ECE438 Lecture Friday March 20, 2023

New topic: speech processing

VOIP - current techniques are quite primitive

Module 4.1.1

1. Analysis-only
   a. speech recognition
   b. speaker verification
   c. speaker identification

2. Synthesis-only
   a. speech-based human-computer interaction
   b. customer support via phone
   c. text-to-speech
      (useful for visually impaired individuals)

3. Analysis followed by synthesis
   a. digital telephony
      1. VOIP
      2. mobile telephony
      3. (and line telephony
   b. speech recording & playback
   c. interactive speech systems

1. Alexa
2. Siri

Speech processing approaches

1. Waveform coding

   doesn't take into account specific categories of speech i.e works for any audio waveform

   examples:

   a. PCM
      i) quantize
      ii) convert to binary

   b. DPCM

   c. ADPCM

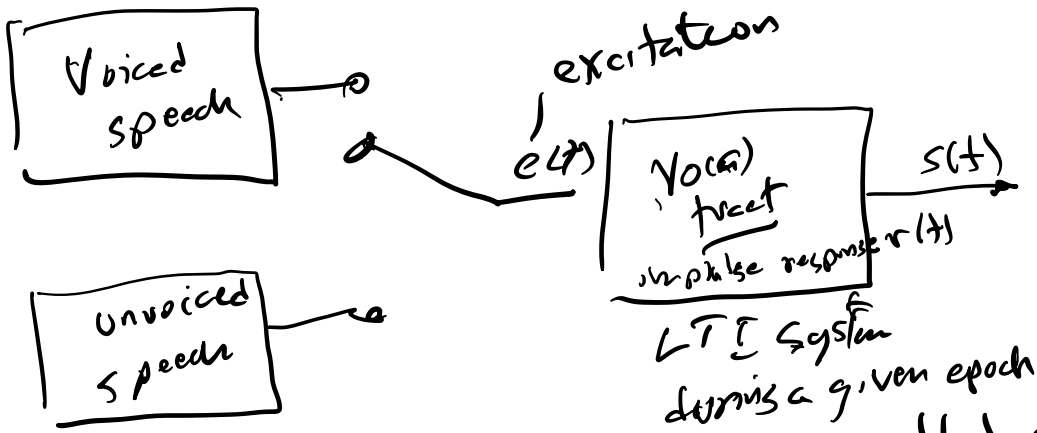
   d. Filterbanks and wavelets

      (will discuss this in ECE 439

2. Speech-based models

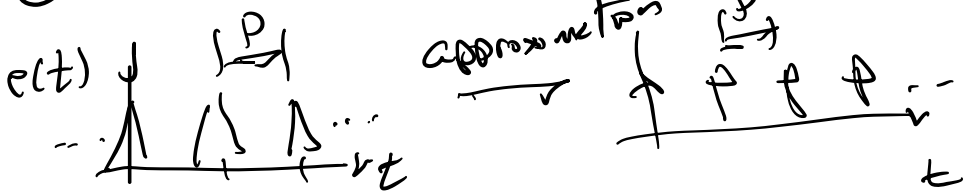
   a. exploit specific characteristic of speech (spoken English)

Simple model for spoken languages (primarily English)



interval for a given phoneme is called an epoch

Consider a voiced phoneme



$\rightsquigarrow e(t) = \text{rep}_P[\delta(t)]$

$E(f) = \frac{1}{P}\,\text{comb}_{\frac{1}{P}}[1] = \frac{1}{P}\sum_{k=-\infty}^{\infty}\delta\left(f - \frac{k}{P}\right)$

$\Rightarrow s(t) = v(t) * e(t)$

$v(t)$: vocal tract impulse response

$* \quad s(t) = \sum_{k=-\infty}^{\infty} v(t - kP)$

$$S(f) = V(f)E(f)$$

$$= \frac{1}{p} \sum_{k=-\infty}^{\infty} V\left(\frac{k}{p}\right) \delta\left(f - \frac{k}{p}\right)$$

$$= \frac{1}{p} \text{comb}_{\frac{1}{p}}\{V(f)\}$$