SIO 209: Signal Processing for Ocean Sciences
Class 16

Florian Meyer

Scripps Institution of Oceanography
Electrical and Computer Engineering Department
University of California San Diego

SCRIPPS INSTITUTION OF OCEANOGRAPHY

UC San Diego
JACOBS SCHOOL OF ENGINEERING

0

Speech Processing

- **Concatenated tube model of the speech production process**

- Reflection coefficients characterize the acoustic impedance contrast between adjacent tubes with different diameters

J. L. Flanagan, "Speech Analysis, Synthesis and Perception." Springer-Verlag, 1972

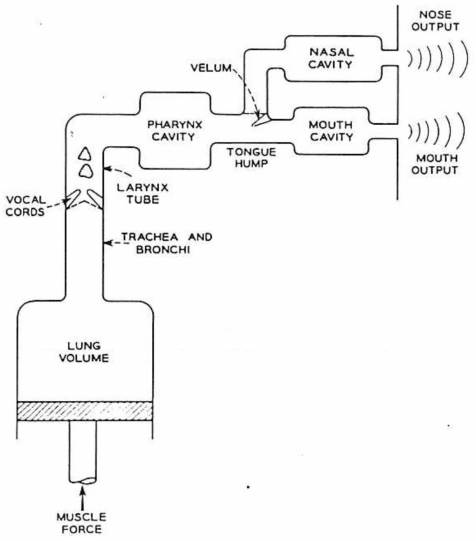
L. Rabiner and B. Gold, "Theory and Application of Digital Signal Processing." Prentice Hall, 1975

L. Rabiner and R. Schafer, "Digital Processing of Speech Signals." Prentice Hall, 1978

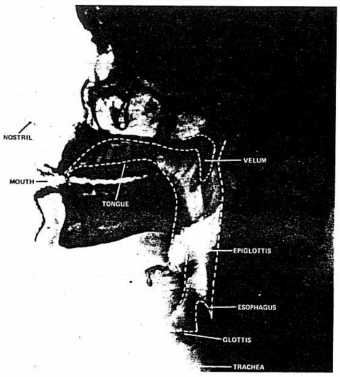
J. Markel, "Digital Inverse Filtering – A New Tool for Formant Trajectory Estimation." IEEE Trans. Audio and Electroacoustics, 1972

1

The Vocal Tract



Schematic diagram of the human speech production mechanism (After Flanagan).



X-ray of a male vocal tract (After Flanagan).

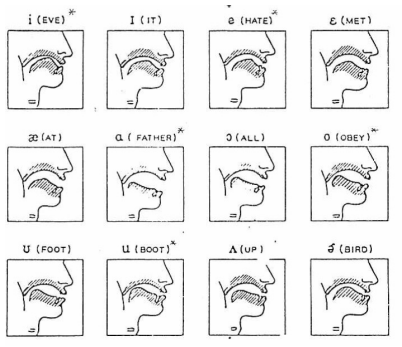
The human vocal tract is a nonuniform acoustical tube that extends from the glottis to the lips. It is about 17 cm long in an adult male and therefore its first quarter-wave resonance occurs at a frequency given by

$$F_1 = \frac{1}{4}\frac{c}{l} = \frac{1}{4}\frac{34{,}000\ \text{cm/sec}}{17\ \text{cm}} = 500\ \text{Hz} \qquad (12.1)$$

Its nonuniform cross-sectional area depends strongly on the position of the articulators and varies from 0 cm² at closure to about 20 cm². The vocal tract has certain normal resonant modes of vibration, called *formants*, that depend heavily on the exact position of the articulators.

2

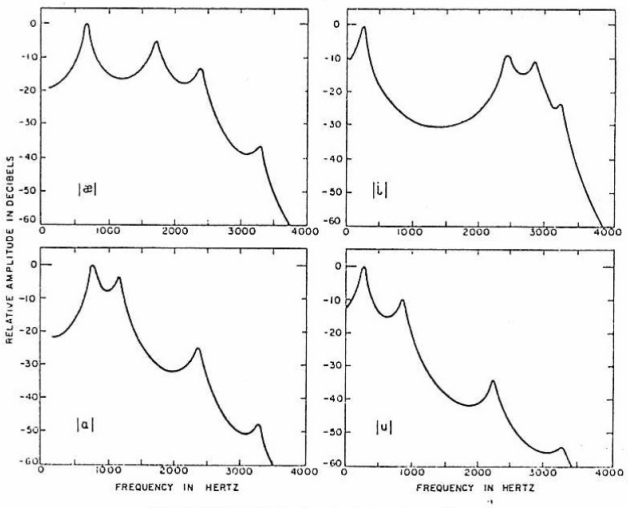
Vowels



Formant Frequencies for the Vowels

Typewritten Symbol for Vowel	IPA Symbol	Typical Word	F_1	F_2	F_3
IY	i	(beet)	270	2290	3010
I	ɪ	(bit)	390	1990	2550
E	ɛ	(bet)	530	1840	2480
AE	æ	(bat)	660	1720	2410
UH	ʌ	(but)	520	1190	2390
A	ɑ	(hot)	730	1090	2440
OW	ɔ	(bought)	570	840	2410
U	ʊ	(food)	440	1020	2240
OO	u	(boot)	300	870	2240
ER	ɚ	(bird)	490	1350	1690

Schematic vocal tract profiles and table of formant frequencies for several vowels



Frequency transmission characteristics of several vowels

3

Vowels

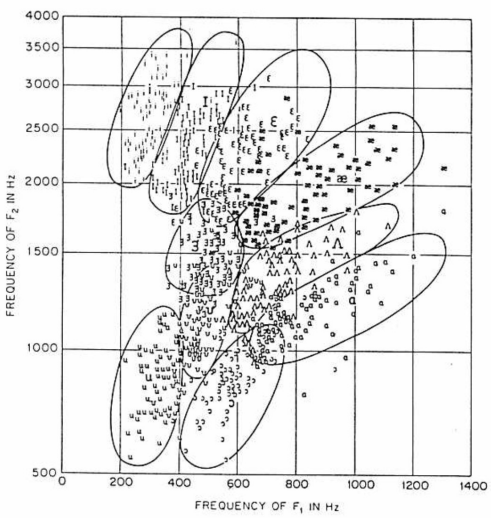


Fig. 3.4 Plot of second formant frequency versus first formant frequency for vowels by a wide range of speakers. (After Peterson and Barney [11].)

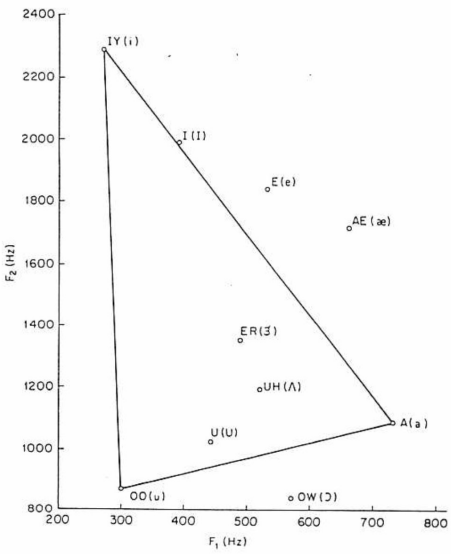
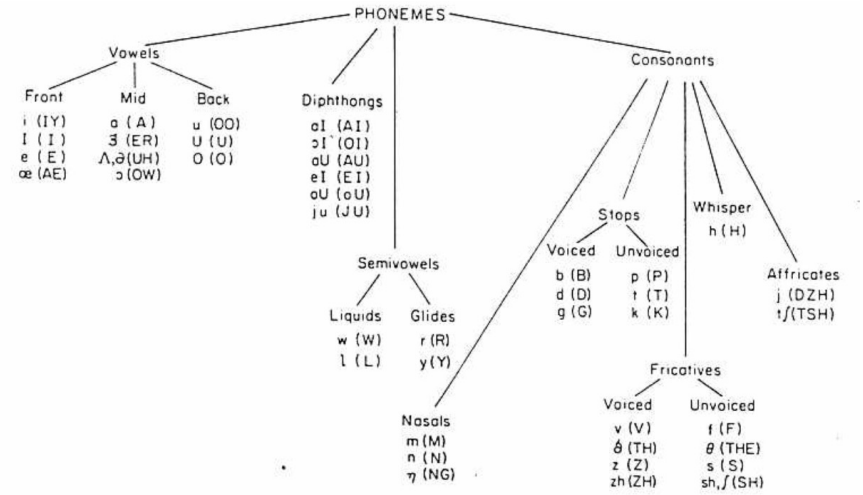


Fig. 3.5 The vowel triangle.

4

Phonems

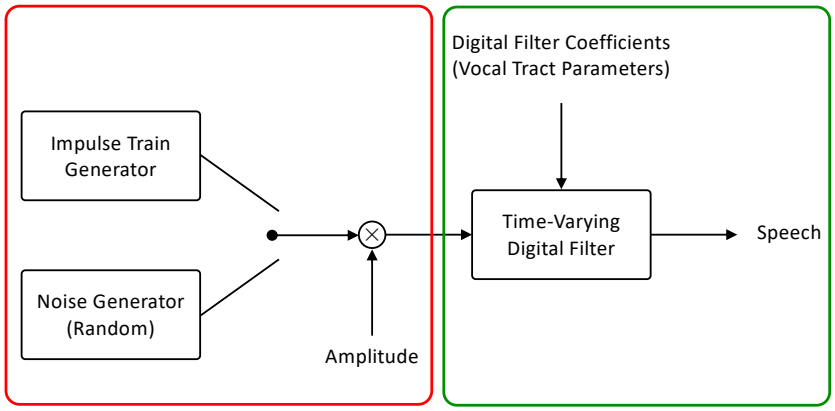
Phonems in American English



5

Example: Homomorphic Speech Processing

- Model of speech production



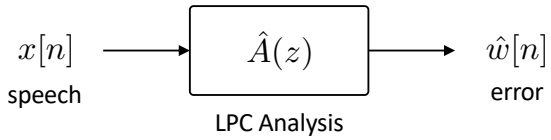
A vocoder extracts parameters of excitation signal and the vocal tract filter; transmitting these parameters (instead of the sampled speech signal) can reduce data rate significantly

6

6

Speech Transmission

- Direct digitization of speech results in a data rate of 64 kB/s since typically 8 Bits/sample are used and $f_s = 8\text{kHz}$ (speech is bandlimited to $4 - 5\text{kHz}$)



- We take segments ("frames") that are 25 ms long, i.e., there are 40 segments/sec
- For each segment we need to obtain
  - the $\hat{A}$ vector (10–14 coefficients)
  - voiced/unvoiced decision
  - pitch period
  - amplitude

⟶ Results in a data rate of $1.0 - 2.4\text{kb/s}$

7