Let us take as model for the speech source a linear dynamic system characterized by a memory function which, driven by a forcing function, gives the speech signal at its output. The forcing function is a quasi-periodic train of laryngeal excitations which follow one after another, by \( t > 0 \), in intervals equal to the pitch period and have so short a duration that at the end of each pitch period equal zero. As the memory function seems to have such characteristics that at the end of each pitch period the response of each single glottal pulse also equals zero, the latter can be expressed with the 'identical' functions or elementary waves

\[
x_k(t; a_k) = \begin{cases} x_k(t; a_k) & \text{if } T_k \leq t < t_k \\ 0 & \text{otherwise} \end{cases}
\]

(1)

where \( a \) is an aggregate of waveform parameters, \( T_k \) is the value of the pitch 'period' in the \( k \)th elementary wave and \( k \) is chosen arbitrarily.

From the speculation about the nature of the speech signal, made above, it follows that each single elementary wave contains no periodic components and is statistically independent. Hence each vocal segment of the speech signal, defined as a phone (Fig. 1), can be regarded as a non-stationary random function \( X(t; a) \) or random speech wave presented by the set of its trials

\[
x(t; a_1), x(t; a_2), x(t; a_3), \ldots, x(t; a_k), \ldots
\]

(2)
which occur in the fixed interval \( (t_2, t_3 + T_k) \), where \( T_k = C^v \geq T_{ak} \) and whereas \( k \) is chosen arbitrarily, it may be chosen in an orderly manner as well.

Let us take samples \( x(t_m; a) = x_{km} \) of the \( k \)\textsuperscript{th} elementary wave in its successive phases of development \( t_m(m = 0, 1, 2, 3, \ldots n) \) at constant sampling intervals \( \Delta t = T_m - t_{m-1} = 1/2 T_{ak} \), where \( T_{ak} \) is the signal bandwidth, which are sufficiently long to ensure the independence of the samples. Then each elementary wave is presented with the set of samples

\[
x_{km}^j(t_m; a) = x_{km}, x_{km+1}, x_{km+2}, \ldots x_{km+\delta T_{ak}}, \ldots x_{2km}
\]

where the samples \( x_{km} \) in the points of zero-crossing and the samples with numbers of sampling between \( k = T_{ak}/\Delta t \) and \( n = T_{ak}/\Delta t \) (See Eq. 1) equals zero.

If the sections of the random speech wave \( x(t; a) \) in the fixed phases of development \( t_m \) are grouped to form ensembles

\[
X(t_m; a) = x_{1m}, x_{2m}, x_{3m}, \ldots x_{km}, \ldots x_{2km}
\]

then the ensembles or random variables \( X(t_m; a) \) form the system of random variables \( X^*(t; a) \)

\[
X^*(t; a) = (X(t_1; a), X(t_2; a), X(t_3; a), \ldots X(t_m; a), \ldots X(t_n; a))
\]

1. Unidimensional Subactivities. \( X^* = X^*_1 + X^*_2 + X^*_3 \) where \( X^*_1 = -x_{km} \), \( X^*_2 = -\delta T_{ak} x^*_2 \) and \( X^*_3 = -\delta T_{ak} x^*_3 \) where \( \delta = C^v \) is scale reduction coefficient which takes values between zero and unity, \( 0 < \delta < 1 \). The unidimensional components form sets of measurements which generate the statistical parameters (mean, variance, etc.) of the distributions of amplitude and pitch period, and the duration of the speech sound. The space filling properties of the portrayal \( X^* \) are given by the mixed product \( V_{X^*} = X^* \times X^*_1 \). The unit vector \( \hat{x} = X^*/|X^*| \) of \( X^* \) gives its directional properties.

2. Two Dimensional Subactivities. \( X^* = X^*_1 + X^*_2 + X^*_3 \) which generate a set of subportrayals: the plot \( \hat{e}_p(t) \) of mean peak amplitude \( (e_p = x_{km}/2) \) as time or mean.
amplitude envelope, the plot $T_0(t)$ of pitch 'period' vs time and the family of waveforms of the trials of the random function.

It is known,⁹ that if we let the total number of samplings per speech sound tend to discontinuity then the system of random variables (Eq. 5), presented by the pattern $X^*$, becomes equivalent to the random speech wave $X(\tau, a)$. Hence any visual representation of speech based on this principle¹⁰,¹¹ should be considered as reliable (Fig 3).

![Fig. 3. Waveform (WF) Portrayal of the vowel [d] as in [drip] build from the high speed oscillogram shown in Fig. 1. Total number of elementary waves $h = 11$. Recorder scale [sec. $10^{-3}$]. Duration $82.10^{-3}$ sec ($\delta = 0.032$).

The pitch 'period' normalized portrayal $B^*$ of the random function generates its mean, variance, correlation function and power spectra. Assuming that each trial $x_k(\tau, a_k)$ exists between two successive positive-going zero-crossings, the portrayal $B^*$ can be obtained after prearranging of the time-domain samples $x_{km}$ in a new phase-angle basis according to the rule given by the matrix $B$

$$ B = \| B_{km} \| \quad (k = 1, 2, 3 \ldots h) (M = 1, 2, 3 \ldots N) $$

where $B_{km} = \| x_{km} \ldots x_{k(m+\delta_k)} \|$

and where $\pm \delta_k \approx \pm n(k)/2\pi$ is an entire digit.

But we may choose another way for analysis of the waveform portrayal by which no pitch normalization is needed; we can use parameters for direct evaluation of the successive waveforms like the crest factor $C$ and the form factor $F$. These parameters, together with the slope of the pulse front $S = \text{Peak/Rise time}$,

---

⁹) Пугачев, В. С., „Теория случайных функции“, Физматгиз, Москва (1960), p. 204.
can be applied also in regard to the mean envelope $\bar{V}_p(t)$ and to the plots of the another waveform parameters vs time: $T_0(t)$, $C(t)$, $F(t)$, etc.

Some preliminary results of the application of the methods described in this paper to the practical problems of speech analysis can be considered as encouraging. Five stressed Bulgarian vowels, uttered by speaker A. M., in nonsense syllables, was subjected to speech analysis by the waveform. Sampling with $\Delta \tau = 0.0001$ sec was accomplished manually from their Waveform (WF) Portrayals shown in Fig. 4. The mean product-factor of the waveform $P = \bar{C} \cdot \bar{F}$, presented in the right-hand side of Fig. 4., seems to match the complexity of the waveform and, since it is non-dimensional, it appears that it is closely correlated to the phonemic value of these vowels. The observation that the sudden change in voice effort during the stress...
results in corresponding changes in the waveform\textsuperscript{12} makes it reasonable to suggest that the product \((C_{\text{P}0} \cdot C_{\text{I}0})\) would be effective by digital evaluation of stress. The envelope form of \(I_{\text{I1}}\)\textsuperscript{13} can be evaluated by the product factor of the envelope \(P_{\text{I}0}\), the ratio between its value in the modulated and non-modulated segments of the carrier vowel sound being greater or equal to 1.55. It is suggested that the most important acoustic cue of \(I_{\text{I1}}\) is the slope \(S_{\text{I}0}\) of the plot of pitch vs time,\textsuperscript{14} the ratio between the slopes of phones, \(I_{\text{I1}}\) and \(I_{\text{I1}}\), with similar waveforms and amplitude envelope forms being found to be 7.5.

\[
\begin{align*}
Q(x_{km}, \tau_m, t_k) \\
x^* &= -x_{km}x_0^* \\
x^*_r &= -\frac{\tau_m}{m}x_0^* \\
x^*_t &= \delta(t_{k-1} - t)x_0^*
\end{align*}
\]

Fig. 5. Computing circuit of the process of statistical analysis of speech by waveform.

It has been shown recently that speech recognition is possible when a computer is presented with short samples of the acoustic waveform, the samples being processed without preliminary analysis.\textsuperscript{15} It is hoped that if a computer operates with the input speech wave according to the ideas set forth in this paper (Fig. 5), the efficiency of the process of mechanical recognition as well as the quality of its output would be improved.

\textsuperscript{13} Christov, P., “Experiment for Changing the Envelope Form of Vowels”, 5\textsuperscript{e} C.I.A., Reports, Liège (1965), A11.