

## 모음 열을 이용한 발화 검증

유일수, 노용완, 홍광석

성균관대학교 정보통신공학부 휴먼컴퓨터연구실

### An Utterance Verification using Vowel String

Il-Soo Yu, Yong-Wan Roh, Kwang-Seok Hong

School of Information and Communication Engineering, Sungkyunkwan University

gildda@nate.com, elec1004@hotmail.com, kshong@skku.ac.kr

#### Abstract

The use of confidence measures for word/utterance verification has become an essential component of any speech input application. Confidence measures have applications to a number of problems such as rejection of incorrect hypotheses, speaker adaptation, or adaptive modification of the hypothesis score during search in continuous speech recognition. In this paper, we present a new utterance verification method using vowel string. Using subword HMMs of VCCV unit, we create anti-models which include vowel string in hypothesis words. The experiment results show that the utterance verification rate of the proposed method is about 79.5%.

사용하는 anti-model로는 음소나 반 음소를 사용하는 방법 등이 제안되었다[2]. 그러나 사용되는 반 음소 모델은 잘 훈련된 음소 모델을 갖고 있어야 하는 것을 전제로 하고 있다.

본 논문에서는 신뢰도 계산에 사용되는 anti-model을 인식 후보 단어에 내재하고 있는 모음 열을 이용하여 구성하는 방법을 제안한다. anti-model로써 모음 열을 이용하면 모델을 만들기 위한 특별한 훈련이나 준비 없이 가변어휘 음성인식기를 그대로 사용하여 신뢰도 계산을 할 수 있다는 장점이 있다. 본 논문에서 제안한 발화 검증 방법의 성능을 평가하기 위해서 VCCV 단위의 가변어휘 음성인식기를 구현하고 이를 사용하여 그 성능을 평가하였다.

#### I. 서론

지난 수년간 음성인식 시스템의 성능은 음성 다이얼링, 구술에 의한 정보검색 등 많은 응용이 가능할 정도로 발전하고 있다. 이러한 시스템들은 사용자가 협조적이고 조심스럽게 시스템의 어휘를 사용할 때 좋은 성능을 보이고 있다. 그러나 사용자의 말에 OOV(out of vocabulary)를 포함하거나 시스템이 받아들이는 문법과 차이가 나면 인식 성능이 급격하게 떨어진다. 특히 사용자가 시스템을 처음 사용할 때는 주저하거나 OOV 단어를 말하거나 단편적 소리를 내어 음성인식에 대해서 문제가 된다. 실제의 응용에 있어서 사용자 발화의 20 ~ 30%가 상술된 문법을 벗어나고 있다. 유용한 시스템은 여분의 말과 잡음을 무시할 수 있어야 하고 인식된 음성이 얼마나 신뢰할 수 있는지 결정할 수 있어야 한다[1].

신뢰도 계산 방법으로는 scaled likelihood에 의한 신뢰도 계산, posterior probability에 의한 신뢰도 계산, normalized posterior probability에 의한 신뢰도 계산, entropy에 의한 신뢰도 계산, filler/garbage model과의 비교에 의한 신뢰도 측정, lattice를 이용한 신뢰도 측정 등 다양한 방법들이 이용되고 있다[4]. 신뢰도 계산을 위해서

#### II. 발화 검증

HMM 기반의 음성인식 시스템에서 우도 비 신뢰도 척도들이 폭넓게 사용되고 있다. 이 척도들은 가설된 단어 뿐만 아니라 filler, garbage 모델 또는 anti-model 들에 기반하고 있다. 가설 모델에 대한 score는 대수 우도 비로 이들 filler/anti-model의 score와 결합된다. 세그먼트  $n$ 에 대해서 가장 일반적으로 사용되는 신뢰도 척도는 다음의 대수 우도 비에 의해 표현된다[1].

$$LLR = \log \frac{P(O|H_0)}{P(O|H_1)} = \log \frac{P(O|\lambda_n^m)}{P(O|\lambda_n^a)} \quad (1)$$

여기서  $P(O|H_0)$ 는 target 모델  $H_0$  로 주어진 상태에서 관측된 음성 세그먼트의 확률이고,  $P(O|H_1)$ 는 anti-model  $H_1$  로 주어진 상태에서 관측된 음성 세그먼트의 확률, 그리고  $\lambda_n^m$  와  $\lambda_n^a$  는 각각 세그먼트  $n$ 에 대한 상응하는 모델과 anti-model이다.

세그먼트 길이에 독립적이기 위해서 식 (1)에 정의된

LLR은 일반적으로 다음 식과 같이 음성 세그먼트의 길이 ( $l_n$ )에 의해 정규화 된다[1].

$$LLR_n = \frac{1}{l_n} \log \frac{P(O/\lambda_n^m)}{P(O/\lambda_n^a)} \quad (2)$$

구문 상에서 신뢰도 척도는 전체 세그먼트에 걸친 신뢰 score를 합하고 구문의 전체 길이 (L)에 의해 정규화 하는 세그먼트에 기초한 척도로부터 얻을 수 있다[1].

$$CM = \frac{1}{L} \sum_n \log \frac{P(O/\lambda_n^m)}{P(O/\lambda_n^a)} \quad (3)$$

여기에서 CM은 confidence measure 이다.

신뢰 척도는 score를 다루고 종종 추가적인 rescoreing pass로 구현된다. rescoreing pass의 목적은 인식된 가설을 받아들이거나 거부를 결정하거나 또는 다음의 모듈에 의해 사용되는 인식된 가설을 위한 신뢰 score를 제공하기 위함이다. 그림 1은 신뢰 척도 계산의 기본 개념을 나타내고 있다[1].

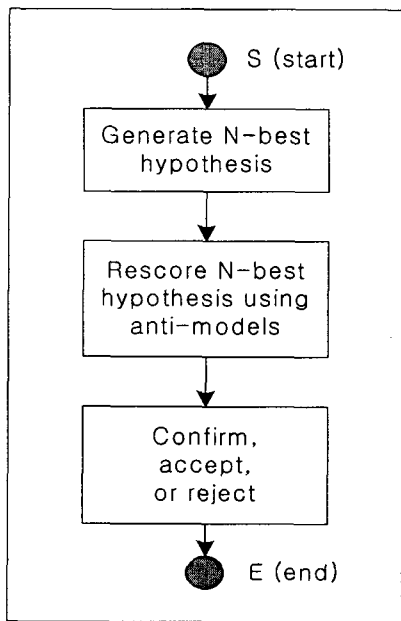


그림1. 신뢰 척도 계산을 위한 예

그림 2는 음소 filler 모델을 사용하여 핵심어 검출 방식을 이용하여 비 등록어를 거절시키는 방법이다. 핵심어 검출 방식은 핵심어(등록어) 모델과 filler(garbage) 모델을 사용하는 연결단어 인식 알고리즘을 기반으로 하고 있다. 여기서 filler 모델은 핵심어에 해당되지 않는 비 핵심어와 비 음성(묵음 또는 잡음) 구간들을 표현하는데 사용된다. 구성된 네트워크에서 인식된 결과는 등록어들

음소들의 열로 나타나게 된다. 즉, “묵음 + (등록어 및 음소들의 열) + 묵음”과 같은 형태가 된다. 여기에서 단어 페널티를 잘 조정하면 입력된 음성이 등록어 이면 인식된 결과는 “묵음 + (등록어 및 약간의 음소들의 열) + 묵음”으로 나타나게 되고, 비 등록어이면 인식된 결과는 “묵음 + (등록어 및 다수의 음소들의 열) + 묵음” 또는 “묵음 + (다수의 음소들의 열) + 묵음”으로 나타나게 된다. 발화 검증 시스템에서는 가변어휘 단어인식 시스템의 단어 페널티와 인식된 결과의 삽입된 음소들의 개수를 이용하여 비 등록어를 거절시킬 수 있다. 삽입된 음소들은 필러 모델을 뜻하며 삽입된 음소가 많다는 것은 인식 결과에 핵심어가 없다는 의미이다[2].

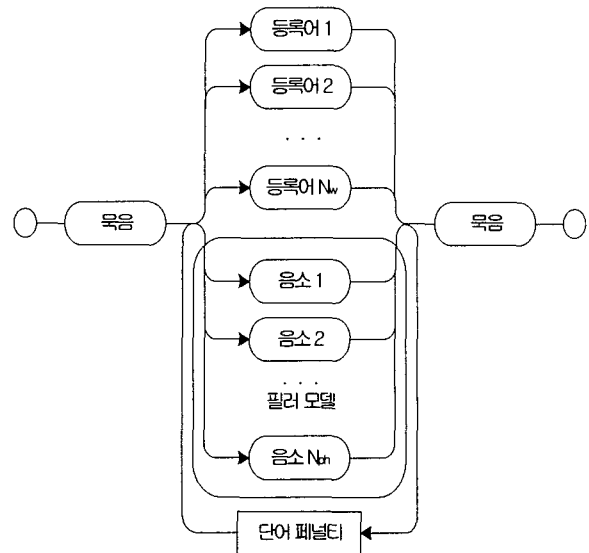


그림 2. 가변어휘 단어인식 시스템의 네트워크

### III. 모음 열을 이용한 발화 검증

본 논문에서는 그림 3과 같은 가변어휘 음성인식기를 이용한 발화 검증 시스템에서 anti-model을 인식 후보 단어의 모음 열을 사용하여 발화 검증하는 방법을 제안한다.

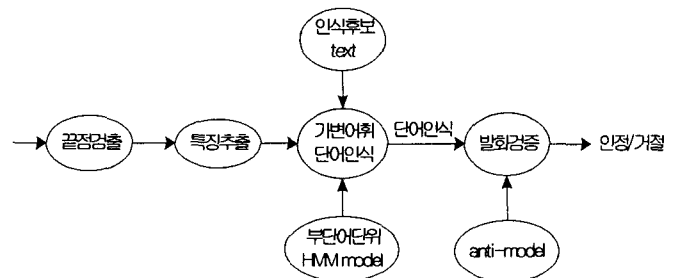


그림 3. 가변어휘 인식기와 발화 검증 시스템

그림 3에 본 논문에서 사용한 발화검증 시스템의 구조를 나타내었다. 가변어휘 단어인식을 위한 부 단어 단위 HMM 모델은 VCCV 단위를 기본으로 하여 구성하였다. 또한 본 논문에서 제안하는 발화 검증을 위해서 인식후보 단어의 모음 열을 anti-model에 적용한다. 표 1에 anti-model로 사용한 반 모음 모델 목록을 나타내었다. 또한 그림 4에는 모델의 구성 예를 나타내었다.

표 1. 반 모음 모델 목록

모 음										
단모음	아	애	어	오	우	으	이			
이중모음	야	애	여	와	왜	요	위	위	유	의

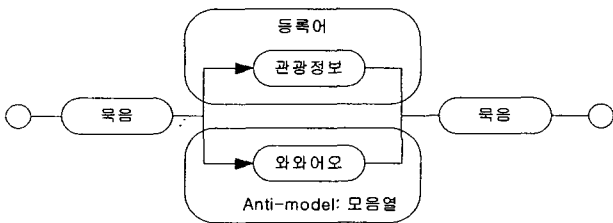


그림 4. 모델 구성 예

본 논문에서 사용한 평가 척도는 다음과 같은 변형된 식을 사용하였다.

$$\frac{\log P(O/H)}{\log P(O/H_v)} < T_h \quad (4)$$

여기에서 단어 모델을  $H$ 라고 할 때  $H_v$ 는 단어를 구성하고 있는 모음 열 모델로 정의한다. 또한  $T_h$ 는 임계값으로 정의한다. 임계값 이하이면 발화 검증된 것으로 판정하고, 그 이상은 거부한다.

#### IV. 실험 및 결과

본 논문에서 사용한 음성 데이터는 사무실 환경에서 20~30대 남녀 화자가 단음절 데이터 521종, 108개의 성이 포함된 성명 데이터 1,145종, PBW 1,001종, PRW 4,088종, 4연속 숫자 음성 35종이다. 이 데이터를 이용하여 CV, VCCV, VC 단위의 HMM을 구성하여 가변어휘 음성인식기를 구성하였다[7].

표 2에 발화 검증기의 성능을 나타내었다. 사용한 데이터는 영어 문장 인식을 위한 영어 문장을 한국어 가변어 음성인식기를 이용하여 인식하기 위해서 영어-한국어 음차 변환을 하여 인식하고 발화 검증을 실험한 결과이다.

여기에서 등록단어 검증은 인식 대상 등록어를 제대로 검증한 경우이고, 비 등록 단어 거부는 비 등록어에 대하여 거부한 경우이고, 등록단어 거부란 인식대상 등록어를 말했는데 거부한 경우, 그리고 비등록 단어 검증이란 비 등록 단어인데 검증한 경우를 말한다. 표 1에서 가장 좋은 결과는 초기치를 0.97로 하였을 때 올바른 검증이 79.5%  $\{=(83.3+75.7)/2\}$  정도의 성능을 나타내고 있음을 알 수 있다.

표 2. 발화 검증의 성능(%)

종류 초기치	등록 단어 검증	비등록 단어 거부	등록 단어 거부	비등록 단어 검증
	0.94	66.7	89.2	33.3
0.97	83.3	75.7	16.7	24.3
1.0	83.3	67.6	16.7	32.4
1.03	100	51.4	0	48.6

표 3. 발화 검증의 성능(%)

종류 초기치	등록 단어 검증	비등록 단어 거부	등록 단어 거부	비등록 단어 검증
	0.94	50	34.3	50
0.97	80	28.6	20	71.4
1.0	80	23	20	77
1.03	80	14.3	20	85.7

표 3에는 우리말의 단어에 대한 발화 검증 실험 결과이다. 실험결과 초기치 0.97일 때 올바른 검증이 54.3  $\{=(80+28.6)/2\}$ %로 문장보다 올바른 검증률이 훨씬 저하되었다. 따라서 발성 길이 정보를 이용한 정규화 기법이 도입된다면 좀더 나은 결과를 보일 것으로 예측된다.

#### V. 결론

본 논문에서는 인식을 위한 단어 내의 모음 열을 anti-model로 이용한 발화 검증 방법에 대하여 제안하여 실험하였다. 기존의 발화 검증을 위한 anti-model은 모델을 만들기 위해 특별히 훈련하거나 준비가 필요하지만 제안된 방법은 가변어휘 음성인식기만 있으면 발화 검증

이 가능하다. 또한 발화 검증 단계에서 삽입된 음소의 개수 등을 이용하는 등의 처리과정을 거치지 않고 문턱 값에 의해서만 검증을 하는 효율적인 방법이다.

추후의 과제는 발성 음의 길이 정보를 이용한 실험을 보완하여 평가하고, 연속 음성인식에 적용해 그 성능을 확인 하고자 한다.

#### <감사의 글>

본 연구는 한국과학재단 목적기초연구(R05-2002-001007-0)지원으로 수행되었음.

#### 참고문헌

- [1] Jean-Claude Junqua, "Robust Speech Recognition in Embedded Systems and PC Applications", Kluwer Academic Publishers, 2000.
- [2] 김기태 외 4인, "가변어휘 단어 인식에서의 미등록어 거절 알고리즘 성능 비교", 한국 음향학회지, 제 20권 2호, pp.27-34, 2001.
- [3] 김철 외 4인, "정규화 신뢰도를 이용한 핵심어 검출 성능향상", 한국 음향학회지, 제 21권 4호, pp.380-386, 2002.
- [4] 김영수, "DB 질의 음성 대화 시스템을 위한 단어 신뢰도 평가", 서강대학교 석사학위논문, 2000.
- [5] M.-W. Koo and S.-J. Lee, "An Utterance Verification System based on Subword Modeling for a Vocabulary Independent Speech Recognition System", Eurospeech 99, pp.291-294, 1999.
- [6] X. Huang, A. Acero and H.-W. Hon, "Spoken Language Processing", CMU, 2000.
- [7] 윤재선, 홍광석, "VCCV 단위를 이용한 어휘독립 음성인식 시스템의 구현," 한국 음향학회지, 제 21권 2호, pp160-166, 2002.