

스펙트럼과 스펙트로그램의 이해

성균관대학교 의과대학 강북삼성병원 이비인후과학교실
진 성 민

= Abstract =

Introduction to the Spectrum and Spectrogram

Sung Min Jin, MD

Department of Otolaryngology, Kangbuk Samsung Hospital, School of Medicine, Sungkyunkwan University, Seoul, Korea

The speech signal has been put into a form suitable for storage and analysis by computer, several different operation can be performed. Filtering, sampling and quantization are the basic operation in digitizing a speech signal. The waveform can be displayed, measured and even edited, and spectra can be computed using methods such as the Fast Fourier Transform (FFT), Linear predictive Coding (LPC), Cepstrum and filtering. The digitized signal also can be used to generate spectrograms. The spectrograph provide major advantages to the study of speech. So, author introduces the basic techniques for the acoustic recording, digital signal processing and the principles of spectrum and spectrogram.

KEY WORDS : Spectrum · Spectrogram.

음성검사법은 이학적 검사, 공기 역학적 검사, 청각심리검사, 성대진동검사, 음향학적 검사 그리고 신경생리학적 검사로 크게 분류할 수 있다. 이들 중 음향학적 검사에 포함되어 있는 스펙트럼과 스펙트로그램에 대하여 소개 하고자 한다.

음성신호의 저장 및 분석

음성신호에 대한 분석은 아날로그(analogue) 신호에 대한 분석과 디지털(digital) 신호에 대한 분석으로 대별되나 최근 들어서는 주로 컴퓨터를 이용하여 디지털 신호를 분석하는 방법이 널리 이용되고 있다. 따라서 컴퓨터를 이용하여 디지털 신호를 분석하기 위하여, 연속된 아날로그 신호인 음성신호를 컴퓨터로 입력하여 숫자화 시켜 저장하는 과정이 필요한데, 이 과정을 digital signal processing 이라 한다(Fig. 1)¹⁾. 예를 들어 사람의 목소리를 컴퓨터로 저장하는 과정은 AD conversion(analogue to digital conversion) 이 되며, 컴퓨터에 저장된 소리가 스피커를 통해 나

오게 되면 DA conversion(digital to analogue conversion)이라 표현 할 수 있다. 음성신호를 컴퓨터에 저장하기 위해서는 중요한 두 가지 요소를 알고 있어야 하는데, 표본 추출율(sampling rate)과 양자화(quantization)가 그것이다.¹⁾²⁾

1. 표본 추출율(Sampling rate)

표본 추출율은¹⁾²⁾ 주파수의 정보와 관련된 요소로서, 연속적인 형태의 신호(analogue signal)인 원래 음성신호에 1초당 몇 개의 점을 선택하여 이를 숫자화시켜 컴퓨터에 정보를 저장할 것인가 하는 것이다. 따라서 단위는 헤르츠(Hz)를 사용한다(Fig. 2).

연속적인 형태의 신호인 음성신호를 컴퓨터에 입력할 때 주파수와 관련된 정보를 얼마만큼 선택할 것인가, 다시 말해 연속적인 형태의 신호인 원신호의 값을 어떻게 하면 정보의 손실이 없이 숫자로 표현할 수 있는가를 정해야 하는데, 결론부터 이야기 하면, 분석하고자 하는 최대 주파수 영역의 2배에 해당하는 값을 표본 추출율로 선택하면 된다. 예를 들어 저장된 신호를 5,000Hz까지 분석하고 싶다면, 1초당 10,000개의 주파수와 관련된 값을 취하면 된다. 즉 표본 추출율을 10,000Hz로 정하면 되는 것이다. 이는 분석하고자 하는 최대 주파수 영역의 2배에 해당하는 값을 표본 추출율로 선택하면 컴퓨터로 입력된 신호는 원래의 아날로그 신호와 동일한 정보를 갖게 된다는 Nyquist's Sampling

논문접수일 : 2008년 9월 24일
심사완료일 : 2008년 11월 7일
책임저자 : 진성민, 110-746 서울 종로구 평동 108
성균관대학교 의과대학 강북삼성병원 이비인후과학교실
전화 : (02) 2001-2267 · 전송 : (02) 2001-2273
E-mail : strobojin@hanmail.net

theory에 따른 것으로, 일반적으로 음악 CD를 제작할 때 정하는 표본 추출율은 인간의 가청영역(약 20~20,000Hz)의 두배에 해당하는 44,000Hz까지의 표본 추출율을 이용하게 된다.¹⁾²⁾

2. 양자화(Quantization)

표본 추출율이 주파수에 대한 정보와 관련된 내용이라면, 양자화는 진폭(amplitude), 즉 에너지와 관련된 요소이다. 이는 진폭(amplitude)에 관련된 정보를 얼마나 자세히 담을 것인가를 결정하는 과정으로 연속된 진폭 값을 개개의 숫자로 자르는 과정이라 할 수 있으며, 단위는 비트(bit)를 사용하게 된다(Fig. 3).¹⁾²⁾ 그런데, 양자화 값은 컴퓨터로 신호를 입력할 때 그 수치를 정해주는 표본 추출율과 달리 경험에 따른 선택을 하게 되고, 일반적으로 분석기계 자체에 그 값이 정해져 있어 검사자가 따로 그 값을 정해주시는 않는다. 보통 사람의 목소리는 8-bit 이상의 conversion을 하게 되고, 음악 CD는 16-bit 이상의 conversion을 하여, 사람의 목소리는 28단계(=256 quantization level) 이상의 값을, 음악 CD는 216단계(=65,536 quantization level) 이상의 값을 취하게 된다.¹⁻³⁾

음파(Sound Wave)

음파는 가로축이 시간이고 세로축이 진폭으로 표현되는 그래프로 나타낼 수 있고, 음파는 단순파(simple wave, sine wave)와 복합파(complex wave)로 대별 할 수 있다.²⁾

1. 단순파(Simple wave, sine wave)

단순파는²⁾ 음파의 가장 단순한 형태로서 일정한 모양이

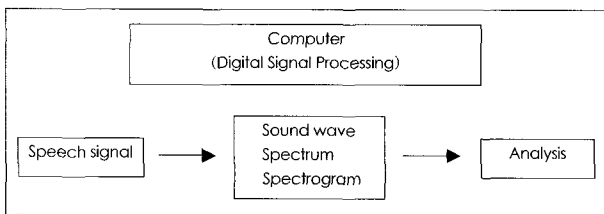


Fig. 1. Acoustic recording and digital signal processing.

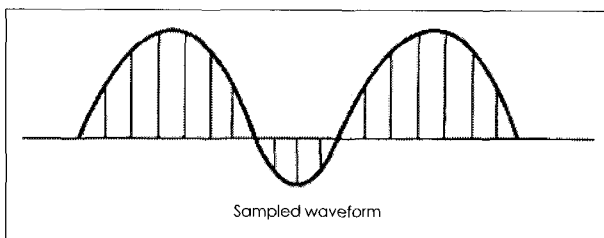


Fig. 2. Illustration of sampling of a waveform. Samples are taken at the points marked by vertical line.

반복되는 주기파(periodic wave)이며, 그림에서와 같이 주기(period), 주파수(frequency, cps, Hz), 진폭(amplitude, dB) 그리고 위상차(phase difference)를 관찰 할 수 있다 (Fig. 4).

2. 복합파(Complex wave)

복합파는²⁾⁴⁾ 두 가지 이상의 요소가 복합된 형태의 파형으로 사람의 목소리를 비롯한 대부분의 주변에서 들을 수 있는 소리이다. 이는 앞서 언급한 바와 같이 일정한 패턴이 반복되는 주기파와 그렇지 아니한 비주기파(aperiodic wave)로 나누어 볼 수 있다. 이때 모든 주기파는 그것이 아무리 복잡한 형태를 가지고 있다 할 지라도 일정수의 단

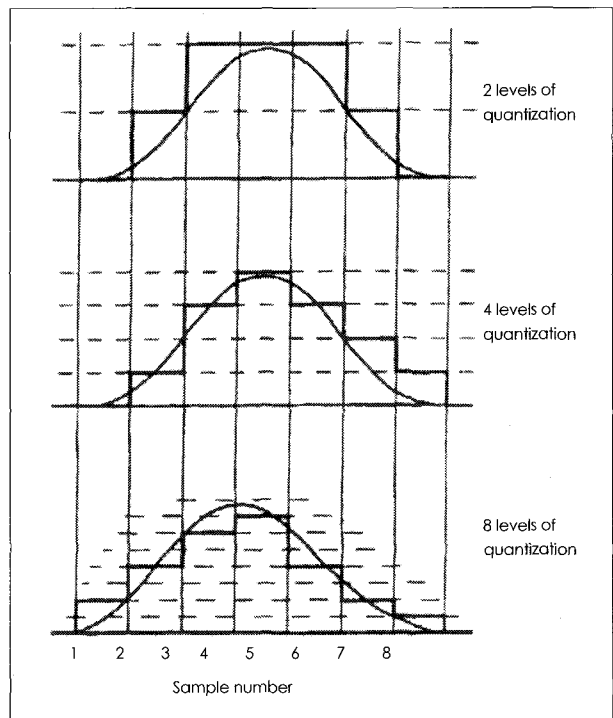


Fig. 3. Illustration of the effects of using different quantization levels to represent a smooth waveform. As the number of levels increase, the fit to the smooth curve improves.

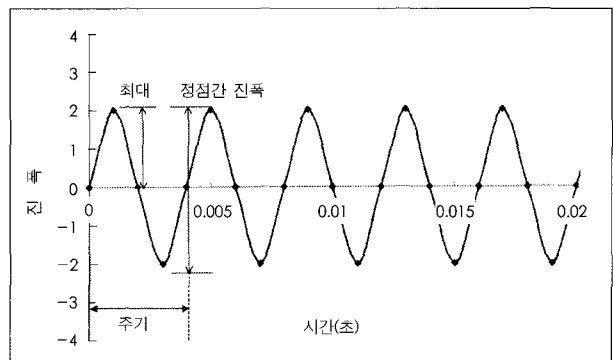


Fig. 4. Illustration of the sine wave.

순파의 합으로 이루어진 것이며, 이들 단순파 각각의 주파수의 최대공약수는 복합파의 주파수가 된다(Fig. 5).

스펙트럼(Spectrum)

앞서 언급한 단순파의 경우 그 파형 만으로도 그 파형이 갖는 진폭과 주파수를 쉽게 알 수 있지만, 여러 개의 단순파가 합쳐진 복합파는 그 복합파가 어떤 단순파들로 구성되어 있는지를 알기가 쉽지 않다. 따라서 이 복합파가 어떤 단순파들이 합쳐져서 이루어진 것인가를 보여주는 방법이 스펙트럼이다.²⁾³⁾⁵⁾⁶⁾

하얀색의 햇빛을 프리즘을 통해 나누어보면 빨강부터 보라까지의 색깔들이 나누어져 있는 것을 관찰할 수가 있는데, 이것이 빛의 스펙트럼인 것과 같이 소리 역시 스펙트럼을 통해 분석하게 되면 그 소리의 각각의 음파의 내용들을 알 수 있으며, 이는 프리즘을 이용한 빛의 스펙트럼과 같은 원리인 것이다.

음파를 그래프로 표시하면 x축은 시간, y축은 진폭인 그래프로 표시되나, 이 음파를 분석하여 스펙트럼으로 표시하면 x축은 주파수, y축은 진폭인 그래프로 표시된다. 따라서 아래 그림과 같은 실선의 복합파를 스펙트럼으로 분석하면 세가지 종류의 주파수와 진폭을 갖는 단순파로 이루어진

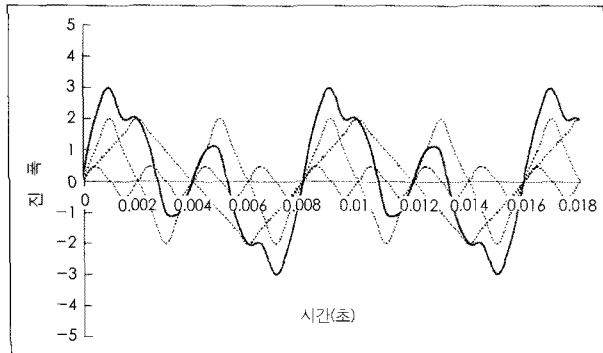


Fig. 5. Illustration of the complex wave.

것을 알 수 있다(Fig. 6).²⁾

일반적으로 음성검사실에서 시행하는 음성신호 분석은 정확성과 신뢰성이 높은 디지털 신호 분석방법을 이용하는데, 스펙트럼 분석에 있어 디지털 신호에서 시간 함수를 주파수 함수로 변환하는 방법은 Fourier 변환과 z변환이 있다.

Fourier 변환을²⁻⁴⁾ 이용하여 분석하는 방법으로는 불연속 푸리에 변환(discrete Fourier transform, DFT)이 있는데, 이는 디지털 신호가 연속된 신호가 아니고 불연속적인 일련의 점의 형태이기 때문에 이를 분석함에 붙여진 이름이라고 이해하면 되며, 이와 같은 변환 방식을 더 빠르게 계산하기 위해 최근에는 고속 푸리에 변환(fast Fourier transform, FFT) 방식을 사용하고 있다. 이와 같은 FFT에 의해 얻어지는 스펙트럼은 아래 그림과 같이 각각의 주파수 별로 에너지가 어떻게 분포되어 있는지를 쉽게 알 수 있다(Fig. 7).

z 변환을²⁾³⁾ 이용하여 얻어지는 스펙트럼은 아래 그림과 같은 LPC(linear predictive coding) 스펙트럼을 얻을 수 있고, LPC 스펙트럼에서는 포먼트에 대한 정보를 보기가 유용하다(Fig. 8).

스펙트럼이 만들어지는 과정을 알아보면, 우선 디지털 녹음 방식으로 얻어진 음성신호는 앞서 언급한 바와 같이 숫자화되어 저장되며, 연속된 신호를 일련의 점들로 표현하게 된다. 이 점들을 일정한 숫자로 잘라서 몇 개의 점(윈도우 길이=점들이 나타내는 시간)을 한 단위로 할 것인가 결정해야 한다. 이때 FFT에서는 2의 제곱(ex. 8, 32, 64, 128, 256...)의 개수로 점들로 잘라서 분석을 하게 되고, 이와 같이 잘라진 점의 개수가 정해지면, 점의 개수 즉 윈도우의 길이의 역수로 취한 값이 여파기의 대역폭이 된다. 이 과정이 끝나고 난 후 앞서 잘라진 점들이 여파기를 통과하면서 각 주파수 대역폭에서 음파의 구성성분을 여과하여 각 주파수 대역의 진폭 수준이 계산된다. 그 다음 각 윈도우에 비친 음향신호에 대한 스펙트럼이 만들어지는데 이를 파워 스펙트

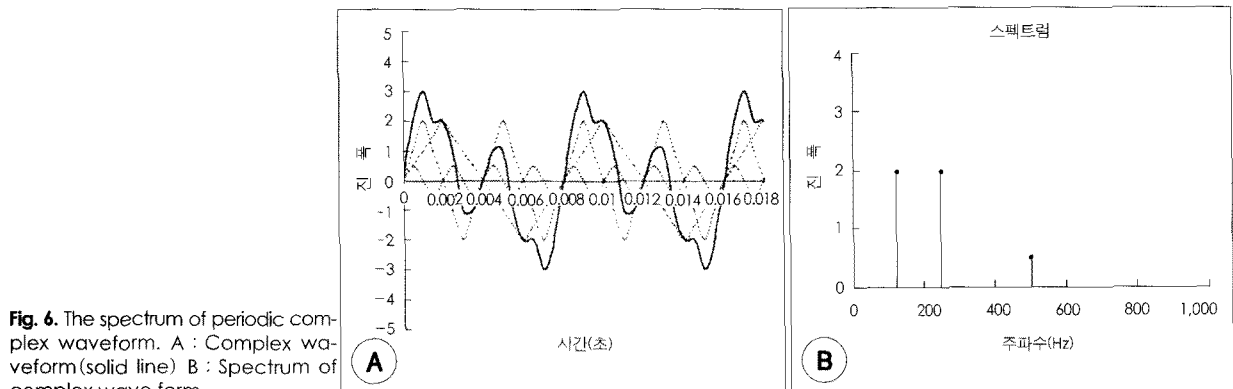


Fig. 6. The spectrum of periodic complex waveform. A : Complex waveform(solid line) B : Spectrum of complex wave form.

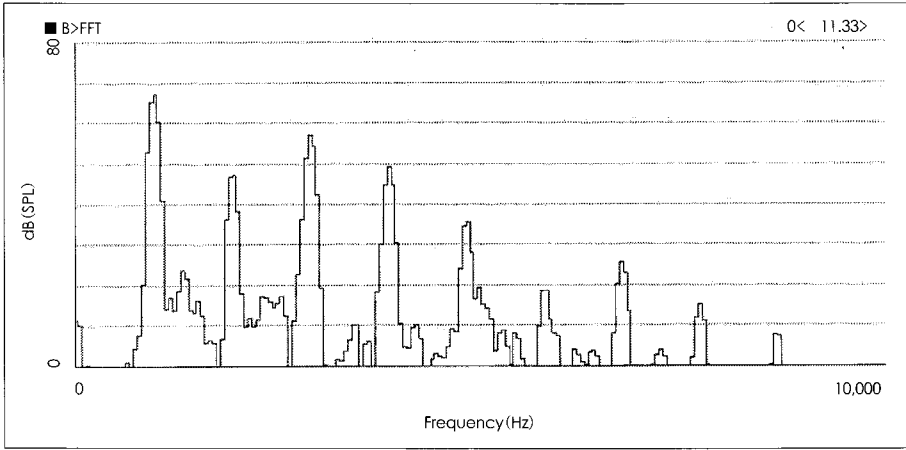


Fig. 7. Illustration of the Fourier spectrum.

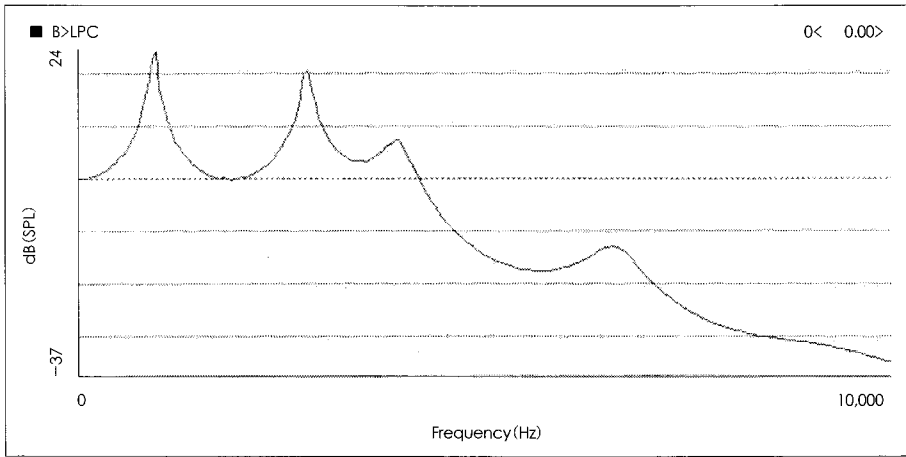


Fig. 8. Illustration of the LPC (linear predictive coding) spectrum.

럼(power spectrum)이라 한다(Fig. 7).²⁾ 앞의 내용을 숫자로 예를 들어 표현을 해보면 다음과 같다. 예를 들어, 1초의 소리를 10,000Hz 표본 추출율(=5,000Hz까지 분석)로 10,000개의 점을 찍어 정보로 저장하고, 256point(=0.0256초)씩 잘라서 분석하고자 한다면, 여과기의 주파수 대역폭은 윈도우 길이를 역수로 취한 값이 되므로, sampling rate/윈도우 길이=10,000/256이 되어 주파수 대역폭은 39 Hz가 된다.

따라서 39Hz의 주파수 대역폭을 갖는 여과기는 5,000Hz 내에 128개가 존재하는 셈이 된다(5,000Hz/39Hz=128개). 이들 대역여과기를 잘라진 점들이 통과하면서 필요로 하는 부분의 각 주파수 대역의 진폭수준 계산되고 각 윈도우에 비친 음향신호에 대한 스펙트럼이 만들어진다. 이에 따라 앞서 보여준 FFT power spectrum이 만들어지고, 각 주파수 대역별로 에너지 값을 얻을 수 있게 된다.²⁾

스펙트로그램(Spectrogram)

우리가 일반적으로 분석하는 목소리는 시간에 따라 그

질도 변하고, 그 크기도 변화하는 매우 복잡한 것이다. 다시 말해 목소리는 시간에 따라 그 구성 성분의 종류와 양이 변화 한다. 따라서 스펙트로그램은 시간의 개념 없이 음파를 구성하고 있는 내용물의 질과 양을 보여주는 스펙트럼과 달리 시간의 축을 덧붙여 주파수와 진폭의 시간에 따른 변화를 보여주는 삼차원 적인 그림이라 할 수 있다. 따라서 스펙트럼은 x와 y의 두 축으로 구성되어 있는 반면, 스펙트로그램은 x, y, z의 세 축을 가지고 있다. 따라서 스펙트로그램은 x축을 시간, y축을 주파수, z축을 진폭(강도)로 표시 한다. 이때 z축은 스펙트로그램상에서 색깔의 진하기로 표시되어 진하면 진폭(강도)가 큰것으로, 연하면 작은것으로 표시된다.²⁻⁴⁾

스펙트로그램이 만들어지는 과정을 알아보면, 우선 디지털 녹음 방식으로 얻어진 음성신호는 숫자화되어 저장되며, 연속된 신호를 일련의 점들로 표현하게 된다. 이 점들을 일정한 숫자로 잘라서 몇 개의 점(윈도우 길이=점들이 나타내는 시간)을 한 단위로 할 것인가 결정해야 한다. 이때 FFT에서는 2의 제곱(ex. 8, 32, 64, 128, 256...)의 개수로 점들로 잘라서 분석을 하게 되고, 이와 같이 잘라진 점의 개

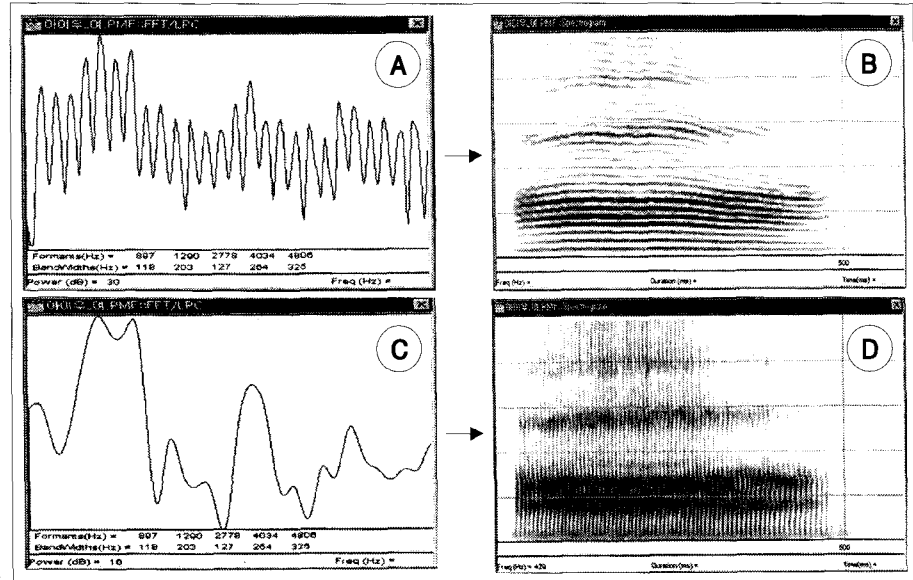


Fig. 9. Spectrum and spectrogram
A, B : Spectrum. C, D : Spectrogram.

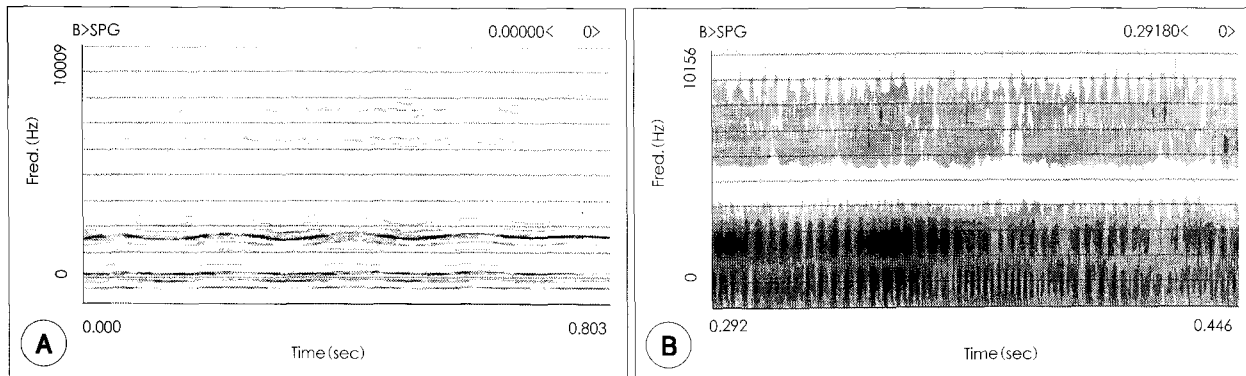


Fig. 10. Narrow band pectrogram (A) and wide band spectrogram (B).

수가 정해지면, 점의 개수 즉 윈도우의 길이의 역수로 취한 값이 여과기의 대역폭이 된다. 이 과정이 끝나고 난 후 앞서 잘라진 점들이 여과기를 통과하면서 각 주파수 대역폭에서 음파의 구성성분을 여과하여 각 주파수 대역의 진폭 수준이 계산된다. 그 다음 각 윈도우에 비친 음향신호에 대한 스펙트럼이 만들어진다. 이와 같이 앞서 스펙트럼이 만들어지는 과정과 동일한 과정을 거친 후, 이 스펙트럼을 시간순으로 늘어 놓으면 스펙트로그램이 완성된다(Fig. 9).²⁾³⁾

스펙트로그램은 음파를 분석하는 주파수 대역폭에 따라 협대역 스펙트로그램(narrow band spectrogram)과 광대역 스펙트로그램(wide band spectrogram)으로 나뉜다(Fig. 10).

협대역 스펙트로그램은 일반적으로 45Hz 이하의 좁은 주파수 대역폭을 이용하며, 광대역 스펙트로그램은 300Hz 이상의 넓은 주파수 대역폭을 이용하여 분석하게 된다. 각각의 스펙트로그램이 일견하여 구분 되는 점은 협대역 스펙트로그램에서는 가로줄의 가는 선들이 관찰되는 반면, 광대역 스펙트로그램에서는 세로줄이 관찰되고, 가로로 굵은 띠

(포먼트)들이 관찰 된다는 점이다.³⁾

협대역 스펙트로그램에서 주파수 대역폭이 좁다는 것은 그만큼 더 촘촘한 여과기를 사용한 것이므로, 분석하고자 하는 음향신호를 구성하고 있는 각 주파수 성분을 하나하나 세밀하게 걸러내서 그 음향신호가 어떤 주파수 성분으로 구성되어 있는지를 세밀하게 알아 낼 수 있다. 광대역 스펙트로그램에서 주파수 대역폭이 넓다는 것은 그만큼 더 성긴 여과기를 이용한 것이므로 분석하고자 하는 음향신호를 구성하고 있는 각 주파수 성분을 하나하나 세밀하게 걸러낼 수는 없지만, 특별히 어떤 주파수 대역에서 강한 음향 에너지가 몰려 있는지를 일목요연하게 보여주며, 또 광대역 스펙트로그램은 짧은 윈도우로 음향신호를 분석한 것으로 시간과 관련된 정보를 세밀하게 알 수 있고, 이러한 이유로 광대역 스펙트로그램에서는 세로선을 관찰 할 수 있다. 따라서 협대역 스펙트로그램에서는 음향신호가 가진 주파수 관련 정보를 섬세하게 알 수 있지만, 시간과 관련된 정보는 잘 알 수 없는 반면, 광대역 스펙트로그램에서는 음향신호가

가진 시간 정보는 잘 알 수 있지만 주파수 관련된 정보를
섬세하게 알 수는 없다.

일반적으로 말소리 연구에서는 포먼트 정보와 시간 정보
가 매우 중요하기 때문에 협대역 스펙트로그램 보다는 광
대역 스펙트로그램이 많이 사용된다.

중심 단어 : 스펙트럼 · 스펙트로그램.

REFERENCES

1) Kent RD, Read C. *The acoustic analysis of speech. 2nd ed. Delmar:*

Singular/Thomson Learning;2002. p.53-103.
 2) 신지영. 말소리의 이해, 음성학. 음운론 연구의 기초를 위하
 여. 한국문화사;2000. p.125-62.
 3) Baken RJ, Orlikoff RF. *Clinical measurement of speech and voice.*
San Diego: Singular/Thomson Learning;2000. p.225-96.
 4) Titze IR. *Principle of voice production. Englewood cliffs: Prentice-*
Hall;1994. p.112-60.
 5) Borden GJ, Harris KS. *Speech science primer. 2nd ed. Baltimore/*
London: Williams & Wilkins;1984. p.89-113.
 6) Sataloff RT. *Voice science. San Diego: Plural Publishing;2005. p.185-*
201.