



## 음성인식 기반 응급상황관제\*

### Emergency dispatching based on automatic speech recognition

이 규 환 · 정 지 오 · 신 대 진 · 정 민 화\*\* · 강 경 희 · 장 윤 희 · 장 경 호

Lee, Kyuwhan · Chung, Jio · Shin, Daejin · Chung, Minhwa · Kang, Kyunghee · Jang, Yunhee · Jang, Kyungho

#### Abstract

In emergency dispatching at 119 Command & Dispatch Center, some inconsistencies between the ‘standard emergency aid system’ and ‘dispatch protocol,’ which are both mandatory to follow, cause inefficiency in the dispatcher’s performance. If an emergency dispatch system uses automatic speech recognition (ASR) to process the dispatcher’s protocol speech during the case registration, it instantly extracts and provides the required information specified in the ‘standard emergency aid system,’ making the rescue command more efficient. For this purpose, we have developed a Korean large vocabulary continuous speech recognition system for 400,000 words to be used for the emergency dispatch system. The 400,000 words include vocabulary from news, SNS, blogs and emergency rescue domains. Acoustic model is constructed by using 1,300 hours of telephone call (8 kHz) speech, whereas language model is constructed by using 13 GB text corpus. From the transcribed corpus of 6,600 real telephone calls, call logs with emergency rescue command class and identified major symptom are extracted in connection with the rescue activity log and National Emergency Department Information System (NEDIS). ASR is applied to emergency dispatcher’s repetition utterances about the patient information. Based on the Levenshtein distance between the ASR result and the template information, the emergency patient information is extracted. Experimental results show that 9.15% Word Error Rate of the speech recognition performance and 95.8% of emergency response detection performance are obtained for the emergency dispatch system.

**Keywords:** automatic speech recognition, emergency dispatching, word segmentation, semantic analysis

#### 1. 서론

현행 시·도 소방본부 또는 소방서의 종합상황실에 배치된 전산·통신요원인 수보요원은 화재 등 각종 재난 발생의 신고 접수와 통보·전달 및 출동의 지령 등의 업무를 수행하고 있다. 수보요원은 ‘긴급구조표준시스템’에 따라 신고 접수를 수행하고 있으나, 구급 상황의 경우 ‘119구급상황관리센터 상담 매뉴얼’의 ‘도입부 수보 프로토콜’을 활용하도록 되어 있다. 따라서 수보요원은 ‘긴급구조표준시스템’에 따라 신고 접수 업무를 수행

하되, 구급 상황의 경우에는 개요/환자의 연령/환자의 의식/환자의 호흡/환자의 심정지/호흡 정지 등 등을 파악하고, 특히 개요의 긴급 출동상황인 경우 이를 지령해야 한다. 그러나 이들의 업무 수행에 필수적으로 활용하는 ‘긴급구조표준시스템’과 ‘수보 프로토콜’이 연계성을 갖지 못함으로써 수보업무의 효율성이 떨어지고 있다[10]. 따라서 수보요원의 업무를 효과적으로 지원하고 위기 상황 대응의 신속성과 정확성을 높이기 위한 기술 개발로 음성인식 기술을 수보요원 프로토콜에 적용해 보고자 한다.

\* 본 연구는 국민안전처 소방안전 및 119구조 구급기술연구개발사업(MPSS-소방안전-2015-71)의 연구비 지원으로 수행되었습니다.

\*\* 서울대학교, mchung@snu.ac.kr, 교신저자

Received 30 May 2016; Revised 20 June 2016; Accepted 22 June 2016

전 세계적으로 응급상황관제시스템은 크게 두 가지 유형, 즉 프로토콜(protocol) 형태의 MPD(Medical Priority Dispatch)와 가이드라인(guideline) 형태의 CBD(Criteria Based Dispatch)로 구분되며, 국가별 또는 지역별 여건에 따라 사용되는 응급상황관제 시스템이 다르다. 현재 사용하는 프로토콜 형태의 MPD방식은 프로토콜 시퀀스를 따르는 대화가 이루어지기 힘든 단점이 있고, 가이드라인 형태의 CBD방식은 교육 훈련을 많이 받아야 하고 경험이 풍부해야 하는데 현실적으로 그렇지 못한 단점이 있다.

음성인식 기술을 응급상황관제시스템에 적용하면 대화중에 자동으로 환자정보가 검출되도록 개발되어 현실적인 어려움을 개선할 수 있을 것이다[8]. 수보요원은 신고대화에 집중할 수 있고 응급출동 판단에 필요한 정보제공을 받기 때문에 업무 부담이 경감되고, 응급관제시스템은 응급출동에 대한 효율성이 높아지고, 훈련이 부족한 수보요원일지라도 음성인식 기술의 정보제공으로 원활하게 신고접수 업무를 수행할 수 있을 것이다.

먼저 현재 응급상황관제 시스템에 음성인식 기술을 적용 방법에 대해서 살펴본다. 그리고, 응급관제시스템의 대용량 음성인식기 개발에 대해서 살펴본다. 대규모 말뭉치 어절 처리 방식, 음향모델, 언어모델, 후처리 방법을 기술하고, 마지막으로 응급상황관제시스템의 성능을 살펴볼 것이다.

## 2. 음성인식기술 적용 방법

음성인식 기술을 응급관제시스템에 적용하기 위해, 2014년도 00지역 실제 구급 신고 전화중 개인정보 제거한 6,600건에 대해 전사하였다.

구급 신고 전화 전사의 주요목적은 증상표현에 대해서 분석하기 위해서이다. 신고자의 증상 표현의 경우, 동일 증상에 대한 다양한 표현이 존재하고 다른 증상임에도 동일하거나 유사한 표현이 존재한다. 수보·상담 요원의 신속한 증상 판단을 돕고 오판을 방지하기 위해 각종 구급 증상에 대한 표현들을 분석하고 해당 구급 증상의 전형적이고 일반적인 증상표현 목록을 사용하였다[9]. 전형적이고 일반적인 증상 표현은 신고자의 환자 정보 자동추출 템플릿 구성에 사용한다.

시스템에서 신고전화는 자동으로 저장되는데 8kHz, 8bits, mono 형식 혹은 8kHz GSM6.10 코딩형식과 같이 다양한 형식으로 저장되어 있다. 파일 형식 뿐 아니라 채널 분리가 되어 있지 않고 신고자와 수보요원의 음성이 겹쳐 있기 때문에, 실제 음성 데이터를 음성인식 실험이나 음향모델 학습용으로 사용하기에 적합하지 않다. 따라서 실험에 사용될 음성 파일은 실제 전사 시나리오를 기반으로 녹음하여 실험에 사용한다.

구급 신고 전화통화는 주변잡음, 신고자의 감정 흥분 상태, 신고자의 비정상적인 발화가 많고 이 같은 발화를 인식하는 경우 인식오류가 많아지므로 비교적 안정적인 발화를 하는 수보요원의 발화만을 인식한다. 또 수보요원의 통화내용만으로는 환자정보를 음성인식 기술로 추출해 낼 수 없기 때문에 전사 시나리오에 기반하여 녹음 시에 환자정보를 포함시켜 사용한다.

제안되는 응급상황관제시스템은 신고접수 대화중 수보요원이 환자정보 포함한 발화를 바로 인식하고, 인식결과로부터의 미추출 후처리 방법을 이용하여 신고환자의 증상정보를 추출하고, 추출정보에 대해서 응급/비응급 예측판단 할 수 있도록 하였다.

## 3. 응급상황관제시스템을 위한 연속음성인식기

본 연구는 대화체 음성을 그 대상으로 하며, 인식된 텍스트를 기반으로 다양한 지식 정보를 추출하는 음성기술 서비스로 수보요원 도우미 역할을 한다. 자연어 대화체 음성을 인식하기 위해 대용량 음성인식기 모델 개발과 녹취록 분석에서 사용되는 후처리 기법이 접목되어 사용되고 응급상황관제시스템에 적용하기 위한 한국어 연속음성인식 시스템을 개발하였다. 연속음성인식모델 개발을 위해 00 소방안전본부에서 제공한 2014년 4월 8일~12월 31일까지의 수보단계 신고 접수 음성 6천6백 건에 대한 전사(Transcription) 데이터와 일반 뉴스, SNS, 일반 대화문장 관련 데이터 등을 함께 이용하였다. 그리고 연속음성인식 개발 툴로 Kaldi toolkit[3]을 사용하고, 서비스 처리를 위해 <그림 1>과 같은 서버-클라이언트 시스템을 구성하였다[1].

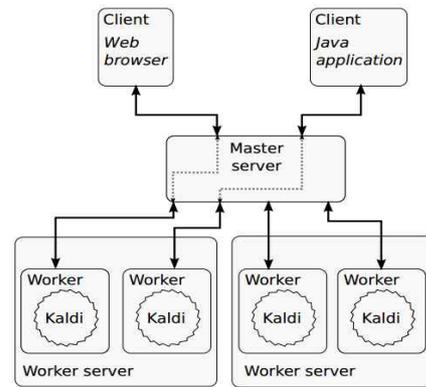


그림 1. 인식기 클라이언트-서버 구조  
Figure 1. client-server architecture in ASR

<그림 1>에서 보는 바와 같이, 클라이언트 부분은 응급상황관제 수보요원의 신고접수 프로그램 내에 내재되는 형태로 설계하고, 서버 부분은 클라이언트와 송수신을 담당하는 마스터 서버와 음성인식을 담당하는 워커 서버로 분리하여 음성인식 서버 확장이 용이 하도록 설계 하였다. 연속음성인식을 위해 심층 신경망(Deep Neural Network) 모델을 이용하는 Kaldi-mnet2-online 디코더를 이용하였다[4].

### 3.1. 어절 분할(Word Segmentation)

한국어에 있어서 띄어쓰기는 음성인식 개발자들에게 고민거리 중 하나이다. 음성인식에서 띄어쓰기 단위는 바로 발음사전의 단위가 되기 때문인데 이를 해결하고자 형태소 해석기를 이용하여 적절한 띄어쓰기를 해 왔었다. 그러나, 현재 대규모 말뭉치가 사용되면서 고유명사나 약어 등이 많이 포함될 수밖에 없기

때문에 형태소 해석기 사용이 어렵게 되고 있다. 특히 블로그, 페이스북 등 인터넷 상에서 대용량의 말뭉치에는 각종 심볼, 축약된 단어, 신조어, 조합형만 표현가능한 문자, 맞춤법에 맞지 않거나, 영어와 한글이 같이 포함되어 있고 자동 수집하는 과정에서 띄어쓰기가 무시되고 수집되는 경우가 많다. 심볼, 약자, 영어 등은 정체를 통해 어느 정도 해결할 수 있지만 더욱 문제점은 말뭉치가 띄어쓰기 없이 수집되어 수작업으로 어절 분리를 할 수 없다는 것이다. 구글에서는 앞서 제기한 여러 문제를 해결하기 위해 문자(한글), 숫자, 영문자 등을 구분하지 않고 모두 심볼로 처리하면서 대용량의 텍스트 말뭉치로부터 통계적 빈도수로 인식단위를 추출하는 WordPiece Model[5] 이라는 개념을 사용했다. WordPiece Model은 모든 음절의 통계치를 기본으로 시작하여 음절을 붙여가는 방식인데 이 방법은 구글에서와 같이 대용량의 텍스트 말뭉치를 보유하고 있어야 의미 있는 인식단위가 추출되고 많은 계산량을 필요로 하기 때문에 실제 사용하기 어려움이 있다. 본 연구에서는 구글의 WordPiece Model과는 다르게 음절이 아닌 어절단위 Unigram과 Bigram 빈도수를 이용하여 긴어절을 분할하고자 한다.

보유한 텍스트 말뭉치는 낭독체 스타일의 말뭉치가 UTF-8 형식으로 약 22 Gbytes와 대화체 스타일의 말뭉치 약 900 MBytes 용량을 사용하였다. 전체 보유한 말뭉치에 대해서 Unigram과 Bigram의 빈도수를 분석해 보았다. 구해진 Unigram 23,604,725 개 Bigram 242,864,185개로 인식 사전의 인식단위 후보로 모두 사용하기에 너무 많아 이를 효과적으로 줄이기 위해서 휴리스틱한 방법을 사용하였다.

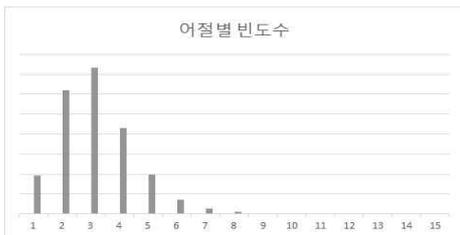


그림 2. 어절별 빈도수  
Figure 2. Eojeol frequency

먼저 말뭉치에 있는 긴 어절들은 Bigram 빈도수가 지극히 작기 때문에 전체 말뭉치 Bigram에서 빈도수 10 이상과 13음절수 이상을 임계값으로 가지치기 하여 17,348,003 개의 Bigram을 얻었고, 여기에는 1,621,803개의 Unigram이 포함되어 있다. 전체 Unigram 빈도수 분포<그림 2>를 보면 2음절에서 4음절의 분포가 전체의 약 78%를 차지하고 있음을 이용하였다. Unigram에서 5음절 이상인 어절은 제외하면 1,249,073개의 Unigram을 얻었다.

최종적으로 얻어진 Unigram, Bigram 빈도수는 어절 분할 score 계산에 사용된다. 어절 분할 알고리즘을 코드로 보면 아래와 같다.

```
def EojeolSeg(text):
    def search(text, prev=""):
        if text == "":
            return 0.0, []
        def candidates():
            for A, B in divide(text):
                A_score = log10(score(A, prev))
                B_score, B_words = search(B, A)
                yield (A_score+B_score, [A] + B_words)
            return max(candidates())
        result_score, result_words = search(text)
        return result_words

    def score(word, prev=None):
        if prev is None:
            if word in Unigram:
                return Unigram[word] / TOTAL
            else:
                return 1.0 / TOTAL
        else:
            bigram = (prev, word)
            if bigram in Bigram and prev in Unigram:
                return Bigram[bigram] / TOTAL / score(prev)
            else:
                return score(word)
```

어절 분할 알고리즘을 이용하여 긴 어절 ‘방어운전하시구요’를 입력으로 주었을 때 ‘방어 운전 하시구요’로 출력되는 예제이다.

```
score 계산: log10(Pr(어절))
입력문자열: 방어운전하시구요.(8음절)
띄어쓰기 총 가지수: 128가지중 최대 scores값 조합을 선정
"방" "어운전하시구요" : -3.97-24.43=-28.4
"방어" "운전하시구요" : -4.66-21.43=-26.09
"방어운" "전하시구요" : -12.43-18.43=-30.86
"방어운전" "하시구요" : -15.43-4.66=-20.09
"방어운전하" "시구요" : -18.43-6.03=-24.46
"방어운전하시" "구요" : -21.43-3.84=-25.57
"방어운전하시구" "요" : -24.44-3.07=-27.51
"방어" "운전" "하시구요" : -4.66-3.38-4.66=-12.7

출력 문자열: 방어 운전 하시구요
```

### 3.2. 언어모델

요즘은 DNN을 언어모델에도 많이 적용하기는 하지만 차원이 높은 벡터가 입력으로 사용되면 학습시간이 많이 소요된다. 이

를 극복하고자 하는 연구[11]가 있으나 본 연구에서는 형태소 해석을 하지 않고 어절 분할 결과를 이용하기에 적합하지 않다. 한국어 연속음성인식을 위한 언어모델은 영역별 말뭉치 데이터로부터 추출한 403K개 유사어절 형태의 인식어휘를 기반으로 SRI LM(SRI Language Modeling Toolkit)[6]을 사용하여 5-gram으로 구성되었다. 언어모델 훈련에 사용된 말뭉치 데이터는 <표 1>과 같다.

**표 1. 언어모델 말뭉치**  
**Table 1. Language Model Text Corpus**

텍스트 영역	텍스트 파일크기 (CP949 encoding기준)	비율 (%)
news	5.8 Gbyte	42.3
일반 대화(영화자막)	1.3 Gbyte	9.5
sns	6.6 Gbyte	48.2
응급상황 관제관련	2.5 Mbyte	-
총합	13.703 Gbyte	100

<표 1>에서 보이는 언어모델 훈련용 텍스트 데이터는 인터넷 뉴스, SNS(Social Network Service) 게시된 글, 위키(Wiki) 게시된 글, 영화 자막(대화) 등의 일반영역의 수집된 데이터들과 00 소방안전본부에서 제공한 2014년 4월 8일~12월 31일까지의 수보 단계 신고 접수 음성 6,600건에 대한 전화음성 텍스트 전사(Transcription) 데이터로 구성되었다. 이들 데이터는 텍스트 정규화 및 정제 작업을 거쳐 어절 분할 알고리즘을 통해 언어모델 학습용 데이터로 구성된다.

결과 언어모델은 각 텍스트 영역별 언어모델을 따로 구성하여 언어모델 Interpolation 방법을 이용해 통합하였다. 응급상황 관제관련 말뭉치가 다른 영역에 비해 많이 부족하여 통합해서 언어모델을 개발하면 응급상황관제 관련 어절 빈도수가 적어 가지치기 될 수도 있기 때문이다.

**표 2. 언어모델 파라미터 개수**  
**Table 2. Language Model Parameters**

1-gram	403,223
2-gram	49,596,627
3-gram	11,923,472
4-gram	707,960
5-gram	75,523

언어모델의 파라미터는 <표 2>와 같다.

결과 언어모델을 응급상황관제 관련 말뭉치를 이용한 PPL 측정 시, 평균 100 미만의 값을 얻을 수 있었다.

### 3.3. 음향모델

본 연구에서 개발된 한국어 연속음성인식 음향모델 단위는 40개의 한국어 음소를 기반으로 문맥 의존 트라이폰(Context

dependent Triphone) 모델로 구성하였다. 음향모델 구조는 DNN-HMM(Deep Neural Network - Hidden Markov Model) 형태로 구성되었다[3].

음성 데이터는 8kHz sampling rate, signed 16bits 데이터로 총 1,372시간이며 테스트 데이터로 6시간 (학습에 사용하지 않은 약 5,888문장 발화파일) 그리고 훈련 데이터로 약 1,366 시간을 사용하였다.

**표 3. 음향모델 학습데이터**  
**Table 3. Acoustic Model Training DB**

발화종류	영역	문장	시간
대화체+낭독체	로봇잡음/채팅 건강 일정관리	73,333	72.8
대화체+낭독체	서울말,상담,계획,채팅, 건강,일정관리	201,973	255.8
낭독체	검색기기	319,991	489.1
낭독체	음향모델 학습용(뉴스,신문,방송..)	578,796	555.1
계		1,174,093	1,372.8

훈련 음향모델은 DNN-HMM 형식으로 HMM은 8,961개의 tied state 모델로 구성되었고, DNN 구조는 3개의 은닉층(Hidden Layer)과 입력층 그리고 출력층, 총 5개의 층으로 구성되었다. 입력특징 벡터의 기본 단위는 25msec 윈도우와 10msec 단위 슬라이딩을 통해 주파수 도메인에서 MFCC(Mel-Frequency Cepstrum Coefficient) 40차원을 추출하고 현재 프레임과 이전 프레임 2개, 이후 프레임 2개 총 5개 프레임의 200 차원 특징벡터와 칼디(kaldi) 온라인 인식기에서 사용하는 iVector 모델의 출력 특징벡터 100차원을 포함 총 300차원의 벡터를 구성 후, LDA 변환 후 (차원 감소 없음) 입력 벡터로 사용하였다. <표 4>은 음향모델의 전체 구조를 설명한다. <표 4>에서 DNN의 은닉층에 사용된 비선형 함수(Nonlinear component)인 Pnorm, p=2를 사용하였다[7]. 비선형 함수 Pnorm을 이용한 Pnorm 네트워크 학습 후 안정성 확보를 위해 정규화(normalize) 과정이 추가되었다. 이후, 감소된 차원들은 각 은닉층 별 설정된 splice index 값에 따라 출력 벡터들이 합쳐져 최종 출력벡터로 구성된다. 그 다음 은닉층에서 선형변환을 통해 차원이 2,750차원 그리고 마지막 층에서는 8,961차원으로 늘어나는 구조를 가지고 있다.

**표 4. 연속음성인식시스템 DNN-HMM 구조**

**Table 4. DNN-HMM Structure in the Continuous Speech Recognition System**

Input Layer	1단계: Spliced 입력벡터차원 = 300 (MFCC 40 x 5 + iVector 100) splice context: t-2, t-1, t, t+1, t+2 사용 2단계: LDA Transform 적용 (차원 감소 없음)
-------------	--

Hidden Layer 1	입력: 300차원 Layer1 뉴런수(노드수): Linear component : 2750 차원 Nonlinear component : Pnorm & Normalize (2750 => 275) 275차원 출력: 550차원 (splice context: t-1, t+2)
Hidden Layer 2	입력: 550차원 Layer2 뉴런수(노드수): Linear component : 2750 차원 Nonlinear component : Pnorm & Normalize (2750 => 275) 275차원 출력: 550차원 (splice context: t-3, t+3)
Hidden Layer 3	입력: 550차원 Layer3 뉴런수(노드수): Linear component : 2750 차원 Nonlinear component : Pnorm & Normalize (2750 => 275) 275차원 출력: 550차원 (splice context: t-7, t+2)
Output Layer	입력: 550차원 Layer4 뉴런수(노드수): Linear component : 2750 차원 Nonlinear component : Pnorm & Normalize (2750 => 275) 275차원 출력: 8,961 차원 (softmax component)

응급상황관제시스템 개발에서 개발된 연속음성인식 시스템의 인식단어 오류율은 <표 5>과 같다. 인식어휘 403K를 기반으로 개발된 5-gram 언어모델을 이용하여 연속음성인식 네트워크를 구성하였고 5,888 문장 발화 테스트 데이터를 사용하여 성능을 측정하였다. 연속음성인식기 성능 측정에 사용되는 사전의 크기는 40만 어휘이다.

표 5. 연속음성인식 성능  
Table 5. Continuous speech recognition performance

사용기술	학습데이터 양	WER (단어 오류율)
GMM-HMM	1,366 시간	15.12%
DNN-HMM	1,366 시간	7.72%

### 3.4. 의미 추출

본 연구에서 의미 추출하는 방법으로 심플한 레벤슈타인 거리 (Levenshtein Distance) 측정방법[2]을 사용하였다. 레벤슈타인 거리 측정값은 두 문자열이 얼마나 다른지를 나타내는 거리 값으로써 주로 철자 검사, 간단한 의미분석, DNA 분석, 표절 검사 등에 널리 사용되는 기술이다. 두 문자열 a, b 사이의 레벤슈타인 거리는 다음과 같이 정의된다.

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 2_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

여기서  $2_{(a_i \neq b_j)}$  는 두 문자열 중 a 문자열의 i번째 문자가 b 문자열의 j번째 문자와 같을 경우 ( $a_i = b_j$ ), 값은 0과 같고, 다르면 값이 2와 같음을 의미한다. 일반적인 레벤슈타인 거리 측정에서의 패널티 값은 1을 보편적으로 많이 사용하지만 본 연구에서는 문자열이 달라질 경우 거리 값이 더 커지는 2를 선택하였다.

두 문자열 a, b 사이의 레벤슈타인 거리 측정값을 이용해 다음과 같이 레벤슈타인 거리비율을 구할 수 있다.

$$lensum = |a| + |b|$$

$$ratio = \frac{(lensum - lev_{a,b})}{lensum}$$

여기서 |a|와 |b|는 두 문자열의 길이를 의미한다. 구해진 레벤슈타인 거리비율은 핵심인식대상별 문턱값(Threshold Value)과 비교분석하여 의미추출에 사용된다. 문턱값은 추출대상 단어의 길이에 따라 차등을 두었는데, 예를 들어 성별, 연령, 신고내역, 신고자본인확인, 주증상 등의 범주내 단어 추출에서는 0.9 값 이상으로 세팅하여 단어가 정확하게 일치하지 않을 경우 추출하지 않도록 하였고, 그 외의 경우, 추출 단어가 길 경우에 0.8~0.88까지로 설정하였다.

본 연구에서 레벤슈타인 거리측정 알고리즘을 적용한 의미추출 구현방법은 <그림 3>과 같다.

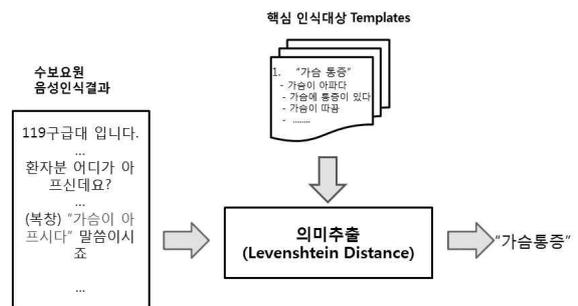


그림 3. 레벤슈타인 거리기반 의미추출  
Figure 3. Semantic extraction based on Levenshtein distance

<그림 3>에서 보는 바와 같이 레벤슈타인 거리기반 의미추출 알고리즘은 핵심인식대상에 대한 템플릿과 119 상황실 수보요원 발화의 음성인식결과와 문장들에 대한 비교분석을 통해 핵심인식대상을 추출한다. 핵심인식대상은 "119 구급상황관리센터 상담 메뉴얼"의 구급대 출동대상 범주 분류표에서 추출된 31가지 주증상 <표 6>, 응급출동 <표 7>과 인적사항 <표 8>를 추출할 수 있도록 설계되었다.

**표 6.** 31개 범주 환자 주증상  
**Table 6.** 31 Categories with patients symptom

1. 두통	11. 심계항진	21. 비출혈
2. 흉통	12. 심정지	22. 질출혈
3. 복통	13. 경련	23. 그 밖의출혈
4. 요통	14. 발작	24. 고열
5. 분만진통	15. 실신	25. 저체온증
6. 그 밖의 통증	16. 오심/구토	26. 현기증/어지러움/ 현훈
7. 의식장애	17. 설사/변비	27. 마비
8. 기도이물	18. 배뇨장애	28. 전신쇠약
9. 호흡곤란	19. 객혈	29. 정신장애
10. 호흡정지	20. 토혈	30. 그 밖의 이물질
		31. 기타/골절/열상/ 염좌/철과상/탈구

**표 7.** 응급출동  
**Table 7.** Emergency response

신고내역	목땀, 목졸림, 질식, 물에 빠짐
환자상태	호흡 여부: 호흡있음/호흡없음 의식 여부: 의식있음/의식없음

**표 8.** 인적사항  
**Table 8.** Personal information

신고자 본인확인	유/무
성별	성별: 남성/여성
연령	0-4세, 5-14세, 15세-39세, 40-69세, 70세 이상

핵심인식대상 추출을 위해서는 각 대상에 대한 템플릿을 구성하였다. 31가지 환자 주증상을 표현/의미하는 템플릿들은 구급 신고전화에서의 한국어 증상 표현연구[9]의 증상별 언어표현을 기반으로 구성하였다. 어떤 증상의 언어표현이 여러 주증상 표현에 포함되면 여러 주증상이 검출될 수 있다. 이럴 경우 현실적으로 수보요원은 표시된 여러 주증상을 보고 질문을 통해 주증상 대상을 줄여 나가는 방법이 적합하다. 아직은 소규모의 전사 말뭉치로 조사 분석되었기 때문에 31개 주증상 표현 템플릿을 정교하게 설계하기에는 한계가 있기 때문이다.

구급 신고 전화에 응급출동 템플릿 구성은 핵심어 중심으로 템플릿을 구성하였다. 핵심어들은 상당히 위급상황을 나타내는 단어들로 다른 주증상이나 인적 사항 등을 알아볼 여유가 없이 핵심어가 검출이 되면 바로 응급출동 해야 한다. 응급 출동에서 특이한 점은 31가지 주증상 표현에 모두 공통으로 포함될 수 있는 호흡/의식 여부가 포함되어 있다. 호흡이나 의식이 없다는 것은 심정지등 매우 위급한 상태일수 있기 때문에 응급출동에 포함시킨다.

인적사항에 대한 템플릿은 간단한 신상정보와 본인여부이다. 구급 신고접수에서 본인여부가 중요한 이유는 신고접수를 본인이 직접 했으면 일단 호흡과 의식은 있는 것으로 판단하고 응급일 확률은 낮아지기 때문이다. 성별이나 나이를 짐작할 수 있는 단어들에 대해서 하나의 단어에서 2가지 핵심 정보를 추출할 수 있도록 템플릿을 구성하였다.

할아버지 => 남성, 70세  
아줌마 => 여성, 40-69  
옆집 사람이, 술취한 사람이 => 본인 아님

#### 4. 실험 환경 및 결과

##### 4.1. 실험환경

실험은 두 가지 측면에서 시행한다. 첫 번째는 40만 어휘 대용량 음성인식기의 자체 성능을 실험하기 위해 음성인식물을 알아보고, 두 번째는 인식결과에서 의미추출 성능을 알아보기 위해 응급출동 검출성능을 실험한다.

실험을 위해 전사한 6,600건(신고접수)에서 구급활동일지와 국가응급환자진료정보망(NEDIS: National Emergency Department Information System) 정보를 연계해서 주증상과 구급환자분류가 확인된 68건을 추출하였다. 68건의 신고통화에 대해서 전사 수정작업을 하였다. 전사 수정작업은 개인정보는 삭제되었기 때문에 가상의 주소로 치환하여 실제 전사기반으로 자연스럽게 수정하는 작업과, 환자 정보를 자연스럽게 포함하도록 수정하는 작업을 병행했다. 환자 정보 수정은 2가지 정보를 포함하도록 수정해야 하는데 ‘어느 템플릿에 템플릿 상태가 어떻다’ 형식으로 이루어졌다. 의미추출이 인식결과와 템플릿들간의 레벤슈타인 거리 기반으로 이루어지기 때문이다.

수보요원: 예 119입니다.  
신고자: 할아버지가 길에 쓰러져 있어요  
수보요원: 예 어디예요? (실제 전사)  
=> 수보요원: 예 할아버지가 어디에 쓰러져 계세요?  
[성별: 남자, [나이: 70대]

수정된 전사 스크립트는 68개 시나리오에서 총 818개의 문장으로 구성되었고, 이는 녹음하여 테스트 입력 음성파일로 사용하였다.

음성인식물 실험에는 818개 음성파일을 입력으로 사용하여 인식물을 실험하였고, 의미추출 성능을 알아보기 위해서 68개 시나리오별로 입력하여 응급출동 검출이 되는지 실험을 하였다.

표 9. 구급 신고전화에서의 한국어 증상 표현 예시

Table 9. Korean symptoms represent an example of the ambulance call report

핵심인식 대상(예시)	실제 증상 표현	증상표현	템플릿 내용
심정지	<p>쓰러졌다. 쇼크상태                      숨을 잘 못 쉰다.                      마비가 와서 몸을 못 움직인다.                      자살 / 헛바닥 나눔 / 피 나눔                      응급하신데요 / 의식이 없음                      숨을 못쉬고 하네                      의식이 없다.                      호흡이 안 되다. 의식이 없다.                      사람이 목을 매다.                      심장이 안 좋다                      쓰러져 있다.                      의식이 없다, 입에서 거품이 난다                      질식했다, 음식물이 목에 걸렸다                      정신을 못 차린다, 숨을 잘 못 쉬는 거 같다                      목을 매다                      숨을 안 쉰다                      쓰러졌다, 의식이 없다                      쓰러졌다, 의식이 없다                      정신이 없는 거 같다                      의식이 없다                      의식과 호흡이 없다                      초점이 없다, 숨을 이상하게 쉰다                      각혈한다                      숨을 안 쉬는 거 같다                      의식이 없다, 돌아가신 거 같다                      숨을 잘 안 쉰다                      기력이 없다, 숨을 안 쉬는 거 같다                      쓰러졌다, 의식은 있다                      숨을 안 쉬는 거 같다                      의식, 호흡이 없다                      의식이 없는 것 같다, 숨을 안 쉰다                      목을 메었다                      숨 쉬기가 곤란하다                      숨을 안 쉰다, 이상한 소리를 낸다                      숨을 안 쉰다                      쇼크가 왔다                      의식이 없다                      죽었다, 의식이 없다                      의식이 없다                      쓰러졌다, 의식이 없다, 호흡은 조금 있는 것 같다                      호흡이 없다, 의식이 없다                      죽은 거 같다, 의식이 없다, 죽사한 거 같다                      의식이 없다, 숨을 안 쉬는 거 같다                      숨을 몰아쉰다, 불러도 대답을 안 한다                      자살 시도했다                      쓰러졌다, 의식이 없다, 호흡은 있다                      숨을 안 쉰다                      의식이 없다, 호흡도 없다                      허리, 다리 다치다.                      또 이상하다. 식은땀 나다.                      막 떨고 있다.                      알 수 없음                      숨을 안 쉬시는 것 같다.                      숨이 가쁘다.</p>	<p>쓰러지다                      쇼크                      숨을 안 쉬다                      숨을 (잘) 못 쉬다                      숨을 몰아쉬다                      숨을 이상하게 쉬다                      숨이 가쁘다                      숨 쉬기가 곤란하다                      마비가 오다                      몸을 못 움직이다                      호흡이 안 되다                      호흡이 없다                      헛바닥 나오다                      의식이 없다                      목을 매다                      자살                      심장이 안 좋다                      입에서 거품이 나다                      질식하다                      정신이 없다                      정신을 못 차리다                      초점이 없다                      기력이 없다                      이상한 소리를 내다                      대답을 안 하다                      죽다                      죽사한 것 같다                      피 나다                      음식물이 목에 걸리다(원인)                      각혈한다</p>	<p>쓰러지-                      쇼크                      숨을 안 쉬-                      숨을 잘 못쉬-                      숨을 몰아쉬-                      숨을 이상하게 쉬-                      숨이 가쁘-                      숨 쉬기가 곤란하-                      마비가 오-                      몸을 못 움직이-                      호흡이 안되-                      호흡이 없-                      헛바닥이 나오-                      의식이 없-                      목을 매-                      자살하-                      심장이 안 좋-                      입에서 거품이 나-                      질식하-                      정신이 없-                      정신을 못 차리-                      초점이 없-                      기력이 없-                      이상한 소리를 내-                      대답을 안 하-                      죽어-                      죽사한 것 같-                      피 나-                      음식물이 목에 걸                      각혈하-</p>

자동으로 응급출동이 검출되는 <그림 4>의 시나리오로 실험을 하였다. 출력 결과는 응급/비응급/예측실패 3종류로 출력된다. 예측실패는 환자 정보 추출에 실패했거나, 입력 데이터에 응급/비응급 판단을 위한 환자정보가 없을 때 출력하도록 시스템을 설계하였다.

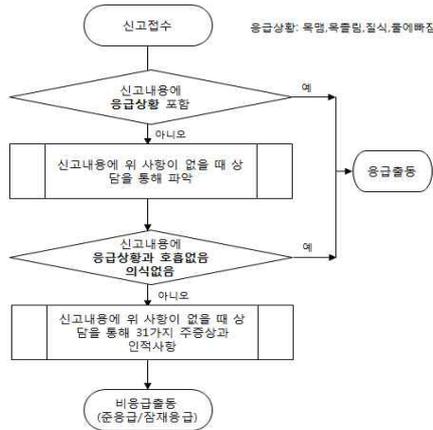


그림 4. 응급출동 시나리오 순서도  
Figure 4. Flowcharts of Emergency response scenarios

#### 4.2. 실험결과

테스트 입력 파일의 혼잡도(Perplexity)는 약 74 정도이고 음성 인식 결과는 <표 10>과 같다.

표 10. 인식 성능  
Table 10. Recognition performance

WER 9.14% [ 348 / 3806, 39 ins, 73 del, 236 sub ]
SER 24.21% [ 198 / 818 ]
(WER: word error rate, SER: sentence error rate)

학습 시 인식률 <표 5>와 비교하면 혼잡도에 비해 단어 오류율이 높지만 이는 대화체 음성의 특성과 한국어에서 어절 분할 방법을 사용했기 때문이다. 문장인식률과 아래의 단어 오류유형을 볼 때 인식결과로부터 의미추출 하기는 충분하다고 생각된다.

그분 상태 <b>어떠 시냐고요</b> 그분 상태 <b>어떠 시냐구요</b>
환자가 <b>숨 쉬기</b> 힘들고 의식이 <b>없다구요</b> 환자가 <b>숨쉬기</b> 힘들고 의식이 <b>없다고요</b>
잠시 침착 하시고 몇 동 몇 <b>호 예요</b> 잠시 침착 하시고 몇 동 몇 <b>호 예요</b>

응급출동상황 검출성능은 아래 <표 11>과 같다.

표 11. 응급출동 검출 성능

Table 11. Emergency response detection performance

응급	95.8% [ 46/48, 2 예측실패 ]
비응급	90 % [ 18/20, 2 예측실패 ]

응급출동 검출 성능은 의미추출 성능과 직결된다. 인식결과와 응급출동 템플릿과 비교분석해서 의미추출 추출하여 응급출동 여부를 판단하기 때문인데, 문장인식률(<표 10>)과 함께 고려했을 때 응급출동 검출 성능은 고무적인 결과라 할 수 있다. <표 11>에서 예측 실패는 시스템이 요구하는 환자정보 부족 또는 의미 정보추출 실패일 때 시스템에서 판단하지 않고 수보원의 판단에 맡기게 된다. 이는 실제 녹취 전사 기반이기 때문에 통화 내용에 시스템이 요구하는 환자정보가 부족한 경우임이다.

#### 5. 결론

본 연구에서는 대용량 음성 인식기를 개발하여 응급관제시스템에 적용함으로써, 수보 프로토콜과 신고 접수업무와의 연계성을 높여 수보업무의 효율성을 높이고자 시도하였다. 대용량 음성 인식기를 위한 발음모델은 40만 어휘 크기로 사전을 구성하였으며, 음향모델은 DNN-HMM 형식으로 개발하였다. 텍스트 맞춤치는 어절 분할 알고리즘 적용한 맞춤치를 이용하여 5-gram 언어모델을 개발하였다. 개념 검증(Proof of concept)으로 개발한 시스템으로서 성능실험 결과는 인식단어 오류율 9.14%와 응급출동 예측성능 95.8%를 보여서, 수보요원의 업무에 충분히 도움을 줄 수 있을 것으로 판단된다.

실제 더 많은 구급통화를 전사 분석하여 시스템을 안정화할 필요가 있고, 특히 의미 추출을 위한 추가 증상 표현 연구와 방법론 연구가 필요하다. 응급출동만이 아니라 31가지 다른 범주에 대해서도 응급출동 예측판단에 도움을 줄 수 있는 예측 모델을 연구할 계획이다.

#### 감사의 글

본 연구는 2015년도 국민안전처 소방안전 및 119구조·구급기술 연구개발사업(MPSS-소방안전-2015-71)의 연구비 지원을 받은 ‘한국형 응급상황관제시스템의 모색 및 개발 연구’ 과제의 일환으로 수행되었습니다.

#### 참고문헌

- [1] Alumäe, T. (2014). Full-duplex Speech-to-text system for Estonian. *Proceedings of Baltic HLT* (pp. 3-10).
- [2] Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707-710.

- [3] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., & Vesely, K. (2011). The Kaldi speech recognition toolkit. *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding*.
- [4] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., & Vesely, K. (2011). Online decoding in Kaldi. Retrieved from [http://kaldi-asr.org/doc/online\\_decoding.html](http://kaldi-asr.org/doc/online_decoding.html) on February 25, 2016.
- [5] Schuster, M. & Nakajima, K. (2012). Japanese and Korean voice search. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 5149-5152).
- [6] Stolcke, A., Zheng, J., Wang, W., & Abrash, V. (2011). SRILM at sixteen: Update and outlook. *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding*.
- [7] Zhang, X., Trmal, J., Povey, D. & Khudanpur, S. (2014). May. Improving deep neural network acoustic models using generalized maxout networks. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 215-219).
- [8] Chung, M., Lee, K., Shin, D., Chung, J., & Kang, K. (2016). Applying Speech Recognition Technology to Emergency Dispatch. *Proceedings of the Korean Institute of Fire Science and Engineering Conference* (pp. 77-78). (정민화·이규환·신대진·정지오·강경희 (2016). 음성인식기술의 응급상황관계 적용, *한국화재소방학회 춘계학술대회 논문집*, 77-78.)
- [9] Jang, Y., Kang, K., Kim, J., & Kim, K. (2016). A study of Korean symptom expression in emergency call. *Proceedings of the Korean Institute of Fire Science and Engineering Conference*, 75-76. (장윤희·강경희·김준태·김경혜 (2016). 구급 신고 전화에서의 한국어 증상 표현 연구, *한국화재소방학회 춘계학술대회 논문집*, 75-76.)
- [10] Kang, K. (2015). A comparison of information gathering and protocol in dispatching. *Proceedings of the Korean Institute of Fire Science and Engineering Conference* (pp. 119-120). (강경희 (2015). 구급 신고 접수와 수요요원의 도입부 수요 프로토콜의 비교, *한국화재소방학회 춘계학술대회 논문집*, 119-120.)
- [11] Kim, K., Lee, D., Lim, M., & Kim, J. (2015). Input Dimension Reduction based on Continuous Word Vector for Deep Neural Network Language Model. *Phonetics and Speech Sciences*, 7(4), 3-8. (김광호·이동현·임민규·김지환 (2015). Deep Neural Network 언어모델을 위한 Continuous Word Vector 기반의 입력 차원 감소. *발소리와 음성과학*, 7(4), 3-8.)
- **이규환 (Lee, Kyuwhan)**  
서울대학교 인문정보연구소  
서울시 관악구 관악로 1  
Tel: 02-880-9039  
Email: whannylee@snu.ac.kr  
관심분야: 음성인식, 음성언어처리, 컴퓨터 기반 언어교육
  - **정지오 (Chung, jio)**  
VSWORKS 대표  
Email: jiochung@vsworks.co.kr  
관심분야: 음성인식, 음성언어처리, 신호처리
  - **신대진 (Shin, Daejin)**  
Eidware 대표  
Email: djshin@eidware.com  
관심분야: 음성인식, 음성언어처리, 컴퓨터 기반 언어교육
  - **정민화 (Chung, Minhwa)** 교신저자  
서울대학교 언어학과  
서울시 관악구 관악로 1  
Tel: 02-880-9195 Fax: 02-882-2451  
Email: mchung@snu.ac.kr  
관심분야: 음성인식, 음성언어처리, 컴퓨터 기반 언어교육
  - **강경희 (Kang, Kyunghee)**  
가천대학교 응급구조학과  
인천광역시 연수구 함박피로 191  
Email: khkang@gachon.ac.kr  
관심분야: 응급의학
  - **장윤희 (Jang, Yoonhee)**  
인하대학교 한국어문학과  
인천광역시 남구 인하로 100  
Email: jyh@inha.ac.kr  
관심분야: 국어학
  - **장경호 (Jang, Kyungho)**  
인하대학교 사회교육과  
인천광역시 남구 인하로 100  
Email: kjang@inha.ac.kr  
관심분야: 통화금융, 계량경제학