ARCHIVES OF ACOUSTICS
12, 3-4, 159-175 (1987)

SIMILARITY FUNCTIONS OF ACOUSTIC PATTERNS AS INDICES OF THE OBJECTIVE QUALITY EVALUATION OF SPEECH TRANSMISSION

CZESŁAW BASZTURA

Institute of Telecommunication and Acoustics
Wrocław Technical University
(50-317 Wrocław, ul. B. Prusa 53/55)

This study is concerned with a new, objective quality evaluation of speech transmission through telephone lines. The conception of the studies is based on the assumption of equivalency between the results of subjective and objective measurements. A special purpose of the study was to investigate the efficiency of the similarity function of the acoustic patterns of signals as an indice of the objective quality evaluation of speech transmission. Altogether, the usefulness of 8 functions of similarity were investigated, such as the Hamming, Euclid, Minkowski, Tanimoto, Chebyshev, Camberra distances, χ-square and directional cosine. The obtained preliminary results indicate for some similarity functions good agreement with subjective results obtained by the subjective method.

1. Introduction

The telephone communication is the fundamental kind of speech-based communication. The development of telephone networks, accompanied by the simultaneous increase in the quality requirements set by users of these systems, makes more and more distinct the necessity of elaborating effective methods for quality measurements of telephone systems and lines. The quality and faithfulness of transmission of messages by telephone links depend both on subjective factors brought about by users and also on the objective physical parameters of tele-transmission systems.

Keeping in with CCITT recommendations, the national and international telephone systems are standardized, in terms of their quality, according to the criterion of loudness [21, 22].

Another important criterion of the quality evaluation of telephone transmission is the clearness or intelligibility as measured by the articulation method [16, 17]. The clearness measurement by the articulation method is carried out by a team of operators. The measurement consists in the transmission and detection of appropriate articulation lists made up from elements of speech (logatoms, words). On the basis of these two basic criteria of loudness and clearness, many methods have been developed, which are objectivized to a greater or lesser extent [1, 2, 3, 7, 9, 11, 12, 16, 17, 18, 19]. Most of them involve numerous faults preventing their use on a larger scale.

The difficulties occurring in the solution of the objectivization problem consist in the fact that the objective measurement system should take into account not only the main characteristics of the standard (reference) system, the physical characteristics of the speech signal, but also the physiological properties of hearing, determining the subjective sensation of loudness or clarity (intelligibility). Since the measurement methods concerned with the quality of speech transmission which are based on the criterion of clearness, or that of intelligibility, are the methods which ensure better agreement with users' evaluations, most objective methods, including the present one, are now devoted to the objectivization of measurements, taking as the reference (comparative) measure subjective measurements of clearness [1, 6, 9, 17].

The complexity of the problem of objectivization of subjective evaluations of clearness or intelligibility implies a different approach to its solution. This is manifested, e.g., in the applied method, in the number and kind of studied interferences, in the ways of the signal parametric representation, measurement methods and algorithms of objective estimation.

T. B. Barnwell, in studies [1, 2, 3], carried out investigations on the effective objective measure correlated with subjective evaluations. He studied a relatively wide spectrum of distortions arising in both discrete transmission (APCM, ADPCM, LPC etc), and analogue transmission (band filtration, noise interference, clippering, echo etc.). As the subjective measure, Barnwell applied the DAM measure (Diagnostic Acceptability Measure) [19]. As objective measures, he used 4 classes of measures:

– the measure of spectral distance, where the distortions were measured in the domain of frequency;

– the parametric measures of distance, where the parametric distances were measured between the patterns of distorted and undistorted speech signals;

– noise measures used for noised signals;

– complex measures which were a linear combination of other measures.

The results which Barnwell obtained ensured a large degree of agreement between objective and subjective measures. Their basic fault is the complicated character of measurement methods and the resulting difficulty involved in applying them in more general operations.

Among other studies, one should mention, e.g., those by W. J. Hartmann [9], L. R. Rabiner [11], D. L. Richards [12] and R. Viswanathan et al. [18].

In most general terms, objective methods for the quality evaluation ef speech transmission can be divided into:

– methods using a test (natural or artificial) signal, and

– methods which do not a test signal.

Previously, most methods for the quality evaluation of speech transmission were based on the use of a test signal as a carrier of information about the degree of signal distortion. Moreover, this signal was a natural or artificial one with the statistics of certain parameters being close to the natural-speech signal [1, 2, 3, 9, 11, 18]. Some papers were concerned with calculation methods based on the results of physical measurements of the parameters of a telephone line. On this basis, it was possible to determined objective measures as correlated with subjective reference measures [7, 12].

In elaborating an objective method for the quality evaluation of the transmission of a speech signal, one should bear in mind the fact that subjective clearness measurements can be replaced only if the principle of equivalency between the results of subjective and objective measurements is satisfied. The comparative basis for this equivalence is the subjective measure, which is understandable if one takes into account the fact that human beings are users of the studied tele-transmission systems. The way of the articulation of transmitted signals and also the perception of these signals by hearing organs also exert an influence on the quality evaluation of transmission. To make the results of subjective measurements to depend to the largest extent only on the physical parameters of a telephone line, logatom lists are used as the linguistic test material, ensuring the elimination of semantic information [20].

The performed studies were based on the a priori assumption saying that there exists some correlation between the objective measure M_r and the subjective measure of the quality evaluation of speech transmission M_s. As the objective measure M_r, the probability degree between the test signal at the output from the investigated line and the signal assumed as a standard one was assumed (Fig. 1). In particular, the

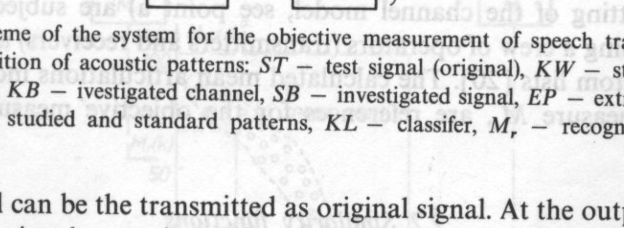


Fig. 1. Block scheme of the system for the objective measurement of speech transmission using the measure of recognition of acoustic patterns: ST – test signal (original), KW – standard channel, SW – standard signal, KB – ivestigated channel, SB – investigated signal, EP – extractor of parameters, OB, OW – studied and standard patterns, KL – classifer, M_r – recognition measure

standard signal can be the transmitted as original signal. At the output of the studied line, the test signal contains information about distortions and interferences introduced by telephone lines, having an effect on the transmission quality, i.e., on the evaluation of the quality of the line.

In investigations of this type, the fundamental problem is to separate out such information; in other words, to create specific acoustic patterns of a signal at the output of the studied line OB and to compare it with the standard pattern OW. Some of the indices serving in determining the mutual relations of the images OB and OW are known as similarity functions [4, 8, 14], applied in the automatic recognition of speech and voices.

2. Research method

2.1. Main assumptions

For further considerations, the following basic assumptions and definitions were presumed:

a) the investigated objects are typical telephone lines alternatively called telephone channels, represented to ensure the repeatability of subjective and objective measurements, by an adjustable physical model of a telephone channel [5];

b) the reference criterion of the quality of speech transmission, and at the same time the quality criterion of the k-th telephone channel, is the logatom articulation index measured according to ISO recommendations [20];

c) the objective measurement (estimation) of articulation is the procedure aimed at some equivalent transformation determined from the set K of channels (in fact, from sets of measurable groups of parameters describing these channels), or otherwise, patterns (x, y) and the values of the indices M_r;

d) it was assumed that the principle of maximum agreement of physical conditions should be applied with objective measurements and the existing conditions for subjective reference measurements;

e) the principle was assumed that resulting from the previous condition the test signal will be a signal of natural speech.

The strategy of the performed results is shown in Fig. 2. In the objective part, there occurs the process of calculating the recognition measure M_r namely the similarity function of the standard patterns OW and the investigated ones OB. Particularly, the recognition measure M_r can be determined in normalized form (maximum similarity = identity $\rightarrow M_r = 100\%$). In the subjective part, the same links (for the setting of the channel model, see point a) are subject to subjective measurements using a crew of operators (transmitters and receivers) and by means of appropriate logatom lists [20]. The calculated mean articulations index, assumed as the subjective measure M_s, are references for the objective measure M_r.

2.2 Similarity functions

Similarity functions serve in determining the relation between two patterns, i.e. the standard pattern of the signal OW and the investigated image OB. Similarity functions can have the form of the distance function $d^\alpha(x, y)$ or of the nearness function $b^\alpha(x, y)$ (where α denotes the kind of the possibility function). Between the distance and nearness functions, with some assumptions, there exist unique relations, such as

$$b^{\alpha}(x, y) = \frac{1}{d^{\alpha}(x, y)} \quad d^{\alpha}(x, y) \neq 0, \tag{1}$$

$$b^{\alpha}(x, y) = \exp[-d^{\alpha}(x, y)]. \tag{2}$$

In automatic speech recognition, the distance functions, also called briefly distances, are more broadly used. Below, expressions are given for the distance and nearness functions used in the studies.

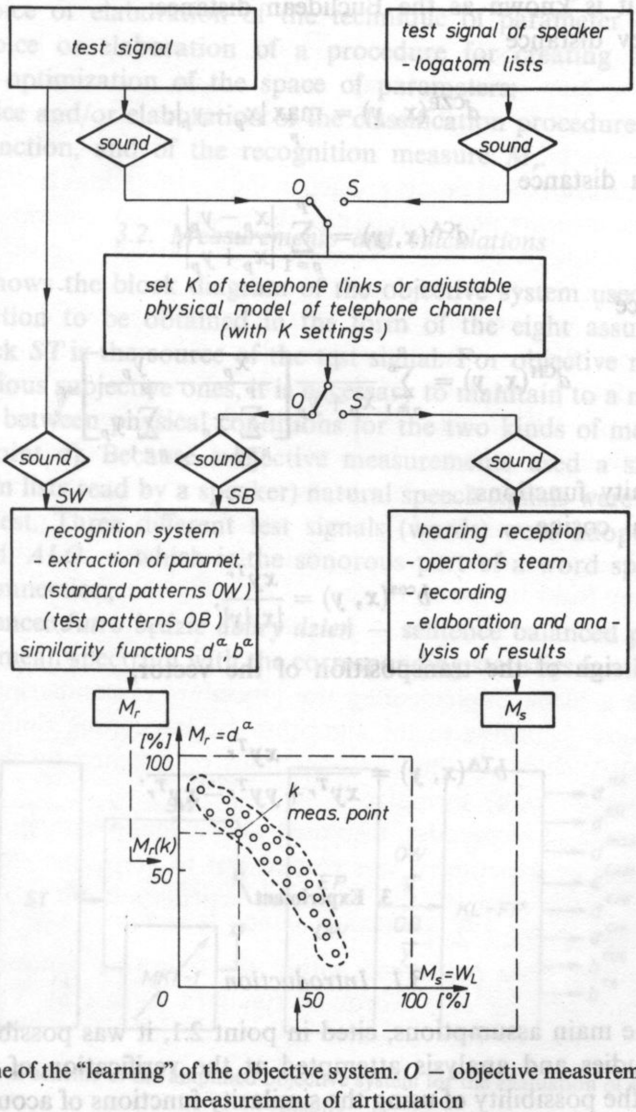


Fig. 2. Block scheme of the "learning" of the objective system. O — objective measurements, S — subjective measurement of articulation

Distance functions:

a) Minkowski distance

$$d^{\mathrm{MI}}(x, y) = \Big[\sum_{p=1}^{P} |x_p - y_p|^r\Big]^{1/r} \qquad r \geqslant 1, \tag{3}$$

where $p = 1, 2, \ldots, P$, where P is the dimension of the parameter vector constituting the image, x_p is the p-th component of the vector which is the standard pattern (OW) an y_p is the p-th component of the vector which is the investigated pattern (OB) for $r = 1$ $d^{\mathrm{MI}}(x, y)$ it is known as the Hamming, or street distance [15], and for $r = 2$, it is known as the Euclidean distance;

b) Chebyshev distance

$$d^{\mathrm{CZE}}(x, y) = \max_p |x_p - y_p|, \tag{4}$$

c) Camberra distance

$$d^{\mathrm{CA}}(x, y) = \sum_{p=1}^{P} \frac{|x_p - y_p|}{|x_p + y_p|}, \tag{5}$$

d) χ^2-distance

$$d^{\mathrm{CH}}(x, y) = \sum_{p=1}^{P} \frac{1}{x_p + y_p}\left[\frac{x_p}{\sum_{p=1}^{P} x_p} - \frac{y_p}{\sum_{p=1}^{P} y_p}\right], \tag{6}$$

and the proximity functions:

e) directional cosine

$$b^{\cos}(x, y) = \frac{xy^{T_r}}{|x|\,|y|}, \tag{7}$$

where T_r is the sign of the transposition of the vector.

f) Tanimoto

$$b^{\mathrm{TA}}(x, y) = \frac{xy^{T_r}}{xy^{T_r} + yy^{T_r} - xy^{T_r}}. \tag{8}$$

3. Experiment

3.1. *Introduction*

By taking the main assumptions, cited in point 2.1, it was possible to carry out experimental studies and analysis attempted at the verification of the hypothesis concerned with the possibility of using the similarity functions of acoustic patterns in the objective evaluation of the speech transmission quality. The first, at the same time the fundamental, problem was the creation of the space of observations (the space of parameters) which was representative from the point of view of the information of interest. In the process of evaluation of the speech transmission quality, the space of observations should contain information about the degree of distortion and interference of the transmitted speech signal. In the carrying out of the experiment, the first step consisted in choosing the appropriate test signal.

The other steps included:

a) the choice of physical quantities representing the desired set of information;

b) the choice of a set (or sets) of parameters of the signal;

c) the choice or elaboration of the technique of parameter extraction;

d) the choice or elaboration of a procedure for creating acoustic patterns including the optimization of the space of parameters;

e) the choice and/or elaboration of the classification procedures and the related probability function, and of the recognition measure M_r.

3.2. *Measurements and calculations*

Figure 3 shows the block diagram of the objective system used, permitting the similarity function to be obtained in the form of the eight assumed probability functions. Block *ST* is the source of the test signal. For objective measurements to replace the tedious subjective ones, it is necessary to maintain to a maximum degree the agreement between physical conditions for the two kinds of measurements (see chapter 2.1, point d). Because subjective measurements used a signal of natural speech (logatom lists read by a speaker) natural speech signals were also included in the objective test. Three different test signals (words) were adopted;

a) the word: *ALO* – which is the sonorous part of a word spoken in making a telephone connection,

b) the sentence: *Jutro będzie dobry dzień* – sentence balanced phonetically and in terms of the mean spectrum with the corresponding statistics for the Polish speech [10],

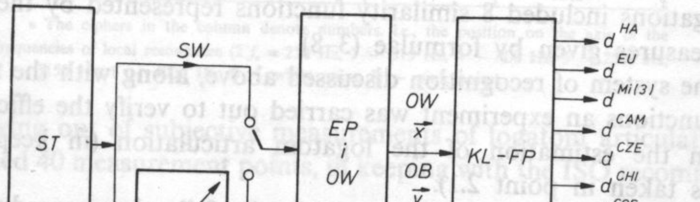


Fig. 3 Simplified block scheme of the amplified objective system for the evaluation of speech transmission quality

c) the sentence: *Sprawdzenie przydatności funkcji podobieństwa do oceny jakości transmisji sygnału mowy.*

The above mentioned test signals were recorded by 3 male speakers (with each speaker recording all 3 words). MKT — 1 is a physically adjustable model of a telephone link permitting (in repeatable way) the setting of various values of the main sets of parameters for typical telephone links [5, 21, 22]. In the further part of this paper, these sets of parameters will be called measurement points. The investigations included 40 measurement points which represented just as many corresponding typical telephone links. The sets of parameters contained the following quantities:

– the frequency band;
– the signal-to-noise ratio (white and pink noise): [s/noise];
– the signal-to-hum ratio; [s/hum]; and
– different types of the amplitude response of the channel the position of local resonances [corrector].

The full set of measurement points applied in the investigations is shown in Table 1.

In choosing the measurement points, e.g., the principle of uniform distribution of these points was followed so that it could cover as broad as possible a range of logatom articulation the (10–97%, see Table 1).

EP and *OW* is a measurement block extracting selected physical parameters of the standard signal *SW* and the investigated one *SB*, and forming the standard patterns *OW* of data in the form of the vector of parameters $\mathbf{x}$ and that of the investigated patterns *OB* of data in the form of the vector of parameters $\mathbf{y}$.

In the present investigations, the extractor of the parameters $\mathbf{x}$ and $\mathbf{y}$ was a system implementing the procedure of measurement of the distribution of time intervals between zero crossings of a speech signal. This kind of parametrization of the signal was chosen because of its simple implementation in a real time and the sufficient efficiency from the point of view of carrying the characteristic features of a speech signal, namely distortions, interference, individual features etc. [4, 6, 7, 8].

KL — FP is a block implementing the procedure of comparing standard and investigated images according to the algorithm for calculating similarity functions. The investigations included 8 similarity functions represented by the distance and nearness measures given by formulae (3–8).

Using the system of recognition discussed above, along with the assumed eight similarity functions an experiment was carried out to verify the efficiency of these functions in the estimation of the logatom articulation (in keeping with the assumptions taken in point 2.1).

The investigatons were performed within the following procedure:

a) the casette recording of 3 words of standard signals by 3 speakers;

b) the transmission of the recorded signals by an adjustable model of the telephone channel MKT – 1, i.e., by 40 settings of it (measurement points — see Table 1);

Table 1. Set of parameters for measurement points used in the experiment

K	band boundary frequency [Hz]	S/noise [dB]	S/hum [dB]	corrector * ch-ki·	logatom articulation W_L [%]
1	2	3	4	5	6
1	200-4000	> 40dB	40	-	96.3
2	200-4000	B+12	40	2,3,7,8	86.8
3	200-4000	B+6	40	2,3,7.8	78.2
4	200-4000	B+3	40	2,3,7,8	70.4
5	200-4000	B 0	40	2,3,7.8	64.1
6	200-4000	B-3	40	2,3,7.8	32.4
7	300-4000	B+9	40	—	74.3
8	300-4000	B+3	40	—	83.7
9	300-4000	B-3	40	—	41.6
10	300-4000	B-9	40	—	29.2
11	600-2000	>40	40	—	93.4
12	600-2000	B+12	40	—	83.7
13	600-2000	B+6	40	—	77.2
14	600-2000	B+3	40	—	56.1
15	600-2000	B 0	40	—	58.5
16	600-2000	B-3	40	—	36.7
17	400-2500	B+15	40	—	85.6
18	400-2500	B+6	40	—	72.2
19	400-2500	B+3	40	—	699.5
20	400-2500	B 0	40	—	53.8
21	400-2500	B-2	40	—	51.4
22	400-2500	B-4	40	—	42.4
23	400-2500	B-6	40	—	19.1
24	400-2500	B-9	40	—	18.4
25	400-2500	B-12	40	—	9.1
26	400-2500	R-3	40	—	27.9
27	400-2500	R-6	40	—	55.4
28	400-2500	R-9	40	—	26.2
29	400-2500	>40	40	—	80.8
30	400-2500	>40	40	—	82.9
31	400-2500	>40	+3	3,4,6.7	84.5
32	400-2500	B+12	+6	3,4,6.7	64.2
33	400-2500	B+12	+3	3,4,6,7	70.7
34	400-2500	B+6	-2	3,4.6,7	48.7
35	400-2500	B+6	-3	3,4,6,7	62.0
36	400-2500	B+6	+6	3,4,6,7	59.0
37	400-2500	B+3	+6	3,4.6,7	58.8
38	400-2500	B+3	+3	3,4.6.7	65.2
39	400-2500	B-3	-3	—	32.1
40	400-2500	R-6	0	-	19.9

* The ciphers in the column denote numbers, i.e., the position on the axis of the frequencies of local resonances ($2\ f_0 = 250$ Hz, 3 — 315 Hz, 4 — 400 Hz, 5 — 2500 Hz, 7 — 3150 Hz, 8 — 4000 Hz, B — white noise, R — pink noise).

c) the carrying out of subjective measurements of logatom articulation for the above-mentioned 40 measurement points, in keeping with the ISO recommendations [20];

d) the measurement of the parameters distributions of time intervals between the zero — crossings of a speech signal of standard signals and those sent over the MKT — 1 and the formation of the vectors

$$X = \{x_1, x_2, \ldots, x_p \ldots x_p\}^T r, \qquad P = 7. \tag{9}$$

Table 2. Values of the components of the vector of standard signals x and the patterns y of the measurement points. Patterns of expressions of three speakers for the third word, k – channel number, W_L – articulation measured subjectively

k		$x_{k,0}$							
		$x_{k,1}$ $y_{k,1}$	$x_{k,2}$ $y_{k,2}$	$x_{k,3}$ $y_{k,3}$	$x_{k,4}$ $y_{k,4}$	$x_{k,5}$ $y_{k,5}$	$x_{k,6}$ $y_{k,6}$	$x_{k,7}$ $y_{k,7}$	W_L [%]
standard k=0	1	1448	411	163	266	444	206	48	100
	2	2077	661	234	380	299	40	18	
	3	2139	645	253	461	428	108	32	
1	1	1365	681	251	335	471	243	28	96
	2	2262	1191	311	416	361	62	80	
	3	2294	1087	298	507	464	64	41	
2	1	1287	1420	544	511	189	170	12	93
	2	1633	1576	427	473	102	29	6	
	3	1629	1666	580	698	108	54	11	
3	1	2574	1089	347	368	420	181	2	86
	2	2551	1865	442	450	304	39	1	
	3	2591	1812	453	530	409	38	1	
4	1	1796	1487	469	413	269	109	3	78
	2	2712	2362	635	731	122	52	6	
	3	2796	2341	615	690	199	62	2	
5	1	2032	1609	518	413	269	27	9	69
	2	2993	2681	731	409	122	12	6	
	3	3029	2591	690	503	199	34	12	
6	1	2317	2041	646	418	226	28	4	58
	2	3081	2912	850	398	92	12	2	
	3	3185	2873	815	500	175	12	3	
7	1	2514	2344	753	381	134	16	2	32
	2	3223	3097	872	364	62	11	3	
	3	3318	3145	386	468	123	11	2	
8	1	2022	1479	473	415	380	31	12	74
	2	2801	2104	569	478	242	3	2	
	3	2823	1966	508	585	355	11	7	
9	1	2297	1818	560	397	252	111	19	61
	2	3029	2583	683	408	143	71	21	
	3	2950	2398	671	570	244	84	40	
10	1	2643	2342	679	338	119	20	4	41
	2	3355	2979	829	321	61	11	2	
	3	2980	880	438	102	16	7		

As an example, Table 2 gives values of the standard vector for 3 speakers, the 3rd word and control vectors (for 10 measurement points);

e) the setting in motion of the programme of the classifer $KL - FP$ reading in the vectors $\boldsymbol{x}$, and the calculation of 8 similarity functions and the polynomial approximation for the calculated similarity functions and the values of the subjectively measured logatom articulation W_L;

f) analysis of the results.

The calculations of the similarity functions and the polynomial approximation were implemented on MC Odra 1304, using purpose-constructed subprogrammes.

The first of the elaborated subprogrammes implemented the calculations of the distance between the investigated vectors $\boldsymbol{y}$ from the standard one $\boldsymbol{x}$.

The other subprogramme implemented the polynomial approximation of points lying on the plane: the logatom articulation $W_L = M_s$, given by the distance function $d^\alpha = M_r$. The least squares method was used by polynomials of different degrees, from 1 to 5. In this way, for each word and each similarity function, the following dependence was obtained:

$$W_L = [s^\alpha(x, y)]. \quad (10)$$

Another subprogramme made it possible to represent graphically the inverse dependence

$$d^\alpha(x, y) = f(W_L) \quad (11)$$

approximating too, the points by least squares method by polynomials with degrees of 1 to 5. Dependencies (10) and (11) make it possible to verify the degree of agreement of subjective measurements ($W_L = M_s$) and objective ones ($d^\alpha = M_r$)

4. The research results and conclusions

The performed experimental results, calculations and analyses made it possible to obtain quite an ample set of results. The chosen characteristic results will be presented in greater detail, including the functional relations $W_L = f(d^\alpha)$, making it possible to evaluate the usefulness of the studied functions d^α.

Figures 4 and 5 show some typical graphic dependencies on the function approximating the dependence $W_L = f(d^\alpha)$ for chosen kinds of distances and for words 1 and 3. Comparing the usefulness of the particular functions as arguments facilitating the mapping of the logatom articulation W_L dependence 10 one can state that:

a) The Hamming distance gives good mapping W_L for all test signals, particularly for the approximation with a third-degree polynomial.

E.g., for the word "ALO"

$$W'_L = 0.929\cdot10^2-0.121\cdot10^{-1}d^{\mathrm{HA}}+0.151(d^{\mathrm{HA}})^2-0.631\cdot10^{-10}(d^{\mathrm{HA}})^3. \quad (12)$$

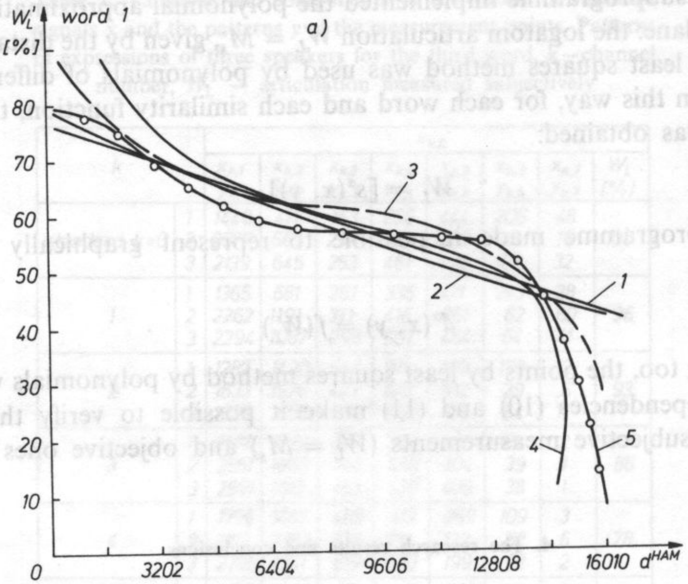
b) For Euclidean and Minkowski distances ($r = 3$)(MI(3)) the best agreement in the mutual mapping between M_r and M_s was obtained just as for the Hamming distance for the longest word and for all higher degrees of the approximating polynomials. For example, for the third word and third degree of the polynomial it was obtained that

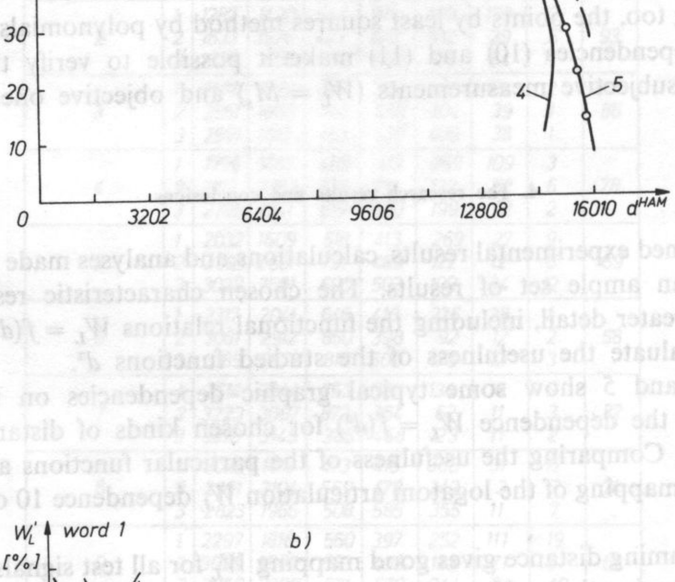
$$W'_L = 90.2-0.184\cdot10^{-1}d^{\mathrm{EU}}+0.364\cdot10^{-5}(d^{\mathrm{EU}})^2-0.239\cdot10^{-9}(d^{\mathrm{EU}})^3, \quad (13)$$

and

$$W'_L = 89.6-0.203\cdot10^{-1}(d^{\mathrm{MI(3)}})+0.448\cdot10^{-5}(d^{\mathrm{MI(3)}})^2-0.327\cdot10^{-9}(d^{\mathrm{MI(3)}})^3. \quad (14)$$

c) The Chebyshev distance, just as the Minkowski distance, gives good unique


W_L'
word 1
[%]
a)
80
70
60
50
40
30
20
10
0
3202
6404
9606
12808
16010
d^{HAM}
1
2
3
4
5


W_L'
word 1
[%]
b)
80
70
60
50
40
30
20
10
0
2024
4047
6071
8094
10118
d^{EUK}
1
2
3
4
5

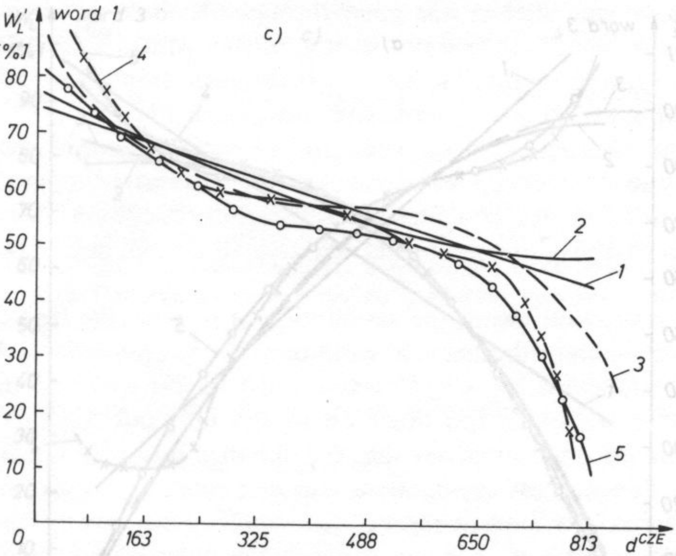


Fig. 4. Plots of functions approximating the logaton articulation $W_L = f(d^\alpha)$ for word 1 and polynomial degrees a) Hamming distance, b) Euclidean distance, c) Chebyshev distance

mapping W_L, particularly for the third word and higher degrees of the approximating polynomial.

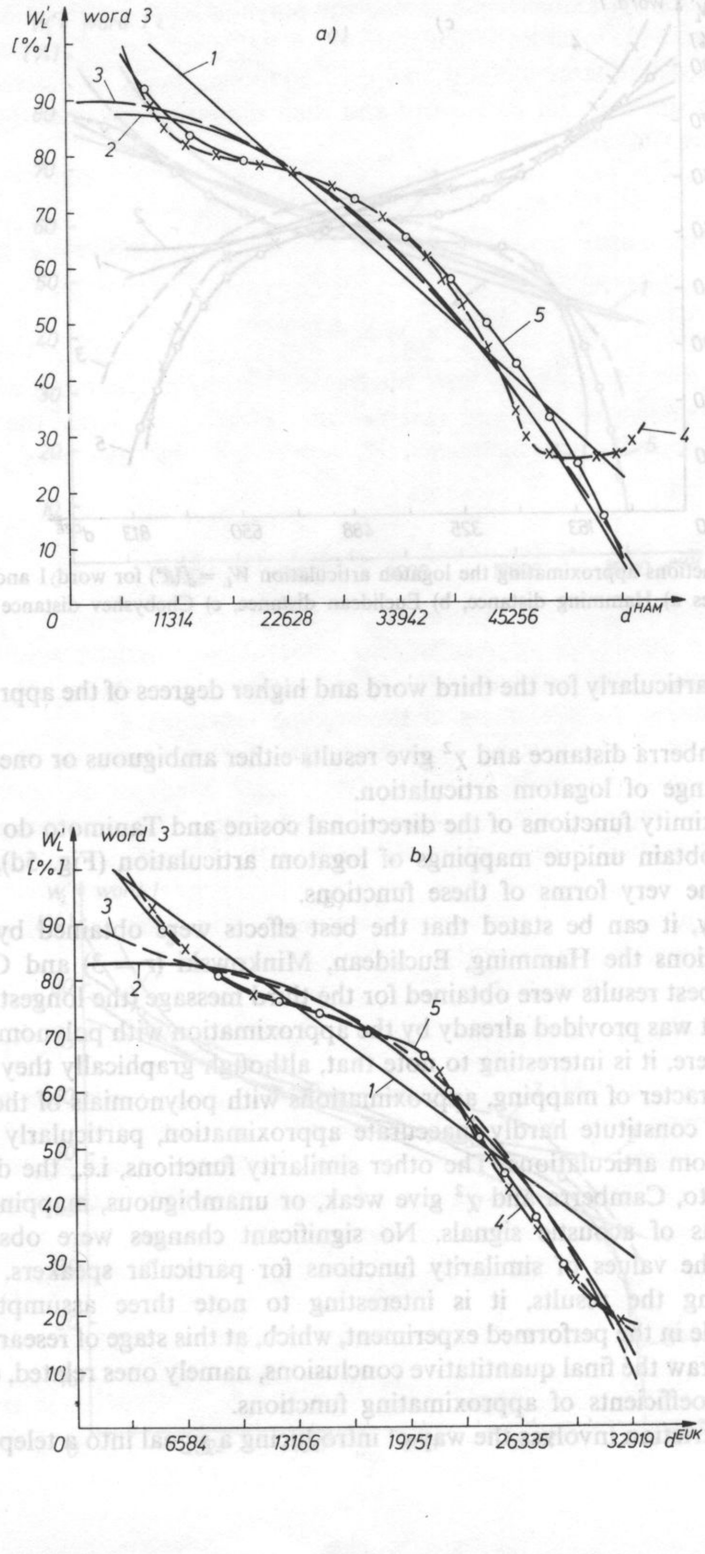
d) The Camberra distance and χ^2 give results either ambiguous or ones covering too small a range of logatom articulation.

e) The proximity functions of the directional cosine and Tanimoto do not make it possible to obtain unique mappings of logatom articulation (Fig. 5d), resulting mainly from the very forms of these functions.

In summary, it can be stated that the best effects were obtained by using as similarity functions the Hamming, Euclidean, Minkowski ($r = 3$) and Chebyshev distances. The best results were obtained for the third message (the longest one), and good agreement was provided already by the approximation with polynomials of the third degree. Here, it is interesting to note that, although graphically they represent the desired character of mapping, approximations with polynomials of the first and second degrees constitute hardly inaccurate approximation, particularly for small and large logatom articulations. The other similarity functions, i.e., the directional cosine, Tanimoto, Camberra and χ^2 give weak, or unambiguous, mappings for the applied patterns of acoustic signals. No significant changes were observed for differences in the values of similarity functions for particular speakers.

In evaluating the results, it is interesting to note three assumptions and restrictions made in the performed experiment, which, at this stage of research do not permit one to draw the final quantitative conclusions, namely ones related, e.g. to the values of the coefficients of approximating functions.

The first restriction involves the way of introducing a signal into a telephone link


W'L [%]
word 3
a)
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
11314
22628
33942
45256
dHAM
W'L [%]
word 3
b)
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6584
13166
19751
26335
32919
dEUK

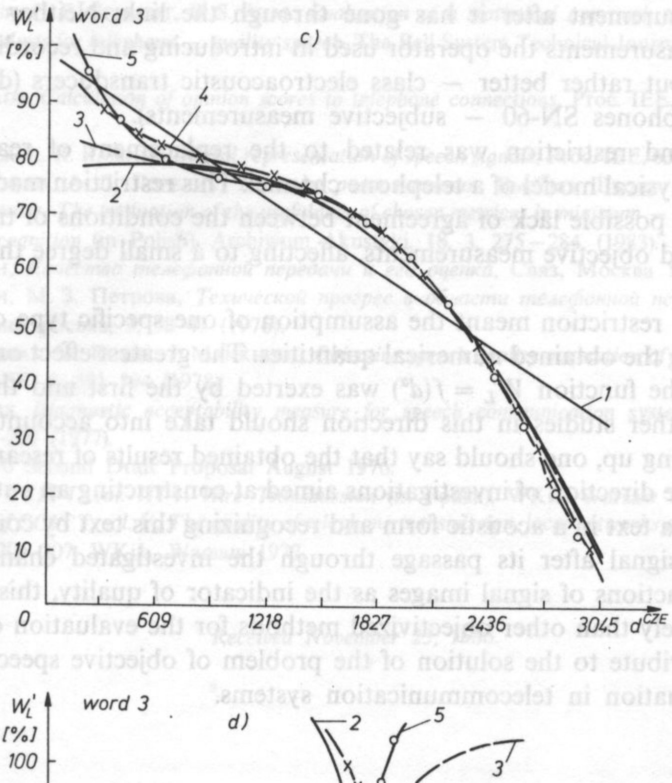


Fig. 5. Plots of functions approximating the logatom articulation $W_L = f(d^\alpha)$ for word 3 and polynomial degrees 1-5: a) Hamming distance, b) Euclidean distance, c) Chebyshev distance, d) Tanimoto proximity function

and its measurement after it has gone through the link. Neither subjective nor objective measurements the operator used in introducing and recording test signals, telephones, but rather better – class electroacoustic transducers (dynamic microphone, headphones SN-60 – subjective measurements).

The second restriction was related to the replacement of real links by an adjustable physical model of a telephone channel. This restriction made it possible to eliminate the possible lack of agreement between the conditions of transmission for subjective and objective measurements, affecting to a small degree the results of the studies.

The third restriction meant the assumption of one specific type of parametrization, affecting the obtained numerical quantities. The greatest effect on the numerical values and the function $W'_L = f(d^\alpha)$ was exerted by the first and third restriction, therefore further studies in this direction should take into account these aspects.

In summing up, one should say that the obtained results of research confirm the validity of the direction of investigations aimed at constructing an automatic devices transmitting a text in a acoustic form and recognizing this text by comparing it with a standard signal after its passage through the investigated channel. Using the similarity functions of signal images as the indicator of quality, this method could more effectively than other objectivized methods for the evaluation of transmission quality contribute to the solution of the problem of objective speech transmission quality evaluation in telecommunication systems.

References

[1] T. P. Barnwell, *Objective measures for speech quality testing*, JASA, **66**, 6, 1658–1663 (1979).

[2] T. P. Barnwell, *Correlation analysis of subjective and objective measure for speech quality*, Proc. IEEE Int. Conf. on Acoustic Speech and Signal Proc., New York 1980.

[3] T. P. Barnwell, *Comparison of parametrically different objective speech quality measures using correlation analysis with subjective quality results*, Proc. ICASSP 80, New York 1980.

[4] Cz. Basztura, *Automatic speaker recognition by zero-crossing analysis of the speech signal*, Chapt.: Man machine communication In: Speech Analysis and Synthesis, 5.I. Polish Academy of Sciences, IPPT PAN, 9–40, Warszawa 1980.

[5] Cz. Basztura et. al., Model of the telephone channel MKT–1 in Polish, Report I-28/R-139/79, Wrocław Technical University 1978.

[6] Cz. Basztura, *Automatic recognition of sound patterns as the method for the evaluation of speech transmission quality in telecommunication channels* (in Polish), Biuletyn Techniczny Ministerstwa Łączności 1, 1982.

[7] Cz. Basztura, *Objective statistical estimation of the articulation of speech transmitted through telephone links* (in Polish), Proc. of XXX Open Seminar on Acoustics, Gdańsk 503–506, 1983.

[8] Cz. Basztura, W. Majewski, *The effect of chosen parameters of a telephone channel on voice identification* (in Polish), Archiwum Akustyki, **16**, 4, 405–416 (1981).

[9] W. J. Hartman, *Objective performance measures for voice systems*, 2nd Symp. a Techn. Exhibition on Electromagn. Compact Proc. IEEE, 521–524 Montreux, June 1977.

[10] W. Majewski, *A phonetic test for subjective measurements of reference attenuation* (in Polish), Przegląd Telekomunikacyjny, **8**, 237–240 (1979).

[11] L. R. Rabiner, C. E. Schmidt, B. S. Atal, *Evaluation of a statistical approach to voices univocal – silence analysis for telephone – quality speech*, The Bell System Technical Journal, **56**, 3, 455–481 (1977).

[12] D. I. Richards, *Calculation of opinion scores to telephone connections*, Proc. IEE, **121**, 5, 313–323 (1974).

[13] R.W. Schafer, L. R. Rabiner, *Digital representation of speech signals*, Proc. IEE, **63**, 662–675 (1975).

[14] В. А. Скрипкин, А. Л. Горелик, *Методы розпознавания*, Высшая Школа, Москва 1977.

[15] R. Tadeusiewicz, *The evaluation of the usefulness of chosen metrices in minimum – distance methods of speech recognition* (in Polish), Archiwum Akustyki, **18**, 3, 275–284, (1983).

[16] Г. В. Вемян, *Качество телефонной передачи и его оценка*, Свяэ, Москва 1970.

[17] Г. В. Вемян, М. З. Петрова, *Техической прогрес в области телефонной передачи мечевых сигналов*, Електросвяз, 3, 38–41 (1976).

[18] R. Visawanthan, W. Russel, J. Makkhoul, *Objective speech quality evaluation of narrowband LPC Vocoders*, IEEE, **6**, 591–594 (1978).

[19] W. D. Voiers, *Diagnostic acceptability measure for speech communication systems*, Proc. IEEE ICASP, 204–207 (1977).

[20] ISO/DP 4870 Second Draft Proposal August 1976.

[21] *Green Book CCITT, vol. III-1: Wire Transmission* (in Polish), WKiŁ Warsaw 1971.

[22] *Green Book of CCITT, vol. V: The quality of telephone transmission, local networks and telephone sets* (in Polish), 204–207, WKiŁ, Warsaw 1977.

Received November 25, 1986.