CS578- Speech Signal Processing
Lecture 2: Production and Classification of Speech Sounds

Yannis Stylianou

University of Crete, Computer Science Dept., Multimedia Informatics Lab
yannis@csd.uoc.gr

Univ. of Crete, 2008 Winter Period
OUTLINE

1. Anatomy and Physiology of Speech Production
   - Larynx
   - Vocal Tract
   - Categories of sound by source

2. Spectrographic analysis of Speech

3. Elements of Language

4. Prosody of Speech

5. Perception of Speech

6. Acknowledgments
A Simple View

Diagram showing the process of sound production in the vocal tract. The diagram includes the following components:

- **Power Supply**
- **Lungs**
- **Larynx**
- **Pharynx**
- **Modulator**
- **Oral Cavity**
- **Nasal Cavity**

The sound waves travel through the vocal tract and are modified by the various parts. The sound is then released into the air, creating different types of sound:

- **“sh”**
- **“o”**
- **“p”**

The diagram also includes graphs of periodic puffs, noise, and impulsive sounds, illustrating the different types of sound waves produced.
Cross sectional view
DOWNWARD-LOOKING INTO THE LARYNX: VOCAL FOLDS

**Left:** Voicing,  **Right:** Breathing

![Diagram](image-url)
Vocal Folds vibration
BERNOULLI’S PRINCIPLE IN THE GLOTTIS

Vocal Folds

Step 1

Step 2

Step 3

(a)

Looking Downward

Step 1

Step 2

Step 3

Step 4

Step 5

Step 6

(b)
Glottal airflow velocity

![Diagram showing glottal airflow velocity with phases labeled: Pitch Period, Closed Phase, Open Phase, Return Phase.](image)
SOFTER, TYPICAL, AND RELAXED GLOTTAL FLOW
Other vocal folds configurations

Left: Whispering,  Middle: Voicing  Right: Whispering voicing
Other forms of vibration

- **Creaky voice:**
  - vocal folds very tense
  - only a portion of them in oscillation
  - harsh-sounding voice
  - high and irregular pitch

- **Vocal fry**
  - folds are massy and relaxed
  - abnormally low and irregular pitch
  - secondary pulses during open phase

- **Diplophonia**
  - extra flaps
  - secondary pulses during the closed phase
Other forms of vibration

- **Creaky voice:**
  - vocal folds very tense
  - only a portion of them in oscillation
  - harsh-sounding voice
  - high and irregular pitch

- **Vocal fry**
  - folds are massy and relaxed
  - abnormally low and irregular pitch
  - secondary pulses during open phase

- **Diplophonia**
  - extra flaps
  - secondary pulses during the closed phase
Other forms of vibration

- Creaky voice:
  - vocal folds very tense
  - only a portion of them in oscillation
  - harsh-sounding voice
  - high and irregular pitch

- Vocal fry
  - folds are massy and relaxed
  - abnormally low and irregular pitch
  - secondary pulses during open phase

- Diplophonia
  - extra flaps
  - secondary pulses during the closed phase
**Examples**

**Upper panel:** vocal fry,  
**Lower panel:** diplophonia
By saying Vocal Tract we mean:

- Oral cavity: from the larynx to the lips, and the Nasal cavity
- Oral tract: 17cm for male voice, shorter for females
- Its purpose is to spectrally “color” the source and generate new sources for sound production
Vocat tract shapes
Spectral Shaping

Vocal tract is often approximated by a linear filter with:

- Formant frequencies
- Formant amplitude
- Formant bandwidth

Assuming a stable vocal tract and only with poles filter:

\[
H(z) = \frac{A}{\prod_{k=1}^{N_i} (1 - c_k z^{-1})(1 - c_k^* z^{-1})} = \sum_{k=1}^{N_i} \frac{A_k}{(1 - c_k z^{-1})(1 - c_k^* z^{-1})}
\]
Spectral Shaping

Vocal tract is often approximated by a linear filter with:
- Formant frequencies
- Formant amplitude
- Formant bandwidth

Assuming a stable vocal tract and only with poles filter:

\[
H(z) = \frac{A}{\prod_{k=1}^{N_i} (1 - c_k z^{-1})(1 - c_k^* z^{-1})} = \sum_{k=1}^{N_i} \frac{A_k}{(1 - c_k z^{-1})(1 - c_k^* z^{-1})}
\]
Let the excitation of vocal tract, $h[n]$, be:

$$u[n] = g[n] \ast p[n]$$

then, the output speech, $x[n, \tau]$, is given by:

$$x[n, \tau] = w[n, \tau]\{h[n] \ast (g[n] \ast p[n])\}$$

and

$$X(\omega, \tau) = \frac{1}{P} \sum_{k=-\infty}^{\infty} H(\omega_k) G(\omega_k) W(\omega - \omega_k, \tau)$$
HARMONICS AND FORMANTS
Ways to categorize speech sounds

- Vocal fold state:
  - Voiced
  - Unvoiced
- Oral tract state:
  - Plosives
  - Fricatives

Also: voiced and unvoiced plosives (/b/, /t/), voiced and unvoiced fricatives (/z/, /f/), whispered unvoiced
“Which tea party did baker go to?”
Outline

1 Anatomy and Physiology of Speech Production
   - Larynx
   - Vocal Tract
   - Categories of sound by source

2 Spectrographic analysis of Speech

3 Elements of Language

4 Prosody of Speech

5 Perception of Speech

6 Acknowledgments
Short Time Fourier Transform, STFT

STFT:

\[ X(\omega, \tau) = \sum_{n=-\infty}^{\infty} x[n, \tau] e^{-j\omega n} \]

where

\[ x[n, \tau] = w[n, \tau] x[n] \]

Spectrogram:

\[ S(\omega, \tau) = |X(\omega, \tau)|^2 \]
Short Time Fourier Transform, STFT

STFT:

\[ X(\omega, \tau) = \sum_{n=-\infty}^{\infty} x[n, \tau] e^{-j\omega n} \]

where

\[ x[n, \tau] = w[n, \tau] x[n] \]

Spectrogram:

\[ S(\omega, \tau) = |X(\omega, \tau)|^2 \]
NARROWBAND SPECTROGRAM

\[ x[n] \]

\[ w[n, \tau] \]

\[ x[n, \tau_1] = w[n, \tau_1] x[n] \]

\[ |X(\omega, \tau_1)| \]

Horizontal Striations
Wideband Spectrogram
SPECTROGRAM ON SPEECH

(a) Amplitude

(b) Frequency (Hz)

(c) Frequency (Hz)
SPECTROGRAM ON SPEECH; ANOTHER EXAMPLE
Do we know better now?

to classify sounds by looking in time of in frequency domain for
- periodic, noisy, impulsive sources?
- shape of vocal tract?
1. **Anatomy and Physiology of Speech Production**
   - Larynx
   - Vocal Tract
   - Categories of sound by source

2. **Spectrographic Analysis of Speech**

3. **Elements of Language**

4. **Prosody of Speech**

5. **Perception of Speech**

6. **Acknowledgments**
Phonemes’ map

- Phonemes
  - Vowels
    - Front: i (i), I (I), e (e), æ (@), e (E)
    - Center: æ (R), Λ (A)
    - Back: a (a), o (o), U (U), u (u)
  - Semi-Vowels
    - Liguids: r (r), l (l)
    - Glides: w (w), y (y)
  - Consonants
    - Affricates: tʃ (tS), dʒ (J)
    - Diphthongs: aɪ (Y), aʊ (W), ɔɪ (O), ju (JU)
    - Nasals: m (m), n (n), ŋ (G)
    - Fricatives
      - Voiced: v (v), ð (D), z (z), ð (Z)
      - Unvoiced: f (f), θ (T), s (s), ʃ (S)
    - Plosives
      - Voiced: b (b), d (d), g (g)
      - Unvoiced: p (p), t (t), k (k)
  - Whispers: h (h)
**Vowels**

- **Source:** Quasi-periodic puffs of airflow
- **System:** Each vowel phoneme corresponds to a different vocal tract configuration.
**Nasals**

- **Source:** Quasi-periodic puffs of airflow
- **System:** Air flows mainly through the nasal cavity and oral tract being constricted
NASALS: TIME AND SPECTROGRAM

(a) Amplitude

(b) Frequency (Hz)

Time (s)

Time (s)
Fricatives

- **Source:**
  - *Voiced:* vocal-folds vibrate
  - *Unvoiced:* vocal-folds are relaxed and not vibrating

- **System:** Oral tract being constricted by tongue at the back, center, or front of the oral tract, or at the teeth or lips
Fricatives’ profile

f (for)  T (thin)  s (see)

S (she)  h (he)  v (vote)

D (then)  z (zoo)  Z (azure)
Fricatives: Time and Spectrogram

(a) Amplitude vs. Time (s)
   - Voice Bar

(b) Frequency (Hz) vs. Time (s)

(c) Amplitude (dB) vs. Frequency (Hz)
   - Harmonics
Voiced:

- **Source:** vocal folds are vibrating ("voice bar")
- **System:** Oral tract being constricted by tongue at the back, center, or front of the oral tract, or at the teeth or lips

Unvoiced:

- **Source:** vocal folds are not vibrating
- **System:** Oral tract being constricted by tongue at the back, center, or front of the oral tract, or at the teeth or lips
Plosives’ profile

p (pay)  t (to)  k (key)

b (be)  d (day)  g (go)
Voice Onset Time

Unvoiced Plosive

(a)

Silence

Burst

Aspiration

Voicing

VOT

Voiced Plosive

(b)

Voice Bar

Burst

Voicing

VOT
Plosives: time and spectrogram

(a)

(b)

Amplitude

Frequency (Hz)

Time (s)

Amplitude (dB)

Frequency (Hz)

Burst

Voice Bar

Aspiration

Harmonics
Semi-vowels

w (we)  y (you)  r (read)  l (left)
TRANSPORTATIONAL SPEECH SOUNDS: “BOY”
1. **Anatomy and Physiology of Speech Production**
   - Larynx
   - Vocal Tract
   - Categories of sound by source

2. **Spectrographic Analysis of Speech**

3. **Elements of Language**

4. **Prosody of Speech**

5. **Perception of Speech**

6. **Acknowledgments**
Prosody of speech

As prosody of speech we refer to:
- Rhythm
- Fundamental frequency contour (pitch)
- Loudness
Stressed speech

“Please do this today”:
1. Anatomy and Physiology of Speech Production
   - Larynx
   - Vocal Tract
   - Categories of sound by source

2. Spectrographic Analysis of Speech

3. Elements of Language

4. Prosody of Speech

5. Perception of Speech

6. Acknowledgments
Outline

1. Anatomy and Physiology of Speech Production
   - Larynx
   - Vocal Tract
   - Categories of sound by source

2. Spectrographic Analysis of Speech

3. Elements of Language

4. Prosody of Speech

5. Perception of Speech

6. Acknowledgments
Most, if not all, figures in this lecture are coming from the book:

**T. F. Quatieri:** Discrete-Time Speech Signal Processing, principles and practice
2002, Prentice Hall

and have been used after permission from Prentice Hall