ARCHIVES OF ACOUSTICS
13, 1–2, 3–11 (1988)

TUTORIAL

REVIEW OF THE TRANSMISSION OF SOUND FROM AIR TO WATER

H. W. JONES and H. W. KWAN

Technical University of Nova Scotia, Institute of Acoustics, Halifax, Nova Scotia, Canada

This paper reviews the physical principles relating to the transmission of sound from air to water. The major emphasis is on the effect of the air-water interface on the transmission of sound to an underwater receiver. A limited consideration only is given to non-flat surfaces and the transmission at grazing angles of incidence.

Artykuł przedstawia fizyczne podstawy transmisji dźwięku z powietrza do wody. Główny nacisk położony na efekty graniczne i ich wpływ na rejestrację przechodzącej fali dźwiękowej pod wodą. Pewne problemy związane z pofalowaną powierzchnią wody są też dyskutowane.

1. Introduction

The transmission of sound from air to water has been the subject of many publications as is shown by references [1-5]. The physical problem relates to two situations; that in which the interface is supposed to be flat i.e. by comparison to the wave length of sound and that in which the surface is supposed to be disturbed by waves. The emphasis in this paper is on the consideration of the first case. There has been much discussion in the literature of the general problem of a situation which relates to a curved wavefront incident on a boundary. This problem has been considered in various circumstances for electromagnetic waves [6, 7] as well as acoustic waves [13]. The topic of interfacial effects has been fashionable in recent years as it has a connection with transmission of noise over the ground. In this regard there is a substantial literature, of which, ATTENBOROUGH's work [8] might be representative and also convenient for its listing of the major references.

2. The transmission of sound into water from a point source

If we follow URICK's approach [2] then we see (Fig. 1) that sound can arrive at the receiver by several paths. First, there is the direct, refracted, path which is shown as *OAR* in Fig. 1. Second, there is the indirect path from bottom reflection, i.e. *OBCR*. Finally, there is a path which may be associated with surface, layer or inhomogeneous waves following paths such as *ODR*. This surface wave might be important at depths small compared with the wavelength of the sound.

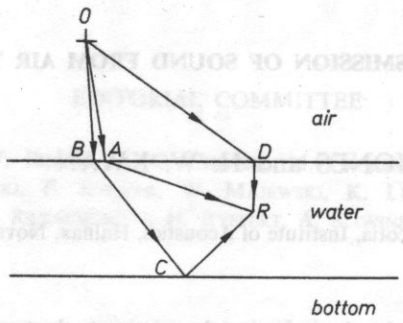


FIG. 1. The contribution to the sound field at a point in the Sea made by a source O in air

If we concern ourselves first with the direct path then we note that there has been considerable discussion of this topic in the literature. HUDIMAC [9] proposed a consideration based on the assumption that the sound reaching the particular point underwater could be obtained from a ray treatment; see Fig. 2. Essentially it was assumed that an element of the wavefront at some point on the surface can be assumed to be plane and the transmission coefficient at some angle (θ_0 in Fig. 2) can then be used to obtain the contribution to the sound pressure at an element dR. It is shown on these assumptions that:

$$\frac{I}{E} = \frac{\varrho_1 c_1 \varrho_2 c_2 \sin^2 \theta_0}{\pi(\varrho_2 c_2 + \varrho c_1 B)^2 (h + dc_2/Bc_1)(h + dc_2/B^3 c_1)}, \tag{1}$$

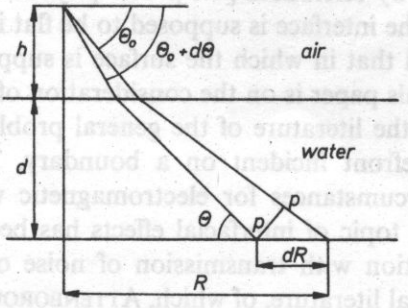


FIG. 2. The geometry of the ray paths. The source is at a height h above the water surface

where

$$B = \{1+[1-(c_2^2/c_1^2)]\,\mathrm{ctn}^2\,\theta_0)\}^{1/2},$$

I, is the sound intensity at R, and E is the output power of the source. HUDIMAC calculated the intensity for a particular case, see Fig. 3.

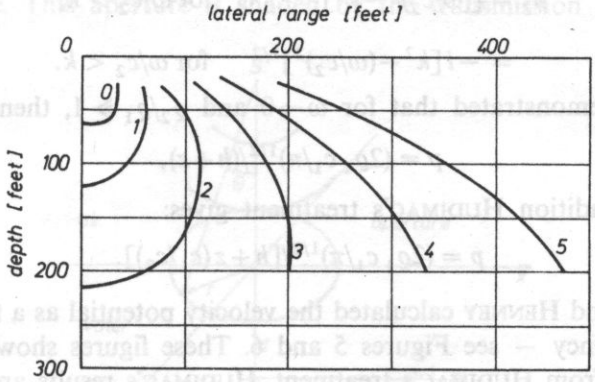


FIG. 3. Plot of iso-intensity lines *vs* range and depth for a 1 watt simple source at 25 feet. Intensities are given in dB re 1 watt/cm^2. Curve O: 120 dB, 1: − 125 dB, 2: − 130 dB, 3: − 135 dB, 4: − 140 dB, and 5: − 145 dB (After HUDIMAC)

WEINSTEIN and HENNEY [1], about eight years later attempted the problem using wave theory to obtain sound pressure at a receiver in the cordinate system in Fig. 4. Its is assumed that there is a point source in air with a velocity potential given by:

$$\phi = (1/R)\exp(-i\omega R/c_1), \tag{2}$$

and in the water the velocity potential is given by:

$$\phi = \int^{\infty} [2\beta_2/(\sigma\beta_1+\beta_2)]\,\{\exp[-i(\beta_2 z+\beta_1 h)]\}\,[J_0(kr)\,k/(i\beta_2)]\,dk, \tag{3}$$

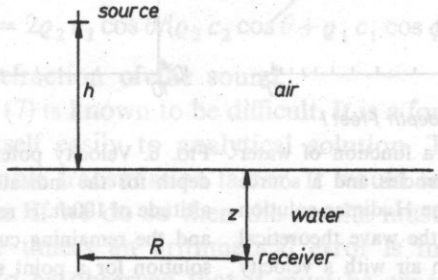


FIG. 4. Coordinate system

where

$$\delta = \varrho_2/\varrho_1,$$

$$\beta_1 = [(\omega/c_1)^2 - k^2]^{1/2} \quad \text{for } \omega/c_1 > k,$$

$$= -i[k^2 - (\omega/c_1)^2]^{1/2} \quad \text{for } \omega/c_1 < k,$$

$$\beta_2 = [(\omega/c_2)^2 - k^2]^{1/2} \quad \text{for } \omega/c_2 > k,$$

$$= -i[k^2 - (\omega/c_2)^2]^{1/2} \quad \text{for } \omega/c_2 < k.$$

These authors demonstrated that for $\omega \to 0$ and $\varrho_2/\varrho_1 \gg 1$, then:

$$p = (2\varrho_1 c_1/\pi)^{1/2}/(h+z), \tag{4}$$

for the same condition HUDIMAC's treatment gives:

$$p = (2\varrho_1 c_1/\pi)^{1/2}/[h + z(c_1/c_2)]. \tag{5}$$

WEINSTEIN and HENNEY calculated the velocity potential as a function of water depth and frequency – see Figures 5 and 6. These figures show a curve for the results obtained from HUDIMAC's treatment. HUDIMAC's results are consistent with the wave treatment for large values of ω and large values of z.

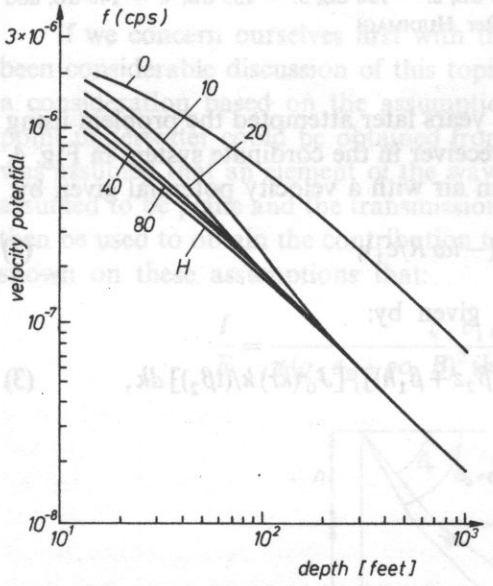


FIG. 5. Velocity potential as a function of water depth for the indicated frequencies and a source altitude of 25 ft. *H* refers to the Hudimac solution and the remaining curves to the wave theoretical solution for a point source in air with a velocity potential of unity at 1 cm (After WEINSTEIN and HENNEY)

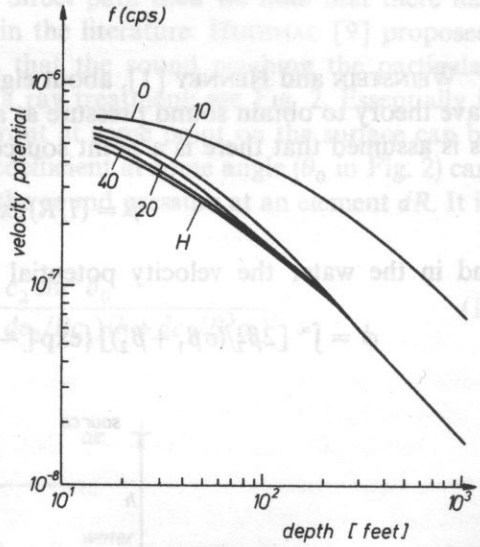


FIG. 6. Velocity potential as a function of water depth for the indicated frequencies and a source altitude of 100 ft. H refers to the Hudimac solution and the remaining curves to the wave theoretical solution for a point source in air with a velocity potential of unity at 1 cm (After WEINSTEIN and HENNEY)

3. Discussion of ray and wave theory results

The findings of the authors of references [1 and 9] are to be expected in that the problem which is being considered is essentially a diffraction problem – see Fig. 7. The incident sound field consists of spherical (or near spherical in the presence of wind) wave fronts impinging on an aperture shown somewhat diagrammatically at the water-surface. This aperture is shaded by the transmission coefficient of the

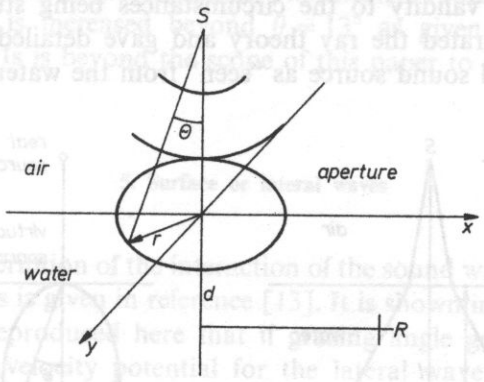


FIG. 7. The geometry of the aperture

sound at the interface; the transmission is a function of and is symmetrical about a vertical axis. Seen in these terms we are left with the solution of the Fresnel-Kirchhoff integral as in optics. Taking the diffraction integral as given in BORN and WOLF's book [10]

$$U(P) = -iA/(2\pi)\int\int (1/rs)\,[\exp ik(r+s)]\,[\cos(n,r)-\cos(n,s)]\,dS, \tag{6}$$

then for this case,

$$U(P) = -iA/(2\pi)\int\int T(1/rs)\,[\exp ik(r+s)]\,[\cos(n,r)-\cos(n,s)]\,dS, \tag{7}$$

where T is the sound pressure transmission coefficient at the interface i.e.:

$$T = 2\varrho_2 c_2 \cos\theta/(\varrho_2 c_2 \cos\theta + \varrho_1 c_1 \cos\phi),$$

ϕ being the angle of refraction of the sound.

The solution of Eq. (7) is known to be difficult. It is a form of Fresnel diffraction which does not lend itself easily to analytical solution. The situation is further complicated by the variable transmission factor. It would seem to be convenient to use a numerical solution. If we do so then the results must coincide with those of WEINSTEIN and HENNEY unless an arithmetical error is made.

At this point it is to be observed that it must be expected that the HUDIMAC solution converges to that of the WEINSTEIN and HENNEY solution because as the

wavelength becomes short it is in the nature of the diffraction integral that its result will be similar to ray theory. This point is well demonstrated in the general literature. If further demonstration of the point is necessary, then to the extent that data are available, reference [3] supports the conclusion that the wave or diffraction solution is, in general, the required solution. It should be observed however, that the diffraction solution is only valid for the circumstance that the coherence length of the sound is sufficient for it to encompass all the phase changes required by the geometry of the aperture. This assumption is often made in noise studies without the necessary demonstration of its validity to the circumstances being studied.

URICK [2] elaborated the ray theory and gave detailed consideration of the position of the virtual sound source as "seen" from the water, see Fig. 8. He notes

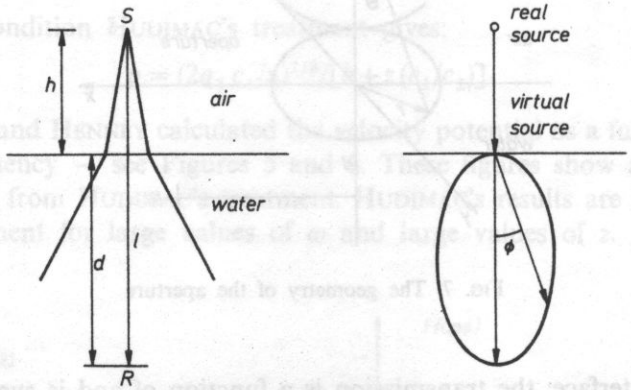


FIG. 8. The receiver (a) beneath the source and (b) in the farfield at distances much greater than the height of the source

that the apparent source height is the same as the apparent depth which is used in elementary optical theory. As the treatment URICK uses is in principle that advanced by HUDIMAC it provides an approximation similar to his. As URICK points out this approximation (accepting θ_{max} is 13°) can be valuable in that it allows a simpler calculation of the sound pressure at a point in the water. His analysis arrives at the equation

$$I_\omega = I_a[4n^2/(d+nh)^2], \tag{8}$$

where I is the intensity, d the depth in the water, h the height of the source and $n = c_1 \cos\theta/c_2 \cos\phi$. URICK shows that for far field, Eq. (8) becomes

$$I_\omega = 4I_a n^2 \cos^2\phi/1^2, \tag{9}$$

after some approximation. This means the source can ce replaced by a dipole located at the surface and radiating as $\cos^2\phi$. The equivalent source has an intensity of $4n^2$ times the real source. As URICK points out this is a valuable approximation in that it allows rapid calculation of the sound intensity under the water.

4. The transmission of sound through a rough interface

This is essentially a problem relating to effect of the change of the angle of incidence associated with a non-flat surface. This is a problem possibly statistical in which the shape of the water surfaces is required. This allows a suitable generalization to be made [11, 12]. The rough surface effect is two-fold. First, the transmission factor (T) is changed locally by the surface shape and the modified value is required in the Fresnel-Kirchhoff integral equation. The effective radius of the aperture of the surface (see Fig. 7) is increased beyond $\theta = 13°$ as given by the total external reflection condition. Is is beyond the scope of this paper to deal with the details of this situation.

5. Surface or lateral waves

An excellent description of the interaction of the sound wave at the surface when total reflection occurs is given in reference [13]. It is shown in an argument which is too detailed to be reproduced here that if grazing angle geometry (see Fig. 9) is considered then the velocity potential for the lateral wave is given by

$$|\phi_{lat}| = 2\exp[-kd(\sin^2\theta - n^2)^{1/2}]/[(\varrho_2/\varrho_1)(r^2+h^2)^{1/2}], \quad (10)$$

subject to:

$$\sin\theta = r/[(r^2+h^2)^{1/2}] > n, \text{i.e. } \theta > 13°,$$

where $n = c_1/c_2$.

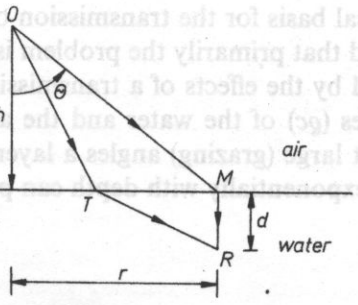


FIG. 9. Paths of wave penetration into the lower medium for the case: $n < 1$

This shows that for a given θ the surface wave decays as $\exp-(d/\lambda)$ below the surface. Similarly, the wave decays as θ increases. Urick evaluated the sound pressure at different depths and his data are reproduced below (see Fig. 10). It is to be noted that roughness of the waves on the water surface will change the situation so that

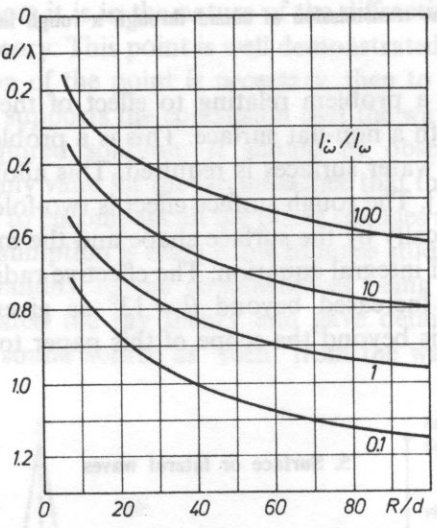


FIG. 10. Ratio of the intensity I_ω of the lateral wave to the intensity I_ω of the refracted wave, using coordinates of d/λ and R/d, where d = receiver depth, R – receiver horizontal range, and λ – wavelength

sin $\theta > n$ condition, will be satisfied in "unusual" regions on the surface. A satisfactory theory for the maintenance, transmission and decay of surface waves in these circumstances does not appear to exist.

6. Conclusion

The essential physical basis for the transmission of sound from air to water has been reviewed. It is noted that primarily the problem is one of diffraction through an aperture which is shaded by the effects of a transmission coefficient which is related to the physical properties (ϱc) of the water and the air, the angle of incidence and surface perturbations. At large (grazing) angles a layered or surface wave exists and this wave which decays exponentially with depth can provide significant sound levels close to the surface.

References

[1] M. S. WEINSTEIN and A. G. HENNEY, J. A. S. A. **37**, 5, 899–901 (1965).
[2] R. J. URICK, J. A. S. A. **52**, 3, 993–999 (1972).
[3] R. W. YOUNG, J. A. S. A. **50**, 5, 1392–1393 (1971).
[4] D. H. TOWNE, J. A. S. A. **44**, 1, 65–76 (1968).
[5] A. R. WENZEL, J. A. S. A. **55**, 5, 956–963 (1974).
[6] A. SOMMERFIELD, Ann. d. Physik, **28**, 665 (1909).

[7] B. Van der Pol, Physica **2**, 843 (1935).
[8] K. Attenborough, S. I. Hayek and J. M. Lawther, J. A. S. A. **68**, 5, 1493–1501 (1980).
[9] A. A. Hudimac, J. A. S. A. **29**, 8, 916–917 (1957).
[10] Born and Wolfe, Principles of optics, 3rd edition, Pergamon Press, pp. 382.
[11] S. C. Lubord and P. M. Hurdle, J. A. S. A. **60**, 5, 1048–1052 (1976).
[12] I. A. Urusovskii, Sov. Phys. Acoust., **10**, 3, 287–293 (1965).
[13] L. M. Brekhovskikh, Waves in layered media, chapter IV, Academic Press, 1960.

Received on May 24, 1988.

ARCHIVES OF ACOUSTICS
13, 1–2, 13–43 (1988)

TUTORIAL

SCANNING ACOUSTIC MICROSCOPY OF POLYMERIC MATERIALS AND BIOLOGICAL SUBSTANCES

ROMAN GR. MAYEV

Center of Acoustic Microscopy, USSR Academy of Sciences Institute of Chemical Physics, USSR, Moscow, 117334 Kosygin str. 4

For the last few years a new method of visualization and quantitative analysis of physico-mechanical properties and microstructure of heterogenetic media – the acoustic microscopy – has been intensively elaborated in the world. In this method waves of ultrasound and hypersound range are used as an analysis factor. It allows to use this method for investigating a wide variety of opaque materials and goods and obtaining information about their inner structures as well as for optically transparent materials in which the contrast between different structures is practically absent. In both cases an investigator receives information that is quite different from that obtained with the help of other methods, namely, the distribution of local physico-mechanical properties for example, bulk compression, shift, etc in the material of the sample.

We give a review of the results of the work on methods and means of acoustic microscopy, worked out in the Centre for Acoustic Microscopy of the USSR Academy of Sciences for investigation of polymeric composites and biological objects as well as the results of analogical investigations of the loading scientific centres in the world.

We set forth common physical basis and principles of getting acoustic images as well as the methods of studying microstructures and mechanical properties of heterogenetic objects with the help of acoustic microscope.

W ciągu ostatnich lat rozwijana jest na świecie akustyczna mikroskopia – nowa metoda wizualizacji i analizy jakościowej własności fizykomechanicznych i mikrostruktury ośrodków niejednorodnych. W metodzie tej stosowane są fale o częstotliwościach ultra- i hiperdźwiękowych. Pozwala to na badanie zarówno różnych nieprzezroczystych materiałów i przedmiotów i uzyskanie informacji o ich wewnętrznych strukturach, jak również badanie materiałów optycznie przezroczystych, w których nie ma praktycznie kontrastu pomiędzy różnymi strukturami. W obu przypadkach badacz uzyskuje informacje całkowicie różne od otrzymanych za pomocą innych metod, a mianowicie rozkład lokalnych własności fizykomechanicznych np. ściśliwość objętościowa, przesunięcie w materiale próbki.

W artykule podajemy przegląd wyników prac wykonanych w Centrum Mikroskopii Akustycznej Akademii Nauk ZSRR i obejmujących akustyczną mikroskopię kompozytów

polimerowych i obiektów biologicznych oraz wyniki analogicznych badań uzyskanych przez przodujące ośrodki naukowe na świecie.

Opracowaliśmy zarówno podstawy fizyczne i zasady uzyskiwania obrazów akustycznych, jak również metody badania mikrostruktur i własności mechanicznych ośrodków niejednorodnych za pomocą mikroskopii akustycznej.

1. Introduction

Today there are several sufficiently effective methods for investigating physico-chemical morphology and the local distribution of microstructures in polymeric mixtures and biological substances. The most important among them are optical and electron microscopic methods.

The present work is an analysis of the possibilities of a fundamentally new method, scanning acoustic microscopy SAM, for studying the physico-mechanical microstructure and dynamic processes in materials of the most diverse nature.

2. Principles of Scanning Acoustic Microscopy (SAM)

Let us first recall the principles of acoustic microscopy (Fig. 1). A high radio frequency pulse from a piezoelectric transducer is focused by an acoustic lens in a liquid coupling medium. The lens is in the form of a spherical cavity at the end of the soundguide. On interacting with the sample a focused beam is partially reflected from the sample and partially passes through it. In the former case we have a reflecting microscope; and in the latter case, after the beam passes through a second lens, we have a transmission acoustic microscope. Subsequent operations include the scanning of the sample relative to the focal region, recording the signal at every point and storing the signal in memory unit, synchronizing the sweep with the scanning, and using the signal to control the intensity of the electron beam. These operations result in the formation of an acoustic image on the display screen.

The interaction of sound wave and the sample yields information that is quite different from that obtained with the help of optical and electron microscopy. This is due to the fundamental difference in the physical nature of ultrasound which manifests itself in the form of elastic deformation waves in a medium. The mechanical nature of ultrasound makes it possible to obtain new information on the mechanical properties of an object with the help of SAM, as distinct from other methods.

The most important characteristics of the SAM method are its resolving power and depth of penetration into the sample. They depend on the frequency of ultrasound, the performance of the lens system, the nature of immersion medium and the properties of the material under investigation etc. With an increase in resolution, the depth of penetration of ultrasound into the sample decreases. Therefore, the frequency of ultrasound should be selected according to the given type of sample and

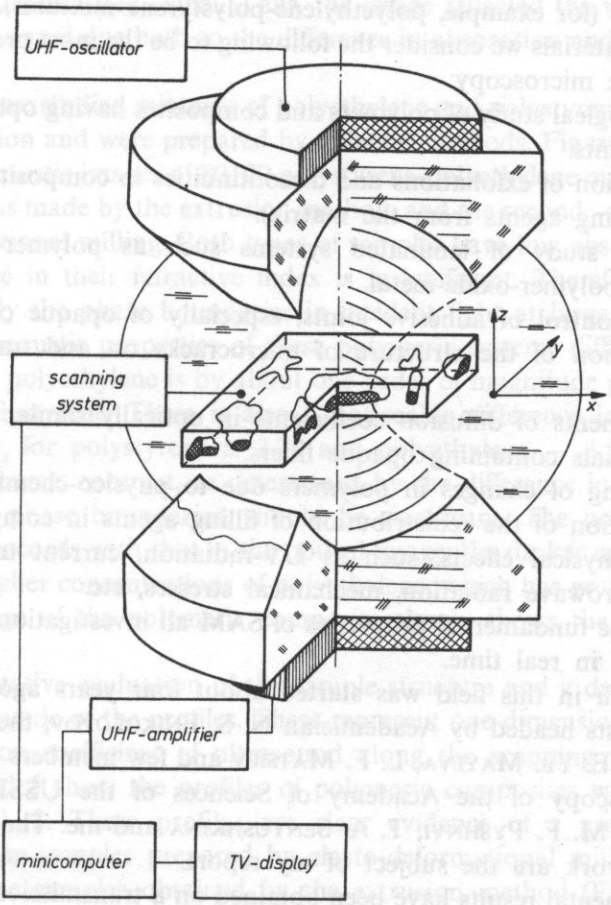


FIG. 1. Principal scheme of transmission scanning acoustic microscope (SAM)

the purpose of the investigation, with a compromise being made between the resolving power and the degree of penetration.

The method of scanning acoustic microscopy is highly sensitive to the presence of various heterogeneities to the appearance of discontinuities, exfoliation and phase boundaries. Owing to mismatch in acoustic impedance there appear strong reflections from the interfaces.

3. SAM of Polymer Composite Materials

Acoustic microscopy has been used in studying the morphology of opaque polymers and composites, in evaluating the elastic properties of deformed polymers, as well as in the investigation of polymeric mixtures whose components have similar

optical properties (for example, polyethylene-polystyrene mixtures). For polymeric and composite materials we consider the following to be the most promising areas of applying acoustic microscopy:

1. Morphological study of polymers and composites having optically similar or opaque components.
2. Investigation of exfoliations and discontinuities in composites (for example, exfoliation of filling agents from the matrix).
3. Adhesion study of laminated systems such as polymer-polymer, polymer-metal, and polymer-oxide-metal.
4. Quality control of adhesive joints, especially of opaque composites.
5. Investigation of the structure of microcracks on and under the sample surface.
6. Measurements of diffusion coefficients in optically similar polymers or in composite materials containing opaque fillers.
7. Monitoring of changes in polymers due to physico-chemical effects.
8. Investigation of the redistribution of filling agents in composites resulting from different physical effects, such as UV-radiation, current transmission and exposure to microwave radiation, mechanical stresses, etc.

Owing to the fundamental properties of SAM all investigations can be carried out directly and in real time.

Our research in this field was started about four years ago by a group of polymer specialists headed by Academician N. S. ENIKOLOPOV; the group included D. D. NOVIKOV, E. Yu. MAYEVA, L. F. MATSIEV and few members of our Center of Acoustic Microscopy of the Academy of Sciences of the USSR: V. M. LEVIN O. V. KOLOSOV, M. F. PYSHNYI, T. A. SENYUSHKINA and me. The most important results of this work are the subject of my report.

The experimental results have been obtained on a transmission raster scanning microscope of our own design [1]. The microscope operated at 0.450 GHz, having a resolution of 3μm. An ultra-high-frequency generator was used for the excitation of ultrasound with the aid of a film transducer.

After the ultrasound is focused on the sample it was received by a confocal lens system. This system consisted of two soundguides made of monocrystalline sapphire oriented along z-axis and acoustic lenses at both ends in the form of spherical cavities with a 300 μm radius of the curvature and a 0.8° aperture. The signal from the receiving transducer was fed, following its amplification, through a computer into a memory unit. The sample was secured in the holder of the scanning system. During a period of $\tau = 8$ s the scanning system made it possible to obtain an acoustic image of the sample cross section with a scanning area of from 0.7×0.7 mm to 2×5 mm. In order to get three-dimensional structure one could also change the position of a selected cross section with respect to the sample thickness within the limits of ± 0.5 mm. Following preliminary numerical treatment of the results on a computer the outgoing data was displayed on a TV-screen as a half-tone image and in the form of profiles and histograms. To obtain the image we used only the differences in the

amplitude of the transmitted signal. Thus, the image reflected the variations in the weakening of the signal due both to the difference in absorption and to reflection or scattering.

In our work we studied mixtures of polyethylene and polystyrene which differed in their composition and were prepared by different methods. Figures 2 and 3 show the optical and acoustic images of 80:20 polystyrene-polyethylene mixtures. The first set of mixtures was made by the extrusion method, and the second – by the method of elasto-deformational milling. Both types of samples have low absorption of light, and the difference in their refractive index is insignificant. Therefore, the optical images reveal only the phase boundaries in incident light at large angles. On the other hand, the acoustic properties of these polymeric systems differ considerably: the absorption in polyethylene is by about one order of magnitude greater than the absorption in polystyrene. There is also a noticeable difference in their acoustic impedance: the v_s for polystyrene is 2.28 and polyethylene – 1.91, respectively. Therefore, the acoustic contrast, as determined by the difference in the absorption and reflection at phase boundaries, should be significant. The position of phase boundaries fully accords with that in the optical image: the darker areas correspond to the zones of higher concentrations of polyethylene, which has greater absorption. The acoustic image of the polymeric composite clearly shows the distribution of phases.

For a quantitative evaluation of the sample structure and a description of the inclusions, use is made of the profiles. These represent one-dimensional distribution of the transmission coefficient of ultrasound along the scanning path.

Figures 4 and 5 show the profiles of polymeric composites which have been shown in Figs. 2, 3. These profiles are clear evidence of a greater degree of homogeneity of the samples prepared by elasto-deformational milling (Fig. 5), as compared with the samples obtained by the extrusion method (Fig. 4). On such a profile that is above the line corresponding to the polyethylene absorption, there are peaks which correspond to polystyrene inclusions.

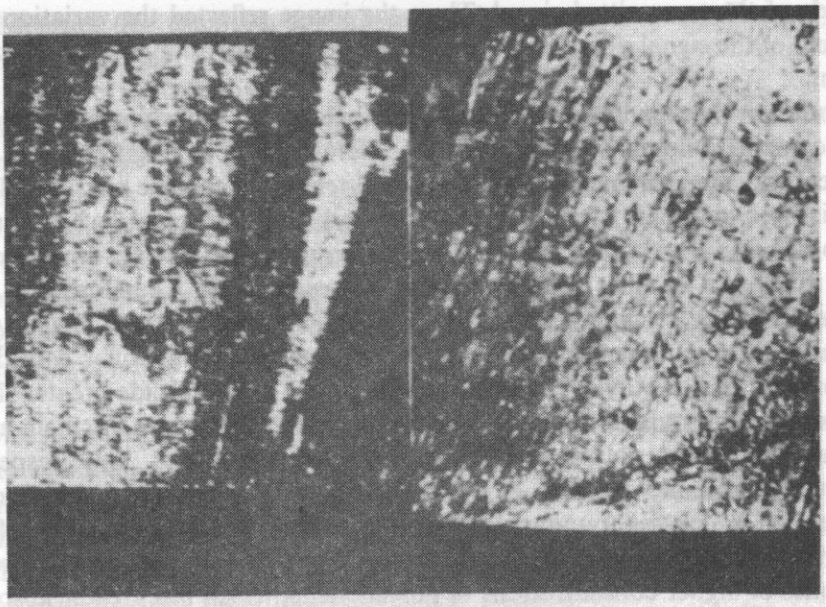
For two-component compositional materials one can by way of illustration offer a simple morphologic description of one inclusion in another. When scanning a sample which is the matrix of substance A with inclusion of substance B, the amplitude of the transmitted ultrasonic wave reflects the variations of attenuation in the focal region of SAM. These variations are determined by the dimensions and physicomechnical properties of heterogeneity. Thus, for inclusion B located in the focal region and having dimensions l_B along the acoustic axis, the amplitude of the signal from the receiving transducer can be described by the following equation:

$$A = A_0 \left(\frac{\varrho_A \varrho_B C_A C_B}{\varrho_A C_A + \varrho_B C_B} \right)^2 \exp(\alpha_A - \alpha_B) l_B,$$

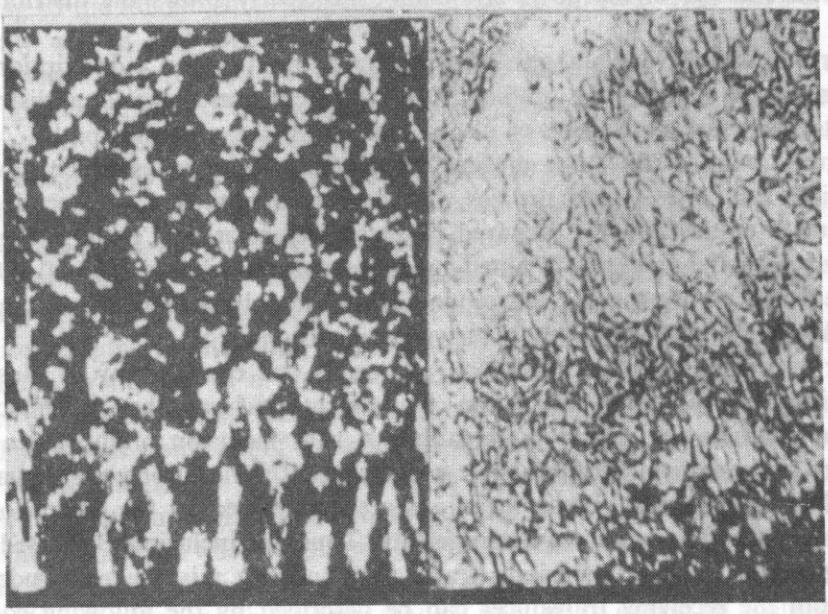
where

α_A absorption coefficient of ultrasound in A medium,

α_B absorption coefficient of ultrasound in B medium,



a)



b)

FIG. 2, 3. Acoustic and optical images of 50:50 polystyrene-polyethylene mixtures (a) — mixtures were made by the extrusion method, (b) — by the method of elasto-deformational milling

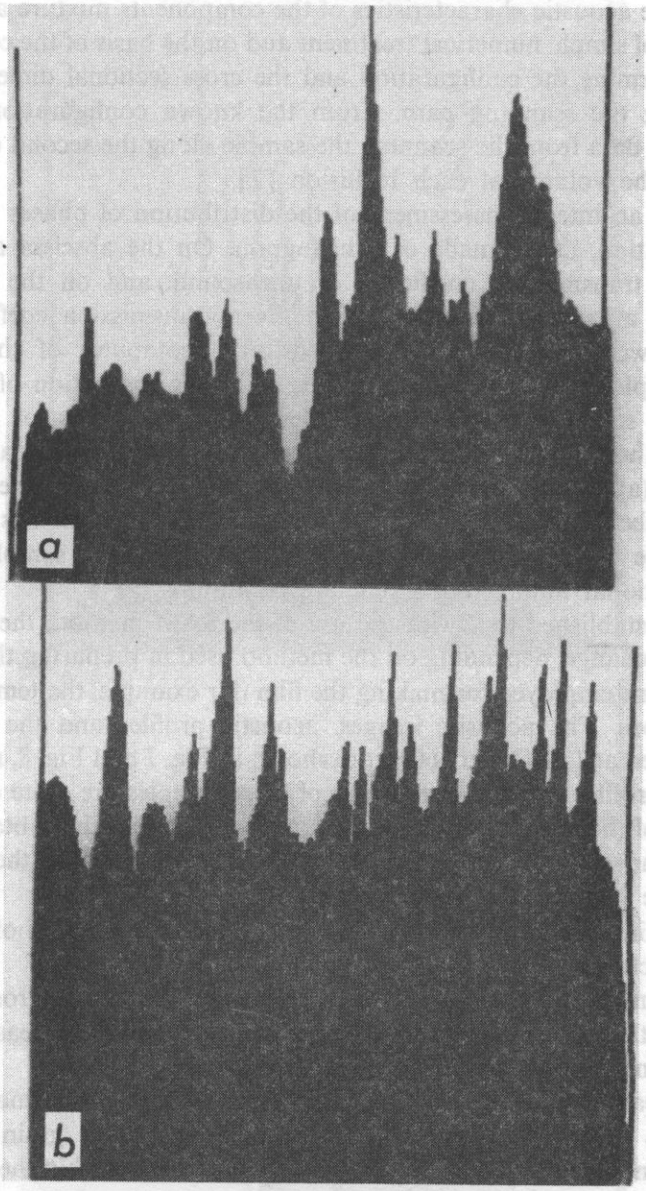


FIG. 4, 5. Profiles of polymeric composites (a) and (b)

ϱ_A density of A medium,
ϱ_B density of B medium,
C_A velocity of ultrasound in A medium,
C_B velocity of ultrasound in B medium,
A_0 amplitude of the signal from the receiving transducer in the absence of inclusion,
A amplitude of the signal when inclusion is present in the focal region.

Thus, if the acoustic characteristics of the components mixture are known, one can, by means of simple numerical treatment and on the basis of the configuration of the peaks, determine the configuration and the cross sectional dimensions of each inclusion along the scanning path. From the known configuration of the cross section and the data from the scanning the sample along the second coordinate one can calculate the volume of each inclusion [2].

To obtain an integral assessment of the distribution of phases in the samples under investigation, use is made of a histogram. On the abscissa are plotted the values for the transmission coefficient of ultrasound, and on the ordinate – a fraction of the area of the sample with a given transmission coefficient.

Fig. 6 shows the histograms obtained on a computer of the two above-mentioned samples of polymeric composite. Since the absorption of ultrasound in polystyrene is smaller than in polyethylene, the properties of inclusions are determined by the wing on the histograms to the right of the main peak. The wing on the histograms (a) is much shorter, and the histogram itself is narrower than (b). This indicates that the heterogeneity of the dimensions of the inclusions in the sample obtained by the extrusion method (b) is greater than in the sample prepared by elasto-deformational milling under cooling conditions (a).

We have established that with the use of the SAM method, the dimensions of heterogeneities change, depending on the method used in preparing the samples and on the conditions employed for making the film (for example, the temperatures used during extrusion). The acoustic images, acoustic profiles and the histograms of samples obtained at 130°C and 160°C are shown in Fig. 7 and Fig. 8, respectively. As is evident the profiles and the histograms of these samples are quite different. Thus, the histogram of the sample obtained at 160°C is wider than that obtained at 120°C. The shift and the widening of histograms to the right indicate that the dimensions of the polystyrene inclusions have increased [1].

The following characteristics are usually of interest in a morphological analysis of similar structures:

– the number of grains in the field of view of the microscope and the coordinates of their mass centers; the ratio of the sum of grain areas to the sample area, i.e. percent ratio of components within the field of view;

– the areas, perimeters and the dimensions of the grains (maximal distance between points belonging to the grain), the criterion of the grain configuration;

– the homogeneity of the distribution of grain matter within the field of view of the microscope; calculation of the degree of the homogeneity.

And if the grains form agglomerated structures, in relation to the agglomerates, of interest are all the above-mentioned characteristics as well as data on the composition of each agglomerate: which grains are part of each agglomerate, the average number and dispersion of grains in the agglomerates.

We have worked out a series of software for calculating all these characteristics. We formed the criterion of homogeneity as the relation: $G = (\bar{S}^2/N)^{1/2}$; where, $\bar{S}^2$ – dispersion of the moving average (two-dimensional distribution of the average

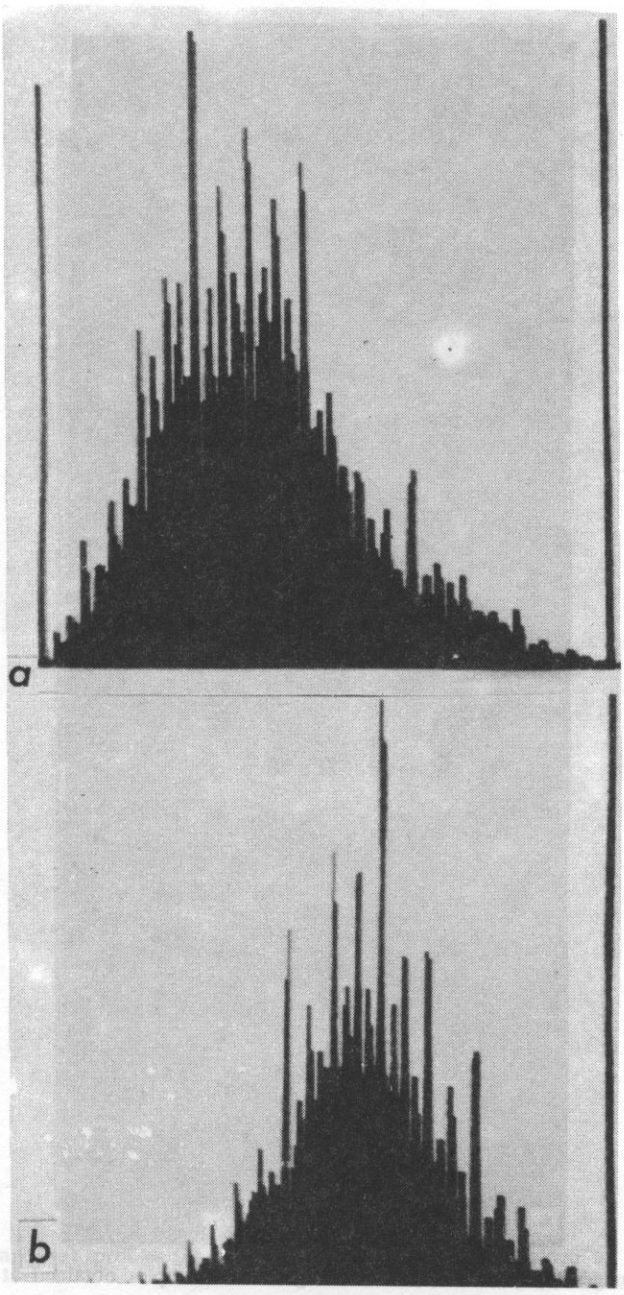


FIG. 6. Histograms of polymeric composites (a) and (b)

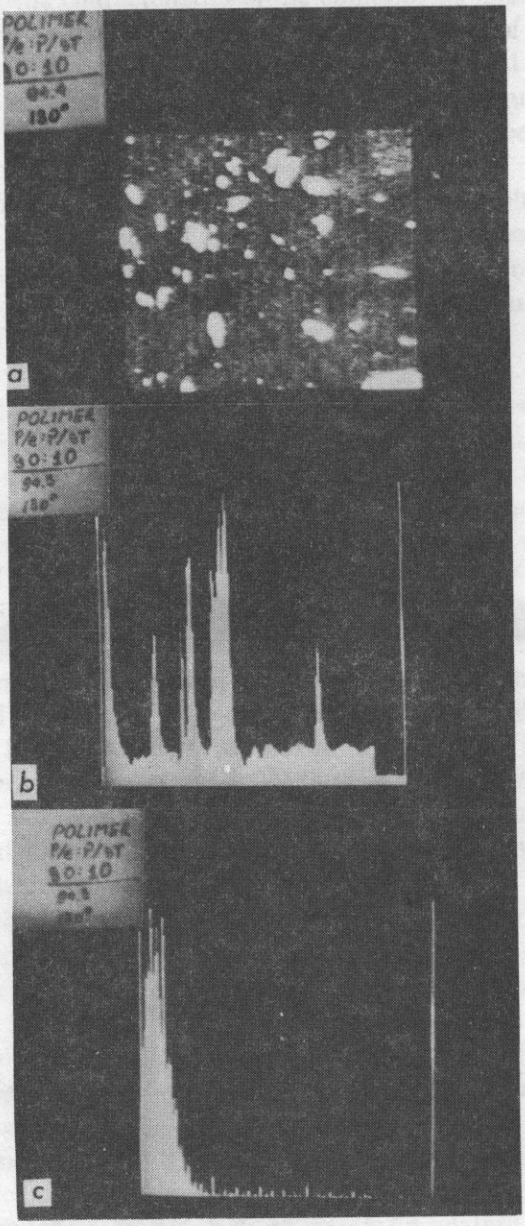


FIG. 7. Acoustic images (a), profiles (b) and histograms of samples obtained at 130°C, respectively

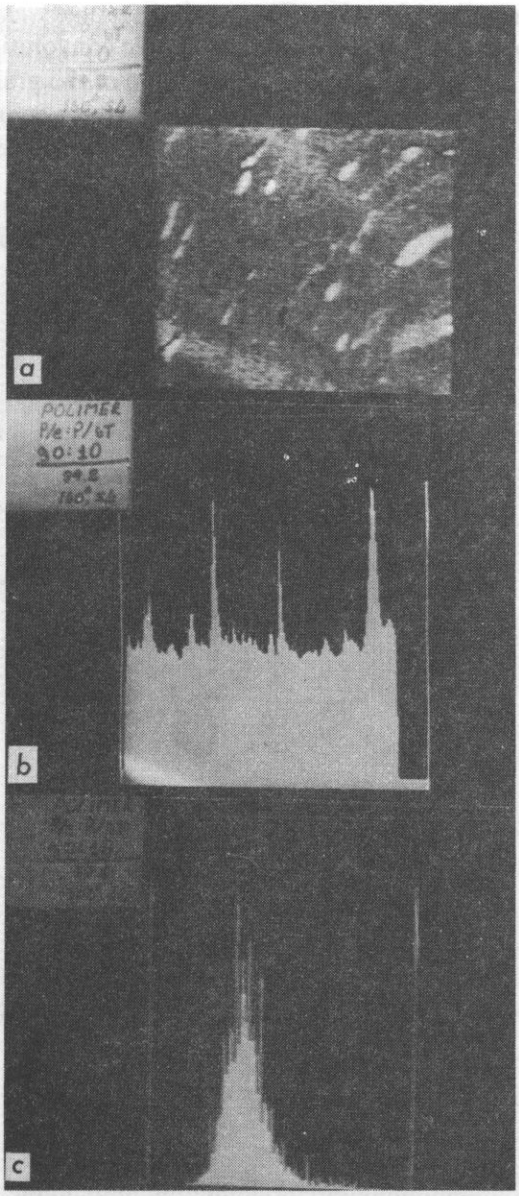


FIG. 8. Acoustic images (a), profiles (b) and histograms of samples obtained at 160°C, respectively

local value), N – normalized function. The moving average was determined from the equation $S(x, y) = q(x, y) * p(x, y)$; where $*$ – cyclic convolution, $p(x, y)$ – binary image, [$p(x, y) = 1$, if the point (x, y) belongs to one of the grains, in all other cases $p(x, y) = 0$;] $q(x, y)$ equal: $q(x, y) = 1$, if the point (x, y) belongs to the square of the area S_{sq} with the center at the zero of the coordinates system, in all other cases

FIG. 9. Result of threshold processing of acoustic image of polymeric composites

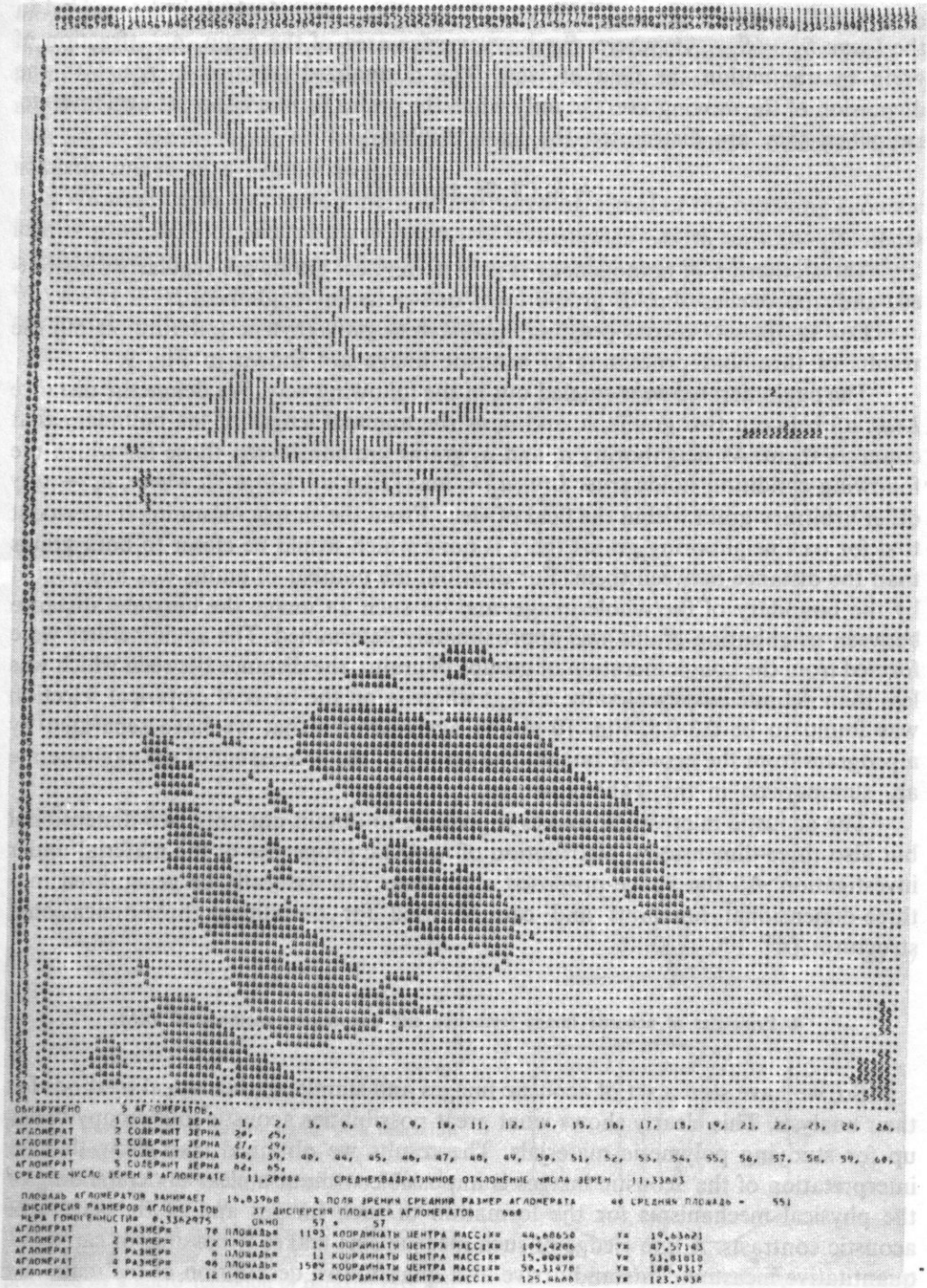


FIG. 10. Agglomerates that have been separated by a program from the acoustic image of polymeric composites

$q(x, y) = 0$. The average with respect to square, the area of which is determined in the form: $S_{sq} = \bar{S}_{\text{grain}}/\Delta$; where, $\bar{S}_{\text{grain}}$ – average area of the grain, Δ – fraction of grain matter within the field of view. The normalized function is equal to the dispersion of the moving average only when the grains form a coherent agglomerate, i.e., when they are distributed the least regularly:

$$N = [\bar{S}_n S_{sq}^2 \Delta(\Delta - 1)]^{1/2}$$

where $\bar{S}_n$ – area of the sample.

Thus, criterion of homogeneity is zero if identical grains are regularly distributed, and approaches zero if grains form one coherent agglomerate.

For samples of mixed polymers polystyrene-polyethylene (mixture 50:50) the results of threshold processing of acoustic image are shown in Fig. 9.

The agglomerates were singled out in the following way. We define the distance $L(a_i, a_j)$ between two grains, a_i and a_j as the minimal length of the boundary that connects them. As neighboring a_i and a_j grains are considered those for which the following condition holds true: $L(a_i, a_j) < \max[L(a_i, a_j), L(a_j a_k)]$, where, a_k – any other arbitrary grain within the field of view. From the simple reasoning it is evident that for true neighboring grains there is none which would be closer to both grains than the distance between them. For dilation, the number of grains was augmented by the boundary of the whole image, and on such an image the maximal distance between neighboring grains was approximately determined. The agglomerates were formed from the conglomeration of groups of grains, the distance between which was less than definite empirical value $\alpha L_{\max}$; where, α on the basis of empirical selection was found to be 0.3-0.5. Fig. 10 shows agglomerates that we have separated by a program from the acoustic image. The results of the processing of the agglomerate are summarized in the Table [3].

The acoustic microscope makes it possible to obtain not only two-dimensional but also three-dimensional distribution of acoustic properties of the material under investigation. All the above-proposed algorithms can naturally be generalized into three-dimensional functions and can be used for describing three-dimensional structures [4].

4. Principles of acoustic image formation and quantitative methods of SAM

So, we have seen a set of acoustic images and simple mathematical methods for their analysis. This clearly shows what great possibilities acoustic microscopy opens up for studying polymeric materials. The results we obtained also show that an interpretation of the acoustic images is impossible without a clear understanding of the physical mechanisms for the formation of such images and of the nature of acoustic contrasts. A knowledge of such mechanisms makes it possible to carry out quantitative measurements and arrive at a quantitative description of the materials under investigation. A comparison of the output signal of an acoustic microscope, its

amplitude and phase with the amplitude and the phase of the reference signal in liquid yields information on the velocity of sound, acoustic impedance, extinction and geometric characteristics of the sample, its thickness, curvature, and inclination angle of the surface.

So, let us now return to the subject of acoustic microscopy and consider how the acoustic image of a sample is formed.

We shall begin by considering how the outgoing signal of the receiving acoustic lens is produced in general. A piezoelectric transducer is used as a receiver. It is a linear receiver. For an electric signal to be generated in the transducer, the incident wavefront must be parallel to its surface; in other words, the acoustic rays after refraction on the lens surface must be normal with respect to the transducer's surface.

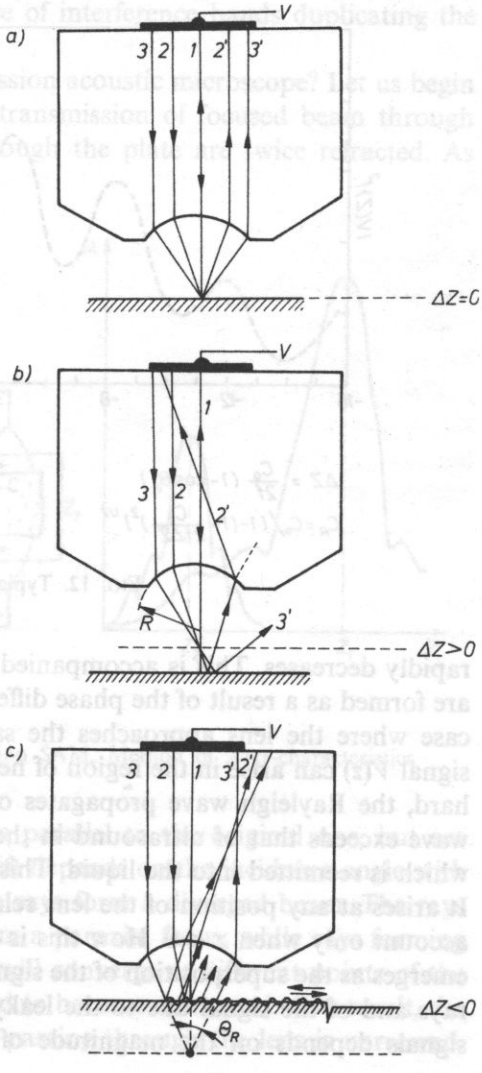
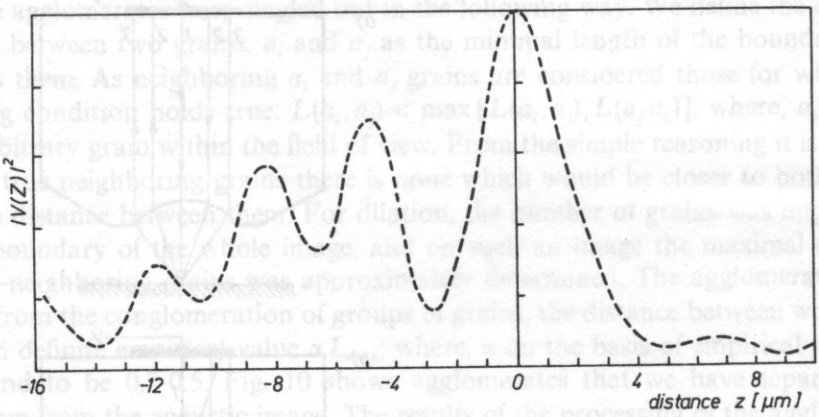


FIG. 11. Formation of output signal in a reflection SAM. z– the focal length of a lens. In case: a) $\Delta z = 0$ – output signal is formed by the integrity of all the refracted rays, b) $\Delta z > 0$ – only paraxial beams contribute to output signal, c) $\Delta z < 0$ – the output signal emerges as the superposition of the signal produced by a mirro–reflection of paraxial rays and of the signal due to the leaky surface Rayleigh wave

In addition, all incident rays reaching the transducer must be in the same phase. Otherwise, there will be interference of signals due to different rays, and the resulting signal will be weakened. First, we shall consider how a signal is formed in a reflecting microscope (Fig. 11). If the boundary of the sample lies in the focus, then the outgoing signal is determined by the integral refractive index for all incidence angles θ from $\theta = 0$ to $\theta = \theta_m$; where, θ_m – one half of the divergence angle (Fig. 11 a). If the lens is moved away from the sample (b), the cone of rays received by the transducer rapidly narrows down. Simultaneously the intensity of the signal entering the transducer decreases. Let us examine the curve describing the amplitude of the outgoing signal $V(z)$ as the function of the distance between the lens and the sample (Fig. 12). With an increase of z in the region $z > 0$ the intensity of the outgoing signal



$$\Delta Z = \frac{C_0}{2f}(1-\cos\Theta_R)$$

$$C_R = C_0/[1-(1-\frac{C_0}{2f\Delta Z})^2]^{1/2}$$

FIG. 12. Typical $V(z)$-characteristic

rapidly decreases. This is accompanied by finely divided, shallow oscillations which are formed as a result of the phase difference of rays following different paths. In the case where the lens approaches the sample, a different dependence of the output signal $V(z)$ can arise in the region of negative values of z. If the sample is sufficiently hard, the Rayleigh wave propagates on its surface. If the velocity of the Rayleigh wave exceeds that of ultrasound in the immersion liquid, a surface wave is formed which is reemitted into the liquid. This is the so-called leaky surface Rayleigh wave. It arises at any position of the lens relative to the sample. However, it is taken into account only when $z < 0$. How this is done is shown in Fig. 11c. The output signal emerges as the superposition of the signal produced by a mirror reflection of paraxial rays and of the signal due to the leaky surface wave. The phase difference of these signals depends on the magnitude of the shift z: owing to their interference the

dependence $V(z)$ has a correct system of maxima and minima. The distance between the neighbouring maxima and minima is unambiguously related to the velocity of the Rayleigh wave on the surface of the sample [5]. The formation of the leaky surface Rayleigh wave is extremely important for the formation of acoustic images in the reflection mode. The formation of the leaky surface Rayleigh wave leads to the presence of interference bands in the vicinity of sharp heterogeneities and on distorted surfaces and so on. It is the cause of the inversion effect of acoustic contrast when there is a small shift of the lens. How interference bands are formed, for instance, near the defect in the sample surface is shown in Fig. 11c. Owing to the reflection from heterogeneities there arises not only a direct but also an inverse leaky surface Rayleigh wave which can also be received by the transducer. Since the phase of this wave depends on the position of the lens' axis relative to the heterogeneity, scanning of the lens leads to the appearance of interference bands duplicating the contour of the heterogeneity.

How is the image formed in the transmission acoustic microscope? Let us begin with the refraction effect and consider the transmission of focused beam through a thin plate (Fig. 13). The rays passing through the plate are twice refracted. As

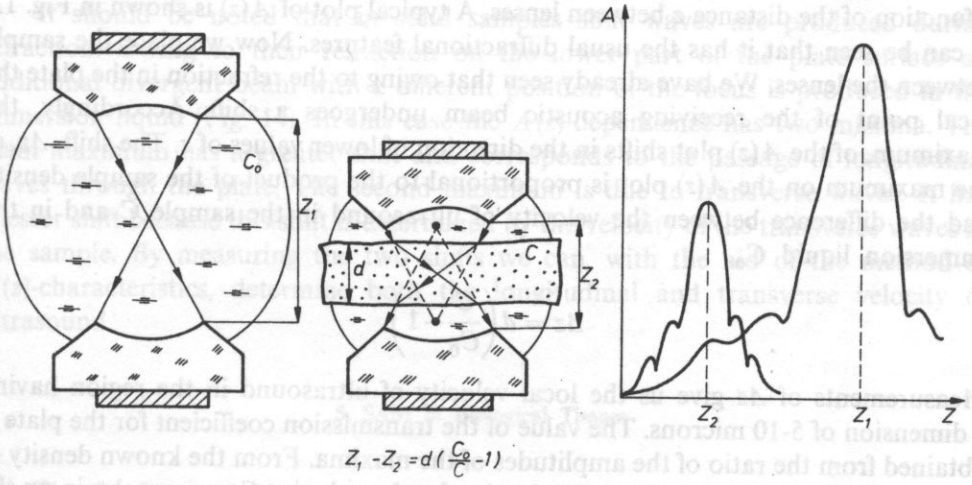


FIG. 13. Formation of output signal in transmission SAM. Method of $A(z)$-characteristics

a result, they leave the plate in the direction parallel to the original one, but are shifted relative to it. The magnitude of the shift depends on the incidence angle with the plate. After passing through the plate the rays form a diverged beam. The rays passing close to the lens axis converge to form a paraxial focus; while rays forming large incidence angles with the plate surface will converge at different points of the acoustic axis. This causes the focus to shift and to become defocused. As a result, the cone of rays received by the transducer after passing through the lens is narrowed,

and the output signal becomes less intense. Changes in the intensity of the signal due to the refraction effect will be determined by the ratio of the ultrasound velocity in the sample to that in the liquid. In an analogous way the rays cone registered by the receiving lens will become narrow owing to phase aberrations. These aberrations are explained by the fact that incident rays at different angles to the sample travel along different paths in the sample. Therefore, the rays have different phases and can mutually weaken the signals generated by them in the output piezoelectric transducer. The magnitude of phase aberrations is also determined by the difference of sound velocity in the liquid and in the sample. Therefore, the phase aberrations cause an additional contrast of acoustic images. There is yet another source of acoustic contrast, namely the reflection at both ends of the sample. This reflection depends on the magnitude of acoustic impedance at the observation point. For samples whose acoustic impedance differs slightly from that of water, such mechanism of acoustic contrast is of little significance. For such samples as polymers and biological tissues the formation of acoustic images is mainly due to the difference in local attenuation of sound.

In order to establish quantitatively the local mechanical properties of the sample we have developed a new special method of $A(z)$-characteristics [4]. Plot $A(z)$ describes the dependence of the output signal A of the receiving lens as a function of the distance z between lenses. A typical plot of $A(z)$ is shown in Fig. 13. It can be seen that it has the usual diffractional features. Now we place the sample between the lenses. We have already seen that owing to the refraction in the plate the focal point of the receiving acoustic beam undergoes a shift. Accordingly, the maximum, of the $A(z)$ plot shifts in the direction of lower values of z. The shift, Δz, of the maximum on the $A(z)$ plot is proportional to the product of the sample density and the difference between the velocity of ultrasound in the sample C and in the immersion liquid C_0.

$$\Delta z = d\left(\frac{C}{C_0} - 1\right).$$

Measurements of Δz give us the local velocity of ultrasound in the region having a dimension of 5-10 microns. The value of the transmission coefficient for the plate is obtained from the ratio of the amplitudes of the maxima. From the known density of the sample ϱ and by using the derived value for the velocity C, we can obtain on the basis of the known values of impedance and the transmission coefficient – the local coefficient of ultrasound attenuation. By using the method of $A(z)$–characteristics we measured the velocity of ultrasound in thin polymer films. For this we used films made of polyarylate and its block copolymer silar. The thickness of the films varied in the range 9–17 μm. For the samples of polyarylate the velocity of ultrasound was found to be 2,710 and for silar 1,020 km/s, respectively. A knowledge of these parameters makes it possible, in principle, to investigate the distribution of the local physico-mechanical properties of these materials within a wide range of concentrations.

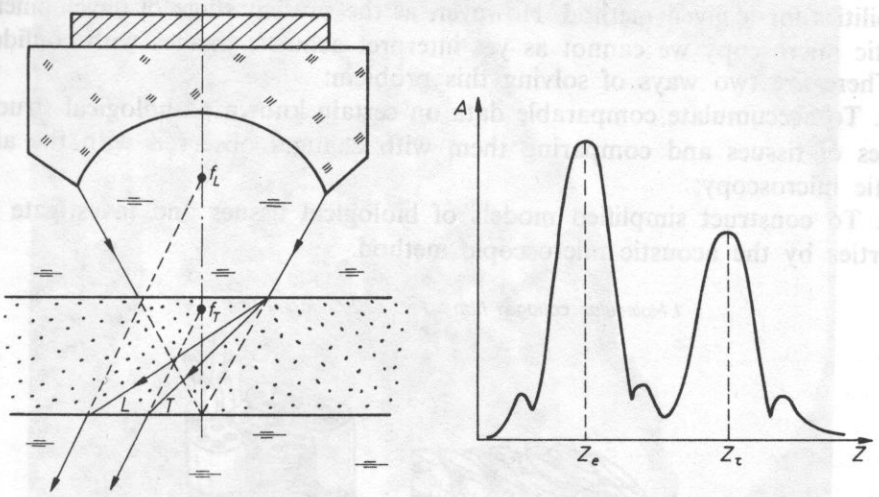


FIG. 14. Formation of two maxima of microscope output signal: for longitudinal waves (L) and shear waves (T)

It should be noted that in solid samples shift waves are produced during refraction. Owing to their refraction on the lower part of the plate surface an additional divergent beam with a different position of the focus is produced in the immersion liquid (Fig. 14). In this case the $A(z)$-dependence has two maxima. The main maximum has a greater shift and corresponds to the passage of longitudinal waves through the plate. The second maximum is due to transverse waves. It has a lesser shift because this shift is determined by the velocity of the transverse waves in the sample. By measuring the two shifts we can, with the aid of the method of $A(z)$-characteristics, determine both the longitudinal and transverse velocity of ultrasound.

5. SAM of Biological Tissues

The development of the quantitative methods of scanning acoustic microscopy opens up great possibilities for research in one more area: the study of the viscoelastic properties of biological materials on a histological and cytological level. For already more than 10 years scanning acoustic microscopy has been used in biological research.

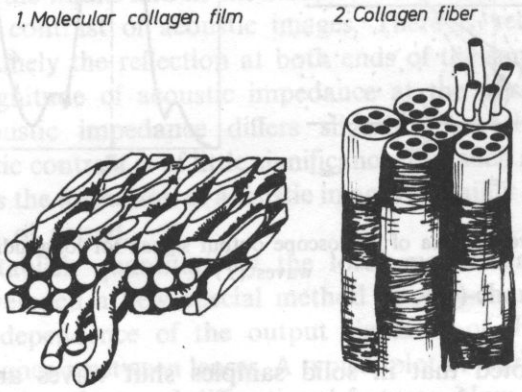
The experience gained to date in interpretation acoustic images of collagen tissues shows that the contrast provided by acoustic microscopy depends on the size of the collagen fibers, their density and water-content, and the degree of order and orientation of the fibers relative to the observation axis (Fig. 15). The fact that there is such a large number of parameters affecting the contrast opens up wide diagnostic

possibilities for a given method. However, at the present stage of development of acoustic microscopy we cannot as yet interpret acoustic images with confidence.

There are two ways of solving this problem:

1. To acccumulate comparable data on certain known pathological structural changes of tissues and comparing them with changes observed with the aid of acoustic microscopy;

2. To construct simplified models of biological tissues and investigate their properties by the acoustic microscopic method.



Materials
Molecular collagen films were formed by drying acetic acid solution of molecular pig skin collagen at 4°C during 48 hours. Collagen fibers from rat tail tendon were under study. 10 μ collagen fibers were prepared using kryotome Leitz 1321.

FIG. 15. Structure of various collagen fibers

The first approach requires extensive acoustic microscopic studies of pathological tissues.

In our work we devoted most of our attention to a study of the viscoelastic properties of models of collagen-rich tissues. First we studied the acoustic images of normal tissues. (Fig. 16). This figure shows the acoustic and optical images of a lateral slice of human skin. The acoustic image was obtained in our laboratory with the aid of transmission acoustic microscope [1] operating on a frequency of 450 HMz and having a resolution of 3μm. The sample, 7μm thick, was in a non-fixed position and was not treated with a dye. The picture gives a good view of three layers. The first is a dark band in the surface of skin is the corneous layer. It consists of a densely packed keratin protein. Then follows an acoustically transparent layer of epithelian cells. And finally there is a structurally heterogeneous derma, which is a dense network of collageneous and elastinic fibers submerged into an intercellular liquid.

There is a general similarity of acoustic and optical picture. Outlines of the objects and boundaries of shape inhomogeneities coincide on both images. One may suppose that the contrast of acoustic images on this picture is determined by the

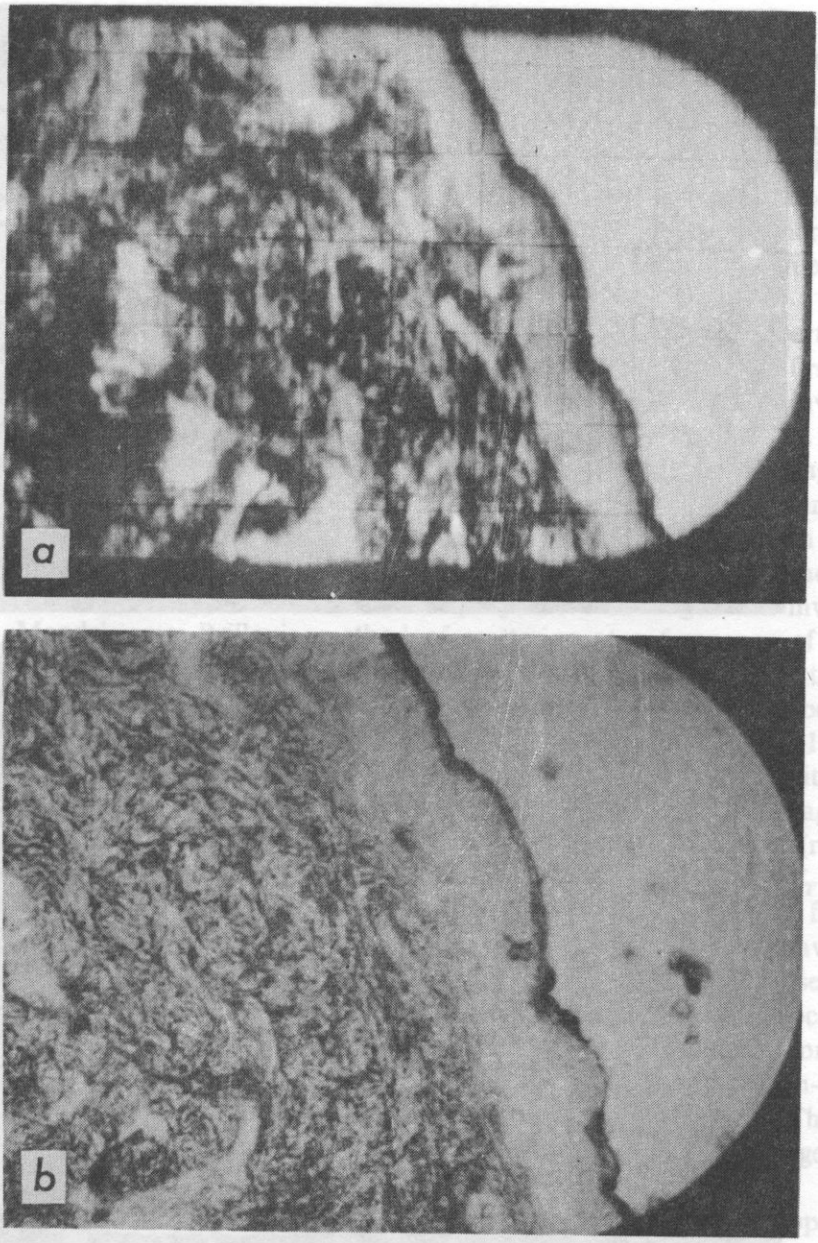


FIG. 16. Acoustic (a) and optical (b) images of a slice of human skin

magnitude of the local coefficient of attenuation of ultrasound in soft tissues correlates with the concentration of fibrilar protein in them. Therefore, a system of dark and light regions in acoustic image should reflect the distribution of collagen (or keratin) in the sample. For a quantitative evaluation we can use the acoustic profile (Fig. 17). In this Figure we can see a liver slice image with bubble air and two vessels

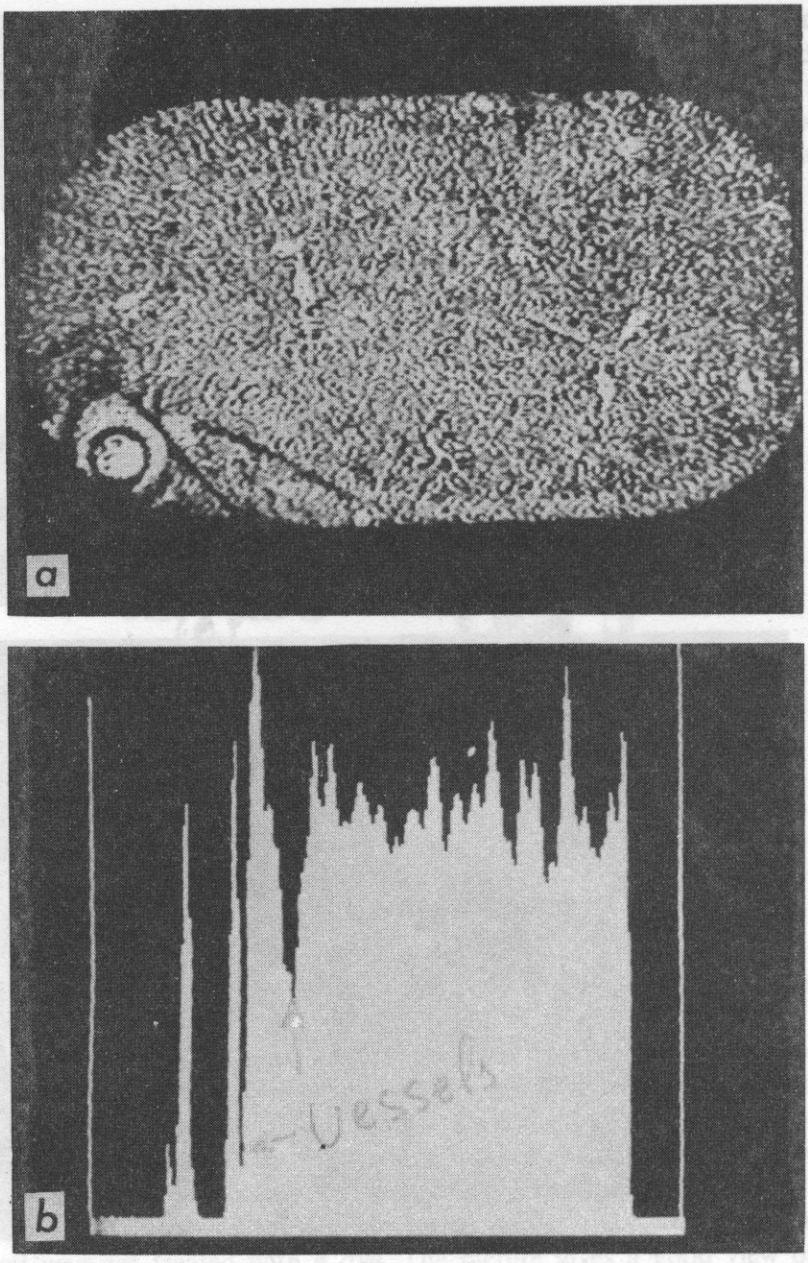


FIG. 17. Acoustic image and profile of liver slice (with the air bubble and two vessels)

and profile, which can give us a precise information. Using this profile as in polymers case we can determine the local concentration of collagen and arrive at a quantitative description of collagen inside the tissue. And herein lies our main difficulty: we can provide a quantitative description of the distribution but we have yet to find a biologist or medical specialist who would be interested in such a description.

Figure 18 shows acoustic and optical images of a longitudinal and a transverse slice of a human sclera. The poor definition of the optical images is due to the fact that the sample was in a non-fixed state and undyed. In both the acoustic and optical images one can see packed bundles of collagen fibers. On the optical image one can see only the outline of the boundaries of the fibers; while on the acoustic image one can see different gradations of grey coloring, which makes it possible to decipher the density of the fibers and the local concentration of collagen. Numerically the concentration of collagen can be estimated by using acoustic profiles as was mentioned above.

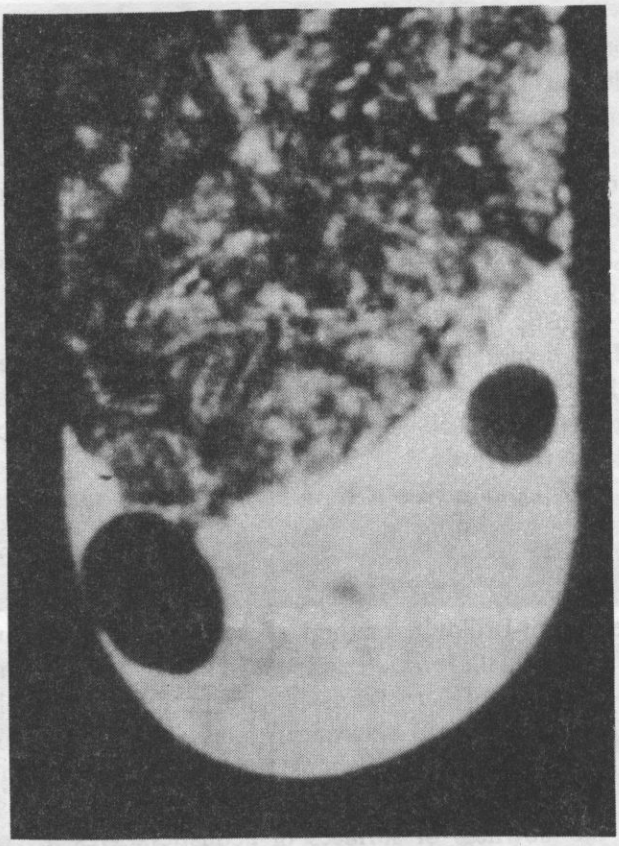
Since the acoustic properties of tissues are determined by the properties of the components of the tissues and their mutual orientation, we are interested first of all in the mechanical properties of the components, collagen in particular. We investigated mechanical properties of collagen-fibers with our acoustic microscope. Different authors (Fig. 19) have measured velocity of ultrasound along the entire length of the fiber ($V_s = 2.1$ km/s) and across it ($V_s = 1.73$ km/s) by means of a scanning laser acoustic microscope operating at a frequency of 100 MHz (wavelength in water $\sim 15\mu$m). These results indicate a strong anisotropy of the elastic modules of collagen. The anisotropy of the viscoelastic properties of collagen was investigated by the Mandelstamm-Brillouin method of scattering at a frequency of 10 GHz (wavelength in water $\sim 0.15\mu$m). Measurements of the longitudinal and transverse waves yielded five values of elastic constants for partially dried collagen fiber, on the assumption that collagen is a transversely masotrophic medium (Fig. 19).

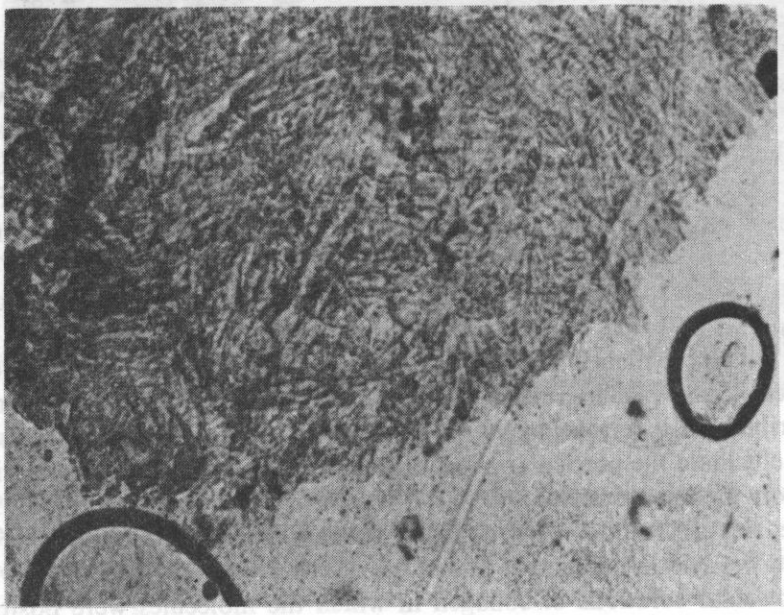
By using the values of these constants we calculated the velocity of the longitudinal and transverse waves propagating at different angles in collagen fiber. Fig. 20 shows the slowness of surfaces of acoustic waves in collagen (reciprocal velocities $1/V_L$, $1/V_T$; where V_L – phase velocity of the longitudinal wave, V_T – phase velocity of the shear wave. Here, the z axis corresponds to the axis of the fiber. The results show a strong anisotropy in the velocity of the longitudinal wave (phase velocity of the longitudinal wave along the fiber $V_{L\parallel} = 3.64$ km/s, phase velocity across the fiber $V_{L\perp} = 2.94$ km/s) and a rather weak anisotropy in the velocity of the shear wave ($V_{T\parallel} = V_{T\perp} = 1.563$ km/s, $V_T = 1.639$ km/s at 49° angle). For natural collagen fiber the transverse mode in the spectrum of Mandelstamm-Brillouin scattering was absent; i.e., there was no propagation of the shear waves. This means that there is a strong dependence of the viscoelastic properties of collagen on its saturation with water.

All the above-mentioned factors – the dependence of the elastic properties on the direction and packing density of fibers and their size, and water concentration – contribute to a certain extent to the contrast of the acoustic images of tissues.

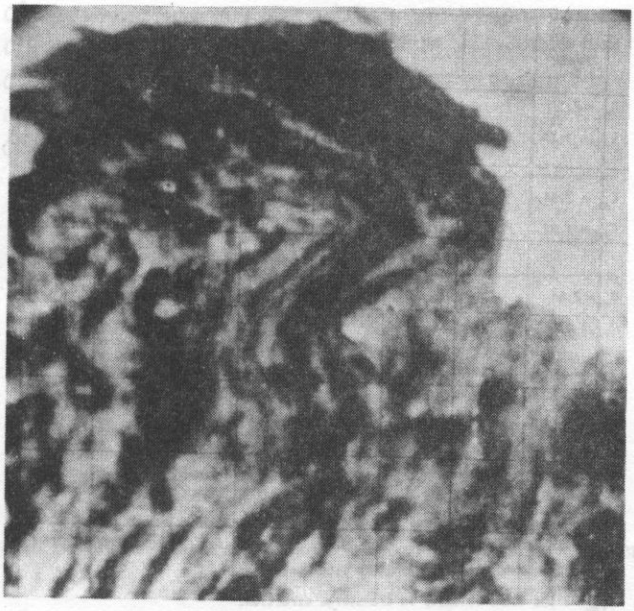
To elucidate the particular effect of each of these parameters we carried out our studies on modeled systems.

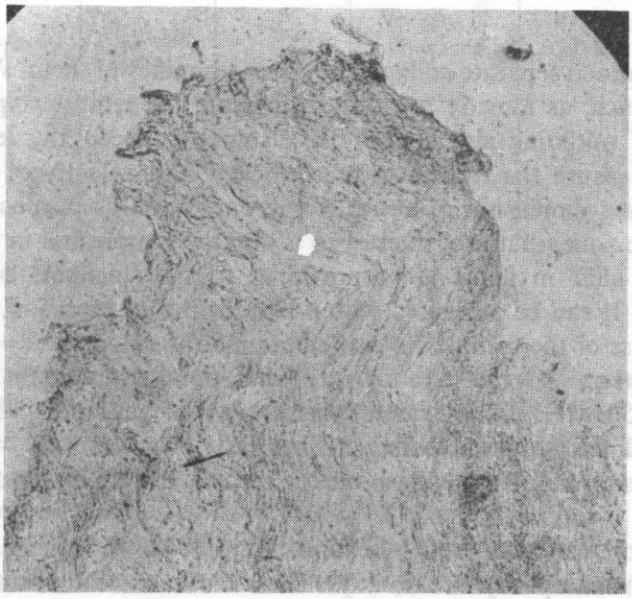
The simplest model of a collagen tissue slice is a homogeneous layer. We examined the gelatin-collagen layers in the dehydrated state. The layers were formed from molecular solutions of collagen in which the molecules were positioned with their long axis being in the plane of the layer.





a





b

FIG. 18. Acoustic and optical images of a longitudinal (a) and a transverse (b) slice of a human sclera

SOUND VELOSITY, KM/SEC	THE METHOD	AUTHORS	OBJECT
C_{ax}= 2.11 C_{rad}= 1.73	SLAM 100 MHz	S.LEES, J.HEELEY, J.AHERN	FIBERS OF RAT'S TAIL TENDON
C_{ax} = 2.64 C_{rad}= 1.89	BRILLOUIN LIGHT SCATTERING 10 GHz	HARLEY, D.JAMES, A.MILLER	FIBERS OF RAT'S TAIL TENDON
C_{ax}= 2.64 C_{rad}= 1.89	BRILLOUIN LIGHT SCATTERING 10 GHz	CUSAK S. MILLER A.	FIBERS OF RAT'S TAIL TENDON
C_{ax} = 1.82 C_{rad}= 1.69	IMPULS ECHO technique 10 MHz	S.LEES, J.HEELEY P.CLEARY	FIBERS OF BOVINE TI- BIA MATRIX
C = 1.73 without specifing direction of propagation	SLAM 100 MHz	S.GOSS W.D.O'BRIEN	FIBERS OF MOUSE TAIL TENDON

C_{ax} is the sound velocity along the fiber axis.

C_{rad} is the sound velocity in transverse direction.

FIG. 19

Next we examined separate collagen fibers and bundles of them. With the aid of the $A(z)$ plots which, as already mentioned, indicate the relationship between the amplitude of the outgoing signal and the distance between the lenses, we can simultaneously measure the transmission coefficient and velocity of sound within a local zone having dimension of about 5 microns. For this purpose we plot the $A(z)$-graph for the zone (that is of interest to us) on the sample and in the absence of the sample. The shift in focus is proportional to the thickness of the sample: $\Delta z = d(C/C_0 - 1)$. It can be measured with a high degree of precision, and, since we know the thickness of the sample we can determine the velocity: $C = C_0(\Delta z/d + 1)$.

We have determined the velocity and ultrasound attenuation coefficient at a frequency of 450 MHz in a 2-micron-thick layer of gelatin, a 10-micron-thick collagen film and in 10-micron collagen fiber [4].

A gelatin film of uniform thickness can be separated from a photo film in water at $t = 90°C$.

Collagen films were prepared by evaporating at 4°C a molecular solution of pig's skin collagen for a period of two days. The molecular solution was obtained by extracting collagen from pieces of pig's skin with a 0.1 M solution of glacial acetic acid followed by precipitation of protein with cooled ethanol.

Collagenous fibers were separated from the longitudinal slices of the tendon from a rat's tail. The slices were obtained on a refrigerated Leitz microtome, the temperature of the table and the knife $T = -20°C$.

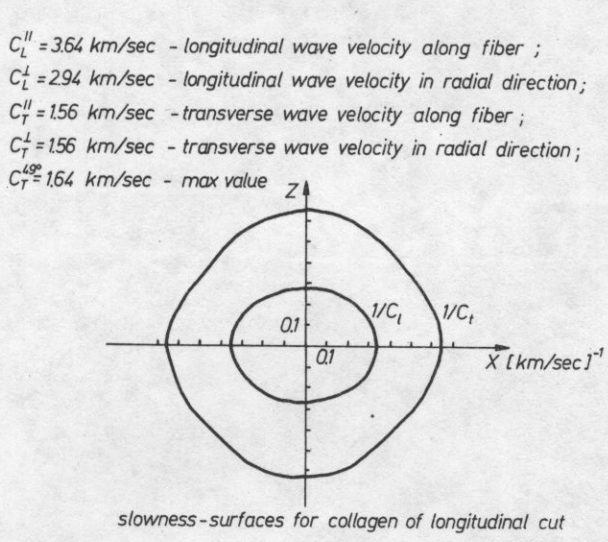


FIG. 20. Grafics of slowness-surfaces of acoustic waves in collagen

Table 1

Sample	Thickness [μm]	Density [g/cm^3]	Velocity of ultrasound [km/s]	Attenuation coef. (10^{-17}) [s^2/cm]
Gelatin	2	1.2	2.9±0.3	420±40
Collagen film	10	1.3	2.4±0.3	445±45
Collagen fiber	10	1.35	2.8±0.3	660±60

In Tab. 1 there are summarized the values for the ultrasound velocity and attenuation coefficients for these samples measured at 450 MHz.

The proposed approach can be regarded only as an initial stage in obtaining the quantitative characteristics of the local viscoelastic properties of tissues. However, even now this method enables us to estimate the effect of a number of parameters, such as the pH, temperature, etc, on the biochemical properties of tissues.

In this report I have described the results of first steps of acoustic microscopic research performed in our Center in only two fields: the physics of polymeric mixtures and biophysics. The various experimental results obtained with the aid of the transmission microscope, and the wide possibility of obtaining quantitative data with this microscope, in our opinion, convincingly demonstrate the rich scientific potentialities inherent in the principle of raster acoustic microscopy.

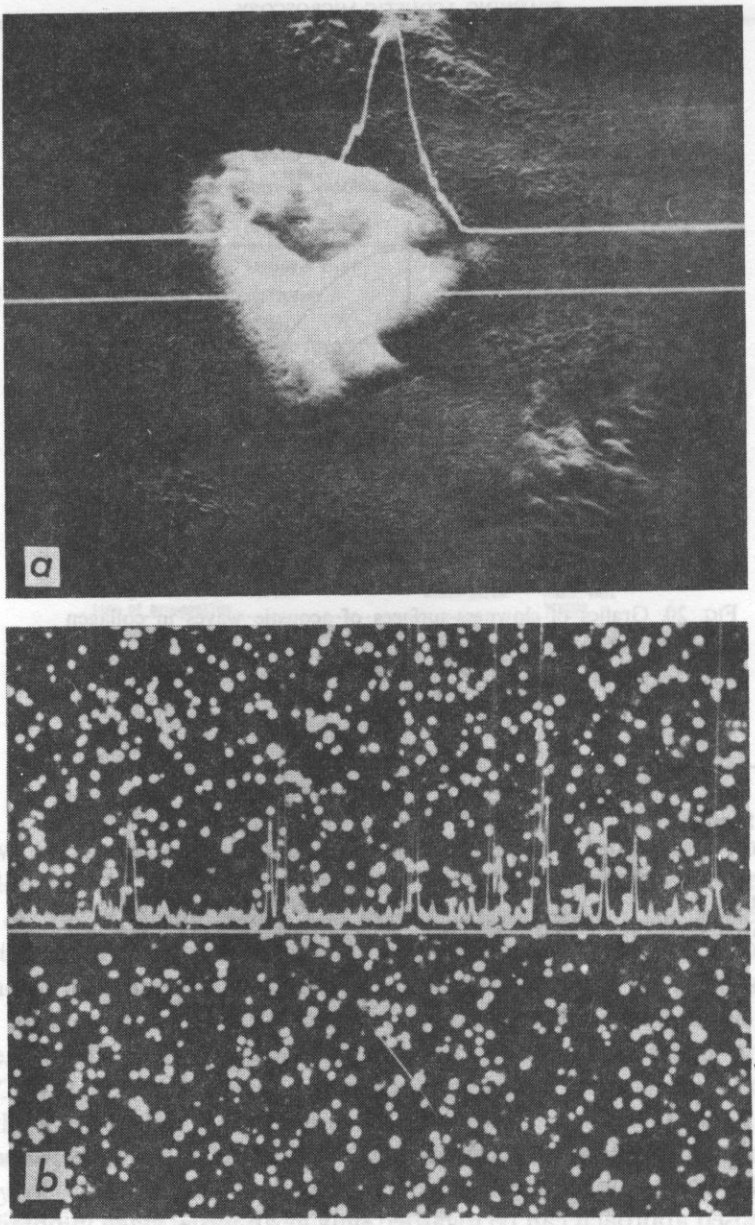
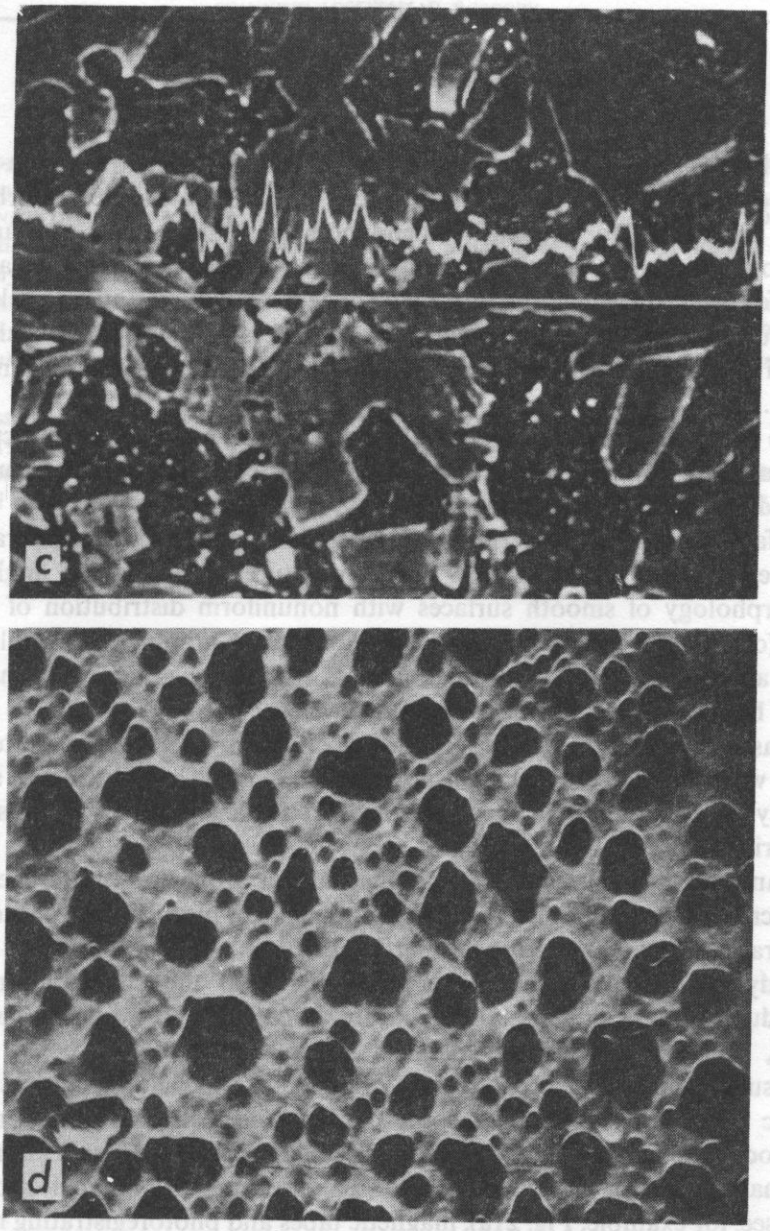


FIG. 21. Acoustic images of different materials. a) laminated structure ($SiO_2-Si-SiO_2$-substrate-Si) $f = 0.4$ GHz. Degree of penetration $\Delta z = -10$ μm. The microscope is focused on the boundary between the substrate and the layer to a depth of 10 μm from the sample surface reflecting optical rays. The acoustic image shows a defect at the interface. The same is confirmed by the one-dimensional profile of the refracted signal amplitude along the scanning line, marked on the photo; b) a photo film, representing the celluloid base, coated by gelatin matrix with silver crystals of $2 \div 3$ μm. Frequency $f = 1.3$ GHz. The image was obtained in the layer close to the surface $\Delta z = -0.5$ μm. One of the parameters determining the quality of photomaterials is the homogeneity degree of silver grains distribution in the matrix, their relative orientation, the absence of agglomerates etc. Optical methods do not allow to carry out such an analysis due to the specific nature of the object. Such a control can be timely realized by SAM methods. Since the acoustic properties of silver crystals differ greatly from those of the matrix and the immersion



media, the image obtained is highly contrasting and its information content is very high. The profile of the refracted signal amplitude along the scanning line, marked on the photo, was applied to the image; c) high-temperature superconducting ceramics. $f = 1.3$ GHz, focal length $\Delta z = -2\mu$m. At the depth of 2 μm of optically non-transparent material one can clearly see the structure of ceramics with sharp boundaries of pores between crystals and the inner structure of pores. The profile of the refracted signal amplitude along the scanning line, marked on the photo, was applied to the image; d) polymer composite (polyacrylate-silar-rubber). $f = 1.3$ GHz. Focal length $\Delta z = 5$ μm. The photo clearly displays the distribution of the two types of fillers (large-sized and finely divided) in the polymer matrix at the depth of 5 μm (Acoustic images were obtained by the specialists from the Center of Acoustic Microscopy, the USSR Academy of Sciences, O. V. Kolosov, L. F. Matsiev and T. A. Senyushkina)

6. Practical Applications of Acoustic Microscopy

The method of acoustic microscopy turns out to be very sensitive to the presence in the object of any kind of inhomogeneities as well as to the disturbance of uniformity, since due to acoustic impendances mismatch there appear strong reflections on boundaries. At the present time the acoustic microscopy allows to reveal the following defects: adhesion disturbance, exfoliations microcracks pores, foreign inclusions, deviations from the given thickness of the layer in multilayered system and coatings, technological deviations of sizes, orientation and grain distribution.

So, the following trends in the development of the methods of refraction and transmission acoustic microscopy for the investigation of surface and subsurface structures, different materials seem to us very perspective:

– surface topography (determination of steps height, the width of cracks and the character of stress fields around them, curvature radius, wedge angles etc.);

– morphology of smooth surfaces with nonuniform distribution of acoustic properties (characterization of separate components of granular and multilayered structures, acoustic images of internal planes, structures, grains, the analysis of thin-filmed heterogeneous objects);

– measurements of local values of the velocity of propagation and attenuation of Rayleigh waves in materials by acoustic microscopy using cylindrical and spherical lenses; study of distribution of local anisotropic elastic properties of crystals and other materials;

– quantitative measurements of nonmechanical properties by the methods of acoustic microscopy, among them local measurements of piezoelectric, photoelectric, high-temperature superconducting properties of films;

– study of dynamic phenomena connected with the change in materials properties due to physical factors (temperature, UV-, IR-, SHF-radiation) as well as mechanical, chemical and pharmacological factors.

The results, which have been recently obtained in this field, allow to state that the acoustic microscopy methods become the working instrument of scientist and engineer. Today we can speak about the first concrete practical applications of the method: quality control of semiconductor engineering and microelectronics instruments (see, for example, Fig. 21a), magnetic tapes and photoregistrating materials (Fig. 21b), technological and physicomechanical control of alloys, ceramics (Fig. 21c), polymeric composite materials with optically close or nontransparent components (Fig. 21d), poly- or monocrystalline films, adhesive joints, seams, reinforcing and protective coatings, antireflecting and paint and varnish layers, biomedical objects and some others.

References

[1] *Proceedings of the first joint Soviet-West Germany Int. symp. on microscope photometry and acoustic microscopy in science*, [Ed.] R. G. Mayev, M. Hoppe, Moscow 1985.

[2] N. S. Enikolopian, O. V. Kolosov, E. Yu. Lagutenkova, R. Gr. Mayev, D. D. Novikov, *Investigations of heterogeneity of polymeric composite materials by methods of acoustic microscopy*, Soviet physics, Doklady **292**, 6, 1418–1422 (1987).

[3] L. A. Piruzyan, O. V. Kolosov, R. Gr. Mayer, V. M. Levin, T. A. Senyushkina, *Acoustic microscopy of organic and biological materials*, Soviet Physics. Doklady, **280**, 4–6, 1115–1117 (1985).

[4] *Proc. of the 5-th Soviet-West Germany symp. on new methods for microscopy in biology and medicine*, Moscow 1987, 141–151.

[5] A. Briggs, *An introduction to scanning acoustic microscopy*, Microscopy handbooks 12, Alden Press Oxford 1985.

[6] O. V. Kolosov, V. M. Levin, R. Gr. Mayev, T. A. Senyushkina, *The use of acoustic microscopy for biological tissue characterization*, Ultrasound in Medicine and Biology, **13**, 8, 477–483 (1987).

Received on January 25, 1988

ARCHIVES OF ACOUSTICS
13, 1-2, 45-53 (1988)

AN ACOUSTIC MICROPROBE

JERZY LITNIEWSKI

Department of Ultrasonics, IFTR, Polish Academy of Sciences
(00-049 Warszawa, ul. Świętokrzyska 21)

In the paper the structure of an acoustic microprobe has been presented of which the resolution does not depend on the receiving transducer dimensions. With this microprobe the acoustic pressure amplitude can be measured with the resolution below the wave length in liquid media and at the surface of solids, in the frequency range from 30 to 40 MHz. The method of finding the resolving power of the microprobe by measuring the response function for a point source has been described.

Resolution of the microprobe has been estimated as better than 25 μm what is about 1/2 of the wave length in water and 1/7 of the longitudinal wave length in such materials as aluminium or glass.

The directional response pattern which is necessary for taking measurements at the non-planar surfaces of solids, has been found.

The possibility of applying the microprobe to measuring the pressure amplitude distribution at the surface of lenses used in acoustic microscopy has been demonstrated.

W pracy przedstawiona została budowa sondy akustycznej, której rozdzielczość nie zależy od rozmiarów przetwornika odbiorczego, umożliwiającej wykonywanie pomiarów amplitudy ciśnienia akustycznego z rozdzielczością poniżej długości fali w ośrodkach ciekłych i na powierzchni ciał stałych w granicach częstotliwości 30 do 40 MHz. Opisana została metoda znajdowania rozdzielczości mikrosondy poprzez znalezienie funkcji odpowiedzi sondy na źródło punktowe.

Rozdzielczość mikrosondy oszacowana została na lepszą od 25 μm co stanowi ok. 1/2 długości fali w wodzie i ok. 1/7 długości fali podłużnej w takich materiałach jak aluminium, szkło.

Znaleziona została charakterystyka kierunkowa sondy niezbędna do uwzględniania przy prowadzeniu pomiarów na niepłaskich powierzchniach ciał stałych.

Pokazana została możliwość zastosowania mikrosondy do określania rozkładu amplitudy ciśnienia na powierzchniach soczewek używanych w mikroskopii akustycznej.

1. Introduction

Measurements of pressure distribution of the ultrasound field, for the field shape to be exactly reconstructed, should be carried out with the resolution of at least $\lambda/2$. In the case of the measurements in liquids, when conventional methods are applied

(e.g. hydrophones), this becomes practically impossible for the frequencies over 10 MHz. For the measurements of pressure distribution (vibration amplitudes) at the surfaces of solids the holographic method [1] seems ideal. It needs, however, a large power ultrasound field.

There exists a possibility of constructing a very sensitive acoustic probe of which the resolving power would not be determined by the dimensions of the piezoelectric receiving transducer.

In 1980 W. DÜRR, D. A. SINCLAIR and E. A. ASH proposed a method for measuring the acoustic pressure in a liquid and obtained the resolution of $\lambda/4$ for the frequency 4.5 MHz. The similar concept of measurement has been applied in the spike-type acoustic microprobe constructed by me (Fig. 1). Such microprobe would make it possible to measure the acoustic pressure distribution with the resolution of $\lambda/2$ in fluid media and at the surface of solids in the frequency range from 30 to 40 MHz (This corresponds to the wave lengths of 40–50 μm in water and to the longitudinal wave length equal to about 170 μm in such materials as glass and aluminium).

2. Structure of the spike-type acoustic probe

The scheme of the probe is shown in the Fig. 1. It consists of an aluminium cone-shaped spike, with the lithium iodate transducer, operating at resonance frequency of 36 MHz, glued to its base. The spike end is spherically shaped, with radius estimated at 40 μm (measured with the microscopy, by comparison).

The vertical angle of the cone equals $\alpha = 40°$. The ultrasound wave falling on the spike parallelly to its symmetry axis is totally reflected from the side (the angle of total internal reflection of the longitudinal wave on the boundary H_2O-Al is 13.5°). Only the spike end transmits the acoustic signal to the surface of the transducer.

From geometrical estimations it follows that only the region with perimeter of about 18 μm is active. As it will occur further this magnitude is close to the microprobe resolution.

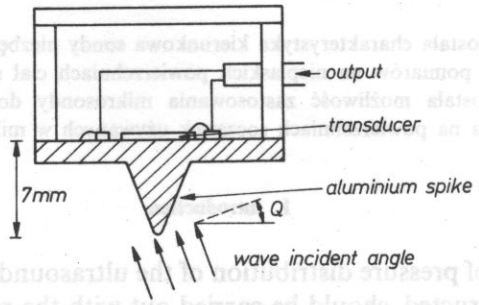


FIG. 1. A schematic diagram of the acoustic microprobe

The spherical shape of the spike end limits its active region on one hand, and allows for better measurement accuracy for non-axially falling waves on the other. These features are indispensable for the measurements of pressure distribution at non-planar surfaces of solids.

3. Measurements of the resolving power of the microprobe

To determine the resolving power of the microprobe the following method has been applied. As it is most generally known the image of an object (u) can be expressed as a convolution of an object function (u_p) and a response function for a point source (image of a point) (h)

$$u(x, y) = u_p(x, y) \otimes h(x, y). \tag{1}$$

If the function $h(x, y)$ could be found then it would be possible to calculate the resolving power of the system from this function.

As the imaged object the pressure distribution in the focal plane of a lens used in acoustic microscopy has been used. This kind of distribution is described by the function type $J_1(x)/x$ [3].

In the calculations I have assumed that the object function is equal to the function describing the theoretical distribution. Actually, this function differs from its theoretical description what implies that the value of resolution, found in this way, can only be smaller than the real resolution.

The theoretical pressure distribution in the focal plane has been approximated with the Gauss function

$$u_p = e^{-(x^2+y^2)/b^2}. \tag{2}$$

This results in considerable simplifications in calculations and have negligible influence on the obtained resolving power of the probe.

The value of b in (2) has been chosen so that the Gauss function would assume the value $1/e \simeq 0.368$ for the same value of the coordinate $(x^2+y^2)^{1/2}$ for which the function which describes the theoretical pressure distribution in the focal plane is also equal to $1/e$. The b value found in this way is

$$b = 0.82\frac{\lambda f}{d},$$

where f – focal length, d – lens diameter and λ – wave length in water, equal to 38 μm for the frequency 40 MHz.

Analogically, the measured pressure distribution is approximated with the Gauss function

$$u = e^{-(x^2+y^2)/a^2}, \tag{3}$$

where the value of a shoud be found from measurements.

To find the function $h(x, y)$, the functions (2) and (3) were substituted to (1) and then both sides of (1) were subjected to Fourier transform. Taking advantage of the fact that the Fourier transform of the convolution of two functions is equal to the product of the Fourier transforms of these functions and of the fact that the Fourier transform of the Gauss function is also the Gauss function (3), the Fourier transform of the function $h(x, y)$ can be found in a simple way and the function itself can be calculated with the inverse transform

$$h(x, y) = \frac{a^2}{b^2} \frac{1}{a^2 - b^2} e^{-(x^2+y^2)(a^2-b^2)}. \tag{4}$$

With the use of the Rayleigh criterion of resolution, from the impulse response function $h(x, y)$ the resolving power of the probe δ can be found. This resolving power should be equal to the diameter of a circle on which the function h assumes the value equal to 0.81 of the value of this function in the point ($x = 0$, $y = 0$). The resolution found in this way equals the resolution denoted as the value of coordinate $(x^2 + y^2)^{1/2}$ for which the function h would assume the value zero for the first time, if the calculations for the distribution of type $J_1(x)/x$ were used instead of the Gauss function.

$$\delta = 0.92(a^2 - b^2)^{1/2}. \tag{5}$$

To find the numerical values of the resolving power, the pressure distribution in the lens focus has been found with the use of the microprobe. In the measurement the system of an acoustic microscope working at the frequency of 40 MHz and at the impulse length 0.5 μs [4] was used. The investigated object (lens) has been connected to the emitting system. The signals received by the microprobe were transmitted to the receiving system of the microscope. Between the lens and the probe there was water. After finding the lens focus the microprobe has been shifted in the focal plane with the mechanical shifting system of the microscope, and the received signal has been recorded with the XY plotter.

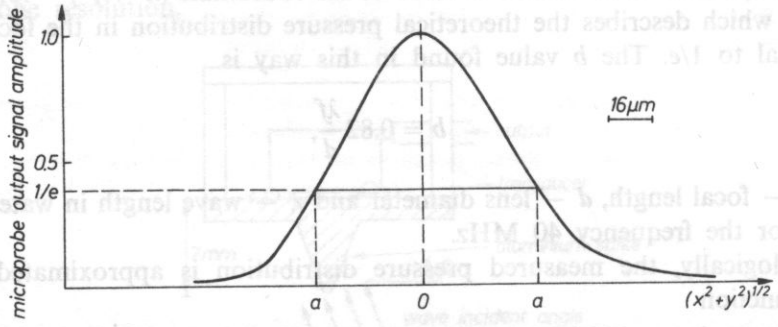


FIG. 2. Distribution of the pressure amplitude in the focal plane of an acoustic microscope lens, measured with the microprobe. Lens dimensions: radius of curvature 3.1 mm, half-angle $\alpha/2 = 50°$. The lens radiates into water with the frequency 40 MHz

The value of a found from Fig. 2 has been substituted to Eq. (5), what resulted in finding the resolving power of the microprobe to be equal about 25 μm.

In the measurements at the surfaces of solids the liquid is applied to wet the surface (e.g. oil). The measurements are carried out at the moment when the probe touches the surface or when it is just above the surface (about 5 μm). In both cases due to the application of a liquid coupling medium, the resolution of the microprobe should be comparable with the one found for the measurements in liquid.

4. The directional response pattern of the microprobe

The measurements of the relation of the signal received from the microprobe to the angle of incidence of the wave on the spike have been carried out. Measurements were always made in the same point of the surface relatively to the transducer which emitted the wave. The results are shown in Fig. 3, together with the theoretically

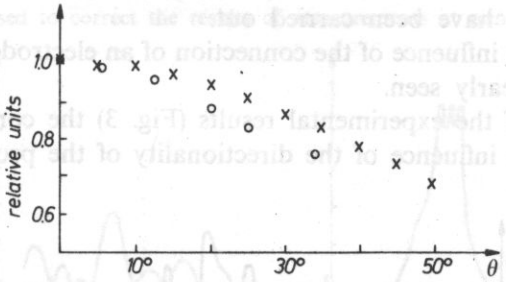


FIG. 3. Directional response pattern of the microprobe: measured points (o) and calculated from Eq. (6) (×)

calculated directional response pattern of the microprobe. Dimensions of the active surface of the probe are about 0.12 λ (λ – wave length in the material of the probe). Therefore, in the investigations of the relation of the signal from the probe to the incidence angle of the measured wave, the results obtained in the paper [5] can be applied. These results concerned the directionality of the longitudinal wave source, with dimensions not greater than the wave length radiating into an elastic semispace. They enable to describe the directional response pattern of the microprobe by the following formula

$$b(\theta) = \frac{v^2 \cos\theta (1 - 2v^2 \sin^2\theta)}{(1 - 2v^2 \sin^2\theta) + 4v^2 \sin^2\theta \cos\theta (1 - v^2 \sin^2\theta)^{1/2}}, \tag{6}$$

where $v^2 = (c_T/c_L)^2$, and c_T, c_L are the velocities of the transverse and longitudinal waves, respectively.

The calculated directional characteristic is presented in Fig. 3. The similar, cosinusoidal relation of both curves to the angle supports the statement that the active spike of the probe can be considered as a nearly point-type source which radiates into an elastic semispace.

For growing angles θ the experimental curve departs more and more from the directional response pattern of a source. It seems that this is caused by the fact that the values resulting from the directional characteristic for the given angle θ are greater than the pressure in the transducer plane, averaged by the transducer. The spike end radiates into the cone-shaped region rather than into a semispace, what also influences the above mentioned discrepancies.

5. Measuring capability of the microprobe

When defining the resolving power of the probe its measuring performance in fluids was also presented. The below described examples prove the possibility of pressure amplitude measurements at flat and spherical surfaces of solids.

The measurements at the surface of a glass cylinder with an ultrasound transducer glued on have been carried out.

In the Fig. 4 the influence of the connection of an electrode on the operation of the transducer is clearly seen.

With the use of the experimental results (Fig. 3) the correction factor which compensates for the influence of the directionality of the probe (Fig. 5) has been

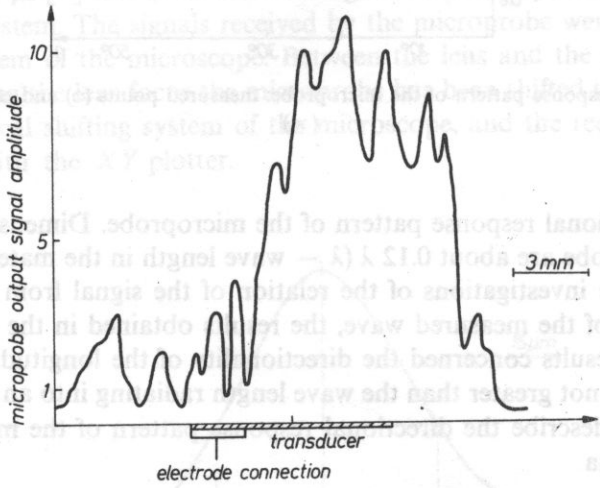


FIG. 4. Pressure amplitude distribution measured with the microprobe at a flat surface of a glass plate of thickness 20 mm. On the other side of the plate a piezoelectric transducer have been glued on. In the figure the position of the transducer and the point of connection of the cable to the transducer is marked. In the amplitude distribution the influence of the connection can be seen. The measurement was carried out at frequency 36 MHz

found. In the correction factor the influence of transverse waves emerging in the spike has not been taken into account because generation of such waves hadn't been noticed in the measurements. The surface waves arising at the spike surface have been damped by applying a wax layer on the spike. The measurement of pressure amplitude distribution at the surface of spherical acoustic lenses used in acoustic microscopy have also been carried out. It has been found that optimal signals were obtained when the lens cap was filled with machine oil. Two lenses with different cap position in the transducer field have been tested. The first, with diameter of 4.5 mm,

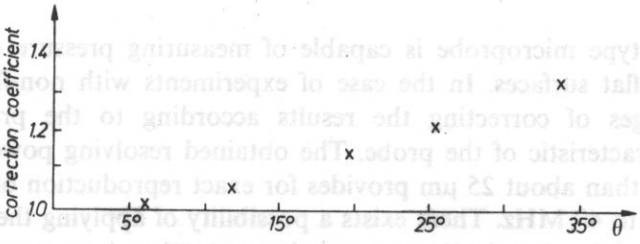


FIG. 5. Plot of the correction coefficient of the microprobe found on the basis of the experimental curve from Fig. 3, used to correct the results of measurements at non-planar surfaces

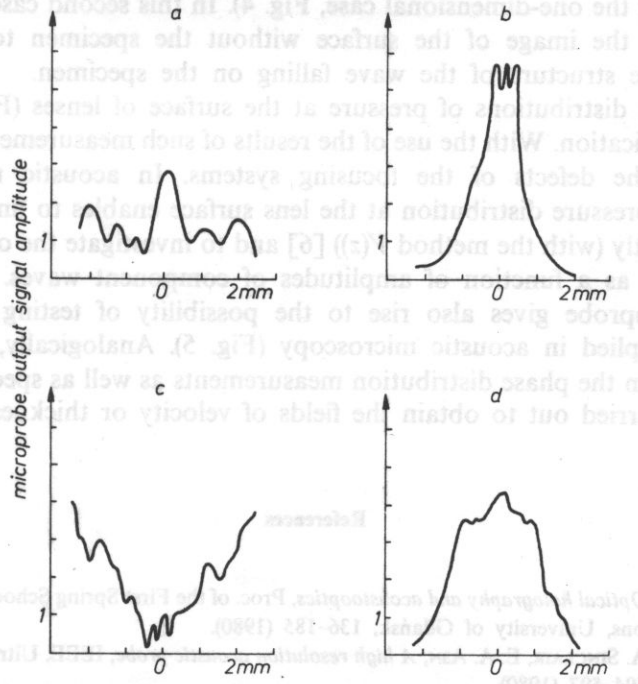


FIG. 6. Distributions of pressure amplitudes at the surface of an acoustic microscope lens of curvature radius 3.1 mm, measured for diverse lens positions in the transducer field: a) z — distance lens–transducer $= 0.42\ d^2/\lambda$, where d — transducer radius, λ — wave length in the medium between the transducer and the lens; b) $z = 1.21\ d^2/\lambda$; c) $z = 0.47\ d^2/\lambda$; d) $z = 1.38\ d^2/\lambda$. Correction according to the plot from Fig. 5 has been applied

was placed in the region between the near and far field ($z \simeq d^2/\lambda$); the second with diameter of 6 mm was in the middle of the near field ($z = 1/2\, d^2/\lambda$). By using the capabilities of the electronical system of the microscope to change the operating frequency within the range of 30–40 MHz it was possible to change the pressure distribution at the lens surface [4]. The results obtained with the correction factor taken into account are shown in Fig. 6.

6. Conclusions

The spike-type microprobe is capable of measuring pressure distributions in liquids and at flat surfaces. In the case of experiments with non-flat surfaces the necessity emerges of correcting the results according to the previously found directional characteristic of the probe. The obtained resolving power of the probe which is better than about 25 μm provides for exact reproduction of wave pressure distribution up to 40 MHz. There exists a possibility of applying the microprobe in place of the receiving lens in the transmission acoustic microscope in forming the microscopic images of surfaces or planar specimens irradiated by a plane wave (analogically to the one-dimensional case, Fig. 4). In this second case it is necessary first to obtain the image of the surface without the specimen to eliminate the influence of the structure of the wave falling on the specimen.

Measuring distributions of pressure at the surface of lenses (Fig. 6) is a very important application. With the use of the results of such measurements it is possible to eliminate the defects of the focusing systems. In acoustic microscopy the knowledge of pressure distribution at the lens surface enables to find the reflection coefficient exactly (with the method $V(z)$) [6] and to investigate the oscillation of the $V(z)$ curve [4] as a function of amplitudes of component waves.

The microprobe gives also rise to the possibility of testing the quality of transducers applied in acoustic microscopy (Fig. 5). Analogically, using a phase detection system the phase distribution measurements as well as specimen visualization can be carried out to obtain the fields of velocity or thickness distributions.

References

[1] A. Śliwiński, *Optical holography and acoustooptics*, Proc. of the First Spring School on Acoustooptics and Applications, University of Gdańsk, 136–185 (1980).

[2] W. Dürr, D. A. Sinclair, E. A. Asn, *A high resolution acoustic probe*, IEEE, Ultrasonics Symposium Proceedings, 594–597 (1980).

[3] J. W. Goodman, *Introduction to Fourier optics*, Mc Graw Hill, New York 1968.

[4] J. Litniewski, *An acoustic microscope in measurements of mechanical properties of surface layers*, Archives of Acoustics, **11**, 3, 299–312 (1986).

[5] L. FILIPCZYŃSKI, *Measurements of longitudinal and transverse waves radiated by a compressional source into elastic semispace*, Proc. of Vibration Problems, **5**, 2, 89–93 (1964).

[6] K. LIANG, G. S. KINO, B. T. KHURI-YAKUB, *Material characterization by the inversion of V(z)*, IEEE Trans. on Sonics and Ultrasonics, vol. SU-32, **2**, 213–224 (1985).

Received on August 7, 1987.

ARCHIVES OF ACOUSTICS
13, 1–2, 55–65 (1988)

TIMBRE DIFFERENCES OF AN INDIVIDUAL VOICE IN SOLO AND IN CHORAL SINGING

TOMASZ ŁĘTOWSKI, LIDIA ZIMAK, HALINA CIOŁKOSZ-ŁUPINOWA

Sound Recording Department Chopin Academy of Music (00–368 Warszawa, Okólnik 2)

The object of the study was to analyze differences between solo and choral production of the same singer. Six female and six male chorus singers sang the same musical material solo, in vocal groups, and as a choir. Choral productions of the individual singers were obtained by recording their solo productions made under playback of the choir sound. Spectral analysis and listening tests were used to analyze differences between recorded samples. The results showed highly significant differences between both types of vocal production. The difference increased with the increasing degree of vocal training of the performer.

Celem niniejszej pracy było porównanie widma głosu śpiewaka solisty z widmem głosu tego samego śpiewaka śpiewającego w chórze. W ramach pracy dokonano szeregu nagrań produkcji solowej grupy śpiewaków, nagrań tej samej grupy osób występujących jako chór oraz nagrań wyizolowanych głosów pojedynczych chórzystów. Wyizolowania głosów pojedynczych chórzystów dokonano metodą opartą o technikę playbacku. W rezultacie przeprowadzonych porównań słuchowych oraz analiz widmowych próbek głosów stwierdzono istnienie znaczących różnic pomiędzy barwą głosu tego samego wykonawcy podczas śpiewu solowego i śpiewu w chórze. Wielkość tych różnic uzależniona jest od stopnia wyszkolenia wokalnego śpiewaka.

1. Introduction

A choir is a group of people simultaneously singing the same musical piece. Choral singing involves blending the voices of several singers together and, therefore, presents different requirements for performers than solo singing. The timbral quality of a choral production depends on the spectral characteristcs of the individual voices and their mutual compatibility. The extremely rich timbre of choral singing is the result of a phenomenon known as the chorus effect. This effect consists of the constant differences in spectral timbre, loudness, and pitch of the same note as performed at the same time by a number of performers. The rich, full timbre of large choirs is quite unique and is not matched by any other type of music ensembles. The chorus effect is not merely a mixture of various voices. Choral singing needs to be

a smooth, well-blended production of a group of co-performing singers rather than a cluster of individual artistic performances. Therefore, it is important to know the type of spectral characteristics of solo voices which contribute most to high quality choir production. The present study approaches this question by investigating the timbre and spectral characteristics of both the choir as a whole and individual performers singing solo and in the choir (choral voice).

Although the term "sound quality" with respect to choral production is widely used, there has almost been no attempt to study this phenomenon from the acoustic point of view or to determine the influence of an elementary choir unit – a single voice – on the timbre of the whole ensemble. Previous studies of the human voice were concerned primarily with various aspects of speech production and perception and with the quality of a solo performance by highly trained professional singers. There have been only a few studies addressing the problem of choral singing and no attempt at all to relate the quality of an individual voice to the quality of the choir performance as a whole.

There is a general agreement among voice teachers that the quality of voice production depends on a large number of acoustical and artistic factors, e.g., the character of the produced vibrato, the relation between loudness and pitch over the entire voice scale, and the spectral character of the vocal sounds. In the case of vocally trained singers the formant regions of sung vowels are distinctly different from those of the spoken vowels [5]. Moreover, the spectra of sung vowels indicate the existence of a substantial concentration of energy in the region 2500–3000 Hz. This concentration is especially strong in the case of male operatic voices.

In 1934 BARTHOLOMEW [1] was the first to show the dependence of high-quality vocal production on the presence of additional acoustic energy in the region of 2800 Hz in the spectrum of sung vowels. This concentration of energy is commonly called the "singer's formant" [4] and is relatively independent of a sung vowel and its pitch. As a rule, it falls between the third and the fourth formant of a spoken vowel, making the singer's voice full and glamorous. SUNDBERG [7, 8] related the existence of the singer's formant to a lower position of the larynx in singing which, together with the widened pharynx, creates an additional resonating space with a resonant frequency of 2500–3000 Hz. The lowered larynx position also favors a darker voice timbre appreciated by most singers. In the case of untrained voices the singer's formant does not normally occur. On the contrary, there appears a tendency to elevate the larynx position in direct proportion to the pitch of the sound being sung [5] which makes higher notes sounding tinnier.

Although the presence of a strong singer's formant is not a necessary condition for good voice quality, e.g., in the case of soprano voice [1, 2, 3], it enables a voice to stand out from the background of the orchestra or that of the accompanying choir [8]. The prominence of one voice against the sound of the whole choir is not, however, a desirable feature in choral singing. On the contrary, in choral singing it is necessary to blend a single voice with the sound of the whole choir.

It has been observed that the same people sing differently while performing solo

or in choir. So far, however, only two studies have addressed this issue. GOODWIN [4] studied soprano singers performing in various ensembles and found that soprano voices are notably weaker when performing in choir than solo. Notably, the energies of the second and the third formant were reduced while no substantial shifts in formant frequencies were observed. ROSSING *et al.* [6] studied bass\baritone voices and reported that solo and choir singing showed differences in both articulation and phonation. In particular, they found that solo in comparison with choir singing yielded a more prominent "singer's formant" and substantially reduced energy of the fundamental.

The purpose of the present study was to compare spectral and timbral differences between solo and choral production of the same person and to relate the overall timbre of the whole choir to that of an individual performer. The success in carrying out this task depends primarily on the method which is applied to isolate a single choral voice from the group production. It should be kept in mind that a necessary condition of a good choral production is that all voices fuse with one another. To satisfy this requirement each of the performers must hear, in natural proportions, the production of the other members of the choir. Therefore, in order to record only a single choral voice the studied voice must be isolated in such a way that the singer can still hear the sound of the choir just as he would normally hear it in the course of collective music making. To satisfy these conditions a method of recording a single, isolated choral voice has been developed by recording the voice of a single chorus singer singing under the playback of the choir performance.

2. Verification of the method used for recording of the isolated choral voices

To verify the desired character of the isolated choral voices a pilot study was performed which involved comparison of a natural choir sound with that made by the mixing of two separate recordings. A group of five sopranos from the chamber vocal ensemble "Warsaw Madrigal Singers" took part in the experiment. The music material was the first eight bars from the soprano part of Jan Maklakiewicz's song "A Lullaby". Instead of the lyrics, the singers sang the vowel (a). The voices of individual singers and various group of singers were recorded by the single-microphone technique using an AKG C-12 microphone set on the axis of the sound source at the height of the performers' heads. The block diagram of the recording equipment is shown in Fig. 1. Rather than placing microphone among the choir ensemble the above recording technique was applied to assure a good blend of all chorus voices in the recordings. The recording session included five steps:

(a) recording of the whole soprano group (recording I),
(b) recordings of the solo productions,
(c) recordings of the whole soprano group minus one person,
(d) recording of the whole soprano group (recording II),

(e) recordings of the single performer singing under the respective recording made in step (c).

Recordings made in steps (b), (c), and (e) were repeated for each of the five sopranos as an isolated voice. The sound intensity level of the original production was kept approximately constant at 75–80 dB. The recordings were made in Studio S-1 of the Chopin Academy of Music in Warsaw. To warm-up singers and to assure reliable singing each musical example was sung with repetition with only the repetition being used for the purpose of this study. The playback was provided with Beyer DT-48 earphones.

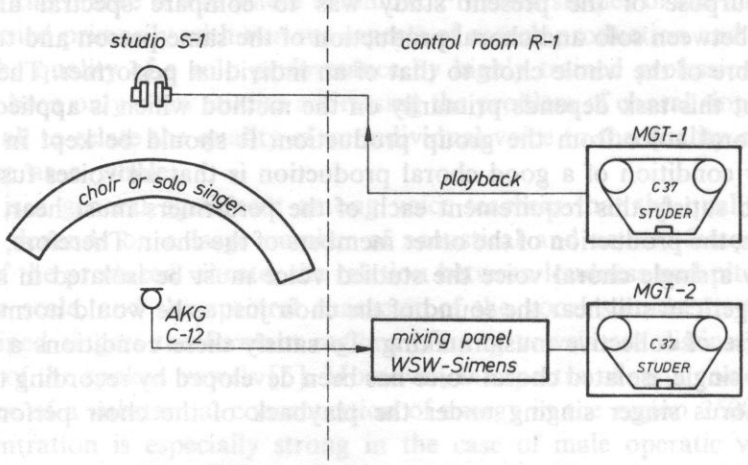


FIG. 1. Block diagram of the recording equipment

In the course of making recordings in steps (b), (c), and (e) one could easily notice the difference between solo and choral singing of the same singer and the difference between the sound produced by the various four-voice groups recorded in step (c). After all the recordings had been made, the respective recordings (c) and (e) were mixed to simulate the sound of the whole soprano group as recorded in steps (a) and (d). The new, mixed recordings were presented to the performers who found them indistinguishable from the original recordings of the whole soprano group. That meant that persons singing under the playback condition always managed to sing as if in the choir. This was confirmed by statements made by all of the performers that sound of playback heard through the earphones gave them the awareness of a "close" contact with the rest of the group. This made it possible to match individual voice production to that of the other part of the ensemble.

The formal listening test employed three alternative forced choice (3-AFC) procedure. Each test trial included recordings I and II of the whole ensemble, and one of the recordings made during the mixing session. The test consisted of 15 triads incorporating all permutations of the tested stimuli. The order of trials in the test were randomized. The listening group consisted of ten experienced listeners and included five sound engineers and five conductor-choirmasters. Each listener was

asked to indicate in each of the triads the stimulus which sounded most different from two others. The sounds were reproduced by Tannoy-Lockwood loudspeakers with sound intensity level equal to that of the original sound. The listeners listened individually to a single presentation of the test during which each listener's responses were collected.

The individual listener's responses appeared to be randomly distributed with the number of responses indicating a mixed recording not exceeding 40%. This result indicates that the recordings achieved by the mixing technique were not different to the listeners from those of the whole ensemble. Therefore, it may be concluded that for the individual choral voice the recordings made with the described method were perceptually equivalent to those of the same person singing directly in the ensemble.

3. Procedure and results

The whole, 17-person choir (ten women and seven men) "Warsaw Madrigal Singers" participated in the main experiment. The choir consisted of five soprano, four alto, five tenor, and three basso voices. The recordings were made with singers positioned in two rows (with men standing in the second row on the platform). The recording technique was the same as described in Section 2. The music material was an excerpt from the first movement of Johann Sebastian Bach's chorale "St. Matthew Passion" – "Befiehl du deine Wege". This four-bar fragment was sung with repetition. Similarly, as in the pilot experiment, the first four bars ensured that the singers had warmed-up properly while the recording consisted only of the repetition itself. The recording session included the following steps:

(a) recording of the whole choir,
(b) recordings of the individual voice groups (sopranos, altos, tenors, and bassos),
(c) recordings of the solo productions (for technical reasons they were made only for 12 performers: three of each voice group),
(d) as (c) but under the playback of the recording made by the whole ensemble,
(e) as (c) but under the playback of the recording made by the respective voice group.

For each of the twelve selected performers three sample recordings were obtained: solo production, voice isolated from the whole choir, and voice isolated from the voice group. Those recordings provided material for later comparative study in which comparisons between the voice of the same person singing solo and under various forms of playback were carried out. The respective recordings have been labeled from 1 to 12 according to the following scheme:

sopranos – from 1 to 3,
altos – from 4 to 6,
tenors – from 7 to 9,
bassos – from 10 to 12.

Each recording has also been labeled by the letter "*a*", "*b*", or "*c*" indicating solo

singing, singing under the playback of the whole choir, and singing under the playback of the given voice group, respectively. The recording of the whole choir was denoted by the symbol "*ch*" while the recordings of the main vocal groups were denoted as indexed "*g*".

To make comparisons between the physical properties of the voice sample „*a*", "*b*", and "*c*", the average spectrum for each of the samples was calculated. Each of the recorded samples was analyzed with a set of one-third octave filters (B-K 1612). The analysis was limited for practical reasons to 100–6300 Hz. For each filter the time-amplitude envelope was recorded on the B-K 2305 graphic level recorder. Subsequently, the plots of spectrum levels for all the center frequencies were calculated at five arbitrarily chosen time prints t_i approximately equally distributed along the time axis. The points t_i were chosen in terms of the particular sound produced at these points. The spectra for all five time points were subsequently averaged with a Wang 600-12 computer. As a result 41 different voice spectra were obtained (36 spectra for single voices and 5 spectra for group performances). Examples of some characteristic spectra are shown in Fig. 2.

The sound spectra such as those presented in Fig. 2 were used later to calculate

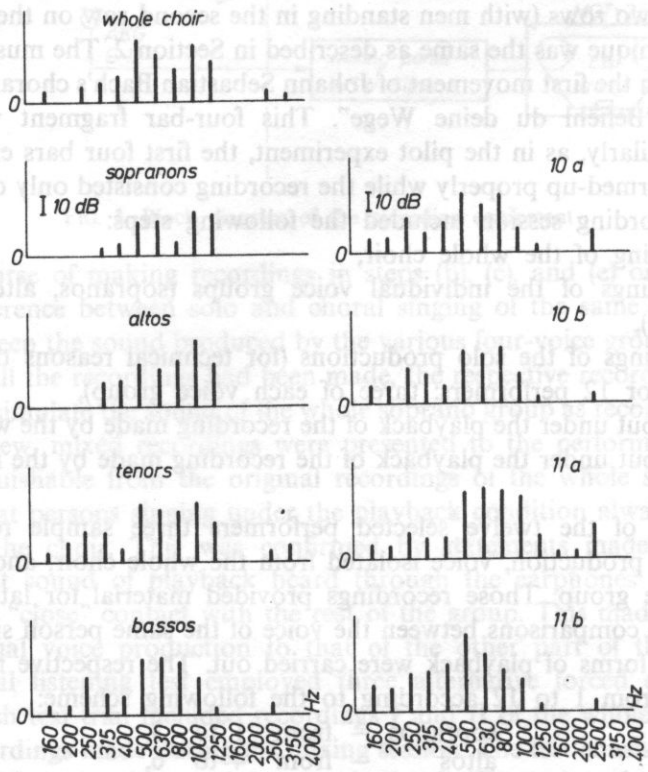


FIG. 2. One-third octave average spectra of vocal recordings made for the choir and four separate voice groups (sopranos, altos, tenors, and bassos). On the left side of the diagram there are four examples of the individual spectra obtained for two bass singers (voices 10 and 11) performing (a) solo and (b) under the playback of choir sound

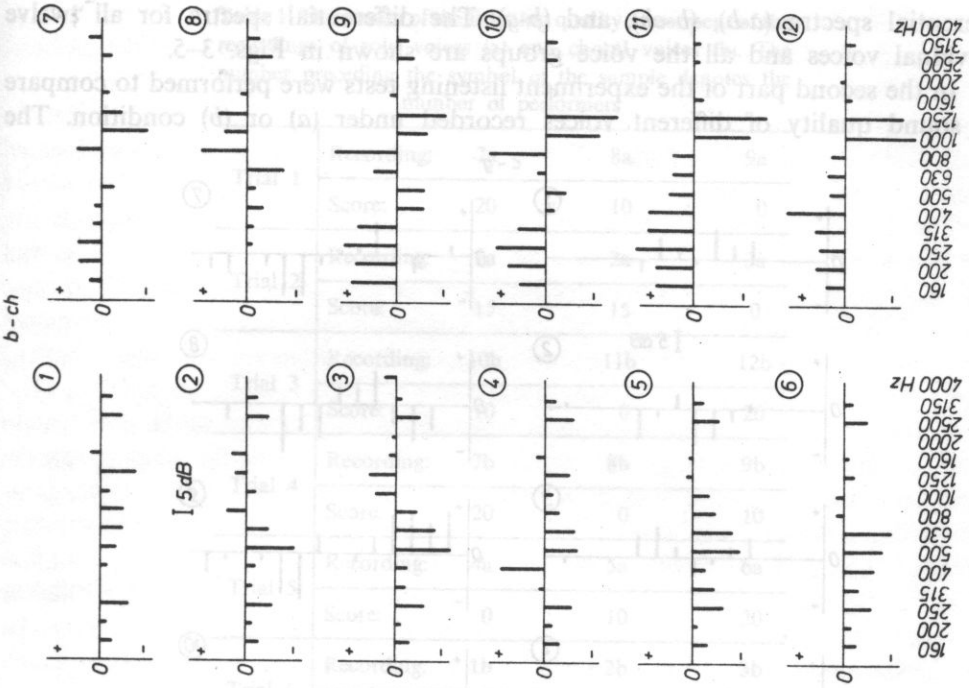


FIG. 4. Differential spectra "*b–ch*"

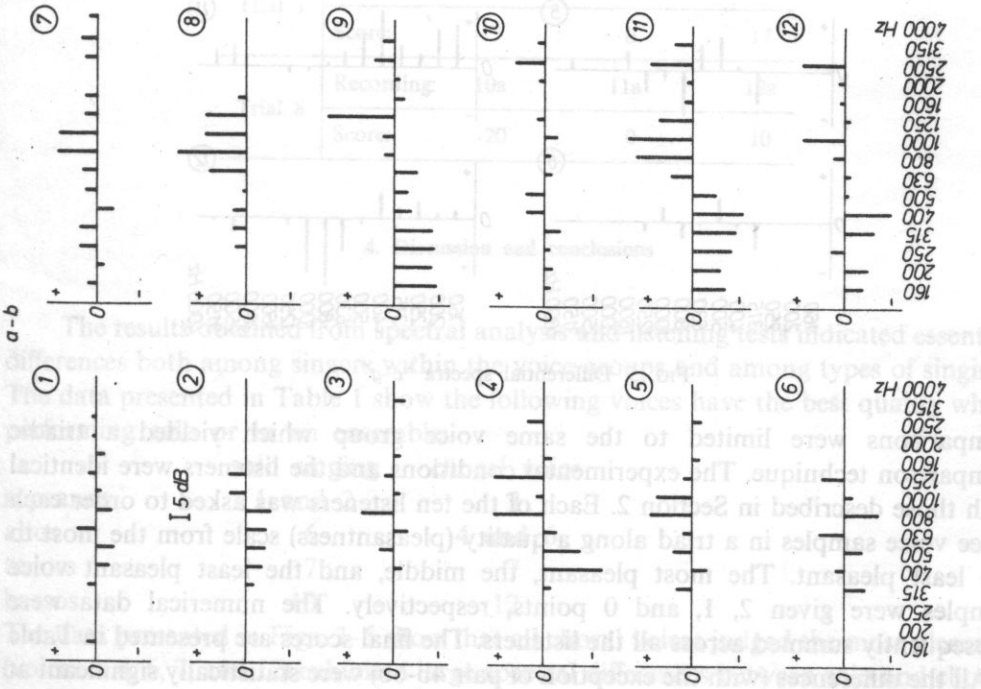


FIG. 3. Differential spectra "*a–b*"

differential spectra (*a-b*), (*b-ch*), and (*b-g*). The differential spectra for all twelve individual voices and all the voice groups are shown in Figs. 3–5.

In the second part of the experiment listening tests were performed to compare the sound quality of different voices recorded under (*a*) or (*b*) condition. The

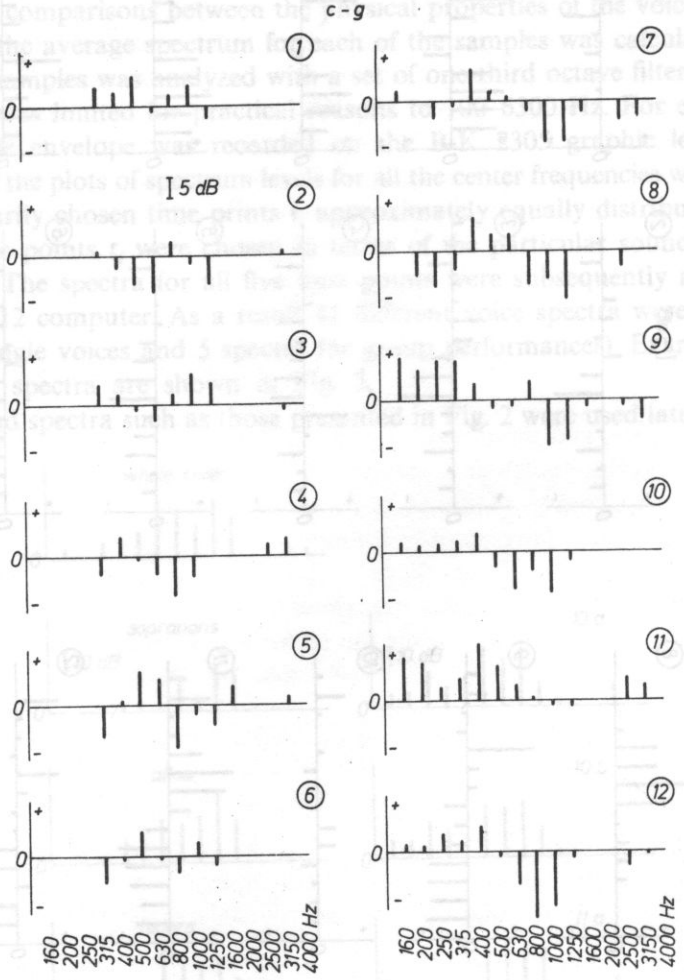


FIG. 5. Differential spectra "*c–g*"

comparisons were limited to the same voice group which yielded a triadic comparison technique. The experimental conditions and the listeners were identical with those described in Section 2. Each of the ten listeners was asked to order each three voice samples in a triad along a quality (pleasantness) scale from the most to the least pleasant. The most pleasant, the middle, and the least pleasant voice samples were given 2, 1, and 0 points, respectively. The numerical data were subsequently summed across all the listeners. The final scores are presented in Table 1. All the differences (with the exception of pair 4b-6b) were statistically significant at the level 0.05.

Table 1. The results of the listening quality assessment of the recordings of solo voices (a) and choral voices (b). The number preceding the symbol of the sample denotes the number of performers

Trial 1	Recording:	7a		8a		9a
	Score:	20	:	10	:	0
Trial 2	Recording:	1a		2a		3a
	Score:	15	:	15	:	0
Trial 3	Recording:	10b		11b		12b
	Score:	10	:	0	:	20
Trial 4	Recording:	7b		8b		9b
	Score:	20	:	0	:	10
Trial 5	Recording:	4a		5a		6a
	Score:	0	:	10	:	20
Trial 6	Recording:	1b		2b		3b
	Score:	0	:	20	:	10
Trial 7	Recording:	4b		5b		6b
	Score:	13	:	0	:	17
Trial 8	Recording:	10a		11a		12a
	Score:	20	:	0	:	10

4. Discussion and conclusions

The results obtained from spectral analysis and listening tests indicated essential differences both among singers within the voice groups and among types of singing. The data presented in Table 1 show the following voices have the best quality while performing solo or in an ensemble:

	solo singing	choral voice
sopranos	1 and 2	2
altos	6	4 and 6
tenors	7	7
bassos	10	12

The data presented in Figs. 3–5 show that all choral voices judged the most pleasant (voices 2, 4, 6, 7, and 12) exhibited large spectral differences between solo and choral modes of singing.

The fact that all the female singers participating in the experiment had no previous vocal training whereas four male singers (voices 7, 8, 10 and 11) had been vocally trained affected to a large extent the specific spectra of individual voices and the voice groups (Fig. 2.) Within the 50 dB dynamic range, the spectra of the female voices seldom extended beyond 1600 Hz, whereas the male voices showed components up to 4000 Hz.

The differential spectra show that among soprano voices the voice 2 was the most elastic, i.e., the most different in recordings "*a*", "*b*", and "*c*". For this voice the most favorable situation was singing along with the whole ensemble. This was expressed by considerable enrichment of the respective spectrum "*b*" in comparison to the other spectra of the same voice. In the case of two other soprano voices, voices 1 and 3, no such tendency was observed.

In the alto group, the most pleasant voice was voice 4. This was the only female voice to show the existence of the singer's formant in solo singing mode. A characteristic nasal timbre was observed in that voice during solo production. Such timbre is observed in cases when a spectrum which is basically limited to low frequencies includes one or two components at much higher frequencies. In turn, the lack of hoarseness and nasality in choral performance reflected the great flexibility of the voice and good matching to singing in the mixed ensemble. It may be hypothesized that it was the presence of low male voices in the ensemble which induced this voice to make the timbre darker (softer, more mellow).

Among the tenor group, voices 7 and 9 did not show any large differences between the spectra of samples "*a*", "*b*", and "*c*". Vocally trained voice 8 exhibited a rich spectrum in solo singing but it appeared to be much thinner as a choral voice. Probably, the singer tried to adapt as much as he could to the sound of the choir, causing a weakening of the higher components of the sound.

In the basso group voice 11 appeared to be the most different among all analyzed voices. Perhaps of importance was the fact that this person had been vocally trained and also had been with the choir for a relatively short time. Therefore, the singer could not blend the timbre of his voice to that of the choir and tried to dominate his vocal group. Voices 10 and 12 sounded especially thin in the case of the recording made under the playback of the voice group. Most probably, this resulted from their attempt to produce a uniform sound with voice 11 which was unflexible and not susceptible to changes in the conditions of singing.

The results of this study confirm the differences in the behavior of singers with trained and untrained voices. While singing in the ensemble, untrained voices were characterized by strengthening the extreme components of the voice spectrum compared with the spectrum of solo singing samples. In choral production these voices sounded brighter, more stable, and more colorful. On the other hand, the vocally trained singers sounded usually darker and poorer while performing in choir than in solo singing. It may be hypothesized that choral performance is a mobilizing factor for untrained voices. The trained singers, however, are forced to adapt their natural, rich sound to the timbre of the ensemble. This unfavorably affects the timbre and quality of their voices.

In the case of solo singing, persons with vocally untrained voices usually sang faster and more nervously than those with trained voices. These singers were unable to remember the tempo of the previously recorded sample of the whole choir, or to sustain longer notes to their end. Persons with trained voices had no trouble in maintaining the same tempo in solo production as it had been in the recording of the whole ensemble.

In summary, (a) the possibility exists for recording isolated choral voice by a method of singing under playback of the other part of the ensemble; (b) solo (group) singing made under playback of other voices is quite natural for choir singers which makes possible the application of such a method for recording choir performances by multitrack techniques; (c) vocally untrained singers tend to operate with a richer (brighter) voice while singing in the ensemble than in solo mode: (d) sharp (bright) voices, not able to maintain the timbre uniform with the rest of the ensemble, exert a considerable effect on the behaviour of the other voices making them sound much thinner than they might; and (e) the voices of vocally trained persons sound richer (more powerful) in solo singing than in choral performances, where the singer's formant is frequently totally missing, but nevertheless the presence of such voices on the ensemble is often an essential factor mobilizing the other performers.

References

[1] W. T. BARTHOLOMEW, *A physical description of "good voice quality"*, J. Acoust. Soc. Am., **6**, 25–33 (1934).

[2] G. BLOOTHOOFT, *Spectrum and timbre of sung vowels*, Doctoral Dissertation, University of Amsterdam, 1985.

[3] G. BLOOTHOOFT, R. PLOMP, *The sound level of the singer's formant in professional singing*, J. Acoust. Soc. Am., **79**, 2028–2033 (1986).

[4] A. GOODWIN, *Acoustic study of individual voices in choral blend*, J. of Research in Singing, **3**, 25–36. (1980).

[5] T. ROSSING, *The science of sound*, Addison-Wesley, Reading, MA, 1982.

[6] T. ROSSING, J. SUNDBERG, S. TERNSTRÖM, *Acoustical comparison of voice use in solo and choir singing*, J. Acoust. Soc. Am., **79**, 1975–1981 (1986).

[7] J. SUNDBERG; *Articulatory interpretation of the "singing formant"*, J. Acoustic. Soc. Am., **55**, 838–844 (1974).

[8] J. SUNDBERG, *The acoustics of the singing voice*, Scientific American, **236**, 3 (1977).

Received on January 12, 1987.

ARCHIVES OF ACOUSTICS
13, 1–2, 67–88 (1988)

REAL TIME SIGNAL PROCESSING TECHNIQUES IN A DUAL BEAM SONAR SYSTEM FOR FISH STOCK ASSESSMENT

R. S. MITCHELL, A. STEPNOWSKI*, J. DROBNIK, D. R. GREEN

Seastar Instruments Ltd. Argo Bldg. Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada B2Y 3Y9

*Technical University of Gdańsk, Institute of Telecommunications, 80–952 Gdańsk, Poland

Advances in VLSI technology including A/D converters, large high speed memories and digital signal processing devices, combined with a dramatic improvement in high performance single board computer SBCs, have radically altered design possibilities for sonar systems [1, 2]. It is now possible to centralize the system functions in one controller and transfer much of the signal processing from hardware to software. Furthermore, digital signal processing techniques may be employed to improve the system's performance and reliability. In this paper, we describe a new stock assessment dual beam sonar designed to take advantage of modern computer architectures and digital signal processing techniques. We also discuss some original processing enhancements to the basic dual beam concept both in hardware and software which have been built into the sonar system. The system is flexible so that the same hardware is employed for various modes. The transfer of functions from hardware to software combined with a hihgly efficient transmitter has yielded a low cost compact system.

Rozwój technologii VLSI bardzo dużej skali integracji, a w szczególności przetworników A/D, bardzo dużych i szybkich pamięci oraz układów do przetwarzania sygnałów w połączeniu z dramatycznym rozwojem mikrokomputerów, w radykalny sposób zmienił możliwości projektowania systemów sonarów [1, 2]. Obecnie stało się możliwym zgrupowanie większości funkcji w systemie jednego kontrolera oraz szerszego programowego przetwarzania sygnałów w miejsce przetwarzania sprzętowego. Należy dodać, że zastosowanie cyfrowego przetwarzania sygnałów poprawia jakość i niezawodność systemu. W artykule podana została ocena dwuwiązkowego sonaru zaprojektowanego pod kątem korzyści jakie oferuje nowoczesna architektura komputerowa i technika cyfrowego przetwarzania sygnałów. Przedyskutowane zostaną pewne nowe oryginalne modyfikacje, tak programowe jak i sprzętowe, przetwarzania sygnałów w porównaniu z klasycznym podejściem do dwuwiązkowego systemu sonaru. Opracowany system jest elastyczny tak, że opracowany hardware może mieć wiele zastosowań. Przeniesienie funkcji operacyjnych z rozwiązań sprzętowych do programowych (software) w połączeniu z bardzo skutecznym nadajnikiem pozwoliło opracować system tani i zwarty (compact).

1. Introduction

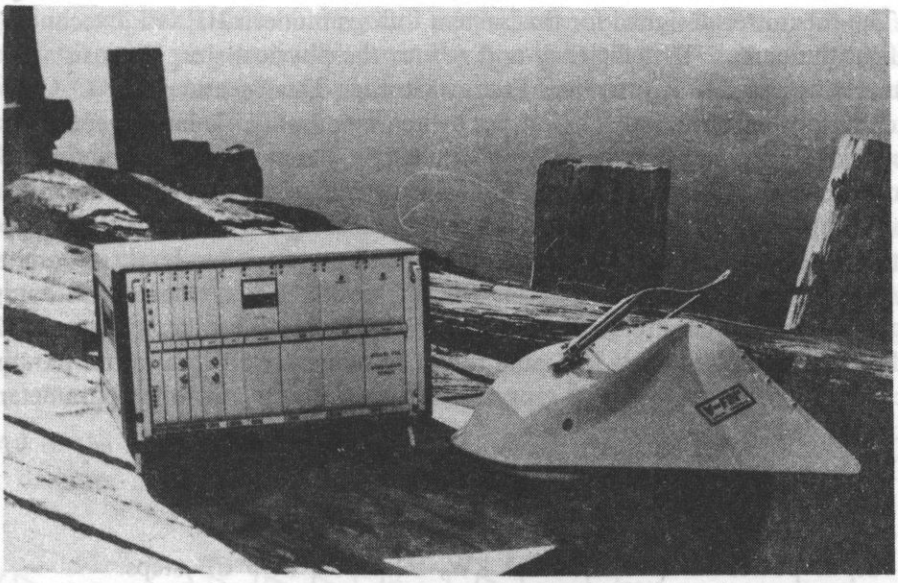
The utilization of dual beam configuration in hydroacoustic systems has proven to be an accurate and reliable means for target strength estimation from received echos. Conventional sonar systems for quantitative fish stock assessment either have limited real time capabilities or require dedicated extensive separate hardware to perform echo counting, target strength estimation and/or echo integration [3, 4].

Over the last few years, there have been major improvements in VLSI technology resulting in the availability of low cost digital signal processing devices and high performance microcomputers. The benefits to the system design engineer have been profound. Large reductions in hardware real estate and enhancements in system performance and flexibility have resulted. The availability of high performance 16 bit microprocessors has changed the design philosophy of microprocessor systems from that of a few dedicated task oriented machines arranged in a sequential or pipeline fashion to that of a high speed parallel bus with one or more high performance machines. Complementing the improvements in machine architecture have been advances in software, namely real time executives enabling a single processor to perform many different tasks simultaneously. Furthermore, international standards for the multiprocessor bus systems have resulted in the availability of numerous inexpensive single board computers.

This paper describes the application of these advances in hardware and software to the design of dual-beam sonar system for fish stock assessment. A high speed A/D converter in conjunction with digital signal processing devices results in relaxes the specifications of the analog hardware, allows the use of a digital Finite Impulse Response (FIR) filter, and transfers the Time Variable Gain (TVG) function provision requirements to a digital processor which results in improvement of the receiver performance and in the overall system flexibility. Emphasis on software design has yielded a flexible system that is compact, low cost and user-friendly, and which combines all basic functions required for acoustic surveying of fish populations, namely: echo-sounding (fish-echo detection), echo-integration, echo-counting and target strength estimation.

2. System architecture

A photograph and block diagram of the dual beam sonar system is shown in Figure 1 and 2 respectively. The central single board computer (SBC), based on the 68010 microprocessor is the heart of the system. Its responsibilities include transmitter and receiver control as well as data collection and real time analysis. A host personal computer, IBM PC/XT or compatible is also connected to the 16 bit Versatile Modular European (VME) based system. The PC provides a user-friendly, menu driven interface between the operator and the central computer. The PC may



Sonar unit Towed Body with transducer

FIG. 1. Dual beam sonar system (Host Computer Workstation not shown)

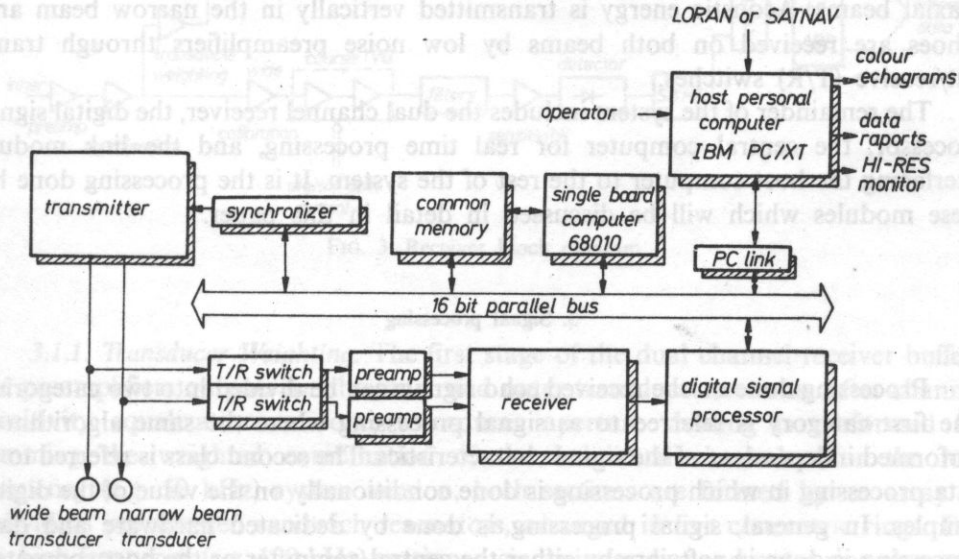


FIG. 2. Dual beam sonar system block diagram

take control of the system for testing and diagnostics. Various peripherals may be connected to the PC for storage of raw or processed data, graphics presentation, user interface and navigation equipment interfacing. A typical configuration consists of a 20 Mbyte hard disk, a high resolution colour monitor, a colour chart recorder and a Loran C serial interface.

The trasmitter designed for this system utilizes modern HEXFET technology to significantly increase the efficiency and reduce the physical size. It consists of three sub-modules: AC/DC Converter, Energy Storage Device and DC/AC Converter (switching power amplifier). The output switching amplifier is implemented in a full wave bridge configuration, and is controlled by the synchronizer module. When commanded by the central computer, the synchronizer will generate necessary control (appropriate transmit gates) to the output bridge network, so the high power burst will be routed to the transducer load. The output power level is regulated by the AC/DC converter the output of which is controlled by the central computer using an eight bit digital to analog converter (DAC). The central computer also loads the desired carrier frequency and ping rate. Consequently all transmitter parameters are computer programable. The ranges of these computer controlled parameters are as follows:

ping rate $f_p = 0.1$–10 Hz,
pulse length $\tau = 0.1$–10 ms,
carrier frequency $f = 30$–450 kHz,
output power $P = 0.075$–2.4 kW, selectable in 3 dB steps.

The transducer, which is physically mounted inside the towed body, consists of two sections arranged in concentric configuration, to form the narrow and wide coaxial beams. Acoustic energy is transmitted vertically in the narrow beam and echoes are received on both beams by low noise preamplifiers through transmit/receive (T/R) switches.

The remainder of the system includes the dual channel receiver, the digital signal processor, the central computer for real time processing, and the link module interfacing the host computer to the rest of the system. It is the processing done by these modules which will be discussed in detail in this paper.

3. Signal processing

Processing done on the received echo signals can be divided into two categories. The first category is referred to as signal processing where the same algorithm is performed independent of the signal characteristics. The second class is referred to as data processing in which processing is done conditionally on the value of the digital samples. In general, signal processing is done by dedicated hardware and data processing is done in software by either the central computer or the host computer.

The echoes received by the sonar transducer are subjected to two stages of signal processing. The first stage is the dual-channel receiver analog preprocessing (signal conditioning) which performs equalization of sensitivities in both channels, transducer aperture weighting, course TVG, filtering, demodulation, sampling and analog to digital conversion (ADC). In the second stage, the multiplexed digital receiver outputs are applied to a hardware digital signal processor (DSP) whicl performs the digital TVG correction, digital filtering in FIR filter, sampling rate

decimation, amplitude thresholding and preliminary bottom, detection. The output of the digital signal processor is buffered in a first-in first-out (FIFO) queue, which in turn is served by the central computer where real time data processing of the echoes is performed.

The central computer can operate in a number of modes to implement echo counting and integration algorithms with or without target strength estimation. The computer also has direct control of the receiver and signal processor to set up the required configuration for the each mode of operation. The output of the central computer is sent to the host PC through a common large memory also on the system bus.

3.1. Analog signal processing

All of the analog signal processing is done on the dual channel receiver, the block diagram of which is shown in Figure 3.

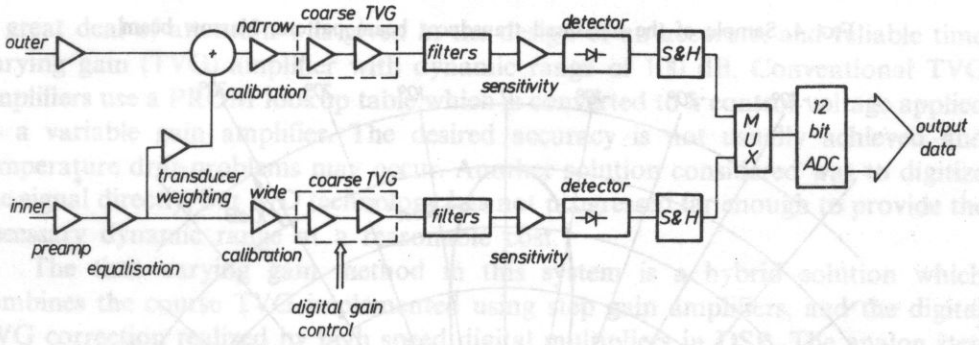


FIG. 3. Receiver block diagram

3.1.1. Transducer Weighting. The first stage of the dual channel receiver buffers the preamplifiers outputs for the inner and outer transducer sections. After channel sensitivity equalization is done, transducer aperture shading is performed by summing the weighted combination of the individual channels. For the low frequency (e.g. 50 kHz) system version, both sections are formed by an array of individual pre-stressed sandwich resonators, arranged in five concentric rings. For the high frequency (e.g. 120 kHz) version, a single inner disc and outer ring elements constitute the transducer's aperture. In the present system design only two-step shading was applied, to reduce the side lobe level in receive narrow beam, by appropriate sensitivity weighting in the variable gain buffer amplifiers of the receiver. However, optional weighting capabilities are provided for transmit beam pattern and also multi-step weightings may be applied for both transmit and receive beams.

The measured narrow beam and wide beam patterns for 50 kHz transducer are shown in Figure 4 and figure 5, respectively. The unweighted beam patterns have

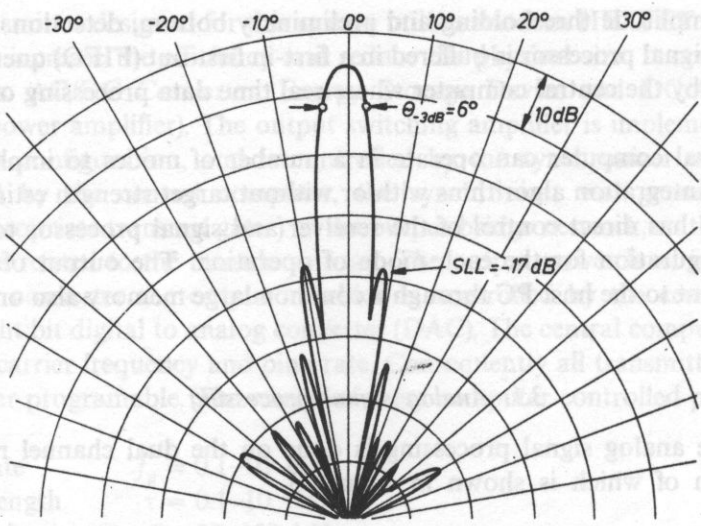


FIG. 4. Sample of the measured transducer beam pattern. Narrow beam

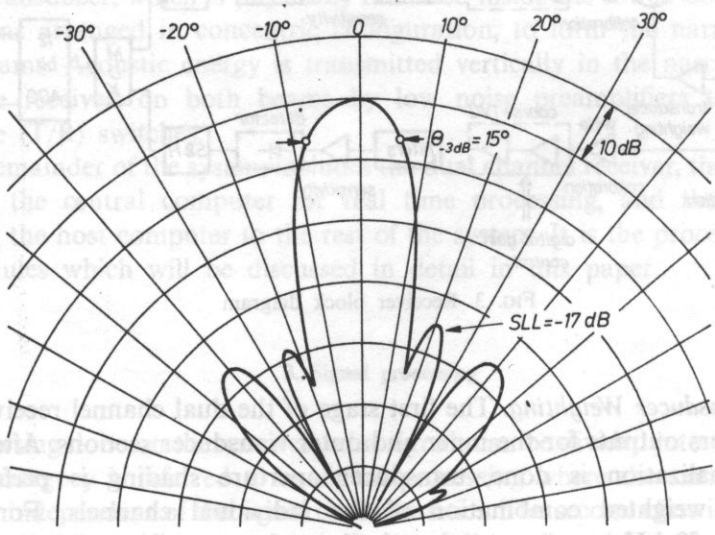


FIG. 5. Sample of the measured transducer beam pattern. Wide beam

side-lobes at approximately 17 dB. The beamwidth of the wide beam is approximately 15 degrees, while of the narrow beam is about 6 degrees both in transmit and receive modes. The simple binomial weighting applied for the receive narrow beam reduces the side-lobe level to about 20 dB.

3.1.2. Time variable gain. The received echoes must be progressively amplified to compensate for the geometric spreading and attenuation of the acoustic wave as it

travels to the target and back. For single fish targets, when TS estimation or fish counting is applied, the targets are considered as isotropic scatterers, and the range (R) dependent gain required to correct for the two way transmission loss is [5]:

$$G(R) = 40 \log \frac{R}{R_{\text{ref}}} + 2\alpha(R - R_{\text{ref}}) \quad [\text{dB}], \tag{1}$$

where R_{ref} reference range equals TVG start range, α attenuation coefficient.

For dense concentrations of fish (schools and scattering layers) when echo-integration is applied, the scattering from the plane is considered, which implies the one way compensation of spreading loss with range (i.e. $1/R^2$ low) and consequently the required time varied gain function has the form:

$$G(R) = 20 \log \frac{R}{R_{\text{ref}}} + 2\alpha(R - R_{\text{ref}}) \quad [\text{dB}]. \tag{2}$$

A great deal of attention was given in the design of an accurate, and reliable time varying gain (TVG) amplifier with dynamic range of 100 dB. Conventional TVG amplifiers use a PROM lookup table which is converted to a control voltage applied to a variable gain amplifier. The desired accuracy is not usually achieved and temperature drift problems may occur. Another solution considered was to digitize the signal directly, but A/D technology has not progressed far enough to provide the necessary dynamic range at a reasonable cost.

The time varying gain method in this system is a hybrid solution which combines the course TVG implemented using step gain amplifiers, and the digital TVG correction realized by high speed digital multipliers in DSP. The analog step amplifiers provide fixed gain steps of 10 dB, from 0 to 100 dB. The computer loads in a RAM a function which changes the analog gain in steps of 10 dB as a function of range. The coarseness is therefore at most 5 dB at a given range. The gain is increased by changing the resistance in the feedback network of an inverting amplifier. Two amplifier stages with dynamic range of 50 dB each are cascaded. The result is a precise, stable TVG amplifier with dynamic range of 100 dB.

3.1.3. Filtering and demodulation. The output of the TVG amplifiers is filtered using conventional multiple feedback bandpass filters and applied to a wide dynamic range envelope detector. The bandwidth of the filters is fixed at the bandwidth suited to the minimum pulse length of the transmitted signal (0.1 ms for 120 kHz, 0.3 ms for 50 kHz).

3.1.4. Digitization. The echoes from each beam are simultaneously sampled and multiplexed through a common A/D converter. The sampling rate was fixed at 80 kHz (40 kHz for each channel). The maximum bandwidth of the receiver is 10 kHz. The 12 bit converter is applied with least significant bit of 2.5 mV.

3.2 Digital signal processing

After the signals have been digitized, they are multiplexed through a common digital signal processor. The 12 bits of data for each channel corresponds to a maximum dynamic range of 74 dB. A block diagram of the digital signal processor is shown in Fig. 6. The first two sections of the processor are essentially part of the receiver. The filtered data is thresholded to remove noise and the bottom echo is detected. The data is buffered and stored in a FIFO buffer.

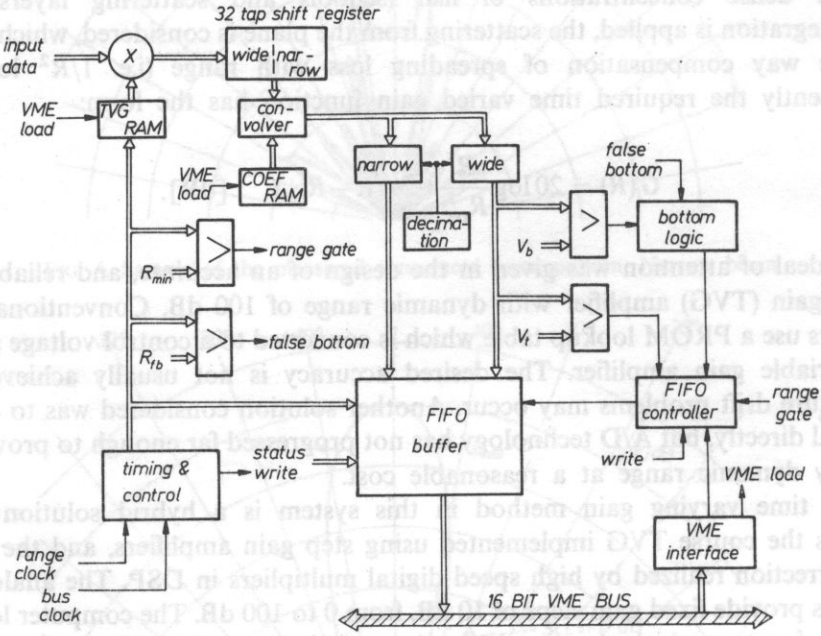


FIG. 6. Digital signal processor block diagram

3.2.1. Precision TVG *multiplier.* As mentioned in Section 3.1.2. the precision requirements of the TVG amplifier are transferred from the analog to the digital processing. A 12 bit correction value is used to adjust the signal level -5 to $+5$ dB to remove the coarseness of the analog TVG. Naturally, like the gain in the analog amplifiers, the correction value is time dependent and is also loaded in a RAM from the central computer.

The above scheme for TVG is very flexible: the same analog and digital hardware can be used for any spreading model, attenuation coefficient and reference range and consequently, arbitrary TVG functions which cover two decades of range are possible. The central computer simply loads the desired gain function and correction in the receiver and digital signal processor during system initialization. The attenuation constant α, and the TVG start and stop range are also programmable by the user.

3.2.2. Digital FIR *filter.* To optimize the echo detection (maximize S/N ratio) and simultaneously provide the best available resolution, the sonar system should adjust the receiver bandwidth to the different sounding pulse lengths in use. Typically, the pulse length ranges from a hundred microseconds to a tens of milliseconds, and consequently the receiver bandwidth, which should be approximately the inverse of pulse length, must vary also in the range of 1.5 to 2 decades. Conventional systems fix the number of pulse lengths and select either different filters or change the passive components of the filters. In either case, the variable pulse lengths result in redundant hardware.

Variable bandwidth is easily accomplished by a digital type filter. In the system, the bandwidth of the analog portion of the receiver is fixed at 10 kHz, which matches the minimum pulse length of 100 μs. The signals are digitized at a rate of 40 kHz so that a digital finite impulse response (FIR) filter can be used to decrease the bandwidth for longer pulse lengths. After TVG correction, the samples for each channel are stored in shift registers which are multiplexed through a common digital FIR filter. The order of the digital filter is 32 taps. For shorter pulse lengths, less than 32 samples are available, but the filter requirements are not as stringent.

The central computer sets the pulse length of the transmitted signal and loads the filter coefficients that match the pulse length into a RAM during system initialization. The FIR architecture was chosen for two reasons. Firstly, the number of samples in the received envelope is fixed and secondly the same digital multiplier used for TVG correction is used for the convolution of the filter impulse response and the delayed data.

The filter output is applied to time decimation logic which will select 1 of N samples, effectively reducing the sampling rate. The value of N is set by the central computer and can range from 1 to 16. The sampling rate can range from 2.5 to 40 kHz. Depending on the system mode of operation, the sampling rate is changed.

3.2.3. Hardware threshold detection. The filtered output for each channel is buffered and applied to a digital comparator which rejects all samples below the signal to noise (S/N) threshold, V_t. The S/N threshold is 12 bits wide and set by the central computer. The level of the threshold used for signal detection is different for each mode of operation. The central computer loads V_t in the digital signal processor and all samples for each channel are compared with this threshold. The comparator is disabled after transmission until a desired blanking range has been reached. The minimum range value is set by the central computer.

3.2.4. Bottom processing. Parallel to the S/N threshold hardware is preliminary bottom processing. The samples from the wide beam channel are thresholded with another threshold V_B to detect the echo from the bottom. The bottom threshold is loaded by the central computer. The computer can disable the bottom processor until a certain range, referred to as start of the bottom window, has been reached. This is necessary so that a large target at close range is not erroneously classified as the bottom return. If a bottom echo has not been detected within the bottom

window, the digital signal processor will generate an artificial false bottom echo which is written into the FIFO buffer. Once a bottom has been detected, the digital signal processor is disabled until the next transmission.

The hardware bottom processing serves only as preliminary bottom detection. One echo is tagged as the bottom and is used by the central computer for further bottom processing as well as optional bottom lock processing for echo integration.

3.2.5. FIFO *data buffering.* All samples exceeding the S/N threshold which are received after the blanking period and before the bottom echo are stored as an alarm block in a FIFO RAM buffer. The voltage amplitudes for each channel as well as the range are written in memory. An additional status byte is written for each alarm.

Buffer status indicators such as empty, half full and full are accessible by the central computer so data collection routines may be implemented. The buffer can contain 256 echo samples greater than the S/N threshold.

If the data rate begins to increase beyond an acceptable rate, the central computer can invoke a number of safeguards so that the buffer does not fill and data is lost. These safeguards include: time decimation, increasing the threshold, filling a temporary buffer and decreasing the ping rate.

4. Real time data processing

The central computer initiates the transmission and the first processing interval after transmission is reserved for direct memory access (DMA) data transfer of the central computer data to the host computer. The software range window from the minimum range to the bottom echo is used for data collection and real time processing. After the bottom echo has been detected, the real time processing is completed for the current ping and the output data is formatted and stored in the common memory for subsequent processing by the host computer.

The system has the following modes of operation:

i) Target strength estimation with optional fish sizing/echo counting.
ii) Echo integration.
iii) Echo integration with TS estimation.

Each of the two computer's share a $512k \times 8$ common memory which is mapped on the system bus.

All of the programs are written using the C programming Language [6].

4.1 Data collection and common memory

After transmission of the sounding pulse, the central computer polls the buffer status on the digital signal processor for non-empty condition. If there is data available, the FIFO is emptied into a large data buffer in the common memory. It is this larger buffer, called the raw data queue, that is used by the central computer for real time processing.

The common memory is used for storing data structures and variables, system parameters and lookup tables, and mailbox communication between two computers.

4.2. Target strength estimation

Routines are executed in real time by the central computer, (68010), which calculates the target strength (TS) and back scattering cross section (σ_{bs}) statistics from the individual fish echo narrow and wide beam amplitudes contained in the raw data queue. The theoretical derivation of target strength and back scattering cross section algorithms from the dual beam system are included in Appendix I. The TS and σ_{bs} data are calculated for each accepted single fish echo using formulae (I.9) and (I.11) and various data outputs are generated and displayed.

4.2.1. Echo detection and classification. The hardware amplitude S/N threshold (sec 3.2.3.) and software time windowing are used as a decision criteria for detection and parameter estimation of the echo pulses. Discrimination of the echoes as those from single or multiple fish targets is done. For each detected echo, an output echo data block is generated and stored in common memory and is available for further processing and/or transfer to the host computer.

A. Extraction of echo pulses – peak amplitude determination. The raw data queue is comprised of alarm blocks which are linked together. Sequential alarm blocks are searched for either a gap in range or a hardware tagged bottom alarm block. A gap is detected when successive alarms have ranges that differ by more than 1 sampling interval. As the raw data queue is being searched for end of the current echo pulse, the peak amplitudes and the pulse length for each beam is determined. Simultaneously, the pulse lengths are compared with a maximum pulse length (determined from the sounding pulse) for multiple fish target detection. If the pulse length is less than the maximum length, the amplitudes are stored in the echo pulse record and the echo pulse is initially classified as a single fish target. Once a gap is detected, the echo pulse record is complete and contains:

1) array of narrow-beam amplitudes (limited to maximum length),
2) array of wide-beam amplitudes (limited to maximum length),
3) narrow-beam peak amplitude,
4) wide-beam peak amplitude,
5) pulse length for each beam,
6) range for maximum peak,
7) single/multiple flag.

The echo pulse record is ready for further analysis for valid single fish detection. The above algorithm is illustrated by the flowchart in Figure 7.

B. Time windowing – pulse width determination. Once the echo pulse has been extracted and formatted in an echo pulse record with single target status, it is subjected to final pulse width check for classification as single, multiple or noise. The half-peak amplitude (−6 dB) pulse width criterion is illustrated by the flowchart in

Fig. 8. The peak amplitudes in the echo pulse record are halved and the amplitude arrays are searched to determine the number of contiguous samples for which the half-peak amplitude levels are exceeded. The resultant pulse lengths are then compared to a preset time window, adjustable automatically for varying pulse lengths. The time window resolution is the period of the carrier frequency, (i.e. 20 μs for 50 kHz, 8 μs for 120 kHz) and there is no ambiguity due to the synchronism of the transmit gate with the carrier frequency. The adjustable number of periods and sampling rate results in increased system flexibility and reliability of the pulse width classification criterion as the time window adapts optimally to the different operating conditions (e.g. S/N ratio).

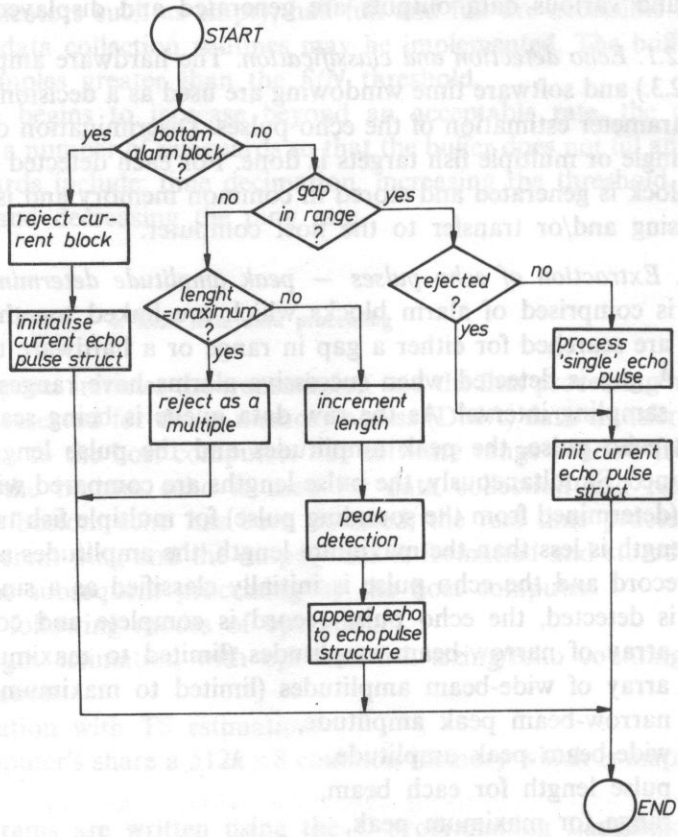


FIG. 7. Process alarm block in Raw Data Queue

Finally, if the half-peak amplitude pulse lengths are within the allowed time window, the echo pulse record is classified as a single fish target. If either the pulse length is greater than the maximum pulse length or less than the minimum pulse length, the echo is rejected as multiple or noise respectively.

C. Amplitude rato determination — beam pattern threshold. Those echo pulse records which are classified as single fish targets are subjected to beam pattern threshold criteria. The ratio of the squared peak-amplitude voltages (I.3) is compared with a computer-set beam pattern threshold T_B (I.4). Those echoes for which the squared voltage ratio does not exceed -3 dB are rejected. In addition, any

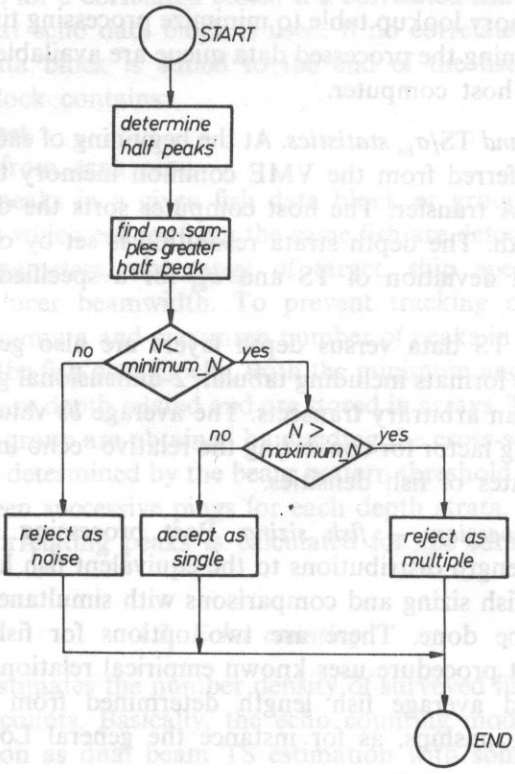


FIG. 8. Time windowing flowchart

"suspicious" echoes ($V_n > V_w$) are also rejected. As a result, the sampled volume becomes better defined and confined to the central portion of the beams. This assures the validity of the dual beam theory and permits the removal of the bias in final target strength estimates [7].

Those echo pulse records which pass the beam pattern threshold are finally considered as single fish targets, and due to their redundancy are reduced to an echo data block which contains:

1) ping number,
2) range,
3) narrow beam peak amplitude,
4) wide beam peak amplitude.

The output echo data blocks are linked together for each ping to form the processed data queue which is analysed after the detection of the bottom echo.

4.2.2 TS *and* σ_{bs} *calculation.* For each current ping after the bottom echo is processed the central computer modifies field 3 and 4 in each output echo data block from peak amplitudes to σ_{bs} and TS respectively using formulae (I.9) and (I.10). The logarithm is derived using a memory lookup table to minimize processing time. The list of output echo data block forming the processed data queue are available in common memory for transfer to the host computer.

4.2.3. Data transfer and TS/σ_{bs} *statistics.* At the beginning of each ping, the processed data queue is transferred from the VME common memory to the host computer memory using DMA transfer. The host computer sorts the data according to the data value and depth. The depth strata resolution is set by operator. The average value and standard deviation of TS and σ_{bs} for a specified fish population are calculated.

Histograms of TS data versus depth layers are also generated and may be presented in various formats including tabular, 2-dimensional graphics and 3-dimensional graphics for an arbitrary transects. The average *bs* value is used to calculate the integrator scaling factor for converting the relative "echo integrator output" into the absolute estimates of fish densities.

4.2.4. TS *data conversion — fish sizing.* Post processing routines convert the estimated target strength distributions to the equivalent fish length distributions so that both acoustic fish sizing and comparisons with simultaneous trawl-caught fish length data may be done. There are two options for fish length distribution estimation. The first procedure uses known empirical relationship between average target strength and average fish length determined from a least-mean-square analysis. Some relationships, as for instance the general Love equation [7]:

$$\mathrm{TS} = 19.1 \log L - 0.9 \log f - 62 \tag{3}$$

are stored in memory lookup tables. The second method utilizes the TS$-L$ relation in general form [5]:

$$\mathrm{TS} = m \log L + b. \tag{4}$$

This allows the operator to specify and enter arbitrary values of the regression coefficients. This method also allows for some post processing corrections of TS/L distributions.

4.2.5. Echo tracking — grouping of correlated peak amplitudes A post processing tracking routine is performed on successive transmissions to increase the reliability of the average target strength estimate by minimizing the effect of fish orientation (aspect). The output echo data blocks in the processed data queue from successive transmissions are sorted to form groups of echo data blocks defined as fish data blocks presumably corresponding to the same fish as seen in successive pings (the fish

data block can be thought of as fish echo traces as seen on the echogram records). The grouping of the data into the ordered sequence of peaks is based on the temporal/spatial correlation of the peaks. Peaks are said to correlate if they are seen in successive pings and have range tolerance within predefined limits [8].

For each echo data block in the current processed data queue, the list of fish data block is searched for a correlated block. If a correlated fish block is found, it is appended and the next echo data block is used. If no correlated fish data block is found, a new fish data block is added to the end of the list.

The fish data block contains:

1. Range of target.
2. Array of TS from each ping.

The number of peaks in a given fish data block or group (equivalent to the number of pings from which echoes from the same fish are detected) is a function of a several system parameters, viz: range of target, ship speed, ping repetition frequency and transducer beamwidth. To prevent tracking of false echoes and merging of groups, minimum and maximum number of peaks in a group are used to window the width of the fish data blocks. Both the minimum and maximum number of peaks in the group are depth related and are stored in arrays. The average value of number of peaks in a group are obtained by dividing the cross-sectional diameter of the beam at the angle determined by the beam pattern threshold and by the distance the ship moves between successive pings for each depth strata. Similarly the range tolerance used for correlating peaks is calculated for the same depth strata and stored in an array.

4.3. Echo counting

Echo counting estimates the number density of surveyed fish populations from the single fish echo counts. Basically, the echo counting mode utilized the same hardware configuration as dual beam TS estimation with some software options. Conventional echo counter systems using only a single beam are biased with some uncertainties in the sampled volume determination. There is an advantage in determining the sampled volume with a dual beam system, especially when the beam pattern threshold T_b is introduced to control the beam angle. Furthermore, the grouping of peaks in successive transmissions into the "fish" data blocks in the tracking routines supplies ready data on number of single fish.

The sampled volume in a single transmission for each specified depth strata is calculated and extended to the volume sampled during the processing interval corresponding to one fish trace in the echo tracking routine. Fish number data from the echo tracker program are sorted in depth and stored in two-dimensional array indexed by range and TS/length classes.

The counting number density is formed by dividing the fish number array elements by the sampled volume and normalizing to the standard volume. The output data constitutes the fish number density averaged over the given transects as well as fish number density by depth.

4.4. *Echo integration*

Echo integration is the second fundamental mode of the system operation, allowing estimation of the average density, $\bar{d}$, of the total surveyed fish population over specified transects. This in turn, allows the estimation of the total fish quantity or biomass by multiplying the density estimate by the volume of water occupied by the surveyed population.

The same system hardware is used for echo integration and TS estimation. Due to the very high sampling rate required for the digital filter, the signal is oversampled for echo integration. The filter output rate is reduced by digital decimation. The amount of reduction corresponds to the sampling period equal to half of the pulse length in use. The system's echo-integration software performs the integration and averaging functions in a digital manner on the sampled narrow beam channel output signals. The echo integration theory is presented in Appendix II.

Before the echo-integration survey is started, the programmable parameters must be entered in the host computer by the operator. Some of the more important parameters include:

1) number of pings in sequence or sequence length in time or distance intervals,
2) S/N threshold,
3) bottom window,
4) surface blank range,
5) surface or bottom locked mode,
6) range dependent layering,
7) overall range gate,
8) scaling constants.

After the system parameters have been setup, the system will integrate all returns for each range layer for all pings in the sequence. Also, bottom tracking is automatically done. Once the specified number of pings has elapsed, the blocks of relative or absolute densities are computed using predetermined scaling factors and formulae II.5 and II.6. The sequence index is incremented, the ping counter is reset and the processing begins for the next sequence.

In echo integration surveys, bottom detection and tracking are especially important for distinguishing between fish and bottom echoes to prevent the integration of large bottom returns. For bottom tracking, the system combines the elements of hardware (section 3.2.2) and software bottom processing. Once a bottom has been initially detected in hardware (signal exceeds the bottom threshold V_b) the programmable bottom window is centred automatically or manually around this range for the next ping. The amplitude of the hardware-detected bottom is used to determine the software bottom value, which may be set at -1 to -3 dB level of the hardware bottom. The software bottom provides a more reliable estimate of the actual bottom. All samples which form the leading edge of the bottom echo pulse are deleted and not integrated. The automatic bottom tracking is normally selected as soon as a bottom echo is within the bottom window. A computer set "false bottom" is injected if a bottom echo has not been detected. In bottom locked mode system

operation, the bottom layer may optionally be extended upward to some desired distance from the bottom.

The echo integration processing cycle is finished at the end of each sequence of pings. The data is sent to the host computer as a block of relative (M_j) or absolute (d_j) acoustic estimates of fish density in predefined depth layers. The host computer averages the estimates of fish density over all layers to give an overall fish density estimate.

Echo-integration may be performed in two different manners, ping (surface) locked or bottom locked.

4.5 Simultaneous echo integration and TS *estimation*

The third optional mode of system operation involves combining the two fundamental modes, which is desirable in some applications. In this case, the data are collected and TS estimation processing done to obtain an average value of σ_{bs} from those returns classified as single targets. Echo integration processing is performed on all echoes and fish densities are computed using σ_{bs} estimated from the same fish population. There are some hardware constraints that must be overcome to obtain reliable estimates. The first concern is the TVG function; TS estimation on single targets requires a $40 \log R + 2\alpha R$ while echo integration requires $20 \log R + 2\alpha R$. As mentioned earlier, the computer can generate arbitrary coarse gain function in the receiver. For this mode of operation the hardware TVG function of $30 \log \mathrm{R} + 2\alpha R$ is used and subsequent adjustment for each process is done by the central computer. The second concern is the S/N threshold set in the digital signal processor. The hardware threshold is set on relatively low level (appropriate for echo integration) and the data is thresholded again in the TS estimation routine with a higher threshold.

5. Conclusions

The application of digital signal processing techniques and the utilization of a high performance central microcomputer has resulted in a dual beam sonar system that is compact, versatile and low cost. The system hardware is used for echo sounding, echo counting, echo integration and *TS* estimation. Furthermore, the hardware can be configured so that simultaneous echo integration and TS estimation can be performed. The user interface to the system is through a user friendly, menu driven personal computer making the system both easy to use and learn.

This system represents the latest generation of equipment for fish stock survery and assessment. In addition to colour echograms, it provides virtually real time estimation of fish stocks, including 3-*D* plots of fish abundance versus position, histograms of fish size distributions and printed reports of size and quantity of fish by time or position interval. Full control of data presentation is available so that

data can be manipulated and presented as desired. Both raw and processed data can be stored on the PC's mass storage device. Data can be reprocessed onshore to provide comprehensive cruise reports without any additional hardware.

6. Appendix I – target strength estimation in a dual beam sonar

The target strength, TS, estimation method utilized in the dual-beam system can be categorized as 'direct in situ technique'. In this meaning, target strength data is extracted from free swimming fish of surveyed population by removing the beam pattern factor from the each individual fish echo. The applied dual beam target strength measurement technique, which we refer to as dual beam signal processing, uses the known dual beam concept [9], however its implementation is partly original both in hardware and software.

Acoustic energy is transmitted on a narrow beam $b_T(\theta, \phi)$, and received on both narrow $b_{RN}(\theta, \phi)$ and wide beams $b_{RW}(\theta, \phi)$. The voltage waveforms of a received echo at the output of each receiver channel, assuming a time varying gain of $40 \log R + 2\alpha R$, are given by:

$$V_N(t) = k_N{}^{1/2} \sigma_{bs}(\theta, \phi) b_T{}^{1/2}(\theta, \phi) b_{RN}{}^{1/2}(\theta, \phi) \cos(\omega_0 t + N_\phi), \tag{I.1a}$$

$$V_W(t) = k_W{}^{1/2} \sigma_{bs}(\theta, \phi) b_T{}^{1/2}(\theta, \phi) b_{RW}{}^{1/2}(\theta, \phi) \cos(\omega_0 t + W_\phi), \tag{I.1b}$$

Parameters k_N and k_W are sonar system constants for the narrow and wide beam channels respectively, which include the linear form of the source level SL of the sonar transmitter and overall voltage response VR of the receiver. The target back-scattering cross section at position (θ, ϕ) is represented by $\sigma_{bs}(\theta, \phi)$. Neglecting the time dependence of the received waveforms and assuming the circular symmetry of the transducer and thus its associated beam patterns $b(\theta, \phi) \triangleq b(\theta)$, the respective voltage squared amplitude ratio becomes:

$$\frac{V_N^2}{V_W^2} = \frac{k_N}{k_W} \frac{b_{RN}(\theta)}{b_{RW}(\theta)}. \tag{I.2}$$

If we further assume equal system constants $k_N = k_W$, equation (I.2) simplifies to the form:

$$\frac{V_N^2}{V_W^2} = \frac{b_{RN}(\theta)}{b_{RW}(\theta)}. \tag{I.3}$$

Equation (I.3) reveals the principal advantage of the dual beam configuration. The ratio of the squared narrow and wide beam amplitudes is equivalent to the ratio of the respective beam patterns for each received echo. Thus an unambiguous measure of the target position in the beam, referred to the acoustical axis of the dual beam transducer is obtained due to the unique relationship of the beam patterns ratio values and their associated variable angle. This is true under the assumption that the

target observed in successive transmissions has the same reflective beam pattern (e.g. isotropic scatterers) which not necessarily must be valid for fish targets wich possess directional back-scattering beam patterns $\sigma_{bs}(\theta, \phi)$ of variable value for different dish aspect [10]. The ratio for specified angle $\theta = \theta_{TH}$, has the sense of beam pattern threshold:

$$T_B = \frac{b_{RN}(\theta)}{b_{RW}(\theta)}\bigg|\theta = \theta_{TH}, \tag{I.4}$$

which bounds the sampling volume to the cone with angle, θ_{TH} as those echoes which have the ratio of squared voltages greater than the threshold will be detected. The cone with angle θ_{TH} is such that the wide beam pattern value is close to its maximum value $b_{RW} \simeq 1$, which in turn gives the approximate ralation:

$$b_{RN}(\theta) = \frac{V_N^2}{V_W^2}. \tag{I.5}$$

Once the beam pattern factor is known, the back scattering cross section of the target can be found. Solving equation (I.1a) for σ_{bs} we obtain:

$$\sigma_{bs} = \frac{V_N^2}{k_N b_T(\theta) b_{RN}(\theta)}. \tag{I.6}$$

If we assume that the transmit beam pattern and narrow receive beam patterns are approximately equal, especially when confined to small angles (in mainlobe – see Fig. 3) formula (I.6) simplifies to:

$$\sigma_{bs} = \frac{1}{k_N} \frac{V_N^2}{b_{RN}^2(\theta)}. \tag{I.7}$$

Comparing equation (I.7) with equation (I.2) we see:

$$\sigma_{bs} = \frac{V_N^4}{V_W^2} \cdot \frac{k_N^2}{k_W^2} \frac{1}{k_N b_{RW}^2(\theta)}. \tag{I.8}$$

Recalling that system constant is equal and the wide beam pattern is unity in the angles of interest, we obtain the simple expression for back scattering cross section:

$$\sigma_{bs} = \frac{1}{k_W} \frac{V_W^4}{V_N^2}. \tag{I.9}$$

Thus, by simultaneously measuring the narrow and wide beam voltages, the unknown beam pattern factor can be removed and as a result the back scattering cross-section of observed targets can be obtained. Consequently, the target strength TS, which is related to the back scattering cross section by the simple logarithmic relation:

$$TS = 10 \log \sigma_{bs}, \tag{I.10}$$

can be extracted from the received echoes using the following equation:

$$TS = 2W - N - K, \tag{I.11}$$

where $W = 20 \log V_N$ [dB], $N = 20 \log V_W$ [dB], $K = 10 \log k_W$ [dB].

Both the back scattering cross section and target strength primary data measured for individual fish are averaged over any specified population. Depending on the system's mode of operation each of these values estimated from the corresponding distributions (histograms) is used. The average target strength data are mostly used for comparison with and/or fish length statistics for fish sizing purposes. On the other hand, the average back scattering cross section is used to convert the relative echo-integration data into the absolute estimate of fish density, i.e. to scale the integrator output (see section 4.4).

7. Appendix II – echo integration theory

Under the assumption of non-coherent first-order scattering model for the collective echo from fish concentration, and with the TVG set at $20 \log R + 2\alpha R$, the running time average of the squared echo envelope over the range layer ΔR insonified in a given ping is proportional to the volume density of the scatterers (fish) in this layer [11]:

$$\int_{t}^{t+\Delta t} v_e^2(t)\, dt = \text{const}\, d_{\Delta R}. \tag{II.1}$$

If, after this first integration, the averaging over the specified surveyed transect is done, the average fish density in layer ΔR is estimated by:

$$d_{\Delta R} = (\text{const})^{-1} M, \tag{II.2}$$

where M is the mean integrator output over the transect and the proportionality constant is the product of the sonar system constant C_s and the average σ_{bs} of fish surveyed in layer ΔR. Thus, if the constant C_s has been measured and the σ_{bs} is known, or can be estimated (sec. 4.2), the average integrator output yields the absolute estimate of fish density or biomass depending on the form of σ_{bs}, i.e., related to the 'pure' TS, or to the TS_{1kq} [12, 5].

For each ping, the consecutive samples within each range layer must be squared and summed. For the jth layer on the kth ping, the partial sum of N squared samples with sampling interval i is given by:

$$S_{jk} = \sum_{i=1}^{N} (v_i)_{jk}^2. \tag{II.3}$$

In successive pings, the partial sums are added and after acquiring the data for all pings in a specified sequence of p pings, the transect sums are calculated for each layer:

$$S_j = \sum_{k=1}^{P} S_{jk} = \sum_{k=1}^{P} \sum_{i=1}^{N} (v_i)^2_{jk}. \tag{II.4}$$

After averaging these values by division by the total number of pings in the transect and by the number of samples in the layer, the "integrator output" for the jth layer is obtained, representing the relative estimate of fish density in the layer:

$$M_j = \sum_{k=1}^{P} \sum_{i=1}^{N} (v_i)^2_{jk}/n_j, \tag{II.5}$$

where $n_j = \sum_{k=1}^{P} n_{jk}$, is the total number of samples in layer j for p pings. These relative estimates are finally converted to the absolute fish density when scaled by appropriate overall system constant:

$$d_j = C^{-1} M_j, \tag{II.6}$$

where d_j – is the acoustic fish density estimate for the jth layer, $C - C_s \sigma_{bs}$ is the overall scale constant comprising the sonar system parameters and the average back scattering cross section of surveyed fish.

References

[1] KNIGHT *et al.*, *Digital signal processing for sonar*, Proceedings of the IEEE, **69**, 11 (1981).

[2] F. GOODENOUGH, *ADC Chips leap ahead both in speed and accuracy*, Electronic Design, (1986).

[3] ANON., *Model 181 dual-beam processor with Acquire and Target Strength Software*, Operator's Manual, Biosonics Inc., Seattle June (1985).

[4] R. BREDE, *Simrad QD Echo-Integrator, in*: FAO Fisheries Circular No. 778, Acoustic System for the Assessment of Fisheries, FAO, Rome 1984.

[5] A. STEPNOWSKI and J. BURCZYŃSKI, *The analysis of the calibration constant in the hydroacoustic system of fish abundance estimation*, In: Meeting on Hydroacoustical Methods for the Estimation of Marine Fish Populations, June 1979, vol. 2, Part a, [ed.] by J. B. Suomala, Cambridge, Massachusetts, C. S. Draper Laboratory Inc., 325–336, 1981.

[6] B. W. KERNIGHAN and D. M. RITCHIE, The C programming language, Prentice-Hall, New Jersey 1978.

[7] J. J. TRANYONOR, N. J. WILLIAMSON, *Target strength measurements of Walleye Pollock (Theragra Chalcogramma) and a simulation study of the dual beam method*, ICES/FAO Symposium on Fisheries Acoustics, Bergen, Norway June 1982, FAO Fisheries Report No. 300, FAO, Rome 1983.

[8] L. M. DICKIE, R. G. DOWD and P. R. BOUDREAU, *An echo counting and logging system ECOLOG for demersal fish size distributions and densities*, Can. J. Fish Aquat. Sci., **40**, 487–498 (1981).

[9] J. E. EHRENBERG, *A review of "in situ" target strength estimation techniques*, Proceedings of the ICES/FAO Symposium on Fisheries Acoustics, Bergen, Norway 1982.

[10] K. G. FOOTE, *Rather high frequency sound scattering by swimbladdered fish*, J. Acoustic Society Am., 78, 688–700 (1985).

[11] J. BURCZYŃSKI, A. STEPNOWSKI, R. SALAMON, W. MARTIN, *The hydroacoustic system on R/V*

"Professor Siedlecki" and its use in fish stock assessment by integration of thin depth layers, Rapp. P.-v. Reun. Cons. int. Explor. Mer, 170, 142–151 (1977).

[12] J. D. R. BAYONA, *Differences in the scaling of echo integrator survey results by Fisheries Research Institutes*, FAO Fisheries Circular No. 778, 107–127 FAO Rome 1984.

Received on June 1, 1987.

ARCHIVES OF ACOUSTICS
13, 1–2, 89–108 (1988)

ULTRASONIC SYSTEM FOR NONINVASIVE MEASUREMENT OF HEMODYNAMIC PARAMETERS OF HUMAN ARTERIAL-VASCULAR SYSTEM

TADEUSZ POWAŁOWSKI

Ultrasonics Department Institute of Fundamental Technological Research, Polish Academy of Sciences (00–049 Warszawa, Świętokrzyska 21)

This paper presents the working principle of an ultrasonic system constructed for the simultaneous noninvasive measurement of the blood flow velocity and, the diameter of the blood vessel. A bi-directional c.w. Doppler flowmeter was used to measure the blood flow velocity. The echo method was used to measure the blood vessel diameter and its changes. The values of the parameters measured were transfered to the computer connected on line with the ultrasonic measuring system. A programme was elaborated for computer analysis of a number of hemodynamic parameters determined from the measured blood flow velocity and the instantaneous diameter of a blood vessel. They are the blood flow rate, the blood pressure, the vascular input impedance and the elasticity of arterial vessel walls. Connected to a computer, the ultrasonic measuring system was used in examinations of the carotid arteries.

W pracy przedstawiono zasadę działania opracowanej aparatury ultradźwiękowej do równoczesnego, nieinwazyjnego pomiaru prędkości przepływu krwi i średnicy naczynia krwionośnego. Do pomiaru prędkości przepływu krwi użyto dwukierunkowego dopplerowskiego przepływomierza fali ciągłej. W pomiarze średnicy naczynia krwionośnego i jej zmian zastosowano metodę echa. Wartości mierzonych parametrów przesyłane były do komputera, który połączony był „on line" z ultradźwiękowym aparatem pomiarowym. Opracowano program analizy komputerowej szeregu parametrów hemodynamicznych wyznaczanych na podstawie mierzonej prędkości krwi i chwilowej średnicy naczynia krwionośnego. Są to prędkość objętościowa krwi, ciśnienie krwi, wejściowa impedancja naczyniowa, elastyczność ścianek naczyń tętniczych. Połączony z komputerem ultradźwiękowym system pomiarowy został zastosowany do badań tętnic szyjnych.

1. Introduction

The recent years saw a serious increase in the number of diseases of the vascular system. According to 1983 data, about 40% of all deaths is caused by diseases of the circulatory system. Of these, 40% are caused by atherosclerosis [15]. As a result of

atherosclerotic vessel diseases, the lumen of the blood vessel narrows or closes. Most often, this occurs in large or medium arteries most significant for the human organism, namely the coronary, cranial and renal arteries, the aorta, the arteries of the lower limbs and mesenteric arteries. Atherosclerosis can develop for a long time, originating in the early years of life. Often clinical symptoms appear very late when the disease is much advanced and it is too late to treat its results. Therefore, there is the constant need for doing research on the methods and diagnostic equipment permitting the early identification of pathological changes in man's arterial vessels. In this direction, much progress was due to the introduction in medical diagnosis of the noninvasive ultrasonic technique, in particular ultrasonic Doppler flowmeters permitting the estimation of blood flow rate in blood vessels. They were widely applied, e.g., in examinations of patency of extracranial carotid arteries [22] and the arteries of the lower limbs [11].

The ultrasonic Doppler method does not ensure full diagnosis of man's vascular system. This results mainly from the fact that blood flow rate is just one of the many hemodynamic parameters describing the state of the circulatory system. The equally important parameters include the blood pressure, the complex vascular input impedance and the elasticity of arterial vessel walls.

An example of the new approach to diagnosis of the vascular system is the noninvasive ultrasonic method and system constructed at the Department of Ultrasonics, Institute of Fundamental Technological Research, Polish Academy of Sciences, for the noninvasive examination of the blood flow rate the vascular input impedance and elasticity of the carotid arteries [18, 19].

The vascular input impedance is defined by the ratio between the blood pressure and the blood flow rate for successive harmonic frequencies of the work of the heart [1, 4, 5]. In the method in question, the blood pressure is determined from displacements of the arterial blood vessel [18, 19]. These displacements are measured with accuracy up to 0.03 mm over the same time and vessel cross-section as the blood velocity. The impedance is calculated by the discrete Fourier transform of the time courses of the blood pressure and flow rate. It is implemented on a MERA–60 computer connected on line to the ultrasonic measuring system.

Measurements of displacements of the arterial vessel walls can also serve for evaluation of the elasticity of arteries. It is very significant from the point of view of the diagnosis of the complex of diseases of blood vessels, called arteriosclerosis. This group includes all changes in arteries which lead to fibrosis of part or whole of the arterial wall, and, as a result, to a change in its elasticity.

The purpose of this study is to present the principal part of the system constructed for the examination of the blood flow rate, vascular input impedance and elasticity of arterial walls. It is an ultrasonic meter of blood velocity, wall displacements and blood vessel diameter. This paper also discusses the basic assumptions and dependencies adopted in computer analysis in the determination of the vascular input impedance and elasticity of the carotid arteries from blood flow velocity and the instantaneous diameter of a blood vessel.

2. Method and system for digital measurements of displacements of blood vessel walls

The noninvasive ultrasonic echo method was used to measure displacements of blood vessel walls. Information on the displacement amplitude of a vessel wall is obtained by measuring the distance between the ultrasonic probe and the examined wall of a blood vessel. Changes in the diameters of peripheral arterial vessels under increased blood pressure are of the order of 0.1 mm. The investigation of so small displacements require high accuracy in tracing and measuring the position of the echo detected from the vessel wall.

In 1972 HOKANSEN [10] proposed an idea of tracing and measuring displacements of a blood vessel wall by means of the ultrasonic echo method. Applying an analog system of his own construction, he measured movements of walls of the femoral artery. In 1985, HOEKS [9] proposed a different conception of the measurement. For this purpose he used a multi-gate pulsed Doppler flowmeter. Using it, he measured the relative changes in the diameter of the common carotid artery. Moreover, this method raises large objections about the accuracy of representing vessel wall displacements.

This study presents a method for digital measurements of displacements of walls of a blood vessel and its diameter. It is an extension of the measurement conception proposed in 1982 by GROVES, POWAŁOWSKI and WHITE [7]. The general idea of this method is shown in Fig. 1.

An ultrasonic probe set perpendicularly to the blood vessel (see Fig. 4) emits towards it at intervals T_p impulses of the ultrasonic wave. The measurement of the position of the vessel wall with respect to the ultrasonic probe consists in counting clock impulses over the time between the trigger impulse initiating the transmitted impulse and the rising slope of the echo detected from the vessel wall. The time measured by the digital method is the basis for calculating the instantaneous distance d between the ultrasonic probe and the surface of the blood vessel wall. For a successive nth measurement cycle, this distance is

$$d_n = \frac{c N_n}{2 f_z}, \tag{1}$$

where N_n is the number of clock impulses counted in the nth measurement cycle, f_z is the frequency of the clock impulses and c is the ultrasonic wave velocity in the medium investigated.

A level comparator was used to determine unambigously the time when the echo slope occurs. In the comparator, the echoes detected from biological structures are transformed into a series of rectangular signals. The measurement of the position of a chosen echo slope is repeated for every cycle of work of the impulse transmitter. It is assumed in the further description of the measurement that the traced slope of the echo E2 overlaps the end of the preceding echo E1. These two echoes displace in

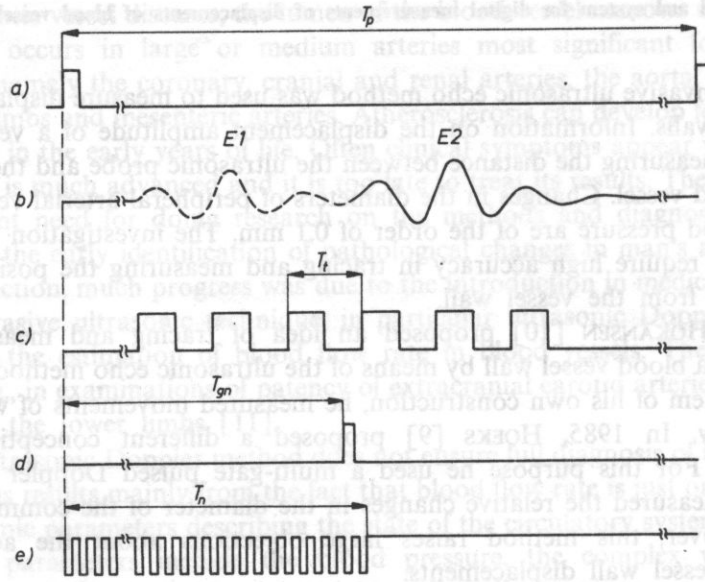


FIG. 1. The principle of the digital measurement of the time variable position of the blood vessel wall: a) trigger impulses, b) echos detected from the external and internal blood vessel walls, c) echos at the output of the level comparator, d) tracing gate, e) clock impulses

the same direction at the same velocity. A situation resembling the above one can occur for echos detected from the external and internal surfaces of a blood vessel wall.

To identify the chosen echo slope, a tracing gate is generated before it. The logical unit of the system stops the time measurement with the clock impulse which occurs after the appearance of the first rising echo slope after the tracing gate. In a current measurement cycle, the position of the tracing gate depends on the position of the echo slope traced in a preceding cycle, and is

$$T_{gn} = T_{n-1} - t_0, \tag{2}$$

where T_{n-1} is the time measured digitally between the trigger impulse and the traced echo slope in the $(n-1)$th measurement cycle, T_{gn} is the time delay of the tracing gate with respect to the trigger impulse in the nth measurement cycle and t_0 is a constant time shift. The time shift t_0 is implemented digitally in the form of N_0 clock impulses.

Fig. 2 shows two conceptions of the detection of the traced echo slope after the appearance of the tracing gate. If the tracing system responds to the positive level of the echo at the comparator output (Fig. 2a), the time shift t_0 of the tracing gate can

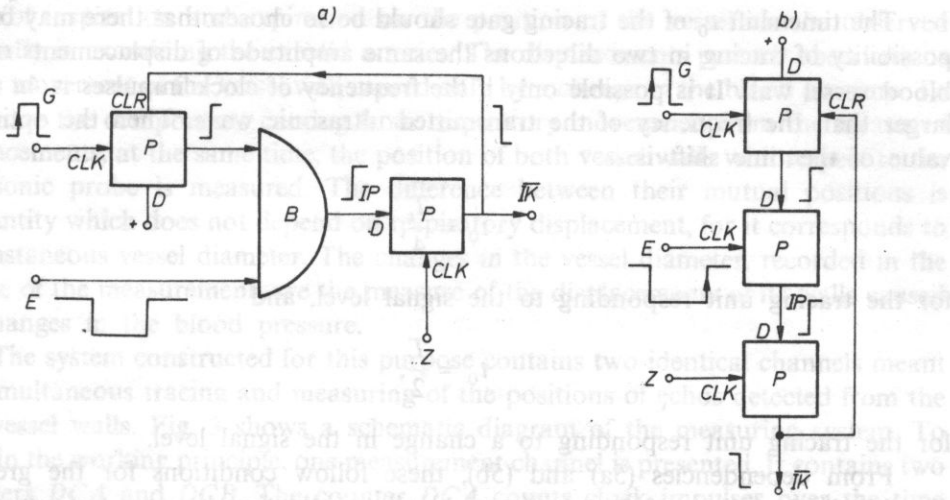


FIG. 2. Two ways of the identification of the positive echo slope by the tracing system: a) the unit responding to a positive echo level, b) the unit responding to changes in the echo level from low to high. *B* – *AND* gate, *P* – *D* Flip Flop, *E* – echo of the comparator output, *G* – tracking gate, *Z* – clock

be contained within the following limits:

$$T_z \leqslant t_0 < \frac{T_t}{2}, \tag{3}$$

where T_t is the ultrasonic wave period and T_z is the clock signal period.

The version of the tracing unit of this type was proposed by GROVES et al. [7]. There is, however, another solution in which the time shift t_0 can be twice as much. This applies to the tracing system which responds to a change in the level from a low to a high one of the echo signal at the comparator output (Fig. 2 b). Then, the time shift of the tracing gate can be

$$T_z \leqslant t_0 < T_t. \tag{4}$$

The difference between these two solutions can readily be explained by assuming that the tracing gate is shifter with respect to the traced echo slope by the time $T_t/2 < t < T_t$. It often occurs in the course of the positive phase of the echo preceding the traced slope. At the comparator output, when the tracing gate appears, there is a high level. For the first system (Fig. 2 a), it is false information about the occurrence of the traced echo slope, and the tracing unit generates the impulse *IK* which ends the time measurement. In the second solution (Fig. 2 b), the tracing unit does not respond to the high level of the comparator. It gives a signal for ending the measurement only when at the comparator output there is a change in the signal level from a low to a high one, corresponding to the appearance of the traced echo slope.

The time shift t_0 of the tracing gate should be so chosen that there may be the possibility of tracing in two directions the same amplitude of displacements of the blood vessel wall. It is possible only if the frequency of clock impulses is $4n$ times larger than the frequency of the transmitted ultrasonic wave. Then the optimum value of the time shift is

$$t_0 = \frac{T_t}{4}, \tag{5a}$$

for the tracing unit responding to the signal level, and

$$t_0 = \frac{T_t}{2}, \tag{5b}$$

for the tracing unit responding to a change in the signal level.

From dependencies (5a) and (5b), these follow conditions for the greatest possible displacements of the traced vessel wall over the time between two successive impulses from the transmitter of the measuring system. They are equal respectively to 1/8 and 1/4 of the wavelength of the transmitted ultrasonic wave in the medium under study. For the frequency of the transmitted wave of the order of MHz, these are very small displacements of the order of hundredths or tenths of a millimetre. Hence, there follows a general condition which should be satisfied by the measuring system in tracing the displacements Δu occurring over the time Δt:

$$\frac{cf_p}{2mf_t} \geqslant \left(\frac{\Delta u}{\Delta t}\right)_{\max}, \tag{6}$$

where f_t is the frequency of the transmitted wave, f_p is the repetition frequency of the transmitted impulses and m is a factor whose value depends on the time delay t_0 and is 4 for $t_0 = T_t/4$, or 2 for $t_0 = T_t/2$. The satisfaction of the above condition is restricted in the range of selecting the frequency ratio f_p/f_t. For, on the other hand, the values of the two frequencies are conditioned by the necessity of obtaining the needed resolution and range of measurement.

In the measuring system described here, meant mainly for the examination of the carotid arteries, the frequency of the transmitted ultrasonic wave is 6.75 MHz, whereas the repetition frequency of the transmitted impulses is 18 KHz. The maximum velocity of wall displacements caused by the blood pressure in these arteries does not exceed a dozen or so mm/s. From condition (6), the measuring system makes it possible to map fully the movements of vessel walls if the two above ways of detecting the traced echo slope are applied. On the other hand it should be noted that the tracing unit responding to a change in the signal level permits the tracing of displacements which are twice as fast compared with the unit responding to the signal level. Due to this, it ensures more stable measurement, in particular if there are additional displacements of the blood vessel with respect to the ultrasonic probe. One of the factors which cause the changes is the respiratory motion. The

effect of inspiration and expiration on the position of the vessel can be observed distinctly in examining the carotid arteries. The displacements perturb observations of the movement of the vessel wall caused only by a change in the blood pressure. To eliminate the components coming from respiratory movements from the measured displacements, at the same time, the position of both vessel walls with respect to the ultrasonic probe is measured. The difference between their mutual positions is a quantity which does not depend on respiratory displacement, for it corresponds to the instaneous vessel diameter. The changes in the vessel diameter, recorded in the course of the measurement, are the measure of the displacements of its walls caused by changes in the blood pressure.

The system constructed for this purpose contains two identical channels meant for simultaneous tracing and measuring of the positions of echos detected from the two vessel walls. Fig. 3 shows a schematic diagram of the measuring system. To explain the working principle, one measurement channel is presented. It contains two counters *DCA* and *DCB*. The counter *DCA* counts clock impulses over the time interval between the trigger impulse releasing the transmitter of the impulse wave and the echo front detected from the blood vessel wall. At the same time, in the counter *DCB*, clock impulses are subtracted from the content of the counter *DCB* recorded in a preceding cycle. The difference value at the output of the counter *DCB* is compared with the programmed quantity N_0 in the digital comparator C_d. The comparator C_d generates the tracing gate *G* as soon as a numerical value equal to N_0

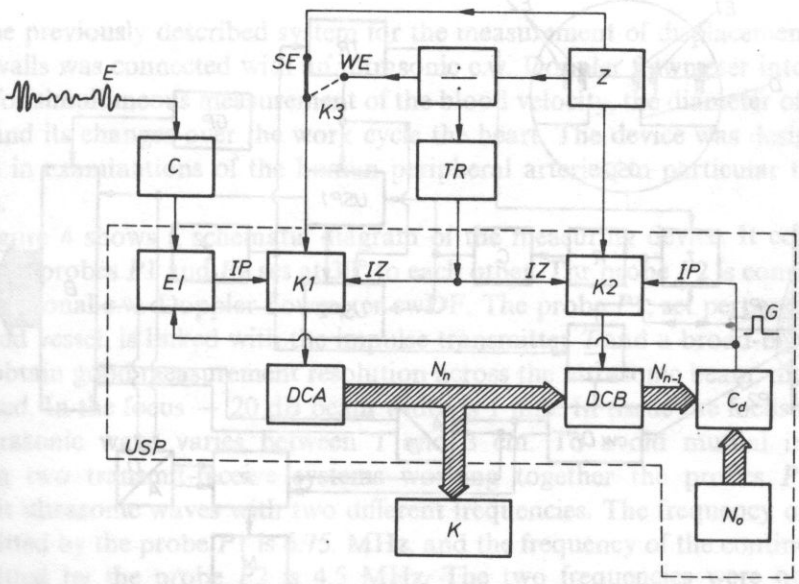


FIG. 3. A schematic diagram of the digital system for tracing and measuring the position of the blood vessel wall: *WE* – choice of the echo slope, *SE* – tracing the chosen echo slope, *TR* – trigger, *Z* – clock

appears at the output of the counter *DCB*. At the moment that the rising slope of the gate *G* appears, the sum of the clock impulses counted by the counter *DCA* is smaller by N_0 than the number of clock impulses counted by this counter in a preceding cycle. This means that, in accordance with the measurement method adopted (formula 2), the tracing gate *G* is displaced with respect to the position which the traced echo slope took in a preceding cycle.

The gate *G* is supplied at the input of the echo identification unit *EI*, to which echos detected from the blood vessel walls are supplied through the level comparator *C*. The unit *EI* generates the impulse *IP* (see Fig. 2), when the first echo slope after the gate *G* occurs. This impulse eliminates by the logical keys *K*1 and *K*2 the clock impulse from the inputs of the counters *DCA* and *DCB*. After rewriting the contents of the counter *DCA* into the computer *K* and the counter *DCB*, the counter *DCA* is reseated. A new measurement cycle begins with another trigger impulse *IZ*. The digital data obtained at the output of the counter *DCA* are transformed in the computer into information about the instantaneous position of the surface of the blood vessel wall with respect to the ultrasonic probe.

The tracing of the position of the chosen echo slope requires that the tracing gate *G* should be set up before it. It is only then that the above-described process of automatic tracing of the chosen echo slope can take place.

The initial position of the tracing gate *G* with respect to the echo slope is set digitally by introducing into the counter *DCA* of such a number of impulses which corresponds to the position of the chosen echo slope with respect to the trigger

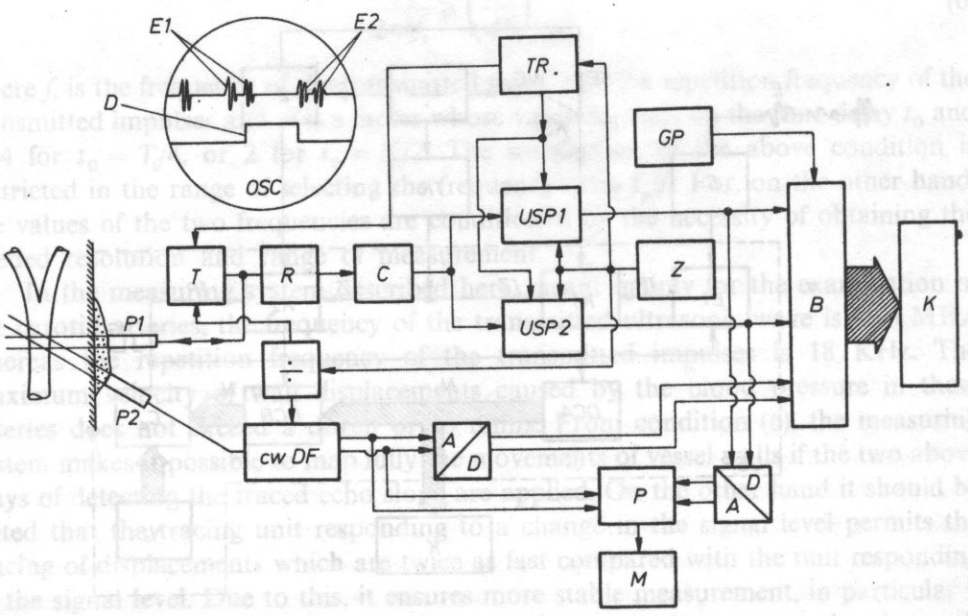


FIG. 4. A schematic diagram of the system for simultaneous measurement of blood velocity, vessel diameter and displacement of blood vessel walls

impulse. The current position of the gate is controlled by its presentation along with echos on the measuring system's screen. To set precisely the position of the tracing gate *G*, into the counter *DCA*, through the key *K3*, clock impulses are replaced by slowly variable ones. The time their introduction is manually controlled, with the echo identification unit *EU* switched off. The rewriting of the current numerical value from the counter *DCA* to the counter *DCB* and the generation of the tracing gate *G* by the comparator C_d take place in the same way as in the course of the automatic tracing of the echo, i.e., synchronically to the trigger impulses.

To measure simultaneously the displacements of the two walls of the blood vessel and the vessel diameter, two symmetrical channels for tracing and measuring the positions of the echo, called later *USP*, were applied. Each of them contains two 12-bit counters *DCA* and *DCB*, the digital comparator C_d and the echo identification unit *EI*. The inputs of these two channels are connected in parallel with the level comparator *C*. The measured data from the outputs of the two channels are assigned to the buffers from which they are then entered into computer for further processing and computations (see Fig. 4).

In the system in question, the delay time between the trigger impulse and the echo detected from the vessel wall is measured by counting clock impulses of 27 MHz. In effect, this permits the representation of the amplitude of displacements of the blood vessel walls an accuracy up to 0.03 mm.

3. Meter of blood velocity wall displacements and blood vessel diameter

The previously described system for the measurement of displacement of blood vessel walls was connected with an ultrasonic c.w. Doppler flowmeter into one joint set-up for simultaneous measurement of the blood velocity, the diameter of the blood vessel and its changes over the work cycle the heart. The device was designed to be applied in examinations of the human peripheral arteries, in particular the carotid arteries.

Figure 4 shows a schematic diagram of the measuring device. It contains two ultrasonic probes *P*1 and *P*2 set at 30° to each other. Thr probe *P*2 is connected with a bi-directional c.w. Doppler flowmeter cwDF. The probe *P*1, set perpendicularly to the blood vessel, is linked with the impulse transmitter *T* and a broad-band receiver *R*. To obtain good measurement resolution across the ultrasonic beam, the probe *P*1 is focused. In the focus – 20 dB beam width is 1 mm. In tissue the focusing zone of the ultrasonic wave varies between 1 and 3 cm. To avoid mutual interference between two transmit-receive systems working together the probes *P*1 and *P*2 transmit ultrasonic waves with two different frequencies. The frequency of the wave transmitted by the probe *P*1 is 6.75. MHz, and the frequency of the continuous wave transmitted by the probe *P*2 is 4.5 MHz. The two frequencies were obtained by dividing the clock frequency of 27 MHz, respectively, by 4 and 6.

The repetition frequency of the transmitted ultrasonic wave impulses is 18 kHz.

This provides the measurement range of 4.2 cm into the patient's body. The echos obtained in the course of the measurement at the output of the receiver R are transformed into rectangular signals by the level comparator C. The echos formed in this way are fed to two digital units for tracing and measuring the position of the echo from the front wall (the channel *USP* 1) and from the back wall (the channel

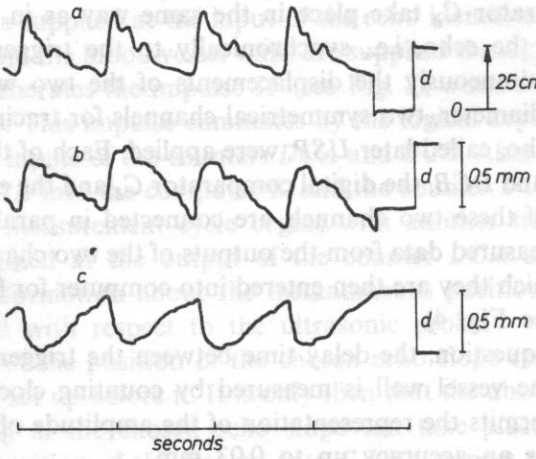


FIG. 5. Time courses of the blood velocity (a) and displacements of the back (b) and front (c) vessel walls recorded in the course of measurement in the common carotid artery in a 40-year-old man

USP 2) of the blood vessel. The 11-bit data obtained at the outputs of the two systems are fed to the output buffers B of the unit, along with the two 8-bit pieces of information about the blood velocity measured simultaneously in two directions. These data are then fed into the computer K, with the sampling frequency FP imposed by the generator GP. This frequency is adjusted depending on the working frequencies of the examined patient's heart. It is so chosen that there are 64 data of each of the measured quantities for the mean period of the work period of the heart. This condition results from the 2^n – point fast Fourier transform applied in the further computer data analysis.

Data on the blood velocity are obtained from the measured difference (Doppler) frequency between the frequency of the transmitted wave and that detected from the flowing blood by the probe P2 of the flowmeter cwDF. The mean Doppler frequency f_{zc} measured by the zero-crossings method is proportional to the mean velocity v_s of the blood flow through the cross-section of the blood vessel, according to the dependence

$$V_s = a\frac{cf_{zc}}{2f_n \cos\theta}, \tag{7}$$

where c is the ultrasonic wave velocity in the medium under study, f_n is the frequency of the transmitted ultrasonic wave, θ is the angle between the direction of the transmitted and detected ultrasonic wave and the axis of the blood vessel, and a is the proportionality coefficient.

The proportionality coefficient in formula (7) depends on the velocity profile of the blood flow studied and on the ratio between the ultrasonic beam width and the blood vessel diameter [6, 17]. Assuming that the mean profile of the blood flow rate in the cross-section of the blood vessel is contained between a parabole and a flat profile, and that the ultrasonic beam is wider than half the vessel diameter, in the first approximation the coefficient a would be 0.85 [6, 17].

In determining the blood velocity from the Doppler frequency f_{zc} measured by the flowmeter, it was assumed that the angle θ in dependence (7) is 60°. This follows from the constant angle 30° between the probes $P1$ and $P2$, and from the assumption that the probe $P1$ is set perpendicularly to the axis of the vessel. It was assumed that the perpendicular setting of the probe $P1$ with respect to the blood vessel would be indicated by obtaining the maximum amplitude of the echoes from the two walls of the vessel. In the course of the measurement these echoes are obtained on the oscilloscope screen *OSC* of the device. At the same time, the screen shows the measured diameter of the vessel under study in the form of a gate.

The data obtained in the measurement of the blood velocity, the diameter of the blood vessel and displacements of its two walls are presented in the form of analog courses on the recorder P or the memory monitor M. The analog recording serves for controlling the measurement data obtained before they are fed into a computer for further analysis. The analog-to-digital (A/D) and digital-to-analog (D/A) converters were applied to transform the measured quantities into digital and analog values (Fig. 4). Figure 5 shows, as an example, the blood flow rate and displacements of the front and back walls of the artery recorded in the course of measurement in the common carotid artery.

The measurement of the internal diameter of the blood vessel requires good resolution of the transmit-receive impulse system. For this purpose, apart from the previously mentioned focusing of the probe $P1$, in the measurements, a narrow transmitted impulse was used with a duration of 0.3 μs (2 high-frequency cycles of the transmitter), corresponding to its length of 0.45 mm in tissue. As a result it made it possible, in the case of the common carotid arteries, to obtain single echoes from the external and internal surfaces of the blood vessel walls. Fig. 6 shows the measured displacements of the internal and external displacements of the surface of the back wall of the common carotid artery with its thickness. In the case studied it was 1 mm.

The single echos obtained from the two surfaces of the vessel walls made it possible to measure its internal diameter. Having at the same time, information on the blood velocity across the measured internal cross-section of the vessel, it is possible to determine on this basis the blood flow rate. The ultrasonic impulse Doppler method applied so far for this purpose does not permit the so precise measurement of the internal diameter of the blood vessel. This mainly results from the fact that the information on the vessel diameter is taken from the measured spatial distribution of the velocity profile of blood flow [2, 6] which for low velocities close to the vessel wall is falsified as a result of the filtration of the Doppler signal coming from the displacing blood vessel walls. A general view of the measuring system in question is shown in Fig. 7.

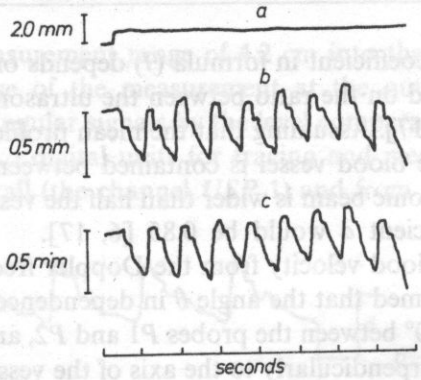


FIG. 6. The thickness of the carotid artery wall (a) and displacements of the external (b) and internal (c) surfaces of the wall studied, recorded in the course of the measurement. The mean internal diameter of the vessel examined was about 8 mm

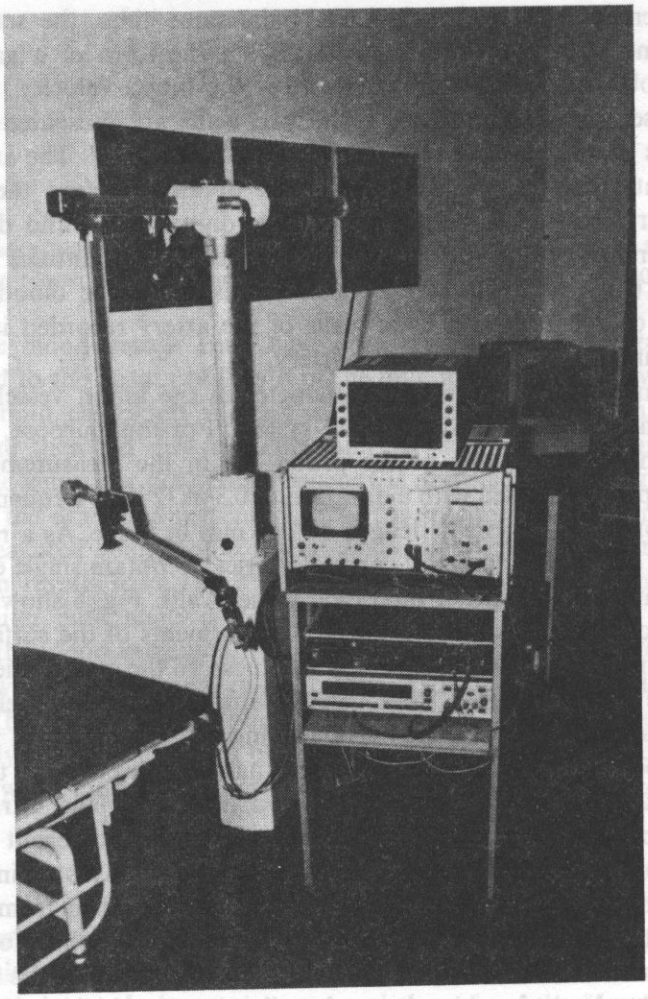


FIG. 7. A general view of the ultrasonic system

4. Computer analysis of hemodynamic parameters

The ultrasonic measuring system described above was used in examination of the carotid arteries. The data obtained at the output of the ultrasonic device are fed into a MERA-60 computer (corresponding to PDP-11) and written on a floppy disk. The number of data recorded each time in the course of the measurement contain information on 640 successive values of each of the four parameters measured simultaneously. They are: the digitally measured times t_{n1} and t_{n2} between the transmitted impulses and the echoes from the front and back walls of the blood vessel and the Doppler frequencies f_{2c1} and f_{2c2} measured for blood flow away from and towards the ultrasonic probe.

According to earlier determinations, the sampling frequency with which data are fed into a computer is so set that 64 data on each of the measured quantities correspond to one mean work cycle of the heart. This means that the number of recorded data comes from about 10 work cycles of the heart.

Along with the above-mentioned data, the value of the sampling frequency F and those of the systole pressure P_s and the diastole pressure P_d measured with a cuff in the brachial artery, with the patient in supine position, are recorded on a computer disk. The data set formed in this way is the basis for further computer analysis of the following hemodynamic parameters:

a) the blood flow rate Q,

b) the blood pressure P,

c) the input vessel impedance Z,

d) the relative change in the blood vessel diameter over the work cycle of the heart $\Delta D/D$,

e) the pulse wave velocity c_t,

f) the coefficient of rigidity of the blood vessel wall α.

The flow diagram of the computer analysis is shown in Fig. 8. From n successive measured data, respectively from dependencies (1) and (7) the instantaneous numerical values of the blood flow velocity $v_1(k)$ and $v_2(k)$ away from and towards the ultrasonic probe, and the distances $d_1(k)$ and $d_2(k)$ between the ultrasonic probe and the internal surfaces of the front and back walls of the blood vessel are determined. In turn, these data serve to determine k successive values of the internal vessel diameter $D(k)$ and the blood flow rate $Q(k)$.

Further computer analysis is performed separately for particular work cycles of the heart. The indicator which identifies a successive cycle is the maximum value of the rate Q occurring in the systole phase. The beginning of the studied cycle K_0 is established at the beginning of the systole phase. For each cycle, the beginning of the analysis is equally shifted with respect to the maximum value of the blood flow rate Q.

The programme of computer analysis assumed the possibility of determine the blood flow rate $Q(k\Delta t)$ and the blood vessel diameter $D(k\Delta t)$ on the basis of a chosen number IL of successive work cycles of the heart. Programmatically, this is implemented according to the following algorithms:

$$Q(k\Delta t) = \frac{1}{IL}\sum_{m=1}^{IL}\frac{\pi v_m(k\Delta t)\,D_m^2(k\Delta t)}{4}, \tag{8}$$

$$D(k\Delta t) = \frac{1}{IL}\sum_{m=1}^{IL} D_m(k\Delta t), \tag{9}$$

where $k = 1, 2, \ldots, 64$ is the number of a successive datum in the work cycle of the heart, m is the number of a successive cycle, $\Delta t = 1/FP$ and $v_m(k\Delta t)$ is the studied linear blood velocity in the mth cycle of the heart.

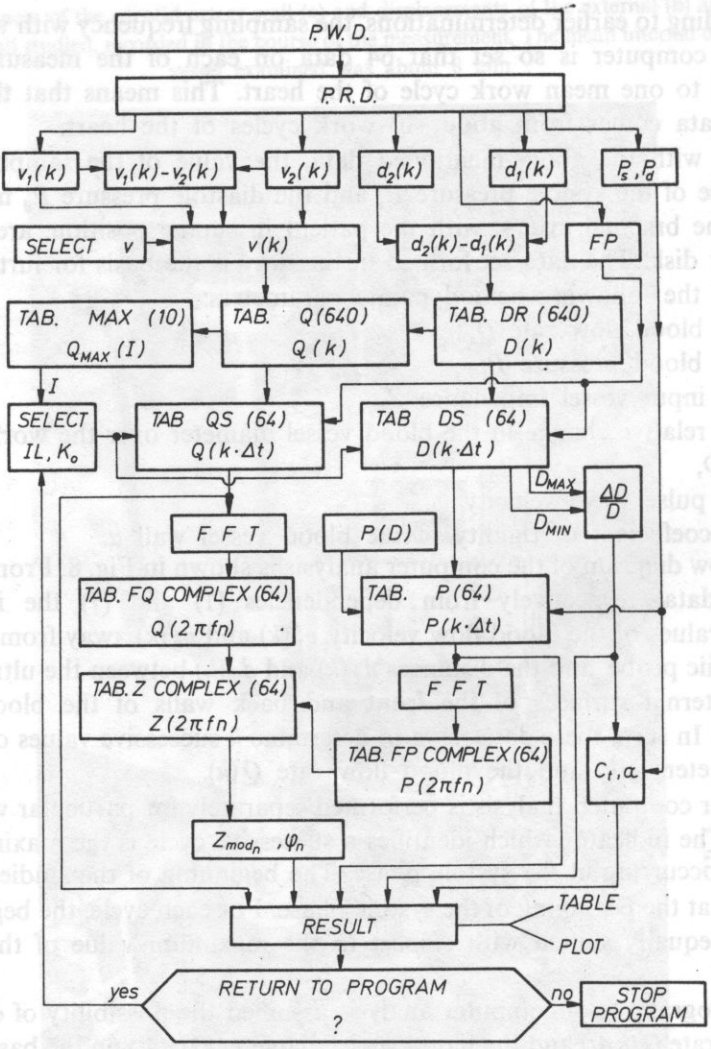


FIG. 8. The network of operations of computer analysis, *PWD* — programme for data recording on a disk, *PRD* — programme for reading data from the disk

According to the adopted network of operations the quantities $Q(k\Delta t)$ and $D(k\Delta t)$, described by formulae (8) and (9) form the basis for determining the mean values of the other hemodynamic parameters.

The blood pressure is another parameter determined from measurement data. It is calculated from instantaneous changes in the blood vessel diameter over a work cycle of the heart. The calculations assumed an exponential dependence between the blood pressure P and the blood vessel cross-section area S, given by the following function:

$$P = P_0 \exp(\gamma S), \tag{10}$$

where P_0 and γ are constant coefficients.

After transformation, this function becomes

$$P(D) = P_d \exp\left[\frac{D^2 - D_d^2}{D_s^2 - D_d^2} \ln \frac{P_s}{P_d}\right], \tag{11}$$

where D_s and D_d are the vessel diameters for the systole pressure P_s and the diastole pressure P_d. The pressure determined in this way is calibrated in absolute units by the systole pressure P_s and the diastole pressure P_d measured in the brachial artery. The values of the two pressures are respectively subordinated to the maximum $D_{\max}$ and minimum $D_{\min}$ diameters of the blood vessel. The blood pressure over the work cycle of the heart, calculated on this basis, is expressed by the following dependence

$$P(n\Delta t) = P_d \exp\left[\frac{D^2(n\Delta t) - D_{\min}^2}{D_{\max}^2 - D_{\min}^2} \ln \frac{P_s}{P_d}\right]. \tag{12}$$

The instantaneous blood pressure $P(k\Delta t)$ and the blood flow rate $Q(k\Delta t)$ are the basis for determining the impit vessel impedance Z. Assuming linearity of the vascular system studied, the impedance Z is calculated as the ratio between the Fourier transforms of the above-mentioned discrete time courses over the work cycle of the heart:

$$Z(2\pi fn) = \frac{P(2\pi fn)}{Q(2\pi fn)} = Z_{\mathrm{mod}_n} e^{j\varphi_n}, \tag{13}$$

where $n = 0, 1, 2, \ldots, k$, f is the frequency of the heart rate, φ_n is the phase and Z_{mod_n} is the modulus. In the calculating, the algorithm of the fast Fourier transform *FFT* [3] was used. In its final form, the input vessel impedance is represented by the modulus Z_{mod_n} and the phase φ_n for successive harmonic n frequencies of the heart rate. For $n = 0$, the impedance represents the mean resistance of the vascular system in question.

Another group of the investigated parameters are related to the elasticity of the walls of arterial blood vessels. One of them is the relative change in the vessel diameter $\Delta D/D$ over in the work cycle of the heart. It is defined in the following way:

$$\frac{\Delta D}{D} = \frac{D_{\max} - D_{\min}}{D_{\min}} \cdot 100\%, \tag{14}$$

where D_{max} and D_{min} are the maximum and minimum blood vessel diameters over the vessel cycle of the heart.

The parameter described by formula (14) depends on the elasticity of the vessel wall, but it cannot be its measure, since it does not into account the increase in the blood pressure causing changes in the vessel diameter. In the analysis, it was introduced solely for cognitive purposes.

A more objective index of the elasticity of arterial vessel walls is the pulse wave velocity c_t determined from the volume elastic modulus K [1]

$$c_t = \sqrt{\frac{K}{\varrho}} = \sqrt{\frac{1}{\varrho}\frac{(P_s - P_d)}{S_s - S_d}S_d}, \tag{15}$$

where ϱ is the blood density, S_s and S_d are the cross-section areas of the blood vessel for the systole pressure P_s and the diastole pressure P_d.

The pulse wave velocity c_t is an index very generally used in the literature for evaluation of the elasticity of the human arterial-vascular system [8, 12]. This velocity depends on the rigidity of blood vessel walls and increases with a person's age. c_t is measured by the method of two sensors set usually at two mutually distant points of the vascular system. This permits only an overall evaluation of the elasticity of the vascular system.

According to some authors [12, 16], the rigidity of the walls of the vascular system, and, thus, the pulse wave velocity, too, are affected by the blood pressure in the vascular system. This fact can be explained by the existence of a nonlinear function between the vessel wall displacements and the blood pressure which causes them. So far in the literature, there has been no agreement about the degree and character of this nonlinearity [13, 14]. In this situation in the evaluation of the vessel wall elasticity an additional index α was introduced. It results from the previously adopted exponential dependence between the blood pressure and the transverse dimensions of the blood vessel (formula (10)). The coefficient α determined from dependence (10) has the following form:

$$\alpha = \frac{S_d}{S_s - S_d}\ln(P_s/P_d), \tag{16}$$

where S_s and S_d are the vessel cross-section areas for systolic P_s and diastolic P_d pressures.

The hemodynamic parameters mentioned so far were preliminarly investigated in the common carotid arteries for a group of 43 healthy persons aged between 9 and 64 years. The age of the persons examined was divided into five groups. The results of the measurements are given in Table 1. In addition, a statistical analysis of the parameters studied, performed without a division into age groups, permits the following conclusion to be drawn:

1) The mean blood flow rate Q_{med} over the cardiac cycle in the common carotid artery in adults aged between 19 and 64 varied between 400 and 620 ml/min. Its mean value was 499 ± 75 ml/min.

2) The relative change in the diameter of the common carotid artery $\Delta D/D$ over the cardiac cycle was greatest in children and decreased with the age of the patients. Investigation of the change $\Delta D/D$ as a function of age by means of linear regression gave the following dependence:

$$\Delta D/D = 15.507 - 0.189\,x \quad [\%], \tag{17}$$

Table 1. The hemodynamic parameters determined noninvasively in the common carotid arteries of healthy persons in five age groups the mean blood flow rate Q_{med}, the mean blood pressure P_{med} from formula (12), the relative change in the artery diameter $\Delta D/D$, the pulse wave velocity c_t formula (15) and the coefficient of rigidity α formula (16). The systole pressure P_s and the diastole pressure P_d were determined with a manometer in the brachial artery

Age group (years)	9–16	19–30	32–40	41–50	52–64
Number of persons	7	9	8	8	11
Mean age	12.3	24.4	36.8	44.9	56.2
S.D. (years)	2.3	4.6	3.2	2.9	3.9
P_s	104.3	117.8	116.9	110.0	122.7
S.D. (mmHg)	10.6	7.9	7.0	10.0	13.8
P_d	60.7	77.2	80.0	70.6	80.0
S.D. (mmHg)	6.1	7.1	5.9	8.6	8.7
P_{med}	82.2	97.3	97.2	90.5	101.5
S.D. (mmHg)	6.4	6.1	5.8	9.2	10.5
Q_{med}	364.3	497.8	511.3	540.0	459.0
S.D. (ml/min)	53.2	87.9	74.9	59.8	61.9
$\Delta D/D$	13.89	11.0	7.54	6.85	5.24
S.D. %	1.83	2.02	1.79	1.05	1.14
c_t	4.52	4.97	5.84	6.26	7.44
S.D. (m/s)	0.61	0.38	0.71	0.49	0.74
α	1.84	1.87	2.55	3.18	3.98
S.D.	0.43	0.37	0.68	0.51	0.62

S.D. – Standard deviation

where x is the age of the persons examined in years. The coefficient of correlation for this linear dependence was 0.877.

3) The pulse wave velocity c_t determined from formula (15) increased linearly with the age of the patients according to the regression line:

$$c_t = 3.457 + 0.0677\,x. \tag{18}$$

The coefficient of correlation for this dependence was 0.877

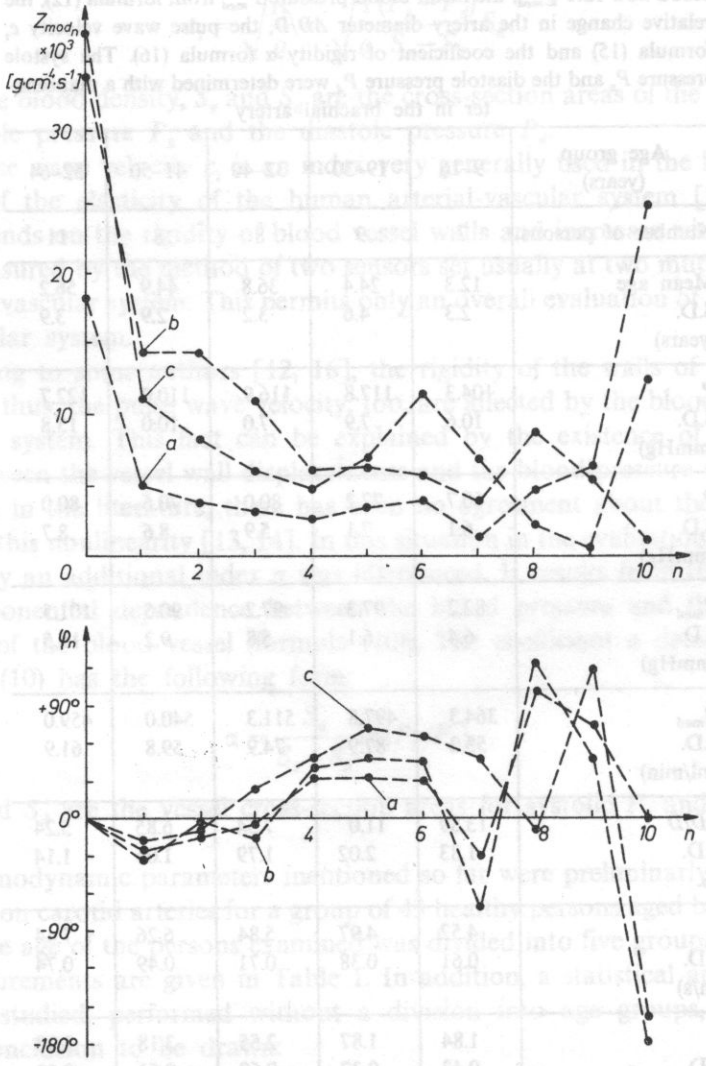


FIG. 9. The modulus (Z_{mod_n}) and the phase (φ_n) of the input vascular impedance determined noninvasively in the common carotid artery (a), in the internal carotid artery (b) and the external carotid artery (c) in a 40-year-old man

4) The coefficient of rigidity α increased linearly as a function of the age of the persons examined. It is described by the regression line:

$$\alpha = 0.8581 + 0.0523\,x. \tag{19}$$

The coefficient of correlation for this function was 0.832.

Using the measuring system described here, preliminary investigations of the input vessel impedance were also performed in the extracranial carotid arteries (Fig. 9). They indicate that for the first few harmonics the modulus and phase of the impedance strongly depend on the inertia, compliance and resistance of the vascular system studied. The first step towards the determination of these values was a computer simulation of the input vessel impedance in the common carotid artery by means of the impedance of a substitute circuit containing elements representing inertia compliance, vascular resistance and peripheral resistance [19]. Further research is under way on the interpretation of the vascular impedance [21].

The preliminary research results presented above indicate the large usefulness of the measuring system for noninvasive evaluation of the human arterial-vascular system, in particular for diagnosis of the carotid arteries.

Acknowledgement

I am very grateful to Mr Zbigniew TRAWIŃSKI, M.Sc., for great help in constructing the measuring system described in this study.

References

[1] D. H. BERGEL, *Cardiovascular fluid dynamics*, Vol. 1, Academic Press 1972.

[2] K. BORODZIŃSKI, L. FILIPCZYŃSKI, A. NOWICKI, T. POWAŁOWSKI, *Quantitative transcutaneous measurements of blood flow in carotid artery by means of pulse and continuous wave Doppler methods*, Ultrasound in Med. and Biol., **2**, 189–193 (1976)

[3] E. O. BRIGHAM, The fast Fourier transform, Prentice Hall, Inc.

[4] D. A. MC DONALD, M. G. TAYLOR, *The hydrodynamics of the arterial circulation*, Progress in Biophysics and Biophysical Chemistry, **9**, 105 (1959).

[5] D. A. MC DONALD, *Blood flow in arteries*, E. Arnold, London 1960.

[6] L. FILIPCZYŃSKI, R. HERCZYŃSKI, A. NOWICKI, T. POWAŁOWSKI, *Blood flow-hemodynamics and ultrasonic Doppler measurement methods*, PWN 1980.

[7] D. H. GROVES, T. POWAŁOWSKI, D. N. WHITE, *A digital technique for tracking moving interfaces*, Ultrasound in Med. and Biol., **8**, 2, 185–190 (1982).

[8] M. B. HISTAND, M. ANLIKER, *Influence of flow and pressure on wave propagation in the cannine aorta*, Circulation Research, **32**, 524 (1973).

[9] A. HOEKS, C. RUISSEN, P. HICK, R. RENEMAN, *Transcutaneous detection of relative changes in artery diameter*, Ultrasound in Med. and Biol., **11**, 51–59 (1985).

[10] D. HOKANSEN, D. MOZERSKY, D. SUMNER, D. STRADNESS, *A phase-locked echo tracking system for recording arterial diameter changes in vivo*, J. Appl. Physiol., **32**, 728–733 (1972).

[11] K. J. HUTCHISON, K. OBERLE, J. A. SCOTT, *A comparison of Doppler ultrasonic waveforms processed by zero crossing and spectrographic techniques in the diagnosis of peripheral arterial disease*, Angiology, **32**, 4, 277–289 (1981).

[12] T. KENNER, *Flow and pressure in the arteries*, from Biomechanics: Its Foundations and Objectives, Prentice Hall.

[13] G. J. LANGEWOUTERS, K. H. WESSELING, W. J. GOEDHARD, *The static elastic properties of 45 human thoracic and 20 abdominal aortas in vitro and the parameters of a new model*, J. Biomechanics, **17**, No 6, 425–435 (1984).

[14] P. van LOON, W. KLIP, E. BRADLEY, *Length-force and volume-pressure relationships of arteries*, Biorheology, **14**, 181–201 (1977).

[15] M. MARKIEWICZ, B. RYMER, *Certain aspects of atherosclerosis pathogenesis*, Wiadomości Lekarskie, PZWL, **36**, 21 (1983).

[16] J. MEGERMANN, J. HASSON, D. WARNOCK, G. LITALIEN, W. ABBOT, *Noninvasive measurements of nonlinear arterial elasticity*, Am. J. Physiol., **250**, H181–H188 (1986).

[17] T. POWAŁOWSKI, *An ultrasonic c.w. Doppler method of measurements of the blood flow velocity*, Archives of Acoustics, **6**, 3, 287–306 (1981).

[18] T. POWAŁOWSKI, B. PEŃSKO, Z. TRAWIŃSKI, L. FILIPCZYŃSKI, *An ultrasonic transcutaneous method of simultaneous evaluation of blood pressure, flow rate and pulse wave velocity in the carotid artery*, Abstracts of Satelite Symp. of the XIIIth World Congress of Neurol., Aachen, M–68 1985.

[19] T. POWAŁOWSKI, B. PEŃSKO, *Noninvasive ultrasonic method of pressure and flow measurement for estimation of hemodynamical properties of cerebrovascular system*, Archives of Acoustic, 3 (1985).

[20] T. POWAŁOWSKI, B. PEŃSKO, Z. TRAWIŃSKI, E. DURIASZ, *Vessel wall elasticity in the carotid artery estimated from transcutaneous ultrasonic measurements*, Abstracts of Symp. UBIOMED VII Eisenach, DDR, 134–135 1986.

[21] T. POWAŁOWSKI, A. MAŁEK, W. STASZKIEWICZ, *Noninvasive ultrasonic diagnostics of cerebrovascular system by means of input impedance method*, Abstracts of Symp. UBIOMED VII, Eisenach, DDR, 136–137 1986.

[22] R. S. RENEMAN, A. G. HOEKS, *Doppler ultrasound in the diagnosis of cerebrovascular disease*, Research Studies Press, 1982.

Received on December 21, 1986.

ARCHIVES OF ACOUSTICS
13, 1–2, 109–126 (1988)

A NONINVASIVE ULTRASONIC METHOD FOR THE ELASTICITY EVALUATION OF THE CAROTID ARTERIES AND ITS APPLICATION IN THE DIAGNOSIS OF THE CEREBRO-VASCULAR SYSTEM

TADEUSZ POWAŁOWSKI, BOGUMIŁ PEŃSKO

Ultrasonics Department, Institute of Fundamental Technological Research, Polish Academy of Sciences.
(00–049 Warsaw, ul. Świętokrzyska 21)

This study presents a comparative evaluation of the methods applied so far to describe the elasticity of the blood vessel walls. It was shown that one of the possible solutions in the description of the dependence between the vessel cross-section and the blood pressure is a logarithmic function. The authors assumed the logarithmic wall rigidity coefficient α resulting from this function as the index of the mechanical properties of the walls of the carotid arteries examined. The value of this coefficient was determined noninvasively from the ultrasonic measurement of the instantaneous diameter of the common carotid artery and the systolic and diastolic pressures measured with a manometer in the brachial artery. The studies were carried out for a group of 43 healthy persons aged between 9 and 64, and for a group of 9 persons aged between 53 and 62 in whom arteriosclerotic changes were found by the X-ray arteriography in the extracranial carotid arteries. The results obtained indicate a linear increase in the coefficient α with age of healthy persons. For the ill group the mean value of the coefficient α was about 50% higher than that for the healthy in the same age group.

W pracy przeprowadzona została ocena porównawcza stosowanych dotychczas metod opisu elastyczności ścianek naczyń krwionośnych. Wykazano, że jednym z możliwych rozwiązań w opisie zależności między przekrojem naczynia i ciśnieniem krwi jest funkcja logarytmiczna. Wynikający z niej logarytmiczny współczynnik sztywności ścianki α przyjęty został przez autorów jako wskaźnik własności mechanicznych ścianek badanych tętnic szyjnych. Wartość tego współczynnika wyznaczano nieinwazyjnie na podstawie ultradźwiękowego pomiaru chwilowej średnicy tętnicy szyjnej wspólnej oraz ciśnień: skurczowego i rozkurczowego mierzonych manometrem w tętnicy ramiennej. Badania przeprowadzono dla grupy 43 osób zdrowych w wieku od 9 do 64 lat oraz dla grupy 9 osób w wieku od 53 do 62 lat, u których stwierdzono metodą arteriografii rentgenowskiej zmiany miażdżycowe w tętnicach szyjnych pozaczaszkowych. Uzyskane wyniki wskazują na liniowy wzrost współczynnika α z wiekiem badanych osób zdrowych. Dla grupy osób chorych średnia wartość współczynnika α była o około 50% większa niż u osób zdrowych w tej samej grupie wiekowej.

1. Introduction

Studies on the blood vessel elasticity are very significant in the diagnosis of human vascular systems. Changes which occur in the vessel walls as a result of the organism's aging or diseases of the vessel walls (arteriosclerosis) cause an increase in their rigidity. This has a negative effect on the blood circulation mechanism in which the vessel elasticity plays a very important role. An increase in the rigidity of the arterial vessel walls is also a factor which contributes to the development of arteriosclerosis, the most serious disease of human vascular system.

Depending on the measurement methods applied, various indices are used to evaluate the vessel wall elasticity. Commonly they are: the pulse wave velocity [1, 5, 13, 28], the relative vessel diameter change [25], the elastic (pressure-strain) modulus E_p [1, 8, 19, 20, 21, 26] BERGEL's incremental modulus E_{inc} [2, 3, 4, 12, 18, 32]. All the above mentioned indices are functions of blood pressure [2, 5, 13, 16, 32]. The effect of blood pressure on the vessel wall rigidity makes difficult the comparative evaluation of the mechanical properties of the walls in persons with different blood pressure.

To take into account the effect of the blood pressure on the vessel wall rigidity, it is necessary to know the functional dependence between the cross-section of the vessel and the blood pressure inside it. This dependence was studied by many authors [12, 14, 15, 17, 30], however, the functions proposed by them cannot be used in noninvasive vessel elasticity measurements. This is due to the fact that their description requires coefficients whose values can be determined only by invasive methods.

This study carried out a comparative evaluation of the methods applied so far to describe the blood vessel wall elasticity. By analysing the dependence between the vessel cross-section area and the blood pressure, it was shown that one of the possible solutions is a logarithmic function. Its determination requires two reference points which can be found by the noninvasive method. This function was applied by the authors in the noninvasive measurement of the input vessel impedance [23, 24], to determine the blood pressure from the instantaneous values of the diameter of the common carotid artery.

Taking the logarithmic dependence between the artery cross-section and the blood pressure as the basis of analysis of the wall elasticity in the common carotid arteries, a new elasticity index was defined. It is a logarithmic wall rigidity coefficient α. This coefficient was applied by the authors in the comparative evaluation of the elastic properties of the walls of the common carotid arteries examined. Its determination requires knowledge of the artery diameter and the blood pressure in systole and diastole. These quantities were measured noninvasively. The instantaneous diameter of the common carotid artery was measured with a pulsed ultrasonic tracing system, constructed by the authors, which was connected "on line" with a computer. The systolic and diastolic blood pressures were measured with a manometer in the brachial artery. The examinations were performed on healthy

persons of different age and on persons with pathological changes in the extracranial carotid arteries.

Apart from the coefficient α, for comparison, the relative diameter change $\Delta d/d$ and the elastic modulus E_p were determined.

2. Elasticity of the blood vessel

It is generally assumed that human blood vessels are elastic. This means that each change in the blood pressure inside the vessel is accompanied by a change in its dimensions, depending on the mechanical properties of the blood vessel walls. In a living organism blood vessels are strongly longitudinally extended (about 1.5 times with respect to their length outside the organism) and fixed to the surrounding tissue. This excludes almost completely the possibility of wall motion along the vessel axis (longitudinal motion). Therefore, it is generally assumed [5, 32, 1], that the only mechanical reaction of the vessel to a change in the blood pressure is a change in its transverse dimensions. Hence, there result, variously defined, elasticity coefficients often applied in the literature to characterise the properties of the blood vessel walls.

In 1960 PETERSON [21] introduced the elastic modulus E_p defined as:

$$E_p(\text{acc. to Peterson}) = \Delta p/(\Delta d/d), \tag{1}$$

where Δp is a pressure change causing a change in the external diameter d of the vessel by Δd.

Analogously to formula (1), the coefficient E_p is also used for the internal diameter of the vessel, according to the dependence [26]:

$$E_p = \frac{(p_s - p_d)\, d_d}{(d_s - d_d)}, \tag{2}$$

where d_s and d_d denote the internal vessel diameters for the systolic p_s and diastolic p_d pressures, respectively.

In the 1961 BERGEL introduced [2] an incremental elastic modulus in the form

$$E_{\text{inc}} = 2(1-\sigma^2)\frac{R_i^2 R_e \Delta p}{(R_e^2 - R_i^2)\Delta R_e}, \tag{3}$$

where R_i and R_e are the internal and external radii of the blood vessel, respectively.

Table 1 lists the values of the moduli E_p and E_{inc} published by different authors. The quite large discrepancy of the results presented in the table results mainly from the different conditions of the experiments, (in vitro, in vivo – with exposed artery, in vivo – studied from the surface of the body) and very different techniques of the measurements of the blood pressure. The results cited were obtained for pressure close to 100 mmHg. In two cases they apply to a wider pressure range.

Most of experimental studies carried out for years in the world indicate that the

Table 1. Elasticity moduli E_p and E_{inc} determined in common carotid artery by different authors

Author	Conditions of experiment	Subject	Blood pressure [mmHg]	Age (years)	E_p	E_{inc}
					10^6 dyn/cm^2	
PETERSON 1960 [21]	in vivo	dog	—	—	2.3–4.2*	—
BERGEL 1961 [2]	in vitro	dog	40–220	—	—	1–12.2
			100	—	—	6.4
PATEL 1963 [19]	in vivo	dog	< 120	—	2.88*	—
PATEL 1964 [20]	in vivo	man		28–69 mean value		
			100	45	6.0*	—
GOW 1968 [8]	in vivo	dog	~ 100	—	2.1*	—
ARNDT 1968 [1]	in vivo*	man		24–34 mean value	0.32–0.58*	—
			~ 100	28	0.46**	—
HAYASHI 1980 [12]	in vitro	man	100	< 40	—	4.5
			100	40–50	—	3.0
NEWMAN 1982 [18]	in vitro	rabbit	100	—	—	7.0
WEIZSACKER 1982 [32]	in vitro	rat	45–112	—	—	2.5–25
			100	—	—	15
RILEY 1984 [26]	in vivo(*)	man		7–25 mean value	0.32–1.61**	—
			~ 80	16	0.64**	—
BOROVETZ 1986 [4]	in vitro	dog	~ 100	—	—	5.0

* acc. to formula (1), ** acc. to formula (2)

(*) ultrasonic measurements from the body surface

reaction of the walls to a change in the blood pressure is nonlinear, or even strongly nonlinear. This means that the elastic indices given by formulae (1), (2), (3) are function of pressure – this greatly reduces their usefulness in evaluating the state of blood vessels.

So far there is no unambiguous agreement as to the analytical form of the nonlinear function connecting the blood pressure with a change in the transverse dimensions of the blood vessel. Most functions proposed in the literature are purely empirical, as they are based on the authors' own experimental data. These functions are greatly different from one another and often have a very complex mathematical form. Table 2 lists formulae describing recent attempts to represent analytically the relations between the dimensions of the blood vessel and the blood pressure. The table also includes the exponential relation, proposed in 1985 by the present authors [23, 24] between the blood pressure and the squared vessel radius, and the equivalent logarithmic function between the vessel cross-section area and the blood pressure. The exponential form was applied for noninvasive determination of the course of the blood pressure in the carotid arteries from ultrasonic measurements of the instantaneous diameters of these arteries. With reference to cross-section area S of a cylindrical vessel in which the blood pressure p is greater than zero, this function can be represented in the following form

$$p = p_0 \exp(\gamma S), \tag{4}$$

where p_0 and γ are constant coefficients.

Earlier suggestions regarding the exponential character of the stress – strain dependeces in blood vessel can be found e.g. in studies by FUNG [6], SIMON [29], TANAKA [31] or GHISTA [7].

Formula (4) can be justified by using GREEN's theory [9, 10] which describes the behaviour of bodies under elastic strains with large amplitudes. The mathematical model of the blood vessel assumed an axially symmetric cylinder in the state of plain strain with large (finite) amplitudes. The cylinder is built of homogenous, elastic and incompressible material. Assuming moreover that the deformation of such a cylinder can be treated as constant axial elongation and homogenous transverse inflation causing a change in the internal and external vessel radii from their initial values a_2 and a_1 to the current values r_2 and r_1, it can be shown [29] that the pressure p inside the vessel is given by the formula

$$p = 2A \int_{r_1}^{r_2} e^{kI} [Q^2/\lambda^2 - 1/Q^2] \, dr/r, \tag{5}$$

where I is the first invariant of the strain tensor given by the expression

$$I = \lambda^2 + Q^2/\lambda^2 + 1/Q^2.$$

λ is the ratio between the deformed and undeformed axial coordinates of the cylinder and Q is the ratio between the undeformed and deformed radial coordinates of the cylinder.

Table 2. Nonlinear functions between vessel dimensions and blood pressure, proposed in the recent literature

LOON et al. [15]: V – volume, p – pressure

$V = V_0 + (V_m - V_0)(1 - e^{-\alpha p})$

V_0 by $p = 0$ mmHg, V_m – max. valve

HAYASHI et al. [12]: R – radius, p – pressure

$$R/R_0 - 1 = \frac{1}{\beta}\ln(p/p_0)$$

R_0 for $p_0 = 100$ mmHg

STETTLER et al. [30]: S – cross-section, p – pressure, z – distance from the heart

$$S(p, z) = S_0(z)\exp\left[\frac{p - p_0}{c(p_0, z)\,c(p, z)\,\varrho}\right]$$

$c(p, z) = (c_0 + Bp)\,g(z)$

$p_0 = 100$ mmHg

MEISTER [17]: S – cross-section, p – pressure, z – distance from the heart

$$S(p, z) = S(p_0, z)\exp\frac{1}{\varrho g^2}\left[\frac{2a_3 p + a_2}{\Delta c(p)} + \frac{4a_3}{\Delta^{3/2}}\operatorname{arc\,tg}\frac{2a_3 p + a_3}{\Delta^{1/2}}\right]$$

$c(p) = a_1 + a_2 p + a_3 p^2$

$g(z) = c_1 + c_2 z, \quad \Delta = 4a_1 a_3 - a_2^2, \quad p_0 = 100$ mmHg

LANGEWOUTERS et al. [14]: S – cross-section, p – pressure

$$S(p) = S_m\left(\frac{1}{2} + \frac{1}{\pi}\operatorname{arc\,tg}\frac{p - p_0}{p_1}\right)$$

p_0 determined from the half-width value of S_m

p_1 determined from the half-width value of vessel compliance

POWAŁOWSKI et al. [24]: p – pressure, R – radius

$$p(R) = p_d \exp\left\{\frac{R^2 - R_d^2}{R_s^2 - R_d^2}\ln\frac{p_s}{p_d}\right\}$$

S – cross-section, p – pressure

$$S(p) = S_d\left[1 + \left(\ln\frac{p}{p_d}\right)\Big/\left(\frac{S_d}{S_s - S_d}\ln\frac{p_s}{p_d}\right)\right]$$

S_d, R_d – for diastolic pressure p_d; S_s, R_s – for systolic pressure p_s

In the formula (5) use was made of the exponential form of the constitutive equation, which, according to SIMON [29], can be written in the form

$$\partial W/\partial I = A\, e^{kI}, \tag{6}$$

where A and k are parameters to be determined from the boundary conditions and W is the density function of the strain energy introduced by GREEN [10].

Expression (5) can be reduced to a form more convenient for numerical calculations:

$$p = -C \int_{x_1}^{x_2} e^{\beta(x+1/x)}\,[1+1/x]\,dx, \tag{7}$$

where

$$x = Q^2/\lambda = 1+\text{const}/r^2, \tag{7a}$$

$$\beta = k/\lambda, \tag{7b}$$

$$C = (A/\lambda)\, e^{-2\beta}, \tag{7c}$$

The incompressibility of the material of the blood vessel walls can be expressed by the equation

$$a_1^2 - a_2^2 = \lambda(r_1^2 - r_2^2), \tag{8}$$

where a_1 and a_2 are the external and internal radii of the vessel for $p = 0$, r_1 and r_2 are its external and internal radii for the pressure p, and λ is the relative elongation which is constant as a function of pressure (according to SIMON [29] $\lambda = 1.532$).

The dependence of the variable x in formula (7) on the squared radius only (formula (7a)) suggests that blood pressure is a function of the cross-section area of the blood vessel. It follows from the incompressibility condition (8) that the function $p = f(S)$ has the same form for the external and internal cross-section of the vessel.

The course of the function p (S) calculated numerically from formula (7) is shown in Fig. 1. The marked experimental points come from a study by SIMON et al. [29], who measured the external radius of the vessel r_1 as a function of pressure for a canine abdominal aorta in vitro. The reference points for the calculations were $S_1 = 60.9$ mm^2, $p_1 = 46.2$ mm Hg, $S_2 = 84{,}9$ mm^2, $p_2 = 199$ mm Hg. The exponential function described by formula (4) was plotted through the same two points. The course of the exponential function distinctly agreed with numerical calculations, and the maximum difference between the values of p calculated from formulae (4) and (7) did not exceed 0.7%.

To show if the exponential function (4) approximates the real experimental data well, the coefficient of determination of the curve R^2 was calculated for data in Fig. 1, using the definition [27]

$$R^2 = 1 - \frac{\sum_i (\tilde{y}_i - y_i)^2}{\sum_i (\bar{y} - y_i)^2}, \tag{9}$$

where y_i is the value of y measured experimentally for the independent variable x equal to x_i, $\bar{y}$ is the mean of N measurements, and $y_i = f(x_i)$ if $f(x)$ denotes the approximating function applied. The coefficient R^2 for the exponential curve in Fig. 1 was 0.9868.

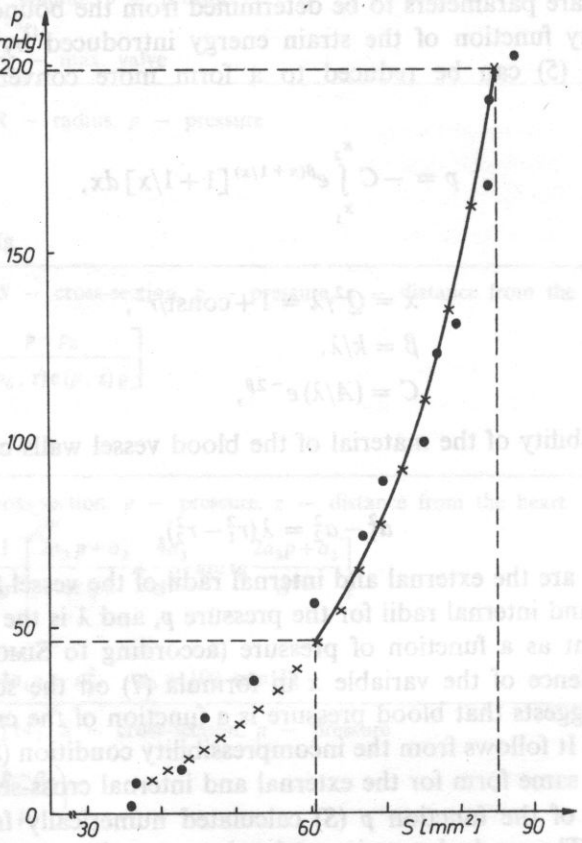


FIG. 1. The dependence between the blood pressure p and the vessel cross-section S: (o) – experimental points according to SIMON et al. [29] for a canine abdominal aorta, (x) – the results of numerical calculations (formula (7)), (solid line) – the exponential dependence from formula (4)

The exponential dependence (formula (4)) studied so far here means that there is a logarithmic function between the vessel cross-section S and the blood pressure p, in the form

$$S = (1/\gamma)\ln(p/p_0). \tag{10}$$

On the basis of the results of experimental studies presented by LOON et al [15] for human common carotid artery and by SIMON et al [29] for a canine abdominal aorta, comparative analysis was carried out of the nonlinear functions $S = f(p)$ described in the literature, including the logarithmic function proposed here. For comparison, a linear dependence between the vessel cross-section and pressure was

also assumed. The results of the comparative analysis are shown in Fig. 2 and 3 and in Table 3. Using the least squares method it was shown that, apart from the linear one, all the functions studied describe very well the results of experimental studies in physiological pressure range from 25 mmHg to 200 mmHg.

There are only slight differences between the courses of particular nonlinear functions. Accordingly, it seems fully justified to assume the function $S = f(p)$ in the logarithmic form proposed here.

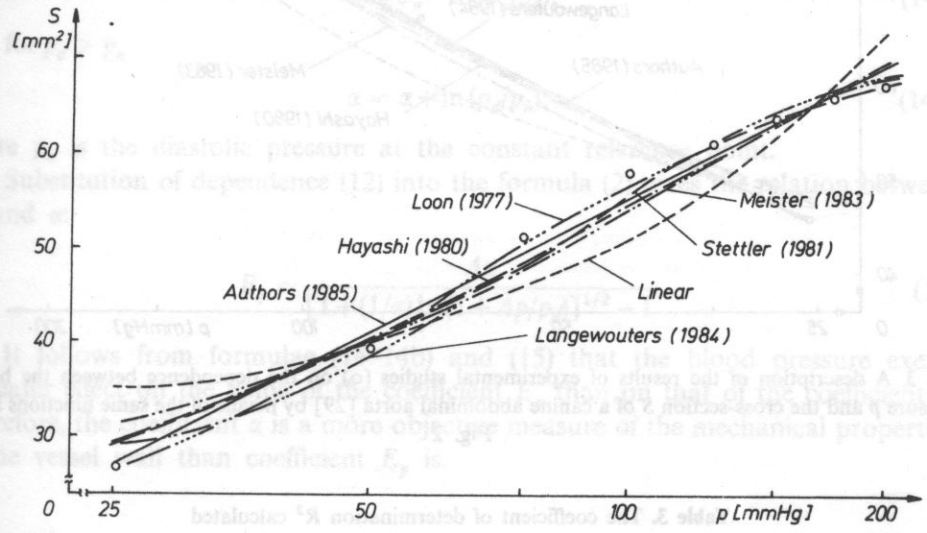


FIG. 2. A description of the results of experimental studies (o) on the dependence between the blood pressure p and the cross-section S in human common carotid artery [15] by means of the functions proposed in the literature (Table 2)

As only one of studied nonlinear functions $S = f(p)$, the logarithimc function can be determined on the basis of any two measurement points. The same applied to the equivalent exponential dependence between the blood pressure and the blood vessel cross-section. Substitution in formulae (4) and (10) the two pairs of values: p_d, S_d and p_s, S_s, corresponding successively to these points, gives the following expressions

$$p = p_d \exp[\alpha(S/S_d - 1)], \tag{11}$$

and

$$S = S_d[1 + (1/\alpha)\ln(p/p_d)]. \tag{12}$$

The coefficient α which occurs in formulae (11) and (12) is defined in the following way

$$\alpha = \frac{S_d \ln(p_s/p_d)}{S_s - S_d} = \frac{S_d \ln(1 + \Delta p/p_d)}{S_s - S_d}, \tag{13}$$

where $\Delta p = p_s - p_d$.

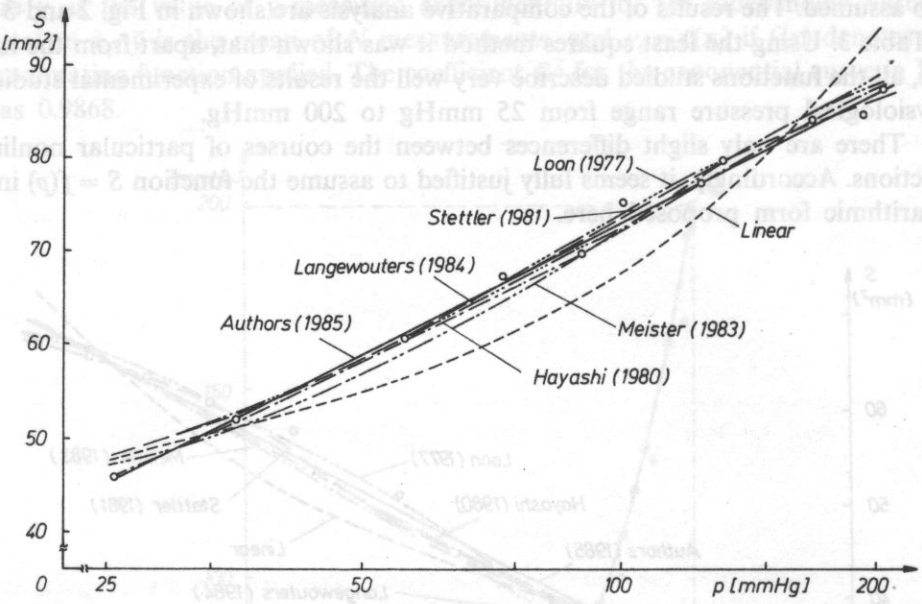


FIG. 3. A description of the results of experimental studies (o) on the dependence between the blood pressure p and the cross-section S of a canine abdominal aorta [29] by means of the same functions as in Fig. 2

Table 3. The coefficient of determination R^2 calculated for experimental data approximated by different non-linear functions

Functions acc. to:	Common carotid man (LOON, 1977)	Abdominal aorta dog (SIMON, 1971)
LOON* [15]	0.9974	0.9950
HAYASHI* [12]	0.9722	0.9873
STETTLER [30]	0.9613	0.9917
MEISTER [17]	0.9635	0.9894
LANGEWOUTERS [14]	0.9862	0.9945
Authors [23]	0.9868	0.9942
Linear	0.8653	0.9185

*) after transformation to the form $S = f(p)$

In noninvasive studies, the values p_d and S_d represent the pressure and the vessel diameter in the diastole, and p_s and S_s are the corresponding quantities in the systole (Table 2) [24].

The coefficient α can be called a logarithmic rigidity coefficient of the blood vessel wall. It follows from the formulae (10) and (13) that the value of the coefficient α depends on the choice of the reference point (p_d, S_d). This means that it is a function

of the diastolic pressure p_d. The effect of the pressure p_d on the value of the coefficient α can be described in the following way:

for $p_d = p_n$

$$\alpha = \alpha_n, \tag{14}$$

for $p_d < p_n$

$$\alpha = \alpha_n - \ln(p_n/p_d), \tag{14a}$$

for $p_d > p_n$

$$\alpha = \alpha + \ln(p_d/p_n), \tag{14b}$$

where p_n is the diastolic pressure at the constant reference point.

Substitution of dependence (12) into the formula (2) gives the relation between E_p and α:

$$E_p = \frac{\Delta p}{[1+(1/\alpha)\ln(1+\Delta p/p_d)]^{1/2}-1}. \tag{15}$$

It follows from formulae (14–14b) and (15) that the blood pressure exerts a greater effect on the value of the coefficient E_p than on that of the coefficient α. Therefore, the coefficient α is a more objective measure of the mechanical properties of the vessel wall than coefficient E_p is.

3. Results and discussion

The elasticity measurements were carried out in human common carotid artery using ultrasonic equipment constructed by the authors [22]. It permits the simultaneous transcutaneous measurement of the instantaneous blood velocity and the instantaneous internal diameter in the blood vessel examined. The vessel diameter was determined by the echo method. Its instantaneous value was measured using a digital system tracking echos from both walls of the blood vessel [11, 22].

The accuracy of the measurements of the vessel wall displacement was 0.03 mm. The data were registered and analysed on a PDP-11 computer connected "on line" with the ultrasonic device.

The studies were carried out on two groups of persons. The first group (43 persons) did not show any pathological changes in the region of the arteries examined. The second group (9 persons – 12 arteries) included patients with arteriosclerosis changes identified by X-ray angiography in the region of the carotid arteries. In all the persons examined, the maximum and minimum values of the vessel diameters d_s and d_d were measured, and then subordinated respective to the values of blood pressure is systole and diastole. The values of systolic and diastolic pressure p_s and p_d were measured with a manometer in the brachial artery, at the height of the

neck, with the patient in supine position. The diameter of the common carotid artery was measured transcutaneously from a point located about 2–3 cm away from bifurcation into the internal and external carotid arteries.

The measurements results were grouped depending on the age of the persons examined (Tables 4 and 5). This age varied in the healthy group from 9 to 64 years, and in the pathological group, from 53 to 62 years. For each patient, from measurements of the diameter and pressure, the following quantities were determined: the relative diameter change $\Delta d/d = (d_s - d_d)/d_d$; the elasticity coefficient E_p (formula (2)); and the logarithmic rigidity coefficient α (formula (13)).

Fig. 4 shows the values of the coefficient α obtained as a function of the age of the persons examined. Circles represent healthy arteries, and crosses mark those with

Table 4. The logarithmic function parameters p_0, γ, α (formulae (16) and (18)) determined on the basis of two measured points (for diastolic p_d and systolic p_s pressures) for healthy and ill persons

Age group (years)	Healthy persons					Ill persons
	9–16	19–30	32–40	41–50	52–64	53–62
Meam Age S.D. (years)	12.3 2.3	24.4 4.6	36.8 3.2	44.9 2.9	56.2 3.9	57.2 3.0
Number of persons	7	9	8	8	11	9
p_d S.D. (mmHg)	60.7 6.1	77.2 7.1	80.0 5.9	70.6 8.6	80.0 8.7	83.1 4.6
p_s S.D. (mmHg)	104.3 10.6	117.8 7.9	116.9 7.0	110.0 10.0	122.7 13.8	150.0 15.4
p_0 S.D. (mmHg)	10.53 4.32	12.87 5.49	7.64 4.79	3.30 1.63	1.72 0.90	0.87 0.86
γ S.D. (cm^{-2})	7.65 2.10	5.09 1.12	6.28 2.28	8.22 2.57	10.10 2.55	12.15 4.40
α S.D.	1.84 0.43	1.87 0.37	2.55 0.68	3.18 0.51	3.98 0.62	6.16 2.29

S. D. – standard deviation

pathological changes. The rigidity coefficient increased linearly with the patient's age in the healthy group, according to the simple regression formula:

$$\alpha = 0.858 + 0.0523\,x, \tag{16}$$

where x is the age in years.

Table. 5. Elasticity modulus E_p (formula 2) for common carotid arteries for healthy persons in different age groups

Age group (years)	9–16	19–30	32–40	41–50	52–64
Modulus E_p	0.423	0.499	0.681	0.776	1.091
S.D. (10^6 dyn/cm^2)	0.118	0.075	0.158	0.118	0.207

S.D. – standard deviation

The coefficient of the linear correlation was 0.832. The values of the coefficient α varied from (1.84 ± 0.43) on average in the age group 9–16 years to (3.88 ± 0.62) on average in the group 52–64 years. The values of the rigidity coefficient α in the pathological group were significantly different ($p < 0.01$) from those in the healthy group (older than 50), namely 6.16 ± 2.29 on average.

The mean values of diastolic pressures in particular age group of adults were similar. The same applied to the healthy and ill person in the same age group (Table 4).

It follows from the formulae (14)-(14b) and the values of p_d in Table 4 that the differences in the diastolic pressure for particular age groups influence only slightly the value of the coefficient α (not more than 15%). This effect was practically negligible ($< 1\%$) for the compared groups of healthy and ill persons. The large difference in the coefficient α between these groups can result only from structural changes in the walls of the arteries examined.

Fig. 5 shows the results of measurements of the relative diameter change in the common carotid artery as a function of the patient's age. The ratio $\Delta d/d$ decreases with increasing age of the persons examined, from the mean value 13.9% for the are group 9–16 years to the mean value of 5.32% for the group 52–64 years. The functional dependence of ratio $\Delta d/d$ on the age can be determined from the previously discussed logarithmic rigidity coefficient α. By assuming the linear dependence of the coefficient α on age: $ax + b$ (see formula (16)) and substituting it in formula (13), the following expression is obtained:

$$\frac{\Delta d}{d} = \frac{\ln(p_s/p_d)}{2\alpha} = \frac{\ln(p_s/p_d)}{2(ax+b)}, \tag{17}$$

where x is age in years.

Evaluation of the pressure ratio p_s/p_d and coefficient α as a function of age for

the healthy persons (formula (16) and Table 4) indicates that the relative diameter change in the carotid artery is mainly affected by age. Therefore in the first approximation, the following regression function can be assumed for this group of persons:

$$\Delta d/d = 1/(3.8885 + 0.2370\,x), \tag{18}$$

where x is age in years.

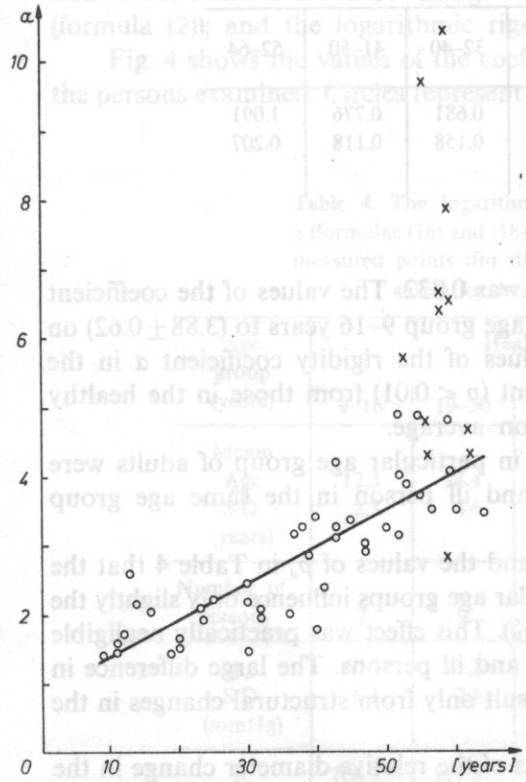


FIG. 4. The logarithmic rigidity coefficient α formula (13) determined for the common carotid arteries of healthy (o) and ill (x) persons of different age. The solid line represents the regression function describing the course of the values of the coefficient α for healthy persons, formula (16)

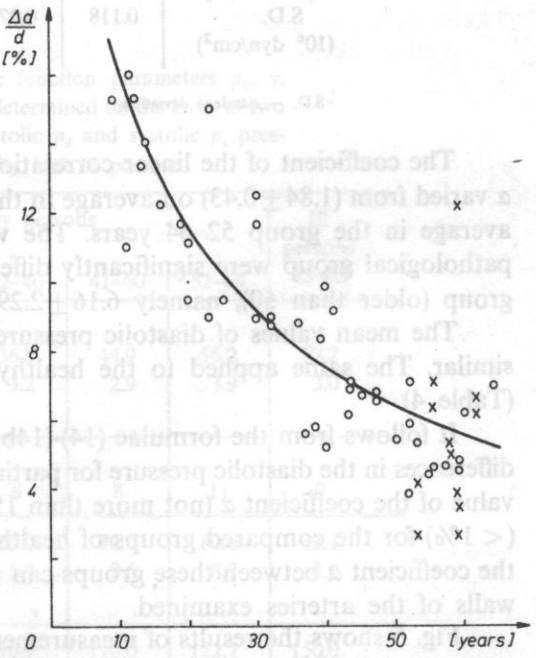


FIG. 5. The relative change $\Delta d/d$ in the diameter of the common carotid artery as function of the age of healthy (o) and ill (x) persons examined. The solid line represents the regression function describing the distribution of experimental points for healthy persons, formula (18)

The determination coefficient R^2 found for this function was 0.7713.

The values of the ratio $\Delta d/d$ obtained from the present measurements are very close to those published by RENEMAN et al. [25]. However, the value of $\Delta d/d$ cannot be recognized as the index of changes occurring in the vessel wall, in view of the dependence between the relative change in the blood vessel diameter and the blood pressure in the systolic and diastolic phases. An example is the lack of significant

difference between the relative in the diameter change of the common carotid artery between the ill and the healthy in the same age group from 52 to 64 years (see Fig. 5). The mean value of the ratio $\Delta d/d$ was respectively 5.48% for ill persons and 5.32% for healthy persons. The values of the pressure difference between the systolic and diastolic phases were different between the two groups. For the healthy group the mean systolic pressure p_s was (122.7 ± 13.8) mmHg and the mean diastolic pressure p_d was (80.0 ± 8.7) mmHg. For ill persons these pressures were respectively $p_s = (150 \pm 15.4)$ mmHg and $p_d = (83.1 \pm 4.6)$ mmHg (Table 4).

In the available literature, the blood vessel elasticity is most often characterized by the elasticity coefficient E_p (formula (1) or (2)) or by E_{inc} (formula(3)).

The values of the coefficient E_p (formula (2)) calculated from the present authors' measurements for 5 age groups are shown in Table 5. Comparison of them with Table 1 shows that the present results are close to those obtained by ARNDT et al. [1] and RILEY et al. [26], who applied a measurement method resembling the one used here.

It follows from Table 5 that E_p increases as a function of the patient's age, just as the coefficient α does (Fig. 4, Table 4). The mutual relation between the values of these two coefficients depends on the diastolic pressure (formula (15)).

For the group of healthy persons with different diastolic pressure, the coefficient α correlated better with age than the coefficient E_p did. The correlation coefficient R determined for the linear dependence of E_p and α on age was for the examined group 0.809 for E_p and 0.8193 for α.

4. Conclusions

1. The evaluation of the elasticity of blood vessel walls requires the description of the functional dependence between the vessel cross-section S and the blood pressure p. On the basis of the results of experimental research presented by LOON et al. [15] for human common carotid artery and by SIMON et al. [29] for canine abdominal aorta, comparative analysis was carried out of the function $S = f(p)$ described in the literature, including a logarithmic function proposed by the present authors. Using the least squares method, it was shown that, apart from the linear one, all the functions examined describe very well ($R^2 > 0.96$) the results of experimental research, over the pressure range from 25 mmHg to 200 mmHg.

2. As the only one of the nonlinear functions studied, the logarithmic function between the vessel cross-section and the blood pressure, proposed by the present authors can be determined by the noninvasive method. Its determination requires two pairs of values: pressure – cross-section, which can be found from ultrasonic measurements of the vessel diameter for the systolic and diastolic pressures, which pressures are determined noninvasively with a cuff.

3. The evaluation of the elasticity of blood vessel walls involved the application of the logarithmic rigidity coefficient α (formula (13)) proposed by the present

authors. For the logarithmic function between the vessel cross-section and the blood pressure, it is a more objective index of the mechanical properties of the vessel wall than the previously used coefficient E_p (formula (2)).

4. The noninvasive studies carried out in the common carotid artery in healthy persons aged between 9 and 64 years, indicate an increase in the rigidity of the artery wall as a function of age. The value of the coefficient α increased linearly with age (formula (22)). The coefficient of linear correlation was equal to 0.832.

5. For ill persons with arteriosclerotic changes in the extracranial carotid arteries, the value of the coefficient α was significantly different ($p < 0.01$) than that for healthy persons in the same age group (from 52 to 64 years). The results obtained for healthy and ill persons confirm the view that a factor which favours the development of sclerosis is an increase in the vessel wall rigidity. The mean value of the coefficient α for ill persons was 6.16 ± 2.29, whereas for healthy persons it was 3.98 ± 0.62.

6. The measurements of the relative diameter change $\Delta d/d$ carried out for healthy persons also indicate the dependence of this quantity on age (Fig. 5, formula (24)).

7. Between healthy and ill persons in the same age group there was no significant difference in the relative diameter change of common carotid artery. The mean value of the systolic pressure in ill persons was greater by about 27 mmHg than that for healthy persons. In both groups the diastolic pressures had similar values.

8. The preliminary results of clinical studies indicate that the method used for evaluating the elasticity of blood vessel walls can be highly significant in the diagnosis of extracranial carotid arteries, in particular in examinations of arteriosclerosis.

References

[1] J. O. ARNDT, J. KLAUSKE, F. MERSCH, *The diameter of the intact carotid artery in man and its change with pulse pressure*, Pflugers Arch. ges. Physiol., **301**, 230–240 (1968).

[2] D. H. BERGEL, *The static elastic properties of the arterial wall*, J. Physiol., **156**, 445–457 (1961).

[3] D. H. BERGEL, *The dynamic elastic properties of the arterial wall*, J. Physiol., **156**, 458–469 (1961).

[4] H. S. BOROVETZ, A. M. BRANT, T. K. HUNG, *Arterial wall biomechanics*, Proceedings of Fifth International Congress on Mechanics in Medicine and Biology, Bologna 1986, pp. 67–70.

[5] F. J. CALLAGHAN, L. A. GEDDES, C. F. BABBS, J. D. BOURLAND, *Relationship between pulse-wave velocity and arterial elasticity*, Med. and Biol. Eng. and Comput., **24**, 248–254 (1986).

[6] Y. C. B. FUNG, *Stress-strain history relations of soft tissues in simple elongation*, in: Biomechanics, Its Foundations and Objectives, Prentice Hall, 1972, 181–208.

[7] D. N. GHISTA, G. JAYARAMAN, H. SANDLER, *Analysis for the non-invasive determination of arterial properties and for the transcutaneous continuous monitoring of arterial blood pressure*, Med. and Biol. Eng. and Comput., **16**, 715–726 (1978).

[8] B. S. GOW M. G. TAYLOR, *Measurements of viscoelastic properties of arteries in the living dog*, Circulation Research, **23**, 1, 111 (1968).

[9] A. E. GREEN, W. ZERNA, *Theoretical elasticity*, Clarendon Press, Oxford 1963.

[10] A. E. GREEN, J. E. ADKINS, *Large elastic deformations*, Clarendon Press, Oxford 1970.

[11] D. GROVES, T. POWAŁOWSKI, D. N. WHITE, *A digital technique for tracking moving interfaces*, Ultrasound in Med. and Biol., **8**, 2, 185–190 (1982).

[12] K. HAYASHI, H. HANDA, S. NAGASAWA, A. OKUMURA, K. MORITAKE, *Stiffness and elastic behaviour of human intracranial and extracranial arteries*, J. Biomechanics, **13**, 2, 175–184 (1980).

[13] M. B. HISTAND, M. ANLIKER, *Influence of flow and pressure on wave propagation in the canine aorta*, Circulation Research, **32**, (1973).

[14] G. J. LANGEWOURTERS, K. WESSELING, W. GOEDHARD, *The static elastic properties of 45 human thoracic and 20 abdominal aortas in vitro and the parameters of a new model*, J. of Biomechanics, **17**, 6, 425–435 (1984).

[15] P. LOON, W. KLIP, E. BRADLEY, *Length-force and volume-pressure relationships of arteries*, Biorheology, 14, 181–201 (1977).

[16] J. MEGERMAN, J. HASSON, D. VARNOCK, G. LITALIEN, W. ABBOT, *Noninvasive measurements of non linear arterial elasticity*, Am. J. Physiol., 250, 181–188 (1986).

[17] J. J. MEISTER, *Mesure par échographie Doppler et modèlisation théorique de l'effet de troubles cardiaques sur la pression et le débit artériels*, Thèsis, EPFL, Lausanne 1983.

[18] D. NEWMAN, J. GREENWALD, *The effect of smooth muscle activity on the static and dynamic elastic properties of the rabbit carotid artery*, in: Cardiovascular System Dynamics, Plenum Press, NY-London 1982.

[19] D. J. PATEL, F. M. FREITAS, J. C. GREENFIELD, D. L. FREY, *Relationship of radius to pressure along the aorta in living dogs*, J. Appl. Physiol. **18**, 6, 1111–1117 (1963).

[20] D. J. PATEL, J. C. GREENFIELD, D. L. FRY, *In vivo pressure-length-radius relationship of certain blood vessels in man and dog*, Pulsatile Blood Flow Proceedings, Editor E. O. Attinger, New York 1964.

[21] L. PETERSON, R. JENSEN, J. PARNEL, *Mechanical properties of arteries in vivo*, Circulation Res., **8**, 622–639 (1960).

[22] T. POWAŁOWSKI, *An ultrasonic apparatus for noninvasive measurement of hemodynamical parameters of human arterial system*, Archives of Acoustics, **13**, 1–2 (1988).

[23] T. POWAŁOWSKI, B. PENSKO, Z. TRAWIŃSKI, L. FILIPCZYŃSKI, *An ultrasonic transcutaneous method of simultaneous evaluation of blood pressure, flow rate and pulse wave velocity in the carotid artery*, Abstracts of Satellite Symp. of the XIIIth World Congress of Neurol., Aachen, M-68, 1985.

[24] T. POWAŁOWSKI, B. PENSKO, *Noninvasive ultrasonic method of pressure and flow measurements for estimation of hemodynamical properties of cerebrovascular system*, Archives of Acoustic, **10**, 3, 303–314 (1985).

[25] R. RENEMAN, T. MERODE, P. HICK, A. MUYTJENS, A. HOEKS, *Age related changes in carotid artery wall properties in men*, Ultrasound in Med. and Biology, **12**, 6, 465–471 (1986).

[26] W. A. RILEY, R. W. BARNES, H. M. SCHEY, *An approach to the noninvasive periodic assessment of arterial elasticity in the young*, Preventive Medicine, **13**, 169–184 (1984).

[27] J. T. ROSCOE, *Fundamental research statistics for the behavioural sciences*, Holt Rinehard Winston, New York.

[28] W. SCHIMMLER, *Untersuchungen zu Elastizitäts – Problemen der Aorta*, Archiv für Kreislauffor-schung, Band 47, Heft 3–4 (1965).

[29] B. SIMON, A. KOBAYASHI, D. STRANDNESS, C. WIEDERHIELM, *Large deformation analysis of the arterial cross section*, J. of Basic Eng., **93D**, 2 (1971).

[30] J. STETTLER, P. NIEDERER, M. ANLIKER, *Theoretical analysis of arterial hemodynamics including the influence of bifurcations*, Annals of Biomedical Engineering, **9**, 145–164 (1981).

[31] T. TANAKA, Y. FUNG, *Elastic and inelastic properties of the canine aorta and their variation along the aortic tree*, J. Biomechanics, **7**, 389–395 (1974).

[32] H. WEIZSACKER, A. PASCALE, *Anisotropic passive properties of blood vessel walls* in: Cardiovascular System Dynamics, Plenum Press, NY-London 1982, pp. 147–362.

Received on November 5, 1987.

ARCHIVES OF ACOUSTICS
13, 1–2, 127–135 (1988)

EVALUATION OF A COMPUTER-MODEL FOR PVDF-TRANSDUCERS OF ARBITRARY CONFIGURATION

JAN SOMER, JAN CORSEL AND HANS VANDERVOORT

Biophysics Dept., University of Limburg, Maastricht, The Netherlands

The electro-mechanical behaviour of PVDF can be described by the modified Mason equivalent circuit. It is possible to extract all necessary parameters and their frequency dependence from electrical input-impedance measurements and subsequent fitting-procedure with the computer-model. A chain-parameter matrix can readily be obtained for the equivalent circuit and matrix-multiplication enables the computation of any transfer function.

Elektromechaniczne zjawiska zachodzące w folii PVDF mogą być opisane przez zmodyfikowany układ zastępczy Masona. Pomiar elektrycznej impedancji wejściowej, a następnie dopasowanie wyników pomiarów do modelu komputerowego umożliwia wyznaczenie wszystkich żądanych parametrów i ich zależności od częstotliwości. Macierz parametrów może być z łatwością wyznaczona z otrzymanego układu zastępczego, a metoda macierzowa pozwala na obliczenie dowolnej funkcji przenoszenia.

1. Introduction

For many years transducers for medical ultrasound and other purposes were based on the application of piezoelectric ceramics. Particularly in using these ceramics for medical diagnostic purposes, where very short pulses are required, the high acoustic impedance has always caused problems.

Moreover, especially in more sophisticated constructions such as phased and linear arrays, suppressing the effects of vibrational modes and coupling coefficients other than just wanted has always been difficult.

Probably due to these shortcomings of piezoelectric ceramics the welcoming of the new piezoelectric plastic PVDF (polyvinylidenefluoride) has been accompanied with a lot of optimism. The much lower acoustic impedance of this plastic would naturally match much better to that of water and biological tissues, resulting in more efficient transmission of acoustic energy into the body, automatically yielding an

effective damping which is desired for wide-band behaviour. It was obviously expected that the better matching would sufficiently compensate for the unfortunately much lower coupling factor of PVDF as compared with e.g. PZT-5A. Even if this were true, still the relatively low dielectric constant is responsible for a high electrical input impedance of PVDF-transducers, which may amount 100 to 200 times as much as for comparable PZT-transducers. This results in very high voltages required for generating a practically useful acoustic output power. In other words, the transmission sensitivity in terms of the ratio of acoustic output power to squared input voltage is considerably lower than for comparable PZT-transducers.

Of course, in the light of the semi-conductor technology high driving voltages are most inconvenient. As a remedy against this trouble it was suggested to construct transducers by stacking several PVDF-layers with alternating poling directions [1, 2] as shown in Fig. 1. Then, when connected as shown, these layers are acoustically

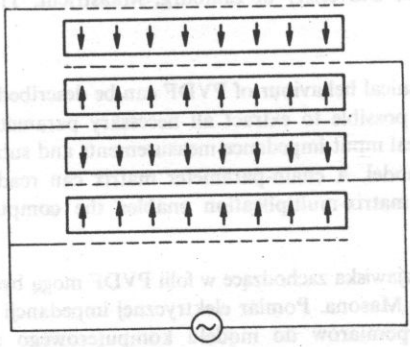


FIG. 1. Stacked transducer configuration with alternating poling directions. Electroded in such a way that the piezoelectric layers are electrically in series and acoustically in parallel

in series and electrically in parallel. In this way we can lower the electrical input impedance with a factor n^2, where n is the number of active layers, as compared with a one-layer PVDF-transducer of similar overall thickness. In this way, more electrical input power is achieved at a lower voltage and, consequently, a corresponding acoustical output power, depending on the coupling factor and the matching to the medium. This idea is by no means new, since already in 1926 it was LANGEVIN who patented such a system using quarts layers [3, 4].

Unfortunately, such a technique to improve the transmision sensitivity will lower the reception sensitivity by a factor n. Complicated electrode-switching may perhaps lead to a compromise yielding a practical transmission sensitivity combined with a best possible reception sensitivity [2].

It will be appreciated that finding the best configuration for a particular purpose would be an almost impossible task since the number of possible combinations is very high, especially when also matching layers and backing are involved. When realizing that bonding several thin metallized PVDF films together, with bonding

layers of let us say 1 μm and of perfect quality, which can hardly be tested, the above conclusion seems more than justified.

The purpose of this study is to evaluate a reliable computer model describing the electro-acoustic behaviour of piezo-electric PVDF-film. Then, using the method of chain-parameter matrix multiplication the performance of any desired transducer-configuration can be computed as a function of e.g. frequency. In this way we might be able to find the theoretical optimum for any transducer-system before trying to realize it in practice.

2. Equivalent circuit for a single layer

It would be most attractive if the values of all parameters, necessary to fully describe the transducer's behaviour, could be extracted from simple measurements of the electrical impedance as a function of frequency.

For this purpose test objects were used consisting of nominally 25 μm thick PVDF-film electroded in such a way that an accurately defined circular area of 6 mm diameter was poled. The measured electrical input impedance as a function of frequency showed remarkable features. Fig. 2 shows the frequency dependence of such a test object where the impedance is assumed to consist of a parallel circuit of a capacitance and a resistance as given in Fig. 3.

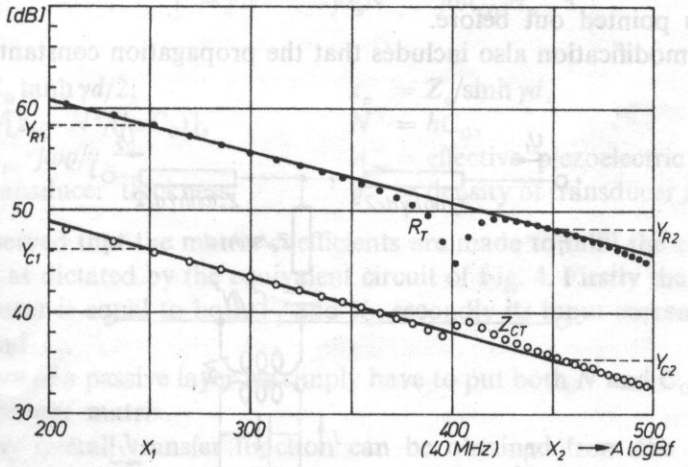


FIG. 2. The behaviour of 20 log R_T and 20 log Z_{CT} as a function of frequency at logarithmic frequency scale (Bode-plots). R_T (asterisks) and Z_{CT} (circles) represent the measured elecrical input-impedance Z_T of a single-layer PVDF-transducer in air, with $Z_{CT} = 1/(\omega C_T)$, according to Fig. 3. The parameters y_{R1}, y_{R2}, y_{C1} and y_{C2} are the coordinates at x_1 and x_2 respectively, of the straight lines through the impedance curves. The constants A and B can be obtained from the data in Fig. 5 and are chosen such that $A \log(Bf) = 400$ corresponds to $f = 40$ MHz

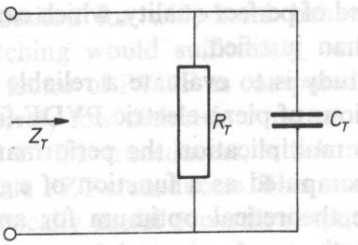


FIG. 3. The electrical impedance circuit representation used for both the measurements and the computer model

When both R_T and $Z_{CT} = 1/(\omega C_T)$ are displayed in dB on a logarithmic frequency scale, so-called Bode-plots, the overall behaviour is as a straight line. Frequency-independent resistance and capacitance would show a horizontal and a 6 dB/octave line respectively. Since this is not true we can conclude that both the dielectric losses and the dielectric constant are frequency dependent. Fortunately enough this straight line behaviour in Bode-plot representation enables us to readily describe these parameters mathematically as a function of frequency.

Around the mechanical resonance frequency, here about 40 MHz, the piezoelectric activity manifests itself as a deviation from the straight lines of both C_T and R_T. This invites us to adopt the well known Mason equivalent circuit, however modified, to represent this behaviour [5, 6]. By modification we mean that the clamped capacitance C_0 should be considered complex in order to account for the dielectric losses through its real part. Both real and imaginary part of C_0 are then frequency dependent as pointed out before.

Further modification also includes that the propagation constant γ is assumed

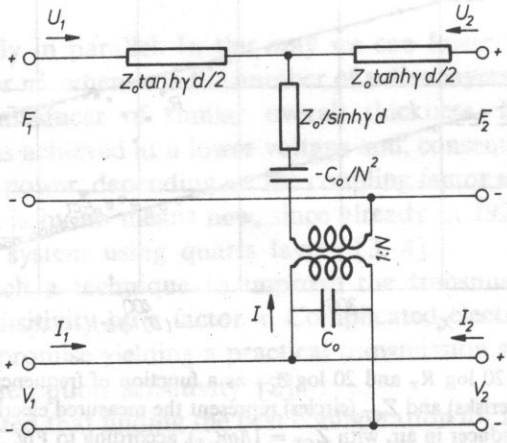


FIG. 4. Four-port equivalent circuit for a single-layer transducer, based on the Mason-model and adapted for application of chain-parameter matrices. Since the real part α of the propagation coefficient γ can not be neglected, the hyperbolic functions will not reduce to trigonometric functions

complex, $\gamma = \alpha + j(\omega/v)$, since mechanical losses, accounted for through α, should also be taken into account. Since C_0 is complex, also $N = hC_0$ is complex. For the time being h, the piezoelectric constant is assumed real. Later on it will be shown that both h and the mechanical loss coefficient α are also frequency dependent, although again obeying simple mathematical expressions.

Fig. 4. represents this equivalent circuit, drawn as a four-port rather than the usual three-port configuration [7].

3. Transfer functions of multi-layer transducers

In classical network theory the use of the chain-parameter metrix for calculating transfer functions of cascaded two-port networks is very common. The extension to four-port networks can be done straightforwardly, employing the basic equations from which the equivalent circuit was derived [7]. Then, transfer functions of multilayered transducers as in Fig. 1, can be computed by multiplication of the 4×4 chain-parameter matrices of all individual layers, including possible matching (passive) layers.

The chain-parameter matrix $[A]$ for the circuit of Fig. 4 takes the form

$$[A] = \begin{bmatrix} yz_s+1 & z_s(yz_s+2) & -yz_sN & 0 \\ y & yz_s+1 & -yN & 0 \\ 0 & 0 & 1 & 0 \\ -yN & -yz_sN & j\omega C_0 yz_p & 1 \end{bmatrix},$$

with

$z_s = Z_0 \tanh \gamma d/2$, $z_p = Z_o/\sinh \gamma d$,
$y = 1/[z_p - N^2/(j\omega C_0)]$, $N = hC_0$,
$Z_0 = A_{pe} \cdot j\omega\varrho/\gamma$, A_{pe} = effective piezoelectric area,
d = transducer thickness, ϱ = density of transducer material.

It can be observed that the matrix coefficients are made to fulfil the electrical input requirements as dictated by the equivalent circuit of Fig. 4. Firstly the input voltage of the transducer is equal to both V_1 and V_2, secondly its input current I equals the sum of I_1 and I_2.

In the case of a passive layer we simply have to put both N and C_0 equal to zero to find the proper matrix.

Now, any overall transfer function can be obtained from the expression

$$\begin{bmatrix} F_1 \\ U_1 \\ V_1 \\ I_1 \end{bmatrix} = [A_t] \cdot \begin{bmatrix} F_2 \\ -U_2 \\ V_2 \\ -I_2 \end{bmatrix}$$

with F = force, U = particle velocity, V = electrical voltage, I = electrical current, where

$$[A_t] = [A_b] \cdot [A_n] \cdot \ldots \cdot [A_i] \cdot \ldots \cdot [A_1] \cdot [A_f],$$

with

$[A_b]$ = chain-parameter matrix of back matching layer,
$[A_i]$ = chain-parameter matrix of i^{th} active layer,
$[A_f]$ = chain-parameter matrix of front matching layer.

4. Estimation of transducer parameters

The crucial step in the whole process is the estimation of the parameters h, α, v_s (sound velocity in transducer material), and the co-ordinates determining the straight lines through R_M and Z_{CM} (the computed impedance curves for the model), similar to the measured values R_T and Z_{CT} as in Fig. 2.

A computer programme has been created which enables us to display both the model-impedance curves and the measured ones simultaneously. Then each of the above mentioned parameters can be altered as desired and again all curves are displayed. After some trials a set of values for h, α, v_s and the co-ordinates y_{C1}, y_{C2}, y_{R1}, y_{R2}, x_1 and x_2 (see Fig. 2) will be obtained, which produce model-curves for R_M and Z_{CM} coinciding with the measured R_T and Z_{CT} as well as possible. We learn from Fig. 5 that such a procedure can lead to a remarkably close fit between measured and model impedance curves. Fig. 5a shows in detail the fit-accuracy around the resonance region whereas Fig. 5b gives the overall view showing a fit within some tenths of a decibel.

It should be realized that no information whatsoever can be extracted from this procedure about the values of h and α outside the resonance region. In order to estimate any frequency dependence a further investigation is necessary. In order to avoid uncertainties, already measured and fitted samples had to be used for this purpose. Several two- and four-layer transducers were realized by bonding with epoxy-resin whereby bonding thicknesses of ca. 1μm were obtained. As expected the new resonance frequencies were one half and one quarter of the original respectively. Again the actual impedance curves were measured and through the matrix-multiplication method corresponding models were computed. Then, by the fitting-method the new h and α values were determined for these composite samples at the pertinent resonance frequencies.

In Fig. 6 the right side scale shows A_1 rather than α. However, since $A_1 = \alpha d/2$, where d is the single layer thickness, there is only a constant factor involved, which does not influence the frequency dependence.

Through each (averaged) single layer value of both A_1 and h, and the corresponding values for the composite transducers, straight lines have been drawn

V_s	2.14550E+03	A	3.32193E+02
h	2.08039E+09	B	0.40000E−06
$\alpha d/2$	0.66277E−01	d	0.26000E−04
Y_{C1}	4.31350E+01	D	0.60000E−02
Y_{C2}	3.47960E+01	ϱ	1.80000E+03
Y_{R1}	5.59200E+01	x_1	3.00000E+02
Y_{R2}	4.74340E+01	x_2	4.60000E+02

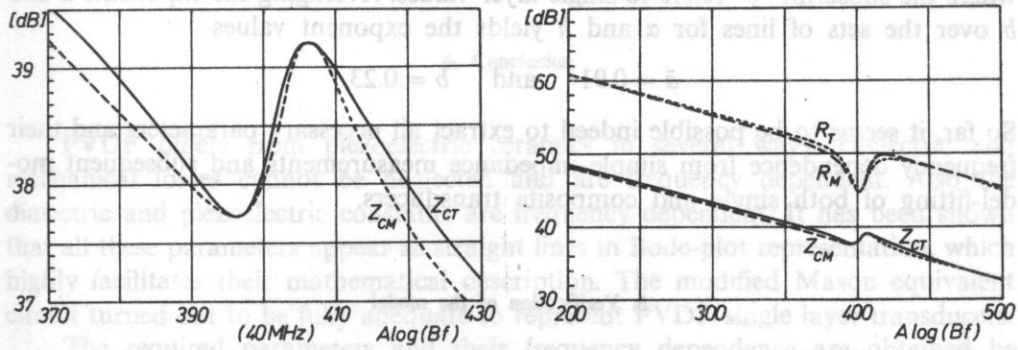


FIG. 5. Final fit-result of experimental and model impedance functions of a 25 μm thick PVDF-transducer of 6 mm diameter in air. Left: In the resonance frequency region (about 40 MHz) right: Frequency range from 10 to 80 MHz. The parameters at the right side are varied until the best fit has been obtained. These are: the sound propagation velocity v_s in m/sec, the piezoelectric constant $h\,(=h_{33})$ in V/m, the mechanical loss coefficient α in Neper/m multiplied by half the thickness and y-coordinates (see text and Fig. 2). The other parameters are all determined or chosen in advance like the constants A and B, used for the logarithmic frequency scale, the thickness d and the diameter D, both in m, the density ϱ in kg/m^3 and the x−coordinates (see text and Fig. 2)

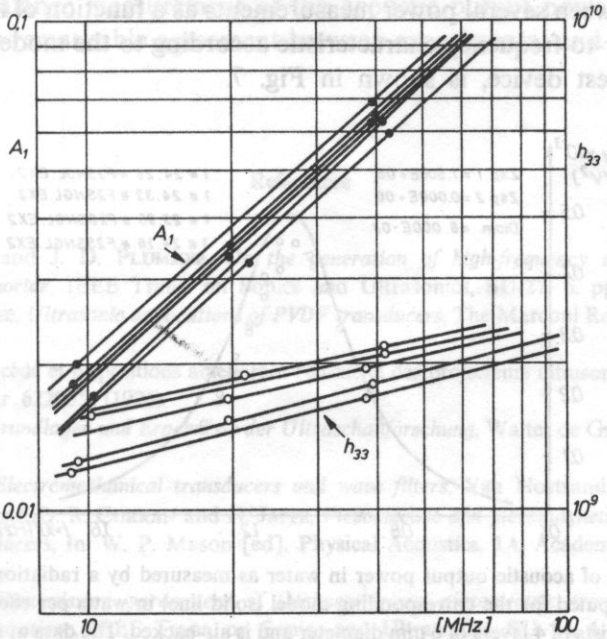


FIG. 6. Values for $A_1 = \alpha d/2$ (dots, left side scale) and h (circles, right side scale), all normalized to single layer conditions 1) around 10 MHz: values found for 4-layer transducers; 2) around 20 MHz: values found for 2-layer transducers; 3) around 40 Mz: averaged values of the single layers constituting the corresponding composite ones

on a double logarithmic scale. The parallelism between these lines indicates that h and α can readily be described as follows

$$\alpha(f) = \alpha_0 (f/f_0)^a \quad \text{and} \quad h(f) = h_0 (f/f_0)^b,$$

where the subscript "0" refers to single layer values. Averaging the exponents a and b over the sets of lines for α and h yields the exponent values

$$\bar{a} = 0.91 \quad \text{and} \quad \bar{b} = 0.23.$$

So far, it seems to be possible indeed to extract all necessary parameters and their frequency dependence from simple impedance measurements and subsequent model-fitting of both single and composite transducers.

5. Verification of the model

In order to avoid sources of error like radiation patterns and their artefacts, acoustic output power measurement with a radiation force balance seems to be the most appropriate method for verification of the developed computer-model.

One of the analyzed four-layer test objects was mounted in a holder to form an air-backed transducer radiating in water with the "perfect" absorber of the balance (in the form of a cone reflecting totally under a 90 degrees angle) in front of it. Comparison between several power measurements as a function of frequency and the calculated power-to-frequency characteristic according to the model as evaluated for this particular test device, is shown in Fig. 7.

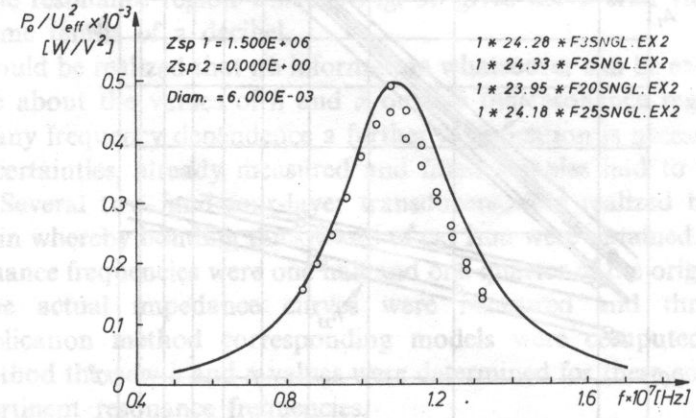


FIG. 7. Comparison of acoustic output power in water as measured by a radiation force balance circles and the values computed for the corresponding model (solid line) in watts per effective voltage squared. The transducer consists of 4 layers of 6 mm diameter and is air-backed. The data at the right side show the order in which the four layers are stacked, the thickness in μm of each layer and the pertinent data-files. For instance F3NGL.EX2 is the data-file obtained by means of the fitting-process for sample 3 according to Fig. 5, etc.

The agreement between measured and calculated results is most satisfactory. Of course, power measurement of only one test object cannot be considered a sufficiently reliable verification. Nevertheless, our approach to model-evaluation seems very promising.

6. Conclusion

PVDF differs from piezoelectric ceramics in several ways. Dielectric and mechanical losses cannot be neglected and are frequency dependent. Also the dielectric and piezoelectric constants are frequency dependent. It has been shown that all these parameters appear as straight lines in Bode-plot representation, which highly facilitates their mathematical description. The modified Mason equivalent circuit turned out to be fully adequate to represent PVDF single layer transducers.

The required parameters and their frequency dependence are obtained by finding the best possible fit between the input-impedance-to-frequency characteristics of the device and the computer model.

A four-port network approach leads to a 4×4 chain-parameter metrix fully representing this equivalent circuit mathematically. It is particularly suited for calculating any transfer function of stacked transducers as described, with or without matching layers, on the basis of matrix multiplication.

Radiation balance measurement of the acoustic output power of a four-layer transducer shows remarkable agreement between experimental and computer model results.

References

[1] R. G. SWARTZ and J. D. PLUMMER, *On the generation of high-frequency acoustic energy with polyvinylidene fluoride*, IEEE Trans. on Sonics and Ultrasonics, **SU-27**, 6. pp. 295–303 (1980).

[2] H. R. GALLANTREE, *Ultrasonic applications of PVDF transducers*, The Marconi Rev., **45**, 224, pp. 49–64 (1982).

[3] P. LANGEVIN, Procédé et dispositions améliorant l'efficacité des projecteurs ultrasonores piëzo-électrique, French Patent nr 622035. (1926).

[4] E. HIEDEMANN, *Grundlagen und Ergebnisse der Ultraschallforschung*, Walter de Gruyter and Co., p. 19, Berlin 1939.

[5] W. P. MASON, *Electromechanical transducers and wave filters*, Van Nostrand, Princeton 1948.

[6] D. A. BERLINCOURT, D. R. CURRAN and H. JAFFE, *Piezoelectric and piezomagnetic materials and their function in transducers*, In: W. P. Mason [ed], Physical Acoustics, **1A**, Academic Press, New York 1964.

[7] E. K. SITTIG, *Transmission parameters of thickness-driven piezoelectric transducers arranged in multilayer configurations*, IEEE Trans. on Sonics and Ultrasonics, **SU-14**, 4, pp. 167–174 (1967).

Received on December 14, 1987.

ARCHIVES OF ACOUSTICS
13, 1–2, 137–145 (1988)

IMPEDANCE OF THE SEMI-INFINITE UNBAFFLED CYLINDRICAL WAVE-GUIDE OUTLET

A. SNAKOWSKA, R. WYRZYKOWSKI

Institute of Physics, Pedagogical University Rzeszów, ul. Rejtana 16

The paper presents the exact formula for the impedance of the outlet of a semi-infinite cylindrical wave-guide derived by considering the propagation of an arbitrary Bessel mode towards the outlet and accounting for the generation of all permissible mode due to the diffraction at the open end. For this purpose, the formula of acoustic potential as well as the expressions for the reflection and transformation coefficient were used.

The results of numerical calculations of the real and imaginary part and the moduli of impedance for the diffraction parameter *ka* in the range 0–20 were presented on graphs.

Wprowadzono ścisłą relację dla impedancji półnieskończonego falowodu cylindrycznego na podstawie teorii uwzględniającej dowolny mod propagacyjny w kierunku otwartego końca falowodu i wszystkich możliwych modów generowanych na tej nieciągłości falowodu. Wyniki zilustrowano w funkcji parametru *ka* w zakresie 0–20.

Introduction

In the practical applications of acoustics, the phenomena occurring at the open ends of wave-guides seem to be important because we come across such elements in different acoustical equipment, e.g. measuring pipes, acoustic horns, tubes. The investigations of the problem were introduced by Lord Rayleigh [1] who calculated the impedance of the outlet provided additionally with an infinitely rigid acoustic baffle and assuming that only the plane wave propagates towards the end. As a result, Rayleigh obtained the well known "correction for the open end". The further step in solving the problem was made when H. LEVINE and J. SCHWINGER [2] derived the acoustical potential of incident plane wave inside the unbaffled semi-infinite cylindrical pipe. However, they neglected the "higher modes effect", i.e. they assumed that only the plane wave is reflected. It is obvious that such assumption is valid only when the wave length is not smaller than the diameter of the pipe, what strongly restricts the applications of the results. In 1948 [3] WAJNSHTEJN developed an analytical theory of the acoustic field of semi-infinite cylindrical

wave-guide applying the factorization method of solving the Wiener–Hopf integral equation. In his paper he worked out the exact formula for the acoustical potential inside the wave-guide making use of some analogies between acoustical and electromagnetic waves. The same results were obtained later by Snakowska and Wyrzykowski [4] who consequently applied to the problem the theory of acoustical field.

In this paper we calculate the impedance of the outlet of the semi-infinite cylindrical wave-guide for any z-axis symmetric Bessel mode propagating towards the end. For this purpose the exact formula for the acoustic field potential [3], [4] has been used.

The obtained numerical results are presented on graphs.

Index of symbols

a radius of wave-guide,
A_l, B_n amplitudes of Bessel modes,
D outlet area,
$H_n^{(1)}(\ )$ n-th order Hankel's function of first kind,
$J_n(\ )$ n-th order Bessel's function,
l, n indices of Bessel modes,
$L_+(\ ), L_-(\ )$ factors of L analytic in upper and lower complex half plane,
N index of the highest Bessel mode allowed in the considered wave-guide,
$N_n(\ )$ n-th order Neuman's function,
p acoustical pressure,
P apparent power,
R_{ln}, R_{II} reflection and transformation coefficients,
$S(\ )$ function describing transformation coefficient R,
v radial wave number $v = \sqrt{k^2 - w^2}$,
w partial wave number,
γ_n partial wave number of n-th Bessel's wave mode,
ζ_l impedance of the outlet for the l-th mode incident,
$v_{\bar{n}}$ normal velocity of vibration,
θ_{ln} phase of the transformation coefficient R_{ln},
$\varkappa$ diffraction parameter,
μ_n n-th zero of Bessel's function $J_1(\ )$,
ϱ_0 medium density,
Σ wave-guide surface,
$\Phi(\ )$ acoustic potential,
$\Psi(\ )$ acoustical potential discontinuity on Σ,
Ω outlet surface.
Other symbols used in the text are typical and are not listed here.

2. Basic formulae

We will consider the cylindrical wave-guide with an infinitely thin and rigid wall Σ which, in suitable coordinates, can be described as follows:

$$\Sigma = \{(\varrho, z): \varrho = a,\ z \geqslant 0\}.$$

To simplify the problem, we assume the z-axis symmetry (which means that the acoustic potential $\Phi(\bar{r}, t)$ does not depend on the angle variable φ in cylindrical coordinates) and the dependence on time in the form $\exp(-i\omega t)$.

The time dependent wave equation

$$\left(\Delta - \frac{1}{c^2}\partial_{tt}\right)\Phi(\bar{r}, t) = 0, \tag{1}$$

takes thus the following form:

$$\left(\frac{1}{\varrho}\partial_\varrho(\varrho\partial_\varrho) + \partial_{zz} + k^2\right)\Phi(\varrho, z) = 0. \tag{2}$$

The assumption that the wall Σ is perfectly rigid leads to the following boundary condition:

$$\partial_n \Phi|_\Sigma = \partial_\varrho \Phi|_\Sigma = 0, \tag{3}$$

which means that the normal component of vibration velocity vanishes at the wave-guide wall.

The solution of the problem consists in finding the function $\Phi(\varrho, z)$ which satisfies Eq. (2) for the boundary condition (3) and, moreover, the Sommerfeld's conditions of radiation [6]. The detailed investigations leading to the solution are enclosed in [3, 5].

The application of the three-dimensional Green free space function and factorization method to the equation of acoustic potential leads to the expression [3, 5]:

$$\Phi(\varrho, z) = \frac{ai}{4}\int_0^\infty \psi(z')\,dz' \int_{-\infty+i\eta}^{\infty+i\eta} v \begin{cases} H_0^{(1)}(v\varrho)\,J_1(va) \\ H_1^{(1)}(va)\,J_0(v\varrho) \end{cases} e^{iw(z-z')}\,dw, \qquad \begin{matrix} \varrho > a, \\ \varrho < a, \end{matrix} \tag{4}$$

with the boundary condition taking form of the integral equation [3, 5]:

$$\int_0^\infty \psi(z')\,dz' \int_{-\infty+i\eta}^{\infty+i\eta} v^2\,H_1^{(1)}(va)\,J_1(va)\,e^{iw(z-z')}\,dw = 0, \tag{5}$$

where $\psi(z)$ defines the potential discontinuity on the Σ surface

$$\psi(z) = \Phi(\varrho, z)|_{\varrho \to a_+} - \Phi(\varrho, z)|_{\varrho \to a_-}, \tag{6}$$

v being the radial wave number $v = \sqrt{k^2 - w^2}$. The potential discontinuity can be interpreted as the density of the surface sources on Σ.

Further development of the factorization method leads to the following expression for the acoustic potential $\Phi_l(\varrho, z)$ [3, 4]:

$$\Phi_l(\varrho, z) = A_l \frac{J_0\left(\mu_l \frac{\varrho}{a}\right)}{J_0(\mu_l)} e^{-i\gamma_l z} + \sum_{n=0}^{\infty} B_n \frac{J_0\left(\mu_n \frac{\varrho}{a}\right)}{J_0(\mu_n)} e^{i\gamma_n z}. \tag{7}$$

The index l points at the fact that we consider the case of one simple Bessel mode incident. It is well known from the theory of infinite cylindrical wave-guide that the radial wave number of such a mode must be equal to $\gamma_n = \sqrt{k^2 - \left(\frac{\mu_n}{a}\right)^2}$, where μ_n is the n-th zero of the Bessel function $J_1(\,)$ and, moreover, to have γ_n real, the diffraction parameter $\varkappa = ka$ must be not smaller than μ_n. This leads to the conclusion that the index N of the highest mode which can propagate without scattering must fulfill the following condition $\mu_N \leqslant ka < \mu_{N+1}$.

The first component in Eq. (7) represents the l-th Bessel mode which, according to the assumption, propagates towards the wave-guide outlet, where it is partly reflected (a component with $n = l$ index under the sum sign) and, due to diffraction, is transformed into an infinite number of Bessel modes (other components under the sum sign). Analyzing carefully the exponential expression under the sum sign, we can see that for a fixed diffraction parameter ka only a certain number of components will represent the modes which can propagate along the wave-guide because starting from $N+1$, the exponents will became negative real numbers and thus the corresponding components of the sum will represent a disturbance attenuated exponentially with increasing z. Since these disturbances are not the energy carrying waves, they will be neglected in further considerations of impedance.

Reflection and transformation coefficients

According to previous assumptions in further developments, we will take into account the following expression for the acoustic potential inside the wave-guide [3, 4]:

$$\Phi_l(\varrho, z) = A_l \left[\frac{J_0\left(\mu_l \frac{\varrho}{a}\right)}{J_0(\mu_l)} e^{-i\gamma_l z} + \sum_{n=0}^{\infty} R_{ln} \frac{J_0\left(\mu_n \frac{\varrho}{a}\right)}{J_0(\mu_n)} e^{i\gamma_n z} \right], \tag{8}$$

because we usually describe the diffraction phenomena on the outlet introducing the so-called reflection ($R_{ll} = B_l/A_l$) and transformation ($R_{ln} = B_n/A_l$, $n \neq 1$) coefficients. Detailed calculations [4], [5] lead to the following expression:

$$R_{ln} = \frac{-2\gamma_l}{\gamma_l + \gamma_n} \left(\prod_{\substack{i=0 \\ i \neq l}}^{N} \frac{\gamma_i + \gamma_l}{\gamma_i - \gamma_l} \prod_{\substack{i=0 \\ i \neq n}}^{N} \frac{\gamma_i + \gamma_n}{\gamma_i - \gamma_n} \right)^{1/2} e^{\frac{1}{2}[S(\gamma_l) + S(\gamma_n)]}, \tag{9}$$

$S(w)$ being equal

$$S(w) = \frac{2w}{\pi} \int_{ka}^{i\infty} \frac{1}{w^2 - w'^2} \left(\operatorname{tg}^{-1} \frac{N_1(a\sqrt{k^2 - w'^2})}{J_1(a\sqrt{k^2 - w'^2})} - n\pi \right) dw'. \tag{10}$$

Effective calculation of the values of R_{ln} coefficients as functions of the diffraction

parameter ka are only possible by numerical methods because the $S(w)$ function cannot be expressed analytically. In the calculations the generally accepted definition of modules and phase of the wave reflection and transformation coefficient is used [3]:

$$R_{ln} = -|R_{ln}|\, e^{i\theta_{in}}. \tag{11}$$

Knowing the explicit expressions for those coefficients, we finally obtained the explicit form of acoustic potential inside the wave-guide, which is necessary to calculate the acoustic impedance of the outlet.

4. Outlet impedance

To calculate the outlet impedance of the wave-guide, we shall use the formula of apparent acoustical power [7]

$$P = \int_{\Omega} \vartheta_{\bar{n}}(\varrho, 0)\, p(\varrho, 0)\, \mathrm{d}\sigma, \tag{12}$$

which is the surface integral over the outlet from the normal component of the velocity of vibration $\vartheta_{\bar{n}}$ and acoustic pressure p.
The required impedance is related to the apparent power P by the formula [7]:

$$\zeta = \frac{P}{D\langle \vartheta_{\bar{n}}^2 \rangle}, \tag{13}$$

where $\langle \vartheta_n^2 \rangle$ is the quadratic mean of the velocity at the outlet. The two quantities under the integral (12) are connected with the acoustic potential as follows:

$$p = -i\omega\varrho_0 \Phi, \tag{14}$$

$$\vartheta_{\bar{n}} = \partial_{\bar{n}} \Phi. \tag{15}$$

From simple calculation we get

$$p(\varrho, 0) = -i\omega\varrho_0 \sum_{n=0}^{N} (\delta_{ln} + R_{ln}) \frac{J_0\left(\mu_n \frac{\varrho}{a}\right)}{J_0(\mu_n)}, \tag{16}$$

$$\vartheta_{\bar{n}}(\varrho, 0) = iA_l \sum_{n=0}^{N} (-\delta_{ln} + R_{ln})\gamma_n \frac{J_0\left(\mu_n \frac{\varrho}{a}\right)}{J_0(\mu_n)}, \tag{17}$$

which leads to the following form of the impedance:

$$\zeta_l = -\omega\varrho_0 \frac{\sum_{n=0}^{N} |R_{ln}|^2 \gamma_n - (1 + 2i\, Im\, R_{ll})\gamma_l}{\sum_{n=0}^{N} |R_{ln}|^2 \gamma_n^2 + (1 - 2\, Re\, R_{ll})\gamma_l^2}. \tag{18}$$

The above result was obtained on the basis of the orthogonality property of the weighted Bessel functions [8]:

$$\int_0^a J_0\left(\mu_n \frac{\varrho}{a}\right) J_0\left(\mu_l \frac{\varrho}{a}\right) \varrho \, d\varrho = \frac{1}{2} \delta_{ln} a^2 J_0^2(\mu_l). \tag{19}$$

The real and imaginary part of the impedance are referred to the specific impedance of environment $\varrho_0 c$:

$$\operatorname{Re} \zeta_l^w = k \frac{\gamma_l - \sum_{n=0}^{N} |R_{ln}|^2 \gamma_n}{\sum_{n=0}^{N} |R_{ln}|^2 \gamma_n^2 + (1 - 2 \operatorname{Re} Rll) \gamma_l^2}, \tag{20}$$

$$\operatorname{Im} \zeta_l^w = k \frac{2 \gamma_l \operatorname{Im} R_l}{\sum_{n=0}^{N} |R_{ln}|^2 \gamma_n^2 + (1 - 2 \operatorname{Re} R_{ll}) \gamma_l^2}. \tag{21}$$

For $N = 0$ we get the case considered by Levine and Schwinger [2]. Expression (18) takes then the well-known form

$$\zeta_0^w = \frac{1 + R_{00}}{1 - R_{00}}, \tag{22}$$

R_{00} being the reflection coefficient of the plane wave.

5. Conclusions

The computer calculations of the real and imaginary part as well as moduli of the outlet impedance have been performed for the diffraction parameter ka varying within the range [0,20]. In Figs. 1 and 2 we compare the values of the acoustic impedance (real and imaginary part) of the unbaffled wave-guide for the plane wave outlet incident, respectively, with the results obtained by Rayleigh (Fig. 1) and by LEVINE and SCHWINGER [2].

As it can be seen from Fig. 1, the values of acoustic resistance computed by using the exact formulae are 1.5–2.0 times smaller than those obtained by Lord Rayleigh, although that difference decreases for the value of the diffraction parameter $ka \sim 3.0$. From the physical point of view it is obvious, because the baffle provides better radiation conditions. It is interesting to note that the acoustic reactance proceeds similarly only for $ka < 1.3$, for which value the two curves intersect and afterwards the values computed along the exact formula are about twice as large as Rayleigh's ones.

In Fig. 3 presenting the acoustic reactance for the plane wave and succeeding five Bessel modes incident, the following regularities can be seen:

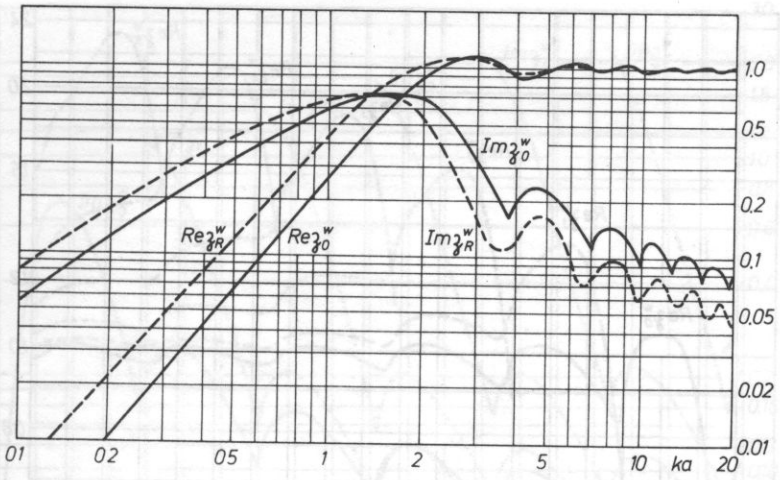


FIG. 1. Acoustic resistance and reactance of the unbaffled wave-guide outlet for the plane wave incident calculated after taking into account the higher Bessel modes which appear, due to diffraction phenomena (continuous line), compared with Rayleigh's resistance (dashed line)

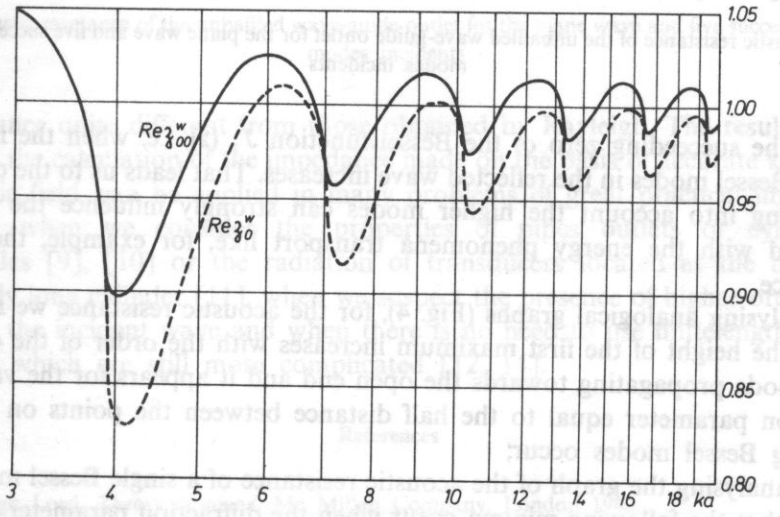


FIG. 2. Acoustic resistance of the unbaffled wave-guide outlet for the plane wave incident calculated after taking into account (dashed line) and neglecting (continuous line) the higher Bessel modes which appear on the open end, due to diffraction phenomena

– the acoustic reactance of the succeeding Bessel modes increases and the maxima appears for such values of diffraction parameter *ka* for which the adequate mode appears;

– analysing the diagram of the acoustic reactance of a single Bessel mode, it can be seen that the following maxima occur when the diffraction parameter *ka*

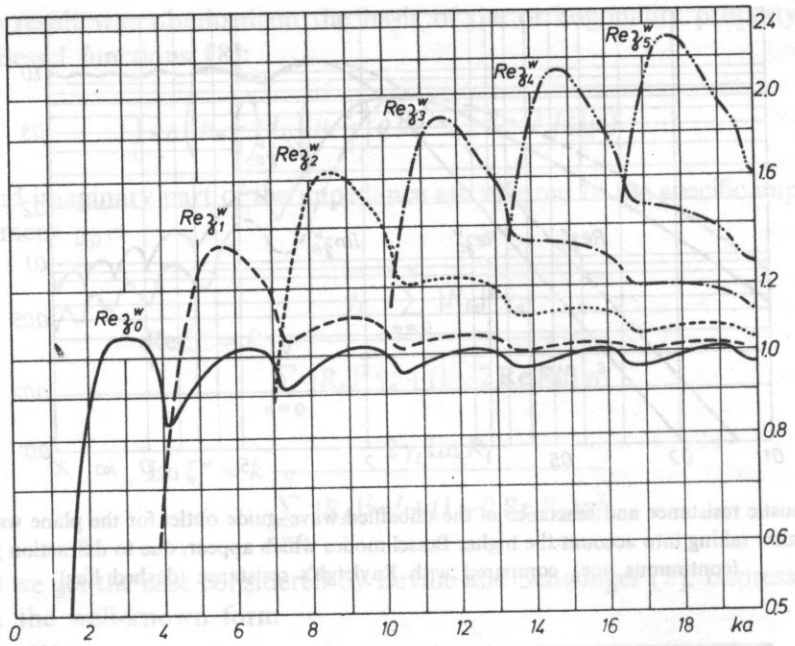


FIG. 3. Acoustic resistance of the unbaffled wave-guide outlet for the plane wave and five succeeding Bessel modes incidents

exceeds the succeeding zero of the Bessel function $J_1(z)$, i.e. when the number of allowed Bessel modes in the reflected wave increases. That leads us to the conclusion that taking into account the higher modes can strongly influence the quantities connected with the energy phenomena transport like, for example, the acoustic impedance.

Analysing analogical graphs (Fig. 4), for the acoustic resistance we notice that

– the height of the first maximum increases with the order of the considered Bessel mode propagating towards the open end and it appears for the value of the diffraction parameter equal to the half distance between the points on which the following Bessel modes occur;

– analysing the graph of the acoustic resistance of a single Bessel mode, it can be seen that the following minima occur when the diffraction parameter *ka* exceeds the value for which the next Bessel mode appears.

The obtained diagrams show that for wave-length shorter than the diameter of the wave-guide, in the presence of higher order Bessel modes the values of impedance differ considerably from those obtained for a plane wave by Rayleigh. It is obvious that for such a case the plane wave approximation can lead to important errors. It is well known that practically the generation of an ideal plane wave is very difficult, especially for the wave-guide with a large diameter in comparison with the wave length. In such a case we must consider the incident wave as a superposition of all allowed Bessel modes and it is possible that their contribution would lead to a value

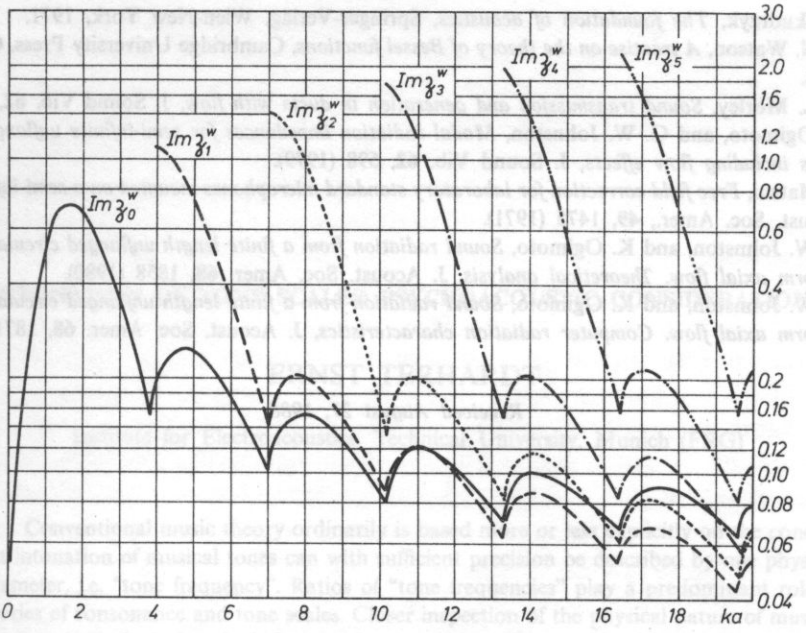


FIG. 4. Acoustic reactance of the unbaffled wave-guide outlet fot the plane wave and five succeeding Bessel modes incidents

of impedance quite different from those obtained by Rayleigh. The results of that paper, i.e. the calculation of the impedance made on the basis of accurate knowledge of acoustic field, can be applied in many problems of great practical importance, especially when we consider the properties of pipes outlets or cylinder-like wave-guides [9], [10] or the radiation of transducers located at the bottom of a relatively long cylinder [11], when we suspect the presence of higher-order Bessel modes in the incident wave and when there is no need to use finitelength cylinder formulae, which are still more complicated [12, 14].

References

[1] Rayleigh Lord, *Theory of sound*, Mc Millan Company, London 1940.

[2] H. Levine, and J. Schwinger, *On the radiation of sound from an unflanged circular pipe*, Phys. Rev., **73**, 383 (1948)

[3] L. A. Wajnshtejn, *The theory of diffraction and the factorization method (Generalized Wiener-Hopf technique)*, Golem 1969.

[4] A. Snakowska, and R. Wyrzykowski, *Calculation of the acoustical field of a semi-infinite cylindrical wave-guide by means of the Green function expressed in cylindrical coordinates*, Archives of Acoustics, **11** (1986).

[5] A. Snakowska, and R. Wyrzykowski, *Wybrane zagadnienia teorii dyfrakcji*, Wydawnictwo WSP Rzeszow 1984.

[6] B. Noble, *Methods based on the Wiener-Hopf technique for the solution of partial differential equations*, Pergamon Press, London-N. York, 1958.

[7] E. Skudrzyk, *The foundation of acoustics*, Springer-Verlag, Wien-New York, 1971.
[8] G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge University Press, Cambridge, 1958.
[9] C. L. Morley, *Sound transmission and generation in ducts with flow*, J. Sound Vib. **62**, 37 (1971).
[10] K. Ogimoto, and G. W. Johnston, *Modal radiation impedances for semi-infinite unflanged circular ducts including flow effects*, J. Sound Vib. **62**, 598 (1979).
[11] E. Matsui, *Free field correction for laboratory standard microphones mounted on a semi-infinite rod*, J. Acoust. Soc. Amer., **49**, 1475 (1971).
[12] G. W. Johnston, and K. Ogimoto, *Sound radiation from a finite length unflanged circular duct with uniform axial flow. Theoretical analysis*, J. Acoust. Soc. Amer. **68**, 1858 (1980).
[13] G. W. Johnston, and K. Ogimoto, *Sound radiation from a finite length unflanged circular duct with uniform axial flow. Computer radiation characteristics*, J. Acoust. Soc. Amer. **68**, 1871 (1980).

Received August 27, 1988.

ARCHIVES OF ACOUSTICS
13, 1-2, 147-156 (1988)

INTONATION OF TONE SCALES: PSYCHOACOUSTIC CONSIDERATIONS

ERNST TERHARDT

Institute for Electroacoustics, Technical University, Munich (FRG)

Conventional music theory ordinarily is based more or less explicitly on the concept that intonation of musical tones can with sufficient precision be described by one physical parameter, i.e. "tone frequency". Ratios of "tone frequencies" play a predominant role in theories of consonance and tone scales. Closer inspection of the physical nature of musical tones, and particularly of pitch-dependent auditory effects, reveals, however, that the aforementioned classical concept is insufficient. A psychoacoustically-oriented dualistic approach to intonation is suggested maintaining that "correct" intonation of a musical tone interval basically depends on (1) harmonic purity, and (2) sensory purity. Harmonic purity depends on memorized pitch-interval templates which in turn are partly of natural, partly of cultural origin. Sensory purity largely depends on perception of fluctuations, and its basic aspects are independent of cultural effects. It is concluded that optimal intonation is a compromise which at every moment of a musical performance must be achieved by active evaluation of the two aforementioned criteria. Theories of perception of pitch and fluctuations readily explain why this is so, and provide promising tools for achieving that compromise.

1. Introduction

The problem of tone scale intonation has been discussed and attacked for some thousand years with remarkably little success. This paper attempts to provide that problem with some new conceptual aspects which may be helpful when modern psychoacoustic methods are considered along the way to a final solution. Actually, scale intonation can hardly be properly discussed without taking into consideration how tone scales as such may have been developed. Fortunately, it is not essential to know every detail of that development; rather it is regarded sufficient to have a concept about what the route of development could have been, on the basis of a number of universal and consistent auditory criteria. In that sense, the introduction to the intonation problem following in the next paragraph itself is part of a proposed concept.

The pitch dimension, which basically is continuous, was dissected into discrete

pitch categories quite early in history. As is well known, musical tones bear certain relationships to each other not only in terms of their pitch differences but also in terms of certain additional qualities. Those qualities appear to have provided the criteria for tone categorization from the very beginning of music. As was made quite obvious by Pythagoras' principle of tone scale generation, i.e. concatenation of fifths and octaves, the most pronounced tonal affinities, i.e. octave- and fifth-affinity, provide the criteria necessary and sufficient to explain why tone scales could hardly have developed in a different manner than they actually did. In this view, the development from the pentatonic through the diatonic to today's chromatic scales appears straightforward and cogent. This is so, at least, if one for a moment ignores the intricacies of intonation and just considers tone categories. Probably, the main mistake inherent in most of the classical tone scale theories is an intermingling of the aspects of tone category and tone intonation. As will be further pointed out below, tone intonation is much more complex than ordinarily has been assumed; in particular, it is not just a matter of small-integer "tone-frequency" ratios.

Thus when we first ignore intonation problems, the development and typical features of tone scales may be seen as follows. Each of the three scales shown in Fig. 1 are an ordered collection of tone categories, i.e. notes, arranged according to their height; and the notes of each scale have been chosen according to the criteria of octave- and fifth-affinity, respectively. However, neither the pentatonic nor the diatonic scales are conclusive in a sense. While in these scales for any arbitrarily chosen tone another can be found which is in an octave relationship, this is not true in each case for fifth-third-, etc. relationships. This can generally be expressed by saying that in these two scales *interval width* (i.e. the number of steps encompassed by two tones) is not an unequivocal indicator of *interval quality* (where quality means the type of tone affinity pertinent to a particular interval category; e.g. the qualities of

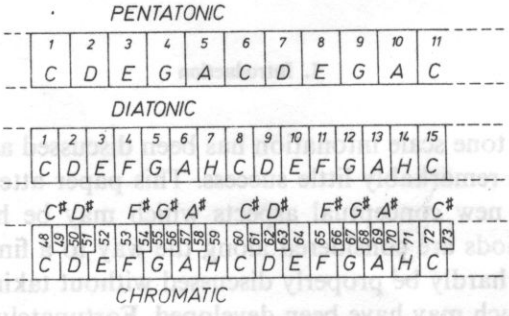


FIG. 1. Schematic representation of the pentatonic, diatonic, and chromatic tone scales. They can be regarded as representing different states of one and the same line of development, i.e. selection of notes in terms of octave- and fifth-affinity. While octave periodicity is implemented in all of them, the first two scales are not conclusive with respect to providing fifths and other intervals. The chromatic scale represents the final state of development, as interval width i.e. number of steps between two tones is an unequivocal indicator of interval quality (i.e. "octaveness", "fifthness", etc.), no matter which note is chosen as an interval's basis

"octaveness", "fifthness", etc). In this sense it can be said that today's 12-note chromatic scale in fact is conclusive, as interval width unequivocally indicates interval quality, no matter which tone is chosen as an interval's basis.

Actually it is that characteristic of the chromatic scale that determines its advantage of being "well-tempered". The aforementioned way of expressing the chromatic scale's unique features is different from, yet musically more relevant than saying that the chromatic scale is obtained by subdividing the octave into twelve "equal steps". The insufficiency of the latter statement is that it does not say in what respect the steps are equal.

Conceptually, a musically relevant theory of tone scales and intonation can hardly be achieved unless the basic difference between "scale" and "intonation" is understood and strictly observed. Here the "Three Worlds Concept" put forward by POPPER [6] and ECCLES [3] is extremely helpful, as it provides to the problem a perfectly fitting frame. That concept maintains that there are three basically different areas ("worlds"), into which all human experiences can be assigned. These are: 1) the world of physical/chemical processes and states, i.e. the "real world", it is called "World 1"; 2) the world of sensory experiences in the widest sense ("World 2"), and 3) the world of information, in particular of symbolically represented products of the human brain ("World 3").

In fact, music exists in three fundamentally different representations, each of which pertains to one of the three worlds (Fig. 2). In World 1, music exists as sound; in World 2, as auditory sensation, in World 3, as a score. In that conceptual frame, a musical tone scale as such is a symbolic representation of tone categories, i.e. notes;

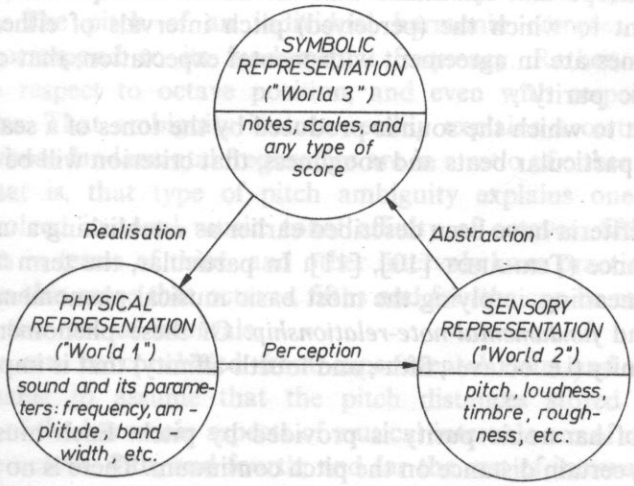


FIG. 2. The three "worlds" of musical reality. In this concept, a musical tone scale is a collection of notes, i.e. symbolic representation of tones; it is pertinent to "World 3". Intonation of a scale is a collection of corresponding physical parameters such as part-tone frequencies and amplitudes ("World 1"). Auditory sensation such as pitch, and roughness provide the decisive criteria of intonation ("World 2")

it is pertinent to World 3. The scale's intonation implies all the physical parameters (mainly, but not only, frequencies) which are correlated with the auditory sensations produced by the sound these are mainly the pitch and beat sensations, as will be pointed out below; those physical parameters are obviously pertinent to World 1. Finally, the auditory sensations mentioned are pertinent to World 2. The theory of tone scales must be concerned with the relationships and interactions between those three manifestations of musical tones. Those relationships can generally be termed *realisation, perception,* and *abstraction* (Fig. 2).

In the frame of that concept we can now start the discussion of tone scale intonation with the help of the following two statements:

a) Intonation is dependent on auditory perception and abstraction. It is neither purely by mathematical (numerical) nor purely by phisical arguments that one can decide about intonation. Intonation of a tone scale is optimal if it satisfies the complex mechanisms of auditory perception and abstraction at any instant of a musical performance.

b) Intonation as such means specification of those physical sound parameters which affect the perceptual and abstract criteria mentioned. Besides the "fundamental frequency" of tones, virtually every other parameter such as spectral envelope and sound pressure level must be taken into consideration.

2. Criteria of intonation

The most promising, yet little recognized, concept of intonation appears to be the dualistic concept that maintains that intonation is dependent on:

1) the extent to which the (perceived) pitch intervals of either successive or simultaneous tones are in agreement with mental expectation; that criterion will be called "harmonic purity";

2) the extent to which the sounds produced by the tones of a scale are free from disturbances, in particular beats and roughness; that criterion will be called "sensory purity".

These two criteria have been described earlier as establishing a useful concept of musical consonance (TERHARDT [10], [11]. In particular, the term "harmony" was given a special meaning, implying the most basic musical phenomena: *tonal affinity, compatibility,* and *fundamental-note-relationship.* Of these phenomena it is in particular tonal affinity (i.e. octave-, fifth-, and fourth-affinity) that is important for scale intonation.

The basis of harmonic purity is provided by pitch. Each musical interval is represented by a certain distance on the pitch continuum. There is no justification for presuming that a particular interval quality corresponds to one and the same distance on every level of the pitch continuum. Rather, the pitch distance corresponding to a particular music interval in general will be dependent on absolute pitch height. The actual pitch distance by which a particular musical interval is

represented on the continuum depends on the metrics by which pitch height is measured. Fortunately, for the theory of musical intervals and intonation it is not necessary to make a decision on the metrics. It is sufficient to be aware that to each musical interval quality corresponds a certain pitch distance that even may be a function of pitch height, and that can be compared with corresponding memorized distances.

The decision of whether or not a given tone interval is "harmonically pure" will depend, among other effects, on the precision by which the corresponding pitch distances are represented in memory. What can be said about the origin and the precision of that representation?

On the basis of psychoacoustic evidence, there is no indication that *every* interval of today's chromatic scale is of natural origin, i.e. either acquired in basic perceptual processes or "hard-wired" in the auditory nervous system. Research on children's intonation in singing appears to indicate a considerable amount of learning in early life. While young children may show a good sense of melodic contour, they usually care little about tone- and interval-categories, let alone precise intonation (DOWLING [2]). Appreciation of harmony appears to develop relatively late (SHUTER-DYSON [8]). It will thus appear reasonable to assume that for instance second- and perhaps even third-intervals, and in particular their intonation, are to a considerable extent culturally acquired.

On the other hand, there are a number of solid arguments in favour of the presumption that the most basic intervals, i.e. octave, fifth, and fourth, are acquired, or predominantly determined, by auditory spectrum analysis in the perception of natural speech (cf. TERHARDT [11]). Additional support to the basic and natural character of the octave, fifth, and fourth comes from pitch ambiguity of harmonic complex tones. The pitch of an individual harmonic complex tone does *not unequivocally* correspond to its fundamental frequency. Rather, there exists an ambiguity with respect to octave position, and even with respect to fifth- and fourth-confusions. That ambiguity in turn readily explains a certain similarity of musical tones whose fundamental frequencies are in a ratio of either 2:1, 1:2, 3:2, 2:3, 4:3, or 3:4. That is, that type of pitch ambiguity explains one of the crucial phenomena involved in tonal music: tonal affinity of octaves, fifths, and fourths. Pitch confusions in terms of third- and other intervals have practically never been observed. It is in this sense that octaves, fifths and fourths can be regarded as more "natural" than the rest of intervals.

If *anything* can be concluded from experimental data and observations, it appears reasonable to assume that the pitch distances stored in memory as representations of the harmonic aspects of musical intervals are of natural origin in the cases of the octave, fifth, and fourth; and for the rest of intervals are culturally dependent. If this is so, then it obviously does not make sense to raise the question of what the "natural" intonation is of thirds, seconds, etc. Experiments designed to determine the optimal intonation of those culturally dependent intervals will necessarily reflect only the listeners' mean previous intonation experiences. Results of

that type of experiment in fact confirm that view (for an extensive review see BURNS and WARD [1]).

With respect to the octave, fifth and fourth, the memorized pitch distances corresponding to them may be affected by actual intonation of real musical instruments as well. Therefore, the harmonic purity of those intervals should be regarded as possibly dependent on both a natural and a cultural component. A sort of competition between these two components may be observed in auditory tuning of octaves of successive tones by a trained musician: while his/her "natural" (and thus "naive") evaluation will ordinarily produce a considerable stretch (i.e. a fundamental-frequency ratio larger than 2:1; cf. WARD [14], WALLISER [13], TERHARDT [11]), cultural experience in ensemble playing may prevent him from stretching the octave by too great an amount. Evidence for that conclusion may be seen for example in recent results by MAKEIG [5].

Of the two intonation criteria, i.e. harmonic purity and sensory purity, the first can be regarded as functionally most relevant; in particular, it applies to both successive and simultaneous tones. Sensory purity, which virtually is dependent on fluctuation effects, i.e. beats, will in many cases provide another strong criterion of intonation which is significant only with simultaneous tones. As was extensively pointed out by HELMHOLTZ [4], musical chords composed of harmonic complex tones attain the highest sensory purity if the fundamental frequencies of the complex tones are in ratios of small integer numbers, e.g. 4:5:6 for the major triad. If the fundamental frequencies depart just a little therefrom, beats can be heard which ordinarily are disliked as being indicative of "mistuning". Harmonic purity is hardly affected by such a small amount of mistuning. As is common experience, slight mistuning does not disturb or even destroy the essential "musical message" but can considerably reduce the sensory pleasantness of a sound. In that sense harmonic purity may be regarded as "functional", while sensory purity is "cosmetical".

In general, the intonation which provides maximal harmonic purity is not necessarily identical with the intonation producing maximal sensory purity. For example, when two successive tones are tuned to give an optimal octave, their fundamental frequencies will turn out to be in a ratio slightly greater than 2:1. As a consequence, the same tones, when sounding simultaneously, may produce audible beats, i.e. less than optimal sensory purity. The latter would be achieved by tuning the tones exactly in the ratio 2:1, which however would render harmonic purity less than optimal. Obviously, intonation in general is a compromise.

So far we have been concerned only with the intonation criteria which depend on tone intervals, as opposed to individual tones. Naturally, intonation of individual, isolated tones depends on absolute-pitch recognition and thus is confined to listeners having absolute pitch. Since only a small percentage of musical listeners possess that ability, and since the development of musical scales is essentially dependent on tone intervals rather than on recognition of individual tones, the latter case is disregarded in the present study.

The criteria of intonation and their role in musical scales can thus be summarized as follows:

Intonation of successive tone intervals is exclusively governed by harmonic purity, i.e. matching of pitch distances to corresponding "templates" stored in memory.

The "templates" which correspond to octaves, fifths, and fourths, probably are of natural origin, i.e. independent of previous musical experiences. The "templates" corresponding to the rest of intervals probably are essentially dependent on previous experience and learning; i.e. they can be developed only if musical scales already exist.

Intonation of simultaneous tone intervals is dependent both on harmonic and sensory purity. In many cases, the latter will provide the most sensitive criterion, and in principle it will apply to all intervals, be they of natural or cultural origin.

Since, with a given musical sound and intonation, it is not always possible to fully satisfy both the criteria of harmonic and sensory purity, intonation generally is a compromise. This implies that there does not exist such a thing as a fixed ideal intonation.

To further understand the advantage provided by the present approach it will be helpful to critically discuss the classical concept of describing intonation merely in terms of "tone frequency" alone.

3. Criticism of the "tone-frequency concept of intonation"

In musical acoustics and music theory, the intonation of tones is ordinarily described by "the" frequency, in the sense of "oscillations per second". However, to make that concept fully valid and significant, two preconditions must be fulfilled. The first is a physical one: the oscillation frequency as such must be defined with sufficient precision; i.e. the tone's oscillations must be strictly periodic. The second condition is a psychoacoustical one: the pitch sensation must be *solely* dependent on the oscillation frequency, i.e. pitch must not depend, for instance, on the tone's spectral composition and sound pressure level. If either one or both of these conditions are violated, any discussion of intonation based on "tone frequency" becomes more or less inadequate. In fact it turns out that for certain types of tones (in particular, percussive tones such as of the piano) the first condition is violated; and the second condition is violated for practically every type of tone.

Concerning piano tones, it is well known that their spectra are slightly inharmonic (SCHUCK and YOUNG [7]). This means that the period of the tone's entire oscillation is not identical with that of the first partial alone: rather, the former is much longer. As a piano tone's pitch corresponds approximately to the period of its first partial, one can also say that its entire physical period is by far longer than that corresponding to its pitch. In other words, in this case the concept of associating the tone's pitch with its physical period fails profoundly. It is thus apparent that neither

harmonic nor sensory purity can be properly accounted for by the concept of "tone frequency". In particular, sensory purity of the piano tone scale cannot consistently be evaluated by the lowest partial frequencies alone, since it depends on beats between higher partials as well; therefore also precise specification of the frequencies of higher partials is required.

Musical tones which are with sufficient precision periodic are produced by instruments with a steady energy supplement, i.e. strings, horns, woodwinds, and organ pipes. The higher partials of those tones are harmonics, i.e. their frequencies are with sufficient precision determined by that of the fundamental, i.e. they are just integer multiples of the latter. Therefore assessment of sensory purity (beats of simultaneous part tones) can be accomplished largely by knowledge of the fundamental frequencies alone (that in turn are identical with oscillation frequencies in that case). However, beats are further dependent on the amplitudes of partials and,

Table 1. Typical magnitudes of effects relevant to intonation. Upper part: "numerical effects", i.e. phenomena which have been considered in terms of the "tone frequency approach". Middle part: piano string inharmonicity as an example of a physical effect. Lower part: aural effects which are pitch dependent and thus relevant to "harmonic purity" of intonation. Further descriptions and explanations of the latter effects can be found, e.g. in TERHARDT et al. [12].

	equivalent frequency cents	difference percent
Numerical effects		
Pythagorean comma, $(3/2)^{12}:2^7$	23.4	1.36
synthonic comma (departure of II-VI from pure fifth in just intonation), 81/80	21.5	1.25
difference between Pythagorean and natural major third, $(3/2)^4:5$	21.5	1.25
difference between natural and tempered fifth $(3/2):2^{7/12}$	2.0	0.11
difference between natural and tempered major third	−13.7	−0.79
Physical effect		
inharmonicity of aurally relevant partials of piano-bass-strings (typical example)	30	1.8
Pitch-dependent, i.e. aural effects		
just noticeable pitch difference of musical tones (typical example)	5	0.3
aural pitch shifts of true harmonics of complex tones: 1st harmonic	−20 ... 0	−1 ... 0
higher harmonics	0 ... 160	0 ... 10
difference between true and nominal pitch of musical tones: low tones	−50 ... 0	−3 ... 0
high tones	0 ... 50	0 ... 3
octave enlargement (octave matching of successive tones)	9 ... 50	0.5 ... 3

moreover, the perceived pitch of any complex tone depends in a complex way on the entire part-tone spectrum. In general, the pitch of a musical tone cannot be precisely determined by the fundamental frequency (or oscillation frequency) alone; rather, the frequencies and amplitudes of many partials play a significant role as well (cf. TERHARDT [12]. Therefore the second of the two aforementioned preconditions is not fulfilled with either type of musical tone.

Naturally the relevance and consequences of these arguments depend on the magnitude of the respective effects. Table 1 presents typical magnitudes of some numerical, physical, and auditory effects pertinent to intonation. The figures indicate that in fact the "mistuning" introduced by inharmonicity of piano strings (i.e. a physical effect) and by pitch shifts (i.e. departures of pitch from supposed nominal values) are at least of the same order of magnitude as classical numerical intonation effects such as for instance the pythagorean comma. It is thus apparent that the classical method of describing intonation just by "tone frequency" can only roughly account for perceptually relevant intonation criteria.

4. Conclusions

Whatever objections may seem justified against the details of the present approach to tone scale development and intonation, one conclusion appears to be quite safe: There is not even theoretically such a thing as an ideal fixed intonation which can be described by "tone frequencies" without making further specifications. Optimal intonation in every case and instant is a compromise dependent on partly contradictory criteria. Optimal intonation can neither be regarded as fixed nor can it be sufficiently specified by just one frequency per tone. Optimal intonation of tonal music must be flexible, i.e. adapting to momentary requirements. Any rigid assignment of frequencies to tones can thus serve only as an abstract reference pattern from which the optimal intonation in every instant will depart more or less distinctly. What type of intonation (i.e. just, pythagorean, or equally tempered) is used as a reference pattern, is of secondary importance, though equally tempered intonation appears to be most convenient for that purpose.

The dualistic concept of harmonic and sensory purity sketched in the present study may provide a systematic solution to the intonation problem. To take full advantage of that approach a theory of pitch perception to evaluate harmonic purity, and of beat- and roughness-perception to evaluate sensory purity are required. The pitch theory must in particular account for pitch shift effects such as those mentioned in Table 1. The virtual-pitch theory (TERHARDT [9], TERHARDT [12]) meets these criteria to a considerable extent. Since perception of beats and roughness is also well understood, it will appear that, in spite of the complexity of effects involved, we are beginning to understand musical tone scale intonation and likewise we are beginning to be able to predict quantitatively the physical sound parameters yielding optimal intonation.

Acknowledgments

The author gratefully acknowledges many helpful comments on the manuscript given by Dr A. JAROSZEWSKI and Dr I. LINDEVALD. This study was worked out in the Sonderforschungsbereich 204 Gehör, München, supported by the Deutsche Forschungsgemeinschaft.

References

[1] E. M. BURNS and W. D. WARD, *Intervals, scales, and tuning*, In: D. Deutsch [ed], The Psychology of Music, Academic Press, New York 1982.

[2] W. J. DOWLING, *Melodic information processing and its development*, In: D. Deutsch [ed], The Psychology of Music, Academic Press, New York 1982.

[3] J. C. ECCLES, *The understanding of the brain*, New York, McGraw-Hill, 1973.

[4] H. HELMHOLTZ, *Die Lehre von den Tonempfindungen*, Braunschweig, Vieweg 1963.

[5] S. MAKEIG, *Octave tuning — two modes of perception*, Res. Sympos. on the Psychol. and Acoustic of Music, Lawrence, Kansas 1982.

[6] K. POPPER, *Objective knowledge*, University Press, Oxford 1972.

[7] O. H. SCHUCK, and R. W. YOUNG, *Observations on the vibrations of piano strings*, J. Acoust. Soc. Am., **15**, 1-11 (1943).

[8] R. SHUTER-DYSON, *Musical ability*, In: D. Deutsch [ed], The Psychology of Music, Academic Press, New York 1982.

[9] E. TERHARDT, *Zur Tönhöhenwahrnehmung von Klängen*, Acustica, **26**, 173-199 (1972).

[10] E. TERHARDT, *Ein psychoakustisch begründetes Konzept der Musikalischen Konsonanz*, Acustica, **36**, 121–137 (1976).

[11] E. TERHARDT, *The concept of musical consonance, a link between music and psychoacoustics*, Music Perception, **1**, 276–295 (1984).

[12] E. TERHARDT, G. STOLL and M. SEEWANN, *Algorithm for extraction of pitch and pitch salience from complex tonal signals*, J. Acoust. Soc. Am., **71**, 679-688 (1982).

[13] K. WALLISER, *Über die Spreizung von empfundenen Intervallen gegenüber mathematisch harmonischen Intervallen bei Sinustönen*, Frequenz, **23**, 139–143 (1969).

[14] W. D. WARD, *Subjective musical pitch*, J. Acoust. Soc. Am., **26**, 369-380 (1954).

Received August 21, 1987.

ARCHIVES OF ACOUSITCS
13, 1 2, 157 189 (1988)

HALF-PLANE EDGE AND RIGHT ANGLE WEDGE AS ELEMENTS CAUSING DIFFRACTION IN URBAN AREA

ELŻBIETA WALERIAN

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

This study analyzed the acoustic fields generated by interactions between acoustic waves and the edge of a half plane and a right angle wedge. Using known solutions of the diffraction of a monochromatic wave on a half-plane and a right angle wedge, they were written in a form permitting simultaneous analysis of three wave types: plane, cylindrical and spherical. Approximate forms of solutions were adopted and the ranges of their applicability analyzed. In the space around the chosen obstacle, its efficiency was calculated with respect to a free field, for wavelengths and distances of interest in urban acoustics.

W artykule poddano analizie pola akustyczne jakie powstają na skutek oddziaływania fal akustycznych z krawędzią półpłaszczyzny i ostrzem klina o kącie rozwarcia równym kątowi prostemu. Wykorzystując znane rozwiązania dyfrakcji fali monochromatycznej na półpłaszczyźnie i ostrzu klina zapisano je w postaci pozwalającej na jednoczesną analizę trzech typów fal: fali płaskiej, cylindrycznej i kulistej. Przyjęto przybliżone postacie rozwiązań, dokonując analizy zakresu ich stosowalności. W przestrzeni wokół wybranych przeszkód obliczono ich efektywność w stosunku do pola swobodnego, dla długości fal i odległości stanowiących przedmiot zainteresowania akustyki urbanistycznej.

Basic notation

V part of the acoustic potential Ψ of the monochromatic wave dependent on the spatial coordinates: $\Psi = V\exp(-i\omega t)$,
k wave number: $k = 2\pi/\lambda = 2\pi f/c = \omega/c$, λ — wavelength, f — frequency, c — wave velocity, ω — angular frequency,
n normal to the barrier surface,
ϱ_0 radial coordinate of the source position,
ϱ radial coordinate of the position of the observation point,
ϕ_0 angular coordinate of the source position,
ϕ angular coordinate of the position of the observation point,
z_0 z-th coordinate of the source position,
z z-th coordinate of the position of the observation point.

Introduction

The notion of diffraction may be understood to mean all deviations from the laws of geometrical acoustics in the process of interactions between acoustic waves with obstacles whose acoustic properties are different from those of the ambient medium. In the simplest case, this is a step-like change in acoustic properties which occurs on the edge of a half-plane.

In practice diffraction is, e.g., a phenomenon which determines the efficiency of flat acoustic screens in the shadow area. However, the use of such screens as a measure against the noise propagation from highway requires considering changes in the acoustic field not only in the shadow area, but also throughout the space of the screen.

In general, wishing to describe the acoustic field in complex urban systems, it is first necessary to gain knowledge of the elementary processes forming the acoustic fields, including wave reflection and diffraction.

For urban systems it is acceptable to describe the acoustic field basing on the laws of geometrical acoustic, with a correction for the diffraction occurring on edges of the types of half-plane and wedge, e.g. diffraction on the half-plane type edge occurs for the flat acoustic screens mentioned above. Diffraction at the right angle wedge occurs at house corners, balconies etc. In the case of a depressed highway diffraction occurs at the wedge of the slope, this time at the wedge whose opening angle depends on the inclination angle of the slope, and which may be different from a right one.

Generally, highways can be recognized as the main noise source urban area. In the first approximation, for large distances, the highway can be considered as a source of plane waves. At shorter distances, the model of a linear source is assumed for it. Very close to the highway distinguishing individual vehicles, the waves from them are considered spherical waves. Using more complex models of highway, the acoustic field can be treated as one composed of elementary waves, e.g. plane, cylindrical and spherical.

Considering the two basic elements at which diffraction occurs (the half-plane and wedge) and three elementary wave types (plane, cylindrical and spherical), this study analyzes the acoustic fields for these cases.

Section 1 of the study presents a general description of the structure of the acoustic field for chosen cases. The starting point were the known solutions of the diffraction problem [1, 2] in which it is possible to extract the geometrical and diffraction parts. The study shows that the asymptotic forms of these solutions for the cases in question can be written in the form of the sum of waves forming independent pairs related to the real source and the image sources representing the waves reflected from the obstacle. The pair of waves related to one source consists of the geometrical wave occurring only in limited space around the barrier, and the diffraction wave present throughout the space.

The above description differs from the geometrical diffraction theory proposed

by KELLER [3, 4] in that, instead of diffraction rays describing the total effect of interaction with the obstacle, it introduces diffraction waves related to particular sources of geometrical waves. This makes it possible to explain the nature of the field at the geometrical boundaries and close to them, which cannot be done on the grounds of the geometrical diffraction theory.

Section 2 gives explicit forms of the component geometrical and diffraction waves distinguished in the description. It discusses in detail the applicability ranges of the asymptotic forms of solutions which provided the basis for the description introduced and which are different for particular cases. The regions where the conditions of asymptotic approximation are satisfied at the same time for all the cases presented coincide with the applicability range of the geometrical diffraction theory.

Taking into account the conditions met in urban area, section 3 compares the efficiency of two kinds of obstacle (half-plane and wedge) for three types of wave (plane, cylindrical and spherical) in the area of the geometrical shadow. Also, it distinguishes area where the presence of obstacle does not cause any significant disturbance and those where, as a result of interference between the waves in them, alternating increases and decreases in the total field amplitude occur.

The fields were analyzed for monochromatic waves from which complex acoustic signals are made. The description of the field used the acoustic potential which is linearly related to the acoustic pressure of the monochromatic wave.

1. Structure of the acoustic field around the half-plane and wedge

By using the appropriate approximations of the exact solutions of the problems of the wave-obstacle interactions, the acoustic potential expressions describing the total field around the obstacle can be written in the form of the sum of the geometrical and diffraction parts of the potential [1, 2].

The geometrical part of the acoustic potential can be obtained on the basis of the laws of geometrical acoustics in the form of the sum of geometrical waves coming from the real source and those from the image sources represented by the waves reflected from the obstacle surfaces.

The areas where geometrical waves are present are determined by the obstacle size. In view of this, the total geometrical field, which is the sum of geometrical waves, is discontinous. This discontinuity is compensated by the diffraction part of the acoustic potential.

After the appropriate rearrangement it was possible to represent the diffraction part of the potential in the form of the sum of diffraction waves, each related to the source of a geometrical wave. Around the geometrical boundary, where the geometrical wave from one of the sources vanishes, the related diffraction wave takes maximum values compensating the jump of the geometrical wave, providing at the same time, the continuity of the field.

On the ground of a description of the field containing the related pairs of geometrical and diffraction waves coming from the same source, one obtains a uniform description of the structure of the acoustic field around the obstacle independent of the type of wave and the kind of obstacle.

1.1. Structure of the acoustic field around the obstacle

The acoustic field of a monochromatic wave with the frequency f, around the ideal hard half-plane, can be determined by the solution of the Helmholtz equation for the acoustic potential

$$(\Delta^2 + k^2) V = 0, \tag{1}$$

with the boundary condition

$$\frac{\partial V}{\partial n} = 0. \tag{2}$$

This equation can be written in the following form:

$$V = V^g + V^d, \tag{3}$$

where V^g is the geometrical part of the acoustic potential and V^d is the diffraction part of the acoustic potential.

In a cylindrical coordinate system (Fig. 1) the equation of the half-plane

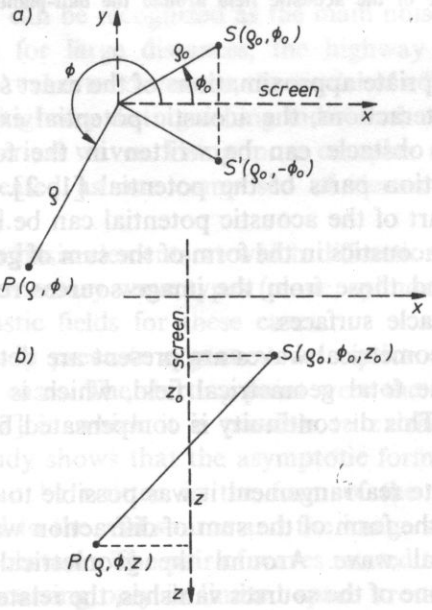


FIG. 1. The coordinate system applied: the screen is a half-plane with the equation $y = 0$, $0 \leqslant x < \infty$, or $\phi = 0$, P — observation point, S — source

disturbing the acoustic field, has the form

$$\phi = 0, \tag{4}$$

the geometrical part of the potential can be written in the form

$$V^g = \eta(\pi-\beta)\, V^i(R)+\eta(\pi-\beta')\, V^i(R'), \tag{5}$$

where

$$\beta = \phi-\phi_0, \tag{6}$$

$$\beta' = \phi+\phi_0, \tag{7}$$

$$R = R(\beta), \quad R' = R(\beta'), \tag{8}$$

R is the distance between the observation position and the source S and R' is the distance between the observation point and the image source S'.

The function $\eta(x)$ is the step function

$$\eta(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leqslant 0. \end{cases} \tag{9}$$

The potential $V^i(R)$ represents a wave incident from the source S (Fig. 2), whereas the potential $V^i(R')$ represents that from the source S', namely a wave reflected from the half-plane $\phi = 0$.

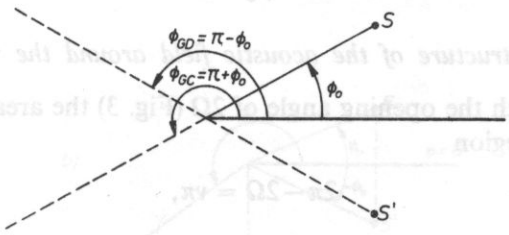


FIG. 2. The position of the shadow boundary $\phi_G = \phi_{GC} = \pi+\phi_0$ and that of the boundary of the reflected wave $\phi_G = \phi_{GD} = \pi-\phi_0$

The diffraction part of the potential can be written in the form of the sum of two diffraction waves

$$V^d = V^d[R(\beta)]+V^d[R(\beta')]. \tag{10}$$

The potential $V^d[R(\beta)]$ represents the wave generated as a result of the interaction of a wave from the source S with the edge of the half-plane, and the potential $V^d(R(\beta'))]$ – the wave generated by the interaction between a reflected wave and the edge, i.e. the interaction between the wave from the image source S' and the edge.

It follows from above that each of the sources, both S and S', are sources of geometrical and diffraction waves. Geometrical waves occur only in certain areas around the half-plane, diffraction waves are present throughout the space. The direct

wave $V^i(R)$ from the source S exists in the area

$$0 < \phi < \pi + \phi_0, \tag{11}$$

where

$$\pi + \phi_0 = \phi_{GC} \tag{12}$$

is the shadow boundary (Fig. 2). The region

$$\phi_{GC} < \phi < 2\pi \tag{13}$$

is the area of the geometrical shadow where there are no geometrical waves.

The reflected wave $V^i(R')$ coming from the source S' exists in the region

$$0 < \phi < \pi - \phi_0, \tag{14}$$

where

$$\pi - \phi_0 = \phi_{GD} \tag{15}$$

is the boundary of the reflected waves (Fig. 2).

In relation to the geometry of the system, the solutions obtained are symmetrized with respect to the half-plane $\phi = \pi$, hence, it is possible to limit the analysis of the field, assuming the source position to be within the interval

$$0 < \phi_0 < \pi. \tag{16}$$

1.2. Structure of the acoustic field around the wedge

For a wedge with the opening angle of 2Ω (Fig. 3) the area of the acoustic field description is the region

$$2\pi - 2\Omega = \nu\pi, \tag{17}$$

where

$$\nu > 1, \tag{18}$$

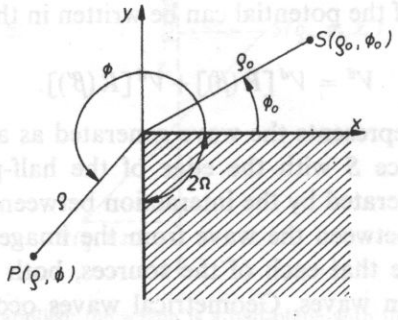


FIG. 3. The position of the right angle wedge in the coordinate system applied

i.e. the region outside the wedge. For wedges for which the parameter ν satisfies condition (18), the formulae given below are valid if the appropriate value of ν is put in plane of $\nu = 3/2$ represented a right angle wedge.

The acoustic field which would emerge as a result of the interaction between the acoustic wave and the ideal rigid right angle wedge, is described by the acoustic potential satisfying the Helmholtz equation (1), with the boundary conditions (2), which in this case must be satisfied on two half-planes forming the wedge (Fig. 3).

Just as for a single half-plance, this potential can be written in the form of sum (3) of the geometrical part V^g and the diffraction part V^d of the potential. The geometrical part V^g (Fig. 4) consists of three waves: the wave ($V^i(R)$) coming from the real source S, the wave ($V^i(R')$) reflected from the half-plane $\phi = 0$, coming from the source S', and the wave ($V^i(R'')$) reflected from the half-plane $\phi = 3\pi/2$ coming from

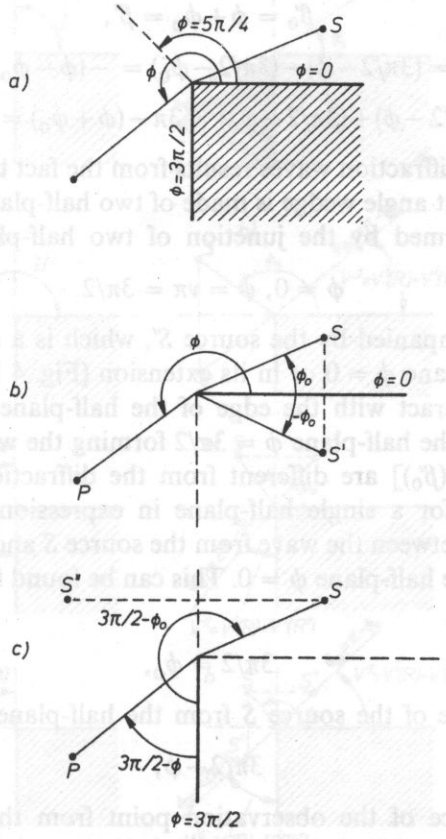


FIG. 4. The sources of the waves making up the acoustic field around the right angle wedge: a) the half-planes $\phi = 0$ and $\phi = 3\pi/2$ making up the wedge; b) the position of the source S' representing the wave $V^i(R')$ reflected from the half-plane $\phi = 0$; c) the position of the source S'' representing the wave $V^i(R'')$ reflected from the half-plane $\phi = 3\pi/2$

the source S''

$$V^g = \eta(\pi-\beta)\eta(3\pi/2-\phi)V^i(R)+\eta(\pi-\beta')\eta(3\pi/2-\phi)V^i(R') + \eta(\pi-\beta'')\eta(3\pi/2-\phi)V^i(R''), \quad (19)$$

$$\beta = \phi-\phi_0, \quad (20)$$

$$\beta' = \phi+\phi_0, \quad (21)$$

$$\beta'' = (3\pi/2-\phi)+(3\pi/2-\phi_0) = 3\pi-(\phi+\phi_0), \quad (22)$$

$$R = R(\beta), \quad R' = R(\beta'), \quad R'' = R(\beta''). \quad (23)$$

The diffraction part of the potential is the sum of the four waves

$$V^d = V^d[R(\beta_0)+V^d[R(\beta'_0)]+V^d[R(\beta_{3\pi/2})]+V^d[R(\beta'_{3\pi/2})], \quad (24)$$

$$\beta_0 = \phi-\phi_0 = \beta, \quad (25)$$

$$\beta'_0 = \phi+\phi_0 = \beta', \quad (26)$$

$$\beta_{3\pi/2} = (3\pi/2-\phi)-(3\pi/2-\phi_0) = -(\phi-\phi_0) = -\beta, \quad (27)$$

$$\beta'_{3\pi/2} = (3\pi/2-\phi)+(3\pi/2-\phi_0) = 3\pi-(\phi+\phi_0) = 3\pi-\beta' = \beta''. \quad (28)$$

The existence of four diffraction waves results from the fact that the wedge is made of two half-planes. A right angle wedge is made of two half-planes. A right angle can be recognized as one formed by the junction of two half-planes (Fig. 4):

$$\phi = 0, \phi = \nu\pi = 3\pi/2. \quad (29)$$

The source S is accompanied by the source S', which is a specular reflection of the source S in the half-plane $\phi = 0$ or in its extension (Fig. 4 b). The waves from these two sources S, S' interact with the edge of the half-plane $\phi = 0$. This interaction occurs in presence of the half-plane $\phi = 3\pi/2$ forming the wedge therefore the waves $V^d[R(\beta_0)]$ and $V^d[R(\beta'_0)]$ are different from the diffraction waves $V^d[R(\beta)]$ and $V^d[R(\beta')]$ occurring for a single half-plane in expression (10).

The interaction between the wave from the source S and the half-plane $\phi = 3\pi/2$ is similar to that of the half-plane $\phi = 0$. This can be found from expressions (27) and (28), where:

$$3\pi/2 = \phi_0, \quad (30)$$

is the angular distance of the source S from the half-plane $\phi = 3\pi/2$ (Fig. 4c), and

$$3\pi/2-\phi, \quad (31)$$

is the angular distance of the observation point from this plane.

Thus, in analogy to (25) and (26) for the half-plane $\phi = 0$, expressions (27) and (28) are the difference and the sum of the angular distances of the source and observation point from the half-plane $\phi = 3\pi/2$.

As a result of a specular reflection of the real source S in the half-plane $\phi = 3\pi/2$

there occurs the source S'' (this reflection may occur in the half-plane $\phi = 3\pi/2$ or its extension). The result of the interaction between the wave from the source S and the half-plane edge $\phi = 3\pi/2$ in the presence of the half plane $\phi = 0$ is the diffraction wave $V^d\ [R(\beta_{3\pi/2})]$. All the four diffraction waves (24) occur throughout the space around the wedge, on the other hand, geometrical waves only do so in certain areas determined by the source position ϕ_0. For this reason, it is convenient to distinguish four regions of the source position (Fig. 5).

If the source is in region I (Fig. 5 (I))

$$0 < \phi_0 < \pi/2, \tag{32}$$

the direct wave $V^i(R)$ occurs in the region

$$0 < \phi < \pi + \phi_0, \tag{33}$$

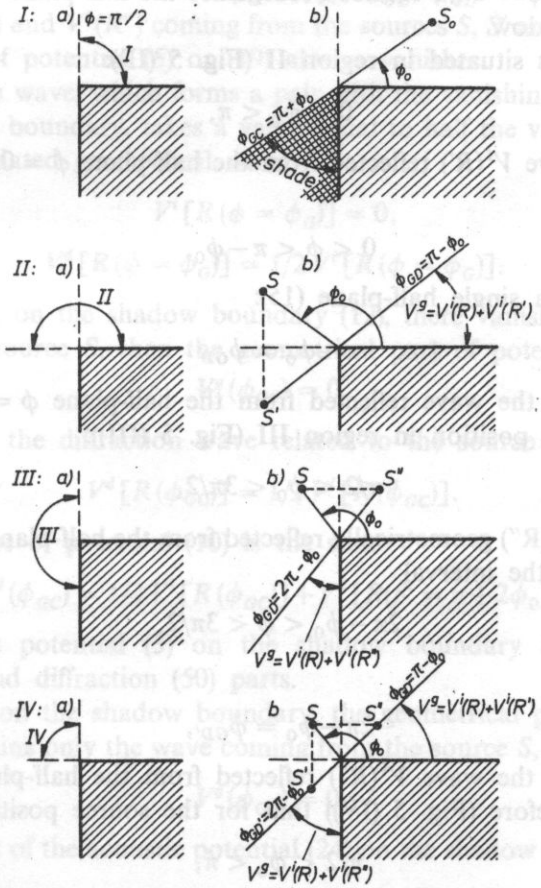


FIG. 5. The four regions of the source position distinguished here: I — $(0 < \phi_0 < \pi/2)$, II — $(0 < \phi_0 < \pi)$, III — $(\pi/2 = \phi_0 < 3\pi/2)$, IV — $(\pi/2 < \phi_0 < \pi)$. *a*) the region of the real source position, *b*) region of an image sources position corresponding to the position of a given real source and areas where particular geometrical waves occur

where, just as for a single half-plane (12)

$$\pi+\phi_0=\phi_{GC}, \tag{34}$$

is the shadow boundary. In the space

$$\phi_{GC}<\phi<3\pi/2 \tag{35}$$

the geometrical shadow occurs, where there are no geometrical waves. For the right angle wedge, this area is smaller than that for a single half-plane, since the part of the space

$$3\pi/2<\phi<2\pi \tag{36}$$

is occupied by the wedge itself.

If the source is situated in region I (32), the wave $V^i(R'')$ geometrically reflected from the half-plane $\phi=3\pi/2$ cannot occur, since the half-plane $\phi=3\pi/2$ is within the geometrical shadow.

If the source is situated in region II (Fig. 5 (II)):

$$0<\phi_0<\pi, \tag{37}$$

then occurs the wave $V^i(R')$ reflected from the half-plane $\phi=0$. It appears in the interval

$$0<\phi<\pi-\phi_0, \tag{38}$$

where, just as for a single half-plane (15):

$$\pi-\phi_0=\phi_{GD} \tag{39}$$

is the boundary of the wave reflected from the half-plane $\phi=0$.

For the source position in region III (Fig. 5 (III)):

$$\pi/2<\phi_0<3\pi/2, \tag{40}$$

there is the wave $V^i(R'')$ geometrically reflected from the half-plane $\phi=3\pi/2$, and its occurrence area is the interval

$$2\pi-\phi_0<\phi<3\pi/2, \tag{41}$$

where

$$2\pi-\phi_0=\phi_{GD'}, \tag{42}$$

is the boundary of the wave $V^i(R'')$ reflected from the half-plane $\phi=3\pi/2$.

It follows therefore (Fig. 5 (IV)) that for the source position in region IV:

$$\pi/2<\phi_0<\pi, \tag{43}$$

in the space around the wedge there is no shadow area, but there are two reflected waves: the wave $V^i(R')$ reflected from the half-plane $\phi=0$ (in region (38)) and the wave $V^i(R'')$ reflected from the half-plane $\phi=3\pi/2$ (in region (41)).

In view of the geometry of the system, the descriptions of the acoustic field are symmetrical with respect to the half-plane $\phi = 5\pi/4$ (Fig. 4a). Therefore, the complete analysis of the field requires to consider the source position only in the interval

$$0 < \phi_0 < 5\pi/4. \tag{45}$$

1.3. Acoustic field on geometrical boundaries

In the assumed coordinate system, the geometrical boundaries are half-planes with equations $\phi = \text{const}$. They are the boundary ϕ_{GC} of the shadow (12), (34) the boundary ϕ_{GD} of the wave reflected from the half-plane $\phi = 0$ (15), (39), and, in addition, for the wedge, the boundary ϕ_{GD} of the wave reflected from the half-plane $\phi = 3\pi/2$ (42).

On the geometrical boundaries $\phi_G = \phi_{GC}$, ϕ_{GD}, $\phi_{GD'}$ one of the geometrical waves $V^i(R)$, $V^i(R')$ and $V^i(R'')$ coming from the sources S, S' and S'' making up the geometrical part of potential (5) or (19) always vanishes.

The diffraction wave, which forms a pair with the vanishing geometrical wave, on the geometrical boundary, takes a value equal to half the value of the acoustic potential of the related geometrical wave,

$$V^i[R(\phi = \phi_G)] = 0, \tag{46}$$

$$V^d[R(\phi = \phi_G)] = 1/2\, V^i[R(\phi = \phi_G)]. \tag{47}$$

For the half-plane, on the shadow boundary (12), there vanishes the wave $V^i(R)$ coming from the source S, then the geometrical part of potential (5) becomes

$$V^g(\phi_{GC}) = 0. \tag{48}$$

At the same time, the diffraction wave related to the source S, takes the form

$$V^d[R(\phi_{GC})] = 1/2\, V^i[R(\phi_{GC})]. \tag{49}$$

The diffraction part of potential (10) is the sum

$$V^d(\phi_{GC}) = 1/2\, V^i[R(\phi_{GC})] + V^d[R(\beta' = \pi + 2\phi_0)]. \tag{50}$$

The total acoustic potential (3) on the shadow boundary is the sum of the geometrical (48) and diffraction (50) parts.

For a wedge, on the shadow boundary, the geometrical part of the acoustic potential (19) contains only the wave coming from the source S, which vanishes, i.e.

$$V^g(\phi_{GC}) = 0. \tag{51}$$

The diffraction part of the acoustic potential (24) on the shadow boundary takes the form

$$V^d(\phi_{GD}) = 1/2\, V^i[R(\phi_{GC})] + V^d[R(\beta'_0 = \pi + 2\phi_0)] + V^d[R(\beta_{3\pi/2} = -\pi)] + V^d[R(\beta'_{3\pi/2} = 2\pi - 2\phi_0)] \tag{52}$$

Just as for a single half-plane, the total acoustic potential on the shadow boundary consists only of the diffraction part (52).

For the case of a single half-plane on the boundary of the wave $V^i(R')$ reflected from the half-plane $\phi = 0$, the geometrical part of potential (5) takes the form:

$$V^g(\phi_{GD}) = V^i(R(\phi_{GD})]. \tag{53}$$

In turn, the diffraction part of potential (10) takes the form

$$V^d(\phi_{GD}) = V^d[R(\beta = \pi - 2\phi_0)] + 1/2\, V^i[R'(\phi_{GD})]. \tag{54}$$

The total acoustic potential on the boundary of the reflected wave is the sum of potentials (53) and (54).

For the case of a wedge on the boundary of the wave $V^i(R')$ reflected from the half-plane $\phi = 0$, the geometrical part of potential (19) takes the form of (53). In turn the expression of the diffraction part (24) is as follows:

$$V^d(\phi_{GD}) = V^d[R(\beta_0 = \pi - 2\phi_0)] + 1/2\, V^i[R'(\phi_{GD})] + V^d[R(\beta_{3\pi/2} = 2\phi_0 - \pi)] + V^d[R(\beta'_{3\pi/2} = 2\pi)]. \tag{55}$$

The total potential is the sum of potentials (53) and (55).

In addition, in the case of a wedge, there can occur the boundary (42) of the wave $V^i(R'')$ reflected from the half-plane $\phi = 3\pi/2$. The geometrical potential (19) takes the form

$$V^g(\phi_{GD'}) = V^i(R(\phi_{GD'})], \tag{56}$$

and the diffraction part (24) of the potential becomes

$$V^d(\phi_{GD'}) = V^d[R(\beta_0 = 2\pi - 2\phi_0)] + V^d[R(\beta'_0 = 2\pi)] + V^d[R(\beta_{3\pi/2} = 2\phi_0 - \pi)] + 1/2\, V^i[R''(\phi_{GD'})]. \tag{57}$$

The total field is the sum of potentials (56) and (57).

2. Explicit forms of acoustic potentials describing the acoustic field around the half-plane and right angle wedge for plane, cylindrical and spherical waves

The first step towards the description of the acoustic field shown in Sect. 1 is the division of the expression describing the potential into the geometrical and diffraction parts (3).

The division of an exact solution describing the acoustic potential around the half-plane into the geometrical (5) and diffraction (10) parts is only possible for the incident plane wave. In the case of cylindrical and spherical waves, this division can be peformed only for the approximate form of solution which is obtained as the condition is satisfied [2]

$$kR_1 \gg 1, \tag{58}$$

where R_1 is the shortest distance between the source and the observation point calculated through the edge.

R_1 can be treated as the parameter which takes a form which depends on the wave type. Considering the fact that the plane wave is the limiting case of a wave radiated by the source which is in infinity, in the assumed coordinate system the distance R_1 is

– for the plane wave

$$R_1 = \varrho, \tag{59}$$

– for the cylindrical wave

$$R_1 = \varrho_0 + \varrho, \tag{60}$$

– for the spherical wave

$$R_1 = [(\varrho_0+\varrho)^2+(z-z_0)^2]^{1/2}. \tag{61}$$

According to expressions (58) and (60), for the cylindrical wave, the division of the potential into the geometrical and diffraction parts is possible in the case of system in which the source or the observation point are far from the diffraction edge (with respect to the wavelength). For the spherical wave, according to expressions (58) and (61), it is possible, just as for the cylindrical wave, when the source or the observation point are far from the edge and also when the source and the observation point are distant enough along the z-axis ($k|z-z_0| \gg 1$).

In the case of interaction between the waves and the wedge the division of the acoustic potential into the geometrical (19) and diffraction (24) parts can be only made when the approximate form of an exact solution is used. The approximation conditions are following:

– for the plane wave

$$k\varrho \gg 1, \tag{62}$$

– for cylindrical and spherical waves

$$\frac{k\varrho_0\varrho}{R_1} \gg 1. \tag{63}$$

Since the source of the plane wave is in infinity, it can be said, in general, that if there is the interaction between three types of wave plane, cylindrical and spherical, and the wedge, the potential can be divided into the geometrical and diffraction parts only if the source and the observation point are at the same time far from the wedge.

It follows from the above that for all the six cases considered (two kinds of obstacle half-plane and wedge, for which each time there is the diffraction of three wave types plane, cylindrical and spherical) the common area where solutions can be divided into the geometrical and diffraction parts is the area which lies far from the diffraction edge. It would be applied only to systems in which the source is also far enough from the edge.

To move over from the symbolic-qualitative description of the acoustic field to

a quantitative one, we need to know explicit forms of functions describing the components of the geometrical parts of the potential (5), (19) and the diffraction parts (10), (24).

The geometrical parts of potentials are made up of waves radiated by sources whose positions are only different in the angular coordinate. Thus the distance $R(\alpha)$ between observation point and the particular source differs only in an angle α. Therefore, for every type of wave: plane, cylindrical, spherical propagation from these sources can be described by the same expression where the right quantity $R(\alpha)$ must be put. This will be presented in Sect. 2.1.

Then step leading to the description given in Sect. 1 is the analysis of simplified expressions describing the components of the diffraction part of the potentials. These simplified expressions are valid in areas appropriately distant from the geometrical boundaries tied up to successive sources. Moreover, it is possible to describe the diffraction, waves on the geometrical boundaries by means of simple expressions.

In the case of not too large area in which simplified expressions for diffraction waves are not valid on the basis of the field continuity, principle the field in this area can be determined by extrapolation from the value taken in the area where approximation is valid up to the value taken on the boundary itself. In this way, one obtains a description of the field throughout the space around the barrier.

The appropriate explicit forms of functions describing the diffraction waves in areas far from the geometrical boundaries and on the various boundaries are shown in points 2.2 and 2.3.

The applicability range of the description presented and its accuracy are discussed in subsection 2.4.

2.1. Explicit forms of the geometrical part of the acoustic potential

In the expression for the geometrical part of the potential for the half-plane (5) it is necessary to substitute the explicit forms of functions describing the waves

$$V^i(R(\alpha)], \tag{64}$$

where

$$\alpha = \beta, \beta'. \tag{65}$$

$R(\alpha)$ is successively the distance of the observation point from the sources S and S'. In the expression for the geometrical part of the potential for the wedge (19) it is necessary to substitute the explicity forms of function (64) for

$$\alpha = \beta, \beta', \beta''. \tag{66}$$

where $R(\alpha)$ is, successively, the distance of the observation point from the sources S, S' and S''.

The explicit forms of function (64) for chosen types of wave, in the assumed coordinate system, are as follows

– for a plane wave:

$$V^i[R(\alpha)] = \exp[ikR(\alpha)], \tag{67}$$

$$R(\alpha) = -\varrho\cos\alpha, \tag{68}$$

– for a cylindrical wave:

$$V^i[R(\alpha)] = \sqrt{\frac{2}{kR(\alpha)}}\exp\{i[kR(\alpha)-\pi/4\}, \tag{69}$$

$$R(\alpha) = [\varrho_0^2+\varrho^2-2\varrho_0\varrho\cos\alpha]^{1/2}, \tag{70}$$

– for a spherical wave:

$$V^i[R(\alpha)] = \frac{\exp[ikR(\alpha)]}{kR(\alpha)}, \tag{71}$$

$$R(\alpha) = [\varrho_0^2+\varrho^2-2\varrho_0\varrho\cos\alpha+(z-z_0)^2]^{1/2}. \tag{72}$$

2.2 Explicit forms of the diffraction part of the acoustic potential

From well-known expressions [1, 2] for the diffraction part of the acoustic potential for chosen kinds of obstacles (a half-plane and right-angle wedge) and three types of waves plane, cylindrical and spherical, the diffraction waves related to sources distant by $R(\alpha)$ from the observation point can be represented in the form

$$V^d[R(\alpha)] = P(\alpha,\nu)\frac{\exp[i\{k[R_1-R(\alpha)]+\pi/4\}]}{\sqrt{2\pi k\varrho}}\times d[R(\alpha)]\,V^i[R(\alpha)], \tag{73}$$

$$P(\alpha,\nu) = \frac{1}{2\nu}\sin(\pi/\nu)\frac{1}{\cos(\pi/\nu)-\cos(\alpha/\nu)}. \tag{74}$$

where $d[R(\alpha)]$ is a coefficient depending on the type of the wave. Formulae (73) and (74) are valid if conditions (58), (62) and (63), and the inequality

$$k[R_1-R(\alpha)] \gg 1. \tag{75}$$

are satisfied. Condition (75) means that the observation point must be far from the geometrical boundaries: the shadow boundary ϕ_{GC} (12), (34), the boundary ϕ_{GD} of the wave reflected from the half-plane $\phi = 0$ (15), (39), and the boundary $\phi_{GD'}$, of the wave reflected from the half-plane $\phi = 3\pi/2$ (42).

Expressions (73) and (74) are valid for wedges with an opening angle satisfying condition (18). A single half-plane is a special case of a wedge with the opening angle

$$2\Omega = 0. \tag{76}$$

Hence, from expressions (17), for the half-plane

$$\nu = 2. \tag{77}$$

In the case of the half-plane, expression (73) must be applied to determine the two

diffraction waves $V^d[R(\alpha)]$ which are part of the diffraction part of potential (10) for

$$\alpha = \beta, \beta'. \tag{78}$$

Then the diffraction part of the acoustic potential can be rewritten in relation to the incident wave $V^i(R)$ coming from the real source S:

$$V^d = P(\phi, \phi_o, \nu = 2)\frac{\exp\{i[k(R_1 - R) + \pi/4]\}}{\sqrt{2\pi k\varrho}} d(R)\, V^i(R), \tag{79}$$

$$P(\phi, \phi_0, \nu = 2) = -\frac{1}{2}\left[\frac{1}{\cos\dfrac{\phi - \phi_0}{2}} + \frac{1}{\cos\dfrac{\phi + \phi_0}{2}}\right]. \tag{80}$$

Depending on the incident wave type, the coefficient $d(R)$ takes the forms:

– for a plane wave

$$d(R) = 1, \tag{81}$$

– for a cylindrical wave

$$d(R) = \sqrt{\frac{R}{\varrho_0}}, \tag{82}$$

– for a spherical wave:

$$d(R) = \frac{R}{\sqrt{R_1 \varrho_0}}. \tag{83}$$

From formulae (59), (67), (68) and (81) the explicit form of the diffraction part of the acoustic potential for a plane wave is as follows:

$$V^d = P(\phi, \phi_0, \nu = 2)\frac{\exp[i(k\varrho + \pi/4)]}{\sqrt{2\pi k\varrho}}. \tag{84}$$

Hence, it follows that the diffraction part of the acoustic potential for a plane wave interacting with half-plane can be regarded as a wave of a cylindrical type generated by the edge. The amplitude of this wave decreases as a root of the distance from the edge ϱ. It differs from the cylindrical wave only by the directional coefficient $P(\phi, \phi_0, \nu = 2)$. In a superposition with the geometrical part of the acoustic potential (5), this wave gives the total acoustic field which appears as a result of the interaction between the plane wave and the half-plane.

On the ground of the reciprocity theorem it is possible to interchange the source and the observation point positions. Thus, for cylindrical and spherical waves, according to (58) it can be set that it is the source which is far from the edge. Then the coefficients $d(R)$, (82) and (83), can be interpreted as measures of the relative curvature of the wave fronts interacting with the edge. The emerging diffraction

waves are cylindrical-type waves deformed by the coefficient $d(R)$, with the directional coefficients $P(\phi, \phi_0, \nu = 2)$ (80), generated by the diffraction edge.

For the wedge, the diffraction part of the acoustic potential (24) contains four diffraction waves $V^d[R(\alpha)]$. To calculate them, it is necessary to use expression (73) for

$$\alpha = \beta_0, \beta'_0, \beta_{3\pi/2}, \beta'_{3\pi/2}. \tag{85}$$

Moreover, inequality (75) must be satisfied for all $R(\alpha)$.

Using the relations between the angle α occurring for the wedge, (25)–(28), the diffraction part of the acoustic potential (24) can be rewritten in relation to the wave $V^i(R)$ coming from the real source S:

$$V^d = P(\phi, \phi_0, \nu = 3/2) \frac{\exp\{i[k(R_1 - R) + \pi/4]\}}{\sqrt{2\pi k\varrho}} d(R)\, V^i(R), \tag{86}$$

$$P(\phi, \phi_0, \nu = 3/2) = \frac{2}{3}\sin(2\pi/3) \left[\frac{1}{\cos\frac{2\pi}{3} - \cos\frac{2(\phi - \phi_0)}{3}} + \frac{1}{\cos\frac{2\pi}{3} - \cos\frac{2(\phi + \phi_0)}{3}} \right]. \tag{87}$$

It follows hence that if the waves of the three chosen types interact with the wedge, the diffraction part of the acoustic potential (86) differs from the diffraction part for half-plane (79) just in the directional coefficient $P(\phi, \phi_0, \nu)$. For the half-plane it has the form of (80), for the wedge that of (87). The other conclusions concerned with the interpretation of the diffraction part of the acoustic potential remain valid. For the six analyzed cases of wave interaction with obstacles the fulfilling of conditions (58), (62), (63) and (75) gives the diffraction part of acoustical potential the same as that derived on the ground of geometrical diffraction theory.

2.3. Explicit forms of the acoustic potential on geometric boundaries

The geometrical parts of the acoustic potential $V^g(\phi_G)$ on the geometrical boundaries for the three chosen types of waves and two kinds of obstacles can be obtained by substituting the appropriate forms of geometrical waves (64) in expressions (48), (51), (53) and (56).

In the expressions describing the diffraction parts of the acoustic potential $V^d(\phi_G)$ on the geometrical boundaries, one of the components of the diffraction waves is always determined by the explicit form of the potential of the geometrical wave which vanishes on a given boundary (46), (47). If condition (75) is satisfied, the other diffraction waves occurring in expressions (50) and (54) can be determined from formula (73).

E.g. on the shadow-boundary, for the half-plane the total acoustic potential is equal to the diffraction part $V^d(\phi_{GD})$ (50). In expression (50), the first term is obtained, depending on the wave type, from one of expressions (67)–(72). In turn, the other term can be determined from (73) only if inequality (75) is satisfied for it.

In the case of the wedge, in expressions (52), (55) and (57) describing the diffraction parts of the potential on the geometrical boundaries, there are components related to the vanishing geometrical wave (40), (47) and those which, after satisfying condition (75), can be determined by expression (73). Moreover, there are also terms of the form

$$V^d[R(\alpha = n\pi, \nu)], \tag{88}$$

$$n = 0, \pm 1, \pm 2, \ldots, \tag{89}$$

which cannot be described by means of expressions (73) and (74) because the singularity occurs in expression (74) for $\alpha = n\pi$. In this case, the directional coefficients in the form

$$P(\alpha = n\pi, \nu), \tag{90}$$

should be replaced by their boundary values which can be obtained from the exact solution of the Helmholtz equation for required boundary conditions [1]

$$P(\alpha = n\pi, \nu) = -(1/2\nu)\operatorname{ctg}(\pi/\nu). \tag{91}$$

For the half-plane:

$$P(\alpha = n\pi, \nu = 2) = 0 \tag{92}$$

i.e., according to expression (50), this problem does not exist. In turn for the right angle wedge,

$$P(\alpha = n\pi, \nu = 3/2) = -(1/3)\operatorname{ctg}(2\pi/3). \tag{93}$$

This makes possible to use formula (52), (55) and (57) for determining the diffraction parts of the potential $V^d(\phi_G)$ on the geometrical boundaries wherever condition (75) is satisfied for components of diffraction waves without singularities.

2.4. Accuracy of the approximate formulae applied – ranges of applicability

The division of an exact solution describing the acoustic potential around the half-plane into the geometrical and diffraction part is only possible for the incident plane wave [2]. The geometrical part of the potential is described by expansion (5). In turn, the diffraction part in the exact solution consists of two diffraction waves, which, in keeping with the notation adapted here, have the form

$$V^d[R(\alpha)] = -\operatorname{sgn}(\pi - \alpha) f(w)\, a[R_1, R(\alpha)]\, V^i[R(\alpha)], \tag{94}$$

where for the plane wave interacting with the half-plane

$$\alpha = \beta, \beta', \tag{95}$$

$$a[R_1, R(\alpha)] = 1/2. \tag{96}$$

The function

$$f(w) = \frac{F(w)}{F(0)}, \tag{97}$$

represents the reduced Fresnel integral

$$F(w) = \int_{w}^{\infty} \exp(i\mu^2)\, d\mu, \tag{98}$$

where

$$w = \sqrt{k[R_1 - R(\alpha)]}, \tag{99}$$

is the square root, calculated in wavelengths, of the difference in the paths passed by the waves to the observation point directly from the source ($R(\alpha)$) and through the edge (R_1).

For the cylindrical and spherical waves interacting with the half-plane, the solution in the form of the sum of the geometrical part of the acoustic potential (5) and the diffraction part (10) whose component diffraction waves are given by expression (94) is the approximate form of the solution which is valid if condition (58) is satisfied.

– for cylindrical wave:

$$a[R_1, R(\alpha)] = \sqrt{\frac{R(\alpha)}{2[R_1 + R(\alpha)]}}, \tag{100}$$

– for spherical wave

$$a[R_1, R(\alpha)] = \sqrt{\frac{R^2(\alpha)}{2R_1[R_1 + R(\alpha)]}}. \tag{101}$$

On the geometrical boundaries (12) and (15) $\phi_G = \phi_{GC}, \phi_{GD}$ for diffraction waves related to the vanishing geometrical waves for all the three types of waves, one obtains the equalities:

$$R(\phi = \phi_G) = R_1, w[R_1, R(\phi = \phi_G)] = 0. \tag{102}$$

Hence, after substituting expressions (102) successively in formulae (100), (97) and (94) it can be seen that equality (47) is valid on the geometrical boundaries.

If condition (75) is satisfied,

$$w[R_1, R(\alpha)] = \sqrt{k[R_1 - R(\alpha)]} \gg 1, \tag{103}$$

the reduced Fresnel integral (97) can be replaced by the first term of a series expansion:

$$f(w) = \frac{\exp[i(w + \pi/4)]}{\sqrt{\pi}\, w}\left[1 + \sum_{n=1}^{\infty} \frac{\prod_{j=1}^{n}(2j-1)}{(2iw^2)^n}\right]. \tag{104}$$

Then, the expression for the diffraction part of the acoustic potential is obtained in the form of (79). This can be done for

$$|w| \geqslant 3 \tag{105}$$

when the error is of the order of the absolute value of the first omitted term, i.e. $(2\sqrt{\pi}\, w^3)^{-1} \sim 0.01$.

For a plane wave interacting with a half-plane, the angular width Δ can be determined for the regions around the geometrical boundaries ϕ_G for $|w| = 3$ inside which expression (79) is not valid

$$\phi_G - \Delta < \phi < \phi_G + \Delta, \tag{106}$$

$$\Delta = 2\arcsin\left(\sqrt{\frac{4.5}{k\varrho}}\right), \tag{107}$$

$$k\varrho \geqslant 4.5. \tag{108}$$

For cylindrical and spherical waves, if condition (58) is initially satisfied, the angular width Δ of the regions around the geometrical boundaries (for $|w| = 3$) where formula (79) is not valid, is

$$\Delta = 2\arcsin\left(\sqrt{\frac{4.5\,(kR_1 - 4.5}{k\varrho_0\, k\varrho}}\right), \tag{109}$$

$$kR_1 \geqslant 4.5. \tag{110}$$

For plane and cylindrical waves, and also a spherical one (for $z = z_0$), Table 1 shows the angular widths Δ of the regions, around the geometrical boundaries, inside which expression (79) cannot be applied for preset parameters $k\varrho$ and $k\varrho_0$. Also, Table 1 shows the approximate values of the absolute distances of the observation point and the source from the edge for two chosen frequencies $f = 500$ and 1000 Hz as typical of noise in urban area.

It can be said in general (Table 1) that for decreasing values of the parameters $k\varrho$ and $k\varrho_0$, the region in which the approximate expression (79) is valid becomes narrower. Thus for small parameters $k\varrho$ and $k\varrho_0$ effective use of the approximate expressions is impossible.

In the case of the wedge, for the exact solution to be divided into the geometrical and diffraction parts, assumptions (62) and (63) must be satisfied. At the same time if condition (75) is fulfilled the diffraction waves can be described by formula (73). It is more difficult to determine the angular width of the regions where these conditions are not met as in the case of a single half-plane. On the other hand, it is possible to estimate them as being of the same order as those in the case of interaction between the chosen wave types and the half-plane.

Table 1. The angular width of regions Δ (107), (109), around the geometrical boundaries, inside which for given parameters $k\varrho$ and $k\varrho_0$ expression (79) is invalid, and the distances ϱ, ϱ_0 for two frequencies f

$k\varrho$	$k\varrho_0$	Δ [°]	ϱ [m]	ϱ_0 [m]	ϱ [m]	ϱ_0 [m]
			$f = 500$ [Hz]	$f = 500$ [Hz]	$f = 1000$ [Hz]	$f = 1000$ [Hz]
1000	∞*	8	100	∞	50	∞
1000	1000	3	100	100	50	50
500	∞	11	50	∞	25	
500	1000	14	50	100	25	50
500	500	15	50	50	25	25
250	∞	15	25	∞	12.5	∞
250	1000	17	25	100	12.5	50
250	500	19	25	50	12.5	25
250	250	22	25	25	12.5	12.5
125	∞	22	12.5	∞	6.3	∞
50	∞	34	5	∞	2.5	∞
25	∞	50	2.5	∞	1.3	∞
10	∞	84	1	∞	0.5	∞

* denotes a plane wave

3. Efficiency of the half-plane and wedge as obstacles disturbing the acoustic field

Comparative analysis of the interaction between the three chosen wave types and a half-plane and a wedge can be carried on only for the systems in which the source and the observation point are far from the diffraction edges. Exactly, such situations are met in urban systems. In these systems, if the observation point is far enough (75) from the geometrical boundaries, the diffraction part of the acoustic potential can be determined for a half-plane from expression (79), for a wedge from expression (86).

Knowing the geometrical parts of potentials (5), (19) and the diffraction parts (10), (24), it is possible to determine the acoustic field on the geometrical boundaries. The diffraction wave related to the vanishing geometrical wave takes the value according to (47). For the wedge, the diffraction wave having directional coefficient with a singularity takes the value according to (93). The other diffraction waves, which are not related to the disappearing geometrical wave, can be determined according to (79) if condition (75) is satisfied.

If the widths Δ of the regions around the geometrical boundaries (Table 1) are not too large, in keeping with the principle of continuity of the acoustic field, extrapolation can be carried out between the value of the acoustic potential on the geometrical boundary and the last of the values calculated according to the

approximation conditions. In this way, a description of the field throughout the region around the obstacle is obtained. To determine quantitatively the disturbance caused in the free field by the presence of a half-plane or a wedge, the obstacle efficiency is introduced:

$$IL = -20\log\frac{|V|}{|V^i|}, \tag{111}$$

where V is the total acoustic potential of the field disturbed by the obstacle and V^i is the acoustic potential of the free field (of the incident wave). With such a definition of the efficiency of an obstacle, the fact that it takes positive values at the observation point means a decrease in the sound pressure level caused by the presence of the obstacle; the fact that it takes negative values represents an increase in the sound pressure level caused by the presence of the obstacle.

The obstacle efficiency in the case of a half-plane ($\nu = 2$) and a wedge ($\nu = 3/2$), regarded as ideally rigid, is a function of the position of the observation point (ϕ, k_0, z) and that of the source (ϕ_0, k_0, z_0) with respect to the diffraction edge,

$$IL = IL(\phi, \phi_0, k\varrho, k\varrho_0, z, z_0, \nu). \tag{112}$$

It follows from the considerations in subsection 1.2 that the dependence of the obstacle efficiency on particular parameters can be determined analytically only in the shadow area:

$$\phi_{GC} < \phi < \nu\pi. \tag{113}$$

In this area, the acoustic potential contains only the diffraction part which can be represented in the form of one diffraction wave (79) for the half-plane, (86) for the wedge, related to the phase and amplitude of the incident wave ($V^i(R)$).

In the regions where there is interference between geometrical and diffraction waves only numerical analyse is possible. The efficiency is calculated as a function of one parameter with the other fixed.

3.1. Efficiency of the half-plane and wedge in the shadow area

The shadow area occurs at the same time in the case of both obstacles a half-plane ($\nu = 2$) and a wedge ($\nu = 3/2$) if the source is situated in the angular interval $(0, \pi/2)$ (Fig. 5 (II)). Then, the obstacle efficiency according to (79) and (86) has the form

$$IL = 20\log\frac{\sqrt{2\pi k\varrho}}{|P(\phi, \phi_0, \nu)|} - 20\log[d(R)]. \tag{114}$$

Taking into acount expressions (81)–(83), the obstacle efficiency for the three chosen wave types can be given by:

– for a plane wave

$$IL_p = 10\log(2\pi k\varrho) - 20\log|P(\phi, \phi_0, \nu)|, \tag{115}$$

– for a cylindrical wave

$$IL_c = IL_p - 10\log(R/\varrho_0), \tag{116}$$

– for a spherical wave

$$IL_s = IL_p - 10\log(R^2/R_1\varrho_0). \tag{117}$$

It follows from formula (115) that the obstacle efficiency for a plane wave in the shadow area is the greater the farther from the diffraction edge the observation points and the farther it is from the shadow boundary. In the first case the value of the parameter $k\varrho$ defining the relative position of the observation point (calculated in wave lenghts) is large. In the second case the absolute value of the directional coefficient $P(\phi, \phi_0, \nu)$ [5] is small. The fact that the obstacle efficiency in the shadow area for cylindrical and spherical waves is greater or smaller than the efficiency for a plane wave depends on whether the quotients in the second terms of formulae (116) and (117) are greater or smaller than unity. Moreover, for each of the three wave types the inequality occurs:

$$IL_j(\phi, \phi_0, k\varrho, k\varrho_0, k|z-z_0|, \nu = 2) - IL_j(\phi, \phi_0, k\varrho, k\varrho_0, k|z-z_0|, \nu = 3/2) > 0, \quad j = p, c, s, \tag{118}$$

meaning that for the same positions of the source and the observation point the efficiency of a half-plane in the shadow area is always greater than that of a right angle wedge. At the same time, the shadow area for the half-plane is always greater than that for the wedge, since part of the shadow area which occurs in the case of the half-plane is occupied by the wedge itself.

3.2. *Numerical examples*

The calculations were made for two types of obstacle, a half-plane ($\nu = 2$) and a right angle wedge ($\nu = 3/2$). Three types of incident waves were assumed: plane, cylindrical and spherical (for $z = z_0$). For cylindrical and spherical waves, a symmetrical system ($k\varrho, k\varrho_0 = 250, 500, 1000$) and nonsymmetrical one ($k\varrho \neq k\varrho_0$) were taken, with $k\varrho$ and $k\varrho_0$ occurring in three combinations of values used in the symmetrical system. Four positions of the source were chosen, belonging to successive regions distinguished in Fig. 5:

$\phi_0 = 10°$ with the source in region I (32),

$\phi_0 = 55°$ with the source in region I (32),

$\phi_0 = 90°$ with the source on the boundary between region I (32) and region II (37),

$\phi_0 = 135°$ with the source in region III (40) and, at the same time, in region IV (43).

Table 2 lists the parameters for which the calculations were made.

The positions of the source in region I:

$$\phi_0 = 10°, 55°$$

Table 2. The values of parameters and wave types for which the obstacle efficiences were calculated

angular source position	obstacle kind	wave type	$k\varrho_0$	$k\varrho$	$k\varrho/k\varrho_0 = m$
$\phi_0 = 10°, 55°,$ $90°, 135°$	$\nu = 2, 3/2$	plane	∞	250	0
			∞	500	0
			∞	1000	0
		cylindrical	250	250	1
			500	500	1
			1000	1000	1
			1000	250	1/4
			1000	500	1/2
			500	250	1/2
		spherical	250	250	1
			500	500	1
			1000	1000	1
			1000	250	1/4
			1000	500	1/2
			500	250	1/2

are related to the presence of the shadow area for both of the obstacle. With the position of the source:

$$\phi_0 = 90°,$$

the shadow area occurs only for the half-plane and it disappears for the wedge. The position of the source:

$$\phi_0 = 135°$$

causes the appearance of the shadow for the half-plane. In the case of the wedge, there is no shadow, on the other hand, there is a wave reflected from the half-plane $\phi = 3\pi/2 = 270°$. Table 3 shows the appropriate geometrical boundaries ϕ_{GC}, ϕ_{GD}, $\phi_{GD'}$, and in view of this, it is possible to distinguish four areas (Fig. 6):

– area A:

$$0 < \phi < \phi_{GD} \tag{119}$$

(which always occurs for the numerical examples)

– area B:

$$\phi_{GD} < \phi < \begin{cases} \phi_{GC} & \text{(for all the positions of the source in the case of the half-plane and for the wedge for the source position } \phi_0 = 10°, 55°, 90°\text{),} \\ \phi_{GD'} & \text{(for the wedge, for } \phi_0 = 135°\text{),} \end{cases} \tag{120}$$

Table 3. The positions of the geometrical boundaries for the numerical examples for the given angular position of the source ϕ_0

ϕ_0 [°]	ν	shadow boundary $\phi_{GC} = \pi + \phi_0$ [°]	boundary of wave reflected from half-plane $\phi = 0$ $\phi_{GD} = \pi - \phi_0$ [°]	boundary of wave reflected from half-plane $\phi = 3\pi/2 = 270°$ $\phi_{GD'} = 2\pi - \phi_0$ [°]
10	2	190	170	–
	3/2	190	170	–
55	2	235	125	–
	3/2	235	125	–
90	2	270	90	–
	3/2	–	90	270
135	2	315	45	–
	3/2	–	45	225

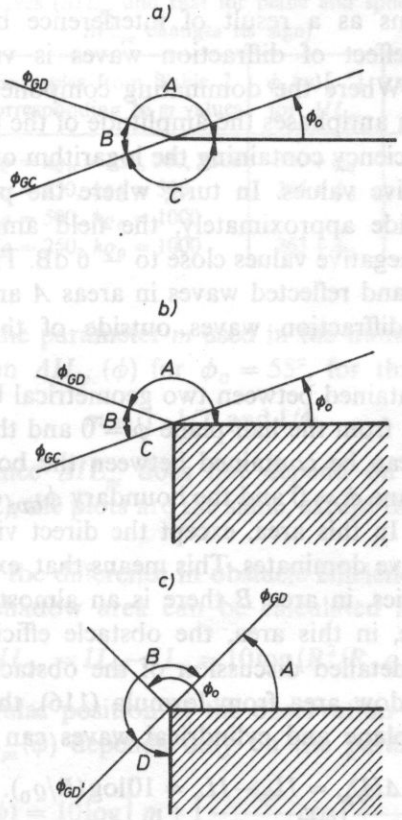


FIG. 6. Areas *A*, *B*, *C*, and *D* around the half-plane (*a*) and around the wedge (*b*), (*c*) distinguished in terms of the character of the function IL (ϕ) describing the efficiency of the obstacle

– area C:

$$\phi_{GC} < \phi < \begin{cases} 360^\circ & \text{(for the half-plane),} \\ 270^\circ & \text{(for the wedge),} \end{cases} \tag{121}$$

– area D (which occurs only for the wedge for $\phi_0 = 135^\circ$)

$$\phi_{GD} < \phi < 270^\circ. \tag{122}$$

In the numerical examples large values of the parameters $k\varrho$ and $k\varrho_0$ describing the relative positions of the observation point and the source we assumed, therefore outside of the direct vicinity of the geometrical boundaries the amplitudes of the diffraction waves are small [5]. This explains the structure of the acoustic fields in the distinguished areas A, B, C and D and the resultant shape of the function $IL(\phi)$ describing the obstacle efficiency in these areas.

Area C – the shadow area – is one where only diffraction waves occur. In this area the obstacle efficiency is positive and increases while deepening into it.

In areas A and D the acoustic field has the same structure since in the areas the direct wave ($V^i(R)$) and one of the reflected waves ($V^i(R')$ or $V^i(R'')$) occur as the dominating component waves.

The total field forms as a result of interference between geometrical and diffraction waves. The effect of diffraction waves is visible only close to the geometrical boundaries. Where the dominating component waves (the direct and reflected ones) interfere in antiphases the amplitude of the total field is close to zero, therefore the obstacle efficiency containing the logarithm of the field amplitude (111) can take the large positive values. In turn, where the phases of the dominating component waves coincide approximately, the field amplitude doubles and the obstacle efficiency takes negative values close to – 6 dB. The latter fact confirms the dominance of the direct and reflected waves in areas A and D, and simultaneously the small influence of diffraction waves outside of the direct vicinity of the geometrical boundaries.

Area B is the one contained between two geometrical boundaries: the boundary ϕ_{GD} of the wave reflected from the half-plane $\phi = 0$ and the shadow boundary ϕ_{GC}. For the wedge, area B can be contained between the boundary ϕ_{GD} of the wave reflected from the half-plane $\phi = 0$ and the boundary $\phi_{GD'}$ of the wave reflected from the half-plane $\phi = 270^\circ$. In this area, except the direct vicinity of the geometrical boundaries, the direct wave dominates. This means that, except the direct vicinity of the geometrical boundaries, in area B there is an almost undisturbed field of the incident wave. Therefore, in this area, the obstacle efficiency is equal to zero.

Passing to a more detailed discussion of the obstacle efficiency in the areas distinguished, in the shadow area from formula (116), the difference between the obstacle efficiencies for plane and cylindrical waves can be calculated:

$$\Delta IL_{pc} = IL_p - IL_c = 10\log(R/\varrho_0). \tag{123}$$

For a given angular position of the source (ϕ_0) the difference between the obstacle

efficiencies for plane and cylindrical waves, as a function of the angular position of the observation point (ϕ), depends only on the parameter

$$m = k\varrho/k\varrho_0, \tag{124}$$

then

$$\Delta IL_{pc}(\phi) = 5\log[1+m^2-2m\cos(\phi-\phi_0)]. \tag{125}$$

As a function of the angle ϕ, $\Delta IL_{pc}(\phi)$ is at first positive, then deeper into the shadow area it takes negative values. This means that deeper into the shadow the obstacle efficiency for cylindrical wave becomes greater than that for a plane wave:

$$\begin{aligned} &\Delta IL_{pc}(\phi) > 0 \quad \text{for} \quad \phi_{GC} < \phi < \phi_k(m), \\ &\Delta IL_{pc}(\phi) < 0 \quad \text{for} \quad \phi > \phi_k(m). \end{aligned} \tag{126}$$

Table 4 gives the approximate values of the angles $\phi_k(m)$ for which ΔIL_{pc} changes its

Table 4. The values of the angles $\phi_k(m)$ for which in the shadow area the difference between the obstacle efficiency for plane and cylindrical waves (ΔIL_{pc} and that for plane and spherical waves ΔIL_{ps} changes its sign)

m	Examples from Table 2 corresponding to m values	$\phi_k(m)$[°] for ΔIL_{pc}	$\phi_k(m)$ [°] for ΔIL_{ps}
1	$k\varrho = k\varrho_0 = 250, 500, 1000$	$300+\phi_0$	$270+\phi_0$
1/2	$k\varrho = 250$, $k\varrho_0 = 500$ $k\varrho = 500$, $k\varrho_0 = 1000$	$285+\phi_0$	$254+\phi_0$
1/4	$k\varrho = 250$, $k\varrho_0 = 1000$	$263+\phi_0$	$248+\phi_0$

sign, for three values of the parameter m used in the numerical examples (Table 2). The plots of the function $\Delta IL_{pc}(\phi)$ for $\phi_0 = 55°$, for three values of m:

$$m = 1,\ 1/2 \text{ and } 1/4,$$

are shown in Fig. 7. Since ΔIL_{pc} does not depend on ν for the two obstacles considered ($\nu = 2$, $\nu = 3/2$), the plots are the same, except that for the wedge they end for $\phi = 270°$.

From formula (117), the difference in obstacle efficiency between the plane and spherical waves in the shadow area can be calculated from the formula

$$\Delta IL_{ps} = IL_p - IL_s = 10\log(R^2/R_1\varrho_0). \tag{127}$$

As a function of the angular position of the observation point ϕ, for $z = 20$, with given values of ϕ_0, $\Delta IL_{ps}(\phi)$ depends only on the parameter m:

$$\Delta IL_{ps}(\phi) = 10\log\left[m+1-\frac{4m}{m+1}\cos\left(\frac{\phi-\phi_0}{2}\right)^2\right]. \tag{128}$$

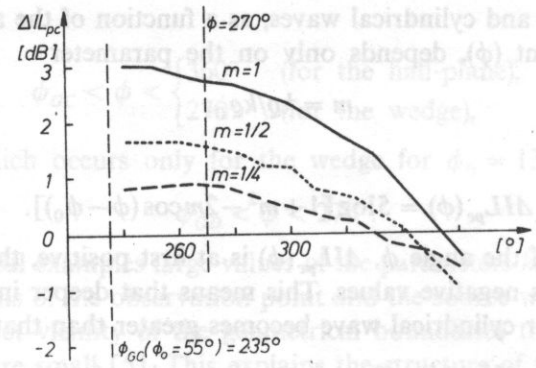


FIG. 7. The difference $\Delta IL_{pc}(\phi)$ 125 between the obstacle efficiencies for plane and cylindrical waves for different values of the parameter $m = 1,\ 1/2$ and $1/4$

The difference in the obstacle efficiency between the plane and spherical waves in the shadow area changes its sign for the angles $\phi_k(m)$ (Table 4):

$$\begin{aligned} \Delta IL_{ps}(\phi) &> 0 \quad \text{for} \quad \phi_{GC} < \phi < \phi_k(m), \\ \Delta IL_{ps}(\phi) &< 0 \quad \text{for} \quad \phi > \phi_k(m). \end{aligned} \tag{129}$$

The plots of $\Delta IL_{pk}(\phi)$ for $\phi_0 = 55°$ are shown in Fig. 8.

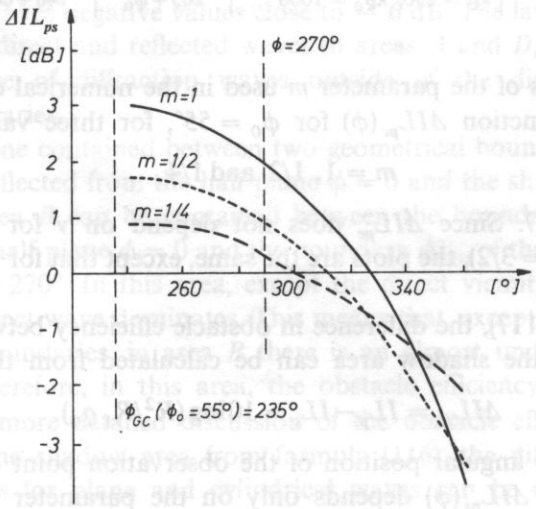


FIG. 8. The difference $\Delta IL_{ps}(\phi)$ (128) between the obstacle efficiencies for plane and spherical waves for different values of the parameter $m = 1,\ 1/2$ and $1/4$

From formula (118) the efficiency of the half-plane in the shadow area is always greater than that of the right angle wedge:

$$\Delta IL_v = IL(v=2) - IL(v=3/2) = 20\log\left[\frac{|P(\phi,\phi_0,v=3/2)|}{|P(\phi,\phi_0,v=2)|}\right] > 0. \qquad (130)$$

The plot of difference (130) as a function of the angular position of the observation point ϕ for $\phi_0 = 10°$, 55°, is shown in Fig. 9.

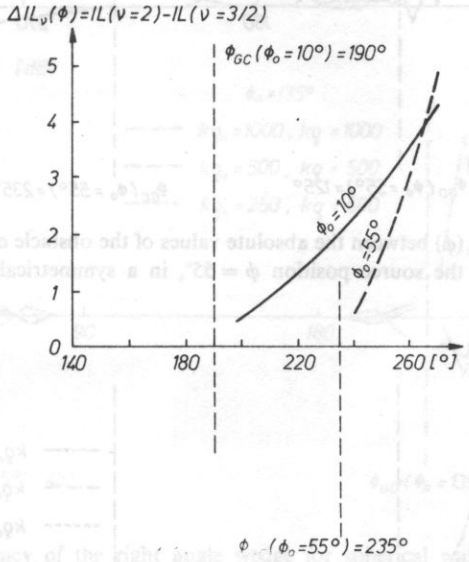


FIG. 9. The difference $\Delta IL\,(\phi)$ (130) between the efficiencies of the half-plane and the right angle wedge in the shadow area for two source positions $\phi_0 = 10°$, 55°

Analyzing the obstacle efficiencies in areas A and D for the chosen range of parameters (Table 2), it can be said that for a given wave type, the same position of the source (ϕ_0, $k\varrho_0$) and the observation point (ϕ, $k\varrho$) the efficiency of the half-plane and the right angle wedge does not show differences with an accuracy up to hundreths of a decibel:

$$IL_j(v=2) \cong IL_j(v=3/2), \quad j = p,c,s. \qquad (131)$$

It follows from comparison of the efficiency for the cylindrical wave IL_c and that of the spherical wave IL_s in areas A and D, that for the same positions of the source and the observation point the positions of the maxima and minima of the efficiency are the same. On the other hand the absolute values of the efficiency $|IL_c|$ for the cylindrical wave are almost always greater than those of the efficiency $|IL_s|$ for the spherical wave:

$$\Delta|IL|_{cs} = |IL_c| - |IL_s| > 0. \qquad (132)$$

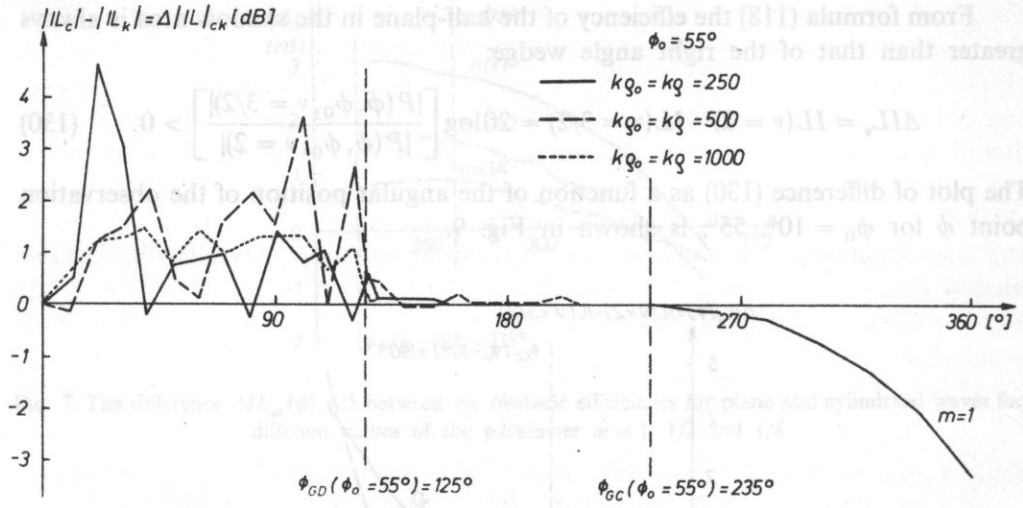


FIG. 10. The difference $\Delta|IL|_{ck}(\phi)$ between the absolute values of the obstacle efficiency for cylindrical and spherical waves for the source position $\phi = 55°$, in a symmetrical system $kq = kq_0$

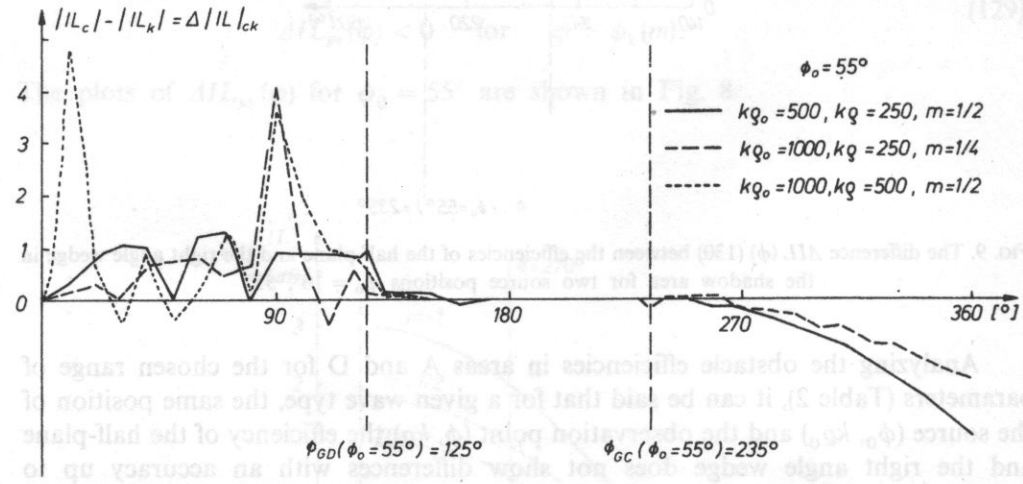


FIG. 11. The difference $\Delta|IL|_{ck}(\phi)$ between the absolute values of the obstacle efficiency for cylindrical and spherical waves for the source position $\phi_0 = 55°$, in a nonsymmetrical system $kq \neq kq_0$

This is shown in Figs. 10 and 11 for symmetrical and nonsymmetrical positions of the source and the observation point with respect to the diffraction edge, for $\phi_0 = 55°$. The fact that for the cylindrical wave in areas A and D the absolute values of the efficiency $|IL_c|$ are greater than those of $|IL_s|$ for the spherical wave, can be explained by the existence of a greater degree of spatial correlation between the dominating component waves which form the field in the case of a cylindrical wave.

For the numerical examples, area D, where the dominating components are the direct wave ($V^i(R)$) and the geometrical wave reflected from the half-plane $\phi = 270°$ ($V^i(R'')$) occurs for $\phi_0 = 135°$, i.e., for the position of the source on the axis of symmetry of the wedge. Hence, the acoustic fields, and, thus, the efficiencies, in areas D and A are the same (see Fig. 12). For comparison, Fig. 13 shows the efficiency of the half-plane for $\phi_0 = 135°$. It can be seen from these figures that the efficiencies of the half-plane and the wedge in area A are the same. Moreover, it can be seen that in the case of the half-plane, area D does not occur, and is replaced by the shadow area C.

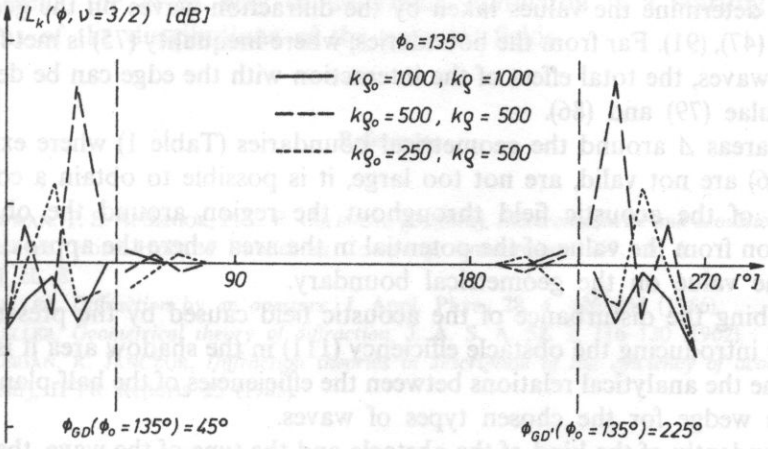


FIG. 12. The efficiency of the right angle wedge for spherical waves for $\phi_0 = 135°$

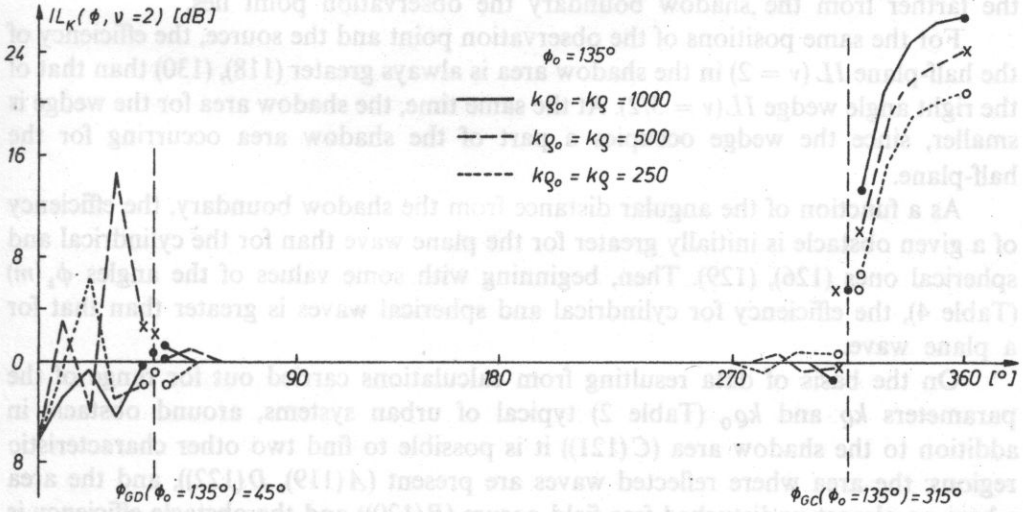


FIG. 13. The efficiency of the half-plane for spherical waves for $\phi_0 = 135°$

4. Conclusion

For the presented six cases of interaction between successively plane, cylindrical and spherical waves, and the edges of an ideal rigid half-plane and an ideal rigid wedge a uniform description of the acoustic field structure is achieved. It consists of the geometrical part (5), (19) and the diffraction part (10), (24) of the acoustic potential. It applies to systems in which, from conditions (58), (62) and (63), both the source and the observation point are far from the diffraction edge.

In the diffraction part of the potential, diffraction waves defined by formulae (73) and (74) and tied up to the appropriate geometrical waves were found. This made it possible to determine the values taken by the diffraction waves on the geometrical boundaries (47), (91). Far from the boundaries, where inequality (75) is met for all the diffraction waves, the total effect of the interaction with the edge can be determined from formulae (79) and (86).

If the areas Δ around the geometrical boundaries (Table 1) where expressions (79) and (86) are not valid, are not too large, it is possible to obtain a continuous description of the acoustic field throughout the region around the obstacle by extrapolation from the value of the potential in the area where the approximation is valid to the value on the geometrical boundary.

Describing the disturbance of the acoustic field caused by the presence of an obstacle by introducing the obstacle efficiency (111) in the shadow area it is possible to determine the analytical relations between the efficiencies of the half-plane and the right angle wedge for the chosen types of waves.

Independently of the kind of the obstacle and the type of the wave, the obstacle efficiency IL in the shadow area (114) is the greater the farther from the diffraction edge the observation point and the source are, and the deeper into the shadow area the farther from the shadow boundary the observation point lies.

For the same positions of the observation point and the source, the efficiency of the half-plane $IL(\nu = 2)$ in the shadow area is always greater (118), (130) than that of the right angle wedge $IL(\nu = 3/2)$. At the same time, the shadow area for the wedge is smaller, since the wedge occupies a part of the shadow area occurring for the half-plane.

As a function of the angular distance from the shadow boundary, the efficiency of a given obstacle is initially greater for the plane wave than for the cylindrical and spherical ones (126), (129). Then, beginning with some values of the angles $\phi_k(m)$ (Table 4), the efficiency for cylindrical and spherical waves is greater than that for a plane wave.

On the basis of data resulting from calculations carried out for range of the parameters $k\varrho$ and $k\varrho_0$ (Table 2) typical of urban systems, around obstacle in addition to the shadow area (C(121)) it is possible to find two other characteristic regions: the area where reflected waves are present (A(119), D(122)), and the area where an almost undisturbed free field occurs (B(120)) and the obstacle efficiency is zero.

In areas where reflected waves occur (A, D) the obstacle efficiency can take large values at points where the dominating component waves (direct and reflected) interfere in antiphases. At points where they interfere in the same phases, the barrier efficiency takes negative values of about -6 dB.

The comparative analysis performed makes it possible to draw conclusions about the structure of the acoustic fields around the obstacle in question. In real urban system where the dominating phenomenon forming the acoustic field, is the diffraction at an edge, or a right angle wedge, from the formulae here it is possible to determine the obstacle efficiency in regions where the approximation conditions are satisfied. In areas for which it is necessary to consider interaction with additional planes present, the given way of describing diffraction is a starting point for calculations of the distributions of the acoustic fields.

References

[1] J. J. BOWMAN, T. B. A. SENIOR, P. L. E. USLENGHI [editors], *Electromagnetic and acoustic scattering by simple shapes*, North-Holland Publishing Company, Amsterdam 1969, ch. 6.

[2] Ref. [1], ch. 8.

[3] J. B. KELLER, *Diffraction by an aperture*, J. Appl. Phys., **28**, 4, 426–444 (1956).

[4] J. B. KELLER, *Geometrical theory of diffraction*, J. A. S. A. **52**, 2, 116–130 (1962).

[5] E. WALERIAN, R. JANCZUR, *Diffraction theories in description of the efficiency of acoustic sciences* [in Polish], IFTR Reports 25 (1985).

Received February 25, 1987.

ARCHIVES OF ACOUSTICS
13, 1-2, 191-202 (1988)

DETERMINATION OF CIRCULAR MEMBRANE PARAMETERS FROM ITS RESONANCE FREQUENCIES

MARIUSZ ZIÓŁKO

Institute of Automatic Control, AGH (30-059 Kraków, al. Mickiewicza 30)

Circumferential forces, speed of elastic wave and dissipation of energy factor are calculated from resonance frequencies of circular membranes made of tantalum and nickel-chromium steel. The results obtained for a mathematical model without dissipation are compared with the results obtained for a model with dissipation of energy. The assignment of the coefficient of partial differential equation is presented for this second case. An algorithm applied to computer simulation of membrane vibrations is based on the "leap frog" difference method.

Znając częstotliwości rezonansowe membran kołowych można obliczyć ich obwodowe siły napięcia, prędkości propagacji fal sprężystych i współczynnik dyssypacji energii. Obliczenia przeprowadzono posługując się danymi uzyskanymi dla membrany tantalowej i membran ze stali niklowo-chromowej. Porównane są wyniki obliczeń dla modelu matematycznego bez dyssypacji z rezultatami dla modelu uwzględniającego rozpraszanie energii. Dla tego drugiego przypadku przedstawiony jest również sposób identyfikacji współczynników równania różniczkowego cząstkowego na podstawie danych eksperymentalnych. Algorytm zastosowany do komputerowego modelowania drgań membran opiera się na metodzie różnicowej „leap frog".

1. Introduction

Euler gave a mathematical model of circular membrane vibrations [1, 4]. This is the differential equation of the hyperbolic type for axial strain as a function of three variables: time t and polar co-ordinates, that is the distance from the centre of membrane and angle ϕ

$$\frac{1}{c^2}\frac{\partial^2 z}{\partial t^2} - \frac{\partial^2 z}{\partial r^2} - \frac{1}{r}\frac{\partial z}{\partial r} - \frac{1}{r^2}\frac{\partial^2 z}{\partial \phi^2} = 0, \tag{1}$$

where the constant coefficient c is the speed of elastic wave.

It is most frequently assumed that the membrane is fixed stiff on the circumference. This means that the boundary conditions for equation (1) are equal to zero. Thus the solution of the homogeneous differential equation (1) presents nondamped vibrations as a result of nonzero initial conditions. If they are well assumed it is possible to simulate steady and symmetrical axial strains. Using the classical method of separation of variables [1] [2] we obtain the relation

$$c = \frac{\omega_i R}{x_i} \tag{2}$$

which enables to calculate the speed of elastic wave c. R is the radius of membrane, x_i are roots of Bessel's function of first kind and zero order, ω_i are resonance frequencies. Next, from the speed of elastic wave it is possible to calculate the circumferencial force. This well known formula enables to calculate the value of the force which is difficult to measure. Usually we can measure a few resonance frequencies. For each of them we obtain from (2) an estimator for the speed of an elastic wave. Differences between these estimators usually differ at the second decimal place. This justifies a verification of the usability of a more complex mathematical model with dissipation of energy.

2. Mathematical model of membrane

Assuming the axial symmetry of forces which deform membrane and introducing into equation (1) the term for dissipation of energy and term for forced vibration, we obtain nonhomogeneous partial differential equation of hyperbolic type

$$\frac{1}{c^2}\frac{\partial^2 z}{\partial t^2} - \frac{\partial^2 z}{\partial r^2} + a\frac{\partial z}{\partial t} - \frac{1}{r}\frac{\partial z}{\partial r} = u, \tag{3}$$

where z – transverse strain of membrane [m], c – speed of elastic wave travelling along the radius of membrane [m/s], a – positive coefficient of dissipation of the energy [s/m^2], u – force vibration function of time t and space variable r [m^{-1}], t – time [s].

To obtain a unique solution of the equation (3) we assume initial conditions

$$z(r,0) = 0, \quad \left.\frac{\partial z}{\partial t}\right|_{t=0} = 0, \tag{4}$$

and boundary conditions

$$\left.\frac{\partial z}{\partial r}\right|_{r=0} = 0, \quad z(R,t) = 0, \tag{5}$$

where R [m] is the radius of membrane. The first boundary condition follows from

the axial symmetry of strain and the second one means that the membrane is fixed stiff on circumference.

Using the classical method we assume that the solution of initial boundary value problem (3), (4), (5) can be described as an infinite series

$$z(r,t) = \sum_{i=1}^{\infty} T_i(t)\, R_{i(r)}. \tag{6}$$

A function forcing the vibration of membrane with frequency ω can also be written as an infinite series with separable variables

$$u(r,t) = \sum_{i=1}^{\infty} p_i R_i(r) \sin \omega t, \tag{7}$$

where p_i are constant coefficients.

Computing the partial derivatives of function (6) and substituting them into (3) we obtain conditions

$$\frac{1}{c^2}\frac{T_i''}{T_i} + a\frac{T_i'}{T_i} - \frac{p_i}{T_i}\sin\omega t = \frac{R_i''}{R_i} + \frac{1}{r}\frac{R_i'}{R_i}, \tag{8}$$

where the upper indexes denote the differentiation with respect to time for the left side of (8) and differentiation with respect to the space variable for the right side of (8). Denoting the value of both sides by $-k_i^2$ we obtain two sets of ordinary differential equations

$$R_i'' + \frac{1}{r}R_i' + k_i R_i = 0, \tag{9}$$

$$\frac{1}{c^2}T_i'' + a\,T_i' + k_i^2\,T_i = p_i \sin\omega t. \tag{10}$$

The boundary condition for equation (9) is obtained from (5)

$$R_i'(0) = 0, \qquad R_i(R) = 0. \tag{11}$$

In this way we improve the constrains for basis functions R_i. It follows that function u must fulfil conditions

$$\left.\frac{\partial z}{\partial r}\right|_{r=0} = 0, \qquad z(R) = 0,$$

for convergence of series (7).

From (4) we obtain the initial conditions for equation (10)

$$T_i(0) = 0, \qquad T_i'(0) = 0. \tag{12}$$

The solution of differential equation (9) is Bessel's function of first kind and zero

order which can be defined either by

$$R_i(r) = \sum_{j=0}^{\infty} (-1)^j \frac{k_i^{2j}}{j!j!}\left(\frac{r}{2}\right)^{2j} \tag{13}$$

or in an equivalent form

$$R_i(r) = \sum_{j=0}^{\infty} a_j (k_i r)^{2j},$$
$$a_0 = 1, \quad a_{j+1} = -\frac{a_j}{4j^2}. \tag{14}$$

From (13) we find that the first condition of (11) is always satisfied. The second condition requires

$$k_i = \frac{x_i}{R}, \tag{15}$$

where x_i are roots of Bessel's function.

The solution of equation (10) with the initial conditions (12) has the form

$$T_i = A_i e^{-\alpha t} \sin(\omega_i t + \phi_i) + B_i \sin(\omega t + \psi_i). \tag{16}$$

The frequency of damping vibrations is given by

$$\omega_i = c\sqrt{k_i^2 - \frac{a^2 c^2}{4}}. \tag{17}$$

Introducing auxiliary variable

$$s_i = \sqrt{\left(k_i^2 - \frac{\omega^2}{c^2}\right)^2 + a^2 \omega^2}, \tag{18}$$

we obtain initial amplitude for damping vibrations

$$A_i = \frac{\omega p_i}{\omega_i s_i}. \tag{19}$$

The phase displacement of these vibrations is obtained from formula

$$\phi_i = \arcsin\frac{a\omega_i}{s_i} \tag{20}$$

and the damping coefficient from

$$\alpha = \frac{ac^2}{2}. \tag{21}$$

The amplitude of steady vibration with forced frequency is given by

$$B_i = \frac{p_i}{s_i} \tag{22}$$

and its phase displacement

$$\psi_i = -\arcsin\frac{a\omega}{s_i}. \tag{23}$$

The greatest amplitudes of the steady vibrations occur for resonance frequencies

$$t_i^r = \frac{c}{2\pi}\sqrt{k_i^2 - \frac{a^2c^2}{2}} \tag{24}$$

and are given by

$$B_i^r = \frac{p_i}{a\omega_i}, \tag{25}$$

where ω_i is defined by (17). The coefficient of dissipation must be small enough for the i-th resonance vibrations to occur, that is

$$a < \frac{\sqrt{2}\,k_i}{c}. \tag{26}$$

3. Determination of membrane parameters

The formula which connects resonance frequencies with the parameters of membrane is obtained from (15) and (24). Assuming that we have managed to measure N first resonance frequencies, which we denote now by ω_i, we obtain the set of equations

$$\omega_i^2 = c^2\left[\left(\frac{x_i}{R}\right)^2 - \frac{a^2c^2}{2}\right]. \tag{27}$$

This means that the resonance frequencies are the functions of three parameters: speed of elastic wave, coefficient of dissipation and radius of membrane. The Jacobian determinant of function (27) has the form

$$\mathrm{J} = \begin{vmatrix} 2c\left(\frac{x_1}{R}\right)^2 - 2a^2c^3 & -c^4a & -\frac{2c^2x_1^2}{R^3} \\ 2c\left(\frac{x_2}{R}\right)^2 - 2a^2c^3 & -c^4a & -\frac{2c^2x_2^2}{R^3} \\ \cdots\cdots\cdots\cdots\cdots & \cdots\cdots & \cdots\cdots \\ 2c\left(\frac{x_N}{R}\right)^2 - 2a^2c^3 & -c^4a & -\frac{2c^2x_N^2}{R^3} \end{vmatrix}. \tag{28}$$

Only two columns of determinant (28) are linearly independent, therefore only two parameters can be calculated from resonance frequencies. As it is easy to measure the radius of membrane, so we will calculate, from (27), the speed of elastic wave and coefficient of dissipation.

From two arbitrary equations (27) we obtain

$$c_{ij} = R\sqrt{\frac{w_i^2 - w_j^2}{x_i^2 - w_j^2}}, \quad 1 \leqslant i,\ j \leqslant N,\ i \neq j, \tag{29}$$

where c_{ij} is one of the estimates of the speed of elastic wave. For N resonance frequencies, the average value of estimates is equal to

$$c = \frac{2(N-2)!}{N!} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} c_{ij}. \tag{30}$$

Next we can calculate the force stretching the membrane [N/m]

$$F = qdc^2, \tag{31}$$

where q – mass density of material of membrane [kg/m^3], d – membrane thickness [m].

From (27) we can calculate estimates for the coefficient of dissipation

$$a_i = \frac{\sqrt{2}}{c}\sqrt{\frac{x_i^2}{R^2} - \frac{\omega_i^2}{c^2}} \tag{32}$$

and obtain finally

$$a = \frac{1}{N}\sum_{i=1}^{N} a_i.$$

If it is possible to measure the amplitudes B_i^r of vibrations of the centre of membrane, we can calculate additionally the amplitudes of forced vibrations [1/m]

$$p_i = B_i^r ac\sqrt{\frac{x_i^2}{R^2} - \frac{a^2c^2}{4}}. \tag{33}$$

4. Measurement of resonance frequencies

The vibrations of circular membrane, stretched with the same force on the whole circumference, can be stimulated by a sonic waves. In this way, the resonance frequencies of membrane were assigned. A variable frequency generator supplied a loudspeaker and was connected with frequency meter. The sonic waves from the loudspeaker reached microphone through the membrane fixed between the stretching rings. The signal from the microphone was amplified and next measured by the digital voltage meter. Experiments were started from the possible lowest frequency

and afterward it was increased gradually. The first resonance frequency was found when the sound intensity indicated by the microphone had the well-marked greatest value. Next, by increasing the frequency of generator once again, the other resonance vibrations were found.

Measurements were carried out for membranes made of tantalum and nickel-chromium steel. The first five resonance frequencies were found for every membrane. The results are presented in Table 1.

Table 1. Measurement results

Membrane No	1	2	3	
Material	tantalum	steel	steel	
Radius [mm]	27.8	27.8	40	
Thickness [mm]	0.025	0.127	0.127	
Roots of Bessel's function	Resonance frequencies [Hz]			Amplitude [mm]
2.40483	434	1043	695	1.25
5.52008	1050	2594	1940	0.54
8.65373	1658	4115	3130	0.34
11.79153	2264	5630	4285	0.25
14.93092	2870	7141	5428	0.19

Sometimes the vibrations of the center of membrane have amplitude great enough to be measured. For this purpose the measuring position was equipped additionally with a micrometer screw. Its end was placed above the center of the membrane. The junction of micrometer screw with the metal membrane was signalled by an ohmmeter as an electric short circuit. After putting the membrane into resonance vibrations, the micrometer screw was dropped until the junction of the micrometer screw with the membrane during its greatest deflection. Next, the generator was switched off and the micrometer screw was dropped once again until it reached the membrane. The difference between the positions of micrometer screw in both these cases was equal to the amplitude of membrane vibrations. These measurements were made only for the third membrane which had the greatest amplitudes of vibration.

Substituting the values of the resonance frequencies into (2) we obtain for Euler equation (1) the estimates of coefficient c. The results of these calculations for data from Table 1 are presented in Table 2. For every membrane there were measured five resonance frequencies, therefore we obtained five estimates of parameter c. The last line of Table 2 presents the mean values of wave speeds.

From formula (29) we obtain for each membrane ten estimates of the speed of elastic wave. The results are presented in Table 3 and their mean values are written in the last line. The circumferencial forces were calculated in accordance with (31).

Table. 2. Speeds of elastic waves calculated from Euler's model [m/s]

No of resonance frequency	Membrane No		
	1	2	3
1	31.52	75.76	72.63
2	33.23	82.08	88.33
3	33.47	83.06	90.90
4	33.54	83.40	91.33
5	33.58	83.54	91.37
average	33.07	81.57	86.91

Table 3. Speeds of elastic waves [m/s] and circumferencial forces [N/m] calculated from the model with dissipation

Estimate No	Membrane					
	1		2		3	
	speed	force	speed	force	speed	force
1	33.61	469	83.49	7120	91.62	8580
2	33.62	469	83.64	7150	92.27	8700
3	33.62	469	83.72	7160	92.06	8660
4	33.63	469	83.74	7160	91.81	8610
5	33.63	469	83.72	7160	92.63	8770
6	33.62	469	83.77	7170	92.16	8680
7	33.63	469	83.77	7170	91.84	8620
8	33.62	469	83.79	7170	91.83	8620
9	33.63	469	83.78	7170	91.60	8570
10	33.64	470	83.77	7170	91.43	8540
average	33.63	469	83.72	7160	91.93	8640

Table 4. Values of dissipation coefficients [s/m^2]

No of resonance frequency	Membrane No		
	1	2	3
1	1.27	0.622	0.567
2	1.29	0.660	0.588
3	1.27	0.659	0.495
4	1.29	0.626	0.514
5	1.24	0.594	0.631
average	1.27	0.632	0.559

The mass density for tantalum was assumed $q = 16.6$ [g/cm^3], while for nickel-chromium steel $q = 8.045$ [g/cm^3].

Table 4 presents the values of the dissipation coefficient calculated from (32) for five resonance frequencies and their mean values. We obtain the damping coefficient by putting mean values of c and a into formula (21). For data presented in Tables 3 and 4 we obtain values

$$\alpha_1 = 718\,[1/s], \quad \alpha_2 = 2215\,[1/s], \quad \alpha_3 = 2362\,[1/s].$$

For the third membrane were measured the amplitudes of vibrations additionally (Table 1), therefore it was possible to compute amplitudes of forced vibrations according to (33). The results are presented in Table 5.

Table 5. Amplitudes of forced vibrations

No of resonance frequency	Amplitudes [1/m]
1	0.349
2	0.376
3	0.375
4	0.377
5	0.364

5. Computer modelling of membrane vibrations

The "leap frog" difference method is frequently used to solve numerically the partial differential equation of hyperbolic type. In the adequate distribution of knots of the space grid the "wave character" of the occurred phenomena is taken into account. The second valuable advantage of this method is the simplicity of its algorithm. For these reasons the leap frog method was used to solve numerically equation (3) which is the mathematical model of vibrating membrane.

Dividing the membrane along its radius into N segments, we define the arrangement of knots of discrete space

$$\Delta = \left\{(r_l, t_k): r_l = R-(l-1)h, t_k = (k-2)\frac{h}{c};\ h = \frac{2R}{2N-1};\ l = 1, 2, \ldots, N+1;\right.$$
$$\left. k = 1, 2, \ldots, \frac{Tc}{h}\right\}, \quad (34)$$

where T denotes the final time. The knots are distributed in such a way that the coefficient of derivative in equation (3) with respect to membrane radius is limited,

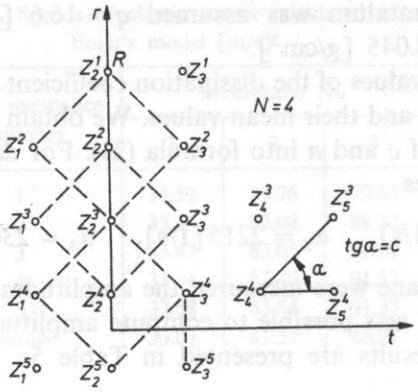


FIG. 1. Distribution of knots in space grid

that is $r_l \neq 0$. Thanks to that, the coefficients of difference equation are also limited. The derivatives from equation (3) are approximated by difference quotients in the following terms

$$\frac{\partial z}{\partial t} \sim (z_{k+1}^l - z_{k-1}^l)\frac{c}{2h}, \quad \frac{\partial^2 z}{\partial t^2} \sim (z_{k+1}^l - 2z_{k-1}^l + z_{k-1}^l)\frac{c^2}{h^2},$$
$$\frac{\partial z}{\partial r} \sim (z_k^{l-1} - z_k^{l+1})/2h, \quad \frac{\partial^2 z}{\partial r^2} \sim (z_k^{l-1} - 2z_k^l + z_k^{l+1})/h^2. \tag{35}$$

For the interior knots defined by (34) the approximate solution of the differential equation (3) is calculated from formula

$$z_{k+1}^l = z_{k-1}^l w_1 + z_k^{l+1} w_2 + z_k^{l-1} w_3 + w_4 u_{(R-(l-1)h, kh/c)}, \tag{36}$$

where the coefficients have values

$$w_4 = \left(\frac{ac}{2h} + \frac{1}{h^2}\right)^{-1}, \quad w_1 = \left(\frac{ac}{2h} - \frac{1}{h^2}\right) w_4,$$
$$w_2 = \left(\frac{1}{h^2} - \frac{1}{2h[R-(l-1)h]}\right) w_4, \quad w_3 = \left(\frac{1}{h^2} + \frac{1}{2h[R-(l-1)h]}\right) w_4.$$

For the boundary knots we obtain from (5)

$$z_k^1 = 0, \quad z_k^{N+1} = z_k^N, \quad k = 3,4,\ldots \tag{37}$$

The leap frog method is a three-step difference scheme. Therefore, at the beginning we put values into the first two steps

$$z_1^l = z_2^l = 0, \quad l = 1, 2, \ldots, N+1. \tag{38}$$

In this way we take into account the initial conditions (4).

The mesh width results from definition (34)

$$\Delta r = \frac{2R}{2N-1}, \quad \Delta t = \frac{2R}{c(2N-1)}. \tag{39}$$

The number of all knots is

$$I = (2N^2+N-1)\frac{Tc}{2R}. \tag{40}$$

It means that the time of computations increases considerably for large N. To obtain correct results of computer modelling of membrane vibrations with the frequency f [Hz] it is sufficient to take

$$N \approx 40\frac{Rf}{c}. \tag{41}$$

Vibrations with the fifth resonance frequency for the membrane number 3 were simulated taking $N = 100$, in other words

$$\Delta r = 402.01\,[\mu m], \quad \Delta t = 4.373\,[\mu s].$$

For this example the time-constant of unsteady state, equal to inverse of α, is equal to 0.42 [ms]. Therefore we can assume that unsteady state vanishes after time 1.7 [ms].

6. Conclusions

In Table 2 there are presented the results of calculations for the classical mathematical model without dissipation. The estimates for the speed of propagation of elastic wave were obtained for each resonance vibration. Differences between them and their mean values are considerable, especially for low frequencies. These differences amount from a few percents for the first and second membrane until 16% for the third membrane. On the other hand, the estimates of speed of elastic wave calculated for the model with dissipation (Table 3) have small deviations (less than 1%). The mean values of speed of elastic waves for the model with dissipation are greater than the mean values for the model without dissipation. The differences are considerable, they are equal to a few percents.

The coefficient of dissipation introduced as a new parameter into the mathematical model, enable to fit better the mathematical model to the experimental data. This possibility exists in general, even if there is no physical justification for such a treatment. However, we must remember that the formula (32) can be used only if $k_i \geqslant \omega_i/c$.

For the first membrane there are small differences between the estimates of coefficient of dissipation (Table 4). The greatest difference between the mean value

and the estimate, occurs for the fifth resonance frequency of third membrane. Its value is equal to 13%.

The generator of electric sinusoidal oscillations had constant amplitude for all frequencies. Therefore the amplitudes of forced vibrations presented in Table 5 are not far each from the other.

The additional advantage of taking into account the dissipation of energy consists in obtaining formulae for numerical solution of the differential equation with better property of numerical stability.

References

[1] I. MALECKI, *Teoria fal i układów akustycznych*, PWN, Warszawa 1964, pp. 472–475.
[2] E. SKUDRZYK, *Die Grundlagen der Akustik*, Springer-Verlag, Wien 1954, pp. 53, 134, 332.
[3] G. N. WATSON, *A treatise on the theory of Bessel function*, University Press, Cambridge 1962, p. 5.
[4] Z. ŻYSZKOWSKI, *Podstawy elektroakustyki*, WNT, Warszawa 1966, p. 182.

Received April 1, 1987.