

論 文

포르만트 VOCODER에 의한 한국어
음성합성에 관한 研究

正會員 許 康 仁* 正會員 李 大 寧**

A Study on the Synthesis of Korean
Speech by Formant VOCODER

Kang In HUR*, Dai Young LEE**

要 約 본 論文에서는 포르만트 VOCODER를 구성하여 한국어 음성합성에 대한 연구를 보고하였다. 음성 합성 파라미터로서는 1) 스펙트럼 모멘트법으로 구한 포르만트 F1, F2, F3 및 평균성도 길이에서 선택한 F4, F5. 2) AMDF를 이용한 최적 Comb법으로 구한 피치 주파수. 3) 단시간 평균에너지와 단시간 평균 진폭 4) Fant가 발표한 포르만트 주파수에서 대역폭 결정식. 5) 단시간 영교차율에 의한 유/무성 결정. 6) Rosenberg가 제시한 음원파형. 7) 가우시안 백색잡음원이다. 합성 결과는 원음성과 매우 잘 일치함을 보였다.

ABSTRACT This paper describes a method of Korean speech synthesis using formant VOCODER. The parameters of speech synthesis are as follows, 1)formant F1, F2, and F3 by spectrum moment method and F4, F5 using the length of vocal tract. 2)pitch frequencies obtained by optimum Comb method using AMDF. 3)short time average energy and short time mean amplitude. 4)The decision method of bandwidth reported by Fant. 5)voiced/unvoiced discrimination using zerocrossing. 6)excitation wave reported by Rosenberg. 7)gaussian white noise. Synthesis results are in fairly good agreement with original speech.

I. 序 論

人間이 發聲하는 言語를 理解하고, 意思를 傳達할 수 있는 口(口)과 귀(耳)를 갖는 컴퓨터의 登場으로 因하여 多様な 言語패턴에 適用하기 위한 많은 研究가 發表되어 一部는 實用化가 되고 있으나 보다 合理的이고 進歩的인 技法을 찾기 위한 研究들이 現在에도 繼續되고 있다.

* 東亞大學校 電子工學科
Dept. of Electronic Engineering, Dong-A
University, Pusan.

** 慶熙大學校 電子工學科
Dept. of Electronic Engineering Kyung Hee
University, Seoul.

論文番號 : 89-67(接受1989. 9. 20)

一般的으로 컴퓨터를 利用한 音聲通信에는 連續的인 音聲 信號를 分析하여 離散的인 言語記號로 變換시키는 音聲認識, 음성의 意味를 理解하는 自然言語 理解 및 原音聲 信號를 再生 시키는 音聲合成으로 區分 되어진다.

效果的인 音聲의 認識이나 合成을 遂行하는 데에는 우선 願하는 音聲의 重要的 特性을 나타내는 特徵 파라메터들을 明確하게 分析하는 것이 基礎이자 前提가 되는 것이다.

이에 關聯되는 主要特徵 파라메터들로서는 피치(pitch)周期, 短時間 平均 에너지(Short time mean energy), 短時間 平均 진폭(Short time mean amplitude), 平均 전력(Mean power), 포르만트(Formant), 帶域幅(Bandwidth) 및 有/無聲 檢出을 위한 短時間 零交叉率(Short time zerocrossing rate) 등이 檢討되어진다.

이러한 音聲波形的 生成 過程을 크게 區分하면 音源의 生成, 聲道에 의한 調音 및 입술(脣)과 코(鼻孔)에 의한 放射의 3 過程으로 나눌 수 있다.

音響管 모델의 聲道部分은 調音器官에 의해 音을 形成하는 部分이며, 이 聲道에서 聲道 傳達 函數의 極으로 定義되며, 스펙트럼상의 主要 피크(Peak)값으로 나타나는 것이 特徵 파라메터 포르만트(formant)이다.

따라서 해당된 音聲의 포르만트를 正確하게 抽出하는 것이 自然言語를 나타내는 音聲의 重要的 情報를 찾는 것이다.

이들 포르만트는 平均길이가 약 17cm인 成人 男性의 聲道인 경우에 5 KHz 以內에서 4~5個가 存在하며, 母音의 音韻性은 낮은 周波數에서 3個의 포르만트로 나타내는 것이 一般的이다.

Fant와 Schaffer & Rabiner 등은 音聲의 포르만트의 모델에서 時間的으로 變化하는 第1, 第2, 第3 포르만트 周波數를 利用하여 꽤 性能이 좋은 合成音聲을 얻어낸 바 있었다.¹⁾

이러한 포르만트 周波數의 推定방법에서 스펙트럼 패턴의 大局的인 形狀에 注目하며 스펙트럼을 몇 個의 帶域으로 分割하여 各 帶域의 第1次 모멘트로서 포르만트 周波數를 推定하는

스펙트럼 모멘트법이 있다.^{2) 3) 4) 5) 6) 7)}

스펙트럼 모멘트法의 利用에 있어서는 各 帶域에 포르만트가 오직 1개만 하도록 分離하면 正確한 포르만트를 推定할 수 있다.

따라서 本 研究는 다른 方法에 비해 正確하게 포르만트를 抽出할 수 있는 스펙트럼 모멘트法을 基礎로 하여 韓國語 音聲의 合成을 하며 遂行上의 問題點과 이의 解決을 위한 理論的 基礎를 檢討하였다.

韓國語의 音聲 合成은 다음과 같은 파라메터들로서 터미널 아날로그 포르만트 合成器(Terminal analog formant synthesizer)에 의한 포르만트 VOCODER를 構成하여 實驗하였다.⁸⁾

合成 파라메터

- 1) 스펙트럼 모멘트法으로 구한 포르만트 F1, F2, F3 및 미리 設定된 F4, F5
- 2) AMDF를 利用한 最適 Comb 法으로 구한 피치(pitch) 周波數
- 3) 短時間 平均 에너지와 平均 振幅
- 4) Fant가 提示한 포르만트값에서의 帶域幅 決定式
- 5) 零交叉率(Zero crossing rate)에 의한 有/無聲 決定
- 6) Rosenberg가 提示한 音源 波形
- 7) 가우시안 白色 雜音源

合成한 音聲은 單母音, 連續母音, 紋章音聲이다.

II. 포르만트(Formant) 周波數의 推定

II-1. Cepstrum 分析

音聲波形 $s(t)$ 는 時間的이나 에너지 面에서 大部分을 차지하는 有聲音으로 表現되며 이 때의 音源을 周期 T [sec]의 펄스列로 近似시키면 다음과 같이 나타난다.

$$s(t) = \sum_{n=-\infty}^{\infty} S_0(t) \delta(t-nT) \quad (1)$$

여기서 $S_0(t)$ 는 음성생성계의 임펄스 응답이고

$\sum_{n=-\infty}^{\infty} \delta(t-nT)$ 는 周期 T의 펄스列이다.

式(1)의 兩邊을 푸리에(Fourier) 變換하여 周波數 領域으로 表現하면 다음가 같다.

$$s(w) = s_0(w) \left[\frac{\sin\{(2N+1) \frac{wT}{2}\}}{\sin \frac{wT}{2}} \right]^2 \quad (2)$$

여기서 $s(w)$ 와 $s_0(w)$ 는 各各 $s(t)$ 와 $s_0(t)$ 의 電力 스펙트럼이다. 또 []안은 標本化 函數이며 周期 T의 逆數 $w=2\pi/T$ 일 때마다 線스펙트럼 構造를 나타낸다.

式(1)에서 音聲波形 $s(t)$ 는 音聲生成系의 입력 應答 $s_0(t)$ 가 周期 T마다 加算된 것이고, 式(2)는 音聲生成系의 스펙트럼 包絡 $s_0(w)$ 와 周波數 $w=2\pi/T$ 의 高周波에서 線 스펙트럼 構造를 보인다.

時間 領域에서의 convolution은 周波數 領域에서 積으로 變換되므로 式(2)에서 兩者를 分離하기 위해서는 먼저 對數를 取하여 積을 和로 變換하여야 한다.

$$\begin{aligned} \log_e s(w) \\ = \log_e s_0(w) + 2 \log_e \left[\frac{\sin\{(2N+1) \frac{wT}{2}\}}{\sin \frac{wT}{2}} \right] \end{aligned} \quad (3)$$

式(3)의 右邊 第 1 項 $s_0(w)$ 는 音聲 生成系의 電力 스펙트럼이고, 形態는 w 에 對해 比較的 淸淸히 變化한다.

第 2 項의 []는 角周波數가 $2\pi/T$ 의 整數倍의 線 스펙트럼이며, w 에 對해 急激한 變化를 한다.

따라서 w 를 時間 t 로 變換하면 $s(w)$ 의 周波數 成分은 低周波 成分이 되고 []의 周波數 成分은 周期 T인 高周波 成分으로 解析될 수 있다.

여기서 다시 逆 FFT를 取하면 스펙트럼 上에서 第 1 項은 원래의 音聲 波形의 電力 스펙트럼의 包絡으로, 第 2 項은 피치 高周波의 線 스펙트럼으로 分離할 수가 있다.

이 原波形의 電力 스펙트럼에 對數를 取하여 이 對數 電力 스펙트럼을 人力 波形으로 變換한 結果를 Cepstrum이라 부른다.

이 때 周波數에 對應한 變數를 quefrequency라 하며 Cepstrum上에서 quefrequency의 大小(高低)에 따라 信號를 分離하는 操作을 lifter라고 한다.

즉, Cepstrum을 quefrequency의 大小(境界는 T)에 의해 liftering 한 스펙트럼의 包絡 成分에서 低 quefrequency 成分을 抽出하여 다시 FFT 한 後 對數化하면 $s(w)$ 를 抽出할 수 있다.

또 피치 成分은 Cepstrum 上에서 周期 T로 나타나므로 이 quefrequency에서 피치 抽出도 구할 수 있다.

그림 1에 스펙트럼 包絡의 抽出過程을 보았다.

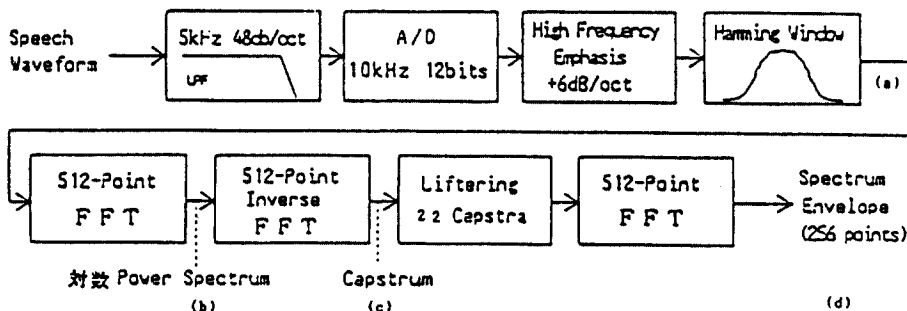


그림 1. Cepstrum 分析에 의한 對數 스펙트럼 包絡의 抽出 手順
Flow diagram of log spectrum envelope by Cepstrum analysis

II-2. 스펙트럼 모멘트

스펙트럼의 振幅을 $A(i)$, $A(i)$ 의 中心周波數를 $f(i)$, ($i=1, 2, \dots, N$)로 할 때 K 次 스펙트럼 모멘트는 다음과 같이 定義된다.

$$M_0 = \sum_{i=1}^N A(i)$$

$$M_1 = \sum_{i=1}^N A(i) f(i) / M_0 \quad (4)$$

$$M_k = \sum_{i=1}^N A(i) (M_1 - f(i))^k / M_0$$

M_1 는 스펙트럼의 重心周波數이고, M_2 는 M_1 의 周邊에서 스펙트럼의 펼쳐진 程度, M_3 는 M_1 의 周邊에서 스펙트럼의 非對稱性을 反映하는 파라메터이다.

그림 2는 單一 피-크를 갖는 4 種類의 스펙트럼에 대하여 M_1, M_2, M_3 를 計算하여 나타낸 것이다.

여기서 M_3 의 符號는 화살표 方向을 나타내고 있다.

그림 2의 上段의 (a), (b)와 같이 스펙트럼의

左右對稱이며, 피-크가 帶域의 中心에 있을 境遇에는 M_1 은 피-크에 一致하여 $M_3=0$ 이다.

이 때 $\sqrt{M_2}$ 는 피-크의 帶域幅을 나타내고 있다.

下段의 (c), (d)는 스펙트럼이 非對稱이며, 피-크가 帶域의 中心에 있지 않는 境遇이다.

이 때 M_1 는 피-크에 一致하지 않으며, (c)와 같이 피-크가 M_1 보다 低周波側에 있는 境遇에는 $M_3 < 0$, (d)와 같이 高周波側에 있는 境遇에는 $M_3 > 0$ 으로 한다.

그러나 그림에서 구한 電力 스펙트럼을 그대로 모멘트로서 구하면 負의 db 값을 갖는 成分이 實際의 스펙트럼 패턴에 비해 너무 작게 評價되어진다.

게다가 帶域에 따라서 $M_0=0$ 으로 될 可能性이 있으며, 計算에서도 좋지 않는 境遇가 많다.

또 $M_0 < 0$ 으로 된 境遇에는 $M_0 > 0$ 일 때와 M_3 의 符號가 바뀔 수도 있다.

그러므로 모멘트 計算을 하기 前에 스펙트럼에 前處理를 하여 $M_0 > 0$ 이 되게 할 必要가 있다.

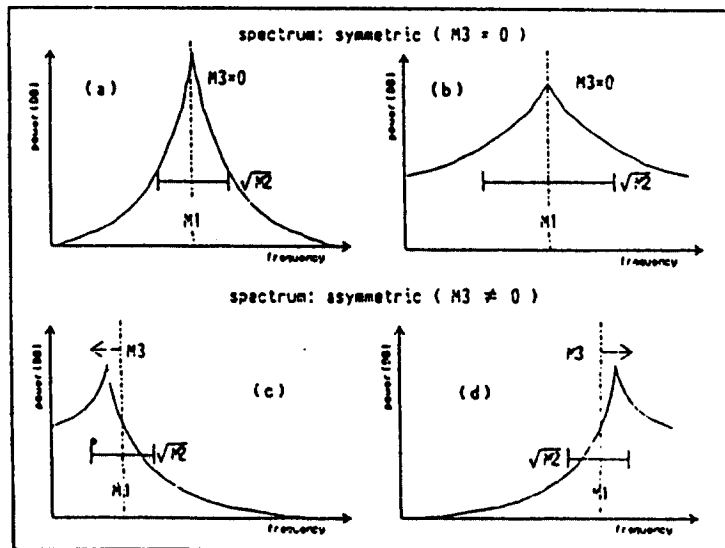


그림 2. 스펙트럼 形狀과 스펙트럼 모멘트
Spectrum pattern and spectrum moments

즉, 스펙트럼의 振幅의 變化幅을 [0.1]로 正規化 하는 것이다.

II-3. 帶域 分割

10 Khz 샘플링을 하였을 때 分析되어지는 帶域은 0~5 KHz이다.

스펙트럼 全體의 M_1 을 FC로 하고, FC 보다 낮은 帶域의 M_1 을 FLC, 높은 帶域의 M_1 을 FHC로 한다.

3 帶域(FLC, FC), (FC, FHC), (FHC, 5Khz)의 M_1 을 計算하여 順序的으로 FLCC, FHCC, FHHC로 한다(그림 3).

그림 3에서, 高木等은 最初의 3個의 帶域을 (0Khz, FLCC), (FLCC, FHCC), (FHCC, FHHC)로 定하였다.

本 研究에서는 첫번째 帶域의 FLCC 값을 FLC와 FLCC의 사이에서 여러 實驗을 通해 設定하였으며, 두번째 帶域의 FLCC 값은 첫번째 帶域에서 구한 第 1 포르만트의 0次 推定值 $f_1^{(0)}$ 의 스펙트럼의 3 dB 点의 周波數 값으로 했다. 그리고 두번째 帶域의 FHCC는 FHCC와 FHC의 사이에서 設定하였다.

세번째 帶域의 FHCC 값은 두번째 帶域에서 구한 第 2 포르만트 0次 推定值 $f_2^{(0)}$ 의 스펙트럼의 3 dB 周波數 값으로 했으며, FHHC는 FHC

와 FHHC의 사이에서 設定되었으며 이 때 구해진 값을 第 3 포르만트의 0次 周波數 $f_3^{(0)}$ 로 하였다.

II-4. 分割 帶域의 修正

前節에서 分割된 帶域內의 스펙트럼은 大部分의 境遇가 非對稱이므로 M_1 즉 第 0次 推定值는 포르만트 周波數의 推定值로서 良好한 값이라고 할 수 없다.

따라서 帶域을 修正하여 推定值를 피-크에 近似시켜야 한다. 이 境遇 第 3次 모멘트 M_3 의 符號가 스펙트럼 패턴의 非對稱을 나타내는 것을 利用하여 $M_3 < 0$ 일 때 스펙트럼의 피-크는 M_1 의 低域側에 存在하고, $M_3 > 0$ 일 때 M_1 보다 高域側에 存在하므로 M_3 의 符號에 따라서 現在의 帶域을 스펙트럼의 低域側 혹은 高域側으로 移動 시킨다.

移動量은 帶域幅에 比例하도록 $\sqrt{M_2}$ 의 값으로 決定한다. K次 推定值가 計算된 경우 다음 式으로 新帶域을 구한다.

$$\frac{M_3}{\text{新帶域}} = (f_1^{(k)} - C \sqrt{M_2}, f_1^{(k)} + \sqrt{M_2})$$

$$\frac{M_3}{\text{新帶域}} = (f_1^{(k)} - \sqrt{M_2}, f_1^{(k)} + C \sqrt{M_2})$$

$$C = 1.5 \text{ (實驗值)} \quad (5)$$

구해진 新帶域의 M_1 은 다음의 推定值 $f_1^{(k+1)}$ 로 한다(그림 4).

즉 帶域의 上限, 下限의 周波數를 離散 스펙트럼의 要素番號로 고쳐서 計算한다.

이 때 最大에서는 스펙트럼의 要素間隔이 1/2의 誤差가 發生하며, 帶域幅의 最小值는 262.5 mel(256点의 스펙트럼에서 26点分의 幅, 實驗值)로 했다.

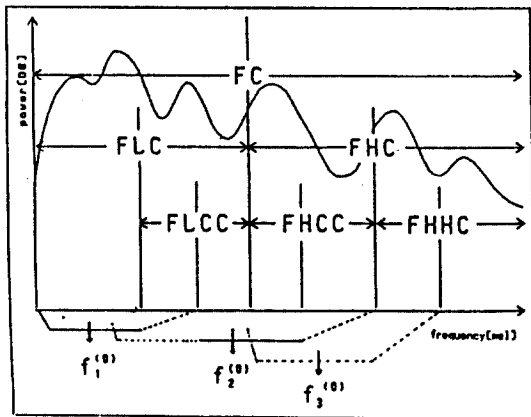


그림 3. 帶域 分割
Devlsion of the frequency Region

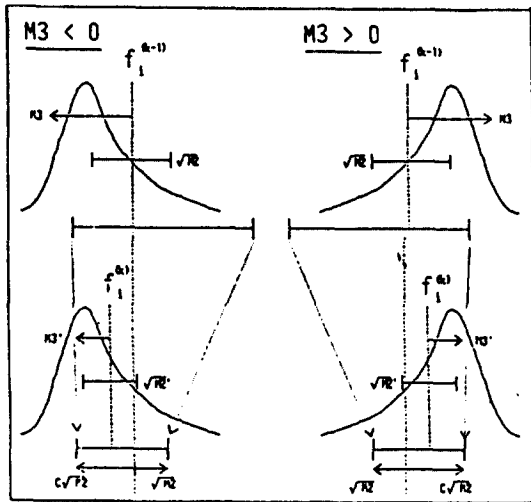


그림 4. 分割 帶域의 修正
Adjustment of a frequency region

II-5. 收斂 判斷

推定値는 帶域 修正의 反復에 의하여 스펙트럼의 피-크에 近似化 되지만 新帶域 計算의 誤差로 인하여 피-크의 周邊에서 振動하는 것을 收斂했다고 判斷한다.

그러므로 正確한 收斂 判斷을 위해 現在의 推定値 $f_i^{(k)}$ 와 하나前의 推定値 $f_i^{(k-1)}$ 의 차를 $D_i^{(k)}$ 로 한다.

$$D_i^{(k)} = f_i^{(k)} - f_i^{(k-1)}, \quad i=1, 2, 3; k=1, 2, 3, \dots \quad (6)$$

$D_i^{(k)}$ 의 符號가 變할 때에는 推定値의 移動 方向이 變한 것으로 判斷하며, 2回 連續으로 $D_i^{(k)}$ 의 符號가 變할 때 즉,

$$\begin{aligned} \text{sgn}(D_i^{(k-1)}) &\neq \text{sgn}(D_i^{(k-2)}) \\ \wedge \\ \text{sgn}(D_i^{(k)}) &\neq \text{sgn}(D_i^{(k-1)}) \end{aligned} \quad (7)$$

여기서,

$\text{sgn}(x) = 0 (x=0), -1 (x<0), +1 (x>0)$: 부호함수

로 된 때에는 $f_i^{(k)}$ 와 $f_i^{(k-1)}$ 의 平均値를 第 i 포르만트 周波數의 推定値로 한다(그림 5).

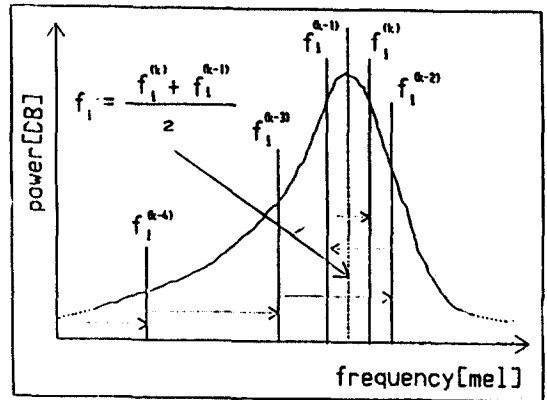


그림 5. 帶域 修正의 收斂 判斷
Convergence Judgement of the region adjustment.

$$f_i = (f_i^{(k)} + f_i^{(k-1)}) / 2 \quad (8)$$

그러나 隣接 포르만트의 推定値가 一致하지만 推定値가 隣接 포르만트와 交差하는 경우에는 1段階前의 推定値를 取한다.

II-6. 스펙트럼의 底 올림

스펙트럼의 變動이나 포르만트 패턴 以外의 凹凸이 모멘트 計算에 미치는 影響을 弱화하기 위해 正規化된 스펙트럼에 一定한 底 올림을 한다.

底 올림으로 스펙트럼 패턴의 相對的인 比率이 鈍化되어 主 피-크만이 計算에 使用되도록 한다.

그러나 底 올림量이 너무 많으면 音韻에 의한 스펙트럼의 잘못으로 모멘트에 影響을 미치므로 適當한 값을 取해야 한다.

本 研究에서는 底 올림量의 最適値는 스펙트럼의 動的範圍(最大値와 最小値의 比)를 規準化하여 計算 했으며 男性은 0.1~0.8 正道의 底 올림量에 對해 거의 變動이 없었으며 여기서는 0.1로 設定하였다.

여성은 實驗 結果 底 올림量을 1.0으로 設定하였다.

II-7. 모멘트법에 의한 포르만트 周波數의 flow chart

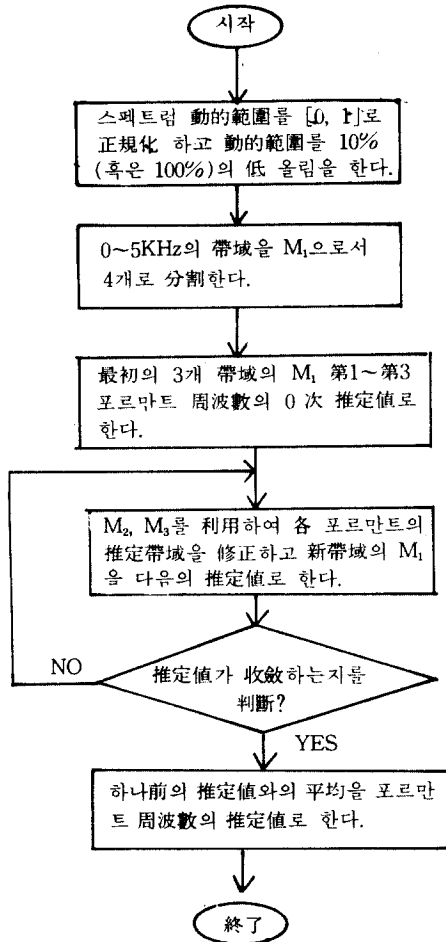


그림 6. 모멘트법에 의한 포르만트 추출 흐름도
Flow diagram of formant extraction by moment method.

III. 피치(pitch)周波數의 抽出

音源에서 有聲音源의 特性은 話者의 個人性, 情緒性 및 音聲의 自然性 등이 크게 寄與를 한다.

이러한 有聲音源의 反復되는 周期를 피치(pitch) 周期라고 한다. 피치 周期가 갖는 言語的

인 要素로서는 單語나 音節의 악센트, 句나 文章의 抑揚패턴 等이다.

音聲 分析合成系에서는 音聲波形에 對하여 10m~20ms 程度의 分析 프레임(frame) 內에서 平均의 特性을 구하며 피치 抽出에 對해서도 分析 프레임 內에서 平均값을 取한다.

VOCODER의 發明以來 提案된 피치 抽出法은 여러가지가 있다. 그 중에서도 平均振幅差函數를 利用하여 最適 Comb法으로 音源의 피치周波數를 구하였다.

平均振幅差函數의 定義는 다음과 같다.

$$D_n = \frac{1}{N} \sum_{k=0}^{N-1} |S_k - S_{k-n}| \quad (9)$$

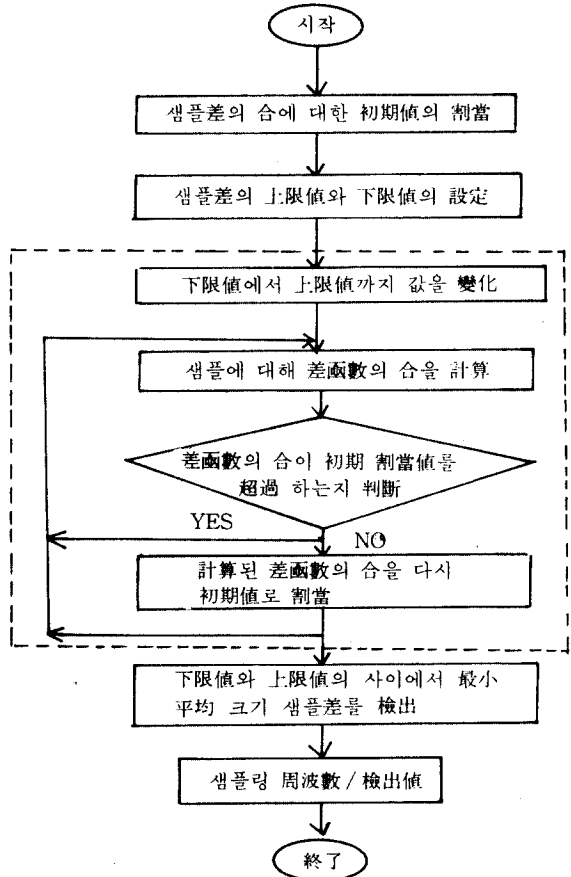


그림 7. 최적 Comb법에 의한 피치 주파수 추출 흐름도
Flow diagram of pitch frequency by optimum Comb method.

즉, 離散信號 샘플 S_k 周邊에서 서로 n 만큼 떨어져 있는 두信號값의 차들을 구하여 합한 것으로 이 값이 적을수록 k 만큼 떨어져 있는 두信號는 서로 類似하다는 것이다.

따라서 피치 周期를 P 로 할 때 $n=0, P, 2P, \dots$ 位置에서 D 는 極小值를 가진다.

最適 Comb法에 의한 피치 檢出의 原理는 다음과 같다.

平均振幅差函數의 合에 의한 初期值를 割當한다.

그리고 샘플差의 上限 값과 下限 값을 決定하여 下限 값에서 上限 값으로 값을 變化 시키면서 差函數의 合이 割當한 初期值 값을 超過하지 않는 k 의 값을 구한다.

이 k 의 값을 샘플링 周波數에 대해 나누어 주면 피치 周波數를 구할 수 있다.

그림 7에 흐름도를 보였다.

IV. 短時間 에너지와 平均 振幅

音聲 信號의 振幅은 時間에 따라 變化하며 無聲音의 세그먼트(segment)의 振幅은 有聲音의 세그먼트의 振幅보다 一般的으로 매우 적게 나타난다.

그러므로 有聲音의 세그먼트와 無聲音의 세그먼트를 識別할 수 있는 根據가 必要하며, 이것은 短時間 에너지와 平均 振幅으로 解決할 수가 있다.

音聲 信號의 短時間 에너지는 다음과 같이 定義 된다.

$$E = \sum_{n=-\infty}^{\infty} [x(m) w(n-m)]^2 \quad (10)$$

위 式은 다음과 같이 나타낼 수 있다.

$$E = \sum_{n=-\infty}^{\infty} x^2(m) \cdot h(n-m) \quad (11)$$

여기서 $h(n) = w^2(n)$

임펄스 應答 $h(n)$ 혹은 窓函數 $w(n)$ 의 選擇은

短時間 에너지에 關係되며 여기서는 Hamming 窓을 使用하였다.

$$h(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi n / (n-1)), & (0 < n < N-1) \\ 0 & \text{그 외} \end{cases} \quad (12)$$

이러한 에너지 函數는 有聲音에서 無聲音이 되는 時刻과 逆으로 無聲音에서 有聲音이 되는 時刻를 指示하는데 重要하며 良好한 音聲(信號對 雜音比가 높은 音聲)에서 에너지는 無音과도 識別하는데 使用할 수가 있다. 式(10)에서 短時間 에너지 函數는 큰 信號 레벨에 對해서도 信號 레벨을 2乘하므로 매우 敏感한 分布로 나타난다.

즉 $x(n)$ 의 샘플 사이의 變動이 매우 強調된다.

이러한 問題를 緩和하기 위해 短時間 平均 振幅函數를 利用할 수가 있다.

短時間 平均 振幅函數의 式은 다음과 같다.

$$M_n = \sum_{m=-\infty}^{\infty} x(m) w(n-m) \quad (13)$$

式(13)은 信號를 2乘 하는 대신에 絶對值를 取하고 있다.

短時間 에너지와 平均 振幅을 구하는 block diagram은 그림 8과 같다.

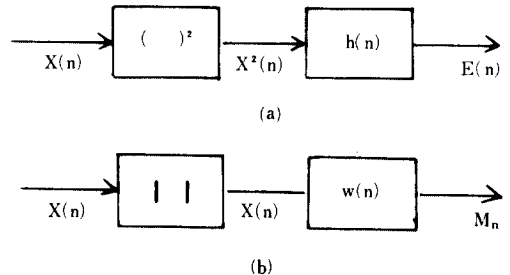


그림 8. 短時間 에너지와 短時間 平均振幅의 블록도
Block diagram representation of
(a) the short time energy
(b) the short time average magnitude

V. 短時間 平均 零交差率

離散時間 信號를 取扱하는 경우 隣接하는 샘플이 正負의 다른 符號값으로 될 때 일어나는 것을 零交差率이라고 한다.

零交差가 일어나는 單位 時間當의 回數를 零交差率이라고 하며 信號에 包含된 周波數 成分의 簡單한 測度이다.

音聲信號는 帶域幅이 넓은 信號이므로 스펙트럼의 性質 評價는 短時間 平均 零交差率로 表現한다.

離散時間 信號의 短時間 平均 零交差率의 定義는 다음과 같다.

$$Z = \sum_{m=-\infty}^{\infty} \frac{|\text{sgn}[x(m)] - \text{sgn}[x(m-1)]|}{w(n-m)} \quad (14)$$

여기서

$$\text{sgn}[x(n)] = \begin{cases} 1, & x(n) > 0 \\ -1, & x(n) < 0 \end{cases}$$

$$w(n) = \begin{cases} 1/2N, & 0 < n < N-1 \\ 0, & \text{그 외} \end{cases}$$

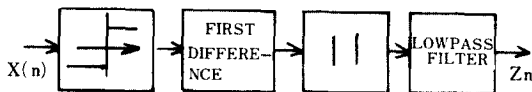


그림 9. 短時間 平均 零交差率의 블럭圖
Block diagram representation of short-time average zero-crossing

VI. 音聲合成 시스템

VI-1. 시스템의 概要

音聲은 聲帶의 振動 혹은 口腔內에서 좁혀진 部分에 의해 생겨진 亂流가 音源이 되고 聲道의 音響傳達 特性에 의해 周波數 成分이 變形되어 입술에서 放射 되어진다.

이러한 모델에 基礎를 둔 音聲 合成法을 生成源 處理에 의한 音聲合成이라고 한다.

生成源 處理에 의한 音聲合成은 18世紀에 機械的 構造로 實現된 것이 있으며 電氣的 構造 된 것은 1939년에 Dudley 등이 提示한 바 있고 이것은 VODER로 이름 붙여졌다.

이러한 電氣的 音聲 合成器의 發明 以來 이것에 多樣한 變形을 加한 合成器들이 考案되게 되었다.

上述한 VODER는 채널形에 속한다.

이것은 周期的인 펄스 音源과 白色 雜音源의 두가지 音源이 있으며 그 出力을 帶域 필터로 分割하고 각 필터 채널의 出力 레벨을 調節하여 合成 音聲의 周波數 스펙트럼 包絡 特性을 制御하고 있는 것이 特徵이다.

本 研究에서는 보다 簡單하고 實用性의 可能性을 檢討하기 위하여 前節에서 分析한 音聲의 特徵 파라메터들과 터미널 아나로그 포르만트 合成器를 利用하여 포르만트 VOCODER를 構成하였으며 全體 構成圖는 그림 10에 나타낸 것과 같다.

이 시스템에서는 먼저 音聲 波形을 스펙트럼 모멘트法으로 抽出한 포르만트 F1, F2, F3와 帶域幅 B1, B2, B3, 聲道의 平均 길이로써 미리 決定된 포르만트 F4, F5와 帶域幅 B4, B5, 零交差率 및 短時間 平均 에너지를 音聲 파라메터 파일에 記錄시킨다.

最適 Comb法으로 피치를 抽出하고 이 피치 周波數와 短時間 平均 에너지 및 零交差率에 의한 有/無聲度를 參考하여 Rosenberg가 提示한 近似 聲帶音源 波形을 生成시켜 音源 파일에 記錄한다.

터미널 아나로그 포르만트 合成器는 音聲 파라메터 파일의 값을 順序的으로 읽고 零交差率과 短時間 平均 에너지로서 有/無聲을 判斷하여

音聲의 프레임에서는 포르만트 周波數, 帶域幅 및 音源波로서 音聲을 合成시키게 된다.

無聲으로 判斷된 프레임에서는 미리 準備된 雜音源 파일의 雜音 波形을 읽어 그 프레임의 平均 에너지의 크기를 갖는 雜音을 生成하게 된다.

이렇게 合成된 音聲은 D/A 變換(10KHz, 12bit)하고 低域 통과 필터를 通過시키면서 再生 音聲 波形을 만들게 된다.

VI-2. 音源 모델과 生成

合成音에 대해 音源의 影響은 매우 크며 소리의 高低, 強弱, 有/無聲의 比率 및 音質의 부드

러움 등과의 關聯性이 있다. 通常的인 會話 때에는 波形은 正弦波에 가까우며 高調波 成分이 나타나지 않으나 화를 내거나 強하게 말하면 三角波에 가깝고 高調波 成分이 많이 包含되어 있다.

따라서 合成音의 音質을 向上 시키기 위해서는 音源을 '어떻게 制御할 것인가?'가 問題가 된다.

音聲合成에 대해 有聲音源의 경우 最適의 音源 波形에 對한 研究로서 BELL 研究所의 Rosenberg 는 그림 11에 나타난 音源波形들로서 合成音에 대한 評價를 한 바 있으며 그림 11의 C 波形이 最適이라고 發表했다.

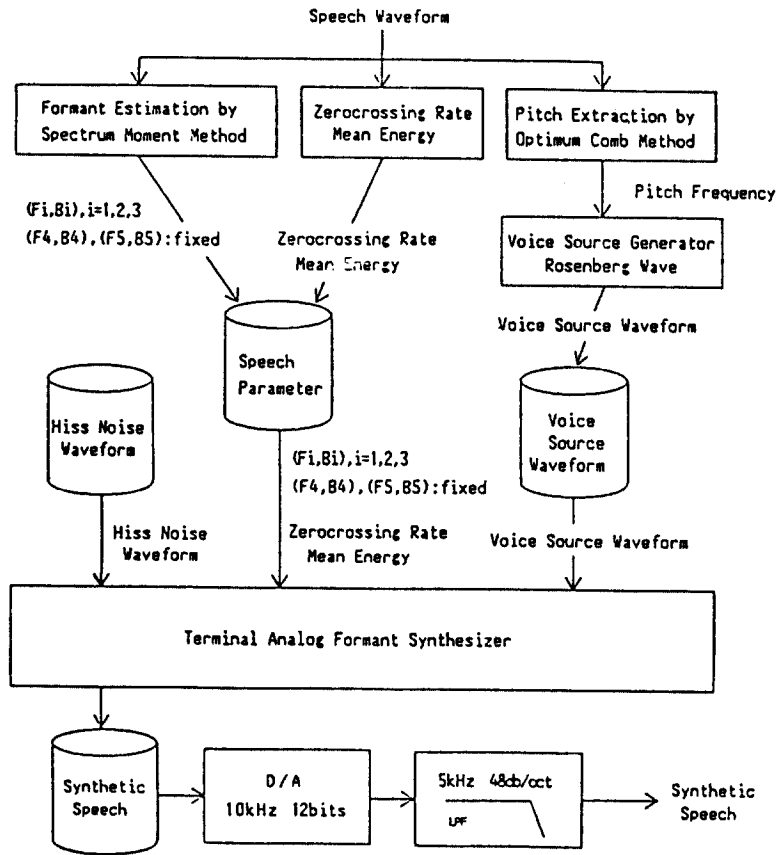


그림 10. 포르만트 VOCODER의 블록圖
Block diagram of a formant VOCODER

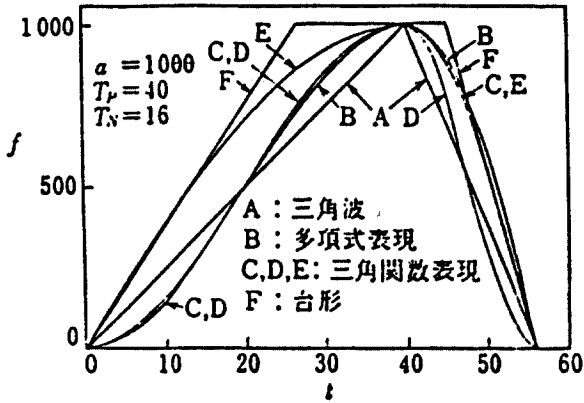


그림 11. Rosenberg에 의해 實驗된 音源波形
An experimented excitation wave by Rosenberg

本 研究에서도 Rosenberg의 最適 近似 音源波形을 利用하였으며 關係式은 다음과 같다.

$$G(n) = \begin{cases} 0.5[1 - \cos(\pi n / N1)] & , 0 < n < N1 \\ \cos[\pi(n - N1) / 2N2] & , N1 < n < N1 + N2 \\ 0 & , \text{그 외} \end{cases} \quad (15)$$

無聲音源으로서는 平坦한 스펙트럼 特性을 갖는 白色 雜音源을 利用하였다.

이러한 音源波의 生成은 音源의 有聲度, 平均 에너지 및 피치를 利用하였다.

이러한 音源波의 生成은 音源의 有聲度, 平均 에너지 및 피치 周波數로서 生成시킨다.

音源의 有聲度(變數名: DUV)로서 有聲音源의 強度(變數名: CV)와 無聲音源의 強度(變數名: CU)를 定하였다.

$$CV = \begin{cases} 1 & , DUV < 0.24 \\ 10^{*}(-DUV + 0.36) & , 0.24 < DUV < 0.36 \\ 0 & , 0.36 < DUV \end{cases}$$

$$CU = \sqrt{1 - CV^2} \quad (16)$$

그림 12~15들은 音聲을 合成한 例들이다. 點線의 左側 部分들은 原音聲의 平均 電力, 피치 周波數, 포르만트 軌跡과 帶域幅 및 原音聲 波形이고 右側 部分들은 有聲度에 따라 크기가 變化된 Rosenberg의 音源 波形과 合成된 波形을 보였다.

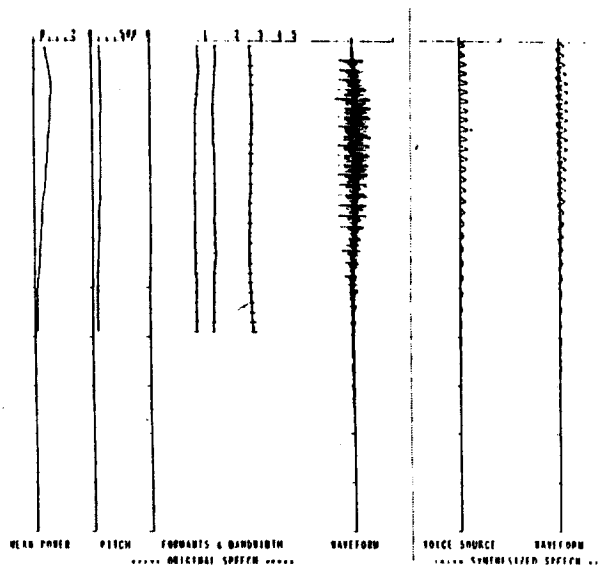
그림 12(a), (b)는 各各 韓國語 單母音 /아 (a) /, /이(i) /의 結果이며, 그림 13은 連續母音 /아이우에오(aiueo) /의 結果이다.

그림 14는 文章音聲 /이것은 책입니다 /의 結果이고 그림 15는 文章音聲 /대한민국은 민주 공화국이다 /의 結果이다.

單母音을 合成하여 聽取한 結果 /이(i) /는 不分明하게 聽取가 되었으나 그 外의 單母音들은 原音聲의 音韻性을 매우 잘 保存하고 있었다.

/이(i) /音이 不分明하게 들리는 이유는 /이(i) /의 第 3 포르만트가 成人 男性의 聲道 길이로서 數學的으로 計算된 포르만트 F4(3, 500 Hz)보다 크거나 거의 같은 값으로 되므로 混同이 일어나기 때문이다.

連續母音의 聽取 結果에서는 原音聲의 音韻性을 매우 잘 保存하고 있었으며 文章音聲의 聽取 結果에서는 一般 子音形은 거의 聽取가 되었으나



(a)

VII. 合成 實驗의 結果 및 考察

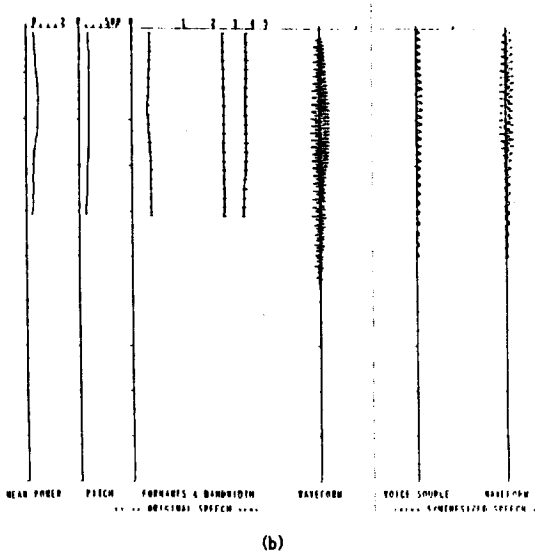


그림 12. 포르만트 VOCODER로서 合成한 例 (a) /아/
(b) /이/
A synthesis speech by formant VOCODER
(a) /a/ (b) /i/ (man)

破擦性的子音 /ㅈ/ 의 경우는 確實하게 聽取가 되지 않음을 알 수 있었다.

이것은 破擦性的子音은 發聲 時間이 짧아 分析區間을 調節하면 聽取가 良好하게 될 수 있을 것으로 思料된다.

VIII. 結 論

本 研究는 스펙트럼 패턴의 情報만을 利用하여 스펙트럼 모멘트法으로 韓國語 音聲에 대한 포르만트 周波數를 抽出하여 포르만트 VOCODER를 構成하여 合成實驗을 한 結果를 報告하였다.

序論에서 나타낸 合成 파라메터로서 單母音, 連續母音, 文章音聲을 合成하여 다음의 結果를 얻었다.

- 1) 單母音들은 原音의 音韻性을 잘 따르고 있었으나 /이(i)/ 音의 F3은 平均 聲道길이로서 結定된 F4에 비해 높은 경우도 있어 聽取가 不分明 하였다.

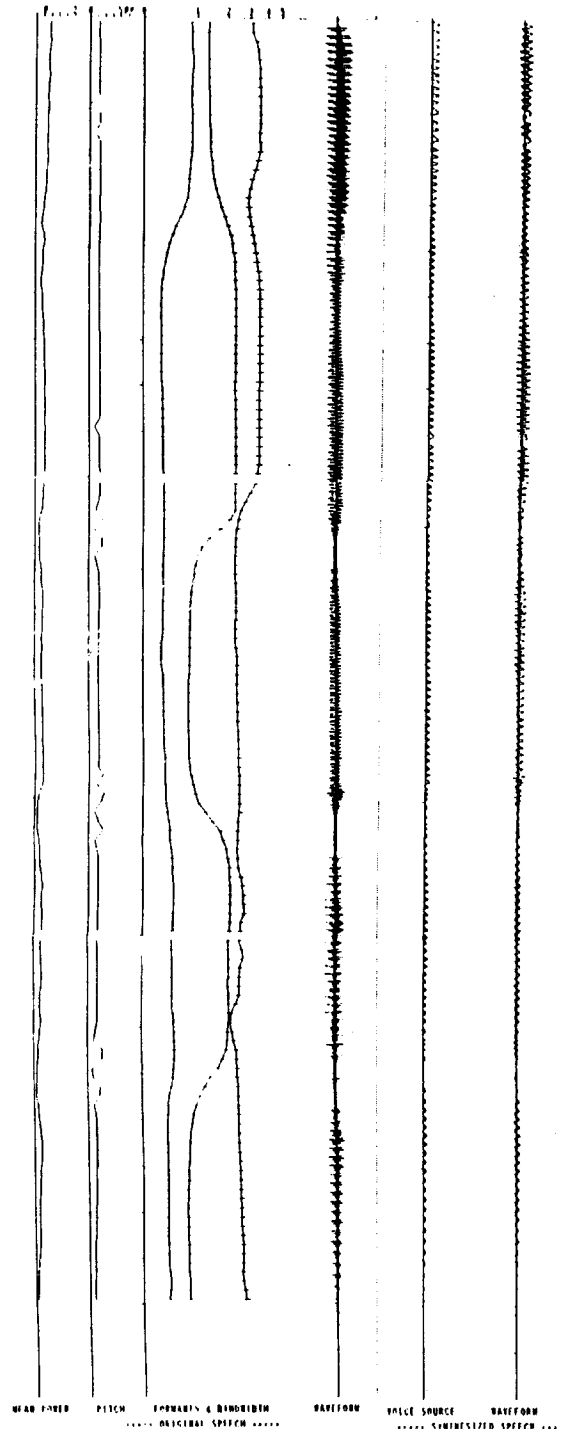


그림 13. 포르만트 VOCODER로서 合成한 例 /아 이 우 에 오/ (男性)
A synthesis speech by formant VOCODER /aieuo/(man)

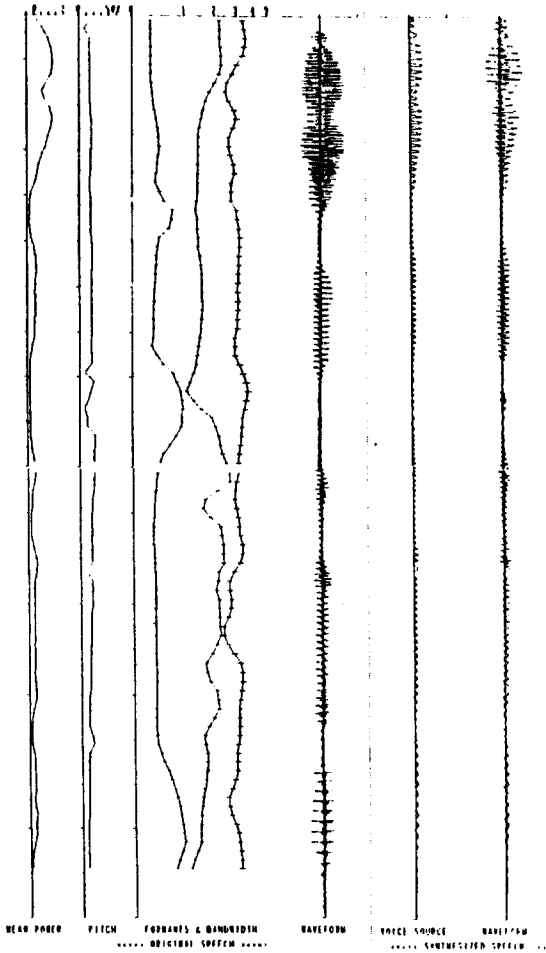


그림 14. 포르만트 VOCODER로서 合成한 例
/ 이것은 책입니다 / (男性)
A synthesis speech by formant VOCODER
/ this is a book / (man)

- 2) 連續母音(아이우에오)에서는 原音의 音韻性을 매우 잘 따르고 있었으며 單母音의 경우와는 달리 /이(i)/도 聽取가 可能하였다. 이것은 母音 1字를 發音할 時의 F3 보다 連續發音의 F3가 적게 나타난 때문에 생각된다.
- 3) 文章音聲의 合成結果 一般子音型은 거의 聽取가 되었으나 破擦性의 子音들은 確實하게 聽取가 되지 않았다.

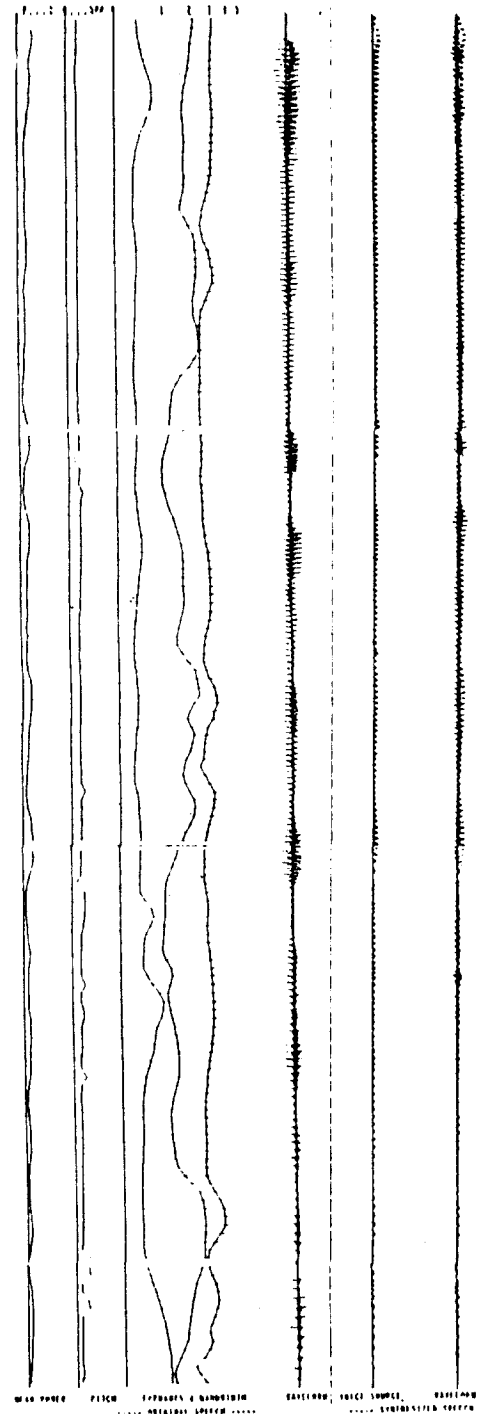


그림 15. 포르만트 VOCODER로서 合成한 例
/ 대한민국은 민주공화국이다 / (男性)
A synthesis speech by formant VOCODER /

이것은 無聲 破熱部の 경우 프레임 周期와 窓의 길이를 調整하면 良好한 聽取가 可能할 것으로 생각한다.

本 論文에서 構成된 포르만트 VOCODER 시스템으로서 音聲合成의 妥當性을 確認할 수 있었다.

參 考 文 獻

1. L.R. Rabiner, R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.
2. 鈴木誠史, 角川晴夫, 中田和男, “モーメント計算によるホルマント周波數の抽出”, 日本音響學會誌, 第19券, 第3號, pp. 106-114, 1963年
3. 中津井護, 鈴木誠史, “逆フィルタとモーメント計算による周波數抽出と合成音による評價”, 日本音響學會誌, 第26券, 第5號, pp. 211-221, 1970.
4. 日淵徹, 北脇信彦, 粕谷英樹, 城戸健一, “スペクトルモーメントを用いた音聲の特徴推出”, 電子通信學會電氣音響研究會資料, EA70-25, 1971-1.
5. 日淵徹, 粕谷英樹, 城戸健一, “スペクトルモーメントを用いた單語音聲の音素群への區分化”, 日本音響學會講演論文集, pp. 411-412, 昭和46年11月.
6. 高木一幸, 板橋秀一, “スペクトルモーメントを利用した音聲のホルマント周波數推定法”, 電子情報通信學會技術研究報告, SP88-80, 1988.
7. Kazuyuki Takagi and Shuichi Iahasi, Formant frequency estimation by moment calculation of the speech spectrum, JASA., S1, Vol.84, J6, p.22, Fall, 1988.
8. J.L. Flanagan, Speech Analysis Synthesis and Perception, Second, Expanded edition, Springe-Verlag, 1972.



許 康 仁(Kang In HUR) 正會員
1955年 2月20日生
1980年 2月: 東亞大學校 電子工學科 卒業
1982年 2月: 東亞大學校 大學院 電子工學科(工學碩士)
1983年 3月~現在: 慶熙大學校 大學院 電子工學科(博士課程)
1984年 9月~現在: 東亞大學校電子工學科 助教授

1988年 9月~1989年 8月: 日本筑波大學 電子情報工學系 客員 研究員



李 大 寧(Dai Young LEE) 正會員
1940年 3月18日生
1968年 9月~1970年 3月: 캘리포니아州 立大學院(工學碩士)
1976年 9月~1979年 9月: 延世大學校大學院電子工學科(工學博士)
1971年 9月: 慶熙大學校工科大学電子工學科 助教授
1977年 3月: 慶熙大學校工科大学電子工學科 副教授

1982年 3月~現在: 慶熙大學校工科大学電子工學科 教授
1988年 1月~現在: 本 學會理事