Acoustic and Biomedical Engineering

The Use of the Phase Locking Information in the Human Auditory System for Frequencies Above 5 kHz

A. Sęk* and M. Bukała
Institute of Acoustics, Faculty of Physics, Adam Mickiewicz University
Umultowska 85, 61-614 Poznań, Poland

Accurate allocation of neural impulses to the same phase (i.e. phase locking) in the auditory system, ceases for frequencies above 5 kHz. However, some recent works suggest that it may have a much higher value. A discrimination of harmonic complex and inharmonic complex sound, formed using sound harmonic complex, in which all components were shifted towards higher frequency by the same number in hertzs, was analyzed. Fundamental frequency was $F_0 = 1$ kHz and signals were bandpass filtered by a fixed filter center frequency of which was $11F_0$ and bandwidth $5F_0$. Discrimination threshold was $\Delta F = 0.089F_0$ for 10 normal-hearing subjects. However, replacing the sinusoidal components with the noise bands brought about a significant increase in thresholds. The largest increase was observed for 700 Hz bandwidth. The replacement of sinusoidal components with noisebands reduces information conveyed by phase locking. The differences in excitation pattern for harmonic complex and inharmonic complex signals, for the average threshold, did not exceed 0.7 dB. Therefore they could not be a useful cue for harmonic complex and inharmonic complex discrimination. A simplified model of phase locking showed that the substitution of sinusoids with bands of noise significantly reduced number of intervals between successive neural spikes corresponding to the virtual pitch of harmonic complex and inharmonic complex sounds. This degradation of discrimination suggests that the main source of information about the pitch of harmonic complex and inharmonic complex, especially for sinusoidal components, was the phase locking.

DOI: 10.12693/APhysPolA.123.1106
PACS: 43.66.Fe, 43.66.Hg

1. Introduction

Perception of pitch or frequency discrimination is usually considered in the light of the two main theories: the place theory and the time theory [1]. The properties of the basilar membrane in cochlea are the basis of the place theory. The membrane behaves similarly to a frequency analyzer and therefore it is usually considered as a system of overlapping bandpass filters called the auditory filters [2, 3]. In the case of simulation with a single sinusoid, a single maximum of amplitude envelope of the basilar membrane vibration is observed. This is the so-called excitation pattern and its maximum location, i.e. the distance the from the oval window, strictly depends on the frequency of the sinusoid. The position (place) of that maximum also determines the pitch of the perceived sound. The assignment of each frequency to a very specific and unique place on the basilar membrane is the basis for the so-called tonotopic organization. When the auditory system is stimulated with a sound composed of many sinusoids (e.g. a harmonic complex sound), many maxima of displacement can be observed on the basilar membrane and many different pitches can be assigned to the sound. It is only possible when the distance between individual components of a complex sound are not smaller than the bandwidth of auditory filters centered at frequencies of these components. In this case, each individual component is assigned to a separate auditory filter. It is said that in such a situation the components were resolved in the auditory system. Then, the signal at the output of each auditory filter, stimulated with a single sinusoid, is also close to a sinusoidal signal. However, in the case of a harmonic complex sound, the distances between adjacent components are the same and equal to the fundamental frequency, while the bandwidth of auditory filters increases with the center frequency. Therefore, for a certain number of components (usually above 9), the assignment of individual auditory filter for each component is not possible. Starting from component number 10–11, more than one component are input signals to a single auditory filter, leading to a complex waveform at its output. Such components are called the unresolved ones [1].

If the input waveform to a single auditory filter consists of several sinusoids with different frequencies, the output signal from the filter is characterized by a slowly varying amplitude envelope superimposed on a much more rapid changes in the instantaneous values. The changes in the instantaneous values occurring at a frequency close to the center frequency of the filter. These rapid changes in the instantaneous values are called the temporal fine structure (TFS). The information related to the amplitude envelope is represented at the level of the auditory nerve in neural firing rate (number of spikes per second). However, the information about the values of instantaneous acoustic pressure changes is included in the distribution of neural impulses in time (phase locking), i.e. in the time intervals between successive pulses [4], which is a basic assumption of the time theory [1, 5]. Phase

*corresponding author; e-mail: oleksek@amu.edu.pl

locking, the assignment of neural impulses to always the same phase of the signal, is the best in the frequency range up to about 4–5 kHz and significantly worsens for higher frequencies. It is assumed that the human auditory system does not use the information conveyed by phase locking in a range above 5 kHz. The supporting arguments for this assumption comes mainly from behavioral studies [1]. Experimental facts, such as a significant increase in the Weber fraction for just noticeable differences (JNDs) in frequencies above 5 kHz or the lack of opportunity to properly reproduce the musical intervals in this frequency range, suggest a negligible effect of phase locking on pitch perception. However, some physiological studies by Heinz [6], Recio-Spinoso [7], and Temchina [8] suggest the existence of a "residual" assignment of neural impulses to always the same phase of the signal (the residual phase locking) for frequencies above 5 kHz. There is a lot of evidence that the discrimination of sinusoidal signals (tones), especially those with a constant amplitude over time, deteriorates with an increase in frequency above 5 kHz [9]. However, it is difficult to translate the conclusions drawn from these experiments onto complex signals. Ritsma [10–12] noted that the components of complex sounds with frequencies above 5 kHz did not produce pure virtual pitch, while Oxenham et al. [13] have shown that the components of complex tones above 6 kHz produced clearly audible musical pitch and intervals.

The subject area of pitch perception, frequency discrimination, and the possibility to use phase locking information for frequencies above 5 kHz is recently an important issue [13] considered also in the context of speech intelligibility [14–19]. Moore and Sęk [20] analyzed the possibility of using the information conveyed by the temporal fine structure for harmonic complex sounds with high fundamental frequency, $F_0 = 800$ Hz or 1000 Hz. They studied the possibility to distinguish harmonic complex (H) stimuli from the inharmonic complexes (I). Inharmonic signals (I) have been constructed on the basis of harmonic signals (H) by shifting all components by the same number in hertzs towards higher frequencies. All stimuli were filtered by a fixed bandpass filter with a center frequency of $14F_0$, bandwidth of $5F_0$, and symmetric slopes of 30 dB/octave. Two observation intervals were presented in random order, contained stimuli HHHH or HIHI, and the task of the listener was to identify the interval in which the change in pitch corresponding to the HIHI interval was present.

Due to the bandpass filtering, spectral envelopes of the H and I signals were extremely close to each other and prevented the discrimination of H and I sounds based on the differences in the excitation patterns produced by these signals. Additionally, they had a very low level of 20 dB SL (for $14F_0$) and were presented at a background of the threshold equalizing noise (TEN, [21]) with a 15 or 10 dB below the signal (H and I) level. It ruled out potential detection cues connected with nonlinear distortions while multitone presentation, as suggested by Oxenham et al. [13]. The difference between adjacent frequency components of sound H and I was always equal to F_0, and the initial phase of the components was random. Therefore, the amplitude envelopes had the same repetition rate and were completely independent for each of the H and I tones within HHHH and HIHI. However, their long-term average spectra (LTAS) were very similar and as such, the comparison of the spectra did not provide an effective source of discrimination of sound H and I. The results of the experiments showed that the vast majority of the subjects were able to discriminate between H and I sound for both $F_0 = 800$ and $F_0 = 1000$ Hz with a center frequency of the bandpass filter of 11.2 kHz and 14 kHz, respectively, that implies the lowest component within the filter, equal to 9.6 kHz and 12 kHz, respectively.

Moore and Sęk [20] analyzed output signal from the auditory filter at a center frequency of 9.6 kHz for both H and I signal. They found that the distance between the peaks of adjacent instantaneous extremes within adjacent maxima in the amplitude envelope of these signals were significantly different. Since the probability of generating a neural spike is a monotonic function of amplitude, the highest probability of the spike generation coincides with peaks in instantaneous values located in amplitude envelope maxima. Moore and Sęk [20] concluded that a significant set of time intervals, and theirs average values, provided very effective information about the frequency of H and I sounds, and therefore led to the discrimination of the sounds.

Another factor that could allow for H and I sound discrimination could be the difference in the excitation patterns produced by these sounds. Therefore, Moore and Sęk [20] examined the differences in excitation patterns produced by H and I sounds for the averaged threshold ΔF (= 143 Hz, for $F_0 = 800$ Hz). They found that the largest difference did not exceed 0.6 dB and occurred in the band of 8–9 kHz which is closest to the lowest component within the bandwidth of the fixed bandpass filter. An additional experiment on the discrimination of intensity of this component showed that the mechanism of the difference in excitation pattern did not play an important role in the discrimination.

The series of experiments presented above undoubtedly showed that the H and I sounds discrimination could not be based on the changes in the signal spectrum and changes in the excitation pattern, discrimination of potentially resolved lowest component, or combinational tones (too low signal level and masking noise). Moore and Sęk [20] therefore concluded that the only source of information about the frequency of the signal (as it changes when H and I tones were switched) was the information about the time distribution of extreme instantaneous values (peaks) occurring within maxima of amplitude envelope and time distribution of neural pulses correlated with them. This means that the neural impulses synchronization (phase locking) with a specific phase of an acoustic wave may reach much higher frequencies, about 10 kHz or more, than the previously assumed 5 kHz [1, 5, 24, 25]. This statement is quite controversial because of the well-established knowledge on the physiology of hearing. However, it is worth to perform a series of tests which could also lead to the verification of Moore and Sęk thesis and provide some further arguments for higher frequency limit of the phase locking, especially in a harmonic complex tone pitch discrimination task. This paper is concerned with this issue.

2. Aim of the study

The above series of experiments by Moore and Sęk [20] was based on sinusoidal signals which are determined signals with a properly defined analytical form. The time structure of such signals, despite the random starting phases, is periodic, which could consequently lead to the generation of neural impulses fully correlated with the extremes of the instantaneous values, even for high frequencies. Therefore, a question can be asked whether the disruption of periodic temporal fine structure of H and I sounds (in which the envelope is still periodic) can lead to a significant degradation of their discrimination. If so, this could be another confirmation of Moore and Sęk's claims on a much higher limit of the occurrence or possibility of using the information contained in the phase locking.

Significant reduction of both a periodicity of the temporal fine structure and the amplitude envelope of signals used by Moore and Sęk [20] can be obtained by replacing the sinusoidal components with a non-overlapping bands of noise at center frequencies corresponding to the frequencies of sinusoidal components. Therefore, the research conducted in this study was almost identical to this presented by Moore and Sęk [20]. In the current study, the fundamental frequency F_0 (center frequency of the 'fundamental' band of noise) for the H sounds was equal to 1000 Hz and the center frequency of the fixed passband filter, with a $5F_0$ bandwidth, was constant and amounted to $11F_0$. This filter was characterized by a 30-dB slope per octave outside of the passband. The lowest component within the filter passband had a frequency of 9 kHz. Sounds I were constructed on the basis of the H sound in which the frequencies, or the center frequencies of all noiseband, were shifted upwards by the same number in hertzs, ΔF. The same fixed bandpass filter was used to H and I signals.

The first experiment in this study was a repetition of the Moore and Sęk's investigations using sinusoidal signals. This allowed to verify the previous results and collect a set of discrimination thresholds (ΔF) of H and I sounds which were treated as a sort of reference data. The thresholds were gathered for a group of 10 normally hearing subjects, aged under 27 years. After these initial studies, sinusoidal signals were replaced by bands of noise at a constant relative bandwidth of 3% of the center frequency (of each band) or bands of constant absolute bandwidth of 700 Hz. Such a bandwidth selection helped to avoid an overlapping of the noisebands, even for the highest center frequencies.

3. Materials and method

3.1. Method

The study was based on a modified computer program and method presented by Sęk and Moore [26]. The modification allowed replacing the sinusoidal signals with noise bands. In accordance with the 2AFC procedure, the subjects were presented two observation intervals in a random order. One of them contained the HHHH signals and the second one contained the HIHI signals. Each of the H and I tones lasted 200 ms including 20 ms of rise and fall time, in accordance to the cosine square function, and was separated from the adjacent sound (H or I) with a 100 ms interval of silence. The time between observation intervals was 300 ms. The overall level of signals was equal to 20 dB SL, so that it minimized the risk of combinational tones. In addition, signals were presented against a background noise, the threshold equalizing noise (TEN, [21]), at the level of 15 dB below the total signal level. This noise, with rise and decay times of 20 ms (cosine square function), began 300 ms before the first signal and ended 300 ms after the second interval.

The use of high frequency sound is often associated with significant changes in loudness resulting from fluctuations of headphones frequency characteristics. Therefore, in the studies of this type stimuli, level randomization is often used. In this study, in addition to a constant level for each stimulus (20 dB SL), their randomized levels were also used (20 dB SL ± 3 dB) independently for each H and I sound in HHHH and HIHI series.

In this experiment, the task of the subject was to identify the interval which included changes concerning the presentation of the HIHI series. The experiment began with a high value of ΔF (usually equal to half of the fundamental frequency), so as to allow the subjects to discriminate the stimuli presented in the initial stage of each experimental run. Two successive correct responses brought about a reduction of the frequency difference ΔF and one incorrect answer brought about an increase of the difference. The single measurement determined eight turnpoints. The changes in the frequency difference ΔF was multiplied (or divided) by 1.25^3 until the first turnpoint, then by 1.25^2 to a second turnpoint, and by 1.25^1 for the remaining six turnpoints on which the threshold was calculated. The threshold values were obtained on the basis of five independent measurements. If the subject within a single measurement of threshold discrimination demanded, under the adaptive procedure, a difference ΔF greater than 500 Hz (half of the fundamental frequency), the adaptive method of selecting the frequency difference was abandoned and testing was continued using the method of constant stimuli. The subjects were presented with 40 pairs of HHHH and HIHI signals with a constant frequency shift, $\Delta F = 500$ Hz, and the number of correct answers was recorded. It was then used to estimate the discrimination threshold using the following extrapolation procedure. The 2AFC procedure was used with an adaptive procedure (2d/1u). In this method the threshold corresponds to a probability of the correct answer of 71% which, in turn, corresponds to the value of the detectability $d' = 0.78$ [27]. It was assumed that the experimental detectability d' is proportional to the frequency shift ΔF. Thus such value of ΔF was established for which d' would amount to 0.78. For example, if the probability of a correct answer of a subject in the experiment was $p = 70\%$, then the corresponding detectability would amount to $d' = 0.74$. Therefore, the estimated value of ΔF is equal to $(0.78/0.74) \times 0.5 F_0$. This extrapolation, however, was used only several times.

3.2. Subjects

Ten subjects (aged up to 27) took part in experiments. All of them were volunteers and they have normal hearing in a range up to 14 kHz, which was analyzed by Grason-Stadler GSI61 audiometer. Because for most of them it was the first psychoacoustic investigation, a series of training sessions were conducted before the experiment for about 4–6 hours to familiarize the subjects with their task and make the test procedure more clear to them.

3.3. Apparatus

The signals were generated using the Sound Blaster Audigy audio card with a 24-bit resolution at 48 kHz sampling rate controlled by a PC with an installed test proposed by Sęk and Moore [26]. They were presented monaurally via Sennheiser HD 580 headphones to the ear which was chosen by the subject. During the study, the subjects were in a soundproof booth equipped with the computers periphery necessary for answering test questions.

4. The use of simplified models

4.2. Simplified model of phase locking

In the present paper a simplified model of phase locking was proposed in order to analyze the distributions of time intervals between the adjacent impulses propagating in the neurons of the auditory nerve. The output signals from the auditory filter at a center frequency of 10 kHz were analyzed, while the auditory system was simulated by H and I sounds without masking noise. The impulse response of the auditory filter was assumed to be the fifth order of the Euler gamma function [28–30]. The impulse response was convolved with the time course of the H an I signals to get the output of the auditory filter. Regardless of the input signal, the output from the auditory filter was a stationary signal that was resampled using sampling rate of 5×10^6 Hz (interpolation using standard MATLAB *interpft* function [31]). This was done to increase the accuracy of determination of intervals between local maxima. Then, this signal was normalized so that its peak values were equal to 1. Neural impulse was assigned to each local maximum higher than 0.65. This lead to a situation in which, on average, within one extreme of amplitude envelope, 3–4 impulses were generated. Finally the distances between the impulses were calculated which fell in the adjacent extremes of amplitude envelope. However, the intervals between adjacent extreme values of the temporal fine structure (corresponding to frequencies close to 10 kHz) were not taken into account. It was done due to the fact that the maximum discharge rate of a single neuron did not exceed a few hundred pulses per second. Analysis of the distribution of distances between the impulses focused mainly on the 500–1500 Hz range which accounts for the repetition rate of the amplitude envelope and the virtual pitch of the applied signals.

4.2. The excitation pattern model

The calculation of the difference in excitation patterns evoked by the H and I signals for the averaged threshold difference ΔF was done using the model described by Moore et al. [32]. The analyzed spectra of signals with a resolution of 2 Hz were obtained using standard MATLAB function (psd [31]) and were treated as an input for the calculation of excitation patterns. Due to the use of a masking (TEN) noise, averaged spectra obtained separately for 50 H and 50 I sounds were used to determine the excitation patterns.

5. Results

5.1 Experimental data

Individual results for all subjects as well as averaged values are presented in a histogram in Fig. 1. For each subject, each of the three bars represents the data gathered for sinusoidal signals (hatched area), noisebands with a bandwidth of 3% (open area), and noisebands with a bandwidth of 700 Hz (filled area) that created complex signals. Preliminary studies carried out for 5 subjects showed that when taking into account five repetitions of each measurement, there was no statistically significant difference for stimuli presentation at a constant (20 dB SL) and randomized level (20 dB SL ±3 dB). Therefore, this essential part of the study results of which are illustrated in Fig. 1, was conducted for a constant level of stimuli, i.e. 20 dB SL.

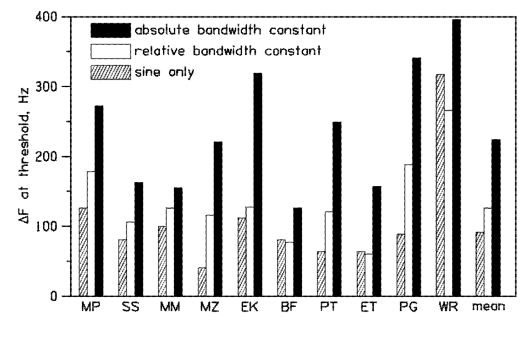

Fig. 1. Discrimination thresholds of H and I signals for 10 subjects, and for different types of components used in H and I sounds: sinusoids (hatched bars), noisebands with a bandwidth of 3% (open bars), and noisebands with a bandwidth of 700 Hz (filled bars).

The results of the frequency discrimination task for high frequency signals are usually characterized by a marked scatter across the subjects participating in the experiment. This is fully confirmed by the gathered thresholds. However, a significant trend in the data appears to be that for all subjects the frequency discrimination thresholds reach the highest value for the H and I signals consisting of the noise of constant absolute bandwidth (700 Hz). Much lower values of these thresholds were obtained for signals composed of noisebands with a constant relative bandwidth (3%). These thresholds, however, were in most of cases higher than the thresholds obtained for sinusoidal components. It is worth to emphasize that the vast majority of the data was collected using the adaptive procedure. The method of constant stimuli and the estimation of the threshold values based on the probability of correct answers was used for subject WR (3% bandwidth) and for the WR and PG subjects for 700 Hz bandwidth. Those were 7 out of 150 thresholds illustrated in Fig. 1.

The gathered data, as illustrated in Fig. 1, were subjected to a within-subject analysis of variance [33] in which the type of signal that makes up the H and I sounds (i.e. sinusoidal signals, 3% and 700 Hz noisebands) was the analyzed factor. This analysis showed that the type of signals used to generate H and I stimuli was highly statistically significant (Snedecor's variable $F = 25.52$, and probability $p < 0.001$) which fully confirms the conclusions drawn previously about the significant differences between the discrimination thresholds for different types of components. Next, a post-hoc Tukey's test (HSD) with respect to the mean values for all subjects was conducted. This analysis showed that the mean thresholds for each type of component signals were pairwise different and these differences were statistically significant ($p < 0.004$). Further Tukey's HSD testing was concerned with the data gathered for individual subjects. It was shown that for all subjects, the discrimination threshold for 700 Hz noisebands and sinusoidal components were statistically significant. Significant differences were also revealed for discrimination threshold measured for two types of noiseband for all subjects. However, statistically significant differences between the discrimination thresholds for complex signals composed of 3% noisebands and discrete sinusoids was found for four subjects only.

Based on the above analysis, it is clear that discrimination of H and I sounds was significantly depend on the type of components of complex signals. The discrimination was the most difficult (highest threshold values) for noisebands of constant bandwidth and the easiest (lowest thresholds) for sinusoidal components. The use of noisebands significantly disrupted the regularity of occurrence of extreme values of instantaneous signal peaks within the maxima in envelope amplitude. Thereby, it disrupted the generation of similar intervals between neural impulses corresponding to those extremes. This made pitch assessment as well as pitch discrimination more difficult, leading to such a significant increase in discrimination thresholds compared to the situation when the components of these sounds were sinusoidal signals.

The main purpose of using sinusoidal components in H and I sounds was to repeat measurements previously done by Moore and Sęk [20]. The result obtained in the present experiment ($\Delta F = 0.089F_0$) fully confirmed the possibility of discrimination of H and I sounds which contained only components above 9 kHz. Moreover, the mean score obtained in this study is lower than the one obtained by Moore and Sęk who reported a threshold $\Delta F = 0.18F_0$. Moore and Sęk observed that the subject with the highest absolute threshold scored the highest discrimination thresholds, too. It should be noted that the absolute thresholds of the subjects participating in the present experiment fell within a 0–20 dB SPL range, whereas in the case of Moore and Sęk this range extended from 8 dB SPL to 42 dB SPL. Taking into account the higher age of listeners in the case of the cited study (24–50 years), it can be assumed that at least some of them suffered from mild cochlear hearing loss associated with aging. A loss of this type can affect the use of the information contained in the temporal fine structure which was clearly showed by Hopkins and Moore [34].

5.2. Analysis of excitation pattern

Regardless of the confirmation of the results presented by Moore and Sęk [20], it was worth to check if for the signals used in this study, the difference of excitation patterns produced by H and I sounds was a potential cue in the discrimination task (the place theory). Zwicker and Fastl [35] suggested that if the difference of excitation patterns caused by two signals in any frequency area is greater than 1 dB, than the signals should be assessed as different. Therefore, in the next step of this study, the calculation of excitation patterns evoked by H and I signals was done, taking into account the averaged discrimination threshold ΔF for each type of components used. These calculations were conducted based on the excitation pattern model proposed by Glasberg and Moore [36]. Because the TEN as a masker was an important part of all stimuli, the excitation patterns in each case was calculated for averaged spectra of 50 signals. The results of these analyses are shown in Fig. 2.

The top line of the figure shows excitation pattern for H (solid line) and I (dashed line) signals, respectively, while the lower row presents the differences between these excitations. Subsequent columns of the figure illustrate the data for all types of components used, i.e. sinusoidal components (SIN), the noisebands of constant relative bandwidth (3%), and the noisebands of constant bandwidth (700 Hz). As it can be seen from the bottom panels of Fig. 2, for all component types the difference in excitation patterns did not exceed 1 dB. This clearly means that the differences in the patterns did not provide sufficient information allowing discrimination of H and I sounds. It is worth noting that the highest difference in calculated excitations (= 0.7 dB) was observed for 3% noiseband at a frequency slightly above 8 kHz. However, differences

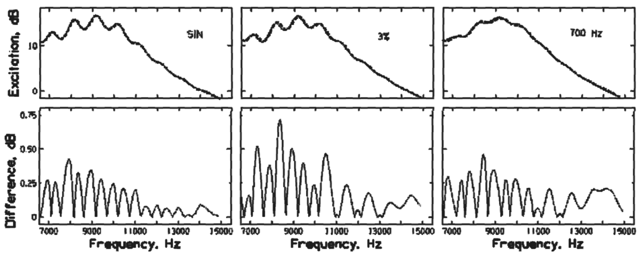

Fig. 2. The top line shows the excitation patterns evoked by the H (solid line) and I (dashed line) signals, while the lower line illustrates the differences between those patterns. Successive columns illustrate the data for different type of signal components: sinusoidal (SIN), noiseband with a bandwidth of (3%) and (700 Hz) respectively. The patterns and their differences were determined for the discrimination threshold averaged across the subjects.

in excitations for H and I sounds composed of sinusoids and 700 Hz noisebands were quite similar and and did not exceed 0.5 dB. Therefore, it is difficult to notice any correlation between differences in excitation patterns and the measured discrimination thresholds. This means that excitation pattern differences provide a kind of information to the auditory system that did not allow, or it is not sufficient, to distinguish between H and I signals. The mutual relationship between excitation pattern differences for each specific type of components of H and I sounds does not reflect the discrimination thresholds for these components. The thresholds measured for the 700 Hz noisebands are significantly larger than for other signals. Therefore, it can be concluded that the differences of these excitation pattern did not play a significant role in the discrimination of H and I sounds.

5.3. Analysis of phase locking

A simplified model of the phase locking was applied to three types of signals components, taking into account both H and I sounds. In the case of I sounds, the frequency shift ΔF, was the average value of the threshold for the appropriate signal type across the subjects. The considerations of this model were limited to the 500–1500 Hz range which represents not only the virtual pitch associated with a missing fundamental component (fundamental band) but also the repetition rate of amplitude envelope of the used sounds. For each type of component, 50 H and 50 I signals with a duration of 200 ms each were analyzed, as was the case during the listening test. Distributions (histograms) of the time intervals between neural impulses obtained in these analyses are shown in Fig. 3. Successive panels of this figure show results for each type of components used, i.e. for sinusoids, 3% and 700 Hz noisebands, respectively. In each part of the figure, the solid line shows a histogram obtained for H and the dotted line for I sound, respectively. To illustrate properly the distribution of registered intervals, the maximum values of the vertical axes are adjusted separately for each component type.

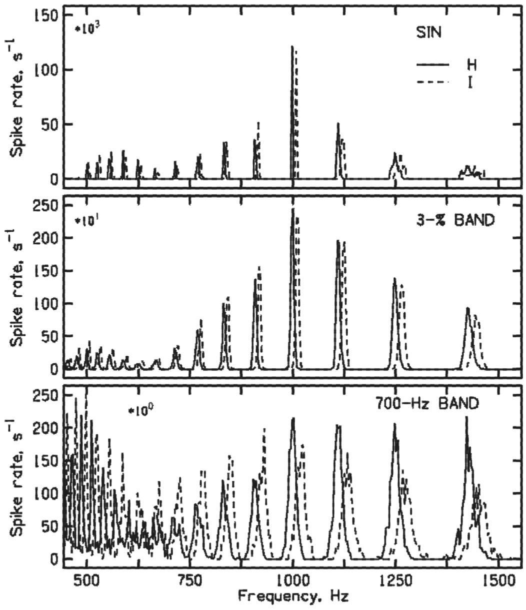

Fig. 3. Histograms of the time intervals between the distributions of neural impulses obtained for the frequency range of 500-1500 Hz for H and I tones. The successive lines of the figure illustrate the data obtained for each type of signal components, such as the sinusoids, 3% and 700 Hz noiseband. These distributions were determined for the discrimination threshold averaged across the subjects. Note, that the vertical axis is scaled in each panel in a different way.

It can easily be seen that the number of time intervals corresponding, or being close, to the frequency of 1 kHz is the largest when the components of the signals H and I are sinusoidal. This number is almost 100 times larger than the number of intervals corresponding to 1 kHz for 3% noisebands. It is also almost 1000 times larger than the number of analogue intervals for the 700 Hz noisebands. The number of intervals corresponding to 1 kHz fully correlate with the obtained discrimination thresholds. In the case of sinusoidal signal, the vast majority of intervals between impulses supplied almost a clear information about the (virtual) pitch of H and I sound so that the discrimination was possible and occurred for a relatively small frequency difference ΔF. A significant difference in the number of intervals corresponding to 1 kHz for 3% noisebands correlates quite well with the result of Tukey's HSD post-hoc test in which there was a statistically significant difference between the discrimination thresholds for these type of signals. The difference in the number of intervals for complex sounds composed of 3% and 700 Hz noisebands also corresponds with the results of Tukey's test. Statistical significance of differences between the discrimination thresholds for these signals were found only in four out of ten subjects.

To sum up this part of the analysis it can be stated that a full correlation between the number of intervals corresponding to frequency close to 1 kHz with the discrimination thresholds was found. This clearly suggests that the time distribution of neural impulses could be the most important factor enabling such a good H and I sounds discrimination, especially in the case when they were composed of sinusoidal signals, and much worse if they were the noisebands.

6. Discussion

This study confirmed earlier reports by Moore and Sęk [20, 26, 37] that the discrimination of harmonic sound (H) consisting of high harmonics (above 9th) and inharmonic sound (I) consisting of the same harmonics shifted towards higher frequency by the same number of hertzs is possible and can be done by most (if not all) of normally hearing listeners. The average result of the present experiment is even lower (i.e. better discrimination was noticed) probably due to the younger group of subjects with lower absolute thresholds. However, the essential aim of the study was to determine whether the phase locking, i.e. precise allocation of neural impulses to the same phase of the stimulus, can provide information allowing discrimination signals frequency components of which exceed 5 kHz. The gathered results seem to fully confirm that the auditory system can effectively use the information conveyed in phase locking evoked by signals at frequencies of 9 kHz and higher.

The generation of the neural representation of acoustic stimuli in neurons of auditory nerve is fairly well recognized [24, 38, 39] and there are numerous of statistical models for simulation tuning of neurons, the rate of its discharge versus level, the suppression effect, etc. A single neuron of the auditory nerve can generate up to several hundred impulses per second [24] which is always caused by a temporary depolarization of the inner hair cell [40]. The rate of those discharges is far from the possibility to replicate frequencies above 9–10 kHz. However, the probability of a sufficient instantaneous depolarization of the inner hair cell, and thus the generation of an impulse in the neuron, is monotonically dependent on the amplitude of the signal [39]. Therefore, it seems that the generation of such an impulse for the extreme instantaneous values in the maximum amplitude envelope is quite probable. So, if the amplitude envelope has a quasi-periodic nature, then impulses in a single neuron can be expected for the extreme instantaneous values of the maxima of amplitude envelope. These impulses do not need to appear in the adjacent extremes of the envelope when its repetition rate is more than a few hundred times per second. However, if a single neuron is observed, then an impulse occurring within it falls in a local maximum of envelope. The discharge of a single neuron does not reflect the signal envelope due to a limited discharge rate. However, an innervation of each point of the basilar membrane is very dense, i.e. there are at least several afferent neurons for each inner hair cell. Therefore, it can be expected that each maximum of amplitude envelope will have its explicit representation in the course of discharge of a group of neurons. Such a situation can be observed when the H and I tones are composed of sinusoidal components. However, if the H and I tones are composed of bands of noise, as some of the sounds used in this study, it is difficult to expect that they may have a periodic or quasi-periodic amplitude envelope. It is also difficult to expect a clear location of the extreme instantaneous values within the envelope maxima. In such a situation, even if the impulses are assigned to the instantaneous peak values, the distances between them in the time domain are approximately random and do not convey explicit information about the frequency (pitch) of the signal. Therefore, the use of noisebands components of H and I sounds resulted in a significant disruption of neural discharges and the regularity of time intervals between them. This in turn brought about an increase in difficulty in assessment of pitch and a marked increase in discrimination thresholds which was observed in the experiment.

The interpretation presented above is confirmed, to a large extent, by a simplified model of phase locking information analysis, i.e. the presented analysis of time intervals between the impulses that can be generated by the extreme instantaneous values appearing in the maxima of amplitude envelope. There was a significant correlation between the number of intervals for a frequency close to the repetition rate of the envelope and the virtual pitch corresponding to H and I sounds with experimentally determined discrimination thresholds. This correlation is an important argument suggesting that, at least for the sinusoidal components in H and I sounds, phase locking has provided very efficient information on the pitch of signals allowing for a good discrimination. Distorting the phase locking information by introducing a narrow and then a very broad (but still non-overlapping) bands of noise resulted in a significant reduction in the number of intervals corresponding to frequencies close to 1 kHz and to the virtual pitch of H and I sounds. The consequence of this reduction was a significant deterioration of H and I signal discrimination and more than twofold increase in the thresholds.

The analysis of the excitation pattern based on the concepts of auditory filters [2, 3, 36] showed that regardless of the type of components used in the H and I sounds, the maximum difference of excitation for an averaged threshold discrimination was less than 1 dB, which is a broadly accepted detection criterion proposed by Zwicker [41]. Excitation pattern differences do not correlate in any way with the set of thresholds. This allows to state that the excitation pattern did not play a crucial role in the discrimination.

To sum up, it can be stated that the discrimination of H and I sounds was made based on the information carried by the phase locking. As the used components had a very high frequency (9 kHz and above), it seems that the auditory system can effectively use the information contained in the phase locking well above the commonly accepted limit of 5 kHz. This can occur especially when the signals are composed of sinusoidal components. This is in line with the suggestions of Moore and Sęk [20] and the concept of residual phase locking [6–8].

7. Conclusions

Results of this study allow to formulate the following conclusions:
1. It is possible to distinguish harmonic complex (H) and inharmonic complex (I) sounds (at a fundamental frequency of 1 kHz) in which high frequency components (9 to 13 kHz) are only present.
2. Replacement of sinusoidal components with 3% or 700 Hz noisebands significantly impairs discrimination of H and I sounds, leading to a much higher discrimination threshold.
3. The discrimination of H and I sounds, regardless of the types of components, is not mediated by the differences in excitation patterns evoked by these sounds. The differences did not exceed the value of 0.7 dB (being usually much less) for the average threshold of discrimination.
4. Number of time intervals between the adjacent neural impulses corresponding to the frequency close to F_0 was a hundred times higher for the H and I signals composed of sinusoids than the signals consisting of noisebands.
5. The presented data demonstrate the possibility of using temporal information contained in the temporal distribution of neural discharges (phase locking) in the frequency range of above 9 kHz.

Acknowledgment

This work was supported by a grant from the National Science Center No. N N 518 502139.

References

[1] B.C.J. Moore, *An Introduction to Psychology of Hearing*, 1 ed., PWN, Poznań 1999 (in Polish).

[2] R.D. Patterson, *J. Acoust. Soc. Am.* **55**, 802 (1974).

[3] R.D. Patterson, *J. Acoust. Soc. Am.* **59**, 640 (1976).

[4] A.R. Palmer, I.J. Russell, *Hear. Res.* **24**, 1 (1986).

[5] W.M. Hartmann, *Signals, Sound, and Sensation*, 1 ed., American Institute of Physics, Woodbury (NY) 1997, p. 298.

[6] M.G. Heinz, H.S. Colburn, L.H. Carney, *J. Acoust. Soc. Am.* **110**, 2065 (2001).

[7] A. Recio-Spinoso, A.N. Temchin, P. van Dijk, Y.H. Fan, M.A. Ruggero, *J. Neurophysiol.* **93**, 3615 (2005).

[8] A.N. Temchin, A. Recio-Spinoso, P. van Dijk, M.A. Ruggero, *J. Neurophysiol* **93**, 3635 (2005).

[9] C.C. Wier, W. Jesteadt, D.M. Green, *J. Acoust. Soc. Am.* **61**, 178 (1977).

[10] R.J. Ritsma, *J. Acoust. Soc. Am.* **34**, 1224, (1962).

[11] R.J. Ritsma, *Int. Audiol.* **2**, 34 (1963).

[12] R.J. Ritsma, *J. Acoust. Soc. Am.* **35**, 1241 (1963).

[13] A.J. Oxenham, C. Micheyl, M.V. Keebler, *J. Acoust. Soc. Am.* **125**, 2189 (2009).

[14] K. Hopkins, B.C.J. Moore, *J. Acoust. Soc. Am.* **125**, 442 (2009).

[15] C. Lorenzi, L. Debruille, S. Garnier, P. Fleuriot, B.C.J. Moore, *J. Acoust. Soc. Am.* **125**, 27 (2009).

[16] A. Sęk, M. Bukała, in: *XIX Conference on Acoustics and Biomedical Ingeneering,* ALNUS, Zakopane, 2012, p. 18.

[17] S. Santurette, T. Dau, *J. Acoust. Soc. Am.* **129**, 282 (2011).

[18] J. Kociński, A. Sęk, P. Libiszewski, *Acta Phys. Pol. A* **119**, 996 (2011).

[19] A. Wicher, E. Ozimek, E. Szymiec, *Acta Phys. Pol. A* **118**, 179 (2010).

[20] B.C.J. Moore, A. Sęk, *J. Acoust. Soc. Am.* **125**, 3186 (2009).

[21] B.C.J. Moore, M. Huss, D.A. Vickers, B.R. Glasberg, J.I. Alcántara, *Brit. J. Audiol.* **34**, 205 (2000).

[22] B.C.J. Moore, B.R. Glasberg, M.J. Shailer, *J. Acoust. Soc. Am.* **75**, 561 (1984).

[23] H.E. Gockel, B.C.J. Moore, R.P. Carlyon, C.J. Plack, *J. Acoust. Soc. Am.* **121**, 373 (2007).

[24] J.O. Pickles, *An Introduction to the Physiology of Hearing*, 2nd ed., Academic Press, London 1988.

[25] A.R. Palmer, in: *Hearing*, Eds. M.P. Haggard, E.F. Evans, Churchill Livingstone, Edinburgh 1987, p. 838.

[26] A. Sęk, B.C.J. Moore, *Int. J.Audiol.* **53**, 58 (2012)1.

[27] M.J. Hacker, R. Ratcliff, *Percept. Psychophys.* **26**, 168 (1979).

[28] T. Dau, B. Kollmeier, A. Kohlrausch, in: *Psychoacoustics, Speech and Hearing Aids*, Ed. B. Kollmeier, World Scientific, Singapore 1996 p. 196.

[29] A. Sęk, B.C. Moore, *J. Acoust Soc Am.* **111**, 2783 (2002).

[30] A. Sęk, B.C. Moore, *J. Acoust Soc Am.* **113**, 2801 (2003).

[31] MathWorks Inc., 2005.

[32] B.C.J. Moore, B.R. Glasberg, T. Baer, *J. Audio Eng. Soc.* **45**, 224 (1997).

[33] P. Lane, N. Galwey, N. Alvey, *Genstat 5. An Introduction*, Clarendon Press, Oxford 1987.

[34] K. Hopkins, B.C.J. Moore, *J. Acoust. Soc. Am.* **122**, 1055 (2007).

[35] E. Zwicker, H. Fastl, *Psychoacoustics — Facts and Models*, 2 ed., Springer-Verlag, Berlin 1999, p. 307.

[36] B.R. Glasberg, B.C.J. Moore, *Hear. Res.* **47**, 103 (1990).

[37] B.C.J. Moore, A. Sęk, J. *Acoust. Soc. Am.* **129**, 3206 (2011).

[38] J.O. Pickles, in: *Frequency Selectivity in Hearing*, Ed. B.C.J. Moore, Academic Press, London 1986, p. 51.

[39] A.R. Palmer, in: *Hearing*, Ed. B.C.J. Moore, Academic Press, San Diego 1995, p. 75.

[40] J.F. Ashmore, *J. of Physiol.* **338**, 323(1987).

[41] E. Zwicker, H. Fastl, *Psychoacoustics – Facts and Models*, Springer-Verlag, Berlin 1990, p. 286.