

線型 豫測法에 依한 音聲信號의 分析과 그 應用 方案 (Analysis of Speech Signals by Linear Prediction and It's Application)

- 音程에 따른 韓國語 單母音의 特性變化를 中心으로 -

金 明 圭 *
(Kim, Myoung Kyu)

要 約

韓國語의 主要 單母音에 對하여 音程을 變化시켜 가면서 線型 豫測法에 따른 模型 스펙트럼과 聲道의 形態를 推定함으로써 音의 高低에 따른 特性 變化를 分析하였다.

또 分析한 結果를 利用하여 音素의 組合에 依한 音聲合成方案을 提示하고 實驗에 依해 그 妥當성을 立證하였다.

Abstract

In this paper, the effect of tone variation of speech signals is discussed by showing the variations of the linear prediction model spectra and the estimated vocal tract shape for Korean vowels.

As an application of the analysis results, a speech synthesis scheme by combination of phonemes is also discussed based on experimental results.

I. 序 論

1960年代末, 線型豫測法(linear prediction method)이 音聲信號에 適用된 이후^{[1]~[3]} 音聲信號處理에 關한 研究가 國內外에서 急激한 進展을 보이고 있고^{[5], [6], [12], [13]} 高性能의 디지털 프로세스가 普及됨에 따라 그 應用 範圍가 점차로 擴大되고 있다.^[4]

線型豫測法은 音聲信號가 지나는 情報量을 줄이는 效率인 方法으로서 이것을 使用하면 정해진 容量의 記憶裝置에 더 많은 音聲情報를 收錄할 수 있고 주어진 通信線路에 더욱 많은 信號를 실어 送受信할 수 있는 등 여러 가지 잇점이 있다. 그러나 예를 들어 컴퓨터 音聲應答 시스템^[5]에서와 같이 많은 어휘에 대한 音聲合成이 必要한 境遇에는 LPC(linear predictive

coding)에 의해 情報를 收錄한다해도 너무 큰 容量의 記憶裝置가 要求되므로 實用上의 問題點을 가진다.

Atal과 Hanauer에 따르면^[7], 線型豫測法에 依해 얻은 파라미터들을 適當히 變換함으로써 音色 또는 音調가 다른 音聲을 合成할 수 있는데, 만일 基本音(音素)에 대한 파라미터만을 記憶시키고 合成 過程에서 組織적으로 變化시킴으로서 많은 종류의 소리를 合成하게 된다면 위와 같은 問題點에 解決될 것이다. 이를 위해서는 各 音에 對한 充分한 分析이 要求되는데 音을 決定하는 要素中 強弱과 長短은 그 意味가 明白하고 獨立인 處理가 可能하여 問題되지 않으나, 音의 高低와 音素 및 音色은 周波數 領域에서 相互關聯이 있어 獨立의 取扱이 困難하며 LPC 파라미터가 차지하는 情報量의 大部分이 이것들을 나타내기 위한 것임을 감안할 때, 이에 對한 分析은 重要한 意味를 가지는 것으로 判斷된다.

그러므로 本稿에서는 韓國語 單母音을 中心으로

a) 音의 高低에 따른 基本 周波數와 線型豫測 스펙트럼의 變化를 觀察하여 分析하고,

* 正會員, 西江大學校 理工大學 電子工學科
(Dept. of Electronics Engr., Sogang Univ.)

接受日字: 1981年 4月 8日

(※ 이 논문은 1980년도 문교부 학술연구조성비에 의하여 연구된 것임.)

b) 그 결과를 음성合成에適用, 音素의 組合에 依 로 定義되는 豫測誤差信號에 對하여 $\sum_{n=0}^{N+M-1} e^2(n)$ 을 最 小化하는 豫測係數 $\{a_{Mi}\}$, $i=1, \dots, M$ 로 分析된 合成의 可能性을 提示하고자 한다.

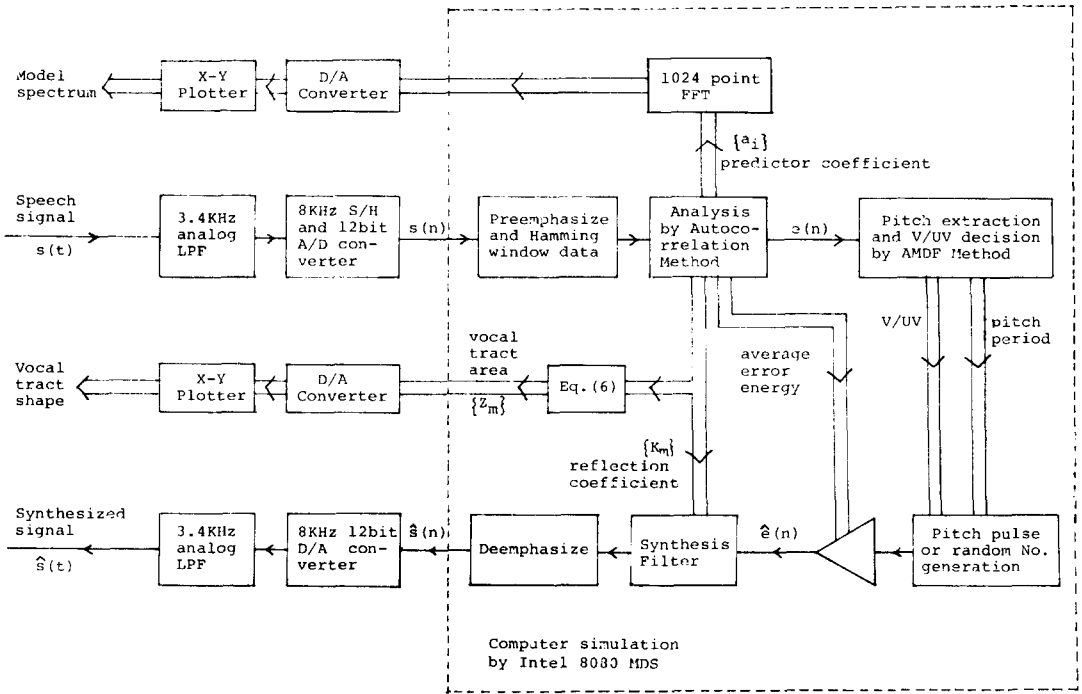


그림 1. 全系統의 블록선도

Fig.1. Block diagram of the overall system.

II. 線型豫測法에 依한 音成信號 처리방법

本稿에서 使用한 信號處理方法을 說明하기 爲하여 먼저 全系統의 블록선도를 그림 1에 紹介하는데 그림에 포함된 컴퓨터 서물레이션은^[12]를 參考, 修正하였고 各 部分은 다음과 같다.

a) 母音 信號의 에너지는 大部分 3.4 KHz 이내에 있으므로 72dB/octave의 遮斷特性을 갖는 PCM用 필터를 使用하고, 8 KHz로 標本化, A/D, D/A 變換하여 入出力을 處理하였다.

b) 하나의 分析 프레임(analysis frame)에 포함되는 데이터 수 N을 200(25ms)으로 하고 40개의 데이터를 overlapping 하였다.

c) 한 프레임의 入力 데이터를

$$s(n), 0 \leq n \leq N-1 \text{로 하고}$$

$$e(n) = \sum_{i=0}^M a_{Mi} s(n-i), \text{ 단 } a_{M0} = 1 \quad (1)$$

$$E(z) = A_M(z) S(z) = \left(\sum_{i=0}^M a_{Mi} z^{-i} \right) s(z) \quad (2)$$

터 $A_M(z)$ 를 定하는 Autocorrelation Method를 使用하였다.

d) 필터의 次數 M은 標本化 周波數가 8KHz 이므로 가장 適當한 값인 12로 하였다.^{[8][13]}

e) 豫測誤差信號를 分析하면 有聲·無聲音(V/UV; voiced or unvoiced)의 情報과 有聲音인 境遇 基本周波數(pitch frequency)를 알 수 있는데 이를 위하여 AMDF(average magnitude difference function)에 依한 方法을 使用하였다.^[9]

f) 式 (2)에서

$$S(z) = \frac{1}{A_M(z)} E(z) \quad (3)$$

이므로, $e(n)$ 과 같은 平均에너지를 갖는 周期的인 펄스(有聲音인 境遇) 또는 白色雜音(無聲音인 境遇)으로 $e(n)$ 을 代身하여 合成 필터 $1/A_M(z)$ 을 通過시키면 合成된 音聲信號 $\hat{s}(n)$ 을 얻는다.

g) 聲帶의 振動波形과 입술에서의 放射를 考慮하면 入力信號를 preemphasize 하여 分析할 境遇 $1/A_M(z)$

이 聲道의 傳達函數라는 物理的 意味를 가지므로 이를 위하여 遮斷 周波數가 100 Hz 인 $1 - 0.92 z^{-1}$ 의 傳達函數로 表現되는 디지털 필터를 使用하고 合成信號는 그 反對의 필터로써 deemphasize 하였다.

h) 標準化 周波數 f_s , 聲道의 길이 L , 音波速度 c 사이에

$$M = \frac{2L f_s}{c} \quad (4)$$

의 關係가 成立하도록 M 을 擇하고, 聲道를 均一한 斷面續을 가지고 M 개의 圓柱가 連結된 것으로 생각하면

$$Z_m = \frac{1 - K_m}{1 + K_m} Z_{m-1} \quad (5)$$

에 依해 斷面의 Z_m 을 推定할 수 있다.^[7] 여기서 K_m 은 反射係數로서

$$a_{m-1, i} = \frac{a_{m-i} - K_m a_{m, m-i}}{1 - K_m^2}$$

$$K_m = a_{m, m} \quad m = M, M-1, \dots, 1 \quad (6)$$

$$i = 0, 1, \dots, m-1$$

에 依하여 필터의 次數와 無關係하게 定해지고, 聲道의 길이를 17 cm 以下로 볼 수 있어 $M=12$ 로 하여 求한 反射係數中 K_m , $m=1, 2, \dots, 8$ 만을 使用, 聲道의 模樣을 推定하였다.

i) 聲道의 周波數 應答에 해당하는 $A_M(e^{j\omega})$ 는

$$\int_{-\pi}^{\pi} \ln |A_M(e^{j\omega})|^2 \frac{d\theta}{2\pi} = 0 \quad (7)$$

를 만족하므로^[8, chap 6] 正規化되어 있다. 그러므로 여기서는 1024 point FFT를 使用, 各音의 模型 스펙트럼을 구하였다.

j) 스펙트럼의 差異를 定量的으로 表現하기 위하여 코우더 (coder)의 評價指數로 使用되는 log likelihood ratio를 利用하였다.^[10] 즉 比較하고자 하는 두개의 스펙트럼 $A_R(e^{j\omega})$ 와 $A_T(e^{j\omega})$ 의 差는 各各에 對한 FFT와 式 (7)을 써서

$$\ell = \log \left[\frac{\int_{-\pi}^{\pi} \frac{A_T(e^{j\omega})}{A_R(e^{j\omega})} \left| \frac{d\omega}{2\pi} \right|^2}{\int_{-\pi}^{\pi} \frac{A_T(e^{j\frac{2\pi i}{k}})}{A_R(e^{j\frac{2\pi i}{k}})} \left| \frac{d\omega}{2\pi} \right|^2} \right] \approx \log \left[\frac{1}{k} \sum_{i=0}^{k-1} \left| \frac{A_T(e^{j\frac{2\pi i}{k}})}{A_R(e^{j\frac{2\pi i}{k}})} \right|^2 \right]$$

단, $k=1024$ (FFT의 point 數)

를 計算하여 그 尺度로 삼았다. 이것은 基準 필터 $A_R(z)$

에 對한 誤差信號의 에너지 α 와 比較되는 信號를 $A_R(z)$ 에 인가했을 때의 誤差信號 에너지 β 에 對해

$$\ell = \log \left(\frac{\beta}{\alpha} \right) \quad (9)$$

의 意味를 가진다.

III. 實驗結果 및 檢討

먼저 각 母音의 여러 音程에 對해 pitch를 抽出, 比較한 結果, 같은 音程에 對한 pitch 周波數는 發聲者나 音素에 關係없이 唯一하게 決定됨을 確認하였다. 그 結果는 表 1과 같고 이것은 音樂에서 定한 各音의 基本周波數와 같거나 (女子의 境遇) 그 半이 된다. (男子의 境遇)

다음에 音聲信號의 周波數 特性과 音程과의 關係를 알기 위해 “어”의 “아래술”에 該當하는 入力信號의 4ier 變換을 그림 2에 보이고 같은 母音에 對

표 1. Pitch 週期와 周波數

Table 1. Pitch period and frequency

	Pitch Period (no. of samples)		Pitch Freq. (Hz)
	man	woman	man
Low SOL	82	41	98
Low LA	73	36	110
Low SI	65	32	123
Mid DO	61	31	131
Mid RE	54	27	148
Mid MI	48	24	167
Mid FA	46	23	174
Mid SOL	41	20	195

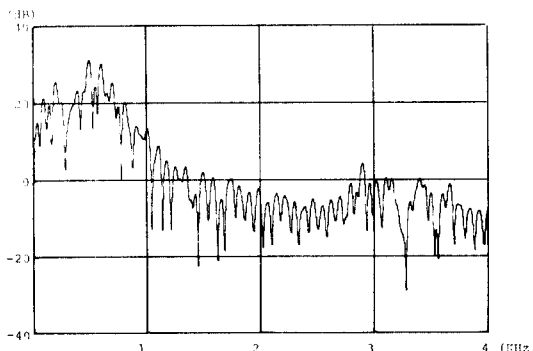


그림 2. “어” “아래술”에 對한 入力信號의 스펙트럼
Fig. 2. Input spectrum of “어” “Low Sol”.

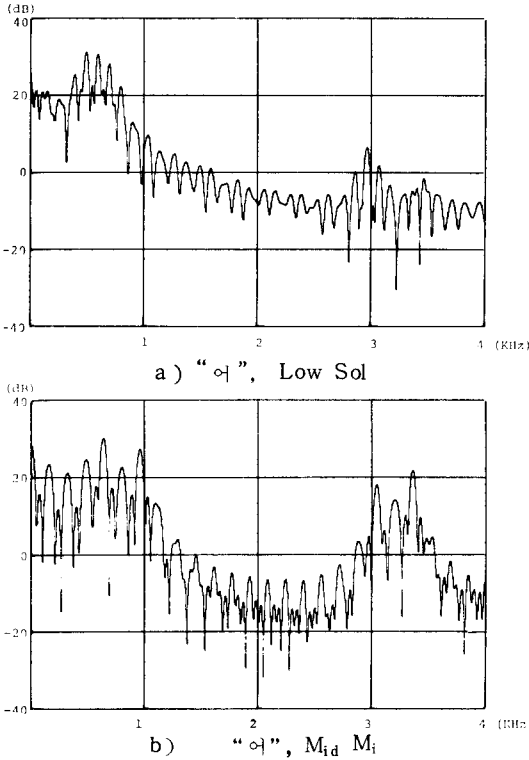
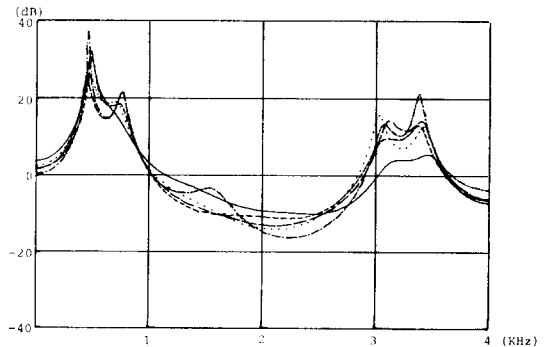
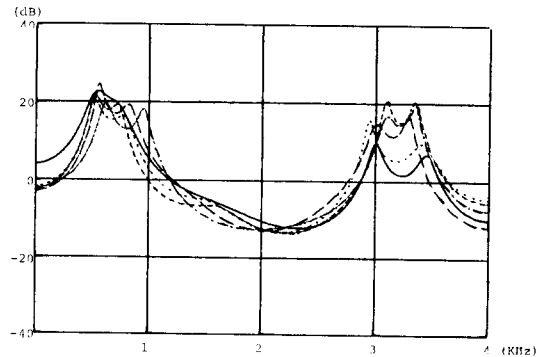
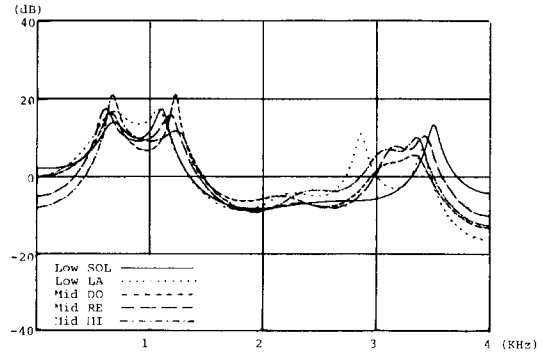


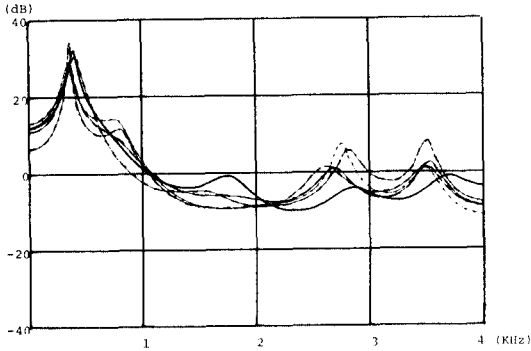
그림 3. 合成된 信號의 스펙트럼
Fig. 3. Spectrum of synthesized signals.

한 두개의 음程 “아래 솔”과 “가운데 미”에 대하여 그림 1의 시스템을 사용, 合成된 波의 fourier 變換을 그림 3에 보였다. 이 두개의 그림에서, 音聲에 對한 음程의 意味는 周波數 特性의 全般的인 形態에 依하지 않고, 部分的인 極大值 사이의 周波數 間隔에 依해서 決定되며 사람의 귀는 그것에 매우 敏感하다는 것을 알 수 있다. 音盤의 回轉數를 달리 하면 音程뿐 아니라 音聲 全體가 달라져 알아 듣기 힘들게 되는 것은 이러한 理由에 依한 것이다. 그러나 preemphasis 에 依한 效果를 考慮한 聲道의 周波數 應答特性은 信號 스펙트럼의 包絡線과 같으므로 高音과 低音을 發聲 할때 聲道의 形態가 달라지는 點을 감안하면 음程에 따라 스펙트럼의 包絡線도 變化될 것으로 豫想할 수 있다.

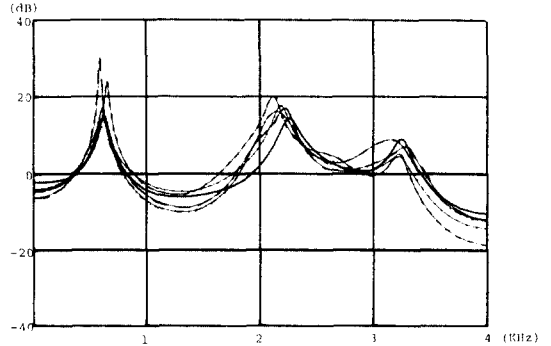
그러나 똑같은 音도 發聲者에 따라 많은 스펙트럼상의 差를 가지기 때문에 위의 變化를 알기 위해서는 同一한 사람의 音聲을 利用해야 한다. 그러므로 正確한 음程을 發音할 수 있는 20代 男性을 發聲者로 하여 韓國語 單母音 8개의 各各에 對하여 普通의 音聲에 많이 包含되는 “아래 솔”부터 “가운데 미”사이의

다섯 개의 음程을 分析하여 聲道의 周波數 應答를 얻은 것이 그림 4이다. 그림에서는 음程에 따라 매우 적은 變化를 보이는데, 우선 이 變化가 發聲狀態의 不規則性에 의한 것이 아님을 밝히고 또 定量的인 傾向을 分析하기 위해서 式(8)의 尺度를 使用한 結果는 다음과 같다.

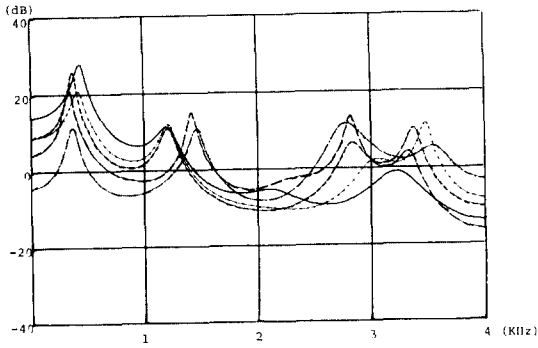




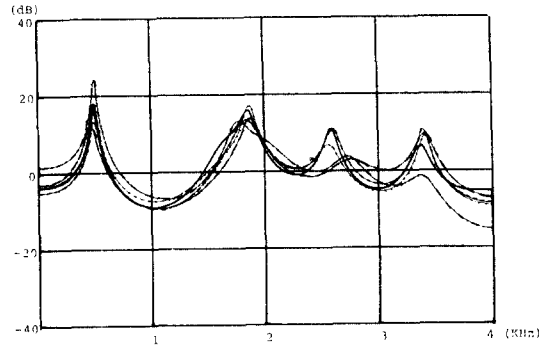
(d) “우”



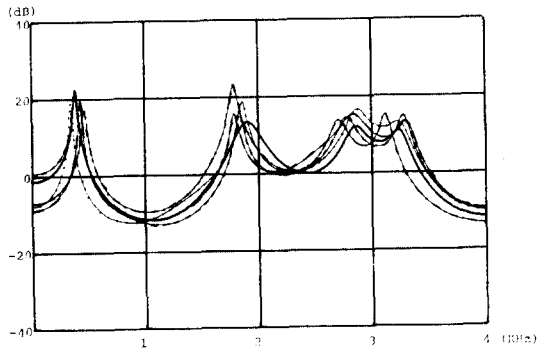
(g) “애”



(e) “으”



(h) “에”



(f) “이”

그림 4. 各 母音에 對한 綫型豫測 模型 스펙트럼
Fig. 4. Linear prediction model spectrum for each vowel.

數들 5 비트로 均一하게 量子化할때 생기는 스펙트럼의 變化 ℓ 을 計算한 結果, 모두 $10^{-4} \sim 10^{-3}$ 의 範圍에 있었는데 이와 比較하면 그 크기의 程度를 알 수 있다.

結局 그림 4는 音程에 따라 變化하는 聲道の 周波數 應答으로 볼 수 있는데 이를 觀察하여 대체로 다음의 事實을 알 수 있다.

a) Formant 周波數와 帶域幅은 音程에 따라 간단히 表現할 수 있는 傾向을 가지고 變化하지 않는다. 帶域幅이 가지는 意味로 볼 때 낮은 音에 대해 좁은 帶域幅을 가진 것으로 推定할 수 있으나 본 實驗結果에 비추어 역시 綫型豫測法에 依한 formant 帶域幅의 推定은 正確한 結果를 얻지 못함을 알 수 있다.^[4]

b) 높은 周波數 部分이 낮은 쪽에 비하여 變化가 큰 것은 에너지의 分布가 많은 部分의 變化가 적음을 意味한다.

서로 다른 時間에 發聲된 열개의 同一한 音의 同一한 音程사이의 ℓ 을 平均하여 0.07을 얻었고, 그림 4에 보인 各 스펙트럼을 母音에 따라 다르나 音程의 差가 클수록 ℓ 이 커지는 現象을 뚜렷이 볼 수 있으며, 대체로 한 音에 0.1 程度의 差가 加算되는 傾向을 보였다. 예를 들어 몇 개의 音을 擇하여 各各의 反射係

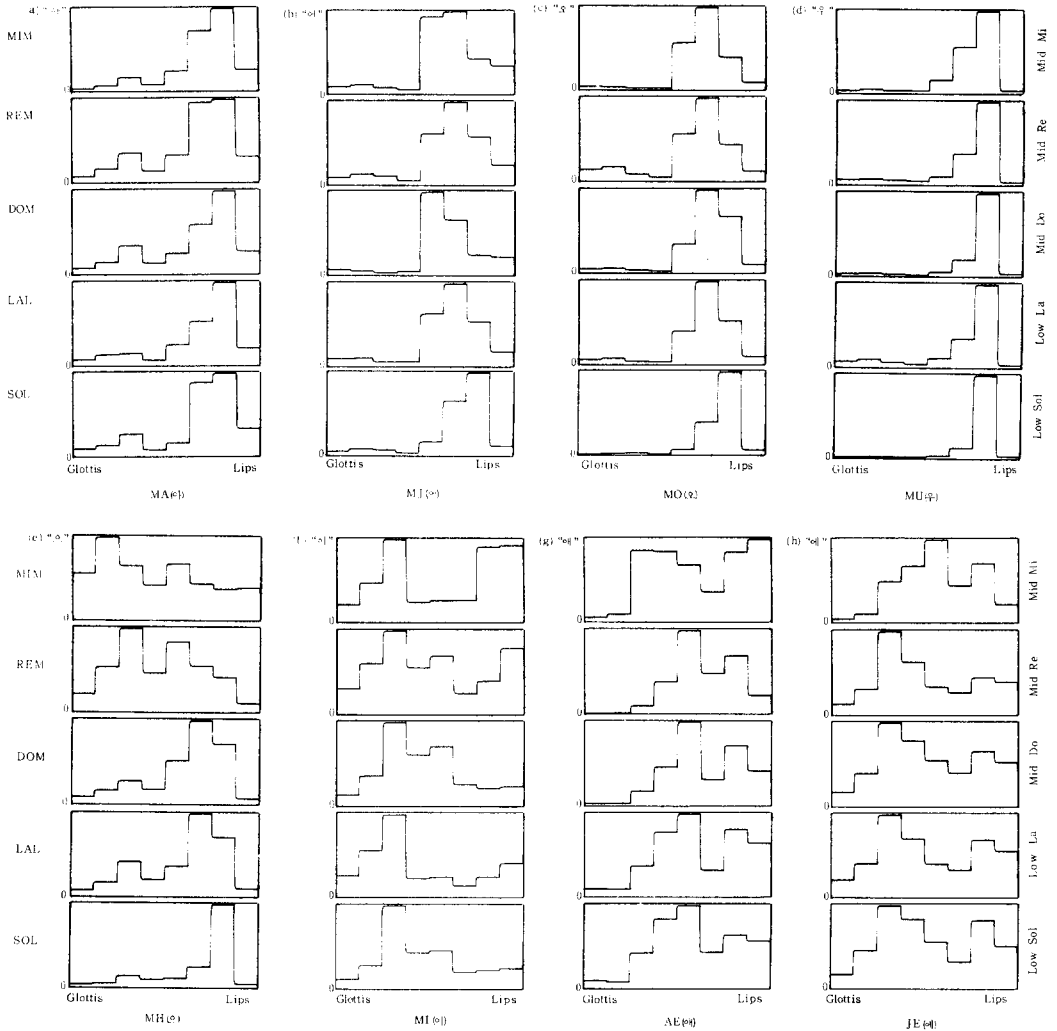


그림 5. 각 모음에 대해 推定된 聲道 斷面積
Fig. 5. Estimated vocal tract area for each vowel.

c) 모음의 種類에 따라 變化되는 傾向이 다르다.

그러므로 聲道の 周波數 應答에 依하여 各 音程의 特徵을 밝히는 것은 어려운 問題이다. 그러나 前記한 바와 같이 音程에 따른 파라미터의 變化는 聲道の 形態에 起因한다고 볼 수 있으므로 이를 分析하기 위하여 式 (5)를 使用, 聲道の 形態를 推定한 結果 그림5를 얻었다. 그림 5는 그림 4에 比하여 變化가 뚜렷하며 그 傾向을 쉽게 把握할 수 있다. 勿論 式 (5)을 얻는 過程에서 취한 假定과 聲道の 길이에 依한 誤差等으로 正確하다고 할 수는 없으나 그림 5는 實際의 發聲狀況과 매우 類似한 傾向을 보이고 있다.

以上을 綜合하면 사람은 聲帶의 振動周波數를 調節

하기 위하여 自然스럽게 聲道の 形態를 變化시키나 聲道の 周波數 應答에는 큰 影響이 없도록 하는 것으로 判斷된다.

다음에는 위의 結果를 音聲合成에 利用하는 方案으로서 各 모음의 한 音を 分析하여 얻은 豫測係數를 使用하고 pitch 周波數만을 各 音程에 맞도록 變化시켜 다른 音程의 信號를 合成함으로써 같은 파라미터로서 合成될 수 있는 音의 範圍를 알아 보았다. 그 結果 特性上 明確히 規定할 수 없고 또 모음에 따라 다르다고 생각되나 聽覺에 依存한 試驗 結果 半옥타브 内에서는 거의 區別할 수 없고 한 옥타브 程度까지도 認識에는 지장이 없는 程度로 判明되었다. 더 넓은 音域

에 適用할때 音質이 低下되는 理由를 時間軸에서 解析하면 하나의 pitch 펄스에 對한 應答 波形的 持續 時間이 너무 짧거나 길어서 原來의 波형과 다르게 되기 때문이며 이는 前述한 formant 帶域幅과 密接한 關係가 있는 것으로 보인다.

普通的 對話에 나타나는 音程의 範圍가 한 옥타브 程度임을 考慮할 때, 위의 結果는 各 單母音 音素에 對하여 하나 또는 두개의 音程에 對한 파라메터만을 記憶시켜 모든 音程의 母音을 自然스럽게 合成할 수 있음을 말한다.

이의 實證을 爲하여 對話의 一部分을 分析한 파라메터중 母音에 속하는 프레임의 豫測係數를 一定한 音程의 正常狀態 母音에서 얻은 똑같은 豫測係數로서 代置하여 合成해 본 結果 前述한 內容과 一致함을 보았다. 이로써 母音에 對하여 音素 組合에 依한 音聲合成의 可能性을 實證한 것으로 볼 수 있다.

參考로 위의 두가지 實驗에서 얻은 合成信號를 本稿에 同封한 錄音 테이프에 收錄하였다.

IV 結 論

音聲信號가 가지는 不規則性和 合成된 信號의 質은 評價하는 基準이 聽覺에 依存할 수 밖에 없으므로 因하여 明確한 結果를 얻기 어려우나 대체로 本 實驗의 結果를 要約하면 다음과 같다.

a) Pitch 周波數는 發聲者나 母音의 種類에 關係없이 音程에 依해 決定된다.

b) 同一한 母音의 다른 音程을 發生하기 위해 사람은 聲道의 形態를 꽤 많이 變化시키거나 周波數 特性으로는 그 變化를 觀察하기가 容易하지 않다.

c) 同一한 母音의 合成을 위하여 半 옥타브 또는 한 옥타브 內에서는 같은 豫測係數를 使用하여도 들기에 어색함이 없다.

以上的 結果를 使用하고 聲道形態의 時間的인 變化를 감안하여 音素사이의 過渡狀態를 解決하면 各 音素에 대한 적은 量의 情報를 利用, 組合함으로써 많은 어휘의 音聲을 合成할 수 있을 것으로 期待된다. 이를 爲하여 子音에 關해서도 이와 같은 分析이 必要하다고 생각된다.

參 考 文 獻

- Saito S., Itakura F. ; The Theoretical Consideration of Statistically Optimum Methods for Speech Spectral Density. Report No.3107, Electrical Communication Laboratory, N. T. T., Tokyo(1966).
- Atal B. S., Schroeder, M. R. ; Predictive Coding of Speech Signals. Proc. 1967. Conf. Communi. and Process, pp. 360~361(1967).
- Atal B. S. ; Speech Analysis and Synthesis by Linear Prediction of the Speech Wave, J. Acous. Soc. Am. 47, 65 (A) (1970).
- L. R. Rabiner, R.W. Schafer : Digital Processing of Speech Signals, Prentice - Hall (1978).
- L. R. Rabiner, R. W. Schafer ; Digital Techniques for Computer Voice Response: Implementations and Applications, proc. IEEE, Vol. 64, pp 416 ~ 433, April 1976.
- Jonathan Allen; Synthesis of Speech from Unrestricted Text, proc. IEEE, Vol. 64, pp. 433~442, April, 1976.
- B. S. Atal, S. L. Hanauer ; Speech Analysis and Synthesis by Linear Prediction of the Speech Wave, J. Acoustics. Soc. Am. Vol. 50, pp. 637~655, Aug. 1971.
- J. D. Markel, A. H. Gray ; Linear Prediction of Speech, Springer - Verlag, 1976, New York.
- M. J. Ross, H. L. Schaffer, A. Cohen, R. Freudberg, and H. J. Manley ; Average Magnitude Difference Function Pitch Extractor, IEEE Trans. Acoust. Speech and Signal Proc. Vol. ASSP- 22, pp. 353~362, October, 1974.
- R. J. Mc Aulay ; Maximum Likelihood Spectral Estimation Using State Variable Techniques, proc. RADC Spectrum Estimation Workshop, May, 1978, pp. 63 ~ 68.
- R. E. Crochiere, J. M. Tribolet, L. R. Rabiner ; An Interpretation of the Log Likelihood Ratio as a Measure of Wave from Coder Performance, IEEE Trans. on ASSP, Vol. ASSP-28, No. 3, June : 1980. pp. 318 ~ 323.
- C. K. Un, et al. ; Voice Band Digital Encoding and Transmission, SRI Project 3733 Report, Stanford Research Institute, June, 1975.
- 吳永換 ; 基本母音의 線形豫測 스펙트럼에 미치는 豫測次數의 影響, 대한 전자공학회 추계종합 학술 발표대회 논문집, Vol. 3, No. 2, 1980.