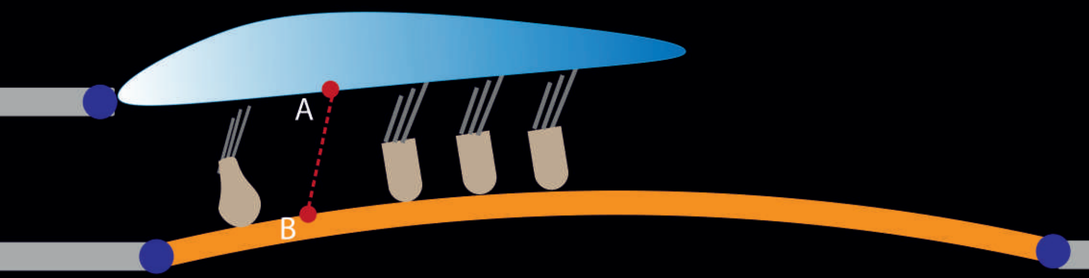
Piotr Kleczkowski

Innovative Material for Studying Auditory Perception



A Set of Videos Explaining Processes in the Organ of Hearing

AGH UNIVERSITY PRESS
KRAKOW 2022

Piotr Kleczkowski

Innovative Material for Studying Auditory Perception

A Set of Videos Explaining Processes in the Organ of Hearing

AGH UNIVERSITY PRESS

KRAKOW 2022

Published by AGH University Press

e-ISBN 978-83-67427-14-2

Head of Publishing of AGH University Press: *Jan Sas*

Reviewers: *prof. dr hab. inż. Andrzej Dobrucki*
prof. dr hab. inż. Anna Snakowska

Author's affiliation:
AGH University of Science and Technology
Faculty of Mechanical Engineering and Robotics
Department of Mechanics and Vibroacoustics

Technical editor: *Agnieszka Rusinek*

Desktop publishing: *Wydawnictwo JAK*

Cover design: *Paweł Sepielak*

AGH University Press (Wydawnictwa AGH)
al. A. Mickiewicza 30, 30-059 Kraków
tel. 12 617 32 28, 12 636 40 38
e-mail: redakcja@wydawnictwoagh.pl
www.wydawnictwo.agh.edu.pl

Contents

Introduction

The idea to develop this innovative learning material stems from the author's experience in teaching a one-semester course on auditory perception for 15 years (first, in Polish; then, in English). The subject area is such that nearly all of the concepts that are taught benefit from some form of graphical presentation. In some cases, static pictures (however elaborate) are unable to convey essential facts; thus, animated graphics are indispensable. This has become clear from the numerous questions and doubts that have been expressed by my students over the years.

The set of videos itself is not a course; it does not contain complete knowledge on auditory perception that is presented in a systematic way; it must be used together with an appropriate course or textbook. This is merely an aid to any course that presents the science of human hearing in a comprehensive manner.

This set of videos is not comprehensive – many more can be designed and produced; it merely covers some topics of an auditory perception course where animated graphics seem to be most beneficial (either for understanding some specific mechanisms and effects that occur in the organ of hearing or illustrating the complex three-dimensional geometry of the structures in this organ). It is meant to serve as a visual aid for the Practical Psychoacoustics course that is held at the Master's level of the Acoustical Engineering major at AGH University of Science and Technology in Krakow, Poland. It can also be used as an educational aid for most other academic courses on human auditory perception (often referred to as psychoacoustics).

There are some fine videos that are available on the internet, but these primarily focus on three-dimensional presentations of the cochlea. Two such videos that the author considers to be particularly valuable are referenced in this material.

This learning material in this set consists of videos as well as comments. Each video has its own comments; however, the comments together do not provide complete and systematic knowledge. On the other hand, knowledge from an appropriate course or textbook is necessary for comprehending most of the comments. In some of the comments, basic facts are briefly reviewed.

The author has taken care that the visual material reflects the contemporary science on human hearing; however, the underlying science is often far from being complete. The material is meant to demonstrate and explain the character of selected processes that are involved in human hearing in a purely qualitative and simplified way. In particular, the material is not meant to convey any quantitative information; therefore, no physical units are provided in the pictures (with a few exceptions).

Technically, the videos were prepared with the use of Adobe Illustrator, Adobe After Effects, Adobe Premiere Pro, Adobe Media Encoder, and Matlab.

This work was supported by the following program:

Zintegrowany Program Rozwoju Akademii Górniczo-Hutniczej w Krakowie, Projekt współfinansowany ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Wiedza Edukacja Rozwój 2014-2020, Nr umowy: POWR.03.05.00-00-Z307/17, nr zadania 1, pozycja budżetowa nr 458.

Piotr Kleczkowski, April 2022.

Comments on videos

Video 1. Sensory receptors and neural pathways in human body

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video1.mp4

Afferent neurons carry information from the sensory receptors in the various organs (in this example, from the skin to the central nervous system; i.e., the brain and spinal cord), whereas efferent neurons carry motor information away from the central nervous system to the muscles and glands of the body.

In the video, a sensory receptor in the hand generates a nerve impulse (presented as a red dot), which is transmitted to the brain through an afferent neural pathway. Then, the information is processed in the brain, and a nerve impulse (the red dot) is sent to the muscles in the hand through the efferent neural fibers. The effect of this neural signal is an action: quickly moving the hand away from an object.

Video 2. Inner structure of cochlea

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video2.mp4

A simplified model of an unwound cochlea; it is similar to a tube that is longitudinally divided into three spaces by two membranes: the thick basilar membrane (orange), and the thin Reissner's membrane (brown). Thus, three canals are formed – also called ducts or *scalae* (of Latin origin – singular: *scala*). From top to bottom: the vestibular canal, the cochlear canal (scala media), and the tympanic canal. The cochlear canal is closed near the apex, so it has no connection with the others. The vestibular and tympanic canals are connected at the far end of the cochlea (on the right in the video); this site is called the apex of the cochlea. The connection is clearly marked by the red arrow that appears nearer to the end of the video. The vestibular and tympanic canals are filled with a fluid that is called perylimph, and the cochlear canal is filled with endolymph. The dotted gray element on the left of the basilar membrane is a model of a flat part of the bone called the modiolus. The dotted gray element on its right is a model of the protrusion of the temporal bone that encapsulates the entire cochlea.

The frontal end of the tube appears to be open in the video; this is intended to make the interior visible. In reality, this end is encapsulated by a bone. There are two openings in the bone (this is not shown in Video 2), and both are covered with elastic membranes. It is through one of these openings where the mechanical vibrations of the stapes are transmitted to the perilymph that fills the vestibular canal. This action is explained in Video 3 (in a simplified form).

Video 3. Pressure along basilar membrane transmitted through perylimph

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video3.mp4

The movement of the stapes footplate toward the interior of the cochlea (inward) exerts positive pressure on the perilymph in the cochlea. The hydromechanical effects are complex, but a simple model assumes that the pressure is evenly distributed along the length of the basilar membrane; so, the membrane is deflected downward. Video 3 illustrates the even distribution of the pressure along the basilar membrane in response to the movement of the stapes footplate.

This simple model is justified by the fact that the speed of sound in water is about 1480 m/s. Considering that the length of the basilar membrane is slightly more than 30 mm, one can conclude that less than half of a wavelength is formed along the basilar membrane even at the highest audible frequency of 20 kHz, thus justifying the simple model at all except the highest frequencies. When the stapes pulls back (i.e., outward), the process is reversed, and the basilar membrane deflects upward. In other words, each cycle of a stimulus evokes a complete cycle of the up-and-down movement of the basilar membrane. The model in the video shows the downward and upward movements of the entire basilar membrane. In reality, this is impossible because of the stiff constraints of the basilar membrane along its edges (as can be seen in Video 3). A more realistic deflection is shown in the basilar membrane transverse section in Video 15.

Video 4. Spatial rotation of basilar membrane

▶▶ https://winntbg.bg.agh.edu.pl/skrypty4/0603/video4.mp4

A simplified model of an unwound basilar membrane is shown. The true basilar membrane has a higher ratio of its length to its width; however, such a view would make further insights more difficult to follow. This model is shown rotating from different perspectives in space, demonstrating its three-dimensional shape. At the end of the movement, the basilar membrane is shown in its unwound side view. This presentation is made under a simplifying assumption that the regular position of the basilar membrane is horizontal; i.e., when treated as a flat object, its surface position is horizontal. The left-hand end of the basilar membrane is at the base of the cochlea, while the right-hand end is at its apex. Under such assumptions, the side view will be the basis for the analyses of the vibrations on the basilar membrane that are included in the following videos (where this view will be further simplified to merely a horizontal line).

Video 5. Traveling wave along basilar membrane – static view

▶▶ https://winntbg.bg.agh.edu.pl/skrypty4/0603/video5.mp4

The mechanical properties of the basilar membrane are not uniform along its length; the membrane is narrower by a factor of about 5, and it is about 100 times stiffer at its base. Geometric proportions (not to actual scale) can be seen in Videos 2, 3, and 4. When cyclic pressure is exerted along such a body (as in Video 3), the cyclic deflection has a specific character; it occurs first at the stiffer end of the body and proceeds toward the end with the lower stiffness. This specific movement is called a traveling wave; this is presented in a three-dimensional model that corresponds to the models in Videos 2, 3, and 4. The orange wave-like element is the model of the basilar membrane (where the traveling wave propagates). The picture of the traveling wave is static, with just its front shifting toward the apex of the cochlea. In reality, the entire wave has a longitudinal movement; this will be clearly shown in upcoming two-dimensional models. The vertical deflection in this model is disproportionately large. This method of presentation is consistent with further presentations, where analyses of the vertical vibration of individual points along the basilar membrane are the objective.

The longitudinal movement of the traveling wave is presented in a 3D video in a supplementary material from the Howard Hughes Medical Institute:

https://www.youtube.com/watch?v=dyenMluFaUw

This shows a three-dimensional view of the external, middle, and inner ear, followed by a simplified three-dimensional view of the basilar membrane and its vibrations that result from several tonal stimuli (individually and simultaneously).

In this video, traveling waves are shown to be short as compared to the length of the basilar membrane. When compared to the literature, this is unrealistic; however, it is likely that this type of presentation was a compromise between the faithfulness to real data and the clarity of the picture.

Video 6. Traveling wave along basilar membrane – longitudinal movement

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video6.mp4

This and the following videos are two-dimensional, as this type of presentation is more appropriate for clearer demonstration. The starting point for the two-dimensional views of the traveling waves on the basilar membrane is a side view of the basilar membrane (i.e., the last picture in Video 4). The entire width of the pictures in the following videos does not represent the entire length of the basilar membrane – only a part of it.

The specific movement in the video is called a traveling wave; this is the basilar membrane responding to the force that is exerted over its length by the pressure of the perilymph in response to the vibration of the foot of the stapes that compresses this fluid. The horizontal speed of the traveling wave is scaled down in an effort for the picture to be easily followed by the eye and analyzed.

It can be noticed that the lengths of the subsequent cycles of the traveling wave get shorter as the wave travels along the basilar membrane.

As mentioned in the introduction, the shape of the envelope and the density of the waveform may not demonstrate real values. In fact, our knowledge on the topic is not complete. In the literature, traveling waves of the cochlea are described under two conditions: passive, and active (i.e., including the cochlear amplifier). The envelopes in all of the videos in this set are meant to present the active state. Regarding the amplitude, the vertical axis has no units, and the values should only be treated as qualitative; real deflections of the basilar membrane are in an order of nanometers. The longitudinal speed of the basilar membrane is arbitrarily chosen so that the movement can be followed by the eye, as are the frequencies of the vibrations.

Video 7. Vertical movement of chosen point on basilar membrane

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video7.mp4

This video contains the same moving plot as was featured in Video 6; however, the envelope of the traveling wave (dashed green line) is now included – this shows that the envelope remains unchanged for a constant tonal excitation. The envelope clearly shows the location of the largest amplitude of the vibration.

When there is a traveling wave in response to a tonal stimulus, individual points along the basilar membrane vibrate vertically – all at the same frequency (which is equal to the frequency of the stimulus). However, the phases of the harmonic movement change from point to point. The movement of a chosen point is marked as a red dot.

Note that the point with the highest amplitude of the vibration was chosen; the frequency of the vibration at this point is referred to as the characteristic frequency. This means that the frequency is characteristic for that particular point. Each point along the basilar membrane has its own characteristic frequency; i.e., the frequency of the stimulus at which the maximum amplitude of the traveling wave occurs. In fact, characteristic frequencies may change their positions along the length of the basilar membrane depending on the sound level.

It should be noted that there is no linear relationship between the value of the deflection of the basilar membrane and the amplitude of the acoustic stimulus.

The movement of the red dot in this video (and in the next one) follows the harmonic motion that is described by the sine or cosine functions.

Video 8. Vertical movement of chosen point – horizontally zoomed-in view

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video8.mp4

The same movement as was featured in Video 7 is shown here; however, this video zooms in on the view of a segment of the basilar membrane.

Video 9. Simultaneous vertical movements of six points along basilar membrane

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video9.mp4

The same traveling wave from Video 7 is shown with the movement of six points that are shown in the vicinity of the characteristic frequency. It can be noticed that each point vibrates at the same frequency (i.e., the frequency of the stimulus), and the neighbor of any chosen point vibrates in anti-phase. Such a phase relationship is the result of the choice of the points. Should they be chosen closer together, there would be some phase shift but not an anti-phase relationship. Should the points be chosen at an appropriate distance (approximately twice as long as in the example), they would move in phase.

Video 10. Two traveling waves along basilar membrane

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video10.mp4

This video continues the series that began with Video 6; however, this one illustrates two simultaneous traveling waves that are excited by two tonal stimuli of different frequencies. Two distant maxima of amplitudes indicate different characteristic frequencies. The traveling wave with the maximum on the left (closer to the basal end of the basilar membrane) is the response to the tonal stimulus of the higher frequency. The cycles of the traveling wave are denser at higher characteristic frequencies; this is a result of the complex contradicting effects of different frequencies of vibrations and the different velocities of the traveling waves. The latter are higher at the base than they are at the apex. There are different estimations in the literature, but a ratio of 10:1 is likely. As was declared in the introduction, the proportion of the densities of the shown waveforms does not present a real proportion – these are just arbitrarily assumed in order to illustrate the phenomenon in a qualitative way.

Video 11. Two traveling waves along basilar membrane with vertical movements of chosen points

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video11.mp4

Video 10 is supplemented by the vertical movement of two points along the basilar membrane at characteristic frequencies and their respective envelopes. The difference in the frequencies of the vibrations can now clearly be seen – this was not evident in Video 10.

Video 12. Three simultaneous traveling waves and vertical movements of points at characteristic frequencies

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video12.mp4

Video 11 is extended by introducing a third traveling wave, which is a response to a third tonal stimulus at yet another higher frequency. Besides the different frequencies of the vibrations of the respective points, strong irregularities can be noticed in the envelopes – especially at the two higher frequencies; this is the supposed effect of the interactions of the traveling waves. Some close points on the basilar membrane respond to two (or even three) of the stimulus tones. This interaction effect is not seen in the traveling wave that represents the lowest frequency (the one furthest to the right), as there is little influence of the two traveling waves that correspond to the two higher frequencies at this segment of the basilar membrane. The interferences that are shown are the result of a simple linear summation model. The real effect is likely to be similar but may differ from this picture, as a linear summation is a simplification of the complex mechanics that are involved. It seems that there have been no investigations published on this topic.

Video 13. Three simultaneous traveling waves with one of smaller amplitude

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video13.mp4

This video presents the same model as can be seen in Video 12, but the amplitude of the stimulus of the intermediate frequency is smaller this time (by 12 dB). It can be noticed that the traveling wave that corresponds to this stimulus contains a more severe disturbance from the interaction. At the same time, this stimulus disturbs the traveling wave of the higher frequency (on its left) to a lesser extent.

Video 14. Bending of stereocilia in organ of Corti

▶▶ https://winntbg.bg.agh.edu.pl/skrypty4/0603/video14.mp4

The orange line along the picture is the transverse section of the basilar membrane. Such a transverse section corresponds to one point along the length of the basilar membrane when seen from its side. On its left, the basilar membrane is attached to a flat gray object; this is a model of a stiff constraint. In fact, this constraint is part of the bone to which the basilar membrane is attached along its length on one side: the modiolus. The modiolus is a screw-like bone that constitutes the axis around which the cochlea is wound with its two and five eighth turns. On the right, there is another gray constraint; this is a model of the protrusion of the temporal bone that encapsulates the entire cochlea. Between the basilar membrane and the stiff constraints of the bone, there are joints (often referred to as pivot points) – these are shown as blue. This two-dimensional presentation may be compared with the three-dimensional one from Video 2. The two joints that are attached to the basilar membrane in Video 14 correspond to the two lines from Video 2 (along which, the basilar membrane is attached to the bones at its left and right edges).

Above the basilar membrane, there is a light-blue object; this is the transverse section of the tectorial membrane. The tectorial membrane on its left is attached to another protrusion of the modiolus.

Points A and B and the line that joins them are merely meant to show the relative positions of any pair of opposite points on two membranes when both membranes are in a neutral position; i.e., when there is no deflection of the basilar membrane.

When the basilar membrane moves upward (as the effect of the vibrations that can be seen in Videos 7–9, 11–13, and 16), the geometry of this system acts in such a way that the displacement of the basilar membrane results in a radial shearing motion between the basilar membrane and the tectorial membrane. In other words, the opposing surfaces of the two membranes shift in opposite directions. This can be seen in the bending of the A-B line from a vertical position to an angled one. This movement exerts shearing forces on the tips of the stereocilia at the tops of all four of the hair cells that can be seen in the cross-section. The shearing force that acts on the stereocilia of the hair cell on the left (called an inner hair cell) is mainly exerted by the movement of the fluid (the endolymph). The shearing force that acts on the three stereocilia from the right (called the outer hair cells) is exerted by the mechanical connection between the stereocilia and the tectorial membrane. The effect of the shearing force is that the stereocilia are bent to the right; i.e., in the direction of the highest row of the stereocilia.

When the basilar membrane moves downward in response to the opposite phase of the air pressure, the relative movement of the surfaces of the two membranes is also opposite. Thus, the shearing force acts in the opposite manner – bending the stereocilia to the left; i.e., in the direction of the lowest row of the stereocilia.

Video 15. Opening of ion channels in stereocilia

▶▶ https://winntbg.bg.agh.edu.pl/skrypty4/0603/video15.mp4

The bending of the stereocilia that was demonstrated in Video 14 initiates another process: the depolarization of a hair cell.

In Video 15, a magnified view of a hair cell can be seen. The body of the hair cell is shortened (a real hair cell is longer) so that some of its elements can be better seen. On top of the cell body, two stereocilia are shown; each one represents an entire row of stereocilia of similar heights (there are three or four rows). The top of the lower stereocilium and the higher stereocilium are joined by the tip link (a specific ligament).

In a neutral state (i.e., when there is no tonal stimulus), there is a negative potential inside the hair cell body; this is supported by ion pumps in the cell wall. At the same time, the endolymph surrounding the hair cells has a positive potential; this is supported by ion pumps in the wall of the cochlea (the tissue that is called *stria vascularis*).

This video shows details of the depolarization process that takes place in the hair cell. When the bending of the stereocilia occurs, the tip link that joins a higher stereocilium with a neighboring lower one pulls the upper surface of the lower stereocilium upward. In effect, a channel is open through which positive potassium and calcium ions flow into the stereocilium and further into the hair cell body (thus depolarizing it). For a while, the charge inside the hair cell grows positive; this can be seen by the greater number of positive ions than negative ones.

Below the hair cell, a synapse of a neuron that belongs to the auditory nerve can be seen. Depolarization causes a chemical neurotransmitter to be released into the tiny space between the hair cell and the neuron's synapse. This is followed by the production of a neural spike in the neuron.

The action that is presented in Video 15 occurs in both the inner and outer hair cells (even though there are substantial differences in their roles in the process of hearing). About ten afferent neurons are attached to each inner hair cell, but only one is attached to each outer hair cell.

A good summary of the function of the organ of Corti is provided in the video that was prepared by NeurOreille (S. Blatrix, G. Rebillard, and R. Pujol) and is available at the following link:

http://www.cochlea.eu/en/cochlea/organ-of-corti

Here, the complete two-dimensional transverse section of the organ of Corti is presented; this shows the up-and-down movement of the basilar and tectorial membranes, the shearing force, the bending of the stereocilia, and the repolarization of the hair cells.

Video 16. Three simultaneous traveling waves with timings of neural spikes

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video16.mp4

When the deflection of the basilar membrane is sufficient, the generation of a neural spike (or impulse) may occur (as was explained in Video 15). The neural spike is produced in a neuron that is attached to a hair cell; it is then transmitted to the central auditory system through neural fibers and a couple of neural centers. In this video, the likely moments of the generation of neural spikes are visualized. The period of time when the generation of a spike in one neuron occurs is marked symbolically with a red arrow. This relationship between the vibration of a point on the basilar membrane and the timing of the spikes is justified by the mechanics in the organ of Corti, but it is a hypothetical one. Experimental data is available on the relationship between the value of the input acoustic stimulus and the timing of the spikes; this is visualized in Videos 17 and 18.

Video 17. Rule of phase locking observed in individual neurons

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video17.mp4

Each neuron tends to generate neural spikes at one specific phase point of the cycle of the tonal stimulus. This attribute is called "phase locking," as each neuron is "locked" to some value of the phase. This has been confirmed in a number of experiments. It is likely that a similar phase relationship also holds true between the oscillatory vertical movement of the basilar membrane and the moments of the spikes (as was presented in Video 16). This effect works up to input frequencies of about 4 kHz; above this range, the phase locking breaks down. The phase locking is not very accurate, as there are some time irregularities (jitters); however, it still provides useful timing information for the higher stages of the auditory system.

In the video, a zoomed-in segment of a sinusoidal acoustic stimulus is shown along the horizontal axis; the phase reference point was assumed as the local peak of the stimulus. The firing in each neuron occurs in some fixed phase; i.e., some fixed time that is related to this peak. This can occur before the peak (this is called "phase lead" in the video), at the peak (this is called "phase coincidence"), or after the peak (this is called "phase lag").

It is important to note that the firing, in fact, does not take place during the positive cycles of the acoustic tonal input but during the negative cycles (compare Videos 3 and 14). The effect is presented as if it happened during the positive cycles to stay consistent with most of the literature (and with Video 16).

The point of the firing during the cycle of the stimulus is demonstrated in the form of a colored dot. This is similar to the way that selected positions along the basilar membrane were presented in previous videos; however, the horizontal position of the dot marks the point in time here, not along the line.

A magenta point appears first in the video; this is the symbol for one of the neurons that is characterized by an early time of the firing. "Early" here means early within the cycle; i.e., with a big phase lead as related to the peak of the tonal stimulus. The moment of the firing is marked by a magenta arrow. Simultaneously, a horizontal magenta line appears; this demonstrates the time that is left in the cycle until the maximum of the stimulus occurs (i.e., the phase lead of the spike).

The next to be shown is a beige point; this is the symbol for another neuron that is characterized by an early time of the firing. In this case, the phase lead is not as big as in the case of the "magenta" neuron. Again, the beige neuron fires with a phase lead. The value of that lead is marked by a horizontal beige line (as was done before with the magenta "neuron").

Then, a green point appears; this is the symbol for a neuron that fires in synchrony with the maximum of the tonal stimulus. This neuron fires at the moment when the stimulus reaches its maximum. The firing is phase coincident with the maximum (this is shown by a green arrow).

Finally, a red point can be observed; this happens to fire after the peak has passed and the acoustic pressure of the stimulus is within the falling phase. A horizontal red line shows the phase lag.

Video 18. Rule of phase locking observed in four neurons

▶▶ https://winntbg.bg.agh.edu.pl/skrypty4/0603/video18.mp4

This video is similar to the previous one; however, one can observe the firing behavior of all four example neurons from the previous video in one cycle of the input tonal stimulus. Two subsequent cycles are presented in order to show that the phase locking repeats from cycle to cycle.

Video 19. Frequency separation on basilar membrane – two partly overlapping traveling waves

▶▶ https://winntbg.bg.agh.edu.pl/skrypty4/0603/video19.mp4

When two tones with similar frequencies are heard (such that both stimuli fall into one auditory filter), they cannot be resolved; we then perceive one tone with beatings (in the same manner that can be observed in a waveform that is the result of summing up two respective sinewaves). When the difference between the frequencies is wide enough so that both stimuli fall into separate auditory filters (especially non-neighboring ones), we can perceive two separate tones.

The effect of bandpass filtering in the basilar membrane can be explained by the model of traveling waves and characteristic frequencies.

The video illustrates two traveling waves that originate from two input tones that are not far apart in frequency but far enough apart so that they can be separated. The vibrations at two characteristic frequencies (shown by two red dots) demonstrate that they vibrate at fairly constant amplitudes, thus showing no clear signs of amplitude modulation (i.e., beatings). The frequency of the vibration of the point on the left is somehow greater than that of the point on the right (which is consistent with the general traveling wave model).

Video 20. Frequency separation on basilar membrane – two closely overlapping traveling waves

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video20.mp4

Two traveling waves originate from two input tones that are very similar in frequency; they are close enough to fall into one auditory filter and, thus, cannot be separated. The vibrations at two characteristic frequencies that are close (shown by two red dots) demonstrate that they vibrate at varying amplitudes; this clearly shows amplitude modulation (i.e., beatings). The frequency of the vibration of the traveling wave on the left is close to that of the traveling wave on the right because of the closeness of the frequencies of their respective tonal stimuli.

Video 21. Measurement of width of critical band – noise with constant spectrum density level

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video21.mp4

A parameter of noise that is related to its power is its spectrum density level (L_0):

$$L_0 = 20\log\left(\frac{\frac{P}{B}}{10^{-12}}\right) \tag{1}$$

where P is the noise power, and B is the bandwidth of the noise (i.e., L_0 is the level of power per bandwidth ratio).

A narrow band of noise is presented to the listener. The band of noise is centered at 2 kHz and is illustrated by a light-green rectangle. The level of noise is shown in the vertical axis of the diagram. The center frequency of the noise and its bandwidth are marked on the horizontal axis. Concurrently a tone at a frequency of 2 kHz is presented; however, its amplitude is gradually increased from low levels. At low levels, the tone is inaudible because it is masked by the band of the noise. At some level of the tone (when it reaches its threshold of masking), the tone emerges as being audible. The level of the tone when this happens is marked by an orange dot on the vertical axis (along with its accompanying value).

This experiment is repeated a number of times; each time, the bandwidth of the noise is increased, but its L_0 remains unchanged. Increasing the bandwidth with L_0 being kept constant can be seen as the widening of the light-green rectangle (with its height remaining constant). Although L_0 does not change, the total noise power does indeed increase when the bandwidth is raised.

In the video, the bandwidth of the noise is increased to the following steps: 50, 100, 200, 400, 800, and 1600 Hz. Up to 400 Hz, each increase is followed by an increase in the masking threshold of the tone. At 400 Hz of noise bandwidth, this is equal to 75 dB; however, when the bandwidth is increased further (to 800 and 1600 Hz), the threshold does not increase any further (although the power of the masking noise and, thus, its level is increased by 3 dB each time the bandwidth is doubled).

This is considered to be proof that auditory filters (the widths of which being expressed by the respective values of the critical bands [CB]) do indeed exist. This is also considered the method to determine their widths. Our hearing system analyzes sound in separate auditory filters; a hypothetical auditory filter is presented as a dark-green bell-shaped curve in the picture. When there is a tone that is accompanied by noise in one auditory filter, the tone must compete with the noise in order to be audible (but only with the noise within that auditory filter). Auditory filters attenuate sounds that are beyond their passbands; thus, these sounds do not interfere with the sounds that are inside each auditory filter. This explains the results of the experiment. As long as the band of noise is narrower that the width of the filter (i.e., the CB), increasing its bandwidth increases its power inside the filter as well as its masking potential; therefore, the threshold of the audibility of the tone is raised. When the band of noise exceeds the limits of the CB, its power inside this band does not increase any more (as it is passed through neighboring auditory filters). Despite the fact that the total power of the noise still increases, the power inside the CB does not (so, the masking threshold of the tone is not increased).

Video 22. Measurement of width of critical band – noise with constant power

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video22.mp4

As in the previous experiment, a narrow band of noise that is centered at frequency f_0 is presented to the listener. The band of noise is shown as a blue rectangle. This time, it is more convenient to consider the spectrum density; i.e.:

$$N_0 = \frac{P}{B} \tag{2}$$

and not its level that can be seen in Equation (1). The value of N_0 is presented on the vertical axis, and the bandwidth of the noise is shown on the horizontal axis. This time, the spectrum density of noise N_0 is not constant – it is continuously decreased; however, its bandwidth is increased proportionally so that the total power of the noise remains constant:

$$P = BN_0 \tag{3}$$

The continuous change of the band of the noise is shown in the video. There is no concurrent tone presented, and the listener's task is to evaluate the subjective loudness of the noise.

As the proportion of N_0 to B is changed from a high value (i.e., a tall narrow band) to lower values, the subjective assessment of the loudness by listeners is constant (two sones). A sone is a subjective measure of loudness; loudness in sones is indicated as the height of the orange bar in the indicator that is shown on the right side in the picture.

However, when the proportion of N_0 to B reaches a certain value, the subjective loudness gradually increases and continues to do so with each further lowering of the N_0 to B proportion; i.e., when the spectrum density decreases but the bandwidth increases.

This may seem to be a paradox, as the power of the noise is kept constant all of the time according to Equation (3). Intuitively, we tend to expect that, with the constant power of a noise, the perceived loudness will remain constant as well. This seemingly counter-intuitive effect is considered to be further proof of the existence of auditory filters and the means for determining their widths (i.e., the critical band). The supposition is that, if our hearing system analyzes sounds in separate auditory filters and the outputs of the filters are transmitted to the central auditory system through independent channels, then the filters should independently contribute to the total assessment of the loudness. Thus, as long as the band of noise passes through just one auditory filter, its loudness is perceived as being constant regardless of its spectral shape (assuming that it has constant power). As soon as some part of noise is passed through a neighboring filter as a consequence of broadening its bandwidth, however, this neighboring filter provides a new contribution to the perception of the loudness. This effect is strong enough to overwhelm some decreases of the noise power in the filter where the noise was first present (when the same power is shared among the neighboring filters).

In the video, extending the bandwidth of the noise beyond one auditory filter (marked in dark green) can clearly be seen.

Video 23. Audiovisual presentation of beatings within range of 0–60 Hz

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video23.mp4

Two tones with equal amplitudes are presented to the listener. The frequency of the first one is kept constant at $f_0 = 500$ Hz, but the frequency of the other is continuously changed at a linear rate; i.e.:

$$f(t) = f_0 + \delta f t \tag{4}$$

where δf is the rate of the change in hertz per second.

The frequencies of both tones are depicted by two orange bars that are placed along the horizontal axis (which is the frequency axis). The numbers under the frequency axis denote the values of $\Delta f = f - f_0$.

As Δf increases, the bar that denotes the frequency of the higher tone continuously shifts; this indicates the current value of its frequency. Simultaneously, the audio layer of the presentation conveys an acoustic signal that is produced by the superposition of the two tones. This signal is perceived by the ear as beatings, as both tones fall into one auditory filter. The momentary value of Δf in the audio layer corresponds exactly to the value of Δf that is depicted graphically in the video layer.

At around 15 Hz, there is the upper limit of the human perception of such beatings; this value is indicated on the horizontal axis. Above this value of Δf, the perception of the beatings changes to the perception of a rough tone. According to the literature, the most dissonant distance of two frequencies is around one fourth of the critical band (CB). As the CB is 110 Hz at 500 Hz, the most dissonant range of Δf is expected at about the middle of the frequency range that is encompassed in this audiovisual presentation. The upper end of Δf corresponds to slightly above half of the CB.

Video 24. Audiovisual presentation of beatings within range of 60–570 Hz

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video24.mp4

As in the previous demonstration, two tones with equal amplitudes are presented to the listener. The frequency of the first one is kept constant at f_0 = 500 Hz, and the frequency of the higher tone is continuously changed at a linear rate of 60 through 570 Hz; i.e., Δf begins at the point where the previous demonstration was stopped. The reason for splitting these two demonstrations is to allow for the slower change of Δf in the previous presentation in order to hear the different phases of the beatings. The rate of the change of Δf is faster in the current demonstration. The splitting also allows for marking the values on the frequency axis in a more readable way.

Along the frequency axis, several values of Δf are marked (expressed as consecutive widths of critical bands [CBs]). At some point of Δf (but within one CB), the perception of the rough tone gives way to two separate tones (still rough). At some further point of Δf, the two separate tones become smooth. According to the literature, this transition should take place when Δf exceeds one CB. Individual perceptions from this video may not strictly follow these rules – partly because of the frequency characteristics of the headphones that are used not being flat enough. Close to the end of the range, there is a value of Δf that is equal to one octave; at this point, a highly consonant sound is perceived (which is the expected effect).

Video 25. Visual presentation of beatings within a range of 0–15 Hz

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video25.mp4

This video contains the waveform of the stimulus that was used in Video 23. The width of the picture corresponds exactly to one second of the signal, and this one-second-wide segment of the signal is shifted to the right as time progresses. The shift is synchronous with real time and was used in both the audio and video layers of Video 23. Only a range of 0–15 Hz is covered in the visualization, as the picture moves too quick to be followed by the human eye at the higher frequencies of the beatings. On top of this, a phantom picture of the waveform that shifts either to the left or right can be seen; this is due to the occurrence of video sampling rate aliasing. These effects can already be observed close to a 15-Hz difference in the frequencies in this video.

Video 26. Model of simultaneous masking on basilar membrane – 500 Hz masker

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video26.mp4

Two tonal stimuli at 500 and 1000 Hz produce traveling waves; the amplitude of the 500-Hz tone (the masker) is constant. The vertical movements of two points along the basilar membrane at characteristic frequencies of 1000 and 500 Hz are shown by the blue dots. At the beginning of the video, only the 500-Hz stimulus is present so its traveling wave can be seen.

The amplitude of another tone at 1000 Hz (the signal) is increased continuously from a value of 0 to be equal to that of 500 Hz. In fact, the increase in the amplitude in the video is not linear; it has been made exponential for a more appealing presentation.

The envelope of the traveling wave that is induced by 500 Hz is shown in red, and it is constant. The changing envelope of the traveling wave that is induced by 1000 Hz is shown in green. The sum of the two envelopes is also shown (in blue).

When the amplitude of the 1000-Hz tone is small, the movement of the point at this characteristic frequency is dominated by the traveling wave that is induced by 500 Hz. The point vibrates at a frequency of 500 Hz. The 1000 Hz-tone is masked and is inaudible. As the amplitude of the 1000-Hz tone increases, the movement of this point turns to a superposition of two vibrations. For a considerable amount of time, the movement is not sinusoidal; only when the contribution from the traveling wave that is induced by 1000 Hz becomes considerably higher than that of the traveling wave that is induced by 500 Hz, the movement becomes dominated by a sinusoidal vibration at 1000 Hz. The change of the character of the vibration at this point only takes place when the green envelope considerably exceeds the red envelope. At this value of the amplitude of the 1000-Hz tone, it exceeds the threshold of the masking and becomes audible.

The video presents a simple model of simultaneous masking on the basilar membrane. The signal can be heard only if the amplitude of its traveling wave at its characteristic frequency clearly exceeds the amplitude of the traveling wave of the masker. Otherwise, the vibration at the characteristic frequency of 1000 Hz is dominated by the contribution from the masker (500 Hz); so, the 1000-Hz frequency cannot be detected by the human ear.

As in some of the earlier videos, a reservation must be expressed: this is a simplified model.

Video 27. Model of simultaneous masking on basilar membrane – 1000 Hz masker

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video27.mp4

The same stimuli that can be found in Video 26 are used; this time, however, the roles are swapped between 500 and 1000 Hz (1000 is the masker, and 500 Hz is the signal). The amplitude of the 1000-Hz tone is constant. At the beginning of the video, only the 1000-Hz stimulus is present. The amplitude of the 500 Hz signal is increased continuously from a value of 0.

The point at the characteristic 500-Hz frequency vibrates at this frequency (even at very low amplitudes); no interaction with the traveling wave of the 1000-Hz masker can be observed. This is the case because of the steeply descending part of all of the traveling waves' envelopes.

The conclusion is that the 1000-Hz tone is inefficient at masking the 500-Hz tone. A comparison of Videos 26 and 27 reveals the general rule of simultaneous masking: “low masks high,” meaning that low frequencies tend to mask high frequencies much more efficiently that the other way around.

Video 28. Equal loudness contours

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video28.mp4

The graph in this video is a set of curves that are referred to as equal loudness contours (or curves). If the frequency of a tonal stimulus is swept through all audible frequencies in such a manner that its level exactly follows one of the curves, the perceived loudness is the same over the entire sweep (regardless of the frequency). The curves are specified in ISO standard 226:2003 on the basis of numerous experiments that featured the participation of young people. These curves are marked in red, while the blue curve represents the earlier ISO standard.

The movement along one of the curves is shown by the black dot. In each position of this dot, the loudness is the same. The position of the dot determines both the frequency of the tonal stimulus (on the horizontal axis) and the level of the tone (on the vertical axis). As long as the frequency and level are equal to those that are indicated by the black dot, the perceived loudness is the same. The same rule applies to any position along the other “red” curves.

The dot slides along the red curve that is labeled “60.” This expresses the fact that, at a frequency of 1 kHz, the level that is indicated by this curve is 60 dB. The markings of the other curves follow the same rule.

The shape of a curve (which is different at each sound level) is the result of the sensitivity of the human ear changing with the frequency and sound level.

The concept of equal loudness contours is closely related to a specific measure of the human perception of loudness (called the loudness level). One unit of the loudness level is called a phon. This measure is meant to be only used with tonal (or narrowband) stimuli. The rule can be simply explained with the help of this video (in the case of the curve with the sliding black dot).

At any position of the black dot, the loudness level equals 60 phons.

The actual sound level at each frequency is that which is indicated by the black dot on the vertical axis. However, the loudness level at any position of the black dot is equal (e.g. along the red curve it is 60 phons, as the red curve indicates 60 dB at 1 kHz). Audio was not added to this video, as a true presentation of constant loudness would require the accurate control of the level of the sound that is delivered to the listener's ear.

For any other curve, the same rule applies with accordingly changed values of the loudness level (indicated at 1 kHz). For any tone that does not belong to any of the curves, a specific and appropriate equal loudness contour must be used.

The graph that is used in the video was made by Lindosland and is in the public domain:

https://en.wikipedia.org/wiki/Image:Lindos1.svg

Video 29. Interaural time delay

▶▶ https://winntbg.bg.agh.edu.pl/skrypty4/0603/video29.mp4

A plane acoustic wave passes around the listener's head; it has a definite direction such that each particular phase of it arrives first to his/her right ear and later to his/her left ear. A simple geometrical analysis shows that there is a simple relationship between the angle of incidence and the delay between the times of arrival:

$$\Delta T\,[\mathrm{ms}] = 0.257\,(\theta + \sin\theta) \tag{5}$$

where θ is the angle of incidence, and $\theta = 0$ when the sound arrives frontally. This simple relationship cannot be used at frequencies that are above around 1500 Hz, as ΔT produces ambiguous hints with the cycles of pressure being closer in space. It is also difficult to discriminate ΔT at low frequencies (below 200 Hz).

Video 30. Interaural intensity difference

▶▶ https://winntbg.bg.agh.edu.pl/skrypty4/0603/video30.mp4

A listening situation is presented that is similar to that of Video 29. Another effect is visualized; it is based on the fact that, when an acoustic wave passes around an object and half of its wavelength is similar or less than the size of the object, an acoustic shadow is formed on the side of the object that is opposite to the sound incidence. Within this acoustic shadow, the sound intensity is lower than it is outside of it.

Some slight effect of the acoustic shadow can be observed at frequencies of a couple of hundred Hz. A more pronounced effect begins for frequencies that are above about 1.8 kHz, and a really strong effect (up to 20 dB) can be observed starting at 6 kHz.

During the first part of the video, half of the wavelength of the sound is close to the diameter of the listener's head. The effect of an acoustic shadow can be observed (although it is not a strong one). In the second part, the frequency of the sound is doubled – this can be seen in halved distances between the wave peaks. The acoustic shadow is clearly stronger.

Both the interaural time delay and interaural intensity difference need both ears in order to be observed; therefore, they are called binaural cues.

Video 31. head-related transfer functions

https://winntbg.bg.agh.edu.pl/skrypty4/0603/video31.mp4

When sound arrives at a human ear, the shape of its time waveform is changed as an effect of the attenuation, diffraction, and reflections that are caused by arms, the head, and the pinna (the last element introduces especially complicated effects due to its complex three-dimensional shape). These effects are frequency-dependent and, together, act as a type of filter. The frequency characteristics of this filter are referred to the head-related transfer function (often abbreviated as HRTF). These effects depend on the direction of the sound incidence; so, the shape of the filter curve depends on the direction. In fact, any direction of incidence has its own characteristic filter. This is demonstrated in the video.

Some example directions of incidence are shown as wide green arrows, and the frequency characteristics of the filters that are active in these directions are shown in the forms of graphs (where frequency is on the horizontal axis, and amplitude is on the vertical one). The phase components of the filters are not shown. It can be seen that these filter shapes differ substantially (especially at middle and high frequencies), and they tend to include very sharp and narrow notches.

Bibliography

1. Blauert J. (1996). *Spatial Hearing: Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, USA.
2. Brungart D.S., Rabinowitz W.M. (1999). Auditory localization of nearby sources. Head-related transfer functions, J. Acoust. Soc. Am., **106**, 1465–1479.
3. Chatterjee M., Zwislocki J.J. (1997). Cochlear mechanisms of frequency and intensity coding. I. The place code for pitch, Hear. Res., **111**, 65–75.
4. Egan J.P., Hake H.W. (1950). On the masking pattern of a simple auditory stimulus, J. Acoust. Soc. Am., **22**, 622–630.
5. Feth L.L., O'Malley H. (1977). Two-tone auditory spectral resolution, J. Acoust. Soc. Am., **62**, 940–947.
6. Gardner M.B., Gardner R.S. (1973). Problem of localization in the medial plane: effect of pinnae cavity occlusion, J. Acoust. Soc. Am., **53**, 400–408.
7. Gelfand S.A. (2010). *Hearing: An Introduction to Psychological and Physiological Acoustics*, Informa Healthcare, Colchester, UK.
8. Goldstein E.B. (2010). *Sensation and Perception*, Wadsworth, Cengage Learning, Belmont, CA, USA.
9. Hartmann W.M. (1998). *Signals, Sound and Sensation*, Springer-Verlag, New York.
10. Heijden M., Kohlrausch A. (1994). Using an excitation-pattern model to predict auditory masking, Hear. Res., **80**, 38–52.
11. Howard D.M., Angus J.A.S. (2009). *Acoustics and Psychoacoustics,* Fourth Edition, Focal Press, Oxford, UK.
12. Jennings A.R. (2005). *On Mechanisms for the Analysis of Spectral and Temporal Envelope Shape in the Human Auditory System* [Ph.D. thesis], Newcastle University, Newcastle upon Tyne, UK.
13. Kleczkowski P., Plewa M., Pluta M. (2011). Masking a frequency band in a musical fragment played by a single instrument, Acta Phys. Pol. A, 119, 991–995.

14. Kleczkowski P., Pluta M. (2012). Frequency discrimination in a simultaneous two-tone signal, Acta Phys. Pol. A, **121**, A120-A125.
15. Lutfi R.A. (1983). Additivity of simultaneous masking, J. Acoust. Soc. Am., **73**, 262–267.
16. Moore B.C.J. (ed.), (1995). *Hearing*, Academic Press, London.
17. Moore B.C.J. (2002). Interference effects and phase sensitivity in hearing, Phil. Trans. R. Soc. Lond. **A 360**, 833–858.
18. Moore B.C.J. (2003). *An Introduction to the Psychology of Hearing*, Fifth Edition, Academic Press, London.
19. Ozimek E. (2018). *Dźwięk i jego percepcja. Aspekty fizyczne i psychoakustyczne*, Wydawnictwo Naukowe PWN, Warszawa.
20. Plack C.J. (2014). *The Sense of Hearing*, Psychology Press, New York.
21. Rabinowitz W.M., Bilger R.C., Trahiotis C., Neutzel J. (1980). Two-tone masking in normal hearing listeners, J. Acoust. Soc. Am., **68**, 1096–1106.
22. Rose J.E., Brugge J.F., Anderson D.J., Hind J.E. (1967). Phase-locked response to low frequency tones in single auditory nerve fibers of the squirrel monkey, J. Neurophysiol., **30**, 769–793.
23. Ruggero M.A., Rich N.C., Recio A. (1996). The effect of intense acoustic stimulation on basilar-membrane vibration, Aud. Neurosci., **2**, 329–345.
24. Ruggero M.A., Shyamla N.A., Temchin A.N., Recio A. (2000). Mechanical bases of frequency tuning of the cochlea: comparison of basilar membrane vibrations and auditory-nerve-fibre responses in chinchilla, Proc. Natl. Acad. Sci., **97**, 11744–11750.
25. Van Opstal J. (2016). *The Auditory System and Human Sound-Localization Behaviour,* Academic Press, London.
26. Warren R.M. (2008). *Auditory Perception. An Analysis and Synthesis*, Cambridge University Press, Cambridge, UK.
27. Wightman F., McGee T., Kramer M. (1977). Factors influencing frequency selectivity in normal and hearing-impaired liteners. In: *Psychophysics and Physiology of Hearing*, E.F. Evans, and J. Wilson, Eds., Academic Press, London, 295–306.
28. Yost W.A. (2008). *Fundamentals of Hearing. An Introduction*, Emerald Group, Bingley, UK.
29. Zwicker E., Fastl H. (2007). *Pychoacoustics. Facts and Models*, Springer, Berlin – Heidelberg.

Brief English-Polish dictionary of most important terms

absolute threshold of hearing	–	bezwzględny próg słyszenia
afferent fiber	–	włókno dośrodkowe (aferentne)
auditory filter	–	filtr słuchowy
band of noise	–	pasmo szumu
bandwidth	–	szerokość pasma
basilar membrane	–	błona podstawna
beating	–	dudnienie
cochlea	–	ślimak
critical band (CB)	–	pasmo krytyczne
efferent fiber	–	włókno odśrodkowe (eferentne)
endolymph	–	endolimfa
equal loudness contours, isophonic curves	–	krzywe jednakowego poziomu głośności
hair cell	–	komórka rzęsata (słuchowa)
level	–	poziom
masking	–	maskowanie
masking threshold	–	próg maskowania
neural impulse, neural spike	–	impuls neuronowy
oval window	–	okienko owalne
perilymph	–	przychłonka, perylimfa
phase locking	–	synchroniczność fazowa
phon	–	fon

Reissner's membrane	–	błona przedsionkowa (Reissnera)
roughness	–	szorstkość
round window	–	okienko okrągłe
simultaneous masking	–	maskowanie równoczesne
sone	–	son
sound level	–	poziom dźwięku
sound pressure level, SPL	–	poziom ciśnienia akustycznego
spectrum density level, spectrum level	–	poziom gęstości mocy
stapes, stirrup	–	strzemiączko
stereocilia	–	rzęski
stimulus	–	bodziec
tectorial membrane	–	błona pokrywkowa, nakrywkowa
temporal bone	–	kość skroniowa
threshold	–	próg
traveling wave	–	fala biegnąca

e-ISBN 978-83-67-427-14-2