

음성 신호의 디지털화와 대역폭 축소의 방법에 관하여 [II]—Vocoding

(On Speech Digitization and Bandwidth Compression Techniques[II]—Vocoding)

股 鍾 官*
(Un, Chong Kwan)

要 約

本 論文은 음성신호의 디지털화와 대역폭 축소에 관한 一部¹⁾에 이은 二部 論文이다. 몇가지 近來에 開發된 Vocoding 方法, 즉 linear predictive coding (LPC), formant vocoding, residual excited linear prediction (REL P) vocoding, 그리고 adaptive predictive coding(APC)에 관하여 論하였다. 本 論文에서는 音聲傳送에 있어서의 대역 制限 方法 中 지금 가장 효과가 있는 LPC 方法을 重點적으로 提議하였다. 또한 현재 처하고 있는 문제점들과 해결책을 托의하였다.

Abstract

This paper is a sequel of the previous paper¹⁾ on speech digitization and bandwidth compression techniques. Several recently developed vocoding techniques, that is, linear predictive coding(LPC), formant vocoding, residual excited linear prediction(RELP) vocoding, and adaptive predictive coding(APC) are discussed. Throughout the paper emphasis is placed on the LPC approach that is presently the most promising technique in speech compression. In addition, current problems and possible solutions are discussed.

1. 序 論

音聲 通信에서 대역폭을 가장 効果적으로 줄이는 方法은 音聲 信號를 直接量子化하지 않고, 音聲 信號의 重要한 性格탄을 一定한 時間(10 ms~30 ms)마다 抽出하여 送信하고 受信機에서는 이 情報를 利用하여 音聲을 다시 構成하는 方法이다. 이 方法을 쓰는 音聲 信號의 符號器를 vocoder, 또는 分解-合成시스템(analysis-synthesis system)이라고 한다.

이 論文은 音聲 信號의 디지털화와 대역폭 縮小方法에 關한 第 II 部 論文으로서 앞서 第 I 部¹⁾에서 取扱한 ADPCM 과 ADM 에 이어 vocoding 方法이 論議된다. 特別 지금까지 알려진 vocoder 中 가장 音質이 좋고 又

즈음 急速히 發達되는 디지털技術과 關聯하여 많은 利點이 있는 linear predictive coding(LPC)^{2,3)} 方法을 重點의으로 論한다. Vocoding 方法에는 傳送速度를 800 bps 까지 내릴 수 있는 formant vocoder⁴⁾, 2.4~4.8 kbps 程度의 LPC vocoder,^{2,5~7)} 시스템이 雜音이나 周圍環境에 影響을 받지 않는 6~9.6 kbps 의 residual excited linear prediction(RELP) vocoder,⁷⁾ 그리고 8~16 kbps 의 adaptive predictive coder (APC)^{8,10)} 등이 있다. 本 論文에서는 LPC 의 原理를 利用하여 위에서 列擧한 여러 가지 vocoder 를 어떻게 만들수 있나를 討議하고 現在까지 音聲 信號의 디지털化 研究에서 題起된 代表의인 問題點들을 論議하겠다.

Vocoder 의 原理를 理解하기 爲해서는 音聲이 어떻게 發生되고, 이것을 디지털信號 處理에서 어떻게 model 化하나를 알아야 하므로 먼저 이에 關하여 簡單히 記述한다.

* 正會員, 한국과학원 전기 및 전자공학과
(Dept. of Electrical Science KAIS)
接受日字: 1978年 9月 4日

2. 音聲 發生 Model^{11,12)}

音聲은 크게 나누어 有聲音(voiced sound)과 無聲音(unvoiced sound)의 두 種類가 있다. 有聲音은 聲帶를 振動시켜서 나오는 air pressure의 準周期的인 pulse가 聲道(vocal tract)를 刺戟시켜 나온다. 한편 無聲音은 聲道를 壓縮시켜 만드는 random noise와 같은 性格의 air turbulence에 依해서 發生된다. 有聲音과 無聲音의 代表的인 波形과 그의 spectrum이 그림 1과 2에 各各 그려져 있다. 人間の 發聲器官은 시스템 理論的인 關點에서 보면 聲道는 時變 線型 filter와 같고 air pressure의 pulse나 air turbulence는 이 filter를 excite시키는 source 信號가 된다. Vocal tract filter를 形成하기 爲해서는 音聲을 分析하여 必要한 parameter들을 求해야 되는 데, 分析方法에 따라 vocoding 方法이 달라지게 된다. 낮은 傳送速度의 vocoder나 디지털信號 處理에 使用되는 音聲 發生 model이 그림 3에 그려져 있다. Excitation source를 modeling하기 爲하여 音聲을 實際로 發生시키는 方法이 여기에서도 使用되는데 有聲音은 pulse generator를 使用하

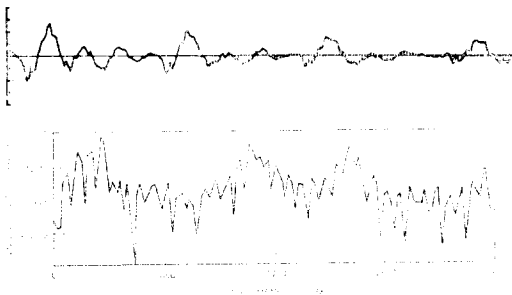


그림 1. 유성음 음성파형(上)과 FFT spectrum(下)
Fig 1. Temporal(top) and Spectral (bottom) Waveforms of Voiced Sound

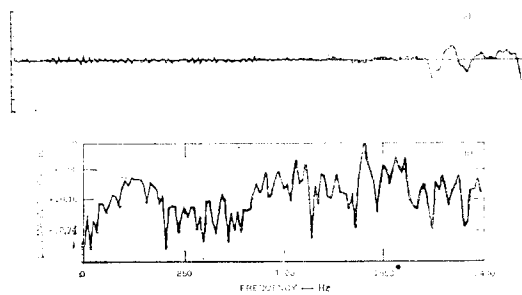


그림 2. 무성음 음성파형(上)과 FFT Spectrum(下)
Fig. 2. Temporal(top) and spectral(bottom) waveforms of unvoiced sound.

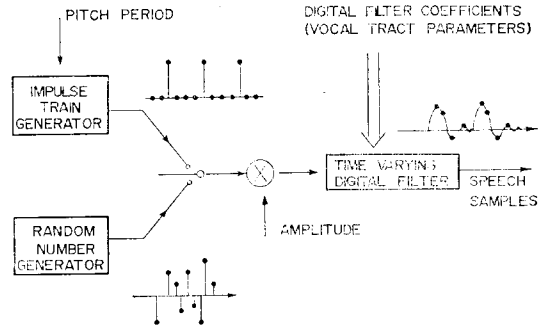


그림 3. 디지털 신호처리에 있어서의 음성발생 model
Fig. 3. Model for production of speech in digital signal processing

고, 無聲音은 random noise generator를 使用한다. 따라서 두 source 중 하나를 使用하기 爲해서는 먼저 音聲波型을 分析하여 有聲音인지 無聲音인지를 決定 [Voiced/Unvoiced (V/UV) decision]해야 하고, 有聲音인 境遇에는 周期(pitch period)를 決定해야 한다.

音聲波型은 그림 1에서 보여진 바와 같이 一般的으로 時間에 따라 變하지만 淸淸히 波形이 變하는 性格을 가지고 있다. 따라서 音聲 信號를 짧게 部分(10 ms ~ 30 ms)으로 나누면 나누어진 各各의 部分은 거의 時間에 따라 變하지 않는 信號처럼 보이게 된다. 그러므로 vocoder에서는 音聲 信號를 quasi-stationary 信號로 간주하고 時間波形(time-waveform)을 約 20 ms 程度로 나누어 每 20 ms의 音聲을 分析하여 filter의 係數와 pitch에 關한 情報(즉 V/UV 決定 및 周期)를 얻게 된다.

3. Linear Predictive Coding(LPC)

音聲을 周波數領域(frequency domain)에서 分析하였을 때 音聲 spectrum은 急히 變하는 纖細한 spectrum 構造가 淸淸히 變하는 包絡線(envelope)에 依해 變調된 것처럼 나타난다. (그림 1 참조). 그림에서 이 纖細한 spectrum 構造는 excitation 信號에 依한 것인데 spectral harmonics는 音聲의 周期(pitch 또는 fundamental frequency)에 該當되고 包絡線의 構造는 vocal tract filter의 spectrum에 該當되는 것이다. LPC는 이 包絡線 構造를 推定하는 한 方法으로서 지금까지 研究된 여러 가지 方法 중 가장 正確하고 效果의으로 計算할 수 있는 것으로 알려져 있다.

音聲을 LPC 方法으로 分析하여 聲道 filter의 係數들을 얻으려면 먼저 filter의 전달함수의 일반적인 model이 必要한데, 그것은 pole만을 갖는 digital

recursive linear filter 로써 表現할 수 있다.

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^p a(i)z^{-i}} \quad (1)$$

여기에서 $a(i)$ 는 豫測係數로서 바로 聲道의 特性을 나타내는 parameter 가 된다. 위의 聲道의 filter 전달 함수를 보다 더 正確하게 나타내려면 pole 뿐만 아니라 zero 도 必要하다. 特히 한글의 “口”이나 “ㄴ”같은 鼻音(nasals)은 聲道의 side branch 인 nasal cavity를 통해서 나오게 되는 데 이 경우에는 zero 들이 더욱 必要하다. 그러나 pole 은 물론 zero 들까지 전달함수에 包含시키면 filter 의 係數를 求하는 것이 대단히 複雜하게 된다. 이 點에 關해서는 나중에 다시 論하겠다.

LPC 分析에서는 N sample 의 입력 音聲 sequence $\{s(n)\}$ 이 주어졌을 때 式 (1)의 inverse filter(즉 $1/H(z)$ 또는 $A(z)$)의 出力 에너지를 最少化 시키는 係數 $\{a(i)\}$ 의 값들을 어떻게 求하는 가가 問題의 焦點이 된다. LPC 는 分析器(analyzer)의 入力 音聲 信號를 어떻게 實義하고 豫測誤差(prediction error)를 어떻게 最少化하는 가에 따라 covariance 方法⁵⁾과 autocorrelation 方法^{6, 12)}의 두 가지로 區分된다. Covariance 方法은 工部에서 取扱된 ADPCM 의 適應豫測器의 豫測係數를 求하는 方法과 같은 것으로, 이미 討議 되었으므로 다시 反復하지 않고, 여기에서는 LPC vocoder 에서 實際로 많이 쓰고 있는 autocorrelation 方法만 論하겠다.

Autocorrelation 方法의 特徵은 入力信號를 window function 을 써서 잘라 豫測誤差信號를 發生시키고 誤差 에너지를 時間(또는 sample 數)에 無關하게 最少化시키는 點이다. window function 은 여러 種類가 있지만 主로 다음의 Hamming window 를 많이 쓴다.

$$W_n = 0.54 - 0.46 \cdot \cos \frac{2\pi n}{N-1}, \quad n=0, \dots, N-1 \quad (2)$$

여기에서 N 은 window 의 길이(즉 sample 數)이다. window 된 音聲 信號를 가지고 誤差信號, 즉

$$e(k) = s(k) - \sum_{i=1}^p a(i)s(k-i) \quad (3)$$

를 發生시켜 [本論文 I部 참조] truncate 시키지 않고 에너지를 最少化시키면 다음과 같은 P simultaneous autocorrelation 의 式을 얻게 된다.

$$\sum_{k=1}^p a(k)R(|i-k|) = R(i), \quad i=1, 2, \dots, P \quad (4)$$

이때 最少化된 誤差 에너지를

$$E_{min} = R(0) - \sum_{k=1}^p a(k)R(k) \quad (5)$$

이다. 여기서

$$R(i) = \sum_{n=0}^{N-1-i} s(n)s(n+|i|), \quad R(i) = R(-i)$$

이다. 式 (4)를 行列式으로 쓰면 다음과 같다.

$$\begin{pmatrix} R(0) & R(1) & R(2) & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & R(p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & \dots & \dots & R(0) \end{pmatrix} \begin{pmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{pmatrix} = \begin{pmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{pmatrix} \quad (6)$$

위의 $p \times p$ 行列式을 autocorrelation 行列이라 하는 데 이 行列의 特性은 對稱的이고 diagonal element 가 모두 같다. 이러한 行列을 Toeplitz matrix 라고 하며, inverse matrix 를 求하는데 있어 covariance matrix 의 inversion 을 위한 Cholesky 方法보다 훨씬 計算이 적게 所要되는 改良된 Levinson 의 方法¹⁴⁾을 쓴다.

Autocorrelation 方法의 큰 利點은 filter 의 係數 $\{a(i)\}$ 를 Toeplitz matrix 를 使用하여 求했을, 境遇 filter 의 安定度(stability)가 언제나 保障된다는 點이다. 한편 Covariance 方法을 利用하여 求한 係數로 구성한 filter 는 때로는 不安定한데, 이와 같은 filter 는 人爲的으로 安定하게 만들어 주어야 한다.¹⁵⁾ 또 한가지 autocorrelation 方法의 長點은 filter 의 係數들을 送信하는데 있어서 $\{a(i)\}$ 의 變形인 反射係數(reflection coefficients) $\{k(i)\}$ 를 쓸 수 있다는 점이다. $\{k(i)\}$ 는 $\{a(i)\}$ 와 달리 dynamic range 가 작아서 ($|k(i)| \leq 1$) 符號化하는 데 bit 의 數를 줄일 수 있고 filter 의 安定度를 쉽게 檢査할 수 있다.

4. LPC 의 原理를 利用한 Vocoder 의 例

앞서 序論에서 言及한 바와 같이 LPC 의 原理를 利用한 vocoding 의 方法이 여러가지 있는데 여기에서는 代表的인 例로 傳送速度가 2.4 kbps 인 LPC vocoder 와 1,200 bps 인 formant vocoder, 6~9.6 kbps 인 RELP vocoder, 그리고 8~16 kbps 인 APC 에 關하여 記述하겠다.

(A) Pitch-excited LPC vocoder^{2, 5-7)}

Autocorrelation 方法에 依한 LPC vocoder 의 블록도가 그림 4에 그려져 있다. 이 블록도의 重要한 部分들을 說明하면 다음과 같다.

a. Low-pass filtering

音聲을 디탈化하기 前에 帶域 制限을 시키야 하는 데 upper cut-off 周波數는 vocoder 의 使用 目的과 傳

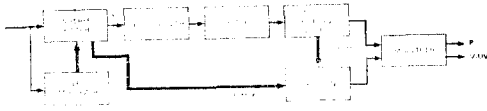


그림 6. LPC inverse filtering 과 AMDF를 사용한 pitch extractor

Fig 6. A pitch extractor based on LPC inverse filtering and AMDF.

인 境遇에는 그의 周期를 測定해야 한다. pitch extraction 은 pattern 認識을 하는 것으로서 vocoder 研究가 始되던 後부터 繼續 研究가 되어 왔지만 아직까지 完璧한 시스템을 만들 수 없는 어려운 分野이다. pitch extraction 의 代表的인 例로는 autocorrelation 또는 그의 變形인 average magnitude difference function (AMDF)을 使用하는 方法^{15,16)}, parallel processing 과 같은 time-domain 에서 하는 方法¹⁷⁾, data reduction 方法¹⁸⁾, cepstrum 方法¹⁹⁾, LPC inverse filtering 을 使用하는 方法^{20,21)}, 등을 들 수 있다. 위의 여러 方法들 중 나중의 두 方法이 가장 效果의이다. 특히 近來에 研究된 LPC inverse filtering 에 依하여 얻은 誤差 信號를 가지고 AMDF 波形을 形成하여 얻은 pitch extraction[그림 6 참조]은 正確한 結果를 얻을 뿐만 아니라 hardware 構成時 multiplier 가 必要 없는 長點이 있다.

e. Parameter 符號化 및 傳送

LPC 分析器가 抽出한 parameter 들, 즉 10개 程度의 反射係數 $\{k(i)\}$, 有聲音/無聲音 (V/UV) 分別, 周期(P), 그리고 Gain(G)은 傳送되기 前 먼저 量子化되어야만 된다. $\{k(i)\}$ 를 除外한 다른 parameter 들은 線型量子化하여 符號化하면 傳送速度가 約 400~500 bps 程度로 充分하다. 그러나 $\{k(i)\}$ 를 符號化하는 데에는 적어도 約 2,000 bps 가 必要하므로 vocoder 의 全體 傳送 速度(約 2,400 bps)의 大部分이 $\{k(i)\}$ 를 符號化하는 데 쓰여지고 있다. 따라서 傳送 速度를 줄이거나 音質을 改善하기 爲해서는 $\{k(i)\}$ 를 效果的으로 符號化하는 問題가 重要하다. 지난 數年 동안 提示된 方法에는 Log area ratio 量子化 方法²²⁾, equal area 量子化 方法²³⁾, arc sine 方法²³⁾, 그리고 piecewise linear irear quantization²⁴⁾ 등이 있다.

f. 音聲 合成(Speech Synthesis)

傳送된 音聲의 parameter 들을 受信器에서 받으면 이들을 利用하여 音聲을 合成하게 된다. 受信機는 作用上 크게 excitation source 와 合成 filter 의 두 部分으로 나눌 수 있다. excitation source 로서는 pulse

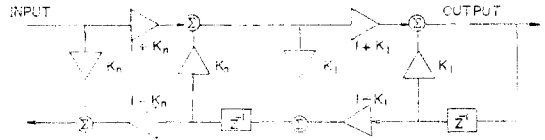


그림 7. LPC vocoder 合成器의 lattice 型 (上)과 ladder 型 filter(不)

Fig. 7. Lattice (top) and ladder(bottom) filters for LPC vocoder.

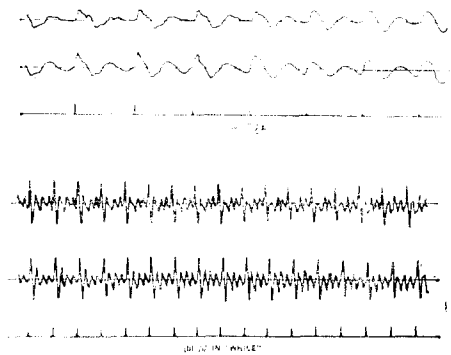


그림 8. LPC Vocoder 입력음성(上) 합성음성(中) 파형의 비교 (아래에 보여진 pulse train 이 excitation signal 임)

Fig 8. Comparison of input(top) and synthesis (middle) speech of LPC vocoder, (The pulse train shown at the bottom is the excitaion.)

generator 와 random noise generator 를 使用하고, V/UV 情報에 따라 switch 가 合成 filter 와 두 source 중 하나에 連結된다. 勿論 pulse generator 에서 나오는 pulse 의 發生 間격은 受信된 pitch 周期 P 에 依해서 定해진다.

合成 filter 는 數式的으로 式 (1)과 같고 係數 $\{a(i)\}$ 를 갖고 構成할 수 있으나 실제 '는 反射계수 $\{k(i)\}$ 로 構成한다. 그 理由는 受信된 係數가 $\{k(i)\}$ 이므로 $\{a(i)\}$ 로 다시 바꾸는 번거로움을 避할 수 있고 $\{k(i)\}$ 로 構成된 filter 가 構造上 round-off 誤差를 減少시키는 長點이 있기 때문이다. $\{k(i)\}$ 를 使用하여 構成한 lattice 型과 ladder 型의 filter 가 그림 7에 그려져 있다. 두 型을 比較하여 볼 때, ladder 型은 lattice 型보

다 round-off 誤差가 적은 反面 multiplier를 많이 使用하기 때문에 計算 時間이 많이 걸리고 hardware가 더 複雜한 短點이 있다.²⁴⁾ 마지막으로 合成 filter의 出力은 受信된 gain에 依하여 振幅이 調整된 後 D/A 變換器에 依해서 analog 信號로 바꾼 다음 pre-emphasis의 反對作用인 de-emphasis를 해주면 原來의 音聲과 같은 信號를 얻을 수 있다. 原來의 音聲과 LPC vocoder에 依해서 얻어진 合成된 音聲의 波形이 그림 8에 比較되어 있다.

參考로 지금까지 論한 LPC vocoder의 計算에 所要되는 時間을 생각하여 보면, 勿論 이 시스템은 real

time으로 動作을 해야 하므로 音聲의 parameter들을 每 20 ms마다 送信해야 될 境遇 모든 計算을 그 동안에 마쳐야만 될 것이다. 多幸히 요즘 많이 쓰이는 Schottky TTL MSI나 LSI는 速度가 빨라서 이러한 半導體 부품으로 시스템을 構成하면 LPC 分析, pitch extraction, 音聲合成을 모두 10 ms 안에 할 수 있다.²⁵⁾ 또한 近來에는 vocoder 시스템을 micro-processor로도 많이 만들고 있는데²⁶⁾ real time 시스템이 되기 위해서는 速度가 빠른 bipolar microprocessor를 使用하여야 한다.

〈第15卷, 6號계속〉