

본 문

음성 신호의 부호화

연속되는 시간에 따라 변화하는 연속 함수인 음성 신호는 디지털로 표현하기 위해서 사이언스 샘플 $X_a(nT)$ 를 만들기 위하여 주기적으로 표본화 되어 있다. 그러므로 음성 신호를 디지털(Digital)로 표현하기 위하여 유한한 값(finite value)으로 그들을 양자화할 필요가 있다.

컴퓨터를 사용하여 취급할 경우 다량의 음성 신호의 메모리가 필요하게 된다. 여기서 음성 신호를 능률적인 부호화로 저장할 필요가 있다. 만일 analog 음성 신호가 band limit 하고 sampling rate 가 Nyquist frequency 의 2배이면 저장할 수 있다.]

3. 음성의 통계적 모델

음성 신호를 디지털로 논하기 위해 ergodic random noise로 가정 한다. 이것은 확률적인 관점에서는 어떤 모델을 정의 하는데 유용한 결과를 얻을 수 있다.

$X_a(t)$ 를 Continuous time random process의 sample function이라고 하면 그와 연속적 sample은 discrete-time random process의 continuous sample로 보일 수 있다. analog signal의 특성은 first order probability density $P(x)$ 와 random process의 autocorrelation function으로 다음과 같다.

$$\phi_x(\tau) = E[X_a(t)X_a(t+\tau)] \quad (2.1)$$

이 analog power spectrum은 $\phi_x(\tau)$ 의 Fourier Transform이다.

$$\phi_x(f) = \int_{-\infty}^{\infty} \phi_x(\tau) e^{-j\Omega\tau} d\tau \quad (2.2)$$

Sample random signal $X_n(t)$ 에 대해서 주어진 discrete-time signal은 auto-correlation function을 갖게 된다. 즉

$$\begin{aligned} \phi(m) &= E[X(m)X(m+m)] \quad (2.3) \\ &= \phi_x(mT) \end{aligned}$$

여기서 $\phi(m)$ 이 $\phi_x(\tau)$ 의 한 표본화된 버전이며 $\phi(m)$ 의 power spectrum은

$$\begin{aligned} \phi(e^{j\Omega T}) &= \sum_{m=-\infty}^{\infty} \phi(m) e^{-j\Omega T m} \\ &= \frac{1}{T} \sum_{m=-\infty}^{\infty} \phi_x(\Omega + \frac{2\pi}{T} K) \quad (2.4) \end{aligned}$$

이러한 표본화된 신호의 power spectrum이고 음성의 random process model의 식이다.

4. Instantaneous 양자화

무한 sequence 음성화형 $\{X(n)\}$ 은 적당한 비율로 low-pass filter 하고 표본(sample) 된다. Digital Communication에 대한 연속적인 표본(sample)들을 전송하기 위해서 용한 진폭에 대한 유한 개의 양자화로 표현되어야 한다.

Sampling과 양자화를 포함한 symbol 플런 $\{X(n)\}$ 과 양자화 $C(n)$ 으로 주일한다.

이것은 양자화된 진폭이 $\{X(n)\} = \{Q(n)\}$ 이며 Δ 는 quantizer의 quantization step size이다. 여기서 $C'(n)$ 과 $C(n)$ 이 error가 없다면 출력 $X(n) = X(n)$ 이다. 대부분의 경우 quantized samples은 아진수로 사용된다.

B-bit binary는 2^B 의 quantization level로 나누어 지며 전송할 때 요구되는 정보량이나 digital 값을 저장하는 것은

$$I = B \cdot F(s) = \text{Bit rate in bits per second} \quad (2.5)$$

$F(s)$ 는 sampling rate (sample/sec)

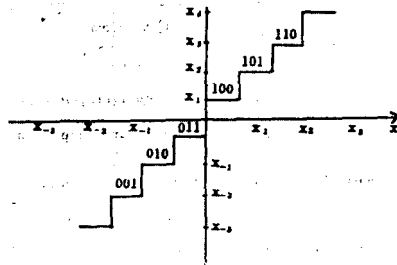
B 는 bits/sample

Sample $\{X(n)\}$ 은 finite의 진폭과 길이에 포함되어 있다.

$$|X(n)| \leq X_{max} \quad (2.6)$$

(그림 2-1)은 8-Level quantizer를 보여주고 있다.

Sampling 진폭은 잘 amplitude의 구간을 포함한 step으로 나누어 의해 양자화 된다.



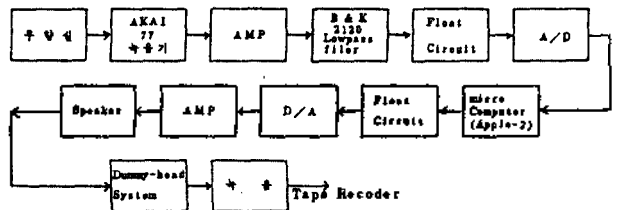
(그림 2-1) 3bit quantizer step size

제 3 장 실험 및 결과고찰

제 1 절. 실험 장치

(그림 3-1)은 본 실험에 사용된 Block-Diagram이며 먼저 무한에서 마이크로폰을 가지고 tape recorder 1 channel로 4명의 화자에 의해 발생된 문장(12일짜리 서울에 간다)이 녹음 된다. 그 녹음된 문장은 증폭을 한 다음 Low pass filter로 통과시키고 A/D converter의 TTL Level 전압을 맞추기 위하여 Floating을 시킨 다음 A/D converter로 (conversion time 30 μ -sec) 8bit sampling한 data를 μ -computer로 받는다.

이때 PCM된 data는 DM, ADM, DPCM, ADPCM 방식들로 시험 데이터를 한 다음 5-bit ADPCM 방식에 의하여 문장을 합성 하였다.



(그림 3-1) ADPCM, DPCM, ADM, DM을 얻기 위해 interface한 Block-Diagram

합성된 음은 재생시간이 1.8초이고 압축율이 45dB인 13×4.4×3.5(m) 크기의 저장체의 장에서 하였다.
제 2 절 결과 및 고찰

1. PCM(pulse code modulation)

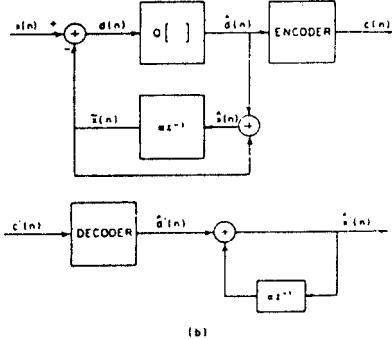
30 μsec conversion time 을 갖는 SNR을 42dB, 8bit 부호화로 행하였다 시뮬레이션 (simulation) 하여 합성된 "길"음이 <그림 3-4>에 나타나 있다.



<그림 3-2> 합성된 "길"음

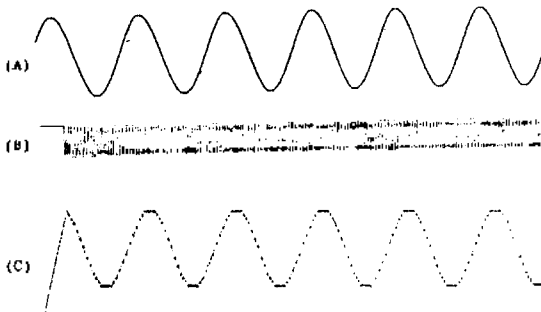
2. 델타 변조(Delta modulation) 방식

앞에서 표시한 것 처럼 PCM방식은 많은 memory가 필요하다 여기서 어느 한정된 memory로 많은 정보를 기록하지 않으면 안된다 그 방법으로 sampling rate를 낮게 하지 않고 그 sampling 주파수에 대하여 bit를 양자화하는 방법이다. 아래 Block Diagram은 본 연구에서 사용한 실험 Block Diagram이며 그 결과 시뮬레이션 (simulation) 하여 얻어진 그래프가 <그림 3-4>에 나타나 있다.



<그림 3-3> 델타 변조 방식의 encoding.(a) Decoding(b) Block Diagram.

△변조 방식은 1bit로 양자된 정보로서 <그림 3-4> (a)에서 1bit의 부호열 Q(n)이 △변조 방식의 (b)로 입력된다.



<그림 3-5> △변조 방식에 의하여 생성된 sin파형

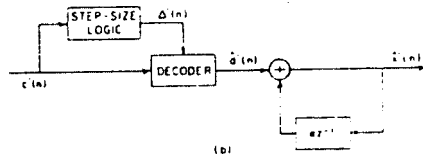
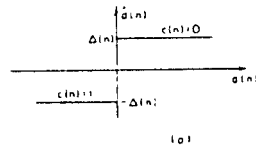
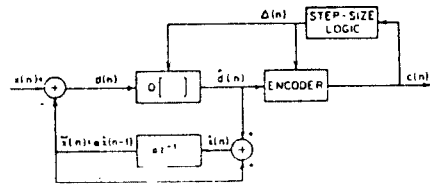
그림(A)은 PCM data의 sin파형이며 (B)는 1bit 부호화를 하였을 때 각 sampling점에서 양자된 △변조 coding data이다. (C)은 이 때 합성된 파형이다.

이 그림에서 첫 부분과 peak to peak에서 overload noise에 의하여 많이 왜곡(distortion)됨을 알 수 있다

3. ADM(adaptive delta modulation)

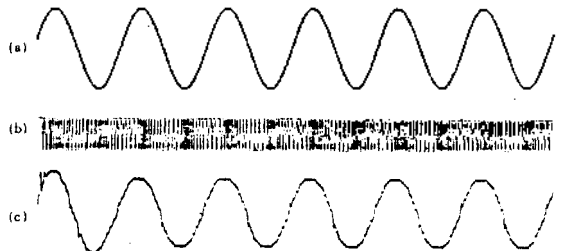
△변조 방식은 진폭의 값 △를 일정하게 하므로써 △의 값을 크게 하면 granular noise가 되고 작게 하면 overload noise가 되므로 그 △의 값을 가변해서 최적 △의 값을 정한다.

<그림 3-6>은 본 연구에서 사용한 Block-Diagram이다.



<그림 3-5> ADM의 Encoding Decoding. Block Diagram

그 결과 얻어진 시뮬레이션 (simulation) 그래프가 아래 그림에 나타나 있고 DM 방식 보다는 performance가 향상되었다



<그림 3-6> 합성된 sin파형

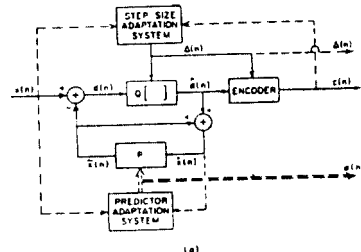
(a)는 PCM signal 이고
(b)는 ADM에 의하여 양자된 coding signal
(c)는 합성된 data이다.

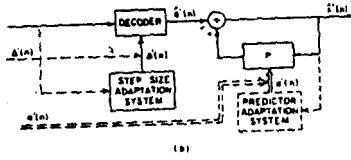
4. ADPCM(Adaptive Differential Pulse Code Modulation)

차별 부호화 방식의 대표적인 ADPCM 방식은 Sample 간 사이의 차분 신호를 적응 부호화 방식이다.

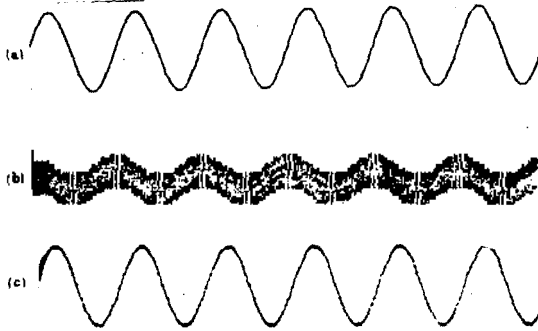
이 방법은 차분 신호값의 양자화폭 Δi를 앞의 Sample값의 결과로부터 정하는 방법이다.

<그림 3-7>은 본 연구에서 사용한 Block Diagram이다.



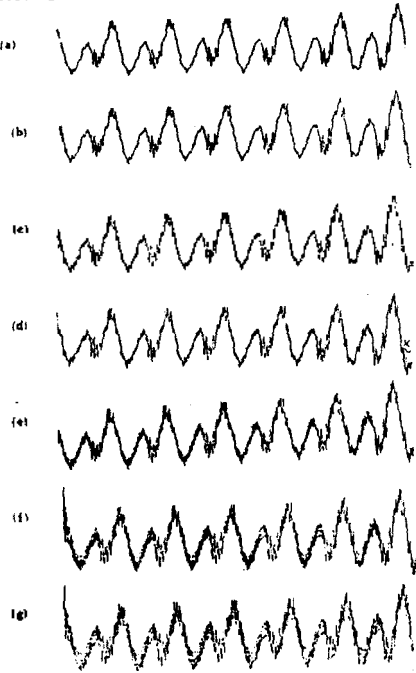


(a) <그림 3-7> ADPCM의 Block Diagram



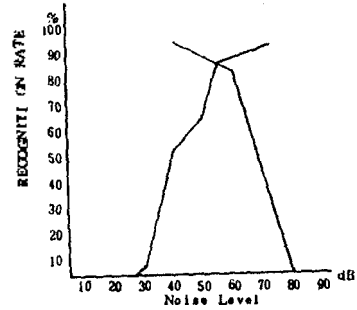
(a) <그림 3-8> ADPCM 방식에 의하여 합성된 sin파

- (a) 8bit, conversion time 30 μ sec로 변환된 PCM sin 합성파
- (b) PCM DATA 1/8로 압축된 ADPCM의 Coding DATA
- (c) 합성된 ADPCM의 sin파선
- (그림 3-8)은 Δ_{min} Δ_{max} 를 각각 1.35 (d) 2.35 (e) 3.35 (f) 5.35 (g) 8.35 (h) 10.35로 하여 얻어진 음성파형 "싱"의 합성파이다.



(그림 3-9) ADPCM 방식에 의한 "싱"의 각 Δ_{min} 과 Δ_{max} 에 Noise에 의한 일그러짐 모양

인식률과 음압과 백색소음과의 관계



(그림 3-10) 인식률과 음압과 White Noise와의 관계
음압과 백색소음과의 관계는 직 접선이 있다. 음압이 56dB의
또 white noise가 60dB일때 두 음압 factor의 가장 좋은 인식
율을 얻을 수 있다.

제 4 장 결 론

ADPCM(Adaptive Differential Pulse Code Modulation) 방식으로 분
할을 합성된 음에 대한 평가법으로서 더머헤드(dummy-head sys-
tem)을 사용하여 음압 레벨(sound pressure level)과 백색소음(whi-
tish noise)에 따라 인공어 "12일파지 서술에 간다"의 화자인식을
조사하여 부음향 factor 사이의 trade off point가 56dB이고 이 점
(point)에서 가장 좋은 청취조건을 얻을 수 있었다.

이렇게 해서 얻은 인식은 음성 이해(speech understanding) 음성
인식(speech identification)을 위한 중요한 파라메터(parameter)
로 사용할 수 있다.

참 고 문 헌

1. J.-L. Flanagan, Voice Man and machines Bell Lab, Murray
hill
2. 이동희, 음향음성학, 1975.
3. 石井聖光, 音聲 昭和 52年
4. 中田初男, 音聲情報處理의 基礎 昭和 56.
5. 中田初男, 音聲デバイス 응용의 實際 昭和 56.
6. 鈴木八十二, ティソル音聲合成器의 設計, 1982.
7. 박순영, 오디오통계에서 정보도 특성에 관한 연구, 1982.
8. A.H.Gray, Quantization and Bit allocation in speech pro-
cessing IEEE Asp 1976.
9. Rabiner, A comparative performance study of served pitch
Detection algorithm IEEE Asp 1976.
10. Nefli, Moller, pitch detection by DATA Reduction IEEE 1976.
11. JAVANT, Digital coding of speech wave forms : PCM, DPCM,
and DM Quantizers, IEE 1984.
12. 윤종원, A study of the comparative performance of adap tive
Delta modulation systems, IEEE 1980.
13. 中田初男, 音聲의 合成과 認識
14. 김순철, 인공어 음성의 분석과 자동인식에 관한 연구
15. 차일환, '개인 식별에 관한 연구', 1982.