Ann.Bull.RILP(1986)No.20, 83-102

A NEURO-SYNAPTIC MODEL OF AUDITORY MEMORY AND PITCH PERCEPTION

Kenji Itoh

Summary

A model of temporal information processing in the auditory nervous system was presented on the basis of the interaction of peripheral fast synaptic potentials as distortion products as well as of slow synaptic potentials as auditory memories in order to analyze the mechanism of pitch perception for complex sounds. The system was composed of the cochlea, including the inner and outer hair cells, cochlear nucleus, inferior colliculus, medial geniculate body and primary auditory cortex. The model system could extract the pitches of combination tones of complex sounds using the interaction of in trains of fast synaptic potentials from outer cells with resonance frequencies of the primary tones. The extraction mechanism of pitches from the temporal information in neural activities was discussed in four process, i.e., the mechanical-to-neural tansduction of the acoustic signal, the extraction of the periodicity in the auditory nerves, the temporal to spatial translation of nerve firings and the determination of pitches, in relation to other phenomena in the perception of complex tones, e.g., hearing of AM sounds, pitch sifts caused by maskers, roughness and tonal dissonance.

1. Introduction

The pitch perception of complex sound is a fundamental basis of aural communication in humans as well as animals[1]. When we hear a complex sounds, some additional low pitches as well as the pitches of primary tones are perceived[2]. The most prominent additional pitches in a complex sound of two components are the quadratic difference tone and the cubic differnce tone (CDT)[3].

The two difference tones are perceived as pitches of f_2-f_1 and $2f_1-f_2$, respectively, in a complex tone with the lower frequency component f_1 and the higher frequency component f_2. As the ratio of f_2 to f_1, increases toward 2.0 from 1.0, the CDT, as well as the quadratic difference tone, decreases rapidly at high frequencies, or slowly at low frequencies[4]. The differnce tones can be cancelled by adding a probe tone of the appropriate level and phase[5]. The difference tones can be derived mathematically from the cubic distortion product of f_1+f_2[6]. The strength of primary tone itself reduces in the complex tone[7].

The origin of such distortion has been studied in the auditory system from the peripheral to the central levels, e.g., nonlinear transmissions in the middle ear, nonlinear wave propagation on the basilar membrane[8], cochlear microphonics (CMs) in hair cells[9] and excitation pattern overlaps in critical bands[10]. Recently, Strube (1986)[11] has proposed a model of the basilar membrane with a nonlinear resonance quality to probe the basilar membrane theory of combination tone (Zwicker, 1982)[12].

In this paper, a model for the generation of the combination tones will be examined at the neuro-synaptic level to observe the mechanism of pitch perception for complex sounds on the basis of the pulse-analogue circuit model of information processing in the auditory system [13].

2. The Model

2.1 Auditory System

The present model, as given in Fig.1, consists of the cochlea (CL), the cochlear nuclkeus (CN), the inferior colliculus (IC), the medial geniculate body (MG), and the auditory cortex (AC) with reciprocal connections with the auditory reticularis thalami (AR). The CL, which is a mechanical-to-neural transducer, outputs through afferent nerves to the CN. The IC intermediates the CN output to the MG, which project afferently to the AC. The MG innervates efferently on the IC which has efferent connections with the SO. The SO outputs through "olivo-cochlear bundles" to the CL.

2.2 Cochlea

Figure 2 gives a simplified cell circuit model of the CL as a pai of arrays composed of two types of hair cells, i.e., outer hair cells (OHCs) and the inner hair cells (IHCs). Each hair cell in both the OHC and IHC arrays has its own characteristic frequency, which as a monotonous function of the position in the array due to the sharp resonance of its hairs [14] as well as the broad resonance of the basilar membrane [15]. A group of OHCs are innervated by an afferent spiral fiber which terminates mainly on the endings of afferent radial fibers attached to an IHC with a lower characteristic frequency than the OHCs. Each OHC is innervated by a bundle of efferent radial fibers [16].

2.3 Performance of the Model System

The model system deals with pitch information in four processes: the mechanical-to-neural transduction of the acoustic signal (N-process), the extraction of the periodicity in the nerves (R-process), the temporal to spatial translation of nerve firings (S-process) and the determination of pitches (P-process).

These four processes are performed as follows.

1) Each OHC generates a train of fast post-synaptic potentials (f-PSPs) at the endings of the spiral afferent fibers when the CM analogous to the component of the characteristic frequency filtered from the input signal exceeds its own threshold which may be modified by slow inhibitory PSPs from the radial efferent fibers [17] [18].

2) The spiral afferent fibers conduct the combined PSP wave of such f-PSP trains from the OHCs with different characteristic frequencies within the group to the endings of radial afferent fibers.

3) The radiant afferent fibers generates pulse trains with a density corresponding to the combined PSP wave and to a PSP

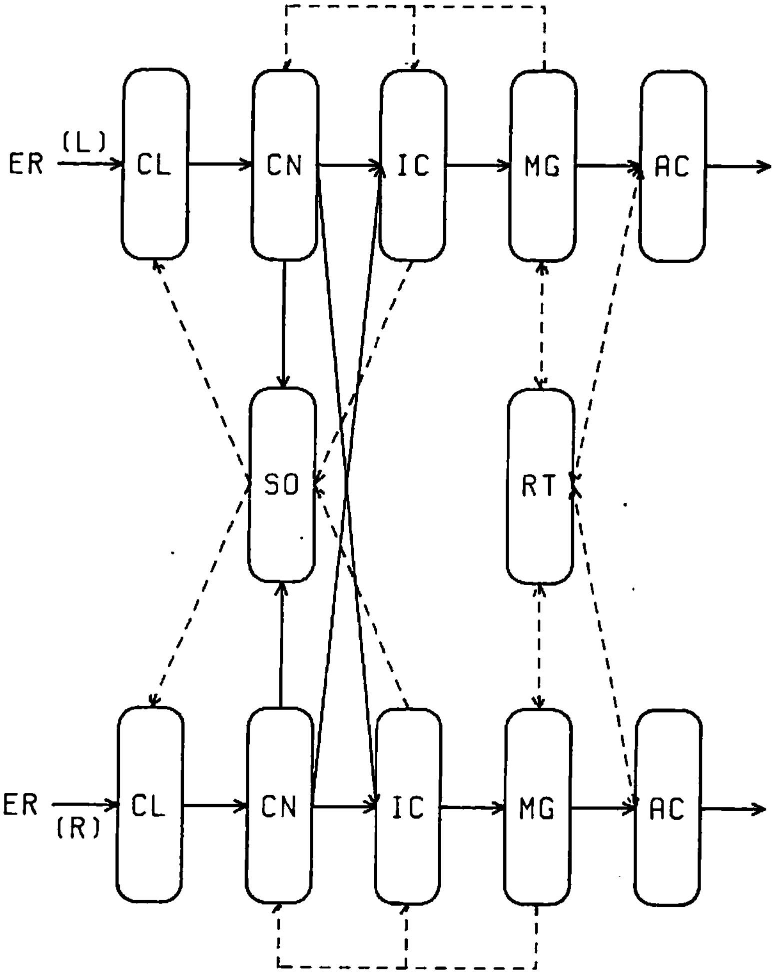

Fig.1. Blockdiagram of the auditory system for pitch perception. AC: auditory cortex; CL: cochlea; CN: cochlear nucleus; ER: ear input; IC: inferior colliculus; MG: medial geniculate body; RT: auditory reticularis thalami; SO: superior olive. Solid lines: afferent pathways; broken lines: efferent pathways.

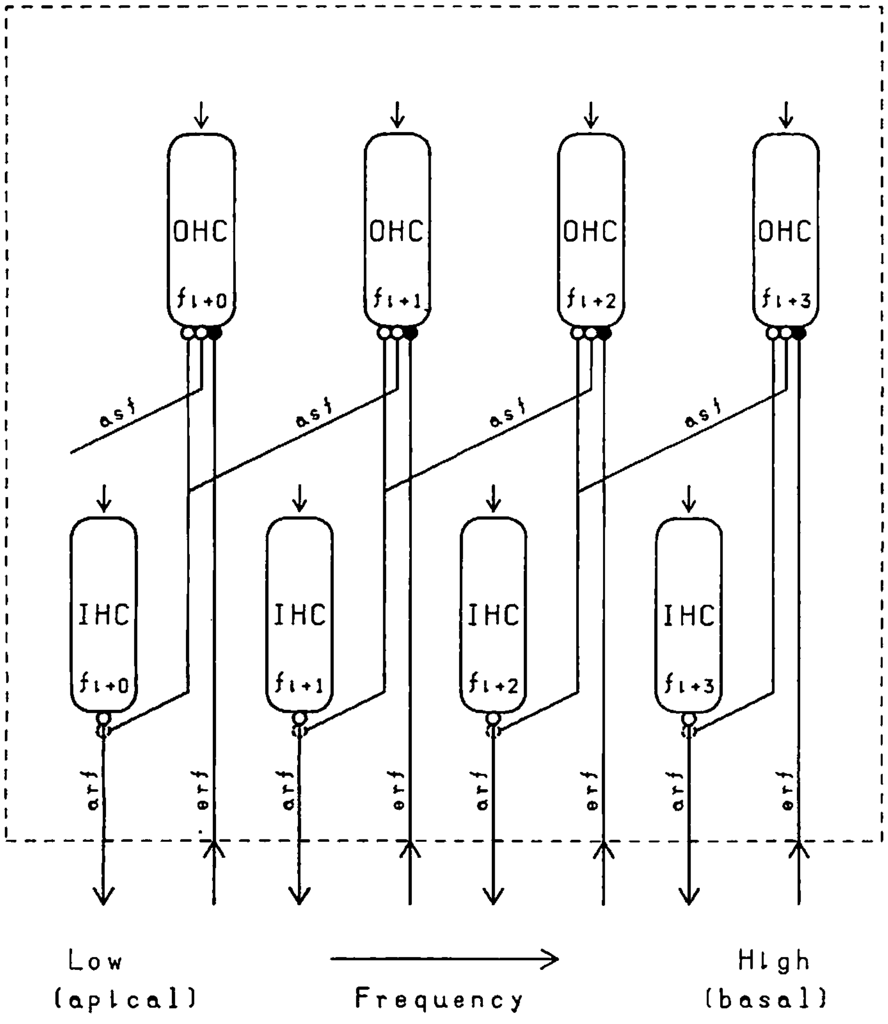

Fig.2. Schema of the cell circuits in the cochlea. IHC: Inner hair cell; OHC: outer hair cell; arf: afferent radial fiber; asf: afferent spiral fiber; erf: efferent radial fiber; open circle: excitatory synapse; filled circle: inhibitory synapse.

wave from a IHC which generates a train of pulses of f-PSPs when the CM analogous to a (rectified) superimposed wave of the lower components in the input signal exceeds its own threshold (N-process) [19].

4) The CN extracts the periodicities in the pulse train sent from the CL (R-process) [20].

5) The data on such periodicities are sent to the IC, where the IC neurons integrate the data within their sharing frequency region [21].

6) The IC output of the temporal firing pattern is sent to the MG, where such temporal firing pulses generate slow post-synaptic potentials (s-PSPs) which are integrated to convert the temporal pattern to a spatial pattern in auditory memory (S-process) [22].

7) The MG output for the spectral information in the auditory memory is projected to the AC, which determines the pitch frequencies and intensities by picking the local peaks in a spatial domain (P-process) [23].

3. Simulation

In order to examine the performance of the model as explained in the previous section with real data, a simulation was performed with the following parameters.

3.1 Post synaptic potential

The time course unit for PSPs, V(t) is simplified in the model as follows [13],

$$V(t)=\begin{cases}(1-\cos(2\pi t/2Tc))/2 & (0\leq t<Tc)\\ \exp(-(t-Tc)/\tau) & (Tc\leq t)\end{cases} \tag{1}$$

where Tc is the duration of the conductance change and τ is the discharging time constant.

When the PSP is generated repetitively, the PSP is

$$V(t)=\begin{cases}(1-\cos(2\pi t/2Tc))/2 & (0\leq t<Tc)\\ (\exp(-(t-Tc)/\tau)-Vo)/(1-Vo) & (Tc\leq t<T)\end{cases} \tag{2}$$

$$Vo=\exp(-(T-Tc)/\tau) \tag{3}$$

$$V(t+T)=V(t) \tag{4}$$

where T is the period of the repetitive stimuli.

3.2 Combination Tone

In the present model, the distortion products for the difference tones are not extracted from a cubic function, but from some interactions (beats) between one of the harmonics of the repetitive f-PSPs caused by the lower primary tone and the funda-

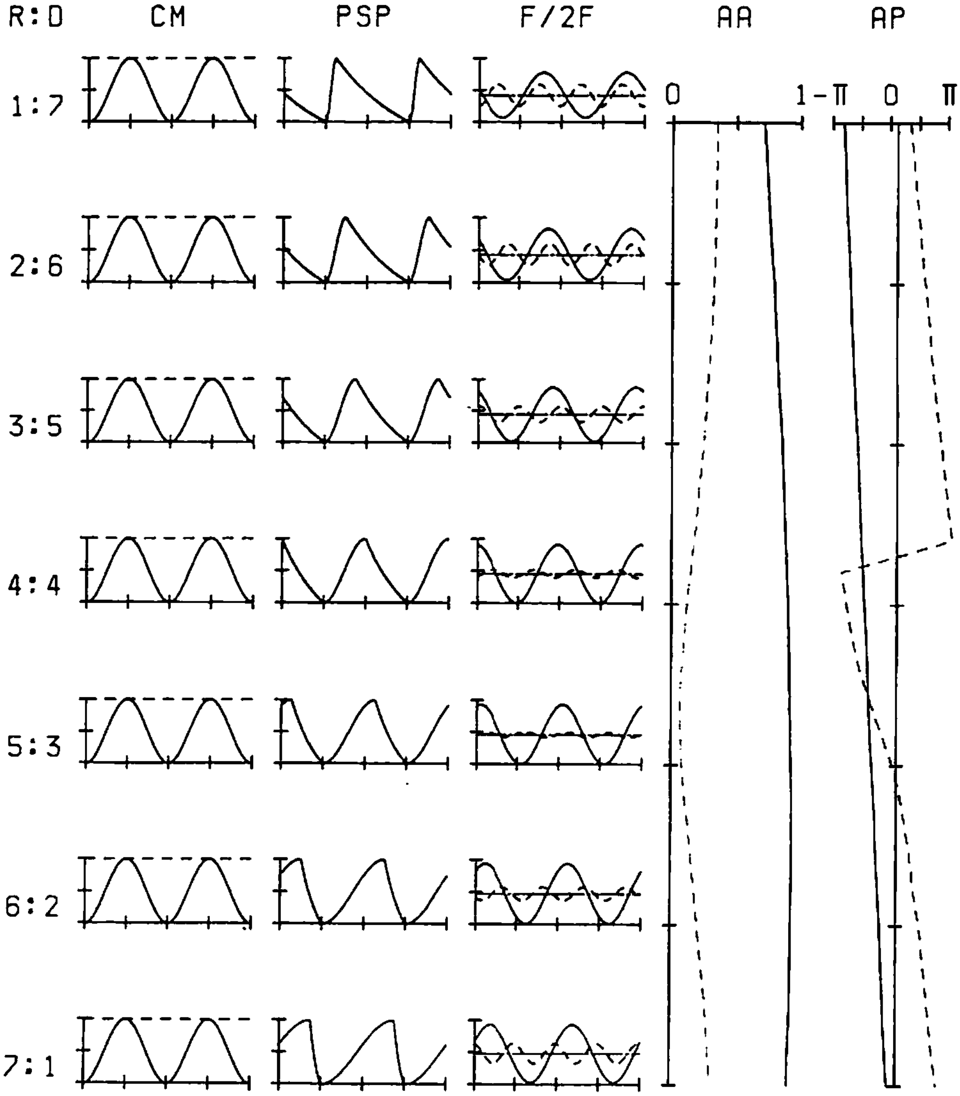

Fig.3. Unit PSPs. AA: amplitudes of fundamental (solid line) and first-order harmonic (broken line) components; AP: phases of fundamental and first-order harmoonic components. CM: cochlear microphnics; F/2F: fundamental and first-order harmonic components; PSP: post-synaptic potentials; R:D: ratio of PSP charging and discharging duration; broken line in CM: threshold of cochlear microphonics.

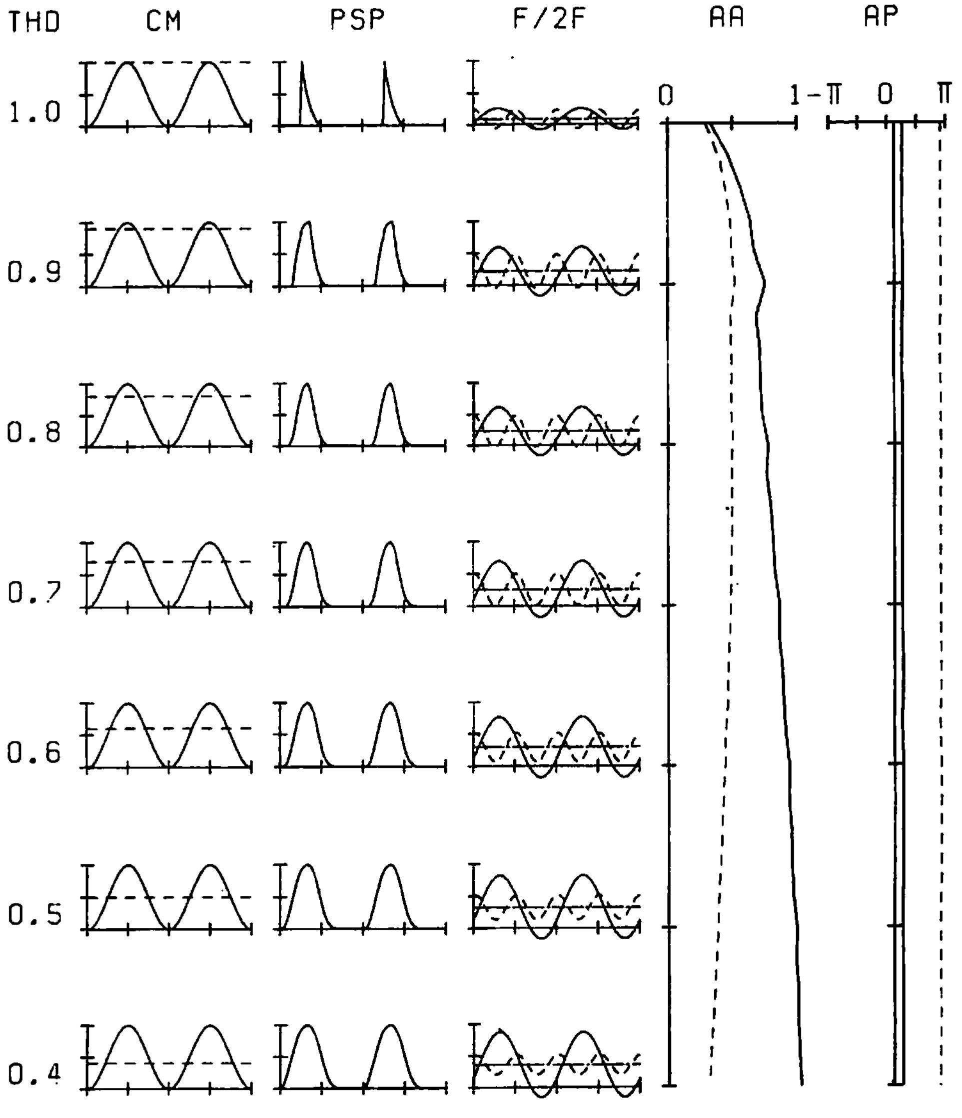

Fig.4. Combined PSPs. THD: relative threshold. PSP rise time Tc = 120 microsec; PSP discharging time = 480 microsec. (a) f = 250 Hz; (b) f = 1000 Hz; (c) f = 2500 Hz.

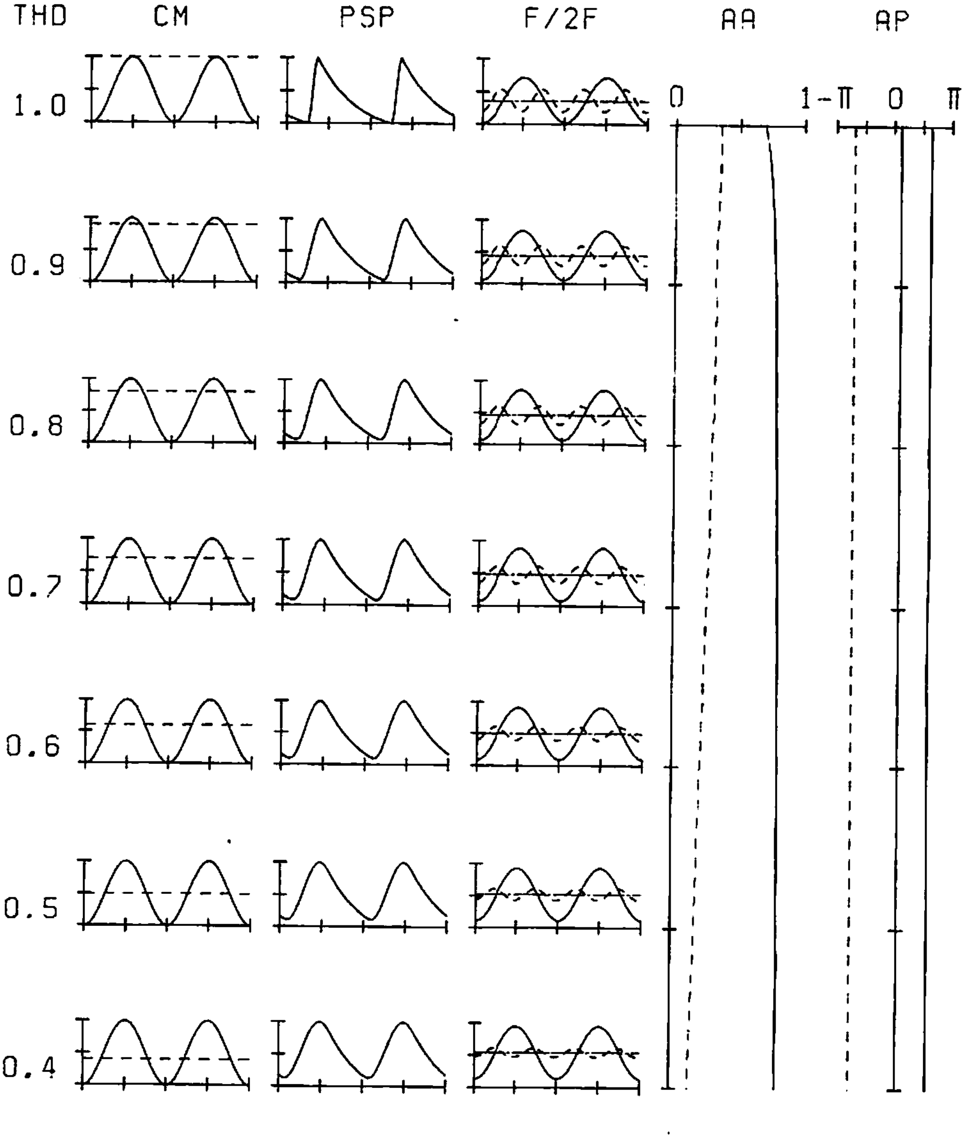

Fig.4(b). f = 1000 Hz.

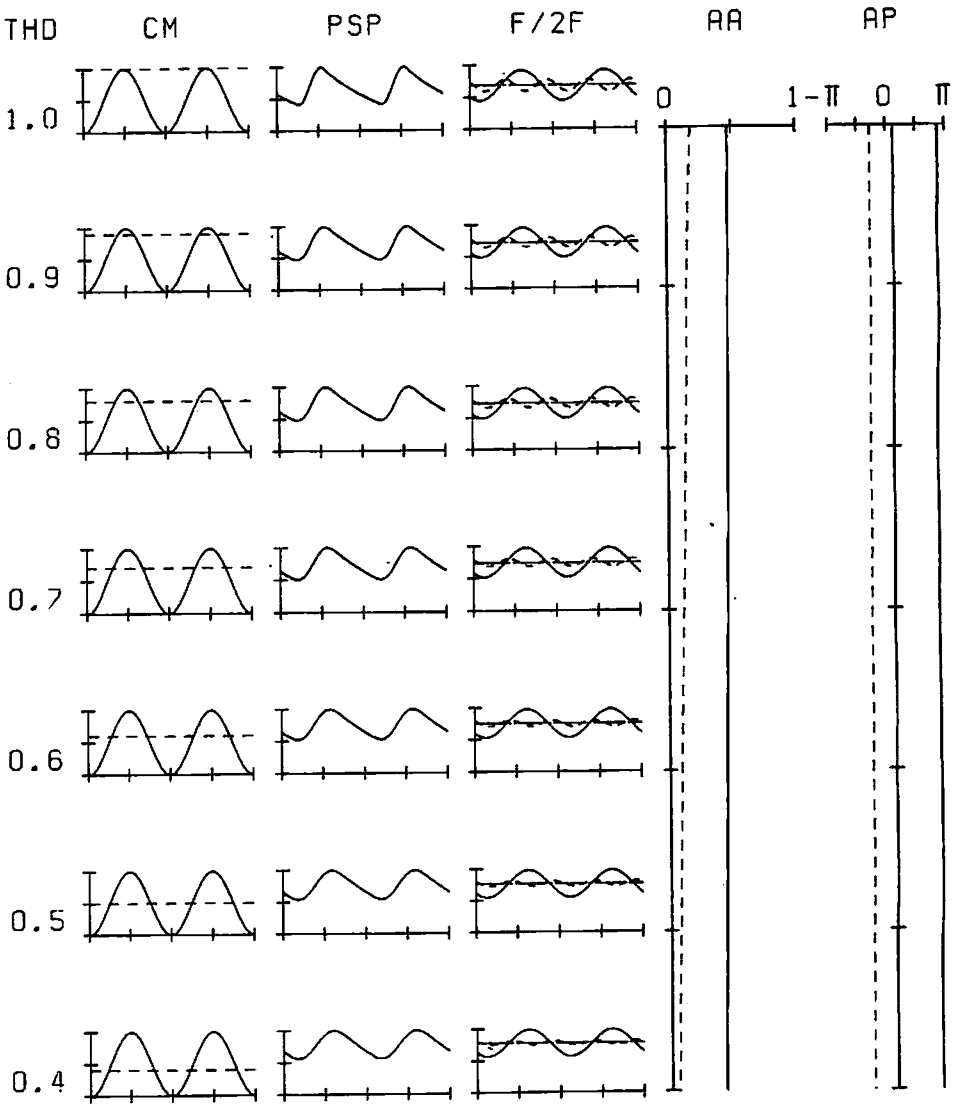

Fig.4(c). f = 2500 Hz.

mental component or one of the harmonics of the repetitive f-PSPs caused by the higher primary tone. When the repetitive f-PSPs of the lower and higher primary tones, P_1 and P_2, are represented as follows,

$$P_1 = a_{10}p_{10} + a_{11}p_{11} + a_{12}p_{12} + \ldots \quad (5)$$

$$P_2 = a_{20}p_{20} + a_{21}p_{21} + a_{22}p_{22} + \ldots \quad (6)$$

where $a_{10}p_{10}/a_{20}p_{20}$, $a_{11}p_{11}/a_{21}p_{21}$ and $a_{12}p_{12}/a_{22}p_{22}$ are the fundamental component, first order harmonics and second order harmonics, respectively, combinations of the higher components in either P_1 or P_2 and the lower components in the other P_1 or P_2 can generate a beat with the frequency of the difference between the two components. Some examples follow

$$k_2a_{20}p_{20} - k_1a_{10}p_{10} \quad (7)$$

$$k_1a_{11}p_{11} - k_2a_{20}p_{20} \quad (8)$$

$$k_1a_{12}p_{12} - k_2a_{21}p_{21} \quad (9)$$

Here, k_1 and k_2 are the frequency-dependent coefficients of the synaptic contacts of the spiral afferent fibers in the OHCs for P_1 and P_2, respectively. The first and second example above correspond to the quadratic difference tone $f_2 - f_1$ and the CDT $2f_1 - f_2$, respectively. Other combinations of the two components can generate distortion products different from the above examples.

4. Results

4.1 Harmonics of the PSPs

Figure 3 shows the fundamental as well as the first order harmonic of the PSPs with various charging durations when the OHC/IHC threshold is fixed at the peak level of the CM. Figure 4 shows the fundamental as well as the first order harmonic of the combined PSPs with various CM thresholds when the f-PSP charging duration was fixed at three levels of .04T, .12T and .3T.

4.2 Combination Tones

Figure 5 shows examples of the system responses when the complex tones of two sinusoids were added. In the figure, pairs of local peaks, with the period of the difference tones and pairs of the primary tones are shown. In Fig. 6, the changes in the system response are indicated when the ratio of f_2/f_1 increases from 1.1 to 1.7 in the case of f_1 = 1000 Hz(levels of two tones: 100%). In Fig. 6, the changes in the system response are indicated when the intensity of f_2 decreases from the level of f_1 in the case of $f_2/f_1 = 1.1$.

4.3 Cancellation of the Combination tone

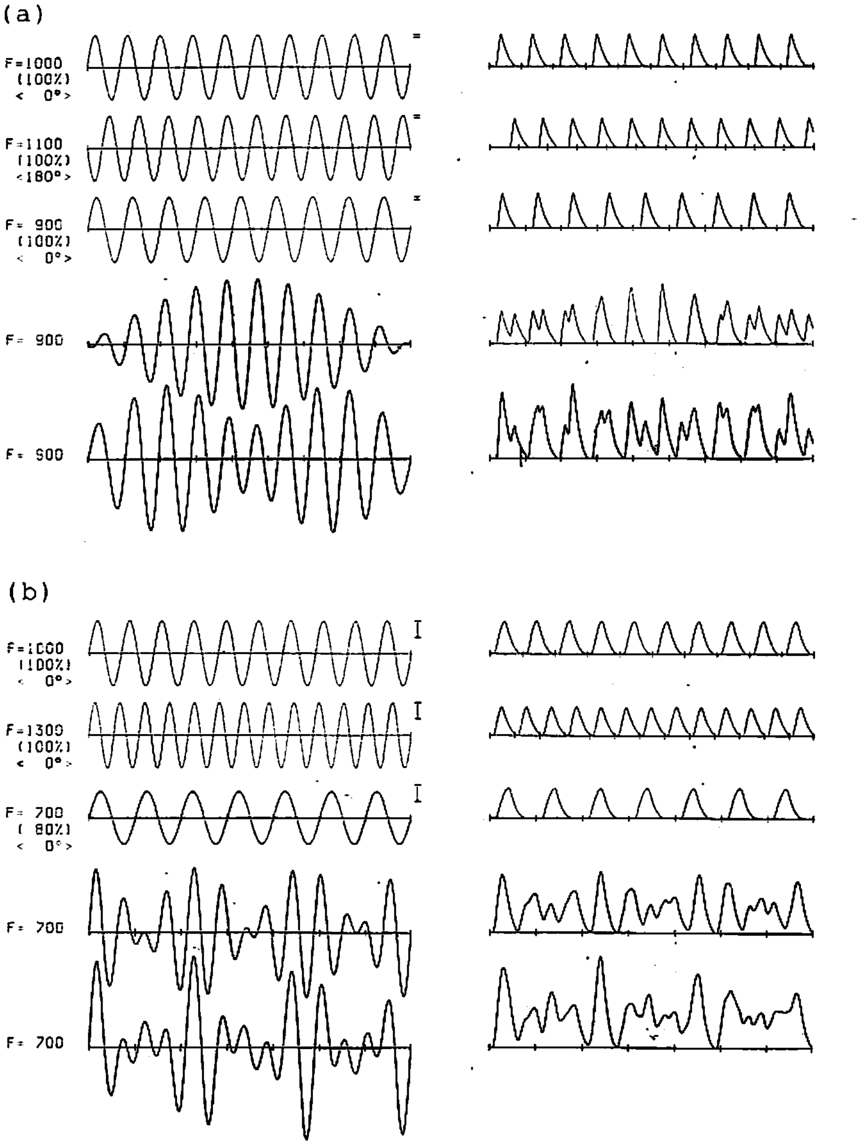

Fig.5. Combination tone and the combined PSP. Left column: original waves of primary tones and the complex tones; right column: combined PSPs for primary tones and superimposed PSPs of the combined PSPs. Lower two traces are for the complex tones of two components and three components, respectively. (a) f2/f1 = 1100/1000 Hz, (b) f2/f1 = 1300/1000 Hz.

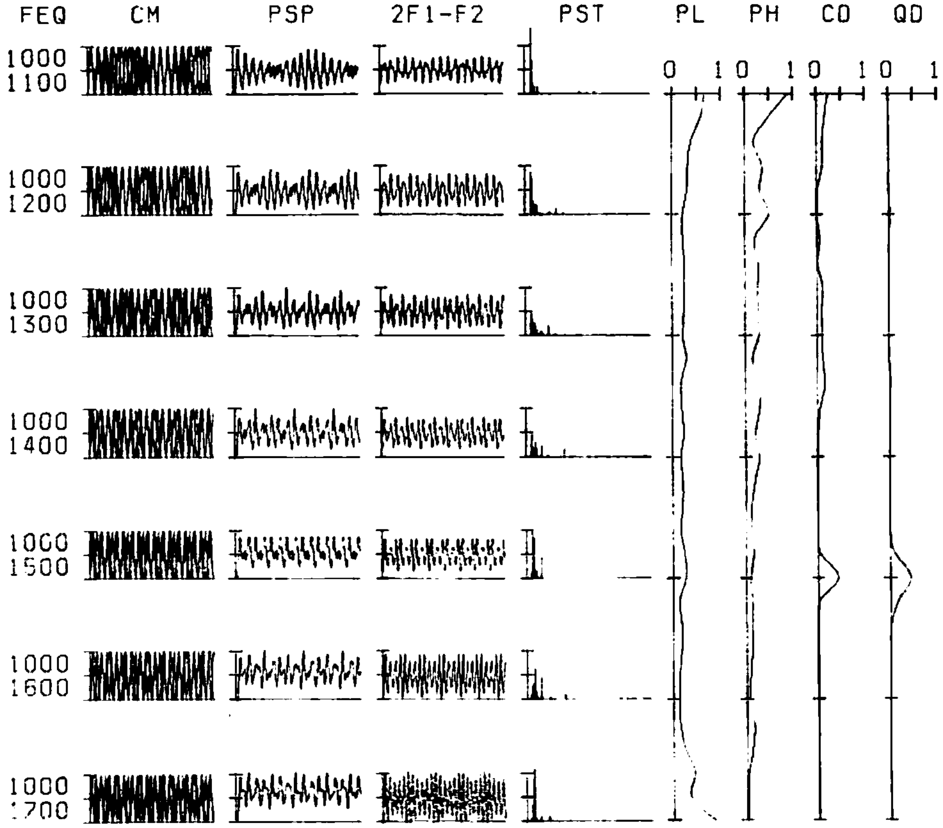

Fig.6. System responses. CM: cochlear microphonics; PSP: superimposed PSP wave of two combined PSPs for primary tones; 2F1-F2: sum of first-order harmonic component of the combined PSP for lower primary tone and fundamental component of the combined PSP for higher primary tone; PST: peak interval histograms of the superimposed PSPs for the complex tone (upward bars) and for the complex tone with cancellation tone. PL: intensity at the lower primary tone; PH: intensity at the higher primary tone; CD: intensity at the cubic difference tone; QD: intensity at the quadratic difference tone in the PST. Solid line: data for complex tone; broken line: data for complex tone with cancellation tone. (a) response for various higher primary tones (lower primary tone = 1000 Hz).

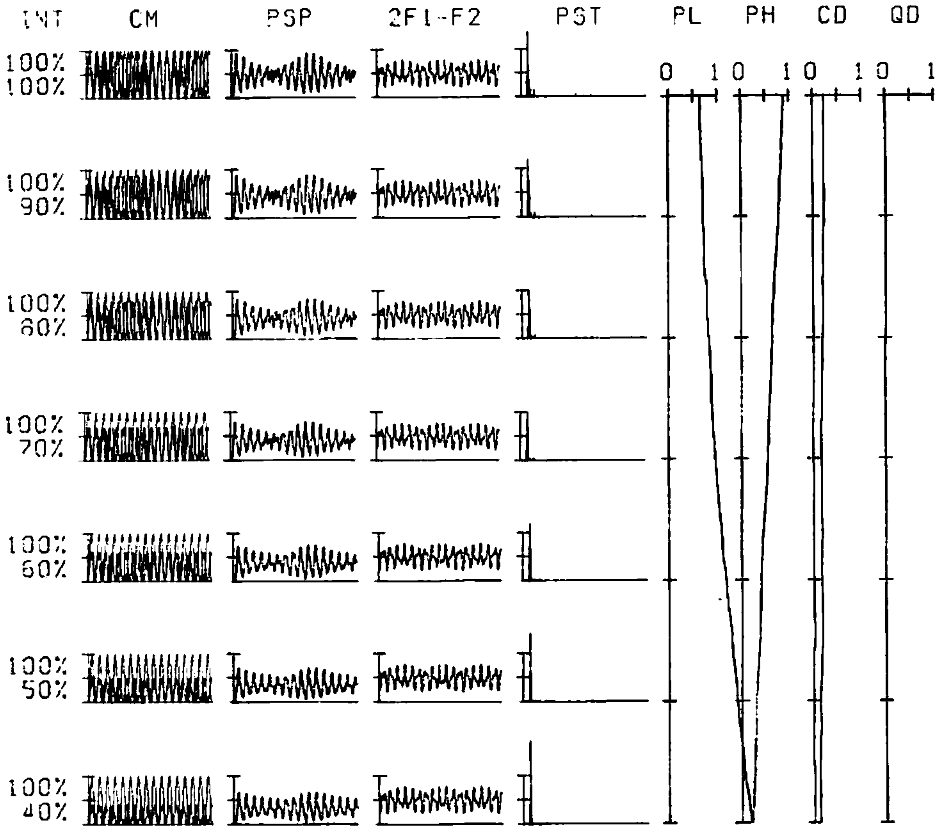

Fig.6(b). System responses for higher primary tones with various intensities relative to lower primary tone ($f_1/f_2 = 1.1$; $f_1 = 1000$ Hz).

Figure 7 shows an example of the system response when a cancellation tone with the frequency of the combination tone and a complex tone are added. In the figure, the estimated combination tone as the beat of $k_1a_{11}p_{11} - k_2a_{20}p_{20}$ is also illustrated.

5. Discussion

A model of temporal information processing in the auditory system was presented on the basis of interactions among post synaptic potentials in order to analyze the mechanism of pitch perception especially for complex tones. The model could simulate the pitch extraction process for complex tones of two components and the cancellation process of combination tone by adding an appropriate tone.

The origin of distortion products, such as the quadratic and cubic difference tones, has been investigated physiologically in various portions of the auditory system. However, the cochlear microphnics do not show periodicity at the frequency of the cubic difference tone, except at the third turn of the cochlea (Dallos 1969)[24]. The basilar membrane can not be observed to vibrate at the frequency of the combination tone (Wilson & Johnstone 1973)[25]. These data indicate that the nonlinearity of distortion products emerges after the cochlear microphonics at least for high frequency tones.

Goldstein and Kiang (1968)[26] recorded the activities of auditory nerves when complex tones were added. They observed periodicities in the firing pattern of these nerves at the characteristic frequency of the combination tone if the tone was lower than 5 kHz. Rose et al (1974)[27] have observed some phase-sensitive neurons in the anteroventral cochlear nucleus of the cat. The synaptic interaction of the auditory nerves presented in our model occurs between the cochlear microphonics and the firings of the auditory nerves.

In our model, there is some spatial summation of the OHC activities, but the major afferent fibers innervate on IHCs without branching. If the IHCs are also involved in the process of complex tones, the mechanism of the interaction is among the components of the complex tone. For lower frequencies (less than 500 Hz) the superposition of cochlear microphonics corresponding to these components is possible because the tuning curves of the cochlear microphonics are broadened at low frequencies[28]. If the IHCs also play a role in generating such distortion products, some improvement of the coordination between the IHCs and OHCs should be made in the present model[29].

Our model includes an assumption about the interaction between the OHC activities and the afferent radial fibers, caused by some electrical interferences from the OHC outputs on the radial fibers in the habenulla area beneath the IHCs, in order to explain the discrepancy between the major afferent innervation on IHCs and the major efferent innervation on OHCs. Recently, Brown and Nuttall (1984)[30] have proposed a model of such interaction between OHCs and IHCs based on the micromechanics of complex variations in the OHC stereocilia-tectorial membrane influencing the IHC stereocilia through the activities of actin, as

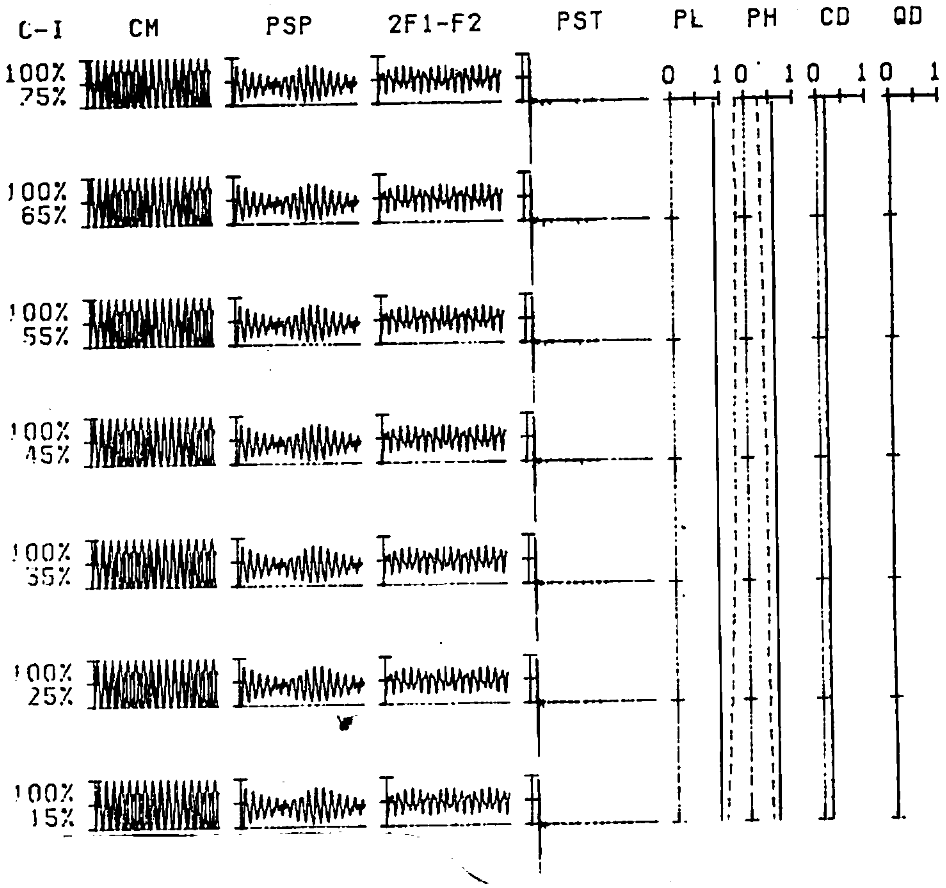

Fig.7(a). System responses for complex tone with cancellation tone with various intensities relative to lower primary tone (phase = o°) (f_1/f_2 = 1.1; f_2 level = 75%).

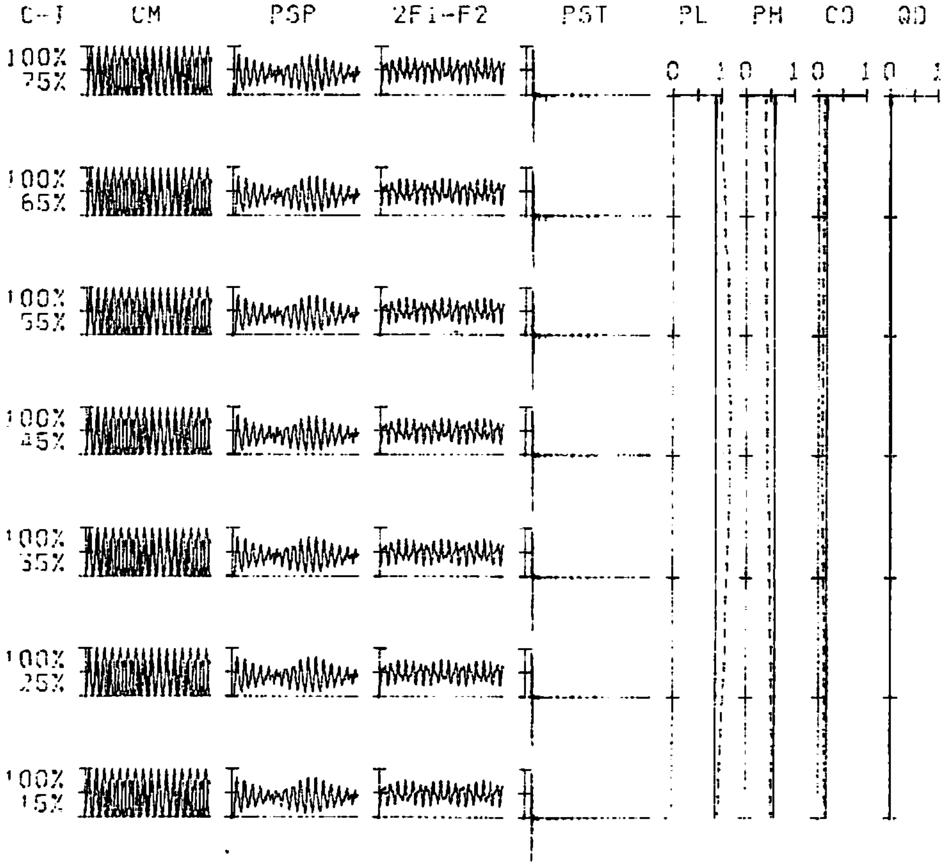

Fig.7(b). The same as in Fig.7(a) (phase = 60°).

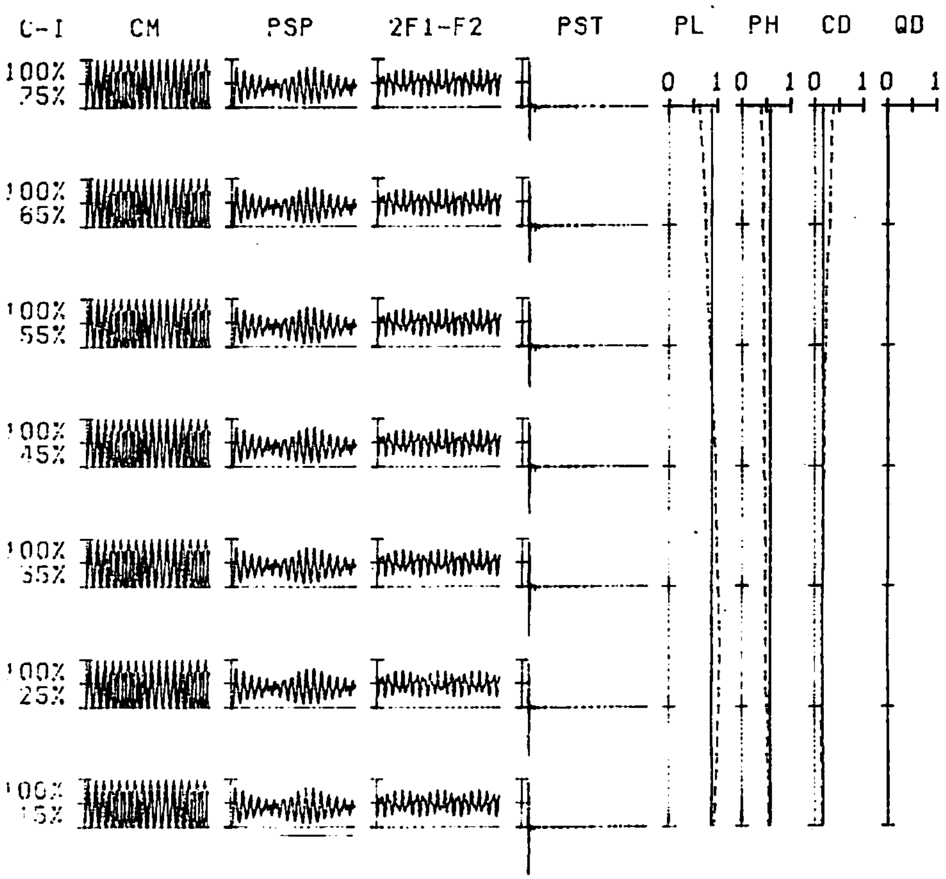

Fig.7(c). The same as in Fig.7(a) (phase = 120°).

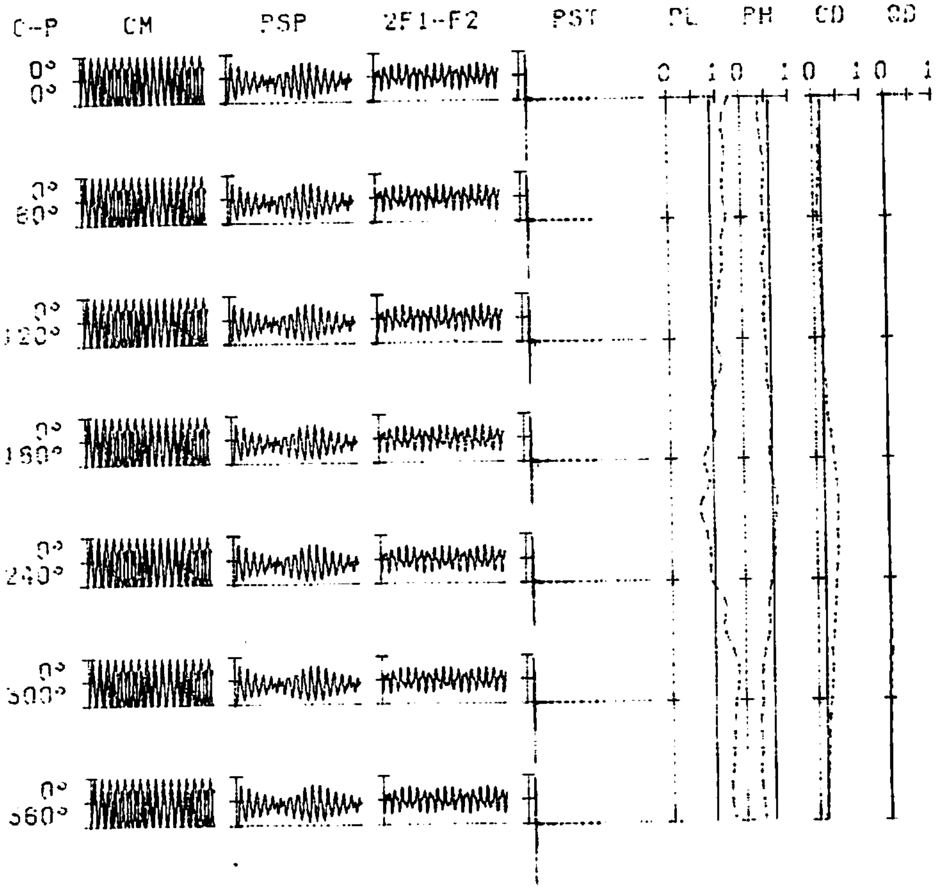

Fig.7(d). System responses for complex tone with cancellation tone with various phases relative to lower primary tone (level = 50%).

well as myocin.

The extraction of pitches from the temporal information in the neural activities [31] simulated in the model should be analyzed in relation to other phenomena in the perception of complex tones, e.g., hearing of AM sounds [32], pitch shifts caused by maskers [33], roughness and tonal dissonance [34], in order to apply this model to complex tones of more than two components.

Acknowledgements

The valuable suggestions of Dr K. Ohgushi and Prof H. Fujisaki are greatfully acknowledged. Supported in part by the Scientific Research Grant (No. 61870097).

References

1. Suga, N. and J. Horikawa (1986) Multiple time axes for representation of echo delays in the auditory cortex of the mustached bat. J. Neurophysiol. 55, 776-805.
2. Terhardt, E. (1974) Pitch, consonance, and harmony. J. Acoust. Soc. Am. 55, 1061-1069.
3. De Boer, E. (1976) On the "residue" and auditory pitch perception. In Kaidel, W.D. and W.D. Neff (eds) Handbook of Physiology, Vol.V/3. Auditory system - clinical and special topics, Springer: Berlin.
4. Goldstein, J.L. (1967) Auditory nonlinearity. J. Acoust. Soc. Am. 41, 676-689.
5. Zwicker, E. (1983) Level and phase of the $(2f_1-f_2)$-cancellation tone expressed in vector diagrams. J. Acoust. Soc. Am. 74, 63-66.
6. Plomp, R. (1976) Aspects of tone sensation - a psychophysical study. Academic Pr: London.
7. Greenwood, D.D. (1986) What is "synchrony suppression" ?. J. Acoust. Soc. Am. 79, 1857-1872.
8. Schroeder, M.R. (1969) Relation between critical bands in hearing and the phase characteristics of cubic tones. J. Acoust. Soc. Am. 46, 1488-1492.
9. Weber, E.G. (1966) Electrical potentials of the cochlea. Physiol. Rev. 46, 102-127.
10. Humes, L.E. (1985) An excitation pattern algorithm for the estimation of $(2f_1-f_2)$ and (f_2-f_1) cancellation level and phase. J. Acoust. Soc. Am. 78, 1252-1260.
11. Strube, H.W. (1986) The shape of the nonlinearity generating the combination tone $2f_1-f_2$. J. Acoust. Soc. Am. 79, 1511-1518.
12. Zwicker, E. (1981) Dependence of level and phase of the $(2f_1-f_2)$-cancellation tone on frequency range, frequency difference, level of primaries, and subject. J. Acoust. Soc. Am. 70, 1277-1288.
13. Itoh, K. (1984) A neuro-synaptic model of bilateral interaction in auditory nervous system. Trans. IECE E67, 12-18.
14. Holton, T. and Hudspeth AJ (1983) A micromechanical contribution to cochlear tuning and tonotopic organization. Sci-

ence 222, 508-510.

15. Cancelli, C., S. Dangelo, M. Masili and R. Harvono (1985) Experimental results in a physical model of the cochlea. J. Fluid Mech. 153, 361-388.
16. Spoendlin, H. (1978) The afferent innervation of the cochlea. In Naunton, R.F. and C. Fernandez (eds) Evoked electrical activity in the auditory nervous system. Academic Pr: New York.
17. Furukawa, T. and S. Matsuura (1978) Adaptive rundown of excitatory post-synaptic potentials at synapses between hair cells and eight nerve fibres in the goldfish. J. Physiol. 276, 193-209.
18. Art, J.J., R. Fettiplace and P.A. Fuchs (1984) Synaptic hyperpolarization and inhibition of turtle cochlear hair cells. J. Physiol. 356, 525-550.
19. Anderson, D.J. (1973) Quantitative model for the effects of stimulus frequency upon synchronization of auditory nerve discharges. J. Acoust. Soc. Am. 54, 361-364.
20. Langner, G. (1981) Neuronal mechanisms for pitch analysis in the time domain. Exp. Brain Res. 44, 450-454.
21. Ehret, G. and M.M. Merzenich (1985) Auditory midbrain responses parallel spectral integration phenomena. Science 227, 1245-1247.
22. Itoh, K. (1985) A neuro-synaptic model of the auditory masking and unmasking process. Biol. Cybernetics 52, 229-235.
23. Whifield, I.C. (1980) Auditory cortex and the pitch of complex tones. J. Acoust. Soc. Am. 67, 644-647.
24. Dallos, P. (1969) Combination tone $2f_l-f_h$ in microphonic potentials. J. Acoust. Soc. Am. 46, 1437-1444.
25. Wilson, J.P. and J.R. Johnstone (1973) Basilar membrane correlates of the combination tone $2f_1-f_2$. Nature 241, 206-207.
26. Goldstein, J.L. and N.Y.S. Kiang (1968) Neural correlates of the aural combination tone $2f_1-f_2$. Proc. IEEE 56, 981-992.
27. Rose, J.E., L.M. Kitzes, M.M. Gibson and J.E. Hind (1974) J. Neurophysiol. 37, 218-252.
28. Schmidt, R.A. and J.J. Zwislocki (1978) Low-frequency neural and cochlear-microphonic tuning curves in the gerbil. J. Acoust. Soc. Am. 64, 502-507.
29. Woolf, N.K. and A.F. Ryan (1985) Ventral cochlear nucleus neural discharge characteristics in the absence of outer hair cells. Brain Res. 342, 205-218.
30. Brown, M.C. and A.L. Nuttall (1984) Efferent control of cochlear inner hair cell responses in the guinea-pig. J. Physiol. 354, 625-646.
31. Ohgushi, K. (1978) On the role of spatial and temporal cues in the perception of the pitch of complex tones.
32. Rose, G.J. and R.R. Copranica (1985) Sensitivity to amplitude modulated sounds in the anuran auditory nervous system. J. Neurophysiol. 53, 446-465.
33. Stoll, G. (1983) Pitch shift of pure and complex tones induced by masking nise. J. Acoust. Soc. Am. 77, 188-192.
34. Clynes, M. (1982) Music, mind, and brain. Plenum: New York.