

피치 변환을 사용한 실시간 음성 변환 시스템

Real-time Voice Change System using Pitch Change

김 원 구

군산대학교 전자정보공학부

Weon-Goo Kim

School of Electronic and Information Eng., Kunsan National University

E-mail : wgkim@kunsan.ac.kr

요 약

In this paper, real-time voice change method using pitch change technique is proposed to change one's voice to the other voice. For this purpose, sampling rate change method using DFT (Discrete Fourier Transform) method and time scale modification method using SOLA (Synchronized OverLap and Add) method is combined to change pitch. In order to evaluate the performance of the proposed method, voice transformation experiments were conducted. Experimental results showed that original speech signal is changed to the other speech signal in which original speaker's identity is difficult to find. The system is implemented using TI TMS320C6711DSK board to verify the system runs in real time.

1. 서론

정보 및 통신 문화가 급속히 발달함에 따라 의사 전달의 중요한 수단인 음성 신호 처리에 관한 연구가 활발히 진행되고 있다. 음성 신호 처리에 관한 연구는 크게 음성 부호화, 음성 인식, 음성 합성 및 음성변환으로 나눌 수 있다. 이중 음성 부호화, 음성 인식 및 음성 합성은 응용 분야가 다양하여 최근 수년간 활발히 연구되는 분야이다. 이에 비하여 음성 변환은 변환시킬 음성 특징 변수가 제한되어 있으며 응용분야가 비교적 제한되어 활발하지는 않으나 꾸준히 연구가 진행되고 있는 음성 신호 처리의 한 분야이다.

음성 변환(voice change)은 음성 신호로부터 구한 특징 파라미터를 다른 값으로 변환시킨 후 다시 합성하여 원래 음성과는 다른 음성을 얻는 기법을 말한다[1-8]. 이러한 방법중 대표적인 것으로 발음의 속도를 변화시키는 시간축 변환(time scale modification), 억양을 변화시키는 피치 변환(pitch modification)과 포먼트 등을 변화

시켜 특수한 효과음을 발생시키는 기법 등을 들 수 있다. 이러한 음성변환 기법들은 어학 학습기, 가정용 VCR, 음성 암호화 장치, 특수한 효과음을 얻기 위하여 사용되고 있다.

본 연구의 목표는 피치 변환을 이용한 실시간 음색 변환 알고리즘의 개발이다. 본 연구에서는 원래 화자의 목소리와 다른 여가가지 음색의 음성을 만들기 위하여 피치 변환을 이용한 음색 변환 알고리즘을 제안하여 음성 신호의 피치를 다양하게 변화하는 것이 가능하도록 하였다. 이것을 위하여 본 연구에서는 표본화율 변화 방법과 시간축 변환 방법인 SOLA(Synchronized OverLap and Add)[2] 방법을 결합한 피치 변환 방법을 제안하였다.

본 연구의 구성은 다음과 같다. 서론에 이어 2절에서는 피치 변환을 이용한 음색 변환 알고리즘에 관하여 설명하고 3절에서는 남녀 음성 신호를 이용한 실험 및 결과에 대하여 알아보고 4절에서 결론을 맺는다.

2. 음색 변환 알고리즘

2.1 피치 변경에 의한 음색 변환 알고리즘

피치 변경에 의한 음색 변환 방법은 원래 화자의 목소리와 다른 여가가지 음색의 음성을 만들기 위하여 음성 신호의 피치를 다양하게 변화시키는 것이다. 이것을 위하여 본 연구에서는 피치 변환 기법의 하나인 표본화율 변화 방법과 시간축 변환 방법인 SOLA 방법[2]을 결합한 음색 변환 방법을 사용하였다. 음색 변환 알고리즘의 블록도는 그림 1과 같다.

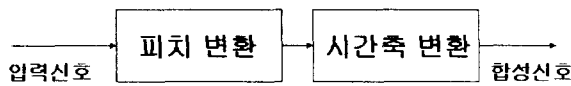


Figure 1. construction of voice change algorithm by pitch change

2.1.1 피치 변환 기법

음성 신호의 특징 변수중 성도 모양에 따른 포먼트(formant) 정보, 목소리의 높낮이와 음색을 나타내는 피치(pitch) 정보는 사람에 따라 독특한 특징을 나타내므로 매우 중요하다.

그림 2는 피치 변환 기법의 블록도이다. 여기서 P1, P2는 각각 입력 및 출력 신호의 피치 정보이고 f1, f2는 신호의 표본화율을 나타낸다. 입력 신호의 표본화율 f1을 표본화율 변환기(sampling rate converter)로 표본화율을 f2로 변환시키면 원 신호의 피치 정보 P1은 변화되지 않는다. 그 후 표본화율이 변환된 신호를 원 신호의 표본화율 f1으로 재표본화(resampling)하면 피치 정보는 P1에서 P2로 변화된다. 그러므로 피치 보정 방법은 표본화율을 변환시킨 후 출력 신호의 표본화율과 입력 신호의 표본화율을 같게

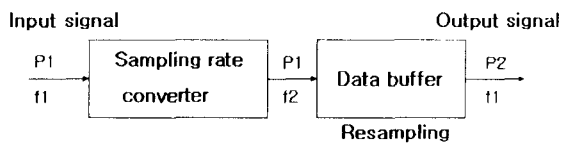


Figure 2. block diagram of pitch change technique

구함으로서 구현될 수 있다.

본 연구에서는 표본화율 변환 방법으로 단구간 푸리에 변환 방법을 이용하였다. N 개의 입력 신호를 M 개의 신호로 변환하는 방법은 다음과 같다. 우선 N 개의 음성 신호 $s(n)$ 의 이산 푸리에 변환 $S(n)$ 이라고 할 때 보간 ($M > N$) 또는 간축 ($M < N$)된 신호의 이산 푸리에 변환 $S'(n)$ 은 다음과 같이 만든다.

$$S'(m) = \begin{cases} S(m), & 0 \leq m < N/2 \\ 0, & N/2 \leq m \leq M - N/2 \text{ if } M > N \\ S(m - M + N), & M - N/2 < m < M \end{cases} \quad (2-1a)$$

$$S'(m) = \begin{cases} S(m), & 0 \leq m < N/2 \\ 0, & m = M/2, \text{ if } M < N \\ S(m - M + N), & M/2 < m < M \end{cases} \quad (2-1b)$$

보간 또는 간축된 음성 신호 $s'(n)$ 은 역 푸리에 변환에 의하여 변환된다.

2.1.2 시간축 변환 기법

피치 변환에 의하여 변환된 음성 신호는 입력 신호와 출력 신호간에 시간적인 차이가 발생한다. 다시 말하면 N 개의 음성 신호의 피치를 M 으로 변경한 경우, $M > N$ 이면 음성 신호의 샘플 수가 증가하게 되고 $M < N$ 이면 음성 신호의 샘플 수가 감소되어 출력된다. 그러나 이러한 방법은 실시간 시스템으로 구현될 경우 DSP 프로세서가 N 분 길이의 입력 신호를 처리하여 $M(M > N)$ 분 길이의 음성 신호로 출력해야 하는 모순이 발생한다. 따라서 피치 변환 기법을 시간축으로 구현하기는 불가능하다. 본 연구에서는 이러한 문제를 해결하기 위하여 음성 신호의 시간축 변환 방법(time scale modification)중의 하나인 SOLA 방법[2]을 사용하였다. 음성 신호의 시간축 변환 방법은 음성 신호의 기본 주파수와 성도 모델 스펙트럼을 보존하여 원래의 신호 특성은 그대로 유지하면서 발음속도만 변화시키는 것이다.

그림 3은 SOLA 알고리즘에 대한 설명이다. 음성 신호 $s(n)$ 을 시간축상으로 a 만큼 변환시킨 신호를 $y(n)$ 이라 할 때, 입력 신호로부터 매 Sa 샘플마다 N 개의 신호를 얻어내어 출력신호 $y(n)$ 과 동기를 맞추기 위해 상관 관계가 최대가 되는 점으로 현 프레임의 이동한 후 overlap하여

더함으로써 S_s 샘플마다 합성된 출력 신호 $y(n)$ 이 생성된다.

SOLA 알고리즘은 모든 연산이 시간 축에서만 이루어지므로 반복적인 연산을 수행해야하는 LSEE-MSTFT(Least Square Error Estimation-Modified Short Time Fourier Transformation) [9] 보다 계산상의 장점이 있으며, 성능 또한 우수하여 고음질의 음성 신호를 합성할 수 있어 실시간 시스템에 적합한 방법이다.

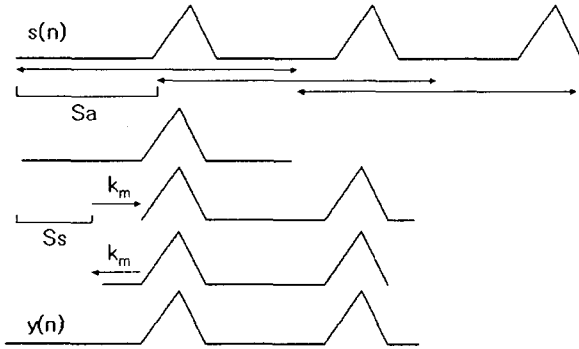


Figure 3. SOLA algorithm

3. 실험 및 결과 고찰

본 실험에서는 제안된 음성 변환 알고리즘을 구현하고 그 동작을 확인하고자 컴퓨터 모의 실험을 수행하였다. 실험에 사용된 음성 데이터는 여러 가지 음소가 포함되어 있는 문장을 선정하여 2명의 남성 및 여성 화자로부터 수집하였다. 실험에 사용한 음성 데이터는 비교적 조용한 일반 사무실 환경과 약간의 잡음이 존재하는 실외에서 이루어졌다. 녹음은 디지털 테이프 녹음기를 사용하였다. 녹음된 음성은 8kHz 16비트로 A/D 변환되어 실험에 사용되었다.

피치 변경에 의한 음색 변환 시스템의 전체 시스템 구성 블록도는 그림 4와 같다. 8kHz로 샘플링되어 들어오는 입력 음성은 256 샘플 단위로 분석 구간을 잡는다. 현재 프레임 신호는 표본화율 변환기에서 $N:M$ 의 비율로 보간($M>N$) 또는 간축($M<N$)되어 M 개의 음성 샘플 $s'(n)$ 을 출력한다. 이때 분석 구간의 이동은 $M/2$ 로 한다. 표본화율이 변환된 음성 신호 $s'(n)$ 과 출력 신호 $y(n)$ 과의 상호 상관 함수를 구하여 그 값이 최대가 되는 lag k_m 을 구한다. 그리고 그 위치에 표본화율이 변환된 음성 신호 $s'(n)$ 을 출력신호 $y(n)$ 과 overlap and add 방식으로 더한다. 출력 신호 $y(n)$ 은 입력 신호의 분석 구간 이동과 같은 $M/2$ 씩 이동하며 출력된다.

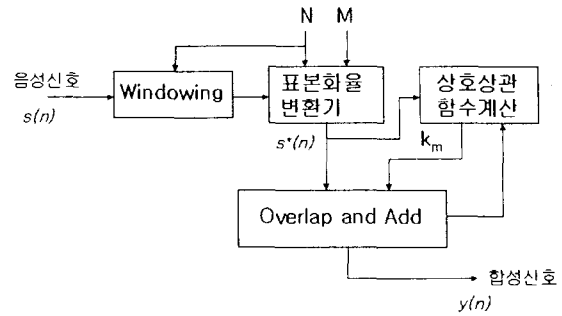


Figure 4. block diagram of pitch change system by pitch change

그림 5, 그림 6은 피치 변경에 의하여 변형된 음성을 보여준다. 그림 6은 남성 화자의 음성 신호 피치를 확대(b) 및 축소(c)한 것이고 그림 7은 여성 화자의 음성 신호 피치를 확대(b) 및 축소(c)한 것이다. 그림에서 알 수 있듯이 피치 확대 및 축소할 경우, 음성 신호의 전체적인 길이와 포락선은 변화가 없고 단지 피치만 확대 또는 축소된 것을 알 수 있다.

제안된 방법을 실시간 시스템으로 구현하기 위하여 Texas Instrument사의 TMS320C6711 DSK 범용 신호 처리 보드를 사용하여 구현하였다. 구현된 실시간 시스템은 실시간으로 동작되는 것이 확인되었다.

4. 결론

본 연구에서는 범용 오디오 프로세서를 이용한 음색 변환기 구현을 위한 실시간 음색 변환 알고리즘 개발을 목표로하여 다양한 음색 변환이 가능한 실시간 음색 변환 알고리즘을 개발하였다.

본 연구에서는 원래 화자의 목소리와 다른 음색의 음성을 만들기 위하여 피치 변환을 이용한 음색 변환 알고리즘을 제안하여 음성 신호의 피치를 다양하게 변화하는 것이 가능하도록 하였다. 이것을 위하여 본 연구에서는 표본화율 변화 방법과 시간축 변환 방법인 SOLA 방법을 결합한 피치변환 방법을 사용하였다.

제안된 방법을 Texas Instrument사의 MS320 C6711DSK 범용 신호 처리 보드를 사용하여 구현하여 실시간으로 동작되는 것이 확인되었다.

음성 신호를 사용한 실험에서 피치가 변형된 음성은 원래의 화자와 특성을 찾을 수 없는 목소리로 변형된 것을 확인할 수 있었다.

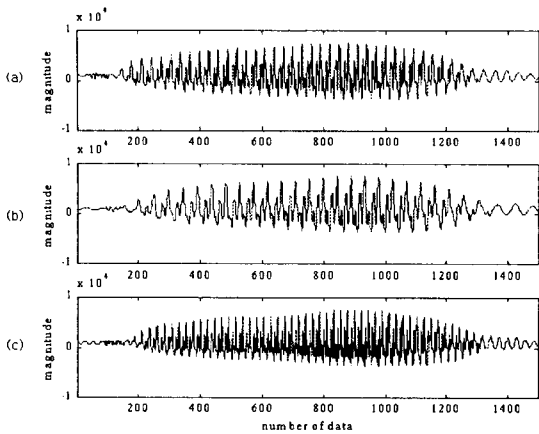


Fig. 5 voice changed speech signal(male)
 (a) original speech (b) pitch expanded by 1.5 times (c) pitch compressed by 0.75 times

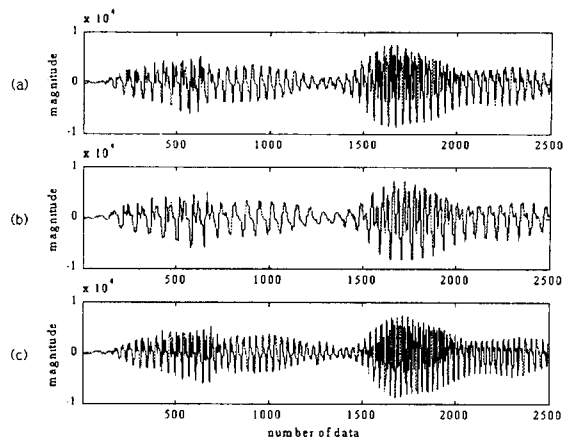


Fig. 6. voice changed speech signal(female)
 (a) original speech (b) pitch expanded by 1.5 times (c) pitch compressed by 0.75 times

감사의 글

본 연구는 한국과학재단 목적기초연구 (R05-2003-000-12043-0) 지원으로 수행되었음.

5. 참고문헌

- [1] S. Roucos and A. M. Wilgus, "High quality time-scale modification for speech," *proc. of ICASSP*, vol. 1, pp. 493-469, (1985)
- [2] J. Makhoul and A. E. Jaroudi, "Time-scale modification in medium to low rate speech coding," *proc. of ICASSP*, vol. 1, pp. 1705-1708, (1986)
- [3] E. Hardam, "High-quality time scale modification of speech signals using fast synchronized-overlap-add algorithm," *proc. of ICASSP*, vol. 1, pp. 409-412, (1990)
- [4] E. Moulines and F. Charpentier, "Pitch Synchronous Waveform Processing Techniques for Text-to-speech Synthesis using Diphones," *Speech Communication*, vol. 9 (5/6), pp. 453-467, (1990)
- [5] E. Moulines and J. Laroche, "Non-parametric techniques for pitch-scale and time-scale modification of speech," *Speech Communication*, vol. 16, pp. 175-205, (1995)
- [6] R. J. McAulay and T. F. Quatieri, "Speech

- transformations based on a sinusoidal representation," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. 34, No. 1, pp. 1449-1464, December, (1986)
- [7] T. F. Quatieri and R. J. McAulay, "Shape invariance time-scale & pitch modification of speech," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. 40, No. 3, pp. 497-510, March, 1992.
- [8] L. R. Rabiner and R. W. Schafer, "Digital Processing of Speech Signal", Prentice-Hall Inc., 1978.
- [9] D. W. Griffin and J. S. Lim, "Signal estimation from the modified short-time fourier transform," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. ASSP-32, pp. 236-243, April, (1984)