Audio Signals

Karanveer Mohan and Stephen Boyd

EE103
Stanford University

September 23, 2015

Acoustic pressure

- mean atmospheric pressure is around 10^5N/m^2
- *acoustic pressure* $p(t)$ is instantaneous pressure minus mean pressure
- we can only hear variations in $p(t)$ on submillisecond and millisecond time scale
- $\mathbf{rms}(p)$ corresponds (roughly) to loudness of sound
- $\mathbf{rms}(p) = 1$N/m^2 is ear-splitting (~ 120 dB SPL)
- $\mathbf{rms}(p) = 10^{-4}$N/m^2 is barely audible (~ 14 dB SPL)
- Sound Pressure Level (SPL) of acoustic pressure signal p is $20\log_{10}(\mathbf{rms}(p)/p_{\rm ref})$, $p_{\rm ref} = 2 \times 10^{-5}$ N/m^2

Vector representation of audio

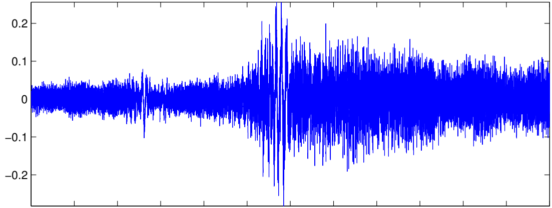
- vector $x \in \mathbf{R}^N$ represents audio (sound) signal (or recording) over some time interval
- x_i is (scaled) acoustic pressure at time $t = hi$:

$$x_i = \alpha p(hi), \quad i = 1, \ldots, N$$

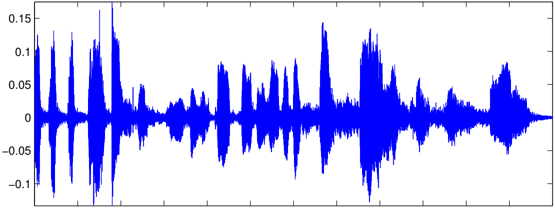
- x_i is called a *sample*
- $h > 0$ is the sample time; $1/h$ is the sample rate
- typical sample rates are $1/h = 44100$/sec or 48000/sec ($h \approx 20\mu$sec)
- for a 3-minute song, $N \sim 10^7$
- α is scale factor
- *stereophonic* audio signal consists of a left and a right audio signal

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 $x_1, \dots, x_k$ 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 $|a_1|, \dots, |a_k|$
- 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, ...
- coefficients $a_1, \dots, a_k$ 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)

Musical tones

- 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 $1/4$ 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

- $f = 440$Hz is middle A
- one *octave* is doubling of frequency
- $f = 880$Hz is A above middle A; $f = 220$Hz is A below middle A
- each musical *half step* is a factor of $2^{1/12}$ in frequency
- middle C is frequency $f = 2^{3/12} \times 440 \approx 523.2$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

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

Periodic signals

- periodic signal

$$p(t) = \sum_{k=1}^{K} \left(a_k \cos(2\pi f k t) + b_k \sin(2\pi f k t)\right)$$

- k is called *harmonic* or *overtone*
- f is frequency
- a_k, b_k are *harmonic coefficients*
- any periodic signal can be approximated this way (Fourier series) with large enough K

Timbre

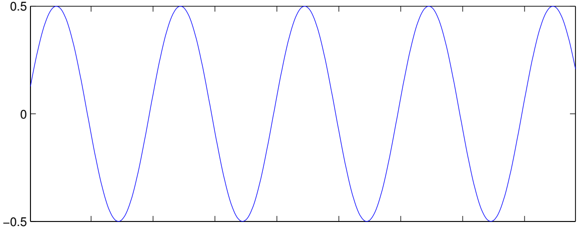
- timbre (quality of musical tone) is determined by *harmonic amplitudes*

$$c_1 = \sqrt{a_1^2 + b_2^2}, \quad \ldots \quad c_K = \sqrt{a_K^2 + b_K^2}$$

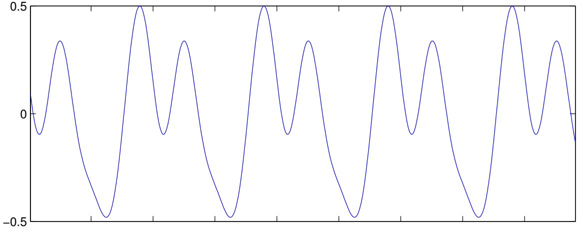
- $c = (1, 0, \ldots, 0)$ (pure sine wave) is heard as pure, boring tone
- $c = (0.3, 0.4, 0.2, 0.3)$ has same pitch, but sounds 'richer'
- with different harmonic amplitudes, can make sounds (sort of) like oboe, violin, horn, piano, ...

Various timbres, same pitch

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

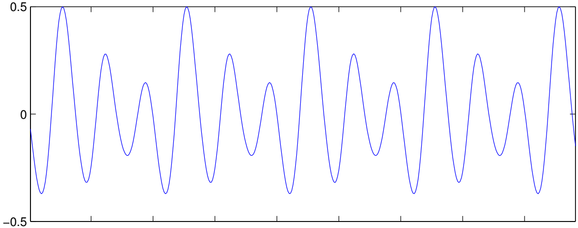


$c = (0.7, 0.6, 0.3, 0.04)$ (play)



Various timbres, same pitch

$c = (0.21, 0.4, 0.9, 0.05, 0.05, 0.05)$ (play)



$c = (0.3, \ldots, 0.3) \in \mathbf{R}^{10}$ (play)