EE3731C: Signal Processing Methods

Lecture II-6:

Example Applications

National University of Singapore

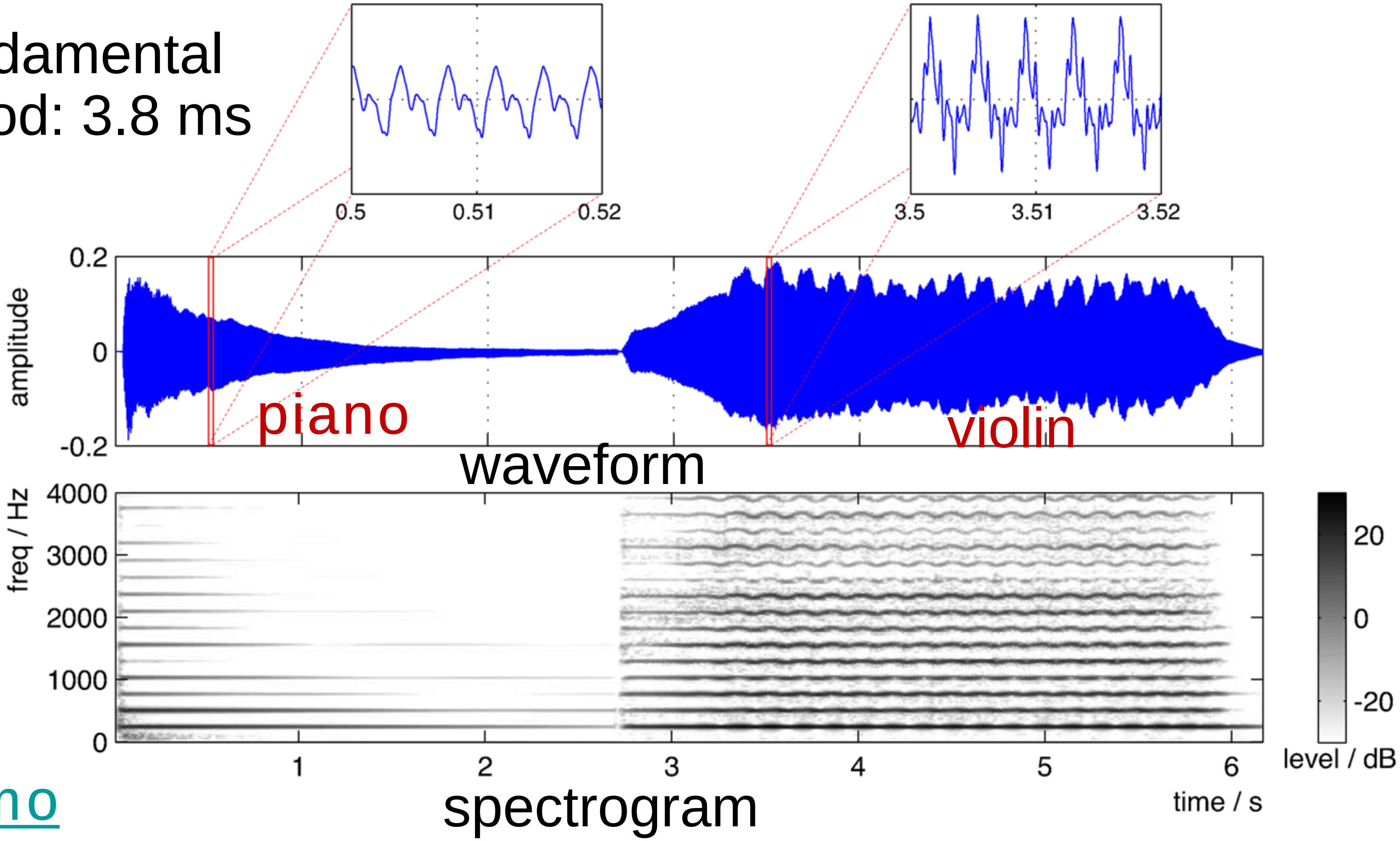
Outline

- Audio Processing
  - –Music signal processing
  - –Speech processing
  - –Echo cancellation
- Visual Processing
  - –Image/video denoising
  - –Image/video compression
  - –Medical image processing
- Audio-Visual Signal Processing

Music Signal Processing: Pitch

- Middle C (262 Hz)

Fundamental period: 3.8 ms



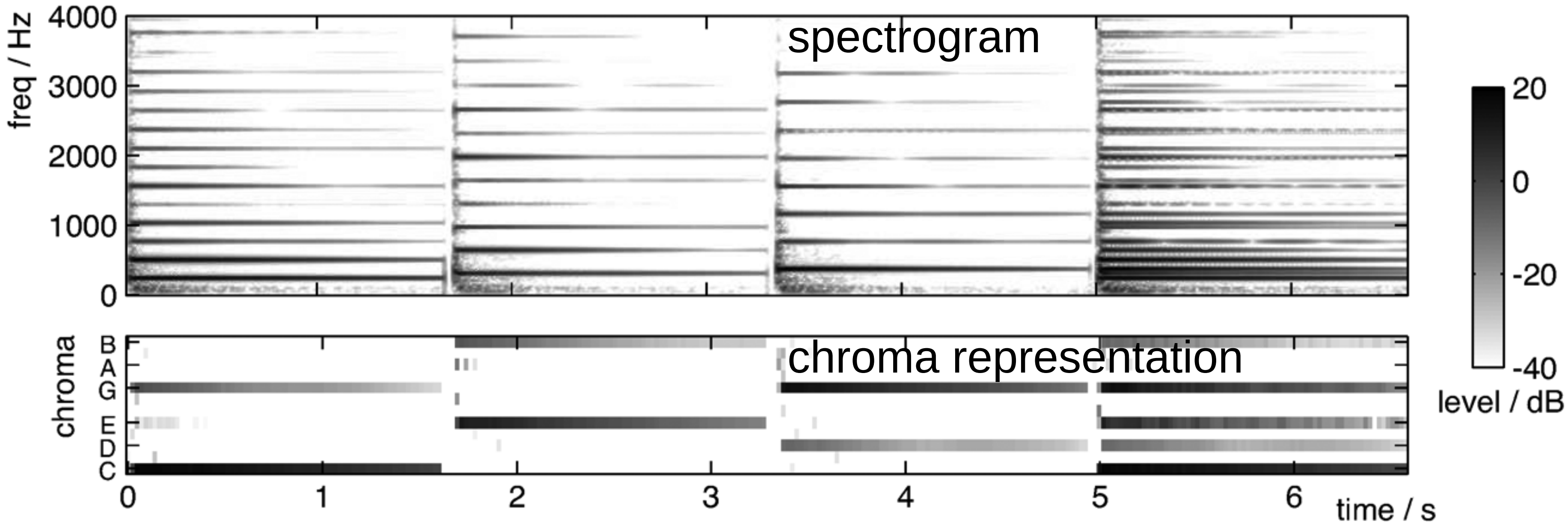
Demo

Music Signal Processing: Chroma

- Chroma
  - A particular pitch can be specified by the concatenation of a chroma and an octave number. The lowest note on a piano is A0 (27.5 Hz), the highest note is C8 (4186 Hz), and middle C (262 Hz) is C4.
- Chroma features
  - A powerful representation for music audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones (or chroma) of the musical octave.
  - It encodes the short-time energy distribution of the underlying music signals over the 12 chroma bands, which correspond to the 12 traditional pitch classes.

Music Signal Processing: Harmony

- Another essential aspect of music is harmony, the simultaneous presentation of notes at different pitches.



Middle C (262 Hz)

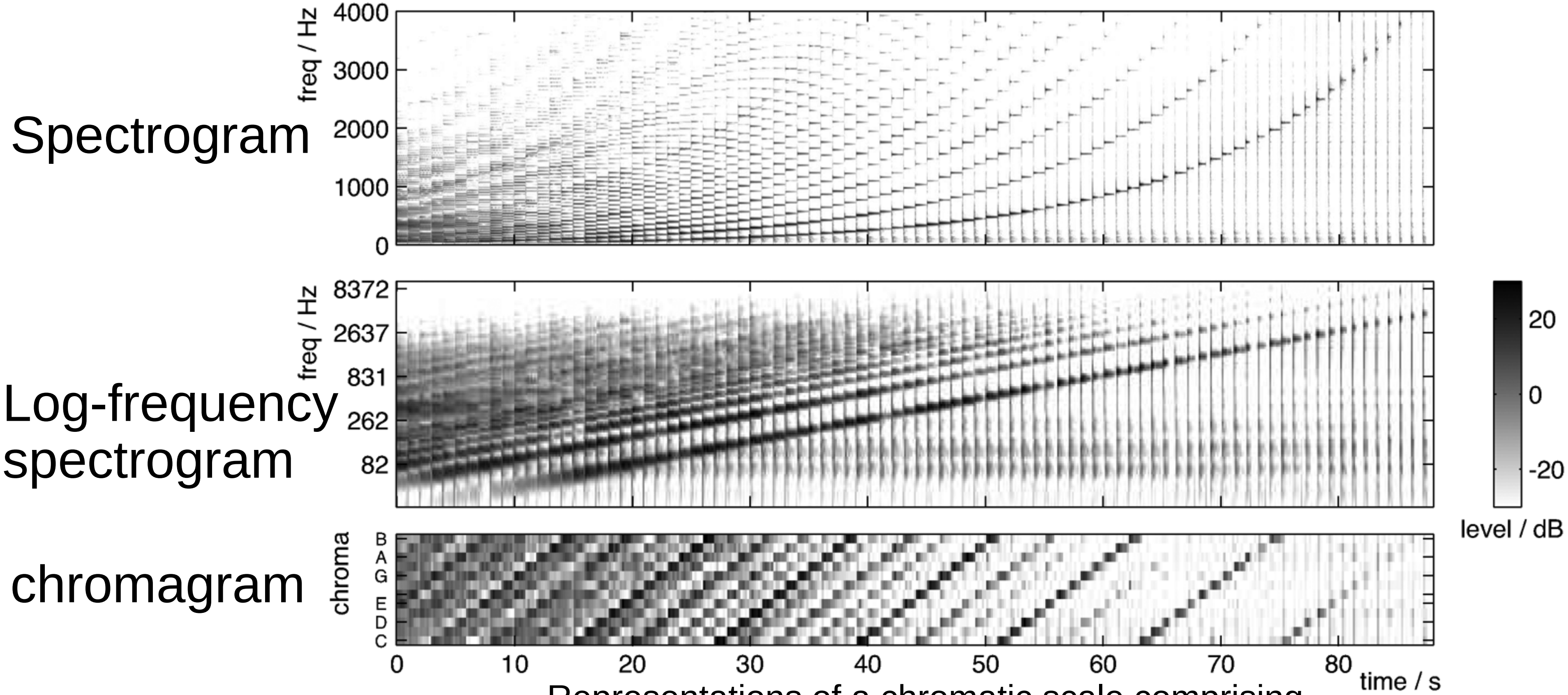
E (330 Hz)

G (392 Hz)

All three notes together - C Major triad

Representations

- Time–Frequency Representations
- Log-Frequency Spectrogram
- Time–Chroma Representations



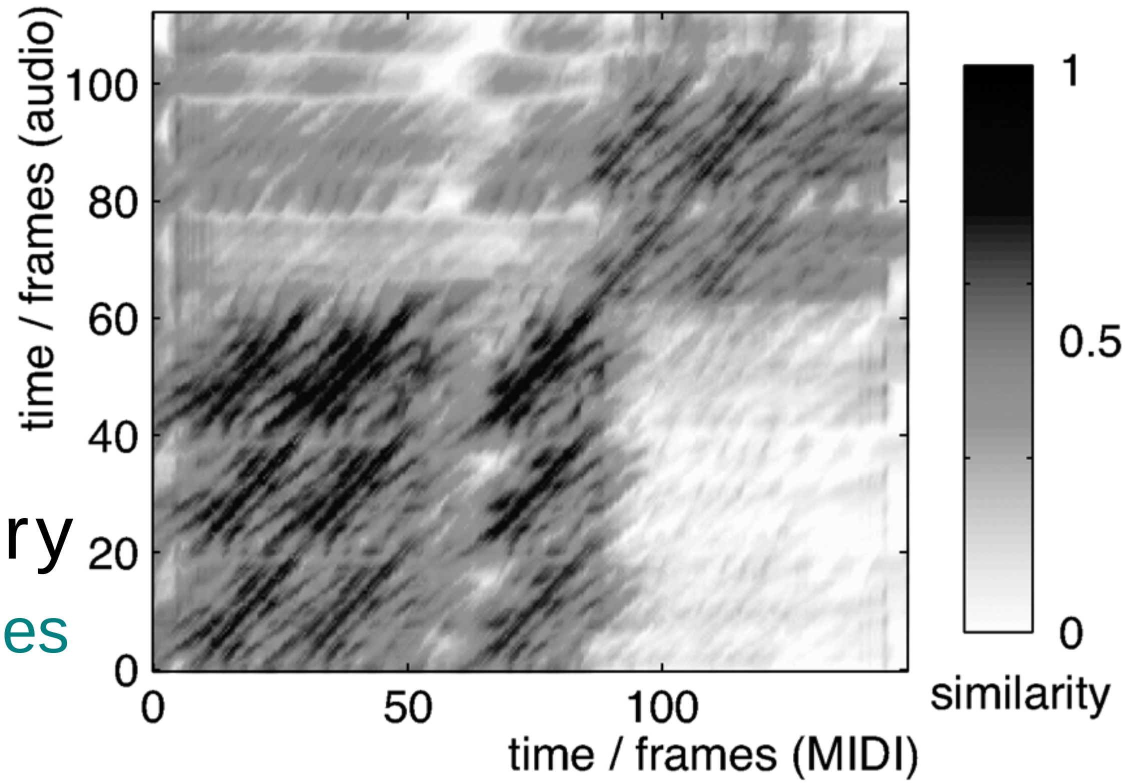
Representations of a chromatic scale comprising every note on the piano from lowest to highest

Toolboxes

- Chroma Toolbox:
  - Contains MATLAB implementations for the extraction of various musically meaningful features (e.g., pitch-based and chroma-based) from waveform based audio signals
  - http://www.mpi-inf.mpg.de/resources/MIR/chromatoolbox/
- Chroma Feature Analysis and Synthesis
  - http://www.ee.columbia.edu/~dpwe/resources/matlab/chroma-ansyn/

Example Applications

- Chord Recognition
- Synchronization and Alignment
- "Cover Song" Detection
  – Music Information Retrieval Evaluation eXchange (MIREX)
- Structure Recovery
  – Find off-diagonal ridges and segment into repeating phrases



Similarity matrix comparing a MIDI version of "And I Love Her" horizontal axis) with the original Beatles recording (vertical axis).

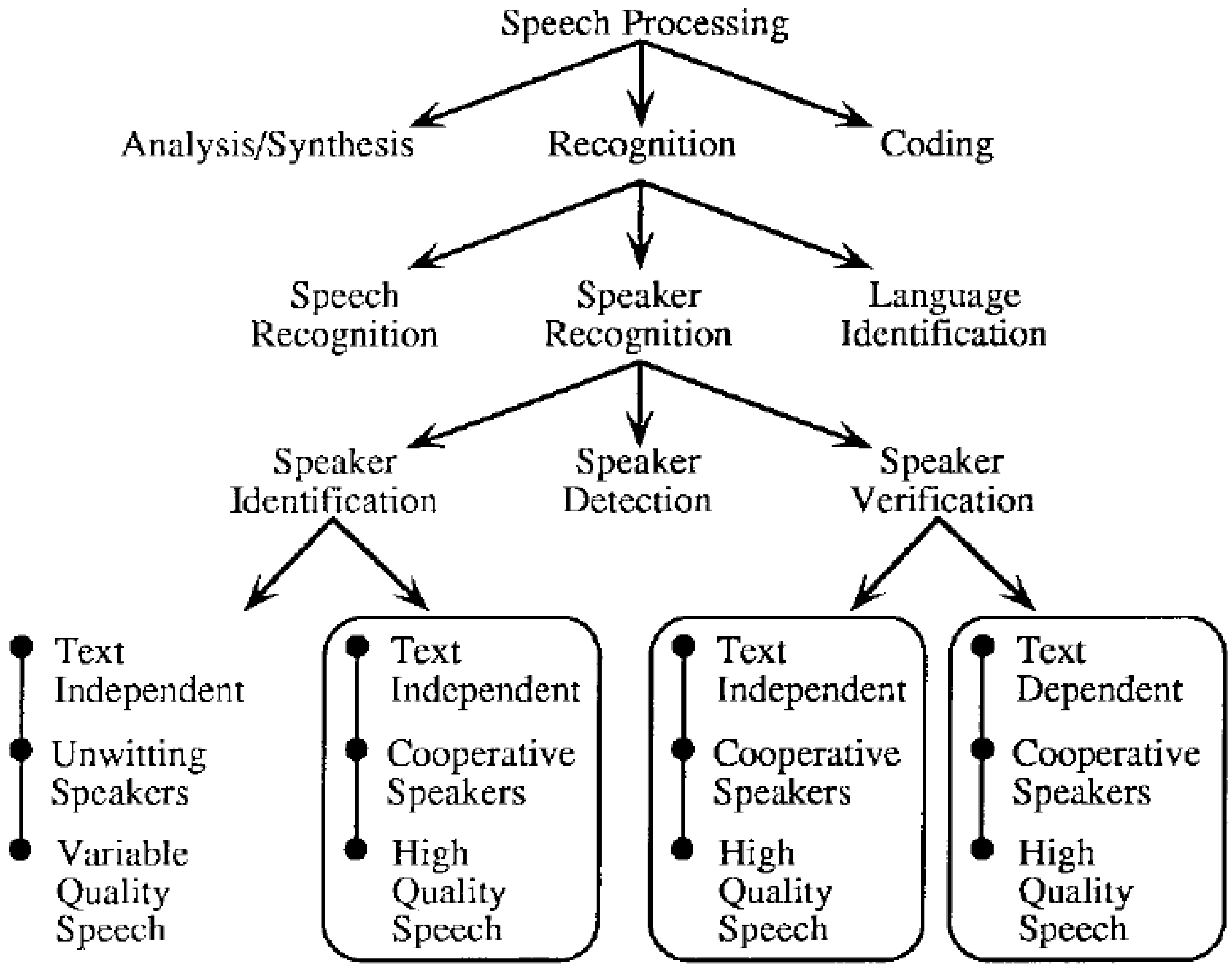
Example Applications

- Music coding for efficient storage and transmission of music signals
- Noise reduction and distortion equalization
- Music synthesis, pitch modification, audio mixing, audio morphing, and audio editing
- Music transcription and content classification, music search engines
- Music sound effects

Online Resources

- International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL)
  - Music Information Retrieval Evaluation eXchange (MIREX)
  - Networked Environment for Music Analysis (NEMA)
  - Structural Analysis of Large Amounts of Music Information (SALAMI)
- Music retrieval demo
  - http://www.rotorbrain.com/foote/musicr/

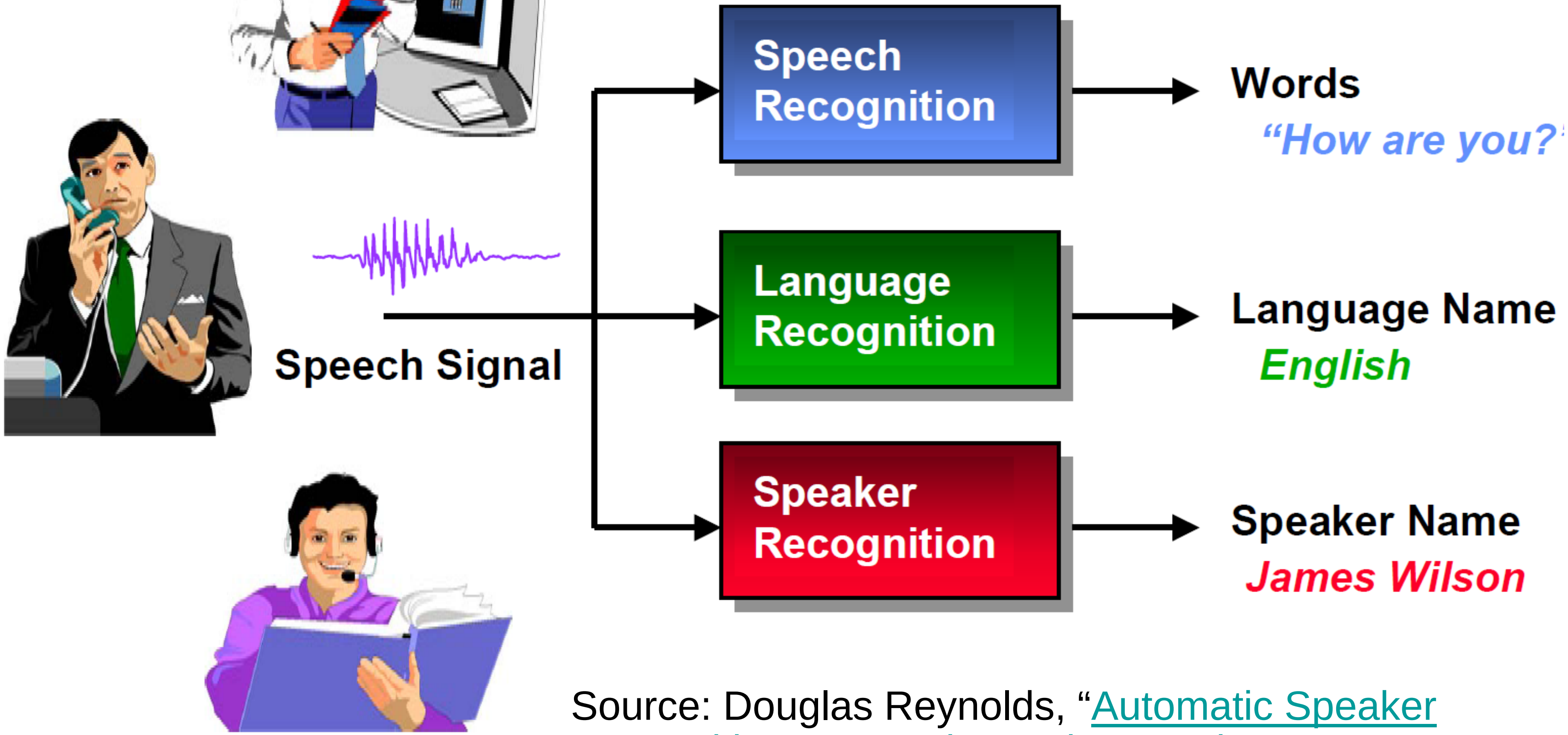
Speech Processing



Source: J.P. Campbell, "Speaker Recognition: A Tutorial"

Speech Processing

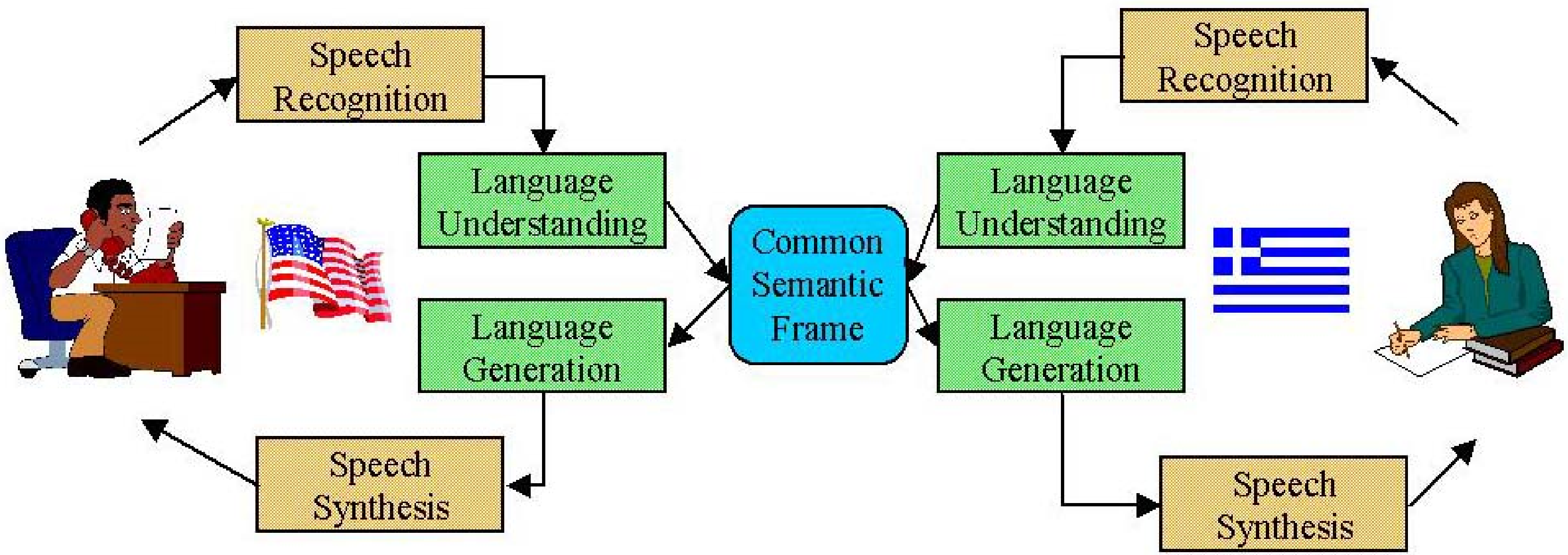
Goal: Automatically extract information transmitted in speech signal



Source: Douglas Reynolds, "Automatic Speaker Recognition - Acoustics and Beyond"

Speech Recognition

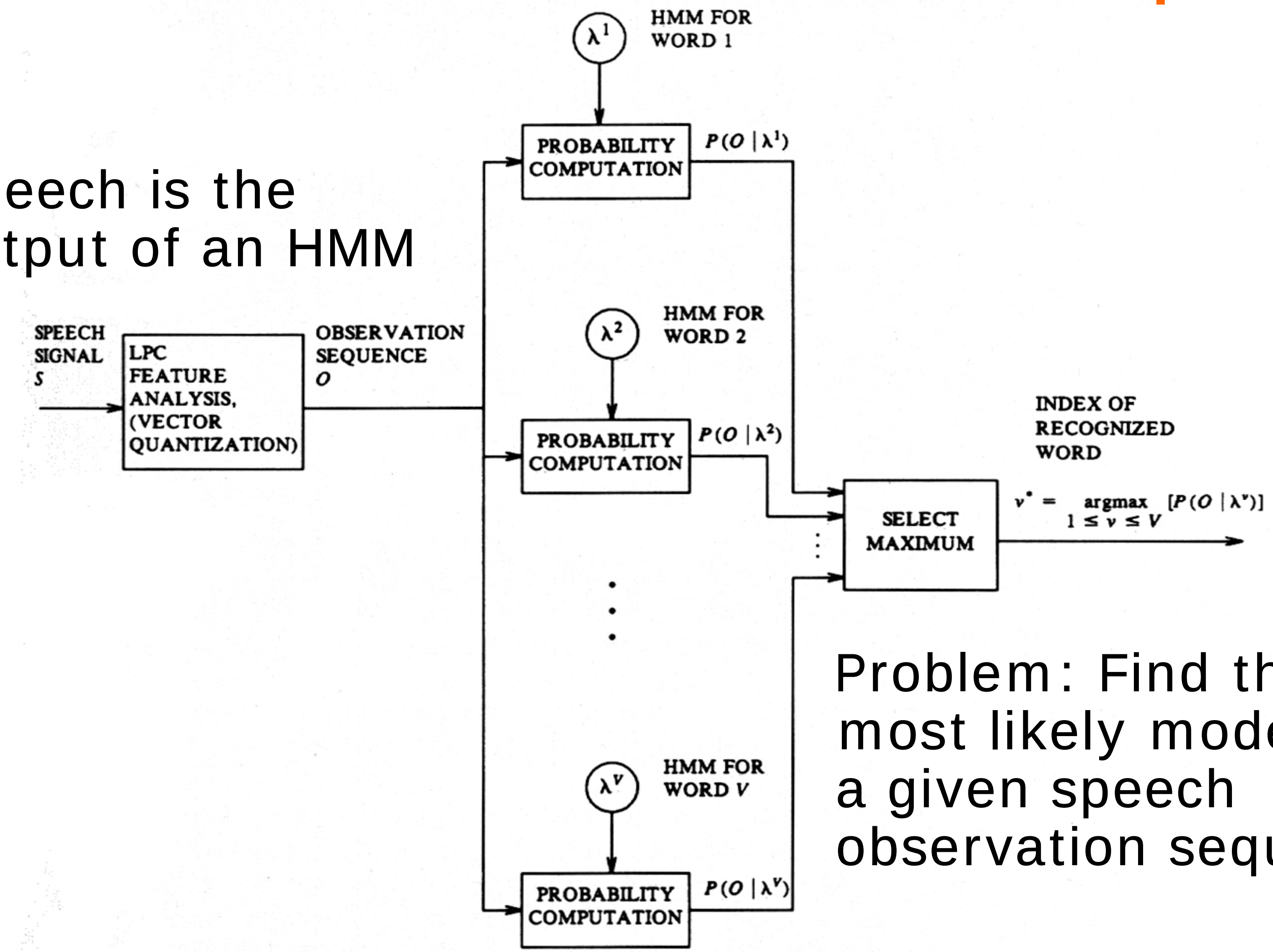
- Speech recognition: translation of spoken words into text
- It can be speaker independent or speaker dependent



From: http://oxygen.lcs.mit.edu/Speech.html

Hidden Markov Models for Speech

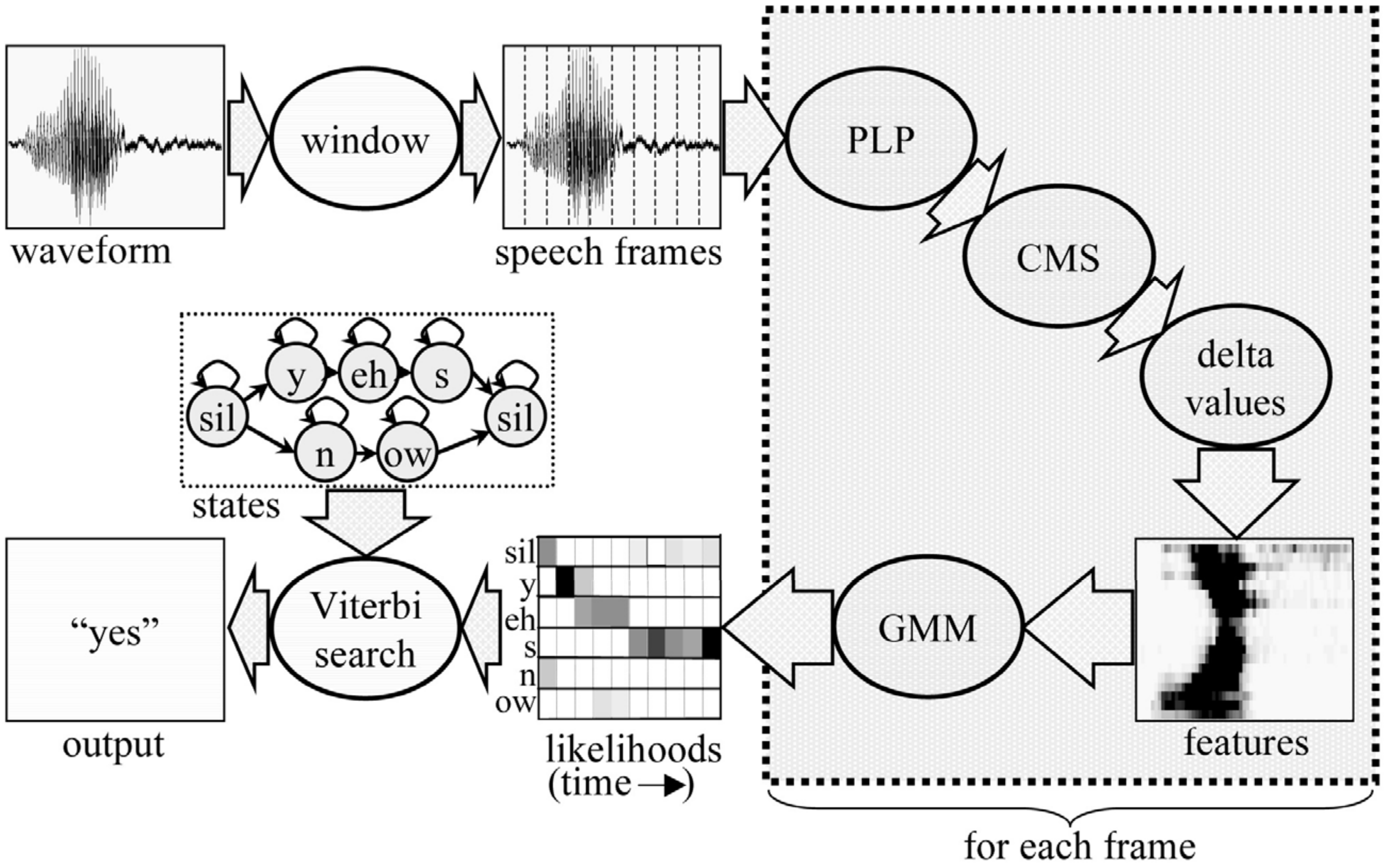
Speech is the output of an HMM



Problem: Find the most likely model for a given speech observation sequence

Block diagram of an isolated word HMM recognizer

Speech Recognition



From: Encyclopedia of Information Systems, 2002

Speaker Recognition

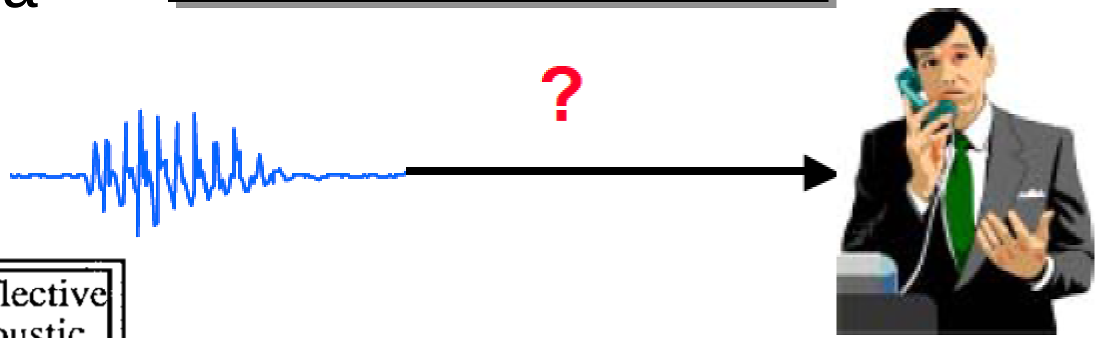
- Speaker recognition encompasses verification and identification.
- Automatic speaker verification is the use of a machine to verify a person's claimed identity from his/her voice.
- In automatic speaker identification, there is no *a priori* identity claim, and the system decides who the person is, or that the person is unknown.

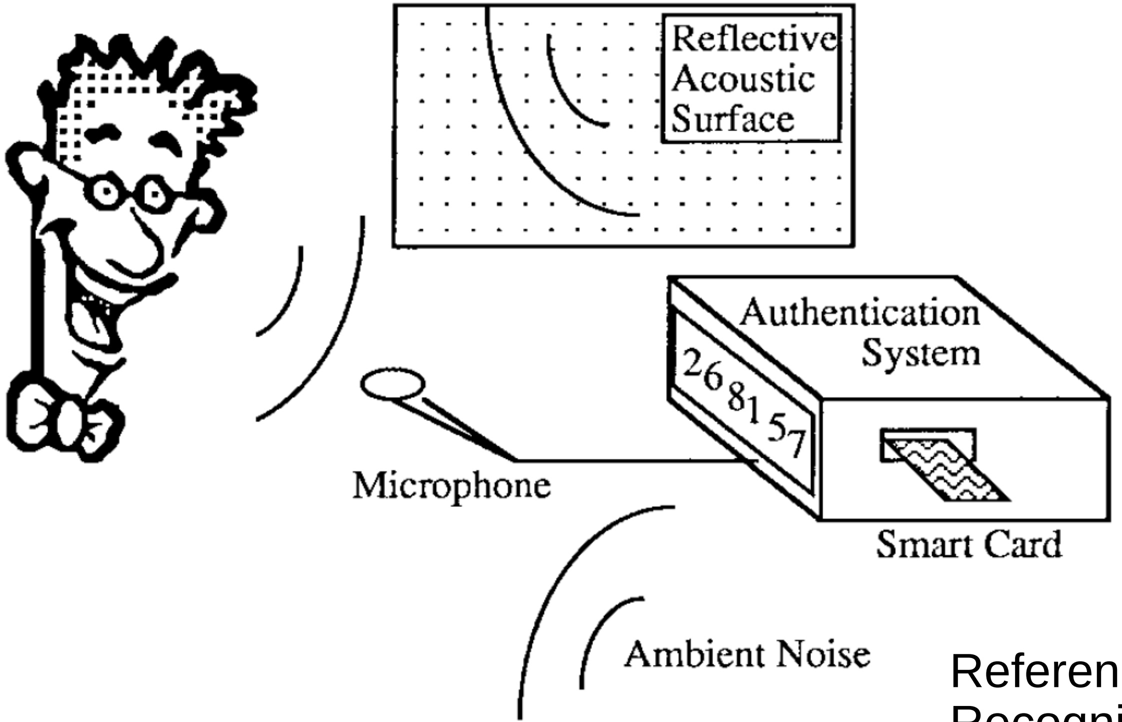
Reference: J.P. Campbell, "Speaker Recognition: A Tutorial"

Speaker Verification

Problem: Deciding whether the speech belongs to a claimed identity

Is this Bob's voice?

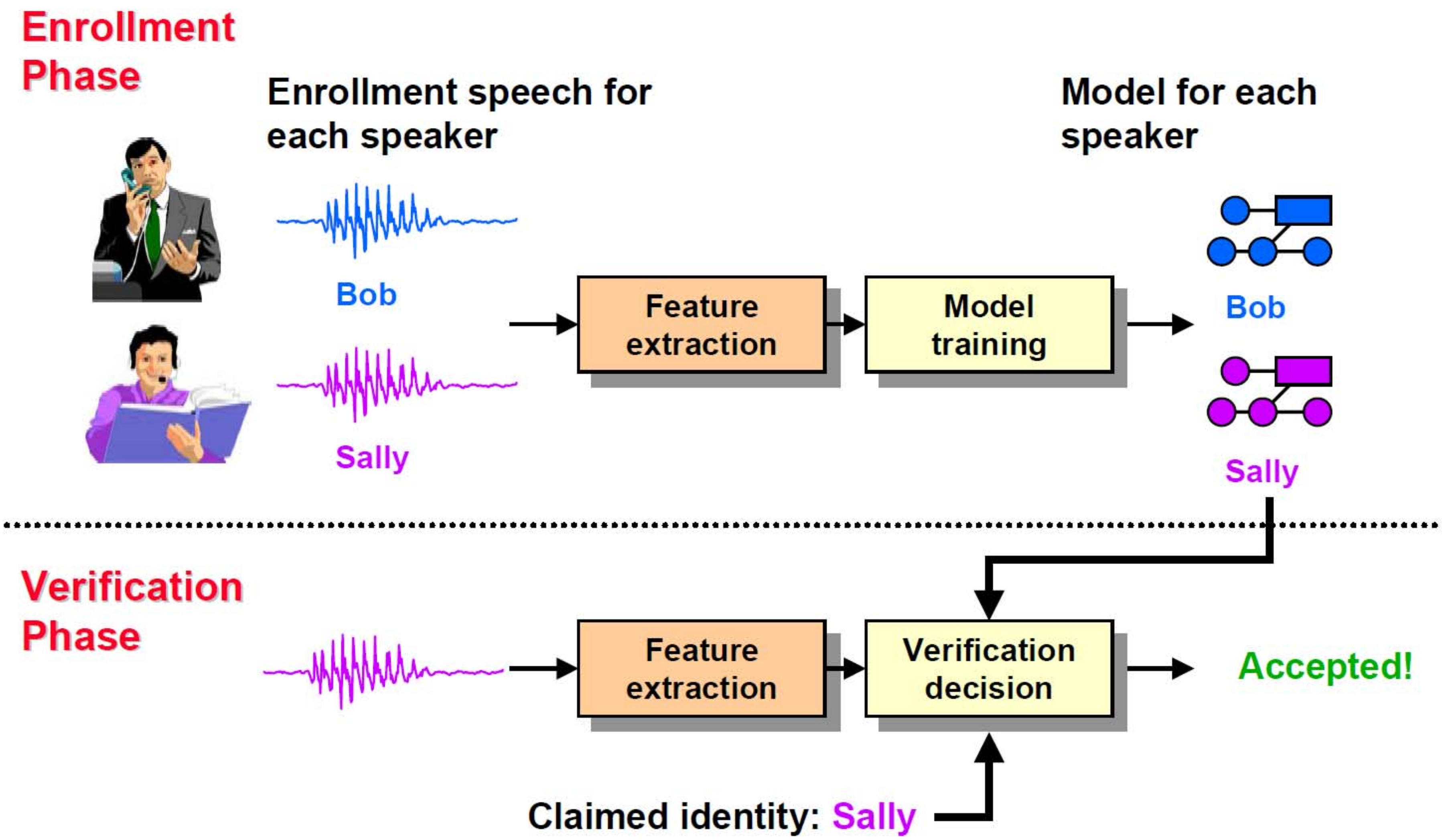




Typical speaker-verification setup

Reference: J.P. Campbell, "Speaker Recognition: A Tutorial"

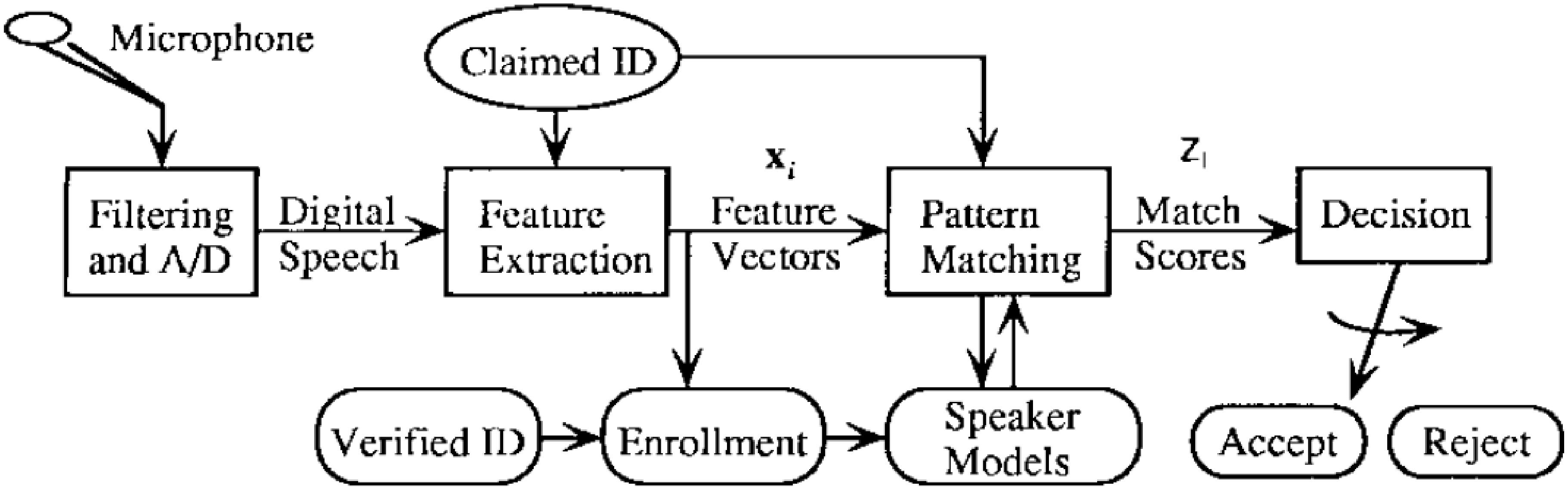
Speaker Verification: Two Phases



Source: Douglas Reynolds, “Automatic Speaker Recognition - Acoustics and Beyond”

Speaker Verification

- Generic speaker-verification system



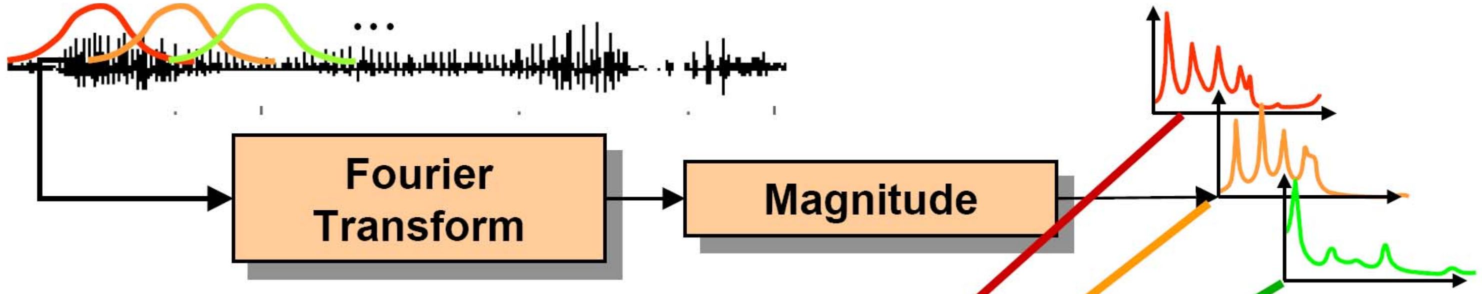
- Pattern matching
  - –Template models
  - –Stochastic models

Reference: J.P. Campbell, “Speaker Recognition: A Tutorial”

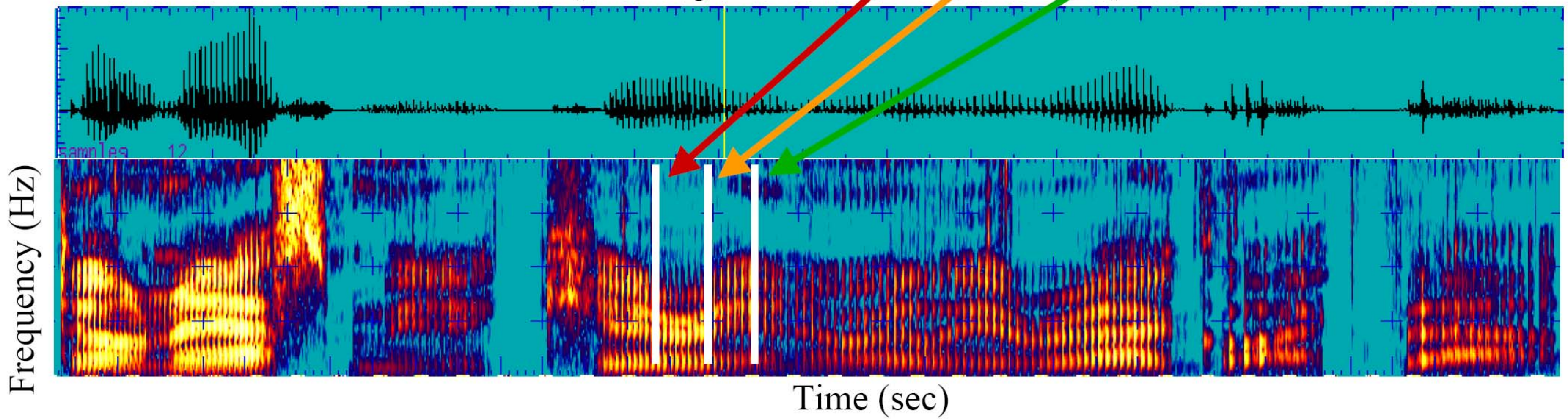
Features for Speaker Recognition

Speech is a continuous evolution of the vocal tract

- Need to extract time series of spectra
- Use a sliding window - 20 ms window, 10 ms shift

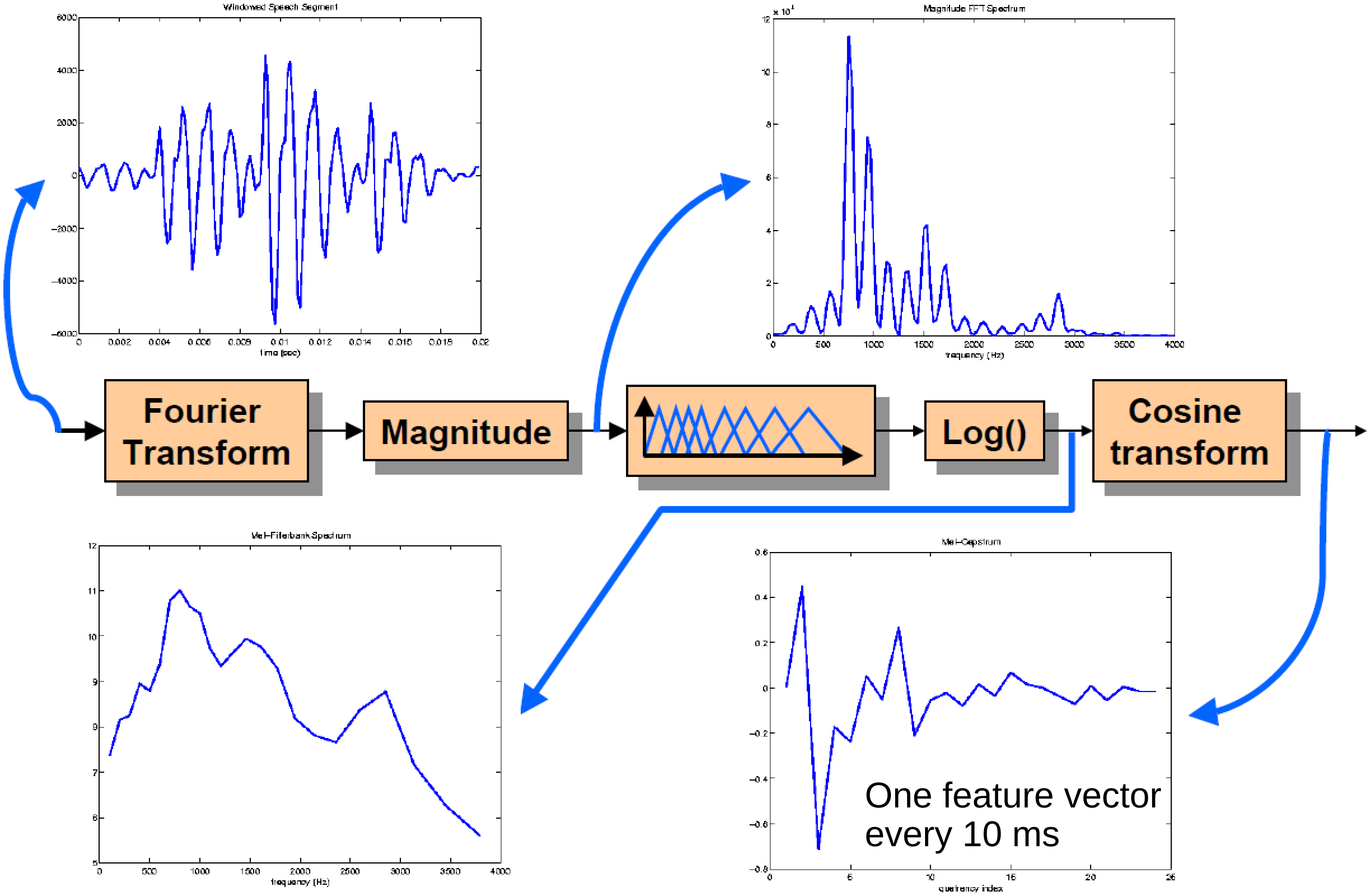


- **Produces time-frequency evolution of the spectrum**

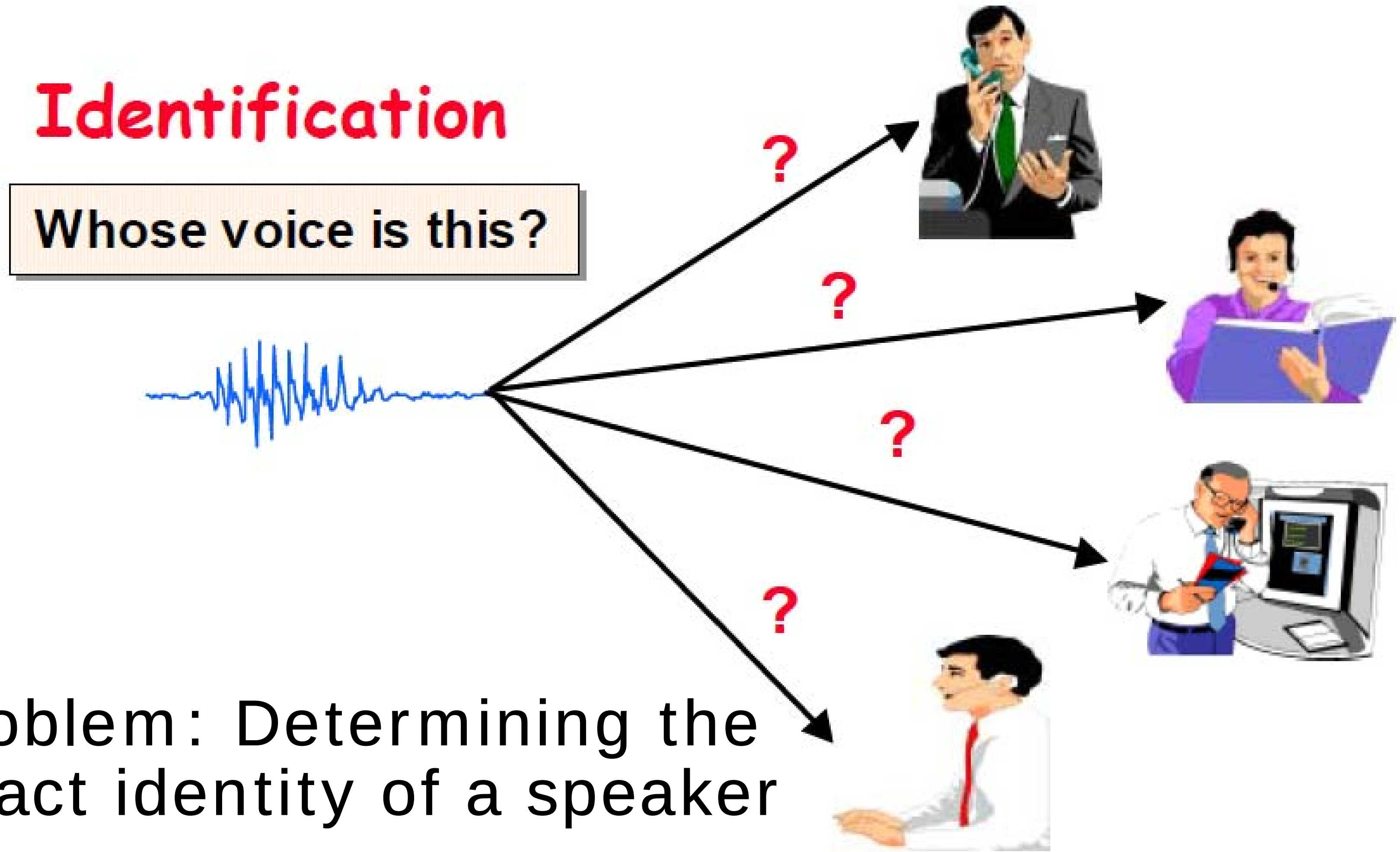


Source: Douglas Reynolds, "Automatic Speaker Recognition - Acoustics and Beyond"

Features for Speaker Recognition



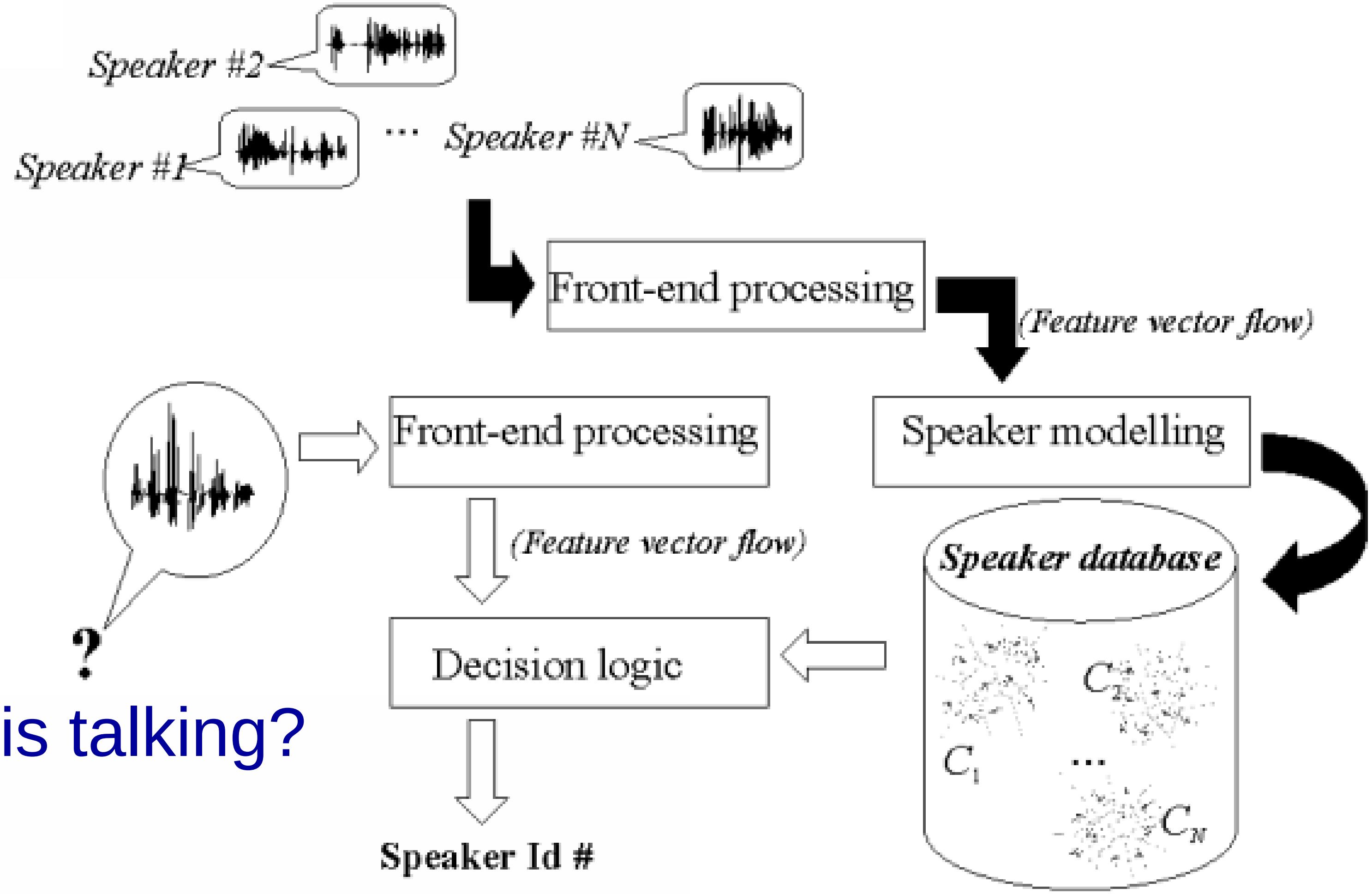
Speaker Identification



Problem: Determining the exact identity of a speaker

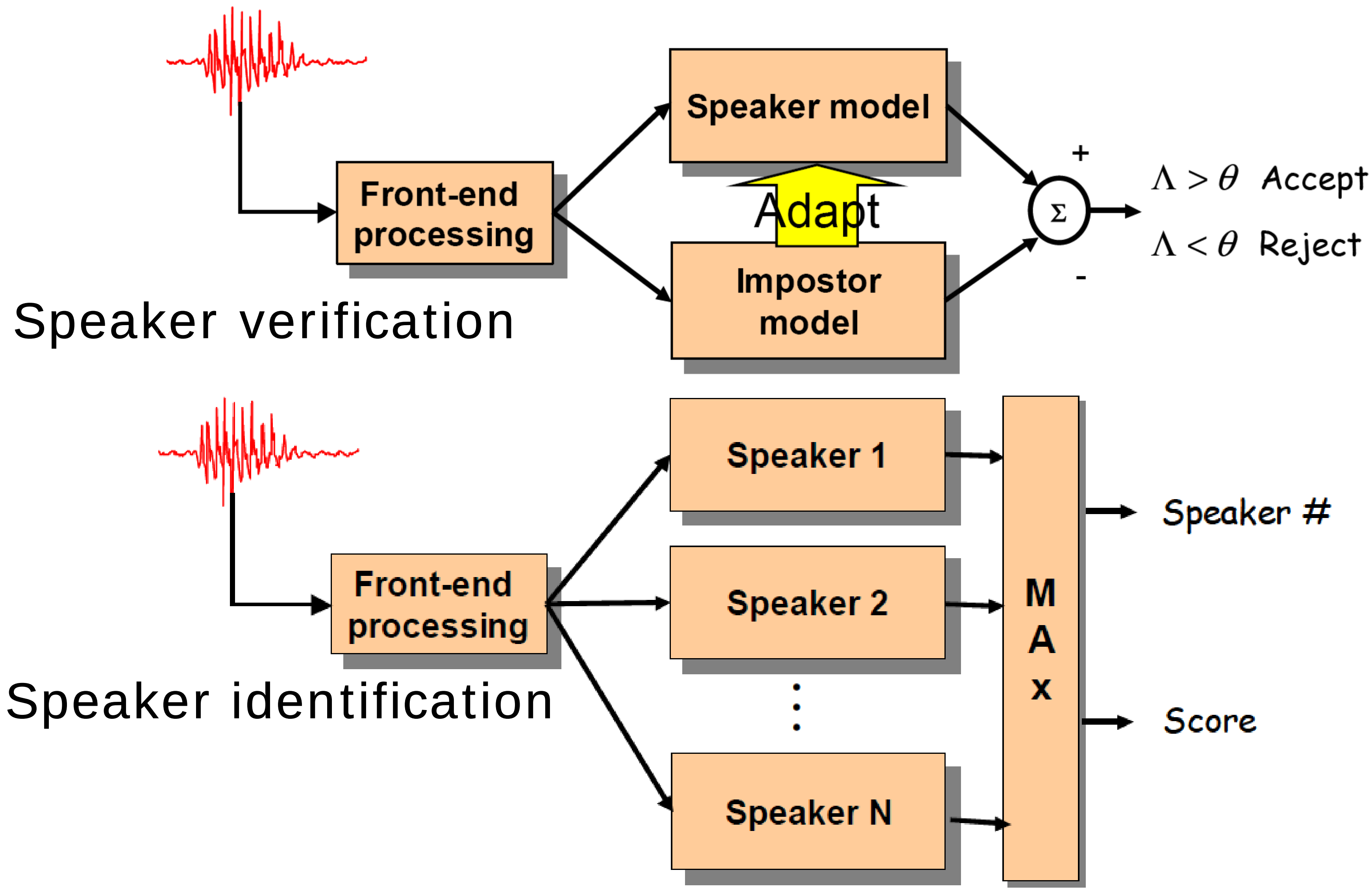
Source: Douglas Reynolds, “Automatic Speaker Recognition - Acoustics and Beyond”

Speaker Identification



Who is talking?

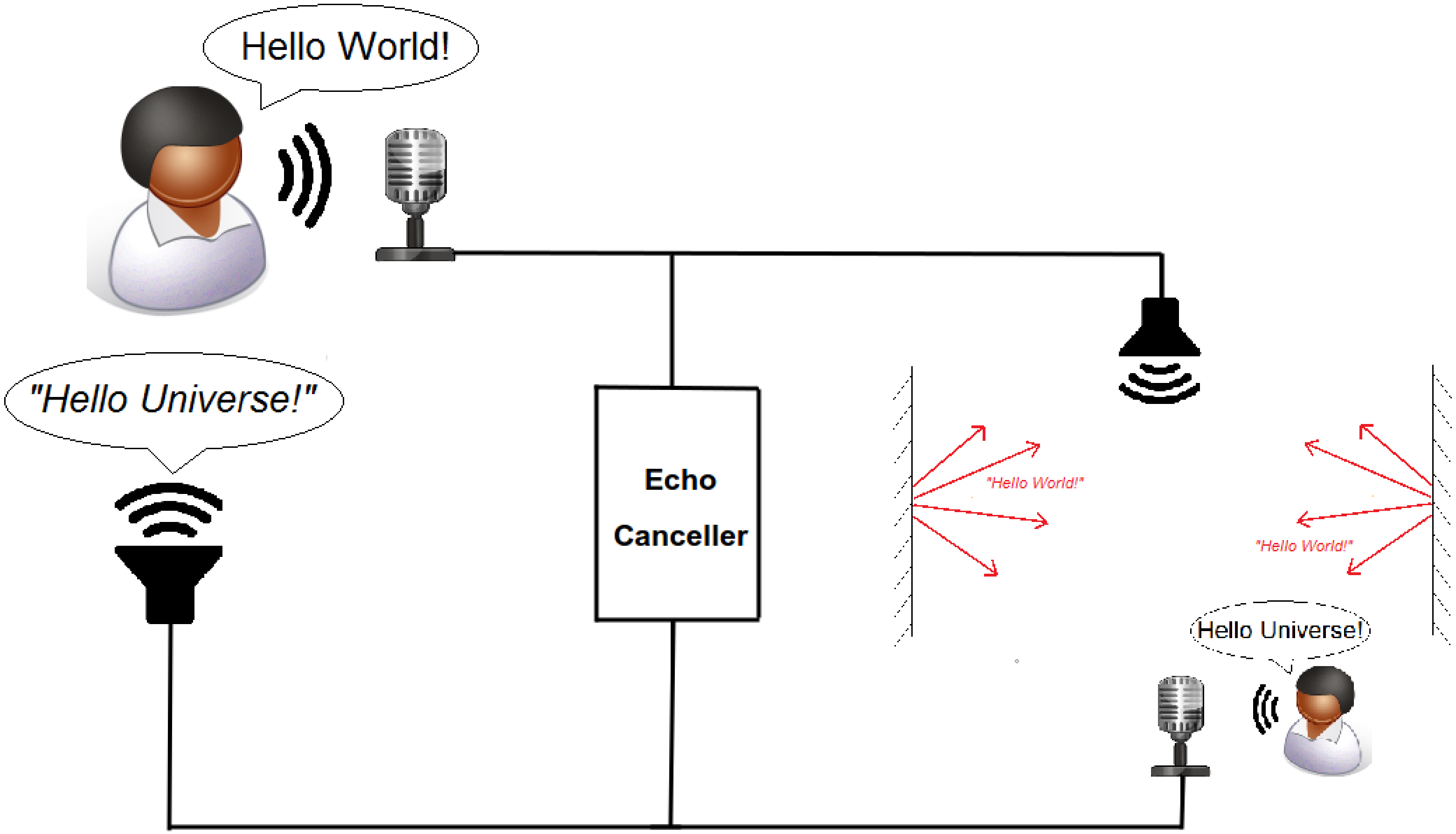
Speaker Recognition Systems



Source: Douglas Reynolds, "Automatic Speaker Recognition - Acoustics and Beyond"

Echo Cancellation

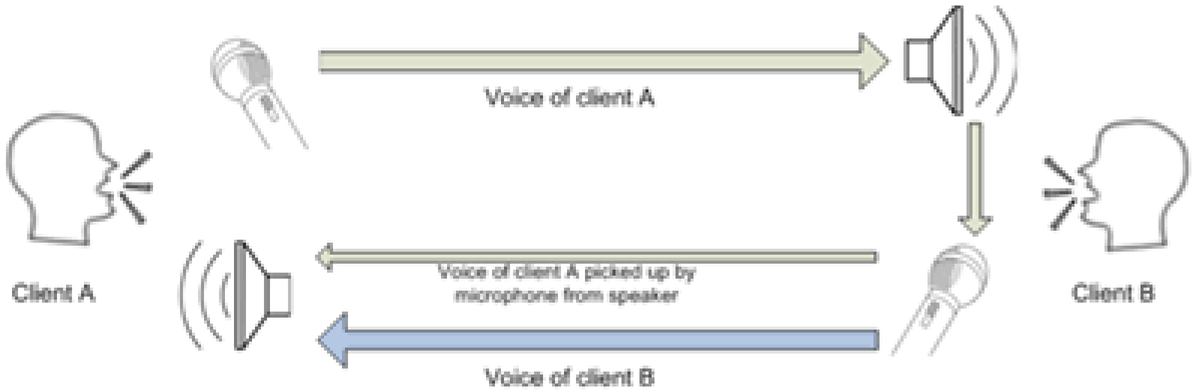
- An echo is what results when a signal bounces off an object, which is perceived some time later as a lower volume version of the original signal.



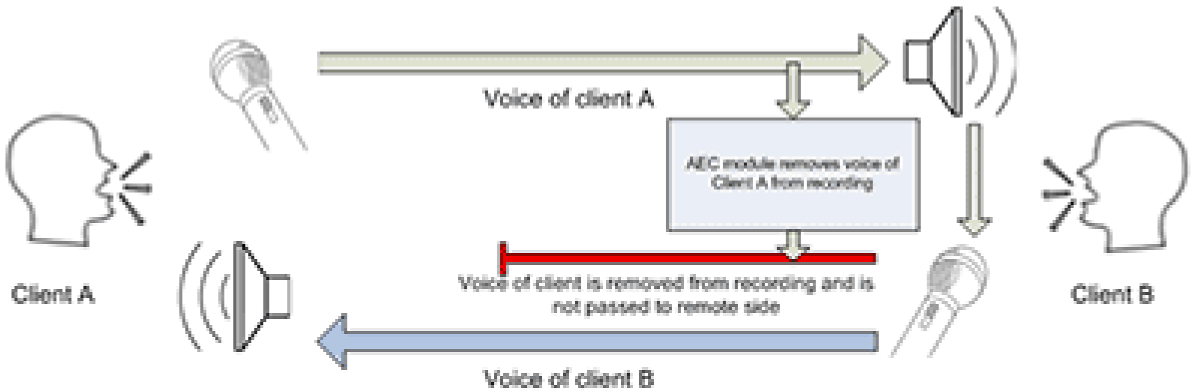
Echo Cancellation

- Echo cancellation: the process of removing echo from a voice communication in order to improve voice quality.

Client A hears his/her own voice.



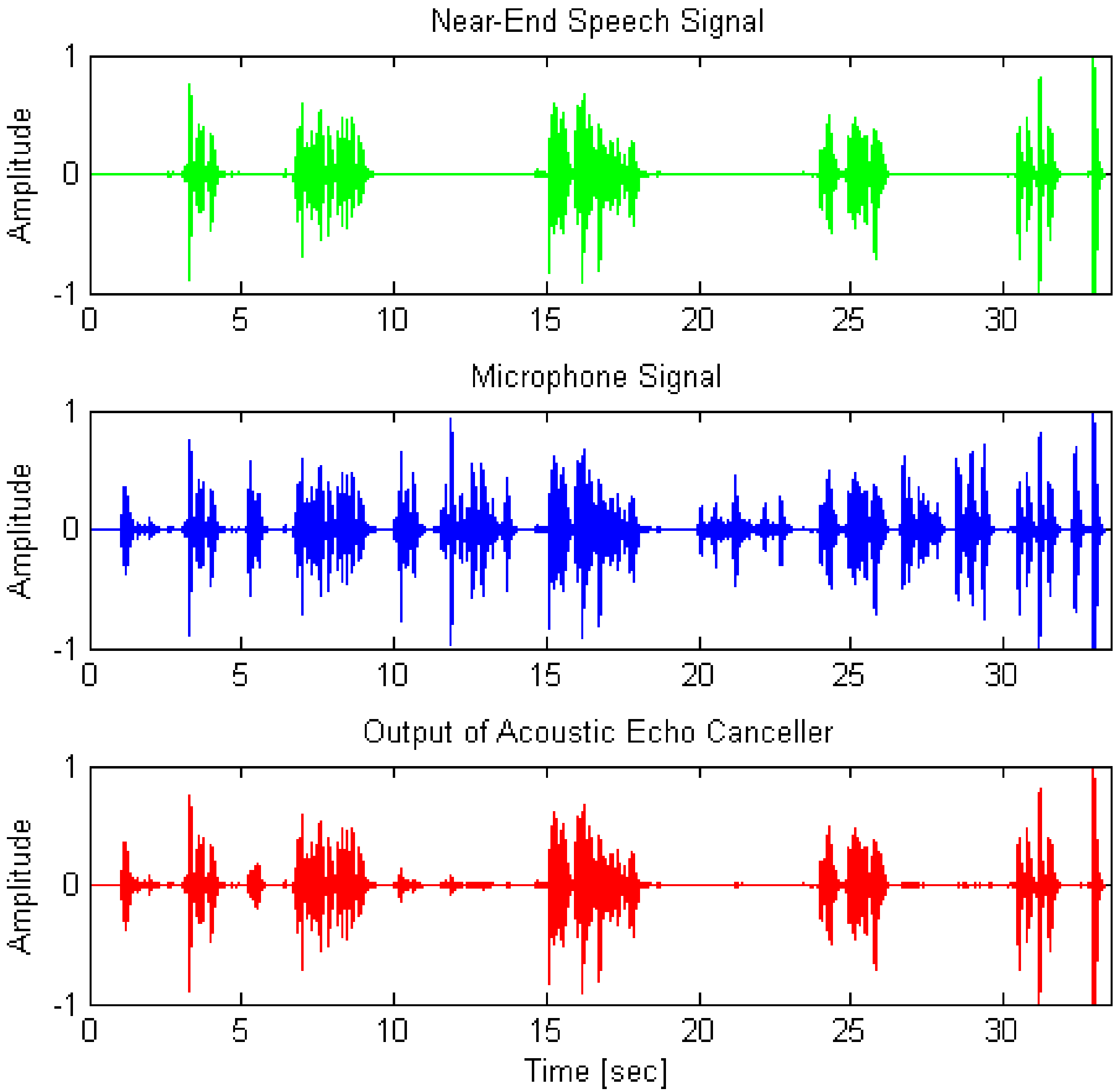
Acoustic Echo Cancellation (AEC)



Echo Cancellation: Example

Microphone signal contains both the near-end speech signal and the far-end echoed speech signal.

The goal of AEC is to cancel out the far-end speech, such that only the near-end speech is transmitted back to the far-end listener.

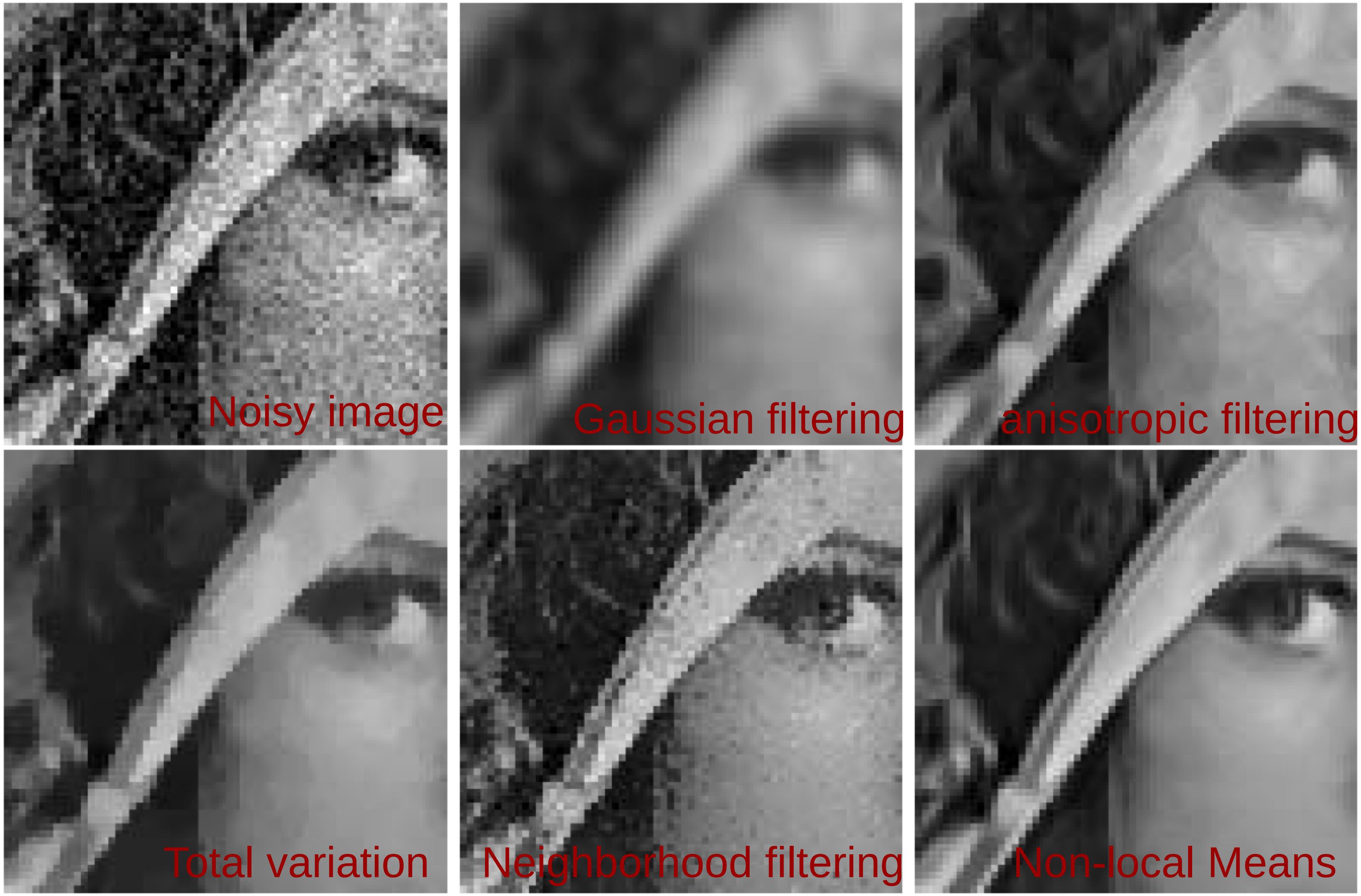


From: http://www.mathworks.com

Online Resources

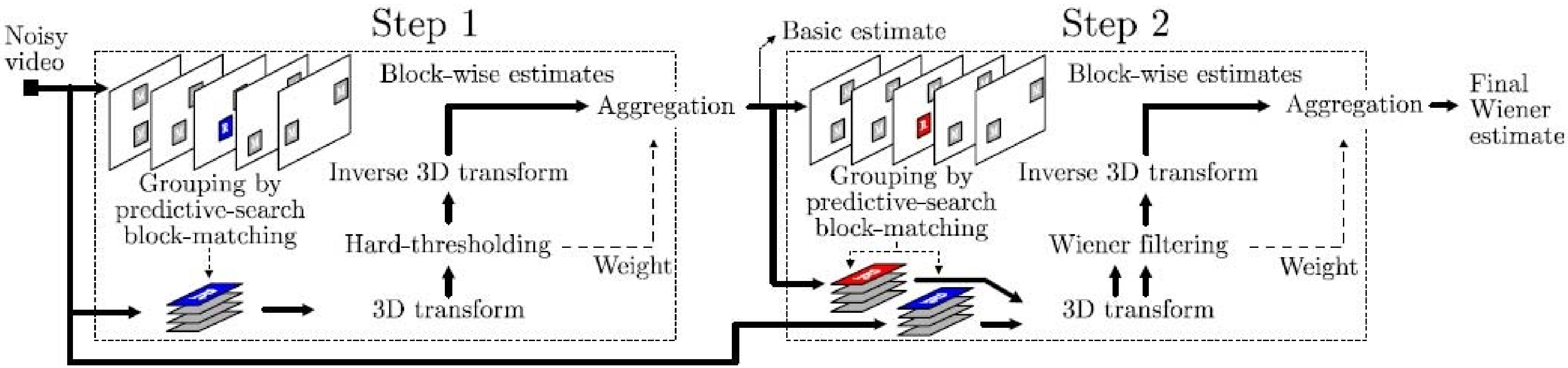
- MATLAB Demo and Examples:
  - Acoustic echo cancellation using adaptive filters
  - Speech Processing using MATLAB
  - Audio processing examples
- Video Demos
  - https://www.youtube.com/watch?v=c2Xy21dAxOU
  - https://www.youtube.com/watch?v=RJmVk56Bl94
- Tutorials
  - Speech Recognition Tutorial
  - Tutorial on Hidden Markov Models

Image Denoising

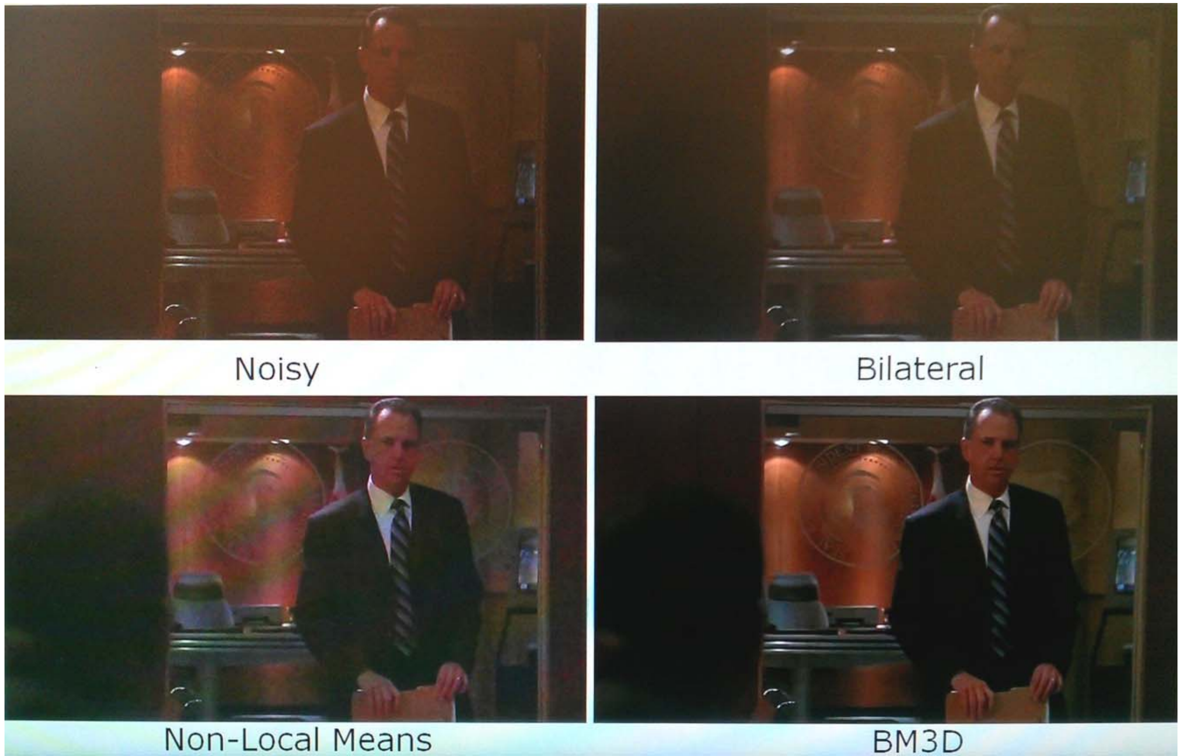


From: A. Buades, "A non-local algorithm for image denoising," CVPR 2005.

Video Denoising



[Block-matching and 3D filtering (BM3D)]

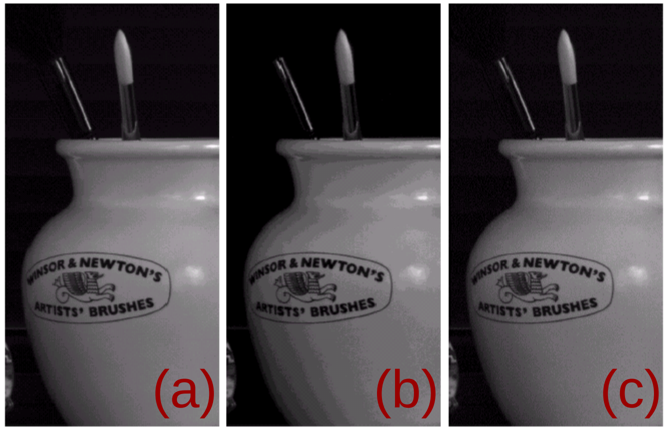


[Video Demo#1]

[Video Demo#2]

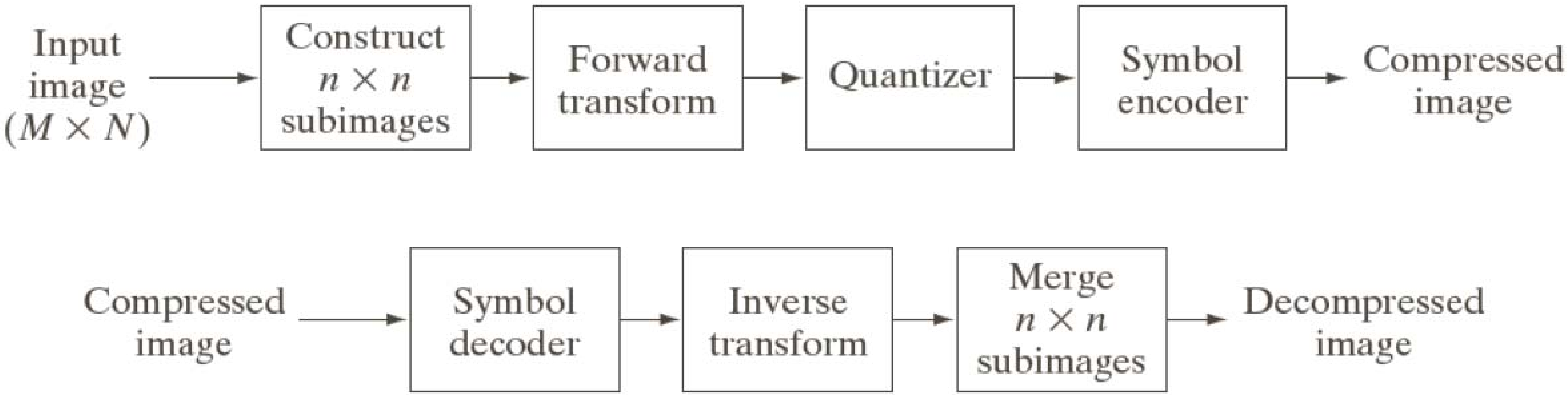
Image/Video Compression

- Goals of compression:
  - –Remove redundancy
  - –Reduce irrelevance



(a) Original image. (b) Uniform quantization to 16 levels. (c) IGS quantization to 16 levels.

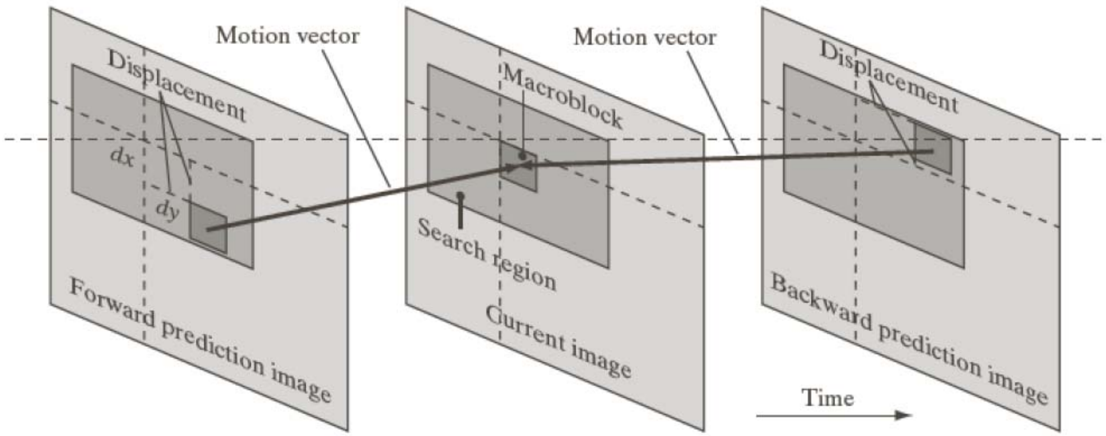
- Image compression

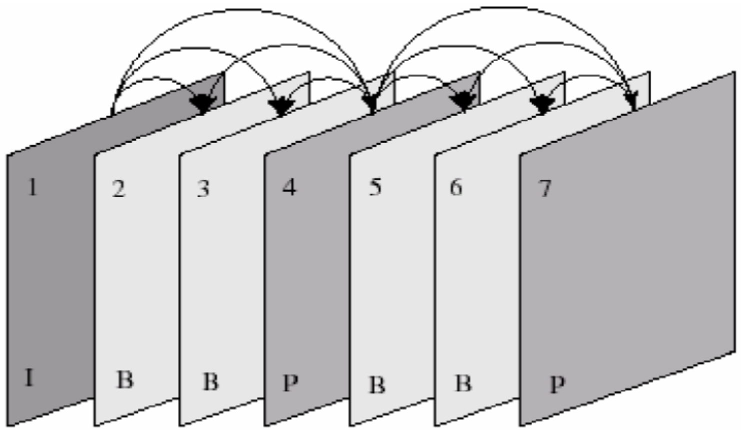


A Block transform coding system

Image/Video Compression

- Video Compression
  - –Simplest way: compress each frame individually
  - –Besides spatial redundancy, also exploit temporal redundancy
  - –Motion compensation





Encoding order: 1 4 2 3 7 5 6

Image Compression Standards, Formats, and Containers

- **Still Image**
  - **Binary**
    - CCITT Group 3
    - CCITT Group 4
    - JBIG (or JBIG1)
    - JBIG2
  - **Continuous Tone**
    - JPEG
    - JPEG-LS
    - JPEG-2000
- **Video**
  - DV
  - H.261
  - H.262
  - H.263
  - H.264
  - MPEG-1
  - MPEG-2
  - MPEG-4

JPEG

JPEG 2000

Image Processing in the Clinical Workflow

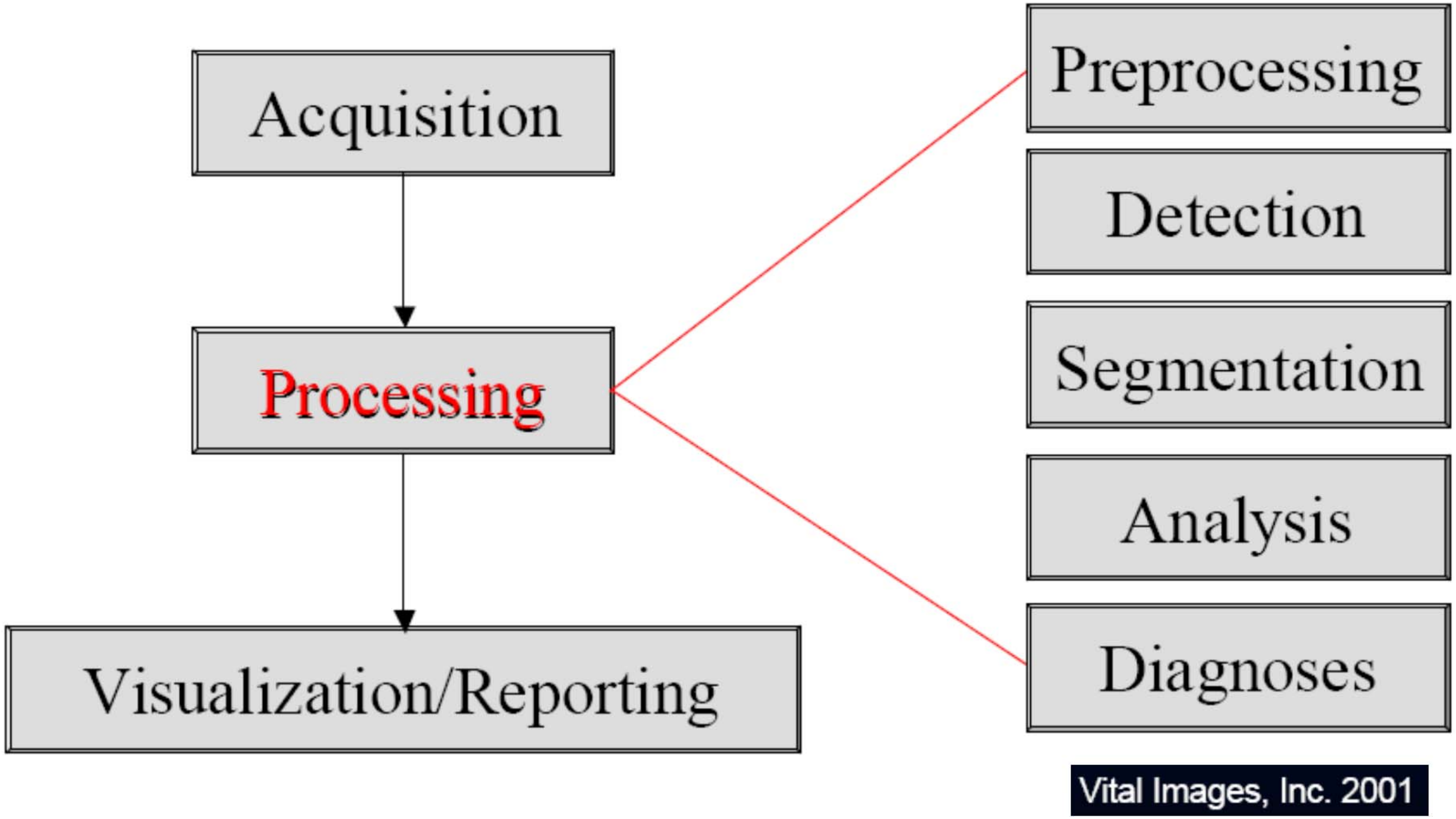
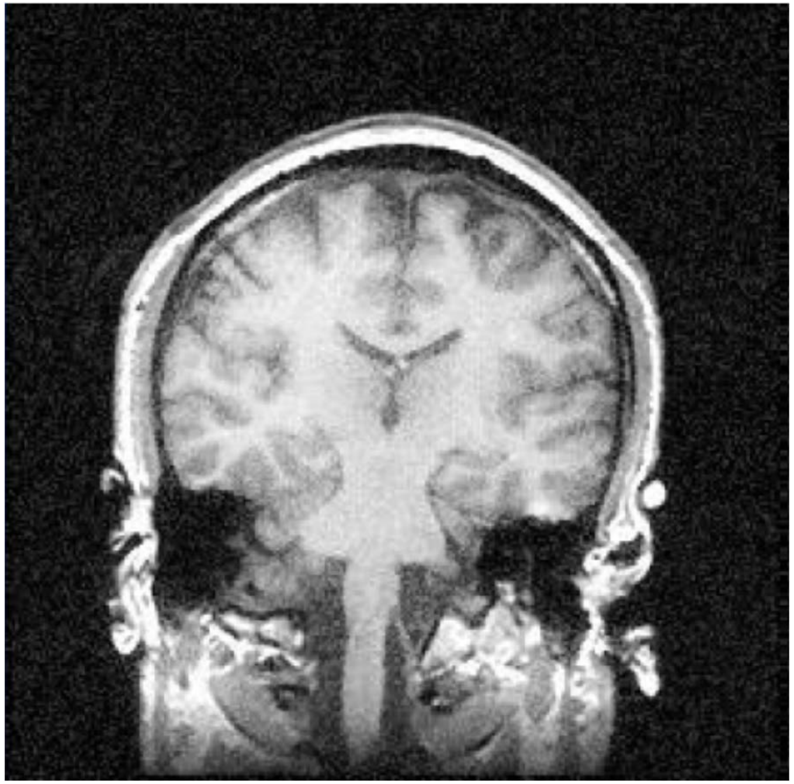


Image Processing

- Preprocessing
  - Denoising, inhomogeneity correction, etc.
- Image registration/alignment
- Detection
  - Finding objects of interest: organs, nodules, tumors, etc.
- Segmentation
  - Extract the exact boundaries (or masks) of anatomical structures
- Data analysis
  - Measurement: volume, shape, ejection fraction, perfusion parameters, etc.
- Classification/diagnosis

Preprocessing: Denoising



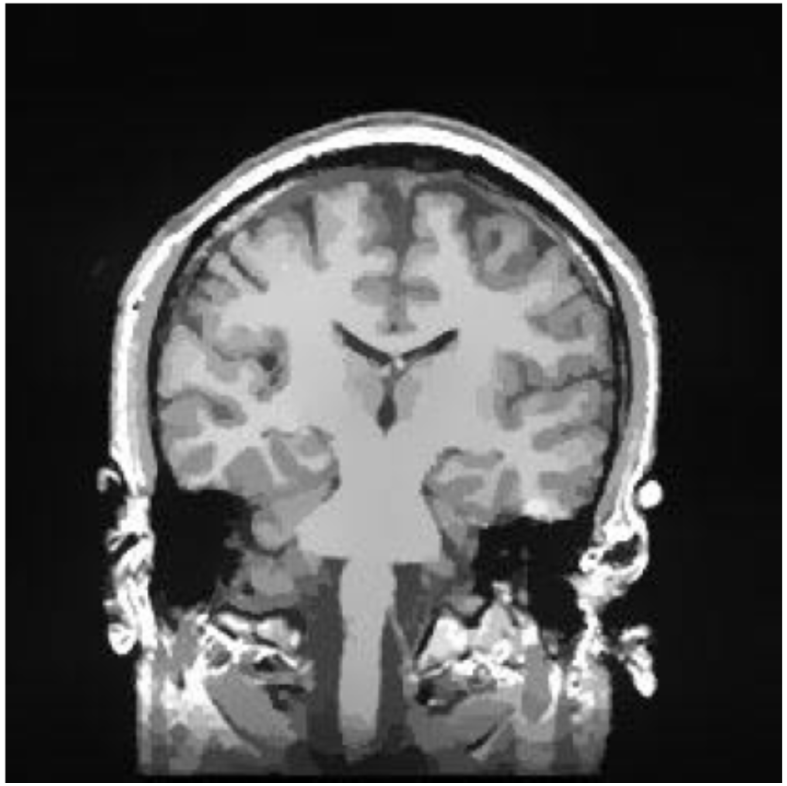
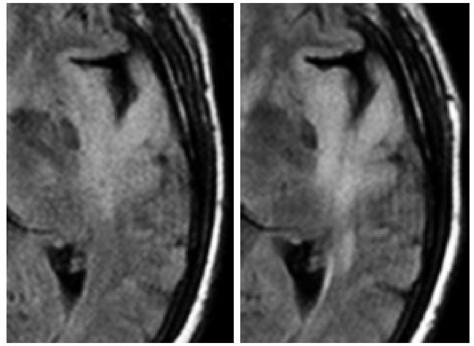


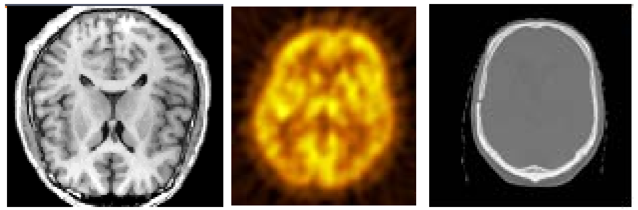
Image Registration

- Temporal evolution
  - ➢ Precise comparison of images of the same patient taken at different times.

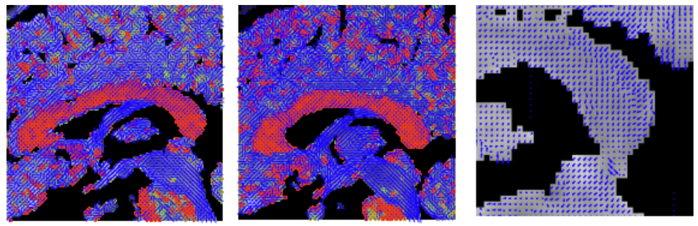


1st scan 2nd scan

- Fusion of multimodal images
  - ➢ Images of the same patient obtained with different modalities



- Inter-patients comparison
  - ➢ Images of different patients



a) b) c)

- Atlas superposition
  - ➢ Mapping patients images to an atlas

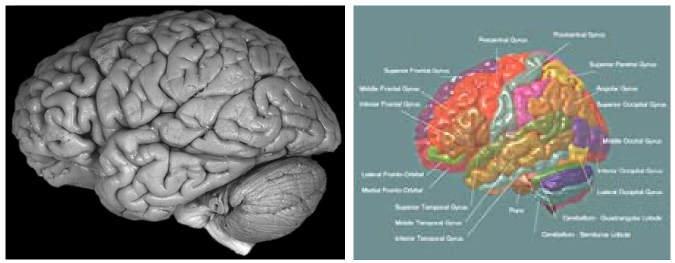
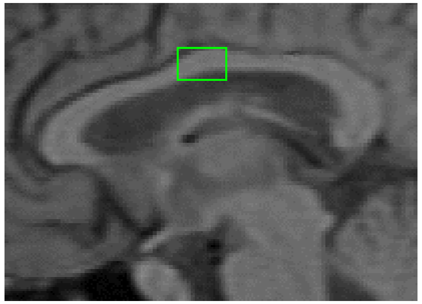
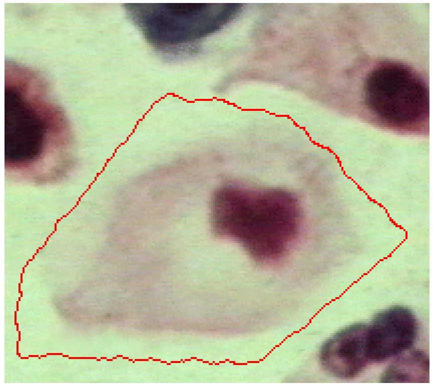


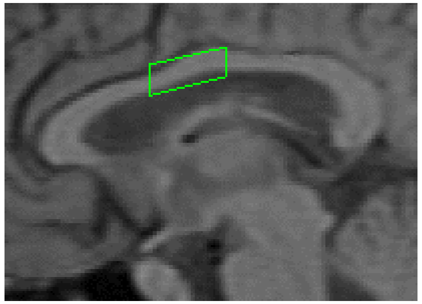
Image Segmentation

Geodesic snake

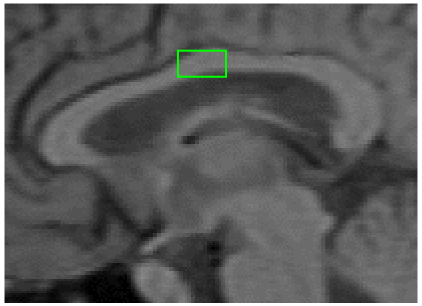


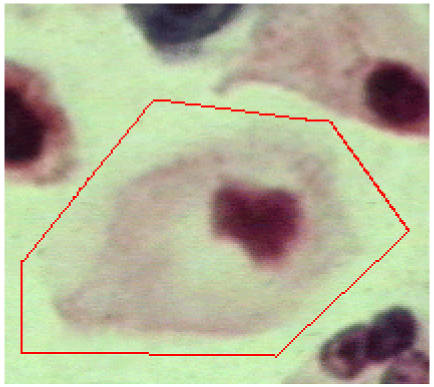


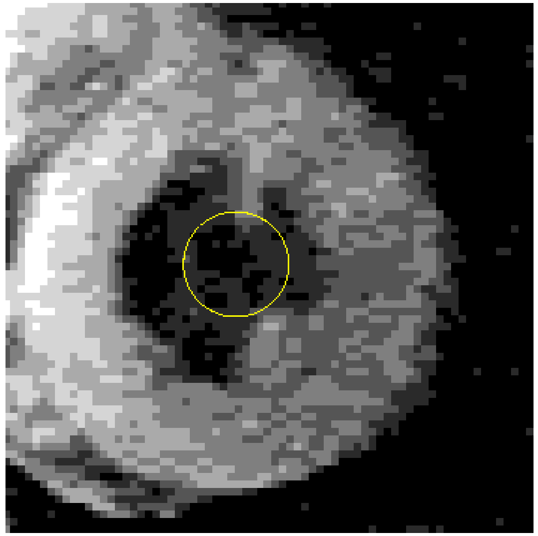
Generalized GVF



Region-aided Geometric Snake





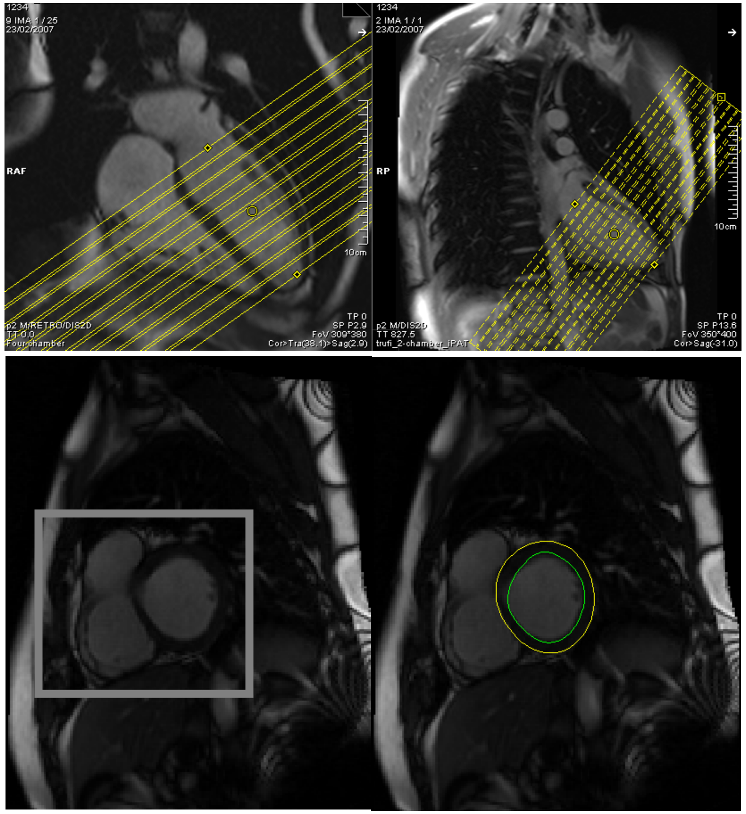


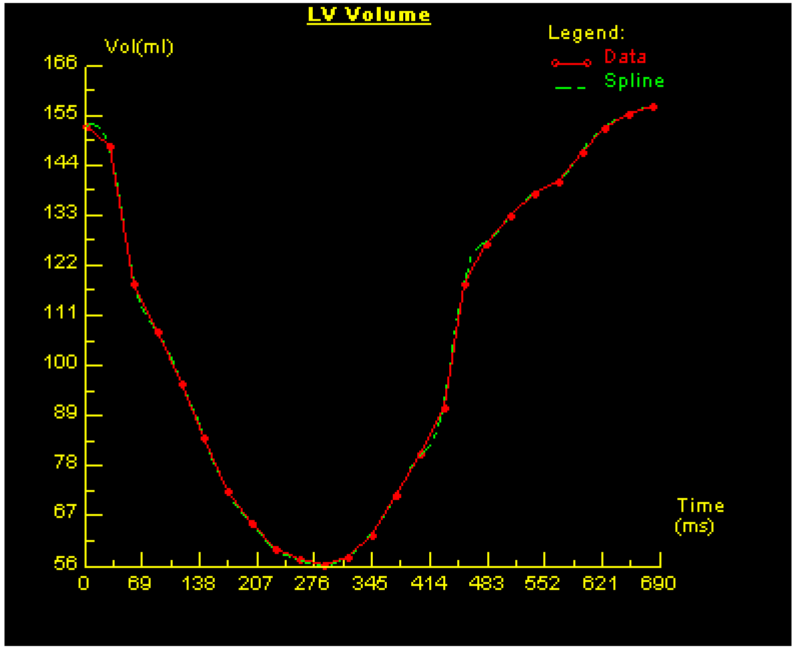
GVF snake

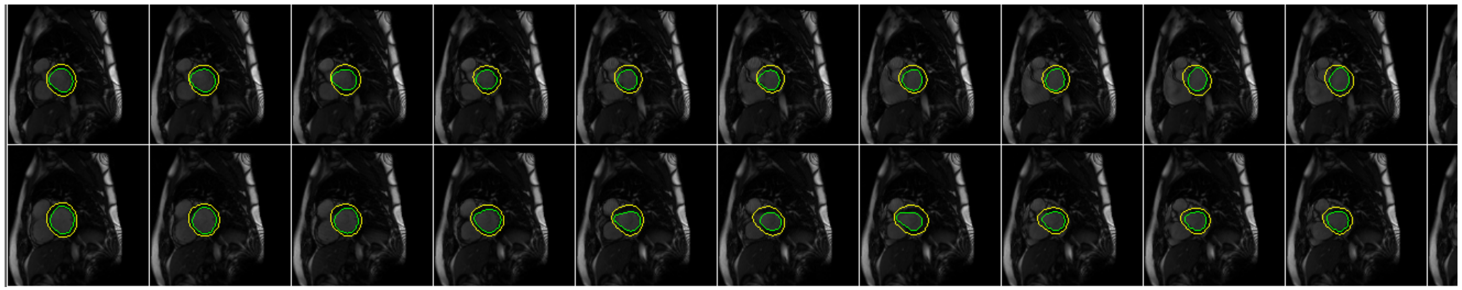
Sources:
http://iacl.ece.jhu.edu/projects/gvf/

http://www.cs.bris.ac.uk/staff/xie/snakes/

Fully Automatic Segmentation of the Left Ventricle in Cardiac Cine MRI







Audio-Visual Signal Processing