

# Modelling the effect of the missing fundamental with an attractor neural network\*

Lubica Beňušková†

*Institute for Developmental Neuroscience, Vanderbilt University, Nashville, TN 37203*

**Abstract.** In order to investigate the use of attractor neural networks (ANNs) in modelling real phenomena of human perception, we developed and tested the ANN model of spectrum-invariant pitch recognition that was aimed at cortical mechanisms of pitch perception. We tested the ANN model in computer experiments which were arranged in a way compatible with the three pitch identification psychophysical experiments with human subjects. The results of computer and actual experiments are qualitatively the same, and quantitatively they depart at most by 10% which can be considered to be in good agreement. It is shown that the present ANN model can provide an explanation for the different efficacies of lower and upper harmonics in conveying the missing fundamental pitch percept, by means of one universal cortical mechanism. In my model, a complex tone evokes the pattern of activity that consists of a particular combination of 'isofrequency' stripes in the ANN that represents the relevant part of the primary auditory cortex (AI). Recognition of the pitch is interpreted as an auto-associative retrieval of the corresponding stripe-like template. The numbers of active neurons in the stripes that represent lower frequencies are greater than the numbers of active neurons in the stripes for the higher frequencies. Thus, when the network is presented with the frequency complex containing either lower or higher harmonics, different efficacies of lower and upper harmonics in conveying the missing fundamental pitch percept are manifested. The proposed ANN model can be easily generalized for the spectrum-invariant recognition of unharmonic sounds, for instance speech.

## 1 Introduction

The study of stochastic nonlinear co-operative systems, for instance the Kirkpatrick and Sherrington models of spin-glasses [1], has resulted in the emergence of a new paradigm in modelling the brain functions, the theory of attractor (Hopfield) neural networks (ANNs) [2, 3]. During the past decade, theoretical analysis of ANNs has made substantial progress [3]. Compared to the large volume of literature on theoretical investigations of ANNs there is relatively little emphasis on studies of their biological or psychological plausibility. Several suggestive analogies between ANN and biological computational processes have been made, for example, for olfaction [4, 5], visual processing [6], and hippocampus [7]. Since these models remain untested at the neurophysiological level, the relevance of ANN to neurobiology is still unclear. By contrast, an exemption is the ANN model of Kleinfeld and Sompolinsky [8] that is capable of generating cyclic sequences of states. When applied to a small neural circuit of the mollusc *Tritonia* that controls the swim rhythm, the model accounted for the experimentally observed output and for the mean operating characteristics of the individual neurons.

We have chosen the phenomena of musical pitch perception as one of the many possible instances where the ANN model can be confronted with the results of actual experiments. In the real world, different frequencies from the same source appear in different neural channels, divided in space and in time, and in addition similar frequencies from different sources appear in the same neural channels, related in time and in space. From this pattern, the auditory system must reconstruct the percepts of the individual sources. What are the mechanisms which the auditory system use to accomplish this task? One of the 'tools' which is used by auditory system for this task is the mechanism of musical pitch perception. Pitch perception can be considered to be a labeling process by which the auditory system collects certain frequencies and tags them with a pitch identifier, and simultaneously collects another set of frequencies and tags them with another identifier, and so on [9]. The

---

\* *Network: Comput. Neural. Syst.* **5**, 333–349, 1994. † On leave of absence from Neurophysics Group, Institute of Animal Biochemistry and Genetics, Slovak Academy of Sciences, 90028 Ivanka pri Dunaji, Slovakia.

lowest frequency in the complex tone is called the *fundamental frequency*,  $f_0$  [10]. The higher frequencies, that are all integral multiples of  $f_0$ , are called the *harmonics* or overtones of the fundamental frequency. The second harmonic has a frequency  $f_0 \times 2$ , the third harmonic has a frequency  $f_0 \times 3$ , and so on. A tone's height, *pitch*, is determined by the tone's fundamental frequency, and expressed as a note in musical notation. Sounds like those produced by speech, slamming a door, or crumpling paper, are musically unharmonic, i.e. the higher frequencies are not integral multiples of the sound lowest frequency. Although there is a dominant tendency to fuse together harmonically related components into a single entity, for any complex sound, no matter how unharmonic, there is some time duration short enough that the components will be integrated into a single entity too. There are many puzzling phenomena in pitch perception but we will deal only with one of them; that is *the effect of the missing fundamental*, known also under other names – virtual pitch, periodicity pitch, low pitch, musical pitch, or salient pitch [9]. The curious effect of the missing fundamental concerns the ability of our auditory system to perceive pitches that correspond to the fundamental frequencies of complex tones while those fundamentals are physically totally absent.

One of the most influential theories of pitch perception was the frequency detection theory of Helmholtz [11]. According to him, the cochlear spectrum analyzer maps the energy of the fundamental frequency tonotopically, and pitch is associated with the spatial position of the fundamental frequency. If there is no fundamental frequency present, the nonlinearities in the periphery will introduce it. It may seem that the peripheral mechanisms are indeed responsible for the effect of the missing fundamental since there is a recent physiological evidence that when the complex stimulus contains only upper harmonics, there are clear correlates of the missing fundamental in the firing pattern of auditory nerve fibers [12] and in the neurons of the cochlear nucleus [13].

However, there are several experimental findings that strongly undermine the idea about the peripheral mechanisms being *always* responsible for the effect of the missing fundamental. First, the perception of virtual pitch is here even when the intensities of harmonics are so low that they cannot convey difference tone distortions [14], or second, when there is an external masking by other frequencies [15]. Third, very strong argument against all theories which explain the perception of pitch of the missing fundamental as a peripheral phenomenon comes from the work of Houtsma and Goldstein [16]. They showed that the corresponding musical pitch can be generated only by two upper harmonics which are presented dichotically, i.e. each one of them is delivered through a separate ear. However, each of them presented monotonically leads to the corresponding pure tone percept. It is clear that with the dichotic stimulus there is no possible way for the two components to fall into the same peripheral (aural) channel. Therefore some recent theories of pitch perception [17–19] propose central mechanisms which perform a template-matching process underlying the perception of the pitch of a complex tone.

To this class of models belongs the hypothesis introduced and tested in my study; that the pitch recognition is essentially an auto-associative recall from memory, in the sense of a cooperative re-creation of the network activity configuration which represents the stored memory template for the given tone. Thus, memories are meant to be perceptual templates for the complex tones that are stored in an ANN via the synaptic weight matrix. The prescription for storage of patterns in ANN is a Hebb-like rule [20] that is the distribution of efficacies of synaptic weights is a result of 'experience' with particular set of patterns. Different sets of patterns yield different distributions of efficacies. The ANN stores templates that can be retrieved when noisy or incomplete version of original patterns are presented at the input. It is not unreasonable to assume that sensory experience with various sounds results in the formation of templates for individual sounds on which our perceptual recognition is based, and we will show that this process can be modelled with ANN.

For this purpose, we will design a network representation of musical tones that is in agreement with the current neurobiological data as well as with the psychophysical experiments with human subjects. The model will be tested by means of computer simulations and its performance will be quantitatively compared with the results from actual psychophysical experiments evaluating the effect of the missing fundamental. It will be argued that the present model can provide an explanation for the different efficacies of lower and upper harmonics in conveying the missing fundamental pitch percept, by means of one universal mechanism that emerges as a result of realistic tone representations in the auditory cortex. we will also discuss the neurophysiological experiments that can directly test the proposed model.

## 2 The model network

The neural network model which is used in computer simulations of the effect of the missing fundamental is the stochastic ANN [3] modified for storage and retrieval of patterns with a low level of overall network activity according to Amit *et al.* [21]. The basic description of the model ANN is as follows:

(1) In the network, there are  $N$  neurons and to each neuron a binary variable  $S_i = \pm 1$ ,  $i = 1 \dots N$ , is assigned.  $S_i$  represents the state (output) of the  $i^{th}$  neuron.  $S_i = 1$  whenever a neuron is active at a given time instant, and  $S_i = -1$  whenever it is inactive. The current state of the network is represented by  $N$ -dimensional binary vector  $\{S_i\} = (S_1, S_2, \dots, S_N)$ .

(2) Each neuron is connected to every other.  $J_{ij}$  denotes the efficacy of the synaptic input from the neuron  $j$  on the neuron  $i$ .  $J_{ij} > 0$  for an excitatory input, and  $J_{ij} < 0$  for an inhibitory input. A neuron does not have a direct feedback, e.g.  $J_{ii} = 0$ .

(3) We will deal with non-orthogonal memory patterns in which the ratio of active vs. non-active neurons is significantly lower than 50%. For such highly correlated or biased patterns we will use the modification proposed by Amit *et al.* [21]. Here, the prescription for synaptic efficacies  $J_{ij}$  is the modified Hebb-like rule

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p (\xi_i^\mu - b)(\xi_j^\mu - b) , \quad (1)$$

where  $N$ -dimensional binary vectors  $\{\xi_i^\mu\} = (\xi_1^\mu, \xi_2^\mu, \dots, \xi_N^\mu)$ ,  $\mu = 1, \dots, p$ , represent chosen memory patterns, and  $-1 < b < 1$  is the bias parameter for which it holds

$$b = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu , \quad \mu = 1, \dots, p . \quad (2)$$

(4) A global control on the dynamics of the network, which prevents too high or too low activity, is achieved by imposing a finite energy cost on fluctuations away from the optimal mean level of activity [21]. Now an energy function has a form

$$E = -\frac{1}{2} \sum_{i,j,j \neq i}^N J_{ij} S_i S_j + \frac{g}{2N} \left( \sum_{i=1}^N S_i - Nb \right)^2 . \quad (3)$$

Here the coefficient  $g > 0$  measures the strength with which this *soft* constraint is imposed. The overall postsynaptic potential of the  $i^{th}$  neuron is

$$h_i = \sum_{j=1}^N J_{ij} S_j - g \left( \frac{1}{N} \sum_{j=1}^N S_j - b \right) . \quad (4)$$

If the constraint parameter  $g$  is large enough, the results become essentially independent of  $g$ . One can find [21] that by the time  $g = 10$ , the asymptotic behavior is reached for all values of the bias  $b$ . In all my simulation, we used the value of  $g = 12$ . The theoretical storage capacity  $\alpha$  of network storing low activity patterns with softly constrained dynamics is  $0.12 \leq \alpha \leq 0.18$ . The exact value depends on the value of  $b$ . Thus, the maximal number of memory patterns which can be retrieved in the network with  $N$  neurons is  $p = \alpha N$ .

(5) The network dynamics is asynchronous, i.e. at every time instant only one randomly chosen neuron updates its state according to its current postsynaptic potential (4). A probabilistic rule for the  $i^{th}$  neuron state transition is expressed by the relation [3, 22]

$$Pr[S_i(t)] = \frac{1}{1 + \exp[-2\beta h_i(t-1)S_i(t)]} ; \quad \beta = T^{-1} . \quad (5)$$

$0 < T < 1$  is the analog of temperature and represents the level of noise during neuronal relaxation. An interpretation of stochastic asynchronous dynamics is as follows. Between generations of the two consecutive action potentials at least an absolute refractory period, which is about 1–2 ms [23], has to pass. In the model, the time interval  $\Delta t = 1$ –2 ms is divided into  $N$  subintervals in such a way that:  $\Delta t = N \delta t$ . At every time instant  $\delta t$  randomly chosen (with equal probability  $\frac{1}{N}$ ) neuron updates its state according to the probabilistic rule (5). Therefore there is a guarantee that every neuron pauses at least 1–2 ms (absolute refractory period)

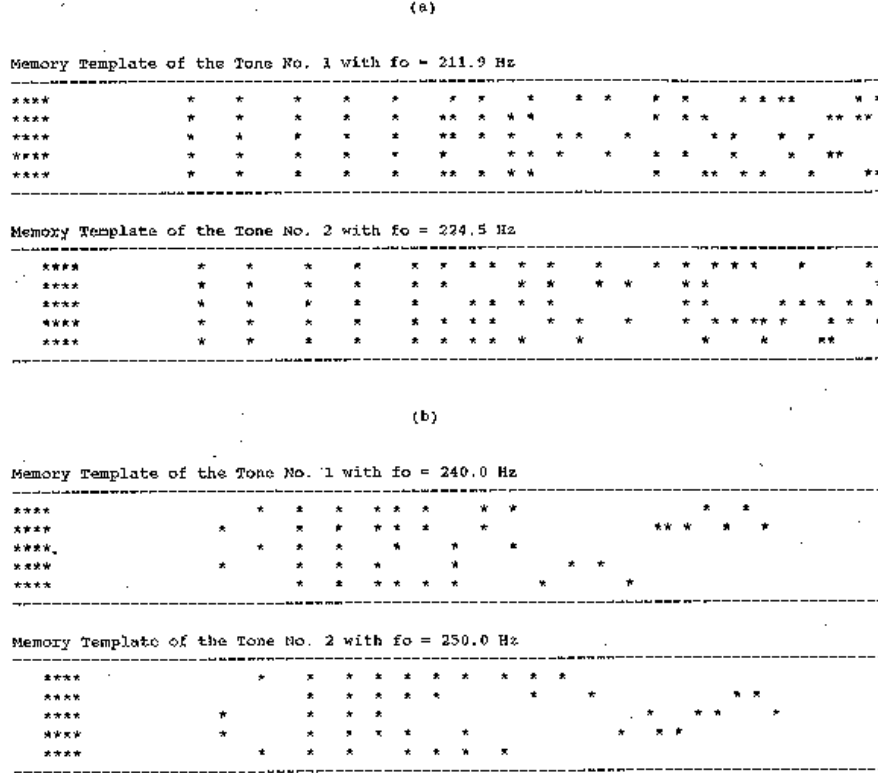


Figure 1: Examples of the complex tone representation that is employed in computer simulations of the effect of the missing fundamental. The tone representation is a stripe-like pattern of activity with more neurons representing low frequencies than representing high frequencies. 'Isofrequency' stripes are orthogonal to the low-to-high frequency gradient. An asterisk means that the neuron in the 'isofrequency' stripe is active while blank means that the neuron is not active. The first four lines of asterisks on the left represent the fundamental frequency 4-stripe. Each next tone from the scale is shifted to the right since it covers higher frequencies. Here only part of the whole network is shown because of the available space. (a) Examples of the tone templates used in computer simulations of the first two experiments, bias  $b = -0.580$ . (b) Examples of the tone templates from the third experiment, bias  $b = -0.723$ .

between two consecutive generations of action potentials. At the same time during the time period of 1–2 ms  $N$  neurons are updated.

(6) An evolution of the network is a sequence of its states in time. It is a trajectory in the state space going through the vertices of  $N$ -dimensional hypercube with an overall number of  $2^N$  vertices. When the time evolution of a stochastic network ends with oscillations around one memory state, it is interpreted as a recall of the given pattern from memory. In this context the recall from memory is a process in which the network re-creates the same pattern of its activity, that was repeatedly spread in it sometimes in the past during the process of learning. Particular pattern of activity is inner representation of corresponding external stimulus. An observable variable is the overlap  $m^\mu(t)$  of the current state  $\{S_i(t)\}$  with individual memory states  $\{\xi_i^\mu\}$ , i.e.,

$$m^\mu(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu S_i(t), \quad \mu = 1, \dots, p. \quad (6)$$

An overlap  $m^\mu$  is a measure of proximity for two configurations in the state space. By following the time evolution of overlaps, we follow the time evolution of the network. The memory states  $\{\xi_i^\mu\}$  are the attractors of the deterministic network without noise. In a stochastic network the configuration will oscillate around one of the memory states, which is in the state space close to the initial configuration.

### 3 Design of the tone representation for the ANN

In the ANN model, the inner representation of any external stimulus corresponds to the pattern of activity of the whole relevant neuron ensemble. In search of a particular geometry of the stored inner representations of the tones, we must follow these conditions:

- (1) Neurons in the modelled nervous structure must have many interconnections, monosynaptic and/or polysynaptic, since the auto-associative properties of ANN are based on the inter-connectivity of the elements.
- (2) These synaptic connections must be plastic, i.e. they undergo changes in efficacy as a consequence of a novel experience.

There is experimental evidence that experience with sounds leads to changes in synaptic efficacies of neurons in several fields of auditory cortex and in one part of auditory thalamus of young and adult mammals [24]. Since we are able to memorize and recognize new sounds throughout the whole life like the animals do, it does not seem unreasonable to extrapolate these findings to humans. Concerning inter-connectivity, neurons in any part of the cerebral cortex have many short-range and long-range vertical and horizontal interconnections [23]. Local interconnections in the auditory thalamus might exist but at present they are not documented, thus we take into account that the two above conditions are met by the auditory cortex.

The frequency representations in the nuclei and cortex of the auditory system have been revealed by determining the neurons' tuning curves. The tuning curve expresses the relationship between the pure tone frequency (i.e. the fundamental frequency alone) and the minimal sound pressure level (SPL) needed to elicit a neuron's response. The frequency at which each neuron has the lowest threshold is called the *best (characteristic) frequency* of the neuron [10].

Microelectrode studies in the cat [25, 26] have revealed several tonotopic organizations in auditory cortex. Best frequencies progress in an antero-posterior direction from low to high in the anterior area (A), from high to low in the primary auditory area (AI), from low to high in the posterior area (P), and from low to high in the ventroposterior (VP) area. Neurons with *similar best frequencies* within each of fields A, AI, P, and VP define '*isofrequency*' stripes that are oriented orthogonal to the low-to-high best-frequency gradients. In anesthetized animals, the vast majority of recorded neurons in areas AI and A (cortical layers III and IV) appear to be narrowly tuned with short-latency responses. Long-latency, broadly-tuned (over several octaves) responses are rarely encountered in these fields. The majority of neurons in fields P and VP are also narrowly tuned and respond in somewhat longer minimum latencies compared to neurons in A and AI. Adjacent to these four primary auditory fields are other auditory responsive areas, including AII where the neurons have much broader tuning and longer response latencies than in AI. To summarize, fields A and AI in auditory cortex exhibit specific characteristics (predominance of narrowly tuned neurons in a precise tonotopic arrangement), field AII exhibits nonspecific characteristics (predominance of broadly tuned neurons and lack of precise tonotopy), and areas P and VP exhibit both specific and nonspecific characteristics. Tonotopic and non-tonotopic fields of the auditory cortex have also been described in monkeys [27, 28].

The subdivisions of auditory cortex vary from species to species [26–28]. The primary auditory cortex (AI) has been identified in all mammalian species studied so far [29], and in all of them the representation of frequencies is tonotopic. Neurons with similar best frequencies within AI define '*isofrequency*' stripes that are oriented orthogonal to the low-to-high best-frequency gradients. We will use the representation of complex tones equivalent to stripe patterns in which every stripe represents one range of frequencies. Neurons in the '*isofrequency*' stripe will be active whenever the corresponding harmonic is present in the spectrum of a given complex tone.

In the construction of particular stripe patterns for the complex tones, further important constraints have to be taken into account.

- (1) Spectra of the tones that are produced by various musical instruments are highly variable in the occurrence and relative intensity of their harmonics [30]. There is a general rule, that on average the intensity of a harmonic decreases with the ordinal number of the harmonic. For instance, for the violine tone  $g=192$  Hz that has 30 harmonics in its spectrum, the relative intensity of individual harmonics higher than the ordinal number 10, is less than 20% of the intensity of any of lower harmonics. In the case of piano tone  $c = 128$  Hz which contains 14 harmonics, the same holds for harmonics higher than ordinal number 7, and so on.
- (2) It is known from psychophysical experiments that accuracy of frequency (pitch) discrimination is a function of frequency. There are individual differences, but in general it is a broad U-shaped dependency, which can be measured very precisely. For instance, humans identify the frequency range of  $(200 \pm 1)$  Hz with the pitch of the frequency equal to 200 Hz, the resolution for 800 Hz tone pitch is  $(800 \pm 2.4)$  Hz, and the resolution for the 12,000 Hz tone pitch is  $(12000 \pm 84)$  Hz, and so on [31]. Since for the frequencies  $\geq 200$  Hz the resolution

decreases as the frequency increases, the frequencies closer to the low border are called resolved and high frequencies (several thousands of Hz) are called unresolved.

Difference in the psychophysical frequency resolution for the low and high frequencies can be related to the cortical surface area devoted to low versus high frequencies. This distribution is known to vary among species. It is interesting to note that in macaque monkey, it seems that low frequencies occupy more space than high frequencies (A. Morel, personal communication). At present, the precise proportion of low versus high frequencies cannot be evaluated in humans, although it is known that in human primary auditory cortex (AI) there is a low-to-high frequency gradient, too [32].

From the viewpoint of the model and its tone representation, these psychophysical and biological relations have an important consequence; that is the number of active neurons in the stripe that represents lower frequencies will be greater than the number of active neurons in the stripe for the higher frequencies. This is a result of both, relatively higher intensities of the lower harmonics of tones (and of naturally occurring sounds), and relatively higher sensitivity to pitch discrimination in lower frequency regions. Higher intensities of lower harmonics imply that more neurons are recruited to process the signal, and in turn more neurons devoted to process particular frequency implies higher resolution. Examples of the complex tone representations that are used in computer simulations of the effect of the missing fundamental are shown in figure 1.

## 4 Simulation of the effect of the missing fundamental

My computer experiments were arranged in a way which is compatible with the three pitch identification psychophysical experiments of Houtsma and Smurzynski [33] and Goldstein et al. [34]. These psychophysical studies deal with pitch identification for complex tones with many successive harmonics which lack the fundamental frequency. The question is raised whether the pitch percept of the missing fundamental is mediated only by low-order resolved harmonics, or whether it can be also conveyed by high-order unresolved harmonics.

Pitch identification was studied as a function of (1) the ordinal number of the lowest harmonic in the acoustic stimulus, and (2) the total number of the harmonics present in the stimulus. For both conditions, the results of psychophysical experiments are presented as the percentage of correctly identified pitches by human subjects. The results of computer simulations are presented as the percentage of correct retrievals of the corresponding tone templates by the model ANN.

We present the results obtained with such value of  $T$  which leads to the best quantitative agreement with the empirical data. In the simulations of the first two experiments  $T = 0.7$ , and in the simulation of the third experiment  $T = 0.5$ . Other values of  $0 < T < 1$  lead to qualitative agreement with the empirical data, i.e. the overall tendencies are preserved but quantitatively the data depart more than 10%.

### 4.1 The first experiment

For the comparison of real data with the result of computer simulations, it is necessary to give a short summary of the design of psychophysical experiments. In the first experiment of Houtsma and Smurzynski [33], four musically experienced subjects listened to complex tones comprising only upper harmonics. Subjects had to identify the note they heard by pressing the appropriate key on a keyboard. The missing fundamental of the presented harmonic complexes was one out of seven possible frequencies: 211.9, 224.5, 237.9, 252.0, 267.0, 282.9, and 299.7 Hz. Pitch identification runs were made for the conditions in which the lowest harmonic number  $n(h)$  was changed from  $n(h) = 7$  through  $n(h) = 19$  in increments of three. The complexes contained 11 successive harmonics of equal amplitude. For each value of  $n(h)$  five runs of 63 trials were taken for each subject, e.g. the total number of trials for every lowest harmonic number was 315.

A 30-dB SPL pink-noise signal formed a constant background to the tones that were raised 20 dB above it. If we hold a pure tone's frequency constant and increase its amplitude  $a$ , the tone's loudness increases. The following equation converts the root mean square pressure,  $p = \frac{a}{\sqrt{2}}$ , into the quantity called sound pressure level (SPL) that is measured in decibels,  $dB$  [35]:

$$SPL = 20 \log \left( \frac{p}{p_0} \right) dB . \quad (7)$$

Here,  $p_0 = 2 \cdot 10^{-5} \text{ N/m}^2$  is a reference pressure which is close to the threshold for hearing in the frequency range 1,000–4,000 Hz. The sound energy  $W = p^2$ . Let  $W_1$  and  $W_2$  represent two values of sound energy, such that  $W_1 = W_2 \times 100$ . Then, according to equation (7), the difference between  $SPL_1$  and  $SPL_2$  is equal to 20

dB. In pink noise, the mean square pressure,  $p^2(f)$ , is inversely proportional to the center frequency  $f$  of the given octave, i.e.  $p^2(f) = \frac{K}{f}$ , where  $K$  is a constant. One octave is a band of frequencies from  $f_0$  to  $2 \times f_0$ . When we calculate the integral SPL for each octave band, we find out that the band SPL for successive octave bands is equal to  $K \ln 2$ , and therefore independent of the center frequency  $f$  [36]. Thus, in the ANN comprised of the 'isofrequency' stripes, the pink noise will be introduced as random activation of neurons with the uniform distribution across the frequencies.

The ANN used in my computer simulation (described in section 2) has  $N = 500$  neurons, that comprise a grid of 100 'isofrequency' stripes with 5 neurons in each of the stripes (see figure 1(a)). There are seven memory patterns, with the fundamental frequencies equal to the tones employed in the psychophysical experiment. Each of the complex tones memory templates is comprised of the fundamental frequency 4-stripe and 29 successive harmonic stripes. The network is small and therefore it can capture general properties of frequency resolution, in the sense of different discrimination sensitivity for different frequencies, only qualitatively. In particular, the stripe resolution in the frequency range 304–1499 Hz is 40 Hz, for the range 1500–3896 Hz it is 82 Hz, for 3897–5888 Hz it is 128 Hz, and for 5888–10240 Hz it is 256 Hz. These numbers have been chosen arbitrarily so that the network size is manageable and the tone patterns are represented in such a way that the two successive harmonics in one tone pattern do not overlap. In addition to decreasing frequency resolution for the stripes representing higher frequencies, the number of active neurons representing high frequencies is relatively smaller than the number of neurons representing lower frequencies. This is captured in a following way (see figure 1(a)): the harmonics with the ordinal number from 2 to 6 are represented by 5 active neurons in the corresponding frequency stripe, harmonics No. 7–11 have 4 (randomly chosen) active neurons, harmonics No. 12–14 have 2 active neurons, harmonic No. 15 has only 1 neuron active, No. 16 has 4 active neurons, No. 17. has 3 active neurons, and harmonics above ordinal number 17 have 2 active neurons in the stripe. These particular numbers have been chosen arbitrarily so that the results of the computer simulation fit the experimental data also quantitatively. The fundamental frequency is an exception – it is represented by 20 active neurons in 4 stripes. This is because the fundamental frequency presented alone as a pure tone, yields the percept of corresponding pitch. In the ANN model, the presentation of the fundamental frequency alone must evoke neural activity of the whole template in 100% of cases, and therefore the initial overlap of the input with the corresponding memory state must be big. In the frame of this model, pitch is represented by the whole memory template, and not by the fundamental frequency alone. Simulations of the template retrieval when the fundamental frequency was presented alone indeed yield 100% correct retrieval.

In the criterion for the correct retrieval of the tone template three conditions must be met simultaneously:

- (1) The overlap between the final network state (after 10 neuronal relaxations) and the desired tone template is the largest compared to the overlaps with all other tone templates.
- (2) The overlap between the activity retrieved at the place of the missing fundamental 4-stripe is the largest compared to overlaps with other missing fundamental 4-stripes.
- (3) The final overlap with the desired fundamental 4-stripe is greater than 0.25.

Conditions 1 and 2 follow logically from the notion that pitch identification is an associative recall of the corresponding template. The third condition has been added arbitrarily so that the number of correct identifications fits the numbers observed experimentally.

The computer simulations were done for 6 inputs, i.e. complexes of 11 harmonics with the lowest harmonic number  $n(h)$  being equal to 7, 10, 13, 16, or 19, like it was done in the actual experiment. For every lowest harmonic number 315 runs of retrieval were ran. The network reaches equilibrium fast, so that after 10 relaxations the average final overlap with the corresponding memory pattern does not improve.

Figure 2 illustrates the input state of the neural network and the last state in the sequence of 10 network state transitions. The signal with SPL that is 20 dB above the SPL of another signal has the energy that is 100 times bigger. Thus, the input contains also randomly activated neurons, distributed uniformly across the frequencies since they are supposed to represent the pink noise. The number of randomly activated neurons is proportional to the ratio between the energy of the signal and the energy of the pink noise. Figure 3 illustrates the evolution of the overlaps (see equation 6) between the current network state and all the memory states during the tone retrieval which is illustrated in figure 2.

Interpretation of the time it takes to identify the pitch, depends on what we consider to be a length of the time interval  $\Delta t$  in asynchronous dynamics. In the case  $\Delta t$  is equal to the absolute refractory period (1–2 ms) then the retrieval of pitch is achieved, on average, 5–10 ms after arrival of the stimulus to the network, what corresponds to 5 neuronal relaxations. If  $\Delta t$  is equal to the relative refractory period then the retrieval time would be approximately ten times longer, e.g. 50–100 ms. This can be considered as a sufficiently short time

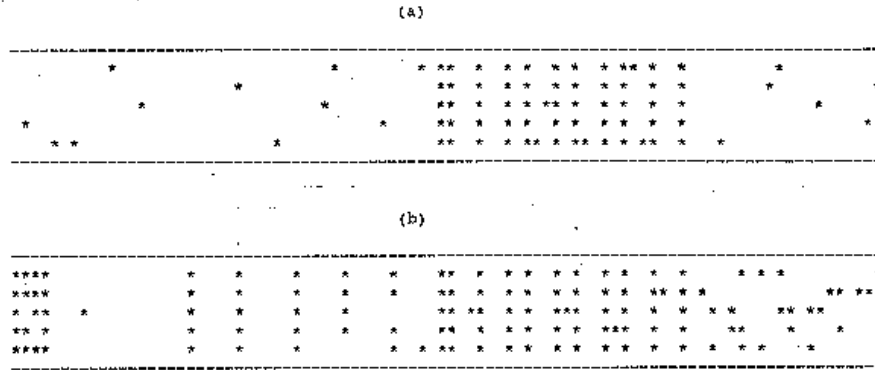


Figure 2: Example of the tone retrieval in the first experiment. An asterisk means that the neuron in the 'isofrequency' stripe is active while blank means that the neuron is not active. (a) Input state of the network is the activity evoked by the frequencies corresponding to 11 harmonics (7–17) from tone No. 1, plus pink noise that activates neurons randomly (see text). (b) Final state of the network after 10 relaxations corresponds to the whole template of tone No. 1. The ANN has  $N = 500$  neurons, and the temperature in neuronal relaxation is  $T = 0.7$ . The coefficient of the soft constraint on relaxation dynamics is  $g = 12$ .

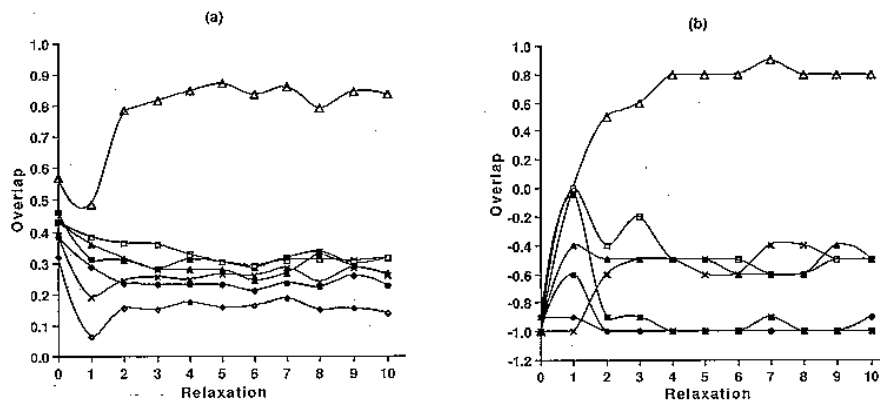


Figure 3: Evolution of the overlaps between the current network state and all the memory patterns during the retrieval of tone No. 1 as illustrated in figure 2. (a) Evolution of overall overlaps. (b) Evolution of partial overlaps corresponding to that part of the representation which covers fundamental frequencies showing that the activity in the fundamental 4-stripe is also retrieved.



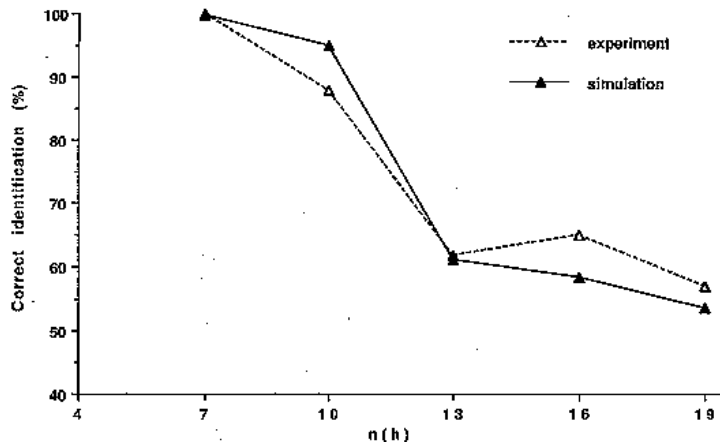


Figure 4: Identification of the pitch of the missing fundamental as a function of the lowest harmonic number  $n(h)$  for the results of actual experiment and my computer simulation from the first experiment. The input stimulus always contains 11 successive harmonics but the ordinal number of the lowest harmonic,  $n(h)$ , varies from 7 to 19 in increment of 3. Each point is the average of 315 trials, i.e. 45 trials for each of the seven tones. The ANN has  $N = 500$  neurons, the temperature in neuronal relaxation is  $T = 0.7$ , and  $g = 12$ .

for accomplishing perceptual task.

Figure 4 shows the percent of correct pitch identification plotted against the lowest harmonic number  $n(h)$  as it was obtained in the computer simulations and in the actual experiment. From the plot of the experimental results [33] we took the average values of those two subjects which participated also in the second experiment (M.H. and J.S.). The results of the actual and computer experiments are in a good quantitative agreement. They show the sharp and monotonic score drop in the region of the lowest harmonic number between  $n(h) = 7$  and  $n(h) = 13$ . This implies that for complexes containing low-order harmonics, the salience of the fundamental pitch percept degrades when harmonic order increases (i.e., when less and less aurally resolved harmonics are present). The performance does not degrade all the way down to the chance level (equal to 14%) as harmonic order  $n(h)$  increases. Instead it reaches an asymptotic level well above chance that is independent of harmonic order.

## 4.2 The second experiment

The aim of the second pitch identification experiment of Houtsma and Smurzynski [33] was to examine the effect of the number of harmonics in the frequency complex on the salience of the pitch of the missing fundamental. The experimental procedure was the same as the one used in the first experiment, except that the total number,  $m$ , of successive harmonics in each sound varied. The lowest harmonic number  $n(h)$  was kept fixed at the values of  $n(h) = 10$  and then of 16. The results of actual and computer experiment are shown in figure 5.

Figure 5 shows the percent of correct pitch identification plotted against the total number of harmonics,  $m$ . As we can see, there is a good quantitative agreement between the experimental and simulated results. The main feature of the data is that performance improves with increasing the number of components  $m$  in the sound, until it reaches an asymptote when adding more harmonics does not improve the performance. For higher harmonics starting with  $n(h) = 16$ , the curves are shifted towards the region of lower percentage of correct identification.

## 4.3 The third experiment

The effect of the number of harmonics  $m$  in the frequency complex on the salience of the pitch of the missing fundamental for different ordinal numbers of the lowest harmonic  $n(h)$ , was investigated also by Goldstein et al. [34]. Listener (N.H.) was presented with eight frequency complexes without the fundamental frequencies and had to identify the pitch he heard. The missing fundamental of these 8 sounds were at 240, 250, 266.7, 281.25, 320, 337.5, 360, and 375 Hz. There was no background noise. Each run included 50 runs for each of

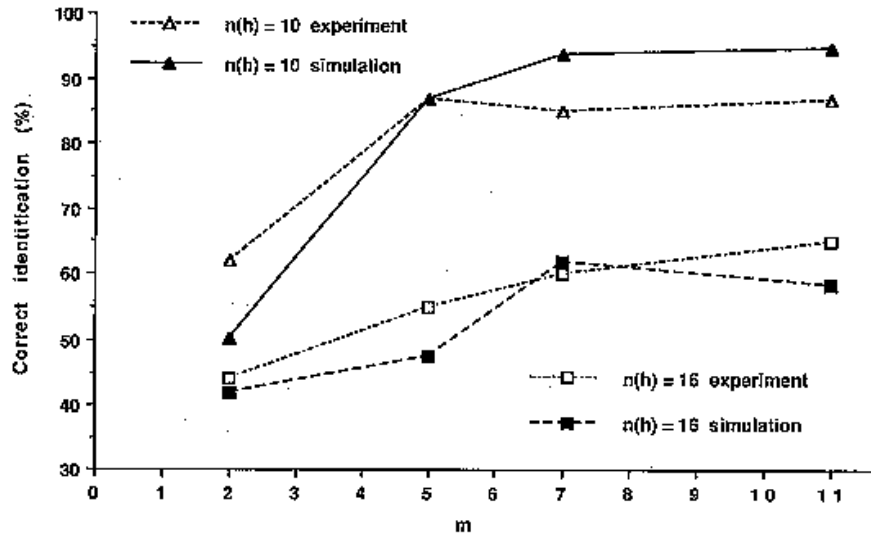


Figure 5: Identification of the pitch of the missing fundamental as a function of the number of harmonics  $m$  for the results of actual experiment and my computer simulation from the second experiment. The ordinal number of the lowest harmonic is fixed, but the number of harmonics,  $m$ , in the stimulus varies. The upper pair of curves corresponds to  $n(h) = 10$  and the bottom pair of curves corresponds to  $n(h) = 16$ . Each point is the average from 315 trials. The ANN has  $N = 500$  neurons, the temperature in neuronal relaxation is  $T = 0.7$ , and  $g = 12$ .

the eight notes. Percents of correct pitch identification with complex tones comprising two or three successive harmonics are shown in figure 6, together with the results of my computer simulations.

For simulation of this experiment new tone templates had to be developed. First, because the used values of the missing fundamental frequencies differed from the values used in Houtsma's and Smurzynski's experiments. Second, the performance for the corresponding total numbers of harmonics, e.g.  $m = 2, 3$ , and the corresponding lowest harmonic number, e.g.  $n(h) = 10$ , in the second and third experiment differ. In the frame of the present ANN model, this difference implies different tone templates. This is not unreasonable, because the subject participating in the third experiment was not the same as any of those in the first two experiments, and they might have had different tone templates reflecting individual differences in musical experience.

Examples of these memory templates are shown in figure 1(b). The network has 455 neurons in 91 stripes with total of 5 neurons in one stripe. There are 8 tone templates each comprised of a fundamental frequency stripe with 20 active neurons, and 21 harmonics in which the number of active neurons decreases with increasing ordinal number of harmonic. There are two neurons active in the stripes corresponding to the harmonics No. 2-3, 5 active neurons for the No. 4-5, 4 active neurons for No. 6-7, 3 active neurons for No. 8-9, 2 active neurons for No. 8-9, and 1 active neuron for the harmonics with the ordinal number greater than 9. The resolution decreases with increasing frequency in such a way that in the interval 375-1499 Hz the resolution is 60 Hz, in the interval 1500-3896 Hz the resolution is 80 Hz, for 3897-5888 Hz it is 120 Hz, and for 5889-10240 Hz it is equal to 256 Hz. All these numbers have been chosen arbitrarily so that the network size is manageable, the two successive harmonics in one tone pattern do not overlap, and the templates lead to the quantitative fit with the experimental results.

Criterion for the correct tone template recall was the same as the criterion used in the previous two computer simulations. Simulations were run for the same values of  $n(h)$  and  $m$  as were used in the actual experiment. There is a good quantitative agreement between the experimental and simulated results. The data show that the performance is better for  $m = 3$  than for  $m = 2$ , i.e. the pitch identification improves with increasing number of components in the sound. For both values of  $m$ , the performance is better for the complexes that begin with the lower harmonic number  $n(h)$ .

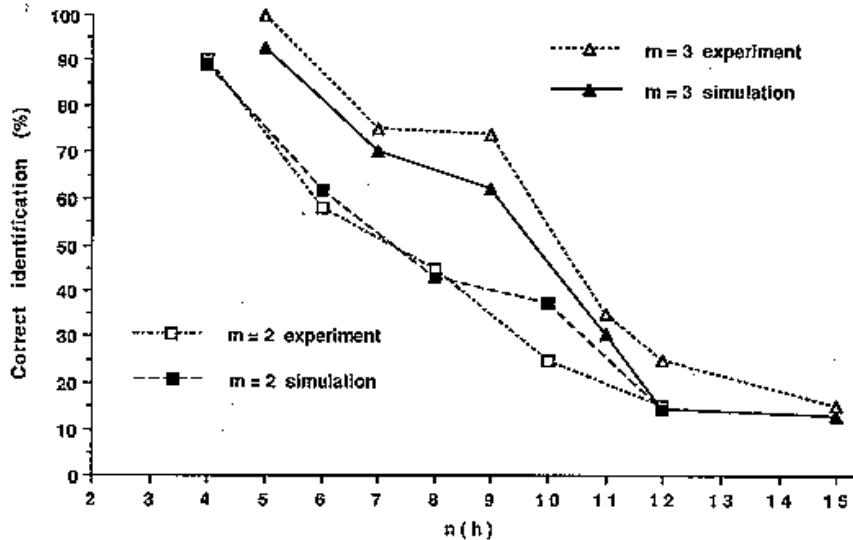


Figure 6: Identification of the pitch of the missing fundamental as a function of the lowest harmonic number  $n(h)$  for the two numbers of harmonics,  $m = 2$  and  $m = 3$  for the results of actual experiment and my computer simulation from the third experiment. Each point is the average from 400 trials, i.e. 50 trials for each of the eight tones. The ANN has  $N = 455$  neurons, the temperature in neuronal relaxation is  $T = 0.5$ , and  $g = 12$ .

## 5 Discussion

Houtsma and Smurzynski [33] pointed out that the results of the first psychophysical experiment, represented in figure 4, suggest a behaviorally based separation of a complex tone's harmonics: those of low order and those of high order, with the separation somewhere between the 10<sup>th</sup> and 13<sup>th</sup> harmonic. According to them this separation might imply either two distinct and separate pitch mechanisms in the auditory system, one for aurally resolved frequencies [17, 18] and the second for unresolved frequencies [37]; or a single mechanism that performs differently for these two kinds of frequencies [38]. An explanation offered by the present ANN model is that there is one cortical mechanism and the differences in its behavior emerge as the consequence of the characteristics of the tone inner representation. Namely, the 'strength' in representation for the resolved and unresolved harmonics differs. The quality of retrieval in terms of the magnitude of the final overlap is proportional to the initial overlap between the input and the corresponding tone template. Thus, the bigger part of the representation devoted to the resolved harmonics conveys the pitch stronger. Second, the frequency resolution in the stripe representation of tones decreases with the ordinal number of harmonics. Therefore there are relatively large overlaps between the representations of individual tones themselves, in that template region which covers unresolved frequencies (see figure 1). Both factors worsen the retrieval of the desired tone template when only high frequencies are presented.

Houtsma and Smurzynski [33] explain the results illustrated in figure 5 (the second experiment) in terms of Goldstein [17] optimum processor theory of pitch perception. According to it, the more harmonics are present in the sound, the less statistical uncertainty exists about the corresponding missing fundamental. The explanation for the general improvement of pitch identification with increasing number of components in terms of the ANN model is similar. Because in this model the pitch perception is an auto-associative recall from memory, the bigger the fragment of memory present at the input the better retrieval from memory.

In the case of the third experiment (figure 6), the accuracy of predictions generated by the ANN model is comparable with the predictions obtained by the optimum processor theory of Goldstein *et al.* [34]. In terms of the ANN model the explanation is again simple, e.g. the bigger the fragment of memory present at the input the better is retrieval from memory. The differences between the results from the second and third experiment for  $m = 2, 3$ , and the lowest harmonic number  $n(h) = 10$ , are attributed to the differences in the tone representations (see figure 1). Namely, worse performance in the third experiment is a consequence of more sparse activity representing harmonics above the ordinal number 10, compared with templates used in the first and second experiments.

Like the optimum processor theory of Goldstein [17] and the virtual pitch theory of Terhardt [18], the ANN model is based on the premise that perception is a template-matching process. Unlike these mathematical models, the ANN model bears strongly on templates that are constructed in accordance with the known neurobiological data for representation of sound frequencies in the primary auditory cortex (AI). When modelling psychophysical experiments that involved different subjects, different tone representations had to be developed. This is not surprising since individual experiences with sounds differ, and so must the inner representation. To capture this more realistically, tone templates should be different for each tone and not only for each individual. As opposed to strictly mathematical optimum theories, the ANN model explains the spectrum-invariant pitch recognition in neurobiological terms, and therefore can offer a basis for designing and interpretation of biological experiments that may directly help in elucidating the actual neural mechanisms that we use in hearing.

Related to the present ANN model is the connectionist neural network for chord classification of Laden and Keefe [19] that also performs matching of templates in the process of simulation of pitch perception. This feed-forward network is able to classify chords well above the chance level also in the case of the missing fundamental or when other harmonics are missing. However, for the connectionist model, the strict quantitative comparison with experimental results has not been done. There is also a difference in how the connectionist model explains the spectrum-invariant pitch perception; pitch identification is a result of generalization from incomplete input by means of hierarchical bottom-up connections, which modify according to the back-propagation algorithm. On the contrary, the ANN model implies that there is a filling-in process in the primary sensory cortex. When only a part of the harmonic complex is present, neuronal activity would appear at the place of the missing fundamental representation in the primary auditory cortex. At present it is possible to test directly this prediction by means of a biological experiment in which the neuronal activity is measured at the 'isofrequency' stripe representing the missing fundamental while the dichotic stimulus consists only of two upper harmonics in order to eliminate the spectral cues conveyed from periphery. The appearance or absence of activity would directly prove or disprove the proposed model. Although, the geometry of the tone representation is based on the data that hold for the primary auditory cortex, it must be admitted that other geometries could also serve the role of the tone template. Therefore, similar experiments should be carried on in parallel on other auditory structures, like other cortical auditory areas and auditory thalamus, in order to see whether and where the filling-in process actually happens.

At present, there is some experimental evidence in favor of the proposed ANN model. In one study, properties of neurons in AI of adult Japanese monkeys were evaluated in response to combinations of successive higher harmonics without the fundamental frequency  $f_0$  [39]. The same neuron responded both to its best frequency  $f_0$  and to the combinations of successive higher harmonics derived from its  $f_0$ , but not to any of these harmonics presented individually. However, since the frequency complexes were presented monotonically, the introduction of the energy corresponding to the missing  $f_0$  by the spectral cues in the periphery cannot be excluded. In another study, Langner et al. [40] performed the measurements of neuromagnetic fields (NMFs) of human auditory cortex in response to complex tones lacking the fundamental frequency. A special synthesis of the tones to minimize the spectral cues was employed. They found that the source of NMF moved to the regions of lower frequencies which were not present in the stimulus. The third line of evidence comes from the studies of patients following bilateral lesions of auditory cortex. Bharucha et al. [41] found that the identification of the pitch of two-harmonic complex tones was significantly worse when the fundamental  $f_0$  was missing in the stimulus (65%) than when  $f_0$  was present (92%). The authors conclude that normal recognition of virtual pitch relies on the integrity of the primary auditory cortex. However, in the chronic stage following ischemic stroke, it seems that non-primary cortical areas and/or subcortical structures are sufficient to account for above-chance performance (65%).

To conclude, it is important to mention that the proposed ANN model for spectrum-invariant pitch recognition can be applied also to invariant recognition of unharmonic sounds; for instance speech. Humans are able to recognize words pronounced with different accents and at different cadence in terms of frequencies. In speech, each syllable is an unharmonic frequency complex that activates particular combination of frequency stripes in the primary auditory cortex. Any harmonic or unharmonic sound can form template that serves as a basis for spectrum-invariant pattern recognition.

## References

- [ 1] Kirkpatrick S and Sherrington D 1978 Infinite-ranged models of spin-glasses *Phys. Rev.* **B17** 4384–4403
- [ 2] Hopfield J J 1982 Neural networks and physical systems with emergent collective computational abilities *Proc. Natl. Acad. Sci. USA* **79** 2554–2558
- [ 3] Amit D J 1989 *Modeling Brain Function. The World of Attractor Neural Networks* (Cambridge: Cambridge University Press)
- [ 4] Gelperin A, Hopfield J J and Tank D W 1985 The logic of *Limax* learning *Model Neural Networks and Behavior*, ed A I Selverston (New York: Plenum Press)
- [ 5] Baird B 1986 Nonlinear dynamics of pattern formation and pattern recognition in the rabbit olfactory bulb *Physica D* **22** 150–175
- [ 6] Koch C, Marroquin J and Yuille A 1986 Analog 'neuronal' networks in early vision *Proc. Natl. Acad. Sci. USA* **83** 4263–4267
- [ 7] Treves A and Rolls E T 1990 Neuronal networks in the hippocampus involved in memory *Preprint* University of Oxford
- [ 8] Kleinfeld D and Sompolinsky H 1989 Associative network models for central pattern generators *Methods in Neuronal Modeling. From Synapses to Networks*, ed C Koch and I Segev (Cambridge: MIT Press)
- [ 9] Hartmann W M 1988 Pitch perception and the segregation and the integration of auditory entities *Auditory Function. Neurobiological Bases of Hearing*, ed G M Edelman, W Einar Gall and W M Cowan (New York: John Wiley and Sons)
- [10] Goldstein E B 1989 *Sensation and Perception, 3rd ed.* (Belmont: Waldsworth)
- [11] Helmholtz von H L F 1963 *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Music* (Braunschweig: F. Vieweg und Sohn)
- [12] Horst J W, Javel E and Farley G R 1986 Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms *J. Acoust. Soc. Amer.* **79** 398 – 416
- [13] Kim D O, Rhode W S and Greenberg S R 1986 Responses of the cochlear nucleus neurons to speech signals: neural encoding of pitch, intensity and other parameters *Auditory Frequency selectivity* ed B C J Moore and R D Patterson (London: Plenum)
- [14] Goldstein J L 1967 Auditory nonlinearity *J. Acoust. Soc. Amer.* **41** 676 – 689
- [15] Licklider J C R 1954 Periodicity pitch and 'place' pitch *J. Acoust. Soc. Amer.* **26** 137(A)
- [16] Houtsma A J M and Goldstein J L 1972 The central pitch origin of the pitch of complex tones: evidence from musical interval recognition *J. Acoust. Soc. Amer.* **51** 520–529
- [17] Goldstein J L 1973 An optimum processor theory for the central formation of the pitch of complex tones *J. Acoust. Soc. Amer.* **54** 1496 – 1516
- [18] Terhardt E 1974 Pitch, consonance and harmony *J. Acoust. Soc. Amer.* **55** 1061–1069
- [19] Laden B and Keefe D H 1991 The representation of pitch in a neural net model of chord classification *Music and Connectionism* ed P M Todd and D G Loy (Cambridge: MIT Press)
- [20] Hebb D O 1949 *Organization of Behavior* (New York: John Wiley and Sons)
- [21] Amit D J, Gutfreund H and Sompolinsky H 1987 Information storage in neural networks with low levels of activity *Phys. Rev. A* **35** 2293 – 2303
- [22] Peretto P and Niez J-J 1986 Stochastic dynamics of neural networks *IEEE Trans. Syst. Man Cybern.* **16** 73–83
- [23] Kandel E R and Schwartz J H 1985 *Principles of Neural Science, 2nd ed.* (New York: Elsevier)
- [24] Weinberger N M and Diamond D M 1988 Dynamic modulation of the auditory system by associative learning *Auditory Function. Neurobiological Bases of Hearing* ed G M Edelman, W Einar Gall and W M Cowan (New York: John Wiley and Sons)
- [25] Merzenich M M, Knight P L and Roth G L 1975 Representation of cochlea within primary auditory cortex in the cat *J. Neurophysiol.* **28** 231–249
- [26] Reale R A and Imig T J 1980 Tonotopic organization in auditory cortex of the cat *J. Comp. Neurol.* **192** 265–292

- [27] Morel A, Garraghty P E, Schwaber M K, Burch-Sims P and Kaas J H 1991 Tonotopic organization, architecture and connections of auditory cortex in macaque monkeys *Soc. Neurosci. Abstr. I* **17** 305
- [28] Morel A and Kaas J H 1992 Subdivisions and connections of auditory cortex in owl monkeys. *J. Comp. Neurol.* **318** 27–63
- [29] Aitkin L M, Irvine D R F and Webster W R 1980 Central neural mechanisms of hearing *Handbook of Physiology – The Nervous System III* (Washington, D.C.: Am. Physiol. Soc.)
- [30] Meyer E and Buchmann G 1933 Die klangenspektren der musikinstrumente *Berliner Berichte, physical. mathem. Klasse* **32** 735-778
- [31] Špelda A 1958 *Úvod do akustiky pro hudebníky* (Praha: Státní nakladatelství krásné literatury, hudby a umění )
- [32] Lauter J L, Herscovitch P, Formby C and Raichle M E 1985 Tonotopic organization in human auditory cortex revealed by positron emission tomography *Hearing Res.* **20** 199–205
- [33] Houtsma A J M and Smurzynski J 1990 Pitch identification and discrimination for complex tones with many harmonics *J. Acoust. Soc. Amer.* **87** 304-310
- [34] Goldstein J L, Gerson A, Srulovicz P and Furst M 1978 Verification of the optimal probabilistic basis of aural processing in pitch of complex tones *J. Acoust. Soc. Amer.* **63** 486–497
- [35] Ford R D 1970 *Introduction to Acoustics* (Amsterdam: Elsevier)
- [36] Pierce A D 1981 *Acoustics. An Introduction to its Physical Principles and Applications* (New York: McGraw-Hill)
- [37] Schouten J F 1949 The residue and the mechanism of hearing *Proc. Kon. Acad. Wetenschap.* **43** 991–999
- [38] Srulovitz P and Goldstein J L 1983 A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum *J. Acoust. Soc. Amer.* **63** 486–497
- [39] Riquimaroux H, Takahashi T and Hashikawa T 1993 Temporal pitch might be already place-coded together with place pitch in the primary auditory cortex of the Japanese monkey *Soc. Neurosci. Abstr.* **19** 1423
- [40] Langner G, Samps M, Heil P, McEvoy L K, Hari R and Ahonen A 1993 Spatial representation of periodicity pitch in the human auditory cortex *Soc. Neurosci. Abstr.* **19** 1423
- [41] Bharucha J J, Tramo M J and Zatorre R J 1993 Abstraction of the missing fundamental following bilateral lesions of auditory cortex *Soc. Neurosci. Abstr.* **19** 1687