음성 인식 기술 평가 동향

유하진*, 김동현**, 육동석**
*서울시립대학교 컴퓨터과학부
**고려대학교 컴퓨터학과

On the Evaluation of Speech Recognition Systems

Ha-Jin Yu*, Donghyun Kim**, and Dongsuk Yook**
[a]School of Computer Science, University of Seoul
[b]Department of Computer Science, Korea University

hjyu@uos.ac.kr, {kaizer, yook}@voice.korea.ac.kr

Abstract

We present a survey on the evaluation methods of speech recognition technology and propose a procedure for evaluating Korean speech recognition systems. Currently there are various kinds of evaluation events conducted by NIST and ELDA every year. In this paper, we introduce these activities, and propose an evaluation procedure for Korean speech recognition systems. In designing the procedure, we consider the characteristics of Korean language, as well as the trends of Korean speech technology industry.

Ⅰ. 서론

최근 음성인식 기술이 실용화를 향하여 발전하고 있고, 음성인식 서비스의 필요성이 대두되면서 음성인식 기술을 자사의 서비스에 도입하려는 업체가 늘어나고 있다. 국내에도 대학, 국책연구소, 기업연구소 등 많은 곳에서 음성인식 연구가 진행되고 있으나, 음성인식 기술을 필요로 하는 업체에서는 어느 곳의 기술이 자신이 원하는 수준을 만족하는 지 알 수 없어 여러 기술을 자신의 시스템에 도입하는 과정에서 많은 시행착오를 겪어 시간을 낭비하고 있거나, 기술력의 차이를 잘 알지 못하는 상태에서 무조건 외국 업체의 기술을 고가로 도입하는 경우가 발생하게 된다. 또한, 각각의 업체에서 필요할 때마다 평가 (BMT; benchmark test)를 함으로써 시간과 비용을 낭비하고 있고, 기술을 판매하는 업체에서도 여러 사용자에 대응하여 반복되는 평가 일정으로 인력과 비용을 낭비하고 있다. 따라서 공신력 있는 기관에서 객관적인 평가를 시행하여 수요자가 어느 기술이 자사의 요구에 가장 적합한 것인지에 대한 정보를 알 수 있다면, 수요자와 공급자가 모두 BMT를 위해 투자하는 비용과 노력을 줄일 수 있을 것이다. 또한, 공동의 평가를 통해서 수요자는 음성인식 기술의 현재 상황 판단과 미래의 예측을 할 수 있어 자사 서비스의 개발 방향을 정할 수 있으며, 기술을 개발하는 곳에서도 평가를 통하여 각 업체가 개발한 기술의 장단점을 파악하여 자사의 기술수준을 높일 수 있고 선의의 경쟁 및 학술발표회를 통해 국내 음성인식 산업의 전체 수준을 향상시킬 수 있다.

미국에서는 이미 NIST (National Institute of Standards and Technology)를 주관으로 하여 1987년부터 평가가 시작되어 현재 다양한 분야에서 계속적인 평가와 개선이 이루어지고 있다. 한국어 음성 인식의 경우 영어 음성인식 기술이 대부분 적용될 수 있으나, 한국어의 고유한 특성이 가지는 문제점이 있으므로 이러한 특성을 잘 평가할 수 있는 방안이 필요하다. 한국어의 두드러진 특징으로는 동사의 다양한 활용과 단음절로 이루어져 명사에 연결되는 조사, 또한 단음절로 이루어진 숫자음이 있다. 따라서, 고립단어, 낭독체 또는 대화체 음성의 대용량 연속음성인식 기술 평가와 숫자음 인식 평가가 필요하다. 또한, 음성인식이 실제로 활용되기 위해서는 잡음에 대한 처리가 필수적이므로 다양한 잡음에 대한 평가가 이루어 져야 한다. 그 밖에 화자 또는 환경 적응의 유무와 적응데이터의 양에 따른 평가도 요구된다.

본 논문에서는 한국어 음성 인식 분야의 기술 평가를 위한 조사 및 제안사항을 기술한다. 2장에서는 미국의

NIST 및 유럽의 ELDA에서 실행하고 있는 기술 평가에 관한 조사결과를 정리하고, 3장에서는 한국어 음성언어 기술 평가 절차 및 방법을 제안한다.

Ⅱ. 기존의 음성언어기술평가

본 절에서는 미국 NIST(National Institute of Standards and Technology)의 Speech 그룹 및 유럽의 ELDA (Evaluations and Language resources Distribution Agency)에서 주관하여 매년 실행하고 있는 음성 정보 처리 분야의 기술 평가를 살펴본다. 조사 자료는 각각의 웹사이트에서 수집되었다.[1][2]

2.1음성인식분야

HUB3 (1992-1995)

HUB3의 목적은 화자 독립 무제한 어휘 음성 인식과 녹음에 사용된 마이크로폰에 독립적인 강인한 음성 인식 기술 평가이다. 학습 데이터에는 제한이 없지만, 개인 데이터를 사용할 경우 LDC에 공개해야 한다. 평가 데이터는 두 채널의 스테레오 데이터로서 채널 1은 Sennheiser HMD-410 마이크로폰을 사용하였고, 채널 2는 세 종류 이상의 알려지지 않은 마이크로폰 데이터를 사용하여 15문장을 20명이 발성하였다. 문장의 경계와 화자의 성별이 알려진, 낭독체 음성(read speech)에 대하여 다음과 같은 조건에서 각각 인식 결과를 평가하였다.

- P0 : 환경 적응 알고리즘을 허락하고 채널 2 데이터 인식 결과를 평가
- C0 : 환경 적응 알고리즘을 허락하고 채널 1 데이터 인식 결과를 평가
- C1a : 학습 데이터를 WSJ0과 WSJ1로 제한하고 채널 2 데이터 인식 결과를 평가
- C1b : 학습 데이터 데이터를 WSJ0과 WSJ1로 제한하고 채널 1 데이터 인식 결과를 평가
- C2a : 환경 적응 알고리즘을 사용하지 않고 채널 2 데이터 인식 결과를 평가
- C2b : 환경 적응 알고리즘을 사용하지 않고 채널 1 데이터 인식 결과를 평가

이러한 복잡한 평가 구분은 사용하는 알고리즘의 효율성을 측정하기 위한 것이다. 예를 들면, C0는 기본 시스템 (baseline system)의 성능을 측정할 수 있고, P0와 C2a를 비교하면 사용한 환경 적응 알고리즘의 효율성을 측정할 수 있다.

HUB4 (1996 - 1999)

HUB4의 목적은 화자 독립 무제한 어휘 음성 인식에서 빠른 인식 알고리즘, 개체명 태깅 (named entry tagging)과 영어, 중국어, 스페인어 등에 관한 인식률 평가이다. 평가 데이터는 단일 채널로 녹음된 2.5 시간의 음성 데이터로 60%가 TV 방송, 40% 라디오 방송이며, 또한 50%가 앵커의 뉴스 방송이고, 50%가 뉴스 매거진 내용이다. 음성데이터는 다음과 같이 분류되어 평가된다.

- F0: 앵커 낭독 음성
- F1: 대화체 음성
- F2: 전화 채널을 통한 음성
- F3: 배경 음악이 있는 음성
- F4: 왜곡이 심한 상태의 음성
- F5: 외국인 음성

HUB5 (1997-2001)

HUB5의 목적은 전화 대화체 음성 인식에서 효율적인 인식 알고리즘 개발이다. 학습 데이터는 전화상의 대화체 음성인 SwitchBoard corpus, CallHome corpus로 전화선상의 통화자로부터 수집되어 표준 전화 코덱 저장 방식 (8kHz sampling, 8-bit mu-law encoding)으로 저장되었다.

평가 데이터는 다음과 같이 세가지로 나누어진다.

- EvalSet-1 : CallHome 코퍼스로부터의 20개 대화문
- EvalSet-2 : Switchboard 코퍼스로부터의 20개 대화문
- EvalSet-3 : 100번의 대화 turn이 있는 Switchborad 코퍼스의 모든 대화문

평가는 각 시스템의 단어오류율과 세 가지 평가 데이터에 따른 성능으로 하였다.

Rich Transcription(2002-2003)

자동 번역 및 자동 메타 데이터 추출을 통한 음성 인식 기술을 평가한다. 학습 데이터로는 (1) 방송 뉴스 (2) 전화 대화 (3) 회의 내용 등 자연적으로 제한 없이 나눈 사람들 간의 대화 내용이 사용된다. 평가 데이터로는 다음의 세 가지가 사용된다.

- 방송뉴스 : HUB4와 유사한 60분 분량의 방송 데이터
- 전화데이터 : HUB5와 유사한 300분 분량의 전화데이터 (화자 바뀜과 화자 ID에 대한 정보는 제공하지 않음)
- 회의 내용: NIST, CMU 등이 보유한 80분 분량의 회의 내용

평가를 위해서는 NIST SCLITE 스코어링 프로그램을

이용한 표준 문자 오류율을 측정하고, 화자 분할 스코어링 프로그램을 이용한 화자 변화 감지와 화자 구별을 기록한 메타 데이터에 대한 평가를 한다.

2.2 음성언어이해 및 정보검색 분야

Topic Detection and Tracking (1998-2004)

다양한 언어의 뉴스매거진과 방송뉴스에서 토픽과 관련된 내용을 검출해 내고 처리할 수 있는 기술에 대한 평가를 목적으로 한다. 학습 데이터로는 LDC에서 제공하는 TDT 코퍼스를 사용하고, 평가 데이터로는 LDC에서 제공하는 Mandarin Chinese 리소스, Arabic 리소스를 사용한다.

평가는 다음과 같은 방식에 대하여 실시된다.

- *Story Segmentation* : 토픽과 관련된 문단들의 변화를 감지하기
- Topic Tracking : 예제 Story와 유사한 Story를 탐지하기
- Topic Detection : 같은 토픽을 다루는 Story를 구별해 그룹화하기
- First Story Detection : 새롭게 나타난 토픽을 검출하기
- Link Detection : 두 개의 Story가 토픽과 관련해 상관이 있는지 검출하기

Spoken Document Retrieval (1997-2000)

자동 음성 인식과 정보 검색 기술을 이용한 방송 뉴스 데이터 검색 기술을 평가한다. 학습 데이터로는 Linguistic Data Consortium(LDC)에서 제공하는 뉴스 텍스트 자료와 HUB 4에서 학습 데이터로 사용된 음성 자료및 이전 테스트의 토픽에 대한 평가로 사용되었던 정보 검색(IR) 리소스를 사용하며, 평가 데이터로는 LDC에서 수집한 같은 주제와 내용으로 구분된 뉴스 Story와 DARPA TDT 수행평가에서 사용한 방송 뉴스 코퍼스를 사용한다.

평가는 검색평가와 음성인식평가로 이루어지며, 검색평가는 NIST가 제공하는 Precision/Recall 스코어링 프로그램을 사용하고, 음성 인식 평가는 NIST가 제공하는 Story에 대한 단어오류율을 평가하는 음성인식 스코어링 프로그램을 사용한다. 오디오 파일로부터 토픽에 대한 검색 결과는 토픽에 대한 순위가 있는 Time-tag 리스트로 이루어진다

2.3 Speaker Recognition (1996-2003)

전화 대화 음성에서의 화자 인식, 구별, 추적에 대한 기술을 평가한다. 학습 데이터로는 각 화자에 대해 단일 대화인 2분 분량의 대화문들을 사용하고, 평가 데이터로는 NIST가 제공하는 SwitchBoard-2 코퍼스. (8비트, m-law 연속음), 세 개의 전화 세션으로 구성된 AHUMADA 코퍼스와 NIST의 CallHome 코퍼스를 사용한다.

평가는 다음 네 가지로 구성된다.

- One-speaker detection : 주어진 음성 세그먼트에 대상 화자가 맞는지 결정
- Two-speaker detection : 전화 대화상의 양쪽 화자가 맞는지 결정
- Speaker tracking : 대상 화자의 대화 세그먼트를 찾아 간격 시간을 결정
- Speaker segmentation : 모르는 화자들 간의 음성 대화 세그먼트를 확인하여 간격 시간을 결정

2.4 오로라(Aurora) 프로젝트

유럽에서는 ELDA (Evaluations and Language Resources Distribution Agency)에서 음성인식에 필요한 자료를 수집, 배포하고 평가하는 일을 맡고 있다. ELDA는 ELRA (The European Language Resources Association) 집행부 역할을 하며, ELRA 는 음성자료의 보급과 평가를 목표로 1995년 비영리기구로 설립되었다.

오로라 프로젝트의 목적은 DSR (Distributed Speech Recognition) 시스템에서의 특징추출 소프트웨어의 표준을 만들고, 배경잡음 하에서의 전처리 알고리즘을 평가하며 음성인식 알고리즘의 강건성(robustness)을 비교 평가하는데 목표를 두고 있다.

오로라 프로젝트를 위한 음성자료로는 TI digits 숫자음 및 Wall Street Journal (WSJ0) 데이터에 여러 SNR의 잡음을 인공적으로 부가하여 사용하고, Entropic의 HTK를 표준 인식기로 채택하였다. 또한, 다양한 속도와 환경에서 주행 중인 자동차에서 녹음한 핀란드어, 스페인어, 독일어, 덴마크어, 이탈리아어 등의 음성이 있다.

III. 한국어 음성 인식 평가 방법

본 장에서는 한국어 음성 인식 평가를 위한 세부 항목과 방법에 대해 기술한다. 평가에 사용될 음성 데이터는 한국어 숫자 음성과 대용량 어휘 고립 및 연속 음성 데이터를 기본으로 하며, 훈련용 데이터와 개발용 데이터를 이용하여 평가 시스템을 준비하도록 한다. 평가는

제한된 시간 동안 제공된 평가 데이터를 가지고 수행되며 인식된 결과를 온라인으로 제출하여 참여 그룹간의 성능을 평가하게 된다.

3.1평가목적

한국어 음성 인식 평가는 현재 국내 음성인식 기술의 수준 평가 및 기술 교류를 통하여 한국어 음성인식 기술의 발전을 도모하려는데 목적이 있다. 또한 평가와 학술 세미나를 통하여, 음성인식 관련 새로운 이론의 검증과 보급을 효과적으로 할 수 있도록 한다.

3.2평가항목

평가는 크게 기본 인식률 평가와 적응 인식률 평가로 나눌 수 있는데, 기본 인식률 평가는 숫자 음성 인식 평가와 연속 음성 인식 평가로 나누어진다. 적응 인식률 평가는 기본 인식률 평가 대상에서 적응 기법을 이용했을 때의 인식 성능을 평가하는 방법이다.

3.2.1기본인식률평가

(1) 고립 단어 음성 인식

고립 단어 음성 인식은 현재 가장 실용화에 접근하여 있으므로, 지명, 업체명, 명령어, 숫자 등을 대상으로 다양한 형태의 평가가 이루어지고 있다. 현재 전화를 통한 음성의 이용분야가 가장 많으므로 유무선 전화 음성을 사용하며, 어휘의 수를 대/중/소 용량(1,000~100,000)으로 구분한다.

(2) 연속 숫자음 인식 평가

한국어 숫자 음성 인식 기술은 많은 응용 가능성에도 불구하고, 영어의 숫자 음성과 비교하여 높은 오류율을 보이고 있어 인식률의 개선이 필요하다. 평가에서는 우선 단음절어로 이루어진 11 종류의 한국어 숫자 음성 인식 기술 평가한다.

E1. 순수 숫자 (예, 공/영/일/이/삼/사/오/육/칠/팔/구)
- 4연 숫자음 : 연속으로 4자리 숫자를 발성한 데이터를 인식하는 방법.
- 알려지지 않은 길이의 순수 숫자 : 연속된 숫자음의 발성으로 숫자의 자리 수에 차이가 있는 데이터를 인식하는 방법.
E2. 자유 숫자 (예, 삼천이백구십국에 삼천이백이번)
-숫자 단위와 함께 쓰인 문장 인식 : 숫자들 사이에 단위가 들어있는 연속음성 데이터를 인식하는 방법.
-연속음성 속에서 숫자음 검출 및 인식 : 연속음성 데이터에서 숫자음성 구간만 검출하여 인식하는 방법.

(3) 연속 음성 인식 평가

화자 독립의 한국어 대용량 어휘 음성 인식을 위한 인식기 개발을 위하여, 연속 음성 인식에서 효율적인 인식 알고리즘을 평가한다.

E1. 낭독체 음성 인식
NIST HUB3 또는 HUB4 형태의 제한된 조건에서 음성 인식의 핵심 기술을 평가하고, 한국어 대용량 연속 음성 인식을 위한 핵심 기술의 현재 상태를 측정한다.
E2. 대화체 자연어 음성 인식
NIST HUB5 또는 Rich Transcription 형태의 무제한 대화체에 대한 음성 인식을 평가 하는 것으로 무제한 어휘의 자연스런 대화체 음성 인식 기술을 평가 한다.

3.2.2조건별인식률비교
기본 인식률 평가는 그림 1 에서 볼 수 있는 세 가지 하부 평가로 구별할 수 있다. 첫째는 도로변, 자동차 운전 중 잡음 등 가산잡음(Additive noise)이 섞인 데이터에 대한 인식률 평가로 각기 다른 환경 잡음과 다른 잡음 세기(dB)의 데이터를 사용하여 인식률을 평가 받는다. 둘째는 채널잡음(Channel noise)이 섞인 데이터에 대한 인식률 평가로 16KHz인 다른 종류의 마이크로 녹음한 데이터와 8KHz인 여러 조건의 채널 잡음과 가산잡음이 함께 섞인 데이터에 대한 인식률 평가이다. 위 두 가지 경우 모두 잡음정보가 알려진 경우와 알려지지 않은 경우로 나누어 평가할 수 있다. 셋째는 microphone array를 사용하여 기본 인식률을 측정하는 평가로 길이 50cm인 정사각형 안에 10cm 간격으로 마이크가 놓여진 Small scale 경우와 길이 2m인 정사각형 안에 20cm 씩 마이크가 놓여진 Large scale인 경우로 나눠서 녹음한 데이터에 대해 인식해야 할 음성 source가 알려진 경우와 알려지지 않은 경우에 대한 인식률을 평가한다.

3.2.3적응인식률평가방법
기본 인식률 평가 대상에서 적응 인식률 평가 대상으로 지목된 항목에 대해 표1과 같은 조건에서의 적응 인식률 평가를 한다.
적응 인식 평가 항목은 표1과 같이 세가지 조건으로 되어 있으며 적응하지 않은 기본 인식률과의 비교를 통해서 평가한다.

Unlimited Adaptaton	On-line Adaptaton	Off-line Adaptaton	Non Adaptaton

Evaluation Set										숫자 (Digit)		문장 (Sentence)	
										4연숫자(4)	자유숫자(N)	낭독체(R)	대화체(C)
Close-talking (Ct) Clean Data										Ct-4D	Ct-ND	Ct-RS	Ct-CS
Additive Noise (An)	Type \ SNR			10	15	20	25	30	Known Tag (Kt)	AnKt-4D	AnKt-ND	AnKt-RS	AnKt-CS
	Gaussian noise (G)												
	Car (C)												
	Airport (A)												
	Babble (B)								Un-known Tag (Ut)	AnUt-4D	AnUt-ND	AnUt-RS	AnUt-CS
	Restaurant (R)												
	Street (S)												
	Train (T)												
Channel Noise (Cn)	16 KHz	Other Microphone 1 (M1)							Known Tag (Kt)	CnKt-4D	CnKt-ND	CnKt-RS	CnKt-CS
		Other Microphone 2 (M2)											
		Other Microphone 3 (M3)											
		Other Microphone 4 (M4)											
	8 KHz	Landline (L)							Un-known Tag (Ut)	CnUt-4D	CnUt-ND	CnUt-RS	CnUt-CS
		Cordless (C)											
		Speaker phone (S)											
		Mobile phone (M)	G	C	A	B	R	S	T				
Microphone Array (Ma)	Small scale (S)			Known location (K)						MaSK-4D	MaSK-ND	MaSK-RS	MaSK-CS
				Unknown location (U)						MaSU-4D	MaSU-ND	MaSU-RS	MaSU-CS
	Large scale (L)			Known location (K)						MaLK-4D	MaLK-ND	MaLK-RS	MaLK-CS
				Unknown location (U)						MaSU-4D	MaSU-ND	MaSU-RS	MaSU-CS

그림 1 평가 항목

표 1 적응인식평가항목

	Descriptoin
Base line (A0)	Non adaptation
Off-line adaptation (A1)	Development data를 이용한 Supervised batch adaptation.
On-line adaptation (A2)	Unsupervised incremental adaptation
Unlimited adaptation (A3)	On-line, Off-line 제한 없는 adaptation.

표 2 기타 평가 항목

	Description
Real time (Rt) recogintion	Real time decoding evaluation
Perplexity (Pp) estimation	Language model perplexity evaluation

3) 기타 평가 방법

그 밖의 인식기의 성능 비교를 위해 표2의 항목에서 인식기 평가를 한다. 실시간 인식 성능 평가는 허용된 인식률에서의 인식시간 평가이다. 예를 들어 숫자 음성 인식의 경우 95% 이상 인식률일 때 인식시간을 측정하여 평가할 수 있다.

3.3데이터

1) 훈련 데이터 (Training Data)

훈련용 음성 데이터는 500 시간 분량의 Close-talking clean 데이터로 한국어 음성 인식 평가가 공고 되었을 때 참가 등록자에 한해 배포된다. 그리고 기본 인식기의 언어 모델 생성을 위해 1 기가바이트 분량의 텍스트

파일도 함께 배포된다.

3.3 데이터

1) 훈련 데이터 (Training Data)

훈련용 음성 데이터는 500 시간 분량의 Close-talking clean 데이터로 한국어 음성 인식 평가가 공고 되었을 때 참가 등록자에 한해 배포된다. 그리고 기본 인식기의 언어 모델 생성을 위해 1 기가바이트 분량의 텍스트 파일도 함께 배포된다.

2) 개발용 데이터 (Development Data)

개발용 데이터는 평가를 대비하기 위한 예비 데이터로 Additive noise, Channel noise, 그리고 Microphone array 환경에서 수집된 데이터를 각각 제공한다. 데이터 분량은 평가용 데이터 보다 긴 시간의 각 조건에 해당하는 데이터로 전사 자료(Transcription)가 함께 제공된다.

3) 평가용 데이터 (Evaluation Data)

한국어 음성인식 평가에 사용되는 평가용 데이터로 제한된 기간에 배포되어 평가용으로 사용된다. Close-talking, 개발용 데이터와 같은 여러 조건의 데이터이다.

3.4제출과결과

1) 제출

평가용 데이터 제공 후 제한된 시간 이내에 최종 학습된 인식기를 제출하고, 인식결과 및 시스템 환경에 대한 보고서를 제출한다.

2) 인식 평가

평가 방법은 연속숫자음 인식의 경우 단어오류율(Word error rate)과, 단어열오류율(String error rate)을 산출하여 평가하고, 연속 음성 인식의 경우는 단어오류율(Word error rate)과, 문장오류율(Sentence error rate)를 산출하여 평가한다. 그리고 기타 평가 방법에서 제시되었던 실시간 인식률과 언어모델의 Perplexity를 평가할 수 있다.

3.5기술발표

대회 결과 발표 및 기술 토론을 위한 워크숍을 개최하여 한국어 음성 인식을 위한 새로운 기술 방안을 논의한다.

3.6 화자인식 평가

화자인식은 음성인식과 근본적인 원리가 유사하므로 음성인식과 같은 방법으로 평가할 수 있다. 화자인식은 화자인증과 화자식별로 구분되며, 또한 화자식별도 사칭자 거부 기능 여부에 따라 나눌 수 있으므로 세가지의 평가가 가능하다.

응용분야에 따라 학습에 참여하는 화자수도 달라질 수 있으므로 30명, 100명, 500명 등으로 나누고, 학습 및 인식에 사용하는 음성 길이도 다양하게 평가할 수 있다. 텍스트 종속은 텍스트 독립에 비하여 적은 양의 데이터로도 높은 인식률을 얻을 수 있으므로 학습 문장 길이를 5음절, 10음절, 15음절, 20음절 등과 같이 하고, 문장독립은 학습에 필요한 데이터를 10초, 30초, 5분, 10분, 20분 등으로 할 수 있다.

또한 음원에 따라 다양한 성능의 차이를 얻을 수 있으므로 PC용 마이크 (다이나믹, 콘덴서), 전화 서버용 (유선, 무선, 휴대폰) 등 다양한 음원을 선정할 수 있다. 그 밖의 평가 절차는 음성인식의 경우와 같이 실행할 수 있다.

V. 결론

본 논문에서는 현재 미국과 유럽에서 시행중인 음성인식 기술의 평가 방법을 살펴보고, 한국어 음성인식 평가를 위한 방법을 제안하였다. 음성인식기술의 평가는 국내 기술 발전을 위해서는 필요한 것이지만, 평가를 받는 당사자에게는 대단히 중대한 일일 수 있다. 대규모 기관의 평가가 다른 기관에 비하여 상대적으로 낮게 나올 경우 해당 연구실의 존폐를 위협하는 결과를 초래할 수 있기 때문이다. 그렇지만, 본문에서 논의된 바와 같이 음성인식 기술은 어휘수, 잡음, 채널, 음원, 화자수 등 다양한 변수에 의해 다양한 결과가 나올 수 있으므로 수요자의 입장에서 자사의 요구조건에 맞는 기술을 선택하게 하는데 큰 도움을 줄 수 있을 것이다. 음성인식 기술의 평가는 기술발전, 수요자 선택권 향상 등 여러 가지 목적이 있을 수 있으므로 목적에 맞도록 평가기준 및 절차 등을 신중하게 선택하는 것이 가장 중요한 일이라고 할 수 있다. 따라서 지속적으로 여러 전문가의 의견을 수렴하여 가장 합리적이고 실용적인 방법을 선택해야 하겠다.

감사의 글

본 연구는 SiTEC의 용역에 의해 대한음성학회에서 수행되었습니다.

참고문헌

[1] http://www.nist.gov/speech

[2] http://www.elda.org/article52.html