한국음향학회지 제31권 제7호 pp. 427~434 (2012)
The Journal of the Acoustical Society of Korea Vol.31, No.7 (2012)
http://dx.doi.org/10.7776/ASK.2012.31.7.427

WSOLA 기반의 음성 시간축 변환을 위한 고속의 정규상호상관도 계산

A Fast Normalized Cross-Correlation Computation for WSOLA-based Speech Time-Scale Modification

임상준, 김형순

(Sangjun Lim and Hyung Soon Kim)

부산대학교 전자전기공학과

(접수일자: 2011년 12월 28일; 채택일자: 2012년 8월 4일)

초 록: WSOLA 방식은 음성 신호의 시간축 변환을 위한 고음질의 효율적인 알고리즘으로 알려져 있다. WSOLA의 계산량은 두 신호 파형 사이의 유사도를 평가하는 반복적인 정규상호상관도 계산에 집중되어 있다. 본 논문은 WSOLA 계산량 감축을 위해 고속의 정규상호상관도 계산 방법을 제안하며, 제안된 방법에서는 미리 계산된 합 테이블을 통해 인접한 구간에서의 반복적인 정규상호상관도 계산의 중복성을 제거한다. 정규상호상관도의 분모 부분은 시간축 변환 비율에 관계없이 높은 중복성을 가지는데 반해, 분자 부분은 보다 낮은 중복성을 가지며 중복 정도가 시간축 변환 비율과 최적 이동값에 의해 영향을 받기 때문에 고속 계산을 위해 보다 복잡한 알고리즘이 요구된다. 시뮬레이션 결과, 제안된 방법이 기존의 WSOLA와 완전히 동일한 음질을 유지하면서도 시간축 압축의 경우 약 40%, 그리고 1/2배속 및 1/3배속으로의 시간축 신장의 경우 각각 약 47% 및 52%의 실행시간을 감소시킴을 보인다.
핵심용어: 시간축 변환, WSOLA, 고속의 정규상호상관도 계산
투고분야: 음성처리 분야(2.4)

ABSTRACT: The overlap-add technique based on waveform similarity (WSOLA) method is known to be an efficient high-quality algorithm for time scaling of speech signal. The computational load of WSOLA is concentrated on the repeated normalized cross-correlation (NCC) calculation to evaluate the similarity between two signal waveforms. To reduce the computational complexity of WSOLA, this paper proposes a fast NCC computation method, in which NCC is obtained through pre-calculated sum tables to eliminate redundancy of repeated NCC calculations in the adjacent regions. While the denominator part of NCC has much redundancy irrespective of the time-scale factor, the numerator part of NCC has less redundancy and the amount of redundancy is dependent on both the time-scale factor and optimal shift value, thereby requiring more sophisticated algorithm for fast computation. The simulation results show that the proposed method reduces about 40%, 47% and 52% of the WSOLA execution time for the time-scale compression, 2 and 3 times time-scale expansions, respectively, while maintaining exactly the same speech quality of the conventional WSOLA.
Key words: Time-scale modification, WSOLA, Fast normalized cross-correlation computation
ASK subject classification: Speech Signal Processing (2.4)

I. 서 론

시간축 변환(time-scale modification)은 음성 신호의 음높이, 음색 등의 주요한 특징들은 그대로 유지하면서 발화 속도만을 빠르게 또는 느리게 변경하는 기술을 의미하며, 음성부호화 전처리, 어학 교육, 노래방 기기 등 여러 응용에 사용된다. 발화 속도 변경을 위해 음성 신호의 전체적인 시간축을 단순히 줄이거나 늘이면 스펙트럼 정보가 왜곡되어 마치 변조된 음성처럼 들리게 된다. 시간축 변환 기술은 스펙

***Corresponding author:** Hyung Soon Kim (kimhs@pusan.ac.kr)
Department of Electronics Eng., Pusan National University, Busan, 609-735, Republic of Korea
(Tel: 82-51-510-2452)

트럼 정보의 왜곡을 최소화 하면서 발화 속도만을 변경시키고자 하는 것으로서, 크게 시간 영역 접근법과 주파수 영역 접근법으로 나누어진다. 그 중에서 시간 영역 접근법은 음성과 같이 단일 기본주파수를 가지는 신호에 대해 신호의 주기적인 특성을 활용함으로써, 주파수 영역 접근법에 비해 상대적으로 적은 계산량으로도 우수한 성능을 나타낸다.

시간 영역에서의 시간축 변환 기술의 대표적인 예로는 Synchronized OverLap and Add(SOLA),[1] OverLap-Add technique based on Waveform Similarity(WSOLA),[2] Pitch Synchronized OverLap and Add(PSOLA)[3] 등이 있다. 그 중에서 SOLA와 WSOLA는 서로 유사한 방법으로 두 가지 모두 고음질을 나타내지만, 매 프레임의 중첩 구간이 고정되어 있지 않은 SOLA에 비해 출력 신호의 중첩 구간이 50%로 일정한 WSOLA가 계산량 면에서 좀더 유리하다. PSOLA는 피치 주기에 따라 적응적인 윈도우 크기를 이용함으로써 음질은 매우 뛰어나지만, 피치 구간 추정의 정확도에 따라 성능이 좌우되며 정확한 피치 추정을 위해 계산량이 증가되기 때문에 실시간 처리에는 잘 사용되지 않는다. 그 대신 피치 구간 추정을 오프라인 환경에서 처리할 수 있는 음성합성기의 발화속도 및 음정 조정에 주로 사용된다. 따라서 음성의 실시간 시간축 변환에는 WSOLA를 쓰는 것이 가장 적합하다.

WSOLA가 다른 시간축 변환 방식에 비해서는 계산량 면에서 매우 효율적이지만, 각종 모바일 기기와 같이 상대적으로 계산능력이 떨어지는 장치에서 WSOLA를 적용하기 위해서는 추가적인 계산량 감축이 도움이 된다. WSOLA는 정규상호상관도를 이용하여 합성 지점을 탐색하는 과정을 거쳐야 하는데 이 과정이 전체 계산의 90% 이상을 차지한다. 따라서 이 과정의 계산량을 줄이는 것이 전체 계산량 감축의 관건이 된다. 이와 관련하여 신호주기 추정을 이용하는 WSOLA 계산량 감축 방식이 제안되었는데,[4] 피치 정보를 이용하여 정규상호상관도의 계산 범위를 줄여서 계산량은 줄어들지만 음질이 다소 떨어지는 단점이 있다.

본 연구에서는 고속의 정규상호상관도 계산 방식을 적용하여 본래의 WSOLA의 고음질 결과를 완전히 동일하게 유지하면서 계산량을 감소시키는 방법을 제안한다.[5,6] 기본 아이디어는 근접 구간에서의 정규상호상관도 계산에 중복연산이 있기 때문에 이를 제거하여 계산량을 감소시키는 것이다. 정규상호상관도 분모의 에너지 계산은 근접 구간에서 배속에 관계없이 항상 계산량을 감소시킬 수 있고, 분자는 저배속에서만 중복이 생기기 때문에, 고배속일 경우 분모의 계산량만 줄이고 저배속의 경우 분모와 분자의 계산량을 모두 감축시킨다. 결과를 확인하기 위해 1분 크기의 문장 발화의 WSOLA 실행 시간을 측정하였으며, 합성결과 파형은 완전히 동일하면서도 실행시간을 감소시킴을 확인하였다.

본 논문의 구성은 다음과 같다. 서론에 이어 II장에서는 기존의 WSOLA에 대해서 설명하고 III장에서는 고속의 정규상호상관도 계산 방식을 제안한다. IV장에서는 제안된 방식들의 실험 결과를 보여주고 마지막으로 V장에서 결론을 맺는다.

II. WSOLA

중첩가산(OverLap and Add, OLA) 방법을 따르는 대부분의 시간축 변환 방법은 원 신호와 시간축 변환된 신호에서 단시간 푸리에 변환 결과의 거리 함수를 최소화 시키는데 기반을 둔다.[7] 원 신호의 길이를 줄이거나 늘이기 위해서 등간격으로 일정 구간을 중첩 가산하는 간단한 해결책은 원 신호의 위상 관계를 파괴시킨다. 따라서 신호의 유사성을 최대화하여 연속성을 보장하는 방식으로 구한 최적 이동값 $\triangle_k$에 의해 출력 신호의 윈도우 위치를 이동시키는 것이 WSOLA의 개념이다. 참고로 연속성을 추구하는 것에서는 SOLA와 동일하지만 출력 신호와의 동기화를 최대화하고 WSOLA는 입력 신호와의 유사도를 최대화 하는 것이 차이점이라고 할 수 있다.

입력 신호 $x(n)$에 대해서 출력 신호 $y(n)$는 다음 식과 같이 주어진다.

$$y(n)=\frac{\sum_k \nu(n-kS)x\left(n+\tau^{-1}(kS)-kS+\triangle_k\right)}{\sum_k \nu(n-kS)} \quad (1)$$

여기서 S는 중첩 길이에 해당하는 값이고 윈도우는

항상 절반 중첩이기 때문에 윈도우 길이는 $N=2S$이다. 그리고 k는 합성 프레임의 인덱스를 나타내므로 $\triangle_k$는 k번째 프레임에서의 최적 이동값을 나타낸다. 시간 매핑 함수 $\tau^{-1}(kS)$는 균일한 시간축 변환에서 kS에 상수가 곱해진 형태로 나타낼 수 있고, $\tau^{-1}(kS)=kS\alpha$로 표현할 때 α는 시간축 변환 비율이 되어 1을 기준으로 그보다 클 경우 빠른 속도로의 변환이고 1보다 작을 경우 느린 속도로의 변환이 된다.[8] 윈도우 함수 $\nu(n)$는 중첩 가산 시에 $\sum_k \nu(n-kS)=1$의 조건을 만족시키는 Hanning 윈도우를 사용하는 경우가 일반적이므로 다음과 같이 간략화된다.

$$y(n)=\sum_k \nu(n-kS)x(n+kS\alpha-kS+\triangle_k) \tag{2}$$

최적의 $\triangle_k$를 구하기 위해서 두 신호를 평가하는 기준으로 정규상호상관도를 사용하며 그 식은 다음과 같다.

$$NCC(k,\triangle)=\frac{\sum_{n=0}^{2S-1} R_k(n)C_k(n+\triangle)}{\left(\sum_{n=0}^{2S-1} R_k^2(n)\right)^{1/2}\left(\sum_{n=0}^{2S-1} C_k^2(n+\triangle)\right)^{1/2}} \tag{3}$$

여기서 식(3)에서 비교하는 두 신호는 식(4)와 같이 정의할 수 있다.

$$\begin{aligned} R_k(n)&=x(n+(k-1)S\alpha+\triangle_{k-1}),\ \ 0\le n\le 2S-1 \\ C_k(n+\triangle)&=x(n+kS\alpha-S+\triangle),\ \ 0\le n\le 2S-1 \end{aligned} \tag{4}$$

식(4)에서 $R_k(n)$는 k번째 프레임의 기준 신호를 의미하고 $C_k(n+\triangle)$는 k번째 프레임의 유사도 판별을 할 신호 즉, $\triangle$에 의해 이동하는 비교 신호를 나타낸 것이다. 최적 이동값 $\triangle_k$는 식(3)을 최대화하는 값으로 다음과 같이 구할 수 있는데,

$$\triangle_k=\arg\max_{\triangle}[NCC(k,\triangle)],\quad |\triangle|\le\triangle_{\max} \tag{5}$$

이 때 비교를 위한 범위는 $\triangle_{\max}$에 의해 정해지고, $\triangle_{\max}$이 클수록 계산량이 증가한다.

III. 제안된 계산량 감축 방법

WSOLA의 계산량은 식(3)의 반복 계산에 집중되어 있고 전체 계산량에서 차지하는 비율은 90% 이상이다. 식(3)은 다수의 곱셈과 덧셈으로 이루어지는데 $\triangle$가 이동하는 구간, $[-\triangle_{\max},\triangle_{\max}]$만큼 반복해서 계산하게 된다. 따라서 식(3)의 계산량을 줄이는 것이 WSOLA의 계산량 감소로 직결된다. 식(3)을 살펴보면 근접한 구간 내에서 중복성을 가짐을 알 수 있고 이러한 중복에 대해서 계산을 피하는 방법인 합-테이블(sum table) 방법을 사용하면 계산 횟수를 줄일 수 있다.[5] 식(3)을 분모와 분자의 수식으로 나눠서 각각의 계산량 감축 방법을 이하에 기술한다. 참고로 분모 부분의 계산량 감축에 대해서는 [5]의 아이디어를 그대로 채용하여 [6]에서 발표한 바와 동일하다.

3.1 분모의 계산량 감소[6]

식(3)의 분모는 각 신호들의 에너지이다. 그런데 $NCC(k,\triangle)$을 최대화하는 $\triangle$인 $\triangle_k$를 구하는 과정에 있어서 $\triangle$와 상관없이 계산되는 기준 신호 $R_k(n)$의 에너지는 상수 성분으로 볼 수 있어 식(5)에서는 계산할 필요가 없다.

비교 신호 $C_k(n+\triangle)$의 에너지는 인접 구간에 대해서 그림 1에서 보듯이 중복 계산을 한다. 이 경우에 합-테이블 방법을 쓰면 인접 구간 에너지에 대해서

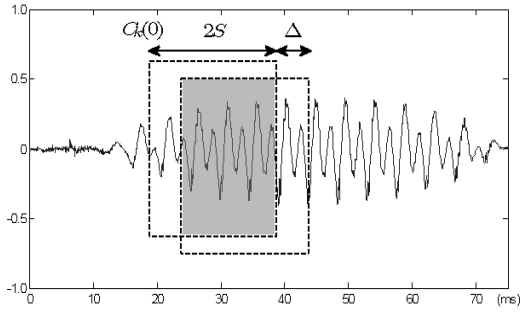

그림 1. 인접 구간인 $C_k(0)$과 $C_k(\triangle)$의 에너지 중복 계산
Fig. 1. Redundant calculation of energies of $C_k(0)$ and $C_k(\triangle)$.

중복 계산을 제거할 수 있다. 합-테이블 방법은 먼저 에너지 계산을 하기 전에 다음과 같이 합 테이블을 만드는데 회귀식의 형태로 계산한다.

$$s_{C_k}(u)=\sum_{n=-\triangle_{\max}}^{u} C_k^2(n) = \begin{cases} C_k^2(u)+s_{C_k}(u-1) & (-\triangle_{\max} \le u \le 2S-1+\triangle_{\max}) \\ 0 & (u<-\triangle_{\max}) \end{cases} \tag{6}$$

이제 식(6)의 테이블을 이용하면 $C_k(n+\triangle)$ 의 에너지는 다음 식과 같이 정리될 수 있다.

$$\sum_{n=0}^{2S-1} C_k^2(n+\triangle)=\sum_{n=0}^{2S-1+\triangle} C_k^2(n)-\sum_{n=0}^{-1+\triangle} C_k^2(n) = s_{C_k}(2S-1+\triangle)-s_{C_k}(-1+\triangle) \tag{7}$$

이 테이블을 활용하면 인접한 구간에 대해 에너지를 구하는 계산량은 표 1과 같이 감소되며 전체 정규상호상관도 계산의 대략 50%가 감소됨을 알 수 있다.

3.2 분자의 계산량 감소

정규상호상관도의 분자 부분은 상호상관 함수이다. 정규상호상관도의 분모 부분과는 달리 상호상관 함수는 경우에 따라 중복 계산이 발생할 수도 있고 발생하지 않을 수 있다. 저배속 변환에서는 그림 2(a)

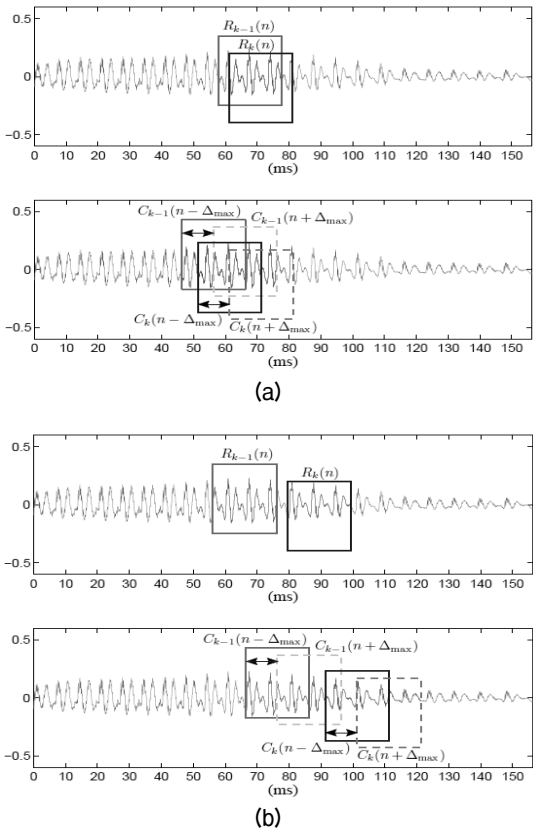

그림 2. 저배속과 고배속에서의 중복 계산 비교
(a) 저배속의 WSOLA
(b) 고배속의 WSOLA

Fig. 2. Comparison of redundant calculation in the time-scale expansion and time-scale compression.
(a) WSOLA based time-scale expansion
(b) WSOLA based time-scale compression

표 1. 기존의 WSOLA 방법과 제안한 합 테이블 방법의 정규상호상관도 분모 부분 산술 계산 비교.

Table 1. The comparison of arithmetic operations in denominator of NCC between the conventional WSOLA method and proposed sum-table method.

	Conventional method		Proposed method	
	Add/Sub	Mul	Add/Sub	Mul
$\sum_{n=0}^{2S-1} R_k(n)C_k(n+\triangle)$	$(2S-1)\times(2\triangle_{\max}+1)$	$2S\times(2\triangle_{\max}+1)$	$(2S-1)\times(2\triangle_{\max}+1)$	$2S\times(2\triangle_{\max}+1)$
$\sum_{n=0}^{2S-1} R_k^2(n)$	$2S-1$	$2S$	0	0
sum table of $\sum_{n=0}^{2S-1} C_k^2(n+\triangle)$	0	0	$2S+2\triangle_{\max}$	$2S+2\triangle_{\max}+1$
$\sum_{n=0}^{2S-1} C_k^2(n+\triangle)$	$(2S-1)\times(2\triangle_{\max}+1)$	$2S\times(2\triangle_{\max}+1)$	$2\triangle_{\max}$	0
Total	$(2S-1)\times(4\triangle_{\max}+3)$	$2S\times(4\triangle_{\max}+3)$	$4S\triangle_{\max}+4S+2\triangle_{\max}-1$	$4S\triangle_{\max}+4S+2\triangle_{\max}+1$
ex) $S=160,\ \triangle_{\max}=80$	103,037	103,360	51,999	52,001

에서 보는 바와 같이 인접 프레임들에 대해 상호상관 함수의 중복계산 부분이 존재한다. 하지만 그림 2(b)에서 보는 바와 같이 고배속의 변환에서는 매번 새로운 신호들에 대해 계산이 이루어지므로 중복을 고려할 여지가 없다. 따라서 저배속의 시간축 변환의 경우에만 중복 계산을 제거하기 위해 정규상호상관도의 분자 부분에 대해서도 다음 식과 같이 합 테이블을 구성한다.

$$s_{R_k,C_k}(u+1,\tau) = \begin{cases} R_k(u)C_k(u+\tau)+s_{R_k,C_k}(u,\tau) & \begin{pmatrix} 0 \le u \le 2S-1 \\ -\triangle_{\max} \le \tau \le \triangle_{\max} \end{pmatrix} \\ 0 & otherwise \end{cases} \quad (8)$$

언급한 것처럼 저배속의 변환에서 현재 프레임의 합 테이블은 이전 프레임이나 다음 프레임의 합 테이블과 중복을 가질 수 있다. 앞서 나타낸 그림처럼 중복을 가질 경우 중복 크기만큼의 합 테이블 값을 가져오면 가져오는 만큼의 계산을 피할 수 있다.

합테이블의 중복 구간을 구하기 위해서 먼저 기준 신호 $R_k(n),\ (0 \le n \le 2S-1)$과 $R_{k-1}(n),\ (0 \le n \le 2S-1)$의 중첩 길이를 구하면 식(4)에 의해 $(2S-1+(k-2)S\alpha+\triangle_{k-2})-((k-1)S\alpha+\triangle_{k-1})$로 계산되며, 이 값이 음수일 경우 중첩 구간이 없는 것이므로 실제 중첩 길이 $L_1(k)$는 다음 식과 같이 표현된다.

$$L_1(k)=\max((2S-1-S\alpha-\triangle_{k-1}+\triangle_{k-2}),0) \quad (9)$$

그 다음 상호상관 곱의 또 하나의 성분인 $C_k(n+\triangle)$에 대해서 이전 프레임의 $C_{k-1}(n+\triangle)$와 중첩 구간을 구한다. 식(4)에 의해 $R_{k-1}(n)$과 $C_{k-1}(n+\triangle)$ 사이의 시간 차이는 $\alpha S-S-\triangle_{k-2}+\triangle,\ (-\triangle_{\max} \le \triangle \le \triangle_{\max})$의 구간을 가지고 $R_k(n)$과 $C_k(n+\triangle)$가 $\alpha S-S-\triangle_{k-1}+\triangle, (-\triangle_{\max} \le \triangle \le \triangle_{\max})$의 구간을 갖는다. 따라서 중첩이 있는 경우를 고려한 구간 $L_2(k)$는 다음 식과 같이 나타난다.

$$L_2(k)=\min\begin{pmatrix}(2\triangle_{\max}-\triangle_{k-1}+\triangle_{k-2}),\\(2\triangle_{\max}+\triangle_{k-1}-\triangle_{k-2})\end{pmatrix} = 2\triangle_{\max}-|\triangle_{k-1}-\triangle_{k-2}| \quad (10)$$

이전 프레임에서의 합 테이블이 기준 신호 $R_{k-1}(n)$로부터 $C_{k-1}(n+\triangle)$와의 시간 차이로 이뤄진 $s_{R_{k-1},C_{k-1}}(u,\tau)$이고 현재 프레임의 $R_k(n)$과 $C_k(n+\triangle)$로부터 생성된 $s_{R_k,C_k}(u,\tau)$도 마찬가지이다. $R_{k-1}(n)$과 $R_k(n)$이 중첩되는 구간을 식(9)에서 구하고 $C_{k-1}(n+\triangle)$와 $C_k(n+\triangle)$이 중첩되는 구간을 식(10)에서 구했다. 식(9)와 식(10)에 의해 중첩되는 구간은 동일한 상호상관 함수 곱이기 때문에 다시 계산하지 않고 그림 3과 같이 이전 프레임, 즉, $s_{R_{k-1},C_{k-1}}(u,\tau)$의 합 테이블 값을 $s_{R_k,C_k}(u,\tau)$에 가져다 쓸 수 있다.

정규상호상관도 분자 성분의 전체 중복량 R_{L_1,L_2}는 다음 식으로 표현되며 식(9)과 식(10)의 곱의 전체 프레임 합에 해당하는데, 저배속에서 배속이 낮을수록 $R_{k-1}(n)$과 $R_k(n)$의 중복이 많아지고, 또한 최적 이동값 $\triangle_{k-2}$과 $\triangle_{k-1}$의 차이가 적을수록 중복이 많아진다.

$$R_{L_1,L_2}=\sum_k L_1(k)L_2(k) \quad (11)$$

IV. 실험 및 결과

4.1 실험 환경

기존 WSOLA의 중복을 제거하여 계산량 감소를 확인하는 실험을 위한 음성 데이터베이스로 원광대 국어공학 센터에서 구축한 한국어 Phonetically Balanced Sentence(PBS) DB를 사용하였다. PBS DB 내의 화자 중 임의의 남성화자 3명, 여성화자 3명을 선택하고 발성된 문장들을 연결해서 각 화자에 대해서 1분 가량의 발성이 되도록 하였다. 각 문장들은 대략 4초에서 11초 가량의 길이를 가지는데 문장의 처음과 끝에 존재하는 무음 구간에 대하여 처음과 끝에 100 ms 가량을 각각 제거하였다. 100 ms 가량의 무음 구간에는 마이크를 켜는 잡음이 있고 제거한 길이 내에 포함되도록 하였다. 음성 데이터는 잡음이 없는 상태에서 녹음되었고 16 kHz로 샘플링되었으며 16비트로 양자화되어 있다.

하드웨어는 인텔의 Core2Duo E8500 CPU와 3.25 GB 램의 PC이고, 소프트웨어는 윈도우 XP 32비트의

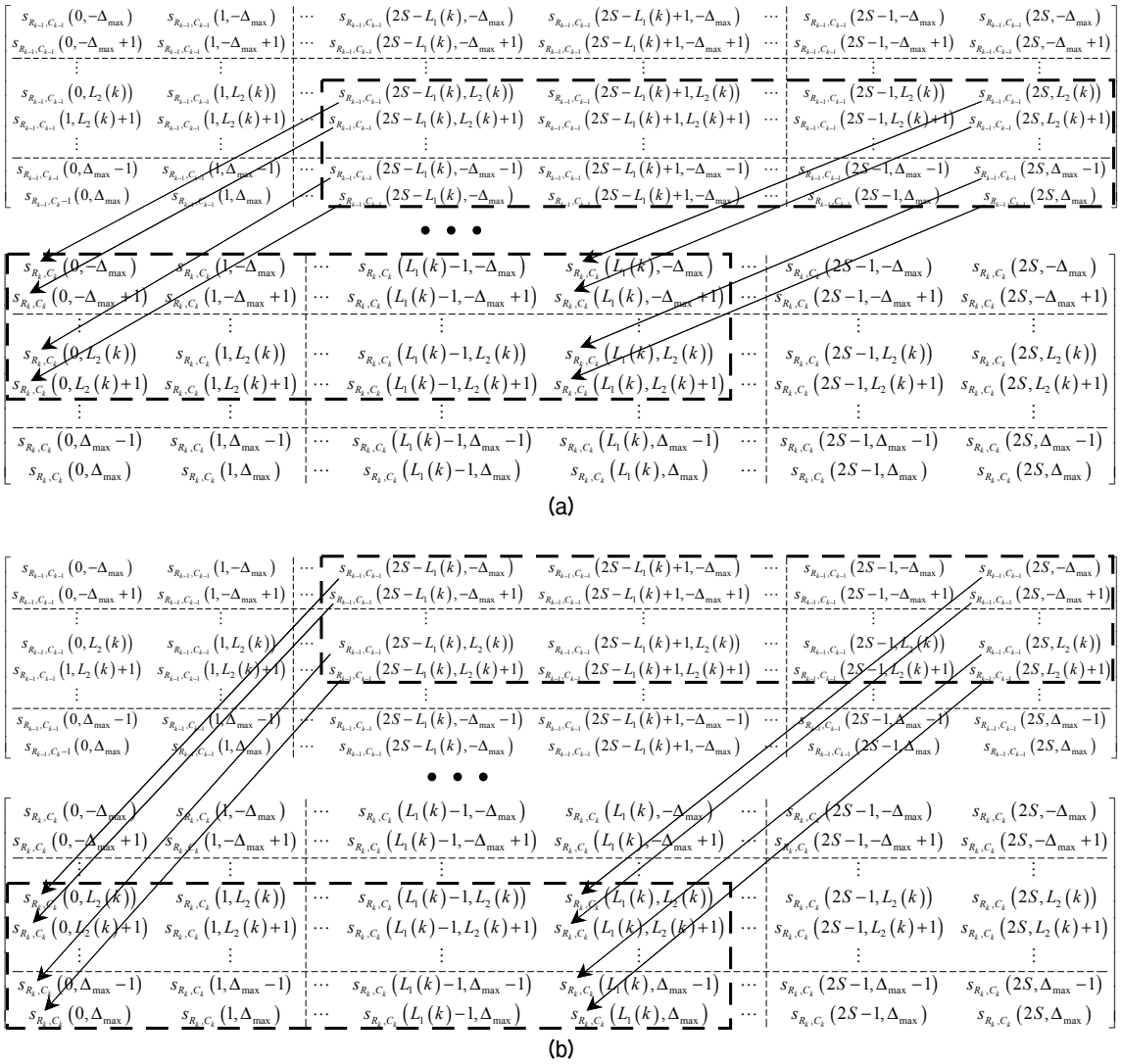

그림 3. 정규상호상관도 분자 합 테이블의 중복 구간
(a) 식(10)에서 $\triangle_{k-1} > \triangle_{k-2}$인 경우
(b) 식(10)에서 $\triangle_{k-1} < \triangle_{k-2}$인 경우

Fig. 3. Redundant region of sum tables for the numerator of NCC.
(a) In case of $\triangle_{k-1} > \triangle_{k-2}$ in eq (10)
(b) In case of $\triangle_{k-1} < \triangle_{k-2}$ in eq (10)

운영체제에서 마이크로소프트 사의 visual studio 2008 버전 내의 C언어로 작성한 소스코드를 디버그 모드에서 실행하여 실행결과를 고해상도 시간 측정 하였다.

4.2 실험 결과

본 논문에서 제안된 방법의 성능을 평가하기 위해 기본적인 WSOLA 방식, 분모의 중복만을 제거한 WSOLA 방식, 분모와 분자의 중복성을 모두 제거시킨 WSOLA 방식의 3가지에 대해 3, 2, 1/2 및 1/3 의 4가지 시간축 변환 비율로 비교하였다. 윈도우는 20 ms Hanning 윈도우를 쓰고 $\triangle_{\max}$는 중첩 길이의 절반인 5 ms로 하여 4가지 시간축 변환 비율에 대해 동일하게 적용했다. 남성화자와 여성화자의 실행시간 측정 결과를 그림 4에 제시하였으며 화자별 결과에 표준편차를 10배 크게 하여 오차막대(error bar)로 함께

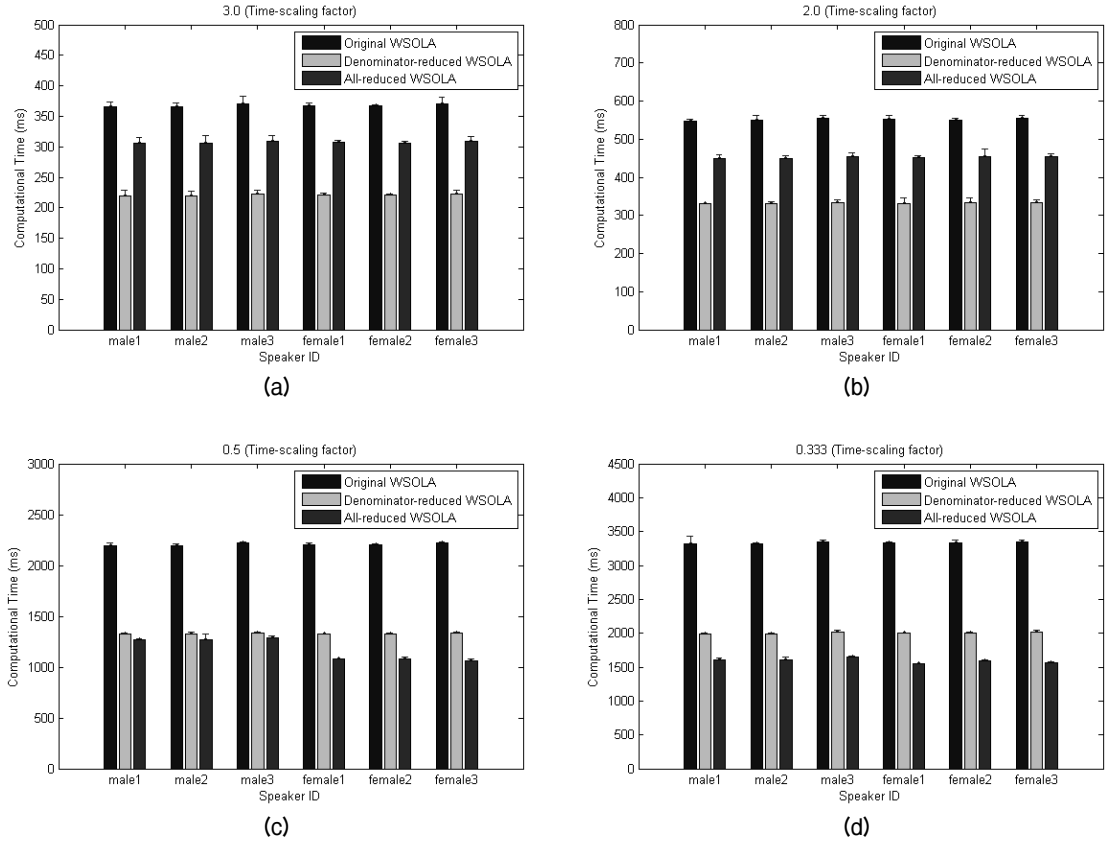

그림 4. 서로 다른 방법들의 실행 시간 측정 결과(ms) (오차막대는 표준편차의 10배 값)
(a) 3배속, (b) 2배속, (c) 1/2배속, (d) 1/3배속
Fig. 4. The execution time (ms) of different methods (Error bars represent std. deviation × 10).
(a) α = 3, (b) α = 2, (c) α = 1/2, (d) α = 1/3

표시하였다. 분모 부분의 계산량을 감소시키는 것만으로 배속에 관계없이 40% 가량의 실행 시간을 줄일 수 있으며 저배속의 변환에서 분자 부분의 중복을 제거해서 1/2 배속에서 남녀 각각 42%, 51%를, 1/3배속에서 남녀 각각 51%, 53% 감소시켰다. 따라서 전체 평균을 고려하면 1/2배속에서 47% 1/3배속에서 52% 감소시켰다. 1/2배속에서 여성화자의 실행 시간 감소폭이 남성화자에 비해 큰데 이것은 식(10)에 의한 중복 구간이 많이 발생했기 때문이다. 결과적으로 고배속으로의 변환은 분모에 대해서만 합-테이블 방법을 써서 중복을 제거하는 것이 적절하고, 효과적이다. 그리고 저배속으로의 변환은 분모와 분자의 중복을 모두 제거하는 것이 전체 계산량 감소에 효과적이다. 이 같은 결과는 전체 합성 결과의 파형이 기존의 WSOLA 방식과 100% 동일한 가운데에 이루어졌음을 확인하였다.

V. 결 론

본 논문에서는 대표적인 시간축 변환 기술인 WSOLA의 계산량을 줄이는 방법에 관해 연구하였다. 제안한 방식의 기본 아이디어는 합 테이블을 이용하여 반복적인 정규상호상관도 계산의 중복성을 제거하는 것이다. 고배속에서는 정규상호상관도의 분모 부분에 대해서만 계산의 중복성이 있고, 저배속에서는 정규상호상관도의 분모와 분자 모두에 계산의 중복성이 있음을 고려하여, 본 논문에서는 고배속과 저배속 각각에 대해 최적의 계산량 감축 방식을 제시하였다. 실험 결과 제안된 방식이 이전의 시간축 변환 계산량 감축방식들과는 달리 기존의 WSOLA 방식과 완벽히 동일한 음질을 나타내면서도, 시간축 변환 비율에 따라 대략 40%에서 50% 정도의 계산량을 감축시킴을 확인하였다. 제안 방식에서

추가로 요구되는 메모리 용량도 얼마 되지 않기 때문에 모바일 기기와 같이 계산처리 능력이 상대적으로 떨어지는 장치에서의 음성의 시간축 변환에 효과적으로 사용될 수 있을 것으로 기대된다.

감사의 글

이 논문은 지식경제부 및 한국산업기술평가관리원의 QoLT기술개발사업의 일환으로 수행되었습니다(과제번호: 10036438).

참고문헌

1. S. Roucos and A. M. Wilgus, “High quality time-scale modification for speech,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* (*ICASSP*), Tampa, FL, pp. 493-496, 1985.
2. W. Verhelst and M. Roelands, “An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* (*ICASSP*), Minneapolis, MN, pp. 554-557, 1993.
3. E. Moulines and F. Charpentier, “Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones,” *Speech Commun.*, vol. 9, no. 5-6, pp. 453-467, 1990.
4. D. S. Kim, Y. H. Lee, H. K. Kim, S. H. Choi, J. W. Kim, M. B. Kim, “Complexity reduction of WSOLA-based time-scale modification using signal period estimation,” *Communications in Computer and Information Science*, vol. 120, pp. 155-161, 2010.
5. J. Luo and E. E. Konofagou, “A fast normalized cross-correlation calculation method for motion estimation,” *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control*, vol. 57, no. 6, pp. 1347-1357, 2010.
6. 임상준, 정용원, 김형순 “WSOLA 기반의 음속 변환을 위한 고속의 정규상호상관도 계산,” *2011 한국음성학회 가을 학술대회 발표논문집*, 85-86쪽, 2011.
7. D. W. Griffin and J. S. Lim, “Signal estimation from modified short-time Fourier transform,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 2, pp. 236-243, Apr. 1984.
8. S. Grotit, Y. Lavner, Time-scale modification of audio signals using enhanced WSOLA with management of transients, *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 106-115, Jan. 2008.

저자 약력

▸ 임 상 준(Sangjun Lim)
2009년 8월: 부산대학교 전자전기공학부 졸업(공학사)
2009년 9월 ~ 현재: 부산대학교 대학원 전자전기공학과(석사과정)
〈주관심 분야〉 음성합성, 음성인식

▸ 김 형 순(Hyung Soon Kim)
1983년 2월: 서울대학교 전자공학과 (학사)
1984년 2월: 한국과학기술원 전기및전자공학과 (박사과정 조기진학)
1989년 2월: 한국과학기술원 전기및전자공학과 (박사)
1987년 ~ 1992년: 디지콤정보통신연구소 선임연구원
1992년 ~ 현재: 부산대학교 전자공학과 교수
〈관심분야〉 음성인식, 음성합성, 음성신호처리