A digital cochlear model as a base of anthropomorphic speech processing*

Petrovsky A.A.[1,2], Likhachov D.S.[1]
1) Belarusian State University of Informatics and Radioelectronics, Computer Engineering Department, 220027, Minsk, P.Brovki st., 6 (Belarus), E-mail: likhachov@bsuir.by
2) Bialystok Technical University, Real-Time Systems Department, 15-351, Bialystok, ul. Wiejska 45A (Poland), E-mail: palex@bsuir.by

Abstract: *According to antisymmetry of basilar membrane (BM) movements, a new mathematical model of cochlea is derived using viscous cochlear fluid theory, and then transformed into a digital cochlear model with bilinear transformation. The frequency responses are found to be quite consistent with the experimental data, especially the high frequency slope is much more improved. A new cochlear map and 3 dB bandwidth characteristics for cochlear filter banks are obtained and presented, which will make applications of cochlear model more quantitative and accurate. Due to simplicity of its structure and reality of its characteristics, it will be proved the model can be used easily in speech processing system.*

Keywords**:** Cochlear model, basilar membrane, tuning cochlear filter.

1. INTRODUCTION

As a central part of auditory system, cochlea acts electrically as a highly overlapped bandpass filter bank. New results are continuously given out and knowledge of cochlea is gradually accumulated, but the mathematical model of cochlea is far from satisfaction. H.B.Allen, et al. [1,2] used Green's function method to find a fluid pressure with assumption of inviscous cochlear fluid and finally acquired a two-dimensional cochlear model represented by a integration equation. The amplitude frequency response is not very satisfied, especially low frequency slope is not as shart as desired, structure of the model appears complicated either.

Cochlear characteristics can be influenced by many fartors, but BM vibration and cochlear fluid motion should be two main factors. In the past years, much work was focused on BM properties, such as nonlinearity and activity of BM vibration [3-5], but cochlear fluid properties were paid little attention. S.Koshigoe, et al.[6] did some researches on cochlear fluid viscosity, but the work was incomplete. A two-dimensional mathematical model of cochlear with viscous fluid motion was developed in [7,8], but high frequency slope of model's amplitude frequency response is not as sharp as desired, meanwhile the time variable was not discretized either.

Based on our previous results, a new mathematical models of cochlea is developed and transformed into a digital form using bilinear transformation. The model looks much simpler structure and comes to be a typical bandpass filter. The amplitude frequency response of the model is quite consistent with the experimental data [9-11]. The cochlear map which relates the center frequency of cochlear filters with BM locations are acquired and presented in a mathematical formula, and 3 dB bandwidth characteristics which relates the 3 dB bandwidth of cochlear filters with BM locations are also formulated.

2. COCHLEAR MODEL WITH CONTINUOUS TIME AND CONTINUOUS SPACE

According to [7], viscous cochlear fluid gradient along the BM can be represented as:

$$\frac{\partial p}{\partial x}=-\rho\left[\frac{\partial u_s}{\partial t}+\gamma\left(\frac{\pi}{H}\right)^2 u_s\right]\left[1-\frac{4\pi}{H^2}\sum_{i=1}^{\infty}\frac{\sin\left(\frac{(2i-1)\pi x}{2L}\right)}{\lambda_i^2(2i-1)^2}\right] \qquad (1)$$

where ρ and γ denote fluid density and fluid viscosity, L and H denote BM length and half height, p and u_s denote fluid pressure and stapes velocity, and x and t represent spatial coordinate (from base to apex along the BM) and time coordinate respectively, $\lambda_i^2=\left[(2i-1)\frac{\pi}{2L}\right]^2+\left(\frac{\pi}{H}\right)^2$.

Due to the antisymmetry of BM movement in vertical direction [2,12], it can be assumed that fluid pressure zero in helicortema, i.e.

$$p(x,y,t)|_{x=L}=0 \qquad (2)$$

where y is spatial coordinate vertical to BM.

Integrating (1) with boundary condition (2), we can obtain a fluid pressure distribution as follows:

$$p=-\rho\left[\frac{\partial u_s}{\partial t}+\gamma\left(\frac{\pi}{H}\right)^2 u_s\right]\left[x-L+\frac{8L}{H^2}\sum_{i=1}^{\infty}\frac{\cos\left(\frac{(2i-1)\pi x}{2L}\right)}{\lambda_i^2(2i-1)^2}\right] \qquad (3)$$

In reference [7], (1) was derived with the assumption $u_s=A\sin(\omega t+\varphi)\sin\left(\frac{\pi y}{H}\right)$, i.e. input signal was assumed to be sinusoid signal. Because speech signal can be represented with a finite sinusoid series [13] (it is obvious that results given here are also applicable to any other kinds of signals which can be represented by a finite sinusoid series) and (1) was obtained by solving a linear partial differential equation [7], (3) is applicable to speech

* This work was supported in part by the Bialystok Technical University under the grant W/WI/3/03

signal according t_0 linear superposition.

From [14], it is known the following relation exists:

$$p(x,t) = -\frac{1}{2}\left[k(x)z(x,t) + r(x)\frac{\partial z(x,t)}{\partial t} + m(x)\frac{\partial^2 z(x,t)}{\partial t^2}\right] \qquad (4)$$

where $p(x,t)$ denotes fluid pressure on the BM and $z(x,t)$ BM vertical displacement, $k(x)$, $r(x)$ and $m(x)$ represent stiffness, damping and mass of BM respectively. Combining (3) with (4), we can get a continuous space and continuous time cochlear model as follows:

$$k(x)z(x,t) + r(x)\frac{\partial z(x,t)}{\partial t} + m(x)\frac{\partial^2 z(x,t)}{\partial t^2} = $$
$$= 2\rho\left[\frac{\partial u_s}{\partial t} + \gamma\left(\frac{\pi}{H}\right)^2 u_s\right]\left[x - L + \frac{8L}{H^2}\sum_{i=1}^{\infty}\frac{\cos\left(\frac{(2i-1)\pi x}{2L}\right)}{\lambda_i^2(2i-1)}\right] \qquad (5)$$

3. COCHLEAR MODEL WITH DISCRETE TIME AND DISCRETE SPACE

In eq(5), let:

$$\alpha = \gamma\left(\frac{\pi}{H}\right)^2 \qquad (6a)$$

$$F(x) = 2\rho\left[x - L + \frac{8L}{H^2}\sum_{i=1}^{\infty}\frac{\cos\left(\frac{(2i-1)\pi x}{2L}\right)}{\lambda_i^2(2i-1)^2}\right] \qquad (6b)$$

then it follows

$$k(x)z(x,t) + r(x)\frac{\partial z(x,t)}{\partial t} + m(x)\frac{\partial^2 z(x,t)}{\partial t^2} = \left(\alpha u_s + \frac{\partial u_s}{\partial t}\right)F(x) \qquad (7)$$

We first discretize space coordinate in (7). To do so, let $x = k\Delta x$, where $0 \le k \le M$ and $x = L/M$, M is the number of BM segments after discretized. (7) is transformed into the following form after discretization:

$$k_k z_k(t) + r_k\frac{dz_k(t)}{dt} + m_k\frac{d^2 z_k(t)}{dt^2} = \left(\alpha u_s + \frac{\partial u_s}{\partial t}\right)F_k \qquad (8)$$

Transfer function of (8) is

$$H_k(s) = \frac{F_k(s+\alpha)}{k_k + r_k s + m_k s^2} \qquad (9)$$

If bilinear transformation is applied to (9), i.e. let

$$p = \frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}} \qquad (10)$$

where T denotes sampling interval, then transfer function for discrete space and discrete time cochlear model can be acquired

$$\hat{H}(z) = \frac{\left[(\alpha T^2 + 2T) + 2\alpha T^2 z^{-1} + (\alpha T^2 - 2T)z^{-2}\right]F_k}{(k_k T^2 + 2r_k T + 4m_k) + (2k_k T^2 - 8m_k)z^{-1} + (k_k T^2 - 2r_k T + 4m_k)z^{-2}} \qquad (11)$$

With respect to speech signal, sampling frequency is generally no less than 8kHz, i.e. $T \le 1.25*10^{-4}$, the value of α is also no more than 10 (refer to Appendix), so all of the items with T^2 in the numerator of (11) can be omitted, and the following transfer function can be got

$$\hat{H}(z) = 2TF_k(1-z^{-2})/ \\ /\left(\begin{array}{l}(k_k T^2 + 2r_k T + 4m_k) + (2k_k T^2 - 8m_k)z^{-1} + \\ +(k_k T^2 - 2r_k T + 4m_k)z^{-2}\end{array}\right) \qquad (12)$$

It can also be put into a much simpler form as follows

$$\hat{H}(z) = A_k\frac{a_{0k}(1-z^{-2})}{1 + b_{1k}z^{-1} + b_{2k}z^{-2}} \qquad (13)$$

where

$$A_k = \frac{2TF_k}{a_0 a_{0k}}, \quad a_{0k} = \frac{1-b_{2k}}{2},$$

$$b_{ik} = b_k/a_k, \quad b_{2k} = c_k/a_k,$$

$$a_k = k_k T^2 + 2r_k T + 4m_k, \quad b_k = 2k_k T^2 - 8m_k,$$

$$c_k = k_k T^2 - 2r_k T + 4m_k.$$

Note that (13) is actually a transfer function of a bandpass digital filter. From (13), we can get amplitude frequency response (Fig.1) and phase frequency response (Fig.2) for the cochlear model.

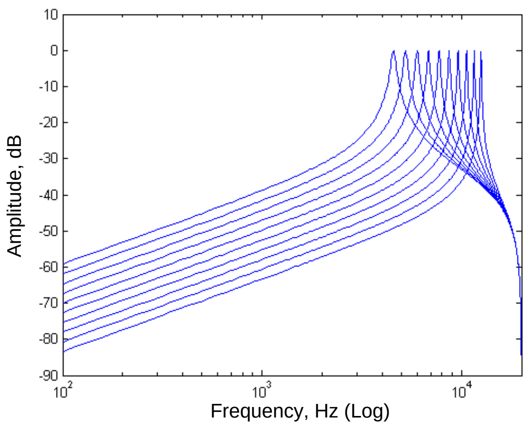

Fig.1. – Amplitude frequency responses of cochlear model for ten equally spaced points on the BM (0.6 cm to 1.63 cm from base to apex, the interval is approximately 0.12 cm)

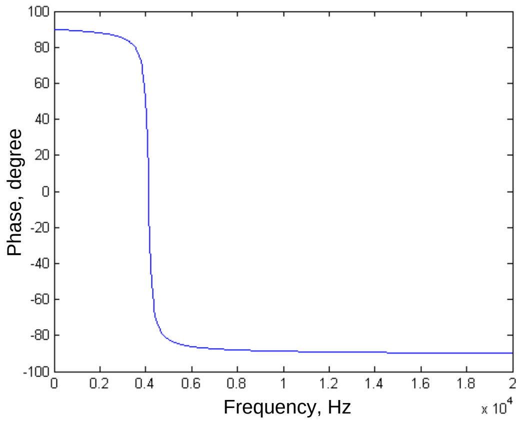

Fig.2. – Phase frequency response of cochlear model for a location near to BM apex (approximately 1.7 cm from the base)

4. COCHLEAR MAP AND 3 dB BANDWIDTH CHARACTERISTICS

From [15], we can get the following transfer function for a analog bandpass filter:

$$H(p)=\frac{(\Omega_0/Q)p}{p^2+(\Omega_0/Q)p+\Omega_0^2}=\frac{\Delta\Omega p}{p^2+\Delta\Omega p+\Omega_0^2}, \quad (14)$$

where Ω_0 is angular central frequency of analog bandpass filter, and $\Delta\Omega$ is 3 dB bandwidth of the filter. If bilinear transformation (10) is applied to (14) (note the following relation exists in bilinear transformation

$$\Omega=\frac{2}{\Delta t}tg\frac{\omega\Delta t}{2}. \quad (15)$$

where ω denotes angular frequency of digital filter), then (14) can be changed into

$$H(z)=a_0\frac{1-z^{-2}}{1+b_1z^{-1}+b_2z^{-2}} \quad (16)$$

where

$$a_0=\frac{\sin\omega_0T}{2Q+\sin\omega_0T} \quad (17a)$$

$$b_1=-\frac{4Q\cos\omega_0T}{2Q+\sin\omega_0T} \quad (17b)$$

$$b_2=\frac{2Q-\sin\omega_0T}{2Q+\sin\omega_0T}. \quad (17c)$$

$$Q=q={\omega_0}/{\Delta\omega}$$

$\omega_0=2\pi f_0$ (f_0 is central frequency of digital filter in Hz)

In bilinear transformation, if $\Delta\omega_0$ is small around the central frequency ω_0, then relation between Ω and ω can be considered to be linear, so the following formula can be obtained from (15),

$$\frac{\partial\Omega(\omega)}{\partial\omega_0}=\sec^2\frac{\omega_0\Delta t}{2},\ \sec x=\frac{1}{\cos x} \quad (18)$$

Let

$$q={\omega_0}/{\Delta\omega} \quad (19)$$

then from (15) and (18), we can get the following relation

$$Q=\frac{\sin\omega_0\Delta t}{\omega_0\Delta t}q. \quad (20)$$

Substituting (20) into (17) and using (19), we can get [16]

$$a_0=\frac{\Delta\omega T}{2+\Delta\omega T} \quad (21a)$$

$$b_1=-\frac{4\cos\omega_0T}{2+\Delta\omega T} \quad (22b)$$

$$b_2=\frac{2-\Delta\omega T}{2+\Delta\omega T} \quad (22c)$$

Let $f=\frac{f}{F_s}=fT$ (F_s is sampling frequency) which is a normalized frequency of called digital frequency, then $\omega T=2\pi fT=2\pi f=\omega$ which is digital angular frequency in radian, in this way, (21) can be changed to the following

$$a_0=\frac{\Delta\omega}{2+\Delta\omega} \quad (22a)$$

$$b_1=-\frac{4\cos\omega_0}{2+\Delta\omega} \quad (22b)$$

$$b_2=\frac{2-\Delta\omega}{2+\Delta\omega} \quad (22c)$$

It can be found there are only two independent variables in (22) because $b_2=1-2a_0$.

Comparing (13) with (16), it is not difficult to find cochlear map and 3 dB bandwidth characteristics as follows

$$\cos\omega_{0k}=\frac{-b_{1k}}{1+b_{2k}} \quad (23)$$

$$\Delta\omega_{0k}=\frac{2(1-b_{2k})}{(1+b_{2k})} \quad (24)$$

where ω_{0k} denotes central angular frequency for *k*-th cochlear filter.

Based on (23) and (24), cochlear map and 3 dB bandwidth characteristics can be calculated, and they are presented in Fig.3 and Fig.4 respectively.

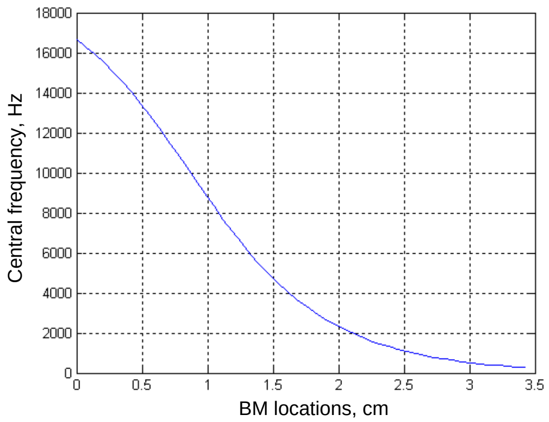

Fig.3. – Cochlear map. (Vertical coordinate denotes the central frequency of cochlear filters, and horizontal coordinate denotes BM locations from base to apex which have a biggest displacement)

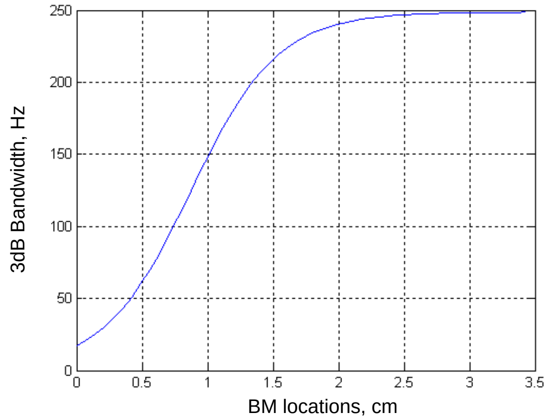

Fig.4. – 3 dB bandwidth characteristics for cochlear filter banks. (Vertical coordinate denotes 3 dB bandwidth of cochlear filters, and horizontal coordinate denotes BM locations from base to apex which have a biggest displacement)

5. TUNING COCHLEAR FILTER BANK

Let value $g = 2\cos\omega_0$ we can express (22a)-(22c) in the followiwing form [21, 22]:

$$b_1 = (a_0 - 1)g \tag{25}$$

$$b_2 = 1 - 2a_0 \tag{26}$$

So, transfer function for *k*-th cochlear filter can be described by following function:

$$H_k(z) = a_{0k}\frac{1 - z^{-2}}{1 + (a_{0k} - 1)g_k z^{-1} + (1 - 2a_{0k})z^{-2}}, \tag{27}$$

where a_{0k} and g_k – parameters, which define filter bandwidth and central frequency correspondingly.

For defined value q and central frequency ω_{0k} the coefficients are following:

$$a_{0k} = \frac{\omega_{0k}}{2q + \omega_{0k}} \tag{28}$$

$$g_k = 2\cos\omega_{0k} \tag{29}$$

For filter bandwidth $\Delta\omega_k$ and central frequency ω_{0k} the coefficients are:

$$a_{0k} = \frac{\Delta\omega_k}{2 + \Delta\omega_k} \tag{30}$$

$$g_k = 2\cos\omega_{0k} \tag{31}$$

A functioning of the *k*-th tunable cochlear filter can be described with following equations [23]:

$$\begin{aligned} w(n) &= x(n) - g_k v(n-1) + 2v(n-2), \\ v(n) &= a_{0k} w(n) + g_k v(n-1) - v(n-2), \\ y(n) &= v(n) - v(n-2), \end{aligned} \tag{32}$$

where *x(n)* – number of the input set, *y(n)* – number of the output set. So, the calculation of each element of the output set requires two multiplication and five additions.

6. DISCUSSION

In Fig.1, BM is evenly divided into 300 segments (M=300). If BM length is L=3,5 cm, then ten amplitude frequency responses in Fig.1, are associated with the BM locations whose first point is 0.6 cm from the base and the last point is 1.63 cm with the space interval being 0.12 cm. The figures presented here look much more consistent with the experimental data [10] [11], especially the high frequency slope has obviously been improved. It can also be seen that there are some differences among the peaks of curves. With the reduction of central frequency, pear value is gradually reduced, on the other hand, the peak value of curves will be gradually reduced with the increment of distance of locations from the BM base, but those differences are very small according to the computation, especially for those curves which have a high central frequency, the differences can almost be unidentified.

From the viewpoint of computation, those differences result from coefficient A_k of (13) for $A_k = \frac{F_k}{r_k}$ and r_k is assumed to be constant. From (6b), it can be found than F_k will be gradually reduced with the increment of x because there are only two items related to x, one is x itself and the other is cosine, the latter will reduce nonlinearly with the increment of x, the reduction of cosine is faster than increment of x, especially when x becomes large.

From the other viewpoint, it seems that BM displacement will reduce with reduction of excitement frequency (assume excitement amplitude is invariant). Because low frequency signals excite BM apex, so it has to travel a longer distance along the BM and degradation seems to take place.

In Fig.2, a phase frequency response is presented. It is associated with a point near to BM apex, approximately

3.3 cm from the base (stapes). It is almost a linear function in linear frequency scale.

Cochlear map is defined to be a relation between excitement frequency or central frequency of cochlear filter and BM location which has a biggest displacement, excitement with different frequency excites a biggest displacement in different BM locations. From Fig.3, we find that relation between excitement frequency and location with biggest displacement is not exactly linear, that is to say, for those cochlear filters equally spaced on the BM, the differences between their central frequencies are not equally spaced in a linear frequency coordinate.

Cochlear map is conventionally deduced from experimental data [17]-[20], here it is deduced directly from the mathematical model of cochlear, and the result is quite consistent with the experimental data.

3 dB bandwidth characteristics is defined to be a relation between 3 dB bandwidth of cochlear filters and BM locations with the biggest displacement. From Fig.4, we can see that bandwidth of cochlear filters is almost increased linearly with the increment of distance from BM base to BM center, but from center to apex, bandwidth is almost invariant. It seems that we can use a filter banks with same bandwidth to process speech signal for speech signal excites BM locations from center to apex.

It is believed that this result will make the application of cochlear model more quantitative and accurate.

7. AN EXAMPLE OF IMPLEMENTATION COCHLEAR MODEL FOR SPEECH CODING

Speech analysis process can be represented as shown in Fig.5 [24]. The approach considered here involves main ideas from papers [25, 26]. The human auditory system model is used for improving spectral peaks selection.

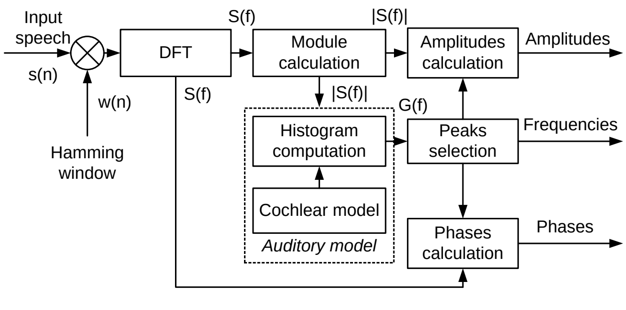

Fig. 5. – Speech analysis scheme

Speech synthesis procedure presumes generating sinusoids according to frequency and phase parameters, weighed by amplitude and then summed to produce a frame of synthesized speech as shown in Fig.6. The resulting speech signal is formed by overlapping and adding each frame of the synthesized speech.

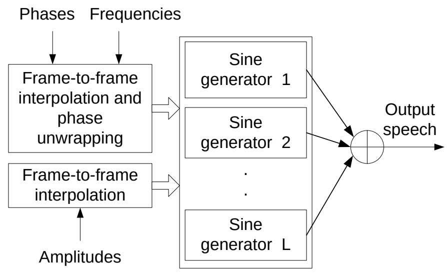

Fig. 6. – Speech synthesis scheme

8. CONCLUSION

Based on previous cochlear mechanics put forward by the authors, a news cochlear model is obtained using both antisymmetry of BM movement and theory that speech signal can be represented by a finite sinusoid series. By using bilinear transformation, the model is changed into a second-order digital cochlear model, that is to say, cochlea can be modeled with a second-order difference equation. The frequency slope is sharpened and thus looks more realistic.

According to characteristics of bilinear transformation, a new cochlear map and 3 dB bandwidth characteristics are acquired. The cochlear map is also quite consistent with the experimental data, except in the BM apex. The cochlear model put forward in this paper is believed to simple in structure, and therefore easy to apply, but it is far from complete and much work can be done because BM nonlinearity and other special properties of BM is not included, this model makes such inclusion feasible.

9. REFERENCES

[1] B.Allen, "Two-dimensional cochlear fluid model: new results", J.Acoust. Soc. Am. 61., 1971, pp.110-119.

[2] B.Allen, M.M.Sondhi, "Cochlear macromechanics: time domain solutions", J.Acoust. Soc. Am. 66., 1979, pp.123-132.

[3] R.Diependaal, "Nonlinear and active cochlear models: analysis and solution methods", Ph.D.thejsis, Delft University, 1988.

[4] J.W.Matthews, "Mechanical modeling of nonlinear phenomena observed in the peripheral auditory system", Ph.D.thesis, Washington University, St.Louis, MD, 1980.

[5] S.T.Neely, D.O.Kim, "A model for active elements in cochlear biomechanics", J.Acoust. Soc. Am. 79, 1986, pp.1472-1480.

[6] S.Koshigoe, W.K.Kwok, and A.Tubis, "Effects of perilymph viscosity on low-frequency intracochlear pressures and the cochlear input impedance of the cat", J.Acoust. Soc. Am. 74, 1983, pp.486-492.

[7] W.G.Wan, C.X.Fan, "A new solution for cochlear macromechanics", Acustica 75, 1991, pp.79-82.

[8] W.G.Wan, C.X.Fan, "A new cochlear model: viscous fluid motion", J.Acoust. Soc. Am. 89, 1991, pp.1865.

[9] L.Robles, M.A.Ruggero, N.C.Rich, "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves and response phases", J.Acoust. Soc. Am. 80., 1986, pp.1364-1374.

[10] W.S.Rhode, "Observation of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique", J.Accoust. Soc. Am. 49, 1971, pp.1218-1231.
[11] W.S.Rhode, L.Robles, "Evidence from Mossbauer experiments for nonlinear vibrations in the cochlea", J.Accoust. Soc. Am. 55, 1974, pp.588-596.
[12] J.B.Allen, "Cochlear modeling", IEEE ASSP Magazine 1, 1985, pp.3-29.
[13] R.J.McAulay, T.F.Quatieni, "Speech analysis/synthesis based on a sinusoidal representations", IEEE Trans. on ASSP 34, 1986, pp.744-754.
[14] Monderer, "Exploring the space-time structure at the output of a cochlear model", Ph.D. thesis, Columbia University, New York, 1988.
[15] D.Johnson, J.Johnson, and H.Moore, A bandbook of active filters, Prentice-Hall, Inc. Englewood Gliffs, N.J., 1980.
[16] A.A.Petrovsky, "Methods and microprocessor means of processing broadband and fastflowing processes in real time", Nauka i Tekhnika, Minsk, 1988.
[17] S.Puria, J.B.Allen, "A parametric study of cochlear input impendance", J.Acoust. Soc. Am. 89, 1991, pp.287-309.
[18] D.H.Eldredge, J.D.Miller, and D.A.Bohne, "A frequency-position map for the chinchilla cochlea", J.Acoust. Soc. Am. 69, 1981, pp.1091-1095.
[19] D.Greenwood, "A cochlear frequency-position function for several species -- 29 years later", J.Acoust. Soc. Am. 87, 1990, pp.2592-2605.
[20] M.C.Liberman, "The cochlear frequency map for the cat: Labeling auditory nerve fibers of known characteristics frequency", J.Acoust. Soc. Am. 72, 1982, pp.1441-1449.
[21] A.A.Petrovsky, J.A. Ganushkin. “Syntheses of digital bandpass filters. Radiotechnika”, Radiotekh. Electronik, 1985, vol. 10, pp. 24-25, 1985.
[22] A.Petrovsky. “The synthesis of high order digital bandpass filter with tunable center frequency and bandwidth”, VIII Europ. Sig. Proc. Conf. (EUSIPCO 96). Trieste, Italy, 10-13 September 1996, pp. 1527-1530.
[23] Alexander A. Petrovsky, Jaroslaw Baszun. “Design of Digital Tunable Bandpass Filters: Synthethis and Applications”, Instytut Informatyki, Politechnika Bialostocka, Kwartalnik electroniki i telekomunikacji, 2001, pp. 5-25.
[24] Likhachov D.S., Petrovsky A.A. “Improved auditory-based speech coding using psychoacoustic model based on a cochlear filter bank and an average localized synchrony detection”, CISIM’2003, pp.11-19.
[25] A. M. A. Ali et al., “Robust auditory-based speech processing using the average localized synchrony detection”, IEEE Transactions on Speech and Audio Processing, vol.10, no.5, July 2002, pp.1447-1459.
[26] Oded Ghitza, “Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition”, IEEE Transactions on Speech and Audio Processing, vol.2, no.1, part II, 1994 - pp. 115-132.