Departamento de Ingeniería Electrónica y Comunicaciones
Universidad Zaragoza

IEEE SAM 2018 SHEFFIELD (UK)

Direction of Arrival Estimation with Microphone Arrays Using SRP-PHAT and Neural Networks

David Díaz-Guerra Aparicio, José Ramón Beltrán Blázquez

Department of Electronic Engineering and Communications

University of Zaragoza

C/ María de Luna 1 - 50018, Zaragoza, Spain.

{646988, jrblebla}@unizar.es

Abstract: The Steered Response Power with phase transform (SRP-PHAT) is one of the most employed techniques for Direction of Arrival (DOA) estimation with microphone arrays due its robustness against acoustical conditions as reverberation or noise. Among its main drawbacks is the growth of its computational complexity when the search space increases. To solve this issue, we propose the use of Neural Networks (NN) to obtain the DOA as a regression problem from a low resolution SRP-PHAT power map. The NNs can learn and exploit the information of the acoustic reflections of the room where the array is located with a training method that can be easily performed by an end user without technical knowledge.

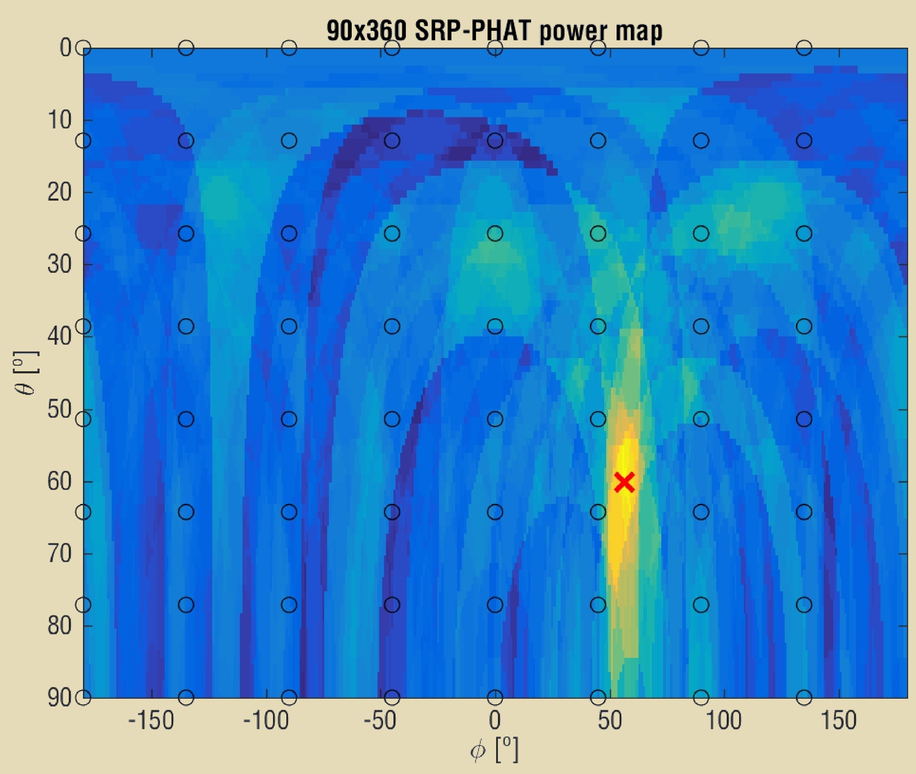
The SRP-PHAT algorithm

The Steered Response Power (SRP) of a sensor array is defined as the power of the output of an array steered to the desired direction using a delay-and-sum beamformer. It can be written in terms of the Cross-Correlation functions between sensors as:

$$P(\boldsymbol{\theta}) = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} R_{nm}(\Delta\tau_{nm}(\boldsymbol{\theta}))$$

where N is the number of sensors, $\boldsymbol{\theta}$ is the desired direction, R_{nm} is the Cross-Correlation Function between the sensors n and m, and $\Delta\tau_{nm}(\boldsymbol{\theta})$ is the time difference of arrival between the same sensors. Typically, R_{nm} is computed as the Generalized Cross-Correlation using the PHAT transform, which leads to the SRP-PHAT algorithm.

To perform a DOA estimation using the SRP-PHAT algorithm we must compute $P(\boldsymbol{\theta})$ for a fine enough grid of directions $\boldsymbol{\theta}$. The figure shows a high-resolution SRP-PHAT power map whose maximum would be lost by an 8x8 power map.

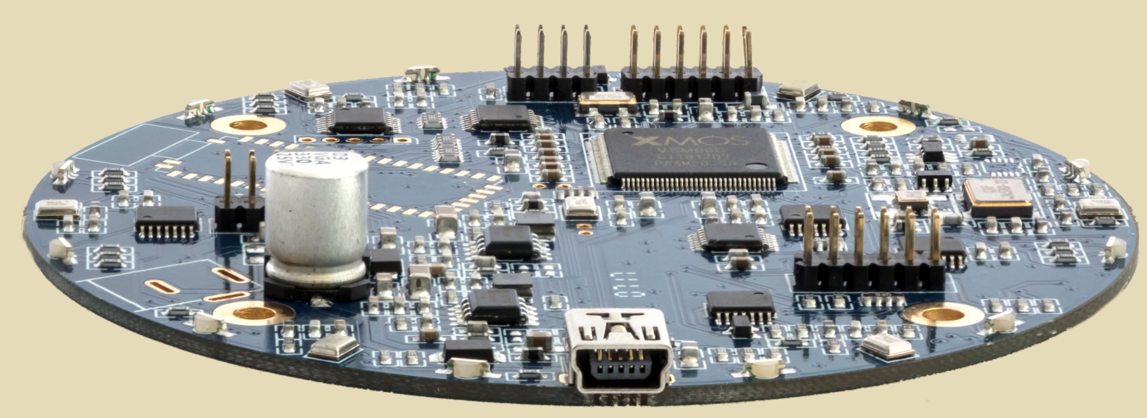


Proposed DOA estimator

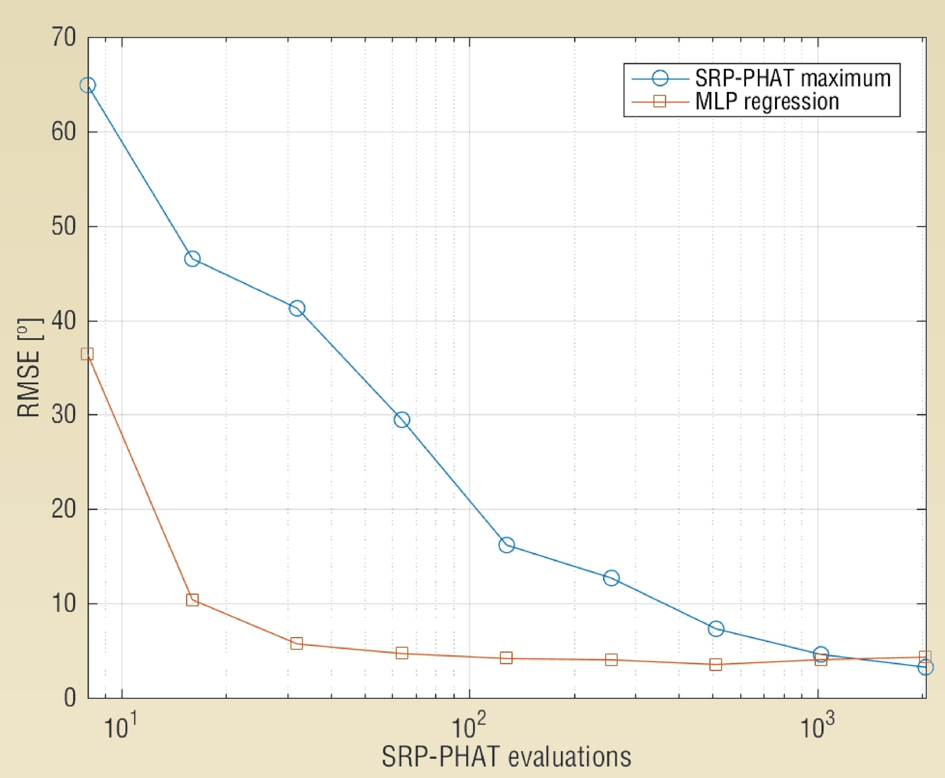
In order to reduce the computational complexity of the SRP-PHAT algorithm we propose to reduce the number of directions where $P(\boldsymbol{\theta})$ is computed and infer the DOA estimation using a neural network instead of looking for the maxima of the SRP-PHAT function. Specifically, we use a MLP with two hidden layers with 128 perceptrons. As we have focused in 2D DOA estimation with a circular microphone array, our output layer has only 2 perceptrons that represent the DOA estimation in spherical coordinates.

To create the dataset to train the network, we placed the array, a miniDSP UMA-8 with 6 MEMS microphones equispaced in a circumference of approximately 90 mm in diameter, in the center of a conference table. With this configuration, we perform a recording of 5 minutes at 44.1 kHz emitting white noise through the speaker of a smartphone while walking around the table at different heights.

We divide the recording in frames of 1024 samples with an overlap of 512 samples and apply the SRP-PHAT algorithm to get 2 power maps, the first with high resolution (90x360) and the second with a lower one. We use the former to obtain the position of the sound source and the last as input of the network. Finally, we randomly permuted the frames and took 18,000 for training and 6,000 for test.



Results



The figure on the left shows the angular Root Mean Square Error (RMSE) in the test dataset for different power map resolutions. It can be seen that the DOA inferred by the MLP from a power map of resolution 8x8 (64 SRP-PHAT evaluations) has an error very similar to taking the maximum of a 32x32 power map (1024 SRP-PHAT evaluations), but using 16 times fewer evaluations of the SRP-PHAT functional. The figure on the right shows the 8x8 power map corresponding to figure of the first section. Despite the real DOA is not captured, the MLP is able to estimate the DOA estimation with an error of only 1.7º.