helping deaf persons, acoustic fields,
acoustic signals analysis, acoustic signals interpretation,
acoustic environment description

Juliusz Lech KULIKOWSKI*

ACOUSTIC ENVIRONMENT ANALYSERS AS TOOLS FOR DEAF PERSONS HELPING

The paper presents a concept of acoustic field analyser as a tool for helping deaf people, able to perceive visual and/or tactile signals only. The acoustic field is considered as a sum of acoustic waves carrying information about events occurring in a monitored physical space. Recognition of acoustic signals together with identification of their sources give a possibility to recognise events of vital importance to a deaf person. There are presented here general problems of acoustic environment analysers' design, particular attention being paid to acoustic sources localisation in physical space.

1. INTRODUCTION

According to some opinions expressed by psychological and psycho-neurological specialists, more than 80% of information from the external world acquired by a normal human brain is received by its visual tract; while the corresponding rate of aural information reaches, approximately, 15%. However, for a deaf person this relatively low external world perception impairment level means impossibility of a quick and adequate reaction to unexpected events arising in his environment and manifested by acoustic signals. Therefore, even if the above-given rates have not been quite precisely evaluated, they illustrate the weight of difficulties arising in communication between deaf people and their environment. Let us remark that deaf people constitute the third most numerous group of disable persons after the blind and the motion-impaired ones; helping them is thus a problem of high social importance.

It is also well known that, within certain limits, inefficiency of some perceptual tracts can be compensated by increased efficiency of the other ones. Till the time when the problem of disabled people helping will be effectively solved by a progress reached in natural or high-quality artificial implants applications, it remains to solve it partially by substitution of the impaired organs' functions by adequately formed functions of other, efficient ones [3]. For this purpose adapting systems, as shown in Fig. 1, can be used. In particular, the impairment of acoustic signals perception can compensated, up to a certain degree, by an *acoustic environment analyser* (AEA) acquiring, recognising and interpreting sounds and rendering the corresponding communicates in an admissible (visual, tactile, etc.) form to the user. This solution can be realised in two basic variants:

* Polish Academy of Sciences, Institute of Biocybernetics and Biomedical Engineering, 4 Ks. Trojdena Str., 02-109 Warsaw, Poland, jlkulik@ibib.waw.pl

first, limited to interpretation of physical sounds only, and second, oriented to interpretation of speech signals. In the future both variants, probably, will be combined within unified acoustic signals analysing systems. The idea of AEA arose and first experiments for proving it have been realised in the Dept. of Biomedical Information Processing Methods IBBE PAS [6]. In this paper our attention will be focused on some basic concepts of the first-type AEAs design. In particular, we would like to stress out the difference between acoustic environment analysis problems and those ones of speech recognition or human voice identification described in the literature [1,5].

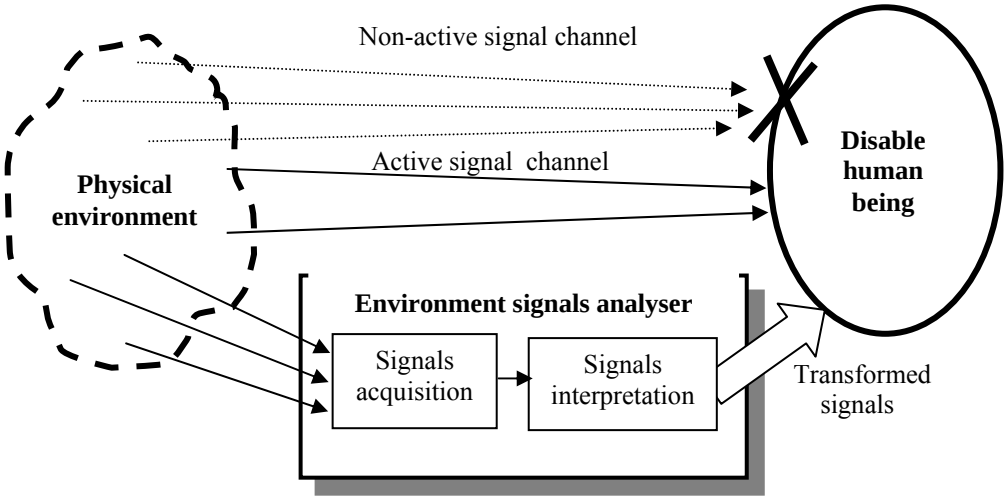


Fig.1. An adapting system for a disabled person.

2. ACOUSTIC ENVIRONMENT

Acoustic environment can be considered on several levels:

- on a *physical level*, when acoustic field and its propagation aspects are taken into account,
- on an *informational level*, when sounds are analysed as carrying information signals,
- on a *semantic level*, when some events causing generation of acoustic signals are taken into account, and
- on a pragmatic level, when a *weight of importance* is assigned to the acoustic signals.

From a physical point of view *acoustic field* ***F*** can be defined as a superposition of *acoustic waves* penetrating a given 3D space-area S. Each acoustic wave is described by a pair

$$\boldsymbol{w}_i = [\boldsymbol{n}_i(x,y,z,t),\ f_i(x,y,z,t)] \qquad (1)$$

where i is an index assigned to the wave, $\mathbf{n}_i(x,y,z,t)$ is a *normal vector* indicating an instantaneous wave propagation direction at a fixed point $(x,y,z) \in S$, t being a value on a real time axis, $f_i(x,y,z,t)$ describes a pressure at the given point and at the time t. It will be assumed that the acoustic field is generated by a finite number I of sound sources, and, thus

$$\mathbf{F} = \sum_i \mathbf{w}_i\ ,\ i = 1,2,...,I. \qquad (2)$$

A dependence of $\mathbf{n}_i(x,y,z,t)$ on t is caused by source movement, while this of $f_i(x,y,z,t)$ describes sound as a sort of vibrations. Therefore, the time-variations in those cases are of different. character. Acoustic wave $\mathbf{w}_i$ is called *steady* if its normal vector does not depend on time. Correspondingly, acoustic field will be called *steady* if it is generated only by steady sound sources.

Acoustic wave is called *flat* if its normal vector is constant in the given space-area. Acoustic wave generated by a point-source located in a physically homogenous space-area is *spherical* in the vicinity of the source and becomes approximately flat in limited space-areas at long distances of the source. If (ξ_i, η_i, ζ_i) are spatial co-ordinates of the i-th sound source, r_i is the distance between (ξ_i, η_i, ζ_i) and (x,y,z), c is the sound propagation velocity in the given physical environment (being assumed to be constant in S), than

$$f_i(x,y,z,t) = K_i \cdot f_i(\xi_i, \eta_i, \zeta_i, t - \tau_i) \qquad (3)$$

K_i being an *attenuation factor* depending on r_i and τ_i – a *time-delay* given by

$$\tau_i = \frac{r_i}{c} \qquad (4)$$

In a more general case, when acoustic wave is not propagated along a straight line or the sound velocity is not constant in S the expression (4) should be modified in the sense that an *effective physical distance* δ_i among the source and the current point (x,y,z) instead of r_i is taken into account.

Each acoustic wave can be considered as an *information carrying factor*. From their *time-variations* point of view acoustic waves can be divided into two classes:

a) those described by *deterministic functions* $f_i(x,y,z,t)$,
b) those being instances of some *non-deterministic processes*.

It might seem that the first class is of low interest for us, deterministic functions carrying no information. However, it is not quite so, because even if an acoustic wave carries a deterministic sound a useful information may be contained in its normal vector indicating a direction to the source.

Next classification step of non-deterministic processes leads to the following sub-classes:

b^1) described by *stationary stochastic* processes,
b^2) described by *non-stationary stochastic* processes,
b^3) described by *irregular non-stochastic* processes (like *fractals*, etc.).

Examples of the b^1 sub-class are: harmonic functions of fixed but previously non-determined amplitude, frequency and/or phase, stochastic noise of fixed power spectrum, etc.

The sub-class b^2 contains, in particular, two important sub-classes of *limited time-duration* functions:

$b^{2,1}$) of a form determined up to a finite set of parameters,
$b^{2,2}$) of a non-determined form.

Example of the $b^{2,1}$ subclass is a door-bell sound of fixed frequency, random starting time-instant and random time-duration. Examples of the $b^{2,2}$ sub-class are: dog's barking, child's cry, human voices heard from behind the door, etc.

In general, it seems reasonable to describe a separate acoustic signal at a given point (x,y,z) as a time-function of the form:

$$f_i(x,y,z,t) = E_i(t) \cdot \varphi_i(t) \qquad (5)$$

where $E_i(t)$ is a non-negative slowly-varying "envelope" representing a rough signal's structure, while $\varphi_i(t)$ is an oscillating "filler" representing its fine structure (for the sake of simplicity spatial arguments in both factors have been omitted). A rough classification of acoustic signals thus can be based on the properties of $E_i(t)$. In particular, there can be distinguished:

1. long-time durable signals:
   - of constant amplitude,
   - of ascending amplitude,
   - of descending amplitude,
   - of fluctuating amplitude,
2. short-time durable signals:
   - single
   - serial:
     - periodically repeated,
     - non-periodically repeated,
     - of a fixed envelope form,
     - of various envelope forms.

On a semantic level of acoustic environment analysis its relationships with a real world are considered. In general, no strong correspondence between physical properties of acoustic signals and their semantic interpretation exists. However, we try to establish, at least, a general frame of such correspondence. For this purpose we shall consider a *real world* as a *composition of events* occurring in a (3+1)D physical space (one dimension being assigned to time). It is evident that only a limited part of the whole physical space is accessible for observation; this part, containing a set of perceptible physical or biological *objects* arranged in a certain way, will be called a *scene*. The objects can be characterised by their *features*, like: *individual identifiers, spatial allocation, physical states*, etc. Some subsets of objects can be also considered as *collections* satisfying certain *relationships*; in such case the collection can be considered as a, higher-level, *composite object*. The objects are not steady in time: they may arise and/or disappear, change their features, realise various forms of action, form new relationships with other objects and/or leave the former ones, etc. Each change or action of this type is called an *event*. Our attention is here focused on a class of events that are manifested by emission of specific sounds. Therefore, the objects of interest at the same time become sources of acoustic signals. A sequence of scenes, linearly ordered in time and representing some events occurring in the sector of physical space under observation will be called an *animated acoustic scene*.

Semantic interpretation of acoustic environment consists in spatial allocation and assigning names of events to the received acoustic signals. For this purpose three types of data should be taken into account:

1. normal vectors of acoustic waves indicating the directions to signal sources;
2. spectral and/or time-variation characteristics of acoustic signals;

3. spatial and/or temporal relationships between acoustic signals considered as components of the same animated acoustic scene.

In practice not all events are of equal importance for the users. A rough classification of events leads to the following classes:

- ordinary events (OE), of low importance level,
- extraordinary events (EE), of moderate importance,
- highly important extraordinary events (HIE).

Typical examples of OE are: a typical acoustic noise heard from a window, people speaking in the next room, etc. Examples of EE are: a sound of a door bell, sudden increasing of a human voice level, dog's barking, etc. Finally, examples of HIE are: fire-alarm signals, sound of explosion, loud human voices or cries, intensive knocking to the door, etc. The above-described signals envelope characteristics in connection with acoustic sources localisation should be a sufficient basis for acoustic signal importance-weight evaluation. However, in a more general case acoustic signal's importance may depend not only on its physical characteristics, but also on a situation it arises, as well as on subjective weight assigned to it by the user.

3. THE STRUCTURE OF ACOUSTIC ENVIRONMENT ANALYSERS

Human sensory systems make us able to perceive surrounding animated scenes. Up to a certain level this process of perception may be unconsciously controlled by our mind, able to select some events and to neglect the other ones. The criteria of selection can be spontaneous or intentionally inspired, the mechanism of selection working, in both cases, according to a principle of information processing effort reduction to an admissible minimum. As a consequence, from a depth of information processing point of view AEAs can be divided into three groups:

- local signalling devices,
- central monitoring systems,
- acoustic scene virtual modelling systems.

Local signalling devices are widely used, not only for disabled persons helping; they emit standard light-signals carrying information about simple and well-defined events: opening the door by somebody, water boiling in the kettle, an electric device being switched on or off, a telephone call, etc. The list of events as well as their spatial allocation in such case can be strongly fixed and, thus, the flexibility of such systems is rather low. And still, assuming that disabled persons are able to perceive the given type of signals, such systems in many every-day situations may help them quite effectively. A big inconvenience of such systems in the case of deaf persons consists in the necessity of a permanent visual contact between the signalling device and the user for an effective light-signals perception.

Central monitoring systems can be considered as the next step in disabled person's environment accommodation to his needs and abilities. In this case a set of input devices: sensors, TV cameras, microphones, etc. could be used to acquire signals from the environment. Signals are then transmitted to a central point where they can be processed and presented to the user in an available form. Flexibility of such systems could be much higher than in the former case. First, microphones or TV cameras can capture signals from large areas; second, the type of events signalised is not strongly limited. Of course, there remain some constraints connected with the type

of physical signals that can be perceived by a system user. For example, a TV monitoring system is useless to a deaf person while his ability to perceive events manifested by sounds can be restored and extended by an installation of microphones distributed, say, in several points of his flat. Central monitoring gives also the possibility of signals selection according to the interests of the user. He can switch off the unnecessary input devices, to change their observation sectors, etc. In central monitoring systems signals recognition is realised on low levels only while interpretation of events remains to be performed by the user.

Virtual scene modelling is the most advanced method of acoustic environment within a given area of interest monitoring, analysing, interpreting and rendering the results to a disabled user in a synthetic and admissible form. Like in central monitoring systems the problem of human-system interaction arises here, as well. In order to make a deaf user free of permanent observation of a visual device signalling acoustic events it seems desirable to deliver him information in two steps:

a) signalling new-detected events (if possible, with their assumed importance-level) by a vibrator or other type of tactile device to the user in order to draw his attention to a visual device,
b) delivering more detailed information about the detected and recognised event in visual or textual form.

Therefore, an animated acoustic scene can be presented to the user by a sequence of messages describing in real time the detected events. However, this description should be given in a language free of meaningless to deaf people terms, like: *clang, tinkle, buzz, whistle*, etc. It is preferred rather to use the terms indicating directly sources or events causing emission of the corresponding sounds. It is also necessary to distinguish between incidental and durable (continued) events. Therefore, a permanently actualised list of detected events, like:

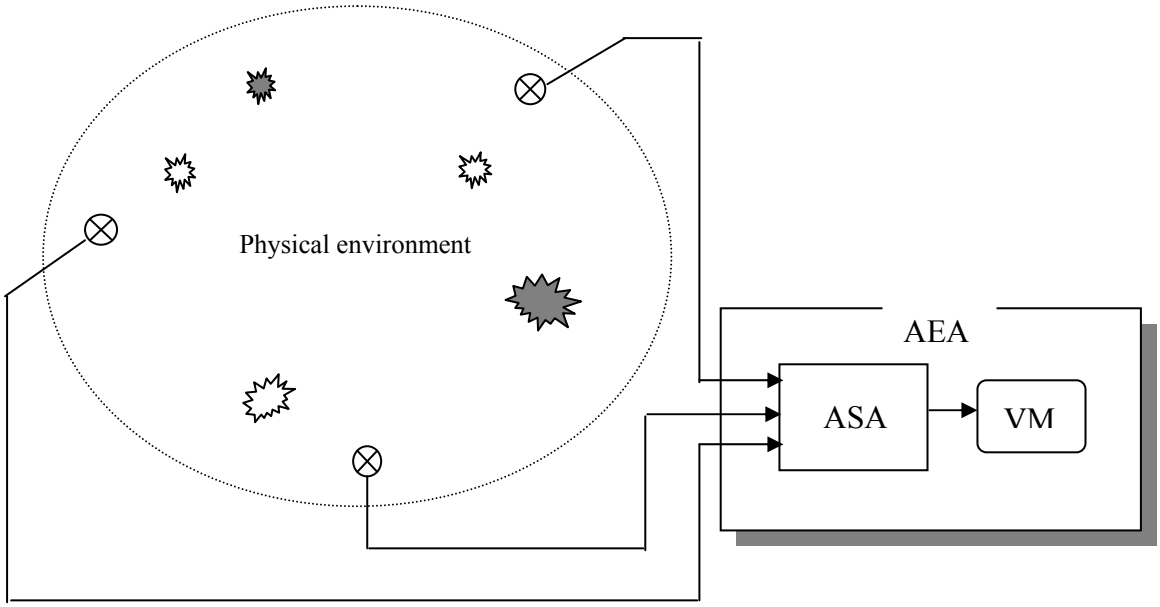
…….
A CAR PASSING THE STREET
HUMAN VOICES FROM A TV SET IN THE NEXT ROOM
RINGING A DOOR BELL
……..

at any time-instant should be presented. It represents a virtual model of animated acoustic scene making a deaf person able to focus his attention on the events of particular interest.

Till now, it is a general idea rather than an advanced project. The concept is based on sound recognition and acoustic source positioning and identification methods whose theoretical backgrounds are, generally, known [2,4]. However, some specific aspects of the problem make it a non-trivial one.

4. ACOUSTIC SOURCE IDENTIFICATION PROBLEMS

Signal source identification means source localisation and assignment to an assumed object causing signal emission. As it has been mentioned above, in AEAs acoustic source identification is a step to signal's importance-weight evaluation. A general scheme of acoustic signals acquisition in AEA is shown below.



- Active acoustic source - Non-active acoustic source - Microphone

AEA – Acoustic environment analyser, ASA – Acoustic signals analyser, VM – Visual monitor

Fig.2. A scheme of acoustic signals acquisition.

Acoustic source localisation can be based on a reverse multi-point stereophonic effect. When a signal $f_i(x_i, y_i, z_i, t)$ is emitted by a source C_i located in $(x_i, y_i, z_i,)$ then, according to (3), it is received by a microphone M_p, p=1,2,…, located in (ξ_p,η_p,ζ_p), as

$$f_p(\xi_p,\eta_p,\zeta_p, t) = K_p \cdot f_i(x_i, y_i, z_i,), t - \tau_p) \qquad (6)$$

where τ_p is given by (4). If two or more signals of this type are received by microphones and delivered to an acoustic signals analyser then differences between the time delays τ_1, τ_2, etc., can be evaluated. According to (4) the differences are proportional to the differences of the corresponding distances:

$$\delta_{pq} = r_{pi} - r_{qi} = c \cdot (\tau_p - \tau_q) \qquad (7)$$

It is well known that if A_p, A_q are two fixed points on a plane, the distance between them being δ_{pq}, then the locus of points the difference of their distances to A_p and A_q being δ_{ij} = const, when $|\delta_{ij}| \le \delta_{pq}$, is a hyperbola, as shown in Fig. 3. The section A_p - A_q is called a basis of a set of hyperbolae corresponding to various δ_{ij} satisfying the above-mentioned inequality. This fact is often used in navigational systems, where three radio-transmitters allocated in three points forming two non-co-linear bases and emitting synchronised harmonically modulated signals make possible allocation of a radio-receiver in a cross-point of two hyperbolae. In this case the differences of phases of received signals are used instead of time-delays. However, using a similar method to acoustic source localisation is, in general, impossible. For an explicit object localisation the shortest

acoustic wave length should be higher than the shortest basis δ_{pq}. Practically, the distances between microphones installed in a typical flat can not be larger than about 8 m. This means that the shortest acoustic wave-lengths that could be used in source allocation, assuming that c = 340m/s, should correspond to the frequencies not higher than 42.5Hz. The problem is that a lot of natural acoustic sources work in larger diapasons covering much higher frequencies. Explicit localisation of such sources using a phase-difference method is thus impossible.

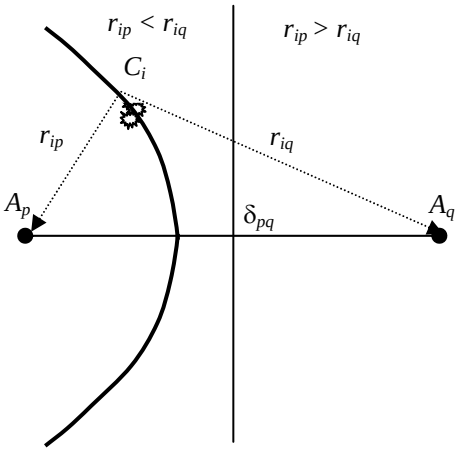


Fig.3. Principle of acoustic source localisation.

The problem can be solved if the differences $\tau_p - \tau_q$ of time-delays are used instead of the differences of phases. As it has been already mentioned, acoustic signals are usually stochastic and, even, irregular in form. For a given pair of microphones allocated in A_p and A_q the corresponding signals, in general, will have the following form:

$$f_p(\xi_p, \eta_p, \zeta_p, t) = K_p \cdot f_i(x_i, y_i, z_i, t - \tau_p) + z_p(t) \qquad (8a)$$

$$f_q(\xi_q, \eta_q, \zeta_q, t) = K_q \cdot f_i(x_i, y_i, z_i, t - \tau_q) + z_q(t) \qquad (8b)$$

where $z_p(t)$, $z_q(t)$ represent all other additional signals coming from other sources.

For evaluation the difference $\theta_{pq} = \tau_p - \tau_q$ a cross-correlation function approach can be used. For this purpose it should be calculated:

$$R_{pq}(\theta) = \frac{1}{Q_p \cdot Q_q} \int_{-T}^{T} f_p(\psi_p, t) \cdot f_q(\psi_q, t - \theta) dt \qquad (9)$$

where

$$Q_p = \int_{-T}^{T} [f_p(\psi_p, t)]^2 dt \qquad (9a)$$

(Q_q being defined in a similar way), ψ_p (ψ_q) represent the corresponding spatial co-ordinates, and T is chosen so that it covers the minimum time-duration of signal's envelope. The cross-correlation function $R_{pq}(\theta)$ should be calculated for $-\Delta_{pq} \le \theta \le \Delta_{pq}$ where $\Delta_{pq} = \delta_{pq} \cdot c$ is the time of acoustic signal propagation along the basis $A_p - A_q$.

An estimate of θ_{pq} is then given as the time of such time-shifting between the signals that maximises the cross-correlation function:

$$\theta^*_{pq} = \{\theta_{pq}: arg\ max_\theta R_{pq}(\theta)\} \qquad (10)$$

This simple idea leads to satisfactory results under the assumption that the side-signals $z_p(t)$, $z_q(t)$ are relatively small. Otherwise, $R_{pq}(\theta)$ will contain several local maxima corresponding to various acoustic sources emitting acoustic waves simultaneously. Another problem is connected with fuzziness of the cross-correlation function maxima. This, in fact, means that θ_{pq} can't be evaluated but within a finite interval. Consequently, the acoustic source can be localised within a finite area, as shown in Fig. 4. Here, as in many other navigational systems a general principle holds: the shorter is the time of signal monitoring, the larger is its localisation area. This principle is of particular interest when a moving source is to be positioned. In certain cases the amplitudes of received signals may carry additional information useful in source identification.

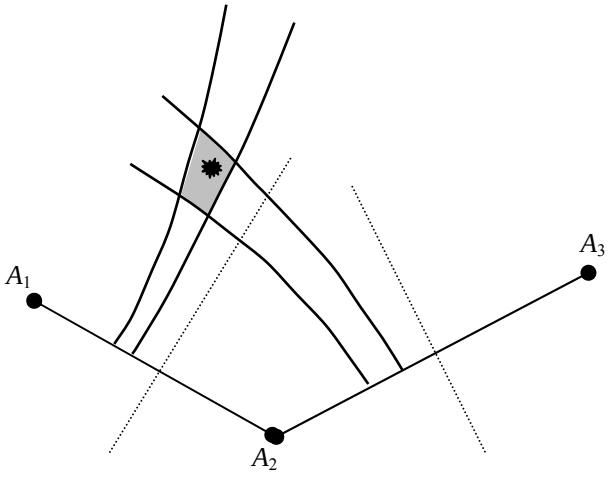


Fig. 4. Supposed area of acoustic source position.

5. CONCLUSIONS

Acoustic environment analysis is a new approach to deaf persons helping based on artificial intelligence methods: acoustic signals recognition, sources identification, environment description, etc. Realisation of this idea is also based on our observations of natural hearing system which makes us able to recognise events manifested by sounds and to localise their sources. Artificial hearing systems, by no doubt, will not be as perfect as the natural ones. However, they may help disabled persons and this is why it seems reasonable to continue efforts to their realisation.

BIBLIOGRAPHY

[1] BASZTURA C., Rozmawiać z komputerem. WPN FORMAT, Wrocław, 1993.
[2] KARLSEN B. L., Spatial localisation of speech segments. Ph. D. Thesis, Aalborg University, 1999.
[3] KULIKOWSKI J.L., Problemy komunikacji z otoczeniem i ochrony bezpieczeństwa osób niepełnosprawnych. Kr. Konferencja Naukowa „Telemedycyna II", Łódź, 2001.
[4] PULKKI V., Virtual sound source positioning using vector base amplitude panning. J. Audio Eng. Soc., vol. 45, no. 6, 1997, pp. 456-466.
[5] STARONIEWICZ P., LISIAK P., Struktura i funkcjonowanie modułu automatycznego rozpoznawania mowy opartego na HMM i diafonach. W: Komputerowe systemy rozpoznawania KOSYR (pod red. M. Kurzyńskiego i in.). Pol. Wrocławska, Wrocław, 1999.
[6] WŁOSKOWICZ D., Metody i systemy multimedialne wspomagania osób niepełnosprawnych. Opr. wewn. IBIB PAN, Warszawa, 2002.