SCI220 – Foundations of Musical Acoustics
Cogswell Polytechnical College
Fall 2008

Week 4 – Class Notes

Psychoacoustics / Hearing (MA)

The Human Ear and Its Response

The Mechanism of the Human Ear

The task of the ear is to convert incoming air-pressure fluctuations into electrical nerve impulses for processing by the brain. The ear consists of three sections: the outer ear, middle ear and inner ear.

The *outer ear* consists of the pinna (or auricle) and auditory canal (or meatus)and ends at the eardrum (or tympanic membrane). The *pinna* is the visible part standing out from your head and its inner portion funnels short-wavelength sound towards the eardrum. The *auditory canal* is nearly 1cm across and 2.5-3cm long and allows for the eardrum to be set where it can be protected from dirt and sharp objects. The *eardrum* is a thin disc consisting of fibrous tissue that bulges in and out as a response to alternations in air pressure.

The *middle ear* is the enclosed chamber immediately behind the eardrum. It is connected to the throat by the *Eustachian tube*, which serves as a mechanism to equalize the pressure on both sides of the eardrum. Unequal pressures reduce listening acuity by distending the eardrum. Two small membrane-covered openings, the *oval* and *round windows*, communicate the middle and inner ears, as well as confine the *perilymph* to the inner ear.

The middle ear contains the *ossicles.* They are three bones (*malleus* or hammer, *incus* or anvil, *stapes* or stirrup) that act as levers that transfer mechanical vibrations from the eardrum to the inner ear. For moderate-amplitude sounds, the maleus moves approximately 1.3 times as far as the stapes. Meaning that the stapes exerts 1.3 times as much force on the oval window as the eardrum does on the malleus. This can be explained as a lever system in which a smaller force must move through a larger distance on one end while a larger force moves through a smaller distance in the other, so that both ends account for the same amount of work done.

The middle ear also contains two muscles, the *tensor tympani* which is attached to the malleus and increases the tension in the eardrum and the *stapedius* which pulls the stapes sideways and reduces the mobility of the ossicle chain. Both actions are intended to reduce sound transmission through the middle ear and protect the inner ear from damage caused by extremely loud sounds. The stapedius is activated by the *acoustic reflex* within 10-20ms for sound levels above 90-100dB. The tensor tympani contracts as a much general reaction that takes 10 times as long. This two protection mechanisms provide up to 20dB or more attenuation for frequencies below 1KHz.

Because the force F_s from the stapes is applied to the very small area S_{ow} of the oval window, it produces a large pressure p_p in the perilymph: $p_p = F_s / S_{ow} = 1.3 F_m / S_{ow} = 1.3(p_e S_e) / S_{ow}$. (Subscripts m and e refer to malleus and eardrum). Because the area of the eardrum is approximately 20-25 times larger than than of the oval window (which is about 3mm²), the pressure variations produced in the perilymph will be some 30 times greater than those in the eardrum. Intensity depends on the square pf amplitude, so this suggests that we can hear sounds approximately 1000 times less intense (30dB lower in sound level) than if the oval window received its vibrations directly from the air.

The *inner ear* occupies a labyrinth of passages within the temporal bone. It consists of the *semicircular canals*, whose role is to sense changes in orientation of the head and the *cochlea*, which is tapering tube coiled some 3 times like a snail shell (unrolled it would be some 3.5cm long).

Inside the cochlea we find the cochlear partition that separates it into two regions, the *scala vestibuli* and the *scala tympani*, along its entire length except for a small opening called the *helicotrema* at the farthest end from the middle ear. All of these regions are filled with perilymph. The partition has a hollow center called the *cochlear duct* or *scala media* filled with the more viscous *endolymph*. The *vestibular* or *Reissner's membrane* is thin, light, and rides along with motions of the adjoining fluid and has no acoustica importance.

The *basilar membrane* is the structure whose vibrations stimulate the hair cells to produce nerve impulses carrying information about sound to the brain. The *Organ of Corti* rides loosely on the inner part of the basilar membrane and contains more than 20,000 hair cells that initiate signals on the individual fibers of the *auditory* nerve when disturbed by the vibrations of the basilar membrane.

For low frequencies there is sufficient time to move from the scala vestibuli to the scala tympani, but for higher frequencies the fluid near the oval window must push down on the basilar membrane to communicate its motion quick enough to the round window.

The tendency for higher-frequency vibrations to disturb the basilar membrane closer to the windows is encouraged by the fact that this end is relatively narrow, stiff, and light, while the end near the helicotrema is wider, more lax, and massive.

There is no direct path from the inner ear to the brain via the auditory nerve, however signals are mixed and partially processed at several points before finally being presented to the *auditory cortex* in the higher parts of the brain for conscious interpretation. Sound perception depends on both the brain and the inner ear.

Limits of Audibility and Discrimination

Human ears are most sensitive to sounds with frequencies from 2-5KHz. A sound level of 0dB (intensity $I = 10^{-12}$ W/m^2) is considered to be the faintest audible sound, but for most this level is 10-20dB or more (depending on the frequency). The audible frequency range is from 20Hz to 20KHz, but this is not a fixed boundary. This range can vary with age and damage by exposure to loud sounds. Frequencies below 30Hz are difficult to hear, but under special conditions it is possible to hear sinusoids around 15Hz. However, for frequencies at this range they are not so much heard as they are felt.

Just Noticeable Difference (JND) – minimum change required in any stimulus (sound level or frequency) before the difference is reliably detected by a human listener.

The JND for sound levels is determined in the following manner:

Two oscillators provide signals X and Y, both sinewaves at the same frequency but at slightly different levels. The output of X is set at a fixed level (say 40dB) at the listener's ears but change the output of Y from time to time. The listener hears X and Y alternatively, and is required to tell which one of the pair seems louder. If the sound level is clearly different (40dB vs 60dB) the judgment made will be correct 100% of the time. If both levels are the same (at 40dB), the listener will have a 50% choice.

However, if both signals are very close in level (say 40.1 and 41 dB) the level of assertion will be considerably smaller. There is a transition from total assurance to complete uncertainty (from 40 to 60dB).

Although the JND tends to be slightly larger for lower frequencies and lower intensities, it is a fair approximation for most sounds of musical interest to say that the JND for sound level is between 0.5dB and 1dB. Meaning the intensity must change 15-30% before a difference is noticed.

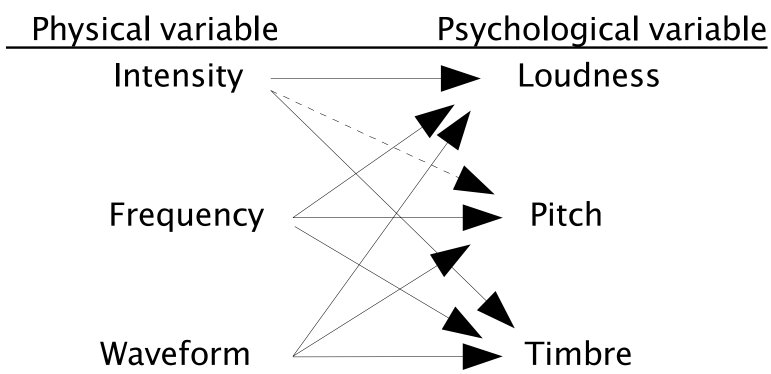
For frequency JND's, there are alternating tones of equal loudness with one signal kept at a fixed frequency. The other signal is varied in frequency to determine the minimum frequency difference to reliably judge which one has a higher pitch.

In summary, for sinusoids with frequencies below 1KHz the JND is 1Hz and increases above this limit. Beyond 5KHz, the JND rises as our pitch judgment becomes quite poor; above 10KHz pitch discrimination vanishes.

Characteristics of Steady Single Tones

If the physical description of a particular sound is characterized by its amplitude (intensity), vibration frequency, and waveform, it is then said that the psychological perception is characterized by its loudness, pitch, and timbre.

Timbre is the "tone color" of a sound. The psychological impression of what characterizes a tone besides its loudness and pitch. It is a complex and multidimensional variable since intensity, frequency, waveform, and duration affect it in different and independent ways.



Loudness and Intensity

Loudness is primarily dependent on intensity. The mechanism is such that large air-pressure variations cause larger amplitude motions of the eardrum, middle-ear bones, oval window, perilymph, and basilar membrane. The more violent the motions are in the basilar membrane, the more stimulation of the hair cells and thus more nerve impulses that are sent to the brain.

Nerve cells do not carry voltage changes depending on the strength of the stimulus, rather they function as an on/off switch in which all pulses are the same. In order to differentiate the number of impulses changes according to the strength of the stimulus; more pulses for stronger stimulation. The sensation of loudness is thus interpreted by the brain as the average rate of impulse arrival on the auditory nerve.

The unit of measurement for loudness is the *sone.* A loudness of 1 sone is defined to be that of a 1000Hz sine wave at 40dB sound level, and all other are compared to that. Because the ear is not equally sensitive to all frequencies, any frequency other than 1000Hz has its own loudness-intensity relationship.

For frequencies of musical interest (above 50dB), it is fair to say that every 10dB increase in sound level (every multiplication of intensity by 10) will double the loudness in sones. For example, a group of 10 people singing the same note will sound about as loud as a soloist, a chorus of 100 will sound four times as much.

Pitch and Frequency

A strong sensation of pitch is determined mainly by its frequency. Our ears assign any periodic waveform very nearly the same pitch as a sine wave with the same repetition frequency.

The *mel* is the unit used to measure pitch differences. The mel scale is that which a 1000Hz tone is assigned 1000 mels and all other tones are assigned a number of mels tellig how they sound when compared to 1000. This implies that other pitches are to be regarded as larger or smaller compared to 1000.

This analogy to sones for loudness is irrelevant to musical acoustics because 1) it is based on a misconception that musical pitch can be measured on a prothetic scale (say measuring large and small), and 2) a true mel scale can be obtained only by judging pitch outside of any musical framework, with subjects with no musical experience.

For practical purposes, pitch judgments based on the octave by well-trained musicians is more important. Every octave interval represents an equal change in musical pitch throughout the range of pitches that can be reasonably judged and have musical importance.

Pitch and Loudness Together

Frequency and loudness - large changes in intensity can produce shifts as much as a semitone in pitch perception, even when the frequency remains constant. Low pitches tend to sound a bit lower, and high pitches higher, when made very loud; medium pitches are hardly affected.

Intensity and loudness - sounds with the same intensity but different frequency may have different loudness levels.

To represent how loudness depends on both intensity and frequency, we use a *Fletcher-Munson diagram,* and each of its contours are *equal-loudness curves.* Each contour represents a family of sine waves with different combinations of intensity and frequency so that the all sound equally loud.

Loudness level is the limited information about the loudness of any sound, involving only comparative judgment but no magnitude estimation. The loudness level, in *phons,* is the same number as the sound level of a 1000Hz sine wave whose loudness equals that of the sound in question.

The *Phon* is a numerical label to identify the loudness level of any sound; defined by the sound level of a 1000Hz sine wave of equal loudness.

Loudness is determined in two steps. First, we obtain the loudness level from the intensity level and frequency on the Fletcher-Munson diagram and then go from the loudness level in phons to the corresponding actual loudness in sons by checking a simple graph.

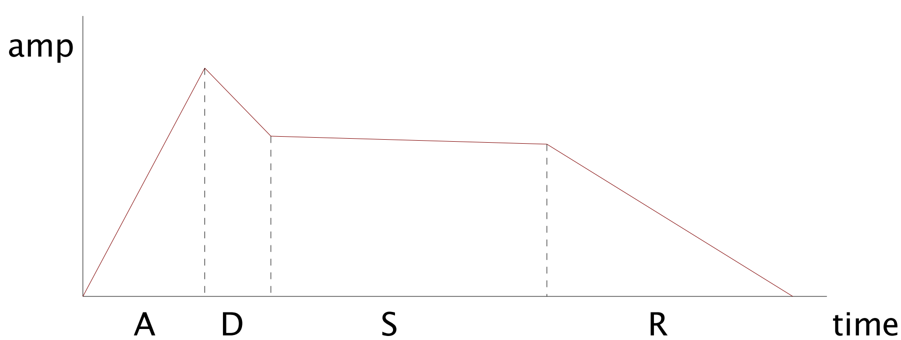
Remember:
Loudness level (phons)
Loudness (sones)

Timbre and Instrument Recognition

Waveform recognition is not sufficient for instrument identification. Instrument recognition depends on hearing the transients or the attacks and decays (envelope of a sound) of a rapidly changing waveform.

The duration of the initial transients on any instrument vary depending on the instruments range. Oboe attack time range from 20-30ms, trumpets have 30-40ms attack times, and flutes/ violins have 70-90ms attack times.

Envelope (ADSR)
Attack – Decay – Sustain – Release



Types of Pitch Judgment

Perfect or Absolute pitch: ability to identify the pitch of a tone heard in isolation, or to sing a named pitch on demand without am external reference.

Thought to be either hereditary or learned behavior.

Comparative or relative pitch: limited sense of hearing in which when listening to two tones it can be decided whether they are the same pitch or which one is higher or lower than the other.

Pitch Perception Mechanisms

Pitch perception theories

Place Theory: associates pitch with a different place in the basilar membrane. Different frequencies preferentially stimulate different places on the membrane and thus different nerve endings. The amplitude of motion of the membrane varies with position for several frequencies. The place of maximum amplitude serves as an index of the driving frequency and thus of pitch. It also corresponds for different harmonic components as they excite different parts of the membrane, and take in account their relative strengths and not the phases.

This theory, however possess some problems. First, it does not describe how we assign pitch to a complex tone, only suggesting that we perceive many pitches together in which each one excites a specific region in the basilar membrane much and that at this peak is a strong Fourier component. With the missing fundamental phenomenon, it does not describe why we hear a fundamental pitch when given a particular harmonic recipe even when the region where this fundamental pitch is not excited.

Periodicity Theory: supposes that the messages from the cochlea to the brain contain more than just information on strengths of the harmonic components. Rather, some information or original waveform is retained. The simplest way to retain this information is to have each nerve ending fire preferentially during one part of the cycle of basilar membrane oscillation. However, this works well for lower frequencies up to 1KHz. It must be supposed that the brain has some way of measuring the time of separation between the period of the stimulating sound wave quite accurately.

Periodicity theory does explain the missing fundamental phenomenon. It takes the overlapping excitation curves so that many high harmonics excite the same part of the basilar membrane. This combination of several high harmonics can make a response waveform with prominent peaks whose period is the same as that of the series fundamental. In this light, a period-measuring mechanism in the brain could get the same message as it would from the fundamental and thus report the same pitch.

While both theories have their corresponding strong and weak points, each one by itself is insufficient to explain our complex hearing mechanism. Newer theories contain certain elements of both theories and expand them with other ideas.

Modern Pitch Perception Theory

Pattern Recognition: ability and tendency to perceive complex stimuli as forming standard patterns; for sound, especially, the assignment of pitch by fitting an actual Fourier spectrum as closely as possible to a harmonic series template.

Pattern search and recognition accounts nicely for our perception of pitch in notes from the piano, chimes, and bells (sounds with slightly inharmonic components).

Critical Band

Range of frequencies within which any two signals strongly stimulate a common portion of the basilar membrane.

See last weeks notes on special intervals and pitch matching.

Combination Tones

The presence of difference tones indicates that a nonlinearity is introduced in the sound transmission path. Since the air itself does not contribute significantly to nonlinear distortion unless sound levels are far above 100dB, the remaining candidates are the eardrum, middle ear, and inner ear. It is difficult to state exactly where the nonlinearity resides, but it has been long accepted that a nonlinearity is present in the ear.

See last weeks notes on heterodyne component.

Loudness and Masking

The impression of loudness derived from any one critical band depends only on the total energy received in that band, and any two well-separated critical bands each make an independent contribution to total loudness. That is, within a critical band intensity is additive, but beyond a critical bandwidth loudness is additive.

Masking: ability of one sound to obscure the presence of another, especially one weaker and higher in frequency. It can manifest itself in three ways: simultaneous, forward, and backward masking.

Simultaneous masking occurs if two sounds are heard simultaneously and one is masked by the other.

In forward masking, a weak sound emitted soon after the end of a louder sound and is masked by the louder sound. In backward masking, a weak sound just before a louder sound can be masked by the louder sound.

Timbre

It is likely that if several harmonic components fall within a single critical band, it must be only their combined strength that principally determines any contribution to timbre sensation from that part of the basilar membrane. Otherwise, we would have to consider as many as 50 or 100 harmonic components that have individual effects on timbre perception. This happens when components are about 15% within one another, but this is not certainly true for the first and second harmonics. Up until the sixth harmonic there is significant individual contribution to tone perception. Above the seventh harmonic, however, there is significant overlap and merging of components.

This suggests that there are six to seven critical bands supplying independent information plus a few more from the overlapping components (summing to 10-15 total critical bands) that provide information for determining timbre.

At high intensity levels, 80-100dB, the era is relatively sensitive to all harmonics the timbre is perceived as rich and solid. When intensity is reduced, 40-50dB, the energy distribution is relatively the same for all harmonics, however the ear's sensitivity decreases far more rapidly for lower components than for the others. Thus, we hear not only a sound at a softer volume but also a thinner tone quality as well.