Sound Localization for a Humanoid Robot by Means of Genetic Programming

Rikard Karlsson, Peter Nordin and Mats Nordahl
Complex Systems Group, Chalmers University of Technology, S-412 96, Göteborg, Sweden, Email: nordin,tfemn@fy.chalmers.se

Abstract. A linear GP system has been used to solve the problem of sound localization for an autonomous humanoid robot, with two microphones as ears. To determine the angle to the sound source, an evolved program was used in a loop over a stereo sample stream, where the genetic program gets the latest sample pair plus feedback from the previous iteration as input. The precision of the evolved programs was dependent on the experimental setup. For a sawtooth wave from a fixed distance the smallest error was $8^\circ$. When letting the distance to the same source vary the error was $23^\circ$. For a human voice at varying distances the error was up to $41^\circ$.

1 Introduction

The purpose of this paper is to investigate sound localization using GP. The system is intended to be used in a humanoid robot equipped with two microphones. Sound localization is performed by a small binary machine code program on an unprocessed stereo stream of sampled sound. Since the system is intended for a humanoid robot, some degree of similarity to human sound localization is desired (see [1] for another study of this problem in a robot context). Because of the limited CPU power on-board the robot the computational requirements need to be minimized.

The GP system needs to generalize from a limited set of training data. The system also needs to be able to localize many different kinds of sounds, sounds from all directions in the horizontal plane, sounds with different intensity, sounds from all distances, sounds when echoes are present, and sounds with background noise present.

As a comparison, we first give a short description of the human auditory system and its sound localization capabilities.

1.1 The auditory system

The ability to localize sounds is an important part of the auditory system and has been essential to our survival. Sound propagates through a medium as a longitudinal wave. On its way to the eardrum it passes the outer ear, called the pinna, and the auditory canal. The longitudinal wave in the air is transferred
to the eardrum, or tympanic membrane. The signal is amplified on its way by the ossicles, which work as levers from the tympanic membrane to the smaller area of the oval window. The movement of the oval window is transferred into a longitudinal wave in the liquid-filled cochlea.

The longitudinal wave propagates down the long stretched basilar membrane in the cochlea. The membrane is narrower and thinner, and thus more sensitive to higher frequencies, near the oval window and responds to lower frequencies further down the membrane. Between 35 000 and 50 000 neurons are distributed along the basilar membrane. Each has a characteristic frequency related to the frequency causing maximal displacement of the basilar membrane at its location. This organization, where neighbouring neurons have similar response, is called tonotopy. It has a frequency range from 200 Hz to 20 000 Hz. Frequencies lower than 200 Hz are coded in terms of time of neural firing, by phase locking between the sound wave and the neuron firing. Between 200 Hz and 4 000 Hz both phase locking and tonotopy are used, while only tonotopy is used above 4 000 Hz.

1.2 Human sound localization

Let us now investigate the cues for sound localization present right at the eardrums. Consider a dummy head in an anechoic (echo-free) chamber with one sound source. The sound wave from the sound source propagates through the air with a velocity of 340 m/s and reaches the dummy head and its pinnae, auditory canals and eardrums. If the dummy head is facing the sound source at an angle it will take the sound longer to reach one ear than the other. This difference in time of arrival is known as the interaural time delay (ITD). If the sound is continuous the ITD can instead be determined by comparing phases of the sound signals at the eardrums.

If we know the relative phases and the wavelengths of the waves, we can only determine uniquely which wave lags the other if we know that one wave is always delayed by less than half a wavelength with respect to the other. This puts a limit on the frequency. When the sound comes from the left or right the delay is maximal. The distance between the ears is approximately 20 cm, which means that the wavelength needs to be longer than 40 cm. This corresponds to frequencies smaller than 850 Hz, for air with a sound velocity of 340 m/s.

Another cue that can be used to determine the direction to a sound source is the interaural intensity difference (IID), since the head casts a sound shadow and reduces the amplitude of a passing sound wave. Sound waves longer than the width of the head are strongly diffracted, and frequencies larger than 1700 Hz are strongly shaded.

The pinnae also filter the acoustic spectrum in a direction-dependent way. This filter is called head-related transfer function (HRTF). It can be seen as a function that maps intensity to intensity and takes the frequency of the sound and the direction to the sound source as arguments. The HRTF provides another potential cue for sound localization.

For localization in the horizontal plane ITD and IID are major cues for humans, while the HRTF is a cue for localization in the vertical planes. In