論文96-33B-1-11

직교 벡터 공간 변환을 이용한 음성 개성 변환

(Voice personality transformation using an orthogonal vector space conversion.)

李 起 承 *, 朴 軍 宗 *, 尹 大 熙 *

(Lee Ki Seung, Park Kun Jong, and Youn Dae Hee)

요 약

본 논문에서는 직교 벡터 공간 변환을 이용한 새로운 음성 개성 변환 알고리즘을 제안하였다. 음성 개성 변환이란 임의 화자(source)가 가지고 있는 몇개의 특징 변수를 다른 화자(target)의 특징 변수로 변환하는 기법이다. 본 논문에서는 LPC 켑스트럼 계수와 여기 신호의 스펙트럼, 그리고 피치 궤적을 변환하여 음성 개성 변환을 구현하였다. LPC 켑스트럼 계수의 변환을 위해 직교 벡터 공간 변환 기법이 제안되었다. 이 기법은 KL(Karhunen-Loève)변환을 이용한 principle component의 분리와 최소 자승 오차를 갖는 선형 좌표 변환을 통해 LPC 켑스트럼의 변환을 수행한다. 또한, 화자간의 운율적인 특징을 변환하기 위해 피치 궤적 변환 기법이 제안되었다. 피치 궤적 변환을 위하여 먼저 두 화자간의 기준 피치 패턴의 작성하고 기준 패턴간의 대응 관계를 추 정한 후 이를 이용하여 source 화자의 피치 패턴이 target 피치 패턴으로 변환되도록 하였다. 컴퓨터를 이용한 모의 실험 결과 제안된 알고리즘은 객관적인 평가와 주관적인 평가에 있어서 우수한 성능을 나타내었다.

Abstract

A voice personality transformation algorithm using orthogonal vector space conversion is proposed in this paper. Voice personality transformation is the process of changing one person's acoustic features (source) to those of another person (target). In this paper, personality transformation is achieved by changing the LPC cepstrum coefficients, excitation spectrum and pitch contour. An orthogonal vector space conversion technique is proposed to transform the LPC cepstrum coefficients. The LPC cepstrum transformation is implemented by principle component decomposition by applying the Karhunen-Loève transformation and minimum mean-square error coordinate transformation(MSECT). Additionally, we propose a pitch contour modification method to transform the prosodic characteristics of any speaker. To do this, reference pitch patterns for source and target speaker are firstly built up, and mapping rule is estimated. These are used to convert source speaker's pitch pattern to target speaker's one. The experimental results show the effectiveness of the proposed algorithm in both subjective and objective evaluations.

I. 서 론

인간이 자신의 의사를 표현하기 위해 가장 흔히 사용하는 방법은 직접 발성하는 것이다. 음성은 입술, 성도(vocal tract), 성대 등 발성 기관의 복잡한 상호작용에 의해 발생되지만, 신호처리 측면에서는 성도를 모델링한 디지탈 필터에 주기적인 펄스나 백색 잡음과

* 正會員, 延世大學校 電子工學科, 音響, 音聲 및 信號處理研究室

(Yonsei Univ., electronics Eng., A. S. S. P. Labs)

接受日字: 1995年10月24日, 수정완료일: 1995年12月20日

같은 여기 신호 (excitation signal)를 통과시키므로써 발생된다고 모델링될 수 있다[4].

음성 변환(voice transformation)[1,2,3,8,9,11,12,13]이란 이러한 음성 신호가 가지고 있는 몇개의 특징 변수를 변환함으로써 본래의 음성과는 다른 변환된 음성을 합성하는 방법이다. 이러한 방법중 대표적인 것으로는 발음의 속도를 변환시키는 시간축 변환(time scale modification)[3,8,9,11], 톤 (tone)을 변환시키는 피치 변환(pitch modification)[8,9], 포만트 주파수를 변환하여 특수한 효과음 을 발생시키는 변환 기법(donaldduck-effect)등이 있다. 이러한 기법들은 어학 학습기[11], 발음 교 정 시스템이나 텔레비젼 방송국에서의 특수한 효과를 내기위한 목적으로 사용되고 있다. 본 논문 에서는 이러한 음성 변환 알고리즘을 종합적으로 적용하여 한사람이 발음한 음성을 다른 사람이 발성한 음성처럼 들리도록 변환하는 음성 개성 변환(voice personality transformation)알고리즘[1,2,4]을 제안하였다. 음성 개성 변환은 문자-음성 변환 시스템(text-to-speech system)에서 발성되 는 기계적인 합성음을 사용자에게 친숙한 사람의 음성으로 변환하거나[8] 음성 인식 시스템에서의 화자 적응(speaker adaptation)기법[4]등에 이용될 수 있다.

이러한 음성 개성 변환을 구현하기 위해서는, 크게 학습과정(training stage)과 변환과정 (transformation stage)이 필요하다. 학습과정은 변환하고자하는 음성데이타(source speech data) 와 변환에의해 얻고자하는 음성데이타(target speech data)를 미리 학습 데이타로 저장하고 이들 데이타로부터 특징 변수를 추출하여 두 화자간의 변수 대응관계(mapping rule)를 추정하는 과정 이다. 변환 과정은 학습 과정에서 얻어진 대응 관계에 따라 주어진 source 음성으로 부터 변환된 target 음성를 얻는 과정이다. 따라서 음성 개성 변환이 효과적으로 이루어지기 위해서는 특징 변 수의 추출과 대응 관계의 추정이 매우 중요하다. 음성 개성 변환에 널리 쓰이고 있는 특징 변수 는 동일한 음소에 대해서도 각 화자마다 독특한 특성을 나타내는 변수들로서, 성도(vocal tract)의 특성을 나타내는 성도 전달 함수와 피치, 발음 속도와 같은 운율 정보(prosody information)로 분 류할 수 있다. 성도 전달 함수의 변환 방법으로, Abe에 의해 제안된 벡터 양자화를 이용한 음성 변환 알고리즘은 특징 변수로 벡터 양자화된 LPC 계수를 사용하여, source 및 target 화자로 부 터 얻어진 코드 벡터간의 대응 관계를 통해 음성 변환을 수행하였다. Nam등에 의해 제안된 음성 변환 기법[1]은 특징 변수로 LPC 켑스트럼을 사용하고, 신경망(Neural Network)을 이용하여 대응 관계를 모델링 하였다. 이들 방법은 비교적 적은수의 어휘에 대해서는 우수한 변환 성능을 보이 지만, target 화자의 특징 변수가 코드북에 포함된 코드 벡터만으로 표현되므로 제한된 음소에 의 해서만 변환이 수행된다는 단점이 있다. 개선된 방법으로, Valbret[2]등은 source 켑스트럼과 taget 켑스트럼간의 대응관계를 분류화된 선형 변환식(classified lnear transform equation)으로 표현하 는 LMR(Linear Multivariate Regression) 기법과 source 화자의 스펙트럼 엔벨로프를 주파수상의 warping 함수에 의해 target 화자의 엔벨로프로 변환시키는 DFW(Dynamic Frequency Warping) 기법을 제안하였다. 그러나 이러한 알고리즘등은 공통적으로, source 및 target 화자의 특징 변수 와 대응 관계 표현이 동일한 벡터 공간상에서 이루어지므로, 화자 고유의 특성을 반영하여 음성 개성 변환을 구현하지 못한다는 단점을 갖는다.

따라서 본 논문에서는 한 화자의 특징변수 추출과 변환 과정에서 기존의 방법과는 다른 새로 운 방법을 제안하였다. 먼저 각 화자에 대한 LPC 켑스트럼 계수를 직교 벡터 공간상의 하나의 신호 벡터로 모델링하는 기법을 제안하였다. 직교 벡터 공간을 나타내는 각각의 직교축은 source 및 target 화자의 LPC 켑스트럼의 상관 행렬(correlation matrix)을 고유분해(eigen-decompostion) 하여 얻어지는 주요 고유벡터(principle eigenvector)로 표현하였다. 따라서 임의의 화자에 대한 LPC 켑스트럼 벡터는 그 화자로부터 얻어진 고유벡터들의 선형조합(linear combination)으로 표 현된다. 음성 개성 변환은 이러한 화자마다 고유하게 얻어지는 고유벡터를 변환함으로써 구현되 며, 동일한 음소에 대한 source 및 target 화자의 켑스트럼 벡터가 벡터 공간상에서 서로 다른 위 치에 존재할 수 있으므로 최소 자승 공간 변환(minimum mean square error coordinate transformation)[7]을 통하여 source 화자의 켑스트럼 벡터가 target 화자의 켑스트럼 벡터 위치로 정확히 옮겨지도록 하였다. 이러한 방법은 특정 화자의 특징 변수를 그 화자로부터 얻어진 직교 벡터 공 간상에서 표현하므로 화자 고

유의 특성을 반영하여 변환을 수행할 수 있으며, LPC 켑스트럼을 해당 화자로부터 추출된 주요한 직교 벡터들로만 표현하므로, 신호 표현에 필요한 차원 (dimensionality)을 줄일 수 있다는 장점을 갖는다. 또한, 고유벡터 변환계수의 직교성 (orthogonality)에 의하여 LMR 기법에 사용된 선형 변환식에 비해 간략화된 변환식을 이용할 수 있다.

한편 완전한 음성 개성 변환을 위해서는 성도 전달 함수의 변환과 함께 음성의 운율적인 정 보(prosodic information)의 변환도 동시에 이루어져야 한다. 기존에 발표된 운율 변환 기법으로, Abe등은 source/target 화자가 가지고 있는 피치를 몇개의 대표값으로 표현하고, 대표값간의 대응 관계를 이용하여 단구간 피치값만을 변환시키는 방법과, 영상 처리에 사용되는 히스토그램 등화 (histogram equalization)기법을 통해 source 화자의 피치에 대한 확률 분포함수를 target 화자의 확률 분포함수로 변환시키는 기법이 Nam등에 의해 제안되었다. 한편, Valbret[2]등은 음성 합성기의 운율 제어로 널리 사용되고 있는 PSOLA (Pitch Synchronous OverLap and Add)기법을 음성 개성 변환에 적용하여 운율 변환을 구현하였다.

이러한 알고리즘등은 주로 단구간 피치의 변환에 목적을 두고 있으나, 실제 운율 정보는 단 구간 특성보다 한문장 전체에 걸쳐서 나타나는 피치의 움직임등에 좌우되므로, 본 논문에서는 운 율 변환을 구현하기 위해 피치궤적(pitch contour)을 변환하는 기법을 제안하였다. 제안된 방법은 알고리즘의 간편성을 위하여 피치궤적을 장구간 특성과 단구간 특성으로 나누어 변환을 수행하도 록 하였다. 이중 장구간 특성은 학습 데이타의 유성음 구간에서 추정된 피치의 평균값으로 표현하였으며, 이에 따라 변환된 음성의 평균 피치값이 target 음성의 평균값과 동일해지도록 스케일 팩터로 source 화자의 피치를 늘리거나 줄임으로서 장구간 피치 변환을 구현하였다. 단구간 특성 의 변환은 변환 source와 target 화자에 대해 미리 기준 피치궤적(reference pitch contour)을 작 성하고 학습 과정에서 기준 피치궤적간의 대응 관계를 Maximum likelihood 함수를 통해 추정하 여 이를 변환 과정에 사용하도록 하였다. 이를 통하여 변환된 음성은 장구간 특성과 단구간 특성 이 함께 변환되어, 보다 target 화자에 근접한 운율 특성을 얻게 된다.

본 논문의 구성은 다음과 같다. 서론에 이어 2장에서는 제안된 음성 개성 변환 알고리즘의 전체적인 구조를 제시하며 3장에서는 LPC켑스트럼의 변환과 운율 정보의 변환 기법에 대해 살펴 보고 4장에서는 합성 방법을, 그리고 5장에서는 제안된 알고리즘을 이용하여 변환된 음성에 대해 성능을 평가하고 마지막으로 6장에서 결론을 맺었다.

Ⅱ. 제안된 음성 개성 변환 알고리즘

입력된 source 음성 신호로 부터 원하는 target 음성으로의 변환을 시키기 위해서는 그림 1 과 같이 크게 3단계의 과정이 필요하다. 첫번째 과정은 음성 신호의 분석(analysis)과정으로 각 화자의 음성 신호로부터 특징 변수를 추출하는 과정이다. 본 논문에서는 성도 전달 함수 및 운율 정보의 변환을 위하여 LPC 켑스트럼과 기본 주파수를 추출한다. 이때 하나의 분석 프레임(frame) 은 32msec의 길이를 갖으며 프레임 레이트는 8msec로 설정하였다. 또한 스펙트럼 추정시의 누설 (leakage)을 줄이기 위해 분석시 hamming window를 사용하였다.

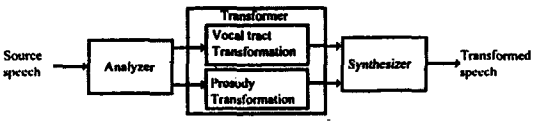


그림 1. 제안된 음성 변환 시스템의 블록도
Fig. 1. Blockdiagram of the proposed voice transformation system.

두번째 과정은 변환과정으로 분석 과정에서 얻어진 source 화자에 대한 특징 변수를 target 화자의 특징 변수로 변경한다. 본 논문에서는 특징 변수로 사용된 LPC 켑스트럼 계수 와 여기 신호의 스펙트럼에 대해 변환을 수행한다. 이때 필요한 변환 규칙(transformation rule)은 학습 과 정(training stage)을 통해 얻어진다. 학습 과정에서는 source와 target 화자로부터 동일한 음성 데 이타를 취한다음 두 음성간의 대응 관계를 추정하여 변환 규칙을 얻게 된다.

마지막으로 합성(synthesis)과정에서는 변환된 LPC켑스트럼과 여기 신호를 이용하여 단구간 음성 신호를 얻고, 여기서 얻어진 각 프레임 단위의 변환 음성 신호는 SOLA(synchronized overlap-add)기법[3]을 이용하여 전체 음성 신호를 구성하게 된다.

Ⅲ. 변환 과정

본 논문에서는 음성 신호의 특징 변수로 사용되는 LPC켑스트럼 계수를 직교 기저 벡터 (orthogonal basis vector)의 선형 조합에 의해 표현되는 신호 벡터로 모델링 하였다. 이때 사용되 는 기저 벡터는 한 화자의 LPC 켑스트럼 벡터를 고유 분해하여 얻어지는 고유 벡터이다. 따라서 신호 벡터의 표현에 필요한 기저 벡터는 각 화자마다 고유하게 얻어진다.

성도 전달 함수의 변환은 이러한 모델링 기법에 근거하여 이루어진다. 즉 그림 2에서 나타낸 것처럼 source 화자의 고유 벡터 공간에서 존재하는 source 화자에 대한 LPC 켑스트럼 벡터를 target 화자의 고유 벡터 공간으로 옮김으로서 성도 전달 함수의 변환을 구현할 수 있게 된다. 이 때 필요한 것은 source 및 target 화자에서 얻어진 학습 데이타로 부터 고유 벡터를 취하는 과정 과 soure 벡터 공간상의 임의 벡터를 target 벡터 공간상의 적절한 위치로 옮기는 선형 변환식을 추정하는 과정이다. 이 과정들에 대해 살펴보면 다음과 같다.

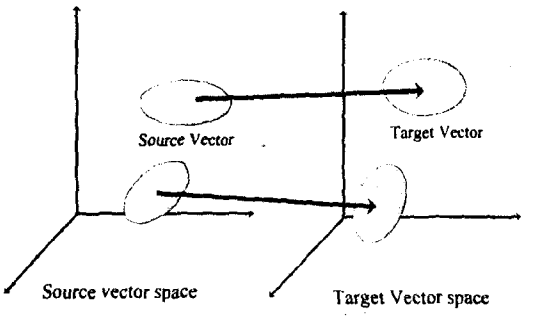


그림 2. 켑스트럼 벡터 변환의 기하학적 표현
Fig. 2. Graphical explanation of the vocal tract transformation.

1. 직교 기저 벡터의 추정

source 및 target 화자로 부터 얻어지는 LPC 켑스트럼 계수를 행렬(column matrix)로 나타내면 아래와 같다.

$$\mathbf{C}_i^s = [\, c_i^s(1)\ c_i^s(2)\ \ldots\ c_i^s(P)\ ]^T,\ i=1,2,\ldots,N^s \quad (1\text{-a})$$

$$\mathbf{C}_j^t = [\, c_j^t(1)\ c_j^t(2)\ \ldots\ c_j^t(P)\ ]^T,\ j=1,2,\ldots,N^t \quad (1\text{-b})$$

여기서 $\mathbf{C}_i^s$와 $\mathbf{C}_j^t$는 각각 source 및 target 화자에 대한 i-번째, j-번째 켑스트럼 벡터를 나타내며 $c_i^s(n)$은 n번재 켑스트럼 계수를 나타낸다. 각 켑스트럼 벡터에 대한 상관행렬(correlation matrix)은 아래와 같다.

$$\mathbf{R}^s = \frac{1}{N^s}\sum_{i=1}^{N^s}(\mathbf{C}_i^s - \mathbf{m}^s)(\mathbf{C}_i^s - \mathbf{m}^t)^T \quad (2\text{-a})$$

$$\mathbf{R}^t = \frac{1}{N^t}\sum_{i=1}^{N^t}(\mathbf{C}_i^t - \mathbf{m}^t)(\mathbf{C}_i^t - \mathbf{m}^s)^T \quad (2\text{-b})$$

N^s 및 N^t는 각각 source 및 target 화자의 전체 학습 데이타수를 나타내며 $\mathbf{m}^s$, m^t는 각각 source 및 target 화자의 켑스트럼 평균 벡터를 나타낸다. 위의 상관행렬이 positive definite 하다면 이 행렬에 대한 고유 분해(eigen-decomposition)는 다음과 같이 주어진다.

$$\mathbf{R}^s = \mathbf{Q}^s\ \boldsymbol{\Lambda}^s\ \mathbf{Q}^{sT} \quad (3\text{-a})$$

$$\mathbf{R}^t = \mathbf{Q}^t\ \boldsymbol{\Lambda}^t\ \mathbf{Q}^{tT} \quad (3\text{-b})$$

이때 $\mathbf{Q}^s, \mathbf{Q}^t$는 각각 source 및 target화자에 대한 고유 벡터 행렬(eigenvector matrix)을 나타내며 $\boldsymbol{\Lambda}^s$, $\boldsymbol{\Lambda}^t$는 고유 벡터에 대응되는 고유치(eigenvalue)로 구성된 대각행렬(diagonal matrix) 이다. 직교 기저 벡터는 고유 벡터중에서 고유치 λ_i^s, λ_i^t가 임계값 λ_{th}이상인 고유 벡터이다. 따라서 source와 target화자에 대한 기저 벡터 집합 V^s, V^t는 다음과 같이 나타낼 수 있다.

$$V^s = \{\mathbf{e}_i^s : \lambda_i^s \geq \lambda_{th}\} \quad (4\text{-a})$$

$$V^t = \{\mathbf{e}_i^t : \lambda_i^t \geq \lambda_{th}\} \quad (4\text{-b})$$

이때 임계값 λ_{th}은 기저 벡터 집합에 포함된 고유 벡터만으로 LPC 켑스트럼을 구성하였을때 청각상 왜곡이 느껴지지 않도록 설정하였다. 위의 과정으로 얻어진 기저 벡터를 이용하여 임의의 LPC 켑스트럼 벡터를 나타내면 아래식과 같이 KL(Karhunen-Loève)급수 전개 형태로 표현된다.

$$\mathbf{C}_i^s \cong \sum_{n=1}^{N} s_n^i \mathbf{e}_n^s \quad (5\text{-a})$$

$$\mathbf{C}_j^t \cong \sum_{m=1}^{M} t_m^i \mathbf{e}_m^t \quad (5\text{-b})$$

여기서 N, M은 기저 벡터의 총 갯수를 나타내며 s_n^i, t_m^i는 각각 n번재, m번째 기저 벡터에 대한 크기

를 나타낸다. 식 (5)로 표현되는 LPC 켑스트럼 벡터는 본래의 값과 약간의 차이를 나타 내는데 그것은 직교 기저 벡터 집합이 임계치 이상의 고유 벡터만으로 구성되어 있기 때문이다. 그러나 실험적인 결과에 따르면 충분히 작은(<0.01) 고유치를 갖는 고유 벡터를 제외하는 경우에 는 식(5)로 표현되는 LPC 켑스트럼과 본래의 켑스트럼간에 차이가 거의 없음을 알 수 있었다.

2. 제안된 모델링의 해석

본 논문에서는 특정 화자의 LPC 켑스트럼을 나타내기 위해 해당 화자의 LPC 켑스트럼에 대 한 상관 행렬을 고유 분해하여, 여기서 얻어지는 고유 벡터를 선형 조합하여 표현하는 방법을 사용하였다. 2차원 공간상의 임의 랜덤 벡터에 대해, 상관 행렬의 고유 벡터를 구하면 그림 3과 같이 벡터간 직교성(orthogonality)을 유지하면서 랜덤 벡터가 최대의 에너지를 갖는 방향으로 향하 게 된다[14]. 이를 LPC 켑스트럼 벡터에 적용하면, 각 고유 벡터가 서로 직교하면서 최대 에너지를 갖는 벡터 방향으로 존재함을 알 수 있다.

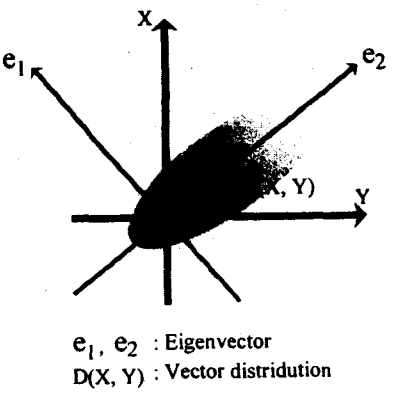


그림 3. 고유 벡터와 신호 분포와의 관계
Fig. 3. Relation between signal distribution and its eigenvector.

따라서 본 논문에서 제안된 고유 벡터에 의한 LPC 켑스트럼 표현은 화자가 가지고 있는 LPC 켑스트럼 벡터의 분포 형태를 고려한 모델링 방법으로, 음성 변환시 source 화자에 대한 고 유 벡터를 target 화자의 고유 벡터로 변환하는 것은 source 화자의 LPC 켑스트럼 벡터의 분포를 target 화자의 분포 형태로 변환함을 의미한다. 이러한 모델링에 의해 변환된 음성의 LPC 켑스트 럼은 확률 분포 함수적으로 target 음성 신호의 유사한 특성을 갖게 된다.

또한, 신호 표현에 사용되는 기저 벡터를 화자의 고유 특성에 따라 사용하게 되므로 LPC켑 스트럼을 그대로 변환 변수로 이용하는 기존의 방법에 비해 몇가지 장점을 갖게된다. 첫번째로 신호 표현에 필요한 차원(dimensionality)을 감소시킬 수 있다. 즉, 그림 4에 제시한 바와 같이 임 의 화자에 대한 LPC 켑스트럼의 에너지와 고유 벡터의 변환 계수 에너지 분포는 변환 계수쪽이 더욱 낮은 차수쪽에 집중함을 알 수 있다. 0.01이상의 에너지를 갖는 계수의 수를 신호 표현에 필 요한 최소의 차원수라고 가정할때, 실험적인 결과에 따르면 LPC 켑스트럼 계수의 경우는 약 15 차원이, 고유 벡터의 변환 계수는 약 11차원이 얻어짐을 알 수 있었다.

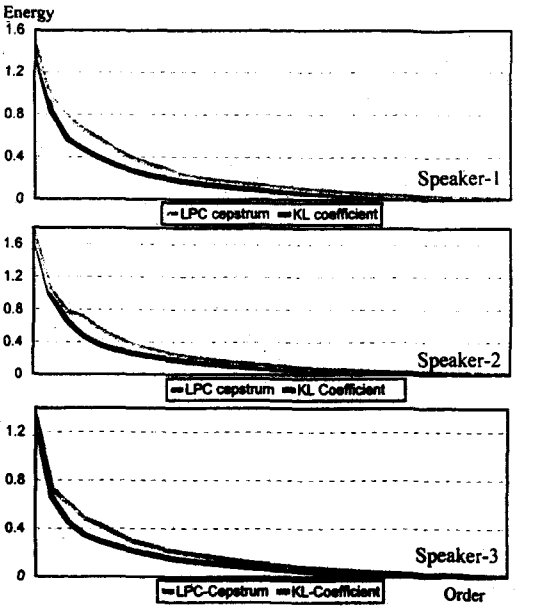


그림 4. LPC 켑스트럼 계수와 변환 계수의 에너지 분포
Fig. 4. Energy distributions of LPC cepstrum and transform coefficients.

두번째로 고유 벡터의 변환 계수는 계수간의 상관성이 존재하지 않게되어[6] 변환식이 간단하 게 구성되는 장점을 갖는다. 즉, 변환에 사용되는 고유 벡터의 변환 계수는 자신과의 상관성만 존 재하게 되어 다른 변환 계수간의 상관 성분을 고려하지 않아도 최적의 변환식을 얻을 수 있다. 이 경우, 변환 변수의 차원이 N인 경우 상관성이 존재하는 LPC 켑스트럼의 변환시 N^2의 연산이 필요하나, 고유 벡터의 변환 계수를 이

용하는 경우 N번의 연산이 필요하게 되어 변환시 계산량을 크게 감소 시킬 수 있다.

3. 선형 변환식의 추정

최적의 LPC 켑스트럼 계수의 변환은 특정 음소에 대해 변환된 source 화자의 LPC 켑스트럼 과 target 화자의 LPC 켑스트럼간의 차이가 최소화되도록 하는 것이다. 본 논문에서는 앞서 제시 한 모델링 방법에 근거하여 source 화자의 벡터 공간상의 신호 벡터를 이에 대응되는 target 화 자의 벡터 공간상 벡터로 이동시킴으로서 LPC 켑스트럼 변환을 구현하게 된다. 이러한 이동은 선형 변환식(linear transformation)에 의해 이루어지며 변환식은 학습 과정에서 얻어진다.

학습에 사용되는 음성 데이타는 source 및 target 화자가 동일하게 발음한 몇개의 문장으로 구성된다. 학습 음성 데이타로부터 먼저 LPC분석 과정을 통하여 LPC 켑스트럼 계수를 추출한다. 본 논문에서는 자기상관 방법을 이용하여 LPC 계수를 추정하고 이를 켑스트럼으로 바꾸는 방법 을 이용하였다. 여기서 얻어지는 LPC 켑스트럼 열(sequence)은 두화자에 대해 공통적인 음소를 포함하고 있지만 두 화자의 발음 속도가 다를 수 있으므로 음소의 위치가 시간적 불일치할 수 있다. 따라서 DTW(dynamic time warping)[4]를 통해 시간적으로 동일한 음소가 대응되도록 하였다.

시간축 정열된 source/target LPC 켑스트럼 벡터는 다음 과정으로, 고유 분해를 통해 얻어진 기저 벡터에 사영(projection)된다

$$s_n^j = \sum_{i=1}^{P} e_n^s(i) c_j^s(i) , \qquad 1 \le n \le N. \tag{6-a}$$

$$t_m^j = \sum_{i=1}^{P} e_m^t(i) c_j^t(i) , \qquad 1 \le m \le M. \tag{6-b}$$

여기서 s_n^j 및 t_m^j는 켑스트럼 벡터 $\mathbf{C}_j^s$ 및 $\mathbf{C}_j^t$를 구성하는 n번째 및 m번째 기저 벡터의 크기를 나타낸다. 이 값은 source/target 벡터 공간에 놓여있는 켑스트럼 벡터의 각 좌표값을 나타 낸다. 이때 동일시간에서의 source 및 target 켑스트럼 벡터는 미리 시간 정렬이 되어 있는 상태 이므로 동일한 음소를 나타내지만 각각의 벡터 공간상에서는 서로 다른 위치에 존재할 수 있다. 따라서 target 벡터 공간상의 목표 위치로 source 벡터를 이동시키기 위해서는 각 좌표의 위치를 변경할 필요가 있다. 이를 위해서 본 논문에서는 아래와 같은 최소 자승 오차 좌표 변환 (minimum mean sqaure error coordinate transformation) 방법[7]이 사용되었다.

$$\hat{t}_m^j = \sum_{n=1}^{N} h_{mn} s_n^j + o_m , \qquad 1 \le m \le M. \tag{7}$$

$\hat{t}_m^j$는 m번째 기저 벡터에 대한 변환값을 나타낸다. 최적의 변환 계수 h_{mn} 및 o_m 값은 변환값 $\hat{t}_m^j$와 시간 정렬된 target 켑스트럼 벡터 t_m^j간의 평균 자승 오차(mean square error)가 최소화 되도록 얻어진다. 평균 자승 오차는 아래와 같다.

$$\xi = \frac{1}{N_s} \sum_{j=1}^{N_s} \sum_{m=1}^{M} (\hat{t}_m^j - t_m^j)^2 \tag{8}$$

평균 자승 오차가 h_{mn}, o_m 값에 대해 convex함수 형태를 갖으므로 최적의 h_{mn}^*, o_m^*은 아래식을 만족한다.

$$\frac{\partial \xi}{h_{mn}} = 0 , \quad \frac{\partial \xi}{o_m} = 0 \tag{9}$$

식 (7)과 식(8)을 식 (9)에 대입하면 아래와 같다.

$$\sum_{j=1}^{N_s} t_m^j s_n^j = \sum_{j=1}^{N_s} \sum_{k=1}^{N} h_{mk}^* s_k^j s_n^j + \sum_{j=1}^{N_s} o_m^* s_n^j \qquad 1 \le m \le M, \ 1 \le n \le N. \tag{10}$$

$$\sum_{j=1}^{N_s} t_m^j = \sum_{j=1}^{N_s} \sum_{k=1}^{N} h_{mk}^* s_k^j + \sum_{j=1}^{N_s} o_m^* \qquad 1 \le m \le M. \tag{11}$$

식(10)과 식(11)은 아래와 같이 행렬식 형태로 표현할 수 있다.

$\mathbf{A}\mathbf{x}_m = \mathbf{b}_m$ 행렬 $\mathbf{A}, \mathbf{x}_m, \mathbf{b}_m$의 각 성분을 나타내면 아래와 같다.

$$\begin{bmatrix} a_{11} & \cdots & a_{1N} & a_{1N+1} \\ \vdots & \ddots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NN} & a_{NN+1} \\ a_{N+11} & \cdots & a_{N+1N} & a_{N+1N+1} \end{bmatrix} \begin{bmatrix} x_1^m \\ \vdots \\ x_N^m \\ x_{N+1}^m \end{bmatrix} = \begin{bmatrix} b_1^m \\ \vdots \\ b_N^m \\ b_{N+1}^m \end{bmatrix}, \quad 1 \le m \le M. \tag{12}$$

$$a_{kn} = \sum_{j=1}^{N_s} s_k^j s_n^j , \qquad 1 \le k \le N, \ 1 \le n \le N.$$

$$a_{nN+1} = a_{N+1n} = \sum_{j=1}^{N_s} s_n^j , \qquad 1 \le n \le N.$$

$$a_{N+1N+1} = N_s$$

$$x_n^m = h_{mn}^* , \qquad 1 \le n \le N , \ 1 \le m \le M.$$

$$x_{N+1}^m = o_m^*$$

$$b_n^m = \sum_{j=1}^{N_s} t_m^j s_n^j , \qquad 1 \le n \le N , \ 1 \le m \le M.$$

$$b_{N+1}^m = \sum_{j=1}^{N_s} t_m^j$$

식 (12)에 따라 최적 변환 계수 $h^*_{mn}o^*_m$는 아래의 행렬식으로 구해진다.

$$\mathbf{x}_m = \mathbf{A}^{-1}\mathbf{b}_m , \qquad 1 \le m \le M. \qquad (13)$$

위의 식중 행렬 **A**의 대각 성분과 N행, N열 성분을 제외한 나머지 성분은 앞장에서 제시한 바와 같이 KL변환 계수의 직교성에 따라 0이 되며 이에 따라 행렬 $\boldsymbol{x}$는 간단하게 구해질 수 있다. 학습 과정에서 산출되는 $h^*_{mn}o^*_m$은 변환 과정시 식 (7)에 대입되어 을 얻는데 사용되며 변환 켑스트럼은 아래식 (14)와 같이 target 화자에 대한 기저 벡터와 f_m^j의 선형조합으로 얻어진다. 이 과정을 그림 5에 나타내었다.

$$\hat{C}_j^t = \sum_{m=1}^{M} f_m^j e_m^t + m^t \qquad (14)$$

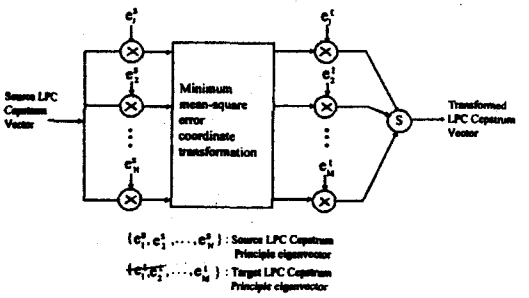


그림 5. 켑스트럼 벡터 변환의 블록도
Fig. 5. Bolckdiagram of the vocal tract transformation.

1) 클래스 분류에 의한 최적 변환

위에서 제시한 방법은 모든 켑스트럼 벡터에 대해 단 하나의 변환 계수만을 사용하여 변환을 수행한다. 그러나 source 화자의 LPC켑스트럼에 대해 클래스 분류를 하고 분류된 각 클래스에 대해 최적의 변환 계수를 사용한다면 좀더 우수한 변환 성능을 기대할 수 있다. 이 경우 입력된 source 켑스트럼 벡터에 대해 클래스 분류를 위한 탐색(search)과정이 필요하므로 부가적인 계산 이 요구되지만, 적절한 클래스 수를 선택함으로서 계산량을 크게 늘리지 않으면서 향상된 성능을 얻을 수 있다. 본 논문에서는 켑스트럼 벡터의 클래스 분류를 위해 LBG알고리즘[10]을 사용하였으 며 벡터간의 거리는 아래식을 이용하였다.

$$d_{ij} = \frac{1}{N}\sum_{n=1}^{N}(s_n^i - s_n^j)^2 \qquad (15)$$

위의 식에서 보듯이 본 논문에서는 클래스 분류도 켑스트럼 벡터에 대해 직접적으로 이루어 지는 것이 아니고, 고유 벡터 공간상의 거리를 이용하여 이루어지도록 하였다. 이 경우 학습 과정 과 변환 과정에 대해 살펴보면 다음과 같다. 학습 데이타에 대해 LBG알고리즘을 적용하면 각 클 래스의 중심 벡터(centroid vector)가 생성되고 모든 학습 데이타에 대해 클래스 분류를 수행한다. 이때 동일한 클래스로 분류된 벡터들로만 학습 데이타를 구성하여 위의 과정 (6)~(13)을 수행하고 이를 클래스 수만큼 반복적으로 수행하여 클래스마다 최적의 변환 계수들을 얻는다. 변환시에는 먼저 source 화자의 켑스트럼 벡터를 기저 벡터에 사영시키고 여기서 얻어진 값을 이용하여 클래 스 분류를 한후 해당 클래스에 해당하는 $h^*_{mn}o^*_m$를 이용해 변환을 수행한다.

본 논문에서는 최적의 클래스수를 설정하기 위하여 클래스수를 증가시키며 음성 변환의 성능 을 평가하였다. 이 결과를 그림 6에 제시하였다. 그림 6에 제시된 바와 같이 클래스 수가 증가함 에 따라 변환 켑스트럼과 taget 켑스트럼간의 차이가 감소하지만 클래스 수가 32개를 넘어서면서 감소량이 현저하게 완만함을 알 수 있다. 따라서 본 논문에서는 계산량과 전체적인 변환기의 성 능을 고려하여 클래스수를 32개로 설정하였다.

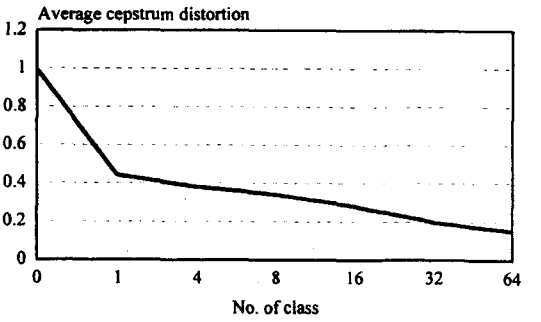


그림 6. 클래스수에 따른 변환음과 target음성간의 왜곡
Fig. 6. Distortion between the transformed and target speech corresponding to number of class.

4. 운율 변환

음성의 운율적인 특징(prosody characteristic)은 발음 속도, 피치의 시간적 변화, 에너지 궤적, 포만트(formant) 주파수의 시간적 변화로 설명될 수 있으며[8,9] 한 개인의 독특한 발성 스타일 (speaking

style)을 나타내는 경우가 많다. 따라서 특정 화자의 운율적인 특징을 다른 사람의 것 으로 변경할 수 있다면 음성의 개성 또한 변경할 수 있을 것이다. 그러나 한사람의 발성 스타일 은 위에 제시한 변수외에도 음향적 특징으로는 설명될 수 없는 다양한 요소에 의해 좌우되므로 완전한 운율 변환을 수행하기는 매우 어렵다. 본 논문에서는 운율적 특징을 비교적 크게 좌우하는 피치 궤적만을 변환 파라메터로 사용하였다.

이러한 운율 변환을 위해서는 성도 전달 함수의 변환 과정과 마찬가지로 특징 변수의 추출과 두 화자에 대한 대응 관계의 추정이 필요하다. 운율 변환의 특징 변수인 피치 궤적을 본 논문에서는 크게 장구간 특성(long time characteristic)과 단구간 특성(short time characteristic)으로 나누어 표현하였다. 이 중 장구간 특성은 화자가 가지고 있는 피치의 전체적인 특성으로 본 논문에 서는 유성음 구간에서 추정된 피치의 전체 평균값으로 나타내었다. 따라서 장구간 피치 특성의 변환은 source 화자의 평균 피치값이 target의 값과 동일한 값을 갖도록 source 화자의 각 피치값에 스케일 펙터 β를 곱함으로서 이루어지게 된다. β는 아래와 같은 과정으로 얻게 된다.

$$\hat{p}^{(t)} = \beta p^{(s)} \tag{16}$$

$$E(\hat{p}^{(t)}) = \beta E(p^{(s)}) = p_{ave}^{(t)} \tag{17}$$

$$\beta = \frac{p_{ave}^{(t)}}{p_{ave}^{(s)}} \tag{18}$$

여기서 E(x)는 변수 x의 평균값을 나타내며 $p_{ave}^{(s)}$, $p_{ave}^{(t)}$는 각각 학습 데이타로부터 얻어진 source, target 화자의 평균 피치값을 나타낸다.

장구간 특성의 변환은 음성의 전체적인 억양을 변환할 수 있으나 구간별 피치 패턴의 특성을 함께 변환하지 못하는 단점이 있다. 따라서 피치 패턴의 지역적인 특성(local characteristic)을 반 영한 단구간 특성의 변환이 함께 이루어져야 한다. 그러나 한 문장에 대한 전체 피치 궤적을 변 환하는 경우, 피치 궤적의 패턴이 매우 다양하게 나타나므로 각각의 경우에 대한 최적 변환 패턴을 찾기는 매우 복잡하다. 이를 위해 본 논문에서는 주어진 문장으로부터 운율 단위(prosody segment)를 분할하여 변환을 수행하는 방법을 제안하였다. 이때 전체 피치 궤적으로부터 운율 단 위를 분리하기 위해 레이블링 과정이 필요하다. 본 논문에서 사용한 레이블링 방법은 비교적 큰 정밀도가 요구되는 학습과정에서는 입력된 문장으로부터 수작업에 의해 운율구(prosody phrase) 를 분할하였으며 변환시에는 그림 7에 제시한 바와같이 피치값이 하강에서 상승으로 변화하고 에너지 궤적의 급격한 변화가 일어나는 지점을 기준으로 피치 궤적을 분할하였다. 그림 7의 P(n)은 주어진 문장에 대한 n번째 운율 단위를 나타낸다.

분할된 각 운율 단위는 DTW에 의해 시간 정렬되어 source 피치 궤적에 대응되는 target 피치 궤적을 얻는다. 시간 정렬된 각 운율 단위는 위에서 구한 beta에 의해 장구간 변환이 이루어 지고 target 피치 궤적과 장구간 변환된 값과의 차를 얻는다.

$$P_d(L) = P_t(L) - \beta P_s(L) \tag{19}$$

여기서 $P_s(L)$과 $P_t(L)$은 각각 time index L에서의 source, target 피치 궤적을 나타낸다. 차궤적 $P_d(L)$은 장구간 변환후에 생성되는 잔차 피치 궤적(residual pitch contour)으로 본 논문 에서는 이를 벡터 양자화에 의해 표현하였다. 이와 동시에 $P_s(L)$ 또한 벡터 양자화를 수행하여 피치 궤적에 대한 대응 관계는 각 벡터의 인덱스값으로 표현한다. 각기 다른 길이를 갖는 피치 궤 적에 대해 코드북을 구성할 경우 코드 벡터의 수가 크기가 증가하므로 본 논문에서는 피치 궤적 을 동일한 길이로 정규화하여 코드북을 구성하였다. 따라서 벡터 양자화를 수행하기전에 피치 궤적의 길이를 동일하게 맞추는 interpolation/decimation 과정이 필요하다. 두화자의 피치 궤적간 대응 관계를 나타내기 위해 본 논문에서는 likelihood 함수를 사용하였다. 주어진 source 화자의 피 치 궤적이 j번째 코드 벡터로 양자화되었다면 최적의 잔차 피치 궤적은 아래식으로 주어진다.

$$P_d^{j*} = \arg\max_k p(P_d^k | P_s^j) \ . \tag{20}$$

P_d^k는 잔차 피치 궤적에 대한 k번째 코드 벡터를 나타낸다. 또한 p(|)는 조건 확률 분포함 수를 나타내며 이 값은 학습과정에서 각각의 P_s^j에 대해 얻어진다. 위의 식은 결국 임의의 source 피치 궤적이 주어졌을 때 가장 발생 빈도가 높은 잔차 피치 궤적을 최적의 피치 궤적으로 선택함을 알 수 있다.

변환 과정에서는 위에서 구한 조건 확률 분포함수

및 스케일 펙터 β를 이용하여 각 프레임에 대한 변환 피치값을 얻는다.

$$\hat{P}^{(t)}(i)=\beta P_s(i)+P_d^*(i)\ ,\ P_d^*(i)=\arg\max_k(P_d^k|\ \hat{P}_s(i)) \quad (21)$$

여기서 $\hat{P}_s(i)$는 벡터 양자화된 source 피치 궤적을 나타낸다. 본 논문에서는 source 화자에 대한 피치값으로부터 식 (21)에 따라 최적의 변환 피치값을 얻고 이 값과 본래 피치값과의 비 α를 구하여 여기 신호의 스펙트럼을 주파수축에서 스케일링함으로서 피치 변환을 구현하였다.

$$\alpha=\frac{\hat{p}^{(t)}(m)}{p^{(s)}(m)}\ . \quad (22)$$

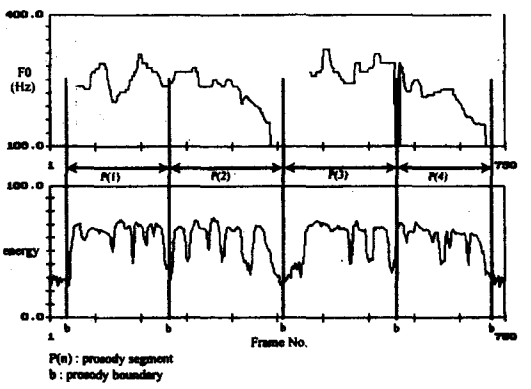


그림 7. 운율 단위의 분할
Fig. 7. Splition of the prosody segment.

Ⅳ. 합성 과정

변환된 음성 신호의 합성은 먼저 변환 켑스트럼 벡터로 부터 성도 전달 함수를 구하고 여기 에 스케일링된 여기 신호의 스펙트럼을 곱한후 퓨리어 역변환을 수행하여 단구간 변환 음성 신호 $\hat{s}(n-mS)$를 얻는다.

$$\hat{s}(n-mS)=\mathrm{IDFT}\{\ \hat{H}_t(\omega,mS)E(\alpha\omega,mS)\} \quad (23)$$

위식에서 S는 분석 프레임 레이트를 나타내며 m은 분석 시간 인덱스를 나타낸다. 또한 $\hat{H}_t(\omega,mS)$는 변환된 LPC 켑스트럼으로부터 얻어진 성도 전달 함수를, $E(\alpha\omega,mS)$은 스케일링된 여기신호의 스펙트럼을 나타낸다. 변환된 단구간 음성 신호로부터 전체 음성 신호를 합성하기 위하여, Griffin과 Lim에 의해 제안된 LSEE-MSTFT[15](Least Square Error Estimation-Modified Short Time Fourier Transform)기법을 이용할 수 있으나, 여기 신호의 변환에 따른 위상 및 길 이 정합 문제를 고려해야 한다. 이는 여기 신호 스펙트럼의 확장 또는 축소에 따른 합성 창함수 의 축소 또는 확장에 따라 전체 음성 신호의 길이가 변경되며, 피치 변환에 따른 피크 위치의 변 경에 따라 인접 프레임간의 위상이 불일치하게 되기 때문이다. 에 사용되는 창함수가 이때 스케 일링된 여기신호는 기본 주파수의 변경에 따라 인접 프레임간의 위상이 서로 불일치하게 된다. 따라서 본 논문에서는 각 프레임간의 위상 정합을 위해 SOLA (synchronized overlap-add)기법[3] 을 사용하여 최종적인 변환 음성을 합성하였다.

Ⅴ. 모의 실험 및 결과

제안된 음성 변환 알고리즘의 성능을 평가하기 위해서 몇명의 화자를 대상으로 음성 변환을 수행하여 성능을 평가하였다. 실험에 사용된 음성 데이타는 연세대학교 신호처리 연구실에 재학 중인 4명의 남성 화자로부터 수집하였으며 각 음성 데이타는 수집된 화자의 영문 이름 첫글자를 따서 KKS, KHG, SJT, YJH로 나타내었다. 모의 실험시 조건을 표1에 나타내었으며 실험에 사용한 음성 데이타는 한국어에서 사용 빈도수가 높은 음소를 골고루 포함하고 있는 문장들로서 아래 와 같다.

문장(1) 음성은 인간과 기계 사이에 가장 효과적인 정보 전달 수단의 하나이다.
문장(2) 새터는 문화적으로 본다면 서울의 중심부이다.
문장(3) 현대는 문화 예술의 시대이다.
문장(4) 차의 법도는 사람의 흐트러진 마음가짐을 바로 잡아 준다.
문장(5) 오늘은 어제의 열매이며 내일의 씨앗이다.
문장(6) 새터는 교통이 편리한 커다란 대학촌이다.
문장(7) 분수처럼 흩어지는 푸른 종소리.
문장(8) 파란 이파리 사이로 함초롬한 꽃망울이 피어난다.

이중 문장(1)~(5)는 변환 규칙의 생성에 필요한 학습 데이타로 사용하였으며 문장(6)~(8)은 실 제 변환을 수행하는데 사용하였다. 음성 변환은 비교적 큰 켑스트럼 거리(cepstrum distance)를 가지고 있는

KHG-KKS 음성데이타와 SJT-YJH간에 수행을 하였다. 객관적인 성능 평가를 위한 척도로 본 논문에서는 식(24)으로 주어지는 target 음성과 변환 음성간의 평균 켑스트럼 왜곡 (average cepstrum distortion)을 이용하였다.

$$D_c = \frac{1}{N_s} \sum_{t=1}^{N_s} \sum_{k=1}^{P} (c_i^t(k) - \hat{c}_i(k)^t)^2 \qquad (24)$$

표 1. 실험 조건
Table 1. Experimental condition.

A/D conversion	8KHz, 16bit
LPC order	12
LPC cepstrum order	20
No. of training sentence	5
No. of training words	32
codebook size for LPC cepstrum	32
codebook size for pith contour	32
Pitch estimation method	clipped autocorrelation method

$c_i^t(k)$, $\hat{c}_i(k)$은 각각 시간 정렬된 target 화자의 켑스트럼 계수와 제안된 알고리즘에 의해 변환된 켑스트럼 계수를 나타낸다. 화자 KHG-KKS간의 변환과 SJT-YJH간의 변환 두가지 경우 에 대한 평균 켑스트럼 왜곡을 표2에 제시하였다. 표에 보듯이 클래스 분류를 하지않은 경우 (class수=1)에도 켑스트럼 왜곡은 변환전과 비교하여 KHG-KKS변환시 58.1%, SJT-YJH변환시에 44.3%로 감소하였음을 알 수 있다. 클래스수가 32인 경우에는 각각의 변환에 대해 80.2%, 63.4% 로 높은 감소율을 나타낸다. 이는 결국 제안된 알고리즘에 의해 source 화자의 켑스트럼 벡터가 target 화자의 켑스트럼 벡터에 가까워짐을 나타내고 있다.

그림 8에 동일한 음성 데이타에 대한 source화자, target화자 및 제안된 알고리즘에 의해 변 환된 음성의 스펙트로그램을 나타내었다. 변환음은 target화자의 음성과 대체적으로 비슷한 포만 트 주파수와 기본 주파수를 나타내고 있다. 그러나 스펙트로그램상으로 변환음은 source화자나 target화자의 음성에 비해 약간의 왜곡이 발생하였음을 알 수 있다.

(a) Source

(b) Target

(c) Transformed

그림 8. 스펙트로 그램 비교
Fig. 8. Spectrogram comparison.

객관적인 평가와함께 제안된 알고리즘의 주관적인 성능 평가를 위해 ABX 테스트를 수행하 였다. ABX 테스트는 청취자에게 먼저 A, B두 화자의 음성을 들려주고 세번째로 변환된 음성을 들려주어 마지막으로 청취한 음성이 어느쪽에 가까운지를 A, B두 화자중에 선택하는 것이다.

ABX 테스트시의 실험 대상은 음성 신호 처리 연구와 실험에 대한 경험이 많은 연세대학교 대학 원생 15명으로 구성하였다. 이에 대한 실험 결과를 표3에 나타내었다. 표3에 나타난 적중률은 KHG-KKS변환시 82.7%, SJT-YJH변환시 69.1%로서 전자의 경우에는 비교적 높은 값을 나타내 지만 후자의 경우에 값이 조금 떨어짐을 알 수 있다. 이는 평균 켑스트럼 왜곡에서도 KHG-KKS 변환이 우수한 성능을 나타낸것으로도 짐작할 수 있으나 가장 큰 원인은 발성 스타일의 차이에 있다고 생각된다. KHG, KKS 두화자의 발성 문장은 전체적으로 피치의 궤적도 완만하고 발음속 도도 비슷하게 나타나지만 SJT, YJH 화자에 있어서는 평균 피치 변화율이 1.3으로 상당히 적은 편이나 구간별 피치 궤적의 차이가 크고, 발음 속도에서도 큰 차이

를 보였으며 각 화자마다 독특 한 발성 스타일을 나타내고 있다. 이러한 사실은 ABX 테스트시에 청취자의 선택 기준이 LPC 켑 스트럼 파라메터나피치와 같은 음향적인 특성보다는 발성 스타일에 좌우되어 객관적인 평가 기준 상으로 target 음성에 가까워도 청취상으로는 source에 가까운 경우가 발생할 수 있음을 의미한 다. 따라서 보다 완전한 개성 변환을 위해서는 음향적인 특징 변수이외의 발성 스타일을 좌우하 는 에너지 궤적, LPC 켑스트럼 계수의 동적인 특성등도 함께 변환을 수행해야함을 알 수 있다.

표 2. 평균 켑스트럼 왜곡
Table 2. Average cepstrum distortion.

No. of class	KHG-to-KKS	SJT-to-YJH
no conversion	0.9941	0.8790
1	0.4402	0.5109
8	0.3391	0.3940
16	0.2822	0.3312
32	0.1965	0.3190

표 3. ABX 테스트 결과
Table 3. Result of ABX test.

Experiment	correct identification ratio
KHG-to-KKS	82.7%
SJT-to-YJH	69.1%

VI. 결 론

본 논문에서는 한사람의 음성을 다른 사람이 발성하는 것처럼 변환하는 음성 개성 변환 기법 의 한 방법으로 벡터 공간의 개념을 도입한 새로운 알고리즘을 제안하고 모의 실험을 통해 성능 을 평가해 보았다. 임의 화자에 대한 LPC켑스트럼 벡터를 고유 벡터 공간상의 신호 벡터로 모 델링하고 벡터 공간의 변환과 좌표 변환을 이용하여 성도 전달 함수의 변환을 구현하였으며 피치 궤적의 변환을 통해 개인의 운율적인 특징도 함께 변환하였다. 제안된 알고리즘의 객관적 성능 평가를 위해 변환음과 taget 음성간의 켑스트럼 왜곡을 측정하였으며 변환전과 비교하여 80%정 도 왜곡이 감소함을 알 수 있었으며 주관적 성능 평가를 위한 ABX 테스트에 있어서도 청취자의 상당수가 변환음을 target 음성으로 인식하고 있음을 알 수 있었다.

제안된 알고리즘은 운율 변환의 파라메터로 피치만을 이용하고 있으나 보다 완전한 변환을 위해서는 발음 속도, 포만트 주파수의 시간적 변화, 에너지 궤적등을 변환 파라메터에 포함시켜야 하며 변환음에서 들리는 왜곡을 감소하기 위해 새로운 여기 신호 모델링 방법과 변환 기법등이 제안되어야 할 것으로 생각된다.

참 고 문 헌

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

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

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

[4] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall Inc, 1978.

[5] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *proc. of ICASSP*, vol. 2, pp. 355-358, 1993.

[6] G. Strang, *Linear Algebra and its applications*, Academic Press Inc., 1980.

[7] S. A. Zahorian and A. J. Jagharghi, "Minimum mean-square error transformations of categorical data to target positions," *IEEE Trans. on Signal Processing*, vol. 40, No. 1, pp. 12-23, January, 1992.

[8] E. Moulines and F. Charpentier, "Pitch Synchronous Waveform Processing Techniques for Text-to-speech Synthesis using Diphones," *Speech Communica vtion*, vol. 9 (5/6), pp. 453-467, 1990.

[9] B.E. Caspers and B.S. Atal, "Changing Pitch and Duration in LPC synthesized speech using Multipulse Excitaion," *J. Acoust. Soc. America, supple. 1*, vol. 73, p.S5, spring, 1983.

[10] Y. Linde, A. Buzzo, and R. M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. on Communications*, vol. COM-28, pp. 84-95, January, 1980.

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

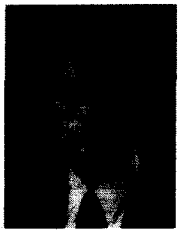
[12] Ki Seung Lee, Dae Hee Youn, and Il Whan Cha, "Voice personality transformation using an orthogonal vector space conversion," *proc. of EUROSPEECH '95*, Madrid, pp. 427-430, 1995.

[13] 이기승, 윤대희, 차일환, 박군종, "직교 벡터 공간 변환을 이용한 음성 개성 변환", 제8회 신호처리 합동학술 대회 논문지 pp. 104-107

[14] Rafael C. Gonzales and Paul Wintz, *Digital Image Processing*, Addison-Wesley Publishing Company, 2nd edition, pp. 125.

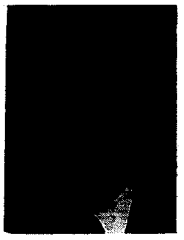
[15] D. W. Griffin and J. S. Lim, "Signal Estimation from the Modified Short-Time Fourier Transform", IEEE Trans. on Acoust, Speech, Signal Processing, vol. ASSP-32, pp. 236-243, Apr., 1984.

저 자 소 개



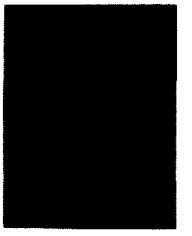
李 起 承(正會員)

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



朴 軍 宗(正會員)

1959년 7월 11일생. 1982년 연세대학교 전자공학과 졸업(공학사). 1984년 연세대학교 대학원 전자공학과 졸업(공학석사). 1991년 ~ 현재 동양 공업 전문 대학 전자통신과 교수. 1994년 ~ 현재 연세대학교 전자공학과 박사과정. 주관심분야는 음성 신호 처리, 컴퓨터 구조등임