

A bio-inspired auditory perception model for amplitude-frequency clustering

Paolo Arena^a, Luigi Fortuna^a, Mattia Frasca^a, Gaetana Ganci^a and Luca Patané^a

^a Dipartimento di Ingegneria Elettrica, Elettronica e dei Sistemi, Università degli Studi di Catania, Viale A. Doria 6, 95125 Catania, Italy.

ABSTRACT

In this paper a model for auditory perception is introduced. This model is based on a network of integrate-and-fire and resonate-and-fire neurons and is aimed to control the phonotaxis behavior of a roving robot. The starting point is the model of phonotaxis in *Gryllus Bimaculatus*: the model consists of four integrate-and-fire neurons and is able of discriminating the calling song of male cricket and orienting the robot towards the sound source. This paper aims to extend the model to include an amplitude-frequency clustering. The proposed spiking network shows different behaviors associated with different characteristics of the input signals (amplitude and frequency). The behavior implemented on the robot is similar to the cricket behavior, where some frequencies are associated with the calling song of male crickets, while other ones indicate the presence of predators. Therefore, the whole model for auditory perception is devoted to control different responses (attractive or repulsive) depending on the input characteristics. The performance of the control system has been evaluated with several experiments carried out on a roving robot.

Keywords: auditory perception, spiking neural network, resonate-and-fire neurons, robot navigation

1. INTRODUCTION

The auditory system is an important sensory apparatus for living beings but for a long time the use of this sensorial stimulus in robotics has been set aside with respect to other sensory processes like vision and touch. Indeed the auditory signal can be a rich source of information; in fact several animals take advantages of the sensory system to locate a sound source and interpret the information carried by the sound signal in order to win their daily survival battle. For example the spatial localization of the auditory stimulus is a fundamental skill of the barn owl (i.e. *Tyto Alba*),¹ a nocturnal predator that adopt an acoustic prey localization during hunting. Another example is the ability of the bat (i.e. *Pipistrellus Chiroteri*)² that localizes preys and obstacles in the environment emitting a sound and analyzing the reflected echoes. The bat is able to acquire information about the speed of the prey through the Doppler effect, emitting a constant frequency signal. Moreover it can evaluate the distance from the target based on the time difference between the emission and the echo reception of a frequency modulated signal.

On the other hand the auditory system is fundamental also for preys. A demonstration is the *noctuid moths*³ that recognize the sound emitted by bats and react depending on the distance from the predator. Mostly important is also the rule of the auditory system in humans since ability to speak is the base of the interpersonal relations.

The sound contains several information that can be summarized in three points:

Who - identity of who sends the signal;

What - the message contained in the sound;

Where - the location of the source.

Further author information:

Send correspondence to Prof. Paolo Arena, E-mail: parena@diees.unict.it

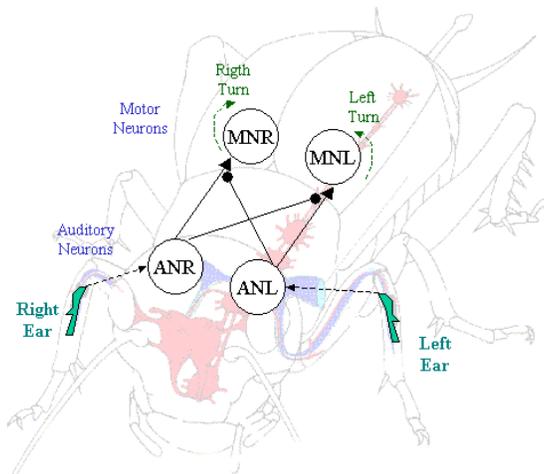


Figure 1. Simplified neural scheme designed to reproduce the cricket phonotaxis.

The localization of the sound source (i.e. directional hearing) can be carried out following several different methodologies.⁴ The most used are: the interaural time difference, that computes the arrival time difference of the sound in the two ears as a function of the origin direction, and the interaural level difference that computes the intensity difference of the sound perceived by the two ears.

In the following section a model based on cricket phonotaxis will be presented in details. Neuro-physiological studies have allowed the modelling of the auditory system; the model proposed in this paper consists of three layers. The first layer, made of integrate-and-fire neurons, consists of two auditory neurons, mutually inhibited and connected to two microphones equipped on the robot at a distance of 10 cm. The mutual inhibition creates a kind of *winner-take-all* competition, in which the winning neuron is the one associated with the highest amplitude signal. Therefore, the two neurons discriminate the amplitude difference between the two microphones. The second layer consists of an array of *resonate-and-fire* neurons tuned at different frequencies. This layer is devoted to frequency identification. The third layer consists of motor-neurons, modelled as integrate-and-fire neurons. Therefore the multi-layer spiking network that has been realized, performs the sensing-perception-action cycle, following the biological mechanisms of the cricket phonotaxis. In the next section a brief description of the cricket phonotactic system is given, in section 3 the bio-inspired auditory perception model for amplitude-frequency clustering is described in details and in section 4 the experimental results are reported. Finally section 5 draws the conclusions and shows some possible future developments.

2. CRICKET PHONOTAXIS

In nature several strategies are adopted by animals to communicate with others, a well-known example is the cricket song.

Male crickets open and close wings rhythmically, causing a vibration generating sound. This song is perceived by female crickets that are able to recognize the song and localize the source in order to reach the male. The cricket song typically consists of multiple groups of 4 syllables, named chirps, emitted with a gap of 20 ms followed by a 340 ms period of silence. The syllable is a 20 ms burst of a 4.7 khz sine wave. The cricket females are particularly selective to the Syllable Repetition Interval (SRI) in the calling song. The neural circuit underlying cricket phonotaxis was subject to several studies.^{5,6} The neural scheme taken into account in this section is a simplified model of the *Gryllus Bimaculatus*'s auditory process, proposed in Ref.⁷ The simplest model is constituted by only four neurons as schematized in Figure 1.

The auditory neurons (AN) represent the first stage of processing in the cricket's prothoracic ganglion in which a pair of neurons receives direct input from the auditory nerve. The motor neurons (MN) produce an output signal to turn left or right depending on which auditory neuron fires first. Each AN presents a direct excitatory synapse to the corresponding MN and an inhibitory synapse to the opposite MN; the connection scheme is similar to the control system of Braitenberg's vehicles.¹³ The AN-MN excitatory synapses exhibit depression, so the response is optimal for a signal with an appropriate temporal patterning. This basilar model is able to simulate the peculiar characteristics of the cricket behavior in terms of calling song recognition and sound source localization. In literature this simple model was refined including other elaboration stages as reported in Ref.⁸ Moreover the problem of sensorimotor integration of different behaviors can be investigated. It is known that the cricket combines its sound response with other sensorimotor activities such as optomotor reflex. For this reason different sensory systems must be integrated.⁹ Preliminary experiments that validate the feasibility of the application of auditory systems on robots were realized by B. Webb and R. Quinn as illustrated in Ref.^{10, 11}

In the next section a new model for the implementation of phonotaxis in bio-inspired robots will be proposed. The aim is to formulate a scheme based on nonlinear dynamical systems modelling neuron dynamics, able to localize a specific sound source and to discriminate different emission frequencies associated with appetitive or aversive stimuli. For example in crickets the calling song produces an attractive behavior, while the sound emitted by bats that are cricket predators induces an escape behavior.¹² While neural circuits responsible for cricket phonotaxis were clearly identified, the same does not hold for neuronal networks for phonophobic behavior. It is only hypothesised that that should be some similar neural circuitry guiding escape reactions from hearing stimuli. Based on these considerations, in this paper the same neural networks is used for both the behaviors, if a suitable clustering phase is introduced between the sensing and the motor stage of the whole network, made of biologically plausible neuron models. The solution, from an engineering point of view, adds flexibility to the bio-drawn phonotactic network, useful to build relevant perception systems.

3. A PHONOTAXIS NEURAL MODEL WITH AMPLITUDE-FREQUENCY CLUSTERING

Starting from the basic characteristics of the cricket phonotaxis neural scheme discussed in the previous section, a new model has been proposed to include an amplitude-frequency clustering. This model is based on a network of integrate-and-fire and resonate-and-fire neurons and is aimed to control the phonotaxis behavior in a roving robot. A complete scheme of the new model is shown in Figure 2, the structure is characterized by three different layers.

In the first layer two auditory neurons are employed to elaborate the signals coming from the sensory system (e.g. microphones equipped on the robot). The ANs are modelled as leaky integrate-and-fire (IF) neuron,^{14, 15} as follows:

$$\begin{cases} \dot{x} = -kx + \sum I_{syn} + I_{dis}y \\ y(x) = \begin{cases} 1 & \text{if } x > 1 \\ 0 & \text{if } x < 0 \\ y(x) & \text{otherwise} \end{cases} \end{cases} \quad (1)$$

where x is the membrane potential, y is the neuron output (i.e. a train of spikes), $\sum I_{syn}$ is the sum of the afferent synaptic currents, I_{dis} is the discharge current and k weights the leaky discharge. The parameter values of the IF neurons are given in Table 1. The IF neurons ANL and ANR are mutually inhibited so that there is a winner-take-all-effect. The winning neuron is the one associated with the highest amplitude signal; in this way, the two neurons discriminate the amplitude difference between the two microphones.

Therefore, the first layer is devoted to the localization of the sound source through the identification of the direction. The second layer consists of an array of resonate-and-fire neurons (RN)¹⁶ tuned at different frequencies. This layer is devoted to frequency clustering. The first layer is connected to the second layer through depression synapses that have been designed to associate a single spike to a burst coming from the corresponding AN.

The equation of the depression synapse is reported in following:

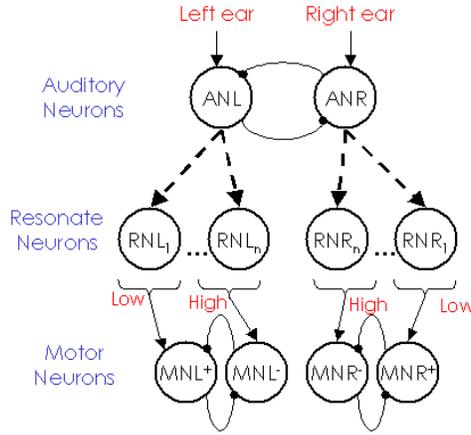


Figure 2. Scheme of the neural model implementing the phonotaxis with the multi-frequency discrimination. Each circle represents a neuron, an arrow indicates an excitatory synapse, a line with a dot is used for an inhibitory activity while a dashed arrow indicates an excitatory synapse with depression. The left (labelled L) and right (labelled R) MNs are associated to action of avoidance (MN^-) or approach (MN^+).

$$\begin{cases} \dot{V}_{Dep} = -a_1 V_{Dep} + a_2 AN_{out} \\ RN_{in} = PWL(a_3 AN_{out} - V_{Dep}) \end{cases} \quad (2)$$

where a_i are parameters (see Table 2), V_{Dep} is the depression effect and PWL is a nonlinear function that limits the synaptic current in $[0,1]$.

In the second layer the Morris-Lecar model^{17,18} is adopted to implement the resonate neurons, the equations of the model are the following:

$$\begin{cases} \dot{V} = k_f [I + g_l (V_t - V) + g_k w (V_k - V) + g_{Ca} m_\infty (V) (V_{Ca} - V)] \\ \dot{\omega} = k_f [\lambda(V) (\omega_\infty(V) - \omega)] \end{cases} \quad (3)$$

where

$$\begin{aligned} m_\infty(V) &= \frac{1}{2} \left(1 + \tanh \frac{V - V_1}{V_2} \right) \\ \omega_\infty(V) &= \frac{1}{2} \left(1 + \tanh \frac{V - V_3}{V_4} \right) \\ \lambda(V) &= \frac{1}{3} \cosh \frac{V - V_3}{2V_4} \end{aligned}$$

V and ω are the state variables of the system, I is the input and $V_i, V_t, g_l, g_k, g_{Ca}$ and k_f are parameters of the model (see Table 3).

The Morris-Lecar neuron emits spikes only if stimulated with a train of spikes showing a particular inter-spiking frequency; the neuron characteristic frequency can be tuned acting on the parameter k_f (see Table 4). In Figure 3 the system behavior is shown. The membrane potential (V) shows damped oscillations around the stable equilibrium point, when the frequency of the external stimulus corresponds to the neuron characteristic frequency, the system is forced toward a limit cycle and emits spikes.

The third layer of the control scheme defines the actions that the robot will accomplish, based on the amplitude-frequency clustering. It consists of motor-neurons, modelled as integrate-and-fire neurons.

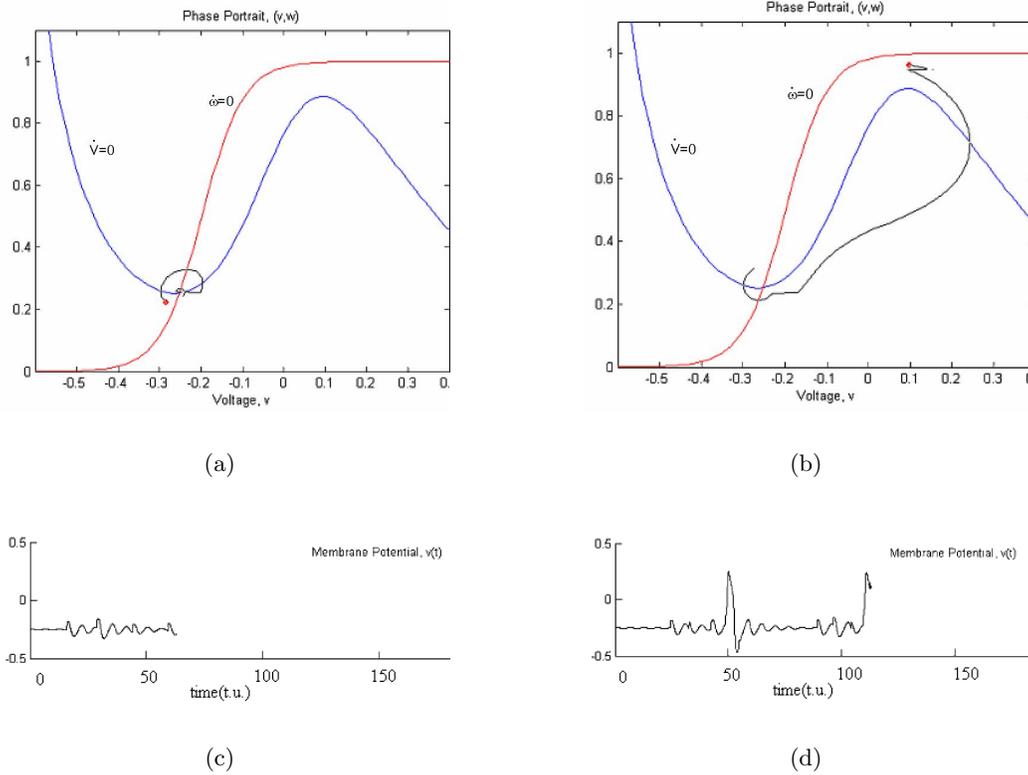


Figure 3. Behavior of the Morris-Lecar model: (a)-(b) phase portrait of the state variable V and ω , the evolution of the system is shown in black, the nullclines $\dot{V} = 0$ and $\dot{\omega} = 0$ are also shown; (c)-(d) time evolution of the membrane voltage (V). When the frequency of the stimulus is uncorrect the system shows damped oscillations (a)-(c) while when the correct stimulus is applied, the system emits spikes (b)-(d).

The system is able to recognize a wide frequency range; in the proposed scheme a low frequency signal produces an attractive behavior and activates the MN^+ , while a high frequency signal triggers an avoidance action and activates the the MN^- .

To obtain more complex behaviors each RN can be associated to a specific network of MNs that will be dedicated to accomplish a particular motor activity.

Table 1. Parameters of the IF neurons.

	mutual inhibition weight	k	I_{dis}
AN	-40	0.7	3
MN	-40	0.1	2

Table 2. Synaptic Parameters.

	a_1	a_2	a_3
AN-RN synapse	-0.08	2	3

Table 3. Parameters of the RNs.

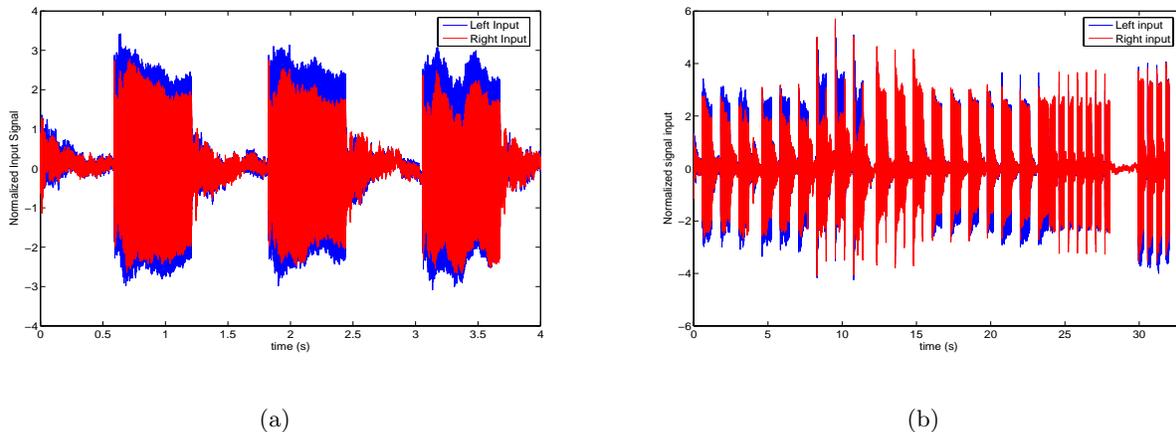
	V_1	V_2	V_3	V_4	V_t	V_k	V_{Ca}	g_{Ca}	I
RN	-0.01	0.15	-0.198	0.1	-0.5	-0.7	1	2	0.255

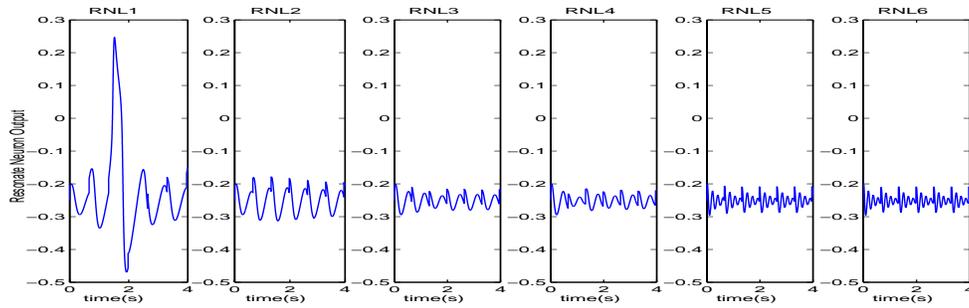
Table 4. Parameters of the Resonate-and-Fire neurons constituting the second layer of the network. Each RN is activated by a characteristic input oscillation frequency.

	k_f	Oscillation frequency
RN_1	$2.418 \cdot 10^{-4}$	0.85-0.73 Hz
RN_2	$2.818 \cdot 10^{-4}$	1.01-0.84 Hz
RN_3	$3.236 \cdot 10^{-4}$	1.14-0.96 Hz
RN_4	$3.636 \cdot 10^{-4}$	1.29-1.07 Hz
RN_5	$4.836 \cdot 10^{-4}$	1.69-1.44 Hz
RN_6	$5.654 \cdot 10^{-4}$	1.84-1.50 Hz

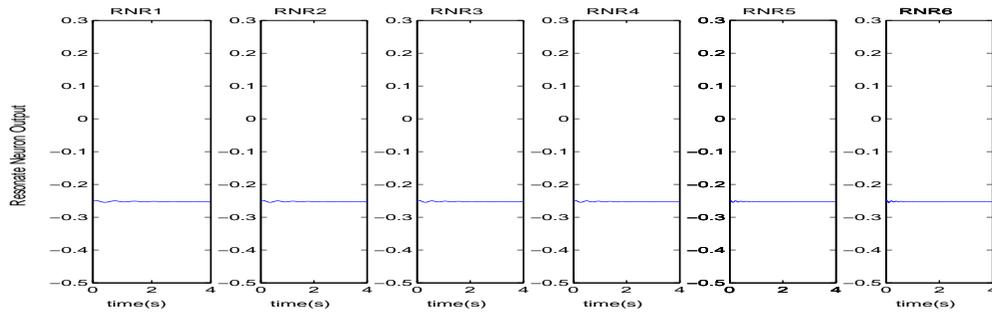
4. EXPERIMENTAL RESULTS

The model for amplitude-frequency clustering that has been realized, shows different behaviors associated with different characteristics of the input signals (amplitude and frequency). The control system was tested on a roving robot to reproduce a behavior similar to that shown by female crickets in which some auditory patterns are associated with the calling song of male crickets, while other ones indicate the presence of predators. Therefore, the network for auditory perception is devoted to control different responses (attractive or repulsive) depending on the input sound characteristics. The description of the roving robot and the results obtained with the application of the control scheme are reported in following. The robot used for the experiments is a dual drive Lego roving robot. It was equipped with two microphones connected to a PC through a sound acquisition board that samples the signals with a frequency of 44100Hz . The auditory perception network was developed in C++ code and the entire processing was performed by the PC. The sound pattern generated with an audio reproducer

**Figure 4.** (a) Sound signals acquired from the two microphones equipped on a roving robot during the first acquisition. The amplitude of the left signal is predominant respect to the right input. (b) Collection of 8 acquisitions (windows of 4 s) processed during the experiment.



(a)



(b)

Figure 5. Trends of the membrane potential of the RNs (i.e. 2^{nd} layer of the spiking network) coupled with the AN. (a) Only the neuron RNL_1 resonates at the frequency of the input signal. (b) The RNs of the right side are not stimulated by the ANR.

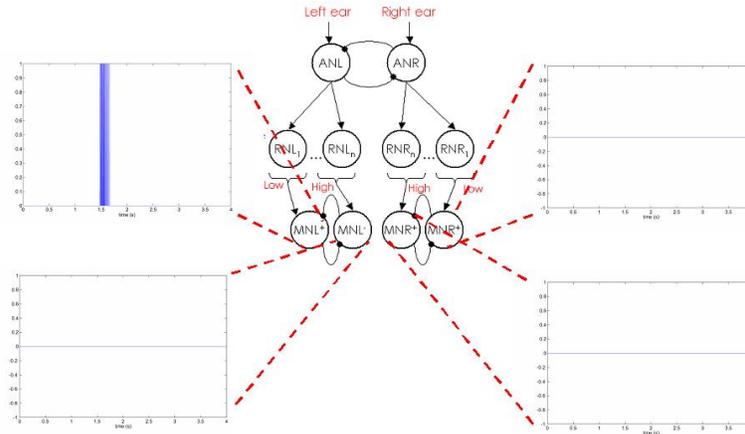


Figure 6. Robot response modulated by an external sound stimulus, a left approaching action is triggered to reach the attractive sound source.

is emitted by a speaker. The commands generated by the PC for the Lego robot are transmitted through an infrared device. In experimental tests, the robot was allowed to move in an indoor flat surface as shown in Figure 8, where a speaker is visible in the lower part of the figure. The audio stream is acquired in windows of 4s that are successively processed. The first microphone output acquisition is shown in Figure 4 (a); the complete audio stream, collection of 8 sound windows, acquired during the experiment is shown in Figure 4 (b). As it can be noticed two different sound patterns are given: low frequency (attractive signal) in the first part and high frequency (repulsive signal) in the second part of the experiment (starting from $t = 24s$). After the acquisition of the first data window, the sound signals are filtered by a bandpass filter used to eliminate the high frequency oscillations in order to process only the characteristics of the input signals relevant to the phonotaxis problem as the SRI. The auditory sensory stimuli are successively processed by the ANs, in this case the ANL wins the competition and emits spikes. Therefore, in the first layer the direction of the incoming sound is identified.

Subsequently, the AN output modulated by the depression synapses, is elaborated by the 2nd layer. The output of the RNs is given in Figure 5 (a). As can be noticed only one RN (i.e. RNL_1) is correctly stimulated by the input with a frequency of 0.81 Hz (see Table 4). All the RNs of the right side are not stimulated because the ANR has been inhibited by the winner neuron ANL (see Figure 5 (b)). The RNs labelled from 1 to 3 are associated to MN^+ and the other from 4 to 6 are associated to MN^- . In the final step the information processed by the auditory perception model is used to modulate the robot behavior. As shown in Figure 6, the amplitude and frequency of the external stimuli are associated to a left approaching response triggered to reach the sound source. When the robot after a sequence of approaching actions, reaches the target, the emitted sound is manually switched to a high frequency pattern that is associated to a dangerous source as shown in Figure 7 (a). In this case the network recognizes the new pattern; in fact a different RN is activated by the emitted sound of a frequency of 1.54 Hz to which the RNL_5 and RNL_6 are sensible (see Figure 7 (b)). The whole trajectory followed by the roving robot controlled by the auditory perception model is shown in Figure 8. The first part of the path is a sequence of approaching movements (i.e. solid arrows) towards the attractive sound source, while the second pathway (i.e. dashed arrows) indicates the escaping behavior produced by a repulsive sound source. Finally the phonotaxis control scheme can be realized in hardware for a real-time processing and the circuit could be equipped on autonomous roving robots. The sensing capabilities can be also improved including several microphones that can discriminate the location of the sound source in shortest time and with high precision.

5. CONCLUSIONS

In this work the problem of recognition and localization of a sound source is discussed. The starting point is the solution adopted by the cricket that is able to identify an auditory stimulus and to reach the sound source. The auditory process has been modelled with a network of neurons with the aim to realize a bio-inspired

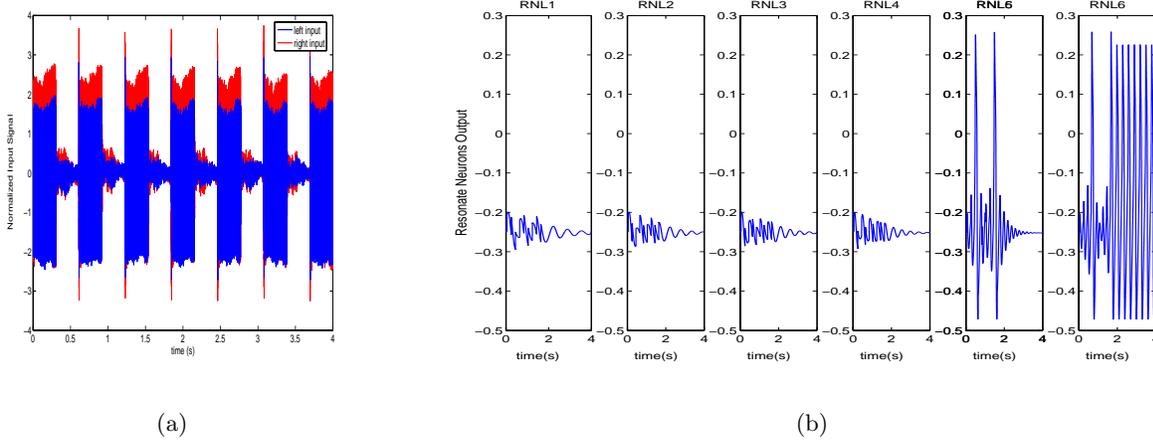


Figure 7. (a) Sound inputs acquired when the the frequency of the sound pattern emitted by the speaker has been increased. (b) The high frequency source stimulates the RNL_5 and RNL_6 that are associated to a left avoidance action.

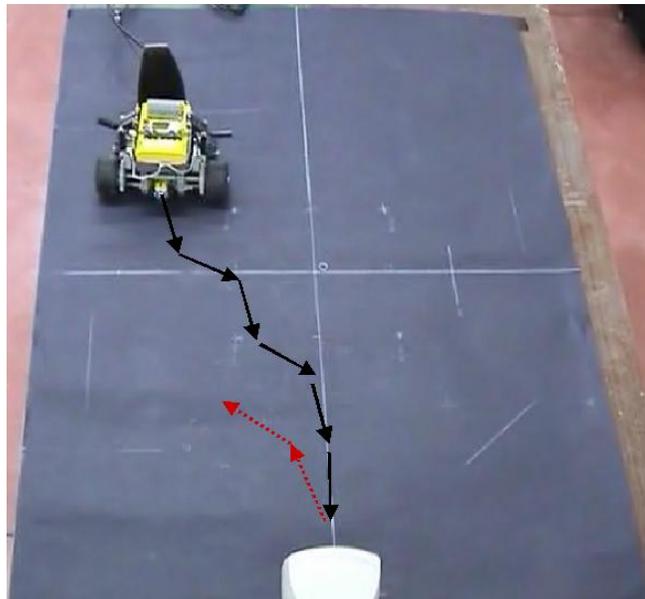


Figure 8. Trajectories followed by the robot when a sound source is active. The first path, solid arrows, is generated by an attractive source (i.e. low SRI); when the sound source is switched to a dangerous pattern (i.e. high SRI), the robot escape away (dashed arrows).

auditory system for robotic applications. The proposed model consists of three functional blocks. The first layer is constituted by a pair of IF neurons mutually coupled in order to realize a winner-take-all competition, the winning neuron is associated to the direction of the source. A second processing stage is executed in the second layer where an array of resonate-and-fire neurons is used to discriminate different frequencies. The third layer is constituted by motoneurons used to produce actions. The control system has been tested on a Lego roving robot with good results. The robot has been able to locate the sound source and to identify the auditory pattern in order to accomplish the correct action (attractive or escape behavior). The experiments, presented in this work, have been accomplished with a software implementation of the perception model. A further development will be an hardware realization of the control system. The spiking network could be implemented in an analog circuit with operational amplifier or the dynamic differential equations constituting the whole system could be written in discrete time for a digital implementation through micro-controllers.

ACKNOWLEDGMENTS

The authors acknowledge the support of the European Commission under the project FP6-2003-IST2-004690 SPARK “Spatial-temporal patterns for action-oriented perception in roving robots”.

REFERENCES

1. M. Konishi, *Neuroethology and Behavioral Physiology*, springer ed. F. Huber and H. Markl, 1983, ch. Neuroethology of acoustic prey localization in the barn owl.
2. V. Walzer, H. Peremans, and J. Hallam, “One tone, two ears, three dimensions: A robotic investigation of pinnae movements used by rhinolophid and hipposiderid bats,” *J. Acoust Soc*, 2004.
3. R. Payne, K. Roeder, and J. Wallmann, “Directional sensitivity of the ears of noctuid moths,” *J. Exp Biol*, vol. 44, pp. 17–31, 1966.
4. Y. Cohen and G. Gifford, “Neural mechanisms of sound localization,” in *Proceedings from the 8th Annual Research Symposium of the SHHH*, 2001.
5. K. Schildberg, “Temporal selectivity of identified auditory interneurons in the cricket brain,” *J Comp Physiol*, vol. 155, pp. 171–185, 1984.
6. K. Schildberg, F. Huber, *Cricket behavior and neuroobiology*, Cornell university press ed. Huber,F., Moore,T.E. and Loher W., 1989, ch. Central auditory pathway:neural correlates of phonotaxis behavior.
7. B. Webb and T. Scutt, “A simple latency dependent spiking neuron model of cricket phonotaxis,” *Biological Cybernetics*, vol. 82(3), pp. 247–269, 2000.
8. R. Reeve and B. Webb, “New neural circuits for robot phonotaxis,” *Philosophical Transactions of the Royal Society A*, vol. 361, pp. 2245–2266, 2002.
9. B. Webb and R. Harrison, “Integrating sensorimotor systems in a robot model of cricket behaviour,” in *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, vol.4196,2000, pp.113–124.
10. A. Horchler, R. Reeve, B. Webb, and R. Quinn, “Robot phonotaxis in the wild: A biologically inspired approach to outdoor sound localization,” in *Proceedings of ICAR*, 2003.
11. B. Webb, R. Reeve, Horchler, and R. Quinn, “Testing a model of cricket phonotaxis on an outdoor robot platform,” in *Proceedings of TIMR03*, 2003.
12. K. Imaizumi and G. S. Pollack, “Neural coding of sound frequency by cricket auditory receptors,” *The Journal of Neuroscience*, vol. 19(4), pp. 1508–1516, February 1999.
13. Braitenberg V., 1984, *Vehicles: Experiments in synthetic psychology*, MIT Press.
14. C. Koch, *Biophysics of Computation, Information Processing in Single Neurons*. Oxford University Press, 1999.
15. E. Izhikevich, “Neural excitability, spiking, and bursting,” *International Journal of Bifurcation and Chaos*, vol. 10, pp. 1171–1266, 2000.
16. E. Izhikevich, “Resonate-and-fire neurons,” *Neural Networks*, vol. 14, pp. 883–894, 2001.
17. C. Morris and H. Lecar, “Voltage oscillations in the barnacle giant muscle fiber,” *Biophysical Journal*, vol. 35, pp. 193–96213, 1981.
18. F. Hoppensteadt and E. Izhikevich, *Brain Theory and Neural Networks*, MIT press, cambridge ed. Arbib MA, 2002, vol. 181–186.