

技術解説

성문파의 추정법

Estimation of Glottal waveform

이 정 철*·안 수 길*
(Jungchul Lee*, Souguil Ann*)
(*서울대학교 전자공학과)

I. 서 론

일반적으로 합성기를 이용하여 구한 합성음은 어둡고 생동감이 없어진다. 이는 인간의 발성구조에 대한 부정확한 모델링 혹은 음원과 성도의 특성에 대한 부정확한 추정과 구현으로 인해서 성도 전달함수와 음원이 실제와 달라지고 변화에 대한 대응이 미흡하기 때문이다. 음성파형은 음원과 성도의 전달함피턴스가 결합된 함수이므로 합성음의 자연성과 개인성 확보를 위해서는 성도의 전달함수 및 음원을 정확히 추정하고 특성을 이해하는 것이 필수적이다. 일반적으로 음원과 성도는 거의 상호작용이 없다고 가정한다. 이 경우 음원과 시스템이 독립적으로 분석가능하며 각각의 특성을 개별적으로 시뮬레이션할 수 있게 된다. 그러나 임펄스열과 저역통과 필터를 이용하여 유성음원을 만들게 되면 유성음원의 스펙트럼 포락선이 거의 균일하게 되어 실제 유성음원의 파형에서 볼 수 있는 zero들을 구현하기 어렵다. 따라서 위의 문제를 해결하기 위해 Rothenberg, Fant, Fujisaki 모델들이 제안되어 사용되고 있지만 아직 합성음의 자연성은 제한되고 있는 실정이다. 현재도 성대, 성도의 특성 및 상호 작용을 잘 정의 할 수 있는 모델개발에 관한 연구가 진행중이다.

본 고는 유성음 발생과정에서 발생기관에서의 성

문의 생리적 구조와 동작원리, 성문파의 추정방법, 발생환경에 따른 성문파의 특성변화에 관해 이미 발표된 여러 연구결과들을 살펴보고 기존에 제안된 성문파 모델들의 특성을 비교분석함으로써 합성음의 문제점을 보완하는 원천기술을 살펴보고자 한다.

II. 성문파의 추정방법

A. 음향학적인 계산

가. 발성구조

음성은 호흡기관과 음식물을 씹는 기관의 계획되고, 정형화된 움직임에 의해 만들어지는 음향신호이다. 공기는 폐압으로 인해 기관을 지나 인두 혹은 목구멍으로 방출되는데 가장 여린 모음을 발생할때의 폐압은 4cm H₂O정도가 되며 아주 세게 발생할때의 폐압은 대략 20cm H₂O정도가 된다. 기관의 상부는 후두로 덮여 있는데 연발조직이 성대라고 하는 두장의 인대와 근육을 싸고 있다. 그리고 성대 사이의 틈(구멍)을 성문(glottis)라고 한다. 유성음은 성대의 떨림운동으로 생성되며 이 방법에서의 한 조음을 발성이라고 한다.

성량적으로 발생과정은 다음과 같다. 초기에 질량이 있고 팽팽한 성대가 서로 맞물려 있다고 가정하자. 성문아래 압력이 횡적인 가속으로 뿔사이를 벌려

높을 정도로 상승되면 구멍사이로 공기의 흐름이 생기게 된다. Bernoulli관계에 따라 지역적인 압력이 감소하게 되고 성대가 원위치로 복귀하려는 힘이 작용하게 된다. 성대가 맞붙게 되면 공기의 흐름이 줄게 되고 다시 지역압이 성문밑 음압값에 근사하게 되어 다시 이완 과정을 반복하게 된다. 성대의 질량과 탄성 그리고 성문밑 압력이 진동주기를 결정한다. 이 주기는 일반적으로 성대의 고유주기보다 짧다.

성대의 떨림에 의해 만들어지는 구멍의 가변면적은 성대 다음의 음향시스템을 여기시키는 공기의 준주기펄스를 만들게 된다. 과거 수십년에 걸쳐 성대의 떨림운동에 대해 매우 상세히 연구가 수행되었다. 구강 뒷부분에 45도 되게 거울을 정착시키고 적절한 주파수로 조정된 stroboscope를 통해서 보게 되면 고정된 혹은 느리게 변하는 성문의 떨림 패턴을 상세한 부분까지 관찰할 수 있다. 더 발전된 기술은 Far-sworth에 의해 제안된 것으로서 고속의 촬영기를 이용하여 초당 4000장의 동화상을 찍는 것이다. 그림 1과 같이 렌즈와 거울을 이용하여 성대에 강한 빛을

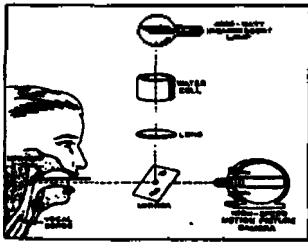


그림 1. 성문의 고속 촬영기술.

비추고 큰거울의 구멍을 통해 촬영한다. 이 방법으로 구한 성인 남성의 성문떨림 한 주기 사진 6장을 그림 2에 보였다.

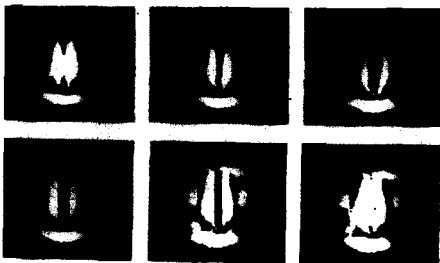


그림 2. 성문떨림의 연속된 한 주기(약 8msec)

성문을 통과하는 공기의 흐름용량에 대한 시간함

수는 성문의 열린 면적에 관계가 있다. 일반적인 피치의 음성일 경우는 성문파를 삼각파로 근사시킬 수 있으며 duty factor는 0.3~0.7 정도이다. 따라서 성문파는 고주파 성분이 두드러진 주파수 스펙트럼을 갖고 있으며 고주파대역에서는 12db/octave로 진폭이 감소한다.

성문파는 단일 발생자의 경우에도 피치와 세기에 따라 매우 다양하게 변한다. 피치가 낮고 약한 음의 경우 성문밑 음압이 낮고 성대의 duty factor는 높으며 성문파의 진폭도 낮다. 피치가 높고 쉼 음의 경우는 성문밑 음압이 높고 duty factor는 작으며 성문파의 진폭이 커진다.

나. 성문 임피던스 모델

발성구조에 대해 정량적으로 분석하고 성문파의 음향적 특성을 추정한다. 몇가지 추정에는 성문밑 음압, 성문기관의 크기, 성문면적의 시간 함수에 대한 지식이 기본이 된다. 한 성인 남성의 성문 밑 기관의 크기를 그림 3에, 이 구조의 전기적 등가회로를 그림 4에 나타내었다.

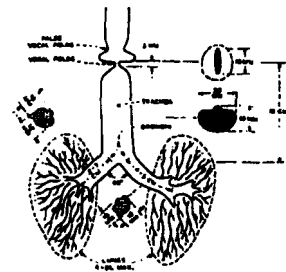


그림 3. 인간의 성문밑 시스템의 모형도.

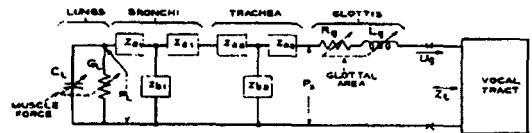


그림 4. 성문밑 시스템의 등가회로.

성문에서의 지역압변화에 의해 성대의 준주기적인 열림과 닫힘이 일어나는데 이는 임피던스($R_g + j\omega L_g$) 값을 변화시키게 되며 공기흐름을 변조하게 된다. 따라서 성도로 전달되는 공기는 펄수의 형태가 된다. 기관지와 기관의 단면적이 성문에 비해 크므로 여기

시의 압력강하는 무시가능하다. 따라서 성문밀 압력은 폐압과 거의 같다. 성문밀압력의 대부분이 소비되는 동안 성문의 가변면적은 시변 임피던스이다. 성문밀 위치에너지가 glottal volume velocity pulse 형태의 동적에너지로 효과적으로 바뀌게 된다. 수천 Hz 이하의 주파수에서는 성문 임피던스의 주성분을 저항으로 볼 수 있다.

시변의 성문 임피던스와 성문면적, 성문핵이 몇가지 유형의 특성을 살펴보자. 성문 임피던스의 초기 추정을 위해서 성문의 저항에 대한 인덕턴스의 비가 성문면적변화주기에 비해 작다고 ($L_g/R_g \ll T_0$) 가정한다. 이 경우 성문파는 계속적으로 형성되는 안정상태의 연속으로 간주할 수 있으며 성문저항을 추정하는데 이용한다. 성문의 크기가 음향자극신호의 파장에 비해 작고, 공기의 흐름속도가 음속에 비해 매우 낮고, 일정하며, 점막에 의한 손실이 없다고 가정한다. 이때 관계식은

$$P_1 - P_2 = \frac{\rho U^2}{2} \quad (1)$$

입자속도는

$$u = \left[\frac{2(P_1 - P_2)}{\rho} \right]^{1/2} \quad (2)$$

성문저항을 R_g^* 라 두면

$$R_g^* = \frac{\rho u}{2A} = \frac{\rho U}{2A} = \frac{(2\rho(P_1 - P_2))^{1/2}}{2A} = \frac{2\rho P_s}{2A} \quad (3)$$

이 된다. 그러나 실제로 점막에 의한 손실이 존재하므로

$$R_g = R_g^* + K \left(\frac{\rho U}{2A} \right) \quad (4)$$

이다. Van den Berg가 실험적으로 구한 저항은

$$R_g = \frac{P_s}{U} = \frac{12\mu d}{lw^3} + 0.875 \cdot \frac{\rho U}{2(lw)} \quad (5)$$

이다. 여기서 μ 는 viscosity 상수이다. (5)식에 의해서 P_s 와 A 값으로 volume flow U 값을 구하게 된다. 이상은 안정된 상태에서 구한 값이다. 성문면적이 시변함수인 초어젖음때의 $U_g(t)$:

$$U_g(t)R_g(t) + \frac{d}{dt} [L_g(t)U_g(t)] = P_s \quad (6)$$

단 $L_g(t) = \rho d / A(t)$.

의 1차 비선형 미분함수식에 의해 구하게 된다. 그런데 시정수 L_g/R_g 가 피치주기 T_0 보다 작으므로 성문파는 (5)식으로 부터 근사적으로 구할 수 있다. 성문에 대한 고속촬영사진으로 부터 측정된 성문면적 파형, 음성의 세기와 기관압에 대한 직접측정치를 이용하여 (5)식의 관계로부터 구한 성문파를 그림 5에 나타내었다.

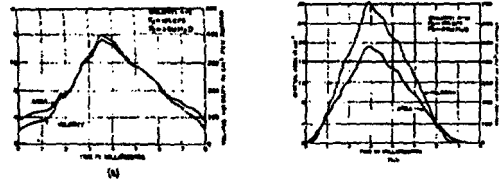


그림 5. 성문면적과 계산된 성문파.

다. One mass model (Self oscillation model)

성대를 그림 6와 같이 하나의 이동 가능한 덩어리로 나타낼 수 있다. 성대의 두께는 d , 길이는 l 이다. 덩어리가 횡적으로 x 만큼 이동하면 성문면적 A_g 와 volume flow U_x 를 변화시킨다. 안정상태에서의 성대면적을 A_{g0} 라 한다. 역학적 진동기는 성문밀 압력과 성문에서의 Bernoulli의 압력작용에 의해 힘을 받는다. Bernoulli압력은 U_x^2 에 의해 결정되고 U_x 는 성문의 비선형, 시변의 음향임피던스에 의해 정해진다.

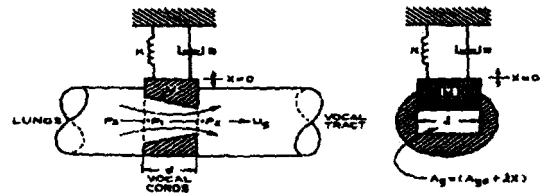


그림 6. 성대의 one mass model.

성문의 음향임피던스는 R_g, R_k, L_g 로 정의된다. 이 값들은 $A_g(t)$ 에 의해 결정되고 R_k 는 $|U_x|$ 에 의해 결정된다. $A_g(t)$ 는 P_s 와 U_x 와 성대음 식임과 힘의 함수를 나타내는 비분량방식으로 관계가 정의된다.

라. Two mass model

초기의 one mass 모델에서는 유원과 성도사이의 관계에 대한 고려없이 피치, 성문파, glottal duty factor, 성도형태, 성문면적 등과 같은 파라미터를 사

용해서 합성정보를 추출했다. 여기에서 얻어진 성분면적의 파형과 성분파는 고속의 동화상과 역필터에

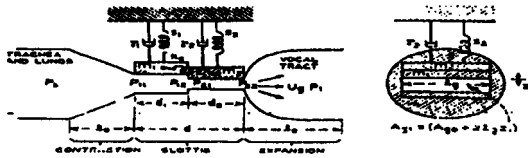


그림 7. 성대의 two mass model.

서 얻어진 파형과 동일했다. 그러나, 음원과 성도 사이에 나타난 음향적인 관계와 같은 자세한 성대의 생리학적인 기능에 대한 설명이 미흡했고, 성도의 포만트주파수 이상에서의 capacitive input load의 진동에 대해서는 규명할 수 없었다. 또 폐와의 연관성, 성대 가장자리에서의 위상천이에 대한 설명도 없었다. 이와 같은 문제점의 해결을 위해 two mass 모델이 도입되게 되었다. 그림 7과 같이 성대를 2개의 mass로 구성된 self-oscillating source로 모델링 했다. 성문의 음압분포는 Bernoulli 방정식을 이용하여 구하는데 van den Berg의 실험치에 의한 glottal flow는 (7)식과 같다.

$$U_g(t) = A_g(t) \left(\frac{2\rho}{K} \right)^{1/2} \quad (7)$$

단 $U_g(t)$: glottal flow, $A_g(t)$: glottal area,
 ρ : air density, K : kinetic resistance,

이로부터 음압분포에 대한 식을 유도하면,

$$P_3 - P_{11} = 1.37 \frac{\rho}{2} \left[\frac{U_0}{A_{01}} \right]^2 + \int_0^L \frac{\rho}{A_c(x)} dx \cdot \frac{dU_0}{dt}$$

$$P_{11} - P_{12} = 12 \frac{\mu l_0^2 d_1}{A_{01}^3} U_0 + \frac{\rho d_1}{A_{01}} \cdot \frac{dU_0}{dt}$$

$$P_{12} - P_{21} = \frac{\rho}{2} U_0^2 \left(\frac{1}{A_{02}^2} - \frac{1}{A_{01}^2} \right)$$

$$P_{21} - P_{22} = 12 \frac{\mu l_0^2 d_2}{A_{02}^3} U_0 + \frac{\rho d_2}{A_{02}} \cdot \frac{dU_0}{dt}$$

$$P_{22} - P_1 = -\frac{\rho}{2} \left[\frac{U_0}{A_{02}} \right]^2 \cdot 2 \frac{A_{02}}{A_1} \left[1 - \frac{A_{02}}{A_1} \right] \quad (8)$$

위 음압분포에 대한 식을 이용하여 연속시스템으로 모델링하기 위해 R, L값을 정의하면

$$R_c = 1.37 \frac{\rho}{2} \frac{|U_0|}{A_{01}^2}, \quad L_c = \int_0^L \frac{dx}{A_c(x)}$$

$$R_{v1} = 12 \frac{\mu l_0^2 d_1}{A_{01}^3}, \quad L_{01} = \frac{\rho d_1}{A_{01}}$$

$$R_{12} = \frac{\rho}{2} \left(\frac{1}{A_{02}^2} - \frac{1}{A_{01}^2} \right) |U_0|$$

$$R_{v2} = 12 \frac{\mu l_0^2 d_2}{A_{02}^3}, \quad L_{02} = \frac{\rho d_2}{A_{02}}$$

$$R_c = -\frac{\rho}{2} \cdot \frac{2}{A_{02} A_1} \left[1 - \frac{A_{02}}{A_1} \right] |U_0| \quad (9)$$

으로 정의되고 이를 이용하여 연속시스템을 구성하면 그림 8과 같다.

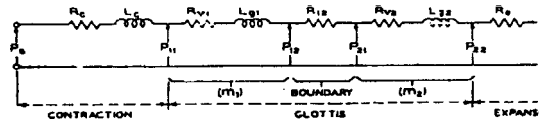


그림 8. 성문의 동가회로.

그리고 이에 대한 미분방정식은 $Z_v(t) = 0$ 이라고 하면

$$(R_{K1} + R_{K2}) |U_g| U_g + (R_{v1} + R_{v2}) U_g + (L_{01} + L_{02}) \frac{dU_g}{dt} - P_3 = 0 \quad (10)$$

의 관계식에 의해 P_3 , A_g , cord tension 값이 주어지면 $U_g(t)$ 를 구하게 된다. 이 모델은 성분면적, 성대 인장력, 안정상태의 성분면적, 성대 아래, 뒷가장자리간의 위상차, 포만트주파수보다 높은 진동주파수(가성) 등 인간의 성대작용에 보다 가까운 요소의 구현을 가능하게 하였다. 규칙합성과 같은 기술에 이 모델을 응용하면 유성음의 완전한 합성을 가능하게 해줄 수 있다.

B. 역필터 이용법

음원 성분파는 유성의 분석 및 합성에 매우 중요하다. 그런데 최종의 음성파형은 음원과 성도의 전달인 피드백이 결합된 함수이므로 음원을 정확히 추정하는 데는 성대의 모양이 중요한 역할을 한다. 성분파는 동화상을 이용하여 성대의 열면면적을 측정하는 방법으로 구할수 있지만 성분이 측정하기 어려운 지점에 있고 특히 동작중의 성분을 촬영하기가 쉽지 않다. 그래서 성도전달함수를 추정하고 이로부터 구한 역필터를 이용하여 성분파를 구하는 방법이 제안되었다.

가. Miller의 방법

1차 포만트주파수에 대한 역필터를 이용하여 성문파를 구하고자 하는 것으로 음성의 스펙트럼을 분석하여 1차 포만트의 위치와 공진주파수에서의 Q값을 구한다. 그리고 성문이 닫혔을 때의 volume velocity가 0이라 가정하에서 구한 소자값으로 역필터를 구성한다. 실험은 그림 9와 같은 시스템으로 수행한다. 제1포만트 주파수에 국한시키기 위해서 성능이 좋은 가변저역통과필터로 고주파 성분을 제거한다.

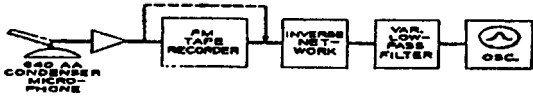


그림 9. 실험장치 구성도.

그러나 높은 피치의 신호에 대해서는 이 시스템의 성능이 감소하므로 성문파의 급격한 변화를 추정하는 어렵다.

나. Pneumotachograph 이용법

일반적으로 음향신호를 저주파특성이 좋은 압력감응 마이크로 녹음하고 녹음된 음을 그 순간의 성도 전달특성으로 구한 역필터에 통과시킴으로써 성문파를 얻게된다. 그런데 성도전달함수는 몇개의 복소쌍의 pole과 0 주파수에서의 zero를 가지고 있다. 따라서 역필터는 복소쌍의 zero와 0 주파수에서의 zero를 가져게 되므로 다음의 제약점을 가진다.

- (1). volume velocity waveform은 bias 값을 가진다. 즉 zero flow의 표시가 없다.
- (2). 저주파 잡음에 큰 영향을 받는다.
- (3). 잔폭고정이 어렵다.
- (4). 조음기관이 움직일 때 공기의 이동이 생겨서 저주파 잡음성분이 생긴다.
- (5). 연속음의 경우 동적으로 변하는 필터파라미터 값들을 알고 있어야 한다.
- (6). 성도의 수학적 모델이 정확하지 않을 때 오차가 발생한다.

(1), (2)항은 0주파수에서의 pole에 의한 것이고 (3)항은 입과 마이크와의 거리, 방향에 영향을 받기 때문이다. (4)항은 조음기관이 움직일 때 생긴 성분도 역필터링하기 때문에 생기는 문제로서 예전 문헌을 0.1초에 1cm 움직일 때 0.12 liter /sec의 공기 흐름이 부가적으로 발생하기 때문이다. 그런데 pneumotachograph (PTG) 마스크를 사용하여 음압대신 입에서의 volume velocity waveform을 구한다면 역필터에서 pole이 없어지게 되고 따라서 문제점 (1),

(2)항이 해결된다. 그리고 입과 코에서 방출되는 공기의 모든 흐름을 측정하므로 잔폭고정문제 (3)항이 해결된다. 그러나 (4), (5), (6)항의 문제점은 여전히 남게 된다. 그러나 PTG의 출력을 역필터링에 사용하기 위해서는

- (1). 발생할 때의 음압이나 공기흐름에 방해가 되지 않아야 하며
- (2). 출력이 volume velocity의 선형함수이고
- (3). 방출된 음향신호에 왜곡이 적어야 하며
- (4). 응답시간이 성문주기에 비해 짧아야 한다.

표준의 PGT 마스크는 방출된 음향신호의 왜곡이 심하고 지체공명으로 인해 응답시간에 심한 제약이 있다. 또한 높은 음향 임피던스를 가지므로 성도의 원래 포만트주파수를 변화시키고 필터링 현상까지 생긴다. 이런 제약점을 해결하기 위해 Rothenberg는 마스크저항, 선형성, 응답시간이 개선된 그림 10와 같은 구조의 새로운 PTG를 제안하였다. Volume velocity로 구한 성문파가 저주파 잡음이 적고 voice 의 onset특성이 뚜렷함을 보인다.

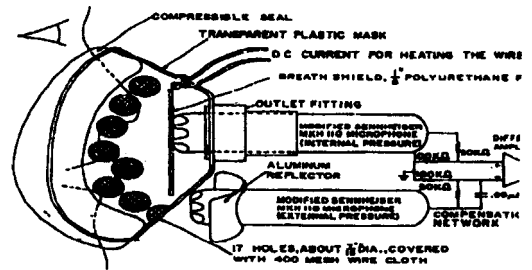


그림 10. Rothenberg pneumotachograph.

다. 신호처리에 의한 방법

만일 성문파나 성도전달함수 중 어느 하나가 정확히 정의된다면 다른 하나가 가정된 모델의 한계내에서 구해질 수 있다. 음성발성모델을 그림 11(a)와 같이 정의할 수 있다. 여기서 성문여기모델 신호 $e(n)$ 은 압력과 같은 물리적 신호가 아니고 성문에서의 volume velocity 신호 $U_g(n)$ 을 생성해 내는 성문모델 필터 $G(z)$ 의 수학적인 입력신호를 나타낸다. 유성음의 경우 $e(n)$ 은 주기적인 임펄스열로 나타난다.

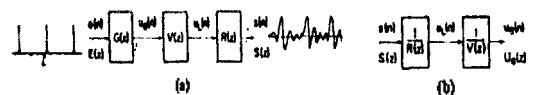


그림 11. (a) 음성발성의 선형모델 (b) 성문의 역필터 개념도.

성도모델 $V(z)$ 는 all-poll. 모델로

$$V(z) = [1 + \sum_{i=1}^k c_i z^{-i}]^{-1} \quad (11)$$

로 정의한다. 보통 저주파에서 방출임피던스 $R(z)$ 를

$$R(z) = 1 - z^{-1} \quad (12)$$

로 근사시킨다. 이상의 모델을 근거로 성문의 역필터링은 그림 11(b)와 같이

$$U_g(z) = \frac{S(z)}{V(z)R(z)} \quad (14)$$

로 개념적으로 정의 될 수 있다. 실제 분석에 들어가기 전에 고려할 사항으로는 U_g 의 정확한 dc 값을 구할 수 없다는 점과 녹음장치에 의한 극성의 반전, 전파 지연 등이 있지만 여기서는 이를 고려하지 않는다. 먼저 유효 구동함수를

$$Q(z) = U_g(z)R(z),$$

으로 정의한다. $R(z)$ 의 특성으로 인해 $q(n)$ 의 평균값은 0이다. 성문폐쇄가 안정상태라고 가정했을 때 음성신호와 성문파 그리고 유효 구동함수의 이론적 시간 관계를 그림 12에 나타내었다. L_c 와 L_o 를 성문폐쇄와 열림 위치라고 정의한다. 만일 $L_c \leq n < L_o$ 에서 $U_g(n)=0$ 라면 $q(n)$ 는 $L_{c+1} \leq n < L_o$ 에서 $q(n)=0$ 가 된다. 음성의 발생모델은

$$s(n) = \sum_{i=1}^k c_i s(n-i) + q(n) \quad (15)$$

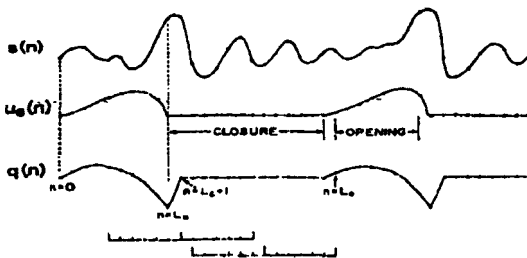


그림 12. 음성, 성문파, 유효 구동함수간의 시간관계.

의 형태가 된다. 그런데 $L_c \leq n < L_o$ 에서 성문이 폐쇄라 가정했기 때문에

$$s(n) = -\sum_{i=1}^k c_i s(n-i) \quad (L_{c+1} \leq n < L_o) \quad (16)$$

이 된다. 만일 M 차의, 단 $M \geq K$, 분석필터를

$$A(z) = \sum_{i=1}^M a_i z^{-i} \quad (a_0=1) \quad (17)$$

라 두면 $L_{c+1} \leq n < L_o$ 에서는 (16)과 (17)식에 의해

$$\varepsilon(n) = \sum_{i=1}^M (a_i - c_i) s(n-i) \quad (18)$$

만일

$$a_i = \begin{cases} c_i & \text{for } i=1, \dots, K \\ 0 & \text{for } i=K+1, \dots, M \end{cases} \quad (19)$$

의 조건이 만족되면 $\varepsilon(n)=0$ 이 된다. $s(n-M)$ 에서 $s(n+N-M-1)$ 까지의 분석 window에서 covariance 방법으로 $A(z)$ 를 구하기 위해서 전체자승오차를

$$\alpha_M(n) = \sum_{j=n}^{n+N-M-1} \varepsilon^2(j) \quad (20)$$

로 구하며 분석필터계수는 이 $\alpha_M(n)$ 을 최소화 함으로써 얻게 된다. 만일 n_1 에서 처음으로 $\alpha_M(n)=0$ 가 된다면 성문폐쇄점은

$$L_c = n_1 - 1 \quad (21)$$

로 정의 되고 이후 n_2 에서 $\alpha_M(n)=0$ 이 된다면 성문열림점이

$$L_o = n_2 + N - M - 1 \quad (22)$$

로 정의된다. 그런데 $\alpha_M(n)=0$ 는 단지 이론적으로 가능할 뿐이다. 그리고 $\alpha_M(n)$ 이 실제시스템의 gain을 고려하지 않은 것이므로 정확도를 높이기 위해 정규화한 값

$$\eta(n) = \alpha_M / \alpha_0 \quad (23)$$

을, 단 α_0 는 입력신호에너지, 사용하는 것이 좋다. 따라서 η 가 충분히 작으면 분석필터계수가 성문이나 방출에 의한 영향이 없는 성도모델 $V(z)$ 를 나타낸다고

본다. 이런 조건에서 성문파는

$$\hat{Q}(z) = S(z) / \hat{V}(z) = S(z) A(z) \quad (24)$$

$$U_c(n) = \hat{Q}(z) / (1-z^{-1}) \quad (25)$$

로 추정된다. 실제음성으로 부터 성문파를 구하는 역 필터시스템의 block diagram을 그림 13에, 그리고 분석결과를 그림 14에 나타내었다. 그런데 이 방법은 성문의 물리적 의미가 포함되지 않으며 입력신호의 녹음환경 등과 같은 여러 인자에 대해 매우 큰 영향을 받는 단점이 있다. 또한 성문폐쇄구간이 최소한 분석차수의 2배 이상이 되어야 하므로 높은 피치의 음성이나 폐쇄구간이 짧은 음성에 대해서는 이 모델이 적합하지 못하다.

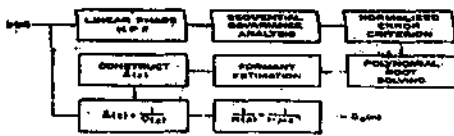


그림 13. 성문의 역필터 블록도.

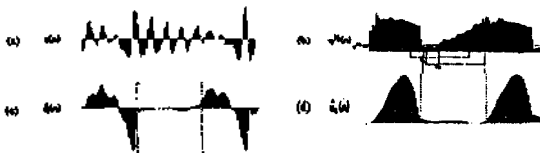


그림 14. (a) 음성신호 (b) 정규화된 오차신호
(a) 유효 구동함수 (b) 성문파.

C. 정합장치 이용법

여기서 제시하는 방법은 입술에서의 음향출력을 이용한다. 그러나 음파에서 성도가 미치는 영향을 나중에 제거하는 대신에 성도가 성문파를 처음부터 최초의 왜곡으로 전송한다는 조건을 설정함으로써 역 필터링을 제거하였다는 점에 큰 차이가 있다. 따라서 좀더 간단한 배치로 목적을 달성할 수 있다. 한쪽 끝은 열려있고 다른 한쪽 끝은 닫혀있는 단면적 5cm, 길이 17cm인 딱딱한 벽을 가진 균일한 관에서 단면적이 작은 구멍으로 관이 음향적으로 여기된다고 가정하자. 또 관의 임펄스 응답이 무한히 긴 시간이라 가정하자. 이때 출력은 입력과 임펄스 응답과의 convolution으로 얻어지므로 성문파의 주기가 3~12msec임을 고려할때 역필터링을 해도 원래의 입력신호를 복원할

수 없게 된다. 만일 이상적 균일성도가 먼 끝에서 정합된 상태로 단단한 벽의 균일관에 연결되어 있다고 가정하면 음원이 거의 무한 길이의 균일관으로 공급 되는 것과 같아진다. 따라서 지연을 고려하지 않는다면 이 관 안의 어디에서든 마이크를 이용하여 입력과 동일한 파형을 얻을 수 있게 된다. 이 경우는

- (1) 성도가 균일관으로 근사되고
- (2) 성도의 벽이 단단하며
- (3) 평면파로 전파되고 이상 기체임을 가정한 것이다.

이 중 (2), (3)항의 가정은 이미 대부분의 역필터링이나 성도에서의 전파해석에 사용되어 왔다. 이러한 가정이 부정확할 경우에는 단지 first-order perturbation이 발생한다고 알려져 있다. 또한 성도가 균일관이 아니더라도 무반향관을 연결할 경우 입력과 거의 동일한 파형을 얻을 수 있음을 보이는 실험결과를 이용해서 Sondhi는 그림 15과 같은 배치를 만들었다. 확장된 관은 내부직경 1 inch, 길이 6 ft의 황동으로 만들었고 반향없는 음향관은 거의 원추형이며 3 ft 정도 길이의 섬유유리로 만들었다. 직경 1/4 inch의

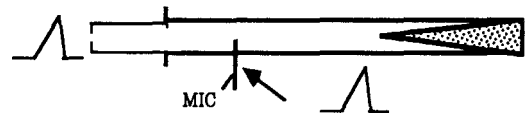


그림 15. 무반향 음향관.

electret 마이크를 개방된 끝에서 1 ft 정도에 위치시켰고 상대 진동을 방해하는 관의 입력을 제거하기 위해 음향관 주변에 작은 구멍을 만들었다. 이 방법의 잇점은 저렴하며 실시간으로 성문파를 얻을 수 있고 주변잡음에 덜 민감하여 일반 실험실에서도 실험이 가능한 점이다.

Ⅲ. 남녀의 성문파 비교분석

Sondhi의 금속관을 이용하여 10명의 성인 남녀를 대상으로 1) 정상음 2) 부드러운 소리 3) 쉿 소리 4) 기성 5) Creaky음 6) 숨가 혹은 감소하는 억양 7) 3음절 중 하나에 상세분준 음성들에 대한 성문파를 구하였다. 이러한 데이터를 이용하여 성문파의 rms intensity, 기본주파수, 위상스펙트럼 그리고 세기 스펙트럼 등이 다양한 분석을 하였다. 그 결과 기본주파수가 시간에 따라 변하면 음원은 두가지 다른 원리에

따라 움직임을 알 수 있다. 그 한가지는 성문파의 각 고조파 간의 관계는 일정하게 유지되지만 스펙트럼의 포락선은 기본주파수의 함수가 되어 주파수와 진폭이 변한다. 다른 하나는 각 고조파간의 관계가 기본주파수의 함수가 되어 변하지만 스펙트럼의 포락선은 변하지 않는 것이다. 남여의 성대파를 분석한 결과는 다음과 같다.

(1)정상음

- A. 스펙트럼 : 1200Hz 이상에서 octave 단위로 급격히 감소
- B. 성문파형 : 여성이 더 대칭적
- C. 기본주기와 2차 harmonic의 위상차 :
남성 : 170도-180도,
여성 : -150도 - -160도
- D. 기본주파수 : 남성(121-139Hz),
여성(155-209Hz)

(2)부드러운 소리

- A. 스펙트럼 : 더 가파르게 감소
- B. 성문파형 : 더 대칭적

(3)큰소리

- A. 스펙트럼 : 고주파에너지 증가
- B. 성문파형 : closing 구간이 짧고 급격함

(4)가성

- A. 스펙트럼 : -20dB/octave, 1200Hz부근에서 기울기는 완만
- B. 성문파형 : 남자(위상반전), 여성(기본주파수 증가)

(5)creaky voice

- A. 스펙트럼 : 느슨한 기울기
- B. 성문파형 : 비정형적
- C. 기본주파수 : 주기마다 기본주파수의 큰 변화(30-90Hz)

(6)의문문

- A. 스펙트럼 : 전체 spectral envelope를 일정하게 유지
- B. 성문파형 : closing의 점유도가 기본주파수에 비례

(7)평서문

- A. 스펙트럼 : 주파수변화에 큰 영향을 받지 않는다.
- B. 기본주파수 : 의문문에 비해 변화가 작다.

(8)강세에 따른 변화

의문형과 평서형의 조합으로 해석 가능

IV. 성문파의 모델링

양질의 음성합성 또는 정확도가 높은 음성인식을 위해서는 성도 뿐만 아니라 음원의 특성을 설명할 수 있는 수학적 모델링이 주요한 연구사항이 되었다. 성문파의 모양과 주기는 매우 다양하게 변한다. 이런 변화가 음성의 자연성과 음질에 어느 정도 영향을 미치는가는 중요한 연구과제이다. 일반적으로 합성기에서는 -12db/octave의 특성을 가지는 고정된 형태의 성문파를 사용하는데 실제 성문파의 특성을 복제하는데 유연성이 없으므로 합성음의 자연성이 떨어지게 된다.

여기서는 삼각파 모델의 특성분석과 이미 제안된 Rosenberg model, Hedelin model, Fant model, Fujisaki model 들을 비교검토한다.

A. 삼각파 모델

일반적으로 중간정도의 피치와 세기를 가진 음성의 환경에서 성문파가 그 모양에서 삼각파로 근사화시킬 수 있다. 그러므로 삼각파의 스펙트럼 특성이 유성음원과 관련성을 가지고 있다. Dunn, et al. 이

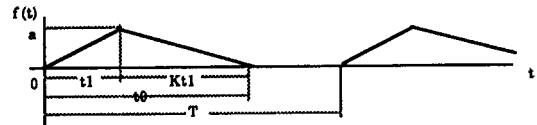


그림 17. K=1, K=11/12인 두 삼각파의 크기 스펙트럼.

실제 성문파의 파형과 스펙트럼과의 관계에 대해 상세히 연구를 수행하였다. 그림 16과 같은 삼각파의 경우 Laplace 변환은

$$F(s) = \frac{a}{s^2} \left[\frac{1}{\tau_1} - \left(\frac{1}{\tau_1} + \frac{1}{\tau_2} \right) e^{-s\tau_1} + \frac{1}{\tau_2} e^{-s(\tau_1+\tau_2)} \right] \tag{26}$$

스펙트럼의 zero는

$$(e^{-(K+1)s\tau_1} - (K+1)e^{-s\tau_1} + K) = 0 \tag{27}$$

식에서 구하게 된다. $x = e^{-s\tau_1}$ 이라 두면

$$x = e^{-s\tau_1} = e^{-(K+1)s\tau_1} \tag{28}$$

k=1일때와 k=11/12(혹은 12/11) 일때의 |F(jw)|의 스펙트럼크기를 그림 17에 나타내었다. 그림 17은 삼각파가 대칭일때 스펙트럼이 sin² x/x²의 모양을 보이기! 비대칭으로 변하면 스펙트럼상의 골짜기를

메꾸는 현상이 일어남을 보여준다. 그리고 대칭성이 조금만 변해도 스펙트럼이 크게 변함을 알 수 있다.

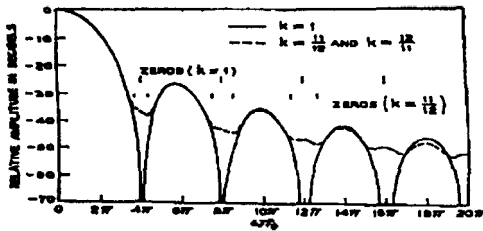


그림 18. 각 성문파 모델에 의한 성문파형

B. 삼각파 모델과 변형된 모델과의 특성비교

삼각파의 경우에는 일차 미분치의 불연속점이 3개이다. 반일 rounding이나 smoothing으로 이러한 불연속점을 제거하면 어떤 효과가 나타날까? 예로서 half sine, half ellipse, raised cosine을 사용할 경우 앞 2개는 2개의 불연속점을 가지고 나머지 하나는 불연속점이 없다. 이들의 스펙트럼 특성은 Half-sine wave

$$f(t) = a \sin \beta t, \quad 0 \leq t \leq \frac{\pi}{\beta}, \quad \beta = \frac{\pi}{\tau_0}$$

$$= 0, \quad \text{elsewhere} \quad (29)$$

$$F(\omega) = \left[\frac{\beta a}{\beta^2 - \omega^2} \right] (1 + e^{-j\omega \pi / \beta})$$

zero는

$$\omega = \pm \frac{(2n+1)\pi}{\tau_0} = \pm (2n+1)\beta, \quad n=1,2,\dots$$

Half-ellipse

$$f(t) = \frac{4}{\pi \tau_0} \left[1 - \left(\frac{2t}{\tau_0} \right)^2 \right], \quad |t| \leq \frac{\tau_0}{2}$$

$$= 0, \quad \text{elsewhere} \quad (30)$$

$$F(\omega) = \frac{2J_1(\omega \tau_0 / 2)}{\omega \tau_0 / 2}$$

zero는

$$\omega = J_1(\omega \tau_0 / 2) \text{의 근이다.}$$

Raised-cosine

$$f(t) = a(1 - \cos \beta t), \quad 0 \leq t \leq \frac{2\pi}{\beta}, \quad \beta = \frac{2\pi}{\tau_0}$$

$$= 0, \quad \text{elsewhere} \quad (31)$$

$$F(\omega) = a \left[\frac{\beta^2}{j\omega(\beta^2 - \omega^2)} \right] (1 + e^{-j2\omega \pi / \beta})$$

zero는

$$\omega = \pm n\beta = \pm \frac{2n\pi}{\tau_0}, \quad n=2,3,\dots$$

이 결과 rounding이나 펄스의 모양에 작은 변화도 zero 패턴과 성문파의 스펙트럼에 큰 영향을 미침을 알 수 있다.

C. 성문파의 zero가 인지에 미치는 영향

성문파의 모양과 주기는 매우 다양하게 변한다. 이런 변화가 음성의 자연성과 음질에 어느 정도 영향을 미치는가는 중요한 연구과제이다. 반일 성문파의 zero가 포먼트 주파수와 가까이 위치하게 되면 스펙트럼을 변화시키고 따라서 인지에도 영향을 미친다. 성도 자극음은 연속음에서 그 파형이 매우 빨리 변하므로 인지에 영향을 준 정도로 오래 안정되지 않는다. 그리고 발성자는 청각의 feedback을 통해서 포먼트 주파수의 원하지 않는 억압을 최소화하도록 성문파를 교정한다. Rosenberg의 정형화된 성문파를 이용한 합성음의 청취실험 결과는 성문파의 세부적인 부분을 제거해도 음질의 저하는 가져오지 않음을 보인다. 또한 매 주기마다의 상세한 파형보다는 몇 주기에 걸친 평균값의 변화가 인지에 더 중요한 영향을 미침을 보인다.

D. 기존 성문파 모델

가. Rosenberg model과 Hedelin model

Rosenberg는 성문파의 진폭, 지속시간, 기울기(skew)의 변수들을 사용해서 2개의 trigonometric segment들로 수학적 모델을 만들었다. 여기에 성문의 적주파, drift를 부가한 것이 Hedelin 모델이며 LPC-vocoder에 이용하였다.

나. Fant model

음원과 필터의 곱으로써 음성 신호를 정의하는 것은 이론적인 해안이 있다. 형태 진동에서의 구조적 음향적 작용을 무시한다면, 이렇게 음원과 필터의 함수적 관계와 기능을 구분해야 하는지에 대한 문제에 접하게 된다. 음원을 정의하는 한가지 방법은 상도 임피던스를 난락회로로 보았을 때 성문을 통한 volume velocity로 정의될 수 있고, 이 성문 흐름

(glottal flow)은 성도 공진(vocal tract resonance)의 진동 요소(oscillatory component)들을 포함하고 있기 때문에, 이 방법으로 정의된 음원은 필터 합수에 종속적이라는 것이 명백하다. 이 성문파를 다음과 같이 모델링 할 수 있다.

성문이 개방될 때

$$U = \frac{1}{2} U_0 (1 - \cos \omega_0 t) \quad (32)$$

성문이 닫힐 때

$$U = U_0 [K \cos(\omega_0 t - \pi) - K + 1] \quad (33)$$

다. Ananthapadmanabha model

위의 모델들에서 중요시했던 연속성을 지키기 않을 경우 문제점이 발생한다. 그런데 이러한 효과(conner effect)는 예외적이라기 보다는 일정한 범칙에 가깝다는 것이 밝혀 졌다. 이에 따라 파라볼릭 함수로 모델링한 terminal return phase를 변수에 포함시켜 모델링한 것이 Ananthapadmanabha 모델이다.

라. LF(Liljencrants Fant)model

LF 모델에서 고려되어지는 변수들 Tp, Te, Ta, Ee가 주어졌을 때 이로부터 성문파를 유일하게 결정할 수 있다.

성문이 개방될 때

$$E(t) = E_0 e^{\alpha t} \sin \omega_0 t \quad (34)$$

성문이 닫힐 때

$$E(t) = \frac{-E_0}{\epsilon t_0} [e^{-\epsilon(t-t_0)} - e^{-\epsilon(t_0-t_0)}] \quad (35)$$

마. Fujisaki model

성문파의 미분치가 다항식(polynomial)의 segment들로 구성되어지는 모델을 제안했는데, 이 모델을 이용하면 변수들의 수를 확장하기가 쉽고, 모델링에서의 레벨수를 자세히 나타낼 수 있다. 대부분의 모델에서 open phase duration(W), pulse skew(s), time interval from glottal closure to maximum negative flow(D)를 제어하는 3개의 시간 변수들을 가지고 있으며, 또한 성문 개방시의 경사(A), 성문이 닫힐 때의 경사(B), 성문이 닫혀가는 부분의 경사(C)를 제어하는 3개의 진폭들을 가지고 있다. 여기서

A변수는 다른 모델들에서 공통적인 것은 아니지만, 제안된 성문모델들을 비교 검토하면서, Fujisaki 모델과 같은 새로운 성문파 모델의 특성을 개방시 중요한 변수들이 되므로 이것을 포함시킨다. 이 모델의 수식은 식 36과 같다. 이 성문 모델을 도입하면 기존의 LPC에서 약 3~4dB의 prediction error가 감소한다.

$$\begin{aligned}
 g(t) &= \frac{2A + R\alpha}{R} t + \frac{A + R\alpha}{R} t^2 & 0 < t < R \\
 &= \alpha(t-R) + \frac{3B - 2F\alpha}{F^2} (t-R)^2 - \frac{2B + F\alpha}{F^3} (t-R)^3 & R < t < W \\
 &= C - \frac{2(C-\beta)}{D} (t-W) + \frac{C-\beta}{D^2} (t-W)^2 & W < t < W+D \\
 &= \beta
 \end{aligned}$$

$$\text{단, } \alpha = \frac{4AR - 6FB}{F^2 - 2R^2}, \quad \beta = \frac{CD}{D - (T - W)} \quad (36)$$

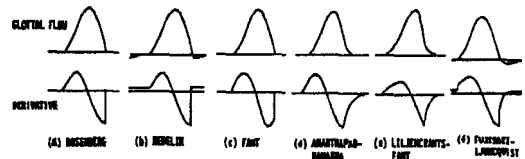


그림 16. 삼각파를 이용한 성문파의 근사.

V. 결 론

지금까지 성문파의 물리적, 생리적 생성원리, 음향학적인 추정, 역필터를 이용한 추정, 무반향관을 이용한 추정 그리고 음운환경 및 발생자에 따른 특성변화등에 관해 정리 기술하였다.

본 연구를 통하여 유성음원의 속성과 추정방법 및 특성변화를 보임으로써 합성음의 다양함, 즉 자연성과 개인성 구현을 위해서는 음원의 정확한 추정이 필수적임을 알 수 있다. 아직까지 국내에서는 성문파에 대한 관심이 부족했던게 사실이지만 텍스트 형태로 저장된 데이터베이스를 엑세스해서 음성으로 서비스해주는 음성출력부 개발이나 자동동역전화 개발에 필요한 원천기술로서 연구개발의 활용도는 매우 높다.

따라서 앞으로는 한국어를 대상으로 유성음원의 추정과 분석, 그리고 이의 모델링과 구현에 대한 연구가 절실하다고 생각된다.

참 고 문 헌

1. R.L. Miller, "Nature of the vocal cord waves," *J. Acoust. Soc. Amer.*, vol.31, pp.667-677, 1959.
2. D. Y. Wong, J. D. Markel, and A. H. Gray, Jr. "Least squares glottal inverse filtering from the acoustic speech waveform," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-27, pp.350-355, Aug. 1979.
3. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*. New York : Springer, 1976.
4. M. M. Sondhi, "Measurement of the glottal waveform," *J. Acoust. Soc. Amer.*, vol.57, pp.228-232, 1975.
5. J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, Second Edition, New York : Springer Verlag, 1972.
6. R. B. Mosen, A. M. Engebretson. "Study of

- variations in the male and female glottal wave," *J. Acoust. Soc. Amer.*, vol.62, pp.228-232, 1977.
7. M. Rothenberg, "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," *J. Acoust. Soc. Amer.*, vol.53, pp.1632-1645, 1973.
8. J. L. Flanagan, K. Ishizaka, and K. L. Shipley, "Synthesis of speech from a dynamic model of the vocal cords and vocal tract," *B.S.T.J.*, vol.54, no.3, pp.485-506, 1975.
9. S. Parthasarathy and D. W. Tufts, "Excitation-synchronous modeling of coiced speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp.1241-1249, Aug. 1987.
10. G. Fant, J. Liljencrants, and Q. G. Lin, "A four-parameter model of glottal flow," *STL-QPSR 4/1985*, pp.1-13, 1985.
11. Fujisaki, H. and M. Ljungqvist, "Proposal and evaluation of models for the glottal source waveform," *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, pp.1605-1608, 1986.

▲이 정 철

1961년 10월 7일생

1984년 3월 : 서울대 전자공학과 졸업(공학사)

1988년 2월 : 서울대 대학원 전자공학과 졸업(공학석사)

현 재 : 서울대 대학원 전자공학과 박사과정

1985년 9월~현재 : 한국전자통신연구소 신호처리연구실(연구원)



▲안 수 길 : 현 서울대학교 전자공학과 교수 (9권 1호 참조)