

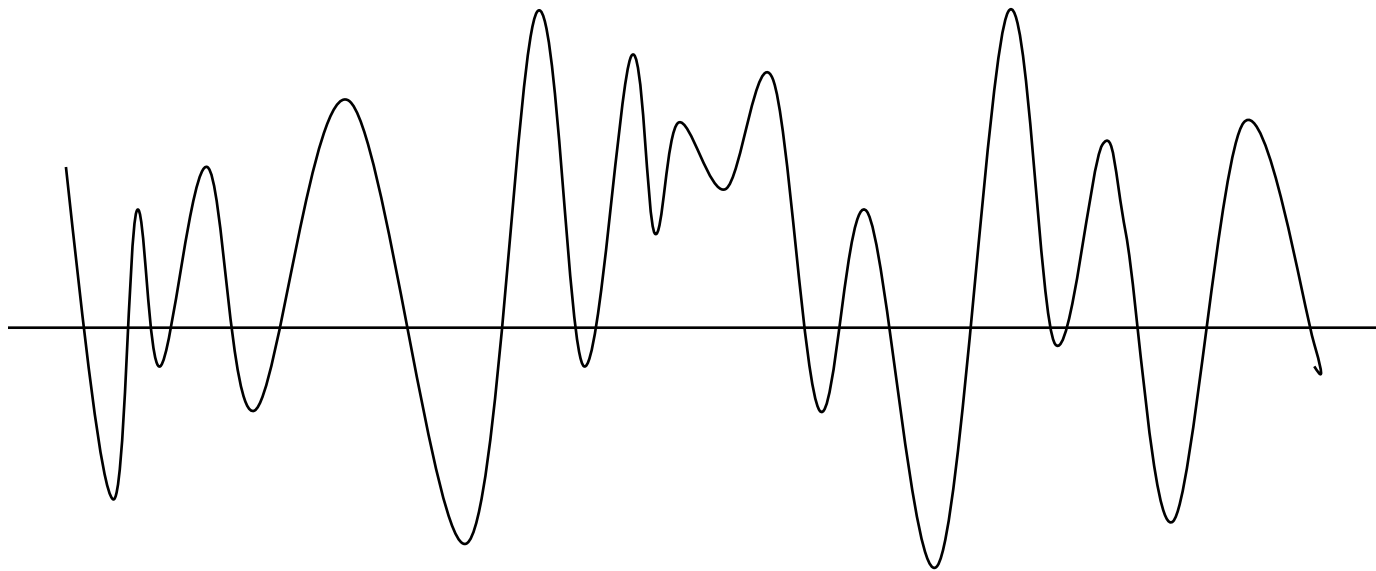
---

# **Voice Traffic Characteristics**

# Voice Communication Characteristics

---

- Speech produces a signal which varies slowly in time (4 kHz bandwidth)



# Voice Coding

---

- The voice processing comprises two steps:
  - Speech analysis
    - converts an analog voice signal into a digital form.
  - Speech synthesis
    - converts a digital voice data into its analog form.
- Two methods used for voice processing
  - waveform coding
    - Pulse Code Modulation (PCM)
    - Code-excited Linear Prediction Coding (CELP)
  - vocoding

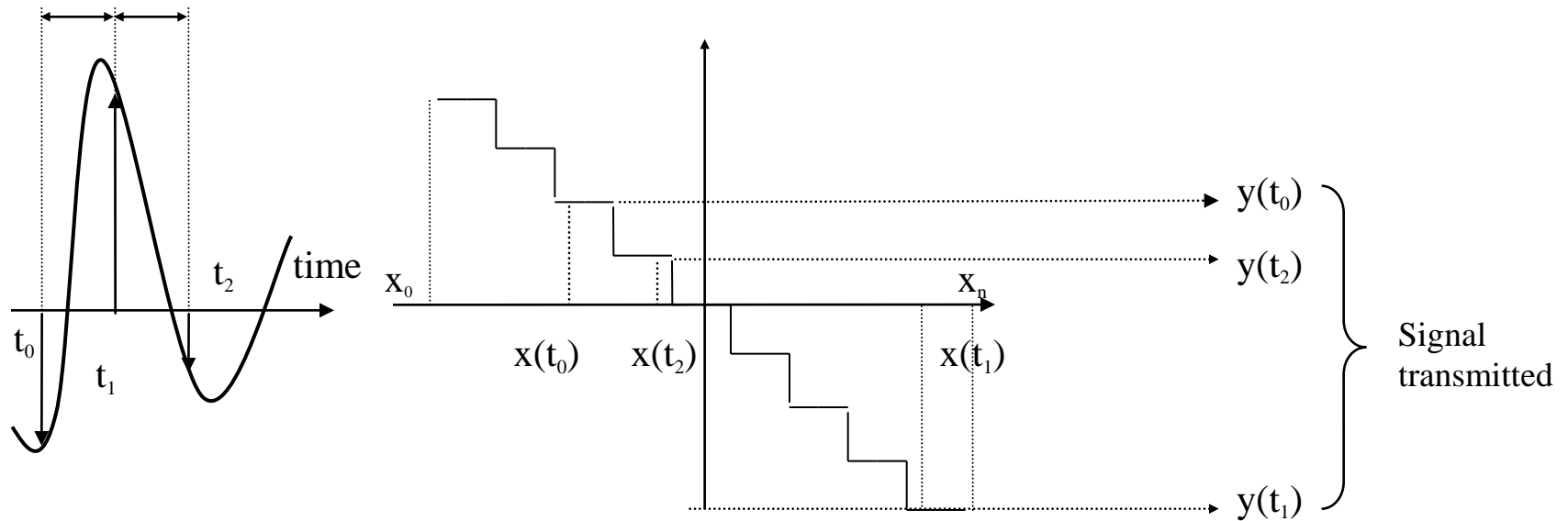
# **PCM**

---

- Signal is sampled at regular intervals
  - sampling rate = 8000 Hz (Nyquist Rate)
- Samples are quantized and transmitted
- 8 bits/sample  $\Rightarrow$  64 kb/s

# Sampling and Quantization

Sampling period



Voice Signal

Quantization function

# Voice Quality Measure: Signal-to-Noise Ratio

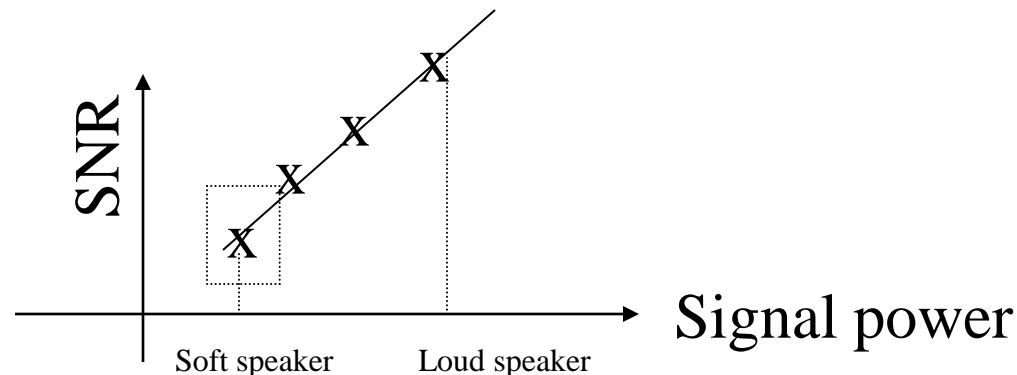
- Quantization is a source of degradation (noise):
- May be measured by:

$$\text{SNR} = \frac{\sum_{k=1}^N \int_{X_{k-1}}^{X_k} x^2 p(x) dx}{\sum_{k=1}^N \int_{X_{k-1}}^{X_k} (x - y_k)^2 p(x) dx} = \frac{\sigma_x^2}{\sigma_q^2}$$

- Where:
  - $p(x)$  the probability density function of the signal
  - $X_k$ , the decision level

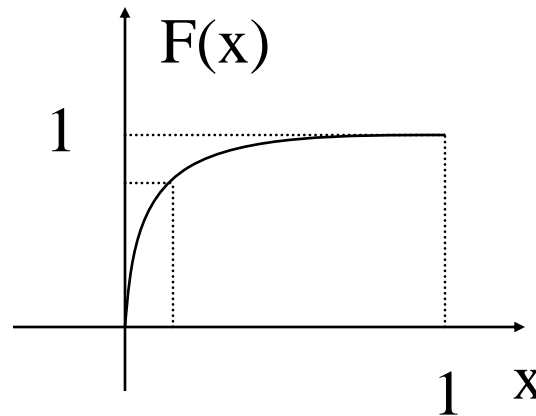
# Uniform Quantizer

- Interval between consecutive decision levels is constant:  $X_k - X_{k-1} = \Delta$  (constant).
- Problem:
  - SNR is not constant
  - depends on the amplitude
    - The soft speaker is penalized more than a loud speaker



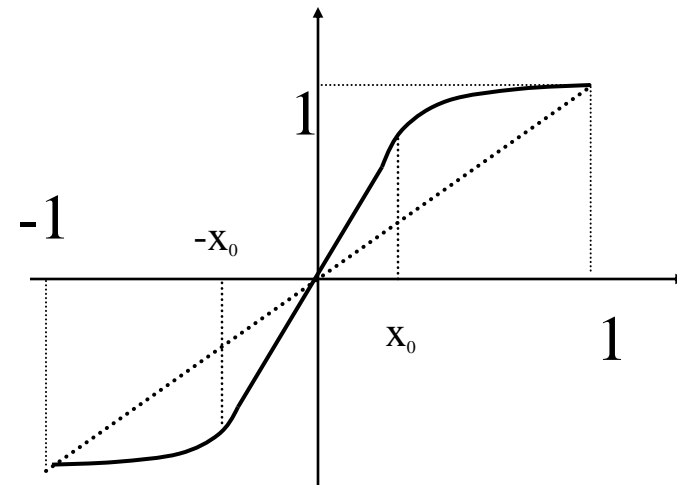
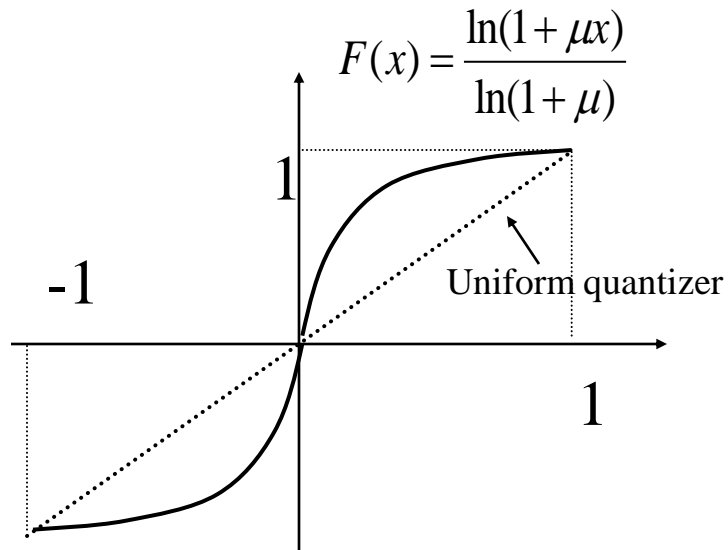
# Non uniform quantizer

- The idea is to apply to the signal a function  $F$ , before using a uniform quantizer
- Two laws are commonly used:
  - A-Law (Europe)
  - $\mu$ -Law (North America)



# Non uniform quantizer

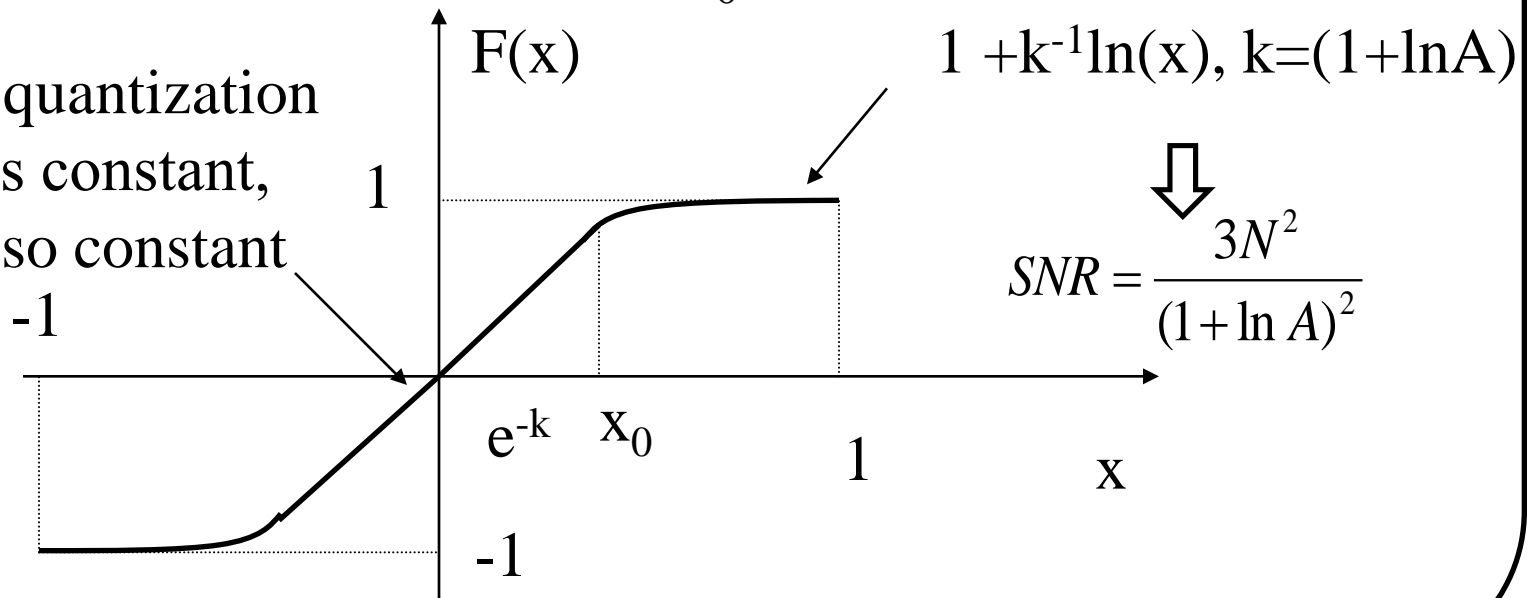
- $\mu$ -Law (North America)
- A-Law (Europe)



# A-Law

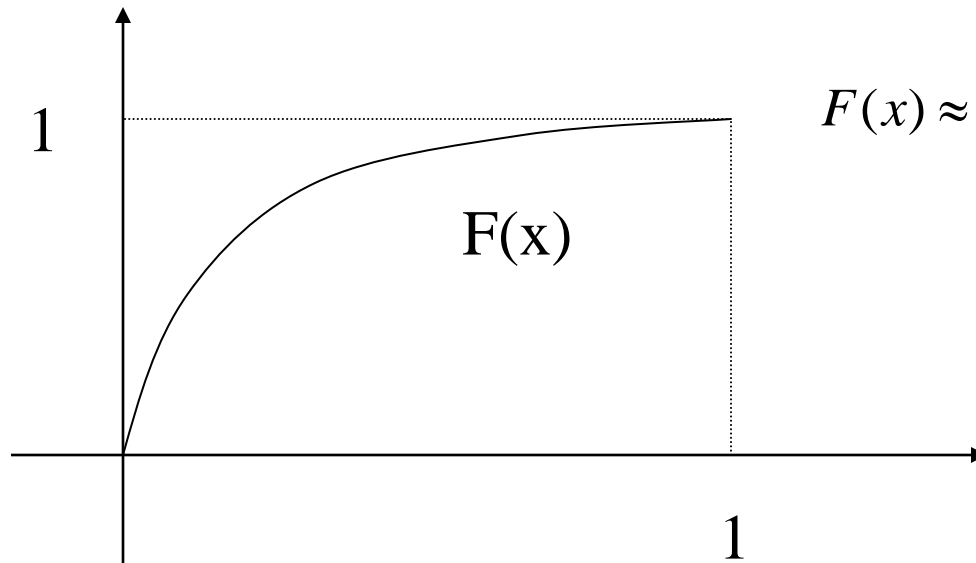
- $F(x) = (A \cdot x) / (1 + \ln A)$  if  $0 \leq x \leq x_0$  with  $x_0 = e^{(1-k)}$  and  $A = 1/x_0$
- $F(x) = (1 + \ln(Ax)) / (1 + \ln A)$  if  $x_0 \leq x \leq 1$

Uniform quantization  
since  $F'$  is constant,  
 $\Delta(x)$  is also constant



# $\mu$ -Law

$$F(x) = \frac{\ln(1 + \mu x)}{\ln(1 + \mu)} \quad 0 \leq x \leq 1$$
 This is not really of the correct form but for  $\mu \gg 1$  and  $\mu x \gg 1$ , you find

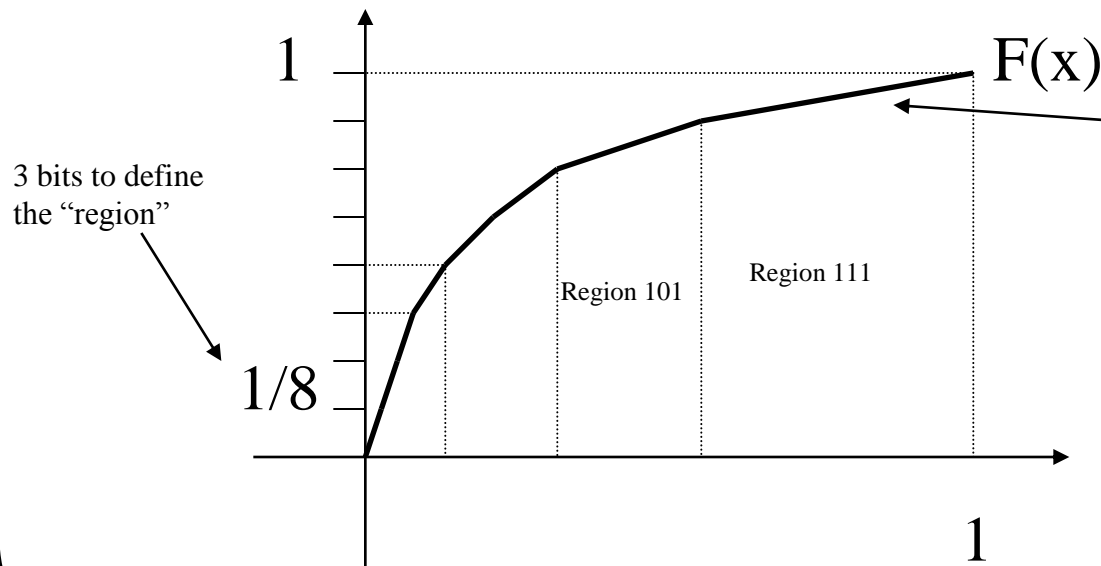


$$F(x) \approx 1 + \frac{\ln x}{\ln \mu} \quad \Rightarrow \quad k \approx \ln \mu$$

$$\Rightarrow SNR \approx \frac{3N^2}{(\ln \mu)^2}$$

# Segmented approximation

- To decrease complexity of processing:



Segmented approximation now,  $F'$  is constant on a segment so  $\Delta$  is and the bigger the slope is the smaller  $\Delta$  is.

# ***Adaptive Differential PCM***

---

- Takes advantage of the slow rate of change in the voice signal:
  - Quantizes and transmits the difference between consecutive samples
  - May use linear prediction of the signal

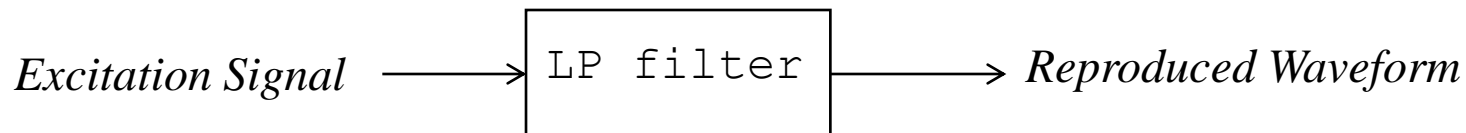
# CELP

## “Code Excited Linear Predictive”

---

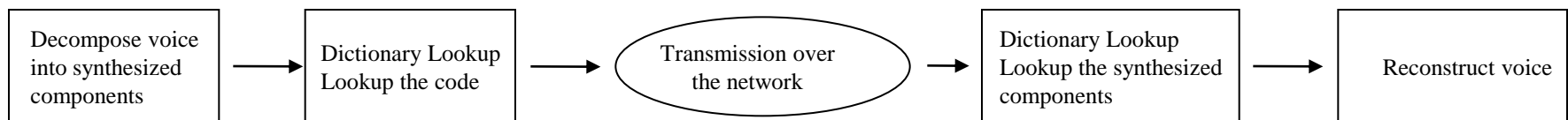
- Coder:
  - Voice is analyzed in frames of 10-30ms, and represented by:
    - synthesis filter
      - updated by linear prediction
    - excitation signal
      - optimally selected so as to minimize a “perceptually weighted measure of distortion)
      - makes use of a codebook
  - A data frame is produced and transmitted

- Decoder:



# VoCoding

- Vocoding (parametric):
  - This scheme works only for human voice. It is not optimized for other audio signals.
  - The parameters (called data frame) are:
    - Coefficients that define vocal tract resonance characteristics (indication whether the source is voiced or not)
    - excitation energy
    - pitch value



# VoCoding

---

- For very low bit rates ( $\cong 2\text{Kb/s}$ )
- Based on modeling the speech production mechanisms rather than the waveforms
  - Speech is processed in frames of 10-25ms
  - Distinction between *voiced* frames and *unvoiced* frames
    - Voiced speech: vocal cords vibrating (e.g., vowels)
    - Unvoiced speech: vocal cords held firm without vibration (e.g., consonants)
- Speech is represented by the following parameters
  - Coefficients that define vocal tract resonance characteristics
  - Excitation energy
  - Pitch value

## **VoCoding (2)**

---

- Low Quality
  - “unnatural”, “buzzy” character
- Works only for human speech
  - Not optimized for other audio signals.
- Little current interest
  - No international standard created

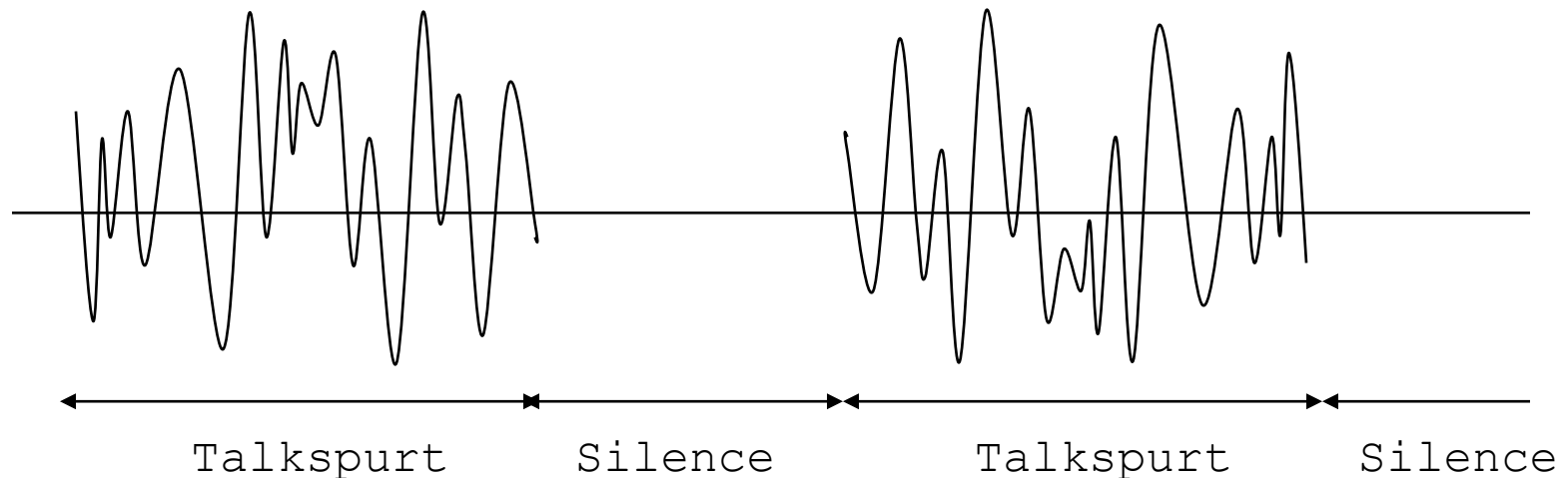
# Voice Coding Standards

	ITU Standard	Compression Method	Data Rate	MOS Score	Delay
PCM	G.711	PCM	64 kbps	4.4	0.75 ms
	G.726	ADPCM	32 kbps	4.2	1 ms
	G.728	LD-CELP	16 kbps	4.2	3-5 ms
CELP	G.729	CS-ACELP	8 kbps	4.2	10 ms
	G.729a	CS-ACELP	8 kbps	4.2	10 ms
	G.723.1	MP-MLQ	6.3 kbps	3.98	30 ms
	G.723.1	ACLEP	5.3 kbps	3.5	30 ms

- MOS (Mean Opinion Score): subjective measure of voice quality  
*Scores of 4 to 5 are deemed toll quality; 3 to 4, communication quality; and less than 3, synthetic quality.*
- CELP: Code book Excited Linear Prediction
- LD: Low Delay
- CS - ACELP: Conjugate Structure - Algebraic CELP
- MP-MLQ: Multi-Pulse Excitation with a Maximum Likelihood Quantizer

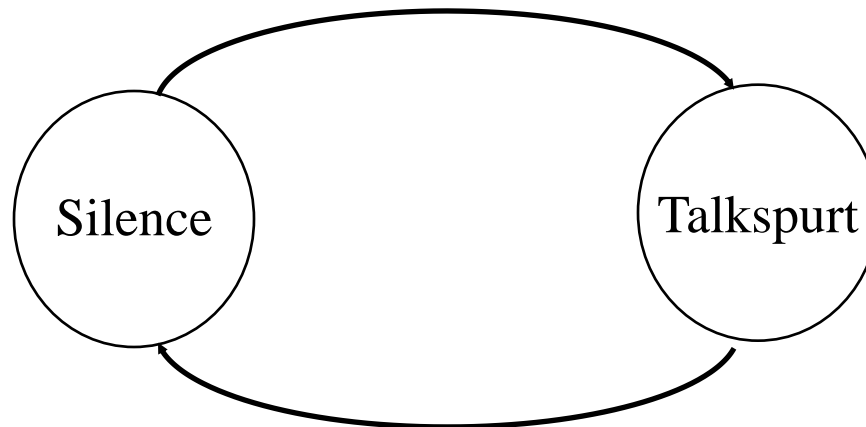
# Speech Activity

- Speech Alternates between two states:
  - Silence (no speech is present)
  - Talkspurt (speech is present)



# Speech Activity (2)

---



Average time  $\approx 1.8$  sec

Average time  $\approx 1.2$  sec

One speaker talking : 64 - 73 %  
Both speakers talking: 3 - 7 %  
Both speakers silent : 33 - 20 %

# Silence Suppression

---

- Voice Activity Detector (VAD)
  - When silence is detected, background noise is transmitted
  - When speech is detected, full fixed bit rate stream is transmitted
- About 60% reduction in data rate
  - Resulting traffic is no longer constant bit rate
  - Statistical Multiplexing gain may be significant