

技術解説

고음질 합성을 위한 피치변경법

On the Pitch Alteration Methods for a
High Quality Speech Synthesis배 명 진
(Myung Jin Bae)* 이 연구는 1992년도 한국과학재단 연구비 지원비에 의한 결과임.
* 과제번호 : 92-21-00-06

요 약

고음질 합성을 위해서는 과형부호화법이 바람직하다. 과형부호화법을 규칙에 의한 음성합성기법에 적용하기 위해서는 메모리 용량의 문제와 피치변경법이 해결되어야 한다. 메모리 용량의 문제는 최근 반도체 기술에 의해 극복되어 왔으며 이제 음원피치변경의 문제가 남아있다. 따라서 본 논문에서는 성도 포먼트의 특성은 변화시키지 않고, 음원피치를 변경시키는 문제에 대해 정리하였다. 먼저 기존에 제안된 몇가지 기법들의 장단점들을 열거한 다음에, 우리 연구실에서 제안했던 방법들에 대해 논의하고자 한다.

I. 피치변경의 필요성

음성합성시스템은 그 기준에 따라 여러가지로 분류할 수 있는데 먼저 분석된 데이터를 그대로 합성해 이용하는 분석에 의한 합성(synthesis by analysis) 법과 규칙에 의해 음성을 발생시켜 합성하는 규칙에 의한 합성(synthesis by rule) 법으로 구분할 수 있다 [14].

데이터 규모에 따른 합성단위로는 文章單位, 單語單位, 音節單位, 變異音單位, 그리고 音素單位로 나눌 수 있다. 예를들면 家電製品의 使用法案內나 錄音案內事項에 대해서는 보통 文章單位로 합성을 하고, 時間 및 氣候案內에는 單語單位의 합성법이 適用된다. 또한 狀況이 複雜하고 모든 分野에 適用되어야

한 合成技法은 音節以下의 合成單位가 바람직하게 된다[14-15][18][21].

합성을 위해 出力되는 데이터의 處理方式에 따라서는 메모리형 合成法과 傳送型 合成法으로 區分할 수 있다. 메모리형 合成法은 分析한 데이터를 메모리에 貯藏시켜두고 必要에 따라 다시 合成시키는 方法으로, 모저(Mozer)法, 피치單位分節法 등이 있다 [14][18]. 따라서 메모리에 貯藏되는 데이터形式이 畚地當 一定비트數로 制限되지않으며, 分析時의 時間은 合成時의 時間과 반드시 一致시킬 必要도 없다. 또한 메모리의 誤譯率은 無視될 程度이기 때문에 이를 防止하기 위한 別途의 技術이 거의 要求되지 않는다.

한편 傳送型 合成法은 傳送채널의 效率을 높이기

위해 分析후 바로 傳送을 目標로하여 處理하고 受信측에서 다시 調合하여 合成하는 方法이다. 이러한 合成法은 公衆電話網의 보코더(voice coder decoder) 시스템에 오랫동안 應用되어 왔다. 따라서 傳送到에 適合하도록 데이터량이나 그 處理時間이 單位時間當一定하게 處理되어야 하며, 傳送到중에 받게될 背景雜音의 影響에 比較的 強靱해야 한다[24-25].

지금까지 傳送型 合成法으로 研究된 方法들은 크게 波形符號化, 信號源符號化, 混成符號化法이 있으며[18][22], 貯藏메모리나 전송채널의 帶域幅節約을 위해서는 信號源符號化法을, 音質을 높이기 위해서는 波形符號化法을 주로 使用하고 있다.

波形符號化法은 音聲情報을 發聲모델에 따라 分離하지않고 波形自體의 殘餘成分을 除去한 後에 符號化하는 方法이며 PCM, ADPCM, ADM등이 提案되어져 있다[6]. 最近에는 디지털信號處理專用칩의 製造技術과 波形符號化法의 分析 및 合成알고리즘이 잘 開發되어 32Kbps傳送率을 갖는 ADPCM의 標準化가 實現되어 졌다. 그렇지만 波形符號化法은 人間の 個性과 感情을 나타내는 勵起(excitation)情報과 意思傳達을 나타내는 聲道濾波器의 포먼트(formants)情報을 分離하지 않고 處理하기 때문에 音源을 變更시켜야 하는 音節單位나 音素單位의 規則에 의한 合成技法으로는 바람직하지 못하다.

信號源符號化法은 勵起情報과 濾波器情報을 分析時에 分離시켜서 獨立的으로 符號化하는 方法으로서 LPC, PARCOR, LSP 등의 알고리즘이 提案되어 있다[16][22]. 이들 알고리즘은 傳送率을 10Kbps 以內로 낮출 수 있기 때문에 채널帶域幅이나 메모리效率의인 符號化法이 된다. 또한 分析時에 抽出된 生成모델의 勵起情報나 濾波器情報를 合成時에 人爲的으로 變更시킬 수 있기 때문에 音節單位나 音素單位의 規則에 의한 合成技法에 適用이 容易하다. 그러나 分析時에 각 成分을 分離하고, 다시 그 情報를 利用해서 合成하기 때문에 分析時의 誤差와 合成時의 誤差가 합해져서 合成音質은 自然性이나 明瞭性이 크게 떨어지게 된다.

信號源符號化에서의 메모리效率성과 波形符號化의 明瞭성과 自然성을 適當히 維持하기 위해 이 두가지 符號化法을 混成시킨 方法이 混成符號化法이며, VELP, MPLPC, RELP, CELP 등의 方法이 提案되어져 있다[24]. 그렇지만 混成符號化法에서는 聲道の 濾波器情報를 符號化하는데 信號源符號化法을 適用하고, 勵起情報의 符號化에는 波形符號化法을

주로 適用하고 있다. 이 때문에 勵起情報를 變更시켜야 하는 音節單位나 音素單位의 合成法에 適用하기에는 어려움이 따른다.

最近 電子工學의 눈부신 發達로 인해 메모리의 製造技術은 10년동안 集積化率을 1000배로 成長시켰다. 16M-비트의 容量을 갖는 메모리를 音聲合成에 利用할 境遇에, 바이트단위로 쓰기 위해 8개를 使用하면 64kbps의 對數-PCM 波形符號化法으로 合成하여도 $(16 \times 10^6) \times 8 / (64 \times 10^3) = 2000$ 초의 音聲 데이터를 貯藏할 수 있는 많은 量이 된다. 따라서 단지 메모리량을 줄이기 위해 信號源符號化用 合成法을 採擇하는 것은 바람직하지 못하다[36][40].

大部分의 音聲合成 應用分野에서는 合成單位를 한 가지로 固定해서 使用하지 않고 몇가지 合成技法들을 混成하여 適用하고 있다. 예를들어 保險會社의 納入回數를 照會하거나 銀行의 殘額을 照會하는 등의 境遇에 있어서, 全般的인 節次나 口常의 案內에 대해서는 文章單位로 分析에 의한 合成法에 의해 放送하고, 殘額이나 納入回數 등에는 單語나 音節單位의 規則에 의한 合成法을 통해 放送한다. 이러한 境遇, 보통 文章單位의 合成에는 明瞭성을 維持하기 위해 波形符號化法이나 混成符號化法을, 그리고 單語나 音節單位의 合成에는 規則合成이 가능하겠끔 信號源符號化法을 使用해야 한다[14]. 이 때문에 簡單한 應用에 대해서도 同一한 데이터베이스를 使用할 수 없게 된다. 또한 이러한 境遇에 波形符號化法과 信號源符號化法이 連結되는 部分에서는 波形符號化의 音源피치가 自然스럽게 連結되지 못하고, 信號源符號化에서는 合成時에 收斂速度가 要求되기 때문에 自然성과 明瞭성이 低下될 수도 있다. 이러한 問題點을 解決하려면 波形符號化法에 의해 規則合成이 可能하도록 波形符號化法에서 音源피치를 變更시킬 수 있어야 한다.

따라서 本 論文에서는 波形符號化法에 의해 規則合成이 可能하도록 音源피치 變更에 대한 技法들을 研究하였다. 우선 既存에 提案된 波形符號化法에서의 피치變更法들을 簡單히 紹介하고 그 限界성을 나눈 다음에 本 研究室에서 波形符號化法과 피치檢出法들의 原理를 利用하여 提案했던 두가지 피치變更法들을 整理하여 說明하고자 한다.

그 內容들은 波形的 週期를 簡單히 半分하는 週期半分法을 適用하여 時間領域上에서 波形的 피치週期를 變更하는 피치週期 半分法과 時間-周波數領域의 變換特性을 利用한 켄스트립 피치變更法이다. 이들

에 대해 原理와 處理過程을 자세히 說明하면서, 實驗考察을 통해 長短點과 向後 改善해야할 內容들을 다루게 된다.

II. 既存의 피치週期變更法

一般的인 線型豫測시스템에서는 勵起임펄스들의 時間間隔을 變更함으로써 有聲音의 피치가 簡單히 變更되고, 持續時間(duration)은 豫測係數의 更新率을 變更시킴으로 바뀌게 된다. Caspers와 Atal은 이러한 線型豫測技法의 觀點에서, 멀티펄스勵起源의 피치週기를 變更하기 위해 영·追加나 削除 技法을 適用하였다. 또한 持續時間은 피치週기들을 追加 또는 除去함으로 變更하였다[30].

그렇지만 混成符號化法인 멀티펄스-LPC法에서는 임펄스들 사이의 間隔調節에 의해 피치가 簡單히 變更될 수 없다. 왜냐하면 멀티펄스LPC法에서는 각 임펄스들이 이전에 찾아진 모든 임펄스들의 效果를 考慮하여 最少의 誤謬가 되도록 計算되기 때문이다. 임펄스位置의 變更은 合成波形的 歪曲을 招來할 뿐이다.

Stella와 Charpentier는 멀티펄스勵起源을 聲道 모델과 勵起源의 한 結合으로 보는 觀點에서, 이러한 問題를 解決하였다. 그들은 우선 멀티펄스分析法으로 다이폰(diphone)單位的 데이터壓縮을 하였고, 合成時에는 原來形態로 다이폰合成을 한 다음에, 位相(phase)보코더를 適用하여 韻律을 變更하였다[31].

Varga와 Fallside는 波形 編輯部分과 線型豫測에 의한 合成部分으로 構成된 混成技法을 適用하여 波形符號化의 피치變更을 試圖하였다[32]. 피치週기를 延長하기 위해 音聲標本의 一部分과 線型豫測 濾波器係數를 사용하여 一般的인 合成濾波器에 의해 合成하고 連結部分을 平滑化하였다. 合成濾波器的 初期條件으로 이들을 使用하면, 合成濾波器는 別途의 새로운 勵起源없이 필요한 音聲의 나머지部分을 簡單히 發生할 수 있지만, 피치週기를 줄이는 境遇에는 單純이 한 피치區間波形的 一部分을 除去하고 平滑化하는 것으로 피치變更을 試圖하였기 때문에 피치가 짧은 女性이나 어린 話者의 境遇에는 스펙트럼歪曲이 많아진다.

Quatieri와 McAulay는 그림 2-1과 같이 周波數領域에서 位相을 保存하는 피치변경법을 提案하였는데 [39] 이것은 入力된 音聲에 대해 振幅 및 位相스펙트럼을 抽出하여 別途로 處理하는 方法이다. 振幅스펙

트럼에 대해서는 두드러진 스펙트럼 봉우리들을 抽出한 다음에 이것을 피치變更率(ρ) 만큼 인터플레이션하여 振幅스펙트럼의 피치를 變更시킨다. 位相스펙트럼에 대해서는 時間領域에서 구한 피치 開始時間(pitch onset time)에 該當하는 位相을 除去하고 나서 피치가 變更되었을 때의 새로운 피치 開始時間의 位相을 더해줌으로서 새로운 位相을 構成하게 된다.

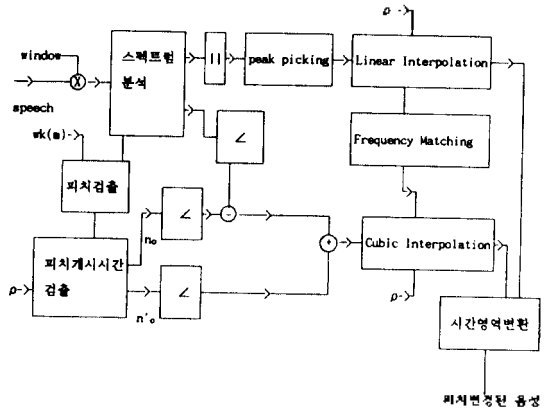


그림 2-1. 피치변경에 대한 분석 및 합성
Fig 2-1. Analysis and synthesis for pitch alteration.

이러한 方法은 波形的 꼴을 그대로 維持하기 때문에 프레임單位로 分析處理하는 通常의 處理法에서 隣近 프레임간의 連結이 아주 容易해 진다는 長點이 있다. 그렇지만 피치變更時에 피치週기와는 별도로 피치의 開始時間을 供給해 주어야 하고, 또한 振幅스펙트럼상에서 두드러진 봉우리 爲 1로 하모닉스의 인터플레이션을 遂行하기 때문에 스펙트럼의 歪曲이 높아진다는 短點이 있다.

III. 피치半分法에 의한 피치變更

이 方法은 音聲의 發聲모델에 根據하여 人爲的으로 變更하려는 피치週기의 2배波형을 線型豫測合成法에 의해 만든 다음에 그 波形的 週기를 半分하는 技法을 適用하였다[33][35]. 먼저 音聲信號의 發聲 모델에 대해 알아보고 여기서 音聲의 基本週기를 半分하는 波形的 半分技法과 함께 피치制御法에 대해

說明하기로 한다. 그다음 實際의 音聲에 대해 처리한 結果를 提示하기로 한다.

1. 有聲音의 피치半分法

有聲音의 振幅스펙트럼 S(K)는 基本周波數 Fo의 高調波마다 값이 存在하는 라인 스펙트럼의 形態를 띤다. 發聲모델에 따라 聲門, 聲道の 特性 H(K)와 이것을 勵起하는 聲帶의 振動特性을 E(K)라 하면, 有聲音의 스펙트럼 S(K)는 다음과 같이 類推할 수 있다.

$$S(k) = E(K) H(K) = \sum_{l=0}^N \delta(k-lFo) \cdot H(k) \quad (3-1)$$

여기에서 有聲音의 基本周波數 Fo를 알고있다면 基本周波數가 2배로 늘어난 有聲音의 스펙트럼 S'(k)를 다음과 같이 구할 수 있다.

$$S'(K) = S(K) \cdot \sum_{l=0}^{N/2} \delta(k-2lFo) \\ = \sum_{l=0}^N \delta(k-2lFo) \cdot H(k) \quad (3-2)$$

스펙트럼 S'(K)는 原來的의 有聲音스펙트럼 S(K)에서 基本周波數를 2배로 늘린 것이 된다. 주어진 有聲音에 대해 基本周波數를 2배로 늘리는 것은 時間領域에서 有聲音의 피치를 半分(halving)하는 것이 된다. 時間·周波數關係에 따라 有聲音의 波形 s(n)에서 피치 p를 半分해 보면 다음과 같다.

$$s'(n) = s(n) \times \sum_{l=0}^p \delta(n \cdot lp/2) \quad (3-3)$$

이 된다. 여기서 有聲音의 피치는 $p = 1/Fo$ 이고 다른 피치검출법들을 통해 이미 이에 대해 알고 있다고 假定한다. 또한 有聲音 s(n)은 피치單位로 週期函數이므로 식(3-3)을 다시 쓰면 다음과 같이 簡略化될 수 있다.

$$s'(n) = s(n) + s(n-p/2) + s(n-p) + s(n-1.5p) + \dots + \\ = P[s(n) + s(n-p/2)] \quad (3-4)$$

여기서 P는 現在의 音聲프레임에서 피치週期 反復回數이다. 이렇게 時間領域上에서 音聲信號의 피치를 半分하는 方法을 지금부터는 피치週期半分(pitch ha-

lving)法이라고 定義한다.

2. 波形에서의 피치 週期伸張

피치半分法은 有聲音의 피치를 2의 指數函數로 줄일 수 있지만, 그 사이의 週기로 變更시키기는 어렵다. 이 때문에 半分法을 使用하여 피치週기를 連續的으로 變更시키는 法을 提案해야 한다. 피치半分法을 適用하여 피치를 線型的으로 줄이려면 피치를 늘리고나서 피치 半分하면 된다. 有聲音의 피치를 p, 늘리는 길이를 L이라하면 피치半分法을 함께 使用하여 줄일 수 있는 피치區間 p'은 다음과 같다.

$$p' = (p + L)/2 \\ = p/2 + L/2 \quad (3-5)$$

따라서 늘리는 標本數 L을 調節하면 現在의 피치와 半分된 피치사이의 값으로 피치週기를 變更시킬 수 있게 된다.

이제는 피치週기를 L-標本만큼 늘리는 方法에 대해 考慮한다. 有聲音의 스펙트럼에서 각 포먼트의 音우리는 얼마간의 帶域幅을 갖게된다. 이것을 時間領域에서 살펴보면 한 피치週期內에서의 포먼트를 週기로 하여 發振하면서, 포먼트帶域幅에 의해 時間에 따라 制動이 發生한다. 이것을 시스템적인 側面에서 考慮하면 聲道の 調音氣管은 安定된 시스템이기 때문에 聲帶의 振動으로 勵起되어진 다음에 一定時間이 經過하면 漸次減衰되고 더 이상의 勵起가 없으면 音聲波形이 영에 到達하게 된다.

이 때문에 有聲音波形은 피치區間 사이에 聲道の 共鳴現象이 나타나며 이것은 安定시스템인 聲道の 特性을 나타내기 때문에 다음피치가 나타날 때까지는 그 波形의 振幅이 漸次減衰하는 模樣이 된다. 즉, 有聲音의 다음피치가 發生하기 前에 波形振幅은 기의 영에 近接하게 됨으로, 피치週기를 늘리려면 이 部分에 영을 插入하면 되고, 이 境遇에 스펙트럼에 歪曲을 最少化할 수 있게 된다.

3. 線型豫測合成에 의한 피치週期の伸張

피치를 늘리기 위해 週期の 끝部分에 單純이 영값을 插入하게 되면, 安定된 聲道の 特性을 充分히 나타내기前에 聲門의 새로운 勵起가 始作되는 發聲의 境遇에는 明瞭性이 크게 低下될 수 있다. 이러한 境遇의 例로는 짧은 피치를 갖는 女性이나 어린이 發聲 그리고 聲道の 길이가 다른 發聲에 비해 길게 모델링

되는 鼻音 또는 有聲音 /이/의 波形을 들 수 있다. 이때 스펙트럼의 變曲을 最少化하는 한 方法으로는 피치를 늘리는 部分에 영값을 넣지않고 聲道の 特性을 延長시켜주면 된다.

이제 피치週期를 늘리는 方法을 考慮한다. 音聲信號는 生成모델에 根據하여 線型豫測에 의해 다음과 같이 all-pole 모델로 合成될 수 있다.

$$\hat{s}(n) = e(n) + \sum_{i=1}^M a_i s(n-i) \quad (3.6)$$

여기서 $e(n)$ 은 過去值들의 線型組合에 의해 現在の 標本값이 豫測될 때 나타나는 豫測誤謬이지만, 合成 시스템에서는 聲門의 特性을 나타내는 勵起源으로 分類할 수 있다. 有聲音의 勵起源은 피치週期の 頂點 스퀘어로 近似되며, 係數 a_i 는 聲道濾波器의 特性을 나타낸다. 故로 有聲音을 發聲하기 위해 聲門이 열려 한 피치週期가 始作되는 位置에서는 豫測誤謬가 最大값이 되고, 다음 피치가 始作되기 以前의 部分에서 豫測誤謬값은 거의 영이 된다.

따라서 한 피치區間의 끝部分은 聲道の 特性을 支配적으로 나타냄으로써 勵起源을 意味하는 誤謬값은 영으로 近似될 수 있고, 또한 이 部分의 音聲標本값을 식 3-6의 過去값 $s(n-i)$ 으로 하면 피치週期가 延長된 새로운 音聲標本값이 豫測될 수 있게 된다.

4. 結果考察

各 音聲試料에 대해 그림 3-1과 같이 處理하였다. 먼저 피치 p 를 구한 다음에 變更된 피치 p' 을 얻기 위해서는 $L = 2(p' - p/2)$ 개의 標本값을 線型豫測 合成法으로 만들어 한 피치가 끝나는 곳에 插入한다. 그 다음에 피치半分法을 適用하면 피치週期가 調節된 波形이 얻어지게 된다.

評價의 目的으로 피치週期 伸張時에 영값을 代入했을 境遇의 스펙트럼歪曲과 線型豫測 合成法으로 波形을 生成했을 境遇의 스펙트럼歪曲을 남여 各各에 대해 測定하여 表 3-1에 提示하였다. 이때 스펙트럼값은 原來 스펙트럼과 比較되었으며, 比較하기 前에 quefrequency에서 피치週期の 80% 以上成分을 除去한 後에, 다음 式에 의해 스펙트럼의 類似性을 百分率로 測定하였다.

$$R(fr) = \frac{\sigma_{os}}{[\sigma_o \sigma_s]^{1/2}} \times 100 \quad (3.7)$$

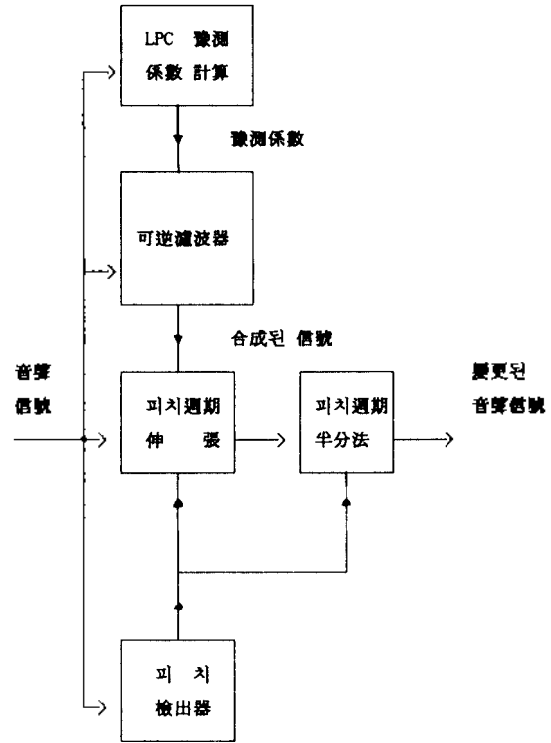


그림 3-1. 피치變更에 대해 提案한 處理過程의 블록도
Fig 3-1. Processing block diagram for altering the pitch of voiced speech.

여기서 σ_o 는 原來音聲의 聲道스펙트럼의 標準偏差이고, σ_s 는 피치變更시킨 聲道스펙트럼의 標準偏差이며, σ_{os} 는 두 스펙트럼사이의 偏差값이다.

線型豫測 合成法으로 生成된 標本값을 피치週期の 200%로 늘리고자하는 部分에 插入했을 境遇가 영값을 插入했을때 보다는 약 5% 程度의 스펙트럼歪曲이 改善되었다. 또한 女性話者의 境遇에 비해 男性話者의 發聲에서 1% 정도가 優秀하게 얻어졌다.

마찬가지로 피치를 줄이는 境遇에 대해서도 스펙트럼歪曲의 程度를 남여 各各에 대해 測定하여 表 3-2에 提示하였다. 피치週期를 壓縮한 境遇에 대해서는 피치의 끝部分을 부조간 削除했을 境遇와 本 研究에서 提案한 線型豫測 合成과 週期半分法을 適用한 境遇를 比較하였다. 피치週期를 80%로 壓縮하였을 때 半分法에 비해 스펙트럼歪曲을 平均 8% 程度로 改善

하였고, 女性話者 보다는 男性話者인 境遇가 3% 程度로 優秀하게 얻어졌다.

표 3-1. 男性 및 女性話者에 대해 피치를 200%로 신장시킨 境遇의 스펙트럼곡

Table 3-1. Spectrum distortion measurement for altering the pitch of a male and a female speaker as 200%

發聲者: 28세 男性話者와 24세의 女性話者(數字音 5번씩 發聲)

발 성	200% 피치신장		
	zero 삽입	LPC합성삽입	개 선(%)
남 성	87.85	93.58	+ 5.73
여 성	92.38	96.58	+ 4.2

표 3-2. 男性 및 女性話者에 대해 피치를 80%로 줄인境遇의 스펙트럼곡

Table 3-2. Spectrum distortion measurement for altering the pitch of a male and a female speaker as 80%.

發聲者: 28세 男性話者와 24세의 女性話者(數字音 5번씩 發聲)

발 성	80% 피치압축		
	삭 제 法	주기반분법	개 선(%)
남 성	77.73	87.42	+ 9.69
여 성	78.48	84.69	+ 6.21

IV. 켈스트럼分析에 의한 피치變更法

이 方法은 quefrequency의 낮은 部分이 聲道의 特性을 나타내면서 켈스트럼의 振幅이 빨리 減少하고, 또한 피치週期 가까이에서는 켈스트럼振幅이 거의 영이되는 性質을 利用하여 피치를 變更하는 것이다 [34][36]. 따라서 變更하려는 피치週期만큼의 영값을 quefrequency上的 거의 영이되는 部分에 插入 또는 削除하는 技法을 適用하였다. 먼저 音聲信號의 發聲 모델에 따른 컨벌루션의 性質에 대해 알아보고, 호모머피處理를 통해 다-컨벌루션을 취하는 켈스트럼分析技法과 그 特性에 대해 알아본다. 그런 다음 音聲의 基本週期를 變更하는 피치制御法에 대해 說明하면서 實際의 音聲에 대해 處理한 結果를 考察하기로 한다.

1. 켈스트럼의 性質

音聲信號의 信號源符號化法으로 켈스트럼分析을 使用하는 것은 켈스트럼상에서 音源情報과 濾波器情

報를 쉽게 分離해서 符號化할 수 있기 때문이다. 그림 4-1에서 처럼 켈스트럼의 낮은 時間部分은 聲道, 聲門필스와 放射情報를 나타내며, 높은 時間部分은 勵起로 인한 것이다. 켈스트럼의 必要한 部分抽出은 周波數不變 線型濾波器라고 하는 다음의 켈스트럼窓函數 $l(n)$ 에 의해 選擇할 수 있다.

$$l(n) = \begin{cases} =1 & |n| < n_0 \\ =0 & |n| > n_0 \end{cases} \quad (4-1)$$

여기서 n_0 는 피치週期 N_p 보다 적게 選擇된다. 平滑化된 對數함수는 켈스트럼에 窓函數를 適用하여 구할 수 있다. 이 平滑化된 對數스펙트럼은 入力音聲프레임의 共鳴構造를 나타내고 線型的으로 포먼트周波數와 一致한다. 또한 聲道, 聲門필스와 放射效果가 調화된 特性을 나타내는 켈스트럼의 낮은 時間成分은 quefrequency의 增加에 따라 빠르게 減少된다.

결국 窓函數가 適用된 複素켈스트럼은 必要한 成分을 復元하기위해 逆特性시스템에 의해 處理되어야 한다. 이 때의 波形은 입력스應答 $h(n)$ 을 나타내게 된다.

만일 켈스트럼窓函數 $l(n)$ 을 勵起成分이 남도록 다음과 같이,

$$l(n) = \begin{cases} =0, & |n| < n_0 \\ =1, & |n| \geq n_0 \end{cases} \quad (4-2)$$

選擇한다면, 봉우리에 대한 높은 時間部分만을 調査하는 것이 되며, 이 部分에서 피치와 有聲音의 推定이 可能해 진다.

有聲音인 境遇에 音聲프레임의 基本週期位置에서 켈스트럼에 봉우리가 나타나지만, 無聲音의 켈스트럼에는 이러한 봉우리가 나타나지 않는다. 이러한 特性을 이용하여 지금의 音聲프레임이 有聲 또는 無聲音인지를 決定할 수 있고, 有聲音의 基本週期檢出을 可能하게 해 준다.

그렇지만 有聲音이라도 켈스트럼상에 뚜렷한 봉우리가 없거나 낮은 레벨이 될 수 있다. 이것은 켈스트럼 計算過程이 音聲波形의 一部區間만을 취해서 處理해야 하기 때문에 窓函數를 適用해야 하며 이에 따른 窓函數 길이나 形態의 影響을 받게 된다. 즉, 窓函數 길이가 音聲피치의 半倍길이보다 짧다면, 켈스트럼상에 週期性이 나타날 수 없게 된다. 한편 窓函數가 피치에 비해 길어지면 켈스트럼振幅의 始作에서

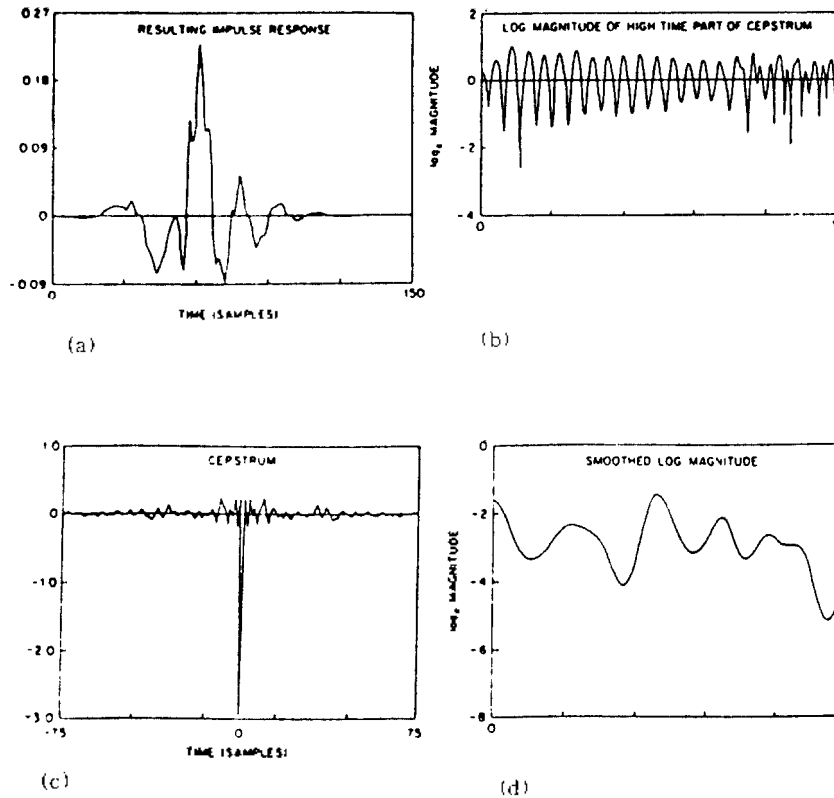


그림 4-1. 켈스트럼에 대한信號源符號化
Fig 4-1. Source coding for the cepstrum.

말까지의變化가 더욱 커지게 되고, 分析모델로부터 偏差가 커지게 된다. 따라서 窓函數를 適當히 維持하기 위한 한가지 方法은 이전의 피치推定에 根據하여 推定된 피치길이의 2배로 窓函數길이를 可變適用하면 된다.

有聲음일지라도 켈스트럼상에 두드러진 봉우리가 나타나지 않는 境遇는 極端으로 帶域制限이 되는 境遇이다. 즉, 音聲信號中 有聲閉鎖(voiced stop)音은 수백 Hz이상의 周波數에서 高調波構造가 별로 나타나지 않는, 極端으로 帶域制限된 境遇이다. 일례로 有聲信號가 純粹한 正弦波形態를 나타내면 그 對數스펙트럼상에 하나의 봉우리만 存在하게 된다. 이때 對數스펙트럼에 週期的인 高調波가 나타나지 않기 때문에 켈스트럼상에는 봉우리가 存在하지 않게 된다.

켈스트럼이 피치와 有聲음을 잘 表現하지 못한 境遇에는 信號源符號化時에 誤譯로 나타나서 合成音의 音質이 低下된다. 이것을 補償하려면 重-交叉率이나

에너지와 같은 다른 情報을 追加해야하는 등 檢出過程이 複雜하게 된다.

2. QUEFREQUENCY상에서 피치變更

켈스트럼 信號源符號化法에 의해 音聲을 合成하고자 할 때는 quefrequency상의 낮은 쪽 時間部分을 濾波器情報로 取한 다음 逆컨벌루션을 통해 임펄스特性을 구한다. 그리고는 音源情報에 該當하는 피치펄스와 컨벌루션을 통해 音聲信號의 部分을 合成하게 된다. 이렇게하면 符號化效率를 높일 수 있게 되어 낮은 傳達率에서도 良質의 合成이 可能하고, 임펄스特性과 컨벌루션되는 피치週기를 變更하여 韻律을 調節할 수 있기 때문에, 規則에 의한 合成法에 適用할 수 있는 符號化技法이 될 수 있다.

그렇지만 앞에서 說明된 것처럼 周波數上에서 피치週期 整數倍의 位置에 두드러진 봉우리가 나타나지 않는 境遇에는 信號源符號化의 誤譯가 發生하게 되고 이에 의해 合成된 音聲은 音質의 低下를 招

來하게 된다. 여기서 濾波器情報과 音源의 勵起情報을 別途로 分離하지 않고 캡스트럼自體를 符號化하는 波形符號化法을 適用하면 自然音의 音質을 最大한 維持할 수 있게 된다. 이렇게 할 때는 符號化에 必要한 메모리規模가 膨大해 지고, 音源파치의 變更改 어려워지게 된다.

符號化에 必要한 메모리의 問題는 序論에서도 言及하였듯이 現在의 技術水準으로 克服이 可能하다. 나머지問題로는 規則에 의한 合成이 이루어지도록 音源피치週期の 變更改가 可能해야 한다.

캡스트럼을 다시 보면 영-quefrequency 近方에 大部分의 캡스트럼값이 모이게 되고, quefrequency가 增加됨에 따라 그 값이 빠르게 減少된다. 특히, 피치週期가 가까이에서는 거의 영이 된다. 이 때문에 變更改하려는 週期만큼을 캡스트럼값이 거의 영이되는 部分에 插入하거나 削除시킴으로써 캡스트럼의 性質에는 거의 影響을 주지 않으면서 音源피치의 週期를 쉽게 變更改시킬 수 있게 된다.

quefrequency上에 영값을 插入 또는 削除하여 피치週期를 變更改시키고자할 때, 現在 分析中인 音聲의 피치를 事前에 알고 있다면 영-값代置를 피치週期 가까이에서 하는것이 바람직하다. 그러나 分析中인 音聲의 窓函數內에서 時間에 따른 音源피치의 變化가 進行中인 境遇에는 피치週期近方의 캡스트럼봉우리가 一定幅을 維持하는 것이 보통이다. 따라서 피치週期變更改를 위해 영값을 代置하는 位置는 영 quefrequency에서는 멀리떨어져게 하면서도 피치펄스에 너무 近接시키지 않는것이 바람직하다.

聲道の 周波數特性中에서 第一포먼트의 周波數는 250Hz이상으로 알려져 있고, 音源의 基本周波數는 最大 400Hz까지로 알려져 있다. 즉, quefrequency上에서 4msec(=1/250Hz)까지는 聲道の 特性을 나타내기 때문에, 피치變更改는 4msec 以上에서부터 피치週期까지에 영값을 代置하는것이 바람직하게 된다.

그림 4-2에는 캡스트럼分析을 통해 聲道の 特性을 維持하면서 피치를 變更改하는 過程을 그림으로 例示하였다. 그림 4-2(a)의 波形은 女性話者가 發聲한 /아/部分의 波形이고, 그 對數스펙트럼을 그림 4-2(b)에 나타내었다. 이에 대한 캡스트럼을 구한 다음에 피치週期를 20標本數만큼 늘리기위해 4msec에서부터 영값을 插入한 캡스트럼과 그의 스펙트럼을 그림 4-2(c)와 (d)에 나타내었다. 그림 4-2의 (b)와 (d)를 比較해보면 스펙트럼의 外廓線은 그대로 維持되면서 基本周波數가 願하는 값으로 낮아졌음을 알 수

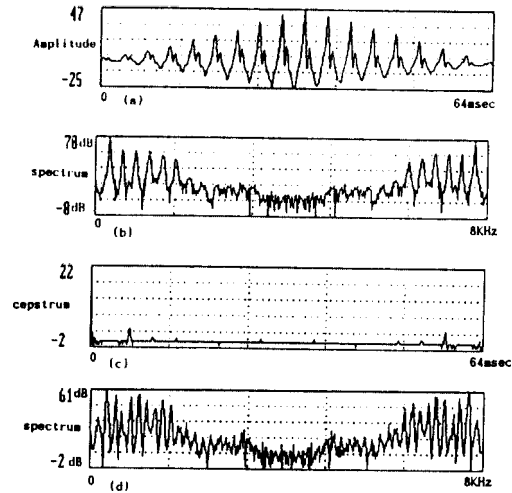


그림 4-2. 캡스트럼分析에 의한 피치變更改의 일례
Fig 4-2. One example altering the pitch by the cepstrum analysis.

있다.

3. 피치檢出時 考慮事項

信號源符號化法과는 달리 波形符號化法에서 피치를 變更改하려면 그 發聲者의 피치變更改를 事前에 알고 있어야 한다. 이것은 發聲者의 抑揚이나 感情의 變化가 中心된 피치를 基準으로 하여 피치의 相對的인 變化로 나타나기 때문이다. 특히 波形符號化에서는 發聲者의 個性과 메시지情報을 保存하기 때문에 音質의 明瞭性이 優秀하다. 따라서 피치變更改時에는 發聲者가 주로 使用하는 피치週期를 基準으로 피치를 變更改시킬 必要가 있다. 그러므로 正確한 피치檢出이 先行되어야 한다.

피치檢出은 캡스트럼을 遂行하기 前이나 그 處理過程에서 이루어져야한다. 그리고 캡스트럼을 計算할 때 窓函數에 따른 影響이 크기 때문에, 그影響을 最少化하기 위해서는 窓函數가 適用되는 時點을 聲門펄스의 勵起部分에 一致시키는 것이 좋다. 이렇게 하면 波形의 位相情報을 잃지않아야 하며 이 境遇에는 時間領域 피치檢出法이 바람직하게 된다.

그렇지만 合成을 위해 波形을 編輯하는 境遇에는 피치檢出이 반드시 自動化될 必要는 없으며, 피치檢出法과 함께 눈으로 피치를 檢出하는 半自動法이나 눈으로만 찾는 手動法으로 處理하여도 된다.

4. 結果考察

한 프레임의 길이를 512-標本으로 하여 그림 4-3에
시와 같이 處理하였다. 각 프레임에 대하여 피치檢出
法을 適用하여 피치를 구한 다음에, 푸리에變換
(FFT)을 遂行하여 周波數領域으로 變換하였다. 그
리고나서 對數函數를 취하여 振幅스펙트럼을 구한 다음
에 다시 逆푸리에變換(IFFT)을 하였다. 이렇게 구
해진 켈스트림값에 피치週期를 變更改하기 위해 영
quefrequency에서는 밀리 먼어지게 하면서 피치週期에
니부 近接하지 않도록 變更改하는 피치週期만큼의
영을 反復的으로 插入하거나 削除하였다.

男女가 各 5번씩 發聲한 數字音에 대해 각 프레임
별로 스펙트럼歪曲을 구하여 표 4-1에 나타내었다.
이 표에서 音聲 피치의 變更改率은 原來音聲의 스펙트
럼과 比較하기 위해 피치의 伸張과 壓縮을 同時에 遂
行하였다. 예를들어 피치週期를 50%로 壓縮한 다음
에 다시 200%로 伸張시켜 원래음聲의 스펙트럼과 比
較하여 그 誤差의 平均을 나타내었다. 켈스트림에 의
한 피치變更改法은 音聲波形에 대해 直接 適用한 피치
半分法에 비해 全般的으로 스펙트럼 歪曲이 작았으
며, 피치주기를 50%에서 200%로 變更改한 境遇에 스
펙트럼歪曲은 男性話者가 0.8% 程度이고, 女性話者
는 1.47% 程度로 아주 優秀하게 나타났다.

표 4-1. 피치변경율에 따른 스펙트럼왜곡

Table 4-1. Spectrum distortion according to pitch alteration rates.

피치변경율(%)	남성화자(28세)	여성화자(21세)
90%에서 111%로	0.17%	0.21%
80%에서 125%로	0.31%	0.41%
70%에서 142%로	0.47%	0.74%
60%에서 166%로	0.63%	1.08%
50%에서 200%로	0.80%	1.47%

V. 位相補償法

켈스트림에 의한 피치週期 變更改法은 時間과 周波
數領域을 往復하면서 處理하기 때문에 計算이 複雜하
다. 그렇지만 디지털 信號處理 專用칩의 도움으로 이
러한 問題는 克服될 수 있다. 또한 音聲合成을 위해
서는 合成用 데이터베이스가 必要한데 데이터베이스
를 作成할 때 켈스트림값을 貯藏해둔다면 合成時의
處理速度를 減少시킬 수 있다. 또 다른 短點으로는
피치變更改時에 켈스트림상에서 피치링스의 位置를 시
프트시키기 때문에 이것을 스펙트럼領域으로 變換했
을 때 그 位相情報을 잃어버리게 된다. 이것은 單純
히 複素켈스트림을 適用한다고 하여도 解決되지 않
는다. 왜냐하면 피치週期の 變更改는 스펙트럼의 高調
波의 位置를 變更改시키기 때문이다. 따라서 位相을 復
元하는 問題가 남게된다.

變更改된 켈스트림을 다시 FFT한 후에 指數函數를
取하면 基本周波數가 變更改된 스펙트럼을 얻을 수 있
다. 이때 位相歪曲을 補償하기 위하여, 그림 5-1과 같
이 時間領域의 피치변경법인 零插入-피치半分法을
使用한다. 따라서 이것은 位相保存이 良好한 영상입-
피치반분법의 位相과 스펙트럼의 振幅歪曲이 적은
켈스트림 振幅스펙트럼을 함께 使用하는 方法이다
[41].

이와 같이 時間領域에서 피치 變更改된 信號를 周波
數領域으로 變換시킨 後에, 이에대한 位相을 檢出하
여, 켈스트림 피치變更改法에 의해 變更改된 振幅스펙트
럼과 結合시킨다. 켈스트림 피치변경법에 의해 얻어
진 진폭 스펙트럼을 C(K)이라 하고, 영상입 피치반
분법에 의해 피치변경된 波形 h(n)의 位相스펙트럼
을 ϕ_r 라 하면 位相이 補償된 피치變更改 스펙트럼은 다
음과 같다.

$$A(K) = C(K) \cdot e^{j\phi_r} \tag{5-1}$$

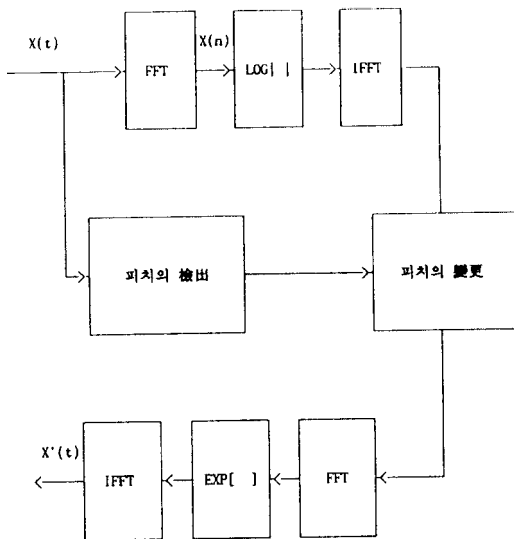


그림 4-3. 피치變更改에 대해 提案한 켈스트림法의 處理過程
블럭도

Fig 4-3. block diagram of cepstrum method for alter-
ing the pitch.

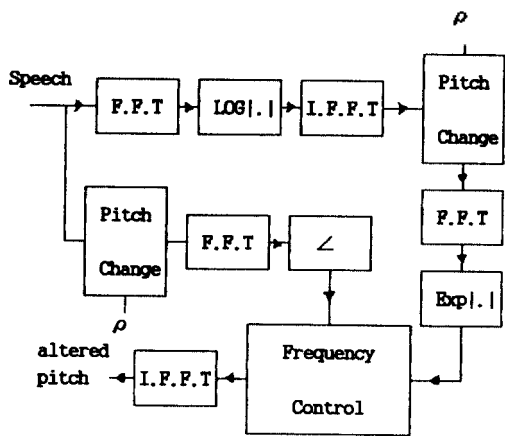


그림 5-1. 위상이 보상된 피치변경법
Fig 5-1. A pitch alteration method compensated the phase.

이에 대한 역변환 값을 $s'(n)$ 라 하면 다음과 같이 나타낼 수 있다.

$$s'(n) = \frac{1}{N} \sum_{k=0}^{N-1} A(k) e^{i(2\pi/N)kn} \quad (5-2)$$

이때 N 은 窓函數 길이이며, $s'(n)$ 은 位相이 保存된 狀態로 피치가 變更된 音聲波형이 된다.

VI. 결 론

모든 設備가 自動化되고 고도로 機械化됨에 따라 音聲을 통하여 작업 設備를 制御 및 運用하려는 音聲信號處理에 관한 研究가 活潑히 進行되고 있다. 특히, 音聲信號의 特徵만을 貯藏하여 狀況에 맞는 音聲應答을 自由로 이룰 수 있는 合成技法에 관한 研究가 많이 提案되고 있다. 合成單位로는 文章單位, 單語單位, 音素單位등의 調音結合 알고리즘이 提案되어지고 있으며 合成單位를 코딩법으로 具現시키는 技法으로는 波形코딩법, 소스코딩법, 혼성코딩법이 提案되어져 있다.

無制限 單語合成에 대한 合成單位로는 音聲데이터의 메모리容량의 限界性 때문에 音素單位나 音節單位的 規則合成 알고리즘에 대한 많은 研究가 進行되고 있다. 또한 合成 코딩技法에는 無制限單語의 合成을 위해 메모리容량의 效率성과 音源피치의 變更이 容易한 소스코딩에 의한 合成技法이 普通 採擇되고

있다. 그렇지만 最近의 半導體技術은 칩당 16Mbit 規模의 메모리를 이미 實用化에 成功하였고, 메모리 칩당 4Mbit의 容량은 現在 市販中에 있다. 이러한 規模의 메모리容량을 使用하게 되면 지금까지의 合成 코딩법에서 데이터량에 따른 코딩기법의 制限性은 除去될 수 있다.

따라서 本 論文은 데이터량의 規模에 限界를 갖지 않으면서 自動 應答시스템에서 高音質을 提供해줄 수 있는 規則合成用 波형코딩 合成技法에 대해 알아 보았다. 우선 既存에 提案된 波형符號化에서의 피치變更法들을 簡單히 紹介하고 그 限界性을 다룬 다음에 本 研究室에서 波형符號化와 피치檢出法들의 原理를 利用하여 두가지 피치變更法들을 提案했던 것을 整理하여 說明하였다. 먼저 波形的 週期를 簡單히 半分하는 週期半分法을 適用하여 時間領域上에서 波形的 피치週기를 變更하는 方法을 說明한 다음에 時間-周波數領域의 變換特性을 利用한 캐스트럼에 의한 피치變更法에 대해 整理하였다.

피치週期半分法은 時間領域에서 바로 處理되기 때문에 피치가 變更되어도 波形的 位相情報가 維持된다. 따라서 한 피치週期單位로 合成時에 波형編輯이 容易하다. 또한 피치週期半分法은 피치檢出이 이루어졌을 境遇에 既存의 線型豫測合成法을 波形的 延長된 部分에만 適用하므로 處理時間이 N 번의 덧셈과 N 번의 시프트演算만 要求된다.

反面, 피치週期半分法은 聲門과 聲道の 特性을 原來스펙트럼과 같게 維持하지 못하고 스펙트럼의 歪曲을 招來하게되는 短點이 있다. 原來의 스펙트럼을 基準으로 하였을 때 男女話者의 모든 境遇에 百分率로 平均 90.5%를 追從하였으며, 피치를 늘린 境遇보다는 중간境遇가 스펙트럼歪曲이 9% 程度 높게 나타났다.

캐스트럼에 의한 피치變更法은 音聲波형에 대해 直接 適用한 피치半分法에 비해 全般적으로 스펙트럼 歪曲이 작았으며, 피치주기를 50%에서 200%로 變更한 境遇에 스펙트럼歪曲은 男性話者가 0.8% 程度이고, 女性話者는 1.47% 程度로 아주 優秀하게 나타났다.

참 고 문 헌

- 7-1. 음성 신호 처리
1. J.L. Flanagan, 2nd ed., *Speech Analysis, Synthesis and Perception*, NY : Springer-Verlag, 1972.
2. G. Fant, *Speech Sounds and Features*, 1972.

3. A.V. Oppenheim and R.W. Schaffer, *Digital Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1975.
 4. R.W. Schaffer and L.R. Rabiner, "Digital representations of speech signals," *Proc. IEEE*, vol.63, No.4, pp.662-667, April 1975.
 5. J.D. Markel and A.H. Gray, Jr., *Linear Prediction of Speech Signals*, Springer Verlag, 1976.
 6. L.R. Rabiner and R.W. Schaffer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
 7. R.W. Schaffer and J.D. Markel, Eds. *Speech Analysis*, New York, IEEE Press, 1979.
 8. W.A. Lea, Ed., *Trends in Speech Recognition*, Prentice Hall, 1980.
 9. W.E. Cooper and J.P. Cooper, *Syntax and Speech*, Harvard University Press, 1980.
 10. R.A. Cole, Ed., *Perception and Production of Fluent Speech*, Lawrence Erlbaum Associates, 1980.
 11. W.E. Cooper and J.M. Sorensen, *Fundamental Frequency in Sentence Production*, Springer Verlag, 1981.
 12. J.P. Haton, *Automatic Speech Analysis and Recognition*, D. Reidel Publishing Company, 1982.
 13. L.H. Witten, *Principles of Computer Speech*, Academic Press, 1982.
 14. G. Bristow, *Electronic Speech Synthesis*, McGraw-Hill, 1984.
 15. R.D. Mori and C.Y. Suen, *New Systems and Architectures for Automatic Speech Recognition and Synthesis*, Springer-Verlag, 1985.
 16. S. Saito and K. Nakata, *Fundamentals of Speech Signal Processing*, Academic Press, 1985.
 17. F. Fallside and W.A. Woods, Ed., *Computer Speech Processing*, Prentice-Hall, 1985.
 18. T.W. Parsons, *Voice and Speech Processing*, McGraw-Hill, 1986.
 19. J.S. Perkell and D.H. Klatt, Ed., *Invariance and Variability in Speech Processes*, Lawrence Erlbaum Associates, 1986.
 20. G. Bristow, *Electronic Speech Recognition*, Collins, 1986.
 21. E.J. Yannakoudakis and P.J. Hutton, *Speech Synthesis and Recognition Systems*, Ellis Horwood Ltd., 1987.
 22. P.E. Papamichalis, *Practical Approaches To Speech Coding*, Prentice-Hall, 1987.
 23. A. Waibel, *Prosody and Speech Recognition*, Morgan Kaufmann Publishers, 1988.
 24. A.N. Ince, *Digital Speech Processing - Speech Coding, Synthesis, and Recognition*, Kluwer Academic Publishers, 1992.
 25. J.R. Deller, J.G. Proakis, J.H.L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Co., 1993.
- 7-2. 피치 변경
30. B. E. Caspers and B. S. Atal, "Changing pitch and duration in LPC synthesised speech using multiple excitation," *J. Acoust. Soc. Amer.*, suppl., vol.73, No.1, pp.S5, Spring, 1983.
 31. M. G. Stella and F. J. Charpentier, "Diphone synthesis using multiple coding and a phase vocoder," in *Proc. IEEE ICASSP'85*, pp.740-744, 1985.
 32. A. Varga and F. Fallside, "A technique for using Multipulse Linear Predictive Speech Synthesis in Text-to-Speech Type Systems," *IEEE Trans. Acoust., Speech, Signal processing*, Vol. ASSP 35, No.4, pp.586-587, April 1987.
 33. 강동규, 김을재, 배명진, 안수진, "음성 파형의 halving 기법에 의한 화성코딩의 피치변경에 관한 연구," *한국음향학회 추계발표회(국제음향학회) 논문집*, pp. 107-111, 1990년 11월 10일.
 34. 배명진, 이비숙, 이해군, 안수진, "웹스트림 분산에 의한 음성 파형코딩의 피치변경에 관한 연구," 제4회 신호처리합동 학술대회 논문집, 제4권, 제1호, pp.304-309, 9월 1991.
 35. M. Bae, H. Yoon, S. Ann, "On Altering the Pitch of Speech Signals in Waveform Coding - Alteration Method by the LPC and the Pitch Halving-," *J., Acoustics of Korea*, Vol.10, No.5, pp.11-19, October 1991.
 36. 민경중, 배명진, 유희상, 안수진, "음성 파형 코딩의 음원 변경에 관한 연구," *음향학회 학술발표회 논문집*, Vol.10, No.1(s), pp.45-49, Nov.9, 1991.
 37. 박찬수, 김형태, 배명진, 안수진, "음성파형 분절의 스트리밍 기법에 관한 연구," *음향학회 학술발표회 논문집*, Vol.10, No.1(s), pp.7-10, Nov.9, 1991.
 38. 이비숙, 배명진, 안수진, "선, 후방향 LPC에 의한 음성 파형 분절의 연결 부분 스트리밍," *음향학회 학술발표회 논문집*, Vol.10, No.1(s), pp.15-20, Nov.9, 1991.
 39. T.F. Quatieri, R.J. McAulay, "Shape Invariant Time-Scale and Pitch Modification of Speech," *IEEE Trans., Signal Processing*, Vol.40, No.3, pp. 497-510, March 1992.
 40. M. Bae, M. Lee, "On a Pitch Change of the Waveform Coding by the Cepstrum Analysis for

- Speech Waveforms," J., Acoustics of Korea, Vol. 11, No.4, pp.14-21, August 1992.
41. U. Kim, J. Lee, M. Bae, S. Ann, "On a Pitch Change of the Waveform Coding by the Cepstrum Analysis for Speech Waveforms-on the Phase Compensation for the Pitch Change-," J., Acoustics of Korea, Vol.11, No.5, pp.51-58, October 1992.

▲배 명 진 : 11권 1E호 참조