

모음의 음향적 특징

양 병 곤 (동의대 영문과)

1. 음원-여파기 이론

음원-여파기 이론(Source-Filter Theory)이란 허파로부터 나온 기류가 음향적으로 근원(source)이 되는 후두(larynx)와 여파기(filter) 역할을 하는 성도(vocal tract)를 거쳐 인간의 음성 생성이 일어난다는 이론이다. 그림1은 이 과정을 보인다.

그림 1

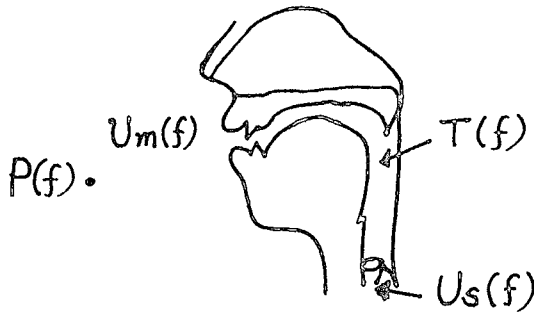
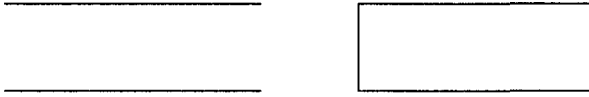


그림1에서 보는 바와 같이 후두에서 성대가 떨리면서 유성음이나 마찰음의 원천이 되는 기류가 생기고, 이것이 구강이나 비강을 거치면서 서로 다른 음향적 에너지가 걸러져서 음성이 생성된다. 공명체인 성도는, 마치 체가 일정한 크기의 알맹이만을 통과시키는 것처럼, 생성된 음원 중에서 일정한 주파수의 에너지를 통과시킨다. 다시 말해서, 혀의 위치, 턱을 벌리는 정도, 입술 모양 등에 의해 성도 모양이 달라질 때마다 다른 크기의 체가 되어, 어떤 주파수에서는 소리 에너지의 전달을 억제하고 다른 주파수에서는 최대의 에너지를 허용한다. 그래서 기류에 방해없이 성도를 통과하는 모음의 경우, 허파로부터 성대를 통과하기까지는 동일한 음이지만, 다른 모양의 성도를 지나면서, 즉, 다른 모양의 여파기를 거치면서 에너지의 분포가 변해 다른 모음으로 들리게 된다.

이러한 관계를 물리적으로 살펴보면, 그림1에서 주파수 f 에서의 성문하 음원의 진폭함수를 $U_s(f)$, 입술 밖의 진폭함수를 $U_m(f)$ 라고 하면 전달함수는 $T(f) = U_m(f) / U_s(f)$

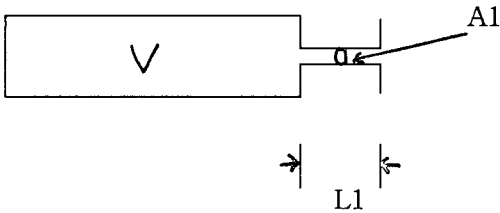
그림3.



$$C/2L, C/2L, 3C/2L, \dots$$

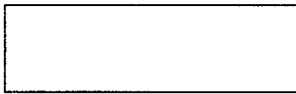
C는 상온에서의 소리의 속도로 34000 cm/sec 이다.

또한, 그림4처럼 한쪽 끝이 좁게 개방된 헬름홀츠 공명기로 된 관의 최저주파수는 그림4.



$f_1 = C / \left\{ 2\pi \sqrt{\frac{VL_1}{A_1}} \right\}$ ($\pi = 3.14$) 성문부위는 진동할 때는 제외하고 보통 닫혀 있는 그림5와 같은 균일관으로 비유할 수 있다.

그림5.



물리학에서 이런 관의 자연 주파수는 폭보다는 길이가 길다고 가정했을 때 다음과 같이 나타낸다.

$$f = C/4L, 3C/4L, 5C/4L, \dots$$

여기서 $C = \lambda * f$ 이다. λ 는 파장이다. $C/4L$ 이므로 $\lambda = 4L$ 이므로 성도의 길이는 파장에 해당한다. 이 관계를 모음에 적용하여 예상주파수를 구하거나 성도의 길이를 추정할 수 있다. 구체적으로 보면 그림5는 영어의 schwa모음인 [어]의 균일관과 비슷

하다. 보통 미국남성의 성도길이는 17cm이므로

$$F1 = C/4L = 34000/4 \times 17 = 500 \text{ Hz}$$

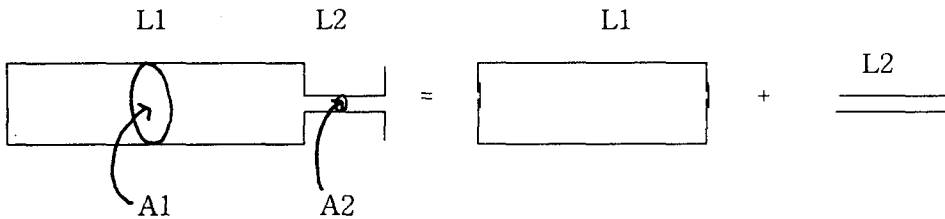
$$F2 = 3C/4L = 3 \times 34000/4 \times 17 = 1500 \text{ Hz}$$

$$F3 = 5C/4L = 5 \times 34000/4 \times 17 = 2500 \text{ Hz}$$

만약 성도의 길이 L이 짧아지면 (한국인 남성의 경우 16 cm) 형성음 값이 높아진다. 15 cm길이의 여성의 성도일 때는 그 값이 F1=567 Hz, F2=1700 Hz, F3= 2833 Hz, F4=3967 Hz . . . 등으로 나타난다. 또한, 성도에 공기보다도 빠른 속도로 전달되는 헬륨이나 산소로 차있다면 주파수가 상승한다.

모음 [i]의 추정 모형은 그림6과 같다.

그림6.



여기서는 두 개의 관의 합에다가 헬륨홀츠의 공명 주파수도 포함해야한다.

그래서

$$F1 = C / \left\{ 2\pi \sqrt{\frac{L1A1L2}{A2}} \right\}, \quad C/2L1, C/L1, C/2L2, C/L2 \dots$$

$$L1 = 9 \text{ cm}, L2 = 8 \text{ cm}, A1 = 5 \text{ cm}^2, A2 = 0.5 \text{ cm}^2$$

이 값에서 계산하면 F1= 202 Hz, F2= 1890 Hz (C/2L1), F3= 2125 Hz (C/2L2) 등이 나온다. 공식에서 F2는 L1에 영향을 받고 F3는 L2에 영향을 받으며 C=λ*f의 관계를 활용하면, 앞에서와 같이, 형성음 주파수값을 이용하여 인강과 구강내 성도의 길이를 추정해낼 수 있다. 인강(LB)은 성문에서 연구개까지의 지역을 말하며, 구강(LF)은 앞니에서 인강벽까지를 말한다. 각각의 길이는 [i]모음의 제2형성음(F2)과 제3형성음(F3)을 수집한 뒤 다음의 공식에 의해 추정할 수 있다.

$$LB = C / (2 \times F2)$$

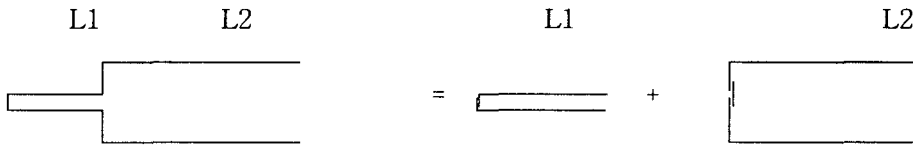
$$LF = C / (2 \times F3)$$

구한 값은 모델의 단순성 때문에 추정치에 불과하지만, Fant는 스웨덴 여성의 인강이 남성의 인강보다 2.1cm 더 짧고, 여성의 구강은 남성의 구강보다 약 1.25cm 더 짧음

을 추정 한 값이 실제 X-선 촬영으로 구한 자료와 잘 맞아 떨어졌음을 밝힌 바있다.

모음 [a]는 위의 물리적인 단순 모델로 분석해보면 그림7과 같다.

그림7.



인강이 구강보다 좁으면, 두개의 한쪽이 닫힌 관에 해당한다. 여기서 발생하는 주파수는 두 관 안에서 나온 주파수의 합에 해당한다.

$$C/4L1, 3C/4L1, 5C/4L1. \dots, \quad C/4L2, 3C/4L2, 5C/4L2. \dots$$

성인 남자의 값은 보통 $L1 = 8 \text{ cm}$, $L1 = 9 \text{ cm}$ 이다. 따라서

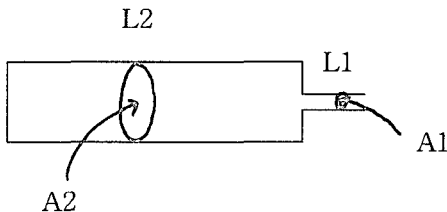
$$1063, 3188. \dots, \quad 944, 2833. \dots$$

이것을 순서대로 배열하면

$$F1=944 \text{ Hz}, F2= 1063 \text{ Hz}, F3=2833 \text{ Hz}. \dots \text{등으로 된다.}$$

모음 [U]의 경우는 그림8과 같이 도식화할 수있다.

그림8.



이것도 헬름홀쯔 공명관에 해당하므로

$$F1= C/(2\pi\sqrt{\frac{L1A1L2}{A2}}), \quad C/2L2, C/L2 \dots \dots$$

$$L1= 2 \text{ cm}, L2= 15 \text{ cm}, A1= 0.3 \text{ cm}^2, A2= 5 \text{ cm}^2$$

이 값에서 계산하면 $F1= 242 \text{ Hz}$, $F2= 1133 \text{ Hz}$ ($C/2L2$), $F3= 2266 \text{ Hz}$ ($C/L2$) 등이 나온다. 지금까지 추정된 모음의 음향적 특성은 실제 발화할 때 사람의 구강벽면의 소리의 반향이 완전하지 않아 약간의 에너지가 빠져 나가고, 원순음 등의 경우에는 관의 길이가 더 길어지므로 반드시 정확한 것이라고 말할 수는 없다.

2. 음향적 특성과 조음적 특성과의 관계 규칙

이러한 음향적 특성을 조음적 특성과 연관시켜 규칙화하면 다음과 같다. 중모음 [어]에서 보았듯이, 성문에서 입술까지 성도의 전체 길이와 형성음 주파수 관계는 다음 규칙에 따른다.

길이 규칙. 모음 형성음의 평균 주파수는 성도 길이에 반비례한다. 다시 말해, 성도가 길수록, 평균 형성음 주파수도 더 낮다. 유아의 성도 길이는 성인 남자의 약 반이 된다. 따라서, 유아가 중모음을 발음할 때 그 모음의 형성음은 남성의 두배수가 된다. 즉, F1은 $500 \times 2 = 1000\text{Hz}$, F2는 $1500 \times 2 = 3000\text{Hz}$, F3은 $2500 \times 2 = 5000\text{Hz}$ 등이 된다. 또한, 여성의 성도는 남성의 성도 보다 약 15% 더 짧다. 그 비율은 약 5/6 이다. 5/6의 역수는 6/5이므로 이것은 '1.2배 인수가 되어, 남성의 중모음의 형성음에 곱했을 때, 여성의 값은 20% 더 높아진다: F1은 $500\text{Hz} \times 1.2 = 600\text{Hz}$, F2는 $1500 \times 1.2 = 1800\text{Hz}$, F3 은 $2500 \times 1.2 = 3000\text{Hz}$ 등이 된다.

좁힘과 F1 관계에는 구강 좁힘에 대한 것과 인강 좁힘에 대한 두 가지 규칙이 있다.

구강 F1 좁힘 규칙: F1의 주파수는 성도의 구강 전반부가 좁혀질수록 낮아진다. 따라서, 균일관 형태에서 혀가 위로 혹처럼 튀어나와서 구개 중간 부분에 가까울 때 또는 구개 앞부위에 가까울 때는 F1은 중간 500 Hz 보다는 주파수가 더 낮아지는 경향이 있다. 이것은 또한 입술이나 이에 형성된 좁힘에도 적용된다.

인강 좁힘/F1 규칙 :F1은 인강이 좁혀질수록 높아진다.

따라서, 후설 모음 [o]에서 [a]로 갈수록 인강이 점점 더 좁혀지므로 이 모음들의 F1이 500Hz보다 점점 더 높아지게 된다.

다음으로는 F2 주파수와 성도 좁힘간의 관계를 보이는 규칙을 살펴본다. F2의 주파수는 혀의 좁힘이 전설모음에 대해서는 구강 앞부분에 있고, 후설 모음에 대해서는 혀의 뒷부분에 있다. 먼저 뒷부분을 좁힐 때는

후설 좁힘/F2 규칙: F2 주파수는 후설부분을 좁힐수록 낮아진다.

예를 들어, 혀를 연구개를 향해 혹처럼 만들어 구강도를 좁혀서 [u]처럼 발음하면 F2 주파수가 낮아진다. 균일관이었을 때의 F2의 주파수 1500 Hz는 모음 [u]에서 약 800Hz로 낮아진다. 후설 모음 계열에서 다음 모음인 [o]는 [u] 보다 혀를 덜 좁히고 입술도 덜 둥글게 한다. 따라서, F2는 좁힘의 정도에 비례하므로 [u] 보다는 [o]에 대해서 보다 높은 주파수를 갖는다. 보통 남성의 [o]의 F2는 약 900 Hz이다. 이와는 반대로 전설부위를 좁힐 때 F2는

전설 좁힘/F2규칙: F2의 주파수는 전설을 좁힐수록 높아진다.

전설모음은 혀의 좁힘 정도가 가장 적은 [ae]에서 가장 큰 [i]로 갈수록 높아진다. F2의 주파수는 중모음 F2위치에서 약간 더 높은 1700 Hz 에서 가장 좁힌 전설 모음

[i]의 고주파에 까지 올라간다. 여기서 F2는 보통 남성에서는 약 2200 Hz의 주파수에 해당한다.

마지막으로 입술을 둥글게 할 때는

원순화 규칙: 모든 형성음 주파수는 입술을 둥글게 하여 좁힐수록 낮아진다.

위에서 제시한 규칙들은 성도 모양으로 부터 형성음을 예측하거나 형성음의 수치를 해석하는데 대체로 사용할 수있다. 그러나, 좁힘점이 하나일 때는 잘 적용되나 여러 개가 될 때는 적용될 수도 있고 안될 수도 있다. 예를 들면, 원순화의 정도 차이는 전설 모음에 적용되었을 때는 강한 효과를 보이지만, 후설 모음에 첨가했을 때는 별로 큰 변화를 볼 수없다.

3. 영어모음의 형성음

지금까지 이론적으로 살펴본 모음의 음향적 특징을 실제 수집한 자료를 통해 살펴보자. 그림9는 미국영어의 여덟 개 모음을 성인 남성이 발음했을 때, 그 모음들이 갖는 세 형성음의 평균치를 나타낸 것이다.

그림9.

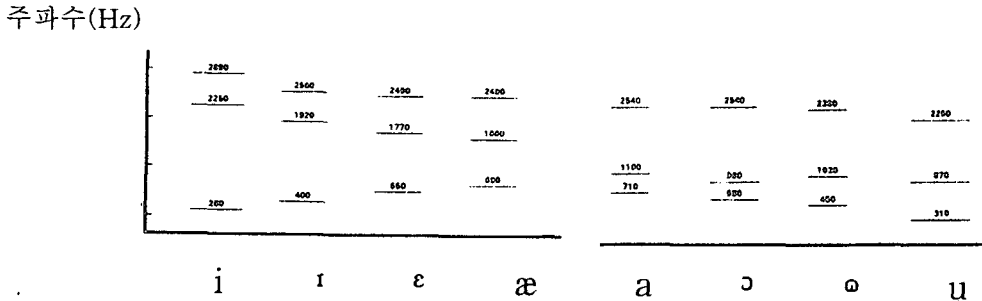
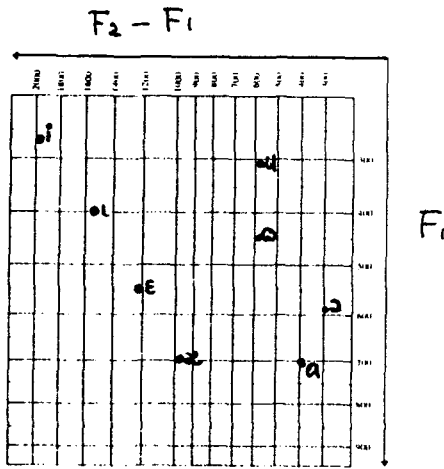


그림9에서, 전설모음은 비교적 높은 F2값을 갖고 후설모음은 그보다 낮은 F2값을 가져 모음의 전설성은 F2와 관련이 있음을 알 수 있다. 한편 고모음의 F1값은 매우 낮고 저모음의 F1값은 상당히 높아서 모음 높이는 F1에 관련됨을 보여준다. 이와 같은 두 형성음의 관계는 전통적인 모음의 기술과 비슷하다. 실제로 세로축에 F1값을 가로축에 F2에 F1을 뺀 값을 표시하면 전통적인 모음 사각도와 매우 비슷한 그림 10과 같은 모음 도표를 얻을 수 있다.

그림10.

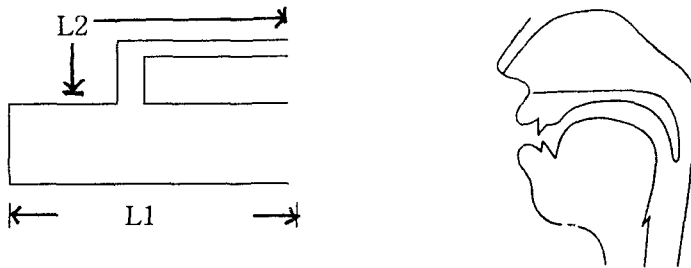


후설모음의 경우 F2값으로 모음의 구분이 분명치 않으나, 그림10에서 보인 것처럼, 수평축을 F1과 F2값의 차이로 대체하므로써 보다 더 전설모음의 구분이 명확하게 된다. 즉, 전설모음에서는 두 형성음의 간격이 크고 후설모음에서는 작게 나타나, 전설성이나 후설성이라는 청각적 개념에 더 접근하게 되었다. 이렇게 조음에 기초한 모음의 전통적 기술이 여러가지 첨단 기구를 사용하여 얻은 음향적 분석 자료와 대체로 상응한다. 따라서, 형성음을 이용한 모음도표가 많이 사용되는데, 그 이유는 모음의 높이는 혀의 높이보다 F1값에 의해 더 정확히 결정되고, 전설후설의 차원도 F1과 F2값의 차이로 혀의 실제 위치를 간단히 표현할 수 있기 때문이다. Lindblom과 Sundberg(1971)도 F1은 턱의 변화 F2는 혀의 변화 F3은 원순성과 연관됨을 밝힌 바있다.

4. 모음의 비음화 및 이중모음

모음의 비음화는 목젓을 내려서 비강으로의 통로를 열어준다. 결국 입과 코에서 소리가 방출되는 그림11과 같은 효과를 낸다.

그림11.

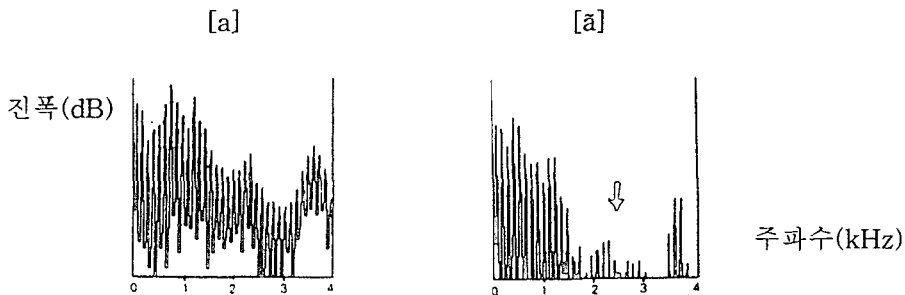


음향학적으로 보면 결과지로 추가의 공명이 생겨서 주파수역에 추가로 정점이 생

성된다. 구강모음에서는 형성음의 평균 간격이 $C/2L_1$ 마다 있다. 그래서 3000 Hz 사이에는 성도길이(L_1) 17 cm이므로 보통 1000 Hz마다 생겨서 모두 세개가 있다. 그런데 비모음에서는 형성음의 평균간격은 $C/2(L_1+L_2)$ 마다 있게된다.

$L_1 = 17$ cm, $L_2 = 12$ cm 이므로 그 평균 간격은 약 600 Hz 마다 생겨서 3000Hz 안에는 약 4 - 5 개의 공명대가 발생한다. 이 추가 공명은 주파수역에 영점이나 반공명 또는 함몰의 형태로 나타난다. 비음화된 모음은 여파기 구실을 하는 성도가 연구개를 내려서 비강을 열고 구강과 비강으로 이중화되어 작용하므로 일반 모음에서 발견되는 비교적 단순한 형성음이 나타나지 않는다. 이러한 이중 공명 체계는 낮은 공명 주파수들의 에너지는 증가시키고 나머지 공명 주파수들의 에너지는 감소시키는 반공명 (Antiresonance)작용을 한다. 그림12는 비음화 모음의 성도의 모양과 모음 [a]의 비음화되기 전,후의 스펙트럼을 비교한 것이다. 성도의 모양은 연구개가 내려와 성대를 통과한 공기가 구강뿐 아니라 비강으로도 지날 수 있음을 보여 준다. 이렇게 비음화된 모음의 스펙트럼은 반공명 작용에 의해 약 1500 Hz 이상의 주파수에서는 급격한 에너지 감소가 있음을 보여 준다.

그림 12.



특히 비음화의 효과는 제1형성음에 나타난다. 예를 들어 F1과 F2가 멀리 떨어져 있는 전설모음에서는 1500 Hz까지 두개의 주파수 정점이 나타난다. 그 반대로 F1과 F2가 서로 근접해 있는 후설모음에서는 1500 Hz안에 세 개의 정점이 나타난다.

모음 중 일정한 조음에 의해 음색이 변하지 않는 것을 단모음 (monophthong) 이라 하고 모음이 두 개 또는 세 개 결합한 음을 각각 이중모음 (diphthong), 삼중모음 (triphthong)이라 하며 이것들을 복모음(compound vowel) 이라 부른다. 그러나 이중모음은 이렇게 간단히 정의되기 보다는 실제로 상당히 논란이 많으며 측정과 분석에도 문제점이 많은 연구 대상 가운데 하나다. 예를 들면, 이중모음은 모음의 시작부터 끝까지 계속적으로 변하기 때문에 측정에 상당한 어려움이 있어서, 조직적인 측정방법을 통해 연구해놓은 자료가 별로 없다. 이중모음에 흔히 나타나는 전이음 [w,j]는 모음

[u, i]와 비슷한 성도 모형을 갖고 있는데 좁힘 모양은 모음보다도 더 좁게한다. 그 좁힘은 소음을 일으킬 정도로 되지는 않는다. 모음보다 더 좁기때문에 F1은 해당 모음보다는 더 낮고 진폭도 더 낮다. 전이음은 항상 다른 모음과 결합되어 나타난다. 연결된 모음의 위치를 형성음이 부드럽게 이동한다. 말소리지 제25-6호의 한국어 이중 모음의 음향학적 연구를 참고하기 바란다.

5. 화자변이 및 모음의 정규화

실제 생활에서 음성은 화자내에서나 화자 사이에서 상당히 다양하게 나타난다. 예를 들면, 화자는 결코 물리적으로 똑같은 발음을 두 번이상하지 않는다 (화자내 변이). 또한 서로 다른 두 화자도 조음적으로나 음향적으로나 똑같이 발음하지는 않는다 (화자간 변이). 이런 화자 변이는 주로, 방언적, 사회 언어학적 차이와 같은 언어적 요소와, 신체구조, 나이, 성별, 발화자의 정서 상태 등과 같은 비언어적 요소로 구분된다.

앞에서 보였듯이, 형성음 주파수는 화자의 성대 전체길이에 반비례하며, 나이와 성별에 따라 성대 길이가 달라서 형성음 값도 달라진다. 또한 인강과 구강의 길이 비율도 화자간 변이의 요소가 된다. Chiba와 Kajiyama (1941)는 여덟살 소녀의 구강길이는 성인남자보다 30% 더 짧으며 소녀의 인강길이는 성인남자보다 56% 더 짧다고 했다. 따라서 모음을 음향적으로 분석할 때는 이러한 비언어학적인 요소의 개입을 최대한 줄여서 순수 언어학적인 자료를 얻도록 노력해야한다. 이러한 다른 화자들이 발성한 모음의 규칙적인 음향 차이를 통계적으로 극복하기위한 노력의 일환으로 모음의 정규화가 있다.

현재까지 남녀 사이와 같은 성별과 나이 또는 감정적인 변화에 따라 변하는 모음형성음을 정규화시켜 보려는 여러가지 수학적인 형성음의 정규화 알고리즘이 제안되어 왔다. 지금까지의 연구방향을 조음적인 방식과 청각적인 단위 변화 방식으로 대별할 수 있다. 조음적인 방식 중에는 Lindblom의 균일축적방식으로서, 이는 주로 성도의 길이 차이에 의해 이런 형성음의 변화가 생기게된다고 가정하고 이 차이를 보상해 주는 방식이다. 예를 들어, 남자의 성도 길이는 약 17 cm에 해당하고 여자의 성도 길이는 약 16 cm에 해당하므로 이를 남자를 기준으로 하여 여자의 성도를 1 cm 늘리거나 여자의 성도를 기준으로 삼는 방식을 말한다. 이에 반해 Fant는 [i]모음 등에서 보이는 성도에서도 구강부위와 인강부위의 비율에 있어서 남녀 차이가 현저하다는 것을 인정하여 모음마다 다른 축척인수를 설정하여야한다는 비균일 축적방식을 제안하였다. 한편 물리적인 단위와 청각적인 단위와의 차이에 초점을 둔 학자들은 남녀 차이 등은 오히려 물리적인 단위를 청각적인 단위로 바꾸면 극복될 수 있다는 주장을 하였다. 구체적으로 보면, 인간의 지각체계는 1 KHz 이상에서는 급속도로 변별력이 감소하며, 형성음간

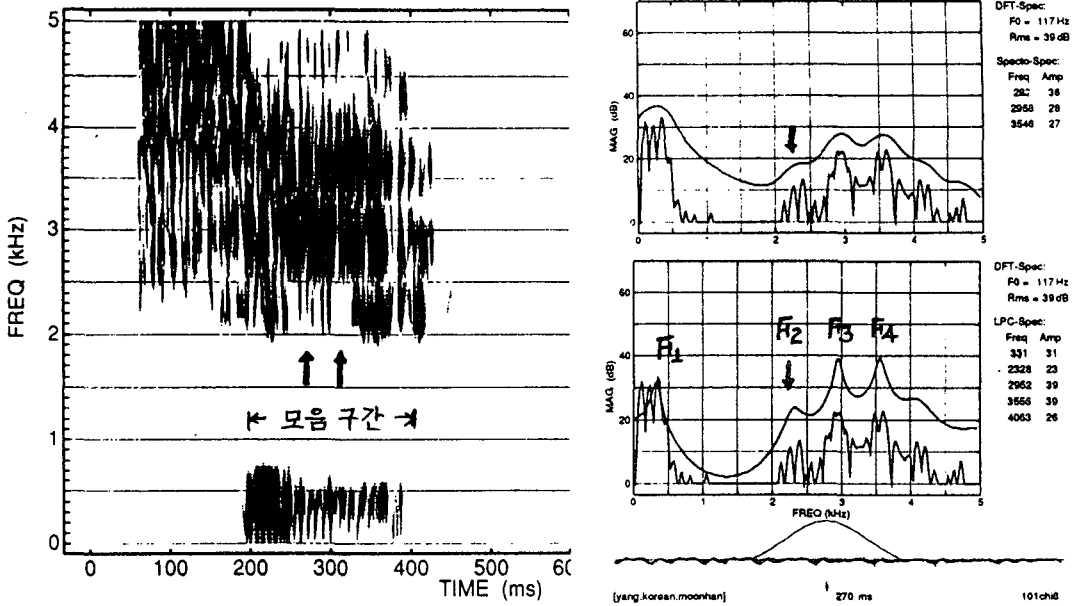
의 구분 능력이 나쁘다. 따라서 음향적 자료의 고주파 부분의 간격을 줄여주는 청각단위인 멜이나 바크로 바꾸어서 정규화시키는 방안이 제안되어 있다. 실제, 한국어 및 영어에 대한 정규화를 적용하여 나온 결과를 보면, 음향학적 수치를 Log나 Bark단위로 바꾸었을 때 정규화가 개선되는 것을 볼 수 있었다. 그 평가에 대해서는 Disner의 논문에 잘 나타나있다. Disner는 언어간 또는 언어내의 모음자질 관계를 어기지 않는 정규화가 중요하다는 지적을 했다. 다시 말해서, 아무리 좋은 정규화라도 본래의 청각적인 관계를 어기는 자료 변환은 옳지 않다고 지적했다.

다음으로는 화자 사이의 변이의 근본적인 원인을 규명하기 위하여 제안된 적응분산이론(Theory of Adaptive Dispersion)과 비약이론(Quantal Theory)에 대해 알아본다. 적응 분산이론은 화자가 발성할 때 전달하는 환경에 따라 적응하여 음성을 발화하는 점을 중시하였다. 특히, 화자는 음성의 음향적 불변 요소 조절뿐만 아니라 충분한 지각적인 대조에 초점을 둔다는 것이다. 예를 들면, 시끄러운 환경에서는 음성을 높이고 더 정확히 발성하려고 애쓰며 (hyper-speech mode), 조용하고 대화하기 좋은 환경에서는 완화된 (hypo-speech mode)로 발성하게 된다. Nooteboom도 단어의 발음은 화자가 청자의 지각이나 이해 과정을 내부적으로 의식하며 말한다고 지적했다. 다시 말해서, 화자도 계속해서 지각 과정과 연관시켜서 발성을 조절한다는 것이다. 한편, Stevens는 그의 비약이론에서 조음적 변수와 음향적 변수의 관계에 대한 견해를 제시하였다. 즉, 조음 활동에는 그 자연스러운 경계내에서는 성도의 음향적 산출에 거의 변화가 없는 일반적인 상태나 영역이 있으며, 반면에 그 경계에서 조금만 벗어나도 엄청난 음향적 변화가 생겨난다는 것이다. 이는 성도의 상태와 스펙트럼 특성 사이에는 일대일 대응 관계가 성립하지 않으며, 더 나아가서 조음기관의 변화와 음향 신호상의 변화에는 비선형적인 관계를 가진다는 것이다. 따라서, 주어진 음향 자료에서 성도 모양을 예측할 수는 있지만, 그 예측이 유일한 해답이 아닌 경우도 많다. 따라서, 조음적 특성과 음향적 특성의 관계 기술은 반드시 일치하지는 않는다는 점에 유의해야한다.

6. 모음의 음향적 분석

지금까지의 모든 내용을 염두에 두고 모음의 음향적 특징을 분석할 때 유의하여야 할 사항들을 생각해보자. 그림13은 30대 한국인 화자가 발성한 [치]음의 실제 스펙트럼 그래프이다.

그림13.



여기서 240ms 지점의 스펙트럼을 살펴보면 평균포락선에 의한 추정값을 보이는 그림 12.A에서는 F1은 283 Hz, F2=2246 Hz, F3=2994 Hz로 추정되었다. 그런데 270ms 지점에서는 두번째 부분이 흐려져서 사라지고 포착이 되지 않았다. 그러나 다행히보다 정밀한 프로그램인 그림12.B의 LPC-Spec 예측치에서는 이것이 잘 포착되었다. 이때 이러한 기계적인 문제점을 잘 인식하고 형성음의 정확한 위치를 예상하고 그 범위에서 찾아나 가야 한다.

또한 그림11의 예에서와 같이 비음화로 인해 모음 [아]의 제1형성음 근처에 생긴 추가의 에너지 띠나, 제3형성음 지역의 반공명으로 인한 형성음 소실 등, 이들 값을 해석할 때 매우 주의해야 한다. 실제 컴퓨터로 음성을 분석해보면 애매한 부분이 상당히 나타난다. 특히 앞에서 언급했듯이, 화자의 신체구조, 나이, 성별, 발화자의 정서 상태 등과 같은 비언어학적인 요소의 개입을 최대한 줄여서 순수 언어학적인 자료를 얻도록 노력해야 한다. 이 때는 정확한 발음에 의하여 발생된 음성으로 이미 분석한 표준자료등을 이용하여 판단을 하는 것이 현명하다. 예를 들어, 영어나 한국어의 단모음에 관한 성별 나이별로 수집한 자료를 참고하여 예상되는 형성음 주파수를 염두에 두고 수집을 하는 것이 좋으며, 만약 컴퓨터 자동 분석 프로그램을 개발할 때도 마찬가지로 그 범위에 대한 자료를 입력하여 예상 밖의 자료 제시가 되지 않도록 유의해야 한다. 특히 남녀 변수를 고려하지 않고 한꺼번에 처리하여 평균을 낸다면 그 결과치는 어느 그룹도 대표하지 못하는 별로 의미없는 결과가 될 것이다. 또한 주파수 및 지속시간의 측정에서도 그 값이 인간의 음성 지각 체계와 관련시켜 제시하고 또한 해석해야 한다. 예를 들어, Hz 단위나 ms(천분의 일초)를 소숫점 이하 여러 자리까지 처리하는 것을 보는데 이는 사람이 변별할 수 있는 주파수와 시간 단위는 제한되어 있다는 지각특성을 고려해 볼 때 별로 의미 없는 수치가 될 것이다.