Audio Signals

Stephen Boyd

ENGR108
Stanford University

September 6, 2021
mean atmospheric pressure is around $10^5 \text{N/m}^2$

**acoustic pressure** $p(t)$ is instantaneous pressure minus mean pressure

we perceive small fast variations in $p(t)$ as sound

$rms(p)$ corresponds (roughly) to loudness of sound

$rms(p) = 1 \text{N/m}^2$ is ear-splitting ($\sim 120$ dB SPL)

$rms(p) = 10^{-4} \text{N/m}^2$ is barely audible ($\sim 14$ dB SPL)

Sound Pressure Level (SPL) of acoustic pressure signal $p$ is

$$20 \log_{10}(rms(p)/p_{ref})$$

$p_{ref} = 2 \times 10^{-5} \text{ N/m}^2$
Vector representation of audio

- vector $x \in \mathbb{R}^N$ represents audio (sound) signal (or recording) over some time interval
- $x_i$ is (scaled) acoustic pressure at time $t = hi$:
  \[ x_i = \alpha p(hi), \quad i = 1, \ldots, N \]
- $x_i$ is called a sample
- $h > 0$ is the sample time; $1/h$ is the sample rate
- typical sample rates are $1/h = 44100/\text{sec}$ or $48000/\text{sec}$ ($h \approx 20\mu\text{sec}$)
- for a 3-minute song, $N \sim 10^7$
- $\alpha$ is scale factor
- stereophonic audio signal consists of a left and a right audio signal
Examples

Instrumental (play)

Speech (play)
Scaling audio signals

- If $x$ is an audio signal, what does $ax$ sound like? ($a$ is a number)
- Answer: same as $x$ but louder if $|a| > 1$ and quieter if $|a| < 1$
  - $2x$ sounds noticeably louder than $x$
  - $(1/2)x$ sounds noticeably quieter than $x$
  - $10x$ sounds much louder than $x$
  - $-x$ sounds the same as $x$

- A volume control simply scales an audio signal
- For this reason, the scale factor usually doesn’t matter
- Example
  - Play $x$
  - Play $2x$
  - Play $(1/2)x$
  - Play $-x$
Linear combinations and mixing

- Suppose \( x_1, \ldots, x_k \) are \( k \) different audio signals with same length.
- Form linear combination \( y = a_1 x_1 + a_2 x_2 + \cdots + a_k x_k \).
- \( y \) sounds like a mixture of the audio signals, with relative weights \(|a_1|, \ldots, |a_k|\).
- Forming \( y \) is called mixing, and \( x_i \) are called tracks.
- Producers do this to produce finished recordings from separate tracks for vocals, instruments, drums, \ldots
- Coefficients \( a_1, \ldots, a_k \) are adjusted (by ear) to give a good balance.
- Typical number of tracks: \( k = 48 \).
Mixing example

▸ tracks
  – drums (play)
  – vocals (play)
  – guitar (play)
  – synthesizer (play)

▸ mix 1: $a = (0.25, 0.25, 0.25, 0.25)$ (play)
▸ mix 2: $a = (0, 0.7, 0.1, 0.3)$ (play)
▸ mix 3: $a = (0.1, 0.1, 0.5, 0.3)$ (play)
suppose $p(t)$ is an acoustic signal, with $t$ in seconds
it is periodic with period $T$ if $p(t + T) = p(t)$ for all $t$
(in practice, it’s good enough for $p(t + T) \approx p(t)$ for $t$ in an interval at least 1/8 second or so)
its frequency is $f = 1/T$ (in 1/sec of Hertz, Hz)
for $f$ in range 100–2000, $p$ is perceived as a musical tone
  – frequency $f$ determines pitch (or musical note)
  – shape (a.k.a. waveform) of $p$ determines timbre (quality of sound)
Musical notes

- $f = 440\text{Hz}$ is middle A
- one octave is doubling of frequency
- $f = 880\text{Hz}$ is A above middle A; $f = 220\text{Hz}$ is A below middle A
- each musical half step is a factor of $2^{1/12}$ in frequency
- middle C is frequency $f = 2^{3/12} \times 440 \approx 523.2\text{Hz}$ (C is 3 half-steps above A)
- in Western music, certain consonant intervals have frequency ratios close to ratios of small integers
**Frequency ratios and musical intervals**

<table>
<thead>
<tr>
<th>half steps</th>
<th>name</th>
<th>frequency ratio</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>unison</td>
<td>$2^{0/12} = 1$</td>
<td>play</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>$2^{1/12} = 1.0595$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>$2^{2/12} = 1.1225$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>minor 3rd</td>
<td>$2^{3/12} = 1.1892 \approx 6/5$</td>
<td>play</td>
</tr>
<tr>
<td>4</td>
<td>major 3rd</td>
<td>$2^{4/12} = 1.2599 \approx 5/4$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>perfect 4th</td>
<td>$2^{5/12} = 1.3348 \approx 4/3$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>$2^{6/12} = 1.4142$</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>perfect 5th</td>
<td>$2^{7/12} = 1.4983 \approx 3/2$</td>
<td>play</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>$2^{8/12} = 1.5974$</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>$2^{9/12} = 1.6818$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>$2^{10/12} = 1.7818$</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>$2^{11/12} = 1.8877$</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>octave</td>
<td>$2^{12/12} = 2$</td>
<td>play</td>
</tr>
</tbody>
</table>
Periodic signals

- periodic signal (with period $1/f$, frequency $f$)

\[ p(t) = \sum_{k=1}^{K} (a_k \cos(2\pi f kt) + b_k \sin(2\pi f kt)) \]

- $k = 1$ terms are called the fundamental
- for $k > 1$, $k - 1$ is called harmonic or overtone
- $a_k, b_k$ are harmonic coefficients
- any periodic signal can be approximated this way (Fourier series) with large enough $K$
timbre (quality of musical tone) is determined by harmonic amplitudes

\[ c_1 = \sqrt{a_1^2 + b_1^2}, \ldots \quad c_K = \sqrt{a_K^2 + b_K^2} \]

- \( c = (1, 0, \ldots, 0) \) (pure sine wave) is heard as pure, boring tone
- \( c = (0.3, 0.4, 0.2, 0.3) \) has same pitch, but sounds ‘richer’
- with different harmonic amplitudes, can make sounds (sort of) like oboe, violin, horn, piano, \ldots
Various timbres, same pitch

pure 220hz tone, $c = 1$ (play)

$c = (0.7, 0.6, 0.3, 0.04)$ (play)
Various timbres, same pitch

\[ c = (0.21, 0.4, 0.9, 0.05, 0.05, 0.05) \] (play)

\[ c = (0.3, \ldots, 0.3) \in \mathbb{R}^{10} \] (play)