論 文

고해상 피치검출을 이용한 한국어 음성신호의 음소분리

正會員 李 應 求* 正會員 李 斗 秀*

Segmentation of the Korean speech signals into phonetic units using the super resolution pitch determination

Eung Gu Lee* Doo Soo Lee* *Regular Members*

要 約

본 논문에서는 고해상 피치검출을 이용해서 정확한 피치를 찾고 각 피치 주기에서의 상관함수와 문턱값을 비교하여 한국어 음성신호를 음소단위로 분리하는 알로리듬을 제안한다. 제안된 알고리듬의 특성은 정확하고 고신뢰도를 갖으며, 변이구간이나 무음구간도 구분할 수 있다. 이 알고리듬은 음소단위로 분리하여 코드북을 설계하는 벡터양자화와 음성인식 분야에 적용된다. 본 논문에서 제안한 알고리듬은 PC 386 / DX 상에서 386 / MATLAB으로 실행한 결과 피치주기를 정확히 찾고 음소별로 분리가 가능함을 알 수 있다.

ABSTRACT

This paper is presented the phonetic segmentation algorithm of the Korean speech signals which is finded the exact pitch using the super resoluton pitch determination and is compared corss-correlation to threshold each pitch period. The features of the proposed algorithm are infinite resolution and high reliability, and also can separate transient or silent segment. The algorithm is instrumental to speech processing applications which require vector quantization and speech recognition. The presented algorithm is implemented by 386-MATLAB on PC 386 / DX and is verified the exact pitch period and the phonetic segmentation of speech signals.

I. 서 론

음성처리 분야에서 피치(pitch)와 유 / 무성음 판별은 중요한 요소로 사용되고 있다. 피치는 음성의 자연스러움에 큰 영향을 미치기 때문에 인간의 청각은 피치의 변화에 민감하게 반응하게 되며, 따라서 피치는 합성음성의 음질을 좌우하는 중요한 요소이다. 또한 음성인식에 있어서도 정확한 피치의 검출은 화자에 따른 영향을 줄일 수 있으며 인식의 정확도를 높일 수 있게 된다. 그러나 정확한 피치의 검출은 음성 신호가 갖고 있는 몇가지 특성으로 인하여 어려운 과제로 남아 있다. 즉 음성신호 자체가 원칙적으로 nonstationary process이며 화자에 따른 음성신호

*漢陽大學校 電子工學科
Dept. of Electronic Eng., Hanyang University
論文番號 : 93-29

특징에 차이가 크며 유 / 무성음 구간은 상호작용등으로 피치의 정확한 검출을 어렵게 하고 있다. 중요하면서도 어려운 분야인 이유로 많은 피치 검출 알고리듬이 제시되어 왔다.[1]-[10] 이런 알고리듬을 크게 세 부류로 구분하면 시간영역, 주파수영역, 그리고 혼합영역에서 구하는 방법으로 나눌 수 있다. 이렇게 구분한 각각의 알고리듬은 실제에 있어 피치 검출 정확도, 유 / 무성음 판별의 정확도, 동작속도, 알고리듬의 복잡성, 하드웨어 구성의 적합성과 비용등의 이유로 몇가지 방법만이 널리 이용되고 있다. 피치 검출은 신호의 스펙트럼, 세기 그리고 피치에 의해 결정되는 음성의 불규칙성과 다양함 때문에 복잡하다. nonstationary를 없애기 위하여 short analysis window가 사용되야 한다. 그러나 가능한 피치값(3-4 cotaves)의 범위가 광역이기 때문에 해석창은 여러주기와 유 / 무성음이 혼합된 세그먼트를 포함하고 있어야 한다. 피치값과 상관없이 음성 신호를 샘플링 하기 때문에 양자화 잡음이 발생하여 해상도와 정확도에 제한을 받는다.

피치 검출 알고리듬은 기존의 알고리듬과 유사하지만 새로운 모델을 제시해서 이들 문제점을 대부분 극복하고 있다. 이 모델은 두 인접하고 중첩되지 않으며 피치 간격을 갖는 두 신호의 유사도를 평가해서 시간영역에서의 고해상 피치를 정확히 추출해 낸다.[11] 추출된 피치값이 고해상이고 정확하다면 피치 동기 스펙트럼 해석이 가능하다.[12] 본 논문에서는 정확한 피치주기에 대한 상관함수와 문턱값을 계산하고 이 값들을 비교해서 무음, 유 / 무성음 그리고 음소를 분리하는 알고리듬을 제시한다. 이 알고리듬의 결과는 잡음에 강하고 고해상도를 갖는다.

Ⅱ. 피치 결정 알고리듬과 피치 추출 과정

두 신호 x(t), y(t)를 다음과 같이 정의한다.

$$x(t)=s(t)w(t) \quad [t_0, t_0+\tau]$$
$$y(t)=s(t+\tau)w(t) \quad [t_0, t_0+\tau] \qquad (1)$$

여기서 s(t)는 음성신호 w(t)는 길이가 τ인 구형창을 나타낸다. 따라서 두 신호 x(t)와 y(t)는 구간 [[t_0, $t_0+\tau$]에서만 값을 갖는다.

음성의 해석구간은 $t=t_0$에서 시작하고 $\tau=T_0$에서 정확한 피치주기를 갖는다면 첫번째 신호 x(t)는 s(t)의 $[t_0, t_0+T_0]$구간의 신호이고 둘째 신호 y(t)는 $[t_0+T_0, t_0+2T_0]$ 구간의 신호가 된다. 두 연속된 신호의 피치주기는 유사하기 때문에 그것들은 다음과 같이 진폭 변조된 신호로 가정할 수 있다.

$$x(t)=a(t_0)y(t)+e(t) \quad [t_0, t_0+T_0] \qquad (2)$$

여기서 $a(t_0)$는 $t=t_0$에서 glottal pulse에 따라 변하는 진폭변조계수이다. 오차항 e(t)는 두 주기신호 사이의 차이를 나타내며 두 세그먼트 x(t)와 y(t)의 유사도를 최대화하기 위하여 e(t)를 구간 $[t_0, t_0+\tau]$에서 최소함으로서 시간 간격 $\tau=T_0$를 $t=t_0$에서의 피치주기로 정의한다. 최적이론을 적용하기 위하여 cost function C를 정규자승오차로 표현한다.

$$C=\frac{\int_{t_0} [x(t)-a(t_0)y(t)]^2 dt}{\int_{t_0} [x(t)]^2 dt} \qquad (3)$$

C를 최소화하기 위해서는 계수 $a(t_0)$에 대하여 미분한 값이 영이 되도록 한다. 이때 τ값을 피치주기로 설정한다. 실질적으로 음성의 밴드폭이 제한되어 있으므로 τ는 $T_{0min} \leqq \tau \leqq T_{0max}$로 제한될 수 있다.

$$T_0=\underset{\tau,\, a(t_0)}{\operatorname{argmin}} \{C\} \qquad (4)$$

결국 최적변조이득은 $a(t_0)=(\mathbf{X}, \mathbf{Y}) / |\mathbf{X}|^2$, $\mathbf{X}$는 x(t)를 $\mathbf{Y}$는 y(t)를 벡터로 표현한 것이고 $(\mathbf{X},\mathbf{Y})$는 두 벡터의 내적이다. 그리고 $|\mathbf{X}|^2=(\mathbf{X}, \mathbf{X})$는 x(t)의 에너지이다. $a(t_0)$의 최적치를 식(3)에 대입하면

$$C=1-\rho^2(\mathbf{X}, \mathbf{Y}) \quad [t_0, t_0+\tau] \qquad (5)$$

여기서 $\rho(\mathbf{X}, \mathbf{Y})$는 $\mathbf{X}$, $\mathbf{Y}$사이의 상호상관계수이다.

$$\rho(\mathbf{X}, \mathbf{Y})=\frac{(\mathbf{X}, \mathbf{Y})}{|\mathbf{X}|\,|\mathbf{Y}|} \quad [t_0, t_0+\tau] \qquad (6)$$

그러므로 t_0에서 피치주기 T_0는 상호상관계수의 최대치로 계산될 수 있다.

$$T_0=\operatorname{argmax}\{\rho(\mathbf{X}, \mathbf{Y})\} \; T_{0min} \leqq \tau \leqq T_{0max} \qquad (7)$$

그래서 식(4)의 정규자승오차의 최소화는 피치주기

값에 대한 $\rho(\mathbf{X}, \mathbf{Y})$의 최대화와 같아진다. $\rho(\mathbf{X}, \mathbf{Y})$는 창함수의 길이에 따라 변하는 값이므로 두 신호의 상대적인 시지연(time lag)의 함수로 표현된 자기상관함수와 혼동해서는 안된다.

식(6)의 실질적인 해는 신호 s(t)를 샘플주기 T로 일정하게 양자화해서 디지탈신호처리로 구할 수 있다. 이 경우 피치주기는 샘플주기에 의해 결정되는 유한 해상도를 갖는 값이 된다. 여기서 이 주기를 정수피치주기라 하고 음성 샘플 $\mathbf{S}[n_1:n_2]$ $(1 \leqq n_1 < n_2)$의 벡터를 다음과 같이 정의한다.

$$\mathbf{S}[n_1:n_2]=(Sn_1, \cdots, Sn_{1+k}, \cdots, Sn_2)^T \ ; \ k=0, 1, \cdots, n_2-n_1$$
$$S_i=S(t+t_0)_{t=(i-1)T} \ ; \ i=1, 2, \cdots \qquad (8)$$

음성신호가 양자화되었기 때문에 식(1)의 x(t), y(t) 세그먼트는 모든 시간 t에 대하여 얻어질 수 없다. 대신 그것들은 두개의 n차원 벡터 $\mathbf{X}_n(i_0)=(x_1, \cdots, x_j, \cdots, x_n)^T$와 $\mathbf{Y}_n(i_0)=(y_1, \cdots, y_j, \cdots, y_n)^T$로 변형해서 표현한다.

$$\mathbf{X}_n(i_0)=\mathbf{S}[1:n], \ \mathbf{Y}_n(i_0)=S[n+1:2n] \qquad (9)$$

여기서 i_0는 시간첨자 t_0에 대한 샘플첨자이다. 식(4)에 의하여 $t=t_0$에서 최적 정수주기 N_0는 정규이산자승오차함수의 최소화로 구해진다.

$$N_0=\underset{n,\ a(t_0)>0}{\operatorname{argmin}} \left[C=\frac{\sum_{j=1}^{n}[x_j-a(t_0)y_j]^2}{\sum_{j=1}^{n}x_j^2} \right] \qquad (10)$$

실질적인 주기값 N_0는 $N_{min} \leqq n \leqq N_{max}$ 범위에 대하여 구한다. N_{min}과 N_{max}는 각각 T_{0min}과 T_{0max}에 대응한다. 식(10)의 최적화는 식(7)의 이산형으로부터 유도되고 $\mathbf{X}(i_0)$와 $\mathbf{Y}(i_0)$의 상호상관계수 $\rho_n(\mathbf{X}, \mathbf{Y})$는 $N_{min}, N_{max}]$에서 n에 대하여 최대치를 구해야 한다.

$$N_0=\underset{n}{\operatorname{argmax}} \ \rho_n(\mathbf{X}(i_0), \mathbf{Y}(i_0)) \quad N_{min} \leqq n \leqq N_{max} \qquad (11)$$

$\rho_n(\mathbf{X}(i_0), \mathbf{Y}(i_0))$는 식(6)과 같이 구해진다. 식(11)의 해는 $[N_{min}, N_{max}]$ 구간에서 $\rho_n(\mathbf{X}, \mathbf{Y})$를 계산하므로 구할 수 있고 그것의 최대치도 찾을 수 있다.

정수 주기 N_0는 샘플주기에 따른 반올림오차를 포함하는 한정된 해상도를 갖고 평가했다. 샘플의 유리수로 표현된 정확한 주기는 N으로 나타내고 $N=T_0/T_s$로 주어진다. T_0는 피치주기, T_s는 샘플주파수이다. 또한 N의 정수 피치를 N_I로 주기버림오차를 $\beta=N-N_I(0 \leqq \beta<1)$로 표시한다. 이 부분구간에서 무한 해상도를 갖는 정확한 주기를 얻기 위한 절차는 다음과 같다. 이상적으로 식(1)의 x, y 구간을 두 벡터 X, Y로 양자화 했을 때 T_0의 구간에 대하여 y구간의 샘플은 x구간과 같은 상대적인 값을 요구한다. 그러면 두 주기의 샘플은 그들 사이의 상호관계수값이 커지게 된다. 이것은 $\mathbf{Y}_n(i_0)$와 $y(t)(t_0 \leqq t \leqq t_0+T_0)$의 시작점이 일치하는 그런 n의 값이 존재한다는 의미이다.

$$y_1=y(t_0)=s(t_0+T_0) \quad t_0 \leqq t \leqq T_0 \qquad (12)$$

분명히 고정된 임의의 샘플주기 때문에 정확한 주기 N은 일반적으로 정수가 아니다. 그래서 식(9)의 $\mathbf{Y}(i_0)$의 정의에서 볼 때 $N=N_0$와 일치하지 않는다면 식(12)의 조건을 만족하는 정수 n을 찾을 수 없다.

이와같은 동기문제를 해결하기 위하여 벡터 $\mathbf{Y}_{NI}(i_0+\beta)$의 첫번째 값을 식(12)와 같이 $y(t_0)$에 일치해야 한다. 그러나 $\mathbf{Y}_N(i_0+\beta)$의 요소는 $\beta=0$가 아니면 s(t)의 일정한 샘플로 직접 이용할 수 없고 보간법에 의해 평가될 수 있다. 음성신호의 선형예측이론은 주기결정에 사용되기 때문에 선형보간이 충분하며 $\mathbf{Y}_N(i_0+\beta)$는 다음과 같이 두 벡터 $\mathbf{Y}_{NI}(i_0)$와 $\mathbf{Y}_{NI}(i_0+1)$의 선형 결합에 의하여 근사화될 수 있다.

$$\mathbf{Y}_{NI}(i_0+\beta) \simeq (1-\beta)\mathbf{Y}_{NI}(i_0)+\beta\mathbf{Y}_{NI}(i_0+1) \qquad (13)$$

만약 $n=N_I$라면 식(10)의 $\mathbf{Y}_N(i_0)$는 $\mathbf{Y}_{NI}(i_0+\beta)$로 대치되고 여기서 최소화는 $0 \leqq \beta<1$ 범위에서 실행되며 β^*로 표현되는 β의 최소치는 벡터 $\mathbf{X}_{NI}(i_0)$와 $\mathbf{Y}_{NI}(i_0+\beta)$사이의 상호상관계수를 최대화함으로서 식(11)과 같이 계산된다.

$$\beta^* = \underset{\beta}{\operatorname{argmax}} \ \rho_{NI}\{\mathbf{X}(i_0), \mathbf{Y}(i_0+\beta)\} \qquad (14)$$

최대 상관계수는 식(12)를 만족하는 주기에 대해 두 벡터 $\mathbf{X}_{NI}(i_0)$와 $\mathbf{Y}_{NI}(i_0+\beta^*)$를 일직선에 일치시키

므로 β^*를 구할 수 있다. 식(14)의 최적화는 orthogonal projection theorem[11]을 사용해서 최적치 β^*를 구할 수 있다.

$$\beta^*=\frac{(\mathbf{X}(i_0),(\mathbf{Y}(i_0+1))|(\mathbf{Y}(i_0)|^2-(\mathbf{X}(i_0),(\mathbf{Y}(i_0))(\mathbf{Y}(i_0),(\mathbf{Y}(i_0+1))}{(\mathbf{X}(i_0),(\mathbf{Y}(i_0+1))[|(\mathbf{Y}(i_0)|^2-(\mathbf{Y}(i_0),(\mathbf{Y}(i_0+1))]+(\mathbf{X}(i_0),(\mathbf{Y}(i_0))[|(\mathbf{Y}(i_0+1)|^2-(\mathbf{Y}(i_0),(\mathbf{Y}(i_0+1))]} \tag{15}$$

여기서 첨자 N_1는 간편하게 하기 위하여 생략하였다. β의 정의에 따라 정확한 주기 $T_{opt}=(N_1+\beta^*)$이다. 정수 N_1는 주기값 N_0을 기반으로 결정된다.

$$N_1=\min\{N_0, N_1\} \tag{16}$$

N_1은 $\rho_{N0}(\mathbf{X}(i_0), \mathbf{Y}(i_0))-\rho_{N1}(\mathbf{X}(i_0), \mathbf{Y}(i_0)$을 최소화하는 N_0의 인접한 정수로 정의한다.

$$N_1=N_0+\text{sign}[\rho_{N0+1}(\mathbf{X}(i_0), \mathbf{Y}(i_0))-\rho_{N0-1}(\mathbf{X}(i_0), \mathbf{Y}(i_0))] \tag{17}$$

여기서 $\rho_N(\mathbf{X}(i_0), \mathbf{Y}(i_0)$는 $N_{min}, N_{max}]$ 밖에서는 0으로 가정한다. 결국 상호상관계수의 최대치 ρ^* 는 다음과 같다.

$$\rho^*=\rho_{N1}(\mathbf{X}(i_0),(\mathbf{Y}(i_0+\beta))$$
$$=\frac{(1-\beta^*)\mathbf{X}(i_0),(\mathbf{Y}(i_0))+\beta^*(\mathbf{X}(i_0)\mathbf{X}(i_0+1)}{[|(\mathbf{X}(i_0)|^2((1-\beta^*)^2|\mathbf{Y}(i_0)|^2+2\beta^*(1-\beta^*)(\mathbf{Y}(i_0),(\mathbf{Y}(i_0+1))+\beta^{*2}(\mathbf{Y}(i_0+1)|^2)]^{1/2}} \tag{18}$$

정수피치 N_0T로부터 정확한 주기평가 T_{opt}를 구하기 위하여 단지 두개의 내적 $(\mathbf{X}(i_0), \mathbf{Y}(i_0+1))_{N1}$와 $(\mathbf{Y}(i_0), \mathbf{Y}(i_0+1))_{N1}$만을 식(15)의 β^*을 구하기 위하여 계산하며 $(\mathbf{X}(i_0), \mathbf{Y}(i_0))_{N1}$와 $|\mathbf{Y}(i_0)|^2$는 정수피치를 계산하는 과정에서 $\rho_{N1}(\mathbf{X}(i_0), \mathbf{Y}(i_0))$을 위하여 이미 계산되었고 $|\mathbf{Y}(i_0+1)|^2$은 $\mathbf{Y}(i_0)^2$로부터 쉽게 구할 수 있다. 그래서 평가된 주기의 해상도 향상과 샘플링 주기 반올림오차의 제거는 정수피치 평가에 대하여 상대적으로 덜 복잡하게 구할 수 있다. 어떤 경우에는 β^*가 [0, 1) 을 벗어날 때도 있다. 이런 경우는 정수피치 N_0가 참값으로부터 한 샘플에 의해 유도될 때 발생할 것이다. 그런 경우 정수피치는 $\beta^* \geq 1$일때 1 증가시키고 $\beta^* <0$이면 1 감소시킨다. 그런 후 β는 식(15)으로 재평가 된다.

$\rho_n(\mathbf{X}, \mathbf{Y})$는 $[N_{min}, N_{max}]$ 구간에서 $n=kN_0(k>1)$의 상관계수값이 $n=N_0$의 값 보다 더 큰 값을 갖는 경우가 생긴다. 이것은 신호가 변이구간 또는 음성의 시작되는 구간에서 nonstationary하기 때문에 생기고 이 경우 피치를 보상해야 한다. 피치를 보상하기 위하여 상관계수가 일정한 문턱값 이상의 피치 후보를 찾아서($n_1, n_2, \cdots, n_M$), 이 중 피치간격이 가장 큰 값(n_M)을 상관계수의 길이로 해서 각 피치후보에 대하여 상관계수를 구하고 처음으로 문턱값 이상인 값을 보상된 피치로 한다. 이 때 문턱값은 실험상 얻어진 값으로 보통 0.8−0.85를 취한다.

$$\mathbf{X}=[1:n_M],\ \mathbf{Y}=[n_k:n_k+n_M-1]$$
$$\rho(\mathbf{X}, \mathbf{Y})>TH1\ ;\ TH1=0.8-0.85$$
$$N_0=n_k\ ;\ k=1, 2, \cdots, M \tag{19}$$

한편 음성구간내에서 피치 사이의 편차는 제한되어 있다. 이 편차는 보통 피치값의 ±15% 이내이고 결코 25%를 초과할 수 없다. 그러므로 과도현상이 안정화된 후(예: 음성의 3-5주기) 정수피치를 모든 구간$[N_{min}, N_{max}]$ 대신 단지 전피치의 주변에서만 찾는다. 이 같은 탐색구간의 축소, 한편으론 계산량의 축소는 피치값에 상관없이 smooth pitch contour를 갖는다.

유 / 무성음 판정과 피치추출값의 선택은 상호상관계수값과 문턱값 $T_{low}(t)$의 비교로 이루어진다. 문턱값의 설정은 음성구간 특히 음성 시작점에서 피치를 검출하기 위해서는 충분히 작아야 하며 고상관값에서는 불규칙하기 때문이 무성음 구간동안 잘못 찾는 것을 피하기 위해 충분히 큰 값을 갖도록 해야 한다. 다른 소리, 다른 화자 그리고 배경 잡음 때문에 상관값의 변화에 대처하는 문턱값의 고정값을 결정하는 것은 상당히 어렵다. 좀더 좋은 방법은 각 순간에 문턱값을 그때 구한 인접 피치주기의 상관계수값에 비례하는 값으로 채택한다.

두개의 경계치 $T_{low}(t)$와 T_{high}를 정의한다. $T_{low}(t)$는 음성구간의 문턱치이고 T_{high}는 무성음구간의 문턱값이다. $T_{low}(t)$는 각 피처에서 다음식에 따라 계산된다.

$$T_{low}(t)=\max\{T_{min} : T_{max}\} \quad (20)$$

여기서 T_{min}은 $T_{low}(t)$의 최소 경계치로 사용된 고정값이고 T_{max}는 현재 음성구간에서 구한 최대 상관계수에 비례하는 값이다. $T_{low}(t)$는 적어도 최소경계치 T_{min}과 같아야 하지만 일반적으로 유성음구간에서 상관값이 증가함에 따라 $T_{low}(t)$는 증가한다. 실험을 통해 얻은 경계치는 각각 $T_{min}=0.8$, $T_{high}=0.85$, $T_{max}\{\rho(t)\}$이다. 그림 1은 음성신호 "백두산"에 대한 피치주기, 그림 2는 문턱값과 상호상관계수 그리고 그림 3은 음성신호를 나타낸다.

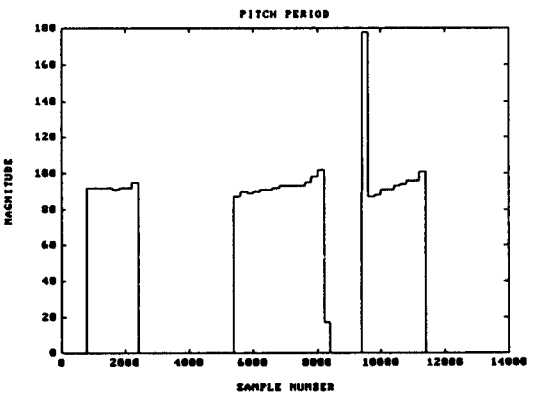


그림 1. 피치주기 "백두산"
Fig. 1. Pitch period "Pack tu san"

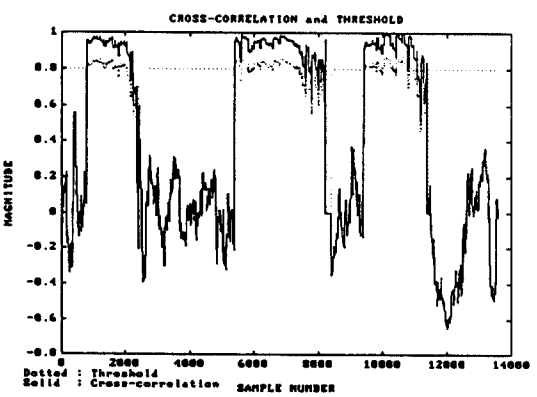


그림 2. 문턱값 "백두산"
Fig. 2. Thresold "Pack tu san"

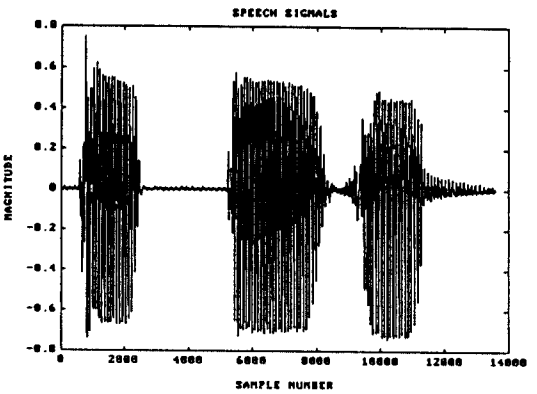


그림 3. 음성신호 "백두산"
Fig. 3. Speech signals "Pack tu san"

Ⅲ. 음소 분리 과정

Ⅱ장에서 구한 피치주기와 문턱값을 비교해서 유성음과 무성음 및 변이구간으로 구별할 수 있다. 음소분리과정을 전체 흐름도로 나타내면 그림4와 같이 표시된다. 먼저 음성구간과 무음구간을 구분하기 위하여 한 피치주기에 대한 에너지 즉 norm값을 문턱값으로 설정한다. 일반적으로 무음구간은 음성구간 보다 에너지값이 현저하게 작으므로 아주 작은 값으로 설정한다. 본 논문에서는 입력데이타의 평균 norm을 계산한 실험결과를 토대로 문턱값을 0.01로 선정했다.

일정한 주기를 갖는 유성음구간에서 상호상관계수가 문턱값 $T_{low}(t)$보다 큰 값을 갖고 무성음구간이나 변이구간에서는 상호상관계수가 문턱값 $T_{low}(t)$보다 작은 값을 갖는다. 그래서 상호상관계수와 문턱값 $T_{low}(t)$를 비교해서 유성음구간을 구분한다. 음성의 변이구간에서는 일시적으로 주기가 변하고 상호상관계수가 문턱값 $T_{low}(t)$보다 작아지나 다음 음소부분으로 정상상태가 되면 상호상관계수는 문턱값 $T_{low}(t)$보다 커진다. 그러나 무성음구간에서는 상호상관계수가 $T_{low}(t)$보다 계속 낮아지고 주기는 $N_{min} \leq N \leq N_{max}$ 범위를 벗어나게 된다. 음성은 약 20msec 구간에서 stationary하므로 해석프레임을 N_{max}(20 msec)로 설정하고, $N_{max}/2$구간씩 겹쳐서 알고리듬을 수행한다.

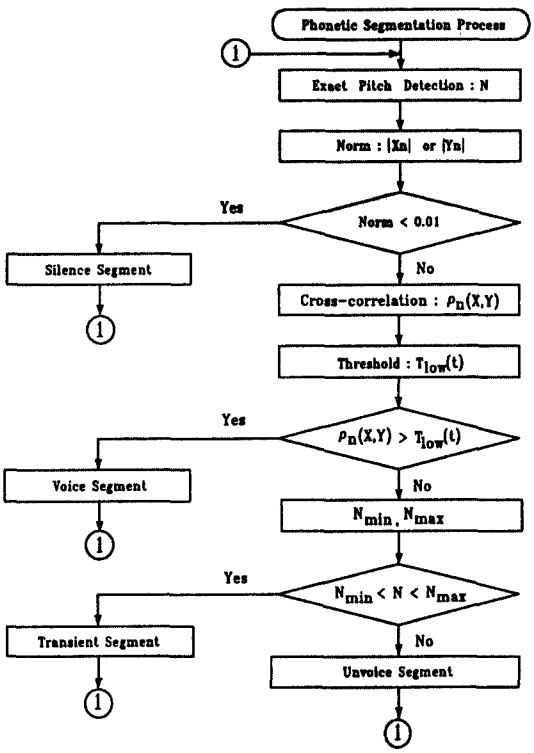


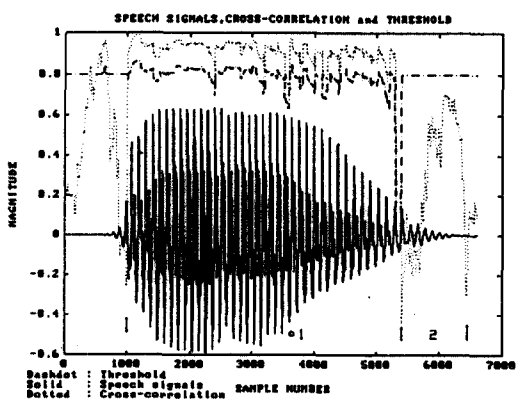
그림 4. 음성분리과정 흐름도
Fig. 4. Flow chart for phonetic segmentation process

Ⅳ. 실험 및 검토

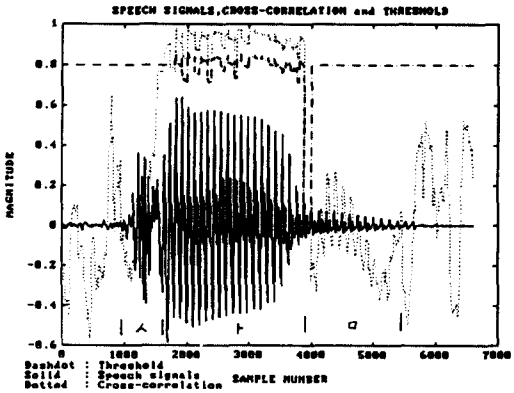
본 연구에서는 음성 입력 장치로 마이크로폰을 통해 음성신호를 ±5V 레벨로 증폭한 다음 Global Lab 상에서 DT2821 보드로 12bit Sampling하여 입력한다. Sampling rate는 10Khz로 하였고 DT2821 입력단에 4Khz의 밴드폭을 갖는 아날로그 저역통과필터를 두었다. Global Lab 상의 데이타 화일을 MATLAB의 MAT화일로 변환한 다음 PC 386 / DX에서 MATLAB으로 알고리듬을 수행하였다.

음성 데이타로 1음절, 2음절, 3음절 단어로 구분하여 실행하였다. 1음절인 경우 숫자 "영"부터 "구"까지의 10개의 데이타, 2음절인 경우 "한글" 등 10개의 데이타, 3음절인 경우 "설악산" 등 10개의 데이타를 받아 실험하였다. 단음절인 경우 음절 사이의 상호영향을 받지 않으므로 정확하게 구분되었다. 그러나 "일", "영", "이"와 같이 "ㅇ+모음"이나 "모음+ㅇ"은 "ㅇ" 자음에 대한 특성때문에 아주 짧은 구간에서 상관계수값이 변함을 알 수 있다. 유성자음과 모음구간에서는 일정한 피치값을 갖는 신호를 나타내고 무성자음구간에서는 피치값을 갖지않고 랜덤신호에 가까운 특성을 나타냄을 볼 수 있었다. 위의 실험결과는 다음 그림 5에 표시하였다.

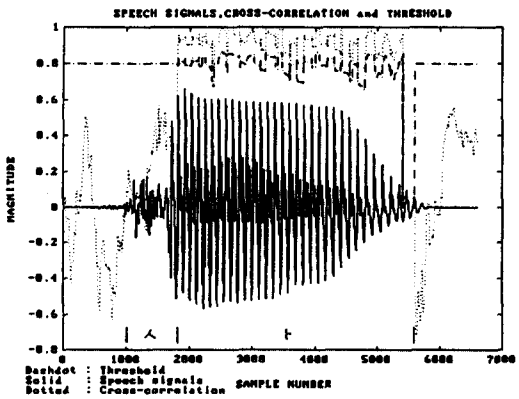
다음절 단어의 경우 화자의 발성 속도에 따라 음절과 음절사이에서 자음의 음소 구별이 부정확한 경우가 발생한다. 실질적으로는 자음구간은 모음구간보다 상당히 짧은 기간에 나타나므로 첫째음절의 종성과 둘째음절의 초성 사이에서 화자의 발성속도가 빠를 때 두 자음이 겹쳐서 구분하기 어렵게 된다. 그러나 실험을 통하여 초당 4음절 이상의 속도로 발성한 경우 구분이 되지 않았으나 실질음성을 들어보면 사람이 구분하기 힘든 경우이므로 위의 문제는 고려대상이 되지 않는다. 그림 6은 2음절과 3음절단어의 음소분리를 나타낸 것이고 그림 7은 발성속도에 따른 음소분리 가능성 여부를 표시한 것이다. 그림 7의 (a)는 초당 1.5 음절로 분리 가능한 것이고 그림 7의 (b)는 초당 4음절로 분리할 수 없는 경우이다. 음소분리과정을 통해 얻은 결과를 확인하기 위하여 각각의 음소구간을 Hypersignal-Workstation DSP Software에서 각 음소 구간별로 나누어 DT2821의 D / A변환기를 통해 출력된 음성을 반복해서 들어본 결과, 유성음구간에서는 음소 단위의 발음을 정확히 알 수 있었으나 무성자음인 경우 정확하게 구분 되지 않고 새로운 음소임을 확인할 수 있다. 예를들어 "한"인 경우 "ㅎ" 구간, "ㅎ"+"ㅏ"구간, "ㅎ"+"ㅏ"+"ㄴ" 구간으로 나누어 들어 볼때 초, 중, 종성의 발음을 확인할 수 있었다. 음성의 시작부분과 변이



(a)"일"
(a)"Ir"

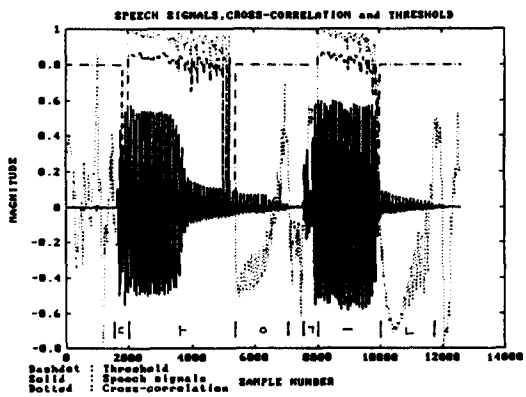


(b)"삼"
(b)"sam"

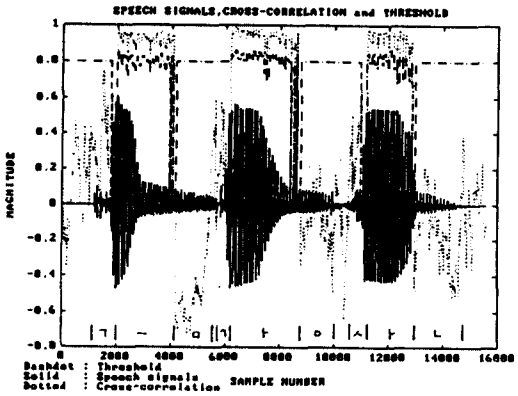


(c)"사"
(c)"Sa"

그림 5. 단음절 단어의 음소분리
Fig. 5. Phonetic segmentation of a syllable word

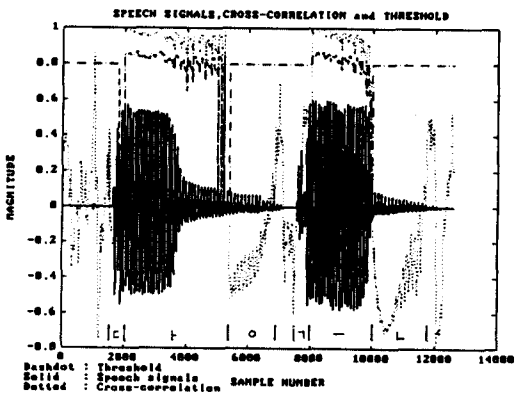


(a)"당근"
(a)"Tang keun"

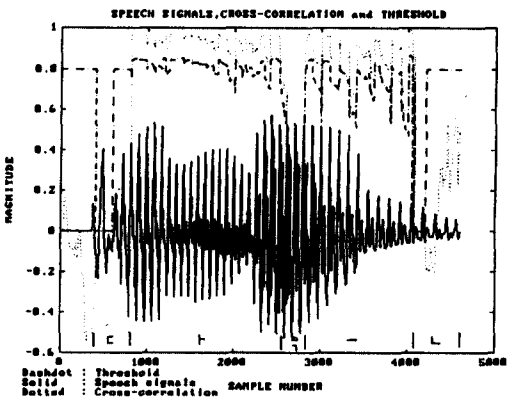


(b)"금강산"
(b)"Keum kang san"

그림 6. 2음절 또는 3음절 단어의 음소분리
Fig. 6. Phonetic segmentation of two or three syllable word



(a)1.5음절 / 초
(a)1.5 syllabes / sec



(b)4음절 / 초
(b)4 syllables / sec

그림 7. 발성속도에 따른 음소분리
Fig. 7. Phonetic segmentation with pronunciation rate

구간에서 피치 주기의 변화가 심하므로 음소분리가 조금씩 다른 특성을 갖기는 하지만 한, 두개의 피치 차는 음성합성시 거의 구분이 되지 않으므로 크게 문제가 되지 않는다. 유성음 구간에서는 인접신호들 간의 상관도가 높으므로 정확하게 음소분리가 가능하였다.

V. 결 론

본 논문에서는 음성 인식 또는 벡터 코딩 이론에 적용될 한국어 음성의 음소분리를 위한 새로운 알고리듬을 제시하고 PC 386 / DX상에서 MATLAB으로 시뮬레이션하고 실험을 통해 확인하였다. 음성신호에 대한 피치주기는 정확히 찾을 수 있었으나 변이 구간과 음성의 시작부분에서의 피치주기의 변화로 음소분리에 오차가 조금 발생하였다. 그런데 이 오차는 그 구간에서의 문턱값을 조절해서 오차를 개선할 수 있었다.

본 논문에서는 문턱값을 설정할 때 현재의 상호상관계수에 일정한 비율로 선정하였는데 인접한 상호상관계수들 사이의 예측값으로 적응해 나가도록 선정하면 훨씬 더 정확한 음소분리를 할 수 있을 것이다. 제안된 알고리듬을 기반으로 실시간 처리를 위한 그 관련된 DSP 알고리듬으로의 변환과 하드웨어 설계가 앞으로의 연구과제로 요구된다.

참 고 문 헌

1. A. M. Noll, "Cepstrum pitch determination," *J. Acoust, Soc. Amer.*, vol. 41, pp.293-309, Feb. 1967.
2. M. N. Sondhi, "New methods of pitch extraction," *IEEE Trans. Audio Electroacoust.*, vol. AKU-16 , pp.262-266, June 1968.
3. B. Gold and L. R. Rabiner, "Parallel processing techniques for estimating pitch periods of speech in the time domain," *J. Acoust. Soc. Amer.*, vol. 46, pp.442-448, Aug. 1969.
4. J. D. Markel, "The SIFT algorithm for fundamental frequency estimation," *IEEE Trans. Audio and Electioacoust.*, vol. AU-20, pp.367-377, Dec. 1972.
5. J. N. Maksym, "Real-time pitch extraction by adaptive prediction of the speech waveform," *IEEE Trans. Audio and Electroacoust.*, vol. AU-21, no.3, June 1973.
6. M. J. Ross et al., "Average magntude difference function pitch extractor," *IEEE Tranc., Acoust., Speech, Signal Processing*, vol. ASSP-22, pp.353 -362, Oct. 1974.
7. L. R. Rabiner et al., "A comparative performance study of several pitch detection algorithms," *IEEE Tranc., Acoust., Speech, Signal Processing*, vol. ASSP- 24, no.5, Oct. 1976.
8. L. R. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE Trans., Acoust., Speech, Signal Processing*, vol. ASSP-25, no.1, Feb. 1977.
9. S. Seneff, "Real-time harmonic pitch detector," *IEEE Tranc. Acoust., Speech, Signal Processing*," vol. ASSP-26, pp.358-365, Aug. 1978.
10. M. Lahat et al., "A spectral autocorrelation method for measurement of the fundamental frequency of noise corrupted speech," *IEEE Tranc., Acoust., Speech, Signal Processing*, vol. Assp-35, no.6, June 1987.
11. Y. Medan et al., "Super resolution pitch determination of speech signals," *IEEE Tranc., Signal Processing*, vol. 39, no.1, pp.40-48, Jun. 1991.
12. Y. Medan and E. Yair, "Pitch synchronous spectral analysis scheme for voiced speech," *IEEE Tranc., Acoust., Speech, Signal Processing*, vol. ASSP- 37, no.9, pp.1321-1328, Sept. 1989.

李 應 求(Eung Gu Lee) **正會員**

1961년 10월 28일생

1985년 : 한양대학교 전자공학과 졸업(학사)

1987년 : 한양대학교 대학원 전자공학과 졸업 석사

1987년~현재 : 한양대학교 대학원 전자공학과 박사과정 재학중

※주관심분야 : 디지탈 신호처리, 음성및 영상시스템, 의용공학 등임

李 斗 秀(Doo Soo Lee) **正會員**

1946년 7월 3일생

1968년 : 전북대학교 전자공학과 졸업(학사)

1970년 : 전북대학교 대학원 전자공학과 졸업(석사)

1973년 : 전북대학교 대학원 전자공학과 졸업(박사)

현재 : 한양대학교 전자공학과 부교수

※주관심분야 : 디지탈 신호처리, 음성및 영상시스템, 의용공학, 원격제어시스템등임