다양한 발성에 따른 다중음성 합성 시스템

박현영, 김명, 배명진
숭실대학교 정보통신공학과

Mutiple-Speech Synthesis System according to Various Utterance

HyunYoung Park, Myoung Kim, MyoungJin Bae
Dept. Information and Telecommunication Engr, Soongsil Univ

Abstract

음성 합성이란 기계적인 장치나 전자회로 또는 컴퓨터 모의를 이용하여 자동으로 음성파형을 생성해 내는 것으로 정의한다. 음성 합성에 대한 연구는 다른 음성에 관련된 기술들보다 가정 먼저 연구된 기술이다. 음성 합성기는 PC의 보급이 확대되고 통신 시장이 커짐에 따라 그 응용 분야가 점차 확대되어 가고 다양한 방식의 음성 합성 기법에 관한 연구가 이루어지고 있다. 일반적으로 자연스러운 대화를 할 때나 글을 읽을 때의 음성에는 피치, 지속시간, 에너지 등의 운율 정보가 포함되어 있다. 따라서, 문장을 합성하는 경우 운율정보를 합성음에 반영하면, 보다 명확한 의미 전달과 다양한 발성변환이 가능해 진다. 본 논문에서는 시간영역에서 PSOLA 합성방식에 의한 피치 변경과 지속시간 변경을 이용하여 다양한 발성변환에 따른 다중음성 합성기를 구현하였다.

Keyword : 음성 합성, PSOLA, 피치, 지속시간, 발성변환

1. 서론

음성을 인간과 기계 사이의 정보전달 수단으로써 효율적으로 사용하기 위해서는 고음질의 합성음뿐만 아니라 다양한 음색을 갖는 합성음을 필요로 한다. 기존의 음성합성 방식은 단일음성 합성 시스템으로 단일 화자의 음성을 입력으로 받아 단일음성으로 합성된 음성을 출력하는 방식이다. 본 논문에서는 다양한 발성에 따른 다중 음성 합성기를 구현 하고자 한다. 이는 단일 화자의 음성을 입력으로 받아 피치와 지속시간 등의 운율을 조절하여 여러 화자의 발성으로 합성된 음을 출력하는 방식이다. 이와 같은 다중 음성 합성기는 다양한 분야에 응용 할 수 있다. 예를 들어, 운동 경기장 등에서 한 사람의 응원으로 여러 사람이 응원하는 효과를 낼 수 있는 응원 합성기, 생일이나 파티장 등에서의 축하 합성기, 돌림노래 합성 시스템, 영화의 효과음 등의 다양한 분야에 응용 할 수 있다.

2. AMDF 피치시점 검출법

AMDF(average magnitude difference function) 법은 자기상관함수에서 X(n)과 X(n-k)곱 대신에 다음 식(1)과 같이 절대값으로 정의된다.

$$AMDF(k)=\sum_{m=-\infty}^{\infty} |x(m)-x(m-k)| \qquad (1)$$

AMDF법은 자기상관함수법에서 수행하는 곱연산을 절대값과 차분으로 대신하기 때문에 상대적으로 빠르다는 장점을 가지고 있다. 이러한 이유로 실시간에 많이 적용한다. 자기상관함수법에서는 피치주기 배수에 최대값을 이루지만, AMDF법에서는 피치주기 배수에 최소값을 갖는다.

3. PSOLA 피치 조절

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

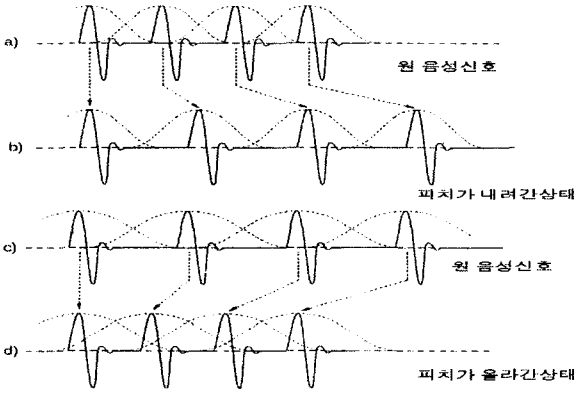

그림 1. PSOLA 피치 조절

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 곱함으로써 다음 식(2)와 같은 피치 단위로 분해된 샘플열들을 얻는다.

$$S_{analysis}(n)=W_{analysis}(m-n)S(n) \qquad (2)$$

$S_{analysis}(n)$: 피치주기 단위의 ST 신호

$W_{analysis}(n)$: 분석 윈도우 함수

m : m번째 피치

$S(n)$: 원 음성 파형

분석과정에서의 ST신호의 열은 원래의 음성 샘플의 피치단위로 배열되어있다. 따라서 피치를 변경하기 위해서는 이 간격들을 변경할 피치 간격들로 재배열하면 된다. 다음 식(3)은 피치가 변경된 신호를 나타낸 것이다.

$$S_{synthesis}(n)=S_{analysis}(n-m_a) \qquad (3)$$

$S_{synthesis}(n)$: 피치가 변경된 ST신호

m_a : 변경할 피치 간격

따라서 피치를 높일 때는 ST신호의 간격을 작게 배열하고, 피치를 낮출 때는 ST신호의 간격을 크게 배열하면 된다. 하지만 이런 순차적인 배열사이에서 정확한 피치 동기화를 유지하는 것이 중요하다. 이렇게 재배열된 ST신호에서 겹쳐지는 부분을 더해주면 된다.

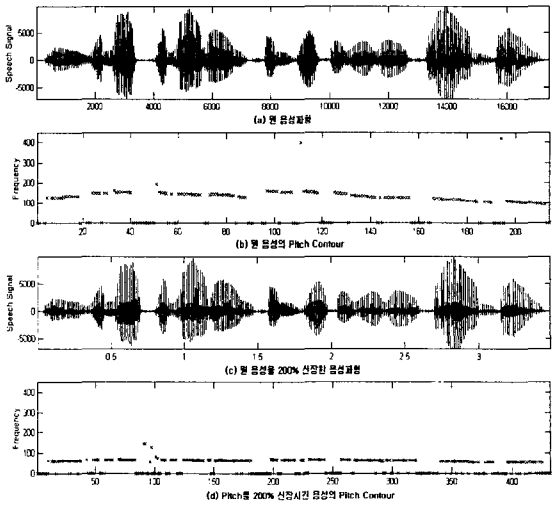

그림 2. 피치를 200% 신장한 예

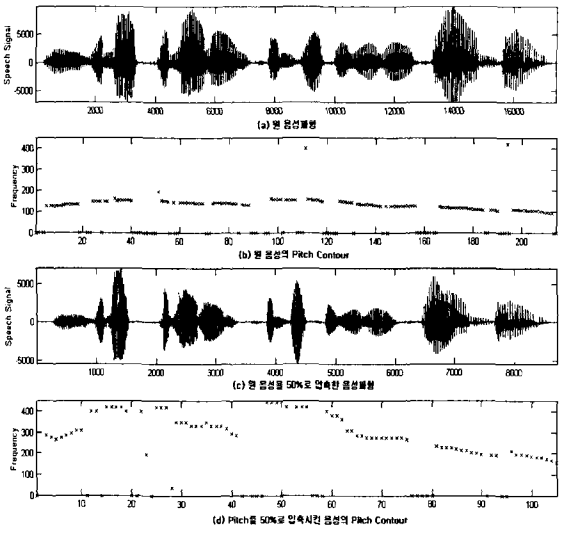

그림 3. 피치를 50% 압축한 예

그림 2, 그림 3은 피치주기 단위의 PSOLA 합성방식을 적용하여 피치를 변경한 예를 나타낸 것이다. 원래 음성의 pitch contour와 피치 변경된 음성의 pitch contour를 나타내고 있다.

4. PSOLA 지속시간 조절

피치 변경된 음성신호는 시간축이 변화기 때문에 원음성의 지속시간도 변한다. 화자의 발성을 변환하기 위해서는 피치 변경 이외에 지속시간도 피치 변경되기 이전의 음성과 같게 해주어야 한다.

따라서 본 논문에서는 PSOLA 방식을 이용하여 지속시간을 조절하였다. 지속시간 조절은 피치단위로 재배열된 신호를 반복 삽입하여 지속시간을 늘이거나 삭제하여 지속시간을 줄인다.

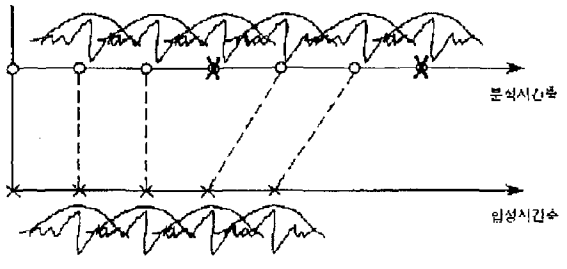

그림 4. PSOLA 지속시간 조절

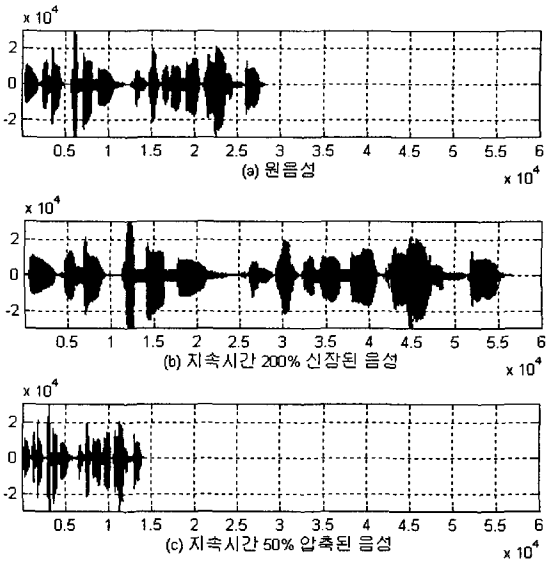

그림 5. 지속시간을 변경한 예

그림 4는 PSOLA 알고리즘을 이용하여 지속시간을 조절하는 방식을 나타내는 그림이다. 그림 5는 /인수네 꼬마는 천재 소년을 좋아한다./라는 한 문장 전체의 지속시간을 200% 신장한 음성과 50% 압축한 음성을 나타내고 있다.

5. 다중음성 합성 구현 알고리즘

그림 6은 PSOLA 방식을 이용하여 다양한 발성 변환에 따른 다중음성 합성 시스템 구현 알고리즘이다. 단일 음성을 입력받아 AMDF 방식을 이용하여 피치시점을 검출하고 변경율에 따라서 각각 피치를 재배열 하여 피치 변경된 합성음을 얻을 수 있다. 이와 같이 시간축에서 피치 변경된 음성은 지속시간이 변하게 되므로 PSOLA 지속시간 조절 방식을 이용하여 원래의 피치 변경되기 이전의 음성과 지속시간을 같게 만들어 주어 지속시간은 일정하게 유지하면서 각각의 피치 변경된 음성을 합성해 냈다. 이와 같이 피치 변경된 합성음에 각각 Delay를 두어 다중음성 합성 시스템을 구현 하였다. 그림 7은 VC++ 프로그램을 이용하여 구현한 다중음성 합성 시스템 프로그램이다.

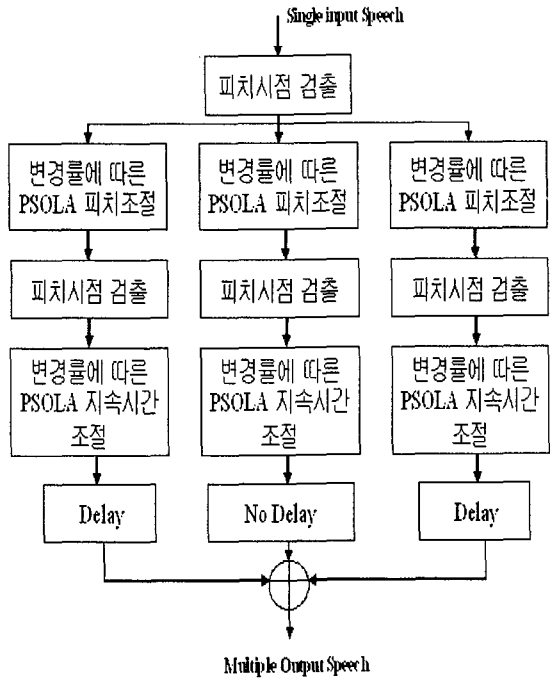

그림 6. PSOLA방식을 이용한 다중 음성합성 블록도

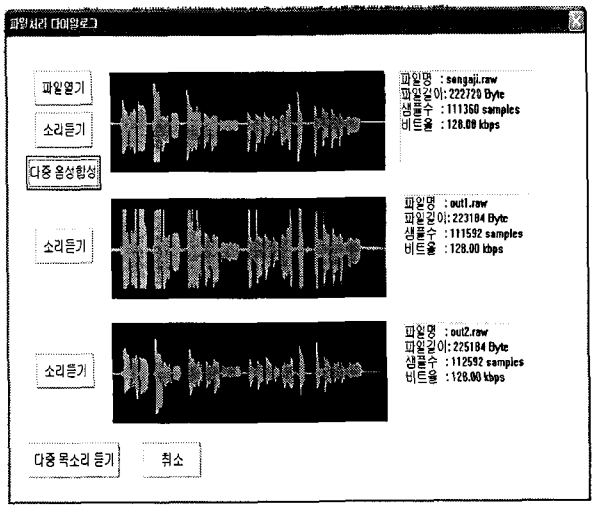

그림 7. 다중음성 합성 시스템 프로그램

6.. 결론 및 고찰

멀티미디어의 급속한 보급으로 인하여 인간과 인간 사이의 정보전달 뿐만 아니라 인간과 기계 사이의 정보전달도 매우 중요해 지고 있다. 이에 따라 인간의 가장 자연스러운 정보전달 수단인 음성을 인간과 기계 사이의 정보전달 수단으로 이용하려는 연구가 음성합성, 음성인식 등의 여러 분야에서 빠른 속도로 진행되고 있다. 음성 합성기는 PC의 보급이 확대되고 통신 시장이 커짐에 따라 그 응용 분야가 점차 확대되어 가고 다양한 방식의 음성 합성 기법에 관한 연구가 이루어지고 있다. 본 논문에서는 PSOLA 합성방식을 사용하여 다양한 발성변환에 따라 기존의 단일음성 합성 방식을 탈피하여 다중음성 합성 시스템 구현에 관하여 제안하였다. 시간축 영역에서 변경율에 따라 PSOLA 합성법을 이용하여 피치를 변경하였고, 각각 지속시간이 달라진 피치 변경된 음성을 다시 PSOLA 지속시간 변경법을 이용하여 원래의 지속시간으로 피치 변경된 합성음을 생성하고, 각각 Delay를 두어 단일음성 다중음성 합성 시스템을 구현하였다. 이와 같이 다양한 발성변환에 따른 다중 음성 합성 시스템은 운동 경기장 등에서 한 사람의 응원으로 여러 사람이 응원하는 효과를 낼 수 있는 응원 합성기, 생일이나 파티장 등에서의 축하 합성기, 돌림노래 합성 시스템, 영화의 효과음 등의 다양한 분야에 응용 할 수 있다. 향후 다양한 합성음을 얻기 위한 연구와 에너지 조절을 통한 음질의 개선에 관한 연구도 필요할 것이다.

참고문헌

[1] L.R.Rabiner, R.W.Schafer "Digital Processing of Speech Signals", PRENTICE HALL

[2] 배명진, 이상효 "디지탈 음성분석." 동영출판사

[3] 박 원 "음성신호의 실시간 운율 조절에 관한 연구." 숭실대학교 석사학위 논문 2001년 6월

[4] T.Takagi. E. Miyasaka, " A Speech Prosody Conversion System with a high Quality Speech Analysis-Synthesis Method", Proc EUROSPEECH93, pp.995-998, September 1993

[5] H.Valbret, E.Moulines, and J.P.Tubach, "Voice transformation using PSOLA technique", Speech Communication, vol. 11, pp.175-187