말소리와 음성과학 제6권 제2호 (2014.06.30)
ISSN 2005-8063, pp. 9~20
www.speechsciences.or.kr
http://dx.doi.org/10.13064/KSSS.2014.6.2.009

한국인을 위한 영어 말하기 시험의 컴퓨터 기반 유창성 평가

Computer-Based Fluency Evaluation of English Speaking Tests for Koreans

장 병 용[1] · 권 오 욱[2]
Jang, Byeong-Yong · Kwon, Oh-Wook

ABSTRACT

In this paper, we propose an automatic fluency evaluation algorithm for English speaking tests. In the proposed algorithm, acoustic features are extracted from an input spoken utterance and then fluency score is computed by using support vector regression (SVR). We estimate the parameters of feature modeling and SVR using the speech signals and the corresponding scores by human raters. From the correlation analysis results, it is shown that speech rate, articulation rate, and mean length of runs are best for fluency evaluation. Experimental results show that the correlation between the human score and the SVR score is 0.87 for 3 speaking tests, which suggests the possibility of the proposed algorithm as a secondary fluency evaluation tool.

Keywords: speaking fluency, pronunciation, SVR, regression

1. 서 론

말하기의 유창성을 판단하기 위해서는 많은 지식과 데이터가 필요하다. 특히 외국어의 유창성 판단에는 여러 복합적인 판단 기준이 필요하다. 이러한 어려움에도 불구하고, 외국어 말하기의 유창성 판단에 관한 시험은 늘어나고 있는 추세이고, 이는 외국어 능력의 등급 기준으로 종종 사용되곤 한다. 현재 이러한 유창성 판단은 해당 외국어의 원어민 또는 비원어민 외국어 교사가 채점을 하고 등급을 매기고 있다. 사람이 직접 채점을 하면, 정확성이나 타당성에서 많은 장점이 있지만, 시간과 비용 면에서 많은 제약이 따르기 마련이다. 또한 이러한 평가는 어느 정도 평가 기준이 제시되어 있지만, 평가자의 주관적인 성향이나 판단 기준의 따라 채점 점수의 차이가 존재한다. 그러므로 유창성 평가 점수에 대한 객관성을 증명하는 것은 매우 어려운 일이다. 본 논문에서는 유창성 평가에서 나타나는 이러한 문제들을 개선하기 위하여 자동 유창성 평가 도구를 개발하고자 한다. 물론 사람이 평가하는 정도의 성능을 가진 도구를 개발하기는 아직 어렵지만, 음성학적 및 언어학적으로 타당한 기준과 평가를 가지고, 자동으로 유창성을 개략적으로 판단하고 등급을 분류할 수 있다면, 사람이 평가할 때 소요되는 시간과 비용을 절약할 수 있고, 평가 결과의 객관성 향상에 기여할 수 있을 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 음성학 분야에서의 유창성 관련 연구 내용을 소개하고, 3장에서는 유창성 평가를 위한 알고리듬을 제안하고, 4장에서는 실험 방법 및 결과를 보여주고, 5장에서 결론을 맺는다.

2. 배경 이론

자동 유창성 평가 도구를 개발하기 위해서는 먼저 유창성에 관한 음성학 분야의 배경 지식이 필요하다. 유창성을 판단하기 위한 음성학 분야의 연구는 예전부터 활발하였다. Fillmore는 유창성의 구성요소 4가지를 다음과 같이 정의하였다[1].

1) 적은 묵음(pause)으로 구체적으로 말할 수 있는 능력
2) 응집력 있고 논리적으로 말하는 능력
3) 다양한 문맥과 상황에서 적절히 말할 수 있는 능력

1) 충북대학교, byjang@cbnu.ac.kr
2) 충북대학교, owkwon@cbnu.ac.kr, 교신저자
이 논문은 한국연구재단의 지원금으로 수행되었습니다. (지원번호: 2012R1A1A2042381).

접수일자: 2014년 3월 11일
수정일자: 2014년 4월 30일
게재결정: 2014년 5월 13일

4) 독창적인 생각을 말할 수 있는 능력

Chambers는 유창성의 양적, 질적 연구를 바탕으로 유창성이란 단어의 정의를 확립하고, 외국어 말하기 평가의 지침을 제시하였다[2]. 여기서 '유창한 말하기'란 언어사용이 부드럽고, 빠르며, 자연스러움을 뜻한다고 설명하고 있다. 유창성을 판단하는데 중요한 요소는 발화 속도(rate of speech), 묵음의 빈도와 위치, 더듬음(hesitation)이고, 이 요소들은 시간적으로 정량화가 가능하다고 설명하고 있다. 이러한 연구를 바탕으로 Kormos는 10가지의 시간적 특징과 3가지의 어휘적 특징이 유창성 판단에 미치는 영향에 대하여 실험하였다[3]. Kormos가 실험한 10가지 시간적 특징(temporal variables)은 다음과 같다.

- 발화 속도(speech rate)
- 조음 속도(articulation rate)
- 발화 시간 비율(phonation-time ratio)
- 연속 발화 평균 길이(mean length of runs)
- 분당 묵음 빈도(number of silent pauses per minute)
- 묵음 평균 길이(mean length of pauses)
- 분당 비유창성 빈도(number of disfluencies per minute)
- 강세 단어 빈도(number of stressed words)
- 강세 단어 비율(ratio of words and stressed words)

시간적 특징은 대부분 발화 속도와 묵음, 그리고 강세와 관련이 많다. 여기서 비유창성(disfluency)은 더듬거나 반복적인 발화를 나타낸다.

또한, 3가지 어휘적 특징은 다음과 같다.

- 총 단어 수(total number of words)
- D-value
- 정확성(accuracy)

여기서 D-value는 Malvern이 제안한 어휘력을 측정하는 방법[4]이라고 설명하고 있다. 이 특징들 중 Kormos는 발화 속도, 연속 발화 평균 길이, 강세 단어 빈도, 정확성이 유창성과 관련이 깊은 특징이라고 설명하고 있다[3]. 하지만 화자별 분석 결과로부터 이러한 특징들과 채점자의 점수는 단순히 비례하지 않는다고 보여주었다. 이는 유창성 평가에는 여러 특징의 복합적인 결합이 필요하다는 것을 의미한다.

유창성 판단에 중요한 또 하나의 특징은 발음 특징이다. 비원어민의 외국어 발화뿐만 아니라 원어민의 발화 또한 발음에 의하여 유창한 정도가 다르게 보이는 것이 사실이다. 같은 내용의 발화임에도 아나운서의 발화가 일반인의 발화보다 더욱 유창해 보이는 것을 간단한 예로 들 수 있다. Neumeyer [5]은 발음평가를 위하여 다음과 같은 6가지 특징을 추출하고 각 특징의 성능을 평가하였다.

- 전체 평균 로그 우도(global log likelihood)
- 국소 평균 로그 우도(local log likelihood)
- 로그 사후 확률(log posterior score)
- 음소 인식 정확도(phone recognition accuracy)
- 음소 지속 시간 점수(segment duration score)
- 음절 시간(syllabic timing)

여기서 음절 시간은 발화 속도와 동일한 특징이다. Neumeyer은 상관분석을 통해 이러한 특징들 중 로그 사후 확률, 음소 지속 시간 확률이 발음을 평가하는데 밀접한 관련이 있다고 설명하고 있다[5].

지금까지 보여준 음성학적, 공학적 연구를 통해 유창성을 판단할 수 있는 중요한 특징들을 알 수 있다. 하지만 이러한 특징들을 자동으로 추출하고, 결합하여 하나의 유창성 점수를 계산하는 연구는 쉽게 찾을 수 없다. 본 논문에서는 음성 인식기술을 기반으로 하여 유창성 평가에 기여가 높은 특징을 추출하고, 이 특징들을 효과적으로 결합하여 하나의 유창성 점수를 자동으로 계산해주는 유창성 평가 알고리듬을 제안하고, 성능 평가를 실시하고자 한다.

3. 제안 알고리듬

3.1. 전체 구조

본 논문에서는 유창성을 평가하기 위하여 음소 정렬기와 특징 추출, support vector regression (SVR)을 이용한 회귀 분석을 사용한다. 또한 이 알고리듬을 작동하기 위하여 음향 모델, 원어민 음소 특징 모델, SVR 학습 모델이 필요하다. <그림 1>은 이러한 유창성 평가를 위한 학습 구조도와 평가 구조도이다. 먼저, 유창성 평가 모델 학습에서는 원어민 음성 데이터베이스(DB)와 공개 음성인식 도구인 hidden Markov model toolkit (HTK)를 이용하여 음소 정렬기에 사용되는 음향 모델을 생성하고, 학습기에서 생성된 원어민 음성 데이터베이스의 음소 정렬 결과를 이용하여 특징 추출에 사용되는 음소 지속 시간 및 로그 우도 확률 모델을 생성한다. 이 후 한국인의 말하기 시험 데이터베이스의 학습 그룹의 발화 음성과 전사 텍스트는 학습기에서 생성한 음향 모델과 음소 정렬기를 통하여 음소열 및 음소 시간이 정렬되고, 이 결과와 원어민의 음소 모델을 이용하여 특징을 추출한다. 추출한 특징 벡터와 평가자의 점수는 SVR 학습기를 통해 SVR 모델을 생성한다. 다음 유창성 평가에서 한국인의 말하기 시험 데이터베이스의 검증 그룹의 발화 음성과 전사 텍스트가 입력되면, 유창성 평가 모델 학습에서 생성된 모델을 이용하여 음소 정렬기와 특징 추출, SVR을 거쳐 유창도 점수가 산출된다. 본 논문에서는 SVR을 이용하여 산출된 유창도 점수와 평가자의 점수의 상관 분석을 통하여 본 논문에서 제안하는 알고리듬의 성능을 확인하였다. <표 1>은 예비실험 결과로부터 선택한 11개의 특징을 요약하여 보여준다.

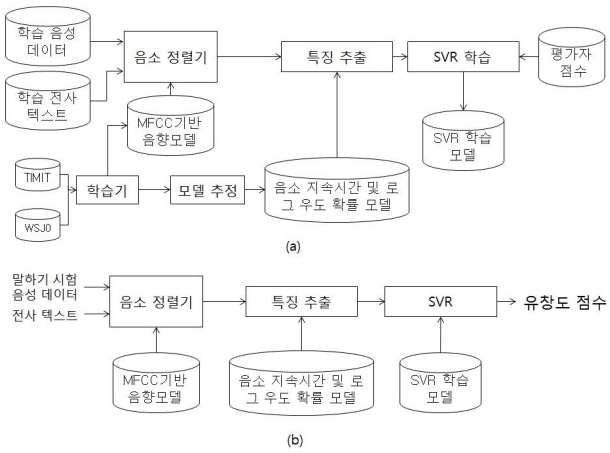


그림 1. 유창성 평가 구조도
(a) 모델 학습 (b) 유창성 평가
Figure 1. Architecture of fluency evaluation
(a) Model learning (b) Fluency evaluation

표 1. 추출한 특징 목록
Table 1. List of extracted features

순번	약어	명칭	설명
1	SR	발화 속도	발화내 시간당 음절 개수
2	AR	조음 속도	묵음 제외한 발화내 시간당 음절 개수
3	PR	발화 시간 비율	총 발화 시간 중 실제 발화한 시간 비율
4	LR	연속 발화 평균 길이	0.25초 이상 묵음 사이의 평균 음절 개수
5	numUP	빈 묵음 빈도	시간당 빈 묵음 개수
6	lenUP	빈 묵음 평균 길이	0.2초 이상인 빈 묵음의 평균 길이
7	LLH	국소 평균 로그 우도	음소별 로그 우도
8	GLH	전체 평균 로그 우도	모든 음소에 대한 평균 로그 우도
9	LPS	사후 음소 확률	음소별 사후 확률
10	PDS	음소 지속 시간 점수	음소 지속시간의 모델링 점수
11	PLS	음소 로그 우도 점수	음소 로그우도의 모델링 점수

3.2. 음소 정렬기(phoneme aligner)

앞 절에서 유창성 평가를 위한 특징 추출 이전 음소 정렬기를 이용하여 음소열 및 음소 시간을 정렬한다고 설명하였다. 이는 음성 인식 기술을 기반한 유창성 평가 특징 추출에서 음성 데이터베이스와 전사 텍스트(transcribed text)를 이용하여 음소열 및 음소 시간을 정렬(align) [6]함으로써 강력한 음성 인식기를 대체하는 방법이다. 이러한 방법이 사용된 것은 다음과 같은 이유 때문이다. 본 연구에서 사용한 Wall Street Journal (WSJ) Corpus [7]와 HTK [6]를 이용한 인식기의 영어 단어 인식 성능은 대략 92% 이지만 이는 원어민 발화에 관한 성능이고, 한국인의 영어 발화에 대한 인식률은 떨어질 수밖에 없다. 한국인의 영어 발화에 대해 강력한 인식기를 구현하기 위해서는 한국인의 영어 발화에 관한 대용량 음성 데이터베이스와 언어 모델링 기술 등이 필요하다. 따라서 본 논문에서는 강력한 음성 인식기가 존재한다고 가정하고, 전사 텍스트를 정렬함으로써 특징 추출 작업을 하였다.

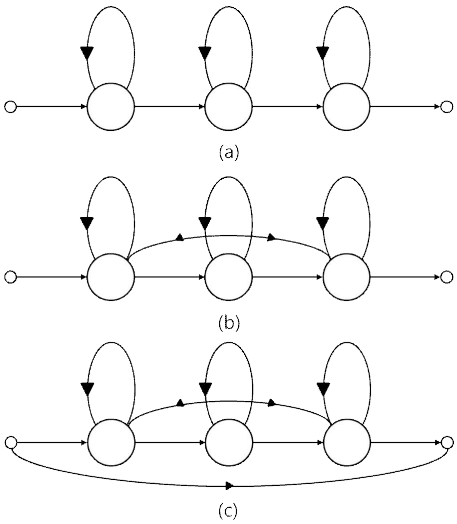


그림 2. HMM 음향 모델 구조
(a) 음소 (b) 긴 묵음 (c) 짧은 묵음
Figure 2. HMM acoustic model topology
(a) phone (b) silence (c) short pause

본 논문에서 사용한 음소 정렬기의 음향 모델은 기본적으로 HTK Recipe [8]를 이용하여 생성하였다. 음향 모델을 생성하기 위한 학습기의 데이터베이스는 WSJ0 [7]과 TIMIT [9]를 사용하였고, 음향 모델은 mel frequency cepstral coefficient (MFCC) 39차의 특징벡터와 트라이폰(triphone) 단위의 3개의 상태를 갖는 좌우 구조(left-to-right topology)의 HMM으로 구성하였으며 <그림 2>와 같다. 음소의 혼합 수는 8개이며, 묵음의 혼합 수는 16개를 사용하였다. 묵음 모델은 긴 묵음(silence)과 짧은 묵음(short pause)으로 확장하여 생성하였으며, 이는 사전(dictionary)의 단어별 발음열 마지막에 긴 묵음과 짧은 묵음을 각각 발음열로 두어 학습하였다. 즉 <그림 3>와 같이 한 단어당 발음이 최소 두 개가 존재하는 모양이 되고, 사전은 cmudict [10]를 이용하였다. 사전 단어 수는 약 13만 단어이고, 단어당 발음이 두 가지이므로 총 사전 크기는 약 26만 단어가 된다.

```
A42128      ey f ao r t uw w ah n t uw ey t sil
A42128      ey f ao r t uw w ah n t uw ey t sp
AAA         t r ih p ah l ey sil
AAA         t r ih p ah l ey sp
AABERG      aa b er g sil
AABERG      aa b er g sp
AACHEN      aa k ah n sil
AACHEN      aa k ah n sp
AAKER       aa k er sil
AAKER       aa k er sp
...
```

그림 3. 긴 묵음과 짧은 묵음을 적용한 발음 사전 일부
Figure 3. Part of pronunciation dictionary using silence and short pause

음소 정렬기는 학습기에서 사용한 음향 모델과 사전 및 혼합 수를 사용하지만, 한국인의 영어 발화에서 나타나는 사전 외(out of vocabulary) 단어가 나타날 수 있다. 더듬거림이나 반복 등이 영어가 아닌 단어로 발성될 수 있기 때문이다. 이를 처리하기 위하여 본 논문에서는 배경(background) 음향 모델을 추가하였다. 이는 이전에 학습하여 생성한 음향 모델과는 별도로 학습하였으며, 음성(speech)과 묵음(silence)만을 구분하여 음성 모델(speech model)을 생성하였다. 음성 모델의 HMM은 음소 모델과 동일한 구조를 갖도록 하였다. 음성 모델을 학습하기 위하여 모든 음소를 'SPEECH'로 변경하고, 사전에 'SPEECH'와 'silence'만을 포함시킨 후 학습을 하여 'SPEECH'에 대한 모델을 생성한다. 학습된 'SPEECH' 모델을 이전에 생성한 음향 모델에 추가하여 최종적인 음향 모델을 완성하였다. 이 음성 모델은 <그림 4>와 같이 말하기 시험의 전사 텍스트에 존재하는 사전 외 단어를 대체하여 사용된다. 음성 모델 적용 전 전사 텍스트 네 번째 줄에 'smartphone'은 사전에 없는 단어로 음성 모델 적용 후 '+SPEECH+'로 대체된다. 하지만 두 번째 줄에 있는 'na'의 경우 'nature'의 더듬음이지만 사전에 'na'의 발음이 존재하면 음성 모델로 대체하지 않는다. 채워진 묵음 'uh'도 'na'와 마찬가지로 사전에 존재하는 발음으로 음소 정렬을 하게 된다.

```
81  690 Yes, I like taking pictures.
129 979 uh I like the photo with  +<H>na+ nature and,
897 1500  some great people in history. uh So,
159 848 I have compact digital camera and my smartphone.
786 1464  In my house, I have DSLR recently.
125 1024  Because they have uh memory by taking pictures,
1020  1465  they have more, they'll have more,
```

(a)

```
81  690 Yes I like taking pictures
129 979 uh I like the photo with na nature and
897 1500  some great people in history uh So
159 848 I have compact digital camera and my +SPEECH+
786 1464  In my house I have DSLR recently
125 1024  Because they have uh memory by taking pictures
1020  1465  they have more they'll have more
```

(b)

그림 4. (a) 원래 전사 텍스트 (b) 배경 음향모델을 사용한 텍스트
Figure 4. (a) Original transcribed text, (b) text using background acoustic model

3.3. 유창성 평가 특징

2절에서 제시된 유창성 관련 특징을 추출하기 위해서는 음절의 개념이 포함된다. 한국어의 경우 보통 한 글자가 한 음절인 경우가 대부분이지만 영어의 경우 특정 단어의 음절의 수를 판단하는 것은 어려운 일이다. 그렇기 때문에 컴퓨터로 음절을 판단할 수 있는 기준이 필요하다. 본 논문에서는 하나의 음절은 모음 1개만을 포함한다고 가정하고, 발화내의 모음 개수를 음절 개수로 간주하였다. 즉, 다음 15개의 모음('ah', 'ey', 'iy', 'ay', 'ih', 'aa', 'ae', 'er', 'aw', 'uw', 'ao', 'eh', 'ow', 'oy', 'uh')이 포함된 구간을 각각 하나의 음절 단위로 판단하였다. 음절 개수를 모음의 개수로 대치함으로써 특징을 더욱 간단하고 효과적으로 추출할 수 있게 된다. 그리고 유창성 평가를 위한 특징을 추출하기 위하여 앞서 설명한 음소 정렬기를 사용하여 텍스트를 정렬한다. 이렇게 정렬된 텍스트를 이용하여 본 논문에서 추출한 특징 및 방법은 다음과 같다.

- 발화 속도 (SR) : 음절의 개수를 총 발화 시간(초 단위)으로 나눈 값으로 발화 속도를 구하는 특징이다. 여기서 총 발화 시간은 발화 안의 묵음을 포함시킨다[3]. 하나의 발화에 대한 발화 속도는 다음과 같다.

$$SR = \frac{N_S}{t_{end} - t_{beg}} \tag{1}$$

여기서 N_S는 발화 내 모음의 개수이고, t_{end}는 발화의 끝나는 시간이고, t_{beg}는 시작 시간이다.

- 조음 속도 (AR) : 조음 속도는 발화 속도와 비슷한 개념이지만, 묵음을 제거한 시간을 사용한다[3]. 이를 적용하여, 조음 속도 (AR)을 구하는 식은

$$AR = \frac{N_S}{t_{end} - t_{beg} - \sum_{j=1}^{N_{UP}} d_{UP}(j)} \tag{2}$$

로 표현된다. 묵음은 빈 묵음(unfilled pause)과 채워진 묵음(filled pause)으로 나뉘는데, 빈 묵음은 소리가 없는 묵음이고, 채워진 묵음은 더듬음이나 반복(repetition) 등이다. 두 개의 묵음 모두 중요한 역할을 하지만 본 논문에서는 검출하기 쉬운 빈 묵음만을 고려하여 특징을 추출하였는데, 여기서 $d_{UP}(j)$는 j번째 빈 묵음의 지속 시간(duration)이고, N_{UP}는 모든 빈 묵음(unfilled pause)의 개수로, 모든 빈 묵음의 지속 시간을 제거함으로써 발화 속도와 구분된다.

- 발화 시간 비율 (PR) : 총 발화 시간 중 실제 발화한 시간 비율을 나타내는 특징으로 빈 묵음을 제외한 시간을 총 발화 시간으로 나누어서 구하고[3], 이는 다음과 같은 수식으로 표현된다.

$$PR = \frac{t_{end} - t_{beg} - \sum_{j=1}^{N_{UP}} d_{UP}(j)}{t_{end} - t_{beg}} \tag{3}$$

- 연속 발화 평균 길이 (LR) : 이 특징은 0.25초 이상의 빈 묵음 사이의 음절 개수의 평균이다. 0.25초는 cut-off point라고 하며, cut-off point가 0.25초 보다 짧으면 파열음(plosive)이 묵음으로 간주되고, 0.25초 보다 길면 생략되는 양이 많아진다[3]. 이 특징은 화자가 얼마나 연속적으로 발화 했는가를 판단할 수 있는 값으로, 발화 샘플의 시작과 끝이 항상 0.25초 이상의 빈 묵음으로 구성되어 있다면 다음과 같이 표현될 수 있다.

$$LR = \frac{N_S}{N_{UP,0.25} - 1} \tag{4}$$

이 때, N_S는 위에서 설명한 것과 마찬가지로 발화 내 모음의 개수이고, $N_{UP,0.25}$는 0.25초 이상의 빈 묵음의 개수이다.

- 빈 묵음 빈도 (numUP) : 이 특징은 빈 묵음의 빈도를 나타내는 값으로 빈 묵음의 개수를 총 발화 시간으로 나눠서 구한다[3]. 하지만 빈 묵음의 지속 시간을 고려하기 위하여 sigmoid 함수를 설계하여 적용한다. Sigmoid 함수를 적용한 빈 묵음 빈도 계산식은 다음과 같이 표현된다.

$$numUP = \frac{\sum_{j=1}^{N_{UP}} f(d_{UP,j})}{t_{end} - t_{beg}} * 60 \tag{5}$$

여기서 $d_{UP,j}$는 j번째 빈 묵음의 지속 시간이고, $f(.)$는 sigmoid 함수이다. Sigmoid 함수는 <그림 5>와 같이 모델링 하였고, 이와 같은 모델링을 통해 0.2초 이하와 3초 이상의 묵음에 대한 지속 시간이 numUP 값에 미치는 영향을 감소시키는 효과가 나타난다.

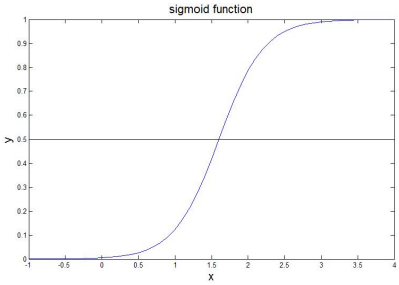


그림 5. Sigmoid 함수
Figure 5. Sigmoid function

Sigmoid 함수의 기본 식 $f(x)$는

$$f(x) = \frac{1}{1 + \exp(-a(x-b))} \tag{6}$$

로 표현되며, 빈 묵음의 지속 시간 중심이 1.6초이고, 0.2초 이하 또는 3초 이상일 때 마진(margin) 1%의 변별력을 가지는 함수를 설계하면, $a = 3.65, b = 1.6$의 파라미터를 가진다.

- 빈 묵음 평균 길이 (lenUP) : 이 특징은 빈 묵음의 평균 길이를 나타내는 값으로 0.2초 이하의 묵음은 초 묵음(micro-pause)으로 더듬거림으로 간주되지 않기 때문에 0.2초 이상의 빈 묵음만을 고려한다[3]. 이 값은 다음과 같은 수식으로 표현된다.

$$lenUP = \frac{\sum_{j=1}^{N_{UP}} d_{UP,j}}{N_{UP}} \tag{7}$$

여기서 N_{UP}는 빈 묵음의 개수이고, 빈 묵음의 지속 시간, 개수 모두 0.2초 이상의 빈 묵음만 고려한다.

- 국소 평균 로그 우도 (LLH) : 이 특징은 발음평가를 위한 특징 중 하나로써 음소별 로그 우도 값이다. 음소는 모든 음소의 값을 사용하지 않고, 발음에 영향이 큰 모음만 사용한다. 로그 우도 값은 두 가지 방법으로 평균을 계산할 수 있는데, 먼저 각 음소 당 시간으로 평균을 내는 국소 평균(local average)은 다음과 같은 식으로 계산한다[5].

$$LLH = \frac{1}{N_S} \sum_{i=1}^{N_S} \frac{l_i}{N_{F,i}} \tag{8}$$

여기서 N_S는 발화내의 모음 개수이고, l_i는 i번째 음소의 로그 우도 값이고, $N_{F,i}$는 i번째 음소의 프레임 개수이다.

- 전체 평균 로그 우도 (GLH) : 위 특징과 같은 로그 우도 값이지만 모든 음소를 총 시간으로 평균을 내는 전체 평균(global average)은 다음과 같이 계산한다[5].

$$GLH = \frac{\sum_{i=1}^{N_S} l_i}{\sum_{i=1}^{N_S} N_{F,i}} \tag{9}$$

- 로그 사후 확률(LPS) : 이 특징은 사후 확률 값으로써 인식기를 통해 인식된 음소가 얼마나 큰 비중을 가지는지를 판단하여 정확한 발음을 했는지를 판단하는 값이다[5]. 일반적으로 t번째 프레임의 정렬된 음소 q_t의 사후 확률 $p(q_t|x)$는

$$p(q_t|x) = \frac{p(x|q_t)p(q_t)}{\sum_{k=1}^{N_{Pho}} p(x|\omega_{k,t})p(\omega_{k,t})} \tag{10}$$

로 표현된다. x는 음성 인식기의 특징 벡터로서 MFCC [11]이고, q_t는 음소 정렬 결과에서 t번째 프레임의 음소 클래스이며, $\omega_{k,t}$는 t번째 프레임에서 k번째 경쟁 음소 클래스를 나타낸다. 이때 사후 확률은 편의상 N_{Pho}개의 모든 음소 확률을 이용하지 않고, N-best 음소 정렬 결과에서 구한 N개의 음소 확률 $p(x|\omega_{k,t})$를 이용하여 근사적으로 계산된다. 이를 바탕으로 사후 확률 특징은 다음과 같이 구한다.

$$LPS = \frac{1}{N_F} \sum_{t=1}^{N_F} \log(p(q_t|x)) \tag{11}$$

$p(q_t|x)$는 t번째 프레임의 사후 확률이고, N_F는 프레임의 개수이다. 이 때 음소 정렬 결과에서 t번째 프레임에 모음

이 존재할 때만 계산한다.

- 음소 지속 시간 점수 (PDS) : 이 특징은 원어민의 음소 지속 시간(phone duration) 확률 모델을 이용하여 비원어민의 발음을 평가하기 위한 것으로 다음과 같이 표현된다[5].

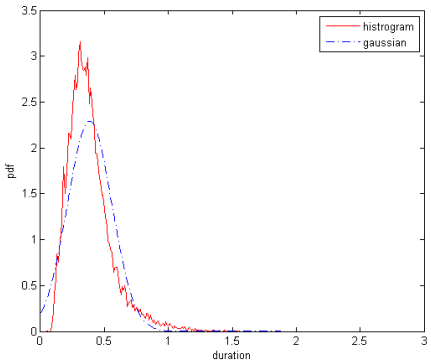
$$PDS = \frac{1}{N_S}\sum_{i=1}^{N_S}\log(p(f(d_i)|q_i)) \quad (12)$$

N_S개의 모음 q_i의 로그 확률 값을 모두 더하는데 이 때 화자마다 발화 속도가 다르기 때문에 이를 고려하기 위하여 함수 $f(d_i)$를 적용하는데 이는 다음과 같이 정의된다.

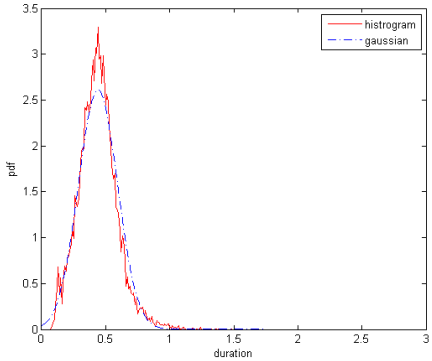
$$f(d_i) = d_i \bullet SR \quad (13)$$

SR은 화자별 발화 속도이고, d_i는 i번째 모음의 지속 시간이다. 이로 인해 각기 다른 발화 속도를 가진 화자들을 대상으로 정규화(normalization)효과를 얻을 수 있다. 그리고 이 특징에서는 확률 모델 $p(x)$를 정의해야 되는데, 본 연구에서는 가우시안 확률 모델을 사용하였다.

<그림 6>(a)는 음소 'iy'에 대하여 원어민 모델의 히스토그램(histogram)과 가우시안 확률 모델을 함께 나타낸 그래프이다. 여기서 'iy'는 ARPABET [12] 형식으로 나타낸 모음의 발음이다. 실선이 히스토그램을 확률 밀도 함수로 변환한 그래프이고, 점선이 가우시안 확률 모델로 일반화 시킨 그래프이다. 지속 시간 모델과 가우시안 확률 모델이 완벽히 일치하진 않지만, 가우시안 확률 모델의 경우 평균과 분산만으로 확률 모델을 모델링할 수 있어서 이를 사용하였다. 다음 <그림 6>(b)와 <그림 6>(c)는 차례대로 'aa'와 'uw'의 히스토그램과 가우시안 확률 모델의 그래프이고, 총 15개의 모음에 대한 모델을 구하여 PDS 특징 추출 계산에 사용된다. 음소마다 차이는 있지만 가우시안 확률모델이 지속 시간 모델과 비슷하다는 것을 확인할 수 있다. 감마 분포가 가우시안 분포보다 실제 히스토그램에 더 유사하지만, 감마 분포는 계산의 복잡성을 고려하여 본 논문에서는 가우시안 분포를 사용하였다.



(a)



(b)

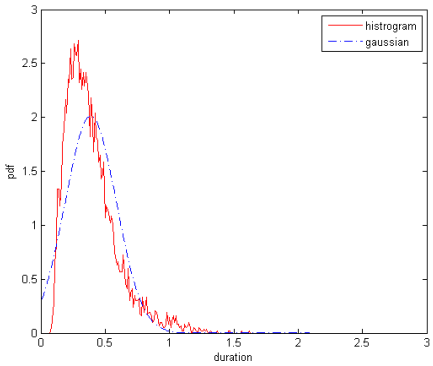
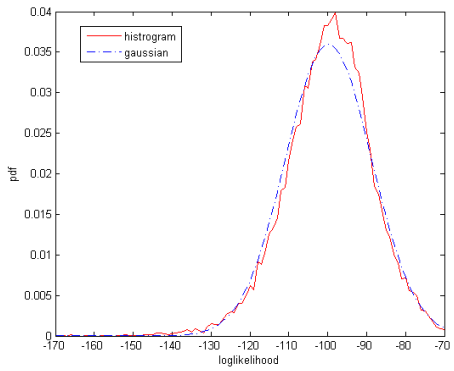


그림 6. 음소 지속 시간에 대한 히스토그램과 가우시안 모델 (a) 'iy' (b) 'aa' (c) 'uw'
Figure 6. Histogram of phone duration and the corresponding Gaussian model (a) 'iy' (b) 'aa' (c) 'uw'

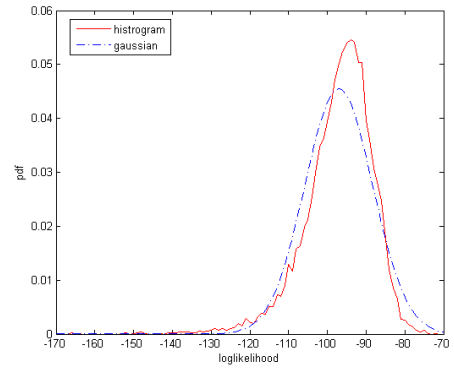
- 음소 로그 우도 점수 (PLS) : 음소 지속 시간 점수 (PDS)와 비슷한 개념으로 로그 우도에 대하여서도 원어민 모델을 만든 후 그 확률 값을 특징으로 사용하였다. 음소 지속 시간 점수와 차이점은 발화 속도를 고려하지 않고, 지속 시간이 아닌 로그 우도를 이용하여 확률 모델을 만들고, 이 확률 값을 특징으로 사용하였고, 수식으로 표현하면 다음과 같다.

$$PLS = \frac{1}{N_S}\sum_{i=1}^{N_S}\log(p(l_i|q_i)) \quad (14)$$

l_i는 음소 q_i에 로그 우도 값으로 원어민의 확률 모델의 로그 확률 값을 사용한다. PDS와 마찬가지로 원어민의 확률 모델을 결정해야 하는데, 이 또한 가우시안 모델을 사용하였고, 히스토그램과 가우시안 모델을 나타낸 그래프는 <그림 7>에서 확인할 수 있다. PDS 모델보다 PLS 모델이 더욱 가우시안과 유사한 분포를 가지는 것을 확인할 수 있다.



(a)



(b)

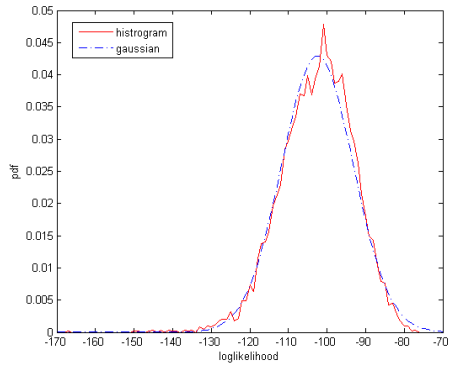


그림 7. 음소 로그 우도 히스토그램과 가우시안 모델 (a) 'iy' (b) 'aa' (c) 'uw'

Figure 7. Histogram of phone log likelihood score and the corresponding Gaussian model (a) 'iy' (b) 'aa' (c) 'uw'

3.4. Support vector regression (SVR)

Support vector machine (SVM)은 Vapnik [13]이 개발한 패턴 식별 방법으로 데이터 마이닝 분야는 물론 얼굴 인식과 같은 패턴인식 응용 분야에도 널리 사용되고 있는 방법이다. 이후 Muller [14], Drucker [15] 등에 의해 회귀 분석(regression)에서도 좋은 성능을 나타낸다고 알려져 있다. SVR의 기본 목표는 <그림 8>과 같이 학습 데이터의 support 벡터를 이용하여 ε의 편차를 갖는 함수 $f(x)$를 결정하는 것이다. 이 때 오류를 최소화하고, 비용 함수와 우도를 고려하여 라그랑지(Lagrange) 기법을 이용하면 최적인 $f(x)$를 결정할 수 있다.

$$f(x)=\sum_{j=1}^{J}\alpha_j k(x,x_j)+b \tag{15}$$

이때 α_j는 j번째 데이터의 가중치, J는 전체 데이터 개수, $k(x,x_j)$는 사상된 공간에서 거리 관계를 보존하도록 정의된 커널(kernel) 함수, b는 바이어스(bias)이다.

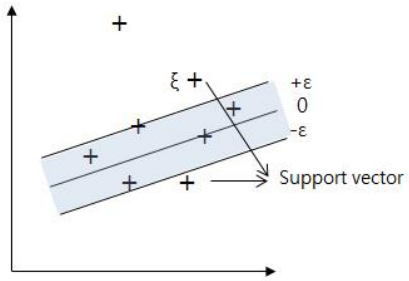


그림 8. ε의 편차를 갖는 함수 [16]

Figure 8. Function on ε deviation [16]

커널 함수는 여러 종류가 존재하는데, 본 논문에서는 다음과 같이 벡터의 내적$(\bullet)$으로 정의된 선형 커널(linear kernel)을 이용하였다.

$$k(u,v)=u\bullet v \tag{16}$$

이렇게 정의된 $f(x)$를 이용하여 검증 데이터의 회귀 분석 값을 산출하게 되는데[16], 이는 <그림 9>과 같이 도식화 할 수 있다. 여기서 $x_j(j=1,2,3,....J)$는 학습 데이터, x는 검증 데이터, $\Phi(x)$는 데이터로부터 추출한 특징 벡터를 나타낸다.

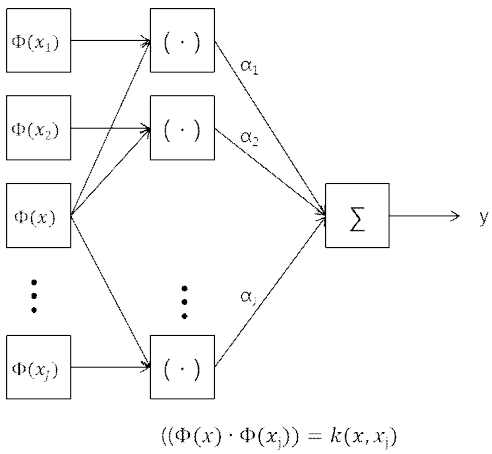


그림 9. SVR의 회귀 분석 구조도 [16]

Figure 9. Architecture of regression analysis by SVR [16]

4. 실험 결과

본 논문에서는 3절에서 언급한 음소 정렬기 결과를 기반으로 하여 특징을 추출하고, SVR을 이용하여 특징들을 결합하여 유창도를 계산하였다.

4.1. 시험 문제 유형 및 음성 데이터베이스

유창성 평가를 위한 한국인 영어 발화 DB는 일반 대학생 48명(여성 26명, 남성 22명)이 3회에 걸쳐 치룬 모의 유창성 평가 시험 데이터를 사용하였으며, 각 회당 시험은 총 5개의 과제로 이루어져 있고, 과제별 유형 및 기본 녹음 시간은 <표 2>와 같다. 총 음성 DB는 8시간 50분 정도의 분량을 가진다. 부록 A에 1회차 영어 말하기 시험 문제를 제시하였다.

표 2. 과제별 유형 및 녹음 시간
Table 2. Task and recording time of each task

과제 번호	과 제	녹음 시간 (초)
1	주어진 문장을 보고 읽기	9
2	주어진 주제에 관한 자신의 생각 말하기	60
3	주어진 그림으로 이야기 구성(3인칭)	50
4	주어진 상황에 맞는 이야기 구성(2인칭)	50
5	주어진 도표 또는 그래프 분석하여 말하기	50

표 3. 유창성 평가 항목
Table 3. Fluency evaluation list

채점 영역 (가중치)	평 가 항 목
전체(5)	전체적인 유창성 정도는 얼마인가?
발음(1)	발음이 명확하여 정확히 알아들을 수 있는가?
	강제, 억양, 리듬을 자연스럽게 구사하고 있는가?
유창성(1)	발화 속도가 자연스럽게 유지되고 있는가?
	발화 중 휴지나 간투사(non-speech)가 많아 부자연스럽지 않은가?
	같은 단어나 어구를 반복하여 문장의 흐름이 끊기지 않는가?
언어사용 (1)	문법에 오류가 없는가?
	다양하고 적절한 어휘와 표현을 사용하고 있는가?
	영어만 사용하고 있는가?
	완결된 문장으로 말하는가?

4.2. 평가자 채점 결과

학습과 검증에 사용할 채점 점수는 한국인 대학 영어 강사 1명과 한국인 이중 언어(bilingual) 대학생 2명이 시행하였다. 1번 과제를 제외한 모든 과제는 10개의 평가 항목을 1-5점으로 채점하였다. 채점은 5단계로 분류된 샘플을 합의 후 선정하여 독립적으로 채점을 진행하였다. 각 평가 항목은 <표 3>과 같고, 보고 읽기 형식의 1번 과제는 주어진 한 문장을 단순히 읽는 유형이고 문장이 짧아서 본 논문의 유창성 평가 실험에는 사용하지 않는다. 각 항목은 모두 5점 만점이고, 전체적인 (holistic) 유창성 평가 항목은 가중치 5를 곱하여 25점 만점으로 총 70점 만점의 점수가 계산되는데, 이를 일반적이고 직관적인 점수 분포를 보기 위하여 100점 만점으로 환산하였다.

평가자간의 상관 분석을 위해 IBM SPSS Statistics를 이용한 통계 분석을 실시하였다. 통계 분석 방법은 이변량 교차상관 분석(cross correlation)을 시행하였고, 상관 계수는 Pearson 계수[16]를 사용하였다. 3명의 평가자간 상관 계수는 <표 4>와 같고, 회차별 평가자의 상관 계수 평균은 <표 5>와 같다.

표 4. 평가자간 상관 계수
Table 4. Correlation between human raters

구분	평가자1	평가자2	평가자3
평가자1	1.00	0.83	0.80
평가자2	·	1.00	0.87
평가자3	·	·	1.00

표 5. 회차별 평가자간 상관 계수
Table 5. Correlation between human raters for each test set

회차	1회	2회	3회
상관계수	0.81	0.85	0.84

상관 계수는 -1에서 1의 범위를 가지게 되고, 본 논문의 평가자간 상관 계수는 0.80 이상이다. 특히, 이중 언어 사용자인 평가자2와 평가자3 사이의 상관 계수는 0.87로 매우 높은 값을 나타내고 있다. Kormos [3]의 실험에서는 외국어 강사 3명과 원어민 3명의 유창성 평가의 평가자간 상관 계수가 각각 0.78, 0.72가 나왔다. 그리고 Neumeyer [5]의 발음 평가 실험에선 10명의 평가자 중 가장 좋은 상관 계수를 갖는 평가자 5명을 선발하여 사용하였는데, 이 5명의 평균 상관 계수가 0.80의 값으로 두 논문과 비교하여 본 논문의 평가자간 상관 계수는 높은 편에 속한다.

평가자 점수를 이용하여 과제별 평가 기여도를 구할 수 있는데, 이는 3명의 평가자 점수의 평균 값과 각 평가자의 과제별 점수의 평균의 상관 계수를 이용하여 구하고, 결과는 <표 6>과 같다. 과제 1번과 평균 점수와의 상관 계수가 0.81로 다른 유형의 과제보다 낮은 것으로 보아 과제 1번의 유형이 유창도 평가에 좋지 않음을 판단할 수 있다.

표 6. 평가자 평균 점수와 평가자내 과제별 상관 계수
Table 6. Correlation between each rater's score and its mean score

과제번호	1	2	3	4	5
상관계수	0.81	0.94	0.95	0.94	0.94

<그림 10>는 평가자 점수의 히스토그램으로 평가자간 점수의 분포가 비슷한 양상을 보인다. 이는 평가자 모두 적절한 범주로 채점을 했다고 볼 수 있는 결과이다.

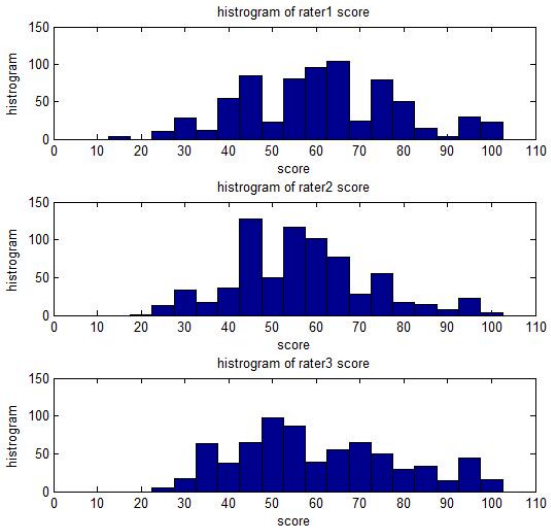


그림 10. 3명 평가자 점수의 히스토그램
Figure 10. Histogram of 3 rater scores

4.3. 특징 분석 결과

특징은 학습에 사용되었던 원어민 발화와 유창성 평가를 위한 한국인의 영어 발화를 인식한 결과를 이용하여 추출하였다. 각 특징들의 성능을 검증하기 위하여, 평가자의 채점 점수와 특징들과의 상관 계수를 각 회당, 총 3회 구하였고, 그 결과는 각각 <표 7>에 나타난다. <표 7>에서 볼 수 있듯이 SR, AR, LR 특징이 채점 점수와 높은 상관계수를 가진다. 특히 SR의 경우 가장 높은 상관계수를 보이고 있는 것으로 보아 가장 중요한 특징이라고 판단할 수 있겠다. 전반적으로 발음을 평가하는 특징인 LLH, GLH, LPS, PDS, PLS의 경우 상관계수가 낮게 나왔지만, 평가 점수 중 발음 평가 부분은 2항목 점수로 비중은 70점 만점에 10점으로 다소 낮기 때문에 나타나는 결과로 보인다. 이에 확인을 위해 발음 평가 항목 점수만을 가지고 상관계수를 분석하였지만, 현재 본 논문에 반영된 발음 관련 평가 항목 중 해당하는 항목은 1개뿐이고, 화자별 변별력도 높지 않아 기대하는 상관계수의 상승은 없었다. 이에 발음 관련 평가에 더욱 큰 변별력을 갖는 평가 항목, 특징 추출이 필요하다고 판단하였다. 발음 평가 특징들만 비교하면, PDS와 GLH가 좋은 성능을 보이고 있다.

표 7. 채점 점수와 특징들의 상관 계수
Table 7. Correlation between human score and features

특징	SR	AR	PR	LR	num UP	len UP	LLH	GLH	LPS	PDS	PLS
1회	0.81	0.67	0.72	0.76	-0.58	-0.42	0.07	0.13	0.04	-0.18	0.10
2회	0.77	0.69	0.68	0.71	-0.53	-0.36	0.06	0.11	0.06	-0.13	0.08
3회	0.71	0.57	0.66	0.67	-0.49	-0.36	0.12	0.17	0.06	-0.23	0.14

4.4. 자동 유창성 평가 결과

유창도를 계산하기 위하여 48명의 화자는 6명씩 8개의 그룹으로 분할한다. 분할된 8개의 그룹 중 7개의 그룹을 학습에 사용하고 1개의 그룹을 검증에 사용하게 된다. 이 때 학습 그룹의 데이터는 정량화를 하고, 학습 데이터의 평균과 분산을 이용하여 검증 그룹의 데이터에 적용하여 실험하였다. 이 때 모든 그룹이 검증에 사용될 수 있게 교차 검증 기법(cross validation)을 사용하여 총 8번의 실험을 하였다. 학습에는 화자의 발화에서 뽑은 특징들을 SVR의 입력 데이터로 사용하고, 3명의 평가자 점수의 평균값을 SVR의 목표 값(desired value)으로 사용하였으며, SVR 커널은 선형 커널을 사용하였고, 비용 상수(cost constant)는 1로 설정하였다. 이 비용 상수는 어느 정도 범위 내에선 결과에 큰 영향을 미치지 않지만, 너무 작은 값으로 설정하면 산출되는 유창도 값의 분산이 너무 작아져 점수가 한 영역대로 몰리는 현상이 나타날 수 있다. 검증은 학습한 SVR 모델을 이용하여 검증 그룹인 6명의 유창성 평가 특징으로 유창도를 산출하였고, 이 유창도 점수와 평가자의 유창도 점수의 상관 계수를 통하여 성능을 확인하였다. 그리고 교차 검증을 시행한 각 검증 데이터의 상관계수의 평균을 최종 성능 값으로 사용하였다.

<표 8>은 임의의 한 검증 그룹의 평가자 점수와 SVR에서 산출한 유창도 점수를 나타낸 것이다. 전체적으로 점수의 분포가 비슷한 양상을 보이고 있지만, 완전히 일치하게 비례하지는 않다는 것을 확인할 수 있다.

표 8. 평가자와 SVR의 유창도 점수
Table 8. Fluency score of raters and SVR

화자번호	1	2	3	4	5	6
평가자 점수	33.8	84.4	97.6	87.6	46.0	42.6
SVR 점수	42.3	70.8	83.7	67.1	47.2	46.7

<표 9>는 과제별 상관 계수 결과로 각 과제만 사용하여 유창도를 계산하였을 때 그 과제의 평가자 점수와 얼마나 상관을 갖는지를 나타낸다. 이 값은 전체적으로 비슷한 성능을 나

타내고 있으며, 1번 과제는 제외하고 4번 과제인 '주어진 상황에 맞는 이야기 구성하기' 유형이 가장 낮은 성능으로 나타남을 확인할 수 있다. <표 10>은 본 논문에서 사용하지 않기로 하였던 1번 과제를 제외한 4개의 과제를 사용하여 회차별로 유창도를 계산하고, 각 회차에 해당하는 평가자의 점수와의 상관계수를 구하여 나타내었고, 유창성 계산에 1번 과제의 부적합성을 보이기 위해 1번 과제를 포함한 결과 또한 나타내었다. 1번 과제의 상관 계수는 <표 9>에서 0.66으로 다른 과제와 비교하여 낮은 점수를 확인할 수 있고, 또한 <표 10>에서 볼 수 있듯이 1번 과제를 합하여 유창도를 산출하였을 때 성능이 다소 떨어지는 것을 볼 수 있다. 회차별 상관 계수 결과로 산출한 최종 성능은 약 0.01정도의 편차를 가지고 있고, 1회차의 성능이 가장 높은 것으로 나타나고 있다. 여기서 <표 10>의 값은 각 회차별 SVR 점수와 평가자 점수를 비교한 것으로, 전체적인 성능을 나타내기에는 다소 무리가 있다. 본 논문에서는 평가자간의 성능과 SVR의 성능을 나타내기 위해 <그림 11>을 나타내었다.

표 9. SVR과 채점 점수의 과제별 상관 계수
Table 9. Correlation between SVR and human score of each task

과제번호	1	2	3	4	5
상관계수	0.66	0.80	0.79	0.76	0.78

표 10. SVR과 채점 점수의 회차별 상관 계수
Table 10. Correlation between SVR and human score for each test set

회 차	1회	2회	3회
상관계수	0.88	0.88	0.84
상관계수(1번 과제 포함)	0.84	0.85	0.82

<그림 11>은 SVR 점수와 평가자 점수의 평균과의 상관 관계와 각 평가자 점수와 평가자 점수의 평균의 상관 관계를 산포도(scatter plot)를 통해 나타낸 것이다. x축은 공통적으로 평가자의 평균 점수고, y축은 각 평가자의 점수, SVR의 점수를 나타낸다. 이 그래프에서 각 데이터들의 기호(symbol)는 (a), (b)에서 각각 평가자와 회차를 상징한다. 산포도의 데이터 기호들이 x와 y가 비례하게 그려놓은 직선에 집중되어 있을수록 오차가 작은 것을 뜻하며, 평가자 점수인 (a)의 데이터들이 SVR 점수인 (b)의 데이터보다 직선에 집중되어 있음으로 평가자의 점수가 더 오차가 작음을 알 수 있지만, 본 실험의 목표였던 평가자의 점수에 SVR 점수가 많이 근접해 있다는 것을 확인할 수 있다.

채점 점수와 특징들의 상관계수를 분석한 결과 중 계수 값이 높은 것으로 나타난 3개의 특징(SR, AR, LR)만을 이용하여 SVR을 학습하고 유창성 점수를 추정한 결과, SVR 점수와 평가자 점수의 상관계수가 0.80로 나타났다. 특히 SR, AR, LR 특징은 약간의 음성인식 오류가 있더라도 인식결과만으로 음절 개수는 충분히 추정 가능하고, 묵음 구간은 에너지 검출만으로 구할 수 있으므로, 전사파일이 없어도 유창성 평가가 가능한 자동 유창성 평가 도구 개발의 가능성을 보여주었다.

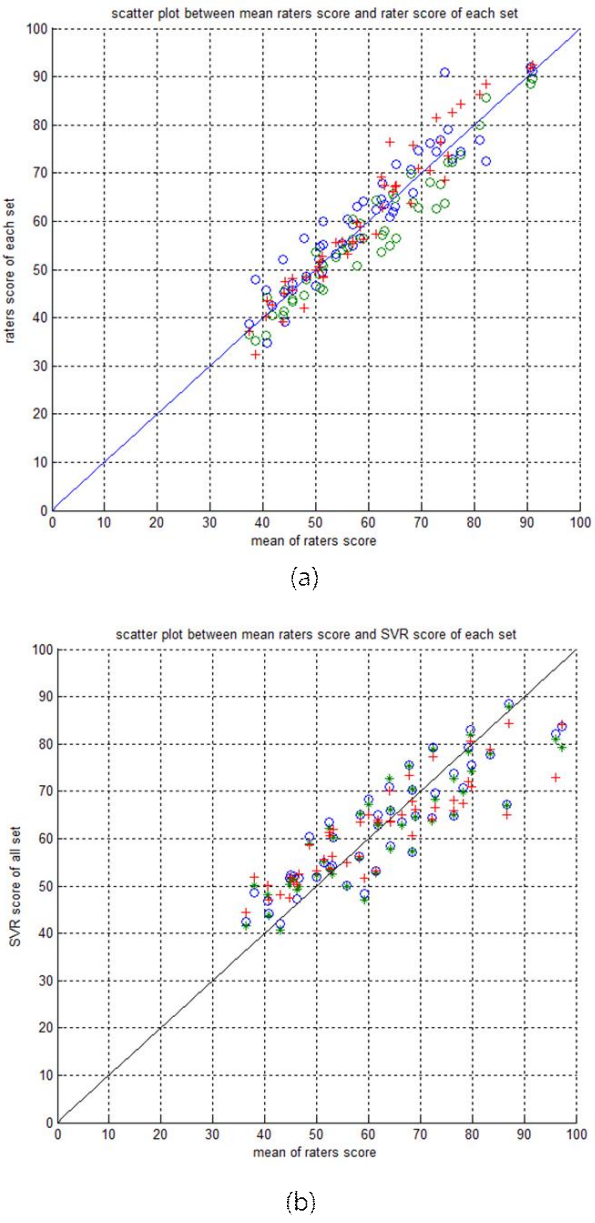


그림 11. 유창성 평가 점수 산포도
(a) 평가자 평균 점수와 각 평가자 점수
(b) 평가자 평균 점수와 SVR 점수
Figure 11. Scatter plot of fluency scores (a) between rater mean score and each rater score, (b) between rater mean score and SVR score

5. 결론

본 논문에서는 유창성 평가를 위한 특징들을 소개하고, 각 특징들을 SVR을 이용하여 결합하여 유창도를 산출하여, 평가자 점수와 상관분석을 통해 회차별, 과제별 성능을 확인하였다. SVR 점수의 성능을 확인하기 전에 평가자간, 평가자내 상관 계수를 통해 검증한 결과 평가자간 상관 계수는 이와 비슷한 실험의 다른 논문에서의 상관 계수보다 다소 높은 성능을 나타내었으며, SVR 점수의 과제별 상관 분석 결과에서 1번을 제외한 모든 과제가 0.76이상으로 좋은 성능을 보였으며, 회차별 분석에서는 1회차의 상관계수가 0.88로 가장 높은 성능의 유창성 평가 산출을 보여주었다. 또한 보고-읽기 유형인 1번 과제의 경우 평가자내 상관 계수에서 낮은 값을 보여주었는데, SVR에서 1번 과제를 포함하여 유창도 점수를 산출하였을 때 1번 과제를 제외하였을 때보다 낮은 상관 계수가 나타나 1번 과제의 유창성 평가에 부적합함을 다시 한 번 확인할 수 있었다. 본 논문의 각 특징들과 평가자 점수와의 상관 분석에서 상관 계수는 각 특징 마다 큰 차이를 보였으나, 이를 모두 결합하여 유창도를 산출하였을 때, 세트별 평균 0.87의 성능을 보여주고 평가자와 SVR 산포도를 비교함으로써 SVR을 이용한 유창성 평가 특징의 결합이 적합함을 확인할 수 있었다. 이는 유창성 평가시 응시자 사전 분류 등의 용도로 활용하기에는 부족함이 없는 성능이라고 판단된다.

본 논문에서는 SVR을 이용한 유창성 평가용 특징 결합 방법을 제시하고 그 적합성을 확인함으로써, 향후 자동 유창성 평가 도구의 실제 응용 가능성을 보여 주었다. 이를 기반으로 좀 더 구체적이고 우수한 성능의 유창성 자동 평가 도구를 개발할 계획이다.

감사의 글

이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2012R1A1A2042381). 본 논문에서 사용된 영어 말하기 유창성 평가용 음성 데이터베이스를 제공해 준 SK텔레콤 성장기술원에 감사를 드립니다.

참고문헌

[1] Fillmore, C. J. (1979). On fluency. *Individual differences in language ability and language behavior*, 85-101.

[2] Chambers, F. (1997). What do we mean by fluency?. *System*, 25(4), 535-544.

[3] Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145-164.

[4] Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. *British Studies in Applied Linguistics*, 12, 58-71.

[5] Neumeyer, L., Franco, H., Digalakis, V., & Weintraub, M. (2000). Automatic scoring of pronunciation quality. *Speech Communication*, 30(2), 83-93.

[6] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., & Woodland, P. (2006). *The HTK Book* (for HTK version 3.4). *Cambridge University Engineering Department*, 2(2), 2-3.

[7] Paul, D. B., & Baker, J. M. (1992, February). The design for the Wall Street Journal-based CSR corpus. *In Proceedings of the Workshop on Speech and Natural Language* (pp. 357-362). Association for Computational Linguistics.

[8] Vertanen, K. (1994). HTK Wall Street Journal Training Recipe. http://www.keithv.com.

[9] Garofolo, J. S. (1988). Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database. *National Institute of Standards and Technology (NIST)*, Gaithersburgh, MD, 107.

[10] Lenzo, K. (2007). The CMU pronouncing dictionary.

[11] Imai, S. (1983). Cepstral analysis synthesis on the mel frequency scale. *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 93-96), IEEE.

[12] Shoup, J. E. (1980). Phonological aspects of speech recognition. *Trends in Speech Recognition*, 125-138.

[13] Haykin, S. (1999). *Neural Networks,* Prentice Hall.

[14] Müller, K. R., Smola, A. J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., & Vapnik, V. (1997). Predicting time series with support vector machines. *In Artificial Neural Networks — ICANN'97* (pp. 999-1004). Springer Berlin Heidelberg.

[15] Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, 155-161.

[16] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222.

[17] Kendall, M. G. (1948). Rank correlation methods.

[18] Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14(4), 423-441.

[19] Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics,* 17(1), 84-119.

• **장병용 (Jang, Byeong-Yong)**
충북대학교 제어로봇공학전공
충북 청주시 흥덕구 내수동로 52(개신동)
Email: byjang@cbnu.ac.kr
관심분야: 음성인식, 유창성
현재 제어로봇공학과 석사 재학 중

• **권오욱 (Kwon, Oh-Wook)** 교신저자
충북대학교 전자공학부
충북 청주시 흥덕구 내수동로 52(개신동)
Tel: 043-261-3374
Email: owkwon@cbnu.ac.kr
관심분야: 음성인식, 감정인식, 음성신호처리
2003~현재 충북대학교 전자공학부 교수

부록

A. 영어 말하기 시험 문제 (1회차)

1. Read this sentence.
(Please read this sentence that you see.)
“Every man for himself, and the Devil take the hindmost.”

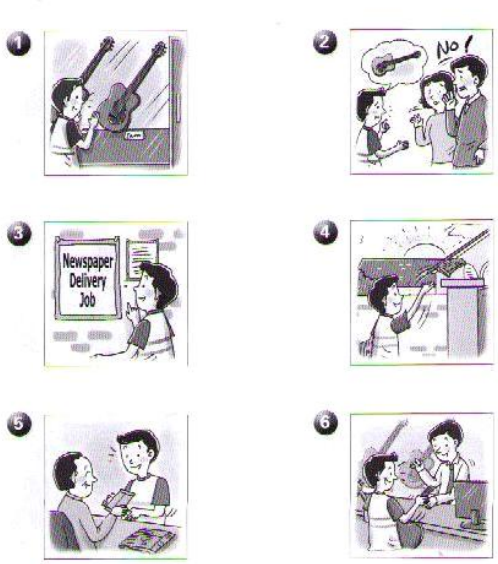
2-0. Answer the question!
(You will be asked four questions about your ideas on taking pictures. Do you like to take pictures?)

2-1. Answer the question!
(What kind of photos do you enjoy looking at?)

2-2. Answer the question!
(What kind of camera do you have?)

2-3. Answer the question!
(Why do people take photos?)

3. Answer the question!
(Describe the given illustrations in order to make a story.)



4. Answer the question!
(One of your friends is shorter than others. She worries about it. What would you like to say to her?)

5. Answer the question!
(Describe the graph or chart using the adjectives more, most, less or least.)