This article was originally published in the *Encyclopedia of Neuroscience* published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

http://www.elsevier.com/locate/permissionusematerial

Auditory guided behavior depends on the ability of the nervous system to construct an accurate representation of the sounds in the environment. Toward this end, there are two primary functions of the auditory system. The first function is to determine what it was in the environment that produced a sound. The second function is to determine where in space the sound occurred. Knowledge of both the spatial position of a sound’s source and what produced it facilitates the initiation of an appropriate behavioral response. For example, a predator like the cat might move toward the sound produced by a mouse rustling in the leaves while the mouse might freeze or move away from the sounds of the approaching cat.

The ability to locate sensory stimuli is not unique to the auditory system as it is also shared by the visual and somatosensory systems. The locations of visual and tactile objects are encoded directly in these systems due to the topographic, spatial organization of the receptors, the rods and cones of the retina, and the mechanoreceptors of the skin, respectively. In contrast, in the auditory system, the hair cells of the cochlea are designed to encode sound frequency and intensity and cannot sense sound location directly. Therefore, sound location must be computed centrally in the auditory system based on the neural representations of the spectral and temporal characteristics of the acoustic stimuli arriving at the two ears.

In order to understand the neural mechanisms of sound location we first have to identify the acoustic information bearing elements of the sounds arriving at the two ears that specify location. There are three primary acoustical characteristics of sounds, or cues, which are used to determine the spatial position of a sound source. These acoustical cues are physically different because they are created by different ways that the propagating sound waveforms interact with the head and external ear. The fact that there are three distinct cues naturally suggests that three different kinds of neural mechanisms are needed to encode them. Indeed, the cues are encoded in anatomically separate, but parallel, pathways in the ascending auditory system. The evidence that each cue is encoded in a different pathway comes from experimental data from disparate scientific disciplines including neurophysiology, behavioral psychology, and comparative neuroanatomy.

### Three Main Acoustical Cues to Sound Source Location

The acoustical cues for sound localization are generated by the interaction of the propagating sound waves with the anatomical structures supporting the receptors of the ear, namely the head and the external ear (or pinnae), and not by the topographic organization of the receptors themselves. Examples of the three cues are shown in Figure 1. Two of the cues, interaural time delays (ITDs) and interaural level differences (ILDs), result directly from the fact that the two ears themselves are sampling the sound waves at different locations due to their positions on opposite sides of the head. These two cues specify primarily sound location in azimuth. The third cue, spectral-shape, results from the way that the sound waves are distorted or filtered by the complicated convolutions and folds of the pinna. Spectral shape cues specify primarily sound location in elevation, and whether a sound is in the front of or behind the observer. Interpretation of the roles of the various auditory nuclei that process the sound localization cues requires knowledge of what the cues are and how those cues change with changes in source location.

#### Interaural Time Delays

ITDs result because the ears are physically separated in space by the head. Therefore, the direction-dependent differences in path lengths that sound must travel to reach each ear from the source will generate different times of arrival of the sound at the two ears, or ITDs (Figure 1(b)). For example, a sound on the right side will arrive at the right ear a short time before it arrives at the left ear, while a sound from directly ahead will arrive at the ears at the same time. The range of ITDs is directly proportional to head diameter so species with larger heads, hence larger distances between the two ears, exhibit a larger range of ITDs. For example, in cats, which have head diameters of 6–7 cm, ITDs increase monotonically with azimuth along the horizontal plane up to a maximum of 350–400 μs (Figure 1(d)). The diameter of the human head is nearly twice that of the cat, on the order of 15 cm. Consequently, the maximum ITD in humans is ~800 μs. Yet smaller mammals, like some species of bats and rodents, have head diameters ~1 cm, resulting in maximum ITDs as small as 40 μs. Such small ITDs are not thought to be useful for accurate sound localization.

#### Interaural Level Differences

The fact that the two ears are separated by an obstacle, the head, results in the ILD cue to sound location.
For sounds of high frequency (or short wavelength), the head essentially creates an acoustic shadow for the far ear as sounds with wavelengths on the order of the diameter of the head and smaller are reflected off the near side of the head. Consequently, the resulting sound arriving at the ear farthest from the source is effectively attenuated thereby creating direction-dependent differences in the amplitudes, or levels, of the sounds that reach the two ears (Figure 1(c)). The sound level is greatest at the ear nearest the source, and lower in the ear farthest from the source. The ILD cue is a rather complicated function of both sound location and, at any one location, sound frequency (Figure 1(e)). As expected, ILDs are small in magnitude for low-frequency sounds and increase in magnitude for high-frequency sounds (Figure 1(e)). Therefore, ILDs are primarily useful for localization of high frequencies. ILDs that are large enough to be useful for sound location are shown for the cat, but the same three cues are available in all species, although the magnitude of the cues will be different. (a) A broadband transient sound is presented from a loudspeaker at (40°, 0°). (b) The resulting acoustical responses to the transient near the eardrums in the left and right ears. The difference in the onset times of the sounds at each ear yields the interaural time difference (ITD) cue and difference in amplitudes reflect the interaural level difference (ILD) cue. (c) The frequency spectrum of the acoustical responses in (b). The spectral shape cues are captured by the changes in the shapes of the spectra as a function of source location. The ILD cue is the difference in sound level computed at each frequency. (d) The ITD plotted as a function of sound source azimuth. ITDs are minimal near the midline and maximal for sources to either side. (e) The joint spatial and frequency dependence of ILDs for sources varying along the horizontal plane. ILDs are a complicated function of azimuth and frequency for high stimulus frequencies. For illustrative purposes, sources were restricted to the horizontal plane and only in the frontal hemisphere for frequencies between 1 and 25 kHz.

**Figure 1** The acoustical cues to sound location. In the example, the cues are shown for the cat, but the same three cues are available in all species, although the magnitude of the cues will be different. (a) A broadband transient sound is presented from a loudspeaker at (40°, 0°). (b) The resulting acoustical responses to the transient near the eardrums in the left and right ears. The difference in the onset times of the sounds at each ear yields the interaural time difference (ITD) cue and difference in amplitudes reflect the interaural level difference (ILD) cue. (c) The frequency spectrum of the acoustical responses in (b). The spectral shape cues are captured by the changes in the shapes of the spectra as a function of source location. The ILD cue is the difference in sound level computed at each frequency. (d) The ITD plotted as a function of sound source azimuth. ITDs are minimal near the midline and maximal for sources to either side. (e) The joint spatial and frequency dependence of ILDs for sources varying along the horizontal plane. ILDs are a complicated function of azimuth and frequency for high stimulus frequencies. For illustrative purposes, sources were restricted to the horizontal plane and only in the frontal hemisphere for frequencies between 1 and 25 kHz.
Monaural Spectral Shape

The ITD and ILD cues are not useful for sound localization in elevation because their values change little with variations in source elevation. The monaural spectral shape cues, however, do change systematically with source elevation. Spectral shape cues arise from direction- and frequency-dependent reflection and diffraction of the pressure waveforms of sounds by the head, torso, and pinna that result in broadband spectral patterns, or shapes, that change with location. The most prominent direction-dependent features in the spectra are the deep notches that occur at some locations (e.g., 5.5 and 9.5 kHz for the right and left ear, respectively, in Figure 1(c)). These spectral notch cues are used both for localization in elevation as well as facilitating the distinction between sound sources in front from sound sources behind of an observer. Spectral cues are created primarily for high-frequency sounds whose wavelengths are on the order of the dimensions of the pinna and its convolutions. The frequency at which the spectral notch occurs varies systematically with sound source elevation, with lower elevations yielding low-frequency notches and higher elevations yielding progressively high-frequency notches.

Neural Processing of Localization Cues in Three Parallel Pathways

For these three acoustical cues to location to be useful for spatial perception, there must be neural mechanisms in the ascending auditory pathways that effectively encodes each of them. Each of the three different cues is encoded in different areas of the auditory brain stem (Figures 2–4). That is, there are three parallel pathways through the brain stem that appear to encode separately the ITD, ILD, and spectral cues. Neurons in these pathways are the earliest places in the ascending auditory pathways to respond selectively to each of the three cues.

ITDs Are Processed in the Medial Superior Olive

The neurons comprising the medial superior olive (MSO) receive excitatory inputs from both ears via the bushy cells of the anteroventral cochlear nucleus (AVCN) on both sides (Figure 2(a)). The bushy cells of both sides receive excitatory inputs from the cochlea via the auditory nerve fibers (ANFs) via specialized synaptic terminals called Endbulbs of Held. The AVCN is organized tonotopically. Most MSO neurons are sensitive to ITDs in sounds presented to the two ears. ITD sensitivity in MSO results from a three-stage process as first hypothesized by a model proposed by Jeffress. First, the afferent inputs to MSO neurons from the ANFs and from the bushy cells of the AVCN carry timing information in the form of phase-locked neural responses. Phase locking refers to the tendency of a neuron to fire action potentials at particular phases of an ongoing periodic sound waveform, such as the sinusoidal waveforms that are typically used in physiological studies of the auditory system. Phase locking is the mechanism by which the peripheral auditory system keeps track of the times of occurrence of the ongoing amplitude fluctuations in sounds. Second, the MSO neurons behave like coincidence detectors, responding maximally only when action potentials from the SBC inputs from the left ear and the right ear arrive nearly simultaneously at the MSO neuron. Third, the axons of the afferent inputs to MSO from the bushy cells form delay lines along one axis of the nucleus due to differences in the neural path length from the bushy cells to the MSO neuron. Differences in neural path length result in differences in neural conduction times to the MSO which are then offset by a physical ITD cue. Since the projections to the contralateral MSO have a naturally longer path length than those to the ipsilateral MSO, most MSO neurons primarily encode ITDs produced by sounds on the contralateral side where the ITD can compensate for the longer conduction time from the contralateral ear (Figure 2(b)). In addition, different MSO neurons have different length delay-line inputs, so different MSO neurons will be maximally sensitive to different ITD values. Moreover, MSO neurons have a bias toward low-frequency sounds, as most MSO neurons are sensitive to predominantly low sound frequencies. Phase locking is best for neurons sensitive to low-frequency sounds. Behaviorally, ITDs are most useful for localization for low-frequency sounds. Most MSO neurons send excitatory afferents to the ipsilateral dorsal nucleus of the lateral lemniscus (DNLL) and the ipsilateral central nucleus of the inferior colliculus (ICC). As a consequence, low-frequency sensitive neurons in the DNLL and IC encode the ITDs of sounds primarily on the contralateral side.

In recent years the Jeffress model has been challenged on several grounds, chiefly by the finding that the MSO also receives inhibitory input from both sides. Current research efforts are aimed at understanding the role of this inhibition and what modifications to the simple model are needed.
ILDs Are Processed in the Lateral Superior Olive

Lateral superior olive (LSO) neurons also receive afferent inputs from both ears (Figure 3(a)). The input from the ipsilateral ear is conveyed via auditory nerve fibers to the bushy cells of the AVCN, which then send an excitatory projection to the ipsilateral LSO. The afferent input from the contralateral ear also comes from bushy cells of the AVCN but in this case the bushy cells project across the midline to the neurons of the medial nucleus of the trapezoid body (MNTB). The bushy cell synapse onto MNTB neurons forms the massive calyx of Held; the large size of this unique synapse results in very fast and secure synaptic transmission. MNTB neurons are glycinergic, so their projection to the LSO has an inhibitory effect. In contrast to the MSO, both the MNTB and the LSO are biased predominantly to neurons that are sensitive to high sound frequencies where the ILD cues themselves are prominent. Because of the interplay of the ipsilateral excitation from the bushy cells and contralateral inhibition from the MNTB, LSO neurons essentially compute the difference between the neural representation of the levels of the sounds present at the two ears. Hence, LSO neurons encode ILDs. The sound level at each ear is believed to be encoded by a rate code, in that the number of action potentials elicited over a given time period is proportional to the level of the sound at the ear. Because LSO neurons are excited by sounds at the ipsilateral ear but inhibited by sounds at the contralateral ear, they will respond best to ipsilateral sound sources that produce ILDs favoring the excitatory ear (Figure 3(b)). LSO neurons project bilaterally to the DNLL and the ICC. In order to achieve the contralateral...
representation of space that is a basic feature of sensory and motor areas of the brain, the high-frequency sensitive neurons of the ventromedial LSO send excitatory projections to the contralateral DNLL and ICC, while the low-frequency sensitive dorsolateral LSO sends predominantly inhibitory projections to the ipsilateral DNLL and ICC. As a result high-frequency sensitive neurons in the DNLL and ICC respond best to contralateral sound sources that produce ILDs favoring the contralateral ear.

**Spectral Shape Cues Are Processed in the Dorsal Cochlear Nucleus**

Less secure is our understanding of the processing of spectral shape cues. It is hypothesized that the DCN analyzes the spectral shape cues created by the interaction of the sound with the external ear in a direction-dependent fashion. The principal output neurons of the DCN are the fusiform and giant cells (Figure 4). Neural responses of the fusiform and giant cells are similar, so they are often considered as a single neural group. As a neural circuit, the DCN is more complicated than the MSO and LSO/MNTB circuits. As such, it is easier to break down the DCN circuit into a simple qualitative model (Figure 4). In this general model, the type IV neurons, the fusiform and giant cells, represent the principal output neurons of the DCN that project to the contralateral ICC. The type IV neurons are selective for the broadband spectral notches due to the way they integrate excitatory and inhibitory inputs from a variety of sources. The main excitatory inputs arise from the ANFs. Due to the ANF inputs, the type IV neurons are tonotopically organized and are predominantly sensitive to high sound frequencies. Type IV neurons also receive inhibitory inputs from two different neurons. First, the type II neurons, corresponding to the tubuloverentral neurons of the DCN, receive an excitatory input.
from the auditory nerve and send a strong inhibitory input to the type IV neurons. Type II neurons are sensitive to a narrow range of frequencies that is similar to that of the type IV neurons. Type II neurons therefore effectively inhibit type IV responses to narrowband sounds. On the other hand, the wide-band inhibitor neuron, or W neuron corresponding to the D-Stellate neurons of the posteroventral cochlear nucleus (PVCN), receives excitatory inputs from many ANFs spanning a broad range of frequency sensitivity. W neurons in turn project a strong glycinergic inhibitory input to the type II neurons but a much weaker inhibitory input to the type IV neurons. Consequently, W neurons respond best to broadband sounds, which causes a strong inhibition of the type II neurons. Therefore, broadband sounds effectively release the strong inhibition on the type IV neurons from the type II. This complicated inhibitory network from the type II and W neurons, along with a single narrow frequency-tuned excitatory input from the auditory
nerve, allows type IV neurons to respond selectively to broadband spectral notches. When the frequency of the spectral notch cue to sound elevation is aligned with the best frequency of a type IV neuron, that neuron’s response is maximally inhibited. The type IV neurons of the DCN send an excitatory projection to the contralateral ICC.

Further Evidence for Processing of Localization Cues in Three Parallel Pathways

The physiological studies reviewed above showed that there are three different nuclei in the auditory brain stem, the MSO, LSO, and DCN, that selectively process each of the three cues to sound location. In support of those studies, there is also evidence from other scientific disciplines that points more to the functional aspects of the encoding of the ITD, ILD, and spectral cues by those three circuits.

Comparative neuroanatomical studies in a large number of mammalian species have found that animals with large heads (like primates) tend to have a large and well-developed MSO, whereas the MNTB and LSO are relatively small and underdeveloped. In contrast, animals with small heads (like bats and most rodents) tend to have a small underdeveloped (in some cases nonexistent) MSO while the MNTB and LSO are large and well developed. Behaviorally, small-headed animals are not able to localize low-frequency sounds based on ITDs, likely due to the fact that the ITDs are quite small and the MSO, which processes ITDs, is also relatively small. On the other hand, mammals with large heads can localize accurately low-frequency sounds based on ITDs because the ITD cues are quite large and the MSO is also well developed. In contrast, small mammals can localize very high-frequency sounds presumably due to the fact that high-frequency sounds yield useful ILD cues to location and the MNTB and LSO which processes ILD cues are well developed. Small mammals tend to be able to hear higher frequencies than large mammals. However, large mammals can also localize high-frequency sounds. The data together implicate the MSO as necessary for encoding ITDs and the MNTB and LSO necessary for encoding ILDs.

Finally, the ability to localize a sound positioned along the horizontal plane, where ITD and ILD cues are predominantly used, is severely impaired in experimental animals where the input or output pathways to the MSO and LSO or the MSO and LSO themselves were selectively lesioned. These experiments show that the interaction of the afferent inputs to the MSO and LSO are necessary for the localization of sound based on ITDs and ILDs. Moreover, in animals where the DCN was selectively lesioned, sound localization ability was impaired chiefly for sources that varied in elevation while localization was nearly normal for sources varying in azimuth.

There are three primary acoustical cues to the spatial location of a sound source. Two cues, ITD and ILD, depend on the fact that the two ears are physically separated in space by the head, which acts as an acoustical barrier. ITD and ILD cues vary maximally for sources along the horizontal plane. Neurons in the MSO and LSO encode the ITD and ILD cues, respectively. Data from comparative as well as lesion studies support the separate processing of those two cues. Spectral shape cues are used for localizing sources varying in elevation. Spectral shape cues are encoded by neurons in the DCN. Lesion studies reveal that the DCN is critical for sound localization in elevation. The MSO, LSO, and DCN send projections to the ICC in such a way that neurons in the ICC encode predominantly sounds in the contralateral field. Neurons at higher auditory stations, including the medial geniculate body in the thalamus and the primary auditory cortex, have also been found to be sensitive to the three acoustical cues to location. However, the manner in which these cues are represented systematically in these higher-order areas as well as how sound source location itself is represented in the auditory system remains elusive.

See also: Auditory Cortex Structure and Circuitry; Auditory Scene Analysis; Auditory System: Central Pathways; Auditory Localization; Cochlear Development.

Further Reading