Digital Speech Processing—Lecture 11

Modifications, Filter Bank Design Methods

1

Modifications to STFT

- modifications to short-time spectrum can be fixed (non-time varying) or time-varying
- fixed modification (no variability with n)

$$\hat{X}_n(e^{j\omega_k}) = X_n(e^{j\omega_k}) \cdot P(e^{j\omega_k})$$

- assume inverse DFT of $P(e^{j\omega_k})$ exists, then

$$p[n] = \frac{1}{N}\sum_{k=0}^{N-1} P(e^{j\omega_k})\, e^{j\omega_k n}$$

- where N is the number of frequencies at which $P(e^{j\omega_k})$ is evaluated

2

Fixed Modifications using FBS

- using FBS methods we get

$$\hat{y}[n] = \sum_{k=0}^{N-1} X_n(e^{j\omega_k}) P(e^{j\omega_k}) e^{j\omega_k n}$$

$$= \sum_{k=0}^{N-1}\left[\sum_{m=-\infty}^{\infty} w[n-m]\, x(m) e^{-j\omega_k m}\right] P(e^{j\omega_k}) e^{j\omega_k n}$$

$$= \left[\sum_{m=-\infty}^{\infty} w[n-m]\, x(m)\right] \sum_{k=0}^{N-1} P(e^{j\omega_k}) e^{j\omega_k (n-m)}$$

$$= \sum_{m=-\infty}^{\infty} w[n-m] x[m] N\, p[n-m]$$

$$\hat{y}[n] = Nx[n] * \left[w[n] \cdot p[n]\right]$$

3

Fixed Modifications using FBS

$$\hat{y}[n] = Nx[n] * \left[w[n] \cdot p[n]\right]$$

- $\hat{y}[n] = x[n]$ convolved with $\left[w[n] \cdot p[n]\right]$ => equivalent to linear filtering operation on $x[n]$
  --ideally want $\tilde{y}[n] = x[n] * p[n]$
  $\Rightarrow$ need duration of $p[n] \ll$ duration of $w[n]$
- $p[n]$ is a periodic sequence of period N (sampled in frequency at N frequencies)
- if $w[n]$ is longer than N => repetitive structure in $p[n] \cdot w[n] = h_p[n]$ => need to avoid long filters (IIR) but instead use RW so that

$$h_p[n] = p[n] \cdot w[n] \approx p[n], \quad 0 \le n \le N-1$$

4

Time-Varying Modifications

- represent time-varying modification as

$$\hat{X}_n(e^{j\omega_k}) = X_n(e^{j\omega_k}) \cdot P_n(e^{j\omega_k})$$

- with time-varying IR, $p_n(m)$, defined as

$$p_n[m] = \frac{1}{N}\sum_{k=0}^{N-1} P_n(e^{j\omega_k}) e^{j\omega_k m}$$

- solve for $\hat{y}[n]$ as

$$\hat{y}[n] = \sum_{k=0}^{N-1} X_n(e^{j\omega_k}) P_n(e^{j\omega_k}) e^{j\omega_k n}$$

$$= \sum_{k=0}^{N-1} e^{-j\omega_k n} \sum_{m=-\infty}^{\infty} x[n-m] w[m] e^{j\omega_k m} P_n(e^{j\omega_k}) e^{j\omega_k n}$$

5

Time-Varying Modifications

$$\hat{y}[n] = \sum_{k=0}^{N-1} e^{-j\omega_k n} \sum_{m=-\infty}^{\infty} x[n-m] w[m] e^{j\omega_k m} P_n(e^{j\omega_k}) e^{j\omega_k n}$$

$$= \sum_{m=-\infty}^{\infty} x[n-m) w[m] \sum_{k=0}^{N-1} P_n(e^{j\omega_k}) e^{j\omega_k m}$$

$$= \sum_{m=-\infty}^{\infty} x[n-m] w[m] N p_n[m]$$

$$\hat{y}[n] = N \sum_{m=-\infty}^{\infty} x[n-m]\left[w[m] p_n[m]\right] = Nx[n] * \left[p_n[m] w[m)\right]$$

- modified output is the window $w[m]$ weighted by the modification $p_n[m]$ and convolved with the input $x[n] \Rightarrow$ window 'time limits' effects of modifications and prevents smearing in time
- for FBS spectral modifications lead to convolving the original signal with a time-limited, window weighted version of the time response due to the modification

6

Fixed Modifications-OLA

- for a fixed modification we again have

$$\hat{X}_n(e^{j\omega_k}) = X_n(e^{j\omega_k})P(e^{j\omega_k})$$

- the basic OLA method gives

$$y[n] = \sum_{r=-\infty}^{\infty}\left[\frac{1}{N}\sum_{k=0}^{N-1} Y_r(e^{j\omega_k})e^{j\omega_k n}\right]$$

$$\hat{y}[n] = \sum_{r=-\infty}^{\infty}\frac{1}{N}\sum_{k=0}^{N-1} Y_r(e^{j\omega_k})P(e^{j\omega_k})e^{j\omega_k n}$$

$$= \frac{1}{N}\sum_{r=-\infty}^{\infty}\sum_{k=0}^{N-1}\sum_{\ell=-\infty}^{\infty} x[\ell]w[rR-\ell)e^{-j\omega_k \ell} P(e^{j\omega_k})e^{j\omega_k n}$$

$$= \frac{1}{N}\sum_{\ell=-\infty}^{\infty} x[\ell]\left[\sum_{k=0}^{N-1} P(e^{j\omega_k})e^{j\omega_k(n-\ell)}\right]\left[\sum_{r=-\infty}^{\infty} w[rR-\ell]\right]$$

7

Fixed Modifications-OLA

$$\hat{y}[n] = \frac{1}{N}\sum_{\ell=-\infty}^{\infty} x[\ell]\left[\sum_{k=0}^{N-1} P(e^{j\omega_k})e^{j\omega_k(n-\ell)}\right]\left[\sum_{r=-\infty}^{\infty} w[rR-\ell]\right]$$

$$\hat{y}[n] = \sum_{\ell=-\infty}^{\infty} x[\ell]\,p[n-\ell]W(e^{j0})/R = \left(\frac{1}{R}\right)W(e^{j0})\left[x[n]*p[n]\right]$$

- $\hat{y}[n]$ is the convolution of $x[n]$ with the time response of the spectral modification ($p[n]$)--with no window modifications
- need to use larger FFT sizes to account for (prevent) aliasing due to duration of $p[n] \Rightarrow N + N_0 - 1$ where N is the window size, N_0 is the duration of $p[n]$

8

Time Varying Modifications-OLA

- for the case of a time-varying modification, using OLA, we obtain

$$\hat{y}[n] = \sum_{r=-\infty}^{\infty}\frac{1}{N}\left[\sum_{k=0}^{N-1} Y_r(e^{j\omega_k})P_r(e^{j\omega_k})\right]e^{j\omega_k n}$$

- which after a great deal of manipulation can be put in the form

$$= \frac{1}{N}\sum_{\ell=-\infty}^{\infty} x[\ell]\sum_{r=-\infty}^{\infty} w(rR-\ell)\left[\sum_{k=0}^{N-1} P_r(e^{j\omega_k})e^{j\omega_k(n-\ell)}\right]$$

$$= \sum_{\ell=-\infty}^{\infty} x[\ell]\sum_{r=-\infty}^{\infty} w[rR-\ell]\,p_r[n-\ell]$$

9

Time Varying Modifications-OLA

- if we make the substitutions $q = n-\ell$, or $\ell = n-q$, we get

$$\hat{y}[n] = \sum_{q=-\infty}^{\infty} x[n-q]\sum_{r=-\infty}^{\infty} p_r[q]w[rR-n+q]$$

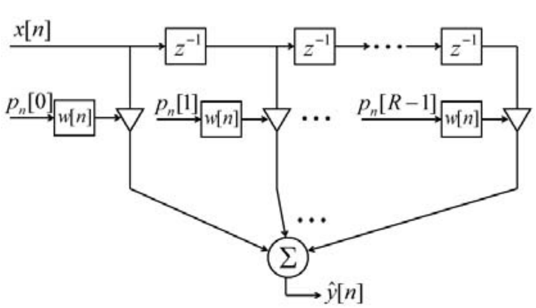
- if we let $\hat{p}[n-q,q] = \hat{p}[m,q] = \sum_{r=-\infty}^{\infty} p_r[q]w[rR-m]$

$$\hat{y}[n] = \sum_{q=-\infty}^{\infty} x[n-q]\hat{p}[n-q,q]$$

- $\hat{p}[m,q]$ is the convolution of $p_r[q]$ and $w[r]$ for all q => each coefficient of the time response due to the time-varying modification is smoothed by the window

10

Time Varying Modifications-OLA



11

Modifications with FBS and OLA

- OLA-any modification is first "**bandlimited by the window**" then acts as a true convolution on the input signal
- FBS-any modification is first "**time limited by the window**" and could change instantaneously

12

Additive Modifications-Quantization

- important to understand effects of additive, signal independent modifications to short-time spectrum as might occur for quantization

$$\hat{X}_n(e^{j\omega_k}) = X_n(e^{j\omega_k}) + E(e^{j\omega_k})$$

where the additive sequence is

$$e[n] = \sum_{k=0}^{N-1} E(e^{j\omega_k}) e^{j\omega_k n}$$

- for FBS method we get

$$\hat{y}[n] = \sum_{k=0}^{N-1}\left[X_n(e^{j\omega_k}) + E(e^{j\omega_k})\right] e^{j\omega_k n}$$
$$= y[n] + e[n]$$

13

Additive Modifications-Quantization

=> additive spectral modification yields additive signal

- for OLA method we get

$$\hat{y}[n] = \sum_{r=-\infty}^{\infty} \frac{1}{N} \sum_{k=0}^{N-1}\left(Y_r(e^{j\omega_k}) + E_r(e^{j\omega_k})\right) e^{j\omega_k n}$$
$$= y[n] + \sum_{r=-\infty}^{\infty}\left[\frac{1}{N}\sum_{k=0}^{N-1} E_r(e^{j\omega_k}) e^{j\omega_k n}\right]$$
$$= y[n] + \sum_{r=-\infty}^{\infty} e_r[n]$$

$\Rightarrow$ larger additive signal because of overlap between analysis frames

- for HW there is about 4 times larger additive signal

14

STFT Summary

1. $X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} w[\hat{n}-m]x[m]e^{-j\hat{\omega}m}$ -- STFT
2. $w[n]$ is the analysis window
3. $X_n(e^{j\hat{\omega}})$ can be considered output of bandpass filter, translated to 0 frequency or normal Fourier Transform of the sequence $x(m)w(n-m)$
4. use the sampling theorem to define sampling rate in time and frequency domain representations of the window => 2-4 times higher sampling rates than for stationary signal
5. two synthesis procedures evolved

FBS => $y[n] = \sum_{k=0}^{N-1} X_n(e^{j\omega_k})e^{j\omega_k n}$ -- sum together the bandpass filter outputs

OLA => $y[n] = \sum_{r=-\infty}^{\infty} \frac{1}{N}\sum_{k=0}^{N-1} Y_r(e^{j\omega_k})e^{j\omega_k n}$

- with $Y_r(e^{j\omega_k}) = X_{rR}(e^{j\omega_k})$ => windowed segments spaced by R samples in time are overlapped and added together
- FBS and OLA are dual methods

15

Digital Filter Bank Designs

16

Digital Filter Bank Designs

- composite frequency response approximates flat magnitude and linear phase
- showed ideal conditions for perfect reconstruction
  - choose same $w[n]$ for all channels of filter bank
  - $w[n]$ has zeros equally spaced at N sample intervals

17

Uniform Filter Bank Designs

- simple design procedure for uniform bank of filters
  - choose number of filters (this is determined by the desired frequency spacing)
  - design LPF with desired frequency resolution and with zeros appropriately spaced in time
- problems with this ideal case:
  - may want non-uniform filter spacing (model ear characteristics)
  - certain frequency bands (e.g., 0-100 Hz) often omitted (almost no essential speech information)
  - no design procedure for LPF allows simultaneous constraints on both time and frequency response => it is very difficult to design the desired filters

18

Modified Filter Bank Structure

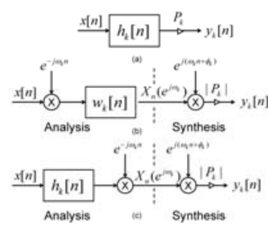
- each channel in FB scaled by complex gain $P_k = |P_k| e^{j\phi_k}$

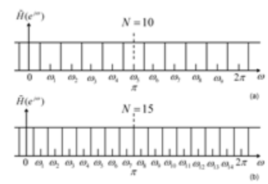
$$y[n] = \sum_{k=0}^{N-1} y_k[n] = \sum_{k=0}^{N-1} P_k X_n(e^{j\omega_k}) e^{j\omega_k n}$$

$$\tilde{h}[n] = \sum_{k=0}^{N-1} P_k h_k[n] = \sum_{k=0}^{N-1} |P_k| w_k[n] e^{j(\omega_k n + \phi_k)}$$

- P_k provides adjustment of gain and phase for each channel to make the overall composite filter bank better approximate the ideal filter bank

Modified Filter Bank Structure





FB Designs-Step 1

- Step 1--choose set of analysis frequencies, $\omega_k, 0 \le k \le N-1$
  assume symmetry in frequency => $\omega_{N-k} = 2\pi - \omega_k$
  assume symmetry in time => $w_k[n] = w_{N-k}[n]$
- giving $X_n(e^{j\omega_k}) = X_n^*(e^{j(2\pi-\omega_k)}) = X_n^*(e^{-j\omega_k})$
- can show that for N even we get

$$\tilde{h}[n] = P_0 w_0[n] + \sum_{k=1}^{N/2-1} 2|P_k| w_k[n] \cos(\omega_k n + \phi_k) + P_{N/2} w_{N/2}[n](-1)^n$$

- and for N odd we get

$$\tilde{h}[n] = P_0 w_0[n] + \sum_{k=1}^{(N-1)/2} 2|P_k| w_k[n] \cos(\omega_k n + \phi_k)$$

- given set of frequencies $\{\omega_k\}$, $0 \le k \le N-1$, we need to design set of lowpass filters (or analysis windows) $\{w_k[n]\}$ with the desired frequency resolution and the desired composite response

Effect of Omitted Channels

- channels are often omitted because they contain virtually no useful information (e.g., channel around ω=0 or around $\omega=\pi$)
- to see the effect of omitting channels from the filter bank, assume uniform frequency spacing, $\omega_k = 2\pi k / N$, and identical analysis windows, $w_k[n] = w[n]$

$$\tilde{h}[n] = w[n] \sum_{k=0}^{N-1} P_k e^{j2\pi kn/N}$$

- defining $p[n] = \sum_{k=0}^{N-1} P_k e^{j2\pi kn/N}$, we get

$\tilde{h}[n] = w[n]\, p[n]$ where $p[n]$ is periodic with period N.

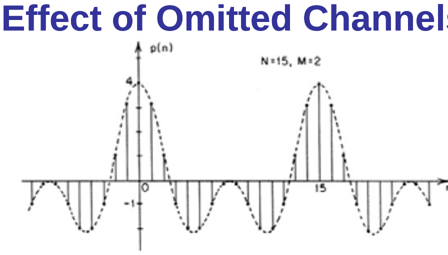
Effect of Omitted Channels

- to omit channels, we set $P_k = 0$ for those channels k that are omitted, e.g., omit channel 0 ($k = 0$), and all channels above $\omega_M = 2\pi M / N$, $P_k = 0$ for $k > M$.
- in this case we get

$$p[n] = \sum_{k=1}^{M} e^{j2\pi kn/N} + \sum_{k=N-M}^{N-1} e^{j2\pi kn/N} = \sum_{k=1}^{M} (e^{j2\pi kn/N} + e^{-j2\pi kn/N})$$

$$p[n] = \frac{\sin\left[\frac{\pi}{N}(2M+1)n\right]}{\sin\left[\frac{\pi}{N} n\right]} - 1$$

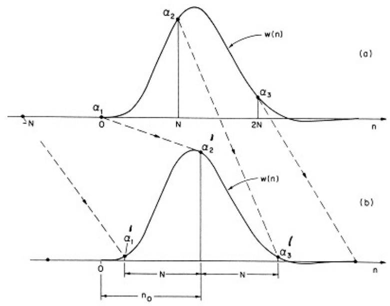
Effect of Omitted Channels



$$p[n] = \frac{\sin\left[\frac{\pi}{N}(2M+1)n\right]}{\sin\left[\frac{\pi}{N} n\right]} - 1$$

- $p[n]$ periodic with period N=15
- pulse amplitudes and widths depend on N and M
- for $M=(N-1)/2$ - all channels included-$p[n]$ is an impulse train spaced every N=15 samples

Effects of Imperfect LPF Design



$$\tilde{h}[n] = w[n]p[n]$$

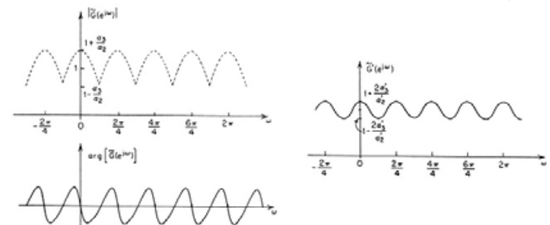
- assume we can delay $w[n]$ by arbitrary amount n_0 giving $w[n-n_0]$

Case 1: $\tilde{h}[n] = \alpha_2\delta[n-N] + \alpha_3\delta[n-2N]$

Case 2: $\tilde{h}(n) = \alpha_1'\delta[n-n_0] + \alpha_2'\delta[n-N-n_0] + \alpha_3'\delta[n-2N-n_0]$

25

Effects of Imperfect LPF Design



- to achieve a shift of n_0 samples, let

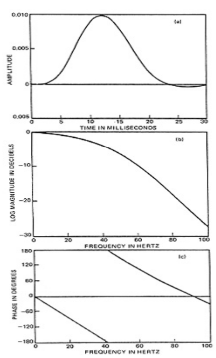
$$P_k = e^{-j2\pi k n_0/N}, \quad 0 \le k \le N-1$$

$$p[n] = \sum_{k=0}^{N-1} e^{j\frac{2\pi}{N}k(n-n_0)} = N\sum_{r=-\infty}^{\infty}\delta[n-n_0-rN]$$

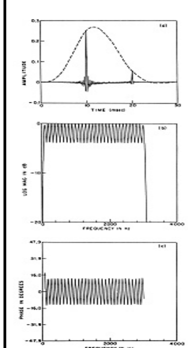
26

IIR Filter Banks

- assume $F_s = 10{,}000$ samples/sec
- assume uniform filter bank with spacing 100 Hz => N=10,000/100=100, $\omega_k = 2\pi k/100,\ 0 \le k \le M$
- analysis range of 100 Hz to 3000 Hz => M=30 channels
- Bessel (maximally flat delay) filters derived from 6th order analog filter
- 30 msec duration impulse response => 300 samples at 10,000 Hz SR
- 60 Hz filter bandwidth



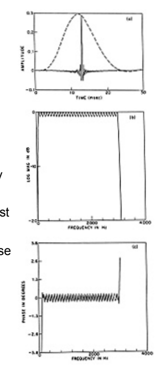
IIR Filter Banks



- for M=30, N=100, we get

$$p[n] = \frac{\sin(0.61\pi n)}{\sin(0.01\pi n)} - 1$$

- broadening of IR from 30 channels
- since IR duration >$2N$ samples, there are at least 2 strong peaks in $\tilde{h}[n]$
- significant ripple in composite frequency response (4 dB, 25 degrees)
- using delay to equalize amplitudes of first and third pulses, get better frequency response (0.8 dB ripple, 0.6 degrees phase deviation)



Summary of FB Design

1. determine **filter spacing** and **number of filters**
2. **design filter** to meet frequency selectivity for each channel
3. **evaluate *w(n)*** and choose delay to minimize ripple
4. evaluate **composite response** and iterate solution if response does not meet specs

29

FIR Filter Banks

- exact linear phase designs when $w(n) = w(L-1-n)$
- good design methods--windowing, optimal equiripple designs

Window Design Method

- design ideal LPF as

$$W_d(e^{j\omega}) = e^{-j\omega n_d} \quad |\omega| \le \omega_p$$
$$= 0 \quad otherwise$$

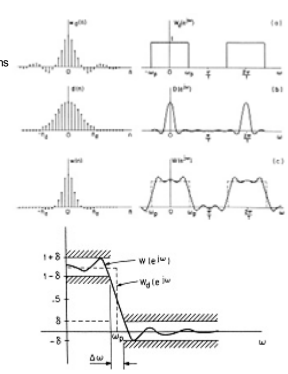
- where $n_d = (L-1)/2$ for L-point window
- can solve for ideal IR as

$$w_d[n] = \frac{1}{2\pi}\int_{-\omega_p}^{\omega_p} e^{-j\omega n_d}e^{j\omega n}d\omega = \frac{\sin\left[\omega_p(n-n_d)\right]}{\pi(n-n_d)}$$

- use finite duration window to truncate $w_d(n)$ giving

$$w[n] = w_d[n]d(n-n_d), \quad -n_d \le n \le n_d$$
$$W(e^{j\omega}) = W_d(e^{j\omega}) \otimes D(e^{j\omega})$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} W_d(e^{j\theta})D(e^{j(\omega-\theta)})d\theta$$

- this leads to a non-zero transition region between the passband and stopband and ripples in both bands



Window Design Properties

1. transition region, $\Delta\omega$, inversely proportional to L
2. $W(e^{j\omega})$ antisymmetric around ω_p
3. peak errors in passband and stopband nearly equal
4. approximation error greatest near ω_p

31

Kaiser Window Designs

- Kaiser window designs close to optimal

$$d[n] = \frac{I_0\left[\alpha\sqrt{1-(n/n_d)^2}\right]}{I_0(\alpha)}, \quad |n| \le n_d$$
$$= 0 \quad \textit{otherwise}$$

- $I_0(\alpha)$ is the zeroth order Bessel function of the first kind
- α is a tradeoff between transition width and peak approximation error

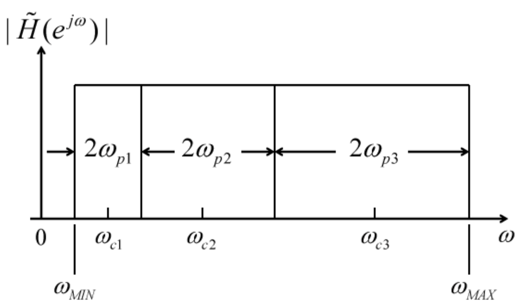
$$L = \frac{-20\log_{10}\delta - 7.95}{14.36\Delta f} + 1$$
$$\Delta f = \Delta\omega / (2\pi)$$

Procedure for design:

1. δ and Δf chosen => L
2. α computed from formula

32

Non-Uniform Filter Banks



33

Ideal Non-Uniform Designs

- filter bank bandpass filter response of form

$$h_k[n] = P_k w_k[n] e^{j\omega_k n}, \quad 0 \le k \le N-1$$

- consider designs using a common window, $d[n]$, for all channels
- the composite frequency response is

$$\tilde{H}(e^{j\omega}) = \sum_{k=0}^{N-1} P_k W_k(e^{j(\omega-\omega_k)})$$

- using the same design window for each channel gives

$$W_k(e^{j(\omega-\omega_k)}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_{dk}(e^{j(\theta-\omega_k)}) D(e^{j(\omega-\theta)}) d\theta$$

34

Ideal Non-Uniform Designs

- solving for $\tilde{H}(e^{j\omega})$ gives

$$\tilde{H}(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\sum_{k=0}^{N-1} P_k W_{dk}(e^{j(\theta-\omega_k)})\right] D(e^{j(\omega-\theta)}) d\theta$$

- letting

$$\tilde{H}_d(e^{j\omega}) = \sum_{k=0}^{N-1} P_k W_{dk}(e^{j(\omega-\omega_k)})$$

- gives

$$\tilde{H}(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{H}_d(e^{j\omega}) D(e^{j(\omega-\theta)}) d\theta$$

- where $\tilde{H}_d(e^{j\omega})$ is the desired composite frequency response

35

Ideal Non-Uniform Designs

- assume $P_k = 1, 0 \le k \le N-1$, and that the bandwidths and center frequencies of $W_{dk}(e^{j(\omega-\omega_k)})$ are such that the entire frequency range $-\pi \le \omega \le \pi$ is covered, giving

$$\tilde{H}_d(e^{j\omega}) = e^{-j\omega n_d}, \quad -\pi \le \omega \le \pi$$
$$\tilde{H}(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-j\omega n_d} D(e^{j(\omega-\theta)}) d\theta = d[n_d] e^{-j\omega n_d}$$

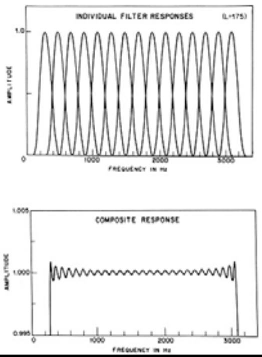
- giving a composite impulse response of

$$\tilde{h}[n] = d[n_d]\delta[n-n_d)$$

- this says that if the composite desired response is flat with linear phase then the actual composite response of the filter bank, using filters all designed with the same window, is also ideal--independent of how the center frequencies and bandwidths are distributed, and no matter what filter design window is used
- this says that perfect reconstruction is theoretically possible using FIR filters with an arbitrary distribution of center frequencies and bandwidths

36

FIR Filter Bank Examples



- $F_s = 9.6$ kHz
- want 15 filters to cover range from 200-3200 Hz
- lowpass cutoff frequency is

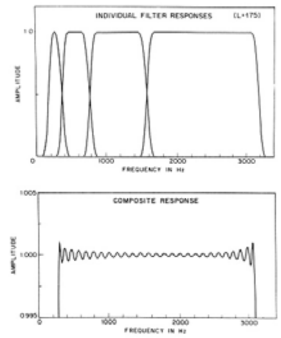
$$F_p = \frac{\omega_p}{2\pi T} = \frac{3200-200}{2(15)} = 100 \text{ Hz}$$

- center frequencies

$$F_k = \frac{\omega_k}{2\pi T} = 200k + 100, 1 \le k \le 15$$

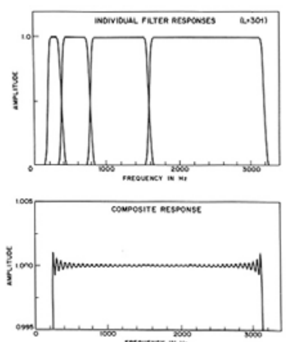
- desire 60 dB attentuation => α=5.65 with 200 Hz transition bands => $L = 175$

FIR Filter Bank Examples



- F_s=9.6 kHz
- want 4 octave band filters => bandwidth doubles with each successive filter => bandwidths of 200, 400, 800 and 1600 Hz, with center frequencies of 300, 600, 1200 and 2400 Hz
- use L=175

FIR Filter Bank Examples

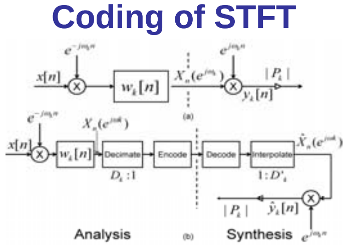


- want narrower transition bands
- use larger value of L--301

Applications of STFT

- vocoders => voice coders, code speech at rates much lower than waveform coders
- removal of additive noise
- de-reverberation
- speed-up and slow-down of speech for speed learning, aids for the handicapped

Coding of STFT

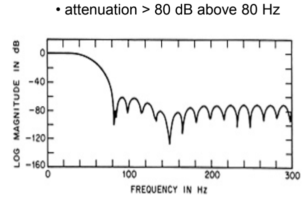


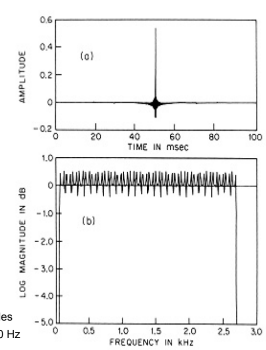
- elements of STFT
  1. set of $\{\omega_k\}$ chosen to cover frequency range of interest
  2. $w_k(n)$-set of lowpass analysis windows
  3. P_k - set of complex gains to make composite frequency response as close to ideal as possible

$\Rightarrow$ goal is to sample STFT at rates lower than $x(n)$

Coding of STFT

- attenuation > 80 dB above 80 Hz





- System example

$F_s = 12195$ samples/sec

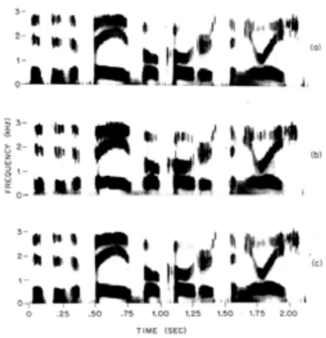
N=128

$\omega_k = 2\pi k / 128$, $F_k = 95.273k$ Hz

$w_k[n] = w[n]$ -- linear phase FIR filter, length 731 samples

$P_0 = 0$, $P_k = 0$, $28 < k < 100$ -- preserved band up to 2690 Hz

Coding of STFT



- part a is the original
- part b has no quantization, but there is fuzziness in the spectrogram due to echo in the composite IR => the system sounds reverberant
- P_k used to adjust phase and improve overall sound quality

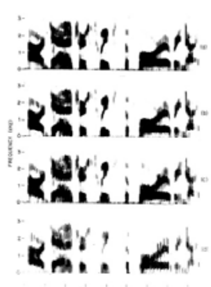
43

Coding of STFT

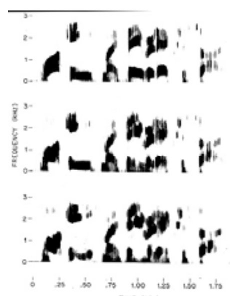
- quantizing STFT's
  - BR=sampling rate x bits/sample
  - channel sampling rate depends on bandwidth of lowpass filter, e.g., 80 Hz BW => SR ≥160/sec
  - using lower SR => aliasing in time
  - narrower bandwidth LP => reverberance because of "holes" in spectrum => need to increase number of channels

44

Coding of STFT

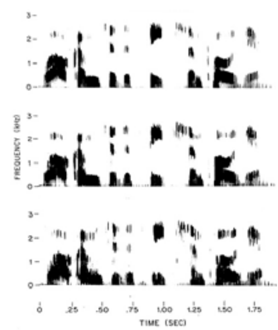


aliasing in time due to lower SR-160/100/80/60



narrower bandwidth LP => 80/53/36 Hz => formant distortions

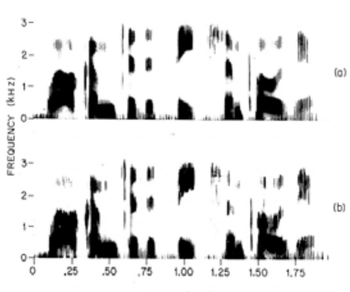
Coding of STFT



- channel coding and quantization
- use ADM with 1 bit/sample
- 28 channels-with real and imaginary outputs
- BR=2 x 28 x SR=56 SR
- for ADM SR=5-10 times Nyquist rate
- SR=500 Hz (part a), 375 Hz (part b), 250 Hz (part c)

46

Coding of STFT



- non-uniform coding and quantization
- 28 channels
- 100/sec SR (gives small amount of aliasing)
- coding log magnitude and phase using 3 bits for log magnitude and 4 bits for phase for channels 1-10; and 2 bits for log magnitude and 3 bits for phase for channels 11-28
- total rate of 16 Kbps

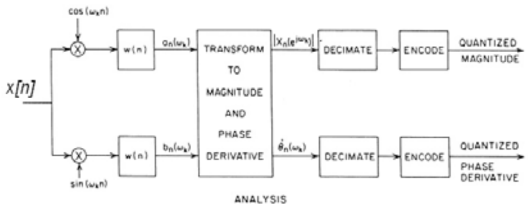
47

STFT Coding Summary

- sample channels at sufficient rate, quantize to 12 bits/sample => perceptually perfect reproduction of *x(n)* possible with bit rates of 100 Kbps
- reduced bit rate => quantize more coarsely, reduce sampling rate => 16 Kbps rates possible
- remove redundancy of the speech signal by higher order modeling prior to STFT operations
- cannot use SNR to evaluate STFT since degradations are perceived as modifications of speech quality and intelligibility-not additive noise => spectrograms, subjective quality testing

48

Phase Vocoder



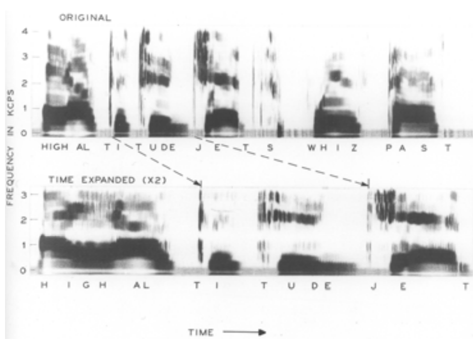
used for speed-up and slow-down of speech

intro | phase vocoder | channel vocoder | speeded-up speech | slowed down speech

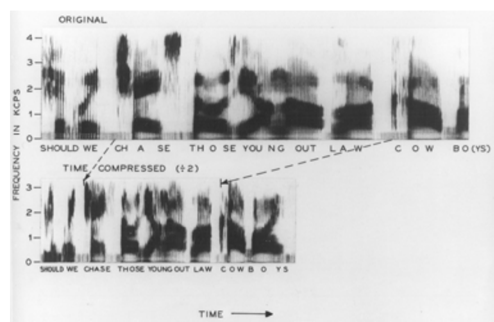
Additional Examples of Rate Changes in Speech

- Male Speaker
  - Original rate
  - Speeded up
  - Speeded up more
  - Slowed down
  - Slowed down more
- Female Speaker
  - Original rate
  - Speeded up
  - Speeded up more
  - Slowed down
  - Slowed down more

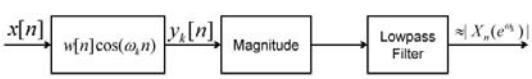
Phase Vocoder Time Expanded



Phase Vocoder Time Compressed



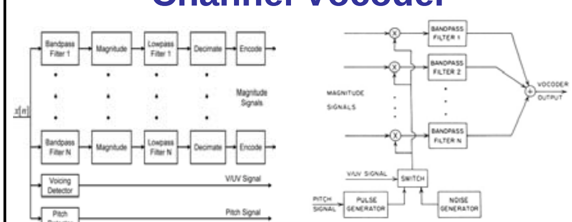
Channel Vocoder



- interpret STFT so that each channel can be thought of as a a bandpass filter with center frequency ω_k
- magnitude of STFT can be approximated by envelope detection on the BPF output (FWR and LPF)
- ***analyzer***-bank of channels; need excitation info (the phase component) => V/UV detector, pitch detector
- ***synthesizer***-channel signal control channel amplitude; excitation signals control detailed structure of output for a given channel; V/UV choice of excitation source

=> highly reverberant speech because of total lack of control of composite filter bank response

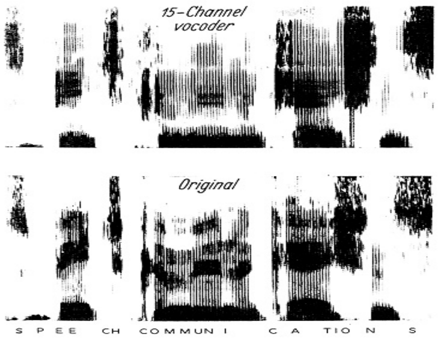
Channel Vocoder



- 1200-9600 bps
- 600 bps for pitch and V/UV
- easy to modify pitch, timing

Channel Vocoder-2.4 kbps

Channel Vocoder



55