

# 스펙트럼包絡 修正法에 의한 韓國語母音의 聲質變換에 관한 研究

이 기 영, 최 창 석\*

관동대학교 전자통신공학과  
\*영지대학교 정보통신공학과

## A Study on Voice Quality Conversion for Korean Vowels Using Spectrum Envelope Correction Method

Ki Young Lee, Chang Seok Choi\*

Dept. Electr. Comm. Eng. in Kwan Dong Univ.  
\*Dept. Inform. Comm. Eng. in Myong Ji Univ.

### 요 약

본연구에서는 스펙트럼포락의 변경에 의해 음성의 개인성이 변환될 수 있다는데 착안하여 스펙트럼포락 수정법에 의한 聲質變換에 관하여 연구하였다. 실험에서는 남성화자와 여성화자가 각각 발성한 한국어 모음을 대상으로 스펙트럼포락 수정법을 적용하여 스펙트로그램과 청취시험을 비교 검토함으로써 聲質變換의 성능을 확인하였다.

### 1. 서 론

음성은 포함된 정보를 크게 두가지로 나누면 언어정보와 비언어적 정보로 나눌 수 있다. 비언어적 정보란 음성의 의미 전달하는 언어적 정보 이외의 개인성, 정서성 및 심리적 인상 등의 여러가지 정보를 의미하며, 이러한 비언어적 정보를 聲質(voice quality)이라는 용어로 표현하고 있다 [9]. 이 성질이라는 음성의 정보는 음성의 여러가지의 특징 파라미터와 많은 관계가 있으며, 이관계를 조사한 연구와 음성에 포함된 정보 중에서 주로 비언어적 정보를 처리하는 성질변환의 연구에 관심이 집중되고 있다.

음성인식분야에서 최종적으로 필요로 하는 기술중의 하나는 불특정화자를 대상으로 한 음성인식이지만, 아직까지 불특정화자의 음성인식에 대해서는 아직까지 실험시점이 연구되고 있는 실정이다. 불특정화자 음성인식의 문제는 음성을 발생시키는 화자가 바뀔 때 따라 성도스펙트럼 성분이나 피치주기성분 등으로 나타나는 개인성이 서로 다르기 때문에 발생하므로 음성의 개인성을 정규화하는 것을 목표로 화자 적응을 이용한 음성인식이 연구되고 있다. 음성의 개인성을 변화시키는 면에서 공통적 방법인 음성합성분야의 성질변환에 관한 연구는, 자동번역기로 하여금 번역된 언어음성을 사용자의 음성으로 변환시켜 출력하도록 한다든가, 멀티미디어의 휴먼인터페이스화 등 그 응용성의 다양함 때문에 많은 관심이 집중되고 있다.

음성의 식별을 위한 특징의 추출은 운율적 정보와 함께 음향적 정보가 포함되어 있기 때문에 어려운 과제이다. 대표적 운율적 정보는 피치주파수의 궤적으로 나타나며, 음향적 정보는 포인트의 궤적으로 나타난다. 성질변환을 수행하기 위해서는 바로 이상과 같은 운율정보와 음향정보를 모두 변경해야 가능해진다. 그러나 아직까지 운율정보를 완전히 파악하기는 어려운 문제로 남아 있으며, 현재에는 음향정보를 대상으로 한 성질변환의 연구가 주로 이루어지고 있다.

운율정보인 피치를 변환하는 방법의 연구는 1982년 Senef [1], 1983년 Atal 등 [2]이 발표하였으며, 성도특성으로 나타나는 포인트의 변환방법의 연구는 1987년 Goncharoff 등 [3], 시간간격이변환방법의 연구는 1984년 Griffin 과 Lim [4], 시간길이, 주파수 및 피치를 모두 변환하는 방법의 연구는 1986년 Quatieri 등 [5]이 제시하였다. 성질변환에 관한 연

구는 1988년 Abe [6] 가 화자 적응기법인 사상코드북을 이용한 방법으로 시작하였으며, 이방법의 학습과정에서는 스펙트럼, 파워 및 피치주파수를 이용하였고, 합성과정에서는 L-PC 보코더를 이용하였다. 1990년 Ueda [7]는 피치변경시 발생하는 스펙트럼편곡을 감소시켜 보다 자연스런 음성으로 성질변환하는 방법을 제안하였다. 1991년 Tubach [8]는 PSO LA와 LMR/ITW 를 이용하여 더욱 고품질의 음성으로 성질변환하는 방법을 발표하였다.

본연구에서는 음성의 스펙트럼포락으로 나타나는 음향정보의 처리와 운율정보인 피치의 변경에 의해 성질변환하는 방법에 관하여 연구하였다. 이방법에서는 화자마다 서로 다르게 나타나는 스펙트럼포락의 변동폭은 개인성을 변화시키는 주요인인으로 존재한다는 점에 착안하여, 화자에 따라서 다른 성도스펙트럼포락 사이의 변동폭을 실험모델로 가정하고 이를 이용하여 스펙트럼포락을 수정하는 스펙트럼포락 수정법(SPECOM)에 의하여 화자의 개인성을 표준화자의 것으로 성질변환하는 방법을 제안하였다. 실험에서는 남성화자와 여성화자가 각각 발성한 한국어 모음을 대상으로 스펙트럼포락 수정법을 적용하여 스펙트로그램과 청취시험을 비교 검토함으로써 성질변환의 성능을 확인하였다.

### 2. 성질변환의 원리

성질변환의 기술적기초가 되는 분야는 음성의 분석합성에 있다. 이것은 자연음성을 분석하여 성도의 진동이나 성도에서의 공진동, 음성의 성질을 결정하는 특징파라미터를 각각 독립으로 시계열형태로 추출하여 그 특징파라미터에 따른 합성장치를 동작시킨다.

먼저, 종래의 전형적인 방법으로 선형예측분석에 의한 분석합성을 설명한다. 그림 1 은 잔차구동방식이라고도 불리우며, 합성장치의 음원으로서 분석시 입력음성으로 부터 분리된 잔차파형을 보존했다가 이것을 이용하는 방법이다.

음성의 높이를 변화시키기 위해서는 유성음부의 잔차파형의 주기를 변경한다. 또한 성도의 특징을 변경하는 것은  $H(w)$ 의 모양을 변형하여 새로운 성도특성을 지닌 필터로 구성한다. 변경된 잔차파형을 변경한 성도필터에 입력하면, 성질변환된 음성파형을 얻는다.

### 3. 제안된 성질변환

입의의 화자가 발성한 음성을 정규화하기 위해서는 음성에 포함된 언어적인 정보외에 비언어적인 정보를 추출하여 표준화자 음성의 비언어적 정보로 정규화할 필요가 있다. 본연구에서는 캡스텀해석을 이용하여 음성에 포함된 비언어적 정보인 개인성을 추출하여 표준화자의 것으로 수정하는 방법으로 화자정규화를 수행하고자 한다. 다음의 그림 2 는 스펙트럼포락 수정법을 이용한 화자정규화의 원리

도를 보이고 있다.

3-1. 스펙트럼포라 수정법 I

캡스트럼 해석에서 저역 quefrancy 성분을 푸리에 변환하면 성도특성을 지닌 스펙트럼포라를 추출할 수 있다. 임의의 화자 A 가 발생한 음성  $X^{(A)}$  을 표준화자 B 가 발생한 스펙트럼포라  $X^{(B)}$  으로 정규화할 수 있는 선형모델 H 가 있다고 가정하면 대수화된 스펙트럼영역에서 화자 A 의 스펙트럼포라  $X^{(A)}$  는 표준화자 B 의 스펙트럼포라  $X^{(B)}$  가 대수화된 선형모델  $\log|H|$  만큼 변동된 것으로 생각할 수 있다. 즉, 표준화자 B 의 스펙트럼포라  $X^{(B)}$  로 정규화된 스펙트럼포라  $X^{(A \rightarrow B)}$  를 다음 식과 같이 나타낼 수 있다.

$$\log|X^{(A \rightarrow B)}| = \log|X^{(A)}| + \log|H| \quad (1)$$

여기서, 대수화된 선형모델  $\log|H|$  은 화자 A 와 화자 B 의 두 스펙트럼포라 사이의 변동폭이다.

임의의 화자의 성도특성을 표준화자의 성도특성으로 정규화하는 방법으로 임의화자가 발생하는 음성의 캡스트럼에서 추출되는 스펙트럼포라와 표준화자의 캡스트럼에서 추출된 스펙트럼포라의 변동폭을 구하고 입력되는 임의화자 음성의 스펙트럼포라에 변동폭 만큼 스펙트럼포라를 수정하면 임의화자의 성도특성을 표준화자의 성도특성으로 정규화할 수 있다. 다음은 스펙트럼 포라의 변동폭을 나타내는 대수화된 선형모델  $\log|H|$  를 구하는 식이다.

$$\log|H| = \log|X^{(B)}| - \log|X^{(A)}| \quad (2)$$

또한, 이식은 캡스트럼 계수가 대수화된 스펙트럼영역에서 갖는 선형관계를 나타낸 것으로 다음 절에서 다루게 될 캡스트럼을 이용한 피치변경방법에서도 유효하게 적용된다.

3-2. 스펙트럼포라 수정법 II

본절에서는 화자의 정규화를 위한 두 번째 단계인 피치주기의 변경에 대하여 서술하고자 한다. 음성의 캡스트럼 해석은 음성  $s(t) \langle \cdot \rangle S(k)$  가 준주기적인 음원  $g(t) \langle \cdot \rangle G(k)$  에 의해 구동되는 성도조율 등가필터  $x(t) \langle \cdot \rangle X(k)$  의 응답이 가장한 것이며, 대수를 이용하여 음성을 성도스펙트럼 성분과 피치주기성분으로 분리하였다.

$$c(\tau) = F^{-1}[\log|X(k)| + F^{-1}[\log|G(k)|] \quad (3)$$

여기서,  $F^{-1}$  는 역푸리에 변환이며, 우변의 좌항으로 나타나는 성도스펙트럼 성분은 0-4[ms] 범위의 저역 quefrancy 영역에 존재하며, 우항에 나타나는 음원의 피치주기는 4[ms] 이상의 고역 quefrancy 영역의 분주리검출에 의해 추출할 수 있다.

캡스트럼해석결과로부터 다시 음성을 생성하고자 할 때는 quefrancy 상의 저역영역을 성도스펙트럼성분으로 취한

다음 역컨벌루션을 통해 임펄스특성을 구하고, quefrancy 상의 고역영역에 나타나는 음원성분에 해당하는 피치주기필스와 이전에 구한 임펄스특성과 컨벌루션하면 음성을 얻을 수 있다. 여기서, 피치주기를 변경하고자 할 때는 최종의 컨벌루션을 하기 전에 음원의 피치주기를 변경시켜 주면된다. 그런데 여기서 피치주기필스는 컨벌루션에 의해 성도특성인 임펄스특성으로 하여금 피치주기성분을 포함한 위상으로의 역할을 하게 된다.

본연구에서는 시간축상의 피치변경 방법으로 기존의 명주파/제거법을 이용하였으며, 시간축상에서 피치주기가 변경된 음성의 위상을 추출하여 피치변경된 음성을 생성하는데 사용하였다. 즉, 식(1)로부터 얻은 성도특성이 정규화된 스펙트럼포라  $X^{(A \rightarrow B)}(k)$  을 변경된 성도특성 등가필터로 하고 시간축상에서 피치변경이 이루어진 임의 화자의 음성의 스

펙트럼  $S^M(k)$  의 위상을 이 필터에 통과시켜 성도특성과 피치주기특성이 정규화된 음성  $s^{(A \rightarrow B)}(t)$  을 얻는다. 이를 식으로 나타내면 다음과 같다.

$$s^{(A \rightarrow B)}(t) = F^{-1}[ X^{(A \rightarrow B)}(k) \angle S^M(k) ] \quad (4)$$

그런데 식(2) 양변의 대수를 벗기면 성도특성을 정규화할 수 있는 선형모델 H 를 다음과 같이 구할 수 있다.

$$H = \frac{X^{(B)}}{X^{(A)}} \quad (5)$$

여기서, 선형모델 H 는 화자 A 와 화자 B 의 두 스펙트럼포라 사이의 변동비율이며,  $X^{(A)}$  와  $X^{(B)}$  는 캡스트럼해석에 의한 스펙트럼포라가 아니라 LPC 해석에 의해 추정된 성도 스펙트럼이다. 따라서, 선형모델 H 를 구하는 과정이 캡스트럼해석에 의해 스펙트럼포라를 구할 때 보다 간단해진다.

또한, 식(1) 양변의 대수를 벗기면 성도특성이 정규화된 스펙트럼포라  $X^{(A \rightarrow B)}(k)$  를 구할 수 있으며 이를 역푸리에 변환하면 성도특성과 피치주기특성이 표준화자의 것으로 정규화된 음성을 얻을 수 있다.

$$s^{(A \rightarrow B)} = F^{-1}[ X^{(A)} H ] \quad (6)$$

여기서,  $X^{(A)}$  는  $X^{(A)}$  와 달리 성도특성을 나타내는 캡스트럼해석이나 LPC 해석에 의한 스펙트럼포라가 아니라 화자 A 가 발생한 음성을 피치주파/제거법에 의해 시간축상에서 피치변경한 후 푸리에 변환한 스펙트럼이다.

그림 3 은 식(5)의 LPC 스펙트럼해석에 의해 정규화 선형모델 H 를 구하는 경우의 스펙트럼포라 수정법을 이용한 화자정규화의 구성도를 보인다.

4. 實驗 및 考察

본 연구의 실험데이터는 한국의 남성화자와 여성화자가 발생한 한국어 모음으로 하였으며, 한국어 모음은 "아",

"어", "오", "우", "으", "이", "에(애)" 등의 7 개의 기본모음으로 하였다. 샘플링 주파수는 10kHz 이며, 창길이는 256 샘플로 하였고, 프레임 주기는 100 샘플로 하였으며, 1-0.95의 권강조필터를 거쳐 14차의 LPC 해석을 하였다.

그림 4 는 가장 왼쪽의 파형부터 여성화자의 원음성파형이고 차례로 여성음성의 피치주기에 의해 변환된 파형, 스펙트럼포라 수정법 I 에 의해 여성음성을 남성음성으로 변환한 파형, 스펙트럼포라 수정법 II 에 의해 여성음성을 남성음성으로 변환한 파형 및 남성화자의 원음성파형을 각각 나타내고 있다.

그림 4 의 스펙트로그램을 관찰해 보면, 첫 번째 파형과 두 번째 파형의 스펙트로그램은 그골과 언덕부분이 유사하나, 각 음성을 청취하였을 때 음의 고저가 다르다는 점을 확인하였다. 또한, 세 번째 파형과 네 번째 파형의 스펙트로그램은 다섯 번째 파형의 스펙트로그램과 유사하며, 각 음성을 청취하였을 때에도 거의 유사한 것으로 확인하였다. 이를 확인하기 위하여 각 음성과 변환음성을 녹음기에 저장하였다.

5. 結論

본연구에서는 임의 화자음성의 개인성을 표준화자의 것으로 변환시키는 성질변환방법에 관하여 연구하였으며, 그 방법으로 스펙트럼포라 수정법을 제안하였고, 한국어 모음을 대상으로 실험을 수행하여 스펙트로그램과 청취시험을 통하여 비교 검토한 결과, 이방법에 의해 음성의 개인성을 변환할 수 있음을 확인하였다.

References

[1] S.Seneff, "System to independently modify excitation and/or spectrum of speech waveform without explicit pitch extraction," IEEE Trans. Acoust., Speech Signal Processing, Vol. ASSP-30, No. 4, pp. 566-578, Aug. 1982  
 [2] B.E.Caspers, B.S.Atal, "Changing pitch and duration in LPC synthesized speech using multipulse excitation," J. Acoust. Soc. Amer., suppl., Vol. 73, No. 1, pp. S5, Spring, 1983  
 [3] M.A.Jasiuk, V.Geocharoff, J.N.Danoulakis, "Improved speech modification method," ICASSP 87, pp.1465-1468, 1987  
 [4] D.W.Griffin, J.S.Lim, "Signal estimation from modified short-time fourier transform," IEEE Trans. Acoust., Speech Signal Processing, Vol. ASSP 32, No. 2, pp. 236-243, Apr. 1984

[5] T.F.Quatieri, R.J.McAulay, "Speech transformations based on a sinusoidal representation," IEEE Trans. Acoust., Speech Signal Processing, Vol. ASSP-34, No. 6, pp. 1449-1464, Dec. 1986  
 [6] M. Abe, S. Nakamura, K. Shikano, H. Kuwabara, "Voice conversion through vector quantization," ICASSP 88, pp. 655-658, 1988  
 [7] T. Takagi, T. Umeda, "Voice quality conversion with correction of spectral distortion by pitch manipulation, and its subjective evaluation," 電子情報通信學會論文誌 A Vol. J73-A, No. 3, pp. 387-396, Mar. 1990  
 [8] H. Valbret, E. Moulines, J.P. Tubach, "Voice transformation using PSDLA technique," EUROSPEECH 91, Genova, Italy, pp. 345-348, Sep. 1991  
 [9] T. Takagi, "Voice quality conversion," 텔레비전學會誌 Vol. 47, No. 12, pp. 28-32, 1993

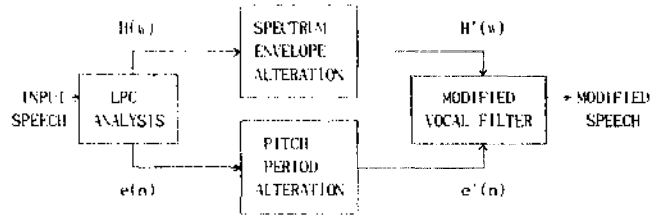
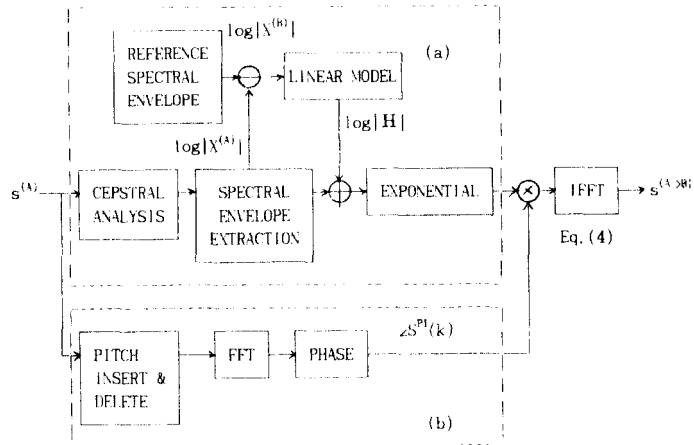


그림 1. 일반적인 성질변환의 구성도  
 Fig. 1. Block diagram of general voice conversion



(a) Spectrum envelope correction  
 (b) Pitch alteration

그림 2. 스펙트럼포락 수정법 1 에 의한 성질변환의 원리도  
 Fig. 2. Principle diagram of voice conversion using spectrum envelope correction 1

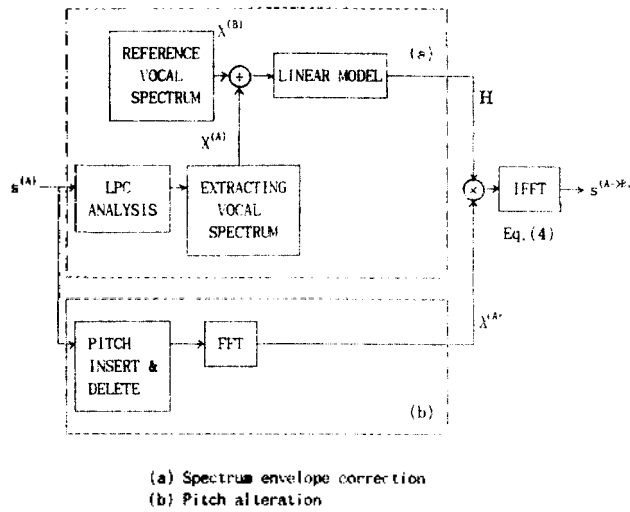


그림 3. 스펙트럼포락 수정법 II 에 의한 성질변환의 구성도  
 Fig. 3. Block diagram of voice conversion using spectrum envelope correction method II

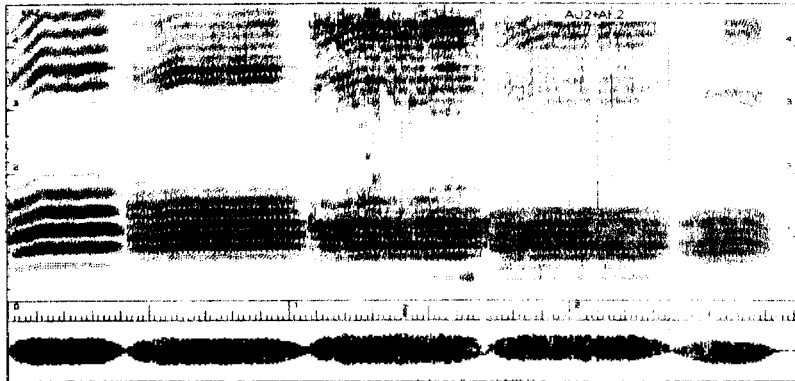


그림 4. 변환음성과정의 비교  
 Fig. 4. Comparison of modified speech waveforms