

Relationships between rhythm and fluency indices and listeners' ratings of Korean speakers' English paragraph reading

Hyunsong Chung*

Department of English Education, Korea National University of Education, Chungbuk, Korea

Abstract

This study investigates the relationships between rhythm and fluency indices and listeners' ratings of the rhythm and fluency of Korean college students' English paragraph reading. 17 university students read and recorded a passage from "The North Wind and the Sun" twice before and after three months of English pronunciation instruction. Seven in-service and pre-service English teachers in graduate school assessed the rhythm and fluency of the utterances. In addition, the values of 14 indices of rhythm and fluency were extracted from each speech and the relationships between the indices and the listeners' ratings were analyzed. The rhythm indices of the speakers in this study did not differ significantly from those of native English speakers presented in previous studies in %V, VarcoV, and nPVIV, but were higher in ΔV, ΔC, and VarcoC and lower in speech rate. The level of rhythm and fluency demonstrated by Korean college students was comparable, at least in terms of objective values for certain indices. The fluency indices, such as percentage of pauses, articulation rate, and speech rate, significantly contributed more to predicting both rhythm and fluency ratings than the rhythm indices.

Keywords: rhythm, fluency, rhythm index, fluency index

1. 서론

이 논문은 한국 대학생들이 발화한 영어 읽기 발화를 리듬(rhythm) 및 유창성(fluency) 지수를 사용해 측정 후, 그 발화에 대한 청자의 리듬 및 유창성 평가에 지수가 어떤 기여를 하는지 분석한 연구이다. 유창성 평가는 응용언어학 분야에서는 주로 제 2언어의 발화를 평가하는 데 초점을 두고 있으며, 해당 언어를 얼마나 몰 흐르듯이 부드럽고 쉽게 발화하는지 평가한다(Derwing et al., 2004; Isaacs & Thomson, 2013). 유창성은 이해

가능성(intelligibility), 이해도(comprehensibility), 외국어 말투(accentedness)와 함께 제 2언어 발화를 평가할 때 가장 많이 사용되는 평가 도구이다.

이해가능성과 이해도, 외국어 말투는 화자의 발화와 청자의 상호작용에 초점을 두고 평가가 이루어지는 반면, 유창성의 경우에는 화자의 발화 능력에 초점을 두고 평가가 이루어지다가, 최근 응용언어학 분야에서 청자의 판단을 반영하여 평가를 수행하는 방향으로 흐름이 바뀌었다(Thomson, 2015).

하지만, 여전히 음성학 분야에서는 객관적인 수치로 유창성

* hchung@knue.ac.kr, Corresponding author

Received 20 November 2022; Revised 10 December 2022; Accepted 11 December 2022

© Copyright 2022 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

을 평가하려는 시도가 주를 이루고 있고, 응용언어학 분야에서는 청자의 판단에 초점을 두고 평가하려는 흐름이 더 크다고 볼 수 있다.

유창성을 평가할 때 리듬을 유창성의 하위 요소로 보는 경향이 있지만, 리듬은 특정 언어의 리듬이 강세 기반을 보이는지, 음절 기반을 보이는지, 모라 기반을 보이는지 언어의 리듬 체계를 분류할 때 사용하는 도구로, 유창성을 측정하는 지수와 리듬을 측정하는 지수가 다른 연구가 많다(Thomson, 2015; White & Mattys, 2007). 물론, 발화 속도나 조음 속도와 같이 리듬 요소나 유창성 요소에 겹쳐져 있는 요소도 있는 것은 사실이다. 리듬 평가 또한 컴퓨터 등의 측정 도구가 발달되지 않았을 때는 청취 또는 인상적 평가에 의지해 왔는데, 오늘날 측정 도구가 발달하면서, 구체적인 지수를 통해 언어의 리듬 체계를 분류하려는 시도가 활발하게 진행되고 있다(Deterding, 2001; Grabe & Low, 2002; Ling et al., 2000; Mariano & Romano, 2011; Ramus et al., 1999; White & Mattys, 2007).

리듬과 유창성 측정 도구로 다양한 지수를 사용해 왔는데, 많이 사용되는 것을 중심으로 표 1과 표 2로 정리하였다.

표 1. 리듬 측정 지수
Table 1. Rhythm indices

관련 연구	이름	설명
Ramus et al. (1999)	%V	전체 발화에서 모음 구간이 차지하는 비율
	ΔV	모음 구간들의 표준 편차
	ΔC	자음 구간들의 표준 편차
Ling et al. (2000)	nPVIV	발화 속도에 정규화된 모음의 변이 지수
	rPVIC	자음의 변이 지수
Deterding (2001)	VI	음절의 변이 지수
Dellwo (2006)	VarcoC	자음 구간의 평균 길이에 대한 표준 편차의 비율
White & Mattys (2007)	VarcoV	모음 구간의 평균 길이에 대한 표준 편차의 비율
	발화 속도	초당 음절 개수(전체 발화에 휴지 구간 포함)
Chung et al. (2008)	기능어 비율	휴지 구간을 제외한 전체 발화 길이에서 기능어 구간이 차지하는 길이 비율

Adapted from Chung et al. (2008) with CC-BY-NC.

표 2. 유창성 측정 지수
Table 2. Fluency indices

이름	설명
발화 속도	초당 음절 개수(전체 발화에 휴지 구간 포함)
조음 속도	초당 음절 개수(전체 발화에 휴지 구간 제외)
휴지 개수	전체 발화에 포함된 휴지의 개수

유창성은 일반적인 측정 방법이기 때문에 별도의 관련 연구를 제시하지 않았다. ‘PVI’는 ‘Pairwise Variability Index’를 지칭하는 것으로, 인접 자음 또는 모음들 간의 길이 변이 지수를 측정하는 것이다. ‘nPVI’는 ‘normalized PVI’를 지칭하는 것으로, 발화 속도에 대해 지수를 정규화하기 위하여, 전체 평균 길이를

구하지 않고, 두 인접 모음 또는 자음의 길이 차이를 그 두 음의 평균으로 나누어 계산한 후, 그 계산된 값의 전체 평균을 백분율 지수로 나타낸 것이다. ‘rPVI’는 ‘raw PVI’를 말하는 것으로 두 인접음의 평균으로 나누지 않고, 각 인접음의 길이 차이를 모두 더한 후, 전체 평균을 구한 것이다. ‘rPVI’ 지수와 같이 단순히 길이 차이의 전체 평균을 구하게 되면 전체 발화에서 초반 발화와 후반부 발화의 속도가 다를 때, 속도 차이를 반영한 인접음 간의 변이 정도를 반영하기 어렵기 때문에 발화 지점의 발화 속도 차이에 따라 지수가 왜곡될 수 있다. ‘VI’는 ‘Variability Index’를 가리키는데, 각 음절의 길이를 전체 평균으로 나눈 후, 그 차이의 전체 평균을 구하는 방식이다. 표 1의 지수 중, ‘%V’는 전체 발화에서 모음이 차지하는 길이 비율을 나타낸다. 비강세 모음의 약화가 제대로 실현된 발화라면 전체 발화에서 모음이 차지하는 비율이 낮고, 그렇지 않으면 모음이 차지하는 비율이 높기 때문에 그 수치가 낮을수록 강세 기반을 보이는 발화라고 할 수 있다. ‘기능어 비율’도 기능어는 기본적으로 강세를 받지 않는 경우가 많기 때문에 발화 전체에서 차지하는 기능어 길이 비율이 낮을수록 강세 기반 리듬 경향을 보인다고 할 수 있다. 나머지 지수들은 인접음들 간의 길이 변이를 나타내는 지수이기 때문에 그 값이 클수록 강세 기반 리듬 경향을 보이는 것이다.

리듬 및 유창성에 관한 평가는 평가 점수의 예측 모델을 만들 기보다는 모국어가 다른 화자들 간의 리듬의 차이를 분석하는 연구가 주로 진행되어 왔다. Chung(2013)에서는 한국어 영어 학습자들이 타 언어 화자와 영어로 담화를 나눌 때의 리듬의 변화를 VI와 발화 속도를 사용해 측정하였다. 분석 결과, 담화 중 대화 상대방의 모국어 배경에 따른 VI값의 변화는 통계적 차이가 없는 것으로 나타났고, 발화 속도는 상대방이 영어 비 원어민 화자일 때, 원어민 화자와 대화할 때보다 더 느려지는 것으로 분석되었다.

Choe(2019)는 부산에 거주하는 화자들에게 ‘Time’지에 실린 기사를 읽고 녹음하게 한 후 다섯 문장을 추출해 %V, ΔV, ΔC, VarcoV, VarcoC, rPVIV, rPVIC, nPVIC, nPVIV, 발화 속도, 휴지 개수 등을 측정하였다. 추출된 지수를 원어민과 비교했을 때, 부산 화자들은 발화 속도에 대한 정규화가 반영되지 않은 ΔV, ΔC, rPVIV, rPVIC 등에서는 영어 원어민 화자들에 비해 더 강세 기반 리듬 성향을 보이는 것으로 나타났다. 하지만, 발화 속도에 대해 정규화를 반영한 nPVIV 지수에서는 영어 원어민 화자들보다 그 값이 낮아, 상대적으로 음절 기반 리듬에 가까운 것으로 나타났다.

Kim(2021)에서는 리듬 지수 중, 모음 지수인 %V, VarcoV 및 nPVIV를 활용해 한국인 영어 학습자와 북미 영어 원어민 간의 지수 차이를 분석하고, 그것이 영어 원어민 청자의 외국어 말투 (accentedness) 평가와 어떤 상관성을 갖는지 분석하였다. %V, VarcoV, nPVIV 모두 외국어 말투 청취 평가에서 외국어 말투가 없는 집단(no accent)과 외국어 말투가 강한 집단(strong accent), 외국어 말투가 없는 집단과 외국어 말투가 약하게 있는 집단(weak accent)을 잘 구분해 주는 지표로 역할을 하고 있음을 발

견하였다. 외국어 말투가 강한 집단과 외국어 말투가 약하게 있는 집단을 구분하는 경우에는 %V는 기여하지 못하고, VarcoV와 nPVIV만 기여하는 것으로 나타났다. 이 연구에서 외국어 말투가 없는 집단은 외국어 말투가 약하게 있는 집단이나, 외국어 말투가 강한 집단에 비해 %V는 낮았고(40.4), VarcoV(61.3)와 nPVIV(67.4)는 높았다.

Chung et al.(2008)에서는 108개 낭독체 문장과 94개 단어 및 구를 녹음한 후, 위에 사용된 리듬 및 유창성 측정 지수를 사용하여 리듬 평가 모델을 만든 후 모델에서 예측한 자동 평가 점수와 인간 평가자들의 평가 점수와의 상관관계를 구하였다. 실험 결과 자동 평가 점수와 인간 평가자들의 평가 점수 간의 상관관계가 적지 않은 것으로 나타나, 리듬 자동 평가의 가능성을 보여주었다.

Kim & Jang(2019)에서는 휴지의 개수와 위치 및 발화 속도, 화자의 능숙도가 청자의 유창성 평가에 미치는 영향을 모델링하였는데, 휴지 개수와 휴지의 위치가 유창성 평가에 더 큰 영향을 주는 것으로 나타났다.

본 논문에서 이와 같은 이론적 배경과 선행 연구를 바탕으로 대학생들에게 발음 수업을 한 학기 동안 실시한 후, 영어 지문을 읽고 녹음하여, 청자의 리듬 및 유창성 평가를 진행하였고, 녹음된 음성 파일을 분석하여 선행 연구에서 사용된 다양한 리듬 및 유창성 지수를 추출하였다. 추출된 지수가 청자의 리듬 및 유창성 평가와 어떤 관련성이 있고, 최적의 평가 예측 모델을 구축하는데 기여하는 지수는 무엇인지 분석하였다.

2. 연구 방법

본 연구를 위해 총 17명의 발화를 녹음하였고, 그에 대한 청취 평가를 진행하였다. 녹음된 발화의 리듬 및 유창성 지수를 추출한 후 그 지수와 청취 평가 간의 관련성을 분석하였다.

2.1. 녹음 지문

녹음 지문은 선행 연구와의 비교를 위해 ‘The North Wind and the Sun’ 지문을 사용하였다. 이 녹음 지문은 The International Phonetic Association(1999)에 언어 간의 전사 비교를 위해 가이드로 제시되어있는 지문이기도 하다. Grabe & Low(2002)와 Mariano & Romano(2011)에서도 이 영어 지문을 비롯해, 이를 각국의 모국어로 번역하여 본인의 모국어로 읽은 발화에서 리듬 지수를 추출하여 언어 간의 리듬 지수를 서로 비교하고 있다.

2.2. 참여자

녹음에는 K대학교에 재학하고 있는 17명의 대학생이 참여하였다. 학생들은 ‘영어 발음 연습’을 수강한 학생들로 수업 첫 시간에 과제로 읽기 지문을 읽고 녹음해 오도록 했고, 학기가 마무리되기 직전 동일한 지문을 다시 녹음하도록 했다. 녹음은 Praat(Boersma & Weenink, 2022)를 활용하도록 했는데, 19명이 제출한 녹음 중 44,100 Hz 표본화(sampling)로 녹음되지 않은 2명의 녹음은 청취 평가에서 오류의 원인이 되기 때문에 분석 대

상에서 제외해, 최종적으로 17명의 사전 녹음과 사후 녹음을 분석 대상으로 해 모두 34개의 녹음 파일을 수집하여 분석하였다.

청취 평가는 K대학교 대학원에 재학 중인 7명의 현직 교사 및 예비 교사들이 진행하였다. 평가자를 원어민으로 하지 않고 비원어민 교사로 한 것은, 실제로 교사들이 말하기 평가의 평가자로 참여하고 있기 때문에 굳이 원어민 평가자로 고정할 필요가 없다는 판단 때문이었다.

2.3. 연구 절차

녹음에 참여한 학생들은 한 학기 총 15주간 진행된 수업에서 첫 4주는 영어 발음의 원리에 대한 강의를 한 주 세 시간씩 수강한 후, 이후 11주 동안 연구자와의 일대일 발음 훈련에 참여하였다. 강의에서는 영어 자·모음의 조음, 억양구(intonational phrase), 단어 강세(word stress), 문장 강세(sentence stress), 핵강세(nuclear accent), 영어의 리듬, 리듬과 강세에 따른 모음의 약화와 연음, 맥락과 의미에 따른 억양의 변화에 대해 수업을 진행하였다. 일대일 발음 훈련에서는 학생들이 일 인당 매주 약 10분씩 정해진 글을 읽으면, 연구자가 분절음, 리듬, 억양 등에 대한 피드백을 주었고, 다음 주에 피드백의 내용이 잘 반영되었는지 확인하였다. 연구에 사용된 녹음 파일은 수업 첫 주에 사전 녹음을, 마지막 주에 사후 녹음을 해 제출하도록 하였다.

청취 평가를 위해서 Praat 스크립트를 이용해 청취 실험 도구를 제작하였다. 녹음된 전체 문장을 들려주고 청취 평가를 진행하는 것이 가장 이상적이지만, 평가자의 피로도를 고려하여 전체 발화 중 비교적 그 길이가 긴 두 번째 문장만을 들려주고, 리듬과 유창성을 9개의 리커트 척도를 사용해 점수가 높을수록 리듬 및 유창성이 좋은 것으로 평가하게 하였다. ‘리듬’은 발화의 장단과 강세 및 비강세 모음이 얼마나 잘 구현되는지 집중하여 평가하도록 했고, ‘유창성’은 얼마나 막힘없이 자연스러운 속도로 발화하는지 평가하도록 했다. 한 사람이 리듬 평가와 유창성 평가를 위해 각각 34개, 총 68개의 발화를 평가하였고, 발화는 Praat의 ‘PermuteBalancedNoDoublets’ 기능을 사용해 사전, 사후 구분 없이 임의의 순서로 들려주었다. 자극은 한 번씩만 제시하였지만, 평가 도구 내에서 평가자가 필요할 경우 하나의 자극을 세 번까지 다시 들을 수 있도록 하였다. 리듬과 유창성 평가는 동시에 진행하지 않고 각각 별도로 진행하였다.

녹음 파일은 발화 전체를 FAVE-align 방식을 사용해 음소 층위와 단어 층위로 강제 정렬한 후, 정확한 분석을 위해서 모든 부분을 연구자가 눈으로 확인하고 정렬을 수정하였다. 발화 중 말실수나 머뭇거림이 있는 경우에는 모두 휴지(sp)로 수정 입력하였다. 정렬을 완료한 후 Praat 스크립트와 Python을 활용해 리듬 및 유창성 지수를 추출하였다. 리듬 지수로는 선행 연구에 사용된 지수 중 연구에서 자주 언급되는 10개 지수를 추출하였는데, %V, ΔV, ΔC, VarcoV, VarcoC, nPVIC, rPVIC, nPVIV, rPVIV, 기능어 비율을 추출하였다. 유창성 지수는 휴지를 포함한 전체 발화 길이, 휴지를 제외한 전체 발화 길이, 음절 수, 휴지 전체 길이, 기능어 전체 길이를 먼저 계산한 후, 발화 속도(speech rate), 조음 속도(articulation rate), 휴지 개수, 휴지 비율

등 4개 지수를 추출하였다. 이 중 휴지 비율은 선행 연구에서는 잘 다루지 않은 지수이다. 기능어 비율은 유창성 지수에서 도출해야하기 때문에 추출한 후 리듬 지수에 포함하였다.

청취 평가와 리듬 및 유창성 지수를 사전과 사후로 나누어, 우선 사전 청취 평가와 사후 청취 평가에 통계적으로 유의한 차이가 있는지 살펴보고, 이후 리듬 및 유창성 지수와 리듬 및 유창성 청취 평가의 관련성을 분석하였다. 사전, 사후의 차이를 분석하기 위해서 대응 표본 *t*-검정과 Wilcoxon signed-rank 검정을 실시하였고, 관련성 분석을 위해서는 선형 혼합 효과(linear mixed effect) 분석을 실시하였다. 통계 분석은 R(R Core Team, 2022)을 이용해 진행하였다.

3. 연구 결과

3.1. 청취 평가 분석

평가 참여자 7명의 평가자 간 신뢰도 분석을 위해서 ‘psych’ 패키지(Revelle, 2022)에 포함된 ‘cohen.kappa’ 기능을 활용하였다. 리듬의 Kappa 계수는 0.2, 유창성 평가의 Kappa 계수는 0.3으로 매우 저조한 신뢰도를 보여주었다. 하지만, 재평가를 할 수 없어서 평가 결과를 그대로 사용하기로 하였고, 7명의 평균 값을 평가 점수로 입력하였다. 평가 점수의 사전, 사후, 전체 점수에 대한 기술 통계는 표 3과 같고, 그림 1은 박스 플롯으로 시각화한 것이다.

기술 통계를 보면 리듬 평가 점수는 사전 평균 점수에 비해 사후 점수가 향상되었지만, 유창성 평가 점수는 미미하게 감소하였음을 알 수 있다. Shapiro-Wilk 검정 결과 리듬 평가(사전: $W=0.939, p=.301$; 사후: $W=0.970, p=.815$)와 유창성 평가(사전: $W=0.949, p=.437$; 사후: $W=0.949, p=.445$) 모두에서 정규성을 충족해 대응 표본 *t*-검정을 실시하여, 발음 수업 전후 점수 간에 유의미한 차이가 있는지 분석하였다. 분석 결과 영어 발음 수업 이전의 리듬 평가 점수(평균=5.89; 표준 편차=1.03)와 수업 이후의 리듬 평가 점수(평균=6.24; 표준 편차=1.13)에는 유의미한 차이가 없는 것으로 나타났다[$t(16)=-1.241, p=.233$]. 유창성 점수 분석 결과에서도 영어 발음 수업 이전의 유창성 평가 점수(평균=6; 표준 편차=1.08)와 수업 이후의 유창성 평가 점수(평균=5.91; 표준 편차=1.26)에는 유의미한 차이가 없는 것으로 나타났다[$t(16)=.307, p=.763$].

표 3. 리듬 및 유창성 청취 평가 기술 통계

Table 3. Descriptive statistics of rhythm and fluency ratings

평가	차수	평균	표준편차
리듬	사전	5.89	1.03
	사후	6.24	1.13
	전체	6.07	1.08
유창성	사전	6	1.08
	사후	5.91	1.26
	전체	5.95	1.16

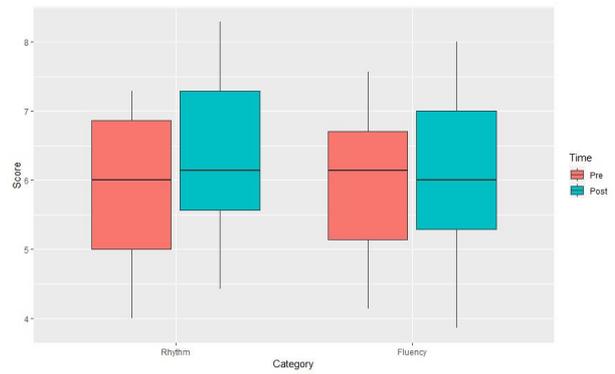


그림 1. 리듬 및 유창성 청취 평가 박스 플롯
Figure 1. Boxplot of rhythm and fluency ratings

3.2. 리듬 및 유창성 지수 분석

리듬 및 유창성 지수 분석을 위해 선행 연구에서 언급된 지수 중 자주 사용되는 14개 지수에 대해 각 지수 별로 값을 추출하였고 추출한 지수의 기술 통계는 표 4와 같다.

Shapiro-Wilk 검정 결과 VarcoC(사전: $W=.845, p=.009$; 사후: $W=.926, p=.184$)와 휴지 비율(사전: $W=.891, p=.048$; 사후: $W=.956, p=.562$), 기능어 비율(사전: $W=.805, p=.002$; 사후: $W=.966, p=.742$)에서는 정규성을 충족하지 못해 ‘coin’ 패키지(Hothorn et al., 2008)에 포함된 비모수 통계인 Wilcoxon signed-rank 검정을 실시하고, 그 외 지수는 정규성을 충족해 대응 표본 *t*-검정을 실시하였다.

표 4. 리듬 및 유창성 지수 기술 통계
Table 4. Descriptive statistics of rhythm and fluency indices

지수	차수	평균	표준편차
%V	사전	40.51	2.67
	사후	41.53	2.41
	전체	41.02	2.56
ΔV	사전	65.23	11.23
	사후	69.15	10.78
	전체	67.19	11.02
ΔC	사전	81.55	10.86
	사후	77.92	8.69
	전체	79.73	9.86
VarcoV	사전	62.44	8.11
	사후	66.49	10.07
	전체	64.47	9.23
VarcoC	사전	56.62	4.96
	사후	56.66	4.45
	전체	56.64	4.64
nPVIC	사전	65.35	4.71
	사후	65.67	3.27
	전체	65.51	3.99
rPVIC	사전	92.54	11.77
	사후	90.3	9.03
	전체	91.42	10.39
nPVIV	사전	66.04	7.13
	사후	64.37	6.44
	전체	65.2	6.74
rPVIV	사전	72.4	12.41
	사후	72.79	11.93
	전체	72.59	11.99
발화 속도	사전	3.19	.44
	사후	3.17	.38
	전체	3.18	.4
조음 속도	사전	3.94	.42
	사후	4.04	.38
	전체	3.99	.4
휴지 개수	사전	18	3.95
	사후	17.18	3.54
	전체	17.59	3.72
휴지 비율	사전	19.14	5.34
	사후	21.69	4.84
	전체	20.42	5.18
기능어 비율	사전	26.71	1.83
	사후	26.46	1.26
	전체	26.58	1.55

표 5. 'Model1h'의 'R' 코드 예시
Table 5. Sample R code for modeling

```

model1h = lmer(RhythmScore ~ pctV+VarcoC+nPVIC
+rPVIC+nPVIV+SR+AR+NumSil+pctSil+pctFW
+(1|File), data=wind_wide)
nPVIV_R <- summary(model1h)
capture.output(nPVIV_R, file = "model1h")

```

분석 결과 영어 발음 수업 이전의 VarcoV(평균=62.44; 표준편차=8.11)와 수업 이후의 VarcoV(평균=66.49; 표준편차=10.07)에 유의미한 차이가 있는 것으로 나타났다($t(16)=3.14, p=.006$). 대응 표본 t -검정을 한 지수 중 그 외 지수에는 사전, 사후 평균 지수 간에 통계적으로 유의미한 차이가 있는 경우는 없

었다. Wilcoxon signed-rank 검정을 실시한 VarcoC와 휴지 비율, 기능어 비율에서도 통계적으로 유의미한 차이가 없었다.

3.3. 리듬 및 유창성 지수와 청자 평가의 관련성

리듬 및 유창성 지수와 청자의 리듬과 유창성 평가의 관련성을 분석하기 위해 'lme4' 패키지(Bates et al., 2015)와 'lmerTest' 패키지(Kuznetsova et al., 2017), 'nlme' 패키지(Pinheiro et al., 2022)를 사용해 선형 혼합 효과(linear mixed effect) 분석을 실시하였다. 상관관계가 큰 지수를 분석 모델에서 제거하기 위해 다중선형분석(multicollinearity)을 실행하였다. 상관관계가 0.5 이상인 지수는 모델에 포함되지 않도록 제거한 후, 지수와 리듬 및 유창성 평가 점수 간의 관련성 분석을 위해 각각 13개씩 26개의 모델을 구성하였다.

표 5는 모델을 구축하기 위한 'R' 코드 중 'model1h'의 예시이다. 모델의 통계 결과는 'nPVIV_R'이라는 파일로 저장하였다. 선형 혼합 효과 분석 결과 26개 모델 중 리듬 평가에 대해서 8개의 모델에서 유의미한 지수의 기여가 발견되었고, 유창성 평가에 대해서는 9개의 모델에서 유의미한 지수의 기여가 발견되었다. 개별 모델 중 어떤 모델이 평가 점수를 예측하는데 상대적으로 더 적합한지 비교하기 위해서 선형 혼합 효과 분석 결과 추출된 AIC(Akaike information criterion)와 BIC(Bayes information criterion) 중, BIC를 사용하였다. BIC는 AIC에 비해 독립 변수의 수가 많을수록 더 강력한 불이익을 주는 구조이기 때문에, 더 효율적인 모델을 구축하기 유리하기 때문이다. BIC의 값이 낮을수록 더 적합성이 좋은 모델이다(Chakrabarti & Ghosh, 2011). 표 6과 표 7은 리듬 평가와 유창성 평가에 대한 각 모델의 적합성을 BIC와 AIC를 통해 보여주고 있다. 순서는 BIC가 낮은 것부터 오름차순으로 배열하였다.

각 모델별로 어떤 지수가 유의미한 영향을 미치고 있는지, 그 추정치(estimate)는 얼마인지 알아보기 위해 리듬 평가와 유창성 평가 각각에 대해 적합성이 가장 좋은 4개의 모델을 그림 2와 그림 3과 같이 플롯으로 표현하였다.

표 6. 리듬 평가에 대한 모델의 적합성
Table 6. Model fit on rhythm ratings

모델	BIC	AIC
Model1h	114.75	94.912
Model1d	115.26	96.945
Model1e	116.40	96.554
Model1b	117.64	100.849
Model1i	118.39	103.126
Model1f	118.65	95.757
Model1c	121.35	104.563
Model1a	124.98	99.032

표 7. 유창성 평가에 대한 모델의 적합성
Table 7. Model fit on fluency ratings

모델	BIC	AIC
Model2f	102.72	79.830
Model2i	104.54	89.273
Model2b	107.02	90.225
Model2j	108.82	87.451
Model2a	108.88	82.928
Model2e	109.50	89.652
Model2l	110.76	87.866

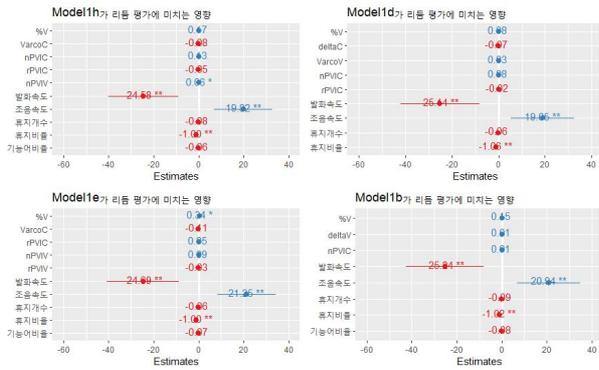


그림 2. 리듬 평가 모델의 적합도 플롯
Figure 2. Plots of the model fit on rhythm ratings

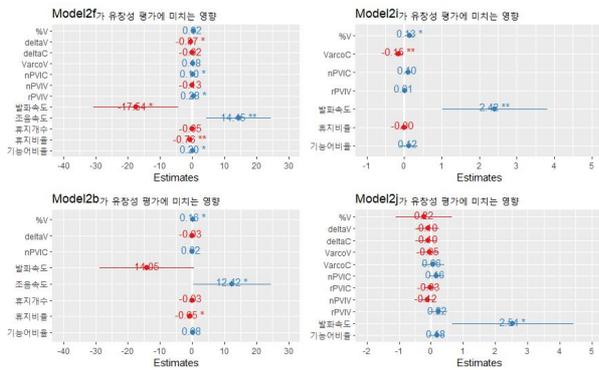


그림 3. 유창성 평가 모델의 적합도 플롯
Figure 3. Plots of the model fit on fluency ratings

리듬 평가에 대한 예측 모델 중 model1h에서는 휴지 비율($p=.004$), 조음 속도($p=.004$), 발화 속도($p=.004$), nPVIV($p=.036$) 등 4개의 지수가 유의미한 기여를 하는 것으로 나타났다. Model1d에서는 휴지 비율($p=.004$), 조음 속도($p=.008$), 발화 속도($p=.005$) 등 3개의 지수가, model1e에서는 휴지 비율($p=.005$), 조음 속도($p=.003$), 발화 속도($p=.005$), %V($p=.011$) 등 4개의 지수가, model1b에서는 휴지 비율($p=.007$), 조음 속도($p=.005$), 발화 속도($p=.006$) 등 3개의 지수가 유의미한 기여를 하는 것으로 나타났다.

유창성 평가에 대한 예측 모델 중 model2f에서는 기능어 비율($p=.057$), 휴지 비율($p=.013$), 조음 속도($p=.009$), 발화 속도($p=.019$), rPVIV($p=.026$), nPVIC($p=.041$), ΔV ($p=.046$) 등 7개의 지수가,

model2i에서는 발화 속도($p=.002$)와 VarcoC($p=.009$), %V($p=.046$) 등 3개 지수, model2b에서는 휴지 비율($p=.038$), 조음 속도($p=.044$), %V($p=.031$) 등 3개 지수, model2j에서는 발화 속도($p=.011$)만이 유의미한 기여를 하는 것으로 나타났다.

4. 논의 및 결론

청취 평가 결과를 보면 우선 평가 참여자 간의 신뢰도가 높지 않다는 것은 그만큼 청취 평가가 쉽지 않다는 것을 보여주고 있다. 사전에 리듬과 유창성의 개념에 대해서 분명히 설명하였음에도 불구하고, 평가자 신뢰도가 높지 않았고, 사후 평가자 면담에서도 평가가 쉽지 않았다는 의견을 들을 수 있었다. 평가자들 간의 신뢰도를 높이기 위해서는 훈련 발화를 통해 평가 훈련이 충분히 이루어져야 할 것이다.

리듬 및 유창성 평가 점수 간의 사전, 사후 평균 점수를 비교했을 때, 통계적으로 유의미한 차이를 발견하지 못하였다. 이것은 영어 발음 연습 수업을 한 학기, 총 15주 동안 진행하였지만, 리듬과 유창성 향상에는 크게 기여하지 못하였음을 보여주고 있다. 발화 분석을 통해 도출된 객관적인 리듬 및 유창성 지수 또한 VarcoV를 제외하고는 사전, 사후 평균 점수 간에 통계적으로 유의미한 차이가 없어서 평가자들과 유사한 경향을 보이고 있다.

하지만, 주목해야 할 점은 리듬 및 유창성 지수가 선행 연구에서 영어 원어민들의 지수를 분석한 결과에 비해 못하지 않다는 것이다. 표 8은 이 연구의 지수값과 선행 연구에서 추출된 값을 비교한 것이다.

표 8. 리듬 지수 비교
Table 8. Comparison of rhythm index values

지수	연구	본연구	1	2	3	4	5
%V		41.02	40.1	38.2	-	38	-
ΔV		67.19	46.4	45.2	-	49	40 초반
ΔC		79.73	53.5	-	-	59	50 후반
VarcoV		64.47	-	-	-	64	-
VarcoC		56.64	-	-	-	47	-
nPVIV		65.2	-	79	57.2	73	50 후반
rPVIV		72.59	-	-	-	-	60 후반
발화 속도		3.18	-	-	-	5.2	-

위 표에서 ‘1’은 Ramus et al.(1999), ‘2’는 Ling et al.(2000), ‘3’은 Grabe & Low(2002), ‘4’는 White & Mattys(2007), ‘5’는 Mariano & Romano(2011)를 가리킨다. Grabe & Low(2002)와 Mariano & Romano(2011)는 본 연구에서 사용한 것과 동일한 지문을 읽은 것을 분석한 결과이다. Ramus et al.(1999)에서는 이 연구와 동일한 지문을 읽은 것이 아니라, 뉴스체의 문장을 읽고 분석한 것이고, Ling et al.(2000)에서도 동일한 지문을 읽지는 않고, 짧은 문장 10개를 영국 화자들이 읽은 것을 분석한 것이다. White & Mattys(2007)도 중간 정도 길이의 문장 5개를 읽은 것을 분석한 결과이다. Mariano & Romano(2011)에는 정확한 지

수의 값이 제시되지 않고 그림의 그래프에 표시만 되어 있어서 그림을 보고 근사치를 제시하였다.

위 표를 보면 %V의 경우 본 연구와 다른 연구의 값이 비교적 낮은 수준에서 비슷한 수치를 보이고 있고, ΔV 와 ΔC 는 본 연구의 수치가 상당히 높은 것을 알 수 있다. VarcoV는 거의 동일하고, VarcoC는 다소 높으며, nPVIV의 경우 중간 정도의 수준이고, rPVIV는 다소 높고, 발화 속도는 많이 느린 것으로 분석할 수 있다. 발화 속도를 제외하고 비교가 가능한 모든 지수에서 본 연구의 지수값이 더 높거나 중간 정도의 수준인 것을 알 수 있다. 이미 대부분의 지수에서 참여자들의 발화가 원어민과 다르지 않은 수준을 보이고 있었지만, 평가자들은 ‘9’점 척도 중 ‘6’점 내외의 평균 점수를 준 것으로 볼 때, 상당히 엄격하게 평가한 것으로 볼 수 있다.

리듬 및 유창성 지수 중 어떤 요소들을 결합한 모델을 구성하여 청취 평가 점수를 잘 예측할 수 있는지 분석했을 때, 리듬 평가 예측 모델에서는 휴지 비율, 조음 속도, 발화 속도, 기능어 비율이, 유창성 평가 예측 모델에서는 이에 더해 %V, ΔV , VarcoC, nPVIC, rPVIV 등이 유의미한 기여를 하는 것으로 분석되었다. 기존 선행 연구에서 예측력이 높다고 알려진 nPVIV와 %V, VarcoC 등은 청취 평가를 예측하는 데 있어서는 일부 모델에서만 유의미하게 기여하는 것으로 나타났다.

본 연구의 평가자들은 리듬과 유창성을 평가할 때 모음 구간이나 자음 구간의 변이 지수 등 리듬 지수보다는 휴지 비율, 조음 속도, 발화 속도 등의 유창성 지수에 더 민감하게 반응해 평가하는 것으로 판단할 수 있다.

본 연구의 결과가 시사하는 영어 교육적 함의는 일정 수준의 대학생들은 이미 지수 측면에서 영어 원어민과 비슷한 정도의 리듬과 유창성을 보이고 있다는 사실과, 리듬과 유창성을 더 향상시키기 위해서는 유창성 요소인 휴지, 조음 속도, 발화 속도를 최적의 상태로 구현할 수 있도록 교육하는 것이 바람직하다는 것이다. 적절한 끊어 읽기와 속도 조절을 할 수 있는 활동을 통해 리듬과 유창성을 향상시킬 수 있을 것이다.

또한, 영어 발화의 평가를 위해서는 지속적이고 반복적인 평가자 훈련이 필요하고, 평가자 훈련을 통해 리듬과 유창성의 개념과 척도에 관한 합의가 이루어질 수 있어야 한다.

본 연구가 가지고 있는 제한점은, 가장 중요한 평가자 간 신뢰도 확보가 이루어지지 못해 연구의 결과를 정당화하거나 일반화하기는 어렵다는 점이다. 영어 원어민 화자의 평가 결과와 비교해 보지 못한 것도 이 연구의 아쉬운 점이다. 또, 음성 자료의 양이 17명에 한정되어 있다는 점이다. 하지만, 다른 연구들이 짧거나 중간 정도 길이의 문장을 읽은 연구가 많은 반면, 이 연구에 사용된 ‘The North Wind and the Sun’은 아주 길이가 긴 다섯 개의 문장으로 구성된 문단이어서, 지수 측면에서는 일반화에 무리가 없을 것으로 판단된다. 또, 리듬 지수 중 ‘VI’는 음절 구분을 반자동화하는 데 어려움이 있어서, 이번 연구의 분석 대상에 포함시키지는 못했다. 추후 이번 연구에서 미진한 부분을 보완해 추가 분석을 실시할 예정이다.

감사의 글

본 연구에서 사용된 FAVE-align 방식의 강제 정렬 도구는 성신여자대학교 윤태진 교수님의 도움으로 만든 것입니다. 도구를 제공해 주신 윤태진 교수님께 감사드립니다.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer (version 6.2.17) [Computer program]. Retrieved from <https://www.fon.hum.uva.nl/praat/>
- Chakrabarti, A., & Ghosh, J. K. (2011). AIC, BIC and recent advances in model selection. In P. S. Bandyopadhyay, & M. R. Forster (Eds.), *Philosophy of statistics: Handbook of the philosophy of science* (pp. 583-605). Amsterdam, Netherlands: Elsevier.
- Choe, W. K. (2019). The realization of English rhythm by Busan Korean speakers. *Phonetics and Speech Sciences*, 11(4), 81-87.
- Chung, H. (2013). A study on the rhythm of Korean English learners' interlanguage talk. *Phonetics and Speech Sciences*, 5(3), 3-10.
- Chung, H., Jang, T. Y., Yun, W., Yun, I., & Sa, J. (2008). A study on automatic measurement of pronunciation accuracy of English speech produced by Korean learners of English. *Language and Linguistics*, 42, 165-196.
- Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.), *Language and language processing: Proceedings of the 38th Linguistic Colloquium* (pp. 231-241). Frankfurt, Germany.
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54(4), 655-679.
- Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29(2), 217-230.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 515-546). Berlin, Germany: De Gruyter Mouton.
- Hothorn, T., Hornik, K., van de Wiel, M. A., & Zeileis, A. (2008). Implementing a class of permutation tests: The coin package. *Journal of Statistical Software*, 28(8), 1-23.
- Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10(2), 135-159.
- Kim, M. S., & Jang, T. Y. (2019). Pauses and speech rates in assessing fluency of English speech. *Korean Journal of Linguistics*, 44(3),

315-339.

- Kim, S. (2021). Foreign accentedness rating and rhythm in L2 English. *The Mirae Journal of English Language and Literature*, 26(4), 101-120.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1-26.
- Ling, L. E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43(4), 377-401.
- Mariano, P., & Romano, A. (2011). Rhythm metrics for 21 languages. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1318-1321). Hong Kong, China.
- Pinheiro, J., Bates D., & R Core Team, (2022). nlme: Linear and nonlinear mixed effects models (version 3.1-157) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://cran.r-project.org/web/packages/nlme/>
- R Core Team (2022). R: A language and environment for statistical computing (version 4.2.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.
- Revelle, W. (2022). psych: Procedures for personality and psychological research (version 2.2.9) [Computer software]. Evanston, IL: Northwestern University. Retrieved from <https://CRAN.R-project.org/package=psych>
- The International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the international phonetic alphabet*. Cambridge, UK: Cambridge University Press.
- Thomson, R. I. (2015). Fluency. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 209-226). Chichester, UK: John Wiley & Sons.
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35(4), 501-522.

- 정현성 (Hyunsong Chung) 교신저자
한국교원대학교 영어교육과 교수
충북 청주시 흥덕구 강내면 태성탑연로 250
Tel: 043-230-3554
Email: hchung@knue.ac.kr
관심분야: 실험음성학, 영어발음교육, 파닉스 교육

리듬 및 유창성 지수와 한국 화자의 영어 읽기 발화 청취 평가의 관련성

정 현 성

한국교원대학교 영어교육과

국문초록

이 논문은 리듬 및 유창성 지수와 한국 대학생이 읽은 영어 발화의 리듬 및 유창성 청취 평가가 어떤 관련성을 가지는지 분석한 연구이다. 이를 위해 대학생 17명이 'The North Wind and the Sun' 문단을 읽고 영어 발음 수업 사전, 사후 두 차례에 걸쳐 녹음한 것을, 대학원에 재학 중인 현직 영어 교사 및 예비 영어 교사 7명이 리듬 및 유창성에 대한 청취 평가를 진행하였다. 또, 선행 연구에서 언급된 리듬 및 유창성 지수 중 14개 지수를 사용해 각 발화 자료의 지수를 추출하여 지수와 청취 평가의 관련성을 분석하였다. 지수를 분석한 결과 %V, VarcoV, nPVIV에서는 선행 연구의 영어 원어민 지수와 거의 비슷한 양상을 보였고, ΔV , ΔC , VarcoC에서는 원어민보다 높았으며, 발화 속도는 원어민보다 느렸다. 한국 대학생들이 일부 리듬 지수에서는 영어 원어민과 비슷한 양상을 보인다고 할 수 있다. 리듬 및 유창성 지수를 사용하여 리듬 평가 점수와 유창성 평가 점수를 예측할 수 있는 최적의 모델을 탐색한 결과, 리듬 및 유창성 평가 모두에서 유창성 지수인 휴지 비율, 조음 속도, 발화 속도 등이 리듬 지수들보다 평가 점수를 예측하는데 더 큰 기여를 하는 것으로 나타났다.

핵심어: 리듬, 유창성, 리듬 지수, 유창성 지수

참고문헌

- 정현성(2013). 타언어 화자와의 담화 상에 나타난 한국인 영어 학습자의 리듬. *말소리와 음성과학*, 5(3), 3-10.
- 정현성, 장태엽, 윤원희, 윤일승, 사재진(2008). 한국인 영어 학습자의 발음 정확성 자동 측정방법에 대한 연구. *언어와 언어학*, 42, 165-196.