

音聲分析, 모델링

安 秀 吉

서울대학교 電子工學科

I. 서론

電子工學도가 다루게 되는 연구대상은 대부분이 자동적으로 電子回路로 치환이 되는 그러한 것이 아닌 경우가 많다. 즉 電子工學분야에서 어떠한 대상을 분석하고 설계할 때 그대상은 電子回路로 자동적으로 대체되는 것만은 아니고 현실적인 취급대상은 일반적으로는 좀 더 복잡한 것인 데 필요에 의해서 일부분의 양상만을 보기로 하고 간단화하게 되는 것이다. 즉 우리의 경우는 대상의 외관, 크기, 무게, 경도, 맛 등은 다 버리고 전자적인 면에 대해서만 대신할 수 있는 model을 만들어 간단화하고 있다.

인간의 발성기구의 modeling 경우도 이 model에 해당되는 물리적기구(電子回路)가 현실의 복잡한 발성기구를 특정시각의 입장에서만 대신하고 있는 것이다. 이 model은 실제기구의 기능을 대체할 수 있는 중에서 가장 간단한 것을 쓰지만 필요에 따라서는 더 복잡한 것이 사용될 때도 있다.

인간이 목소리를 내고 있을 때 이 소리를 transducer에 의해서 전기파형으로 바꿀 수 있는 데, 이때 出力되는 전기信號와 똑 같은 파형을 계속 낼 수 있는 등가회로가 있다면 그것이 바로 발성기구를 회로적으로 modeling한 것이라고 말할 수 있을 것이다. 인간의 목소리는 비교적 복잡하기 때문에 이와 비슷한 파형을 내기 위해서 model은 최소한 8차의 공진회로급 등가회로가 필요하다는 것이 밝혀져 있다. 이 때 세부효과의 차는 다 무시하고 지금 발생하고 있는 말의 언어학적인 내용을 알아듣기에 충분한 정도의 유사기능을 얻기 위해서 그렇다는 뜻이다.

사실 여러사람이 같은 “아” 소리를 내고 있다고 생

각하면 보는 각도에 따라서는 파형이나 spectrum에서 같은 부분보다는 틀린 부분이 많을 경우가 더 많지만 우리의 귀에는 같은 “아” 가 들리게 된다. 따라서 電子回路를 써서 파형이나 spectrum의 미세부분은 상당히 달라도 사람 귀에 대강 같은 “아” 소리로 느껴지게 되는 가장 간단한 등가회로를 갖고 발성기구를 modeling하게 되는 데 이 model은 사람의 발성을 담당하는 성대(vocal cord), 그리고 인후나 구강, 입술로서 구성되는 聲道(vocal tract) 등 실제대상을 여러가지 각도에서 전폭적으로 대신하는 것이 아니고 단지 언어학적 각도에서 보았을 때의 bare minimum 등가를 이룩하게 되는 것으로 화자(speaker)의 목소리의 潤기나 低음 우렁참, 또는 톤(tone)을 완전하게 나타내는 데는 불충분한 경우가 많다. 사실상 이 대체품은 파형이나 spectrum에서 가장 중요한 대세만을 둥글둥글 따라가서 미세부분은 상당히 다르고 곡선도 비교적으로 간단한 편이다. 그런데도 “아” 로 들리는 것이다.

전화장치는 그러한 bare minimum정보를 전달하는 장치의 하나의 예로서 사람들이 말하는 언어학적 내용을 간신히 알기에 족한 주파수대역만을 취급하고 있기 때문에 소프라노 가수의 노래소리를 전화로 들어 그 소리의 질을 평가하려 해서는 안될 것이다. 전화라는 장치의 목적은 어디까지나 사람 목소리 속에서 언어학적 정보라는 일부만을 겨우 족하게 전달하는 것을 목적으로 삼고 있는 것이다.

인간의 聲道を modeling하는 시도는 과거 기계를 이용하여 音聲을 합성하려고 했던 때부터 시작하였고 말 할 수 있을 것이다. 그후 통계학, 音聲학, 디지털 信號처리 등의 이론을 바탕으로 聲道を 정확히 modeling하고 이를 이용한 音聲의 새로운 분석 기법

을 개발하려는 노력이 최근 20년간 활발히 이루어졌으며 또한 많은 성과가 있었다.

聲道の 정확한 modeling, 및 이를 이용한 音聲의 분석기법은 音聲信號처리, 즉 音聲합성, 音聲인식 등과 같은 적용분야에 가장 기초가 되는 부분이며 그 정확성에 따라 이러한 적용분야들의 성공여부가 결정된다고 해도 과언이 아니다. 聲道を 정확히 modeling하게 되면 音聲합성시 가장 자연스런 音聲信號를 만들 수 있고, 聲道model의 파라미터들은 시간적으로 충분히 천천히 변하므로 音聲 coding(부호화)에 매우 효과적이다. 또한 서로 다른 발음은 서로 다른 聲道の 모양에서 기인하므로 聲道の 정확한 modeling은 특정발음을 정확히 인식하는데 있어서의 관건이 된다.

우선 聲도를 음향학적으로 고찰해 보면 聲道는 아래 그림과 같은 일종의 음향tube(관)로 가정할 수 있다.

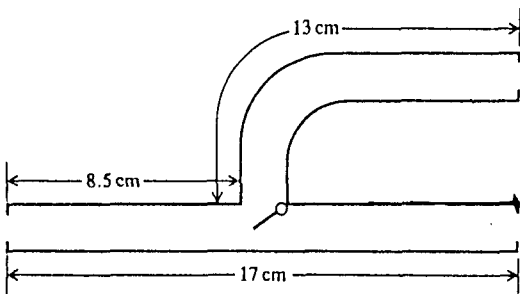


그림 1. 音聲의 음향 tube model

대부분은 무손실인 음향 tube로 가정하며 또한 AR(Auto Regressive)model의 한계를 만족시키기 위해 비강의 효과가 없는 단일통로 음향tube로 modeling한다. 이 음향tube는 전체통로에 따라서 계속 단면이 원형이며 평면파동의 가정을 만족하는 것으로 modeling한다. 본논문에서는 音聲信號를 기초로 한 聲道の 통계적인 modeling에 중점을 둔 것이다.

聲道の 통계적인 model 중 가장 성공적인 model 이 線型豫測 model이다. 音聲의 線型豫測 model은 1960년대 Fant를 시작으로 해서 Flanagan에 의해서 발전되었고 1970년대에 Itakura, Atal 등에 의해 확립되었다. 線型豫測 model은 몇가지 단점이 있음에도 불구하고 성공적인 model로 평가받고 있으며 따라서 현재 音聲信號처리 분야에서 가장 많이 이용

되는 model이다. 그러므로 이 글에서는 주로 線型豫測 model을 중심으로 개발된 音聲의 분석기법들을 소개하고자 한다. 더불어 최근에 연구되고 있는 몇몇 modeling에 대해서도 간단히 언급할 것이다.

Ⅱ. 音聲의 분석합성 model

그림 2는 音聲이 線型豫測 model의 블록도이다. 필터-入力 $e(t)$ 는 유성음의 경우는 주기 P의 impulse 列로 그리고 무성음의 경우는 백색잡음으로 modeling 된다. 성대(glottis)에서 나오는 Glottal 파형 $U_g(t)$ 는 100Hz 근처에 cut off frequency(저지주파수)를 갖는 two pole 저역필터로 modeling된다.

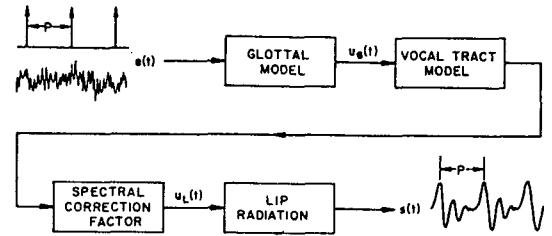


그림 2. 音聲발생기구의 線型model

그리고 다음을 갖는 聲道부분은 몇개의 2차공명기(resonator) 들이 직렬로 연결되어 있는 것으로 modeling된다. 이들 각각의 공명은 해당중심주파수 및 대역폭을 갖는 formant(공진봉우리)로 정의된다. 聲道model은 따라서 모든 주요formants(1st formant, 2nd formant, 3rd formant 등)를 고려하고 있으므로 이들 극점들을 보상한 후(고주파성분이 더 잘 방사되는 특성을 갖고 있는 바) 입술 방사 model(lip radiation model)을 거치게 된다. 위의 model을 z변환의 형태로 쓰면 다음과 같다.

$$S(z) = E(z) G(z) V(z) L(z) \quad (1)$$

여기서 $G(z)$ 와 $L(z)$ 는 각각 다음과 같은 형태를 갖는다.

$$G(z) = \frac{1}{(1 - e^{-\sigma T} z^{-1})^2} \quad L(z) = 1 - z^{-1} \quad (2)$$

여기서 T는 信號의 sampling 간격이다. 식(2)에서 보면 lip radiation은 일종의 고역필터-로 modeling 되었음을 알 수 있다. 전극model(all pole model) V(z)가 k개의 formants로 구성되었을 경우 아래와 같은 형태로 쓸 수 있다.

$$V(z) = \frac{1}{\prod_{i=1}^k [(1 - 2e^{-c_i T} \cos(b_i T)z^{-1} + e^{-2c_i T} z^{-2})]} \quad (3)$$

식(3)에서 i번째 formant의 중심주파수와 대역폭은 각각 $F_i = b_i/2\pi$, $B_i = C_i/\pi$ 로 주어짐을 알 수 있다.

$$S(z) = E(z) \frac{1}{A(z)} \quad [\text{합성 model}] \quad (4)$$

$$A(z) = \sum_{i=0}^M a_i z^{-i} \quad (a_0 = 1) \\ = \frac{1}{G(z) V(z) L(z)} \quad (5)$$

따라서 $M \geq 2k+1$ 이어야 한다.

여기서 A(z)는 전영(all zero) 필터-이고 흔히 역필터-(inverse filter)라고 불리운다. 왜냐하면 순방향으로 이 filter에 excitation E(z)를 인가하면出力단자에 音聲信號 S(z)가 出力되고 역으로 이 필터-에 音聲 S(z)이 入力되었을 경우 아래와 같이 出力-는 음원의 정보 E(z)를 얻을 수 있기 때문이다.

$$E(z) = S(z) A(z) \quad [\text{분석 model}] \quad (6)$$

III. 線型豫測 모델

Norbert Wiener에 의해 확립된 線型豫測 이론이 Itakura와 Atal등에 의하여 처음으로 音聲에 적용되었다. 제 2장의 6식을 시간영역에서 표시하면 다음과 같다.

$$e(z) = \sum_{i=0}^M a_i s(n-i) = s(n) + \sum_{i=1}^M a_i s(n-i) \quad (7)$$

$$\hat{s}(n) = \sum_{i=1}^M a_i s(n-i) \quad (8)$$

로 정의하면 $\hat{s}(n)$ 은 s(n)의 M차線型豫測치가 되고 $l(n) = s(n) - \hat{s}(n)$ 은 예측誤差로 해석할 수 있다. 線型豫測필터-특성을 F(z)라 하면 식(8)로부터

$$\hat{s}(n) = F(z)s(z), \quad F(z) = -\sum_{i=1}^M a_i z^{-i} \quad (9)$$

이 되고 따라서

$$E(z) = S(z)[1 - F(z)] = S(z)A(z) \quad (10)$$

가 된다. 이 관계는 그림 3에 나와 있다.

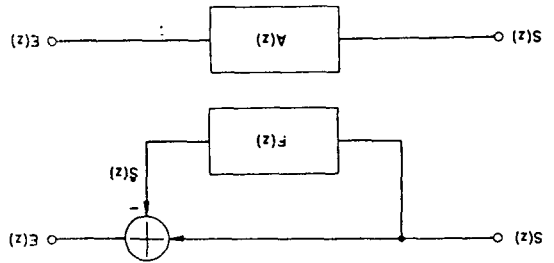


그림 3. 線型豫測 model. 예측기(상도)와 역필터

$e(n)$ 이 s(n)와 $\hat{s}(n)$ 사이의 誤差信號임으로 $e(n)$ 을 最少化하는 방향으로 係數 a_i 들을 결정하는 것이 타당하다. 여러가지 최적화중에서 誤差의 제곱의 합을 最少化시키는 方法이 가장 많이 사용된다. 즉 總誤差제곱의 합 a는

$$\alpha = \sum_{n=n_0}^{n_1} e^2(n) = \sum_{n=n_0}^{n_1} \sum_{i=0}^M \sum_{j=0}^M a_i s(n-i) a_j s(n-j) \quad (11)$$

여기서 n_0 와 n_1 은 분석이 이루어 지는 구간을 나타낸다.

$$c_{ij} = \sum_{n=n_0}^{n_1} s(n-i)s(n-j) \quad (12)$$

를 새로 정의하면

$$\alpha = \sum_{i=0}^M \sum_{j=0}^M a_i c_{ij} a_j \quad (13)$$

이 된다.

α 를 最少化하는 a_i 들을 찾기 위해서는 α 를 a_i 로 미분하여 영으로 놓으면 된다.

즉

$$\frac{\partial \alpha}{\partial \alpha_k} = 2 \sum_{i=0}^M a_i c_{ik} = 0$$

따라서

$$\sum_{i=0}^M a_i c_{ik} = -c_{0k} \quad k = 1, 2, \dots, M$$

이 M개의 미지수를 포함한 M개의 방정식을 풀면 a_i 를 구할 수 있다.

그런데 이 경우 c_{ij} 를 어떻게 정의하느냐에 따라 autocorrelation 방법과 covariance 방법으로 나뉘어진다.

Autocorrelation 방법의 경우, N개의 음성 sample $\{s(n)\} = \{s(0), s(1), \dots, s(N-1)\}$ 이 주어졌다고 가정할 때 $n_0 = -\infty, n_1 = \infty$ 로 설정하고, $n < 0, n \geq N$ 에 대해서는 $s(n) = 0$ 로 가정하고 c_{ij} 를 구한다.

즉

$$c_{ij} = \sum_{n=-\infty}^{\infty} s(n-i) s(n-j) = r(|i-j|) \quad (14)$$

covariance 방법의 경우, $n_0 = M, n_1 = N-1$ 을 설정한다. 즉

$$c_{ij} = \sum_{n=M}^{N-1} s(n-i) s(n-j) \quad (15)$$

다음 장에서는 음성에 적용되어질 수 있는 몇가지의 線型豫測기법에 대하여 간략히 소개할 것이다.

IV. 여러가지 線型豫測기법

1 Maximum likelihood (ML)

짧은 시간동안 음성의 spectrum이 알려지지 않는 몇몇 파라미터에 의해 결정되어진다고 가정하자. ML 방법은 이들 파라미터들을 그 model에 의해 합성된 data가 실제 음성信號와 같아질 수 있는 가장 높은 확률을 주는 파라미터들로 결정하는 방법이다. 어떤 信號 $\{x(n)\}$ 이 gaussian이면 변수 $x(0), x(1), \dots, x(N-1)$ 까지의 joint 를 정의할 수 있다. 이것은 우리가 구하고자 하는 파라미터들과 변수 $x(0), x(1), \dots, x(N-1)$ 의 함수가 된다. 이러한 변수들의 특정한

값이 관측되었다면 이러한 관측의 가능성을 가장 높게 하는 파라미터 값들의 확률밀도함수를 최대화하는 것이다. 그러나 이 ML기법은 N이 조금만 커져도 문제가 매우 비線型적이 되고 따라서 풀기가 어려워진다. 따라서 많은 가정을 하게 되는 데 그중의 한 예를 들자면 확률밀도함수를 다음과 같이 가정한다.

$$P[x(0), x(1), \dots, x(N-1)] = (2\pi\sigma)^{-N/2} \exp(-\alpha / 2\sigma^2)$$

여기에서
$$\alpha = \sum_{n=-\infty}^{\infty} \left[\sum_{i=0}^M a_i x(n-i) \right]^2 \quad (16)$$

단, $x(n) = 0$ for $n < 0$ and for $n > N-1$

위와같은 가정은 결국 前章에서 설명한 autocorrelation 방법으로 이끈다.

2. 최소분산 기법

여기서는 음성信號의 통계적특성에 대해 특별한 가정을 하지 않는다. 식(7)의 誤差信號의 분산을 구하면 아래와 같이진다.

$$E[e^2(n)] = \sum_{i=0}^M \sum_{j=0}^M a_i a_j E[x(n-i)x(n-j)] \quad (17)$$

여기에서 $E[\cdot]$ 는 기대치연산자이다. 정상적(stationary)信號에 대하여

$$E[x(n-i)x(n-j)] = \rho(i-j) = \rho(j-i) \quad (18)$$

즉 autocorrelation 함수가 된다.

$$\text{결국 } E[e^2(n)] = \sum_{i=0}^M \sum_{j=0}^M a_i \rho(i-j) a_j \quad (19)$$

이 된다.

최소분산기법은 이 誤差信號의 분산을 最少化하는 것이다. 결국 이 문제는 $\rho(i-j)$ 를 어떻게 推定하느냐에 따라 역시 autocorrelation 방법과 covariance 방법으로 나뉘어질 수 있다. 한 예로 아래와 같이 $\rho(i-j)$ 를 예측하면 covariance 방법이 된다.

$$\rho(i-j) = \frac{1}{N-M} \sum_{n=M}^{N-1} x(n-i)x(n-j) = \frac{c_{ij}}{(N-M)}$$

3. Prony의 기법

만약 $\{x(n)\}$ 이 단일 pitch 구간동안의 음성信號라

고 가정하자. 이는 all pole filter의 단위 sample 信號를 入力했을 때의 出力이라고 생각할 수 있다. 즉

$$X(z) = E(z)/A(z) = 1/A(z) \quad (20)$$

위의 식을 부분분수 분할하면

$$1/A(z) = \sum_{i=1}^M \frac{u_i}{1-z_i z^{-1}} \quad (21)$$

이 됨으로 이를 역변환하면 $x(n) = \sum_{i=1}^M u_i (z_i)^n$ 이 된다. 따라서 2M개의 변수 u_i 와 z_i 를 구하기 위해서 2M개의 音聲 data가 필요하게 된다. 또한 이전 pitch구간의 과도현상의 影響과 비강으로 부터 기인하는 zero의 影響을 고려하면 식(20)은

$$X(z) = \frac{P(z)}{A(z)} = \sum_{i=0}^{M-1} P_i z^{-i} / \sum_{i=0}^M a_i z^{-i} \quad (22)$$

로 쓸 수 있다. 식(22)를 시간영역에서 표현하면

$$\sum_{i=0}^M a_i x(n-i) = \sum_{i=0}^{M-1} P_i \delta_{n,i} \quad (23)$$

이 된다. 여기에서 $\delta_{n,i}$ 는 Kronecker delta 이다. 식(23)은 결국 아래식을 의미한다.

$$\sum_{i=0}^M a_i x(n-i) = 0 \quad n = M, M+1, \dots, N-1 \quad (24)$$

식(24)에서 model이 실제 音聲信號를 정확히 예측하지 못했을 경우를 가정하면 식(24)는 다음과 같이 쓸 수 있다.

$$\sum_{i=0}^M a_i x(n-i) = e(n) \quad n = M, M+1, \dots, N-1 \quad (25)$$

따라서 이는 결국 전장의 covariance方法和 일치하게 된다.

4. Correlation 일치 기법

Correlation 일치기법은 音聲信號 $\{x(n)\}$ 와 all pole filter $\sigma/A(z)$ 의 unit sample response, 각각의 autocorrelation 함수들이 가능한 한 많은 점에서 일치하게 하는 기법이다. 구하고자 하는 model을 $H(z) = \sigma/A(z)$ 라고 하면 入力이 unit sample 信號임으로

$$\sum_{i=0}^M a_i h_{n-i} = \sigma \delta_{n,0} \quad (26)$$

가 성립한다. 여기서 $\{h_n\}$ 은 $H(z)$ 의 역치환이 causal하다고 가정한다. 식(26)으로 부터

$$\sum_{i=0}^M a_i \rho(i-j) = \sigma h_{-j}, \quad \rho(i-j) = \sum_{n=0}^{\infty} h_n h_{n+|i-j|} \quad (27)$$

를 유도할 수 있다. 그런데 $\{h_n\}$ 이 causal하다고 가정하였으므로

$$\sum_{i=0}^M a_i \rho(i-j) = \begin{cases} 0 & j > 0 \\ \sigma^2 & j = 0 \end{cases} \quad (28)$$

이 된다.

될 수 있는 한 많은 점에서 합성filter의 出力과 音聲信號 $\{x(n)\}$ 각각의 autocorrelation 함수 $\rho(j)$ 와 $r(j)$ 가 일치하도록해야 함으로

$$\rho(j) = r(j), \quad j = 0, 1, \dots, M \quad (29)$$

여기에서 $r(i-j) = \sum_{n=0}^{N-1-|i-j|} x(n)x(n+|i-j|)$ 로 놓으면 이는 결국 autocorrelation方法和 일치하게 된다.

5. PARCOR(Partial correlation)

여기서는 前方線型豫測기와 後方線型豫測기가 사용된다. 前方線型豫測誤差와 後方線型誤差를 각각 $X_n^+(n)$ 과 $X_n^-(n)$ 이라고 하면 각각 다음과 같이 정의된다.

$$\begin{aligned} x_m^+(n) &= x(n) - \left[-\sum_{i=1}^m a_{m_i} x(n-i) \right] \\ &= \sum_{i=1}^m a_{m_i} x(n-i) \quad \text{단 } a_{m_0} = 1 \end{aligned} \quad (30)$$

$$\begin{aligned} x_m^-(n) &= x(n-m-1) - \left[-\sum_{i=1}^m b_{m_i} x(n-i) \right] \\ &= \sum_{i=1}^{m+1} b_{m_i} x(n-i) \quad \text{단 } b_{m,m+1} = 1 \end{aligned} \quad (31)$$

前方豫測誤差와 後方豫測誤差의 總誤差제곱은 각각 다음과 같이 주어진다.

$$\alpha_m = \sum_{n=n_0}^{n_1} [x_m^+(n)]^2, \quad \beta_m = \sum_{n=n_0}^{n_1} [x_m^-(n)]^2 \quad (32)$$

주어진 파라미터들은 이들 誤差제곱을 最少化시키도록 결정되어진다. 그런데 이와같이 前方豫測誤差와

後方豫測誤차를 이용하게 되면 새로운 파라미터를 정의하여 filter를 구성할 수 있는 데 그것이 바로 PARCOR 係數 k_i 이다. 아래와 같이 $A_m(z)$ 와 $B_m(z)$ 를 정의하면

$$A_m(z) = \sum_{i=0}^m a_{mi} z^{-i} \quad \text{단 } a_{m0} = 1$$

$$B_m(z) = \sum_{i=1}^{m+1} b_{mi} z^{-i} \quad \text{단 } b_{m,m+1} = 1 \quad (33)$$

파라미터 k_i 와 더불어 합성filter는 다음과 같은 구조를 갖는다.

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z)$$

$$B_m(z) = z^{-1} [k_m A_{m-1}(z) + B_{m-1}(z)]$$

$$k_m = -\frac{1}{\beta_{m-1}} \sum_{n=n_0}^n x_{m-1}^+(n) x_{m-1}^-(n) \quad (34)$$

위의 식을 시간영역에서 표시하면

$$x_m^+(n) = x_{m-1}^+(n) + k_m x_{m-1}^-(n)$$

$$x_m^-(n) = k_m x_{m-1}^+(n-1) + x_{m-1}^-(n-1) \quad (35)$$

이 되고 이것에 대한 그림이 그림 4이다.

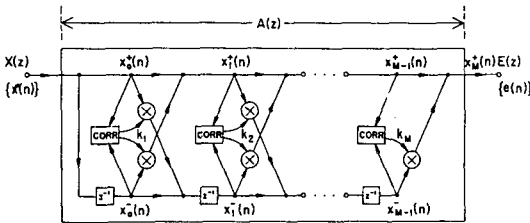


그림 4. parcor을 이용한 역filter의 구조

만약 autocorrelation 方法을 이용하게 되면

$$\alpha_m = \beta_m = \sum_{n=-\infty}^{\infty} [x_m^+(n)]^2 = \sum_{n=-\infty}^{\infty} [x_m^-(n)]^2$$

이 되고 따라서

$$k_m = -\frac{\sum_{n=-\infty}^{\infty} x_{m-1}^+(n) x_{m-1}^-(n)}{\left(\sum_{n=-\infty}^{\infty} [x_{m-1}^+(n)]^2 \sum_{n=-\infty}^{\infty} [x_{m-1}^-(n)]^2 \right)^{1/2}} \quad (36)$$

이 된다.

V. 音聲信號의 時變模型

지금까지 설명한 聲道model은 짧은 구간동안 시스템이 時不變이라고 가정하였다. 즉 짧은 구간에서 音聲信號가 정상적이라고 가정한 것이다. 그러나 音聲信號의 많은 부분이, 예를 들면 默音과 마찰음의 시작점과 유성음/무성음, 무성음/유성음의 변이구간은 매우 빨리 변하고, 정상상태의 모음구간에서도 聲道 시스템은 정상적이긴 하지만 時不變은 아니다.

따라서 既存의 時不變model에 의해서는 만족한 결과를 얻을 수가 없었다. 따라서 상기한 바와 같은 성질의 音聲信號를 model하는 方法으로 파라미터에 의한 時變model이 제시되었다. 대표적인 方法으로는 時變線型豫測, Kalman filter를 이용한 方法, 격자구조를 이용한 순환ARMA方法, 全極 最少自乘格子方法, ARMA LS 方法등이 있으나 본논문에서 소개하기에는 너무 방대하여 線型豫測과 관련하여 時變線型豫測에 대하여만 살펴보고자 한다.

방금 설명한 바와 같이 급격히 끝나는 자음 또는 연음구간등은 조음기관의 구조가 처리하는 音聲data 구간내에서 급격히 변함으로 線型豫測기법으로는 system parameters를 정확히 구할 수가 없다. 따라서 時刻적으로 파라미터가 변하는 model을 고려하게 되었다. 전극model의 경우 식(7)을 다시 쓰며는

$$s(n) = -\sum a_i s(n-i) + Gu(n) \quad (37)$$

여기서 $u(n)$ 은 入力이고 G 는 이득이다. 以前章에서는 a_i 가 10~30msec동안 상수값을 갖는 다고 가정을 하였다. 時變線型豫測方法의 경우에는 예측係數들이 시간과 함께 변할 수 있도록 허용하여야 한다. 즉 (37)은 아래와 같이 쓸 수 있다.

$$s(n) = -\sum_{i=1}^p a_i(n) s(n-i) + Gu(n) \quad (38)$$

일반적으로 時變예측係數는 어떤 시간함수 집합 $\{U_k(n), K=0, \dots, q\}$ 의 線型결합에 의하여 아래와 같이 modeling된다.

$$a_i(n) = \sum_{k=0}^q a_{ik} u_k(n) \quad (39)$$

흔히 사용되는 시간함수의 종류는 다음과 같다.

(1) 시간의 전력함수 집합 $u_k(n) = n^k$ (40)

(2) 시간의 삼각함수 집합 $u_k(n) = \begin{cases} \cos(kwn) & k: \text{짝수} \\ \sin(kwn) & k: \text{홀수} \end{cases}$

$\omega = \pi / N$ (41)

$$\phi_{kl} = \begin{bmatrix} c_{kl}(1,1) & c_{kl}(1,2) & \cdots & c_{kl}(1,m) \\ c_{kl}(2,1) & c_{kl}(2,2) & \cdots & c_{kl}(2,m) \\ \vdots & \vdots & \ddots & \vdots \\ c_{kl}(m,1) & c_{kl}(m,2) & \cdots & c_{kl}(m,m) \end{bmatrix} \quad (48)$$

$0 \leq k \leq q, \quad 0 \leq l \leq q$

식(38)과 식(39)로부터 예측방정식은 다음과 같이 주어진다.

$$\hat{s}(n) = - \sum_{i=1}^M \left[\sum_{k=0}^q a_{ik} u_k(n) \right] s(n-i) \quad (42)$$

既存의 線型豫測方法과 같이 係數의 推定을 위해 總誤差 自乘을 구하면

$$\alpha = \sum_n e^2(n) = \sum_n \left\{ s(n) + \sum_{i=1}^M \sum_{k=0}^q a_{ik} u_k(n) s(n-i) \right\}^2 \quad (43)$$

와 같이 되고 이를 最少化하는 a_i 의 방정식은 아래와 같이 쓸 수 있다.

$$\sum_{i=1}^M \sum_{k=0}^q a_{ik} c_{kl}(i, j) = -c_{0l}(0, j) \quad (44)$$

$1 < j < M, \quad 0 < l < q$

여기서

$$c_{kl}(i, j) = \sum_n u_k(n) u_l(j) s(n-i) s(n-j) \quad (45)$$

이다. $q=0$ 인 경우 $u_0(n)=1$ 임으로 時變線型豫測 $C_{00}(i, j)$ 는 時不變線型豫測 상관함수와 같다. 위의 식으로 부터 시간 n 에 따라 時不變線型豫測方法과 같이 係數를 推定하는 方法은 covariance方法과 autocorrelation方法으로 나뉜다.

1. Covariance 方法

covariance 方法의 경우 계산구간이 $n=M$ 부터 $N-1$ 까지임으로 식(44)을 아래와 같이 정의된 vector와 행렬을 이용하여 쓸 수가 있다.

$$a_i = [a_{1i}, a_{2i}, \dots, a_{mi}]^T \quad 0 \leq i \leq m \quad (46)$$

$$\Psi = [c_{0i}(0,1), c_{0i}(0,2), \dots, c_{0i}(0,p)]^T \quad (47)$$

$0 \leq i \leq q$

식(45)에서 보면, $\phi_{kl} = \phi_{lk} = \phi_{kl}$ 이므로 식(49)는 다음과 같이 쓸 수 있다.

$$\phi a = -\psi \quad (49)$$

여기서 ϕ 는 $(M \times M)$ 대칭을 가진 $(q+1) \times (q+1)$ 블록 대칭행렬이 된다.

따라서 구하고자 하는 파라미-터는

$$a = -\phi^{-1} \psi \quad (50)$$

이 된다.

2. Autocorrelation 方法

時不變方法에서와 같이 처리하려는 data를 무한시간구간에서 data를 windowing한다고 가정하면 식(45)는 autocorrelation함수로 정의될 수 있다.

$$c_{kl}(i, j) = \sum_{n=-\infty}^{\infty} u_k(n-i) u_l(n-j) s(n-i) s(n-j) \\ \equiv r_{kl}(i-j)$$

이러한 정의와 함께 식(49)의 행렬은 대칭이고 블록Toeplitz가 된다.

時變 covariance方法을 위한 誤差自乘 最少化의 한계는 단지 과거 M 개의 sample로부터 예측될 수 있는 音聲 sample에서만 더 해지는 반면 時變 autocorrelation 方法은 전체시간구간에서 誤差를 最少化 한다. 따라서 時不變의 경우와 같이 구간의 끝부분에서 나타나는 불연속에 의한 線型豫測係數의 distorsion이 時變係數에 나타난다. Autocorrelation 方法에 의하여 推定된 係數내의 이러한 distorsion은 구간의 끝 data에 의하여 심하게 影響을 받을 수도 있고 받지 않을 수도 있다. 이러한 distorsion을 줄이기 위하여 autocorrelation 方法에서는 window를 사용한다. 그러나 windowing 효과는 autocorrelation 方法에서 끝부분의 影響을 줄 수 있을 지는 모르지만 音聲信號의 또 다른 시간 變

移를 만든다. 따라서 이러한 windowing은 두가지 문제의 원인이 된다. 첫째 이 방법은信號의 windowing에 의하여 원인이 되는 system parameter의 시간변이에 민감함으로 時變線型 豫測에 의한 係數의 推定에 나쁘게 작용한다. 둘째는 windowing은 분석구간 전체의 誤差의 상대적인 weight에 影響은 준다.

즉, windowing은 data의 끝 부분은 보다 작지만 중앙data의 信號에너지는 무척 크다. 이러한 이유로 window의 사용은 두가지方法에 모두 이롭지 않다. 그리고 autocorrelation 方法의 경우 window를 사용하지 않는 경우에도 끝부분의 影響이 원인이 되어 推定하는데 distorsion을 가져온다.

따라서 covariance方法이 autocorrelation 方法보다 더 有利함을 알 수 있다. 그러나 식(38)의 model에 의한 信號를 완전하게 나타내기 위해서는 이득과 入力信號도 推定해야만 한다.

Ⅵ. 결론 및 향후연구방향

지금까지 설명한 線型豫測을 중심으로 한 聲道 model이 여러가지 장점을 갖고 있기는 하나 그를 이용한 여러가지 응용에서 아직도 미흡한 성능을 보여주는 예가 없지 않다. 이들은 모두 분석시 실제와 상이한 제한적가정에서 부터 기인함으로 보다 물리적으로 타당한 model을 사용하고 또한 현재까지 개발된 기술의 한계적인 聲道推定을 위한 각종parameter 誤差를 줄이기 위해 물리적인 모형을 구현하여 물리 음향학적 분석方法과 信號처리기술의 접목을 통한 연구가 시도되어야 할 것으로 판단된다. 이렇게 함으로서 유성음의 주기성으로 인한 推定誤差 lip radiation 특성에 의한 誤差, 聲道の 실제값을 알지 못하는 데서 오는 信號처리결과와 평가제약등을 모두 극복 할 수 있음으로서 聲道推定目的뿐 아니라 音聲信號분석을 위한 보다 진보된 기술을 도출해 낼 수 있을 것으로 생각되며 특히 音聲信號의 時變특성 분석을 위한 modeling에도 유용한 결과를 가져올 것으로 판단된다.

參 考 文 獻

[1] J.L. Flanagan. speech analysis,

synthesis and perception. 2nd Ed, New York. Springer-verlag. 1972.

[2] L.R Rabiner and R.W.Schafer, Digital processing of speech signals, Englewood Cliffs, N.J,Prentice-Hall, 1978.

[3] F.Itakura and S. Saito, A statistical method for estimation of speech spectral density and formant frequencies, Trans. IECE Japan, vol. 53 A, pp.34-42. 1970.

[4] B.S.Atal and S.L.Hanauer Specch analysis and synthesis by linear prediction of the speech wave. *J.Acoust. Soc. Amer.* vol50. pp.637-655. Aug. 1971.

[5] J.D.Markel and A. H. Gray, Jr. Linear prediction of speech. New York, Springer Verlag. 1976.

[6] J. Makhoul Linear prediction: a tutorial review *Proc. IEEE.* vol. 63. pp. 561-580. Apt. 1975.

[7] G.E. Kopec, A.V. Oppenheim. and J. M. Tribolet, Spcech analysis by homomorphic prediction, *IEEE Trans. Acoust. speech. and Signal Processing.* vol.ASSP 25. pp. 40-49. Feb. 1977.

[8] B.S. Atal and M.R. Schroedor. Linear prediction analysis of spcech based on a pole-zero representation, *J. Acoust. Soc. Amer.* vol.64. pp.1310-1318. Nov. 1978.

[9] K.H. Song and C.K. Ün. Pole-zero modeling of speech based on high-order pole model fitting and decomposition method. *IEEE Trans, Acoust, Specch, and Signal Proccsing,* vol. ASSP-31, pp.1556-1565. Dec. 1983.

[10] K. Steiglitz, On the simultaneous estimation of poles and zeros in speech analysis, *IEEE Trans, Acoust, specch, and Signal Processing,* vol.ASSP-25. pp.229-234, June. 1977.

[11] I.S. Konvalinka and Matausek, Simultaneous estimation of poles and zeros in speech analysis and ITIF-iterative inverse filtering algorithm.

- IEEE Trans. Acoust. Speech. and Signal Processing.* vol ASSP 27, pp.485-492, Oct. 1979.
- [12] H. Morikawa and H. Fujisaka, Adaptive analysis of speech based on a pole-zero representation. *IEEE Trans. Acoust. Speech, and Signal Processing.* vol ASSP-30. pp.77 88, Feb. 1982.
- [13] Y. Miyanaga, N.Miki, N. Nagai, and K. Hatori, A speech analysis algorithm which eliminates the influence of pitch using the model reference adaptive system. *IEEE Trans. Acoust. Speech, and Signal Processing.* vol ASSP-30, pp. 88 96. Feb. 1982.
- [14] L.A. Liporace. "Linear estimation of nonstationary signals. *J. Acoust. Soc. Amer.* vol. 58. pp. 188-1295, 1975.
- [15] F. Casacuberta and E. Vidal, "A nonstationary model for the analysis of transient speech signals," *IEEE Trans. Acoust. Speech, and Signal Processing.* vol. ASSP-35. pp. 226-228. 1987.
- [16] K. Kodera, R. Gendrin, and C. De Villedary. Analysis of time-varying signals with small B-T values." *IEEE Trans. Acoust. Speech, and Signal Processing.* vol. ASSP-26. pp. 64-76. 1978.
- [17] M.G. Hall, A.V.Oppenheim and A.S. Willsky. "Time varying parametric modeling of speech." *Signal Processing.* vol 5. pp. 267-285. 1983.
- [18] Y.T.Lee and H.F.Silverman, "On a general time varying model for speech. : in Proc. ICASSP, May 1988. pp.91-95.
- [19] K.S.Nathan and H.F.Silverman, "High resolution characterization of formants in vowel-consonant transition," in Proc. ICASSP, Apr.1990, pp. 353 356.
- [20] B.S.Atal and S.L.Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.* vol.50, pp. 637-665,1971. ㉸

筆 者 紹 介



安 秀 吉

1930年 4月 17日生

1957年 2月 서울大 통신공학과 졸업

1959年 2月 서울大 전자공학과 대학원 졸업(공학석사)

1976年 2月 서울大 전자공학과 대학원 졸업(공학박사)

1959年 1月~ 1960年 프랑스 CEN Saclay 연구소 연구원

1960年 ~ 1963年 서울대학교 전자공학과 강사

1964年 ~ 1968年 France CNES를 위해서 인공위성지상추적국 CNES 2002호 설계제작

1969年 ~ 현재 서울대학교 전자공학과 교수

1992年 ~ 현재 IEEE 사무총장

주관심분야: 음성신호처리, 검출 및 추정, 통신이론, 회로망 및 시스템 해석