論文 86-23-4-21

實時間 有聲音·無聲音·無音 識別裝置의 構成에 關한 硏究

(A Study on Implementation of Real Time Voiced/Unvoiced/Silence Discrimination System)

房 萬 源*, 崔 甲 石**

(Man Won Bang and Kap Seok Choi)

要 約

本 論文은 音聲信號의 零交差 回數와 周波數 에너지 스펙트럼 分布를 利用한 有聲音·無聲音·無音識別裝置를 構成하여 實驗하였다. 零交差 回數의 測定에는 周波數-電壓 變換型의 새로운 時間幅 分類器를 提案하여 使用하였다.

實驗 結果 이 方式은 有聲音 區間에서 임펄스性 雜音의 영향이 제거되었고, 實時間 處理가 可能하였으며, 識別裝置의 構成이 비교적 간단하였다.

Abstract

In this paper, the implementation of a voiced/unvoiced/silence discrimination system is presented. The algorithm is based on the zerocrossing rate and the spectral energy distribution of speech. In measuring zerocrossing rate, a new frequency-to-voltage conversion type interval filter is used. Expermental results show that with the proposed algorithm the effect of impulse noise in voiced intervals can be removed.

I. 序 論

音聲認識의 前處理에 必要한 세그멘테이션(segmentation)이나 피치(pitch) 抽出등의 音聲分析에 있어서 重要한 問題의 하나는 音聲의 各 分析區間을 有聲音·無聲音·無音으로 分類하는 것으로 이에 대한 여러가지 方法들이 提案되고 있다.[1,8]

Noll과 Markel의 方式은[3,4] 音聲의 時間波形의 周期性에 주목한 方法으로서 有聲音 區間에서도 音源声帶 진동의 불규칙 진동등에 의해서 時間波形上의 周期構造가 일시적으로 붕괴되어 약해지는 區間, 즉 有聲音의 시작이나 종단에서의 分類가 곤란하게 된다.

패턴인식이나 線形豫測法에 속하는 Atal과 Rabiner의 方法은[5,7] 몇개의 다른 分析에서 얻어진 파라메터를 이용하여 各 class의 표준패턴과의 유사도에 의하여 分類하는 方法으로서 他 方式에 의한 處理를 할 必要가 있고 파라메터의 種類도 서로 다른 것으로서 그 數가 비교적 많다.

연관된 대수 스펙트럼의 周波數와 時間에 대한 2차원 푸리에 變換으로 定義되는 2차원 켑스트럼(cepstrum)法은[8] 音聲의 대수 스펙트럼 포락에 주목한 方法으로서 캡스트럼의 分析으로 얻어진 4개 이상의 파라메터를 이용하여 音源의 불규칙성에 의한 時間波形에서 피

*正會員, 關東大學 電子工學科
(Dept. of Elec. Eng., Kwan Dong Univ.)
**正會員, 明知大學校 電子工學科
(Dept. of Elec. Eng., Myong Ji Univ.)
接受日字 : 1985年 8月 20日

치周期構造가 不安定한 有聲音 區間에서도 分類가 可能한 方法이다. 그러나 계산량이 상대적으로 增加하며 他 方式에 의한 處理를 병렬로 할 必要가 있고 實時間에 간단히 識別하는 方法들은 아니다.

비교적 간단하며 單時間의 音聲情報로서 實時間 識別이 可能한 方法은 音聲信號의 零交差 回數를 利用하는 方法과 高域과 低域의 에너지 크기를 비교하는 方法들이 있다.[9] 零交差 回數를 利用하는 方法은, 音聲信號가 零點을 交差하는 比率이 有聲音은 30〔回/10ms〕以下에 存在하는 확률이 높고, 無聲音은 30〔回/10ms〕以上에 存在하는 확률이 높은 性質을[7] 이용한 것으로, J. Wiren과 H. L. Stubbs는 時間幅 分類器 (interval filter)를 사용하여 識別하였다.[10] 高域과 低域의 에너지 크기를 비교하는 方法은 有聲音의 에너지가 주로 낮은 周波數 영역에 分布되어 있고, 無聲音의 에너지가 높은 周波數 영역에 分布되어 있는 性質을 이용하여 識別하는 方法이다. 그러나 有聲音의 聲帶音源에는 임펄스列의 영향으로 피치周期 때마다 큰 에너지를 갖는 高周波 成分이 수반되므로 이 때문에 有聲音 區間을 無聲音 區間으로 識別할 可能性이 있다.[2] 이 問題를 해결하기 위하여 Knorr는 比較器의 出力을 2～4〔ms〕정도 遲延시켜서 識別誤差를 감소시키고 임펄스性 雜音의 영향을 제거하였다.[1] 그러나 遲延回路의 사용으로 인하여 最大 4〔ms〕의 時間遲延이 發生하고, 이 때문에 有聲音의 시작과 종단에서 識別誤差가 나타나게 된다.

本 論文에서는 周波數－電壓 變換型의 새로운 時間幅 分類器를 提示하여 零交差 回數를 測定하고, 이것을 에너지 比較法에 의하여 얻어진 結果와 組合함으로써 識別誤差와 遲延時間을 감소시켜 有聲音·無聲音의 識別이 實時間에 處理되도록 하였으며, 音聲區間과 無音區間의 識別은 入力音聲의 최저 에너지 레벨을 測定하여 미리 設定된 기준치와 비교하는 方法을 이용하였다.

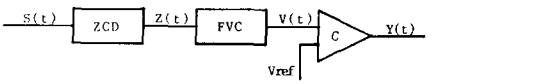
Ⅱ. 時間幅 分類器

零交差波의 測定에는 時間幅 分類器(interval filter)를 利用할 수 있다. 時間幅 分類器는 미리 決定된 時間幅과 인가된 信號의 時間幅을 비교하는 것으로서,[11] 周波數－電壓 變換器를 使用한 時間幅 分類器를 그림 1에 보였다.

그림 1에서 入力 音聲信號를 S(t)라 하고, S(t) 의 直交信號를 $\sigma(t)$라 하면 S(t)와 $\sigma(t)$를 直交 成分으로 하는 벡터는 다음 式(1)로 表示된다.[11]

$$S(t)-j\sigma(t)=a(t)e^{-j\varphi(t)} \qquad (1)$$

여기에서



ZCD; zerocrossing detector. FVC; frequency-to-voltage converter.
C ; comparator. Vref; reference voltage.

그림 1. 周波數－電壓 變換型 時間幅 分類器의 構成圖
Fig. 1. Block diagram of frequency-to-voltage converter type interval filter.

$$\left.\begin{aligned} a(t)&=[S^2(t)+\sigma^2(t)]^{\frac{1}{2}}\\ \varphi(t)&=\tan^{-1}\frac{\sigma(t)}{S(t)}\end{aligned}\right] \qquad (2)$$

와 같이 주어진다. 式(1)은 복소 平面上에서 定義된 함수이며, S(t)는

$$S(t)=a(t)\cos\varphi(t) \qquad (3)$$

와 같이 a(t)와 $\cos\varphi(t)$로 표시할 수 있다.[12]

零交差波를 z(t)라 하면,

$$z(t)=A\cos\varphi(t) \qquad (4)$$

단 A : 일정진폭

式(3)에서 a(t)는 零交差波의 포락선에 대응해서 대부분 $a\geq 0$으로 되고 $\varphi(t)$는 벡터 회전각이 된다. 따라서 $d\varphi/dt$는 순시각 周波數가 되며 零交差波의 情報는 전부 이 속에 포함된다. 여기에서 a(t)와 $\cos\varphi(t)$의 스펙트럼은 중복되지만 a(t)의 最大 變化速度가 $\cos\varphi(t)$의 最小 變化速度보다 작으므로 이들을 서로 분리할 수 있고, S(t)의 零交差波 發生으로 나타나는 高調波는 적당한 濾波器에 의하여 거의 제거할 수 있으므로 a(t)가 一定한 진폭 A로 되며, $\cos\varphi(t)$만이 原信號 대로의 理想的인 零交差 信號 z(t)로 된다.

周波數－電壓 變換器의 出力을 V(t)라 하면,

$$V(t)=K\cdot X(t) \qquad (5)$$

여기에서 X(t)는 $d\varphi/dt$이고, K는 周波數－電壓 變換係數이다. 따라서 V(t)는 式(5)와 같이 $d\varphi/dt$에 의해서 決定되며, 진폭 A와는 無關하게 된다.

레벨 比較器의 出力 Y(t)는

$$\left.\begin{aligned} Y(t)=1 &: V(t)\geq V_{ref}\\ 0 &: V(t)<V_{ref}\end{aligned}\right] \qquad (6)$$

여기에서 V_{ref}는 時間幅 設定 基準電壓이다. 그러므로 時間幅 分類器의 V_{ref}를 30〔回/10ms〕에 設定하면 有聲音·無聲音의 識別이 可能하게 된다.

Ⅲ. 識別 시스템의 構成

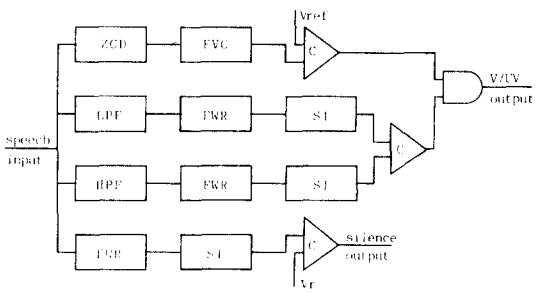
有聲音의 周波數 스펙트럼은 피치 周波數 간격의 線

스펙트럼이 되고, 스펙트럼의 포락은 聲道의 공진특성에 대응하여 存在하고 있으므로 有聲音에 의한 聲帶의 진동은 비교적 周期的이며 낮은 周波數 영역에 그 에너지가 分布되어 있다.

無聲音의 경우 그 音源이 亂流性 雜音源 혹은 破裂音源으로 보이는 音聲으로 되고, 그의 스펙트럼은 연속적인 線 스펙트럼이 되며 스펙트럼 포락은 聲道의 공진특성에 의존한다. 따라서 無聲音의 音源은 音調를 좀히는 위치 부근에서 發生한 雜音이기 때문에 無聲音의 에너지는 有聲音의 에너지에 비하여 비교적 높은 周波數 영역에 分布되어 있다. 그러므로 低周波 成分의 에너지와 高周波 成分의 에너지를 비교함으로서 有聲音과 無聲音의 識別이 可能하다.[1]

한편 零交差波의 測定으로 얻어지는 結果는 에너지 成分과는 無關하며, 단지 瞬時角 周波數 成分 $d\varphi/dt$ 과의 關係만을 가지고 있으므로 에너지의 비교에 의하여 얻어진 結果와 組合하게 되면 識別誤差의 감소와 더불어 實時間 處理가 확실하게 이루어 질 수 있다.

그림 2에 時間幅 分類器와 에너지 比較法을 利用한 有聲音·無聲音·無音 識別 시스템을 提示하였다. 그림 2에서 時間幅 分類器는 零交差波 檢出器, 周波數-電壓 變換器 및 比較器로써 構成하였다.



ZCD : zerocrossing detector.
FVC : frequency-tovoltage converter.
LPF : lowpass filter.
HPF : highpass filter.
FWR : full wave rectifier.
S I : short time intergrator.
C : comparator.
Verf : interval filter reference voltage.
Vr : silence reference voltage.

그림 2. 有聲音·無聲音·無音 識別裝置의 構成圖
Fig. 2. Block diagram of voiced/unvoiced/silence discrimination system.

零交差波 檢出器는 式(3)의 $a(t)$와 $\cos\varphi(t)$의 스펙트럼이 중복되므로 혼변조에 의해 찌그러짐이 發生하며, 音聲波에 혼입하는 시스템 雜音이 클리핑(clipping)에 의해서 強調되기 때문에 品質이 低下된다. 이러한 결함을 제거하기 위하여 0레벨 대신 固定레벨을 갖도록 하였다.[11]

周波數-電壓 變換係數 K는[14]

$$K = V_{REFF} \cdot R_1 \cdot C_2 \qquad (7)$$

로서, V_{REFF}는 變換 基準電壓이며, R_1은 180[KΩ], C_2는 500[PF]로 設計하였고, 零交差 回數가 30[回/10ms]일 경우를 레벨 比較器의 基準電壓으로 設定하여, 周波數-電壓 變換器의 出力이 이 값보다 클 경우 無聲音, 작을 경우는 有聲音으로 判定하도록 하였다.

한편 高域과 低域의 에너지 비교는 低域通過 濾波器, 高域通過 濾波器, 全波整流回路, 平滑回路 및 레벨 比較器를 利用하였다.

低域通過 濾波器는 차단 周波數를 1[KHz]에 設定하였고, 6극의 butter worth-VCVS(Voltage controlled Voltage Source) 濾波器로 하였으며, 高域通過 濾波器는 차단 周波數를 5[KHz]에 設定하였고 6극의 1[dB]리플을 갖는 chebyshev cascade VCVS 濾波器로 하였다.

그리고 音聲區間과 無音區間의 識別은 미리 人力 音聲信號의 최저 에너지 레벨을 測定하여 측정된 값보다 15[dB]정도 낮은 값을 기준치로 設定하고, 이 기준치를 초과할 경우를 音聲存在區間, 기준치 以下일 경우를 無音區間으로 하는 에너지 檢出 方式[3]을 이용한 것으로써 信號 對 雜音比가 양호한 高品質(S/N比 60 dB 以上)의 音聲信號를 대상으로 하여 全波整流回路, 平滑回路 및 레벨比 較器를 사용하여 간단하게 構成하였다.

Ⅳ. 實驗 및 結果

有聲音·無聲音의 識別 實驗을 하기 위하여 試作한 時間幅 分類器를 그림 3에 提示하였다.

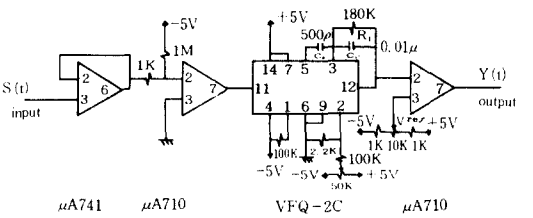


그림 3. 周波數-電壓 變換型 時間幅 分類器
Fig. 3. Realization of frequency-voltage converter type interval filter.

그림 3에서 周波數-電壓 變換器는 최대 0.02%의直線性을 갖는 Datel社의 VFQ-2C를 使用하였고, 基準

電壓 V_{REFF}는 −5 V로 하였다.

또한 3〔KHz〕 入力의 경우 V(t)의 계산값은 1.35(V)였으나 測定된 값은 1.37〔V〕로서 0.02〔V〕의 誤差가 나타났으며, 이것은 變換器의 直線性에 기인한 것으로 생각된다. 따라서 레벨 比較器의 基準電壓 V_{ref}는 1.37〔V〕에 設定하여 實驗하였다.

그림 4 는 音聲區間·無音聲區間의 識別回路로서 無音區間의 實測 雜音電壓은 10〔mv_{P-P}〕였고, 有音區間의 最小 電壓은 100〔mv_{P-P}〕 以上이었으므로 回路에서 V_r 값은 安全値를 고려하여 50〔mv〕에 設定하였다.

또한 시스템에 인가하기 前에 미리 12〔KHz〕의 차단 周波數를 갖는 4 극의 butter worth 低域通過 濾波器를 사용하여 信號內에 포함되어 있는 高周波 雜音을 제거하였다.

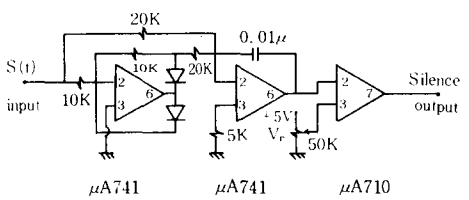


그림 4. 音聲區間·無音區間 識別 回路
Fig. 4. Realization of voice/silence discriminator.

實驗은 2 현상 메모리 스코프를 사용하여 시스템에 인가된 音聲 入力信號와 識別出力을 각각 메모리시키고 메모리 된 出力을 폴리그래프(polygraph)로 기록하였다.

觀測된 結果를 그림 5～9에 나타내었다.

그림 5, 6은 "전", "자"에 대한 有聲音·無聲音 識別結果이며, 遲延時間을 觀測하기 위하여 入力 音聲信號에 대한 時間幅 5ms/cm로 變化시켜 觀測한 結果를 그림 7 에, 1ms/cm로 變化시켜 觀測한 結果를 그림8에 나타내었다. 그림 9 는 연속된 音聲 "초생달"에 대한 有音·無音 區間의 識別出力이다.

實驗 結果 본 시스템은 音聲信號의 實時間 識別이 可能하였고, 有聲音 區間에서 피치 周期때 마다 나타나는 高周波 成分을 時間幅 分類器에 의해서 제거할 수 있었다.

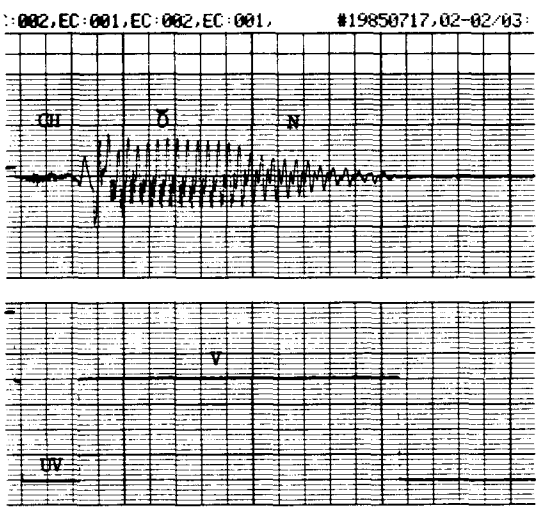


그림 5. 유성음/무성음 식별 결과
위 : 원음성 "전" 수직축 : 1V/cm
수평축 : 50ms/cm
아래 : 식별 출력 수직축 : 2V/cm
수평축 : 50ms/cm

Fig. 5. V/UV decision. Top display: original sheech "chon". Vertical: 1V/cm, horizontal: 50ms/cm. Bottom display: V/UV output signal. Vertical: 2V/cm, horizontal: 50ms/cm.

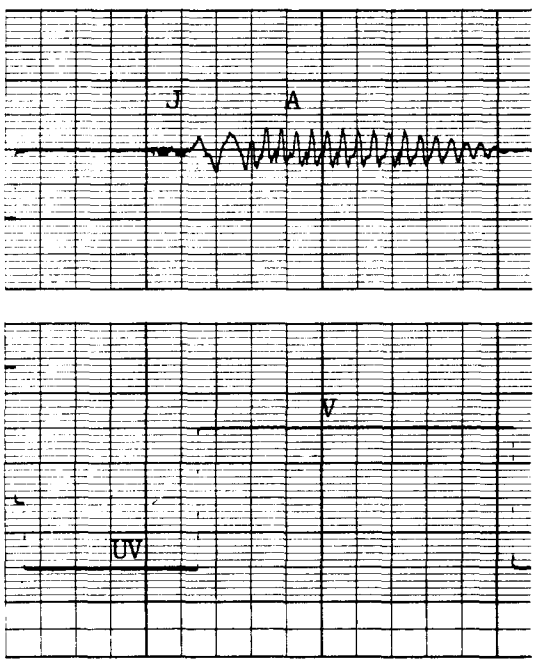


그림 6. 유성음/무성음 식별 결과
위 : 원음성 "자" 수직축 : 1V/cm
수평축 : 50ms/cm
아래 : 식별출력 수직축 : 2V/cm
수평축 : 50ms/cm

Fig. 6. V/UV decision. Top display: original speech "ja."
Vertical: 1V/cm, horizontal: 50ms/cm.
Bottom display: V/UV output signal.
Vertical: 2V/cm, horizontal: 50ms/cm.

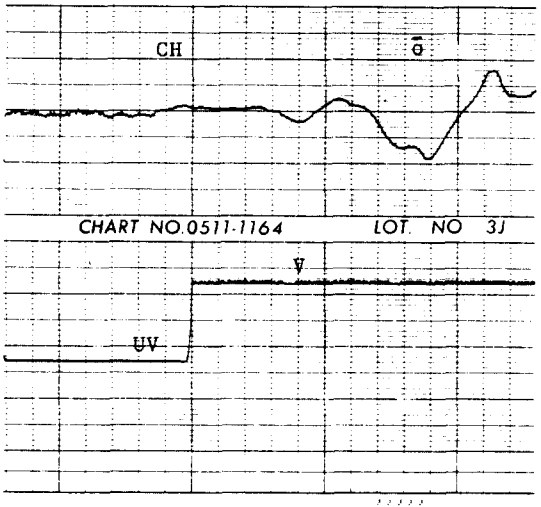


그림 7. 유성음/무성음 식별결과
위 : 원음성 "전" 수직축 : 1V/cm
수평축 : 5ms/cm
아래 : 식별 출력 수직축 : 2V/cm
수평축 : 5ms/cm

Fig. 7. V/UV decision. Top display: original speech "chon".
Vertical: 1V/cm, horizontal: 5ms/cm.
Bottom display: V/UV output signal.
Vertical: 2V/cm, horizontal: 5ms/cm.

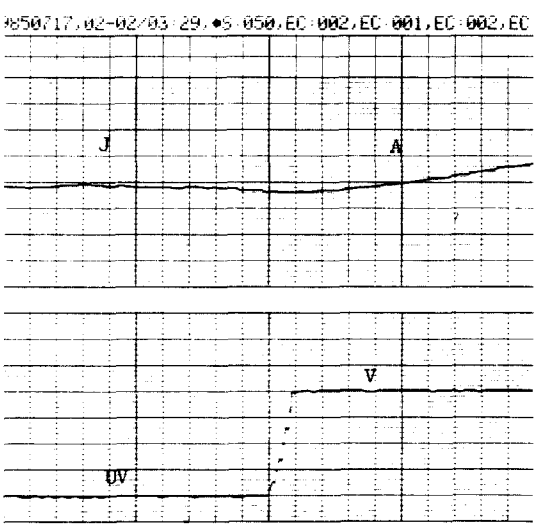


그림 8. 유성음/무성음 식별 결과
위 : 원음성 "자" 수직축 : 1V/cm
수평축 : 1ms/cm
아래 : 식별 출력 수직축 : 2V/cm
수평축 : 1ms/cm

Fig. 8. V/UV decision. Top display: original speech "ja".
Vertical: 1V/cm, horizontal: 1ms/cm.
Bottom display: V/UV output signal.
Vertical: 2V/cm, horizontal: 1ms/cm.

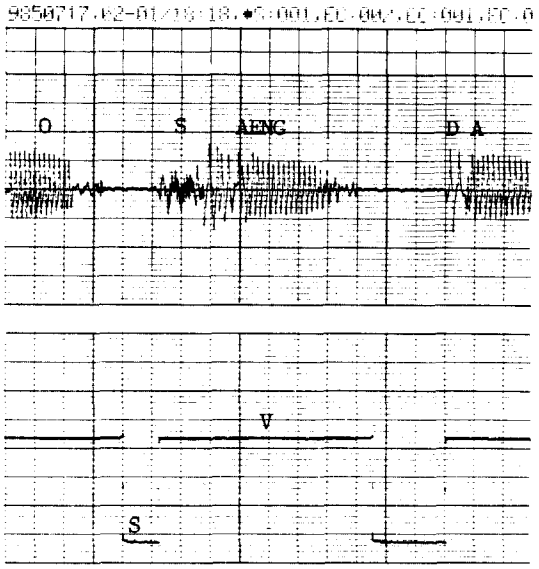


그림 9. 유음/무음구간 식별 출력
위 : 원음성 "초생달" 수직축 : 1V/cm
수평축 : 0.1s/cm
아래 : 식별출력 수직축 : 2V/cm
수평축 : 0.1s/cm

Fig. 9. Voice/Silence decision. Top display: original speech "cho saeng dal".
Vertical: 1V/cm, horizontal: 0.1s/cm.
Bottem display: voice/silence output signal.

V. 結 論

本 硏究에서는 音聲信號의 에너지 스펙트럼 分布와 零交差 回數를 이용한 有聲音·無聲音·無音 識別裝置를 構成하여 임펄스 成分의 영향 제거와 遲延時間 감소에 대하여 實驗的으로 考察한 結果 다음과 같은 結論을 얻었다.

1) 時間幅 分顯器와 에너지 分析法으로 얻어진 結果를 AND 게이트로 組合하는 것에 의해서 有聲音 區間에서 聲帶振動의 임펄스列에 의한 識別誤差와 임펄스性 雜音의 영향이 제거되었다.

2) 時間幅 分類器를 使用한 것에 의하여 實時間 識別이 可能하였다.

3) 音聲 存在區間의 檢出은 信號 對 雜音比가 양호한 경우 入力 音聲의 에너지를 檢出하여 入力 音聲의 최저 에너지 레벨보다 15[dB] 낮은 값으로 기준치를 設定하는 것으로 可能하였다.

4) 識別裝置의 構成이 간단하므로 實時間 處理를 目的으로 하는 音聲認識 등에 有效한 方法이다.

參 考 文 献

[1] S.G. Knorr, "Reliable Voiced/Unvoiced Decision," *IEEE Trans. ASSP-27,* pp. 263-267, Jun., 1979.

[2] 有井浩二, 村田正, 滑川敏彦, "實時間 有聲一無聲一無音 識別 システム", 信學會 全國大會 論文集, pp. 710～710～717, 1982.

[3] A.M. Noll, "Cepstrum Pitch Determination", JASA, vol. -42, no. 2, pp. 293-309, 1967.

[4] J.D. Markel, "The SIFT algorithm for fundamental frequency estimation", *IEEE Trans, AU,* vol. AU-20, no. 5, pp. 367-377, Dec., 1972.

[5] L.R. Rabiner, M.S. Sambur, "Application of an LPC distance measure to the voiced-Unvoiced-Silence detection problem", *IEEE ASSP.* vol. ASSP-25, no. 4, pp. 338-343, Aug., 1977.

[6] L.J. Siegel, A.C. Bessey, Voiced/Unvoiced /Mixed excitation classification of speech", *IEEE ASSP,* vol. ASSP-30, no. 3, pp. 451-460, Jun., 1982.

[7] B.S. Atal, L.R. Rabiner, "A Pattern revognition approach to Voiced-Unvoiced-Silence classification with applications to speech recognition", *IEEE Trans Proc.,* vol. ASSP-24, no. 3, pp. 201-212, Jun., 1976.

[8] 小畑秀文, "高レベル雜音中の 音聲の有聲/無聲判定の 最適化 手法", 信學論, vol. J65－D, no. 11, pp. 1406～1413, nov. 1982.

[9] L.R. Rabiner, R.W. Schafer, *Degital processing of speech signals,* Prentice-Hall, London, pp. 120-135, 1978.

[10] J. Wiren, H.L. Stubbs, "Electronic binary selection system for phoneme classification", JASA, vol. 28, no. 6, pp. 1082-1091. Nov. 1956.

[11] 三浦種敏監修, 聽覺と音聲, 電子通信學會, 東京, pp. 384～396, 1980.

[12] William J. Done, R. Lynn Kirlin, "Speech Zero-Crossing Rate Compression for Bandwidth Compression", *IEEE Tran. ASSP,* vol. ASSP-23, no. 5, Oct., 1979.

[13] 新美康永, 音聲認識, 共立出版, 東京, pp. 67～72, 1979.

[14] Datel, デイテル總合 カタログ no. 4 1984/85, 東京, pp. 210～211, 1984.

[15] D.E. Johnson,J.L. Hilbur, *Rapid Pratical Design of Active Filters,* John Wiley & Sons, New York, pp. 6-133, 1975.