Journal of Korean Institute of Intelligent Systems
Vol. 25, No. 6, December 2015, pp. 585-591
http://dx.doi.org/10.5391/JKIIS.2015.25.6.585

ISSN(Print) 1976-9172
ISSN(Online) 2288-2324

JKIIS

전사텍스트를 이용한 반자동 레이블링 구현

Implement of Semi-automatic Labeling Using Transcripts Text

원동진* · 장문수**† · 강선미***
Dong-Jin Won, Sun-Mee Kang, and Moon-soo Chang†

*서경대학교 전자컴퓨터학과, **컴퓨터과학과, ***전자공학과
*School of Electronic Computer, Seokyeong University
**†School of Computer science, Seokyeong University
***School of Electronic Engineering, Seokyeong University

요 약

구어 연구를 위한 전사 과정에서 문자로 표현된 발화를 녹음 음성에 연결해주는 작업을 레이블링이라고 한다. 기존 레이블링 도구들은 대부분 수동으로 작업이 이루어진다. 제안하는 반자동 레이블링은 자동화 모듈과 수동 조정 모듈로 구성된다. 자동화 모듈은 G.Saha 알고리즘을 활용하여 음성구간을 추출하고, 기구축된 발화텍스트의 발화 수와 발화의 길이 정보를 이용하여 발화구간을 예측한다. 본 논문에서는 기존 수동 도구의 정확성을 유지하기 위하여 자동 레이블링된 발화구간을 보정하기 위한 수동 조정 사용자 인터페이스를 제공한다. 제안하는 반자동 레이블링 알고리즘으로 구현한 도구는 기존 수동 레이블링 도구와 비교하여 작업 속도가 평균 27% 향상되었다.

키워드 : 전사, 레이블링, 발화, 구어, 사용자 인터페이스

Abstract

In transcription for spoken language research, labeling is a work linking text-represented utterance to recorded speech. Most existing labeling tools have been working manually. Semi-automatic labeling we are proposing consists of automation module and manual adjustment module. Automation module extracts voice boundaries utilizing G.Saha's algorithm, and predicts utterance boundaries using the number and length of utterance which established utterance text. For maintaining existing manual tool's accuracy, we provide manual adjustment user interface revising the auto-labeling utterance boundaries. The implemented tool of our semi-automatic algorithm speed up to 27% than existing manual labeling tools.

Key Words : Transcription, Labeling, Utterance, Spoken Language, User Interface

Received: Jun. 15, 2015
Revised : Aug. 24, 2015
Accepted: Aug. 24, 2015
†Corresponding author
cosmos@skuniv.ac.kr

1. 서 론

구어(spoken language)와 관련된 연구에서 녹음된 음성 데이터를 문자로 표현하는 작업을 전사(transcription)라고 한다. 그리고 넓은 의미의 전사에는 문자화가 완료된 데이터에 언어 분석을 위한 정보, 즉 음성, 음운, 형태소 등을 추가하는 작업이 포함된다. 전사 작업은 주로 컴퓨터를 이용하는데, 최근의 자료들이 멀티미디어화 되면서 음성이나 화상 데이터와 연동된 전사 자료가 만들어지고 있다. 이러한 자료를 만들기 위해서는 녹음된 음성구간과 텍스트를 동기화시키는 과정이 필요한데, 이 작업을 레이블링(labeling)이라고 한다. 레이블링된 전사 자료는 음성학이나 언어학적 분석뿐만 아니라 음성 자료를 이용하는 모든 연구에서 매우 유용하게 활용되고 있다.

최근에 구축된 외국의 멀티미디어 전사 자료들은 레이블링이 되어 있는 경우가 종종 발견되지만, 제작비용이 많이 들고 제작과정에서 전문 인력이 많이 필요하기 때문에 많은

본 논문은 한국연구재단 연구자지원사업(NRF-2011-32A-B00202)에서 지원하여 연구하였음.

전사 자료들이 음성 자료와 텍스트 자료가 별개로 구축되고 있는 실정이다. 또한 과거에 구축된 모든 전사 자료들은 레이블링을 위한 도구가 개발되어 있지 않았기 때문에 모두 레이블링 되지 않은 상태로 존재한다[1].

구어 자료 중에는 아동 구어 연구를 위한 자료도 상당히 많다[2]. 아동기는 언어 발달상에서 매우 중요한 시기이며 이 때 산출되는 언어 자료는 구어 연구에서 매우 중요한 자료이다. 외국의 경우에는 2010년대 초반부터 아동 언어 전사 데이터베이스에도 레이블링된 자료가 구축되고 있으나 전체 자료량에 비교하면 아직 미미한 수준이다. 그나마 새로 구축되는 자료 중에는 레이블링된 경우가 종종 있으나, 기구축된 자료는 여전히 음성 자료와 텍스트 자료가 따로 제공되고 있다.

최근에 연구 개발된 레이블링 도구는 음성인식을 이용하는 경우[3]도 있는데, 음성 인식 기술은 성인 데이터를 이용하여 개발된 것들이다. 아동 발화는 짧은 성도, 작은 성대, 혀의 움직임에 의해 성인보다 높은 기본 주파수와 포먼트 주파수를 가진다[4]. 따라서 음성인식을 이용하여 자동 레이블링을 구현하게 되면 아동 발화에 대한 성능은 제한적일 수밖에 없다. 또한 아동 발화는 성인 발화에 비해 발화 길이가 대체로 짧고, 문장의 완성도가 성인과 비교하여 떨어진다. 이러한 특징은 통계를 기반으로 하는 최근의 음성인식 기술에서 아동 음성에 대한 성능을 보장하기 어렵게 한다.

국내 아동 구어 연구에서 전사 작업의 대부분은 전용도구를 사용하지 않고 워드프로세서를 사용하여 이루어지고 있으며, 레이블링된 전사 자료는 공개된 적이 없다[5].

본 논문에서는 기구축된 텍스트 전사 자료를 레이블링하기 위해 자동과 수동 모듈로 구성된 반자동 레이블링 알고리즘을 제안한다. 그리고 아동 언어를 자동화 대상에 포함시키기 위하여 본 논문에서는 음성 인식 기술을 사용하지 않고, 기본적인 음성 신호 정보와 기구축된 전사 텍스트 정보를 이용하여 자동화 알고리즘을 구성한다.

그리고 자동화로 인한 오류를 보정하기 위해 수동 조정 인터페이스를 개발한다. 이 인터페이스는 오류 수정 기능뿐만 아니라 기존 수동 도구들의 단점을 보완하여 작업의 효율성을 향상시키는 기능을 제공한다.

본 논문의 구성은 2장에 기존의 레이블링 도구에 대한 설명과 문제점을 제시한다. 3장은 이를 개선하기 위한 반자동 레이블링의 자동 모듈을, 4장에는 수동 모듈에 대하여 기술한다. 그리고 5장에서는 시스템의 성능평가를 위한 실험과 결과를 검토하고 6장에서 결론을 맺는다.

2. 기존 연구

국외에서 수동 레이블링을 지원하는 도구로는 CLAN[6], Transcriber[7], Praat[8] 등이 있고, 자동 레이블링 도구로는 SPPAS[4]가 있다.

CLAN은 해외에 있는 아동 언어 데이터뱅크인 CHILDES에서 제공하는 도구로 명령어 기반의 도구이다. 레이블링을 위한 미디어 연결 기능은 명령어와 시간값을 같이 입력하여 설정한다. 명령어 기반의 작업은 모든 명령어를 외워야 하는 불편함이 있으며, 시각화가 되어 있지 않아 사용자가 현재 하는 작업의 결과를 알기 어렵게 한다.

Transcriber는 전사와 레이블링 작업을 동시에 할 수 있는 도구이다. 음성데이터를 텍스트로 옮기기 위해 타이핑을 하면서 단축키를 이용하여 레이블링한다. 이러한 작업 방식은 키보드만을 사용하여 사용자가 하나의 입력 도구에 집중할 수 있게 한다. 그러나 다양한 기능들의 단축키를 알아야 하며, 한번 입력된 구간은 직접적인 수정이 되지 않고 삭제 후에 재입력해야 한다.

Praat는 파형, 스팩토그램, 피치, 포먼트 등의 다양한 음성 정보를 제공한다. 이러한 정보들은 음성의 구간을 확인하는데 매우 유용하다. 인터페이스로는 키보드와 마우스의 사용을 모두 지원하고 입력 및 수정도 지원하여 사용자의 편의성도 제공한다. 그러나 기구축된 전사 자료의 레이블링을 위한 자동화 모듈이 따로 제공되지 않는다. 따라서 기존 전사자료를 레이블링할 때도 새로 전사를 시작할 때와 마찬가지로 텍스트를 입력해 주어야 한다.

SPPAS는 전사 텍스트와 음성 인식 기술을 이용하여 자동 레이블링하는 도구이다. 기본적으로 묵음구간을 제외한 음성 부분을 인식하고 전사 텍스트를 확인하여 레이블링하게 된다. 다중 언어를 지원하지만 해당 언어의 사전과 음성 모델 학습이 필요하다. 이는 많은 리소스를 요구하고, 아동 발화의 음성 인식에는 성능을 보장하지 않는 문제가 있다.

국내에서 개발한 전사 및 레이블링 시스템으로는 자동 음성분할 및 레이블링 시스템의 구현[9]과 메아리 1.0[10]이 있다. 두 도구는 연구 목적으로 개발되어 사용자 편의성을 고려하지 않고 기능 위주의 도구이다. 그리고 외부에 공개가 되지 않아 직접적으로 사용하기는 어렵다. CosmoScribe 2.0[11]은 한국어 구어 연구를 위해 개발된 전사 도구지만 레이블링을 지원하지 않는 문제점이 있다.

3. 발화구간 예측 알고리즘

제안하는 반자동 레이블링은 자동화 단계와 수동 조정 단계로 구성된다. 자동화 단계는 다시 음성구간 추출 알고리즘과 발화구간 예측 알고리즘으로 나뉜다.

여기서 음성구간(voice boundary)이란 일정 크기 이하의 소리를 나타내는 묵음(silence) 구간을 제외한 나머지 구간을 말한다. 그리고 발화(utterance) 구간은 하나의 발화가 시작해서 끝나는 구간을 말한다. 하나의 발화구간 안에는 여러 개의 음성 및 묵음 구간이 존재할 수 있다. 발화는 구어 연구에서 각기 다른 범위로 정의되는데, 가장 기본적인 발화의 의미는 발화자가 쉬지 않고 한 번에 말한 음성을 의미한다. 대화형식의 발화에서는 한 턴(turn) 즉, 발화자가 바뀌는 순간을 기준으로 발화를 나눈다.

자동화 단계에서는 먼저 발화구간이 아닌 음성구간을 추출한다. 발화는 음성의 집합으로 볼 수 있기 때문에 추출된 음성구간들은 전사텍스트를 이용해 합성 또는 분할로 발화구간 후보들을 구한다. 그리고 시간 비율을 이용한 점수 계산을 통하여 발화구간 후보들 중에서 최적값을 찾는다.

자동 레이블링의 최적값은 4장에서 설명할 수동 조정 인터페이스를 이용하여 미세 조정을 거쳐서 최종 레이블링 값으로 완성된다. 그림 2는 제안하는 레이블링 알고리즘의 흐름도를 나타내고 있다. 본 장에서는 자동화 단계인 음성구간 추출과 발화구간 예측 알고리즘에 대해서 설명한다.

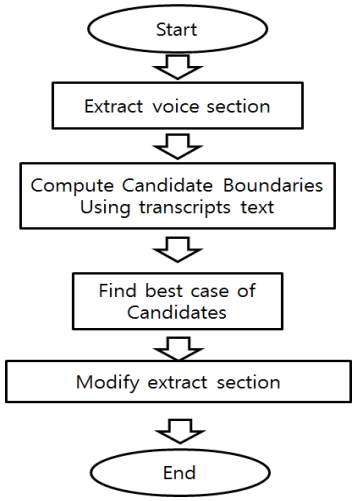

그림 1. 제안한 반자동 레이블링 흐름도

Fig. 1. Flow chart of the proposed Semi-aoutomatic Labeling

3.1 음성 구간 추출

자동 레이블링의 첫 단계는 음성구간 추출이다. 음성구간 추출을 위해 먼저 묵음이 아닌 음성을 검출해야 하는데, 본 논문에서는 음성검출 알고리즘으로 G.Saha의 알고리즘[12]을 사용한다. 이 알고리즘은 확률밀도함수를 이용하여 음성 신호를 검출한다.

음성 신호 검출 알고리즘들은 프레임별로 음성임을 판단하고 묵음 프레임은 검출하지 않고 음성 신호로 판단된 프레임만 검출하는 형식이다. 본 논문에서는 음성 프레임이 아닌 음성구간의 시작과 끝 시간을 구한다. 음성구간을 구할 때, 녹음 데이터의 중간부분만 계산할 때가 있다. 이를 위해, 시작 시간을 설정하여 계산에 이용한다. 녹음 데이터에서 검출된 음성 프레임들은 각각 계산된 시간값을 부여받는다. 음성 프레임은 연속성을 판단하여 첫 프레임의 시간값은 시작 시간, 끝 프레임의 시간값은 끝 시간으로 정해 음성구간을 추출한다.

3.2 전사텍스트를 이용한 발화구간 예측

자동 레이블링의 두 번째 단계는 추출된 음성구간과 기구축된 전사텍스트의 길이 정보를 이용하여 가능한 발화구간의 후보를 만들어 내고, 그 후보들의 점수를 계산하여 최적값을 가진 후보를 찾는다.

3.2.1 레이블링 후보 생성

본 논문에서는 음성구간들을 조합하여 발화구간을 예측하기 위하여 기구축된 발화 텍스트의 길이 정보를 이용한다. 발화텍스트는 전문가들에 의해 전사된 발화의 텍스트이다. 말하는 사람이나 상황에 따라 발화 음성의 길이가 다르지만, 하나의 발화 내에서 음절들의 길이의 변화는 크지 않다. 본 논문에서는 전체 텍스트에서의 발화 수와 한 발화 내에서의 음절의 수를 이용하여 음성구간들을 조합하여 발화구간을 예측하고자 한다.

이를 위하여 몇 가지 인자가 사용되는데, 이 인자값들로 인해 다수의 발화구간 후보가 나오게 된다. 후보 생성에 사용하는 인자는 다음과 같다.

- 최소 발화 시간
- 발화 내 최대 휴지 시간
- 발화 간 최소 휴지 시간
- 최소 구간 비(R)

음성 검출에서 묵음을 제외한 모든 소리를 음성으로 간주하기 때문에, 여기에는 음성뿐만 아니라 여러 가지 생활 잡음도 포함된다. 단답형의 짧은 발화도 일정 이상의 시간을 가지기 때문에 그 이하의 길이를 가지는 소리는 잡음으로 간주한다. 최소 발화 시간은 이 잡음을 제거하기 위한 인자이다.

발화 내 최대 휴지 시간은 발화를 하는 도중에 생기는 짧은 쉬는 구간(pause)의 시간을 정한 것이다. 발화는 쉬지 않고 말하는 음성의 집합이지만 중간에 길지 않게 말을 쉬는 경우가 있을 수 있다.

발화 간 최소 휴지 시간은 발화자 간의 발화나 한 발화자가 오래 쉬고 발화를 하는 턴 사이의 휴지 시간을 의미한다.

최소 구간 비(R)는 음성구간과 전사텍스트의 길이를 비교할 때 허용하는 최소의 비이다. 음성구간과 전사텍스트의 비는 일반적으로 정확히 일치하지 않기 때문에 허용하는 오차 범위이다.

전사텍스트의 발화 수는 고정되어 있고, 위의 인자들로 인해 각 발화들이 서로 다른 비율로 음성 구간을 구성할 수 있기 때문에 다수의 레이블링 후보가 생성된다. 레이블링 후보집합 C는 식 (1)과 같이 나타낼 수 있다.

$$C = \{C_1, C_2, \cdots, C_i, \cdots, C_s\} \tag{1}$$

C_i는 i번째 레이블링 후보를 나타내며, 이것은 전사텍스트와의 비율 계산에 의해 할당된 n개의 발화구간으로 구성되며, 식 (2)와 같이 나타낼 수 있다.

$$C_i = \{c_{i1}, c_{i2}, \cdots, c_{ik}, \cdots, c_{in}\} \tag{2}$$

전사텍스트와 음성구간을 이용하여 발화구간 후보를 계산할 때, 추출된 음성구간은 시간값이고 발화텍스트는 문자수이다. 둘의 기준을 맞추기 위하여 두 값을 비율로 변경한다. 전사텍스트에 n개의 발화가 존재한다면 p번째 발화텍스트의 비율은 식 (3)로 계산된다. $U(p)$는 p번째 발화텍스트의 음절 수를 의미한다.

$$u_p = \frac{U(p)}{\sum_{x=1}^{n} U(x)} \tag{3}$$

추출된 음성구간이 m개이면, q번째 음성구간의 비율은 식 (4)로 계산된다. $V(q)$는 q번째 음성구간의 길이를 의미한다.

$$v_q = \frac{V(q)}{\sum_{y=1}^{n} V(y)} \tag{4}$$

계산된 n개의 발화텍스트 비율과 m개의 음성구간 비율은 식 (5)와 식 (6)처럼 각각 발화텍스트 비율 집합 U와 음성구간 비율 집합 V로 나타낸다.

$$U = \{u_1, u_2, \cdots, u_n\} \tag{5}$$

$$V = \{v_1, v_2, \cdots, v_m\} \tag{6}$$

그림 3은 자동 레이블링 후보를 생성하는 중에 하나의 발화구간 c_{ia}를 추출하는 알고리즘의 흐름도이다. 발화텍스트의 발화 수 n개를 기준으로 순차적으로 계산하게 된다.

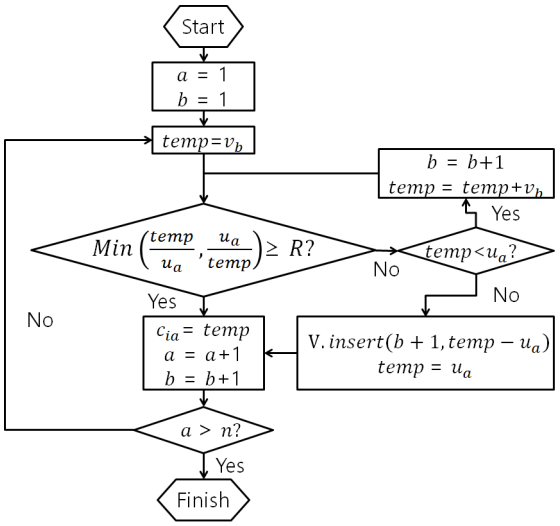

그림 2. 발화구간 후보 계산 흐름도

Fig. 2. Flow chart of the Extracting candidate of utterance boundary

발화텍스트 비율 u_a와 음성구간 비율 v_b의 비가 최소 구간비 R보다 크거나 같으면 두 값이 일치하는 것으로 판단하여 발화구간 c_{ia}에 계산된 $temp$값을 대입한다. 그러나 두 값의 비가 R보다 작으면, 두 가지 경우로 나뉜다. 계산된 $temp$값이 u_a보다 작은 경우는 음성구간의 합이 발화텍스트보다 짧은 상태로 판단하여 $temp$에 다음 음성구간인 v_{b+1}를 더한 다음 다시 비교를 시도한다. 반대로 계산된 $temp$값이 u_a보다 큰 경우는 너무 긴 음성구간이 하나로 잡힌 경우로 판단하여 발화텍스트의 비율 u_a만큼만 $temp$에 대입하고, 남은 음성구간 즉, $temp - u_a$는 새로운 음성구간으로 삽입하여 알고리즘을 진행한다.

3.2.2 최적의 후보 선택

앞 절에서 생성한 레이블링 후보군 C에 대해서 기구축된 발화텍스트와의 적합도를 계산하여 최적의 후보를 찾는다. 식 (7)은 적합도를 계산하는 식을 나타내고 있다. C_i의 모든 원소들과 발화텍스트의 비율을 비교하여 그 값을 모두 합한 다음, 총 발화 수 n으로 나눈 값을 적합도로 정의한다. 모든 레이블링 후보에 대해서 적합도를 계산하고 나서, 가장 큰 값을 가진 후보를 자동 레이블링의 결과로 산출한다.

$$Scoring(C_i, U) = \frac{\sum_{z=1}^{n} (c_{iz}/u_z)}{n} \tag{7}$$

그림 4는 음성구간을 추출한 상태와 발화구간을 예측하여 자동 레이블링이 끝난 상태를 비교하여 나타내고 있다. 그림 5는 이 구간에 대한 전사텍스트를 나타낸 것이다. 여러 개로 나뉜 음성구간이 하나의 발화구간으로 묶여진 것을 볼 수 있다.

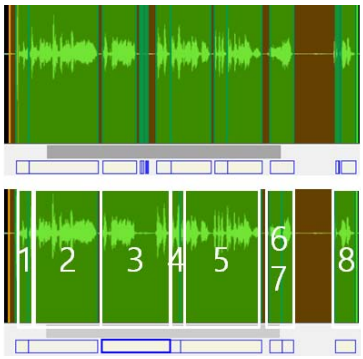

그림 3. 음성구간 추출(상단)과 발화구간 예측 후(하단)의 모습

Fig. 3. Extracting voice boundaries(top) and result of predicting utterance boundaries(bottom)

	Speaker	Utterance
1	CHI	깡총.
2	MOT	깡총 하는 거야 그렇구나.
3	MOT	선호가 깡총 그러는 거야?
4	CHI	어.
5	MOT	토끼가 깡총하는 거야?
6	CHI	어.
7	MOT	아.
8	MOT	먹지대?

그림 4. 그림 4에 사용된 전사텍스트

Fig. 4. Transcription text for Fig. 4

4. 수동 조정 사용자 인터페이스

3장의 알고리즘에 의해 예측된 발화구간은 여러 가지 요인으로 인해 발화와의 오차가 발생할 수 있다.

첫째, 잡음으로 인하여 음성이 아닌 구간이 검출되는 경우다. 잡음이 음성으로 잡힐 경우 발화구간을 예측할 때 잡음도 발화의 일부로 인식하여 잘못된 발화 검출이 될 가능성이 높다. 둘째, 연속된 대화에서 음성구간에 대한 판단이 잘못되는 경우가 있다. 여러 발화가 하나의 음성구간으로 검출될 수도 있다. 이 경우에는 전사텍스트로 예측하더라도 정확도가 떨어지게 된다. 셋째, 발화 내에서 발화 속도가 현저하게 변화하는 경우가 있다.

이렇게 발생한 오차는 사용자가 수동으로 수정할 필요가 있다. 이 작업은 기존 수동 도구를 사용하여 레이블링을 하는 것과 마찬가지로 높은 집중력과 많은 시간이 요구될 수 있다. 본 논문에서는 작업자의 피로를 줄이고, 수동 조정 작업을 편리하게 할 수 있는 사용자 인터페이스를 개발한다.

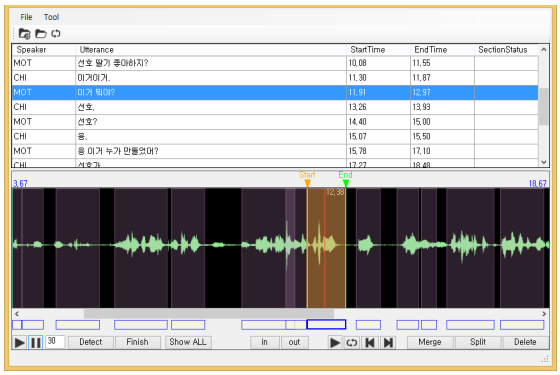

그림 5. 수동 조정 사용자 인터페이스

Fig. 5. Manual adjustment User Interface

수동 조정 사용자 인터페이스는 작업자에게 익숙한 윈도우 환경을 통해 제공된다. 개발된 사용자 인터페이스는 음성의 파형, 재생 위치, 선택된 발화구간 등을 보여준다. 음성을 파형으로 제공하게 되면 사용자는 음성을 눈으로 확인할 수 있어서 구간을 조정하는데 도움이 되고 피로도를 줄일 수 있다[13]. 레이블링된 발화구간들의 표시를 하단에 사각형 형태로 제공하고, 파형에 투명한 배경으로도 제공하여 사용자가 구간을 확인할 수 있게 한다.

제안하는 수동 조정 사용자 인터페이스는 마우스의 이용에 중점을 둔다. 마우스의 사용은 사용자의 행동을 그대로 반영[14]하여 키보드의 단축키보다 직관성이 뛰어나고 사용자가 인터페이스를 숙지할 때 유용하다[15].

5. 실험 및 평가

본 논문의 자동 레이블링 알고리즘은 기존 수동 도구로 이루어지고 있는 전사 환경을 개선하기 위해 제안하였다. 제안한 알고리즘의 성능을 평가하기 위해 대표적인 기존 수동 도구들과 비교 실험을 한다. 비교 대상은 제안하는 알고리즘과 같이 발화 단위로 레이블링을 하는 Transcriber와 음성 관련 분야에서 가장 보편적으로 사용하고 있는 Praat를 선정한다.

실험은 제안하는 알고리즘으로 구현한 도구를 포함한 세 가지 레이블링 도구를 실험자들이 사용하여, 그 결과에 대해 작업 속도를 비교하고, 제안하는 레이블링 방식에 대한 만족도를 조사하는 방식으로 진행한다. 그리고 실험자를 전문가와 비전문가 그룹으로 나누어 실험하여 각 그룹에 대한 효과를 평가하고자 한다.

실험에 사용할 데이터는 여러 가지 상황을 고려하기 위하여 성인과 아동이 모두 등장하는 장면의 대화형 음성 파일을 선정하고, 주변 잡음이 많은 것과 없는 것을 모두 포함하도록 한다.

실험의 참가인원은 11명이고, 음성 데이터는 7개 파일에서 실험을 위해 20발화씩 발췌하여 사용한다.

5.1 작업 속도

속도 측정을 위한 실험값은 전사가 완료된 20발화의 음성을 듣고 레이블링을 완료하는데 걸리는 시간을 사용한다. 표 1은 각 음성 파일에 대한 작업시간을 나타낸다.

표 1. 각 파일의 20발화 작업 속도(초)

Table. 1. Every file's work speed of 20 utterances(sec)

File	Group	Transcriber	Praat	Proposal
1	A	375	277	260
	B	382	424	307
2	A	362	314	282
	B	486	428	343
3	A	424	357	361
	B	617	500	366
4	A	365	337	361
	B	511	446	372
5	A	307	314	248
	B	428	312	287
6	A	326	293	296
	B	434	432	314
7	A	350	293	315
	B	468	312	342
Avg. (sec)	A	363	312	303
	B	509	405	334
	Total	456	371	322
Perfor-mance	A	0.83	0.97	1
	B	0.65	0.82	1
	Total	0.70	0.86	1

작업 속도 평가는 수동 레이블링에 대한 제안하는 레이블링의 작업 효율성에 관한 평가이다. 표 1의 결과는 전문가 그룹(그룹 A)과 비전문가 그룹(그룹 B)에 따라 효율성이 다르게 나타나는 것을 보여준다. 파일 3,4,6번은 발화가 겹쳐서 나타나거나 제 3자의 음성이 섞여서 나타나서 전문가의 정밀한 판단이

필요한 데이터이다. Praat의 사용에 익숙한 그룹 A는 이들 파일에서 제안하는 도구보다 Praat의 작업속도가 빨랐다. 그러나 두 도구 모두 숙련되지 않은 그룹 B는 제안하는 도구 쪽이 더 적은 시간이 걸린 것으로 나타났다.

전체적인 평균속도는 두 그룹 모두 제안하는 도구의 속도가 빠른 것으로 나타나 작업 효율성이 향상된 것으로 판단된다. 예외적으로 7번 파일은 두 그룹 모두 Praat가 빠른 것으로 나타났다. 이것은 연속되는 발화가 많을 경우 자동 레이블링에 오류가 증가하는 것으로, 향후 알고리즘을 개선할 필요가 있다. 그리고 Transcriber의 경우에는 처음 예상과는 달리, 수정 기능의 불편함으로 인하여 거의 모든 상황에서 가장 느린 결과를 나타내었다.

5.2 반자동 레이블링의 만족도

수동 방식이나 반자동 방식의 레이블링은 사용자에 의해 최종적으로 결과물이 생성되기 때문에 이에 대한 정확도 평가는 무의미하다. 본 논문에서는 작업 과정과 결과물에 대한 사용자 만족도 설문 조사 방식으로 평가를 진행하고자 한다.

표 2. 만족도 설문

Table. 2. Satisfaction Questionnaire

	Contents	Group A	Group B
1	Uses the semi-automatic labeling speed up the work more than manual labeling?	3.75	4.14
2	Uses the semi-automatic labeling make the work more accurate?	3.00	3.85
3	Is the semi-automatic labeling steps easy to learn and use?	4.00	4.28
4	Is the manual adjustment user interface easy and convenient to use?	3.75	4.00
5	Does the Auto-Labeling help to the work that labeling utterance boundaries?	2.50	3.42
	Total	3.40	4.03
		3.71	

조사 항목은 속도 향상 정도, 정확도 향상 정도, 학습의 용이성, 수동 조정 도구의 편의성, 자동 모듈의 유용성 등 다섯 가지 항목으로 구성한다. 표 2는 실험자에게 5점 만족도 설문으로 평가한 것을 전문가(A)와 비전문가(B) 그룹으로 나누어 정리한 것이다.

결과물에 대한 질문 중 1번(작업 속도 향상)의 경우 그룹 A와 그룹 B 각각 3.75와 4.14의 점수로 대체적으로 만족하였다. 2번(정확도 향상)에서 그룹 A는 3.00으로 만족도가 보통인데, 음성 파형 외의 부가적인 정보가 제공되는 Praat에 익숙한 전문가에게는 다소 불편한 것으로 나타났다.

작업 방식에 관련된 3번(레이블링 방식 학습과 사용의 편리성)과 4번(수동 조정 인터페이스 편의성)의 경우, 그룹 A와 그룹 B가 모두 만족하는 경향으로 나타났다.

자동화 알고리즘에 대한 만족도(5번)는 그룹 A에서 다소 떨어지는 것으로 나타났다. 점수가 2.50밖에 안되는데, 사용자 코멘트에서 세밀한 구간의 검출이 다소 아쉬웠다고 나왔다. 이것은 반자동 레이블링에서 수동 조정 과정이 필요한 것을 의미한다.

6. 결론 및 향후 연구

구어 연구에서는 음성을 텍스트로 변환하는 전사 작업이 필요하다. 본 논문에서는 전사 자료를 위한 반자동 레이블링 방법을 제안하고 그 도구를 개발하였다. 제안한 레이블링은 자동 발화구간 검출 과정과 검출된 발화 구간을 발화텍스트에 맞춰 미세 조정을 하는 수동 레이블링 과정으로 구성된다.

기존 레이블링 도구들은 새로 구축되는 전사 자료에 초점을 맞춰 개발되었기 때문에 기구축된 전사텍스트를 음성 자료에 레이블링하여 멀티미디어 전사 자료로 재구축하는 데에는 적합하지 않았다. 제안한 도구는 기구축된 전사텍스트를 이용하여 음성 데이터에서 자동으로 발화 구간을 검출하고, 이것을 수동 레이블링 도구를 이용하여 조정함으로써, 기존 전사 자료를 멀티미디어 전사 자료로 빠르게 재구축할 수 있다.

또한, 기존 도구와의 비교 실험에서, 제안한 알고리즘으로 구현된 레이블링 도구는 기존의 수동 레이블링 도구와 비교하여 작업 속도 면에서 평균 27%정도 향상되었다.

자동 발화구간 예측에서 발화구간 검출 기능이 잡음이나 기타 소음이 많은 음성 파일의 경우 정확하지 않아 스펙트로그램이나 피치 같은 추가적인 음성 정보를 이용하여 개선할 계획이다.

References

[1] TalkBank, "TalkBank Transcript Browser," Available: http://talkbank.org, [Accessed: Feb 2, 2015].

[2] CHILDES, "CHILDES Transcript Browser," Available: http://childes.psy.cmu.edu/browser, [Accessed: Feb 2, 2015].

[3] Bigi, Brigitte, "SPPAS: a tool for the phonetic segmentations of Speech," *The eighth international conference on Language Resources and Evaluation*, vol. 8, pp. 1748-1755, 2012.

[4] Sharmistha S. Gray, et al., "Child Automatic Speech Recognition for US English: Child Interaction with Living-Room-Electronic-Device s," *WOCCI 2014*, poster session, 2014.

[5] Jiyoung Shin et al., "Developing a Korean Standard Speech DB," *Journal of the Korean society of speech sciences*, vol. 7, no. 1, pp. 139-150, 2015.
[6] CHILDES, "Using CLAN," Available: http://child-es.psy.cmu.edu/clan/, [Accessed: Feb 2, 2015].
[7] Claude Barras, et al., "Transcriber: a free tool for segmenting, labeling and transcribing speech." *First international conference on language resources and evaluation (LREC)*. pp. 1373-1376, 1998.
[8] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer," Available: http://www.praat. org, 2009, [Accessed: Feb 2, 2015].
[9] Jongmo Sung and Hyung Soon Kim, "Implemen- tation of the Automatic Speech Segmentation and Labeling System," *The Journal of The Acoustical Society of Korea*, vol. 16, no. 5, pp. 50-59, 1997.
[10] Kang-Chun So, "A Study on the Method of Computational Processing of Dialectal Sound Data," *The Society of Korean Language and Literature*, vol. 142, pp. 7-30, 2006.
[11] Sun-dong Kwak and Moon-soo Chang, "CosmoScriBe 2.0 : The development of Korean transcription tools," *Journal of Korean Institute of Intelligent Systems*, vol. 24, no. 3, pp. 323-329, 2014.
[12] G. Saha, Sandipan Chakroborty, and Suman Senapati, "A new silence removal and endpoint detection algo-rithm for speech and speaker recognition applications," *Proceedings of the 11th National Conference on Communications (NCC),* pp. 291-295, 2005.
[13] Dong-jin Won and Moon-soo Chang, "An Improvement of Audio controller in Transcription Tool," *Proceedings of KIIS Spring Conference*, Vol. 22, no. 2, pp. 121-122, 2012.
[14] Donald A. Norman, *The Design of Everyday Things* , Basic Books, 2002.
[15] Tekla S. Perry, and John Voelcker, "Of mice and me-nus: designing the user-friendly interface," *IEEE Spectrum*, vol. 26, no. 9, pp. 46-51, 1989.

저 자 소 개

원동진(Dong-Jin Won)
2011년 : 서경대학교 컴퓨터과학과 졸업
2011년~현재 : 서경대학교 대학원 전자컴퓨터학과 석사과정

관심분야 : HCI, Image Processing, Signal Processing
Phone : +82-10-7499-0938
E-mail : wondongjin87@gmail.com

장문수(Moon-Soo Chang)
1992년 : 고려대학교 전자전산공학과 공학사
1994년 : 고려대학교 전자공학과 공학석사
2001년 : 동경공업대학 지능시스템과학전공 공학박사
2000년~2003년 : 한국전자통신연구원 선임연구원
2003년~현재 : 서경대학교 컴퓨터과학과 부교수

관심분야 : Natural Language Understanding, Knowledge Minig, HCI
Phone : +82-940-7754
E-mail : cosmos@skuniv.ac.kr

강선미(Sun-Mee Kang)
1981년 : 고려대학교 전자공학과 공학사
1988년 : 독일 Erlangen-Nuernberg 전기전자공학과 Diplom
1992년: 고려대학교 전자공학과 공학박사
1994년~1997년 : 고려대학교 산업대학원 객원조교수
1997년~현재 : 서경대학교 전자공학과 부교수

관심분야 : Digital Signal Processing, Speech Recognition, Pattern Recognition
Phone : +82-940-7737
E-mail : smkang@skuniv.ac.kr