2001-04-16
dpwe@ee.columbia.edu

Speech & Audio Processing & Recognition

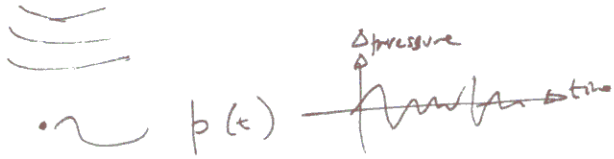
http://www.ee.columbia.edu/~dpwe/courses/e6820-2001-01/

What's it about?

SOUND — Mechanical vibration

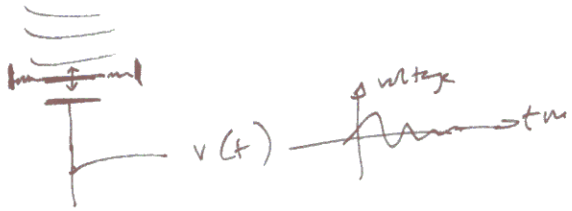
↓ air moves

↓ radiating wave (Δpressure)



SOUND: pressure variations at a point in space

↓ push against surface (diaphragm of mic)

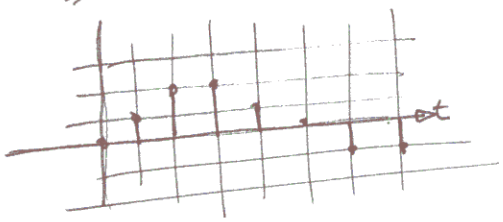


↓ voltage v(t)

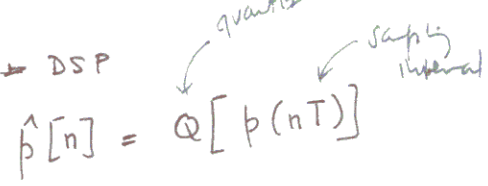
↓ processing

DIGITAL: representing sound on a computer → series of numbers

⇒ discrete time, discrete value



→ DSP



+ Can convert back to sound:

Speaker passes current through magnetic field
→ force
→ moves "cone"
→ pushes air
→ radiated wave . . .

Why Sound? because we hear

Why hearing? because it's useful (evolutionary advantage)

- info at a distance
- consistent / reliable
- correlates with physical force
- complements vision
- common in higher animals

footsteps ex →

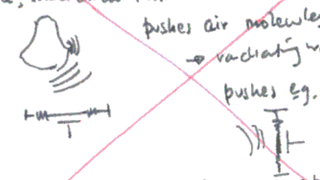
- "what" info? depends on what we care about → ψ

2001-01-15
dpwe@ee.columbia.edu
SAPR – Intro ②

What is Sound?

- air vibration → small changes in pressure

hence, mechanical motion



→ electrical signal ∝ pressure $p(t)$

Computers
- sample $p(t)$ at discrete times
- quantize to finite levels → $p[n] = Q(p(nT))$

Generate sound by moving speaker ∝ voltage

Speech & Audio Processing & Recognition

is founded on Acoustic Communication:

eg. source-channel model:

message → signal → channel → receiver → decoder

(speech [illegible]) — synthesis — Audio (processing) — recognition

---

~~[illegible]~~

~~Audio is interesting because we hear~~

~~We hear because it's useful~~

~~i.e. evolutionary advantage~~

- ~~information at a distance~~
- ~~relatively consistent / reliable~~
- ~~correlated with physical force~~
- ~~complementary to vision.~~
- ~~universal ?~~

~~⇒ what we hear reflects what is useful~~

ECOLOGICAL constraints.

What is the scope of audio processing?

(?)

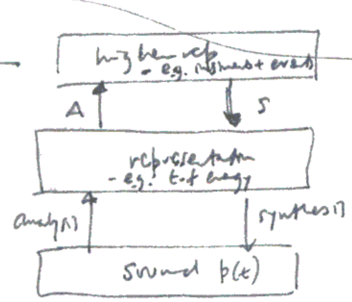
(info) subset of audio

	Sp synth	GSM etc	Vx mod	ASR
speech				
music – sing – jazz – rock	• synths • phys. models	WBAC	. reverb . time scale mod … SFX	• music ID • transcription
AUDIO env. – footsteps – cars – wind	• foley • models	⋮	⋮	• robots • Audio CBR
synth – classic – SFX	• Sound design			• alarm detect.
	synthesis	compression	modification	recognition

recognition: classify, understand (subset)

PROCESSING

Representation



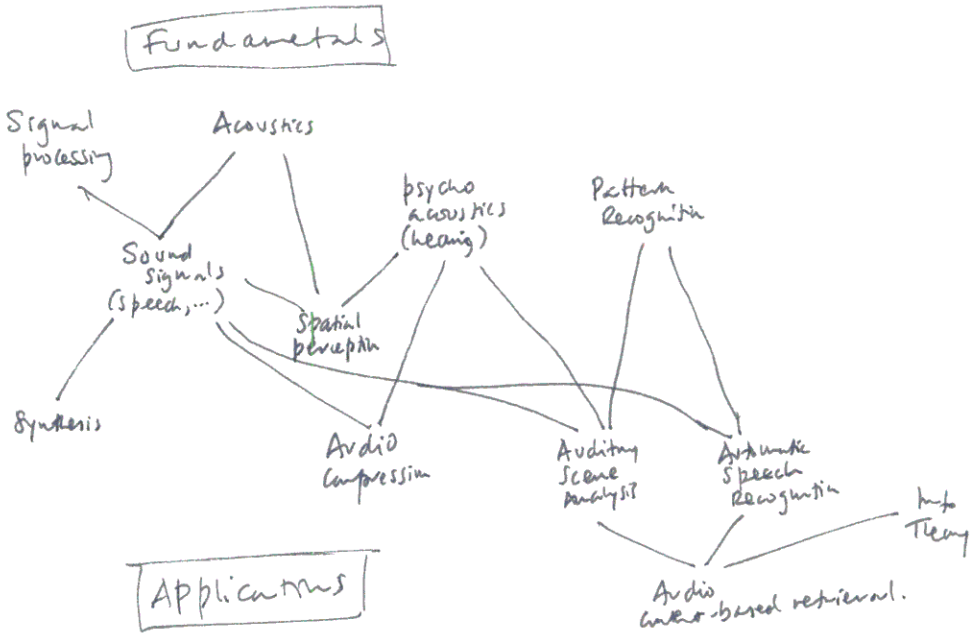
General info proc model

examples: speech recog
music synth

Representations:
- permit modifications (pitch + timing)
- make something useful explicit (spectrogram)
- efficient (compress)

2001-01-15
dpwe@ee.columbia.edu
SAPR - Intro (2)

Course overview



Goals

(1) Audio Proc - tools & techniques

(2) Advanced DSP - example applications
- specific techniques
} apply beyond audio

(3) Practical experience - computer examples
- reading papers
- ~~projects~~ projects & HW

Structure

(1) ~~Lectures~~ Class
- lectures
- demonstrations
- discussions < papers / ideas
- project stuff
} student presentations

(2) Homework - book problems
- computer problems - Matlab
- paper summaries
- build up a web site

+ exams - fundamentals
- qualitative

(3) Project - any specific aspect (some ideas ...)
- lit review
- PRACTICAL - Matlab (eg.)
- several pieces
- evaluation
- presentation (midterm?)
- start early - ideas now!

HW 1
- GTM info
- Matlab:
  - wavfile
  - plot
  - time of day
- initial prog. struggles
- some F+G

2001-01-14
dpwe@ee.columbia.edu
SAPR - Overview (1)

Course overview - summarize some of the key points
→ see where we're going
→ start thinking

- Slidepack
- Sndex

(1.) Speech signal
- waveform - timescale - sounds
- spectrogram
- t/f tradeoff
- cup bubbles

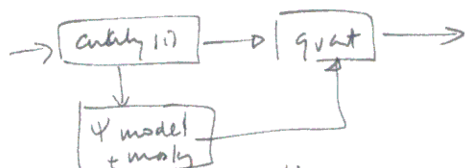
Spectrogram
- t-f energy c/w ear — t vs. f tradeoff.
- variation in time
- phonemes - stops, vowels, fricatives, glides
- linguistic variation.

- sgram pair
- segmentation.
+ wavesurfer
- timit examples → seg
- t/f tradeoff

RT sgram
SNACK?

→ ψ ? • SWS
• restoration
• vowel seq.
• pulsation

(2) WBAC : Structure:



- coding : samples = rate + bits
- Q noise
- subband → channel matching
- ψ masking - ~~chan~~ noise shaping

- examples at various levels + sgrams.

- blk diag
- masking image.
dspB
+ Sndex - SR
- NMR
- basic mpeg example
- vary bit rates
- look at sgram.

(3) ASR - basic blocks
- component models words → grammar → prons → acous phones
basic eq^n $W^* = \max_M (p(X|M))$
SEARCH

- ASR intro slides
+ SPRACHdemo
- clean example
- ...plus added noise

(4) Audio CBR - mixtures
- search ⟨ by example / by attributes
- muscle fish

- musclefish slide
+ demos
- musclefish demo.