BEARING ESTIMATION WITH TIME-DELAY NEURAL NETWORKS

Brigitte Colnet and Jean-Claude Di Martino

CRIN-INRIA CNRS
Nancy. FRANCE
e-mail : bcolnet@loria.fr , dimartin@loria.fr

ABSTRACT

In this paper we present a neuromimetic approach to bearing estimation issue. The proposed method is based on time-delay neural networks. This kind of network is well suited to take into account constraints encountered in signal processing: it deals with the dynamic nature of signal and discovers acoustic and temporal features. According to the propagation model of plane waves, the network has to relate the delays between sensors to enable source localisation. The time-delay neural network approach is encompassed in a successive-refinement method. Thus, accuracy is increased while the number of networks to look at the whole horizon is reduced.

1. INTRODUCTION

We propose a neural network approach to the bearing estimation issue. To detect and locate far field sources, important information is provided by the time propagation delays between array sensors. As shown Waibel in [1], time-delay neural networks (TDNN) deal with dynamic nature of signal and take into account temporal relationships among acoustic events. Constraints tied to the propagation model and encountered in our application are the same as those encountered by authors in [1] and are well supported by the TDNN approach. Indeed, we have to represent relationships between spacio-temporal events. Moreover, these spacio-temporal relationships are time-invariant.

This work was supported by the CTSN (Underwater Study and Search Centre) DGA Toulon under grant 924863009 CNRS/DCAN/DRET

The localisation method presented here lies on a successive-refinement process based on time-delay neural networks. The performance of this method on real signals confirms encouraging results obtained in simulation tests.

2. PROPAGATION MODEL

We assume that waves emitted by a far field source are plane when they impinge a linear array of equally spaced sensors. The knowledge of the array geometry and of the propagation delays allows us to locate the source. Indeed, let θ be the direction of arrival of signal, the delay between two adjacent sensors is expressed by

$$\tau = \frac{d \sin \theta}{c} \tag{1}$$

where d is the distance separating the two sensors and c is the celerity in the propagation homogeneous medium. In equation (1) τ does not depend on time. This relation is time-invariant.

3. PRINCIPLE OF THE METHOD

We implement a two-level process to deal with the source localisation issue. First, we take advantage of time-delay neural network that has been studied to phoneme recognition in [1]. Some constraints in speech recognition are almost the same as those encountered in signal processing to bearing estimation:

- neural network has to take into account the dynamic nature of signal: data to process represent the temporal evolution of the signal received on the array sensors,

0-7803-2431-5/95 $4.00

- it has to represent temporal relationships between acoustic events. In the case of plane waves impinging on a linear array of equally spaced sensors, the time when one wave impinges on a given sensor has to be related to the time it reaches other sensors,
- invariance under translation in time has to be provided because spacio-temporal relationships are time-invariant (see equation (1)): wave propagation delays do not change over time because the array geometry is supposed to be fixed while the emitting source does not move.

We build a set of time-delay neural networks that focus on some parts of the whole horizon (from -90 to $+90$ degrees). A TDNN attached to an angular sector indicates in which part of this angular sector the source is located, if any.

Second, the neural networks are used in a successive-refinement approach.

4. TDNN ARCHITECTURE

Consider a TDNN focused on angular sector $[\theta, \theta + \Delta\theta]$ divided in n equal parts of size $\frac{\Delta\theta}{n}$ degrees. A four-layer network is built:

4.1. Input layer

The input layer is supplied with the sampled signal received by the array sensors. The basic unit of a TDNN is a delayed unit. Each sensor is associated to one delayed unit. The number of delays introduced for one unit is chosen to be equal or greater to the maximum delay between the two end sensors. Because coefficients are the samples of signal on sensors at a given time, introducing delays is like considering signal through a temporal observation window sliding over data. Input data are normalised to lie in the interval $[-1, +1]$.

4.2. Hidden layers

The first hidden layer has $nh1$ units expanded out temporally. The second hidden layer is made up of $nh2$ delayed units. The number of delays for the hidden units depends on the number of input delays and on the size of the window frame sliding on the previous layer.

4.3. Output layer

It is obtained by integrating over time the $nh2$ hidden units. The number of output units is equal to $nh2$ and $nh2 = n$, the number of parts we divide the angular sector focused by the TDNN.

4.4. Connections

The originality of TDNN lies in delayed units and their feed-forward connections with shared weights. A frame window scans the input layer. Each unit in this window is connected to one frame of $nh1$ units in the first hidden layer. The frame window size is chosen to integrate low level temporal events of signal arriving on sensors: it depends on the propagation delay between two adjacent sensors. In the second hidden layer, each frame of $nh2$ units receives activation from a larger sliding window on the first hidden layer (see figure 1). This allows the network to process temporal events on a larger scale. Additionally, since we need to link temporal events whatever time they occur, connection weights between units corresponding to the same time shift are forced to have the same value. They are updated with the average of all corresponding time-delayed weight changes computed using the standard back-propagation algorithm [2].

4.5. Training process

The TDNN dedicated to angular sector $[\theta, \theta + \Delta\theta]$ is trained to activate the one of its n output units that corresponds to the angular sector from which signal comes. It is supplied only with signal coming from $[\theta, \theta + \Delta\theta]$. Training is performed with simulated data. For sensor k at time t, the equation of signal is:

$$s_k(t) = \sum_{i=1}^{M} a_i \sin(2\pi f_i t) + \eta_k(t) \qquad (2)$$

where

a_i, f_i are respectively the amplitude and the frequency of the i^{th} component of signal with $f_i \in [f_{min}, f_{max}]$,

M is the number of components,

$\eta_k(t)$ is white noise.

Recognition process is achieved with real data.

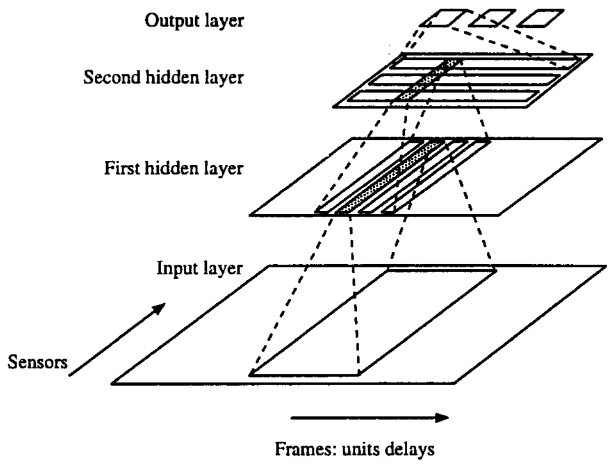

Figure 1: TDNN architecture for bearing estimation

5. SUCCESSIVE-REFINEMENT APPROACH

The originality of this work is the way neural networks are encompassed in a successive-refinement approach (fig. 2). First, a TDNN Γ_1 looks at the whole horizon of 180 degrees. Γ_1 has n_1 output units that allow it to locate a source with a coarse accuracy of $\frac{180}{n_1}$ degrees. If a source is detected in angular sector Sa_i, i ranging from 1 to n_1 (output number i of Γ_1 is activated), we use a more specialised network Γ_{2i}. The latter has been trained to locate more accurately a source being in Sa_i. It is a n_2-output-unit network. This reduces the search space to an angular sector of $\frac{180}{n_1 \times n_2}$ degrees. Moreover, a third level allows us to obtain the desired accuracy.

Experiments have shown that three refinement levels are enough to obtain accuracy within the limit of the 3-dB beamwidth [4]. The total number of networks used is $1 + n_1 + n_1 \times n_2$. An interesting feature is that all of these networks have the same architecture; they only differ in the number of output units.

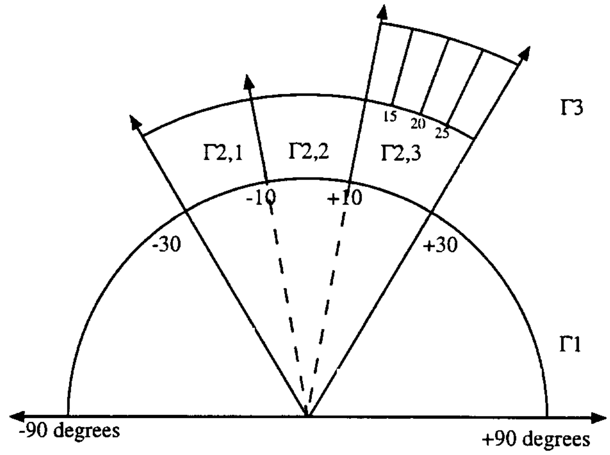

Figure 2: Refinement model

6. EXPERIMENTS AND RESULTS

6.1. Experiment framework

Array features
8 sensors far apart from 0.1 meter,
array length = 0.7 m,
sound celerity in air = 340 m/s,
sampling frequency = 20000 Hz,

Source features
The frequency of the emitting sources ranges from 500 to 2000 Hz.

6.2. Experiment results

In this experiment, we trained 13 neural networks with simulated data according to equation (2). The first network has $n_1 = 3$ output units. It separates the whole horizon in three 60-degree angular sectors. Then, in the second step of refinement process, three networks, each one having $n_2 = 3$ output units, refine the search to 20-degree angular sectors. Nine other four-output networks allow us to locate the emitting source with a 5-degree accuracy. Then, we trained only 13 networks while 37 multi-layer perceptrons were used to get the

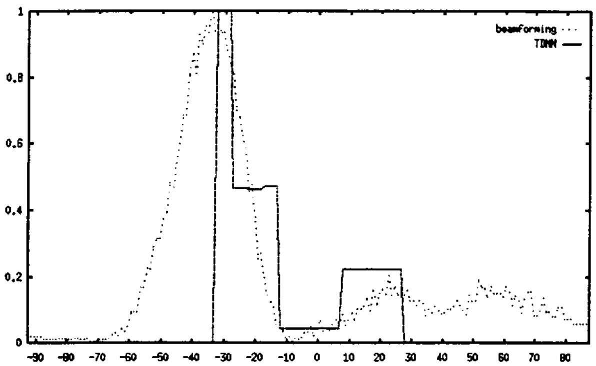

Figure 3: Results obtained on real signal. *Direction of arrival of signal is -34 degrees. The frequency emission is 1000 Hz.*

same accuracy in [3]. Results obtained on real data with this method are compared to beamforming in figure 3.

7. CONCLUSION

This method takes advantage of the TDNN architecture :

- A TDNN has the ability to handle the dynamic nature of signal. Signal parameters such as frequency, phase and amplitude do not matter. Thus, combinatorial explosion due to large class parameters is avoided [5].

- Networks are supplied with raw temporal data: there is no pre-processing step.

- They are robust to noise: good results are obtained with noisy real data.

The successive-refinement approach allows us to lower the number of networks to train without reducing localisation accuracy. By the way, training networks can run in parallel.

In further work we will study how the network encodes the relationships it discovered between acoustic events. Another point is the adaptation of the architecture of TDNN. Indeed, it seems promising to orientate the scanning window according to the direction of arrival of signal.

8. REFERENCES

[1] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, K.J. Lang *Phoneme recognition using time-delay neural networks*, IEEE transactions on acoustics, speech and signal processing, vol. 37, no. 3, pp. 328-339, March 1989.

[2] D.E Rumelhart, J.L McClelland, *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1 MIT press 1988.

[3] B. Colnet, J.P Haton, *Far field array processing with neural networks*, Proceedings of ICASSP 94 vol. 2,pp 281-284.

[4] W.S Burdic, *Underwater acoustic system analysis*, Prentice-Hall, 1984.

[5] R. Rastogi, P.K Gutpa, R. Kumaresan, *Array signal processing with interconnected neuron-like elements*, Proceedings of ICASSP 87, pp 2328-2331.