ARCHIVES OF ACOUSTICS
13, 3–4, 205–218 (1988)

AUTOMATIC VOICE RECOGNITION IN OPEN SETS

CZESŁAW BASZTURA, WOJCIECH MAJEWSKI, JERZY JURKIEWICZ

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

Every system of automatic voice recognition can be divided into three parts: the voice source, the measuring system (the system of parameter extraction) and the classifier. The object of interest of the present paper is the classifier, in which emphasis is laid upon the procedure capable of recognizing voices in open sets. The methodology of investigations, analysis of the problem and the recognizing algorithm has been presented. Also the experimental results accounting for choice optimization and extraneous voice discrimination as well as the problem of choosing the threshold values for a given recognition strategy has been discussed.

Każdy system automatycznego rozpoznawania obrazów można podzielić na trzy części: źródło sygnału, układ pomiarowy (układ ekstrakcji parametrów) oraz klasyfikator. Przedmiotem zainteresowania niniejszej pracy jest klasyfikator z punktu widzenia procedury przydatnej do rozpoznawania głosów w zbiorach otwartych. Przedstawiono metodykę badań, analizę problemu i algorytm rozpoznawania oraz omówiono wyniki eksperymentów uwzględniających optymalizację wyboru, dyskryminujących obce głosy, oraz wybór wartości progowych dla określonej strategii rozpoznawania systemu.

1. Introduction

In the problem of automatic recognition of the acoustic (and not only acoustic) patterns the cases appear when it can not be assumed a priori that the currently recognized object or it's representing pattern belongs to a fixed set of object or pattern classes. In the process of speech recognition the number of recognized linguistic units is limited in general to a given set of classes (closed set), whereas because of practical reasons, it is not always possible to assume the analogical limitation in the task of recognizing the speaker's voice [1, 6, 7]. Then the problem arises of developing a recognition algorithm capable of dealing with the open set of patterns, i.e. an algorithm which wouldn't need the assumption that an input voice pattern of an unknown speaker must belong to the given set. The concept of such approach to the voice recognition problem has been presented in the paper [6]. Here the analysis of this problem and the complete recognition algorithm together with the experimental results is described and discussed. One of the main targets of the present work was to carry out the probability analysis of errors and risk concerned with making a decision for an algorithm of recognition in open sets as a function of the approximation method of extraneous voice patterns distribution and the discrimination threshold.

2. Methodological assumptions and recognition procedure

Analogically to the majority of automatic recognition systems in automatic voice recognition (AVR) as the description of utterances the patterns are used. A pattern is an ordered set of numerical parameters belonging to a specified parameter space (observation space) χ^K (K denotes the space dimension). These parameter sets form some specified distributions characterised by the probability densities $Q(\mathbf{x}|m)$ where $\mathbf{x}$ is a parameter vector (voice pattern) in a multi – dimensional space, and m is an index of speaker's voice or, generally, a class index.

In the recognition process the classical Bayesian decision criterion accounts for the probability with which the recognised pattern $\mathbf{y}$ represents the class m, i.e. $P(m|\mathbf{y})$ [4, 5].

This probability is connected with conditional probability densities by the Bayes relation

$$P(m|\mathbf{y}) = \frac{Q(\mathbf{y}|m)P_m}{\sum_{l=1}^{M} Q(\mathbf{y}|l)P_l}; \quad m = 1, 2, \ldots, M \tag{1}$$

where M – number of classes, P_m – probability of occurrence for a pattern belonging to the m–th class. The problem of recognition in the classical approach resolves itself into finding the minimum of risk $R_m(\mathbf{y})$ concerned with assigning the pattern $\mathbf{y}$ to the class m:

$$R_m(\mathbf{y}) = \sum_{l=1}^{M} C_{m,l} P(l|\mathbf{y}); \quad m = 1, 2, \ldots M, \tag{2}$$

where $C_{m,l}$ — loss matrix element which denotes the value of loss resulting from assigning the pattern from the class l to the class m [4].

Taking into account the fact that for a given vector pattern $\mathbf{y}$ the denominator in (1) is constant, the expression (2) can be rewritten in the form(1)

$$R_m(\mathbf{y}) = \sum_{l=1}^{M} C_{m,l} Q(\mathbf{y}|l) P_l; \quad m = 1, 2, \ldots M. \tag{3}$$

(1) by factoring the constant out and neglecting it.

2.1. *The Bayesian decision criterion for open sets*

In the case of recognition in open sets the set of voice classes of the speakers consists of a subset M of the known recognized classes (closed set) and one multiobject class corresponding to the complement of the subset of known speakers' voices, which is named the ground or the class of extraneous voices.

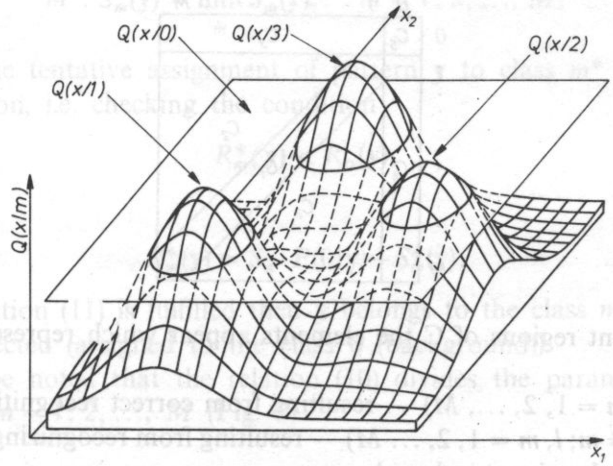


Fig. 1. Distributions of conditional probability densities in case of a two-dimensional space ($K = 2$)

For the subset of known pattern classes $m = 1, 2, \dots M$ it can be assumed that the conditional distributions are normal distributions with the conditional probability density $Q(\mathbf{x}|m)$ expressed by the relation (4) (cf. Fig. 1):

$$Q(\mathbf{x}|m) = (2\pi)^{\frac{-K}{2}} |\mathbf{B}_m|^{\frac{-1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{W}_m)^{Tr}\mathbf{B}_m^{-1}(\mathbf{x}-\mathbf{W}_m)\right\} \tag{4}$$

where

$$B_m = \frac{1}{I_m - 1}\sum_{i=1}^{I_m}(\mathbf{W}_m - \mathbf{x}_{m,i})(\mathbf{W}_m - \mathbf{x}_{m,i})^{Tr} \tag{5}$$

$\mathbf{B}_m$ – covariance matrix of the intra-class dissipations and

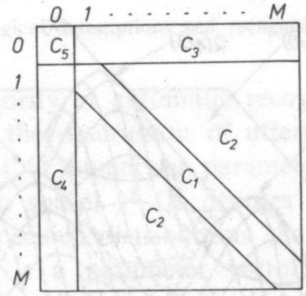
$$\mathbf{W}_m = \frac{1}{I_m}\sum_{i=1}^{I_m}\mathbf{x}_{m,i} \tag{6}$$

is the mean vector for a class, $m = 1, 2, \dots M$, M – number of classes (speakers' voice), $i = 1, 2, \dots I$, I – number of repetitions of an utterance for the m-th class in the training sequence TS, K – dimension of the parameter vector, T_r – denotation of the vector transposition. The conditional distribution of the background ($m = 0$) ([2])

$Q(\mathbf{x}|0)$ is in general case a multimodal distribution with a significant number of modes, tending to infinity. In general its measurement or assumption of an analytical form is impossible.

Before analyzing some possibilities of approximation of the distribution $Q(\mathbf{x}|0)$ the structure of the loss matrix should be examined.

In the loss matrix $\mathbf{C}$ the following regions can be distinguished:



In the subsequent regions of C the elements appear which represent the following values:

C_1 ($C_{m,m}$; $m = 1, 2, \ldots, M$) – resulting from correct recognition of pattern m,

C_2($C_{l,m}$; $l \neq m$; $l, m = 1, 2, \ldots M$) – resulting from recognizing a representative of class m as belonging to class l,

C_3($C_{0,m}$; $m = 1, 2, M$) – resulting from rejecting a representative of class m considering it as not belonging to the closed set,

C_4 $C_{l,0}$;($l = 1, 2, \ldots, M$) – resulting from assigning on extraneous patterns class representative to class l from the closed set,

C_5(3) ($C_{0,0}$) – resulting from the correct rejection (assigning to the background) of a voice representative from ouside the closed set.

Taking into account the above described structure of the matrix $\mathbf{C}$ the relation (2) can be transformed as follows:

$$R_0(\mathbf{y}) = C_{00}\, P(O|\mathbf{y}) + \sum_{i=1}^{M} C_{0l}\, P(l|\mathbf{y}); \quad m = 0, \tag{7}$$

$$R_m(\mathbf{y}) = C_{m,0} P(O|\mathbf{y}) + S_m(\mathbf{y}); \quad m = 1, 2, \ldots, M, \tag{8}$$

where

$$S_m(\mathbf{y}) = \sum_{i=1}^{M} C_{m,j}\, P(l|\mathbf{y}); \quad m = 1, 2, \ldots, M, \tag{9}$$

---

(2) Introduction of an additional class $m = 0$ necessitates for an appropriate modification of the summation range in Eqs. (1), (2), (3).

(3) The regions C_1 and C_5 represent the losses resulting from the correct decision, therefore usually the zero values are assumed in these regions, or if there are some special reasons the gains can appear here (with the opposite sign).

is the risk of a decision that **y** represents the class m from the closed set. Numerically this value is proportional to the value of risk (cf. Eq. (2)).

If in the region C_4 all the values are equal to a constant ($C_{m,0} = S_4$; $m = 1, 2, \ldots, M$; it will be assumed that this condition is valid from now on), then the recognition process in open sets can be divided into two stages

1) recognition in the closed set, i.e. finding

$$m^*: S_m^*(\mathbf{y}) = \min_{\hat{m}} S_m(\mathbf{y}); \quad m = 1, 2, \ldots, M, \tag{10}$$

what denotes the tentative assignment of pattern **y** to class m^* and

2) verification, i.e. checking the condition

$$R_m^*(\mathbf{y}) < R_0(\mathbf{y}), \tag{11}$$

where

$$R_m^*(\mathbf{y}) = S_4 \cdot P(0|\mathbf{y}) + S_m^*(\mathbf{y}). \tag{12}$$

If the condition (11) is fulfilled then **y** belongs to the class m^* otherwise the pattern **y** is rejected (assigned to the class 0 (background)).

It should be noted that the relation (10) divides the parameter space into M regions χ_m^K; $m = 1, 2, \ldots, M$ (Fig. 2)

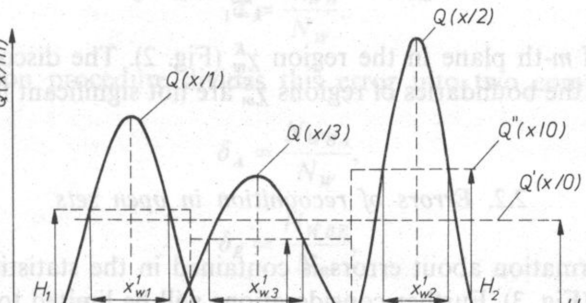


Fig. 2. Illustrations for the method of approximation of $Q(x|0)$ and determination of threshold values in a one-dimensional space $K = 1$, $Q'(x|0)$ – Case 1 (Eq. (39)), $Q''(x|0)$ – Case 2 (Eq. (40))

$$\mathbf{x} \in \chi_m^K \text{ if } S_m(\mathbf{x}) < S_l(\mathbf{x});\ l \neq m,\ l = 1, 2, \ldots M, \tag{13}$$

and the relation

$$S_m(\mathbf{x}) = \min_{l \neq m} S_l(\mathbf{x}) \tag{14}$$

is the equation of the boundary of the region χ_m^K with respect to the vector $\mathbf{x}$. Therefore, recognition in the closed subset is equivalent to finding

$$m^*: \mathbf{y} \in \chi_m^K. \tag{15}$$

Similarly, the verification process (Eq. (11)) in every region χ_m^K specifies the subregion $\chi_{w_m}^K$ (Fig. 2). From the above it results that the exact knowledge of the whole distribution $Q(\mathbf{x}|0)$ is not needed. It is only necessary to know the boundaries of the regions χ_m^K; $m = 1, 2, \ldots M$ or the distribution $Q(\mathbf{x}|0)$ in some neighbourhood of these boundaries. If the values in the regions of matrix $\mathbf{C}$ do not differ significantly, then most frequently the following relation takes place:

$$\mathbf{W}_m \in \chi_{w_m}^K;\ m = 1, 2, \ldots, M, \tag{16}$$

where $\mathbf{W}_m$ as in Eq. (6).

Hence, the region $\chi_{w_m}^K$ includes some neighbourhood of the vertex of distribution $Q(x|m)$; $m = 1, 2, \ldots M$ (Fig. 2). An approximation of $Q(\mathbf{x}|0)$ by M planes, one per each region χ_m^K, can be proposed then

$$Q(\mathbf{x}|0) = H_m(\mathbf{x}), \quad m{:}\mathbf{x} \in \chi_m^K, \tag{17}$$

where

$$H_m(\mathbf{x}) = \cdot h_{m,0} + \sum_{k=1}^{K} h_{m,k} \cdot x_k \tag{18}$$

is the equation of m-th plane in the region χ_m^K (Fig. 2). The discontinuities of such approximation at the boundaries of regions χ_m^K are not significant for the verification process.

2.2. *Errors of recognition in open sets*

The full information about errors is contained in the statistics of recognitions and verifications (Fig. 3). Further considerations will be limited to the global errors of recognition. From the belows quoted directory the types of recognitions and eliminations used in Fig. 3 can be read. The following notations have been introduced: N – number of voice patterns, W – known patterns (belonging to the closed set M), O – extraneous patterns from outside the set, P – patterns correctly initially recognized, B – patterns erroneously initially recognized, A – patterns accepted by the verifier, E – patterns eliminated by the verifier.

EXAMPLE: N_{WPE} denotes the number of patterns recognized by the classifier and then rejected in the verification process.

If in the previously distinguished regions in the loss matrix its elements are equal

$$P_m = \frac{1}{M} P_W;\ m = 1, 2, \ldots, M, \tag{19}$$

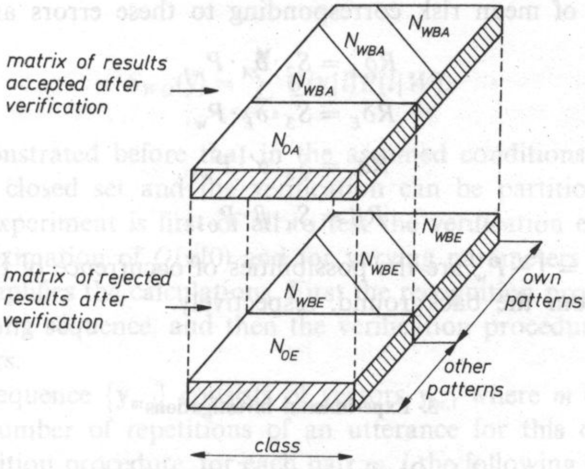


Fig. 3. Recognition and verification statistics in an open test set

where P_W — probability of occurrence of any representative of a closed set, then the following errors can be found (Eqs (20) to (26)). Inside the closed set the statistics of erroneous recognitions are represented by

$$\delta = \frac{N_{WB}}{N_W}. \tag{20}$$

The verification procedure divides this error into two components:

$$\delta_A = \frac{N_{WBA}}{N_W}, \tag{21}$$

$$\delta_E = \frac{N_{WBE}}{N_W}, \tag{22}$$

and introduces the verification errors: type I (error of false rejection)

$$\alpha' = \frac{N_{WPE}}{N_W}, \tag{23}$$

$$\alpha'' = \frac{N_{WPE} + N_{WB}}{N_W} = \alpha' + \delta, \tag{24}$$

$$\alpha''' = \frac{N_{WPE} + N_{WBE}}{N_W} = \alpha' + \delta_E, \tag{25}$$

and type II (error of false acceptation):

$$\beta = \frac{N_{OA}}{N_O}. \tag{26}$$

The components of mean risk corresponding to these errors are

$$R\delta_A = S_2 \cdot \delta_A \cdot P_w, \tag{27}$$

$$R\delta_E = S_3 \cdot \delta_E \cdot P_w, \tag{28}$$

$$R\alpha' = S_3 \cdot \alpha' \cdot P_w, \tag{29}$$

$$R\beta = S_4 \cdot \beta \cdot P_0, \tag{30}$$

where P_w and $P_0 = 1 - P_w$ are the possibilities of occurrence of the voice from the closed set and from the background, respectively.

3. Experimental investigations

3.1. Organisation of experiment

Values of the loss matrix **C** elements can be fixed or can be the parameters of an experiment. There are no reasons to differentiate them in the distinguished regions and it has been assumed that in the regions C_1 and C_5 the values are zero, whereas in the regions C_2, C_3 and C_4 the values are 1, S_3 and S_4, respectively, where S_3 and S_4 are the parameters of the experiment. Since recognition and verification consists in the choice of minimum risk, the addition of an arbitrary constant to the matrix **C** does not change the results of recognition. Further, the scaled matrix **C** will be applied, with the values in regions C_1, C_2, C_3, C_4 and C_5 equal to $-1, 0, S_3 - 1, S_4 - 1, -1$, respectively. Under these assumptions the relations (7), (8) and (9) can be transformed as follows

$$R_0(\mathbf{y}) = -Q(\mathbf{y}|0)P_0 + (S_3 - 1)S_0(\mathbf{y});\ m = 0, \tag{31}$$

$$R_m(\mathbf{y}) = (S_4 - 1)Q(\mathbf{y}|0)P_0 + S_m(\mathbf{y});\ m = 1, 2, \ldots, M, \tag{32}$$

where

$$S_m(\mathbf{y}) = -Q(\mathbf{y}|m)P_m;\ m = 1, 2, \ldots, M, \tag{33}$$

$$S_0(\mathbf{y}) = \sum_{l=1}^{M} Q(\mathbf{y}|l)P_l;\ m = 0, \tag{34}$$

and the verifying relation can be rewritten in the form

$$S_m^*(\mathbf{y}) < -S_4 Q(\mathbf{y}|0)P_0 - (1 - S_3) \cdot S_0(\mathbf{y}). \tag{35}$$

Having introduced $P_l = P(l|W)P_W$, where $P(l|W)$ denotes the conditional probability inside the closed set, the relation (35) can be transformed as follows:

$$S_m^*(\mathbf{y}) < -S_4 Q(\mathbf{y}|0)P_0 - (1 - S_3)S_{WO}(\mathbf{y}), \tag{36}$$

where

$$S_{WO}(\mathbf{y}) = \sum_{l=1}^{M} Q(\mathbf{y}|l)P(l|W). \tag{37}$$

It has been demonstrated before that in the assumed conditions the recognitions procedure in the closed set and the verification can be partitioned.

The aim of experiment is first of all to test the verification errors for various methods of approximation of $Q(\mathbf{x}|0)$ and for varying parameters S_3 and S_4. Such decomposition simplifies the calculations. First the recognition procedure is realised for the whole testing sequence, and then the verification procedure is repeated for various parameters.

The testing sequence $\{\mathbf{y}_{m,i}\}$ consists of vectors $\mathbf{y}_{m,i}$ where m denotes the class index and i – number of repetitions of an utterance for this class.

In the recognition procedure, for each pair m, i the following quantities can be found

$$m^*_{m,i},\ S^*_m(\mathbf{y}_{m,i}),\ S_0(\mathbf{y}_{m,i}), \tag{38}$$

i.e. the three values which allow for to repeat quickly the verification procedure many times according to Eq. (36).

The choice of possible types of planes Eq. (18) has been limited to two simple cases (cf Fig. 2)

$$Q(\mathbf{x}|0) = \mathrm{H}, \quad \mathrm{H}\ \text{– constant}, \tag{39}$$

$$Q(\mathbf{x}|0) = \mathrm{H}_m;\ m: \mathbf{x} \in \chi^K_m. \tag{40}$$

For the first case the threshold value H can be described as

$$\mathrm{H} = \gamma \cdot Q_{\text{śr}}, \tag{41}$$

where

$$Q_{\text{śr}} = \frac{1}{M} \sum_{m=1}^{M} Q(W_m|m), \tag{42}$$

and γ is a coefficient (a parameter of the experiment).

For the second case Eq. (40) two versions of the method of introducing the thresholds H_m for each class can be distinguished

a) $$\mathrm{H}_m = \gamma \cdot Q(W_m|m);\ m = 1, 2, \ldots M, \tag{43}$$

with the same coefficient for all the classes and

b) $$\mathrm{H}_m = \gamma_m \cdot Q(W_m|m);\ m = 1, 2, \ldots M. \tag{44}$$

with the coefficient γ_m chosen individually for each class on the basis of the testing sequence, so that the verification risk component in the given class would be minimum, whilst the values N_{WPE} and N_{OA} are taken into account for each class.

3.2. *The experiment*

The experimental investigation aimed at the verification of the proposed algorithm and the method of analysis for open sets has been realized with the following agreements taken into account:

a) As the parameters in the observation space, the values extracted from a fixed signal (a key phrase) will be used, the efficiency of which for voice recognition has been well established. It has been assumed that the parameter vector will be formed from the distribution of time intervals between the speech signal zero-crossings in the sentence „Jutro będzie ładny dzień" (pol: "It will be a fine day tomorrow") [1, 2, 3]. Value of the dimension of the space K had been initially assumed equal to 7; it has been reduced then to $K = 4$, because parameters with the greatest discrimination force have been chosen.

b) The mean object of interest will be the verification procedure, for the fixed parameters of the signals and of the parameter extraction system.

c) The training sequence TS is the set of vectors $\{\mathbf{x}_{m,i}\}$ $m = 1, 2, \ldots, M$ (M – number of voices in the closed set) $i = 1, 2, \ldots, I_m$ (it has been assumed $I_m = I$ — number of repetitions for each class).

d) It has been assumed that the training sequence will consist of the utterance patterns of $M = 10$ speakers per $I = 10$ repetitions. The test sequence will contain the patterns of 10 speakers × 10 repetitions from the closed set and of 10 speakers × 10 repetitions as extraneous utterances. Together, the open test set consisted then of 200 utterances comming from 20 speakers.

The analysis and recognition has been carried out with the use of programs written on ZX Spectrum microcomputer.

With these programs the following calculations have been realized:

a) The mean vector $\mathbf{W}_m$ for 10 classes from the training set were calculated (Table 1).

b) The estimators of conditional probability densities $Q(\mathbf{y}|m)$ and risk $R_m(\mathbf{y})$ were found. In the Table 2 the conditional densities have been confronted.

c) The values of errors α and β and the value of verification risk as a function of γ were calculated and plotted for the cases given by Eqs (39) and (40). The diagrams of these functions are presented in the Fig. 4a, b, c.

d) The matrices of recognitions and verifications were calculated for various

Table 1. Vectors of class standards

Parameter k	Pattern n: 1	2	3	4	5	6	7	8	9	10	Pattern medium
1	−1.094	−0.037	−0.583	−0.655	0.629	−0.649	−0.196	−0.003	1.017	−0.825	−0.240
2	−0.840	−0.224	−0.478	−0.575	−0.158	−0.690	−0.370	−0.288	0.151	−0.6064	−0.408
3	−1.379	−0.518	−0.652	−0.952	−0.978	−1.165	−0.764	−0.772	−0.100	−0.726	−0.801
4	0.298	−0.493	1.066	−0.749	1.359	−1.238	−0.939	−0.539	−0.588	−0.158	−0.198

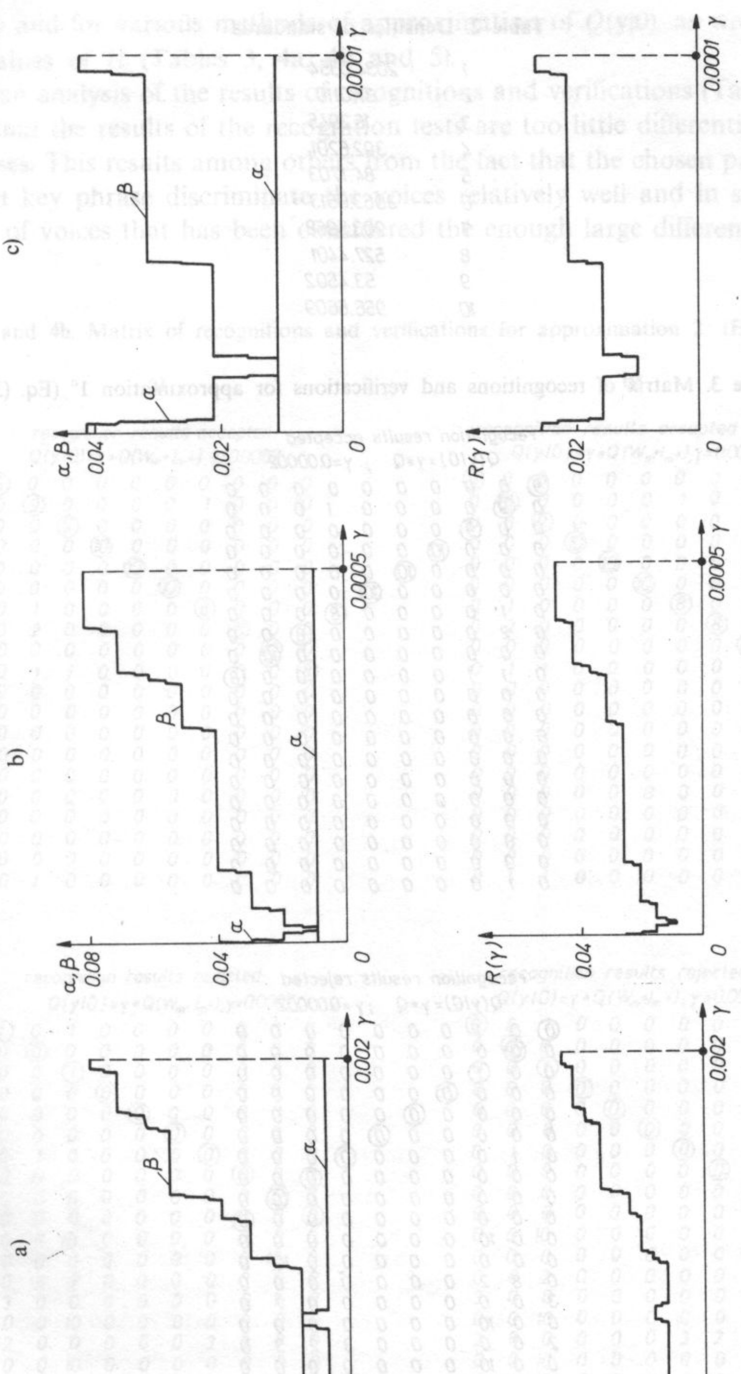


Fig. 4. Graphs of errors α and β and risk function R depending on coefficient γ

Table 2. Densities of standards

1	2034.0354
2	37.0110
3	16.2945
4	392.6201
5	84.1703
6	3063.0513
7	203.5968
8	527.4401
9	53.4502
10	956.6609

Table 3. Matrix of recognitions and verifications for approximation 1° (Eq. (39))

recognition results accepted

$Q(y/0) = \gamma * Q \;;\; \gamma = 0.00002$

9	0	0	0	0	0	0	0	0	0
0	9	0	0	0	0	1	0	0	0
0	0	9	0	0	0	0	0	0	0
0	0	0	10	0	0	0	0	0	0
0	0	0	0	10	0	0	0	0	0
0	0	0	0	0	10	0	0	0	0
0	1	0	0	0	0	8	0	0	0
0	2	0	0	0	0	0	8	0	0
0	0	0	0	0	0	0	0	10	0
0	1	1	0	0	0	0	0	0	8
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0

recognition results rejected

$Q(y/0) = \gamma * Q \;;\; \gamma = 0.00002$

0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	10	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0
0	8	2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	7	0
0	0	10	0	0	0	0	0	0	0
2	0	0	0	0	0	3	2	3	0
0	0	10	0	0	0	0	0	0	0
0	0	3	0	0	0	2	0	5	0
8	0	2	0	0	0	0	0	0	0
0	0	9	0	0	0	0	0	0	0

values of γ and for various methods of approximation of $Q(y|0)$, as well as for the optimal values of H (Tables 3, 4a, 4b and 5).

After an analysis of the results of recognitions and verifications (Table 6) it can be stated that the results of the recognition tests are too little differentiated for the studied cases. This results among others from the fact that the chosen parameters of the present key phrase discriminate the voices relatively well and in such a small testing set of voices that has been considered the enough large differences couldn't

Table 4a and **4b**. Matrix of recognitions and verifications for approximation 2° (Eq. (40), (43))

a)

recognition results accepted

$Q(y|0)=\gamma * Q(W_m+l_m+); \gamma=0.0007$

(6)	0	0	0	0	0	0	0	0	0
0	(9)	0	0	0	0	1	0	0	0
0	0	(9)	0	0	0	0	0	0	0
0	0	0	(10)	0	0	0	0	0	0
0	0	0	0	(10)	0	0	0	0	0
0	0	0	0	0	(10)	0	0	0	0
0	1	0	0	0	0	(8)	0	0	0
0	2	0	0	0	0	0	(8)	0	0
0	0	0	0	0	0	0	0	(10)	0
0	1	1	0	0	0	0	0	0	(8)
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0

recognition results rejected

$Q(y|0)=\gamma * Q(W_m+l_m+); \gamma=0.0007$

(1)	0	1	0	0	0	0	0	0	0
0	(0)	0	0	0	0	0	0	0	0
0	0	(1)	0	0	0	0	0	0	0
0	0	0	(0)	0	0	0	0	0	0
0	0	0	0	(0)	0	0	0	0	0
0	0	0	0	0	(0)	0	0	0	0
0	1	0	0	0	0	(0)	0	0	0
0	0	0	0	0	0	0	(0)	0	0
0	0	0	0	0	0	0	0	(0)	0
0	0	0	0	0	0	0	0	0	(0)
0	0	10	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0
0	8	2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	7	0
0	0	10	0	0	0	0	0	0	0
2	0	0	0	0	0	3	2	3	0
0	0	10	0	0	0	0	0	0	0
0	0	3	0	0	0	2	0	5	0
8	0	2	0	0	0	0	0	0	0
0	0	9	0	0	0	0	0	0	0

b)

recognition results accepted

$Q(y|0)=\gamma * Q(W_m+l_m+); \gamma=0.0003$

(9)	0	0	0	0	0	0	0	0	0
0	(9)	0	0	0	0	1	0	0	0
0	0	(9)	0	0	0	0	0	0	0
0	0	0	(10)	0	0	0	0	0	0
0	0	0	0	(10)	0	0	0	0	0
0	0	0	0	0	(10)	0	0	0	0
0	1	0	0	0	0	(8)	0	0	0
0	2	0	0	0	0	0	(8)	0	0
0	0	0	0	0	0	0	0	(10)	0
0	1	1	0	0	0	0	0	0	(8)
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0

recognition results rejected

$Q(y|0)=\gamma * Q(W_m+l_m+); \gamma=0.0003$

(0)	0	1	0	0	0	0	0	0	0
0	(0)	0	0	0	0	0	0	0	0
0	0	(1)	0	0	0	0	0	0	0
0	0	0	(0)	0	0	0	0	0	0
0	0	0	0	(0)	0	0	0	0	0
0	0	0	0	0	(0)	0	0	0	0
0	1	0	0	0	0	(0)	0	0	0
0	0	0	0	0	0	0	(0)	0	0
0	0	0	0	0	0	0	0	(0)	0
0	0	0	0	0	0	0	0	0	(0)
0	0	10	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0
0	8	2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	7	0
0	0	10	0	0	0	0	0	0	0
2	0	0	0	0	0	3	2	3	0
0	0	10	0	0	0	0	0	0	0
0	0	3	0	0	0	2	0	5	0
8	0	2	0	0	0	0	0	0	0
0	0	8	0	0	0	0	0	0	0

Table 5. Matrix of recognitions and verifications for approximation 2° including optimized coefficient γ_m

recognition results accepted

$Q(y|0) = H_{m*opt}$

9	0	0	0	0	0	0	0	0	0
0	9	0	0	0	0	1	0	0	0
0	0	9	0	0	0	0	0	0	0
0	0	0	10	0	0	0	0	0	0
0	0	0	0	10	0	0	0	0	0
0	0	0	0	0	10	0	0	0	0
0	1	0	0	0	0	8	0	0	0
0	2	0	0	0	0	0	8	0	0
0	0	0	0	0	0	0	0	10	0
0	1	1	0	0	0	0	0	0	8
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0

recognition results rejected

$Q(y|0) = H_{m*opt}$

0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	10	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0
0	8	2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	7	0
0	0	10	0	0	0	0	0	0	0
2	0	0	0	0	0	3	2	3	0
0	0	10	0	0	0	0	0	0	0
0	0	3	0	0	0	2	0	5	0
8	0	2	0	0	0	0	0	0	0
0	0	9	0	0	0	0	0	0	0

appear. What's more, it occurred not profitable to choose the individual optimal threshold (Eq. (44)). It should be clearly underlined here that the methodological correctness of the method of choice of the discriminating quantities has been confirmed, what was the main target of the study; whereas the small number of the background patterns did not allow for significant differentiation of the cases of approximation of $Q(\mathbf{x}|0)$ (cf Eqs. (39), (40), (43) and (44)).

Table 6 Confrontation of experimental results (for $S_3 = S_4 = 1$)

N	Casc 1° (Eq. (39))	Casc 2° (Eq. (43))		Case 2° (γ_m individual) (Eq. (44))
	$\gamma = 0.00002$	$\gamma = 0.0007$	$\gamma = 0.0003$	
N_{WPA}	91	90	91	91
N_{WBA}	6	6	6	6
N_{OA}	1	1	2	1
N_{WPE}	1	2	1	1
N_{WBE}	2	2	2	2
N_{OE}	99	99	98	99
Errors	%	%	%	%
δ	8	8	8	8
δ_A	6	6	6	6
δ_E	2	2	2	2
α'	1	2	1	1
α''	9	10	9	9
α'''	3	4	3	3
β	1	1	2	1

4. Conclusion

The above presented methodological considerations justify the statement that the proposed algorithm of voice recognition in open sets is a flexible procedure that allows for fitting the global recognition characteristics, which are the errors α and β, to a certain operating strategy of the system.

For the chosen parameters patterns describing voices there always exists a possibility of carrying out an optimization of the recognition process by an appropriate choice of background approximation, i.e. by selecting the optimal threshold values H_m.

This is the basic advantage of the method. It has been experimentally verified for the population of 20 voice classes including (10 extraneous ones).

If the tests had been carried out for a larger population of voices especially extraneous, then finding a more distinct optimal threshold H_m could have been expected with regard to smaller granularity. In Fig. 4a the minimum of risk is not univocally determined because of large granularity of data e.g., two flat minima appear what has been already explained above by a too small size of the testing sequence.

Irrespective of the above, in every experiment on automatic voice recognition the following factors will always influence the numerical values: a) choice of the key phrase, b) observation space parameters, c) method of forming the voice patterns.

In a fixed, chosen pattern preparation system these factors can be treated as the values which have a parametric influence. The further studies on automatic voice recognition in open sets wall concentrate on the larger number of classes M, mainly on the background voices, and on other parameters describing voices of speakers.

References

[1] Cz. Basztura, *Automatic speaker recognition by zero-crossing analysis of the speech signal*, In: Speech Analysis and Synthesis (W. Jassem[ed.]), 5 PWN, p. 9–40, Warszawa 1980.

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

[3] Cz. Basztura, J. Jurkiewicz, *The zero-crossing analysis of a speech signal in the short-term method of automatic speaker identification*, Arch. of Acoustics **3**, 3, 185–196 (1978).

[4] J. Z. Cypkin, *Basics of the learning systems* (in Polish) WNT, Warszawa 1973.

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

[6] W. Majewski, Cz. Basztura, *Speaker recognition in open sets*, In: Proceedings of the Tenth International Congress of Phonetic Sciences M.P.R. Van den Broecke and A. Cohen [eds] Foris Publications, p. 322–325, Dordrecht 1984.

[7] A. E. Rosenberg, *Automatic speaker verification: a review*, Proc. IEEE, **64**, 4, 475–487 (1976).

Received on May 15, 1986; revised version on September 22, 1986.

ARCHIVES OF ACOUSTICS
13, 3-4, 221-234 (1988)

STATISTIC MODEL OF SPEECH-LIKE SIGNAL GENERATION(1)

PAWEŁ BIEŃKOWSKI, WOJCIECH MYŚLECKI

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

This paper presents a parametric model of a generator of a signal with probabilistic characteristics of a speech signal. Optimal parameters of the model have been selected for a case of a Polish speech signal and research results of the model in the form of a computer programme are given.

W pracy przedstawiono parametryczny model generatora sygnału o cechach probabilistycznych zbliżonych do wyróżnionych cech probabilistycznych sygnału mowy. Dobrano optymalne parametry modelu dla przypadku sygnału mowy polskiej oraz przedstawiono wyniki badań modelu zrealizowanego w postaci programu na EMC.

1. Introduction

The authors of this paper attempt to determine generation rules of signal with chosen statistic distributions of parameters close to distributions occuring in a speech signal. Such synthetic signals find application in studies concerned with mechanisms of speech signal perception and with speech signal transmission in acoustic and telecommunication channels. A model of generation of a signal with statistic characteristics significant from the point of view of speech transmission quality in telecommunication channels is presented on the basis of own research and research results given in papers [3, 9, 10].

This model includes only long-term probabilistic characteristics of the speech signal. Hence, the speech signal is a realization of a stationary stochastic process, which is an ergodic process, as well. According to the ergodicity definition probabilistic characteristics can be obtained from a singular realization in time $t \to \infty$. In practice we have a time-limited process realization in speech signal

(1) Research performed within problem CPBP O2.03., 9.3.

analysis. Therefore, the notion of weak stationarity was used in this paper. The minimal time after which probabilistic characteristics of the signal do not change significantly was called the minimal time interval of signal stationarity or minimal time interval of signal stationarity with respect to a definite characteristic. Stated above assumptions concern research referred to in the paper as well as studies performed by the authors.

2. Theoretical foundations of the model

The shape of the probability density function of instantaneous amplitudes is one of the basical probabilistic characteristics of a speech signal. It distinguishes clearly the speech signal from other stochastic processes. The probability density function $p(x)$ of a random signal $x(t)$ determines the probability of the event that the signal's values in an arbitrary moment are contained in a definite interval. Probability density functions of instantaneous amplitudes of a speech signal, measured for various languages, show a great consistence of shape. Fig. 1 presents density distributions for 3 languages: English (according to DAVENPORT [2]), Russian (according to SZITOW and BIELKIN [5]), Polish (according to BRACHMAŃSKI and MAJEWSKI [1]).

Two parts can be distinguished in the curve which represents density distribution of instantaneous amplitudes of a speech signal. The first part characterizes the distribution for great amplitude values and can be approximated with the Laplace distribution; while the second one, representing low amplitudes, has the shape of a sharp vertex and is described by a normal, Laplace or delta distribution. The physical interpretation of the shape of the amplitude distribution curve is given by DAVENPORT [2]. The shape vertex of the distribution curve can result from the number of pauses in speech of which the character of the signal is determined by the reverberation and noises in the room, and channel's noises. A different interpretation was presented by RIMSKIJ-KORSAKOW [9]. He assumed that the achieved distribution differs greatly from the normal distribution, because speech, as well as music, is a complex random process. Apart from fluctuations of signal's phrases and amplitudes, related with the incoherence of separate sources of vibrations and transitions between individual sounds (what should give a normal distribution), also relatively slow variance changes occur. In the case of speech they result from a prosodic modulation of voice strength and modulation, stress and tempo, influence of the expiration process etc. Rimskij-Korsakow's hypothesis supplemented with WOLF's consideration [10] can be presented in analytic form as follows.

Let us assume that the amplitude distribution is normal but with dispersion values s varying in chosen time segments of the speech signal:

$$p_i(x; s_i) = \frac{1}{\sqrt{2\pi}\, s_i} \exp\left(-\frac{x^2}{2s_i^2}\right). \tag{1}$$

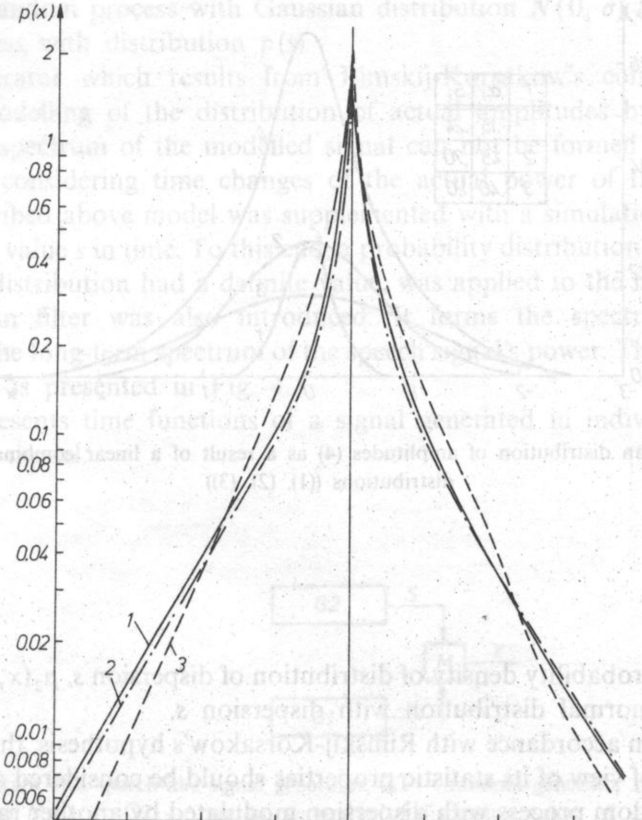


Fig. 1. Probability density distribution of actual values of a speed signal. According to: 1 — DAVENPORT, 2 — SZITOW and BIELKIN, 3 — BRACHMAŃSKI and MAJEWSKI

A weighted sum of these distributions

$$p(x) = \sum_{i=1}^{n} \alpha_i p_i(x; s_i); \sum \alpha_i = 1 \tag{2}$$

gives a distribution different from Gaussian (Fig. 2).

In the case of constant change of dispersion s the total density distribution is described by formula:

$$p(x) = \int_0^\infty p_1(s) p_2(x, s) ds, \tag{3}$$

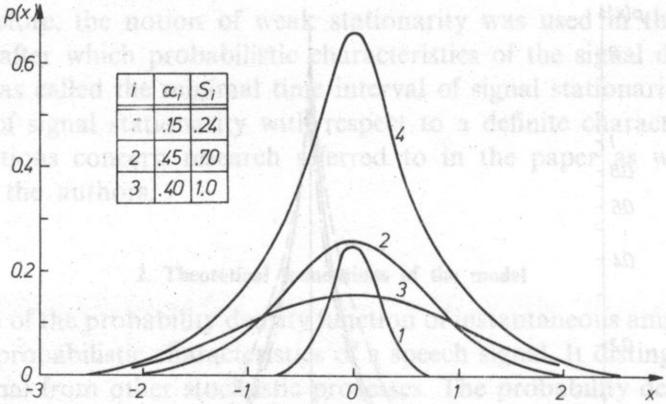


Fig. 2. Non-Gaussian distribution of amplitudes (4) as a result of a linear combination of Gaussian distributions ((1), (2), (3))

where $p_1(s)$ – probability density of distribution of dispersion s, $p_2(x, s)$ – probability density of normal distribution with dispersion s.

Therefore, in accordance with Rimskij-Korsakow's hypothesis, the speech signal from the point of view of its statistic properties should be considered as a realization of a normal random process with dispersion modulated by another random process. Inblin [3] has provided proof for such a hypothesis. He performed theoretical analysis and experimental verification of two excluding one another hypotheses. In the first hypothesis he assumed the speech signal to be a simple exponential random process with time-dependent dispersion: while in the second hypothesis speech was to be a normal random process with varying in time dispersion. The author proved the second hypothesis to be true on the basis of an analysis of the excess coefficient of density distributions in following time intervals of the signal. He also found that probabilities of instantaneous values of the speech signal are described by a normal distribution with constant dispersion in time intervals of about 15–30 ms. These results confirm the Rimskij-Korsakow hypothesis. Inblin's analysis has probabilistic sense and thus it can not be accepted that the distribution of amplitudes is a normal distribution in every time interval of 15–30 ms.

A model of signal generation according to rules can be proposed on the basis of Rimskij-Korsakow's conception. Such a model is shown in Fig. 3.

The following relationship describes it:

$$Y = X \cdot S, \tag{4}$$

where X – random process with Gaussian distribution $N(0, \sigma)$ S – modulating random process with distribution $p(s)$.

The generator which results from Rimskij-Korsakow's conception enables parametric modelling of the distribution of actual amplitudes by changing $p(s)$. Whereas, the spectrum of the modelled signal can not be formed and there is no possibility of considering time changes of the actual power of the signal.

The described above model was supplemented with a simulation of changes of the dispersion value s in time. To this end, a probability distribution $p(T)$ of the time, in which the distribution had a definite value, was applied to the modulated signal S. A Gaussian filter was also introduced. It forms the spectral characteristic according to the long-term spectrum of the speech signal's power. The block diagram of this model is presented in Fig. 4.

Fig. 5 presents time functions of a signal generated in individual functional blocks of the model.

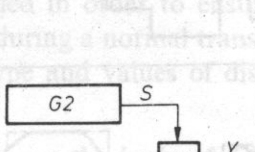


Fig. 3. Block diagram of a speech-like signal generator: G1 – random generator of noise with normal distribution N (0, σ); G2 – random generator of dispersion value, M – modulator

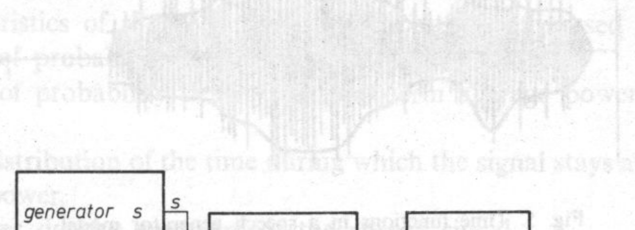


Fig. 4. Block diagram of a signal generator

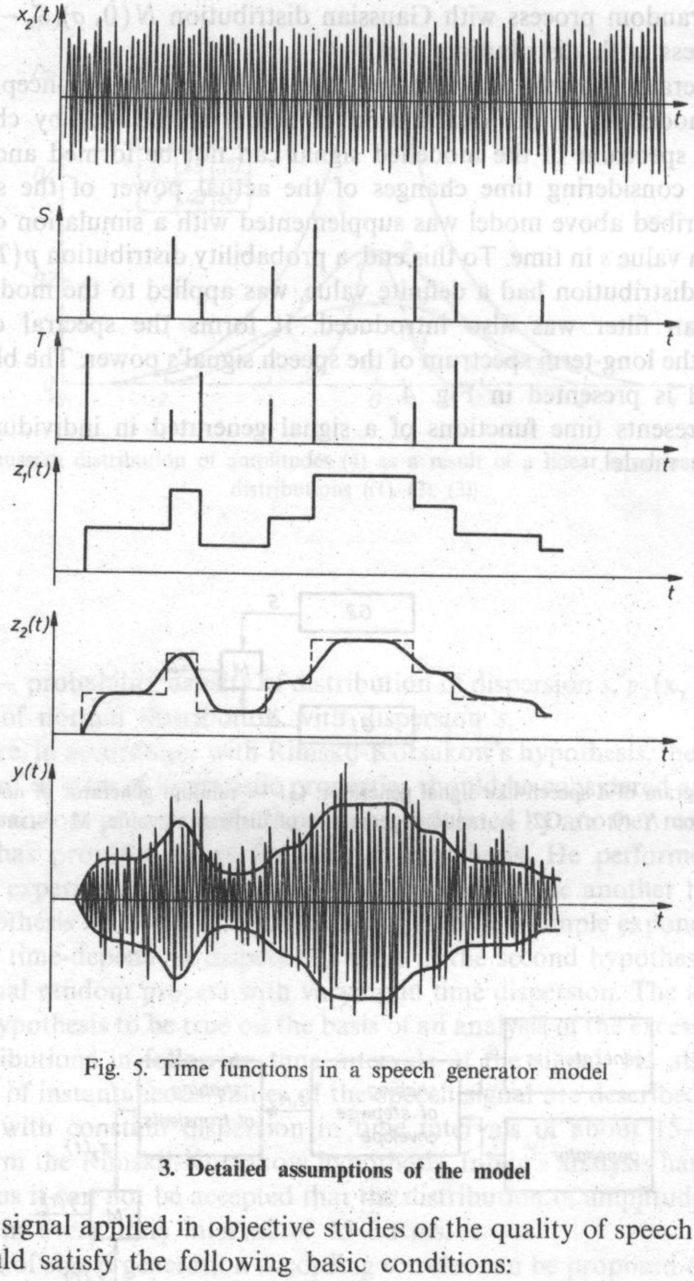


Fig. 5. Time functions in a speech generator model

3. Detailed assumptions of the model

The test signal applied in objective studies of the quality of speech transmission systems should satisfy the following basic conditions:

C1 – Condition of multidimensional space of signal observation. Subjective assessment of speech quality, which is always the final reference measure of an objective quality estimator, is a function defined in a multidimensional space of auditory perception where a numerous set of physical characteristics of the speech signal is analysed [7]. Hence, the test signal has to undergo multidimensional analysis in order to determine an objective estimator of transmission quality. The following relationship describes the process of determining such an estimator

$$w \leftrightarrow \tilde{w} = D\{d^n[y(t),\ y^*(t)]\}, \tag{6}$$

where: w – reference measure of transmission quality subjective measurements; $\tilde{w}$ – objective estimator of quality measure, d^n – distance in a n-dimensional space of signal observation between the transmitted, $y(t)$, and received, $y^*(t)$, test signal, D operator of transformation of d^n into w.

C2 – Condition of the ability to respond to disturbances and distortions of the channel. Possibly to the same extent as a signal of natural speech.

C3 – Condition of convergence of statistic spectral and time characteristics with natural speech characteristics.

This condition has to be satisfied in order to ensure work conditions of the channel close to conditions existing during a normal transmission of a speech signal, because it strongly influences the type and values of disturbances and distortions occuring in the channel.

C4 – Condition of signal generation on the basis of determined rules. When this condition is fulfilled then a standard test signal is obtained and the effect of certain types of disturbances and distortions in the channel on parameters of the output signal can be analysed mathematically.

The signal described in paragraph 2 satisfies condition C1 (i.e. it can be presented in a multidimensional space of observation), because it has several statistical characteristics of a speech signal which can be expressed by:

a) distribution of probability density of instantaneous values,
b) distribution of probability density of short-term average power of the speech signal,
c) probability distribution of the time during which the signal stays at the same level of average power,
d) average power density spectrum of the speech signal.

However, a model which generates such a signal does not include the density of zero – crossings of a speech signal, two- and more dimensional density distributions of amplitudes, the autocorrelation function and a series of other structural characteristics of the speech signal. The model was limited to mentioned above characteristics in order to achieve a test signal with a minimal time interval of stationarity, shorter than that of a natural speech signal; simplify the technical realization of the model and because existing methods of analysis do not make it possible to measure all physical characteristics of a speech signal.

In accordance with condition **C4** the presented signal is generated on the basis of determined rules. These rules can be accomodated to any language, while detailed parameters of the model have to be selected individually.

4. Measurements

Parameters which would ensure convergence of the characteristics of generated signal with the characteristics of the Polish speech signal are necessary to build the model. To this end we have to determine: distributions $p(s)$ and $p(T)$, frequency characteristic of the filter and to elaborate the method of determining function $z_2(t)$ on the basis of function $z_1(t)$ (see Fig. 4).

It was accepted that the distribution of probability density of the generated signal's amplitudes should be consistent with the density distribution of instantaneous amplitudes of the Polish speech signal given by BRACHMAŃSKI and MAJEWSKI [1]; while the average density spectrum of the power of the Polish speech signal should be approximated on the basis of results presented in papers [4, 6, 11]. No data on the distribution of probability density of short-term power $p(s)$ and the probability distribution $p(T)$ of the time the signal stays on the same level of average power can be found in literature. Therefore, suitable experimental studies were performed.

4.1. Research methods

A 5-minute newspaper text read in turn by 20 speakers with a constant average level of sound intensity was the phonetic material. The recording was done in the recording room of the ITA at the Technical University of Wrocław. A UM57 Neumann condenser microphone and M601 SD ZRK tape recorder were used. Irregularities of the microphone's response characteristic did not exceed $\pm$ 2 dB in the band 30 ÷ 15000 Hz; whereas irregularities of the frequency characteristic of the tape recorder's record-play channel did not exceed $\pm$ 2 dB in the band 80 ÷ 10000 Hz. The microphone was placed near the speaker's mouth (in the near field).

The system shown in Fig. 6 was used for measurements of distributions of probability density $p(s)$ and $p(T)$.

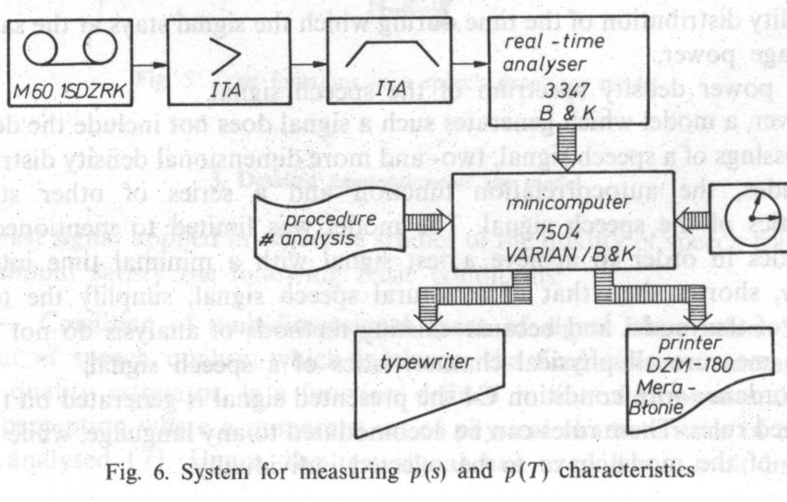


Fig. 6. System for measuring $p(s)$ and $p(T)$ characteristics

The analysed speech signal recorded on a tape recorder was subjected to band-pass filtration after being amplified what reduced the signal's spectrum to a band 200 Hz – 5000 Hz (the band deciding about the information content of the speech signal). The rms value of the signal was detected in the entire band in a 3347 Bruel-Kjaer synchronous analyser. So obtained rms values were sampled at 20 ms intervals and classified in 26 ranges; 2 dB each. At the same time events that following samples of the rms value were contained within the same 2 dB range were counted and classified.

The selection of the time-constant of the rms value detector should be discussed separately. In accordance with the Rimskij-Korsakow conception, on which the proposed model is based, the distribution of the probability density of the rms value of following speech segments had to be determined. INBLIN [3] has determined lengths of these segments as equal to 15–30 ms. The length of the segment was accepted as equal to 20 ms, so the applied time-constant of the rms value detector has to be equal to $\tau = 10$ ms.

The total measuring error of the rms value of a random signal depends on errors of the instrument and statistical inaccuracy due to a finite time of averaging. The total error of the 3347 analyser does not exceed $\pm$ 0.6 dB. The estimation of the statistical error is expressed by the relationship

$$\varepsilon = \frac{1}{2\sqrt{BT}}, \tag{6}$$

where $B-$ width of analysing filter, $T-$ effective time of averaging.
A 68% confidence level was determined from relationship (6) when the condition $BT > 15$ satisfied. For used in measurements values: $B = 5$ kHz and $T = 20$ ms, the reading error is negligible with respect to the observation resolution of the p (s) distribution accepted at 2 dB.

4.2. Results

Measurements of the p (s) distribution were carried out for speech signal intervals of 5, 3, 1 min. Statistic analysis has proved that a 3 minute interval of a speech signal is sufficient to determine the $p(s)$ distribution. $p(s)$ distributions for following voices were normalized in relation to the average value and then were averaged. Fig. 7 presents the distribution of probability density of short-term average power p (s), averaged for 20 men's voices.

The direct application of the achieved p (s) distribution in the model is possible, but that would require a very complicated random-number generator. Therefore, it had to be approximated by a different distribution with known analytical form which would be more simple to generate. The application of the analytical p (s) distribution also makes it possible to perform a mathematical analysis of the effect

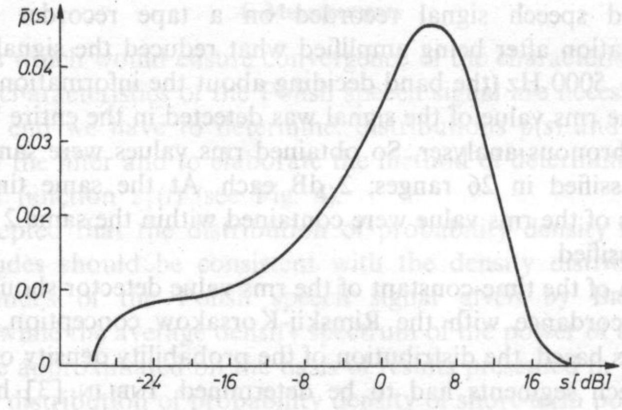


Fig. 7. Distribution of the average probability density of short-term average power of a Polish speech signal

this distribution has on the distribution of the probability density of amplitudes of the output signal. Two types of approximating distributions were considered: Rayleigh's distribution and gamma distribution. The probability density of the output signal's amplitudes for the Rayleigh's distribution in the form

$$p(s) = \frac{s}{\sigma_0^2} \exp\left(-\frac{s^2}{2\sigma_0} \right) \tag{7}$$

achieved from relationship (4) – is expressed by Laplace's distribution

$$p(x) = \frac{1}{2\sigma_0} \exp\left(-\frac{|x|}{\sigma_0} \right). \tag{8}$$

This distribution is frequently used to approximate amplitude densities of a speech signal. The parameter σ_0 in relationship (8) was determined from the approximation of the curve of amplitude density of Polish speech [1] by the Laplace's distribution. The correlation coefficient was used as the similarity measure of these curves. Because the segment corresponding to greater values of amplitudes is the most significant part of the curve of speech amplitude's density, the correlation coefficient was maximized in the interval $0.3 < |x| < 4$. The point 0.3 was determined as shown in Fig. 8. The maximal value of the correlation coefficient ($R_{max} = 0.986$) was achieved for $\sigma_0 = 0.74$; thereby determining the value of the σ_0 parameter for Rayleigh's distribution.

The possibility of applying the gamma distribution in the following form

$$p(x) = \frac{a^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-ax}, \quad a > 0, x > 0, \alpha > 0, \tag{9}$$

where $\Gamma(\alpha)$ – Euler's gamma function, α, a – parameters of the distribution, was considered.

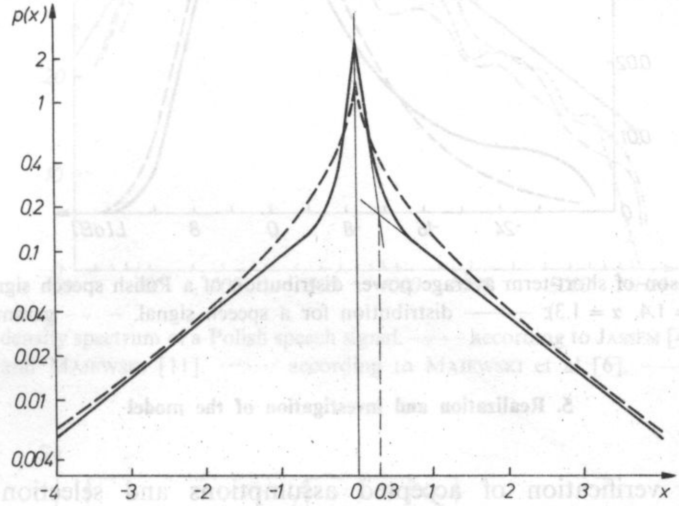


Fig. 8. A comparison of amplitude density distribution of Polish speech with a distribution resulting from the application of the gamma distribution to the signal model; ——— distribution for a speech signal, – – – theoretical distribution

The analytical form of the probability distribution of the output signal was not found, because an integral expression was impossible to solve. It resulted from numerical integration that the maximum value of the correlation coefficient at parameters of gamma distribution equal to $a = 1.4$ and $\alpha = 1.3$ was found between the curve of density distribution for Polish speech and the curve of probability density of the product of the gamma and normal distribution $N(0.1)$. The approximation with gamma distribution was proved to be better, because the correlation coefficient R_{max} was equal to 0.996 (for $0.3 < |x| < 4$). This approximation is shown in Figs. 8 and 9.

From the above considerations it was accepted that the generation of the short-term average power distribution will be done on the basis of the gamma distribution.

The distribution $p(T)$ of the time a signal stayed on the same level of average power was measured at the same time. Achieved distributions were averaged on a set of 10 speakers and approximated by an unilateral normal distribution. The best approximation was obtained at the value of 22.5 of the σ_0 parameter of this distribution.

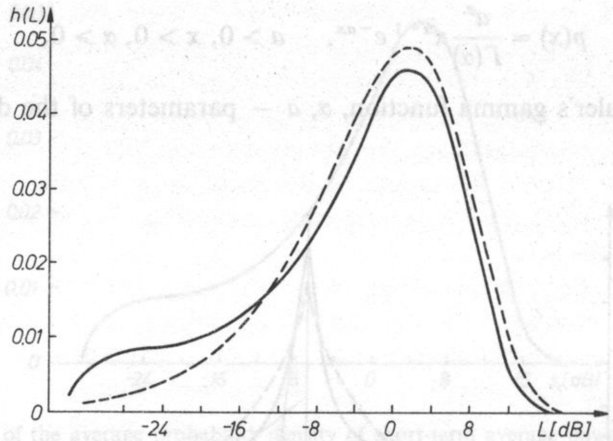


Fig. 9. A comparison of short-term average power distribution of a Polish speech signal with gamma distribution ($a = 1.4$, $\alpha = 1.3$); ——— distribution for a speech signal, – – – gamma distribution

5. Realization and investigation of the model

The final verification of accepted assumptions and selection of model's parameters can be done only on the basis of experimental studies. To this end a numeric parametric model of a signal generator was built in the form of a computer program. Individual blocks of this generator (see Fig. 4) were realized as follows:

"Generator T" and "Gaussian generator" were realized on the basis of the tribution was applied in accordance with conclusions from paragraph 4. Presented in paper [12] algorithm of generation was used.

"Generator T and "Gaussian generator" were realized on the basis of the FPMCRW generator found in the Fortran library of the Odra 1300 system [14].

"Spectrum forming filter"— the power density spectrum of a Polish speech signal was approximated by straight line segments with a slope of 12 and 6 dB/octave, as it is shown in Fig. 10. Connected in a cascade, a low- and a high-pass Butterworth's filter, respectively, were used in the program. "Shaping of a stepwise envelope "and" "shaping of transients" due to a lack of adequate experimental research and the simplicity of the technical realization, the $z_2(t)$ function was formed with a linear approximation between successive values of the $z_1(t)$ function; this is shown in Fig. 11.

The frequency of signal sampling of $f_p = 8$ kHz was accepted in calculations.

Before the experimental work on the optimization of the model was done, the minimum time interval of stationarity of the generated signal was estimated. Hence, the problem was to find how many samples have to be generated by generators $p(T)$ and $p(s)$ in order to stabilize analysed distributions. The minimum time interval of

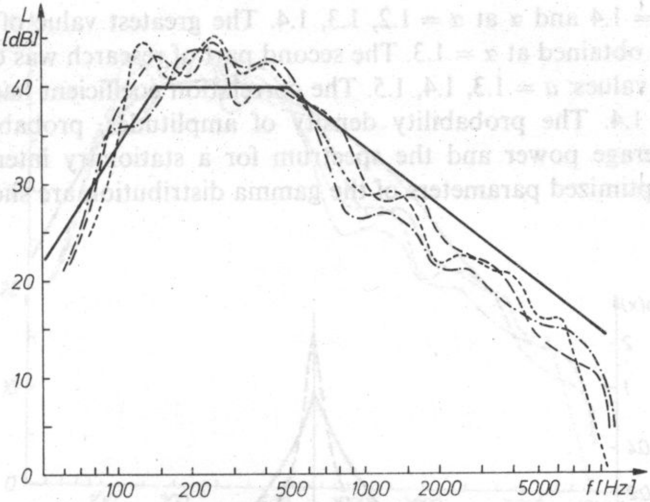


Fig. 10. Power density spectrum of a Polish speech signal. ---- according to JASSEM [4], -·-·- according to ZALEWSKI and MAJEWSKI [11], ······ according to MAJEWSKI et al [6]. ——— approximation

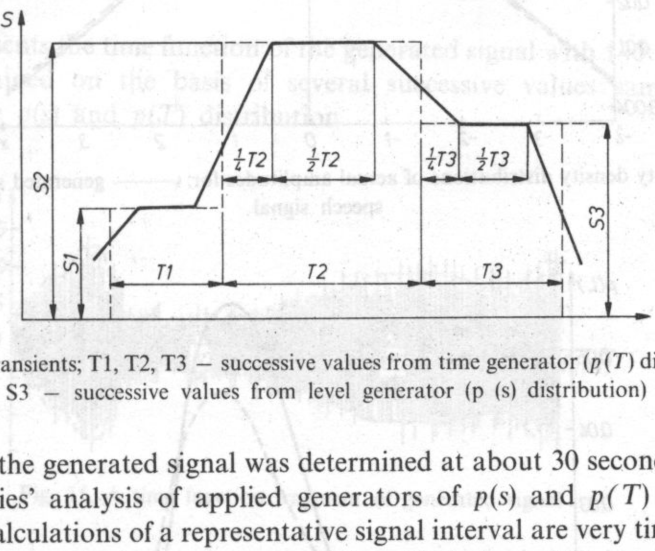


Fig. 11. Shape of transients; T1, T2, T3 — successive values from time generator ($p(T)$ distribution); S1, S2, S3 — successive values from level generator (p (s) distribution)

stationarity of the generated signal was determined at about 30 seconds on the basis of the properties' analysis of applied generators of $p(s)$ and $p(T)$ distributions.

Because calculations of a representative signal interval are very time-consuming, a single program of dynamic, sequential optimization [8] was applied. The correlation coefficient between the probability density curve of instantaneous amplitudes of the generated signal and of natural speech in the range of values of amplitudes $0.3 < |x| < 4$ was accepted as the objective function. The number of investigated parameters was reduced to two, which have the greatest effect on the objective function: parameter α and parameter a of the gamma distribution. In the first part of investigations three 30 s stationary intervals of a signal were generated. The earlier set central value of the parameter of the gamma distribution was accepted at $a = 1.4$ and α at $\alpha = 1.2$, 1.3, 1.4. The greatest value of the correlation coefficient was obtained at $\alpha = 1.3$. The second part of research was done for $\alpha = 1.3$ and following values: $a = 1.3$, 1.4, 1.5. The correlation coefficient had the maximum value for $a = 1.4$. The probability density of amplitudes, probability density of short-term average power and the spectrum for a stationary interval of a signal generated at optimized parameters of the gamma distribution are shown in Figs. 12, 13, 14.

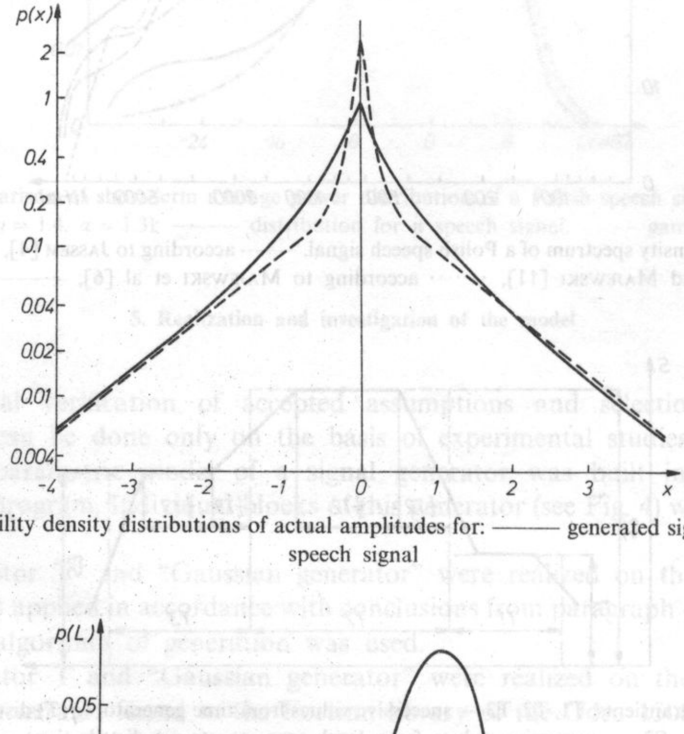


Fig. 12. Probability density distributions of actual amplitudes for: ——— generated signal, ---- Polish speech signal

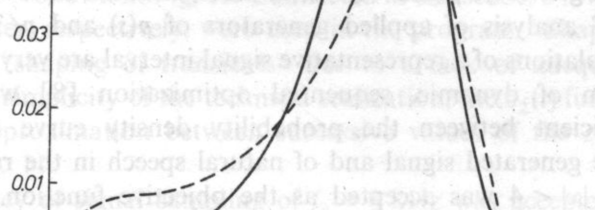


Fig. 13. Probability density distributions of short-term average power of a generated signal ——— and Polish speech signal ----

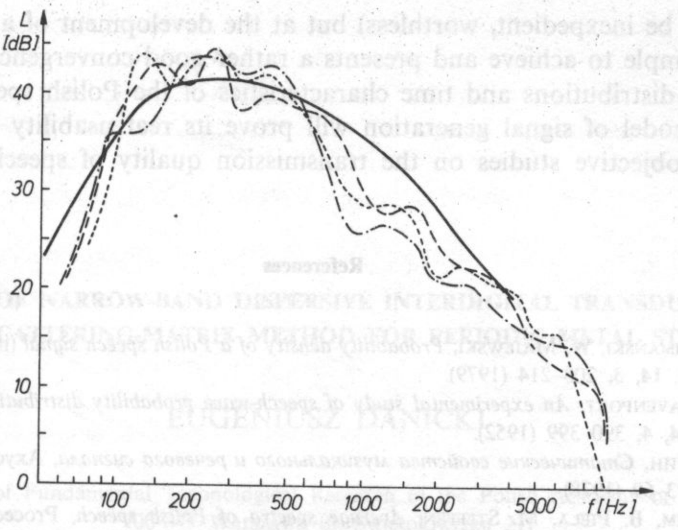


Fig. 14. Power density spectrum of a generated signal ——— and Polish speech signal – – – – according to JASSEM; –·–·– according to ZALEWSKI and MAJEWSKI; ······ according to MAJEWSKI et al.

Fig. 15 presents the time function of the generated signal with 140 mm duration which was obtained on the basis of several successive values sampled by the generator of the $p(s)$ and $p(T)$ distribution.

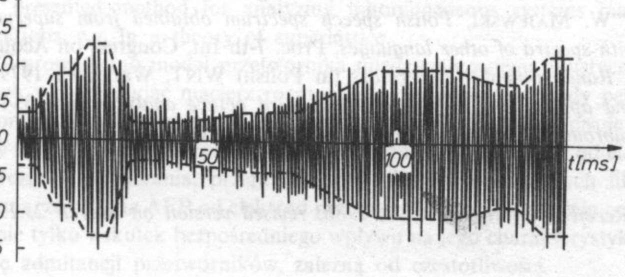


Fig. 15. A time function fragment of generated signal

6. Conclusion

A comparison between probability distribution and characteristics of the generated signal (paragraph 5) and corresponding distributions and characteristics of Polish speech demonstrates a lack of their full compatibility. Yet, our studies were not aimed at the development of a signal with characteristics of the speech signal (this would be inexpedient, worthless) but at the development of a signal which is relatively simple to achieve and presents a rather good convergence with accepted probability distributions and time characteristics of the Polish speech signal. The proposed model of signal generation will prove its real usability when it will be applied in objective studies on the transmission quality of speech.

References

[1] S. Brachmański, W. Majewski, *Probability density of a Polish speech signal* (in Polish) Archiwum Akustyki, **14**, 3, 203–214 (1979).

[2] W. B. Davenport, *An experimental study of speech-wave probability distributions*, J. Acoust. Soc. Amer., **24**, 4, 390–399 (1952).

[3] Я. Инблин, *Статические свойства музикального и речевого сигнала*, Акустический Журнал, **25**, 5, 693–69 (1978).

[4] W. Jassem, B. Piela, M. Steffen, *Average spectra of Polish speech*, Proceedings of Vibration Problems, **2**, 59–71 1959.

[5] П. Г. Капотов, Ю. П. Максимов, В. В. Марков. *Нелинейные искажения речевого сигнала*, Электросвязь, **8**, 29–34 (1976).

[6] W. Majewski, H. B. Rothman, H. Hollien, *Acoustic comparisons of American English and Polish*, Journal of Phonetics, **5**, 247–251 (1977).

[7] B. J. Mc Dermott, *Multidimensional analyses of circuit quality judgements*, J. Acoust. Soc. Amer., **45**, 774–781 (1969).

[8] Z. Polański, *Contemporary methods of experimental studies* (in Polish) WP, Warszawa 1978.

[9] А. В. Римский-Корсаков, *Статические свойства радиовещательного сигнала*, Акустический Журнал, **6**, 3, 360–369 (1960).

[10] D. Wolf, *Analytische Beschreibung von Sprachsignalen*, AEU, **31**, 9, 392–398 (1977).

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

[12] R. Zieliński, *Random-number generators* (in Polish) WNT, Warszawa 1979.

[13] *Instruction and applications of the real time third octave analyzer type* 3347, Bruel-Kjaer, 1971.

[14] *Fortran – subroutine library* (in Polish) Zeszyty Elwro, nr 13044 2.

Received on November 2, 1986; revised version on March 2, 1987.

ARCHIVES OF ACOUSTICS
13, 3-4, 237-253 (1988)

ANALYSIS OF NARROW-BAND DISPERSIVE INTERDIGITAL TRANSDUCERS BY THE SCATTERING-MATRIX METHOD FOR PERIODIC METAL STRIPS

EUGENIUSZ DANICKI

Institute of Fundamental Technological Research of the Polish Academy of Sciences
(00-049 Warszawa ul. Świętokrzyska 21)

A model of the interdigital transducer of surface acoustic wave (SAW) is introduced on the strength of the scattering matrix for a single metal strip of the periodic system deposited on a piezoelectric substrate surface. The model describes properties of the transducer within the wide frequency band and takes into account the SAW reflection of electrical origin from the metal strips. The calculations performed for narrow-band dispersive filters show that the SAW reflections from the transducer electrodes distort the filter response not only as a result of the immediate effect upon the filter frequency characteristic, but also through the frequency-dependent change in the transducer admittance. A presented method for analyzing inhomogeneous systems may have some other applications, e.g. in a theory of superlattice.

Wprowadzono model przetwornika międzypalczastego akustycznych fal powierzchniowych wykorzystując macierz rozpraszania dla jednej elektrody periodycznego układu, położonego na powierzchni podłoża piezoelektrycznego. Model opisuje własności przetwornika w szerokim pasmie częstotliwości i uwzględnia odbicie AFP od elektrod, pochodzenia elektrycznego. Obliczenia, przeprowadzone dla wąskopasmowych filtrów dyspersyjnych wykazują, że odbicia AFP od elektrod przetworników zniekształcają odpowiedź impulsową filtru nie tylko wskutek bezpośredniego wpływu na jego charakterystykę, ale także poprzez zmianę admitancji przetworników, zależną od częstotliwości.

1. Introduction

The commonly used equivalent scheme of a pair of metal strips of an interdigital transducer (IDT) [1] is based upon the theory of bulk-wave transducers. That is why, in order to take account of the phenomena connected closely with SAW, this scheme had to be appropriately modified. Thus the SAW reflection has been considered from the metal strips in the transducer fundamental frequency-band [2] and on the harmonic frequencies [3]. The so-called element factor [4] is introduced that describes quantitatively the phenomenon of generating and detecting the surface wave by the transducer electrodes, this phenomenon being substantially different from the analogous electromechanical transformation for the bulk waves.

A different approach will be assumed below to the theory of interdigital transducers, this approach being based upon the strict theory of SAW. For instance, the above-mentioned element factor results naturally from the theory of periodic metal strips situated on a piezoelectric substrate surface [5] (the case of near-surface waves being considered in [6]).

The IDT model introduced in the present paper is based upon the scattering matrix for the periodic metal strips [7] as well as upon circuit theory. The reasoning that permits the theory [7] to be applied to the non-periodic metal strips within the wide frequency-band including the SAW Bragg reflection band, is similar to that in the considerations presented in [8], though the equivalent scheme will not be derived in the present paper as being non-physical for the surface waves. It will be shown that the IDT model based exclusively upon the scattering matrix for the metal strips is sufficient for analyzing the SAW transducers and filters.

In order to simplify matters it is assumed below that the transducer metal strips have a width equal to the spacing between them (this spacing can vary along the transducer). Account will be taken of the SAW reflection brought about exclusively by the electrical interaction of the conducting metal strip with the wave. The mechanical properties of the metal strips can be taken into consideration in the similar way as in Refs [9–11]. Bulk waves are left out of account.

As mentioned above, a strict theory exists only for the periodic system of metal strips. Extension of its results to the case of non-periodic transducers will be effected by comparing the appropriate results of the theory of non-homogeneous transmission lines, these results being shown in the next Section along with the results of the theory [7] shown in the Section after next. The resulting model of the transducer, as presented in Sections 4 and 5, has been used for analyzing the dispersive delay-lines, the analysis being similar to that presented in [12]. In contrast with [12], in Section 6 are analyzed the narrow-band filters with a considerably longer time response.

2. Non-homogeneous transmission line

We now consider a transmission line (Fig. 1) composed of three sections of different but homogeneous lines having the lengths $l/2$, l' and $l/2$. The impedances of these lines are R or R'. The wave velocity in these lines are denoted by v and v' (for simplification it is assumed that $l/v = l'/v'$ and $R \geqslant R'$).

The relation between voltage and current complex amplitudes on the line terminals is the following ($\alpha = \omega l/v$, ω – angular frequency)

$$\begin{bmatrix} i_{n+1} \\ e_{n+1} \end{bmatrix} = \begin{bmatrix} \cos(\tau) & -j\sin(\tau)/Z \\ -jZ\sin(\tau) & \cos(\tau) \end{bmatrix} \begin{bmatrix} i_n \\ e_n \end{bmatrix}, \tag{1}$$

where the phase shift τ is expressed by $(j=\sqrt{-1})$

$$e^{-j\tau}=\frac{(R+R')^2}{4RR'}[\cos(2\alpha)-\Delta^2-j2\sin(\alpha)(\cos^2(\alpha)-\Delta^2)^{1/2}], \qquad (2)$$

$\Delta=(R'-R)/(R'+R)$, while the wave-impedance is

$$Z=R[(\cos(\alpha)-\Delta)/(\cos(\alpha)+\Delta)]^{1/2}. \qquad (3)$$

This is at the same time the impedance of a semi-infinite chain composed of successively linked two-port networks described by the relation (1).

It should be noted that the wave impedance Z may have a complex value. This occurs when the wave propagating an infinite chain composed of the networks, as depicted in Fig. 1, is in synchronism with the periodic inhomogeneity of such a transmission line. Then $\tau\approx\pi/2$, and as results from (2), the wave number of the wave $r=\tau/(2l)$ also has a complex value ($2l=l+l'$ is the period of the transmission line over the length of which period the wave changes its phase by τ). The frequency range for which the wave number is complex is called a stop-band.

Next we consider the chain (transmission line) composed of different two-port networks presented in Fig. 1 (with different parameters R, R' etc). The network with the number n is described by (1) on assuming the values τ_n, Z_n instead of τ, Z, respectively. We introduce the notion of the wave α^+ propagating to the right and having the wave number r and the complex amplitude a^+, as well as that of the wave α^+ propagating to the left and having the wave number $-r$ and the amplitude a^-. These amplitudes are defined for each network of the chain in the following standard fashion (the wave phases are referred to the centers of networks)

$$\begin{aligned} i_n&=(\alpha^+\exp(j\tau_n/2)-\alpha^-\exp(-j\tau_n/2))/\sqrt{Z},\\ e_n&=(\alpha^+\exp(j\tau_n/2)+\alpha^-\exp(-j\tau_n/2))\sqrt{Z}. \end{aligned} \qquad (4)$$

The relations obtained from (1) and (4), between the amplitudes of the waves α^+ and the waves α^- in the successive networks of the non-homogeneous transmission line under consideration can be written in the form (the equality of currents and voltages on the boundaries of the adjoining networks should be taken into account)

$$\alpha_n^+=T_{n-1,n}\alpha_{n-1}^+\exp[-j(\tau_{n-1}+\tau_n)/2]+\Gamma_{n,n-1}\alpha_n^-\exp(-j\tau_n), \qquad (5)$$

and similarly for the wave α_n^-.

The phenomena described by the above relation and accompanying the wave propagation in non-homogeneous transmission line are the following:

- change in the wave phase after transmission of the network from one edge to the other is τ_n,
- wave transmission through a connection between the adjoining networks. T_{LE} is the transmission coefficient for the wave passing from the network L to the network E,

$$T_{LE}=2(Z_LZ_E)^{1/2}(Z_L+Z_E)^{-1}. \qquad (6)$$

– wave reflection, the reflection coefficient of the wave propagating in the network L from the boundary with the network E is

$$\Gamma_{LE} = (Z_E - Z_L)(Z_L + Z_E)^{-1}. \tag{7}$$

The relations (5)–(7) are also valid in the stop band, where Z_n has a complex value (however note that the above interpretations may not be correct). That is because (4) has been introduced entirely formally (the square-root sign in (6) is of no importance, it must, however, be consistently the same as in (4)). Relations (6) and (7) define the elements of the scattering matrix for the successive networks of the inhomogeneous transmission line.

In conclusion we consider the wave reflection from the boundary between the homogeneous line of the impedance $R = 1$, and the semi-infinite line periodically inhomogeneous, composed of the networks depicted in Fig. 1, and having the wave impedance Z. The reflection coefficient of the wave is

$$\alpha^+/\alpha^- = \Gamma = (1-Z)/(1+Z). \tag{8}$$

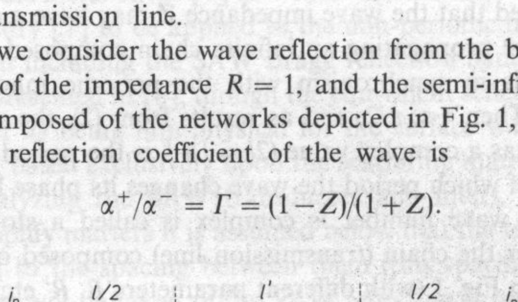


Fig. 1. Transmission line

3. Selected results of the SAW theory

In Ref [7] the periodic system is considered of the metal strips with the period Λ (for simplification the metal strips under study have the width $\Lambda/2$) and situated on the surface of the half-space of a piezoelectric substrate characterized by $\Delta v/v$, v (SAW velocity) and ε_{eff}(effective surface permittivity of the substrate). All the metal strips are grounded with the exception of one having the number m, to which a potential of an angular frequency ω is applied, this potential in the complex form being equal to $V_m \exp(j\omega t)$.

The current that flows into the metal strip of the number n and of the length W is

$$J_n = -j\omega W C_{|n-m|} V_m + W I_n, \tag{9}$$

where the first component is defined exclusively by the dielectric properties of the substrate, and the other component is connected with the surface accustic wave generated at the metal strip m. The quantities occurring in (9) are defined by simplified relation (the simplifications concerning mainly I_0 and E_f)

$$C_n = \frac{1}{\pi}\varepsilon_0(1+\varepsilon_{\text{eff}})\left(n^2-\frac{1}{4}\right)^{-1}, \tag{10a}$$

$$I_n = YV_m \exp(-j|n-m|\tau), \tag{10b}$$

$$Y = E_f/Z, \tag{10c}$$

$$E_f = \omega\varepsilon_0(1+\varepsilon_{\text{eff}})4\pi(\Delta v/v)s(1-s)\sin(\pi s), \tag{10d}$$

$$Z = [(\omega-\omega_3)/(\omega-\omega_1)]^{1/2}, \tag{10e}$$

$$\tau = r\Lambda, \tag{10f}$$

$$r = K/2+\frac{1}{v}[(\omega-\omega_1)(\omega-\omega_3)]^{1/2}, \tag{10g}$$

$$\omega_1 = Kv/2, \tag{10h}$$

$$\omega_3 = \omega_1(1-s(1-s)\Delta v/v), \tag{10i}$$

$$s = \frac{1}{2}\omega/\omega_1. \tag{10j}$$

These relations are valid for $0 < s < 1$, i.e. in the fundamental frequency band.

In the periodic system of metal strips described among others in [7], there occur two waves, a forward-travelling wave of the wave number r and the amplitude α_0 and a backward-travelling wave of the wave-number r–K and the amplitude α_{-1}. These amplitudes are connected by the relation

$$\alpha_{-1}/\alpha_0 = (1-Z)/(1+Z), \tag{11}$$

where Z is defined by (10e).

The relation (11) is made use of in [13] for constructing the boundary conditions for a bounded system of metal strips. Let the wave $\alpha^+\exp(-jkx)$, $x < 0$, $k = \omega/v$, propagating on the free surface of a piezoelectric, falls on the area covered by periodic metal strips ($x > 0$). In the area $x < 0$ account must also be taken of the reflected wave $\alpha^-\exp(jkx)$, while in the area of the metal strips two waves exist $\alpha_0\exp(-jrx)+\alpha_{-1}\exp j(K-r)x$ (the factor $\exp j\omega t$ is omitted in this paper). The boundary conditions for $x = 0$ are [13]

$$\alpha_0 = \alpha^+, \quad \alpha_{-1} = \alpha^-,$$

hence, on consideration of (11) one obtains the relation (8). The analogy with the transmission line discussed above is obvious, particularly when noticing that Z defined by (3) can be written in the form similar to (10e) for $\alpha \approx \pi/2$.

4. Mathematical model of metal strips

Analogies between surface waves and waves in the transmission line have been taken into account when introducing the equivalent scheme for transducer metal strips (Refs [1, 8, 14]). It may be stated here that (5) describes the SAW propagation under the system of metal strips, if the parameters occurring there are defined by the relations (10), (6) and (7). In the case of a non-periodic system of metal strips, in the relation (10) Λ_n should be assumed as equal to the "local period" of this system (which is justified only for a nearly periodic system of metal strips, which take place e.g. in the case of narrow-band dispersive filters). For the consecutive metal strips the parameters Y_n, Z_n, and τ_n will occur.

When supplying the metal strips from an external source is taken into account this leads to the modification of the relation (5). It should then be assumed that under the metal strip a change occurs in the amplitudes of the waves α^+ and α^- by a certain value dependent upon the potential of the metal strip. Similarly, the current flowing to the metal strip depends upon the waves travelling under it. The relevant considerations can be found in [7] and their result, on consideration of (5), is contained in the following relations

$$\alpha_n^+ = T_{n-1,n}(\alpha_{n-1}^+ + V_{n-1}(Y_{n-1}/2)^{1/2})\exp[-j(\tau_{n-1}+\tau_n)/2] + \Gamma_{n,n-1}[V_n(Y_n/2)^{1/2} + T_{n+1,n}(\alpha_{n+1}^- + V_{n+1}(Y_{n+1}/2)^{1/2})\exp[-j(\tau_n+\tau_{n+1})/2]]\exp(-j\tau_n), \quad (12)$$
$$I_n = (\alpha_n^+ + \alpha_n^-)(2Y_n)^{1/2} + Y_n V_n,$$

(similarly for α^-, the square-root sign is to be chosen consistently as in (6)).

The relations are illustrated in Fig. 2. They describe the propagation, generation and detection of SAW in the area of the metal strips, the relation (9) taking additionally into account the mutual capacitance of the metal strips. In order to complete the description the value should be adopted yet of the wave impedance on the free surface that limits the system of the metal strips. The discussion in Section 3 justifies the adoption of this value equal to 1. The error of leaving out of account the wave reflection on the border of two areas with different wave numbers ($r > k$) may be usually disregarded.

It should be emphasized that the above depicted model (Fig. 2) takes no account of either the bulk waves or of the elastic properties of metal strips, and the only cause of wave reflections from the metal strips is the short-circuit effect for electric potential under metal strips (this is so-called $\Delta v/v$ reflection). However, both Ref [9] and Refs [10] and [15] indicate that the mechanical properties of the metal strips may be included into the model through appropriate correction of the parameters that occur there.

In conclusion, it is worth noting that the case under study in [7] and [16] of a bounded periodic system of metal strips a part of which have an established non-zero potential may be described as well by the model in Fig. 2 [14], [17].

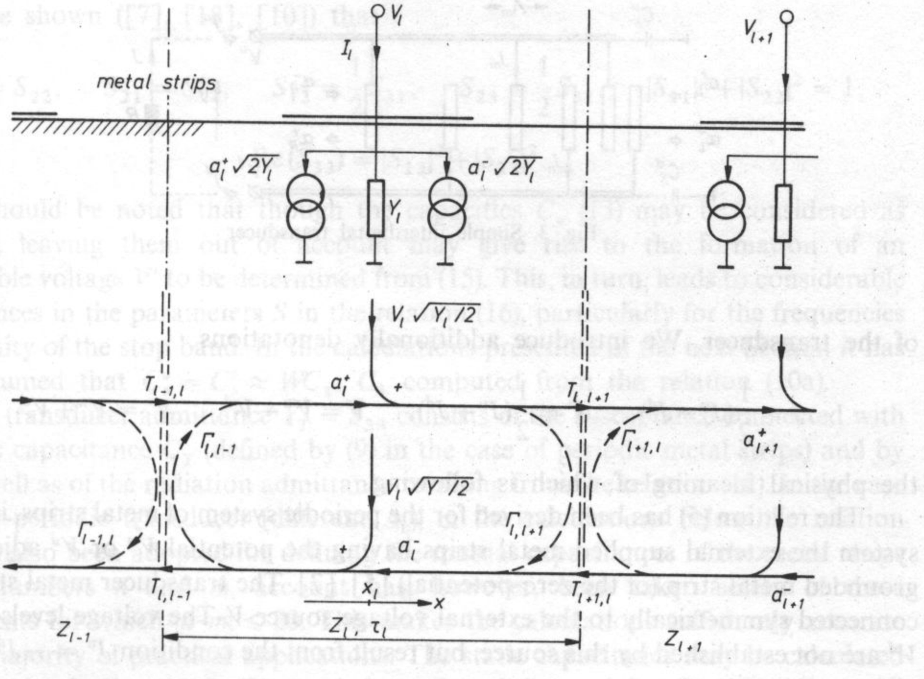


Fig. 2. Metal strip model resulting from the scattering matrix

5. Analysis of interdigital transducers

On the basis of the model introduced in the preceding Section the scattering matrix will be derived below for a simple nearly periodic ($\Lambda_n \approx \Lambda_{n+1}$) transducer (Fig. 3) with the aperture W (the width of the metal strip of the number n is equal to $\Lambda_n/2$).

The quantities $\alpha_{L,R}^{+,-}$ shown in Fig. 3 denote the complex amplitudes of acoustic waves on the left-hand (L) and the right-hand (R) side of the transducer and propagating to the left ($-$) or to the right ($+$). $J^{u,d}$ and $V^{u,d}$ denote both the complex amplitudes of the current flowing to the transducer bars (u – upper one, d – lower one) and the potentials of these bars, respectively. In addition it is convenient to introduce total SAW amplitudes as $A = \alpha\sqrt{W}$ with indices L, $+$ and $-$, of the same meaning as in the case of α ($|A|^2/2$ is the power of the SAW beam of the width W).

In accordance with Fig. 3 one obtains

$$J^{u,d} = \sum_{l=l^{u,d}} J_l + j\omega C_x^{u,d} V^{u,d} \tag{13}$$

where $l^{u,d}$ are the numbers of the metal strips connected to the upper or the lower bar

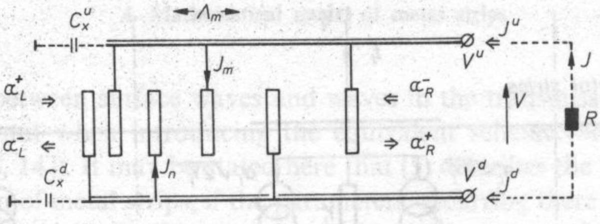


Fig. 3. Simple interdigital transducer

of the transducer. We introduce additionally denotations

$$J = \frac{1}{2}(J^u - J^d), \quad J' = \frac{1}{2}(J^u + J^d), \quad V = V^u - V^d, \quad V' = V^u + V^d \tag{14}$$

the physical meaning of which is following.

The relation (5) has been derived for the periodic system of metal strips, in which system the external supplied metal strips having the potential V^u or V^d adjoins the grounded metal strip (of the zero potential) [5], [7]. The transducer metal strips are connected symmetrically to the external voltage source V. The voltage levels V^u and V^d are not established by this source, but result from the condition $J^u = -J^d$, that is

$$J' = 0. \tag{15}$$

Consequently, the potential V' denotes the magnitude of the non-symmetrical supply to the transducer metal strips, $V' = 0$ when $V^u = -V^d$. In this connection, in the case under investigation of a bounded system of metal strips, situated on a substrate of a finite thickness, the capacity is to be taken into account between the metal strip and the filter housing, this capacity corresponding to the grounded metal strips discussed previously. These are capacities $C_x^{u,d}$ taken into consideration in the relation (13) and in Fig. 3.

The relations (12) which constitute the set of equations for the amplitudes of acoustic waves, currents and potentials of the metal strips in each section of the transducer of a length Λ_n, may be solved recurrently. One obtains relations between the amplitudes $A_{R,L}^{+,-}$ and the quantities as given in (14) these relations being similar to the relation presented below for J'

$$J' = \alpha A^+ + \beta A^- + \gamma V + \delta V'.$$

The condition (15) permits eliminating V'. Ultimately one obtains

$$\begin{aligned} A_R^+ &= S_{11} A_L^+ + S_{12} A_R^- + S_{13} V, \\ A_L^- &= S_{21} A_L^+ + S_{22} A_R^- + S_{23} V, \\ J &= S_{31} A_L^+ + S_{32} A_R^- + S_{33} V. \end{aligned} \tag{16}$$

It can be shown ([7], [18], [10]) that

$$S_{11} = S_{22}, \quad S_{21} = S_{12}^*, \quad S_{13} = \frac{1}{2}S_{31}, \quad S_{23} = \frac{1}{2}S_{32}, \quad |S_{11}|^2 + |S_{22}|^2 = 1,$$

$$\mathrm{Re}\,(S_{33}) = |S_{13}|^2 + |S_{23}|^2.$$

It should be noted that though the capacities C_x (13) may be considered as parasitic, leaving them out of account may give rise to the formation of an appreciable voltage V' to be determined from (15). This, in turn, leads to considerable disturbances in the parameters S in the relation (16), particularly for the frequencies in a vicinity of the stop band. In the calculations presented in the next Section it has been assumed that $C_x^u = C_x^d \approx WC_1$, C_1 computed from the relation (10a).

The transducer admittance $Y_T = S_{33}$ consists of the susceptance connected with the static capacitance C_T (defined by (9) in the case of periodic metal strips) and by (13), as well as of the radiation admittance resulting from the relation (12). In the case of nearly-periodic transducer (different Λ_n), in the calculations below the relation (10a) has also been adopted for defining the mutual capacitances of the metal strips having numbers n and m, account has, however, been taken solely of these components for which $|n-m| \leqslant 20$. This makes the calculations sufficiently accurate for the majority of practical applications. The static capacitance may be calculated separately (neglecting the SAW, that is for $\Delta v/v = 0$) by applying (9), (10), (13) as well as the relation (16) which is reduced then to $J = j\omega C_T V$.

6. Numerical analysis of narrow-band chirp filters

The dispersive filter under analysis (Fig. 4) consists of two transducers, a broad-band one with $n = 3$ pairs of splitted metal strips ($m = 4$ of a metal strip falls to a wave-length), and a dispersive one (with linear dispersion) characterized by the $B = 4$MHz pass-band and duration of the time response $T = 8\mu$ (or $T = 16\mu s$) the center frequency of the filter being 30 MHz. Analysis has been performed of dispersive transducers with $m = 2$ metal strips to a wave-length and with sectional metal strips ($m = 3$ or $m = 4$ metal strips to a wave-length). The calculations have been carried out for filters with an aperture $W = 4$ mm on the substrate of SiO_2YX or $LiNbO_3$YZ. In the calculations the load impedance R has been assumed to be equal to the impedance of the generator supplying the filter ($R = 50\,\Omega$ or $R = 1\,\Omega$).

Fig. 4. Configuration of the down-chirp filter

The filter frequency response $H(f)$ is

$$H_l(f) = 2RS_{l3}^{(1)} S_{3l}^{(2)} [(1+S_{33}^{(1)} R)(1+S_{33}^{(2)} R)]^{-1}, \tag{17}$$

where the superscripts (1) and (2) denote the parameters of a broad-band transducer and a dispersive one, respectively, the subscript $l = 1$ for the down-chirp filter, while $l = 2$ for the up-chirp filter. TTS is left out of account in (17). The impulse response $h(t)$ and the compressed pulse $c(t)$ is calculated with the aid of the FFT algorithm:

$$h_l(t) = F^{-1}(H_l(f)), \quad c(t) = F^{-1}(H_1 H_2).$$

Fig. 5a depicts the calculated frequency characteristic of the delay line composed of two broad-band transducers for a substrate of lithium niobate or quartz. The insertion losses, amounting at center frequency to 32. 7 dB and 79. 5 dB increase by about 1.5 dB to a transducer at the band edges. The same transducers have been assumed in the calculations presented below.

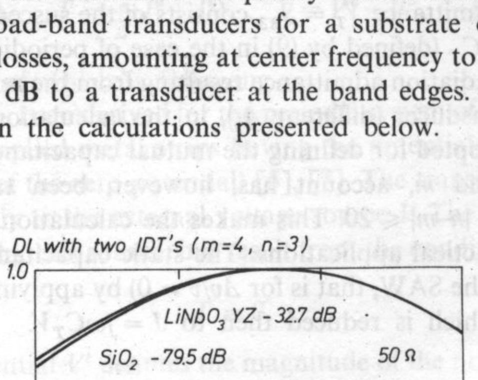


Fig. 5a

For later comparisons Figs 5b and 5c present in a certain sense ideal responses of dispersive filters in which dispersive apodized transducers of periodic metal strips are applied (Ref [19]). The apodization is calculated on the basis of the relation (7.2) given in [7] ($f_0 = 60$ MHz was applied, that corresponds to four ($m = 4$) metal strips on a wave-length at center frequency of the filter). The synchronous SAW reflections from the metal strips in the pass-band of these transducers do not occur. The frequency and time responses presented are rounded as a result of the above-discussed form of the frequency characteristic of the broad-band transducer. The conductance G and the susceptance B of the dispersive transducers are also shown, as well as the shape of compressed pulse.

The calculated responses of similar filters on a substrate of $LiNbO_3$ (Figs 5d and 5e) are more distorted. This results from an appreciably higher admittance value of the dispersive transducers ($G+jB$), this value being dependent upon frequency. By comparing the filter responses calculated at different loads R it is seen that the effect of the factor $(1+RS_{33}^2)$ occurring in (16), must be considerable for $R = 50\,\Omega$. This

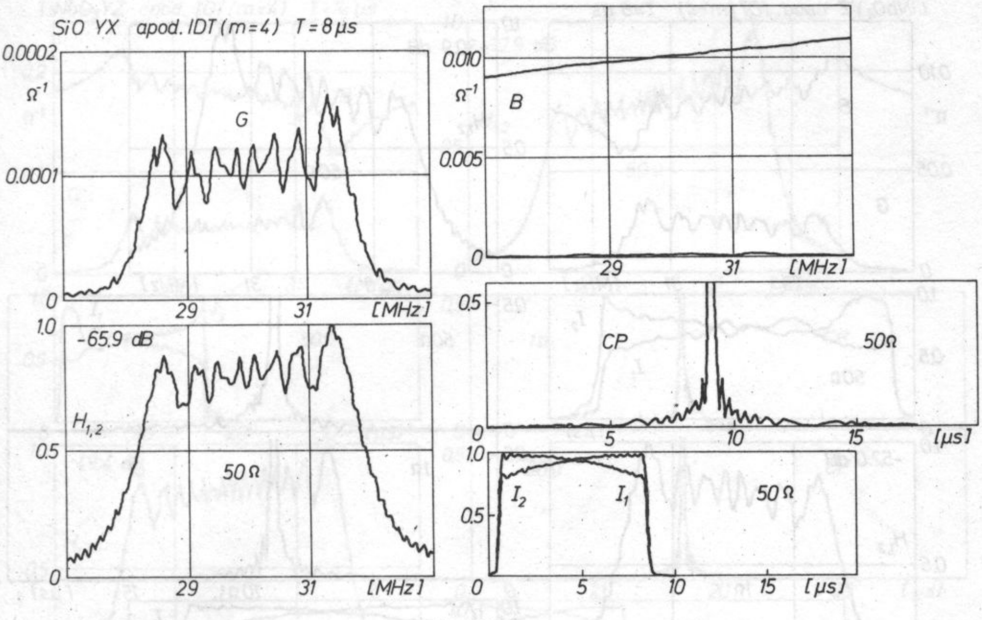

SiO YX apod. IDT (m=4) T = 8 μs
0.0002
Ω⁻¹
G
0.0001
0
29
31
[MHz]
0.010
Ω⁻¹
B
0.005
0
29
31
[MHz]
1.0
-65.9 dB
H₁,₂
0.5
50 Ω
0
29
31
[MHz]
0.5
CP
50Ω
0
5
10
15
[μs]
1.0
I₂
I₁
50 Ω
0.5
0
5
10
15
[μs]

Fig. 5b

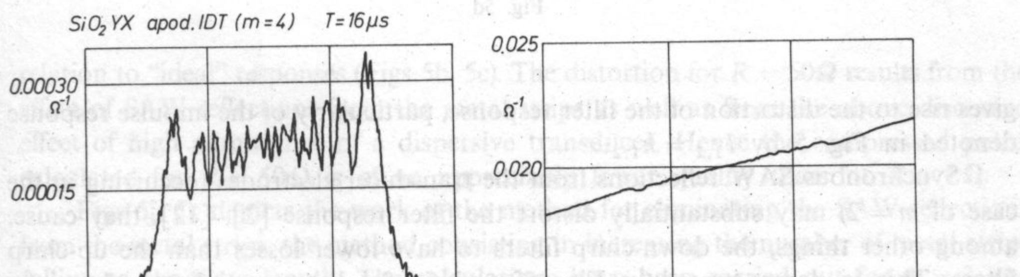

SiO₂ YX apod. IDT (m=4) T = 16 μs
0.00030
Ω⁻¹
0.00015
0
29
31
[MHz]
0.025
Ω⁻¹
0.020
0.015
29
31
[MHz]
1.0
-65.1 dB
H₁,₂
0.5
50Ω
0
29
31
[MHz]
0.5
CP
50Ω
0
5
10
15
[μs]
1.0
50Ω
0.5
I₂
I₁
5
10
15
[μs]

Fig. 5c

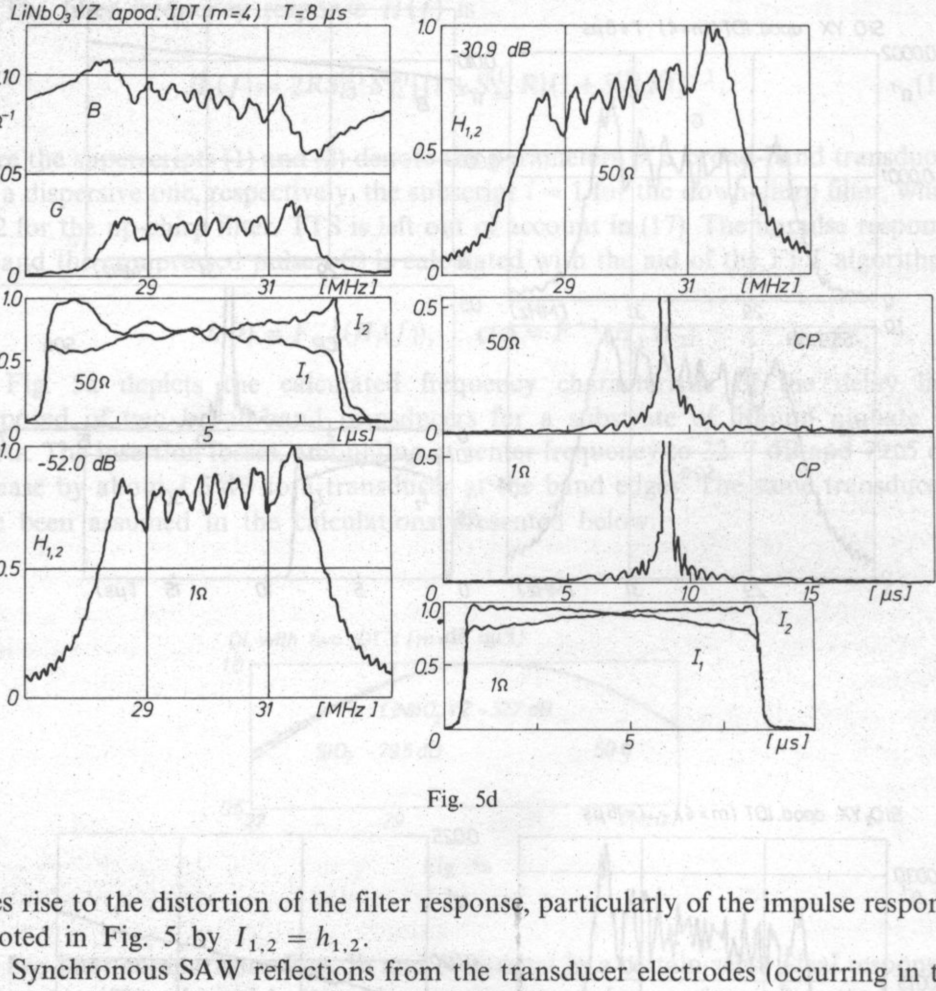


Fig. 5d

gives rise to the distortion of the filter response, particularly of the impulse response denoted in Fig. 5 by $I_{1,2} = h_{1,2}$.

Synchronous SAW reflections from the transducer electrodes (occurring in the case of $m = 2$) may substantially distort the filter response [2], [12], they cause, among other things, the down-chirp filters to have lower losses than the up-chirp filters. This is demonstrated in Figs 5f and 5g for lithium niobate, where the reflections are fairly large as a result of large $\Delta v/v$. For quartz (Figs 5h and 5i) where $\Delta v/v$ is of inconsiderable value, the effect of the SAW reflections upon the filter characteristics is small.

In addition, in Figs 5f–5i the coefficient is shown of SAW reflections from a short-circuited dispersive transducer $\Gamma_0 = |S_{12}|$ and from a non-loaded transducer Γ_v(Γ_v is calculated from (16) on assuming that $J = 0$ and on eliminating V, as $|A_L^-/A_R^+|$). It is noteworthy that Γ_v is smaller than Γ_0. Both these coefficients for quartz are substantially smaller than for lithium niobate, the effect of unidirectional performance of dispersive transducers for quartz practically does not occur either. (Note: only $\Delta v/v$ reflections are taken into account in the paper).

The filter responses as presented in Figs 5f, 5g are considerably distorted in

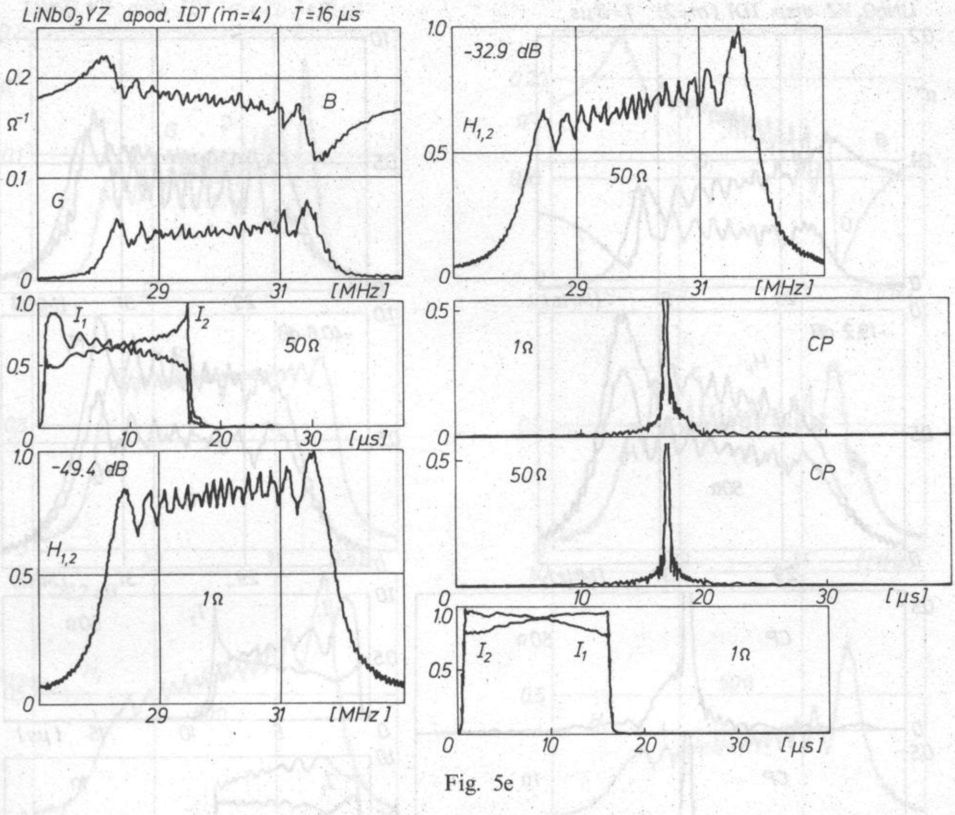


Fig. 5e

relation to "ideal" responses (Figs 5b, 5c). The distortion for $R = 50\,\Omega$ results from the effect of SAW reflections from the metal strips as well as from the above discussed effect of high admittance of a dispersive transducer. Hence the compressed pulse, calculated for $R = 50\,\Omega$ is of an appreciably lower quality than for $R = 1\,\Omega$.

Figs 5j, 5k depicts the work of the method for eliminating the SAW reflections from the metal strips, the method consisting in increasing the number of metal strips falling to the wave-length. The calculations have been carried out for $m = 3$ and $m = 4$ (splitted metal strips). The responses calculated are similar to these presented in Fig. 5d where noteworthy are: low coefficient Γ_0 in both cases, losses for $m = 4$ lower than those for $m = 3$, and insertion losses in general lower than those for the apodized transducers (Fig. 5d).

The presented frequency and time responses of different filters ($T = 8\mu s$ or $16\mu s$, $m = 2$, 3 or 4), calculations of transducer admittance as well as of the pulse after the compression for different load values ($R = 50\,\Omega$ or $1\,\Omega$) permit assessing in general the magnitudes of the parasitic effects (SAW reflections from the metal strips, effect of electrical matching of a transducer with the load, as well as TTS connected with Γ_0 and Γ_v) that occur in arbitrary narrow-band filters (the case of $s < 1$ in the relation (10)).

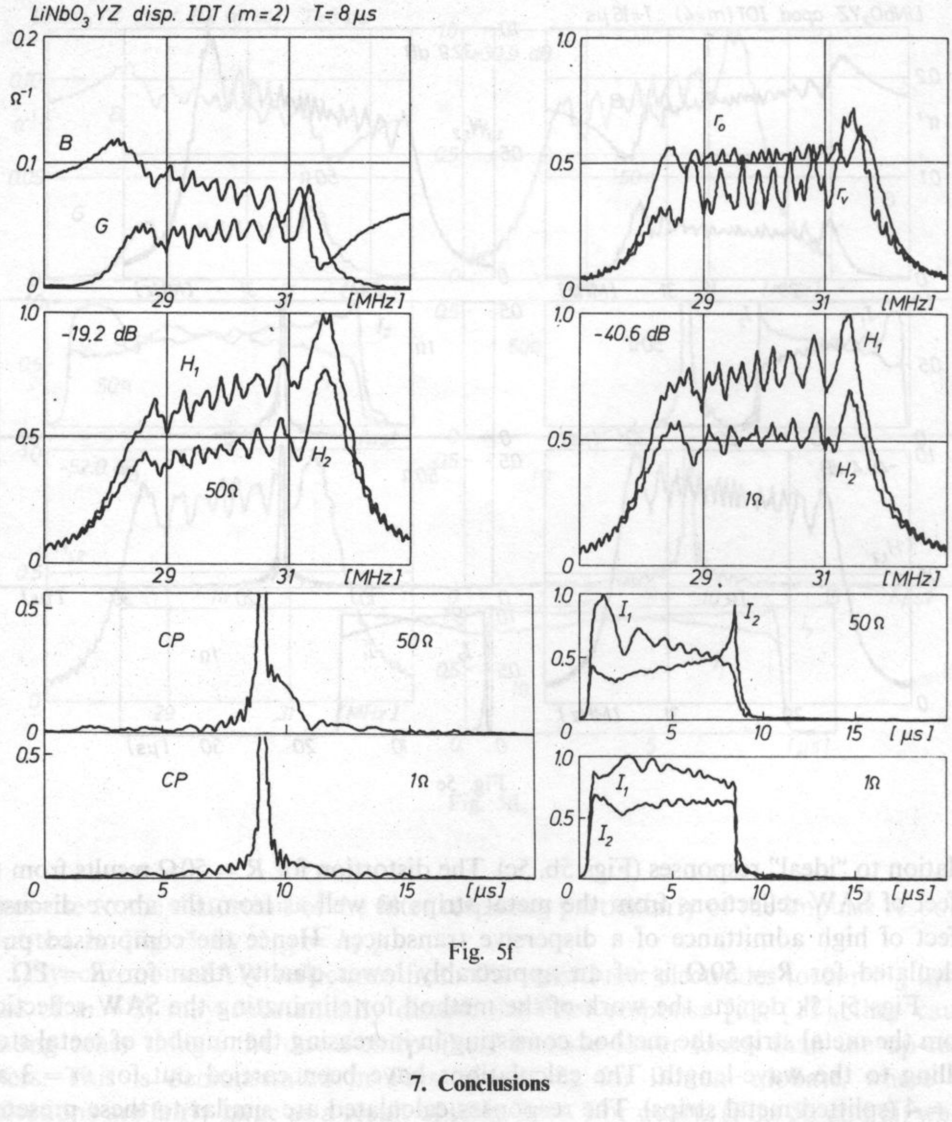


Fig. 5f

7. Conclusions

The above simple model of the interdigital transducer electrodes based upon the scattering matrix, has been derived from the strict theory of periodic metal strips. The model does not include any artificial non-physical parameters and may be readily employed for calculating the scattering matrix of interdigital transducers. The model is open to further modifications such as taking account of mechanical properties of metal strips and bulk waves, and even diffraction [9], [7]. Similarly it is possible to broaden the frequency band of the correctness of the model (for $s > 1$). Application of the above introduced model to the analysis of unidirectional transducers is obvious.

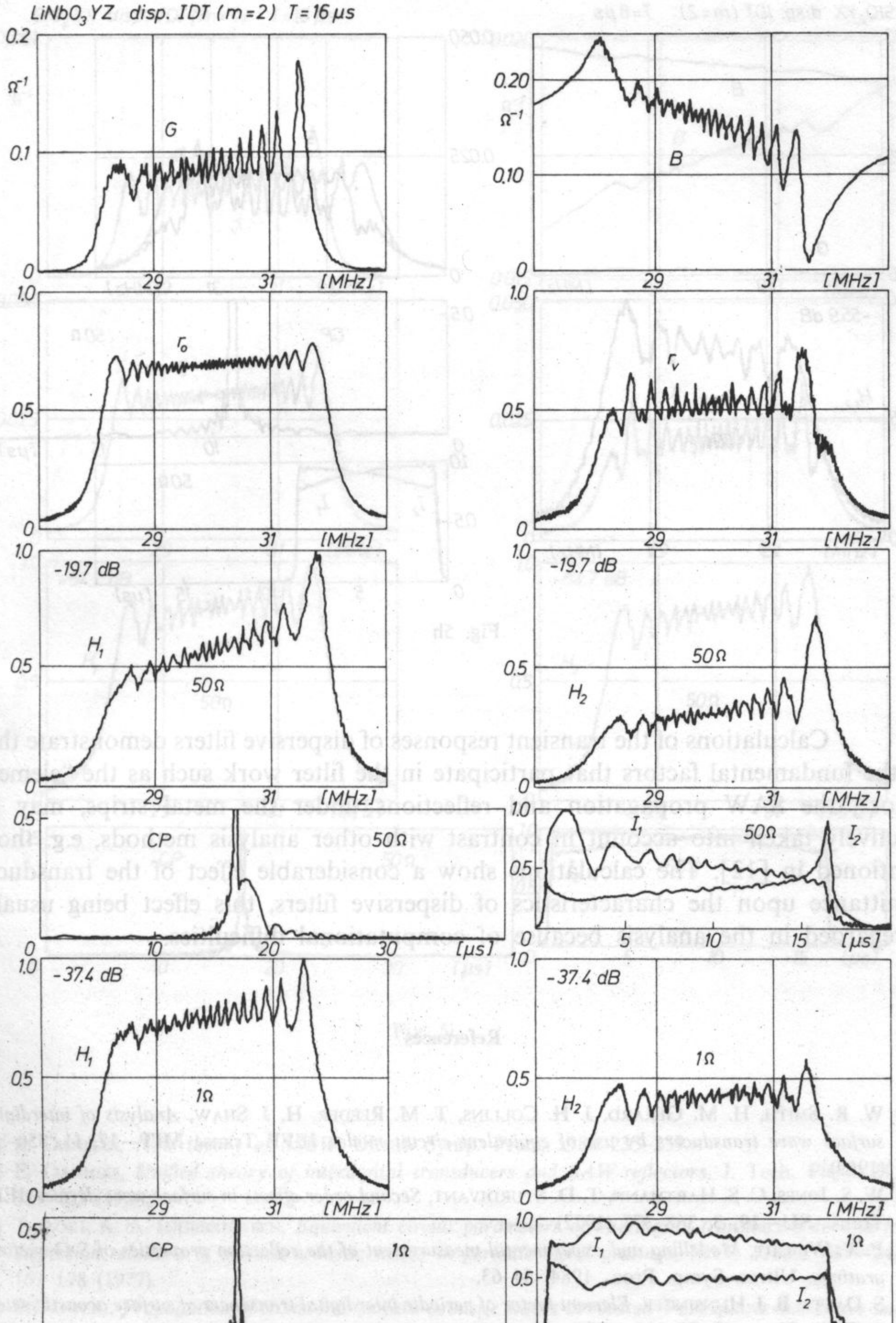

LiNbO$_3$YZ disp. IDT (m=2) T=16 μs
G
B
Ω^{-1}
r_o
r_v
[MHz]
-19.7 dB
H_1
H_2
50Ω
CP
I_1
I_2
[μs]
-37.4 dB
1Ω

Fig. 5g

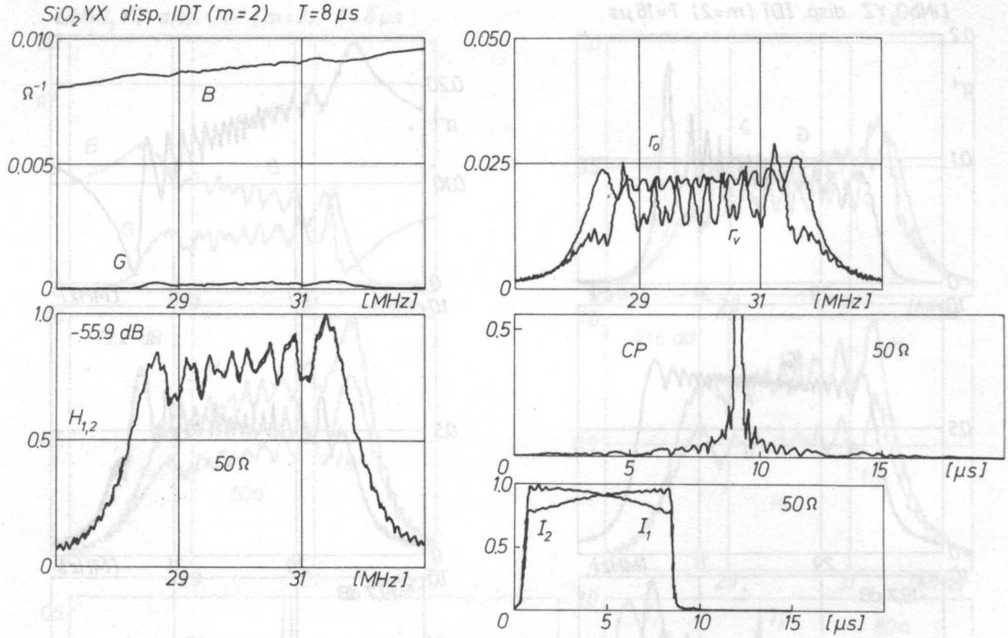


Fig. 5h

Calculations of the transient responses of dispersive filters demonstrate that all the fundamental factors that participate in the filter work such as the "element factor", the SAW propagation and reflections under the metal strips, may be effectively taken into account in contrast with other analysis methods, e.g. those mentioned in [12]. The calculations show a considerable effect of the transducer admittance upon the characteristics of dispersive filters, this effect being usually disregarded in the analysis because of computational difficulties.

References

[1] W. R. Smith, H. M. Gerard, J. H. Collins, T. M. Reeder, H. J. Shaw, *Analysis of interdigital surface wave transducers by use of equivalent circuit model*, IEEE Trans., **MIT−17**, 11, 856–864 (1969).

[2] W. S. Jones, C. S. Hartmann, T. D. Sturdivant, *Second order effects in surface wave devices*, IEEE Trans., **SU−19**, 3, 368–377 (1972).

[3] P. V. Wright, *Modelling and experimental measurement of the reflection properties of SAV metallic gratings*, Ultras. Symp. Proc., 1984, 54–63.

[4] S. Datta, B. J. Hunsinger, *Element factor of periodic interdigital transducer of surface acoustic waves*, J. Tech. Phys., **19**, 89–102 (1978).

[5] E. Danicki, *Theory of periodic interdigital transducer of surface acoustic waves*, J. Tech. Phys., **19**, 89–102 (1978).

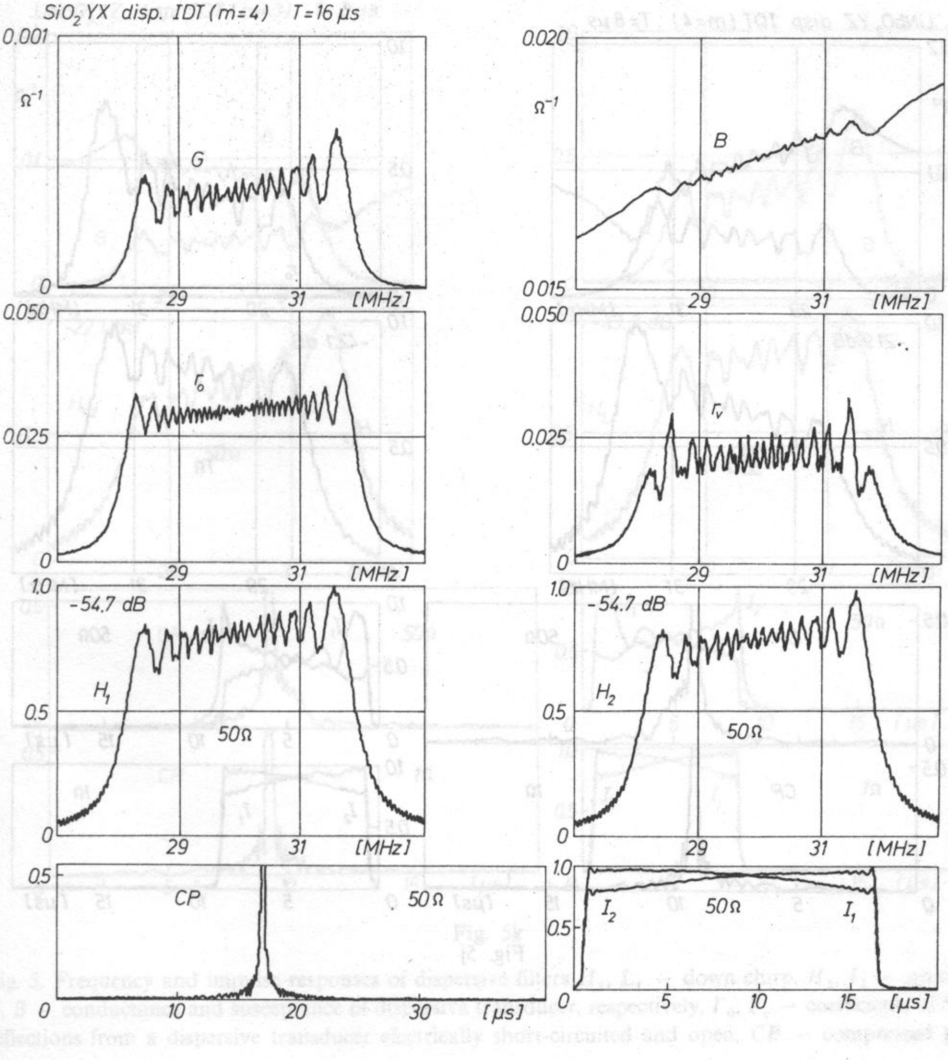


Fig. 5i

[6] E. Danicki, *New theory of SSBW*, Ultras. Symp. Proc., 1980 235–239.

[7] E. Danicki, *Unified theory of interdigital transducers and SAW reflectors*, J. Tech. Phys., **21**, 3, 387–403 (1980).

[8] T. Aoki, K. A. Ingebrigtsen, *Equivalent circuit parameters of interdigital transducers derived from dispersion relations of surface acoustic waves in periodic metal gratings*, IEEE Trans., **SU−24**, 3, 167–178 (1977).

[9] E.Danicki, *Propagation of transverse surface acoustic waves in rotated Y-cut quartz substrates under heavy periodic metal electrodes*, IEEE Trans., **SU−30**, 5 (1980).

[10] D. P. Chen, A. Haus, *Analysis of metal strip SAW gratings and transducers*, IEEE Trans., **SU−32**, 3, 395–408 (1985).

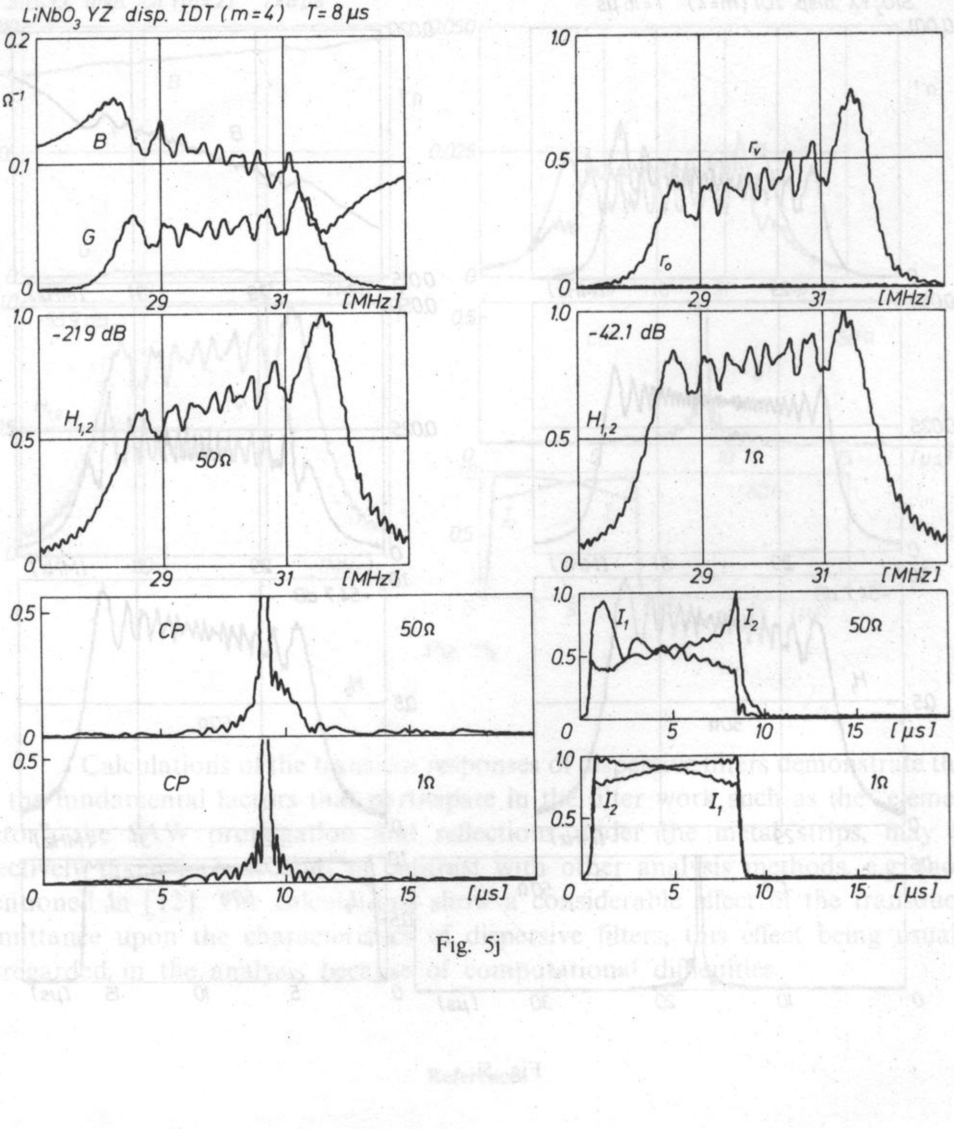


Fig. 5j

[11] E. Danicki, *Perturbation theory of surface acoustic wave reflections from a periodic structure with arbitrary angle of incidence*, Arch. Mech., **36**, 5–6, 623–638 (1984).

[12] B. Lewis, R. G. Arnold, *Electrode reflections, directionality, and passband ripple in wideband SAW chirp filter*, IEEE Trans., **SU – 32**, 3, 409–422 (1985).

[13] M. E. Field, R. C. Ho, C. L. Chen, *Surface acoustic wave grating reflector*, Ultras. Symp. Proc., 1975, 430–433.

[14] E. Danicki, J. Filipiak, *Equivalent network model of interdigital transducers of surface acoustic waves*, J. Tech. Phys., **19**, 2, 173–182 (1978).

[15] T. Kojima, J. Tominaga, *Split open metal strip arrays and their application to resonators*, Ultras. Symp. Proc. 1981, 117–122.

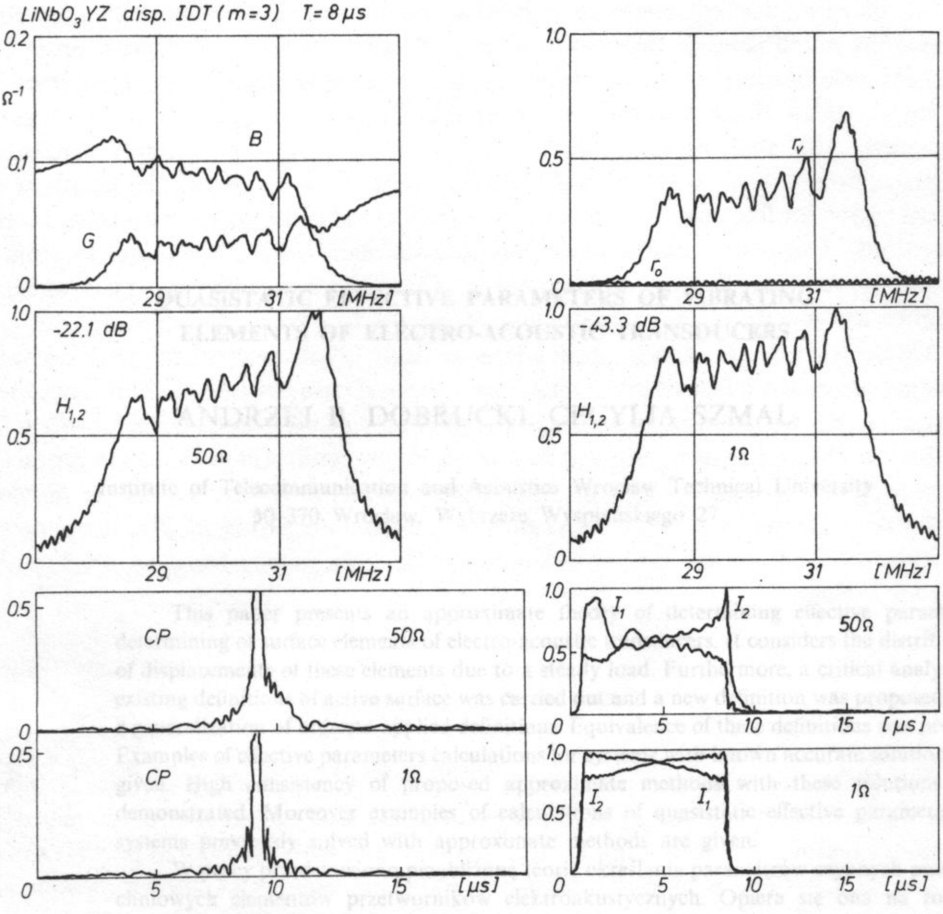


Fig. 5k

Fig. 5. Frequency and impulse responses of dispersive filters H_1, L_1 – down-chirp, H_2, I_2 – up-chirp, G, B – conductance and susceptance of dispersive transducer, respectively, Γ_0, Γ_v – coefficients of SAW reflections from a dispersive transducer electrically short-circuited and open, CP – compressed pulse

[16] E. Danicki, S. Bartnicki, *A method for SAW reflection coefficient determination*, Ultras. Symp. Proc., 1979, 663–666.

[17] J. Filipiak – paper unpublished yet.

[18] C. M. Panasiak, B. J. Hunsinger, *Scattering matrix analysis of surface acoustic wave reflectors and transducers*, IEEE Trans., **SU – 28**, 2, 79–81 (1981).

[19] E. Danicki, J. Filipiak, A. Kawalec, *SAW dispersive delay line utilizing apodised IDT with periodic electrodes*, Electr. Lett., **22**, 19, 976–977 (1986).

Received on August 19, 1988.

ARCHIVES OF ACOUSTICS
13, 3-4, 257-266 (1988)

QUASISTATIC EFFECTIVE PARAMETERS OF VIBRATING ELEMENTS OF ELECTRO-ACOUSTIC TRANSDUCERS

ANDRZEJ B. DOBRUCKI, CECYLIA SZMAL

Institute of Telecommunication and Acoustics Wrocław Technical University
50-370 Wrocław, Wybrzeże Wyspiańskiego 27

This paper presents an approximate theory of determining effective parameters determining of surface elements of electro-acoustic transducers. It considers the distribution of displacements of these elements due to a steady load. Furthermore, a critical analysis of existing definitions of active surface was carried out and a new definition was proposed. It is a generalization of hitherto applied definitions. Equivalence of these definitions was proved. Examples of effective parameters calculations for systems with known accurate solutions are given. High consistency of proposed approximate methods with these solutions was demonstrated. Moreover examples of calculations of quasistatic effective parameters of systems previously solved with approximate methods are given.

W pracy przedstawiono przybliżoną teorię określania parametrów czynnych powierzchniowych elementów przetworników elektroakustycznych. Opiera się ona na rozpatrywaniu rozkładu przemieszczeń tych elementów pod wpływem obciążenia statycznego. Ponadto poddano krytycznej analizie istniejące definicje powierzchni czynnej i zaproponowano nową definicję, będącą uogólnieniem dotychczas stosowanych. Wykazano ekwiwalentność tych definicji. Podano przykłady obliczania parametrów czynnych dla układów, dla których znane są rozwiązania ścisłe i wykazano dużą zgodność zaproponowanych metod przybliżonych z tymi rozwiązaniami.Podano również przykłady obliczeń quasistatycznych parametrów czynnych układów, dla których rozwiązania były również znalezione metodami przybliżonymi.

1. Introduction

A standard electro-acoustic transducer consists of an electromechanical transducer which transforms an electric signal into a mechanical one or a mechanical signal into an electric one; and a superficial element – generally a membrane which radiates acoustic waves when it is a loudspeaker or receives them when it is a microphone. Generally, the method of equivalent electric circuits [7], based on analogies of mechanic, electric and acoustic systems, is used in the analysis of loudspeaker functioning. These analogies can be applied to elements with geometric dimensions much smaller than wave length. This is not much of a problem in the case of the electric part of the transducer in the range of acoustic frequencies, because the length of an electromagnetic wave in this frequency range is equal to tens of kilometers. The analysis of mechanic and acoustic equivalent circuits can cause problems. Sound wave velocity in air is equal to about 340 m/s. So, the length of an acoustic wave for frequencies of several thousand Hz is comparable with transducer's dimensions. And even smaller velocities of mechanical waves are found in surface vibrating systems, such as plates and shells. Therefore, concentrated values of these elements are significant in a frequency range up to several hundred Hz.

Concentrated parameters equivalent to individual mechanical and acoustic elements are called effective parameters. In a general case they depend on frequency, because of the distribution variability of amplitudes of vibrations on the surface for various activation frequencies. Such a distribution depends not only on the geometry of the surface element, but also on the activation method. This makes the analysis additionally complicated. This paper proposes such a definition of effective parameters that their values are nearly independent from frequency for a relatively large frequency band.

2. Definitions of effective parameters

The following effective parameters characterize a vibrating surface element: mass, stiffness and surface. The notions of mass and stiffness are related with Rayleigh's method [4] consisting in checking a system with distributed parameters in relation to a system with one degree of freedom. Elements of such a system are calculated on the basis of a comparison of potential and kinetic energy of a real vibrating system with these energies of a system with one degree of freedom. The kinetic and potential energy of a real vibrating system depends on the distribution of displacements on the shell's surface which, in turn, depends on frequency and on the method of activation. In a general case the equation of motion of a homogeneous shell with — a model of a surface vibrating element of a transducer — can be expressed with [2]:

$$L(\mathbf{u}) - \varrho h \frac{\partial^2 \mathbf{u}}{\partial t^2} = \mathbf{P}[r, \varphi, z(r, \varphi), t], \tag{1}$$

where ϱ — density of material, h — shell thickness, L — differential operator dependent on shell's shape, $\mathbf{u}$ — vector of displacements of the shell's centre surface, $\mathbf{P}$ — vector of activation.

Fig. 1 presents the shell's geometry. Equation

$$z = z(r, \varphi) \tag{2}$$

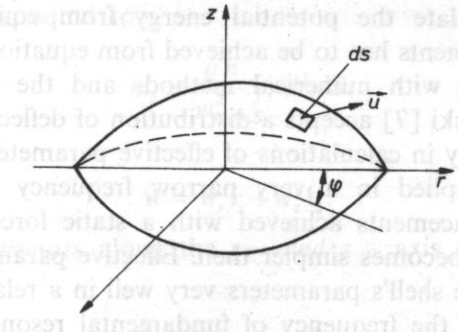


Fig. 1. Geometry of a vibrating shell

describes its shape in the case of axial symmetry. The form of operator L depends on various shapes of the shell.

In case of harmonic excitation

$$\mathbf{P}[r, \varphi, z(r, \varphi), t] = \mathbf{p}[r, \varphi, z(r, \varphi)]e^{j\omega t} \tag{3a}$$

displacement are also sinusoidal

$$\mathbf{u}[r, \varphi, z(r, \varphi), t] = \mathbf{w}[r, \varphi, z(r, \varphi)]e^{j\omega t} \tag{3b}$$

then equation (1) has the following form:

$$L(w) + \omega^2 \varrho h \mathbf{w} = \mathbf{p}[r, \varphi, z(r, \varphi)]. \tag{4}$$

When $\mathbf{p}(r, \varphi, z) = 0$, then the equation of motion is reduced to a eigen problem. Equations (1) and (4) are equations of equilibrium of forces. The first term on the left side describes elastic forces acting on the shell. They are a result of own elasticity. These forces cause a displacement of the shell's elements in relation to the state of equilibrium and impart a certain potential energy to the shell

$$U = \frac{1}{2}\int_s L(\mathbf{w}) \cdot \mathbf{w} dS. \tag{5}$$

Effective stiffness is achieved on the basis of comparison between this energy and the potential energy of a system with one degree of freedom

$$U = \frac{1}{2} k_c w_{\max}^2, \tag{6}$$

where k_c – effective stiffness, $w_{\max}$ – a certain, discriminated displacement of the shell, most frequently equal to the maximum axial deflection, to the displacement of its geometrical centre in the case of activation by a uniformly distributed acoustic pressure, or a displacement at the point of activation by an axial concentrated force.

In order to calculate the potential energy from equation (5), the shell's distribution of displacements has to be achieved from equation (4). This equation is most frequently solved with numerical methods and the solution is frequency dependent. Z. Żyszkowski [7] accepts a distribution of deflections achieved for the first resonance frequency in calculations of effective parameters. Unfortunately this assumption can be applied in a very narrow frequency range. In this paper a distribution of displacements achieved with a static force, i.e. for $\omega = 0$, was accepted. Equation (3) becomes simpler then. Effective parameters calculated from this equation model the shell's parameters very well in a relatively wide frequency range – from zero to the frequency of fundamental resonance.

Neglecting the dynamic term in equation (4) we obtain:

$$L(\mathbf{w}) = \mathbf{p}[r, \varphi, z(r, \varphi)]. \tag{7}$$

Then, equation (5) can be noted as:

$$U = \frac{1}{2}\int_S \mathbf{p}[r, \varphi, z(r, \varphi)]\,\mathbf{w}\,dS. \tag{8}$$

We have to do with an interesting case when the shell is activated by a force concentrated in the medium and acting along the z-axis

$$\mathbf{p}(r, \varphi, z) = P_z \delta[r, \varphi, z(r, \varphi)], \tag{9}$$

where $\delta[r, \varphi, z(r, \varphi)]$ is the δ–Dirac function. Then the potential energy is equal to

$$U = \frac{1}{2} P_z w_{\max}. \tag{10}$$

From comparison of equations (6) and (10) an expression for effective rigidity we see that

$$k_c = \frac{P_z}{w_{\max}}. \tag{11}$$

The second term in equation (4) defines the inertial force related with the kinetic energy of shell's motion.

Kinetic energy of the shell is given by expression

$$T = \frac{1}{2}\varrho h \omega^2 \int_S |\mathbf{w}|^2 dS \tag{12}$$

where

$$\mathbf{w}(r, \varphi)|_\omega \approx \mathbf{w}(r, \varphi)|_{\omega=0}.$$

The effective mass is achieved from a comparison of the kinetic energy determined from expression (12) and the kinetic energy of a system with one degree of freedom

$$T = \frac{1}{2} m_c \omega^2 w_{\max}^2. \tag{13}$$

Thus, effective mass is equal to

$$m_c = \frac{\varrho\, h}{w_{\max}^2} \int_S |\bar{w}|^2\, dS. \tag{14}$$

Because:

$$\mathbf{w} = w_r\, \mathbf{i}_r + w_z\, \mathbf{i}_z, \tag{15}$$

where $\mathbf{i}_r$, $\mathbf{i}_z$ are unit vectors along the $r-$ *and* $z-$ axis, respectively. Then, we have

$$m_c = \frac{\varrho h}{w_{\max}^2} \int_S w_r^2\, dS + \frac{\varrho h}{w_{\max}^2} \int_S w_z^2\, dS. \tag{16}$$

The first term in equation (16) determines the effective mass related to the displacements of the shell in the radial direction, and the second with displacements in the axial direction.

The effective surface is the last among effective parameters which characterize surface systems. Definitions of this quantity given in [5, 7] are not general definitions. According to Żyszkowski the definition refers to plane systems only, while the definition of Makarewicz and Konieczny concerns a specific method of activating the shell, namely – with a plane wave. Therefore, it can not be used for a different activation method i.e. with a spherical wave or by force concentrated in a point. A definition without these shortcomings is presented below.

Effective surface is the surface of a flat piston shifted perpendicularily to its plane by a force of the value equal to the value of a characteristic virtual displacement of the shell due to a uniformly distributed pressure of such a value that performed by it work is equal to the work of forces acting on the shell.

This definition is based on the principle of virtual work [4] and introduces the notion of pressure equivalent to forces really acting on the shell. This pressure does not have to be acoustic pressure; it can also be static pressure.

When the distribution of static displacements achieved from equation (7) is accepted, then a quasistatic effective surface is obtained. Virtual work performed in order to virtually displace the shell by $\delta \mathbf{w}$ is equal to

$$\delta W = \int_S \mathbf{p}[r,\, \varphi,\, z(r,\, \varphi)]\, \delta \mathbf{w} dS. \tag{17}$$

Substituting

$$\delta \mathbf{w} = \delta w_{\max} \frac{\mathbf{w}}{w_{\max}}, \tag{18}$$

we have

$$\delta w = \frac{\delta w_{\max}}{w_{\max}} \int_S \mathbf{p}[r,\, \varphi,\, z(r,\, \varphi)]\, \mathbf{w} dS. \tag{19}$$

The work performed by an uniformly distributed pressure resulting in the same displacement of the shell is equal to

$$\delta W = p\int_S \delta w_n dS = p\frac{\delta w_{max}}{w_{max}}\int_S w_n dS \tag{20}$$

where w_n is the displacement normal to the surface S. We have normal displacement in formula (20), because the force acting on the shell due to an uniformly distributed pressure is perpendicular to the shell's surface. On the basis of a comparison between relationships (19) and (20) we reach an expression for pressure corresponding to force actually acting on the shelll:

$$p = \frac{\int_S \mathbf{p}[r, \varphi, z(r, \varphi)]\, w dS}{\int_S w_n dS}. \tag{21}$$

If such a pressure is exerted on a flat piston with surface equal to the effective surface S_c and shifts it by δw_{max} then work equal to

$$\delta w = pS_c\delta w_{max} \tag{22}$$

is performed.

On the basis of relationships (20), (21) and (22) the effective surface is equal to

$$S_c = \frac{1}{w_{max}}\int_S w_n dS. \tag{23}$$

The following new definition of the effective surface is thus equivalent to the definition given above:

The effective surface is such a surface of a flat piston which ensures a volumetric displacement of the piston equal to the volumetric displacement of the shell.

This definition does not include a direct dependence on the shell activation method. Therefore, it is of more general character than the definition based on the principle of virtual work. It is a well-known fact [6] that the power radiated by the source in the range of low frequencies depends on the volumetric velocity of the source and, on the other hand, it does not depend on its shape. The definition of the effective surface based on the equivalence of volumetric deflections leads to the substitution of a real source by a flat piston which radiated the same power as the source under consideration in an infinitely great baffle. The normal displacement in the equation for effective surface (23) can be expressed with two components − axial and radial:

$$S_c = \frac{1}{w_{max}}\int_S w_n dS = \frac{1}{w_{max}}\Big[\int_S w_z \cos(n, z) dS + \int_S w_r \cos(n, r) dS\Big] \tag{24}$$

$$= \frac{1}{w_{max}}\Big[\int_0^{r(\varphi)}\int_0^{2\pi} w_z r dr d\varphi + \int_0^{r(\varphi)}\int_0^{2\pi} w_r \frac{dz}{dr} r dr d\varphi\Big].$$

In a case of an axial activation of the shell the second component of the sum in expression (24) is very small and can be neglected. If the shell's motion is purely axial, then it is equal to zero exactly. If additionally $w_z(r) = \text{const}$, i.e. the shell vibrates like a rigid piston, then we reach a well known result – the effective surface is equal to the surface of the shell's orthogonal projection onto the plane perpendicular to the z – axis. The second term of the sum appearing in Eq. (24) is decisive in the case of radial vibrations of a cylindrical shell. The Żyszkowski and Makarewicz definitions did not take into account the possibility of evaluation of the active surface of such a system.

3. Effective parameters of circular plates

The equation of the circular plate has been solved analytically. Therefore, strictly theoretical results can be compared with results of the theory based on quasistatic displacements of the shell.

Frequently the circular plate is the vibrating element in electroacoustic transducers. It is a planar system.

The equation for vibrations of a plate has the following form

$$\Delta \Delta u + \frac{\varrho h}{B}\frac{\partial^2 u}{\partial t^2} = \frac{P(r, \varphi, t)}{B}, \tag{25}$$

where $u = we^{j\omega t}$ – displacement of the plate, h – plate's thickness, $B = \dfrac{Eh^3}{12(1-\nu^2)}$ – flexural rigidity, ϱ – density of material, E – Young's modulus, ν – Poisson ratio, $P(r, \varphi, t) = p(r, \varphi)e^{j\omega t}$ – force acting on an unit surface, Δ – Laplace operator.

We will consider a case of $p(r, \varphi) = p = \text{const}$. The system are axi-symmetrical. Operator L from equation (1) has the following form

$$L(u) = -\frac{B}{r}\frac{d}{dr}\left\{r\frac{d}{dr}\left[\frac{1}{r}\frac{d}{dr}\left(r\frac{du}{dr}\right)\right]\right\}. \tag{26}$$

The equation of quasistatic displacements of the plate – equivalent to equation (7) has the following form:

$$\frac{1}{r}\frac{d}{dr}\left\{r\frac{d}{dr}\left[\frac{1}{r}\frac{d}{dr}\left(r\frac{dw}{dr}\right)\right]\right\} = \frac{P}{B}. \tag{27}$$

The solution of this equation depends on boundary conditions. Two typical boundary conditions will be considered:

1. Plate's edge simple-supported

$$w(a) = 0,$$
$$\frac{d^2 w}{dr^2} + \frac{\nu}{r}\frac{dw}{dr} = 0 \quad \text{for } r = a \tag{28}$$

where a — plate's radius.

This equation has the following solution

$$w = w_{\max}\left[1 - \left(\frac{r}{a}\right)^2\right]\left[1 - \frac{1+\nu}{5+\nu}\left(\frac{r}{a}\right)^2\right], \tag{29}$$

where

$$w_{\max} = \frac{pa^4}{64B}\frac{5+\nu}{1+\nu}. \tag{30}$$

The potential energy equivalent to this displacement is equal to:

$$U = \frac{1}{2}\int_S wp\,dS = \frac{1}{2}w_{\max}^2 \frac{64\pi B}{a^2}\frac{(1+\nu)(7+\nu)}{3(5+\nu)^2}. \tag{31}$$

Thus, the effective stiffness corresponding to this potential energy equals:

$$k_c = \frac{64\pi B}{3a^2}\frac{(1+\nu)(7+\nu)}{(5+\nu)^2}. \tag{32}$$

The following equation expresses the kinetic energy corresponding to the quasistatic distribution of velocity on the plate's surface

$$T = \frac{1}{2}\varrho h\omega^2 w_{\max}^2 \int_S \left[1 - \left(\frac{r}{a}\right)^2\right]^2\left[1 - \frac{1+\nu}{5+\nu}\left(\frac{r}{a}\right)^2\right]^2 dS = \frac{1}{2}M_g v_{\max}^2 \frac{(113+36\nu+3\nu^2)}{15(5+\nu)^2}, \tag{33}$$

where $v_{\max} = \omega w_{\max}$ — plate's velocity in the geometrical centre, $M_g = \varrho h\pi a^2$ — geometrical mass of the plate. The effective mass equivalent to this energy is equal to:

$$m_c = \frac{3\nu^2 + 36\nu + 113}{15(5+\nu)^2} M_g. \tag{34}$$

The resonance frequency calculated on the basis of effective parameters is equal to

$$f_{\text{rez}} = \frac{1}{2\pi}\sqrt{\frac{k_c}{m_c}} = \frac{1}{2\pi a^2}\sqrt{\frac{B}{\varrho h}\frac{320(1+\nu)(7+\nu)}{3\nu^2 + 36\nu + 113}}. \tag{35}$$

The accurate solution, corresponding to the dynamic equation of plate's vibrations, contains ordinary and modified Bessel functions [1]. For $\nu = 0.3$ the fundamental frequency of plate's vibrations, calculated from accurate considerations, is expressed by:

$$f_{01} = \frac{1}{2\pi} \frac{(k_{01}\, a)^2}{a^2} \sqrt{\frac{B}{\varrho h}}, \tag{36}$$

where $k_{01}\ a = 2.221$ is the first eigenvalue.

The value corresponding to it in formula (35) is equal to 2.224. Hence, the relative error is equal to 0.15%. It is worth mentioning that active parameters depend on the Poisson ratio which occurs explicitly in boundary conditions. Therefore, the Poisson ratio can be measured indirectly on this basis.

And lastly the effective surface of a supported plate was calculated

$$S_c = \int\limits_S w dS = \pi a^2 \frac{7+\nu}{3(5+\nu)}. \tag{37}$$

For $\nu = 0.3$ its value is equal to 0.459 of the geometrical surface.

2. Plate's edge clamped

In this case boundary conditions have the following form:

$$w(a) = 0, \quad \left.\frac{dw}{dr}\right|_{r=a} = 0. \tag{38}$$

The solution of equation (27) is as follows

$$w(r) = w_{\max}\left[1 - \left(\frac{r}{a}\right)^2\right]^2, \tag{39}$$

where

$$w_{\max} = \frac{pa^4}{64B}. \tag{40}$$

The effective stiffness of this plate equals

$$k_c = \frac{64}{3} \frac{\pi B}{a^2}, \tag{41}$$

whereas, effective mass:

$$m_c = \frac{1}{5} M_g. \tag{42}$$

The resonance frequency, calculated on the basis of these parameters, is equal to

$$f_r = \frac{3.21^2}{2\pi a^2} \sqrt{\frac{B}{\varrho h}}. \tag{43}$$

In this case, the coefficient corresponding to the first eigen value is equal to 3.21, while its value determined from the accurate solution is equal to 3.19. This coefficient does not depend on the Poisson ratio. The Poisson ratio is not present in the boundary conditions.
Therefore, the effective surface:

$$S_c = \frac{1}{3}\pi a^2 \tag{44}$$

is smaller than the analogic surface of a simple-supported plate.

4. Effective parameters of loudspeaker suspension of the voice-coil

Calculation results of effective parameters of non-planar systems will be presented in this paragraph. Apart from simple cases, the form of the Loperator for such systems remains unknown. The distribution of displacements for a quasistatic activation was determined with an approximate method of finite elements [3]. In order to study the influence of chosen geometrical features on active parameters the shape of typical loudspeaker suspension of the voice-coil and of certain model shells was considered in calculations. An unit-value concentrated force, acting on the internal rim of the shell under consideration, was accepted as the activating factor. Fig. 2 presents the shape of the model.

The following standard data were included in calculations:
– Young's modulus $E = 10^9$ N/m^2
– Poisson ratio $\nu = 0.3$
– density of material $\varrho = 1000$ kg/m^3
– the external rigidity of the suspension of the voice-coil is rigidly clamped while the internal rim can move freely along the z–axis only.

The following quantities varied:
– thickness of the suspension of the voice-coil h
– wave length DF
– number of waves n
– wave height A
– internal – R_w and external – R_z radii, at fixed cross-section shape.

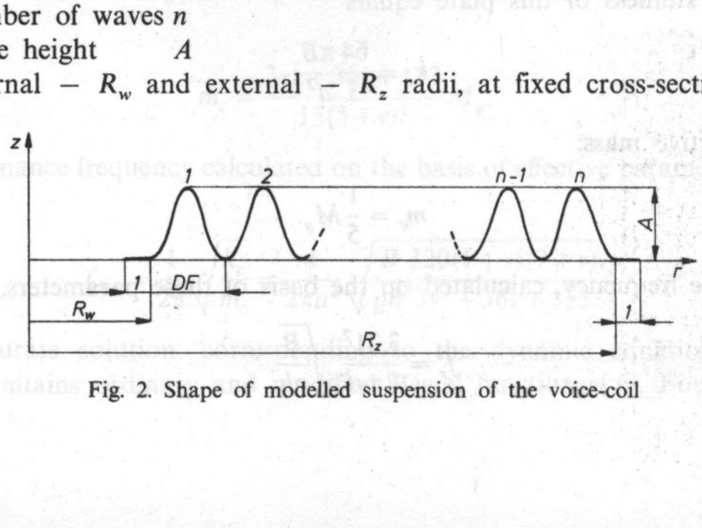


Fig. 2. Shape of modelled suspension of the voice-coil

Table 1. Effective parameters of suspension of the voice-coil

Geometric parameters of the suspension of the voice-coil						Effective parameters of the suspension of the voice-coil				
Lp.	h [mm]	A [mm]	n	DF [mm]	R_w [mm]	R_z [mm]	k_c [N/m]	m_c [g]	m_c/M_g	S_{c_2} [cm²]	$S_c/\pi(R_z^2-R_w^2)$
1	0.35	2	7	5	24	59	1289	1.262	0.2893	41.61	0.4235
2	0.25	„	„	„	„	„	832	0.929	0.2981	42.77	0.4353
3	0.45	„	„	„	„	„	1848	1.584	0.2823	40.72	0.4144
4	0.35	1	„	„	„	„	481	0.978	0.2680	38.19	0.3969
5	„	3	„	„	„	„	2951	1.621	0.3032	45.64	0.4423
6	„	2	5	„	„	49	2219	0.835	0.2996	27.13	0.4304
7	„	„	9	„	„	69	862	1.762	0.2830	58.67	0.4194
8	„	„	7	4	„	52	1923	1.060	0.2968	32.05	0.4300
9	„	„	„	6	„	66	936	1.496	0.2834	52.46	0.4185
10	„	„	„	5	12	47	1669	0.853	0.2750	28.75	0.4116

Calculation results are presented in Table 1. Effective parameters related to shell's displacements along the radius are not given. They are very small in comparison to adequate effective parameters related with shell's displacements along the axis. For example, in the case of suspension of the voice-coil no 1 the value of the ratio of effective mass related to radial displacement of the suspension of the voice-coil and the geometrical mass is equal to $4.258\cdot 10^{-4}$. Whereas, the adequate surface ratio is equal to $1.093\cdot 10^{-3}$. A comparison between these data and results presented in Table 1 proves the error due to the neglection of these parameters to be insignificant.

5. Conclusions

A comparison between results of accurate and quasistatic solution proves that the difference between them does not exceed 1%.

Characteristic features of the accurate solution, such as Poisson ratio-dependence or independence of parameters for example, can also be found in the approximate solution.

Values of effective parameters calculated for a non-planar system, such as a loudspeaker suspension of the voice-coil indicate a strong influence of geometrical features on the value of parameters.

To the suspension of the voice-coil designer the suspension voice-coil stiffness is the most interesting feature. It depends on thickness and increases with an increase of thickness — more than h^1 (pure tension) but less than h^3 (pure bending). The index is not constant and depends on thickness. The average value for considered suspension of the voice-coil was equal to 1.35. Hence, material work conditions are closer to tension.

Stiffness increases with the height of the suspension of the voice-coil wave. A plate has the lowest stiffness. Yet it is not applied in loudspeaker suspension of the voice-coil, because of high nonlinearity. Stiffness decreases with an increase of the suspension of the voice-coil wave length and number of waves *n*.

Also other effective parameters, such as mass and surface, depend on geometrical parameters, but the ratio between them and geometrical mass and surface are nearly independent from shape. Changes of absolute values are related with variable amounts of material used to produce the suspension of the voice-coil and with variable geometrical surface of various suspension of the voice-coil structures.

References

[1] A. Dobrucki, *Foundations of acoustics* (in Polish), Wydawnictwo Politechniki Wrocławskiej, 1987.

[2] A. Dobrucki, C. Szmal, *Calculations of active parameters of loudspeaker membranes and suspensions* (in Polish), Proc. XXXIII OSA, Rzeszów 1986, 47–50.

[3] A.Dobrucki et al., *Computer simulation and investigations of loudspeaker suspension and suspension of the voice-coil* (in Polish), ITA Politechniki Wrocławskiej, Raport I-28/SPR-019/86.

[4] S. Kaliski [ed], *Vibrations and waves in solids* (in Polish), PWN, Warszawa 1966.

[5] R. Makarewicz, J. Konieczny, *Active surface and mass of microphone membranes* (in Polish), Archiwum Akustyki **12**, 4, 313–326 (1977).

[6] С. Н. Ржевкин, *Курс лекций по теории звука*, Изд. Московского Университета, Москва 1960.

[7] Z. Żyszkowski, *Foundations of electroacoustics* (in Polish), WNT, Warszawa, 1966.

Received on December 2, 1986.

ARCHIVES OF ACOUSTICS
13, 3–4, 269–278 (1988)

PERCEPTION ASPECTS OF A RULE SYSTEM FOR CONVERTING MELODIES FROM MUSICAL NOTATION INTO SOUND

L. FRYDÉN

Conservatory of the Swedish Radio, Edsberg, Sweden

J. SUNDBERG, A. ASKENFELT

Department of Speech Communication and Music Acoustics Royal Inst. of Technology
(S–10044 Stockholm 70, Sweden)

The starting point of this project is the "mechanical" impression obtained when a computer rather than a good musician converts the musical notation into the corresponding sound sequences. A basic assumptions is that this mechanical effect can be partly eliminated by introducing "pronunciation rules", which introduce minute, context dependent deviations from the durations, pitches an amplitudes specified in the musical score. The method applied is analysis by synthesis. A computer program developed for the conversion of text to speech (Carlson and Granström [3]) is applied to the singing synthesizer MUSSE (Larsson [4]). In the present, musical version of the conversion program, the input is the melody in musical notation, and the output is the melody, synthetically performed. A set of such pronunciation rules have been formulated for, and tested on traditional Western melodies. One group of rules operates with short time windows of two or three notes. Thus they are applied using as criteria the size of the musical interval formed by adjacent notes, or the difference in duration between two or three adjacent notes. Another type of rules operates with a time window of variable length; the window is limited by the distance between adjacent chord changes. The rules manipulate the duration and amplitude of notes. In other words all these rules introduce discrepancies between what is written in the notation and what is actually played. The effects of these rules are evaluated by means of listening tests with musically sophisticated judges. The results show that the musical acceptability of such synthesized performances can be substantially improved by applying these "pronunciation" rules. Furthermore, our experiences indicate that the magnitude of these discrepancies between what is written and what is supposed to be played is very critical. If the magnitude is so high, that the effect is identified for what it actually is, then the effect generally tends to be musically impossible. If, on the other hand, the magnitude is too small, there is, of course, no effect at all. It seems that the musically useful effects are found in between this threshold of diagnosis and the threshold of perception. The perceptual implications of these rules regarding musical communication will be discussed, and the effects of the rules may be demonstrated by means of tape illustrations.

Praca przedstawia propozycję rozwiązania problemu „sztuczności" i „martwoty" charakteryzujących muzykę produkowaną w sposób syntetyczny przez komputery. Autorzy opierają się na swym długoletnim doświadczeniu w analizowaniu muzyki „żywej" i obserwacji wszelkich odchyleń jej parametrów w zakresie od matematycznie regulowanych równomiernych skal. Proponują wprowadzenie całego szeregu reguł rozbijających w określony sposób równomierność skal w zakresie czasu, natężenia i częstości. Podano opis eksperymentu psychoakustycznego, w którym grupa doświadczonych słuchaczy porównywała i oceniała kompozycje wyprodukowane syntetycznie z użyciem i bez użycia wspomnianych reguł. W większości przypadków wprowadzenie odchyleń od równomierności nadało muzyce bardziej żywy charakter i zostało pozytywnie ocenione przez słuchaczy. Dokonano analizy uzyskanych wyników i wyciągnięto wnioski co do dalszych ulepszeń proponowanego systemu.

1. Introduction

When music is being played, a string of note signs is converted into a sequence of tones. If this conversion is realized in a one-to-one fashion, e.g. by a computer, the result is musically a disaster. This fact raises a question what are the exact relationships between note signs and their corresponding acoustic signals in musical performances? Or, in other words, what do musicians do when they play, and why do they do that?

Previous studies of musical performance have revealed an almost overwhelming complexity (see e.g. BENGTSSON and GABRIELSSON [2]). SEASHORE (1938) measured the actual durations of the tones in several musical performances of the same song. Among other things he found that in none of the singers was there "a slightest approach to an even time for a measure". Also, he found that musicians tend to show a modest degree of reproducibility as regards the performance of a given piece of music.

These findings would seem trivial to a musician. He knows perfectly well that a piece of music can be played in a number of different ways which are all musically acceptable. And yet, the field of liberty for musicians is by no means unlimited; there is a large class of performances which are musically inacceptable. This suggests the existence of *musical performance rules* stating what is musically permitted and what is not in the conversion of notes to tones. If a player violates these rules, he runs the risk of being classified as a poor musician.

These performance rules are scientifically interesting. First of all, they separate music and non-music. Second, the rules would bear witness of auditory, perceptual and cognitive functions involved in the ears' and the brain's processing of the acoustic signal in music listening. Thus the rules may reveal some of the basic requirements for enjoying music.

2. Method

If we want to explore the rules which a musician must normally obey in playing a piece of music we can choose measurements or synthesis-by-rule. In the present project, the latter strategy was chosen. The reason for this is our assumption that the rules are numerous and that they interfere with each other. Also, the pedagogical and artistic experience of one of the authors (LF) had generated a number of hypotheses regarding such rules.

Like in a previous study of musical performance focussing on singing (SUNDBERG [7]) a computer-controlled vowel synthesizer was used (MUSSE, LARSSON [4]) which can generate one part only. In the present experiment the signal characteristics were adjusted so as to be similar to a wind instrument timbre. The pitch frequency changed in steps in accordance with the equally tempered scale, and there was a very slight vibrato only. The amplitude changed in steps of 1/4 dB. The duration was controlled in steps of a time unit corresponding to 0.8 to 1.2 csec depending on the tempo; according to VAN NOORDEN [5] this is accurate enough.

The computer programs used for controlling the synthesizer were 1) a notation program (Askenfelt [1]) by means of which the melody can be written in ordinary music notation on the computer screen; and 2) a text-to-speech program written by CARLSON and GRANSTRÖM [3]. The information encoded in the notation is translated into "vowel sounds" of corresponding duration, pitch frequency and amplitude. The rules are triggered by specific sequences of durations and/or pitches, and they manipulate amplitudes and durations.

3. Rules

Up to now we have formulated a dozen rules and tested seven of them in a listening experiment. We will now present these rules, one by one.

(1) *Pitch and amplitude*. In almost every music instrument, the amplitude increases somewhat with the fundamental frequency, as is illustrated schematically in Fig. 1. Our rule increases the amplitude as function of frequency by 4 dB/octave.

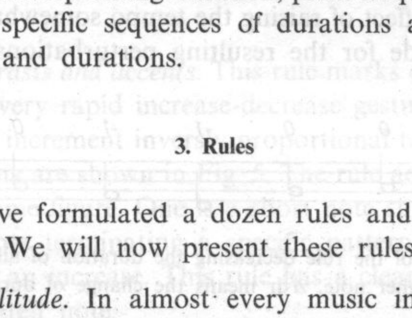


Fig. 1. Schematic illustration of the rule increasing the amplitude of the note as function of its fundamental frequency. The quantity is 4 dB of amplitude increase per one octave rise in fundamental frequency

(2) *Note value and duration.* This rule sharpens the duration contrasts between note values. All whole-notes and half-notes are played with their nominal durations while all quarternotes and eightnotes are played 1 cs too short, and every second sixteenthnote is played 1 cs too short. Thus long notes are played in their full length, and short notes are shortened a percentage which is inversely proportional to the nominal duration, (see Table 1). No compensation is made for the resulting perturbation of the mechanical meter.

Table 1. Shortening of the durations for different note values

Note value	Shortening csec	%
Sixteenth	−0.5	−3.1
Eight	−1.0	−3.1
Quarter	−1.0	−1.6
Half	0	0
Whole	0	0

(3) *Pitch increase and duration.* This rule simply states that the duration of each note is decreased by 1 csec the following note has higher pitch. In sequences of rising pitches the rule has the effect of raising the tempo somewhat, (see Fig. 2). As before, no compensation is made for the resulting perturbations of the meter.

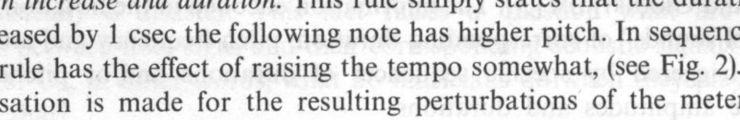


Fig. 2. Schematic illustration of the rule decreasing the duration of all notes which are followed by a higher note. Δ dr means the change of duration

(4) *Leaps and duration.* Our next rule increases the duration of tones terminating a melodic leap which is not followed by another leap of identical direction. The quantity was adjusted to $0.6 * 1^{1/2}$ (cs), where 1 is the size of the interval in semitones, (see Fig. 3).

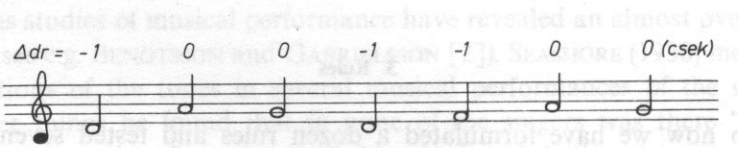


Fig. 3. Schematic illustration of the rule increasing the duration of all notes which terminate a melodic leap. 1 is the size of the leap interval in semitones

(5) *Leaps and pauses.* In instrumental music, particularly when played on bowed instruments, wide melodic leaps are often performed with a micropause during the pitch change. This is the effect of a rule, which decreases the amplitude of the final portion of the tone at a constant rate. The decrease starts at 0.5∗I cs from the end of the note, where I is the interval size in semitones. This rule has a negligible effect on narrow intervals as major and minor second but is quite noticeable in wide intervals, (see Fig. 4).

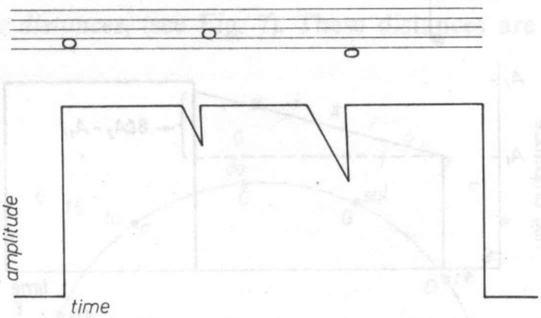


Fig. 4. Schematic illustration of the rule which inserts amplitude decreases towards the end of a note. The decrease has a constant rate. The onset time of the decrease is determined by the interval between the note and the following note

(6) *Note value contrasts and accents.* This rule marks contrasts in note values by accents, i.e. small and very rapid increase-decrease gestures in the amplitude. The rule adds an amplitude increment inversly proportional to the duration of the tone. The details of the accents are shown in Fig. 5. The rule adds an accent in two cases, as exemplified in the same figure. One is a short note surrounded by longer notes. The other case is a note terminating a specific pattern of changes in durations: a decrease followed by an increase. This rule has a clear effect particularly on the short notes after a dotted note.

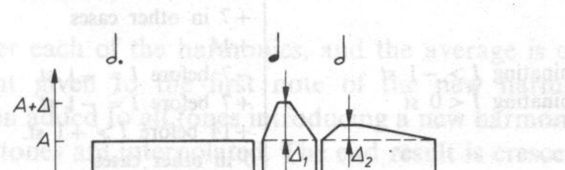


Fig. 5. Schematic illustration of the accents marking certain contrasts in note value

(7) *Amplitude continuity.* If the amplitude of tones changes stepwise, the melodic continuity may be disturbed. Our next rule states that the last amplitude reading of a tone should be corrected by a constant corresponding to 80% (in log. terms) of the amplitude difference between consecutive notes, (see Fig. 6).

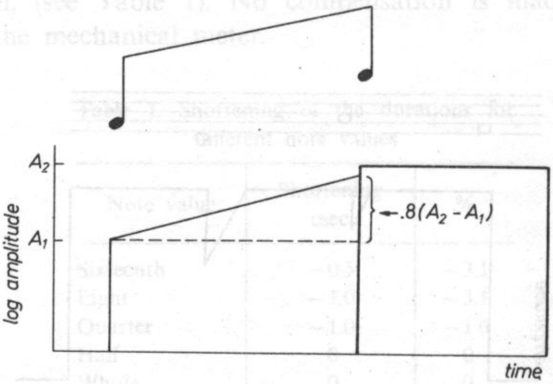


Fig. 6. Schematic illustration of the rule decreasing the amplitude contrasts between adjacent notes

(8) *Scale tone tuning.* The equally tempered scale is rarely used in music practice (SUNDBERG [9]). This rule modifies the scale tone frequencies by multiples of 7 cents which is our incremental step and which is close to one or two DL: s of pitch in a normal listener (RAKOWSKI [6]). Table 2 shows the values used.

Table 2. Deviations in cents from the equally tempered scale. I = interval, st = semitone

Scale tone in semitones above the root of the tonic	Deviation (cent)
1	-14 before $I = -1$ st $+7$ in other cases
2	$+14$
3 terminating $I > -1$ *st* terminating $I < 0$ *st*	-7 before $I = -1$ st. $+7$ before $I = -1$ st $+14$ before $I > +1$ st. 0 in other cases
4	$+14$
5	$+7$
6	$+14$
7	$+7$
8	-7 before $I = -1$ st. $+14$ in other cases
9	$+7$
10	-7
11	$+14$

The rules presented so far operate on a narrow time window of two or three notes only. The following rules use a wider time perspective, namely changes of harmony. The first note appearing over a new harmony is marked in the score. The following rules use such marks cues.

(9) *Chord changes and amplitude.* This rule generates crescendos and decrescendos. The technicalities are as follows. The distance along the circle of fifths from the root of the tonic is first determined for each note. Then, this distance is multiplied by 1.5 for scale tones located on the subdominant side of the circle. We have called the resulting values *tonic distances*, (see Fig. 7). These distances are averaged over all

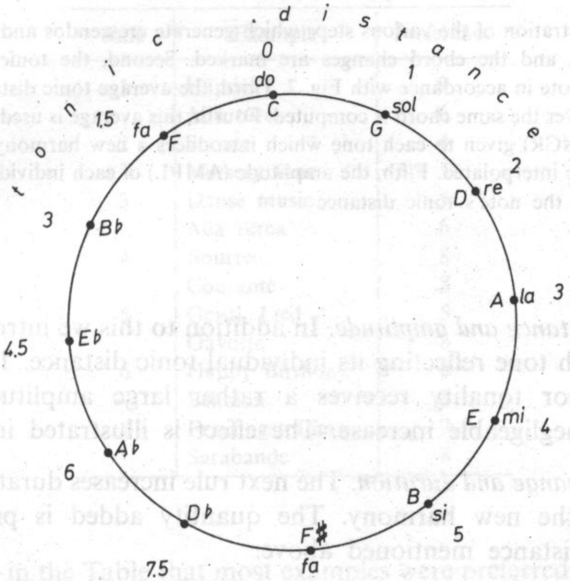


Fig. 7. Definition of tonic distance by means of the circle of fifths

tones appearing over each of the harmonies, and the average is converted into an amplitude increment given to the first note of the new harmony. When such increments have been added to all tones introducing a new harmony, the amplitudes of the intermediate tones are interpolated. The end result is crescendos culminating at harmonic changes introducing a chord which is more remote from the tonic than the preceding chord and decrescendos ending at chords which are less remote from the tonic than the preceding chord. Fig. 8 illustrates how the rule treats a typical cadence. In most cases the rule generates crescendos culminating at chords with the function of a dominant and decrescendos to their following "rest chords". As such sequences of dominant chord — rest chord are used to mark — among other things — the termination of phrases and subphrases in simple tunes, this rule often has the effect of marking the melody structure (cf SUNDBERG and LINDBLOM [8]).

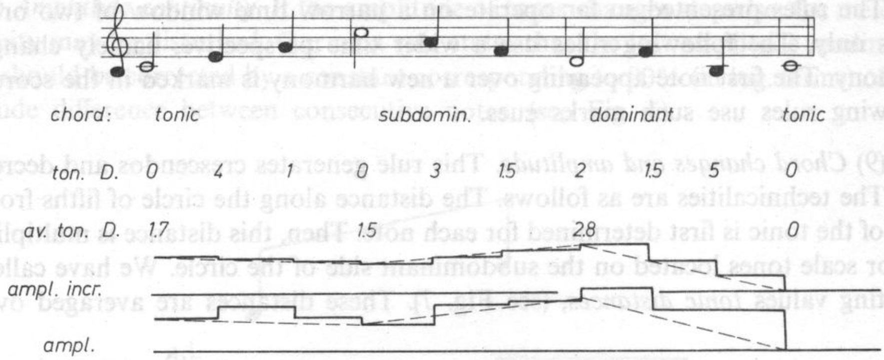


Fig. 8. Schematic illustration of the various steps which generate crescendos and diminuendos. First, the chords are identified, and the chord changes are marked. Second, the tonic distance (TON. D) is determined for each note in accordance with Fig. 7. Third, the average tonic distance (AVTON. D.) over the notes appearing over the same chord is computed. Fourth, this average is used to decide the amplitude increment (AMPL. INCR) given to each tone which introduces a new harmony: the amplitudes of the intermediate tones are interpolated. Fifth, the amplitude (AMPL) of each individual note is increased by an amount reflecting the note's tonic distance

(10) *Tonic distance and amplitude.* In addition to this we introduce an amplitude increment on each tone reflecting its individual tonic distance. This means that e.g. *F #* in *C* major tonality receives a rather large amplitude increase, while a G receives a negligeable increase. The effect is illustrated in the same Fig. 8.

(11) *Chord change and duration.* The next rule increases duration of the first note appearing over the new harmony. The quantity added is proportional to the averaged tonic distance mentioned above.

(12) *Phrase endings and duration.* This rule applies to and therefore mark phrase endings. As yet, we have no algorithm, detecting phrase boundaries automatically, so we mark the boundaries between phrases and subphrases in the input score. Our rule simply adds 4 csec to the note terminating a phrase. For notes terminating subphrases the last 1 cs is used for a pause.

4. Evaluation

A listening experiment test was carried out in order to test the effect on musical acceptability of each of the rules specified above. Thirteen melodies were presented in pairs of performances. In one performance no rule was applied i.e. the performance involved no changes of the nominal durations and amplitudes, while in the other performance one of the seven rules applied. The melodies where chosen so as to expose particularly clearly the effect of the rule to be tested. The tape was presented over headphones to 9 musically highly experienced judges who were asked to mark on a paper which performance in each pair they preferred from a musical point of view. The results are shown in Table 3.

Table 3. Result of listening experiment, where 9 musical experts compared the musical quality of synthetic performances. The numbers indicate how many of the experts who preferred the rule controlled performance. The examples are specified in the **Appendix**

Rule	Example	Preference
1	Händel	9
	Alla turca	6
2	Sonate	4
	Nursery tune	5
3	Danse music	5
	Alla turca	6
4	Bourrée	8
	Courante	8
5	Geistl. Lied	5
	Gavotte	8
6	Happy Birthday	5
	Bellman	8
7	Fruelingstraum	7
	Sarabande	8

It can be seen in the Table that most examples were preferred by more than half of the jury, when one rule is applied. Still, in some cases the scores are quite low. In some cases this low score probably had an unexpected explanation. When rule number 5, which introduces pauses during leaps, was tuned loudspeakers were used, while headphones were used in the test. It seemed that the absence of room reverberation in the headphone listening made the pauses sound too long.

There is a trend that excerpts from the baroque era are rated higher than other excerpts. It cannot be excluded that the reason for this is that our rules are perhaps optimized with respect to this kind of music. However, we believe that the reason is that our examples from the postbaroque periods are more dependent on the harmonic progressions underlying the melody; such examples are likely to suffer more from the fact that none of these seven rules reacts to the harmonic context.

It seems fair to conclude from the test that in a majority of cases each of the seven rules improve the musical acceptability of the performance. Still, a single rule does not appear to improve the quality of a performance to any appreciable extent. Rather the effect is that the mechanical impression of the performance is eliminated.

It is likely that the effect of the individual rule is dependent on the effect of other rules. Therefore we would expect that the summed effect of all rules will improve the performance much more than a single rule.

An important question is to what extent the exact formulation of our rules is critical; can the deviations from nominal durations and the changes in amplitude be replaced by a corresponding randomization? An informal listening test was made where two performances were compared. The first one was in accordance with the rules, while in the second one the durations and amplitudes were varied randomly within the limits set by our rule system; a random number between 0 and 1 was used to determine the quantity of the effects. The result demonstrated that the last mentioned, randomized performance appeared to lack stability in some way. This is not astonishing. It seems obvious that no event, which may catch a devoted listener's attention in an ideal performance, can be explained as random. It appears axiomatic that all which can be noticed in an ideal performance serves the purpose of communication.

5. Discussion

If it is correct to assume that our rules all contribute to the performance of a melody, an interesting question is what purpose these rules serve. Some of the rules probably reflect shear convention while some other might be dependent on properties of human perception and cognition. In any case it is interesting to speculate a little about the possible background of some of the rules.

We believe that some rules derive from the human voice. The reduction of amplitude steps at tone boundaries and the general growth of amplitude with fundamental frequency may be examples of this. Other rules are likely to have a psycho-acoustic background. The principle of lengthening the note which terminates a leap might reflect certain effects studied by VAN NOORDEN [5]; the disruption of a melodic line caused by a wide leap can be eliminated by reducing the speed of playing. A similar effect might be obtained if the tone terminating the leap is lengthened.

There may also be purely psychological foundations for some of the rules. The lengthening both of the notes terminating a leap and of the first note appearing over a new harmony may serve the purpose of convincing the listener that the note, even though unexpected, was indeed intended. Also, the increase in tempo during sequences of rising intervals may have a psychological background. A rising interval may possess an "activating" connotation and is often associated with an upward locomotion. Then, the player may need to stress that this "upward motion" is not hard to execute, and he announces this by increasing the tempo somewhat. Another possible explanation would be that pitch rises are often combined with an increases tempo in excited speech. The rule lengthening the last tone in a phrase would be recognized by speech researchers; in speech this phenomenon is well known and is generally referred to as the principle of final lengthening. Our acquaintance with speech may have programmed us so that lengthening is interpreted as a sign of a final element. If so, the reversion of the rule (shortening of final note in a phrase) would hardly be possible in any performance.

There are, no doubt, more rules than those presented above, particularly if we turn to the performance of polyphonic music. Thus, we do not pretend that our rules are sufficient to produce a musically acceptable performance. Neither do we pretend that the present formulation of the rules is definite and different repertoires will probably need at least partly different rule systems. What we pretend is that the musical quality of a performance can be significantly improved by applying a set of rules.

Another important point is to state that we have not tried to model the multitude of choices which is available to the living musician and which allows him to play the same piece in many different ways, all of which are equally acceptable from a musical point of view. There are, on the other hand, some possibilities to include such a liberty. One way is to allow for personal variations in quantity of the rules. Another possibility is to vary the structural interpretation of the music which the musician performs. Also, we would like to declare that we do not believe that our performance rules must always be obeyed in a good performance. On the contrary, such a performance may be boring in the long run. We believe that musicians should violate one or more of the performance rules as soon as they want to tell something in particular or excite the audience by means of a surprise.

In our work with formulating rules we have observed that the quantity of the rule's effect is very critical. If the effect is too great, the effect becomes obvious in a physical sense, and then the effect is embarrassing. For example if the lengthening of notes terminating leaps is too large, one can hear that this is what happens, and the result is musically unacceptable. The musically proper effect of the rule arises when its effect is just noticeable but not indentifiable for its nature. Perhaps this is something which is essential for art in general: we do not want to be disturbed by information about the technical means behind the piece of art, we just want to enjoy it!

We repeat that these attempts to interpret the rules represent pure speculation. The point is that some of our rules probably have a background of some kind which may be independent of music. We believe that further research on this background will be interesting and rewarding.

6. Conclusions

From the above we conclude the following:

1. It is possible to improve the musical acceptability of a performance by applying a limited set of "pronounciation" rules.

2. Such rules can be discovered by means of a analysis-synthesis approach.

3. Such an approach enables us to formulate new hypotheses as to how the rules tried should be complemented and thus to contribute to knowledge about and scientific understanding of music.

Appendix

Origin of the melody excerpts used in the evaluation test:

J. S. BACH: *Gavotte* from Partita E major for violin solo, BWV 1006; *Courante* from Suite D major for cello solo, BWV 1012; *Bourreé* from Suite C Major for cello solo, BWV 1009; *Sarabande* from Suite c minor for cello solo, BWV 452; *"Dir, dir, Jehova ..."* *Geistl. Lied* BWV 452.

C. M. BELLMAN: *Vila vid denna källa*, Nr 82 in Fredmans epistalar.

G. F. HÄNDEL: *Sonata E major* for violin and continuo, op. 1, no. 15.

K. JULARBO: *Livet i Finnskoga* (*Dance music*).

W. A. MOZART: *Alla turca* from Sonate für Klavier A major, K 331.

F. SCHUBERT: *Frülingstraum* from Winterreise, op. 89, no. 11, D 911; *Sonate* für Violin und Klavier, op. posth. 137:1, D 384.

Traditional: *Happy birthday*.

References

[1] A. ASKENFELT, *Automatic notation of played music: the VISA project*, Fontes Artis Musicae, **26**, 109–120 (1979).

[2] I. BENGTSSON, A. GABRIELSSON, *Rhytm research in Uppsala*, In: Music room acoustics, Publ. 17 issued by the Royal Swedish Academy of Music, Stockholm, pp 19–56.

[3] R. CARLSON, B. GRANSTRÖM, *A phonetically oriented programming language for rule description of speech*, In: Speech Communication, [ed.] G. Fant, Almqvist and Wicksell, Stockholm, vol II, pp 245–253.

[4] B. LARSSON, *Music and Singing Synthesis Equipment MUSSE*, Speech Transmission Laboratory Quarterly Progress and Status Report 1, 38–40 (1977).

[5] L. P. A. S. VAN NOORDEN, *Temporal coherence in the perception of tone sequences*, Druk vam Voorschooten (Diss).

[6] A. RAKOWSKI, *Pitch discrimination at the threshold of hearing*, Proceedings of the 7th International Congress on Acoustics, **3**, 373–376 (1971).

[7] J. SUNDBERG, *Synthesis of singing*, Swedish Journal of Musicology, **60**, 1, 107–112 (1978).

[8] J. SUNDBERG, B. LINDBLOM, *Generative theories in language and music descriptions*, Cognition **4**, 99–122, (1976).

[9] J. SUNDBERG, *In tune or not? A study of fundamental frequency in music practice*, Speech Transmission Laboratory, Quarterly Status and Progress Report 1, 49–78 (1982).

Received September 13, 1988.

ARCHIVES OF ACOUSTICS
13, 3-4, 281-301 (1988)

STATISTICAL SOUND EVALUATION FOR THE IMPROVEMENT OF SOUND INSULATION SYSTEMS

MITSUO OHTA

Faculty of Engineering, Hiroshima University
(Shitami, Saijo-cho, Higashi-Hiroshima, 724 Japan)

HIROFUMI IWASHIGE

Faculty of Education, Hiroshima University
(2-17, Midori-machi, Fukuyama, 720 Japan)

The random noise fluctuation encountered in our living environment, such as street noise, road traffic noise, etc. exhibits various kinds of probability distribution forms apart from the usual Gaussian distribution due to the diversified causes of the fluctuations. From the practical viewpoint of control and regulation of such environmental noise, several statistics, such as median, L_5 and L_{10} (in general, so called L_α sound level), directly connected with the probability distribution form of random noise fluctuation are very often used for evaluation of the human response. Thus, it is essential to establish a systematic method for evaluating the effect of the system change of noise control on the widely-used standard noise index such as L_α. In this paper, general and fundamental considerations for statistical evaluation of transmitted sound waves have been theoretically proposed, when the system characteristic of the sound insulation is changed by any improvement work. The theoretical result was experimentally confirmed not only by the result of the digital simulation technique, but also by actually observed data obtained using reverberation room method. The results of the experiment are in good agreement with our theory.

Praca poświęcona jest badaniu przechodzenia losowej fali akustycznej przez ekran dźwiękochłonny pojedynczy i podwójny. Oryginalnym wynikiem jest porównanie zjawiska przechodzenia energii w modelu z pojedynczą ścianką z modelem z podwójną ścianką. Przedstawiono porównania wyników analitycznych z rezultatami doświadczeń.

Introduction

The random fluctuation of noise and vibration encountered in our living environment, such as street noise, road traffic noise, machine or structure vibration, etc., exhibits various kinds of probability distribution forms apart from a usual

Gaussian distribution due to the diversie causes of the fluctuations. From the practical viewpoint of control and regulation for such environmental noise and vibration pollutions, several of statistics, such as median, L_5 and L_{10} (in general, L_α(($100 - \alpha$) percentile) sound or vibration levels), directly connected with a whole shape of the probability distribution of random noise and vibration fluctuations are very important for an evaluation of the human response [1]. Thus, it must be an essential problem to establish a systematic method for the purpose of evaluating the effect of system change of noise or vibration controls on the standard noise index such as L_α. In this paper, general and fundamental considerations for the statistical evaluation of noise or vibration have been theoretically proposed, when the characteristic of the control system is changed by any improvement work. That is, when an arbitrarily distributed random signal is passed through noise or vibration control systems, a unified statistical treatment for the probability density function of its output energy fluctuation has been proposed in the universal form of an expansion series expression. Hereupon, an input random noise may have arbitrary types of the first and higher order correlations among arbitrarily chosen samples, and furthermore noise or vibration control systems have arbitrary linear characteristics of the finite memory type. For the purpose of finding systematically and universally the effect of the system change on the statistical evaluation quantities L_α of the output signal, the distribution function for the output energy fluctuation observed before the system characteristic is changed, has been taken into consideration as the first term of the unified expansion expression.

Furthermore, in view of the arbitrariness of possible input characteristics, the possible variety of noise or vibration control systems, and the complexity of the mathematical expressions involved and its statistical treatment, a digital simulation technique appears to be a powerful way of experimentally confirming the theoretical expressions. That is, in the simple and basic case when a homogeneous single wall is changed to a double wall by the improvement work as an example of a noise control system, the probability distribution function of the output sound energy fluctuation simulated on a digital computer has been drawn graphically for the comparison between theory and experiment. Finally, the validity and usefulness of our theoretical consideration have been confirmed experimentally by applying it to the actual observed noise data obtained using the reverberation room method. The experiment has been carried out with the single wall and double wall composed of an aluminum panel, using road traffic noise as an arbitrary random incident wave. We have been able to observe a good agreement between theory and experiment in the above two cases.

1. General theory

When a general random sound pressure wave of an arbitrary non-Gaussian distribution type $X(t)$ (let X_j be the sampled value at time point t_j of $X(t)$; $j = 1, 2, \ldots, K$) passes through the time-invariant linear system with an impulse response function $h(t)$ (let b_{ij} be the sampled weighting value of $h(t)$), the transmitted sound pressure wave $Y(t)$ (let Y_i be the sampled value at time point t_i of $Y(t)$; $i = 1, 2, \ldots, N$) can be easily given as the following equation in the discrete form [2, 3]:

$$Y_i = \sum_{j=1}^{K} b_{ij} X_j, \tag{1}$$

where K is the order of linear system.

On the other hand, an N-dimensional probability density function $P_{\mathbf{z}}(\mathbf{z})$ for the transmitted sound pressure wave $Z(t)$ (let Z_i be the sampled value of $Z(t)$; $i = 1, 2, \ldots, N$) after changing the system characteristic from $h(t)$ to $W(t)$ (with sampled weighting value a_{ij}) can be expressed on the basis of the N-variate joint probability density function $P_{\mathbf{Y}}(\mathbf{Y})$ for the above transmitted sound pressure wave $Y(t)$ of the original system $h(t)$ as follows

$$P_{\mathbf{z}}(\mathbf{Z}) = \sum_{r=0}^{\infty} \sum_{r_1+\ldots+r_N=r} \frac{B_r(r_1, r_2, \ldots, r_N)}{r_1!\, r_2! \ldots r_N!} (-1)^{r_1+r_2+\ldots+r_N} \frac{\partial^{r_1+r_2+\ldots+r_N}}{\partial Z_1^{r_1} \partial Z_2^{r_2} \ldots \partial Z_N^{r_N}} P_{\mathbf{y}}(\mathbf{Z}), \tag{2}$$

where the above expansion coefficients are given as:

$$B_0(0, 0, \ldots, 0) = 1,$$

$$B_1(0, 0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0) = \sum_{p_1=1}^{K} (a_{ip_1} - b_{ip_1}) \varkappa_{x1}(0, 0, \ldots, 0, \overset{p_1}{1}, 0, \ldots, 0),$$

$$B_2(0, 0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0, \overset{j}{1}, 0, \ldots, 0) = \sum_{p_1, p_2=1}^{K} (a_{ip_1} a_{jp_2} - b_{ip_1} b_{jp_2})$$

$$\times \varkappa_{x2}(0, 0, \ldots, 0, \overset{p_1}{1}, 0, \ldots, 0, \overset{p_2}{1}, 0, \ldots, 0) + B_1(0, 0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0)$$

$$\times B_1(0, 0, \ldots, 0, \overset{j}{1}, 0, \ldots, 0),$$

$$B_3(0, 0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0, \overset{j}{1}, 0, \ldots, 0, \overset{k}{1}, 0, \ldots, 0) = \sum_{p_1, p_2, p_3=1}^{K} \tag{3}$$

$$\times (a_{ip_1} a_{jp_2} a_{kp_3} - b_{ip_1} b_{jp_2} b_{kp_3}) \varkappa_{x3}(0, 0, \ldots, 0, \overset{p_1}{1}, 0, \ldots, 0, \overset{p_2}{1}, 0, \ldots, 0, \overset{p_3}{1}, 0, \ldots, 0)$$

$$+ B_2(0, 0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0, \overset{j}{1}, 0, \ldots, 0) B_1(0, 0, \ldots, 0, \overset{k}{1}, 0, \ldots, 0)$$

$$+ B_2(0, 0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0, \overset{k}{1}, 0, \ldots, 0) B_1(0, 0, \ldots, 0, \overset{j}{1}, 0, \ldots, 0)$$

$$+ B_2(0, 0, \ldots, 0, \overset{j_1}{1}, 0, \ldots, 0, \overset{k}{1}, \ldots, 0) B_1(0, 0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0) - 2B_1$$

$$\times (0, 0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0) B_1(0, 0, \ldots, 0, \overset{j}{1}, 0, \ldots, 0) B_1(0, 0, \ldots, 0, \overset{k}{1}, 0, \ldots, 0).$$

Hereupon, $\varkappa_{xl}(0, \ldots, 0, \overset{p_1}{1}, 0, \ldots, 0, \overset{p_2}{1}, 0, \ldots, 0, \overset{p_l}{1}, 0, \ldots, 0)$ denotes the l-dimensional correlation function of $X(t)$.

In Eq. (2), it is noteworthy that the N-dimensional probability density function $P_{\mathbf{Y}}(\mathbf{z})$ is not directly related to Z_i itself, but is given by merely substituting $Z_i's$ for random variables $Y_i's$ in the joint probability density $P_{\mathbf{Y}}(\mathbf{Y})$ of a transmitted sound pressure wave $Y_i (i = 1, 2, \ldots, N)$. In the above expression $P_{\mathbf{Z}}(\mathbf{z})$, in a specific case when $N = 1$, $P_z(z)$ can be directly expressed as follows (The usefulness of this probability expression is briefly discussed in **Appendix 1**):

$$P_z(Z) = \sum_{r=0}^{\infty} (-1)^r \frac{B_r(r)}{r!} \frac{\partial^r}{\partial Z^r} P_y(Z), \tag{4}$$

where the expansion coefficients are simplified as follows:

$$B_0(0) = 1,$$

$$B_1(1) = \sum_{p_1=1}^{K} (a_{p_1} - b_{p_1}) \varkappa_{x1}(0, 0, \ldots, 0, \overset{p_1}{1}, 0, \ldots, 0), \tag{5}$$

$$B_2(2) = \sum_{p_1, p_2=1}^{K} (a_{p_1} a_{p_2} - b_{p_1} b_{p_2}) \varkappa_{x2}(0, 0, \ldots, 0, \overset{p_1}{1}, 0, \ldots, 0, \overset{p_2}{1}, 0, \ldots, 0) + B_1(1)^2,$$

$$B_3(3) = \sum_{p_1, p_2, p_3=1}^{K} (a_{p_1} a_{p_2} a_{p_3} - b_{p_1} b_{p_2} b_{p_3}) \varkappa_{x3}$$

$$\times (0, 0, \ldots, 0, \overset{p_1}{1}, 0, \ldots, 0, \overset{p_2}{1}, 0, \ldots, 0, \overset{p_3}{1}, 0, \ldots, 0) + 3B_2(2)B_1(1) - 2B_1(1)^3.$$

From the above theoretical results, Eqs. (2) or (4), we can find that the joint probability density expression of Z_i can be expressed in a universal expansion form in which the joint probability density $P_{\mathbf{Y}}(\mathbf{Y})$ of random process Y_i is taken into the first term (so that it may be convenient for our purpose of research) and its successive derivatives are taken in the second and higher expansion terms. Furthermore, the effect of system change on the resultant distribution form of the transmitted sound pressure fluctuation is explicitly reflected in each expansion coefficient (cf. Eqs. (3) or (5)).

Now, as was reported in the previous papers [9], we can employ the following probability density function of a statistical Hermite series-type expression as $P_y(z)$ for a transmitted sound pressure wave $Z(t)$ of an arbitrary distribution type for the non-changed system $h(t)$:

$$P_y(z) = \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{-z^2/2\sigma_y^2} \sum_{n=0}^{\infty} A_n H_n\left(\frac{z}{\sigma_y}\right), \tag{6}$$

where:

$$A_n \overset{\Delta}{=} \frac{1}{n!}\left\langle H_n\left(\frac{Y}{\sigma_y}\right)\right\rangle,$$

($\langle * \rangle$ denotes a statistical mean operation with respect to random variable $*$). Substituting Eq. (6) into Eq. (4) and applying the well-known relation between the normal Gaussian distribution function and the Hermite polynomial:

$$\frac{1}{\sqrt{2\pi}} e^{-\eta^2/2} H_n(\eta) = (-1)^n \frac{d^n}{d\eta^n} \frac{1}{\sqrt{2\pi}} e^{-\eta^2/2}, \tag{7}$$

we can obtain the probability density function $P_z(z)$ of a transmitted sound wave after changing the system from $h(t)$ to $W(t)$ concerning its impulse response function as:

$$P_z(z) = \sum_{r=0}^{\infty} \sum_{n=0}^{\infty} \frac{B_r(r)}{r!} \frac{A_n}{\sigma_y^r} \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{-z^2/2\sigma_y^2} H_{n+r}\left(\frac{z}{\sigma_y}\right). \tag{8}$$

The probability density function of the sound energy fluctuation is more important than that of the sound presssure wave fluctuation for the purpose of evaluation of noise pollution. By using an integral formula:

$$\int_{-\infty}^{\infty} e^{-B^2 x^2} H_n(X) dx = \begin{cases} 0 & (n\text{: odd}) \\ (1/B)^{n+1}(1-2B^2)^{n/2}\,\Gamma((n+1)/2) & (n\text{: even})' \end{cases} \tag{9}$$

the moment generating function of Laplace transformation type for transmitted sound energy $E(\overset{\Delta}{=} z^2)$ is derived as follows:

$$M_E(\theta) \overset{\Delta}{=} \langle e^{\theta E} \rangle = \sum_{r=0}^{\infty} \sum_{n=0}^{\infty} \frac{B_r(r)}{r!} \frac{A_n}{\sigma_y^r} \frac{(\sqrt{2})^{n+r}}{\sqrt{\pi}}\, \Gamma\left(\frac{n+r+1}{2}\right)$$
$$\times \frac{1}{(1-2\sigma_y^2\theta)^{1/2}} \left(\frac{2\sigma_y^2\theta}{1-2\sigma_y^2\theta}\right)^{\frac{n+r}{2}} \quad (n+r\text{: even}). \tag{10}$$

By applying an integral formula on an associated Laguerre polynomial $L_n^{(\alpha)}(*)$:

$$\int_0^{\infty} e^{-pt} t^{\alpha} e^{\beta t} L_n^{(\alpha)}(at) dt = \frac{\Gamma(n+\alpha+1)}{n!} \frac{(p-a-\beta)^n}{(p-\beta)^{n+\alpha+1}} \tag{11}$$

to the inverse Laplace transformation of Eq. (10), the probability density function $P_E(E)$ of the transmitted sound energy through the changed system $W(t)$ is consequently given in the form of an expansion expression as follows (see Appendix 2):

$$P_E(E)=\sum_{r=0}^{\infty}\sum_{n=0}^{\infty}\frac{B_r(r)}{r!}\frac{A_n}{\sigma_y^r}\frac{(\sqrt{2})^{n+r}}{\sqrt{\pi S}}(-1)^{\frac{n+r}{2}}\cdot\left(\frac{n+r}{2}\right)!\,E^{-\frac{1}{2}}e^{-\frac{E}{s}}L_{\frac{n+r}{2}}^{(-\frac{1}{2})}\left(\frac{E}{s}\right)$$

$$=\sum_{m=0}^{\infty}\left\langle H_{2m}\left(\frac{Y}{\sigma_y}\right)\right\rangle\frac{(-1)^m}{\Gamma\left(m+\frac{1}{2}\right)2^m\sqrt{s}}E^{-\frac{1}{2}}e^{-\frac{E}{s}}L_m^{(-\frac{1}{2})}\left(\frac{E}{s}\right) \quad (12)$$

$$+\sum_{r=1}^{\infty}\sum_{n=0}^{\infty}\frac{B_r(r)A_n(\sqrt{2})^{n+r}}{r!\,\sigma_y^r\sqrt{\pi s}}(-1)^{\frac{n+r}{2}}\left(\frac{n+r}{2}\right)!\cdot E^{-\frac{1}{2}}e^{-\frac{E}{s}}L_{\frac{n+r}{2}}^{(-\frac{1}{2})}\left(\frac{E}{s}\right)$$

with $s\stackrel{\Delta}{=}2\sigma_y^2$.

If we introduce a dimensionless variable $\eta\,(\stackrel{\Delta}{=}E/s)$ for the purpose of obtaining the universal expression, we can easily find the following unified expression:

$$P_\eta(\eta)=\sum_{m=0}^{\infty}\left\langle H_{2m}\left(\frac{Y}{\sigma_y}\right)\right\rangle\frac{(-1)^m}{\Gamma\left(m+\frac{1}{2}\right)2^m}\eta^{-\frac{1}{2}}e^{-\eta}L_m^{(-\frac{1}{2})}(\eta)$$

$$+\sum_{r=1}^{\infty}\sum_{m=0}^{\infty}\frac{B_r(r)A_m(\sqrt{2})^{m+r}}{r!\,\sigma_y^r\sqrt{\pi}}(-1)^{\frac{m+r}{2}}\left(\frac{m+r}{2}\right)!\,\eta^{-\frac{1}{2}}e^{-\eta}L_{\frac{m+r}{2}}^{(-\frac{1}{2})}(\eta). \quad (13)$$

Hereupon, the relation

$$L_n^{(-1/2)}(\eta)=(-1)^n\frac{1}{2^n n!}H_{2n}(\sqrt{2\eta})$$

should be observed.

In this expression, the transmitted sound energy distribution through the original system $h(t)$ is taken as the first expansion term and the effect of characteristic change for the sound insulation due to changing the system. The statistical characteristic of an incident sound wave is reflected in the expansion coefficients of the remainder terms.

2. Step response function for single wall and double wall

2.1. Step response function for single wall

By use of the mass law, the frequency transfer function G_{I} for a single wall is directly given as [12]:

$$G_{\mathrm{I}}(j\omega)=\frac{1}{1+j\omega T},\qquad T\stackrel{\Delta}{=}\frac{m}{2\varrho c}\cos\theta, \quad (14)$$

where m, ϱc and θ denote respectively the surface mass of single wall, the characteristic impedance of air and the incident angle. From the above equation, the step response function $S_I(t)$ is easily derived as follows:

$$S_I(t) = 1 - e^{-t/T}. \tag{15}$$

2.2. *Step response function for double wall*

As reported in our previous paper from the viewpoint of the distributed constant circuit theory, the frequency transfer function G_{II} for a double wall is given as follows [13]:

$$G_{II}(j\omega) = \frac{1}{1 + j\omega(T_1 + T_2) + (j\omega)^2 T_1 T_2 (1 - e^{-j\omega\tau})}, \tag{16}$$

$$T_1 \triangleq \frac{m_1}{2\varrho c}\cos\theta, \; T_2 \triangleq \frac{m_2}{2\varrho c}\cos\theta, \quad \tau_\theta \triangleq \frac{2d}{c}\cos\theta,$$

where m_1, m_2 denote the surface mass of each panel, and d denotes the width of the air gap. Hereupon, the above expression of $G_{II}(s)$ can be rewritten as the following expansion form (see Appendix 3):

$$G_{II}(s) = \frac{1}{(1+T_1 s)(1+T_2 s)} \sum_{n=0}^{\infty} \left[\frac{T_1 T_2 s^2}{(1+T_1 s)(1+T_2 s)} e^{-\tau_\theta s} \right]^n = \sum_{n=0}^{\infty} \sum_{r=0}^{\infty} \frac{1}{r!} \frac{T_2^n}{T_1^{n+2}}$$
$$\times \left(1 - \frac{T_2}{T_1}\right)^r (n+1)(n+2) \dots (n+r) \cdot \frac{s^{2n+r}}{\left(s + \frac{1}{T_1}\right)^{2n+r+2}} e^{-n\tau_\theta s}. \tag{17}$$

Using a formula on the Laplace transform of a Laguerre polynomial (see Eq. (11)), the step response function for a double wall can be explicitly derived in the following expansion form as one of a Laguerre filter in some sense:

$$S_{II}(t) = \sum_{n=0}^{\infty} \sum_{r=0}^{\infty} \frac{1}{r!} \frac{T_2^n}{T_1^{n+2}} \left(1 - \frac{T_2}{T_1}\right)^r (n+1)(n+2) \dots (n+r)$$
$$\times \frac{2n+r-1)!}{\Gamma(2n+r+2)} (t - n\tau_\theta)^2 e^{-\frac{t-n\tau_\theta}{T_1}} L^{(2)}_{2n+r-1}\left(\frac{t - n\tau_\theta}{T_1}\right). \tag{18}$$

Especially, if we focus our attention on the special case when every surface mass is equal (i.e., $m_1 = m_2 = m$ $(T_1 = T_2 = T)$), the above expression of $S_{II}(t)$ can be reduced to the expansion expression given by:

$$S_{\mathrm{II}}(t)=\frac{1}{T^2}\sum_{n=0}^{\infty}\frac{(2n-1)!}{\Gamma(2n+2)}(t-n\tau_\theta)^2\,e^{-\frac{t-n\tau_\theta}{T}}\,L_{2n-1}^{(2)}\left(\frac{t-n\tau_\theta}{T}\right). \tag{19}$$

Fig. 1 shows the step response curves for a single wall with various values of the incident angle θ obtained by use of Eq. (15). On the other hand, Eq. (19) has an infinite expansion form in the form of a Laguerre filter, but in practice we must use only the first finite term. Fig. 2 is introduced to find what number of expansion terms in the expression of $S_{\mathrm{II}}(t)$ (see Eq. (19)) is needed to evaluate $S_{\mathrm{II}}(t)$ as exactly as possible. In this figure a comparison between the step response curve of a single wall $S_{\mathrm{I}}(t)$ in the specific case of $m=m_1+m_2$ and the step response curves of a double wall $S_{\mathrm{II}}(t)$ with m_1 and m_2 is shown, where a single surface mass is taken as $m=5.5\ \mathrm{Kg/m^2}$ $(=m_1+m_2)$ for a single wall and two surface masses $m_i=2.75\ \mathrm{Kg/m^2}$ $(i=1, 2)$ are taken by letting $d=0$ and $\theta=0°$ for a double wall. After confirming an agreement between the two kinds of step response curves ($S_{\mathrm{I}}(t)$ and $S_{\mathrm{II}}(t)$), we can find that it is sufficient to take the first 8 expansion terms of Eq. (19) to evaluate $S_{\mathrm{II}}(t)$. Fig. 3 shows the step response curves of a double wall with various values of the incident angle θ, where the two surface masses are equally taken as $m_1=m_2=3.22\ \mathrm{Kg/m^2}$ and the width of the air gap is set as $d=0.05$ m.

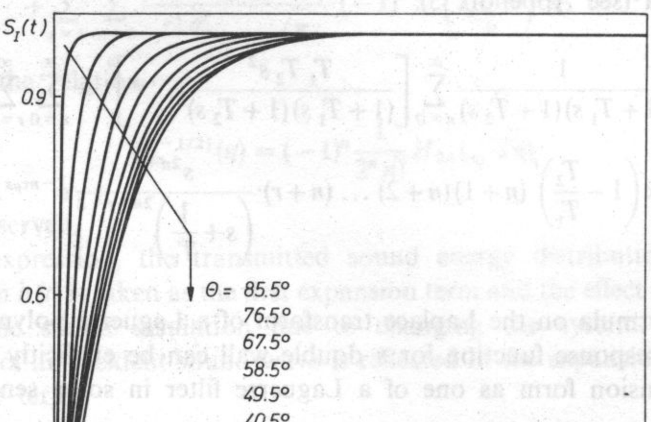


Fig. 1. Step response curves for a single wall with various values of incident angle θ, in the case with $m=3.22\ \mathrm{Kg/m^2}$ (cf. Eq. (15))

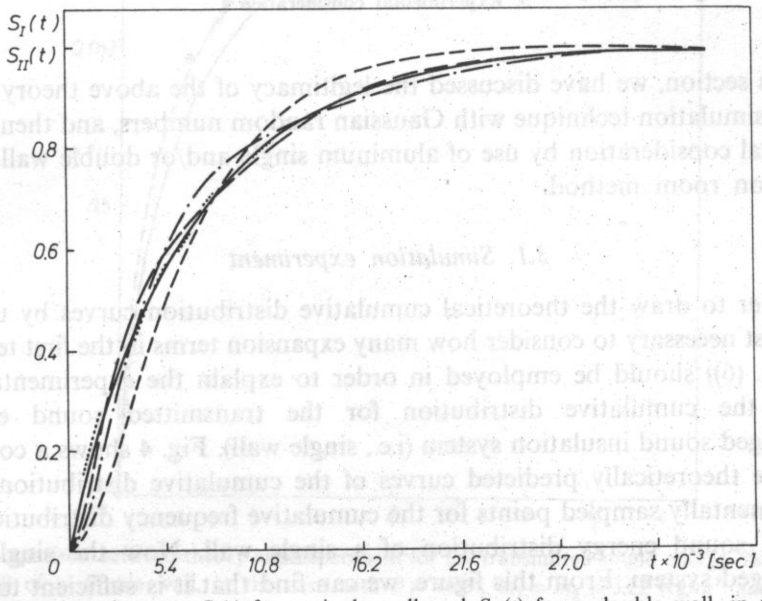


Fig. 2. A comparison between $S_I(t)$ for a single wall and $S_{II}(t)$ for a double wall, in the case with $m_1 = m_2 = 2.75$ Kg/m^2, $d = 0$ m and $\theta = 0°$. Theoretical curve of $S_I(t)$ (cf. Eq. (15)) is shown by (———). Theoretical curves of $S_{II}(t)$ (cf. Eq. (19)) are shown with the degree of approximation n [$n = 0$ (– – – –), $n = 1$ (—·—), $n = 2$ (— —) and $n = 7$ (..........)]

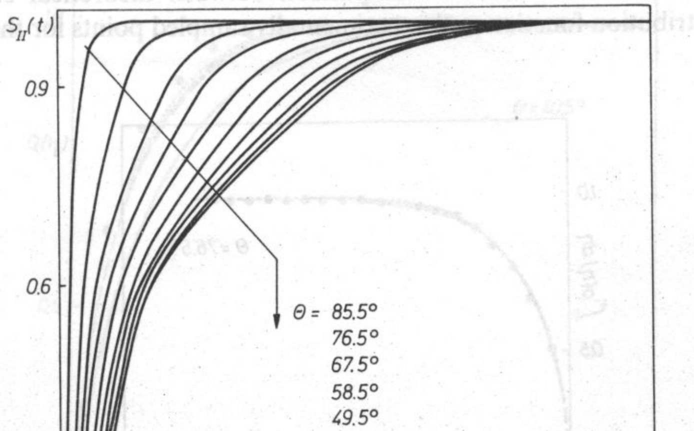


Fig. 3. Step response curves for a double wall with various values of incident angle θ, in the case with $m_1 = m_2 = 3.22$ Kg/m^2, $d = 0.05$ m (cf. Eq. (19))

3. Experimental consideration

In this section, we have discussed the legitimacy of the above theory by use of the digital simulation technique with Gaussian random numbers, and then by actual experimental consideration by use of aluminum single and/or double walls with the reverberation room method.

3.1. *Simulation experiment*

In order to draw the theoretical cumulative distribution curves by use of Eq. (13), it is first necessary to consider how many expansion terms in the first term of Eq. (13) (cf. Eq. (6)) should be employed in order to explain the experimental sample points of the cumulative distribution for the transmitted sound energy of a non-changed sound insulation system (i.e., single wall). Fig. 4 shows a comparison between the theoretically predicted curves of the cumulative distribution function and experimentally sampled points for the cumulative frequency distribution on the transmitted sound energy distribution of a single wall. Now the single wall is a non-changed system. From this figure, we can find that it is sufficient to take the first 3 expansion terms to calculate the distribution curve in the case with $\theta = 76.5°$ (we have confirmed that it is sufficient to take the first three expansion terms to fit the experimental sample points in other cases with various values of θ).

Figures 5, 6, 7 and 8 show a comparison between theoretical curves of the cumulative distribution function and experimentally sampled points for the cumulative

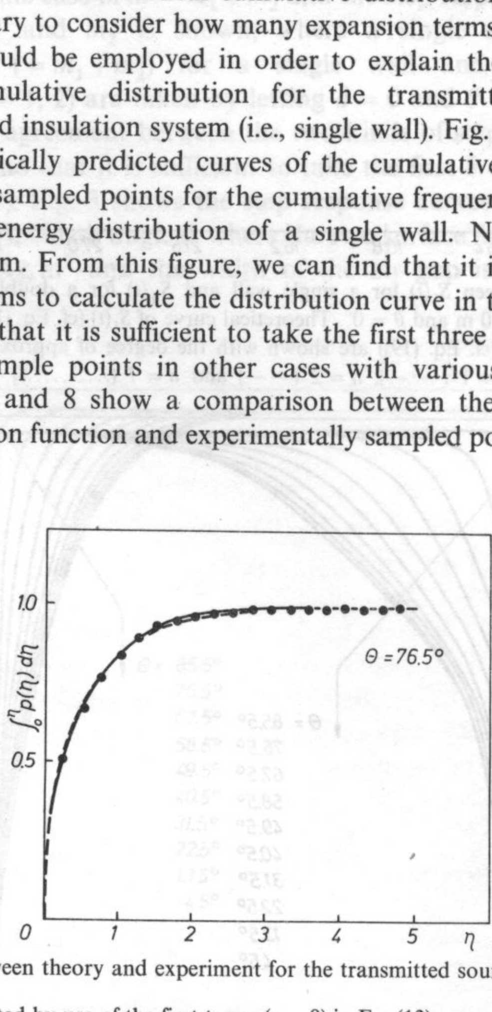


Fig. 4. A comparison between theory and experiment for the transmitted sound energy distribution $Q(\eta)$ $(\overset{\Delta}{=} \int_0^\eta P(\eta)d\eta$; $P(\eta)$ is evaluated by use of the first terms $(r = 0)$ in Eq. (13) corresponding to a single wall) of a single wall, in the case with $\theta = 76.5$ and $m = 3.22$ Kg/m^2. Experimentally sampled points are marked by (●) and theoretical curves are shown with the degree of approximation m [$m = 0$ and 1 (– – – –), $m = 2$ (———)]

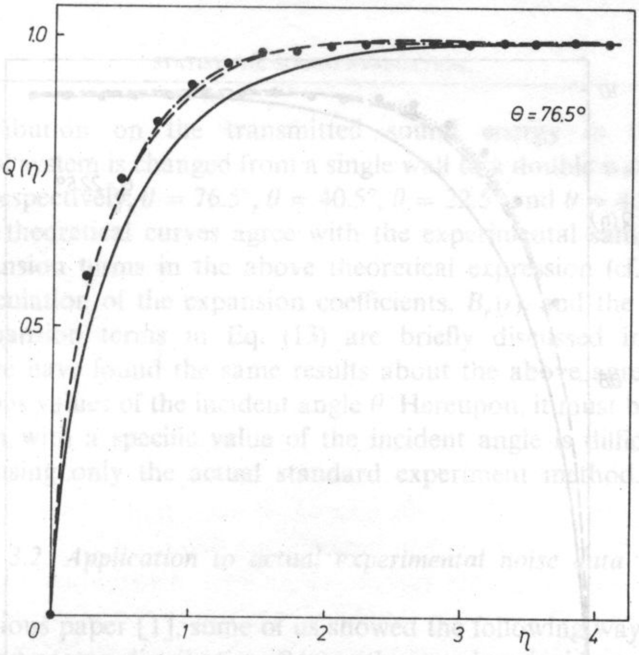


Fig. 5. A comparison between theory and experiment for the transmitted sound energy distribution $Q(\eta)$ (cf. Eq. (13)) of a double wall, in the case with $\theta = 76.5°$, $m_1 = m_2 = 3.22$ Kg/m^2 and $d = 0.05$ m. Experimentally sampled points are marked by (●) and theoretical curves are shown with the degree of approximation r [$r = 0$ (———), $r = 1, 2, 3, 4$ and 5 (– – –), $r = 6$ (— · —)]

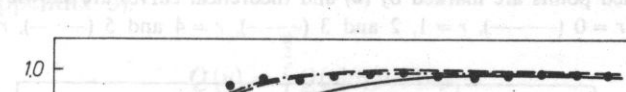


Fig. 6. A comparison between theory and experiment for the transmitted sound energy distribution $Q(\eta)$ (cf. Eq. (13)) of a double wall, in the case with $\theta = 40.5°$, $m_1 = m_2 = 3.22$ Kg/m^2 and $d = 0.05$ m. Experimentally sampled points are marked by (●) and theoretical curves are shown with the degree of approximation r [$r = 0$ (———), $r = 1, 2$ and 3 (– – –), $r = 4$ and 5 (— · —), $r = 6$ (.........)]

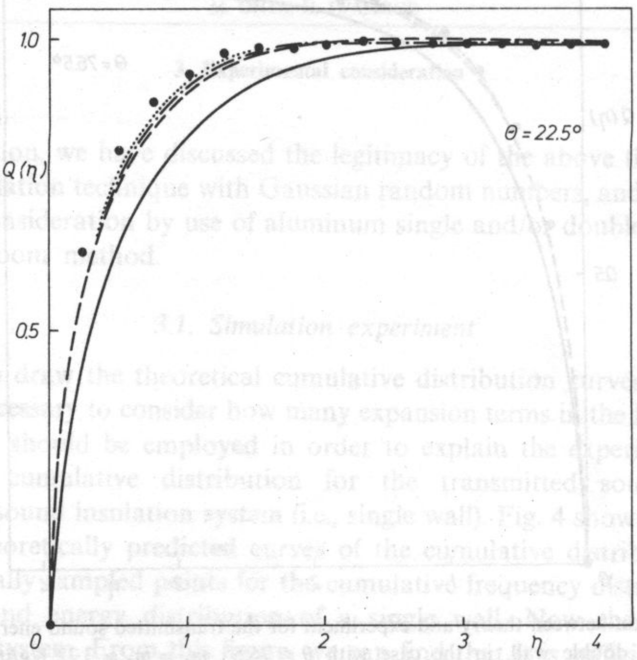


Fig. 7. A comparison between theory and experiment for the transmitted sound energy distribution $Q(\eta)$ (cf. Eq. (13)) of a double wall, in the case with $\theta = 22.5°$, $m_1 = m_2 = 3.22$ Kg/m^2 and $d = 0.05$ m. Experimentally sampled points are marked by (●) and theoretical curves are shown with the degree of approximation r [$r = 0$ (———), $r = 1$, 2 and 3 (– – – –), $r = 4$ and 5 (—·—), $r = 6$ (........)]

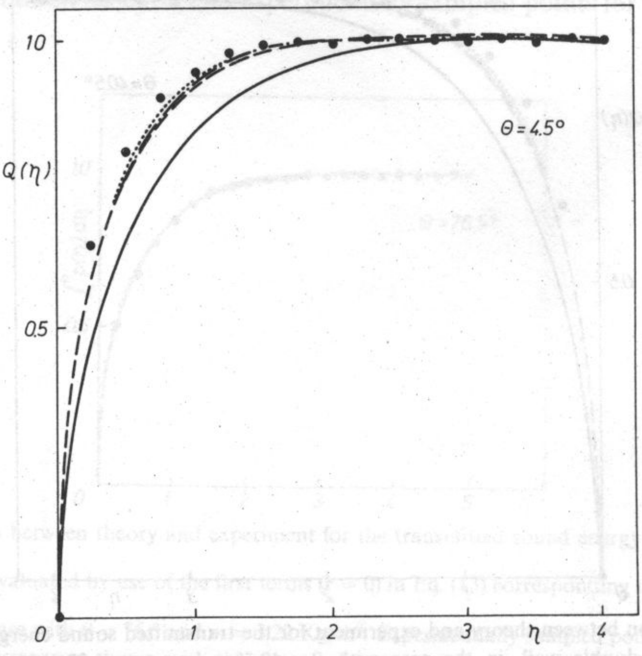


Fig. 8. A comparison between theory and experiment for the transmitted sound energy distribution $Q(\eta)$ (cf. Eq. (13)) of a double wall, in the case with $\theta = 4.5°$, $m_1 = m_2 = 3.22$ Kg/m^2 and $d = 0.05$ m. Experimentally sampled points are marked by (●) and theoretical curves are shown with the degree of approximation r [$r = 0$ (———), $r = 1$, 2 and 3 (– – – –), $r = 4$ and 5 (—·—), $r = 6$ (........)]

frequency distribution on the transmitted sound energy in the case when the noise control system is changed from a single wall to a double wall, by setting the incident angle respectively; $\theta = 76.5°$, $\theta = 40.5°$, $\theta = 22.5°$ and $\theta = 4.5°$. It is obvious that the above theoretical curves agree with the experimental sample points with increasing expansion terms in the above theoretical expression (cf. Eq. (13)) (The methods of calculation of the expansion coefficients, $B_r(r)$, and the decision of the number of expansion terms in Eq. (13) are briefly discussed in Appendix 4). Furthermore, we have found the same results about the above agreement in other cases with various values of the incident angle θ. Hereupon, it must be observed that such a situation with a specific value of the incident angle is difficult to measure practically by using only the actual standard experiment method.

3.2. *Application to actual experimental noise data*

In the previous paper [1], some of us showed the following way of treating the cumulative sound energy distribution $Q(\eta)$ with a random incidence. By letting the transmitted sound energy distribution with a specific incident angle θ be the conditional distribution $P(\eta/\theta)$ with a fixed value of θ, the transmitted cumulative sound energy distribution with a random incidence can be directly expressed as follows (see Appendix 5):

$$Q(\eta) = \int_0^{\frac{\pi}{2}} Q(\eta|\theta)P(\theta)d\theta, \tag{20}$$

where:

$$Q(\eta/\theta) = \int_0^{\eta} P(\eta'/\theta)d\eta' \quad (\eta' = E'/S) \qquad \text{and:}\ P(\theta) = \sin 2\theta.$$

Our experiments corresponding to this case with a random incidence have been done in the laboratory of Hiroshima University by use of an aluminum single wall and/or double wall. Then, we have first recorded both the sound pressure waves in a source room and in a reception room (which are reverberant) by use of a data recorder. Next, we have converted these analogue waves to digital quantities by an A–D converter and then evaluated them with use of a digital computer. We have used road traffic noise actually recorded in Hiroshima City as one example of an arbitrary sound input source. Fig. 9 shows a comparison between the theoretically predicted curves and experimentally sampled points for the cumulative distribution function on the transmitted sound energy for a single wall ($m = 3.22$ Kg/m^2) before our system change. As seen too evidently in this figure, we find that it is sufficient to take the first 3 expansion terms in Eq. (13).

Figure 10 shows a comparison between the theoretically predicted curves and experimentally sampled points for the cumulative distribution on a transmitted sound energy for a double wall ($m_1 = m_2 = 3.22$ Kg/m^2, $d = 0.05$ m) after changing the system from a single wall to a double wall. It is obvious that the present theoretically predicted curves agree with the experimentally sampled points with increasing expansion terms in our theoretical expression.

Hereupon, from the fundamental viewpoint, we have applied a new consideration for a random incidence (cf. Eq. (20)) to the present case. However, in the architectural acoustics field, the coefficient of transmission of the sound intensity with a random incident property is usually given by use of the frequency transfer function of α as follows:

$$\tau = \int_0^{\frac{\pi}{2}} \frac{1}{|\alpha|^2} P(\theta) d\theta, \quad \text{with } P(\theta) = \sin 2\theta, \tag{21}$$

after first averaging $1/|\alpha|^2$ by $P(\theta)$ instead of averaging $Q(\eta|\theta)$ by $P(\theta)$.

Figure 11 shows a comparison between the theoretically predicted curves and experimentally sampled points for the cumulative distribution on a transmitted sound energy by use of the latter simplified method. In the above case, after finding the system characteristics due to averaging Eqs. (15) and (19) by $P(\theta)$, Eq. (13) itself has been used instead of using Eqs. (13) and (20) to draw the theoretically predicted distribution curves. By comparing Fig. 11 with Fig. 10 evaluated by the former exact method, though there is some difference between theory and experiment in this figure, in practice this simplified method is still useful for the case with a rough evaluation. Finally, a practical example of applying the proposed evaluation method to actual design problems has been illustrated in Appendix 6.

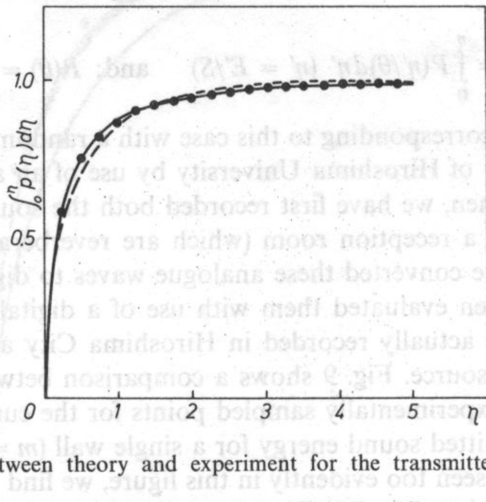


Fig. 9. A comparison between theory and experiment for the transmitted sound distribution $Q(\eta)$ $(\stackrel{\Delta}{=} \int_0^\eta P(\eta)d\eta$; $P(\eta)$ is evaluated by use of the first terms ($r = 0$) in Eq. (13) corresponding to a single wall and the random incident property Eq. (20)) of single wall, in a case with $m = 3.22$ Kg/m^2. Experimentally sampled points are marked by (●) and theoretical curves are shown with the degree of approximation m[$m = 0$ (– – –), $m = 1, 2, 3$ and 4 (——)]

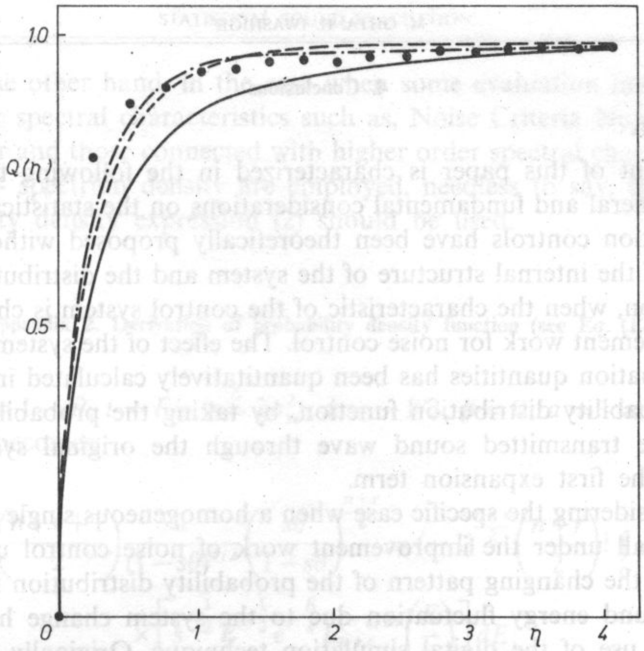


Fig. 10. A comparison between theory and experiment for the transmitted sound energy distribution $Q(\eta)$ (cf. Eqs. (13) and (20)) of a double wall, in the case with $m_1 = m_2 = 3.22$ Kg/m^2 and $d = 0.05$ m. Experimentally sampled points are marked by (●) and theoretical curves are shown with the degree of approximation r [$r = 0$ (———), $r = 1, 2, 3$ and 4 (– – – –), $r = 5$ and 6 (— · —)]

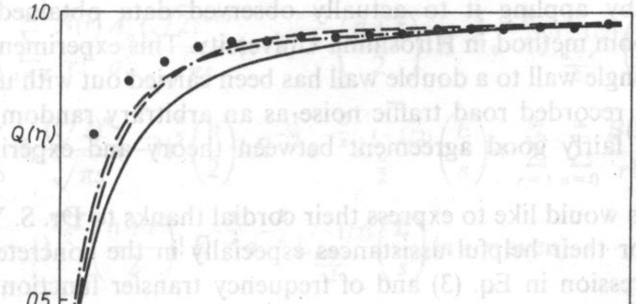


Fig. 11. A comparison between theory and experiment for the transmitted sound energy distribution $Q(\eta)$ of a double wall by use of a simplified method (cf. Eqs. (21) and (13)), in the case with $m_1 = m_2 = 3.22$ Kg/m^2 and $d = 0.05$ m. Experimentally sampled points are marked by (●) and theoretical curves are shown with the degree of approximation r [$r = 0$ (———), $r = 1, 2, 3$ and 4 (– – – –), $r = 5$ and 6 (— · —)]

4. Conclusions

The content of this paper is characterized in the following three points:

(1) The general and fundamental considerations on the statistical evaluation of noise or vibration controls have been theoretically proposed without any special assumptions of the internal structure of the system and the distribution type of the input fluctuation, when the characteristic of the control system is changed by some kind of improvement work for noise control. The effect of the system change on the statistical evaluation quantities has been quantitatively calculated in the expansion form of a probability distribution function, by taking the probability distribution function of the transmitted sound wave through the original system into consideration as the first expansion term.

(2) By considering the specific case when a homogeneous single wall is changed to a double wall under the improvement work of noise control under a specific incident angle, the changing pattern of the probability distribution function on the transmitted sound energy fluctuation due to the system change has been drawn graphically by use of the digital simulation technique. Originally, such a typical situation with a specific incident angle is essentially difficult to be examined only from a usual experimental method. Finally, a good agreement between theory and experiment has been found.

(3) Finally, the validity of our theoretical consideration has been confirmed experimentally by appling it to actually observed data obtained by using the reverberation room method in Hiroshima University. This experiment on the system change from a single wall to a double wall has been carried out with use of aluminum panels, taking a recorded road traffic noise as an arbitrary random incident wave. Consequently a fairly good agreement between theory and experiment has been found.

The authors would like to express their cordial thanks to Dr. S. Yamaguchi and Dr. K. Nagai for their helpful assistances especially in the concrete calculation of probability expression in Eq. (3) and of frequency transfer function in Appendices **2** and **3**.

Appendix 1. Usefulness of probability expression (see Eq. (4))

A typical case with $N = 1$ presented by Eq. (4) is obviously valid, if the statistics such as L_{eq} (equivalent sound pressure level), median, L_5 and L_{10} directly connected with a single-variate probability distribution form are employed as standard noise indexes. But, in order to estimate the expansion coefficients $B_r(r)$ ($r = 1, 2, 3, \ldots$) (cf. Eq. (5)), it should be observed that the statistical information on the bivariate correlation function so-called usual auto-correlation function and higher order multi-variate correlation functions for the input sound pressure fluctuation X_i must be used. On the other hand, in the case when some evaluation indexes connected with the power spectral characteristics such as, Noise Criteria Number and Noise Rating Number and those connected with higher order spectral characteristics (such as, bi- or poly- spectrum density are employed, needless to say, the multi-variate joint probability density expression (2) should be used.

Appendix 2. Derivation of probability density function (see Eq. (12))

Setting $p = 1 - s\theta$, $t = E/s$ ($s \overset{\Delta}{=} 2\sigma_y^2$), $\alpha = -1/2$, $\beta = 0$, $a = 1$ and $n = (n+r)/2$, equation (11) becomes:

$$\Gamma\left(\frac{n+r+1}{2}\right)\frac{1}{(1-s\theta)^{1/2}}\left(\frac{s\theta}{1-s\theta}\right)^{\frac{n+r}{2}} = (-1)^{\frac{n+r}{2}}\left(\frac{n+r}{2}\right)! \int_0^\infty e^{\theta E} \times\left[s^{-\frac{1}{2}}E^{-\frac{1}{2}}e^{-\frac{E}{s}}L^{(-1/2)}_{\frac{n+r}{2}}\left(\frac{E}{s}\right)\right]dE. \tag{A2.1}$$

Thus, by applying the above relationship to Eq. (10), the probability density function $P_E(E)$ can be directly obtained as:

$$P_E(E) = \sum_{r=0}^{\infty}\sum_{n=0}^{\infty}\frac{B(r)}{r!}\frac{A_n}{\sigma_y^r}\frac{(\sqrt{2})^{n+r}}{\sqrt{\pi s}}(-1)^{\frac{n+r}{2}}\left(\frac{n+r}{2}\right)! E^{-\frac{1}{2}}e^{-\frac{E}{s}}L^{(-1/2)}_{\frac{n+r}{2}}\left(\frac{E}{s}\right)$$

$$= \sum_{n=0}^{\infty}A_n\frac{(\sqrt{2})^n}{\sqrt{\pi s}}(-1)^{\frac{n}{2}}\left(\frac{n}{2}\right)! E^{-\frac{1}{2}}e^{-\frac{E}{s}}L^{(-1/2)}_{\frac{n}{2}}\left(\frac{E}{s}\right) + \sum_{r=1}^{\infty}\sum_{n=0}^{\infty}\frac{B(r)}{r!}\frac{A_n}{\sigma_y^r}\frac{(\sqrt{2})^{n+r}}{\sqrt{\pi s}} \times(-1)^{\frac{n+r}{2}}\left(\frac{n+r}{2}\right)! E^{-\frac{1}{2}}e^{-\frac{E}{s}}L^{(-1/2)}_{\frac{n+r}{2}}\left(\frac{E}{s}\right) (n+r\text{: even}). \tag{A2.2}$$

In the first expansion term of Eq. (A2.2), setting $n = 2m$ and using the relationship $\Gamma(n+1/2) = (2n)!\sqrt{\pi}/2^{2n}n!$, we can obtain the probability density expression (12).

Appendix 3. Derivation of Expansion Expression Eq. (17)

Using the relationship:

$$\frac{1}{a-x} = \sum_{n=0}^{\infty}\frac{x^n}{a^{n+1}}, \tag{A3.1}$$

equation (16) can be rewritten as:

$$G_{\mathrm{II}}(s)=\frac{1}{(1+T_1 s)(1+T_2 s)-T_1 T_2 s^2 e^{-\tau s}}$$

$$=\frac{1}{(1+T_1 s)(1+T_2 s)\left[1-\frac{T_1 T_2 s^2}{(1+Tsx)(1+T_2 s)}e^{-\tau s}\right]}$$

$$=\frac{1}{(1+T_1 s)(1+T_2 s)}\sum_{n=0}^{\infty}\left[\frac{T_1 T_2 s^2}{(1+T_1 s)(1+T_2 s)}e^{-\tau s}\right]^n. \tag{A3.2}$$

After setting $T_2=T_1+\varDelta$ in the above equation, and making use of the formula:

$$(a+x)^k=\sum_{r=0}^{\infty}\frac{k(k-1)..(k-r+1)}{r!}a^{k-r}x^r, \tag{A3.3}$$

we have:

$$G_{II}(s)=\sum_{n=0}^{\infty}\frac{T_1^n.(T_1+\varDelta)^n s^{2n}}{(1+T_1 s)^{n+1}(1+T_1 s+\varDelta s)^{n+1}}e^{-n\tau s}$$

$$=\sum_{n=0}^{\infty}\frac{T_1^n(T_1+\varDelta)^n s^{2n}}{(1+T_1 s)^{2n+2}}\left[1+\frac{\varDelta s}{1+T_1 s}\right]^{-(n+1)}e^{-n\tau s}$$

$$=\sum_{n=0}^{\infty}\sum_{r=0}^{\infty}T_1^n(T_1+\varDelta)^n\frac{(-1)^r(n+1)(n+2)..(n+r)}{r!}\frac{\varDelta^r s^{2n+r}}{(1+T_1 s)^{2n+r+2}}e^{-n\tau s}$$

$$=\sum_{n=0}^{\infty}\sum_{r=0}^{\infty}\frac{1}{r!}T_1^{n+r}T_2^n\left(1-\frac{T_2}{T_1}\right)^r(n+1)(n+2)\dots(n+r)\frac{s^{2n+r}}{(1+T_1 s)^{2n+r+2}}e^{-n\tau s}$$

$$=\sum_{n=0}^{\infty}\sum_{r=0}^{\infty}\frac{1}{r!}\frac{T_2^n}{T_1^{n+2}}\left(1-\frac{T_2}{T_1}\right)^r(n+1)(n+2)\dots(n+r)\frac{s^{2n+r}}{\left(s+\frac{1}{T_1}\right)^{2n+r+2}}e^{-n\tau s}. \tag{A3.4}$$

Appendix 4. Evaluation of expansion coefficient and the number of expansion terms in Eq. (13)

First, the value of finite order K in Eq. (1) should be chosen (In our experiment related to Figs. 10 and 11, we have set $K=10$). The sampled weighting values b_i and $a_i (i=1, 2, \dots, K)$ are concretely calculated by using step response curves of the original and the altered systems as:

$$b_i=S_{\mathrm{I}}(t_i)-S_{\mathrm{I}}(t_{i-1}),\quad a_i=S_{\mathrm{II}}(t_i)-S_{\mathrm{II}}(t_{i-1}),\ t_i\overset{\Delta}{=}i\times T, \tag{A4.1}$$

where T is the sampling period (in the present experiment, T is set to 0.003 s.). On the other hand, the statistical properties related to the lower order cumulant functions on the input sound pressure fluctuation can be experimentally evaluated as follows:

$$
\begin{aligned}
&\varkappa_{X1}(0, 0, \ldots, 0, \overset{P_1}{1}, 0, \ldots, 0) = \langle X_{P_1} \rangle, \\
&\varkappa_{X2}(0, 0, \ldots, 0, \overset{P_1}{1}, 0, \ldots, 0, \overset{P_2}{1}, 0, \ldots, 0) = \langle (X_{P_1} - \langle X_{P_1} \rangle)(X_{P_2} - \langle X_{P_2} \rangle) \rangle, \\
&\varkappa_{X3}(0, 0, \ldots, 0, \overset{P_1}{1}, 0, \ldots, 0, \overset{P_2}{1}, 0, \ldots, 0, \overset{P_3}{1}, 0, \ldots, 0) \\
&\qquad = \langle (X_{P_1} - \langle X_{P_1} \rangle)(X_{P_2} - \langle X_{P_2} \rangle)(X_{P_3} - \langle X_{P_3} \rangle) \rangle, \\
&\cdots\cdots\cdots\cdots
\end{aligned}
\tag{A4.2}
$$

by using experimentally sampled data X_i $(i = 1, 2, \ldots)$ on the input sound pressure waves $X(t)$. Thus, substituting the concrete values of b_i, a_i $(i = 1, 2, \ldots, K)$ and Eq. (A4.2) into Eq. (5), the expansion coefficients, $B_r(r)$ $(r = 1, 2, \ldots)$, can be concretely calculated.

Next, let us discuss the decision problem on the number of expansion terms for the probability density expression (13). From the arbitrariness of the weighting functions, $h(t)$ and $W(t)$, for the original and the altered noise control systems, the probability density distribution and correlation functions of input signal $X(t)$, and the functional form of $P_y(y)$ in Eq. (4), we can not find generally and systematically any conclusion for the above decision problem only from the mathematical viewpoint. Principally, this problem must be considered by taking both sides on the theoretical convergency property of the expansion expression and the experimental reliability in an estimation of the expansion coefficients $B_r(r)$ (especially related to the higher order cumulants). In this paper, we have only pointed out in the conclusion section that this problem will have to be considered as future study. However, for the purpose of examining partly the legitimacy of the proposed expression Eq. (4), it can be shown in the following that the well-known Gram-Charlier A type series expansion [8] on the arbitrary type probability expression is contained as a special case in Eq. (4).

Let us especially choose the Gaussian distribution as an arbitrary probability density function, $P_y(z)$, in Eq. (4), as:

$$
P_y(z) = \frac{1}{\sqrt{2\pi}\,\sigma_z} e^{-\frac{(z-\mu_z)^2}{2\sigma_z^2}}, \quad (\mu_z = \langle z \rangle, \quad \sigma_z^2 = \langle (z - \langle z \rangle)^2 \rangle), \tag{A4.3}
$$

by setting $\varkappa_{y1}(1) = \varkappa_{z1}(1) \overset{\Delta}{=} \mu_z$, $\varkappa_{y2}(2) = \varkappa_{z2}(2) \overset{\Delta}{=} \sigma_z^2$ $(\varkappa_{yl}(l) = 0,\ l \geqslant 3)$. At this time, the probability density expression (4) can be easily rewritten as follows (see Eq. (7)):

$$
P_z(z) = \frac{1}{\sqrt{2\pi}\,\sigma_z} e^{-\frac{(z-\mu_z)^2}{2\sigma_z^2}} \sum_{r=0}^{\infty} C_r H_r\left(\frac{z-\mu_z}{\sigma_z}\right). \tag{A4.4}
$$

Hereupon, the expansion coefficient, C_r, is given as:

$$\begin{aligned} C_r &\triangleq \frac{1}{r!\,\sigma_z^r} B_r(r) = \frac{1}{r!\,\sigma_z^r} \frac{\partial^r}{\partial \theta^r} \exp\left\{ \sum_{l=1}^{\infty} \frac{1}{l!} \left[\varkappa_{zl}(l) - \varkappa_{yl}(l) \right] \theta^l \right\} \Bigg|_{\theta=0} \\ &= \frac{1}{r!\,\sigma_z^r} \frac{\partial^r}{\partial \theta^r} \left[\langle \exp(z\theta) \rangle \exp\left(-\mu_z \theta - \frac{1}{2}\sigma_z^2 \theta^2\right) \right] \Bigg|_{\theta=0} \\ &= \frac{1}{r!} \left\langle \frac{\partial^r}{\partial (\sigma_z \theta)^r} \exp\left(z\theta - \mu_z \theta - \frac{1}{2}\sigma_z^2 \theta^2 \right) \Bigg|_{\theta=0} \right\rangle \\ &= \frac{1}{r!} \left\langle \frac{\partial^r}{\partial t^r} \exp\left(\frac{z-\mu_z}{\sigma_z} t - \frac{1}{2} t^2 \right) \Bigg|_{t=0} \right\rangle = \frac{1}{r!} \left\langle H_r\left(\frac{z-\mu_z}{\sigma_z} \right) \right\rangle . \end{aligned} \tag{A4.5}$$

Here, the relationship between the exponential function and the Hermite polynomials:

$$H_r(\xi) = \frac{\partial^r}{\partial t^r} \exp\left(\xi t - \frac{1}{2} t^2 \right) \Bigg|_{t=0} \tag{A4.6}$$

has been used. Furthermore, we can easily find $C_0 = 1$, $C_1 = C_2 = 0$ from $\mu_z = \langle Z \rangle$ and $\sigma_z^2 = \langle (Z - \langle Z \rangle)^2 \rangle$. Thus, equation (A4.4) is reduced accurately to the well-known Gram-Charlier A type series expansion expression, as follows:

$$P_z(z) = \frac{1}{\sqrt{2\pi}\,\sigma_z} e^{-\frac{(z-\mu_z)^2}{2\sigma_z^2}} \left\{ 1 + \sum_{r=3}^{\infty} \frac{1}{r!} \left\langle H_r\left(\frac{z-\mu}{\sigma_z} \right) \right\rangle H_r\left(\frac{z-\mu_z}{\sigma_z} \right) \right\}. \tag{A4.7}$$

Finally, in all figures in this paper, it should be noticed that the tendency of converging to a certain proper cumulative probability distribution curve can be obviously found out and so can be examined too by finding such a saturation tendeny from an experimental viewpoint, as pointed out in section 3.

Appendix 5. Engineering background of Eq. (20)

As is well-known, by letting $E(\theta_i)$ be a transmitted output energy component from the input sound energy with a specific incident angle θ_i, the total energy E of the transmitted sound wave with random incident angles can be expressed from a deterministic viewpoint as follows:

$$E = \sum_i E(\theta_i) P(\theta_i) d\theta \to \int_0^{\pi/2} E(\theta) P(\theta) d\theta, \tag{A5.1}$$

$$P(\theta) = \frac{\cos\theta \sin\theta}{\int_0^{\pi/2} \cos\theta \sin\theta \, d\theta} = \sin 2\theta \quad (0 \leqslant \theta \leqslant \pi/2),$$

where $p(\theta_i)\,d\theta$ denotes the rate of energy component between θ_i and $\theta_i + d\theta$. If the above viewpoint is generalized for the statistical evaluation of the insulation system by newly introducing any kind of probabilistic viewpoint, equation (A5.1) must be first satisfied in the averaged form of energy fluctuation:

$$\langle E \rangle = \sum_i \langle E|\theta_i \rangle P(\theta_i) d\theta \to \int_0^{\pi/2} \langle E|\theta \rangle P(\theta)\, d\theta, \tag{A5.2}$$

by use of the well-known additive property of mean energy. Hereupon, $\langle E|\theta \rangle$ is the conditional expectation of energy with a fixed incident angle θ. For the purpose of establishing some kind of probabilistic law for the total energy fluctuation of the transmitted sound wave with random incident angles, naturally we have to find it in a unified probabilistic expression form of generalizing the above equation (A5.2) Under the above background, equation (20) has been rationally introduced as a unified probabilistic law of transmission with random incident angles.

Appendix 6. An example of applying the present theory to an actual design problem

As an application example of the present statistical method, let us consider a statistical prediction method for the sound insulation effect of single and double walls. Let E_α and E'_α, respectively, be the $(100 - \alpha)$ percentage point of the output sound energy distribution of the original noise control system with $h(t)$ and that of the changed system with $W(t)$. At this time, an improvement of the evaluation index L_α can be quantitatively given as follows:

$$\Delta L_\alpha \overset{\Delta}{=} L_\alpha - L'_\alpha = 10 \log_{10} \frac{E_\alpha}{E'_\alpha}, \tag{A6.1}$$

where E'_α is calculated from the first expansion term of Eq. (12). Now, let us consider a prediction problem of improving the above L_α in a typical case when a sound insulation wall such as single or double walls is newly constructed. In this case, the present theory can be directly used by setting $h(t) = \delta(t)$ (Dirac's δ-function). We have employed the following two kinds of walls as examples of noise insulation systems:

Case A: A single wall with surface mass $m = 3.22$ ($\mathrm{Kg/m^2}$).

Case B: A double wall with two panels of surface mass 3.22 $\mathrm{Kg/m^2}$ (i.e., $m_1 = m_2 = 3.22$ $\mathrm{Kg/m^2}$) and width of air gap $d = 0.05$ m.

Furthermore, the Gaussian random numbers have been used as the standard input sound pressure fluctuation to the noise control system, in the present simulation experiment.

Figure A6.1 shows a comparison between theoretically predicted values and

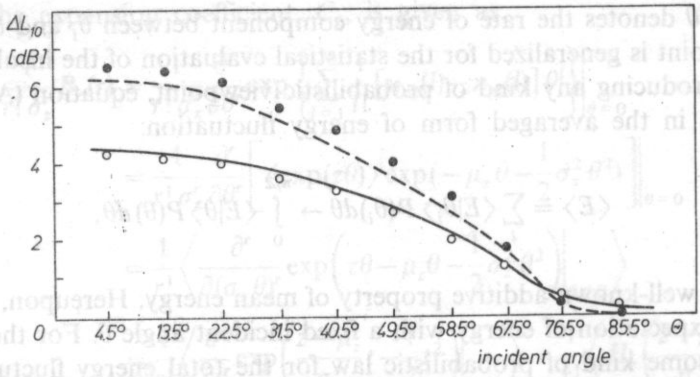


Fig. A6.1. A comparison between theoretically predicted values and experimentally sampled points for ΔL_{10}. Experimentally sampled points are marked by (o) and theoretical values are shown by (——) for Case A. Experimentally sampled points are marked by (•) and theoretical values are shown by (– – – –) for Case B

experimentally sampled points for ΔL_{10} with various values of the incident angle θ, where it is very difficult to measure by using only the actual standard experiment method. It must be observed that the prediction accuracy remains within the range less than a practically permissible error 1 dB.

References

[1] M. OHTA, S. YAMAGUCHI and S. HIROMITSU, *A general theory on probabilistic evaluation of noise and vibration control system and its application to a single wall. A methodological study on statistical evaluation of noise and vibration control systems, Part* 1, J. Acoustical Society of Japan, **34**, 6, 333–341 (1978).

[2] M. OHTA, S. YAMAGUCHI and S. HIROMITSU, *A unified expression for the multivariate joint probability density function of the output fluctuation of an arbitrary linear vibratory system with arbitrary random excitation*, Journal of Sound and Vibration, **56**, 2, 229–241 (1978).

[3] M. OHTA, K. HATAKEYAMA, S. HIROMITSU and S. YAMAGUCHI, *A unified study on the output probability distribution of arbitrary linear vibratory systems with arbitrary random excitation*, Journal of Sound and Vibration, **43**, 4, 693–711 (1976).

[4] H. AKAKIKE, *Fittig autoregressive models for prediction*, Ann. Inst. Statist. Math., **21**, 243–247 (1969).

[5] H. AKAKIKE, *Power spectrum estimation through autoregressive model fitting*, Ann. Inst. Statist. Math., **21**, 407–419 (1969).

[6] G. E. P. BOX and G. M. JENKINS, *Time series analysis forecasting and control*, Holden-Day 1970

[7] M. OHTA, S. YAMAGUCHI and S. HIROMITSU, *A unified study on the multivariate joint probability expression and its linear transitional property for the state variables of stochastic noise environmental system*, Trans. Institute of Electronics and Comm. Eng. of Japan, **60**–A, 9, 844–851 (1977).

[8] S. S. WILKS, *Mathematical statistics*, Princeton University Press, New York, 1963.

[9] M. OHTA, T. KOIZUMI, *General statistical treatment of the response of a nonlinear rectifying device to a stationary random input*, IEEE Trans. on Information Theory, **IT-14**, 4, 595–599 (1968).

[10] M. OHTA, T. KOIZUMI and S. YAMAGUCHI, *A study of intensity fluctuation for sampled random signals having some arbitrary correlation and distribution functions and its digital simulation*, Proceedings of the 18th Japan National Congress for Applied Mechanics, 39–49 (1968).

[11] M. Abramowitz, I. A. Stegun, *Handbook of mathematical functions with formulas, graphs and mathematical tables*, Superintendent of documents, U. S. Government Printing Office, Washington 1964, p. 782

[12] A. London, *Transmission of reverberant sound through single walls*, National Bureau of Standards, 42, June 605–615 (1949).

[13] A. London, *Transmission of reverberant sound through double walls*, J. Acoust. Soc. Amer. **22**, 2, 270–279 (1950).

[14] S. Yamaguchi, M. Ohta, H. Iwashige and S. Hiromitsu, *A methodological study on statistical evaluation of noise and vibration control system* V–*A probabilistic evaluation of transmitted sound wave based on the system change of sound insulation*, Report to the Spring Meeting of the Acoustical Society of Japan, 427–428 (1978).

[15] D. Middleton, *An introduction of statistical communication theory*, Mc Graw Hill Book Company, New York, 1960 p. 141–144

[16] M. Rosenblatt, J. W. Van Ness, *Estimation of the bispectrum*, Ann. Math. Stat., **36**, 1120–1136 (1965).

[17] E. T. Whittaker, G. N. Watson, *Modern analysis*, University Press, Cambridge, 1935.

Received on June 1, 1987; revised version on September 25, 1988.

ARCHIVES OF ACOUSTICS
13, 3-4, 305-309 (1988)

CONTINUITY EFFECTS IN THE PERCEPTION OF SOUNDS[1]

REINIER PLOMP

Institute for Perception TNO, Soesterberg, and Faculty of Medicine, Free University, Amsterdam
The Netherlands

The paper presents the phenomenon of continuity in the preception of sounds in condition of uncomplete information. On the basis of examples of perception of sounds with gliding frequency changes interrupted by noise burst, modulated tones "damaged"in similar way, it is concluded that under specific conditions the hearing system is able to restore the original sound patterns. Other examples of continuity are also given, especially, of alternating tone sequences which demonstrate that perception is not a possive process but a highly active one.

W pracy poruszono problem ciągłości i percepcji dźwięków w przypadku braków w dopływie informacji. Podano przykłady od efektów ciągłości przy percepcji zmian częstotliwościowych przerywanych wysokopasmowym szumem do ciągów impulsów tonu o różnym czasie trwania maskowanym pobodźcowo szumem. W ogólności ciągłość jest zachowana, jeśli widmo dźwięku, który ma być słyszany jako ciągły, jest pokryte widmem szumu. Przedstawione przykłady ilustrują zdolność „restoracyjną" słuchu i wskazują na rolę wyższych ośrodków neuronowych w realizacji tej funkcji. Wynika stąd, że organ słuchu demonstruje właściwości percepcyjne wskazujące na działanie nie tylko biernego , lecz również aktywnego procesu przetwarzania i „uzupełniania" informacji.

It seems to me that there is a large gap between our psycho-acoustical knowledge and that of the perception of music. We know much more about the ear's frequency resolution, the pitch-extraction mechanism, etc., than 30 years ago, but there are still musically highly relevant phenomena which need more attention from the researchers. This holds particularly for the temporal factor in tone perception. We are able to hear the individual voices in a musical performance without realizing that these voices were seemingly intermingled inextricably in the air.

This capacity of "pattern recognition" is not specific, but is a general property of our sense organs. It has had obtained much more attention in vision than in hearing. Figure 1 illustrates how visual elements are ordered in our perception on the basis of their properties and spatial relations. In Gestalt psychology laws have been

---

[1]The paper was presented at the Summer Workshop on Psychoacoustic of Music, Jabłonna/near Warsaw 5-11 July 1982.

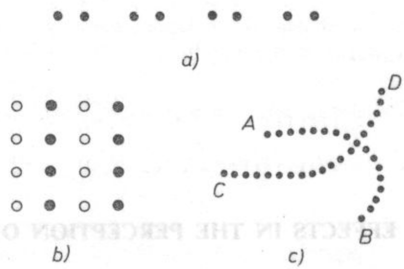


Fig. 1. Illustrations of the Gestalt principles of (a) proximity, (b) similarity and (c) good continuation (from Deutsch [4])

formulated which can explain the ordering or grouping of complex stimuli: a = proximity, b = similarity, c = good continuation, whereas d = common fate, meaning that elements which move in the same direction are perceived together (not illustrated).

A very interesting example is the case of overlapping objects. In Fig. 2a we are unable to see more than a collection of differently shaped elements, but by introducing another object (Fig. 2b) the B's become visible. This example is highly relevant in auditory perception. Due to other, masking, sounds, we may hear only fragments of a voice, but we still perceive a continuous voice signal. Apparently, the ear is able to restore that voice.

This effect can be illustrated strikingly by making use of alternating presentation of tone bursts and noise bursts. Figure 3 represents these signals. Tone bursts of a continuously varying frequency from 500 to 2000 Hz are alternated with a band of noise between 900 and 1100 Hz. As long as the frequency of the tone is below about 900 Hz or above about 1100 Hz, we hear a series of tone bursts which increase step by step in pitch. Between 900 and 1100 Hz, however, the auditory sensation is quite

Fig. 2a. Seemingly unrelated elements (from Bregman [1])
Fig. 2b. The same elements as shown in Fig. 2a, but now seen as B's partly covered by a black object (from Bregman [1])

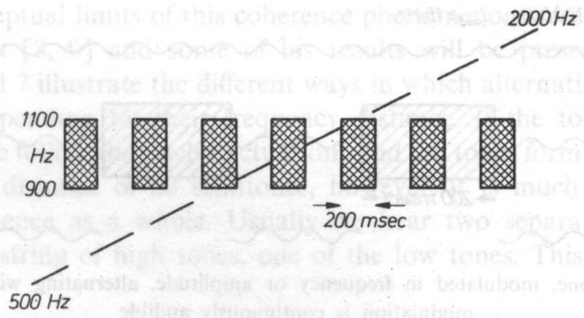


Fig. 3. Tone bursts alternating with noise bursts (noise band 900–1100 Hz). The tone bursts are heard separately for frequencies outside the noise band, but as a continuous tone for frequencies covered by this band

different: The tone is heard without interruptions and with a *continuously* increasing pitch. The effect of the noise bursts is quite different from that of silent gaps between the tone bursts: The noise bursts *connect* the tone bursts into one continuous tone.

The continuity effect seems to have been discovered independently by a number of investigators (MILLER and LICKLIDER [7], THURLOW [12], WARREN et AL., [13], HOUTGAST [5]). It only occurs when the sound spectrum of the signal to be heard continuously is completely "covered" by the spectrum of the noise (for a more correct formulation, see WARREN et AL., [13], HOUTGAST [5]). The maximum signal level at which it still sounds continuously, the pulsation threshold, has proved to be a powerful new measure in auditory research (e.g. HOUTGAST [6]).

An interesting point is: Up to what duration of the noise bursts is the impression of continuity of a tone maintained? This question was investigated by DANNENBRING [2] with alternately rising and falling, or zigzag, frequency glides centred at 1000 Hz. Halfway along a frequency glide taking 2 sec a noise burst of up to 450 milliseconds may be inserted, before the perceived continuity has gone. This value reduces to about 350 milliseconds for a steady-state tone, but increases to 650 milliseconds if the noise bursts are located at the extreme values of the frequency zigzags. These results suggest that the continuity effect of a tone is intensified rather than reduced by varying its frequency.

A tone does not need to be constant in order for the bursts to be heard as a continuous tone. Figure 4 illustrates the case of a frequency-modulated tone alternating with either noise bursts or silent gaps. The difference is large: The modulation is not very distinct when the tone bursts alternate with silent intervals, but becomes very prominent and *continuously* audible when they alternate with noise. This also holds for amplitude-modulated tones.

Repetition of the same signal is not essential for the continuity effect. When a tone burst of 40 milliseconds is immediately followed by a 300 millisecond noise burst, we hear a much longer tone burst, as if it were present during the noise. This can be demonstrated nicely by playing a scale of eight tones, half of which last 300

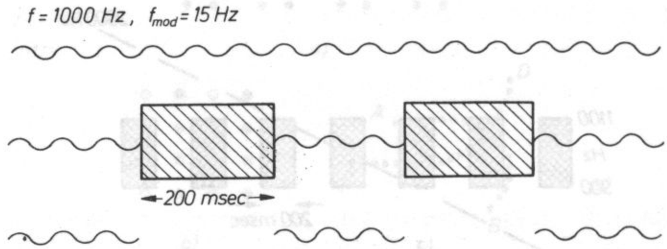


Fig 4. Bursts of a tone, modulated in frequency or amplitude, alternating with noise bursts. The modulation is continuously audible

milliseconds, the other half lasting 40 milliseconds, followed by 260 milliseconds of noise, see Fig. 5. We hear a scale of equal-duration tones, with four interfering noise bursts, not the alternation of long and short tone bursts actually presented. The implications of this effect for the perception of simultaneous notes in music were studied by RASCH [10, 11].

We may conclude from these examples that the hearing system is able to restore sound patterns if the duration of the masked portions does not exceed a few hundreds of milliseconds. The finding that the continuity effect is also operative for sounds varying substantially in time indicates that the phenomenon cannot be explained by peripheral auditory processes, but that central processes are involved. The hearing system tries continuously to restore partly "damaged" sound signals in order to give the most probable representation of our acoustical environment. In this restoration process both more innate Gestalt laws and recognition on the basis of earlier experience play a role.

Another example of continuity is to be found in tone sequences. Successive tones give a much stronger melodic coherence if their frequency distance is small than if it

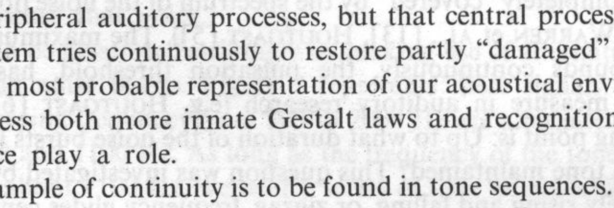


Fig. Scale of eight tones, half of them shortened from 300 to 40 milliseconds, with noise added. We hear a scale of equal–duration tones is large. The perceptual limits of this coherence phenomenon were recently explored by VAN NOORDEN [8, 9] and some of his results will be presented here.

Figures 6 and 7 illustrate the different ways in which alternating tone sequences are perceived, depending on their frequency distance. If the tones are only one semitone apart, we follow the pitch fluctuations and the tones form a coherent whole. For a frequency distance of 20 semitones, however, it is much more difficult to perceive the sequence as a whole. Usually we hear two separate sequences, one consisting of the string of high tones, one of the low tones. This is an example of fission.

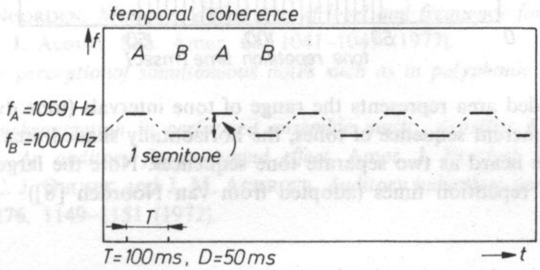


Fig. 6. Alternation of tones with a frequency difference of one semitone. The tone are perceived as a coherent sequence of tones (from VAN NOORDEN [8])

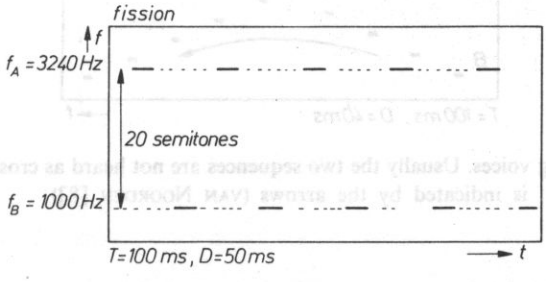


Fig. 7. Alternation of tones with a frequency difference of 20 semi tones. Usually, the tones are perceived as two separate sequences, one with a low, the other with a high pitch (from VAN NOORDEN [8])

Especially with slow tone sequences, there is a range of tone intervals where either temporal coherence or fission can be heard at will. This overlap is given in Fig. 8.

Counterpoint theory advises against crossing voices, see Fig. 9. It is indeed difficult to follow a tone sequence through a cross-over; the arrows indicate what is normally heard. This grouping tendency is so strong that even if the subsequent tones are alternately presented in the right ear and the left ear, we still hear a lower and higher melody (DEUTSCH [3]).

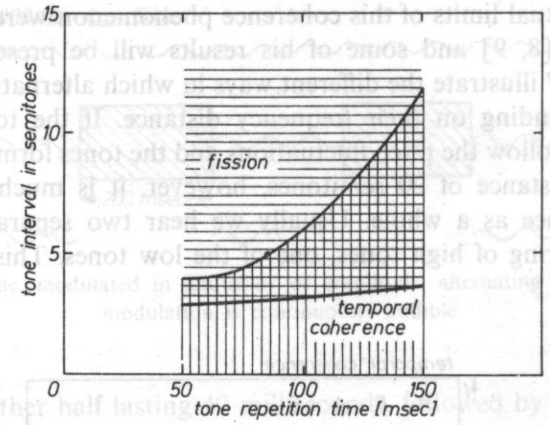


Fig. 8. The vertically shaded area represents the range of tone intervals (tone duration 40 milliseconds) that can be heard as a coherent sequence of tones, the horizontally shaded area represents the range of tone intervals that can be heard as two separate tone sequences. Note the large overlap for the longer repetition times (adopted from van Noorden [8])

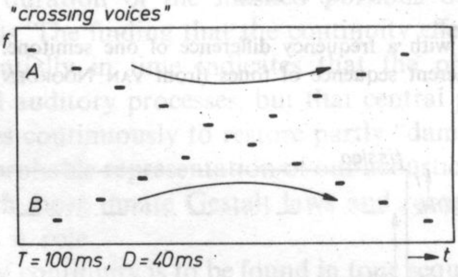


Fig. 9. The case of crossing voices. Usually the two sequences are not heard as crossing each other, but as is indicated by the arrows (VAN NOORDEN [8])

The way in which these tones are grouped in hearing demonstrates that perception is not a passive process but a highly active one, in which the elements are ordered according to patterns apparently preferred by the system. These patterns can be considered to represent the "best guess" of how the acoustical outer world is organized. More research will be required to get a better insight into how these guesses are made and by what factors they are determined.

References

[1] A. S. BREGMAN, *Asking the "what for" question in auditory perception.* In: Perceptual organization, [ed.] M. Kubovy and J. R. Pomerantz, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1981 pp. 99–118.

[2] G. L. Dannenbring, *Perceived auditory continuity with gliding frequency changes*, Doctoral thesis, McGill University, Montreal 1974.

[3] D. Deutsch, *Two-channel listening to musical scales*, J. Acoust. Soc. Amer. **57**, 1156–1160 (1975).

[4] D. Deutsch, *Grouping mechanisms in music*, In: The psychology of music, [ed.] D. Deutsch, Academic Press, New York 1982, pp. 99–134.

[5] T. Houtgast, *Psychophysical evidence for lateral inhibition in hearing*, J. Acoust. Soc. Amer. **51**, 1885–1894 (1972).

[6] T. Houtgast, *Lateral suppression in hearing*, Doctoral thesis, Free University, Amsterdam 1974.

[7] G. A. Miller, J. C. R. Licklider, *The intelligibility of interrupted speech*, J. Acoust. Soc. Amer. **22**, 167–173 (1950).

[8] L. P. A. S. van Noorden, *Temporal coherence in the perception of tone sequences*, Doctoral thesis, Technological University Endhoven 1975.

[9] L. P. A. S. van Noorden, *Minimum differences of level and frequency for perceptual fission of tone sequences ABAB*, J. Acoust. Soc. Amer. **61**, 1041–1045 (1977).

[10] R. A. Rasch, *The perceptionof simultaneous notes such as in polyphonic music*, Acustica **40**, 21–33 (1978).

[11] R. A. Rasch, *Synchronization in performed ensemble music, Acustica* **43**, 121–131 (1979).

[12] W. R. Thurlow, *An auditory figure-ground effect*, Amer. J. Psychol. **70**, 653–654 (1957).

[13] R. M. Warren, C. J. Obusek and J. M. Ackroff, *Auditory induction: perceptual synthesis of absent sounds*, Science **176**, 1149–1151 (1972).

Received on September 13, 1988.

ARCHIVES OF ACOUSTICS
13, 3-4, 313-322 (1988)

THE PITCH CODE IN MUSIC AS COMPARED WITH THE NATURAL LANGUAGE CODE

ANDRZEJ RAKOWSKI

Chopin Academy of Music
(00-368 Warszawa ul. Okólnik 2)

Comparisons are drawn between the formal features of language and music. Basic phonological universals of language and music are compared. Attention is drawn to specific properties of sound material and the general features of human perception mechanisms in the process of aural communication. The general common rules are described governing the organisation of communication codes in language and music.

Dokonano porównania funkcji i cech formalnych języka i muzyki. Zwrócono uwagę na uniwersalny charakter podstawowych zasad tworzenia kodów komunikacyjnych mowy i muzyki. Ukazanie przez autora psycholingwistycznych analogii pomiędzy mową i muzyką przyczynia się do lepszego zrozumienia praw rządzących doborem i organizacją materiału dźwiękowego w muzyce.

1. Introduction

The aim of the present paper is a comparison of language and music as far as their basic communicational functions are concerned. The comparison results in conclusions concerning some features occurring generally in both fields. The considerations are scarce in the detailed part due to the fact that among other limitations, the analysis of musical phenomena has been limited to the way of organizing pitch without mentioning organization in time, loudness and timbre. Also, only those organization systems are discussed which employ the discrete division of pitch continuum, i.e. musical interval systems. Finally, in discussing these features only the basic problems of organizing the material are considered remaining within the field of music phonology. The detailed analysis concerning phonological universals in music is preceded by more general discussion of the relations between language and music.

2. The comparison of functions and formal features of language and music

Before we start to discuss the functions and formal features of both fields compared, let us dwell for a moment on the problem of definitions of language and consequences they may lead to in respect to analogies in question.

Let us show it in the following example; if we define language broadly as "a system of communication or the means by which humans exchange meanings and values" RUWET [23], then its main universal feature (i.e. being "a system of communication etc...") clearly overlaps with those of some other selected systems of communication. On the basis of this definition we can speak of music as a "language", however, we cannot use the expression "the language of ants". If we adopt any other definition of language, the situation may look quite differently.

2.1. Language

According to the trends adopted by various authors, the following definition of a spoken language may be formulated: Language is a device for *transmitting* and *processing* information, using signs which are composed of elementary meaningless units produced by human articularly organs.

Besides the basic function of processing and transmitting information, three other important functions of language can be listed:

Representation (substituting phenomena of the environment by signs);

Expression (expressing the feelings of a speaker); and

Impression (making an impression on listeners and inducing them to take up attitudes or to undertake action).

Finally, three important formal characteristics of language can be mentioned:

Arbitrariness of language signs, held by DE SAUSSURE [24] to be a basic property of language, directly connected with its function of representation. Some doubts have been raised about the phonological aspect of language arbitrariness (JAKOBSON [11]). It should also be remembered that the conventional character of language is in some cases replaced by a more natural one (onomatopeic phenomena). Therefore perhaps it is more correct to understand arbitrariness as an extremely important, but not fully universal, feature of a language sign.

Double articulation of language. We can specify two broad classes or types of articularly systems effective in the process of conveying semantic information. The first type may be described as "systems with single articulation". Here information is transmitted by a series of single, unanalysable units, each having its own meaning. These systems are mostly natural, (i.e. their signs have a direct, biological or physical relation to represented phenomena). Typical signs of this kind are: cries of pain, caresses, raising, voice in anger, the speed and timing of movements etc. Natural signs of single articulatory systems are, to a large degree, universally understood among human societies at very divergent stages of development. They are also more or less universal in the communication of humans with animals and/or between animals. However, many signs of single articulatory systems are purely conventional. A good example of such a conventional sign is nodding the head in approval (in most societies) or in negation (in some others).

In contrast to the situation described above, double articulatory systems use signs which are constructed from a limited number of smaller, meaningless units. The human language, while incorporating some features of natural articulation, may be taken as a representative of the conventional, double-articulatory system. This is perhaps its main and most important formal characteristic. The structure of language as a combinatory-type device utilizing a code of a limited number of elements and rules for combining them, stems from the fundamental properties of the human mind and is effective in overcoming the limitations of memory.

Generative properties of language. This extremely important aspect of language has been forcefully demonstrated by CHOMSKY [4]. Its essence is the innate ability of the users of language to generate an infinite number of new, meaningful utterances. The generative properties of language are perhaps the most "human" feature of this system of communication and best reflect the power of human creative thinking (ROEDERER [21]).

2.2. *Music*

The basic assumption accepted in this paper was that both natural language and music take part in the process of human communication. This assumption, however, may seem to some degree controversial. Though it is obvious that language plays an essential role in communication by transmitting semantic information the question immediately arises: *what is being transmitted in music*? Music is constructed differently than language and we cannot say that its elements are simply and directly signs representing elements of the outer world, as the elements of language are. Then, it may seem doubtful whether any important similarity between language and music really exists.

The problem requires serious consideration. It cannot be considered at length within this paper, however some important facts should be mentioned. Similarity between language and music within the domain of communication stems from the fact that differences in the role played by them in that domain are quantitative rather than qualitative in nature. Let us first consider language. As it has been said its elements are signs representing the reality outside the language itself. They are used for transmitting information about the outer world. However, though most important, it is not the only information that language can transmit. The other information considers not the outer world but the language itself: the form in which its elements have been organised, the logic and beauty of this form, even the beauty of the elements themselves. The above kind of information is of great importance in poetry, particularly in modern poetry, but it is also present in language outside poetry. We may call it "internal aesthetic information".

or less universal in the communication of humans with animals and/or between animals. However, many signs of single articulatory systems are purely conventional. A good example of such a conventional sign is nodding the head in approval (in most societies) or in negation (in some others).

In contrast to the situation described above, double articulatory systems use signs which are constructed from a limited number of smaller, meaningless units. The human language, while incorporating some features of natural articulation, may be taken as a representative of the conventional, double-articulatory system. This is perhaps its main and most important formal characteristic. The structure of language as a combinatory-type device utilizing a code of a limited number of elements and rules for combining them, stems from the fundamental properties of the human mind and is effective in overcoming the limitations of memory.

Generative properties of language. This extremely important aspect of language has been forcefully demonstrated by CHOMSKY [4]. Its essence is the innate ability of the users of language to generate an infinite number of new, meaningful utterances. The generative properties of language are perhaps the most "human" feature of this system of communication and best reflect the power of human creative thinking (ROEDERER [21]).

2.2. *Music*

The basic assumption accepted in this paper was that both natural language and music take part in the process of human communication. This assumption, however, may seem to some degree controversial. Though it is obvious that language plays an essential role in communication by transmitting semantic information the question immediately arises: *what is being transmitted in music*? Music is constructed differently than language and we cannot say that its elements are simply and directly signs representing elements of the outer world, as the elements of language are. Then, it may seem doubtful whether any important similarity between language and music really exists.

The problem requires serious consideration. It cannot be considered at length within this paper, however some important facts should be mentioned. Similarity between language and music within the domain of communication stems from the fact that differences in the role played by them in that domain are quantitative rather than qualitative in nature. Let us first consider language. As it has been said its elements are signs representing the reality outside the language itself. They are used for transmitting information about the outer world. However, though most important, it is not the only information that language can transmit. The other information considers not the outer world but the language itself: the form in which its elements have been organised, the logic and beauty of this form, even the beauty of the elements themselves. The above kind of information is of great importance in poetry, particularly in modern poetry, but it is also present in language outside poetry. We may call it "internal aesthetic information".

We can easily find analogous situation in music, though both the surface structure of events and the relative importance of the two processes are quite different. The counterpart of the "outer" semantic language information is vague, though extremely important system of musical connotations. These connotations, mostly of subconsious and emotional nature play important part in music appreciation and apprehension. Music brings memories, evokes feelings, generates emotions; in some way music tells about or originates phenomena that are outside it, using structures that are mostly conventional and partly natural. The role of music in that process may be understood as that of a system of signs used to convey information from sender (composer, singer, instrumentalist) to receiver (concert listener).

But that is not all, and, as some say, it is not the most important part of information contained in the musical message. Music is the art which comes into being through the organisation of sound structures. The imposition of this organisation on sound is conditioned by highly specific properties of human mind (probably related to those which laid the foundations for the development of language). By structuring sound, creatively inclined individuals acquire satisfaction through feelings of "beauty" and "perfection". As this satisfaction is intended to be shared by listeners, it is necessary to transmit information about the structure imposed on sound. And this "inner" information on the structure of a musical message contributes to the essential, aesthetic value of music. Similarly as mentioned earlier, secondary information contained in language, it may also be called "internal aesthetic information".

Now, if the transmission of information about the structure imposed on sound has to be efficient, several conditions must be fulfilled, very similar to those which govern the transmission of information in language. First, there must be a code known, or at least to some degree familiar to both the sender and the receiver of the information. This code may be arbitrary, or at least partly arbitrary, with limitations imposed on arbitrariness by the properties of the sound source, the auditory discrimination functions and the capacity of human memory. Then, over some degree of complexity, the code should have a permutational (hierarchical) structure. Some of these and other detailed conditions imposed on the pitch code of music by physics and physiology will be discussed later as "basic phonological universals of musical scales". Before that let us compare some of the functions and formal characteristics of language and music to find which of them may be understood as universal in both systems and which are typical of only one of them.

As we have shown, the first of two main functions of language, that of transmitting information, may be considered universal for both language and music. However, the second one, *mental processing* of information, facilitated by the use of language, cannot be attributed to music, because elements of music are derived of a direct representative function in relation to other phenomena. The mental processes of music composition and appreciation are of specific character and cannot be directly compared to those activated by language.

Though not capable of *representation* (except in marginal cases) music is particularly well suited to fulfill two other functions characteristic of spoken language; those of *expression* and of *impression.* According to popular opinion, the function of expression in music is more important than any other function. This is directly related to strong emotional connotations of music. Thus, the ability to express the feelings of the sender and to activate the mind of the receiver, may be considered a common universal of language and music.

Arbitrariness in selecting the details of a code for transmitting information is characteristic of both speech and music. (Limitations imposed upon this feature at the phonological level will be discussed later).

Double articulation, as commonly understood in language, cannot be attributed to music because the latter is deprived of meaningful units and conveys no semantic information. However, another important formal feature of similar character should be mentioned at this point. This feature is universal for all highly developed communicational systems, language and music included. It may be called the *permutational structure*. Following the rules of this structure, smaller units that are limited in number may be used to compose a more numerous, or even unlimited set of larger units. In music, as in language, the permutational structure may be observed on both phonological and syntactic levels.

In spoken language small units are combined to form a vocabulary of signs and signs may be composed into a practically unlimited number of meaningful communications. In music, limited sets of frequency and time ratios are used as the fundamental material for the creation of melodic structures. The combinatory, permutational character of both language and music is their important common feature.

There is another problem worth considering. It illustrates some aspects of the important analogies between language and music, and shows where "creative thinking" in music is really active. In his description of the specific features of language ("langage") DE SAUSSURE [24] explained how virtually all individual acts of speech ("parole") raise the possibility of interferring with established language rules. A single utterance, understandable but not exactly following the rules of a code, may initiate a slight change of the rules, which then, in turn, spreads out through imitations and finally becomes a fixed part of the standard rules itself. As a result, detectable changes can appear in a living language after only a few decades of its use, and after some hundreds of years the language may undergo very serious transformations.

There is an important analogy to this process in music. A musical composition produced in a definite style, (i.e. according to definite rules) combines some elements which are purely schematic or conventional and some which are entirely new (sometimes even breaking the established rules and thereby not accepted by more conservative listeners). The proportion of new elements in a given work is often taken as one of the measures of the talent of the composer and/or of the value of the composition itself. The new formal "inventions" found in masterpieces of great composers enter into the basic set of rules for musical composition and form definite changes in the existing code.

Important conclusions may be drawn from the analogy described above. As far as the formal communicational aspects of language and music are concerned, each new utterance in language corresponds to the creation of a new composition in music. This analogy is surprising! Perhaps it would be more easily accepted if a comparison were made between written musical compositions and works of literature. However, the above analogy correctly illustrates the mechanisms active in the temporal transformation of both language and music.

3. The basic phonological universals of musical scales

In some linguistic literature one may find the opinion that languages of the world differ in unpredictable ways and that in various languages, except for the fact of their double articulation, "everything may be different". However, such an opinion is not shared by the majority of linguists. HALLE [6] points to the invariance of the fundamental set of distinctive features in speech. HOCKETT [8] states that similarities of all known phonological systems of the world are much greater than could possibly be predicted from the physiological limitations of speech organs.

The above statements are reconciliations of the evident fact that all natural languages of the world are constructed according to some universal rules imposed by nature. Three groups of factors seem to be limiting the arbitrary character of language at the phonological level. They are related respectively, to the anatomy and physiology of speech organs (source of sound) to the discriminating power of the ear and to specific properties of the human brain (particularly to the accuracy of memory).

Having this in mind let us now look at the basic phonological system of music; that of pitch organisation in the form of musical scales. The universal features, similarly to those of language, are determined here by three elements:

— The sound source (physical properties of musical instruments).
— The analyser (frequency selectivity of the ear).
— The central processor effective in the formation of the code (typical features of a human mind; specifically, properties of the auditory memory).

3.1. The source. Sound material

The physical properties of musical instruments imposed essential limitations of a seemingly "universal" character upon the construction of a musical code long before these limitations could have been scientifically explained. Their manifestations were either purely spontaneous (as in the acquisition of simple scales based on fifths) or highly sophisticated (as in late Pythagorean or Arabic mathematical systems).

However, the fundamental question still remains: is the construction of musical scales really natural and universal, or rather is it conventional and based on arbitrary assumptions?

HELMHOLTZ [7], who first approached this problem from the scientific point of view, strongly emphasized the concept of a natural basis for the organization of pitch material in music. In his opinion, if physical properties of musical tones were such that between their component frequencies definite harmonic relations were preserved, then these harmonic relations constituted a natural basis for all frequency relations in music. Therefore, according to Helmholtz, the musical scale should employ only "natural" intervals with frequency ratios 2:1, 3:2, 4:3, etc. The immediate consequence of this assumption is the prominence of the harmonic interval system in music with its "natural" superiority over all other musical systems and the condemnation (as "unnatural") of the equally tempered scale.

There are several arguments in favour of the above mentioned "naturalness". The auditory sensation created by two simultaneously sounding complex tones, whose intervalic relation changes continuously from the unison to the octave, contains several distinct salient points. They occur at frequency ratios 1:1, 5:6, 4:5, 3:4, 2:3, 3:5 and 1:2 and correspond to the minima of acoustic beats between the harmonics of two component complex tones. The existence of such perceptually salient points is the usual basis for the formation of categories, for which these selected points can be used as cognitive prototypes (ROSH [22]). It is, therefore, easy to explain why the musical system of a Western culture is based primarily on intervals with simple or almost simple frequency ratios. There are, however, some restrictions which may spoil the picture of such a "natural" system as being universal in music.

As may be obvious from the above description, the appearance of salient points at natural frequency ratios along the interval magnitude continuum requires that two conditions be fulfilled. First, the tones must have a harmonic spectrum; second, they must be heard simultaneously. However, it must be admitted, that the fulfilment of such conditions is not a universal feature of music. Therefore, the question arises as to whether the interval categories based on natural standards may function in music without the constant presence of their cognitive prototypes. The answer should be given separately for two different cases; first, for the case when cognitive prototypes are available for learning, second – for the case when they are absent altogether.

For the first case the answer is obviously "yes". Natural musical intervals whose salience is easily tested by listening to simultaneous complex tones, become overlearned in the social process of the formation of a musical code. They can be reproduced with some accuracy and the categories which form around them are easily recognized (also in monophonic music). Even the secondary properties of simultaneous tones, such as their dissonance character due to the acoustic beats become generalized and may function not only in chords but also in succesion of tones.

The second case, that of music being composed and performed without the constant reference to natural intervals as specifically sounding consonances raises some doubts. There are three hypotheses concerning this problem; we may call them: "the hypothesis of innate factors", "the hypothesis of spectral learning" and "the hypothesis of incidental standards". The first two hypotheses give affirmative answer to the question of the universal character of natural intervals in music. The answer given by the third hypothesis is "no".

According to the "Hypothesis of innate factors", the human brain is genetically prepared to prefer combinations of tones in which frequency ratios correspond to natural intervals. This hypothesis had strong advocates among physicists and philosophers of the past (see PLOMP and LEVELT [13]) and is also held by some present writers (BOOMSLITER and CREEL [2]). Nevertheless its experimental verification was practically impossible for a long time (see Levelt et al. 1966). Only recently some published results of psychoacoustic experiments seem to be in favour of the hypothesis of innate factors to some degree. It was found by DEMANY and ARMAND [5] that 3-month old infants discriminate the melodic interval of an octave as leading to much greater similarity of component tones than the interval of a seventh or a ninth. This experiment leads to interesting conclusions concerning the role of innate and cultural factors at the formation of musical scales.

In the "Hypothesis of spectral learning" (TERHARDT [26]), it is assumed that since natural intervals are constantly present in the spectrum of musical tones and of vowels, so they may become overlearned through listening to such sounds.

The third one, "Hypothesis of incidental standards" is based on the fact that in some non-Western musical cultures the intervalic structure of music bears no relation to the natural overtone order (HUSSMAN [10]). However, in those cultures music has an exclusively monodic character and most of the instruments used are primitive idiophones with nonharmonic sound spectra. The lack of perceptually salient prototypes of natural intervals appears to be an evident reason for the fact that no corresponding categories are formed. Instead, some other incidental standards are accepted and corresponding categories overlearned.

The process of imprinting standard intervals in the long-term memory of musicians and music listeners closely corresponds to learning sounds of speech. From the point of view of auditory perception, simultaneous complex-tone intervals show some analogy to those syllables in which "salient points" are provided directly through the interaction of physical and psycho-acoustic phenomena (e.g. stop-vowel combinations). Melodic intervals are more similar to vowels. In those intervals, like in vowels, no physically salient prototypes are provided for recognition. Their standards may be imagined as arbitrarily chosen isolated points along smooth perceptual continua.

As a conclusion, it could be argued that the hypothesis on the universal character of "natural" harmonic intervals has no full confirmation. Instead, the elements of musical scales may be considered as being partly natural and partly arbitrary.

One thing should be explained as being a constant source of misunderstanding.

The formation of a musical scale can be understood as the division of the perceptual continuum of the pitch difference (or pitch ratio), into discrete, step-like categories. These categories tend to be formed around the salient points along the perceptual continuum. However, the salient points themselves *are not* units of a code. They are prototypes or standards employed in the formation of such units. The units of the perceptual code in music, similarly to those in speech, are categories. The perceptual categories, that are units of the pitch-difference code in music, correspond to segments on the frequency-ratio scale. Within each category various, perceptually different frequency ratios may be accepted as "intonation variants" of the same musical interval (RAKOWSKI, MIŚKIEWICZ [9]). The "within-category" discrimination of musical intervals is inhibited through the mechanism of categorical perception, however, through proper training of listeners in a specific experimental arrangement, this inhibition can be removed (BURNS, WARD [3]). The intonation variants of musical intervals are very important in playing some musical instruments (i.e. the violin). Musicians use these variants for raising or lowering pitch of some notes in order to increase harmonic tensions (RAKOWSKI [16]) The intonation variants in music are counterparts of phonetic variants in speech. Their existence may be considered an universal for the above compared domains.

3.2. *The ear*

The human ear is particularly well suited to discriminate frequency. At frequency 1000 Hz, the number of tones of different frequencies that a trained listener can discriminate within a semitone is more than a hundred (Rakowski [14]).This high selectivity revealed in laboratory conditions decreases significantly during practical listening to music. Still, it is sufficiently high to enable hearing the within-category deviations of musical pitch in a 12–semitone system of a Western music. These are two restrictions which limit this high ability. At first, the frequency selectivity of the ear drops down considerably at low frequencies.The pitch of low contrabass tones is estimated much less accurately than that of violin tones. Secondly, pitch selectivity may be lowered for very short musical tones, particularly when those are presented in rapid sequences. In such cases the additional effect of a pitch shift may occur (RAKOWSKI, HIRSH [18]).

However, frequency discrimination alone is not sufficient for the functioning of a musical pitch code. Equally important is the accuracy in estimating differences in the musical interval magnitudes or frequency ratios. Here the accuracy of the ear is significantly smaller (e.g. see HOUTSMA [9], ATTNEAVE, OLSON [1], RAKOWSKI [15]).

3.3. *Memory*

Some musicologists complain about the lack of interest of avant-guarde musicians in composing and playing quartertone or other microtone music. One of the arguments in promoting microtone interval systems, as the important enrichment of musical language, is the statement the quartertones and even smaller intervals can easily be detected by the ear. Although this statement is true, the reasoning is incorrect. The main obstacle in using microtonal divisions of an octave is not poor auditory discrimination, but limited capacity of the memory. It is very difficult to store in the memory a large number of intervals as units of the same level, e.g. after the division of an octave into 24 or 53 parts. The situation here is quite similar to the one in natural languages which all, due to memory limitations, use quite limited number of phonemes. These limitations are even more severe in the domain of musical pitch than in the language because here the units to be remembered are differentiated only through the unidimensional change of auditory sensation. George A. Miller, in his well-known paper about "the magical number seven" (MILLER 1956) gave the summary of a number of psychological experiments concerning memory. He concluded that the memory for changes of the magnitude of any single-dimensional sensation is limited to a very small number of separate steps of this magnitude: seven plus or minus two. There is no reason why such a unidimensional sensation as "magnitude" of pitch should be an exception to this rule.

There are, however, two important additional phenomena effective in overcoming the limitation of pitch memory in music. These two phenomena were gradually exploited in the long process of the formation of the fully-developed musical code, as means to overcome the contradictions between the great pitch selectivity of the ear and the limited capacity of pitch memory.

The first and most important phenomenon effective in the development of the musical code was the octave similarity of musical pitch. The second phenomenon was the hierarchization of structure into sets of units of various order. A good example of this effect is the acquisition of the chromatic scale through enlargements of the primarily acquired diatonic material.

The role played by the octave phenomenon in music requires some additional explanation. Due to the harmonic structure of most musical sounds, tones separated by the interval of an octave or its multiplicity seem "similar" or, in some aspects, "the same". The octave similarity of tones resulted in a cyclic or "spiral" character of the musical pitch scale (SHEPARD [25]). Such spiral changes in the domain of a given sensation may be regarded as a superposition of two movements, a periodic movement and a steady movement within the sensation area. Under such conditions musical pitch no longer suffered the limitations typical of all unidimensional sensations. It could be organized and memorized along two separate scales: the unprecise — "tone-height" or pitch-range scale covering the total range of pitch used in music and a within-octave scale that could be structured more precisely. The first one corresponds to the steady increase of the sensation magnitude within its full range. The second one corresponds to a periodic movement within the octave range.

The two-dimensional scaling of musical pitch concerns both relative and absolute pitch. There is however an essential difference between the way in which the magnitude of relative and absolute pitch can be fixed in the memory. The within-octave steps of relative pitch (musical intervals) can be fixed in long term memory and their sequences easily recognized as familiar melodies by most people. On the contrary, the within-octave steps of absolute pitch ("chromas", corresponding to the names of musical notes) can be permanently memorized only by a very small percentage of people, the possessors of "absolute pitch". All other people have the ability to assess absolute values of pitch very accurately but these values can be only kept in the short-term memory lasting for about 1–3 minutes (RAKOWSKI MORAWSKA-BÜNGELER [20]).

It results that only the memory for relative pitch memory for musical intervals can be used as the base for the pitch code in music. The vocabulary of this code is subjected to basic limitations of memory concerning the number of units that can be memorized. Here we see the far-reaching analogy with the limitations typical of the codes in all natural languages which, as a rule, use severly limited number of phonemes.

Limitations imposed by the basic properties of long-term memory are important universal feature, common for such codes of human communication as language and music.

References

[1] F. ATTNEAVE, R. K. OLSON, *Pitch as a medium: A new approach to psychophysical scaling*, Am. J. Psychol., **84**, 147–166 (1971).

[2] P. BOOMSLITER, W. CREEL, *Extended reference: an unrecognized dynamic in melody*, J. Music Theory 7, 2–22 (1963).

[3] E. M. BURNS, W. D. WARD, *Categorical perception – phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals*, J. Acoust. Soc. Am. **63**, 456–468 (1978).

[4] N. CHOMSKY, *Language and mind*, Harcourt, Brace and World, New York 1968.

[5] L. DEMANY, F. ARMAND, *The perceptual reality of tone chroma in early infancy*, J. Acoust. Soc. Am. (1984).

[6] M. HALLE, *The sound pattern of russian*, Mouton and Co, The Hague 1959.

[7] H. L. F. VON HELMHOLTZ, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, Vieweg u. Sohn, Braunschweig 1963.

[8] Ch. F. HOCKETT, *The problem of universals in language*, In: J. H. Greenberg[ed.] Universals of language, The MIT Press, Cambridge, Mass. 1963.

[9] A. HOUTSMA, *Discrimination of frequency ratios*, J. Acoust. Soc. Am. **44**, 383 A (1968).

[10] H. HUSSMAN, *Einführung in die Musikwissenschaft*, Quelle u. Mayer, Heildelberg 1958.

[11] R. JAKOBSON, *The phonemic concept of distinctive features*, Proc. 4-th Int. Congr. of Phonetic Sciences, Helsinki 1961.

[12] W. J. M. LEVELT, J. P. VAN DE GEER, R. PLOMP, *Triadic comparisons of musical intervals*, Brit. J. Math. Statist. Psychol., **19**, 163–179 (1066).

[13] R. PLOMP, W. J. M. LEVELT, *Tonal consonance and critical bandwith*, J. Acoust. Soc. Am., **38**, 548–560 (1965).

[14] A. RAKOWSKI, *Pitch discrimination at the threshold of hearing*, Rep. 7-th Intern. Congr. Acoustics, Budapest, 20-H-6 (1971).

[15] A. RAKOWSKI, *Tuning of isolated musical intervals*, J. Acoust. Soc. Am., **59** Suppl. Spring S 50 A (1976).

[16] A. Rakowski, *Intonation variants of musical intervals as counterparts of phonetic variants in language*, Symposium on Music and Cognitive Sciences, IRCAM, Paris, March 1988, (poster presentation 1988).

[17] A. Rakowski, *Some phonological universals in music and language*, Musikometrika **1**, 1–10 (1988).

[18] A. Rakowski, I. J. Hirsch, *Poststimulatory pitch shifts for pure tones*, J. Acoust. Soc. Am., **68**, 467–474 (1978).

[19] A. Rakowski, A. Miśkiewicz, *Deviations from equal temperament in tuning isolated musical intervals*, Arch. Acoustics **10**, 95–104 (1985).

[20] A. Rakowski, M. Morawska-Büngeler, *In search for the criteria of absolute pitch*, Arch. Acoustics, **12** (1987).

[21] J. G. Roederer, *The perception of music by the human brain*, The Humanities Assoc. Rev. **30**, 11–23 (1979).

[22] E. Rosh, *Human categorization*, In: N. Warren ed. Studies in cross-cultural psychology, vol. 1, Academic Press, London 1978.

[23] N. Ruwet, *Language, musique, poesie*, Ed. du Seuil, Paris 1972.

[24] F. Saussure, *Course de linguistique generale*, Paris – Lausanne 1916.

[25] R. N. Shepard, *Circuity in judgements of relative pitch*, J. Acoust. Soc. Am., **36**, 2346–2353 (1964).

[26] E. Terhardt, *A contribution to the theory of pitch, consonance and harmony*, J. Acoust. Soc. Am., **55**, 1061–1069 (1974).

Received on May 9, 1988.

The work was supported by a grant from the Polish Academy of Sciences (CPBP 0203. VII. 8).

ARCHIVES OF ACOUSTICS
13, 3-4, 325-329 (1988)

ANALYSIS OF SOUND ABSORPTION BY A THREE-LAYER RESONANT AND ABSORBING STRUCTURE

KRZYSZTOF WERNERÓWSKI

Faculty of Mechanics, Academy of Technology and Agriculture
(85-763 Bydgoszcz, ul. Prof. S. Kaliskiego 7)

The paper presents an analogous method of calculating resonance frequencies of a three-layer perforated structure. It was found that a model of vibrations with three degrees of freedom can be applied here. The possibility of widening frequency bands by using absorbing layers was developed and investigated. Effective attenuation of harmful noise with relatively low frequencies was stated.

W pracy przedstawiono analogową metodę obliczania częstotliwości rezonansowych trójwarstwowego ustroju perforowanego. Stwierdzono możliwość zastosowania modelu drgań o trzech stopniach swobody. Opracowano i zbadano możliwości rozszerzenia pasm częstotliwości przez zastosowanie warstw pochłaniających. Stwierdzono możliwość skutecznego tłumienia szkodliwego dźwięku o stosunkowo niskich częstotliwościach.

1. Introduction

Practical attenuation of harmful noise of machines and machinery is a very difficult problem. Spectral analysis determines characteristic frequencies and the white noise share.

Only a single resonant-absorbing structure has been worked out theoretically and experimentally in detail [8, 11]. The solution for one frequency with definite attenuation in the surroundings (e.g. in the range of 2.5 octave) is often not sufficient.

Hence, the application of multiple perforated resonant and absorbing structures is necessary. A solution (Fig. 1) effective for three frequencies with an adequately expanded band is frequently sufficient. A relatively high noise suppression effectiveness is ensured by axially distributed decreasing holes.

Calculations of structures with several layers concern multilayer baffles without perforations, mainly [8, 10]. In literature [1–7, 9] referring to analysis of sound absorption of resonant and absorbing structures the problem of multiple solutions is limited just to advise of general character.

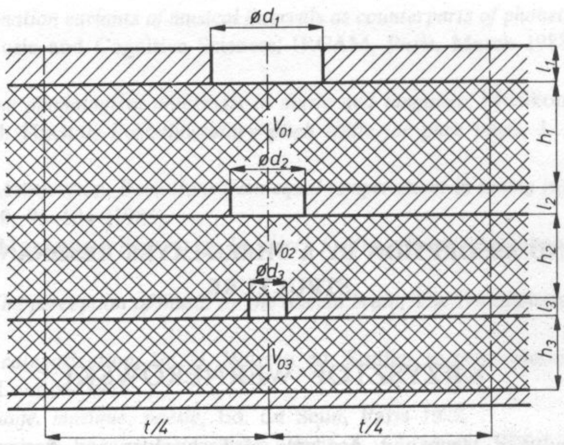


Fig. 1

It is very difficult to silence noise components of machines with several frequencies distinct in the spectrum and relatively low. Only a precisely calculated multiple resonant and absorbing structure secures effective attenuation of very strenuous noise.

There is an analogy between a three-layer perforated structure and a vibrating system with three degrees of freedom.

Air in the holes can be defined as "acoustic masses", while compliances of the gas mixture in the space between plates are inverses of acoustic "elasticities". Of course damping takes place also in the porous material.

2. Calculations

An accurate determination of the size of the vibrating system is the basis for modelling. The following quantities were calculated from the presented in Fig. 1 system's geometry:

acoustic "masses"

$$m_{ai} = \frac{4\varrho_0(l_i + \Delta l_i)}{\pi d_i^2}, \tag{1}$$

corrections for co-vibrating masses

$$\Delta l_i = 0.7 d_i, \tag{2}$$

acoustic stiffnesses

$$k_{ai} = \frac{\varrho_0 \cdot c^2}{v_{0i} - \Delta v_{0i}}, \tag{3}$$

$$V_{0i} = \frac{h_i \cdot t^2}{4} \tag{3a}$$

corrections for the existing co-vibrating mass

$$V_{oi} = 0.14\, d_i^3, \tag{3b}$$

coefficient of "attenuation" i. e. "dissipation" of acoustic energy

$$c_{ai} = \frac{\tilde{\mathbf{P}}_i}{\tilde{\mathbf{V}}_i}, \tag{4}$$

where: ϱ_0 – density of air, in kg m^{-3}, c – speed of sound in the medium, in m s^{-1}, $\tilde{\mathbf{P}}_i$ – effective pressure expressed with the method of complex numbers in the analysed part of the structure, in Pa, $\tilde{\mathbf{V}}_i$ – effective volume velocity expressed with the method of complex numbers in an element of the silencing construction, in m^3 s^{-1}.

The model of a vibrating system with three degrees of freedom is shown in Fig. 2.

Including attenuation (energy dissipation) Lagrange equations of the second type were supplemented to the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} + \frac{\partial \Phi}{\partial \dot{q}_i} = 0, \tag{5}$$

$$L = T - U\Phi = \frac{1}{2}\sum_{i,j=1}^{f} c_j\, \dot{q}_i\, \dot{q}_j, \tag{5a, 5b}$$

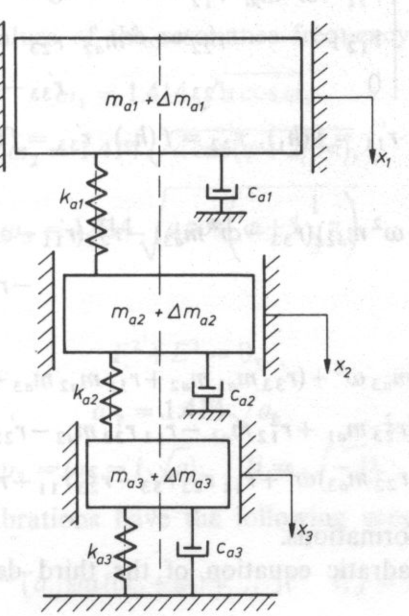


Fig. 2

where: T – kinetic energy, U – potential energy, Φ – Rayleigh's dissipation function, q_1 – generalized coordinates $x_1, x_2, x_3, \dot{q}_1$ – generalized velocities, $x_1 \ldots$ The actual mechanical system is described by the following system of differential equations

$$\begin{aligned} m_{a1}\ddot{x}_1 + c_{a1}\dot{x}_1 + r_{11}x_1 + r_{12}x_2 = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ m_{a3}\ddot{x}_3 + c_{a3}\dot{x}_3 + r_{32}x_2 + r_{33}x_3 = 0, \end{aligned} \tag{6}$$

$$r_{11} = k_{a1}, \tag{6.1}$$

$$r_{12} = -k_{a1} = r_{21}, \tag{6.2}$$

$$r_{22} = k_{a1} + k_{a2}, \tag{6.3}$$

$$r_{23} = -k_{a2} = r_{32}, \tag{6.4}$$

$$r_{33} = k_{a2} + k_{a3}, \tag{6.5}$$

$$r_{13} = r_{31} = 0. \tag{6.6}$$

We seek solutions in the following form

$$x_i(t) = e^{-h_i t} f_i(t), \tag{7}$$

$$h_i = \frac{c_{ai}}{2m_{ai}} \tag{8}$$

The determinant of the characteristic system

$$D(\omega^2) = \begin{vmatrix} r'_{11} - \omega^2 m_{ai} & r_{12} & 0 \\ r_{12} & r'_{22} - \omega^2 m_{a2} & r_{23} \\ 0 & r_{23} & r'_{33} - \omega^2 m_{a3} \end{vmatrix} = 0, \tag{9}$$

$$r'_{11} = f(h_1),\ r'_{22} = f(h_2),\ r'_{33} = f(h_3), \tag{9a, b, c}$$

are with expansion

$$(r'_{11} - \omega^2 m_{a1})(r'_{22} - \omega^2 m_{a2})(r'_{33} - \omega^2 m_{a3}) - r_{23}^2(r'_{11} - \omega^2 m_{a1}) - r_{12}^2(r'_{33} - \omega^2 m_{a3}) = 0. \tag{10}$$

Equation

$$\begin{aligned} -m_{a1}m_{a2}m_{a3}\omega^6 + (r'_{33}m_{a1}m_{a2} + r'_{11}m_{a2}m_{a3} + r'_{22}m_{a1}m_{a3})\omega^4 \\ + (r_{23}^2 m_{a1} + r_{12}^2 m_{a3} - r'_{11}r'_{33}m_{a2} - r'_{22}r'_{33}m_{a1} \\ - r'_{11}r'_{22}m_{a3})\omega^2 + r'_{11}r'_{22}r'_{33} - r_{23}^2 r'_{11} - r_{12}^2 r'_{33} = 0, \end{aligned} \tag{11}$$

is the result of transformations.

A resultant biquadratic equation of the third degree can be noted in the following form

$$\omega^6 + A\omega^4 + B\omega^2 + C = 0, \tag{12}$$

$$A = \frac{r'_{33} m_{a1} m_{a2} + r'_{11} m_{a2} m_{a3} + r'_{22} m_{a1} m_{a3}}{-m_{a1} m_{a2} m_{a3}}, \tag{12a}$$

$$B = \frac{r'_{11} r'_{33} m_{a2} + r'_{22} r'_{33} m_{a1} + r'_{11} r'_{22} m_{a3} - r^2_{23} m_{a1} - r^2_{12} m_{a3}}{m_{a1} m_{a2} m_{a3}} \tag{12b}$$

$$C = \frac{r^2_{23} r'_{11} + r^2_{12} r'_{33} - r'_{11} r'_{22} r'_{33}}{m_{a1} m_{a2} m_{a3}}. \tag{12c}$$

Substituting

$$\omega^2 = x - \frac{A}{3}, \tag{13}$$

we have

$$x^3 + 3Ex - 2F = 0. \tag{14}$$

Real roots have vibroacoustic sense, for

$$F^2 + E^3 > 0, \tag{15}$$

and

$$\varphi = \frac{1}{3} \arccos \frac{F}{a^3}. \tag{16}$$

$$a = \pm\sqrt{|E|} \quad \text{with opposite sign } F \tag{17}$$

there are three positive values of the resonance frequency

$$\omega_1 = 1.414\sqrt{a\cos\varphi}, \tag{18a}$$

$$\omega_2 = 1.414\sqrt{a\cos(\varphi + 2/3\pi)}, \tag{18b}$$

$$\omega_3 = 1.414\sqrt{a\cos\left(\varphi + 1\frac{1}{3}\pi\right)}, \tag{18c}$$

and in the case of

$$F^2 + E^3 = 0, \tag{19}$$

$$\omega_1 = 1.414\sqrt{a}, \tag{19a}$$

$$(\omega_2 = \omega_3 = i\sqrt{a}), \quad (i = \sqrt{-1}). \tag{19b, c}$$

Model of acoustic vibrations have the following general form

$$f_i(t) = e^{-h_i t}\left[a_{ij} \sin(\omega_{it} + \varphi_i) + \ldots\right], \quad i, j = 1, 2, 3. \tag{20}$$

The structure has a relatively wide frequency range of the noise attenuation.

Naturally, constructions filled with sound absorbing material (Fig. 1) are effective in a wider range of frequencies.

The width of a band within which the absorption coefficient does not drop below half the value it has at resonance is important.

Calculations of a three-layer resonant and absorbing structure were carried out on the basis of results of analysis of the analogue model (15 – 19b, c) and the methodology of insulating-sound absorbing structures [8, 11].

The curve (Fig. 3) of the absorption coefficient of the three-layer resonant and absorbing structure depends on the frequency of absorbed sound. Theoretical curves

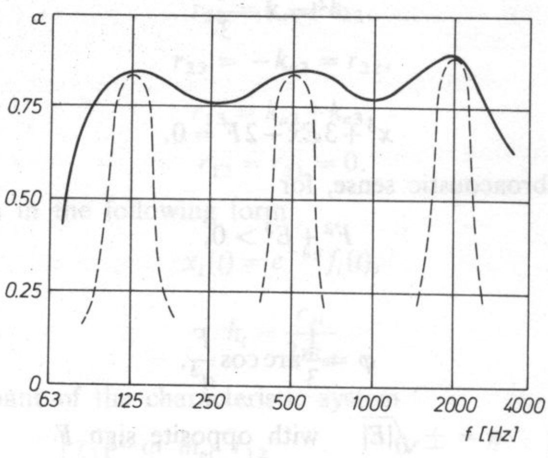


Fig. 3

for pure resonance of structure's sub-assemblies are denote with dashed lines. Continuous attenuation is ensured by the porous material and co-operation of individual Helmholtz resonators.

Of course extremal values are a result of free vibration frequencies of the system. A filled resonant and absorbing structure acts effectively on acoustic waves inciding under various angles.

Simplified model testing was performed and the literature was analysed [7, 11]. The analogous method was found sufficiently consistent with actual functioning.

The three-layer resonant and absorbing structure can function effectively in a band up to ~ 7.5 octave for $a_{rez} = 0.6$.

3. Conclusions

1. The axial distribution of decreasing openings ensures a relatively high effectiveness of noise suppression.

2. An analogy exists between acoustic functioning of a three-layer perforated structure and a vibrating system with three degrees of freedom.

3. A three-layer resonant and absorbing structure functions in a wide band of sound frequencies.

References

[1] Cz. Cempel, *Applied vibroacoustics* (in Polish) PWN, Warszawa – Poznań 1978.

[2] Z. Dąbrowski, A. Hać, S. Radkowski, *Selection of acoustic insulation of the power unit's casing in a K-611 hydraulic excavator* (in Polish), Proc. Noise Control 85, Kraków 1985.

[3] A. Drzymała, H. Herba, S. Klimczyk, J. Kowalski, M. Lewicki, T. Mazurkiewicz, R. Wyrzykowski, *Noise emission reduction for chosen sources in furniture production* (in Polish), Proc. Noise Control 85, Kraków 1985.

[4] Z. Engel, L. Styczniewicz, *Acoustic energy radiation by a system of flat superficial sound sources* (in Polish), Proc. Noise 85, Kraków 1985.

[5] W. Fasold, *Beitrag zur Ermittlung des Einflusses von einzelmen kleinen Öffnungen auf die Schalldämmung, Hochfrequenztechnik u. Elektroakustik*, Band 72, Heft 3 (1971).

[6] W. Gierulski, Dissipation of a sound wave's energy in Polish, Proc. Noise Control 85, Kraków, 1985.

[7] I. Malecki, *Theory of acoustic waves and systems* (in Polish), Warszawa 1964.

[8] J. Sadowski, *Foundations of acoustic insulating power of structures* (in Polish), PWN, Warszawa 1973.

[9] K. Sobol, K. Ziemiański, *Method of finite bands in the analysis of free vibrations of three-layer plate systems* (in Polish), Proc. Noise Control 85, Kraków 1985.

[10] K. Wernerowski, *Problems of noise minimalization in machines on the noise's path of propagation* (in Polish), Proc. Noise Control 85, Kraków 1985.

[11] Z. Żyszkowski, *Foundations of electroacoustics* (in Polish) WNT, Warszawa 1953.

Received in revises version in March 4, 1988.

Acknowledgements

No scientific journal could survive whithout the help of experts who have agreed to act as referees. It is encouraging to note that despite the conditions under which researchers work, a pure and unselfish scientific spirit is preserved by the reference who ensure the very existence of scientific communication.

Because of the large diversity of topics that *Archives of Acoustics* covers and our efforts to shorten the publication process, it is impossible for our Editorial Board members to review all the papers we are receiving.

Here we give a list of colleagues who have kindly agreed to review manuscripts for *Archives of Acoustics* (13 volume). I would like to thank all of them for their assistance and hope that they will continue to give us the benefit of their invaluable advice.

Eugeniusz Danicki
Editor-in-chief

J. Adamczyk
Z. Engel
L. Filipczyński
J. Fyk
J. Filipiak
Z. Jagodziński
A. Jaroch
A. Jaroszewski
J. Kacprowski
Z. Kotulski
R. Markiewicz
A. Markiewicz
M. Niewiarowicz
T. Pałko
W. Pajewski
A. Rakowski
W. Rdzanek
F. Rejmund
A. Sliwiński
M. Sowiński
W. Więźlak
R. Wyrzykowski
I. Żuchowicz-Wodnikowska