2000년도 한국음향학회 학술발표대회 논문집 제19권 제1(s)호

남녀 음성 변환 기술연구

최정규, 김재민, 한민수
한국정보통신대학원대학교 공학부

A Study On Male-To-Female Voice Conversion

Jung-Kyu Choi, Jae-Min Kim, Min-Su Han
School of Engineering , Information and Communications University
kyuro@icu.ac.kr

Abstract

Voice conversion technology is essential for TTS systems because the construction of speech database takes much effort. In this paper, male-to-female voice conversion technology in Korean LPC TTS system has been studied. In general, the parameters for voice color conversion are categorized into acoustic and prosodic parameters. This paper adopts LSF(Line Spectral Frequency) for acoustic parameter, pitch period and duration for prosodic parameters. In this paper, Pitch period is shortened by the half, duration is shortened by 25%, and LSFs are shifted linearly for the voice conversion. And the synthesized speech is post-filtered by a bandpass filter. The proposed algorithm is simpler than other algorithms, for example, VQ and Neural Net based methods. And we don't even need to estimate formant information. The MOS(Mean Opinion Socre) test for naturalness shows 2.25 and for female closeness, 3.2. In conclusion, by using the proposed algorithm, male-to-female voice conversion system can be simply implemented with relatively successful results..

I. 서론

음성변환(Voice Conversion)이란 화자의 개인성정보를 수정하거나 치환하는 기술로서 정의되는데 일반적으로 입력음성을 목적화자가 들리는 것처럼 변환하는 것을 말한다.[1] 최근 문서 음성 변환(Test-to-Speech)시스템의 급증하는 수요로 인해 그 중요성이 커지고 있는데 일반적인 문서 음성 변환 시스템은 한 화자의 음성 DB를 구축하고 무제한 합성을 하는데 음성 DB의 구축은 많은 시간과 노력이 필요하므로 하나 이상의 DB를 작성한다는 것은 매우 힘든 일이다. 따라서 대화시스템과 같이 구성된 음성 DB 이외의 음성을 출력하고자 하는 경우에는 음성 변환 기술이 필수적인 조건이 된다. 본 논문에서는 20대 남성화자의 DB로 구성된 LPC 합성기에서 동일연령층의 여성으로의 음성변환 기술에 관하여 연구하였는데 기존의 VQ나 신경망을 이용한 방법에 비해 연산량이 적고 메모리를 절약할 수 있는 음성변환 필터를 구현해 보았다. 제안된 알고리즘에 의해 합성된 합성음을 주관적인 음질평가 방법인 MOS 평가를 자연성과 음성변환 정도에 대해 실시하였다. II 절에서는 음성변환에 관한 일반적인 내용과 기존의 방법에 대해 알아보고 III 절에서는 남녀 음성의 특징과 이를 나타내는 특징 파라미터와 본 논문에서 제안한 알고리즘에 대해 살펴본다. IV 절에서는 음성변환 실험에 대한 주관적인 평가방법의 결과를 나타낸다.

II. 음성변환

음성변환을 위해 고려해야 할 화자의 개인성 요소는 크

게 음향학적인 요소와 운율적 요소로 나뉘는데 음향학적인 요소는 발성기관의 해부학적인 차이나 발성 기관의 조음 방법 차이에서 나타나는 포만트 주파수나 , 대역폭 등이 있으며 운율적 요소는 기본 주파수 궤적, 지속시간, 피치 등이 있다.[1] 일반적인 음성변환 시스템은 크게 분석부, 변환부, 합성부로 나뉘는데 분석부에서는 매 분석구간마다 변환을 수행할 특징 파라미터를 추출하고 변환부에서는 추출된 파라미터를 목적화자의 특징 파라미터로 변환시키며 합성부에서는 변환부에서 변환된 파라미터들을 이용하여 음성으로 재합성하는 일을 수행한다. 변환방법에 있어서 일반적으로 화자의 모델링을 위해서 VQ(Vector Quantization) 기반의 Codebook이 주로 사용되어져 왔다.[2] 하지만 이러한 방법은 효율적인 사상학습과 학습데이터의 선정에 어려움이 있고 양자화 오류가 발생한다. 이러한 단점들을 해결하기 위해 GMM(Gaussian Mixture Model)이나 신경망(Neural Network) 등에 의한 화자 모델링 방법이 제안되기도 하였으나 복잡한 연산을 수행하게 된다.[3][4]

III. 남녀 음성 변환

A.남성과 여성의 음성

(1) 피치주기

일반적으로 남성과 여성 음성의 가장 큰 특징은 피치주기의 차이라고 말하여 진다. [5] 피치주기는 아주 높은 톤의 아이나 여성의 경우 최저 2ms 에서 아주 낮은 톤의 남성의 경우 최고 25ms 를 나타낸다고 한다. [6] 본 연구에서는 20 대 중반의 음성 DB 를 갖고 동일 연령층의 여성음성으로 변환하는 것을 목표로 하기 때문에 20 대 중반 남성 3 명과 여성 3 명의 평균 피치 주기를 구해 보았는데 그 결과는 표 1 과 같다.

표 1. 20 대 중반 남녀의 평균 피치주기

남성		여성	
남성 1	8.750(ms)	여성 1	3.875(ms)
남성 2	9.500(ms)	여성 2	4.875(ms)
남성 3	9.038(ms)	여성 3	4.250(ms)
평균	9.230(ms)	평균	4.333(ms)

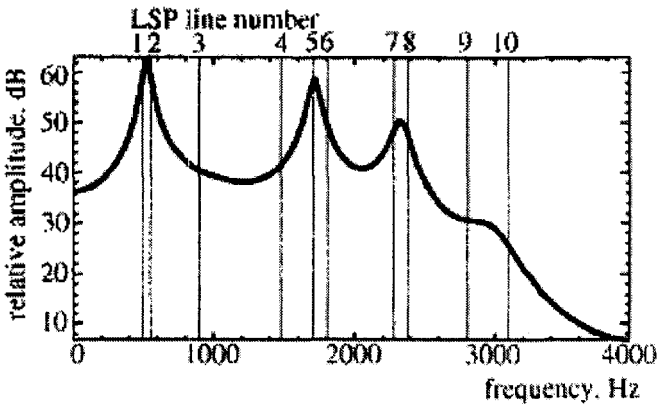
표 1 에 나타난 바와 같이 여성의 피치주기가 대략 남성의 피치주기의 반정도로 나타났다.

(2)지속시간

또한 여성음성의 특징이라 할 수 있는 것으로 일반적으로 여겨지는 것은 남성에 비하여 성도의 길이가 짧고 연약하기 때문에 음의 변화가 심하고 발음속도가 빠르다는 것이다.[5] 발음속도는 사람마다 다르고 말하는 사람의 상태에 따라서도 같은 사람이라도 다르기 때문에 이에 대한 정확한 데이터를 구하기는 어렵지만 본 논문에서는 남성음성을 여성음성에 좀더 가깝게 변화 시키기 위해 실험적으로 평균 지속시간을 남성의 75%정도로 하였다.

(3)포만트

일반적으로 남성 성도가 여성에 비해 큰 공동(cavity)으로 인해 포만트 주파수가 같은 모음일 경우 낮다고 알려져 있다.(Fant, 1975) [5] 본 연구에서는 포만트 정보를 이용하기 위해서 LSF (Line Spectral Frequency)를 채택하였는데 적은 수(10~14 차)의 LSF 값으로도 음성의 스펙트럼 포락 정보-포만트 정보와 대역폭을 포함하여-를 나타내고 조정하기 쉽기 때문이다.[7] 그림 1 은 LPC 스펙트럼과 10 차 LSF 의 관계를 나타낸다.



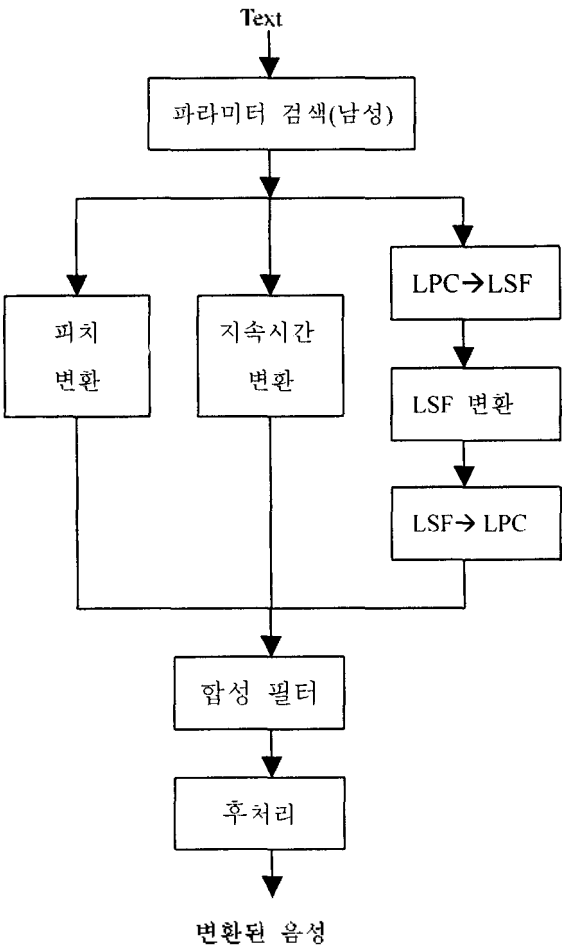
<그림 1. LPC 스펙트럼과 LSF>

본 연구에서는 음성변환을 위해 LSF 값을 linear 하게 shift 하였는데 이는 포만트 정보가 음소마다 다르고 같

은 음소일지라도 사람마다 다 다르기 때문에 모델링하기 힘들고 이는 복수의 음성 DB를 갖는 것과 마찬가지이기 때문이다.

B. 제안된 알고리즘

본 논문에서는 남녀 음성변환을 위해 채택한 파라미터는 앞의 III.A에서 설명한 바와 같이 피치주기, 지속시간, LSF이고 그림 2는 본 논문에서 제안된 알고리즘의 Block Diagram 이다.



<그림 2.제안된 알고리즘의 Block Diagram>

단계별로 살펴보면

- 파라미터 검색단계: 음성변환을 위한 파라미터들, 즉 피치주기, 지속시간, LSF를 프레임별로 검색한다.
- 음성변환단계: 검색된 파라미터들을 변환하는 단계로 피치주기는 0.5배, 지속시간 0.75배, LPC 계수를 LSF로 구하고 30Hz씩 linear하게 shift시킨다. shift된 LSF를 다시 LPC 계수로 계산한다.
- 합성단계: 변환된 파라미터로부터 음성을 합성한다.
- 후처리 단계: 고주파 영역과 저주파 영역에서의 noise를 제거하기 위해 BPF 한다.

IV. 실험 결과

음성변환의 정도와 자연성을 평가하기 위하여 주관적인 MOS 평가를 하였는데 우선 남녀 각각 6인을 선정하여 MOS 평가에 대한 기준을 상세히 설명하고 6문장에 대해 원래 남성음성의 합성음과 변환된 합성음을 random하게 배열하여 들려주고 평가하였다. 첫번째 평가에서는 들려주는 합성음이 남성에 가까운지 여성에 가까운지를 평가하는 실험이었는데 그 평가 기준은 표 2와 같다.

표 2. 음성변환정도 평가기준

평가 점수	평가 기준
1	완전 남성 음성이다
2	남성음성 같다
3	남성인지 여성인지 모르겠다
4	여성음성 같다.
5	완전 여성 음성이다

실험결과 원래 합성음은 1.59, 변환된 음성은 3.20으로 나타났는데 변환된 음성이 여성음성에 가까운 것을 나타냈으나 변환 정도에 있어서는 미흡하였다. 두 번째 자연성 평가에서는 원래 합성음 3.02, 변환음 2.25로 음질의 열화가 발생하는 것을 알 수 있었다.

V. 결론

본 연구에서는 기존의 VQ 기반의 음성변환 알고리즘과는 달리 20대 남성화자의 DB로 구성된 LPC 합성기에서 동일연령층의 여성으로의 음성변환 기술에 관하여 연구하였는데 기존의 연구에 비해 연산량이 적고 메모리를 절약할 수 있는 음성변환 필터를 구현해 보았다.

사용한 feature 들로는 남성과 여성의 가장 큰 차이를 나타내는 피치주기와 지속시간 그리고 스펙트럼 정보를 나타내는 LSF 를 이용하였다. 실험결과에서 나타나듯이 여성음성에 가까운 결과를 나타냈으나 그 정도에서는 미흡하였고 음질에서도 원래 음성에 비해 열화가 발생함을 나타내었다. 이 부분에 대해 음성변환 정도 및 음질 개선을 위한 연구가 지속적으로 행해지고 있다.

참고문헌

[1] 오영환 저, 음성언어정보처리, 홍릉과학출판사, 1998

[2] Abe, M., Nakamura, S., Shikano, K., Kuwabara, H. "Voice Conversion through Vector Quantization", Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on , 1988 , Page(s): 655 -658 vol.1

[3] Stylianou, Y.; Cappe, O.; Moulines, E., "Continuous Probabilitic Transform For Voice Conversion", Speech and Audio Processing, IEEE Transactions on Vol.: 6 2 , March 1998 , Page(s): 131 -142

[4] Nakamura, S.; Shikano, K., "Speaker Adaptation Applied to HMM and Neural Networks",Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on , 1989 , Page(s): 89 -92 vol.1

[5] Tielen, Mirijam Thecla Jacoba, "Male and Female Speech (An Experimental Study of sex-related voice and pronunciation characteristics", 1992

[6] L.R. labiner, R.W.Schafer, "Digital Processing of Speech Signals ", Prentice Hall, 1978.

[7] McLoughlin, I.V.; Chance, R.J., "LSP-based speech modification for intelligibility enhancement", Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on Vol. 2 , 1997 , Page(s): 591 -594 vol.2