Biol. Cybern. 74, 359–366 (1996)

Biological Cybernetics

Neural networks simulating the frequency discrimination of hearing for non-stationary short tone stimuli

Andrej Kral[1], **Vladimir Majernik**[2]

[1] Institute of Pathological Physiology, School of Medicine, Sasinkova 4, SK-811 08 Bratislava, Slovak Republic
[2] Institute of Mathematics, Slovak Academy of Sciences, Stefanikova 49, SK-814 73 Bratislava, Slovak Republic

Received: 15 November 1994/Accepted in revised form: 7 November 1995

Abstract. This paper addresses the question of frequency discrimination of hearing for non-stationary (short) tone stimuli (duration $\leqslant 125$ ms). Shortening of the stimulus duration leads to widening of the frequency spectrum of the tone. It can be shown that for hearing no acoustical uncertainty relation holds and thus some nonlinear elements must be present in hearing physiology. We present neurophysiological and psychoacoustical findings supporting the hypothesis that frequency discrimination of non-stationary short tone stimuli is performed in neural networks of the auditory system. Neural network architectures that could process the temporal and place excitation patterns originating in the cochlea are suggested. We show how these networks (temporal coincidence network processing the temporal code and lateral inhibition network processing the place code) can be combined to show performance consistent with auditory physiology. They might explain the frequency discrimination of hearing for non-stationary short tone stimuli. We show the fitting of psychophysical relations based on these networks with the experimentally determined data.

1 Introduction

The amazing ability of the auditory system to discriminate frequencies of short acoustic stimuli has not been fully explained at the neurophysiological level. The acoustical uncertainty relation and the experimental results in psychoacoustics show that the auditory system is nonlinear with respect to frequency discrimination (see Appendix). In the last decade the new physiological findings in sensory acoustics, especially so-called active hearing (cochlear amplifier) with energy generation during the transduction process within the inner ear, has revolutionized hearing theory (e.g. Keidel 1992) and thrown new light mainly on the frequency discrimination ability of the ear for stationary tone stimuli, i.e. stimuli with a duration longer than 125 ms (Fack 1956). However, the high discriminatory ability of the ear for non-stationary short tone stimuli, i.e. stimuli with durations of less than 125 ms (measured by the difference limen for frequency), is unlikely to be explained in terms of the cochlear amplifier only. Therefore, the assumption seems to be justified that the fine frequency discrimination ability for non-stationary successive tone stimuli is probably performed in the auditory neural network (Majernik and Kaluzny 1979). We present a computer model simulating the neural network of the afferent auditory pathway which might explain this phenomenon.

Here the question is addressed as to how the input neural excitations are further processed in the nervous system so that the fine difference limina for frequency for non-stationary short tone stimuli known from psychoacoustic experiments (Oetinger 1959; Liang and Cistovic 1960; Cardozo 1962) are reached.

Our further considerations are based on the assumption that the afferent pathway consists of several neural networks running in parallel and serving different purposes. The parallelism in afferent auditory pathways is well known (e.g. Irvine 1992). As far as our goal is concerned we assume three parallel networks, which correspond to the classical place and temporal theory of hearing:

1. The network which sharpens (or preserves the sharpness of) the input excitation curve from hair cells (for details see Discussion). This networks preserves the tonotopic map and possibly increases its information content in the convergent afferent auditory pathway. Such a network might have lateral inhibitory couplings (Majernik and Kral 1993).
2. The network devoted to processing of the temporal code. Phase-locked spike sequences represent the input to this network. This network, as will be shown, might consist of temporal coincidence neurons. Its function is to transform the frequency of the phase-locked firings to a place code which could increase the discrimination ability due to interaction with the tonotopic map.

Correspondence to: A. Kral. e-mail: kral@fmed.uniba.sk

3. The network detecting a change in the frequency of the phase-locked spike sequences. This network should respond to tiny changes in frequency of non-stationary short tone stimuli that are not detected in the tonotopic map.

We attempt to simulate the fine discriminatory ability of the auditory system by means of the synergetic cooperation of these networks. Next we describe the architecture and function of these parallel networks.

2 Methods

In what follows we shall use computer simulations of neural networks to find out how the afferent auditory pathway processes the signal originating in the basilar membrane to achieve the frequency discrimination of the human ear.

2.1 Lateral inhibition neural network (LINN)

Lateral inhibition is a ubiquitous phenomenon found at numerous locations in the nervous system. It has been shown to be a mechanism enhancing the amount of information of the afferent signal (Rozsypal 1985; Majernik and Kral 1993). This network is also suitable for preserving the sharpness of tuning characteristic of cochlear nerve units in the higher levels of afferent processing (for details see Discussion). Such a mechanism will be applied to the place code of acoustic stimuli. There is also strong evidence of the usefulness of lateral inhibition from signal processing in hearing aids (cochlear implants: Ifukube and White 1987) and various other models. These neural and sensory facts represent a starting point for our further considerations.

2.1.1 The model. We consider neural network of L one-dimensional layers consisting of N neuron-like elements (processing elements, PEs). The PEs of the kth layer are connected only with the PEs of the $(k+1)$th layer (Fig. 1). Denoting the activity (pulsation) of the xth processing element in the kth layer as $\varphi[x]_k$ then the activity of the ith PE in the $(k+1)$th layer is given by the equation

$$\varphi[i]_{k+1} = f\left(\sum_{i-M}^{i+M} w[x]\,\varphi[x]_k - \Theta\right) \tag{1}$$

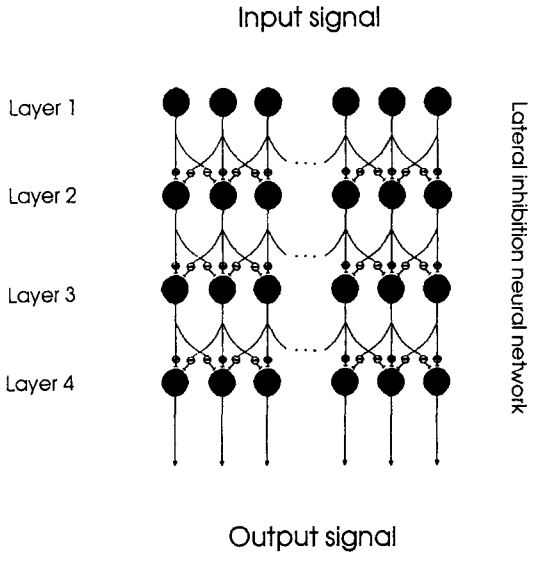

Fig. 1. The architecture of a lateral inhibition neural network (LINN)

where f is the linear threshold function shown in Fig. 2, $w[x]$ are the corresponding connection weights, Θ is the neuron threshold value and M the number of PEs to which it is directly connected. There are different types of one-dimensional network according to the weights chosen in (1). If $w[i]$ is always positive (i.e. the coupling between the ith PE in the kth layer and the ith PE in the $(k+1)$th layer is an excitatory one) and $w[x]$ $(x = i-M,\ i-M+1,\ i-M+2, \ldots, i-1, i+1, \ldots, i+M)$ (i.e. the connection between the xth PE in the kth layer and the ith PE in $(k+1)$th layer) is inhibitory, then we have a lateral inhibition network. Each one-dimensional lateral inhibition neural network (LINN) is given by the following parameters:

1. the excitatory weight $w[i]$
2. the inhibitory weights $w[x]$
3. the threshold value Θ
4. the connectivity, defined as $c = M/N$

According to the choice of these parameters, there is a variety of possible LINNs. In the computer model the real one-dimensional L-layered neural network is represented by a recursive network (i.e. the PEs' outputs are led back to the input of the network). Here the kth iteration in the computer model represents the state of the kth neuron layer of the real network.

The model implementation consists of 1000 PEs, with $c = 8.7\%$ $(M = 87)$, $w[i] = 1.4$, $w[x] = -0.01$, $\Theta = 0.2$. The neuron thus possesses one excitatory connection $(w[i] = 1.4)$ and $M-1$ inhibitory connections $(w[x] = -0.01)$ to other neurons. Such a network matches the proposed synaptic coupling architecture among cortical A cells and type II dorsal cochlear nucleus (DCN) units (Shamma and Symmes 1985). The physiological connectivity is difficult to assess. The upper threshold of the linear threshold function was set to 7.5 (we assume the maximal neural firing rate to be 750 Hz). This architecture shows behavior without 'wrong maxima' (a maximum on the output excitation curve which was not present in the input: Majernik and Kral 1993) or any other form of unwanted activity.

2.2 Temporal coincidence neural network (TCNN)

The second mechanism involved in frequency discrimination of tone signals is based on the analysis of the time patterns of neural discharges coming from the basilar membrane. This mechanism can be applied to tone signals with frequencies up to approximately 5 kHz. These frequencies seem to have greater importance for the sensation of acoustical stimuli than those above 5 kHz. The main information is carried by the former, while the latter probably only add timbre to the sound sensation.

There are numerous studies demonstrating the existence of two mechanisms such as these. Moore (1972), for example, showed that Zwicker's (1970) model leads to the auditory uncertainty relation (see Appendix)

$$\Delta f \cdot t_d \geqslant 0.24$$

where Δf represents the difference limen for frequency and t_d is the duration of acoustical stimuli. The temporal mechanism would not be subject to such a limitation. Siebert (1970) suspected that temporal cues would be much more efficient in pitch determination than place cues. Indeed, Moore (1972) showed that $\Delta f \cdot t_d$ drops below 0.24 at short stimulus durations (10–125 ms) for frequencies up to 5 kHz. This demonstrates that up to 5 kHz there is apparently more information gained

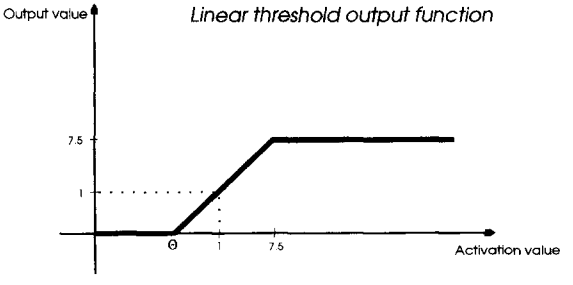

Fig. 2. The output function of the LINN processing element

by auditory frequency analysis than for the higher frequencies. Siebert's hypothesis suggests that the temporal mechanism could be the reason for this phenomenon. This is very plausible due to the fact that hair cells lose their ability to form an alternating current response at approximately the same frequency range, i.e. around 5 kHz (e.g. Cody and Russell 1987). The alternating current response is exploited in temporal coding by giving rise to phase-locked activity spike sequences in afferent neurons. Another finding supporting the participation of temporal coding in frequency discrimination comes from psychoacoustics. It has been shown that the frequency difference limina do not vary with frequency in the way predicted by place models. (For review see e.g. Moore 1993.)

Apparently the temporal information has to be exploited in the neural network of the auditory organ. We suggest a network capable of transforming this temporal code into a place code. This transformation is crucial to the possibility of further interaction between different neural networks.

2.2.1 The model. The simplest mechanism for such a transformation is a temporal coincidence neural network (TCNN). In this network a single PE, a 'neuron', can be tuned to detect a certain firing frequency on its input. A PE has a certain number of 'dendrites' the length of which are in a certain proportion. The PE threshold is higher than a single depolarization. For simplicity a PE with only two dendrites is considered (Fig. 3). If, for the lengths of the dendrites, the equation $l_2 = 2l_1$ holds, then the PE triggers a spike only if the depolarizations coming down the 'dendrites' reach the 'trigger zone' simultaneously (or within a certain small time delay Δt). This is the case when a given depolarization traveling down the longer 'dendrite' coincides with the succeeding one traveling down the shorter 'dendrite'. The PE is tuned to an input firing frequency for which we can write $\varphi = v/l_1$, where v is the constant spike propagation velocity. Such units are reminiscent of the so-called coincidence filter (Jeffress 1948) designed to explain sound localization ability.

In a coincidence network architecture we consider N temporal coincidence processing elements arranged with different l. Each PE receives excitation from 'its' distinct place on the basilar membrane. With this network we can transform discharge frequency (temporal code) into place code. Except for this, irregularities in the spike sequences (e.g., due to spontaneous activity of the hair cells) are eliminated in such a network. The ability to respond to certain frequencies is no longer limited by the biophysical properties of the traveling wave on the basilar membrane (as far as its shape – its envelope – is concerned).

2.3 Frequency change detection

The temporal principle is more suited to the detection of very minute frequency changes than changes in the peak position of the basilar membrane excursions (place principle). The phase locking could not only serve for detecting stimulus frequency, but could constitute input

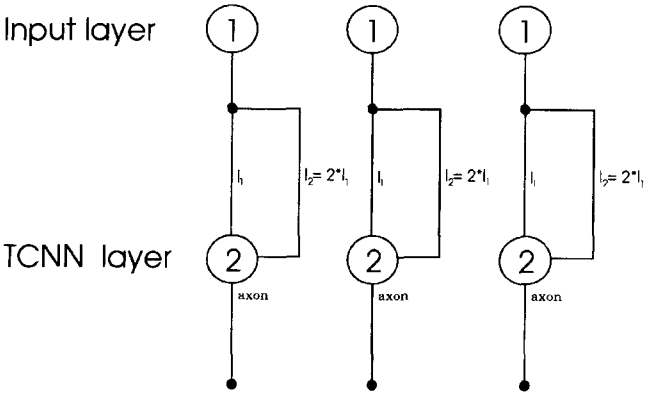

Fig. 3. The temporal coincidence neural network (TCNN). Note the different lengths of (dendrites) which constitute the frequency filter. Thresholds of the TCNN processing elements (PEs) are twice as high as those in the input layer. PEs are arranged according to their l_1 value

to a network designed to detect minute frequency change. The limitation for such recognition is the coding mechanism (we have to be aware of the fact that the relation between spike generation and basilar membrane excursion is a probabilistic one, i.e. the spike sequence generated is not perfectly regular, but occurs in one very specific excursion point on the basilar membrane), and the properties of neurons.

2.3.1 The model. Our hypothetical frequency change recognition network is based on the abovementioned temporal coincidence principle. We would like to have a processing element responding to a frequency change from f_1 to f_2 of two successive tone stimuli. Due to the change in stimulus frequency the interspike distances change as well. Taking a constant spike propagation velocity v, we can write interspike distances

$$d_1 = \frac{v}{f_1} \quad \text{and} \quad d_2 = \frac{v}{f_1 + \Delta f}$$

Our PE should become activated when the interspike distance changes from d_1 to d_2. It follows that $d_2 = d_1 + \Delta d$. Considering a processing element with three 'dendrites' of the abovementioned architecture and the aim of detecting such a change, the lengths of its dendrites should be in the proportion

$$l_1 : l_2 : l_3 = h : (h + d_1) : [h + d_1 + (d_1 + \Delta d)]$$

where l_1, l_2 and l_3 are the dendrite lengths and h is an arbitrary length constant. The network architecture resembles that of a TCNN. PEs are arranged in ascending order according to l_1 (Δl is a constant). One PE receives excitation from 'its' distinct point at the basilar membrane. It detects a frequency change which cannot be resolved at the basilar membrane for short tone stimuli. The network consists of such PEs arranged in ascending fashion according to d_1 (Δd is constant).

3 Results

The behavior of LINN with a broad input excitation curve is depicted in Fig. 4. The shallow flanks of the input excitation curve are suppressed, whereby amplifying the

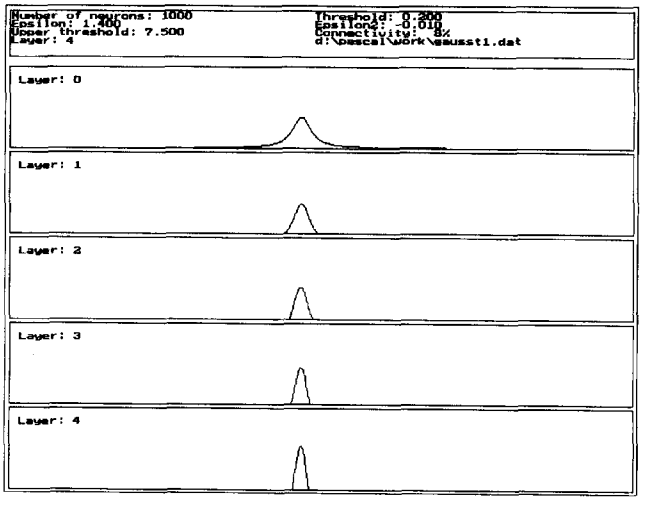

Fig. 4. The results of the LINN. The *top window* contains the network parameters. *Below* are the excitation curves from the corresponding layers of LINN (PEs arranged along the *abscissa*, output values of the PEs on the *ordinate*). The excitation curve is sharpened in the network. Activity spreads over 400 PEs in layer 0, whereas only 26 PEs are left active in layer 4. For details on excitation curve sharpening through an entropy-like quantity and the rate of convergence see Majernik and Kral (1993)

maximum. The resulting output layer activity is further restricted to the region where the maximum in the input excitation curve is located. Four layers roughly correspond to the number of nuclei in the lemniscal afferent pathway with substantial connectivity (cochlear nucleus, inferior colliculus, medial geniculate body and the cortex).

In Fig. 5 we see the results of the computer simulation of a temporal coincidence neural network. Each PE is depicted as a small circle. In the vertical direction (along the ordinate) is the activity of the given PE over time. In the horizontal direction the processing elements are arranged in descending order according to their 'tuned' frequency at a given time. We considered three dendrites for each PE with length proportions as described above. The input excitation was a uniform spike sequence for all units along the frequency axis (as can be expected from basilar membrane physiology). PEs active at the given time (corresponding to the position along the ordinate, i.e. the time axis) are marked in black.

The 9th, 18th, 27th and 36th PEs are activated by spike sequences with $d = 9$ bins (first stimulus). The 13th and 26th PEs are activated by the second simulated stimulus (causing input spikes with $d = 13$). Activity of the 18th and 27th PE is unwanted, because no stimuli with frequencies corresponding to $d = 18$ bins and $d = 27$ bins were actually applied. Such effects are not consistent with the experimental results in psychoacoustics. These responses have to be eliminated in the network. The same is true for the activation of the 26th PE. There are two possible solutions to this problem:

1. Inclusion of inhibitory couplings in the TCNN. The inhibition of PEs with a 'tuned period' that is a multiple of that of the given PE could prevent this activity. Nevertheless, the output of such a network would be biased in the case of two auditory stimuli of the corresponding frequencies. Then 'spurious negativity' of the higher tuned frequency PE would appear.
2. Interaction between the LINN and TCNN. The TCNN would receive excitation from active PEs in the LINN. If excitation from the LINN were necessary for suprathreshold activation of TCNN PEs (the thresholds have to be adjusted), the spuriously active PEs would no longer fire due to lack of excitation from corresponding portions of the LINN. The hypothetical result of such an arrangement is shown in Fig. 6.

The maximum firing frequency for a neuron is approximately 750 Hz (even up to 1000 Hz: e.g., Schmidt 1985). That is why the phase-locked activity cannot exceed this boundary. This is in perfect agreement with the neurophysiological observations, which show that the neurons do not fire in each stimulus cycle but rather fire synchronously with every multiple of it (Rose et al. 1967). This is why we suspect the basilar membrane to be functionally divided (in the region for frequencies up to 5 kHz) into segments corresponding to, for example, 750 Hz bands. Each of the segments considered serves as a bank of input activity generators to its temporal coincidence network (for each PE one generator at the corresponding place). The network processes the input in the

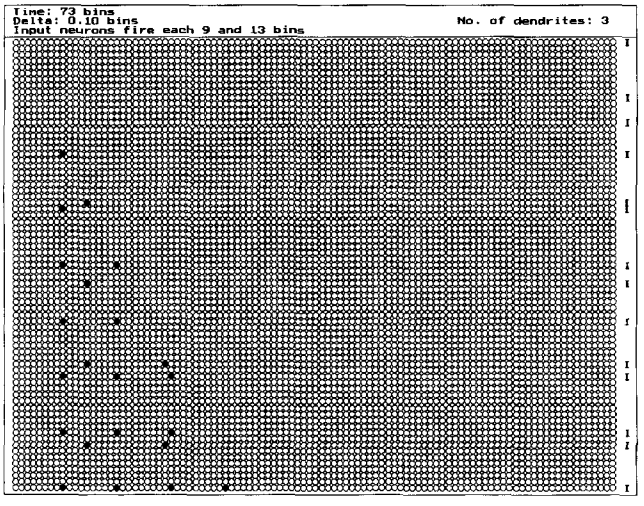

Fig. 5. The results of the TCNN simulation with input firings each 9 and 13 bins (i.e. two coded stimuli of different frequencies corresponding to the interspike intervals 9 and 13 bins). The network parameters are in the *top window*. The window below contains a TCNN spatio-temporal plane (PE arranged according to 'tuned' frequency on the *abscissa*, time on *ordinate*; the PE tuned to 1 bin interspike interval in time 0 at the *upper left corner*). '*I*' symbols in the *right-hand column* indicate appearance of discharges on input to the network at the given moment. Discharging PE are marked as *black circles*. Discharges are distributed in time among the 9th, 18th, 27th, 36th and 13th, 26th processing elements, thus indicating a transformation to place code. For non-stationary short tone stimuli the frequency resolution of this mechanism is supposed to exceed the range of place code resolution. The transformation of temporal code to place code is in agreement with physiological observations. It further enables an information interchange between the place code and the temporal code (originating in the cochlea) in the higher stages of auditory processing

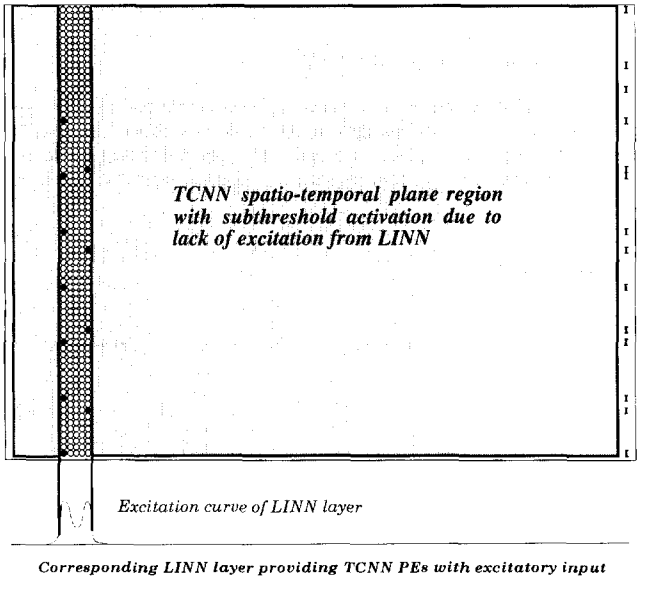

Fig. 6. The effect of excitation from the LINN on TCNN behavior after modification of thresholds in TCNN processing elements. The spuriously active PEs in Fig. 5 now receive only subthreshold activation due to lack of excitation from the corresponding portions of the LINN. Their activity in the TCNN can thus be effectively suppressed by such cooperation

abovementioned manner. To resolve the stimulus frequency from the activity of this network it is necessary to know the exact position of the active processing element as well as the segment to which the PE belongs (the place information). The segmentation architecture of the TCNN is depicted in Fig. 7.

The behavior of the frequency change detection network is shown in Fig. 8. After a change in stimulation frequency coded by interspike distance $d = 11$ bins to 12 bins, the 12th PE in the network discharges. The 'false' activation of a PE when l_1 is a multiple of that of the 'correctly' activated PE (delay of 24 bins, 24th PE) can be eliminated in a similar manner as described for the TCNN.

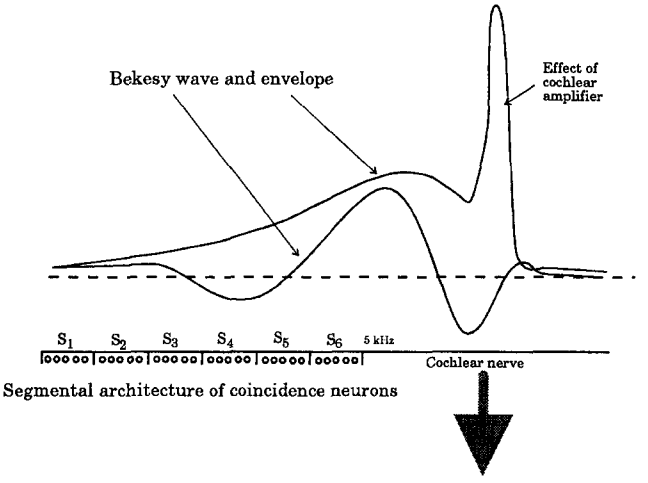

Fig. 7. Hypothetical segmental arrangement of auditory nerve neurons. The segments correspond to a given frequency band

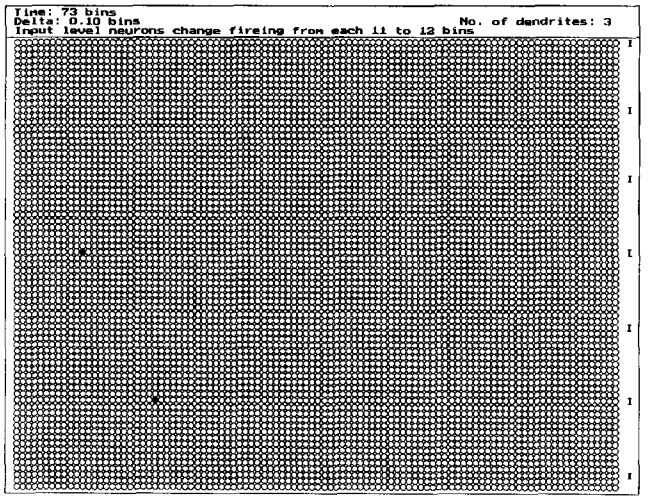

Fig. 8. Frequency change neural network spatio-temporal plane arranged in the same manner as in Fig. 5. Input activity represents a change in the spike-sequence interspike intervals from 11 to 12 bins (note *ticks* on the right of the figure). Some further frequency resolution refinement could be achieved by such a network

4 Discussion

Some neurophysiological justification of the neuroinformatical architecture of our model is to be found mainly in the results of single unit recordings.

The existence of a lateral inhibition network in the afferent pathway is still a subject of discussion. The theory of lateral inhibition in auditory research was introduced by Katsuki (1966). Today it is clear that the frequency tuning of the cochlear nerve neurons is essentially identical with the mechanical tuning of the basilar membrane (for review see Patuzzi and Robertson 1988). Nevertheless, there are numerous signs of lateral inhibitory influences in more central parts of the afferent auditory pathway. Some cochlear nucleus neurons seem to be tuned far sharper than auditory nerve units (e.g. Rhode and Greenberg 1992). DCN units of type II and III show inhibitory regions in the two-dimensional frequency-intensity plane (inhibitory sidebands), are sharply tuned, are weakly responsive to wideband noise and are nonmonotonic in their response. Such units are not present in the cochlear nerve, so they have to be the result of neural inhibition. A possible function of inhibitory sidebands is to preserve spectral selectivity when there is convergence of afferent fibers (Rhode and Greenberg 1992). Phillips (1993) uses a similar argument to explain the nonmonotonic frequency tuning curve (FTC) of neurons in more central parts of the afferent pathway. He points out that this nonmonotony could be the result of subtraction of one FTC by the FTC of other units (most probably the neighboring ones in the tonotopic map). This is in accord with the architecture of a LINN. Evans and Zhao (1993) claim to have found evidence of lateral inhibition in DCN cells III and IV in microiontophoretic studies. Ehret and Merzenich (1988) conclude that lateral inhibition at the level of the inferior colliculus plays a role in frequency resolution. Blackburn and Sachs (1990) suggest that the rate-place representation observed in the response of anteroventral cochlear nucleus (AVCN) choppers at high sound pressure levels may be the result of some further sharpening of the spectral representation derived from lateral inhibitory interactions. We can conclude that nonmonotonic units that demonstrate sharp frequency tuning, inhibitory sidebands and a weak responsiveness to wideband noise like DCN units of type II and cortical A cells (Shamma and Symmes 1985) resemble in behavior a PE of a lateral inhibition neural network. Other evidence comes from psychoacoustic observations. Psychophysical tuning curves determined in forward masking or using the pulsation threshold method are typically sharper than those obtained in simultaneous masking. According to Houtgast (1974) this difference arises because of lateral inhibition (suppression). In simultaneous masking its effect is not seen since any reduction of the masker activity in the frequency region of the signal is accompanied by a similar reduction in signal evoked activity (Moore 1993). Lateral inhibition is, we believe, extremely important if not for sharpening then at least for preservation of frequency selectivity in the convergent afferent auditory pathway. It represents a mechanism of functional separation of

signals in the highly interconnected nervous system (Majernik and Kral 1993). Many models of hearing have exploited lateral inhibition (e.g. Kurogi 1991), showing its usefulness in acoustic signal processing.

On the question of the temporal coincidence, it needs to be emphasized that cortical auditory neurons have very poor steady-state temporal responses to simple tones (Steinschneider et al. 1980). Even more, there seems to be a general trend for a decrease in the temporal resolution capabilities of neurons (e.g. to AM stimuli) located at successively higher levels of the auditory pathway (Schreiner and Langner 1988). This seems to indicate that the temporal code is transformed in the centripetal pathway. One possible transformation is the transformation to a place code, so that integration of these two forms of information can take place.

Keidel's group has found neurons in the inferior colliculus which show multiple peak responses in the peristimulus histograms (discharge rate versus stimulus frequency). The periods of the peaks correspond to the stimulus frequency associated with the first peak (Keidel 1992; Erulkar 1975, p. 190). Similar behavior is to be expected from PEs in the TCNN. David et al. (1969) explained these results by a different neural architecture. They claim to have discovered the so-called clock cells in the inferior colliculus and medial geniculate body and speculated about a decoding mechanism suitable for processing temporal code. They suggested the spike sequences from cochlea could be compared with the activity of clock cells, each of them spontaneously discharging at different frequencies. Although they did not suggest the exact architecture of the neural network, one can assume this might be another possible mechanism. Nevertheless, such a network would have to be much more complex than our temporal coincidence network.

The second line of evidence comes from psychoacoustic experiments. Stevens introduced the 'mel scale' into psychoacoustics as a measure of subjective tone pitch (Zwicker 1975). In what follows we will explain the relation between mel and Hz based on these networks. [The classical explanation of subjective pitch relies on the correspondence between the course of the mel-Hz relation and bark-Hz relation (Zwicker 1975). This explanation seems to be incomplete, leaving no room for temporal mechanisms participating in subjective pitch.]

Pitch as a sensory quantity is related to the neural mechanisms by which it is generated through certain processing of the input signal. To find this mechanism we have first to look at what kind of activity this stimulus evokes in the afferent auditory pathway and then to investigate the relation between the number of periods of tone pulses (stimulus) and the number of neural discharges in our neuron model. Let us denote the upper threshold for discharge frequency of neurons as f_G. The number of neuron discharges at higher stimulus frequencies is thus 2, 3 or m times smaller than the number of the periods of tone stimulus. In an idealized case of neuron respondence the relation between the stimulus frequency f and the number of neural discharges Φ as a function of

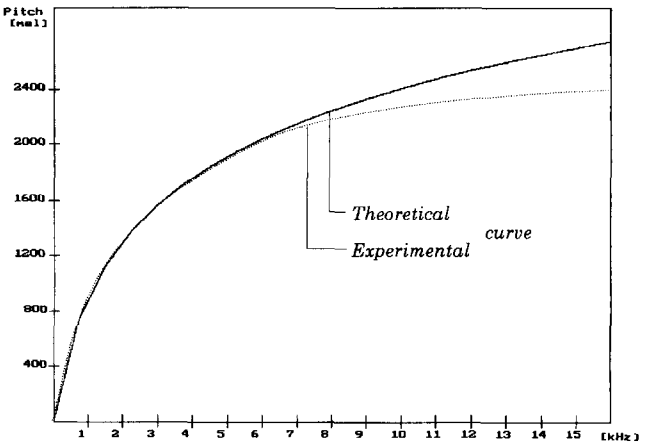

Fig. 9. Theoretical subjective pitch curve according to the behavior of the TCNN. The correspondence with the experimental subjective pitch curve is very good up to 5–6 kHz. The separation in the higher frequency region is irrelevant due to disappearance of an alternating current response in hair cells from 5 kHz above. This is taken as further support of the proposed neural network architecture

frequency can be expressed in the following way:

$$
\begin{aligned}
&M_1 = f && \Phi_1 = f && \text{for } 0 \leqslant f < f_G \\
&M_2 = f_G + f/2 && \Phi_2 = \tfrac{1}{2} f && \text{for } f_G \leqslant f < 2f_G \\
&\vdots \\
&M_m = f_G + \sum_{i=2}^{m-1} \frac{1}{i} f_G + \frac{1}{m} f && \Phi_m = \frac{1}{m} f && \text{for } (m-1)f_G \leqslant f < m f_G
\end{aligned}
\tag{2}
$$

If we assume that the pitch sensation is given by the number of neural discharges, then the subjective pitch scale is an integration over the neural discharges, and so the pitch-frequency function is given by the values of the first column in (2). Figure 9 shows the experimental pitch-frequency function together with the idealized theoretical function (2). We see that the agreement is good up to a frequency of approximately 5 kHz if we take $f_G = 750$ Hz. The overestimation above 5 kHz is due to the fact that above this frequency only the place coding can be exploited in reality.

We have presented two types of artificial neural network that may, by their mutual cooperation, explain the mechanisms for processing the place and temporal codes aimed at frequency discrimination. We have presented neurophysiological and psychophysical findings corresponding to the behavior of such networks. On the basis of these correlations we may state that the model presented might represent one possible explanation of the neuroinformatical structure of the afferent auditory pathway responsible for auditory frequency discrimination of nonstationary tone stimuli.

Appendix

The acoustical uncertainty relation (AUR) states that the product of bandwidth and duration of an acoustical signal is a constant that depends only upon the signal

waveform. It can be expressed mathematically in the following way: Consider a signal $f(t)$ and its Fourier transform $G(\omega)$. The effective signal duration Δt and the effective signal bandwidth $\Delta\omega$ are, according to Gabor (1946), defined by relations

$$\Delta t = 2\sqrt{[\pi\overline{(t-\bar{t})^2}]}$$

$$\Delta\omega = 2\sqrt{[\pi\overline{(\omega-\bar{\omega})^2}]}$$

where

$$\bar{t} = \frac{\int_{-\infty}^{\infty} t|f(t)|^2\,dt}{\int_{-\infty}^{\infty}|f(t)|^2\,dt} \quad \text{and} \quad \bar{\omega} = \frac{\int_{-\infty}^{\infty}\omega|G(\omega)|^2\,d\omega}{\int_{-\infty}^{\infty}|G(\omega)|^2\,d\omega}$$

respectively. Here, the raised sign denotes the mean value of the corresponding quantity. The product of Δt and $\Delta\omega$ satisfies the inequality

$$\Delta t\cdot\Delta\omega \geqslant 1 \qquad \text{(A1)}$$

which expresses AUR. For the Gaussian function

$$f(t) = A\cdot e^{-a(t-t_0)^2}\cdot e^{i(\omega t+\varphi)}$$

where A, a, t_0 and φ are certain constants, the inequality (A1) becomes an equation with the minimal value of the right-hand side just equal to 1.

Several authors have dealt with the determination of AUR either theoretically (Gabor 1946; Wunsch 1962) or experimentally (Oetinger 1959; Liang and Cistovic 1960; Cardozo 1962; Ronken 1970; Majernik and Kaluzny 1979). If one puts the difference limen for frequency (DLf) and the signal duration t_d proportional to the effective bandwidth and effective signal duration, i.e. $\mathrm{DLf} = k_1\cdot\Delta\omega$ and $t_d = k_2\cdot\Delta t$, where k_1 and k_2 are constants, then the *auditory* uncertainty relation can be written in the form (Majernik 1967)

$$(\mathrm{DLf})\cdot t_d = F(f, t_d)$$

When the frequency analysis is performed by a linear analyzer, then for a given signal envelope the function $F(f, t_d)$ should be constant. However, the function $F(f, t_d)$ obtained by fitting the experimental data of DLf as a function of signal duration, given in Majernik and Kaluzny (1979), has the form

$$F(f, t_d) = t_d\left[K_1 + \frac{K_2}{(K_3 + t_d)^2}\right]e^{bf}$$

where f is in Hz, t_d in ms, and $K_1 = 1.85$ Hz, $K_2 = 625.5$ ms, $K_3 = 1.42$ ms and $b = 4\times10^{-4}\,\mathrm{Hz}^{-1}$ (Kaluzny et al. 1985), which by no means a constant but depends strongly on the signal frequency and duration. This points out that the auditory frequency analysis cannot be performed solely by a kind of linear analyzer which is bound to the AUR, but must be improved by means of some nonlinear processes that are probably also in the auditory neural network.

Acknowledgements. The authors would like to thank Professor Ivan Hulin for his support, and Dr. L. Krajcovicova and the anonymous referees for comments which helped to improve the quality of the paper. The work was supported in part by grant No. 1/1169/95 of the Ministry of Education and Science of Slovak Republic.

References

Blackburn CC, Sachs MB (1990) The representation of the steady-state vowel sound (ε) in the discharge patterns of cat anteroventral cochlear nucleus neurons. J Neurophysiol 63:1191–1212

Cardozo BL (1962) Frequency discrimination of human ear. In: Proceedings of the 4th International Congress on Acoustics, Copenhagen

Cody AR, Russell IJ (1987) The responses of hair cells in the basal turn of the guinea-pig cochlea to tones. J Physiol (Lond) 383:551–569

David E, Finkenzeller P, Kallert S, Keidel WD (1969) Reizfrequenzkorrelierte 'untersetze' neuronale Entladungsperiodizität im colliculus inferior und im corpus geniculatum mediale. Pflugers Arch 309:11–20

Ehret G, Merzenich MM (1988) Complex sound analysis (frequency resolution, filtering and spectral integration) by single units of the inferior colliculus of the cat. Brain Res 472:139–163

Erulkar SD (1975) Physiological studies of the inferior colliculus and medial geniculate body. In: Keidel WD, Neff WD (eds) Handbook of sensory physiology. auditory system, vol. V/2. Springer, Berlin Heidelberg New York, pp 147–198

Evans EF, Zhao W (1993) Varieties of inhibition in the processing and control of processing in the mammalian cochlear nucleus. Prog Brain Res 97:117–126

Fack H (1956) Informationstheoretische Behandlung des Gehoers. In: Wincker F (ed) Impulstechnik. Springer, Berlin Heidelberg New York, pp 1–128

Gabor D (1946) Theory of communication. J Int Eng 93:429–457

Houtgast T (1974) Lateral suppression in hearing. PhD Thesis, Free University of Amsterdam. Academische Pers. BV, Amsterdam

Ifukube T, White RL (1987) A speech processor with lateral inhibition for an eight channel cochlear implant and its evaluation. IEEE Trans Biomed Eng 34:876–882

Irvine DRF (1992) Physiology of the auditory brainstem. In: Popper AN, Fay RR (eds) The mammalian auditory pathway: neurophysiology. Springer, Berlin Heidelberg New York, pp 153–231

Jeffress LA (1948) A place theory of sound localisation. J Comp Physiol Psychol 41:35–39

Kaluzny J, Majernik V, Kaluzna M (1985) A contribution to the determination of the frequency discrimination ability. Arch Acoust 10:253–258

Katsuki Y (1966) Neural mechanisms of hearing in cats and monkeys. Prog Brain Res 21A:71–97

Keidel WD (1992) Das Phaenomen des Hoerens: ein interdisziplinaerer Diskurs II. Naturwissenschaften 79:347–357

Keidel WD, Neff WD (eds) (1975) Handbook of sensory physiology. Auditory system, vol. V/2. Springer, Berlin Heidelberg New York

Kurogi S (1991) Speech recognition by an artificial neural network using findings on the afferent auditory system. Biol Cybern 64: 243–249

Liang C, Cistovic L (1960) Dependence of the frequency differential thresholds on the duration of the tone signals. Akust Zhur 6:81–85

Majernik V (1967) Die Unschaerferelation fuer das Gehoer. Fyz cas SAV 2:65–73

Majernik V, Kaluzny J (1979) On the auditory uncertainty relations. Acustica 43:132–146

Majernik V, Kral A (1993) Sharpening of input excitation curves in lateral inhibition. Int J Neural Syst 4:65–80

Moore BCJ (1972) Some experiments relating to the perception of pure tones: possible clinical applications. Sound 6:73–79

Moore BCJ (1993) Frequency analysis and pitch perception. In: Yost WA, Popper AN, Fay RR (eds) Human psychophysics. Springer, Berlin Heidelberg New York

Oetinger R (1959) Die Grenze der Hoerbarkeit von Frequenz und Tonzahlaenderungen bei Tonimpulsen. Acustica 9:430–445

Patuzzi R, Robertson D (1988) Tuning in mammalian cochlea. Physiol Rev 68:1009–1082

Phillips DP (1993) Representation of acoustic events in the primary auditory cortex. J Exp Psychol Hum Percept Perform 19:203–216

Rhode WS, Greenberg S (1992) Physiology of the cochlear nuclei. In: Popper AN, Fay RR (eds) The mammalian auditory pathway: neurophysiology. Springer, Berlin Heidelberg New York, pp 94–152

Ronken DA (1970) Some effects of bandwidth-duration constraints on frequency discrimination. J Acoust Soc Am 49:1232–1242

Rose JE, Brugge JF, Anderson DJ, Hind JE (1967) Phase-locked response to low frequency tones in single auditory nerve fibres of the squirrel monkey. J Neurophysiol 30:769–793

Rozsypal AJ (1985) Computer simulation of an ideal lateral inhibition function. Biol Cybern 52:15–22

Schmidt RF (1985) Fundamentals of neurophysiology. Springer, Berlin Heidelberg New York

Schreiner CE, Langner G (1988) Coding of temporal patterns in the central auditory system. In: Edelman GM, Gall WE, Cowan WM (eds) Auditory function: neurobiological bases of hearing. Wiley, New York, pp 337–361

Shamma SA, Symmes D (1985) Patterns of inhibition in auditory cortical cells in awake squirrel monkeys. Hear Res 19:1–13

Siebert WM (1970) Frequency discrimination in the auditory system: place or periodicity mechanisms. Proc IEEE 58:723–730

Steinschneider M, Arezzo J, Vaughan HG Jr (1980) Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey. Brain Res 198:75–84

Wunsch G (1962) Moderne Systemtheorie. Geest und Portig, Leipzig

Zwicker E (1970) Masking and psychological excitation as consequences of the ear's frequency analysis. In: Plomp R, Smoorenburg GF (eds) Frequency analysis and periodicity detection in hearing. AW Sijthoff, Leiden

Zwicker E (1975) Scaling. In: Keidel WD, Neff WD (eds) Handbook of sensory physiology. Auditory System, vol. V/2. Springer, Berlin Heidelberg New York, pp 401–448