LPC 켑스트럼 거리 기반의 천이구간 정보를 이용한 음성의 가변적인 시간축 변환

Variable Time-Scale Modification of Speech Using Transient Information based on LPC Cepstral Distance

이 성 주* · 김 희 동** · 김 형 순*
(Sungjoo Lee · Hee Dong Kim · Hyung Soon Kim)

ABSTRACT

Conventional time-scale modification methods have the problem that as the modification rate gets higher the time-scale modified speech signal becomes less intelligible, because they ignore the effect of articulation rate on speech characteristics. Results of research on speech perception show that the timing information of transient portions of a speech signal plays an important role in discriminating among different speech sounds. Inspired by this fact, we propose a novel scheme for modifying the time-scale of speech. In the proposed scheme, the timing information of the transient portions of speech is preserved, while the steady portions of speech are compressed or expanded somewhat excessively for maintaining overall time-scale change. In order to identify the transient and steady portions of a speech signal, we employ a simple method using LPC cepstral distance between neighboring frames. The result of the subjective preference test indicates that the proposed method produces performance superior to that of the conventional SOLA method, especially for very fast playback case.

Keywords : time scale modification, transient information

I. 서 론

음성 신호의 시간축 변환(time-scale modification)이란 기본 주파수 및 스펙트럼 형태 등 원래의 신호 특성은 그대로 유지하면서 발음속도만 빠르게 또는 느리게 변환시키는 것이다. 음성의 시간축 변환은 음성부호화 전단계로서 음성신호를 압축하거나, 디지탈 전화 응답 장치에서 메세지 재생속도를 바꾸어 듣거나, 또는 노약자 및 언어 교육을 위한 특수장치 등을 비롯하여 다양한 응용분야를 가지고 있다. 이러한 시간축 변환 방법에는 단구간 Fourier해석에 의한 방법[1], 정현

* 부산대학교 전자공학과
** 한국외국어대학교 정보통신공학과

파 모델에 의한 방법[2][3], synchronized overlap and add (SOLA)방법[4][10], pitch synchronized overlap and add (PSOLA)[5] 방법 등이 있다. 이들 시간축 변환 방법 중 SOLA 방법은 비교적 적은 계산량으로도 우수한 성능을 나타내기 때문에 널리 사용되고 있다. 이 방법은 음성 신호의 시간축 변환 위해 단구간 신호들을 중첩 가산(overlap and add)해서 더하기 전에 상호상관함수를 이용하여 단구간 신호들의 동기를 맞추는 방법을 사용한다. 이러한 SOLA 방법은 실시간 처리가 가능하면서도 비교적 고음질의 변환 신호를 얻을 수 있지만, 시간축 변환의 비가 커짐에 따라 합성음의 명료도가 떨어지는 문제점이 있다. 그 이유 중 하나는 사람이 발음속도를 바꾸어 발음할 때의 음성학적 특성을 고려하지 않고 전체 음성신호를 일률적으로 변환시키는 방법을 사용하기 때문이다. 이 문제의 개선을 위하여 유성음과 무성음의 구별에 의한 가변적인 시간축 변환 방식이 제안된 바 있으며, 일률적인 SOLA 방식에 비해서 개선된 음질을 얻었다고 보고되었다[6].

음성 신호는 일반적으로 정적인(steady) 구간과 천이(transient) 구간으로 나눌 수 있으며, 음성인지(speech perception)에 대한 연구에 따르면 천이 구간에서의 시간 정보가 음성의 인지에 매우 중요한 역할을 한다고 알려져 있다[7][8]. 본 논문에서는 이러한 지식을 고려하여 천이 구간의 시간적 정보를 보존하도록 하는 가변적인 시간축 변환 방법을 제안한다. 이를 위하여 먼저 음성 신호를 천이 구간과 정적인 구간으로 구분하여야 하며, 본 논문에서는 음성 신호로부터 추출한 LPC 켑스트럼 계수를 이용하여 구한 인접 프레임들간의 스펙트럼 거리를 특정 임계치와 비교함으로써 이들 구간들을 구분하였다. 제안된 방식의 성능을 평가하기 위하여 일률적인 SOLA 방법으로 합성된 신호와 함께 청취자 선호도 조사를 실시함으로써 제안된 방식이 우수함을 확인하였다.

본 논문의 구성은 아래와 같다. 서론에 이어 2장에서는 기존의 시간축 변환 방법인 SOLA 방법을 살펴보고, 3장에서는 본 논문에서 제안하는 방법에 대해 설명한다. 그리고 4장에서는 제안된 알고리즘을 적용하여 합성한 음성 신호와 기존의 방법들로 합성된 음성 신호에 대해 선호도 평가를 통해 성능을 비교하였으며, 마지막으로 5장에서 결론을 맺는다.

II. Synchronized OverLap and Add (SOLA) 방법 [4][10]

Synchronized overlap and add (SOLA)방법은 합성하고자 하는 프레임을 앞서 합성된 신호와의 상호상관함수가 최대가 되는 위치에 평균 중첩가산하는 것이다. 단순히 합성하고자하는 프레임을 합성구간만큼 이동한 후 중첩가산하는 방법도 시간축 변환의 한 방법이 될 수는 있지만, 피치주기나 위상정보를 보존하지 못하는 문제점이 있어서 합성음질이 크게 떨어진다.

이에 반하여 최대 상호상관함수의 값을 갖는 위치에 동기를 맞추어 프레임들을 중첩가산하는 SOLA 방법은 피치와 위상정보를 잘 보존하게 되어 비교적 우수한 음질의 소리를 합성한다. 음성신호 $x(n)$을 변환비율 α에 의해 변환시켜 $y(n)$이라는 합성신호를 얻는다고 하자. 이 때, α가 1보다 크다면 시간축으로 신장된 합성신호를, α가 1보다 작다면 시간축으로 압축된 합성신호를 각각 얻게 된다. 음성신호 $x(n)$에서 매 분석구간 S_a마다 N개의 샘플로 구성된 프레임들을 가지고 매 S_s마다 합성신호 $y(n)$를 합성하는데 사용한다면, 분석구간 S_a와 합성구간 S_s

사이에는 $S_s=\alpha S_a$라는 관계가 성립한다. 합성 과정은 매 프레임당 이루어 지고 새로운 분석 프레임이 앞의 과정에서 합성된 프레임에 더해 진다. 알고리즘은 $y(j)=x(j),\ 0\le j\le N-1$, 즉 첫번째 분석 프레임을 합성 프레임으로 초기화하는 과정과 합성신호, $y(mS_s+j)$에 음성신호의 m번째 프레임, $x(mS_a+j)$의 동기를 맞추어 재배열한 후 중첩 가산하는 과정으로 구성된다. 이 때 동기가 일치하는 위치는 다음과 같이 $y(mS_s+j)$과 $x(mS_a+j)$ 사이의 정규화된 상호상관함수를 통하여 얻어 진다.

$$R_m(k)=\frac{\sum_{j=0}^{L-1} y(mS_s+k+j)x(mS_a+j)}{[\sum_{j=0}^{L-1} y(mS_s+k+j)\sum_{j=0}^{L-1} x(mS_a+j)]^{1/2}},\quad -\frac{N}{2}\le k\le \frac{N}{2} \tag{1}$$

여기서, $R_m(k)$은 m번째 프레임에서 정규화된 상호상관함수를 나타내고, L은 $y(mS_s+j)$와 $x(mS_a+j)$사이에 중첩되는 샘플 수이다. k_m이 $R_m(k)$를 최대화시키는 위치라고 한다면, $x(mS_a+j)$은 $y(mS_s+k_m+j)$에 식 (2)와 같이 재배열되어 중첩가산된다.

$$y(mS_s+k_m+j)=(1-f(j))\,y(mS_s+K_m+j)+f(j)x(mS_a+j),\quad 0\le j\le L_m-1 \tag{2}$$
$$y(mS_s+k_m+j)=x(mS_a+j),\quad L_m\le j\le N-1$$

여기서, L_m은 두 신호의 중첩가산구간을 나타내고 $f(j)$는 weighting 함수로 $0\le f(j)\le 1$의 값을 갖는다.

정규화된 상호상관함수는 중첩되는 범위가 잘못된 동기로 두 신호를 중첩가산하게 만들어 음질에 악영향을 미칠 수 있다. 이러한 오류를 막기 위하여 앞서 설명한 L이 프레임 길이의 1/8 이상이 되는 범위에서는 정규화된 상호상관함수의 최대값을 찾도록 한다.

이러한 SOLA 방법은 적은 계산량으로 비교적 우수한 음질을 얻을 수 있는 장점이 있지만, 음성신호를 일률적으로 압축시키거나 신장시키기 때문에 큰 비율로 신호를 변환할 경우 사람이 듣기에 불명료한 소리를 합성하는 문제점을 지닌다.

III. 천이 구간 정보를 이용한 음성의 가변적인 시간축 변환

음성신호의 시간축 변환의 목표는 음성신호의 기본 주파수와 성도 모델 스펙트럼을 보존하여 원래의 신호 특성은 그대로 유지하면서 발음 속도만 변화시키는 것이다. 기존의 시간축 변환 방법 중에서 적은 계산량으로 우수한 성능을 내는 SOLA 방법은 음성 신호를 일률적으로 압축하거나 신장시키기 때문에, 큰 비율로 신호를 변환할 경우 합성음의 명료도가 떨어지는 단점이 있다.

음성신호는 정적인 구간과 천이 구간으로 나눌 수 있으며, 음성인지(speech perception)에 관한 연구 결과에 따르면 천이 구간의 시간적 정보가 소리를 인지하는데 매우 중요하다고 알려져 있다[7]-[9]. 자음-모음-자음으로 이루어진 음절의 처음 부분과 마지막 부분을 절단해 가면서 변화된 음절의 인지도를 시험하는 과정에서 Furui는 청취자가 인지할 수 있는 임계점은 절단된 음

절을 올바르게 인지할 수 있는 절단점 위치의 함수인 경계점으로서 최대 스펙트럼 천이 위치와 관련이 있다는 사실을 발견하였다[7]. 즉, 최대 스펙트럼 천이 위치를 포함하는 대략 10ms 길이의 음성 신호는 음절의 인식에 필요한 중요한 정보를 가지고 있다는 것이다. 또한, 스펙트럼이 얼마나 급격히 변화하는가 하는 정도가 서로 다른 부류의 음성을 분별하기 위한 중요한 특성이라는 사실도 보고된 바 있다[8]. 이 연구에 따르면 자음을 모음으로부터 분별하는 특징은 어느 특정한 시간에서의 스펙트럼의 모양이 아니라 얼마나 급격하게 스펙트럼이 변화하는가에 달려 있다는 것이다. 이와 같이 스펙트럼이 급격히 변화하는 천이 구간은 음성 정보가 많이 함축되어 있으므로, 발음 속도가 바뀌더라도 천이 구간의 시간적 정보를 보존하는 것이 명료한 청취에 도움을 주게 된다.

본 논문에서는 이와 같은 천이 정보를 이용하여 가변적인 시간축 변환을 하는 방식을 제안한다. 이를 위하여 먼저 음성 신호를 천이 구간과 정적인 구간으로 구분하여야 하는데, 본 논문에서는 음성 신호로부터 추출한 LPC 켑스트럼 계수[11]를 이용하여 구한 인접 프레임들 간의 스펙트럼 거리를 특정 임계치와 비교함으로써 이들 구간들을 구분하였다.

음성신호는 매 10 ms마다 20 ms씩 중첩이 되도록 30 ms 크기의 프레임 단위로 LPC 켑스트럼을 구한다, 이 때 i번째 프레임에 해당하는 인접 프레임들간의 스펙트럼 거리 $D(i)$는 다음과 같은 Euclidean 거리를 사용하였다.

$$D(i) = \sum_{k=1}^{p} [c_{i-2}(k) - c_{i+2}(k)]^2 \qquad (3)$$

여기서 $c_i(k)$, $k=1, \cdots, p$는 i번째 프레임의 LPC 켑스트럼 계수를 나타낸다. 식(3)의 $D(i)$ 값이 실험적으로 구한 임계값 D_{th}보다 크면 천이구간으로, 그보다 작으면 정적인 구간으로 결정한다.

그림 1. 음성신호로 부터 천이 구간과 정적인 구간을 구분하는 과정의 예 (a) 음성신호 "음, 저 총무부의 운영관리실 번호가 몇 번 이지요" (b) 인접 프레임들간의 스펙트럼 거리와 임계치 설정 (c) 천이 구간 (1)과 정적인 구간(0)의 구분

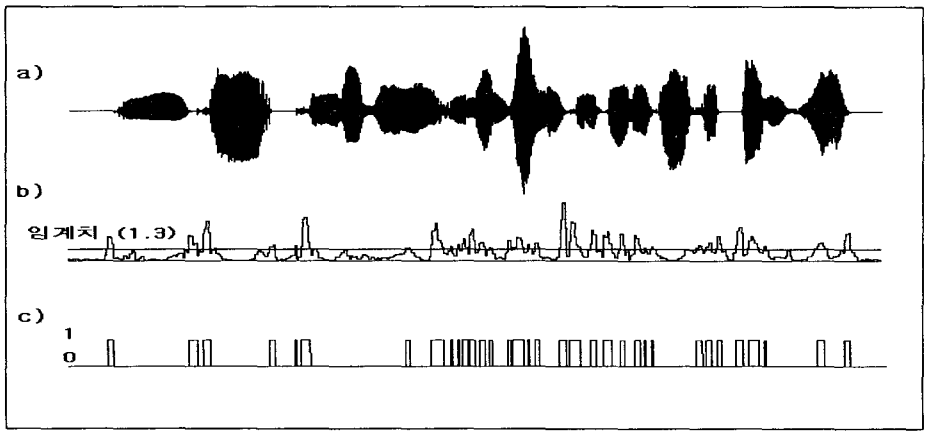


그림 1에 스펙트럼 거리 $D(i)$를 이용하여 음성 신호를 천이 구간 및 정적인 구간으로 구분하는 과정을 나타내었다. 그림 1의 (a)는 음성신호를 나타내고, (b)는 인접 프레임들 간의 LPC 켑스트럼 계수를 분석하여 얻은 스펙트럼 거리와 임계치를 나타낸다. 그리고 (c)에서 1은 천이 구간을, 0은 정적인 구간을 각각 나타내고 있다. 본 논문에서는 천이 구간이 전체 음성 신호의 구간들의 20% 이상을 차지하도록 임계치 D_{th}를 1.3 으로 결정하였다.

본 논문에서는 이상과 같이 음성신호를 정적인 구간과 천이구간으로 구분한 다음, 천이구간의 시간축 정보를 보존하기 위해 정적인 구간에 대해서만 시간축 변환을 수행하는 방법을 사용하였다. 전체 프레임 수를 T라 하고, 천이구간의 프레임 수를 T_t, 정적인 구간의 프레임 수를 T_s라 한다면 T는 다음과 같이 표현될 수 있다.

$$T = T_t + T_s \tag{4}$$

그러므로 변환 비율 α로 일률적으로 시간축 변환된 신호와 제안된 방법으로 변환된 신호는 다음과 같은 관계가 있다.

$$\alpha T = T_t + \alpha_s T_s \tag{5}$$

여기서, α_s는 정적인 구간의 변환 비율을 나타낸다. 위의 식 (5)에 식 (4)를 대입하여 α_s에 대해 정리하면,

$$\alpha_s = \frac{(\alpha-1)T + T_s}{T_s} \tag{6}$$

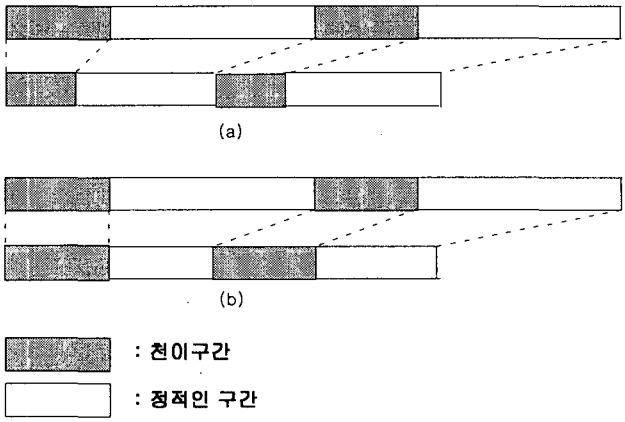
와 같고, $\beta = T_s / T$라 두면 위의 식 (6)은 다음과 같이 정리된다.

$$\alpha_s = \frac{(\alpha-1)+\beta}{\beta} \tag{7}$$

여기서, β는 전체 프레임 수에서 정적인 구간의 프레임 수가 차지하는 비율을 의미한다.

그림 2는 기존의 SOLA 방법과 제안된 방법으로 시간축 변환을 수행할 경우 변환비율, α=0.7에 의해 시간적으로 압축된 결과신호를 각각 나타내고 있다. 그림 2(a)에서 보는 바와 같이 기존의 SOLA 방법으로 시간축 변환 할 경우 천이구간과 정적인 구간이 모두 변환비율, α에 의해 시간축으로 압축되는 반면, 그림 2(b)에서 보는 바와 같이 제안된 방법에서는 천이구간은 그대로 유지하면서 정적인 구간에 대해서만 전체 프레임 수에서 정적인 구간의 프레임 수가 차지하는 비율, β에 따른 가변적인 시간축 변환 비율, α_s에 의해 시간축 변환되어 전체적으로는 목표하는 시간축 변환 속도를 만족시키게 된다.

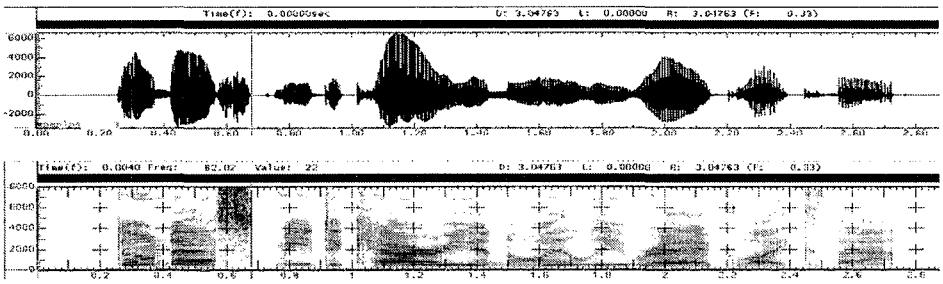
그림 2. 변환비율이 α = 0.7일때, 기존의 SOLA 방법과 제안된 방법의 시간축 변환 예 (a) 원신호와 기존의 SOLA 방법에 의해 시간축 변환된 신호 (b) 원신호와 제안된 방법에 의해 시간축 변환된 신호



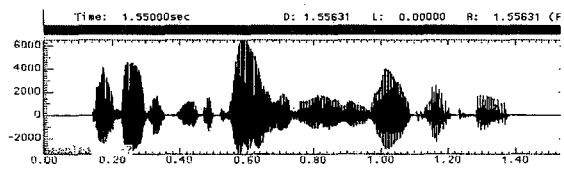
그러나, 이 방법을 그대로 사용하기 위해서는 전체 음성신호에 대해 미리 정적인 구간 및 천이 구간의 구분작업을 수행하여 이들 구간들의 비율을 알아내야 하는데, 이는 변환하고자 하는 음성신호의 길이가 매우 길 경우, 실시간 신호처리 및 메모리의 효율적인 사용 등의 측면에서 바람직하지 못하다. 따라서, 본 논문에서는 일정기간 단위(예를 들면, 1초)마다 정적인 구간 및 천이구간 프레임들을 구분한 다음, 그 기간 내에서의 비율에 의거하여 음성의 시간축 변환을 수행하는 방법을 사용함으로써 이러한 문제를 해결하였다.

그림 3은 일률적인 SOLA 방법과 LPC 켑스트럼 거리를 이용하는 제안된 방법에 의해 시간축으로 압축된 신호와 원신호를 스펙트로그램 상에서 비교한 것이다. 이 경우 시간축 변환 비율, α는 0.5로서 음성신호를 시간적으로 2배 빠르게 압축하는 것을 의미한다. 그림 3에서 보는 바와 같이 일률적인 SOLA 방법이 원신호의 포만트 천이보다 급격한 포만트 천이를 보이는 반면에(그림3(b)), 제안된 방법은 원신호의 포만트 천이특성을 비교적 잘 보존하는 것을 볼 수 있다(그림 3(c)).

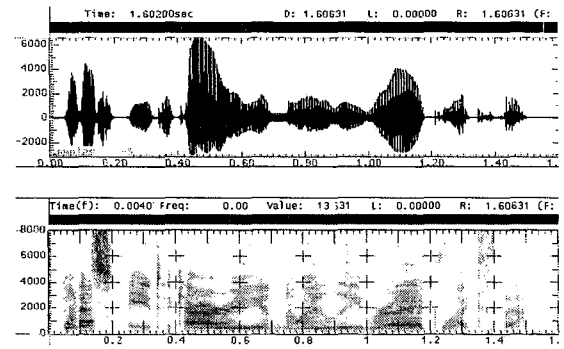
그림 3. α = 0.5로 시간축 변환된 신호의 파형과 스펙트로그램 (a)원신호의 파형과 스펙트로그램 (b)일률적인 SOLA 방법에 의해 시간축 변환된 신호의 파형과 스펙트로그램 (c)LPC 켑스트럼 거리를 이용하는 제안된 방법으로 시간축 변환된 신호의 파형과 스펙트로그램



(a)



(b)



(c)

IV. 실험 및 고찰

본 실험에서는 천이 구간 정보에 따른 가변적인 시간축 변환 알고리즘의 성능을 평가하기 위해 컴퓨터로 모의 실험을 시행하였다. 각 구간에서 천이 구간 여부를 구분하기 위해서 LPC 켑스트럼 계수를 이용하여 인접 프레임 사이의 스펙트럼 거리를 구한 다음, 이 스펙트럼 거리가 특정 임계치(본 논문에서는 임계치를 1.3으로 정하였음)를 넘으면 천이 구간으로, 그렇지 않으면 정적인 구간으로 각각 구분하였다. 그리고 이 정보를 이용하여 천이구간은 그대로 유지하고 정적인 구간에 대해서만 시간축 변환을 수행하여 그 결과를 2장에서 설명한 기존의 시간축 변환 방법 즉, 일률적인 SOLA 방법에 의한 결과들과 비교하여 청취자 선호도 조사를 실시하였다. 모의 실험에 사용된 음성 데이타는 한국전자통신연구원의 부서명 음성 데이터베이스[12]의 일부로서 성인 남성이 비교적 조용한 환경에서 보통 대화 속도로 녹음한 것이다. 실험에 사용된 문장들은 다음과 같다.

(1) 거기 에트리지요? 설비과를 연결해주세요.
(2) 음, 저 총무부의 운영 관리실 번호가 몇 번이지요.
(3) 수고하십니다, 회계과 좀 부탁할까요.
(4) 자산 관리과 좀 바꿔 주시겠어요.
(5) 한국전자통신연구소지요? 근로복지실 전화번호가 어떻게 되나요.

음성 데이타는 16kHz로 샘플링된 16비트 데이타를 8kHz로 down-sampling하여 사용하였고, 음성 신호의 분석은 30ms의 길이에 해당하는 240샘플 단위로 분석 구간을 잡고, 분석시 이동값 S_a 는 10ms의 길이에 해당하는 80샘플로 정하였다. 스펙트럼 거리를 구하기 위해서 $1-0.95\ z^{-1}$ 형태의 전처리(preemphasis) 과정을 거친 다음 인접 프레임들의 LPC 켑스트럼 계수를 구하고, 식 (3)과 같이 이들 계수들의 차를 제곱하여 누적한 것을 스펙트럼 거리로 사용하였다. 이 때 LPC 켑스트럼 계수의 차수는 10차로 정하였다.

기존의 시간축 변환 방법과 성능을 비교하기 위해서 기존의 SOLA 방법을 함께 구현하고 위의 5가지 음성 데이타를 각각 0.5, 0.7, 1.3, 1.5, 및 1.8 배의 5가지 속도로 변환시키면서 청취자 선호도 조사를 실시하여 그 성능을 평가하였다. 청취자 선호도 조사는 어떤 방법이 사람이 듣기에 더 명료한 소리를 합성하는가에 기준을 두고 청취자 자신이 어떤 음성 데이타가 어떠한 방법으로 변환된 신호인지를 알 수 없도록 음성 데이타의 순서를 랜덤하게 주어 실시하였다. 우선, 청취자의 선호도를 조사하기 전에 명료함의 기준을 위해 사람이 보통 속도로 자연스럽게 발음하는 것을 들려주었다. 각 청취자가 두 가지 다른 방식으로 합성된 음성 데이타를 집중하여 듣는데 도움을 주기 위해 헤드폰을 사용하였고, 다른 조사 대상자의 영향을 배제하기 위하여 1명씩 개별적으로 조사하였다. 청취자가 두 가지 방법에 의한 합성음 중에서 어느 쪽이 더 명료한지 판단할 수 없을 경우에는 판단을 유보하도록 허용하였다. 실험대상은 만 23-31세 사이의 부산 대학교 대학원생 20명(남성 17명과 여성 3명)으로 구성하였다. 표 1은 이러한 청취자 선호도 조사의 결과이며, 표에서 '구별안됨'항은 어느 쪽이 더 명료한지 알 수 없다고 판단한 경우를 의미한다.

표 1. 청취자 선호도 조사 결과

변 환 비 율	선 호 도		
	제안된 방법	구별안됨	일률적인 SOLA방법
0.5 배	85 %	4%	11%
0.7 배	58 %	19%	23%
1.3 배	51 %	29%	21%
1.5 배	51 %	27%	22%
1.8 배	53 %	24%	23%
평 균	59.6 %	20.4%	20.0%

표에서 보는 바와 같이 본 논문에서 제안된 방법이 모든 변환 속도에 대해서 기존의 방법에 비해 명료한 소리를 합성하고 있음을 확인할 수 있었다. 특히 변환비율을 0.5배로 빠르게 할 경우, 기존의 SOLA 방법에 비해 훨씬 명료한 소리를 합성할 수 있었다.

V. 결 론

본 논문에서는 음성의 천이 구간의 시간적 정보가 음성 인지에 중요한 역할을 한다는 지식에 기반을 둔 음성의 가변적인 시간축 변환 방법을 제안하였다. 이를 위하여 천이 구간과 정적인 구

간을 먼저 구분하고 천이 구간의 시간축 정보는 그대로 유지하면서 정적인 구간만을 시간축으로 변환시키도록 SOLA 방법을 가변적으로 적용하였다. 제안된 방법의 성능을 평가하기 위하여 기존의 일률적인 SOLA 방법을 사용한 결과와 청취자 선호도를 조사하여 결과를 비교하였다. 실험 결과 제안된 방법이 모든 변환속도에 대해 기존의 방법들에 비해 성능이 우수함을 확인하였다. 특히 변환비율을 0.5배로 빠르게 할 경우, 기존의 SOLA 방법에 비해 훨씬 명료한 소리를 합성할 수 있었다.

실시간 구현의 관점에서 볼 때, 본 논문에서 제안한 방법은 SOLA 방법에 비해 LPC 켑스트럼 계수를 추출하는 과정과 스펙트럼 거리를 구하는 과정이 추가되므로 계산량이 많아지는 단점을 지닌다. 그러나 디지털 전화응답장치와 같이 중저 전송속도의 음성 압축 방식이 함께 사용될 경우에는 음성 압축/재생과정에서 LPC 계수가 이미 추출되므로 계산량 추가가 문제 되지 않는다[13]. 현재 제안된 방식의 계산량을 더욱 줄이는 방법과 음질을 보다 개선하는 방법에 대해 연구가 진행되고 있다.

참 고 문 헌

[1] M. R. Portnoff. 1981. "Time-scale modification of speech based on short-time Fourier analysis." IEEE Trans. *Acoustic., Speech, Signal Processing*, ASSP-29(3), 374-390.

[2] T. F. Quateri and R. J. McAulay. 1992. "Shape invariant time-scale and pitch modification of speech," IEEE Trans. *Signal Processing*, 40(3), 497-510.

[3] T. F. Quateri and R. J. McAuay. 1986. "Speech transformation based on a sinusoidal representation," IEEE Trans. Acoustic. *Speech, Signal Processing*, ASSP-41(6), 1449-1464.

[4] S. Roucos and A. M. Wilgud. 1986. "High quality time-scale modification for speech," in Proc. ICASSP, 493 - 496.

[5] E. Moullines and F. Charpentier. 1990. "Pitch synchronous waveform processing for text-to-speech synthesis using diphones," *Speech Communication,* 9(5/6), 453-467.

[6] 손단영 · 김원구 · 윤대희 · 차일환. 1995. "유/무성음 결정에 따른 가변적인 시간축 변환," *대한전자공학회 논문지*, 32B(5), 788-797.

[7] S. Furui. 1979. "On the role of spectral transition for speech perception," *J. Acoust. Soc. Amer.*, 80, 1016-1025.

[8] K. N. Stevens. 1980. "Acoustic correlates of some phonetic categories," *J. Acoust. Soc. Amer.*, 68, 836-842.

[9] H. S. Kim. 1989. "A study on the use of perceptual information for speech recognition," Ph.D. Dissertation, KAIST.

[10] J. Makhoul and A El-jaroudi. 1986. "Time-scale modification in medium to low rate speech coding," in *Proc ICASSP*, 1705-1708.

[11] L. Rabiner and B. H. Juang. 1993. *Fundamentals of Speech Recognition*, Prentice-Hall, 100-117.

[12] 이영직 · 류준영 · 김상훈 · 황규웅. 1995. "ETRI의 음성 데이터 베이스 구축 현황," *제 12회 음성통신 및 신호처리 워크샵 논문집*, 256-267.

[13] A. M. Kondoz. 1994. *Digital Speech : Coding for Low Bit Rate Communication Systems*, Wiley.

접수일자 : '98. 2. 15.
게재결정 : '98. 3. 21.

▲ 이성주
부산시 금정구 장전동 산 30번지
부산대학교 전자공학과(우 : 609-390)
Tel : (051) 516-4279
e-mail : lsj@hyowon.cc.pusan.ac.kr

▲ 김희동
경기도 용인시 모현면 왕산리 산 89
한국외국어대학교 정보통신공학과(우 : 449-850)
Tel : (0335) 30-4254
e-mail : kimhd@ice.hufs.ac.kr

▲ 김형순
부산광역시 금정구 장전동 산30
부산대학교 공과대학 전자공학과(우 : 609-935)
Tel : (051) 510-2452 Fax : (051) 515-5190
e-mail : kimhs@hyowon.pusan.ac.kr