

자기상관함수에서 위상 성분의 보존에 의한 피치 시점 검출에 관한 연구

함명규, 최성영, 박종철, 배명진

승실대학교 정보통신공학과

On a Pitch Point Detection by Preserving the Phase Component of the Autocorrelation Function

Myungkyu Ham, Sungyoung Choi, Jongcheol Park, Myungjin Bae

Dept. of Telecommunication, Soongsil University

E-mail : mkham@assp.soongsil.ac.kr

요약문

음성신호처리 분야에서 음성신호의 기본 주파수를 정확히 검출할 수 있다면 음성인식을 할 때 화자에 따른 영향을 줄일 수 있으므로 인식의 정확도를 높일 수 있고, 음성합성을 할 때 자연성과 개성을 쉽게 변경하거나 유지할 수 있다. 또한 분석을 할 때 피치에 동기시켜 분석하면 성분의 영향이 제거된 정확한 성도 파라미터를 얻을 수 있다. 위와 같은 피치 검출의 중요성 때문에 피치검출에 대하여 다양한 방법이 제안되었다[1].

본 논문에서는 음성신호의 분석 시 불안정한 구간에 대해 피치 시점을 검출하는 방법을 연구하였다. 음성신호의 분석에 있어서 기존의 자기상관함수법(Autocorrelation Function)은 주기성을 강조할 수 있다는 장점을 가지고 있다. 그러나 자기상관함수는 위상 성분을 보존하지 못한다는 단점을 가지고 있다. 따라서, 자기상관함수를 사용하면서 위상성분을 보존할 수 있는 알고리즘을 제안하고자 한다.

실험결과 피치시점을 수동으로 찾은 경우와 비교하였을 때 약 98% 정도의 정확도를 얻을 수 있었다. 위의 결과와 같이 위상 성분이 보존된 자기상관함수를 사용할 경우 음성합성, 코딩, 인식에서 유용하게 쓰일 수 있다.

1. 서론

음성합성 시스템은 분류기준에 따라 분석된 데이터를 그대로 합성에 이용하는 분석에 의한 합성법과 규칙의 의해 음성을 발생시켜 합성하는 규칙에 의한 합성법으로 구분할 수 있다. 음성합성은 합성단위에 따라서 문장단위, 음절단위, 음소단위 등으로 나눌 수 있다. 예를 들면 가전 제품의 사용법 안내나 녹음 안내 사항에 대해서는 보통 문장 단위로 합성하고, 시간 및 기후안내에는 단어 단위의 합성법이 적용된다. 또한 상황이 복잡하고 모든 분야에 적용되어야 할 합성기법은 음절이하의 합성단위가 바람직하게 된다.

합성을 위해 출력되는 데이터의 처리 방식에 따라서는 메모리형 합성법과 전송형 합성법으로 구분할 수 있다. 메모리형 합성법은 분석한 데이터를 메모리에 저장시켜 두고 필요에 따라 다시 합성하는 방법으로 모저(Mozer)법, 기본 주파수단위 분절법 등이 있다. 이 방법은 메모리에 저장되는 데이터 형식이 번지당 일정 비트수로 제한되지 않으며, 분석시의 시간은 합성시의 시간과 반드시 일치시킬 필요도 없다. 또한 메모리의 저장 오류율은 무시될 정도이기 때문에 이를 방지하기 위한 별도의 기술이 거의 요구되지 않는다[3][4].

부호화 방식에 따라서는 파형 부호화법, 신호원

부호화법, 혼성 부호화법으로 분류할 수 있다. 파형 부호화법은 파형 자체의 잉여성분을 제거한 후에 부호화 하는 방법이며, PCM, ADPCM, ADM 등이 제안되어 있다. 이 부호화법은 인간의 개성과 감정을 대별해 주는 여기 정보와 메시지전달을 나타내는 여파기 정보를 분리하지 않고 처리하기 때문에 음원을 변경시켜야 하는 음절단위나 음소단위의 합성기법으로는 바람직하지 못하다[2].

최근 다양해진 음성서비스 분야에서는 고음질의 합성음을 요구하고 있다. 이러한 고음질 합성방식으로는 파형부호화법이 바람직하다. 그렇지만 파형부호화법을 사용하면 데이터베이스에 저장해야 할 메모리 규모가 방대하고 음원피치의 변경이 어렵다는 문제점이 발생한다. 부호화에 필요한 메모리 문제는 현재의 기술수준으로 충분히 극복이 가능하며, 나머지 문제를 해결하기 위해서는 파형부호화법을 규칙에 의한 합성에 적용하기 위해 음원피치를 변경시킬 수 있어야 한다.

본 논문에서는 피치 불안정 구간에서 피치 시점 검출을 위해 자기상관함수에서의 주기성 강조를 이용하고, 이때 위상성분을 보존 할 수 있는 방법을 제안하고자 한다.

2. 기존의 피치 검출법

피치검출기는 여러 가지 음성처리 시스템에 필수적인 요소이다. 피치검출기의 성능으로는 음성합성시에 여기원의 특성을 나타내어 음질의 자연성을 좌우한다. 그리고, 검출된 피치 변화도는 화자인식용 및 발음 장애자를 위한 보조 시스템용 파라미터로 널리 적용되는 등, 거의 모든 음성 분석-합성 시스템에 널리 쓰이고 있다. 지금까지 이러한 피치검출에 대한 다양한 알고리즘들이 제안되어 왔다. 그렇지만, 대부분의 피치검출기는 잡음이 있는 환경, 대역폭이 제한된 경우, 주파수의 진폭특성이 열화된 환경, 주파수의 위상특성이 변질된 경우 등과 같은 상황에서 신뢰할 만한 결과를 얻지 못하고 있다.

시간영역 검출법은 파형의 주기성을 강조한 후에 결정논리에 의해 피치를 검출하게 된다. 주기성 강조는 성도의 공명현상에 나타나는 포먼트 들의 영향을 제거시키고 여기원의 기본주파수만을 강조하는데 그 목적이 있다. 따라서 주기성 강조가 복잡하면 복잡할수록 결정논리가 간단해진다. 시간영역에서의 방법은 보통 시간영역에서 수행되므로 영역의 변환이 불필요하고, 합, 차, 비교논리 등 간단한 연산만 필요로

하다. 피치 검출의 결정 논리법에서는 강조된 음성 파형의 기본주기를 실험적인 문턱값이나 무게치를 적용하여 결정하는 것이 보통이다. 그 외에도 위상이나 진폭에 따라 몇 가지의 기준 패턴을 만들어 입력된 음성의 파형에서 추출한 패턴과의 유사도를 측정하여 주기성을 결정하기도 한다. 결정 논리가 실험적인 문턱값에 의존하면 검출오차가 확률적으로 커지기 때문에 경우의 수에 따라 결정 문턱값을 신중히 고찰해야 하는 복잡성이 따른다.

주파수 영역의 피치 검출법은 음성 스펙트럼의 고조파 간격을 측정하여 유성음의 기본주파수를 검출하는 방법이다. 일반적으로 스펙트럼은 한 프레임(20~40ms)단위로 구해지므로, 이 구간에서 음소의 천이나 변동이 일어나거나 배경 잡음이 발생하여도 평균화되므로 그 영향을 적게 받는다. 그러나 처리과정상 주파수 영역으로의 변환과정이 필요함으로 계산이 복잡하며, 기본주파수의 정밀성을 높이기 위해 FFT의 포인트 수를 늘리면 그만큼 처리시간이 길어진다.

시간-주파수 혼성영역법은 시간영역법의 계산시간 절감과 피치의 정밀성, 그리고 주파수영역법의 배경잡음이나 음소변화에 대해서도 피치를 정확히 구할 수 있는 장점을 취한 것이다. 이러한 방법으로는 Cepstrum법, 스펙트럼 비교법 등이 있고, 이 방법은 시간과 주파수 영역을 왕복할 때 오차가 가중되어 나타나므로 피치추출의 영향을 받을 수 있고, 또한 시간과 주파수영역을 동시에 적용하기 때문에 계산과정이 복잡하다는 단점이 있다.

3. 실험 및 결과

제안한 피치 변경법의 성능을 평가하기 위해 IBM-PC/pentium(500MHz)에 음성 입·출력용 16bit AD/DA 변환기를 인터페이스하여 사용하였다. 음성 시료로는 24세의 남성, 30세의 남성, 27세의 여성, 29세의 여성에게 아래와 같은 문장들을 각각 5회씩 발성하게 하고 11kHz로 표본화하고 16bit로 양자화하여 저장하였다.

- 발성1 : /인수배 꼬마는 천재소년을 좋아한다./
- 발성2 : /예수님께서 천지 창조의 교훈을 말씀하셨다./
- 발성3 : /여기는 음성 합성 연구실입니다./
- 발성4 : /창공을 헤쳐나가는 인간의 도전은 끝이 없다./

발성5 : /이용해 주셔서 감사합니다./

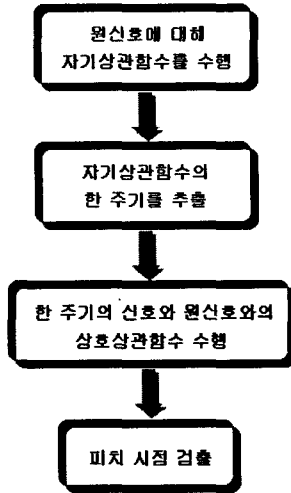


그림 4-1. 처리 블록도

알고리즘의 처리를 위한 분석 프레임의 길이는 256표본을 사용하였다. 그림 4-1은 처리를 위한 블록도를 나타낸 것이다. 그리고, 그림 4-2는 /인수내 꼬마는 천재소년을 좋아한다/의 파형에 대해 프레임 단위의 처리 결과를 나타낸 것이다.

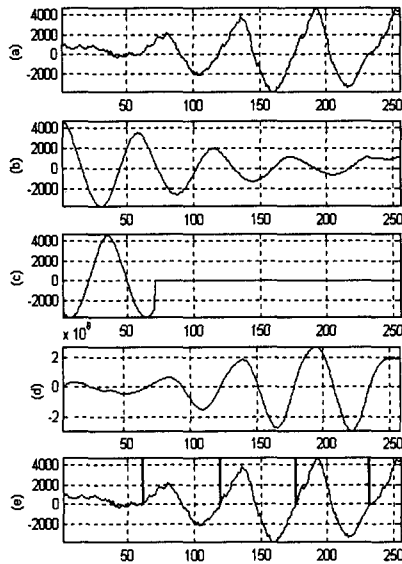


그림 4-2. 처리결과 예

(a) 원음성, (b) 자기상관함수, (c) 자기상관함수의 한주기파형, (d) 한주기를 위용하여 상호상관함수를 사용한 파형, (e) 피치시점

그림 (b)와 (d)를 비교해 보면, 자기상관함수만을 사용하였을 때에는 위상 성분이 없어 피치 주기를 얻는데는 문제가 없지만 피치 시점을 검출하는데는 어려움이 많음을 알 수 있다. 그러나, (d)와 같이 자기상관함수를 하여 얻어진 한 주기의 파형과 원신호를 상호상관함수를 이용하였을 경우에는 주기성이 강조됨과 동시에 위상 성분도 그대로 유지됨을 알 수 있어 불안정한 음성신호의 구간에서도 피치 시점을 검출하는데 좋은 결과를 얻을 수 있었다[5][6].

표 4-1 피치 변경율에 따른 SNR(화자1)

정확도(%)	
수동 방법	100
제안한 방법	97.5
오차	2.5

표 4-2 피치 변경율에 따른 SNR(화자2)

정확도(%)	
수동 방법	100
제안한 방법	98.5
오차	1.5

이렇게 하여 얻은 피치 시점검출의 결과를 표 4-1에 나타내었다. 표 4-1에서 볼 수 있듯이 수동으로 피치를 검출한 결과를 100%로 놓고, 제안한 방법을 사용하여 검출한 결과의 정확도를 비교하였을 때 97.5% 임을 볼 수 있다. 또한, 표 4-2에서는 98.5%의 정확도를 얻을 수 있었다.

5. 결론

음성은 인간이 사용하는 가장 기본적인 의사소통을 위한 수단이며 편리함과 경제상의 측면에서 다른 방법에 비하여 우수한 특성을 지닌다. 인간에 대한 음성신호의 정보전달 측면에 있어서는 신속성 및 용이함은, 음성을 중요한 정보교환의 매체로서 자리잡게 하였고 인간과 기계사이의 정보교환을 위한 MMI(Man-Machine Interface)에서 음성은 더욱 중요한 의미를 가진다.

음성신호처리 분야에서 음성신호의 기본 주파수를 정확히 검출할 수 있다면 인식의 정확도를 높일 수 있고, 음성합성을 할 때 자연성과 개성을 쉽게 변경하거나 유지할 수 있다. 위와 같은 피치검출의 중

요성 때문에 피치검출에 대하여 다양한 방법이 제안되었다.

본 논문에서는 음성신호의 분석에 있어서 기존의 자기상관함수법(Autocorrelation Function)은 주기성을 강조할 수 있다는 장점을 이용하면서 위상 성분을 보존하지 못한다는 단점을 보완할 수 있는 알고리즘을 제안했다.

실험결과 피치시점을 수동으로 찾은 경우와 비교하였을 때 약 98% 정도의 정확도를 얻을 수 있었다.

참고 문헌

- [1] L.R. Rabiner, and R.W. Schafer, *Digital Processing of speech Signals*, Prentice-Hall, 1978.
- [2] A.M. Kondoz, *Digital Speech -Coding for Low Bit Rate Communications Systems-*, John Wiley & Sons, 1994.
- [3] B.E. Caspers, and B.S. Atal, "Changing Pitch and Duration in LPC Synthesised Speech using Multipulse Excitation," *J. Acoust. Soc. Amer.*, Vol.73, No.1, pp.55, 1983.
- [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] 함명규, 이동기, 배명진, "음성선형 PCM파형에서 양자화 오차를 이용한 F1/F0을 검출" 제 10 회 신호처리 합동학술대회, 1997.
- [6] 함명규, 나덕수, 정찬중, 배명진, "비대칭 Weighting을 사용한 음성 피치 변경법" 전자공학 회 하계종합학술대회, 1998.