EE627 - Speech Signal Processing
Lecture 11/12 : Cepstral Analysis Techniques for Speech Recognition

R. Hegde
Dept. of Electrical Engg.
IIT Kanpur

Outline

Signal Combinations

Signals can be combined in 2 ways

i. Linear (Addition)

ii. Convolved (convolution)

How do we separate the individual signals??

i. $x(n) = x_1(n) + w(n)$

Then $X(\omega) = X_1(\omega) + W(\omega) \approx X_1(\omega)$ ← Filter

ii) $x(n) = x_1(n) * w(n)$ then $x_1(n) = ??$

Deconvolution

Cepstrum : i. deconvolve individual components

ii) linearly combine the component signals

$s(n) = e(n) * \theta(n)$

Say $\mathcal{H}$ will follow

$$\mathcal{H}(s(n)) = \mathcal{H}\{e(n) * \theta(n)\} = \mathcal{H}(e(n)) * \mathcal{H}(\theta(n))$$

If $\mathcal{H}(e(n)) \simeq \delta(n)$ && $\mathcal{H}(\theta(n)) \simeq \theta(n)$

Then $e(n)$ and $\theta(n)$ can be "deconvolved"

This leads to Homomorphic systems

Complex Cepstrum

Complex Cepstrum : Retains Phase
Real Cepstrum : Discards Phase

Note: RC = CC under assumption of minimum phase

RC	CC
* Discards phase	* Retains Phase
* Easy to compute	* Difficult to Comp.
* Speech Analysis Recognition	* Vocoders, speech coding

Notations Used

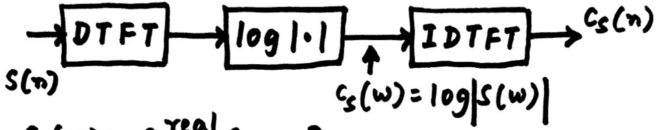
Name	Notation for Signal $x(n)$	Relationship
Complex cepstrum (CC)	$\gamma_x(n)$	
Real cepstrum (RC)	$c_x(n)$	$c_x(n) = \gamma_{x,\text{even}}(n)$
Short-term complex cepstrum (stCC) frame ending at m	$\gamma_x(n;m)$	
Short-term real cepstrum (stRC) frame ending at m	$c_x(n;m)$	$c_x(n;m) = \gamma_{x,\text{even}}(n;m)$

Real Cepstrum

$$c_s(n) = \mathcal{F}^{-1}\{\log|\mathcal{F}[s(n)]|\}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|S(\omega)|\, e^{j\omega n}\, d\omega$$

Note: RC is an even sequence on 'n'



$$c_s(\omega) = Q_*^{real}\{s(n)\} = \log S(\omega)$$

* → indicates deconvolution

real → indicates log of a real number

Real Cepstrum - Contd.

$$C_s(w) = \log|S(w)| = \log|E(w)\,\theta(w)|$$
$$= \log|E(w)| + \log|\theta(w)|$$

$$C_s(w) = C_e(w) + C_\theta(w)$$

Since $C_s(w)$ is periodic its FS

$$d_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} C_s(w)\, e^{-jwn}\, dw$$

More Importantly: $c_s(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} C_s(w)\, e^{jwn}\, dw$

But $C_s(w)$ is REAL

Real Cepstrum for ASR

In Speech Recognition we compute

$$c_s(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} C_s(w) \cos(wn)\, dw$$

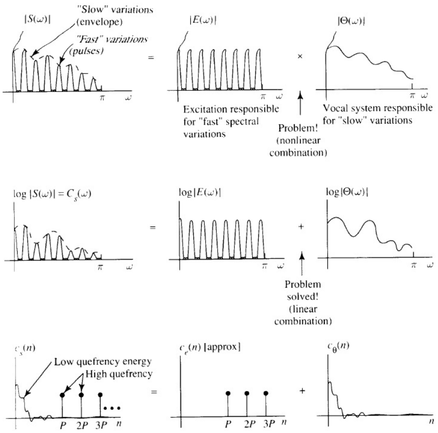
$$c_s(n) = \frac{1}{2\pi} \int_{0}^{\pi} C_s(w) \cos(wn)\, dw$$

Note: IDFT is replaced by DCT in practice

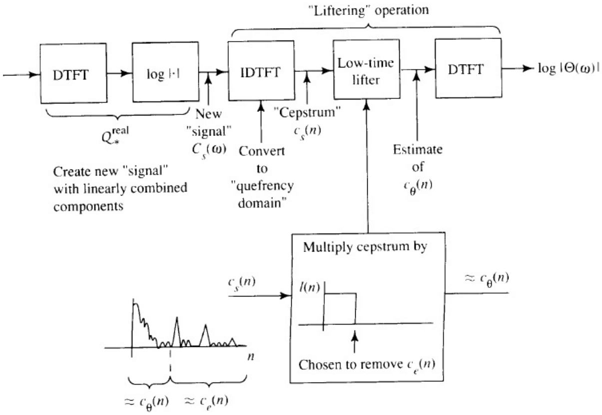
$$\boxed{c_s(n) = c_e(n) + c_\theta(n)}$$

Bottom Line

Linear Combination - Illustration



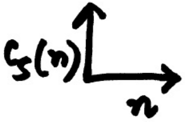
Liftering


"Liftering" operation
DTFT
log |·|
IDTFT
Low-time lifter
DTFT
log |Θ(ω)|
Q_*^{real}
New "signal" $C_s(\omega)$
"Cepstrum" $c_s(n)$
Convert to "quefrency domain"
Estimate of $c_\theta(n)$
Create new "signal" with linearly combined components
Multiply cepstrum by
$l(n)$
Chosen to remove $c_e(n)$
$c_s(n)$
$\approx c_\theta(n)$
n
$\approx c_\theta(n)$ $\approx c_e(n)$

Cepstral Terminology

$c_s(n)$ is the cepstrum

$C_s(\omega)$ is the Spectrum



Spectral	Cepstral
Frequency	Quefrency
Harmonic	Rahmonic
magnitude	Gamnitude
Phase	Saphe
Filter	Lifter
Fundamental	Mundafental

Short Term Cepstrum

We KT

$$c_s(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \{\log|S(\omega)|\} e^{j\omega n} d\omega$$

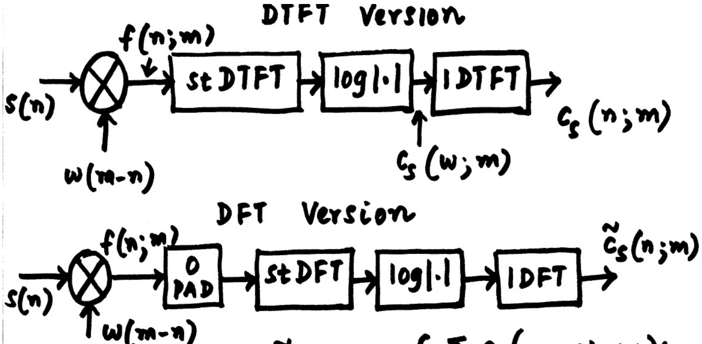
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|\sum_{\forall l} s(l) e^{-j\omega l}\right| e^{j\omega n} d\omega, \forall n$$

For Length 'N' frame ending at time 'm'

$$f(n;m) = s(n)\,\omega(m-n)$$

$$\therefore c_s(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left\{\log\left|\sum_{l=m-N+1}^{m} f(l;m) e^{-j\omega l}\right|\right\} e^{j\omega n} d\omega$$

Short Term Cepstrum - Contd.



$$\tilde{c}_s(n;m) = \begin{cases} \sum_{\forall q} c_s(n+qN;m); & n = 0,1,\cdots,N-1 \\ 0; & \text{otherwise} \end{cases}$$

Note that $\tilde{c}_s(n;m)$ is a periodic version of $c_s(n;m)$

Short Term Cepstrum and Windowing

$s(n) = e(n) * \theta(n)$ $\quad e(n) \rightarrow \boxed{\theta(n)} \rightarrow s(n)$

Consider a speech frame of length 'N' ending at 'm'

$$f_s(n;m) = s(n)\,w(m-n)$$

$$f_s(n;m) = [e(n) * \theta(n)]\,w(m-n)$$

↙ just shows f is derived from s(n)

Can we move the window w(m-n) inside the *

Short Term Cepstrum and Windowing

If we do that then

$$f_s(n;m) \approx e(n)w(m-n) * \theta(n)$$

$$f_s(n;m) = f_e(n;m) * \theta(n)$$

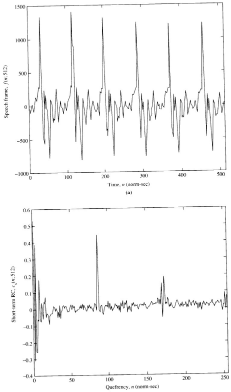
where $f_e(n;m)$ is a frame of $e(n)$ windowed and ending at m

Alternately: $c_s(n;m) = c_e(n;m) + c_\theta(n)$

$\therefore$ $c_e(n;m)$ will appear in $c_s(n;m)$ as a pulse train added to $c_\theta(n)$ the RC

$c_\theta(n)$ decays very quickly wrt 'P' the pitch period

Speech Cepstrum



Cepstrum and Pitch

From previous Fig/Eqn.

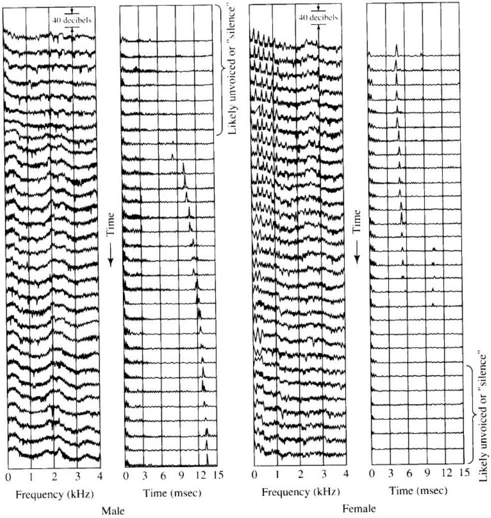
$$c_s(n;m) \approx \begin{cases} c_e(0;m) + c_\theta(0); & n = 0 \\ c_\theta(n); & 0 < n < P \\ c_e(n;m); & n \geq P \end{cases}$$

$$c_e(0;m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |E(\omega;m)| \, d\omega$$

$$c_\theta(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |\theta(\omega)| \, d\omega$$

Locate initial peak in $c_s(n;m)$ which is well separated from $\theta(n)$

Cepstrum and Pitch (Noll)



Liftering

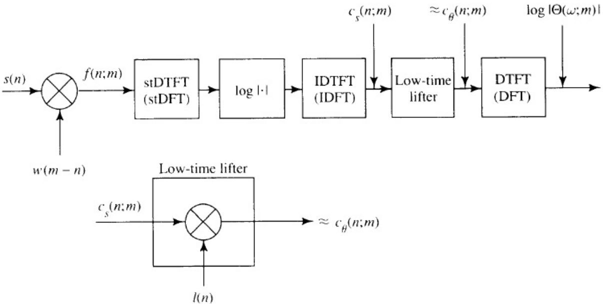
1. Compute the stRC of the speech $c_s(n;m)$ as above.
2. Multiply $c_s(n;m)$ by a "low-time" window, $l(n)$ to select $c_\theta(n)$:

$$c_\theta(n) \approx c_s(n;m)l(n). \tag{6.42}$$

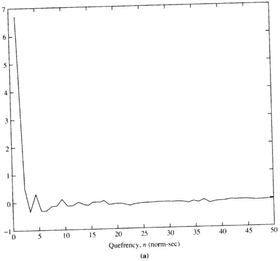
(Note that the lifter $l(n)$ should theoretically be an even function of n, or even symmetric about time $(N-1)/2$ if an N-point DFT is used in the next step.)

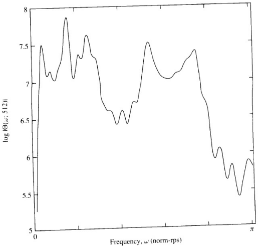
3. To get the estimate of $\log|\Theta(\omega)|$, DTFT (DFT) the estimate of $c_\theta(n)$.

Liftering - Contd.



Liftering - Contd.





Complex Cepstrum

RC not invertible ; NO ROUND TRIP!!

CC: $\hat{x}_s(n) = \mathcal{F}^{-1}\{\log \mathcal{F}\{s(n)\}\}$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log S(\omega)\, e^{j\omega n}\, d\omega$$

log is now the complex logarithm

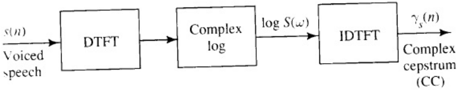
$$\log(z) = \log|z| + j\arg\{z\}$$

$$\log S(\omega) = \log|S(\omega)| + j\arg\{S(\omega)\}$$

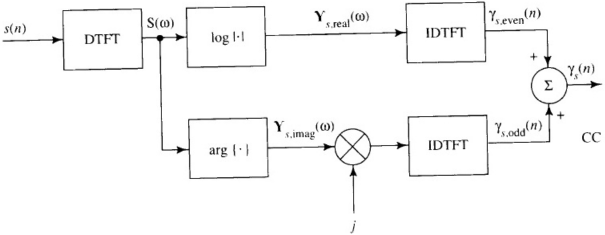
$\arg\{S(\omega)\}$ is the Phase (unwrapped)

But add multiples of 2π to make $\arg\{S(\omega)\}$ an odd fn. of ω

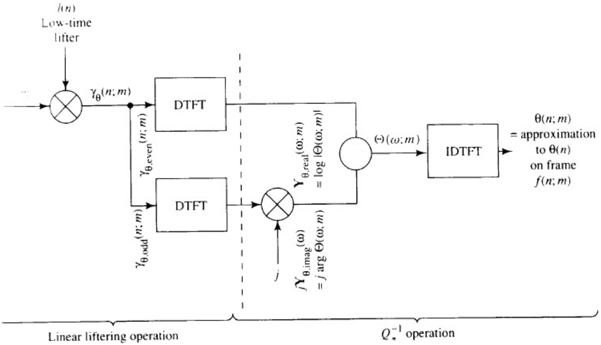
Complex Cepstrum - Contd.



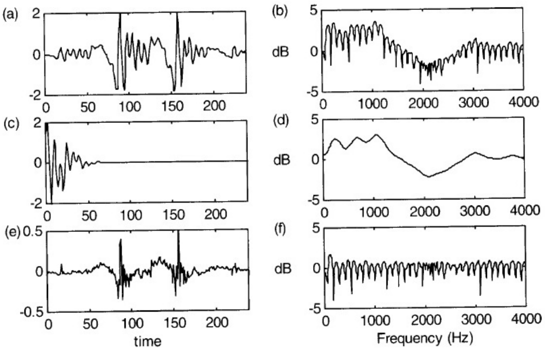
Complex Cepstrum - Contd.



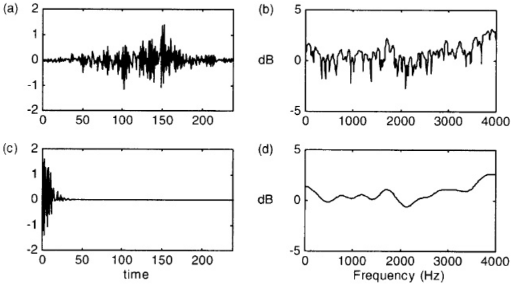
Complex Cepstrum - Contd.



Cepstral Smoothing - Some Points



Cepstral Smoothing - Some Points



References

Deller et. al.
Discrete Time Processing of Speech Signals.
Wiley.

Rabiner and Juang.
Fundamentals of Speech Recognition.
Prentice Hall.