THE USE OF NONLINEAR SIGNAL DECOMPOSITION INTO FUNCTIONAL SERIES FOR SPEECH RECOGNITION *

Alexander M. Krot[a] and Polina P. Tkachova[b]

[a] Institute of Engineering Cybernetics of the National Academy of Sciences of Belarus
6, Surganov Str., 220012, Minsk, Belarus

Tel.: (375) 172 842086, Fax.: (375) 172 318403
[b]Belarusian State University,
4,Skoriny Av., 220050 Minsk, Belarus

Abstract. The nonlinear speech signal decomposition based on Volterra-Wiener functional series is described. The nonlinear filter bank structure is proposed for phonemes recognition solving.
Keywords: nonlinear signal decomposition, Volterra-Winer functional series, nonlinear filter bank structure, phoneme recognition.

1. Introduction

For speech recognition problems solving there are some paradigms and approaches. Among of them we can mention statistical approach based on hidden Markov models [1], [2], [3], [4], nonlinear dynamic method using neural networks [5], [6], algebraic approach [7], linguistic methods [8] and so on. However, in spite of some success in this problem solving there exist difficulties connecting with speech signal peculiarities. Speech signal by its nature has two aspects [4]. Firstly, speech of the person is defined by physical parameters, such as vocal tract length, glottal size and so on. Secondly, the speech producing is impossible without neural control of the articulations, which defines the personal learned abilities such as dialect or regional accents, pronunciation, speed and timing of the articulators. These two reasons find the reflection in speech signal nonlinearity, and it is necessary to implement nonlinear model.

This paper presents the nonlinear speech signal decomposition based on Volterra-Wiener functional series [9], [10]. It is shown the usage this nonlinear decomposition for nonlinear bank filters designing. It is proposed to solve the phoneme recognition problem by means of identification algorithm of these filters.

2. Linear and nonlinear decomposition of signal into series of functions and functionals

It is well known that *music* signal *y(t)* in the time t can be represented by means of Fourier series of the kind:

$$y(t)=\sum_{k=-\infty}^{\infty} Y(\omega_k)e^{i\omega_k t}, \qquad (1)$$

where $Y(w_k)$ are the coefficients of Fourier series as the following:

$$Y(w_k)=\frac{1}{T}\int_0^T y(t)e^{-i\omega_k t}dt. \qquad (2)$$

* This work was supported by International Science and Technology Center (ISTC) under project B-95

Here T is interval of signal observation and $w_k = \frac{2\pi}{T}k$ is angular frequency. The relation (1) is described a linear decomposition of temporal function $y(t)$ into series from orthogonal Fourier functions $exp(iw_k t)=cosw_k t+isinw_k t$. On other hand, music signal as linear process, which is generated by linear dynamic system (LDS) exclusively has a full representation by means of linear series from orthogonal functions Since a stationary LDS may be characterized by means of transfer function $H(w_k)$ the input signal $x(t)$ acting on LDS generates its output signal $y(t)$:

$$y(t)= \sum_{k=-\infty}^{\infty} H(\omega_k)X(\omega_k)e^{i\omega_k t} . \quad (3)$$

As concerning *speech* signal, this signal is the product of strongly nonlinear dynamic systems (NDS), i.e. one is nonlinear process when its harmonic components have actions each other. In connection with this such signal can be represented by means of Volterra-Wiener series as the following [9], [10]:

$$\begin{aligned} y(t)=h_{0+} \sum_{k_1=-\infty}^{\infty} H_1(\omega_{k_1})X(\omega_{k_1},\theta)e^{i\omega_{k_1}t} + \sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} H_2(\omega_{k_1},\omega_{k_2})X(\omega_{k_1},\theta)X(\omega_{k_2},\theta)e^{i(\omega_{k_1}+\omega_{k_2})t} - \\ - D_x \sum_{k_1=-\infty}^{\infty} H_2(\omega_{k_1},-\omega_{k_1}) + \sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}\sum_{k_3=-\infty}^{\infty} H_3(\omega_{k_1},\omega_{k_2},\omega_{k_3})\times \\ \times X(\omega_{k_1},\theta)X(\omega_{k_2},\theta)X(\omega_{k_3},\theta)e^{i(\omega_{k_1}+\omega_{k_2}+\omega_{k_3})t} - \\ - 3D_x \sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} H_3(\omega_{k_1},-\omega_{k_1},\omega_{k_2})X(\omega_{k_2},\theta) + \ldots . \end{aligned} \quad (4)$$

Comparing (3) with (4) we can see that mentioned above linear decomposition is particular case of this nonlinear decomposition similarly LDS is specific case of NDS. It follows from (4) the stationary NDS represents in the form of parallel connection of one-, two-, …, multi-dimensional stationary LDSs with input signals $x(t_1,q)$, $x_2(t_1,t_2,q)= x(t_1,q)\, x(t_2,q),\ldots, x_m(t_1,\ldots, t_m,q)= x(t_1,q)\ldots x(t_m,q)$ respectively.

3. The Volterra-Wiener functionals on finite intervals and their identification algorithm based on measuring the Wiener kernels

When the Wiener method [9], [10] applies in practice for the NDS analysis it is necessary to tell about physical white noise, that is about the process with the finite spectrum which overlaps the bandwidth of the system under study. Moreover, when NDS model based on Volterra-Wiener series is realized on a computer discrete input x_n and output y_n signals and Wiener kernels $h_m[n_1,\ldots,n_m]$ will have a finite duration in time; that is why some refinement of the relation (4) is required [11].

Let us write discrete Volterra-Wiener functionals for the discrete signals x_n and h_0, $h_1[n]$, $h_2[n_1,n_2],\ldots$ given on the finite time intervals, that is $n, n_1,\ldots,n_L= 0, 1,\ldots,N-1$ [11], [12].

To represent one-dimensional sequences y_n and x_n in with the finite length N in frequency domain let us consider discrete Fourier transform (DFT) [13]:

$$X_k = \sum_{n=0}^{N-1} x_n w_N^{nk} , w_N = exp\left(-j\frac{2\pi}{N}\right), k=0,\ldots,N-1 \quad (5)$$

and also its multidimensional analogues, i.e. the m-dimensional DFT's $(m=2,3,\ldots)$ of the kind:

$$H_m[k_1,\ldots,k_m]=\sum_{n_1=0}^{N-1}\ldots\sum_{n_m=0}^{N-1}h_m[n_1,\ldots,n_m]\,w_N^{n_1k_1+\ldots+n_mk_m}\ . \qquad (6)$$

Taking into account (15), (16) and using inverse DFT and its multidimensional analogues let us transform relation (14) to the form [11], [12]:

$$y_n=h_0+\frac{1}{N}\sum_{k_1=0}^{N-1}H_1[k_1]X_{k_1}w_N^{-nk_1}+\frac{1}{N^2}\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}H_2[k_1,k_2]X_{k_1}X_{k_2}w_N^{-n(k_1+k_2)}-$$
$$-\frac{D_x}{N}\sum_{k_1=0}^{N-1}H_2[k_1,N-k_1]+\frac{1}{N^3}\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\sum_{k_3=0}^{N-1}H_3[k_1,k_2,k_3]X_{k_1}X_{k_2}X_{k3}w_N^{-n(k_1+k_2+k_3)}-$$
$$-\frac{3D_x}{N^2}\sum_{k_1=0}^{N-1}\sum_{k_3=0}^{N-1}H_3[k_1,N-k_1,k_3]X_{k_3}w_N^{-nk_3}+\ldots. \qquad (7)$$

We consider index $N\text{-}k_1$ in (17) with respect to modulo N [12], i.e. $(N\text{-}k_1)\,mod N=((N\text{-}\text{-}k_1))$. If x_n is N-pointed sample of stationary Gaussian white noise then

$$h_0=M[y_n]=\frac{1}{N}\sum_{n=0}^{N-1}y_n=\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{N}\sum_{k=0}^{N-1}Y_kw_N^{-kn}=\frac{1}{N^2}\sum_{k=0}^{N-1}Y_kN\delta_{k,0}=\frac{1}{N}Y_0\ ..$$

The identification scheme of discrete NDS is similar to Wiener's circuit for determining m-order kernel [9]. This scheme may be present as follows: white Gaussian noise with zero mean and variance D_x, is given by the inputs of unknown NDS and bank of m complex exponential filters. Then the output signals from the system and bank are multiplied and the result signal is averaged [10].

Let us calculate DFT-image of kernel $h_1[n]$. First obtain signal $y_n^{(1)F}$ from the output of known system that is filter with complex exponential impulse response $w_N^{\tau}=exp\left[-j\frac{2\pi}{N}\tau\right]$:

$$y_n^{(1)F}=\sum_{\tau_1=0}^{N-1}x_{((n-\tau_1))}w_N^{k\tau_1}=\sum_{\tau_1=0}^{N-1}w_N^{k\tau_1}\frac{1}{N}\sum_{m=0}^{N-1}X_mw_N^{-m(n-\tau_1)}=\frac{1}{N}\sum_{m=0}^{N-1}X_mw_N^{-mn}\sum_{\tau_1=0}^{N-1}w_N^{\tau_1(m+k)}=$$
$$=\frac{1}{N}\sum_{m=0}^{N-1}X_mw_N^{-mn}Nd_{m,((N-k))}=X_{((N-k))}w_N^{-((N-k))n}=X_k^*w_N^{kn}\ . \qquad (8)$$

Calculate an averaged signal from the scheme output, consisting of unknown NDS, known system and multiplier:

$$M[z_n^{(1)}]=M[y_ny_n^{(1)F}]=h_0M[X_k^*]w_N^{kn}+\frac{1}{N}\sum_{k_1=0}^{N-1}H_1[k_1]M[X_{k_1}X_k^*]w_N^{n(k-k_1)}+$$
$$+\frac{1}{N^2}\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}H_2[k_1,k_2]M[X_{k_1}X_{k_2}X_k^*]w_N^{-n(k_1+k_2-k)}-\frac{D_x}{N}M[X_k^*]w_N^{nk}\sum_{k_1=0}^{N-1}H_2[k_1,N-k_1]+$$
$$+\frac{1}{N^3}\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\sum_{k_3=0}^{N-1}H_3[k_1,k_2,k_3]M[X_{k_1}X_{k_2}X_{k_3}X_k^*]w_N^{-n(k_1+k_2+k_3-k)}-$$
$$-\frac{3D_x}{N^2}\sum_{k_1=0}^{N-1}\sum_{k_3=0}^{N-1}H_3[k_1,N-k_1,k_3]M[X_{k_3}X_k^*]w_N^{n(k-k_3)}+\ldots. \qquad (9)$$

It was shown in [10], [12] that if $\{x_n\}$ is the sample of white Gaussian noise, then their coefficients X_k are N-pointed sample of Gaussian noise, that is the properties are true:

$$M[X_k]=0;$$

$$M[X_{k_1}X_{k_2}]=D_X d_{k_1,N-k_2};$$

$$M[X_{k_1}X_{k_2}X_{k_2}]=0; \qquad (10)$$

$$M[X_{k_1}X_{k_2}X_{\ 3}X_{k_4}]=D_X^2[d_{\ 1,N-k_2}d_{\ 3,N\ k_1}+\ _{1,N-k_3}d_{\ 2,N-k_4}+d_{\ 1,N\ k_4}d_{\ 2,N-k_3}],\ \ldots$$

$D_X\ \ \frac{1}{N}\sum_{=0}^{N\ 1}X\ \ X_k^*$ - is the variance of partial population of coefficients X_k, and $D_X=ND_x$. Taking into account (10) we transform (9) to the form

$$M[y_n y_n^{(1)F}]=\frac{D_X}{N}H_1[k]+\frac{D_X^2}{N^3}\{\sum_{k_1=0}^{N-1}H_3[k_1,N-k_1,k]+\sum_{k_1=0}^{N-1}H_3[k_1,k,N-k_1]+$$

$$+\sum_{k_2=0}^{N-1}H_3[k,k_2,N-k_2]\}-\frac{3D_X^2}{N^3}\sum_{k_1=0}^{N-1}H_3[k_1,N-k_1,k]=D_x H_1[k]. \qquad (11)$$

Symmetry property of kernel $H_3[k_1,k_2,k_3]$ was used for deduction (11). On the other hand, using (8) we can calculate the mean of partial population:

$$M[y_n y_n^{(1)F}]=\frac{1}{N}\sum_{n=0}^{N-1}X_k^* w_N^{kn}\frac{1}{N}\sum_{m=0}^{N-1}Y_m w_N^{-mn}=\frac{1}{N}Y_k X_k^*. \qquad (12)$$

Taking into account (11), (12) we have that [11], [12]

$$H_1[k]=\frac{Y_k X_k^*}{ND_x}=\frac{Y_k X_k^*}{D_X}. \qquad (13)$$

According to relation (13) $H_1[k]$ is the sample of transfer function $H_1(w)$ for the stationary LDS, identified on the basis of stationary white noise $\{x_n\}$.

The DFT-image of kernel $h_2[n_1,n_2]$ may be calculated in an analogous manner. In this case signal $y_n^{(2)F}$ is the product of outputs for two complex exponential filters, and either is described by the relation (6), that is $y_n^{(2)F}=X_{k_1}^* X_{k_2}^* w_N^{n(k_1+k_2)}$. As a result we have [11], [12]:

$$H_2[k_1,k_2]=\frac{Y_{k_1+k_2}X_{k_1}^* X_{k2}^*}{2ND_x^2}-\frac{h_0}{D_x}N\delta_{k_1,N-k_2}. \qquad (14)$$

We can also show that DFT-image of kernel $h_3[n_1,n_2,n_3]$ is [12]:

$$H_3[k_1,k_2,k_3]=\frac{Y_{k_1+k_2+k_3}X_{k_1}^* X_{k2}^* X_{k3}^*}{6ND_x^3}-\frac{N(H[k_1]\delta_{k_2,N-k_3}+H[k_2]\delta_{k_1,N-k_3}+H[k_3]\delta_{k_1,N-k_2})}{6D_x}.$$

In this paper we use the identification scheme by Wiener kernel measuring for NDS testing by the white noise on finite interval [11], [12] and also the analogous identification scheme for other types testing signals (see, for example Ref. [14], [15]).

4. The synthesizer and recognizer of phonemes of Belarusian language based on nonlinear decomposition

The obtained nonlinear decomposition (4) may be used for identification of group of phonemes (say, sonorous phonemes of Belarusian language) by means of mentioned m-order multidimensional nonlinear filters.

For recognition the following classification is used [16]. All the phonemes of Belarusian language are divided into two groups: the first group has vocal (vowel) ones, the second group has consonant ones.

The vocal phonemes are again divided into labial once (Î, Ó) and nonlabial once (À, Ý, I{Û}). (The sound Û is not considered as individual phoneme since in Belarusian language it meet only after hard consonants and is modification of the phoneme I).

The consonants (due to difficulties in their recognition) are classified on the basis of these two approaches [16].

According to the first approach we are using the nonlinear filters structure consisting of 7 nonlinear Volterra-Wiener filters. Each from them may be stimulated by a different testing signal (for example by white noise, by colored noise, by tone, by tone plus noise etc.) depending on their position in the first scheme. But first approach has essential lack because sonorous group consists of 12 phonemes that it makes difficult to use the identification scheme in practice.

That is why we apply the second approach for more reliable phoneme recognition. According to second approach we have a recognizer (synthesizer) in the form of the nonlinear bank filters consisting of 10 Volterra-Wiener filters which may include from 1 to 6 nonlinear multidimensional (*m*-order) filters (or functionals). Thus, by increasing the channel number in the nonlinear filter structures (or the number of testing signals) we increase a probability of phoneme recognition, generally speaking.

It is important to mention that both approaches to phoneme recognition do not exclude each other and are using combined in phoneme recognition problem solving.

References

1. L.R. Rabiner "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. of the IEEE*, vol. 77, No. 2, pp. 257-286, Feb. 1989.
2. L.R. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*. PTR Prentice Hall Inc., NJ, 1993.
3. H.A. Bourland and N. Morgan, *Connectionist Speech Recognition: A Hybrid Approach*. Kluwer Academic Publishers, Boston M.A., 1994.
4. E.I. Bovbel, P.P.Tkachova and I.E. Kheidorov, "Autoregressive hidden Markov models for isolated words recognition", in *Recent Advances in Information Science and Technology*, World Scientific: Singapore etc., 1998, pp.211-214.
5. *Artificial Neural Networks. Concept and Theory* / Compiled by P. Mehra, B.W. Wah. IEEE Computer Society Press, 1992.
6. B.A. Pearlmutter, "Gradient calculations for dynamic recurrent neural networks: A Survey", IEEE Trans. On Neural Networks, vol. 6, No. 5, pp. 1212-1228, Sept. 1995.
7. Yu.I. Zhuravlev, "Algebraic methods for the construction of the recognition and forecasting algorithms", *Proc. 5th Open German-Russian Workshop on Pattern Recognition and Image Understanding,* 21-25 Sept., 1998, Herrshing, Germany.
8. V.W. Zue, "The use of speech knowledge in automatic speech recognition". *Proc. of IEEE*, vol. 73, No. 11, Nov. 1985.
9. N. Wiener, *Nonlinear Problems in Random Theory*. New York: John Wiley and Sons Inc., 1958.
10. A.S. French and E.G. Butz, "Measuring the Wiener kernels of nonlinear system using the fast Fourier algorithm". Int. J.Control, No. 17, pp. 529-539, 1973.
11. A.M. Krot. *Discrete Models of Dynamic Systems Based on Polynomial Algebra*. Minsk: Nauka i tekhnika, 1990 (in Russian).

12. A.M. Krot and E.B. Minervina, "Identification and modeling of complex system based on series from the orthogonal Wiener-Volterra functionals", in *Advances in Synergetics*, vol. 6, pp. 184-190, 1995.
13. H.J. Nussbaumer, *Fast Fourier Transform and Convolution Algorithms*. Berlin, Springer-Verlag, 1982.
14. A.M. Krot and M.A. Shcherbakov, “Identification of discrete input nonlinear systems for digital chaotic signal processing”. *2nd IMACS International Conference on: Circuits, Systems and Computers (IMACS-CSC'98)*, vol. 2, 1998, pp. 795-797.
15. A.M. Krot, M.A. Shcherbakov and P.P Tkachova “Nonlinear filtering for solving the problem of variability in speech recognition”. *The 5th Open German-Russian Workshop on Pattern Recognition and Image Understanding*, 21-25 Sept., 1998. Herrsching, Germany.
16. P.Ya. Yurgelevich, *Course of Modern Belarussian Language with a Historical Comments*. Minsk: Vysh. Shkola, 1974 (in Belarusian).