THE ANALYSIS OF SPEECH PERCEPTION MECHANISMS
ON THE MODELS OF AUDITORY SYSTEM

V.P. BONDARENKO, V.R. MOOR, A.N. CHABANETS

Research Institute of Automatics and Electromechanics
Tomsk, USSR, 634004

ABSTRACT

This article concerns the model investigations in auditory system. The model is synthesized on the basis of a little number of the raw data, with a limited system complexity, and an element reliability. Hence, the model structure concerns as a hierarchical system with a high potential reliability.

INTRODUCTION

The analysis of speech signal is associated with model investigations in auditory. The auditory system is divided into several schemical levels: mechanical conversion level of sound signal, sensory level and neuron processing level.

The auditory system complexity depends on the signal processing level. If the spinal ganglion cosists of about 30000 neurons, then the brain consists of about I0000000 neurons. The auditory system is a complex hierarchical system. It intends to the prediction systems /I/. There are two ways of the complex system simulation: analytical and synthetical.

An analytical method is based on the determinate of majority common real system parameters and their linkages. The most models of auditory system is based on the analytical approach /2,3,4,5/. The main disadvantage of this approach lies in practical impossibility to take into account the whole information about structure and behaviour of auditory system. Hence, the models obtained are private and explain only partial effects of auditory model operation. The analysis of the complex system must not be done by independent simulation of parts. Hence, the private models can not explain the basic behavioral principles of auditory system.

The synthetical approach is more preferable. It is realised by means of optimum models synthesis, which then approximate to the real system due to nearing objective functors and linkages between model -, and system parameters. The main difficulty of using of the synthetical approach lies in the term "optimum" for the auditory model system. Due to the analysis of the common processing principles of the information in auditory system /6/, we extracted as optimum criterion - minimum losed information of the input signal by its model processing. Besides, an auditory model system should hold a high potential reliability.

PROBLEM STATEMENT

The complexity of the model structure is limited upper

$$S=\sum_{i=1}^{n} N_i \leq S_0 < \infty$$

where N_i - information capacity of the i-th element;

n - generalized number elements.

The elements have threshold of sensitivity

$$\gamma^2 pt = C \leq C_0 > 0$$

where t - reaction time; γ - error of conversion; p - input signal power.

The model elements are not reliable and are not substitute after refusal. Probability of no-failure element operation is limited upper $p \leq p_0 < 1$.

The loss information of the input signal is calculated by functor

$$\delta = \int_W [f(w) - \varphi(w)]^2 \, dw$$

where $w = \Omega \times D$ - signal region; Ω - frequency range; D - dynamical range.

$$\varphi(w) = L^*[a(f)]$$

where L - signal definition operator a(f) in model; L^* - return operator.

The task is the information probability search of the model and the structural element linkage definition. The need of loss minimization defines a decision - making task.

Frequency- and dynamic ranges are divided into subranges, with their own data model element. The common decision task is in /6,7,8/, due to this decision task was synthesized the auditory model system. This model relates to the symmetrical system class m-order /I/. Hence, the system is operationable, till its every part (involving m-elements) is working.

The probability of no-failure model operation is: $R=\sum \prod_{i=1}^{n} p_i^{\omega_i} q_i^{1-\omega_i}$

$$\sum_{i=1}^{n} \omega_i > m$$

where $w_i=0$ - for failured element; $w_i=I$ - for operatable element; p_i - probability of no-failure operation of i-th element; $q_i = 1 - p_i$.

In the present model the common element complexity is equal, hence: $p_i=p_j=p$ so

$$R=\sum_{i=m}^{n} C_n^i p^i q^{n-i}$$

The value estimations R are /I/, by $n \to \infty$, and $\frac{m}{n} < P_0$

$$R \begin{cases} > 1 - exp(-kn) \\ < 1 - exp[-kn + O(\ln n)] \end{cases}$$

where $k = \frac{m}{n} \ln \frac{m}{pn} + (1-\frac{m}{n}) \ln \frac{1-m/n}{1-p}$

Thus, by $\frac{m}{n} < P_0$, the system reliability approximates to 1.

Modes Description

The synthetized model involves: filter system, threshold elements, spaceadding linkages, and time adding filters.

The filter system has the transfer function

$$y(j,\omega,x) = exp[-1{,}44Q^2(ye^x-1)^2 - j5Qe^x y] \quad (I)$$

where x - space coordinate (filter number); Q - gain-bandwidth product of the filter system; $y=w/w_{01}$; w_{01} - resonance filter frequency with number x=0.

The threshold elements are equal - allocated along axis x in several rows. A threshold of element in i-th row is

$$a_i = a_1 \cdot \beta^{i-1}$$

where $\beta > 1$.

The element linkages are defined by functors:

$$g(x_1,t) = \int_{-\infty}^{t}\int_{0}^{x_0} h_1(t-\tau,x_1) h_2(x_1-x) L(\tau_i,x) \, dx \cdot d\tau \quad (2)$$

$$\gamma(x_1,t) = \int_{-\infty}^{t}\int_{0}^{x_0} h_3(t-\tau,x_1) h_4(x_1-x) M(\tau_i,x) \, dx \cdot d\tau \quad (3)$$

where h_1,h_3 - weight functions of the time summation; h_2,h_4 - weight functions of space summation; L,M - threshold element reactions; x_0 - upper filter number level.

The relations analyse (I7,(2),(3) has shown, that the model is not critical to value Q, since the filter time constants and weight function constants are matched. The model parameters estimation can be done by common psychoacoustical and neurodynamical data of auditory human system and by functional model analysis.

Parameter Model Improvement

There are many papers dealing with the subject parameter measurement and parameter estimate of the filter system of the auditory analyzer, but the results are contradictive /9,IO,II/.

The model described agrees well with relative levels of auditory system. It was found that the gain- band width product defines the curve of absolute sensetive level, a threshold curve type, amplitude-, and frequency modulation sensitivity, and etc. It follows, that the frequency group width (critical bandwidth) in auditory system agrees with the band-pass relative filter. Hence,

$$Q = f_{med} / f_{crit}$$

where f_{med} - medium group frequency; f_{crit} - critical frequency band for f_{med}.

Thus, gain- bandwidth product Q depends on filter resonance frequency w_{res}. Since w_{res} is a coordinate function x,

$$\omega_{res}(x) = \omega_{01} e^{-x}$$

we have, that Q is also coordinate function x.

It was found, that the gain- bandwidth means in the frequency range I-IO kc/s agrees with the artical results /IO,II/, and in range below 500 kc/s - with experimental data of Bekesy /9/.

INPUT SIGNAL CODING

Final signal description is defined with expressions (2) and (3). The functor g(x, t) defines the amplitude spectrum of the input signal, and γ (x,t) - differential from the phase spectrum.

In the bandwidth

$$\Delta\omega(x) = \frac{\omega_{res}(x)}{Q(x)}$$

g(x,t) and γ (x,t) can be represented as one count on the coordinate.

From condition (I) these counts must be taken in points of the largest value g(x, t). I.e., the restcounts it is necessary to supress by means of the suppression function ξ (t, x)

$$\xi(t,x_1) = g(t,x) - P_{x1}(t,x) \quad ; \quad (4)$$

$$P_{x1}(t,x) = \beta(x) g(x) \quad ;$$

$$\beta(x) = \begin{cases} <1 & x_1 - \Delta x \leq x \leq x_1 + \Delta x \\ >1 & x_1 - \Delta x > x > x_1 + \Delta x \end{cases} \quad ;$$

where P_x (t,x) - breakpoint in x, becoming by the excitation g (t,x) in the point x; g(x) - model reaction coused by the sine signal.

Such a function is realized on the basis of the known lateral inhibition. A making-decision procedure on the basis (4) is: if ξ (t,x) > 0, so a signal in point x_1 exists, if ξ (t,x) < 0, so - isn't.

Values g(x ,t) and γ (x ,t) - are the result of the first level of the coding of the final description. It's adaptive, since g(x ,t) and γ (x ,t) are defined by input signal structure.

CONCLUSIONS

The model described and a number of signal processing mechanisms are very common with the data mentioned in neurodynamics and psychoacoustics dealing with human sound signal sensibility. At the same time, the model is optimal as to minimum criterion of the loss information with potential reliability, near to I.

The theoretical and the experimental mode investigations provided us to study the perception mechanisms particularly; simutaneous - and sequential mechanisms of disable, to-tone suppression, the vowel attribute determination and etc.

It was found, that the sounds formants are markedly changed during the base tone period, it allows one to obtain the information about the speech signal thin structure.

REFERENCES

I. Fleishman B.S. Theory elements of potential efficiency in complex systems. - M.: Soviet.Radio, I97I. - 223 p.

2. Labutin V.K., Moltchanov A.P. Auditory models. - M.: Energija,I973.- IIO p.

3. Flanagan J.L. Analysis, synthesis and speech perception. - M.: Svjar, I968.- 489 p.

4. Pozin N.V. Neuron structure simulation. - M.: Nauka, I970. - I56 p

5. Temnov V.L. A model of the results description of psychoacoustical experiments with stationary signals. - In the book: Speech signal analysis by human. - L.: Nauka, I97I, p. I8-25.

6. Bondarenko V.P., Razin V.M. On the modeling of human system perception. - Izvestija TPI, I976, v. 266, p. 38-40.

7. Bondarenko V.P., Razin V.M. Modeling of peripherial system perception. - Izvestija TPI, I976, v. 266, p. 4I-44.

8. Bondarenko V.P., Razin V.M. Peripherial structure optimization of perception system. - Izvestija TPI, I976, v. 266, p. 45-48.

9. Bekesy J. Experiments in hearing. - N.Y., I960.

IO. Johnstone B.M., Royle A.J.F. Basilar membrane examined with mossbauer technique. - Science, I968.

II. Boer E., Jongkees N.R. Computer simulation of cochlear filtering. Proceedings of the z-th International Congress on Acoustics, vol. 3, Budapest, I97I, p. 456-460.