NEURAL NETWORKS FOR HIGH PERFORMANCE TIME-DELAY ESTIMATION AND ACOUSTIC SOURCE LOCALIZATION

Ludwig Houégnigan[1], Pooyan Safari[2], Climent Nadeu[2], Mike van der Schaar[1], Marta Solé[1], Michel André[1]

[1]Laboratory of Applied Bioacoustics (LAB), Polytechnic University of Catalonia, UPC Barcelona Tech, Spain
[2]TALP Research Center - Dept. TSC, Polytechnic University of Catalonia, UPC Barcelona Tech, Spain

ABSTRACT

Time-delay estimation is an essential building block of many signal processing applications. This paper follows up on earlier work for acoustic source localization and time delay estimation using pattern recognition techniques in the adverse environment such as reverberant rooms or underwater; it presents unprecedented high performance results obtained with supervised training of neural networks which challenge the state of the art and compares its performance to that of well-known methods such as the Generalized Cross-Correlation or Adaptive Eigenvalue Decomposition.

KEYWORDS

Time-delay estimation, neural networks, source localization, underwater acoustics, room acoustics

1. INTRODUCTION

Time-delay estimation (TDE) is a task as fundamental as spectral estimation and a key step for many popular applications such as sonar and radar direction finding, seismology, biomedicine, satellite navigation or acoustic source localization.

Recent advances in machine learning invite us to revisit classical signal processing problems and theory to renew our understanding of these problems so as to challenge well-established techniques. In that frame, over the past years, the authors of this paper, through different publications [1-3] have proposed original approaches using machine learning and data-specific modelling in order to improve TDE in the context of both air and underwater acoustic source localization (biological sources such as cetaceans, or artificial ones such as pingers, ships, navy sonar, etc).An approach using neural networks is justified by at least 3 fundamental assumptions:

David C. Wyld et al. (Eds) : CSITA, ISPR, ARIN, DMAP, CCSIT, AISC, SIPP, PDCTA, SOEN - 2017
pp. 137– 146, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.70114

Assumption (1): the most widely used methods for TDE have demonstrated their optimality [4-6] in the case of random signals and have been statistically analyzed according to that particular context, e.g. cross-correlation was demonstrated to be optimal for random data at high Signal-to-noise ratio (SNR). However, in many situations, SNR may be relatively low and the signals at stake can be far from random. Rather, they display statistical structure which can be exploited in order to achieve greatly improved results for all kinds of estimation tasks. Machine learning tools- and in particular supervised learning algorithms such as artificial neural networks- offer us a great opportunity to develop robust time-delay estimators that improve largelyon the classical estimators.

Assumption (2): Classical methods for TDE suppose models (see section 2)which typically fail to fully render the complexity of propagation in underwater contexts or in reverberant rooms in air. Using supervised learning is expected to permit to match more closely the available data to its environment.

Assumption (3): In tasks such as localization or angle estimation, much effort is done on tracking solutions such as Extended and Unscented Kalman Filters, or particle filters. Yet, putting more emphasis on reducing a priori the mean error and variance of TDE, by yielding more accurate and consistent estimates, will clearly facilitate the task of any subsequent tracking algorithm.

On the one hand, comprehensive studies on TDE [7-8]were published but none of them, to our knowledge, has ever included supervised learning. On the other hand, little has yet been published on time-delay estimation using supervised learning besides benchmark papers by Shaltaf et al. [9, 10]. With respect to the previously mentioned works, this article intends to progress by addressing the following points:

(1) Including large time delays: hence avoiding to restrict estimation to a narrow range of time-delays which fall within the Nyquist range and could in fact already be addressed by beamforming techniques, a shortcoming in [9,10].

(2) Instead of estimating only a nominal time-delay value (time position of a peak), the neural network was tasked to provide a multidimensional output representing an ideal time-delay response, (cf. section 2 and fig. 3 and 4).

(3) The number of data samples at stake is large, i.e. 8 datasets containing each 400 000 samples. Each of them was evaluated by multiple methods, in order to provide a robust and comparative statistical analysis of the various time-delay estimators at hand under different levels of noise.

2. MODELS AND METHODS FOR TIME-DELAY ESTIMATION

2.1. Ideal free-field model

Methods such as standard cross-correlation and generalized cross-correlation [6, 8, 11] or minimum entropy [12] are based on this model. It proposes to view two signals x_1 and x_2, acquired at two spatially separated sensors, as attenuated and delayed versions of a source signal plagued with additive noise. This model is well-described in [7] and can be described with the following equations:

$$x_1(n) = \alpha_1 s(n - \tau_1) + b_1(n) \text{ (eq.1)}$$

$$x_2(n) = \alpha_2 s(n - \tau_2) + b_2(n) \text{ (eq.2)}$$

where s is the source signal, α_1 and α_2 are attenuation factors due to propagation and b_1 and b_2 represent uncorrelated additive noise .

In this model the sample time-delay τ_{12}between signals x_1 *and* x_2 can be set as:

$$\tau_{12} = \tau_1 - \tau_2 \text{ (eq.3)}$$

2.2 Real reverberant model

The reverberant model entails a higher level of complexity as it assumes that the original signal is deteriorated by multipaths and reverberation (walls, floors, tables in room acoustics, surface, seafloor, or scattering in underwater acoustics). Hence, signals x_1 and x_2 are modelled as convolutive mixtures of the source signal:

$$x_1(n) = h_1 * s(n) + w_1(n) \text{ (eq.4)}$$

$$x_2(n) = h_2 * s(n) + w_2(n) \text{ (eq.5)}$$

where h_1 and h_2 aim to model, with FIR or IIR filters, the channel impulse responses from the source to the positions of sensors 1 and 2, and where "*" indicates convolution and the noises $w_1$and w_2 can be correlated. Such a model can provide a better description of the propagation environment and is typically based on adaptive algorithms such as the Adaptive Eigenvalue Decomposition (AED) [13, 14, 15].

2.3. Supervised estimation and training with neural networks

In this approach, it is assumed that the time-delay can be approximated by the output of a previously trained neural network which receives as input a combination (or a transformation) of signals x_1 and x_2. No particular assumption is made with regard to modelling. The capabilities of neural networks for system identification and interpolation permit to construct a system that minimizes the error between its output and an ideal response represented by a peak at the localization of the correct time-delay.

In this study, the input vector was set to $x = [x_1 x_2]$, a concatenation of vectors x_1 and x_2, and no particular effort was made to construct a transformed or more compact input. In [9, 10] it is also proposed to use the sum of the received signals or a transformation of those signals such as the Discrete Cosine Transform (DCT) to provide a more compact representation and hence to optimize the dimension of the neural architecture. It is proposed here to provide as target the dirac delta function$\delta(n - \tau_{12})$. The neural network then aims, through supervised training, at minimizing the distance between its output and an ideal response composed of a value 1 at the correct time-delay and zeros elsewhere.

3. IMPLEMENTATION AND TEST

3.1 Signal and noise models

Eight artificial datasets containing chirp signals were constructed. Each signal featured random duration (with 10 to hundreds of samples at a sampling rate of 16 kHz, their number being randomly selected from uniform distributions) and varying noise. The variance $\sigma_S{}^2$ of the signal of interest is related to the noise variance of each dataset as described in table 1.1 and 1.2. The noisy dataset aims at mimicking adverse conditions which typically cause failures in time-delay estimators as will be shown in section 4.

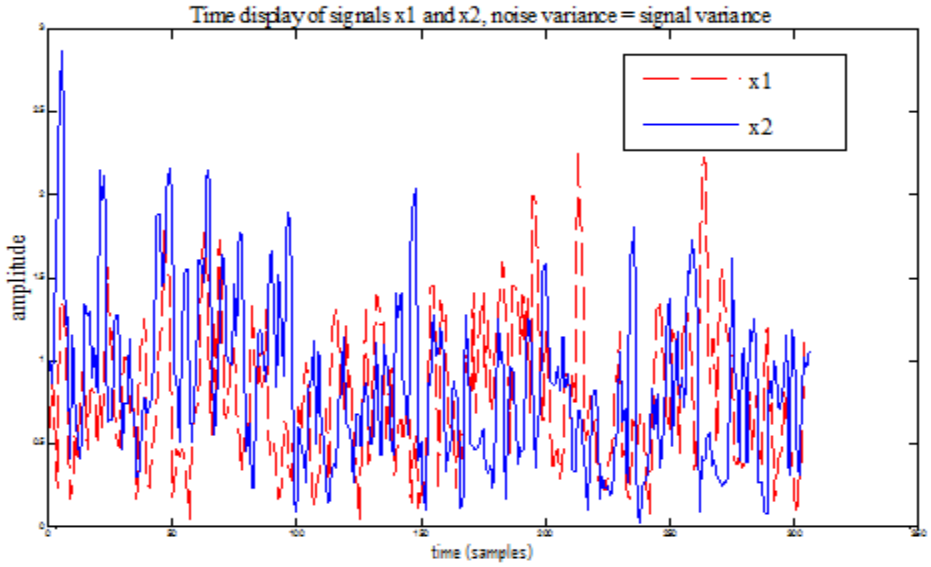

Fig.1 signals x1 and x2 are composed of noisy chirps signals of variable duration

Dataset#	1	2	3	4
Noise var. $\sigma_N{}^2$	0	$0.2\sigma_S{}^2$	$0.4\sigma_S{}^2$	$0.5\sigma_S{}^2$

Table 1.1

Dataset#	5	6	7	8
Noise var. $\sigma_N{}^2$	$0.6\sigma_S{}^2$	$0.8\sigma_S{}^2$	$\sigma_S{}^2$	varying

Table 1.2

Dataset 8 is made of 400000 signals corrupted with additive noise which variance is uniformly taken between $0.2\sigma^2$ and σ^2.

3.2. Neural network parameters

Multilayer perceptrons (“mlps”) architectures including a single hidden layer and 30 hidden units were used. Sigmoid and linear activation functions were respectively used for the hidden and output units. The training procedure was conducted using a standard backpropagation algorithm with a fixed mini-batch size of 100 and 100 epochs. The fixed momentum and weight decay for all the systems were respectively set to 0.9 and 10^{-7}. A sparsity target of 0.05 and a sparsity penalty of 10^{-4} were used for all the networks. L2 regularization norm was set to 10^{-3}.

For each dataset, 19 neural nets were trained with varying learning rates. Among those 19 nets, for the sake of concision, only 4 were selected and are displayed here, namely the nets providing respectively the best (lowest) mean error, the worst (highest) mean error, the best (lowest) variance and the worst (highest) variance, referred to respectively as MLPA, MLPB, MLPC, MLPD. Those nets are then compared to 4 other estimators: the generalized cross-correlation with SCOT and PHAT filters (GCC-SCOT and GCC-PHAT), standard unbiased cross-correlation (XCOR) and the Adaptive Eigenvalue Decomposition (AED). In that sense, the different models already presented in section 2. are all tested here.

For each method, this work is not limited solely to the nominal time-delay, as is the case in most publications, but it also analyzes the delay distribution as indicated by the following two points:

(1) For a nominal time-delay$\hat{\tau}$: the error $\hat{\tau} - \tau_{12}$ between the estimated nominal time-delay and the ideal time delay should be minimized.

(2) Similarity of the output of the neural net with the ideal target $\delta(n - \tau_{12})$ must be evaluated. To that purpose, a similarity measure named $Q_{KL}$was derived from the Kullback-Leibler divergence to assess the resemblance of the output $\mathrm{NN}_{12}$of the neural network to the ideal response, both seen as data distributions:

$$\mathrm{Q_{KL}} = \mathrm{abs}(\mathrm{KL}\big(\mathrm{NN}_{12}, \delta(\,\mathrm{n} - \tau_{12})\big), \text{(eq.6)}$$

where $\mathrm{KL(P, R)}$ is a modified expression of the Kullback-Leibler divergence between two distributions P and R:

$$\mathrm{KL(P, R)} = \quad -\sum_{\mathrm{x}} \mathrm{p(x)} \log(\mathrm{r(x)} + 1) + \sum_{\mathrm{x}} \mathrm{p(x)} \log(\mathrm{p(x)} + 1)\ , \text{(eq.7)}$$

The metric Q_{KL} resembles a distance, inasmuch as it is positive and approaches 0 when the output of the neural net resembles its target. Yet, $\mathrm{Q_{KL}}\ is$ not a true metric since it is not symmetric and does not verify the triangle inequality. It was however found to be a convenient, consistent and compact measurement compared to others such as χ^2,Euclidian or Bhattacharryya distance which were also evaluated.

4. RESULTS

4.1 First overview of results

Figure 3. represents the output of cross-correlation estimator (XCOR), the output of an “mlp” and the ideal response (target). In this plot, where the variance of the noise was set to 0, it can be observed that both the cross-correlation and the neural network match closely the peak value of the target.

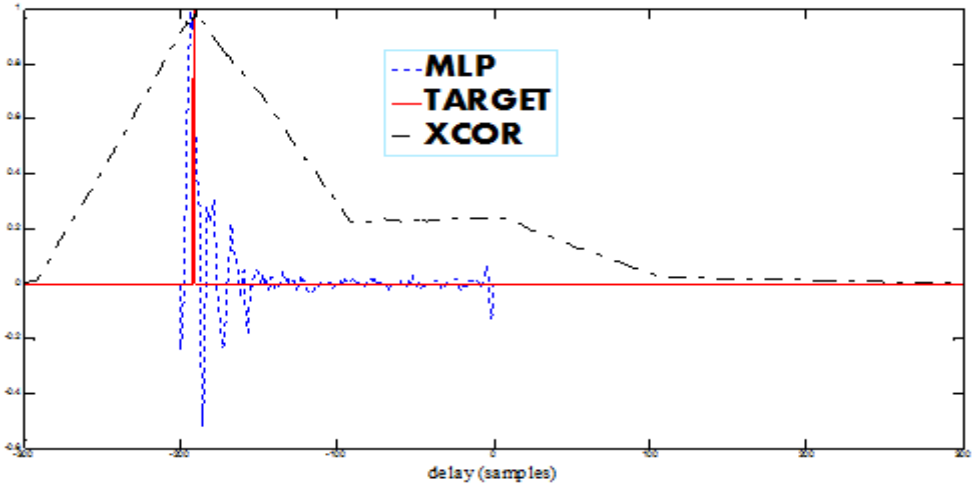

Figure 3. Output of mlp and XCOR estimators with respect to Target. Noise variance= 0

However, the shape of the target distribution is much closer to the shape of the neural network than to that of the cross-correlation. This is adequately described by the following Q_{KL}measures: $Q_{KL}(Target)$=0, $Q_{KL}(MLP) = 0.0022$, $Q_{KL}(XCOR) = 13.82$. Q_{KL}consistently produces low values when the distribution at hand is close to the target and higher values when that distributions is less close.

Figure 4. represents the output of the cross-correlation estimator (XCOR), the output of an "mlp" and the ideal response (target), this time in the presence of variable noise. It can be observed that cross-correlation is performing poorly at estimating the nominal delay whereas the neural network still closely matches the target. Beyond the nominal delay value, the overall shape of the distribution is also slightly affected by noise: the neural network, although it performs much closer to the target than cross-correlation does, is noisier than previously and has more leakage and ripples. This is adequately reflected by the Q_{KL} measures: $\mathrm{Q_{KL}(Target)}$=0, $\mathrm{Q_{KL}(MLP)} = 0.1895$, $\mathrm{Q_{KL}(XCOR)} = 14.30$.

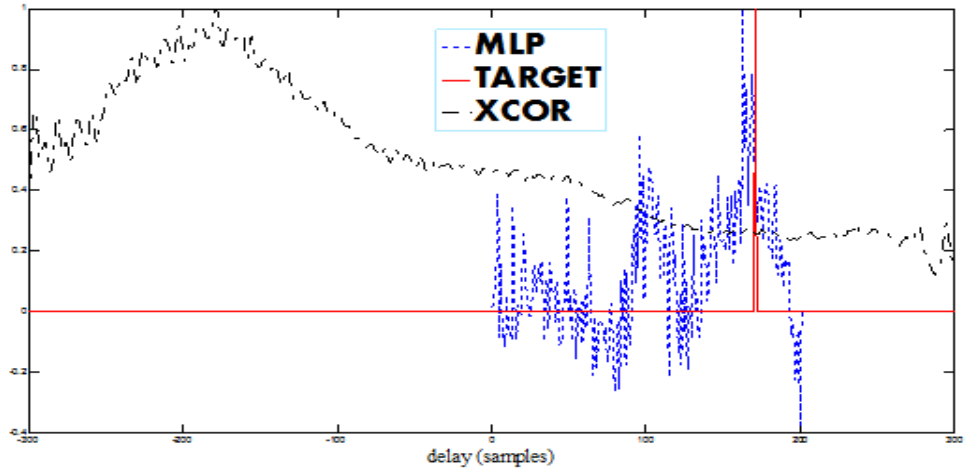

Figure 4. Output of mlp and xcor estimators with respect to target. Variance is variable (dataset 8).

4.2 Statistical significance

Tables 2 and 3 summarize the evolution of the mean of the error and its standard deviation for the tested estimators as variance changes. Among the non-supervised methods, the results obtained for GCC-SCOT, GCC-PHAT, the standard cross-correlation and AED are presented. At a noise variance of 0, SNR is infinite so that cross-correlation performs better than any of the methods under scrutiny, it is indeed an optimal estimator in such rare conditions. All "mlps" perform well: they have low variance and a small bias (one sample). On the contrary, Adaptive Eigenvalue Decomposition performs poorly, probably due to the absence of convolutive mixture.

Table 2. Standard deviation of various estimators for current datasets

	MLPA	MLPB	MLPC	MLPD	SCOT	PHAT	XCOR	AED
$\sigma_N^2 = 0$	1.32	12.30	1.32	12.30	8.16	3.86	0	32.86
$\sigma_N^2 = 0.2\sigma_S^2$	8.89	11.29	8.78	11.60	115.97	115.70	119.86	79.63
$\sigma_N^2 = 0.4\,\sigma_S^2$	16.55	17.74	16.55	18.78	113.59	113.27	114.09	78.34
$\sigma_N^2 = 0.5\,\sigma_S^2$	19.43	21.67	18.22	22.33	110.61	110.21	106.84	78.47
$\sigma_N^2 = 0.6\,\sigma_S^2$	21.82	25.62	20.00	25.88	107.11	106.70	102.86	78.82
$\sigma_N^2 = 0.8\,\sigma_S^2$	26.29	27.92	25.30	30.40	100.01	99.47	102.37	79.08
$\sigma_N^2 = \sigma_S^2$	30.23	31.81	29.33	32.39	94.95	94.24	106.29	78.87
$variable\ \sigma_N^2$	23.35	26.95	23.31	26.95	110.85	110.48	117.78	79.61

Table 3. Mean of the error of various estimators for current datasets

	MLPA	MLPB	MLPC	MLPD	SCOT	PHAT	XCOR	AED
$\sigma_N^2 = 0$	0.98	5.56	0.98	5.56	0.40	0.16	0	43.29
$\sigma_N^2 = 0.2\sigma_S^2$	2.13	3.46	2.19	3.19	196.74	196.38	222.51	207.95
$\sigma_N^2 = 0.4\,\sigma_S^2$	4.86	6.88	4.86	6.69	181.46	181.09	271.72	200.99
$\sigma_N^2 = 0.5\,\sigma_S^2$	6.60	9.23	8.55	9.05	171.46	171.02	292.26	197.93
$\sigma_N^2 = 0.6\,\sigma_S^2$	8.55	12.12	11.15	11.43	161.99	161.57	304.85	195.04
$\sigma_N^2 = 0.8\,\sigma_S^2$	13.09	17.66	16.49	16.57	147.86	147.38	313.32	191.80
$\sigma_N^2 = \sigma_S^2$	18.74	23.27	18.95	19.49	138.76	138.22	312.02	191.41
$variable\ \sigma_N^2$	8.84	11.62	8.84	11.62	171.77	171.35	274.01	199.54

As variance increases, the neural solutions prove to perform consistently better than any of the other methods at stake. The latter display non-monotonic and inconsistent evolutions both on their means and variances. As noise increases, all non-supervised methods face large variance and strongly biased estimates (systematically above 100 samples whereas neural solutions remain below 25). The inconsistency of these responses indicates poor estimators. Failure due to noise has been frequently demonstrated inliterature [13, 15], yet even in high noise it is found that the neural solution remains satisfactory.

Boxplots *(figure 5 to 8)* provide us additionally with a compact understanding of the performance of the various estimators and some additional statistics. For each box, the central mark is the median of the error, edges of the box are set to the 25th and 75th percentiles, the whiskers extend to the most extreme data points and outliers are displayed as triangles beyond the whiskers.

It can be observed that all trained "mlps" systematically outperform all non-supervised methods when noise is present. They also consistently perform with small and controlled bias and variance. It is also remarkable that with no noise the correlation methods perform well. In particular, the standard cross-correlation is unbiased, has no variance and thus no outliers.

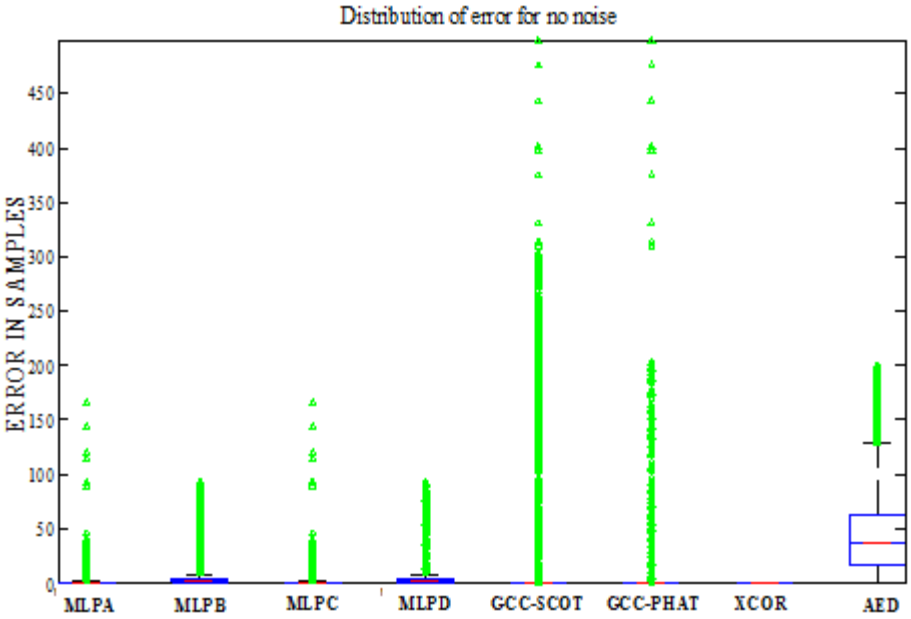

Figure 5. boxplot representation of the error distribution of various estimators when noise is absent.

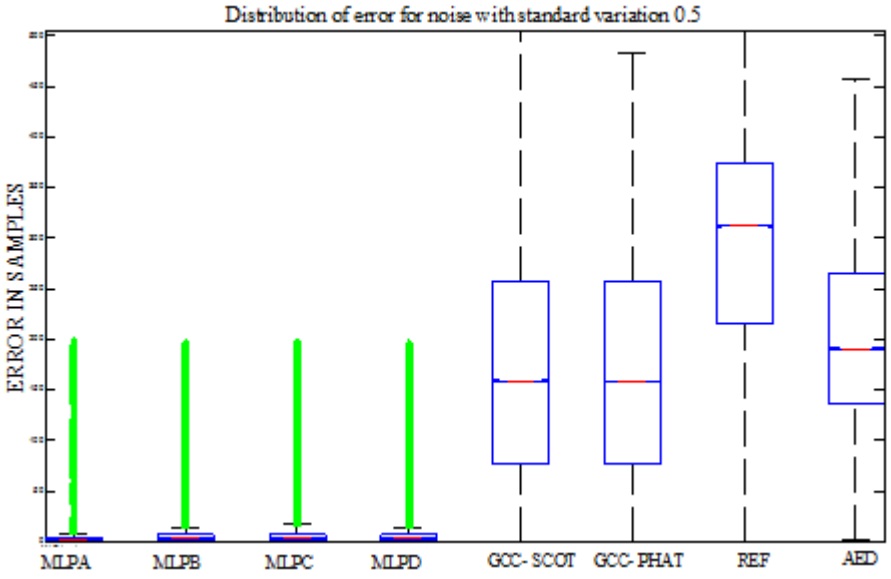

Figure 6. boxplot representation of the error distribution of various estimators when noise variance equals 0.5 signal variance.

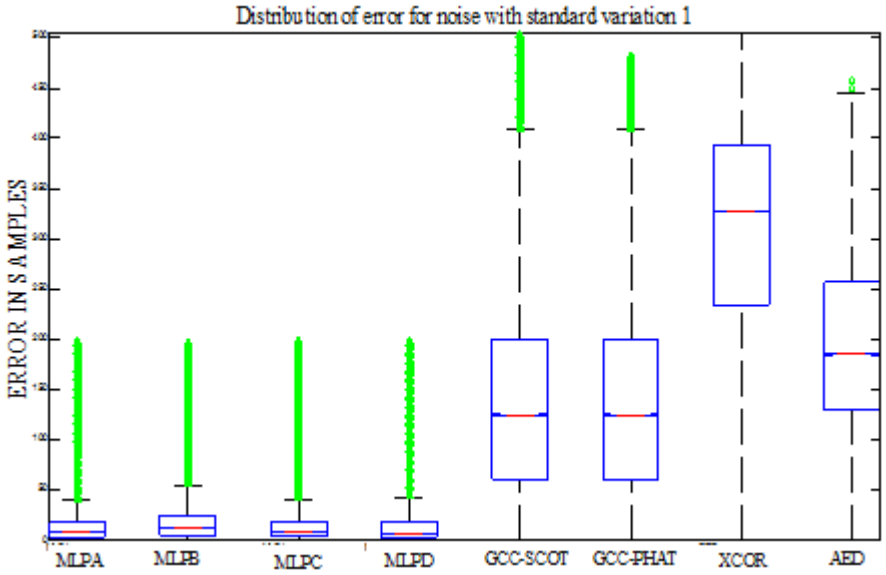

Figure 7. boxplot representation of the error distribution of various estimators when noise variance equals signal variance.

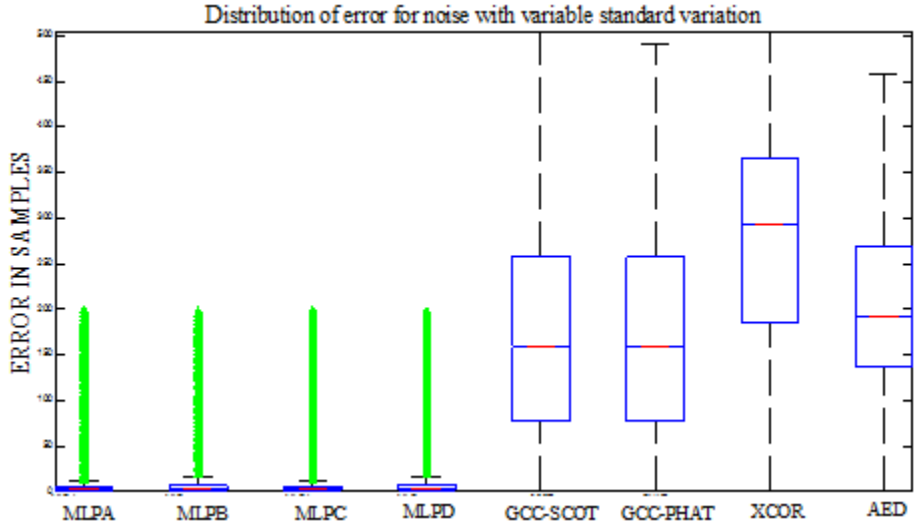

Figure 8. boxplot representation of the error distribution of various estimators with changing noise variance.

5. CONCLUSIONS

In this paper supervised neural networks were used for a successful time-delay estimation and proved to outperform benchmark methods both for the nominal estimation of time-delay and in approximating an ideal time-delay response. As an entry for localization this robust time-delay estimates would produce drastically more consistent location estimates. The integration of these improved time-delay estimators both in underwater and in room acoustics is the object of ongoing research projects.

REFERENCES

[1] Houégnigan, Ludwig, et al. "Space–time and hybrid algorithms for the passive acoustic localisation of sperm whales and vessels." Applied acoustics 71.11 (2010): 1000-1010.

[2] André, M., et al. "Localising Cetacean Sounds for the Real-Time Mitigation and Long-Term Acoustic Monitoring of Noise, Advances in Sound Localization." InTech (2011).

[3] Houegnigan, Ludwig, et al. "Neural networks for the localization of biological and anthropogenic source at neutrino deep sea telescope." OCEANS 2015-Genova. IEEE, 2015.

[4] Carter, G. Clifford. "Coherence and time delay estimation." Proceedings of the IEEE 75.2 (1987): 236-255.

[5] Fowler, Mark L., and Xi Hu. "Signal models for TDOA/FDOA estimation." IEEE Transactions on Aerospace and Electronic Systems 44.4 (2008): 1543-1550.

[6] Carter, G. C. "Time delay estimation for passive sonar signal processing." IEEE Transactions on Acoustics, Speech, and Signal Processing 29.3 (1981): 463-470.

[7] Chen, Jingdong, Jacob Benesty, and Yiteng Huang. "Time delay estimation in room acoustic environments: an overview." EURASIP Journal on applied signal processing 2006 (2006): 170

[8] Huang, Yiteng Arden, Jacob Benesty, and Jingdong Chen. "Time delay estimation and source localization." Springer Handbook of Speech Processing. Springer Berlin Heidelberg, 2008. 1043-1063.

[9] Shaltaf, Samir. "Neural-network-based time-delay estimation." EURASIP Journal on Applied Signal Processing 2004 (2004): 378-385.

[10] Shaltaf, Samir J., and Ahmad A. Mohammad. "Neural networks based time-delay estimation using DCT coefficients." American Journal of Applied Sciences 6.4 (2009): 703.

[11] Knapp, Charles, and Glifford Carter. "The generalized correlation method for estimation of time delay." IEEE Transactions on Acoustics, Speech, and Signal Processing 24.4 (1976): 320-327.

[12] Benesty, Jacob, Yiteng Huang, and Jingdong Chen. "Time delay estimation via minimum entropy." IEEE Signal Processing Letters 14.3 (2007): 157-160.

[13] Benesty, Jacob. "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization." The Journal of the Acoustical Society of America 107.1 (2000): 384-391.

[14] Reed, Feintuch, P. Feintuch, and N. Bershad. "Time delay estimation using the LMS adaptive filterstaticbehavior." IEEE Transactions on Acoustics, Speech, and Signal Processing 29.3 (1981): 561-571.

[15] Carter, G. Clifford, and E. Richard Robinson. "Ocean effects on time delay estimation requiring adaptation." IEEE Journal of oceanic engineering 18.4 (1993): 367-378.