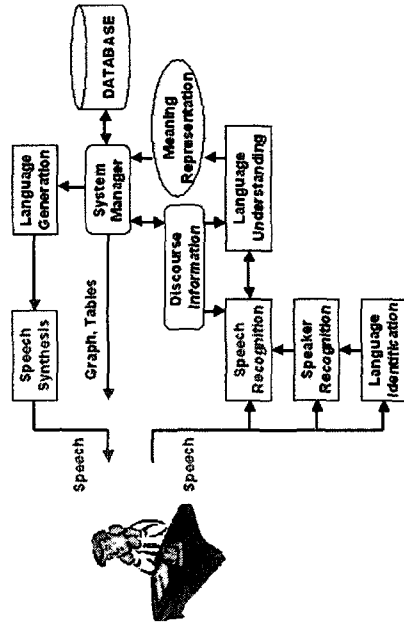


인문계 전공자를 위한 음성정보처리의 기초

2002. 11. 15.

부산대학교 김형순 교수
kimhs@pusan.ac.kr

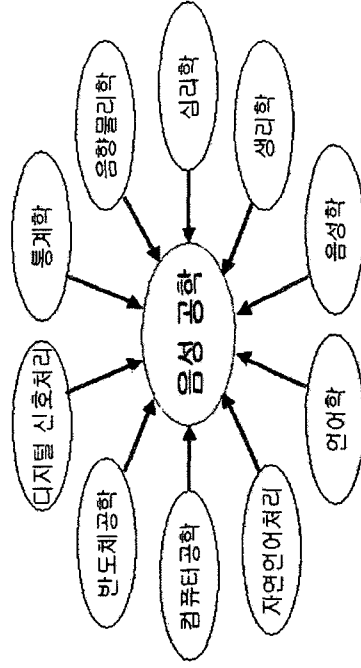
Voice User Interface (VUI)



음성공학의 연구 분야

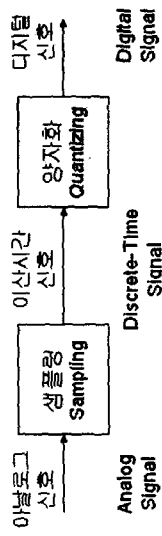
- 음성분석 (Speech Analysis)
- 음성부호화 (Speech Coding)
- 음성인식 (Speech Recognition)
- 음성합성 (Speech Synthesis)
- 화자인식 (Speaker Recognition)
- 음질개선 (Speech Enhancement)
- 기타 : 발화속도 변환, 음색변환 등

음성공학 관련 기술



아날로그-디지털 변환 (A/D Conversion)

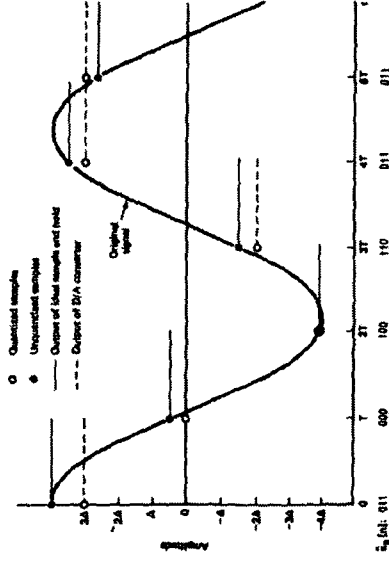
- 아날로그-디지털 변환 (Analog-to-Digital Conversion)



PNU Speech Comm. Lab

아날로그-디지털 변환 (A/D Conversion)

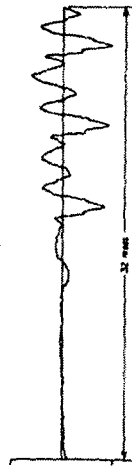
- 아날로그-디지털 변환 (Analog-to-Digital Conversion)



PNU Speech Comm. Lab

아날로그 신호와 이산시간 신호

- 아날로그 신호 (Analog Signal) 또는 연속시간 신호 (Continuous-Time Signal)



- 이산시간 신호 (Discrete-Time Signal)



샘플링 주파수 = 8000 Hz

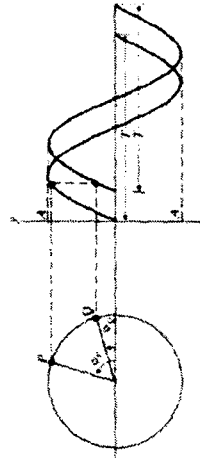
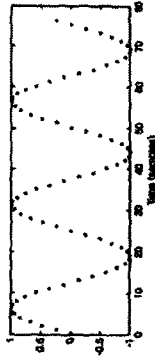
PNU Speech Comm. Lab

이산시간 신호의 예

- 정현파 (sinusoid) 수열

$$x(n) = A_0 \sin(\omega_0 n + \phi_0)$$

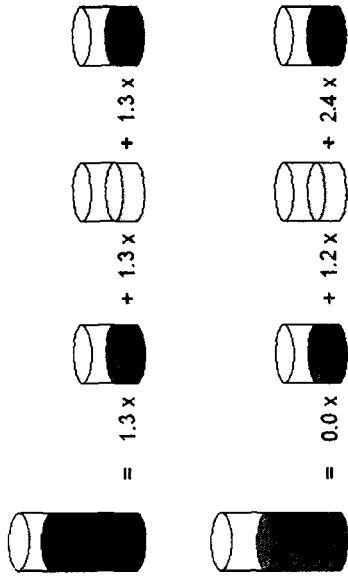
진폭 주파수 위상
진폭 주파수 위상



PNU Speech Comm. Lab

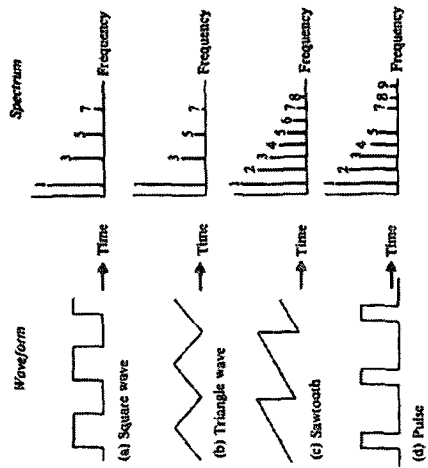
시간 영역과 주파수 영역

■ 스펙트럼의 개념



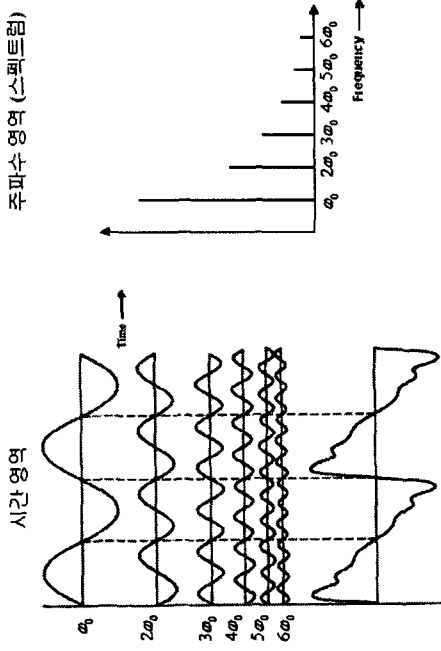
PNU Speech Comm. Lab

시간 영역과 주파수 영역



PNU Speech Comm. Lab

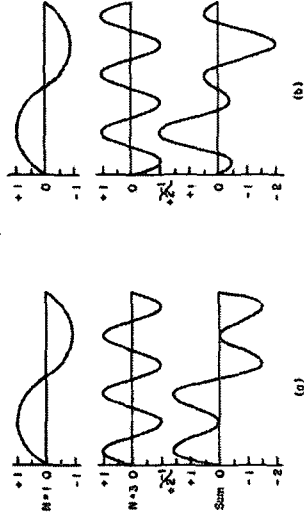
시간 영역과 주파수 영역



PNU Speech Comm. Lab

시간 영역과 주파수 영역

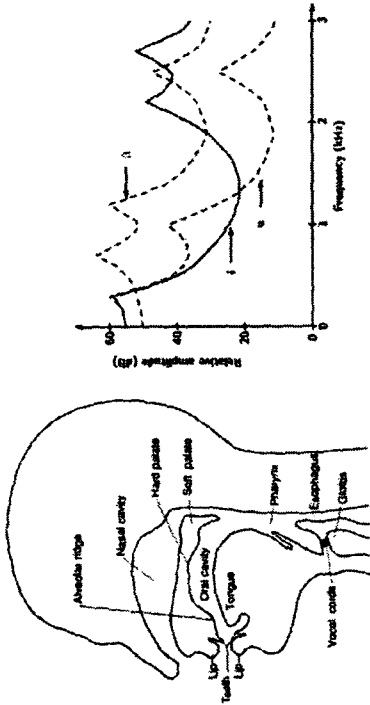
■ 스펙트럼 크기가 같아도 위상이 다르면 파형 모양이 달라짐



- 그러나, 사람은 소리의 위상 특성에 상대적으로 둔감함

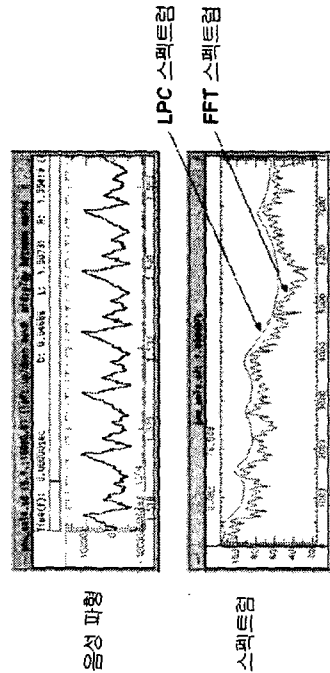
PNU Speech Comm. Lab

음성 발생 기관 및 모음의 스펙트럼



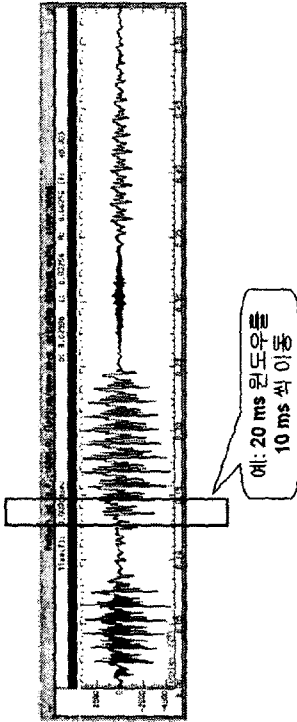
음성신호의 분석

- 대표적인 음성신호 분석 방법
 - FFT (Fast Fourier Transform) 분석
 - LPC (Linear Predictive Coding) 분석



음성신호의 분석

- 음성신호의 단구간(short-time) 분석
 - 음성학적 정보 분석은 비교적 동일한 특성을 가지는 단구간 신호의 분석을 통해 이루어짐



Fast Fourier Transform (FFT) 분석

- Discrete Fourier Transform(DFT)
 - 시간영역에서 제한된 이산 신호의 Fourier 변환

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

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

연관관계 (correlation)

$$\sum_n x(n)y^*(n)$$

복소수 기호

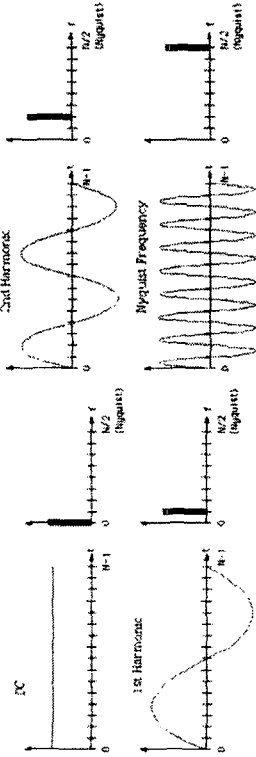
$$e^{j\omega n} = \cos \omega n + j \sin \omega n$$

($\omega = \frac{2\pi}{N}k$ 일 경우)

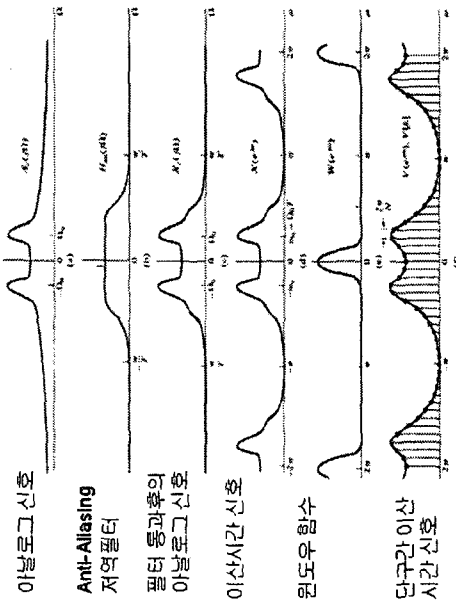
- Fast Fourier Transform (FFT)
 - Discrete Fourier Transform(DFT)의 고속 구현 형태임
 - 예를 들어, N=1024일 경우 DFT의 계산량의 1/100 소요

FFT 분석

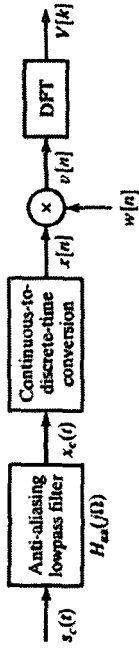
- 정현파 신호의 FFT 분석 예



단구간 Fourier 변환

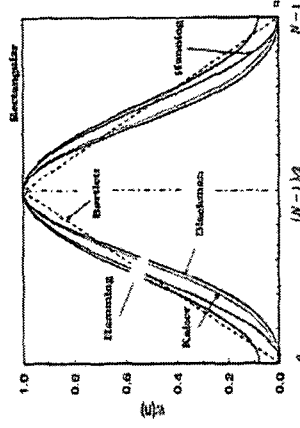


음성신호의 FFT 분석



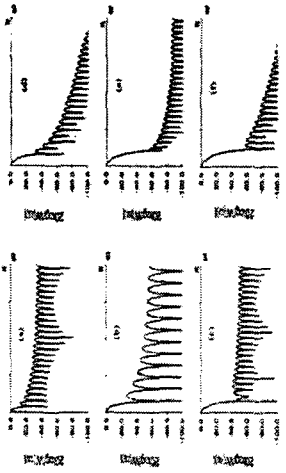
단구간 Fourier 변환

- 윈도우 함수의 역할
 - 시간영역에서의 윈도우 함수들의 형태



단구간 Fourier 변환

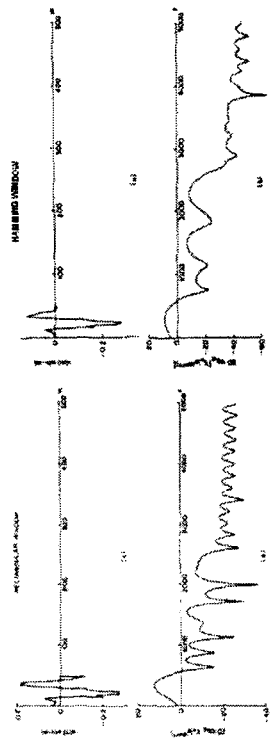
- 윈도우 함수의 역할 (계속)
 - 여러 윈도우 함수들의 주파수 응답특성



PNU Speech Comm. Lab

단구간 Fourier 변환

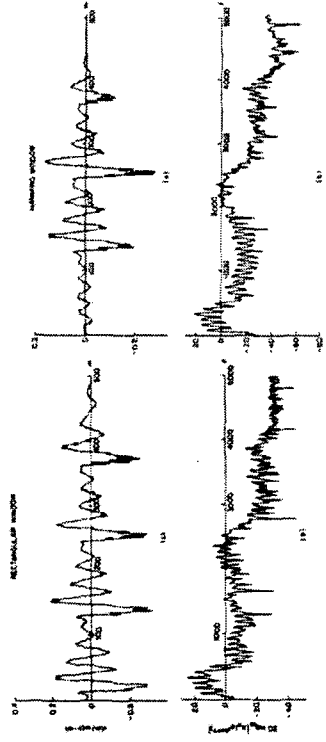
- 음성 스펙트럼 분석의 예 (50 msec 윈도우 크기)



PNU Speech Comm. Lab

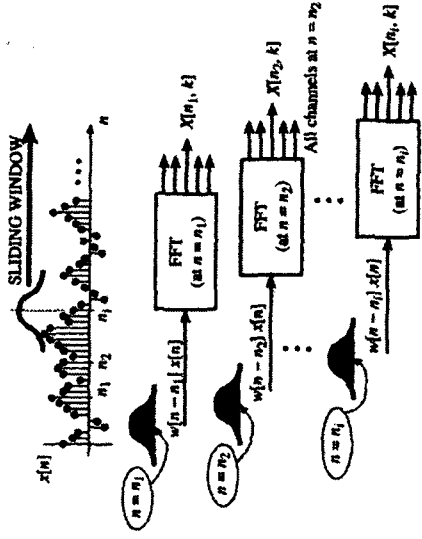
단구간 Fourier 변환

- 음성 스펙트럼 분석의 예 (50 msec 윈도우 크기)



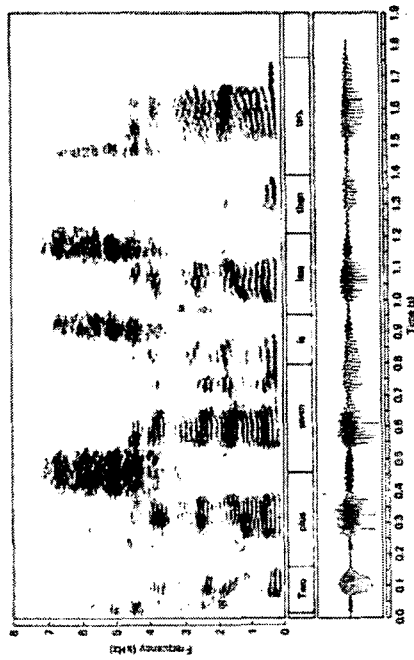
PNU Speech Comm. Lab

단구간 Fourier 변환 : 스펙트로그램 생성과정



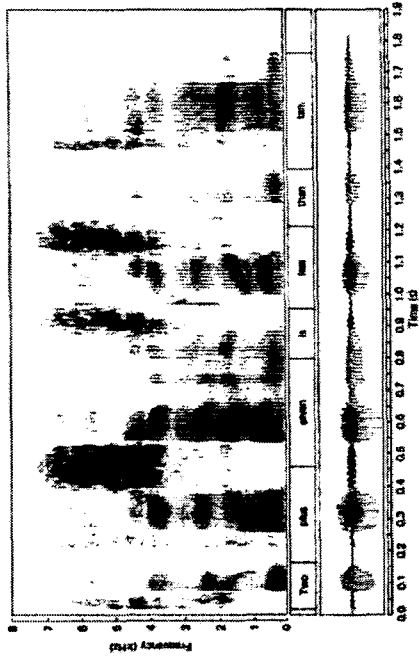
PNU Speech Comm. Lab

협대역 스펙트로그램 (Narrowband Spectrogram)



PNU Speech Comm. Lab

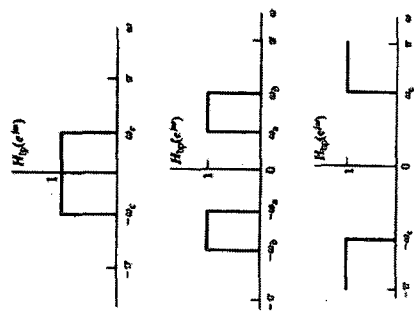
광대역 스펙트로그램 (Wideband Spectrogram)



PNU Speech Comm. Lab

필터 (Filter)

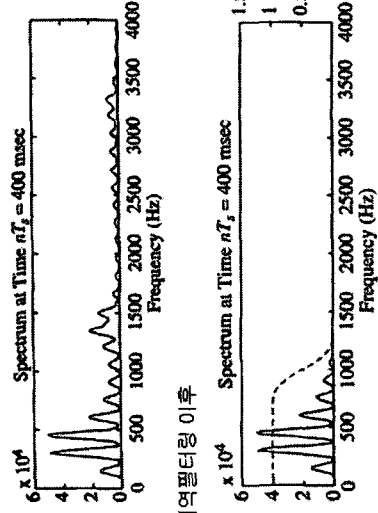
- 이상적인 필터의 예
 - 저역 필터
Lowpass filter (LPF)
 - 대역 필터
Bandpass filter (BPF)
 - 고역 필터
Highpass filter (HPF)



PNU Speech Comm. Lab

저역필터(Lowpass Filter)의 동작 예

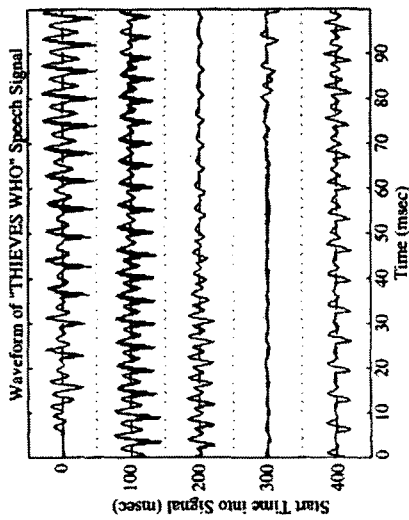
- 필터링 전후의 스펙트럼 특성
 - 저역필터링 이전
 - 저역필터링 이후



PNU Speech Comm. Lab

지역필터의 동작 예

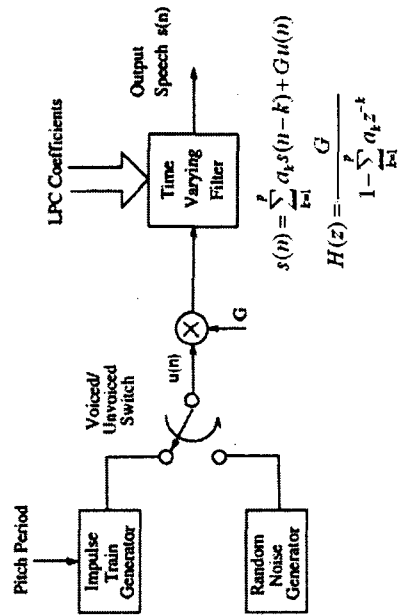
- 필터 통과 전의 음성신호



PNU Speech Comm. Lab

선형예측부호화 (LPC) 분석

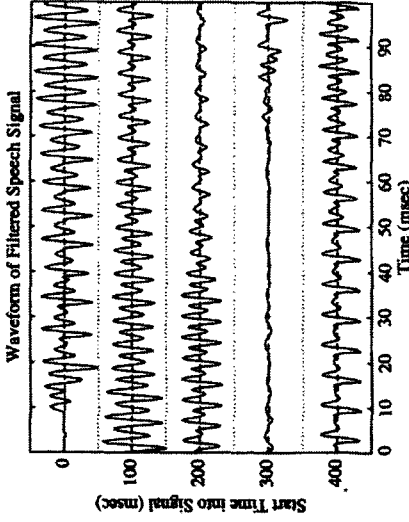
- 선형예측부호화 (Linear Predictive Coding) 모델



PNU Speech Comm. Lab

지역필터의 동작 예

- 지역필터 통과 후의 음성신호



PNU Speech Comm. Lab

LPC 분석

- LPC 분석의 장점
 - 음성신호(특히 모음)에 적합한 모델
 - 여기신호와 성도를 잘 구분해 줌
 - 계산량이 적음
- LPC 분석의 단점
 - 비음, 마찰음 등의 표현이 부정확
 - 잡음환경에 취약
 - 모든 주파수 대역을 동일하게 취급 (음성인지에 관한 지식을 반영못함)

PNU Speech Comm. Lab

LPC 분석

■ LPC 분석 수식

- 현재 신호의 선형 예측

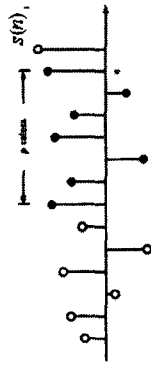
$$\tilde{s}(n) = \sum_{k=1}^p a_k s(n-k)$$

- 예측오차 $e(n) = s(n) - \tilde{s}(n)$

- 단구간 음성 신호 $s(n)$, $n = 0, 1, \dots, N-1$ 가 주어졌을 때,

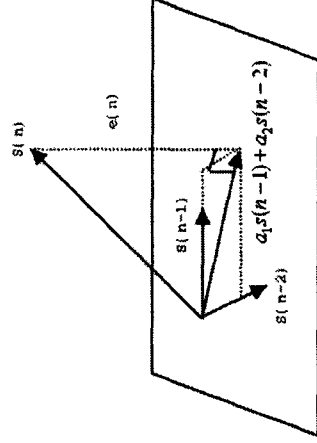
예측오차 에너지 $E = \sum_{n=0}^{N-1} \left[s(n) - \sum_{k=1}^p a_k s(n-k) \right]^2$ 를 최소화하는

$a_k, k = 1, 2, \dots, p$ 를 LPC 계수가 됨



LPC 분석

■ LPC 분석의 기하학적 해석



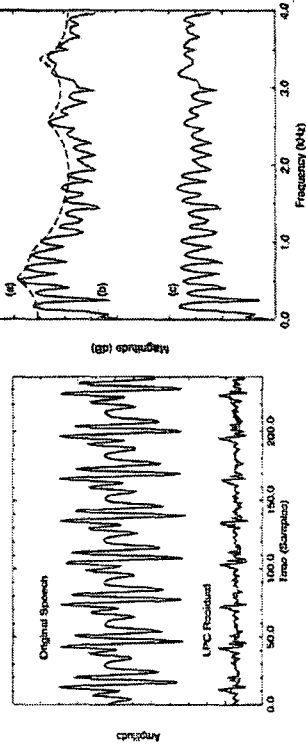
$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k)$ 를 최소화 $\Rightarrow e(n) \perp s(n-k)$

PNU Speech Comm. Lab

LPC 분석

■ LPC 분석 결과의 예

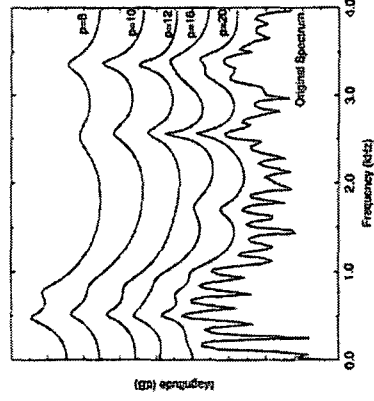
- 원 음성 및 LPC 잔차 신호의 시간/주파수 영역 표현



PNU Speech Comm. Lab

LPC 분석

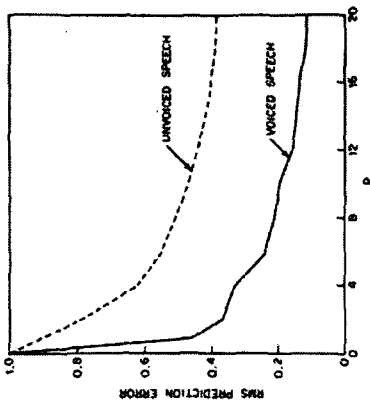
■ LPC 차수에 따른 LPC 스펙트럼 포락선의 모양



PNU Speech Comm. Lab

LPC 분석

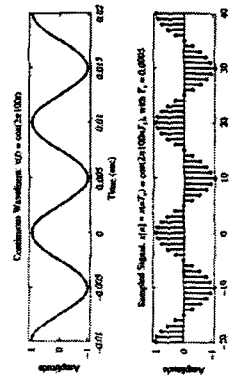
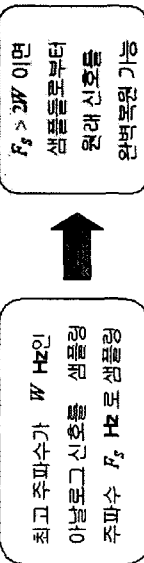
- LPC 차수에 따른 예측 오류의 감소



PNU Speech Comm. Lab

샘플링

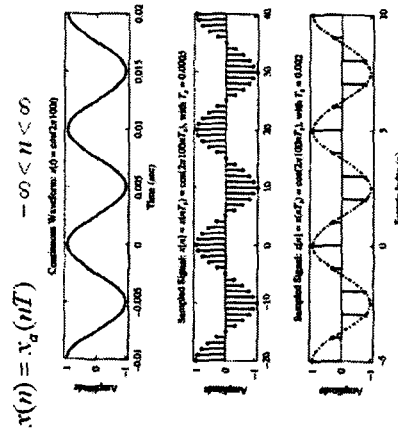
- 샘플링 이론 (Sampling Theorem)



PNU Speech Comm. Lab

샘플링

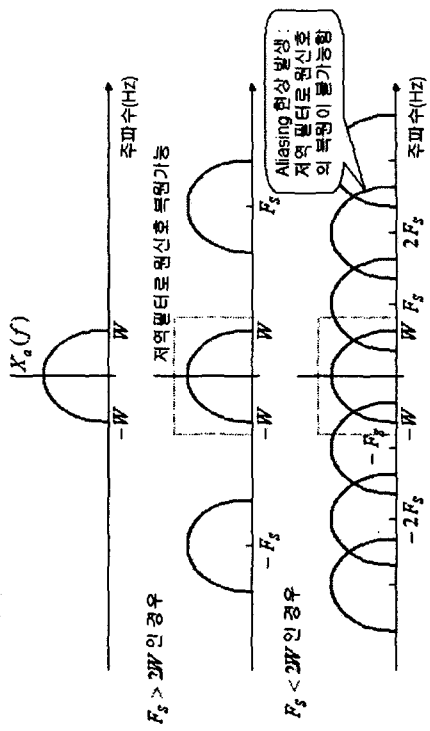
- 샘플링 (Sampling)



PNU Speech Comm. Lab

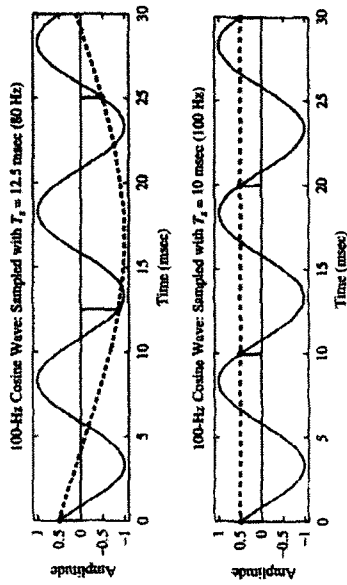
샘플링

- 샘플링의 주파수 영역 해석



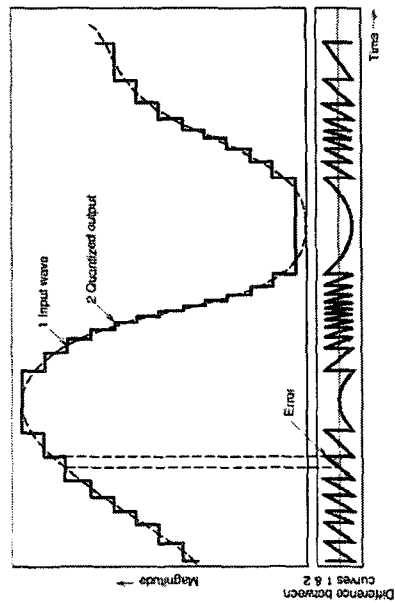
샘플링

■ Aliasing 의 결과 예



양자화 (Quantization)

■ 양자화 과정의 예



샘플링

■ Aliasing 문제의 해결 방안

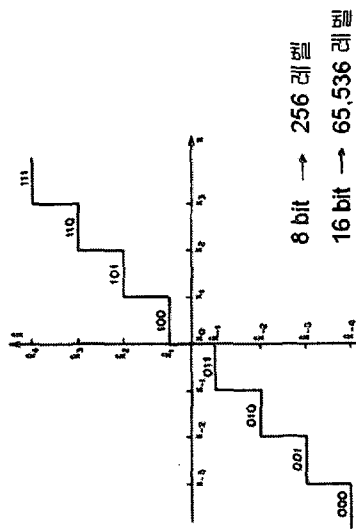
- 샘플링 주파수를 신호의 최대주파수의 2배 이상으로 높인다.
- 샘플링하기 전에 **aliasing**을 막기 위한 저역 필터를 통과시킨다. ("Anti-aliasing 필터" 라고 부름)

■ 샘플링 주파수 선정 예

- 전화망 음성신호: 300-3,400 Hz 대역 \rightarrow 8,000 Hz 샘플링
- CD 오디오: 가청 주파수 대역 20-20,000 Hz \rightarrow 44,100 Hz 샘플링

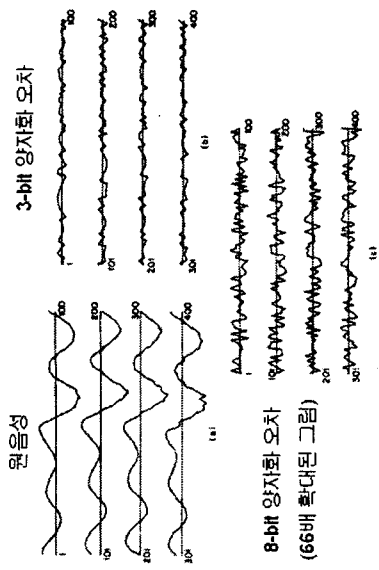
양자화

■ 3-bit (8 레벨) 양자화



양자화

- 양자화 오차 (Quantization Error 또는 Quantization Noise)



PNU Speech Comm. Lab

질문과 답변 (Q & A)

Thank You !

Hyung Soon Kim

Department of Electronics Engineering
Pusan National University
kimhs@pusan.ac.kr

PNU Speech Comm. Lab