# Auditory Perception

Like the visual system, the human auditory system can be divided into two stages:

* Physical reception of sounds
* Processing and interpretation

Like the visual system, the human auditory system has both strengths and weaknesses:

* Certain things cannot be heard even when present
* Processing allows sounds to be constructed from incomplete information

The principal characteristics of sound - as perceived by the listener - are:

* Pitch
* Loudness
* Timbre

## Pitch

The human ear can detect changes in air pressure at rates from 25 to 16,000 cycles a second (approximately).

The faster the rate of change (i.e., the more cycles per second) the higher the *pitch* of the sound we hear.

* Most people have good *relative* pitch recognition
* Only a few people (mainly musicians) have good *absolute* pitch recognition

The relationship between frequency and pitch is not linear.

If we hear two tones, one of which varies at *twice the frequency* of the other, we hear the difference between them as an *octave*.



The smallest pitch/frequency change that can be discerned (the Just Noticeable Difference or JND):

* is roughly constant across the frequency-range for pitch
* varies for frequency, with quite small changes being discernible at low frequencies, but only larger changes being discernible at higher frequencies.

## Loudness

The perceived intensity of a sound depends upon:

* The sound pressure.
* The distance between the source and the listener
	+ sound intensity does not decline linearly with distance.
* The duration of the sound
	+ human beings are very poor at judging the loudness of sounds that are heard for less than about 0.2 seconds.
	+ Such sounds usually appear much quieter than their intensity might suggest.
* Absorption and reflection of the sound by the air and by surrounding objects.
	+ Different frequencies may be absorbed/reflected by different amounts.
* The frequency of the sound
	+ the perceived loudness varies with frequency as well as amplitude.

For example:



## Timbre

Natural complex sounds usually consist of:

* A fundamental tone - a sine-wave with a certain frequency
* A series of harmonics - other sine-waves at higher frequencies, usually at multiples of the fundamental frequency

The quantity and distribution of harmonics and their relative strengths help determine the *timbre* of a sound.

For example:

* one musical instrument might produce a sound which includes lots of harmonics, some of which are almost as strong as the fundamental
* another might produce a sound which comprises just the fundamental plus a few, relatively weak harmonics.

Also:

* one sound might contain only *even-order* harmonics (twice the fundamental frequency, four times the fundamental frequency, etc.)
* another might include *odd-order* harmonics (three times the fundamental frequency, five times the fundamental frequency, etc.).

These differences would contribute to the individual sound of an instrument, helping us to recognise the difference between (e.g.) a violin and a trumpet.

If we hear a complex wave in which the component sinewaves are:

* harmonically related (i.e., exact multiples of the fundamental frequency), we perceive it as a single sound.
* NOT harmonically related, we perceive it as several separate sounds.

Sounds also vary in the way in which their amplitude varies over time.

This is known as the *amplitude envelope*, for example:

* Some sounds start at a high amplitude but then fade in intensity (e.g., percussion instruments)
* Other sounds start at a low amplitude and build in intensity (e.g., some wind instruments)

The amplitude envelope helps us to distinguish between sounds which may have similar harmonic structures.

## Localisation of Sound Sources

Our hearing system allows us to determine the location of sound sources with reasonable accuracy, subject to certain limitations.

* **Stereo hearing** allows us to locate the source of a sound by comparing the sound arriving at each ear and noting differences in:
	+ **amplitude** (interaural intensity), although our sensitivity to amplitudes changes is limited
	+ **time of arrival** (interaural delay) - we can recognise differences of 10 micro-seconds or less between the time of arrival of a sound at each ear.

Stereo hearing works in the horizontal plane only and is least effective in the middle range of audible frequencies.

* **Head movement** allows us to improve the localisation accuracy of stereo hearing.
* **Analysis of reflected versus direct sound** allows us to localise sound in both the horizontal and vertical planes - to a limited extent.
* **Familiarity** affects localisation accuracy - both ways.

Research has shown that human ability to locate sources of sound in the horizontal plane varies from:

* Around 1o or less when the source is directly in front/behind the listener
However, *front-back reversals* are common
* Around 15o or more when the source is to the left or right of the listener.
* Sounds originating within an arc of approximately 70o on either side of the head are localised with least accuracy.



Localisation performance is better for non-musical sounds (e.g., clicks, percussive noises, etc.) than for musical tones.

Localisation errors at the sides of the head are typically 8o for clicks and 5.6o for noises.

Front-back reversals are also less common for clicks and noises.

Localisation varies with frequency:

|  |  |  |
| --- | --- | --- |
| Below 1000 cycles | good | Based on timing/phase differences |
|   |   |   |
| 1000 to 3000 | poor | Neither timing/phase nor intensity differences predominate. |
|   |
| Above 3000 cycles | good | Based on intensity differences |

Localisation of sound in the vertical plane is far less accurate than localisation in the horizontal plane.

Recent research - aimed at establishing guidelines for the use of sound in 3D environments - concluded that the average listener can reliably distinguish only **three** vertical source locations.

Judgement of distance is based partly on intensity - the quieter the sound, the further away the source.

However, distance also affects:

* The audio spectrum of the sound - some frequencies travel better than others
* The balance between reflected and direct sound - the further the sound has travelled, the more likely it is to include a significant percentage of reflected components.

Sound localisation (in both the horizontal and vertical planes) can be improved by tailoring the sound distribution.

This is done using **Head-Related Transfer Functions** (HRTFs).

Ideally, HRTFs should be tailored to suit the individual. However, this is complex and costly.

Researchers are currently trying to develop *non-individualised HRTFs* which will give a useful improvement in localisation accuracy for a substantial percentage of the population.

## Short-term Auditory Memory

Research suggests that the human auditory system includes a short-term store - a kind of mental 'tape loop' that always stores the last few seconds of sound.

This is known as the *Pre-categorical Acoustic Store* or PAS (Crowder & Morton, 1969).

Researchers disagree as to the length of the store. Estimates range from as little as 10 seconds to as much as 60 seconds.

However, there is significant evidence for the existence of such a store.

The existence of this auditory store explains some of the following effects:

* **Recall of Un-attended Material**
	+ A number of studies (e.g., Glucksberg & Cowan, 1970) have been conducted in which subjects were asked to listen to several simultaneous streams of speech or sound, then recall the content of ONE of the streams.
	+ They were either not told in advance which stream they would have to recall, or were deliberately told to concentrate on the wrong stream.
	+ Subjects were able to recall the last few seconds of sound from any of the streams, but could only recall earlier material from the stream to which they had consciously listened.
* **The Recency Effect** (Postman & Phillips, 1965)
	+ If someone listens to a voice reciting a list of digits (or characters, etc.), and is then asked to repeat the digits, he or she will recall the last few digits more reliably than the earlier ones.
	+ Typically the last 3-5 digits are recalled.
	+ The number of digits recalled is roughly constant: if the list is made longer, more digits will be forgotten from the earlier parts of the list, but roughly the same number of digits from the end of the list will be recalled.
* **The Auditory Suffix Effect** (Conrad, 1960)
	+ The recency effect (see above) is most noticeable when the speech or sound is followed by a period of silence.
	+ If a further sound occurs after (e.g.) a list has been spoken, recall is impaired.
	+ Conversely, if speech or sound is followed by complete silence, the period for which the last few seconds of it can be recalled extends significantly.

In short, the human hearing system behaves as if it incorporates a 'tape-loop' that can store a few seconds of sound:

* Sounds are recorded onto the loop as they are heard.
* New sounds are recorded over older sounds, but...
* if no new sounds are heard, the previous recording remains.

New sounds impair recall regardless of their type, but speech (or sounds that are interpreted as speech) cause greater impairment than non-speech sounds.

## Speech and Non-Speech Sound

Research suggests that the human hearing system responds differently to speech and non-speech sounds.

Speech appears to make greater demands on mental resources. Consider, for example:

* The Auditory Suffix Effect (described earlier).
* Studies (such as that by Ayers, Jonides, et.al., 1979) which show that an ambiguous sound causes more disruption to recall and other processes when it is interpreted as speech than when it is interpreted as a musical sound.