



**ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

Ψηφιακή Επεξεργασία Φωνής

**Ενότητα 5η: Ημιτονοειδής Ανάλυση και
Τροποποίηση Φωνής**

Στυλιανού Ιωάννης

Τμήμα Επιστήμης Υπολογιστών

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:

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

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:

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

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

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

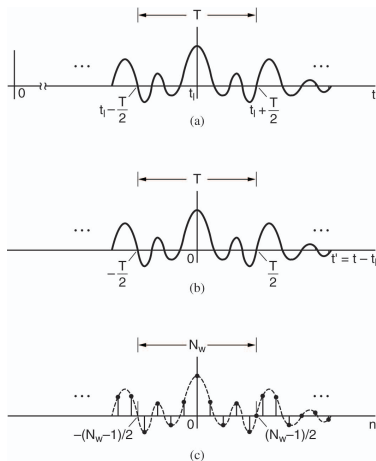
where:

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

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:

$$\begin{aligned}A'_k(t) &= A'_k \\ \Omega'_k(t) &= \Omega'_k\end{aligned}$$

which leads to:

$$\theta'_k(t) = \Omega'_k(t - t_l) + \theta'_k$$

and to:

$$s(t) = \sum_{k=1}^{K'} A'_k \exp(j\theta'_k) \exp[j\Omega'_k(t - t_l)] \quad t_l - \frac{T}{2} \leq t \leq t_l + \frac{T}{2}$$

STATIONARITY ASSUMPTION

We assume stationarity inside the analysis window:

$$\begin{aligned}A'_k(t) &= A'_k \\ \Omega'_k(t) &= \Omega'_k\end{aligned}$$

which leads to:

$$\theta'_k(t) = \Omega'_k(t - t_l) + \theta'_k$$

and to:

$$s(t) = \sum_{k=1}^{K'} A'_k \exp(j\theta'_k) \exp[j\Omega'_k(t - t_l)] \quad t_l - \frac{T}{2} \leq t \leq t_l + \frac{T}{2}$$

STATIONARITY ASSUMPTION

We assume stationarity inside the analysis window:

$$\begin{aligned}A'_k(t) &= A'_k \\ \Omega'_k(t) &= \Omega'_k\end{aligned}$$

which leads to:

$$\theta'_k(t) = \Omega'_k(t - t_l) + \theta'_k$$

and to:

$$s(t) = \sum_{k=1}^{K'} A'_k \exp(j\theta'_k) \exp[j\Omega'_k(t - t_l)] \quad t_l - \frac{T}{2} \leq t \leq t_l + \frac{T}{2}$$

DISCRETE-TIME FORMULATION

Steps to discrete time formula:

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

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

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.

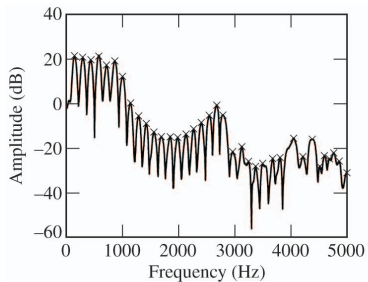
Following the above necessary constraints, for unvoiced speech, and for a window width to be *at least* 20ms, an 100 Hz harmonic structure provides good results.

KARHUNEN-LOÈVE EXPANSION

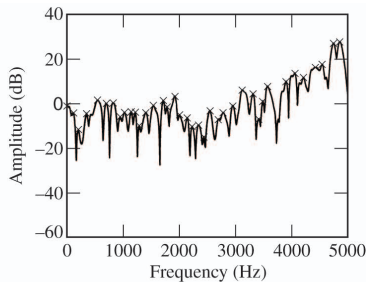
- Karhunen-Loève expansion allows constructing a random process from harmonic sinusoids with uncorrelated complex amplitudes.
- Estimated power spectrum should not vary “too much” over consecutive frequencies.

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



(a)



(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

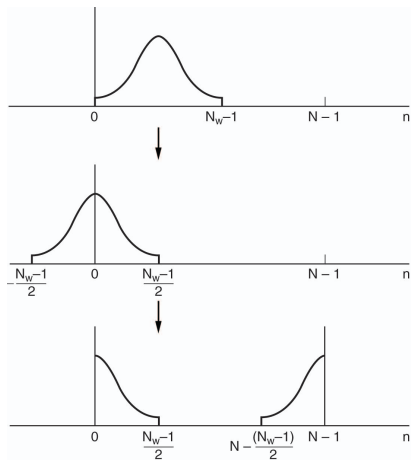
IMPLEMENTATION

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

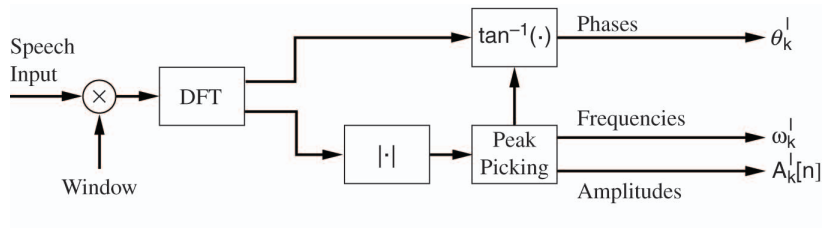
$$\sum_{n=-\infty}^{\infty} w[n] = 1$$

- 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 ...



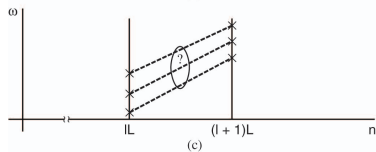
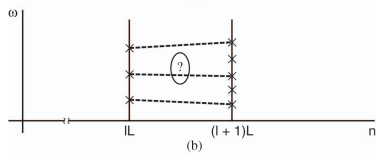
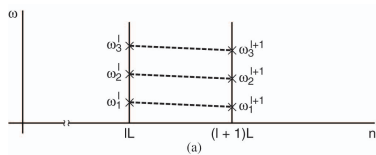
BLOCK DIAGRAM OF THE ANALYSIS SYSTEM



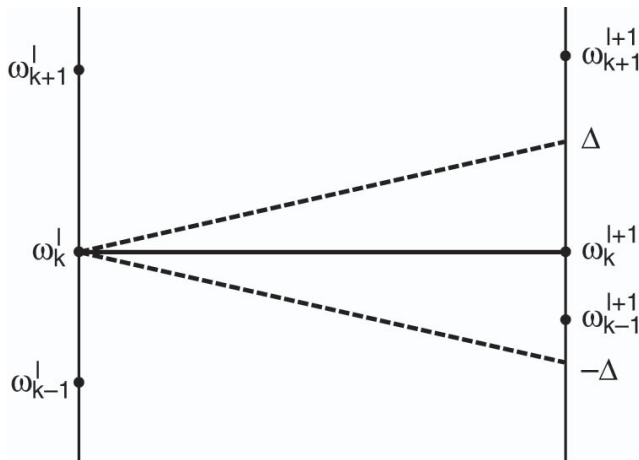
OUTLINE

- 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

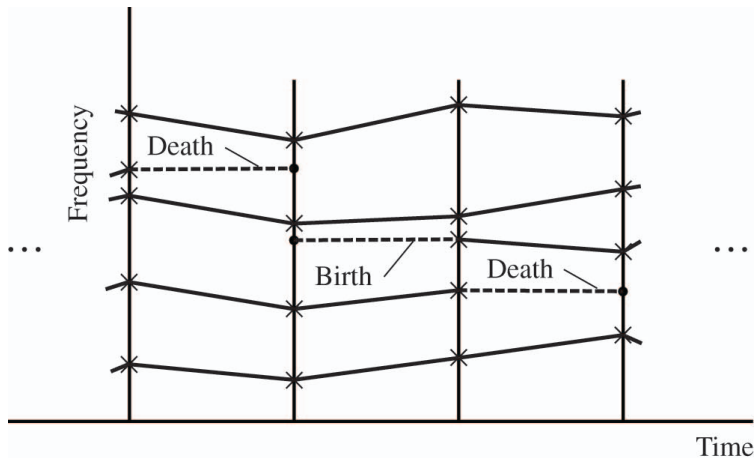
PROBLEM OF FREQUENCY MATCHING



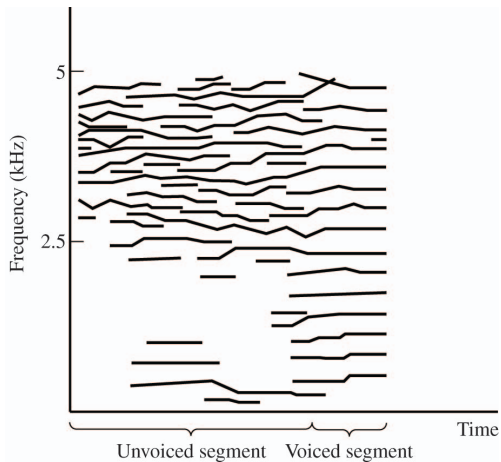
FRAME-TO-FRAME PEAK MATCHING



THE BIRTH/DEATH PROCESS



A BIRTH/DEATH PROCESS IN SPEECH

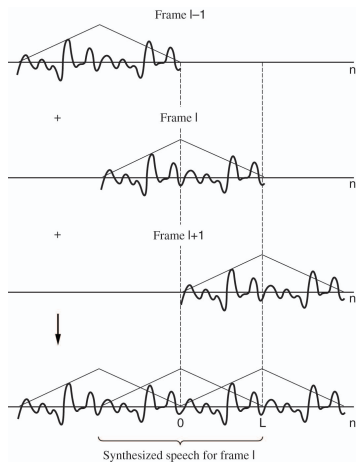


WHY NOT ...

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

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

A SIMPLE SOLUTION: OLA

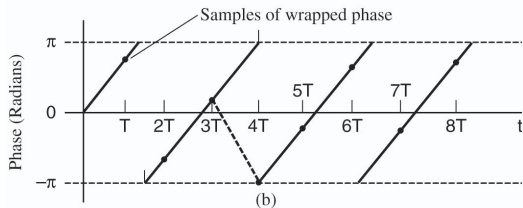
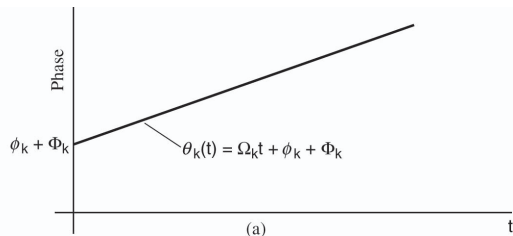


AMPLITUDE INTERPOLATION

Linear Interpolation:

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

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:

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

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:

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

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:

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

FOUR CONSTRAINTS FOR PHASE POLYNOMIAL

There are four constraints

$$\theta(0) = \theta'$$

$$\dot{\theta}(0) = \Omega'$$

$$\theta(T) = \theta'^{+1} + 2\pi M$$

$$\dot{\theta}(T) = \Omega'^{+1}$$

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

We need one more constraint!

FOUR CONSTRAINTS FOR PHASE POLYNOMIAL

There are four constraints

$$\theta(0) = \theta'$$

$$\dot{\theta}(0) = \Omega'$$

$$\theta(T) = \theta'^{+1} + 2\pi M$$

$$\dot{\theta}(T) = \Omega'^{+1}$$

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

We need one more constraint!

FOUR CONSTRAINTS FOR PHASE POLYNOMIAL

There are four constraints

$$\theta(0) = \theta'$$

$$\dot{\theta}(0) = \Omega'$$

$$\theta(T) = \theta'^{+1} + 2\pi M$$

$$\dot{\theta}(T) = \Omega'^{+1}$$

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

We need one more constraint!

FOUR CONSTRAINTS FOR PHASE POLYNOMIAL

There are four constraints

$$\theta(0) = \theta'$$

$$\dot{\theta}(0) = \Omega'$$

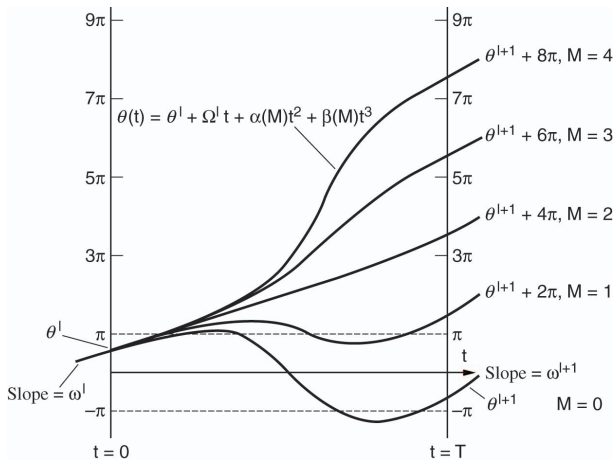
$$\theta(T) = \theta'^{+1} + 2\pi M$$

$$\dot{\theta}(T) = \Omega'^{+1}$$

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

We need one more constraint!

HOW TO CHOOSE M



ESTIMATING M

- Find M that minimizes the criterion:

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

- 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:

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

- 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:

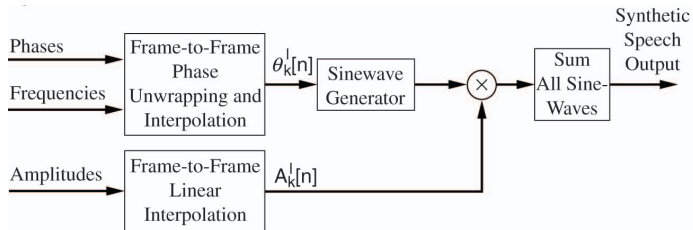
$$f(M) = \int_0^T [\ddot{\theta}(t; M)]^2 dt$$

- 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^*

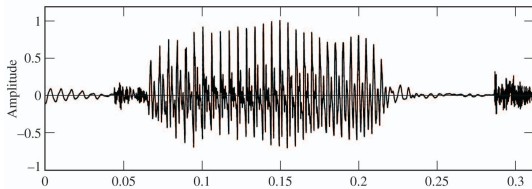
BLOCK DIAGRAM OF THE SYNTHESIS SYSTEM



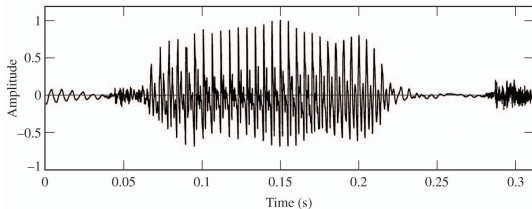
OUTLINE

- 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

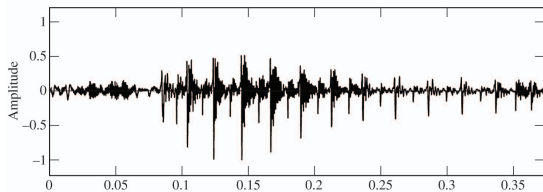


(a)

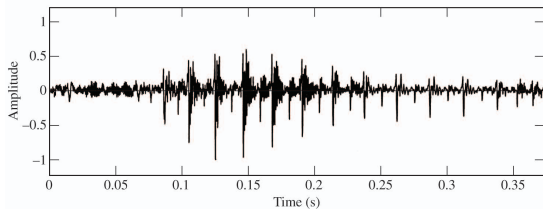


(b)

RECONSTRUCTION EXAMPLE

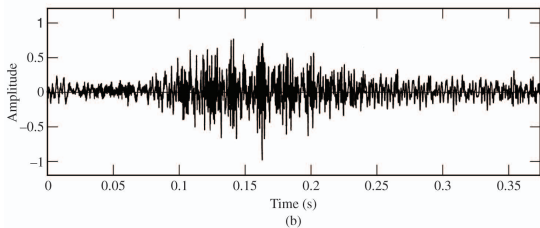
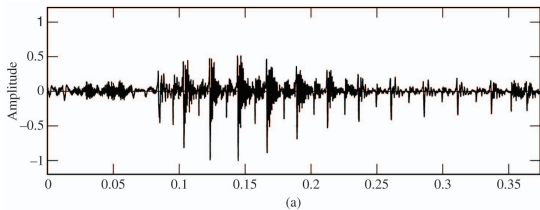


(a)



(b)













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:

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

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:

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

Then:

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

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

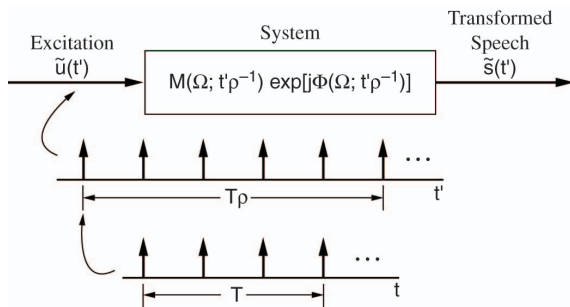
Let's t represent the original articulation rate and t' the transformed rate:

$$t' = \rho t$$

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'

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

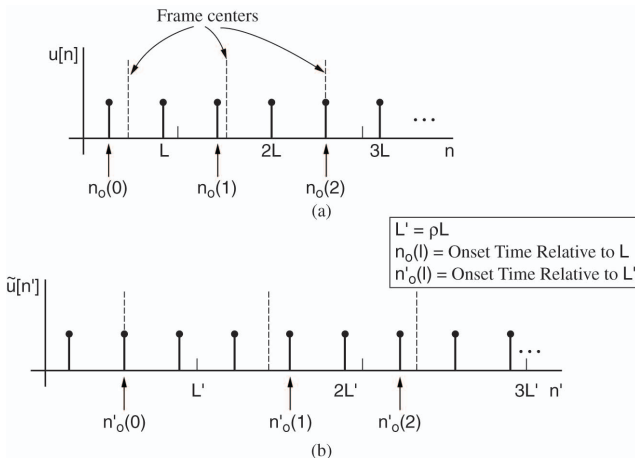
WAVEFORM IN t'

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

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:

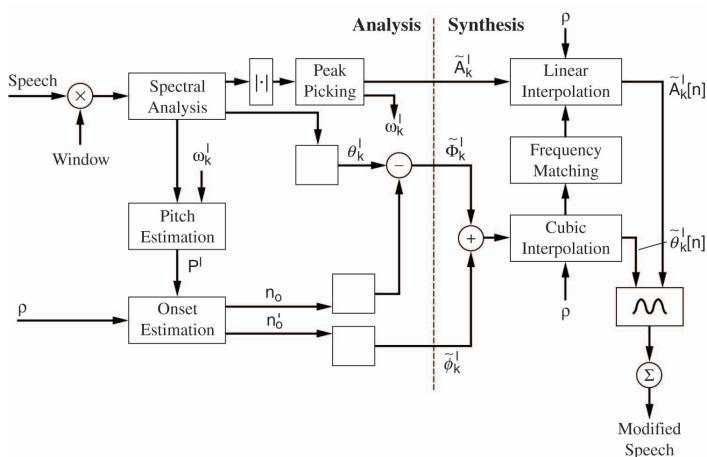
$$\tilde{\phi}_k^l = (n'_o(l) - lL')\omega_k^l$$

where $L' = \rho L$

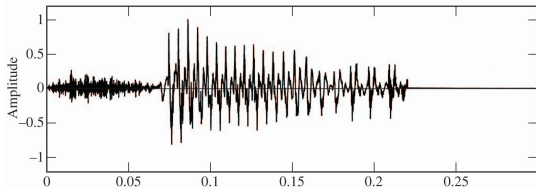
Synthesis is performed in the same way as if no modification is applied:

- Linear interpolation for amplitudes
- Cubic interpolation for phases

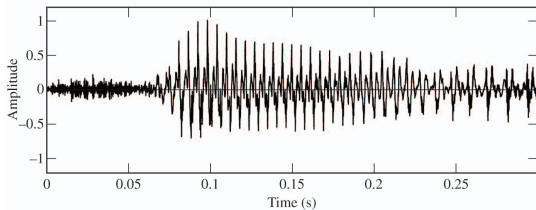
BLOCK DIAGRAM FOR ANALYSIS/SYNTHESIS FOR TIME-SCALE MODIFICATION



EXAMPLE OF TIME-SCALE MODIFICATION
















(a)



(b)

SOUND EXAMPLES

	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.

Τέλος Ενότητας



Ευρωπαϊκή Ένωση
Πρωτόκολλο Συνεργασίας



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΗ ΚΟΙΝΩΝΙΑ ΤΗΣ ΓΝΩΣΗΣ

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα αδειοδότησης

- Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση, Όχι Παράγωγο Έργο 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».

[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>



- Ως **Μη Εμπορική** ορίζεται η χρήση:
 - που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
 - που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
 - που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο
- Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Κρήτης, Στυλιανού Ιωάννης. «Ψηφιακή Επεξεργασία Φωνής. Ημιτονοειδής Ανάλυση και Τροποποίηση Φωνής». Έκδοση: 1.0. Ηράκλειο/Ρέθυμνο 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://www.csd.uoc.gr/~hy578>

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες

Εικόνες/σχήματα/διαγράμματα/φωτογραφίες που περιέχονται σε αυτό το αρχείο προέρχονται από το βιβλίο:

Τίτλος: *Discrete-time Speech Signal Processing: Principles and Practice*

Prentice-Hall signal processing series, ISSN 1050-2769

Συγγραφέας: Thomas F. Quatieri

Εκδότης: Prentice Hall PTR, 2002

ISBN: 013242942X, 9780132429429

Μέγεθος: 781 σελίδες

και αναπαράγονται μετά από άδεια του εκδότη.