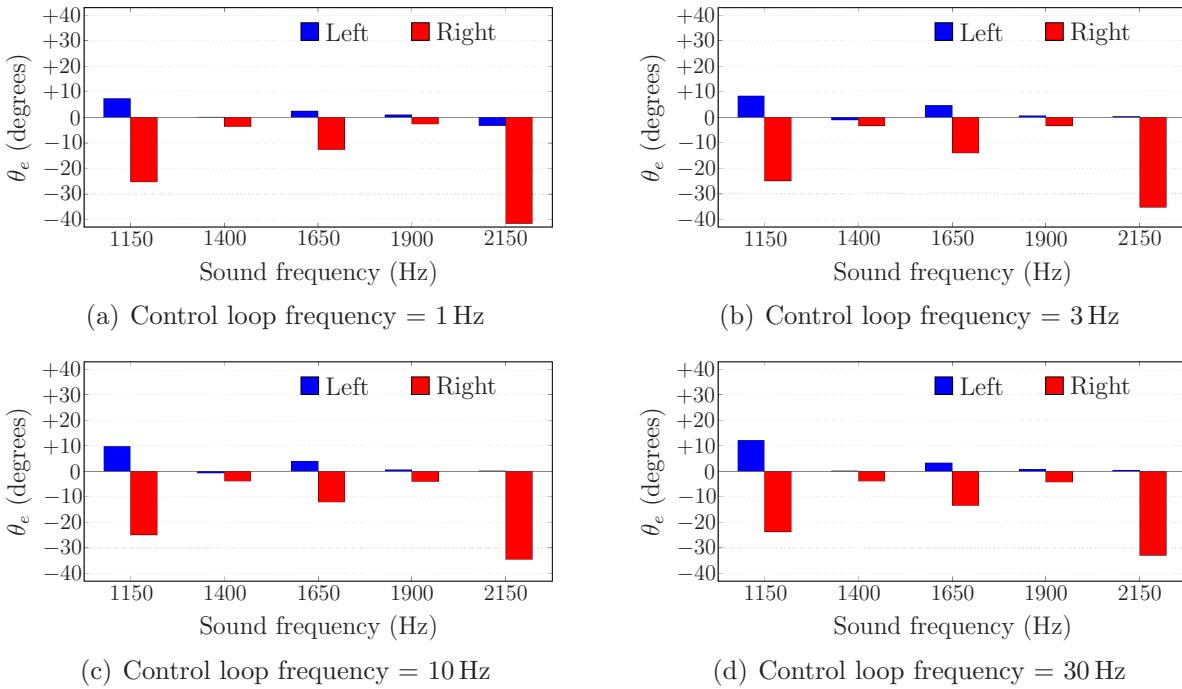
**Figure 3.17:** Example directness plots for the **AVG** and **MAX** decision models at a 30 Hz control loop frequency in response to a 1400 Hz sound frequency.

Fig. 3.18 and 3.19 depict the angular error  $\theta_e$  in degrees between the average headings and the ideal heading, for varying sound and control loop frequencies, for the two decision models **MAX** and **AVG** respectively. Since there are 5 trajectories per sound frequency,  $\theta_e$  for each sound frequency is calculated by taking the mean of the average headings calculated for each of the 5 trajectories via the directional statistics approach [80] and then computing the difference between this mean and the ideal heading. The  $0^\circ$  line represents the ideal trajectory.

From Fig. 3.18, it is evident that for the **MAX** model, the heading error  $\theta_e$  is lower in the 1400 Hz–1900 Hz range and higher at the upper and lower extremes of 1150 Hz and 2150 Hz. There is an overall decrease in  $\theta_e$  as the control loop frequencies increase. This is consistent with the reasoning presented in the previous section that at lower frequencies, if the magnitude of the model’s output is large or small the turning speed of the robot is respectively faster or slower, but in both cases the turn lasts for a relatively longer period of time than when the control loop frequency is higher. For higher control loop frequencies the robot’s turns last for a relatively shorter period of time. Therefore, at high control loop frequencies any overshoots in the robot’s trajectory are quickly



**Figure 3.18:** Trajectory directness for the **MAX** model.



**Figure 3.19:** Trajectory directness for the **AVG** model.

corrected for, while at low control loop frequencies the robot takes a longer time to take corrective action. Thus, at high frequencies the robot's turns are tighter than those at

low frequencies and it is able to localize the sound much better at high control loop frequencies than at low ones. In comparison, for the **AVG** model, it is evident from Fig. 3.19 that the heading error  $\theta_e$  is also lower in the 1400 Hz–1900 Hz range than at 1150 Hz and 2150 Hz, for all the control loop frequencies. However, overall  $\theta_e$  is lower for the **AVG** model than for the **MAX** model. Again, this is consistent with the reasoning that at lower control loop frequencies, even though the response rate of the robot is slower and the turns last for a longer period of time, the **AVG** model compensates for this with relatively smaller outputs. Therefore, overshoots in the robot’s trajectory are smaller in magnitude. At higher control loop frequencies, the robot’s response rate is faster and the turns last for a shorter period of time. This results in overshoots of even smaller magnitude, which are then corrected faster.

### 3.5.3.2 Statistical Analysis

In order to determine whether the performance of a decision model is dependent on the control loop frequencies or not, Pearson’s chi-square ( $\chi^2$ ) test [99] was employed. This is a statistical test for independence that can be used for categorical data to determine whether the categorical variables used have a significant association between them or not. It involves the calculation of the chi-square statistic  $\chi^2$  and the probability  $\rho$  of observing a value for  $\chi^2$  that is at least as extreme as the one calculated with degrees of freedom  $d$ , given that the null hypothesis is true. If  $\rho < 0.05$ , then the null hypothesis is rejected at the 5% statistical significance level.

The null hypothesis formulated here was that the effect of the type of the decision model is independent of the control loop frequency, and the alternative hypothesis was that the effect of the type of the decision model is not independent from the control loop frequency. The categorical variables were the type of decision model, having two values — **MAX** and **AVG**, and the control loop frequency, having four values — 1 Hz, 3 Hz, 10 Hz and 30 Hz. Given these variables, a contingency table was constructed (refer to Table 3.2) with each cell containing the total number of *hits* for a given control loop frequency and a given decision model. The value of  $d$  was calculated as  $(r-1) \cdot (c-1) = (2-1) \cdot (4-1) = 3$ . Table 3.3 lists the values of  $\chi^2$ ,  $d$  and  $\rho$ .

Consulting the table for critical values of  $\chi^2$  [34], i.e. the probability that the observed value of  $\chi^2$  is at least as extreme in a  $\chi^2$  distribution, it was noted that the critical value of  $\chi^2$  for  $d = 3$  at the 5% significance level is 7.815. Since the observed value of  $\chi^2 = 3.4$  was less than the critical value of  $\chi^2 = 7.815$ , it meant that the probability of observing such a value for  $\chi^2$  was greater than 0.05, as confirmed by a calculated  $\rho$ -value of 0.334.

Therefore, the null hypothesis that the type of the decision model is independent from the control loop frequency cannot be rejected at the 5% significance level.

**Table 3.2:** Total number of *hits*.

	1 Hz	3 Hz	10 Hz	30 Hz
<b>MAX</b>	16	26	32	29
<b>AVG</b>	31	32	31	31

**Table 3.3:** Results of the Chi-square test.

$\chi^2$	$d$	$\rho$
3.4	3	0.334

## 3.6 Summary and Conclusions

In this chapter the peripheral auditory model was incorporated into Braitenberg vehicle-like mobile robot, with sensorimotor couplings established between the auditory model and the actuators of the robot. By coupling the ipsilateral output of the auditory model with the contralateral actuator and vice versa, phonotaxis behaviour could be elicited. The phonotactic performance of the Braitenberg “lizard” thus constructed was evaluated in two independent experiments, employing two different Braitenberg vehicles — one employing the basic cross-coupling with the magnitude of the auditory model’s outputs directly driving the actuators, and the other with decision models embedded into the sensorimotor couplings. First, the performance of the basic Braitenberg vehicle was evaluated in direct comparison to the step-control strategy employed by Zhang et al. by repeating their phonotaxis experiment. The robot’s trajectories on the left and the right side were also examined statistically for any significant differences via the Mann-Whitney U test. Second, the performance of the Braitenberg vehicle with decision models was evaluated for two different types of decision models and four different control loop frequencies. The dependence of the type of the decision model on the control loop frequency was statistically examined as well via Pearson’s  $\chi^2$  test of independence.

From the results obtained with the basic Braitenberg vehicle, it is evident that the performance of the continuous control strategy without decision models is comparable

to the step-control strategy with decision models. The phonotactic performance of the Braitenberg vehicle in terms of the number of *hits*, *near hits* and *misses* is quite good in the 1400 Hz–1900 Hz range, and is comparable for both the strategies. In terms of the trajectory directness, again the Braitenberg vehicle performs well in the 1400 Hz–1900 Hz range, with the trajectories being quite direct. From this perspective the performance of the two strategies is comparable as well. This indicates that the auditory model can be used in itself to localize sound, without employing a decision model to re-interpret its outputs. Statistically, it was determined that there was probably no significant difference between the left and right side trajectories for the three sound frequencies examined, from the lower end, the middle and upper end of the 1400 Hz–1900 Hz range.

For the Braitenberg vehicle with embedded decision models, it is evident that the phonotactic performance in terms of the number of *hits*, *near hits* and *misses* is quite good over the 1400 Hz–1900 Hz range. For the **MAX** decision model which is sensitive to variation in the control loop frequency, the performance improves as the control loop frequencies increase due to the stronger compensating effect at higher frequencies, which suppresses the overshoots of the robot. However, at the highest control loop frequency the negative effect of the noise in the auditory signals may overcome the compensatory effect over time, leading to degraded localization performance. For the **AVG** decision model, the performance is better than the **MAX** decision model since the former is less sensitive to variations in the control loop frequency. In terms of the trajectory directness, the **AVG** decision model performs better than the **MAX** decision model, because its averaging effect generates an output of smaller magnitude than the output of the **MAX** decision model, resulting in smaller overshoots for which a smaller corrective action is required. Statistically, it was determined that the effect of the type of the decision models was probably dependent on the control loop frequency, although not significantly.

# Chapter 4

## Modifying Directionality Through Auditory System Scaling

### 4.1 Introduction

As previously described in Sect. 1.1, the peripheral auditory system of lizards has achieved high directionality due to the acoustical inter-coupling of the ipsilateral and contralateral tympani via the Eustachian tube, providing an internal pathway for the transmission of sound. This transmission is strongly influenced by the properties of such internal pathways, as well as by the head size and is therefore frequency dependent. Since the properties of the internal pathways presumably scale with the size of the lizard it is important to understand how physical scaling affects the directional characteristics of the auditory system.

From the auditory model's perspective, it can be hypothesised that any joint scaling of the ear separation and the auditory model parameters should result in a corresponding scaling in the frequency response as well. Provided that the scaling of the separation between the “ears” is consistent with the scaling of the parameters of the corresponding auditory model, a scaled version of a model may exhibit a phonotactic performance equivalent to that of the original unmodified model, albeit at a different frequency range from that of the original model. On the other hand, if the scaling of the separation between the “ears” is inconsistent with the scaling of the parameters of the corresponding auditory model or vice versa, then the scaled version of the model may exhibit a phonotactic performance which is inequivalent to that of the original model, and degraded in terms of the directional acuity. In this chapter, this hypothesis is experimentally investigated by scaling the auditory model parameters to match three different ear separations

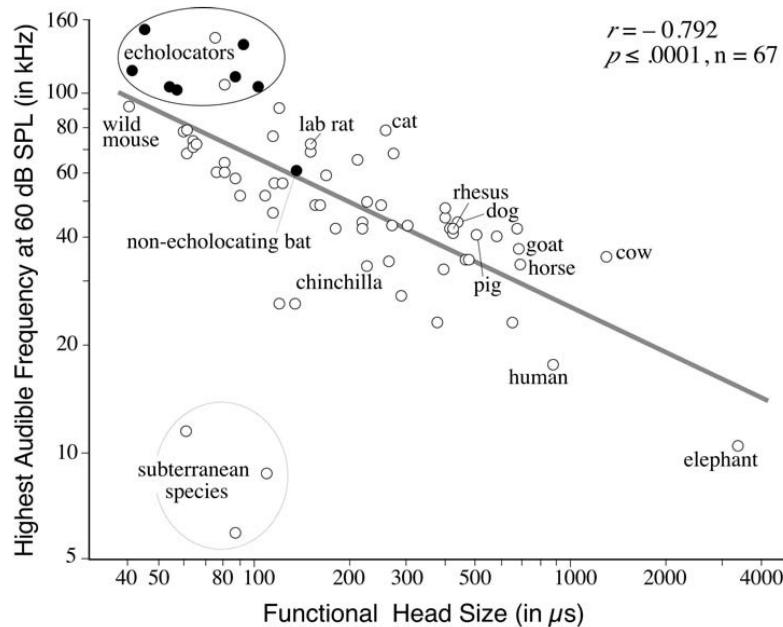
and consequently three different peak sound frequencies, i.e. frequencies around which the model has peak response, and then evaluating the phonotactic performance through a robotic implementation for various degrees of ear separation.

The work presented in this chapter covers and elaborates further upon the work reported by the author in [115].

## 4.2 Auditory System Scaling

There have been extensive studies on the evolution of hearing mechanisms [83] and sound localization abilities of a number of species of animals [44]. Early results however, have been obtained via behavioural experiments and therefore report the behavioural response of the animal, produced by neural processes to auditory stimuli. Interestingly, these studies have reported that it is rather the width of the field of best *vision* and not the functional head size, i.e. the time taken by the sound waves to travel from one ear to the other, which is a good predictor of sound localization acuity in animals [47]. Later, techniques such as laser doppler vibrometry [65, 123] have been employed with animals to study the underlying biophysical mechanisms of their auditory systems, and have allowed researchers to build accurate artificial models of the same. The focus there is on what frequencies can animals *detect* and *localize* rather than their *behavioural response* to them. Traditional explanations of sound localization rely on physical factors such as the head size and the acoustics involved. The two physical cues available for sound localization are the interaural time difference (ITD) and interaural level or intensity difference (ILD or IID). Both of these cues are higher in magnitude for animals with larger heads and consequently greater separation between the ears, because a) sound requires more time to travel between the ears, thereby generating a larger ITD cue, and b) the larger the head, the lower the frequencies that it can block, thereby creating a sound shadow, or a sound intensity difference between the ears and thus a larger ILD cue. This would suggest that having smaller heads implies the ability to hear higher frequencies, and having larger heads implies the ability to hear lower frequencies. This has been confirmed for mammalian species [45]. Figure 4.1 depicts the relationship between functional head size, which is directly proportional to physical head size and consequently to the ear separation, and the highest audible frequency for different mammalian species. However, there have only been a few morphological and physiological studies on the effect of body size on the auditory sensitivity in gekkonoid lizards that have reported that with increase in body size, the auditory sensitivity improves while the octave band-

width of the frequency response of the middle ear decreases [133, 132, 134]. However, to the best of the author's knowledge, there have not been any studies reported in the literature that have investigated the correlation between the head size and directional hearing in lizards.



**Figure 4.1:** The relationship between functional head size and the highest audible frequency for different mammalian species (taken from [45]).

As is evident from Fig. 4.1, among the mammalian species the smaller animals must hear higher frequencies in order to exploit the ILD cue because the smaller the size of the head, the higher the frequency necessary for blocking to generate ILD cues. However, in the case of gekkonoid lizards, the only cue available at the relevant frequencies is the ITD because the ILD cue is almost non-existent due to the diffraction effects described in Sect. 1.1. The auditory periphery of these reptiles has evolved in such a manner that they do not need to hear very high frequencies as do their mammalian counterparts that share similar head dimensions.

Since lizards exhibit different head sizes and consequently auditory peripheries of varying dimensions among the various subspecies, and due to the lack sufficient biological data on the correlation between the head size and auditory directionality, the auditory model can be utilized to investigate the effect of scaling on the directionality. The parameters of the auditory model are matched to an ear separation of approximately 13 mm, with strong directionality over 1000 Hz–2200 Hz with a peak at approximately 1500 Hz. This leads to a two-fold hypothesis. First, for a given auditory model matched to a given ear separation, if the ear separation is physically varied, the directionality

should a) degrade if the ear separation is less than the matched value and b) improve if the ear separation is greater than the matched value, subject to the constraint that the ear separation is less than half the wavelength. This is because, relative to the ITD cues available at the matched ear separation, a greater ear separation implies greater ITD cues while a smaller ear separation implies smaller ITD cues. Second, scaling the auditory model by a given factor should result in a proportional scaling of the frequency response of the model as well. This is because, by scaling the auditory model, it is matched to a proportionally different ear separation, and therefore to a proportionally different magnitude of the binaural ITD cues.

## 4.3 Design and Implementation

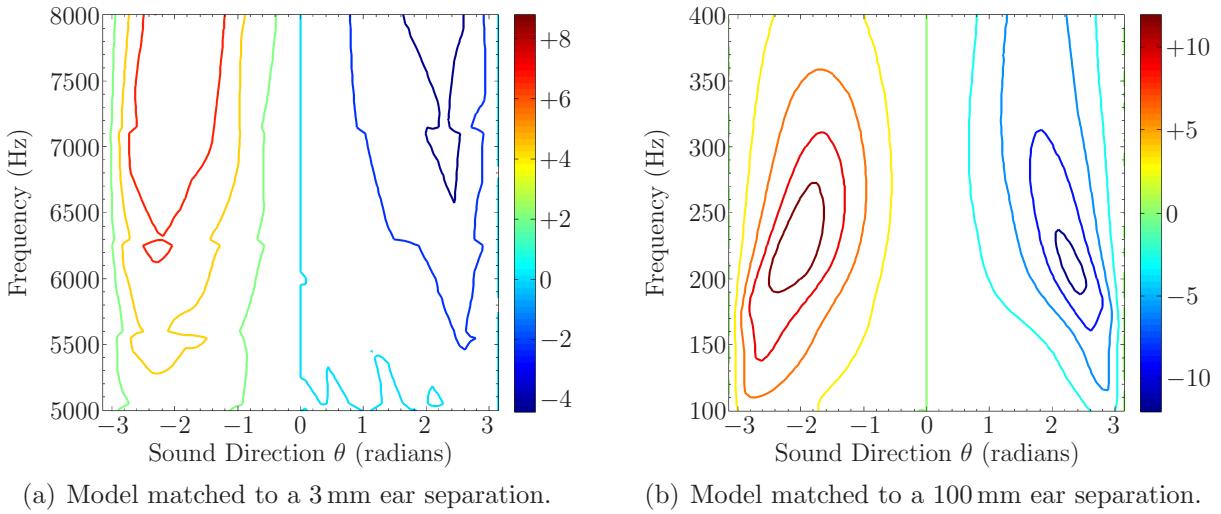
### 4.3.1 Scaled Auditory Models

The parameters of the default auditory model matched to an ear separation of 13 mm were scaled by a fixed factor to obtain its scaled versions. Two such scaled models were generated, one scaled up to match an ear separation of 100 mm by multiplying the default auditory model's parameters by a scaling factor of  $\frac{100 \text{ mm}}{13 \text{ mm}}$  and the other scaled down to match an ear separation of 3 mm by multiplying the default auditory model's parameters by a scaling factor of  $\frac{3 \text{ mm}}{13 \text{ mm}}$ . Figure 4.2 depicts the calculated response of the scaled auditory models in dB in their respective frequency ranges for an angular sound source displacement range of  $[-\pi, +\pi]$ .

### 4.3.2 Robotic Implementation

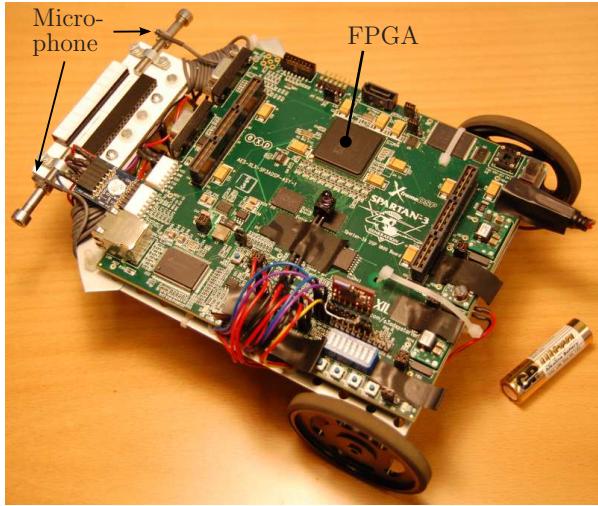
The physical robot consisted of an Xilinx XtremeDSP Starter Platform FPGA board, mounted on a mobile platform assembled with electro-mechanical parts from the Robotics Starter Kit from Digilent Inc.. The sole processing unit on the board was the Xilinx Spartan<sup>®</sup>-3A DSP 1800A, an FPGA chip optimized for DSP applications. The main reason for choosing this particular platform was its generous logic gate count (1.8 million gates). Figure 4.3 depicts the robotic platform.

Two standard 6 V DC motors controlled by two H-bridge motor drivers independently powered two rear wheels, while at the front, an omnidirectional ball caster was employed as a support. Power to the development board was supplied via a 3.7 V, 6 Ah lithium



**Figure 4.2:** The response of the scaled auditory models in dB.

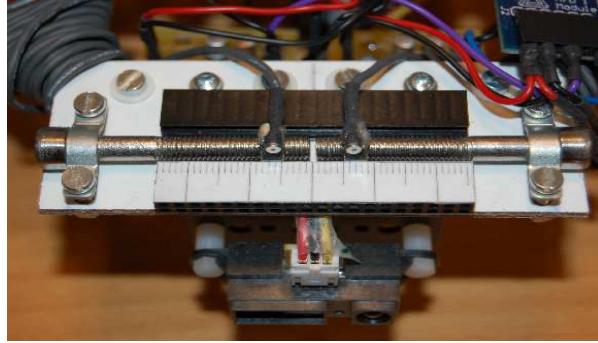
polymer battery, while power to the motors was supplied via a 3.7 V, 920 mAh lithium polymer battery.



**Figure 4.3:** The mobile robot.

The three different auditory model implementations written in the **VHSIC Hardware Description Language** (VHDL) hardware programming language (VHSIC stands for **V**ery-**H**igh-**S**peed **I**ntegrated **C**ircuits) were programmed into this chip. The VHDL design was synthesized (compiled) via Xilinx's proprietary software tool, namely the Xilinx Integrated Synthesis Environment (Xilinx ISE), into hardware-level binary "bit-stream" that contained all the information necessary to properly configure and program the logic into the FPGA chip. Two omnidirectional microphones (model FG-23329-P07

from Knowles Electronics) were mounted at the front of the robot on worm gears (refer to Fig. 4.4), allowing the separation between them to be varied with millimeter precision within the range of 3 mm–120 mm (refer to Fig. 4.4). The voltage signals from these microphones were pre-amplified, digitised and fed into the FPGA chip via a dual channel 12-bit analogue-to-digital converter, where they were processed by the auditory model and the left and right output powers were computed. These power values were fed into a decision model, which was essentially a set of *if–then* control rules identical to that used by Zhang et al. [138], namely, 1) *if* sound appears to come from left (i.e. when the left power value is greater than the right one) *then* turn left on-the-spot at full speed for a fixed period of time, 2) *if* sound appears to come from right (i.e. when the right power value is greater than the left one) *then* turn right on-the-spot at full speed for a fixed period of time, and 3) *if* sound appears to come from the front (i.e. when the left power value is equal to the right one) *then* go forward with full speed for a fixed period of time. The decision model controlled the motors based on this set of rules.

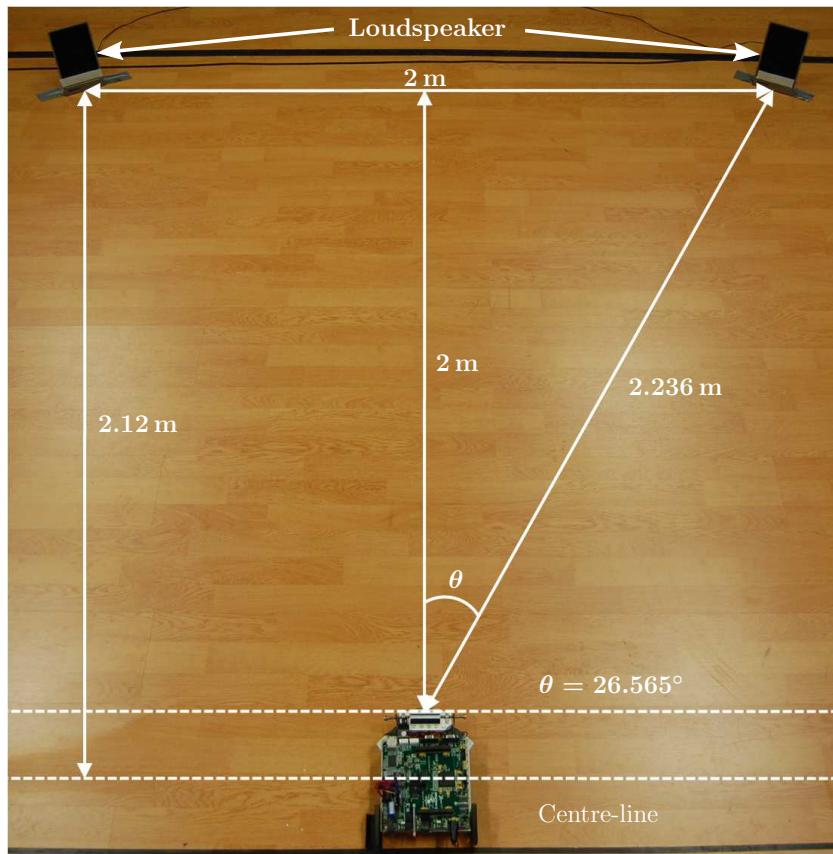


**Figure 4.4:** The microphone assembly.

## 4.4 Experimental Setup

Figure 4.5 depicts the experimental arena. Two loudspeakers (model MD 42491 from Medion AG) were affixed on the arena floor 2 m apart and 2 m from the mobile robot. In order to evaluate the phonotactic performance, three independent experiments were performed for three different separation lengths between the microphones, i.e., the ear separation.

First, the default auditory model matched to a 13 mm ear separation was utilized, so that the auditory model was matched to the default peak frequency of approximately 1500 Hz, i.e. the peak response was around 1500 Hz. Then the ear separation was set to



**Figure 4.5:** The experimental arena.

one of five fixed lengths, namely, 3 mm, 6.23 mm, 13 mm, 36 mm and 100 mm, labelled as *min*, *small*, *normal*, *large* and *max* respectively. These values represented a uniform change on a logarithmic scale in the ear separation with respect to a 13 mm ear separation. For each of the 5 ear separation values, 13 sets of experiments were performed in the frequency range of 1200 Hz–1800 Hz in steps of 50 Hz, within the relevant frequency range of 1000 Hz–2200 Hz of the auditory model. In each set, one of the loudspeakers emitted a continuous tone at the selected frequency and the goal of the mobile robot was to localize and approach the loudspeaker. The path of the robot was tracked from the starting point until one of the three finishing conditions was met. This was done 10 times in total per set, with 5 trials with the left loudspeaker activated and 5 trials with the right loudspeaker activated. To minimise any order effects, the experiments were randomised by a) alternating between the left and right sides for consecutive trials, b) randomising the order of selection of the frequencies by shuffling the frequency values and c) randomising the order of selection of the ear separations by shuffling the ear separation values as well. The final outcome of each trial was classified as either a *hit* if the robot hit the loudspeaker, a *near hit* if the robot passed within a circle of radius

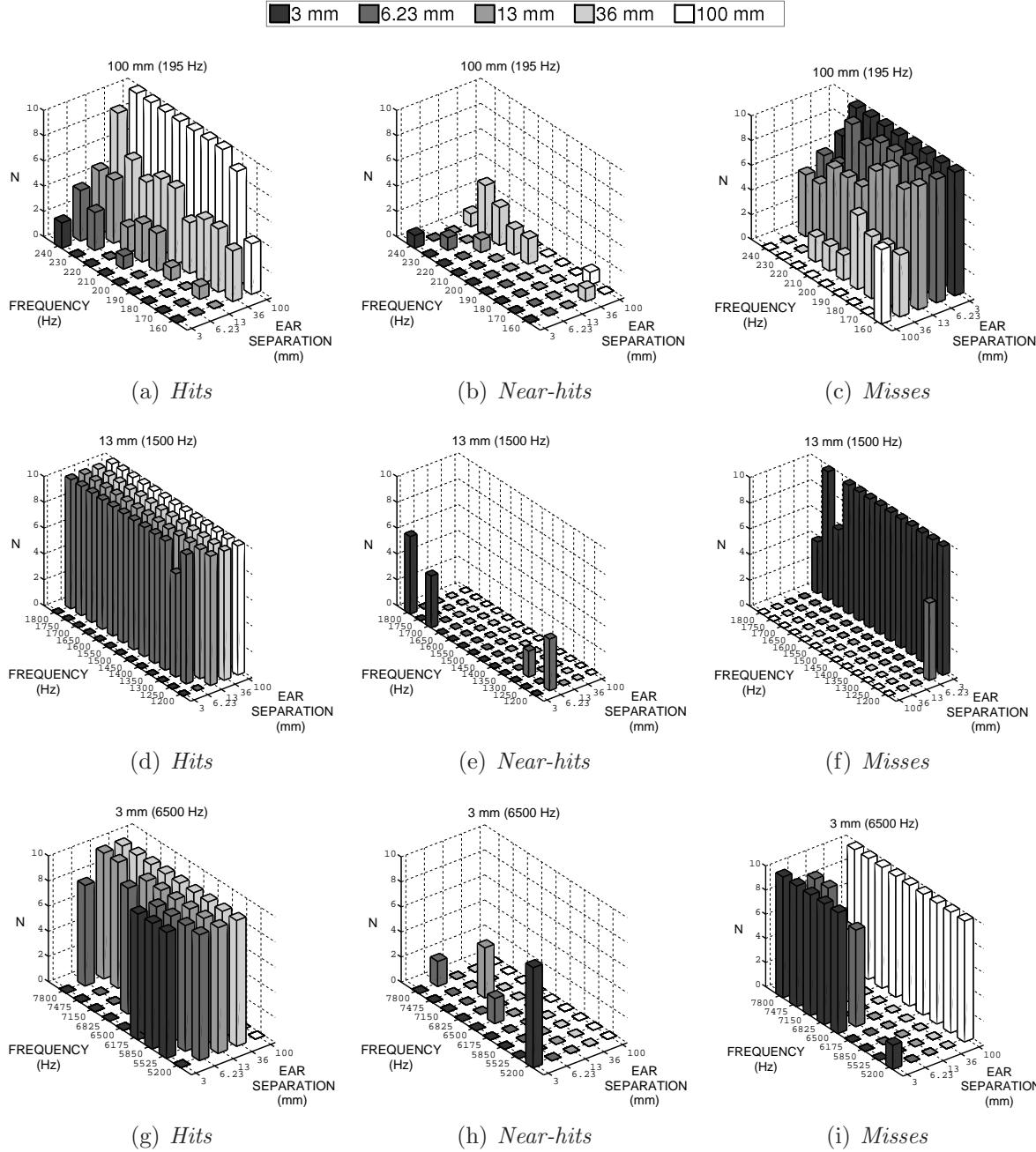
20 cm around the loudspeaker or a *miss* if the robot remained outside the circle while moving either behind the speaker or outside the arena boundaries. The total number of hits, near hits and misses were individually summed over the 10 trials, for each sound frequency and ear separation combination.

Next, the auditory model with its parameters scaled so as to match the *max* ear separation of 100 mm was utilized, so that the peak response of the model was around  $1500 \text{ Hz} \times \left( \frac{13 \text{ mm}}{100 \text{ mm}} \right) = 195 \text{ Hz} \approx 200 \text{ Hz}$ . For each of the 5 ear separations, the experiments described above were repeated for the frequency range of 160 Hz–240 Hz in steps of 10 Hz, resulting in 7 sets of experiments per ear separation.

Finally, the auditory model with its parameters scaled so as to match the *min* ear separation of 3 mm was utilized, so that the peak response of the model was around  $1500 \text{ Hz} \times \left( \frac{13 \text{ mm}}{3 \text{ mm}} \right) = 6500 \text{ Hz}$ . Once again, for each of the 5 ear separations, the experiments described above were repeated for the frequency range of 5200 Hz–7800 Hz in steps of 325 Hz, resulting again in 7 sets of experiments per ear separation.

## 4.5 Results and Discussion

Figure 4.6 depicts the phonotactic performance of the robot with the default auditory model and the two scaled auditory models, for varying ear separations in terms of the number of *hits*, *near-hits* and *misses*. It is evident that the robot exhibits strong directionality for all three scaling factors over their respective frequency ranges subject to the constraint that auditory model is matched to the actual ear separation, consistent with our hypothesis that scaling the auditory model parameters results in a directly proportional scaling in the frequency response, as long as the model is matched to the actual ear separation. In terms of the number of hits and misses, it is evident that in almost all cases the robot performs better when the actual ear separation is equal to or greater than the matched ear separation, and worse when the actual ear separation is less than the matched ear separation. When employing the auditory model scaled to match an ear separation of 100 mm, the phonotactic performance steadily decreases as the actual ear separation becomes smaller than 100 mm (refer to Fig. 4.6(a)). A decrease in the actual ear separation down from 100 mm results in ITD cues that are smaller than those the auditory model is tuned to, and this creates a mismatch between the actual ear separation and the auditory model. When the actual ear separation is identical to the matched value, the performance is excellent, with almost no misses.



**Figure 4.6:** Total number of *hits*, *near-hits* and *misses* by the robot for each combination of sound frequency and ear separation. The auditory model parameters are matched to 100 mm in Fig. (a)–(c), to 13 mm in Fig. (d)–(f) and to 3 mm in Fig. (g)–(i). Note that Fig. (c) and Fig. (f) have reversed ear separation axes for best visibility of all data points.

When the auditory model is scaled to an ear separation of 13 mm, the phonotactic performance decreases as the actual ear separation becomes smaller than 13 mm (refer to Fig. 4.6(d)). Again, a decrease in the actual ear separation down from 13 mm re-

sults in ITD cues that are smaller than those the auditory model is tuned to, and this creates a mismatch between the actual ear separation and the auditory model. When the actual ear separation is increased over the matched value of 13 mm, the phonotactic performance is excellent, with no misses. An increase in the actual ear separation over the matched one results in ITD cues that are greater than those when the actual ear separation is identical to the matched value. Consequently, the directionality of the auditory model is further improved. When the actual ear separation is identical to the matched value, the robot exhibits excellent performance.

When the auditory model is scaled to match an ear separation of 3 mm, the phonotactic performance improves as the actual ear separation increases with respect to the matched value. However, at an actual ear separation of 100 mm, the performance is drastically degraded. This is due to the fact that the phase difference between the sound waves incident from either side, at  $\pm 90^\circ$  relative to the robot, is zero for ear separation of  $\frac{k \cdot \lambda}{2}$ , for all  $k \in \mathbb{N}$ . The wavelength  $\lambda$  of a 6500 Hz sinusoid is  $\frac{340 \text{ ms}^{-1}}{6500} \approx 52 \text{ mm}$ , and the 100 mm ear separation is about twice this wavelength. Therefore, the phase difference between the sound waves at either side varies periodically with the angle of incidence, resulting in the apparent sound direction being incorrectly related to the actual direction for a relatively large number of incidence angles. When the robot is moving, the relative incidence angle and thus the phase difference is also changing continuously. Consequently, the decision model is unable to steer the robot to the sound source.

#### 4.5.1 Trajectory Directness

In order to determine the performance of the robot in terms of its trajectory, the “directness” statistic [113] as given by (3.3). The trajectories were divided into  $n$  vectors ( $n = 13$  in this case) of a fixed length and the “directness” was calculated via the same procedure as described in Sect. 3.4.2.

Average headings were individually calculated for the left and the right side trajectories via the directional statistics approach [80], for each sound frequency and ear separation combination. Fig. 4.7 depicts the angular error  $\theta_{\text{err}}$  in degrees between the average heading and the ideal heading of  $0^\circ$ . It is evident from Fig. 4.7(a) and Fig. 4.7(b) that when the auditory model is scaled up to match an ear separation of 100 mm, the heading errors are quite low for all sound frequencies for an actual ear separation of 100 mm, for both the left and the right side trajectories respectively. In both cases, a decrease in the actual ear separation results in a corresponding increase in the heading

errors over the entire frequency range, due to smaller ITD cues being extracted by the auditory model, resulting in an increasing mismatch between the scaled up auditory model and the actual ear separation.

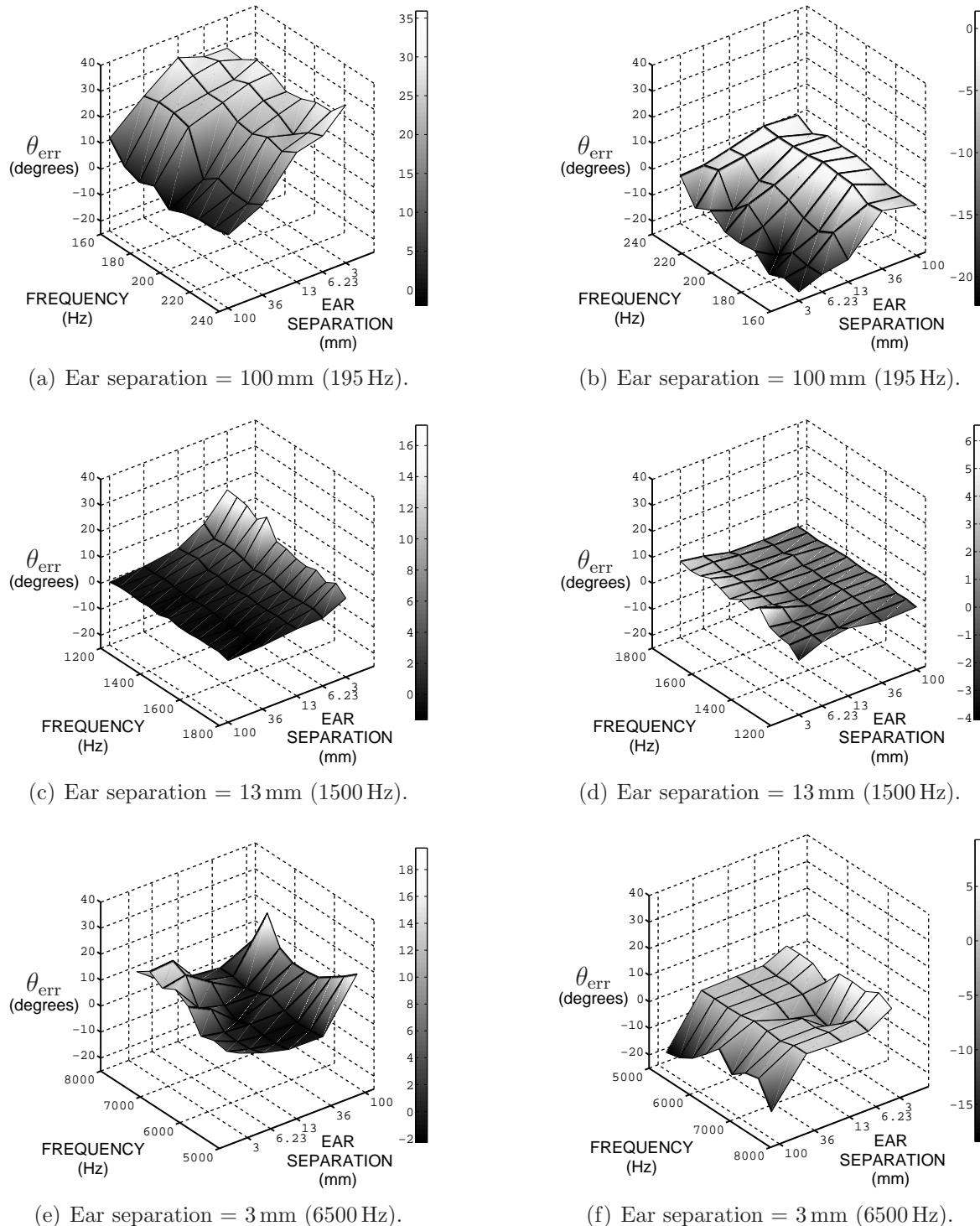
When the default auditory model is utilized, it is evident from 4.7(c) and Fig. 4.7(d) that the heading errors are almost  $0^\circ$  for all sound frequencies when the actual ear separation matches an ear separation of 13 mm. However, when the actual ear separation decreases below 13 mm, the heading errors increase, again due to the smaller ITD cues being extracted by the auditory model, resulting in a mismatch between the auditory model and the actual ear separation. When the actual ear separation increases above 13 mm, the heading errors are again quite low, since the ITD cues are now larger than those expected by the auditory model, and this further improves directionality.

When the auditory model that is scaled down to match an ear separation of 3 mm is utilized, it is evident from 4.7(e) and Fig. 4.7(f) that for all sound frequencies the heading errors decrease as the actual ear separation increases over the matched value of 3 mm, again due to greater ITD cues being extracted by the auditory model. However, when the actual ear separation is 100 mm, the heading errors are large, since this ear separation is close to twice the wavelength of the peak sound frequency, resulting in a mismatch between the actual and the apparent sound direction as described earlier.

These results are consistent with the hypothesis that having relatively greater ITD cues improves directionality, while having relatively smaller ITD cues degrades directionality.

## 4.6 Summary and Conclusions

In this chapter the default auditory model was modified to generate two different versions, one by scaling up the model's parameters, and another by scaling down the model's parameters. The phonotactic performance of the three models thus obtained was experimentally evaluated in terms of the number of *hits*, *near hits* and *misses*, in order to test the hypotheses that — (a) any joint scaling of the auditory model should result in an equivalent scaling of its frequency response, and the scaled model may perform equivalently to the unscaled model in phonotaxis, provided that the scaling of the model's parameters is consistent with the scaling of the corresponding ear separation and (b) if the ear separation for any given auditory model, scaled or otherwise, is varied either above or below the matched ear separation then the model's directionality should either



**Figure 4.7:** Average angular error  $\theta_{\text{err}}$  in the trajectory heading with respect to the ideal heading of  $0^\circ$ . Figures (a), (c) and (e) correspond to the tests performed with the left speaker activated and Fig. (b), (d) and (f) correspond to the tests performed with the right speaker activated.

improve (under the condition that the ear separation is less than half the wavelength) or degrade respectively. The performances of the three auditory models were also examined in terms of the trajectories, via the computation of a “directness” statistic.

It is evident from the frequency responses of the scaled auditory models that when the model’s parameters are scaled, there is a corresponding scaling of the frequency response as well. When the parameters of the auditory model are scaled up to match an ear separation greater than 13 mm, the frequency range is scaled down below the 1000 Hz–2200 Hz range and when the parameters of the auditory model are scaled down to match an ear separation less than 13 mm, the frequency range is scaled up above the 1000 Hz–2200 Hz range. In terms of the number of *hits*, *near hits* and *misses* it is evident that the robotic implementation exhibits excellent performance for both the scaled up and scaled down auditory models, comparable to that of the unscaled model, provided that the scaling of the models is consistent with the actual ear separation. These results are consistent with the first hypothesis regarding the persistence of a scaling law. Furthermore, for any given auditory model matched to a given ear separation, when the actual ear separation is increased above the matched value the performance of the robotic implementation improves as long as the actual ear separation is less than half the wavelength of the frequency of the sound source, and when the actual ear separation is decreased below the matched value the performance of the robotic implementation degrades. This is consistent with the second hypothesis regarding the consequences of any mismatch between the scaling of the model’s parameters and that of the actual ear separation on the directional acuity of the robotic implementation. This hypothesis is further supported through the results obtained in terms of the heading errors. Once again, it is evident that for any given auditory model matched to a given ear separation, when the actual ear separation is increased above the matched value the heading errors of the robotic implementation tend to decrease as long as the actual ear separation is less than half the wavelength of the frequency of the sound source, and when the actual ear separation is decreased below the matched value the heading errors of the robotic implementation tend to increase.

It can be concluded that the auditory scaling law seen in mammalian species also holds for the lizard peripheral auditory system, but further physiological and morphological studies are required to confirm these findings from a biological perspective. When the relationship between the ear separation and the tuned auditory model is consistent the model exhibits strong directionality, and a negative mismatch created by utilizing a relatively smaller ear separation than the matched value degrades the directionality while a positive mismatch created by utilizing a relatively greater ear separation enhances it.



# Chapter 5

## Learning to Represent Sound Direction

### 5.1 Introduction

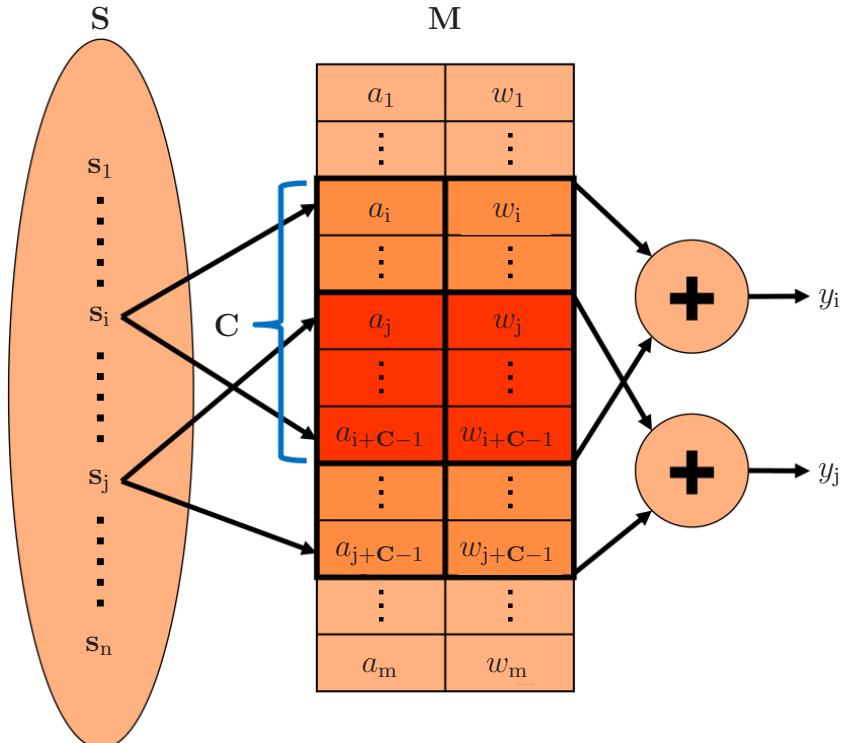
The lizard’s peripheral auditory system model has been studied extensively from a robotics perspective, both in simulation [137] and in real-world implementation as a controller for a sound localizing mobile robot in phonotaxis experiments. These controllers comprised simple *if - then* rules [138, 115] that mapped the auditory cues extracted by the system to motor control outputs that steered the robot towards the sound source. This work has been further extended to Braitenberg vehicle-like [10] implementations [117, 114] (refer to Chap. 3 as well). This simple control strategy works well, but only as long as the robot is allowed to move and approach the sound source. If the translational and/or rotational capability is removed, this strategy fails in consistently localizing the sound source over the relevant frequency range because the relevant cues extracted by the model vary non-linearly as the relative azimuthal position of the sound source changes. Therefore, the extracted auditory cues cannot be directly mapped in a 1:1 manner to the sound direction. Spatial movement defined by the control strategy compensates for this non-linear representation of sound direction by forcing the agent to place the source in the direction of most sensitivity and keep it there, essentially enforcing a 1:1 mapping. In real lizards, the auditory cues extracted by the peripheral auditory system are passed on to the animal’s nervous system for further processing and interpretation. This raises the question, how well can the peripheral auditory system in itself represent the directional information of a sound source? In this chapter we attempt to investigate this question by coupling the peripheral auditory system model with a simplified Cerebellar

Model Articulation Controller [3] neural network and training the network in simulation and in practice to construct a plausible representation of the sound direction.

The work presented in this chapter covers and elaborates further upon the work presented by the author at the International Workshop on Bio-inspired Robots<sup>1</sup> held through April 6–8, 2011 at L’École des Mines de Nantes, Nantes, France. However, the conference proceedings were not formally published by the conference organizers.

### 5.1.1 Cerebellar Model Articulation Controller

The Cerebellar Model Articulation Controller (henceforth referred to as CMAC) architecture was originally proposed by James S. Albus in 1975 with primary application towards control of robotic manipulators. The simplicity and suitability of the CMAC architecture, especially for digital implementations, were key factors in its selection in this work. Furthermore, the CMAC is a universal function approximator. It is based upon theories of how the cerebellum stores sensorimotor mappings [2, 4]. Figure 5.1 depicts the CMAC architecture.



**Figure 5.1:** The CMAC architecture.

<sup>1</sup>[Link to conference abstract](#).

It is basically a sparsely filled memory space  $\mathbf{M}$  containing the network's weights  $w_k$ , each stored in a memory location with address  $a_k$ , where  $k = 1 : m$  and  $m$  is the total number of weights. Given an input vector space  $\mathbf{S}$  with  $n$  vectors, each vector  $\mathbf{s}_k$  in  $\mathbf{S}$  addresses a set of memory locations of width  $\mathbf{C}$ , in  $\mathbf{M}$ . The corresponding output  $y_k$  is calculated by summing up the contents of the set of addressed memory locations. The addressing scheme quantizes each point in a given input vector  $\mathbf{s}_i = (s_1, s_2, \dots, s_p)$  into discrete groups via a quantization function. This divides the input space  $\mathbf{S}$  into  $p$ -dimensional hypercubes. Depending on the number of quantization functions used, different overlapping layers of such hypercubes exist, each shifted from the previous layer to a certain extent. Each vector in the input space  $\mathbf{S}$  is associated with a hypercube in each layer, and in this manner the vector activates a number of hypercubes within the set of all hypercubes. Each hypercube is furthermore associated with a memory location, and thus for a given input vector a set of memory locations, and consequently a set of weights is addressed. The CMAC employs the delta rule as the learning algorithm for updating the weights and the learning error is uniformly distributed over all the weights. For a given training input vector  $s_i$ , the weight update  $\delta w_k$  for a given weight  $w_k$  is given by

$$\delta w_k = \frac{\eta}{\mathbf{C}} \left( d_i - \sum_{k=1}^{\mathbf{C}} w_k \right), \text{ where} \quad (5.1)$$

$\eta$  = learning rate

$\mathbf{C}$  = the number of weights activated

$d_i$  = the target output

$\sum_{k=1}^{\mathbf{C}} w_k$  = the actual network output.

Consider two input vectors  $\mathbf{s}_i$  and  $\mathbf{s}_j$  positioned in  $\mathbf{S}$  in such a manner that the sets  $[a_i \dots a_{i+\mathbf{C}-1}]$  and  $[a_j \dots a_{j+\mathbf{C}-1}]$ , addressed by  $\mathbf{s}_i$  and  $\mathbf{s}_j$  respectively, overlap to a certain extent. The extent of this overlap depends on the width  $\mathbf{C}$ , which is also the generalization parameter of the network. For a given pair of input vectors, the larger the value of  $\mathbf{C}$ , the greater is the overlap and higher is the generalization capability of the network. Alternatively, for a given  $\mathbf{C}$ , the closer  $\mathbf{s}_i$  and  $\mathbf{s}_j$  are to each other, the greater is the overlap, and consequently, closer are the corresponding outputs  $\mathbf{y}_i$  and  $\mathbf{y}_j$ . The

choice of  $\mathbf{C}$  typically depends on the characteristics of the input space  $\mathbf{S}$ , but in order to avoid over-generalization, the condition  $\mathbf{C} \ll m$  must be maintained. The consequence of utilizing such a mapping scheme that creates an overlap between the memory location sets corresponding to different input vectors, such that the closer the input vectors are the greater is the overlap, is that similar inputs tend to generalize and produce similar outputs, while dissimilar inputs tend to dichotomize, resulting in dissimilar outputs. If the input vectors are far enough from each other, the outputs are independent, i.e. any outliers in the input space will be distinctly mapped to a specific output, ensuring that any extreme points in the input space will also be accounted for and represented in the input-output mapping.

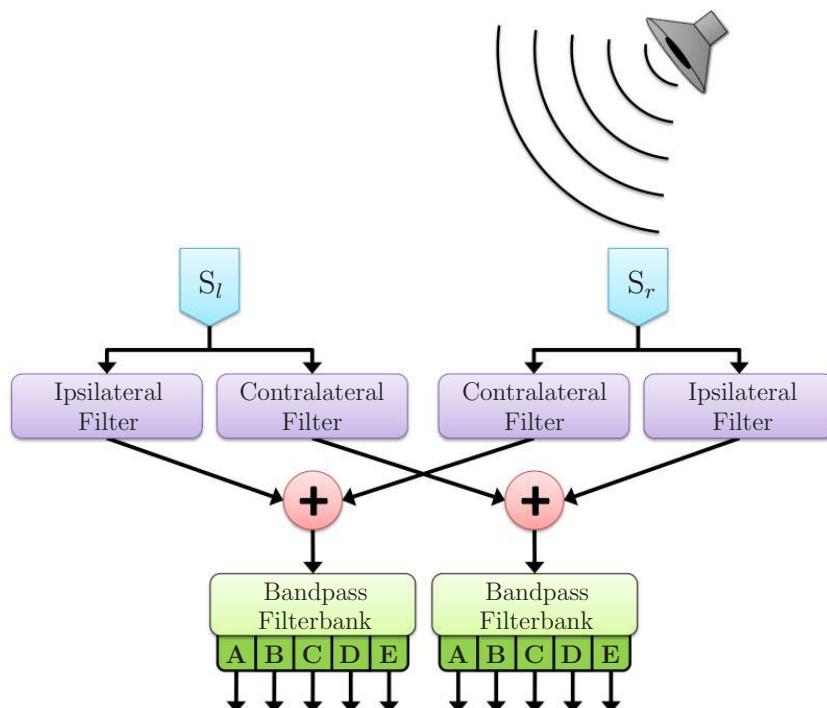
In the originally proposed CMAC architecture, the input vectors address a conceptual memory space rather than a physical memory space because the total addressable memory width increases significantly even for a small increase in the number of dimensions of the input vectors, irrespective of the fact that the conceptual memory space is sparsely filled. If for each point in a vector  $\mathbf{s}_i = (s_1, s_2, \dots, s_p)$  in the input space  $\mathbf{S}$ , there exist  $R$  distinctly separate values, then there are  $R^p$  distinctly separate points in the input space. Therefore, it quickly becomes infeasible from a implementation and computation point of view as the conceptual memory space typically exceeds the available physical memory in most real-world problems where the CMAC needs to be implemented within bounded hardware resources. In such cases, a hash function is then employed to map the conceptual memory addresses into physical memory addresses that address a much smaller memory space that is much more feasible to implement on hardware, at the expense of additional computation. In our case, a hash function does not offer a significant advantage in terms of error minimization and the time and effort needed to implement it cannot be sufficiently justified. We therefore avoid its usage, and treat the conceptual memory space and the physical memory space as one.

## 5.2 Design and Implementation

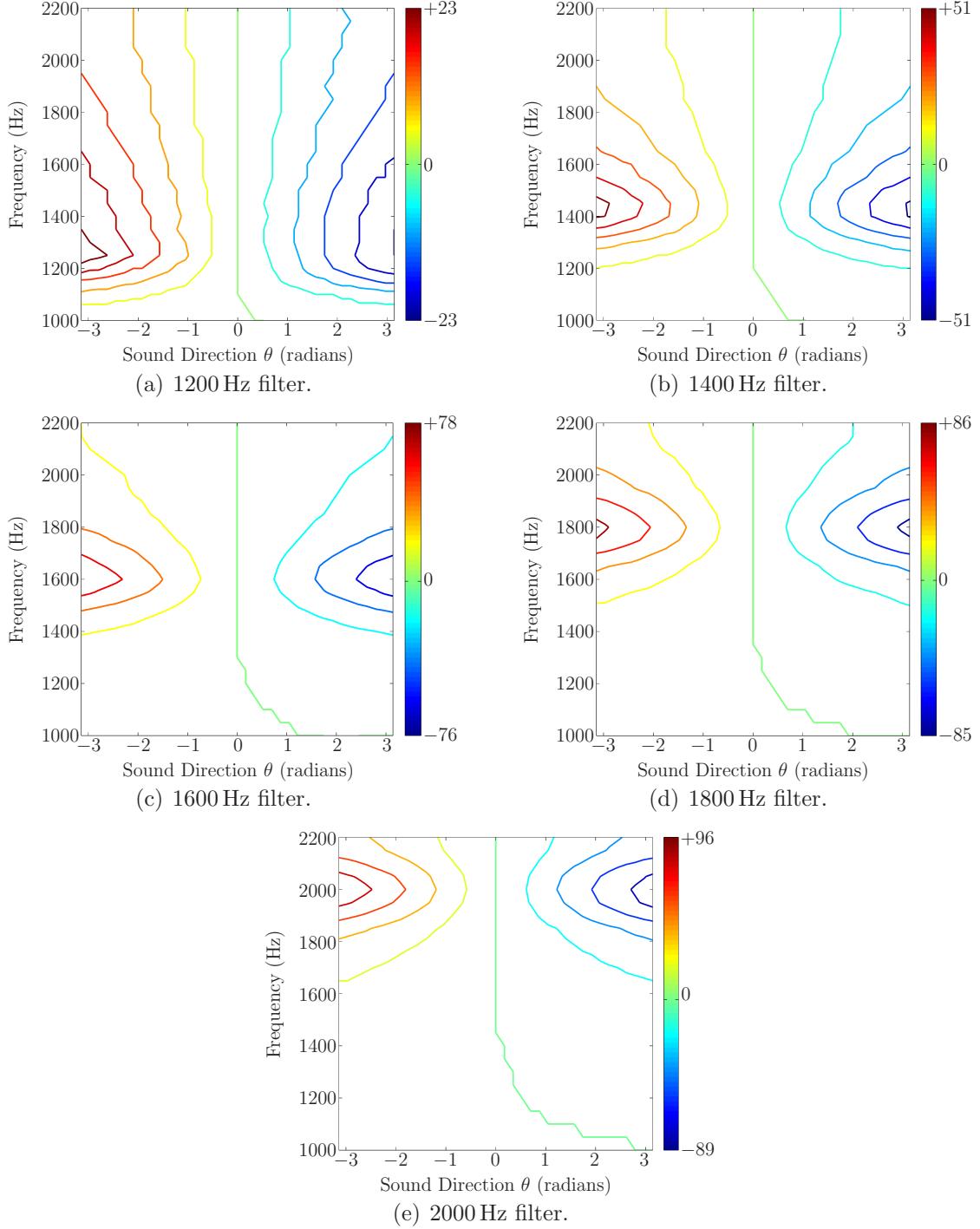
### 5.2.1 The Auditory Model Implementation

The auditory model is coupled to the CMAC neural network and generates input data for it. As described in Sec. 3.2.1, the response of the model to various azimuthal positions of the sound source is different for different sound frequencies (refer to Fig. 3.4). In general, for lower frequencies the response varies slowly while for higher frequencies it

varies sharply, and in both cases the variation is nonlinear. Since the goal is to construct an accurate representation of sound direction, information regarding sound frequency must be made available to the network. Thus, instead of calculating the powers at the outputs of the auditory model, the two left and right output signals  $i_l$  and  $i_r$  of the model are individually coupled to two separate but identical filterbanks as shown in Fig. 5.2 in order to extract the frequency information for a given auditory signal. Each filterbank has five bandpass filters labelled **A**, **B**, **C**, **D** and **E**, with centre frequencies of 1200 Hz, 1400 Hz, 1600 Hz, 1800 Hz and 2000 Hz respectively, and a 3 dB bandwidth of 200 Hz. These values were chosen so as to cover the relevant frequency range of 1000 Hz–2200 Hz while maintaining differentiation of the frequency range into sufficiently distinct groups. Figure 5.3 depicts the average power in the output signals of the bandpass filters for different sound frequencies between 1000 Hz–2200 Hz and different azimuthal positions of the sound source within the range  $[-\pi, +\pi]$ . These output powers were then employed as input when training the CMAC neural network.



**Figure 5.2:** The auditory model augmented with filterbanks which acts as input to the CMAC neural network.



**Figure 5.3:** The relative response in dB of individual bandpass filters in the filterbank.

### 5.2.2 The Learning Architecture

Before the CMAC architecture could be implemented in practice, the generalization parameter  $\mathbf{C}$  must be chosen. Typically,  $\mathbf{C}$  is chosen on the basis of a trade-off between the

nature of the input-output mapping required, the dispersion characteristics of the data points in the input space and the amount of physical memory available to implement the memory space. To the best of the author's knowledge, there is no standard method to determine  $\mathbf{C}$  in the existing literature. Therefore,  $\mathbf{C}$  was determined incrementally in simulation. Different networks were generated in succession, each with a higher value of  $\mathbf{C}$  than its predecessor, for  $\mathbf{C} = 10, 20, 30, \dots, 200$ . Each network was then trained in a supervised manner using input data from the auditory model filterbanks as described in Sect. 5.2.1. As an additional consequence, the use of supervised learning also allowed the generation of a reference "best possible" representation of the sound direction for comparative purposes. However, instead of feeding the power values at all 10 filters simultaneously to the network, only one pair of filters, one from each filterbank but having identical centre frequencies, was selected at a time to act as input. This was necessary because a 10-dimensional input vector would result in a large input space, consequently requiring a large memory space. For example, if each point in  $\mathbf{s}_i = (s_1, s_2, \dots, s_p)$  is an 8-bit value, then  $R = 2^8 = 256$ , and the size of the input space becomes  $R^p = 256^{10}$  or, of the order of  $10^{24}$ . By limiting the number of input space dimensions to 2, the size of the input space is significantly reduced to  $256^2$  or, of the order of  $10^4$ . The width  $N_{total}$  of the maximum addressable memory in terms of the number of memory locations is given by

$$N_{total} = (2^b + \mathbf{C} - 1)^p , \quad (5.2)$$

where  $b$  is maximum number of bits needed to represent a point in  $\mathbf{s}_i = (s_1, s_2, \dots, s_p)$  and  $p$  is the number of dimensions of the input space. Thus, the maximum required memory locations depends on three parameters — (a) the maximum number of bits required to represent the scalar data points in the input vectors, (b) the dimensionality of the input space and (c) the generalization factor  $\mathbf{C}$ . From (5.2), one can see that  $N_{total}$  increases exponentially with the input space dimension  $p$ . And for implementation purposes the highest possible representation for the power outputs from the filterbanks was an 8-bit representation. Utilizing greater than 8 bits was impossible because this would require a minimum of  $(2^9 + 1 - 1)^2 = 262,144$  memory locations, or 19-bit memory addresses. With 19-bit memory addresses, the maximum number of addressable memory locations is  $2^{19} = 524,288$ . The maximum available memory on the robotic platform is 1344 KiB, so for 524,288 memory locations to fit inside the avail-

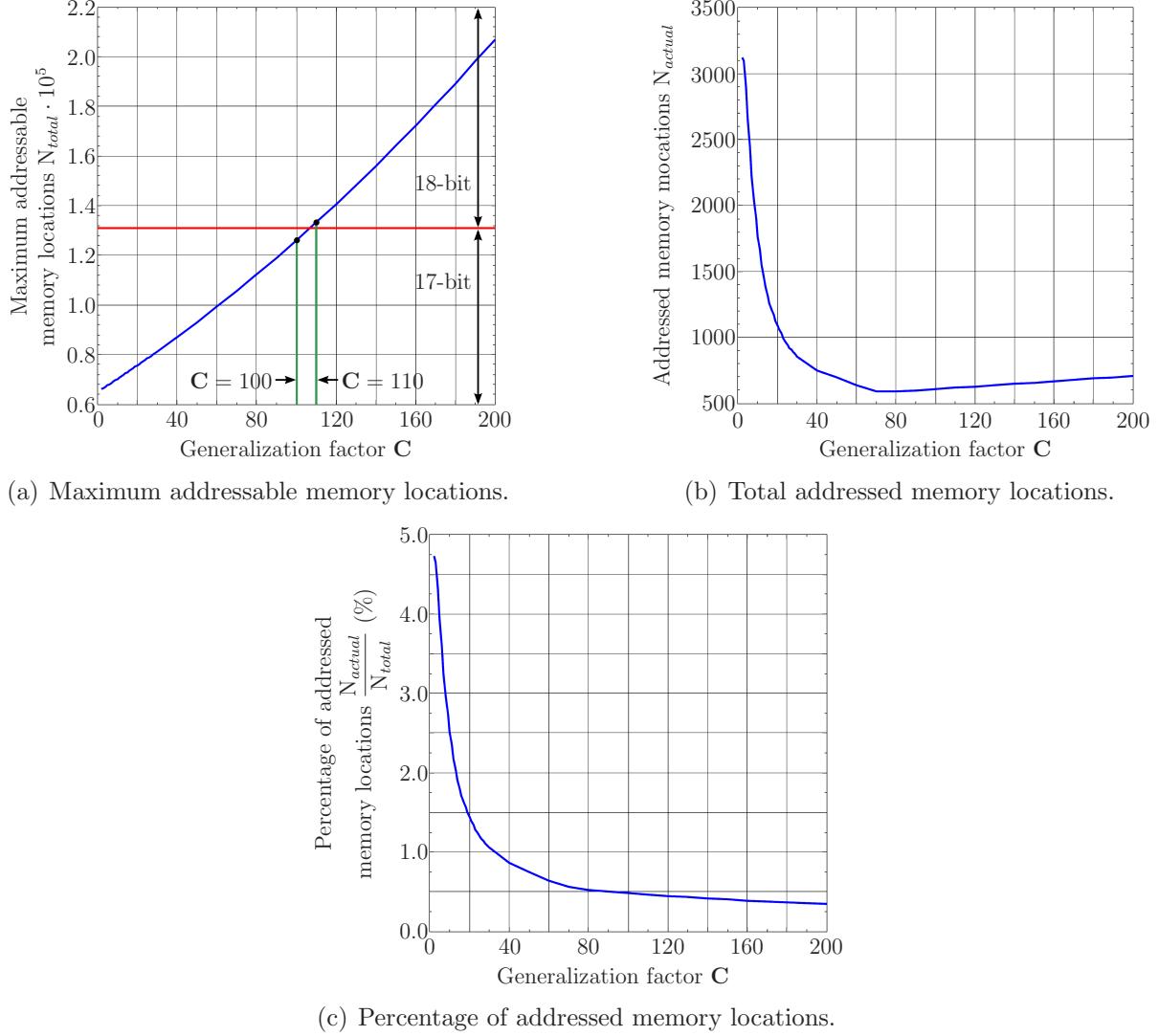
able memory, the maximum possible number of bits available to store each weight is  $\frac{1344 \cdot 1024}{524288} = 2.625 \approx 2$  bits. A 2-bit resolution implies that each weight can only have 4 values, and this is impractical. Therefore, an 8-bit representation was utilized for the power outputs from the filterbanks. After substituting for  $b$  and  $p$ , we obtain the relationship between maximum addressable memory locations, i.e.  $N_{total}$ , and generalization parameter  $C$  as

$$N_{total} = (2^8 + C - 1)^2. \quad (5.3)$$

Figures 5.4(a)–(c) depict the effect of varying  $C$  on the memory requirements. As  $C$  increases the area of each hyper-rectangle also increases, reducing the total number of hyper-rectangles required to cover the input space. Consequently, the number of memory locations actually addressed  $N_{actual}$  also decreases in an exponential manner, as is evident in Fig. 5.4(b). Furthermore, from Fig. 5.4(a) it can be seen that for  $C \leq 100$ , 17 bits are needed to represent the memory addresses, and for  $C > 100$ , 18 bits are needed to represent the memory addresses.

The maximum number of memory locations that can be addressed with 17-bit addresses is  $2^{17} = 131,072$ , which implies that there are 131,072 individual network weights. The range over which a weight can be updated is  $[-2^{n-1}, 2^{n-1} - 1]$ , where  $n$  is the number of bits required to represent the weight. A higher  $n$  implies a greater range, and consequently a higher resolution in the network's outputs. However,  $n$  is bounded by the physical memory available for implementation. As mentioned earlier, the maximum available memory on the robotic platform is 1344 KiB, so for 131,072 memory locations to fit inside the available memory, the maximum possible  $n$  is  $\frac{1344 \cdot 1024}{131072} = 10.5 \approx 10$  bits. In other words, in order to realize 131,072 memory locations in an available space of 1344 KiB, the maximum number of bits required to store each weight is 10. Thus, the usable memory size is  $2^{17} \cdot 10 = 1,310,720$  bits, or 1280 KiB and each weight has a range of  $[-2^9, 2^9 - 1]$ , or  $[-512, 511]$ . On the other hand, if we consider 18-bit addresses, then the maximum number of memory locations that can be addressed is  $2^{18} = 262,144$ . For this value, in order to fit inside the available memory the maximum possible  $n$  is  $\frac{1344 \cdot 1024}{262144} = 5.25 \approx 5$  bits, giving a range of  $[-2^4, 2^4 - 1]$ , or  $[-16, 15]$ . For a fixed-point implementation where the weights will have integer values, the minimum increment for any given weight will be 1, and having such small limits could cause the weights to

overflow much faster during the learning process than when the limits are  $[-512, 511]$ . Therefore, the 17-bit representation for the memory addresses was selected.



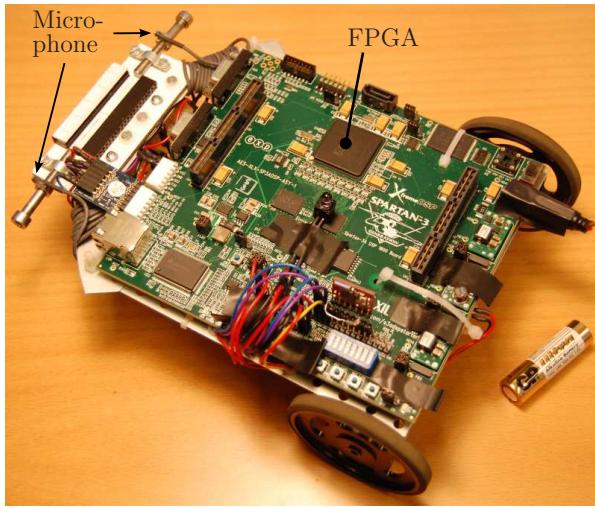
**Figure 5.4:** The effect of the generalization parameter  $C$  on memory requirements.

Utilizing a 17-bit representation narrows the range of possible values of  $C$  to less than 100. From Fig. 5.4(b), it is evident that for  $C \geq 80$ , the number of memory locations actually addressed, or  $N_{actual}$ , tends to increase. Therefore, in order to minimize the amount of memory used, the upper bound on  $C$  was further reduced to 80. After presenting the input training data to the network with  $C = 80$ ,  $N_{actual}$  was determined to be equal to 588. However, limiting the addressed memory to the minimum possible is not a major constraint, since the entire memory space is implemented. But the lower the generalization factor, the lower is the generalization capability of the network, and

it was determined that there was no further improvement in the network's performance for  $C > 80$ . Furthermore, it is evident from Fig. 5.4(c) that the percentage reduction in the number of addressed memory locations is exponential with respect to increasing  $C$ , and after a certain value of  $C$  there is no significant reduction. For  $C = 80$ , only 0.5239 % of the total addressable memory locations are actually addressed. As a matter of fact, it was determined that for  $C \geq 80$  the percentage reduction in the addressed memory locations, between two consecutive values of  $C$  was less than 0.1 %. For these reasons, there was not enough justification for considering values of  $C > 80$ .

### 5.2.3 The Robotic Platform

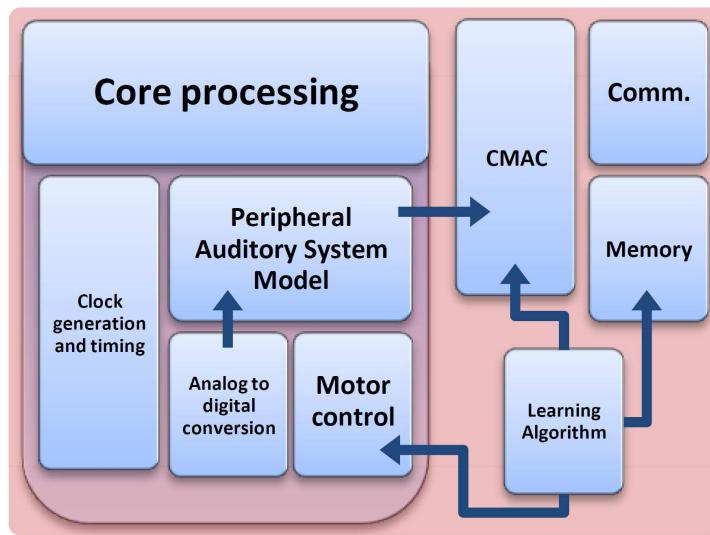
For the robotic trials, a mobile robot, depicted in Fig. 5.5 was constructed. It comprised a Xilinx XtremeDSP Starter Platform FPGA board (produced by Xilinx Inc.) mounted on a mobile platform assembled with electro-mechanical components (from the Robotics Starter Kit from Digilent Inc.). The sole processing unit on the board was the Xilinx Spartan®-3A DSP 1800A FPGA chip. The main reasons for choosing this particular platform were its generous logic gate count (1.8 million gates) and available on-chip memory of 1344 KiB.



**Figure 5.5:** The mobile robot.

The auditory model and the learning architecture were implemented on the FPGA chip using integer representation in the VHDL programming language. The VHDL design was then synthesized or compiled via Xilinx's proprietary software tool, namely the Xilinx Integrated Synthesis Environment (Xilinx ISE), into hardware-level binary

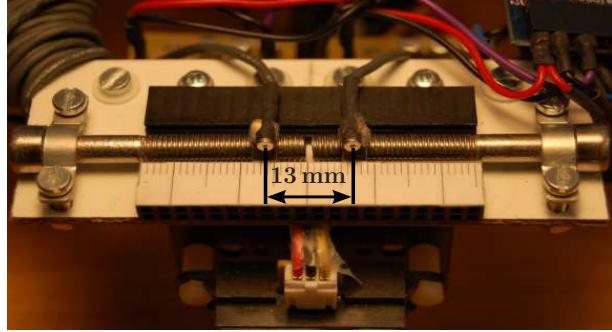
“bitstream” that contained all the information necessary to properly configure and program the logic into the FPGA chip. Figure 5.6 depicts the complete implementation architecture on the FPGA chip. The VHDL code was modular in design, consisting of individual functional modules that were further composed of individual sub-modules, all linked together via a global clock and timing generation mechanism. The design was based around a uni-directional flow of information from one module to another. The modular code eased debugging tasks considerably.



**Figure 5.6:** The modular implementation on the FPGA chip.

Two omnidirectional microphones (model FG-23329-P07 from Knowles Electronics) are mounted at the front of the robot (refer to Fig. 5.7), 13 mm apart. This separation is similar to that found between the tympani of a typical lizard [21] because the auditory model’s parameters have been derived based on measurements with the real animal, and a different microphone separation would create a mis-match between the model and the external stimuli it expects. External sinusoid auditory signals were passed as input to the board via these microphones. A dual channel 12-bit analogue-to-digital converter digitized the signals from the microphones before passing them on to the FPGA chip.

A wireless link was set up between the robot and an off-board PC as a bi-directional communication channel, using a 2.4 GHz RF wireless transceiver (nRF2401A from Nordic Semiconductor Inc.). Two standard 6 V DC motors controlled by H-bridge motor drivers independently controlled two rear wheels, while at the front, an omnidirectional ball caster was employed as a support. Power to the development board was supplied via a 3.7 V, 6 Ah lithium polymer battery, while power to the motors was supplied via a 3.7 V, 920 mAh lithium polymer battery.



**Figure 5.7:** The microphone placement.

## 5.3 Experimental Methods

### 5.3.1 Simulation

#### 5.3.1.1 Experimental Setup

**Training Data Generation.** To generate the training dataset, sinusoidal sound signals (generated in Matlab<sup>®</sup> using a floating point representation with 32-bit precision, with a zero centre amplitude and a peak amplitude of  $\pm 1.0$ ) arriving at the left and right sides of the simulated peripheral auditory system, from a hypothetical pure tone sound source of given frequency and azimuthal position with respect to its centre, were first generated. The signals were appropriately phase-shifted according to the azimuthal position of the source. These signals were then fed to the auditory system model and the scalar output power values at the left and right bandpass filterbanks were obtained as mentioned in Sect. 5.2.1. Two bandpass filters, one from the left filterbank and the other from the right filterbank, with identical cut-off frequencies that were randomly chosen were then selected and their output powers were paired together to form a training vector. The filters were selected based on a shuffled table containing the centre frequencies. This training vector was then labelled with the chosen azimuthal position. In this manner, different training vectors for 25 different sound source frequencies between 1000 Hz–2200 Hz in steps of 50 Hz, 37 different azimuthal positions within the range  $[-\frac{\pi}{2}, +\frac{\pi}{2}]$  in steps of  $5^\circ$ , and 5 filterbank pairs were generated, giving a total of  $25 \cdot 37 \cdot 5 = 4625$  training vectors. All the weights stored in the memory space were initialized to have a value of 0.

**Supervised Learning.** In order to generate a reference “best possible” representa-

tion, a Matlab® simulation with supervised learning was set up. The training dataset was randomly shuffled and training vectors were sequentially fed into a CMAC network with generalization parameter  $\mathbf{C} = 80$  as described in Sec. 5.2.2. The network’s output, which was a summation of the weights stored at the memory locations addressed by each vector, was calculated. This output represented the estimated sound source direction. The weights were then updated according to the gradient descent rule described in Sect. 5.1.1. This procedure was repeated for all training vectors. The training period was 100 epochs, where one epoch meant the presentation of all the training vectors to the network, with the learning rate  $\eta$  set to 1. After each epoch, the network was tested in simulation with a test dataset derived in a similar fashion to the training dataset but independent from it. Since it was not evident beforehand what the sound direction representation should look like, the error metric employed here during the testing phase was the root-mean-square error between the current representation and an “ideal” representation, defined as one in which the outputs from the bandpass filters were mapped directly and linearly to the sound direction. A linear mapping meant that the network’s output could be directly interpreted as the estimated sound direction, making the turning motion of the robot directly proportional to the network’s estimate, similar to the direct mapping between auditory cues and sound direction allowed by having translational and rotational freedom as described in Sec. 5.1. Mathematically, the learning error  $\xi$  is given by

$$\xi = \sqrt{\frac{\sum_{f=1000}^{2200} \sum_{\phi=-\frac{\pi}{2}}^{+\frac{\pi}{2}} |\hat{\theta}_{f,\phi} - \theta_{f,\phi}|^2}{N \cdot M}}, \text{ where} \quad (5.4)$$

$f$  = the sound frequency

$\phi$  = the azimuthal sound source displacement

$\hat{\theta}_{f,\phi}$  = the estimated sound direction for a  
given  $f$  and  $\phi$

$\theta_{f,\phi}$  = the actual sound direction

$M$  = the number of distinct values of  $\phi$  ( $M = 37$ )

$N$  = the number of distinct values of  $f$  ( $N = 25$ ).

**Unsupervised Learning.** A second CMAC neural network, identical to the first, was then trained in an unsupervised manner via reinforcement learning. Again, a randomly chosen vector from the dataset was fed into the network and its output, which was the estimated sound source direction, was calculated. The “robot” was then “rotated” by an angle equal to the calculated output such that it was oriented towards the estimated sound source direction. Then the difference between the left and right power values generated for the new orientation was calculated. The previously activated weights were either decremented, incremented or left unchanged by a fixed amount equal to  $\delta w$  as given in (5.1) depending on whether the difference was greater than, less than or equal to 0, respectively. This procedure was repeated for all training vectors. The training time was again 100 epochs, with the learning rate  $\eta$  set to 1.

The simulations described so far were designed specifically to determine whether a plausible sound direction representation could be generated or not. Therefore, their primary goal was to validate the learning algorithm. However, the details of the auditory model and the filterbank implementation in simulation differed from those in the robotic implementation. The implementation in simulation employed a decimal number representation while the robotic implementation employed a integer number representation. The sinusoids employed in the simulations consisted of 32-bit decimal values bounded within  $\pm 1.0$  while those in the robotic implementation consisted of 12-bit integer values bounded within the approximate range [800, 2600]. Therefore, in the robotic implementation, the filter coefficients of the auditory model and the filterbanks were scaled up in order to avoid their integer outputs either underflowing or overflowing. The filterbank outputs acted as inputs to the CMAC neural network, which was designed for 8-bit integer inputs, as described in Sect. 5.2.2, both in simulation and in the robotic implementation. Therefore the filterbank outputs in the robotic implementation had to be scaled *down* to fit within 8 bits. Conversely, in the simulation the filterbank outputs were decimal values of the order of  $10^{-8}$ , and therefore these had to be scaled *up* and then rounded up or down to integer values to fit within the 8-bit range of [0, 255]. However, the VHDL programming language only allows division by powers of 2 ( $2^k$  where  $k \in \mathbb{N}$ ), while there is no such restriction in Matlab<sup>®</sup>. Therefore the filterbank outputs, i.e. the training dataset, in simulation and in the robotic implementation differed from each other in terms of their bounds. In the simulation, the training dataset was bounded within the range [0, 255], while the training dataset in the robotic implementation was bounded within the range [0, 175]. Essentially, this signified that the training dataset in the robotic implementation was a scaled-down version of the training dataset. In order to increase the upper limit of the bounds in the robotic implementation, the division

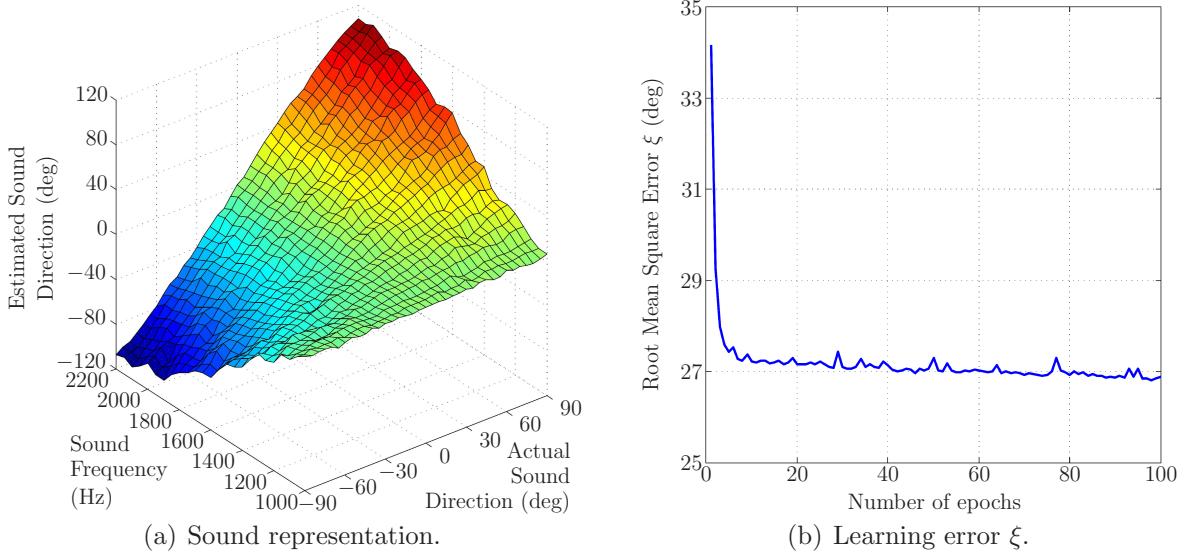
factor was decremented in the VHDL code, but even the minimum possible decrement resulted in the upper bound going over 255. The only option left was to increase the scaling factor in the simulation such that the training dataset in simulation would also lie within the range  $[0, 175]$ , but this was not attempted as it would have meant performing all the simulations again and this was not feasible due to lack of time.

Since the CMAC neural network's operation is dependent on the relative dispersion of data points in the input space, any change in the same could affect the network's performance, and consequently the learning algorithm and the sound direction representation learned. In order to determine the effect of such compression of the training data bounds on the network's ability to learn a representation of sound direction, a third CMAC neural network identical to the previous two was trained in an unsupervised manner via reinforcement learning in simulation. The simulation setup was equivalent to the previous one, but care was taken to ensure that the implementations of the peripheral auditory model, the CMAC architecture and the learning algorithm were all identical to their respective counterparts on the robot. In addition, the auditory signals utilized to generate training data were derived from the real-world digitized audio signals recorded at the outputs of the analogue-to-digital converter for sound frequencies within the 1000 Hz–2200 Hz range and angular displacements of a loudspeaker acting as the sound source within the range  $\left[-\frac{\pi}{2}, +\frac{\pi}{2}\right]$ . Instead of one loudspeaker being placed sequentially at different angular displacements, a configuration of 37 individual passive loudspeakers (from Digilent Inc.) placed 2 m away from the robot's microphones at regular angular intervals of  $5^\circ$  was utilized. All loudspeakers were sourced by a single audio jack of a standard PC sound card, through two 32 channel analogue switches (ADG732 from Analog Devices Inc.). The training time was however reduced to 50 epochs and a trial simulation was run to determine how quickly the error metric, which was also identical to that utilized in the previous simulations, would settle down. It was observed that 50 epochs were sufficient for the error metric to settle down, and thus the training time was set to 50 epochs. The learning rate was again set to 1. The simulation was then reset and executed again.

### 5.3.1.2 Results and Discussion

As mentioned earlier, the CMAC neural network's operation strongly depends on the separation characteristics of data points in the input space, and therefore on their relative magnitudes. Since for different frequencies the variations in the auditory model's outputs over the  $\left[-\frac{\pi}{2}, +\frac{\pi}{2}\right]$  range are not identical, the sound direction representation learned

by the network should reflect these characteristics and deviate from the “ideal” map, where the estimated angles correspond 1:1 to the actual angles. Figure 5.8 depicts the results obtained via supervised learning in simulation, where these characteristics are evident.



**Figure 5.8:** The sound direction representation learned in simulation via supervised learning by the CMAC neural network in terms of the estimated sound direction.

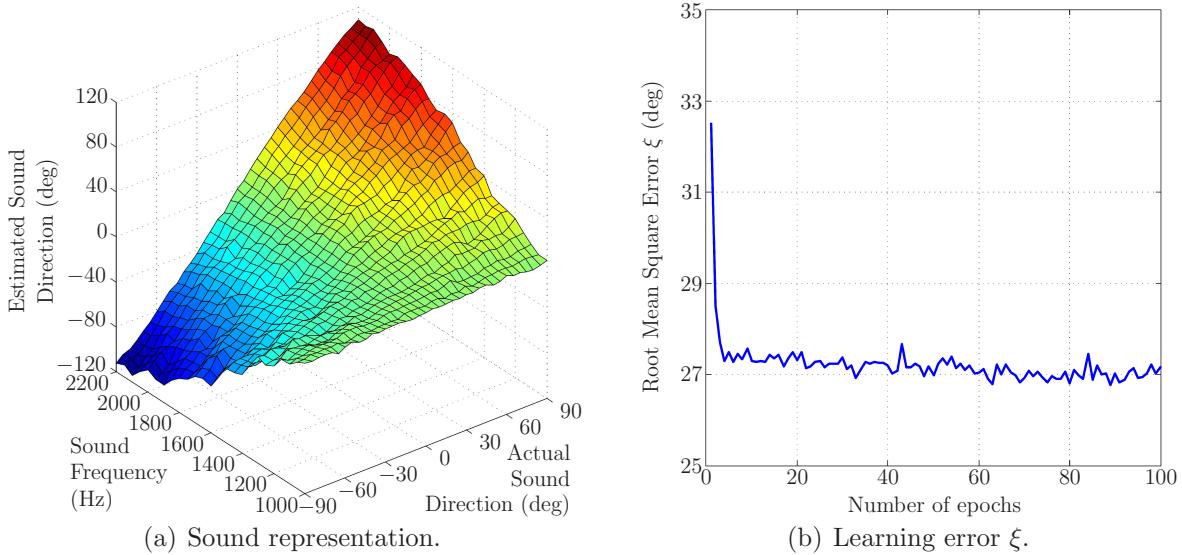
At the lower end of the frequency range the gradient of the estimated sound direction over the  $[-\frac{\pi}{2}, +\frac{\pi}{2}]$  range is extremely flat with respect to the gradient of the actual sound direction, while at the higher end of the frequency range the gradient of the estimated sound direction is highly comparable to the gradient of actual sound direction. While moving from the lower extreme to the higher one, the gradient of estimated sound direction becomes steeper in a gradual manner, reflecting the characteristics of the auditory model’s response. Furthermore, for any given frequency, the gradient of the estimated sound direction is quite linear. From Fig. 5.8(b) it is evident that the network learns quite fast, with the root-mean-square error settling around  $27^\circ$  after only a few epochs. At the higher end of the frequency range the learned representation of sound direction has a relatively steep gradient close to that of the “ideal” representation, and

$$\sum_{\phi=-\frac{\pi}{2}}^{+\frac{\pi}{2}} \left| \hat{\theta}_{f,\phi} - \theta_{f,\phi} \right|^2$$

thus the mean-squared error for a given frequency, i.e. the term  $\frac{1}{M}$  in (5.4), is small in this region, while at the lower end of the frequency range the learned representation is quite flat, such that the mean-squared error for a given frequency is large. Therefore, the overall root-mean-square error  $\xi$  over all the frequencies and az-

imuthal displacements will not be near  $0^\circ$ . This indicates that using the “ideal” sound representation to define  $\xi$  will not provide the most suitable error metric. Unfortunately, prior information about the expected sound representation was not available for the definition of a more suitable error metric.

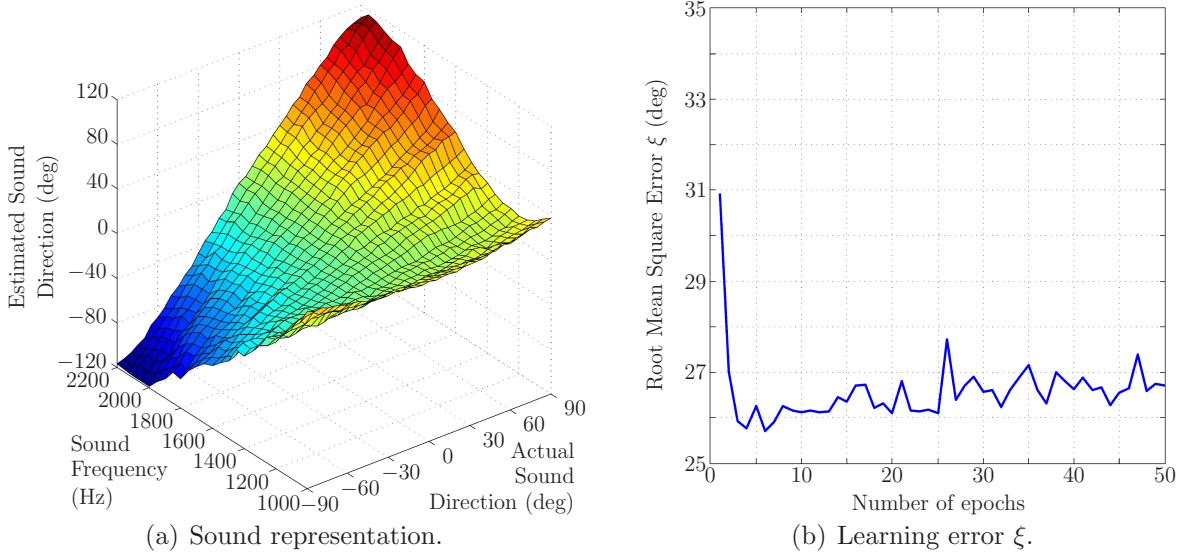
Figure 5.9 depicts the sound direction representation learned via unsupervised learning. It is evident that the sound direction representation is quite similar to that obtained after supervised learning. Once again, the gradient of the estimated sound direction is considerably small with respect to the gradient of actual sound direction at the lower end of the frequency range, while at the higher end of the frequency range the gradient of the estimated sound direction is highly comparable to the gradient of actual sound direction. The gradient of estimated sound direction becomes steeper in a gradual manner while moving from the lower extreme to the higher one, again reflecting the characteristics of the auditory model’s response. Furthermore, for any given frequency, the gradient of the estimated sound direction is quite linear as in the previous case. From Fig. 5.9(b) it is evident that the network learns quite fast as well when utilizing unsupervised learning, with the error metric settling around  $27^\circ$  as in the former case after only a few epochs. For reasons mentioned earlier, the overall root-mean-square error  $\xi$  over all the frequencies and azimuthal displacements will not be near  $0^\circ$ .



**Figure 5.9:** The sound direction representation learned in simulation via unsupervised learning by the CMAC neural network in terms of the estimated sound direction.

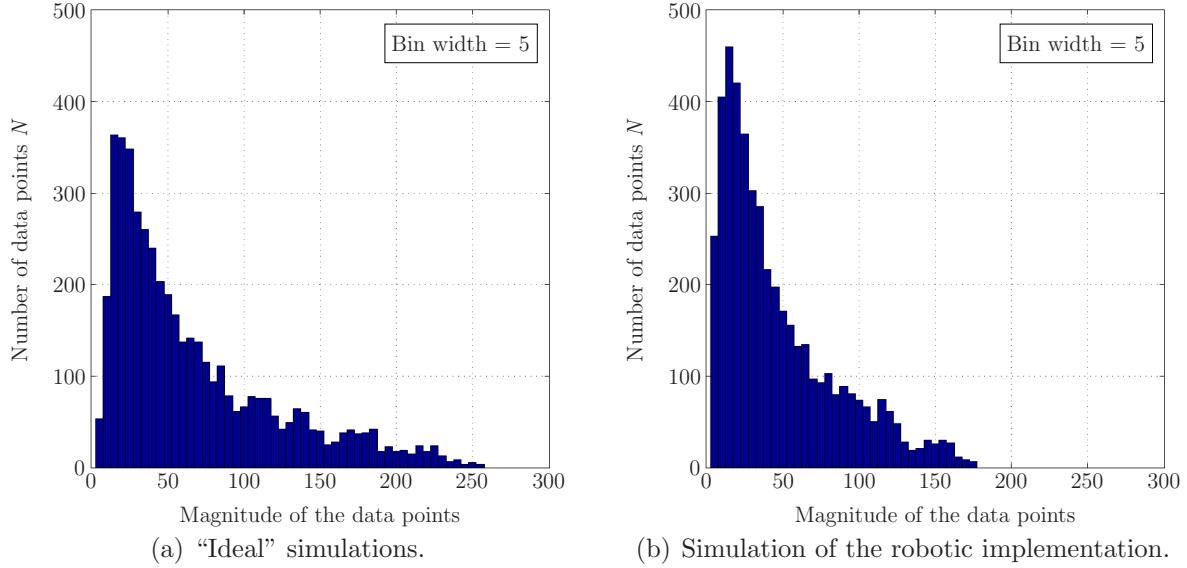
Figure 5.10(a) depicts the sound direction representation learned via unsupervised learning by the simulated robotic implementation. It is evident that the sound direction representation is quite similar to those depicted in Fig. 5.8(a) and Fig. 5.9(a). Once

again, the gradient of the estimated sound direction is considerably small with respect to the gradient of actual sound direction at the lower end of the frequency range, while at the higher end of the frequency range the gradient of the estimated sound direction matches that of the actual sound direction quite well. The gradient of estimated sound direction becomes more and more pronounced in a gradual manner while moving from the lower extreme to the higher one, again reflecting the characteristics of the auditory model's response. Furthermore, for any given frequency, the gradient of the estimated sound direction is quite linear as in the previous cases. However, at the lower extremes of the frequency range, the representational map tends to curve upwards, indicating that the estimated values of the sound direction in this region are not close to  $0^\circ$  as in the previous cases. This effect can be attributed to the compressed bounds of the training data employed in this simulation as described in Sect. 5.3.1. The training data employed in the previous simulations was bounded within the range  $[0, 255]$ , while in this case it was bound within the range  $[0, 175]$ . However, in all cases, the total number of training data points was 4625. Figure 5.11 depicts the distribution of these 4625 data points in both cases.



**Figure 5.10:** The sound direction representation learned via unsupervised learning in the simulation of the robotic implementation of the CMAC neural network in terms of the estimated sound direction.

The training data points that are of the lowest magnitude, i.e. the first bin in Fig. 5.11(a) and in Fig. 5.11(b), correspond to the lower extremes of the frequency range. These data points correspond to the auditory model's response at the lower end of the 1000 Hz–2200 Hz frequency range (refer to Fig. 3.4). In this region, the response is quite



**Figure 5.11:** Distribution of the training data in the “ideal” simulations and the simulation of the robotic implementation.

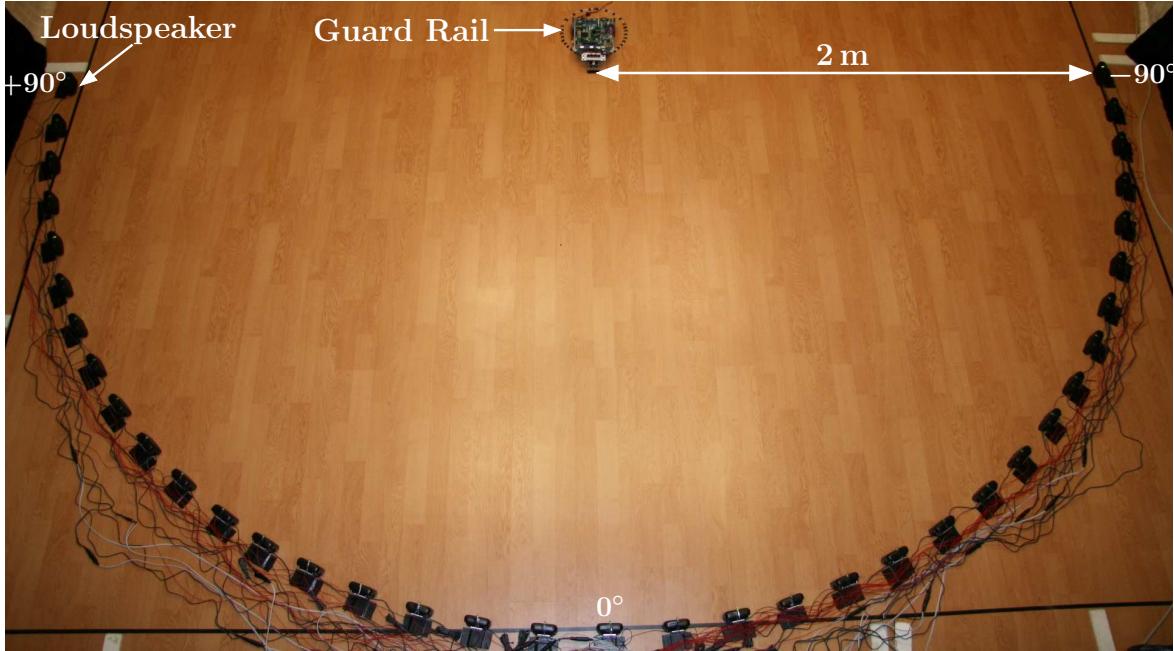
flat and significantly small. Therefore, the filterbanks also produce significantly small outputs as training data when presented with such inputs. Thus the summation of the weights that these training data points activate determines the network’s estimate of the sound direction for the lower extremes of the frequency range. It is evident that in the training data employed in the simulation of the robotic implementation, a larger number of data points (approximately 250) are of the lowest magnitude than those in the training data employed in the “ideal” simulations (approximately 50), i.e. there is a five-fold increase. As a result, during the training process the weights activated by these points are updated a disproportionately more number of times in the simulation of the robotic implementation, than in the “ideal” simulations. In one epoch of learning, the weights that determine the network’s estimate of the sound direction for the lower extremes of the frequency range are updated approximately 5 times as much in the simulation of the robotic implementation, than in the “ideal” simulations. As the learning progresses, this disproportionate update frequency results in the network’s estimate to be either incremented or decremented more times than required, depending on the order in which the training data points are presented to the network. In this particular instance, the order in which the training data points were presented was of course different in each epoch, but resulted in these weights being overall incremented more times than being decremented. This consequently results in the network’s estimate of the sound direction also increasing at the lower extremes of the frequency range, making the representational map to curve upwards as depicted in Fig. 5.10(a).

From Fig. 5.10(b) it is evident that the learning error settles quickly around  $26^\circ$  in a few epochs. Once again, for aforementioned reasons, the overall root-mean-square error  $\xi$  over all the frequencies and azimuthal displacements will not be near  $0^\circ$ . However, it shows an overall increase over the subsequent epochs. This is again due to the rise in the network's estimates of the sound direction at the lower extremes of the frequency range, which results in the mean-squared error for these frequencies increasing as well. Therefore, there is an overall increase in the root-mean-square error  $\xi$ .

### 5.3.2 Robotic Trials

#### 5.3.2.1 Experimental Setup

The robotic trials were conducted inside a sound-dampening chamber of dimension 3 m x 4 m. A total of 37 passive loudspeakers were affixed to the floor, arranged in a semi-circle from  $-\frac{\pi}{2}$  to  $+\frac{\pi}{2}$ , with one speaker every  $5^\circ$ . The robot was placed facing the loudspeakers at the centre of the hypothetical circle created by closing the semi-circle. The radius of the semi-circle was 2 m. Figure 5.12 depicts the experimental arena.



**Figure 5.12:** The experimental arena.

The robot had the freedom to rotate in either direction. It was assumed that the motors and the wheels steering the robot were not identical in their characteristics, and

therefore any rotational movement could also result in the robot drifting by a small amount either laterally or longitudinally or both. Therefore, a circular guard rail was placed around the robot so as to restrict its motion to a rotational one as much as possible.

The online learning trials were performed in an unsupervised manner resembling that of the simulation. A CMAC neural network with the generalization parameter  $\mathbf{C} = 80$  was implemented on the robot. All the weights were initialized to 0. The learning process consisted of two phases — *Estimation* and *Update*, executed in that order. In the *Estimation* phase, a tone of given frequency was emitted from a loudspeaker at given azimuthal displacement. The frequency and azimuthal displacement were chosen based on a shuffled table of all the possible combinations of the two. The robot then sensed the emitted tone and calculated the average power of the outputs of the left and right filterbanks. Two filters, one from each filterbank but with identical cut-off frequencies, were selected and their output powers were fed to the CMAC neural network as inputs. The filter was also selected based on a shuffled table containing the centre frequencies. The network calculated the memory addresses and summed up the addressed weights to generate an estimate of the sound direction. The robot then either rotated with a fixed speed to the left or right, or remained immobile, for a certain amount of time which was directly proportional to its estimate of the sound direction. As mentioned in Sec. 5.2.2, the range over which each individual weight can vary is  $[-512, 511]$ . Since  $\mathbf{C}$  weights are selected for each input vector, the range over which the output of the network can vary is  $[-512 \cdot \mathbf{C}, 511 \cdot \mathbf{C}]$ , or  $[-40960, 40880]$ . The robot's angular speed of rotation was calculated to be  $100^\circ/\text{sec}$  based on the radius of the wheels, thus requiring 1.8 s to complete a  $180^\circ$  turn. The range  $[-40960, 40880]$  was mapped to the estimated sound direction, or equivalently the rotational angle of the robot, by multiplying the network's output with a scaling factor  $F_{\text{scale}} = \frac{180}{|-40960|}$ . It is obvious that there is an unequal division of the range over which the weights can vary, but this was not expected to create any problems. This is because a difference of  $|-40960| - 40880 = 80 = \mathbf{C}$  between the absolute lower and upper bounds implies a angular difference of  $\frac{180 \cdot 40880}{|-40960|} \approx 0.3^\circ$  in the turning angle of the robot, and in reality the mechanics and of the robot's motors and the friction between the floor and the wheels would restrict such fine rotational movements. In fact, it was experimentally determined that the minimum possible angle of rotation, i.e. for which there was visible rotational movement, of the robot was approximately  $1^\circ$ .

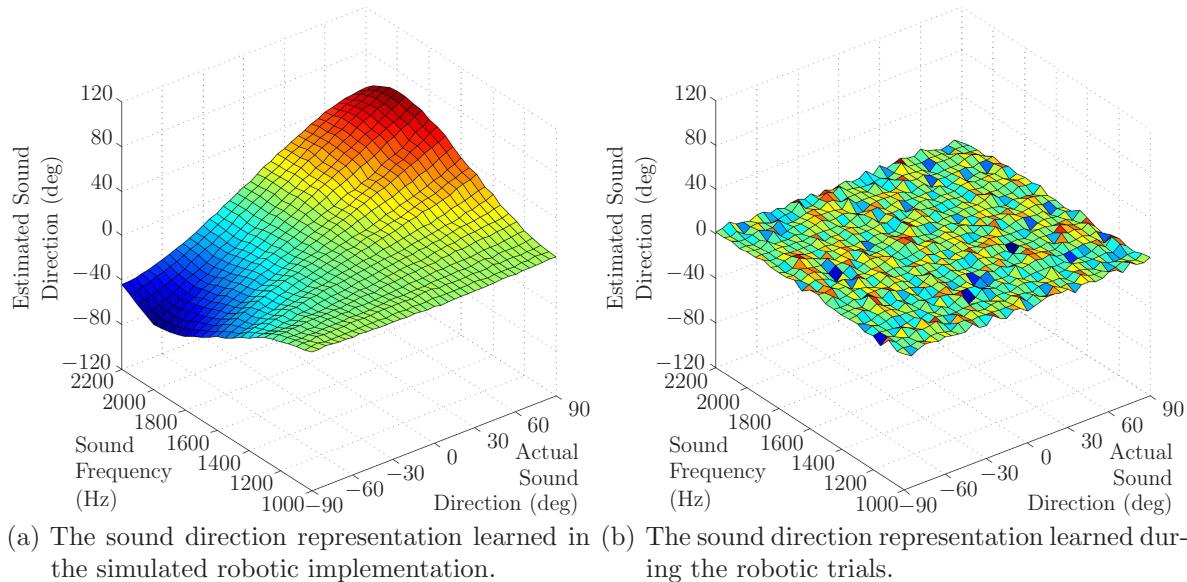
After the rotation was complete, the *Estimation* phase was terminated and the *Update* phase was executed. The emitted tone in this phase was the same as that in the *Estimation* phase. Here, the robot once again sensed the tone and recalculated the av-

verage power of the outputs of the left and right filterbanks. Then the output powers of the two filters selected in the *Estimation* phase were compared to each other. If the output power of the filter from the *left* filterbank was greater than that of the filter from the *right* filterbank then all the **C** weights addressed in the *Estimation* phase were *decremented* by 1, and if the output power of the filter from the *right* filterbank was greater than that of the filter from the *left* filterbank then the weights addressed in the *Estimation* phase were *incremented* by 1. And finally, if the both the output powers were equal, the weights were not modified. An increment or decrement of 1 ensured that the weights would be updated as slowly as possible, minimizing the possibility of overflow too soon. Care was taken to ensure that no individual weight value would overflow or underflow, by checking whether any update operation could result in an overflow or an underflow before the update was applied. If so, then that weight's value was frozen and there were no further updates allowed. This was done to ensure that the network would not “unlearn” the information it had learned so far. Even with such form of protection, if individual weights started to overflow or underflow very early into the learning process due to large increments or decrements, the network could very quickly reach a state where the learning algorithm would essentially settle into a cycle of learning and unlearning.

After the weights were updated, the network's output (i.e. the sum of the weights) for the current sound frequency and angular displacement of the sound source, was transmitted over the wireless link to an off-board PC for debugging and data-logging purposes and the *Update* phase was terminated, and a new *Estimation* phase was started. The training period was again 100 epochs, where one epoch meant the presentation of all the training vectors. Each phase lasted for 3 seconds and thus the actual time taken per training vector was 6 seconds. The remaining time in the *Update* phase was used to transmit the network's output to an off-board PC for debugging and data-logging purposes. Therefore, the total training time per epoch was  $\frac{4625 \cdot 6}{3600} = 7$  hours, 42 minutes and 30 seconds. The learning process was constantly monitored via the wireless link, and after every epoch, the process was frozen, and the robot's sound localization performance was evaluated by presenting a randomly chosen monotone stimulus from the 1000 Hz–2200 Hz range from a randomly chosen loudspeaker and observing the robot's behaviour. Simultaneously, the current sound direction representation as learned by the robot was visualized on the off-board PC.

### 5.3.2.2 Results and Discussion

It was discovered after the first epoch of training that the robot's performance was much worse than expected. The rotational movements of the robot in the subsequent test phase were quite small for all frequencies and angular displacement of the sound source. In order to determine the cause of this behaviour, the sound direction representation learned by the robot was examined and the intermediate representational map learned during the robotic trials was compared with that learned in the implementation-specific simulation to identify any differing characteristics. Figure 5.13 depicts the sound direction representation learned after one epoch, in the implementation-specific simulation and in the robotic trials.



**Figure 5.13:** The sound direction representation learned by the robot after one epoch.

The sound direction representation learned by the robot after one epoch appears to be random and without any logical structure as compared to its counterpart in simulation. The considerably smaller estimates of the sound direction explain the small rotational movements seen during the test phase after the first epoch. Subsequent training epochs did not improve the representational map.

All possible sources of external error were investigated first in order to identify the possible cause. The sound levels at the microphones were measured to ensure that all the loudspeakers were producing sound of equal loudness as measured at the robot. Next, the analogue auditory signals at the output of the microphones were examined for any possible noise that would affect the signal integrity, and also whether they were at the

appropriate voltage level, without any form of clipping. The digitized signals at the output of the analogue-to-digital converter were also examined to ensure that the digital versions were of the appropriate level and were a faithful reproduction of their analogue counterparts. No abnormalities were discovered at this level, and this also confirmed that there were no environmental characteristics such as echoes affecting the propagation of the audio signals.

After confirming the integrity of the auditory signals, the wireless link was examined for sources of errors, considering the possibility that the error might lie in the transmission medium and the post-processing of the data. A constant stream of frames of known data bytes were transferred over the link from the robot to the PC, the transmission was monitored for events such as dropped bytes or frames, framing errors, transmission delays and so on. This also allowed the investigation of the communication module in the robot's code. No errors were evident at this level.

From the above-mentioned investigation, it was concluded that there were no sources of error external to the robot that could be labelled as contributing factors in the failure of the learning process. This focussed the investigation on the robot's two main processing modules, namely the peripheral auditory model and the CMAC implementation, and on the learning algorithm itself (refer to 5.6). These sub-systems were systematically investigated both individually and cumulatively for possible errors as described in the following paragraphs.

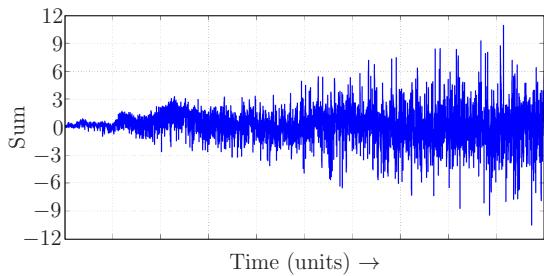
The peripheral auditory model implementation was examined for possible sources of error. To this end, a simulation model identical to the auditory model implementation on the robot in terms of the computations performed and the number representation used, was created and tested. The auditory signals used for testing the simulation model were derived from the digital versions recorded at the outputs of the analogue-to-digital converter for various sound frequencies and angular displacements of the sound source. The model calculated the powers at the output of the filterbanks as described in Sect. 5.2.1. These power values were compared to those obtained by the peripheral auditory model implementation on the robot, and it was confirmed that these were identical for various sound frequencies and angular displacements of the sound source. All the sub-modules in the auditory model such as the ipsilateral and contralateral filter implementations, the signal power calculations and the filterbank implementation (refer to Fig. 5.2), were tested as well for any discrepancies in their respective outputs, and it was confirmed that the simulation model did not deviate from the robotic implementation at the sub-module level either. In this manner, it was verified that the peripheral auditory model implementation on the robot was functioning appropriately.

Next, the CMAC implementation was investigated for possible sources of error. Once again, a simulation model identical to the CMAC implementation on the robot in terms of the computations performed and the number representation used, was created and tested. The simulation model of the peripheral auditory system was coupled to this model, in order to ensure that the correct information was being appropriately passed from one to the other and that the individual modules were cumulatively functioning appropriately. The inputs to the model were the output powers as described in Sect. 5.2.1, calculated by the simulation model of the peripheral auditory system. Since this model was verified as described above to produce outputs identical to those produced by the auditory model implementation on the robot, its outputs were perfectly credible as inputs to the CMAC simulation model. These outputs were fed to the CMAC simulation model as inputs, and its output i.e. the memory addresses for the **C** memory locations activated by the inputs, was calculated. These memory addresses thus obtained were compared with those calculated by the robotic implementation of the CMAC architecture, and it was confirmed that they were identical. The sub-modules in the CMAC implementation responsible for the calculation of the hypercubes and for mapping the hypercubes to the memory locations were also tested for any erroneous calculations and it was confirmed that the simulation model did not diverge from the robotic implementation in the intermediate computational level. In this manner, it was also verified that the CMAC implementation on the robot was functioning appropriately.

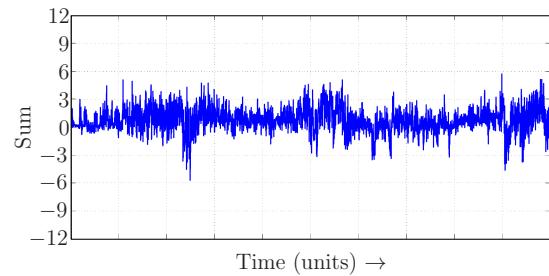
After testing and verifying the modules on an individual and on a cumulative basis, the learning algorithm was finally investigated to determine whether the source of error lay in its implementation on the robot. As before, a simulation model of the learning algorithm was created, identical to the implementation on the robot in terms of the computations performed and the number representation used. The auditory signals were also identical to the digital audio signals recorded at the outputs of the analogue-to-digital converter for various sound frequencies and angular displacements of the sound source. Furthermore, the order in which the auditory signals were presented in the simulation was also identical to that utilized in the robotic trials. It was also verified that the weights stored in the memory space were appropriately initialized to zero. This model coupled together the individual modules as depicted in Fig. 5.6 that were verified to be functioning properly as described in the previous paragraphs, and therefore this was essentially a simulation of the entire learning algorithm as it progressed during the robotic trials.

However, all the individual modules were once again monitored for any discrepancies, as well as all the internal variables that were used by the learning algorithm. The

simulation was run for one epoch, and the internal variables were compared with those in the implementation on the robot, logged during the trials. It was discovered that the network's output in the *Estimation* phase, which was the sum of the weights stored at addressed memory locations, was being erroneously calculated by the learning algorithm. This implied that either the calculations in the summation process or those in the actual update process where the weights were incremented or decremented were flawed. In order to identify the exact source of error, first the update process was disabled such that the weights were not modified under any circumstance and the simulation was reset with all the weights initialized to a fixed value of 512 for test purposes. It was then started and allowed to progress for one epoch, and the network's output was monitored for any alterations, but none were observed. If there were indeed discrepancies in the summation process, the summation of  $C = 80$  weights would have changed. However, the summed value remained constant at  $C \cdot 512 = 80 \cdot 512 = 40960$  throughout the simulation. This excluded the possibility that the summation process was erroneous, and suggested the update process as being flawed. The simulation was reset once again, and this time the weights were initialized to zero. The simulation was started and allowed to progress as usual for one epoch, and the network's output was observed. Figure 5.14 depicts the evolution of the network's output over one epoch of training in simulation and in the robotic trials.



(a) The sum of the weights stored at addressed memory locations in simulation.



(b) The sum of the weights stored at addressed memory locations during the robotic trials.

**Figure 5.14:** The evolution of the output of the network, i.e. the sum of the weights stored at the addressed memory locations, over time.

It is evident from Fig. 5.14(a) that in simulation, the network's output evolves gradually over time, fluctuating with steadily increasing bounds. However, from Fig. 5.14(b) that during the robotic trials, the network's output does not follow the same pattern. Instead, it rises sharply and then fluctuates within approximately constant bounds much smaller than those in simulation. Although the fluctuations in the network's output are to be expected as the learning algorithm increments or decrements the weights according to the gradient of the network's inputs, the width of the boundaries of the fluctuations

will reflect the characteristics of the network's architecture and the learning algorithm. When the first input data pair arrives, the network's output is zero in the *Estimation* phase, since the weights are initialized to zero, and in the subsequent *Update* phase the selected weights are updated. When the next input data pair arrives in the next *Estimation* phase, there may or may not be an overlap between the weights currently selected and the weights that were selected in the previous step. The minimum and maximum overlap between the set of **C** memory locations addressed, and thus the set of weights, by two separate input data pairs is one and **C** respectively. Since each weight is incremented or decremented by a factor of 1, the maximum shift in the magnitude of network's output in the beginning between two consecutively presented input data pairs can only be  $\mathbf{C} = 80$  i.e., if the presentation of both the data pairs results in the selection of identical weights. This of course, depends on the order in which the input data is presented, but this was identical in both simulation and during the robotic trials. The network's output should gradually increase as more and more input data pairs are presented and the weights become larger and larger in magnitude. Furthermore, the magnitude of the shift in the network's output between consecutive presentations of input data pairs should also increase over time for the same reason. These characteristics are evident in the simulation (refer to 5.14(a)) but are not observed during the robotic trials (refer to 5.14(b)). The network's output in fact shifts immediately to values of the order of  $10^3$ . According to the simulation results, shifts of such magnitude can occur only after approximately 250 consecutive presentations of input data pairs. It is logically impossible for a increment or decrement operation to generate such large changes when the increment/decrement factor is just 1. This suggested the possibility of a code compilation error residing in the Xilinx ISE. In order to confirm whether this was indeed a compiler error, the VHDL code was recompiled and programmed into the FPGA chip and the robotic trials were repeated once again, and similar results were obtained. This procedure was repeated multiple times, and each time the results were similar. This behaviour led to the conclusion that the VHDL compiler was flawed in some respect. Tracking down the bug in the compilation process was rendered impossible due to the fact that the Xilinx ISE toolchain comprises several compilers at various levels of abstraction between the VHDL design and the actual hardware level implementation, and furthermore all the software tools are proprietary in nature, and thus their source is not available to the general public. This implies that the programs can only be reverse-engineered at the machine-code level in order to understand the inner functional mechanism, and attempting to recognize and extract the logical structure in this manner would be a extremely complex task requiring a monumental amount of time and effort.

Even if this was achievable, tracking and fixing the actual bug would nevertheless require significant amount of time and effort. Furthermore, this is a task that must be delegated to the developers of the Xilinx software toolchain as they are much more capable, and are nevertheless expected to handle such issues. These findings have been duly reported to the appropriate authorities at Xilinx, and resolution is in progress.

## 5.4 Summary and Conclusions

In this chapter the peripheral auditory system was coupled with the Cerebellar Model Articulation Controller neural network to construct a plausible representational mapping of the sound direction. Two identical neural networks were created. Suitable values for the parameters pertaining to the network architecture were derived in simulation. One network was trained via supervised learning using the gradient descent approach in order to generate a reference representation of sound direction, and the other via reinforcement learning in an unsupervised manner to generate the same for comparative purposes. After validating the approach in simulation, the unsupervised learning approach was implemented on a mobile robot and robotic trials were conducted in order to incrementally learn the sound direction representation online. The sound localization performance of the robot was evaluated at intermediate points during the learning process.

It is evident from the simulation results that the system is able to construct a plausible representation of sound direction with auditory directional information generated by the peripheral auditory system. The representational maps built via supervised and unsupervised approaches are quite similar to each other. Furthermore, the results of additional experiments with the implementation-specific simulation model, which is identical to the robotic implementation with respect to the computational architecture and mechanism, validate the accuracy of the robotic implementation in simulation. However, the robotic trials were not fruitful, as is evident from the representational mapping built by the robot during the trials and the poor localization performance observed during the intermediate test phase. Further systematic investigation of the various components of the robotic implementation revealed abnormalities in the learning algorithm that produced errors in the update operation of the learning algorithm, which resulted in the weights being updated inaccurately. Further analysis of the debugging data logged during the trials led to the conclusion that the cause of the abnormality was an error in the compilation process itself. Given the complexity of the multi-tiered compilation process and

the proprietary nature of the compilation tools in the toolchain, any attempt to resolve such errors was rendered impossible from the user's end, and the task was relegated to the appropriate authorities, pending further investigation.

Based on the simulation results, it can be concluded that quality of the sound direction representation learned by the system is quite good. The representational map clearly reflects the characteristics of the input data from the peripheral auditory model presented to the CMAC neural network. The gradient of the representational map over the frontal semicircular region is quite linear for individual frequencies, and the gradually changing slope of this variation over the relevant range of frequencies is an accurate reflection of the input data characteristics. However, it is imperative to perform robotic trials to obtain definite proof that the simulation results can be accurately replicated in practice.



# Chapter 6

## Sensory Modulation of Motion

### 6.1 Introduction

Tetrapods such as salamanders and lizards exhibit an undulatory locomotion during terrestrial walking, characterized by a lateral bending of the animal's trunk during locomotion [108, 25, 101]. However, it has been observed that these animals tend to stabilize their heads in a way that it is always oriented towards the direction of motion. It can be hypothesized that this is done to minimize the oscillation of the head-centered reference frame for visual and auditory spatial information. In the absence of such stabilization, this reference frame would also oscillate along with the head, resulting in oscillating auditory cues and possibly confusing the animal. However, to the best of the author's knowledge, no behavioural study with either salamanders or lizards has been conducted that could corroborate this hypothesis.

Naturally, it can be argued that if these oscillating auditory cues are indeed a source of confusion to the animal, then it need only stop moving and sense the auditory information while remaining completely still. Once the direction of the sound source has been confirmed, then it could resume locomotion. Repeating such a behaviour in discrete steps would in effect neutralize the need for stabilizing the head. Again however, there has not been any behavioural study conducted that could support this argument. However, localizing a source while being stationary is a substantially more difficult problem than while moving.

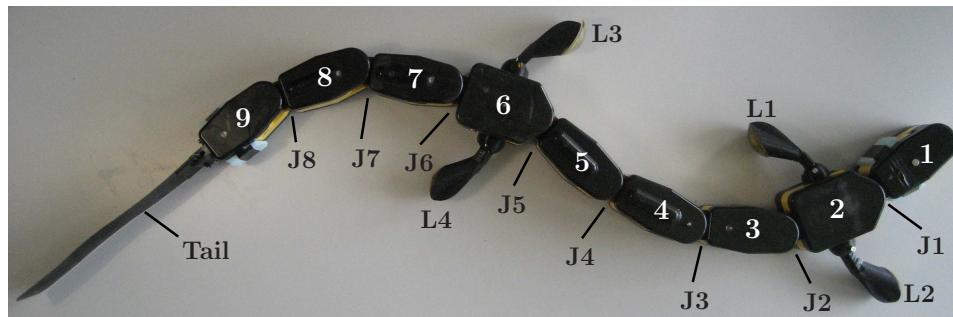
For the purpose of this work, the research question is framed in two parts in the following manner — (a) Is it possible to localize a static sound source during locomotion in the presence of oscillating auditory cues and in the absence of any form of head stabilization? and (b) If so, how well can the sound source be localized? In this chapter

these questions are investigated both in simulation and robotic experiments.

The work presented in this chapter covers and elaborates further upon the work reported by the author in [116].

## 6.2 *Salamandra Robotica II*

*Salamandra Robotica II* is a second generation quadruped mobile robot [63] developed in the Biorobotics Laboratory at the École Polytechnique Fédérale de Lausanne. It is primarily intended to act as a tool for studying the neuroscientific principles involved in quadruped locomotion [56, 57, 59, 60], especially in the salamander which is a model organism in the sense that its locomotor control circuitry is one of the most primitive in the animal kingdom, and has remained unchanged for thousands of years. The animal exhibits a rhythmic undulatory locomotion pattern as a result of the activation of neuronal Central Pattern Generators (henceforth referred to as CPGs) located in its spinal cord [18]. CPGs are known to exist in a number of animals as well as in humans, being responsible for generating different forms of rhythmic movements such as walking and breathing. Artificial CPGs have been implemented on a number of bio-inspired robots as well to generate suitable locomotion patterns and/on gaits. An exhaustive review of CPGs can be found in [58]. Figure 6.1 depicts the salamander robot.



**Figure 6.1:** *Salamandra Robotica II*.

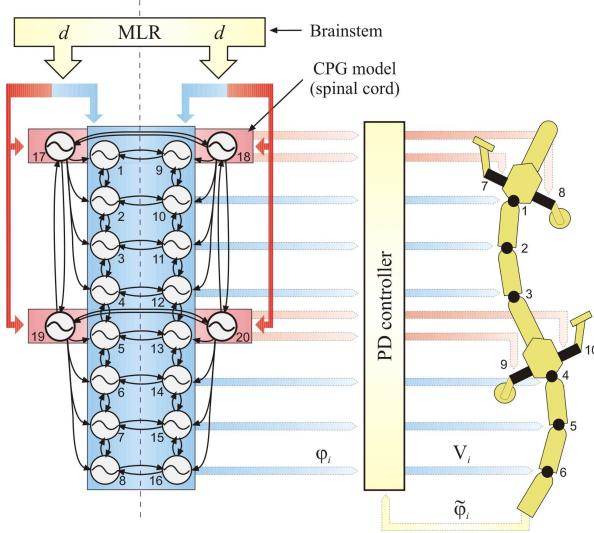
It is composed of nine modules (numbered 1–9), which can be classified into three types – passive modules, simple active modules and compound active modules. Six of the modules (numbered 3–5 and 7–9) are simple active modules and have identical mechanics and electronics. Two are compound active modules (numbered 2 and 6) and have two limbs each (forelimbs L1 and L2 on module 2 and hindlimbs L3 and L4 on module 6) on either side as additional mechanical attachments. The remaining one module (numbered 1) is a passive module, without any actuators built in. Module 1

represents the head, modules 3–5 represent the trunk and modules 7–9 represent the tail of the salamander. Modules 2 and 6 represent the anterior and posterior pelvic girdle respectively.

All active modules have a single servo motor, which controls a revolute joint at the front of the module with its axis of rotation in the vertical plane. All nine modules are mechanically linked to each other via these revolute joints. However, compound active modules have two additional actuators that control two revolute joints for the limbs on the left and right side. Locomotion patterns for the salamander robot are generated by an implementation of a neuronal model of the spinal cord CPG neurons on the robot, along with the subsequent motor drive signals that directly control the servo motors for each module. Simple active modules are driven by two body CPGs each, with one CPG controlling the right side of the body of the robot, while the other controls the left side of the body of the robot. Compound active modules are driven by four CPGs each, with two body CPGs controlling the left and right sides identical to the simple active modules, while the remaining two limb CPGs control the left and right side limbs. The body CPG oscillations for each module are phase-shifted by a certain extent from the body CPG oscillations for the preceding module, and these relative phase shifts dictate the relative angles of rotation of the revolute joints of each module with respect to the preceding module. For any given module, the instantaneous angle of rotation of its revolute joint is directly proportional to the instantaneous difference between the left and right body CPG oscillations. The first module (numbered 1 in Fig. 6.1) is a passive head module. A passive “tail” element is connected to the last module (numbered 9) to generate thrust while swimming underwater and plays no role during terrestrial walking. For the purpose of the work presented here, this element is ignored. Thus there are in total eight inter-module joints (labelled J1–J8). All the limbs have a continuous rotational motion with appropriate phase relationships between the limb movements and the movements of the revolute joint of the compound active module to which they are attached.

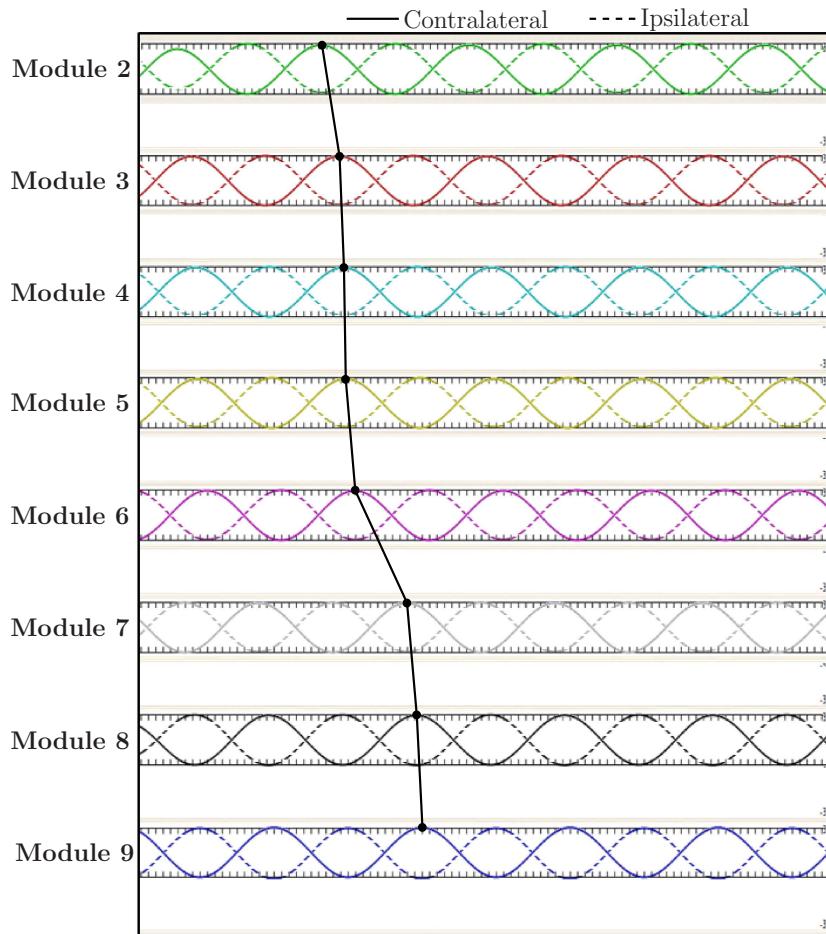
In order to achieve proper undulatory locomotion, proper phase relationships must be maintained between (a) the body CPGs of consecutive modules, (b) the left and right body CPGs of each module, (c) the limb CPGs, and (d) the body CPGs and limb CPGs overall. The left and right body CPG oscillations for each module must always be  $180^\circ$  out of phase with each other. Additionally, forelimbs L1 and L2 also must always be  $180^\circ$  out of phase with each other. Similarly, the hindlimbs L3 and L4 must always be  $180^\circ$  out of phase with each other as well. Finally, diagonally opposite limb pairs i.e. the pair L1 and L4 and the pair L2 and L3, must always be in phase with each other. The

requirement for producing a turning motion is that both the contralateral and ipsilateral side of the body must bend *towards* the desired direction of the turn. This is achieved by increasing or decreasing the amplitude of oscillations of the contralateral body CPGs relative to the oscillation amplitude of the ipsilateral body CPGs. For example, when the contralateral body CPG oscillations are larger in magnitude than the ipsilateral body CPG oscillations, the contralateral side of the body bends in the opposite direction but the ipsilateral side does not bend in the opposite direction, resulting in a turning motion towards the ipsilateral side. The radius of the turn depends on the magnitude of the contralateral oscillations with respect to the magnitude of the ipsilateral oscillations. The limb CPG oscillations need not be modified to evoke turning. Figure 6.2 depicts the CPG model implemented in the robot, with individual CPGs arranged in a double-chain structure as shown. The body CPGs in each module control the corresponding revolute joint via proportional-derivative (PD) controller. Figure 6.3 depicts sample body CPG oscillations on both the contralateral and ipsilateral sides for each module during locomotion in a straight line, highlighting the phase relationships.



**Figure 6.2:** The CPG model (figure by A. Ijspeert, courtesy Biorobotics Laboratory, EPFL).

Naturally, the locomotion mechanism in real salamanders is not strictly the same as in the robot. The salamander's neuromuscular system consists of a number of muscle segments running from head to tail. Terrestrial locomotion is achieved by a standing wave of smooth variation in the extent of expansion and contraction of laterally opposite epaxial and hypaxial muscle fibers all along the animal's body [8, 9, 97], producing lateral bending. This is also observed in lizards [102]. But in principle, this simultaneous expansion and contraction mechanism can be considered equivalent to having each



**Figure 6.3:** Contralateral and ipsilateral body CPG oscillations for modules 2–9 during locomotion in a straight line. The black line highlights the inter-module phase relationships.

discrete revolute joint in the robot being proportionally controlled by the instantaneous difference between the ipsilateral and contralateral body CPG oscillations of the module containing that particular joint. The smooth variation in the extent of expansion and contraction from head to tail is analogous to having varying phase difference in the body CPG oscillations between consecutive modules.

### 6.2.1 Shortcomings in the Robot Model

There are certain shortcomings of the robot that are significant enough to warrant attention. First and foremost, the mechanical design of the robot is quite weak. More concretely, the connectors on the revolute joints that couple individual modules together are in the form of four small metal screws mating with a plastic female counterpart with-

out any fastner to secure the connection. This implies that the undulatory motion can easily stress the connectors to the extent that the internal threads on the female counterpart can erode and the screws can simply slide outwards. Furthermore, the limbs are connected to the limb actuators on the compound active modules via a single metal screw and nut arrangement, which allows the limbs to be longitudinally folded in towards the body during swimming. During terrestrial locomotion, the limbs must be folded out. However, the folding and unfolding actions must be performed manually, and there is no locking mechanism present that can hold the limbs in a folded or unfolded position. This implies that during terrestrial locomotion when diagonally opposite limbs must rotate downwards and lift the body to some extent above the ground, there is no means of guaranteeing that the limbs will be completely unfolded when they strike the ground. Rather, the centrifugal forces generated during the rotational motion of the limb are expected to hold it in an unfolded state throughout the motion. This mechanism works fairly well as long as the robot is moving on a smooth surface such as a smooth wooden floor where the upward rotation of the limbs immediately after the stance phase (when the limbs are in contact with the ground) is unhindered, but it fails rather easily when there are irregularities on the surface, such as small bumps and crevices. The limbs easily tend to fold inwards just before the start of the upward rotation, at which point the limbs are in a position to fold longitudinally inwards but are still in contact with the surface, due to the forward movement of the robot and the additional friction created between the limb and the irregularities. Additionally, the limbs are mechanically quite weak as expected, being constructed entirely of plastic via 3D printing technology. Specifically, the hinges coupling the legs to the compound active modules are quite thin and easily stressed beyond their capabilities during locomotion, leading to serious mechanical damage. Unfortunately, the thermoplastic material used in non-industrial-grade 3D printers such as those typically used in research institutions is structurally weak when used to create small parts and is not suitable for designs in which large and sudden mechanical stress is commonplace.

Secondly, the design of the electronics and the firmware running on the robot is quite illogical in the author's opinion. For example, the electronic motor drivers for the limbs have rather sensitive over-current protection built in, which tends to shut all of the robot's electronic systems down completely in case the limbs are over-stressed or if their movement is hindered in any way. That in itself is sensible, but the only way restart the robot is to reboot all the modules' master microcontrollers — a process possible only by recharging the robot's batteries. This implies that if the robot breaks down, it can take a substantial amount of time to restart it.

Finally the robot has an autonomy of around 20 minutes of continuous operation, which implied that it had to be recharged quite frequently. The time taken to recharge fully was roughly two hours. All these issues presented themselves multiple times during the robotic trials described in this chapter, creating significant delays and rendering the robotic experiments an exercise in patience.

### 6.3 Design and Implementation

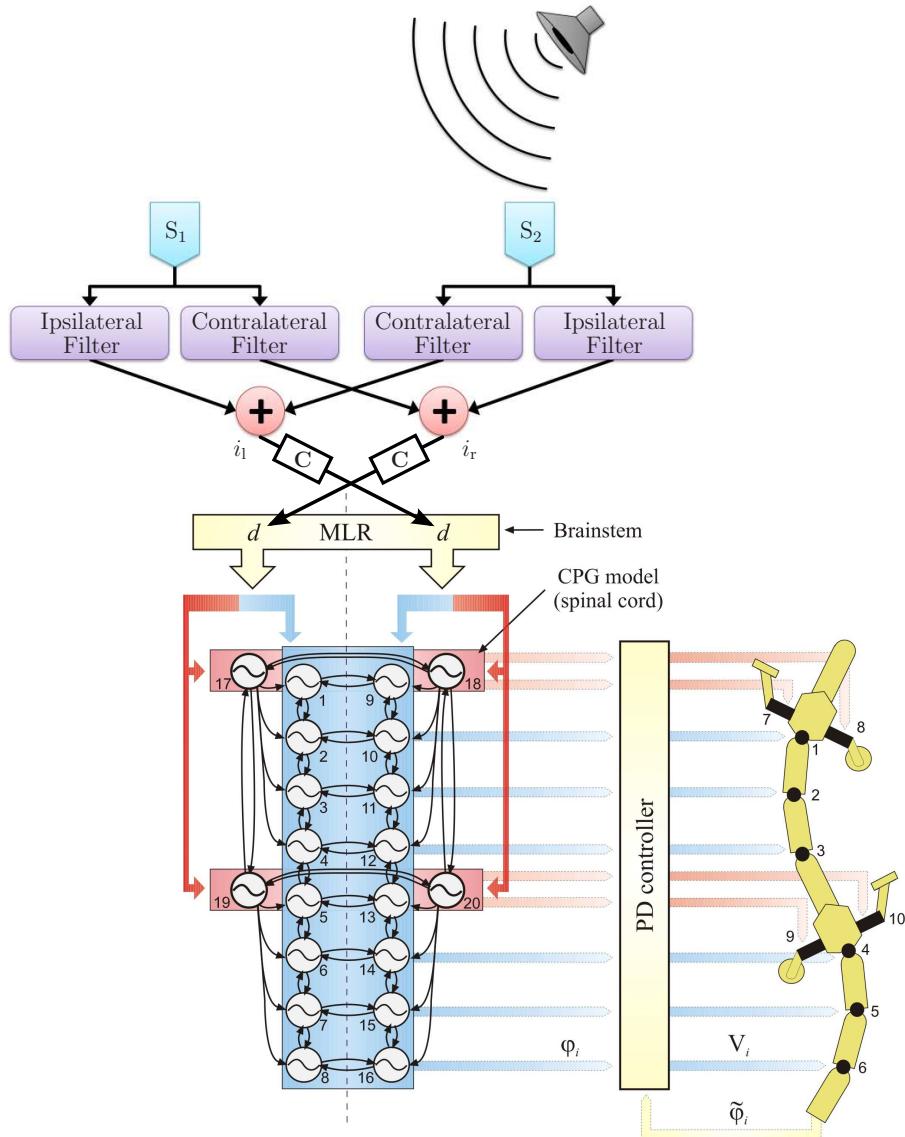
In order to investigate the effect of oscillating auditory cues on sound localization ability, first an appropriate experimental platform must be designed. The lizard's peripheral auditory system model and the CPG model of the salamander's locomotor system are coupled together to create a lizard-salamander hybrid model. As described in Sect. 3.3 Braitenberg vehicles are an attempt to describe conceptually the evolution of different behaviours as a consequence of modifying the structure of the sensorimotor couplings. In order to evoke phonotactic behaviour, the auditory model's outputs are cross-coupled to the drive inputs of the CPG model as shown in Fig. 6.4, effectively implementing a Braitenberg vehicle-like sensorimotor coupling.

The coupling **C** in Fig. 6.4 between the auditory model and the CPG model is realized in the form of the connection given by (6.1), with the *right* sensory output (i.e. the right tympanal vibration  $i_r$ ) modulating the body CPGs to the *left* side and the *left* sensory output (i.e. the left tympanal vibration  $i_l$ ) modulating the body CPGs to the *right* side.

$$\begin{aligned}\mu_l &= \text{sgn}(\|i_r\| - \|i_l\|) \cdot (1 + \alpha \cdot \text{sgn}(\|i_r\| - \|i_l\|)) \\ \mu_r &= \text{sgn}(\|i_l\| - \|i_r\|) \cdot (1 + \alpha \cdot \text{sgn}(\|i_l\| - \|i_r\|))\end{aligned}\tag{6.1}$$

where  $\|i_l\| = \frac{i_l}{i_l + i_r}$  and  $\|i_r\| = \frac{i_r}{i_l + i_r}$  are the normalized sensory outputs of the auditory model.

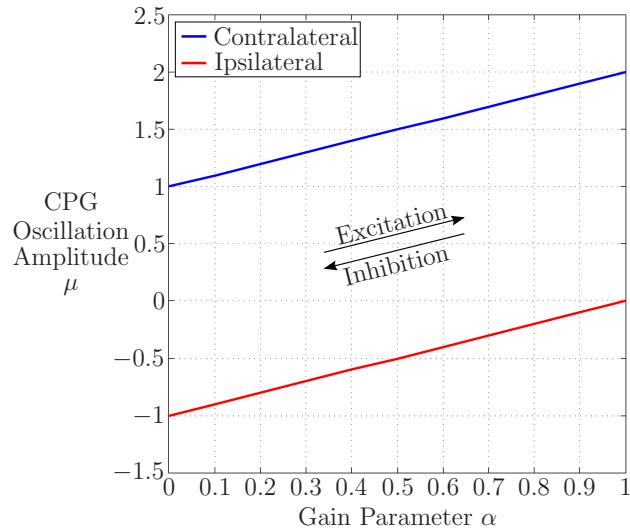
Given that turning motion is produced by having stronger CPG oscillations on the contralateral side than on the ipsilateral side, there are three possibilities — (1) make the contralateral CPG oscillations stronger, keeping the ipsilateral CPG oscillations unchanged, (2) make the ipsilateral CPG oscillations weaker, keeping the contralateral



**Figure 6.4:** The lizard-salamander hybrid model.

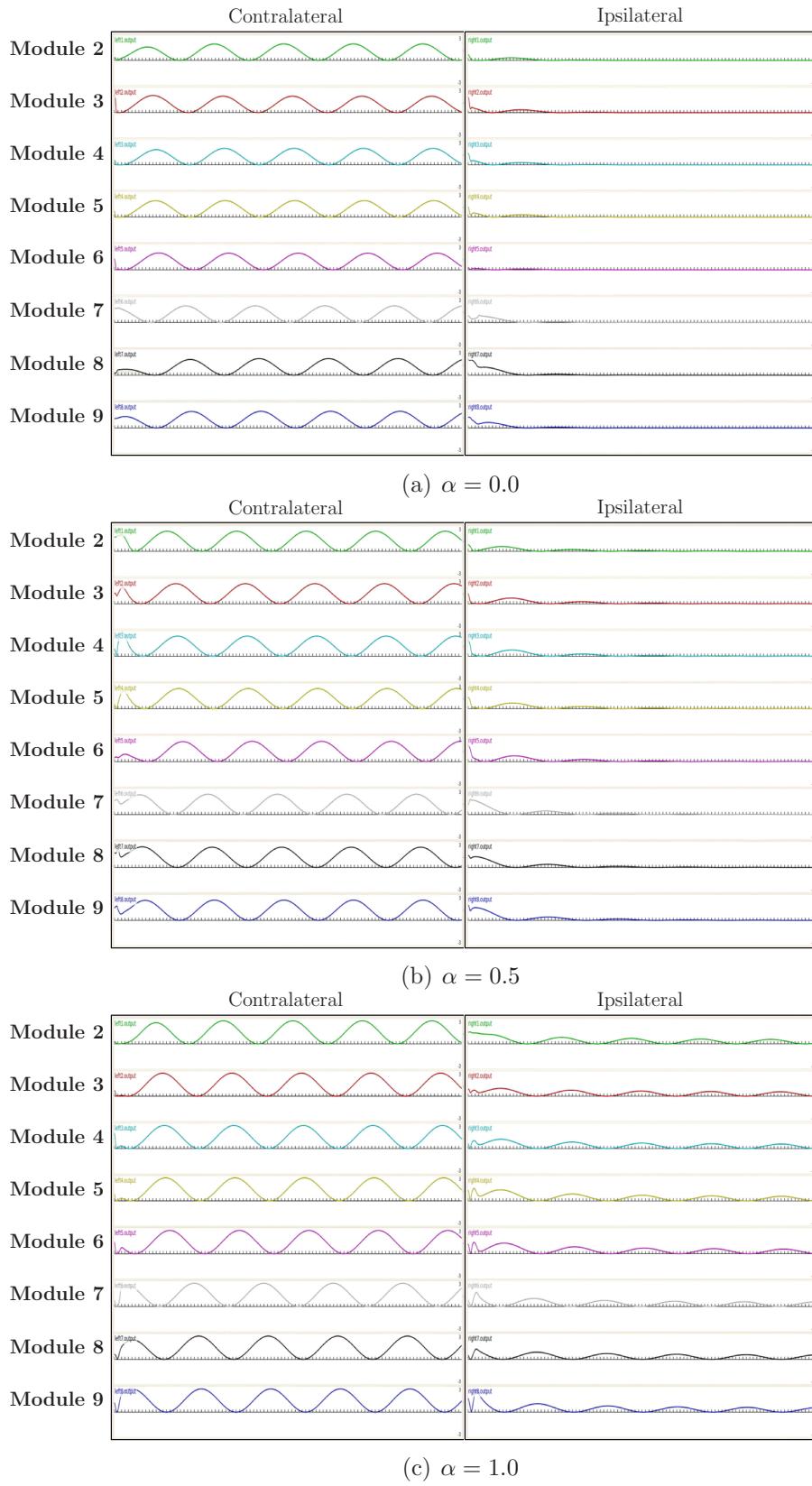
CPG oscillations unchanged, or (3) simultaneously make the contralateral CPG oscillations stronger and the ipsilateral CPG oscillations weaker. Here, the last option is chosen to ensure true Braatenberg couplings. The modulated parameters are the variables  $\mu_l$  and  $\mu_r$  which are proportional to the actual amplitudes of the left and right body CPG oscillations respectively. Decreasing  $\mu_l$  and/or  $\mu_r$  decreases bending of the trunk on the left and/or right side. This attenuates the head oscillations and thus the auditory cue oscillations. On the other hand, increasing  $\mu_l$  and/or  $\mu_r$  increases bending of the trunk on the left and/or right side. This amplifies the head oscillations and thus the auditory cue oscillations. In order to observe the effect of change in the amplitude of auditory cue oscillations and consequently change in the extent of bending of the trunk on the

sound localization ability, a gain parameter  $\alpha$  is defined which governs the extent to which  $\mu_l$  and  $\mu_r$  are modified. In fact,  $\alpha$  also determines how natural the motion of the robot is compared to that of a real salamander. The valid range of values for  $\alpha$  in (6.1) is  $[0.0, 1.0]$ . Since the variation in the response of the auditory system model varies for different frequencies, the sensory outputs  $i_l$  and  $i_r$  are normalized via division with the term  $i_l + i_r$  so as to generalize the sensorimotor coupling over the relevant frequency range of 1000 Hz–2200 Hz, making  $\alpha$  relatively insensitive to changes in sound frequency. Figure 6.5 depicts the effect of  $\alpha$  on  $\mu$  in general. Negative values of  $\mu$  imply inhibition while positive values imply excitation, and the magnitude of  $\mu$  is directly proportional to the level of excitation or inhibition.



**Figure 6.5:** The variation in the amplitude  $\mu$  of the contralateral and ipsilateral CPG oscillations with respect to the gain parameter  $\alpha$ .

Lower values of  $\alpha$  imply *weak* excitation and *stronger* inhibition of the contralateral and ipsilateral body CPGs respectively, and *higher* values imply *strong* excitation and *weaker* inhibition of the contralateral and ipsilateral body CPGs respectively. This results in tighter turns towards the sound source for lower values of  $\alpha$ , because the contralateral side of the body bends more *towards* the sound source and the ipsilateral side of the body bends less *away* from the sound source. Similarly, for higher values of  $\alpha$ , turns towards the sound source are wider because the contralateral side of the body bends more *towards* the sound source and the ipsilateral side of the body bends more *away* from the sound source. Figures 6.6(a), 6.6(b) and 6.6(c) from simulations highlight as an example the effect of varying  $\alpha$ , for  $\alpha = 0.0, 0.5$  and  $1.0$  respectively, on the contralateral and ipsilateral body CPG oscillations when the sound source is placed to the right side of the robot.



**Figure 6.6:** The effect of  $\alpha$  on the contralateral and ipsilateral body CPG oscillations.

It is evident that as  $\alpha$  increases, the inhibition of the oscillations of the ipsilateral body CPGs decreases, generating oscillations. At the same time, the excitation of the contralateral body CPG oscillations increases as well. This implies that the body becomes more flexible and tends to bend more, leading to wider turns.

## 6.4 Experiments and Results

### 6.4.1 Phonotactic Performance in Simulation

#### 6.4.1.1 Experimental Setup

In order to evaluate the phonotactic performance of the lizard-salamander hybrid model, a simple phonotaxis experiment was set up in the realistic physics-based Webots<sup>TM</sup> simulation environment [85], using a simulation model of the salamander robot (refer to Fig. 6.7) augmented with the peripheral auditory system model. It was assumed that the microphones that acted as the “ears” were mounted on the head module with a separation of 13 mm between them, similar to the separation between the tympanal membranes a typical lizard [21]. Since the auditory model’s parameters were derived based on experiments with the real animal, a different microphone separation would create a mismatch between the model and the external stimuli it expects.



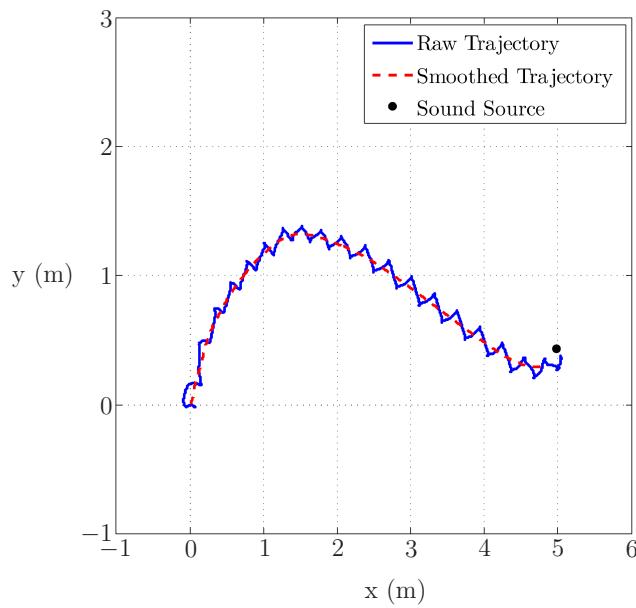
**Figure 6.7:** The Webots<sup>TM</sup> model of *Salamandra Robotica II*.

The goal the robot was to localize and locomote over a smooth surface free of obstacles, towards a simulated sound source of given frequency and angular displacement with respect to its head, placed 5 m away from the robot. However, Webots<sup>TM</sup> does not incorporate any physics-based model of sound propagation and therefore this was

modelled via geometrical means. Assuming the ideal condition of a noiseless transmission medium, the audio signals *expected* to arrive at the ipsilateral and contralateral microphones were generated in the robot's controller itself, with the appropriate phase shift between them calculated on the basis of the separation between the microphones, the sound frequency, the angular position of the sound source and the instantaneous orientation of the head. This was done at each time step, i.e. every 2 ms of virtual time (the minimum possible time step in Webots<sup>TM</sup>), thereby simulating approximately the spatio-temporal dynamics of the audio signals according to the robot's movements and head oscillations. The angular displacement with respect to the robot was varied over the frontal  $[-\frac{\pi}{2}, +\frac{\pi}{2}]$  region (from left to right) in  $5^\circ$  steps, resulting in 37 individual trials. The frequency was chosen to be 1900 Hz from the 1000 Hz–2200 Hz range because at this frequency the peripheral auditory system shows a strong response (refer to Fig. 3.4), and kept constant throughout all the trials. For each trial, the robot's head movements, body CPG oscillations, joint angles, the auditory model outputs  $i_l$  and  $i_r$  and the amplitudes  $\mu_l$  and  $\mu_r$  of the left and right body CPG oscillations were logged. Trials were terminated when the robot either hit the sound source or moved past it, whichever occurred first. If the robot hit the sound source it was recorded as a *hit*, and if it moved past it, it was recorded as a *miss*. For any given values of the user-specified variables in the simulation (sound frequency, angular displacement and  $\alpha$ ) the calculations involved in the sensorimotor coupling in the simulation environment were completely deterministic. The physical parameters of the environment relevant to locomotion were also fixed throughout the trials. Therefore, repeating a trial multiple times always produced identical results, eliminating the need for performing multiple trials for a given set of values of sound frequency, angular displacement and  $\alpha$  and focussing the experiments to be of a purely qualitative nature.

The performance criteria chosen were (a) a “success” map in terms of the *hits* over the entire range of angular displacement and  $\alpha$ , and (b) the absolute angular deviation of the recorded average trajectory heading from the ideal trajectory heading of  $0^\circ$  of the robot. The ideal trajectory heading was defined as a straight line from the robot to the sound source. The average heading of any given recorded trajectory was determined by first dividing the trajectory into 10 parts and computing the heading vector for each part. These heading vectors were then averaged to determine the average heading of the trajectory via the directional statistics approach [80]. Then, the absolute difference between the average heading and the ideal heading was computed. Since the raw trajectories were actually a log of the recorded head movements, and were thus oscillatory in nature, it was necessary to first convert these into smooth trajectories which could

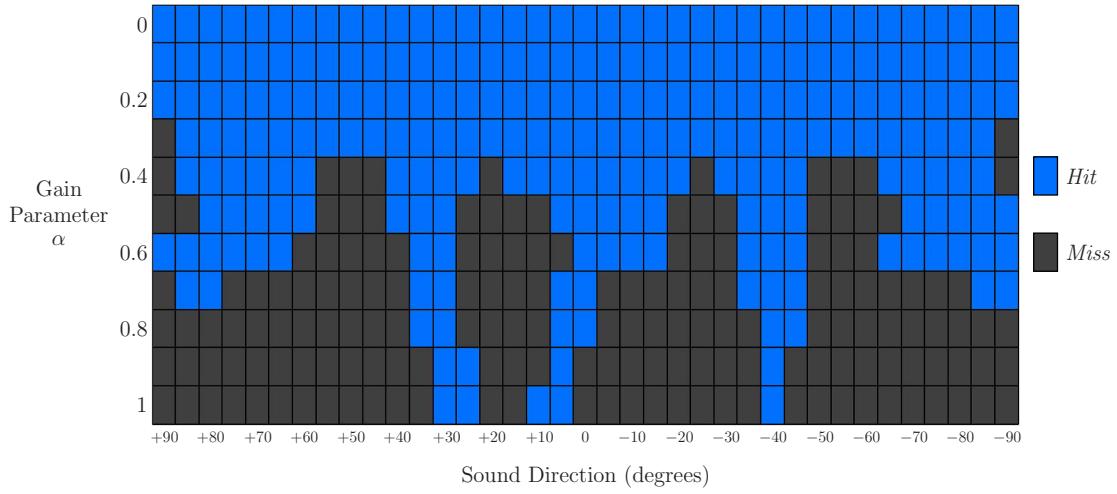
be used for determining the average trajectory headings. The raw trajectories typically consisted of approximately 20,000 data points. These trajectories were passed through a moving average filter with a fixed window size of 2000 data points calculated via trial-and-error to obtain smooth trajectories that sufficiently represented the robot's overall trajectory. Figure 6.8 depicts as an example the raw trajectory and the smoothed trajectory generated by the moving average filter for a trial with sound frequency of 1900 Hz, angular displacement of  $+85^\circ$  (to the right) and gain parameter  $\alpha = 0.4$ .



**Figure 6.8:** An example of a raw trajectory and its smoothed version.

#### 6.4.1.2 Results and Discussion

Figure 6.9 depicts the “success” map. Analyzing the performance from the point of view of variation in  $\alpha$  it is evident that for any given angular displacement, the performance is near perfect for lower values of  $\alpha$ , with very few *misses* over the entire  $[-\frac{\pi}{2}, +\frac{\pi}{2}]$  range. As a consequence of the robot making tighter turns, it is able to orient itself quickly in the direction of the sound source at which point it becomes fairly trivial to localize it. As  $\alpha$  increases the performance drops gradually as the robot makes wider turns, and for high values of  $\alpha$ , there are very few *hits*. Wider turns imply that the robot is not able to orient itself towards the sound source in time, and follows a spiralling trajectory around the sound source without ever coming in contact with it under the condition that the maximum bending of the trunk is insufficient to achieve the required radius of the turn.

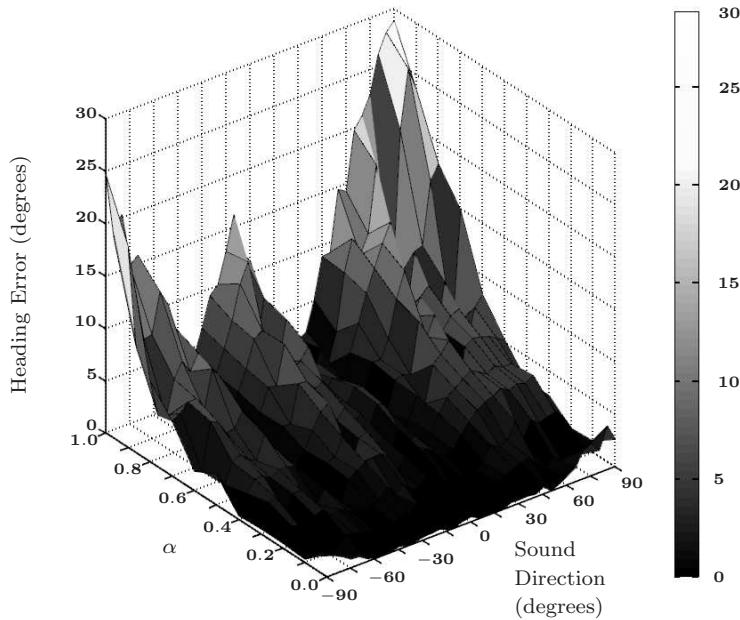


**Figure 6.9:** The phonotactic performance in terms of the “success” map.

From the point of view of the robot’s performance over the range of angular displacement it is evident that as  $\alpha$  increases, the performance starting from the extremes of the angular displacement and progressing inwards within the  $[-\frac{\pi}{2}, +\frac{\pi}{2}]$  range, begins to degrade. At the extremes, the robot must bend maximally in the ipsilateral direction and minimally in the contralateral direction in order to make as tight turns as possible. However, for higher values of  $\alpha$ , the robot’s trunk bends significantly in both directions, making it quite difficult for the robot to orient itself towards the sound source in time as it nears it. On the other hand, for lower values of  $\alpha$ , the trunk bends strongly towards the ipsilateral direction and weakly towards the contralateral direction, allowing tight turns and orienting the robot towards the sound source sufficiently early in time for successful localization. Progressing towards the centre of the range, i.e.  $0^\circ$ , for lower values of  $\alpha$  the robot’s trunk is relatively stiff on either side, and thus it is able to orient itself quickly towards the sound source. Once again, for higher values of  $\alpha$  the robot’s trunk is quite flexible and bends significantly on either side, making it difficult to orient towards the sound source in time. For lower values of  $\alpha$ , the trunk bends more towards the ipsilateral side and less towards the contralateral side, requiring less time to orient itself towards the sound source. However, it was observed that the initial conditions of the CPGs evoked some irregularities in the behaviour for large values of  $\alpha$ . It is evident from Fig. 6.9 that even for high values of  $\alpha$ , the robot’s performance was reasonably good for angular displacements of around  $+30^\circ$  and  $-40^\circ$ . Upon further investigation it was discovered that when the simulation was started, the robot’s body and limb CPGs would fire immediately as expected, but it required some time for them to settle into the correct phase relationship between each other. This caused the robot’s body to start

undulating on-the-spot before the limbs could propel it forward in the proper manner. In fact, the robot would tend to orient itself at approximately  $+25^\circ$  first and then at approximately  $-45^\circ$  before the CPG oscillations settled. This created biases in the simulation model that favoured angular displacements of the sound source that were in the general direction of these initial orientations.

Figure 6.10 depicts the absolute angular deviation of the recorded average trajectory heading from the ideal trajectory heading, over  $[-\frac{\pi}{2}, +\frac{\pi}{2}]$  for  $\alpha = 0.0, 0.1, \dots, 1.0$ . It is evident that for lower values of  $\alpha$  the angular deviation is almost  $0^\circ$ , while for higher values, the angular deviation steadily increases. Tighter turns for low values of  $\alpha$  imply that the robot is able to orient its body towards the sound source much earlier in time, as it nears the sound source. Once the body is oriented directly at the sound source, locomotion occurs in a straight line. Therefore, a relatively smaller percentage of the trial time is spent in turning, and a relatively larger percentage is spent in straight walking, so that the overall deviation in heading is small. Wider turns for high values of  $\alpha$  imply that the robot orients its body towards the sound source much later in time as it nears the sound source. Therefore, a relatively larger percentage of the trial time is spent in turning, and a relatively smaller percentage is spent in walking straight, so that the overall deviation in heading is large.

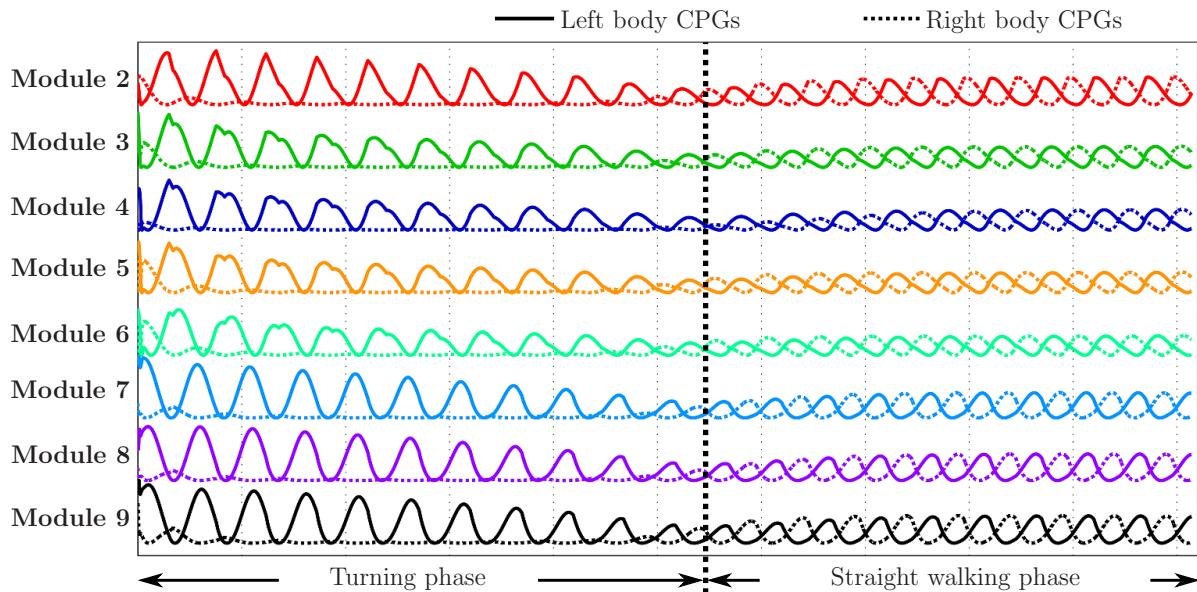


**Figure 6.10:** Absolute angular heading error.

Again, from the point of view of the range of angular displacement, it is evident that for any given value of  $\alpha$  the absolute angular deviation in heading is higher at the extremes and steadily declines when progressing inwards to the centre of the  $[-\frac{\pi}{2}, +\frac{\pi}{2}]$

range. The slope of this declination is low for lower values of  $\alpha$ , and steadily increases as  $\alpha$  increases.

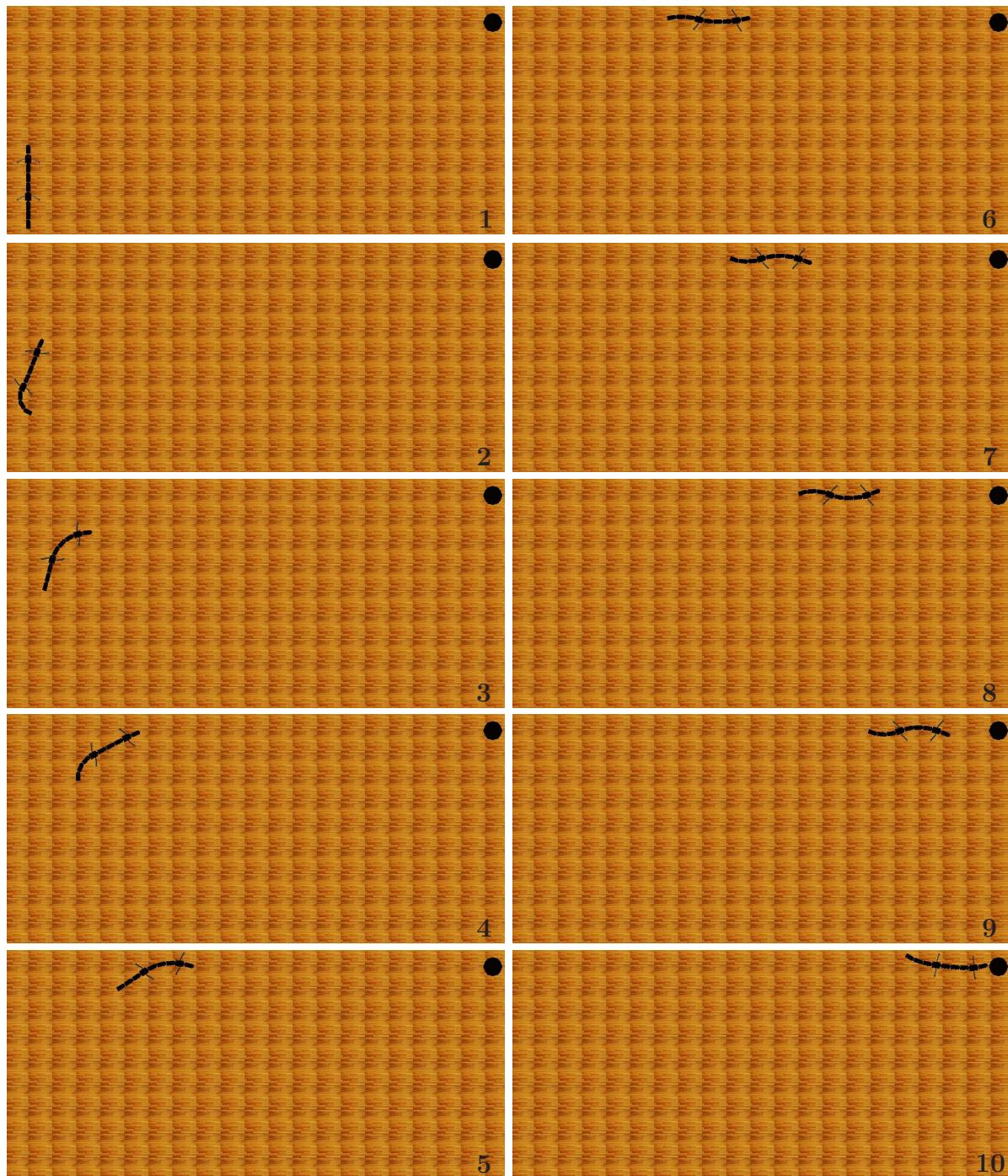
Figure 6.11 depicts as an example the body CPG oscillations for all the modules 2–9 during the locomotion of the lizard-salamander hybrid robot while localizing a sound source (the black dot) of frequency 1900 Hz placed 5 m from the head of the robot at an angular displacement of  $+75^\circ$  (to the right), for  $\alpha = 0.2$ . It is evident that while turning, the contralateral side of the body bends more *towards* the sound source and the ipsilateral side bends less *away* from the sound source, since the contralateral body CPG oscillations are amplified while the ipsilateral body CPG oscillations are attenuated. When the robot has oriented itself completely towards the sound source, the contralateral and ipsilateral bending are equal in magnitude due to equal amplitudes of the contralateral and ipsilateral body CPG oscillations.



**Figure 6.11:** An example of the CPG oscillations for modules 2–9 during the locomotion pattern.

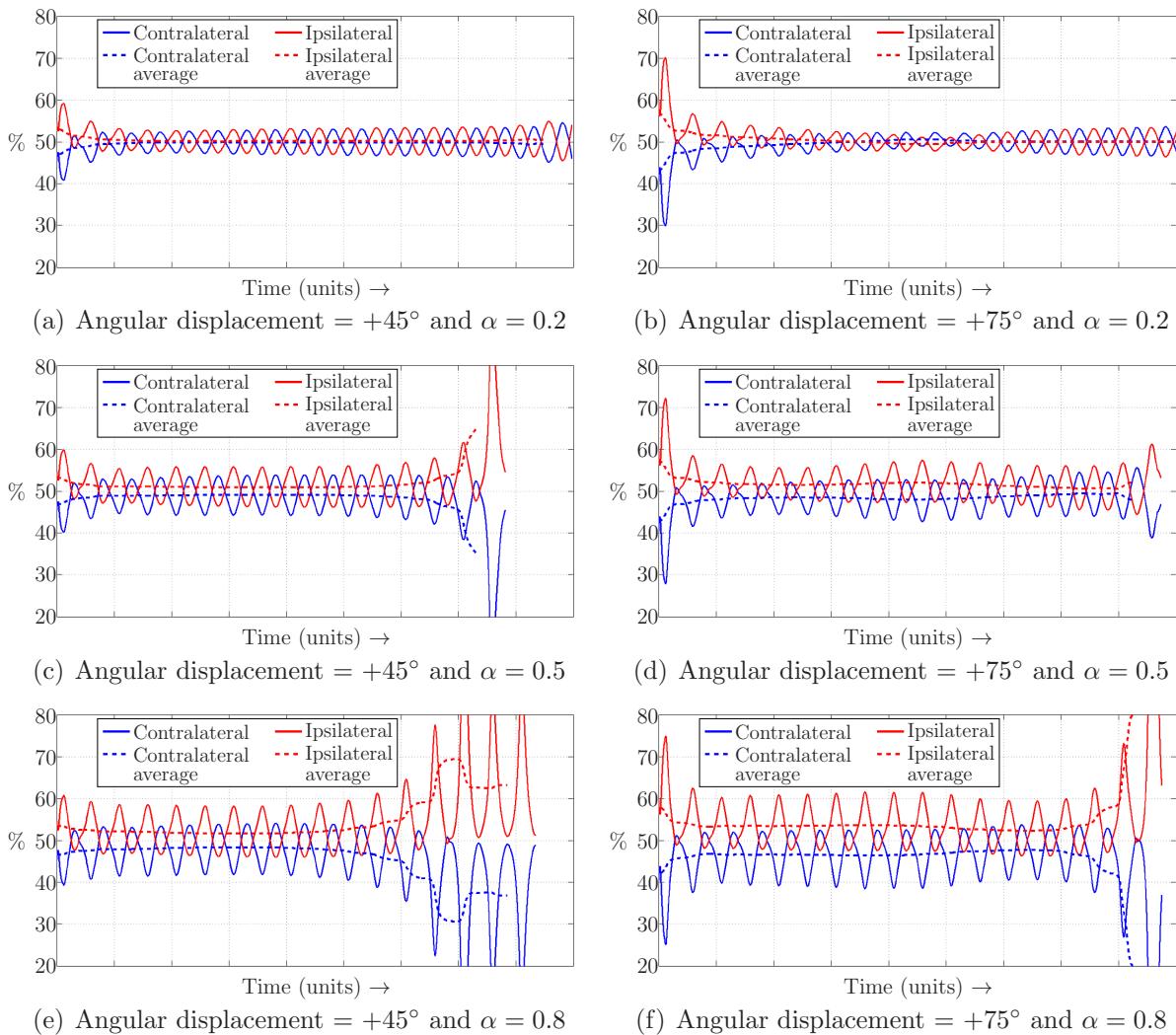
Figure 6.12 depicts the locomotion pattern for the aforementioned example. The locomotion pattern can be broken down into two phases — a turning phase and a straight walking phase. The robot initially turns right as the sound source is always to the right of the robot's head irrespective of its oscillatory movement. This generates asymmetrically oscillating auditory cues, and this asymmetry decreases the longer the robot stays in the turning phase. After some time the robot is oriented such that its body is pointing in the general direction of the sound source, and now the head movement pattern is such that the sound source appears to oscillate to either side relative to the

head. This generates symmetrically oscillating auditory cues centered at the mid-line of the robot.



**Figure 6.12:** Snapshots of the locomotion pattern of the lizard-salamander hybrid robot in simulation. Figures labelled 1–5 depict the turning phase and figures labelled 6–10 depict the straight walking phase.

Figure 6.13 depicts as an comparative example the dynamics of the normalized relative tympanal vibration strengths on contralateral and ipsilateral side, which depend directly on the dynamics of the corresponding oscillating auditory cues generated due to the oscillatory head movements during the robot's locomotion in simulation, for  $\alpha = 0.2, 0.5$  and  $0.8$  and angular displacement of  $+45^\circ$  and  $+75^\circ$  (both to the right). In the beginning of the trials when the robot is turning, on an average the ipsilateral tympanal vibration strength is greater than that on the contralateral side. In these situations, for a given  $\alpha$ , a larger angular displacement of the sound source implies a greater average ipsilateral tympanal vibration strength as compared to the contralateral side. In similar manner, for a given angular displacement, increasing  $\alpha$  leads to greater average ipsilateral tympanal vibration strength than on the contralateral side.



**Figure 6.13:** The dynamics of the relative tympanal vibration strengths on the contralateral and ipsilateral sides during locomotion.

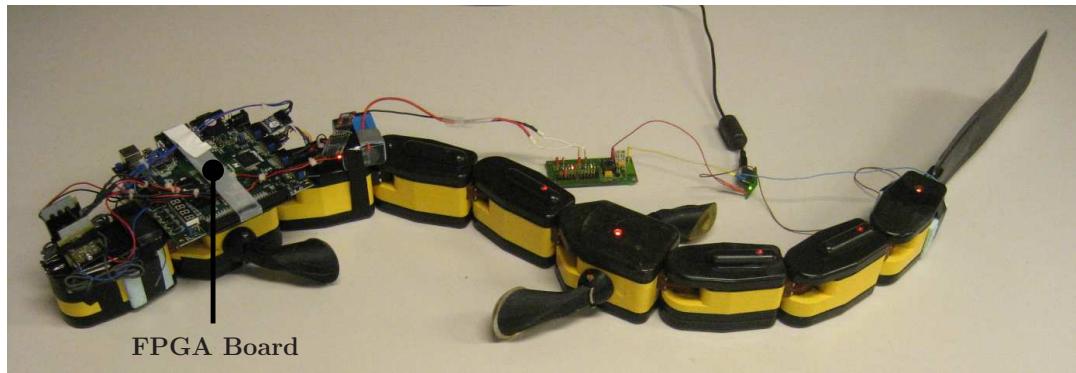
At some point during the trials, the robot tends to orient itself towards the sound source, given that  $\alpha$  is sufficiently low as depicted in Fig. 6.13(a) for example ( $\alpha = 0.2$  in this case). Once that occurs, the head oscillates equally to either side. In this case, the average ipsilateral and contralateral tympanal vibration strengths become equal. However, if  $\alpha$  is increased ( $\alpha = 0.5$  for example) the robot exhibits more pronounced bending of the body on *both* sides while turning, resulting in wider turns. Due to this, the robot is unable to orient itself sufficiently towards the sound source, and the source still remains displaced with respect to the robot. Therefore, the average ipsilateral tympanal vibration strength is still greater than that on the contralateral side as depicted in Fig. 6.13(c). If  $\alpha$  is increased further, the bending of the robot's body is even more pronounced on both the sides, leading to even wider turns. Naturally, the robot is unable to sufficiently orient itself towards the sound source, and now the source is displaced even more with respect to the robot. Therefore, the difference between the average ipsilateral and contralateral tympanal vibration strengths is even greater than before as depicted in Fig. 6.13(e). This trend can be seen for both  $+45^\circ$  and  $+75^\circ$  angular displacement. Finally, for a sufficiently low  $\alpha$  ( $\alpha = 0.2$ ), at some point towards the end of the trial the robot successfully localizes the sound source by coming in physical contact with it. However, for  $\alpha = 0.5$  and  $0.8$ , the robot is unable to localize successfully and starts to move past it, although in close proximity. In such situations, even though the robot is quite close to the sound source, the source is near its maximum displacement of  $+90^\circ$  to the right with respect to the robot. As the robot moves past it, any small change in the head position implies in a dramatic change in the auditory cues and subsequently in the average ipsilateral and contralateral tympanal vibration strengths. This leads to a sharply rising difference between the two at the end of the trials.

## 6.4.2 Robotic Experiments

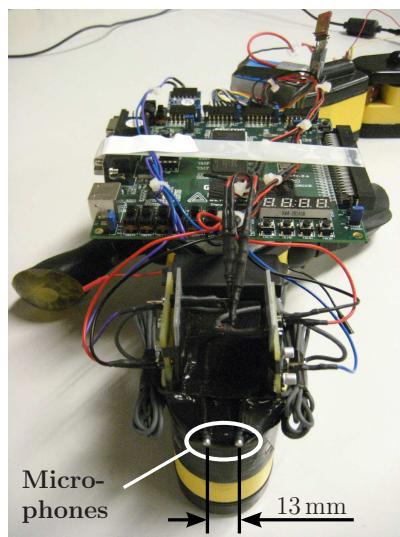
### 6.4.2.1 Experimental Setup

After the characterization of the lizard-salamander hybrid model in simulation, its performance from the point of view of the effect of varying the gain parameter  $\alpha$  on the locomotion trajectory was tested in simulation and via robotic experiments. A Field Programmable Gate Array (FPGA) board (Nexys<sup>TM</sup> 2 from Digilent Inc.) was mounted just behind the head module on the anterior pelvic girdle of the robot as shown in Fig. 6.14(a). The auditory model was implemented on the Xilinx Spartan-3E XC3S500E FPGA chip, with a gate count of 500K, available on the board. External auditory sig-

nals were passed as input to the board via two omni-directional microphones (model FG-23329-P07 from Knowles Electronics) mounted, 13 mm apart (again, similar to the distance between the tympanal membranes of a typical lizard for the aforementioned reasons), on the head of the robot as shown in Fig. 6.14(b).



(a) The hybrid robot.



(b) Microphone placement.

**Figure 6.14:** The lizard-salamander hybrid robot.

A dual channel 12-bit analogue-to-digital converter was utilized to digitize the signals from the microphones before passing them on to the FPGA chip. The processed scalar outputs of the auditory model, i.e. the left and right tympanal vibration strengths  $i_l$  and  $i_r$ , were continuously transmitted via a 2.4 GHz RF wireless transceiver (nRF2401A from Nordic Semiconductor Inc.) to an off-board PC. The calculations involved in the sensorimotor coupling given by (6.1) were performed off-board as well. A supervisory control program on the PC integrated the wireless communication between the robot and the PC and the abovementioned calculations together. It implemented a control

loop that accepted the outputs of the auditory model, calculated the appropriate values for the left and right body CPG oscillations ( $\mu_l$  and  $\mu_r$  respectively) and transmitted these values to the robot via a second 868 MHz RF wireless transceiver (nRF905 from Nordic Semiconductor Inc.). The loop performed data-logging duties as well. This sequence of steps was repeated for any new and valid values for the auditory model outputs received from the robot. This control loop was configured to run as fast as possible, so the most significant delays were the ones caused by the quality of the two wireless links. The execution time for one loop cycle was approximately 4–5 ms excluding only the delays caused by anomalies in the wireless links, which were unpredictable.

Two equivalent phonotaxis experiments were set up in the Webots<sup>TM</sup> simulation environment and in an anechoic chamber to provide a comparison of the robot's performance in simulation and robotic trials. However, there were certain differences between the simulation and the real-world environment that made direct comparisons infeasible. Firstly, the calculations in the FPGA implementation of the auditory model on the robot were less accurate than those in simulation. This can be attributed to the fact that the implementation in simulation was completely devoid of any form of auditory noise, since the environment's programming model does not offer any means to simulate the physics of sound propagation. Therefore, the sound source was simulated geometrically rather than physically, by generating pure sinusoid signals appropriately phase-shifted depending on the spatial position of the two microphones relative to the spatial position of the sound source. Also, there was no model of internal electronic noise typically seen in real-life electronic systems. Secondly, the delays experienced at the interfaces (both at the hardware and software levels) between the two communication channels, the PC and the robot were not modelled in the simulation. This implied that the control loop frequency in simulation was much faster than that in the real-world scenario. Finally, the physical characteristics of the floor of the experimental arena in the simulation, imitating a smooth wooden surface, did not match those of the floor in the anechoic chamber which was composed of multiple interlinked tatami mats. These mats are manufactured from closed-cell bidensity polyethylene with a PVC layer surface and rice-straw embossing in a cross-hatch pattern for high friction. Since the robot's ventral side is always in contact with the surface of the floor during locomotion, it can be argued that this high friction, characteristic of these mats, could affect the locomotion pattern of the robot by restricting its motion to a much greater degree than the smooth wooden floor in simulation. All these differences together restricted the comparison between the robot's performance in simulation and in practice to one of a qualitative nature.

The goal of the robot was again to localize and locomote towards a monotone sound

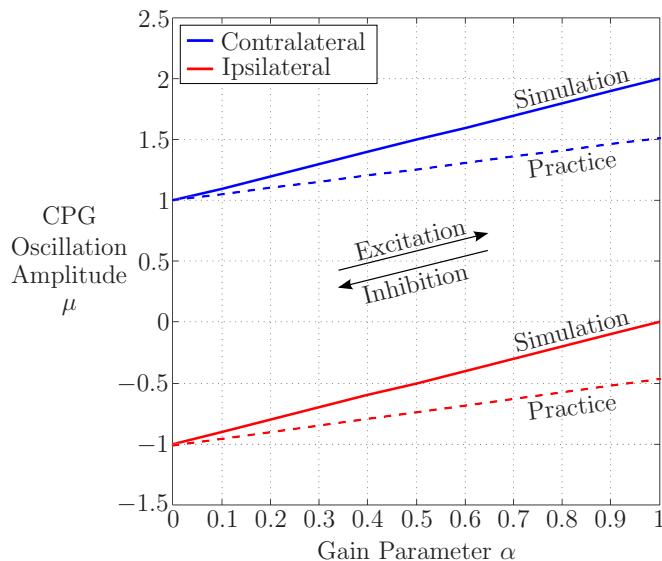
source placed at a fixed angular displacement of  $-70^\circ$  (on the left side) relative to the robot. Two independent sets of trials were performed for two different sound frequencies, namely 1900 Hz and 2000 Hz, chosen from the 1000 Hz–2200 Hz range. The choice of these particular values was motivated by the fact that over the range  $\left[\frac{-\pi}{2}, \frac{+\pi}{2}\right]$  the variation in the auditory model’s response , i.e. the range over which  $i_l$  and  $i_r$  (and subsequently  $\|i_l\|$  and  $\|i_r\|$ ) vary, is greatest around 2000 Hz. This implies that for frequencies in this neighbourhood the auditory model’s response generated during the robot’s oscillatory head movements is of high resolution, and small changes in the head orientation can elicit a different response. This renders the coupling  $\mathbf{C}$  more sensitive to the dynamics of auditory cues, which is required at low values of  $\alpha$  when the head oscillations and thus the auditory cue dynamics are repressed considerably. Unfortunately, the shortcomings mentioned in Sect. 6.2.1, resulting in the robot suffering mechanical damage to the limbs multiple times, and the heavily restricted time frame available to perform the experiments negated any possibility of repeating the experiments for a broad range of sound frequencies. To put matters in perspective, the robot was available for experimentation for only two days, and it was estimated that the overall period of time in which the robot functioned without problems and credible experimental data was recorded was in the order of minutes

In order to demonstrate the qualitative effect of modifying the gain parameter  $\alpha$  on the robot’s trajectory, three independent trials were performed in each set, for different values of  $\alpha$ . For both frequencies,  $\alpha$  was varied within the range [0.0, 1.0] in steps of 0.1. However, for 1900 Hz the choice of  $\alpha$  had to be restricted to 0.1, 0.3 and 0.7, while for 2000 Hz  $\alpha$  was restricted to 0.0, 0.3 and 0.6. The reason for this restriction was two-fold. Firstly, the shortcomings of the robot caused it to break down a great number of times during the trials and the restricted experimental time frame necessitated maximizing the time available and minimizing time spent in repeating a trial where the robot had broken down, until it functioned properly. In fact, a significant number of the trials were unsuccessful as a result of the robot breaking down. For the above-mentioned values for  $\alpha$ , the robot functioned without problems.

Secondly, small changes in  $\alpha$  did not produce clearly distinguishable changes in its locomotion pattern. This was because in the CPG model implementation on the robot, the range of values for the amplitudes of body CPG oscillations were more conservative than those in the simulation, and therefore less sensitive to small changes in  $\alpha$ . Unfortunately, the upper and lower limits of the range of values for the amplitudes of body CPG oscillations were pre-calculated to lie in a “safe zone” in which the robot’s mechanics and electronics would not be over-stressed, and then hard-coded in the firmware of the

robot. This rendered it impossible to increase the range to match that in the simulation.

Figure 6.15 depicts the difference in the variation of  $\mu$  with respect to  $\alpha$  in simulation and on the actual robot. The reason for such a conservative implementation was that certain constraints to maintain mechanical stability had to be observed while actuating the rotary joints of robot, which are not present in the Webots<sup>TM</sup> model of the robot. Therefore, it had to be ensured that the joints, which are constructed primarily of plastic, were not stressed by strong or fast actuation as this could easily result in severe mechanical damage. Therefore, the task of differentiating the locomotion patterns with minor changes between them was rendered considerably difficult, as the patterns appeared almost identical. In order to obtain clearly distinguishable locomotion patterns, the choice of  $\alpha$  was restricted to values sufficiently far apart from each other.



**Figure 6.15:** The difference between simulation and robotic implementation, in the variation of the amplitude  $\mu$  of the contralateral and ipsilateral CPG oscillations with respect to the gain parameter  $\alpha$ .

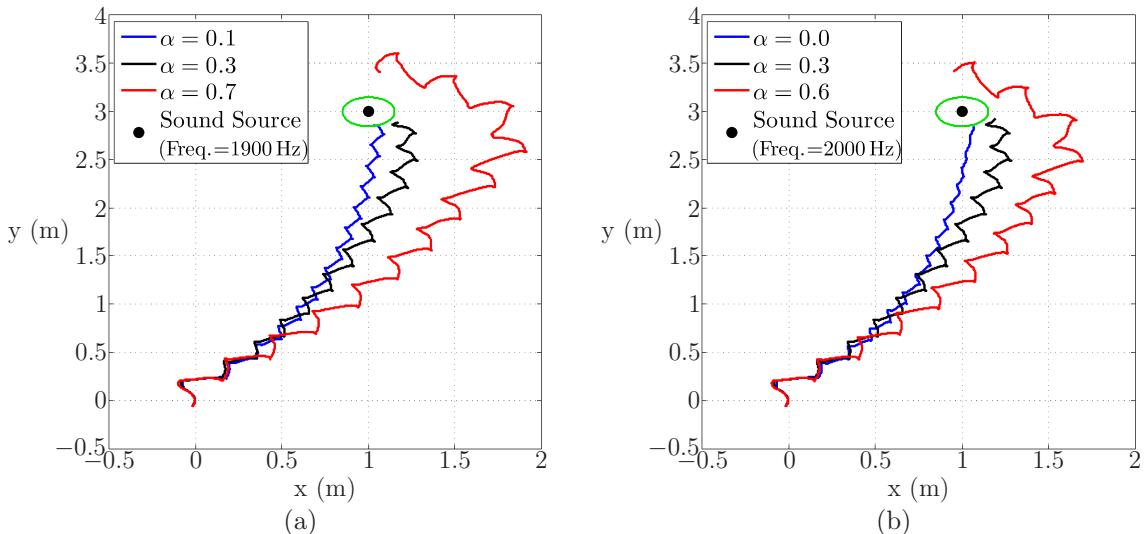
One possible solution could have been to modify the range of values for  $\mu$  in the simulation to match that on the robot. However, as mentioned in Sect. 6.3  $\mu$  is a dimensionless quantity that is proportional to the amplitude of the body CPG oscillations, and is not the actual value of amplitude itself. Therefore, in order to match the range of values for  $\mu$  in simulation to that on the robot, the relationship between  $\mu$  and the actual CPG oscillation amplitude must be known in both cases, so that proper adjustments can be made. In the simulation model the instantaneous amplitude of the CPG oscillations can easily be measured, but this is not possible in the robot model because this feature is not provided by the firmware and would require careful modification of the firmware code, which is not a trivial task by any means and would require significant amount of

time which unfortunately was not available.

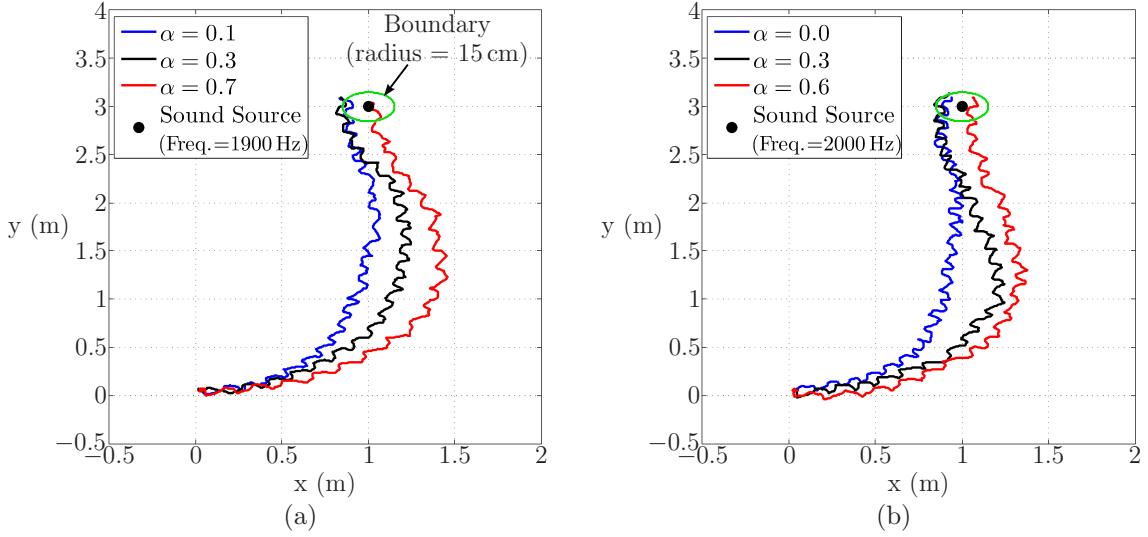
During each trial the head movements, the normalized auditory model outputs  $\|i_l\|$  and  $\|i_r\|$  and the amplitudes  $\mu_l$  and  $\mu_r$  of the left and right body CPG oscillations were logged. A circular boundary centered on the sound source defined the success criteria. Only if the robot entered this area, was it defined as a success. Trials were terminated when the robot either (a) hit the sound source, (b) moved inside the boundary or (c) moved past the sound source without crossing the boundary, whichever occurred first.

#### 6.4.2.2 Results and Discussion

Figures 6.16 and 6.17 depict the robot's head trajectories in simulation and robotic trials respectively. It is evident in both the cases that increasing  $\alpha$  resulted in the robot making wider turns. Therefore, the results of the simulation qualitatively match those obtained in the robotic trials. Also, the trajectories for both the sound frequencies match qualitatively with each other. However, there are some differences between the results from the simulation and those from the robotic trials. In simulation, the effect of increasing  $\alpha$  on the head oscillations is evident — as expected, they become more pronounced as  $\alpha$  increases. This however is not the case in the robotic trials. The head oscillations tend to remain relatively unaffected by changes in  $\alpha$ . This can be attributed to the conservative nature of the CPG model implementation on the robot, which does not permit strong head oscillations.



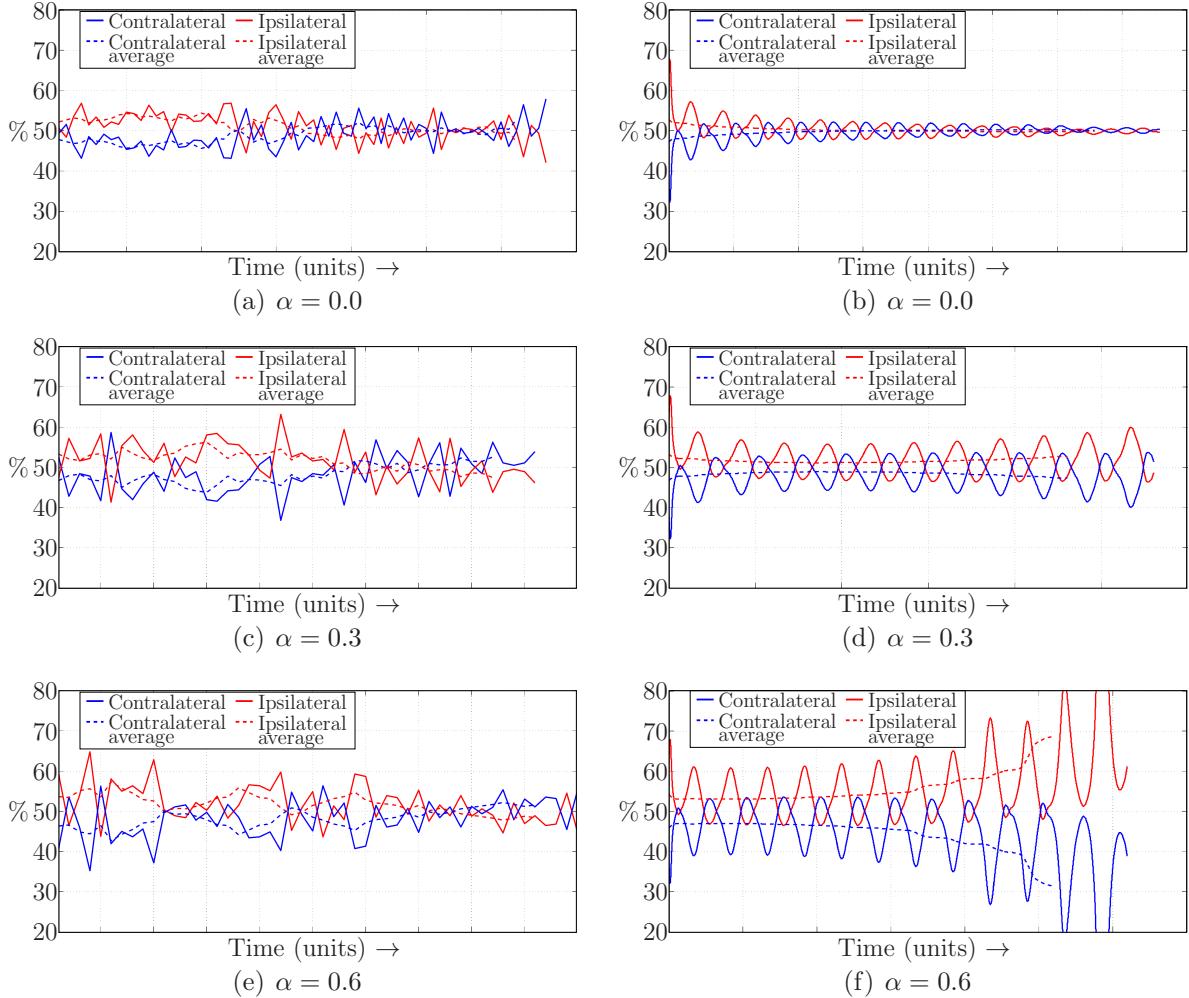
**Figure 6.16:** The head trajectories recorded in simulated trials.



**Figure 6.17:** The head trajectories recorded in robotic trials.

In terms of whether the robot performed successful localization or not according to the definition, it is evident that in simulation the robot was successfully able to localize only for low values of  $\alpha$  ( $\alpha = 0.0$  and  $0.1$ ) for both sound frequencies. For larger values, the robot was unsuccessful in localizing the sound source, although it came quite close for  $\alpha = 0.3$ . For  $\alpha = 0.6$  and  $0.7$ , the turns were too wide to localize the sound source successfully. Conversely, in the robotic trials, the robot successfully localized the sound source in every trial. Again, this reflects the effect of having a more conservative range of values for the amplitudes of body CPG oscillations in the robotic trials. There is less overall bending of the body in the robotic trials than in simulation for increasing  $\alpha$ , allowing for tighter turns during locomotion.

Figure 6.18 depicts as an example the dynamics of the auditory model outputs, i.e. the normalized relative tympanal vibration strengths  $\|i_l\|$  and  $\|i_r\|$  for the ipsilateral and contralateral sides during locomotion in robotic trials, for the sound frequency of 2000 Hz. As expected, in both simulation and the robotic trials the difference between the average ipsilateral and contralateral tympanal vibration strengths is greater in the beginning of the trials when the robot is turning. When the robot has oriented itself in the direction of the sound source, this difference is negligible. After this point, the results in simulation and in the robotic trials differ. Here again, the effect of the conservative range of values for the amplitudes of body CPG oscillations is visible. The difference increases in simulation as the robot fails to orient itself towards the sound source in time, while in the robotic trials the difference reduces as the robot is able to orient itself towards the sound source in time and successfully localizes it.



**Figure 6.18:** The dynamics of the relative tympal vibration strengths on the contralateral and ipsilateral sides during locomotion in simulation and robotic trials. The sound frequency is 2000 Hz. Figures (a), (c) and (e) show results from robotic trials, while Fig. (b), (d) and (f) show results from simulations.

## 6.5 Summary and Conclusions

In this chapter the peripheral auditory system model of the lizard was coupled to a neuronal spinal cord model composed of central pattern generators responsible for generating locomotion patterns. Braitenberg cross-couplings were defined to this end, configured to produce contralateral excitation and ipsilateral inhibition of the amplitudes of the body CPG oscillations in the presence of a sound source. A gain parameter  $\alpha$  was defined which governed the extent of modulate the level of excitation and inhibition. The hybrid model was characterized in simulation in terms of angular displacement of the sound source and  $\alpha$  via phonotaxis experiments with a robotic model of a salamander.

Further comparative phonotaxis experiments in simulation and with the *Salamandra Robotica II* salamander-robot were performed to evaluate the phonotactic performance of the hybrid robot and to observe the effect of varying  $\alpha$  on the locomotion trajectories.

It can be concluded that the robot was able to successfully localize a sound source in both simulation and robotic trials in spite of the head oscillations that produce oscillating auditory cues, given that the gain parameter  $\alpha$  is sufficiently low. However, the results in simulation and robotic trials diverge in terms of the phonotactic performance and trajectory shapes. Successful localization was observed in all robotic trials for all the selected test values of  $\alpha$ , while in simulation only low values of  $\alpha$  resulted in successful localization. This outcome was attributed to the fact that the implementation of the CPG model in simulation was different from the implementation on the salamander-robot. The amplitudes of the CPG oscillations  $\mu$  were more sensitive in simulation to changes in  $\alpha$ , as compared to the CPG oscillation amplitudes in practice. The head oscillations should get stronger for increasing  $\alpha$  as seen in the simulation, rather than stay significantly unchanged as seen in practice. This created an important mismatch between simulation and practice, resulting in weaker performance in the former and a stronger performance in the latter. From the point of view of the head oscillations, the simulation results are more accurate than those in practice. However, the effect of varying  $\alpha$  on the locomotion trajectories was successfully verified in both simulation and practice.

In response to the first part of the research question posed in the beginning of this chapter, the results of the simulation and robotic trials are in favour of the conclusion that it is possible to localize a static sound source during locomotion in the presence of oscillating auditory cues and in the absence of any form of head stabilization. With regards to the question of how well can the sound source be localized, the simulation results appear to suggest that localization performance is quite good depending on the choice of the gain parameter. The robotic trials however, are not a good indicator of the localization performance, since there is a distinct mismatch between the parameters employed in simulation and in the robotic trials. More thorough practical experiments that match the conditions in the simulation must be performed to answer this question concretely.



# Chapter 7

## Three-dimensional Sound Source Localization

### 7.1 Introduction

As is evident from the previous chapters, all the tasks to which the peripheral auditory system was applied were two-dimensional in nature i.e., the goal of the system was to localize a sound source either to its left or right but always in the same two-dimensional plane as the robot. Therefore, two microphones placed horizontally were sufficient to extract the direction-dependent binaural cues containing information about the azimuthal displacement of the sound source. However, the peripheral auditory system coupled with a horizontal two-microphone configuration cannot localize sound in three-dimensional space because such a configuration is not sufficient to extract auditory cues containing information regarding the elevation of the sound source. This raises the question, how can the peripheral auditory system localize a sound source in three dimensions? After all, lizards reside in a three-dimensional world and thus are subjected to sound sources located in three-dimensional space that are relevant to their survival, for example during mating and/or other social interactions. However, to the best of the author's knowledge, there have not been any behavioural studies conducted that can confirm that lizards are capable of performing three-dimensional sound localization or that they utilize elevation cues. Furthermore, no bioacoustical study was uncovered by the author in the literature which reports on the three-dimensional response of the lizard's peripheral auditory system. In this chapter the investigation is restricted to the problem of three-dimensional sound source localization with the peripheral auditory system from an engineering perspective. Preliminary experiments with a simple

bio-inspired approach that utilizes three microphones in conjunction with the auditory model in duplicate to localize sound in three dimensions are presented.

The work presented in this chapter covers and elaborates further upon work accepted for publication at the IEEE International Conference on Robotics and Biomimetics<sup>1</sup>, held through December 7–11, 2011. Due to lack of travel funds, the submission was later retracted.

## 7.2 Three-dimensional Sound Localization

The problem of localizing sound in three dimensions has been extensively studied and a number of approaches have been proposed. The majority of these approaches employ a static two-dimensional multi-microphone arrangement [136, 75, 120] of a varying number of microphones, typically ranging from 4 to 128 or more, usually mounted in an array or a rectangular grid. These techniques rely on measuring the difference between the time of arrival of sound, i.e. the Interaural Time Difference (ITD) cues, at the individual microphones to estimate the location of the sound source. The mechanical assembly as a whole restricts the location of the source to the frontal hemisphere. The main drawback of these approaches is the fact that engineering such complex systems is non-trivial. Utilizing a large number of microphones requires multi-channel audio capturing systems which generate a considerably large quantity of multi-dimensional data. Processing such data is a computationally heavy task, necessitating systems with significant computational resources.

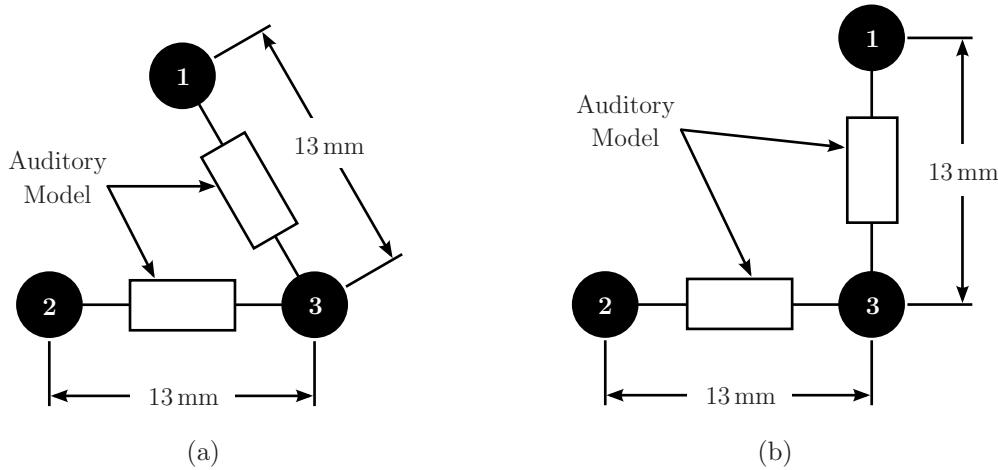
Approaches based on determining the well-known head-related transfer function (HRTF) [17], typically that of a human, have existed for quite some time, and rely on modelling the pre-filtering effects of the pinnae, head and torso [64, 55, 53, 73]. These usually employ just two microphones (but are not limited to that number) placed inside the ears of an artificial human head. These techniques are designed to localize a sound source located in a three-dimensional sphere centred on the microphone arrangement. The latest of such approaches focusses on the miniaturization of the microphones or of arrays of microphones at the silicon level [72, 76].

Since the lizard auditory system model can localize sound in the two-dimensional azimuth plane only, information regarding the location of the sound source in the third dimension, or its elevation, must be added to be able to localize in three-dimensional space. The simplest way of incorporating such information is to have two identical and

---

<sup>1</sup>[Link to conference website.](#)

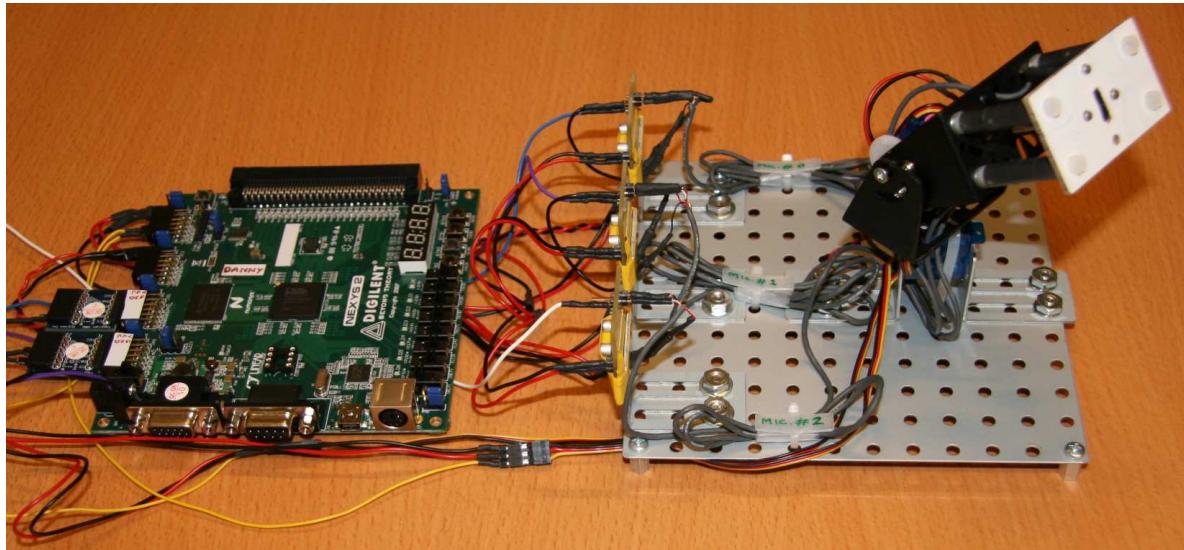
independent auditory system models operating simultaneously, one for localization in the horizontal plane and the other in the vertical plane. Each model requires two microphones to provide audio input to it, but instead of using two microphones per model for a total of four, one can just as well employ three microphones in all, with one of the microphones shared between the two models. This leads to two possible microphone configurations as depicted in Fig. 7.1 — one with the microphones arranged in a square grid, and another where they are arranged in a triangular grid. In this paper we use the triangular grid configuration purely to maintain the configurational symmetry.



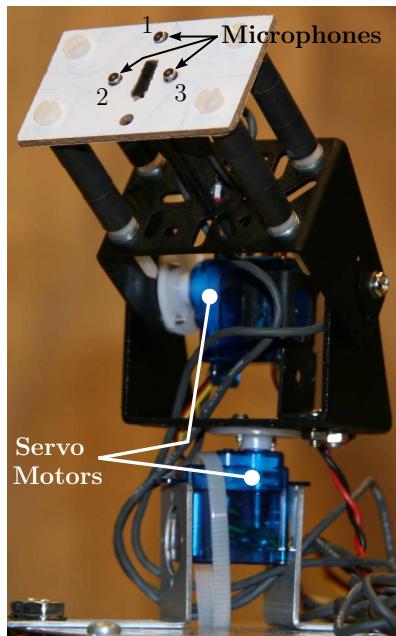
**Figure 7.1:** The possible configurations utilizing three microphones. (a) A triangular grid arrangement. (b) A square grid arrangement.

### 7.3 Design and Implementation

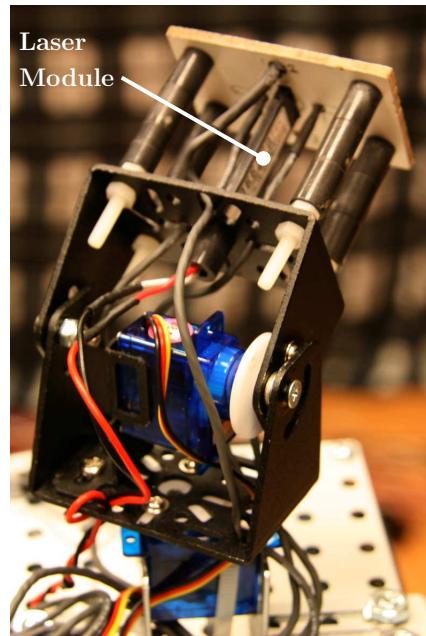
A sound localization system depicted in Fig. 7.2(a) was assembled, consisting of a pan-tilt unit (refer to Fig. 7.2(b)) using off-the-shelf hardware components, interfaced to a Field Programmable Gate Array (FPGA) board (the NEXYS<sup>TM</sup>2 from Digilent Inc.). The pan-tilt unit measured  $132 \times 50 \times 30$  mm in its fully straightened state. The three omnidirectional microphones (model FG-23329-P07 from Knowles Electronics) were fixed in the form of an equilateral triangular grid at the front of the tilt frame as depicted in Fig. 7.2(b). The distance between any two microphones was 13 mm, which is the same as that between the two eardrums of the real animal in question. This is important because the measured parameters of the auditory system model are tuned to the distance between the eardrums of a typical lizard [21], and this relationship must be maintained in the implementation as well. The microphones numbered 2 and 3 acted



(a) The FPGA development board interfaced to the pan-tilt unit.



(b) The front-side of the pan-tilt unit with the three microphones positioned in a triangular grid.



(c) The back-side of the pan-tilt unit depicting the laser module.

**Figure 7.2:** The sound localization platform.

as the horizontal pair, providing audio input to one auditory system model, and those numbered 1 and 3 acted as the vertical pair, providing audio input to the other auditory system model. A laser module was also mounted between microphones numbered 2 and 3 in order to provide a visual aid for determining successful localization by illuminat-

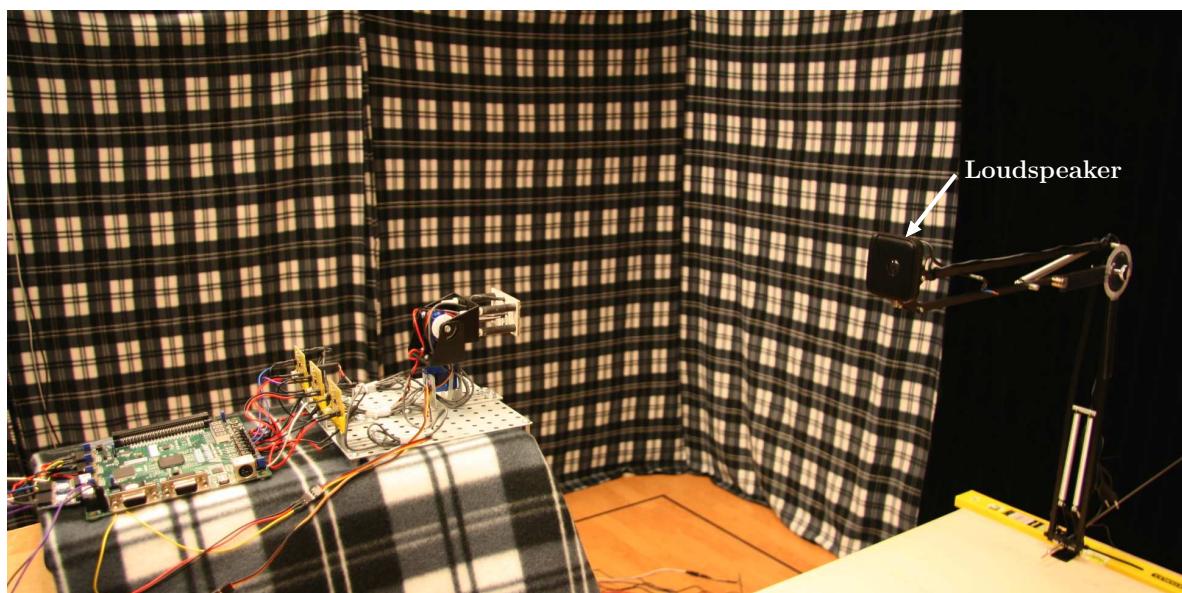
ing the target sound source. The pan-tilt unit was driven by two HITEC HS-55 servo motors. The auditory system models and the servo motor control logic were designed and implemented in VHDL and programmed into the Xilinx Spartan 3E XC3S1200E (with 1.2 million logic gates) FPGA chip available on the board. The pan-tilt unit was interfaced to the development board by coupling the servo motors, the laser module and the microphones whose outputs were digitized via two dual-channel analogue-to-digital converters.

The scalar outputs of each auditory model were the average powers of the filtered ipsilateral and contralateral auditory signals  $i_l$  and  $i_r$  respectively, which are respectively analogous to the average tympanal vibration strengths of the lizard's ipsilateral and contralateral eardrums. The pan servo motor was controlled by the outputs of the auditory system model associated with localization in the horizontal plane and the tilt servo motor was controlled by the outputs of the auditory model associated with localization in the vertical plane. For example, when the sound source was to the left of the two microphones located in the horizontal plane, then  $i_l > i_r$  and the pan servo motor rotated the pan-tilt unit to the left, and when the sound source was to the right of the two microphones located in the horizontal plane, then  $i_r > i_l$  and the pan servo motor rotated the pan-tilt unit to the right. The tilt servo was controlled in a similar fashion.

Initial tests revealed that this strategy introduced mechanical jittering in the movement of the pan-tilt unit when the laser beam was quite close to the loudspeaker, because at this stage both the auditory systems were competing to maintain the balance between  $i_l$  and  $i_r$ , and any overshoot in one plane produced an overshoot in the other plane as well due to the  $60^\circ$  angle between the systems' reference directions. This made it impossible for the laser module to keep the laser beam steady while focussing it on the loudspeaker, and thus impossible to visually confirm whether localization was successful or not. Therefore, in order to minimize these mechanical jitters inherent in such a control strategy, two constraints were introduced — (a) the speeds of both the servo motors were reduced from the default speed of  $357^\circ/\text{sec}$  to approximately  $25^\circ/\text{sec}$  (this was determined via trial-and-error to be a sufficient compromise between reducing experimental time and minimizing the jitters) and (b) a stop criterion was introduced where a servo motor was stopped and locked when the auditory model controlling it had successfully localized the sound source, i.e.  $i_l$  and  $i_r$  were equal in magnitude. For instance, if localization in the horizontal plane was successful before that in the vertical plane, the pan servo motor was locked in its current position, and the tilt servo motor was allowed to run until the localization in the vertical plane was successful as well. This was an unfortunate but necessary hack, that compounded the negative aspects of the chosen strategy.

## 7.4 Experimental Setup

The experimental setup (refer to Fig. 7.3) consisted of a miniature loudspeaker (from Digilent Inc.) mounted on a mechanical assembly that allowed the loudspeaker to be moved in three dimensions. The dimensions of the loudspeaker were  $70 \times 75 \times 27$  mm. The loudspeaker and the pan-tilt unit were placed such that when both were located at  $0^\circ$  azimuth and  $0^\circ$  elevation, the laser beam from the pan-tilt unit was focussed on the centre of the loudspeaker. The positions of the pan-tilt and the loudspeaker assembly were then kept fixed throughout the experiments, and only the loudspeaker was allowed to move relative to its horizontal and vertical swivel points. A monotone sound was continuously played through the loudspeaker and the sound localization system had to then localize the sound. The maximum possible ranges of azimuth and elevation angles within the mechanical constraints of the loudspeaker assembly were  $[-80^\circ, +80^\circ]$  and  $[0^\circ, +60^\circ]$  respectively. Viewing the loudspeaker from the perspective of the pan-tilt unit, negative angles in the horizontal plane signify that the loudspeaker was to the left side of the unit, while positive angles signify that the loudspeaker was to the right side. Similarly, in the vertical plane, positive angles signify that the loudspeaker was above the pan-tilt unit. At the locations defined by an azimuth of  $\pm 80^\circ$  and an elevation of  $+60^\circ$  the distance between the loudspeaker and the microphones was approximately 90 cm, while at the location defined by an azimuth of  $0^\circ$  and an elevation of  $0^\circ$  the distance between the loudspeaker and the microphones was approximately 38 cm.

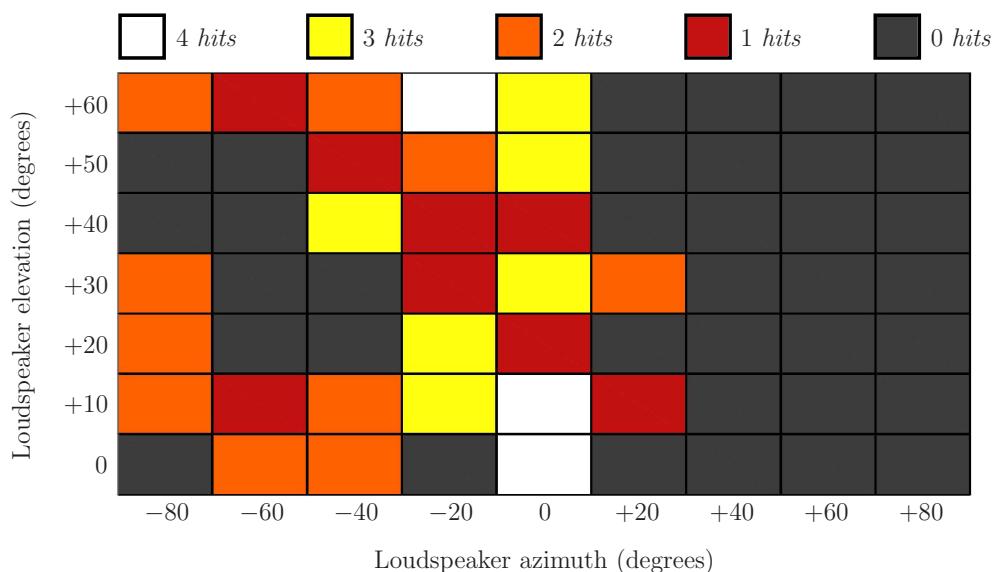


**Figure 7.3:** The experimental setup.

In order to evaluate the localization performance in three-dimensional space, experiments were conducted for various values of the azimuth and elevation angles. The azimuth angle was varied from  $-80^\circ$  to  $+80^\circ$  in steps of  $20^\circ$ , and the elevation angle was varied from  $0^\circ$  to  $+60^\circ$  in steps of  $10^\circ$ . For a given azimuth and elevation angle of the loudspeaker, five consecutive trials were performed. Before each trial, the servo motors were reset such that both motors were rotated to their centre positions in which case the laser beam was directed straight upwards. Localization was considered successful and labelled as a *hit* if the system managed to direct the laser beam at the body of the loudspeaker or within a 5 cm radius around the loudspeaker. The sound frequency was chosen as 2000 Hz from the 1000 Hz–2200 Hz range, because at this frequency the peripheral auditory system shows a strong response (refer to Fig. 3.4). All experiments were conducted in a sound dampening chamber to minimize sound reflections.

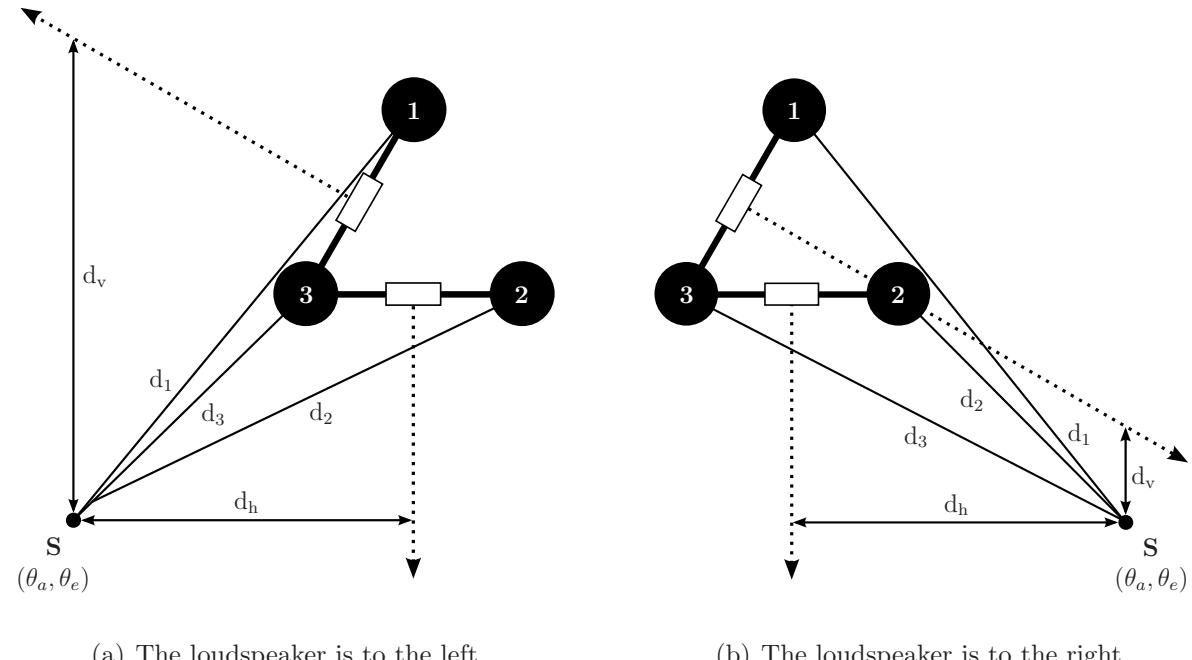
## 7.5 Results and Discussion

Fig. 7.4 depicts the localization performance of the system in terms of the number of *hits*. It is evident that the system performs reasonably well within the  $-20^\circ$  to  $+20^\circ$  range in the azimuth plane, considering the simplicity of the implemented control strategy. Outside this range, the performance suffers significantly and is quite inconsistent. For instance, when the loudspeaker is on the left side of the microphone grid localization is successful for a limited number of values of elevation, but when the loudspeaker is on the right side of the microphone grid, localization fails completely.



**Figure 7.4:** Localization performance in three dimensions in terms of the number of *hits*.

This result can be attributed to the fact that the plane formed by the microphones that are being utilized for localization in the vertical plane is not in fact vertical, but rotated by  $30^\circ$  relative to the vertical towards the right side of the pan-tilt unit as a consequence of the equilateral triangle configuration. In the initial condition, the microphone grid is pointing vertically upwards, and thus the relative position of the loudspeaker in this case is always below the plane formed by the microphone grid irrespective of the relative elevation angle of the loudspeaker, and is either to the left, in the middle or to the right of the grid. Therefore, when the trial starts the microphone grid always moves downwards, either vertically or diagonally. Since the pan and tilt servo motors are controlled independently, either of them can be locked before the other, depending on which auditory model's outputs are balanced first. Viewing the loudspeaker from the perspective of the microphone grid, Fig. 7.5(a) and Fig. 7.5(b) depict the initial conditions, with the loudspeaker respectively to the left side and the right side of the microphone grid. In both figures, the elevations of the loudspeaker are identical.



**Figure 7.5:** The bias introduced into the system due to the independent servo motor locking in the control strategy.

The terms  $d_1$ ,  $d_2$  and  $d_3$  represent the distances between the sound source  $\mathbf{S}$  located at a fixed azimuth  $\theta_a$  and elevation  $\theta_e$  and the microphones 1, 2 and 3 respectively, while  $d_h$  and  $d_v$  represent the horizontal and vertical distances between  $\mathbf{S}$  and the central planes (the plane formed by the dotted line and an axis projecting normally outwards towards the reader) of the microphone pairs 2-3 and 1-3. The central plane of a microphone pair

represents an hypothetical plane on which any given point is equidistant from either microphone of that pair. This means that if at any time during the trial the microphone grid is oriented such that the loudspeaker lies on the central plane of either microphone pair, the outputs of the auditory model connected to that microphone pair are balanced and that particular model has successfully localized the sound source. The central planes for both the pairs intersect at an axis which passes through the centre of the equilateral triangle formed by the three microphones, and when the loudspeaker lies at any point on this axis,  $d_1$ ,  $d_2$  and  $d_3$  are equal to each other and both the auditory models' outputs are balanced. When the microphone grid starts to orient itself towards the loudspeaker, the distances  $d_h$  and  $d_v$  start becoming smaller simultaneously. As soon as the loudspeaker "lands" on either of the central planes, the control strategy locks the servo motor corresponding to that particular central plane.

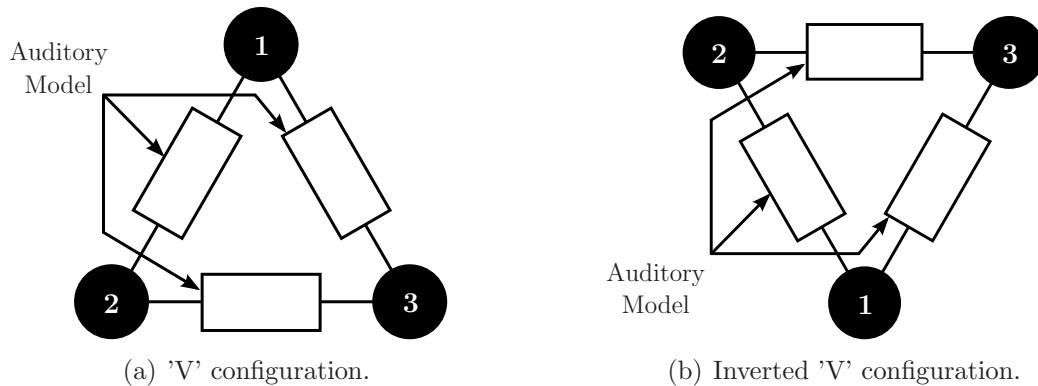
For the same elevation angle,  $d_v$  is relatively larger when the loudspeaker is to the left side than when the loudspeaker is to the right side, since the central plane corresponding to the microphone pair 1-3 is tilted downwards towards **S** when the loudspeaker is to the right side. Therefore, the tilt servo motor runs for a longer period of time, moving the microphone grid in a diagonally straight line towards the loudspeaker, before the loudspeaker lands on the central plane of the microphone pair 1-3. At this point the corresponding auditory model's outputs are balanced, and the servo motor is then locked by the control strategy. On the other hand, when the loudspeaker is to the right side, the tilt servo motor only runs for a relatively short period of time before the loudspeaker lands on the central plane of the microphone pair 1-3. The servo motor is then prematurely locked by the control strategy, with the microphone grid oriented at a location relatively higher above the loudspeaker. Due to the independent locking of the servo motors, the control strategy does not allow for any kind of compensation between the two auditory model's outputs. However, in the absence of this locking mechanism, the two auditory model's would continuously compete with each other, since any imbalance in the outputs of one model would cause the corresponding servo motor to run, possibly disrupting the balance in the outputs of the other model, causing its corresponding servo motor to run in order to compensate for the imbalance created. In this manner, the two auditory models would continuously be taking corrective action against each other, leading to mechanical jitters in the system.

In short, the observed behaviour is the result of the interaction of the non-ideal geometry of the microphone grid with the unfortunate hack of locking the servo motors prematurely.

## 7.6 Summary and Conclusions

In this chapter a very preliminary bio-inspired approach to sound source localization in three dimensions using three microphones was presented. The model of the peripheral auditory system was implemented in duplicate, and employed in conjunction with a pan-tilt unit with the microphones mounted on it in the form of a triangular grid configuration. The three-dimensional sound localization ability of the system was experimentally evaluated. Given the simplicity of the localization control strategy the system performed reasonably well over a limited subset of the experimental range of azimuth and elevation of the sound source. However, it was discovered that the control strategy actually hinders the localization ability of the system.

It can be concluded from the results that the system shows promising results. However, a vertical bias was introduced due to the independent servo motor locking employed in the control strategy. This bias led to unsuccessful source localization performance on the right side. This indicates that the control strategy with two independent auditory models employed here is not suitable for a triangular grid microphone configuration. As an alternative, it may be prudent to employ three microphone pairs instead of two such that two of the microphone pairs are arranged in a 'V' or an inverted 'V' configuration as depicted in Fig. 7.6(a) and Fig. 7.6(b) respectively. These two pairs would together handle localization in the vertical plane, thereby eliminating the vertical bias, while the third horizontally arranged microphone pair would handle localization in the horizontal plane as in the current implementation. This kind of compensatory mechanism may exhibit better localization performance.



**Figure 7.6:** A three microphone pair configuration.

It may also be interesting to employ the square grid microphone configuration, which does not suffer from such a vertical bias, along with the existing control strategy and repeat the experiments in order to investigate the localization performance.



# Chapter 8

## Conclusions

The work presented in this dissertation focusses on the investigation of an existing model of the peripheral auditory system of lizards. In earlier work, a robotic model comprising the auditory model coupled with a simple decision model has been built by Zhang et al. and the phonotactic performance of the robotic model has been evaluated in both simulation and practice, in order to evaluate the directionality of the auditory model. The work presented here builds upon this earlier research, and extends the application of the auditory model to previously unexplored areas through a series of independent experiments conducted with the auditory model. Two different implementations of the model were realized using two different technologies, one employing a digital signal processor which featured serial task execution, and the other employing an FPGA chip, which featured parallel task execution. However, both the implementations exhibited strong directional acuity, leading to the conclusion that the auditory model is quite robust. In the following sections, a summary of the experiments performed is presented along with the contributions made, and possible future directions are briefly discussed.

### 8.1 Summary and Contributions

All previous robotic implementations of the auditory model utilized a step-control strategy, realized by coupling the auditory model's outputs to a decision model that steered the robot towards the sound source. As an alternative, a mobile robot inspired by Braitenberg vehicles was designed and built, with two microphones as sound sensors and two motor driven wheels as actuators. The sensors and actuators were cross-coupled to each other through the auditory model, such that the contralateral output of the auditory

model was scaled and coupled to the ipsilateral motor as its drive speed parameter, while the ipsilateral output of the auditory model was scaled and coupled to the contralateral motor as its drive speed parameter. This rendered the control strategy continuous in nature. The phonotactic performance of the robot was evaluated on either of its sides over the auditory model's relevant frequency range, and the robot's left and right side trajectories were statistically examined to determine any significant differences between them. The performance of the continuous control approach was compared to that of the step-control approach used by Zhang et al., and it was discovered that two approaches were comparable. Statistical analysis revealed that in most cases there were not any significant differences between the left and right side trajectories.

Next, simple decision models were independently embedded into the sensorimotor couplings of the robot in order to investigate their effect on the directional acuity. Two types of decision models were employed — the **MAX** model whose output was the maximum of its inputs taken over a fixed period of time, and the **AVG** model whose output was the average of its inputs taken over a fixed period of time. The time period could be varied by introducing varying delays in the control loop, essentially varying the control loop frequency. The phonotactic performance for the two decision models over the relevant frequency range was evaluated for various control loop frequencies. The dependence of the chosen decision models on the robot's performance was examined statistically. It was discovered that the **MAX** model was quite sensitive to control loop frequencies, and higher frequencies improved its phonotactic performance. On the other hand, the **AVG** model was relatively less sensitive to control loop frequencies, and it exhibited good and steady performance over all the frequencies. Statistical analysis revealed that there was a probability that the decision models were dependent on the control loop frequencies.

After validating the use of Braitenberg sensorimotor cross-couplings as an valid alternative control strategy to using decision models for phonotactic steering, the scaling properties of the auditory model were investigated. A new mobile robot was designed and built which employed the auditory model coupled with a decision model for phonotactic steering. This platform was employed to test the hypothesis that any joint but consistent scaling of the auditory model and the ear separation should result in an equivalent scaling of the frequency response, and to investigate the effect of violating the scaling relationship between the ear separation and the auditory model on the directional acuity of the system. Extensive phonotaxis experiments were performed to test the hypothesis, and the results proved the hypothesis to be true. Consistent scaling of the auditory model and the ear separation resulted in an proportional scaling of the

frequency response of the model. Furthermore, the directional acuity degraded when the scaling of the ear separation was inconsistent with the scaling of the auditory model. From the auditory model's perspective, this indicated that the auditory model was quite robust to scaling, under the condition that the scaling remained consistent.

All the previous robotic models were mobile, and their goal was to localize a sound source by moving towards it. The ability to move freely in space towards a sound source facilitated sound localization by enforcing a direct mapping between the auditory model's outputs and the sound direction. Therefore, the more difficult task of sound localization while remaining in a fixed location was subsequently addressed, in order to investigate the feasibility of localizing sound in this manner. This had been previously attempted by Zhang et al., but only in simulation and only with constant frequency sounds. In this work, the goal was to validate the learning approach in simulation, and then implement it in practice. The auditory model was coupled to a Cerebellar Model Articulation Controller based neural network and the ability of the system to learn a representation of sound direction over the relevant frequency range was investigated in simulation and in practice. It was discovered in simulation that the system was able successfully build a representational map of the sound direction. However, due to an error in the proprietary software tools used to implement the robotic model, all attempts at realizing the learning process online in practice failed.

The auditory model mimics a biological system and one can find innumerable examples of co-ordinating sensory and motor systems in nature. However, all the robotic models employed previously were wheeled robots. Therefore, it was considered prudent to investigate how such a biomimetic sensory system could be combined with a biomimetic motor system. To this end, the auditory model was coupled to a Central Pattern Generator based locomotor model of *Salamandra robotica II* a salamander-like quadruped robot developed in the Biorobotics Laboratory at the École Polytechnique Fédérale de Lausanne, employing Braitenberg cross-couplings between the auditory model's outputs and the ipsilateral and contralateral drive signals for the locomotor model. The strength of the drive signals determined the extent to which the body could bend in either direction. Since the salamander-robot exhibits undulatory locomotion similar to a lizard, characterised by oscillating head movements which result in oscillating auditory cues, the robotic model thus obtained was employed in order to evaluate the sound localization ability of the system in the presence of such oscillatory auditory cues. Extensive phonotaxis experiments were performed in simulation and a small subset of those were performed in practice due to time constraints. The robotic model performed quite well in simulation as well as in practice.

The sound localization capability of the auditory model is restricted to a two dimensional space. In order to investigate the feasibility of sound localization in three dimensional space with the auditory model, a pan-tilt sound localization system was built, with three microphones mounted on it in an equilateral triangular grid. The three microphones formed two pairs, which acted as input to two independent auditory models. One model was employed for horizontal sound localization, and the other for vertical sound localization. The two models independently controlled two servo motors to rotate the pan-tilt unit in three dimensions. Preliminary experiments were performed to evaluate the three dimensional sound localization performance of the system. It was discovered that the system performed well only at a few points in the three dimensional test space. The cause of this behaviour was determined to be the control strategy used for localization, which caused the servo motors to lock prematurely during the trials.

## 8.2 Future Directions

There are several directions that could be pursued in terms of future work. Since the realization of the learning process in practice was unsuccessful due to circumstances that could not be controlled, the question whether online learning of sound representation can be successful in practice is still unanswered. Therefore, as an immediate step, the experiments can be repeated provided that the error in the software tools has been duly rectified by its developers. Alternatively, a different robotic platform based on a digital signal processor can be employed instead of the current one which is based on an FPGA chip, which would require re-implementing the auditory model and the neural network.

In the experiments performed with the salamander robot, the robotic model and the simulation model had marked differences between them that rendered direct comparison between the phonotactic performances of the two impossible. For instance, the CPG model in simulation differs from that on the salamander robot in that the CPG oscillations are more sensitive to  $\alpha$  in simulation than in practice, leading to inconsistent trajectories for the same  $\alpha$ . Also, the mechanical properties of the arena surface and the robot's limbs and modules are not sufficiently accurate in the simulation model, again leading to inconsistent bending patterns and thus inconsistent trajectories. Therefore, it is prudent to modify either the simulation model or the robotic model, to reflect the properties of the other more accurately and perform more comparable and thorough experiments.

In the three dimensional sound localization experiments, the phonotactic performance

was only evaluated for a small set of points in three dimensional space, and thus more extensive experiments are required. Furthermore, the control strategy used was not optimal for the triangular grid configuration and thus a more appropriate strategy is required, such as using three microphone-pairs instead of two with three auditory models in the control. Two auditory models could work in tandem in a compensatory manner to eliminate the vertical bias evident in the current strategy. Eliminating the unfortunate hack employed and utilizing a more principled coupling of the control axes would also alleviate the problem.







# Bibliography

- [1] Akcakaya M. and Nehorai A., “Performance Analysis of the *Ormia ochracea*’s Coupled Ears,” *Journal of the Acoustical Society of America*, vol. 124, pp. 2100–2105, 2008.
- [2] Albus J.S., “A Theory of Cerebellar Function,” *Mathematical Biosciences*, vol. 10, no. 1-2, pp. 25–61, 1971.
- [3] Albus J.S., “A New Approach to Manipulator Control: The Cerebellar Model Articulation Controller (CMAC),” *Journal of Dynamic Systems, Measurement, and Control*, vol. 97, no. 3, pp. 220–227, 1975.
- [4] Albus J.S., “Mechanisms of Planning and Problem Solving in the Brain,” *Mathematical Biosciences*, vol. 45, no. 3-4, pp. 247–293, 1979.
- [5] Asano F., Suzuki Y., and Sone T., “Role of Spectral Cues in Median Plane Localization,” *Journal of the Acoustical Society of America*, vol. 88, pp. 159–168, 1990.
- [6] Ashley-Ross M.A. and Bechtel B.F., “Kinematics of the Transition between Aquatic and Terrestrial Locomotion in the Newt *Taricha torosa*,” *Journal of Experimental Biology*, vol. 207, no. 3, pp. 461–474, 2004.
- [7] Azaria M. and Hertz D., “Time Delay Estimation by Generalized Cross Correlation Methods,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 2, pp. 280–285, 1984.
- [8] Azizi E., “Biomechanics of Salamander Locomotion,” Ph.D. dissertation, University of Massachusetts Amherst, 2005.
- [9] Bennett W.O., Simons R.S., and Brainerd E.L., “Twisting and Bending: The Functional Role of Salamander Lateral Hypaxial Musculature during Locomotion,” *Journal of Experimental Biology*, vol. 204, no. 11, pp. 1979–1989, 2001.

- [10] Braitenberg V., *Vehicles: Experiments in Synthetic Psychology.* MIT Press, Bradford Books, Cambridge, 1984.
- [11] Brandstein M. and Ward D., *Microphone Arrays: Signal Processing Techniques and Applications.* Springer Verlag, 2001.
- [12] Buchanan J.T., “Neural Network Simulations of Coupled Locomotor Oscillators in the Lamprey Spinal Cord,” *Biological Cybernetics*, vol. 66, no. 4, pp. 367–374, 1992.
- [13] Buchanan J.T. and Grillner S., “Newly Identified ‘Glutamate Interneurons’ and their Role in Locomotion in the Lamprey Spinal Cord,” *Science*, vol. 236, no. 4799, pp. 312–314, 1987.
- [14] Cade W., “Acoustically Orienting Parasitoids: Fly Phonotaxis to Cricket Song,” *Science*, vol. 190, pp. 1312–1313, 1975.
- [15] Carter G.C., “Coherence and Time Delay Estimation,” *Proceedings of the IEEE*, vol. 75, no. 2, pp. 236–255, 1987.
- [16] Chen J.C., Yao K., and Hudson R.E., “Acoustic Source Localization and Beam-forming: Theory and Practice,” *EURASIP Journal on Applied Signal Processing*, vol. 4, pp. 359–370, 2003.
- [17] Cheng C.I. and Wakefield G.H., “Introduction to Head-related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space,” *Journal of the Audio Engineering Society*, vol. 49, no. 4, pp. 231–249, 2001.
- [18] Chevallier S., Ijspeert A.J., Ryczko D., Nagy F., and Cabelguen J.-M., “Organisation of the Spinal Central Pattern Generators for Locomotion in the Salamander: Biology and Modelling,” *Brain Research Reviews*, vol. 57, no. 1, pp. 147–161, 2008.
- [19] Christensen-Dalsgaard J., “Directional Hearing in Nonmammalian Tetrapods,” in *Sound Source Localization*, ser. Springer Handbook of Auditory Research, Popper A.N. and Fay R.R., Eds. Springer New York, 2005, vol. 25, pp. 67–123.
- [20] Christensen-Dalsgaard J., “Vertebrate Pressure-gradient Receivers,” *Hearing research*, vol. 273, no. 1-2, pp. 37–45, 2011.
- [21] Christensen-Dalsgaard J. and Manley G.A., “Directionality of the Lizard Ear,” *Journal of Experimental Biology*, vol. 208, no. 6, pp. 1209–1217, 2005.

- [22] Christensen-Dalsgaard J. and Manley G.A., “Acoustical Coupling of Lizard Eardrums,” *Journal of the Association for Research in Otolaryngology*, vol. 9, no. 4, pp. 407–416, 2008.
- [23] Christensen-Dalsgaard J., Tang Y., and Carr C.E., “Binaural Processing by the Gecko Auditory Periphery,” *Journal of Neurophysiology*, vol. 105, no. 5, p. 1992, 2011.
- [24] Cohen A.H., “Evolution of the Vertebrate Central Pattern Generator for Locomotion,” in *Neural Control of Rhythmic Movements in Vertebrates*, Cohen A.H., Rossignol S., and Grillner S., Eds. John Wiley & Sons, New York, 1988.
- [25] Daan S. and Belterman T., “Lateral Bending in Locomotion of Some Lower Tetrapods,” in *Proc. Ned. Akad. Wetten. Ser. C*, vol. 71, 1968, pp. 245–266.
- [26] Delvolv  I., Bem T., and Cabelguen J.-M., “Epaxial and Limb Muscle Activity During Swimming and Terrestrial Stepping in the Adult Newt, *Pleurodeles waltl*,” *Journal of Neurophysiology*, vol. 78, no. 2, pp. 638–650, 1997.
- [27] Delvolv  I., Branchereau P., Dubuc R., and Cabelguen J.-M., “Fictive Rhythmic Motor Patterns Induced by NMDA in an In Vitro Brain Stem–Spinal Cord Preparation from an Adult Urodele,” *Journal of Neurophysiology*, vol. 82, no. 2, pp. 1074–1077, 1999.
- [28] DiBiase J.H., Silverman H.F., and Brandstein M.S., “Robust Localization in Reverberant Rooms,” in *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Verlag, 2001, pp. 157–180.
- [29] Edgecomb R.S., Robert D., Read M.P., and Hoy R.R., “The Tympanal Hearing Organ of a Fly: Phylogenetic Analysis of its Morphological Origins,” *Cell and Tissue Research*, vol. 282, pp. 251–268, 1995.
- [30] Ekeberg  , “A Combined Neuronal and Mechanical Model of Fish Swimming,” *Biological Cybernetics*, vol. 69, no. 5, pp. 363–374, 1993.
- [31] Ekeberg  , Wall n P., Lansner A., Tr v n H., Brodin L., and Grillner S., “A Computer-based Model for Realistic Simulations of Neural Networks I. The Single Neuron and Synaptic Interaction,” *Biological Cybernetics*, vol. 65, no. 2, pp. 81–90, 1991.
- [32] Fay R.R. and Popper A.N., “Introduction to Sound Source Localization,” in *Sound*

- Source Localization*, ser. Springer Handbook of Auditory Research, Popper A.N. and Fay R.R., Eds. Springer New York, 2005, vol. 25, pp. 1–5.
- [33] Fischer B.J., “Optimal Models of Sound Localization by Barn Owls,” in *Advances in Neural Information Processing Systems 20*, Platt J.C., Koller D., Y. Singer, and S. Roweis, Eds. MIT Press, 2008, pp. 449–456.
- [34] Fisher R.A. and Yates F., *Statistical Tables for Biological, Agricultural and Medical Research*. Oliver and Boyd, 1963.
- [35] Fletcher N.H., *Acoustic Systems in Biology*. Oxford University Press, USA, 1992.
- [36] Fletcher N.H. and Thwaites S., “Physical Models for the Analysis of Acoustical Systems in Biology,” *Quarterly Reviews of Biophysics*, vol. 12, no. 1, pp. 25–65, 1979.
- [37] Frolich L.M. and Biewener A.A., “Kinematic and Electromyographic Analysis of the Functional Role of the Body Axis during Terrestrial and Aquatic Locomotion in the Salamander *Ambystoma tigrinum*,” *Journal of Experimental Biology*, vol. 162, no. 1, pp. 107–130, 1992.
- [38] Gao K.Q. and Shubin N.H., “Late Jurassic Salamanders from Northern China,” *Nature*, vol. 410, no. 6828, pp. 574–577, 2001.
- [39] Gelfand S.A., *Hearing: An Introduction to Psychological and Physiological Acoustics*. Informa HealthCare, 2004.
- [40] Grillner S., “Neural Control of Locomotion in Lower Vertebrates: From Behavior to Ionic Mechanisms,” in *Neural Control of Rhythmic Movements in Vertebrates*. John Wiley & Sons, 1988, pp. 1–40.
- [41] Grillner S., Deliagina T., El Manira A., Hill R.H., Orlovsky G.N., Wallén P., Ekeberg Ö., and Lansner A., “Neural Networks that Co-ordinate Locomotion and Body Orientation in Lamprey,” *Trends in Neurosciences*, vol. 18, no. 6, pp. 270–279, 1995.
- [42] Grillner S., Wallén P., Brodin L., and Lansner A., “Neuronal Network Generating Locomotor Behavior in Lamprey: Circuitry, Transmitters, Membrane Properties, and Simulation,” *Annual Review of Neuroscience*, vol. 14, no. 1, pp. 169–199, 1991.
- [43] Haykin S., *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, 1999.

- [44] Heffner H.E. and Heffner R.S., "Hearing," *Comparative Psychology: A Handbook*, pp. 290–303, 1998.
- [45] Heffner H.E. and Heffner R.S., "High-frequency hearing," *The Senses: A Comprehensive Reference, Volume 3: Audition*, pp. 55–60, 2008.
- [46] Heffner R.S., "Comparative Study of Sound Localization and its Anatomical Correlates in Mammals," *Acta Oto-Laryngologica*, vol. 117, no. S532, pp. 46–53, 1997.
- [47] Heffner R.S., "Comparative study of sound localization and its anatomical correlates in mammals," *Acta Oto-Laryngologica*, vol. 117, no. S532, pp. 46–53, 1997.
- [48] Hill K.G. and Boyan G.S., "Directional Hearing in Crickets," *Nature*, vol. 262, no. 5567, pp. 390–391, 1976.
- [49] Hill K.G. and Boyan G.S., "Sensitivity to Frequency and Direction of Sound in the Auditory System of Crickets (gryllidae)," *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 121, no. 1, pp. 79–97, 1977.
- [50] Holland O., "The First Biologically Inspired Robots," *Robotica*, vol. 21, pp. 351–363, 2003.
- [51] Hooper S.L., "Central Pattern Generators," in *eLS*. John Wiley & Sons, 2001.
- [52] Horchler A.D., Reeve R.E., Webb B., and Quinn R.D., "Robot Phonotaxis in the Wild: A Biologically Inspired Approach to Outdoor Sound Localization," *Advanced Robotics*, vol. 18, no. 8, pp. 801–816, 2004.
- [53] Huang J., Kume K., Saji A., Nishihashi M., Watanabe T., and Martens W.L., "Robotic Spatial Sound Localization and its 3D Sound Human Interface," in *Cyber Worlds, 2002. Proceedings. First International Symposium on*, 2002, pp. 191–197.
- [54] Huber F., Moore T.E., and Loher W., *Cricket Behavior and Neurobiology*. Cornell University Press, 1989.
- [55] Hwang D. and Choi J., "Real-Time Binaural Sound Source Localization using Sparse Coding and SOM," in *Intelligent Robotics and Applications*, ser. Lecture Notes in Computer Science, Liu H., Ding H., Xiong Z., and Zhu X., Eds. Springer Berlin / Heidelberg, 2010, vol. 6424, pp. 582–589.
- [56] Ijspeert A.J., "A 3-D Biomechanical Model of the Salamander," in *Proceedings of*

- the Second International Conference on Virtual Worlds*, Heudin J.C., Ed. Springer Verlag, 2000, pp. 225–234.
- [57] Ijspeert A.J., “A Connectionist Central Pattern Generator for the Aquatic and Terrestrial Gaits of a Simulated Salamander,” *Biological Cybernetics*, vol. 84, no. 5, pp. 331–348, 2001.
- [58] Ijspeert A.J., “Central Pattern Generators for Locomotion Control in Animals and Robots: A Review,” *Neural Networks*, vol. 21, no. 4, pp. 642–653, 2008.
- [59] Ijspeert A.J., Crespi A., and Cabelguen J.-M., “Simulation and Robotics Studies of Salamander Locomotion. Applying Neurobiological Principles to the Control of Locomotion in Robots,” *Neuroinformatics*, vol. 3, no. 3, pp. 171–196, 2005.
- [60] Ijspeert A.J., Crespi A., Ryczko D., and Cabelguen J.-M., “From Swimming to Walking with a Salamander Robot Driven by a Spinal Cord Model,” *Science*, vol. 315, no. 5817, pp. 1416–1420, 2007.
- [61] Isley T.E. and Gysel L.W., “Sound-source Localization by the Red Fox,” *Journal of Mammalogy*, vol. 56, no. 2, pp. 397–404, 1975.
- [62] Kagami S., Mizoguchi H., Tamai Y., and Kanade T., “Microphone Array for 2D Sound Localization and Capture,” in *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, vol. 1, 2004, pp. 703–708.
- [63] Karakasiliotis K. and Ijspeert A.J., “Analysis of the Terrestrial Locomotion of a Salamander Robot,” in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, Oct. 2009, pp. 5015–5020.
- [64] Keyrouz F. and Diepold K., “A Novel Biologically Inspired Neural Network Solution for Robotic 3D Sound Source Sensing,” *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 12, pp. 721–729, 2008.
- [65] Khanna S.M., Koester C.J., Willemin J.F., Daendliker R., and Rosskothen H., “Noninvasive Optical System for the Study of the Function of Inner Ear in Living Animals,” in *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, Tuchin V.V., Ed., vol. 2732, 1996, pp. 64–81.
- [66] Kim H.D., Komatani K., Ogata T., and Okuno H.G., “Design and Evaluation of Two-channel-based Sound Source Localization over Entire Azimuth Range for Moving Talkers,” in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ*

- International Conference on*, 2008, pp. 2197–2203.
- [67] Klump G.M., “Sound Localization in Birds,” in *Comparative Hearing: Birds and Reptiles*, ser. Springer Handbook of Auditory Research, Dooling R.J., Fay R.R., and Popper A.N., Eds. Springer-Verlag, 2000, vol. 13, pp. 249–307.
- [68] Knapp C. and Carter G., “The Generalized Correlation Method for Estimation of Time Delay,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 4, pp. 320–327, 1976.
- [69] Knudsen E.I., Blasdel G.G., and Konishi M., “Sound Localization by the Barn Owl (*Tyto alba*) measured with the Search Coil Technique,” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 133, pp. 1–11, 1979.
- [70] Knudsen E.I. and Konishi M., “Mechanisms of Sound Localization in the Barn Owl (*Tyto alba*),” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 133, pp. 13–21, 1979.
- [71] Kobayashi T., Kameda Y., and Ohta Y., “Sound Source Localization with Non-calibrated Microphones,” in *Proceedings of the 14th International Conference on Advances in Multimedia Modeling*. Springer Verlag, 2008, pp. 134–143.
- [72] Kornatowski E., “Localization of Sound Source Direction in Real Time,” in *Advances in Multimedia and Network Information System Technologies*, ser. Advances in Intelligent and Soft Computing, Nguyen N., Zgrzywa A., and Czyzewski A., Eds. Springer Berlin / Heidelberg, 2010, vol. 80, pp. 39–47.
- [73] Kwon B., Park Y., and Park Y., “Sound Source Localization for Robot Auditory System using the Summed GCC Method,” in *Control, Automation and Systems, 2008. ICCAS 2008. International Conference on*, 2008, pp. 241–245.
- [74] Lewinger W.A., Harley C.M., Ritzmann R.E., Branicky M.S., and Quinn R.D., “Insect-like Antennal Sensing for Climbing and Tunneling Behavior in a Biologically-inspired Mobile Robot,” in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, 2005, pp. 4176–4181.
- [75] Li D. and Levinson S.E., “A Bayes-rule based Hierarchical System for Binaural Sound Source Localization,” in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 5, 2003,

- pp. V–521–4.
- [76] Lisiewski A.P., Liu H.J., Yu M., Curran L., and Gee D., “Fly-ear Inspired Micro-sensor for Sound Source Localization in Two Dimensions,” *The Journal of the Acoustical Society of America*, vol. 129, no. 5, p. EL166, 2011.
  - [77] Liu H. and Shen M., “Continuous Sound Source Localization based on Micro-phone Array for Mobile Robots,” in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010, pp. 4332–4339.
  - [78] Lund H.H., Webb B., and Hallam J., “A Robot Attracted to the Cricket Species *Gryllus bimaculatus*,” in *Fourth European Conference on Artificial Life*, 1997, pp. 246–255.
  - [79] Mann H.B. and Whitney D.R., “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other,” *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.
  - [80] Mardia K.V. and Jupp P.E., *Directional Statistics*. John Wiley & Sons, 2000.
  - [81] Mason A.C., Lee N., and Oshinsky M.L., “The Start of Phonotactic Walking in the Fly *Ormia ochracea*: A Kinematic Study,” *Journal of Experimental Biology*, vol. 208, no. 24, pp. 4699–4708, 2005.
  - [82] Mason A.C., Oshinsky M.L., and Hoy R.R., “Hyperacute Directional Hearing in a Microscale Auditory System,” *Nature*, vol. 410, no. 6829, pp. 686–690, 2001.
  - [83] Masterton B., Heffner H.E., and Ravizza R., “The Evolution of Human Hearing,” *The Journal of the Acoustical Society of America*, vol. 45, pp. 966–985, 1969.
  - [84] McClellan A.D., “Control of Locomotion in a Lower Vertebrate, the Lamprey: Brainstem Command Systems and Spinal Cord Regeneration,” *American Zoologist*, vol. 29, no. 1, pp. 37–51, 1989.
  - [85] Michel O., “Cyberbotics Ltd. Webots<sup>TM</sup>: Professional Mobile Robot Simulation,” *International Journal of Advanced Robotic Systems*, vol. 1, no. 1, pp. 39–42, 2004.
  - [86] Michelsen A., “Biophysics of Sound Localization in Insects,” in *Comparative Hearing: Insects*, ser. Springer Handbook of Auditory Research, Hoy R.R., Popper A.N., and Fay R.R., Eds. Springer-Verlag, 1998, vol. 10, pp. 18–62.
  - [87] Michelsen A., “The Tuned Cricket,” *Physiology*, vol. 13, no. 1, pp. 32–38, 1998.

- [88] Michelsen A. and Larsen O.N., “Biophysics of the Ensiferan Ear,” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 123, pp. 193–203, 1978.
- [89] Michelsen A. and Nocke H., “Biophysical Aspects of Sound Communication in Insects,” *Advances in Insect Physiology*, vol. 10, pp. 247–296, 1974.
- [90] Michelsen A., Popov A.V., and Lewis B., “Physics of Directional Hearing in the Cricket *Gryllus bimaculatus*,” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 175, no. 2, pp. 153–164, 1994.
- [91] Michelsen A., Popov A.V., and Lewis B., “Physics of Directional Hearing in the Cricket *Gryllus bimaculatus*,” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 175, pp. 153–164, 1994.
- [92] Miles R.N., Robert D., and Hoy R.R., “Mechanically Coupled Ears for Directional Hearing in the Parasitoid Fly *Ormia ochracea*,” *Journal of the Acoustical Society of America*, vol. 98, no. 6, pp. 3059–3070, 1995.
- [93] Moiseff A., “Binaural Disparity Cues Available to the Barn Owl for Sound Localization,” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 164, pp. 629–636, 1989.
- [94] Moiseff A. and Konishi M., “The Owl’s Interaural Pathway is not Involved in Sound Localization,” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 144, pp. 299–304, 1981.
- [95] Mungamuru B. and Aarabi P., “Enhanced Sound Localization,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, no. 3, pp. 1526–1540, 2004.
- [96] Murray J., Erwin H., and Wermter S., “A Hybrid Architecture using Cross-correlation and Recurrent Neural Networks for Acoustic Tracking in Robots,” *Biomimetic Neural Learning for Intelligent Robots*, pp. 97–97, 2005.
- [97] O’Reilly J.C., Summers A.P., and Ritter D.A., “The Evolution of the Functional Role of Trunk Muscles during Locomotion in Adult Amphibians,” *American Zoologist*, vol. 40, no. 1, pp. 123–135, 2000.
- [98] Payne R.S., “How the Barn Owl Locates Prey by Hearing,” *Living Bird, First*

- Annual of the Cornell Laboratory of Ornithology*, vol. 1, pp. 151–159, 1962.
- [99] Pearson K., “On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling,” *Breakthroughs in Statistics: Methodology and Distribution*, vol. 2, pp. 11–28, 1992.
- [100] Reeve R.E. and Webb B.H., “New Neural Circuits for Robot Phonotaxis,” *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 361, no. 1811, pp. 2245–2266, 2003.
- [101] Ritter D., “Lateral Bending during Lizard Locomotion,” *Journal of Experimental Biology*, vol. 173, no. 1, pp. 1–10, 1992.
- [102] Ritter D., “Axial Muscle Function during Lizard Locomotion,” *Journal of Experimental Biology*, vol. 199, no. 11, pp. 2499–2510, 1996.
- [103] Robert D., “Innovative Biomechanics for Directional Hearing in Small Flies,” *The Biological Bulletin*, vol. 200, no. 2, pp. 190–194, 2001.
- [104] Robert D., Miles R.N., and Hoy R.R., “Directional Hearing by Mechanical Coupling in the Parasitoid Fly *Ormia ochracea*,” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 179, no. 1, pp. 29–44, 1996.
- [105] Robert D., Miles R.N., and Hoy R.R., “Tympanal Mechanics in the Parasitoid Fly *Ormia ochracea*: Intertympanal Coupling during Mechanical Vibration,” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 183, no. 4, pp. 443–452, 1998.
- [106] Robert D., Read M.P., and Hoy R.R., “The Tympanal Hearing Organ of the Parasitoid Fly *Ormia ochracea* (Diptera, Tachinidae, Ormiini),” *Cell and Tissue Research*, vol. 275, pp. 63–78, 1994.
- [107] Rodemann T., Ince G., Joublin F., and Goerick C., “Using Binaural and Spectral Cues for Azimuth and Elevation Localization,” in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, 2008, pp. 2185–2190.
- [108] Roos P.J., “Lateral Bending in Newt Locomotion,” in *Proc. Ned. Akad. Wetten. Ser. C*, vol. 67, 1964, pp. 223–232.
- [109] Roth G., Nishikawa K.C., Naujoks-Manteuffel C., Schmidt A., and Wake D.B.,

- “Paedomorphosis and Simplification in the Nervous System of Salamanders,” *Brain, Behavior and Evolution*, vol. 42, no. 3, pp. 137–152, 1993.
- [110] Roth G., Nishikawa K.C., and Wake D.B., “Genome Size, Secondary Simplification, and the Evolution of the Brain in Salamanders,” *Brain, Behavior and Evolution*, vol. 50, no. 1, pp. 50–59, 1997.
- [111] Schnupp J., Nelken I., and King A., *Auditory Neuroscience: Making Sense of Sound*. The MIT Press, 2010.
- [112] Schöneich S. and Hedwig B., “Hyperacute Directional Hearing and Phonotactic Steering in the Cricket (*Gryllus bimaculatus* deGeer),” *PLoS ONE*, vol. 5, no. 12, p. e15141, 12 2010.
- [113] Schul J., “Song Recognition by Temporal Cues in a Group of Closely Related Bushcricket Species (genus *Tettigonia*),” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 183, pp. 401–410, 1998.
- [114] Shaikh D., Hallam J., and Christensen-Dalsgaard J., “Control of a Braitenberg Lizard in a Phonotaxis Task with Decision Models,” in *Technical Report Series: Proceedings of Towards Autonomous Robotic Systems*, Kyriacou T., Nehmzow U., Melhuish C., and Witkowski M., Eds., University of Ulster. Intelligent Systems Research Centre, University of Ulster, 2009, pp. 48–54.
- [115] Shaikh D., Hallam J., and Christensen-Dalsgaard J., “Modifying Directionality through Auditory System Scaling in a Robotic Lizard,” in *From Animals to Animats 11*, ser. Lecture Notes in Computer Science, Doncieux S., Girard B., Guillot A., Hallam J., Meyer J.A., and Mouret J.B., Eds. Springer Berlin / Heidelberg, 2010, vol. 6226, pp. 82–92.
- [116] Shaikh D., Hallam J., Christensen-Dalsgaard J., and Ijspeert A.J., “Combining Bio-inspired Sensing with Bio-inspired Locomotion,” in *The 5th International Symposium on Adaptive Motion in Animals and Machines (AMAM 2011)*, 2011, pp. 27–28.
- [117] Shaikh D., Hallam J., Christensen-Dalsgaard J., and Zhang L., “A Braitenberg Lizard: Continuous Phonotaxis with a Lizard Ear Model,” in *Bioinspired Applications in Artificial and Natural Computation*, ser. Lecture Notes in Computer Science, Mira J., Ferrndez J., lvarez J., Paz F.de la , and Toledo F., Eds. Springer

- Berlin / Heidelberg, 2009, vol. 5602, pp. 439–448.
- [118] Strumillo P., *Advances in Sound Localization*. InTech, 2011.
- [119] Tamai Y., Kagami S., Amemiya Y., Sasaki Y., Mizoguchi H., and Takano T., “Circular Microphone Array for Robot’s Audition,” in *Sensors, 2004. Proceedings of IEEE*, vol. 2, 2004, pp. 565–570.
- [120] Tamai Y., Sasaki Y., Kagami S., and Mizoguchi H., “Three Ring Microphone Array for 3D Sound Localization and Separation for Mobile Robot Audition,” in *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, 2005, pp. 4172–4177.
- [121] Valin J.-M., Michaud F., Rouat J., and Letourneau D., “Robust Sound Source Localization using a Microphone Array on a Mobile Robot,” in *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 2, 2003, pp. 1228–1233.
- [122] van Trees H.L., *Optimum Array Processing*. Wiley Online Library, 2002.
- [123] Vogel U., Zahnert T., Hofmann G., Offergeld C., and Hüttenbrink K.-B., “Laser Vibrometry of the Middle Ear: Opportunities and Limitations,” in *Middle Ear Mechanics in Research and Otosurgery. Proceedings of the International Workshop on Middle Ear Mechanics and Otosurgery, Dresden, Germany, September 19-22, 1996*, Hüttenbrink K.-B., Ed. Dept. of Oto-Rhino-Laryngology, University Hospital Carl Gustav Carus, Dresden University of Technology, 1997, pp. 128–133.
- [124] von Campenhausen M. and Wagner H., “Influence of the Facial Ruff on the Sound-receiving Characteristics of the Barn Owls Ears,” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 192, pp. 1073–1082, 2006.
- [125] Walker T.J., “Phonotaxis in Female *Ormia ochracea* (Diptera: Tachinidae), A Parasitoid of Field Crickets,” *Journal of Insect Behavior*, vol. 6, no. 3, pp. 389–410, 1993.
- [126] Wallén P., Ekeberg Ö., Lansner A., Brodin L., Tråvén H., and Grillner S., “A Computer-based Model for Realistic Simulations of Neural Networks II. The Segmental Network Generating Locomotor Rhythmicity in the Lamprey,” *Journal of Neurophysiology*, vol. 68, no. 6, pp. 1939–1950, 1992.

- [127] Walter W.G., "An Imitation of Life," *Scientific American*, vol. 182, no. 5, pp. 42–45, 1950.
- [128] Walter W.G., "A Machine that Learns," *Scientific American*, vol. 185, no. 2, pp. 60–63, 1951.
- [129] Webb B. and Harrison R.R., "Integrating Sensorimotor Systems in a Robot Model of Cricket Behavior," in *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, vol. 4196. Society of Photo-Optical Instrumentation Engineers, 2000, pp. 113–124.
- [130] Webb B. and Scutt T., "A Simple Latency-Dependent Spiking-neuron Model of Cricket Phonotaxis," *Biological Cybernetics*, vol. 82, no. 3, pp. 247–269, 2000.
- [131] Weber T. and Thorson J., "Phonotactic Behavior of Walking Crickets," in *Cricket Behavior and Neurobiology*, 1989, vol. 10, pp. 310–339.
- [132] Werner Y.L., Igic P.G., Seifan M., and Saunders J.C., "Effects of Age and Size in the Ears of Gekkonomorph Lizards: Middle-ear Sensitivity," *Journal of Experimental Biology*, vol. 205, no. 20, pp. 3215–3223, 2002.
- [133] Werner Y.L., Montgomery L.G., Safford S.D., Igic P.G., and Saunders J.C., "How Body Size Affects Middle-ear Structure and Function and Auditory Sensitivity in Gekkonoid Lizards," *Journal of Experimental Biology*, vol. 201, no. 4, pp. 487–502, 1998.
- [134] Werner Y.L., Montgomery L.G., Seifan M., and Saunders J.C., "Effects of Age and Size in the Ears of Gekkotan Lizards: Auditory Sensitivity, its Determinants, and New Insights into Tetrapod Middle-ear Function," *Pflügers Archiv European Journal of Physiology*, vol. 456, pp. 951–967, 2008.
- [135] Wever E.G., *The Reptile Ear: Its Structure and Function*. Princeton University Press, 1978.
- [136] Wu W.-C., Hsieh C.-H., Huang H.-C, and Chen O.T.C., "Hearing Aid System with 3D Sound Localization," in *TENCON 2007 - 2007 IEEE Region 10 Conference*, 2007, pp. 1–4.
- [137] Zhang L., "Modelling Directional Hearing in Lizards," Ph.D. dissertation, Maersk Mc-Kinney Moller Institute, Faculty of Engineering, University of Southern Denmark, 2009.

- [138] Zhang L., Hallam J., and Christensen-Dalsgaard J., “Modelling the Peripheral Auditory System of Lizards,” in *From Animals to Animats 9*, ser. Lecture Notes in Computer Science, Nolfi S., Baldassarre G., Calabretta R., Hallam J., Marocco D., Meyer J.A., Miglino O., and Parisi D., Eds. Springer Berlin / Heidelberg, 2006, vol. 4095, pp. 65–76.
- [139] Zhang L., Hallam J., and Christensen-Dalsgaard J., “Modelling Asymmetry in the Peripheral Auditory System of the Lizard,” *Artificial Life and Robotics*, vol. 13, pp. 5–9, 2008.
- [140] Zhang L., Hallam J., and Christensen-Dalsgaard J., “Modelling the Lizard Ear: Directness of Phonotaxis with Noise Distractor,” in *Proceedings of Towards Autonomous Robotic Systems 2008 (TAROS 2008)*, 2008, pp. 193–200.

# List of figures

1.1	Two lizard species with a directional peripheral auditory system. . . . .	3
1.2	The structure of the lizard ear. . . . .	5
2.1	Different mechanisms of directional hearing observed in nature. . . . .	11
2.2	The barn owl ( <i>Tyto alba</i> ). . . . .	13
2.3	The barn owl ear placement. . . . .	13
2.4	Two subfamilies of crickets. . . . .	14
2.5	The cricket ear structure. . . . .	15
2.6	The parasitoid fly <i>Ormia ochracea</i> . . . . .	16
2.7	The anatomy of the <i>Ormia ochracea</i> ear. . . . .	18
2.8	The inter-tympanal bridge and its mechanical model. . . . .	18
2.9	The single point surround microphone TetraMic from Core Sound Inc.. .	21
2.10	One of the robotic tortoises built by W. Grey Walter. . . . .	22
2.11	Two of the different robots used in cricket hearing research. . . . .	23
2.12	A fire salamander or <i>Salamandra salamandra</i> . . . . .	24
2.13	The first version of the salamander robot, <i>Salamandra robotica</i> . . . . .	25
2.14	The auditory task model proposed by Lei Zhang. . . . .	26
3.1	Lumped-parameter model of the lizard's peripheral auditory system. . . .	30
3.2	Directionality of the peripheral auditory system model. . . . .	32

3.3	The digital implementation of the peripheral auditory system. . . . .	33
3.4	The response in dB of the digital implementation of the auditory model. . . . .	34
3.5	Simple Braitenberg vehicles with light sensors. . . . .	35
3.6	The mobile robot platforms. . . . .	36
3.7	The supervisory control loop running on the ARM9 core. . . . .	38
3.8	The experimental arena and the initial positions of the robot. . . . .	39
3.9	Step vs. continuous control in terms of <i>hits</i> , <i>near hits</i> and <i>misses</i> . . . . .	41
3.10	Sample robot trajectories for 1600 Hz tone from the loudspeaker. . . . .	43
3.11	Trajectory directness in step-control vs. continuous control . . . . .	43
3.12	Braitenberg vehicle with embedded decision models. . . . .	46
3.13	The Lego NXT based mobile robot. . . . .	47
3.14	The experimental arena. . . . .	48
3.15	Example trajectories. . . . .	49
3.16	Performance in terms of the number of <i>hits</i> , <i>near hits</i> and <i>misses</i> . . . . .	50
3.17	Example directness plots for the <b>AVG</b> and <b>MAX</b> decision models. . . . .	52
3.18	Trajectory directness for the <b>MAX</b> model. . . . .	53
3.19	Trajectory directness for the <b>AVG</b> model. . . . .	53
4.1	Functional head size versus highest audible frequency. . . . .	59
4.2	The response of the scaled auditory models in dB. . . . .	61
4.3	The mobile robot. . . . .	61
4.4	The microphone assembly. . . . .	62
4.5	The experimental arena. . . . .	63
4.6	Phonotactic performance in terms of the <i>hits</i> , <i>near-hits</i> and <i>misses</i> . . . . .	65
4.7	Average angular error $\theta_{\text{err}}$ in the trajectory heading. . . . .	68

5.1	The CMAC architecture. . . . .	72
5.2	The auditory model augmented with filterbanks. . . . .	75
5.3	The relative response in dB of individual bandpass filters in the filterbank.	76
5.4	The effect of the generalization parameter <b>C</b> on memory requirements. .	79
5.5	The mobile robot. . . . .	80
5.6	The modular implementation on the FPGA chip. . . . .	81
5.7	The microphone placement. . . . .	82
5.8	The sound direction representation after supervised learning. . . . .	86
5.9	The sound direction representation after unsupervised learning. . . . .	87
5.10	The sound direction representation in the simulated implementation. .	88
5.11	The distribution of the training data. . . . .	89
5.12	The experimental arena. . . . .	90
5.13	The sound direction representation learned by the robot after one epoch.	93
5.14	The evolution of the network output over time. . . . .	96
6.1	<i>Salamandra Robotica II</i> . . . . .	102
6.2	The CPG model. . . . .	104
6.3	Inter-module body CPG phase relationships. . . . .	105
6.4	The lizard-salamander hybrid model. . . . .	108
6.5	CPG oscillation amplitude $\mu$ vs. gain parameter $\alpha$ . . . . .	109
6.6	Effect of $\alpha$ on contralateral and ipsilateral body CPG oscillations.	110
6.7	The Webots <sup>TM</sup> model of <i>Salamandra Robotica II</i> . . . . .	111
6.8	An example of a raw trajectory and its smoothed version. . . . .	113
6.9	The phonotactic performance in terms of the “success” map. . . . .	114
6.10	Absolute angular heading error. . . . .	115

6.11 Example CPG oscillations during sound localization in simulation. . . . .	116
6.12 Snapshots of the locomotion pattern in simulation. . . . .	117
6.13 The dynamics of the relative tympanal vibration strengths. . . . .	118
6.14 The lizard-salamander hybrid robot. . . . .	120
6.15 Difference between simulation and robotic implementation for $\mu$ vs. $\alpha$ . .	123
6.16 The head trajectories recorded in simulated trials. . . . .	124
6.17 The head trajectories recorded in robotic trials. . . . .	125
6.18 Vibration strength dynamics in simulation and robotic trials. . . . .	126
7.1 The possible configurations utilizing three microphones. . . . .	131
7.2 The sound localization platform. . . . .	132
7.3 The experimental setup. . . . .	134
7.4 Localization performance in three dimensions. . . . .	135
7.5 The bias due to the control strategy. . . . .	136
7.6 A three microphone pair configuration. . . . .	138
A.1 Directness plots. . . . .	169
B.1 Directness plots for a control loop frequency of 1 Hz. . . . .	174
B.2 Directness plots for a control loop frequency of 3 Hz. . . . .	175
B.3 Directness plots for a control loop frequency of 10 Hz. . . . .	176
B.4 Directness plots for a control loop frequency of 10 Hz. . . . .	177

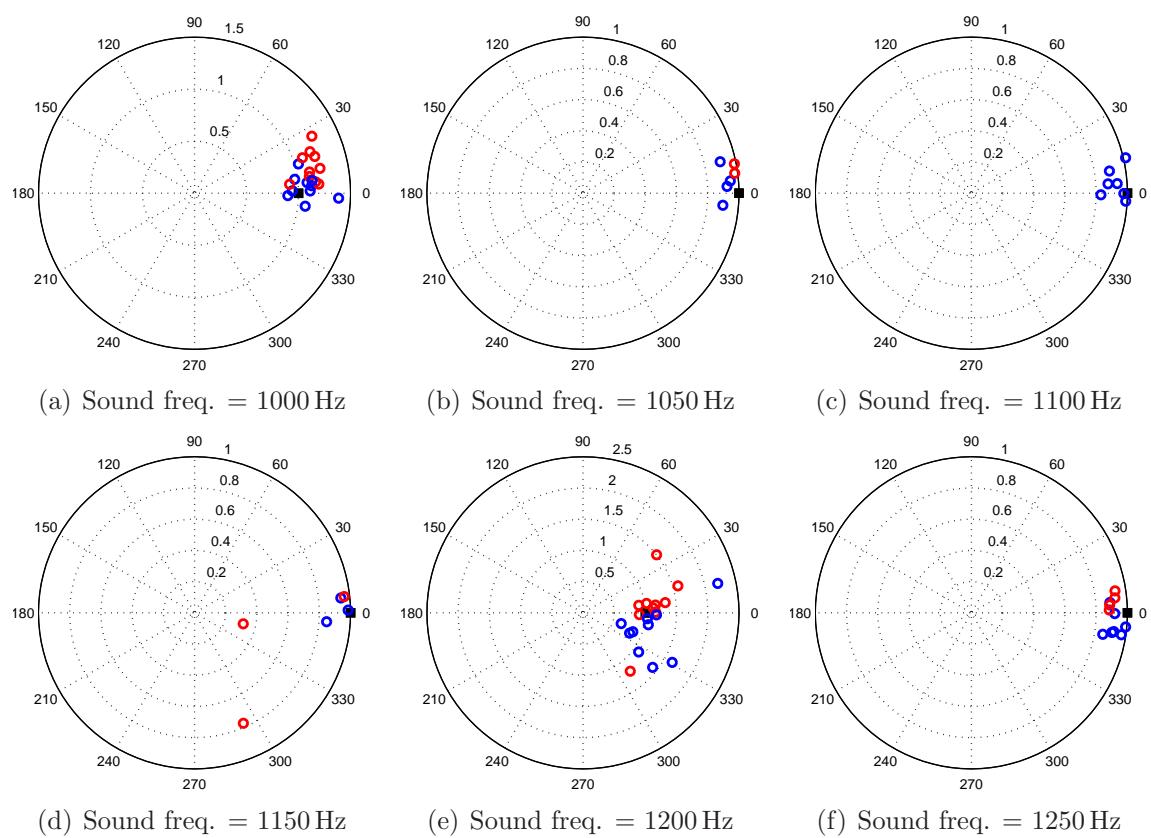
# List of tables

3.1	Mann-Whitney U values. . . . .	45
3.2	Total number of <i>hits</i> . . . . .	55
3.3	Results of the Chi-square test. . . . .	55

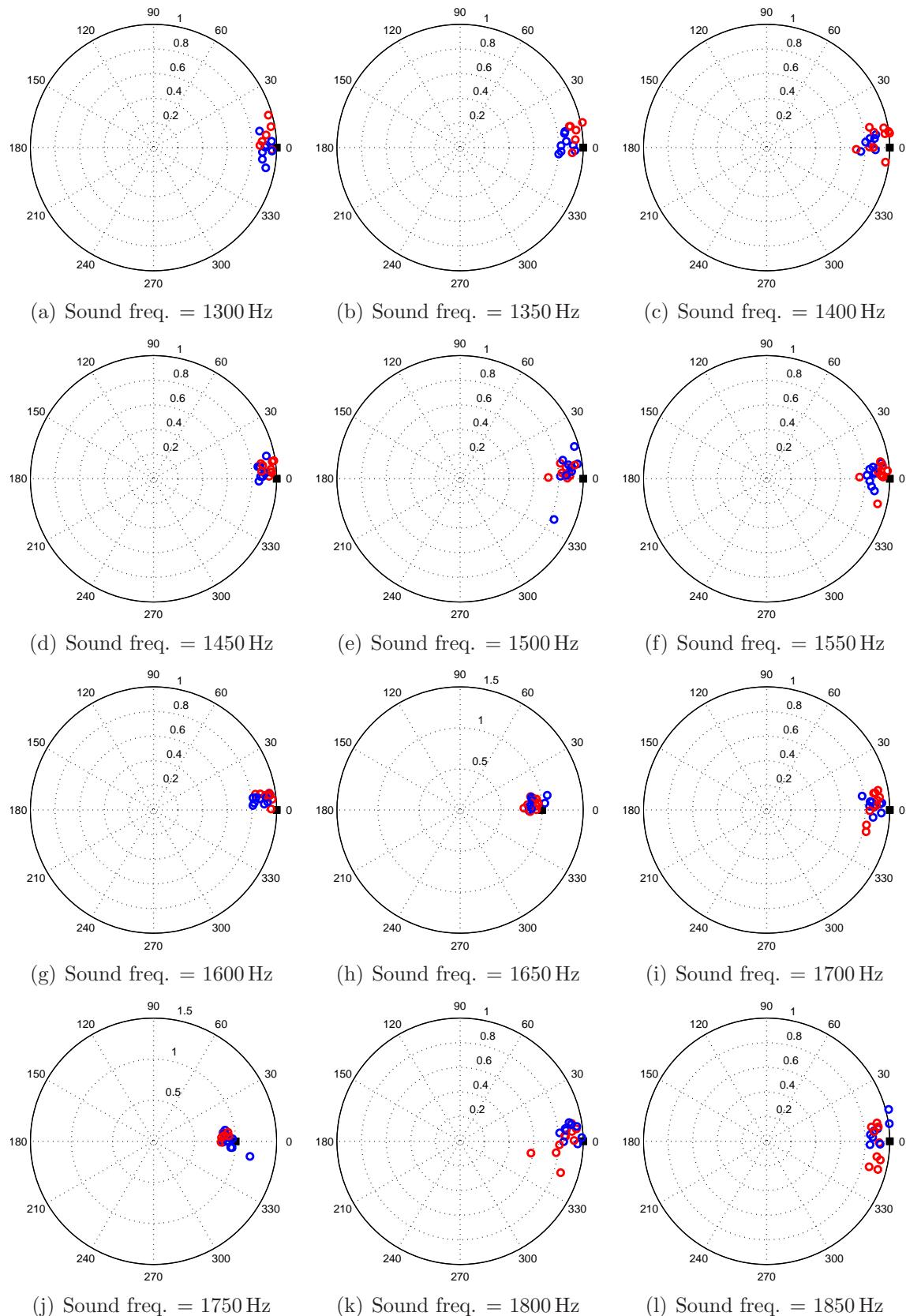


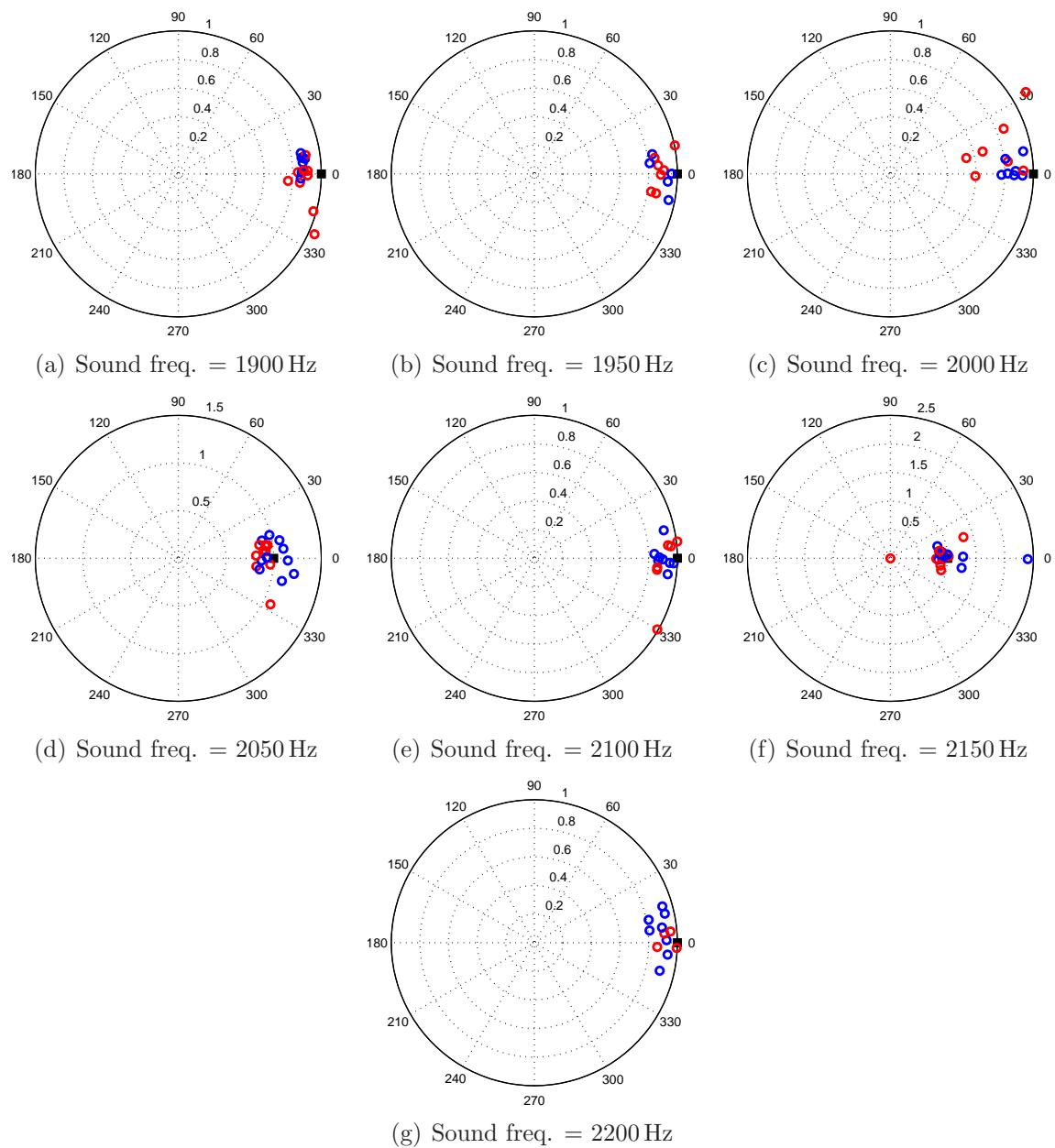
# Appendix A

Figure A.1 depicts the directness plots obtained during the phonotaxis experiments performed with the simple Braitenberg lizard (refer to Sect. 3.4). The trajectories in which the robot's initial position is to the left of the loudspeaker are represented by **blue** markers, while those in which the robot's initial position is to the right of the loudspeaker are represented by **red** markers.



**Figure A.1:** Directness plots.

**Figure A.1:** Directness plots (continued).

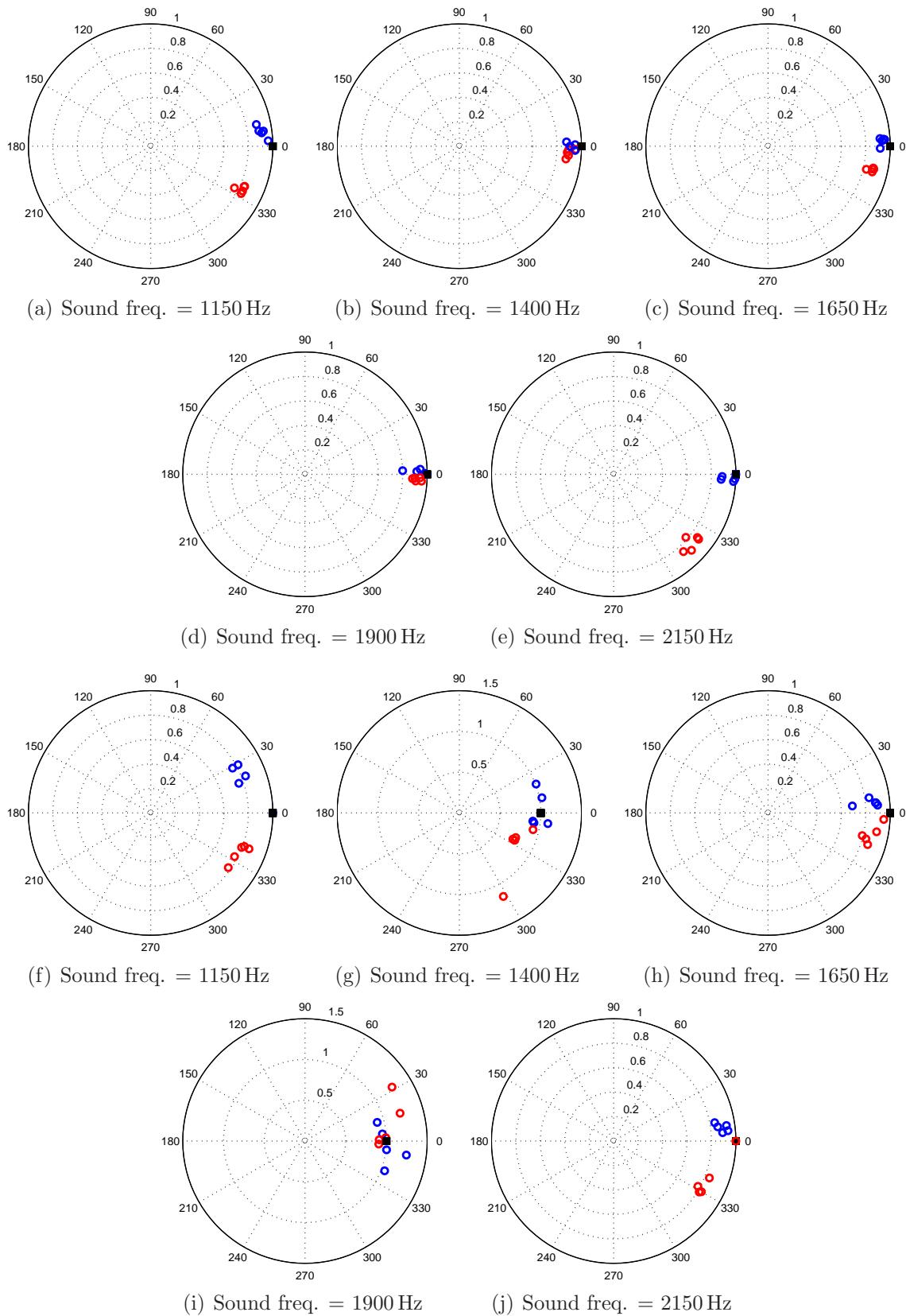


**Figure A.1:** Directness plots (continued).

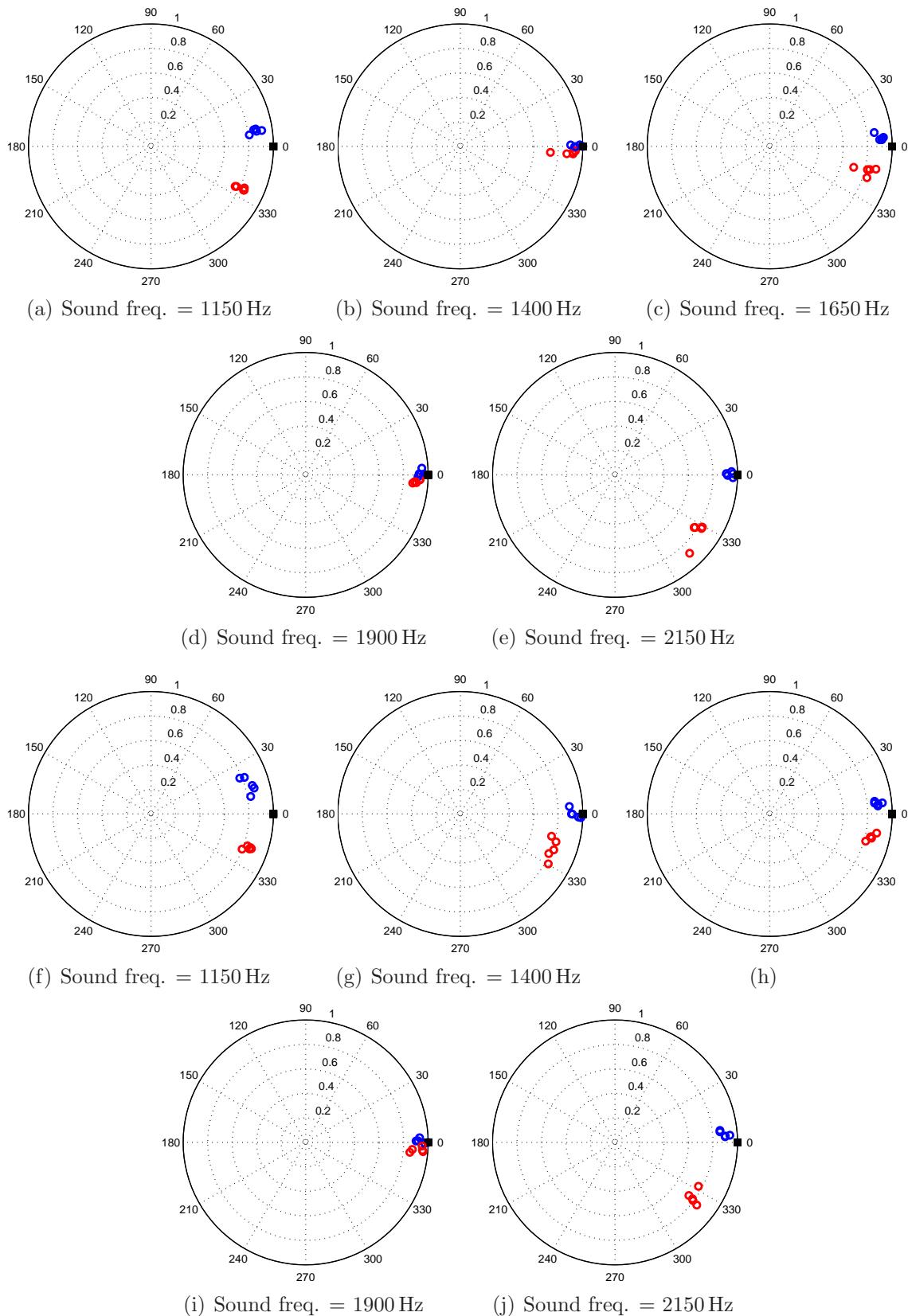


## Appendix B

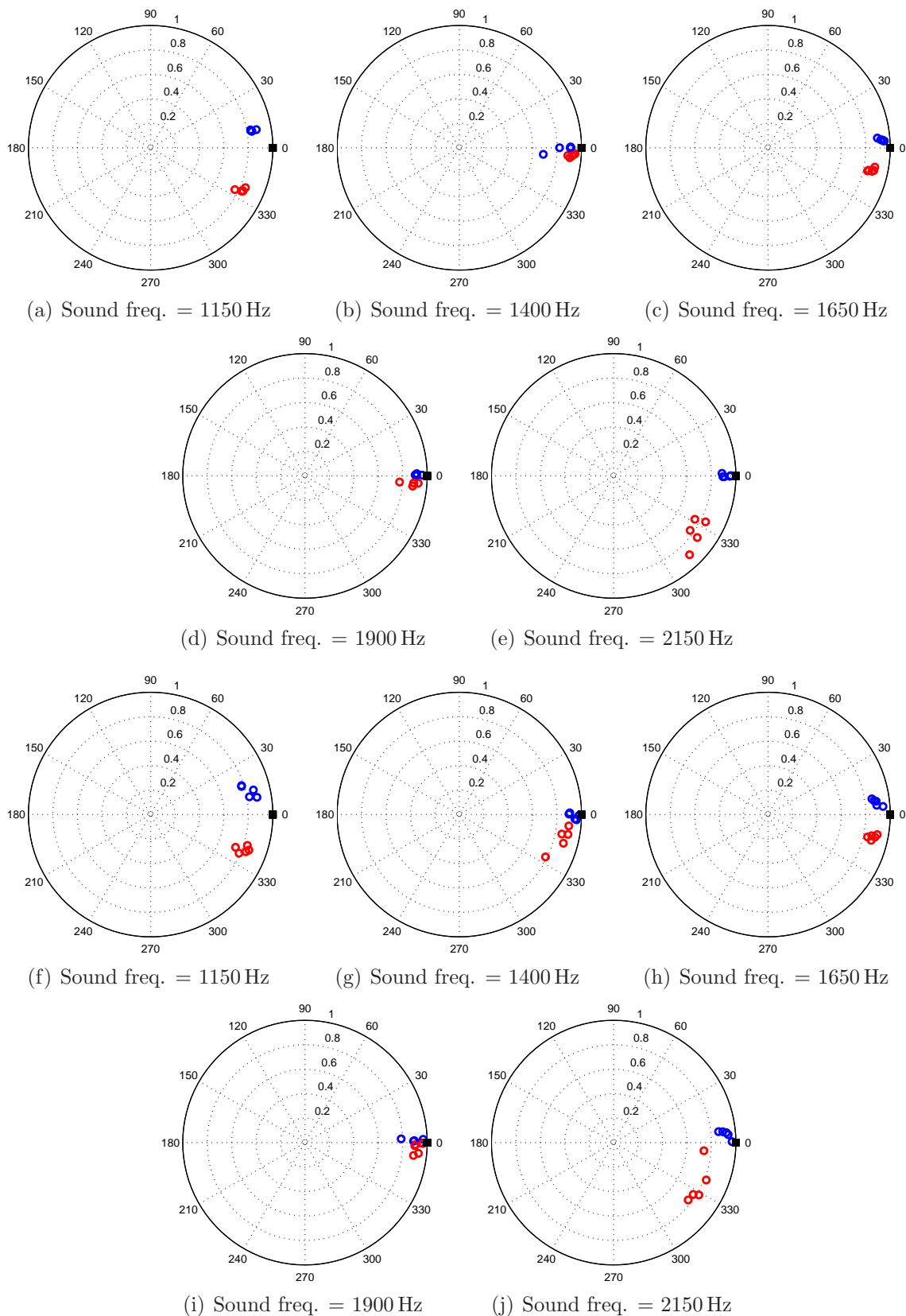
Figures B.1–B.4 comparatively depict the directness plots obtained during the phono-taxis experiments performed with the Braitenberg lizard with the **AVG** and **MAX** decision models embedded into the sensorimotor couplings, for the control loop frequencies of 1 Hz, 3 Hz, 10 Hz and 30 Hz (refer to Sect. 3.5). The trajectories towards the left loudspeaker are represented by **blue** markers, while those towards the right loudspeaker are represented by **red** markers.



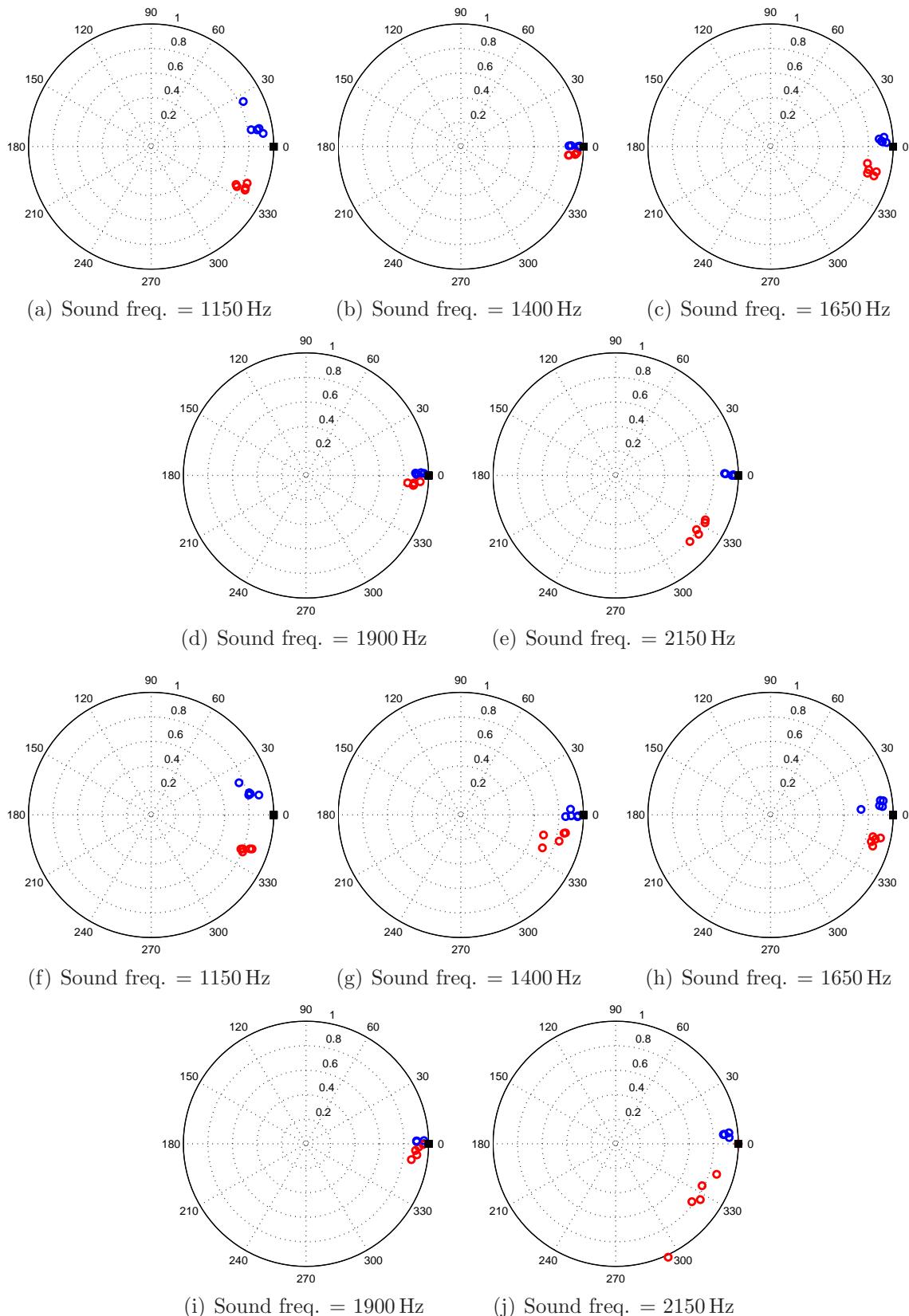
**Figure B.1:** Control loop freq. = 1 Hz. Figures (a)–(e) correspond to the **AVG** model, while Fig. (f)–(j) correspond to the **MAX** model.



**Figure B.2:** Control loop freq. = 3 Hz. Figures (a)–(e) correspond to the **AVG** model, while Fig. (f)–(j) correspond to the **MAX** model.



**Figure B.3:** Control loop frequency = 10 Hz. Figures (a)–(e) correspond to the **AVG** model, while Fig. (f)–(j) correspond to the **MAX** model.



**Figure B.4:** Control loop frequency = 30 Hz. Figures (a)–(e) correspond to the **AVG** model, while Fig. (f)–(j) correspond to the **MAX** model.