

화자 정규화를 위한 비정형 워핑함수 도출에 관한 실험

Experiments on Extraction of Non-Parametric Warping Functions for Speaker Normalization

신 옥 근*
(Ok Keun Shin*)

*한국해양대학교 IT공학부

(접수일자: 2005년 4월 29일; 수정일자: 2005년 5월 26일; 채택일자: 2005년 5월 30일)

화자들 사이의 워핑특성을 알아보기 위해 비정형 워핑함수를 도출하는 실험을 수행하였다. 이를 위해 모음의 MFCC와 LP 스펙트럼을 이용하여 화자별, 음소별 대표 스펙트럼을 선정한 다음 음소별 기준 스펙트럼을 선택하였다. 기준 스펙트럼과 대표 스펙트럼을 스펙트럼의 전체대역에서 DTW로 비교하여 화자별 워핑함수를 구한 다음, 이들을 clustering함으로써 비정형 워핑함수의 집합을 도출하였다. 이 함수집합에서 남성화자와 여성화자의 함수들이 각각 구간선형함수와 파워함수와 유사함을 관찰할 수 있었으며, 이를 근거로 이 함수들을 조합한 하이브리드 워핑함수 집합을 정의하였다. 음소단위의 인식 실험을 통하여 새로 정의된 함수들의 인식률을 시험하였으며 두 함수집합 모두에서 개선된 인식률을 얻을 수 있었다.

핵심용어: 음성인식, 화자정규화, 비정형, 하이브리드, 워핑함수

투고분야: 음성처리 분야 (2.4)

In this paper, experiments are conducted to extract a set of non-parametric warping functions to examine the characteristics of the warping among speakers' utterances. For this purpose, we made use of MFCC and LP spectra of vowels in choosing reference spectrum of each vowel as well as representative spectra of each speaker. These spectra are compared by DTW to give the warping functions of each speaker. The set of warping functions are then defined by clustering the warping functions of all the speakers. Noting that male and female warping functions have shapes similar to piecewise linear function and power function respectively, a new hybrid set of warping functions is defined. The effectiveness of the extracted warping functions are evaluated by conducting phone level recognition experiments, and improvements in accuracy rate are observed in both warping functions.

Keywords: Speech recognition, Speaker Normalization, Non-parametric, Hybrid, Warping function

ASK subject classification: Speech Signal Processing (2.4)

1. 서론

화자 정규화 (VTLN)는 화자들 사이의 성도의 길이와 모양이 달라서 생기는 음향학적 변이를 최소화함으로써 화자독립 음성인식 시스템의 성능을 향상시키는 방법의 한 가지이며 주파수영역의 스펙트럼을 워핑하는 방법이

많이 쓰인다. 이러한 화자 정규화는 워핑 파라미터를 추정하는 방법을 기준으로 분류하면 화자의 발화에서 포먼트의 위치를 찾은 다음 이를 기준 포먼트의 위치와 비교하여 파라미터를 찾아내는 '파라미터 기반 워핑'[1]과 미리 정의된 워핑함수의 집합으로부터 화자별로 최적의 워핑함수를 추정하는 'exhaustive search' 방법으로 나눌 수 있다. 이 두 가지 워핑방법 중 후자의 exhaustive search 방법이 전자보다 더 좋은 성능을 가지는 것으로 알려져 있다. [2,3,4].

Exhaustive search 방법에서 중요한 연구 주제중의 하나는 워핑함수이며, (구간)선형 함수[5,6,7]와 비선형 함수로 대별된다. 이들 중 비선형 함수에 대한 연구는 화자들 사이의 스펙트럼이 비선형적으로 워핑되어 있다는 실험결과[3,4]에 근거하여 Bilinear transform (BLT) 함수[4,8], 지수함수[9,10], 그리고 frequency dependent scale factor[3] 등과 같은 다양한 형태의 함수를 대상으로 계속 연구되어 왔다. 그러나 이런 노력에도 불구하고 대용량 연속음성의 화자 정규화에서 가장 기본적이고 단순한 형태의 함수인 구간선형함수보다 우수한 성능을 갖는 비 선형워핑함수는 찾지 못한 것으로 알려져 있다 [8,10]. 화자들의 스펙트럼이 비선형적으로 워핑되어 있음에도 불구하고 비선형 워핑함수가 우수하지 못한 근본적인 이유는 화자들 사이의 복잡한 스펙트럼 워핑을 소수의 파라미터로 표현할 수 있는 정형화된 함수로 국한하여 모델링한 것에 기인하는 것으로 보인다. 이 함수들이 효율적이지 못한 또 다른 이유는 워핑함수의 집합을 도출하는 대부분의 방법들이 포먼트의 위치만을 이용하기 때문이다. 포먼트들은 워핑함수를 쉽게 추출할 수 있게 하는 장점이 있지만 스펙트럼의 저주파대역에 집중되어 있으므로 고주파 영역의 워핑정보는 전혀 추출할 수 없는 단점이 있다.

본 연구에서는 몇 개의 간단한 파라미터로 워핑함수를 표현하고자 하는 정형화의 제약없이, 스펙트럼의 모든 대역 폭에서 발화 스펙트럼을 직접 비교하여 워핑함수를 추출함으로써 워핑함수가 어떤 모양의 함수인지, 그리고 추출된 이 워핑함수의 성능은 어떤지를 알아보기 위한 실험을 수행한다. 이를 위해 먼저 모음의 음소별 MFCC의 평균에 대응하는 LPC 스펙트럼들을 찾아 화자별 대표 스펙트럼과 음소별 기준 스펙트럼으로 선정한 후 화자들의 음소별 대표 스펙트럼과 기준 스펙트럼을 DTW 알고리즘을 이용하여 비교함으로써 화자별 음소별 워핑함수를 추출하고 이들을 clustering하여 워핑함수의 집합으로 결정한다. 이렇게 구한 워핑함수로 선행연구[11]의 구간선형함수를 교체하여 인식 실험을 수행하여 그 효과를 검토한다.

본고의 구성은 다음과 같다. 먼저 II장에서 본 연구에서 수행한 실험의 개요를 설명한 다음, III장과 IV장에서 화자들의 대표 스펙트럼과 음소별 기준 스펙트럼을 추출하고 이들을 비교하여 워핑함수들의 집합을 도출하는 과정을 설명한다. V장에서는 앞 장의 결과를 정형함수의 조합으로 표현하는 하이브리드 워핑함수의 도출과정에

대해 설명한다. 이어서 VI장에서는 추출된 함수들을 이용하여 각 화자들의 워핑함수를 추정하는 방법에 대해 설명한 다음, VII장에서 워핑함수 추출 및 인식 실험에 대한 여러 조건들과 실험 결과, 그리고 이에 대한 고찰을 기술하고, 마지막의 VIII장에서 결론짓는다.

II. 워핑함수집합 도출의 개요

화자들의 발화 데이터로부터 워핑함수의 집합을 도출하기 위해서는 전체 화자들의 발화의 기준이 되는 음소별 기준 프레임과 각 화자의 발화특성을 표현하는 음소별 대표 프레임을 결정한 다음, 이들을 비교하여 화자들의 발화가 어떻게 워핑되어있는지 알아야 한다. 이를 위해 성도의 특징이 잘 나타나는 모음의 스펙트럼을 추출한 다음, 화자별로 음소별 '대표 (representative) 스펙트럼'을 선정한다. 이어서 이 화자들 중에서 기준 화자를 선택하고, 기준화자의 대표 스펙트럼을 '기준 (reference) 스펙트럼'으로 정한다. 이렇게 구해진 기준 스펙트럼과 각 화자의 대표 스펙트럼을 음소별로 비교하여 화자별, 음소별 워핑함수를 추출한 다음, 화자별로 대표함수를 선택하여 각화자의 워핑함수로 결정한다. 모든 화자들의 함수가 구해진 다음에는 비슷한 함수들을 하나로 묶어 워핑함수 집합을 만든다. 그림 1에 워핑함수 도출과정을 보인다.

음성인식기를 학습시키고 시험하기 위해 필요한 개별 화자들의 워핑함수는 최종적으로 구해진 워핑함수의 집합을 이용하여 추정하며, 본 연구에서는 선행연구[11]에서 제안한 벡터 양자화기를 이용한 방법을 사용한다.

III. 대표 스펙트럼과 기준 화자의 선정

이 장에서는 화자별 대표 스펙트럼들과 음소별 기준 스펙트럼을 선정하는 방법에 대해 기술한다.

3.1. 개별화자의 음소별 대표 스펙트럼

음소별 기준 스펙트럼, 그리고 화자들의 음소별 대표 스펙트럼을 선택하기 위해 직접적인 방법으로 스펙트럼들을 비교하는 것은 여러 가지 면에서 어려운 점이 많다.

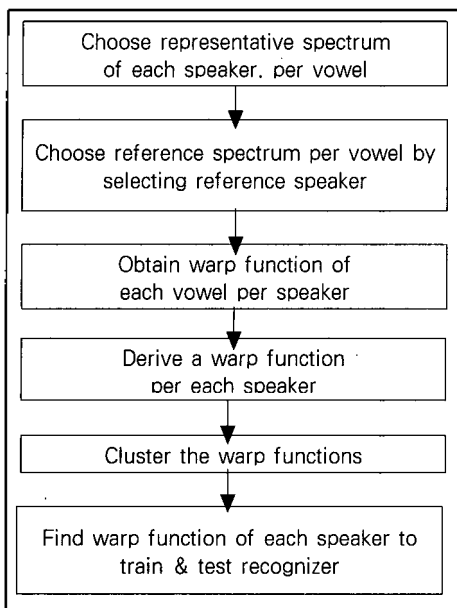


그림 1. 워핑함수집합 추출 과정
Fig. 1. Warping function extraction.

비교해야할 스펙트럼의 수와 종류가 많을 뿐 아니라, 적합한 거리척도와 비교 방법을 고안해야 하기 때문이다.

본 연구에서는 모든 스펙트럼들을 직접적인 방법으로 비교하기보다는, 발화 프레임들로부터 MFCC와 LP (Linear Prediction) 스펙트럼을 같이 이용하여 화자들의 대표 스펙트럼과 기준 스펙트럼을 선정하고 이들을 비교하는 방법을 택한다. 그림 2에 각 화자의 음소별 대표 스펙트럼을 추출하는 과정을 보인다. 먼저 음소별로 발화의 가운데 위치하는 프레임들을 선택함으로써 문맥의 영향을 최소화한다. 이 프레임들로부터 MFCC와 LP 모델을 구하는데 이들은 같은 프레임에서 추출된 것이므로 서로 대응관계가 있다. 각각의 화자에 대해 음소별로 MFCC의 평균을 구하고 이 평균과 가장 가까운 MFCC를 찾은 다음, 이에 대응하는 LP 모델을 찾는다. 이 LP 모델로부터 스펙트럼을 구하여 화자의 음소별 대표 스펙트럼으로 삼는다. LP 스펙트럼은 파워스펙트럼과 달리 불필요한 굴곡이 적은 윤곽을 표현할 뿐 아니라 모음의 포먼트 위치를 포함하는 스펙트럼의 전체대역을 나타낼 수 있어 워핑함수를 추출하는데 알맞은 영역이라 할 수 있다.

3.2 기준화자의 선정

화자들의 음소별 대표 스펙트럼과 비교될 음소별 기준 스펙트럼이 필요하다. 화자단위의 워핑을 전제하므로 적절한 화자를 선택한 다음 이 화자의 음소별 대표 스펙트

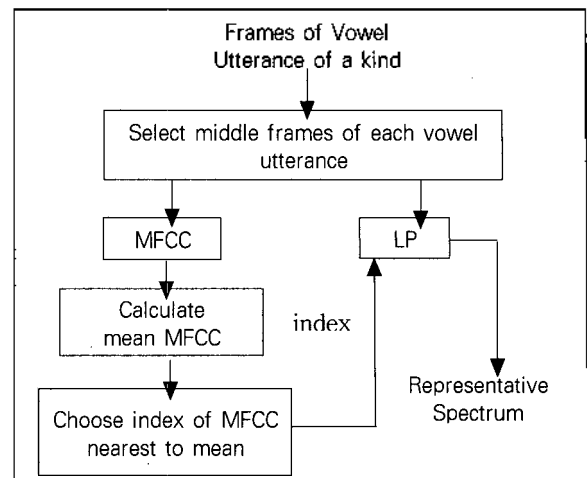


그림 2. 화자별 음소 대표스펙트럼 추출과정
Fig. 2. Selection of representative spectra per vowel, per speaker.

럼들을 기준 스펙트럼으로 결정한다. 이를 위해 음소별로 화자들의 대표 MFCC의 평균을 구하고, 이 음소별 평균과 각 화자의 대표 MFCC 사이의 거리를 구한다. 이 거리의 합이 가장 작은 화자의 스펙트럼을 음소별 기준 스펙트럼으로 선택한다.

IV. 워핑함수 집합의 추출

III장에서 구한 각 음소의 기준 스펙트럼과 화자별 대표 스펙트럼을 DTW로 비교하여 최종적인 워핑함수집합을 도출하는 과정을 설명한다.

4.1 DTW를 이용한 화자별 음소별 워핑함수

DTW (Dynamic Time Warping)[12]는 시간 영역에서 두 신호사이의 거리를 계산하되 시간축상의 워핑의 영향을 최소화하기 위해 다이나믹 프로그래밍 알고리즘을 이용하는 time alignment의 한 가지 기법이다. 본 연구에서는 이 DTW기법을 주파수 영역에 적용하여 음소별 기

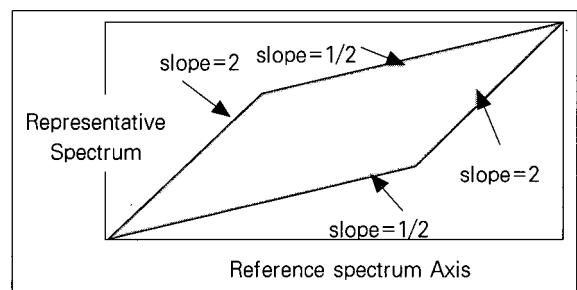


그림 3. DTW 탐색구간
Fig. 3. Search space of DTW.

준 스펙트럼'을 기준으로 각 화자들의 대표 스펙트럼을 정렬함으로써 워핑함수를 구한다. 이 과정을 거쳐 얻어지는 것은 화자별, 음소별 워핑함수이다. 본 연구에서 이용한 DTW 탐색 구간을 그림 3에 보인다.

그림 4에 기준 스펙트럼, 미지의 입력 스펙트럼, 그리고 이들을 DTW로 비교하여 워핑함수를 추출한 다음 워핑한 스펙트럼의 예를 도시하였으며, 그림 5에는 이 때 구한 워핑 함수를 보인다. 그림 4의 스펙트럼들은 관찰을 쉽게 하기 위해 각각의 크기를 적당히 조정한 것이며, 'Reference'는 음소 /ey/의 기준 스펙트럼으로 남성화자(MDBB1)의 것이고 'Input'은 여성화자(FHLM0)의 음소 /ey/의 대표 스펙트럼이다. 이 예의 여성화자의 포먼트가 기준 스펙트럼인 남성화자의 것보다 높은 주파수에 위치함을 볼 수 있다. 그림 5에 실선으로 표현된 '추출된 워핑함수'는 기준 스펙트럼에 비해 'Input' 스펙트럼이 어떻게 워핑되어 있는가를 나타내는 것이다. 따라서 이 화자의 스펙트럼을 정규화하기 위해서는 그림 5에 보인 '추출된 함수'의 역함수로 워핑할 필요가 있다. 이 역함수는 그림의 점선을 기준으로 '추출된 함수'를 선대칭으로 변환한 함수이며, 이렇게 구한 워핑함수로 'Input'을 워핑한 것이 그림 4의 'Warped' 스펙트럼이다. 'Warped' 스펙트럼을 'Reference'와 비교하면 두 스펙트럼의 주요 피크들이 잘 일치함을 알 수 있다.

4.2. 워핑함수들의 집합

4.1 절까지의 과정에서 구해진 것은 화자별, 음소별 워핑함수들이다. 이들로부터 적당한 수의 워핑함수의 집합을 만들어야 하는데 화자별 워핑함수를 만든 다음, 이들로부터 워핑함수의 집합을 만든다. 일반적으로 한 화자의 여러 음소의 워핑함수들 사이에는 편차가 있어 화자를 대표하는 워핑함수를 구하는 것은 쉽지않다. 본 연

구에서는 간단히 음소별 워핑함수의 평균을 취해 화자별 워핑함수를 만든다.

화자별 함수를 구하고 난 다음에는 비슷한 워핑함수들을 하나로 묶어 적당한 수의 최종적인 워핑함수들의 집합을 만든다. 본 연구에서는 각각의 화자별 함수를 하나의 벡터로 취급하고 k-means 알고리즘을 이용하여 clustering함으로써 최종적인 함수의 집합을 결정한다.

V. 추출된 워핑함수의 정형화

앞 장에서 기술한 방법으로 도출한 함수들의 모양은 화자의 성별에 따라 다소 달랐다. 남성화자의 경우 대체적으로 구간선형함수와 비슷하고, 여성화자의 경우에는 파워함수[10], 또는 BLT (Bilinear Transform) 함수[4]와 비슷하다는 것에 주목하여 성별에 따라 서로 다른 정형 워핑함수를 적용하는 간단한 실험을 수행하였다. 이 실험에서 고려한 정형함수들은 구간선형함수, 파워함수, 그리고 BLT 함수등의 세 가지이며 이들 중 상대적으로 성능이 좋은 구간선형함수 (eq. (1))와 파워 함수 (eq. (2))를 최종적으로 선택하여 실험하였다.

$$\omega_L(f) = \begin{cases} af, & f \leq f_0 \\ af + \frac{f_N - af_0}{f_N - f_0}, & f \geq f_0 \end{cases} \quad \text{eq.(1)}$$

$$\omega_b(f) = \left(\frac{f}{f_N}\right)^\beta f_N \quad \text{eq.(2)}$$

여기서 f 는 주파수, f_N 과 f_0 은 각각 Nyquist 주파수와 구간선형 함수의 한계주파수이다.

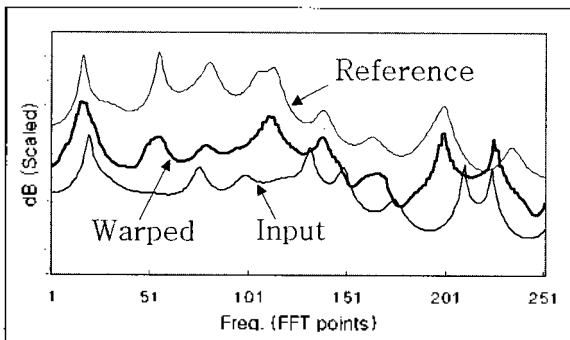


그림 4. 워핑 전후의 화자 FHLM0 및 기준화자(MDBB1)의 모음 /ey/의 LP 스펙트럼
Fig. 4. LP spectra of reference (MDBB1), Input & Warped (FHLM0) for /ey/.

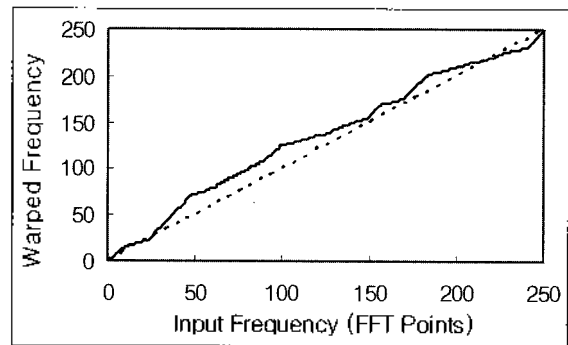


그림 5. 그림 4의 예에서 추출된 워핑함수
Fig. 5. 'Warped' function of Fig. 4.

VI. 벡터 양자화기를 이용한 워핑함수 추정

워핑함수의 집합을 결정한 다음에는 인식모델의 학습 및 시험에 사용될 데이터들의 화자별 워핑함수를 결정해야 한다. 본 연구에서는 선행연구[11]에서 제안한 벡터 양자화기를 이용하여 각 화자들의 워핑함수를 결정한다. 이 방법은 정규화된 특징 벡터들로 학습된 벡터 양자화기를 만든 다음, 시험용 화자의 특징벡터들을 이 양자화기와 비교하여 최적의 함수를 찾아내는 방법이다. 선행 연구에서는 구간선형 워핑함수를 이용하였으며, 이 때의 성능은 포맷트를 직접적인 방법으로 추출하여 워핑한 방법[13]의 성능과 비슷하였다.

VII. 실험 및 고찰

본 연구의 실험 조건은 워핑함수 추출을 위한 부분을 제외하면 선행연구의 조건들과 같다. 따라서 선행연구와 본 연구의 성능 차이는 워핑함수의 차이에 기인한다. 다음에 선행 연구와 본 연구에서 공통적으로 이용한 인식기와 코퍼스에 대해 요약한 다음, 본 연구의 워핑함수 추출결과, 음소인식 실험결과, 그리고 이에 대한 고찰을 차례로 기술한다.

7.1 인식기와 코퍼스

HTK 음성인식 툴킷[14]을 이용하여 음소단위의 인식 실험을 수행하였다. 이 툴킷을 음소별로 3개의 state를 갖는 CDHMM으로 설정하였으며, 각 state는 3개의 GMM로 모델링하였다. TIMIT 음성 코퍼스[15]를 이용하여 실험하였는데, 0.97의 계수를 갖는 1차 프리엠퍼시스 필터로 음성 데이터를 필터링한 다음, 초당 100개의 비율로 25ms길이의 해밍 윈도우를 적용하여 프레임을 추출하였다. 특징 벡터는 각 프레임의 에너지와 12개의 MFCC, 그리고 이들의 1차 및 2차 미분을 포함시킨 39차원의 벡터이다.

워핑 함수추출과 정규화 양자화기 생성을 위해 공통적으로 사용한 음소는 13개의 모음 (iy, ih, eh, ey, ae, aa, ah, ao, ow, uh, uw, ux, er)이다. 이 모음발화들이 학습용으로 쓰일 때는 성비 균형을 위해 TIMIT의 학습용 발화중 임의로 선정한 남녀 각각 136명씩의 화자의 발화로부터 추출하였다.

7.2 워핑함수 집합의 도출

워핑함수를 도출하기 위해 MFCC 및 LPC의 두 가지 종류의 특징벡터를 같이 이용하였다. 이 특징벡터들은 인식 실험을 위해 추출한 것과 같은 방법으로 전처리된 프레임 데이터로부터 구한 12차의 MFCC와 LP 스펙트럼이다. 훈련용화자의 대표 MFCC들을 음소별로 MFCC의 평균과 비교하여 MDBB1을 기준화자로 선택하고 이것과 나머지 화자의 스펙트럼을 DTW로 비교하였다. 워핑함수의 집합은 화자별 워핑함수를 LGB 벡터 양자화기로 clustering 하여 구한 16개의 함수에 기울기가 1인 선형 함수를 합하여 모두 17개의 함수를 갖게 하였다. 이 함수의 집합을 그림 6에 보인다. 이 워핑함수의 집합을 선행연구의 정규화 벡터 양자화기에 적용하여 화자들의 워핑함수를 추정하였으며 이들의 분포를 그림 7에 보인다. 그림 7의 'Warped Function Id'는 그림 6의 함수들을 그림의 위쪽에서부터 차례로 번호를 붙인 것이다. 이들 그림에서 다음을 관찰할 수 있다. 먼저 함수들 사이의 간격 (그림 6), 그리고 빈도의 분포 (그림 7)가 고르지

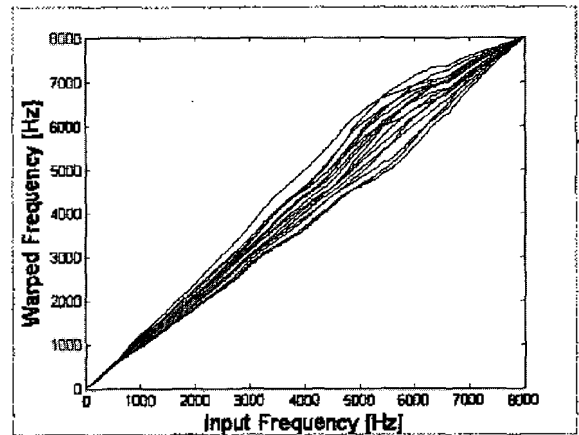


그림 6. 추출된 워핑함수의 집합
Fig. 6. Extracted set of warping functions.

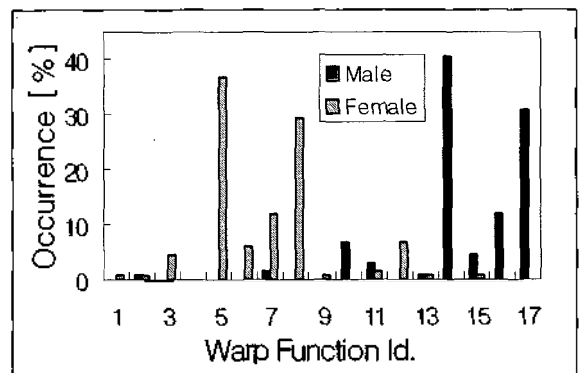


그림 7. 학습용화자들의 워핑함수 분포
Fig. 7. Warp factor distribution of all TIMIT training speakers.

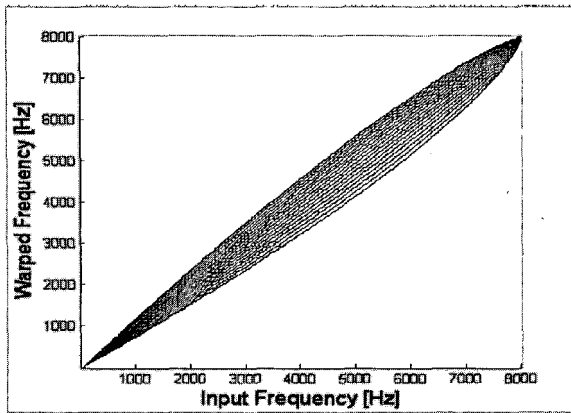


그림 8. 파워 워핑함수 집합
Fig. 8. Power warping function.

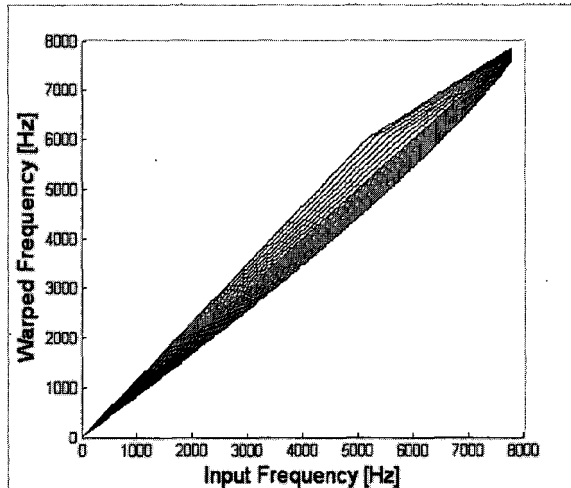


그림 9. 하이브리드 워핑함수의 집합
Fig. 9. Hybrid warping functions.

않다. 이 것은 워핑함수의 도출과정이 최적이지 아니며 더 개선될 여지가 있음을 의미한다. 그림 6에 보인 함수들의 저주파 특성은 비교적 선형적이며 구간선형함수와 유사한 형태이다. 이 선형 구간은 대략 F3까지 유지되지만 그 이상의 주파수 대역에서는 다분히 비선형적이다. 또 남성화자에 해당하는 함수는 구간선형함수와 유사하며, 여성화자의 함수는 파워함수 (그림 8)와 비슷하다는 것을 알 수 있다. 이에 근거하여 그림 9와 같은 정형화된 하이브리드 워핑함수를 정의하고 이를 이용한 워핑을 실험을 수행하였다.

7.3 인식 실험 및 고찰

다음 표1에 LP 모델의 차수에 따른 음소단위의 인식결과를 보인다. LP모델의 차수가 26일 때 구한 워핑함수의 성능이 가장 좋았다. 또 표2에는 워핑하지않은 Baseline 시스템, 구간선형함수 (Piecewise Linear), 본 실험에서 도출된 비정형함수 (Non-parametric), 파워

표 1. LP 모델의 차수에 따른 비정형함수의 인식률

Table 1. Accuracy rates as function of LP orders.

LP Order	Accuracy Rate(%)
24	55.46
26	55.60
28	55.51

표 2. 워핑함수별 인식률

Table 2. Accuracy rates as function of warping functions.

Warping function	Accuracy Rate(%)	Improvement
Baseline	52.79	-
Piecewise Linear	54.91	2.21
Non-parametric	55.60	2.81
Power	54.41	1.62
Hybrid	55.27	2.48

함수 (Power), 그리고 하이브리드함수 (Hybrid)를 이용했을 때의 인식 결과를 보인다. 도출한 비정형함수를 이용했을 때 Baseline 인식기보다 약 2.8%, 그리고 구간선형함수를 이용한 경우보다는 0.6%의 절대 인식률 향상을 보였다. 가장 좋은 성능을 갖는 것으로 알려진 구간선형함수의 인식률 향상이 약 2.2%인 것을 감안하면 0.6%의 향상은 미미한 것은 아니다.

파워함수의 성능은 하이브리드함수의 성능과 비교하기 위한 것으로 표 2의 모든 함수들 중 가장 낮지만 이것과 구간선형함수를 합성한 하이브리드함수의 성능은 이 두 함수들 보다 다소 우수하다. 이 것은 남성과 여성의 워핑특성이 같지 않음을 시사한다.

VIII. 결론

본 연구에서는 화자들 사이의 실제 워핑정도를 알아보기 위해 화자들의 발화 스펙트럼을 전체 주파수대역에서 직접 비교하여 워핑함수의 집합을 도출하고 이를 이용하여 정규화한 다음, 음소단위의 인식 실험을 수행하였다. 워핑함수의 도출에 필요한 기준 스펙트럼과 화자별 대표 스펙트럼을 구하기 위해 MFCC와 LP 스펙트럼을 이용하였으며, DTW를 이용하여 스펙트럼들을 비교해서 워핑함수를 도출하였다. 도출된 워핑함수는 전체적으로 비선형적이었으나 F3이하의 대역에서는 비교적 선형이었다. 또 남성화자와 여성화자의 함수들이 각각 구간선형함수와 파워함수와 유사함을 발견하였으며 이 두

함수들을 합성한 하이브리드 함수 집합을 도출하였다. 인식실험결과, 구간함수와 비교하여 비정형함수는 0.6%, 그리고 하이브리드함수는 0.27%의 절대 인식률향상을 얻을 수 있었다. 이 실험에서 구한 워핑함수의 모양과 함수들 사이의 간격은 고르지 않았으며 훈련용 화자들의 함수 분포도 이상적인 분포와는 거리가 있었다. 좀 더 체계적인 방법으로 접근한다면 더 좋은 인식률 향상을 얻을 수 있을 것으로 기대되며, 도출된 비정형함수의 정형화를 위한 노력 또한 필요하다.

참고 문헌

1. E. Edie and H. Gish, "A Parametric Approach to Vocal Tract Length Normalization", Proc. ICASSP'96, 346-349, 1996.
2. H. Wakita, "Normalization of Vowels by Vocal Tract Length and Its Application to Vowel Identification", IEEE. Trans. on ASSP. Vol. ASSP-25, No. 2, 183-192, 1977.
3. S. Umesh, S. V. B. Kumar, M. K. Vinay, R. Shamar and R. Shinha, "A Simple Approach to Non-Uniform Vowel Normalization," Proc. ICASSP, 517-520, 2002.
4. P. L. Dognin, "A Bandpass Transformation for Speaker Normalization", Ph.D. Thesis, University of Pittsburgh, 2003.
5. L. Lee and R. C. Rose, "A Frequency Warping Approach to Speaker Normalization", IEEE Trans. on Speech and Audio Processing, 6 (1), 49-60, 1998.
6. D. Pye and P. C. Woodland, "Experiments in Speaker Normalization", ICASSP, 1047-1050, 1997.
7. M. Pitz and H. Ney, "Vocal Tract Normalization as Linear Transformation of MFCC", Proc. EUROSPEECH, 1445-1448, 2003.
8. P. Zhan and Alex Waibel, "Vocal Tract Length Normalization for Large Vocabulary Continuous Speech Recognition, Language Technologies Institute Technical Report : CMU-LTI-97-150, Carnegie Melon University, May, 1997.
9. P. Zhan and M. Westphal, "Speaker normalization based on frequency warping", ICASSP-97, Munich, Germany, 1039-1042, 1997.
10. S. Molau, S. Kanthak and H. Ney, "Efficient Vocal Tract Normalization in Automatic Speech Recognition", Proc. ESSV, 209-216, 2000.
11. 신옥근, "연속음성 인식기를 위한 벡터양자화기 기반의 화자정규화", 한국음향학회지, 제 23권 제 8호, 583-589, 2004.
12. L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, (Parentice Hall, New Jersey, 1993.)
13. M. A. Bacciani, *Speech Recognition System Design Based On Automatically Derived Units*, Ph. D. Thesis, (Boston University, 1999.)
14. S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev and P. Woodland, *The HTK Book*. ver. 3.,

Microsoft Corp., 2000.

15. J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallet and N. L. Dahlgren, DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus: CDROM, NIST., 1993.

저자 약력

• 신 옥 근 (Ok Keun Shin)

1981년 서강대학교 전자공학과 졸업 (학사)
 1983년 부산대학교 전자공학과 (공학석사)
 1989년 프랑스 Université de Franche-Comté (공학박사)
 1983년~1995년 한국전자통신연구소 선임연구원
 1995년~현재, 한국해양대학교 IT공학부 부교수
 * 관심분야: 신호처리, 음성신호처리, 음성인식