Peculiarities of Wiener Class Systems and their Exploitation for Speech Signal Prediction

E. Ivanovas, D. Navakauskas

Department of Electronic Systems, Vilnius Gediminas Technical University,
Naugarduko str. 41-413, LT-03227, Vilnius, Lithuania, phone: +370 612 03253, e-mail: edgaras.ivanovas@el.vgtu.lt

Introduction

Speech technologies are widely employed in modern telecommunication systems. Received speech signal quality [1], spoken word recognition accuracy [2] and successful speaker identification correctness [3, 4] strongly rely on speech modelling precision. Even modern audio coders are based on linear models that are used to simulate speech, which in general is non-linear and non-stationary. Therefore, non-perfect modelling occurs.

As majority of real world systems are not linear in nature, nonlinear methods for their modelling are required and application range of those models is very wide. Therefore, in a research literature of a last decade Wiener class systems (see Fig. 1), defined as linear time invariant, dynamic, causal and stable subsystem followed by static nonlinearity, have drawn considerable attention. Several possible parametric and non-parametric identification procedures of Wiener class system were presented quite recently [5, 6, 7].

In this paper we apply Wiener class system consisting of linear dynamic subsystem coupled with a sigmoid nonlinearity to speech modelling. Initial experiments are performed on a Lithuanian vowel "ū". Basing on the results and observations made, we extend our experiments to other vowels taken from different words. The results are presented, where the performance of Wiener system model is opposed to widely employed Linear Prediction Coding (LPC) model. Basing on the found properties Wiener system model is proposed to be used for speech prediction. Presented experimental simulations demonstrate increase of accuracy at selected signal windows thus confirming proposed approach. Application of results in telecommunication systems could possibly decrease data transfer rates, increase speech intelligibility and naturalness of speech synthesis.
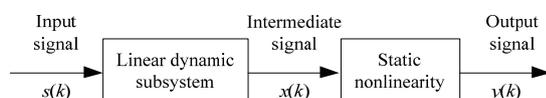

Fig. 1. Wiener class system representation

Setup for speech signal prediction

The Wiener system model we use consists of infinite impulse response dynamic subsystem in the linear part and a sigmoid function in the nonlinear part. The intermediate signal of Wiener system model is expressed as

$$x(k) = \frac{b_0 + b_1 q^{-1} + b_2 q^{-2} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n}} \cdot s(k), \quad (1)$$

here $s(k)$ is an input signal; $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ are dynamic system coefficients; q is time shift operator; n is order of the dynamic system. The used in Wiener system model nonlinearity is defined by

$$f(x(k); \alpha, \beta) = \frac{1 - \exp(\alpha(\beta - x(k)))}{1 + \exp(\alpha(\beta - x(k)))} \equiv y(k), \quad (2)$$

here $y(k)$ is output signal; $f(\cdot)$ is monotonous and invertible sigmoid function; α and β are gain and bias of sigmoid function, respectively. The sigmoid function dependence on these parameters is demonstrated in Fig. 2.

In order to find parameters of the Wiener system model applied to predict signal, a fragment of it is used to generate the training dataset. The output signal is advanced in time via l samples in respect to input signal.
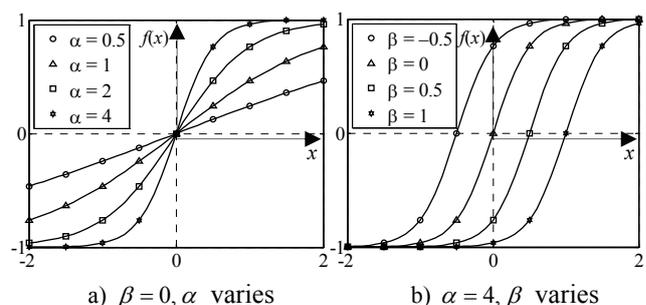

a) $\beta = 0, \alpha$ varies   b) $\alpha = 4, \beta$ varies

Fig. 2. Nonlinear function (2) dependence on values of α and β parameters

Identification of the parameters is done minimizing Mean Square Error (MSE) between the actual advanced signal $s(k+l)$ and the estimated value $\hat{y}(k)$. Parameter optimization is done using Levenberg-Marquardt algorithm. It is iterative procedure and takes much more calculations for identification. However, it can be implemented in field programmable gate arrays that can be used for acceleration of parameter identification [8].

Speech data. The experiments are performed using 16 bit mono signal quality with 11.025 kHz sampling frequency audio records of Lithuanian words "kultūra" and "įmonė". The speech is produced by an adult male Lithuanian speaker. The records are done in a non-noisy environment.

Model initialization. 14th order of the LPC model was chosen for the comparison purposes. The same order was chosen for the linear dynamic part of Wiener system model. The initialization of linear dynamic system coefficients can be done randomly; however, here the LPC coefficients were used. As the bias in signals was removed, we set β to 0 as its initial value. Parameter α is set to value 2, for tangent to be at 45 degrees at point 0.

Investigation of single vowel prediction

The first group of experiments is devoted to see the influence of training procedure on future signal error. For this reason the MSE of learning, testing and validation datasets are observed during the training procedure. The length of the datasets is 256 samples. The validation dataset window is shifted by 256 samples and testing dataset window is shifted by 512 samples to the future in respect to training dataset window (see Fig. 3).

Ten experiments were performed on vowel "ū" from the word "kultūra". The dataset windows were shifted 10 times by 10 samples more in respect to each other. The achieved prediction results are summarized in the Table 1. Moreover, in Fig. 4 there are shown two cases of model identification process that result: a) in Wiener system model perfection on the validation dataset (Experiment 2); b) in LPC model perfection on the training dataset (Experiment 9).
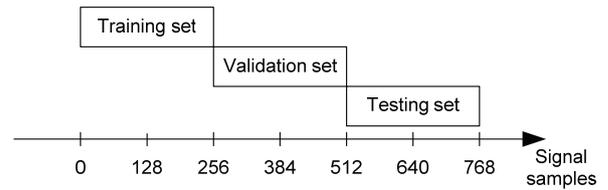

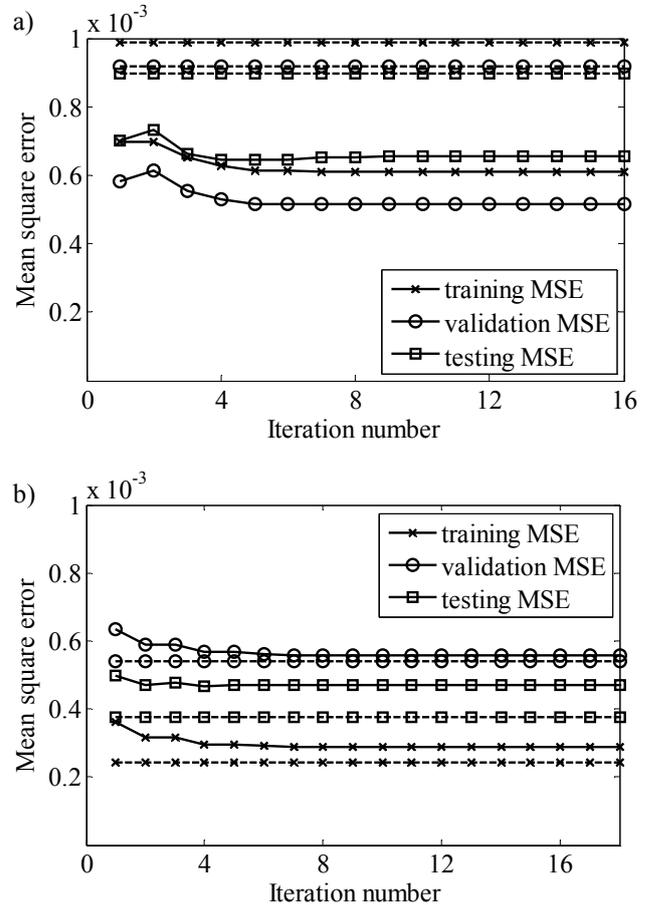**Fig. 3.** The organization of training, validation and testing datasets

Fig. 4. Model identification process mean square errors during: a) Experiment 2; b) Experiment 9. Notations: solid lines with ticks – WS model; dashed lines with ticks – LPC model

Table 1. Vowel "ū" prediction results comparing MSE of LPC and Wiener system models on training, validation and testing data

Experiment number	MSE of LPC model, 10^{-4}			MSE of Wiener system model, 10^{-4}			MSE ratio, %		
	Training dataset	Validation dataset	Testing dataset	Training dataset	Validation dataset	Testing dataset	Training dataset	Validation dataset	Testing dataset
1*	3.1580	**3.5035**	5.2116	3.4889	3.8889	5.3594	110.47	**111.00**	102.83
2	9.8749	9.1845	8.9524	6.0930	5.1494	6.5563	61.70	**56.06**	73.23
3	3.4636	4.1196	6.3259	3.1540	**3.7607**	5.1238	91.06	91.29	80.99
4	10.9890	10.2655	9.3263	7.2881	6.1405	7.7170	66.32	59.81	82.74
5	3.7918	4.9780	7.4858	3.4041	4.4351	5.8389	89.77	89.09	78.00
6	12.3711	**12.2119**	9.0802	8.0191	**7.0434**	**8.3597**	64.82	57.67	92.07
7	3.7408	5.1101	5.9357	3.6890	5.1950	5.6210	98.61	101.66	94.70
8	15.1792	12.0148	**11.2340**	9.6602	7.0248	6.5802	63.64	58.46	**58.57**
9*	**2.4413**	5.4082	**3.7638**	**2.8945**	5.5910	**4.7122**	118.57	103.38	**125.19**
10	**17.3411**	5.9460	7.9325	8.3421	6.0088	5.6714	**48.11**	101.06	71.50
Average	8.2351	7.2742	7.5248	5.6033	5.4238	6.1540	81.31	82.95	85.98
PSD**	5.2969	3.1436	2.1234	2.4327	1.1089	1.09724	22.36	21.18	17.77

* – two best prediction results; ** – population standard deviation; data in bold – extreme values

It is worth to notice, that the smallest errors of LPC model (Experiments 1 and 9) are never outperformed by Wiener system model. This peculiarity suggests an idea that chosen nonlinearity do not suites well for vowel "ū" signal and no reduction of MSE for the training dataset can be possible, as linear dynamic subsystem will always have MSE smaller or equal to the LPC model. To confirm this, a further investigation should be performed on the choice of nonlinearity depending on the initial signal phase in a window.

The results in Table 1 seem to be strongly dependent on the initial phase of the speech signal to be predicted. Comparing means and population standard deviations of both models confirm, that the Wiener system model demonstrates more consistent results on testing dataset – it outperforms LPC model in 8 experiments. Thus the mean of MSE ratio (LPC vs Wiener system models) for the Wiener system model are smaller by approximately 18 % (training dataset), 17 % (validation dataset) and 14 % (testing dataset).

Investigation of multiple vowel prediction

The second group of experiments is devoted to investigate if the use of sigmoid nonlinearity can give advantage on other vowels. For this, prediction errors of LPC, Infinite Impulse Response (IIR) and Wiener system models in the signal windows shifted to future in respect to the training dataset window will be analysed.

The experiment was performed by identifying the parameters of the models from 256 sample size training dataset and employing them to predict a 1025 sample size signal window consisting of a same training dataset and 769 future samples, which give us an insight of prediction accuracy of 70 ms of the future signal. Six experiments were done where vowels from the Lithuanian words "kultūra" and "įmonė" were used. To exploit the peculiarity the MSE on training dataset of Wiener system model was compared with the MSE of IIR model. In case the latter was smaller or equal to the one of the Wiener system model, the dataset windows are shifted by 10 samples. The result of each experiment is 1024 sample size error vector, which was divided to four equal parts and MSE value was calculated for each of the quarter. The resulting data can be seen in Table 2.

The first quarter is actually the data the models were trained on. Because of this and additional 14 parameters
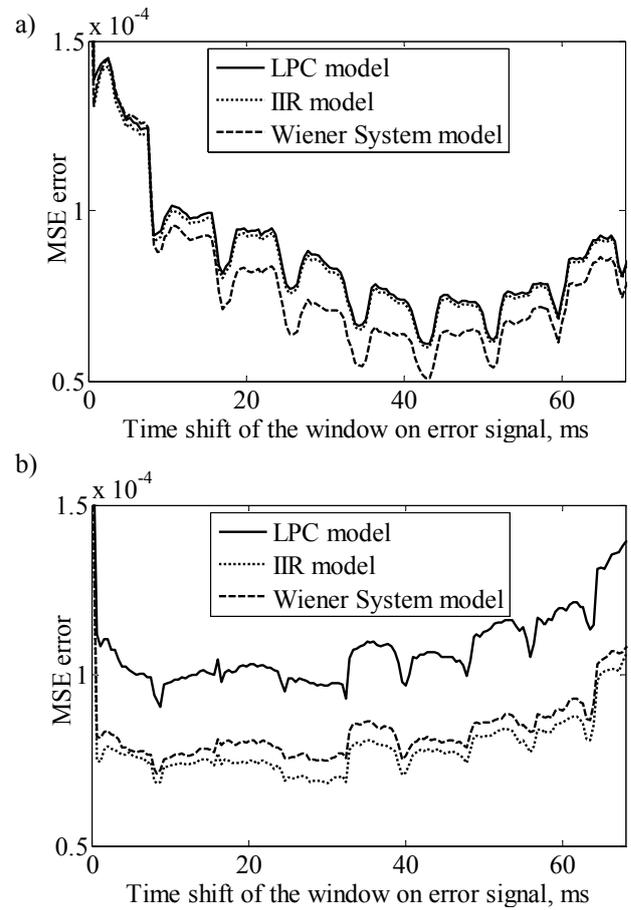

a)

b)

Fig. 5. Influence of MSE calculated from 256 samples on time shift of the window: a) vowel "u"; b) vowel "į"

the results of IIR model has to be better than the ones of LPC model. Meanwhile, the smaller MSE of the Wiener system model were guaranteed by the constraint of the experiment (see data in bold).

The second, third and fourth quarters' MSE shows how well the models perform on future data that have not been seen during the training procedure. The ratios of MSE between LPC and IIR models are very close to 1 in all vowels except "į". That could mean that the additional parameters of a linear subsystem are not very useful for the prediction of future signal.

The Wiener system model performs better than the IIR model predicting future samples of all tested vowels with an exception of "į", too. An additional shift of signal

Table 2. Six vowel prediction results comparing MSE of IIR, LPC and Wiener system models by the quarters of error vectors

Vowel	IIR model MSE / LPC model MSE				Wiener system model MSE / LPC model MSE			
	1st quarter of dataset	2nd quarter of dataset	3rd quarter of dataset	4th quarter of dataset	1st quarter of dataset	2nd quarter of dataset	3rd quarter of dataset	4th quarter of dataset
"u"	0.7976	0.9845	0.9846	0.9917	**0.7740**	**0.8840**	**0.8577**	**0.9178**
"ū"	0.7806	0.9604	0.9788	0.9778	**0.7048**	**0.9297**	**0.9079**	**0.8906**
"a"	0.8483	1.0001	0.9977	0.9991	**0.7830**	**0.9454**	**0.9650**	**0.9949**
"į"	0.8718	**0.7263**	**0.7361**	**0.7569**	0.8269	0.7836	0.7690	0.7710
"o"	0.4290	0.9557	0.9950	1.0032	**0.3938**	**0.8817**	**0.8799**	**0.8491**
"ė"	0.9998	0.9742	0.9935	0.9953	**0.9470**	**0.8702**	**0.8293**	**0.8172**
Average	0.7879	0.9335	0.9476	0.9540	0.7383	0.8824	0.8681	0.8734
PSD*	0.1754	0.0939	0.0948	0.0885	0.1705	0.0518	0.0613	0.0722

* – population standard deviation; data in bold – minimum values comparing corresponding results of IIR vs. Wiener system models

window by 10 samples was tried. It gave smaller MSE ratios of Wiener system model than of IIR model and confirmed the dependence of the results on initial signal phase and not on the vowel. It seems that Wiener system model benefits from its sigmoid nonlinearity where additional parameters of IIR model gave only marginal improvement. The average MSE ratios of a Wiener system model are less about 5 %, 8 % and 8 % for the second, third and fourth quarters respectively.

In Fig. 5 the change of MSE is depicted shifting error calculation window in time. In Fig. 5a it can be clearly seen that the MSE of the Wiener system model is higher in comparison of both other models at a time region of about 7 ms. At that time, the models had to predict signal consisting of about 13 ms of training signal and 7 ms of future signal. The advantage of Wiener system model can be clearly seen starting from the shift about the 10 ms until the end of the data. In Fig. 5b the MSE of vowel "į" modelling is shown. The Wiener system model outperforms the LPC model. However, the IIR model gives better MSE during all the modelling of the signal. Thus, sigmoid nonlinearity does not give an improvement in this case and a different type of nonlinearity might be needed to archive that [5–7].

Conclusions

1. For vowel prediction LPC model could be replaced by Wiener system model. Experiments with six Lithuanian vowels on average gave more than 13 % improvement in MSE of prediction. The average of MSE for vowel "ū" prediction with the Wiener system model were smaller by approximately 18 % for the training dataset, 17 % for the validation dataset and 14 % for the testing dataset.

2. A nonlinearity employed in Wiener system model plays a major role on the performance of the system in comparison with LPC and IIR models. Badly chosen nonlinearity can limit short-term prediction capabilities as confirm results of two experiments with vowel "ū" or

impact long-term prediction as show experimental results with vowel "į".

3. The use of Wiener system model instead of IIR model improves MSE of long-term prediction of five Lithuanian vowels on average by more than 5 %.

A further investigation of Wiener system model should be performed on the choice of nonlinearity taking into account the initial speech signal phase to be predicted.

References

1. **Anskaitis A., Kajackas A.** The Tool for Quality Estimation of Short Voice Segments // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 8(104). – P. 103–108.
2. **Tamulevičius B., Arminas J., Ivanovas E., Navakauskas D.** Hardware Accelerated FPGA Implementation of Lithuanian Isolated Word Recognition System // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 3(99). – P. 57–62.
3. **Šalna B., Kamarauskas J.** Biometric Evaluation of Effectiveness of Different Methods in Speaker Recognition // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 2(98). – P. 67–70.
4. **Ivanovas E., Navakauskas D.** Development of Biometric Systems for Person Recognition: Biometric Feature Systems, Traits and Acquisition // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 5(101). – P. 87–90.
5. **Raich R., Zhou T., Viberg M.** Subspace Based Approaches for Wiener System Identification // IEEE Transactions on automatic control, 2005. – No. 10(50). – P. 1629–1634.
6. **Hu, X. L., Chen, H. F.** Strong Consistence of Recursive Identification for Wiener Systems // Automatica, 2005. – No. 11(41). – P. 1906–1916.
7. **D'Amato A., Teixeira B., Bernstein D.** Semiparametric Identification of Wiener Systems Using a Single Harmonic Input and Retrospective Cost // American Control Conference (ACC), 2010. – P. 4498–4503.
8. **Laptik R., Arminas V., Navakauskas D.** Ant System Implementation using Microblaze: Some Preliminary Results on Efficiency Study // Electronics and Electrical Engineering. – Kaunas: Technologija, 2009. – No. 6(94). – P. 27–30.

E. Ivanovas, D. Navakauskas. Peculiarities of Wiener Class Systems and their Exploitation for Speech Signal Prediction // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 5(111). – P. 107–110.

In this paper Wiener class system consisting of infinite impulse response dynamic subsystem and sigmoid nonlinearity is employed for speech modelling. The experiments on Lithuanian vowel "ū" with different initial phase of the signal are done. The average of mean square errors of prediction with the Wiener system model are smaller approximately by 18 % for the training dataset, 17 % for the validation dataset and 14 % for the testing dataset in comparison to Linear Prediction Coding model. Additional experiments with 6 different Lithuanian vowels on average showed more than 13 % improvement in MSE of prediction with the Wiener system model. Moreover, the use of Wiener system instead of IIR model showed improvement in MSE of long-term prediction for five Lithuanian vowels on average by more than 5 %. Application of results in telecommunication systems could possibly decrease data transfer rates, increase speech intelligibility and naturalness of speech synthesis. Ill. 5, bibl. 8, tabl. 2 (in English; abstracts in English and Lithuanian).

E. Ivanovas, D. Navakauskas. Vinerio klasės sistemų ypatumai ir jų taikymas šnekos signalams prognozuoti // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2011. – Nr. 5(111). – P. 107–110.

Straipsnyje nagrinėjamas Vinerio klasės sistemos, sudarytos iš begalinės impulsinės reakcijos dinaminės posistemės ir sigmoidinio netiesiškumo, taikymas lietuvių kalbai modeliuoti. Pristatyti skirtingų „ū" balsio pradinių signalo fazių tyrimai. Gauti rezultatai patvirtina Vinerio sistemos modelio pranašumą, palyginti su tiesinės prognozės kodavimo modeliu, – Vinerio sistemos modelio mokymo duomenų aibių vidutinė kvadratinė klaida yra mažesnė apie 18 %, patikros – 17 % ir testavimo – 14 %. Papildomais eksperimentais su šešiais balsiais parodoma, kad Vinerio sistemos modelio taikymas vidutiniškai sumažina prognozės vidutinę kvadratinę klaidą ne mažiau kaip 13 %. Be to, Vinerio sistemos modelio taikymas vietoj begalinės impulsinės reakcijos modelio vidutiniškai ne mažiau kaip 5 % sumažina penkių balsių ilgalaikės prognozės vidutinę kvadratinę klaidą. Gautus rezultatus pritaikius telekomunikacijų sistemose, potencialiai galima būtų sumažinti duomenų srautus, padidinti šnekos suprantamumą ir sintezuotos šnekos natūralumą. Il. 5, bibl. 8, lent. 2 (anglų kalba; santraukos anglų ir lietuvių k.).