# Activation Function Study for Wavelet Network

J. Fernando Marar[1,2] E. C. B. Carvalho Filho[2] L. Weigang[3] Leonardo D. Sá[3]

[1]Department of Computer Science
Universidade Estadual Paulista - Unesp -
Av. Luiz Edmundo Coube, s/n,
Bauru-SP, Brazil.

[2]Department of Computer Science
Universidade Federal de Pernambuco
Av. Prof. Luiz Freire, s/n,
Cx. Postal 7851, 50732-970, Recife-PE, Brazil.
e-mail : {jfm|ecdbcf}@di.ufpe.br

[3]Department of Meteorological Science
Instituto Nacional de Pesquisas Espaciais
Av. Dos Astronautas, 1758
Cx. Postal 515, 12201-970, S.J. dos Campos, Brazil.
e-mail : {wei|leo}@met.inpe.br

**Abstract**

The main purpose of this paper is to investigate theoretically and experimentally the use of family of Polynomial Powers of the Sigmoid (PPS) Function Networks applied in speech signal representation and function approximation. This paper carries out practical investigations in terms of approximation fitness (LSE), time consuming (CPU Time), computational complexity (FLOP) and representation power (Number of Activation Function) for different PPS activation functions. We expected that different activation functions can provide performance variations and further investigations will guide us towards a class of mappings associating the best activation function to solve a class of problems under certain criteria.

Keywords : PPS, Neural Networks, Function Approximation, Wavelets.

## 1 Introduction

Wavelet functions have been successfully used in many problems as the activation function of feed-forward neural networks [ZB92],[STK92], [PK93], [aa96]. There are claims that many biological fundamental properties can emerge from wavelets transforms [Dau88]. A number of practical problems are waiting for new approaches to be applied. Activation functions are very important in the process of improving the performance of wavelet networks when used in function approximation and representation.

A family of polynomial wavelets generated from powers of sigmoids (PPS) provides a robust way for designing neural network architectures [MFV96]. The PPS technique combines interesting features of neural networks, wavelet transform and classical polynomial ideas to tackle real world problems.

SPIE Vol. 3077 • 0277-786X/97/$10.00

This paper shortly describes the family of PPS function. It compares the classical radial basis functions to PPS-Radial and PPS-Wavelet functions by a linear model [Ach56]. Other experiments, speech signals, are represented by adaptives PPS Radial and Wavelets neural networks.

## 2 PPS - A Family of Polynomial Powers of Sigmoids

Let $\Upsilon : \mathbb{R} \rightarrow [0,1]$ be a sigmoid function defined by $\Upsilon(x) = \frac{1}{1+e^{-x}}$ . The *nth*-power of the sigmoid function is a function $\Upsilon^n : \mathbb{R} \rightarrow [0,1]$ defined by $\Upsilon^n(x) = (\frac{1}{1+\exp(-x)})^n$. Let $\Theta$ be the set of all power functions defined by 1:

$$\Theta = \{\Upsilon^0(x), \Upsilon^1(x), \Upsilon^2(x), \cdots, \Upsilon^n(x), \cdots\} \tag{1}$$

Radial functions are a special class the functions. Their characteristic feature is that their response decreases or increases monotonically with distance from a central point. The center, the distance scale, and the precise shape of the radial function are parameters of the model, all fixed it is linear. A typical radial function is the Gaussian [FP91]. A PPS radial function is obtained from the first derivative of $\Upsilon(x)$, it is written a linear combination of PPS:

$$\psi_1(x) = -\Upsilon^2(x) + \Upsilon(x) \tag{2}$$

In [MF97] an effective procedure for generating polynomial forms of wavelet functions from the successive powers of sigmoid functions is presented. The resulting functions, referred to as *polynomial wavelets*, represent a robust solution for the construction of neural networks based on wavelets. Now, it is illustrated four examples of PPS-wavelet functions can be defined by:

$$\begin{aligned}
\psi_2(x) &= 2\Upsilon^3(x) - 3\Upsilon^2(x) + \Upsilon(x) && (3)\\
\psi_3(x) &= -6\Upsilon^4(x) + 12\Upsilon^3(x) - 7\Upsilon^2(x) + \Upsilon(x) && (4)\\
\psi_4(x) &= 24\Upsilon(x)^5 - 60\Upsilon^4(x) + 50\Upsilon^3(x) - 15\Upsilon^2(x) + \Upsilon(x) && (5)\\
\psi_5(x) &= -120.\Upsilon^6(x) - 360\Upsilon^5(x) - 390\Upsilon^4(x) + 180.\Upsilon^3(x) - 31\Upsilon^2(x) + \Upsilon(x) && (6)
\end{aligned}$$

This construction technique is interesting because the family of PPS-wavelets comes naturally from a sequence of derivatives of the sigmoid function.

## 3 On Approximation of Functions by RBF and PPS

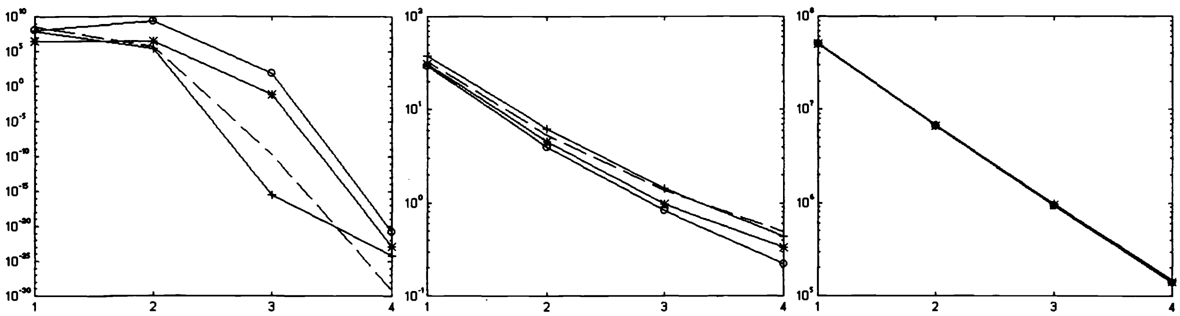
Radial functions are simply a class of functions. In principle, they could be employed in any sort of model (linear or nonlinear) and neural network (single or multi-layer) paradigms. However, in 1988 Broomhead and Lowe [BL88] showed that radial basis function network have traditionally been associated with radial functions in a single-layer network such as shown in Figure 2.

In order to investigate the function approximation , it was selected 10 different functions as defined in Table 1. Four sample sets of 201, 101, 51 and 26 equally spaced points were used to test the approximation on the interval $x \in [-10, 10]$ . In this investigation the functions $\psi_1$, $\psi_2$, $\psi_5$ and the Gaussian function were used to approximate the functions presented in Table 1

| | | | |
|---|---|---|---|
| $f_1(x) =$ | $\frac{\cos(x)-\sin(x)}{1+e^{-x}}$ | $f_2(x) =$ | $\frac{\sin(x)}{x}$ |
| $f_3(x) =$ | $-xe^{\frac{-x^2}{2}}$ | $f_4(x) =$ | $5\left(-\Upsilon^2(x) + \Upsilon^1(x)\right)$ |
| $f_5(x) =$ | $10\left(2\Upsilon^3(x) - 3\Upsilon^2(x) + \Upsilon^1(x)\right)$ | $f_6(x) =$ | $2\Upsilon^1(x) - \Upsilon^0(x)$ |
| $f_7(x) =$ | $8\cos(\Upsilon^1(x)) + 4\sin(\Upsilon^1(x)) - 8$ | $f_8(x) =$ | $\frac{1}{1+x^2}$ |
| $f_9(x) =$ | $\frac{1}{\sqrt{1+x^2}}$ | $f_{10}(x) =$ | $3\left(\Upsilon^1(x+2) + \Upsilon^1(x-2) - 2\Upsilon^1(x)\right)$ |

Table 1: illustration of analytic functions used in this application

The Figure 1(a) shows the least squared error for the functions. The Figure 1(b) shows the time in seconds spent. The Figure 1(c) shows the number of flops for the sets.



(a) (b) (c)

Figure 1: Plots of the squared errors, time spent and flops of the functions versus the sets of with samples with 201, 101, 51 and 26. We used the functions $\psi_1$ = "o", $\psi_2$ = "*", $\psi_5$ = "+" and the Gaussian by "- -" .

The test set 3 with 51 samples was best approximated by function $\psi_5$. For a large number of samples the process is not very good, the LSE is too high. But for samples like 51 and 26 the technique is very adequate.

## 4 PPS-Wavelets and PPS-Radial Neural Networks

The interest for the application of neural networks in function approximation was much enhanced after the results obtained by Hornik [Hor89] and Cybenko [Cyb89]. More recently, many works have been able to show the attractive characteristics of using neural networks, more specifically feedforward neural networks, as universal function approximators [Fun89], [HN89], [Hor89],[Cyb89], [LYPS93], [WG95]. In general, every continuous function can be represented by Equation 7:

$$\tilde{f}(x) = \sum_{j=1}^{m} w_i \psi(d_j(x - t_j)) \tag{7}$$

where $w_i$,$t_i$ e $d_i$ are the weights, shifts and dilations of the mother function $\psi(\bullet)$, respectively [Chu92], and $m$ is the number of basis function employed. In the context of neural networks, this definition can be represented through the architecture illustrated in Figure 2.

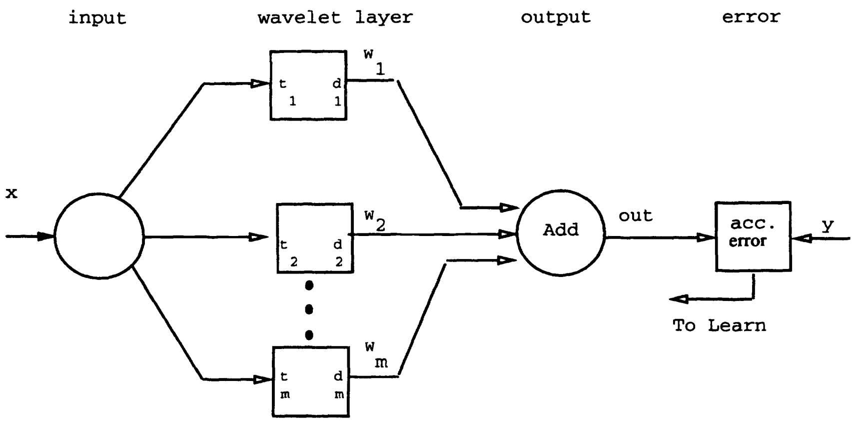



Figure 2: A neural network which employs wavelets as the activation function. The first layer corresponds to the set of adaptable wavelet basis functions. The second (output) layer combines the responses of the basis function and produces an output *out* used to compute the cost function *acc.error*. The error propagates back to the network for the continuation of the learning process if it exceeds a maximum allowed.

In this paper, a feedforward neural network is defined by an input $\{x_n, y_n\}$, where $n$ is the number of patterns, a set of coefficients $(w_j), j = 1..m$ that correspond to the weights of the basis functions, by the elements $(t_j), j = 1..m$, and $(d_j), j = 1..m$, that represent the coefficients of shifts and dilations of the activation function $\psi(d_j(x - t_j))$; and by the variable $(y_i)$ that corresponds to the desired output of the network.

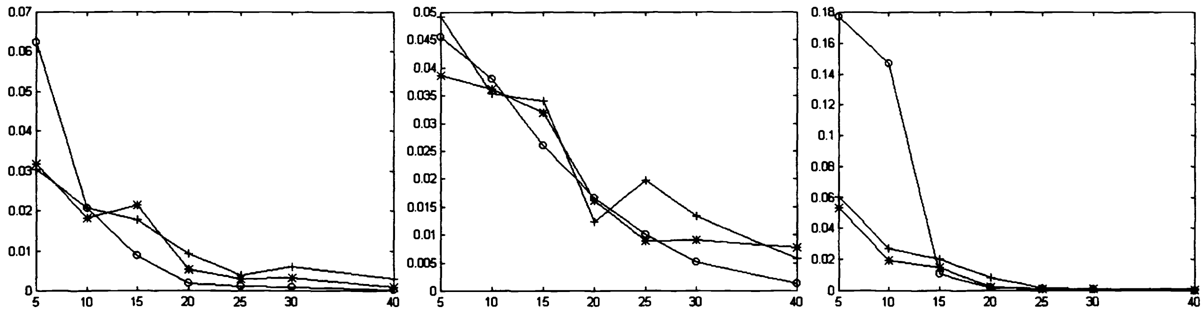
## 5 PPS Activation Functions Applied on Speech Representation

In order to demonstrate the function approximation performance of PPS-Wavelet and PPS-Radial network in comparison with other activation functions, we select the problem of representing three phonemes, "a", "e" and "i", used by Szu at al[STK92]. Those patterns were extracted from speech signal and it can be seen in Figures 4(a), 5(a) and 6(a), it is approximates just that single period. Actually, the spoken signal is a periodic pattern of the phonemes the same techniques.

The main purpose of these experimental results with PPS Function Networks is to carried out practical investigations in terms of approximation fitness(LSE), time consuming(CPU Time), computational complexity(FLOP) and representation power(Number of Activation Function) for different PPS activation functions. Intuitively, we expect that different activation functions can provide performance variation and further investigations will guide us for a class of mapping associating the best activation function to solve a class of problem under certain criteria.

Three PPS neural networks were trained with activation functions $\psi_1$, $\psi_2$ and $\psi_5$. Where $\psi_1$ is a PPS-radial basis function and $\psi_2$,$\psi_5$ are PPS-wavelets. The experiments were carried out with 5, 10 , 15 , 20 , 25, 30 and 40 hidden neurons as described in Figure 2.

The least square errors results obtained from used of the three PPS functions applied to speech signal of the vowel "a", "e" and "i" can be seem in Figure 3(a),(b) and (c) respectively.



(a) (b) (c)

Figure 3: Least square error for representing the vowels "a", "e" and "i" with $\psi_1 = +$, $\psi_2 = *$ and $\psi_5 = o$ activation functions varying the number of hidden neurons.

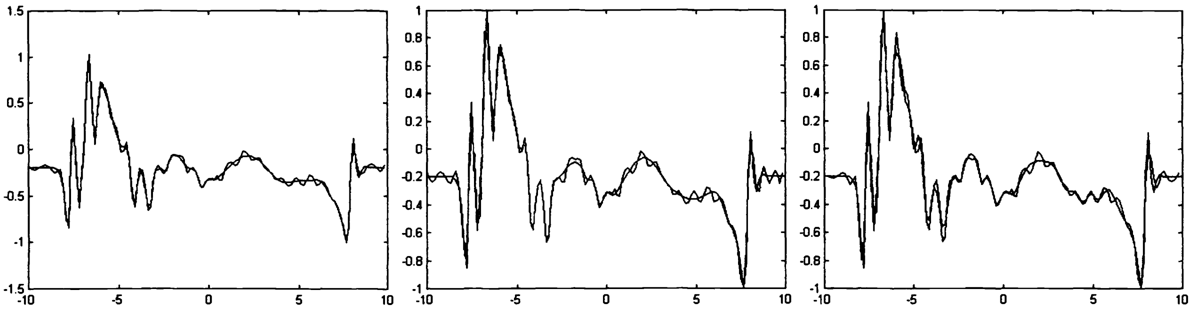
The Figure 3 demonstrates that vowel signal have differents degrees of difficulty of representation. Also, for small number of hidden neurons, the use of different activation functions provides varied performance for the neural network. When increasing the number of hidden neurons, the LSE for different activation functions gets close. The best results were found when using $\psi_5$.

The signal representation of the vowel "a" with 25 hidden neurons, "e" with 40 hidden neurons and "i" with 20 hidden neurons using $\psi_1$, $\psi_2$ and $\psi_5$ are illustrated in Figures 4, 5 and 6 , respectively.

The experimental results confirms the potential application of PPS functions for another real world problems. Theoretical results in Marar [MF97] have demonstrated the robustness of PPS techniques and further use in many correlated areas.

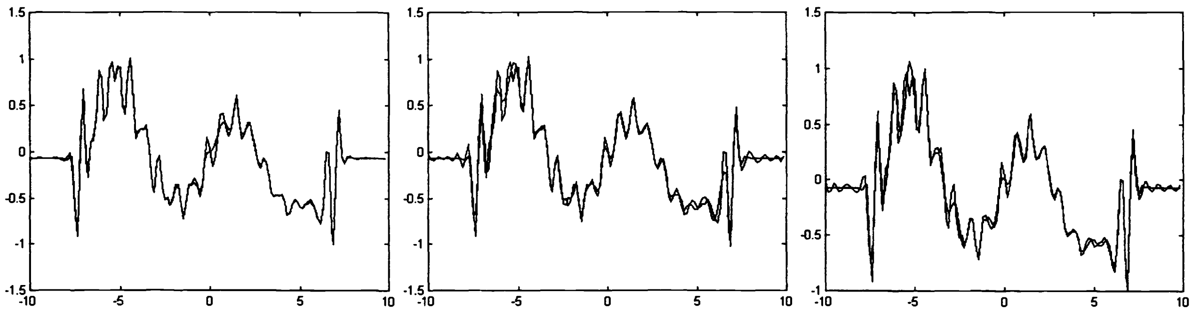
## 6 Conclusions

In this paper, a family of polynomial wavelets was presented constructed from powers of sigmoids. It has been shown, in the context of function approximation, that this set of functions can provide a very good approximation capability, with a fast convergence of the training process. The problem of speech



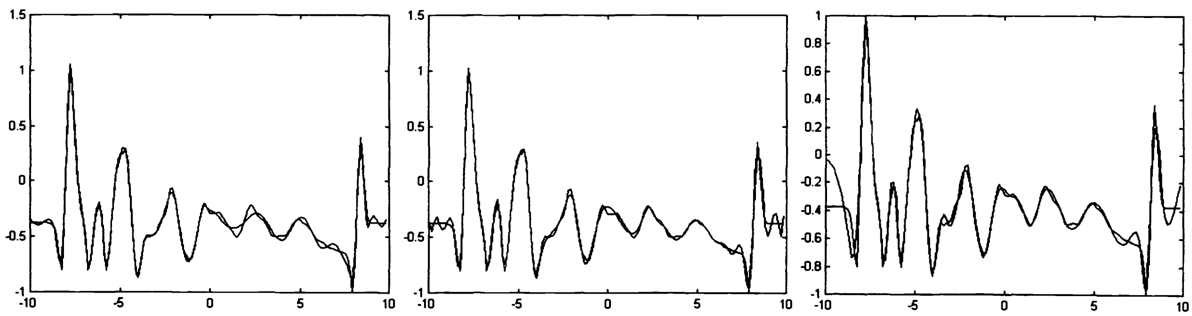

(a) (b) (c)

Figure 4: The signal of the Vowel "a" represented by $\psi_5$, $\psi_2$ and $\psi_1$ respectively




(a) (b) (c)

Figure 5: The signal of the Vowel "e" represented by $\psi_5$, $\psi_2$ and $\psi_1$ respectively




(a) (b) (c)

Figure 6: The signal of the Vowel "i" represented by $\psi_5$, $\psi_2$ and $\psi_1$ respectively

signal representation was investigated with the use of PPS neural networks. The results demonstrated a good fitness on original signals. One of the advantages of using wavelets in the design of neural networks is the reduced number of basis function usually required to give a good approximation to a function [STK92],[PK93],[MF97]. This family of PPS functions also represents a potential scheme for the development of efficient neural network structures in pattern classification problems [STK92].


## Acknowledgements

The authors would like to acknowledge the Brazilian Research Council (CNPq), the Brazilian Federal Agency for Postgraduate Studies (CAPES) and UNESP-Bauru for partially supporting this work. The authors to thank Drs H. Szu, B. Telfer and S. Kadambe for them supplying the data of speech signals.


# References


[aa96] L. Weigang at al. Neural networks adaptive wavelets for prediction of northeastern brazil rainfall monthly anomalies time series. *SPIE-Aerosense/96- 10th Annual International Aerosense Symposium*, 2760- Neural Networks:175–187, 1996.

[Ach56] N.I. Achieser. *Theory of Approximation.* Frederick Ungar Publishing Co., New York, 1956.

[BL88] D.S Broomhead and D. Lowe. Multivariable funcional interpolation and adaptive networks. *Complex Systems*, pages 2:321–355, 1988.

[Chu92] C.S. Chui. *An Introduction to Wavelets.* Academic Press,Inc,Boston, 1992.

[Cyb89] G. Cybenko. Approximation by superposition of a sigmoidal function. *Mathematics of Control, signals and Systems*, 2:303–314, 1989.

[Dau88] J. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoust. Speech,Signal Proc.*, 8, July 1988.

[FP91] F.Girosi and T. Poggio. Networks for learning: A view from the theory of approximation of functions. *Neural Networks: Concepts,Applications and Implementations*, I:110–154, 1991.

[Fun89] K. Funahashi. On the approximate realization of continuos mappings by neural networks. *Neural Networks*, 2:183–192, April 1989.

[HN89] R. Hecht-Nielsen. Theory of the backpropagation neural network. *Proc. IJCNN (Washington) D.C.*, I-593:18–22, June 1989.

[Hor89] K. Hornik. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.

[LYPS93] M. Leshno, L.V Ya, A. Pinkus, and S. Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6:861–867, 1993.

[MF97] J.Fernando Marar and E.C.B.Carvalho Filho. Polynomial powers of sigmoids: Theory and pratical applications. *PhD Thesis - UFPE - Brazil - written in portuguese*, 1997.

[MFV96] J.Fernando Marar, E.C.B.Carvalho Filho, and G.C. Vasconcelos. Function approximation by polynimial wavelets generated from powers of sigmoids. *SPIE-Aerosense/96- 10th Annual International Aerosense Symposium*, 2762- Wavelet Application III:365–374, 1996.

[PK93] Y.C. Pati and P.S. Krishnaprasad. Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations. *IEEE Transactions on Neural Networks*, 4:73–85, January 1993.

[STK92] Harold Szu, Brian Telfer, and Shubha Kadambe. Neural network adaptive wavelets for signal representation and classification. *Optical Engineering*, 31:1907–1916, September 1992.

[WG95] J. Wray and G.G.R. Green. Neural network, approximation theory, and finite precision computation. *Neural Network*, 8:31–37, January 1995.

[ZB92] Qinghua Zhang and Albert Benveniste. Wavelet networks. *IEEE Transactions on Neural Networks*, 3:889–898, November 1992.