CS578- Speech Signal Processing

Lecture on Voice Function Assessment

Yannis Stylianou

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

Univ. of Crete

OUTLINE

Voice Function Assessment

Use of:

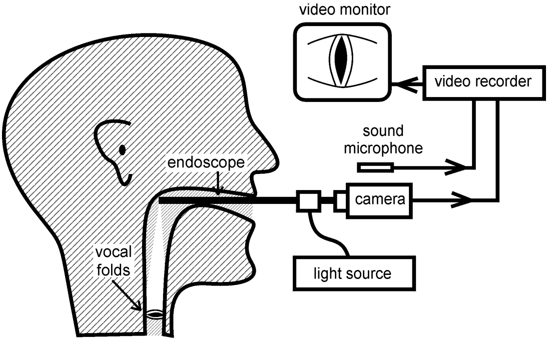
- Voice Production Models
- Algorithms of Signal/Speech Processing
- Special devices (i.e., high speed camera, EGG etc)

for assessing voice disorders and quality of voice in general.

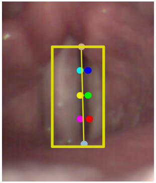
List of devices

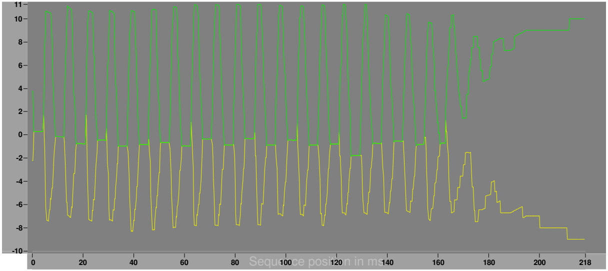
- Endoscopes
- High-speed cameras
- Videokymograps
- Electroglottographs
- Accelerator probes
- Pneumotachographs
- Contact microphones
- Microphones
- Cameras (possibly)

Invasive: Using cameras ...



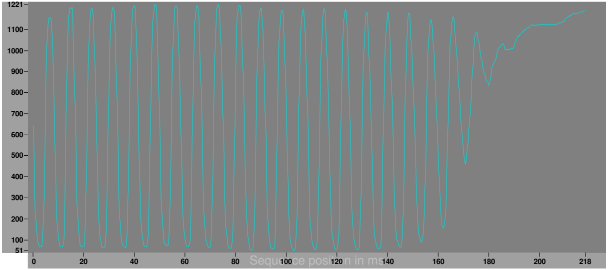
High-Speed (*invasive*)



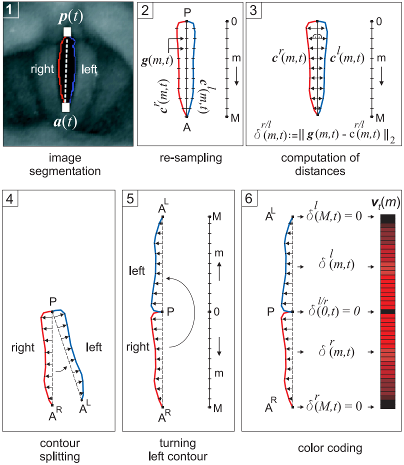


High-Speed (*invasive*)

- Glottal surface

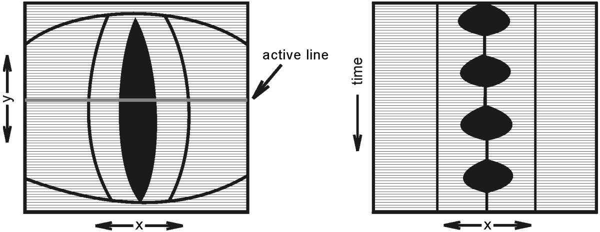


PHONOVIBROGRAPHY - PVG[1]



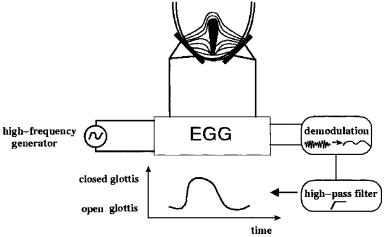
Videokymography[2]

- Principle



Non-invasive

- EGG [3] ▶ Principle



- Contact microphone signals

- Speech

Outline

Measuring Jitter (1/2)

- Local jitter is the period-to-period variability of pitch (%)

$$\frac{\frac{1}{N-1}\sum_{n=1}^{N-1}|u(n+1)-u(n)|}{\frac{1}{N}\sum_{n=1}^{N}u(n)}$$

- Absolute jitter is the period-to-period variability of pitch in time

$$\frac{1}{N-1}\sum_{n=1}^{N-1}|u(n+1)-u(n)|$$

Measuring Jitter (2/2)

- Relative Average Perturbation (RAP): 3 periods (%)

$$\frac{\frac{1}{N-2}\sum_{n=1}^{N-2}\frac{|2u(n+1)-u(n)-u(n+2)|}{3}}{\frac{1}{N}\sum_{n=1}^{N}u(n)}$$

- Pitch Period Perturbation Quotient (PPQ): 5 periods (%)

$$\frac{\frac{1}{N-4}\sum_{n=1}^{N-4}\frac{|4u(n+2)-u(n)-u(n+1)-u(n+3)-u(n+4)|}{5}}{\frac{1}{N}\sum_{n=1}^{N}u(n)}$$

Spectral Jitter Model

Jittered impulse train:

$$g[n] = \sum_{k=-\infty}^{+\infty} \delta[n-(2k)P] + \sum_{k=-\infty}^{+\infty} \delta[n+\epsilon-(2k+1)P]$$

and the corresponding magnitude spectrum:

$$|G(\omega)|^2 = \frac{\omega_0^2}{2} \sum_{k=-\infty}^{+\infty} \left(1+\cos\left[(P-\epsilon)\,k\frac{\omega_0}{2}\right]\right)\delta\left(\omega-k\frac{\omega_0}{2}\right)$$

Spectral Jitter Model

Jittered impulse train:

$$g[n] = \sum_{k=-\infty}^{+\infty} \delta[n - (2k)P] + \sum_{k=-\infty}^{+\infty} \delta[n + \epsilon - (2k+1)P]$$

and the corresponding magnitude spectrum:

$$|G(\omega)|^2 = \frac{\omega_0^2}{2} \sum_{k=-\infty}^{+\infty} \left(1 + \cos\left[(P - \epsilon)\, k \frac{\omega_0}{2}\right]\right) \delta\left(\omega - k\frac{\omega_0}{2}\right)$$

Beat Spectrum

This representation leads to a beat spectrum:

$$1 + \cos\left[(P-\epsilon)\,k\frac{\omega_0}{2}\right] = 1 + \cos\left(k\pi\right)\cos\left(k\frac{\epsilon}{P}\pi\right)$$

with intersections at:

$$\omega_k = \left(k + \frac{1}{2}\right)\frac{\pi}{\epsilon}$$

Log Magnitude spectrum

$$20\log_{10}|G(\omega)| =$$

$$10\log_{10}\left(\frac{\omega_0^2}{2}\left(1+\cos\left[\left(P-\epsilon\right)\omega\right]\right)\right)$$

$$\left[\sum_{l=-\infty}^{+\infty}\delta\left(\omega - l\omega_0\right) + \sum_{l=-\infty}^{+\infty}\delta\left(\omega - (l+\frac{1}{2})\omega_0\right)\right]$$

Harmonic and Subharmonic spectra

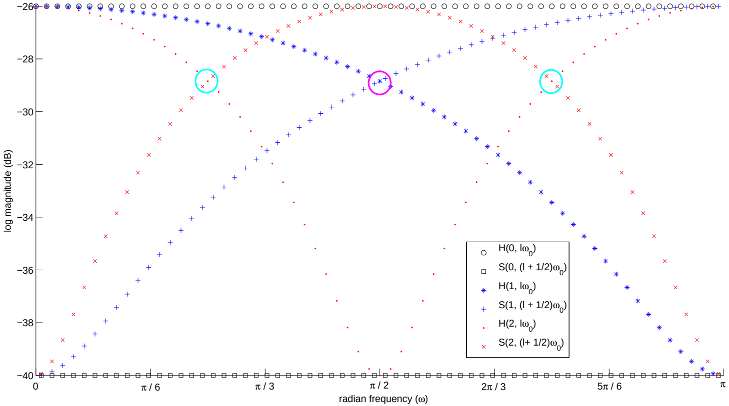
The magnitude spectrum can be split into a Harmonic spectrum:

$$H(\epsilon, l\omega_0) = 10\log_{10}\left(\frac{\omega_0^2}{2}\left(1+\cos\left[(P-\epsilon)\, l\omega_0\right]\right)\right)$$

and a Subharmonic spectrum:

$$S(\epsilon, (l+\frac{1}{2})\omega_0) = 10\log_{10}\left(\frac{\omega_0^2}{2}\left(1+\cos\left[(P-\epsilon)\,(l+\frac{1}{2})\omega_0\right]\right)\right)$$

Harmonic-Subharmonic examples



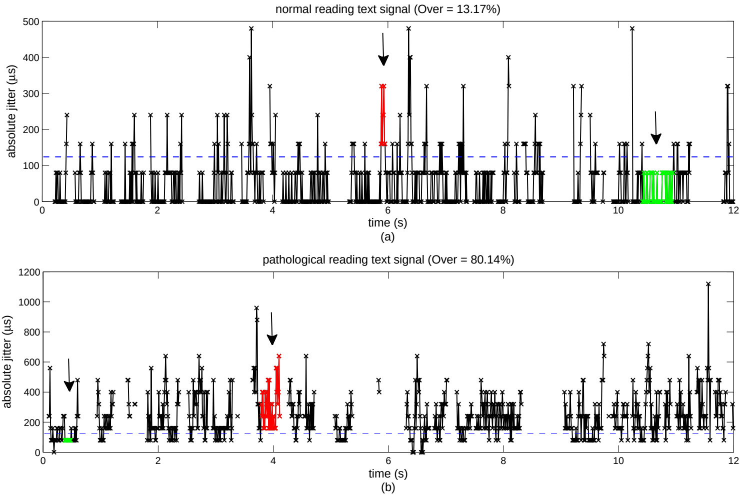
Voice Pathology Detection

AUC (standard error) %		
	MEEI	PdA
MDVP	90.66 (1.42)	70.65 (2.50)
Praat	90.47 (1.44)	62.94 (2.67)
SJE with enhancement	94.82 (0.92)	84.65 (1.92)

Running Speech Database

- MEEI - Rainbow passage: $\text{MEEI}_{\text{Rainbow}}$
- Recordings are limited to 12 seconds (2 first sentences)
- 53 signals from healthy voices and 660 signals from pathological voices
- Sampling frequency: 25 kHz (683 signals), 10 kHz (30 signals), 16 bits
- Onset and offset effects in the voiced areas (2 frames)
- Pitch period measured at a rate of 10ms
- SJE used a window of 4 times the local pitch period

Features


normal reading text signal (Over = 13.17%)
absolute jitter (µs)
0
100
200
300
400
500
0
2
4
6
8
10
12
time (s)
(a)
pathological reading text signal (Over = 80.14%)
absolute jitter (µs)
0
200
400
600
800
1000
1200
0
2
4
6
8
10
12
time (s)
(b)

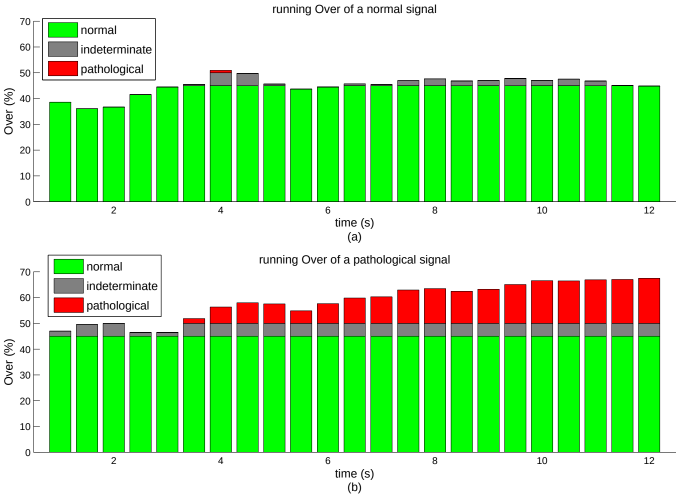
Discrimination Power

AUC (standard error) % ($MEEI_{Rainbow}$, Thr_{SJE})		
Over	Max Over	Max Under
95.69 (0.80)	93.32 (1.10)	91.61 (1.30)

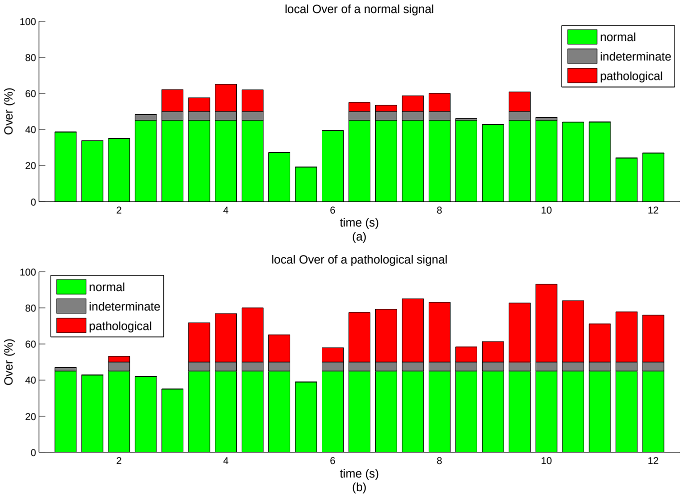
RUNNING AND LOCAL: *Over*

- **Local:**
  Feature is computed using a sliding analysis window of fixed size
- **Running:**
  Feature is computed using gradually extended analysis windows

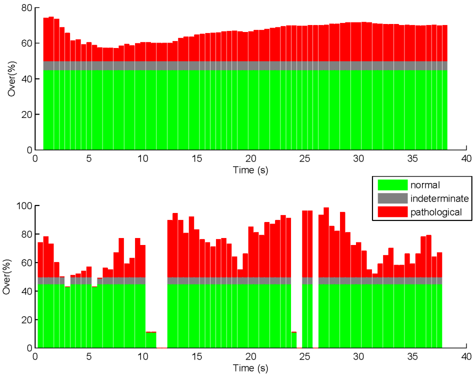
Example from MEEI: Running *Over*



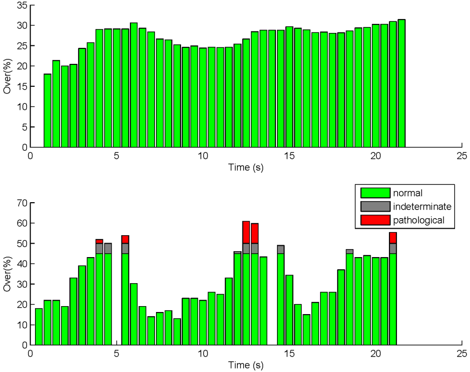
Example from MEEI: Local *Over*



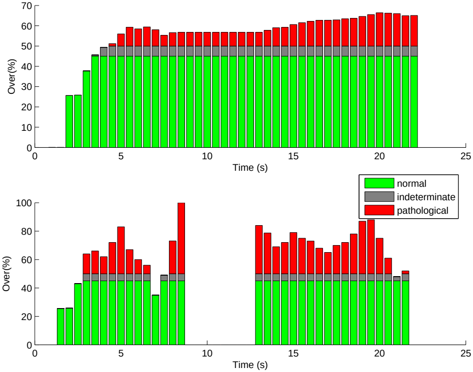
Before treatment



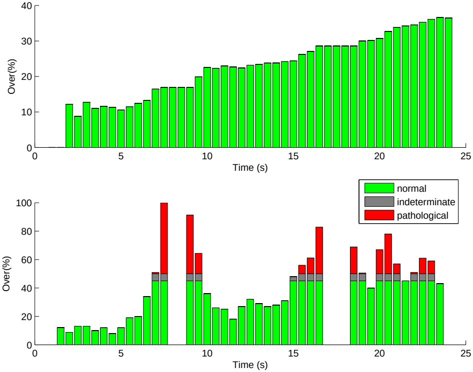
After treatment



Before Treatment



After Treatment



References for the work on SJE

1. Miltos Vassilakis and Yannis Stylianou:
   Spectral jitter modeling and estimation.
   Elsevier Biomedical Signal Processing and Control. Special Issue: M&A of Vocal Emissions, 2009, DOI: 10.1016/j.bspc.2009.02.001
2. Miltos Vassilakis and Yannis Stylianou:
   Voice Pathology Detection based on Short-Time Jitter Estimations in Running Speech.
   Folia Phoniatrica et Logopaedica, Vol.61, No.3, pp 153-170, 2009 DOI: 10.1159/000219951

Outline

Sinusoidal model

$$s(t) = \sum_{k=1}^{K(t)} A_k(t) cos[\theta_k(t)]$$

where

$$A_k(t) = \underbrace{\mu_k(t)}_{\text{excitation}} \cdot \underbrace{M_k(t)}_{\text{vocal track}}$$

and

$$\theta_k(t) = \underbrace{\phi_k(t)}_{\text{excitation}} + \underbrace{\Phi_k(t)}_{\text{vocal track}}$$

$$\phi_k(t) = 2\pi k \int_0^t f_0(\tau) d\tau + \phi_k$$

Jitter and Shimmer

Jitter:

$$f_0(t) = f_0 - \delta \sin(\pi f_0 t + \psi_k)$$

Shimmer:

$$\mu_k(t) = \mu_k[1 + \gamma_k \cos(\pi f_0 t + \chi_k)]$$

so then:

$$s(t) = \sum_{k=-L}^{L} A_k[1 + \gamma_k \cos(\pi f_0 t + \chi_k)]$$
$$e^{j[2\pi k f_0 t + \delta_k \cos(\pi f_0 t + \psi_k) + \theta_k]} w(t)$$

$$s(t) \approx \sum_{k=-L}^{L} A_k e^{j\theta_k} e^{j2\pi k f_0 t} w(t)$$
$$[1 + (\gamma_k \cos(\chi_k) \cos(\pi f_0 t) - \gamma_k \sin(\chi_k) \sin(\pi f_0 t))$$
$$+ j(\delta_k \cos(\psi_k) \cos(\pi f_0 t) - \delta_k \sin(\psi_k) \sin(\pi f_0 t))]$$

Jitter and Shimmer

Jitter:

$$f_0(t) = f_0 - \delta \sin(\pi f_0 t + \psi_k)$$

Shimmer:

$$\mu_k(t) = \mu_k[1 + \gamma_k \cos(\pi f_0 t + \chi_k)]$$

so then:

$$s(t) = \sum_{k=-L}^{L} A_k[1 + \gamma_k \cos(\pi f_0 t + \chi_k)]$$
$$e^{j[2\pi k f_0 t + \delta_k \cos(\pi f_0 t + \psi_k) + \theta_k]} w(t)$$

$$s(t) \approx \sum_{k=-L}^{L} A_k e^{j\theta_k} e^{j2\pi k f_0 t} w(t)$$
$$[1 + (\gamma_k \cos(\chi_k)\cos(\pi f_0 t) - \gamma_k \sin(\chi_k)\sin(\pi f_0 t))$$
$$+ j(\delta_k \cos(\psi_k)\cos(\pi f_0 t) - \delta_k \sin(\psi_k)\sin(\pi f_0 t))]$$

Jitter and Shimmer

Jitter:

$$f_0(t) = f_0 - \delta \sin(\pi f_0 t + \psi_k)$$

Shimmer:

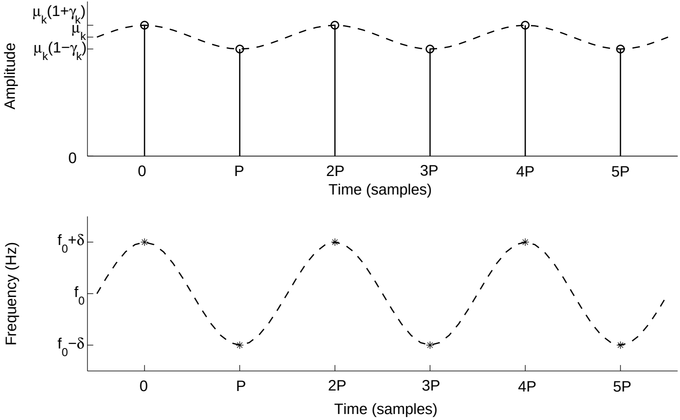
$$\mu_k(t) = \mu_k[1 + \gamma_k \cos(\pi f_0 t + \chi_k)]$$

so then:

$$s(t) = \sum_{k=-L}^{L} A_k[1 + \gamma_k \cos(\pi f_0 t + \chi_k)] \\ e^{j[2\pi k f_0 t + \delta_k \cos(\pi f_0 t + \psi_k) + \theta_k]} w(t)$$

$$s(t) \approx \sum_{k=-L}^{L} A_k e^{j\theta_k} e^{j2\pi k f_0 t} w(t) \\ [1 + (\gamma_k \cos(\chi_k) \cos(\pi f_0 t) - \gamma_k \sin(\chi_k) \sin(\pi f_0 t)) \\ + j(\delta_k \cos(\psi_k) \cos(\pi f_0 t) - \delta_k \sin(\psi_k) \sin(\pi f_0 t))]$$

Models of shimmer and jitter



Shimmer in the upper panel, Jitter in the lower panel.

Parameters estimation

Suggesting:

$$x(t) = \sum_{k=-L}^{L} [a_k + b_k \sin(\pi f_0 t) + c_k \cos(\pi f_0 t)] e^{j2\pi k f_0 t} w(t)$$

and by letting

$$\begin{aligned} b_k &= \rho_{1,k} a_k + \rho_{2,k} j a_k \\ c_k &= \pi_{1,k} a_k + \pi_{2,k} j a_k \end{aligned}$$

then:

$$x(t) = \sum_{k=-L}^{L} a_k e^{j2\pi k f_0 t} w(t)$$

$$[1 + (\pi_{1,k} \cos(\pi f_0 t) + \rho_{1,k} \sin(\pi f_0 t)) + j(\pi_{2,k} \cos(\pi f_0 t) + \rho_{2,k} \sin(\pi f_0 t))]$$

Parameters estimation

Suggesting:

$$x(t) = \sum_{k=-L}^{L} [a_k + b_k \sin(\pi f_0 t) + c_k \cos(\pi f_0 t)] e^{j2\pi k f_0 t} w(t)$$

and by letting

$$\begin{aligned} b_k &= \rho_{1,k} a_k + \rho_{2,k} j a_k \\ c_k &= \pi_{1,k} a_k + \pi_{2,k} j a_k \end{aligned}$$

then:

$$\begin{aligned} x(t) = \sum_{k=-L}^{L} a_k e^{j2\pi k f_0 t} w(t) \\ [1 + (\pi_{1,k} \cos(\pi f_0 t) + \rho_{1,k} \sin(\pi f_0 t)) \\ + j(\pi_{2,k} \cos(\pi f_0 t) + \rho_{2,k} \sin(\pi f_0 t))] \end{aligned}$$

Comparing models

What we want, was:

$$s(t) \approx \sum_{k=-L}^{L} A_k e^{j\theta_k} e^{j2\pi k f_0 t} w(t)$$
$$[1 + (\gamma_k \cos(\chi_k)\cos(\pi f_0 t) - \gamma_k \sin(\chi_k)\sin(\pi f_0 t))$$
$$+ j(\delta_k \cos(\psi_k)\cos(\pi f_0 t) - \delta_k \sin(\psi_k)\sin(\pi f_0 t))]$$

and what we suggest:

$$x(t) = \sum_{k=-L}^{L} a_k e^{j2\pi k f_0 t} w(t)$$
$$[1 + (\pi_{1,k} \cos(\pi f_0 t) + \rho_{1,k} \sin(\pi f_0 t))$$
$$+ j(\pi_{2,k} \cos(\pi f_0 t) + \rho_{2,k} \sin(\pi f_0 t))]$$

Comparing models

What we want, was:

$$s(t) \approx \sum_{k=-L}^{L} A_k e^{j\theta_k} e^{j2\pi k f_0 t} w(t)$$
$$[1 + (\gamma_k \cos(\chi_k) \cos(\pi f_0 t) - \gamma_k \sin(\chi_k) \sin(\pi f_0 t))$$
$$+ j(\delta_k \cos(\psi_k) \cos(\pi f_0 t) - \delta_k \sin(\psi_k) \sin(\pi f_0 t))]$$

and what we suggest:

$$x(t) = \sum_{k=-L}^{L} a_k e^{j2\pi k f_0 t} w(t)$$
$$[1 + (\pi_{1,k} \cos(\pi f_0 t) + \rho_{1,k} \sin(\pi f_0 t))$$
$$+ j(\pi_{2,k} \cos(\pi f_0 t) + \rho_{2,k} \sin(\pi f_0 t))]$$

Final ...

$$
\begin{aligned}
\pi_{1,k} &= \hat{\gamma}_k \cos(\chi_k) \\
\rho_{1,k} &= -\hat{\gamma}_k \sin(\chi_k) \\
\pi_{2,k} &= \hat{\delta}_k \cos(\psi_k) \\
\rho_{2,k} &= -\hat{\delta}_k \sin(\psi_k)
\end{aligned}
$$

which leads to the final solution for the estimates

$$
\begin{aligned}
\hat{\gamma}_k &= \sqrt{\pi_{1,k}^2 + \rho_{1,k}^2} \\
\hat{\delta}_k &= \sqrt{\pi_{2,k}^2 + \rho_{2,k}^2}
\end{aligned}
$$

Validation

- Glottal airflow rate

$$r[n] = A_1 \sum_{k=-\infty}^{+\infty} \delta[n-(2k)P] + A_2 \sum_{k=-\infty}^{+\infty} \delta[n+\epsilon-(2k+1)P]$$

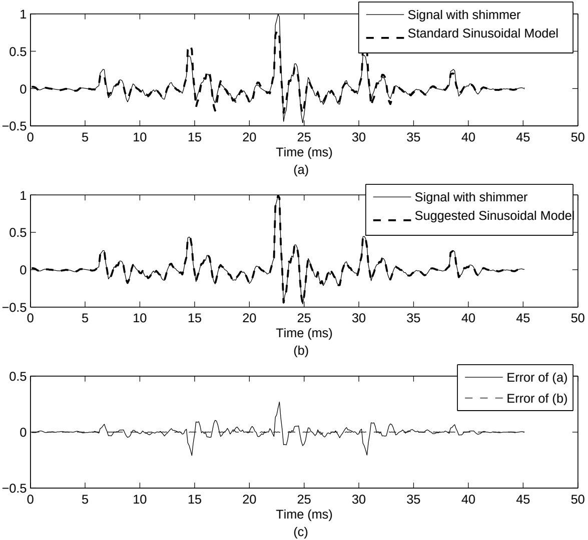
  where $A_1 = A_0 + A$ and $A_2 = A_0 - A$, with A to control the shimmer value while ϵ controls the jitter.

- Glottal airflow frequency response (maximum phase)

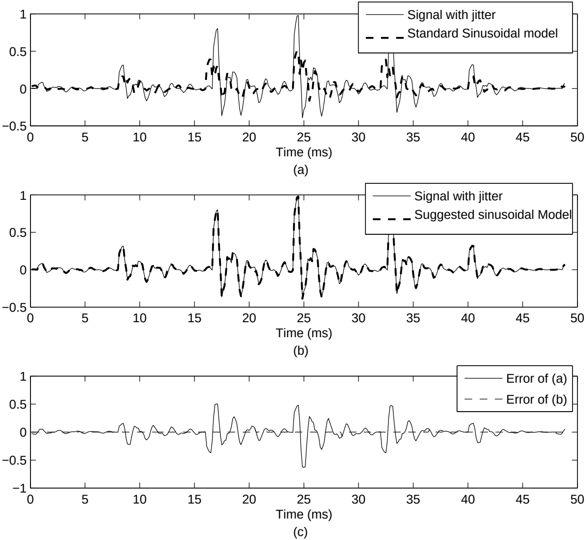
$$G(z) = \frac{1}{(1-bz)^2}$$

- Vocal tract, as an AR filter (using phoneme /a/ from real speech - male speaker)

Validation: Modeling Shimmer

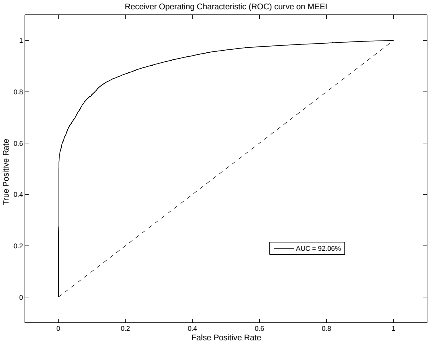


Validation: Modeling Jitter



Evaluation: Voice Pathology Detection (MEEI)

AUC: 92.06% ($L = 20$, FLD, 5000 Monte Carlo repetitions)



Quasi-Harmonic Model, QHM

- The model:

$$s(t) = \left(\sum_{k=-K}^{K} (a_k + tb_k) e^{2\pi j f_k t} \right) w(t)$$

- Instantaneous parameters:

$$\begin{aligned} m_k(t) &= \sqrt{(a_k^R + tb_k^R)^2 + (a_k^I + t' b_k^I)^2} \\ \phi_k(t) &= 2\pi f_k t + atan \frac{a_k^I + t' b_k^I}{a_k^R + tb_k^R} \\ f_k(t) &= f_k + \frac{1}{2\pi} \frac{a_k^R b_k^I - a_k^I b_k^R}{m_k^2(t)} \end{aligned}$$

Quasi-Harmonic Model, QHM

- The model:

$$s(t) = \left(\sum_{k=-K}^{K} (a_k + t b_k) e^{2\pi j f_k t} \right) w(t)$$

- Instantaneous parameters:

$$\begin{aligned} m_k(t) &= \sqrt{(a_k^R + t b_k^R)^2 + (a_k^I + t' b_k^I)^2} \\ \phi_k(t) &= 2\pi f_k t + atan \frac{a_k^I + t' b_k^I}{a_k^R + t b_k^R} \\ f_k(t) &= f_k + \frac{1}{2\pi} \frac{a_k^R b_k^I - a_k^I b_k^R}{m_k^2(t)} \end{aligned}$$

ITERATIVE QHM

Iteratively improve estimation of QHM parameters:

Let's assume that we know at $n = 0$, $f_k(0)$:

1. Compute the $a_{k,n}$ and $b_{k,n}$ through Least Squares using $f_k(n-1)$.
2. Compute instantaneous components, for $k = 1 \ldots K$:
$$\begin{cases} A_k(n) = m_k(0) \\ \Phi_k(n) = \phi_k(0) \\ F_k(n) = f_k(0) \end{cases}$$
3. Move to the next time instant: $n = n + 1$ and go to step 1.

Iterative QHM

Iteratively improve estimation of QHM parameters:
Let's assume that we know at $n = 0$, $f_k(0)$:

1. Compute the $a_{k,n}$ and $b_{k,n}$ through Least Squares using $f_k(n-1)$.
2. Compute instantaneous components, for $k = 1 \dots K$:
$$\begin{cases} A_k(n) = m_k(0) \\ \Phi_k(n) = \phi_k(0) \\ F_k(n) = f_k(0) \end{cases}$$
3. Move to the next time instant: $n = n + 1$ and go to step 1.

Iterative QHM

Iteratively improve estimation of QHM parameters:
Let's assume that we know at $n = 0$, $f_k(0)$:

1. Compute the $a_{k,n}$ and $b_{k,n}$ through Least Squares using $f_k(n-1)$.
2. Compute instantaneous components, for $k = 1 \ldots K$:
$$\begin{cases} A_k(n) = m_k(0) \\ \Phi_k(n) = \phi_k(0) \\ F_k(n) = f_k(0) \end{cases}$$
3. Move to the next time instant: $n = n + 1$ and go to step 1.

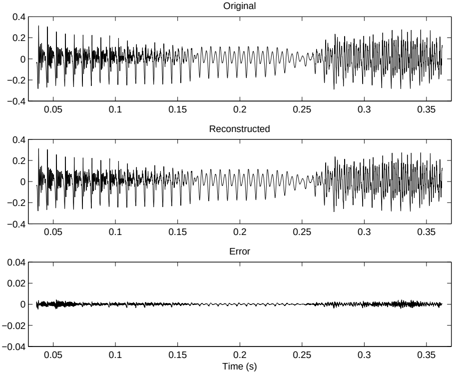
Iterative QHM

Iteratively improve estimation of QHM parameters:
Let's assume that we know at $n = 0$, $f_k(0)$:

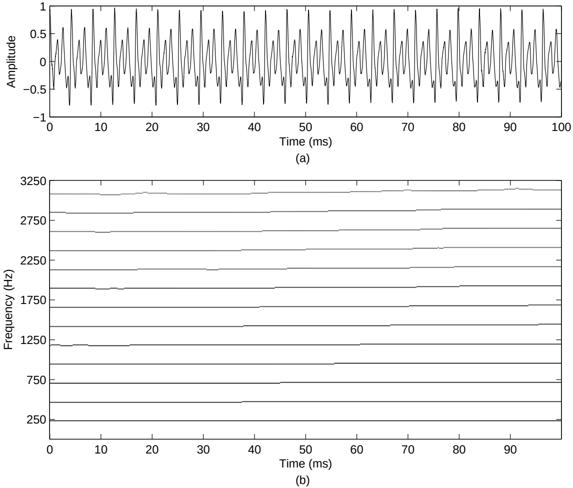
1. Compute the $a_{k,n}$ and $b_{k,n}$ through Least Squares using $f_k(n-1)$.
2. Compute instantaneous components, for $k = 1 \dots K$:
$$\begin{cases} A_k(n) = m_k(0) \\ \Phi_k(n) = \phi_k(0) \\ F_k(n) = f_k(0) \end{cases}$$
3. Move to the next time instant: $n = n + 1$ and go to step 1.

Validation: Signal reconstruction

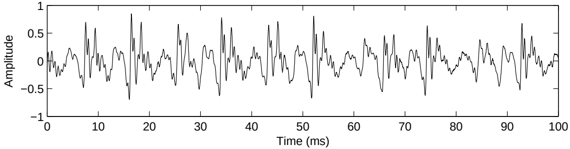
$$s[n] = \sum_{k=1}^{L} A_k[n] \cos\left(\Phi_k[n]\right)$$



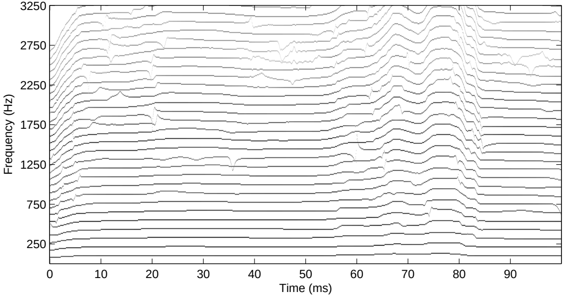
Time-Frequency Example: Normophonic case



Time-Frequency Example: Dysphonic case



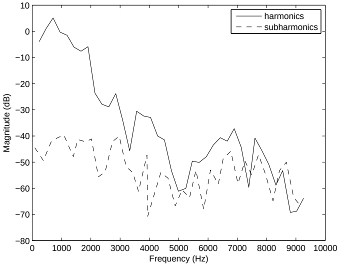
(a)

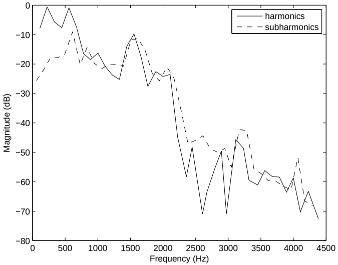


(b)

Harmonics and Sub-harmonics: Frequency

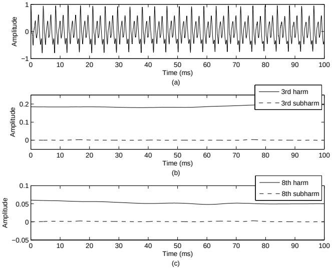
Left: Normophonic speaker, Right: Dysphonic speaker

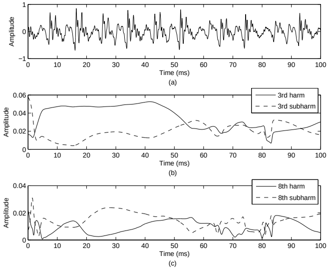




Harmonics and Sub-harmonics: Time

Left: Normophonic speaker, Right: Dysphonic speaker





References for the work on Sinusoidal Model

1. Yannis Pantazis, Olivier Rosec, and Yannis Stylianou:
Adaptive AM-FM Signal Decomposition with Application to Speech Analysis
IEEE Trans. on Audio, Speech and Language Processing, Vol.19, No.2, Feb 2011, pp 290-300.
2. Yannis Pantazis, Miltiadis Vasilakis, and Yannis Stylianou:
Jitter and Shimmer modeling based on a Time-Varying Sinusoidal Model of Speech
3rd Advanced Voice Function Assessment International Workshop, pp. 33-36, May 2009, Madrid

Outline

Define Vocal Tremor

- Vocal Tremor: Involuntary modulations of frequency and/or amplitude in sustained phonation.
- Pathological & Physiological Vocal Tremor.
  - Pathological Tremor: From diseases like Parkinson, essential tremor, etc. $\Rightarrow$ Strong motor synchronization.
  - Physiological Tremor: Natural stochastic modulations in the interval $[2, 15]Hz$ with low amplitude.
- Acoustic Vocal Tremor Attributes:
  - Modulation Frequency: How fast are the modulations.
  - Modulation Level: How strong are the modulations.

Define Vocal Tremor

- Vocal Tremor: Involuntary modulations of frequency and/or amplitude in sustained phonation.
- Pathological & Physiological Vocal Tremor.
  - Pathological Tremor: From diseases like Parkinson, essential tremor, etc. $\Rightarrow$ Strong motor synchronization.
  - Physiological Tremor: Natural stochastic modulations in the interval $[2, 15]Hz$ with low amplitude.
- Acoustic Vocal Tremor Attributes:
  - Modulation Frequency: How fast are the modulations.
  - Modulation Level: How strong are the modulations.

Define Vocal Tremor

- Vocal Tremor: Involuntary modulations of frequency and/or amplitude in sustained phonation.
- Pathological & Physiological Vocal Tremor.
  - Pathological Tremor: From diseases like Parkinson, essential tremor, etc. $\Rightarrow$ Strong motor synchronization.
  - Physiological Tremor: Natural stochastic modulations in the interval $[2, 15]Hz$ with low amplitude.
- Acoustic Vocal Tremor Attributes:
  - Modulation Frequency: How fast are the modulations.
  - Modulation Level: How strong are the modulations.

Define Vocal Tremor

- Vocal Tremor: Involuntary modulations of frequency and/or amplitude in sustained phonation.
- Pathological & Physiological Vocal Tremor.
  - Pathological Tremor: From diseases like Parkinson, essential tremor, etc. $\Rightarrow$ Strong motor synchronization.
  - Physiological Tremor: Natural stochastic modulations in the interval $[2, 15]Hz$ with low amplitude.
- Acoustic Vocal Tremor Attributes:
  - Modulation Frequency: How fast are the modulations.
  - Modulation Level: How strong are the modulations.

Define Vocal Tremor

- Vocal Tremor: Involuntary modulations of frequency and/or amplitude in sustained phonation.
- Pathological & Physiological Vocal Tremor.
  - Pathological Tremor: From diseases like Parkinson, essential tremor, etc. $\Rightarrow$ Strong motor synchronization.
  - Physiological Tremor: Natural stochastic modulations in the interval $[2, 15]Hz$ with low amplitude.
- Acoustic Vocal Tremor Attributes:
  - Modulation Frequency: How fast are the modulations.
  - Modulation Level: How strong are the modulations.

Define Vocal Tremor

- Vocal Tremor: Involuntary modulations of frequency and/or amplitude in sustained phonation.
- Pathological & Physiological Vocal Tremor.
  - Pathological Tremor: From diseases like Parkinson, essential tremor, etc. $\Rightarrow$ Strong motor synchronization.
  - Physiological Tremor: Natural stochastic modulations in the interval $[2, 15] Hz$ with low amplitude.
- Acoustic Vocal Tremor Attributes:
  - Modulation Frequency: How fast are the modulations.
  - Modulation Level: How strong are the modulations.

Define Vocal Tremor

- Vocal Tremor: Involuntary modulations of frequency and/or amplitude in sustained phonation.
- Pathological & Physiological Vocal Tremor.
  - Pathological Tremor: From diseases like Parkinson, essential tremor, etc. $\Rightarrow$ Strong motor synchronization.
  - Physiological Tremor: Natural stochastic modulations in the interval $[2, 15]Hz$ with low amplitude.
- Acoustic Vocal Tremor Attributes:
  - Modulation Frequency: How fast are the modulations.
  - Modulation Level: How strong are the modulations.

Vocal Tremor Estimation

Use of an AM-FM decomposition algorithm based on the adaptive time-varying quasi-harmonic model for speech.

- High resolution in Time-Frequency plane.
- Estimation of Vocal Tremor for any sinusoidal component of speech.
- Time dependent Vocal Tremor estimations.

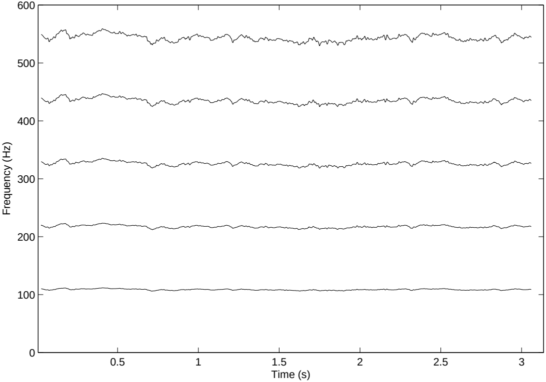
AM-FM Decomposition using aQHM

- Speech is modeled as a sum of AM-FM sinusoids:

$$s(t) = \sum_{k=1}^{K} a_k(t) cos(\phi_k(t))$$

  - K is the number of components,
  - $a_k(t)$ is the instantaneous amplitude of the k^{th} sinusoid,
  - $\phi_k(t)$ is the instantaneous phase of the k^{th} sinusoid, and
  - $f_k(t) = \frac{1}{2\pi}\frac{d\phi_k(t)}{dt}$ is the instantaneous frequency of the k^{th} sinusoid.
- AM-FM decomposition algorithm tries to estimate the instantaneous components.

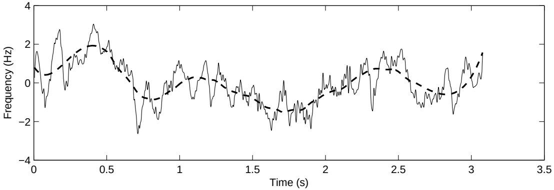
Example of AM-FM decomposition on Speech



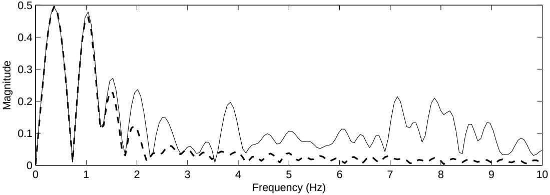
Preprocessing of Inst. Component

- Downsample inst. component to $f_s = 1000Hz$
- Remove the very slow ($< 2Hz$) modulations of the instantaneous component.
- This is performed by Savinzky-Golay smoothing filter.
  - S-G smoothing filter performs a local polynomial regression.
  - S-G filter parameters: 4th order polynomial & 1sec frame size.
  - Advantage: Preserve features of the time-series such as relative maxima, minima and width.

S-G Filter Output



(a)



(b)

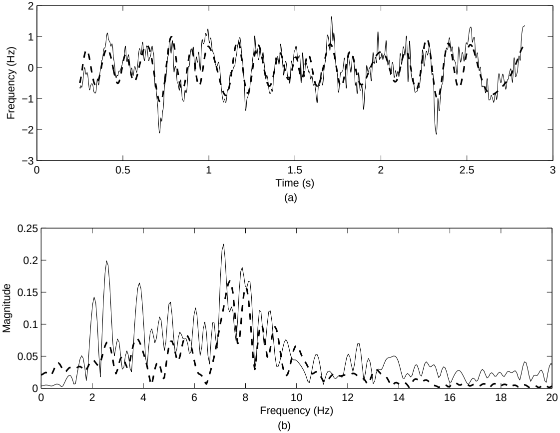
Compute Modulation Frequency & Level

- Assuming that the processed inst. component has a single but time-varying modulation frequency and modulation level.

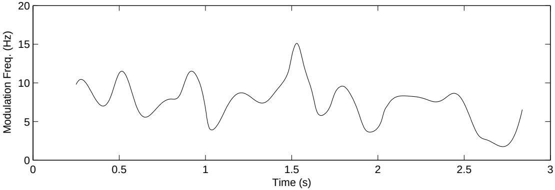
$$x(t) = m(t)cos(\psi(t))$$

- Apply for second time the AM-FM dec. alg. to the processed inst. component.
- Thus,
  - Modulation frequency, $\frac{1}{2\pi}\frac{d\psi(t)}{dt}$, is estimated from the FM component of AM-FM dec. alg.
  - Modulation level, $m(t)$, is estimated from the respective AM component.

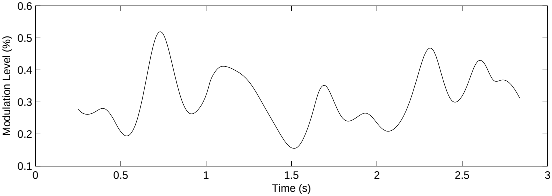
Compute Modulation Frequency & Level



Compute Modulation Frequency & Level



(a)



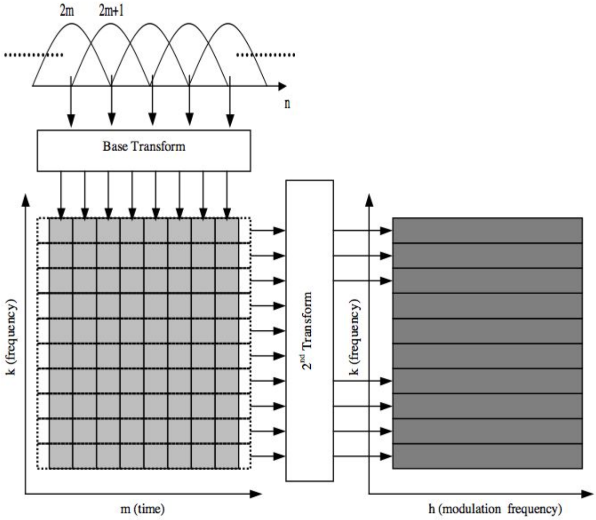
(b)

References for the work on Tremor

1. Yannis Pantazis, Maria Koutsoyannaki and Yannis Stylianou: *A novel method for the extraction of vocal tremor*, MAVEBA-2009, Florence, Italy, 14-16 Dec, 2009.
2. Maria Koutsoyannaki, Yannis Pantazis, Yannis Stylianou, and Philippe Dejonckere: *Tremor in speakers with spasmodic dysphonia*, MAVEBA-2011, Florence Italy, Aug 2011

Outline

Principle


2m
2m+1
n
Base Transform
k (frequency)
m (time)
2nd Transform
k (frequency)
h (modulation frequency)

Our work on Modulation Spectra

1. Maria Markaki and Yannis Stylianou:
   Voice Pathology Detection and Discrimination based on Modulation Spectral Features.
   IEEE Trans. on Audio, Speech and Language Processing.
   TASL.2010.2104141, Jan 2011
2. J.D. Arias-Londono, J.I. Godino-Llorente, M. Markaki, and Y. Stylianou:
   On combining information from Modulation Spectra and Mel-Frequency Cepstral coefficients for Automatic Detection of Pathological Voices.
   Logopedics, Phoniatrics, Vocology (LPV), Nov 2010.
3. Maria Markaki and Yannis Stylianou:
   Discrimination of Speech from Nonspeeech in Broadcast News Based on Modulation Frequency Features
   Speech Communication, Special Issue on Speech Communication on Perceptual and Statistical Audition, July 2010

Outline

Acknowledgments

- My ex-students: Miltos Vasilakis, Yannis Pantazis.
- My colleague: Olivier Rosec, France Telecom
- Juan Ignacio Godino-Llorente, Department of Circuits & Systems Engineering, Universidad Politécnica de Madrid, for the availability of the PdA database.
- Debora Kiagiadaki (MD, ENT) for providing us with the before and after treatment examples.

THANK YOU

for your attention

Outline

J. Lohscheller, U. Eysholdt, H. Toy, and M. Doellinger, “Phonovibrography: Mapping High-Speed Movies of Vocal Fold Vibrations into 2-D Diagrams for Visualizing and Analyzing the Underlying Laryngeal Dynamics,” *IEEE Trans. Medical Imaging*, vol. 27, no. 3, pp. 300–309, 2008.

J. Svec and H. Schutte, “Videokymography: high-speed line scanning of vocal fold vibration,” *Journal of Voice*, vol. 10, no. 2, pp. 201–205, 1996.

N. Henrich, C. d’Alessandro, B. Doval, and M. Castellegno, “On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation,” *Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1321–1332, 2004.