Auditory and Vestibular Systems

Nicholas Cothros

The present review concerns the human auditory system and the human vestibular system. These two systems are remarkably sensitive and precise. Although neophytes may fail to intuit a link between hearing and balance, the two are related on both anatomical and functional levels. The auditory system will be considered first.

Auditory System

The nature of sound. "Sound" is the term given to changes in pressure generated by vibrating molecules in a medium (like air or water). The vibrating tines of a tuning fork, the pulsating membrane of a stereo speaker, or the vibrating vocal cords of a human speaker all give rise to compressions in the air. These compressions are followed by rarefactions. Imagine the vibrating membrane of a stereo speaker. As the membrane oscillates outwards, air is compressed. Before a second compression occurs, the membrane moves inward, producing a rarefaction.

Sound, which is the product of alternating compressions and rarefactions of air over time, can be modeled by a sinusoidal curve. The rhythmic rising and falling of a sinusoid capture the respective compressions and rarefactions. The amplitude of the sinusoid pertains to the loudness of the sound. The period of the sinusoid (how many peaks occur per unit of time) pertain to the frequency or pitch of the sound.

A pure tone (in other words, a sound of just one frequency, like 1000 Hz) can be modeled by a simple sine wave. In daily experience, sounds are far more complex and are composed of multiple frequencies superimposed on one another. A real sound wave looks rather complex. The auditory system manages to dissect a complex sound wave into its component frequencies.

The ear. The human ear comprises three portions, namely, the outer ear, middle ear, and inner ear. The outer ear includes the pinna (the mass of cartilage that most would refer to as the "ear") and the ear canal, which ends at the tympanic membrane or "ear drum". The ear drum is the bridge to the middle ear, an air-filled cavity with three ossicles (tiny bones), called the malleus, incus and stapes. The stapes is attached to another membrane, like the ear drum only smaller, called the oval window. The oval window is attached to an inner ear structure called the cochlea, which is a liquid-filled series of tubes.

Sound enters the ear canal and vibrates the ear drum. The vibration of the ear drum is transferred to the mobile ossicles, whose motion vibrates the oval window. The vibration of the oval window in turn leads to compressions and rarefactions in the liquid inside the cochlea. The perception of sound is tied to the motion of this fluid and how it affects neural activity.

Hearing involves converting changes in pressure (such as compressions and rarefactions in air) to mechanical vibrations. The ear drum does not, however, connect to the cochlea directly. The intervening action of the ossicles performs a number of important functions to make hearing possible. This is considered next.

The ossicles. A major problem confronted by the auditory system is that the ear canal is air-filled and the cochlea, where neural transduction first takes place, is liquid-filled. Liquids have a much higher impedance than air. Most sound is reflected by the surface of water.

This is why those submerged in a pool of water usually cannot hear raised voices outside the water. The ossicles together act as an impedance matcher, faithfully conveying ear drum vibrations to the oval window with minimal loss of information.

Impedance matching is afforded by the lever-like action of the ossicles. The relatively large displacement of the ear drum causes progressively smaller displacements of each subsequent ossicle. The final ossicle, the stapes, produces small yet forceful displacements, concentrated over the relatively small area of the oval window. These vibrations are strong enough to displace the liquid beyond the oval window.

The cochlea. The cochlea is essentially a collection of three parallel liquid-filled tubes: the scala vestibuli, the scala media and the scala tympani. Displacement of the oval window sends a pressure wave through the scala vestibuli. This causes vibration inside the scala media, which houses the organ of Corti.

The organ of Corti consists of hair cells held between two membranes. The tectorial membrane rests on top of the hair cells, with the "hairs" embedded within the membrane; the basilar membrane sits beneath the hair cells. As the pressure wave described earlier travels the length of the scala vestibuli, its vibration is transmitted to the basilar membrane, which also vibrates.

Oscillation of the basilar membrane causes the hair cells and tectorial membrane to osscillate accordingly. The basilar membrane and the tectorial membrane are anchored in different locations within the cochlea, such that the vertical motion of the hair cells results in lateral motion of the stereocilia (the "hairs"). Spring-like filaments hold the hairs together. As the hairs bend toward the kinocilium (the largest hair), the springs are stretched. These springs are in turn mechanically linked to K+ channels, which are pulled open by the motion of the hairs, depolarizing the hair cell. Depolarization of hair cells increases the rate of exctitatory neurotransmitter released onto the afferent nerve fiber, increasing the rate of firing of the nerve fibre. In this manner, a sound signal is sent to the brain.

Coding sound frequency. Sound causes displacement of the ear drum, which ultimately displaces liquid inside the cochlea. The liquid displacement results in the vibration of the basilar membrane and this in turn causes depolarization of hair cells sitting on the membrane. So begins the neural transduction of sound.

One of the important features of a sound wave is its frequency. Humans can detect frequencies from 20 Hz to 20,000 Hz. However, neurons (including receptors like hair cells) are limited in their firing rate. Neurons are not able to fire at frequencies far beyond 1,000 Hz. As a result, the nervous system is ill equipped to represent sound frequency by simply producing a matching firing frequency. Human speech has frequency content of roughly 3,000 Hz, so even the frequency of human speech sounds is too high to be represented by producing a matched frequency of neural firing.

The shape of the basilar membrane helps to solve the problem of coding sound frequency. The basal end of the basilar membrane, near the oval window, is narrow and quite taut, much like the high string of a guitar. Moving towards the apical end, the basilar membrane becomes progressively wider and less taut, like the low string of a guitar.

The consequence of varying the shape of the basilar membrane in this manner is that the location of maximal displacement varies with the frequency of sound. A high frequency sound maximally displaces the basal end of the basilar membrane. A low frequency sound maximally displaces the apical end of the basilar membrane. The nervous system exploits this property by coding sound frequency tonotopically. In other words, the firing of hair cells attached to the basal end of the basilar membrane is interpreted by the brain as high frequency sound. The firing of hair cells attached to the apical end of the basilar membrane is interpreted by the brain as low frequency sound.

Coding sound frequency tonotopically allows the nervous system to overcome the limits of neural firing rate and represent frequencies over a large range. Tonotopy exists throughout the auditory system.

Coding sound amplitude. Another important feature of a sound wave is its amplitude or loudness. Like frequency, amplitude is likely coded by more than simply increasing neural firing rate. Other coding mechanisms involve multiple sets of neurons with different thresholds, such that the loudest sounds are associated with additional recruitment of neurons with the highest threshold.

Perception of loudness may also be partly the result of amplification via the outer hair cells of the organ of Corti. Receptor potentials in an outer hair cell trigger vibrations in the outer hair cell body. The vibrations are oscillations in the outer hair cell's length. These vibrations move the tectorial membrane and help to amplify the sound and increase hearing sensitivity. Interestingly, the reflex loop that occurs between the inner hair cells and the outer hair cells gives rise to otoacoustical emissions.

Coding sound direction. The hearing experience is incomplete without determining the origin of a sound in space. Having two ears is an important component of sound localization as it grants the nervous system an opportunity to compare sounds arriving at each ear.

One criterion the nervous system uses is the difference in sound intensity at each ear. This gives a clue regarding the direction a sound is coming from. The head, which of course lies between the two ears, muffles sound. A sound arriving from the left will sound more intense in the left ear than in the right ear. The more lateralized the sound, the greater the difference in intensity between the ears. The neural circuitry underlying this sort of localization begins in the auditory brainstem.

For example, a sound from the left excites the lateral superior olive on the left side, which excites higher auditory areas. At the same time, a sound from the left actives an inhibitory neuron in the trapezoid body, which inhibits the lateral superior olive on the right side. Insofar as net excitation on the left is greater than net excitation on the right, the brain interprets the sound as coming from the left. Note that sound localization of this sort is best for high frequency sounds (greater than 3,000 Hz). The wavelength of low frequency sounds is too great to cast a "sound shadow" around the head.

The nervous system is also capable of localizing sound by comparing the time at which a sound arrives at each ear. A sound coming from the left will arrive at the left ear before it arrives at the right ear. These timing differences are very small (humans are sensitive to discrepancies as short as 10 microseconds, which corresponds to one degree). Once again, the auditory brain stem plays a key role in this type of sound localization. The medial superior olive contains a set of neurons called "coincidence detectors". As the name suggests, these neurons are most active when neural inputs from each ear arrive at the same time. Any given coincidence detector is innervated by two axons: one from the left ear and one from the right. However, the length of these two axons differs.

Imagine for example a coincidence detector that receives input from the left ear via a short axon and input from the right ear via a long axon. In order for this coincidence detector to be maximally excited, inputs from the left and right must arrive at the same time.

However, the difference in axonal length is such that a sound must arrive at the right ear first in order for the sound inputs from both left and right ear to arrive at the coincidence detector at the same time. If a sound maximally excites this coincidence detector, the sound must have come from the right. The activity of this coincidence detector thus signals to the brain that a sound came from the right. Naturally, the collection of coincidence detectors represents a range of sound directions from the extreme left to the extreme right. Sound localization of this sort is best for low frequency sounds, whose relatively large wavelengths exaggerate the discrepancy between both ears in terms of when sound arrives.

The brain exploits having two ears in order to localize sound. However, this works best for sounds in the horizontal plane. In order to distinguish between sounds coming from above or below, the brain relies on pinna cues. The pinna consists of several convolutions, which reflect sound in different ways, depending on the origin of the sound in the vertical plane. Differences in the sound are manifested only when the sound wave is complex (containing multiple frequencies), however. Pure tones are not reflected in different ways as a function of position in the vertical plane.

Auditory cortices. Sound information ultimately reaches the cerebral cortex. The primary auditory cortex (A1), located on the superior temporal gyrus, is tonotopically organized. "Stripes" of neurons in A1 represent low frequencies at the rostral end and high frequencies at the caudal end. This tonotopy is repeated in all A1 stripes. Some A1 stripes contain neurons that are excited by both ears and others contain neurons that are more excited by one ear than the other. In this manner, different sound directions are also represented in A1.

Two language areas of particular interest are Broca's area and Wernicke's area. Broca's area in the frontal lobe is involved in language production. Wernicke's area at the junction of the temporal and parietal lobes is involved in language perception. Lesions to these areas result in different aphasias. Broca's aphasia is marked by impaired speech production, although the context of the words is important (writing is frequently not affected in Broca's aphasia). Wernicke's aphasia is associated with impaired comprehension and the production of nonsense words.

Hearing loss. Hearing deficits can be conductive, sensorineuronal, or central. Conductive hearing loss is the result of damage to the ear drum, ossicles or basilar membrane as a result of physical trauma or very loud sounds. Infection can fill the middle ear with liquid, causing a conductive deficit.

Sensorineural deficits result from damage to the mechanisms of auditory transduction (something killing the hair cells). Loud sounds can damage the hairs on hair cells. Toxic substances can enter hair cells and kill them. Aging leads to death of microvasculature supplying the inner ear, resulting in death of hair cells. Central deficits result from damage to auditory pathways within the central nervous system. Tumours in the auditory nerve can affect hearing. Similarly, tumours in auditory pathways beyond the auditory nerve can affect hearing.

Vestibular System

The vestibular system is responsible for providing a sense of balance and body position in space. The body is able to translate in three planes (x, y and z) and rotate along three axes (again, x, y and z). The vestibular system informs the brain about motion in six degrees of freedom.

The vestibular labyrinth. Inside the inner ear lies the vestibular labyrinth, which is attached to the cochlea. The labyrinth consists of two otolith organs (the saccule and utricle) and three semicircular canals. The canals are sensitive to rotation. The otoliths are sensitive to linear acceleration and static position of the head relative to gravity. Like the cochlea, the labyrinth projects to the eighth cranial nerve.

The otolith organs. The utricle and saccule are two membranous sacs. A portion of the membrane is thickened and contains hair cells. This portion is called the macula. The hairs project into a gelatinous layer called the otolithic membrane. Calcium carbonate crystals, called otoconia, sit atop the gelatinous layer.

As in the auditory system, a mechanical disturbance excites the hair cells. When the head tilts or accelerates in one direction, the inertia of the otoconia prevents them from moving right away and the hairs bend in a direction that opposes the movement direction. The baseline firing rate of the hair cells is roughly 100 Hz. If the hairs bend towards the kinocilium, mechanically gated K+ channels open and the cell is depolarized, again like in the auditory system. Bending the hairs away from the kinocilium closes the channels and hyperpolarizes the cell. Hyperpolarization or depolarization of the hair cells decreases or increases the rate of action potentials along the afferent nerve.

Within each macula, the kinocilia are oriented in all possible directions and so hair cell activity can signal translations left, right, up, down, backward and forward. Sitting upright, the macula of the utricle lies at the bottom. It can thus sense translations left, right, forward and backward (horizontal translations). The macula of the saccule is oriented in the vertical plane. It is this sensitive to translations in the vertical plane, like movements up, down, forward and backward.

The semicircular canals. Each vestibular labyrinth (left and right) has three canals, which are orientated at right angles to one another. One canal lies in the horizontal plane. Anterior and posterior canals are angled 45 degrees from straight ahead (front-left and back-left for the left canals). The canals resemble liquid-filled rings. A swelling in each canal, called the ampulla, contains a flexible membrane called the cupula, which houses hair cells.

Rotation causes the liquid to move around. The inertia of the liquid prevents it from moving right away and the hairs in the cupula are bent in a direction that opposes movement direction. Firing of the hair cells is much like that in the otolith organs. Activity in all six canals corresponds to rotation is all possible directions. When one canal's hair cells are maximally firing, the contralateral counterpart is maximally inhibited.

The vestibular apparatus plays a key role in the vestibular ocular reflex (VOR), which maintains eye position during head rotation. Clinicians can infer the location of certain brain lesions by irrigating the labyrinth with warm or cold water, artificially inducing a VOR. Eye movements that fail to compensate for the induced sense of head spin indicate brain damage.

Major Themes/Questions

- What is the significance of the hair cell?
- Compare and contrast the auditory and vestibular systems. How do the differ? How are they similar?
- What are the benefits of having two ears?

- What sort of deficits would likely be manifested in patients with auditory brainstem lesions?
- How can lesions in the cerebral cortex affect language? Which areas in particular are critical in this regard?
- How does the auditory system approximate a Fourier transform?
- Following carolic testing of vestibular function, the patient's eyes may move. Also, the patient's eyes may not move at all, or only one eye may move. Which behaviour(s) indicate brain damage?