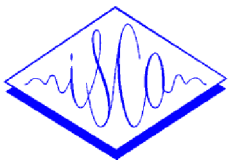

Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 1999)
Firenze, Italy, September 1-3, 1999

THE USE OF NONLINEAR FILTERING FOR THE PROBLEM OF VOICE VARIABILITY SOLVING*

Polina P. Tkachova[a] and Alexander M. Krot[b]
[a]Belarusian State University
4, Skoriny av., 220050, Minsk, Belarus
[b]Institute of Engineering Cybernetics of the National Academy of Sciences of Belarus
6, Surganov Str., 220012, Minsk, Belarus
alxkrot@newman.basnet.minsk.by
Tel.: (375 17) 284 20 86, Fax: (375 17) 231 84 03

Abstract. The approach to the problem of voice variability solving based on nonlinear Volterra-Wiener filtering is proposed.

Keywords: the problem of voice variability, preliminary processing, nonlinear filtering, Volterra-Wiener series.

Introduction

Due to acoustic phrase realization variability generated by various sources phonetic recognition by computer gives only limited results [1], [2]. At first acoustic changeability arises because of changing acoustic situation as well as because of distinguishing spatial orientation and electroacoustic transducer characteristics. Secondly speech signal changeability of the same speaker can arise owing to his psycho-physiological state, speech rate and vocal track state. Thirdly social and linguistic character differences, influence of dialect, size and form of vocal track can result in speech signal changeability from speaker to speaker [2].

Some of these variations influence insignificantly audio articulation when the other can result in very negative consequences [1], [2], [3].

Successful phonetic recognition depends highly on ability of variability controlling. One should not only extract and use information from phonetic variation in the recognition process but it is also necessary to learn to eliminate or depress the acoustic variation not carrying useful information.

Signal presentation based on modeling of aural impression

In last years in choosing design ways of a number of speech recognition systems knowledge of hearing man aid began to go long. An ear has already been noted to carry out peculiar frequency analysis of incoming signal. The form of auditory filters even can be characterized with some reliability. A concept of "critical bands" of audition originally dates from psycho-physiological investigation results [4]. Afterwards its acceptability were confirmed in general by neuro-physiologists studying responses of nerve fibers on tone sounds and combination of pure tones [2], [5]-[8].

The presentation based on modeling of audition retains and underlines signal characters playing substantial role for phonetic recognition [9], [10], [11]. Appropriate using of audition modeling results can really reduce to appearance of more reliable recognition system [2], [6], [10] and [11].

Generally the contemporary speech recognition strategies are based on spectral speech signal presentation derived in many ways from speech production models. These models are enough good for constructing analysis and synthesis systems of speech. The given systems are intended for more precise reconstruction of signal under investigation. Spectral presentation form as digital spectrogram used in synthesis is often considered as being suitable for recognition. However really these problems are completely different [2]. It is evident that human brain recognizes a signal only after preliminary processing by hearing organs. On the basis of distal auditory system investigation information on level of eighth cranial nerve is not considered to be precise copy of amplitude spectrum logarithm [2].

On the basis of above-mentioned and spectral presentation methods suitable fully only in cases of linear stationary signal filtering it is worth while to using nonlinear analysis methods based on functional series of Volterra-Wiener [6], [10], [11], [12] for auditory filtering process

* This work partly was supported by the International Science and technology Center (ISTC) under project B-95 and partly by University of Firenze, Italy

modeling. The signal processing in distal auditory system is shown by a number of researcher [7], [8] to describe by synchronous filtering in auditory nerve with response on stimulus as speech-like vocalic. In connection with this in the present paper method of such the system identification will be proposed for the purpose of constructing adaptive nonlinear model of auditory filter.

Thus the purpose of the present paper is to development of effective digital identification method of the nonlinear system modeling filtering process in distal auditory system for solving the problem of variability in speech signal recognition [9], [11].

Some experiment results concerning human ear perception of voice

The last world wide psychological and physiological experiments provide new information about the principals of human perception at time-frequency domain. It was shown that absolute auditory threshold can be approximated by nonlinear function

$$T_q(f) = 3{,}64\left(\frac{f}{1000}\right)^{-0{,}8} - 6{,}5e^{-0{,}6\left(\frac{f}{1000}-3{,}5\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4 \ (Db).$$

It is known that a band inside which human ear can not distinguish close frequencies, is a "critical" one. Its boundaries can be defined according to

$$BW_c(f) = 25 + 75\left[1 + 1{,}4\left(\frac{f}{1000}\right)^2\right]^{0{,}69} \ (Hz).$$

In practice there are several widely used approximations of critical band, in particular, Bark-scale

$$Bark = 13\arctan(0{,}00076f) + 3{,}5\arctan\left[\left(\frac{f}{7500}\right)^2\right] \ (Bark)$$

and Mel-scale

$$Mel = 2595\log 10\left(1 + \frac{f}{700}\right) \ (Mel).$$

Main frequencies and band widths for ideal filter bank in Bark and Mel scales can be found in Table 1.

However, the absence a single methodological approach to account of all peculiarities of human ear perception of speech makes impossible the problem of voice variability solving. In connection with this general approach to solving this problem will be developed.

The problem of voice variability solving depends strongly on how we can classify optimally the sounds of human voice (or phoneme of speech). The art of classification connects with peculiarities of considered language (see, for example, the Table 2) and, therefore, is very important for successful solving the problem of voice variability.

Within the framework of the proposed approach such classification means the appropriate choice of a nonlinear Volterra-Wiener filter corresponding to each group of phonemes (then each Volterra-Wiener functional corresponds to certain phoneme, see Table 2).

Identification of discrete nonlinear systems based on Volterra-Wiener functionals

The problem of finding the functional relationship that determines the output of a system in response to any relevant input is known as the problem of system identification or system characterization.

Wiener [13] considered the class of causal systems that produce an output with finite mean-square value when their input is Gaussian white noise. The output y(t) of an unknown nonlinear "black-box" system can be approximated by a series of functionals $G_m[h_m, x(t)]$ as

$$y(t) = \sum_{m=0}^{M} G_m[h_m, x(t)]. \qquad (1)$$

where x(t) is an input white Gaussian process with the zero mean value.

Several methods have been presented to find the Wiener kernels of nonlinear systems from given input and output pairs. Lee and Shetzen [14] showed that the kernels can be estimated by input-output crosscorrelation. French and Butz [15] used the fast Fourier transform (FFT) and Walsh transform algorithms for the Wiener kernels calculation. However, there are some difficulties in these methods:

a) A white Gaussian process is unrealizable;
b) Formula for kernels, m≥2, contains Dirac delta functions, when two or more kernel's arguments are equal;
c) The necessary computations increase very rapidly depending on the order of Wiener kernels that are being calculated.

This paper will resolve these problems by investigating discrete systems with discrete inputs generated using FFT algorithm. We will

construct the *G*-functionals for such inputs, and the formula for Wiener kernels and efficient identification algorithm will be presented in frequency domain [10], [12]. For more efficient computation of Fourier coefficients the split-radix FFT algorithms will be used [6], [16], [17].

Forcing functions for nonlinear systems testing

It is necessary that the stimulant system must, on the one hand, be like a random noise to get maximum information about unknown system and, on the other hand, to simplify the G-functionals and procedure of identification on the whole [10].
Taking into consideration these circumstances, select as test input periodic noise approximation

$$x(n) = \sum_{k=0}^{N_x-1} X(k)\exp(j2\pi kn/N_x). \qquad (2)$$

Here $X(k)=A(k)\exp(j\varphi(k))$ are complex Fourier coefficients where the amplitudes A(k) determinate the power spectrum of input and phases $\varphi(k)$ are independent random values with uniform distribution.
For real zero mean signal, the complex valued Fourier coefficients have the following relationship $X(0)=0$, $X(N_x-k)=X^*(k)$. According to Central Limit Theorem, the signal in the form of (2), being a sum of independent random quantities, has a nearly Gaussian distribution for large N_x. For every set $\varphi_l(k)$ random phases, formula (2) determinates the sequence $x_l(n)$ having N samples which may be formed by inverse FFT of coefficients $X_l(k)=A(k)\exp(j\varphi_l(k))$. This inverse FFT (IFFT) is calculated by split-radix FFT for Hermitian symmetric sequences [17].
In order to joint the separate sequences into long random process the following way can be proposed. Every time the set $\varphi_l(k)$ random phases is formed by random permutation of the totality $2\pi i/N_x$, $i=0,...,N_x-1$. In this case for any sequence $x(0)=0$ and joint of the sequences may be reduce to change sign of the blocks of data so that the condition $x(1)>x(0)$ is satisfied.

G-functionals and Wiener kernels in frequency domain

According to method of generation of the test signal the l-th sequence $x_l(n)$ of the random process x(n) is determinated by the set of input Fourier coefficients $X_l(k)$, $k=0,...,N_x-1$, and the corresponding response $y_l(n)$ can be characterized by the set of output Fourier coefficients $Y_l(k)$, $k=0,...,N_y-1$. Therefore, it is possible to rewrite input-output relationship for nonlinear systems in frequency domain as

$$Y_M(k) = \sum_{m=0}^{M} G_m[H_m, X(k)], \qquad (3)$$

where $H_m(k_1,..., k_m)$ is multidimensional discrete Fourier transformation (DFT) of the Wiener kernel $h_m(n_1,..., n_m)$.
Using a Gram-Schmidt orthogonalization procedure, the *G*-functionals can be shown as [10]

$$G_m[H_m, X(k)] = \sum_{\Omega_m} H_m(k_1,...,k_m)\delta^k_{k_1+...+k_m} \prod_{i=1}^{m} X(k_i), \qquad (4)$$

where the summation must extend over the m-D region Ω_m consisting various combination $(k_1,...,k_m)$ from integers $\{0,1,...,N_x-1\}$, such that $k_1\geq k_2\geq...\geq k_m$, $k_i\neq k_j$, and δ^j_i is a Kronecker delta.
The Wiener kernels in frequency domain for model (3) can be determined by minimization the mean square error between the DFT of the system output Y(k) and $Y_M(k)$ - DFT of model response [18]:

$$F = E\{\Delta^T\Delta^*\} \rightarrow \min,$$

where $\Delta=[\delta_1,...,\delta_N]$ is the vector of the complex errors having elements $\delta_k=(Y(k)-Y_M(k))$, and $E\{\cdot\}$ denotes the average operation.
Minimizing this function, the optimal Wiener kernels become [10]

$$H(k_1,...,k_m) = E\{Y(k_1,...,k_m)\cdot\times$$

$$\times \prod_{i=1}^{m} X^*(k_i) \Big/ \prod_{i=1}^{m} A^2(k_i)\}. \qquad (5).$$

In order to construct the estimation of kernel $H_m(k_1,...,k_m)$ which will be suitable in practice let us introduce the periodogram

$$I^l_{y_{x...x}}(k_1,...,k_m) = Y_l(k_1,...,k_m)\exp[-j\sum_{i=1}^{m}\varphi_l(k_i)]. \qquad (6)$$

Then as the estimation of kernel $H_m(k_1,...,k_m)$ may be use

$$\overline{H}_m(k_1,...,k_m) =$$

$$= \sum_{l=1}^{L} I^l_{y_{x...x}}(k_1,...,k_m) \Big/ L\prod_{i=1}^{m} A(k_i) \qquad (7)$$

The statistical properties of the estimation have been investigated in [18]. In particular it has been proved that estimation (7) is unbiased and its variance goes to zero if $L \to \infty$.

Algorithm of identification

Taking into account that the random phases is formed by random permutation of the values $2\pi i/N_x$, $i=0,...,N_x-1$, the equation (6) for periodogram can be rewritten as [10]

$$I^l_{y_{x..x}}(k_1,...,k_m)=Y_l(k_1,...,k_m)\times$$

$$\times exp[-j(2\pi/N_x)\{s^l_{k_1}+...+s^l_{k_m}\} mod N_x]$$

where s_k^l is l-th set of random integers obtained by permutation of the totality $\{0,...,N_x-1\}$.
The calculation of Wiener kernels may be performed more effectively if it is noted that periodogram (8) may take limited number of values $Y_l(k)\exp[-j(2\pi i/N_x)]$, $k=0,...,N_y-1$, $i=0,...,N_x-1$. This allows us in advance to form the array of possible products for every DFT $Y_l(k)$. Thus the algorithm of identification consists of the following steps [10], [11]:

1) Generation of random integers $s^l_1,...,s^l_{N_x}$ by means of permutation of set $\{0,...,N_x-1\}$ and forming the complex Fourier coefficients

$$X(k)=\begin{cases} A(k)\exp(j2\pi s^l_k/N_x), k=0,...,N-1; \\ 0, \quad k=N,...,N_x-N-1; \\ A(N_x-k)\exp(-j2\pi s^l_k/N_x), k=1,...,N. \end{cases}$$

2) Calculation of the l-th block of input by means of the inverse FFT based on the Hermitian-valued split-radix FFT algorithm [6], [16], [17]:

$$x_l(n)=IFFT\{X_l(k)\}, n=0,..., N_x-1.$$

3) Stimulation of the system by the input $x_l(n)$ and registration of the response $y_l(n)$.
4) Calculation of the complex Fourier coefficients by means of the real-valued split-radix FFT algorithm:

$$Y_l(k) = FFT\{y_l(n)\}, k=0,...,N_y-1.$$

5) Definition of the array $Z_l(k,i)$ of all possible values of the periodograms

$$Z_l(k,i)=Y_l(k)\exp(-j2\pi i/N_x), k=0,...,N_y-1,$$
$$i=0,...,N_x-1.$$

6) Forming from the array $Z_l(k,i)$ periodograms

$$I^l_{Y_{x...x}}(k_1,...,k_m)=Z_l(k_1+...+k_m, \{s^l_{k_1}+\cdot$$

$$+...+s^l_{k_m}\} mod N_x), m=1,...,M$$

7) Calculation of the estimation of kernels using eqn. (7).

Since the number C_m of combination $k_1,...,k_m$ containing in the kernel definition region Ω_m increases rapidly with a growth of kernel order m and the number of multiplications, demanding for calculation of array $Z_l(k,i)$ of all possible values of periodograms, do not depend on m, proposed algorithm as compared with methods [14], [15] allows to decrease the number of multiplications to a marked degree.
Actually, the most effective method [15] demands approximately $L(C_1+2C_2+...+mC_m)$ complex multiplications just as proposed method requires only $LN_xN_y/2$ multiplications.

Conclusion

An algorithm for identification of 1-D discrete nonlinear systems [10] (as well as 2-D ones [12]) in terms of the orthogonal series was presented. The process generated by inverse FFT algorithm was used as a test signal. For this input the G-functionals and Wiener kernels was defined in frequency domain. The proposed algorithm offers a significant reduction in computational complexity compared with the known methods as a number of multiplications do not depend on kernel order. The additional reduction on a number of arithmetic operations is achieved by the use of the Hermitian and real-valued split-radix FFT and fast polynomial transform algorithms. These results are used for construction of algorithms and software of digital nonlinear signal filtering for solving the problem of variability in speech recognition [11].

References

1. D.H. Klatt, "The problem of variability in speech recognition and models of speech perceptron", in Variability and Invariance in Speech Processes, J.S. Perkell and D.H.Klatt, Eds. Hillsdale, NJ: Lawrance Erlbaum Assoc., 1985.
2. V.W. Zue, "The use of speech knowledge in automatic speech recognition". Proc. of the IEEE, vol. 73, no. 11, November 1985.
3. G.R. Doddington, "Speaker Recognition - Identifying People by their voices". Proc. of the IEEE, vol. 73, no. 11, November 1985.

4. H. Fletcher, "Auditory patterns", Rev. Mod. Phys., vol. 12, pp. 47-65, 1940.
5. N.Y.-S. Kiang, T. Watanabe, E.C. Thomas and L.F. Clark, Discharge patterns of single fibers in the Cat's auditory nerve, Research Monograph No 35. Cambridge, MA: M.I.T. Press, 1965.
6. A.M. Krot, E.B. Minervina, "Identification and modelling of complex system based on series from the orthogonal Wiener-Volterra functionals". IVth Annual Seminar "Nonlinear Phenomena in complex systems". Minsk, Febr., 1995, pp. 153-158.
7. R. Lyon, "A computational model of filtering, detection, and compression in the cochlea", in Proc. ICASSP-79, pp. 1282-1285, 1982.
8. S. Seneff, Pitch and spectral analysis of speech based on an auditory synchrony model, Ph.D. dissertation, Mass. Inst. Technol. Cambridge, MA, 1985.
9. E.I. Bovbel, P.P. Tkachova, 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.
10. A.M. Krot and M.A. Shcherbakov, "Identification of discrete input nonlinear systems for digital chaotic signal processing", 2nd IMACS International Conference on: Circuis, Systems and Computers (IMACS-CSC'98), vol.2, 1998, pp. 795-797.
11. 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 September 1998, Herrshing, Germany.
12. A.M. Krot and M.A. Shcherbakov, "Identification of discrete 1-D and 2-D input systems for digital signal and image processing", Proc. of the 6th IEEE Workshop on Intelligent Signal Processing and Communication Systems, November 5-6, 1998, Melbourne, Australia, pp. 881-885.
13. N. Wiener, Nonlinear problems in random theory. MIT press, 1958.
14. M. Schetzen, "Nonlinear system modelling based on the Wiener theory". Proc. IEEE, vol 69, pp. 1557-1567, 1981.
15. A.S. French, E.G. Butz, "Measuring the Wiener kernels of nonlinear system using the fast Fourier algorithm". Int. J.Control, vol. 17, pp. 529-539, 1973.
16. A.M. Krot, Discrete models of dynamic systems based on polynomial algebra. Minsk: Nauka i tekhnika, 1990 (the monograph in Russian).
17. A.M. Krot, E.B. Minervina, "Synthesis of FFT split-radix algorithms for real-valued and Hermite-symmetrical series". Radioelectronics & Communication Systems, vol. 12, pp. 10-15, 1989 (a translation from Izv. VUZ, Radioelektronika, vol.32, No.12, pp.12-17, 1989).
18. M.A. Shcherbakov, The digital polynomial filtering: theory and applications. Penza: State Technical University of Penza Publ., 1997 (the monograph in Russian).

Table 1. Main frequency and band width in Bark and Mel scales

Channel No.	Bark-scale		Mel-scale	
	Main frequency, Hz	Band width, Hz	Main frequency, Hz	Band width, Hz
1	50	100	100	100
2	150	100	200	100
3	250	100	300	100
4	350	100	400	100
5	450	110	500	100
6	570	120	600	100
7	700	140	700	100
8	840	150	800	100
9	1000	160	900	100
10	1170	190	1000	124
11	1370	210	1149	160
12	1600	240	1320	184
13	1850	280	1516	211
14	2150	320	1741	242
15	2500	380	2000	278
16	2900	450	2297	320
17	3400	550	2639	367
18	4000	700	3031	422
19	4800	900	2482	484
20	5800	1100	4000	556
21	7000	1300	4595	639

Table 2. Classification of belarusian language consonants

Contribution of voice and noise	Formation place	Formation method	Labial				Lingua							
			Labial-labial		Labial-dental		Front				Middle		Back	
							dental		alveolar					
			hard	media	hard	media	hd	m	hd	m	hd	m	hd	m
Noise	Stop	voiced		'	-	-		-	-	-	-	-	( )	(')
		surd		'	-	-		-	-	-	-	-	( )	(')
	Fricative	voiced	-	-	-	-		'		-	-	-	γ	(γ')
		surd	-	-		'		'						(')
	Affricates	voiced	-	-	-	-	(Z)	Z'		-	-	-	-	-
		surd	-	-	-	-		'		-	-	-	-	-
Sonorous				'	-	-		'		-	-	j( )	-	-
			,	'				'	-	-	-	-	-	-