ARCHIVES OF ACOUSTICS
4, 3 (1979)

PROBABILITY DENSITY DISTRIBUTION OF INSTANTANEOUS VALUES OF A SPEECH SIGNAL IN POLISH

STEFAN P. BRACHMAŃSKI, WOJCIECH MAJEWSKI

Institute of Telecommunication and Acoustics, Wrocław Technical University (50-317 Wrocław)

In the present paper the results of investigating the probability density distribution of instantaneous values of a Polish speech signal are presented. The aim was to obtain data providing an additional criterion for the generation of artificial test signals. The signals are used in objective methods of measuring the quality of Polish speech transmission. The probability density distribution of instantaneous values of a speech signal in Polish was compared with the distributions in English, German, and Russian. The minimum duration of an acoustic signal, sufficient to obtain stationary characteristics of the probability density distribution of instantaneous values of a Polish speech signal, was determined. The distributions obtained were approximated by means of exponential functions.

1. Introduction

Speech is the most common means of transmitting information among people. The quality of the transmission depends on objective physical parameters of the devices used in transmitting speech signals and on subjective factors connected with the transmitter and the receiver of the information transmitted. All measurements of the quality of speech signal transmission should consider subjective factors. Either subjective methods of measurement or the evalution of objective methods should be performed with consideration of subjective factors.

Subjective measurements are laborious, expensive and require a large group of people. They cannot thus be generally and widely used [10]. There is an increasing demand for working objective methods of measuring the quality of speech signal transmission which would provide results similar to the subjective opinion of the users of telecommunication devices.

Most of the methods used or suggested so far are based on a comparison of the parameters of a test signal at the input of the transmission channel with the parameters at the output. As a test signal we usually employ a deterministic signal [7, 8, 13, 14], a Gaussian random signal [3] or a speech signal representing a set of speech sounds, syllables or sentences [10, 13]. In the first two cases, the test signal should approximate the basic characteristics of a speech signal. In the third case, however, the test signal contains the characteristics of a standard speech signal if it is a sample of elements occurring in natural speech. The test signals usually provide consistency of the signal spectrum with the power density of the natural speech spectrum [7, 8, 10, 14].

The results of the investigations in paper [6] show that if

$$f_1(t) \leftrightarrow |F_1(\omega)| \tag{1}$$

and

$$f_2(t) = y(t)\; f_1(t) \leftrightarrow |F_2(\omega)|, \tag{2}$$

where $f_1(t)$ denotes a primary acoustic signal, $|F_1(\omega)|$ is the amplitude spectrum of the signal $f_1(t)$, $f_2(t)$ denotes a secondary acoustic signal, $|F_2(\omega)|$ is the amplitude spectrum of the signal $f_2(t)$, $y(t)$ denotes a function introducing nonlinear distortions of the signal $f_1(t)$, $\leftrightarrow$ denotes the mark of correspondence, then in spite of the fact that

$$\big||F_1(\omega)| - |F_2(\omega)|\big| < \delta, \tag{3}$$

where δ is the least spectral difference perceptible by means of the measuring method used, one reacts to the effect of nonlinear distortions caused by the function $y(t)$ because these distortions cause important changes in the probability density distribution of instantaneous values of the signal. Thus it seems advisable to include in a test signal not only spectral parameters, but also other additional statistical features of a speech signal that are important in signal perception. This concerns mainly the probability density distribution of instantaneous values of a natural speech signal.

Experimental investigations [1, 3] have shown that a speech signal with the statistical features mentioned above may be considered as a stationary ergodic random process $\{X(t)\}$, provided the signal fragment lasts at least several tens of seconds. It follows from the ergodic characteristics that the statistical features of a speech signal and, in particular, the probability density of instantaneous values of a speech signal may be computed by finding the mean value in time of the n-th sample $x_n(t)$ of a random process $\{X(t)\}$. The probability density $p(x)$ of instantaneous values of a random signal is the derivative of the distribution function of the random variable $x(t)$ [2], which may be expressed as

$$p(x) = \frac{dF(x)}{dx} = \frac{P(x \leqslant x(t) < x+dx)}{dx}, \tag{4}$$

where $F(x) = P[x(t) < x]$ is the distribution function of the random variable $x(t)$.

Practically, in analogue methods of signal analysis, the probability density of instantaneous values of a speech signal may be estimated [2] from the formula

$$p(x) = \frac{T_x}{T \Delta x}, \tag{5}$$

where Δx is the finite width of the time interval including x, T_x — the time interval where the signal $x_n(t)$ is within the section with a width equal to Δx, and T — the duration of the analysis $x_n(t)$. In digital methods of signal analysis, $p(x)$ may be estimated from the formula

$$p(x) = \frac{N_x}{N \Delta x}, \tag{6}$$

where N_x is the number of samples in the interval of width Δx and N is the total number of samples.

2. Method of investigation

Measurement of the probability density of instantaneous values of a speech signal in Polish is based on phonetic data obtained from a reading by 11 readers (10 male voices and 1 female voice) of the same newspaper text, at a constant sound intensity level. Long reading at the same level is very tiring, so to avoid changes in sound level intensity caused by readers fatigue, the test signal was recorded in five-minute series. The total time of signal duration for one reader was ten minutes.

A block diagram of the system for measuring the probability density of instantaneous values of a speech signal is shown in Fig. 1. The recordings were

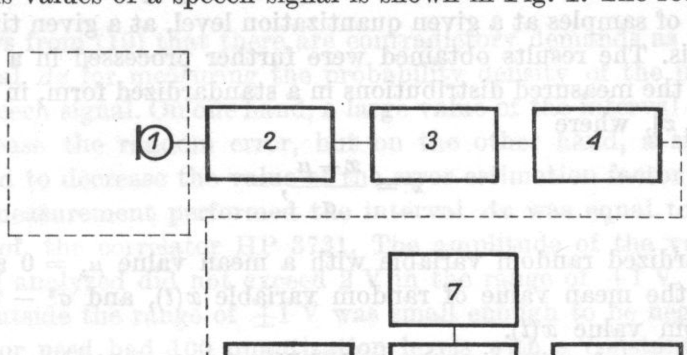

Fig. 1. A block diagram of the system for the measurement of the probability density of instantaneous values of a speech signal

1 – microphone, 2 – amplifier, 3 – bandpass filter (100-6000 Hz, 24 dB/oct.), 4 – tape recorder (v = 38.1 cm/s), 5 – tape recorder (v = 9.05 cm/s), 6 – correlator, 7 – clock, 8 – digital computer

made in an audio monitoring studio at the Institute of Telecommunication and Acoustics of Wrocław Technical University. To record the signal test, a non-directional condenser microphone (Neumann type UM 57) was used with a professional tape-recorder MS 181. The frequency response of the microphone was linear within ± 2 dB from 30 to 15 000 Hz, while the frequency response of the recording system did not exceed ± 1 dB in the frequency band 20-15 000 Hz. The microphone was placed in the nearfield of the speaker. The signal/noise ratio of the whole system was 50 dB.

The speech signal recorded was processed into discrete units by means of a Hewlett-Packard type HP 3721 correlator. The correlator HP 3721 has a constant sampling time of 333 μs [15]. According to the sampling theorem [2], this permits the presentation in digital form of all the information in the analogue signal below an upper limiting frequency of about 1500 Hz. The frequency analysis of the medium spectrum of speech in Polish shows that natural speech contains frequency components over 1500 Hz, but above 6000 Hz they are small enough to be neglected [11]. To perform the analysis of a speech signal in Polish within the range 100-6000 Hz, the frequency of the signal components was diminished four times by decreasing the speed of the magnetic tape by a factor of four. This procedure caused the system dynamic range to decrease by about 2 dB and the frequency response of the readout track did not exceed ± 1 dB in the range 20-10 000 Hz.

The duration of the analysis of speech signal was estimated by means of a clock connected to the correlator. The clock switched the correlator on at the movement of inputting the speech signal. From this moment on the clock measured the time of the signal speech analysis and, having achieved the assigned value of the analysis time, caused cessation of the processing of the signal in the correlator. The result of this processing by the correlator was the total number of samples at a given quantization level, at a given time of speech signal analysis. The results obtained were further processed in a digital computer to give the measured distributions in a standardized form, in a coordinate system $[p(z), z]$, where

$$z = \frac{x-\mu}{\sigma}, \tag{7}$$

z is a standardized random variable with a mean value $\mu_z = 0$ and variance $\sigma_z^2 = 1$, μ — the mean value of random variable $x(t)$, and σ^2 — the variance of the random value $x(t)$.

The accuracy of estimating the probability density of instantaneous values of the speech signal is defined by means of the relative effective error which is the square root of the relative mean square error. The mean square error of the probability density estimation of instantaneous values of a signal is defined [1] as the sum of the variance $D^2[\hat{p}(x)]$ and squared error estimation factor $b^2[\hat{p}(x)]$.

The variance $D^2[\hat{p}(x)]$ of the estimate of the probability density of instantaneous values of a speech signal describes the random part and is defined by the formula

$$D^2[\hat{p}(x)] = \frac{p(x)}{2BT\Delta x}, \tag{8}$$

where

$$p(x) = \lim_{\substack{T\to\infty \\ \Delta x\to 0}} \hat{p}(x)$$

is the probability density of instantaneous values of the speech signal, $\hat{p}(x)$ — the weighted estimate of the function $p(x)$, T — the duration of the speech signal analysis, B — the frequency band width of the speech signal and Δx — a finite interval including x.

The error estimation factor $b[\hat{p}(x)]$ for the instantaneous value of a speech signal describes the systematic part of the error and is defined by the approximate relation

$$b[\hat{p}(x)] \approx \frac{1}{24}\Delta x^2 \ddot{p}(x), \tag{9}$$

where $\ddot{p}(x)$ is the second derivative of the function $p(x)$ with respect to the argument x.

A relative effective error ε of the estimate of the probability density of instantaneous values of the speech signal may be calculated from

$$\varepsilon \approx \left\{\frac{1}{2BT\Delta x p(x)} + \frac{\Delta x^4}{576}\left[\frac{\ddot{p}(x)}{p(x)}\right]^2\right\}^{1/2}. \tag{10}$$

It follows from (10) that there are contradictory demands as to the width of the interval Δx for measuring the probability density of the instantaneous value of a speech signal. On one hand, a large value of the interval Δx is necessary to decrease the random error, but on the other hand, a smaller width Δx is required to decrease the value of the error estimation factor [2].

In the measurement performed the interval Δx was equal to the quantization step of the correlator HP 3731. The amplitude of the voltage of the speech signal analyzed did not exceed 2 V in the range of ± 1 V. The number of samples outside the range of ± 1 V was small enough to be neglected. Since the correlator used had 100 quantization levels with a constant step, in the present investigations the quantization step was 0.02 V.

The accuracy of estimating the probability density of instantaneous values of a speech signal was defined by means of the relative effective error. Examples of its values for the case most dependent on the time T of duration of the speech signal analysis are given in Table 1.

Table 1. The relative error ε [%] for various durations T of speech signal analysis

T [min]	$p(x)$ 0.005	0.01	0.05	0.1	0.5	1.0
1	6.56	4.64	2.08	1.47	0.82	1.85
5	3.08	2.20	0.97	0.69	0.51	1.50
10	2.15	1.50	0.68	0.48	0.48	1.60

3. Results of the experiment

To estimate the minimum duration of the speech signal analysis which is sufficient to obtain a stationary distribution of the probability density of instantaneous values of a speech signal in Polish, distribution measurements for various durations of speech signal analysis were performed. The duration of the analysis started with 1 minute and was then gradually lengthened by 1 minute increments to reach eventually 10 minutes.

The probability density distributions of instantaneous values of a speech signal lasting 1, 5 and 10 minutes are presented in Fig. 2. In this figure the distributions measured for four speakers (three male voices and one female voice) are presented. They are the ones most dependent on the signal duration. In other cases the dependence was only slight, so the analysis was made for durations shorter than 1 minute. It turned out that increasing the duration of the speech signal analysis over 40 seconds caused only slight changes in the probability density distribution of instantaneous values of a speech signal in Polish.

The agreement of the distributions for one speaker and various durations was checked by means of the Smirnov test [4]. It has been shown that at a confidence level of $\alpha = 0.05$ the hypothesis of the agreement of the probability density distributions of instantaneous values of a speech signal in Polish, with test signals lasting at least 60 seconds, cannot be rejected. From these results one may assume that at a given quantization step, those distributions do not change for an analysis lasting at least 60 seconds. Thus, measurement of these distributions should not last less than 60 seconds.

Comparison of the probability density distributions of instantaneous values of a speech signal in Polish for different speakers was performed by means of the Smirnov test. It has been shown that at a confidence level of $\alpha = 0.05$ the population of speakers was homogeneous and the hypothesis of the agreement of the distributions cannot be rejected. Thus the results obtained may help to estimate the mean probability density distribution of instantaneous values of a speech signal in Polish. This distribution is presented in Fig. 3.

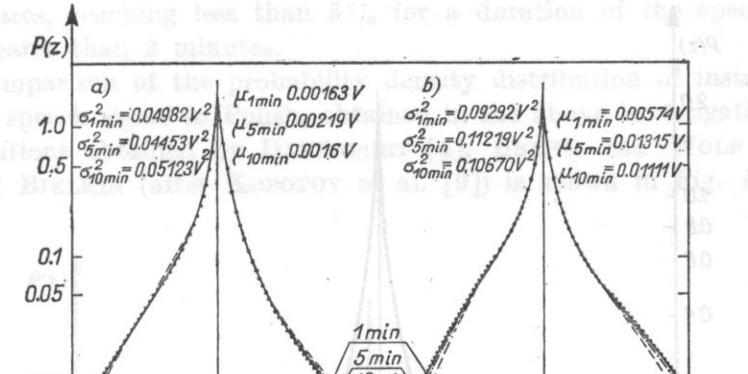

Fig. 2. The probability density distribution of instantaneous values of a Polish speech signal for four speakers and three durations of speech signal analysis
a, b, c – male voices, *d* – female voice

The probability density distribution of instantaneous values of a speech signal in Polish, obtained as a result of the present investigations, may be approximated by

$$p(z) = ae^{-b|z|} + ce^{-d|z|}. \tag{11}$$

Values of the coefficients *a*, *b*, *c* and *d* were found using the least squares method [2, 11]. The formula approximating the empirical probability density distribution of instantaneous values of a speech signal in Polish was obtained in the form

$$p(z) = \frac{0.049}{\sigma_1} e^{-0.194|z|/\sigma_1} + \frac{0.346}{\sigma_2} e^{-1.4|z|/\sigma_2}, \tag{12}$$

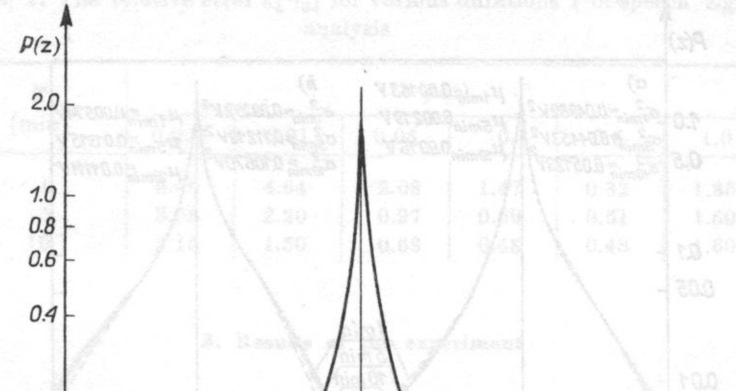

Fig. 3. Mean probability density distributions of instantaneous values of a speech signal in Polish

where $\sigma_1^2 = 0.66\sigma^2$ and $\sigma_2^2 = 0.34\sigma^2$, σ^2 being the variance of the empirical probability density distribution of instantaneous values of a speech signal in Polish.

4. Conclusion

The results obtained show that at a given quantization step the stationary probability density distributions of instantaneous values of a speech signal in Polish are obtained for an analysis duration T of at least 60 seconds. The relative effective error for an analysis time $T = 60$ seconds, and for the method described, does not exceed 7 % for the least favourable case – largest absolute amplitude values. With longer duration of the analysis the relative effective error decreases, reaching less than 5 % for a duration of the speech signal analysis greater than 2 minutes.

The comparison of the probability density distribution of instantaneous values of a speech signal in Polish, obtained in the above investigations, with the distributions obtained by DAVENPORT [4], BREHM and WOLF [3], and SHITOV and BIELKIN (after KOPOTOV et al. [9]) is shown in Fig. 4.

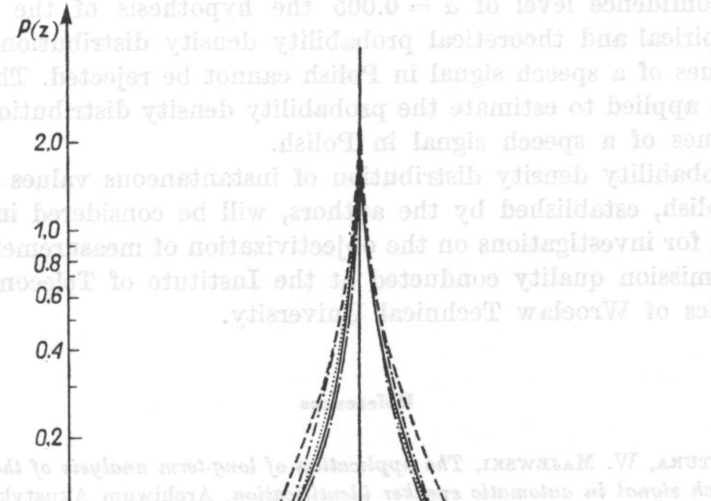

Fig. 4. The probability density distribution of instantaneous values of a speech signal

1 – after Davenport, 2 – after Brehm and Wolf, 3 – after Shitov and Bielkin, 4 – after Brachmański and Majewski, 5 – from relation (12)

To verify the hypothesis of the equivalence of the probability density distribution of instantaneous values of a speech signal in Polish with the distributions for English [4], German [3] and Russian [9], the $\varkappa^2$ consistency test was applied. The $\varkappa^2$ consistency test showed that at a confidence level $\alpha = 0.005$ the hypothesis of the consistency of the probability density distributions of instantaneous values of a speech signal in Polish, compared with the distributions in German and Russian, cannot be rejected.

To verify the empirical consistency of probability density distribution of instantaneous values of a speech signal in Polish with the theoretical distribution calculated from (12), the $\varkappa^2$ consistency test was applied. It was shown that at a confidence level of $\alpha = 0.005$ the hypothesis of the consistency of both empirical and theoretical probability density distributions of instantaneous values of a speech signal in Polish cannot be rejected. Thus, formula (12) may be applied to estimate the probability density distribution of instantaneous values of a speech signal in Polish.

The probability density distribution of instantaneous values of a speech signal in Polish, established by the authors, will be considered in generating a test signal for investigations on the objectivization of measurement of speech signal transmission quality conducted at the Institute of Telecommunication and Acoustics of Wrocław Technical University.

References

[1] Cz. Basztura, W. Majewski, *The application of long-term analysis of the zero-crossing of a speech signal in automatic speaker identification*, Archiwum Akustyki, **13**, 1, 3-15 (1978).

[2] J. S. Bendat, A. G. Piersol, *Methods for the analysis and measurement of random signals*, PWN, Warszawa 1976.

[3] M. Brehm, E. W. Jungst, D. Wolf, *Simulation von Sprachsignalen*, AEU, **28**, 11, 445-450 (1974).

[4] W. B. Davenport, *An experimental study of speech-wave probability distributions* JASA, **28**, 4, 390-399 (1952).

[5] A. Góralski, *Methods of the description and statistical inference in psychology*, PWN, Warszawa 1976.

[6] S. R. Hlibowicki, *Audibility examinations of change of probability density of acoustical signal instantaneous values*, 59th Convention AES, Hamburg 1978.

[7] T. Houtgast, H. J. M. Steeneken, *Evaluation of speech transmission channels by using artificial signals*, Acustica, **23**, 6, 355-367 (1971).

[8] T. Hougast, H. J. M. Steeneken, *Modulation transfer function in room acoustics as a predictor of speech intelligibility*, Acustica, **28**, 1, 66-73 (1973).

[9] P. G. Kopotov, J. P. Maksimov, W. M. Markov, *Nielinieinie iskazhenia riechevovo signala*, Elektrosviaz, **8**, 29-34 (1976).

[10] W. Majewski, Cz. Basztura, S. P. Brachmański et al., *Objectivization of measurements methods of articulation intelligibility*, Report No. I/28/R-199/77, Wrocław 1977.

[11] L. Z. Rumszycki, *Mathematical elaboration of the results of experiments*, WNT, Warszawa 1973.

[12] J. Zalewski, W. Majewski, *Polish speech spectrum obtained from superposed samples and its comparison with spectra of other languages*, Proc. of 7th International Congress on Acoustics, Budapest 1971.

[13] *Acoustics – recommended methods for measuring the intelligibility of speech,* Draft ISO/DP 4870, 1975.
[14] Green Book of CCITE, *Wire transmission,* III-1, WKiŁ, Warszawa 1976.
[15] *Service Manual Model 3721 Correlator,* Hewlett-Packard 1971.

Received on April 13, 1978