# METHODS OF SPEECH SIGNAL PARAMETRIZATION BASED ON GENERALIZING OF LINEAR PREDICTION

A.N. Sobakin

Moscow, USSR

## ABSTRACT

The generalization of speech analysis method on the basis of linear prediction reveals unused potential possibilities of this method and permits to develope new algorithms of evaluating speech signal parameters.

## INTRODUCTION

Modern achievements in the sphere of speech analysis and synthesis are mainly connected with the use of algorithms of speech signal parametrization, that take into consideration in some degree the nature of speech production.

According to Fant's model [1], the speech production consists of excitation signal transformation by the linear dynamic system (LDS), which parameters correspond to the state of vocal tract at the moment of articulation.

The change in the vocal tract state during articulation leads to the LDS parameters modification.

The tracing of these changes is usually carried out by shifting analysis window within which the LDS parameters may be considered to be sufficiently stable. The transfer function of such LDS at the analysis interval has the form of fraction-rational function with zeroes and poles.

The signal at the LDS input is looked upon as a sequence of alternating intervals, corresponding to voice or noise excitation. The whole excitation signal in that case is modulated by the time envelope of the speech signal.

Linear prediction [2-6] as a method of speech signal analysis was worked out on the basis of much more simplified pattern of speech formation, than one described above.The method is based on deriving the LDS parameters according to the speech signal estimates, ignoring transfer zeroes within the analysis interval. The most simple calculation formulas are obtained in the metrical space.

The quality of obtained LDS parameters estimates will essentially depend on the location of the analysis window at the time axis.

If the interval of analysis corresponds either to an interval of noise excitation or to an interval of free LDS oscillations (for example, the interval of vocal cords closure) then it is possible to show, that in that case the estimates will be unbiassed.

But in case when the analysis interval contains one or several pitch impulses, LDS parameters estimates will be biassed. It is explained by the misagreement between the analysis method and the speech signal structure, for example at the voiced intervals of speech.

Thus, the problem of more complete agreement between the analysis method and the speech formation pattern is an urgent issue.

According to the said above, it seems perspective to examine possible linear prediction generalizations, introducing additional parameters and characteristics of the method. Additional degrees of freedom may be used for more complete agreement between the method of analysis and the speech signal structure.

The generalization of linear prediction leads to the algorithm modifications of speech signal parameters estimates, and, in the long run to obtaining new parametrical spaces for analysis and speech recognition.

## GENERALISATION OF LINEAR PREDICTION

The essence of linear prediction is nonrecursive p-order filter that transforms the speech signal counts [x] into residual signal e[n], using weight coefficients $\{A_k\}$:

$$e[n]= \sum_{k=0}^{p} A_k * y_k[n], \qquad (1)$$

$$A_0 =1, \qquad (2)$$

where $y_k[n]=T^k(x[n])=x[n-k], \quad k=0,1,\dots,p;$ (3)

$T^k(\ )$ k-power of the delay operator.

When analysing speech optimal coefficients of filter (I) $\bar{a}_{opt} =\{A_0, \dots,A_p\}$ are determined from condition of minimum residual signal deviation $\{e[n]\}_0^M$ from the coordinate beginning in the metric space $L_2$ within the analyses interval [0,M]:

$$\bar{a}_{opt} =\arg\min F(\bar{a}), \qquad (4)$$

$$\text{where}\quad F(\bar{a})= \sum_{n=0}^{M} e^2[n] \qquad (5)$$

-is a squared quality criterion.

Suggested linear prediction generalization concerns two filter components : extension of operator class , on the basis of which it is formed (3) and generalization of constraint imposed on it's coefficients (2).

Quality functional is also generalized (5), that allows to choose different metric spaces for estimation of analysis parameters.

As seen from (3),the original transformation (1) was formed on linear delay operators,that represent the class of physically realizable linear systems with constant parameters.Principally,it is possible to substitute the original delay operators for a set of any stable operators $U_0, U_1, \dots, U_p$ from the class indicated.

Then proportion (3) is transformed into corresponding cascade form as follows :

$$y_k[n]=U_k(y_{k-1}[n]), \qquad k=0,1,\dots,P; \qquad (6)$$

where $y_{-1}[n]=x[n]$.

Each linear operator (6) is determined in the frequency sphere by the transfer function of fraction - rational type.

According to the speech signal physical characteristics,the choice of transfer function parameters allows to change in necessary direction the structure and features of linear transformation (1).

Thanks to that,the agreement between algorithm analyses and dynamic speech characteristics will be achieved.

The cascade form (6) of transformation (1) also allows examine the corresponding generalized structures of lattice filters [7] on the basis of linear operators specifically chosen.

It is worth-while to note ,that besides cascade form (6) ,the parallel form of the speech signal preliminary transformations can be easily formed on the basis of indicated set of linear operators. Each of the output signals $y_k[n]$ is obtained as the result of application the corresponding operator directly to the input signal x[n].

The condition (2) influences the structure and features of filter (1) not to a lesser degree.

This limitation for parameters of the filter was introduced to eliminate zero solution during the search for quality functional minimum. In essence it can be considered as the constraint on vector $\bar{a}$ coordinate magnitude in (p+1)-dimentional space of parametres.

In general,this constraint may be written down as follows:

$$f(\bar{a})=f(A_0,A_1,\dots,A_p)=0 , \qquad (7)$$

where f()is an arbitrary function of (P+1) variable.

The only condition of choosing the function is zero solution elimination in the problem under consideration.Thus,the equation (7) in (P+1) space of parameters determines a surface,not passing through the coordinate beginning.

The search for an optimal vector of coefficients $\bar{a}_{opt}$ with constraint (7) may be realised on the basis of generalized quality functional $F_r(\bar{a},b)$:

$$F_r(\bar{a},b)= \sum_{n=0}^{M} |e(n)|^r + b * f(\bar{a}), \qquad (8)$$

where b - the Lagrange factor from the set of real numbers, R,r - an integer.

The value of "r" in (8) determines the choice of the metric space $L_2$, where the search for optimal vector of parameters $\bar{a}_{opt}$ is carried out.

Lagrange factor b increases by one the amount of target unknown values and reduces the problem of conditional extremum searching to the search for unconditional extremum for quality functional (8).

As before,condition (4) in which functional (8) was used instead of functional (5) ,determines vector $\bar{a}_{opt}$ and factor b in expanded (P+2)-dimentional metric space $L_{r,1} = L_r * R$ .

Proceeding from condition (4) of the quality functional minimum (8) ,the task of searching filter (1) parameters may be presented as generalization of linear prediction method.

The particulare choice of basic operators (6) of limiting function (7) and characteristical constant determines in each case different algorithms of speech signal analysis and different parametric spaces for their description.

## ON THE CHOICE OF BASIC LINEAR OPERATORS, LIMITING FUNCTION AND METRIC SPACE.

Among three components,that determine the particulare form of analysis algorithm in the formulated task,the most promissing and the most difficult at the same time is the problem of the best choice of basic linear operators (6).

As in classical methods of digital filters design [8], the complexity of this problem for the class of linear systems with infinite impulse responce (IIR-filters) is increasing as compared to the choise of linear operators from the class of linear systems with finite impulse response (FIR-filters).

Let's confine to setting a mathematical problem of choosing operators (6) from the FIR-system class with impulse responses of p-length. In that case the set of transformations (6) is represented by a linear equation system,that is formed with the help of square B matrix of (P+1)*(P+1) size.

B matrix lines are the impulse responses of the basic operators,derived from the above mentioned class of the FIR-systems.

In that case,the set of operators (6) in parallel form is expressed by delay operators (3),and the corresponding vectors of coefficients $\bar{a}$ and $\bar{c}$ for both variants are related to each other by the following linear equation system:

$$\bar{c}' = B' * \bar{a}' \qquad (9)$$

(an accent means transposition).

In case of B matrix inversion,parameters $\bar{a}$ and $\bar{c}$ are equivalent according to the information theory.

However,the latter doesn't mean their equivalence from the viewpoint of their optimum coding for speech transmission and recognition. Thus,the problem of the best choice of basic FIR-systems is formulated as the problem of transformation search (9) (i.e. B-matrix),that brings about the improvement of estimated parameters in the systems of speech transmission and speech recognition.The B matrix choice allows to take into account more completely the speech signal structure and features.

Using the linear operator theory in Gilbert spaces [9] it is possible to approximate any linear operator from the IIR -system class by a linear operator from the FIR-system class.The problem of the optimum choice of basic operators from the class of IIR-systems may be reduced to the above formulated task of FIR-systems .

It doesn't seem possible to examine different variants of condition (7) fuly enought. Let's confine ourselves to 2 types of function f(.).

For predicting methods,the choice of function f(.) in the form of scalar product of weight coefficients $\bar q$ by parameter vector $\bar a$ is a natural generalization of constraint (2):

$$f(\bar a)=(\bar q\ ,\ \bar a) - 1 = 0\ . \qquad (10)$$

Equation (10) with minus one in the left part determines a hyperplane in the space of parameters ,that doesn't pass through the coordinate beginning.

Interesting results are obtained if a square form of the parameter vector is taken as the second limiting function:

$$f(\bar a)=(D\bar a'\ ,\ \bar a) - 1 = 0. \qquad (11)$$

Equation (11) determines the second order plane in the parameter space with the help of D matrix of (P+1)*(P+1) size.

In both cases, the choice of either particulare vector $\bar q$ for condition (10) or D matrix for condition (11) gives aditional degrees of freedom, helping to determine the structure and features of the corresponding estimation algorithm of the speech signal parameters.

The choice of metric space $L_r$ ,i.e. the choice of characteristical number r ,also determines the structure and features of the obtained algorithms.

The most developed and examined algorithms are the estimation algorithms for squared quality criterion in metric space $L_2$ (r=2).

However, the results of theoretical calculations and experimental[10] researches show that modular criterion (r=1) has the advantages in the speech signal analysis .For example ,single excitation pulses don't distore the target values of the LDS parameters and the obtained parameter estimations are nonbiassed.

It seems interesting to examine the minimax quality criterion for r= ∞ and the obtained results of the speech signal investigation, though there arises the necessity to use complex Remez algorithm [11] for estimating parameters.

## THE EXAMPLES OF ANALYSIS ALGORITHMS

In practice the determinating of functional extremum may be carried out in two ways: either on the basis of the equation system that is derived when the quality functional gradient is equal to zero or by adaptive methods [12] in the form of consecutive approximations to target parameters.

The adaptive methods are of the most interest in the sphere of aplied researches.

The system of adaptive equations for determining the LDS parameter estimates in case when m-th coordinate of vector $\bar q$ is equal to one and other coordinates are equal to zero,will look as follows:

$$A_k[n+1]=A_k[n]-g(n)*e_1(n)*y_k[n],\ k=0,\dots,m-1,m+1,\dots,P; \qquad (12)$$

where g(n) is normalizing multiplier.

The equation system (12) reminds of the system of adaptive equations for linear prediction based on the method of the least squares [6].In fact when the first coefficient is equal to one [$a_0$ =1],the forward linear prediction is obtained ,when the last coefficient is equal to one ($a_p$ =1) the backward linear prediction is obtained.

Thus,there exists a principal possibility to work out filters [7] on the basis of generalized linear operators.

The adaptive algorithm obtained for a unitary D matrix will differ from other known methods of estimating in most degree.Condition (11) in that case will mean that the norm of coefficient vector is equal to one:

$$\|\bar a\| = 1. \qquad (13)$$

Equation (13) in parametrical space determines a spheric surface of an unitary radius,within which the search for quality functional extremum is carried out.

The corresponding adaptation equations look as follows:

$$A_k[n+1]=A_k[n]-g(n)\{e_1(n)*y_k[n]-b[n]*A_k[n-1]\}$$

$$b[n+1]=b[n]-g_1(n)*(\sum_{k=0}^{p} A_k^2[n-1] - 1\};\quad k=0,1,\dots,P. \qquad (14)$$

The estimation of coefficient vector, obtained on the basis of equations (13) and (14) is an approximated latent vector value of covariation signal matrix y[0],y[1],...,y[n] that correspond to maximum latent value of this matrix. This algorithm differs from the classical method of linear prediction.

Function e (n), used in equations (12) and (14), is identically equal to residudl signal for squared quality criterion (r=2) and is of the same sign as the residual signal for modular quality critetion (r=1). In these equations normalizing multipliers g(n) and $g_1(n)$ secure the convergence of successive iterations $\bar a[n]$ to the LDS parameters optimal value, determined by condition (4).

The initial value of target parameters in adaptive algorithms (12) and (14) may be equal zero.

## CONCLUSIONS

Suggested generalization of linear prediction allows to develope algorithms of the speech signal parameters estimation, that differ from traditional ones.

Introduced constants of generalized method, at the stage of joint constraint of coefficients and at the stage of preliminary transformations as well, provide additional degrees of freedom, that allow more completely take into consideration the current speech signal characteristics.

The given examples of the adaptive algorithms show the potential abilities of the examined above generalization of linear prediction method, but it is evident, that the problem of the speech signal parametrization is not solved yet.

## REFERENCES

[1] Г. Фант.
Акустическая теория речеобразования.
М. Наука, 1964.
[2] B.S. Atal and M.R. Schroeder.
Adaptive predictive coding of speech signals.
Bell. Syst. Thechn. J.,v.49,pp.1973-1986, oct. 1970.
[3] B.S. Atal and J.R. Hanauer.
Speech analusis by linear prediction of the speech wave.
J.Acoust.Soc.Amer.,v.50,n.2,pp.637-655,Aug.1971.
[4] А.Н. Собакин.
Об определении формантных параметров голосового тракта по речевому
сигналу с помощью ЭВМ.
Акустический журнал АН СССР, т.XVIII, вып. 1, стр. 106-114, 1972.
[5] H. Vakita.
Direct Estimation of the Vocal Tract Shape by Inverse Filtering of
Acoustic Speech Waveform.
IEEE Trans. on Audio Electroacoust.,v.AU-21,n.5.pp.417-427,1973.
[6] Дж. Маккол.
Линейное предсказание. Обзор.
ТИИЭР, т.63, вып.4,стр.561-580. Апр. 1975.
[7] Б. Фридландер.
Решетчатые фильтры для адаптивной обработки данных.
ТИИЭР, т.70,вып.8,стр. 54-98. "Мир",М., 1982.
[8] Л. Рабинер, Б. Голд.
Теория и применение цифровой обработки сигналов.
"Мир", М., 1978.
[9] Н.И.Акиезер, И.М.Глазман.
Теория линейных операторов в гильбертовом пространстве.
"Наука", М., 1966.
[10] E.Denoel,Sjlvay J.P.
Lineat prediction of speech with a least absolute error criterion.
IEEE Trans.on Acoust.,Speech,Signal Processing,v.ASSP-33,n.6,
pp.1397-1403, 1985.
[11] Е.Я. Ремез.
Общие вычислстельные методы чебышевского приближения.
Изд. АН УССР, Киев, 1957.
[12] Я.З. Цыпкин.
Адаптация, обучение и самообучение в автоматических системах.
"Наука", М., 1968.