한국음향학회지 제20권 제3호 pp. 15~23 (2001)

선형다변회귀모델과 LP-PSOLA 합성방식을 이용한 음성변환

Voice Conversion Using Linear Multivariate Regression Model and LP-PSOLA Synthesis Method

권 홍 석*, 배 건 성*
(Hong Seok Kwon*, Keun Sung Bae*)

*경북대학교 전자·전기공학부

(접수일자: 1999년 1월 12일; 수정일자: 2000년 7월 26일; 채택일자: 2001년 3월 20일)

본 논문에서는 임의의 사람이 발성한 음성을 마치 다른 사람이 발성한 것처럼 들리도록 하는 음성변환 기술에 대하여 설명하고, 화자간의 성도 특성과 여기신호 특성 파라미터 변환을 독립적으로 수행하기 위한 변환방법을 실험한다. 성도 특성 파라미터 변환은 입력되는 음성신호에서 LPC (Linear Predictive Cofficient)켑스트럼을 추출하여 선형다변회귀모델에 적용하여 수행하고, 여기신호 특성 파라미터 변환은 잔차신호를 추출하여 LP-PSOLA (Linear Predictive-Pitch Synchronous Overlap and Add) 합성방식을 이용한 화자간의 평균 피치주기 변환으로 수행된다. 실험결과는 선형다변회귀모델과 LP-PSOLA 합성방식을 이용하여 변환된 음성이 대상화자의 음성에 유사함을 보여준다.

핵심용어: 음성변환, LP-PSOLA, 선형다변회귀, 피치주기, DTW (Dynamic Time Warping)
투고분야: 음성처리 분야 (2,4)

This paper presents a voice conversion technique that modifies the utterance of a source speaker as if it were spoken by a target speaker. Feature parameter conversion methods to perform the transformation of vocal tract and prosodic characteristics between the source and target speakers are described. The transformation of vocal tract characteristics is achieved by modifying the LPC cepstral coefficients using Linear Multivariate Regression (LMR). Prosodic transformation is done by changing the average pitch period between speakers, and it is applied to the residual signal using the LP-PSOLA scheme. Experimental results show that transformed speech by LMR and LP-PSOLA synthesis method contains much characteristics of the target speaker.

Keywords: *Voice conversion, LP-PSOLA, Linear multivariate regression, Pitch period, DTW*
Ask subject classification: *Speech signal processing (2,4)*

I. 서 론

음성변환이란 임의의 사람이 발성한 음성을 마치 다른 사람이 발성한 것처럼 들리도록 하는 것을 말한다. 이를 TTS (Text-to-Speech) 변환 시스템과 같이 미리 구축된 데이터베이스를 이용하는 합성기에 적용함으로써 화자마다 데이터베이스를 구축하지 않고 후처리과정으로 음성변환을 수행하여 원하는 음색으로 합성할 수 있다. 또, 음성인식 시스템의 경우에는 화자간 변화를 줄이기 위한 전처리과정으로 화자적응 기술에 이용될 수도 있다. 초기의 음성변환은 벡터양자화 (VQ: Vector Quantization)를 통해 각 화자의 코드북간 대응관계를 이용하는 방법으로 시도되었다. 이 방법은 벡터양자화를 통한 클러스터링

책임저자: 권홍석 (hong@mmirl1.knu.ac.kr)
702-701 대구광역시 북구 산격동 1370번지
경북대학교 전자·전기공학부
(전화: 053-840-8627; 팩스: 053-950-5590)

(clustering) 과정을 원화자와 대상화자 각각의 성도 특성 파라미터에 적용한다. 이렇게 하여 만들어진 두개의 VQ 코드북을 이용하여 원화자의 중심값 (centroid)과 대상화자의 중심값 사이의 관계를 발생 빈도수로 표현되는 대응 코드북 (mapping codebook)을 만들어 음성변환을 수행하였다[1,2]. 그러나 이 방법은 변환된 특징 파라미터가 이산적 (discrete)이기 때문에 음성의 다양한 변화를 표현할 수 없다는 단점이 있다.

본 논문에서는 이런 단점을 보완하고자 화자간 성도 특성의 관계를 설명하는 대응 코드북 대신에 선형 변환식을 사용하고, 음성발생모델에서 여기신호 특성 및 성도 특성을 변환하는 음성변환 방법을 제시한다. 이 방법은 LP-PSOLA (linear predictive-pitch synchronous overlap and add)합성방식이 성도 특성과 여기신호 특성 파라미터를 추출한 후 여기신호에 시간축 변환 (time-scale modification)이나 피치주기 변환 (pitch-scale modification)을 수행한다는 특징을 이용한 것이다. 즉, 각각의 특성 파라미터를 독립적으로 변환한 다음 성도 특성 파라미터를 표현하는 필터에 변환된 여기신호를 통과시켜 음성을 합성함으로써 음성변환을 수행한다. 제시한 방법은 크게 훈련과정과 변환-합성과정으로 구분된다. 훈련과정에서는 원화자와 대상화자의 훈련데이터에서 LPC (linear predictive coefficient)켑스트럼을 추출하여 변환된 음성과 대응되는 대상화자 음성사이의 LPC 켑스트럼간 선형 변환식을 선형다변회귀모델에 적용하여 추정하고 운율정보 변환을 위해 화자간의 평균 피치주기 비를 구한다. 변환-합성과정에서는 원화자의 입력음성에서 추출한 LPC 켑스트럼을 훈련과정에서 추정한 선형 변환식을 이용하여 변환하고, 원화자의 잔차신호를 LP-PSOLA 합성방식에 적용하여 대상화자의 피치주기를 갖도록 변환하면서 합성한다.

본 논문의 구성은 다음과 같다. 먼저 II 장에서는 제시한 음성변환 알고리즘을 훈련과정과 변환-합성과정으로 구분하여 설명하고, 선형다변회귀모델을 이용한 선형 변환식을 추정하는 과정과 평균 피치주기 비를 구하는 과정을 설명한다. III장에서는 LP-PSOLA 합성방식을 이용한 음성변환 과정을 설명한다. IV장에서는 음성변환 알고리즘에 대한 실험 결과를 제시·분석하고 마지막으로 V 장에서 결론을 맺는다.

II. 음성변환 알고리즘과 선형다변회귀모델

음성발생모델의 관점에서 보면 음성신호는 여기신호가 성도 특성을 나타내는 필터를 통과함으로써 발생되는 신호로 볼 수 있다. 성도 특성은 사람의 성도를 전극 (all-pole) 필터로 가정하고 음성신호에 선형예측분석을 적용하여 추정한 LPC 켑스트럼으로 표현할 수 있다. 또, 음성신호를 그 필터에 역으로 통과시킴으로써 얻게 되는 잔차신호는 여기신호 특성을 잘 나타낸다. 따라서 음성발생모델에 잘 부합되는 LPC 켑스트럼을 성도 특성 파라미터로 사용하고 잔차신호를 여기신호 파라미터로 이용한다.

2.1. 음성변환 알고리즘

음성변환 알고리즘은 훈련과정과 변환-합성과정으로 구성된다. 훈련과정에서는 원화자와 대상화자의 훈련데이터에서 LPC 켑스트럼과 잔차신호를 추출한다. 이때 원화자와 대응되는 대상화자의 LPC 켑스트럼 사이에는 선형 관계가 성립한다고 가정하고 서로 대응되는 LPC 켑스트럼쌍을 선형다변회귀모델에 적용하여 선형 변환식을 추정한다. 그리고 화자간 평균 피치주기 비를 구하여 운율정보 변환 관계를 나타낸다. 그림 1은 훈련과정에서 선형 변환식을 추정하는 과정을 단계별로 보여주고 있다.

단계 1 : 원화자와 대상화자의 훈련데이터 음성에서 일정한 길이를 갖는 프레임 단위로 LPC 켑스트럼을 추출한다.

단계 2 : 원화자의 LPC 켑스트럼으로 VQ 코드북을 생성하고 원화자와 대상화자의 LPC 켑스트럼을 DTW (Dynamic Time Warping)로 서로 같은 음소가 대응되도록 시간 정렬하고 대응되는 두 훈련데이터간의 프레임 갯수를 같게 만든다.

단계 3 : 시간 정렬된 원화자의 LPC 켑스트럼을 VQ 코드북으로 부호화하여 동일한 코드워드에 속하는 원화자와 대응되는 대상화자의 모든 LPC 켑스트럼쌍을 분류한다.

단계 4 : 각 코드워드에 속하는 모든 LPC 켑스트럼쌍을 선형다변회귀모델로 선형 변환식을 추정한다.

변환-합성과정은 분석단계, 변환단계 그리고 합성단계의 세가지 단계를 통하여 수행된다. 첫번째 분석단계에서는 입력되는 음성에서 LPC 켑스트럼과 잔차신호를 추출한다. 이때 각 프레임은 PSOLA 합성방식에서 사용되는 분석 피치마크 (analysis pitch-mark)를 기준으로 구한다. 두번째 변환단계에서는 원화자의 LPC 켑스트럼을 선형 변환식으로 변환하고, 잔차신호는 일반적인 시간영역의 PSOLA 합성방식에 적용하여 대상화자의 평균 피치주

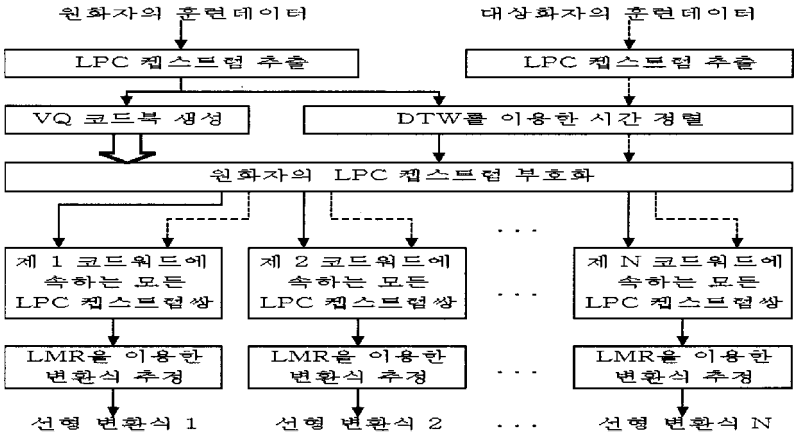


그림 1. 선형 변환식 추정 과정
Fig. 1. Procedure for the estimation of linear mapping.

기를 갖도록 변환시킨다. 마지막으로 합성단계에서는 변환된 LPC 켑스트럼에서 구한 LPC 계수와 변환된 잔차신호를 이용하여 합성함으로써 음성을 변환한다.

2.2. 선형다변회귀모델을 이용한 선형 변환식 추정

임의의 한 변수에 대한 정보를 이용하여 다른 변수의 변화를 예측할 때 다른 변수에 영향을 주는 변수를 독립변수 (independent variable)라 하며, 독립변수의 영향을 받는 변수를 종속변수 (dependent variable)라 한다. 회귀분석은 독립변수로부터 종속변수를 예측하기 위하여 회귀 방정식 (regression equation)이라는 두 변수 사이의 구체적인 함수관계를 규명하는데 이용되는 통계적 분석방법이다. 이때 두 변수 사이의 관계를 선형이라고 가정하고 분석하는 방법을 선형회귀 (linear regression)라고 부르며 특히 다수의 변수를 가지는 선형회귀를 선형다변회귀라고 한다[4].

본 논문에서는 원화자의 M차 LPC 켑스트럼을 독립변수로, 대응되는 대상화자의 M차 LPC 켑스트럼을 종속변수로 보고 화자간 LPC 켑스트럼 사이의 관계를 선형다변회귀를 이용하여 규정지으려 한다. 이때 훈련과정에서 사용된 음성을 벡터양자화하여 각 클러스터마다 선형 변환식을 추정함으로써 전체 자승오차를 줄일 수 있다. 그러나 변환-합성 과정에서의 코드북 검색시간을 고려하여 코드북 크기는 적당하게 하여야 한다.

원화자의 임의의 클러스터에 속하는 LPC 켑스트럼열과 대응되는 대상화자의 LPC 켑스트럼열은 식 (1)과 같이 나타낼 수 있다. 이때 모든 선형 변환식이 각각의 클러스터마다 독립적으로 추정되기 때문에 각 클러스터를 구분하는 인덱스는 편의상 생략하기로 한다.

$$\{C^s\}=\begin{bmatrix} C_1^s \\ C_2^s \\ \vdots \\ C_N^s \end{bmatrix}=\begin{bmatrix} c_{11}^s & c_{12}^s & \cdots & c_{1M}^s \\ c_{21}^s & c_{22}^s & \cdots & c_{2M}^s \\ \vdots & \vdots & \vdots & \vdots \\ c_{N1}^s & c_{N2}^s & \cdots & c_{NM}^s \end{bmatrix}$$

$$\{C^t\}=\begin{bmatrix} C_1^t \\ C_2^t \\ \vdots \\ C_N^t \end{bmatrix}=\begin{bmatrix} c_{11}^t & c_{12}^t & \cdots & c_{1M}^t \\ c_{21}^t & c_{22}^t & \cdots & c_{2M}^t \\ \vdots & \vdots & \vdots & \vdots \\ c_{N1}^t & c_{N2}^t & \cdots & c_{NM}^t \end{bmatrix} \quad (1)$$

여기서 $\{C^s\}$, $\{C^t\}$는 원화자와 대상화자의 LPC 켑스트럼열을 말하며 N은 LPC 켑스트럼열의 갯수를, M은 LPC 켑스트럼의 차수를 나타낸다.

원화자와 대상화자 LPC 켑스트럼의 각 차수 평균을 식 (2)처럼 나타내면 평균 벡터는 식 (3)과 같이 된다. 이때 두 화자의 평균을 뺀 벡터 사이에는 식 (4)와 같은 선형 변환식이 성립하게 된다[3].

$$b_k^s=\frac{1}{N}\sum_{i=1}^{N}c_{ik}^s\ ,\quad b_k^t=\frac{1}{N}\sum_{i=1}^{N}c_{ik}^t \quad (2)$$

$$B^s=[\ b_1^s\ b_2^s\cdots b_M^s],\ B^t=[\ b_1^t\ b_2^t\cdots b_M^t] \quad (3)$$

$$\widetilde{C}^t=\widetilde{C}^s A \quad (4)$$

여기서 ~ 표기는 각 화자의 LPC 켑스트럼에서 평균을 뺀 LPC 켑스트럼을 의미하며, B^s, B^t 는 M차 평균 벡터이고 A는 화자간의 성도 특성 변환 관계를 설명하는 선형 변환식으로 $M\times M$ 행렬이 된다.

훈련에 사용된 모든 LPC 켑스트럼에 대하여 변환된 LPC 켑스트럼과 대상화자의 LPC 켑스트럼간 자승오차는 식 (5)와 같이 표현된다. S값이 최소가 되도록 A행렬을 추정하기 위해 $1 \le l \le M$, $1 \le m \le M$ 의 범위를 가지는 모든 정수 l, m에 대해서 $\partial S/\partial a_{lm} = 0$을 풀면 식 (6), (7)과 같은 방정식이 구해지며 그 해는 식 (8)이 된다. 식에서 윗첨자 T는 전치행렬을 의미하며 원화자의 i번째 LPC 켑스트럼은 식 (9)를 이용하여 대상화자의 LPC 켑스트럼으로 변환할 수 있다.

$$S = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\tilde{c}_{ij}^{t} - \sum_{k=1}^{M} \tilde{c}_{ik}^{s} a_{kj} \right)^2 \tag{5}$$

$$\sum_{k=1}^{M} a_{km} \sum_{i=1}^{N} \tilde{c}_{il}^{s} \tilde{c}_{ik}^{s} = \sum_{i=1}^{N} \tilde{c}_{il}^{s} \tilde{c}_{im}^{t} \tag{6}$$

$$\tilde{C}^{sT} \tilde{C}^{s} A = \tilde{C}^{sT} \tilde{C}^{t} \tag{7}$$

$$A = \left(\tilde{C}^{sT} \tilde{C}^{s} \right)^{-1} \tilde{C}^{sT} \tilde{C}^{t} \tag{8}$$

$$C_i^t = \tilde{C}_i^s A + B^t \tag{9}$$

2.3. 피치주기 변환 비의 추정

음성변환에서는 화자간의 성도 특성 파라미터 외에도 운율정보의 변환도 고려하여야 한다. 음성의 운율적인 특징에는 여러가지가 있지만 본 논문에서는 각 화자의 평균 피치주기를 추출하여 그 비를 LP-PSOLA 합성방식에 적용하여 화자간의 평균 피치주기를 변환한다. 이는 식 (10)을 이용하여 구한 평균 피치주기 비인 β를 원화자의 각 프레임 피치주기에 곱함으로써 이루어지며, 이때 p_{ave}^{s}, p_{ave}^{t}는 각각 원화자와 대상화자의 평균 피치주기를 나타낸다.

$$\beta = \frac{p_{ave}^{s}}{p_{ave}^{t}} \tag{10}$$

III. LP-PSOLA 합성방식을 이용한 음성변환

PSOLA 합성방식은 음성파형을 그대로 연결하는 합성방식으로 피치를 맞추어 분석하고 합성하므로 운율 조절이 용이하며, LPC 계열의 합성방식보다 명료도, 자연성의 향상을 가져올 수 있다. 일반적으로 PSOLA 합성방식은 다음과 같은 3단계로 나뉘어 수행된다. 첫째로 원음성을 단구간 분석신호 (short-time analysis signals)로 나누는 분석단계, 둘째로 추출된 각 단구간 분석신호를 단구간 합성신호 (short-time synthesis signals)로 정렬하는 변환단계, 셋째로 변환된 단구간 합성신호를 중첩가산 (overlap-add)하여 합성하는 합성단계로 나뉜다.

LP-PSOLA 합성방식은 TD-PSOLA (Time Domain PSOLA) 합성방식의 변형된 형태로, TD-PSOLA 합성방식이 시간축 변환과 피치주기 변환을 음성신호에서 수행하는 반면에 LP-PSOLA 합성방식은 잔차신호에서 수행한다는 것이 가장 큰 차이점이다. 즉, TD-PSOLA 합성방식을 수행하기 전에 음성신호에서 각 분석 피치마크를 기준으로 분석 필터를 추정한다. 그리고나서 단구간 분석신호를 분석 필터에 역필터링하여 구해진 잔차신호는 시간축 변환과 피치주기 변환을 위해 합성 피치마크를 기준으로 다시 정렬된다. 마지막으로 변환된 잔차신호를 합성 필터에 필터링함으로써 변환된 음성신호를 얻을 수 있다 [5]. 이때 음성변환을 하기 위해서 합성 필터는 대응되는 분석 필터의 변환된 값을 이용하면 된다.

3.1.1. 분석단계

분석단계는 입력되는 음성신호 $x(n)$을 단구간 분석신호열 $x(s,n)$으로 나누는 과정이다.

$$x(s,n) = h_s(n) x(n + t_a(s)) \tag{11}$$

여기서 s는 단구간 분석신호의 인덱스이고 n은 각 단구간 분석신호안의 샘플 인덱스가 되며, $h_s(n)$은 분석 창함수이고 $t_a(s)$는 s번째 분석 피치마크를 나타낸다. 이때 사용되는 분석 창함수는 해닝(hanning)형이며 그 길이 T는 단구간 분석신호에서 구한 피치주기 $P(s)$에 비례한다. 즉, $T = \mu P(s)$로 비례계수 μ는 일반적으로 2 이상의 값을 가진다. 분석 피치마크 $t_a(s)$는 유성음인 경우 피치에 해당되는 값을 가지며 무성음이나 묵음인 경우에는 일정한 간격을 가진다. 이렇게 구한 모든 단구간 분석신호에서 LPC 켑스트럼열을 구하고 각 프레임별로 나온 단구간 분석 잔차신호열을 중첩가산하여 음성신호와 같은 길이의 잔차신호를 구한다. 그리고나서 식 (11)에서 설명한 것과 동일한 방법으로 잔차신호에서 분석 잔차신호열 $e(s,n)$을 추출한다.

3.1.2. 변환단계

분석단계에서 생성된 단구간 분석 잔차신호열은 원화자 음성의 피치 간격으로 배열되어 있으므로 이 간격을

조절함으로써 대상화자 음성의 피치주기를 갖도록 변환할 수 있다. 즉, 변환하고자 하는 피치 간격인 합성 피치마크를 기준으로 단구간 분석 잔차신호열을 재배열하여 단구간 합성 잔차신호열을 구하게 된다. 이때 합성 피치마크는 원하는 피치주기 변환을 수행할 수 있도록 합성 피치마크 주위에 있는 분석 피치마크의 피치주기에 $1/\beta$을 곱하여 구한다. 이렇게 구한 합성 피치마크를 기준으로 대응되는 각 분석 피치마크의 단구간 잔차신호를 중첩가산하여 피치주기가 변환된 잔차신호를 구한다. 그리고 분석단계에서 추출한 LPC 켑스트럼은 선형 변환식을 이용하여 대상화자의 LPC 켑스트럼으로 변환하고 각 합성 피치마크를 기준으로 정렬한다.

3.1.3. 합성단계

합성단계에서는 피치주기가 변환된 잔차신호에서 단구간 합성 잔차신호를 추출하고 변환된 LPC 켑스트럼에서 구한 LPC 계수를 컨볼루션하여 단구간 음성신호를 얻는다. 최종 변환된 전체 음성신호는 단구간 음성신호를 중첩가산함으로써 구할 수 있다.

IV. 실험 및 고찰

4.1. 선형다변회귀모델을 이용한 파라미터 변환

훈련데이터로서 연구실 환경에서 남성화자 3명과 여성화자 1명으로부터 각 화자에 대해 약 5분 길이의 음성신호를 수집하여 사용하였다. 표 1은 실험에 사용된 실험 조건을 나타내며, DTW와 원화자의 LPC 켑스트럼을 벡터양자화하기 위해 사용되는 거리척도로서 유클리디언 (Euclidean) 거리를 사용하였다.

실험에 사용된 3명의 20대 남성화자를 각각 M1, M2, M3로 나타내고 1명의 20대 여성화자를 F1으로 나타내기로 한다. 실험은 M1에서 M2로, M3에서 F1으로 그리고 M1에서 F1으로 변환을 수행하였다. M1에서 M2로의 변환을 CASE 1, M3에서 F1으로의 변환을 CASE 2, M1에서 F1으로 변환하는 과정을 CASE 3로 칭한다. 표 2는 실험에 사용된 화자의 평균 피치주기로서 자기상관함수를 이용하여 추출한 값을 샘플수로 나타낸 것이며, 표 3은 화자간의 평균 피치주기를 변환하기 위하여 적용된 피치주기 변환계수를 나타낸 것이다.

화자간의 LPC 켑스트럼 변환에 대한 객관적인 평가는 식 (12)로 표현되는 LPC 켑스트럼의 평균 유클리디언 거리차를 이용하였다. 이때 C^a, C^b는 비교하고자 하는 대상의 LPC 켑스트럼을 의미한다. 표 4는 각 CASE에 대하여 벡터 코드북 크기를 변화시키면서 구한 LPC 켑스트럼의 변환율을 보여주고 있으며 이를 그림으로 나타낸 것이 그림 2이다. 여기서 코드북 크기가 증가함에 따라 변환율도 거의 선형적으로 증가하며 대상화자의 LPC 켑스트럼으로 더욱더 접근하게 된다는 것을 알 수 있다. 그러나 코드북 크기를 크게 하면 검색시간이 많이 걸리기 때문에 실험에서는 64개의 코드워드를 가지도록 하였다. 표 5는 실험에 사용된 64개의 코드북 크기에 대해 각 CASE의 평균 거리차를 구한 것으로서 ST (Source/Target)보다는 TTR (Target/Transform)이 훨씬 작아진다는 것을 보여주고 있다. 이는 원화자의 LPC 켑스트럼에 비해 변환된 LPC 켑스트럼이 대상화자에 가깝다는 것을 의미한다.

$$D=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}[\ C_i^a(j)-C_i^b(j)\ ]^2 \qquad (12)$$

그림 3은 각 CASE에 대한 원화자, 대상화자 그리고 변환된 LPC 스펙트럼을 보인 것으로, 실선은 원화자와 대상화자의 LPC 스펙트럼 특성이며 점선은 변환된 계수의 LPC 스펙트럼 특성이다. 그림에서 변환된 계수의 LPC 스펙트럼이 원화자의 LPC 스펙트럼에 비해 대상화자의

표 1. 실험 조건
Table 1. Experimental condition.

샘플링 주파수	8 kHz	LPC 켑스트럼 차수	16차
양자화 레벨	16 bits	프레임 크기	160 샘플 (20ms)
벡터 양자화	LBG 알고리즘 64개의 코드북 크기	프레임 이동 크기	80 샘플 (10ms)
		창함수	해닝 창함수

표 2. 각 화자의 평균 피치주기
Table 2. Average pitch period for each speaker.

화자	M1	M2	M3	F1
평균 피치주기	80.06	66.20	71.75	36.87

표 3. 각 CASE의 피치주기 변환계수
Table 3. Pitch-scale modification factor for each CASE.

변환 대상	CASE 1	CASE 2	CASE 3
피치주기 변환계수	1.21	1.95	2.18

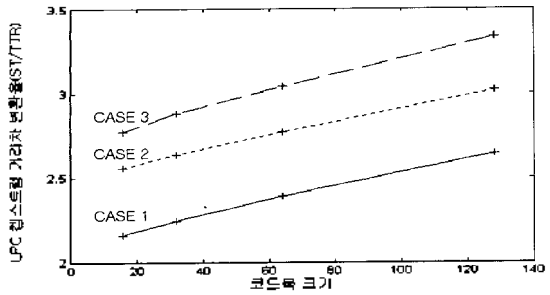


그림 2 . 코드북 크기에 따른 LPC 켑스트럼의 변환율 비교
Fig. 2. Comparison of LPC cepstrum transformation ratio for different codebook size.

표 4. 코드북 크기에 따른 LPC 켑스트럼의 변환율 비교
Table 4. Comparison of LPC cepstrum transformation ratio for different codebook size.

코드북 크기	CASE 1	CASE 2	CASE 3
16	2.1630	2.5579	2.7712
32	2.2484	2.6363	2.8794
64	2.3903	2.7680	3.0396
128	2.6402	3.0158	3.3308

표 5. LPC 켑스트럼의 거리차 비교
Table 5. Comparison of LPC cepstrum distances.

	CASE 1	CASE 2	CASE 3
원화자와 대상화자(ST)	0.7858	1.0863	1.1578
원화자와 변환된 계수(STR)	0.4571	0.6938	0.7769
대상화자와 변환된 계수(TTR)	0.3288	0.3924	0.3809
변환율(ST/TTR)	2.3903	2.7680	3.0396

LPC 스펙트럼에 유사하게 됨을 알 수 있다.

원화자의 LPC 켑스트럼을 대상화자의 LPC 켑스트럼으로 변환하면 그 극점이 단위원 외부에 존재하여 불안정한 특성을 나타내는 경우가 있다. 이런 불안정한 특성을 없애기 위하여 LPC 계수의 극점이 단위원 외부에 존재할 때, 그 극점을 단위원 내부에 존재하는 상반되는 점 (conjugate reciprocal location)으로 이동시켜 시스템의 특성을 안정화시켜서 음성을 합성하는데 사용하였다. 그림 4는 불안정하게 변환된 LPC 계수에 단구간 분석 잔차신호를 통과시켜 얻은 단구간 합성신호와 특성을 안정화시킨 LPC 계수를 통과시켜 얻은 단구간 합성신호를 보인 것이다. 그림 5는 변환된 LPC 계수와 그 계수의 불안정한 특성

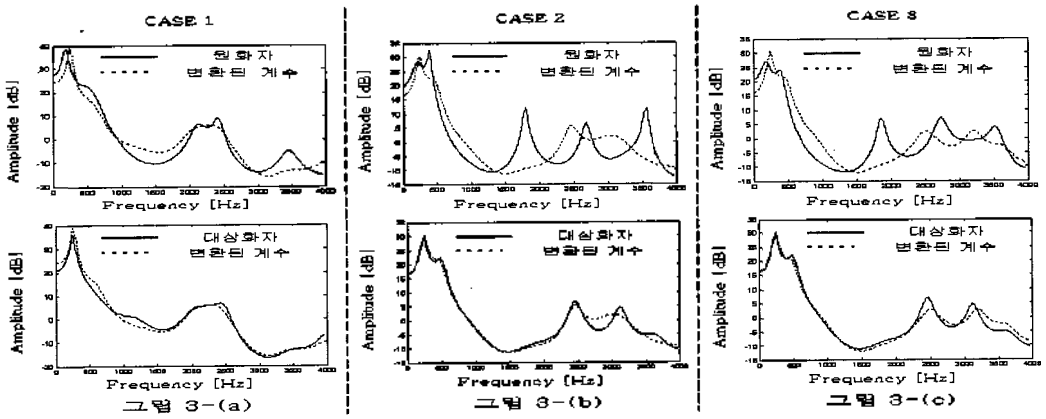


그림 3. 각 CASE에 대한 LPC 스펙트럼 특성 비교
Fig. 3. Comparison of LPC spectrum characteristics for each CASE.

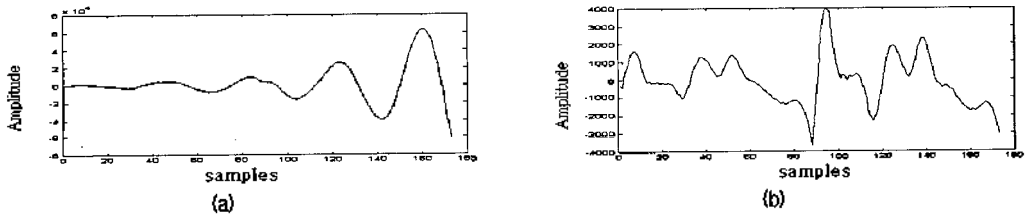
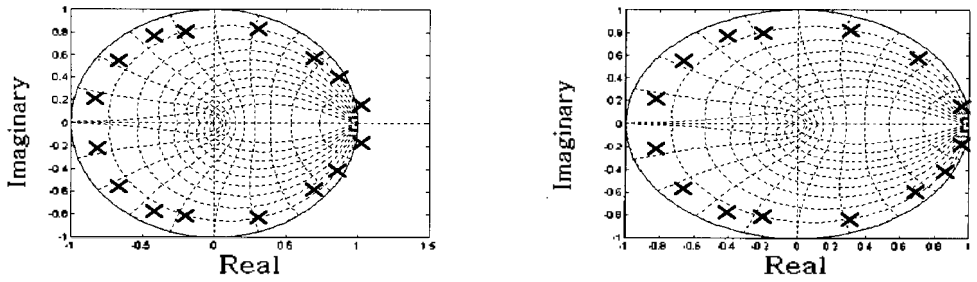
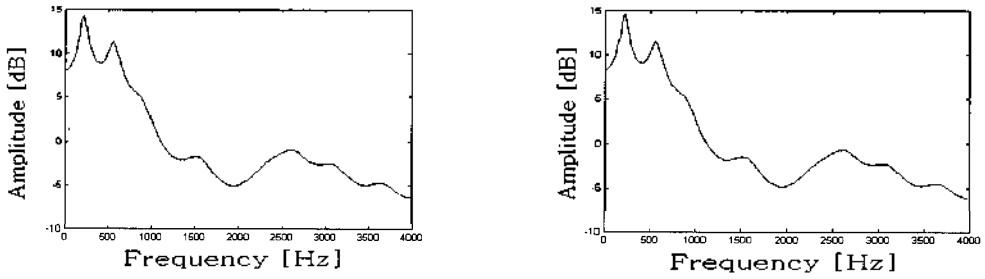


그림 4. (a) 변환된 LPC 계수로 합성한 단구간 합성신호
(b) 안정화시킨 LPC 계수로 합성한 단구간 합성신호
Fig. 4. (a) Synthetic speech signal with transformed LPC coefficients,
(b) Synthetic speech signal with stablized LPC coefficients.



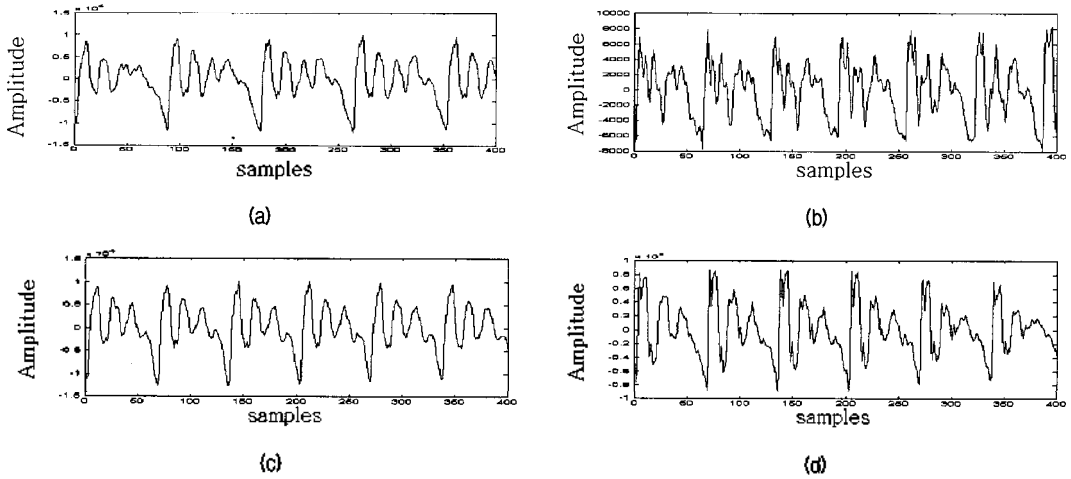
(a) (b)

그림 5. (a) 변환된 LPC 계수의 극점
(b) 안정화시킨 LPC 계수의 극점
Fig. 5. (a) Poles of transformed LPC coefficient,
(b) Poles of stablized LPC coefficient.



(a) (b)

그림 6. (a) 변환된 LPC 계수의 스펙트럼
(b) 안정화시킨 LPC 계수의 스펙트럼
Fig. 6. (a) Spectrum of transformed LPC coefficient,
(b) Spectrum of stablized LPC coefficient.



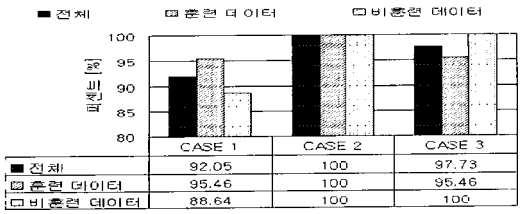
(a) (b)

(c) (d)

그림 7. /으/ 모음에 대한 안정된 구간의 파형
(a) 원화자 (b) 대상화자 (c) 피치주기만을 변환한 경우
(d) 피치주기와 LPC 캡스트럼을 함께 변환한 경우
Fig. 7. Waveforms of stable region for sound /으/.

을 안정화시킨 LPC 계수의 극점을 그린 것이며 그림 6은 변환된 계수의 LPC 스펙트럼과 불안정한 특성을 안정화시킨 계수의 LPC 스펙트럼을 보인 것이다. 그림 4~6에 나타나듯이 변환된 LPC 계수의 극점이 단위원 외부에 존재할 때 그 극점을 단위원 내부로 이동시킴으로써 LPC 스펙트럼은 바뀌지 않고 합성된 음성신호의 왜곡을 제거할 수 있었다[7].

그림 7은 원화자의 /으/ 모음에 대한 안정된 구간의 파



	CASE 1	CASE 2	CASE 3
■전체	92.05	100	97.73
▣훈련 데이터	95.46	100	95.46
□비훈련 데이터	88.64	100	100

그림 8. 각 청취 실험별 적중률
Fig. 8. Correct identification ratio for each listening test.

형, 대상화자의 파형, 원화자의 평균 피치주기만을 변환한 파형 그리고 LPC 켑스트럼과 평균 피치주기를 함께 변환한 파형을 각각 보인 것으로 LP-PSOLA 합성방식을 이용하여 화자간의 피치주기 변환을 수행한 결과를 제시하고 있다. 파형에서 나타나듯이 LPC 켑스트럼과 피치주기를 함께 변환한 파형이 평균 피치주기만을 변환한 파형에 비해 대상화자의 파형과 더 유사함을 볼 수 있다. 이는 성도 특성 변환을 수행하지 않고 피치주기만을 변환시켜서는 대상화자의 음색을 갖도록 변환할 수 없다는 것을 말한다.

4.2. 변환된 음성의 청취 실험

음성변환에 대한 주관적인 성능 평가를 위해 13명의 청취자에게 ABX 실험을 수행하였다. ABX 실험은 실험에 참여한 청취자에게 먼저 A, B 두 화자의 음성을 각각 들려주고 세번째로 변환된 음성 X를 들려주어 이 X 음성이 A, B 두 음성중에 어느쪽에 가까운지를 선택하는 청취 실험이다. 이 실험 결과를 그림 8에 나타내었다.

그림 8에서 CASE 1의 경우가 가장 낮은 적중률을 보이는데, 표 5에 나타나듯이 CASE 1의 평균 LPC 켑스트럼 거리차 변화가 가장 작기 때문이라고 짐작할 수 있으나 반드시 LPC 켑스트럼 거리차 변화 정도에만 좌우되는 것은 아니다. 이는 남성화자에서 남성화자로의 변환에 비해 상대적으로 피치주기 변환계수가 큰 여성화자로의 변환에서 적중률이 높은 결과가 말해 주듯이, 성도 특성 변환에 의한 영향뿐만 아니라 높은 피치주기 변환계수로 변환되어 피치주기가 작아진 음성이 일반적으로 여성화자에 가깝게 들리기 때문이라고 생각된다. 또, 그림 8의 결과는 훈련에 사용된 데이터와 훈련에 사용되지 않은 데이터를 비교했을 때, 적중률에서 큰 차이가 없음을 보여준다. 이는 원화자 임의의 음성을 대상화자의 음성으로 변환할 수 있음을 말한다. 그러나 변환된 음성의 음질은 상대적으로 CASE 1이 가장 좋지만 전체적으로 저하되는 경향을 보였다. 이것은 CASE 1의 경우가 다른 CASE 2, CASE 3에 비해 LPC 켑스트럼의 거리차 변화가 가장 작기 때문이라고 생각되며 전체적으로 음질이 저하되는 것은 피치주기가 변환된 단구간 잔차신호와 변환된 LPC 계수를 컨볼루션할 때 발생되는 왜곡과 심한 피치주기 변환으로 인한 잔차신호의 왜곡에서 기인된다고 생각된다.

V. 결 론

본 논문에서는 한 사람이 발성한 음성을 마치 다른 사람이 발성한 음성처럼 들리도록 변환하는 기법인 음성변환을 수행하였다. 제시한 방법은 음성발생모델에 잘 적용되도록 LP-PSOLA 합성방식을 이용한 것으로, 성도 특성 파라미터인 LPC 켑스트럼의 변환과 여기신호 특성 파라미터인 잔차신호의 피치주기 변환을 동시에 그리고 독립적으로 수행하였다. 성도 특성 변환은 LPC 켑스트럼을 선형다변회귀모델에 적용하였고, 화자간의 운율정보 변환은 여기신호 특성에 해당되는 잔차신호를 추출하여 LP-PSOLA 합성방식으로 화자간의 전체적인 평균 피치주기를 변환함으로써 수행하였다. 실험 결과는 선형다변회귀를 이용하여 변환한 LPC 켑스트럼이 원화자보다는 대상화자의 LPC 켑스트럼에 근접하며, 이는 변환된 성도 특성이 대상화자의 성도 특성에 유사하다는 것을 의미한다고 볼 수 있다. 또, 변환된 계수의 불안정한 특성은 그 극점을 단위원 내부로 이동시킴으로써 안정성을 보장할 수 있었다.

향후 변환된 LPC 켑스트럼과 잔차신호를 컨볼루션할 때 발생되는 왜곡으로 인한 음질 저하를 개선하고 전체적인 피치주기 변환 외에 세밀한 운율정보 변환에 대한 연구가 수행되어야 할 것이다.

참 고 문 헌

1. Hisao Kuwabara, Yoshinori Sagisaka, "Acoustic characteristics of speaker individuality: Control and conversion", Speech Communication, vol. **16**, No. 2, pp. 165-173, 1995.
2. Masanobu Abe, Satoshi Nakamura, Kiyohiro Shikano, Hisao Kuwabara, "Voice conversion through vector quantization", ICASSP'88, pp. 655-658, 1988.
3. H. Valbret, E. Moulines and J. P. Tubach, "Voice transformation using PSOLA technique", Speech Communication, vol. **11**, No. 2-3, pp. 175-187, 1992.
4. Brice carnahan, H. A. Luther, James O. Wilkes, *Applied Numerical methods*, John Wiley & Sons, Inc., 1969.
5. W. B. Kleijn, K. K. Paliwal, Speech coding and synthesis, Chapter 15, Elsevier, 1995.

6. Y. Linde, A. Buzo, R. M. Gray, "An algorithm for vector quantizer design", *IEEE Trans. Commun.*, vol. 28, No. 1, pp. 84-95, Jan., 1980.
7. Alan V. Oppenheim, Roland W. Schafer, Discrete-time signal processing, Prentice-Hall International. Inc., 1989.
8. 최철민, 전범기, 성굉모, "유성음의 잔류신호 변환을 이용한 음색 변환", 한국음향학회 정기총회 및 학술발표회 논문집 제 16권 제 2(s)호, pp. 127-130, 1997.

저자 약력

● 권 홍 석 (Hong Seok Kwon)

1997년 2월: 경북대학교 전자공학과 졸업
1999년 2월: 경북대학교 전자공학과 석사
1999년 3월~현재: 경북대학교 전자공학과 박사과정
※ 주관심 분야: 음성신호처리, 디지털신호처리, 적응신호처리, 오디오코딩 등

● 배 건 성 (Keun Sung Bae)

1977년 2월: 서울대학교 전자공학과 졸업
1979년 2월: 한국과학기술원 전기및전자공학과 석사
1989년 5월: University of Florida 공학박사
1979 3월~현재: 경북대학교 전자·전기공학부 교수
※ 주관심분야: 음성분석 및 인식, 디지털신호처리, 디지털통신, 음성부호화, 웨이브렛 이론 등