۱– مقدمه

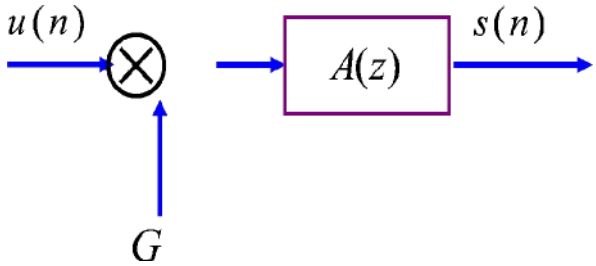
آشنایی با نحوه محاسبه ضرایب LPC

آشنایی با روش محاسبه ضرایب کپسترال LPC

۲– محاسبه ضرایب پیشگویی خطی

همان طور که در بخش قبل گفته شد از این ضرایب برای مدل کردن گفتار استفاده می شود.

به این صورت که مدل سیگنال تحریک (قطار ضربه) را به مدل مسیر صوتی (مدل AR) اعمال کرده و خروجی صوتی را می گیریم. این امر به صورت خیلی خلاصه در تصویر ۱ نشان داده شده است.



تصویر ۱– مدل سازی گفتار بوسیله مدل AR به صورت خیلی ساده

همان طور که گفته شد، مدل AR فرض می کند که سیگنال در زمان n از ترکیب خطی p نمونه قبلی سیگنال محاسبه می شود.

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \ldots + a_p s(n-p),$$

با خلاصه کردن و اضافه کردن عبارت تحریک (Gu(n)):

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + Gu(n),$$

با تبدیل z گرفتن از طرفین:

$$S(z) = \sum_{i=1}^{p} a_i z^{-i} S(z) + GU(z)$$

که در نهایت پاسخ سیستم به صورت روبرو به دست می آید:

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} = \frac{1}{A(z)}.$$

سیگنال اصلی برابر روبر است:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n).$$

سیگنال تخمین زده شده برابر روبرو است:

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k s(n-k).$$

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum^{p} a_k s(n-k)$$

در نتیجه خطا به صورت روبرو به دست می آید:

و A(z) هم به صورت روبرو می باشد:

معادلات آنالیز LPC

در صورتی که دو فرض روبرو را بکنیم:

$$S_n(m) = s(n+m)$$
$$e_n(m) = e(n+m)$$

هدف کمینه کردن سیگنال متوسط مربعات خطا می باشد:

$$E_n = \sum_m e_n^2(m)$$

یا به عبارتی:

$$E_n = \sum_m \left[s_n(m) - \sum_{k=1}^{p} a_k s_n(m-k) \right]^2.$$

برای کمینه شدن خطا En باید مشتق آن نسبت به پارامترها (ak) صفر باشد:

$$\frac{\partial E_n}{\partial a_k} = 0, \qquad k = 1,2,\ldots,p$$

که به دست می آید:

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^{p} \hat{a}_k \sum_m S_n(m-i)S_n(m-k)$$

در صورتی که از این نماد استفاده شود:

$$\phi_n(i,k) = \sum_m S_n(m-i)S_n(m-k)$$

خواهیم داشت:

$$\phi_n(i,0) = \sum_{k=1}^{p} \hat{a}_k \, \phi_n(i,k) \qquad i = 1,2,\ldots,p$$

که مجموعه ای از p معادله و p مجهول می باشد.

کمترین مقدار متوسط مربعات خطا به صورت روبرو به دست می آید:

$$\hat{E}_n = \sum_m s_n^2(m) - \sum_{k=1}^{p} \hat{a}_k \sum_m s_n(m)s_n(m-k)$$
$$= \phi_n(0,0) - \sum_{k=1}^{p} \hat{a}_k \, \phi_n(0,k).$$

دو روش مهم برای حل این معادله ارائه شده است:

١. **روش خودهمبستگی (Autocorrelation):**

در صورتی که تعریف روبرو را در نظر بگیریم:

$$s_n(m)=\begin{cases} s(m+n).w(m), & 0\le m\le N-1 \\ 0, & otherwise. \end{cases}$$

متوسط مربعات خطا برابر است با:

$$E_n=\sum_{m=0}^{N-1+p} e_n^2(m)$$

$$\phi_n(i,k)=\sum_{m=0}^{N-1+p} s_n(m-i)s_n(m-k), \qquad \begin{matrix} 1\le i\le p \\ 0\le k\le p \end{matrix}$$

$$\phi_n(i,k)=\sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k), \qquad \begin{matrix} 1\le i\le p \\ 0\le k\le p \end{matrix}.$$

به این دلیل که $\phi_n(i,k)$ تنها تابعی از i-k می باشد، تابع کواریانس به یک تابع ساده خودهمبستگی کاهش می یابد:

$$\phi_n(i,k)=r_n(i-k)$$

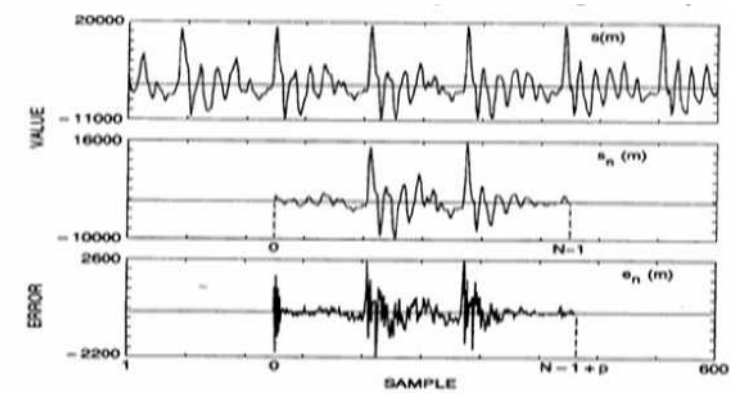
چون تابع خودهمبستگی متقارن است:

$$r_n(-k)=r_n(k)\rightarrow \sum_{k=1}^{p} r_n(|i-k|)\hat{a}_k=r_n(i), \qquad 1\le i\le p$$

$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & ...r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & ...r_n(p-2) \\ r_n(2) & r_n(1) & r_n(0) & ...r_n(p-3) \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & ... \ r_n(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} r_n(1) \\ r_n(2) \\ r_n(3) \\ r_n(p) \end{bmatrix}.$$

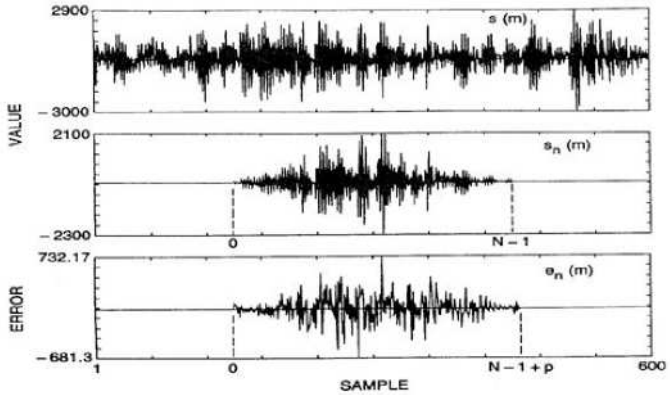
با حل معادله بالا که p معادله و p مجهول است (p مجهول همان ak ها هستند) ضرایب به دست می آیند.

در تصویر ۲ نمونه ای از یک سیگنال صدادار و خطای پیشگویی آن را مشاهده می کنید.



تصویر ٢ – نمونه ای از یک سیگنال صدادار و خطای پیشگویی آن

در تصویر ٣ نمونه ای از یک سیگنال بدون صدا و خطای پیشگویی آن را مشاهده می کنید.



تصویر ٣ – نمونه ای از یک سیگنال بدون صدا و خطای پیشگویی آن

٢. روش کوواریانس:

در این روش بازه محاسبه خطا را ١-N=>m=>٠ در نظر می گیریم.

سپس از گفتار بدون وزن دهی مستقیماً استفاده می کنیم.

$$E_n = \sum_{m=0}^{N-1} e_n^2(m)$$

و $\phi_n(i,k)$ به یکی از دوصورت روبرو تعریف می شود:

$$\phi_n(i,k) = \sum_{m=0}^{N-1} s_n(m-i)s_n(m-k), \qquad \begin{matrix} 1 \le i \le p \\ 0 \le k \le p \end{matrix}$$

$$\phi_n(i,k) = \sum_{m=-i}^{N-i-1} s_n(m)s_n(m+i-k), \qquad \begin{matrix} 1 \le i \le p \\ 0 \le k \le p \end{matrix}.$$

نهایتاً به فرمول زیر می رسیم:

$$\begin{bmatrix} \phi_n(1,1) & \phi_n(1,2) & \phi_n(1,3) & \cdots & \phi_n(1,p) \\ \phi_n(2,1) & \phi_n(2,2) & \phi_n(2,3) & \cdots & \phi_n(2,p) \\ \phi_n(3,1) & \phi_n(3,2) & \phi_n(3,3) & \cdots & \phi_n(3,p) \\ \vdots & \vdots & \vdots & & \vdots \\ \phi_n(p,1) & \phi_n(p,2) & \phi_n(p,3) & \cdots & \phi_n(p,p) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} \phi_n(1,0) \\ \phi_n(2,0) \\ \phi_n(3,0) \\ \vdots \\ \phi_n(p,0) \end{bmatrix}.$$

ماتریس کوواریانس نتیجه شده متقارن است ولی Toeplitz نیست.

می توان این معادله را بوسیله یک سری روش ها به نام تجزیه Cholesky حل کرد.

۳. روش لوینسون–دوربین

فرمول های زیر به صورت جلو رونده محاسبه می شود:

$$E^{(0)} = r(0)$$

$$k_i = \left\{ r(i) - \sum_{j=1}^{L-1} \alpha_j^{(i-1)} r(|i-j|) \right\} \Big/ E^{(i-1)}, \quad (*) \qquad 1 \le i \le p$$

$$\alpha_i^{(i)} = k_i$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}$$

$$E^{(i)} = (1-k_i^2) E^{(i-1)},$$

note: the summation in (*) is ommitted for $i = 1$

به ak ها ضرایب پیشگویی، به kiها ضرایب PARCOR و به gm که در زیر تعریف می شود ضرایب لگاریتم گفته می شود.

$$g_m = \text{log area ratio coefficients} = \log\left(\frac{1-k_m}{1+k_m}\right).$$

این ضرایب قابل تبدیل شدن به هم هستند. یعنی با داشتن یک سری از آن ها می توان سری دیگر را یافت.

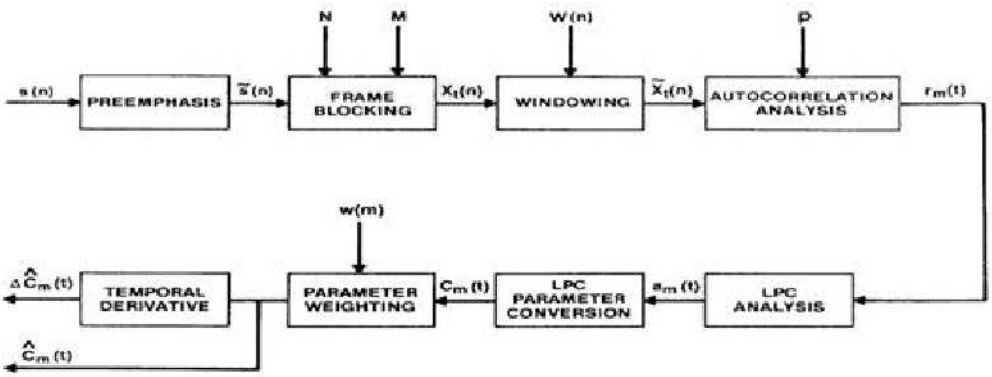
در نتیجه می توان مدل را با هر کدام از این ضرایب به عنوان پارامتر مدل کرد.

۴– ضرایب کپسترال پیشگویی خطی

برای کاربردهای بازشناسی گفتار از پارامترهای LPC ضرایب کپسترال استخراج می شود.

همان طور که در فصل پیش دیدیم با بردن ویژگی ها به حوزه مپستروم، ویژگی های نهایی تقریباً غیرهمبسته می شوند.

خلاصه این کار در تصویر ٤ نشان داده شده است.



تصویر ٤– خلاصه استخراج ضرایب کپسترال پیشگویی خطی

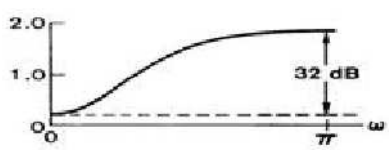
شرح بلوک های تصویر ٤ به صورت زیر است:

- **پیش تاکید:** معمولاً یک فیلتر مشتق گیر می باشد اثر فرکانس های بالا را بیشتر می کند تا بتواند با فیلتر محیط که فرکانس های بالا را تضعیف می کند مقابله کند.

$$H(z) = 1 - \tilde{a} z^{-1}, \qquad 0.9 \le a \le 1.0.$$

$$\tilde{s}(n) = s(n) - \tilde{a} s(n-1).$$

تصویر پاسخ فرکانسی یک فیلتر پیش تاکید در تصویر ٥ آمده است.



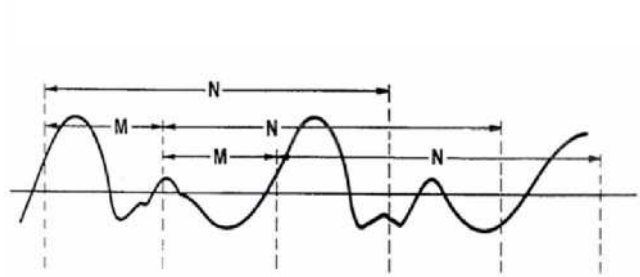
تصویر ٥– فیلتر پیش تاکید با ضریب ٠.٩٥

- **فریم بندی:** مانند دیگر کاربردهای پردازش گفتار ابتدا سیگنال به فریم های کوتاهی تقسیم می شود.

اگر بخواهیم این عمل را فرموله کنیم به صورت روبرو به دست می آید.

$$x_\ell(n) = \tilde{s}(M\ell + n), \qquad \begin{array}{l} n = 0,1,\ldots,N-1 \\ \ell = 0,1,\ldots,L-1. \end{array}$$

خلاصه این عمل را در تصویر ٦ مشاهده می کنید.



تصویر ٦– خلاصه عمل فریم بندی

- **پنجره گذاری:** پس از فریم بندی روی هر فریم یک پنجره ضرب می کنیم تا مرزهای فریم را به سمت صفر سوق دهد تا گسستگی در سیگنال کم شود.

$$\tilde{x}_\ell(n) = x_\ell(n)w(n), \qquad 0 \le n \le N-1.$$

پنجره هایی مثل همینگ و هنینگ استفاده می شوند که در مرز ها به صفر متمایل می شوند.

- **تحلیل خودهمبستگی:** همان طور که گفته شد برای محاسبه ضرایب از تحلیل خودهمبستگی استفاده می شود.
- **تحلیل پیشگویی خطی:** به روش لوینسون–دوربین ضرایب پیشگویی محاسبه می شوند.
- **انتقال به حوزه کپستروم:** بوسیله فرمول زیر و به صورت جلو رونده ضرایب کپسترال LPC محاسبه می شوند.

$$c_0 = \ln \sigma^2 \qquad \sigma^2 \text{ is the gain term in LPC model}$$

$$c_m = a_m + \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k}, \qquad 1 \le m \le p$$

$$c_m = \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k}, \qquad m > p,$$

- **وزن دهی پارامترها:** ضرایب کپسترال درجه پایین به شیب کلی طیف وابسته اند.

  درجات بالای ضرایب کپسترال بیشتر به نویز حساس اند.

  وزن دهی طوری انجام می شود که این حساسیت ها برطرف شود.
- **پردازش زمانی:** همان طور که در فصل گذشته توضیح داده شد، ضرایب دلتا و دلتادلتا به بردار ویژگی اضافه می شوند.

در تصویر ٧ مقادیر معمول تحلیل LPC را مشاهده می کنید.

فرض کنید که:

- N تعداد نمونه های هر فریم
- M تعداد شیفت های بین فریم ها
- P درجه تحلیل LPC
- Q بعد بردار کپسترال LPC
- K تعداد فریم هایی که مشتق های زمانی بر روی آن ها محاسبه می شوند

Parameter	Fs=6.67kHz	Fs=8kHz	Fs=10kHz
N	300 (45 msec)	240 (30 msec)	300 (30 msec)
M	100 (15 msec)	80 (10 msec)	100 (10 msec)
p	8	10	10
Q	12	12	12
K	3	3	3

تصویر ۷–مقادیر معمول تحلیل LPC برای بازشناسی گفتار با توجه به فرکانس نمونه برداری

۵– خلاصه و نتیجه گیری

در انی فصل با مفهوم ضرایب پیشگویی خطی آشنا شدیم.

در این فصل با نحوه محاسبه ضرایب پیشگویی خطی آشنا شدیم.

در این فصل با نحوه محاسبه ضرایب کپسترال پیشگویی خطی آشنا شدیم.

۶ – منابع درس

۱- Rabiner, “Fundumentals of Speech Recognition”
۲- Huang, Acero, “Spoken Language Processing”
۳- Deller, “Discrete-time processing of speech signals”