CS578- Speech Signal Processing
Lecture 5: Sinusoidal modeling and modifications

Yannis Stylianou

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

Univ. of Crete
1 Sinusoidal Speech Model
2 Estimation of Sinewave Parameters
   - Voiced Speech
   - Unvoiced Speech
   - The Analysis System
3 Synthesis
   - Linear Amplitude Interpolation
   - Cubic Phase Interpolation
4 Examples
5 Sound Examples
6 Shape Invariant Time-Scale Modifications
   - The Model
   - Parameters Estimation
   - Synthesis
   - Sound Examples
7 Shape Invariant Pitch Modifications
8 Acknowledgments
9 References
**Source-Filter[1]**

- **Source:**
  \[
u(t) = \text{Re} \sum_{k=1}^{K(t)} \alpha_k(t) \exp[j\phi_k(t)]\]

  where:
  \[
  \phi_k(t) = \int_0^t \Omega_k(\sigma) \, d\sigma + \phi_k
  \]

- **Filter:** \(h(t, \tau)\) with Fourier Transform (FT):
  \[
  H(t, \Omega) = M(t, \Omega) \exp[j\Phi(t, \Omega)]
  \]
Source-Filter[1]

- **Source:**
  \[
  u(t) = \text{Re} \sum_{k=1}^{K(t)} \alpha_k(t) \exp[j\phi_k(t)]
  \]
  where:
  \[
  \phi_k(t) = \int_0^t \Omega_k(\sigma)d\sigma + \phi_k
  \]

- **Filter:** \(h(t, \tau)\) with Fourier Transform (FT):
  \[
  H(t, \Omega) = M(t, \Omega) \exp[j\Phi(t, \Omega)]
  \]
Output speech

\[ s(t) = \text{Re} \sum_{k=1}^{K(t)} A_k(t) \exp[j\theta_k(t)] \]

where:

\[
A_k(t) = \alpha_k(t) M[t, \Omega_k(t)] \\
\theta_k(t) = \phi_k(t) + \Phi[t, \Omega_k(t)]
\]

\[
= \int_0^t \Omega_k(\sigma) d\sigma + \Phi[t, \Omega_k(t)] + \phi_k
\]
Frame-by-Frame analysis
Stationarity Assumption

We assume stationarity inside the analysis window:

\[ A_k^l(t) = A_k^l \]
\[ \Omega_k^l(t) = \Omega_k^l \]

which leads to:

\[ \theta_k^l(t) = \Omega_k^l(t - t_l) + \theta_k^l \]

and to:

\[ s(t) = \sum_{k=1}^{K^l} A_k^l \exp(j\theta_k^l) \exp\left[j\Omega_k^l(t - t_l)\right] \quad t_l - \frac{T}{2} \leq t \leq t_l + \frac{T}{2} \]
We assume stationarity inside the analysis window:

\[ A_k^l(t) = A_k^l \]
\[ \Omega_k^l(t) = \Omega_k^l \]

which leads to:

\[ \theta_k^l(t) = \Omega_k^l(t - t_l) + \theta_k^l \]

and to:

\[ s(t) = \sum_{k=1}^{K^l} A_k^l \exp(j\theta_k^l) \exp\left[j\Omega_k^l(t - t_l)\right] \quad t_l - \frac{T}{2} \leq t \leq t_l + \frac{T}{2} \]
We assume stationarity inside the analysis window:

\[ A^l_k(t) = A^l_k \]
\[ \Omega^l_k(t) = \Omega^l_k \]

which leads to:

\[ \theta^l_k(t) = \Omega^l_k(t - t_l) + \theta^l_k \]

and to:

\[ s(t) = \sum_{k=1}^{K^l} A^l_k \exp(j\theta^l_k) \exp\left[j\Omega^l_k(t - t_l)\right] \quad t_l - \frac{T}{2} \leq t \leq t_l + \frac{T}{2} \]
Discrete-Time Formulation

Steps to discrete time formula:

- Time shift: \( t' = t - t_l \)
- Convert to discrete time:

\[
\begin{align*}
\mathcal{S}[n] &= \sum_{k=1}^{K^l} A_k^l \exp(j \theta_k^l) \exp(j \omega_k^l n) - \frac{N_w - 1}{2} \leq n \leq \frac{N_w - 1}{2}
\end{align*}
\]
Mean-Squared Error

Given the original measured waveform, $y[n]$ and the synthetic speech waveform, $s[n]$, estimate the unknown parameters $A_k^l$, $\omega_k^l$, and $\theta_k^l$ by minimizing the MSE criterion:

$$
\epsilon^l = \sum_{n=-(N_w-1)/2}^{n=(N_w-1)/2} |y[n] - s[n]|^2
$$

which can be written as:

$$
\epsilon^l = \sum_{n=-(N_w-1)/2}^{n=(N_w-1)/2} |y[n]|^2 + N_w \sum_{k=1}^{K^l} \left( |Y(\omega_k^l) - \gamma_k^l|^2 - |Y(\omega_k^l)|^2 \right)
$$

which can be reduced further to:

$$
\epsilon^l = \sum_{n=-(N_w-1)/2}^{n=(N_w-1)/2} |y[n]|^2 - N_w \sum_{k=1}^{K^l} |Y(\omega_k^l)|^2
$$
Given the original measured waveform, $y[n]$ and the synthetic speech waveform, $s[n]$, estimate the unknown parameters $A_k^l$, $\omega_k^l$, and $\theta_k^l$ by minimizing the MSE criterion:

$$
\epsilon^l = \sum_{n=-(N_w-1)/2}^{n=(N_w-1)/2} |y[n] - s[n]|^2
$$

which can be written as:

$$
\epsilon^l = \sum_{n=-(N_w-1)/2}^{n=(N_w-1)/2} |y[n]|^2 + N_w \sum_{k=1}^{K^l} \left( |Y(\omega_k^l) - \gamma_k^l|^2 - |Y(\omega_k^l)|^2 \right)
$$

which can be reduced further to:

$$
\epsilon^l = \sum_{n=-(N_w-1)/2}^{n=(N_w-1)/2} |y[n]|^2 - N_w \sum_{k=1}^{K^l} |Y(\omega_k^l)|^2
$$
Mean-Squared Error

Given the original measured waveform, $y[n]$ and the synthetic speech waveform, $s[n]$, estimate the unknown parameters $A_k^l$, $\omega_k^l$, and $\theta_k^l$ by minimizing the MSE criterion:

$$
\epsilon^l = \sum_{n=-(N_w-1)/2}^{n=(N_w-1)/2} |y[n] - s[n]|^2
$$

which can be written as:

$$
\epsilon^l = \sum_{n=-(N_w-1)/2}^{n=(N_w-1)/2} |y[n]|^2 + N_w \sum_{k=1}^{K^l} \left( |Y(\omega_k^l) - \gamma_k^l|^2 - |Y(\omega_k^l)|^2 \right)
$$

which can be reduced further to:

$$
\epsilon^l = \sum_{n=-(N_w-1)/2}^{n=(N_w-1)/2} |y[n]|^2 - N_w \sum_{k=1}^{K^l} |Y(\omega_k^l)|^2
$$
Karhunen-Loève Expansion

- Karhunen-Loève expansion allows constructing a random process from harmonic sinusoids with uncorrelated complex amplitudes.
- Estimated power spectrum should not vary “too much” over consecutive frequencies.

Following the above necessary constraints, for unvoiced speech, and for a window width to be at least 20ms, an 100 Hz harmonic structure provides good results.
Karhunen-Loève expansion allows constructing a random process from harmonic sinusoids with uncorrelated complex amplitudes.

Estimated power spectrum should not vary “too much” over consecutive frequencies.

Following the above necessary constraints, for unvoiced speech, and for a window width to be at least 20ms, an 100 Hz harmonic structure provides good results.
Karhunen-Loève Expansion

- Karhunen-Loève expansion allows constructing a random process from harmonic sinusoids with uncorrelated complex amplitudes.
- Estimated power spectrum should not vary “too much” over consecutive frequencies.

Following the above necessary constraints, for unvoiced speech, and for a window width to be at least 20ms, an 100 Hz harmonic structure provides good results.
Karhunen-Loève expansion allows constructing a random process from harmonic sinusoids with uncorrelated complex amplitudes.

Estimated power spectrum should not vary “too much” over consecutive frequencies.

Following the above necessary constraints, for unvoiced speech, and for a window width to be \textit{at least} 20ms, an 100 Hz harmonic structure provides good results.
Example
Implementation

• Window width be 2.5 times the average pitch period or 20 ms, whichever is larger.

• Use Hamming window, normalized to one:

\[ \sum_{n=-\infty}^{\infty} w[n] = 1 \]

• Use zero padding to get enough samples of the underlying spectrum (i.e., 1024-point FFT)

• Remove linear phase offset

• Refine your frequency estimates
**Implementation**

- Window width be 2.5 times the average pitch period or 20 ms, whichever is larger.
- Use Hamming window, normalized to one:
  \[ \sum_{n=-\infty}^{\infty} w[n] = 1 \]
- Use zero padding to get enough samples of the underlying spectrum (i.e., 1024-point FFT)
- Remove linear phase offset
- Refine your frequency estimates
Implementation

- Window width be 2.5 times the average pitch period or 20 ms, whichever is larger.
- Use Hamming window, normalized to one:
  \[
  \sum_{n=-\infty}^{\infty} w[n] = 1
  \]
- Use zero padding to get enough samples of the underlying spectrum (i.e., 1024-point FFT)
- Remove linear phase offset
- Refine your frequency estimates
**Implementation**

- Window width be 2.5 times the average pitch period or 20 ms, whichever is larger.
- Use Hamming window, normalized to one:
  \[ \sum_{n=-\infty}^{\infty} w[n] = 1 \]
- Use zero padding to get enough samples of the underlying spectrum (i.e., 1024-point FFT)
- Remove linear phase offset
- Refine your frequency estimates
**Implementation**

- Window width be 2.5 times the average pitch period or 20 ms, whichever is larger.
- Use Hamming window, normalized to one:
  \[ \sum_{n=-\infty}^{\infty} w[n] = 1 \]
- Use zero padding to get enough samples of the underlying spectrum (i.e., 1024-point FFT)
- Remove linear phase offset
- Refine your frequency estimates
Showing the process ...
Block diagram of the Analysis System
Problem of frequency matching
Frame-to-Frame Peak Matching
The birth/death process
A birth/death process in speech
Why not to estimate the original speech waveform on the $l$th frame, directly as:

$$s[n] = \sum_{k=1}^{K^l} A^l_k \cos (n\omega^l_k + \theta^l_k), \quad n = 0, 1, 2, \ldots, L - 1$$
A simple solution: OLA

Frame l−1
+
Frame l
+
Frame l+1

Synthesized speech for frame l
Amplitude Interpolation

Linear Interpolation:

\[ A_k^l[n] = A_k^l + \left( A_{k+1}^l - A_k^l \right) \left( \frac{n}{L} \right) \quad n = 0, 1, 2, \ldots, L - 1 \]
Phase wrapped
Cubic Phase model

\[ \theta(t) = \zeta + \gamma t + \alpha t^2 + \beta t^3 \]
Assuming that vocal tract is slowly varying, and since:

$$\theta(t) = \int_0^t \Omega(\sigma)d\sigma + \phi + \Phi[t, \Omega(t)]$$

So:

$$\dot{\theta}(t) \approx \Omega(t)$$

$$\dot{\theta}' \approx \Omega'$$

$$\dot{\theta}' + 1 \approx \Omega' + 1$$
Assuming that vocal tract is slowly varying, and since:

$$\theta(t) = \int_0^t \Omega(\sigma) d\sigma + \phi + \Phi[t, \Omega(t)]$$

$$\dot{\theta}(t) \approx \Omega(t)$$

So:

$$\dot{\theta}' \approx \Omega'$$
$$\dot{\theta}' + 1 \approx \Omega' + 1$$
Assuming that vocal tract is slowly varying, and since:

\[ \theta(t) = \int_0^t \Omega(\sigma)d\sigma + \phi + \Phi[t, \Omega(t)] \]

\[ \dot{\theta}(t) \approx \Omega(t) \]

So:

\[ \dot{\theta}' \approx \Omega' \]
\[ \dot{\theta}' + 1 \approx \Omega' + 1 \]
There are four constraints

\[
\begin{align*}
\theta(0) &= \theta' \\
\dot{\theta}(0) &= \Omega' \\
\theta(T) &= \theta' + 1 + 2\pi M \\
\dot{\theta}(T) &= \Omega' + 1
\end{align*}
\]

and ... five unknowns (don’t forget M) 
We need one more constraint!
There are four constraints

\[ \begin{align*}
\theta(0) &= \theta' \\
\dot{\theta}(0) &= \Omega' \\
\theta(T) &= \theta' + 1 + 2\pi M \\
\dot{\theta}(T) &= \Omega' + 1
\end{align*} \]

and ... five unknowns (don't forget M)
We need one more constraint!
There are four constraints

\begin{align*}
\theta(0) &= \theta' \\
\dot{\theta}(0) &= \Omega' \\
\theta(T) &= \theta^{l+1} + 2\pi M \\
\dot{\theta}(T) &= \Omega^{l+1}
\end{align*}

and ... five unknowns (don’t forget M)

We need one more constraint!
There are four constraints

\[ \begin{align*}
\theta(0) &= \theta' \\
\dot{\theta}(0) &= \Omega' \\
\theta(T) &= \theta'^{+1} + 2\pi M \\
\dot{\theta}(T) &= \Omega'^{+1}
\end{align*} \]

and ... five unknowns (don’t forget M) 
We need one more constraint!
How to choose $M$

$$\theta(t) = \theta^l + \Omega^l t + \alpha(M)t^2 + \beta(M)t^3$$

- $\theta^{l+1} + 8\pi$, $M = 4$
- $\theta^{l+1} + 6\pi$, $M = 3$
- $\theta^{l+1} + 4\pi$, $M = 2$
- $\theta^{l+1} + 2\pi$, $M = 1$
- $\theta^{l+1}$, $M = 0$

Slope = $\omega^l$

Slope = $\omega^{l+1}$

$t = 0$ to $t = T$
Estimating $M$

- Find $M$ that minimizes the criterion:

$$f(M) = \int_0^T \left[ \ddot{\theta}(t; M) \right]^2 dt$$

- Using continuous variable:

$$x^* = \frac{1}{2\pi} \left[ (\theta^l + \Omega^l T - \theta^{l+1}) + (\Omega^{l+1} - \Omega^l) \frac{T}{2} \right]$$

- $M^*$ is the nearest integer to $x^*$
Estimating $M$

- Find $M$ that minimizes the criterion:

$$f(M) = \int_0^T \left[ \ddot{\theta}(t; M) \right]^2 dt$$

- Using continuous variable:

$$x^* = \frac{1}{2\pi} \left[ (\theta^l + \Omega^l T - \theta^{l+1}) + (\Omega^{l+1} - \Omega^l) \frac{T}{2} \right]$$

- $M^*$ is the nearest integer to $x^*$
Estimating $M$

- Find $M$ that minimizes the criterion:

$$f(M) = \int_{0}^{T} \left[ \ddot{\theta}(t; M) \right]^2 dt$$

- Using continuous variable:

$$x^* = \frac{1}{2\pi} \left[ (\theta^l + \Omega^l T - \theta^{l+1}) + (\Omega^{l+1} - \Omega l) \frac{T}{2} \right]$$

- $M^*$ is the nearest integer to $x^*$
Block diagram of the Synthesis System
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sinusoidal Speech Model</td>
</tr>
<tr>
<td>2</td>
<td>Estimation of Sinewave Parameters</td>
</tr>
<tr>
<td></td>
<td>- Voiced Speech</td>
</tr>
<tr>
<td></td>
<td>- Unvoiced Speech</td>
</tr>
<tr>
<td></td>
<td>- The Analysis System</td>
</tr>
<tr>
<td>3</td>
<td>Synthesis</td>
</tr>
<tr>
<td></td>
<td>- Linear Amplitude Interpolation</td>
</tr>
<tr>
<td></td>
<td>- Cubic Phase Interpolation</td>
</tr>
<tr>
<td>4</td>
<td>Examples</td>
</tr>
<tr>
<td>5</td>
<td>Sound Examples</td>
</tr>
<tr>
<td>6</td>
<td>Shape Invariant Time-Scale Modifications</td>
</tr>
<tr>
<td></td>
<td>- The Model</td>
</tr>
<tr>
<td></td>
<td>- Parameters Estimation</td>
</tr>
<tr>
<td></td>
<td>- Synthesis</td>
</tr>
<tr>
<td></td>
<td>- Sound Examples</td>
</tr>
<tr>
<td>7</td>
<td>Shape Invariant Pitch Modifications</td>
</tr>
<tr>
<td>8</td>
<td>Acknowledgments</td>
</tr>
<tr>
<td>9</td>
<td>References</td>
</tr>
</tbody>
</table>
Reconstruction Example
Reconstruction Example
1 **Sinusoidal Speech Model**
2 **Estimation of Sinewave Parameters**
   - Voiced Speech
   - Unvoiced Speech
   - The Analysis System
3 **Synthesis**
   - Linear Amplitude Interpolation
   - Cubic Phase Interpolation
4 **Examples**
5 **Sound Examples**
6 **Shape Invariant Time-Scale Modifications**
   - The Model
   - Parameters Estimation
   - Synthesis
   - Sound Examples
7 **Shape Invariant Pitch Modifications**
8 **Acknowledgments**
9 **References**
### Sound Examples

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Mixed</th>
<th>Min</th>
<th>Zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td><img src="audio1.m4a" alt="Audio" /></td>
<td><img src="audio2.m4a" alt="Audio" /></td>
<td><img src="audio3.m4a" alt="Audio" /></td>
<td><img src="audio4.m4a" alt="Audio" /></td>
</tr>
<tr>
<td>Female</td>
<td><img src="audio5.m4a" alt="Audio" /></td>
<td><img src="audio6.m4a" alt="Audio" /></td>
<td><img src="audio7.m4a" alt="Audio" /></td>
<td><img src="audio8.m4a" alt="Audio" /></td>
</tr>
<tr>
<td>Male</td>
<td><img src="audio9.m4a" alt="Audio" /></td>
<td><img src="audio10.m4a" alt="Audio" /></td>
<td><img src="audio11.m4a" alt="Audio" /></td>
<td><img src="audio12.m4a" alt="Audio" /></td>
</tr>
<tr>
<td>Female</td>
<td><img src="audio13.m4a" alt="Audio" /></td>
<td><img src="audio14.m4a" alt="Audio" /></td>
<td><img src="audio15.m4a" alt="Audio" /></td>
<td><img src="audio16.m4a" alt="Audio" /></td>
</tr>
</tbody>
</table>
1 Sinusoidal Speech Model
2 Estimation of Sinewave Parameters
   - Voiced Speech
   - Unvoiced Speech
   - The Analysis System
3 Synthesis
   - Linear Amplitude Interpolation
   - Cubic Phase Interpolation
4 Examples
5 Sound Examples
6 Shape Invariant Time-Scale Modifications
   - The Model
   - Parameters Estimation
   - Synthesis
   - Sound Examples
7 Shape Invariant Pitch Modifications
8 Acknowledgements
9 References
Excitation model

We have seen that:

\[ u(t) = \sum_{k=1}^{K(t)} \alpha_k(t) \exp[j \phi_k(t)] \]

where:

\[ \phi_k(t) = \int_{0}^{t} \Omega_k(\sigma)d\sigma + \phi_k \]

Assuming voiced speech and constant frequency in the analysis window, then:

\[ u(t) = \sum_{k=1}^{K(t)} \alpha_k(t) \exp[j(t - t_0)\Omega_k] \quad t \in [0, T] \]
Excitation model

We have seen that:

\[ u(t) = \sum_{k=1}^{K(t)} \alpha_k(t) \exp[j\phi_k(t)] \]

where:

\[ \phi_k(t) = \int_{0}^{t} \Omega_k(\sigma) d\sigma + \phi_k \]

Assuming voiced speech and constant frequency in the analysis window, then:

\[ u(t) = \sum_{k=1}^{K(t)} \alpha_k(t) \exp[j(t - t_0)\Omega_k] \quad t \in [0, T] \]
**Speech model [2]**

Then:

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

where:

\[ A_k(t) = \alpha_k(t) M_k(t) \]

\[ \theta_k(t) = \phi_k(t) + \Phi_k(t) \]

Therefore:

\[ \Phi_k(t) = \theta_k(t) - (t - t_0)\Omega_k \]
Uniform time-scale, by $\rho$

Let’s $t$ represent the original articulation rate and $t'$ the transformed rate:

$$t' = \rho \, t$$

Given the source/filter model:

- System parameters are time-scaled
- Excitation parameters (phase) are scaled in such a way to maintain fundamental frequency.
Uniform time-scale, by $\rho$

Let’s $t$ represent the original articulation rate and $t'$ the transformed rate:

$$t' = \rho \ t$$

Given the source/filter model:

- System parameters are time-scaled
- Excitation parameters (phase) are scaled in such a way to maintain fundamental frequency.
Onset-time model for time-scale

Excitation $\tilde{u}(t')$ → System $M(\Omega; t'\rho^{-1}) \exp[j\Phi(\Omega; t'\rho^{-1})] →$ Transformed Speech $\tilde{s}(t)$

$t' - T \rho$
Excitation function in $t'$

- Time-scaled pitch period:
  \[ \tilde{P}(t') = P(t'\rho^{-1}) \]

- Modified excitation function
  \[ \tilde{u}(t') = \sum_{k=1}^{K(t)} \tilde{\alpha}_k(t') \exp\left[j\tilde{\phi}_k(t')\right] \]

  where:
  \[ \tilde{\phi}_k(t') = (t'\rho^{-1} - t_0')\Omega_k \]

  and
  \[ \tilde{\alpha}_k(t') = \alpha_k(t'\rho^{-1}) \]
System function parameters in $t'$

\[
\tilde{M}_k(t') = M_k(t' \rho^{-1}) \\
\tilde{\Phi}_k(t') = \Phi_k(t' \rho^{-1})
\]
Waveform in $t'$

$$\tilde{\xi}(t') = \sum_{k=1}^{K(t)} \tilde{A}_k(t') \exp \left[ j\tilde{\theta}_k(t') \right]$$

where

$$\tilde{A}_k(t') = \tilde{\alpha}_k(t') \tilde{M}_k(t')$$
$$\tilde{\theta}_k(t') = \tilde{\phi}_k(t') + \tilde{\Phi}_k(t')$$
Onset times estimation

Frame centers

\[ u[n] \]

\[ n_0(0), n_0(1), n_0(2), \ldots \]

\[ L, 2L, 3L, \ldots \]

\[ \tilde{u}[n'] \]

\[ n'_0(0), n'_0(1), n'_0(2), \ldots \]

\[ L', 2L', 3L', \ldots \]

\[ L' = pL \]

\[ n'_0(l) = \text{Onset Time Relative to } L \]

\[ n'_0(l) = \text{Onset Time Relative to } L' \]
Estimating System Phase

Let’s assume that the onset time $n_o(l)$ for the $l^{th}$ frame is known, then:

$$\phi^l_k = \hat{n}_o(l)\omega^l_k$$

where $\hat{n}_o(l) = n_o(l) - lL$.

Then, the system phase is estimated as:

$$\tilde{\phi}^l_k = \theta^l_k - \phi^l_k$$
Let’s assume that the onset time $n_o(l)$ for the $l^{th}$ frame is known, then:

$$\phi^l_k = \hat{n}_o(l)\omega^l_k$$

where $\hat{n}_o(l) = n_o(l) - lL$.

Then, the system phase is estimated as:

$$\tilde{\Phi}^l_k = \theta^l_k - \phi^l_k$$
Estimating Excitation Phase

Let’s assume we know the onset time in the previous frame \( l - 1 \), then the current onset time in \( t' \), is given by:

\[
n_o'(l) = n_o'(l - 1) + J' P^l
\]

and then:

\[
\tilde{\phi}_k^l = (n_o'(l) - lL') \omega_k^l
\]

where \( L' = \rho L \)
Synthesis

Synthesis is performed in the same way as if no modification is applied:

- Linear interpolation for amplitudes
- Cubic interpolation for phases
Block diagram for Analysis/Synthesis for Time-Scale modification
Example of Time-scale modification
Sound Examples

<table>
<thead>
<tr>
<th></th>
<th>0.5</th>
<th>0.8</th>
<th>Orig</th>
<th>1.2</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
</tr>
<tr>
<td>Female</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
</tr>
<tr>
<td>Trumpet</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
<td>![Audio Clip]</td>
</tr>
</tbody>
</table>
Reading paper

Paper:

T. F. Quatieri and R. J. McAulay:
Shape Invariant Time-Scale and Pitch Modification of Speech
pp 497-510, March 1992
1 Sinusoidal Speech Model
2 Estimation of Sinewave Parameters
   • Voiced Speech
   • Unvoiced Speech
   • The Analysis System
3 Synthesis
   • Linear Amplitude Interpolation
   • Cubic Phase Interpolation
4 Examples
5 Sound Examples
6 Shape Invariant Time-Scale Modifications
   • The Model
   • Parameters Estimation
   • Synthesis
   • Sound Examples
7 Shape Invariant Pitch Modifications
8 Acknowledgments
9 References
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
OUTLINE

1 Sinusoidal Speech Model

2 Estimation of Sinewave Parameters
   - Voiced Speech
   - Unvoiced Speech
   - The Analysis System

3 Synthesis
   - Linear Amplitude Interpolation
   - Cubic Phase Interpolation

4 Examples

5 Sound Examples

6 Shape Invariant Time-Scale Modifications
   - The Model
   - Parameters Estimation
   - Synthesis
   - Sound Examples

7 Shape Invariant Pitch Modifications

8 Acknowledgments

9 References
