

Speech Production and Acoustic Properties — Summary Notes

1. Physiological Speech Production

Speech starts from the brain, triggering motor commands to control lungs, vocal folds, and articulators. Air pushed from the lungs passes through the vocal folds to create oscillations that generate sound.

2. The Vocal Folds

Located in the larynx. The glottis is the opening between folds; below is the subglottal region. Frequency depends on fold tension, air pressure, and length/mass of the folds.

3. Linguistic Structure of Speech

Speech is structured as Phones → Syllables → Words → Utterances. Phonetics deals with sound production; Phonology deals with abstract sound patterns.

4. Acoustic Representation

Speech waveforms are digitized as PCM signals. Defined by sampling rate (F_s) and quantization accuracy.

5. Sampling Rate

Nyquist frequency = $F_s/2$. Speech mainly lies in 300–3500 Hz. Common bands: Narrow (8kHz), Wide (16kHz), Super-wide (32kHz), Full (44.1kHz).

6. Quantization

Linear quantization uses equal steps; Logarithmic and μ -law improve low-amplitude clarity.

7. Short-Time Analysis

Speech is non-stationary, analyzed using 20–30 ms windows with Hann/Hamming functions.

8. Spectral Analysis

DFT/FFT converts signals to frequency domain. Spectrograms show time-varying spectral features like F0, formants, harmonics.

9. Zero-Crossing Rate (ZCR)

$ZCR = (1/(2N)) \sum |\text{sign}(x[n]) - \text{sign}(x[n-1])|$. Measures smoothness. High for fricatives (/s/, /f/), low for vowels.

Numerical Example: Zero-Crossing Rate

Given samples: [0.8, 0.4, -0.2, -0.5, 0.1, 0.7, -0.3, -0.6, 0.2, 0.9, 0.3, -0.4, -0.7, 0.5, 0.6, -0.1, -0.8, 0.3, 0.4, -0.5]

Step 1: Convert to signs \rightarrow [+1, +1, -1, -1, +1, +1, -1, -1, +1, +1, +1, -1, -1, +1, +1, -1, -1, +1, +1, -1]

Step 2: Count sign changes = 9 zero crossings

Step 3: ZCR = 9/20 = 0.45 crossings per sample

Step 4: ZCR rate = $0.45 \times 8000 = 3600$ crossings/sec

Interpretation: Low ZCR \rightarrow voiced (smooth), High ZCR \rightarrow unvoiced (noisy).

■ Summary Table

Topic	Key Idea
Speech Production	Brain \rightarrow Lungs \rightarrow Vocal folds \rightarrow Sound
Vocal Folds	Tension + Pressure control frequency
Phonetics vs Phonology	Physical vs Abstract structure
Sampling Rate	Nyquist = $F_s/2$
Quantization	Linear / Log / μ -law companding
Short-Time Analysis	20–30 ms windowed segments
Spectrogram	Shows F0, harmonics, formants
ZCR	Measures smoothness; detects voiced/unvoiced