論文98-35S-4-12

낮은 차원의 벡터 변환을 통한 음성 변환

(Voice Conversion Using Low Dimensional Vector Mapping)

李 起 承 *, 都 元 *, 尹 大 熙 *

(Kee Seung Lee, Won Doh, and Dae Hee Youn)

요 약

이 논문에서는 임의의 화자가 발성한 음성을 다른 화자가 발성한 음성처럼 들리도록 변환하는 음성 변환 알고리즘을 제안하였다. 개인의 음성 특성을 변환하기 위해 성도 전달 함수의 특성을 변환 변수로 사용하였으며, 기존의 기법과 비교하여 적은 계산량으로 고음질의 변환음을 얻기 위한 새로운 방법을 제안하였다. 성도 전달 함수의 변환은 소프트 분류에 의한 선형 변환을 통해 이루어지며, 특징 변수로서 평균 LPC 켑스트럼과 낮은 차원의 벡터 신호로 모형화한 잔차 LPC 켑스트럼을 사용하여 변환할 때의 계산량이 줄어들도록 하였다. 제안된 음성 변환 기법의 성능을 평가하기 위해 2명의 남성 화자와 1명의 여성 화자로부터 수집된 61개의 한글 낱말을 사용하여 변환 규칙을 생성하였으며, 이를 동일 화자가 낭독한 9개의 한글 문장에 적용하여 객관적인 성능 평가와 주관적 청취 테스트를 수행하였다.

Abstract

In this paper, we propose a voice personality transformation method which makes one person's voice sound like another person's voice. In order to transform the voice personality, vocal tract transfer function is used as a transformation parameter. Comparing with previous methods, the proposed method can obtain high-quality transformed speech with low computational complexity. Conversion between the vocal tract transfer functions is implemented by a linear mapping based on soft clustering. In this process, mean LPC cepstrum coefficients and mean removed LPC cepstrum modeled by the low dimensional vector are used as transformation parameters. To evaluate the performance of the proposed method, mapping rules are generated from 61 Korean words uttered by two male and one female speakers. These rules are then applied to 9 sentences uttered by the same persons, and objective evaluation and subjective listening tests for the transformed speech are performed.

I. 서 론

음성 변환[1-3,5,6,8-12,14]이란 음성 신호가 가지고 있는 몇 개의 특징 변수를 변환하여 본래의 음성 신호와는 다른 음성 신호를 합성하는 기법을 말한다. 음성변환에는 음성의 발성 속도를 변환하는 시간축 변환[2], 억양을 변환하는 피치 변환[8], 성도 전달 함수의 특성을 변환하는 포만트 변환[3] 기법 등을 들 수 있다. 이들 변환 기법을 종합적으로 적용한 음성 개성 변환 기법은 입력 화자가 지니고 있는 특징 변수를 목표 화자의 특징 변수와 유사하도록 변환함으로써, 입력 음성 신호를 마치 목표 화자가 발성하는 것처럼 들리도록 변환하는 기법을 말한다[5,6,10-12,14].

음성 변환이 이루어지기 위해서는 크게 학습 과정과 변환 과정이 필요하다. 학습 과정은 주어진 입력 화자와 목표 화자의 특징 변수들 간의 대응 관계를 추정하는 과정으로, 변환에 앞서 미리 습득된 입력 화자와

* 正會員, 延世大學校 信號處理硏究센터
(Center for Signal Processing Research Yonsei Univ.)

接受日字: 1997年12月15日, 수정완료일: 1998年3月30日

목표 화자의 음성 데이터로부터 얻어진다. 변환 과정은 입력된 음성 신호에서 변환 변수를 추출하고 이 변수에 대해 학습 과정에서 추정된 대응 관계를 이용하여 변환을 수행함으로써 구현된다. 음성 변환을 위한 특징 변수는 개인의 특징을 잘 반영하고 있는 변수들로서, 성도 전달 함수의 특성을 나타내는 특징 변수가 그 대표적인 예이다.

음성 변환에 대한 연구는 화자 인식 및 화자 적응 음성 인식에 관한 연구를 기반으로 하고 있다. 대표적인 예로서, Abe 등이 제안한 음성 변환 기법[1]은 화자 적응 음성 인식에 사용된 코드북 매핑 기법을 통해 음성 변환을 수행하였다. Nam등에 의해 제안된 음성 변환 기법[5]은 성도 전달 함수를 나타내기 위한 특징 변수로 LPC 켑스트럼을 사용하였으며, LPC 켑스트럼의 변환에 신경망(Neural Network)을 사용하여 대응 관계를 비선형식으로 모형화하였다. 또한 Valbret 등은 음성 합성기의 운율 제어로 널리 사용되고 있는 PSOLA(Pitch Synchronous Over-Lap Add) 기법[8]을 바탕으로, 성도 전달 함수의 변환을 위해 청각 특성을 고려한 멜(mel) 켑스트럼을 변환하는 기법을 제안하였다[6]. 이 기법은 각 화자에 대한 멜(Mel) 켑스트럼을 다중 정규 분포를 갖는 확률 신호로 모형화하여 대응 관계를 분류화된 선형 변환식으로 표현하였다.

이 논문에서는 기존의 방법들과 비교하여 적은 계산량만으로 고음질의 변환음을 얻기 위한 새로운 방법을 제안하였다. 먼저 각 화자에 대한 성도 전달 함수의 특성을 선형 시불변 특성 변수와 시변 특성 변수로 구분하여 표현하였다. 여기서 시불변 특성 변수는 화자의 발성기관을 반영하는 평균적인 특성을 나타내고, 시변 특성 변수는 음소에 따라 다르게 나타나는 특성을 나타낸다.

제안된 방법에서 시불변 특성의 변환은 단순히 평균 특성의 대체에 의해 구현함으로써 시스템을 전체적으로 간단하게 구성할 수 있는 장점을 지니고, 시변 특성 변수, 곧 음소적 특성의 변환은 변환 행렬에 의한 선형 변환 기법을 바탕으로 이루어진다. 이를 위해 음소에 따라 적응적인 변환을 수행하는 분류화된 변환 기법을 제안하였다. 그러나 분류 구획이 명백히 정의되는 하드 클러스터링 기법이 구획의 경계 부근에서 그릇된 변환을 수행할 수 있다는 단점에 따라, 여기서는 확률적인 분류를 행하는 소프트 클러스터링 기법을 적용하였다. 그러나 이 경우, 모든 클래스에 대한 변환 벡터를 선형 조합의 형태로 표현해야하므로 변환 행렬의 추정과, 변환 과정시 계산량이 크게 증가한다는 단점이 발생한다. 이의 해결을 위해 LPC 켑스트럼 벡터를 고유 분해하여 비교적 큰 고유값을 갖는 고유 벡터들로 기저 벡터들을 구성한 다음, 이 벡터들로 LPC 켑스트럼 벡터를 표현하는 방법으로 정보 감축을 꾀하였다. 이러한 방법을 통하여 고음질 구현에 유리한 소프트 클러스터링 기반 선형 변환의 장점을 살리면서 계산량을 줄일 수 있게 되었다.

이 논문의 구성은 다음과 같다. 서론에 이어 2장에서는 성도 전달 함수의 변환 방법에 대해 알아본다. 3장에서는 제안된 기법들을 실제 음성 신호에 적용하여 객관적, 주관적인 변환 정도를 살펴보고, 변환음의 음질을 평가하기로 한다. 마지막으로 4장의 결론에서는 제안 방법의 고찰과 추후 연구 과제 등을 제시하며 끝맺는다.

Ⅱ. 음성 신호의 모형화 및 변환

음성 신호의 변환을 위해서는 기본적으로, 분석, 변환, 합성의 3단계 과정이 필요하다. 분석 과정은 변환의 대상이 되는 음성의 매개변수를 추출하는 과정이며, 변환은 미리 정해진 규칙에 따라 추출된 매개변수를 변환하는 과정이며, 합성은 변환된 매개변수를 이용하여 음성 신호를 합성하는 과정이다. 각 매개변수의 변환 규칙은 변환에 앞서 미리 구성된 데이터 베이스를 이용한 학습 과정에서 얻어진다. 곧, 변환하고자 하는 화자의 음성 신호와 목표로 하는 화자의 음성 신호를 동일한 단어나 문장에 대해 미리 구성하고, 이들 데이터 간의 대응 관계를 추정하여 이를 변환 규칙의 작성에 이용하는 것이다.

1. 선형 모형을 통한 성도 전달 함수의 모형화

음성 변환이 효과적으로 구현되기 위해서는 두 화자가 동일한 음소를 발성하였을 때 이 두 가지 음성 신호 간에 어떠한 차이를 나타내는지 살펴보는 것이 필요하다. 화자 A, B가 동시에 동일한 음소를 발성한 경우에, 두 화자로부터 추출된 성도 전달 함수의 특성은 동일한 음소 데이터일 경우에도 서로 다른 형태를 나타나는데, 이러한 차이는 크게 두 가지 요인에 의한 것으로 생각할 수 있다. 첫번째로 두 화자간의 발성

기관 차이에서 오는 음향적 특성에 의한 것으로, 이는 모든 음소에 대해 동일하게 나타난다. 두번째는 음소에 의존적으로 나타나는 화자의 특성에 의한 것으로 음소마다 각기 다른 형태로 나타난다[7].

이 논문에서는 이러한 두 가지 특성을 직렬 연결된 선형 시스템으로 모형화하였다. 따라서 임의 화자가 발성하는 특정 음소의 신호는 여기(excitation) 신호가 음향적 특성을 나타내는 선형 시스템과 음소적 특성을 나타내는 선형 시스템을 차례로 통과함으로써 발생된다고 가정하였다. 이를 식으로 나타내면 아래와 같다.

$$H_p^{(s)}(\omega) = H^{(s)}(\omega) L_p^{(s)}(\omega) \tag{1}$$

위 식에서 첨자 s와 p는 각각 화자, 음소의 인덱스로서, $H^{(s)}(\omega)$는 근원 화자에 대한 음향적 특성을 나타내며, $L_p^{(s)}(\omega)$는 근원 화자에 대한 p-번째 음소의 특성을 나타낸다. 이를 켑스트럼 영역의 신호로 바꾸면 아래와 같이 나타낼 수 있다.

$$h_p^{(s)}(n) = h^{(s)}(n) + l_p^{(s)}(n) \tag{2}$$

주어진 성도 전달 함수의 특성으로부터 음향적 특성을 나타내는 신호와 음소적 특성을 나타내는 신호를 분리하기 위해, 여기서는 켑스트럼의 장시간 평균을 이용하였다. 음향적 특성을 나타내는 신호는 모든 음소에 걸쳐 동일하게 나타나므로, 특정 화자의 음성 데이터로부터 켑스트럼 계수를 추출하고, 모든 음소에 대한 평균을 구하면 근사적으로 음향적 특성의 신호를 얻을 수 있다. 곧,

$$h^{(s)}(n) \cong E[\, h_p^{(s)}(n)\,] \tag{3}$$

위 식에서 $E[\,.\,]$는 학습 데이터에 대한 전체 평균을 나타낸다. 이때 필요한 조건은 p-번째 음소적 특성 $l_p^{(s)}(n)$들의 평균이 0이 되어야 한다는 것이다. 이는 충분히 많은 켑스트럼 데이터에 대해 평균을 구하는 경우 음소 정보에 대한 평균은 0으로 근사화할 수 있음이 입증되어 있다[7].

이 논문에서는 $h^{(s)}(n)$을 추정하기 위해 얼마만큼의 학습 데이터가 필요한지를 알아보기 위해 켑스트럼 평균값의 수렴 정도를 관찰하였다. 실험 결과에 따르면, 화자에 따른 차이가 약간 있지만, 대개 2,000 - 3,000개 이상의 켑스트럼 데이터로 얻어진 평균은 거의 일정함을 알 수 있었다. 따라서 음향적 특성을 나타내는 특징 매개변수를 추정하기 위해서는 3,000개 이상의 켑스트럼 데이터가 필요하다.

2. 성도 전달 함수의 변환

제안된 성도 전달 함수의 변환은 음향적 특징 변수의 변환과 음소적 특징 변수의 변환 두 가지로 구분되어 수행된다. 이 과정을 그림 1에 나타내었다. 성도 전달 함수의 특성을 나타내는 변수로서는 선형 예측 계수를 사용하였으며, 이를 로그 영역으로 나타내기 위해 다시 LPC 켑스트럼 계수로 바꾸어 사용하였다. 음향-음소 모형화에 따른 근원 및 목표 화자의 p-번째 음소에 대한 LPC 켑스트럼을 나타내면 다음과 같다.

$$\mathbf{C}_p^{(s)} = \mathbf{C}_p^{(s)} + \overline{\mathbf{C}}^{(s)} \tag{4}$$

$$\mathbf{C}_p^{(t)} = \mathbf{C}_p^{(t)} + \overline{\mathbf{C}}^{(t)} \tag{5}$$

여기서 $\mathbf{C}_p^{(s)}$, $\mathbf{C}_p^{(t)}$는 각각 p-번째 음소에 대한 근원 화자와 목표화자의 LPC 켑스트럼 벡터를 나타내며, $\mathbf{C}_p^{(s)}$, $\mathbf{C}_p^{(t)}$는 음소 의존 특성, $\overline{\mathbf{C}}^{(s)}$, $\overline{\mathbf{C}}^{(t)}$는 시불변 특성으로서, 각 화자에 대한 음향적 특성을 나타내는 변수이다. 성도 전달 함수의 변환의 첫단계는 평균 특성을 변환하는 것이다. 평균 특성이 변환된 근원 화자의 LPC 켑스트럼 벡터 $\tilde{\mathbf{C}}_p^{(t)}$은 아래와 같이 나타낼 수 있다.

$$\tilde{\mathbf{C}}_p^{(t)} = \mathbf{C}_p^{(s)} - \overline{\mathbf{C}}^{(s)} + \overline{\mathbf{C}}^{(t)} \tag{6}$$

여기서 벡터 $\tilde{\mathbf{C}}_p^{(t)}$는 평균적인 특성은 목표 화자와 동일하나, 음소적인 정보는 아직 근원 화자의 특성을 갖고있다. 따라서 음소 의존 특성 $\mathbf{C}_p^{(s)}$과 $\mathbf{C}_p^{(t)}$ 간의 변환 규칙이 필요한데, 이는 근원 및 목표 화자로부터 동일한 음소에 해당하는 변수가 주어졌을 때, 두 변수 간의 관계를 나타낸다. 따라서 변환식을 추정하기 위한 학습 과정에서는 두 화자로부터 추출된 동일한 음성 데이터의 LPC 켑스트럼 벡터들이 필요하다. 그러나 동일한 음성 데이터인 경우에도, 화자간의 발음 속도에서 차이가 있을 수 있기 때문에, 이러한 음소의 시간적 불일치를 보상하기 위해 DTW(Dynamic Time Warping) 알고리즘을 이용하여 LPC 켑스트럼 벡터를 음소적으로 시간 정렬하도록 하였다.

또 한가지 필요한 과정은, 주어진 근원 화자의 LPC 켑스트럼이 어떤 음소에 해당하는가를 판별하는 것이

다. 이는, 음소 의존 특성의 변환 규칙이 동일한 음소의 데이터 단위로 생성되어야 하기 때문이다. 여기서는 비교적 간단한 방법으로서, 근원 화자의 음소 의존 특성 $C_p^{(s)}$가 음소 단위로 결합 정규 분포를 갖는다고 가정하여, 벡터 양자화에 의한 음소 구분을 수행하였다. 이를 위해서 학습 과정에서 근원 화자의 LPC 켑스트럼에서 평균 LPC 켑스트럼을 제거하여 음소 의존 정보를 얻고, LBG 알고리즘을 이용하여 음소 의존 정보에 대한 기준 패턴을 작성하도록 하였다. 음소 판정 시에는 주어진 음소 의존 특성과 기준 패턴들간의 자승 오차를 각각 구하고 이 값이 최소가 되는 기준 패턴의 인덱스를 음소 정보로 사용하였다.

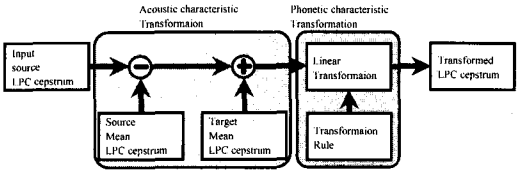


그림 1. 성도 전달 함수 변환의 블록도
Fig. 1. A block diagram of the transformation of vocal tract transfer function.

p-번째 음소에 대한 음소 의존 정보의 변환은 다음과 같은 선형 변환식을 통해 이루어진다.

$$\hat{C}_p^{(t)} = \mathbf{H}_p C_p^{(s)} + \mathbf{O}_p \tag{7}$$

여기서 $\hat{C}_p^{(t)}$는 변환된 음소 의존 정보를 나타내며, $\mathbf{H}_p$는 p-번째 음소에 대한 변환 행렬, $\mathbf{O}_p$는 p-번째 음소에 대한 근원, 목표 화자의 평균값을 보정하기 위한 오프셋 행렬이다. $\mathbf{H}_p$와 $\mathbf{O}_p$는 p-번째 음소 데이터 $\hat{C}_p^{(t)}$와 $C_p^{(t)}$ 간의 평균 자승 오차가 최소화되도록 얻어진다.

이러한 음소 정보의 변환에서 고려되어야 할 문제는 각 음소의 경계면 곧, 전이 구간에서 다소간 성능이 떨어질 수 있다는 점이다. 이는 음소의 구분이 뚜렷하지 못한 전이 구간에서는 음소의 판정시 많은 오차를 나타낼 수 있으며, 이에 따라 선택된 변환 행렬과 오프셋 행렬도 최적의 변환 행렬을 보장할 수 없기 때문이다. 이러한 현상은 인간의 성도 전달 함수 특성이 전이 구간에서도 비교적 완만한 변화 형태를 나타내는 사실과는 반대되는 것으로, 변환음의 품질을 떨어뜨리는 요인이 될 수 있다. 이 논문에서는 이러한 문제점을 해결하기 위해 확률적 방법에 의한 음소 구분과 이를 통한 변환 기법을 제안하였다.

3. 확률적인 분류 변환 기법

벡터 양자화기를 이용한 기존의 분류 기법은 주어진 입력 벡터와 코드북에 포함된 기준 코드 벡터와의 유클리디언 거리를 측정하여, 가장 짧은 거리를 나타내는 코드 벡터 인덱스를 입력 벡터의 클래스로 간주하는 것이다. 따라서 이 방법은 입력 벡터가 주어지면, 유일하게 클래스가 정해진다. 이에 비하여 확률적인 분류는 주어진 입력 벡터에 대하여 각 클래스에 대한 likelihood 값을 구하여 분류를 수행한다. 입력 벡터가 $C_i^{(s)}$로 주어질 때, 이 벡터의 클래스 j에 대한 likelihood는 다음과 같다.

$$p(c=j|C_i^{(s)}) = \frac{p(c=j, C_i^{(s)})}{p(C_i^{(s)})} \tag{8}$$

Bayesian 정리를 이용하면 식 (8)은 다음과 같이 나타낼 수 있다.

$$p(c=j|C_i^{(s)}) = \frac{p(C_i^{(s)}|c=j)p(c=j)}{\sum_j p(c=j, C_i^{(s)})} = \frac{p(C_i^{(s)}|c=j)p(c=j)}{\sum_j p(C_i^{(s)}|c=j)p(c=j)} \tag{9}$$

각 클래스가 동일한 확률로 발생된다고 가정하면, $p(c=j)$는 1/J라 할 수 있다. J는 전체 클래스 수를 나타낸다. 이 경우 (9)는 다음과 같이 나타낼 수 있다.

$$p(c=j|C_i^{(s)}) = \frac{p(C_i^{(s)}|c=j)}{\sum_j p(C_i^{(s)}|c=j)} \tag{10}$$

입력 벡터 $C_i^{(s)}$가 코드 벡터를 중심 벡터로 갖는 결합 정규 분포를 갖는다고 가정하면, 조건 확률 $p(C_i^{(s)}|c=j)$은 아래 식으로 표현된다.

$$p(C_i^{(s)}|c=j) = \frac{1}{(2\pi)^{p/2}|\Sigma_j|^{1/2}} \exp\left\{-\frac{1}{2}(C_i^{(s)} - m_j)^T \Sigma_j^{-1}(C_i^{(s)} - m_j)\right\} \tag{11}$$

여기서 $\mathbf{m}_j$는 j-클래스의 평균 벡터, 곧 j-번째 코드 벡터를 나타내며, Σ_j는 j-클래스에 대한 공분산 행렬을 나타낸다.

이러한 확률적인 클래스 분류를 음소 적응 변환에 도입하면, 변환식도 아래와 같이 전체 음소에 대한 변환 벡터의 선형 조합으로 표현된다.

$$\hat{C}_p^{(t)} = \sum_{j=1}^{J} w_j(C_p^{(s)})\{\mathbf{H}_j C_p^{(s)} + \mathbf{O}_j\} \tag{12}$$

여기서 $w_j(C_p^{(s)})$는 벡터 $C_p^{(s)}$의 j 번째 클래스 가중값

으로, (10)으로 주어지는 $p(c=j|C_i^{(s)})$의 값을 갖는다. (12)에서 행렬 $\mathbf{H}_j$와 $\mathbf{O}_j$는 앞 절과 마찬가지로, $\hat{C}_p^{(t)}$와 $C_p^{(t)}$ 간의 평균 자승 오차가 최소화되도록 구해진다. 이를 만족시키는 최적 행렬 $\mathbf{H}_j^*$와 $\mathbf{O}_j^*$는 다음 식을 만족시킨다.

$$\frac{\partial \varepsilon}{\partial \mathbf{H}_j^*}=0\ ,\quad \frac{\partial \varepsilon}{\partial \mathbf{O}_j^*}=0\ . \tag{13}$$

여기서 $\varepsilon=\mathrm{E}[\|C_p^{(t)}-\hat{C}_p^{(t)}\|^2]$ 이다. 위 식을 만족하는 $\mathbf{H}_j$와 $\mathbf{O}_j$는 다음과 같이 주어진다.

$$\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix}=\begin{bmatrix}\mathbf{R}_{kk} & \mathbf{R}_k\\ \mathbf{R}_k^T & \mathbf{R}_{ww}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{D}\\ \mathbf{P}\end{bmatrix} \tag{14}$$

이를 구성하는 각 행렬은 다음과 같다.

$$\mathbf{R}_{kk}=\begin{bmatrix}\mathrm{E}[w_1w_1C_p^{(s)}C_p^{(s)T}]&\cdots&\mathrm{E}[w_jw_1C_p^{(s)}C_p^{(s)T}]\\ \vdots&\ddots&\vdots\\ \mathrm{E}[w_1w_jC_p^{(s)}C_p^{(s)T}]&\cdots&\mathrm{E}[w_jw_jC_p^{(s)}C_p^{(s)T}]\end{bmatrix}\ (\mathrm{JM}\times\mathrm{JM}) \tag{15}$$

$$\mathbf{R}_k=\begin{bmatrix}\mathrm{E}[w_1w_1C_p^{(s)}]&\cdots&\mathrm{E}[w_jw_1C_p^{(s)}]\\ \vdots&\ddots&\vdots\\ \mathrm{E}[w_1w_jC_p^{(s)}]&\cdots&\mathrm{E}[w_jw_jC_p^{(s)}]\end{bmatrix}\ (\mathrm{JM}\times\mathrm{J}) \tag{16}$$

$$\mathbf{R}_w=\begin{bmatrix}\mathrm{E}[w_1w_1]&\cdots&\mathrm{E}[w_jw_1]\\ \vdots&\ddots&\vdots\\ \mathrm{E}[w_1w_j]&\cdots&\mathrm{E}[w_jw_j]\end{bmatrix}\ (\mathrm{J}\times\mathrm{J}) \tag{17}$$

$$\mathbf{D}=[\,\mathrm{E}[w_1C_p^{(s)}C_p^{(t)T}]\ \cdots\ E[w_jC_p^{(s)}C_p^{(t)T}]\,]^T\ (\mathrm{JM}\times\mathrm{M}) \tag{18}$$

$$\mathbf{P}=[\,\mathrm{E}[w_1C_p^{(t)}]\ \cdots\ E[w_jC_p^{(t)}]\,]^T\ (\mathrm{J}\times\mathrm{M}) \tag{19}$$

$$\mathbf{X}=[\,\mathbf{H}_1\ \mathbf{H}_2\ \cdots\ \mathbf{H}_J\,]^T\ (\mathrm{JM}\times\mathrm{M}) \tag{20}$$

$$\mathbf{Y}=[\,\mathbf{O}_1\ \mathbf{O}_2\ \cdots\ \mathbf{O}_J\,]^T\ (\mathrm{J}\times\mathrm{M}) \tag{21}$$

위 식에서 w_j는 $w_j(C_p^{(s)})$를 간략히 나타낸 것이다. (14)를 살펴보면, 최적 변환 행렬은 (MJ+J)×(MJ+J)의 크기를 갖는 행렬의 역행렬을 구함으로서 얻어짐을 알 수 있다. J는 전체 클래스의 개수, 곧 표현 가능한 대표 음소의 종류수를 나타내며, M은 변환하고자하는 벡터의 차원 수를 나타낸다. 따라서 행렬의 크기는 차원 수와 클래스의 증가에 따라 기하 급수적으로 증가하게 되며, 이에 따른 역행렬 계산의 소요 시간, singularity와 같은 연산 상의 문제를 일으킬 수 있다. 또한 실제 변환 시에도, 단 한 개의 클래스만을 고려한 기존의 방법에 비해 확률적인 변환의 경우는 모든 클래스를 고려해야하므로 계산량이 늘어나게 된다.

이러한 문제를 해결하는 방법은 클래스 수 J와 벡터의 차원 M을 되도록 작은 값으로 줄이는 것이다. 그러나 클래스 수를 줄이는 경우, 표현 가능한 대표 음소 수를 제한하므로 변환의 적응성이 떨어지게 된다. 이는 변환기의 성능 자체를 떨어뜨리는 요인이 될 수 있다. 두 번째로 차원 M을 줄이는 경우로서, 높은 차수에 해당하는 LPC 켑스트럼 벡터의 계수를 제거하는 방법을 생각할 수 있다. 높은 차수에 해당하는 LPC 켑스트럼 계수는 낮은 차수의 계수에 비해 매우 낮은 에너지를 가지고 있기 때문에 이 방법이 타당해 보일 수 있다. 그러나 높은 차수의 정보는 스펙트럼 구조를 세밀하게 표현하는데 필요한 상위 차수 큐퍼런시의 정보를 나타내므로, 이 정보가 손실되면 성도 전달 함수의 세밀한 표현이 어려워진다.

4. 낮은 차원의 신호 모형화

계산량을 줄이기 위한 한 방편으로서, 여기서는 화자 고유의 직교 기저 벡터를 이용하였다. 곧, 특정 화자의 성도 전달 함수를 해당 화자의 음성 데이터로부터 얻어지는 기저 벡터에 의해 표현함으로써, 성도 전달 함수의 세밀한 특성을 보존하면서도 낮은 차원의 벡터 신호로 근사화할 수 있도록 하는 것이다. 최적의 기저 벡터를 구하기 위해 먼저 다음과 같이 LPC 켑스트럼 벡터에 대한 공분산 행렬을 구한다.

$$\boldsymbol{R}_{cc}^{(s)}=E[\{\mathbf{C}_p^{(s)}-\overline{\mathbf{C}}^{(s)}\}\{\mathbf{C}_p^{(s)}-\overline{\mathbf{C}}^{(s)}\}^T] \tag{22}$$

위의 상관행렬이 양정치(positive definite)라면 이 행렬에 대한 고유 분해는 다음과 같이 주어진다.

$$\boldsymbol{R}_{cc}^{(s)}=\boldsymbol{Q}^{(s)}\boldsymbol{\Lambda}^{(s)}\boldsymbol{Q}^{(s)T} \tag{23}$$

이때 $\boldsymbol{Q}^{(s)}=[\,\boldsymbol{q}_1^{(s)}\ \ldots\ \boldsymbol{q}_N^{(s)}\,]$는 고유 벡터 행렬을 나타내며 $\boldsymbol{\Lambda}^{(s)}=diag[\,\lambda_1^{(s)}\ \ldots\ \lambda_N^{(s)}\,]$는 각 고유 벡터에 대응되는 고유값으로 구성된 대각행렬을 나타낸다. 음소 정보를 표현하기 위한 주요 직교 기저 벡터는 행렬 $\boldsymbol{Q}^{(s)}$에 포함된 고유 벡터 중에서 고유값 λ_i^s가 임계값 λ_{th} 이상인 고유 벡터를 사용한다. 이는, LPC 켑스트럼의 분산이 특정 화자내의 음소적인 변동에 의해 주로 일어난다는 가정에 바탕을 둔 것이다. 이 경우,

고유값이 큰 고유 벡터들은 상대적으로 음소적인 변동을 많이 반영하는 기저 벡터들로 간주할 수 있다. 따라서 화자 s의 음소적 변동을 잘 나타내는 기저 벡터 집합 $V^{(s)}$는 다음과 같이 나타낼 수 있다.

$$V^{(s)}=\{\boldsymbol{q}_i^{(s)} : \lambda_i^{(s)}\geq\lambda_{th}\} \tag{24}$$

이때 임계값 λ_{th}는 기저 벡터 집합 $V^{(s)}$에 포함된 고유 벡터만으로 LPC 켑스트럼을 구성하였을 때 청각상 왜곡이 느껴지지 않도록 설정한다. 실험적으로, 전체 에너지의 약 90%를 포함하는 상위 고유 벡터만으로 기저 벡터 집합을 구성하는 경우 청각적인 왜곡은 거의 느낄 수 없었다. 이렇게 얻어진 기저 벡터를 이용하여 음소 의존 벡터 $\boldsymbol{C}_p^{(s)}$를 나타내면 아래와 같이 Karhunen-Loève 급수 전개 형태로 표현된다.

$$\boldsymbol{C}_p^{(s)} \cong \sum_{n=1}^{N} k_p^{(s)}(n)\,\boldsymbol{q}_n^{(s)} \tag{25}$$

여기서 N은 기저 벡터 집합 $V^{(s)}$에 포함되는 기저 벡터의 총 갯수를 나타내며, $k_p^{(s)}(n)$은 근원 화자의 p-번째 음소에 대한 n 번째 기저 벡터의 사영으로서, n 번째 KLT 계수를 나타낸다.

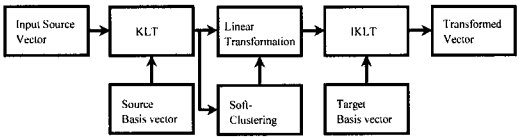


그림 2. 차원 감소 과정을 포함하는 성도 전달 함수 변환의 블록도

Fig. 2. A block diagram of the vocal tract transfer function transformation with dimensionality reduction.

이와 같은 정보 감축 과정이 포함된 성도 전달 함수 변환의 전체 블록도를 그림 2에 제시하였다. 이로부터, 실제 음소 정보의 변환은 평균값이 제거된 LPC 켑스트럼에 직접적으로 수행되는 것이 아니고 이를 KL 변환한 변환 계수에 대해 수행됨을 알 수 있다. 곧 (7)은 실제적으로 다음과 같다.

$$\hat{\boldsymbol{K}}_p^{(t)} = \mathbf{H}_p\,\boldsymbol{K}_p^{(s)} + \mathbf{O}_p \tag{26}$$

위 식에서 $\hat{\boldsymbol{K}}_p^{(t)}, \boldsymbol{K}_p^{(s)}$ 는 각각 근원 화자, 목표 화자의 LPC 켑스트럼에 대한 KLT 계수 벡터를 나타낸다. 마찬가지로 최적의 변환 행렬과 오프셋 행렬 $\mathbf{H}_j^*$와 $\mathbf{O}_j^*$는 KLT 계수 간 평균 자승 오차가 최소화되도록 구해지며, 소프트 클러스트링을 위한 likelihood 값의 계산도 LPC 켑스트럼이 아닌 KLT 계수를 통해 얻어진다.

이러한 낮은 차원의 모형화를 통한 성도 전달 함수의 변환은 특징 벡터의 특성을 충분히 보존하면서 차원수를 줄일 수 있으므로 줄어든 계산량에 따른 성능 저하를 방지할 수 있게 된다.

Ⅳ. 모의 실험 및 결과 고찰

제안된 음성 변환 알고리즘의 성능을 평가하기 위해서 몇 명의 화자를 대상으로 음성 변환을 수행하여 성능을 평가하였다. 실험에 사용된 음성 데이터는 대학원 재학중인 2명의 남성과 방송국의 전문 여자 아나운서로부터 수집하였으며 각각에 대한 음성 데이터는 M1, M2, W로 나타내었다. 음성 데이터는 우리말에서 사용 빈도수가 높은 음소를 골고루 포함하고 있는 61개의 단어들로서 이를 표 1에 제시하였다.

표 1. 실험에 사용된 음성 데이터

Table 1. Speech data used in experiment.

바람, 다리, 받고, 파도, 딸, 틀, 가을, 색동, 동굴, 입술, 까닭, 칼, 자리, 이제, 찌개, 처음, 소리, 쉬파리, 시간, 싸리, 씨알, 나비, 하늘, 동해, 마음, 나리, 글, 근세, 양식, 예닐곱, 둥지, 달, 구리, 달력, 이웃, 이발, 메밀, 세상, 애기, 해방, 취나물, 취향, 위문, 빨래, 뒤, 되풀이, 된장, 외길, 외국, 바다, 말씀, 어머니, 열, 건강, 피리, 노인, 눈동자, 눈사람, 완성, 의사, 고마워

표 2. 실험 조건.

Table 2. Experimental Condition.

A/D 변환	16KHz, 16bit, Linear
LPC 차수	20
LPC cepstrum 차수	30
학습 단어수	62
분석 프레임 길이	480 표본 (30ms)
분석 프레임 이동 거리	48 표본 (3ms)
분석 창함수	Hamming 창함수

표 2에 모의 실험 시의 조건들을 나타내었다. 실험에 사용한 음성 데이터는 비교적 조용한 환경에서 디

지털 테이프 녹음기를 사용하여 녹음하였으며, 이를 다시 표 2에 주어진 샘플링 주파수와 양자화 비트로 A/D 변환하여 실험에 사용하였다. 음성 데이터의 수집은 동일한 단어들에 대해 2차례에 걸쳐 수행되어, 첫번째로 수집된 음성 데이터는 변환 규칙을 작성하는데 이용하였으며, 두번째로 수집된 음성은 성능 평가를 수행하는데 사용하였다.

1. 객관적인 성능 평가

음성 변환의 객관적인 성능 평가는 변환된 음성 신호의 성도 전달 함수 특성과 목표 음성의 성도 전달 함수간 유사 정도를 수치로 나타내는 것이다. 이를 위해, 이 논문에서는 평균 켑스트럼 왜곡 감소율을 이용하였다. 이 값은 변환 전의 근원, 목표 화자의 켑스트럼 거리와 비교하여 변환된 켑스트럼이 목표 화자와 얼마나 유사한지를 백분율로 나타낸 것이다. 이를 식으로 나타내면 다음과 같다.

$$D_{ratio}=\left(1-\frac{D(\hat{\mathbf{C}}^{(t)},\mathbf{C}^{(t)})}{D(\mathbf{C}^{(s)},\mathbf{C}^{(t)})}\right)\times 100\ (\%) \qquad (27)$$

여기서 $D(\mathbf{X},\mathbf{Y})$는 두 벡터 $\mathbf{X}$, $\mathbf{Y}$의 평균 유클리디언 거리를 나타내며, $\hat{\mathbf{C}}^{(t)}$, $\mathbf{C}^{(t)}$는 각각 변환 켑스트럼과 목표 화자의 켑스트럼, 그리고 $\mathbf{C}^{(s)}$ 는 근원 화자의 켑스트럼을 나타낸다. 만일 변환된 LPC 켑스트럼과 목표 화자의 LPC 켑스트럼이 동일하다면 $D(\hat{\mathbf{C}}^{(t)},\mathbf{C}^{(t)})$는 0이 되며 이때의 D_{ratio}는 100의 값을 갖는다. 따라서 변환된 켑스트럼이 목표 화자의 켑스트럼에 가까울수록 D_{ratio}는 100에 근접한 값을 갖는다.

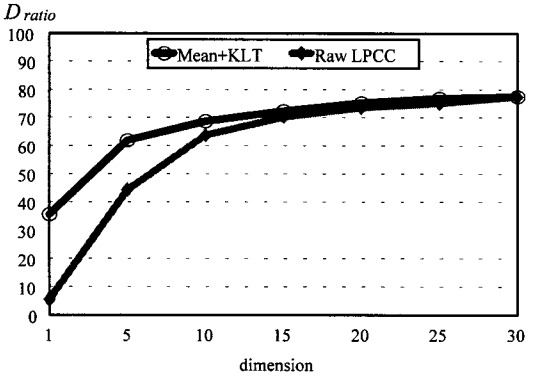


그림 3. 화자 M1-M2 변환시 왜곡 감소율 (class=16)
Fig. 3. Distortion reduction ratio for M1-M2 Conversion.

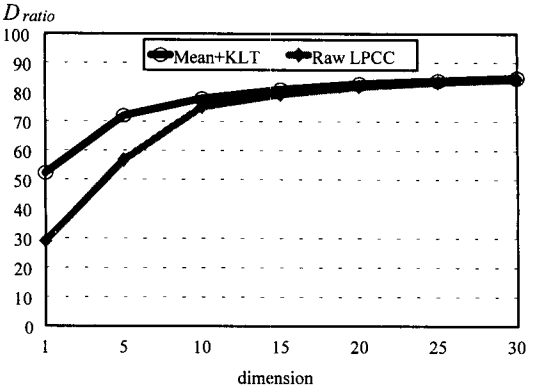
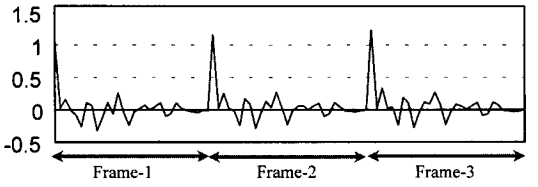


그림 4. 화자 M2-W 변환시 왜곡 감소율 (class=16)
Fig. 4. Distortion reduction ratio for M2-W conversion.

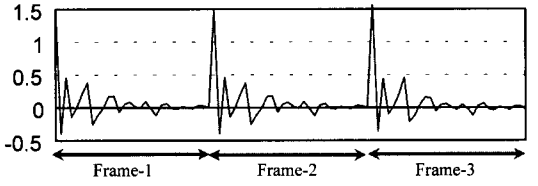
이 논문에서는 제안된 성도 전달 함수의 모형화의 타당성을 알아보기 위해, LPC 켑스트럼 자체를 변환 변수로 사용한 경우와, 평균 LPC 켑스트럼 및 KLT 계수로 표현되는 변환 변수를 사용한 경우 각각에 대해 D_{ratio}를 구하였다. 변환 방법은 두 가지 모든 경우에 대해 2장에서 제시한 확률적인 분류에 바탕을 둔 선형 변환 기법을 이용하였다.

그림 3에 분류 클래스를 16개로 설정한 경우, 화자 M1, M2 간의 각 차원수에 대한 D_{ratio} 값을 도시하였다. 그림에서 볼 수 있듯이, 동일한 갯수의 KLT 계수와 LPC 계수를 사용하는 경우, KLT 계수를 사용하는 경우가 더 큰 D_{ratio} 값을 나타내고 있다. 또한, 차원수가 증가함에 따라 초기에는 D_{ratio} 값이 크게 증가하나, 특정 차원 수에 도달한 후로는 D_{ratio} 값이 완만히 증가함을 나타내고 있다. 이러한 D_{ratio} 값은 KLT 계수의 경우, 약 10 근방에서 완만한 증가를 보이고, LPC 계수의 경우에는 약 15 근방에서 수렴되는 것을 관찰할 수 있다. 이는 높은 차원의 LPC 계수로 표현되는 성도 전달 함수의 특성이 이보다 낮은 차원으로 표현된 KLT 계수로 충분히 근사화될 수 있음을 보여주고 있다.

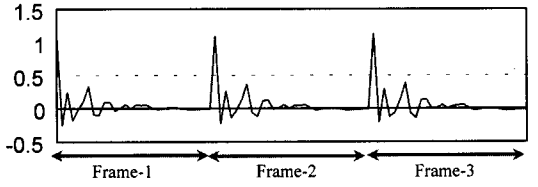
그림 4는 남성-여성간의 음성 변환을 위한 실험으로서, 화자 M2와 화자 W 간의 결과를 제시한 것이다. 이 결과는 화자 M1-M2 간의 변환 결과와 유사하게 나타난다. 차이점은 M1-M2 간의 변환에 비해 다소 높은 D_{ratio}를 보인다는 점인데, 이는 M2-W 간의 LPC 켑스트럼 관계가 제안된 변환 기법에 더욱 적합한 특성을 지니고 있는 것으로 생각된다.



(a) M2 화자의 LPC 켑스트럼



(b) W 화자의 LPC 켑스트럼



(c) 변환(M2→W) LPC 켑스트럼

그림 5. 근원, 목표 화자의 LPC 켑스트럼 및 변환된 LPC 켑스트럼

Fig. 5. Source, target, and transformed LPC cepstrum.

그림 5는 비교적 많은 특성 차이를 나타내는 M2-W 화자간의 변환에 있어서 연속된 몇 개의 프레임에 대한 변환된 켑스트럼을 나타낸 것이다. 변환된 켑스트럼은 목표 화자와 비슷한 형태를 지니고 있음을 알 수 있다.

2. 주관적인 성능 평가

음성 변환의 최종 목표는 변환음이 청취상으로 목표 화자의 음성과 유사하도록 처리하는 것이므로, 변환 음성을 실제로 청취하였을 때 이 음성이 목표 화자의 음성과 얼마만큼 유사한지를 살펴보는 것이 매우 중요하다. 이러한 주관적인 평가를 위해 ABX 테스트를 수행하였다. 실험 대상은 음성 신호 처리 연구와 실험에 대한 경험이 많은 대학원생 18명으로 구성하였으며, 사용된 음성 데이터는 실험에 사용한 61개의 단어 중에서 청취자마다 각기 다른 10개의 무작위로 추출된 단어를 사용하였다. 따라서 실험에 사용된 총 단어 수는 180개이다.

변환 방법은 제안된 평균 LPC 켑스트럼 및 KLT 계수를 확률적인 분류에 의해 선형 변환하는 기법을 사용하였다. 변환시의 조건은 10개의 KLT 계수를 사용하고 클래스 수는 16개로 설정하였다. 확률적인 분류를 위해 사용되는 공분산 행렬은 계산상의 편의를 위해 대각 행렬로 근사화된 형태를 사용하였다.

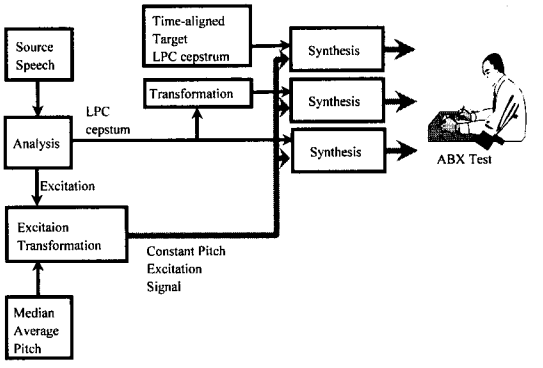


그림 6. 주관적 성능 평가의 실험 과정

Fig. 6. An experiment for subjective evaluation.

이 논문에서 사용한 ABX 테스트는 단지 성도 전달 함수의 변환에 의한 결과만을 얻기 위해, 그림 6에 제시한 바와 같이 주어진 근원 화자의 음성 신호에 대해 몇 가지 부가적인 처리를 수행하였다. 먼저 실험 대상인 두 화자의 평균 피치 값을 2로 나누어 두 화자의 중간 평균 피치를 산출한다. 근원 화자의 여기 신호는 MBE(Multi Band Excitation) 기반 여기 신호의 변환 방법[8][13]에 따라 전 프레임이 중간 평균 피치 값을 갖도록 변환한다. 고정된 피치로 변환된 여기 신호에 대해 근원 화자의 LPC 켑스트럼, 음소적으로 시간 정렬된 목표 화자의 LPC 켑스트럼, 그리고 변환된 LPC 켑스트럼을 통해 얻어진 성도 전달 함수를 통과시켜 3개의 음성 데이터를 작성한다. 이렇게 3가지 LPC 켑스트럼에 대한 음성 신호를 합성하여 ABX 테스트를 수행한다. 이처럼 복잡한 처리를 수행하는 이유는 ABX 테스트시 발생 가능한 판정 오류를 억제하기 위해서이다. 실제 실험시 청취자의 판단 기준은 성도 전달 함수의 특성 뿐만이 아니고, 발성 스타일, 곧 운율 정보에 크게 영향을 받을 수 있기 때문이다. 따라서 여기서와 같이 피치를 일정한 값으로 고정하게 되면, 운율 정보가 제거된 상태라 말할 수 있으므로 각 음성 신호는 성도 전달 함수의 특성에 의해서만 영향을 받게된다. ABX 테스트의 결과를 표 3과

표 4에 제시하였다.

표 3. M1-M2 화자간의 ABX 테스트 결과
Table 3. ABX test results for M1-M2 conversion.

실험 방법	적중률(%)
평균 LPC 켑스트럼만 변환	70.0
평균 및 잔류 특성 모두 변환	72.5

표 3은 화자 M1-M2 간의 변환 결과를 나타낸 것으로, 음향적인 특성을 나타내는 평균 켑스트럼만을 변경한 경우와 잔류 특성을 함께 변환한 경우의 결과를 제시하였다. 각각의 경우 70%, 72.5%의 적중률을 나타내고 있다. 이는 남성 화자간 변환에 있어서는 음향적인 특징 변수가 음성 변환에 있어서 중요한 역할을 차지하는 것으로 생각할 수 있으며, 음소적인 특징 변수는 두 화자에 대해 대체적으로 동일한 특성을 지닌다고 볼 수 있다.

표 4. M2-W 화자간의 ABX 테스트 결과
Table 4. ABX test results for M2-W conversion.

실험 방법	적중률(%)
평균 LPC 켑스트럼만 변환	43.4
평균 및 잔류 특성 모두 변환	86.7

이와는 반대로 M2-W 간의 변환에 있어서는 표 4에서 볼 수 있듯이 적중률 면에서 큰 차이를 나타내고 있다. 특히 잔류 특성이 함께 변환된 경우의 적중률은 86.7%로, 4가지 경우 중 가장 높은 값을 보이고 있다. 이러한 사실은, 앞 절에서 LPC 켑스트럼의 왜곡 감소율 면에서도 M2-W 변환이 더욱 우수한 성능을 나타내었다는 것으로 뒷받침되지만, 기본적으로는 남성-여성간의 음소적 특징 변수의 큰 차이에 원인이 있는 듯하다. 남성-여성간의 평균 LPC 켑스트럼은 많은 차이를 보이지만, 실제로 평균만이 변경된 음성을 청취하였을 때는 두 화자의 중간적인 음색을 느낄 수 있었다. 이로부터 남성-여성간의 성도 전달 함수 특성이 평균 특성 뿐만이 아니고 음소적 특성에 있어서도 많은 차이를 지니고 있다고 판단된다.

실험적으로, KLT 계수와 클래스 수를 각기 다르게 하여 변환에 이용하는 경우, 적중률은 KLT 계수의 차원에는 비교적 높은 영향을 받는 반면, 클래스 수에는 비교적 적은 영향을 받음을 알 수 있었다. 또한 KLT 계수의 경우에는 약 5개 이상의 계수가 사용되는 경우 전 계수(30개)를 사용하는 경우와 청취상으로 거의 차이를 느낄 수 없음을 알 수 있었다. 이는 앞서 제시한 객관적 결과와는 조금 상이한 것으로, 청취상의 유사도는 D_{ratio} 값이 50~60%를 초과할 때, 거의 동일하게 느끼게 됨을 알 수 있다.

Ⅳ. 결 론

이 논문에서는 한 사람의 음성을 다른 사람이 발성하는 것처럼 변환하는 음성 변환 알고리즘을 제안하고, 성능을 평가하였다. 제안된 화자의 특징 변수 추출, 변환 방법에 있어서 기존의 방법과는 다른 새로운 기법을 제안하였다.

먼저 화자의 특징 변수로는 성도 전달 함수의 특성을 사용하였으며, 성도 전달 함수의 특성을 모든 음소에 대해 일관적으로 나타나는 평균적인 특성 변수와, 음소에 적응적으로 나타나는 변수로 구분하여 표현하였다. 또한 음소에 적응적으로 나타나는 변수는, 변환시의 메모리양, 계산량의 감소를 위해 고유 분해를 통한 낮은 차원의 모형화 방법이 적용되었다.

성도 전달 함수의 변환은 주어진 음성 신호에서 평균 특성을 제거하고, 이를 목표 화자의 평균 특성으로 대체한 후, 음소적인 특성은 분류 선형 변환을 통해 이루어지도록 하였다. 제안된 분류 선형 변환은 클래스의 구분이 확률적인 값으로 표현되는 소프트 클러스터링 방법을 사용하였으며, 이에 따라 변환식은 각 클래스에 대한 선형 조합의 형태로 얻어지게 된다.

제안된 방법은 낮은 차원의 변수들만으로 우수한 변환 성능을 나타내었으며, 주관적인 성능 평가의 하나인 ABX 테스트에 있어서도 상당수의 청취자가 변환음을 목표 화자의 음성으로 인식하였다. 제안된 알고리즘은 음성 변환의 변수로서 성도 전달 함수의 특성만을 이용하고 있으나, 보다 완전한 변환음을 얻기 위해서는 발성음의 속도, 포만트 주파수의 시간적인 변화, 에너지 궤적 등 운율과 관련된 변수를 변환 매개변수에 포함시켜야 할 것으로 생각된다.

※ 이 논문은 1996년도 학술진흥재단의 대학부설 연구소 연구비 지원에 의하여 연구되었음.

참 고 문 헌

[1] D. G. Childers, B. Yegnunarayana and K.

Wu, "Voice conversion: factors responsible for quality," *proc. of ICASSP*, vol. 1, pp. 748-751, 1985.

[2] S. Roucos and A. M. Wilgus, "High quality time-scale modification for speech," *proc. of ICASSP*, vol. 1, pp. 493-469, 1985.

[3] P. J. Bloom, "High-quality digital audio in the entertainment industry: an overview to achievements and challenges," *IEEE ASSP Magazine*, pp. 2-25, October, 1985.

[4] G. R. Doddington, "Speaker recognition-identifying people by their voices," *Proceedings of IEEE*, vol. 73, No. 11, pp. 1651-1664, November, 1985.

[5] Il Hyun Nam, "Voice personality transformation," *Ph. D Thesis, Electrical Engineering Rensselaer Polytechnic Institute,* Troy, NY, 1991.

[6] H. Valbret, E. Moulines, and J. P. Tubach, "Voice transformation using PSOLA technique," *Speech Communication*, vol. 11, pp. 175-187, 1992.

[7] Y. Zhao, "An acoustic-phonetic-based speaker adaptaion technique for improving speaker-independent continuous speech recognition," *IEEE Trans. on Speech and Audio processing*, vol. 2, No. 3, pp. 380-394, July, 1994.

[8] E. Moulines and J. Laroche, "Non-parametric techniques for pitch-scale and time-scale modification of speech," *Speech Communication*, vol. 16, pp. 175-205, 1995.

[9] 한동철, 이기승, 윤대희, 차일환, "음성 신호 시간축 변환의 실시간 구현에 관한 연구", 한국 음향학회지, 제14권, 제2호, pp. 50-61, 1995년 4월.

[10] K. S. Lee, D. H. Youn, and I. W. Cha, "Voice personality transformation using an orthogonal vector space conversion," *proc. of EUROSPEECH '95*, Madrid, pp. 427-430, 1995.

[11] H. Mizuno and M. Abe, "Voice conversion algorithm based on piecewise linear conversion rules of formant frequency and spectrum tilt," *Speech Communication*, vol. 16, No. 2, pp. 153-164, 1995.

[12] Y. Stylianou O. Cappe and E. Moulines, "Statistical methods for voice quality transformation," *proc. of EUROSPEECH '95*, Madrid, pp. 447-450, 1995.

[13] W. M. E. Yu and C. F. Chan, "Efficient multiband excitation linear predictive coding of speech at 1.6 kbps," *proc. of EUROSPEECH '95*, Madrid, pp. 685-688, 1995.

[14] 이기승, 박군종, 윤대희, "직교 벡터 공간 변환을 이용한 음성 개성 변환", 대한 전자 공학회 논문지, No. 1, vol.-33B, pp. 96-107, Jan. 1996.

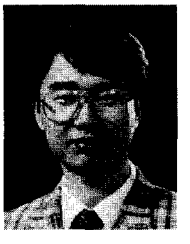
저 자 소 개

李 起 承(正會員)

1968년 1월 25일생. 1991년 연세대학교 전자공학과 졸업(공학사). 1993년 연세대학교 대학원 전자공학과 졸업(공학석사). 1993년 ~ 현재 연세대학교 전자공학과 박사과정 재학중. 주관심분야는 음성 신호 처리, 영상 신호 처리 등

尹 大 熙(正會員)

1951년 5월 25일생. 1977년 연세대학교 전자공학과 졸업(공학사). 1979년 Kansas State University 졸업(공학석사). 1982년 Kansas State University 졸업(공학박사). 1982년 ~ 1985년 Univ. of Iowa 조교수. 1985년 ~ 현재 연세대학교 교수. 주관심 분야는 음성 신호 처리, 적응 신호 처리, 레이다 신호 처리 등임



都 元(正會員)

1969년 5월 19일 생. 1992년 3월 연세대학교 전자공학과(공학사). 1994년 3월 연세대학교 전자공학과(공학 석사). 1994년 -현재 연세대학교 전자공학과 박사 과정. 주관심분야: 음성 신호 처리