Today- Oct. 26th

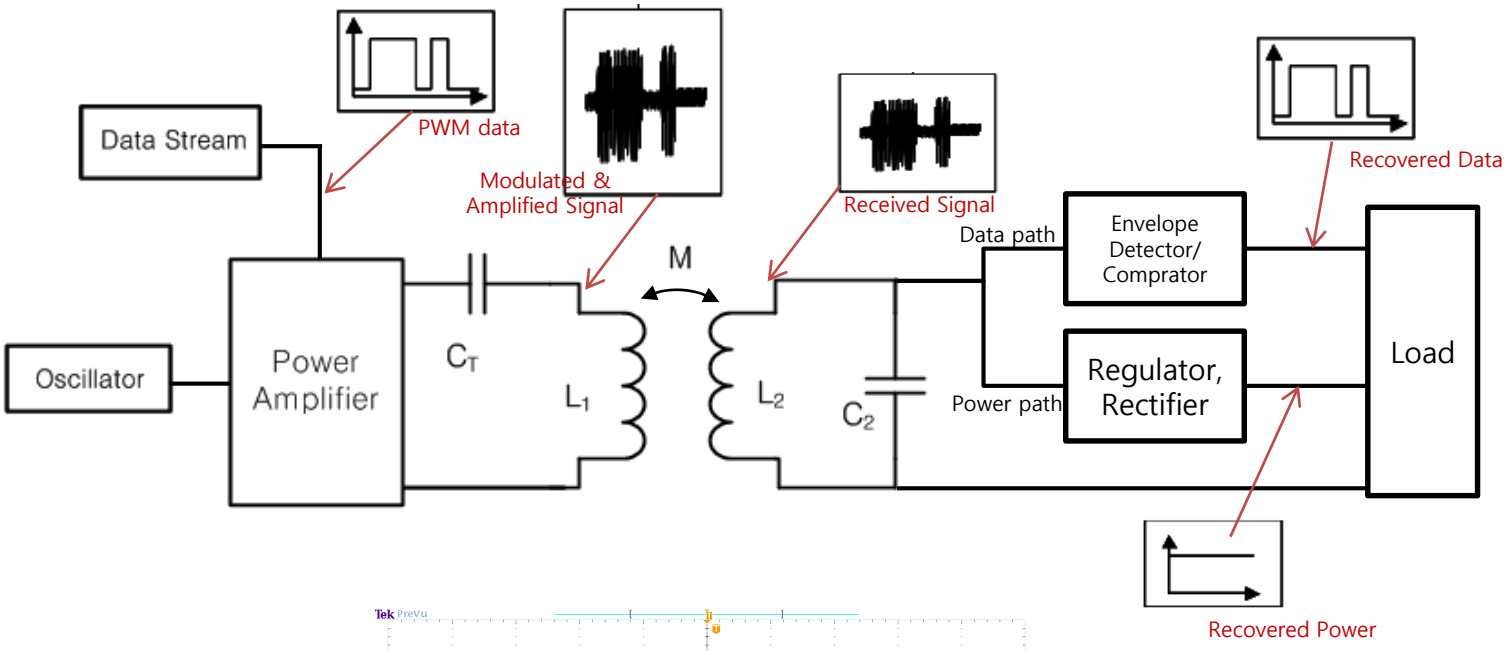
- Question on projects?
- Review
  - Telemetry
  - Formant based Speech Processing
- Speech processing Strategies –continued
  - CA
  - Lessons learned
  - CIS
  - CIS based
  - Fine Structure

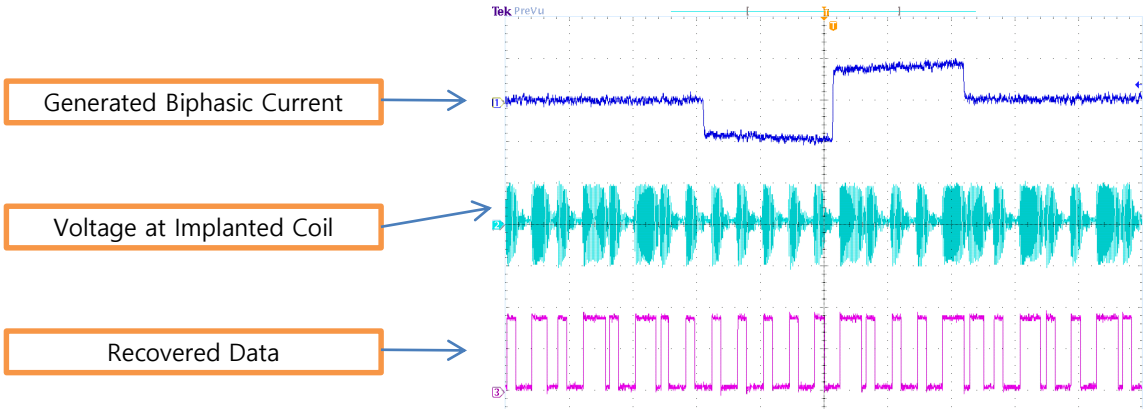
Review

telemetry

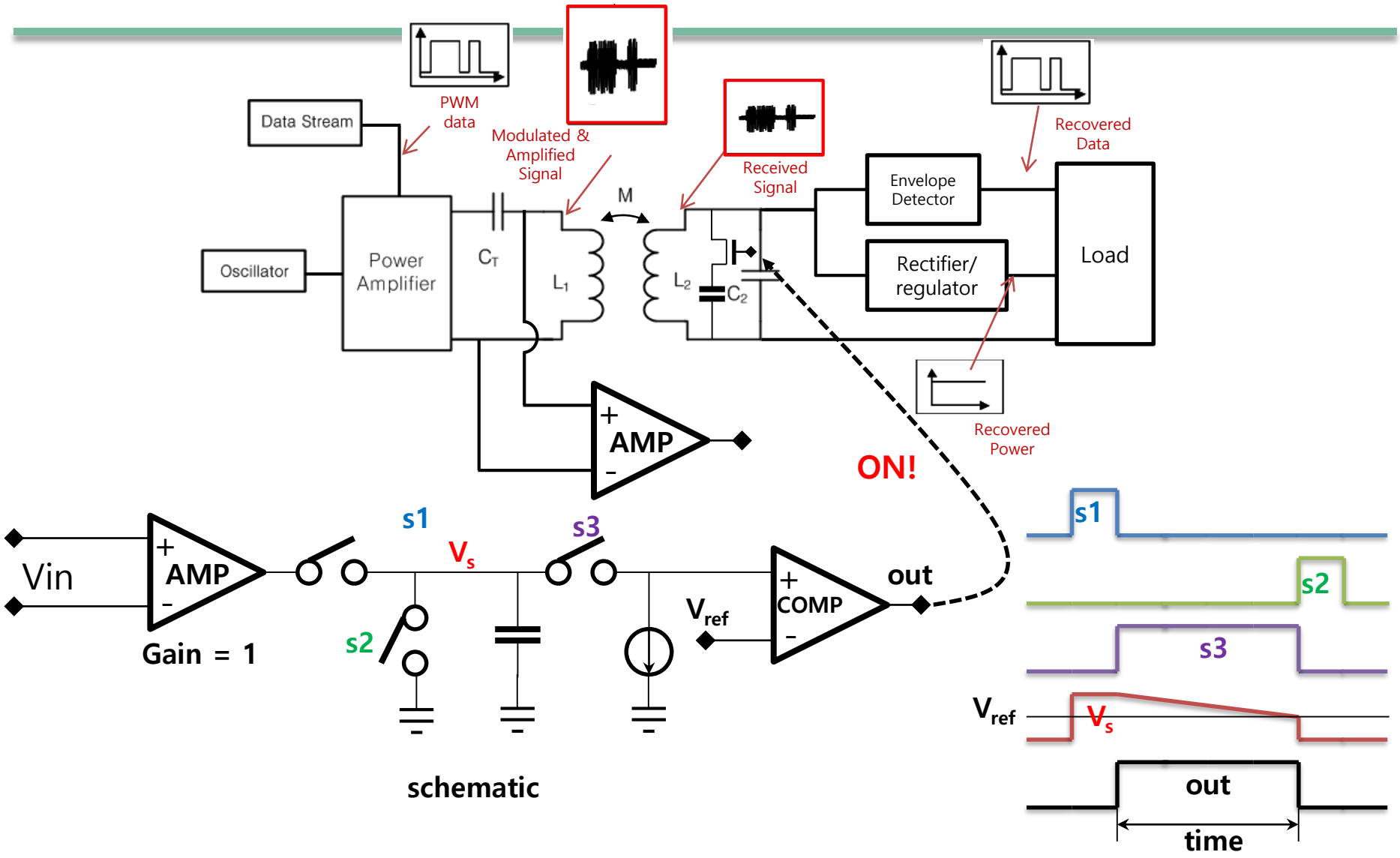
Data Telemetry – Inductive Link

- Downlink using PWM scheme





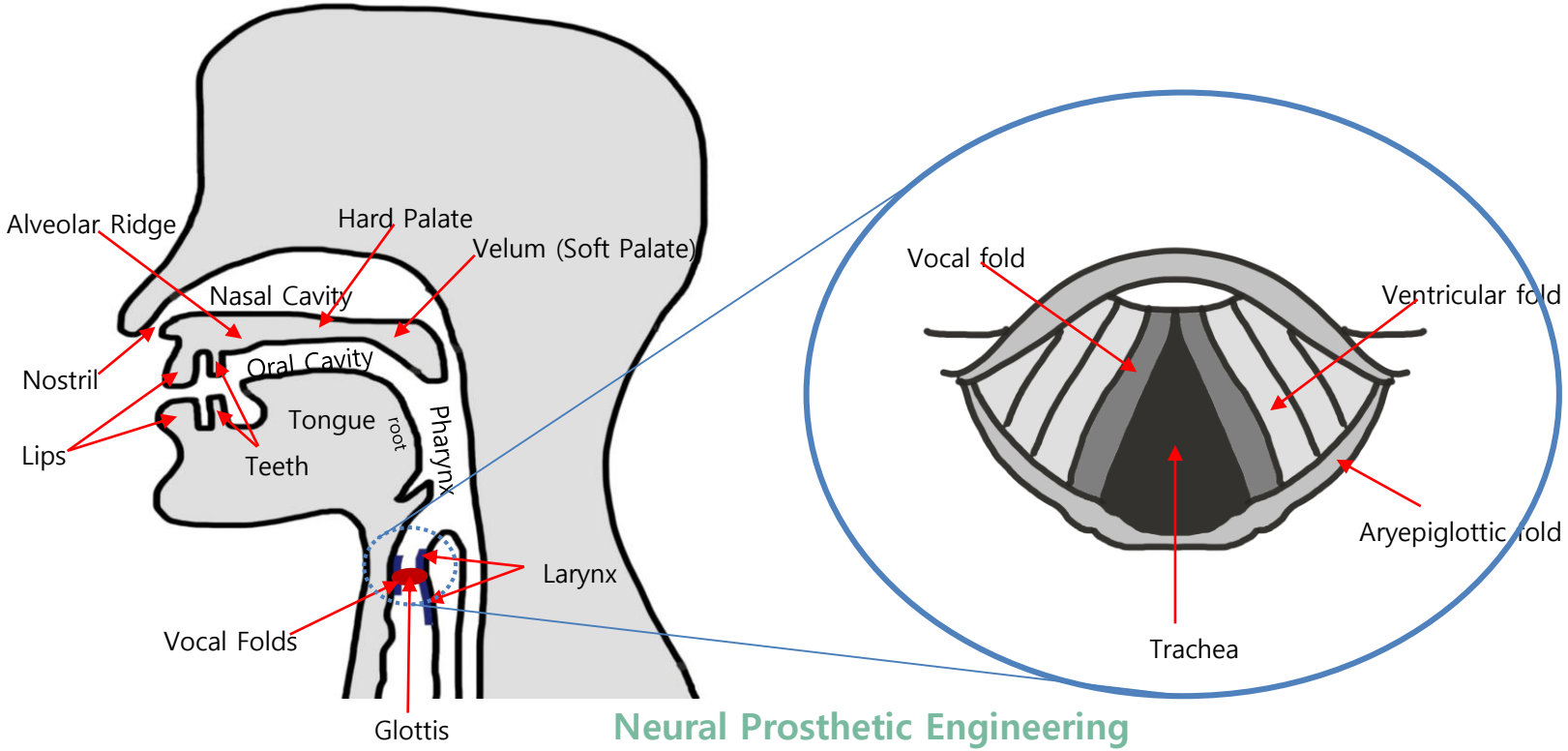
Backward telemetry



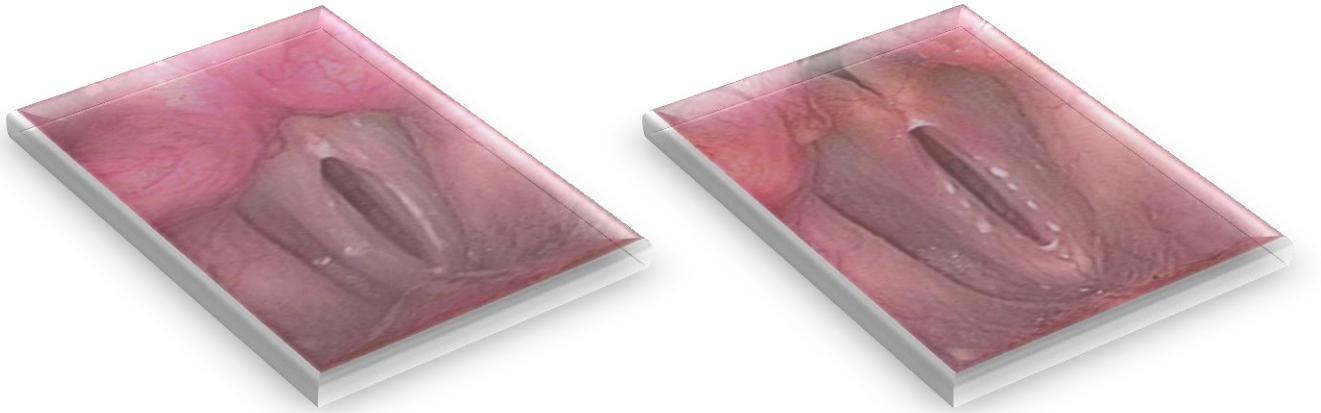
Strategies for Representing Speech Information with Cochlear Implants

Making of voice: Vocal Tract

- Sound source is in the Larynx (Vocal Fold)
- The vocal tract is the cavity where sound is filtered.
- The vocal tract consists of the laryngeal cavity, the pharynx, the oral cavity, and the nasal cavity.
- The average length of the vocal tract in adult humans is 17 cm (male) and 14 cm (female).



Vocal fold at low and high pitches



120 Hz and 200 Hz

http://www.vowelsandconsonants3e.com/chapter_2.html
https://www.youtube.com/watch?v=v9Wdf-RwLcs

Sound: Voiced or unvoiced

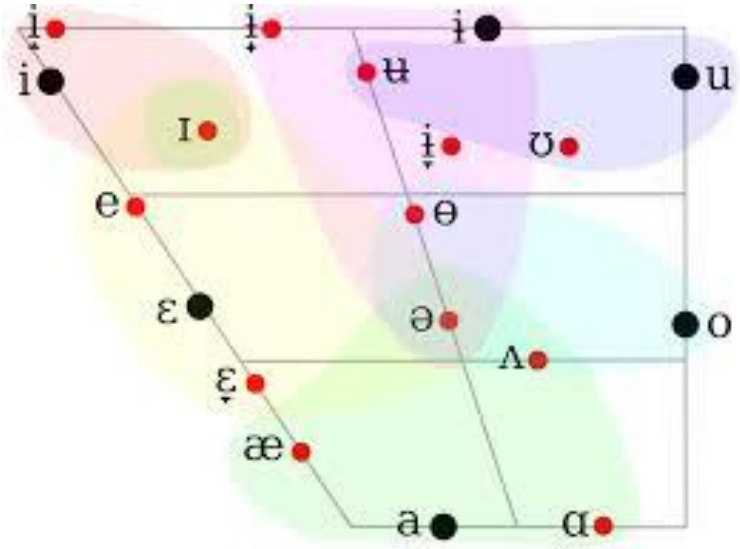
- Voicing means air is forced into the vocal tract.
- All the vowels are voiced sounds.
- Consonants are voiced or unvoiced sounds.
- Voiced sounds are resonant (vibrant).
- Unvoiced sounds are noisy.

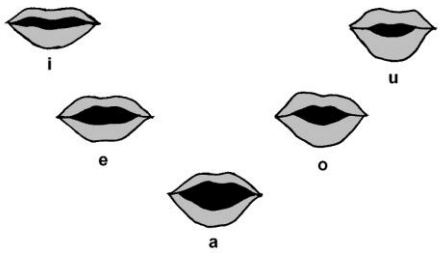
Articulation in Vocal Tract

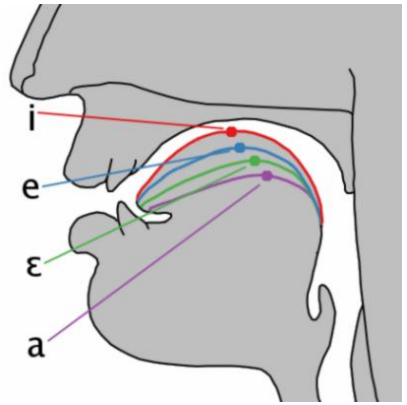
- Place of articulation
  - Where the vocal tract is shut off or narrowed
- Manner of articulation
  - How the vocal tract is articulated
- Voicing
  - Whether air is forced through the larynx

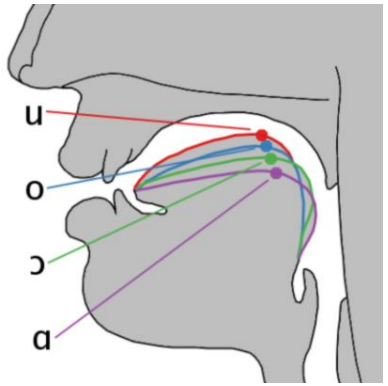
Articulation for Vowels

- Place of the articulation: High(u), Mid(o), Low(a)
- Shape of the lips: Rounded (o) or not (i)









Wikipedia, Wikimedia 2016

Articulation for Consonants

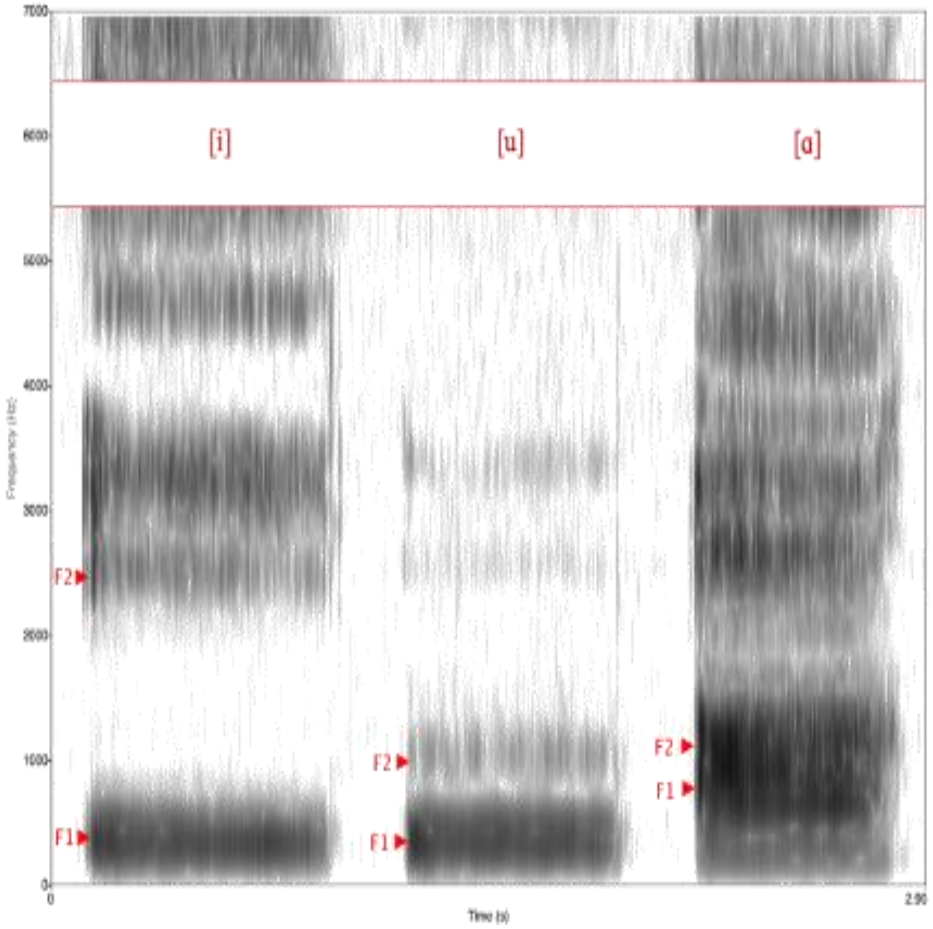
- Stop (plosive): A stop is a consonant in which airflow is completely blocked for a short time
  - [p], [t], [k] / [b], [d], [g]
- Nasals: made by lowering the velum and allowing air to pass into the nasal cavity
  - [m], [n], [ŋ]
- Fricative: airflow is constricted but not cut off completely.
  - [s]/[z]
- Affricative: Stops that are followed immediately by fricatives
  - [ts]/[dj]
- Liquid –consonants in which the tongue produces a partial closure in the mouth, resulting in a resonant, vowel-like consonant,
  - [l], [r]
- Glide –consonants with no stop or friction which consist of a glide (a quick, smooth movement) towards a following vowel.
  - [w], [y]

Formants in spectrogram

- Distinctive frequency components of the sound
- Peaks in the amplitude/frequency spectrum (spectrogram)

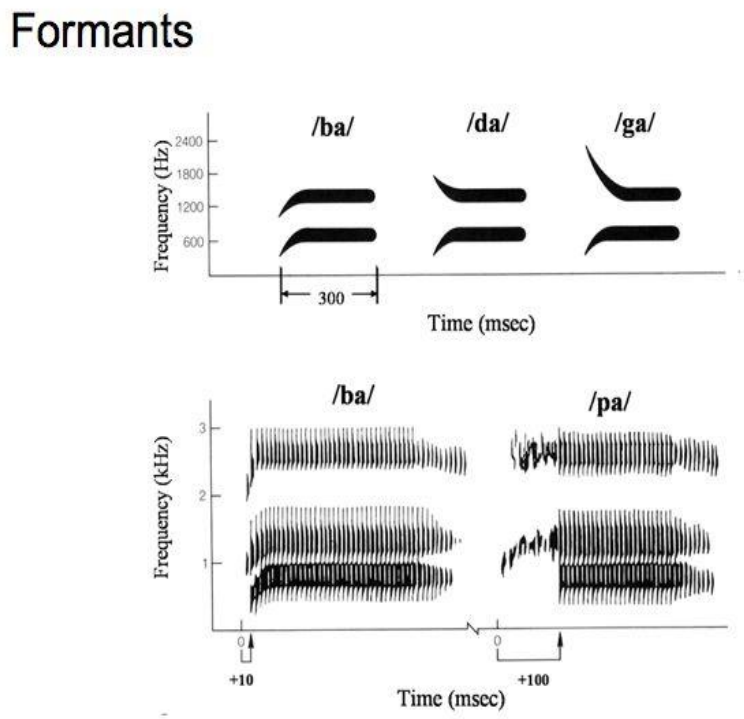
- The formant with the lowest frequency is called F1, the second F2, and the third F3.

- Most often the two first formants, F1 and F2, are enough to disambiguate the vowel.
- An interactive demonstration of this can be found here.
- http://auditoryneuroscience.com/topics/two-formant-artificial-vowels



Formants of consonants

- Nasal and Liquid consonants have added formant (F3) at higher frequencies
- Plosives and Fricatives modify the placement of formants of the vowels
  - Bilabial sounds (b, p) cause lowering of the formants
  - Velar sounds (k and g) show F2 and F3 coming together
  - Albeolar sounds (t and d) cause less systematic changes in neighboring vowel formants



Formants

- The component sounds that build up the phrase "A bird in the hand is worth two in the bush".

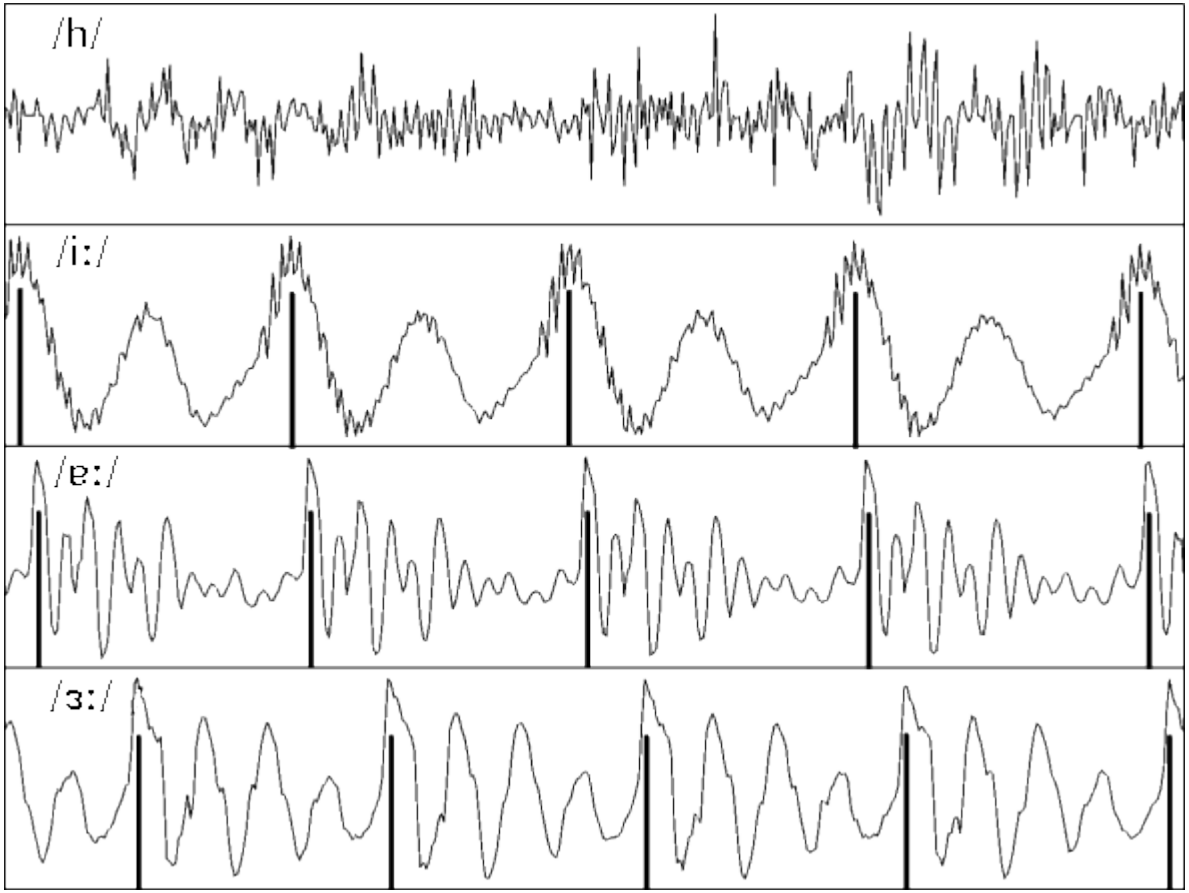
http://www.vowelsandconsonants3e.com/chapter_7.html#

Frequencies of sounds

- C1 32.7 Hz (lowest C on a standard 88 key piano)
- C4 261.64 Hz (middle C on 88 key piano)
- C6 1046.50 Hz (Highest note reproducible by the average female human voice)
- C8 4186 Hz (highest note on 88 key piano)

https://www.youtube.com/watch?v=qNf9nzvnd1k

Sound Waveforms: Voiced or unvoiced



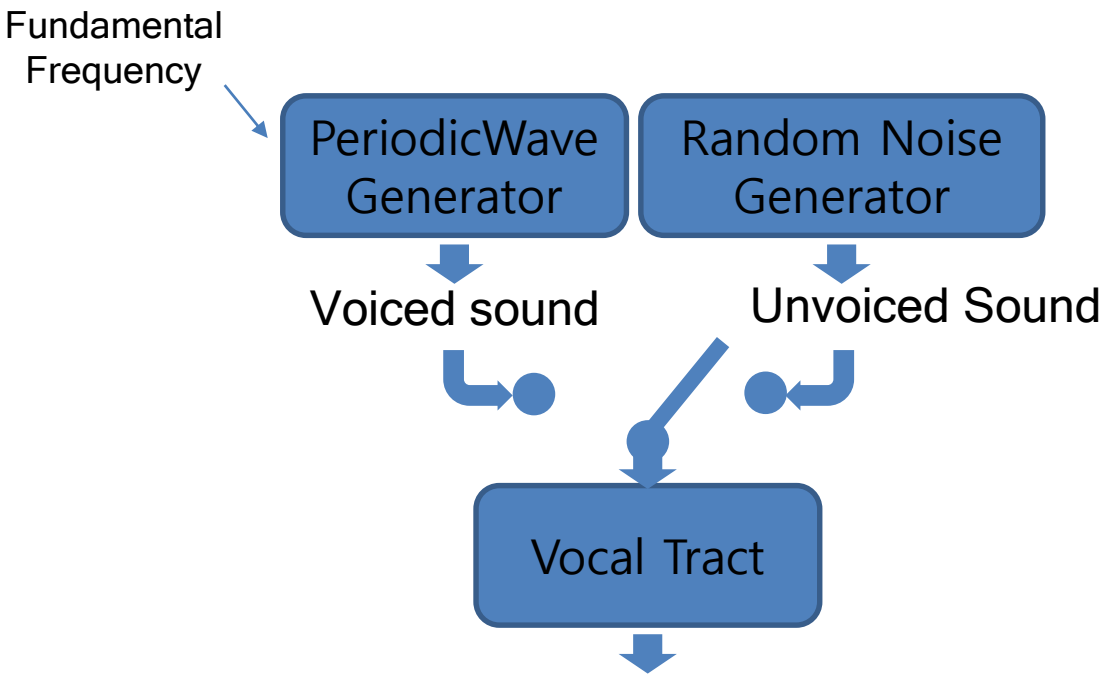
40 msec view

http://clas.mq.edu.au/speech/acoustics/waveforms/speech_waveforms.html

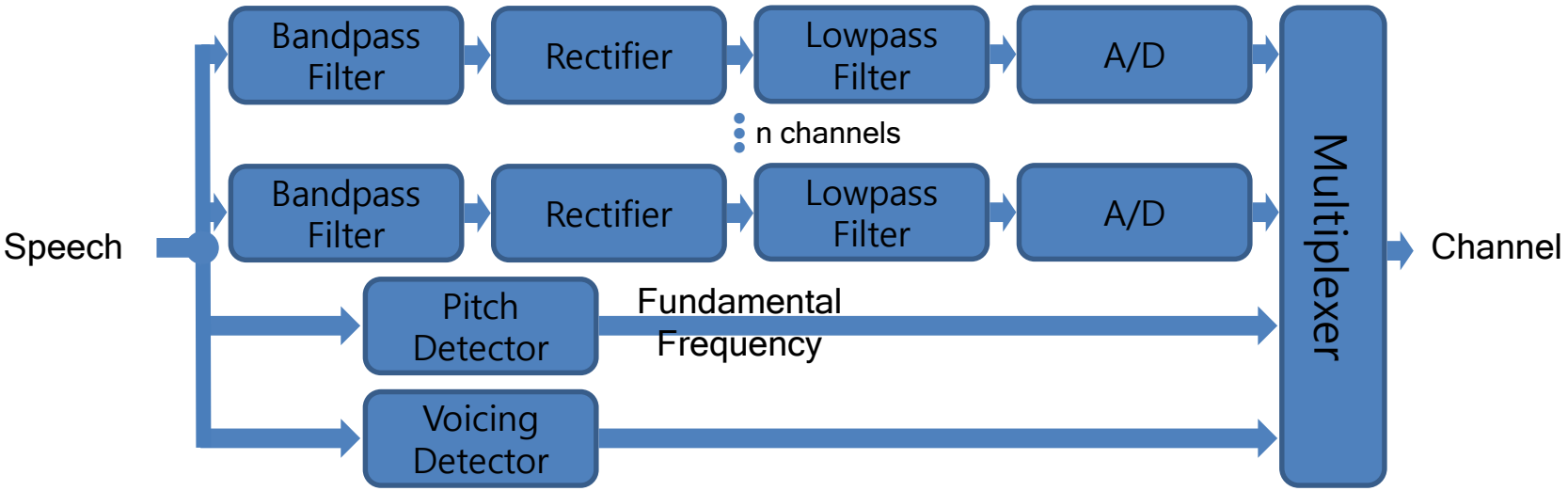
Vocoder

- Vocoder (voice coder)
  - invented by Dudley in the 1930s
  - a means of reproducing an intelligible facsimile of a voice for recorded messages on telephone systems
  - Analysis (encoding) stage / decoding (synthesis) stage
  - A limited set of parameters from speech input in the analysis part → transmitted to the receiver
  - The information rate required for transmission of the parameters is much less than that required for transmission of the unprocessed speech signal

Model for Voice Coding

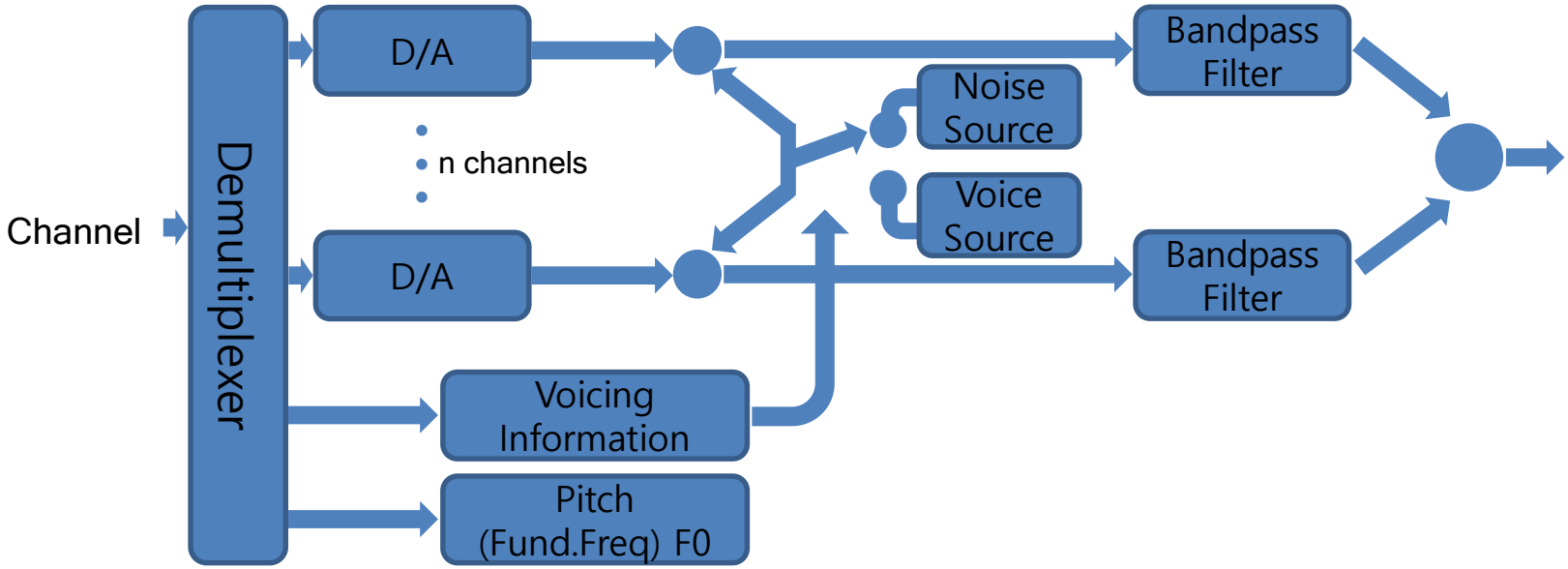


Channel Vocoder : analysis part



- Voicing detector determines whether the sound is voiced or not
- Pitch detector determines the frequency of the glottal openings for the voice sound
- Configuration of the vocal tract is found with a band of bandpass filters and envelopment detector (low pass filters).
- This analysis provides information of the vocal tract at 5-30 msec interval.

Channel Vocoder: synthesis part



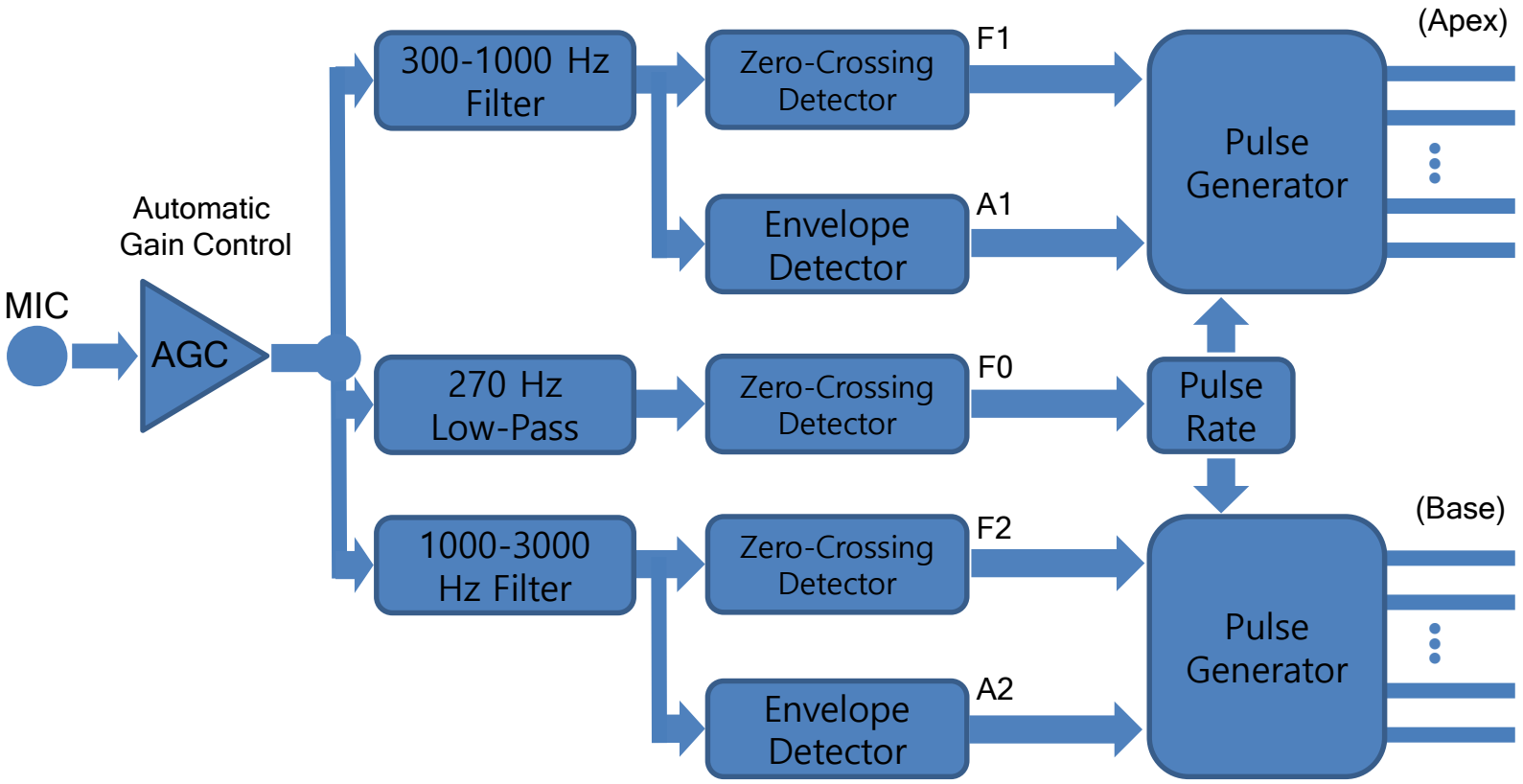
- A synthesized speech signal is formed by summing the outputs of the band pass filters.
- Voicing information is a binary indication.
- Each output is a smoothed envelop energy.

Speech Processing Strategies

Formant based speech Processing Strategies

- Vocoder theory and models played major roles in the early designs.
- Fundamental Frequency (F0) and two formants (F1 and F2) are used
- F0 is the fundamental frequency and determines the stimulation rate
- F1 gives information about vowels
- F2 gives information about consonants

Speech Processing Strategies – F0/F1/F2

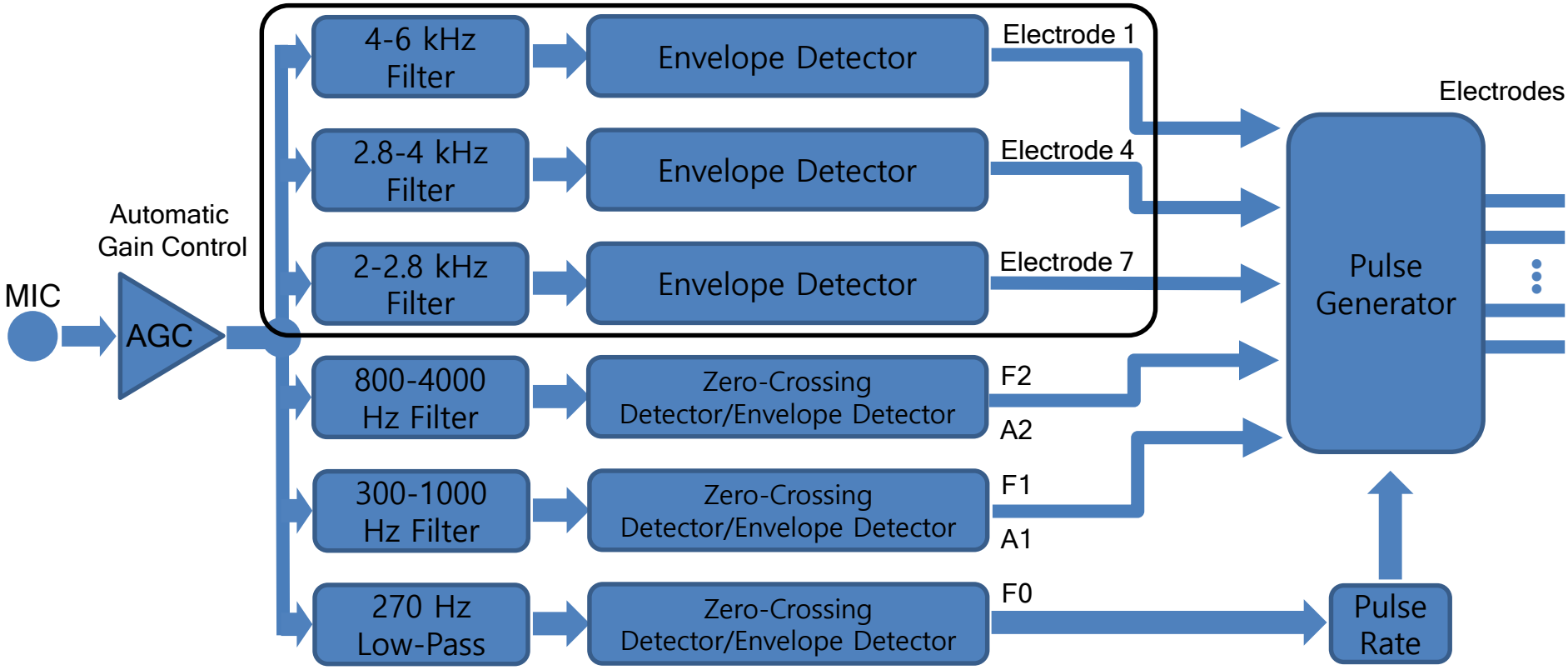


P.C.Loizou, (IEEE Engineering in Medicine and biology, 1999)

MPEAK Speech Processing Strategy

- In addition to formant information, MPEAK extracts channels of higher frequency information from speech
- MPEAK as well as F1/F2 strategies, tend to make errors in formant extraction in noisy environment

Speech Processing Strategies – MPEAK

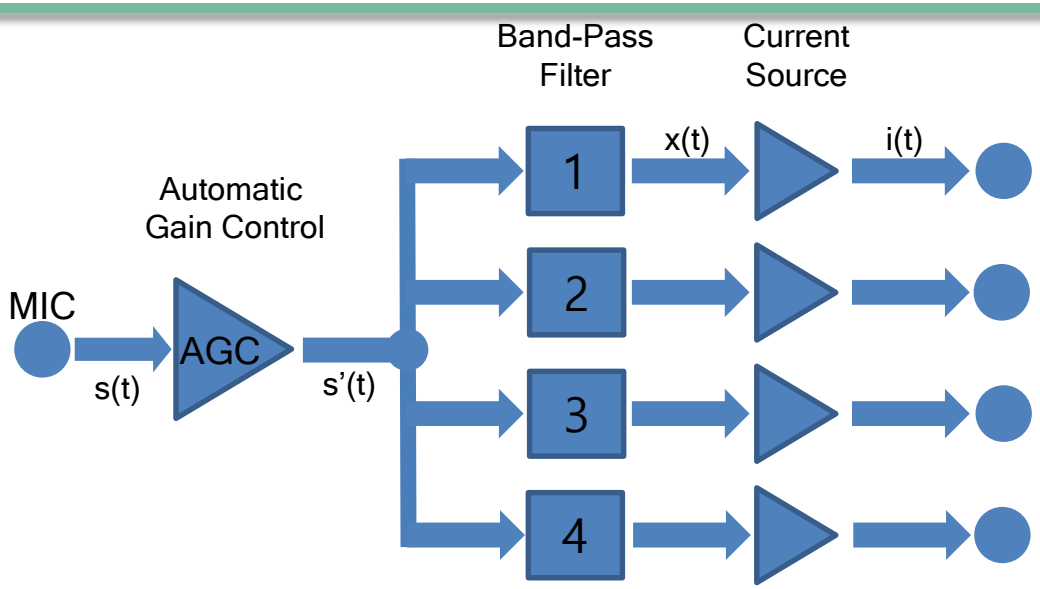


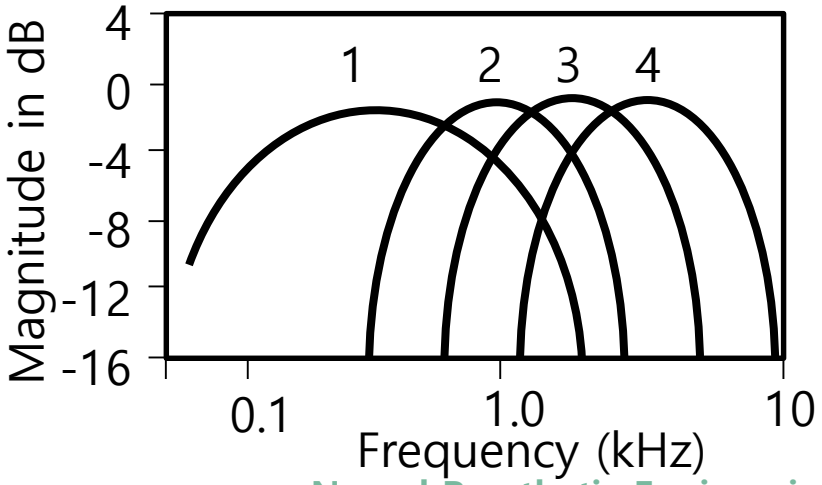
P.C.Loizou, (IEEE Engineering in Medicine and biology, 1999)

Recent Speech Processing Strategies

- Compressed Analog (CA)
- Continuous Interleaved Sampling (CIS)
- ACE and SPEAK (Cochlear)
- Harmony HiRes Virtual Channels (Clarion)

Speech Processing Strategies - CA





Neural Prosthetic Engineering

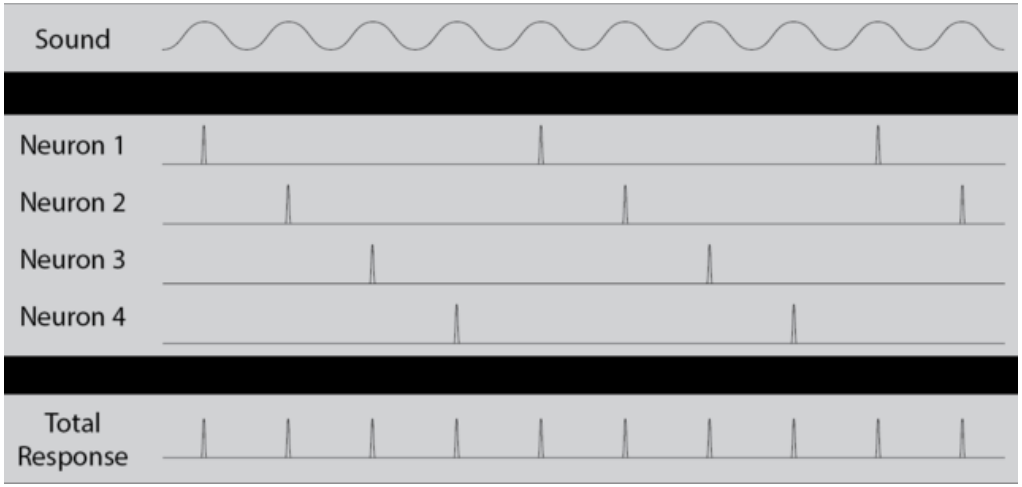
B. Wilson et al., (Nature, 1991)

Lessons learned

- Lessons learned from the formant-based strategies and the CA strategy.
- The amount of information perceived by CI users is much less.
- Perception of electrical stimuli is different from acoustic stimuli.
- Pitch saturation limit= typically around 300 pulses/s for electrical pulses or 300 Hz for electrical sinusoids. Higher rates or frequencies do not produce increases in pitch.
- In normal hearing, different pitches are heard over much wider ranges of rates or frequencies (up to ~5KHz), probably through combinations of rate and place cues ('Volley' theory and Place theory) .

Theories

- Place Code Theory
- Time (Rate) Code Theory
- Volley Theory

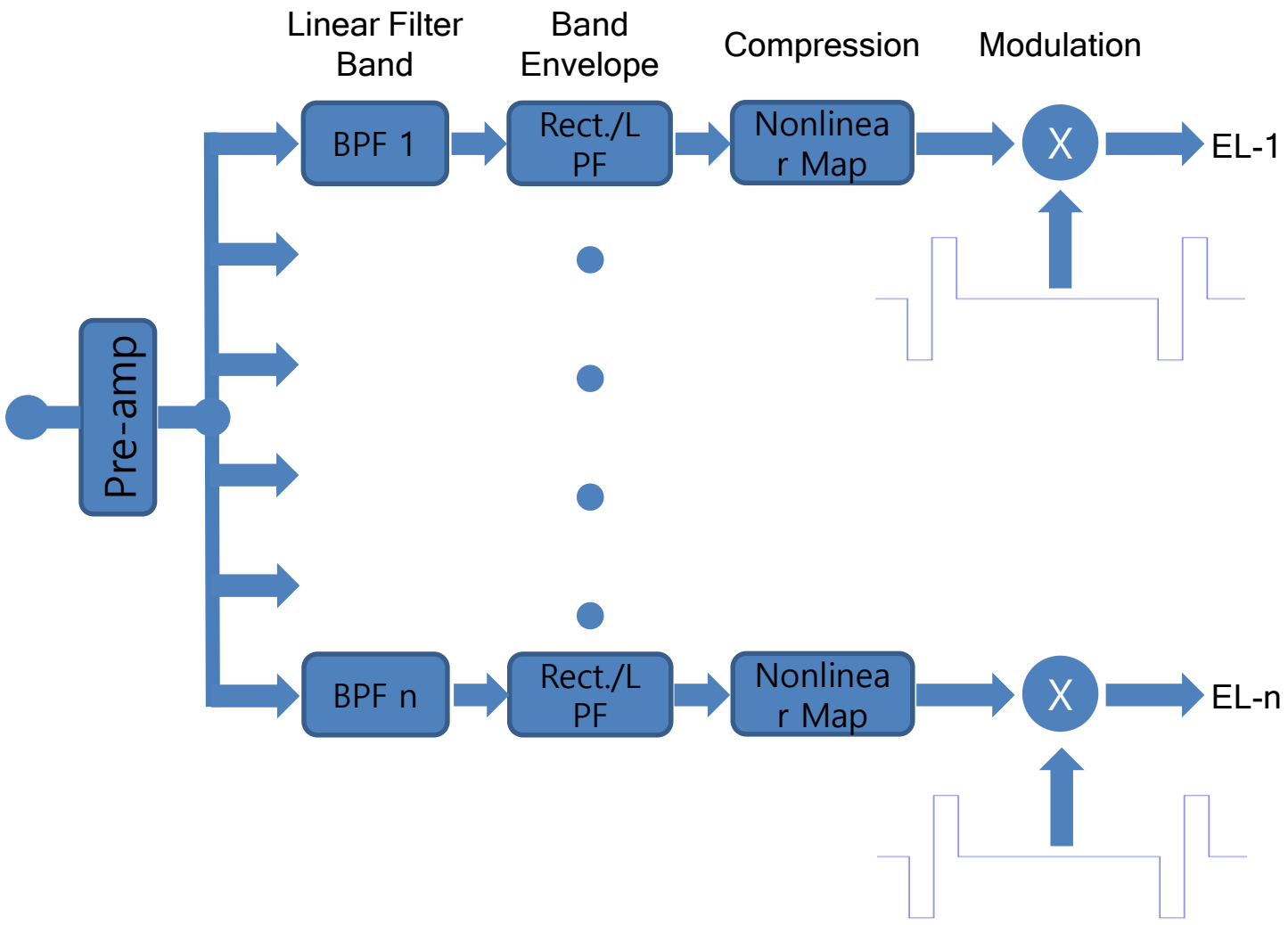


Wilipedia
File:Volley Principle of Hearing.png

CIS (Continuous Interleaved Sampling)

- Pulsatile processing
- Biphasic pulse trains are delivered the electrodes in a non-simultaneous (interleaved) pattern.
- No Patent
- Commercial devices use modified version of CIS

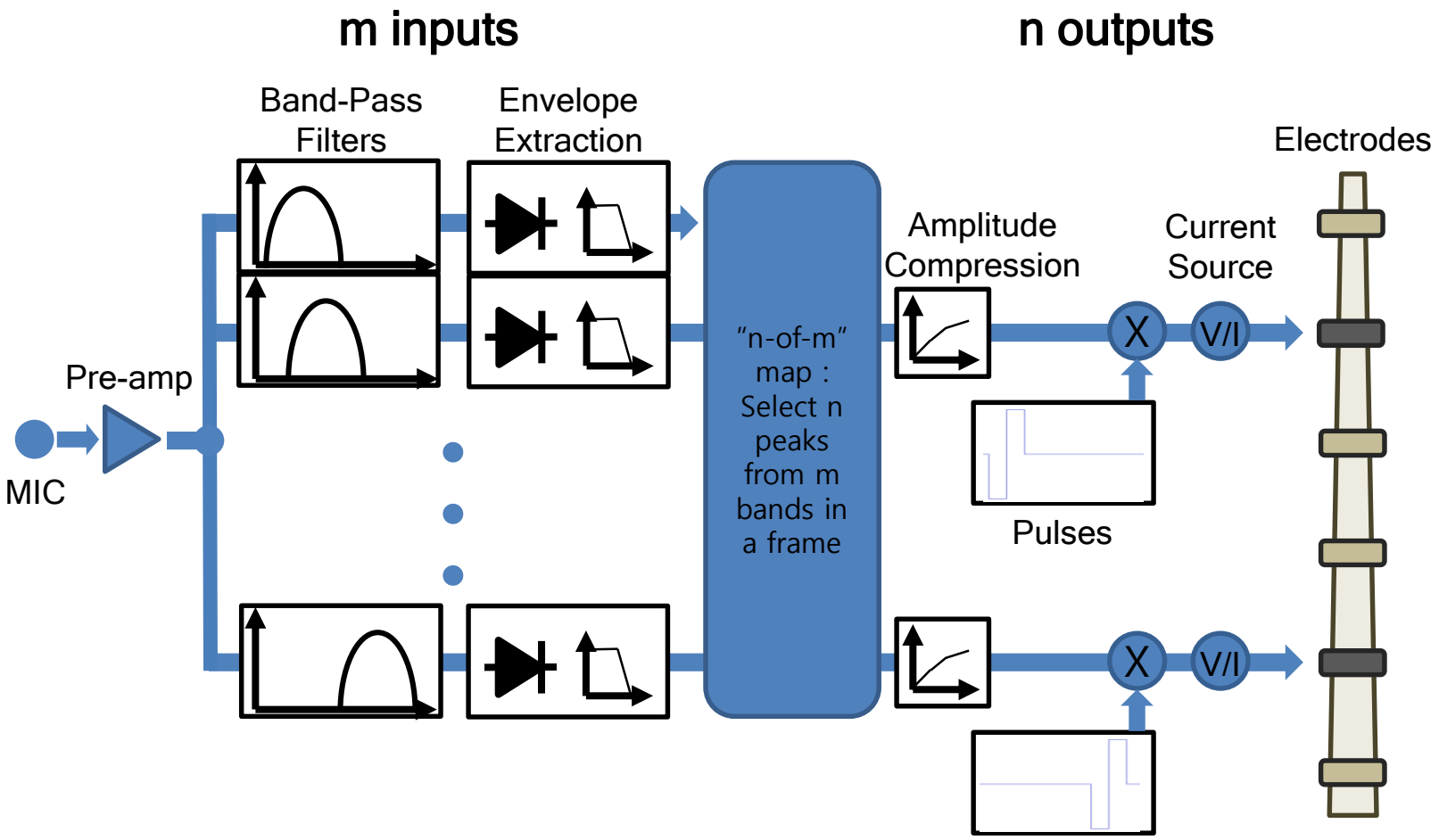
Speech Processing Strategies - CIS



Neural Prosthetic Engineering

B. Wilson et al., (Nature, 1991)

Speech Processing Strategies – n of m ,SPEAK, ACE



F.G.Zeng et al., (IEEE Reviews in Biomedical Engineering, 2008)

Speech Processing Strategies – n of m ,SPEAK, ACE

- **The pre-processing is similar to the CIS strategy**
- **N-of-m strategy has greater number of bandpass filters**
- **The SPEAK strategy selects 6–8 largest peaks and has a fixed 250 Hz per channel rate**
- **The ACE strategy has a larger range of peak selection (8-12) and higher rate (900-1200 Hz) than the SPEAK strategy**

F.G.Zeng et al., (IEEE Reviews in Biomedical Engineering, 2008)

Summary

- Speech Processing Strategies advance with time
- Formant based
- CA
- CIS
- Need to implement finer features (more detailed sounds)
  - Tonal languages
  - Music

Discussion: Fine Structure Representation

- Typical Frequency range of CI frequency filters: 300-8000Hz
- Normal audible frequency range: 20- 20,000Hz

- Low frequency cues (20-50Hz) give prosody information (stress, syllabification)-"Envelope Cues"
- Mid frequency cues (50-500Hz) give segmental information such as consonant manner, voicing, and intonation-"Periodicity Cues"
- High frequency cues (600-10,000Hz) gives consonant place and vowel quality- " Fine Structure Cues"

- Advanced Bionics HiRes is an example of Speech Processing Strategy intended to provide better Fine Structure Cues
- HiRes sample temporal fluctuations up to 2800 Hz across 16 channels
- 16 independent current sources enable simultaneous analog stimulation (SAS) as well as CIS
- "current Steering" provides virtual channel capability (HiRes 120= 15 channels times 8 spectral bands per channel)

- [1] HiResolutin Sound Processing, by Jill.B.Firszt, www.advancedbionics.com
- [2} HiRes Fidelity 120 Sound Processing, Advanced Bionics Technical Report, www.advancedbionics.com
- [3] Rosen, Temporal information in speech and its relevance for cochlear implants, Cochlear Implnat: Acquisition and controversies, ed. B Fraysse, N. Couchard, pp3-26 (1989)

Related Videos

- Hearing CI
- https://www.youtube.com/watch?v=00WOao4kpwM
- CI simulations
- https://www.youtube.com/watch?v=iwbwhfCWs2Q
- A day of a CI user
- https://www.youtube.com/watch?v=pk_7MVqpnIk