به نام خدا

Digital Speech Processing

گروه هوش مصنوعی، دانشکده مهندسی کامپیوتر

گفتار پردازی رقمی

استخراج ویژگی گفتار

حمیدرضا برادران کاشانی

پاییز ۱۴۰۱

سرفصل مطالب

- استخراج ویژگی در حوزه فرکانس
- استخراج ویژگی در حوزه زمان

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس

❖ تبدیل فوریه زمان گسسته (DTFT)

❖ سیگنال زمان گسسته **x[n]** را در نظر بگیرید، **DTFT** آن عبارتست از:

$$X(e^{jw}) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-jwn}, \quad w \in [-\pi, +\pi)$$

❖ **DTFT** یک سیگنال گسسته، خود یک سیگنال پیوسته در فرکانس (پیوسته) ***w*** است.

❖ فرکانس ***w*** بطور پیوسته بین **pi–** و **pi** تغییر می کند، بنابراین برای محاسبه **DTFT** بایستی محاسبات در تعداد بینهایت فرکانس صورت پذیرد که قطعا عملیاتی نمی باشد.

❖ همچنین کامپیوترها بایستی با سیگنال هایی با تعداد نمونه های محدود کار کنند نه تعداد بینهایت نمونه.

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس

❖ تبدیل فوریه گسسته (DFT)

❖ سیگنال زمان گسسته **x[n]** را در نظر بگیرید که شامل ***N*** نمونه است؛

$$x[n]=\{x[0],x[1],\ldots,x[N-1]\}$$

❖ **DFT** سیگنال گسسته **x[n]**، یک سیگنال گسسته تعریف شده در فرکانس (گسسته) w_k است:

❖ به عبارتی **DFT** فرم نمونه برداری شده از سیگنال پیوسته **DTFT** است .

$$X[k]=X(e^{jw})\Big|_{w_k=k.\frac{2\pi}{N}},\quad k=0,1,\ldots,N-1,$$

$$X[k]=\left|X[k]\right|e^{j\measuredangle X[k]}$$

$$\left.\begin{aligned} w_k &= 2\pi\frac{f}{F_s}\\ w_k &= 2\pi\frac{k}{N}\end{aligned}\right\}\rightarrow f=k.\frac{F_s}{N}$$

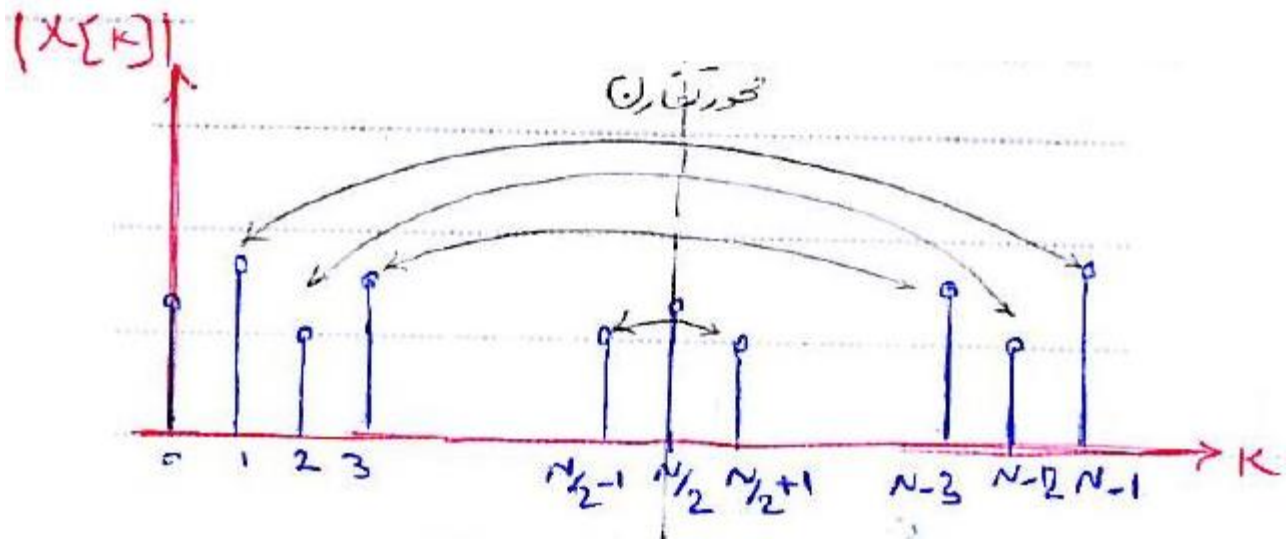
Frequency resolution

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس

❖ تبدیل فوریه گسسته (DFT)

- ❖ حاصل DFT تعداد N عدد مختلط است که در محدوده $k \in [0, N-1]$ است که این محدوده فرکانسی نیز متناظر با $(0, F_s)$ است.
- ❖ اگر سیگنال x[n] حقیقی باشد، نمودار بصورت متقارن حول N/2 است. بنابراین تنها کافی است که N/2+1 نمونه از دامنه DFT را استفاده کنیم.



Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس

❖ تبدیل سریع فوریه (FFT)

- ❖ هدف از الگوریتم های **FFT** کاهش تعداد ضرب و جمع در مقایسه با تبدیل فوریه گسسته **DFT** است.
- ❖ مرتبه محاسباتی **DFT** از مرتبه $O(N^2)$ است که در آن ***N*** طول پنجره آنالیز است.
- ❖ در واقع تبدیل فوریه سریع از طریق تجزیه ماتریس **DFT** به حاصلضرب ماتریس های تنک (**sparse**) که در آن‌ها اکثر داریه‌های ماتریس صفر هستند، محاسبات را تسریع می‌بخشد.
- ❖ الگوریتم کولی- توکی به صورت بازگشتی تبدیل فوریه گسسته را به مسایل کوچک‌تر می‌شکند و زمان مورد نیاز برای انجام محاسبات را به مقدار قابل توجهی کاهش می‌دهد.
- ❖ مرتبه محاسباتی **FFT** از مرتبه $O(N \log_2 N)$ است.

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس

❖ تبدیل فوریه گسسته زمان کوتاه (stDFT یا STFT)

- ❖ تبدیل STFT برای سیگنال هایی غیر ایستا (non-stationary) استفاده می شود.
- ❖ سیگنال غیر ایستا یعنی سیگنالی که مشخصات طیفی آن در طول زمان تغییر می کند، مثلا تغییر محتوای فرکانسی در طول زمان مثل صدای آمبولانس یا گفتار انسان یا ...
- ❖ محاسبه DFT برای یک سیگنال طولانی که با زمان متغیر است، اطلاعات معناداری را نتیجه نداده و مشخص نمی کند که در زمان های مختلف چه تغییرات طیفی رخ داده است.
- ❖ برای حل این مساله، سیگنال به تعدادی فریم کوتاه تقسیم می شود (بین ۲۰ تا ۳۰ میلی ثانیه) و سیگنال گفتار در این بازه ایستا یا شبه ایستا در نظر گرفته می شود.
- ❖ حال DFT بر روی این فریم های کوتاه مدت اعمال شده که به آن STFT می گویند.

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس

❖ تبدیل فوریه گسسته زمان کوتاه (stDFT یا STFT)

❖ سیگنال $x[n]$ را در نظر بگیرید. فریم l سیگنال را با $x_l[n]$ بیان کنیم که طولش N نمونه است.

$$x_l[n] = x[n]\, w[n-(l-1)H], \quad l = 1, 2, \ldots, M \qquad (1)$$

$$x_l[n] = x[n+(l-1)H]\, w[n], \quad l = 1, 2, \ldots, M \qquad (2)$$

$$M = \left\lfloor \frac{L-N}{H} \right\rfloor$$

❖ l شماره فریم و H شیفت فریم هستند. فرضا با همپوشانی فریم ۵۰٪ داریم: $H=N/2$

❖ M تعداد کل فریم ها است و L طول کل سیگنال بر حسب تعداد نمونه ها است.

❖ در رابطه شماره (۱) سیگنال ثابت و پنجره به سمت جلو شیفت داده می شود. در رابطه شماره (۲) پنجره ثابت و سیگنال به سمت عقب حرکت می کند. از لحاظ نتیجه فریم بندی یکسان است.

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس

❖ تبدیل فوریه گسسته زمان کوتاه (stDFT یا STFT)

❖ رابطه stDFT یا بطور خلاصه STFT:

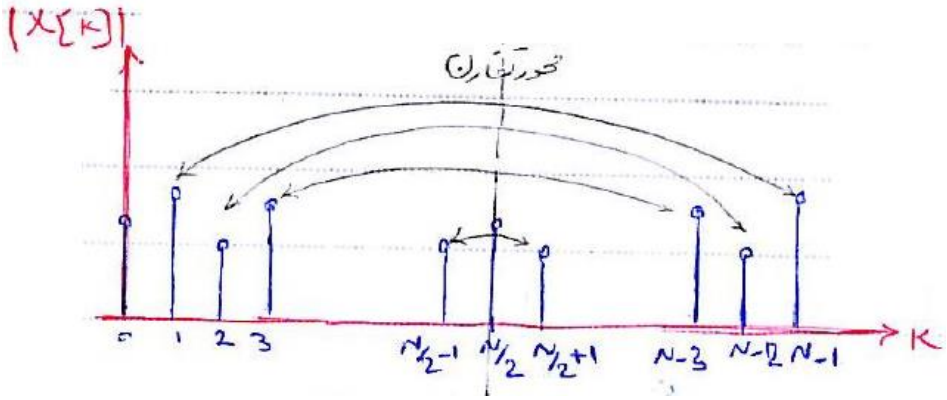
$$X_l[k] = DFT\{x_l[n]\}$$
$$= \sum_{n=0}^{N-1} x[n]\, w[n-(l-1)H]\, e^{-j\frac{2k\pi}{N}n}, \quad 0 \le k \le N-1$$

❖ k اندیس یا بین فرکانسی است و متغیری گسسته.

❖ در پیاده سازی جهت افزایش سرعت به جای DFT از FFT استفاده می شود.

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس



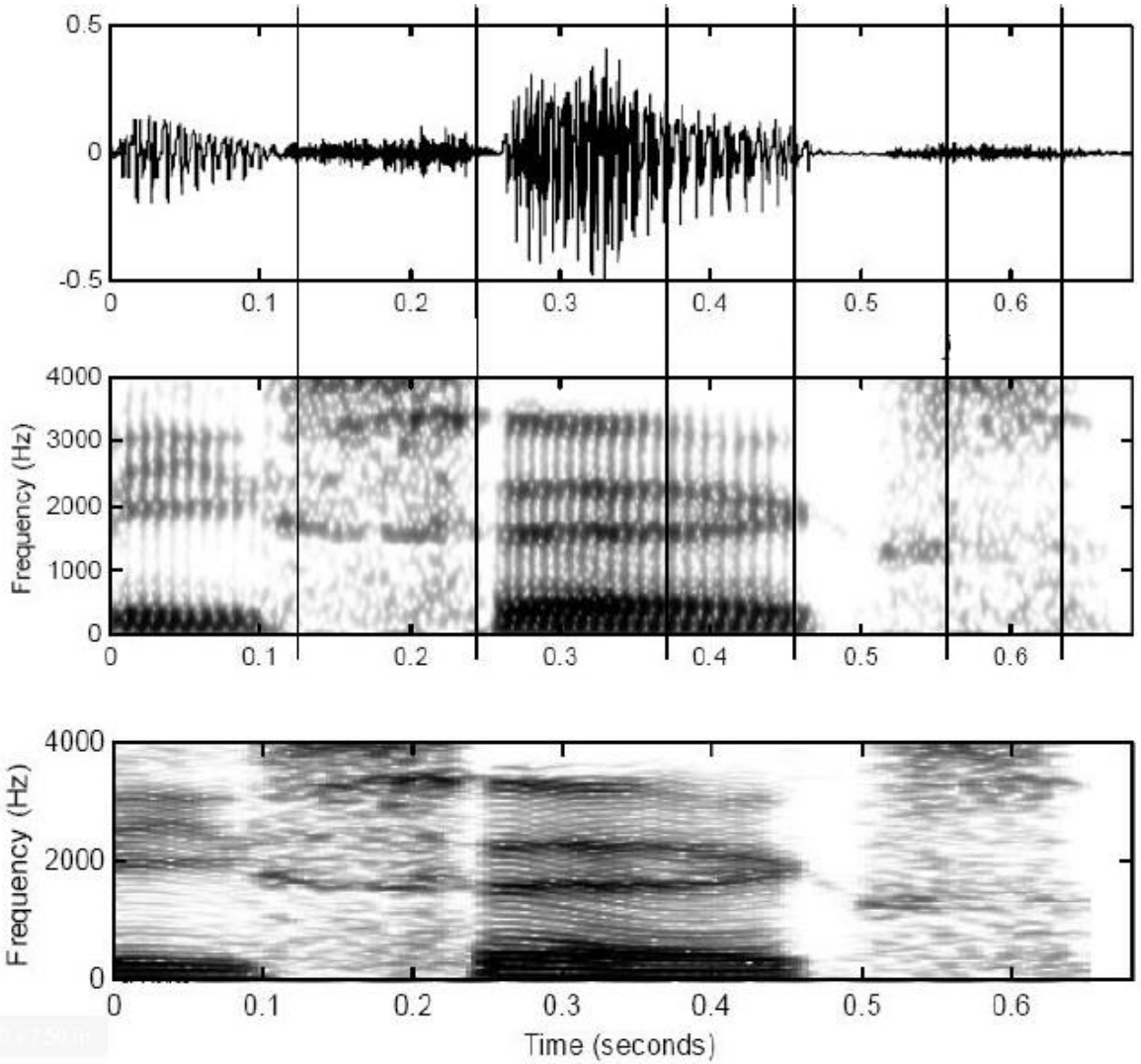
❖ برای رسم اسپکتروگرام کافی است $X_l[k]$ را تبدیل به مقیاس **dB** کنیم:

$$X_l^{dB}[k] = 20\log_{10}\left(\varepsilon + \left|X_l[k]\right|\right)$$

$$Spec = \begin{bmatrix} X_1^{dB}[N/2] & X_2^{dB}[N/2] & \dots & X_m^{dB}[N/2] & \dots & X_M^{dB}[N/2] \\ X_1^{dB}[N/2-1] & X_2^{dB}[N/2-1] & \dots & X_m^{dB}[N/2-1] & \dots & X_M^{dB}[N/2-1] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_1^{dB}[k] & X_2^{dB}[k] & \dots & X_m^{dB}[k] & \dots & X_M^{dB}[k] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_1^{dB}[0] & X_2^{dB}[0] & \dots & X_m^{dB}[0] & \dots & X_M^{dB}[0] \end{bmatrix}$$

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه فرکانس



- طیف باند پهن (wideband)
  - پنجره زمانی کوتاه مثلا ۱۰ میلی ثانیه یا کمتر
  - تفکیک یا رزولوشن زمانی خوب
  - تفکیک فرکانسی کم
- طیف باند باریک (narrowband)
  - پنجره زمانی بلند مثلا ۲۰ میلی ثانیه یا بیشتر
  - تفکیک یا رزولوشن زمانی کم
  - تفکیک فرکانسی زیاد

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

❖ روش شناخته شده در پردازش گفتار و پرکاربرد در حوزه های کدینگ گفتار، سنتز گفتار، شناسایی گفتار و گوینده و ...

❖ ایده اصلی LPC:

❖ یک نمونه از سیگنال گفتار در زمان مشخص را می توان توسط ترکیب خطی از تعدادی از نمونه های قبلی آن تقریب زد.

$$\hat{s}[n] = -\sum_{k=1}^{P} a_k \, s[n-k]$$

❖ P: مرتبه تحلیل LPC است

❖ $\{a_k\}$ ها ضرایب تحلیل LPC است.

❖ هدف: بدست آوردن ضرایب تحلیل LPC است.

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

❖ راه حل کلی بدست آوردن ضرایب LPC:

❖ کمینه سازی مجموع مربعات خطای تخمین : $e\ n]$

سیگنال تحریک یا Excitation
سیگنال باقی مانده یا Residual
خطای پیشگویی یا Prediction error

$$e[n] = s[n] - \hat{s}[n] = s[n] + \sum_{k=1}^{P} a_k\, s[n-k]$$

❖مبنای کار LPC:

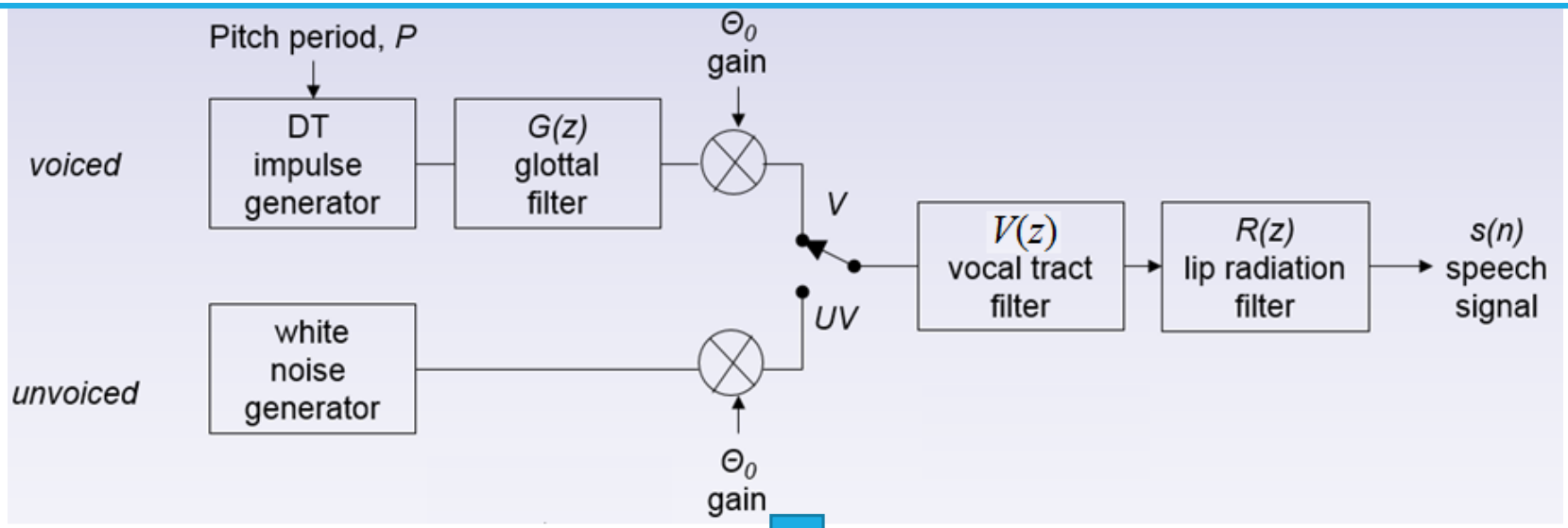
❖ گفتار را می توان به عنوان خروجی یک فیلتر خطی و متغیر با زمان مدل کرد بطوریکه این فیلتر توسط پالس های شبه پریودیک و یا نویز تحریک می شود.

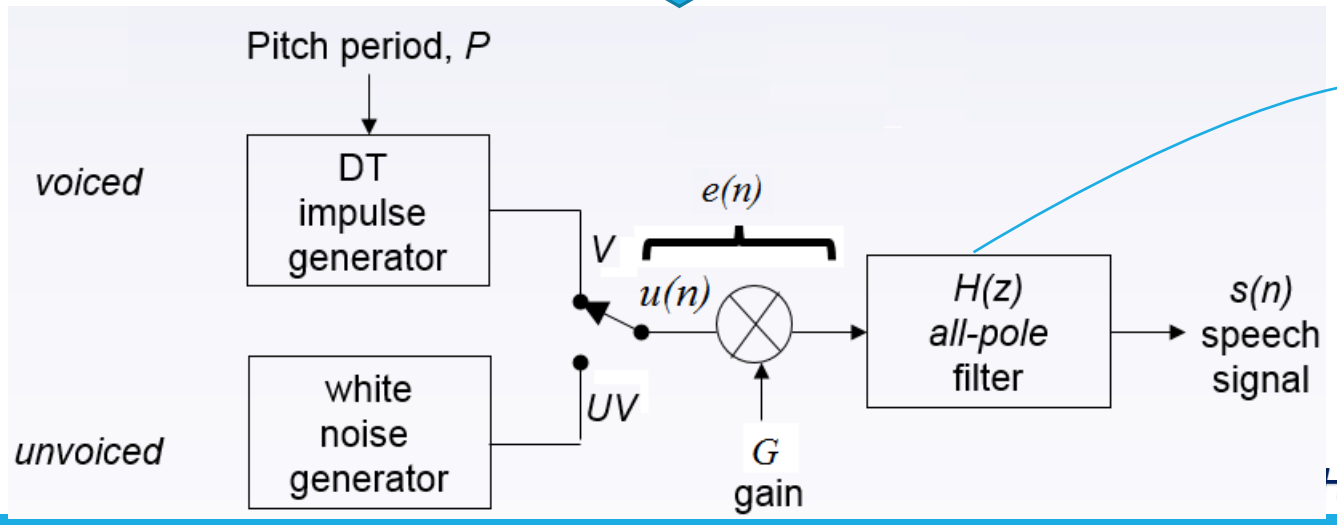
❖ فرض می شود که پارامترهای مدل بر روی بازه آنالیز گفتار ثابت هستند.

❖ در LPC به دنبال تخمین پارامترهای فیلتر (و منبع) در مدل تولید گفتار "منبع-فیلتر" هستیم.

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

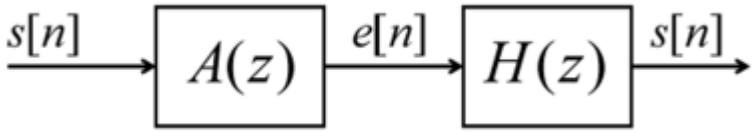




فیلتر تمام قطب بیانگر اثرات فیلترهای چاکنای، مجرای گفتار و لبها است

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)



$$e[n] = s[n] + \sum_{k=1}^{P} a_k \, s[n-k]$$

$$e[n] = G \, u[n]$$

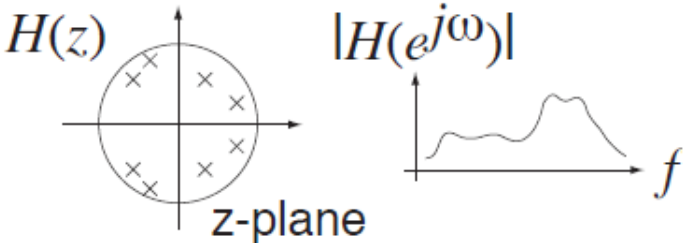
روش LPC مدل کردن یک سیگنال با یک سیستم تمام قطب یا مدل بازگشتی (AR) است

$$E(z) = S(z) + \sum_{k=1}^{P} a_k \, S(z) \, z^{-k}$$

$$GU(z) = S(z)\left(1 + \sum_{k=1}^{P} a_k \, z^{-k}\right)$$

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 + \sum_{k=1}^{P} a_k \, z^{-k}} = \frac{1}{A(z)}$$

A(z) فیلتر معکوس یا inverse filter برای H(z) است



Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

- هدف اصلی LPC: بدست آوردن ضرایب $\{a_k\}$ برای هر فریم زمان کوتاه گفتار مثل s[n]
- راه حل: کمینه سازی مجموع مربعات خطای تخمین $e[n]$

$$E=\sum_n e^2[n]=\sum_n\left(s[n]+\sum_{k=1}^{P}a_k\,s[n-k]\right)^2$$

- کافی است از E نسبت به ضرایب $\{a_i\}$ مشتق بگیریم و برابر با صفر قرار دهیم:

$$\frac{\partial E}{\partial a_i}=0\Rightarrow 2\sum_n\left(s[n-i]\right)\left(s[n]+\sum_{k=1}^{P}a_k\,s[n-k]\right)=0$$

$$\Rightarrow -\underbrace{\sum_n\left(s[n-i]s[n]\right)}_{\phi_{ss}(i,0)}=\sum_{k=1}^{P}a_k\underbrace{\sum_n\left(s[n-i]s[n-k]\right)}_{\phi_{ss}(i,k)}$$

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

$$\Rightarrow \underbrace{-\sum_{n}\left(s[n-i]s[n] \right)}_{\phi_{ss}(i,0)} = \sum_{k=1}^{P} a_k \underbrace{\sum_{n}\left(s[n-i]s[n-k] \right)}_{\phi_{ss}(i,k)}$$

$$\Rightarrow -\phi_{ss}(i,0) = \sum_{k=1}^{P} a_k \, \phi_{ss}(i,k)$$

$$\begin{aligned}
&i=1\Rightarrow\ a_1\phi_{ss}(1,1) + a_2\phi_{ss}(1,2) + \ldots + a_p\phi_{ss}(1,p) = -\phi_{ss}(1,0)\\
&i=2\Rightarrow\ a_1\phi_{ss}(2,1) + a_2\phi_{ss}(2,2) + \ldots + a_p\phi_{ss}(2,p) = -\phi_{ss}(2,0)\\
&\ldots \qquad\qquad \ldots. \qquad\qquad \ldots.\\
&i=p\Rightarrow\ a_1\phi_{ss}(p,1) + a_2\phi_{ss}(p,2) + \ldots + a_p\phi_{ss}(p,p) = -\phi_{ss}(p,0)
\end{aligned}$$

p معادله و p مجهول
معادلات Yule-Walker

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

$$\begin{bmatrix} \phi_{ss}(1,1) & \phi_{ss}(1,2) & \dots & \phi_{ss}(1,p) \\ \phi_{ss}(2,1) & \phi_{ss}(2,2) & \dots & \phi_{ss}(2,p) \\ \dots & \dots & \dots & \dots \\ \phi_{ss}(p,1) & \phi_{ss}(p,2) & \dots & \phi_{ss}(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} -\phi_{ss}(1,0) \\ -\phi_{ss}(2,0) \\ \dots \\ -\phi_{ss}(p,0) \end{bmatrix}$$

❖ چون $\phi_{ss}(i,k) = \phi_{ss}(k,i)$ ، پس ماتریس ϕ_{ss} متقارن (symmetric) و مثبت معین (positive-definite) است، پس ماتریس حتما معکوس دارد.

$$\phi_{p\times p}\ a_{p\times 1} = \psi_{p\times 1} \Rightarrow a = \psi\,\phi^{-1}$$

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

- روش های حل معادلات Yule-Walker :
  - روش اتوکورولیشن یا لوینسون – دوربین
  - روش کواریانس
  - روش Burg

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

❖ روش اتوکورولیشن یا لوینسون – دوربین

❖ هدف: ساده کردن محاسبات $\phi_{ss}(i,k)$

$$\phi_{ss}(i,k)=\sum_{n=-\infty}^{+\infty} s[n-i]\ s[n-k]$$

$$\sum_{n=-\infty}^{+\infty} s[n]\ s[n+(i-k)]$$

$$=\phi_{ss}(0,\ i-k)$$

❖ در روش اتوکورولیشن فرض می شود که سیگنال در یک پنجره همینگ با طول **N** ضرب شده است و خارج از بازه **[0,N-1]** برابر با صفر است.

$$\phi_{ss}(0,\ i-k)=\sum_{n=0}^{N-1-|i-k|} s[n]\ s[n+|i-k|]$$

$$=R_{ss}(|i-k|) \qquad ,\quad 0\le k\le p,\quad 1\le i\le p$$

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

❖ روش اتوکورولیشن یا لوینسون – دوربین

$$\begin{bmatrix} \phi_{ss}(1,1) & \phi_{ss}(1,2) & \dots & \phi_{ss}(1,p) \\ \phi_{ss}(2,1) & \phi_{ss}(2,2) & \dots & \phi_{ss}(2,p) \\ \dots & \dots & \dots & \dots \\ \phi_{ss}(p,1) & \phi_{ss}(p,2) & \dots & \phi_{ss}(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} -\phi_{ss}(1,0) \\ -\phi_{ss}(2,0) \\ \dots \\ -\phi_{ss}(p,0) \end{bmatrix}$$

Toeplitz Matrix

ماتریسی که عناصر روی قطرش یکسان هستند

$$\begin{bmatrix} R_{ss}(0) & R_{ss}(1) & \dots & R_{ss}(p-1) \\ R_{ss}(1) & R_{ss}(0) & \dots & R_{ss}(p-2) \\ \dots & \dots & \dots & \dots \\ R_{ss}(p-1) & R_{ss}(p-2) & \dots & R_{ss}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} -R_{ss}(1) \\ -R_{ss}(2) \\ \dots \\ -R_{ss}(p) \end{bmatrix}$$

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

❖ روش اتوکورولیشن یا لوینسون – دوربین

❖ اگر ماتریس ضرایب در معادلات یول-واکر ساختار Toeplitz داشته باشد، حل این دستگاه معادلات بصورت کارا و با استفاده از روش بازگشتی لوینسون-دوربین و با مرتبه محاسباتی $O(p^2)$ انجام می شود.

❖ در این روش، ضرایب جدیدی به نام ضرایب انعکاسی یا Reflection Coefficient یعنی k_i حاصل می شوند.

❖ نام دیگر ضرایب انعکاسی، ضرایب PARCOR است.

$$E^{(0)} = R_{ss}(0)$$
$$\text{for } i = 1 \text{ to } p$$
$$k_i = R_{ss}(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} R_{ss}(i-j) \Big/ E^{(i-1)}$$
$$a_i^{(i)} = k_i$$
$$\text{for } j = 1 \text{ to } i-1$$
$$a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)}$$
$$\text{end}$$
$$E^{(i)} = (1-k_i^2) E^{(i-1)}$$
$$\text{end}$$
$$\text{for } i = 1 \text{ to } p$$
$$a_i = a_j^{(p)}$$
$$\text{end}$$

$$G^2 = R_{ss}(0) - \sum_{k=1}^{p} a_k R_{ss}(k)$$

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

$$H(z)=\frac{S(z)}{GU(z)}=\frac{1}{1+\sum_{k=1}^{P} a_k\ z^{-k}}=\frac{1}{A(z)}$$

$$or \quad H(z)=\frac{S(z)}{U(z)}=\frac{G}{A(z)}$$

$$H(e^{jw})=\frac{G}{A(e^{jw})}$$

$$H[k]=H(e^{jw_k})\Big|_{w_k=k\,2\pi/N}=\frac{G}{A[k]}$$

$$|H[k]|=\frac{G}{|A[k]|}$$

- محاسبه پوش طیف (spectral envelope) و فرمنت ها در آنالیز LPC
- هدف از محاسبه پوش طیف حذف هارمونیک ها (اثرات دندانه ای در طیف گفتار) و رسیدن به تابع تبدیل مجرای گفتار است.
- با توجه به روابط فوق اندازه DFT تابع تبدیل مجرای گفتار (یعنی $|H[k]|$) معادل پوش طیف حاصله از آنالیز LPC است.

Hamidreza Baradaran Kashani

مراحل محاسبه پوش طیف و فرمنت ها در آنالیز LPC

- محاسبه ضرایب LPC $\{a_k\}$ و گین G برای یک فریم زمان-کوتاه (short-term) سیگنال گفتار
- قرار دادن ضرایب در یک دنباله و اضافه کردن صفر (zero padding) تا حدی که تعداد اعضای دنباله توانی از ۲ شود.

$$a[n]=\{a_0,\ a_1,\ a_2,\ldots,a_p,\ \underbrace{0,0,\ldots,0}_{N-(p+1)}\}$$

- گرفتن تبدیل فوریه از دنباله فوق (محاسبه $A[k]$)

$$A[k]=DFT\{a[n]\}$$

- محاسبه دامنه طیف، معکوس کردن آن و ضرب در گین G

$$|H[k]|=G/|A[k]|$$

- اعمال لگاریتم به دامنه طیف بصورت

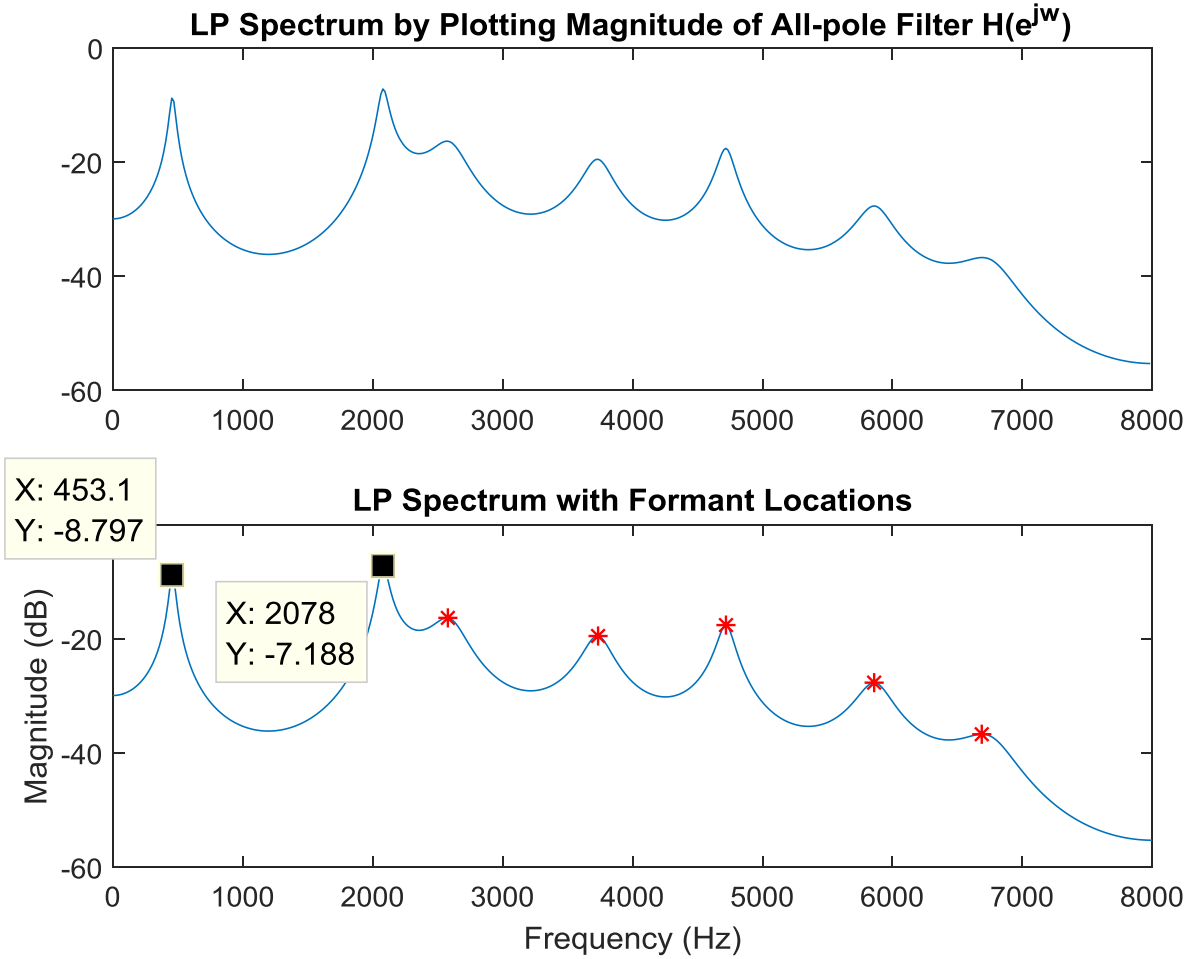
$$20\log_{10}\left(\varepsilon+|H[k]|\right)$$

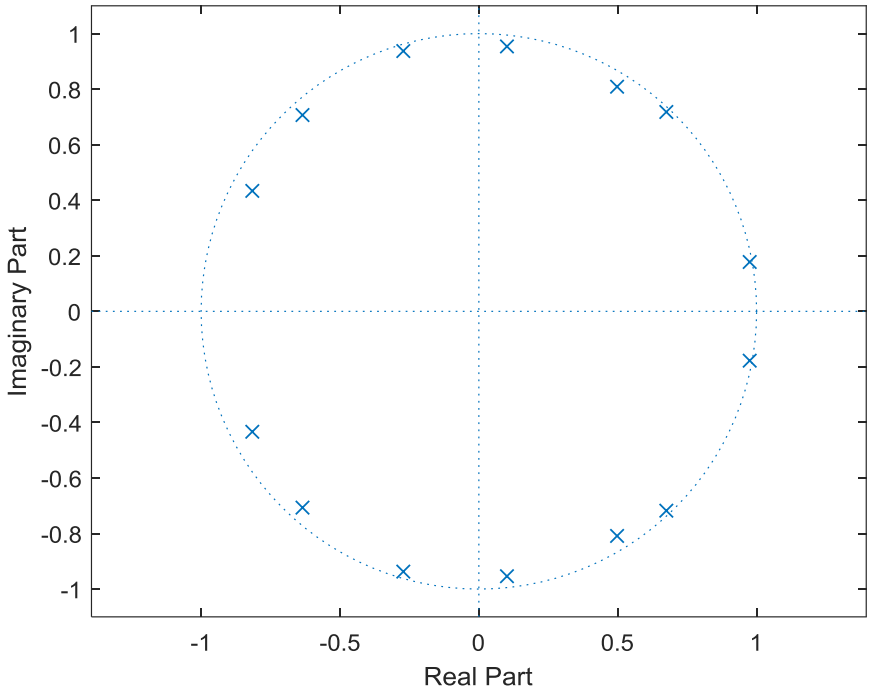
- یافتن پیک های پوش طیف حاصله به عنوان فرمنت ها

Hamidreza Baradaran Kashani

مراحل محاسبه پوش طیف و فرمنت ها در آنالیز LPC

پوش طیف مربوط به یک فریم گفتاری
از صدای واکه /e/ و محل فرمنت ها

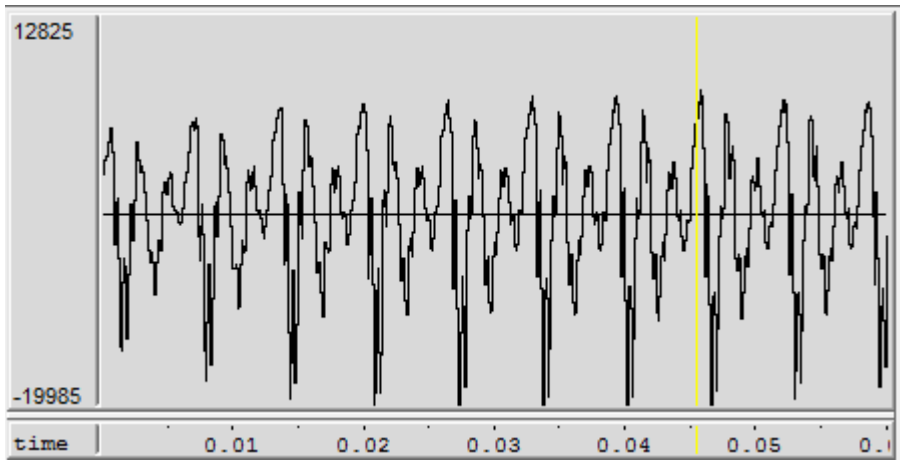


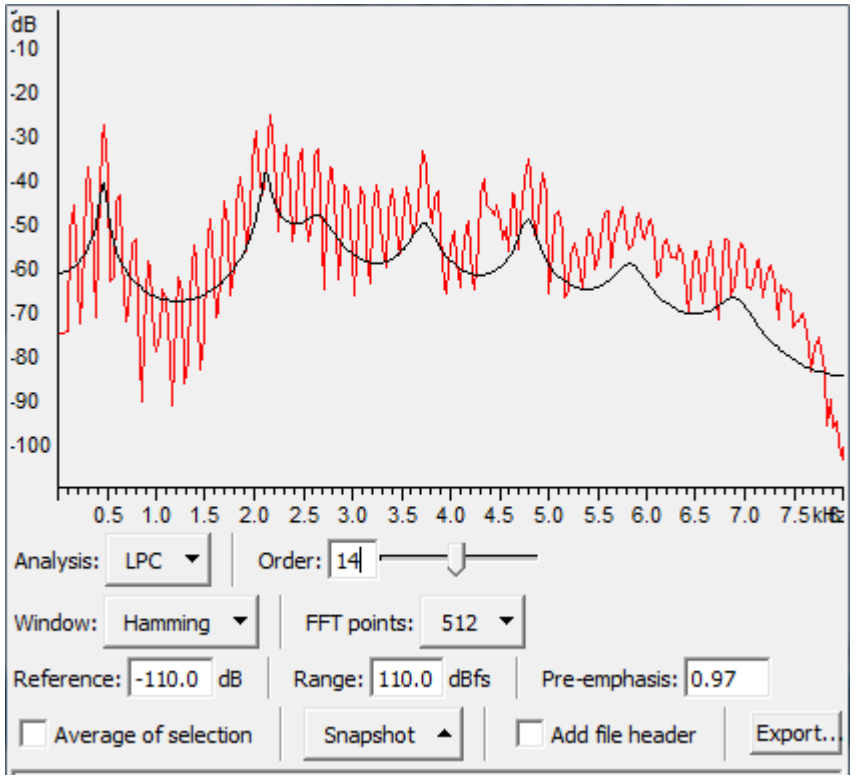


Hamidreza Baradaran Kashani

مراحل محاسبه پوش طیف و فرمنت ها در آنالیز LPC

پوش طیف مربوط به یک فریم گفتاری
از صدای واکه /e/ و محل فرمنت ها





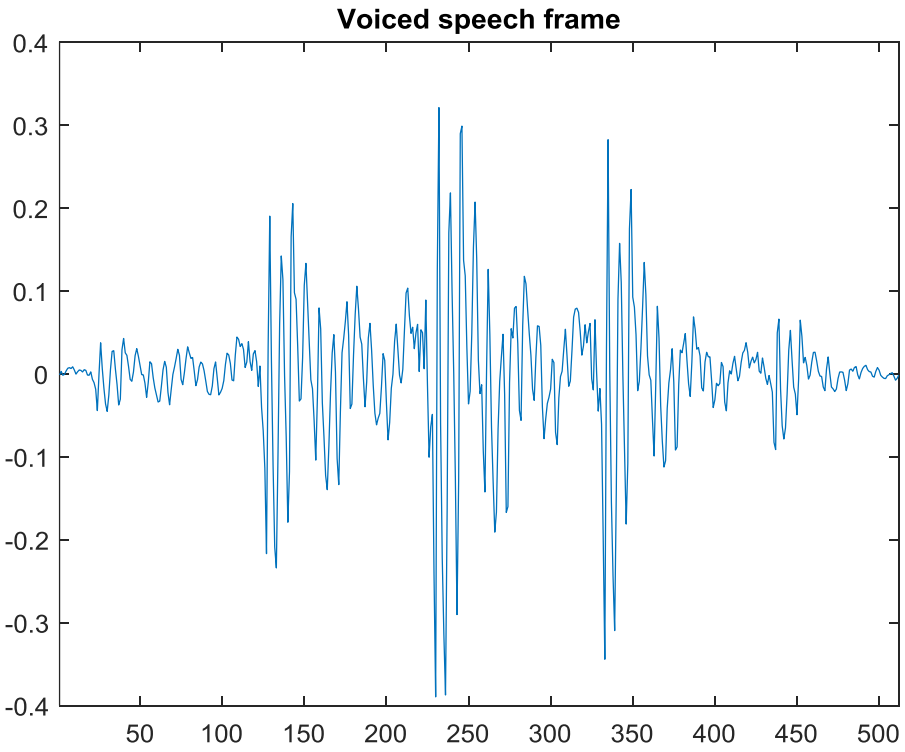
Hamidreza Baradaran Kashani

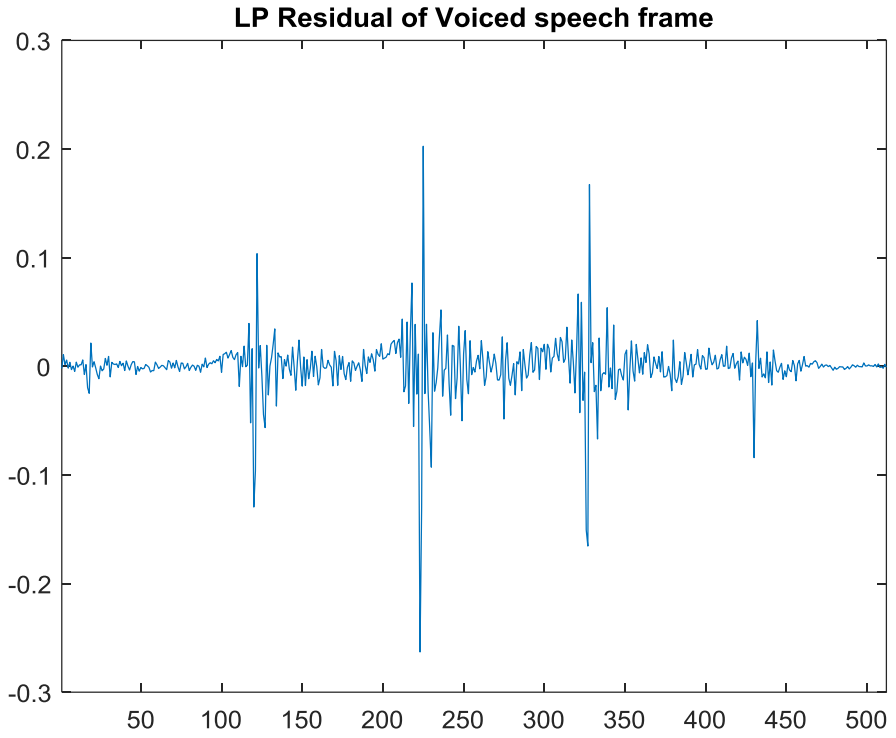
محاسبه سیگنال مانده LP

$$a[n]=\{a_0,\ a_1,\ a_2,\ldots,a_p,\ \underbrace{0,0,\ldots,0}_{N-(p+1)}\}$$

$$e[n]=s[n]\otimes a[n],\qquad \otimes:convolution$$

❖ کانولوشن سیگنال گفتار با دنباله **a[n]** بیانگر ضرایب **LP**، سیگنال مانده **LP** یا همان **e[n]** را نتیجه می دهد.



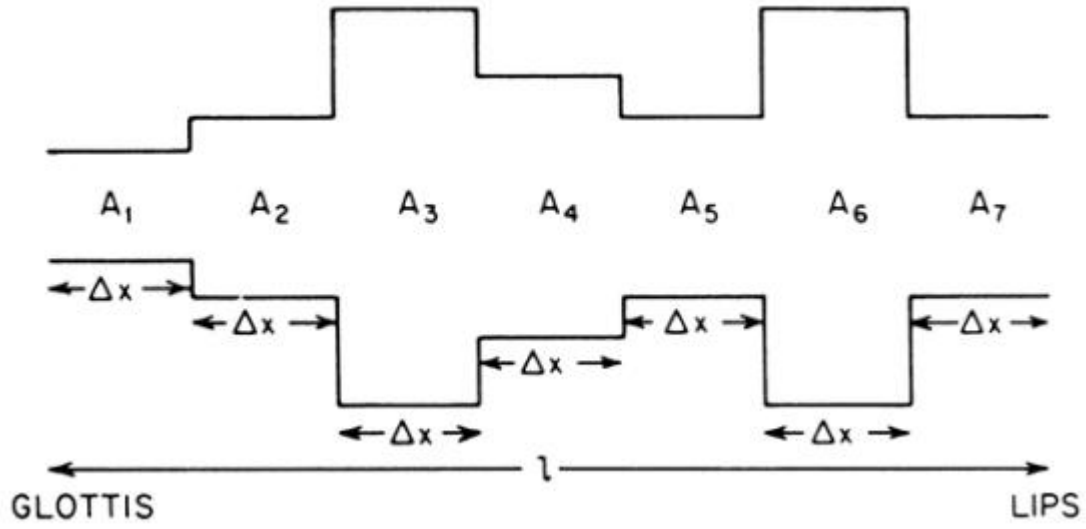


Hamidreza Baradaran Kashani

سایر ویژگی های استخراجی از تحلیل LPC

❖ **ویژگی های PARCO (ضرایب ki)**

❖ **مدل مجرای گفتار (از چاکنای تا لب ها) را می توان با اتصال تیوب هایی با سطح مقطع مختلف مدل کرد.**



$$k_i = \frac{A_{i+1} - A_i}{A_{i+1} + A_i}, \quad -1 \le k_i \le 1,$$

$$1 \le i \le P, \; A_{p+1} = A_p$$

❖ **ویژگی های LAR (Log Area Ratio)**

$$AR_i = \frac{A_{i+1}}{A_i} = \frac{1-k_i}{1+k_i},$$

$$LAR_i = g_i = \log\left(1-k_i/1+k_i\right)$$

Hamidreza Baradaran Kashani

سایر ویژگی های استخراجی از تحلیل LPC

❖ ویژگی های LAR (Log Area Ratio)

❖ ضرایب LAR برای کوانتیزاسیون و فشرده سازی گفتار به کار می روند.

❖ در آنالیز LPC با ضرایب ai به عنوان ریشه های چند جمله ای مخرج سر و کار داریم که این ضرایب بسیار حساس هستند و کوانتیزه کردن آنها باعث جابجایی زیاد فرمنت ها می شود.

❖ حساسیت ضرایب LAR بسیار کمتر است.

$$AR_i = \frac{A_{i+1}}{A_i} = \frac{1-k_i}{1+k_i},$$

$$LAR_i = g_i = \log\left(1-k_i/1+k_i\right)$$

Hamidreza Baradaran Kashani

آنالیز پیشگویی خطی (LPC)

$$H(z)=\frac{S(z)}{U(z)}=\frac{G}{A(z)}$$

$$A(z)=1+\sum_{k=1}^{P} a_k\ z^{-k}$$

$$P(z)=A(z)+z^{-(p+1)}A(z^{-1})$$

$$Q(z)=A(z)-z^{-(p+1)}A(z^{-1})$$

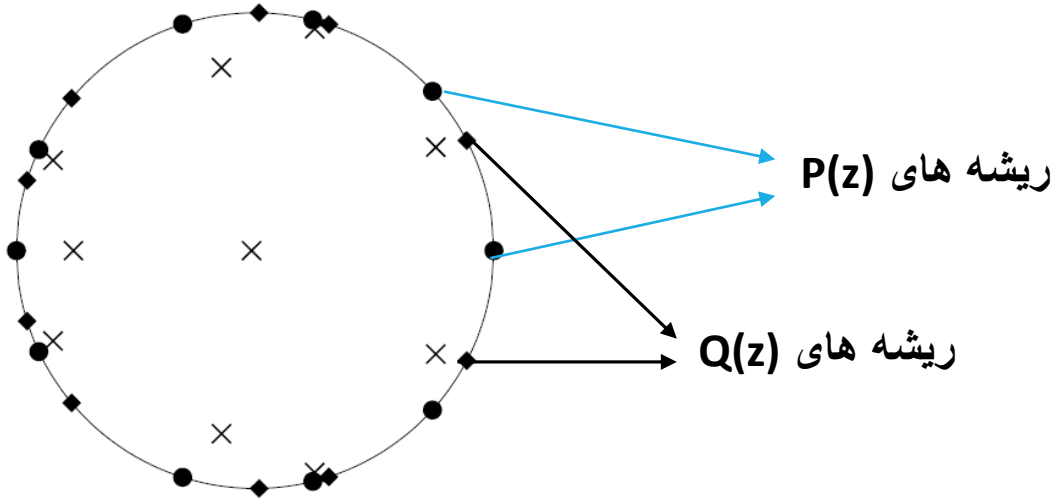
$$A(z)=\frac{P(z)+Q(z)}{2}$$

❖ ویژگی های LSF یا ضرایب جفت طیف خطی

❖ H(z) بیانگر فیلتر مجرای گفتار و A(z) فیلتر معکوس H(z) است.

❖ A(z) را می توان بصورت ترکیب دو چند جمله ای P(z) و Q(z) بصورت زیر نوشت:

❖ ریشه های چند جمله ای P(z) و Q(z) در روابط روبرو همان ضرایب LSF هستند که بصورت یک در میان حول دایره واحد قرار گرفته اند.



LSP or LSF: Line Spectral Pairs or Line Spectral Frequencies

Hamidreza Baradaran Kashani

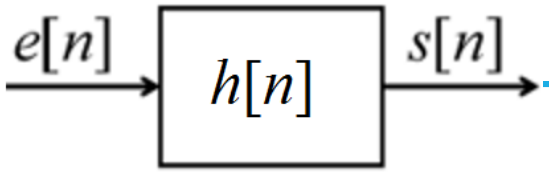
تحلیل کپستروم (Cepstrum Analysis)

- ادبیات استفاده شده جهت پردازش سیگنال در حوزه کپسترال
- نامگذاری: معکوس کردن هجای اول کلمات معادل در حوزه فرکانس

$S[k]$	$C_s[n]$
Frequency	Quefrency
Spectrum	Cepstrum
Filter	Lifter
Harmonic	Rahmonic

Hamidreza Baradaran Kashani

تحلیل کپستروم (Cepstrum Analysis)



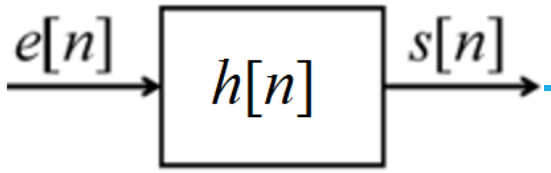
- با در نظر گرفتن مدل منبع-فیلتر:
- هدف تحلیل کپستروم: جداسازی دو بخش سیگنال تحریک و پاسخ ضربه مجرای گفتار از یکدیگر
- ضرایب کپستروم پایدارتر و قابل اعتمادتر از ضرایب LPC هستند.
- ضرایب کپستروم در حوزه جدیدی به نام کیوفرنسی (Quefrency) قرار دارند (مشابه زمان اما کمی متفاوت)

- کپستروم حقیقی: معکوس تبدیل فوریه لگاریتم دامنه تبدیل فوریه (F یعنی تبدیل فوریه)

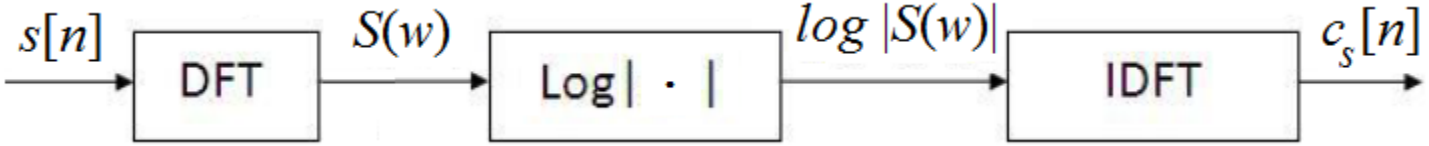
$$C_s[n] = F^{-1}\left\{\log\left|F\left\{s[n]\right\}\right|\right\}$$

Hamidreza Baradaran Kashani

تحلیل کپستروم (Cepstrum Analysis)



❖ روابط تحلیل کپستروم:



$$s[n] = h[n] \otimes e[n]$$

$$S(e^{jw}) = H(e^{jw})\,E(e^{jw})$$

$$S[k] = H[k]\;E[k]$$

$$\left|S[k]\right| = \left|H[k]\right|\;\;\left|E[k]\right|$$

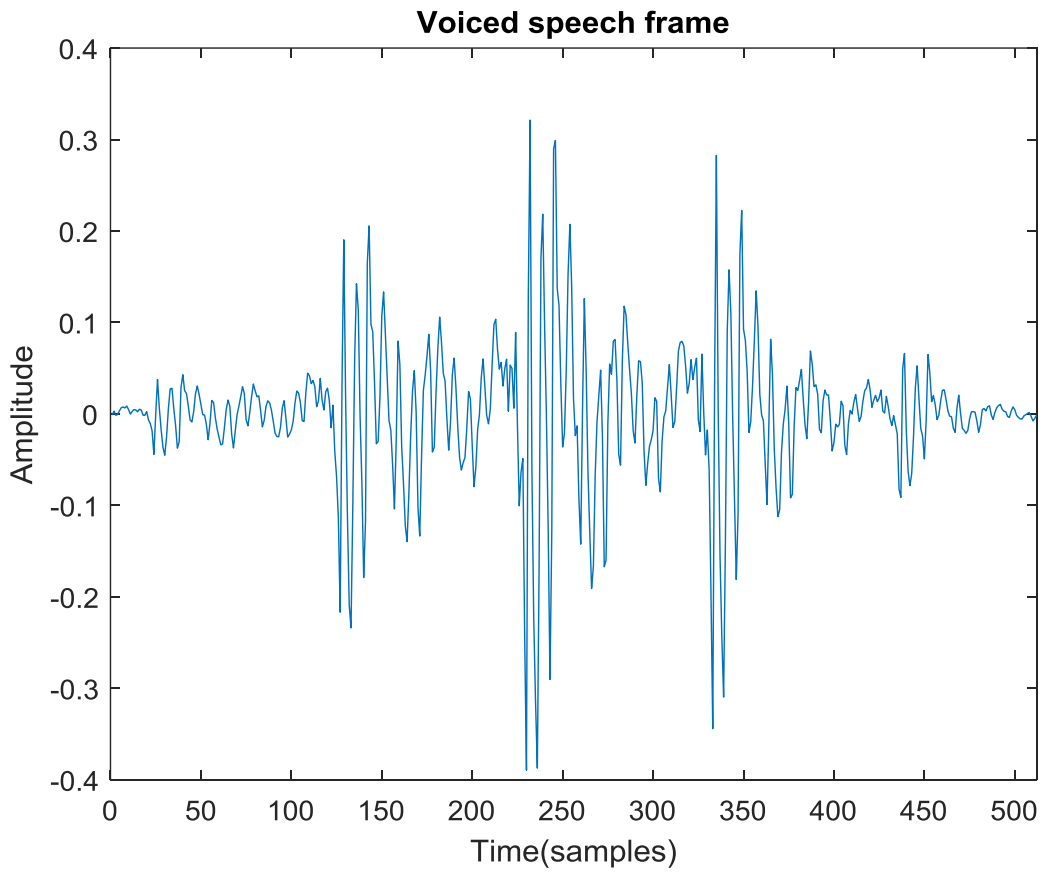
$$\log\left|S[k]\right| = \log\left|H[k]\right| + \log\left|E[k]\right|$$

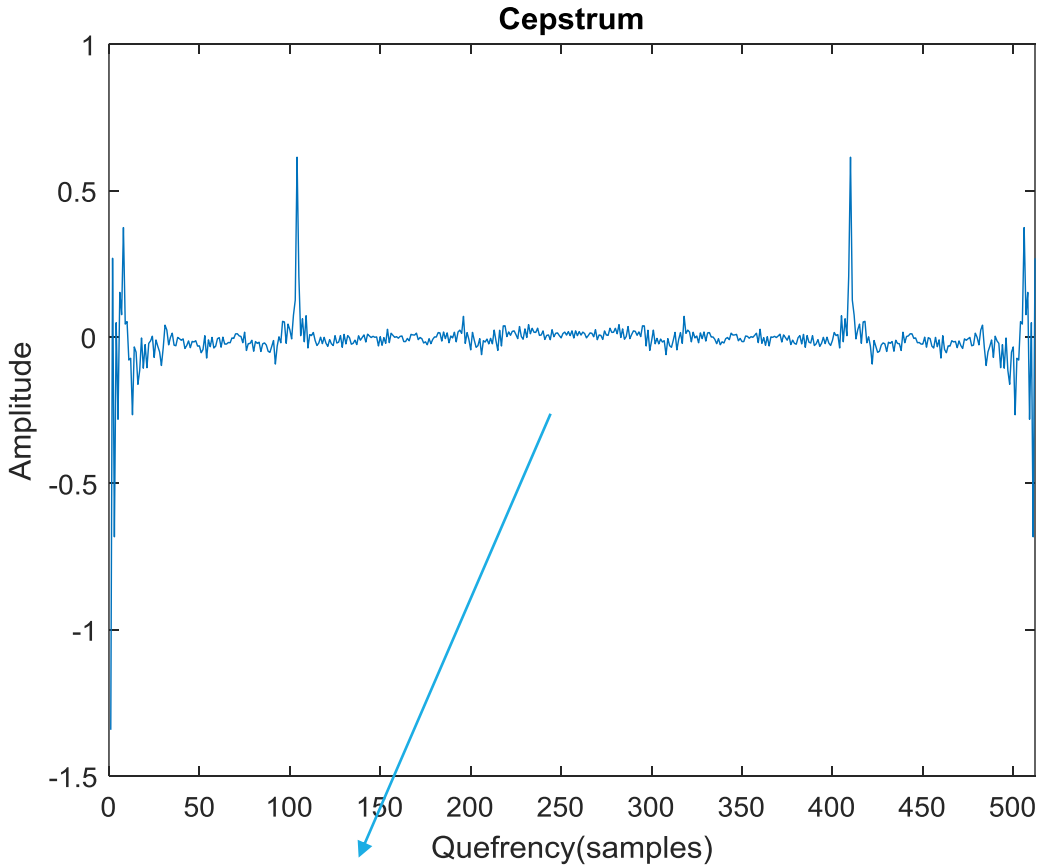
$$F^{-1}\left(\log\left|S[k]\right|\right) = F^{-1}\left(\log\left|H[k]\right|\right) + F^{-1}\left(\log\left|E[k]\right|\right)$$

$$C_s[n] = C_h[n] + C_e[n]$$

Hamidreza Baradaran Kashani

تحلیل کپستروم (Cepstrum Analysis)

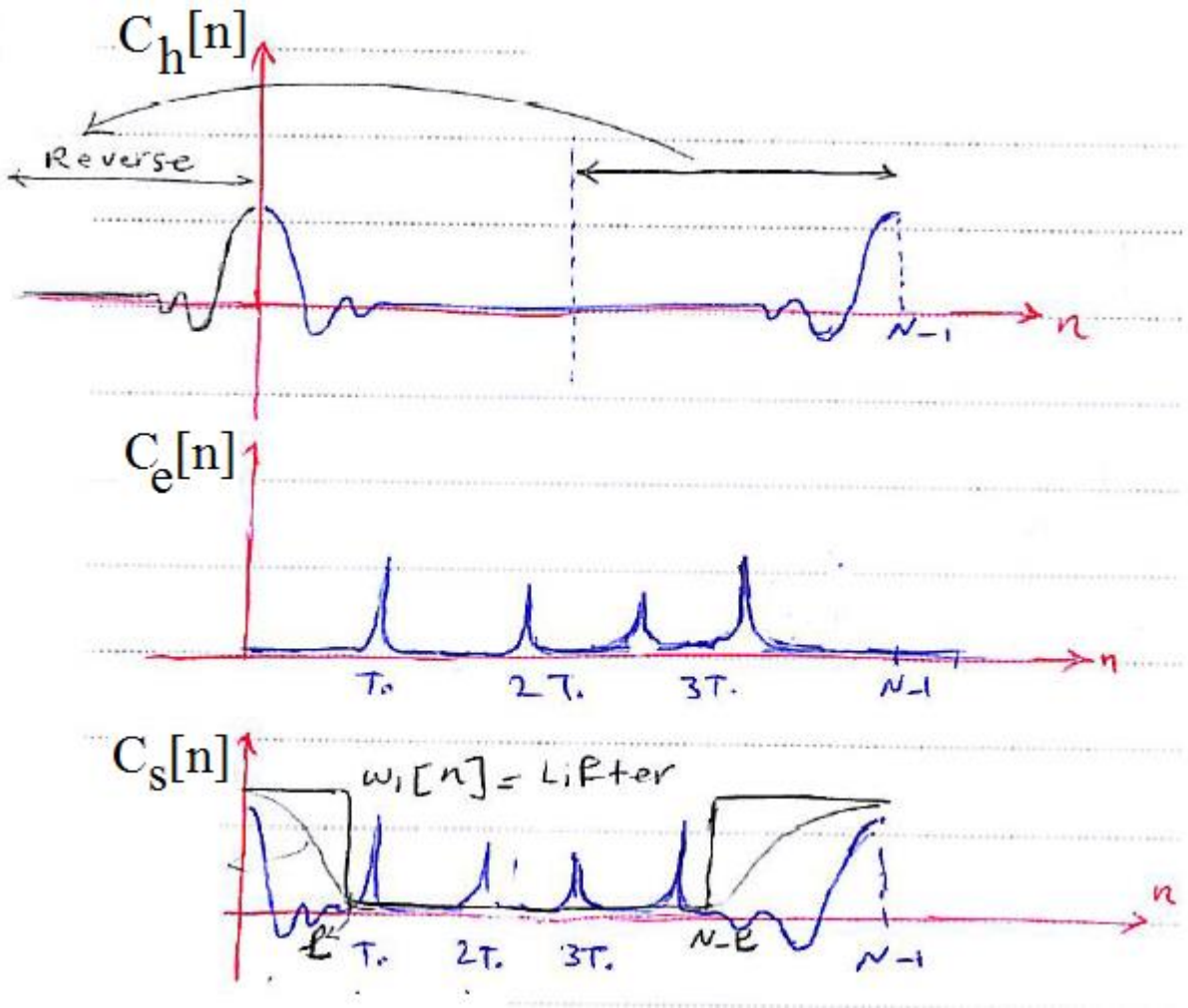




سوال؟
چه اطلاعاتی در ضرایب کپستروم وجود دارند و می توان استخراج کرد؟

Hamidreza Baradaran Kashani

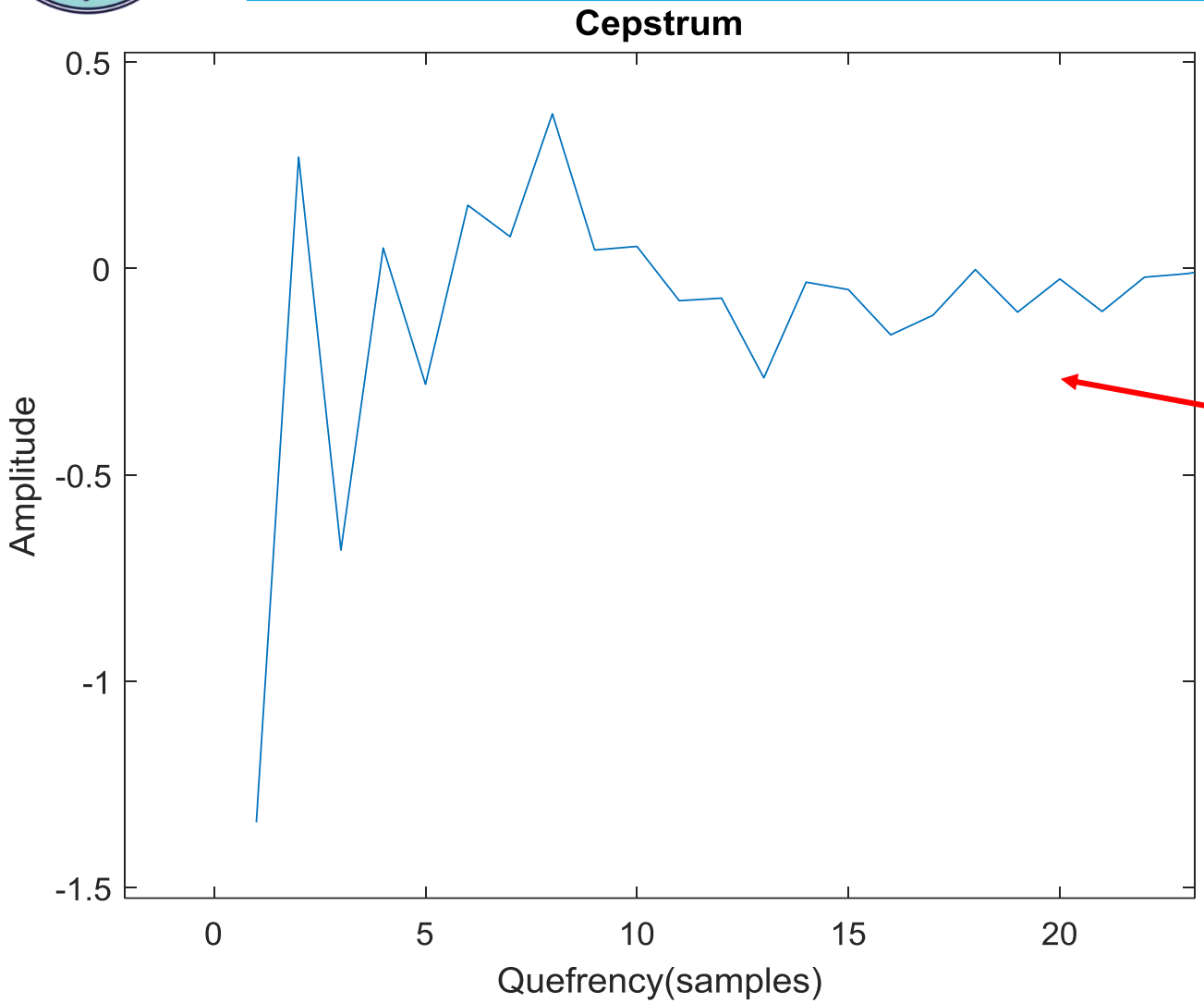
تحلیل کپستروم (Cepstrum Analysis)

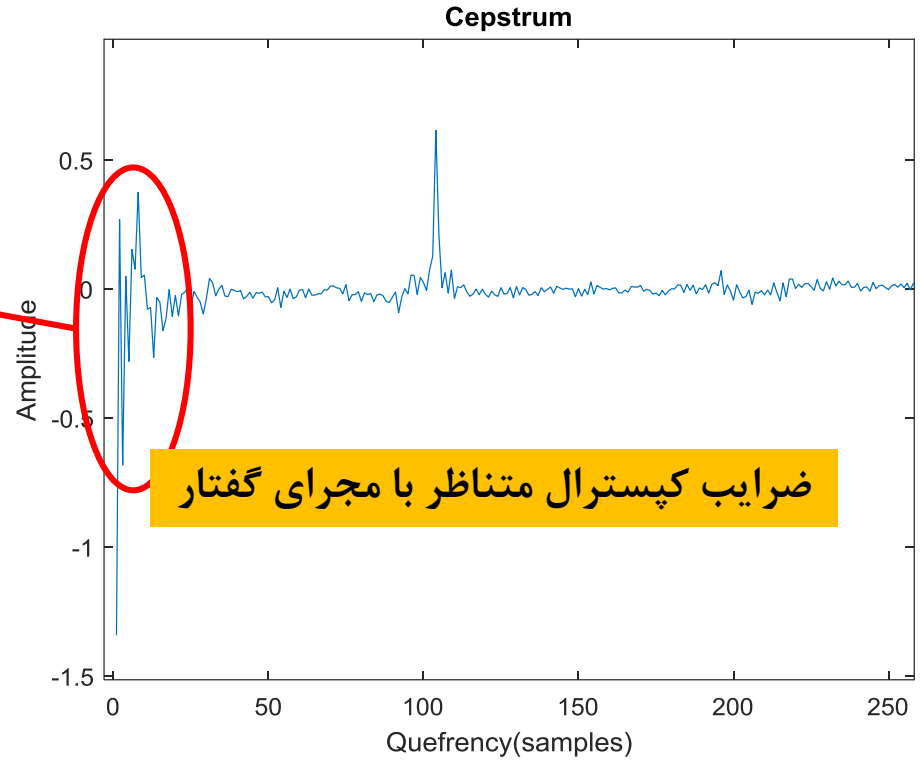


- مشخصات طیفی مجرای گفتار اصولا در ضرایب پایین کپستروم نهفته است.
- مشخصات حنجره و سیگنال تحریک در ضرایب بالاتر کپستروم قرار دارند.
- جداسازی اطلاعات مرتبط با مجرای گفتار از اطلاعات منبع تحریک:
  - اعمال لیفتر پایین گذر (در حوزه کیوفرنسی) برای استخراج ضرایب کپسترال متناظر با مجرای گفتار
  - اعمال لیفتر بالا گذر (در حوزه کیوفرنسی) برای استخراج ضرایب کپسترال متناظر با منبع تحریک
  - لیفتر بایستی نسبت به محور وسط متقارن باشد

Hamidreza Baradaran Kashani

تحلیل کپستروم (Cepstrum Analysis)

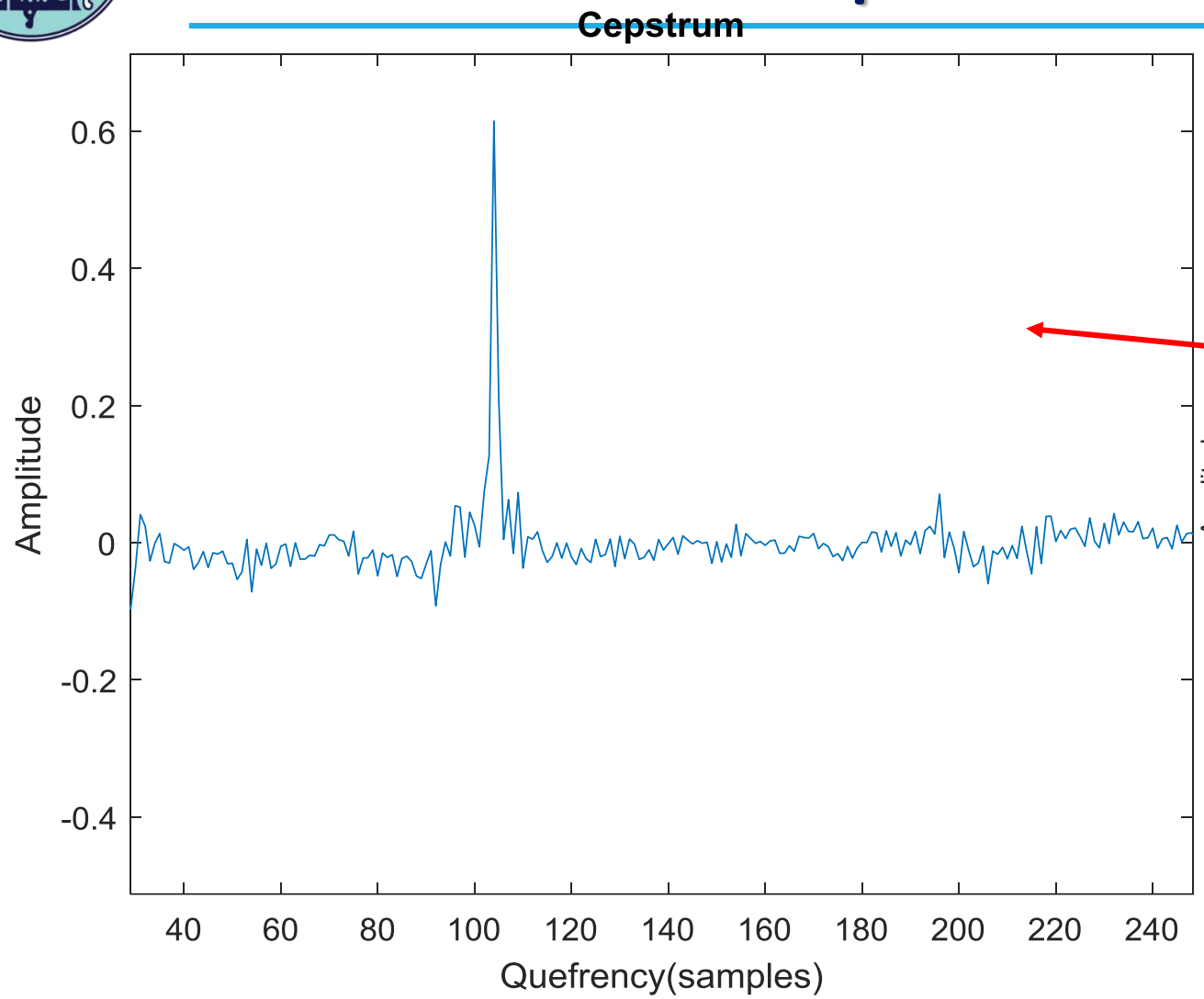


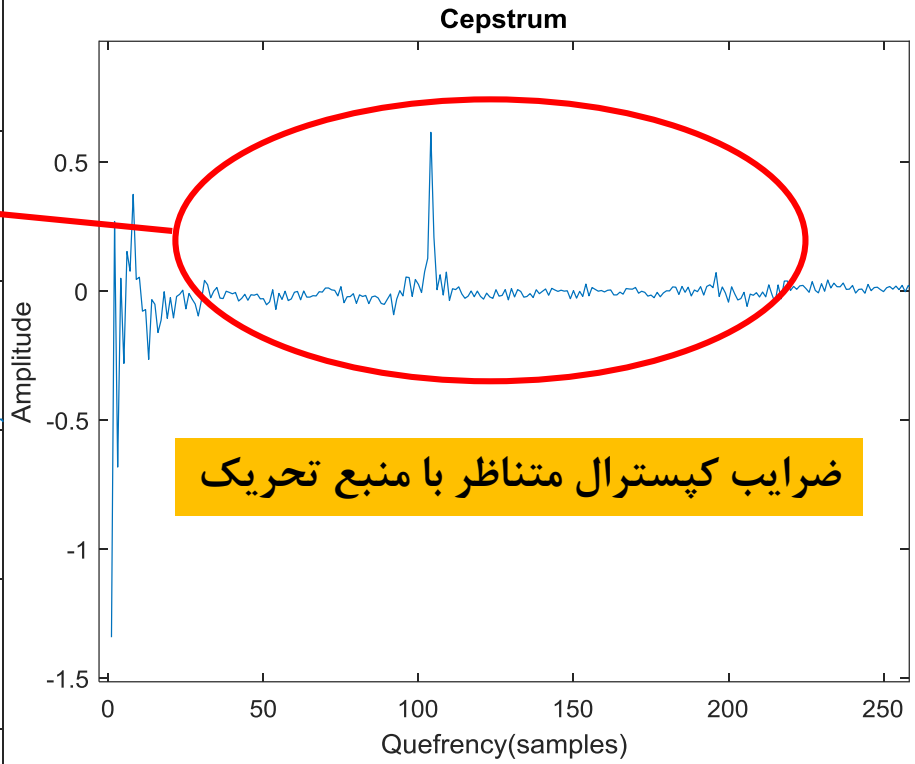


ضرایب کپسترال متناظر با مجرای گفتار

Hamidreza Baradaran Kashani

تحلیل کپستروم (Cepstrum Analysis)

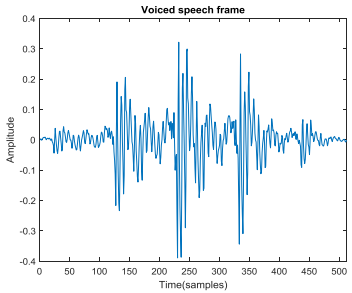




ضرایب کپسترال متناظر با منبع تحریک

Hamidreza Baradaran Kashani

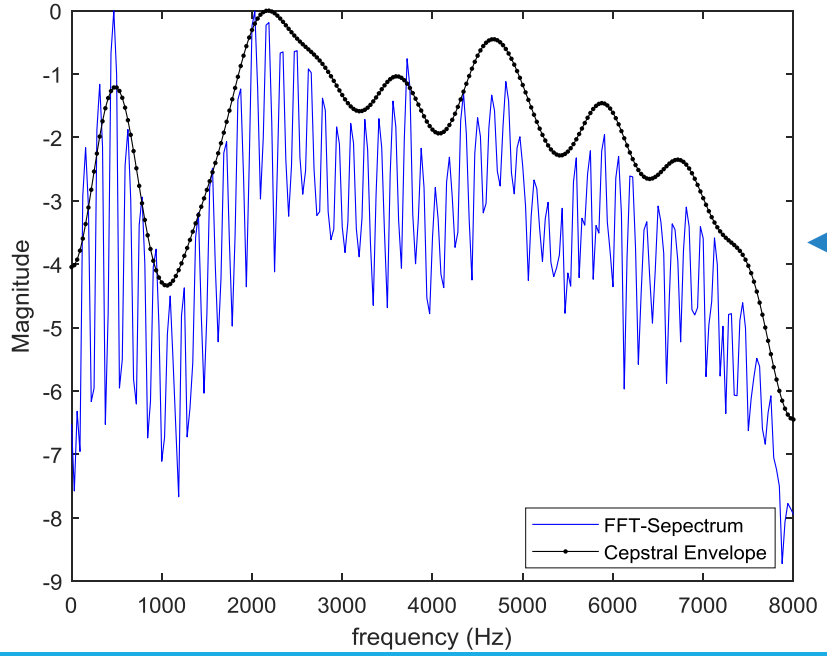
استخراج پوش طیف از ضرایب کپسترال



فریم بندی ← پیش تاکید ← پنجره گذاری ← استخراج ضرایب کپسترال

↓

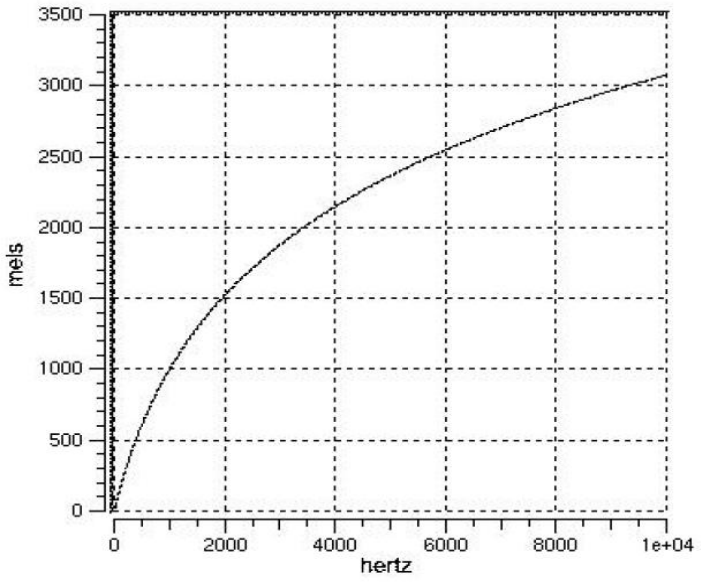
اعمال لیفتر پایین گذر ← تبدیل فوریه گسسته (DFT) ← بخش حقیقی DFT



Hamidreza Baradaran Kashani

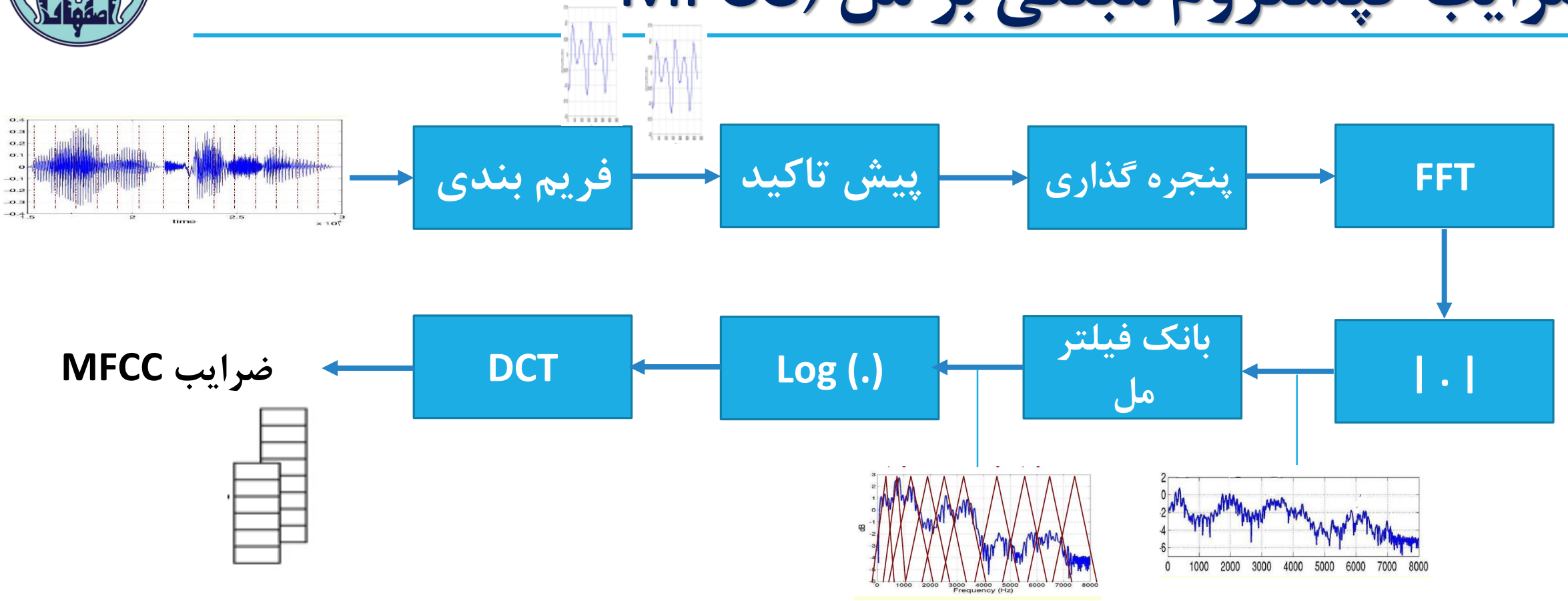
ضرایب کپستروم مبتنی بر مل (MFCC)

- در سال ۱۹۸۰ توسط Mermelstein & Davis
- پرکاربردترین ویژگی در سیستم های شناسایی گفتار
  - البته پرکاربرد در حوزه بازشناسی گوینده؛ بازشناسی زبان، تبدیل گفتار و ...
- مهمترین مشخصه ویژگی های MFCC:
  - شبیه سازی رفتار سیستم شنوایی انسان با استفاده از بانک فیلتر توزیع شده در مقیاس غیرخطی مل
  - گوش انسان حساسیت یکسانی به تمام باندهای فرکانسی نشان نمی دهد:
  - به عنوان مثال: گوش تغییر فرکانس یک تون خالص (pure tone) از ۱۰۰ هرتز به ۲۰۰ هرتز را احساس می کند اما از ۵۱۰۰ هرتز به ۵۲۰۰ هرتز را خیر!



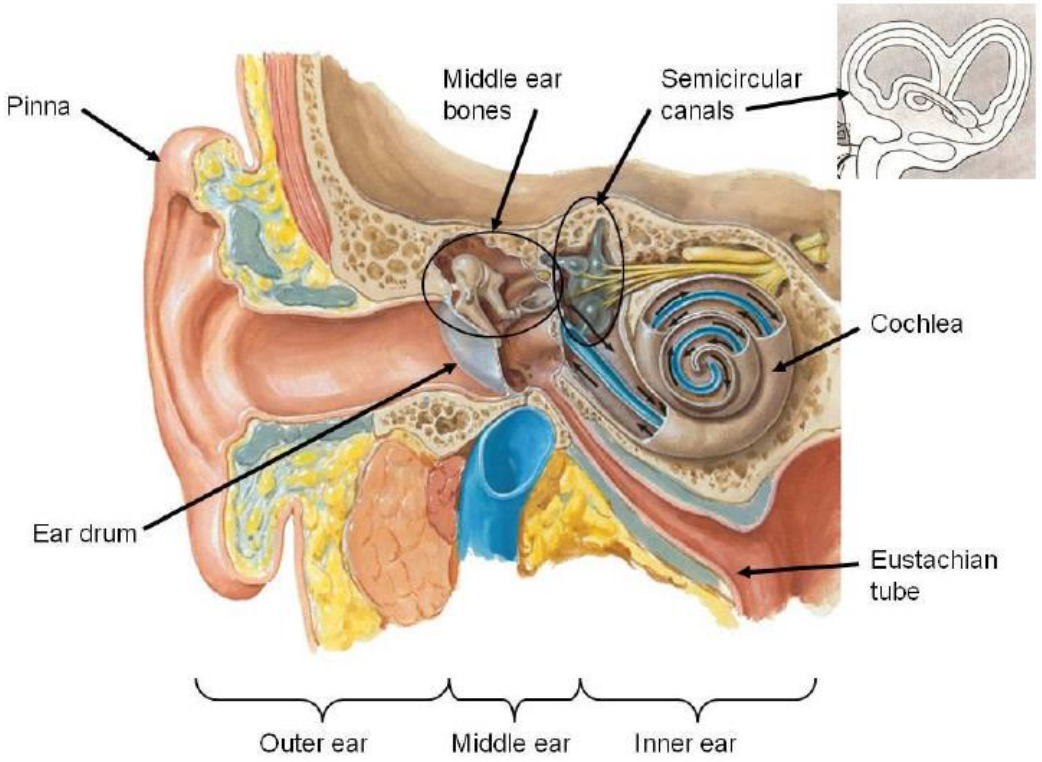
Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)



ضرایب کپستروم مبتنی بر مل (MFCC)

- ضرایب MFCC ویژگی های گفتاری مبتنی بر سیستم شنیداری انسان است.
- گوش یکی از دو مولفه شنیداری اصلی در سیستم درک گفتار است.
  - مولفه دیگر سیستم عصبی شنیداری یا همان مغز است.
- ساختار گوش از سه بخش تشکیل شده است:
  - گوش بیرونی، گوش میانی و گوش درونی
  - در ادامه تنها بر روی گوش درونی متمرکز می شویم.



Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)

❖ گوش درونی

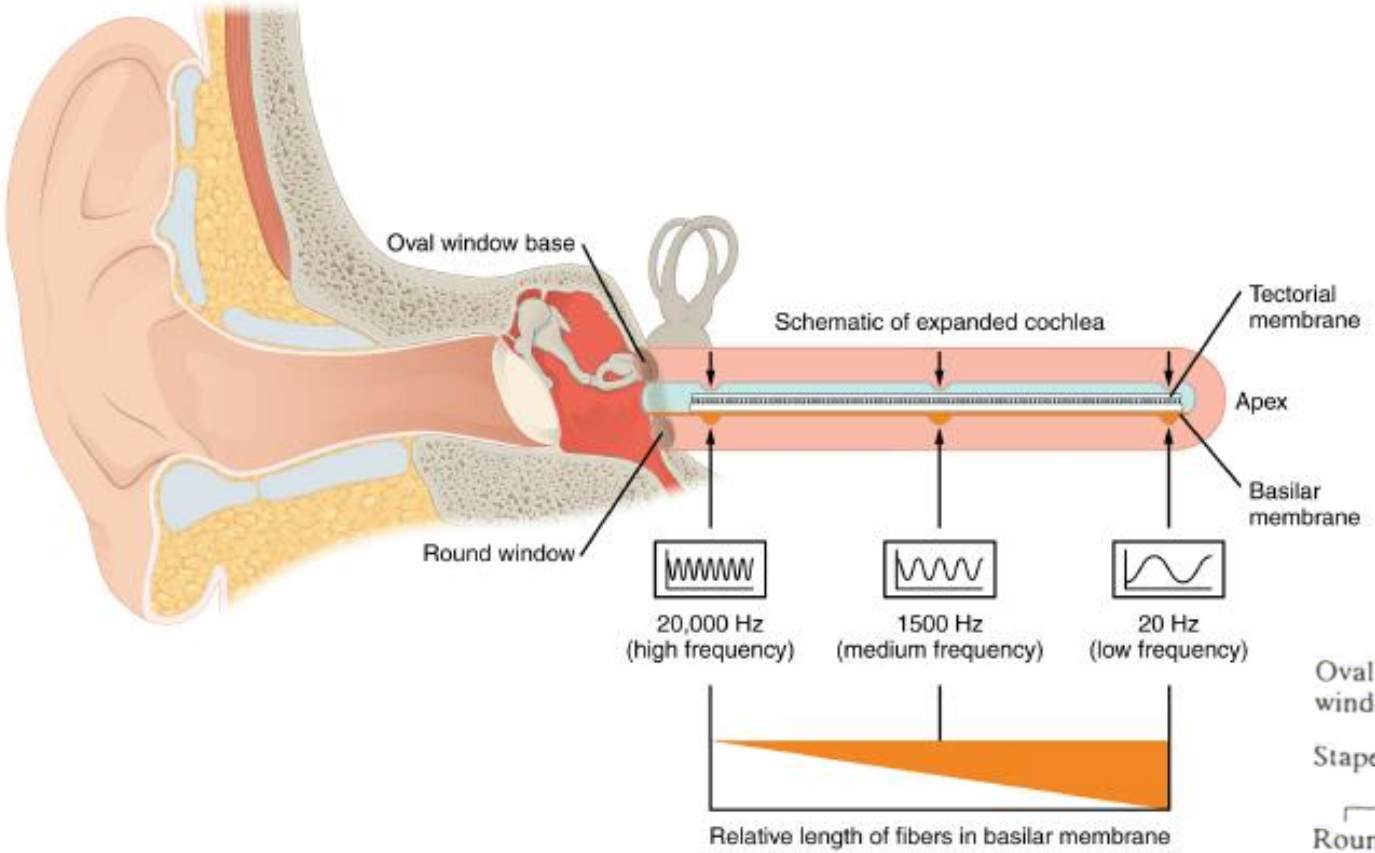
- ❖ پس از عبور امواج صوتی از هوا و انتقال آن از طریق مکانیزم گوش میانی، این امواج به مایعات گوش درونی تحویل می گردند.
- ❖ وظیفه گوش درونی تبدیل امواج مکانیکی صوتی به انرژی الکتریکی لازم برای تحریک اعصاب شنوایی است.
- ❖ گوش درونی بایستی اصوات را از نظر زیروبمی و بلندی طبقه بندی کرده (شبیه به یک بانک فیلتر میانگذر)، تا مغز به کمک این اطلاعات پیام موجود در موج صوتی را تفسیر نماید.
- ❖قسمتی از گوش درونی که مربوط به درک گفتار است و بطو مستقیم با عصب شنوایی در ارتباط است، حلزونی گوش است.

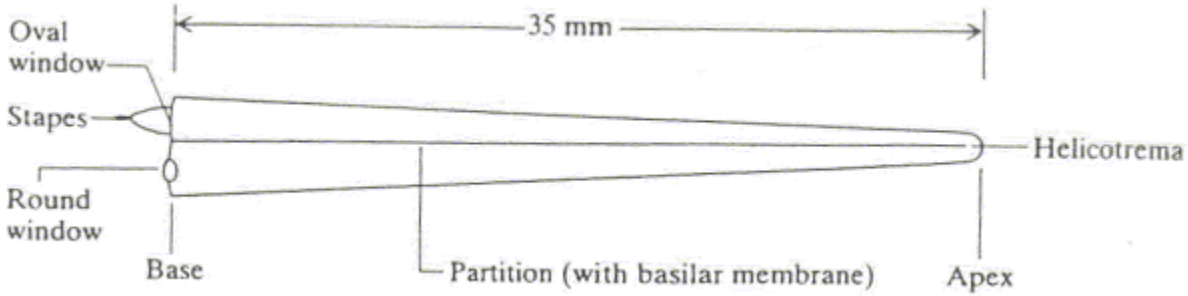
Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)

بخش حلزونی (Cochlea)

- یک لوله مارپیچ با طول ۳.۵ سانتی متر است که ۲.۶ بار به دور خود پیچیده است.
- حلزونی گوش همانند یک بانک فیلتر عمل می کند.

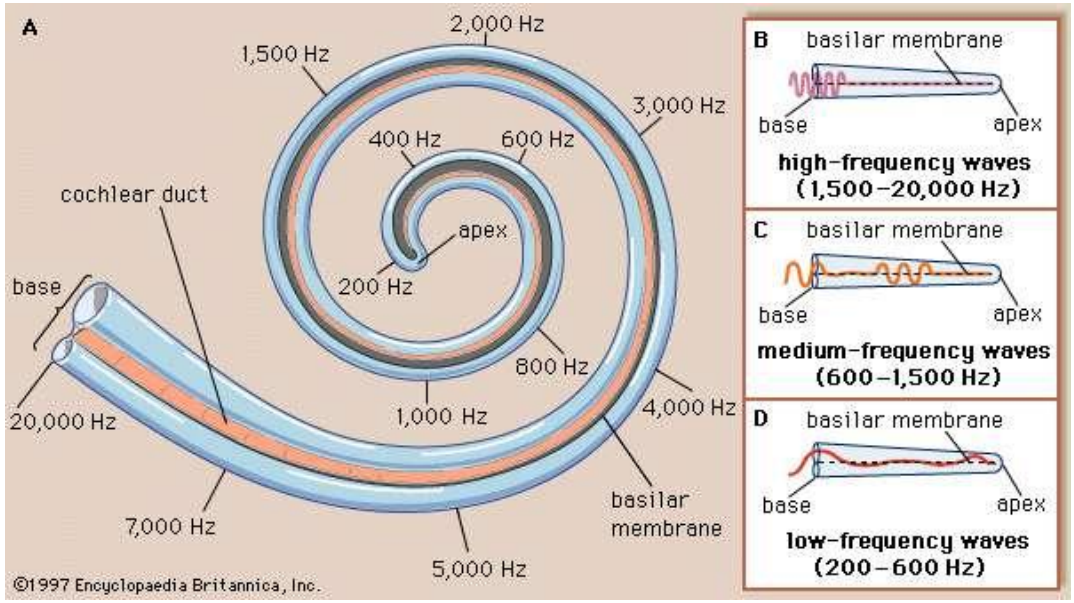




Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)

بخش حلزونی (Cochlea)

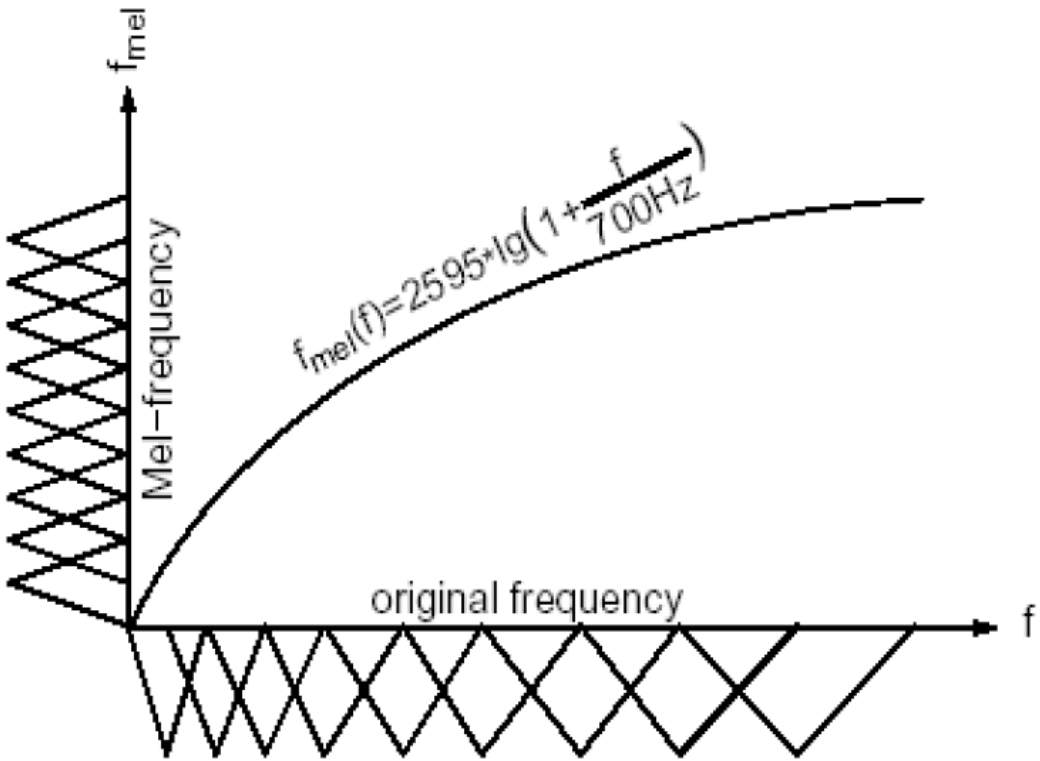


- تئوری تشدید: در این تئوری تصور می شود که در طول غشا پایه (basilar membrane)، الیافی عمود بر محور طولی حلزون قرار گرفته اند که هر کدام بر اساس کشش، قطر و طول خود نسبت به فرکانس خاصی دارای تشدید بوده و در صورت تحریک با آن فرکانس باعث ایجاد پالس عصبی در رشته های عصبی مربوطه می گردند.
- بر اساس این تئوری مکان درک فرکانس در حلزون مهم است.
- الیاف غشا پایه در پیچ قاعده حلزون (نسبت به پیچ راس) کوتاهتر و تحت کشش بیشترند، لذا در برابر فرکانس های بالاتر به حداکثر ارتعاش خود می رسند.
- الیاف غشا پایه در سمت راس حلزون که بلندتر و تحت کشش کمتر هستند، در فرکانس های پایین حداکثر ارتعاش را پیدا می کنند.

Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)

❖ بانک فیلتر مل



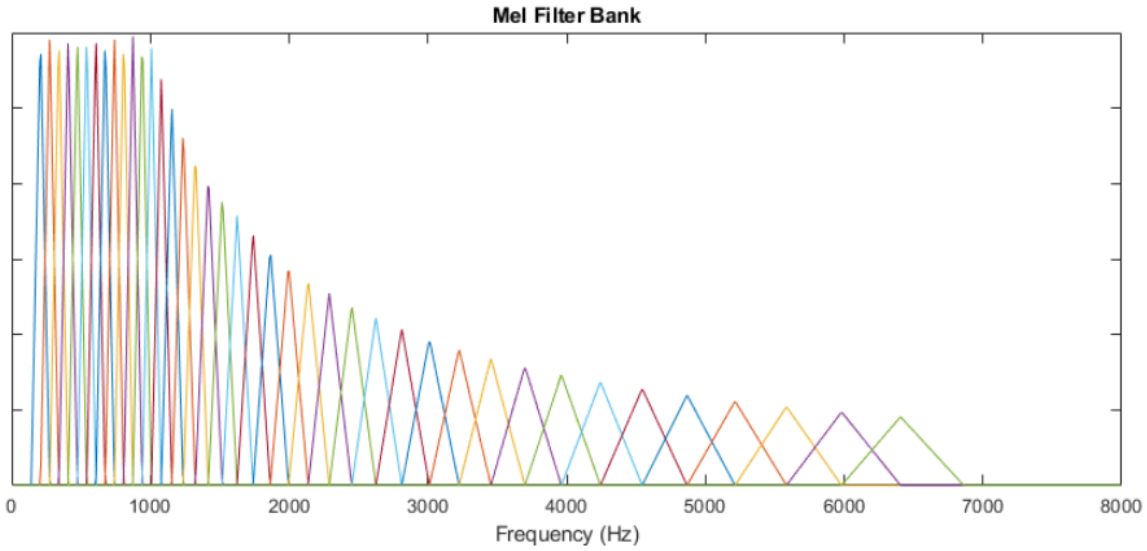
- ❖ بانک فیلتر مل در روش MFCC برای شبیه سازی بخش حلزونی گوش است.
- ❖ محور عمودی در شکل روبرو همان محور طولی حلزونی است.
- ❖ سلول های موجود در بخش حلزونی مثل یک فیلتر میانگذر عمل می کنند. در واقع هر کدام از فیلترهای مثلثی یک سلول شنوایی را شبیه سازی می کند.
- ❖ فیلترها بر روی محور مل بصورت یکنواخت و بر روی محور فرکانس بصورت غیر یکنواخت توزیع شده اند.
- ❖ مقیاس مل تقریبا زیر 1 KHz خطی است و بالای آن لگاریتمی است.

Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)

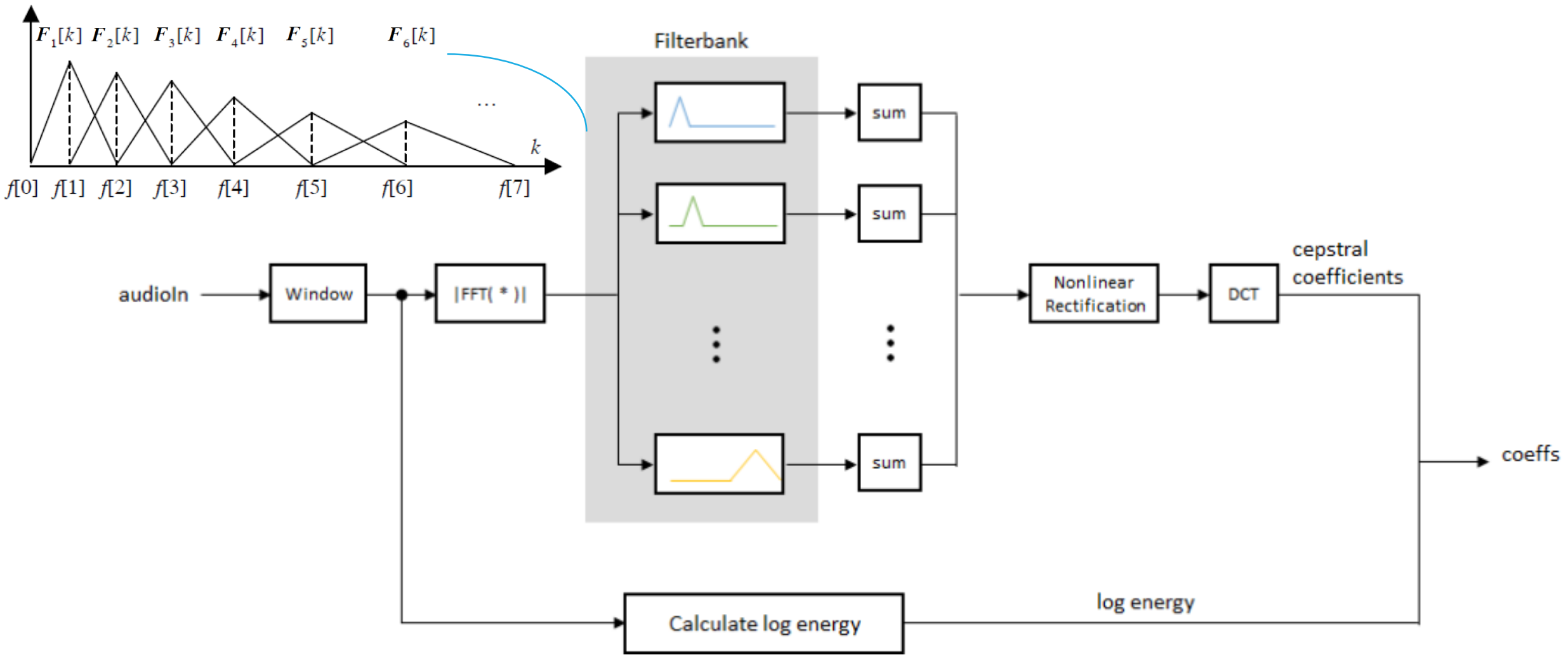
❖ بانک فیلتر مل

- ❖ معمولا دامنه همه فیلترهای مثلثی یکی است. در برخی منابع دامنه فیلترها را به گونه ای می گیرند که سطح زیرشان یک شود.
- ❖ تعداد فیلترها محدود: بین ۲۰ تا ۴۰ فیلتر
- ❖ برای اکثر کاربردها بانک فیلتر بر روی یک محدوده فرکانسی مشخص توزیع می شوند.
  - ❖ مثلا برای گفتار میکروفونی F_low = 60 Hz, F_high=7000 Hz
  - ❖ برای گفتار تلفنی: F_low=100 Hz, F_high=3900 Hz



Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)



Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)

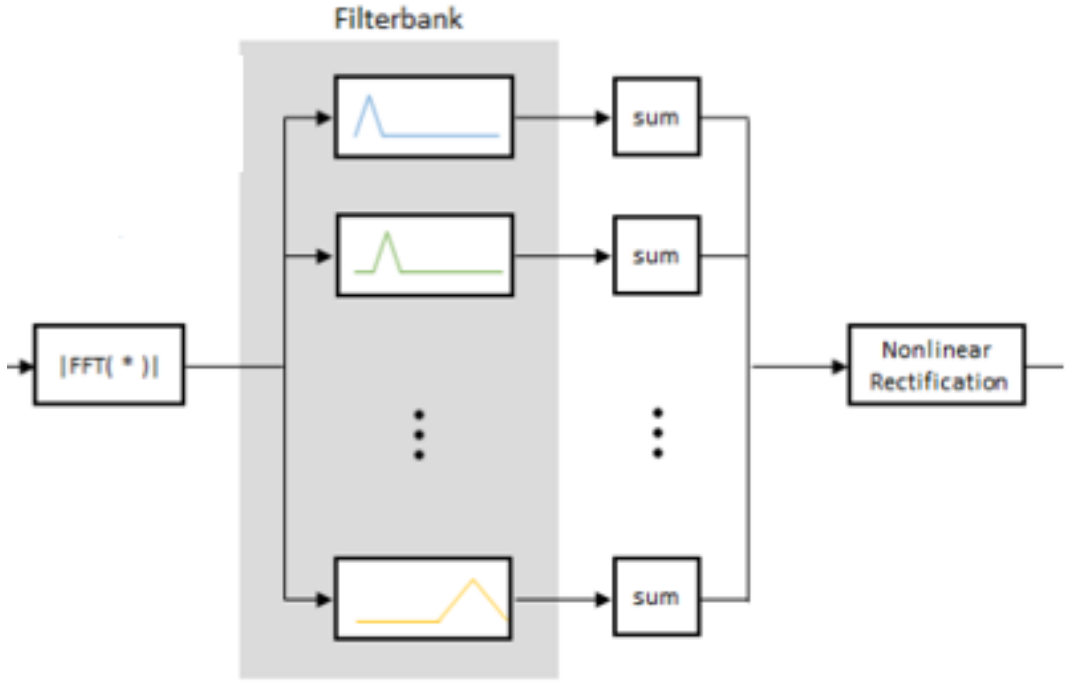
❖ خروجی فیلتر b ام برای فریم گفتاری m ام با دامنه طیف $|S_m[k]|$

$$y_b = \sum_{k=0}^{N/2} |S_m[k]| F_b[k], \quad b = 1, \ldots, Nf$$

❖ با فرض داشتن ۲۴ فیلتر (Nf=24)، تعداد ۲۴ خروجی فیلتربانک خواهیم داشت:

$$\bar{y} = \left[y_1, y_2, \ldots, y_{Nf} \right]^T$$

❖ گوش انسان شدت صوت را بصورت خطی دریافت نمی کند بلکه بصورت غیرخطی است. بنابراین یک تابع غیرخطی مانند لگاریتم بر روی خروجی فیلتربانک ها اعمال می شود.



Nf : تعداد فیلترها در بانک فیلتر

Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)

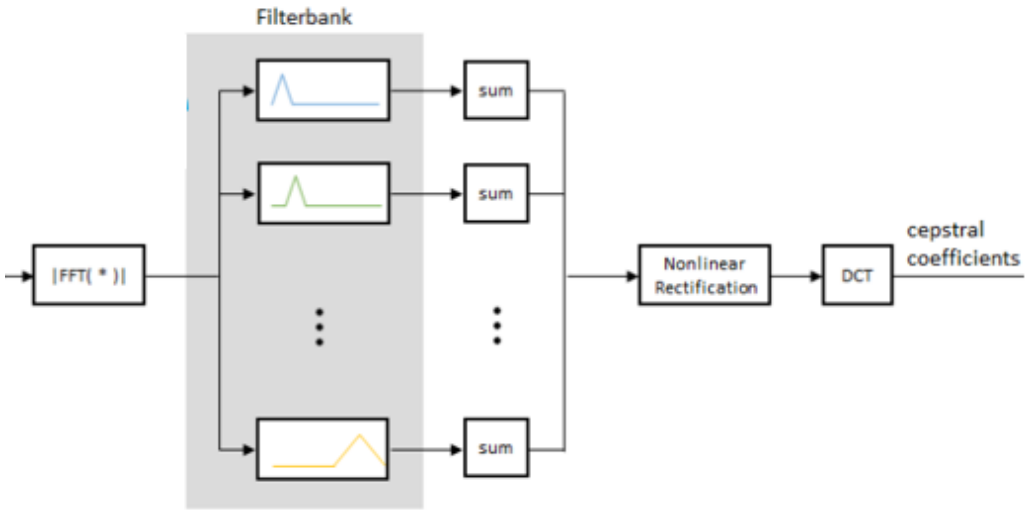
❖ یک ویژگی ساده: لگاریتم خروجی بانک فیلتر یعنی:

$$\log(\varepsilon + \bar{y}) = \left[\log(\varepsilon + y_1),\ \log(\varepsilon + y_2), \ldots, \log(\varepsilon + y_{Nf})\right]^T$$

❖ مشکل اصلی این بردار ویژگی:

❖ مولفه های این بردار با یکدیگر همبستگی (corrolation) دارند و بایستی توسط یک تبدیل متعامد برطرف شود.

❖ اعمال DCT بر روی بردار $\log(\varepsilon + \bar{y})$



$$\begin{bmatrix} \log(\varepsilon + y_1) \\ \log(\varepsilon + y_2) \\ \ldots \\ \log(\varepsilon + y_{Nf}) \end{bmatrix}_{Nf \times 1} \rightarrow DCT \rightarrow \begin{bmatrix} c_0 \\ c_1 \\ \ldots \\ c_{Nf-1} \end{bmatrix}_{Nf \times 1}$$

Hamidreza Baradaran Kashani

ضرایب کپستروم مبتنی بر مل (MFCC)

❖ تبدیل DCT

- ❖ تبدیل **DCT** یک تبدیل متعامد (**orthogonal**) است که اگر سیگنال ورودی حقیقی باشد، سیگنال خروجی حاصل از **DCT** (ضرایب **DCT**) هم حقیقی است، (برخلاف **DFT**)
- ❖ خاصیت فشرده سازی **DCT** بهتر است از خاصیت فشرده سازی تبدیل فوریه گسسته (**DFT**) است. به همین دلیل در فشرده سازهای سیگنالی **JPEG**، **MPEG** و **H.261** استفاده می شود.

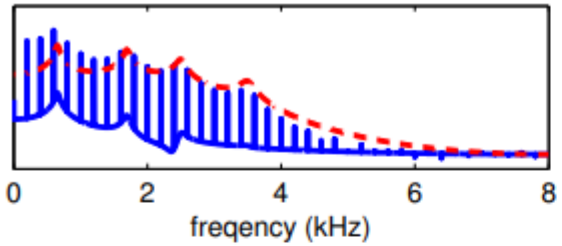
$$c_n = \sum_{b=1}^{Nf} \log(\varepsilon + y_b)\ \mathrm{Cos}\left(\frac{\pi n(b-0.5)}{Nf}\right), \quad 0 \le n \le Nc$$

- ❖ تعداد ضرایب **MFCC** یعنی **Nc** کمتر از تعداد فیلتر ها است. مثلا برای ۲۴ فیلتر از ۱۳ ضریب استفاده می کنند.
- ❖ در برخی موارد به جای ضریب **C0** از لگاریتم انرژی فریم **m** ام استفاده می کنند

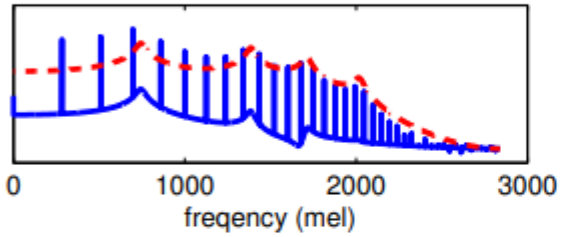
$$c_0 = \sum_{b=1}^{Nf} \log(\varepsilon + y_b) \qquad c_0 = \log\left(\varepsilon + \sum_{n=0}^{N-1} (s_m[n])^2\right)$$

Hamidreza Baradaran Kashani

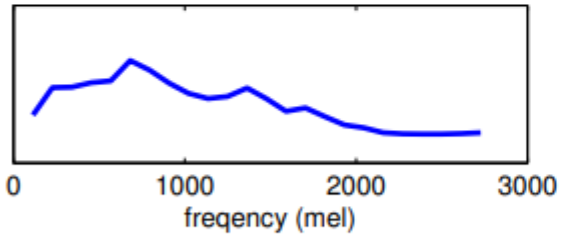
ضرایب کپستروم مبتنی بر مل (MFCC)



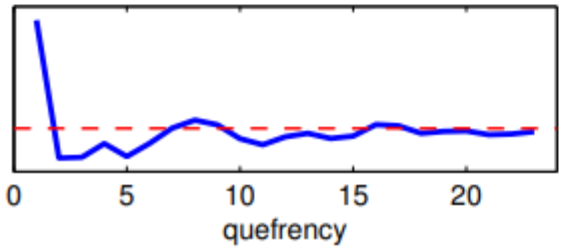
Linear to Mel frequency



Filterbank ($\sim$ 20-25 filters) + log()

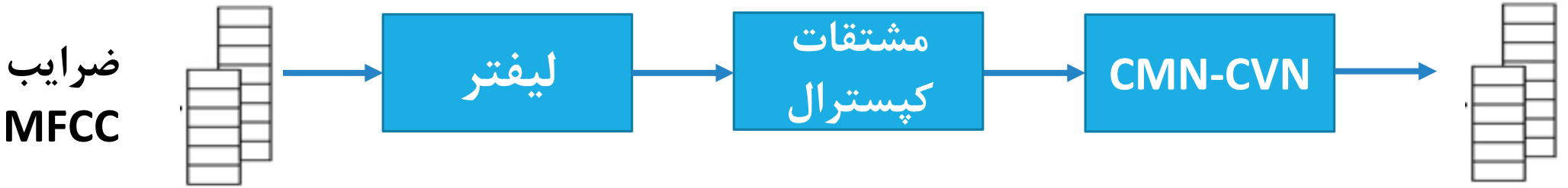


Discrete Cosine Transform



Hamidreza Baradaran Kashani

پس پردازش ضرایب MFCC

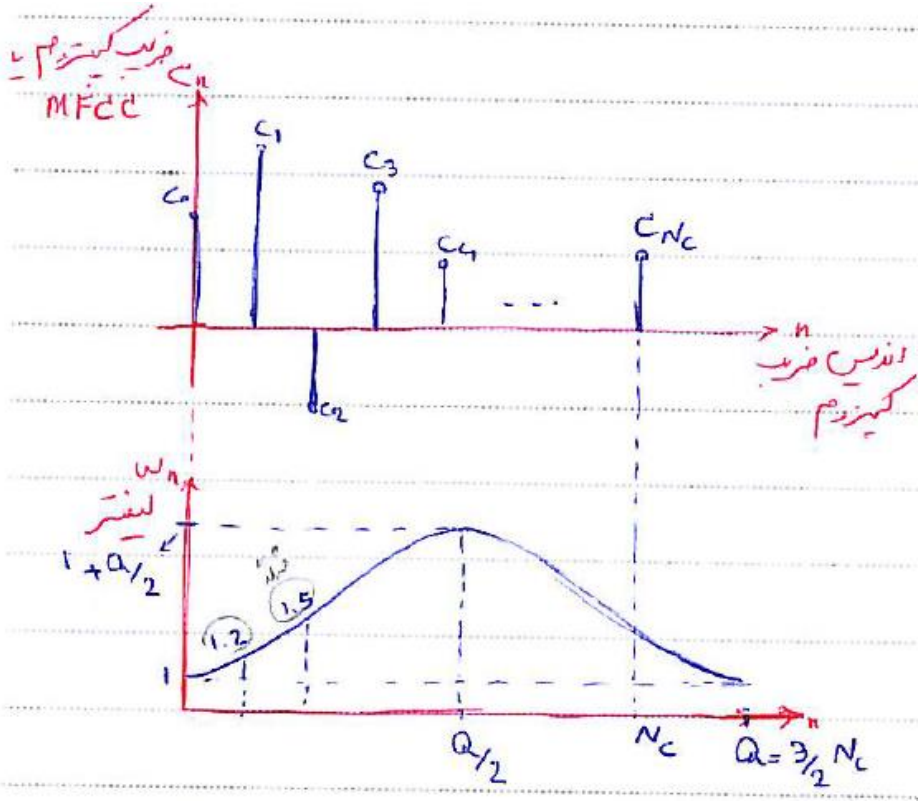


❖ اعمال لیفتر به بردار کپستروم

❖ برای هموار کردن لگاریتم طیف

$$w_n = 1 + \frac{Q}{2}\,\mathrm{Sin}\left(\frac{\pi n}{Q}\right)$$

$$c'_n = c_n\, w_n$$

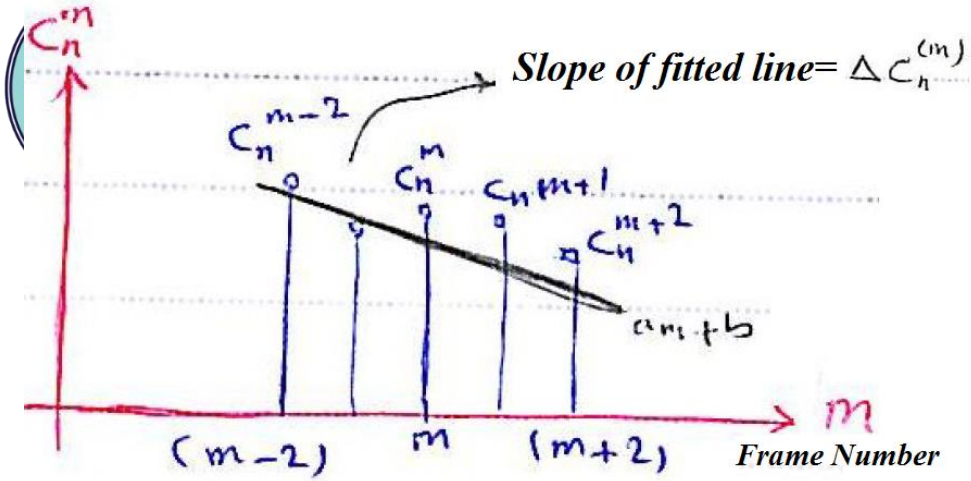


❖ با وجود DCT برای هموارکردن لگاریتم فیلتربانک نیازی به لیفتر احساس نمی شود.

❖ با استفاده از نرمالیزه سازی CMN-CVN عملا اثر لیفتر حذف می شود

Hamidreza Baradaran Kashani

پس پردازش ضرایب MFCC



$$\Delta c_n^{(m)} = \frac{\sum_{\tau=1}^{k} \tau \left(c_n^{(m+\tau)} - c_n^{(m-\tau)} \right)}{2 \sum_{\tau=1}^{k} \tau^2}$$

$$\boldsymbol{c}'^{(m)} = \begin{bmatrix} \boldsymbol{c}^{(m)} \\ \Delta \boldsymbol{c}^{(m)} \\ \Delta\Delta \boldsymbol{c}^{(m)} \end{bmatrix}_{(3*Nc \times 1)}$$

مشتقات کپسترال

- ضرایب کپسترال حاصل از یک فریم گفتاری، حاوی اطلاعات استاتیک گفتار هستند.
- مشتقات زمانی ضرایب کپسترال حاوی اطلاعات پویا و گذاری گفتار هستند.
- محاسبه مشتقات زمانی کپسترال روی فریم های زمانی مجاور
- مشتق اول از رابطه روبرو حاصل می شود.
- مشتق دوم با محاسبه رابطه روبرو بر روی مشتق اول نتیجه می شود و به همین ترتیب ...
- معمولا حداکثر از مشتقات اول و دوم (و گاهی مشتق سوم) استفاده می شود.
- بردار ویژگی کامل **MFCC** از اتصال بردار استاتیک و مشتقات آن حاصل می شود.

Hamidreza Baradaran Kashani

پس پردازش ضرایب MFCC

❖ نرمالیزه سازی ضرایب کپسترال به میانگین و واریانس

$$\underbrace{X_1(k)}_{\text{طیف گفتار اعوجاج یافته}} = \underbrace{H(k)}_{\substack{\text{مشخصه فرکانسی کانال تلفن} \\ \text{یا میکروفون}}} \underbrace{X(k)}_{\text{طیف گفتار تمیز}}$$

$$\underbrace{F^{-1}\left\{\log\left|X_1(k)\right|\right\}}_{\text{کپستروم گفتار اعوجاج یافته}} = \underbrace{F^{-1}\left\{\log\left|H(k)\right|\right\}}_{\text{کپستروم کانال}} + \underbrace{F^{-1}\left\{\log\left|X(k)\right|\right\}}_{\text{کپستروم گفتار تمیز}}$$

❖ اثر وسیله ضبط یا انتقال (نوع میکروفون، نوع کانال تلفن و نوع گوشی تلفن) تا حدود زیادی توسط حذف میانگین از بین می رود. کپستروم کانال در طی زمان ثابت است و با کسر میانگین از بردارهای کپستروم حذف می شود.

Cepstral Mean Subtraction or Normalization= CMS or CMN

Hamidreza Baradaran Kashani

پس پردازش ضرایب MFCC

❖ نرمالیزه سازی ضرایب کپسترال به میانگین و واریانس

$$\mu_n = \frac{1}{M}\sum_{m=1}^{M} c_n'^{(m)}$$

$$\sigma_n = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(c_n'^{(m)} - \mu_n\right)^2}$$

$$\tilde{c}_n^{(m)} = \frac{c_n'^{(m)} - \mu_n}{\sigma_n}$$

❖ M تعداد فریم های گفتاری (غیر سکوت) است.

❖ نرمالیزه سازی بردارهای ویژگی نسبت به انحراف معیار (واریانس)، باعث مقاوم شدن ویژگی نسبت به تغییر ولوم صدا و اضافه شدن نویز محیطی می شود.

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه زمان

Hamidreza Baradaran Kashani

استخراج ویژگی در حوزه زمان

- با فرض بخش بندی سیگنال زمانی گفتار به فریم های کوتاه مدت (short time) ویژگی های زیر را می توان استخراج کرد:
  - میانگین کوتاه مدت
  - انرژی کوتاه مدت
  - دامنه متوسط کوتاه مدت
  - نرخ عبور از صفر
  - اتوکورولیشن

Hamidreza Baradaran Kashani

میانگین و انرژی کوتاه مدت

❖ میانگین کوتاه مدت (short time average)

❖فرضا $s_m[n]$ فریم ***m*** ام سیگنال گفتار با ***N*** نمونه باشد، آنگاه میانگین کوتاه مدت بصورت زیر است:

$$\bar{s}_m = \frac{1}{N}\sum_{n=0}^{N-1} s_m[n]$$

❖ انرژی کوتاه مدت (short time energy)

$$E_m = \frac{1}{N}\sum_{n=0}^{N-1} (s_m[n])^2$$

❖برخی اوقات میانگین یک فریم از سیگنال کسر شده و سپس انرژی محاسبه می شود:

$$s'_m[n] = s_m[n] - \bar{s}_m \qquad E_m = \frac{1}{N}\sum_{n=0}^{N-1} (s'_m[n])^2$$

Hamidreza Baradaran Kashani

میانگین و انرژی کوتاه مدت

❖ انرژی کوتاه مدت (short time energy) – ادامه

❖ برخی واقع به جای انرژی از لگاریتم انرژی استفاده می شود:

$$E_m^{dB} = 20\log_{10}\left(\varepsilon + E_m\right)$$

❖ در برخی مقالات در محاسبه انرژی تقسیم بر تعداد N نمونه را ندارند:

$$E_m = \sum_{n=0}^{N-1} \left(s'_m[n]\right)^2$$

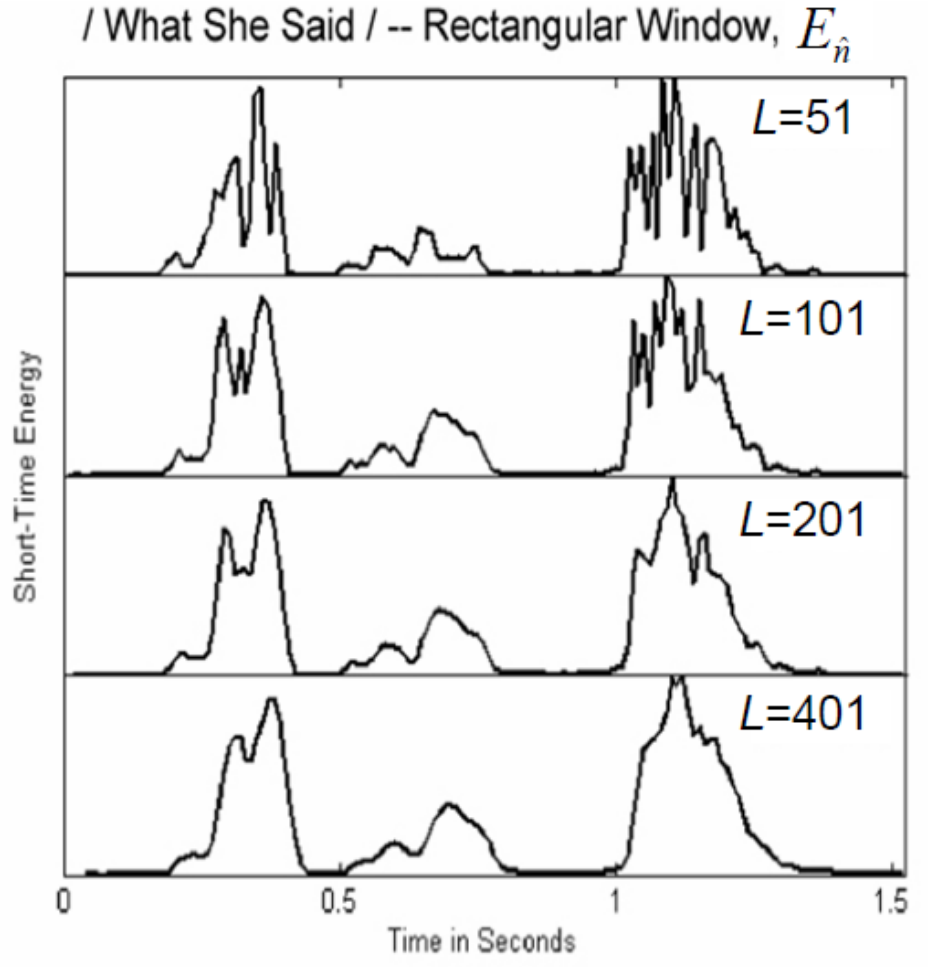
❖ انرژی در فریم های واکدار معمولا بیشتر از فریم های بیواک است.

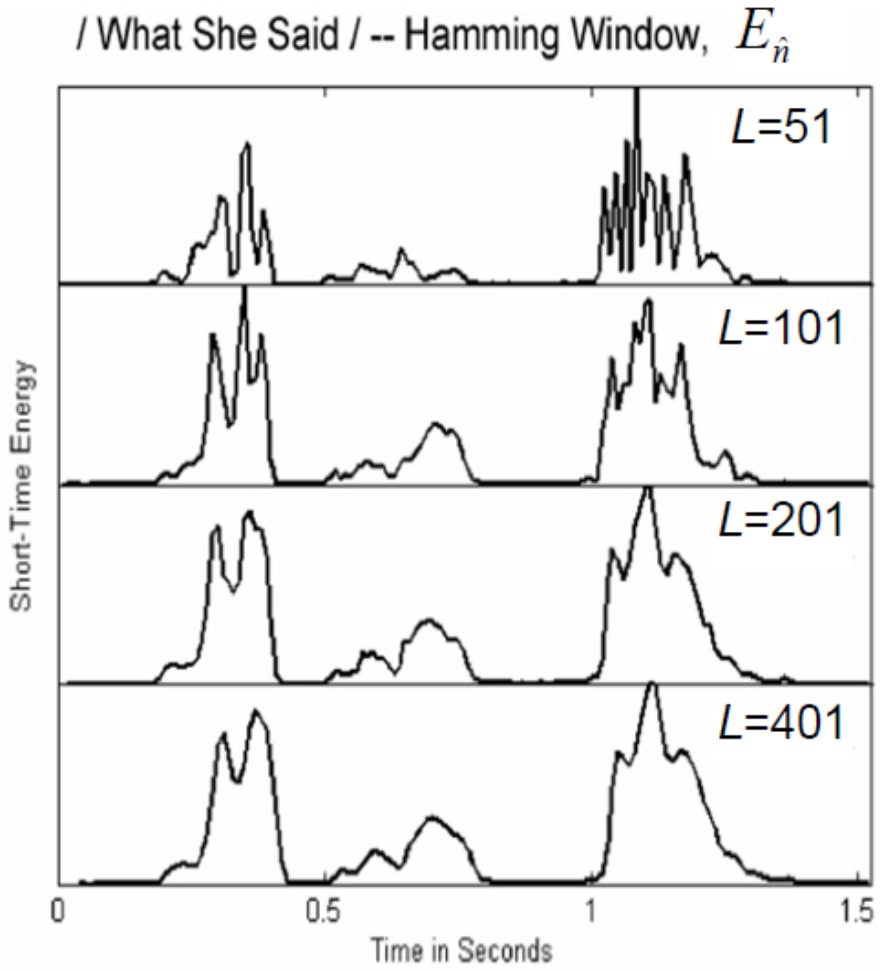
❖ انرژی در واکه ها معمولا بیشتر از همخوان ها است.

❖ برای گفتارهایی با سطح سیگنال به نویز (SNR) متوسط و بالا، ویژگی انرژی کوتاه مدت یک ویژگی ساده و در عین حال کارا جهت استخراج نواحی گفتاری از نواحی سکوت (silence) یا نویزهای ضعیف پس زمینه است.

Hamidreza Baradaran Kashani

میانگین و انرژی کوتاه مدت



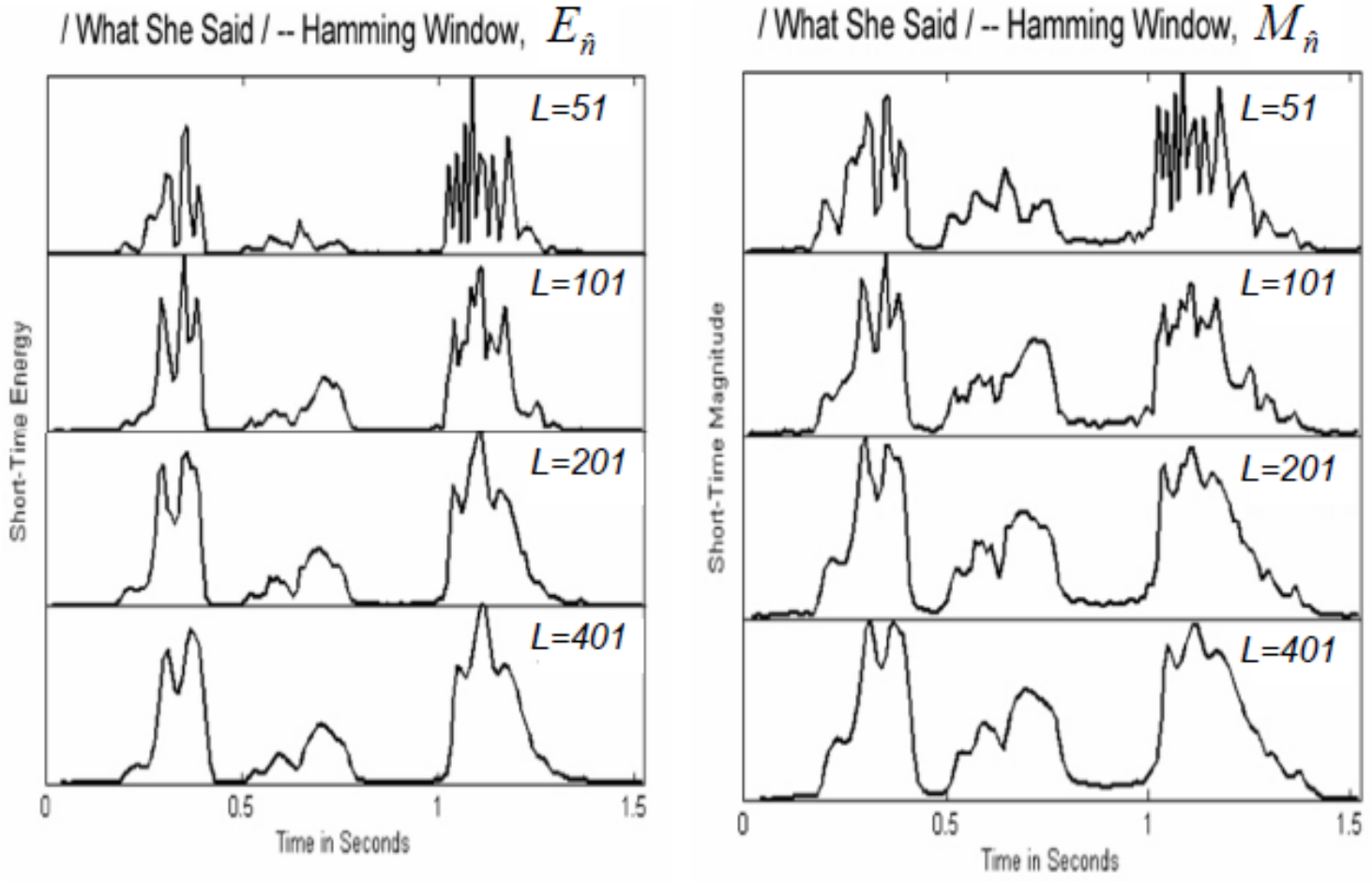


Hamidreza Baradaran Kashani

دامنه متوسط کوتاه مدت

❖ دامنه متوسط کوتاه مدت (short time average magnitude)

$$A_m = \frac{1}{N}\sum_{n=0}^{N-1}\left|s'_m[n]\right|$$



nidreza Baradaran Kashani

نرخ عبور از صفر (ZCR)

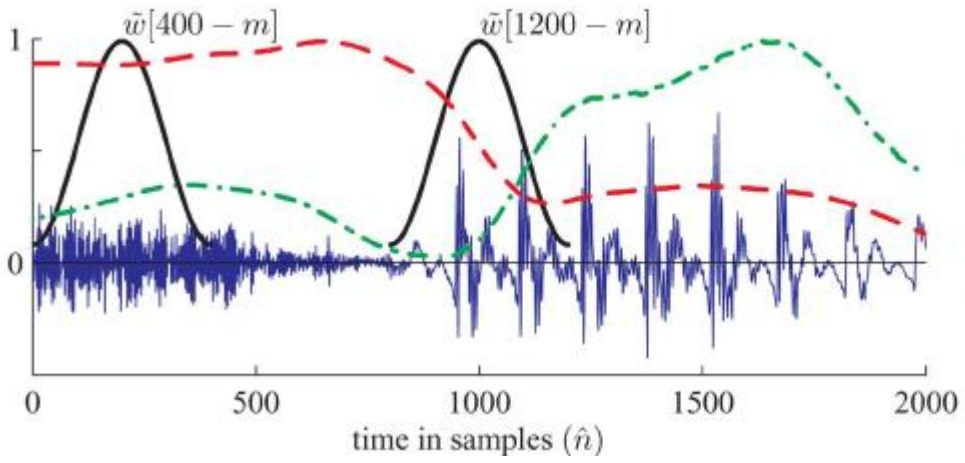
❖ تعداد دفعات عبور از صفر در فریمی که **N** نمونه دارد

$$ZCR_m = \frac{1}{N}\sum_{n=1}^{N-1}\frac{\left|\operatorname{sgn}\left(s'_m[n]\right)-\operatorname{sgn}\left(s'_m[n-1]\right)\right|}{2}$$

$$\operatorname{sgn}(x)=\begin{cases}+1, & if \quad x >= 0\\ -1, & if \quad x < 0\end{cases}$$

❖ هر چه **F0** بیشتر باشد، **ZCR** بیشتر است.

❖ **ZCR** برای واج های بیواک بیشتر از سکوت و واج های واکدار است.



Hamming window with duration *L=401* samples (25 msec at *Fs*=16 kHz)

منحنی قرمز: **ZCR**

منحنی سبز: انرژی

Hamidreza Baradaran Kashani

تابع اتوکورولیشن

❖ اتوکورولیشن به محاسبه شباهت سیگنال با شیفت یافته های آن می پردازد.

$$Cross\ corrolaion: \quad R^{xy}[k] = \sum_{n=0}^{N-1-k} x[n]\, y[n+k]$$

$$Auto-correlaion: \quad R_m^{ss}[k] = R_m[k] = \sum_{n=0}^{N-1-|k|} s'_m[n]\; s'_m[n+|k|]$$

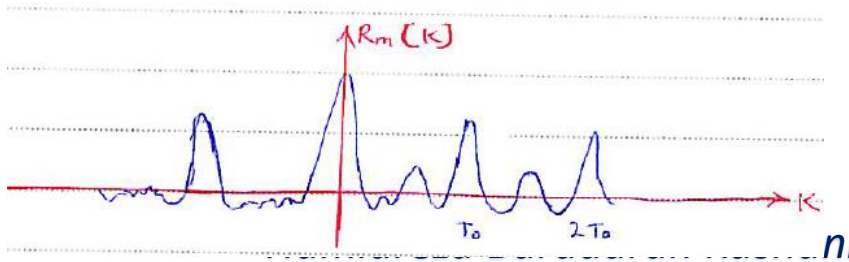
در محاسبه تابع اتوکورولیشن بهتر است میانگین فریم ابتدا کم شود.

❖ نکته: $R_m[k] = R_m[-k]$

❖ تبدیل فوریه $R_m[k]$ همیشه مبت بوده و طیف توان (**power spectrum**) نام دارد

$$P_m[k] = |X_m[k]|^2 = FFT\{R_m[k]\}$$

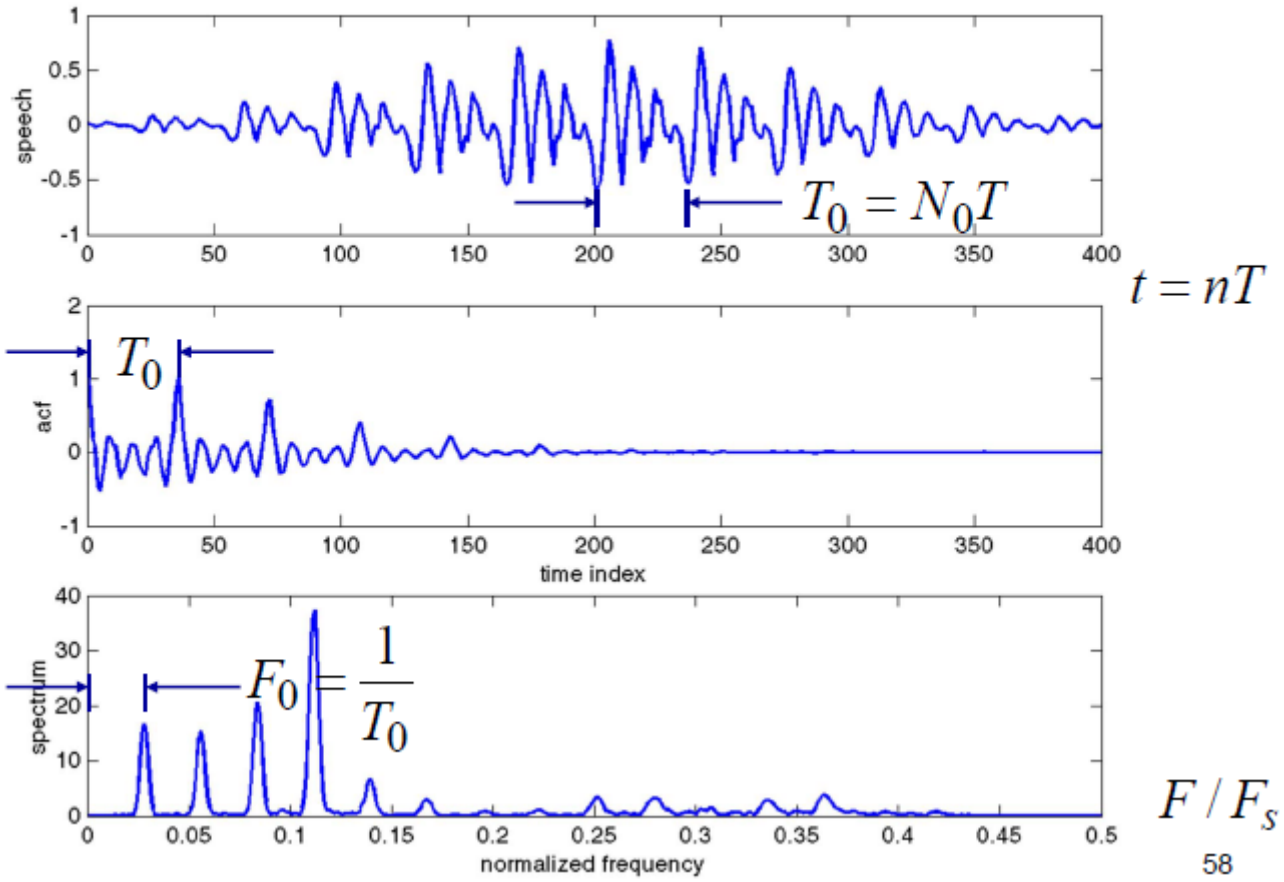
$$k = -(N-1), -(N-2), \ldots, -2, -1, 0, 1, 2, \ldots, (N-1), (N-2)$$



تابع اتوکورولیشن

در اینجا قبل از محاسبه اتوکورولیشن فریم را دریک پنجره همینگ ضرب کرده است.

Voiced (female) *L=401* (magnitude)



58

Hamidreza Baradaran Kashani

تابع اتوکورولیشن

❖ اتوکورولیشن نرمالیزه (Normalized Autocorrelation Function) یا NACF

$$r_m[k] = \frac{R_m[k]}{R_m[0]}, \quad -1 \le r_m[k] \le 1$$

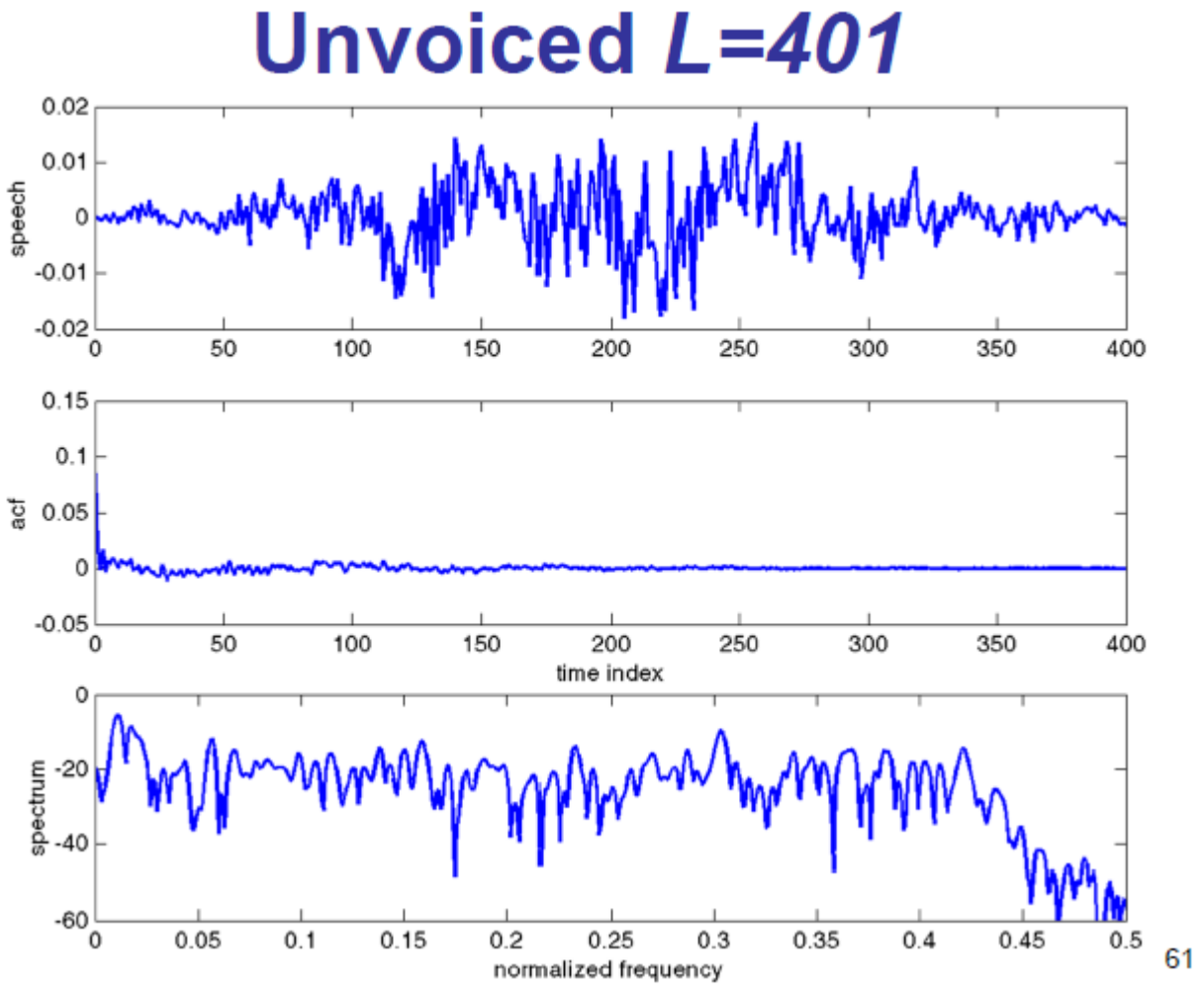
❖ استفاده از NACF برای تصمیم گیری در مورد واکدار یا بی واک بودن

$$r_m[1] = \frac{R_m[1]}{R_m[0]} = \frac{\sum_{n=0}^{N-1-1} s'_m[n]\, s'_m[n+1]}{\sum_{n=0}^{N-1} s'_m[n]\, s'_m[n]}$$

$$\begin{aligned} &if \;\; r_m[1] > thr: \\ &\quad frame\ is\ voiced \\ &else: \\ &\quad frame\ is\ unvoiced \end{aligned}$$

Hamidreza Baradaran Kashani

تابع اتوکورولیشن



Hamidreza Baradaran Kashani

با تشکر از اساتید و همکاران گرامی:

آقای دکتر همایونپور

آقای دکتر کبودیان