

# 피치변경율에 따른 최적의 피치변경법에 관한 연구

박형빈, (\*)조왕래, 김종득, 박 원, 심도식, 배명진

송실대학교 정보통신공학과, (\*)백성대학 전자과

## A Study on the Optimal Pitch Alteration According to the Pitch Alteration Ratio

HyungBin Park, (\*)WangRae Jo, JongDeuk Kim, Won Park, Dosik Sim, MyungJin Bae

Dept. of Telecom. Engr. Soongsil Univ.

(\*)Dept. of Electronic Engr., Byuksung College

hbpark@assp.soongsil.ac.kr, mjbae@saint.soongsil.ac.kr

### 요 약

일반적으로 파형 부호화법에서는 발생자의 개성과 매시지 정보를 보존하여 음질의 명료성이 우수하다. 하지만 파형 부호화법이 규칙에 의한 합성에 적용되도록 하기 위해서는 운율 조절에서 피치 변경법이 요구되어진다.

본 논문에서는 스펙트럼의 왜곡율을 최소화하면서도 위상을 그대로 보존할 수 있는 시간영역에서의 피치변경법을 적용하였다. 이러한 방법들은 시간영역에서만 처리됨으로써 계산량을 줄여 실시간 운율조절에 적합하다. 결과적으로 피치주기를 100%에서 50%로 압축한 경우 평균 스펙트럼 왜곡율은 5.62%이고, 피치주기를 100%에서 200%로 신장한 경우 평균 스펙트럼 왜곡율은 4.20%정도로 비교적 우수한 결과가 얻어졌다.

### 1. 서론

최근 다양해진 음성서비스 분야에서는 고음질의 합성음을 요구하고 있다. 이러한 고음질 합성방식으로는 파형부호화법이 바람직하다. 그러나 파형부호화법을 사용하면 메모리 용량이 증가하고 음원피치의 변경이 어렵다는 문제점이 있다. 부호화에 필요한 메모리 문제는 현재의 기술수준으로 충분히 극복이 가능하다. 나머지 문제의 해결법으로는 파형부호화법을 규칙에 의한 합성에 적용되도록 음원피치를 변경시킬 수 있어야 한다.

지금까지 제안된 피치변경법은 그 처리영역에 따라 시간, 주파수, 시간-주파수 혼성처리법이 있다.

시간영역법에는 Multi-Pulse법, 피치반분법등이 있다.

Caspers와 Atal은 MPLPC에서 펄스사이에 영을 삽입하거나 삭제하는 방법을 제안했으나, MPLPC상의 펄스 열은 피치와 포먼트에 대한 상호 연관을 갖고 있으므로 스펙트럼의 왜곡이 심하다[3]. Varga와 Fallside는 LPC계수를 이용한 피치연장법을 제안했으나, 이 방법은 피치주기를 줄이는 경우에는 단지 파형의 일부분을 소거하고 평활화하는 방법을 사용하고 있기 때문에 스펙트럼의 왜곡이 많이 나타난다[4]. 피치반분법은 임의로 변경하려는 피치주기의 2배 파형을 만든 후에 그 파형의 주기를 반분하는 피치변경법이다[6].

Quatieri와 McAulay는 주파수영역에서 위상을 보존하는 피치변경법을 제안하였는데 이것은 입력된 음성에 대해 진폭 및 위상스펙트럼을 추출하여 별도로 처리하는 방법이다. 진폭스펙트럼에 대해서는 두드러진 스펙트럼 봉우리들을 추출한 다음에 이것을 피치변경율( $\rho$ )만큼 인터플레이션하여 진폭스펙트럼의 피치를 변경시킨다. 위상스펙트럼에 대해서는 시간영역에서 구한 피치 개시시간(pitch onset time)에 해당하는 위상을 제거하고 나서 피치가 변경되었을 때의 새로운 피치 개시시간을 디해줌으로서 새로운 위상을 구성하게 된다. 그렇지만 피치변경시에 피치주기와는 별도로 피치의 개시시간을 공급해 주어야 하고, 또한 진폭 스펙트럼 상에서 두드러진 봉우리 위주로 고조파의 보간을 수행하기 때문에 스펙트럼의 왜곡이 높아진다는 단점이 있다.

시간-주파수 혼성법으로는 켈스트럼의 특징을 이용하여 켈스트럼값이 거의 영이 되는 부분에서 영값을 삽입하거나 삭제함으로써 피치를 변경하는 방법이 있다[5]. 그러나 이 방법 역시 위상의 보존이 어렵다는 문제점을 가지고 있다. Takagi와 Miyasaka가 제안한 시간-

주파수혼성법은 시간영역에서 피치변경을 하였을 때에 나타나는 스펙트럼 왜곡을 스펙트럼영역상에서 LPC포락 을 통해 수정하는 방법이다[7]. 이 방법은 LPC스펙트럼 포락이 갖는 극점에 치중된 시스템 전달 특성 때문에 모든 유성음에 적합하지 못하다는 한계성을 갖는다.

## II. 유성음의 성질

유성음은 폐에서 올라온 공기가 성문을 통하여 배출 될 때 생성되므로 공진을 수반한다. 그리고 성도에서의 공명으로 인하여 에너지가 크고 준-주기적인 형태의 신호가 된다.

이를 주파수영역에서 살펴보면 성도의 공명봉우리에 음성신호의 기본주파수  $F_0$ 가 세세하게 나타나고 있다.

이렇게 성도의 공명봉우리에 해당되는 주파수들을 포 만트라고 하고 가장 낮은 주파수를 갖는 봉우리를 제 1 포만트,  $F_1$ , 이라 한다. 또한 포만트들은 대역폭을 갖게 되므로 시간영역의 한 피치구간에서는 감쇄진동을 하게 된다.

$F_1$ 이 주파수 영역에서 다른 포만트들보다 훨씬 높은 에너지 봉우리를 갖기 때문에  $F_1$ 만을 고려하여 다음과 같이 근사적인 방법으로 성도를 분석할 수 있다.

$$\begin{aligned} h(t) &= \int_{-\infty}^{\infty} F(f) e^{j2\pi ft} df \\ &= \int_{-BW/2}^{BW/2} \cos\left(\frac{2\pi f}{BW}\right) e^{j2\pi ft} df * 2\cos\left((2\pi F_1 t) - \frac{\pi}{2}\right) \\ &= \frac{4BW}{\pi - 4\pi BW^2} \cos\left((2\pi F_1 t) - \frac{\pi}{2}\right) \end{aligned} \quad (2.1)$$

여기서  $F_1$ 는 제 1 포만트의 주파수이고  $BW$ 는  $F_1$ 이 갖는 대역폭이다.

식(2.1)에서는  $F_1, BW$ 가 시간영역에서의 오실레이션 을 결정하는데, 여기서  $F_1 \gg BW$  라면 오실레이션은  $F_1$ 에만 의존하게 된다.

또한 임펄스열 발생기는 피치주기로 할당된 단위 임 펄스의 시퀀스를 발생한다. 다음에 이 신호는 임펄스 응답으로 성문파형  $g(n)$ 을 여기한다.  $g(n)$ 의 형태는 단 적으로 특징지을 수 없지만, Resonberg에 의해 다음과 같이 합성 펄스파형 형태로 제시되었다[1].

$$\begin{aligned} g(n) &= \frac{1}{2} \left(1 - \cos\left(\pi \frac{n}{N_1}\right)\right), \quad 0 \leq n \leq N_1 \\ &= \cos\left(\pi \frac{n - N_1}{2N_2}\right), \quad N_1 \leq n \leq N_1 + N_2 \\ &= 0, \quad \text{otherwise} \end{aligned} \quad (2.2)$$

$g(n)$ 이 유한 길이이므로 전극 모델이 바람직하게 되며,  $G(Z) = Z[g(n)]$ 에 대해 이극형 전달함수로

보통 모델링 하고 있다. 그리고 공기중 방사효과는  $R(z) = R_0(1 - z^{-1})$ 로 나타낼 수 있으며, 이는 고역 필터로 동작하여 성도의 주된 공명효과를 강조시키게 된다.

결국 유성음의  $S_o(n)$ 은 식(2.1)과 식(2.2)가 시간영역에서 컨벌루션 된 것으로 다음과 같이 나타나게 된다.

$$S_o(n) = h(n) * g(n) \quad (2.3)$$

## III. 영교차 단위의 시간축조절 피치변경법

유성음 한 피치주기 동안의 파형 모양은 식 (2.3)의 성문펄스파형과 식 (2.1)의 제 1 포만트 파형이 서로 컨벌루션되어 나타나기 때문에 첫 파형의 봉우리는 성문 파의 모양과 유사하다. 그러므로 이 파형은 인근 봉우 리파형에 비해 영교차 간격이 길고 봉우리의 진폭도 높 게 나타난다.

음성 스펙트럼은 기본주파수의 고조파로 이루어진 여 기스펙트럼과 포만트 공명봉우리가 곱해진 형태를 띠게 된다. 여기스펙트럼에서 고조파 봉우리들의 대역폭은 포만트 봉우리에 비해 아주 좁은 편이다. 따라서 피치 변경시에 여기스펙트럼의 고조파 간격을 변경시키면 음 성스펙트럼에서 제 1 포만트의 위치  $F_1$ 은 기본주파수  $F_0$ 의 거의 정배수를 나타내게 되고, 고조파의 위치에 따라 추정되는 포만트의 위치는 다음 범위내에서 변화 를 관찰할 수 있다.

$$F_1 = F_1 \pm \frac{1}{2} F_0 \quad (3.1)$$

이것을 시간영역에서 설명하면 한 피치 주기내에서 제 1 포만트 파형은 피치주기에 거의 정배수가 되고, 피치를 변경하는 경우에도 피치주기와 제 1 포만트의 주기는 정배수로 일치시키는 것이 필요하다. 시간영역 에서 음성 파형에 대해 쉽게 처리할 수 있는 것은 영교 차 단위로 피치주기를 일치시키는 것이다.

먼저 피치주기를 압축시키기 위해 음성파형의 일부분 을 삭제시키면 한 피치 주기파형의 모서리에서 스펙트 럼 누설현상이 발생되기 때문에 파형의 평요성이 저하 된다. 이것은 피치주기와 파형에 포함된 제 1 포만트의 주기가 불일치되어 나타나는 현상이다. 이러한 경우에 파형 모서리의 스펙트럼누설을 제거하려면 피치 변경된 파형에 대해 시간축 조절을 수행하여 파형의 영교차 점 을 변경된 피치주기에 일치시키면 된다. 이렇게 하면 시간축 조절에 의해 포만트의 주파수가 일부 변경되지 만, 피치주기와 포만트의 주기가 거의 정약수배로 일치 하게 된다.

#### IV. 진폭대칭법에 의한 피치변경법

피치주기를 100%이상 변경하려면 G-peak 사이의 성도 포먼트 정보는 그대로 유지하면서 G-peak의 간격을 신장시켜야한다.

진폭대칭법에 의한 피치변경법은 배주기를 기준으로 피치변경시에 음소의 명료성이 유지되고 버즈음이 발생하지 않다[10][11]. 유성음의 파형 형태는 한 피치주기 내에서 G-peak의 파형이 두드러진 봉우리를 이루고 다음 피치가 시작될 때까지 파형의 진폭이 감소하면서 진동하는 구조를 나타내게 된다. 이러한 파형은 거의 직각삼각형의 윤곽을 이루게 되며 직각삼각형의 밑변의 길이가 피치주기를 구성한다. 피치를 신장시키기위해 먼저 한 피치주기의 파형을 그대로 두고 신장되는 구간에는 파형을 대칭적으로 삽입하면 된다.

$$s'(n) = s(n), \quad n=0,1,\dots,p-1$$

$$= s(p-n), \quad n=p, p+1, \dots, p \cdot r-1 \quad (4.1)$$

여기서  $s'(n)$ 는 피치신장된 파형,  $p$ 는 원래 음성신호  $s(n)$ 의 피치주기,  $r$ 은 피치 신장율이다.

이러한 진폭대칭법에 의한 피치변경법은 피치의 2배 주기까지 신장시킬 동안에 원래 음성파형의 한 피치구간이 그대로 유지되기 때문에 명료성 열하가 거의 없다. 또한 파형의 위상특성도 원래 파형의 피치주기 특성을 그대로 보존하기 때문에 인근 피치와의 파형 연결이 자연스럽게 된다.

#### V. PSOLA 합성방식에 의한 피치변경법

기존의 PSOLA 합성방식은 먼저 원래의 음성 파형을 피치주기 단위로 분해한 다음 분해된 피치 단위에 윈도우 함수를 곱해서 ST(Short-Term)신호의 열로 만든다. 분해된 단위의 운율조절을 하고 이렇게 조절된 단위로 부터 음성을 합성한다.

##### 1) 분석

원래 음성 파형이 유성음인 경우에는 피치단위로 분해한 다음 윈도우 함수를 곱하여 ST신호의 열로 만든다. 무성음인 경우에는 10ms의 주기로 일정하게 분석한다. 분석 윈도우 함수에는 다음과 같은 Hanning, Hamming, Blackman 등의 형이 쓰인다.

이런 윈도우 함수를 원래의 음성 샘플에 곱하므로써 다음 (5.1)식과 같은 피치단위로 분해된 샘플열들을 얻는다.

$$S_{analysis}(n) = W_{analysis}(m-n)S(n) \quad (5.1)$$

$S_{analysis}(n)$  : 피치주기 단위의 ST신호

$W_{analysis}(n)$  : 분석 윈도우 함수

$m$  :  $m$ 번째 피치

$S(n)$  : 원 음성 파형

##### 2) 운율 조절 및 합성

분석과정에서의 ST신호의 열은 원래의 음성 샘플의 피치 단위로 배열되어 있다. 따라서 피치를 변경하기 위해서는 이 간격들을 변경할 피치 간격들로 재배열하면 된다. 다음 (5.2)식은 피치가 변경된 신호를 나타낸 것이다.

$$S_{synthesis}(n) = S_{analysis}(n - m_a) \quad (5.2)$$

$S_{synthesis}(n)$  : 피치가 변경된 ST신호

$m_a$  : 변경할 피치 간격

따라서 피치를 높일때는 ST신호의 간격을 작게 배열하고, 피치를 낮출때는 ST신호의 간격을 크게 배열하면 된다. 하지만 이런 순차적인 배열사이에서 정확한 피치 동기화를 유지하는 것이 중요하다. 이렇게 재배열된 ST신호에서 겹쳐지는 부분은 더해지면 된다.

#### VI. 피치변경율에 따른 최적의 피치변경법

본 논문에서는 스펙트럼의 왜곡율을 최소화하면서도 위상을 그대로 보존할 수 있는 시간영역에서의 피치변경법을 적용하였다. 피치변경율에 따라 스펙트럼 왜곡율이 크게 증가하지 않게 하려면 포먼트 주파수와 기본 주파수를 정수배로 유지시키는 것이 필요하다. 따라서 피치주기를 압축하는 경우에는  $F_1$ 과  $F_0$ 의 운율 정수배로 유지시키시 스펙트럼 왜곡을 최소화하는 영교차단위의 시간축 조절에 의한 피치변경법을 적용하였다[12].

이 방법은 시간축 조절을 제 1 포먼트의 주파수 범위 내로 제한하여 변경율이 높아질 때 발생하는 스펙트럼 왜곡을 줄였다.

또한 피치주기를 신장하는 경우에는 스펙트럼과 위상의 왜곡이 최소가 되는 배주기 파형을 만든다음에 이들을 기준으로 피치를 변경하는 것이 바람직하다. 피치주기를 신장하는 경우에 적용된 진폭대칭법에 의한 피치변경법은 배주기의 피치를 기준으로 피치주기를 신장하는 시간영역법이다[10][11]. 이것은 피치주기를 신장할 때 원래의 한 피치주기 일부 파형을 시간축에 역정렬시키는 것이다. 따라서 성도의 포먼트 성분이 그대로 보존되고, 파형의 위상특성을 그대로 보존할 수 있는 특징이 있다. 하지만 피치주기 변경율이 150%에서 100% 사이에서는 파형이 서로 중첩되는 부분이 증가하

기 때문에 명료성이 저하되는 단점을 가지고 있다. 따라서 본 논문에서는 피치변경율이 150%에서 100% 사이에서는 PSOLA 합성방식에 의한 피치변경법을 적용하였다. 다음 그림6-1은 영교차 단위의 시간축 조절 피치변경법에 의해서 피치주기를 압축시킨 경우의 예를 나타낸 것이다.

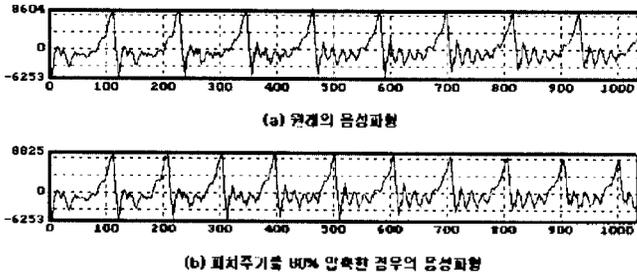


그림6-1. 영교차단위의 시간축 조절에 의한 피치변경법

다음 그림6-2는 PSOLA 합성방식에 의한 피치변경법과 진폭대칭법에 의한 피치변경법에 의해서 피치주기를 신장시킨 경우의 예를 나타낸 것이다.

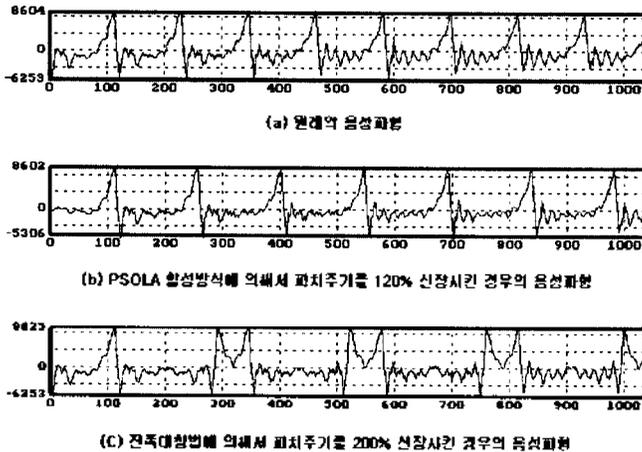


그림6-2. PSOLA방식과 진폭대칭법에 의한 피치변경법

결과적으로 이러한 방법들은 시간영역에서만 처리됨으로써 계산량을 줄여 실시간 운용조절에 적합하다. 또한 그림6-1과 그림6-2에서 볼 수 있듯이 성도의 포먼트 특성의 왜곡은 작고, 파형의 위상특성을 그대로 보존할 수 있다는 장점을 가진다.

VII. 실험 및 결과

본 논문에서 제안한 방법을 시뮬레이션 하기 위해 IBM-PC/586에 마이크 입력이 가능한 16비트 A/D변환기를 인터페이스하여 2명의 남성과 2명의 여성화자를 통해 다음 음성시료를 발생하게 하고 이를 11kHz의 표본화율로 16비트 양자화 하였다.

- 발성 1: /인수네 꼬마는 천재소년을 좋아한다./
- 발성 2: /승실대 정보통신과 음성통신연구팀이다./
- 발성 3: /예수님께서 천지창조의 교훈을 말씀하셨다./
- 발성 4: /공일이삼사오육칠팔구/

다음 그림 7-1은 본 논문에서 제안한 방법을 블록도로 나타낸 것이다.

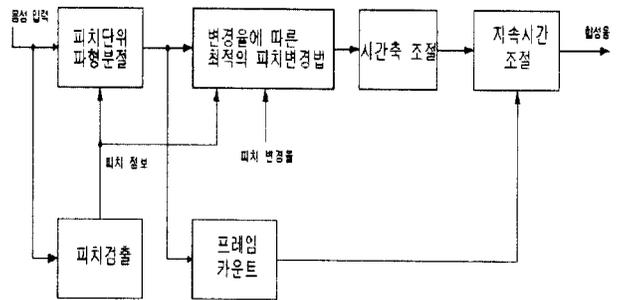


그림7-1. 본 논문에서 제안한 방법의 블록도

시뮬레이션에는 한프레임의 길이를 256표본으로 사용하였다. 먼저 변적비교법[9]을 사용하여 한 피치구간의 음성표본을 피치단위로 자른 다음에 각 피치 구간마다 음성표본을 저장한다. 그런 다음 피치 변경율에 따라 피치주기 압축시에는 영교차 단위의 시간축 조절에 의한 피치변경법을 적용하였다[12]. 또한 피치주기 신장시에는 PSOLA 합성방식에 의한 피치변경법과 진폭대칭법에 의한 피치변경법을 적용하였다[10][11].

본 논문에서는 피치주기를 50%에서 100%까지 변경시켰을 경우와 100%에서 200%까지 변경시켰을 경우에 대해서 스펙트럼 왜곡율을 측정하였다. 스펙트럼의 기준은 피치변경되기 이전의 원래 음성의 스펙트럼이다. 피치를 변경시키면 원래의 스펙트럼과 비교할 수 없기 때문에 배주기를 기준으로 피치주기를 신장시킨 다음에 스펙트럼을 다음의 리프트식에 동과시켜 스므딩시킨 다음에 원래의 음성스펙트럼의 포락과 비교를 통해 스펙트럼 왜곡율을 측정하였다.

$$S_L(k - \frac{F_0}{2}) = \sum_{l=0}^{F_0-1} S(k-l) \tag{8.1}$$

다음 표7-1과 표7-2는 피치주기를 압축한 경우와 신장한 경우에 측정된 화자들간의 평균 스펙트럼 왜곡율을 나타낸 것이다.

결과적으로 피치주기를 100%에서 50%로 압축하는 경우에 화자들간의 평균 스펙트럼 왜곡율은 5.62%이고, 피치주기를 100%에서 200%로 신장하는 경우에 평균 스펙트럼 왜곡율은 4.20%이었다.

## 피치 변경율에 따른 최적의 피치 변경법에 관한 연구

표7.1 피치압축시 평균 스펙트럼 왜곡율

변경율	피치 주기 압축율	
	80%	60%
평균	3.61	7.64

표7.2 피치신장시 평균 스펙트럼 왜곡율

변경율	피치 주기 신장율			
	120%	140%	180%	200%
평균	1.52	2.08	5.14	8.16

### VIII. 결론

일반적으로 음성합성분야에서 고음질의 파형부호화법은 파라미터들을 여기용과 성도여과기용으로 분리하지 않기 때문에 이 파형부호화법을 규칙에 의한 합성 방식으로 사용하기는 어렵다. 따라서 파형부호화법을 규칙 합성에 적용하려면 운율조절을 위해 피치변경법이 필요하다.

본 논문에서는 스펙트럼의 왜곡율을 최소화하면서도 위상을 그대로 보존할 수 있는 시간영역에서의 피치변경법을 적용하였다. 이러한 방법들은 시간영역에서만 처리됨으로써 계산량을 줄여 실시간 운율조절에 적합하다는 장점이 있다. 결과적으로 피치변경율에 따른 최적의 피치변경법을 적용하였을 때 비교적 우수한 결과를 얻을 수 있었다.

본 연구는 정보통신부의 1998년도 대학기초과제 연구지원비에 의해 이루어졌습니다.

### X. 참고문헌

[1] L.R. Rabiner & R.W. Schafer, Digital Processing of speech Signals, Prentice-Hall, 1978.  
 [2] M.R. Portnoff, "Time-Scale Modification of speech Based on Short-Time Fourier Analysis," IEEE, Trans. Acoust. Speech, Signal Processing, Vol.ASSP-29, No.3, pp.374-390, June 1981.  
 [3] M.G. Stalla and F.J. Charpentier, "Diphon Synthesis using Multipulse Coding and a Phase Vocoder", Proc. IEEE ICASSP'85, PP.74 0~744, 1985.  
 [4] A. varga and F. Fallside, "A technique for Using Multipulse Linear Predictive Speech Synthesis in Text-to-speech Type System", IEEE signal processing, Vol.ASSP-35, No.4,

pp.586-587, APRIL 1987.  
 [5] 배명진, 이미숙, 이혜군, 안수길, "퀵스트림 분석에 의한 음성 파형코딩의 피치변경에 관한 연구", 제 4 회 신호처리합동 학술대회 논문집, 제 4 권 1호, pp.304~309, 1991년 9월.  
 [6] 강동규, 김윤재, 배명진, 안수길, "음성합성의 halving기법에 의한 파형코딩의 피치 변경에 관한 연구", 한국 음향학회 추계발표회(국제 음향학회 논문집), pp.107~111, 1990년 11월 10일.  
 [7] T. Takigi and E. Miyasska, "A Speech Prosody Conversion System with a High Quality Speech Analysis-Synthesis Method", Proc.EUROSPPEECH '93, pp.995 ~ 998, September 1993.  
 [8] 배명진, "위상보상된 고조파 스케일링에 의한 음향합성용 피치변경법", 한국음향학회, 한국음향학회지, 제13권 6호, pp.995 ~ 998, September 1993.  
 [9] 배명진, 안수길, "번적 비교법을 이용한 음성신호의 고속 피치 추출", 전자공학회지, 제22권, 2호, pp.13 ~ 17, 1985년 3월.  
 [10] 손상목, 배재욱, 김용, 배명진, 기석철, 김상용, "진폭대칭법에 의한 배주기 피치변경법", 한국통신학회, 제 8 회 신호처리합동학술대회 논문집, 제 8 권, 1호, pp.708~711, 1995년 9월.  
 [11] 조왕래, 손정호, 배명진, "음성신호의 진폭대칭조절에 의한 피치신장법", 한국음향학회, 전기음향 워크샵 논문집, Vol.14, No.4S, pp.170-174, 1995년 10월  
 [12] 문효경, 임성빈, 배명진, 기석철, "영교차 단위의 시간축 조절에 의한 파형코딩의 피치변경법", 1995년도 한국음향학회 학술발표대회 논문집, Vol.14, No.1(s), pp 211-214, Nov.11, 1995