DAT335 – Music Perception and Cognition
Cogswell Polytechnical College
Spring 2009

Week 7 – Class Notes

Space Perception

Introduction

Localization: refers to judgments of the direction and distance of a sound source.
Lateralization: describe the apparent location of the sound source within the head.

The range of lateralization is limited, since most lateralized sounds are perceived somewhere along an imaginary line joining the two ears, but localized sounds can be perceived as coming from any direction.

Monaural: situation where a sound is delivered to one ear only.
Binaural: situations where sound reaches both ears.
Diotic signal: a stimulus arriving at both ears is identical.
Dichotic signal: the sound is different at the two ears.

The direction of sound sources in space are defined by three planes relative to the head.
Horizontal plane: passes through the upper margins of the entrances to the ear canals and the lower margins of the eye sockets.
Frontal plane: lies at right angles to the horizontal plane and intersects the upper margins of the entrances to the ear canals.
Median plane: lies at right angles to the horizontal and frontal planes.

The center of the head serves as a point of intersection of all three planes and defines the origin of the coordinate system for specifying angles of sounds relative to the head. The direction of the sound can be specified by its azimuth and elevation.
Azimuth (Θ)*:* angle produced by projection onto the horizontal plane.
Elevation (δ)*:* angle produced by projection onto the median plane.

There are two aspects of performance for the localization of sounds.

1. How well the perceived direction of a sound source corresponds to its actual direction.
2. How well subjects can detect a small shift in position of a sound source.

The Localization and Lateralization of Pure Tones

Cues for Localization

There are two possible cues as to the localization of a sound source: an interaural time difference (ITD) and an interaural intensity difference (IID). If the IID is specified in decibels, it is then referred to as an interaural level difference (ILD). Both cues are not equally effective at all frequencies.

Low frequency sounds have a relatively long wavelength compared to the size of the head, thus, the sound diffracts or “bends” around the head and results in little or no “shadowing” caused by the head. For high frequencies, the wavelength is short compared to the dimensions of the head, little diffraction occurs.

For a sound source distant from the listener, ILD's are negligible below about 500 Hz, but may be quite large (20 dB) for high frequencies. Sound sources close to the head can cause considerable ILD's at low frequencies.

ITD's range from 0 (sound located straight ahead) to 690 μs (directly opposite one ear). This time difference can be calculated by the path difference between the two ears. Time differences, for sinusoidal sounds, are equivalent to a difference in phase or interaural phase difference (IPD) between the two ears. For low frequency tones, the IPD provides effective and unambiguous spatial information of a sound. However, for high frequency sounds, the IPD provides ambiguous cues. Head movements or movements of the sound source may solve this ambiguity. There is no limit in the ability to detect phase differences between the two ears, however, they become very ambiguous for frequencies above 1500 Hz.

The “duplex theory” introduced by Lord Rayleigh in 1907, states that sound localization is based on ILD's at high frequencies and ITD's at low frequencies. This idea holds rather well for sinusoidal tones, but is not quite accurate for complex sounds.

Performance in Localization and Lateralization Tasks

The minimum audible angle (MAA) is the smallest detectable change in angular position, relative to the subject.

For example, for $\Theta = 0^{\circ}$ a shift of about 1º can be detected for frequencies below 1000 Hz. As the frequency increases (1500-1800 Hz), the performance worsens; this is consistent with the duplex theory. Performance markedly worsens when the reference azimuth is moved away from 0º.

The threshold for detecting ITD changes are smallest when the reference ITD is zero (sound heard in the center of the head). The smallest detectable ITD increases with lower frequencies, although it is roughly constant at about 3º. Above 900 Hz, the ITD threshold increases markedly and above 1500 Hz it is essentially undetectable.

The threshold for detecting changes in the ILD's are smallest when the reference is also zero. Changes in the ILD of about 1 dB can be detected across a wide range of frequencies, although performance worsens for frequencies around 1000 Hz.

Binaural Beats

They are head when a tone of one frequency is presented in one ear, via headphone, and a tone of slightly differing frequency is presented to the other ear. If the frequency difference is equal or less than 2 Hz, the sound appears to move right and left across the head. For higher frequencies, the perception is more diffuse and appears as a fluctuation in loudness or roughness. Binaural beats depend upon an interaction in the nervous system of the neural output from each ear. They serve as a demonstration that the neural discharges in the auditory nerve present phase information of the sound stimulus.

Neural spikes occur at a particular phase of the stimulating waveform. At points in the auditory system where the signals from the two ears are combined, the trains of neural spikes from the two ears superimpose differently depending on the relative phase of the stimuli at the two ears.

The Lateralization of Complex Sounds

The Acuity of Lateralizing Transients

Greatest acuity is achieved for stimulus that provide continuous transient information (i.e. noise and sinusoids) and are relatively long in duration. For tones of shorter duration, acuity is not very good, and cues related to onset and offset transients become more important.

Acuity as a Function of Frequency and the Use of Envelopes

For a train of pulses, the discrimination of lateralization degrades if these clicks were highpass filtered, but it remains unaffected when lowpass filtered. Also, masking with low pass noise produces a marked disruption, while high passed noise had little effect. The discrimination of lateral position on the basis of time delays between the two ears depends largely on the low-frequency content of the clicks.

For sounds with a complex sound, sound lateralization depends on the interaural time delay of the amplitude envelope rather than the time delay in the “fine structure”. Because the ITD of the envelope can be processed even for complex sounds with no components below 1500 Hz indicates that the duplex theory is not completely correct.

Time-Intensity Ratio

If identical clicks are presented to the two ears via headphones, then the sound source is usually located in the middle of the head; the image is said to be centered. If the click on the left ear is made to lead that in the right ear by a small amount, the sound image moves to the left. However, by making the click in the right ear more intense, it is possible to move back toward the the right, so that the image is centered once again. It seems possible to trade time differences for intensity differences. The amount of time difference needed to offset a 1dB difference in level at the two ears is described as the “trading ratio”.

The “latency” hypothesis suggests that the time required to evoke a neural response is shorter for intense sounds, so that intensity differences were transformed into time differences at the neural level. However, some experiments indicate that time and intensity are not truly equivalent.

Binaural Adaptation

For a train of clicks at a high rate, the binaural system appears to process only the onset of the click train; it seems like there is an adaptation to high click rates, so that clicks after the first convey little information for localization. The higher the click rate, the faster the adaptation. Although the later clicks can still be heard, they do not aid in the improvement of localization. “Triggers”, such as low intensity short noise or tone burst, can produce a release from adaptation. Essentially, the release of the binaural adaptation is because the triggers signal a change in the stimulus.

Binaural Interference

The ability to detect or discriminate interaural differences in stimuli centered at one frequency is degraded by the presence of conflicting interaural differences in stimuli centered in another frequency region. This across-channel interference is called binaural interference.

It is similar to modulation detection interference (MDI) in that it shows some asymmetry in frequency. Low-frequency interferers strongly degrade the lateralization of high-frequency targets, while high-frequency interferers have little or no effect on the lateralization of low-frequency targets.

Binaural interference can be interpreted as the fusion of the target and interfering sound into a single sound image. The perceived location of the fused image depends on some form of average of the positions of the target and interfering sounds. If judgments of lateral position are based upon this image, then changes in the position of the target sound will be more difficult to detect when the target is accompanied by an interfering sound that does not change. Furthermore, low-frequency sounds are weighted more heavily than high-frequency sounds in determining the average, also known as “spectral dominance”.

Across-Frequency Integration and the Effect of Bandwidth

For sounds with a small bandwidth, such as sinusoids and narrow noise bands, the cue of ITD can be ambiguous. If the IPD is ambiguous, we tend to perceive the sound in a location corresponding to the smallest IPD. This ambiguity can be resolved by increasing the bandwidth of the stimulus so that it produces more or less independent outputs from several auditory filters. The auditory system, thus, determines the localization of a sound on the basis of the ITD that is common across “channels”, rather than on the basis of the smallest IPD that varies across “channels”.

The Cone of Confusion and the Role of Head Movements

If the head is kept stationary, then a given ITD is not sufficient to define uniquely the position of the sound source in space; there is a cone of confusion such that any sound source on the surface of this cone would give rise to the same ITD.

Ambiguities related to the cone of confusion are resolved by head movements. Head movements provide cues other than changes in ITD and IID. For complex sounds, movements of either the head of the sound source produce changes in the spectral patterning at each ear, and these changes provide the cues for determining the direction of a sound source.

Monaural Localization, The Role of the Pinnae, and HRTFs

Head movements or movements of the sound source are important in resolving ambiguities in the vertical direction. However, our abilities are far greater of what would be predicted if the only information available was related to interaural differences and changes in those differences produced by head movements.

It has been suggested that the pinnae provide information which is used in judgments of vertical location and for the discrimination of front and back, while others suggestions state that the pinnae are important for localization in every direction.

Pinnae modify the spectra of incoming sounds in a way that depends on the angle of incidence of the sounds relative to the head. The head and the pinnae form a complex direction-dependent filter. This filtering is characterized by measuring the spectrum of the sound source and the spectrum reaching the ear drum. The ratio of these two gives the “head related transfer function” (HRTF). The HRTF pattern shows a complex arrangement of peaks and dips which vary systematically with the direction of the sound source relative to the head and which is unique for each direction in space. The spectral changes produced by the head and pinnae can be used to judge the localization of a sound source.

Because of the vastly different shapes and sizes of the head and pinnae, the HRTFs vary with individuals.

The Precedence Effect

In normal listening environments, the sound of a given source reaches our ears via a number of different paths. Some of the sound arrives by a direct path, but a good portion of it reaches our ears after one or more reflections from the surfaces of the room. Despite this, we are not normally aware of these reflections and they appear to have little influence on our judgments of the direction of the sound source. As such, we can still locate the sound source in a very reverberant room where the total energy in the reflected sound may be greater than that reaching our ears by a direct path.

From the results of a particular experiment, we can summarize the phenomenon as follows:

1. Two brief sounds that reach the ears in close succession are heard as a single sound if the interval between them is sufficiently short. The time interval can be as low as 5 ms for single click and as long as 40 ms for complex sounds.
2. If the two successive sounds are heard as fused, the location of the total sound is determined largely by the location of the first sound. This is known as the “precedence effect” or the “Haas effect”.
3. The precedence effect is only shown for sounds of a discontinuous or transient character.
4. The lagging sound can be shown to have a small but demonstrable influence. If the location of the lagging sound departs more and more from the location of the leading sound, it “pulls” the total sound along with it up to a maximal amount (7º) and the progressively becomes less effective.
5. If the interval between the arrival of the two sounds is 1 ms or less, the precedence effect does not operate.
6. If the lagging sound is made sufficiently intense (10-15 dB higher than the leading sound), it overrides the precedence effect and is heard as a separate effect.
7. It is assumed that the precedence effect works most effectively when the leading and lagging sounds are qualitatively similar. However, even when the two sounds have a markedly different spectrum, the leading sound can disrupt the lateralization of the lagging sound.
8. The precedence effect is usually described as resulting from the processing of ITD's and ILD's.
9. The precedence effect can take time to build up.
10. Changes in acoustical conditions can cause the precedence effect to break down temporarily.
11. The precedence effect does not involve a complete suppression of information about echoes. Listeners can easily hear the differences between a sound with echoes and one without echoes. Also, change in the pattern of the echoes can be readily detected. The pattern may supply information about room acoustics and the positions of walls and objects in a room.

It has been suggested that the precedence effect does not depend on a “hard wired” mechanism, rather, it may be partly reflected by high-level cognitive processes.

The precedence effect plays an important role in our perception of everyday sounds. It enables us to locate, interpret, and identify sounds in spite of wide variations in the acoustical conditions in which we hear them. Without it, listening in reverberant rooms would be an extremely confusing experience.

Binarural Masking Level Differences

The masked threshold of a signal can sometimes be markedly lower when listening with two ears than when listening to one. The masking level difference (MLD) of binaural masking level difference (BMLD) is the measure of the improvement in detectability of a signal which can occur under binaural listening conditions. It is the difference in threshold of the signal for the case where the signal and the masker have the same phase and level relationships at the two ears and the case where the interaural phase and/or level relationships of the signal and masker are different.

The MLD implies that the detection and and discrimination of the signals are improved when the signal and masker are not coincident in space. This phenomenon is closely related to the "cocktail party" effect.

Let us define a couple of terms:
"Homophasic" is the condition when the relative phase of the signal at the two ears is the same as the relative phase of the masker. When the phase relationship are opposite, they are defined as "antiphasic".

The Sluggishness of the Binaural Systematically

Binaural sluggishness is the phenomenon in which only slow changes in location of a sound source can be consciously followed. The minimum audible movement angle (MAMA) is the angle through which a sound source has to move for it to be distinguished from a stationary source. For relatively low movement rates, the MAMA is about 5º, but as the rate of movement increases, the MAMA progressively increases to about 21º. Thus, the binaural system is relatively insensitive to movements at high rates.

Models of Binaural Processing

Recent models of binaural processing assume that firing patterns are compared for neurons with corresponding CF's in the two ears. At each CF there is an array of delay lines which can delay the neural spikes from one ear relative to those from the other; each delay line has a characteristic time delay. These are followed by coincidence detectors, which count the number of spikes arriving synchronously from the two ears. The interaural delay of the signal is coded in terms of which delay line gives the highest response in the coincidence detectors.

Influence of Vision on Auditory Localization

Stationary sound sources are usually perceived as having a fixed position in space. If the head is moved, the sound arriving at the two ears changes, but the sound image remains stationary in space. It is thought that somehow the information about the sound image is combined with information about the sound location relative to the head to arrive at a constant percept. Vision, in this case, plays an important role by helping define the position of the head in space.

The perception of auditory spatial cues is strongly influenced by perceived visual orientation. In other words, the highest level of spatial representation involves an integration of information from the different senses.

The Perception of Distance

For familiar sounds, sound level may give a crude indication of distance from the listener. This cue appears to be most effective when multiple sound sources are present, so that comparison of the levels of different sources is possible. It has been shown that the change in intensity that occurs when a listener walks toward a sound source can provide an absolute cue to distance.

Over moderate distances, the spectrum of a complex sound source may be changed, owing to the absorbing properties of the air; high frequencies are more attenuated than low ones. This cue, however, does not provide a cue for absolute distance. For sounds close to the head, larger than normal ILD's provide a cue for distance.

The previously described cues could be used to judge the distance of a sound source in free space. But in the case of listening in rooms with reflecting walls, the ratio of direct to reflected sound and the time delay between direct and reflected sound provide a cue to distance. In spite of perceptual fusion or the precedence effect, we are able to make judgments of distance from the information provided by echoes.

It should be noted that judgments of distance are relatively inaccurate, and errors of the order of 20% are not uncommon for unfamiliar sound sources. Also, for nearby sound sources, the distance tends to be overestimated, while for far away sources the distance tends to be underetimated.