Edward Hess
Term Paper
Psychology 6014A

Auditory Localization

Introduction

Localization in relation to the auditory system refers to the ability to determine the direction and distance of a sound source. Similar to the vision system, the auditory system uses a variety of cues to provide estimations of the position of a sound source, and these cues are often used in conjunction with each other to provide enhanced accuracy. This paper provides a survey of the current understanding of auditory localization. Additionally, a comparison of auditory and visual localization cues will be included, as well as a discussion of how auditory localization may be improved (by looking at examples from the animal kingdom), and how the auditory localization system complements the visual localization system.

What are the cues that we use?

It is convenient to break down auditory localization cues into the following three categories;

i) Binaural cues

ii) Monaural cues

iii) Distance cues

iv) Miscellaneous effects

The Binaural Cues

The important early work in auditory localization revolved around the study of binaural cues (Middlebrooks 1991). It is probably safe to say that Lord Rayleigh, one of the pioneers in this field, felt that humans had two ears for more than just cosmetic reasons. Rayleigh proposed that humans could get localization information in two ways. They could sense intensity differences in a sound source between the two ears caused by the deflection of sound waves traveling to the ear opposite the sound source. Rayleigh also believed that for any sound source originating from a point that wasn't equidistant from both ears, a phase difference would exist, and this phase difference would provide localization information. These two cues are dubbed Interaural Intensity Differences (IIDs), and Interaural Phase Differences (IPD) respectively (see Figure 1).

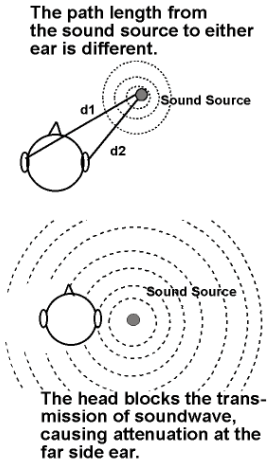

Figure1

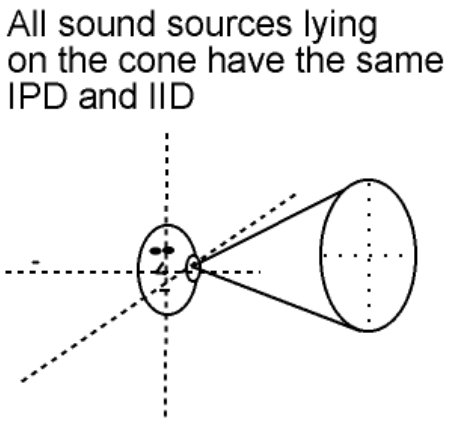

Figure 2

Binaural Cue #1: Interaural Intensity Differences

The Interaural Intensity Difference, as stated previously, is created because the head locks the transmission of sound waves. This mechanism, however, is useful only in certain situations. Sound waves of sufficiently long wavelength (i.e. below approximately 1000 Hz) will travel around the head; therefore a significant difference in intensity between the ears will not be created in these situations (Middlebrooks, 1991). This mechanism also requires that the head differentially blocks the path from the sound source to one ear; in a situation where the sound source is of equal distance to either ear, little intensity difference will be created. This means that IID information is most useful for judgement of azimuth (i.e. horizontal angle), because the horizontal plane has a broad sweep of angles where transmission of sound waves from a sound source to one ear is blocked. This cue is much less useful for determination of elevation.

Binaural Cue #2: Interaural Phase Differences

Lord Rayleigh supported his assertion that phase information was important for determination of location with an experiment using slightly mismatched tuning forks. The tuning forks would create two similar tones (i.e. pitch and amplitude) with continuously varying phase relationship. The perceived location of this "sound source" appeared to oscillate from ear to ear providing confirmation of the importance of phase difference cues. As with the to the IID cues, these phase difference cues are useful under certain circumstances. These cues are most effective at low frequencies (i.e. approximately 1400 Hz and below), and are most effective in providing information about the azimuth of the sound source.

Additional Observations Regarding the Binaural Cues

Lord Rayleigh's theory regarding binaural cues is known as the "Duplex Theory". The binaural localization cues are complementary to each other in the sense that they act over different but somewhat overlapping frequency ranges. It is interesting to note that the reason that neither cue works over the same range is due to the physical properties of both sound waves and the physical dimensions of the head. The IID cues rely on a significant difference in sound level entering each ear, and this is due to the fact that (high frequency) waves cannot bypass the head easily. In the IPD cues it is more desirable to have equal intensity signals from a given sound source at both ears, hence sound waves that are more able to bypass the dimensions of the head are more useful.

It has also been found that localization using these two cues is not very reliable in the frequency range from 1500-3000 Hz. In this range, neither of the binaural cues is in its optimum operating range. Additionally, there are situations where identical IIDs and IPDs may be produced by a given sound source at different locations. Known as the "Cone of Confusion" (see Figure 2), sound sources lying on the surface of a cone projecting from either ear (along the axis of the ear canal) produce these identical intensity and phase differences at the opposite ear (Mills 1972). The Cone of Confusion model makes some theoretical assumptions (i.e. a spherical head and symmetrical ear canals) but front-back and top-bottom confusions have been reported in the literature, providing support for this idea.

The Duplex Theory, although a significant step forward in the understanding of auditory localization, provided an incomplete picture (Duda, 2000). Some of the major shortcomings are listed below:

i) It does not address vertical localization to any appreciable extent.

ii) It fails to account for how the auditory system deals with environmental effects such as echoes and "room modes" (areas in a room where constructive and destructive interference modify the sound distribution significantly.

iii) It does not address discrimination of front and back sound sources.

Monaural Cues

Information about the elevation of a sound source comes primarily from monaural cues. As indicated by the name, monaural cues are available using input into only one ear. These cues are provided by the interaction of the outer ears (known as the pinnae) with incoming soundwaves.

What Do the Pinnae Do?

The pinnae act to transform the complex waveform containing a broad spectrum of frequencies into a different waveform prior to it reaching the eardrum. This is due to constructive and destructive interference created by the reflection of the wave front as it reaches and bounces around the interior of the pinnae. The modifications made by the pinnae result in changes in the intensity of the different frequency components that make up the incoming sound. The brain is able to perform a "spectral analysis" (on the

intensity and frequency relationships) and gain positional information because the pinnae will predictably modify the waveform depending on the location of the sound source.

The pinnae produce the most dramatic changes in intensity as a sound source is moved from the horizontal plane (at a height equal to the ears) vertically. The greatest intensity changes caused by the pinnae occur at a band located at approximately 10,000 Hz. This is known as the "Pinna Notch". The pinnae also provide information related to azimuth as well, although intensity changes do not vary as much as for horizontal changes as they do for vertical changes.

It is important to note that these monaural cues rely on a change in a spectrum of frequencies. Pure tones provide little "spectrum" to work with and will foil-pinnae related cues. In experiments related to measurement of pinnae effects, broadband noise (e.g. "white noise") is typically used as a sound source.

Head Related Transfer Functions

The pinnae provide a very rich source of auditory information. A large amount of work has been devoted to measuring how the pinnae transform sound. This work typically involves measuring Head Related Transfer Functions (HTRFs). Head Related Transfer Functions are functions of four variables (three related to position and one related to frequency) that describe the change in intensity (at the eardrum) caused by a pinna for a given frequency sound wave arriving at the outer ear.

One of the things that make HRTFs interesting is that they may be used to modify a recorded sound source to give it positional information. Given a broadband sound source, tailoring the relative intensity of the frequency components of the sound source can give it a very palpable "location" even when presented to the listener via speakers located at a position far away from the perceived location. The mathematical process of modifying the digital data that makes up a recorded sound using the HRTFs is called convolution. Application of HRTFs is currently seen in commercial products such as sound cards (ESS Technology, 2000) and software.

Pinnae vary significantly from person to person, and hence the way they transform sound varies between individuals as well. Head Related Transfer Functions, as one might expect given this fact, also vary from person to person. They may be considered an "earprint" of sorts. In experiments testing the ability of measured HRTFs to provide accurate positional information, it has been noted that individuals were most successful in "locating" the source of a sound when it was modified using their own HRTFs, i.e. HRTFs recorded using microphones placed in that subject's ears. Non individualized HRTFs (i.e. an HRTF from a representative subject) produced significantly more cases of front to back and up-down confusion than localization experiments using the individually measured transfer functions (Wentzel, 1993).

Distance Related Cues

Monaural and binaural cues provide the listener information about the direction of a sound source. To actually localize a sound source a third component, distance, is required. There are a variety of distance cues, and the main ones are:

i) Sound intensity

ii) Frequency

iii) Movement parallax

iv) Reverberation

Sound Intensity

This cue operates on the basis that the further away a sound source, the lower the intensity of the sound reaching the listener is. In general, the intensity of the sound drops off with the square of the distance because sound from a theoretical "point source" is emitted in a spherical pattern. It is important to note that the surroundings of the listener and sound source can have a significant impact on the intensity of sound. A given room can produce nodes (areas where certain frequency sounds are intensified or attenuated due to the physics of waveform addition). It is also relevant to note that the listener must have an external intensity reference to make inferences about distance based on loudness. Listening to a sound that they are already accustomed to (e.g. somebody talking at a normal conversational level) will provide better localization results than a sound that they are unfamiliar with (Blauert, 1997).

Frequency

The distance of a sound source may also be indicated by the relative frequency distribution of the sound source. As sound travels through the atmosphere and walls, high frequency sounds are differentially absorbed, and a preponderance of low frequencies reach the listener. This information gives some indication of distance. This effect may be observed daily on the streets of any large city; whenever someone drives by with a high-power stereo system playing loud music, generally very little of the high frequency portion of the music extends beyond the cabin. Only the bass notes get through.

Movement Parallax

Movement parallax is the effect that nearer sound sources that are moving shift their perceived position faster than moving sound sources that are further away. This effect is used by filmmakers to give an aural sensation of movement. For example, when a helicopter flies from one side of the screen to the other, the sound of the helicopter blades follows along with the helicopter by panning the sound from the speakers on one side to the speakers on the other side. A helicopter further off in the distance would take longer for the sound to pan from one side to the other.

Reverberation

The idea that reverberation is useful as a distance cue is based on the observation that further away sounds often have a greater proportion of reflected sound. The brain is able to identify reflected sound and hence use it as a distance cue. For sound to be reflected,

however, there must be a surface for the sound waves to be reflected off of, such as the walls of a room. Furthermore, the materials these surfaces are made of have a significant impact on the amount of reflection that occurs. Placing "soft" materials along the walls, such as fabric panels may deaden a "live" (i.e. reverberant) room. Reverberation is very much tied to the qualities of the room, and this fact must impact the utility of reverberation as a distance cue.

More Comments on Distance Cues

The four distance cues commented on here may supplement each other in providing distance information, because they can occur simultaneously. None of these cues are particularly reliable, however, because they all require some type of tenuous reference point. For example, the use of loudness as a distance cue requires that the listener has an understanding of the loudness of the sound source at a given distance, and then must recall that perceived loudness as the sound source changes position. Given that the sound source may change intensity, and that there is no persistent reference in this situation, distance estimation could be extremely variable. It would seem that absolute judgements using such a cue would be difficult; relative changes in distance would probably be more successful using these cues. Additionally, these cues may vary given the context surrounding the listener and the sound source, as explained in the loudness and reverberation. Of the three types of localization cues discussed in this paper, research is the sketchiest in the area of distance localization.

The placement of ears on either side of the head trades accuracy of localization for breadth of coverage. In most mammalian prey species (e.g. rabbits and other rodents), the eyes are placed on opposite sides of the head for the sake of giving the animal a very wide angle of view. Having ears on either side of the head instead of both being directed forward allow a person to detect sounds from all around, alerting that person to the presence of many sound sources out of the field of view. In these situations, the ears provide guidance for the eyes, giving the listener a somewhat rough cue as to the presence of an entity, so a more exacting localization with the eyes may be conducted.

Miscellaneous Localization Cues – The Precedence Effect

There is a phenomenon known as the Precedence Effect, the "Law of the First Wavefront", or the Haas Effect which affects perception of direction (Duda, 2000). This effect is that given a group of equivalent sound sources (e.g. a set of stereo speakers playing the same tone) presented in front of a listener, the sound source that is heard first will be the perceived source of the sound. If both sound sources are heard at the same time, then the sound will be localized at a point halfway between the two sound sources. This effect can be relatively dramatic; a sound source of 8dB lower intensity may still be taken as the location of the sound if it is the first sound heard.

It is not obvious why the Precedence Effect should exist. It would appear that the ability to suppress reverberations is useful in comprehending the complex signal that appears at the ears. In reverberant environments, an inability to suppress indirect sound could make localization of the sound difficult because there would be many perceived sources of the

sound. The shortest path to the ear is the correct direction to localize to; not the myriad other reflected "sources". Fortunately the most direct path is also the quickest route, hence the basis for the effect. A theoretical model of the Precedence Effect has been used successfully in the sound localization system of a robot to help it cope with operation in reverberant surroundings (Huang et. al., 1997).

The Clifton Effect

A very interesting observation was made by Clifton (1987) regarding the human auditory system's capability to adapt to environments that produce a lot of reflected sound. Clifton demonstrated that the suppression of echoes (as described in the Precedence Effect section, above) was a dynamic process. Subjects who were exposed to audible "clicks", one from a speaker to the left of a subject and then one from the right of the subject, that were closely spaced in time. Initially, these clicks were detected as individual events, but after a short period (within a second), the second click was no longer perceived as a separate click but rather contributed to the "spaciousness" of the initial click. Essentially, the auditory system had come to the conclusion that the other clicks should be considered as an ambient quality that should be associated with the initial click. Reversing the signals to the speakers reestablished this pattern of separate clicks that eventually "fused" into a single click that held ambiance information.

This phenomenon indicated that the brain dynamically adjusts to the surrounding auditory environment. Mechanisms in the auditory system exist for sophisticated analysis of the environment. The brain can adjust to the environment and suppress (the

Precedence Effect) and/or synthesize (the Clifton Effect) information to provide better comprehension capabilities.

How Accurate is Auditory Localization?

It is fairly obvious that auditory localization is fairly crude relative to visual localization. Measurements of auditory localization, and of "localization blur" (i.e. the calculation of JNDs for change of position of a sound source) have been undertaken since the 1920's. Results have been variable, but the lower limit of reported values (Blauert, 1997) for localization blur is approximately 1°, which tends towards larger values at higher frequencies.

It is also important to note that the resolution of auditory localization varies depending upon location. It is not equivalent throughout the median and horizontal plains; some arcs are better resolved than others. In the horizontal plane, sounds coming from directly in front of the listener. Localization has the best resolution directly to the front of the listener, slightly less resolved immediately behind the listener, and directly to either side is less resolved than either the front or back conditions (Blauert, 1997).

A Comparison of the Auditory and Visual Localization Mechanisms

To make comparisons between the two systems, it is useful to first consider what visual objects are, and what sonic "objects" are. Objects in the visual world are rocks, trees, houses, birds, people; physical entities that have some cohesiveness. Objects in the sonic

world, one might assert, are sounds that may be associated with a physical entity; the song of a bird, the crash of a cymbal, the chirp of crickets, or the rustling of leaves.

The process of visual localization of objects relies on high-resolution detectors (the retinae) that are mounted on a swiveling head. The head is guided by object recognition, peripheral vision, and swiveling eyeballs to perform a search for an object of interest. Localization is achieved once the head has been swiveled and inclined (and perhaps body position adjusted as well) sufficiently to place the image of the object of interest on the highest resolution area of the retina. Localization in terms of elevation and azimuth has been completed at this point, and precision is orders of magnitude better than in the auditory system (Blauert, 2000).

The auditory system, however, gains little in terms of localization information by swiveling the head. The ears do not function like parabolic microphones (at least significantly), homing in on sonic objects by scanning and determining the direction that produces the greatest intensity signal of the sound of interest. The auditory system relies on the physical interaction of the medium through which sound is transmitted with the body of the listener to gain the majority of the information it gets related to sound source location. Additionally, the auditory system relies on wide separation (relative to the eyes) of the ears to gain information from the phase differences caused by different path lengths to each ear. The closer together the ears are, the smaller the usable frequency range is for detecting phase difference cues.

It is also interesting to compare the nature of the "input devices" themselves. Generally, people have an equivalent number of eyes and ears. The eyes, however, are actually a collection of millions of transducers that all may be directly influenced by the external world. These high-resolution sensors afford several distance cues that have no analogs in the auditory world, such as occlusion, accretion, and deletion. The auditory system is merely composed of two inputs; there are only two eardrums exposed to the outside world. A compound mechanism like the retina would not be useful in the auditory world because of the summative nature of the medium of transmission. The auditory mechanism must decompose the "sound wave soup" arriving at the eardrums to determine the sonic objects that are present, and what the location of these objects is.

It has been established that auditory localization and visual localization function through very different means, but are there some similarities? Obviously, analogies drawn between visual and auditory mechanisms will be a bit of a stretch, but such comparisons can be interesting nonetheless. The most direct analogy that may be drawn between visual and auditory localization cues is that of motion parallax. As previously described, close objects move more rapidly across the auditory field than far objects. Likewise, near objects move across the visual field more quickly that objects that are further away.

Consider the auditory distance cue related to frequency (see section above related to auditory distance cues). Its analog in the visual sense is that of the visual distance cue of atmospheric perspective. Both of these cues relate to distance and are caused by atmospheric interference. They both result in distortion of the original signal that results

in a less sharp image (relative to their respective domains) at further distance than at closer distances.

It is difficult to find many other direct analogies, but some indirect ones exist. Both the visual and auditory localization cues rely on direct comparison of signals reaching each eye/ear to obtain directional information. The directional information they acquire, however, is different. As discussed previously, the binaural cues give information related to azimuth, whereas the binocular cues relate primarily to depth.

How could auditory localization be improved?

Not that it needs to be, but it is entertaining to think about the ways in which human auditory localization could be improved. The visual system has some distinct advantages. Visual information is generally persistent unless someone shuts off the lights. Visual information is constantly being beamed from various objects to the eyes. This is useful in making relative assessments, such as distance estimations.

A way to improve the ability of individuals to estimate distances would be to create "sonic sunshine" that could give continuous feedback to the auditory system. There is anecdotal evidence that blind individuals are able to do this to a limited extent by listening to faint echoes from the taps of their canes. This is known as echolocation and it has been a trick that bats and porpoises have been using for years. Bats emit a series of high intensity, high frequency (40-100,000 Hz) sound pulses and listen for the echoes.

This setup allows bats to fly in complete darkness with ease. Porpoises use a series of clicks to achieve their localization goals.

In general, to find improvements (or at least useful specialization's) related to a given sensory input, one merely has to look around the animal kingdom. Some species has already developed a highly refined version. Swiveling pinnae, bigger pinnae, different frequency sensitivities, it has all been tried. The aforementioned bats often have relatively large pinnae spaced closely together on the front of their head, probably to enhance their ability to echolocate (and audio-locate in general).

How Auditory Localization Complements Visual Localization

Auditory localization is rather crude when compared to visual localization every category related to spatial resolution (Blauert 2000). The auditory system, however, complements the visual localization system nicely. This is true because the auditory system has the following properties:

i) It operates well in the near or total darkness

ii) It is panoramic

iii) It isn't subject to occlusion

The fact that the auditory system relies on variations in air-pressure means that it is available in all lighting conditions (unless someone is in the middle of a vacuum, in which case auditory localization may be the least of their worries). The fact that it is panoramic is due to the positioning of the ears on opposite sides of the head. Occlusion

doesn't occur (at least not nearly as dramatically) like in the visual system. These attributes allow the auditory system to handle a number of situations or perhaps help guide the visual system.

Conclusions

It appears that the cues and mechanisms that are used to perform auditory localization are fairly well understood. Binaural cues, pinnae effects, and the Precedence Effect are well characterized and give a fairly complete picture of how humans perform localization. Currently it is possible to recreate, quite successfully, "synthetic" localization cues using HRTFs, demonstrating that there is a fairly complete understanding of the mechanism related to auditory localization. It does appear that the neurophysiology of auditory localization, however, isn't particularly well understood. Many recent advances (e.g. Keller, 1998; Konishi, M. 1999) have been made, but a complete understanding of the neural mechanisms involved in auditory localization is not currently available. This is most likely the area where the largest advances in understanding will be made in the future.

References

Blauert J., (1997). *Spatial Hearing: The Psychophisics of Human Sound Localization*, (The MIT Press).

Clifton, R. K. (1987). "Breakdown of Echo Suppression in the Precedence Effect," J. Acoust. Soc. Am., Vol. 82, pp. 1834-1835.

Duda R. (2000). "Duda Research: Duplex Problems," http://www.best.com/~duda/Duda.R.B.7.html

Middlebrooks J. C., and D.M. Green, (1991). "Sound Localization by Human Listeners," Annu. Rev. Psychol. 42:135-59.

ESS Technology, 2000. "Canyon3D Technology Backgrounder," http://www.canyon3d.com/techback.html

Goldstein B., (1999). *Sensation and Perception.* (Brooks/Cole Publishing Company).

Huang J., Supaongprapa, T., Terakura I, Wang F., Ohnishi N.,and N. Sugie, "A Model-based Sound Localization system and its Application to Robot Navigation," Robotics and Autonomous Systems 27, 199-209.

Keller C. H., Hartung, K., Takahashi T. T., (1998) "Head-related Transfer Functions of the Barn Owl: Measurements and Neural Responses," Hearing Research 118: 13-34.

Konishi M., 2000. "Study of Sound Localization by Owls and its Relevance to Humans," Comparative Biochemistry and Physiology Part A 126: 459-469.

Wenzel E. M., Arruda M., Kistler D., Wightman F. (1993). "Localization Using Non-individualized Head-related Transfer Functions," J. Acoust. Soc. Am. 94 (1).

Wessels N.K., Hopson J. L., (1988). *Biology.* Random House. pp. 942-3.