1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 1(s) 호

음성신호의 스펙트럼해석 및 모델링

- FFT와 선형예측분석법에 의한 음성신호분석 -

창원대학교 제어계측공학과
조철우

Spectral Analysis and Modelling of Speech Signal

- Analysis by FFT and LP Analysis -

Dept. of Control and Instrumentation Eng.
Cheol-Woo Jo
cwjo@sarim.changwon.ac.kr

요 약

본 논문에서는 음성신호처리의 기초적인 해석법인 FFT와 LP분석법에 대하여 기본적인 이론과 함께 분석과정에서 알아두어야 할 사항들을 정리한다. 아울러 이러한 분석을 실제 음성신호를 대상으로 행함에 있어서 주의해야할 점들을 실제음성을 처리한 그림과 함께 설명한다.

1. 개 요

음성신호를 분석하는 방법은 매우 다양하다. 지금도 여러 가지 해석기법들이 제안되고 있지만 전통적으로 음성분석에 사용되어온 방법은 음성을 주파수 성분으로 분해하여 보는 푸리에 변환기법 및 이를 응용한 스펙트로그램이다. 푸리에변환 및 스펙트로그램에는 음성신호의 음원, 성도, 피치, 음질, 음색등 거의 모든 정보가 나타나 있기 때문에 지금까지 거의 모든 음성학자, 음성공학자들의 기본적인 해석도구로 사용하여 왔다. 또, 선형예측분석기법은 지난 십여년간 음성분석에 기본적인 도구로 사용되어 왔다. 이 방법은 공학적인 목적의 코딩이나 음성인식, 음성합성등 이외에도 성도의 특성변화를 측정하거나 포만드값을 구할 때등 FFT나 스펙트로그램의 결과를 보완해 줄 수 있는 처리방법으로 많이 이용하고 있다. 그러나 많은 사람들이 이러한 분석법을 기존의 소프트웨어등을 통하여 사용하고 있으면서도 기본적인 원리를 이해하지 못한 관계로 결과의 분석에서 어려움을 겪고, 분석결과로부터 잘못된 결론을 도출할 수 있는 가능성도 큰 것이 사실이다. 이러한 면에서 음향학회의 음성학 관련 전공분야의 여러분들에게 좀 더 자세한 내용을 알릴 수 있는 기회가 있으면 좋겠다는 취지에서 여러 분들께서 제안을 하셨고, 이 특강이 이루어 지게 되었다.

이 글에서는 음성의 중요한 신호처리 기법중에서 FFT스펙트럼분석법과 선형예측에 의한 음성모델링에 관하여 이야기하고자 한다. 이 글을 보는 대상을 공학 및 자연과학을 하시는 분들이 아닌 음성학쪽의 분들을 대상으로 하였기 때문에 구체적인 수식이나 프로그램의 전개보다는 개념 및 사례중심으로 설명을 하고자 한다.

2. 디지털신호의 개요

디지털신호는 아날로그신호로부터의 표본화에 의해 얻어진다. 이 기능은 A/D변환기(Analog-to-Digital Converter)를 통해서 수행되는데 음성신호의 경우 PC의 사운드카드가 아날로그신호를 디지털로 변환하는 역할을 하게된다. FFT나 LPC등은 디지털로 변환된 신호에 대하여 적용되기 때문에 원래의 신호가 디지털화 되면서 발생하는 변화에 대하여 알아둘 필요가 있다.

우선 신호를 단위 시간당 얼마만한 개수를 받아들일 것인가에 관한 표본화율에 대하여 알아본다. 표본화율은 Nyquist 정리에 의해서 결정되는데 이는 '어떤 신호의 최대 주파수 성분이 Fs일 때 표본화에 필요한 최소의 주파수는 2Fs가 된다'는 것이다. 즉, 0~4KHz의 주파수 범위를 갖는 신호의 경우 최대 주파수의 두배인 8KHz로 표본화하면 된다. 이 주파수보다 적은 수로 표본화할 경우에는 에일리어싱 현상에 의해 고주파대역에서 왜곡이 발생하게 된다. (그림1) 이렇게 Nyquist율보다 낮은 표본화주파수로 얻어진 신호를 통한

분석결과는 이론적으로 신뢰할 수 없는 것이 된다. 음성신호의 경우 분석하고자 하는 주파수의 범위에 따라서 적절히 선택하면 되므로 8KHz, 10KHz, 16KHz등 여러 가지의 선택이 가능하다.

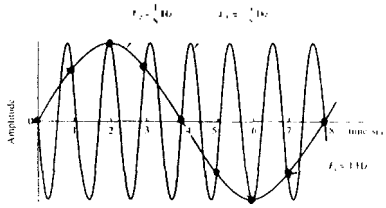

그림1. 신호의 에일리어싱현상

이러한 표본화과정을 주파수 평면에서 보면 원래 아날로그 신호의 스펙트럼성분이 Fs마다 반복적으로 존재하는 형상이 된다. 이 경우 0 ~ Fs/2 까지의 주파수만이 실제로 유효한 주파수가 된다. 즉, 8KHz로 표본화 했을 경우 4KHz까지의 주파수 성분만이 표본화된 신호에 담기있게 된다. (그림2)는 표본화시의 주파수 성분의 변화를 보인 것이다.

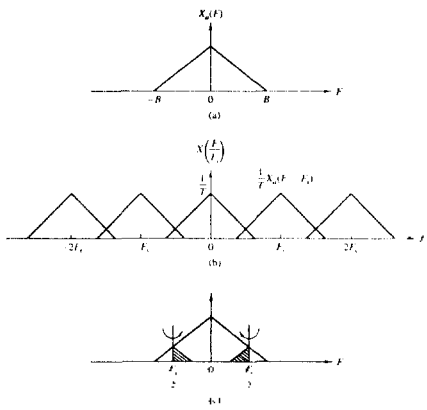

그림2. 표본화시 주파수성분의 변화

3. FFT에 의한 신호분석

FFT를 설명하기 전에 먼저 푸리에변환의 개념에 대하여 설명한다. 푸리에 변환은 프랑스의 수학자인 조셉 푸리에(1768 1830)가 고안한 변환으로 임의의 신호를 정현파 성분으로 분해하여 해석할 수 있는 방법이다.

(1)푸리에 급수의 식(아날로그, 주기함수)

$$c_n = \frac{1}{T}\int_T^0 f(t)\, e^{-jn w_n t}, n=0, \pm 1, \pm 2, \cdots$$

(2)푸리에 변환의 식(아날로그, 비주기함수)

$$X(w) = \int_{-\infty}^{\infty} x(t)\, e^{-jwt} dt$$

(3)DFT의 식(디지탈신호, 주기 및 비주기)

$$X(k) = \sum_{n=0}^{N-1} x(n)\, e^{-j\frac{2\pi nk}{N}}, k=0, 1, \cdots, N-1$$

아날로그 신호의 경우 주기신호에 대하여는 푸리에급수전개에 의해 분석하고, 비 주기신호에 대하여는 푸리에변환에 의해 분석한다. 그러나 디지털 신호의 경우에는 항상 유한한 구간의 신호만을 다룰 수 밖에 없으므로 푸리에 변환은 DFT(Discerete Fourier Transform)이란 형태로 모든 신호에 대하여 적용된다. DFT의 연산속도를 빠르게 하기 위하여 고안된 것이 FFT(Fast Fourier Transform)으로 DFT의 연산과정을 규칙적으로 나누어서 계산속도를 높인 것이다. DFT의 경우는 아날로그 푸리에변환이 무한한 길이의 신호에 대하여 변환을 수행하던 것과는 달리 유한한 개수의 신호에 대하여 변환을 수행하게 되므로 결과에 있어서도 개념적으로 차이가 있다. 쉽게 말하면 푸리에 변환식을 유한한 개수만큼 표본화하여 디지털화한 것이라고도 말할 수 있다. 그러므로 DFT의 결과는 신호의 표본화과정에서와 마찬가지로 Fs를 기준으로 반복적인 형태로 나타나게 된다. 역시 마찬가지로 유효한 주파수 영역은 0 ~ Fs/2까지가 된다. 이 때 Fs는 신호의 표본화 주파수와 같다.

FFT는 DFT의 계산속도를 빨리 하기 위하여 다음과 같이 짝수, 홀수 항으로 연산식을 나누어 분해하여 계산후 합해나가는 방식으로 곱셈연산의 수가 원래 $O(N^2)$에서 $O(NlogN)$으로 줄어들게 된다. (그림3)은 8점의 FFT연산을 위한 분할구조를 보인 것이다. 이것은 radix-2 FFT알고리즘의 구조를 나타낸 것으로 짝수항과 홀수항으로 계속 분할하여 계산함으로써 계산량을 줄이는 방법이다.

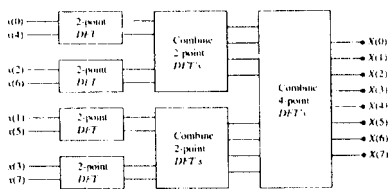

그림3 8점 FFT의 분할구조

이와 같이 짝수 홀수 항으로 나눌 경우 FFT연산에 대한 입력신호표본의 개수는 2^N개가 되어야 한다. 즉, 64, 128, 256, 512, 1024 등이 되어야 한다. 이와 같은 FFT알고리즘을 radix-2알고리즘이라 하며 가장 효과적인 알고리즘이다. 입력신호표본의 개수가 2의 지수승으로 떨어지지 않을 경우도 FFT연산이 가능하지만 계산속도가 떨어지게 된다. 이런 경우 임의의 개수만큼 0을 붙여서 2의 지수승으로 개수를 맞추어 연산할 수 있다.(zero padding)

디지털 신호의 푸리에 변환은 이와 같은 FFT알고리즘에 의해서 이루어 지게 된다.

다음은 널리 사용되는 신호처리 패키지인 MATLAB에서의 FFT함수 사용법에 대한 설명이다.

y=fft(x,N);

여기서 x는 입력신호벡터, N은 신호표본의 갯수, y는 변환된 스펙트럼성분 벡터이다. FFT의 출력의 개수는 입력신호표본의 개수와 같다. 단지 출력은 0 ~ Fs까지의 성분을 포함하고 있고, 결과에서 우리는 0 ~ Fs/2까지만 필요하므로 만약 개수가 256개라면 처음의 128개만 취하여 보면 된다. y벡터는 복소수로 구해지므로 스펙트럼성분의 크기를 보기 위해서는 절대치(abs(y(1:128)))나 데시벨값(20*log(abs(y(1:128))))을 구해야 한다. 위상의 변화를 보기 위해서는 angle(y(1:128))로 구하면 된다. 여기서 fft, abs, log, angle등은 MATLAB에서 기본적으로 제공하는 함수이다.

그런데 이 과정에서 또 한가지 고려해야할 점은 윈도우 함수에 관한 것이다. 앞서 말한 바와 같이 디지털신호의 해석은 전체의 신호를 포함할 수가 없어서 일정한 개수의 신호를 대상으로 하기 때문에 그렇게 자르는 과정에서 주파수 성분에 대한 왜곡이 발생한다.

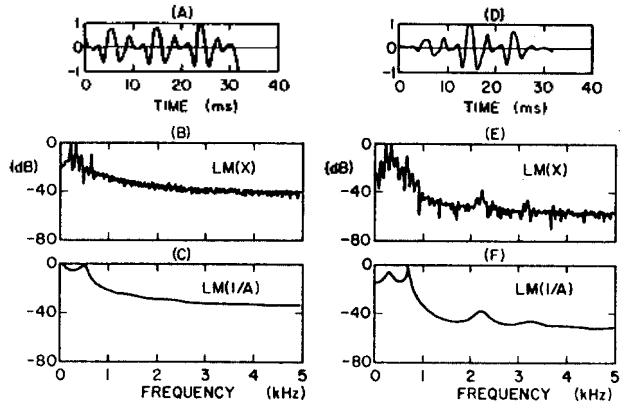

그림4 윈도우함수의 영향
(a) 시간축 윈도우 함수의 적용원
(b) 윈도우를 씌운후 FFT 결과
(c) 윈도우를 씌운 후의 LPC스펙트럼

윈도우함수는 잘라지는 신호의 끝 부분의 크기를 줄여서 잘라짐에 의한 불연속점으로부터 발생하는 주파수 왜곡을 감소시켜주는 역할을 한다. 많이 사용되는 윈도우 함수의 종류는 해밍(hamming), 해닝(hanning), 블랙만(blackmann), 카이저(kaiser)윈도우 등이 있다. 이들중 해밍과 해닝이 많이 사용된다.

(4)

$$h(n)=\begin{cases} 0.54-0.46\cos(2\pi n/N-1), n=0,1,\ldots N-1 \\ 0, otherwise \end{cases}$$

식(4)는해밍윈도우의 식을 나타내며 (그림5)은 각종 윈도우의 형상을 나타낸 것이다.

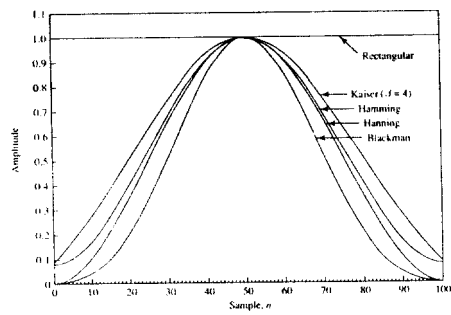

그림5 각종 윈도우 함수의 형상

음성을 FFT한 결과는 유성음의 경우 (그림6) 에서와 같이 주기적인 무늬로 나타나는 피치성분과 울퉁불퉁한 산봉우리 형상으로 나타나는 성도의 성분으로 나뉘어 진다. 무성음의 경우는 피치성분이 없는 잡음성분이 주가 된다.

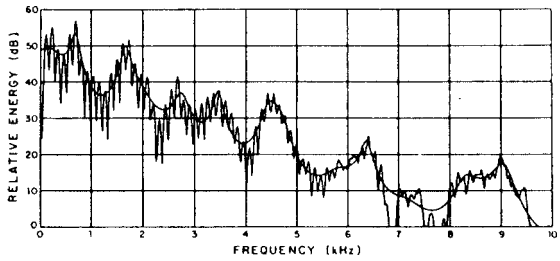

그림6 모음의 FFT결과 및 LPC엔벨로프

스펙트로그램

FFT가 일정구간의 음성표본을 대상으로한 주파수 분석이라면 스펙트로그램은 이러한 FFT를 연속적으로 분석한 것으로 주파수 성분의 변화를 시간적으로 관찰할 수 있게 한 것이다.

스펙트로그램은 지금까지 거의 모든 음성분석과정에 중요한 도구로 사용되어 왔다. 스펙트로그램은 포함되는 표본의 수에 따라서 광대역 스펙트로그램과 협대역 스펙트로그램으로 나뉜다. 협대역 스펙트로그램은 한 프레임의 음성을 분석하는데 필요한 표본의 수가 많으므로 스펙트럼영역에서 피치성분의 주기성을 나타내게 되며 광대역 스펙트럼은 표본의 수가 적으므로 시간영역의 해상도가 높아지 시간영역에서 피치의 주기성분이 관찰되며 주파수영역에서는 해상도가 떨어지게 된다. 여기서 협대역과 광대역이란 말은 주파수 분석을 여러 개의 대역필터로 행한다고 생각할 때 대역필터의 대역폭이 넓고 좁음을 말한다. 따라서 디지털 음성처리의 경우 동일한 표본화율에서 분석 표본수가 작은 경우는 주파수 성분을 나타내는 점수가 작아지므로 각 점이 대표하는 주파수대역은 커져서 광대역이되고 반대의 경우는 협대역이 된다.

4. 선형예측분석(LP Analysis)

음성의 선형예측모델 즉, Linear Predictive Model이란 음성신호를 성대와 성도부분으로 나누어서 이를 대표하는 전달함수를 구하는 기법이다. 이러한 모델에서는 일반적으로 음원을 유성음의 경우 주기적인 임펄스열로, 무성음의 경우는 백색잡음 발생기로 가정하여 이에 해당하는 성도의 전달함수를 수학적으로 구한다.

기본적으로 이 모델링 방법은 음성의 주기적인 성질에 바탕을 두고 한 주기내에서는 과거의 값들을 통해 현재의 값을 예측해 낼 수 있으며 이러한 예측함수의 계수를 구하면 그것이 바로 음성의 특징, 또는 성도의 특징을 나타내는 전달함수라는 것이다.

(그림7)은 음성의 생성모델의 블록도이다.

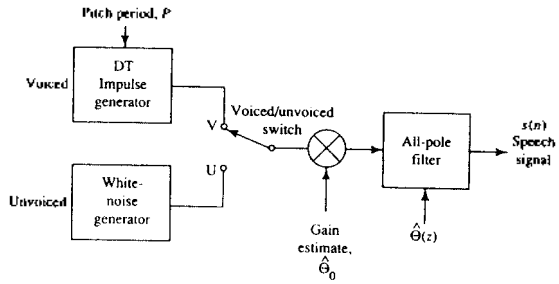

그림7 음성의 생성모델

선형예측모델은 다음 식과 같이 나타난다.

$x(n)$을 음성신호라고 하고 $x'(n)$을 예측된 신호라고 할 때

$$x'(n) = \sum_{i=1,N} a_i x(n-i),$$

N=한 프레임에서 음성 표본의 수

과 같이 주어진다. 이 때 두 신호의 오차 $e(n) = x(n) - x'(n)$의 값을 최소로 하는 a_i를 구하면 이것이 선형예측계수라고 하며 성도의 특성을 나타내는 계수가 된다.

즉, $\frac{\sigma e(n)^2}{\sigma a_i}$ 을 최소로 하는 N개의 a_i를 구하면 된다. 여기서 N개의 방정식이 성립되며 a_i를 구하는 방법은 자기상관법(autocorrelation method), 공분산법(covariance method), 더빈의 방법(Durbin's method)등이 있는데 여기서는 구체적인 수식의 전개는 생략하기로 하고 자세한 사항은 참고문헌[Makhoul][Markel&Gray]을 참고하기 바란다.

MATLAB의 경우(version 5.0이상의 signal processing toolbox)는 lpc라는 함수에 의해 선형예측계수를 구할 수 있다.

나음과 같은 함수를 이용하여 선형예측계수를 구할 수 있다.

a=lpc(x,N);
a: 선형예측계수
N: 분석차수

선형예측분석에 의해 구해진 계수는 입력된 음성신호표본들에 포함된 여러 주기의 음성신호들의 평균적인 성도특성을 나타낸다고 보면 된다. 실제로 음성신호가 여러 주기동안에 약간씩 변하더라도 우리가 분석한 구간내에서는 변하지 않는다고 가정하여 여러 주기의 음성에 대한 특성을 대표치로 구한 것이기 때문에 포함된 여러 주기의 평균적인 성도특성을 나타낸다고 보는 것이 옳다.
이렇게 구한 선형예측계수는 성도의 특성을 나타내기 때문에 결국 FFT분석한 결과의 포락선특성을 나타내는 것으로 볼 수 있다. 따라서 선형예측계수를 통해 구성된 성도의 특성함수에서 봉우리의 위치와 폭을 측정함으로써 음성의 특징주파수인 포만트의 값과 대역폭을 구할 수 있다.
그림6의 실선부분은 선형예측계수로부터 구해진 성도의 스펙트럼을 보여준다.
그림6에서 각 봉우리의 위치에 해당하는 주파수를 포만트라 하며, 봉우리의 폭을 대역폭이라고 한다. 대역폭을 구하는 방법은 포물선보간법에 의한 방법 및 선형예측 방정식의 근으로부터 구하는 방법이 있다.

분석차수의 선택

선형예측분석에서 분석차수를 얼마로 할것인가 하는 것은 분석결과에도 영향을 미친다. 대개는 분석차수를 설정할 때 성도의 길이와 표본화 주파수가 적절한 예측계수의 선택에 영향을 준다. 분석차수를 잘못 선택하면 불필요한 계산시간이 낭비될 수 있을뿐 아니라 원하는 결과를 얻을 수도 없다.
참고문헌에서는 성도의 특성함수 A(z)가 가질 수 있는 시간지연은 소리가 성문으로부터 입술까지 전파하는데 걸리는 시간의 두배정도가 되어야 한다고 한다. 즉, 성도의 길이를 17cm, 소리의 속도를 34cm/ms 라고 했을 경우 가능한 시간지연은 1ms이므로 10KHz의 표본화율이라면 분석차수를 최소 10차로 하여야 한다.[Markel & Gray, pp.154-156) 분석차수는 가능한 한 최소의 차수를 선택하는 것이 바람직하다. 대개 표본화 주파수에 해당하는 값+2~3차의 분석차수를 택하면 된다. 차수가 너무 크면 분석에 필요치 않은 스펙트럼 봉우리를 부각시켜 결과분석에 혼란을 가져올 수 있다. 즉, 10KHz표본화율에서 10차의 분석을 하는 것보다 20차이상의 분석을 한다는 것이 더 좋지는 않다는 것이다.

분석구간간격의 선택

선형예측분석구간간격의 선택은 다음과 같은 두가지 요인에 따른다. 첫째, 전이구간의 선택이 잘 되도록 한다. 분석구간내에서 성도특성의 변화가 거의 없다고 생각되는 범위를 택하면 된다. 유성음의 경우 보통 15~20ms 구간을 선택하며 포함되는 표본의 수N은 표본화 주파수 f_s(kHz) 곱하기 15~20ms 가 된다. 무성음의 경우는 좀 더 작은 구간을 택할 필요가 있는데 보통 10ms가 적당하다고 한다.

윈도우씌우기

윈도우함수를 씌우는 것은 유한개의 음성표본을 잘라내는 데서 생기는 스펙트럼의 오차를 감소하기 위하여 필요한데 분석구간이 길다면 오차는 적어지지만 구간내의 변화를 검출할 수 없게되므로 윈도우를 씌우는 것이 필요하게 된다. 윈도우함수는 FFT의 경우와 동일한 함수들을 사용하면 된다.

프리엠퍼시스(Pre-emphasis)

음성신호로부터 음원신호나 입술의 방사효과를 제거한 순수한 성도의 특성을 얻기 위해서는 이들의 특성을 제거하는 과정이 필요한데 이를 프리엠퍼시스라고 한다. 프리엠퍼시스는 다음과 같은 전달함수를 갖는 필터를 통과시켜 줌으로써 가능하다.

$$H(z) = 1 - \mu z^{-1}, 0.9 \le \mu \le 1$$

μ값은 유성음의 경우는 커지고 무성음의 경우는 아주 작다. 프리엠퍼시스의 효과는 높은 주파수 성분을 상대적으로 높여주는 고역통과필터의 역할을 한다.

5. 결 론

이 글에서는 FFT와 LPC 분석법의 원리를 알아보고 우리가 음성신호를 분석할 경우 고려할

사항들을 알아보았다. 요즘은 멀티미디어 컴퓨터의 보급으로 소리 및 음성을 분석할 수 있는 기기나 소프트웨어가 보편화 되어가고 있으나 전문적으로 분석하고 결과를 해석하기 위해서는 보다 근본적인 과정에 대한 이해가 필요하다고 본다. 이 글이 공학도가 아닌 인문과학등의 분야에서 음성을 응용하고자 하는 사람들에게 도움이 될 수 있기를 바란다.

본 논문의 원고와 발표에 사용한 그림등은 아래 홈페이지에서 다시 볼 수 있다.
http://soback.kornet.nm.kr/~cwjo

참고문헌

<FFT>

L.R.Rabiner, R.W.Shafer, 'Digital Processing of Speech Signals',pp.250-354, Prentice-Hall (1978)

이채욱, '디지탈신호처리-기초와응용-', pp.89-138,청문각(1994)

J.G.Proakis, D.G.Manolakis, 'Digital Signal Processing', 3rd ed., pp.230-499, Prentice-Hall (1996)

V.K.Ingle, J.G.Proakis, 'Digital Signal Processing using MATLAB V.4', pp.40-79,116-181, PWS (1997)

<LPC>

B.S.Atal,S.L.Hanauer, 'Speech Analysis and Synthesis by Linear Prediction of the Speech Wave',pp.13-31, JASA, Vol.50, No.2(part2), (1971)

J.Makhoul, 'Linear Prediction: A Tutorial Review',pp.115-134, Proceedings of IEEE, Vol.63, No.4, April (1975)

L.R.Rabiner, R.W.Shafer, 'Digital Processing of Speech Signals',pp.396-461, Prentice-Hall (1978)

J.D.Markel & A.H.Gray,'Linear Prediction of Speech',Springer Verlag (1980)

J.R.Deller, Jr,J.G.Proakis, J.H.L.Hansen,' Discrete-Time Processing of Speech Signals', pp.260-351, McMillan (1993)

FFT 및 Spectrogram 분석도구

<Shareware 및 Demo판>

CoolEdit: win용 신호분석기

http://www.syntrillium.com/cool.htm

Gram32 : win95용 실시간 스펙트로그램

http://www.winboss.dk/public/shareware/filareas/filearea212.htm

Signalyser: MAC용 음성분석기

http://www.agoralang.com/2410/signalyze.html

Wincecil : win용 주파수,피치분석기

http://gopher.sil.org/FTP/SOFTWARE/DOS/

<상용>

Matlab:

http://www.mathworks.com/

공개판 sptools available

ESPS/Xwave: http://www.entropic.com/

<기타 소프트웨어 패키지관련 정보>

comp.speech archive site

http://svr-www.eng.cam.ac.uk/comp.speech/FAQ Packages.html