6

음성의 음향적 검사
(Sound Spectrographic Analysis)

국립과학수사연구소

홍 수 기

서 론

1. 신호처리와 음성신호 분석

신호처리의 목적은 신호를 변형하여 우리가 원하는 형태로 만드는 것으로 신호를 변환시키는 장치 즉 시스템이 신호에 응답하여 다른 형태의 신호를 만들어 내는 것을 신호처리라 한다.

현재는 음성신호 처리시에 대부분 입력신호인 아날로그 신호(Analog Signal)를 표본화(Sampling)하고 양자화(Quantizing)하여 디지탈 신호(Digital Signal)로 변환한 후 필요한 신호처리를 수행한다. 디지탈 신호를 처리하므로써 정확성, 신뢰성, 처리속도를 증가시키게 되고 전자시스템(Electronic System)의 크기를 줄일 수가 있다.

시간함수를 주파수함수로 변환하기 위한 방법으로는 아날로그 신호인 경우에는 Fourier 변환과 Laplace 변환이 있고, 디지탈 신호인 경우에는 Fourier 변환과 Z 변환이 있다. 주기적인 유한기간의 이산 시간함수를 푸리에 변환을 이용하여 주파수 영역의 이산함수로 변환하여 처리 가능하도록 하는 방법이 이산 푸리에 변환(Discrete Fourier Transform ; DFT)이고 음성신호 분석에 많이 이용되고 있는 스펙트로그램(Spectrogram)이나 FFT 스펙트럼(Spectrum) 등은 DFT에 의한 주파수 분석 결과이다. FFT(Fast Fourier Transform)는 유한기간 함수의 DFT를 계산하는데 필요한 많고 복잡한 연산과정을 줄일 수 있는 알고리즘(Algorithm)이다. Z 변환은 특정한 성질의 결과를 얻기 위해서 이산신호를 처리하는 장치나 알고리즘인 디지탈 필터(Digital Filter)에 이용되며 선형예측 부호화(Linear Predictive Coding ; LPC) 방법은 이러한 특별한 디지탈 필터를 이용하는 방법이다.

음성신호처리가 이용되는 분야로는 음성신호 분석분야, 음성의 합성분야, 음성의 합성에 의하여 얻어진 음성신호의 분석분야로 구분할 수 있다. 음성신호 분석은 화자식별(Speaker Identification), 화자확인(Speaker Verification), 자동 음성 인식 시스템(Automatic Speech recognition System), 및 음성병리학과 언어음성학에서의 음향검사 등에 이용되고 있으며, 음성합성분야는 문자를 인식하여 음성을 출력으로 하는 자동 읽기 시스템과 데이타 복구 시스템에 이용되고, 음성의 합성에 의한 음성신호 분석은 비화전송(Secure Voice Transmission)과 데이타 압축에 의해 훼손된 음성의 강조와 확장 및 압축에 응

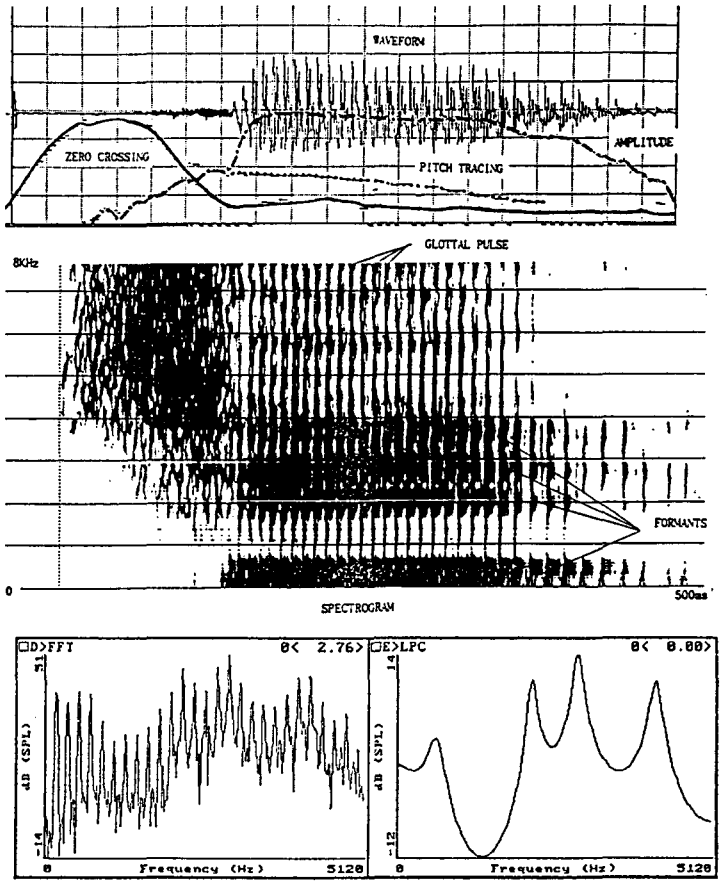

그림 1.

용된다.

음선신호 분석에서는 그림 1에서와 같이 음성신호의 스펙트로그램, FFT 스펙트럼, 시간파형, 진폭 포락선, 영교차 파형, 켑스트럼, LPC에 의한 주파수 분석 등을 이용하여 실험 목적에 맞는 여러가지 음향학적인 검사를 하게 된다.

시간파형은 입력된 음성신호를 분석한 것이 아니라 입력된 음성신호에 의해 생긴 전압변화를 시간변화에 따라 선형 스케일로 나타낸 것이다. 진폭 포락선은 시간파형을 대수적으로 처리하여 전압변화를 데시벨(dB)로 나타낸 시간파형의 대수적 포락선이다. 영교차 파형(Zero Crossing Display)은 진폭이 일정하고 입력 음성파의 영교차 위치에서 진폭의 부호가 반전하는 펄스 파형으로 "s"나 "sh"와 같은 고주파 소리를 감지하는데 뛰어나다.

FFT 스펙트럼은 입력신호의 주파수별 전력크기(Power Magnitude)를 나타낸 것으로 선택된 데이타 그룹의 FFT 분석에 의하여 얻어진다. 켑스트럼은 FFT 스펙트럼에 나타난 진폭을 대수 변환한 뒤 이를 다시 푸리에 변환한 결과로 여러 형태로 존재하나 모든 것이 대수 스펙트럼으로 이를 이용하므로써 음성의 기본주파수와 스펙트럼의 비주기적인 성분들을 분리하는 것이 가능하여 정밀도 높은 분석이 가능하다. LPC에 의한 주파수

포락선은 시간파형의 선형예측 분석에 의하여 자동적으로 보여지는 형태로 포만트에 관한 정보를 잘 나타낸다.

스펙트로그램은 시간에 따라 변화하는 신호를 분석하기에 가장 유용한 주파수 분석 형태로 시간, 주파수 및 진폭을 3차원적으로 나타냈으며 수평축은 시간, 수직축은 주파수를 나타내고 진폭은 검은색의 농도 변화나 색깔로 구분되어진다. 스펙트로그램은 음성신호를 기억시켜 이것을 반복재생하며 주파수 분석을 하여 얻어지는 음성 스펙트럼으로 분석 필터의 중심 주파수는 고정시키고 분석하려는 신호의 주파수를 이동시키면서 분석하는 heterodyne형 분석방법을 이용한다. 분석 필터 중 광대역(일반적으로 300Hz 이상) 필터는 빨리 공명을 시작하여 정상적인 진폭에 이르기까지 소요되는 불변시간〈(1/300)sec〉이 짧아 신호의 시간변수를 분해하는데 사용되고 협대역(일반적으로 45Hz 이하) 필터는 불변시간〈(1/45)sec〉이 길어서 주파수 변수를 보다 정확히 분해하는데 사용된다

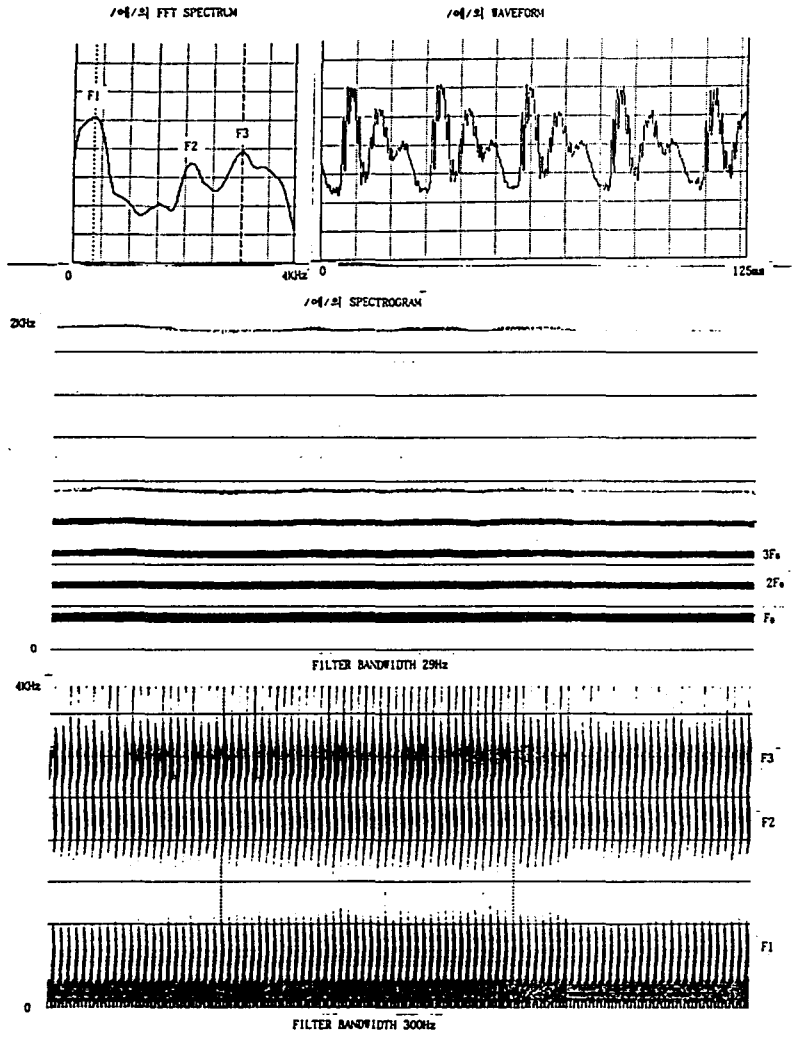

그림 2.

(그림 2). 음성을 스펙트럼으로 분석할 때, 우리 인간의 귀는 주파수에 따라 소리의 크기 변화를 감지하는 능력이 다르며 높은 주파수에서 더 민감하게 감지하므로 실제 인간이 감지하는 음성과 가장 근사한 음성 스펙트럼을 얻고, 높은 주파수에서 에너지가 약화되는 음성의 특징으로 분해능이 저하될 수 있어 대부분 입력신호를 고역강조(Hi-shaping 또는 Pre-emphasis)시켜 분석한다.

2. 모음과 자음의 음향적 특성

음성파의 음원으로는 연속적인 성문(glottis)의 개폐에 의한 주기적인 공기진동과 구강내의 혀, 이, 입술 등에 의해 만들어진 좁은 공간 부위에서 공기류를 불어 넣어서 생긴 잡음성의 공기 진동인 난류로써 구분된다.

진동 가능한 어떤 계가 그 계의 자연 진동수와 거의 같은 진동수를 갖는 주기적인 힘을 받을 때 그 계는 비교적 큰 진폭으로 진동하게 되고 이런 현상을 공명(resonance)이라 한다.

모음과 유성자음은 음원인 성대에 의한 주기적인 공기진동이 공명계인 성도를 공진시킴으로써 발생된 음성파형으로 그림 2에서와 같은 시간에 따라 변화하는 주기적인 복합파형을 가지고 있다. 이 복합파형은 성대진동에 의한 기본 진동음과 그것의 배음들로 구성되어 있으며 배음들의 진동수는 기본 진동의 정수배로 이루어져 있고 각각의 배음들은 다른 진폭을 가지고 있다. 성도의 공진 특성에 의해 몇 개의 배음들은 강조되며 이런 높은 에너지를 가진 배음들을 포만트(Formant)라 하고 이 포만트들에 의해 즉 성도의 공진 특성에 의해 음색이 달라지게 된다. 성도의 공진 특성은 필터처럼 작용하여 그림 3에서와 같이 성도의 크기와 모양에 따라 즉 인강 및 구강의 길이, 성도에서의 좁힘점과 좁힘정도 등에 의해 배음들이 어떤 주파수에서는 강조되고 어떤 주파수에서는 약화된다. 또한 혀와 턱의 움직임에 의한 구강 모양의 변화도 성도의 길이를 변화시킨다.

포만트의 물리적 특성을 이해하고 예상 주파수를 계산하거나 성도의 길이 등을 추정하기 위해 성도를 단순 음향관으로 모델화 할 수 있다.

우선 그림 4와 같은 단일 음향관을 고려해 보면, 관을 따라 전파되어 가고 있는 종파(음파도 종파의 일종임)는 관의 끝에서 반사되어 진행파와 반사파의 간섭에 의하여 정상파를 만들 수 있다. 관의 끝이 닫혀 있으면 반사파와의 위상차가 π가 되어 관의 끝에서는 마디(node)를 이루고 관의 끝이 열려 있으면 그곳에는 배(antinode)를 이루게 되므로 관의 양쪽이 모두 닫혀 있거나 열려있는 경우 가장 낮은 공명주파수(F1)는 C/2L로 여기서 C는 공기 중에서의 소리의 속도이고 L은 음향관의 길이이다. 다음의 공명주파수들(F2, F3, F4, …)은 이 기본 공명주파수의 정수배 즉 nC/2L, n=1, 2, 3, 4, …이고, 또한 한쪽만이 열려 있는 경우 가장 낮은 공명주파수는 C/4L이고 다음의 공명주파수들은 이 주파수의 홀수배 즉 (2n-1)C/4L, n=1, 2, 3, …이다.

모음 /ə/의 성도형태를 한쪽만이 열려있는 단일 음향관으로 가정하면 보통 남자 성인의 성도길이는 약 17cm이고, 공기 중에서 소리의 속도는 상온에서 약 340m/sec 이므로 F1=500Hz, F2=1500Hz, F3=2500Hz, …로 주어진다.

성도를 두 개의 음향관(모델I)의 합으로 가정하면 그림 5와 같이 두 가지 경우로 고려할 수 있다.

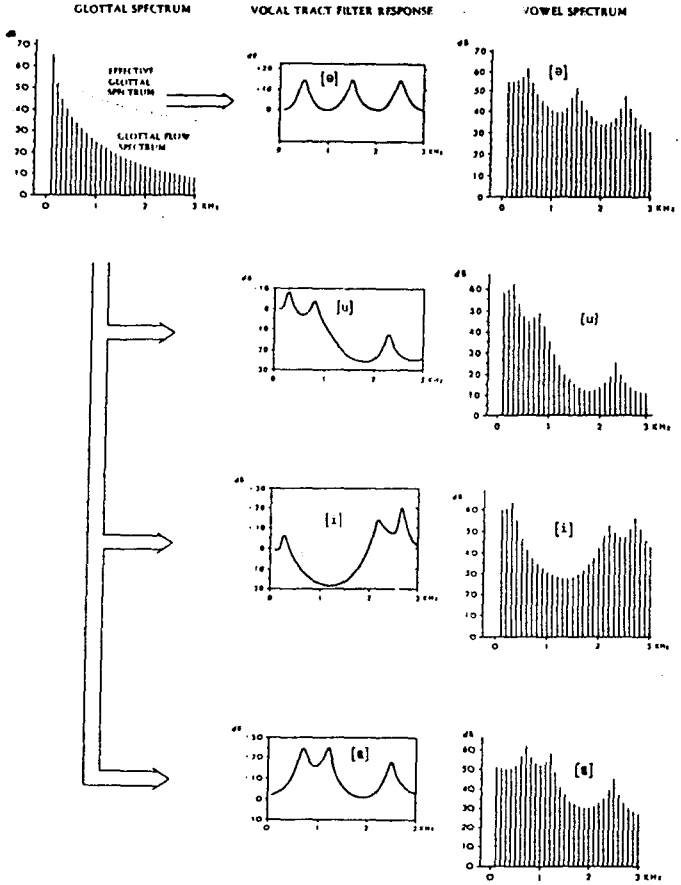
GLOTTAL SPECTRUM
VOCAL TRACT FILTER RESPONSE
VOWEL SPECTRUM
[ə]
[u]
[i]
[ɑ]

그림 3.

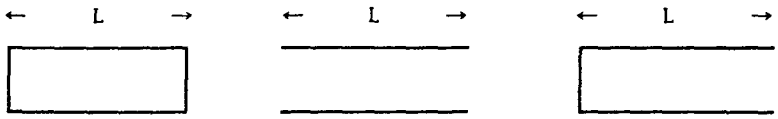
L
L
L

그림 4.

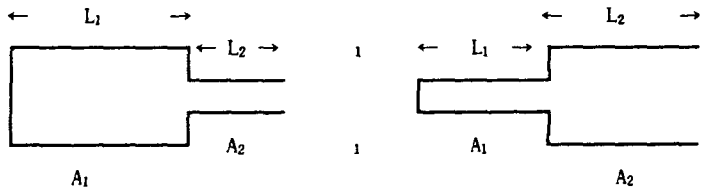
L1
L2
A2
A1
L2
L1
A1
A2

그림 5.

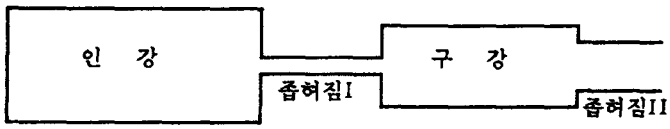

그림 6.

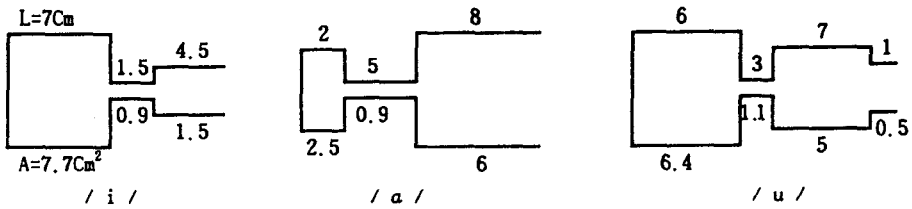

그림 7.

첫번째는 모음 /i/와 /u/의 성도의 형태로 가정될 수 있으며 이 음향관의 가장 낮은 공명주파수는 Helmholze 공명기의 공명 조건에 의해 $F1=C/2\pi(L_1A_1L_2/A_2)^{1/2}$로 주어지고, 여기서 L_1 및 L_2는 관의 길이이고 A_1 및 A_2는 관의 단면적이다. F2이상의 공명주파수들은 기본 공명주파수의 배음에 의해 주어지는 것이 아니라 양쪽이 모두 닫혀 있는 관과 모두 열려 있는 두 개의 균일관에 의한 각각의 공명주파수인 $nC/2L_1$과 $nC/2L_2$에 의해 결정되어진다. 이와 유사하게 두번째는 모음 /a/의 성도 형태로 가정될 수 있으며 한쪽만이 열려 있는 두 개의 음향관으로 고려되어 각각의 공명주파수 $(2n-1)C/4L_1$과 $(2n-1)C/4L_2$에 의해 결정되어진다.

또한 좀 더 복잡한 음향관(모델II)으로써 성도를 인강, 혀에 의해서 좁혀진 부분(좁혀짐 I), 구강 및 입에 의해서 좁혀진 부분(좁혀짐 II)으로 구분하여 그림 6와 같이 모델링 할 수 있다.

이러한 음향관들의 공명주파수들은 공명조건으로부터 매우 복잡한 계산식에 의해 결정되어지는데 모음 /i/, /a/, /u/의 성도형태를 그림 7과 같이 가정할 수 있다.

위에서 추정된 두 가지의 음향관들(모델I과 모델II)에 의한 모음 /i/, /a/, /u/의 공명주파수들[1)2)]과 실제 미국인[3)]과 한국인[4)]에 의해 발음된 것으로부터 측정된 포만트들을 비교하면 다음과 같다.

모음	/i/				/a/				/u/			
포만트	모델I	모델II	영어	한국어	모델I	모델II	영어	한국어	모델I	모델II	영어	한국어
F1(Hz)	202	325	270	270	903	750	730	760	242	310	300	320
F2(Hz)	1890	2300	2290	2190	1020	1197	870	1200	1133	850	870	700
F3(Hz)	2125	2770	3010	3020	2830	2920	2440	2570	2266	2720	2240	2430

성도를 두 부분으로만 구분한 모델I 보다는 좀 더 세분화하여 구분한 모델II에 의한 계산치가 실제의 모음 발음으로부터 구한 실험치와 더 잘 일치함을 보여준다.

또한 3개의 공명주파수 중에서 제 3 공명주파수에서 가장 많은 차이가 생기는데 Dunn (1950)에 의하면 모델II에서 모음 /u/인 경우 구강 부분을 1cm 길게하고 인강 부분을 같은 길이만큼 짧게 했을 경우 F1, F2, F3는 각각 298, 860, 2450Hz로 F1과 F2는 약간 변화하나 F3는 많이 변화한다. 대부분의 유성음들은 수 개의 포만트를 가지고 있고 그

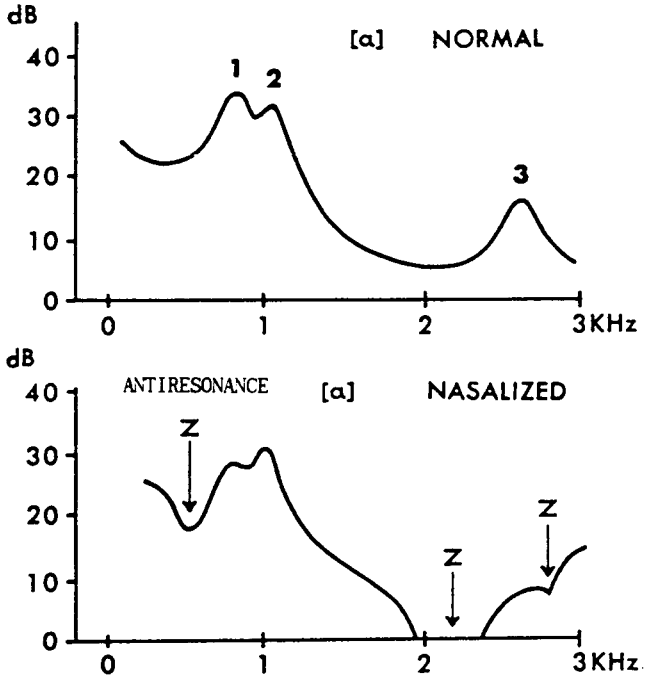

그림 8.

중에서 F1과 F2는 혀의 위치와 밀접한 관계가 있어 언어학적으로 중요한 의미를 갖는다. 혀의 높이가 낮은 모음일수록 F1이 높아지는 경향이 있으며 혀의 위치가 앞쪽일수록 F2가 높아지는 경향이 있으나 일반적으로 이 상관관계는 F1과 혀의 높이만큼 중요한 의미를 갖지는 못한다. 또한 F3 이상의 포만트들은 성도를 구성하는 각 부분들의 미세한 변화에도 크게 영향을 받아 음성의 개인적인 특성을 구별하는데 중요한 의미를 가지고 있다.

모음이 비음화 되어지면 그림 8에서와 같이 F1의 진폭이 감소되고 성도에 지분(상인두, 비강)이 가해지기 때문에 반공진(antiresonance)이 생기게 되며 인강과 구강의 전달과정에서 추가 포만트가 첨가되어진다.

음향음성학적인 연구를 위해서 자음들은 조음 특성에 따라 분류되어진다. 자음의 조음 특성에는 조음 방법에 따른 특성, 조음 장소에 따른 특성, 유/무성에 따른 특성 등이 있다.

자음의 특성에 따라 발음지속시간이 달라지고 스펙트럼에서 잡음성 음의 세기 및 스펙트럼 형태가 달라지게 된다(그림 9). 또한 자음의 특성에 따라 근접해 있는 모음의 포만트나 발음지속시간 등이 영향을 받기도 한다.

3. 음성의 음향적 특성 측정

성도의 공진특성에 의해 에너지가 증가된 배음들로 형성된 포만트는 배음들의 군으로 구성되어 있기 때문에 넓은 주파수 범위를 가지고 있어 일반적으로 중심 주파수를 포만트 주파수라하며 광대역 대역폭(일반 남성인 경우 300Hz) 필터를 이용한 스펙트로그램, FFT 스펙트럼, LPC 주파수 포락선 등을 이용하여 측정할 수 있다. 여성이나 어린 아이들의 음성처럼 피치(pitch)가 높은 경우는 남성에서 보다 더 넓은 대역폭 필터를 사용해야만

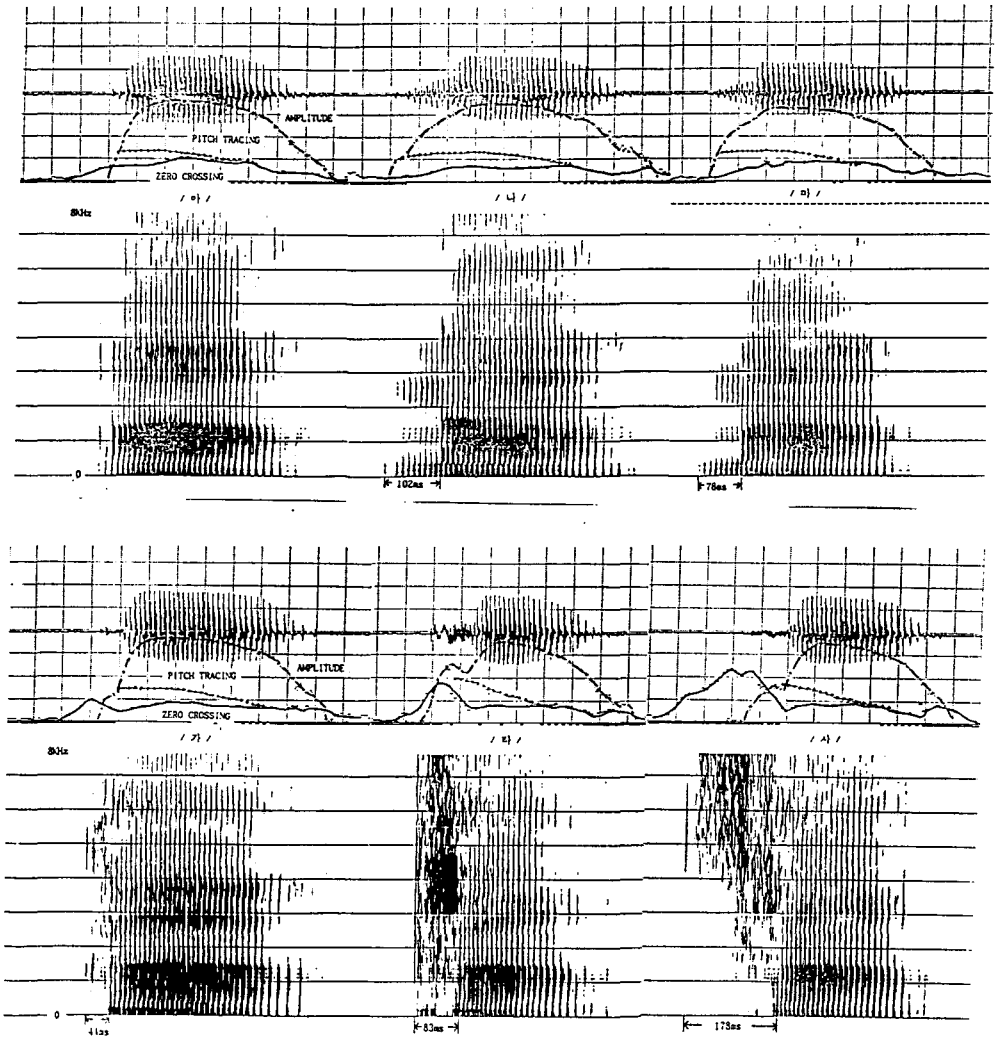

그림 9.

정확히 포만트 주파수를 측정할 수 있다. 포만트의 폭(bandwidth)은 포만트 포락선의 정점에서 약 3dB 낮은 곳 즉 세기가 약 반으로 되는 곳에서의 주파수 폭으로 결정된다.

기본주파수는 원리적으로 음성파형 중에서 가장 낮은 반복 주파수로 시간파형으로부터 직접 추출할 수도 있고 스펙트로그램에서는 광대역 필터를 사용한 경우는 스펙트로그램상에 나타난 성문 펄스인 세로선들의 수와 시간간격을 측정하여 구할 수 있으며 협대역 필터를 이용한 경우에는 스펙트로그램상에 나타난 배음들 사이의 주파수 간격을 직접 측정하므로써 가장 용이하고 정확하게 측정할 수 있고 기본주파수의 변화곡선으로부터 억양의 형태를 직접 볼 수 있다. 또한 성문파형 주기의 빠른 변화로 vocal roughness와 관련되어지는 음향적 특성인 vocal jitter(또는 pitch perturbation)와 기본주파수 변화 속도를 측정할 수 있고 특히 고주파수 쪽에서 측정하면 관측하기가 더 용이하다. 또한 켑스트럼 분석을 기초로한 계산 방법을 이용하거나 선형예측법에 의하여 성도의 음향전달 특성을 역필터 처리하여 얻어진 음원파형에서도 기본주파수를 분석할 수 있다.

유성음 신호의 빠른 진폭 변화를 나타내는 vocal shimmer는 연속적인 주기들간의 평균

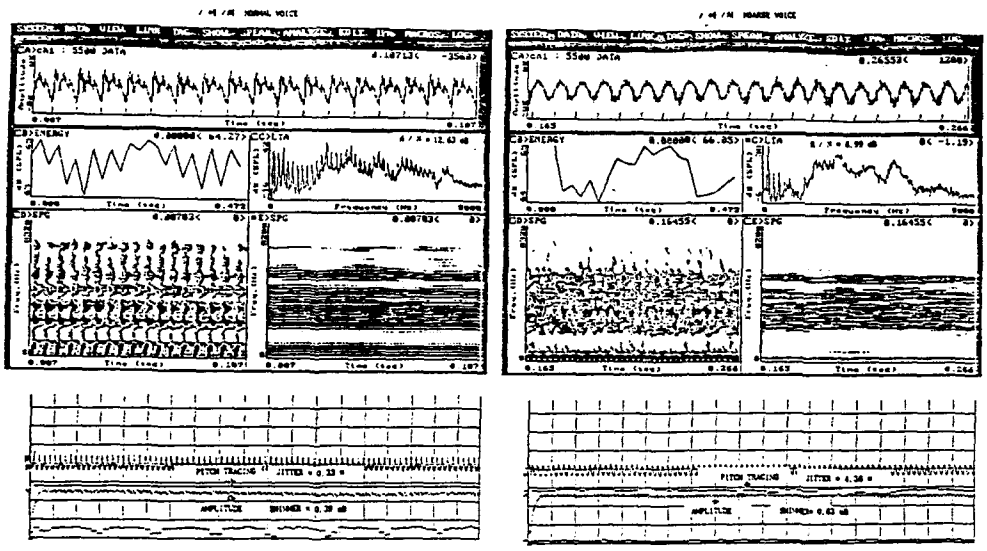

그림 10.

진폭 차이로 정의 되어지며, 이것은 진폭 포락선으로부터 측정할 수 있고 진폭 변화 속도와 발성 시초의 증대 속도와 끝의 감쇄 속도를 구할 수 있다.

스펙트로그램과 시간파형으로부터 모음과 자음의 발음지속시간을 측정할 수 있으며 특히 V.O.T.(voice onset time)는 광대역 스펙트로그램을 이용하면 효과적으로 측정할 수 있다.

신호대 잡음비(H/N)는 협대역 필터를 이용하고 주파수 축을 확대시킨 스펙트럼에서 정확히 측정할 수 있는 양으로 배음들 사이에서의 잡음은 거의 항상 불완전한 성문 폐쇄와 관련되어 있고 종종 jitter와 shimmer와도 관련되어진다. jitter와 shimmer는 sideband를 생기게 하는 원인이 된다. 예로 기본주파수가 120Hz이고 sideband가 약 130Hz라면 10번째 배음의 sideband는 1300Hz로 11번째 배음인 1320Hz에 더 가까이에 있으므로 높은 주파수에서 잡음성인 거친 목소리가 더 잘 감지된다.

hoarseness는 목소리가 여러 측면에서 비정상적이라는 것을 나타내기 위해 사용되는 일반적인 용어이다. hoarseess와 관계가 있는 주된 음향적인 요소들은 잡음성분으로 구성되어 있고 조화성분들은 사라진다(그림 10).

잡음성분의 범위와 에너지는 감지되는 hoarseness의 정도에 따라 변화하고 모음 /u/와 /ɔ/에서 보다 /a/, /e/, 및 /i/에서 더욱 더 명백하다. hoarseness가 적은 경우는 주로 4 KHz이상에 에너지가 분포되어 있다. hoarseness의 정도가 증가함에 따라 에너지는 점점 조화음 범위로 확장되고 기본주파수는 뚜렷한 불규칙성을 나타낸다. hoarseness가 아주 심한 경우 포만트 범위에 속한 배음들의 구조는 잡음성분으로 대치되어 포만트 구조는 사라져가고 성문 펄스는 비주기적으로 변한다. 이러한 hoarseness는 광대역 및 협대역 스펙트로그램으로부터 관측할 수 있으며 신호대 잡음비를 구함으로써 양적인 측정이 가능하다.

References

1) Dunn HK : 4The Calculation of Vowel Resonances and an Electrical Vocal Tract, JASA 22 : 740-

753, 1950
2) 양병곤 : 모음의 음향적 특징, 음성학 학술대회 자료집 *113-124, 1994*
3) Kent RD and Forner LL : *Developmental Study of Vowel Formant Frequencies in an Imitation Test, Readings in Clinical Spectrography of Speech, Singular Publishuing Group. and KAY Elemetrics Corp 392-401, 1978*
4) 박종철 : *Sound Spectrograph*에 의한 우리말 단모음 분석에 관한 연구, 연대 산업 대학원 석사 논문, *1985*