Perceptual attributes of acoustic waves
Pitch (Part I)

**Pitch:** Sonic (*i.e.* perceptual) attribute of sound waves, related mainly to frequency.
Large frequency values result in 'high' pitch while low frequency values result in 'low' pitch.
ANSI definition: “That attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high”

The frequency range of hearing extends from ~20Hz to ~20,000Hz (20KHz). On average, frequencies below 20Hz sound as individual pulses with no definite pitch, while frequencies above 20KHz are inaudible by humans. The frequency hearing range that can give an accurately identifiable pitch sensation extends from ~30Hz to ~5000Hz. The common instrumental pitch range spans most this frequency range (see the figure, below).



The range of maximal pitch accuracy extends ~six octaves above ~60Hz (~B1), to 60\*26 = ~3800Hz (see Fig. 1, below, for an explanation of this formula).

Frequencies above 10kHz give rise to pitch sensations that, although may be distinguishable from other pitch sensations, they are hard to identify and cannot accurately portray direction of pitch change.
[Listen to this high-frequency example](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/pitchdirectionchange.wav) over headphones (uncompressed .wav file).  It includes three 1-second-long tones, played successively and introducing two possible pitch changes: from the first to the second tone and from the second to the third.  Pick, in your opinion, the pattern of pitch changes from the following: SD,  DS,  SU,  US,  UD,  DU,  SS  UU,  DD (S: pitch stays the same;  U: pitch goes up;  D: pitch goes down).  Look at the bottom of the page for the actual frequencies of the three tones.

Complex-tone spectral components with frequencies above 10,000Hz usually represent the 'noisy' portions of musical sounds (bow scrapings, reed attacks, hammer hits, *etc.*) and have timbral (tone color) significance.
According to the American Standards definition of pitch, the pitch sensation represented by ***A4*** corresponds to ***440Hz***.

**Pitch of pure (simple/sinusoidal) tones**

|  |  |
| --- | --- |
| **Dependence on frequency**For simple/pure tones, pitch closely relates to frequency. Similarly to loudness and SIL [Sound Intensity Level], pitch relates to frequency logarithmically; pitch increase by an octave corresponds to frequency doubling. The figure to the right illustrates the frequency / pitch relationship. The same pitch interval (*e.g.* octave) corresponds to an increasingly larger frequency distance(after Campbell and Greated, 2001).  | Pitch Versus Frequency   |
| **Dependence on intensity**The pitch of pure tones also depends on intensity (see the figures to the right and below). In general, increasing the intensity of pure tones decreases the pitch of low frequencies (approx. <300Hz), increases the pitch of high frequencies ( approx. >3000Hz), and has no noticeable effect at middle frequencies.Figure to the right: **(a)** A pure tone of frequency 98Hz has a pitch of G2 when quiet (ppp)and a pitch lower than E2 when loud (fff). **(b)**, **(c)**, and **(d)** show the variations of pitch with intensity for pure tones of frequency 392Hz, 784Hz, and 3136Hz respectively. (Campbell & Greated, 1987; derived from Stevens & Davies, 1939).The figure below offers another, simplified illustration of the average dependence of the pitch of pure tones on intensity.   | Pitch dependence on amplitude |
| http://acousticslab.org/psychoacoustics/PMFiles/PMImages/pitchloud2.jpg   |
|     **Dependence on duration**Pitch also depends on duration. A tone must last more than a minimum amount of time (~10-60ms, depending on frequency and intensity) in order to sound more than a 'click' and convey a clear sense of pitch (see the figure, below). |
| http://acousticslab.org/psychoacoustics/PMFiles/PMImages/pitchduration.jpg |
|  **Dependence on the introduction of a second, 'interference' tone** • Introducing a second, intense tone (or band of noise) at a frequency below that of an existing tone will raise the pitch of the original tone.• Introducing a second, intense tone (or band of noise) at a frequency above that of an existing tone will drop the pitch of the original tone. |

As we've seen, **JND** (just noticeable difference) and **difference threshold** both refer to the smallest perceivable change of a physical variable. **Absolute threshold** refers to the limits in the values of a physical variable beyond which the variable loses its perceptual identity (*e.g.* the absolute thresholds for frequency are 20Hz and 20kHz).

|  |
| --- |
| JND for Pitch: ~0.3-1% of frequency, depending on register (*i.e.* on frequency region - see the figure, below).  Expressed differently, it corresponds to approximately 1/30th of the critical band, 1/12th of an equal-tempered semitone, or 5-8 *cents* (more information on tuning temperaments and 'cents' is beyond the scope of this class).  |
| http://acousticslab.org/psychoacoustics/PMFiles/PMImages/freqJND.jpg |
| http://acousticslab.org/psychoacoustics/PMFiles/PMImages/freqJNDa.jpg |

**Pitch of complex tones**

The dependence of the pitch of periodic complex tones on frequency, intensity, duration, and the introduction of intense 'interference' tones is qualitatively similar to that of pure tones but more complex, due to the variety of frequencies involved. The pitch JND for complex tones may be larger or smaller than that for pure tones depending on spectral context.

**Dependence of pitch on spectrum**

For periodic complex tones with harmonic spectra (*i.e.* spectral components with frequencies that are integer multiples of the frequency of the lowest component, called 'fundamental') the pitch matches (in general) the frequency of the fundamental component, regardless of whether or not this component is perceivable (*i.e.* even when it is too low in level and/or masked) and even if it is not physically present in the tones' spectrum at all.
Phenomenon of the missing fundamental: the observation that the pitch of a complex harmonic tone matches the frequency of its fundamental spectral component, even if this component is missing from the tone's spectrum.
In general, the phenomenon of the pitch of the missing fundamental (one of the perceptions often referred to as virtual pitch) cannot be explained by place (tonotopic) theories of pitch perception (see below) and provides evidence that pitch does not only depend on frequency, intensity, and duration but also on spectral distribution.
Note that various minor deviations from harmonic spectra (up to ~1-2% of frequency) and the way these interact when several instruments perform together in unison change the timbre of the combined sound rather than its pitch, resulting in what is referred to as the 'chorus effect' (richness of ensemble sound due to slow and varying beating rates among the slightly detuned components of the complex tones involved).

**Harmonic complex signals are perceived as a unit** rather than a set of multiple pure tones.
However, the individual harmonics can be heard if we draw attention to them by removing them and re-introducing them. Audio example of drawing attention to individual harmonics (Houtsma et al., 1987).
The throat singers of Tuva (peoples of Tibet and the Siberian grasslands) exploit this phenomenon to create musical passages where one singer appears to produce two pitches simultaneously: one acting as a fixed drone and one performing a sort of melody.  These passages are created from the single fundamental frequency produced (acting as a background drone) and harmonic components associated with this fundamental that are selectively accentuated (acting as the melody line). [See an article on Tuva singing](http://www.sciam.com/article.cfm?articleID=00080AA2-BA32-1C73-9B81809EC588EF21&pageNumber=1&catID=2) for more information.

**The pitch of harmonic complex tones will match that of the fundamental component even if several of the low frequency components (not just the fundamental) are missing from the tone's spectrum.**
[Listen to a pair of complex tones illustrating the phenomenon of the "missing fundamental"](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/300_15-300_10.wav) (.wav file).  Both have fundamental of 300Hz and up to 15 components (ramp spectrum: An = A1/n for all components).  The first tone has all 15 components while the second is missing the first 5. As you can see, removing these first 5 components does not change the tone's pitch.

For complex tones with a fundamental of 100Hz, the pitch sensation changes after removing the first ~15-20 components and starts deteriorating after removing the first ~25-30 components.  For fundamentals of 500Hz and 800Hz, the same observations occur after removing the first ~10-12  and ~4-7 components respectively.
[Listen to this example](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/600_13.wav) (uncompressed .wav file; ~1.2Mb).  It includes 13 harmonic complex tones, with fundamental of 600Hz, played in succession.  The first tone has all harmonic components from the 1st to the 16th and each successive tone drops one component, starting from the fundamental, until only the highest 3 components remain.  What happens to the pitch?

Experiments examining the effect of mistuning some of a harmonic tone's components on the resulting pitch have determined that **the frequency region most important to pitch is between ~400 and ~1500Hz** (*i.e.* mistuning components laying within this region has the most effect on pitch while presence of frequency components within this region results in most clear pitch sensations).  How does this statement relate to what happened to the pitch of the tones in the previous example?

**In the following example, adapted from Smoorenburg (1970)**, two complex tones with 2 components each are presented in succession.
(a) 800Hz + 1000Hz (b) 750Hz + 1000Hz.
When moving from (a) to (b), some listeners hear the pitch going down by following the motion of the first component in each tone (800Hz 750Hz). Explicit rules are employed to track the physical attributes of the two tones and determine the pitch motion, in an example of **analytic listening**.
Other listeners hear the pitch going up, by reconstructing the motion of the (missing) fundamental implied by the two complex tones (200Hz 250Hz). Implicit rules are employed to synthesize a physical attribute that is implied by the rest of each tone's attributes, helping determine pitch motion. This is considered an example of **synthetic listening**.
Listen to the Smoorenburg example

***Additional Analytic/Synthetic Listening Examples (optional materials)***
In [Dannebring's (1974) demonstration](http://acousticslab.org/extra/Week4.htm#Dannenbring1974) (masking noise filling tone gaps - bottom of the Module 3a notes), listeners synthesize the sensation in a form of listening often referred to as synthetic or holistic, based largely on implicit rules (rules we are not explicitly aware of).
If listeners are alerted to the fact that the presented tones have gaps, they may be able to perceive them by directing their attention to separate portions of the total stimulus. This 'directed' form of listening is often referred to as analytic, based largely on explicit rules (rules we are explicitly aware of).
The [McGurk effect](http://acousticslab.org/extra/PMDownloads/McGurkEffect.rm) (discussed in class) offers another example of synthetic listening.  It illustrates that perceptions are the result of an experience-guided synthesis of information from our environment, with our responses to stimuli not necessarily matching the stimuli but reflecting the best way such stimuli fit to our previous experience and the resulting expectations. The effect is so robust that, even when alerted to the stimuli inconsistency, listeners/viewers are unable to listen to the a/v composite analytically.

**Complex tones with inharmonic spectra** (*i.e.* spectra whose frequency components are not integer multiples of a 'fundamental' component) do not elicit a clear, unique pitch sensation.  They may elicit more than one competing pitch sensations, may resemble chords, or may sound as noise, depending on their spectral distribution.
As illustrated in class, however, changing the spectral peaks of inharmonic complex tones without changing the frequencies of their sinusoidal components can result in a distinguishable change in pitch, matching the frequency change of the spectral peaks, providing additional evidence of the dependence of pitch on spectral distribution.
**Click below** to listen to three individual major chords, performed by combinations of flute, clarinet, and oboe.
Clarinet-Flute-Oboe     Flute-Clarinet-Oboe     Oboe-Clarinet-Flute
Now, listen to the "chord melody" example played in class, containing five "notes," each of which is one of the three chords above.  All 3 chords are major, include exactly the same notes (C5, E5, and G5), and have spectral components that are not harmonically related (although the spectra of the individual notes in the chords are harmonic, the spectra of the entire chords are inharmonic).  The frequencies of the components are almost identical among chords but the spectral envelopes (*i.e.* relative intensities of the components) are different and depend on which instrument plays what note (a flute, a clarinet, or an oboe).  The melody you hear (C-E-G-E-C) tracks the position of the flute in each successive chord, because the flute's fundamental frequency provides the spectral peak for each chord's spectrum.  In other words, the melody you hear matches the pitch changes corresponding to the changes in the flute's fundamental frequency.

Theories of pitch

**Place (Tonotopic) theory:** Pitch relates directly to the point of stimulation on the BM. This theory was proposed by Ohm (1843), developed by Helmholtz (1862), and confirmed experimentally by von Békésy (1950s), who won the Nobel prize in medicine (1961) for his contributions to the understanding of hearing.
Click on the following links to see animations (.mov files) of the **basilar membrane motion** for [1000Hz](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/travwave1l.mov), [8000Hz](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/travwave8l.mov) and [1000Hz+8000Hz](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/travwave1_8.mov) (from the '[Ear Works](http://www.neurophys.wisc.edu/h%26b/auditory/fs-auditory.html)' website at the University of Wisconsin, Madison).

The place theory of pitch can explain the pitch of pure tones but needs to be modified to reliably address the pitch of complex tones and related phenomena.
Some observations that challenge the "place" or "tonotopic" theory of pitch include the pitch of the missing fundamental (or virtual pitch) and two related pitch-shift effects (see below).
Attempts to explain virtual pitch in terms of intermodulation distortion products (especially the difference frequency), whether created at the sound source, in the propagation medium, or in some part of the ear, have been countered by observations that cancelling, masking, or in some way removing/inhibiting these intermodulation distortion products does not alter or diminish the strength of the virtual pitch.  In addition, certain artificial spectral manipulations used to create the pitch shift effects described below, create situations where virtual pitch and the difference frequency do not match, further challenging the "intermodulation distortion" explanation of virtual pitch.

|  |  |
| --- | --- |
|        The first pitch shift effect: Shifting all components of a harmonic complex tone by a value: |Δƒ|, results in a shift of the perceived pitch by a value |ΔP|, despite the fact that the frequency spacing between the components (and therefore the difference frequency |Δƒ| among successive components of the complex tone) remains unchanged (|Δƒ|=ƒ0).  The second pitch shift effect: For the harmonic complex tone to the right (continuous spectral lines) the perceived pitch P matches that of the difference frequency ƒ0. Increasing the spacing of the components while keeping the frequency of the center component ***n*** the same, results in a drop in pitch P'< P , although the difference frequency has increased from |Δƒ|=ƒ0 to |Δƒ|=ƒ0+dƒ (broken lines). | http://acousticslab.org/psychoacoustics/PMFiles/PMImages/pitchshifts.gif |

The artificial spectra created in explorations of the two pitch shift effects, above, are slightly inharmonic but not inharmonic enough for the pitch sensation to deteriorate.

An **approximate value of the virtual pitch of slightly inharmonic spectra**, whose fundamental or even several more low components are missing, can be calculated as follows:
Assume the 3-component harmonic spectrum: 800, 1000, 1200Hz (components 4, 5, and 6 of a 200Hz missing fundamental) and a slightly inharmonic version of the same spectrum: 792, 1020, 1170Hz.
The virtual pitch of the harmonic spectrum will correspond to 200Hz, while the virtual pitch of the inharmonic spectrum will correspond, approximately, to the average of the three implied fundamentals: (792/4 + 1010/5 + 1170/6)/3 = (198+202+195)/3 = 595/3 = 198.3Hz.

|  |
| --- |
| An additional challenge to the place (or tonotopic) theory of pitch is that it requires energy in the lowest 5-8 components (depending on fundamental frequency) in order to produce a clear pitch sensation.  These low components are the components that are usually resolved best by the basilar membrane (*i.e.* they lay within separate critical bands) and can therefore provide clear place-related pitch information and/or produce strong intermodulation distortion products. The problem is that clear virtual pitch sensations persist even when the remaining components in a spectrum are not resolvable (see the 'unresolved upper harmonics' in the figure, below). |
| http://acousticslab.org/psychoacoustics/PMFiles/PMImages/BasilarMembrane+CriticalBand.jpg |

This observation led to the **first temporal theory of pitch**, called the **"Residue" theory of pitch**, developed first by Schouten (1938) and later by Walliser (1969).  It states that pitch is determined based on the temporal interaction, at a neural level, of the unresolved, 'residual' upper harmonics in a spectrum.  This theory is challenged by evidence that the frequency region most important to pitch is between ~400 and ~1500Hz (see the discussion on the pitch of complex tones, above).

**Temporal (Periodicity) theory:** Pitch relates directly to the periodicity of a sound's waveform, periodicity that is detected by the auditory system in terms of neural firing patterns. This theory was proposed by Seebeck (1843), developed by Rutherford (1886), worked out by Schouten (1938), and has been reformulated several times since (*e.g.* volley theory of pitch by Wever, 1949).
Click on the following links to see animations (.mov files) illustrating the phenomena of [neural firing synchrony](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/synchrony.mov), [phase locking](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/phaslock.mov), and the associated [volley theory of pitch](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/volley.mov) (from the '[Ear Works](http://www.neurophys.wisc.edu/h%26b/auditory/fs-auditory.html)' website at the University of Wisconsin, Madison).

Phase locking may occur at or below spontaneous firing rates of neurons, possibly assisting with signal detection below firing response threshold levels.

Whether based on the periodicity of the sound signal itself or the periodicity of the sound signal's representation in the auditory nerve (*e.g.* periodicity of neural spike activity), temporal models of pitch assume that the hearing mechanism extracts a signal's periodicity by performing an **autocorrelation function** on the signal.  Qualitatively, autocorrelation functions provide a measure of how well a signal matches a time-shifted version of itself, as a function of the amount of time shift.
Periodicity theories of pitch can explain the pitch of complex tones and phenomena such as the 'missing fundamental' but need to be modified to reliably address pitch of tones above 5kHz (value at which phase locking starts breaking down), phenomena such as the 'pitch-shift' effects, or the observation that our perception of pitch changes with changes in the physical/physiological properties of a listener's basilar membrane (*e.g.* *presbycousis*, related to BM hardening with age).

There are several pieces of evidence for and against both theories, and each theory can explain certain observations but not others. Contemporary hybrid theories of pitch incorporate aspects of both (Plomp, 1964; Terhardt, 1974; Pierce 1990). **Terhardt's theory of 'virtual pitch',** introduces 'previous learning' as a novel component to the theories of pitch, learning that can be based on either temporal or place pitch cues and which provides us with pitch templates that we use when determining the pitch of tones with ambiguous spectra (spectra with missing components, mistuned, *etc.*).  [E. Terhardt's webpage](http://www.mmk.e-technik.tu-muenchen.de/persons/ter.html), although not well designed, provides a wealth of information on pitch theory and perception and is highly recommended.

In general, it appears that the most salient/prominent (in terms of the relationship among loudness, frequency separation, and critical bandwidth) components of a signal's spectrum are the most important carriers of pitch information. In the presence of multiple complex tones, it appears that the components of a given complex tone are perceptually linked together into a single percept that is separate from the other complex tones thanks to similarities in the ways their amplitude and frequency values change with time (i.e. spectral jitter).

Wever (main figure in the 'periodicity theory' camp) and Békésy (main figure in the 'place theory' camp) were close friends. You can read their [short biographies](http://www.parmly.luc.edu/hearing/bekwev.htm), written by their student, Prof. R. Fay, of Loyola University, Chicago.
See a concise and systematic [historical review of pitch theories](http://acousticslab.org/psychoacoustics/PMFiles/PMDownloads/2004_IRCAMPitch.pdf) by Alain de Cheveigné.
(IRCAM, Paris - file originally at <http://recherche.ircam.fr/equipes/pcm/cheveign/pss/2004_ICA.pdf>).

Perceptual attributes of acoustic waves
Pitch (Part II)

|  |  |
| --- | --- |
|   Perception of pitch relations - Unit of pitch    **Mel:** Pitch-height unit and scale devised by S. S. Stevens (1937, 1940) and based on 'twice as high' perceptual judgments. 'Twice as high' amounts to a larger musical interval at high registers than at low registers.  Reference: 1000 mels = pitch of 1000Hz presented 40dB above threshold.(Is the Mel unit analogous to the Phon or the Sone unit of loudness?) | http://acousticslab.org/psychoacoustics/PMFiles/PMImages/mels.jpg |

Octave (pitch chroma / pitch height)

[Use this [virtual music](http://www.bgfl.org/bgfl/custom/resources_ftp/client_ftp/ks2/music/piano/index.htm) keyboard to help you with the concepts in the following section]

Notes separated by octave intervals are characterized (cross-culturally) by a high degree of sameness (if performed sequentially) and  smoothness (if performed simultaneously). This 'sameness' is referred to as: *pitch chroma*.

**Pitch chroma:** The distinctive quality of a specific tone, separating it from the rest of the tones within an octave. It describes perceptual 'differences'/'distances' of pitches within an octave and the perceptual sameness of pitches separated by one or more full octaves. It is reflected in the fact that the different note names (*e.g.* **C, D, E, F, G, A, B, C, D** ...) repeat periodically for every 2/1 increase in frequency (*i.e.* every octave) with the addition of a subscript (*e.g.* C4) to indicate how high or low this pitch is relative to some reference pitch.  In other words, a numeric subscript difference between two notes that share the same pitch chroma (*e.g.* C4 vs. C5) reflects a *pitch height* difference of one or more octaves between two notes.

**Pitch height:** term describing the perceptual 'highness' or 'lowness' of a pitch; it is related to frequency.
The intervals A3 (220Hz) - A4 (440Hz) and A4 (440Hz) - A5 (880Hz) are both octaves, with the three notes (A3, A4, A5) having the same pitch chroma but different pitch heights. In terms of their pitch height, octaves are equidistant perceptually although they are not equidistant in terms of Hz because, as indicated by Fechner's psychophysical law, perceptual and physical magnitudes of stimuli are related logarithmically. (*i.e.* Hertz versus Pitch; Intensity in W/m2 versus Loudness in phons or sones; *etc.*)
Notes represented by different names are different in both pitch chroma and pitch height. Within a single octave, the pitch raises in accordance with the note name (C3 is lower than D3, which is lower than E3, *etc.*). [Why, in your opinion, does octave numbering loop at the note C but note letters loop at the note A?]

Multidimensionality of Pitch (pitch spiral)

Pitch is multidimensional, involving pitch height (one dimension: frequency) and pitch chroma (two dimensions, separating pitches within an octave and linking pitches an octave apart).

**Note:** A variable is uni-dimensional if all its values can fit on a single straight line.
If for example A, B, & C represent 3 values of a variable as points in space then:
a) if |AB| + |BC| = |AC| then the variable is uni-dimensional
b) if |AB| + |BC| ≠ |AC| then the variable is multidimensional

|  |
| --- |
| **Pitch chroma circularity (octave)Pitch Spiral: pitch hight and pitch chroma** |
| The western chromatic scale breaks the octave down into 12 different pitch chromas. The pitch chroma dimensions represent a circularity in pitch perception. Due to this circularity, although the octave interval represents the largest physical distance (*within a single octave*) in terms of pitch height, it represents the smallest perceptual distance in terms of pitch chroma.  Pitch perception therefore *wraps* on the octave, with scales defining sets of different pitch chromas that repeat at different pitch heights for each new octave. This results in what may be called a pitch spiral.The perceptual circularity of pitch is explored in Shepard-tone scales/slides (after Roger Shepard) that present the paradox of a continuously ascending (or descending) pitch. |
| Escher | Shepard scales/slides are the auditory analog of the continuously ascending/descending staircases, explored conceptually by Penrose and artistically by Escher ([click here for more on Escher and his paintings](http://cybermuse.gallery.ca/cybermuse/youth/escher/home/home01_e.jsp)).Listen to two pitch spiral examples (Houtsma et al., 1987). |

**Tuning and music performance & perception**

**Although string players often follow the rather stretched (corresponding to the 'stretched' response of strings and the ear) Pythagorean tuning (with musical scales constructed by moving upwards in 5th intervals (3/2 ratio), until all twelve semitone intervals have been produced), while singers and trombone players often follow just tuning (also stretched but less than Pythagorean tuning), most performers adjust their intonation to the performance context. Regardless of tuning system, the perception of musical intervals seems to be categorical rather than continuous (at least for experienced listeners), tolerating minor deviations from ideal tunings.  In addition, performers seem to systematically stretch/contract intervals based on melodic and harmonic context.**

**In general, listeners appear to have more tolerance for stretched rather than contracted tuning.** **Click here for an example** **(the first melody uses compressed tuning, the second uses stretched tuning, and the third is mathematically correct; Houtsma *et al*., 1987).
Consistent with our preference for slightly stretched, rather than contracted tunings, listeners tend to prefer stretched over numerically precise or contracted octave intervals, and tend to reproduce them as such.**

Key to the 3-tone listening example:  DD  (the tones have frequencies 12025Hz, 12000Hz, and 11975Hz, in this order)

Columbia College, Chicago - Audio Arts & Acoustics