파형 부호화 방식에 의한 음성신호 압축에 관한 연구

°안 동 순, 조 병 모, 차 일 환
연세대학교 전자공학과

The Study on the Speech Signal Compression by the Coding Method

Ann, Dong Soon, Cho, Byoung Mo, Cha, Il Whan
Yonsei University Dept. of Electronic Eng.

Abstract

Speech sound includes personality, move and feeling. Respectively, that is different from man and woman, youn and old.

In this paper, sentence … Sip E Il Ka Gi Seo Ul E Gan Da … A/D conversion time is 50㎲sec.Data are abtained using the microcomputer and compressed by ADPCM Rate of compression is 1/8.

Data compressed by ADPCM are synthesized and compared to the original sound.

Rate of speech identification is analysed using the sound pressure, white noise.

Coding of ADPCM is done for 5 bit.

As the result of fixing starting voltage by 2.6V.

It is acertained that variable value increase in initial speech signal and then process is made by minimum value " 3 ".

From the result of processing, synthesized sound is almost equal to original sound.

Minimum values cause distorition, Dummy Head System is used in this experiment.

요 약

음성은 말하는 사람의 개성, 감정, 정서를 포함하고 있으며 또한 남녀 노소에 따라서 다른 구별되는 소리를 듣게된다.

본 연구에서는 4인의 화자에 의해서 발성된 3초의 (12일까지 서울에 간다) 시간 길이를 갖는 문장을 사용하였다. 변화시간(conversion time)은 50마이크로-초(㎲-sec)로 하였으며 8-bit 마이크로-프로세서(㎲-processor)로 데이타(data)를 받은 다음 ADPCM-method에 의하여 디지탈 음성신호를 처리하였다. Digital 음성신호의 시작점(starting point)를 2.6 volt로 잡아준 결과 ADPCM의 최소화값(minimum value)에 따라서 원래의 음성파형은 초기치에서 변화량(variable volue)이 크게 됨을 확인한 다음 최소화값을 시뮬레이션을 통하여 최적 스텝사이즈(△step size)를 △min = 3으로 처리 하였다.

ADPCM 방식으로 합성된 음은 그 평가방법으로서 더미헤드 시스템(dummy head)을 사용하여 음질을 평가 하였다.

1. 서 론

기계와 인간과의 음성 통신은 음성 파형의 물리적 특성에 관한 과학적인 연구가 활발해 지면서 시작되었다. 1950년 Shannon이 정보이론을 발표하면서 음성의 정보전달에 관한 연구가 활발해지기 시작했다.

그후 델타변조(Delta modulation)에 의한 합성 Voice Coder형 음성 합성기 등이 발표되었고 pollack(1954)에 의한 청각에 의한 인식이 발표되었다. 1960년대의 computer를 이용한 델타 합성에 관한 연구, voice printer 규칙합성 수법, 음성 응답장치 등이 발명되었다. 1970년과 1980년도 사이에는 음성파형을 응용하여 학습의 수단이나 연구의 목적 등에 이용하였다.

본 연구에서는 한국어의 문장음을 ㎲-processor를 이용하여 합성한 다음 그에 대한 평가방법으로써 dummy-head를 이용하여 음향특성이 음질에 끼치는 영향을 고려하므로서 음성식별, 음성이해를 위한 자료로서 사용될 수 있다.

2. 본 론

2-1. 음성 합성방식

음성 합성 방식은 파형부호와 방식, 스펙트럼 부호와 방식, 규칙 합성방식으로 나눌 수 있다. 파형부호와 방식은 파라메타가 파형이며 PCM, APCM, DM, ADM, CSM, DPCM, ADPCM 방식등으로 나누어진다. 스펙트럼 부호화 방식은 PARCOR, LSP, Vocoder 방식등으로 나누어진다.

규칙 합성방식은 합성 음소 방식과 생성규칙에 의한 Spectrum Analog 방식이다. 이러한 방식들은 음성을 합성할 때의 필요한 정보량이 아주 작아진다는 잇점이 있다.

2-2. 음성 합성에 대한 이론

2-2-1. 음성의 통계적 모델

음성 신호 파형은 ergodic random process로 가정하고 처리한다.

random process의 anto correlation은

$$\phi_a(\tau) = E[\, X_a(t)\, X_a(t+\tau)\,] \quad \cdots\cdots\cdots\cdots \langle 2\text{-}1 \rangle$$

analog power spectrum은 $\phi_a(\tau)$의 Fourier trans form이므로

$$\phi_a(\Omega) = \int_{-\infty}^{\infty} \phi_a(\tau) e^{-j} \, d\tau \quad \cdots\cdots\cdots\cdots \langle 2\text{-}2 \rangle$$

Sampling random signal $X_a(t)$의 discrete − time은

$$\phi(m) = E[\, X_a(nT)\, X_a(nT+mT)\,] = \phi_a(mT) \quad \cdots \langle 2\text{-}3 \rangle$$

$\phi(m)$의 power spectrum은

$$\phi(e^{jT}) = \sum_{m=-\infty}^{\infty} \phi(m) e^{-j\,Tm}$$

$$= \frac{1}{T} \sum_{k=-\infty}^{\infty} \phi_a\left(\Omega + \frac{2\pi}{T} k\right) \quad \cdots\cdots\cdots\cdots \langle 2\text{-}4 \rangle$$

이 식은 음성신호의 randnm process model로 표현되고 있으며 원래 음성신호의 아나로그신호가 sampled signal의 power spectrun으로 표시된 것이다.

음성신호로부터 Probability density function을 추정하면 gamma distribution과 Laplacian density는

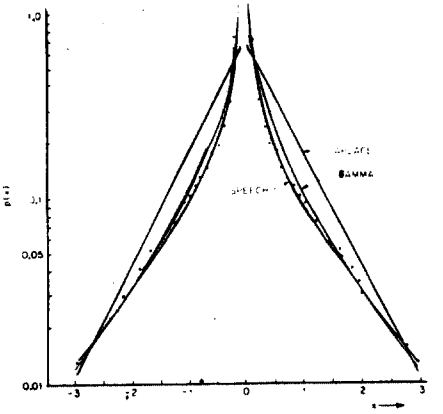
$$P(x) = \frac{1}{\sqrt{2}\,\sigma_x} e^{-\frac{\sqrt{2}\,x}{\sigma_x}}$$

$$P(x) = \left[\frac{\sqrt{3}}{8\pi\sigma_x x}\right]^{\frac{1}{2}} e^{-\frac{\sqrt{3}\,x}{2\sigma_x}} \quad \cdots\cdots\cdots\cdots \langle 2\text{-}5 \rangle$$

시스템에서 quantizer의 입력은

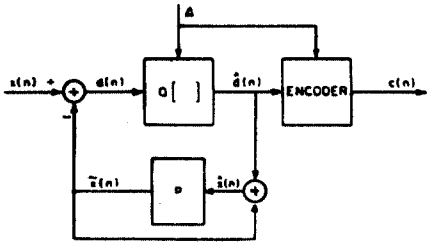
$$d(n) = x(n) - \tilde{x}(n) \quad \cdots\cdots\cdots\cdots \langle 2\text{-}6 \rangle$$

x(n) : unquantized 입력표본치

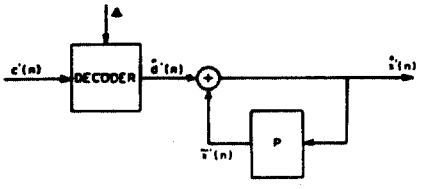


〈그림 2-1〉 음성신호의 실제 진폭의 분포를 Gamma pdf와 Laplacian pdf와 비교한 그림

2-2-2. Differential quantization의 일반적 이론



〈그림 2-2〉 differentlal quantization coder



〈그림 2-3〉 differential quantization decoder

$\tilde{x}(n)$: 입력신호의 추정치, 예견치

예견치는 예견계수 system의 P의 계수이다.

d(n)은 예견되는 error인 P의 출력이다.

양자화된 차분신호는

$$d(n) = d(n) + e(n) \quad \cdots\cdots\cdots\cdots \langle 2\text{-}7 \rangle$$

e(n) : 양자화된 error

예견치 x(n)에 더해진 양자화된 차분신호는

$$x(n) = \tilde{x}(n) + d(n) \quad \cdots\cdots\cdots\cdots \langle 2\text{-}8 \rangle$$

식으로 부터

$$x(n) = x(n) + e(n) \quad \cdots\cdots\cdots\cdots \langle 2\text{-}9 \rangle$$

여기서 양자화된 차분신호가 전송을 위해서 memory에 저장된다.

$$SNR = \frac{E[x^2(n)]}{E[e^2(n)]} = \frac{\sigma_x^2}{\sigma_e^2} \quad \cdots\cdots\cdots\cdots \langle 2\text{-}10 \rangle$$

이것은

$$SNR = \frac{\sigma_x^2}{\sigma_d^2} \cdot \frac{\sigma_d^2}{\sigma_e^2} = Gp \cdot SNR_Q \quad \cdots\cdots \langle 2\text{-}11 \rangle$$

$$SNR_Q = \frac{\sigma_d^2}{\sigma_e^2} \quad \cdots\cdots \langle 2\text{-}12 \rangle$$

이것이 quantizer의 S/N rate이다.

$$Gp = \frac{\sigma_x^2}{\sigma_d^2} \quad \cdots\cdots \langle 2\text{-}13 \rangle$$

를 차분에 기인된 SNR의 gain을 표시 σ_x^2이 고정된 것으로 주어지면 Gp는 분모를 최소화 함으로서 Gp의 값을 최대화 할 수 있다.

즉 prediction error의 변량을 최소화 함으로서 gain을 증가시킬 수 있다.

$\tilde{x}(n)$은 과거 양자화된 값의 선형 결합이다.

$$\tilde{x}(n) = \sum_{k=1}^{p} \alpha_k x(n-k) \quad \cdots\cdots \langle 2\text{-}14 \rangle$$

이 예견치는 finite impulse response filter의 출력이다.

$$P(z) = \sum_{k=1}^{p} \alpha_k Z^{-k} \quad \cdots\cdots \langle 2\text{-}15 \rangle$$

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} \alpha_k Z^{-k}} \quad \cdots\cdots \langle 2\text{-}16 \rangle$$

이 system function의 출력이 재건된 impulse 응답출력이다.

이 system의 prediction error의 variance는

$$\sigma_d^2 = E[d^2(n)] = E[(x(n) - \tilde{x}(n))^2]$$

$$= E[(x(n) - \sum_{k=1}^{p} \alpha_k x(n-k))^2]$$

$$= E\{[x(n) - \sum_{k=1}^{p} \alpha_k x(n-k) - \sum_{k=1}^{p} \alpha_k e(n-k)]^2\}$$

$$\cdots\cdots \langle 2\text{-}17 \rangle$$

predictor의 계수 $\{\alpha_j\}$, $1 \le j \le p$를 선택하기 위해서 σ_d^2을 minimize하고 각 parameter에 대해서 σ_d^2이 나누어지고 또 그 미분은 Zero와 같다.

$$\frac{\partial \sigma_d^2}{\partial \alpha_j} = -2E[(x(n) - \sum_{k=1}^{p} \alpha_k (x(n-k) + e(n-k))] \cdot [x(n-j) + e(n-j)]] = 0$$

$$1 \le j \le P \quad \cdots\cdots \langle 2\text{-}18 \rangle$$

이 식을 보다 더 compact한 식으로 쓰면,

$$E[(x(n) - \tilde{x}(n)\, x(n-j)] = E[d(n)\, x(n-j)] = 0,$$

$$1 \le j \le P \quad \cdots\cdots \langle 2\text{-}19 \rangle$$

식 (2-18) 식은 P방정식으로 확장하면

$$E[x(n-j)x(n)] + E[e(n-j)x(n)]$$

$$= \sum_{k=1}^{p} \alpha_k E[x(n-j)x(n-k)]$$

$$+ \sum_{k=1}^{p} \alpha_k E[e(n-j)x(n-k)]$$

$$+ \sum_{k=1}^{p} \alpha_k E[x(n-j)e(n-k)]$$

$$+ \sum_{k=1}^{p} \alpha_k E[e(n-j)e(n-k)] \quad \cdots\cdots \langle 2\text{-}20 \rangle$$

여기서 $1 \le j \le P$이다.

e(n)이 stationary white noise sequence이고 e(n)이 x(n)과 uncorrelated하다고 가정할 수 있다.

$$E[x(n-j)e(n-k)] = 0, \quad \text{for all } n, j \text{ and } k \quad \cdots\cdots \langle 2\text{-}21 \rangle$$

$$E[e(n-j)e(n-k)] = \sigma_e \delta(j-k)$$

윗식을 사용하면 식 〈 2-15 〉은

$$\phi(j) = \sum_{k=1}^{p} \alpha_k [\phi(j-k) + \sigma_e^2 \delta(j-k)], \quad 1 \le j \le P \cdots \langle 2\text{-}22 \rangle$$

$\phi(j)$는 x(n)의 auto correlation function이다. σ_x^2에 의한 윗 방정식과 normalized된 auto-correlation을 다음과 같이 정의한다.

$$\rho(j) = \frac{\phi(j)}{\sigma_x^2} \quad \cdots\cdots \langle 2\text{-}23 \rangle$$

matrix형태로 식 〈 2-15 〉을 쓰면

$$\rho = c\alpha \quad \cdots\cdots \langle 2\text{-}24 \rangle$$

$$\rho = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(P) \end{bmatrix} \quad \cdots\cdots \langle 2\text{-}25 \rangle$$

$$c = \begin{bmatrix} (1 + \frac{1}{SNR}) & \rho(1) & \cdots\cdots & \rho(p-1) \\ \rho(1) & (1 + \frac{1}{SNR}) & \cdots\cdots & \rho(p-2) \\ \vdots & & & \\ \rho(p-1) & \rho(p-2) & \cdots\cdots & (1 + \frac{1}{SNR}) \end{bmatrix} \quad \cdots\cdots \langle 2\text{-}26 \rangle$$

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} \quad \cdots\cdots \langle 2\text{-}27 \rangle$$

$$SNR = \frac{\sigma_x^2}{\sigma_{e2}}$$

optimum predictor coeffcient의 vector는 윗식의 해에서 얻어진다.

$$\alpha = c^{-1}\rho$$

1/SNR term을 생략하고 p = 1로 가정하면

$$\alpha_1 = \frac{\rho(1)}{1 + \frac{1}{SNR}} \quad \cdots\cdots \langle 2\text{-}28 \rangle$$

이 식은 $\alpha_1 < \rho_1$임을 나타낸다.

optimum Gp를 얻기 위하여

$$\sigma_d^2 = E[\{x(n) - \tilde{x}(n)\}(x(n) - \tilde{x}(n))]$$

$$= E[(x(n) - \tilde{x}(n))\, x(n)] - E[(x(n) - \tilde{x}(n))\, \tilde{x}(n)]$$

$$\cdots\cdots \langle 2\text{-}29 \rangle$$

예견치는 predictor 계수와 uncorrelated하다.

$$\sigma_d^2 = E[(x(n) - \tilde{x}(n))x(n)]$$

$$= E[x^2(n) - E[\sum_{k=1}^{p} \alpha_k(x(n-k) + e(n-k))x(n)]$$

..... 〈 2-30 〉

uncorrelat 한 signal 과 noise 에 대해서

$$\sigma_d^2 = \sigma_x^2 - \sum_{k=1}^{p} \alpha_k \phi(k)$$

$$= \sigma_x^2[1 - \sum_{k=1}^{p} \alpha_k \rho(k)]$$

식 2-8 식으로 부터

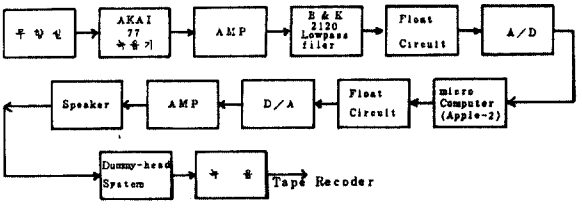
$$(Gp)_{opt} = \frac{1}{1-\alpha_1 \rho(1)}$$ ..... 〈 2-31 〉

optimal Gp 값이 구해진다.

3. 실험 및 결과 고찰

3-1. 실험장치

〈그림 3-1〉은 본 실험에서 사용한 Block-Diagram 이다. 무향실에서 마이크로폰을 가지고 tape recoder 에 1 channel 로 4 명의 화자에 의해서 발성된 문장 (12 일까지 서울에 간다) 을 녹음 시킨다. 그 녹음된 문장을 증폭한 다음 Low pass ficter 로 통과시키고 TTL Level 전압을 맞춘다음 conversion time 50 μ - sec) 8bit 로 sampling 한 Data 를 μ - computer 로 받는다.



〈그림 3-1〉 ADPCM, DPCM, ADM, DM을 얻기 위하여 Interface 한 Block - Diagram.

이 PCM data 는 ADPCM방식에 의하여 문장음 (26 k byte) 을 합성하였다.

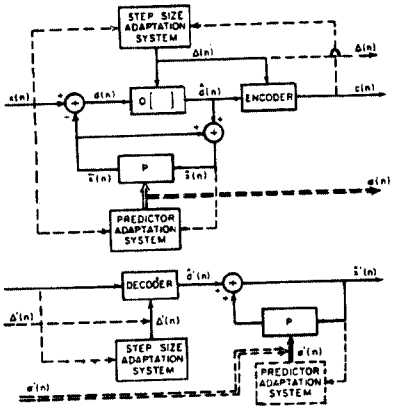
합성된 음은 dummy-head, headphone system 을 통하여 녹음하였다.

녹음조건은 음압과 소음을 각각 20dB, 40 dB, 60 dB, 80 dB로 10 m 멀어진 거리에서 녹음하였다. 피실험자로는 " ADPCM합성음의 인식률 시험표를 만든 다음 24 ~ 30 세의 정상적 청각을 가진 10 인의 대학원생으로 치루었다.

3-2. 결과 및 고찰

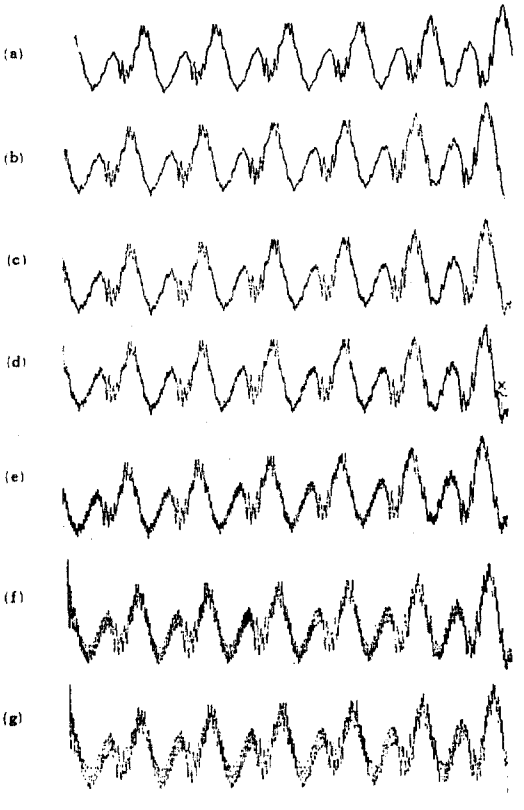
파형 부호화 방식의 대표적인 ADPCM방식은 Sample 값 사이의 차분 신호 적응 부호화 방식이다.

이 방법은 차분 신호값의 양자화폭 △i를 앞의 Sample 값의 결과로 부터 정하는 방법이다.

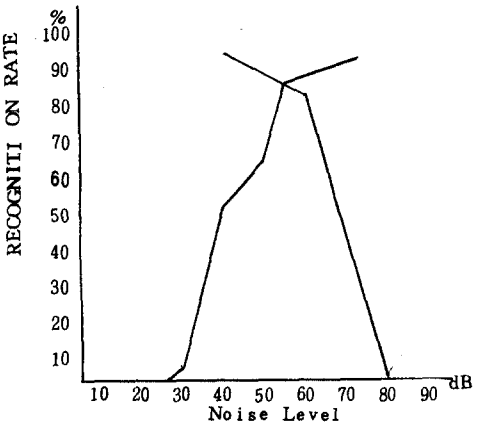


〈그림 3-2〉 ADPCM의 Block-Diagram

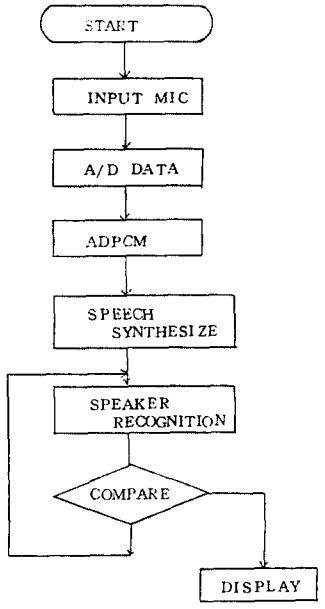
여기서 step siz 는 Jayant 가 제시한 최적 △값을 사용하였으며 △min volue 은 simulation 하여 △min = 3 으로 하였다.



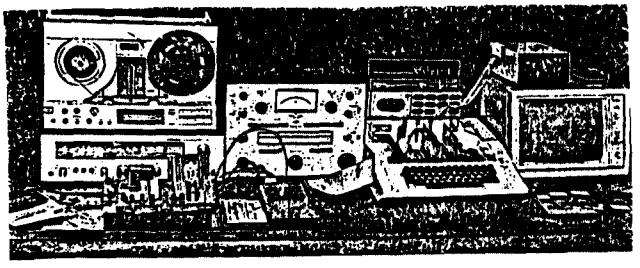
〈그림 3-3〉 ADPCM방식에 의한 합성음 " 십 " 음의 각 △min 과 △max 의 Noise 에 의한 일그러진 모양



〈그림 3-4〉 인식률에 대한 음압과 white-Noise 와의 관계



SYSTEM CONTROL FLOW CHART



〈그림 3-5〉 전체 실험장치도

결 론

음성 신호의 digital coding 중 ADPCM방식으로 문장음을 합성하였다. 그 합성음은 두 음향 factor 사이의 trade off point 가 56dB이고 이점에서 가장 좋은 청취조건을 얻었다.

이렇게 해서 얻은 인식은 speech under standing과 음성인식 (speech identification) 과 음성 intelligence를 향상시키는 parameter로 사용될 수 있다.

참 고 문 헌

1. J.L. Flana GAL, Voice Man and machines Bell Lab, Murroy hill.
2. 石井聖光, 音聲 昭和 52年.
3. 中田和男, 音聲情報處理의 基礎 昭和56.
4. 中田和男, 音聲デベイス活用의 實際 昭和56.
5. 鈴木八十二, テイゾル音聲合成器의 設計, 1982.
6. 박순영, 오디토리움에서 명료도 특성에 관한 연구, 1982.
7. A.H.Gray, Quantization and Bit allocation in speech processing IEEE Assp 1976.
8. JAYANT, Digital coding of sheech wave forms: PCM, DPCM, and DM Quantizers, IEEE 1984.
9. 은종관, A study of the comparative performance of adap tive Delta modulation systems, IEEE 1980.
10. L.R. Rabiner/R.W. Schafer, Digital Processing of Speech Signals 1982.
11. OSMU Fujimura, Kobayasi, Nasalization of vowel in Radiation to Nasals, J.A.S.A. 1958.
12. K.N. Stevens, Speake authentification and Identification Acomparison of Spectrographic and Auditory Presentations of Speech Material, J.A.S.A. 1968.
13. OSMU Fujimura, Nasalization of vowels in Relation to Nasals J.A.S.A. 1958.
14. 차일환, '개인 식별에 관한 연구', 1982.