Chapter 2:
Sound Properties and the Waveform View

Introduction

- This chapter covers:
  - The relationship between perceptual and physical properties of sound.
  - Pitch and frequency, including the range of human hearing and Shepard tones.
  - Loudness and amplitude, including the range of human hearing for loudness expressed in decibels and the relationship between loudness and hearing damage.
  - Timbre and waveform, including basic waveforms such as sine, triangle, square, sawtooth, and noise.
  - Articulation and amplitude envelopes, including bowed/blown envelopes vs. struck/plucked envelopes.
  - Rhythm and amplitude transients, including their use in beat detection.

Sound Properties

- A sound wave is generated by some vibrating source, propagates through a medium as a series of compressions and rarefactions, and is finally received by our ears and brain.

Figure 2.1 A vibrating string produces a series of compressions and rarefactions that are received by the ear, coded as neural impulses, and sent to the brain.

- The mechanics of sound production doesn't provide much information about the sound itself.
- There are musical sound properties to keep in mind:
  - Pitch, loudness, timbre, articulation, and rhythm.
- Philosophical questions:
  - What is music, and what constitutes a "musical" sound?

- The **waveform view** of sound is useful when determining the properties for a given sound.
  - This is a graph of the change in air pressure at a particular location over time due to a compression wave.
  - A change in air pressure over time can be graphed on a simple x–y graph with time being the x and the air pressure being the y.

Table 2.1 **Motion of string and corresponding air pressure in front of string**

	Motion of string	*Air pressure*
1.	Not moving; in regular position	Normal
2.	Moving "forward" to furthest displacement	Higher than normal
3.	Moving "backward" to center position	Normal
4.	Moving "backward" to furthest displacement	Less than normal
5.	Moving "forward" to center position	Normal
. . .	. . . cycle repeats . . .	. . . cycle repeats . . .

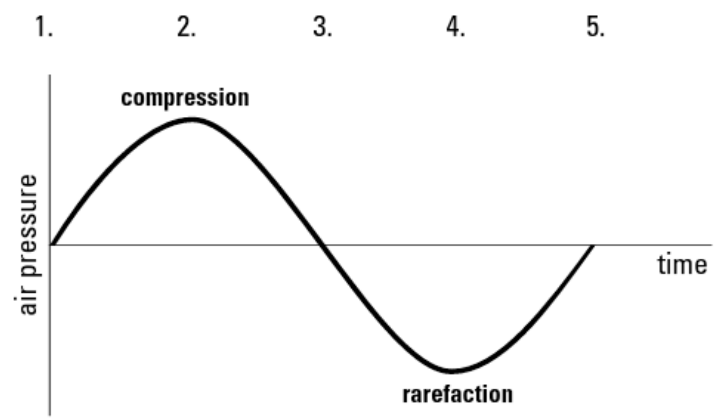

Figure 2.2 Waveform view of sound. The numbers above the graph correspond to the numbered string motions in Table 2.1.

Pitch

- **Pitch** is a measurement of sound made by your brain.
  - This is a perceptual property of sound.
- Because the waveform view shows what is happening to air molecules when they are disturbed by something that vibrates, this is a physical representation, not a perceptual representation of what is happening in your brain.
- The sound properties described earlier are all perceptual properties.
  - In order to use the physical waveform view to understand something about these properties, we need to identify physical properties that are related to them.
- Our ears change the physical properties of sound in various ways, such as emphasizing certain frequencies, and our brain analyzes the perceived sound properties in relation to perceptions it

has encountered before, such as identifying a sound as the harmonic interval of a perfect fifth played on a piano.

- The physical property that is related to pitch is **<u>frequency</u>**.
    - When it comes to the motion of a string, frequency is the rate at which a string moves through a full cycle of motions from center, to forward, to center, to backward, to center and then repeats.
        - These cycles of motion create compression and rarefaction cycles in the air that repeat at the same rate (frequency).
    - Although musical instruments have different physical motions, they all produce the necessary compression and rarefaction cycles at a rate related to the rate of their physical motions.
    - Frequency is measured by the number of cycles of compression/rarefaction that occur per second.
        - The cycles per second (cps) measurement is also referred to as **<u>hertz (Hz)</u>**.
        - This rate can be determined from the waveform view by measuring the amount of time the sound wave takes to go through a compression–rarefaction cycle.
            - This measurement is called the **<u>period</u>** of the waveform and is measured in seconds per cycle.
                - The letter "T" will stand for the period.
                - Period gives us the number of seconds per cycle, so frequency (*f*) is equal to the reciprocal of the period, T.
                    - This is also known as inverting the period.
                - As the period gets smaller, the frequency gets larger, and as the period gets larger, the frequency gets smaller.

$$f = 1 \div T = 1/T$$

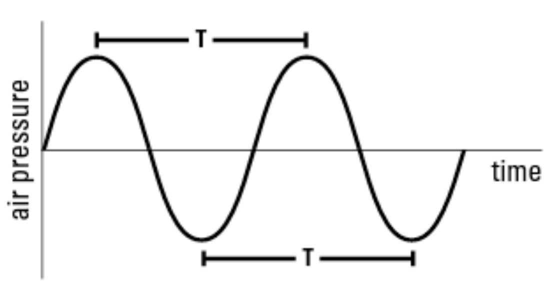

Figure 2.3a Waveform view showing the period of the waveform measured from two different starting points.

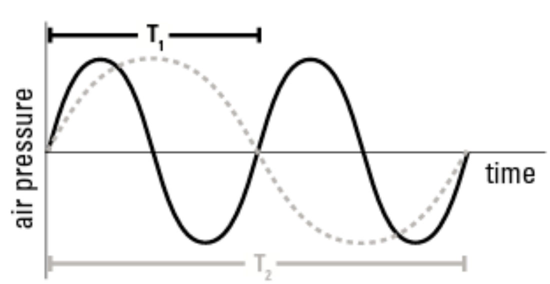

Figure 2.3b Two waves with the same amplitude but different frequencies graphed on the same axis. Period T_2 is twice as long as Period T_1, resulting in half the frequency.

Table 2.2 Some period–frequency relationships

Period	*Frequency*
2 seconds	1 ÷ 2 = 0.5 Hz
1 second	1 ÷ 1 = 1 Hz
½ second	1 ÷ ½ = 2 Hz
0.00227 seconds (2.27 milliseconds)	1 ÷ 0.00227 = 440 Hz (the tuning A)

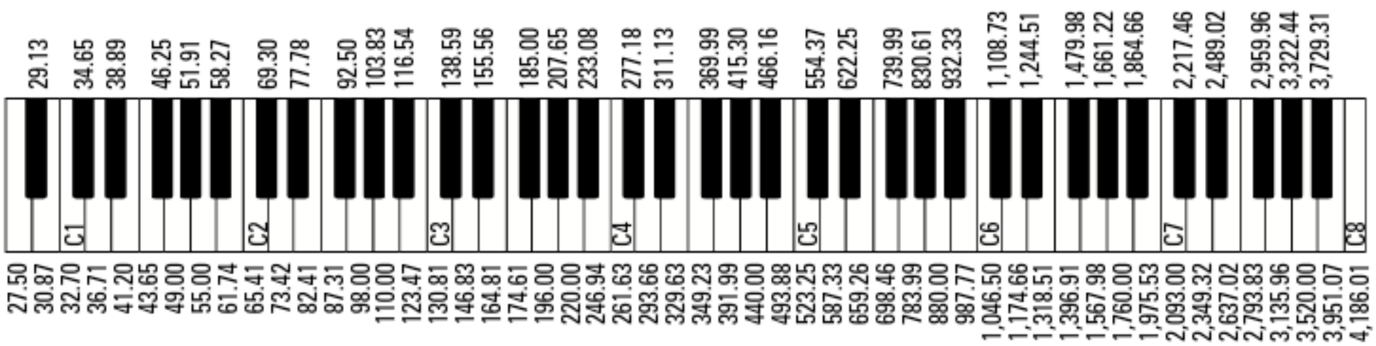

Figure 2.4 The frequencies associated with piano keys. C4 is middle C in this figure.

- While there are a wide range of frequencies that occur in the world, we are only sensitive to a certain range.
    - The frequency range of human hearing is about 20 Hz to 20,000 Hz (20 kHz).
        - Any back and forth motion of an object in a medium creates a compression wave at some frequency, but only those compression waves whose frequencies fall in the 20 Hz to 20 kHz range are really "sound waves."
        - This is an approximate range, and this range will vary from person to person.
            - This range can vary in the same person from year to year because as we age, our sensitivity to high frequencies gradually diminishes.
- Frequencies that are above our hearing range are referred to as **ultrasonic** frequencies.
    - A dog whistle is a good example of an object that produces frequencies this high.
- Frequencies that are below our hearing range are referred to as **infrasonic** frequencies.
    - Elephants can communicate at frequencies below our hearing range.
    - Some organ pipes produce infrasonic frequencies that are felt rather than heard.
- It is possible to fool the ear in various ways with regard to the relationship between frequency and pitch.

    - The cognitive psychologist, Roger Shepard (1929–2022), developed an illusion in which a series of tones appears to rise endlessly, but never leaves a relatively narrow range of frequencies.
        - The tones, in his honor, are referred to as **Shepard tones**.
        - The tones in this illusion are made up of a number of frequencies in octaves.
        - As all the frequencies rise, the higher frequencies gradually fade out and frequencies below them gradually fade in.
        - As the loudness relationships between the octaves change, our ears shift smoothly from the octaves that are fading out to the octaves that are fading in without us being consciously aware of it.
        - A good analogy to this auditory illusion is the visual illusion *Ascending and Descending*, created by M. C. Escher, which was inspired by a design by Lionel Penrose and Roger Penrose.

Loudness

- The perceptual property of **loudness** is related to the physical property of **amplitude**.
    - Amplitude (represented by the letter "A") is determined by how much the air pressure in a compression or rarefaction deviates from the normal air pressure.
        - In the case of a stringed instrument, the harder the string is plucked or bowed, the farther from the normal position the string moves and the greater the deviation in air pressure from the norm in the resulting compressions and rarefactions.
            - Struck instruments such as percussion generate greater amplitude in the same way.
        - For instruments driven by breath, the greater the airflow, the more the air molecules get packed together before the reed, lips, or vocal cords close, and the greater the amplitude of the resultant sound wave.
- On the waveform view, amplitude is measured from the x-axis to the peak (or the trough) so that it represents the deviation of air pressure from normal.

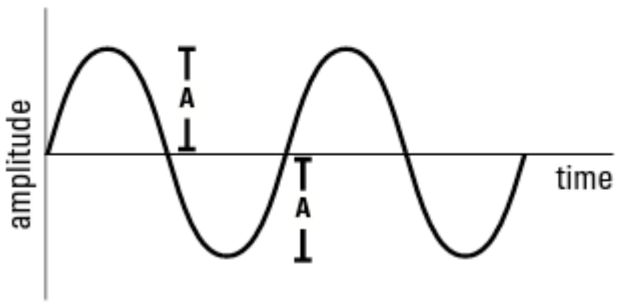

Figure 2.5a Waveform view showing the amplitude of the waveform measured in two different ways.

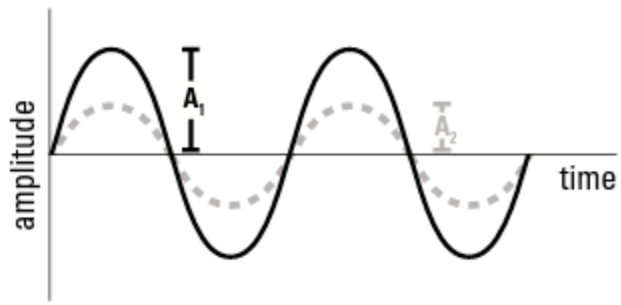

Figure 2.5b Two waves with the same frequency but different amplitudes graphed on the same axis.

- Amplitude measurements are given in relation to a reference value, resulting in a sound pressure level.
  - This level is expressed in units known as decibels (dB), or specifically **<u>decibels of sound pressure level (dB SPL)</u>**.
    - There are many different kinds of decibel measurements in music technology, which is why it is important to remember the “SPL” part of dB SPL.
- As with frequency, the range of human hearing for loudness is limited to just a part of the full range of possible sound pressure levels.
  - The quietest sound we can possibly hear is given as 0 dB SPL and is referred to as the **<u>threshold of hearing</u>**.
    - The “0” does not mean that there is no pressure in the sound wave, just that the sound pressure of the compression wave we’re measuring is the same as the sound pressure of a compression wave that was experimentally determined to be the quietest that humans can hear.
  - The loudest sound that we can bear is approximately 120 dB SPL and is referred to as the **<u>threshold of pain</u>**.
    - Anything above this is both physically painful and damaging to our hearing.
      - It is also important to note that even prolonged exposure to sound pressure levels significantly lower this can still cause hearing damage.

Table 2.3 Sound sources and related sound pressure levels

Sound source	Sound pressure level
Rock music peak	150 dB
Jet engine at 30 meters away	140 dB
Threshold of pain	120 dB
Symphonic music peak	120–137 dB
Amplified rock music at 1–2 meters	105–120 dB
Subway train at 60 meters away	95 dB
Piano played loudly	92–95 dB
Train whistle at 150 meters away	90 dB
Telephone dial tone	80 dB
Chamber music in small auditorium	75–85 dB
Piano played at moderate levels	60–70 dB
Normal conversation at arm's length	45–55 dB
Whisper	30 dB
Threshold of hearing	0 dB

Source: Data from Sallows (2001) and ASHA website.

Loudness and Hearing Damage

- Hearing is both a wondrous and fragile sense.
    - As we get older, our hearing naturally gets worse, with the higher frequencies slowly falling away.
    - Our modern noisy lifestyle can cause our hearing to deteriorate faster.
- Many people have had the experience of leaving a concert and noticing that their hearing is dull or muted.
    - This sensation, referred to as **temporary threshold shift**, often goes away after a few hours or days, but just because the dullness has left doesn't mean that there are no lasting effects.
        - The dullness comes from over-exciting the hair cells in the cochlea that are responsible for high frequency sounds.
            - If the sonic abuse is severe enough, the dull sensation may be a *permanent* threshold shift instead of a temporary one (permanent hearing damage).
- Persistent "ringing" in the ear, called tinnitus, is another possible outcome of hearing damage.

- While there are mechanical devices that can help you cope with the effects of hearing loss, including hearing aids and cochlear implants, there is currently no mechanical device that hears with the sensitivity of your natural hearing system.
    - The best solution to hearing loss is to prevent it in the first place.
        - The first and best way to prevent loudness-induced hearing loss is to avoid loud sounds.
            - There are a number of musical situations that require monitoring to make sure your ears aren't overexposed.

Table 2.4 **Maximum exposure time at various sound pressure levels**

Sound pressure level	*Max. exposure time*
82 dB	16 hours
85 dB	8 hours
88 dB	4 hours
91 dB	2 hours
94 dB	1 hour
97 dB	30 minutes
100 dB	15 minutes
103 dB	7.5 minutes

Source: Data from NIOSH (1998).

- There is concern that personal listening devices (PLDs), such as iPods, can contribute to hearing loss.
    - Guidelines have been published that suggest volume limitations and a time limit on PLD listening per day.
        - In the below rough guidelines, numbers are dependent on the volume output by the specific PLD, the fit of the earphones, and the individual's ears.
            - The greater number of dB SPL that a PLD outputs at a given percentage of maximum volume, the lower the exposure limit would be at that volume.

Table 2.5 **Recommended exposure limits for personal listening devices**

% of PLD max. volume	Exposure limits
10–50	No limit
60	18 hours
70	4.6 hours
80	1.2 hours
90	18 minutes
100	5 minutes

Source: Portnuff and Fligor (2006).

- When loud sounds can't be avoided, earplugs can help to protect your ears.
  - Inexpensive earplugs tend to change the balance of frequencies, and hence the timbre, of sounds that reach your middle and inner ears by reducing higher frequencies more than lower ones.
    - In musical situations, more expensive earplugs that are molded specifically to fit your ears can reduce the loudness of sounds evenly across all frequency ranges.
      - Consider this an investment in the longevity of your musical career and hearing.

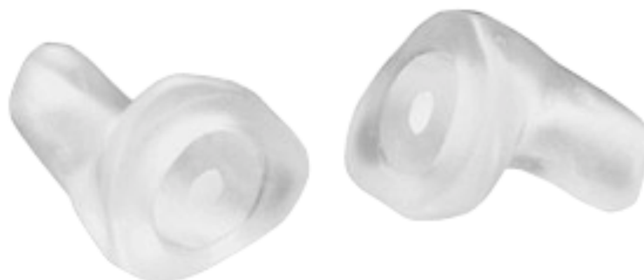

Figure 2.6 Custom-molded earplugs designed especially for music. Different filters can be inserted to reduce sound by 9 dB, 15 dB, or 25 dB. (Copyright © Etymotic Research Inc. Used with permission)

Amplitude and Loudness Perception

- Decibels are used when expressing sound pressure levels because they reduce a wide range of numbers down to a manageable range.
  - Our hearing is very sensitive, resulting in a ratio of sound pressure at the threshold of pain to sound pressure at the threshold of hearing of about 1,000,000 to 1.
    - Small changes in decibel values can reflect rather large changes in sound pressure.
- A discrepancy between physical measurements and perception is the difference in perceived loudness levels for sounds at different frequencies.
  - We are more sensitive to frequencies between about 1 kHz and 5kHz, so those sounds require less intensity, and hence fewer dB SPL, to sound as loud as lower frequency sounds.
    - A number of consonants in our language have significant energy in that range.

Timbre

- The perceptual property of **timbre** is related to the physical property of the shape of the wave, or the **waveform**.
    - Timbre is also related to the physical property of the sound's spectrum.
- Thus far in the discussion of sound, it has been assumed that the vibrating object is moving in the simplest possible way.
    - The shape produced by that simple back and forth motion is called a **sine wave** after the mathematical function that produces such a shape.
    - A real-world vibrating object seldom moves in such a simple fashion.
        - Typically, the back and forth motion will be more complicated, resulting in an equally complicated graph of the changing amplitude over time.

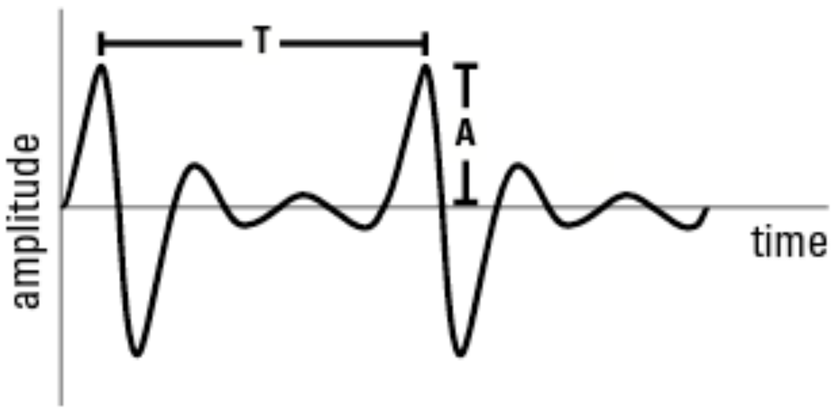

Figure 2.7 Two cycles of a trumpet waveform. Period and amplitude are indicated.

- Because it is difficult to make many generalizations about the waveforms of actual instruments, the waveform view is a bit limited in what it can tell us about timbre.
    - Timbre is a complicated phenomenon and can be influenced by the other sound properties (pitch, loudness, articulation) and the overall sonic context (what other instruments are playing, whether it is noisy, etc.).
- There is a collection of largely artificial waveforms that can be used as a sort of rudimentary timbral vocabulary.
    - The simplest is the sine wave.
    - The other standard waveforms are the **triangle wave**, the **sawtooth wave**, the **square wave**, and a version of the square wave called a **pulse wave**.

  - These waveforms formed the basis for early analog synthesizer sounds, and many current software synthesizers and some hardware synths use analog-modeling techniques to create sound.
  - The term "analog-modeling" is used to describe digital synthesis methods that are designed to mimic original analog synthesis techniques, usually with some modern digital twists.

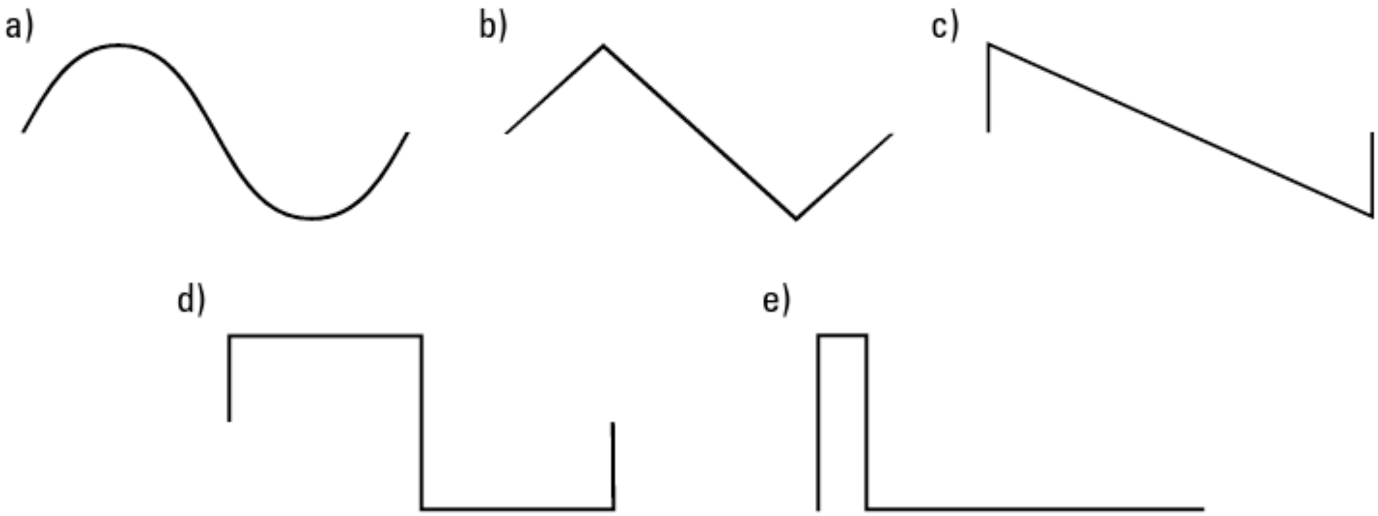

Figure 2.8 Basic waveforms: (a) sine, (b) triangle, (c) sawtooth, (d) square, and (e) pulse.

- Another standard waveform is one that has no regular pattern at all: **noise**.
  - Noise is an important sound component during the attack phase of almost all instruments.
  - Noise was also an important sound source in analog synthesizers and is used today in analog and analog-modeling synthesis.
  - Noise can have a variety of qualities that are usually described using colors, such as **white noise** (very harsh and grating) and **pink noise** (still noisy, but more pleasant).
  - Noise does not have a predictable amplitude pattern, so the below waveforms are only representative.

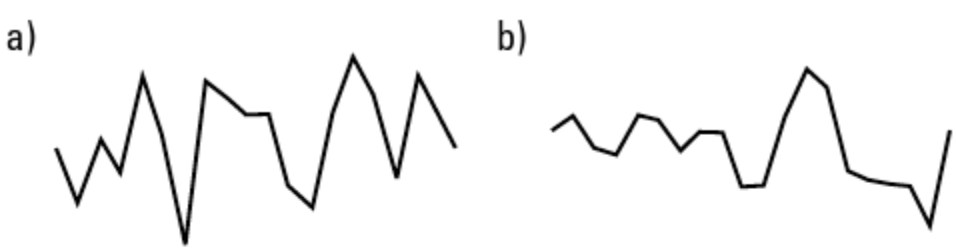

Figure 2.9 Amplitude plots of noise: (a) white noise, and (b) pink noise. The duration for each is about the same as the period of a 440 Hz periodic waveform.

Phase

- When you're looking at a waveform, it may actually start in a different position than an amplitude at zero.
- The position at which the waveform starts is determined by its phase.
  - Phases may be given in degrees with a phase shift of 360 degrees returning the waveform to its normal starting point.
    - In the below example, (a) shows a sine wave starting conventionally with the amplitude at zero and then rising; this corresponds to an initial phase of 0 degrees.
    - (b) shows the sine wave starting with an amplitude at the maximum and then falling; this corresponds to a phase difference of one-quarter of the waveform's period, or 90 degrees.
      - This is equivalent to a cosine wave.
    - (c) shows a sine wave starting with an amplitude of zero and then falling; this corresponds to a phase difference of one-half of the waveform's period, or 180 degrees.

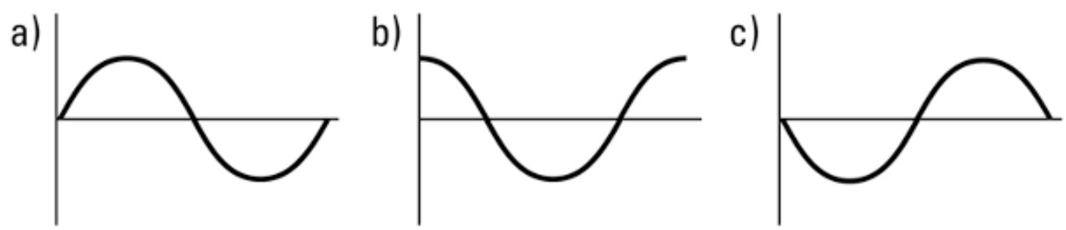

Figure 2.10 Sine wave phases: (a) zero degrees, (b) 90 degrees, (c) 180 degrees.

- Phase becomes important when waveforms are combined together.
  - Combining a sine wave at some frequency with a second sine wave at the same frequency but 180 degrees out of phase results in the two sine waves canceling each other out (**phase cancellation**) and no sound.
    - Phase cancellation can occur when audio recordings are mixed together and when audio is processed through effects.
      - Phase cancellation is an important part of the phaser effect.
        - A phaser changes only the phases of the partials in the audio that it's processing that, when combined with the original signal, results in controllable phase cancellation that colors the sound in an interesting way.

Articulation

- The perceptual property **articulation** refers to how the loudness of the sound changes over time.
    - For example, the loudness of an accented note rises more quickly from science and to a higher maximum loudness than a note that is not accented, and a note that is staccato will have a quick rise and a quick fall off at the end.
- Articulation is not just limited to musical notes; the loudness of the non-musical sounds around us also changes over time.
    - A thunderclap has a sudden jump in loudness followed by a long fall away.
    - A motorcycle roaring toward you has a long, slow increase in loudness followed by a long slow decrease as it passes you and roars on.
- When loudness was discussed earlier, it was related to the amplitude of the individual cycle of a waveform, whose duration is quite short.
    - For example, the period of A 440 Hz (the tuning A) is just over 2 milliseconds in length.
    - The changes in loudness referred to as articulation are taking place over much larger spans of time.
        - Example: An eighth note (at quarter note = 120 bpm) is 0.25 seconds long, which is over 100 times as long as the period of the individual waveform at A 440 Hz.
            - You could fit over 100 cycles of a waveform whose frequency is 440 Hz into that eighth note.
            - Even at the lowest frequency that humans can hear, 20 Hz, the period is 0.05 seconds; you could fit five of those cycles into that eighth note.
- The physical property that is related to articulation is referred to as an **amplitude envelope** because it contains or envelops many repetitions of the waveform.
    - To represent the amplitude envelope, we can continue to use the waveform view (amplitude vs. time), but now we "zoom out" to look at changes in amplitude at the timescale of the note.

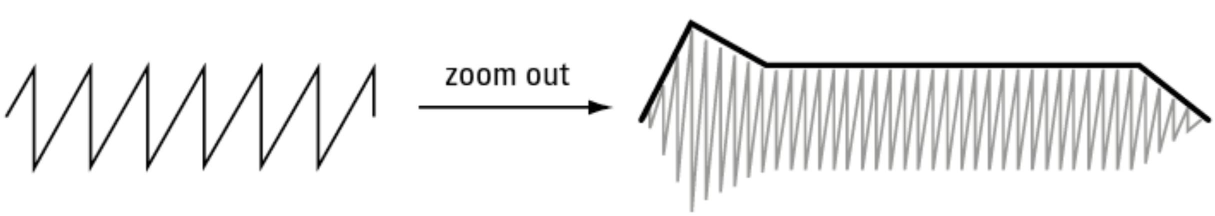

Figure 2.11 A "zoom out" from individual cycles of a waveform to see the amplitude envelope. Only the top of the envelope is usually shown because many waveforms are the same on the top and on the bottom. The frequency is extremely low here—20 Hz—so you can still see the individual waveforms within the envelope.

"Bowed or Blown" Envelopes

- As with waveforms, every instrument's amplitude envelope is a little bit different and it changes depending on how the notes are articulated.
    - However, we can make the distinction between instruments that are "bowed or blown" and instruments that are "struck or plucked."
- Instruments that are bowed or blown can be modeled by an **attack-decay-sustain-release envelope (ADSR)**.
    - This envelope also has its roots in analog synthesis and is widely found in various forms on hardware and software synthesizers.
    - The attack-decay segments roughly model the beginning of an instrument's note, where the amplitude rides from silence to an initial peak and then falls somewhat to a sustained level.
    - The difference in amplitude between the peak and the sustain levels reflects the degree of initial accent on the note.
        - A strong accent will have a greater peak and a greater fall off from the peak to the sustain level, whereas a note that is attacked more gently will have a lower peak and a smaller difference between those two levels.
    - The sustain segment is the most characteristic segment for this envelope model.
        - Only instruments in which the performer continuously supplies energy to the instrument by blowing, bowing, or some other means, will have such a sustain segment.
    - The release portion of the envelope reflects how the sound falls away from the sustain level to silence.
        - In the envelope for a staccato note, this transition will happen very quickly (very small release time).
        - If a note is allowed to "ring into the hall," the transition will take longer (longer release time).

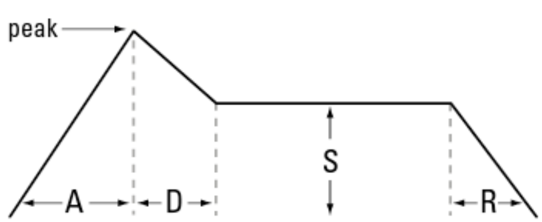

Figure 2.12 An attack–decay–sustain–release (ADSR) amplitude envelope characteristic of bowed or blown instruments.

"Struck or Plucked" Envelopes

- Instruments that are struck or plucked can be modeled by an **attack-release envelope (AR)**.
  - With a struck or plucked instrument, the performer initially imparts energy to the instrument and then allows the vibrations to damp down naturally.
  - The duration of the attack segment reflects the force with which the instrument is activated.
    - It can also reflect the materials that impart the initial impulse.
  - The duration of the release portion of the envelope is related to how hard the instrument was struck or plucked.
    - The harder the strike, the longer the release.
  - The size and material of the vibrating object also impact the release.
    - A longer string or larger drumhead is likely to vibrate longer than short strings and small drumheads.

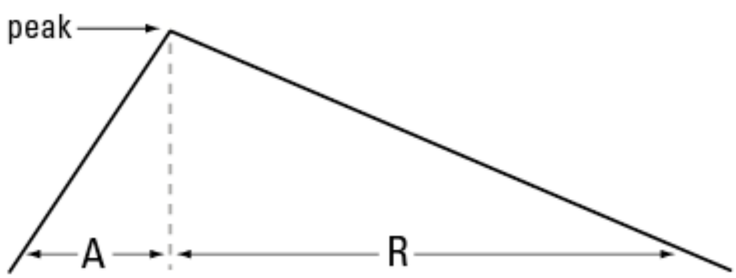

Figure 2.13 An attack–release (AR) amplitude envelope characteristic of struck or plucked instruments.

- These envelope models are not necessarily mutually exclusive.
  - Brass and woodwind instruments can be played with a "bell-like tone."
  - Individual notes in a rapid succession of plucked or struck tones will not seem to be significantly different in articulation from individual notes in a rapid succession of bowed or blown tones.
- One significant difference between the two envelopes is that struck or plucked tones cannot truly be "slurred" and will always have an attack of some kind, whereas the pitch of bowed or blown notes can be changed while the energy is still being imparted to the instrument, allowing the new note to "take over" the envelope of the previous note.
  - In struck or plucked notes, the release of one note can overlap the attack of the next, but the envelope will still be articulated for each note.
    - Example: A pianist's legato technique in which the notes are slightly overlapped to create smooth note-to-note transitions.

- Many hardware and software synthesizers do not have separate controls for these two envelope types.
  - However, the AR envelope can be simulated with an ADSR envelope in which the sustain level is set to 0.
    - The "release" of the AR will then either be determined by the ADSR's decay or a combination of the ADSR's decay and release lengths, depending on the length of the decay segment and how long the note is held down.
- Many synths allow for more complex envelopes with more segments (multiple attacks, multiple decays) and/or different line shapes for the segments.

Rhythm

- **Rhythm** is a perceptual property whose physical counterpart is complex, because rhythm is really a meta-property consisting of multiple notes or sound.
  - There are different levels of rhythm from a single sound that has its own internal rhythm, such as a "bubbling" synthetic sound, to a group of sounds/notes forming a rhythmic pattern, to a group of rhythmic patterns forming a phrase, and so on.
- At the level of a group of sounds/notes, aspects of rhythm can be seen in the waveform view by identifying patterns in the attacks of the notes, referred to as **transient patterns**.
  - The term transient is used because the attack-decay portions of an envelope form a short-term transition from no sound to the sustain or release of a sound.
  - Some sound types, such as drums, form patterns that have strong transients and no sustain, whereas other sounds, such as slurred woodwinds, brass, or strings, form patterns in which the transient can be difficult to see.
  - Viewing transients in the waveform view involves even more "zooming out" than with amplitude envelopes.

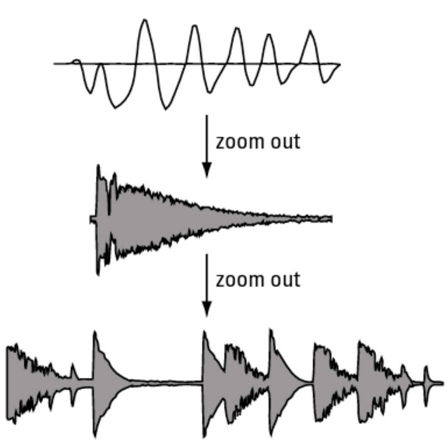

Figure 2.14 A "zoom out" from individual cycles of the waveform of a percussive sound to the amplitude envelope of that single event (AR envelope) to a pattern of percussive transients.

- The analysis of transients as a pattern of beats and bars is a standard gesture in many recording programs and is generally referred to as **beat detection**.
    - This process allows you to manipulate audio as separate logical chunks, the same way you can manipulate notes.
        - These audio beats can be "snapped" to bar and beats, thereby allowing the tempo of the audio to change in a naturally musical way.
    - Many recording programs can extract "groove" information—variations in transient timing and amplitude—and then apply it to other audio files and to MIDI messages.

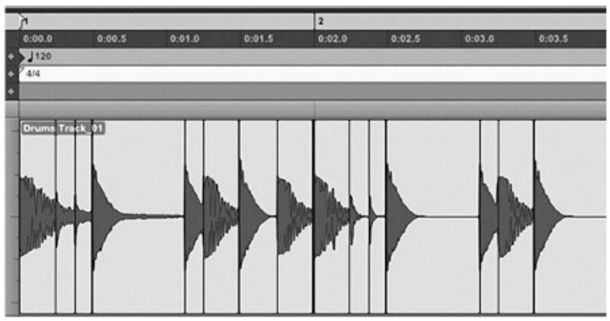

Figure 2.15 The results of transient analysis. The vertical lines indicate identified beats. (Screenshot is provided courtesy of Avid Technology, Inc.)

- Sound that consists of clearly defined transients in a regular pattern is easier for humans and for software to parse into beats and bars.
    - Legato passages where the notes are not re-attacked, and hence have fewer transients, are more difficult for software that relies solely on transient detection to parse.
        - Our perceptual system is more successful here because, in addition to transient detection, we can also bring pitch detection and other techniques to bear on the problem.

Sound Property Summary

Table 2.6 summarizes the perceptual sound properties and their physical counterparts. This table will be further refined in the following chapter.

Table 2.6 **Perceptual and physical properties of sound**

Perceptual properties	*Physical properties*
Pitch	Frequency
Loudness	Amplitude
Timbre	Waveform
Articulation	Amplitude envelope
Rhythm	Transient patterns
