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

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

Source-Filter[1]

- Source:

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

where:

$$\phi_k(t) = \int_0^t \Omega_k(\sigma) d\sigma + \phi_k$$

- Filter: $h(t, \tau)$ with Fourier Transform (FT):

$$H(t, \Omega) = M(t, \Omega) \exp\left[j\Phi(t, \Omega)\right]$$

- Source:

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

where:

$$\phi_k(t) = \int_0^t \Omega_k(\sigma) d\sigma + \phi_k$$

- Filter: $h(t, \tau)$ with Fourier Transform (FT):

$$H(t, \Omega) = M(t, \Omega) \exp\left[j\Phi(t, \Omega)\right]$$

Output speech

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

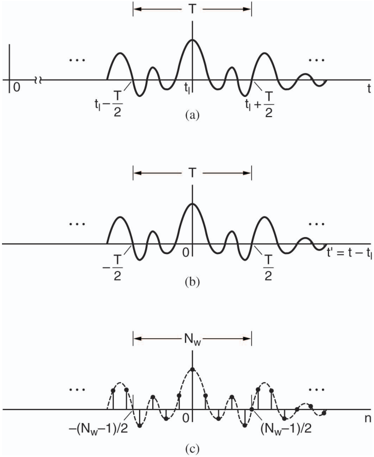
where:

$$\begin{aligned} A_k(t) &= \alpha_k(t) M\left[t, \Omega_k(t)\right] \\ \theta_k(t) &= \phi_k(t) + \Phi\left[t, \Omega_k(t)\right] \\ &= \int_0^t \Omega_k(\sigma) d\sigma + \Phi\left[t, \Omega_k(t)\right] + \phi_k \end{aligned}$$

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

Frame-by-Frame analysis



Stationarity Assumption

We assume stationarity inside the analysis window:

$$
\begin{aligned}
A_k^l(t) &= A_k^l \\
\Omega_k^l(t) &= \Omega_k^l
\end{aligned}
$$

which leads to:

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

and to:

$$s(t) = \sum_{k=1}^{K^l} A_k^l \exp\left(j\theta_k^l\right) \exp\left[j\Omega_k^l(t - t_l)\right] \quad t_l - \frac{T}{2} \leq t \leq t_l + \frac{T}{2}$$

Stationarity Assumption

We assume stationarity inside the analysis window:

$$\begin{aligned} A_k^l(t) &= A_k^l \\ \Omega_k^l(t) &= \Omega_k^l \end{aligned}$$

which leads to:

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

and to:

$$s(t) = \sum_{k=1}^{K^l} A_k^l \exp\left(j\theta_k^l\right) \exp\left[j\Omega_k^l(t - t_l)\right] \quad t_l - \frac{T}{2} \leq t \leq t_l + \frac{T}{2}$$

Stationarity Assumption

We assume stationarity inside the analysis window:

$$\begin{aligned} A_k^l(t) &= A_k^l \\ \Omega_k^l(t) &= \Omega_k^l \end{aligned}$$

which leads to:

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

and to:

$$s(t) = \sum_{k=1}^{K^l} A_k^l \exp\left(j\theta_k^l\right) \exp\left[j\Omega_k^l(t - t_l)\right] \quad t_l - \frac{T}{2} \leq t \leq t_l + \frac{T}{2}$$

Discrete-Time Formulation

Steps to discrete time formula:

- Time shift: $t^{'} = t - t_l$
- Convert to discrete time:

$$s[n] = \sum_{k=1}^{K^l} A_k^l \exp\left(j\theta_k^l\right) \exp\left(j\omega_k^l n\right) \quad -\frac{N_w - 1}{2} \le n \le \frac{N_w - 1}{2}$$

Mean-Squared Error

Given the original measured waveform, $y[n]$ and the synthetic speech waveform, $s[n]$, estimate the unknown parameters A_k^l, ω_k^l, and θ_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:

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

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, ω_k^l, and θ_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:

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

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, ω_k^l, and θ_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:

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

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

- 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.

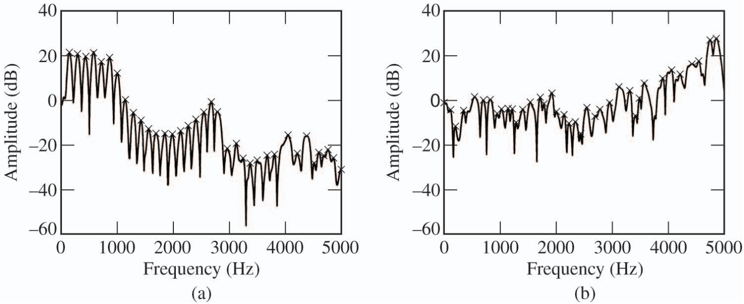
Following the above necessary constraints, for unvoiced speech, and for a window width to be *at least* 20ms, an 100 Hz harmonic structure provides good results.

Karhunen-Loève Expansion

- 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.

Example



40
20
0
–20
–40
–60
0
1000
2000
3000
4000
5000
Amplitude (dB)
Frequency (Hz)
(a)
40
20
0
–20
–40
–60
0
1000
2000
3000
4000
5000
Amplitude (dB)
Frequency (Hz)
(b)

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

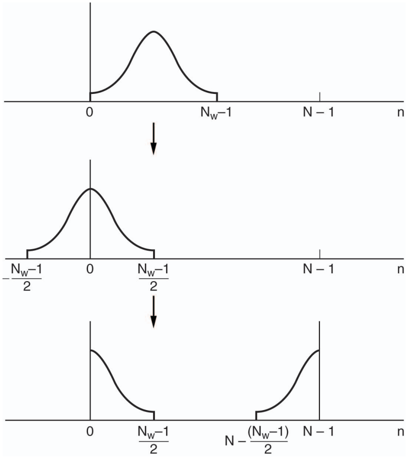
Implementation

- Window width be 2.5 times the average pitch period or 20 ms, whichever is larger.
- Use Hamming window, normalized to one:

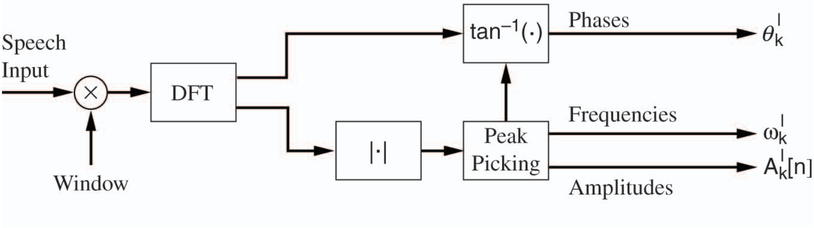
$$\sum_{n=-\infty}^{\infty} w[n] = 1$$

- 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 ...



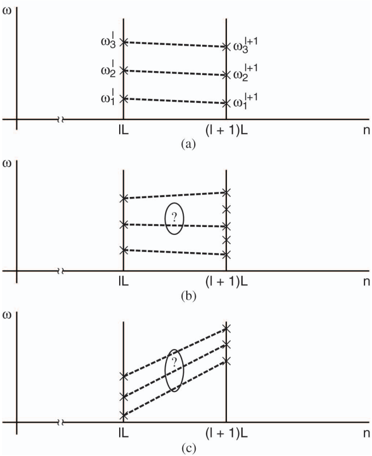
Block diagram of the Analysis System



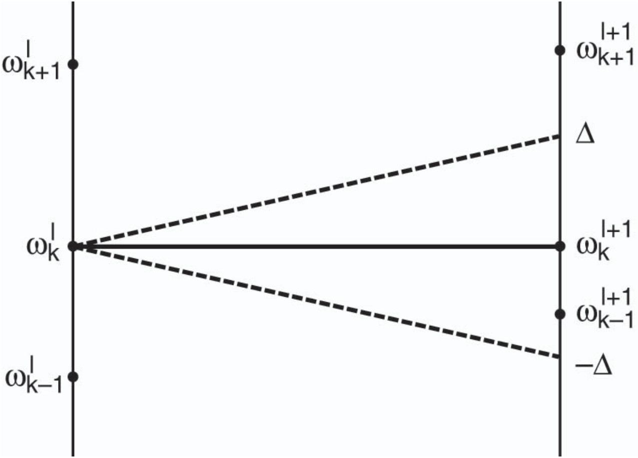
Outline

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

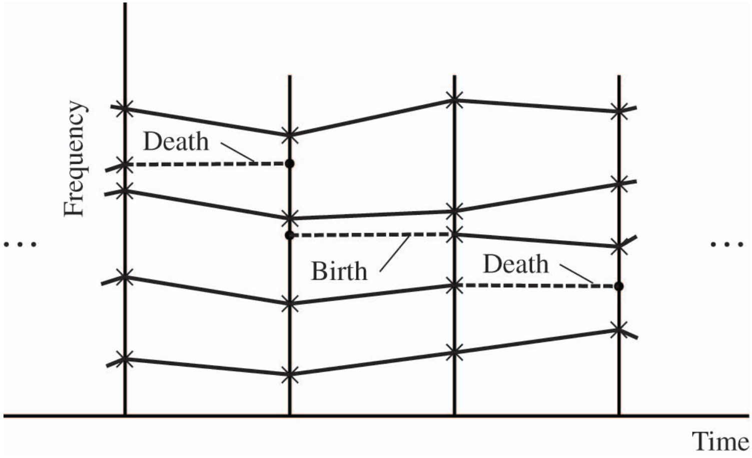
Problem of frequency matching



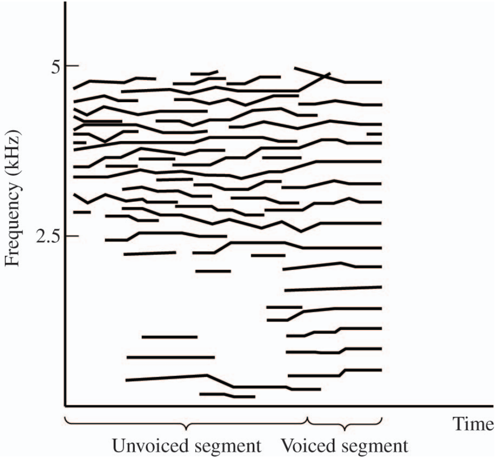
Frame-to-Frame Peak Matching



The birth/death process



A birth/death process in speech

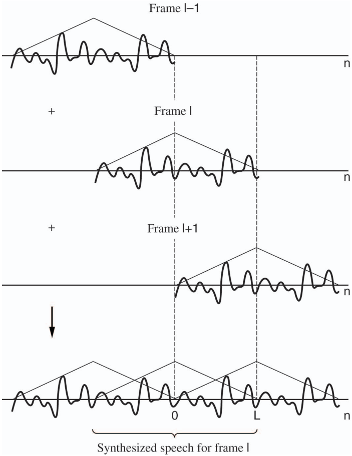


Why not ...

Why not to estimate the original speech waveform on the lth frame, directly as:

$$s[n] = \sum_{k=1}^{K^l} A_k^l \cos\left(n\omega_k^l + \theta_k^l\right), \quad n = 0, 1, 2, \cdots, L-1$$

A simple solution: OLA

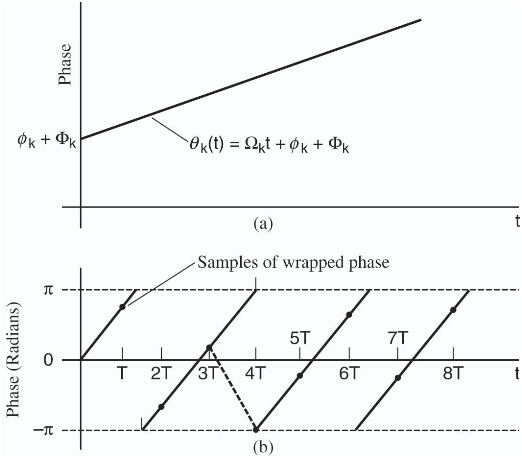


Amplitude Interpolation

Linear Interpolation:

$$A_k^l[n] = A_k^l + \left(A_k^{l+1} - A_k^l\right)\left(\frac{n}{L}\right) \quad n = 0, 1, 2, \cdots, L-1$$

Phase wrapped



Cubic Phase model

$$\theta(t) = \zeta + \gamma t + \alpha t^2 + \beta t^3$$

About the phase derivative

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:

$$\begin{aligned} \dot{\theta}^l &\approx \Omega^l \\ \dot{\theta}^{l+1} &\approx \Omega^{l+1} \end{aligned}$$

About the phase derivative

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:

$$\begin{aligned} \dot{\theta}^l &\approx \Omega^l \\ \dot{\theta}^{l+1} &\approx \Omega^{l+1} \end{aligned}$$

About the phase derivative

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:

$$\begin{aligned} \dot{\theta}^l &\approx \Omega^l \\ \dot{\theta}^{l+1} &\approx \Omega^{l+1} \end{aligned}$$

Four constraints for phase polynomial

There are four constraints

$$
\begin{aligned}
\theta(0) &= \theta^{I} \\
\dot{\theta}(0) &= \Omega^{I} \\
\theta(T) &= \theta^{I+1} + 2\pi M \\
\dot{\theta}(T) &= \Omega^{I+1}
\end{aligned}
$$

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

Four constraints for phase polynomial

There are four constraints

$$
\begin{aligned}
\theta(0) &= \theta^{l} \\
\dot{\theta}(0) &= \Omega^{l} \\
\theta(T) &= \theta^{l+1} + 2\pi M \\
\dot{\theta}(T) &= \Omega^{l+1}
\end{aligned}
$$

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

Four constraints for phase polynomial

There are four constraints

$$
\begin{aligned}
\theta(0) &= \theta^{l} \\
\dot{\theta}(0) &= \Omega^{l} \\
\theta(T) &= \theta^{l+1} + 2\pi M \\
\dot{\theta}(T) &= \Omega^{l+1}
\end{aligned}
$$

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

We need one more constraint!

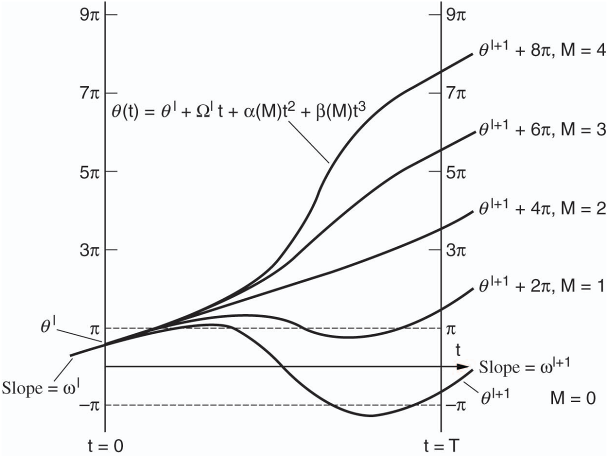
Four constraints for phase polynomial

There are four constraints

$$
\begin{aligned}
\theta(0) &= \theta^{I} \\
\dot{\theta}(0) &= \Omega^{I} \\
\theta(T) &= \theta^{I+1} + 2\pi M \\
\dot{\theta}(T) &= \Omega^{I+1}
\end{aligned}
$$

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

How to choose M



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:

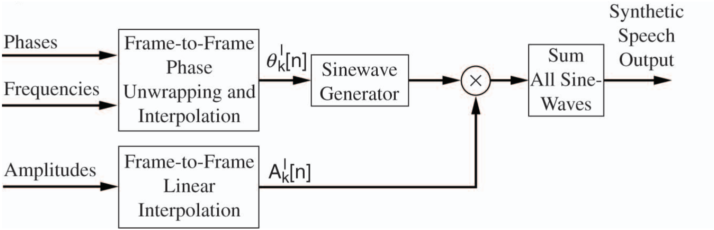
$$f(M) = \int_0^T \left[\ddot{\theta}(t; M)\right]^2 dt$$

- 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^*

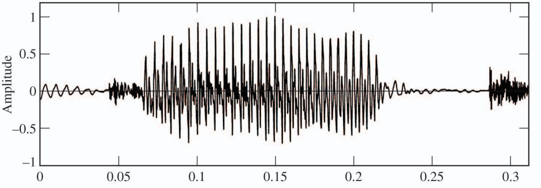
Block diagram of the Synthesis System



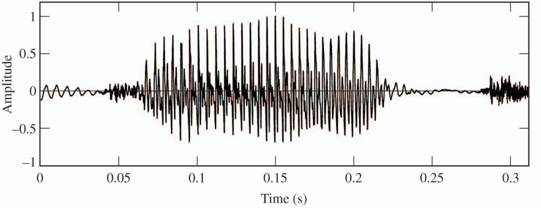
Outline

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

Reconstruction Example

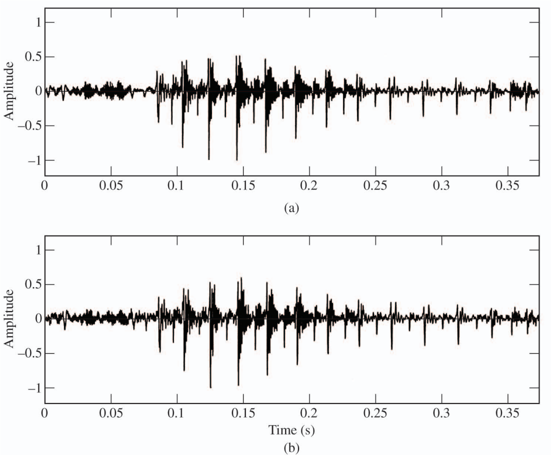


(a)

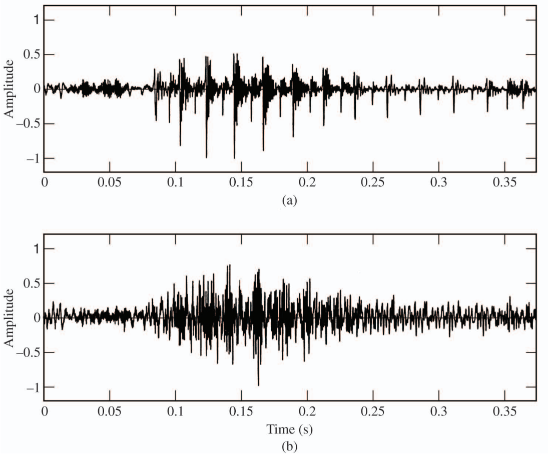


(b)

Reconstruction Example



Magnitude-only Reconstruction Example



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

Sound Examples

	Original	Mixed	Min	Zero
Male				
Female				
Male				
Female				

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

Excitation model

We have seen that:

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

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:

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

Excitation model

We have seen that:

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

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:

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

Speech model[2]

Then:

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

where:

$$\begin{aligned} A_k(t) &= \alpha_k(t) M_k(t) \\ \theta_k(t) &= \phi_k(t) + \Phi_k(t) \end{aligned}$$

Therefore:

$$\Phi_k(t) = \theta_k(t) - (t - t_0)\Omega_k$$

UNIFORM TIME-SCALE, BY ρ

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 ρ

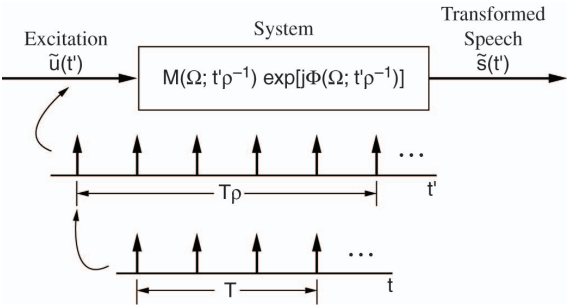
Let's t represent the original articulation rate and t' the transformed rate:

$$t' = \rho\ t$$

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 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'

$$
\begin{aligned}
\tilde{M}_k(t') &= M_k(t'\rho^{-1}) \\
\tilde{\Phi}_k(t') &= \Phi_k(t'\rho^{-1})
\end{aligned}
$$

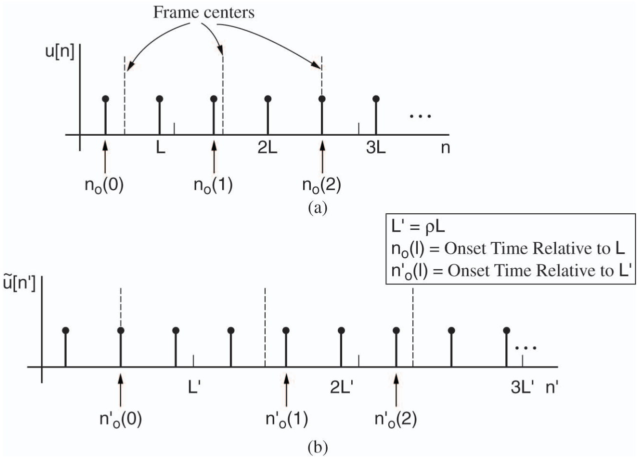
Waveform in t'

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

where

$$\begin{aligned}
\tilde{A}_k(t') &= \tilde{\alpha}_k(t')\tilde{M}_k(t') \\
\tilde{\theta}_k(t') &= \tilde{\phi}_k(t') + \tilde{\Phi}_k(t')
\end{aligned}$$

Onset times estimation



Estimating System Phase

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

$$\phi_k^l = \hat{n}_o(l)\omega_k^l$$

where $\hat{n}_o(l) = n_o(l) - lL$.
Then, the system phase is estimated as:

$$\tilde{\Phi}_k^l = \theta_k^l - \phi_k^l$$

Estimating System Phase

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

$$\phi_k^l = \hat{n}_o(l)\omega_k^l$$

where $\hat{n}_o(l) = n_o(l) - lL$.
Then, the system phase is estimated as:

$$\tilde{\Phi}_k^l = \theta_k^l - \phi_k^l$$

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:

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

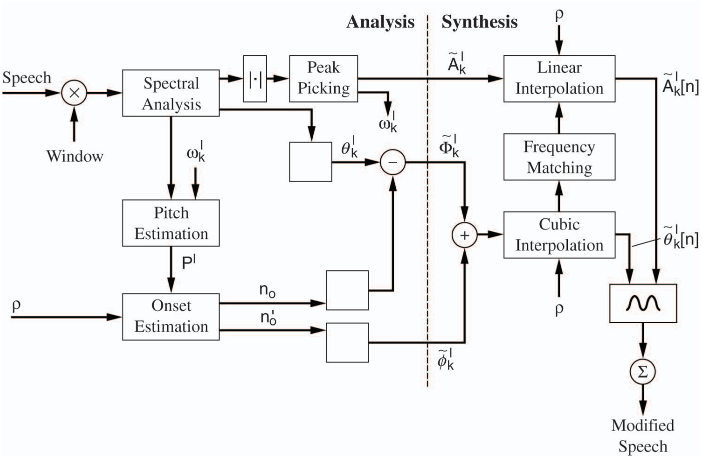
where $L^{'} = \rho L$

Synthesis

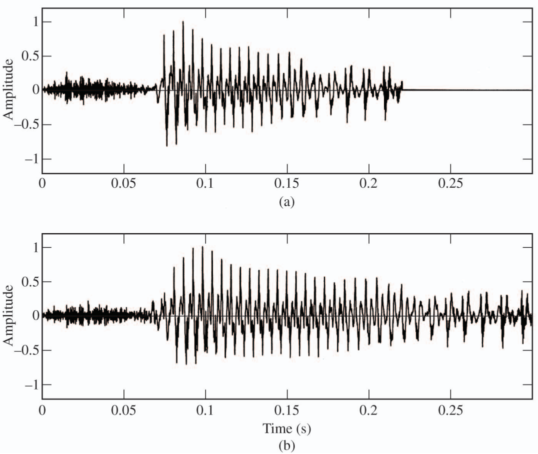
Synthesis is performed in the same way as if no modification is applied:

- Linear interpolation for amplitudes
- Cubic interpolation for phases

Block diagram for Analysis/Synthesis for Time-Scale modification



Example of Time-scale modification



	0.5	0.8	Orig	1.2	1.5
Male					
Female					

	0.75	Orig	1.25
Trumpet			

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

Paper:

T. F. Quatieri and R. J. McAulay:
Shape Invariant Time-Scale and Pitch Modification of Speech
IEEE Trans. Acoust., Speech, Signal Processing, Vol.40, No.3,
pp 497-510, March 1992

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

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

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

R. J. McAulay and T. F. Quatieri, “Speech analysis/synthesis based on a sinusoidal representation,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 744–754, Aug 1986.

T. F. Quatieri and R. J. McAulay, “Shape Invariant Time-Scale and Pitch Modification of Speech,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-40, pp. 497–510, March 1992.