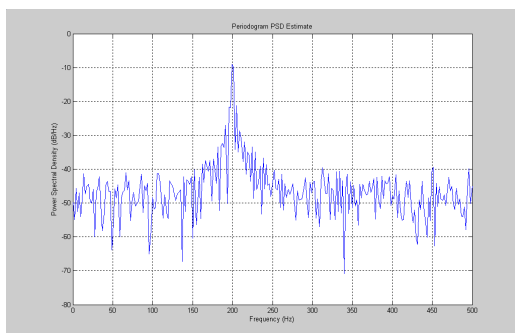


ANALIZA WIDMOWA c.d.

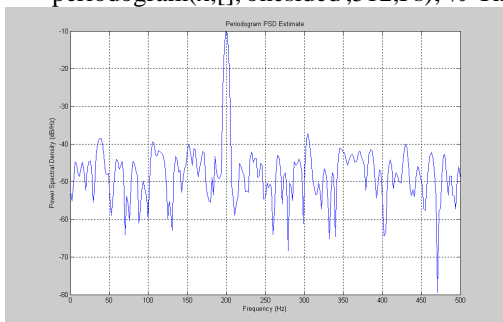
- **Momenty jako parametry zmiennych losowych**
- **Momenty widmowe**
- **Identyfikacja mówcy z zastosowaniem momentów widmowych**
- **Wysokorozdzielcza parametryczna estymacja WGM – widma gęstości mocy**
- **Estymacja WGM na podstawie modelu AR**
- **Pytania z analizy widmowej dla Akustyki Mowy**
- **Zastosowania momentów widmowych**

Na poniższych wykresach widzimy periodogramy otrzymane w MATLABie. Drugim z nich jest periodogram zmodyfikowany z oknem von Hanna. Zwraca on estymatę (1.33) lub (1.34).



```
>> Fs = 1000; t = 0:1/Fs:.3;
```

```
x = cos(2*pi*t*200)+0.1*randn(size(t)); % A cosine of 200Hz plus noise
periodogram(x,[],'onesided',512,Fs); % The default window is used
```



```
>> Fs = 1000; t = 0:1/Fs:.3;
```

```
x = cos(2*pi*t*200)+0.1*randn(size(t)); % A cosine of 200Hz plus noise
periodogram(x,hanning(301),'onesided',512,Fs); % The hanning window is used
```

Na podstawie definicji WGM jako DTFT z ciągu autokorelacji (1.29) estymatorem widnowym w postaci **korelogramu** jest

$$\hat{P}_c(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{r}[k] e^{-j\omega k} \quad (1.35)$$

gdzie $\hat{r}[k]$ oznacza estymatę $r[k]$ otrzymaną na podstawie dostępnych próbek $\{x[1], x[2], \dots, x[N]\}$. Jeżeli o tym sygnale zakładamy jedynie to, że jest on stacjonarny, to możemy znaleźć $\hat{r}[k]$ potrzebne w (1.35) dwoma sposobami: jako estymator standardowy nieobciążony

$$\hat{r}[k] = \frac{1}{N-k} \sum_{n=k+1}^N x[n]x^*[n-k], \quad 0 \leq k \leq N-1 \quad (1.36)$$

i jako estymator standardowy obciążony (lepszy dla średnich i dużych wartości parametru k)

$$\hat{r}[k] = \frac{1}{N} \sum_{n=k+1}^N x[n]x^*[n-k], \quad 0 \leq k \leq N-1 \quad (1.37)$$

przy czym, dla indeksów ujemnych, korzystamy z właściwości

$$\hat{r}[-k] = \hat{r}^*[k] \quad (1.38)$$

W praktyce obliczenia periodogramu i korelogramu wykonuje się za pomocą FFT zamiast DTFT, wymagającej continuum wartości częstotliwości. Stosuje się też uzupełnianie ciągu $\{x[n]\}_{n=1}^N$ zerami, gdy N nie jest potęgą liczby 2.

Momenty jako parametry zmiennych losowych

W poniższych wzorach zastosowano następujące oznaczenia [3]

- ξ – zmienna losowa,
- $f_{\xi}(x)$ – funkcja gęstości prawdopodobieństwa zmiennej losowej ξ ,
- x – wartość zmiennej losowej w zbiorze $\xi(\mathcal{E}) \subseteq R$ (inkluzyja zbiorów),
- \mathcal{E} – zbiór zwany przestrzenią wyników doświadczenia lub przestrzenią zdarzeń elementarnych; elementy $e \in \mathcal{E}$ tego zbioru są nazywane wynikami doświadczenia lub zdarzeniami elementarnymi.

Moment zwykły rzędu r zmiennej losowej ξ to liczba

$$\overline{\xi^r} = \int_{-\infty}^{\infty} x^r f_{\xi}(x) dx$$

Moment centralny rzędu r zmiennej losowej ξ to liczba

$$\overline{(\xi - \bar{\xi})^r} = \int_{-\infty}^{\infty} (x - \bar{\xi})^r f_{\xi}(x) dx$$

gdzie $\bar{\xi} = \overline{\xi^1}$ to moment zwykły rzędu 2. W przypadku zmiennej losowej dyskretnej ξ o skończonej liczbie realizacji $x_k; k = 0, 1, \dots, N-1$; powyższe wyrażenia przyjmują postać

$$\overline{\xi^r} = \sum_{k=0}^{N-1} x_k^r P_{\xi}(x_k)$$

i

$$\overline{(\xi - \bar{\xi})^r} = \sum_{k=0}^{N-1} (x_k - \bar{\xi})^r P_{\xi}(x_k)$$

gdzie $\{P_{\xi}(x_k); k = 0, 1, \dots, N-1\}$ to zbiór prawdopodobieństw przyporządkowanych tym realizacjom.

Momenty unormowane otrzymujemy dzieląc momenty zwykłe i centralne przez $\overline{\xi^0} = \overline{\xi^r} \Big|_{r=0}$. Momentami najważniejszymi i najczęściej wykorzystywanymi do opisu

zmiennych losowych są:

– $\overline{\xi}$ – moment zwykły rzędu 1, nazywany wartością oczekiwaną, wartością średnią lub nadzieją matematyczną zmiennej losowej ξ i oznaczany również $E(\xi)$ lub m_ξ , lub $M(0)$,

– $\overline{\xi^2}$ – moment zwykły rzędu 2, nazywany wartością średniokwadratową zmiennej losowej ξ i oznaczany także $E(\xi^2)$,

– $\overline{(\xi - \overline{\xi})^2}$ – moment centralny rzędu 2 lub inaczej wariancją zmiennej losowej ξ i oznaczany najczęściej σ_ξ^2 .

Często też wykorzystywany jest również parametr $\sqrt{\sigma_\xi^2}$ – pierwiastek z wariancji oznaczany przez σ_ξ , nazywany odchyleniem standardowym (dyspersją) zmiennej losowej ξ . (Dyspersja – inaczej rozrzut.)

Momenty widmowe

Momenty widmowe to **deskryptory** cech sygnału. Wykorzystuje się je do badania **podobieństwa** pomiędzy różnymi obiektami. Oblicza się je najczęściej na podstawie estymaty WGM widma sygnału.

Aby zdefiniować momenty widmowe, w powyższych wzorach zastępujemy x_k przez częstotliwość f_k , $k = 0, 1, \dots, N-1$ oraz $P_\xi(x_k)$ przez widmo gęstości mocy P_k lub jego estymatę \hat{P}_k , liczone za pomocą N -punktowej DFT (FFT). Wówczas, wprowadzając oznaczenie

$$M(r) = \overline{\xi^r} \Big|_{f_k = x_k, P_k = P_\xi(x_k) \text{ lub } \hat{P}_k = P_\xi(x_k)} \quad (1.39)$$

otrzymujemy moment widmowy r -tego rzędu

$$M(r) = \sum_{k=0}^{N-1} P_k f_k^r \quad (1.40)$$

i moment widmowy unormowany r -tego rzędu

$$M_u(r) = \frac{M(r)}{M(0)} \quad (1.41)$$

gdzie $M(0) = \sum_{k=0}^{N-1} P_k$ to moment widmowy rzędu zerowego mający zastosowanie normalizujące i oznacza moc sygnału, i moment widmowy unormowany centralny r -tego rzędu

$$M_{uc}(r) = \sum_{k=0}^{N-1} \frac{P_k [f_k - M_u(1)]^r}{M(0)} \quad (1.42)$$

gdzie

$$M_u(1) = \sum_{k=0}^{N-1} \frac{P_k f_k}{M(0)} \quad (1.43)$$

to moment widmowy unormowany pierwszego rzędu używany do obliczania momentów widmowych centralnych wyższych rzędów – ma interpretację środka ciężkości widma.

A teraz dalsze szczególne przypadki momentów widmowych. Moment widmowy unormowany centralny drugiego rzędu – ma interpretację kwadratu szerokości pasma częstotliwości zajmowanych przez sygnał

$$M_{uc}(2) = \sum_{k=0}^{N-1} \frac{P_k [f_k - M_u(1)]^2}{M(0)} \quad (1.44)$$

Moment widmowy unormowany centralny trzeciego rzędu to niesymetria widma, inaczej skośność (ang. *skewness*)

$$M_{uc}(3) = \sum_{k=0}^{N-1} \frac{P_k [f_k - M_u(1)]^3}{M(0)} \quad (1.45)$$

Parametr widmowy będący miarą spłaszczenia widma (popularnie kurtoza)

$$kurtosis = \frac{M_{uc}(4)}{[M_{uc}(2)]^2} \quad (1.46)$$

Można też inaczej zapisać

$$kurtosis = \frac{1}{N} \sum_{k=0}^{N-1} \frac{(x_k - \bar{x})^4}{\sigma_x^4} \quad (1.47)$$

gdzie x_k to k -ta obserwacja spośród dostępnych N obserwacji, \bar{x} to średnia arytmetyczna ze wszystkich N próbek, σ_x – to odchylenie standardowe (dyspersja, rozrzut) liczone na podstawie obserwacji.

Parametrem służącym jako miara płaskości widma (ang. SFM – *Spectral Flatness Measure*) jest (średnia geometryczna przez średnią arytmetyczną)

$$SFM = 10 \log_{10} \frac{\frac{1}{N/2} \sqrt{\prod_{k=0}^{N/2-1} P_k}}{\frac{1}{N/2} \sum_{k=0}^{N/2-1} P_k} \quad (1.48)$$

przy czym P_k to widmowa gęstość mocy (WGM) obliczona za pomocą N -punktowej DFT (FFT).

Momenty widmowe można także obliczać dla fragmentów widma. Wówczas należy w powyższych wzorach zmienić zakresy sumowania (i iloczynu dla SFM) na numery próbek widma odpowiadające częstotliwościom: f_d – dolnej i f_g – górnej, widma sygnału. Przykładowo, pierwszy moment widmowy znormalizowany $M_u(1)$ – środek ciężkości widma, liczony w zakresie pomiędzy dwoma kolejnymi minimami obwiedni widma, może być interpretowany jako częstotliwość **formantu**, znajdującego się w tym paśmie częstotliwości.

Poniżej jest zamieszczony przykład zastosowania momentów widmowych w praktyce.

Identyfikacja mówcy z zastosowaniem momentów widmowych

Zgromadź zbiory nagrań głosów. Aby sklasyfikować szczególny, docelowy głos spośród głosów w zbiorze odniesienia (referencyjnym), należy [1] zgromadzić głosy o nazwie angielskiej *isochunks* (jednakowe kawałki). *Isochunk* jest definiowany jako segment mowy, w którym mówca, wymawiając go, za każdym razem koniecznie stosuje tę samą wymowę, a więc brzmi on tak samo. Aby móc taki segment wykorzystać, w bazie zbiorów nagrań głosów musi egzystować więcej niż jeden taki segment dla każdego mówcy. Im więcej, tym lepiej z punktu widzenia jego weryfikacji/identyfikacji na podstawie porównania z nagraniami zapisanymi w bazie.

Isochunk może mieć dowolną długość, jednak nie powinien zawierać przerw. Chodzi o to, by miał on z zasady taką samą reprezentację, przy której może on różnić się dla poszczególnych mówców sposobami jego artykulacji. (Artykulacja – ruchy i układ narządów mowy (np. języka, warg, podniebienia miękkiego, więzadeł głosowych) podczas wymawiania poszczególnych głosek.) Czynnikiem krytycznym w selekcji segmentu mowy jako *isochunk*-a jest, by wybrany ciąg brzmiał za każdym razem tak podobnie, jak to tylko możliwe dla danego mówcy i niepodobnie, jak to tylko możliwe, dla innych mówców.

Po wyselekcjonowaniu *isochunks* i ich wydobyciu, następnie, celem wydobycia z nich cech dystynktywnych przydatnych do klasyfikacji, można zastosować następujący algorytm [1], sporządzony w Szwecji na zamówienie banku, a oparty na metodzie momentów widmowych opracowanej w North Carolina State University, USA, którego kolejne kroki zestawiono poniżej.

1. Oblicz DFT (ang. *Discrete Fourier Transform*) w oknie o długości N .
2. Odrzuć część urojoną.
3. Przesuń o jedną próbkę i powtórz N -krotnie kroki 1 i 2.
4. Weź średnią z N transformat i przeskaluj przez pierwiastek trzeciego stopnia. To zmniejszy wpływ pierwszego formantu.
5. Wykonaj interpolację wynikowej średniej za pomocą sześciennego splinu (ang. *cubic spline*), patrz MATLAB – *cubic spline interpolation*, aby otrzymać synchroniczne, ciągłe widmo wysokości dźwięku (ang. *pitch*).

6. Scałkuj ciągłe widmo wysokości dźwięku od 0 do 4000 Hz. To zwróci masę widma

$$mass = \int_0^{4000} S(f)df \quad (1.49)$$

7. Podziel widmo przez jego masę. Dostarczy to poniższej funkcji gęstości prawdopodobieństwa i sprawi, że powierzchnia pod krzywą widma będzie miała wartość jeden, a dla wszystkich pod-powierzchni będzie to wartość z zakresu od 0 do 1.

$$P(f) = \frac{S(f)}{mass} \quad (1.50)$$

8. Scałkuj funkcję gęstości prawdopodobieństwa przemnożoną przez częstotliwość. Otrzymasz pierwszy moment tj. wartość oczekiwaną

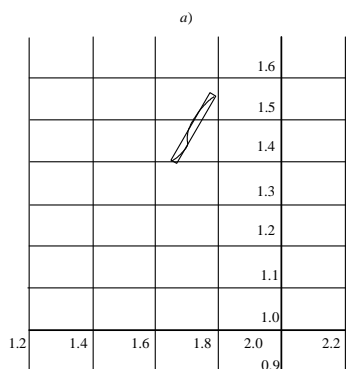
$$m_1 = \bar{x} = \int_0^{4000} f P(f)df \quad (1.51)$$

9. Scałkuj kwadrat różnicy pomiędzy pierwszym momentem i iloczynem częstotliwości przez funkcję gęstości prawdopodobieństwa. Otrzymasz drugi moment, to jest dyspersję wokół wartości oczekiwanej

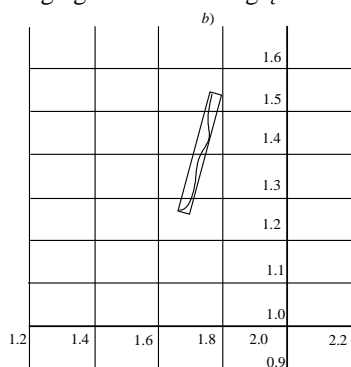
$$m_2 = \sigma^2 = \int_0^{4000} (f - m_1)^2 P(f)df \quad (1.52)$$

10. Powtórz kroki od 1 do 9, gdy liczba próbek pozostawionych w segmencie mowy jest większa od $3N$.
11. Przeskaluj pierwszy i drugi moment, odpowiednio, przez 10^{-3} i 10^{-6} .
12. Sporządź wykres drugiego momentu (oś pionowa) w funkcji pierwszego (oś pozioma). To utworzy ślad (ang. *track*).
13. Nałóż na *track* otaczający go prostokąt o minimalnej wielkości – MER od ang. *minimal enclosing rectangle*. Końcowe wyniki do klasyfikacji są wydobywane z MER.
14. Wydobądź z MER następujące atrybuty: wartości minimalną i maksymalną na osiach poziomej i pionowej (każda z nich reprezentuje jeden narożnik), długości boków po stronie dłuższej i krótszej, współrzędne środkowego punktu prostokąta na osi poziomej i pionowej i kąt orientacji prostokąta względem osi poziomej.

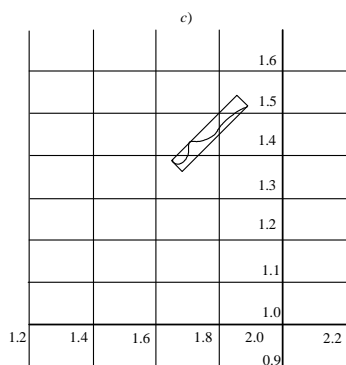
Przykłady i trzy wykresy momentów: drugiego względem pierwszego, zawierające MER są dostępne w [1] i pokazane na rys. 3, 4 i 5. Rozpatrzono tam głos – segment wypowiedzi – znanego szwedzkiego polityka oraz segment tej samej wypowiedzi ale przez profesjonalnego imitatora (naśladowcę), a następnie przez imitatora jego naturalnym głosem. Do klasyfikacji stosowano z sukcesem dyskryminanty z p. 14 (patrz wyżej) oraz miary oparte na odległości Mahalanobisa wprowadzonej w 1936. (patrz Wikipedia). Osiągnięto podobne wyniki dla obu tych metod. Pozwalały one na pozytywną ocenę wyników porównania oryginalnego naturalnego nagrania z imitacjami. Metoda momentów widmowych okazała się niewrażliwa na imitacje oryginalnej wypowiedzi.



Rys. 3. Wykres śladu (ang. *track*) uformowanego przez wykreślenie pierwszego i drugiego momentu względem siebie. Głos słynnego szwedzkiego polityka.



Rys. 4. Wykres śladu (ang. *track*) uformowanego przez wykreślenie pierwszego i drugiego momentu względem siebie. Głos jednego z profesjonalnych imitatorów podczas imitacji słynnego szwedzkiego polityka. Ta sama wypowiedź, jak na rys. 3.



Rys. 5. Wykres śladu (ang. *track*) uformowanego przez wykreślenie pierwszego i drugiego momentu względem siebie. Naturalny głos imitatora z rys. 3 podczas imitacji słynnego szwedzkiego polityka. Ta sama wypowiedź, jak na rys. 3.

W MATLABie oprócz nieparametrycznych działają również inne estymatory WGM, ale należące do parametrycznych. Są to:

- ◆ pburg działający wg metody Burga,
- ◆ pcov działający wg metody kowariancji,
- ◆ peig działający wg metody wektorów własnych,
- ◆ pmcov działający wg zmodyfikowanej metody kowariancji,
- ◆ pmusic wykorzystujący algorytm MUSIC (Multiple Signal Classification),
- ◆ pwelch działający wg metody Welcha,
- ◆ pyulear działający wg metody AR (autoregresji) Yule-Walkera.

Wszystkie te algorytmy, należące do wysokorozdzielczej parametrycznej estymacji WGM, są opisane w Help do MATLABa i opatrzone przykładami i ilustracjami.

Wysokorozdzielcza parametryczna estymacja WGM – widma gęstości mocy

Podstawowe założenie: badany sygnał traktowany jest jako proces stochastyczny. Proces ten jest generowany przez model matematyczny zależny od małej liczby parametrów. Poszukiwane WGM determinują parametry modelu. Parametryczna estymacja WGM składa się z trzech etapów.

1. Wybór odpowiedniego modelu parametrycznego (na podstawie wiedzy *a priori* o procesie).
2. Estymacja parametrów (identyfikacja) modelu.
3. Obliczenie widma na podstawie tak otrzymanych parametrów modelu.

Najczęściej stosowane modele parametryczne (inaczej innowacyjne):

- AR – autoregresywny, autoregresji,
- MA – ruchomej średniej, ang. *moving average*,
- ARMA – autoregresji i ruchomej średniej, mieszany
- suma cisoid (zespolonych sinusoid) z addytywnym szumem białym.

Model AR jest filtrem cyfrowym rekursywnym (IIR od ang. *Infinite Impulse Response*). Jego transmitancja nie ma zer poza punktem $z = 0$. Model kształtuje widmo wyłącznie za pomocą biegunów. Stąd ma angielską nazwę *all-pole model*. Równanie różnicowe (r.r.) – algorytm – modelu AR rzędu p , AR(p) to

$$x[n] = -\sum_{m=1}^p a_m x[n-m] + e[n] \quad (2.1)$$

gdzie $x[n]$ to (dostępna) obserwacja wyjścia modelu, a_1, a_2, \dots, a_p – parametry modelu, $e[n]$ – nieznanne losowe pobudzenie modelu (nie dostępne), innowacja, jest realizacją białego szumu gaussowskiego o zerowej wartości średniej i nieznannej mocy (wariancji) σ^2 .

Model MA jest filtrem cyfrowym nierekursywnym (FIR od ang. *Finite Impulse Response*) rzędu q , MA(q) nie mającym biegunów (ang. *all-zero model*). Równanie różnicowe (r.r.) – algorytm – modelu MA(q) to

$$x[n] = \sum_{m=0}^q b_m e[n-m] \quad (2.2)$$

gdzie $b_0 = 1, b_1, b_2, \dots, b_q$ – parametry modelu, inaczej współczynniki ruchomej średniej.

Model ARMA jest filtrem cyfrowym IIR rzędu $\max(q, p)$. Równanie różnicowe – algorytm – modelu ARMA to kombinacja liniowa (2.1) i (2.2)

$$x[n] = -\sum_{m=1}^p a_m x[n-m] + \sum_{m=0}^q b_m e[n-m] \quad (2.3)$$

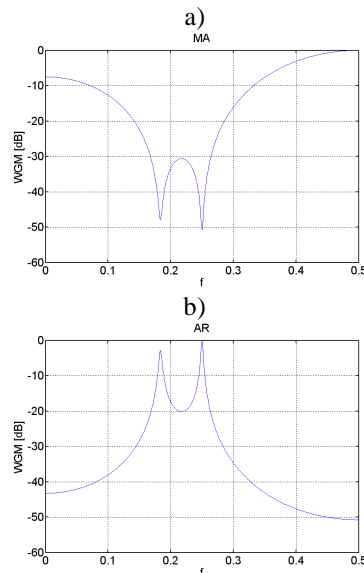
Parametry tego modelu nazywają się tak samo, jak parametry jego składowych AR i MA.

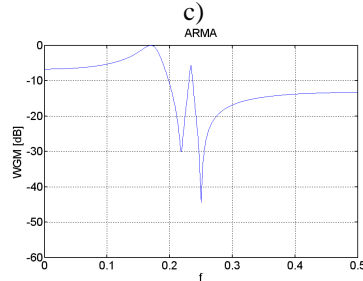
Model „suma zespolonych sinusoid w białym szumie” ma algorytm

$$x[n] = \sum_{i=1}^m A_i \exp(j\omega_i n) + e[n] \quad (2.4)$$

gdzie A_1, A_2, \dots, A_m – amplitudy zespolone (parametry modelu), a $e[n]$ – realizacja białego szumu gaussowskiego o zerowej wartości średniej i nieznannej mocy (wariancji) σ^2 (tak samo, jak wyżej).

Ogólnie parametrami modelu są zarówno zespolone amplitudy jak i pulsacje sinusoid zespolonych. Te ostatnie spełniają związek $-\pi < \omega_1 < \omega_2 < \dots < \omega_m < \pi$.





Rys. 6. Typowe widma modeli parametrycznych: a) MA, b) AR i c) ARMA; dla celów porównawczych maksima znormalizowano do 0 dB.

Na rys. 6 zera transmitancji MA to: $\pm j0.99$ i $-0.4 \pm j0.9$, stąd $H_{MA}(z) = (z^2 + 0.99^2)[(z+0.4)^2 + 0.9^2]$. Bieguny transmitancji AR to $\pm j0.99$ i $-0.4 \pm j0.9$, stąd $H_{AR}(z) = \frac{1}{H_{MA}(z)} = \frac{1}{(z^2 + 0.99^2)[(z+0.4)^2 + 0.9^2]}$. Dla transmitancji ARMA przyjęto na powyższym rysunku położenie zer: $\pm j0.995$ i $-0.2 \pm j0.995$, położenie biegunów: $-0.1 \pm j0.98$ i $-0.5 \pm j0.999$. Przy tym $H_{ARMA}(z) = H_{AR}(z)H_{MA}(z)$. Czytelnikowi pozostawiamy znalezienie tej transmitancji.

Estymacja WGM na podstawie modelu AR

Model AR stosowany jest szeroko w technice fonicznej, np. w modelu artykulacji mowy, w modelowaniu niektórych instrumentów muzycznych i in. Modeluje się również charakterystyki torów transmisyjnych, bezpośrednio na podstawie odbieranego sygnału.

Jeżeli $x[n]$ jest procesem AR rzędu p tzn. $\{x[n]\} \in \text{AR}(p)$, to jego WGM określone jest wzorem

$$P_{\text{AR}(p)}(\omega) = \frac{\sigma^2}{\left| 1 + \sum_{m=1}^p a_m \exp(-j\omega m) \right|^2} \quad (2.5)$$

Aby znaleźć WGM trzeba znać σ^2 i a_m , $m = 1, 2, \dots, p$. Jeżeli obie strony równania (2.1) modelu AR pomnożymy przez $x^*[n-l]$, $l = 0, 1, \dots$ i uśrednimy po zbiorze realizacji operatorem wartości oczekiwanej E , to otrzymamy

$$\underbrace{E\{x[n]x^*[n-l]\}}_{r[l]} = - \underbrace{\sum_{m=1}^p a_m E\{x[n-m]x^*[n-l]\}}_{\text{autokorelacja } r[l-m]} + \underbrace{E\{e[n]x^*[n-l]\}}_{=0} \quad (2.6)$$

Zerowanie się ostatniego składnika oznacza niezależność. Z (2.6) wynikają równania Youle-Walkera

$$r[l] = \begin{cases} -\sum_{m=1}^p a_m r[l-m], & l \neq 0 \\ -\sum_{m=1}^p a_m r[l-m] + \sigma^2, & l = 0 \end{cases} \quad (2.7)$$

Metoda autokorelacyjna

Dysponując estymatą ciągu autokorelacji (por. z (1.10))

$$\hat{r}[l] = \frac{1}{N} \sum_{k=0}^{N-1} x[k] x^*[k-l] \quad (2.8)$$

gdzie N to liczba próbek ciągu obserwacji $x[n]$ i gdzie $l = -(p-1), \dots, 0, \dots, (p-1)$, znajdujemy parametry modelu $AR(p)$ rozwiązując równanie macierzowe

$$\hat{\mathbf{R}} \hat{\mathbf{a}} = -\hat{\mathbf{r}} \quad (2.9)$$

gdzie

$\hat{\mathbf{a}} = [\hat{a}_1 \hat{a}_2 \dots \hat{a}_p]^T$ wektor współczynników autokorelacji (T oznacza tu transpozycję wektora)

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{r}[0] & \hat{r}[-1] & \hat{r}[-2] & \dots & \hat{r}[-p+1] \\ \hat{r}[1] & \hat{r}[0] & \hat{r}[-1] & \dots & \hat{r}[-p+2] \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \hat{r}[p-1] & \hat{r}[p-2] & \hat{r}[p-3] & \dots & \hat{r}[0] \end{bmatrix}$$

$\hat{\mathbf{r}} = [\hat{r}[1] \hat{r}[2] \dots \hat{r}[p]]^T$ – wektor autokorelacji.

Rozwiązanie:

$\hat{\mathbf{a}} = -\hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}$, przy czym

$$\sigma^2 = \hat{r}[0] + \sum_{m=1}^p \hat{a}_m \hat{r}^*[m].$$

Równania Youle-Walkera dla $x[n]$ – stacjonarnego procesu $AR(p)$ o ciągu autokorelacji $r_x[l]$

$$r_x[l] = \begin{cases} -\sum_{m=1}^p a_m r_x[l-m], & l = 1, 2, \dots \\ -\sum_{m=1}^p a_m r_x[l-m] + \sigma^2, & l = 0 \\ r_x^*[-l], & l = -1, -2, \dots \end{cases} \quad (2.10)$$

gdzie σ^2 – to moc ciągu innowacji (szumu białego). Inaczej w postaci macierzowej

$$\begin{bmatrix} r_x[0] & r_x[-1] & \cdots & r_x[-p] \\ r_x[1] & r_x[0] & \cdots & r_x[-p+1] \\ \vdots & \vdots & \cdots & \vdots \\ r_x[p] & r_x[p-1] & \cdots & r_x[0] \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad - \text{układ równań tzw. normalnych}$$

Macierz autokorelacji $(p+1) \times (p+1)$ jest macierzą Toeplitza. Jest to dodatnio określona macierz Hermite'a. Jej szybkie odwracanie daje algorytm Levinsona-Durbina (str. 15).

W praktyce, dysponując obserwacją $x[n]$, najpierw estymujemy ciąg autokorelacji $\{\hat{r}_x[n]\}_{-L}^L$. Następnie formujemy macierz estymat autokorelacji $(p+1) \times (p+1)$, $p \ll L$.

W końcu obliczamy wektor współczynników $\{\hat{a}_m\}_1^p$. Są one nazywane współczynnikami liniowej predykcji (ang. LPC – *linear predictive coding/coefficients*).

Rząd modelu p , to w przybliżeniu liczba zespolonych sinusoid wykrywanych przez $P_{AR(p)}(\omega)$ (2.5).

Model AR(p) stacjonarnego ciągu próbek $x[n]$

Równanie różnicowe procesu autoregresji AR(p) rzędu p

$$x[n] + \sum_{m=1}^p a_m x[n-m] = e[n]$$

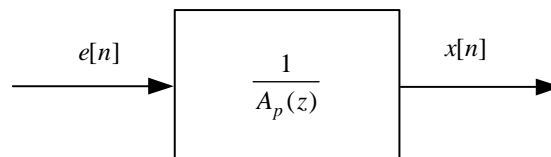
albo

$$\sum_{m=0}^p a_m x[n-m] = e[n]$$

gdzie zawsze $a_1 = 1$,

$e[n]$ – stacjonarny szum biały – ciąg innowacji. Schemat blokowy modelu AR(p) pokazano

poniżej, gdzie $x[n] = e[n] - \sum_{m=1}^p a_m x[n-m]$ i $A_p(z) = \sum_{m=0}^p a_m z^{-m}$; $\{a_m\}_1^p$ – współczynniki modelu AR(p).



Rys. 7. Schemat blokowy modelu AR(p).

Uwaga: **predyktor** działa wg algorytmu

$$\hat{x}[n] = -\sum_{m=1}^p a_m x[n-m] = -a_1 x[n-1] - a_2 x[n-2] - \cdots - a_p x[n-p]$$

z błędem predykcji

$$e[n] = x[n] - \hat{x}[n] = \sum_{m=0}^p a_m x[n-m].$$

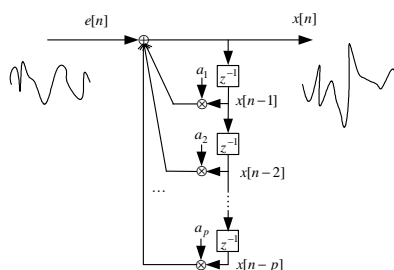
Filtr rekursywny AR(p) – stabilny, przyczynowy (tj. realizowalny w czasie rzeczywistym) i odwracalny

Model AR(p) o strukturze transwersalnej (bezpośredniej) ma transmitancję

$$H(z) = \frac{X(z)}{E(z)} = \frac{1}{1 + \sum_{m=1}^p a_m z^{-m}}.$$

Filtr do niego odwrotny operuje transmitancją

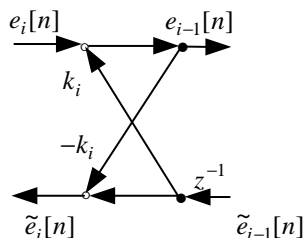
$$H^{-1}(z) = 1 + \sum_{m=1}^p a_m z^{-m}.$$



Rys. 8. Struktura transwersalna modelu AR(p).

Struktura kratowa (ang. *lattice*) Markela i Graya modelu AR(p). Algorytm *i*-tego ogniwa kraty

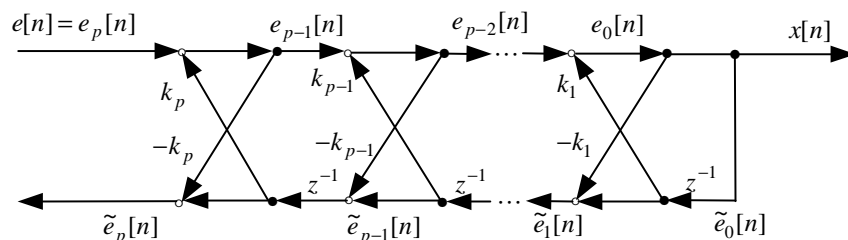
$$\begin{aligned} e_{i-1}[n] &= e_i[n] + k_i \tilde{e}_{i-1}[n-1] \\ \tilde{e}_i[n] &= -k_i e_{i-1}[n] + \tilde{e}_{i-1}[n-1] \end{aligned}$$



Rys. 9. Pojedyncze ogniwo struktury kratowej modelu AR(p).

Na rys. 9 oznaczono sygnały na wyjściach sumatorów. Cały filtr kratowy, pokazany na rys. 10, ma transmitancję

$$H(z) = \frac{X(z)}{E(z)} = \frac{1}{A_p(z)} = \frac{1}{1 + \sum_{m=1}^p a_m z^{-m}}$$



Rys. 10. Struktura kratowa (ang. *lattice*) Markela i Graya modelu $AR(p)$.

Algorytm Levinsona-Durbina obliczania współczynników k_i , $i=1,2,\dots,p$ na podstawie współczynników a_m , $m=1,2,\dots,p$

- (1) $a_m^{(p)} = a_m$
- (2) $k_p = a_p^{(p)}$
- (3) $k_i = a_i^{(i)}$
- (4) $a_m^{(i-1)} = \frac{a_m^{(i)} + k_i a_{i-m}^{(i)}}{1 - k_i^2}$; $m=1,2,\dots,(i-1)$

Rekursję powtarza się dla $i=p, p-1, \dots, 2, 1$, a więc dla malejących indeksów i , aż do obliczenia wszystkich k_i , $i=1,2,\dots,p$.

Przykład 2.

$$\text{Dla } H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}} = \frac{1}{1 + 0.9z^{-1} - 0.64z^{-2} + 0.57z^{-3}}, \quad N=3$$

dostajemy

$$z(1) \quad a_1^{(3)} = a_1 = 0.9, \quad a_2^{(3)} = a_2 = -0.64, \quad a_3^{(3)} = a_3 = 0.576$$

$$z(2) \quad i = p = 3, \quad k_3 = a_3^{(3)} = 0.576$$

$$z(4) \quad m=1, \quad a_1^{(2)} = \frac{a_1^{(3)} + k_3 a_2^{(3)}}{1 - k_3^2} = \frac{0.9 + 0.576 \times (-0.64)}{1 - 0.576^2} \cong 0.7952$$

$$z(4) \quad m=2, \quad a_2^{(2)} = \frac{a_2^{(3)} + k_3 a_1^{(3)}}{1 - k_3^2} = \frac{-0.64 + 0.576 \times 0.9}{1 - 0.576^2} \cong -0.182$$

$$z(3) \quad i = p-1 = 2, \quad k_2 = a_2^{(2)} = -0.182$$

$$z(4) \quad m=1, \quad a_1^{(1)} = \frac{a_1^{(2)} + k_2 a_1^{(2)}}{1 - k_2^2} = \frac{0.7952 - 0.182 \times 0.7952}{1 - (-0.182)^2} \cong 0.673$$

$$z(3) \quad i = p-2 = 1, \quad k_1 = a_1^{(1)} = 0.673$$

Struktura kratowa ma bardzo dobre właściwości numeryczne. Współczynniki tej struktury, k_i , $i=1,2,\dots,p$ nazywa się współczynnikami odbicia (od ang. *reflection coefficients*). Jeszcze inna nazwa to PARCORY (od ang. *partial correlation coefficients*). Są to współczynniki o wartościach $|k_i| < 1$, pozwalające na tworzenie stabilnych filtrów. Nie

wymagają one zmuszenia do przeprojektowania całej struktury kratowej, gdy trzeba zwiększyć liczbę ogniw kraty. Takiej właściwości nie ma struktura transwersalna.

Sygnał mowy można modelować za pomocą struktury kratowej, podobnie jak sygnały instrumentów muzycznych czy traktu głosowy. Można znaleźć WGM tak wygenerowanych sygnałów oraz transmitancję traktu głosowego. W technice fonicznej bardzo często są wykorzystywane struktury kratowe z podanych wyżej powodów. Filtr kratowy wykorzystuje się np. do rekonstrukcji sygnału fonicznego obciążonego szumem i zakłóceniami, metodą interpolacji adaptacyjnej, opierając się na założeniu, że sygnał akustyczny można zamodelować jako proces autoregresji [2].

Pytania z analizy widmowej dla Akustyki Mowy

1. Czym zajmuje się analiza widmowa sygnałów mowy? Co to widmo gęstości mocy (WGM) i jak się to widmo oblicza?
2. Jaki jest związek pomiędzy WGM na wyjściu i na wejściu systemu dyskretnego? Objaśnij użyte symbole.
3. Zdefiniuj periodogram i korelogram. Objaśnij jak się oblicza periodogram Schustera i w jakim celu się go stosuje.
4. Do czego wykorzystuje się momenty widmowe i na jakiej podstawie się je oblicza?
5. Zapisz formuły na najważniejsze momenty widmowe. Objaśnij użyte symbole.
6. Co to jest miara płaskości widma (ang. *Spectral Flatness Measure* – *SFM*)? Jak można ją obliczyć dla fragmentu widma?
7. Dla jakich sygnałów używamy nieparametryczną, a dla jakich parametryczną estymację widma? Wymień etapy parametrycznej estymacji widma.
8. Wymień najczęściej stosowane modele parametryczne i naszkicuj typowe widma dla modeli parametrycznych.
9. Objaśnij w jaki sposób wykonuje się estymację WGM na podstawie modelu autoregresji.
10. Jakie właściwości ma struktura kratowa modelu autoregresji i gdzie znajduje zastosowanie?

Zastosowania momentów widmowych

1. Sterowanie robotów za pomocą głosu. (Słuch robotów bazuje na momentach widmowych.)
2. Rozpoznawanie wad wymowy.
3. Niektóre odmiany aparatów słuchowych (ang. *hearing aids*).
4. Rozpoznawanie emocji w połączeniu z obrazem.
5. Rozpoznawanie emocji w zastosowaniach sądowych i policyjnych.
6. Estymacja wysiłku wokalisty.

Ewa Hermanowicz
20-03-2015