ARCHIVES OF ACOUSTICS
6, 1, 23-44 (1981)

AVERAGING THE FREQUENCY OF THE LARYNX TONE IN THE CORRELATION METHOD OF ITS ESTIMATION USING THE TECHNIQUE OF LINEAR PREDICTION

ANDRZEJ DZIURNIKOWSKI

PESEL (02-106 Warszawa, ul. Pawińskiego 17/21)

This paper presents some problems in the effects of averaging the results obtained from the analysis of a speech signal, related to the use of the correlation method of the estimation of the frequency of the larynx tone. The present considerations are concerned with the analysis of a speech signal using the algorithm of linear prediction. The thesis that the results obtained from averaging depend on the parameters of the analysis assumed and the character of a signal is proposed and the reasons for this phenomenon discussed. This dependence must not be neglected in choosing the methods of speech signal analysis in real investigations.

1. Introduction

The autocorrelation analysis of a speech signal is one of the earliest methods of the estimation of the frequency of the larynx tone (the fundamental frequency). Since, in general, a speech signal is an implementation of a certain stochastic process which for adequately long durations in its selected classes, can be regarded as a stationary signal, it is necessary to take into consideration during the analysis of it the dependencies which arise from this fact and conditions for the estimators of an autocorrelation function of this type of signals.

This paper presents numerical methods of the estimation of the frequency of the larynx tone by means of the autocorrelation analysis using the technique of linear prediction. The direct considerations and examples are based on the analysis of the autocorrelation sequence of the error signal of linear prediction.

It is proposed that the precision of the results obtained using the estimation method assumed depends directly both on the parameters of the analysis assumed and the character of the signal itself in a predetermined interval of the analysis. The averaging of the estimated values of the periods of the larynx tone is an effect of these dependencies.

The present paper discusses this problem and presents a method for its analysis, based on a measure of the deviation of the estimated values of the periods of the larynx tone from their real values, as proposed by the author, and the smoothing coefficient defined on the basis of this measure.

The existence of these factors in the methods for the determination of the value of the periods of the larynx tone presented here must be considered in the determination of the aim of investigation and also calls in question its use in the spectral analysis of a speech signal, synchronised by the larynx tone.

2. Estimation of the autocorrelation function of a discrete stochastic process

All stochastic processes are characterised by giving a n-dimensional probability distribution of random variables for $n \to +\infty$. A human speech signal is one of these processes, experimental investigations of which are in view of practical constraints performed on a signal of finite duration. It is, therefore, impossible to determine precisely all the parameters of the probability distribution on the basis of experimental data. In this situation, functionals defined by selected implementations of random processes obtained from investigation of signals are assumed as the parameters of the probability distribution and are called the estimators of the parameters of processes analysed. In the correlation methods for the investigation of random signals, which are most frequently used to determine the regularity of the structure of a speech signal and for determination of its successive periods, called the larynx tone, the autocorrelation function of the process (or its transform) is assumed as the parameter, which is defined as

$$R_x(t_1, t_2) = E[X_0(t_1) X_0(t_2)], \tag{1}$$

where $X_0(t) = X(t) - E[X(t)]$. For the stationary processes this function depends only on $\tau = t_2 - t_1$. In numerous cases in view of the fact that the variance does not depend on the time t, the normalised autocorrelation function is assumed

$$R_x^u(\tau) = \frac{R_x(\tau)}{\sigma_{x^2}}. \tag{2}$$

Estimation of the parameters of the probability distribution of stochastic processes, based on the estimators, must always satisfy the following requirements: the estimators should be compatible, unweighted, "best" from the point of view of a criterion assumed (e.g. effective where effectiveness is the ratio of the minimum optimum variance to the variance of an estimator assumed) [12]. There is a number of estimators for determination of the autocorrelation function $R(\tau)$ of a stationary process, based on its implementation. The estima-

tor defined by the formula

$$P_x(\tau) = \int_0^{T-\tau} a(t, \tau) X_0(t) X_0(t+\tau)\, dt \tag{3}$$

is most frequently used, where $a(t, \tau)$ is the weight function whose selection affects in a significant way the quality of the estimator. In the case when the mean value of a random process is known, the quantity $P_x(\tau)$ defined by formula (3) is an unweighted estimator of the function $R_x(\tau)$. Determination of the optimum weight function a_{opt} in view of the minimum variance of the estimator $P_x(\tau)$ is here troublesome and requires knowledge of the autocorrelation function $R_x(\tau)$. Therefore, this estimator is quite often determined on the basis of the integral mean as

$$P_x(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} X_0(t) X_0(t+\tau)\, dt \tag{4}$$

which gives a relatively low estimation error [12]. The situation is different when the mean value of a random process is unknown and it is estimated on the basis of the implementation of the process itself. In this case the error of the weight of the estimator cannot be avoided. In order to avoid the weight of this estimator it would be necessary to introduce additionally a coefficient of the general form $1/1-F[R_x(\tau)]$ [12]. Since the function $R_x(\tau)$ is unknown at the time of estimation, therefore in order to avoid the weight first the weighted autocorrelation function is quite often determined and then iterative operations are used to avoid the weight [12]. It is a very complicated process, therefore, many authors consciously or not assume in their investigations estimated values weighted by error. The estimator of the autocorrelation function of random signal is essentially the estimator of the function and not a parameter and also a random function. Therefore, its covariance should be determined in the estimation of the error of such estimator. Given in [12], the formula for the estimation of the covariance of the estimator permits a conclusion that the covariance between the values of the estimator decreases with increasing duration of the process. Therefore, the final values of the estimator in a sequence are insignificant. In view of the above fact and also the relevant relations for the variability of the estimator [12] it is possible to determine in practice the duration τ_{max}, for which it is worthwhile to calculate the estimator [4, 5].

In practice under the pressure of requirements resulting from the assumptions of an experiment the analysis of this type is not frequently carried out despite the fact that only in some cases this approach would have been justified by the requirement of the experiment e.g. when the absolute values of the estimator are not essential, but the character of its variation or the position of its extrema etc. As for a continuous stationary process, for the stationary

sequence $\{X(n)\}$ the estimator defined by the formula

$$P_j = \frac{1}{N-j} \sum_{n=1}^{N-j} x(n)x(n+j) \tag{5}$$

is determined, to which all the remarks concerning the estimation of the autocorrelation function of a random process apply.

In the further part of this paper for the sake of uniformity of notation, the values of the autocorrelation function or its estimators will expressed by $\varrho(j)$ (the coefficients of the autocorrelation function).

3. Numerical methods for the estimation of the frequency of the larynx tone by the autocorrelation analysis using the technique of linear prediction

Only those signals will considered that are represented by the sample sequences $\{x(n)\}$ of the signal, i.e. signals in the form of numerical data, which can be easily processed digitally using computer.

One of the basic methods for the estimation of the frequency of the larynx tone used in the literature is the autocorrelation analysis of the sample sequence of a speech signal, $\{x(n)\}$ carried out on the basis of the relation

$$\varrho_x(j) = \sum_{n=-\infty}^{+\infty} x(n)x(n+j). \tag{6}$$

In practice the autocorrelation coefficients are calculated for the cut-off signal, i.e. a signal for which $x(n) = 0$ for $n < 0$ and $n > N-1$. This signifies that

$$\varrho_x(-j) = 0 \quad \text{for } |j| \geqslant N, \tag{7}$$

and

$$\varrho_x(-j) = \varrho_x - j = \sum_{n=0}^{N-1-j} x(n)x(n+j) \quad \text{for } j = 0, 1, \ldots, N-1.$$

There is a number of algorithms for calculating the autocorrelation coefficients, for example: the algorithm based on the calculation of the Fast Fourrier Transform (FFT) [9, 11] or the algorithm which implements the so called continuous correlation [9]. However, in both these and other methods, the results obtained, which are the basis for the estimation of the frequency of the larynx tone, are, in addition to the estimation error, weighted by the higher harmonic frequencies of F_0 (the effect of formants is significant), particularly the first one, that occur in a speech signal. One of the methods for the estimation of the larynx tone freguency by the autocorrelation technique in a signal from which the effect of formants was eliminated is the one proposed by ITAKURA and SAITO [7], a method

for the estimation of the frequency of the larynx tone, using the technique of linear prediction. The algorithm for calculating the estimated frequencies f_0, as described in [4] is based on the calculation in the discrete time domain of the successive values of the error signal $\{e(n)\}$ (cf. Fig. 1) defined as

$$e(n) = \sum_{i=0}^{M} a_i x(n-i) = x(n) + \sum_{i=1}^{M} a_i x(n-i), \tag{8}$$

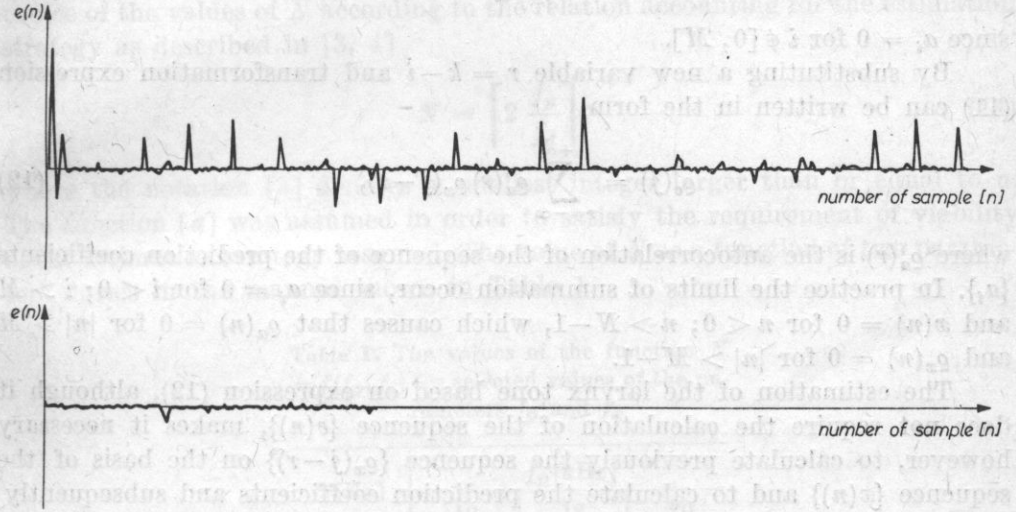

Fig. 1. An example of the error signal $\{e(n)\}$ in the word "pokoju"

where a_i are the coefficients of the inverse filter $A(z)$ of order M, defined in the z-domain as

$$A(z) = \sum_{i=0}^{M} a_i z^{-1}, \qquad a_0 = 1. \tag{9}$$

Given in [4], the analysis of the duration of the period of the larynx tone and the estimation of its frequency is based on the autocorrelation analysis of the sequence $\{e(n)\}$

$$\varrho_e(j) = \sum_{n=-\infty}^{+\infty} e(n)e(n+j). \tag{10}$$

There are a few specific variants of calculating the autocorrelation sequence $\{\varrho_e(j)\}$ using the relations between the coefficients of linear prediction and the real values of a signal, based on a direct implementation of the procedure described by formula (10) or using the autocorrelation function $\{\varrho_x(j)\}$ of the real data $\{x(n)\}$. The second variant uses the relation obtained from transformation

of formula (10) by insertion in it of the expression described by formula (8)

$$\varrho_e(j) = \sum_{n=-\infty}^{+\infty} \sum_{i=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} a_i a_k x(n-i) x(n+j-k). \tag{11}$$

This insertion involved the fact that

$$e(n) = \sum_{i=-\infty}^{+\infty} a_i x(n-i)$$

since $a_i = 0$ for $i \notin [0, M]$.

By substituting a new variable $r = k-i$ and transformation expression (11) can be written in the form

$$\varrho_e(j) = \sum_{r=-\infty}^{+\infty} \varrho_a(r)\, \varrho_x(j-r), \tag{12}$$

where $\varrho_a(r)$ is the autocorrelation of the sequence of the prediction coefficients $\{a_i\}$. In practice the limits of summation occur, since $a_i = 0$ for $i < 0$; $i > M$ and $x(n) = 0$ for $n < 0$; $n > N-1$, which causes that $\varrho_a(n) = 0$ for $|n| > M$ and $\varrho_x(n) = 0$ for $|n| > M-1$.

The estimation of the larynx tone based on expression (12), although it does not require the calculation of the sequence $\{e(n)\}$, makes it necessary however, to calculate previously the sequence $\{\varrho_x(j-r)\}$ on the basis of the sequence $\{x(n)\}$ and to calculate the prediction coefficients and subsequently on their basis the sequence $\{\varrho_a(r)\}$. When using one of the many available techniques of the calculation of the error signal given in [8] it is, however, more convenient to analyse on the basis of relation (10) in the limited summation range

$$\varrho_e(j) = \sum_{n=0}^{N-1-j} e(n) e(n+j), \tag{13}$$

where the sequence $\{\varrho_e(j)\}$ is in practice calculated for $j \leqslant N/2$. This method, which essentially consists in simple calculation of the autocorrelation coefficients, based on the values of the error signal of linear prediction will be the basis of the further considerations.

4. The effect of the method for estimation of the larynx frequency on the precision of results obtained

It follows from (13) that if one takes the sequence $\{\varrho_e(j)\}$ as the direct basis for tracing the duration of the period of the larynx tone by determining the values of $\max_j \{\varrho_e(j)\}$ in a predetermined interval Q depending on f_p (the sampling frequency) and a predetermined period of the analysis of the frequency

of the larynx tone, then the selection of the values of N in formula (13) essentially affects the averaging of the values of $\{f_{om}\}$, which are the successive values of the estimated frequency of the larynx tone.

It is interesting to consider how the values of N are chosen in relation to the frequency f_p and a predetermined analytical range of the larynx tone limited by the lower boundary f_d and the upper boundary f_g and how this affects the averaging of the estimated values of $\{f_{om}\}$, $m = 1, 2 \ldots$, in relation to their real values. In practice only the lower boundary essentially affects the choice of the values of N according to the relation accounting for the estimation strategy as described in [3, 4]

$$N = \left\lceil 2 \frac{f_p}{f_d} \right\rceil, \tag{14}$$

where the notation $\lceil a \rceil$ denotes a smallest integer larger than or equal to a. The function $\lceil a \rceil$ was assumed in order to satisfy the requirement of viability of the estimation strategy assumed. The value of N as a function of two parameters varies in the manner shown in Table 1.

Table 1. The values of the function $N = f(f_p, f_d)$ for selected values of the parameters f_p and f_d

f_d [Hz]	f_p [kHz]		
	10	12	20
20	1000	1200	2000
40	500	600	1000
50	400	480	800
60	334	400	667
80	250	300	500
100	200	240	400

While the lower predetermined frequency f_d affects the determination of the lower boundary of the interval Q, in which $\max_j \{\varrho_e(j)\}$ is traced, the upper frequency f_g is the lower limit of the interval Q; $j \in [L(f_g), N/2]$. The interval thus defined is essential for the effect of averaging the estimated values of f_{om}. This effect can to a lesser or greater degree occur in relation to the real period of the larynx tone.

If T_m denotes the real m-th period of the larynx tone then it can be stated that the longer the period T_m the lesser averaging effect occurs. The shorter the period T_m the greater the effect is, depending in direct proportion on the value of N. The variable N is a function of the permissible value of $T_d = \max_m \{T_m\}$, defined by formula (14). Therefore, the higher frequency of the

larynx tone is examined and at the same time the lower frequency f_d is taken, the greater the effect of averaging the frequency of the larynx tone will be in its estimation based on the autocorrelation method and the technique of linear prediction.

The above problem can be presented in the following way. If T_h denotes the hypothetical period of the larynx tone defined by the number of samples of the signal sampled at the frequency f_p and α denotes the filling coefficient of interval $[0, N]$ of the form

$$\alpha = \frac{N}{T_h} \tag{15}$$

then the function $\alpha(T_h)$ which takes values from the interval $\left[2, \frac{Nf_g}{f_p}\right]$ behaves as in Fig. 2.

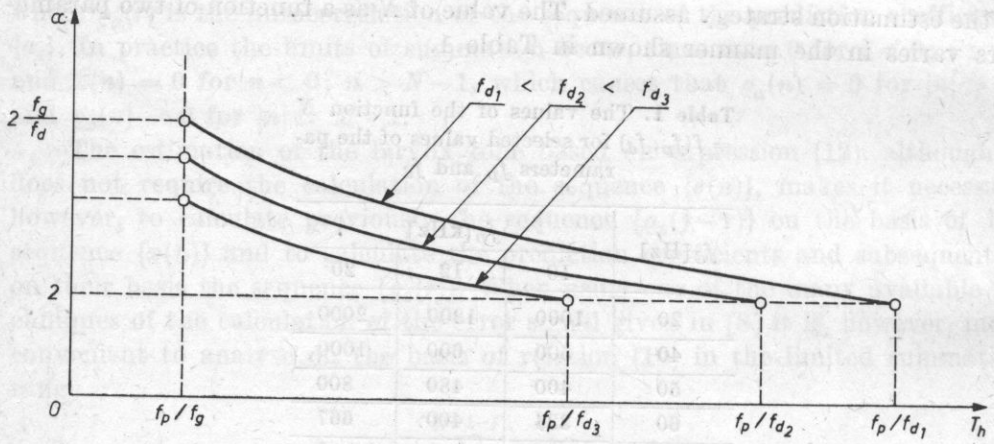

Fig. 2. The curve of the function $\alpha(T_h)$ in relation to the parameter f_d

It can be seen that depending on the frequency f_d assumed and in particular on the possible gap $f_r = f_g - f_d$, the filling coefficient varies taking a maximum value of

$$\alpha_{\max} = 2\frac{f_g}{f_p}. \tag{16}$$

When the function describing the signal is periodic, the coefficient α is insignificant in the calculation of the autocorrelation coefficients, since it is invariable for each successive period of a signal with the duration N. It is significant, however, in the autocorrelation analysis of quasiperiodic signals, e.g. a voiced speech signal. In this case $\alpha \neq$ const. and depending on the real values

of the period of the larynx tone T_m can be expressed by the formula*

$$\alpha_q = \sum_{l=1}^{P}\left[I\left(N-\sum_{m=0}^{l}T_m\right)+\frac{\left(N-\sum_{m=0}^{l-1}T_m\right)I\left(N-\sum_{m=0}^{l-1}T_m\right)}{T_l}I\left(\sum_{m=0}^{l}T_m-N\right)\right], \tag{17}$$

where

$$T_0=0, \quad p=\lceil \alpha_{\max}\rceil, \quad T_{p+1}=+\infty, \quad I(x)=\begin{cases}0 & \text{for } x<0,\\ 1 & \text{for } x\geqslant 0.\end{cases}$$

The coefficient α_q is a function of the successive periods of the larynx tone; $\alpha_q = f(T_0, T_1, T_2, T_3, \ldots, T_{\lceil \alpha_q\rceil})$ and takes the real positive values $\alpha_q \leqslant \alpha_{\max}$.

Formula (17) applies to the real conditions of the autocorrelation analysis of a signal of finite duration. The filling coefficient of a quasiperiodic signal, α_q, significantly affects the averaging of the estimated of the larynx tone, since it reflects the share of a given number of the variable periods in the calculation of the autocorrelation coefficients. There is still another aspect related to the behaviour of the autocorrelation function determined by the sequence of its coefficients, which is connected with a speech signal, and accordingly the error signal, being in most general terms a stochastic signal. When one considers only the voiced speech signal, which is quasiperiodic and therefore interesting for the estimation of the fundamental frequency f_{om}, then for adequately long signals it is often possible to assume that a signal is ergodic. Under this assumption for such signals the autocorrelation function behaves in a specific manner and is then defined by the relation

$$\varrho_x(j) = \lim_{N\to+\infty}\frac{1}{2N}\sum_{n=-N}^{n=N}X(n)X(n+j). \tag{18}$$

The autocorrelation function of an ergodic stochastic process is nonlinear and takes values from a limited range, which results from the fact that

$$\lim_{j\to 0}\varrho_x(j) = \varrho_x(0) = \lim_{N\to+\infty}\frac{1}{2N}\sum_{n=-N}^{n=N}X^2(n) \tag{19}$$

and

$$\lim_{j\to\pm\infty}\varrho_x(j) = \lim_{j\to\pm\infty}\sum_{X(n)}\sum_{X(n+1)}X(n)X(n+j)\varphi\left(X(n), X(n+j), j\right) = E^2[X(n)], \tag{20}$$

* Formula (17) is valid for the analysis of a quasiperiodic signal synchronised by the larynx tone as described in [4].

where φ is the probability density function, while $E[X(n)]$ is the first moment of the random variable $X(n)$* and $\varrho_x(0) = \max q_x(j)$. The possible assumption is here used, that for $j \pm \infty$ interaction of the random variables $X(n)$ and $X(n+j)$, represented by the values $x(n)$ and $x(n+j)$, disappears and in the boundary case they can be treated as statistically independent.

Similar relations apply to the error signal obtained using the technique of linear prediction. The foregoing considerations dealt with the autocorrelation signal analysis based on relation (6). In practice the analysis is carried out on a signal of finite duration using relations (7) and (13). This fact naturally causes a more rapid decrease in the value of the autocorrelation function. It can be expressed in approximation by the relation

$$\varrho'(j) = \beta(j)\,\varrho_x(j), \tag{21}$$

where $\beta(j) = \dfrac{N-1-j}{(N-1)}$; $j \in [L(f_g), N/2]$ and functions as the damping coefficient of the autocorrelation $\varrho_x(j)$ (see Fig. 3).

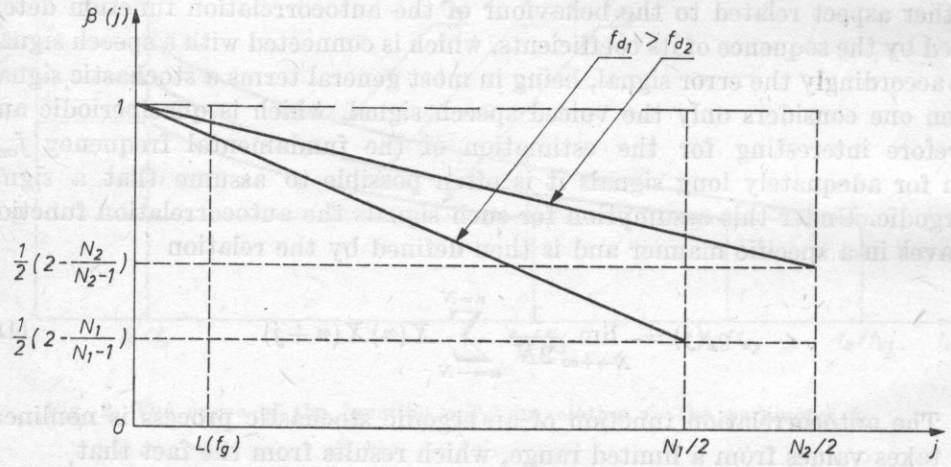

Fig. 3. The dependence of the linear "damping coefficient" of autocorrelation, $\beta(j)$ on the length of the calculation interval

The relation discussed above, which is connected with the random character of the signal, is distinctly reflected in Figs. 4-8. In practice, relation (21) only expresses a trend in the behaviour of the autocorrelation function, which can be disturbed by the quasiperiodicity of the signal. This phenomenon is shown in Figs. 4-8. A more exact consideration of how the signal itself affects the be-

* In the case of quasiperiodicity of the ergodic process

$$E[X(n)] = E[X(n+j)] = \ldots = E[X].$$

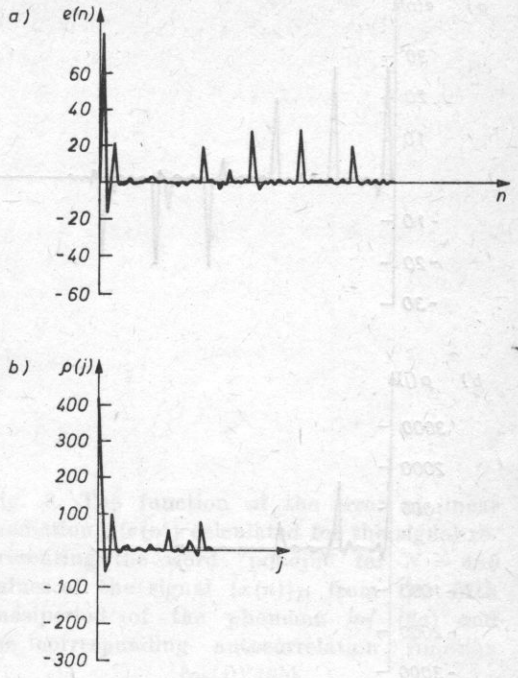

Fig. 4. The function of the error of linear prediction, $\{e(n)\}_D$ calculated for the signal $\{X(n)\}_D$ representing the word "pokoju", for the first $N = 480$ values of the signal $\{X(n)\}_D$ from the beginning of the phoneme /k/ (4*a*) and the corresponding autocorrelation function $\{\varrho_e(j)\}$ (4*b*)

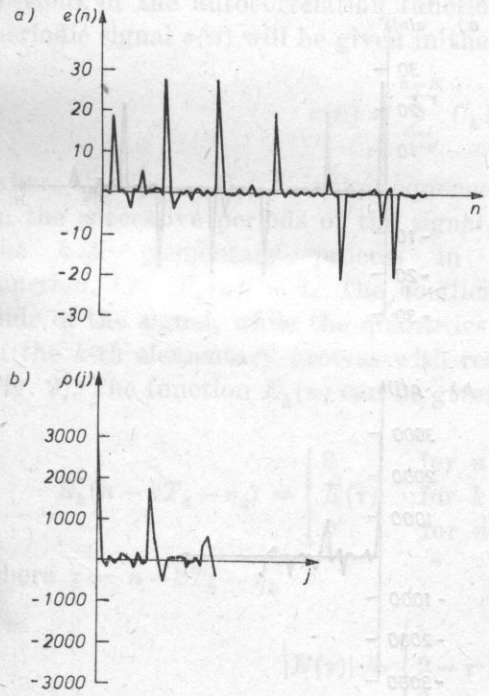

Fig. 5. The function of the error of linear prediction, $\{e(n)\}$ calculated for the signal representing the word "pokoju" for $N = 480$ values of the signal $\{x(n)\}_D$ from the second quasiperiod of the signal of the phoneme /o/ (5*a*) and the corresponding autocorrelation function $\{\varrho_e(j)\}$ (5*b*)

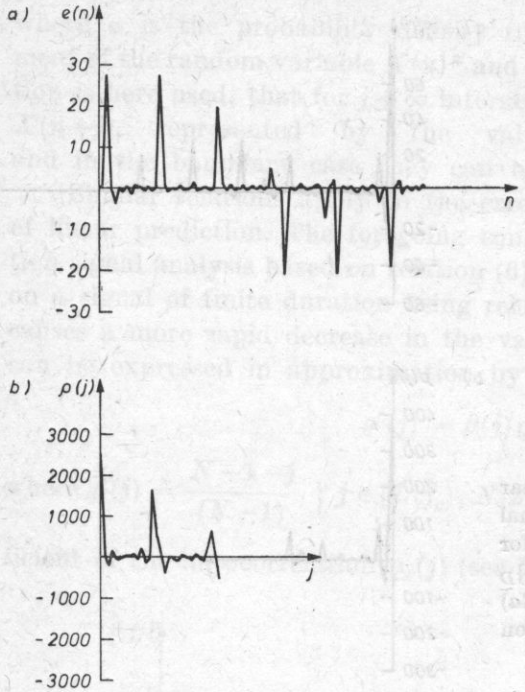

Fig. 6. The function of the error of linear prediction, $\{e(n)\}$ calculated for the signal representing the word "pokoju" for $N = 480$ values of the signal $\{x(n)\}_D$ from the third quasiperiod of the signal of the phoneme /o/ (6*a*) and the corresponding autocorrelation function $\{\varrho_e(j)\}$ (6*b*)

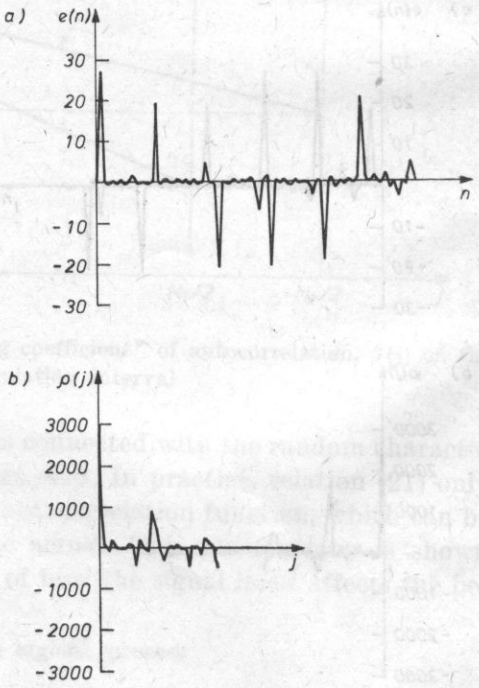

Fig. 7. The function of the error of linear prediction, $\{e(n)\}$ calculated for the signal representing the word "pokoju" for $N = 480$ values of the signal $\{x(n)\}_D$ from the fourth quasiperiod of the phoneme /o/ (7*a*) and the corresponding autocorrelation function $\{\varrho_e(j)\}$ (7*b*)

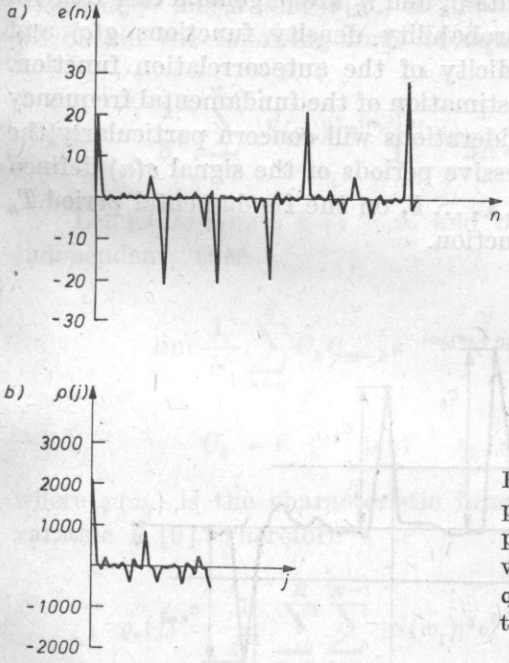

Fig. 8. The function of the error of linear prediction, $\{e(n)\}$ calculated for the signal representing the word "pokoju" for $N = 480$ values of the signal $\{x(n)\}_D$ from the fifth quasiperiod of the phoneme /o/ (8*a*) and the corresponding autocorrelation function $\{\varrho_e(j)\}$ (8*b*)

haviour of the autocorrelation function now follows. To that end the quasi-periodic signal $e(n)$ will be given in the form

$$e(n) = \sum_{k=0}^{k=K} C_k E_k(n - kT_h - \eta_k), \tag{22}$$

where $C_0, C_1, \ldots, C_k, \ldots$ is the sequence of the coefficients of amplitude changes in the successive periods of the signal, while $E_k(n)$ is the function describing the k-th elementary process in the period and is a normalised function, i.e. $|E_k(n)| = 1$. The coefficients C_k characterise changes in amplitude of the signal, while the quantities η_k describe changes in the initial phase of the k-th elementary process with respect to the hypothetical period T_h (see Fig. 9). The function $E_k(n)$ can be given in approximation in the following way

$$E_k(n - kT_h - \eta_k) = \begin{cases} 0 & \text{for } n < kT_h + \eta_k, \\ E(\tau) & \text{for } kT_h + \eta_k \leqslant n \leqslant (k+1)T_h + \eta_{k+1}, \\ 0 & \text{for } n > (k+1)T_h + \eta_{k+1}, \end{cases} \tag{23}$$

where $\tau = n - kT_h - \eta_k$

$$|E(\tau)| = \begin{cases} \tau & \text{for } 0 \leqslant \tau \leqslant 1, \\ 2 - \tau & \text{for } 1 < \tau \leqslant 2, \\ 0 & \text{for } \tau > 2. \end{cases}$$

It can be assumed that the coefficients C_k and η_k are in general case random variables characterised by unknown probability density functions $g(c)$ and $g(\eta)$, and essentially disturb the periodicity of the autocorrelation function. Since the interest here is first of all the estimation of the fundamental frequency of signals of this type, the present considerations will concern particularly the influence of change in duration of successive periods of the signal $e(n)$ defined by the sequence of values of η_k, $k = 0, 1, \ldots k$, on the fundamental period T_h estimated using the autocorrelation function.

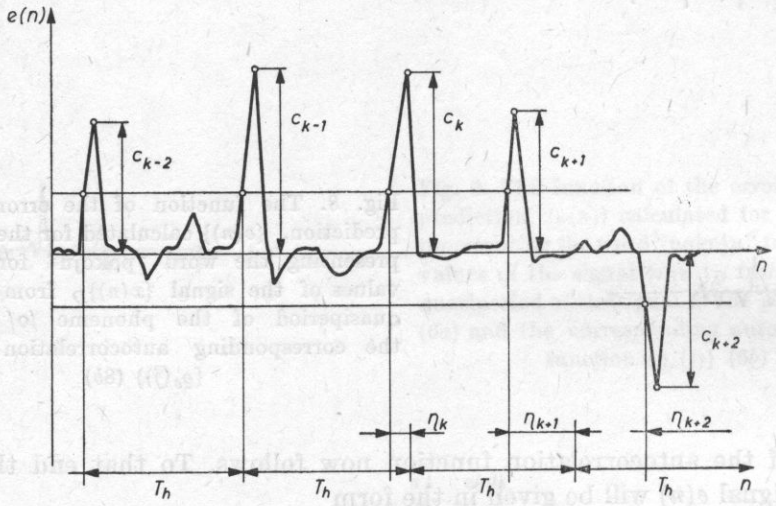

Fig. 9. A random sequence of pulses of a quasiperiodic signal

We insert into formula (18) the signal described by formula (22)

$$\varrho_e(j) = \lim_{N\to\infty} \frac{1}{2N} \sum_{n=-N}^{n=N} e(n)e(n+j)$$

$$= \lim_{N\to\infty} \frac{1}{2N} \sum_{n=-N}^{n=N} \sum_{k=0}^{k=K} C_k E_k(n - kT_h - \eta_k) \sum_{l=0}^{l=K} C_l E_l(n - lT_h - \eta_l + j)$$

$$= \lim_{N\to\infty} \frac{1}{2N} \sum_{k=0}^{k=K} \sum_{l=0}^{l=K} C_k C_l \sum_{n=-N}^{n=+N} E(n - kT_h - \eta_k) E(n - lT_h - \eta_l + j). \quad (24)$$

Substituting the relevant Fourier transforms [10] for the expression $E(n)$, with consideration of delay

$$E(n - kT_h - \eta_k) = \frac{1}{N} \sum_{r=0}^{N-1} S(\omega_r) e^{i\omega_r n} e^{-i\omega_r kT_h} e^{-i\omega_r \eta_k},$$

where $\omega_r = 2\pi r/N$ and $S(\omega_r)$ is the amplitude spectrum of the process $E(n)$, we obtain the following form of equation (24):

$$\varrho_e(j) = \frac{1}{N^2}\sum_{r=0}^{N-1}|S(\omega_r)|^2 e^{i\omega_r j}\left[\lim\frac{1}{2N}\sum_{k=0}^{K}\sum_{l=0}^{K}C_k C_l e^{-i\omega_r(\eta_k+\eta_l)}e^{-i\omega_r(k+l)T_h}\right]. \quad (25)$$

Let us designate $k+l = m$ and note that if C_k and C_l and η_k and η_l are independent, then

$$\lim\frac{1}{N}\sum_{k=0}^{K}C_k C_{m-k}e^{-i\omega_r(\eta_k+\eta_i)} = \begin{cases}\bar{C}^2 & \text{for } m = 0,\\ \bar{C}^2[\varphi(\omega_r)]^2 & \text{for } m \neq 0;\end{cases}$$

$$C_k = C_{-k} \quad \text{and} \quad \eta_k = \eta_{-k}, \quad \varphi(\omega_r) = E[e^{j\eta\omega_r}],$$

where $\varphi(\omega_r)$ is the characteristic function of the distribution of the random variable η [9]. Therefore

$$\varrho_e(j) = \frac{1}{N^2}\sum_{m=0}^{K}\sum_{r=0}^{N-1}|S(\omega_r)|^2 e^{i\omega_r j}e^{-i\omega_r mT_h}\left\{\lim\frac{1}{2N}\sum_{k=0}^{K}C_k C_l e^{-i\omega_r(\eta_k+\eta_l)}\right\}$$

$$= \frac{\bar{C}^2\varrho_0(j)}{2N^2} - \frac{\bar{C}^2}{2N^2}\sum_{m=1}^{K}\sum_{r=0}^{N-1}|S(\omega_r)|^2|\varphi(\omega_r)|^2 e^{-i\omega_r(mT_h-j)}, \quad (26)$$

where $\varrho_0(j)$ is the autocorrelation function of the periodic signal $E(n)$.

It should be noted that relation (26) is valid for $N\to+\infty$, i.e. at the same time $K\to+\infty$, which is not satisfied for investigations in practice, and also under the assumption that $E[e] = 0$. More exact consideration of these problems can be found in paper [1]. The above theoretical considerations lead to the conclusion that the autocorrelation function of the signal described by formula (22) is a complicated probability density function of the random variable η and C and therefore it is difficult to draw conclusions on the behaviour of values of the autocorrelation function without knowledge of these distributions.

Irrespective of the form assumed for the distributions of these variables, the results of analyses based on relation (22) do not have general character, since the shape of these distributions is strongly dependent on the individual characteristics of the speaker. It is, however, an essential result of these considerations that they confirmed strong dependence of the autocorrelation function on the length of the analytical interval and on the real duration of successive periods. The determination of the degree of averaging the frequency of the larynx tone, consisting in tracing $\max_j\{\varrho(j)\}$ in a predetermined analytical interval is therefore dependent on the values

of N and T_h, the latter being a hypothetical duration of the larynx tone period, chosen as the first period within the interval of the signal processed.

It seems justified to introduce a measure which would be a simple relation describing the degree of deviation of the estimated mean duration of the period of the larynx tone, from the real length of the first period within the predetermined l-th interval of the analysis. This measure can be expressed in the form

$$\delta_1^l = 1 - \frac{T_s^l}{T_{rz}^l}, \qquad T_{rz}^l = T_h^l = T_1^l. \tag{27}$$

The averaged estimated length can in approximation be described as the arithmetic mean

$$T_s^l = \frac{\sum_{i=1}^{a_k} T_i^l}{a_k^l} \cong \frac{N}{a_k^l}, \qquad a_k^l = [a_q^l]. \tag{82}$$

If we hypothetically assume the periodicity of the signal in the investigated interval N with a predetermined period equal to a chosen value T_i, $i = 1, 2, \ldots, a_k$, then we can determine the filling of this interval by relevant periods of the signal as $a_i = [N/T_i]$. Thus assuming that $T_1^l = T_{rz}^l \cong N/a_1^l$ we obtain the approximate value δ_1^l of the quantity δ_{1p}^l as

$$\delta_{1p}^l \cong 1 - \frac{a_1^l}{a_k^l}. \tag{29}$$

It should be noted here that expression (29) can be used to determine the approximate value δ_1 defined by formula (27) only for the relatively long analytical intervals with respect to the periods of the larynx tone under analysis (i.e. for high values of a_k). Fig. 10 shows schematically the examples of values taken by the function a_q depending on the distribution of the real successive periods of the larynx tone for a predetermined length of the analytical period N.

It follows from the foregoing considerations that in the autocorrelation analysis the estimated periodicity of the signal deviates in succesive steps of the analysis from the real values of periods of the quasiperiodic signal and depends in most general case on the value of a_k and also on the real period $T_1 = T_{rz}$ itself. This difference can be defined in approximation by the deviation measure δ_1 whose sign shows the direction of this deviation.

The deviation measure δ_1 does not define the measure of averaging irrespective of how this measure will be defined. There is, however, a relation between these two measures, since the averaging measure is a function of the variables (N, T_i^l, δ_1^l).

If we take as the averaging measure a smoothing coefficient of the sequence of the real durations of the period of the larynx tone in the form of a normalised sum of differences between the successive real values and estimated period

durations for their predetermined number L in the form

$$W^L = \frac{1}{L}\sum_{i=1}^{L} |(T^l_{rz} - T^l_s)/T_{rz}| \cong \frac{1}{L}\sum_{i=1}^{L} |\delta^i_{1p}| \tag{30}$$

then the thus defined smoothing coefficient is the arithmetic mean of successive deviation measures. The greater absolute values the coefficient W^L takes the more significant the smoothing of the real values of the sequence $\{T^i_{rz}\}$ becomes. Some conclusions can be drawn, therefore, on the averaging of these values in their estimation by the method of linear prediction, based on determination in a predetermined analytical interval of the averaged values of successive periods of the larynx tone [4]. The value of the coefficient W^L via the absolute values of the deviation measure δ^l_{1p} depends on the ratio a^l_1/a^l_k which to some extent reflects the quasiperiodicity of the signal investigated.

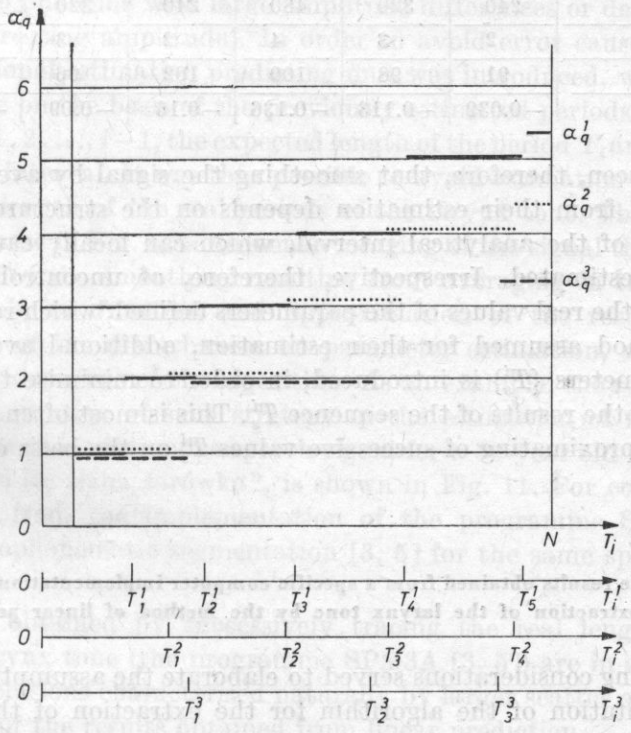

Fig. 10. An example of the values of the function depending on the length and distribution of quasiperiods of a signal

The longer the analytical interval defined by the value of N and the shorter the quasiperiods T_i contained in it, the greater value a_k takes, which affects determination of the averaged duration of the estimated period T_s. At the same time, however, the greater the measure δ_1 can become in the case of local steplike change in the real value of the larynx tone. In turn, for low values of N (short

analytical interval) and long quasiperiods $T_i a_k$ decreases, which causes large averaging errors to occur in the estimation of value of T_s and subsequently significant changes in values of δ_1. This can be demonstrated by the estimation of the period T_s for the two successive sequences of the quasiperiods T_i; the sequence $C_1 = \{88, 95, 103, 107, 111, 119\}$ and the sequence $C_2 = \{89, 115, 85, 108, 112\}$. The results obtained are shown in Table 2.

Table 2. Variation in the deviation of the estimated mean duration of the period of the larynx tone from the real duration of the first period, depending on the length of the analytical period N, for two chosen sequences of values of T_i

Coefficients	Sequence					
	C_1			C_2		
N	240	320	480	240	320	480
a_k	2	3	4	2	3	4
T_s	91	98	100	102	96	99
δ_1	−0.039	−0.113	−0.136	−0.16	−0.09	−0.12

It can be seen, therefore, that smoothing the signal by averaging its real values obtained from their estimation depends on the structure of the signal and the length of the analytical interval, which can locally cause large error of the values estimated. Irrespective, therefore, of uncontrolled smoothing or averaging of the real values of the parameters defined, which results from the analytical method assumed for their estimation, additional averaging of the estimated parameters $\{T_s^i\}$ is introduced, in order to minimise the local errors and smooth out the results of the sequence T_s^i. This is most often done by linear or nonlinear approximating of successive values T_s^i on the basis of the previous values [4].

5. Discussion of the results obtained from a specific computer implementation of the algorithm for the extraction of the larynx tone by the method of linear prediction

The foregoing considerations served to elaborate the assumptions for a computer implementation of the algorithm for the extraction of the larynx tone in a continuous speech signal. This algorithm was written in the Fortran language and used for statistical investigations aimed at elaboration of a method for speaker identification.

This algorithm described in [4] assumed the following implementation conditions. The lower frequency limiting the analytical range of the larynx tone frequency was taken as $f_d = 50$ Hz, which at the sampling frequency used in these investigations defined the length of the analytical segment $N = 480$ samples

according to formula (14) and Table 1, while the upper limiting frequency of the larynx tone was $f_g \cong 333$ Hz.

Thus the coefficient a_k was contained in the interval from 2 to 14 and in the investigation of male voices whose mean frequency F_0 oscillated about a frequency of ~ 120 Hz, this coefficient was close to 5 (Figs. 4-8). The averaging range in a signal of relatively low disturbance in its periodicity should be thus selected that it could be possible to estimate correctly successive values of F_{oi} and at the same time smooth out their variations. Moreover, it may happen during the estimation of the larynx tone by the autocorrelation technique that disturbances in the structure of the signal $\{X(n)\}$ and in $\{e(n)\}$ were so significant as to cause (as in the case in Fig. 4b) "local disturbances" in the work of the unit tracing $\max_j\{\varrho_e(j)\}$. This also happened (in addition to the events shown in Fig. 4) when the analytical segment contained transitions from phoneme to phoneme with large amplitude differences or decaying signals (in terms of decreasing amplitude). In order to avoid error caused by this influence an additional estimation predicting unit was introduced, whose function was to determine on the basis of the previously estimated periods of the larynx tone, $T_{i-n}, n = 1, 2, \ldots, i-1$, the expected length of the period T_i and accordingly of the successive interval. It was thus possible to avoid estimation error equal to half or multiple length of the real periods of the larynx tone. This caused, however, as in the case in Fig. 4, additional averaging of the signal through smoothing in addition to the smoothing resulting from averaging in the calculation of the autocorrelation function and approximation of the results, reducing partly the effect of the disturbances on parameter estimation, which did not occur in the previous steps of the algorithm. As an example, the result obtained from the implementation of such strategy of the estimation of the frequency of the larynx tone in a continuous speech signal for a 30-year old man who said "W pokoju paliła się słaba żarówka", is shown in Fig. 11. For comparison the results obtained from the implementation of the programme SPM3A based on primary microphonematic segmentation [3, 5] for the same speech signal is shown in Fig. 12.

The results obtained by successively tracing the real lengths of single periods of the larynx tone (the programme SPM3A [3, 5]) are in accordance to the previous conclusions characterised naturally by larger scatter of their values than in the case of the results obtained from linear prediction.

The results obtained by linear prediction reflect changes in the mean values of the results of primary segmentation calculated for a time constant of the order of scores of milliseconds, which fully confirms the foregoing conclusions on smoothing and averaging results obtained from the autocorrelation analysis used in the estimation of the larynx tone by the method of linear prediction. The differences at the beginning and at the end of the curves in Figs. 11 and 12 result from relatively high disturbances of the structure of the signal $\{X(n)\}$ and also

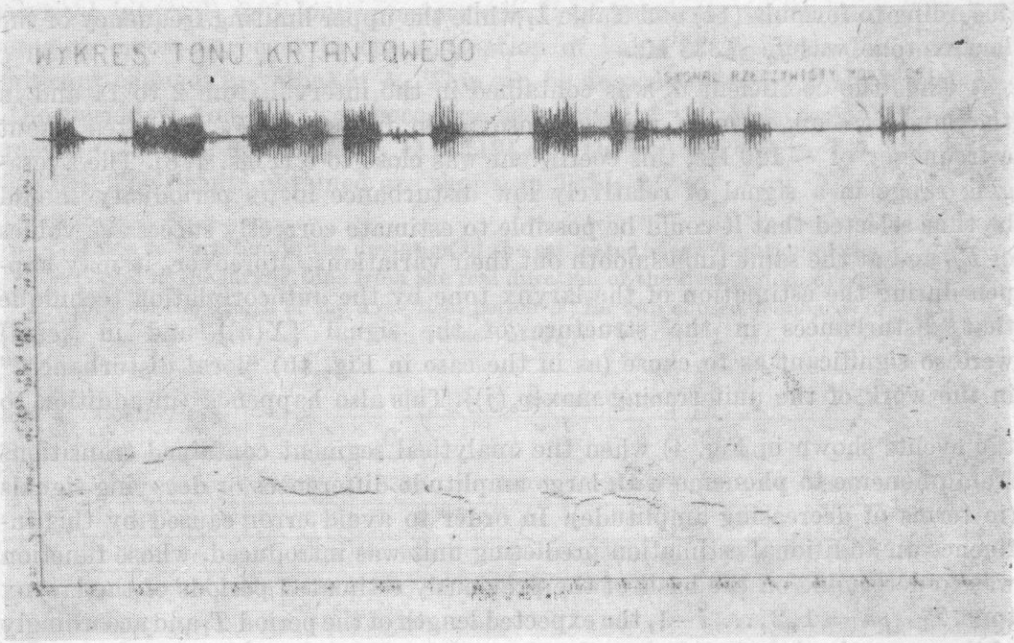

Fig. 11. The estimated pitch contour obtained by the method of linear prediction for the signal representing the sentence "w pokoju paliła się słaba żarówka"

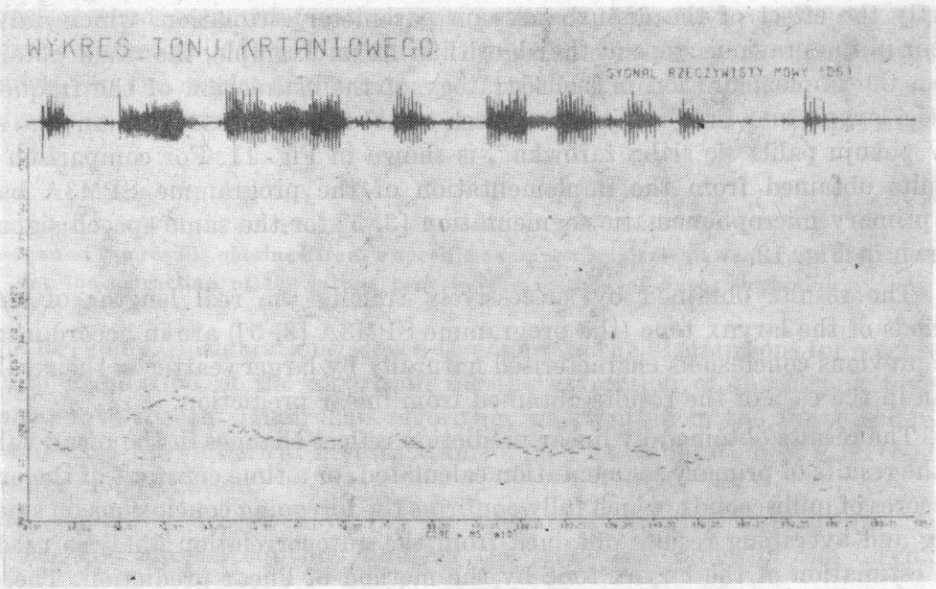

Fig. 12. The estimated pitch contour obtained by primary segmentation of the speech signal representing the sentence "w pokoju paliła się słaba żarówka"

from the assumed threshold values and the analytical intervals of F_0 (cf. [5]) and can be avoided by changing these values or their adequate fitting to a specific signal.

6. Conclusions

It can be stated on the basis of the results obtained by the two methods for the estimation of the parameter F_0 that both are viable from the point of view of their conditions and in agreement with theoretical predictions. These methods can be used for further analysis of human speech for different purposes. While the results obtained from primary segmentation serve for spectral analysis of the signal synchronised by the larynx tone in a system for speech recognition [2], the results obtained from the implementation of the programme based on the method of linear prediction are useless for this type of analysis. They can be used, however, in statistical investigation of the parameter F_0 for speaker identification or verification [6]. At present the author is performing extensive investigations of selected statistical parameters of the distributions of the frequency F_0 obtained by using both methods described above in order to determine the effect of the extraction methods used on the precision of identification and verification of speakers, depending on the classification algorithms employed. Preliminary investigations showed that there are parameters which show smaller scatter when the model of linear prediction is used than in the case when the method of primary segmentation is used, which is essential for the classification of characteristics. This confirms the conclusion that the choice of analytical method essentially affects the results of the investigations of speech signal and accordingly different directions of investigation (e.g. speech recognition or speaker identification require different analytical methods). In the present case the estimation of the frequency of the larynx tone by the method of linear prediction for spectral analysis of a speech signal synchronised by the larynx tone is not suitable despite advantages of the method itself (e.g. easy implementation of the digital algorithm for linear prediction).

References

[1] H. H. Beisner, *Spectrum of the lagged product in crosscorrelation*, JASA, **62**, 4, 916–921 (1977).

[2] L. Bolc, A. Dziurnikowski, K. Jaszczak, G. Kiełczewski, *Sistiema ponimania rieczi SUSY, Materiały konferencyjne: Trudy rasszíriennowo zasiedania rabocziej grupy II* KHBBT *po mietodam rozpoznawania, klasyfikacji i poiska informacji*, IOiK PAN, Warszawa 1976, 219–257.

[3] A. Dziurnikowski, *Primary segmentation of speech sound signals in the* SUSY *system*, Reports of JJUW, **52** (1976).

[4] A. DZIURNIKOWSKI, *Automatic determination of the frequency behaviour of the larynx tone by the method of linear prediction*, Archives of Acoustics, **5**, 1, 31–46 (1980).

[5] A. DZIURNIKOWSKI, *Microphonemes as fundamental segments of a speech wave; Primary segmentation – automatic searching for microphonemes*, Papers of the IJCAI – 75, 476–482, Tbilisi 1975.

[6] A. DZIURNIKOWSKI, *Automatic speaker recognition, problems and methods* (in Polish), Problemy kryminalistyki, **143**, 54–68 (1980).

[7] F. I. ITAKURA, S. SAITO, *Analysis-synthesis telephony based upon the maximum likelihood method*, Proc. 6th Int. Congress on Acoustics, C17–20, Tokyo 1968.

[8] J. D. MARKEL, A. H. GRAY, *Linear prediction of speech*, Springer–Verlag, Berlin, Heidelberg, New York, 1976.

[9] R. K. OTNES, L. ENOCHSON, *Digital time series analysis* (in Polish), WNT, Warszawa 1978.

[10] B. W. PAWŁOW, *Diagnostic investigation in technology* (in Polish), WNT, Warszawa 1967, 142–150.

[11] G. SANDE, *On the alternative method for calculating covariance functions*, Princeton Computer Memorandum, Princeton, New York 1965.

[12] S. J. WILENKIN, *Statistical methods for investigation of automatic control system* (in Polish), WNT, Warszawa 1969, 13–13, 61–62.

Received on August 1, 1979; revised version on August 7, 1980.