

Combination Tones in the Model of Central Auditory Processing for Pitch Perception

Tadeusz ZIEBAKOWSKI

Institute of Manufacturing Engineering, West Pomeranian University of Technology
Al. Piastów 19, 70-310 Szczecin, Poland; e-mail: tezet@zut.edu.pl

(received November 8, 2011; accepted November 9, 2012)

This work addresses the problem of difficulties in classical interpretation of combination tones as non-linear distortions. One of the basic problems of such an interpretation is to point out the sources of these distortions. Besides, these kinds of distortions have numerous “anomalies” which are difficult to explain on the grounds of physics or physiology. The aim of the model presented in this paper is to show that combination tones phenomenon can be explained as an effect of central mechanisms. Most of existing theories of pitch perception focus mainly on virtual pitch perception and do not take into account combination tones as an element of the same mechanism. The proposed model of central auditory processing for pitch perception allows one to interpret in a coherent way both virtual pitches and combination tones phenomena. This model is of a demonstrative nature and gives an introduction to more advanced model. It belongs to the class of spectral models and it will be shown that such a model can be in a simple way extended to spectral – time model which is partially consistent with autocorrelation models.

Keywords: combination tone, virtual pitch, pitch perception, central auditory processing.

1. Introduction

Pitch perception is closely connected with spectral analysis which is for example manifested in the fact that sounds with different frequencies can be ranked by pitch in a proper frequency-related order. The connection between pitch perception and spectral analysis was explicitly noticed by G. OHM who in 1843 announced the psychophysics law which states that in complex sounds individual pitches are distinguished only if the spectrum of the sound contains appropriate spectral components. More precise observations show, however, that this law is not completely correct. Already in the times when Ohm published his law the phenomenon of combination tones was known i.e. additional audible tones to which correspond no frequency components contained in the source sound. More or less at the same time the interpretation problem of the pitch of complex sounds also was noticed which as a result of farther investigations led to discovery of the residue phenomenon and connected with it the phenomenon of virtual pitch. The phenomena of virtual pitch and combination tones can be observed e.g. in the situation when the source of the sound contains two sinusoidal components with different frequen-

cies (two-tone). An occurrence of these phenomena depends however closely on the location of spectral components in the frequency domain. An example of pitch perception image for a two-tone is shown in Fig. 1. Such an image will be called a *psychophysical spectrum*. The basic information about virtual pitches and combination tones will be now briefly presented.

Firstly, the effect of virtual pitch was observed in harmonic complex from which its fundamental frequency had been removed. It turned out that the pitch of fundamental frequency is also perceived when it had been removed from harmonic complex. In the case of two-tone virtual pitches are observed in a narrow band around the frequency of $\omega_2 - \omega_1$. A distinctive feature of virtual pitches is the fact that they are difficult to jam with noise. What is more, in case of two-tone a noise which is not very intensive can even amplify the audibility of virtual pitches. Virtual pitches are much stronger effects in harmonic complexes and their audibility increases along with an increase of the amount of source tones. Initially it was thought that the mechanism of generating virtual pitches can be almost fully understood on the basis of the ear's physiology. The discovery of binaural perception of virtual pitches (HOUTSMA, GOLDSTEIN, 1972; HOUTSMA,

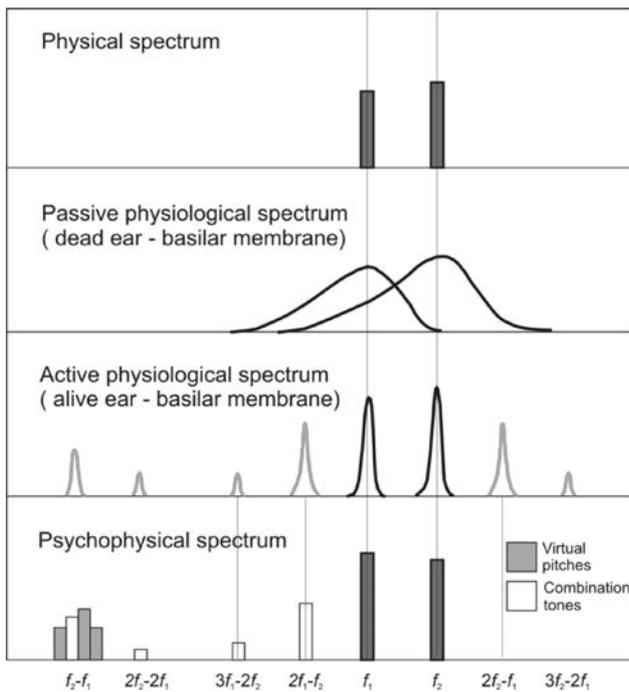


Fig. 1. Differentiation of spectra for two-tone on different levels of auditory process.

2007) was a strong argument for the central origin of the phenomenon and it triggered further development of central theories.

There are many theories (models) of virtual pitch perception, which can be divided into two categories: spectral (place) theories and temporal (time) theories. The basis of spectral theories is the assumption that the basis for pitch identification is an approximate spectral image of the stimulus generated at the level of the inner ear, which is coded by vibrations of the basilar membrane. Such a spectral image will be called as a *physiological spectrum*¹ (see Fig. 1). On the other hand, temporal theories are based on time distribution of spikes generated as a result of sound stimulus in auditory nerve fibres. The best known theories of the first category are presented in GOLDSTEIN (1973), WIGHTMAN (1973) and TERHARDT (1974). At its simplest, the main idea of these theories is to select the best fitted pattern of harmonic complex to a given group of tones. Because they operate in frequency domain they assume resolution of spectrum components. However, a virtual pitch can also occur with unresolved components. These kinds of situations are better dealt with theories of the second category, which are to be found in studies of LICKLIDER (1951), MEDDIS and HEWITT (1991a; 1991b) or MOORE (1997). Pitch estimation in these models is connected with an effect of synchrony in response of auditory nerve fibres to sound stimulus and relies in general on using autocorrelation function

¹This spectrum is different in dependence on if the ear cooperates with nervous system or not. Hence the terms: *active* and *passive physiological spectrum* have been introduced in Fig. 1.

(ACF) for firing rate of spikes in auditory nerve. However, these models too are subject to some limitations, because the effect of synchrony is not observed above 5 kHz which reduces the applicability of temporal theories up to the frequency of 5 kHz. A more comprehensive review of these issues can be found in CHEVEIGNE (2005).

Combination tones contrary to virtual pitches have somewhat different features. They behave as additional real components introduced into sound spectrum: they are subject to beating with tones of near frequency introduced into the stimulus and they can be masked by a noise band centred at their frequency. Frequencies of combination tones are linear combinations of primary tones with low integers. This triggered a hypothesis formulated by HELMHOLTZ (1856), saying that combination tones are nonlinear distortion products generated within the ear². Helmholtz showed that an addition of displacement square factor to equation of harmonic oscillator driven by two sinusoidal forces introduces to the oscillator's vibration spectrum additional components (HELMHOLTZ, 1863). Helmholtz solved this equation by perturbation method and he obtained in the first order of calculations "square" factors with frequencies of $\omega_2 - \omega_1$, $\omega_1 + \omega_2$, $2\omega_1$ and $2\omega_2$, and in the second order of the calculations "the third order" factors with frequencies of $2\omega_1 - \omega_2$, $2\omega_1 + \omega_2$, $2\omega_2 - \omega_1$, $2\omega_2 + \omega_1$, $3\omega_1$ and $3\omega_2$. Helmholtz's hypothesis is generally accepted up till now, although there are facts that can challenge its credibility. Given that combination tones happen as a result of nonlinear distortions it is noteworthy to think that we can only hear so few combination tones. The most prominent combination tones are the tones of odd orders of type $\omega_1 - n(\omega_2 - \omega_1)$ for $n = 1, 2 \dots 6$, $\omega_2 > \omega_1$ in particular $2\omega_1 - \omega_2$ tone and also the difference tone $\omega_2 - \omega_1$ (PLOMP, 1965; 1976; SMOORENBURG, 1972a; 1972b). It is noteworthy to think that the tone of the third order $2\omega_1 - \omega_2$ ($n = 1$) is audible at lower levels of the sound than the tone of the second order $\omega_2 - \omega_1$. Tones above the frequency of ω_1 are not usually audible. The tone of $2\omega_1 - \omega_2$ is particularly not accompanied by the tones of $2\omega_1 + \omega_2$ and $2\omega_2 + \omega_1$ whereas the audibility of $2\omega_2 - \omega_1$ is sporadically reported. Apart from that a certain "anomaly" in perception of combination tone $2\omega_1 - \omega_2$ was observed. If amplitudes of tones ω_1 , ω_2 are identical and equal to x then the nonlinear tone $2\omega_1 - \omega_2$ should increase initially as x^3 whereas in fact it increases smaller than the first power (ZWICKER, 1955; 1968; GOLDSTEIN, 1967; HELLE, 1969/70; SMOORENBURG, 1972a; 1972b; ZWICKER, FASTL, 1973). The audibility of combination tones of type $\omega_1 - n(\omega_2 - \omega_1)$ strongly depends on the ratio of primary tones frequency and is limited

²Some authors consider also so-called *extra-aural combination tones* in contrast to *intra-aural* considered in this work, which sources are outside the ear (LOHRI *et al.*, 2011).

to the range of $\omega_2 : \omega_1$ from about 1.1 to about 1.5 (PLOMP, 1965; SMOORENBURG, 1972a; 1972b). Moreover, their levels increase markedly with decreasing $\omega_2 : \omega_1$ (ZWICKER, 1955; GOLDSTEIN, 1967; GREENWOOD, 1971; SMOORENBURG, 1972a; 1972b). Contrary to virtual pitches, binaural perception of combination tones is not observed psychophysically.

The problems outlined above are more in-depth discussed in PLOMP (1976) and DE BOER (1984; 1991). Singularities in combination tone perception are tried to be explained at the level of the ear's physiology. Some kind of spectral sound analysis is conducted directly in the cochlea of the inner ear. Owing to an appropriate mechanical construction of the cochlea spectral information is directly represented as the maxima of envelope of basilar membrane's vibrations, whereas various frequencies correspond to various places of maximal stimulation of the membrane. In order to determine the sources of nonlinear distortions the mechanisms of both the middle and inner ear were examined. The measurements of transmittance of the middle ear point out to its linear behaviour in a quite wide range of changes of acoustic pressure. However, an analysis of the basilar membrane in the inner ear showed maxima in places corresponding to spectrum components of nonlinear distortions (ROBLES *et al.*, 1997). However, these maxima were shown only in the situation when there was cooperation between the ear and the auditory nervous system.

The research conducted over the past decades shows that sound is not the only source of mechanical energy determining the behaviour of the cochlea. Outer hair cells (OHC's) which are found in the organ of Corti, which are detectors of sound information can also move as a result of the action of efferent neurons which provide information from the brain (ASHMORE, 2008). This feature of OHC's causes that some vibrations of basilar membrane in the cochlea can be amplified while the other can be damped. In this way auditory system realise the feedback between the central and the peripheral levels. Studies conducted by RUGGERO and RICH (1991) indicate that combination tones closely depend on the condition of OHC's. It turns out that characteristic for a membrane controlled by signals from the auditory nerve a nonlinear relation of the basilar membrane's displacement depending on an acoustic stimulus changes into a linear relation when function of OHC's is disturbed. Therefore, some researchers assume that the sensory cells in the cochlea are the source of the nonlinearity (EGUILUZ *et al.*, 2000).

There is also one more source of information regarding combination tones, i.e. otoacoustic emission (OAE). When the ear is stimulated by two-tone burst containing spectral components ω_1, ω_2 , ($\omega_2 > \omega_1$), an emission of sound containing spectral components with frequencies of $|n\omega_1 \pm m\omega_2|$, where n and m are some

small integers, is registered in the external ear canal (the so-called DPOAE). These kinds of studies were conducted on various species of vertebrates, including mammals (e.g. recently MICHALSKI *et al.*, 2011), birds, reptiles and amphibians (e.g. VAN DIJK, MANLEY, 2001). The ears of many of these animals significantly differ from the human ear e.g. the frog's ear does not have a basilar membrane. It is interesting that nevertheless DPOAE $2\omega_1 - \omega_2$ is observed as a dominating component.

Concluding, physiological research has shown that the ear, and in particular the inner ear, is complex, controlled by nervous system electromechanical device which can analyze sounds and also can generate sounds. Most of widely accepted theories of combination tones assume to model the vibrating elements of the inner ear by sound-driven harmonic nonlinear oscillator (HELMHOLTZ, 1863; EGUILUZ *et al.*, 2000; VAN DIJK, MANLEY, 2001). However, these theories do not explain, how locally vibrating elements generate globally combination tones. Furthermore, "the origins of many of the nonlinearities in the mechanics of the cochlea remain a contentious issue" (ASHMORE, 2008).

The indicated difficulties in interpreting combination tones as nonlinear phenomena raise the question of whether or not another explanation of the phenomenon is possible. In the present study an alternative thesis was assumed that combination tones are the product of the central mechanisms which are responsible for pitch extraction. One of the main problems with locating the source of combination tones at the central level is the fact that these tones are to be found at the peripheral level. However, this can be justified by that the feedback between the central and peripheral levels puts combination tones to the ear by OHC (Fig. 2).

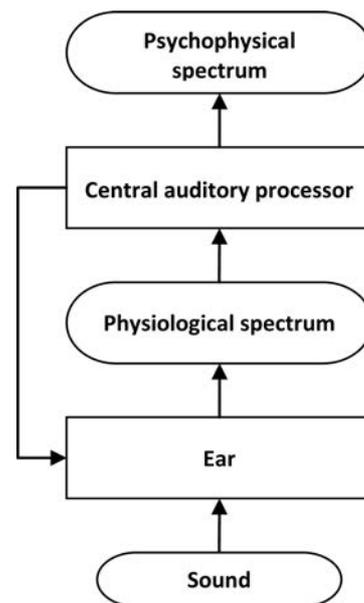


Fig. 2. Schematic diagram of the location of the auditory processor in auditory system.

A simplified model of central auditory processing for pitch perception, which describes generation of both combination tones and virtual pitches, will be presented in the next chapters. Following on from GOLDSTEIN (1973) this model will be called an *auditory processor*. The presented model will belong to the class of spectral models and owing to assumptions made earlier the description of virtual pitches will be far insufficient. The model proposed in this work is of demonstrative nature and gives an introduction to more advanced model. However, an advantage of such a model is easy presentation of the proposed mechanism of combination tone generation. Furthermore, it will be shown that such a model can be in a simple way extended to spectral – time model which in the scope of virtual pitch phenomena will be partially consistent with autocorrelation models. In proposed construction, in an elementary scope, the notions from the theory of group algebras (convolution algebras) will be used.

2. Basic construction elements of model

Sound pitch sensation refers to a very broad class of signals including also some forms of noises. However, in order to simplify our considerations the scope of signals taken into account can be restricted to almost-periodic ones. We will consider amplitude spectra of such signals which we will describe with finite linear combinations of Dirac impulses with nonnegative coefficients:

$$F = \frac{1}{\lambda} \sum_{\omega \in \Omega} F_{\omega} \delta_{\omega} \quad \text{or} \quad F = \frac{1}{\lambda} \sum_{\omega \in \Omega} F_{\omega} \langle \omega \rangle, \quad (1)$$

where $\Omega \subset R$ is a finite set of frequencies and $F_{\omega} \geq 0$, $\langle \omega \rangle := \delta_{\omega}$ is the Dirac measure in the point $\omega \in R$ and $\lambda \in R$ is a normalizing constant such that norm $\|F\| := \frac{1}{\lambda} \sum_{\omega \in \Omega} F_{\omega}$ satisfy $\|F\| < 1$. The space of all discrete measures on R containing such spectra will be denoted by $M^1(R)$. Sets of frequency differ from R which have some group structure G will be also considered and also the spaces $M^1(G)$ of discrete measures on G .

We start our considerations from observation that from mathematical point of view the convolution (i.e. $*$ -operation) describes generation of combination tones, because e.g.:

$$\delta_{\omega_2} * \delta_{-\omega_1} = \delta_{\omega_2 - \omega_1}, \quad \delta_{\omega_1} * \delta_{\omega_1} * \delta_{-\omega_2} = \delta_{2\omega_1 - \omega_2}. \quad (2)$$

If G is a group, then $M^1(G)$ has a well defined convolution³.

³We will use also multiplicative group notation instead of additive. If $g_i \cdot g_j'$ is multiplication of $g_i, g_j' \in G$ then convolution is:

$$\left(\sum_i A_{g_i} \delta_{g_i} \right) * \left(\sum_j B_{g_j'} \delta_{g_j'} \right) = \sum_{i,j} A_{g_i} B_{g_j'} \delta_{g_i \cdot g_j'}.$$

Pitch perception process will be described in the form of “psychophysical” spectrum sequence $\{P_i\}_{-\infty}^{\infty}$ which is generated on the basis of an appropriate sequence of physical spectra $\{F_i\}_{-\infty}^{\infty}$. More precisely, it generally will be defined as an infinite iteration process Ψ , in which spectrum $P^{(n+1)}$ is created on the basis of previous spectra $P^{(k)}$ and $F^{(k)}$ for $k \leq n$:

$$P^{(n+1)} = \Psi \left(\{P_i\}_{-\infty}^n, \{F_i\}_{-\infty}^n \right). \quad (3)$$

The “linear” formula is the one of the simplest possibilities to determine iteration process Ψ :

$$P^{(n+1)} = Q_F^{(n)} * P^{(n)} + \delta_0, \quad (4)$$

where $Q_F^{(n)} \in M^1(X)$ is a factor dependent on physical spectrum $F^{(n)}$. It can be noticed that formula (4) is similar to the formula $S^{(n+1)} = qS^{(n)} + 1$ of generation the $n+1$ partial sum of the geometric series $1, q, q^2, q^3 \dots$. Formula (4) has also its deeper psychophysical meaning connected with the rise and decay of pitch sensation in response to a rectangular pulse stimulus (PLOMP, 1964). Such pitch sensation varies approximately geometrically. It will be farther more thoroughly discussed. One can assume in the simplest case that

$$Q_F = \alpha F^{(n)} + \beta \delta_0, \quad (5)$$

where $\alpha, \beta \in R$ and δ_0 is Dirac measure in zero (convolution identity).

When process (4) is a stationary iterative one i.e. when $F^{(n)} = F = \text{const}$, the natural task is to determine criteria of convergence and points of convergence of such a process. Then, one can notice that for $\|Q_F\| < 1$ mapping $\Psi: P \mapsto Q_F * P + \delta_0$ is contracting-mapping and using contracting-mapping principle the psychophysical limit spectrum P can be determined as a fixed point of Ψ . The situation, in which $F^{(n)} = F \equiv 0$ will be called the *silence state*. In such a case $Q_F^{(n)} = \beta \delta_0$ and from Eq. (4) psychophysical spectrum is

$$P = \frac{1}{1 - \beta} \delta_0. \quad (6)$$

The next step in the proposed construction is to introduce into $M^1(G)$ such a structure which can split pitch sensations into tonal and virtual ones. In the simplest case one can do it by introducing two sets of frequencies: tonal T and virtual V which satisfy $G = T \cup V$. Moreover, a group structure should be introduced in the set G . One can do it, assuming that G is an extension of subgroup V by two element group Z_2 . It can be realized by simple or semi-simple product. If θ denotes generator of group Z_2 then $G = V \cup \theta V$. A division G into tonal and virtual sets implies that any spectrum P can be written down as a sum of two components: tonal P_T and virtual P_V i.e. $P = P_T + P_V$. It is assumed that every physical spectrum F has only tonal part, so $F_V \equiv 0$ and so:

$$F = \theta \frac{1}{\lambda} \sum_{\omega \in \Omega} F_{\omega} \delta_{\omega}. \quad (7)$$

The above considerations will now be illustrated by a simple example.

EXAMPLE. Perception of a single tone in the form of a rectangular pulse (see PLOMP, 1964).

Let $G = Z_2 \times Z$ be a semi-simple product group determined by relation $\theta \langle \omega \rangle \theta = \langle -\omega \rangle$ for integers $\langle \omega \rangle \in Z$ and $\theta \in Z_2$ (the brackets “ $\langle \rangle$ ” denote passing from additive to multiplicative notation i.e. $\langle \omega \rangle \cdot \langle \omega' \rangle = \langle \omega + \omega' \rangle$, $\langle 0 \rangle = 1$, moreover, the symbol $\langle \omega \rangle$ in dependence on the context may denote an element of group G or an element of space $M^1(G)$ with amplitude 1. The tonal spectrum determines values on the elements of $\theta \langle \omega \rangle$ type, but virtual spectrum determines values on the elements of $\langle \omega \rangle$ type. Let $F^{(n)} = F = \theta \langle \omega \rangle$ be a spectrum of one tone with the frequency ω and amplitude 1, and $Q_F = \beta \delta_0 + \alpha F$ where $\alpha = \beta = q/2$ and $0 < q < 1$. Then $\|Q_F\| = \|\alpha F + \beta \delta_0\| \leq \|\alpha F\| + \|\beta \delta_0\| = q < 1$ so process (4) is convergent. Starting from the spectrum of the silence state: $P = \frac{1}{1-\beta} \delta_0$ one gets the next psychophysical spectra for $n = 1, 2, 3, \dots$ on the basis of (4):

$$P^{(n)} = \frac{(2-q)^2 - q^{n+1}}{2(1-q)(2-q)} \langle 0 \rangle + \frac{q(2-q-q^n)}{2(1-q)(2-q)} \theta \langle \omega \rangle. \quad (8)$$

The process $\{P^{(n)}\}$ is convergent to the spectrum $P = \frac{(2-q)}{2(1-q)} \langle 0 \rangle + \frac{q}{2(1-q)} \theta \langle \omega \rangle$. When we reset component $\theta \langle \omega \rangle$ in the physiological spectrum from a certain moment i.e. since some $n_0 F^{(n)} = F \equiv 0$ for $n > n_0$ then process $\{P^{(n)}\}$ comes back to the silence state through the following sequence of psychophysical spectra:

$$P^{(n)} = \frac{2(1-q+2^{-n}q^n)}{(1-q)(2-q)} \langle 0 \rangle + \frac{2^{-n}q^n}{(1-q)} \theta \langle \omega \rangle. \quad (9)$$

The above example has been illustrated in Fig. 3. Let us also observe the role of factor $\beta \delta_0$ in Eq. (5).

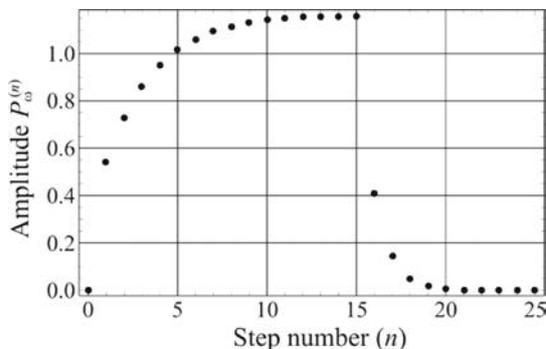


Fig. 3. Pitch perception $P_{\omega}^{(n)}$ of single tone in the form of the rectangular pulse: $F_{\omega}^{(n)} = 1$ for $n = 1 \dots 15$, and: $F_{\omega}^{(n)} = 0$ for $n > 15$ ($q = 0.7$ in Eqs. (8) and (9)).

If $\beta = 0$ at $F \equiv 0$ then $Q_F \equiv 0$ and therefore reset of the physiological spectrum resulted in an instantaneous jump to the silent state, which has not been evidenced in the experiments.

3. Auditory filters

Auditory filters have been introduced to psychoacoustics in order to describe such phenomena as the masking phenomenon, the critical bands phenomenon, or the phenomenon of tuning. These phenomena are observed psychophysically and also are studied on the grounds of physiology of the ear. Although the mechanism of functioning of auditory filters is not fully known, most researchers suppose that it functions at the peripheral level. However, one can assume, as we do in this work, that as in case of combination tones this mechanism is located at the central level and is reflected at the peripheral level through the loop-back. In our model of auditory processor, introduction of filters allows one to reduce the amount of spectral components taking part in processing described by Eq. (4). The mechanism proposed in Eq. (4) makes it possible to generate for every pair of components ω_1, ω_2 , additional spectral components with frequencies of $|n\omega_1 \pm m\omega_2|$, where n and m are any given integers, which in fact is not observed. Apart from this auditory filters will function as a template used to estimate virtual pitch. Filtering will concern the both tonal and virtual sets of frequencies. In the structure of filters presented in Fig. 4 in the tonal part the comb filter is set for generating fundamental tone and harmonics whereas in the virtual part the filter with three passbands is set for only such virtual pitch which comes from neighbouring harmonics. For simplicity, rectangular outlines of filters and equal width of passbands are assumed. The coherence of the model requires to take into consideration two auxiliary passbands around zero frequency: unperceivable subvirtual pitch (virtual zero) and unperceivable zero combination tone (tonal zero).

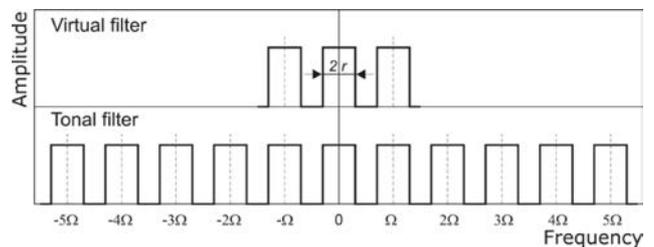


Fig. 4. Filters’ characteristics for given Ω .

The designed processor will be equipped with a collection of virtual-tonal auditory filters which will be indexed (like in harmonic complex) by fundamental frequency Ω . Mathematically, filters can be described by means of filters’ characteristics, i.e. function

$\varphi_\Omega: T \cup V \rightarrow \{0, 1\}$ where Ω runs through the set Ω_V of all the frequencies corresponding to perceived virtual pitches. An assumption is made that values of the set Ω_V create a finite arithmetic progression:

$$\Omega_V = \{\Omega_1, \Omega_2, \dots, \Omega_{\max}\}, \tag{10}$$

where $\Omega_{i+1} = \Omega_i + \Delta$.

The initial Eq. (4) defining pitch perception will be rewritten on respective filters, on the condition that on individual filters the spectra processing occurs independently. Accordingly, a certain partial psychophysical spectrum $P_\Omega^{(n)}$ can be connected with every pair of tonal-virtual filters for given $\Omega \in \Omega_V$ so:

$$P_\Omega^{(n+1)} = \varphi_\Omega \cdot (Q_F * P_\Omega^{(n)}) + \delta_0. \tag{11}$$

The dot denotes multiplying the spectrum (as a measure) by the function. All of partial spectra create the final pitch spectrum according to the rule of superposition:

$$P^{(n)} = \sum_{\Omega \in \Omega_V} P_\Omega^{(n)}. \tag{12}$$

It is easy to show that partial psychophysical spectrum $P_\Omega^{(n)}$ can be non-zero only on filter passbands.

4. Determination of psychophysical spectra for periodic signals

Amplitude spectra of periodic signal are discrete spectra, in which the components are equally spaced by a fixed width $\Delta\omega$. In our model this situation is described by the group $G = Z_2Z$ with the defining relation $\theta\omega\theta = \omega^{-1}$ which⁴ in the additive notation is given by $\theta\langle\omega\rangle\theta = \langle-\omega\rangle$ where $\omega \in Z, \theta \in Z_2$. Group G as a set can be presented as $G = Z \cup \theta Z$ what shows the division of group G into the virtual and tonal parts. The space $M^1(G)$ can be in this case identified with $L^1(G)$, so instead operate on measures we can operate on functions for which the convolution is defined by the formula:

$$f * g(x) = \sum_{y \in G} f(y)g(y^{-1}x), \tag{13}$$

$$f, g \in L^1(G), \quad x \in G.$$

Additionally, owing to the discrete character of the group and due to the limitation of frequency ranges by the auditory filters the consideration of spectra can be narrowed down to finite dimensional linear spaces, so Eq. (11) can be written in the matrix form. Let $N > 0$ be the number of the highest audible component. Let us define the following subset in G :

$$B = \{\langle -N \rangle, \langle 1-N \rangle, \dots, \langle 0 \rangle, \dots, \langle N+1 \rangle, \langle N \rangle, \theta\langle -N \rangle, \theta\langle 1-N \rangle, \dots, \theta, \dots, \theta\langle N-1 \rangle, \theta\langle N \rangle\}. \tag{14}$$

⁴One can also assume that $\theta\omega\theta = \omega$ but relation $\theta\omega\theta = \omega^{-1}$ allows one to make the consideration simpler and more transparent.

This set can be treated as a certain base of linear subspace $W \subset L^1(G)$ generated by this set. In base B the element $Q_F \in W$ can be written as:

$$Q_F = [V_{-N}V_{1-N}, \dots, V_N, T_{-N}, T_{1-N}, \dots, T_N]^T, \tag{15}$$

where the first half is the virtual and the second the tonal part.

Let us now turn to writing Eq. (11) in the matrix form in the base B . Let us now take the element $Q_F \in W$ defined by Eq. (15) and let us write down the operation $Q_F * (\cdot)$ in the base B as mapping $\mathbf{Q}_F: W \rightarrow W$ given by as follows:

$$\mathbf{Q}_F: W \ni f \rightarrow \varphi_B \cdot (Q_F * f) \in W, \tag{16}$$

where

$$\varphi_B(\langle i \rangle) = \varphi_B(\theta\langle i \rangle) = \begin{cases} 1 & \text{for } |i| \leq N, \\ 0 & \text{for } |i| > N. \end{cases}$$

The matrix of this mapping in base B has the following form:

$$\mathbf{Q}_F = \begin{bmatrix} \mathbf{V} & \mathbf{T}^T \\ \mathbf{T} & \mathbf{V}^T \end{bmatrix}, \tag{17}$$

where

$$\mathbf{V} = \begin{bmatrix} V_0 & V_1 & \dots & V_N & & 0 \\ V_{-1} & V_0 & V_1 & \vdots & V_N & \\ \vdots & V_{-1} & \ddots & \ddots & \vdots & \ddots \\ V_{-N} & \vdots & \ddots & \ddots & \ddots & \vdots & V_N \\ & & & V_{-N} & \vdots & \ddots & \ddots & \vdots \\ & & & & \ddots & \vdots & \ddots & \ddots & V_1 \\ 0 & & & & & V_{-N} & \dots & V_{-1} & V_0 \end{bmatrix}, \tag{18}$$

$$\mathbf{T} = \begin{bmatrix} T_0 & T_1 & \dots & T_N & & 0 \\ T_{-1} & T_0 & T_1 & \vdots & T_N & \\ \vdots & T_{-1} & \ddots & \ddots & \vdots & \ddots \\ T_{-N} & \vdots & \ddots & \ddots & \ddots & \vdots & T_N \\ & & & T_{-N} & \vdots & \ddots & \ddots & \vdots \\ & & & & \ddots & \vdots & \ddots & \ddots & T_1 \\ 0 & & & & & T_{-N} & \dots & T_{-1} & T_0 \end{bmatrix}.$$

The characteristics of the filters φ_Ω can be likewise written in the matrix form Φ_Ω (Eq. (9)). In the base $B = \{b_k\}$ (Eq. (14)) it will be diagonal matrices with values of $(\Phi_\Omega)_{k,k} = \varphi_\Omega(b_k)$ on the main diagonal. For

example, for $\Omega = 2$ and an even N matrix Φ_2 has the following form:

$$\Phi_2 = \text{diag} \left(\underbrace{\left[0, \dots, 0, \overset{-2}{1}, 0, \overset{0}{1}, 0, \overset{2}{1}, 0, \dots, 0 \right]}_V, \underbrace{\left[\overset{-N}{1}, 0, 1, 0, \dots, \overset{-2}{1}, 0, \overset{0}{1}, 0, \overset{2}{1}, \dots, 0, 1, 0, \overset{N}{1} \right]}_T \right). \quad (19)$$

The psychophysical spectra $\mathbf{P}_\Omega^{(n)}$ and δ_0 can be written in the form of one column matrices with a length $4N+2$. In view of the above definitions Eq. (11) takes the following form:

$$\mathbf{P}_\Omega^{(n+1)} = \Phi_\Omega \mathbf{Q}_F \mathbf{P}_\Omega^{(n)} + \delta_0. \quad (20)$$

Because $\Phi_\Omega \delta_0 = \delta_0$ then it follows from the preceding equation that $\mathbf{P}_\Omega^{(n+1)} = \Phi_\Omega \mathbf{P}_\Omega^{(n)}$. The problem of how to determine spectra $\mathbf{P}_\Omega^{(n)}$ can be reduced to subspace $W_\Omega \subset W$ determined by projecting images Φ_Ω . Equation (20) can therefore be given by:

$$\mathbf{P}_\Omega^{(n+1)} = \mathbf{Q}_\Omega \mathbf{P}_\Omega^{(n)} + \delta_0, \quad \mathbf{Q}_\Omega := \Phi_\Omega \mathbf{Q}_F \Phi_\Omega. \quad (21)$$

The above considerations will be illustrated in the next section on two-tone problem example.

5. Two-tone pitch perception

Let us consider physical spectrum of two-tone given by:

$$F = \frac{\theta}{2N+1} [F_{-\omega_2} \langle -\omega_2 \rangle + F_{-\omega_1} \langle -\omega_1 \rangle + F_0 \langle 0 \rangle + F_{\omega_1} \langle \omega_1 \rangle + F_{\omega_2} \langle \omega_2 \rangle], \quad (22)$$

where $\omega_2 > \omega_1 > 0$. In an artificial way an imperceptible 0-th tonal component was introduced into the spectrum as an element of the mechanism of the auditory processor. It is assumed that the amplitude of this component has a constant in time value. Amplitude spectrum has property $F_{-\omega} = F_\omega$ so by substituting Expression (22) into Eq. (5) we have:

$$Q_F = \beta \langle 0 \rangle + \theta [\varepsilon \langle 0 \rangle + A(\langle -\omega_1 \rangle + \langle \omega_1 \rangle) + B(\langle -\omega_2 \rangle + \langle \omega_2 \rangle)], \quad (23)$$

where ε, A, B , are calibrated amplitudes $F_0, F_{\omega_1}, F_{\omega_2}$ by factor $\alpha/(2N+1)$. While having Q_F it is possible to determine matrix \mathbf{Q}_F (17). The aim is to determine partial limit spectrum $\mathbf{P}_\Omega := \lim_{n \rightarrow \infty} \mathbf{P}_\Omega^{(n)}$ of sequence of partial psychophysical spectra of two-tone for individual fundamental frequencies $\Omega \in \Omega_V$. These spectra are determined from the following equation derived from Eq. (21):

$$\mathbf{P}_\Omega = (\mathbf{I} - \mathbf{Q}_\Omega)^{-1} \delta_0, \quad (24)$$

where \mathbf{I} is an identity matrix in the base B of the space W .

Let us now consider a tonal filter with the fundamental frequency Ω showed in Fig. 4 and for simplicity let us assume that $2r$ is smaller than auditory frequency resolution so only one component can be located within passband. Depending on the position of two-tone components in relation to the tonal filter the four cases can be differentiated which are showed in Fig. 5. Case 1 is the most interesting one. It can be checked that other cases can be narrowed down to either no perception of tones (Case 2) or to perception of one or two singular tones (Cases 3 and 4).

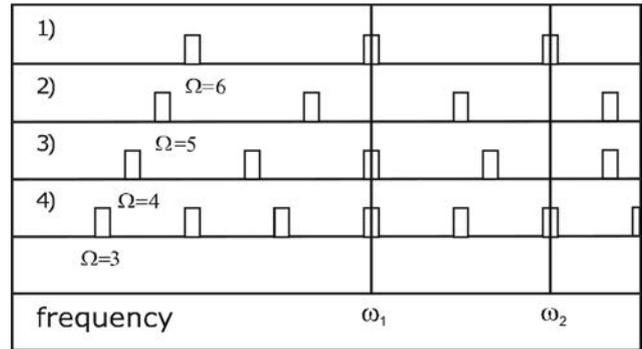


Fig. 5. The possibilities of deploying two-tone in relation to the tonal filter.

Let us now consider Case 1. The psychophysical spectrum determined from Eq. (24) can be presented in the following form:

$$\mathbf{P}_\Omega = [V_{-N}, V_{1-N}, \dots, V_N, T_{-N}, T_{1-N}, \dots, T_N]^T. \quad (25)$$

Let us write non-zero components of this spectrum as functions of amplitudes ε, A and B :

$$\begin{aligned} V_0 &= (1 - \beta)S, & V_{\omega_2 - \omega_1} &= (1 - \beta)R, \\ T_0 &= \varepsilon S, & T_{\omega_2 - \omega_1} &= \varepsilon R, \\ T_{2\omega_1 - \omega_2} &= AR, & T_{\omega_1} &= AS + BR, \\ T_{\omega_2} &= BS + AR, & T_{2\omega_2 - \omega_1} &= BR, \end{aligned} \quad (26)$$

where

$$\begin{aligned} r &= 2AB, \\ s &= (1 - \beta)^2 - \varepsilon^2 - 2A^2 - 2B^2, \\ R &= \frac{r}{s^2 - 2r^2}, \\ S &= \frac{s}{s^2 - 2r^2}. \end{aligned} \quad (27)$$

The spectrum is also presented in Fig. 6. It can be observed that the spectrum contains virtual pitch $\omega_2 - \omega_1$, and in the tonal range apart from fundamental components it also contains zero component and combination tones with the frequencies of $\omega_2 - \omega_1, 2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$.

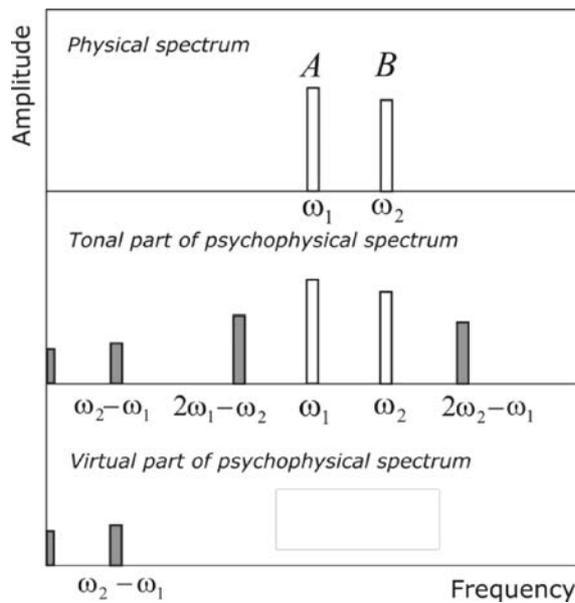


Fig. 6. Partial psychophysical spectra of two-tone – the Case 1.

Let us now go over the mechanism of generating psychophysical image of two-tone in the model constructed above. This one can enrich psychophysical spectrum with components which frequencies are linear combination of the frequencies of physical spectra components with low integers. In order to better illustrate this issue the following diagrams will be used:

$$p_1 \xrightarrow{f_1} p_2 \xrightarrow{f_2} p_3 \cdots p_{n-1} \xrightarrow{f_{n-1}} p_n, \quad (28)$$

where $\{p_i\}_1^n, \{f_j\}_1^{n-1}$ are sequences in the group G such that $p_{i+1} = f_i \cdot p_i$. These sequences will describe non zero components of corresponding spectra $\{P^{(i)}\}_1^n, \{F^{(j)}\}_1^{n-1}$ which satisfy Eqs. (4) and (5). For example, two steps of the process generated by one spectral component described in Sec. 2 can be described by the following diagram:

$$\langle 0 \rangle \xrightarrow{\theta\langle\omega\rangle} \theta\langle\omega\rangle \xrightarrow{\theta\langle\omega\rangle} \langle 0 \rangle. \quad (29)$$

In general, there are many paths that lead to generating a certain spectral component of psychophysical spectrum. However, not every path is possible to be proceeded due to the fact that some components are cut off by the auditory filters. Below, there are examples of paths leading to generation of combination tones.

$$\begin{aligned} \langle 0 \rangle &\xrightarrow{\theta\langle\omega_2\rangle} \theta\langle\omega_2\rangle \xrightarrow{\theta\langle\omega_1\rangle} \langle\omega_2-\omega_1\rangle \xrightarrow{\theta} \theta\langle\omega_2-\omega_1\rangle, \\ \langle 0 \rangle &\xrightarrow{\theta\langle\omega_1\rangle} \theta\langle\omega_1\rangle \xrightarrow{\theta\langle\omega_2\rangle} \langle\omega_1-\omega_2\rangle \xrightarrow{\theta\langle-\omega_1\rangle} \theta\langle 2\omega_1-\omega_2 \rangle, \quad (30) \\ \langle 0 \rangle &\xrightarrow{\theta\langle\omega_2\rangle} \theta\langle\omega_2\rangle \xrightarrow{\theta\langle\omega_1\rangle} \langle\omega_2-\omega_1\rangle \xrightarrow{\theta\langle-\omega_2\rangle} \theta\langle 2\omega_2-\omega_1 \rangle. \end{aligned}$$

Let us note, that both virtual and tonal pitches are generated in different steps of perception process: tonal ones in odd steps and virtual ones in even steps so we can say odd-even pitch rather than tonal-virtual one. First path in (30) shows that combination tone $\omega_2 - \omega_1$ in tonal spectrum exists only if the imperceptible 0-th tonal component in spectrum F is introduced (see also Eq. (26): $T_{\omega_2-\omega_1} > 0$ if $\varepsilon > 0$).

Image presented in Fig. 6 is not, however, fully compatible with this one in Fig. 1. The spectrum presented in Fig. 6. lacks higher combination tones, such as $\omega_1 - n(\omega_2 - \omega_1), n > 1$. On the other hand, the spectrum contains the $2\omega_2 - \omega_1$ combination tone, which in practice is not psychophysically observed. As we know from the Introduction, it occurs, however, in the alive ear i.e. in situation when ear cooperates with nervous system. Thus, an explanation of these phenomena requires to take into account the feedback which in our model has been omitted. We return to the first of these problems in Sec. 7.

6. Extension to spectral-time model

As we know from the Introduction, there are two mechanisms in auditory system responsible for pitch extraction: spectral (place) and temporal (time). These correspond to two kinds of pitch perception models: spectral and temporal. There are attempts to create a model which merges these mechanisms into one whole (e.g. MEDDIS, O'MARD, 1997) but they are not entirely satisfactory (see CHEVEIGNE, 2005).

In this section, we show in outline that our spectral model can be in a simple way extended to spectral – time model. Moreover, it turns out that such an extended model allows one to describe pitch perception process by one mechanism in both frequency and time domain.

The spectral model presented in the previous chapters has several limitations the most important of which include the fact that:

1. It is based on the concept of spectral models and inevitably inherits some of their drawbacks described earlier in the Introduction.
2. In virtual pitch perception it does not take into account phenomena connected with shift of components in harmonic complexes, the so-called I and II shift effects
3. It predicts occurrence of combination tone of the type $2\omega_2 - \omega_1$, which is not, however, psychophysically observed.
4. It does not explain such phenomena as tuning, masking or suppression.

Let us also turn our attention to the fact that our considerations are based on the assumption that the bandwidth of filter passbands is smaller than auditory frequency resolution. If we in our model reject this

assumption that within passbands additional components can be appeared what is shown on Fig. 7. Then:

5. There is the problem of localization of pitch of source tones in the situation when two components are placed within a range of the same passband.

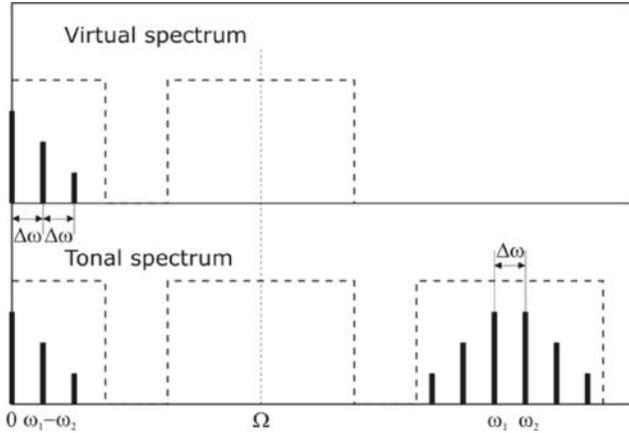


Fig. 7. The psychophysical spectrum in the case when two-tone is placed within a range of the one tonal passband.

As far as virtual pitch perception is concerned the first two problems can be well solved by models based on autocorrelation analysis (LICKLIDER, 1951; MEDDIS, HEWITT, 1991a). We will show now that by moving our construction from frequency domain to time domain we obtain in the range of virtual pitch perception a model consistent with autocorrelation models.

Let us consider the group $G = Z_2R$ with the defining relation $\theta t \theta = t^{-1}$ which in the additive notation is given by $\theta(t)\theta = \langle -t \rangle$ where $t \in R, \theta \in Z_2$. This time instead of applying it for spectra, it will be used for waveforms on a principle similar to that described in Sec. 2. In the time domain the same symbols will be used as in the frequency domain, they will only be spelled with a lower key. The group G determines the psychophysical time domain divided into two parts: the virtual (real time t) and the tonal (time θt). Let us consider the waveform $s(t)$, which by assumption will be appropriately “cut” by means of a window and let us define the physiological waveform $f \in L^1(G)$ in the following form:

$$f(t) \equiv 0, \quad f(\theta t) = s(t). \quad (31)$$

Equation (4) which defines the perception process can be rewritten in the following way:

$$p^{(n+1)} = f * p^{(n)} + \delta_0, \quad (32)$$

where $p^{(n)}$ is the n -th psychophysical waveform. The convolution is given by a formula, similarly to Eq. (13):

$$f * g(x) = \int_{y \in G} f(y)g(y^{-1}x), \quad (33)$$

$$f, g \in L^1(G), \quad x \in G.$$

Let us follow the process of creation of this waveform. We start from $p^{(0)} = \delta_0$. Then, we have $p^{(1)} = f + \delta_0$ and $p^{(2)} = f * f + f + \delta_0$. It can be checked that $f * f$ contains only the virtual part and that:

$$\begin{aligned} f * f(\tau) &= \int_{t \in R} f(\theta(t))f(\langle \theta(t) \rangle^{-1} \langle \tau \rangle) dt \\ &= \int_{t \in R} f(\theta(t))f(\langle -t \rangle \theta(\tau)) dt \\ &= \int_{t \in R} f(\theta(t))f(\theta(t + \tau)) dt \\ &= \int_{t \in R} s(t)s(t + \tau) dt. \end{aligned} \quad (34)$$

We see that the process of generating a psychophysical waveform in its virtual part is therefore in its first steps approximated by means of ACF.

Now, let us return to the beginning of our considerations on the spectral model. Let us note that it has been implicitly assumed that the physiological spectrum which is fed to the input of auditory processor agrees with the physical spectrum of the sound. However physiological spectrum understood as the envelope of vibration of the basilar membrane is a far approximation of the sound spectrum. Moreover, the spectrum reflected in auditory nerve is enriched with additional information connected with the temporal phase synchrony of auditory nerve fibres. Sound information at the level of the auditory nerve can thus be described by a discrete set of amplitudes $\{A_\omega \in R\}$ and a discrete set of phase factors $\{\varphi_t \in [0, 2\pi]\}$ where $\omega \in S$ where $S \subset R$ is a discrete set of frequencies, and t is related to the dependence $t = 1/\omega$. Extension of our spectral model to the spectral-time model can thus be achieved by replacing the real amplitudes F_ω in Eq. (1) or (7) with complex amplitudes $A_\omega e^{i\varphi_{1/\omega}}$. After substitution in Eq. (7) we have:

$$F = \theta \frac{1}{\lambda} \sum_{\omega \in S} A_\omega e^{i\varphi_{1/\omega}} \delta_\omega. \quad (35)$$

We can pass to the time domain by means of formula $f(t) := F(1/t)$. Finally, in our extended model, the filtering described in Sec. 3 should be taken into account. Summing up the psychophysical waveforms for all auditory filters we get as regards virtual pitch the procedure similar to SACF proposed by MEDDIS and HEWITT (1991a).

Concluding we see that virtual pitch can be simultaneously estimated by the spectral and the temporal mechanism. Previously presented spectral model can only estimate the band in which virtual pitch may occur. Temporal mechanism, however, allows one to more

precisely estimate the location of virtual pitch within the band on principle of extraction of the maximum of SACF in the iterative process. Let us add that processing between bands also occurs in the extended model. In the time domain such a processing can be described by the cross-correlation function of signals from two different bands-time windows.

The time extension of the spectral model outlined above does not solve all the problems, in particular the third and the fourth problem and also does not completely solve the fifth problem. However, these problems can be solved by taking into account the feedback between the central and the peripheral auditory system and by a proper choice of the group G . The latest is connected with the problem of encoding sound information in the auditory nerve fibres. Such an encoding can be described by a stochastic process in the frequency-amplitude domain. In our model, the introduction of the additional amplitude dimension requires an “extension” of the group G . Good results can be achieved by using the group $SL(2,R)$. A full consideration of this issue is beyond the scope of this paper. The model based on this group will be the subject of a separate publication.

7. Discussion

An explanation of some “anomalies” of combination tones

The presented model of auditory processor can in an easy way explain the occurrence of combination tones in limited ranges of the frequency quotient of harmonic components of two-tone (approximately 1.1 – 1.5). Combination tones occur when tones ω_1, ω_2 fall into adjacent passbands of a given tonal filter, i.e. in a neighbourhood of frequency quotient $(k+1)/k$ (see Fig. 8). For $k > 1$ these quotients are to be found in

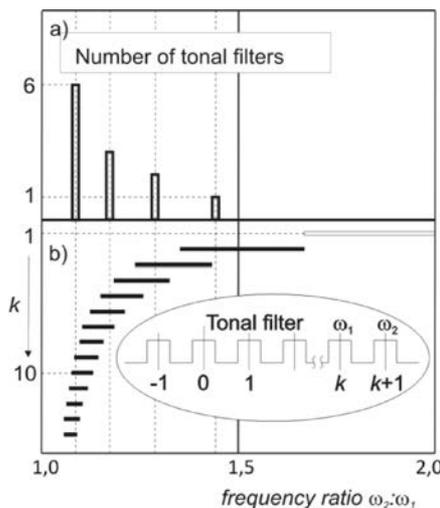


Fig. 8. Pictorial view of participation of tonal filters in creation of combination tone of type $2\omega_1 - \omega_2$.

the range from 11:10 to 3:2 (for $k = 1$ the ratio is 2:1, for which the combination tone would have to occur around the inaudible zero frequency).

Next problem arises when the amplitude of the tone $2\omega_1 - \omega_2$ increases, when frequencies of the tones ω_1, ω_2 are getting closer to each other. If, however, the frequency ratio of the tones $\omega_1 : \omega_2$ is close to unity, the number of auditory tonal filters which can encompass the tones ω_1, ω_2 in adjacent bandwidths increase (see Fig. 8). At the same time the combination tone has a greater intensity, because it is generated by a larger number of filters.

Another problem arises when the amplitude of combination tone $2\omega_1 - \omega_2$ increases depending on the level of stimulus. In the classical interpretation of Helmholtz this tone should increase in a logarithmic scale 3 times faster than source tones at least in the range of low level, which in reality it is not observed. The presented model predicts that the level of combination tone $2\omega_1 - \omega_2$ increases more or less equal to the level of primary tones, what is convergent with empirical data (see Fig. 9).

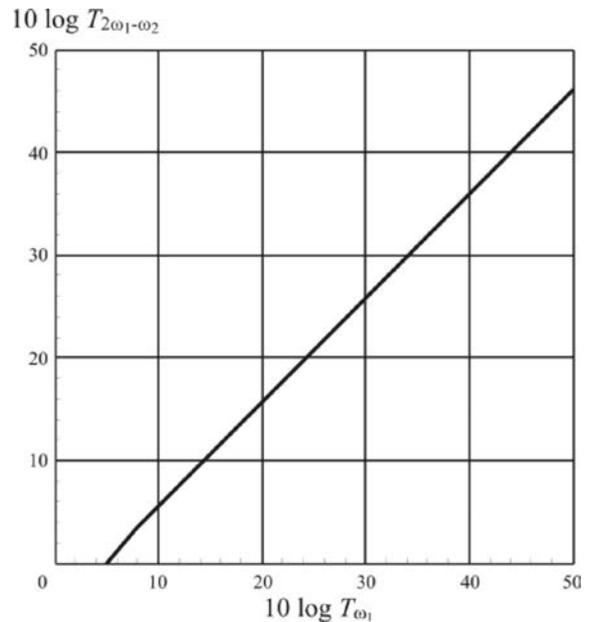


Fig. 9. The level of combination tone $2\omega_2 - \omega_1$ as a function of the level of tone ω_1 ($\beta = 0.4, \varepsilon = 0.4, \omega_2 : \omega_1 = 4 : 3, A = B$).

Tonal versus virtual pitches

The proposed model, according to the description presented in the preceding sections, assumes that the auditory processor has two groups of registers, one of which remembers tonal spectra and the other virtual ones for various sizes of auditory filters (see Fig. 10). The effective pitch sensation as a psychophysical spectrum is the sum of spectra from all tonal and virtual registers. However, as what was mentioned in the In-

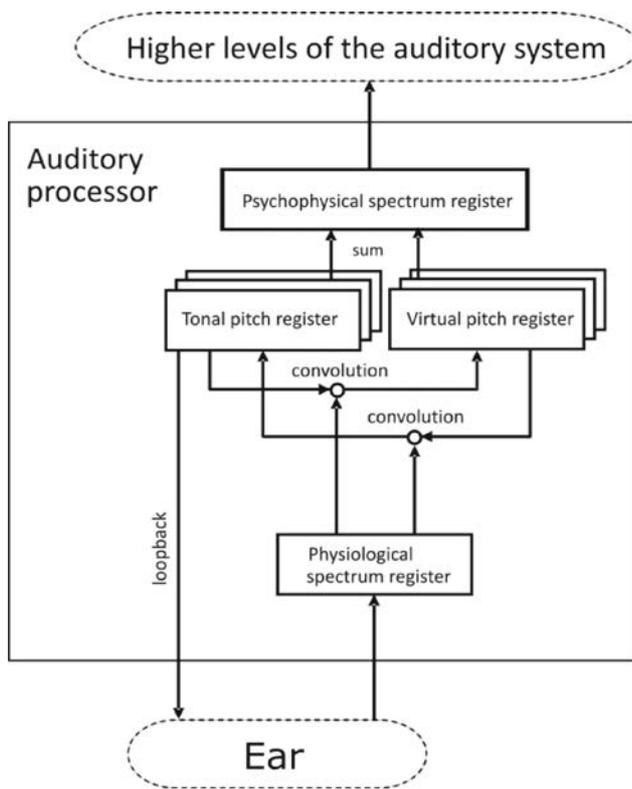


Fig. 10. Schematic diagram of the auditory processor.

roduction, there are large differences in perception of virtual and tonal pitches. In the present model this is explained by the feedback between the auditory processor and the peripheral system. It is assumed that virtual registers – contrary to tonal ones - do not participate in the feedback. Therefore, virtual pitches are not in direct interaction with “real” tones on the peripheral level, owing to which, for example, they are not subject to beat phenomenon with tones of similar frequencies.

The occurrence of higher combination tones, such as $\omega_1 - n(\omega_2 - \omega_1)$, ($n > 1$) can also be explained by feedback of the auditory processor. Let us assume that owing to the feedback the initial spectrum of the two-tone is enriched with the combination tone $2\omega_1 - \omega_2$. By performing similar calculations for three – tone it can be demonstrated that the combination tone $3\omega_1 - 2\omega_2$ is generated on the basis of the rule $2(\cdot_1) - (\cdot_2)$ i.e. $2(2\omega_1 - \omega_2) - \omega_1 = 3\omega_1 - 2\omega_2$. On the same basis further combination tone, such as $\omega_1 - n(\omega_2 - \omega_1)$ are obtained.

Problem of locating the auditory processor

In the proposed model of the central auditory processing of pitch the perception process is described by means of convolution and sound spectra generated by ears. From the mathematical point of view convolution is directly connected with symmetry. Therefore, symmetry described by a group (of symmetry) plays

the main role in the proposed construction. Similar cases are to be found in contemporary physics. Symmetry is the basis of the most fundamental phenomena such as elementary interactions and elementary particles. Symmetry also allows one to simplify some problems in physics e.g. the problem of describing an optical spectrum. If e.g. chemical particles have nontrivial symmetry then on this basis we can in a qualitative way describe the optical spectrum without having to solve differential equations. In analogy to this, in our case symmetry may describe a structure of neural net which is responsible for pitch extraction. One of the difficulties of our analogy is that the physics of the brain is nonlinear contrary to spectroscopy whose physics is linear. In general most of methods based on symmetry do not work on nonlinear systems. We could also put forward a hypothesis that the action of neural net as a macroscopic structure is connected with a linear quantum structure with an inner symmetry on microscopic level. Such a structure could perform signal processing according to the presented description. There is one argument in favour of this hypothesis which is also a conclusion of the proposed model, namely, the structure responsible for signal processing on the central level must be richer and more complex than the neural structure of the brain.

References

1. ASHMORE J. (2008), *Cochlear outer hair cell motility*, *Physiol. Rev.*, **88**, 173–210.
2. BOER E. (1984), *Auditory physics. Physical principles in hearing theory II*, *Physics Reports*, **105**, 3, 141–226.
3. BOER E. (1991), *Auditory physics. Physical principles in hearing theory III*, *Physics Reports*, **203**, 3, 125–231.
4. CHEVEIGNE A. (2005), *Pitch perception models*, [in:] *Pitch: Neural Coding And Perception*, PLACK C.J., OXENHAM A.J., FAY R.R., POPPER A.N. [Eds.], pp. 169–233, Springer Verlag, New York.
5. EGUILUZ V.M., OSPECK M., CHOE Y., HUDSPETH A.J., MAGNASCO M.O. (2000), *Essential nonlinearities in hearing*, *Phys. Rev. Lett.*, **84**, 5232–5235.
6. GOLDSTEIN J.L. (1967), *Auditory nonlinearity*, *J. Acoust. Soc. Am.*, **41**, 676–689.
7. GOLDSTEIN J.L. (1973), *An optimum processor theory for the central formation of the pitch of complex tones*, *J. Acoust. Soc. Am.*, **54**, 1496–1516.
8. GREENWOOD D.D. (1971), *Aural combination tones and auditory masking*, *J. Acoust. Soc. Am.*, **50**, 502–543.
9. HELLE R. (1969/70), *Amplitude and phase of third order auditory difference tone* [in German: *Amplitude und Phase des im Gehör gebildeten Differenztones dritter Ordnung*], *Acustica*, **22**, 74–87.
10. HELMHOLTZ H. (1856), *On the combination tones* [in German: *Über Combinationstöne*], *Ann. Phys. Chem.*, **99**, 497–540.

11. HELMHOLTZ H. (1863), *On the sensation of tone as a physiological basis for the theory of music*, 1st English ed. 1897, Paperback ed. 1954, Dover, New York; Chap.VII, App.XII, [in German: *Die Lehre von Tonempfindungen als physiologische Grundlage für die Theorie der Music*], F.Vieweg, Braunschweig.
12. HOUTSMA A.J.M. (2007), *Experiments on pitch perception: implications for music and other processes*, Archives of Acoustics, **32**, 3, 475–490.
13. HOUTSMA A.J.M., GOLDSTEIN J.L. (1972), *The central origin of the pitch of pure tones: evidence from musical interval recognition*, J. Acoust. Soc. Am., **51**, 520–529.
14. LICKLIDER J.C.R. (1951), *A duplex theory of pitch perception*, Experientia, **7**, 128–133.
15. LOHRI A., CARRAL S., CHATZIOANNOU V. (2011), *Combination tones in violins*, Archives of Acoustics, **36**, 4, 727–740.
16. MEDDIS R., HEWITT M. (1991a), *Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery I. Pitch identification*, J. Acoust. Soc. Am., **89**, 2866–2882.
17. MEDDIS R., HEWITT M. (1991b), *Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery II. Phase sensitivity*, J. Acoust. Soc. Am., **89**, 2883–2894.
18. MEDDIS R., O'MARD L. (1997), *A unitary model of pitch perception*, J. Acoust. Soc. Am., **102**, 1811–1820.
19. MICHALSKI W., BOCHNIA M., DZIEWISZEK W. (2011), *Simultaneous measurement of the DPOAE signal amplitude and phase changes*, Archives of Acoustics, **36**, 3, 499–508.
20. MOORE B.C.J. (1997), *Introduction to psychology of hearing*, 4th edition, Academic Press, London.
21. OHM G.S. (1843), *On the definition of a tone with the associated theory of the siren* [in German], Poggendorf's Annalen der Physik und Chemie, **59**, 497ff.
22. PLOMP R. (1964), *Rate of decay of auditory system*, J. Acoust. Soc. Am., **36**, 277–282.
23. PLOMP R. (1965), *Detectability threshold for combination tones*, J. Acoust. Soc. Am., **37**, 1110–1123.
24. PLOMP R. (1976), *Aspects of tone sensation*, Academic Press, London.
25. ROBLES L., RUGGERO M.A., RICH N.C. (1997), *Two-tone distortion on the basilar membrane of the chin-chilla cochlea*, J. Neurophysiol, **77**, 5, 2385–2399.
26. RUGGERO M.A., RICH N.C. (1991), *Furosemide alters organ of Corti mechanism: Evidence for feedback of outer hair cells upon the basilar membrane*, J. Neurosci., **11**, 1057–1067.
27. SMOORENBURG G.F. (1972a), *Audibility region of combination tones*, J. Acoust. Soc. Am., **52**, 603–614.
28. SMOORENBURG G.F. (1972b), *Combination tones and their origin*, J. Acoust. Soc. Am., **52**, 615–632.
29. TERHARDT E. (1974), *Pitch, consonance, and harmony*, J. Acoust. Soc. Am., **55**, 1061–1069.
30. VAN DIJK P., MANLEY G.A. (2001), *Distortion product otoacoustic emissions in the tree frog Hyla cinerea*, Hearing Research, **153**, 14–22.
31. WIGHTMAN F.L. (1973), *The pattern-transformation model of pitch*, J. Acoust. Soc. Am., **54**, 407–416.
32. ZWICKER E. (1955), *Abnormal amplitude characteristics of nonlinear distortions in the ear* [in German: *Der ungewöhnliche Amplitudengang der nichtlinearen Verzerrungen des Ohres*], Acustica, **5**, 67–74.
33. ZWICKER E. (1968), *The cubic difference tone and the auditory excitation* [in German: *Der kubische Differenzton und die Erregung des Gehörs*], Acustica, **20**, 206–209.
34. ZWICKER E., FASTL H. (1973), *Cubic difference sounds measured by threshold- and compensation-method*, Acustica, **29**, 336–343.