JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지(J. Korea Inst. Inf. Commun. Eng.) Vol. 20, No. 3 : 464~470 Mar. 2016

음성 인식용 데이터베이스 검증시스템을 위한 새로운 음성 인식 성능 지표

지승은 · 김우일*

A New Speech Quality Measure for Speech Database Verification System

Seung-eun Ji · Wooil Kim*

Department of Computer Science & Engineering, Incheon National University, Incheon 22012, Korea

요 약

본 논문에서는 음성의 특성 지표를 이용한 음성 인식용 데이터베이스 검증 시스템의 개발 내용을 소개하고 이 시스템의 핵심 기술인 음성 특성 지표 추출 알고리즘을 설명한다. 선행 연구에서는 본 시스템에 필요한 효과적인 음성 인식 성능 지표를 생성하기 위해 대표적인 음성 인식 성능 지표인 단어 오인식률(Word Error Rate, WER)과 상관도가 높은 여러 가지 음성 특성 지표들을 조합하여 새로운 성능 지표를 생성하였다. 생성된 음성 인식 성능 지표는 다양한 잡음 환경에서 각 음성 특성 지표를 단독으로 사용할 때보다 단어 오인식률과 높은 상관도를 나타내어 음성 인식 성능을 예측하는데 효과적임을 입증 하였다. 본 실험에서는 선행 연구에서 조합에 사용한 이차적인 음성 인식기에서 추출된 음향 모델 확률 값을 GMM(Gaussian Mixture Model) 음향 모델 확률 값으로 대체해 조합함으로써 시스템 구축 시 다른 음성 인식기에 대한 의존성을 감소시킨다.

ABSTRACT

This paper presents a speech recognition database verification system using speech measures, and describes a speech measure extraction algorithm which is applied to this system. In our previous study, to produce an effective speech quality measure for the system, we propose a combination of various speech measures which are highly correlated with WER (Word Error Rate). The new combination of various types of speech quality measures in this study is more effective to predict the speech recognition performance compared to each speech measure alone. In this paper, we increase the system independency by employing GMM acoustic score instead of HMM score which is obtained by a secondary speech recognition system. The combination with GMM score shows a slightly lower correlation with WER compared to the combination with HMM score, however it presents a higher relative improvement in correlation with WER, which is calculated compared to the correlation of each speech measure alone.

키워드 : 단어 오인식률, 상관 계수, 성능 예측, 음성 인식, 음성 인식 성능 지표

Key word : Word error rate, Correlation coefficient, Performance prediction, Speech recognition, Speech quality measure

Received 20 January 2016, **Revised** 29 February 2016, **Accepted** 03 March 2016

*** Corresponding Author** Wooil Kim (E-mail: wikim@inu.ac.kr, Tel:+82-32-835-8459)
Department of Computer Science & Engineering, Incheon National University, Incheon 22012, Korea

Open Access http://dx.doi.org/10.6109/jkiice.2016.20.3.464 print ISSN: 2234-4772 online ISSN: 2288-4165

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/li-censes/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright Ⓒ The Korea Institute of Information and Communication Engineering.

Ⅰ. 서 론

음성 인식 시스템은 시중에 출시되기 전에 시스템의 인식 성능을 검증하기 위해 대량의 음성 데이터베이스를 이용한 평가 작업이 요구된다. 또한 평가에 필요한 데이터베이스 구축 과정에서 데이터베이스를 구성하는 각 음성 데이터가 음성 인식 성능 평가에 적합한지에 대한 검증 작업이 선행되어야 한다. 음성 데이터의 적합성을 판단하기 위해서는 음성 분석 분야 전문가의 자문이 필요하며, 대량의 데이터를 검증하는 과정에 시간과 비용이 요구된다. 이에 따라 음성 데이터베이스를 자동으로 분석하고 적합성을 평가하는 과정을 효과적으로 처리할 수 있는 음성 데이터베이스 검증 시스템이 필요하다.

본 논문에서는 앞에서 기술한 문제를 해결할 수 있는 윈도우즈 환경 기반의 음성 인식용 데이터베이스 검증 시스템의 개발 내용을 소개한다. 본 시스템은 음성 신호에서 추출할 수 있는 여러 가지 음성 특성 지표를 이용하여 인식 성능을 검증한다. 따라서 검증 시스템의 핵심 요소인 다양한 음성 특징 추출 알고리즘을 소개하며 시스템에 적용된 새로운 음성 인식 성능 지표를 제안한다. 효과적인 음성 인식 성능 지표를 생성하기 위해 다수의 선행 연구가 진행되어 왔다[1, 2]. 본 연구에서는 기존의 대표적인 인식 성능 지표인 단어 오인식률(Word Error Rate, WER)과 상관관계(Correlation)가 높은 음성 특성 지표를 분석한 후, 이를 조합하여 새로운 음성 특성 지표를 생성하였다. 기존 연구에서는 다양한 특성 지표를 여러 방법으로 조합, 추가하여 더 효과적인 음성 인식 성능 지표를 생성하기 위해 진행 된 실험 결과를 소개하였다[3-5]. 본 논문에서는 선행 연구의 문제점이었던 특징을 추출하는 단계에서 부차적인 음성 인식기를 사용하는 점을 해결하기 위한 새로운 조합을 제안한다.

본 논문은 다음과 같이 구성된다. II장에서는 개발한 음성 인식용 데이터베이스 검증 시스템에 대해 상세하게 소개하고, III장에서는 본 검증 시스템에 적용한 음성 특징 추출 기능을 특징 별로 설명한다. 그리고 IV장에서는 분석한 음성 특성 지표들을 조합하여 새로운 음성 인식 성능 지표를 생성하는 실험을 간단히 소개한다. V장에서 실험 방법 및 결과를 설명하고 VI장에서 결론을 맺는다.

Ⅱ. 음성 인식용 데이터베이스 검증 시스템 개발

본 연구에서 개발한 음성 인식용 데이터베이스 검증 시스템은 윈도우즈 기반의 그래픽 사용자 인터페이스(Graphic User Interface, GUI)와 윈도우즈, 리눅스 환경에서 작성된 특징 추출 스크립트로 구성되어 있다. 리눅스 서버 상에서 검증할 음성 데이터베이스를 사용하여 스크립트를 실행 시키면 각 음성 데이터의 여러 가지 특징 지표를 출력한 결과 파일 형태로 특정 경로에 저장된다. 그림 1은 본 연구에서 개발한 음성 데이터베이스 검증 시스템의 실행 화면이다.

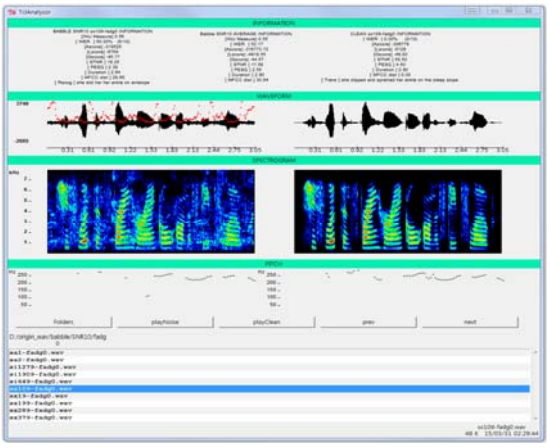

Fig. 1 The developed speech database verification system

2.1. 어플리케이션의 구조

그림 1과 같이 본 시스템은 다섯 개의 패널로 구성된다. 최상단의 창은 Information 패널로서, 왼쪽부터 순서대로 오염된 신호의 특징 정보, 데이터베이스의 평균 특징 정보, 깨끗한 신호의 특징 정보를 출력한다. 두 번째 Waveform 창은 오염된 음성과 레퍼런스 음성의 파형을 출력하며 세 번째와 네 번째 창은 음성 파형의 시간 축과 동일한 축을 사용하여 각각 음성의 스펙트로그램과 피치(Pitch) 정보를 나타낸다. 최 하단의 창에는 사용자가 지정한 경로인 음성 데이터 폴더의 음성 파일 리스트가 출력되며, 각 파일 이름을 클릭하여 해당 음성 파일의 정보를 나타낼 수 있다.

2.2. 시스템의 개발 환경 및 필요 라이브러리

본 시스템에서는 윈도우즈 상에서 제작된 GUI 기

반의 어플리케이션이며 개발 언어로는 Tcl/tk(Tool Command Language and Tk GUI)를 사용하였다[6]. Tcl/tk는 스크립트 언어인 Tcl과 Tk GUI Toolkit을 말하며, Tcl은 인터프리터 스크립트 언어로서 이러한 언어의 대표적인 특징인 멀티 플랫폼에서의 실행 및 확장을 목적으로 선택하였다. 또한 Tcl은 본 시스템에서 음성 특성 지표의 시각화를 위해 사용하는 패키지인 Snack Sound Toolkit[7]을 지원한다. Snack Sound Toolkit은 Cross-platform 라이브러리로서, 오디오 파일을 입력 받아 스펙트럼 분석 및 필터링 등의 음성 신호 처리 기능을 제공하며 이 밖에도 Python과 Ruby와 같은 프로그래밍 언어에 바인딩이 가능하다. 개발한 어플리케이션에서는 Snack Sound Toolkit을 이용하여 음성 데이터의 파형과 주파수 성분을 출력한 스펙트로그램, 음성의 피치 정보를 출력하는데 사용하였다.

2.3. 시스템의 활용 방법 및 데이터 폴더 구조

본 어플리케이션을 사용하기 위해서는 검증 대상이 되는 오염된 음성 데이터베이스와 이에 대한 레퍼런스로서 깨끗한 환경에서 수집된 원래의 음성 데이터베이스가 필요하며, 각 데이터베이스에 대해 스크립트를 실행하여 추출한 특징 파일도 함께 요구된다. 어플리케이션을 실행한 후, 오염된 음성 데이터베이스가 저장된 경로를 입력하면 시스템 하단에 해당 폴더의 파일 리스트가 출력되며 자동으로 입력 파일과 같은 이름의 깨끗한 음성 파일의 특징 정보를 동시 출력한다. 또한 해당 폴더의 데이터베이스 별 음성 특징들의 평균값을 출력하여, 보다 효율적인 음성 데이터 분석이 가능하도록 하였다. 이 기능을 실행하기 위해서 음성 데이터가 저장되는 폴더와 추출한 특징 파일이 저장되는 폴더, MFCC(Mel Frequency Cepstral Coefficient) 계수 유사도[5]의 시각화를 위한 수치를 저장하는 폴더 총 3개의 최상위 폴더가 필요하다. 이 중 본 시스템의 음성 데이터베이스의 폴더 구조는 그림 2와 같다.

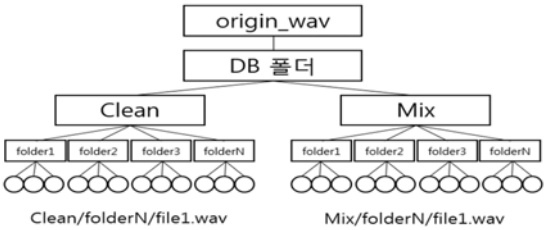

Fig. 2 Speech database folder structure of the verification system

Ⅲ. 음성 특징 추출 기능

본 장에서는 음성 인식용 데이터베이스 검증 시스템의 주 기능인 음성의 특성 지표를 비교하여 음성 인식 성능을 검증하는 단계에 사용되는 특성 지표 추출 기능을 소개한다. 본 시스템의 개발을 위해 선행 연구로 여러 가지 음성 특성 지표를 분석하였고, 다양한 특징 추출 소프트웨어를 이용하여 어플리케이션에 총 11개의 특성 지표를 출력할 수 있게 하였다. 표 1는 본 시스템에 출력 되는 11개의 음성 특성 지표를 나타낸다.

Table. 1 The speech measures to be output to the system

Speech measures	
1	Waveform
2	Spectrogram
3	Pitch
4	Acoustic model score
5	Language model score
6	GMM model score
7	Word error rate
8	SNR
9	PESQ
10	MFCC distance
11	Duration

3.1. 특징 추출 기능 구현 환경

전체 특징 추출 기능은 리눅스 환경에서 Perl과 C Shell 스크립트 언어로 작성되었으며, 특성 지표 중 하나인 PESQ(Perceptual Evaluation of Speech Quality)는 윈도우즈 상에서 오픈 소스[8]를 사용하여 MATLAB 기반의 스크립트로 작성하였다. 또한 그밖에 음성의 다른 특성 지표를 추출하기 위해 HTK(Hidden Markov Model Toolkit)[9]와 SPHINX-3[10]를 이용해 구축한 음성 인식기를 활용하였다.

3.2. 음성 특성 지표 및 추출 방법

3.2.1. 스펙트로그램

시간축의 음성 신호를 주파수 축의 신호로 변환하여 2차원으로 표현한 그래프를 스펙트럼이라고 한다. 스펙

트럼의 가로 축은 주파수, 세로 축은 주파수 성분의 크기로 구성되며 분석 시 음성 신호가 어떠한 주파수 성분으로 구성되어 있는지를 알 수 있다. 스펙트로그램은 스펙트럼을 시간 축으로 재 표현한 것으로 가로 축은 시간을 나타내며 세로 축은 주파수, 주파수 성분의 크기와 밀도를 색으로 나타낸다. 본 시스템에서는 Snack Sound Toolkit을 이용하여 스펙트로그램의 시각화를 구현하였다.

3.2.2. 피치

발화시 음성 신호는 유성음과 무성음으로 구성되며 시간 축에서 유성음은 일정한 주기적인 특성을 가지기 때문에 여러 개의 주기 신호의 합으로 표현할 수 있다. 이러한 유성음은 주기 신호에서 가장 큰 반복 패턴를 나타내는 최대 주파수를 갖는데, 이의 역수를 취한 값이 피치이다. 따라서 피치는 음성의 주기적 특성을 나타내는 지표이며 유성음과 무성음을 구분하기도 하는 등 음성의 많은 특징을 분석할 때 쓰이는 지표이다. 본 시스템에서는 Snack Sound Toolkit을 이용하여 피치의 시각화를 구현하였다.

3.2.3. 음향 · 언어 · GMM 모델 확률 값

깨끗한 레퍼런스 데이터베이스로 생성한 음향 · 언어 · 가우시안 혼합 모델에 대한 우도에 로그를 취한 값으로, 입력된 특징이 훈련된 모델에서 발견될 확률 값이다. 확률인 0~1 사이 값에 로그를 취해서 음수 값으로 나타나게 된다. 입력 된 음성 특징으로 구한 모델 확률 값은 훈련된 이상적인 모델과 유사할수록 큰 값을 갖기 때문에 음성 인식 성능 지표로 쓰일 수 있다. 본 시스템에서는 SPHINX-3과 깨끗한 음성 데이터베이스를 이용하여 구축한 음성 인식기의 음향 모델과 언어 모델의 확률 값을 구하였다. 또한 같은 음성 데이터베이스로 HTK(Hidden Markov Model Toolkit)를 이용하여 GMM 모델을 훈련하고 입력 음성의 GMM 모델 확률 값을 구하였다.

3.2.4. 단어 오인식률

단어 오인식률은 깨끗한 파일의 Transcription과 오염된 음성 파일의 인식 결과를 비교하여 발화한 단어 중 정확히 인식한 단어의 비율을 나타내는 대표적인 음성 인식 성능 지표이며 본 시스템에서는 SPHINX-3 인식기를 사용해 구하였다.

3.2.5. SNR

SNR(Signal-to-Noise Ratio, SNR)은 인식 성능을 판단할 때 대표적으로 쓰이는 음성 특성 지표로서 듣고자 하는 주변 잡음 크기에 대한 음성 신호 크기의 상대적인 비율을 표현한 값이다. SNR은 음성 신호의 크기를 잡음 신호의 크기로 나눈 것에 로그를 취해준 값으로 SNR이 양수일 경우 잡음 신호 보다 음성 신호의 크기가 크다. 따라서 SNR이 큰 음성일수록 음성의 인식 성능이 좋을 거라 표현할 수 있다. 선행 실험에서, SNR은 NIST에서 제공하는 STNR 기법[11]을 이용하여 계산하였다.

3.2.6. PESQ

사람이 주관적으로 음성을 듣고 1~5 등급 사이로 품질을 평가하는 방법인 MOS(Mean Opinion Score)기법은 주관적인 방법이기 때문에 평가자가 필요하다. PESQ는 음성 품질 평가를 위한 자동화된 테스트로써 MOS를 자동화 시킨 방법이다. 본 시스템에서는 윈도우즈 환경의 MATLAB을 이용해 오픈소스로 제공되는 스크립트를 사용하여 오염된 음성의 PESQ 값을 구하였다.

3.2.7. MFCC 계수 유사도

MFCC 특징 추출 기법은 현재 음성 인식 시스템의 특징 추출 기법으로 가장 널리 사용되고 있는 방법이다. 우선 음성 데이터의 아날로그 신호를 푸리에 변환을 통해 주파수 스펙트럼으로 변환한다. 그 후 인간의 청각 시스템을 모방한 주파수 스케일인 Mel-filter Bank 분석을 통해 얻은 계수를 로그를 취한다. 마지막으로 이산 코사인 변환(Discrete Cosine Transform, DCT)을 적용함으로써 Cepstrum 계수로 변환한다. MFCC 계수 유사도는 MFCC 계수의 거리를 측정함으로써 얻을 수 있으며 깨끗한 신호와 잡음에 오염된 신호의 MFCC 특징 계수의 차이를 나타낸다.

3.2.8. 음성 지속 시간

음성 지속 시간은 음성 신호의 길이를 음절의 수로 나눈 값으로 데이터 당 발화 속도를 수치로 표현한 값이다. SPHINX-3 인식기를 통해 각 데이터의 음성 인식 결과를 구하고 CMU(Carnegie Mellon University)에서 제공하는 단어 사전을 이용해 각 인식 결과에서 음절수를 계산하였다.

3.3. 스크립트 구조 및 실행 단계

특징 추출 기능 수행과 추출 후 최종적으로 어플리케이션에 출력 가능한 포맷인 '.info' 확장자의 음성 특징 파일을 생성하기 위해 리눅스 환경에서 스크립트를 구현하였다. 특징 추출에 필요한 바이너리 파일과 라이브러리, 리스트 파일 등을 특정 경로에 저장한 후 최종 스크립트를 수행하면 각각의 특징을 추출하는 스크립트들이 순차적으로 수행된다.

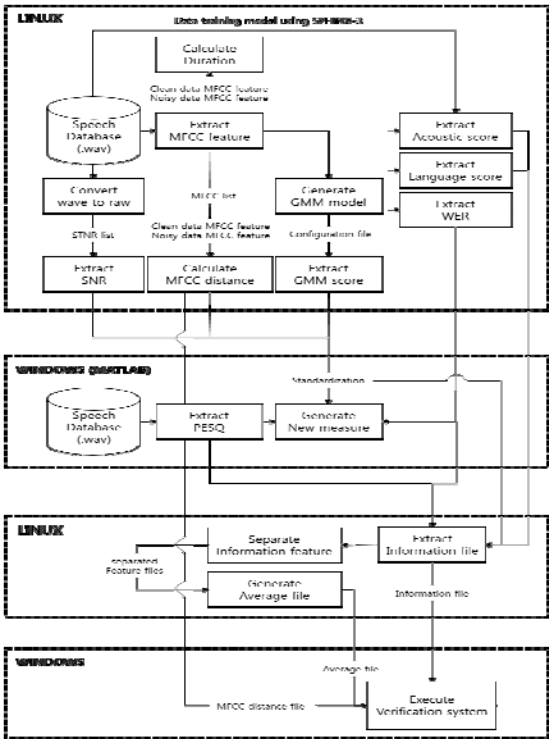

Fig. 3 Data flow diagram of the speech measure extraction algorithm

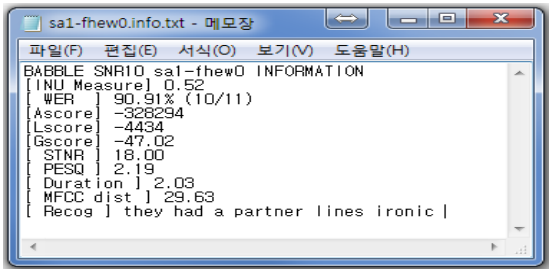

Fig. 4 The result file of the speech measure extraction algorithm

수행 과정의 중간 결과물로 개별 특징 파일이 생성되며 최종적으로 특징 파일들을 다시 불러 어플리케이션에 출력 가능한 형태로 정보 파일이 특정 경로로 저장된다. 그림 3은 시스템에 출력할 음성 정보를 추출하는 알고리즘의 데이터 흐름도를 나타낸 것이며, 그림 4는 최종으로 얻게 되는 음성 데이터 정보 파일이다.

Ⅳ. 음성 특성 지표를 이용한 새로운 음성 인식 성능 지표 생성

현재 개발된 음성 데이터베이스 검증 시스템에 출력되는 음성 특성 지표들 가운데 음성 인식 성능 지표로써 단어 오인식률을 사용한다. 선행 연구에서 단어 오인식률을 대체할 수 있는 새로운 성능 지표를 생성하기 위해, 분석한 여러 가지 특성 지표들 중 단어 오인식률과 상관도가 높은 특성 지표들을 채택하고 조합하여 새로운 성능 지표를 제안했다. 선행 실험으로 단어 오인식률과 상관도가 높은 음성 특성 지표를 알아내기 위해 다양한 음성 특성 지표를 분석하였으며, 그 중 GMM 음향 모델 확률 값, SNR, PESQ, MFCC 계수 유사도를 채택하여 새로운 음성 인식 성능 지표를 생성하였다. 새로운 음성 인식 성능 지표는 각 음성 특성 지표를 0~1의 범위로 정규화하고 단어 오인식률과 상관도에 비례하는 가중치를 적용하여 생성하였다. 아래 식 (1)은 오인식률과의 상관도에 비례하는 가중치 w_i의 계산식을 나타내며, 식 (2)는 새로운 성능 지표 I를 구하는 식이다.

$$w_i = \frac{corr(f_i, WER)}{corr(f_1, WER) + ... + corr(f_n, WER)} \quad (1)$$

$$I = (normal(f_1) * w_1) + ... + (normal(fn) * w_n) \quad (2)$$

Table. 2 Correlation of speech quality measures and WER using acoustic score extracted from SPHINX-3

Measure	Babble noise	Car noise
SNR	-0.69	-0.69
PESQ	-0.73	-0.74
Acoustic score	-0.71	-0.70
MFCC distance	-0.72	-0.71
New measure	-0.7936	-0.7790

표 2는 선행 연구에서 채택된 특성 지표들과 단어 오인식률의 상관도(Correlation Coefficient)와 이를 조합하여 생성한 새로운 성능 지표와 단어 오인식률의 상관도를 보여준다. 생성한 지표는 Babble 잡음, 자동차 잡음 환경에서 모두 각 음성 특성 지표를 단독으로 사용할 때보다 월등한 상관도를 보였다[5].

본 시스템에서의 단어 오인식률과 음향 모델 확률 값은 CMU에서 제공하는 SPHINX-3 음성 인식기를 이용하여 취득된다. 이와 같이 음성 인식기를 이용하여 음성 특성 지표를 추출하기 위해서는 음성 인식 시스템을 별도로 구축해야하는 부가적인 처리 과정이 필요하다. 따라서 본 논문에서는 음향 모델 확률 값을 별도의 음성 인식기 없이 계산 가능한 GMM 모델 확률 값으로 대체하고 이를 조합하여 새로운 음성 인식 성능 지표를 생성한다. 또한 이를 이용하여 성능을 평가하고 시스템에서의 성능 평가 기준인 단어 오인식률을 대체함으로써 음성 인식용 데이터베이스 검증 시스템의 독립성과 편의성을 향상시키고자 한다.

Ⅴ. 실험 및 결과

5.1. 실험 과정

본 실험에서는 Babble 잡음, 자동차 잡음이 각각 5, 10, 15dB의 SNR로 오염된 TIMIT 음성 데이터베이스를 사용하였다. 단어 오인식률의 신뢰도를 위해 음성 파일 하나 당 발화하는 단어의 수가 많은 상위 120개의 파일을 채택하여 오염된 잡음 별로 360개씩 총 720개의 음성 데이터를 사용하였다.

5.2. 실험 결과

표 3은 실험을 위해 채택한 음성 특성 지표들과 단어 오인식률의 상관도를 나타낸다. 이를 조합하여 완성한 음성 특성 지표는 Babble 잡음, 자동차 잡음 환경에서 각각 단어 오인식률과 –0.7766, -0.7691의 상관도를 보인다. 이는 표 2에서 설명된 선행 연구 결과인 SPHINX-3 음성 인식기를 사용하여 추출한 음향 모델 확률 값을 조합하였을 때보다 조금 낮은 상관도이다. 선행 연구에서는 Babble 잡음, 자동차 잡음 환경에서 각각 단어 오인식률과 –0.7936, -0.7790의 상관도를 보였다. 절대치로 평균 0.013 정도 낮은 상관도 보이나, GMM 모델 확률 값을 이용한 실험 결과에서도 음성 인식기를 사용한 선행 실험과 유사하게 각 특성 지표를 단독으로 사용할 때보다 조합하여 생성한 성능 지표가 단어 오인식률과 높은 수치의 상관도를 보이는 것을 확인할 수 있었다. 또한 특성 지표를 단독으로 사용할 때와 조합한 지표를 사용할 때, 오인식률과의 상관도 차이가 선행 연구의 결과보다 더 큰 비율로 상승하는 것을 볼 수 있다. 표 4는 babble 잡음 환경에서 실험한 결과를 보여주며 차례대로, 조합에 사용 되는 개별 특성 지표들과 단어 오인식률의 상관도 평균과, 생성한 성능 지표와 단어 오인식률의 상관도를 비교하여 상승한 정도를 나타낸다. 선행 연구에서는 음성 특성 지표를 조합하여 성능 지표를 생성하였을 때 단어 오인식률과의 상관도가 약 11.27% 상승하였으나 본 실험에서는 약 36.84%로, 보다 높은 상승률을 보였다.

Table. 3 Correlation of speech quality measures and WER using GMM acoustic score

Measure	Babble noise	Car noise
SNR	-0.69	-0.67
PESQ	-0.69	-0.69
GMM acoustic score	-0.26	-0.51
MFCC distance	-0.62	-0.63
New measure	-0.7766	-0.7691

Table. 4 Correlation comparison with WER of speech measures and new combined measure

Correlation with WER	prior combination	New combination
Avg. each measure	-0.71	-0.57
Combination measure	-0.79	-0.78
Increase rate	11.27%	36.84%

Ⅵ. 결 론

본 논문에서는 음성의 특징을 이용한 음성 인식용 데이터베이스 검증 시스템의 개발 내용을 소개하고 이 시스템의 핵심 기술인 음성 특성 지표 추출 알고리즘을 설명하였다. 또한 선행 연구에서부터 본 시스템에 필요한 효과적인 음성 인식 성능 지표를 생성하기 위해 대표적인 음성 인식 성능 지표인 단어 오인식률과 상관도가 높은 여러 가지 특성 지표들을 채택, 조합하여 새로

운 성능 지표를 제안해왔다. 본 실험에서는 선행 연구에서 채택했던 이차적인 음성 인식기를 사용해 추출하는 지표인 음향 모델 확률 값을 GMM 음향 모델 확률 값으로 대체 후 조합함으로써 시스템 구축 시 다른 음성 인식기에 대한 의존성을 줄였다. 새로 조합한 음성 성능 지표는 Babble 잡음, 자동차 잡음 환경에서 모두 각 음성 특성 지표를 단독으로 사용할 때보다 단어 오인식률과 높은 상관도를 나타내어 음성 인식 성능을 예측하는데 효과적임을 입증하였다. 이 결과는 선행 연구의 결과보다 단어 오인식률과 다소 낮은 상관도를 보이지만 향후 다른 부차적인 인식기를 사용하지 않은 독립적인 시스템의 개발 진행을 위해 보다 효과적인 음성 성능 지표로 발전시켜 사용할 수 있음을 확인하였다.

ACKNOWLEDGMENTS

This Research was supported by a grant (12-TI-C01) from Advanced Water Management Research Program funded by the Ministry of Land, Infrastructure and Transport of Korean government.

REFERENCES

[1] S. -Y. Yoon, L. Chen and K. Zechner, “Predicting Word Accuracy for the Automatic Speech Recognition of Non-native Speech,” *Interspeech-2010*, pp. 773-776, 2010.

[2] W. Kim and J. H. L. Hansen, “Phonetic Distance Based Confidence Measure,” *Signal Processing Letters, IEEE* vol. 17, no. 2 , pp. 121-124, Feb. 2010.

[3] S. Ji and W. Kim, “A Study on Speech Measure Analysis for Speech Recognition Accuracy Estimation in Noisy Environments,” *A Conference of Acoustical Society of Korea*, vol. 34, no. 1, pp. 46, May 2015.

[4] S. Ji, J. Cho and W. Kim, “Development of Database Verification System for Automatic Speech Recognition,” *KCC2015*, vol. 34, pp. 719-720, June 2015.

[5] S. Ji and W. Kim, “A Study on Effective Speech Recognition Performance Measure using MFCC Similarity,” *KSCSP-2015*, vol. 32, no. 1, pp.220-222, Aug. 2015.

[6] Tcl Developer Xchange. Tcl/tk Software and download page [Internet]. Available: http://www.tcl.tk/software/tcltk

[7] SNACK Sound Toolkit developed by KTH Royal Institute of Technology. Snack software and tutorial download page [Internet]. Available: http://www.speech.kth.se/snack

[8] Y. Hu and P. C. Loizou, “Evaluation of Objective Measure for Speech Enhancement,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 1, pp. 229-238, Sep. 2008.

[9] Hidden Markov Model Toolkit (HTK) developed by Cambridge University. HTK software and tutorial download page [Internet]. Available: http://htk.eng.ca0m.ac.uk

[10] SPHINX project by Carnegie Mellon University. SPHINX software and tutorial download page [Internet]. Available: http://cmusphinx.sourceforge.net

[11] STNR technique provided by National Institute of Standards and Technology(NIST) [Internet]. Available: http://www.nist.gov/speech

지승은(Seung-Eun Ji)

2015년 2월 인천대학교 컴퓨터공학부 공학사
2015년 인천대학교 컴퓨터공학부 석사과정
※관심분야 : 패턴인식, 음성인식, 휴먼 컴퓨터 인터페이스

김우일(Wooil Kim)

2003년 고려대학교 전자공학과 공학박사
2004년 ~ 2005년 미국 카네기 멜론 대학교 박사 후 연구원
2005년 ~ 2012년 미국 텍사스 주립대 (University of Texas at Dallas) 연구원 및 연구교수
2012년 ~ 현재 인천대학교 컴퓨터공학부 조교수
※관심분야 : 신호처리, 패턴인식, 음성인식, 휴먼 컴퓨터 인터페이스