



Automatic severity classification of dysarthria using voice quality, prosody, and pronunciation features*

Eun Jung Yeo¹ · Sunhee Kim² · Minhwa Chung^{1,**}

¹Department of Linguistics, Seoul National University, Seoul, Korea

²Department of French Language Education, Seoul National University, Seoul, Korea

Abstract

This study focuses on the issue of automatic severity classification of dysarthric speakers based on speech intelligibility. Speech intelligibility is a complex measure that is affected by the features of multiple speech dimensions. However, most previous studies are restricted to using features from a single speech dimension. To effectively capture the characteristics of the speech disorder, we extracted features of multiple speech dimensions: voice quality, prosody, and pronunciation. Voice quality consists of jitter, shimmer, Harmonic to Noise Ratio (HNR), number of voice breaks, and degree of voice breaks. Prosody includes speech rate (total duration, speech duration, speaking rate, articulation rate), pitch (F0 mean/std/min/max/med/25quartile/75 quartile), and rhythm (%V, deltas, Varcos, rPVI, nPVI). Pronunciation contains Percentage of Correct Phonemes (Percentage of Correct Consonants/Vowels/Total phonemes) and degree of vowel distortion (Vowel Space Area, Formant Centralized Ratio, Vowel Articulatory Index, F2-Ratio). Experiments were conducted using various feature combinations. The experimental results indicate that using features from all three speech dimensions gives the best result, with a 80.15 F1-score, compared to using features from just one or two speech dimensions. The result implies voice quality, prosody, and pronunciation features should all be considered in automatic severity classification of dysarthria.

Keywords: dysarthria, automatic severity classification, speech dimensions, machine learning, feature selection

1. 서론

마비말장애는 중추 신경계 및 자율 신경계의 손상으로 말소리 산출과 관련된 근육의 마비 및 약화로 나타나는 말장애이다 (Darley et al., 1969). 마비말장애 진단 및 치료 시 주로 사용되는 지표는 말 명료도(speech intelligibility)이다. 언어재활사는 말 명

료도 평가를 통해 화자가 의사소통 상황에서 얼마나 어려움을 겪고 있는지 확인하고, 그에 따라 알맞은 중재 방안을 모색한다 (Lee et al., 2012). 그러나 임상에서 주로 사용되는 말 명료도 평가는 청지각적 평가로, 주관적일 뿐만 아니라 많은 시간과 노력이 소요된다. 반면 마비말장애 중증도 자동 분류 기술은 추출된 음성 특징을 바탕으로 음성 질환을 진단 또는 분류한다. 해당

* This research is supported by Ministry of Culture, Sports and Tourism and Korea Creative Content Agency(Project Number:R2019080018)

** mchung@snu.ac.kr, Corresponding author

Received 1 June 2021; Revised 14 June 2021; Accepted 14 June 2021

© Copyright 2021 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.

기술이 숙련된 언어재활사의 말 명료도 평가와 유사한 결과를 산출한다면, 객관적, 일관적 결과를 제공함으로써 언어재활사를 보조할 수 있을 것이다(Hernandez et al., 2020a; Kim et al., 2015; Narendra & Alku, 2021). 더 나아가 현재까지 국내에는 마비말장애 진단 목적의 표준화 도구가 없으며, 어떤 평가 요소가 한국어를 사용하는 마비말장애 화자의 말 특성을 대표하는지에 대한 연구는 아직 부족한 실정이다(Hong et al., 2018). 중증도 자동 분류에서 유의미하게 사용되는 특징에 대한 분석은 한국어를 사용하는 마비말장애 화자의 말 특성 연구에도 중요한 통찰력을 제공할 수 있다.

최근 딥러닝 기술이 발전함에 따라 마비말장애 중증도 자동 분류 연구 분야에서도 딥러닝 방법론이 적용되고 있다. 이는 MFCCs(mel frequency cepstral coefficients), log filter bank와 같은 기본적인 음성 특징을 심층 신경망의 입력값으로 사용하는 방법이다(Bhat & Strik, 2020; Janbakhshi et al., 2019). 이 방법론은 데이터 준비 과정이 간단하다는 장점이 있지만, 각 집단 별로 대량의 데이터가 필요할 뿐만 아니라 분류 결과에 대한 해석이 직관적이지 않아 해석을 위한 후처리 과정이 필요하다는 단점이 있다. 반면 음성 특징을 추출하여 머신러닝 분류기의 입력값으로 사용하는 방법도 있다. 이 방법론에서는 분류 집단 간 차이를 반영하는 최적의 특징 셋을 찾는 것이 중요하다. 본 연구의 목적은 마비말장애 중증도 분류를 위한 최적의 특징 셋을 탐색하는 것으로, 두 번째 방법론을 사용한다.

선행연구에서는 마비말장애 중증도 자동 분류를 위해 말 명료도에 영향을 미친다고 알려진 특징들을 중심으로 특징 셋을 구성하였다. 사용된 특징들로는 음질(Narendra & Alku, 2021), 운율(Hernandez et al., 2020a; Hernandez et al., 2020b; Kadi et al., 2013), 조음 관련 특징(Hernandez et al., 2020; Kim et al., 2015; Lansford & Liss, 2014)이 있다. 이처럼 대부분의 선행연구는 한 가지 말 기능 특징에 집중하여 특징 셋을 구성하였다. 그러나 말 명료도는 호흡, 발성, 조음, 공명, 운율 다섯 가지의 말 기능 특징의 복합적인 영향을 받는 척도로(Hong et al., 2018), 한 가지의 말 기능 특징만으로는 충분히 설명되기 어렵다. 반면 Kim & Kim(2012)은 말 명료도 예측을 위해 음성의 장애 특성을 음성의 질, 운율, 음질로 구분하여 다양한 말 기능 특징을 반영하고자 했다. 이때 음성의 질은 조음을 반영하는 특징을 포함한다. 그러나 해당 연구에서 사용된 특징은 임상 현장에서 사용되는 특징들과 거리가 있어 결과에 대한 직관적인 해석이 어렵다는 한계가 있다.

말 명료도는 다양한 말 기능 특징의 종합적인 영향을 받는다. 하지만 대부분의 선행연구는 한 가지 말 기능 특징에 초점을 맞추어 음성의 장애 특성을 제한적으로 반영하였다. 본 논문에서는 음질, 운율, 발음 세 가지 말 기능 특징 조합을 사용하여 장애 특성을 보다 다면적으로 포착하고자 한다. 더 나아가, 중증도 자동 분류의 특징 셋으로 임상 현장에서 마비말장애 중증도 분류 시 사용되는 특징들을 사용함으로써, 청지각적 기준인 말 명료도를 더욱 잘 반영하고, 실험 결과에 대한 직관적 해석을 제공하고자 한다.

논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안하는 방법론을 소개하고, 3장에서는 실험에 사용된 데이터베이스와 통계분석을 제시한다. 4장에서는 실험 결과를 요약하고, 5장에서는 본 연구의 한계점과 향후 진행할 연구를 소개하며 마무리한다.

2. 방법론

2.1. 중증도 자동 분류

본 연구는 세 가지 말 기능 특징인 음질, 운율, 발음 특징의 다양한 조합을 중증도 자동 분류에 사용할 것을 제안한다. 먼저 기본 특징인 MFCCs 외에, 다양한 말 기능 특징을 반영하는 음질, 운율, 발음과 관련된 특징들을 추출한다. 그 다음, 특징 선택 과정을 거쳐 추출된 특징들 중 분류에 도움이 되는 특징들만으로 간추린다. 이때 MFCCs는 음성의 기본 특성으로 특징 선택 과정에서 제외한다. 마지막으로, MFCCs와 선택된 특징들을 머신러닝 분류기의 입력값으로 넣어준다. 분류기는 입력된 특징들을 바탕으로 음성을 비장애, 경도, 경도-중등도, 중등도-중도, 중도 다섯 가지의 중증도로 분류한다. 분류기의 성능은 F1-score로 평가한다. 그림 1은 중증도 자동 분류 개요도이다.

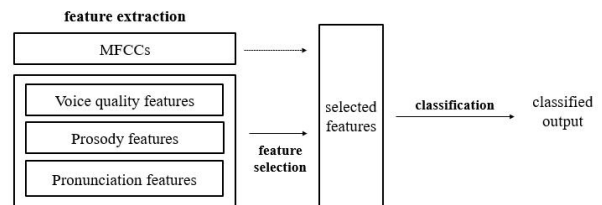


그림 1. 실험 개요도
Figure 1. Design of the experiment

2.2. 특징 추출

2.2장에서는 실험에서 사용된 특징들을 설명한다. 전체 특징 리스트는 표 1에 제시하였다.

2.2.1. MFCCs (Mel frequency cepstral coefficients)

MFCCs는 소리의 고유한 특징을 나타내는 값으로, 음성인식, 음성합성 등 오디오 도메인에서 기본적으로 사용하는 음성 특징이다. MFCCs는 특히 조음 특징을 반영하는 것으로 알려져 있으며, 장애 자동 진단 및 분류 연구에서도 기본 특징으로 사용되어 왔다(Bhat & Strik, 2020; Hernandez et al., 2020, Janbakhshi et al., 2019). 해당 연구에서 MFCCs는 librosa(McFee et al., 2015)로 추출했으며, 발화 당 프레임의 평균값을 사용하였다.

2.2.2. 음질 특징

본 연구에서는 음질 특징으로 jitter, shimmer, harmonic to noise ratio(HNR), voice breaks 개수, voice breaks 정도를 사용하였다. Jitter, shimmer, HNR은 언어 병리학 분야에서 음성 장애를 진단

할 때, voice breaks 관련 특징들은 화자의 음성 유지 능력을 확인할 때 사용되는 특징이다. 모든 음질 특징은 Praat(Boersma & Weenink, 2001)로 추출하였다.

표 1. 특징 리스트
Table 1. Feature list

특징 종류	특징	
스펙트럼	Mel frequency cepstral coefficients (MFCCs)	
음질	Jitter, shimmer, harmonic to noise ratio (HNR), voice breaks 개수, voice breaks 정도	
운율	발화 속도	전체 길이, 말 길이, 말 속도, 조음 속도
	음높이	F0 평균값/표준편차/최솟값/최댓값/중앙값/25분위수/75분위수
	리듬	%V, deltas, Varcos, rPVIs, nPVIs
발음	음소 정확도	자음 정확도(PCC), 모음 정확도(PCV), 전체 음소 정확도(PCT)
	모음 왜곡도	Vowel space area (VSA), Formant centralized ratio (FCR), Vowel articulatory index (VAI), F2-ratio

PCC, percentage of correct consonants; PCV, percentage of correct vowels; PCT, percentage of total correct phonemes.

2.2.3. 운율 특징

본 연구에서는 운율 특징을 발화 속도, 음높이, 리듬으로 세분화하였다. 특징들은 마비말장애 중증도 분류 과제에서 운율 특징의 다양한 조합을 살펴본 Hernandez et al.(2020)에서 사용한 특징 셋을 참고하였다.

발화 속도 특징으로 전체 길이(total duration), 말 길이(speaking duration), 말 속도(speaking rate), 조음 속도(articulation rate)를 사용하였다. 전체 길이는 음성의 길이를, 말 길이는 전체 길이 중 휴지 구간을 제외한 구간을 의미한다. 말 속도는 전체 음절 수를 전체 길이로 나눈 값을, 조음 속도는 전체 음절 수를 말 길이로 나눈 값을 의미한다. 발화 속도 특징들은 모두 Parselmouth (Jadoul et al., 2018)를 사용하여 추출하였다.

음높이 특징으로 F0의 평균값, 표준편차, 최솟값, 최댓값, 중앙값, 25 분위수, 75 분위수를 사용하였다. 음높이 특징들은 모두 Praat로 추출하였다.

리듬 특징은 %V, deltaV, deltaC, Varco-V, Varco-C, VrPVI, CrPVI, VnPVI, CnPVI가 사용되었다. 리듬의 경우, 특정한 음향학적 특징이 없으므로 리듬 요소와 강한 상관관계를 보이는 것으로 알려진 특징들을 사용하였다(Dellwo & Wagner, 2003). %V는 음성 전체 중 모음이 차지하는 비율을, deltaV와 deltaC는 델타값으로 각각 모음 또는 자음 길이의 표준편차를 의미한다. Varco-V와 Varco-C는 발화 속도에 영향을 받는 델타값들을 정규화하기 위해 제안된 값으로, 각각 델타값을 모음 또는 자음의 평균 길이로 나눈 값이다. PVI(pairwise variability index)는 rPVI와 nPVI로 구분되는데, rPVI는 raw PVI를 뜻하는 용어로 VrPVI와 CrPVI는 각각 연속된 모음 또는 자음 간의 간격을 나타낸다. nPVI는 normalized PVI를 뜻하며, 각 rPVI를 정규화한 값이다.

모든 리듬 특징들은 Correlatore(Mairano & Romano, 2010)를 사용하여 추출하였다.

2.2.4. 발음 특징

발음 특징으로 음소 정확도와 모음 왜곡도를 사용하였다. 음소 정확도는 자음 정확도(percentage of correct consonants, PCC), 모음 정확도(percentage of correct vowels, PCV), 그리고 전체 음소 정확도(percentage of total correct phonemes, PCT)로 구성하였다. PCC, PCV, PCT는 국내 조음 장애 평가 도구인 APAC과 U-TAP을 참고하였다. 음소 정확도 특징은 6,000시간 가량의 비장애 음성으로 학습한 음성인식기를 이용하여 추출하였다.

모음 왜곡도 특징으로는 모음 공간 면적(vowel space area, VSA), formant centralized ratio(FCR), vowel articulatory index (VAI), F2-Ratio를 사용하였다. 각 특징에 대한 수식은 Kang et al.(2010)과 Kim et al.(2014)을 참고하였다. 단, 데이터베이스의 한계로 /우/를 비슷한 조음 위치의 /오/로 대체하여 추출하였음을 밝힌다. VSA는 F1-F2 모음 사각형의 면적을 구하는 공식으로, 화자의 조음기관 움직임 범위를 반영한다. FCR과 VAI는 모음 포먼트 중앙화의 지표이고, F2-ratio는 평균 모음의 대표인 /이/와 원순 모음의 대표인 /오/의 F2 값의 비율이다. 모음 왜곡도 특징은 위에 언급된 음성인식기와 Praat로 추출되었다. 음성인식기를 이용해 음성을 음소 단위로 강제정렬하고, Praat 스크립트를 이용해 필요한 모음 구간의 중앙에서 F1, F2 포먼트를 추출하였다. 포먼트 추출 시 여성 음성은 최대 포먼트 주파수를 5,500 Hz, 남성 음성은 5,000 Hz로 설정하였다. 표 2는 각 모음 왜곡도 특징에 대한 산출식을 제시한다.

표 2. 모음 왜곡도 특징 산출식
Table 2. Formula for vowel distortion feature

특징	산출식
Vowel space area (VSA)	$1/2 * [(F2/이) / (F1/에) + (F2/에) / (F1/아) + (F2/아) / (F1/오) + (F2/오) / (F1/이) - (F1/이) / (F2/에) + (F1/에) / (F2/아) + (F1/아) / (F2/오) + (F1/오) / (F2/이)]$
Formant centralized ratio (FCR)	$(F2/오 + F2/아 + F1/이) / (F1/오) / (F2/이 + F1/아)$
Vowel articulatory index (VAI)	$(F2/이) / (F1/아)$
F2-Ratio	$(F2/이) / (F2/오)$

2.3. 특징 선택

특징 선택의 목적은 추출된 특징 중 학습에 불필요한 특징을 제거하는 것에 있다. 본 연구에서는 scikit-learn 툴킷에서 제공하는 recursive feature elimination(RFE)와 extra trees classifier (ETC)를 사용하였다. RFE는 우선 특징 모두를 포함한 다음, 중요도가 낮은 특징들을 하나씩 제거하여 중요한 특징들을 골라내는 알고리즘이다. 중요도는 독립 변수와 종속 변수 간의 관계를 계산하는 로지스틱 회귀 방식을 바탕으로 산출된다. ETC는 앙상블 알고리즘으로, 여러 개의 의사 결정 트리를 형성한 후 각 의사 결정 트리로부터 분류 결과를 취합하여 결론을 내린다.

ETC는 무작위로 데이터를 나눈 다음 최상의 분할을 선택하는데, 각 집단 안에 다른 집단의 데이터가 최대한 섞여 있지 않도록 학습한다.

2.4. 증중도 분류

증중도 분류를 위한 머신러닝 분류기로는 SVM(support vector machine)과 MLP(machine layer perceptron)를 사용하였다. SVM은 장애 발화처럼 소규모의 데이터 분류 과제에서 좋은 성능을 보이는 것으로 알려져 있다(Kadi et al., 2013; Narendra & Alku, 2018). SVM의 하이퍼파라미터인 C와 gamma는 10^{-4} 에서 10^4 사이에서 그리드 서치(grid search)를 통해 최적화하였다. MLP는 딥러닝 모델 중 하나이다. MLP의 하이퍼파라미터인 은닉층 개수(1~5개), 학습률(0.0001~0.1), 옵티마이저(SGD, lbfgs, adam), 활성화 함수(logistic, tanh, ReLU)도 그리드 서치를 통해 최적화하였다.

3. 데이터베이스

3.1. 음성 데이터

본 논문에서는 한국어 마비말장애 데이터베이스로 QoLT 코퍼스(Choi et al., 2012)를 사용하였다. QoLT 코퍼스는 뇌성마비 마비말장애인 대상 음성인식기 개발을 위해 구축되었으며, 고립단어와 문장으로 구성된다. 본 연구에서는 문장만을 사용했는데, 이는 음질 특징, 운율 특징, 발음 특징 모두를 종합적으로 평가하기 위해서는 고립단어보다 문장이 적합하기 때문이다. 화자는 비장애인 화자 10명과 마비말장애 화자 70명으로 구성되며, 화자 당 5개의 문장을 2회 반복하였다. 그러나 본 논문에서는 모음 왜곡도 추출을 위해 모음 /ㅏ/를 포함하고 있지 않은 문장을 제외하고 사용하였다. 이에 따라 실험에는 화자별 4개의 문장을 2회 반복한 자료로, 총 640개(80명×8문장)의 발화가 사용되었다. 표 3은 데이터베이스의 문장 5개를 제시한다.

표 3. QoLT 문장 발화
Table 3. Speech stimuli from the QoLT database

문장 발화
추석에는 온 가족이 함께 송편을 만든다.
갑자기 미국에 있는 오빠 얼굴이 보고 싶다.
어제 하늘이 컴컴해지더니 비가 쏟아졌다.
동생이랑 싸워서 엄마한테 혼났다.
(제외) 시원한 물 한 잔 주세요.

마비말장애 화자 70명은 말 명료도에 기반하여 장애 증중도가 평가되었다. 말 명료도 평가는 5점 척도의 청지각적 평가로 진행되었으며, 2년 이상의 임상 경험이 있는 언어재활사 5명이 실시하였다. 언어재활사는 각 화자가 발화한 문장 8개를 모두 청취한 다음 화자의 말 명료도를 판단하였다. 말 명료도 평가는 신뢰도 평가를 위해 일주일 간격으로 두 차례 진행되었으며, 평가마다 화자의 순서는 무선배치되었다. 그 결과, 평가자 간 신뢰도는 급내상관계수 .95, 평가자 내 신뢰도는 각각 .92, .98, .84,

.93, .93로 양호하였다. 각 화자의 증중도는 해당 화자에 대한 10개의 증중도 평가 결과(5명×2회) 중 최빈치의 점수로 정하였다. 이에 따라 증중도별 화자의 수는 경도 24명, 경도-중등도 24명, 중등도-중도 15명, 중도 화자 7명이었다.

증중도 자동 분류 실험에 사용된 훈련 화자와 테스트 화자는 장애 증중도 집단별 7:3으로 구성하였으며, 훈련과 테스트에서 서로 겹치지 않도록 구성하였다.

3.2. 통계분석

3.2장에서는 추출된 특징들이 증중도 집단 별 유의미한 차이가 있는지 two-way mixed ANOVA와 Bonferroni 사후분석을 통해 알아보았다. ANOVA 분석에서 개체 내 요인은 말 기능 특징으로, 개체 간 요인은 증중도 집단으로 설정하였다. 더 나아가, 분석 결과가 선행연구와 비슷한 양상을 따르는지 살펴보았다.

3.2.1. 음질 특징

음질 특징인 jitter, shimmer, HNR, voice breaks 개수, voice breaks 정도를 장애 증중도 집단별로 비교하였다. ANOVA 분석 결과, 집단 간 유의미한 차이가 있었다, $F(4, 630)=121.87, p<0.05$. 사후분석 실시 결과, 음질 특징 모두 유의미한 효과가 있었지만, 특히 voice breaks 관련 특징들이 뚜렷한 차이를 보였다. voice breaks 개수는 (경도, 경도-중등도), (중등도-중도, 중도)를 제외한 모든 집단에서 유의한 차이를 보였고, voice breaks 정도는 모든 집단에서 유의한 차이를 보였다($p<0.05$). 음질 특징 모두 마비말장애 화자 집단이 비장애 화자 집단보다 높아 선행연구 결과와 유사한 양상이 나타났다(Seo, 2014; Seo & Seong, 2013). 그림 2는 음질 특징의 평균 그래프이다.

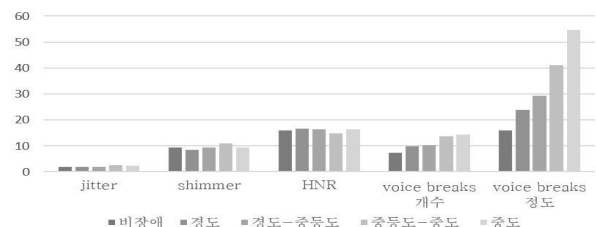


그림 2. 음질 특징 평균
Figure 2. Mean values for voice quality measurements

3.2.2. 운율 특징

발화 속도 특징인 전체 길이, 말 길이, 말 속도, 조음 속도를 장애 증중도 집단 별로 비교하였다. ANOVA 분석 결과, 장애 증중도 집단 별 발화 속도 특징 간에는 유의미한 차이가 있었다, $F(4, 630)=126.89, p<0.05$. 사후검정 결과 전체 길이, 말 길이, 말 속도, 조음 속도 모두 모든 집단 간 유의미한 차이가 나타났다($p<0.05$). 전반적으로 마비말장애 화자는 비장애 화자보다 말 길이와 전체 길이는 길고, 말 속도와 조음 속도는 느렸다. 이는 선행 연구의 보고와 동일한 결과이다(Hong & Byeon, 2014; Lee & Kim, 2012). 그림 3은 발화 속도 특징의 평균 그래프이다.

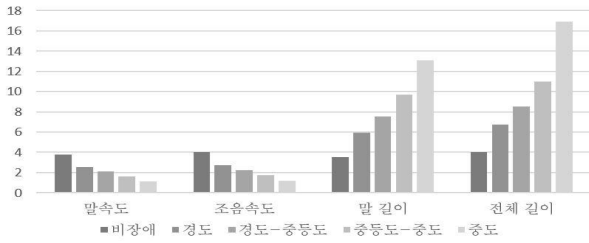


그림 3. 발화 속도 특징 평균

Figure 3. Mean values for speech rate measurements

음높이 특징인 F0의 평균, 표준편차, 중앙값, 최솟값, 최댓값, 25 분위수, 75 분위수를 장애 중증도 집단별로 살펴보았다. ANOVA 분석 결과, 장애 중증도 집단별 음높이 특징 간에는 유의미한 차이가 있었다, $F(4, 630)=16.37, p<0.05$. 사후검정 결과, 최댓값은 (경도, 경도-중등도), (경도-중등도, 중도), (중등도-중도, 중도)를 제외한 모든 집단에서, 75 분위수는 (비장애, 경도), (경도, 경도-중등도), (중등도-중도, 중도)를 제외한 모든 집단에서 유의미한 차이를 보였다($p<0.05$). 반면 F0의 평균, 표준편차, 중앙값, 25 분위수는 중도 집단을 제외한 집단에서 유의미한 차이가 나타나지 않았으며, F0 최솟값은 모든 집단에서 차이를 보이지 않았다. 그림 4는 음높이 특징의 평균 그래프이다.

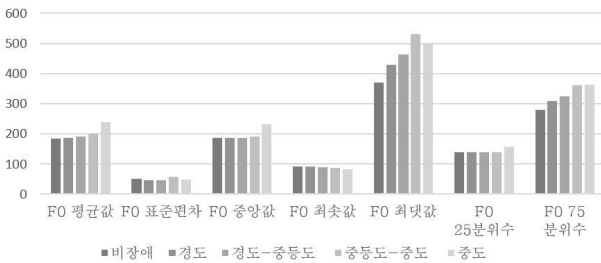


그림 4. 음높이 특징 평균

Figure 4. Mean values for F0 measurements

리듬 특징인 %V, deltas, Varcos, rPVI, nPVI를 장애 중증도 별로 비교하였다. ANOVA 분석 결과, 장애 중증도 집단별 리듬 특징 간에는 유의미한 차이가 있었다, $F(4, 630)=133.72, p<0.05$. 사후검정 결과, deltaV, deltaC, Varco-C, rPVI-V, rPVI-C, nPVI-C에서 모든 집단 간 유의미한 차이가 나타났다($p<0.05$). 마비말장애 집단의 리듬 특징은 비장애 집단보다 높아 선행연구와 유사한 경향을 보였다(Hernandez et al., 2020). 그림 5는 리듬 특징의 평균 그래프이다.

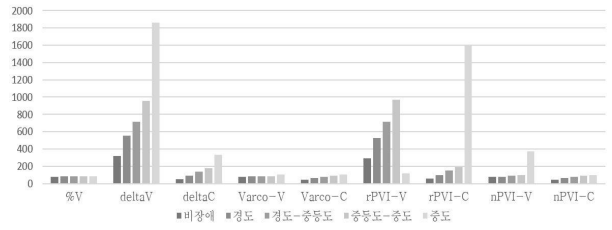


그림 5. 리듬 특징 평균

Figure 5. Mean values for rhythm metrics

3.2.3. 발음 특징

음소 정확도 특징인 PCC, PCV, PCT를 장애 중증도별로 살펴 보았다. ANOVA 분석 결과, 장애 중증도 집단별 음소 정확도 특징 간에는 유의미한 차이가 있었다, $F(4, 630)=8,101.36, p<0.05$. 사후검정 결과, PCC, PCV, PCT 모두 모든 집단 간 차이는 유의미했다($p<0.05$). 음소 정확도 특징 모두 마비말장애 집단이 비장애 집단보다 낮았으며, 장애 중증도가 심할수록 더 낮아지는 경향이 나타났다. 이는 분절음 산출 능력이 말 명료도와 강력한 상관관계를 보인다는 선행연구와 동일한 양상이다(Clarke & Hoops, 1980; Lee et al., 2012; Whitehill & Ciocca, 2000). 그림 6은 음소 정확도 특징의 평균 그래프이다.

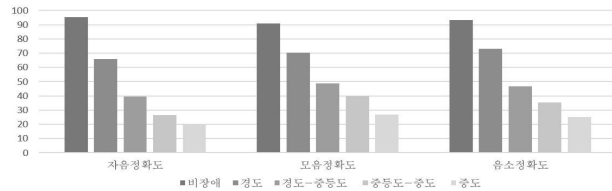
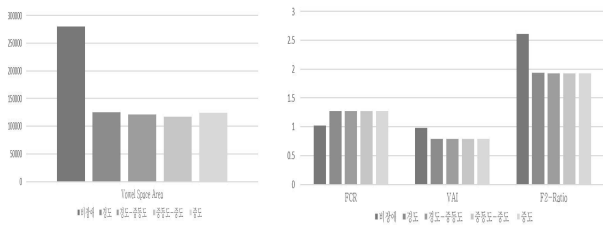


그림 6. 음소 정확도 특징 평균

Figure 6. Mean values for percentage of correct phonemes

모음 왜곡도 특징인 VSA, FCR, VAI, F2-ratio를 장애 중증도 집단 간 비교하였다. ANOVA 분석 결과, 장애 중증도 집단별 모음 왜곡도 특징 간에는 유의미한 차이가 있었다, $F(4, 630)=840.31, p<0.05$. 사후검정 결과, VSA, FCR, VAI, F2-Ratio 모두 비장애 집단과 경도 집단은 다른 집단과 유의미한 차이를 보였지만($p<0.05$), 장애 집단 간에는 유의미한 차이가 나타나지 않았다. 마비말장애 화자는 비장애 화자보다 좁은 모음 공간 면적, 높은 FCR, 그리고 낮은 VAI와 F2-Ratio를 보였다. 이는 마비말장애인의 모음이 비장애인보다 모음 공간 면적이 축소되고, 모음 포먼트가 중앙화되며, 모음 대비가 감소된다는 선행연구의 보고와 일치하는 결과이다(Kang et al., 2010; Kim et al., 2014). 그림 7은 모음 왜곡도 특징의 평균 그래프이다.



FCR, formant centralized ratio; VAI, vowel articulatory index.

그림 7. 모음 왜곡도 특징 평균

Figure 7. Mean values for vowel distortion measurements

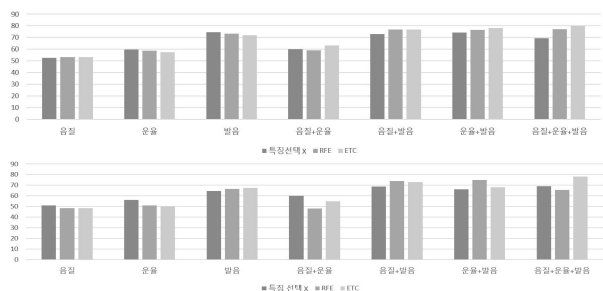
3.2.4. 데이터 분석 정리

ANOVA 분석 결과, 음질 특징, 운율 특징, 발음 특징 모두 집단 간 유의미한 결과를 보였다. 사후분석 결과 모든 집단 간 유의미한 차이를 보인 특징은 다음과 같았다: degree of voice breaks, 전체 길이, 말 길이, 말 속도, 조음 속도, deltaV, deltaC, Varco-C, rPVI-V, rPVI-C, nPVI-C, PCC, PCV, PCT. 비장애 화자와 마비말장애 화자 간 뚜렷한 차이를 보인 특징은 모음 왜곡도 특징인 VSA, FCR, VAI, F2-ratio였다. 더불어, 데이터베이스 내 특징들의 양상은 선행 연구 결과와 유사하였다.

4. 실험 결과

4.1. 중증도 자동 분류

우선, 특징 선택 알고리즘의 필요성을 확인하기 위해 특징 선택 적용 전후의 F1-score를 살펴보았다. 그 결과, 한 개의 말 기능 특징만을 사용한 경우 특징 선택을 적용하지 않았을 때 성능이 더 높았지만, 두 개 이상의 말 기능 특징을 사용했을 때에는 특징 선택 과정을 거쳤을 때 더 높은 성능이 나타났다. 이는 분류에 사용되는 특징 개수가 많다고 해서 반드시 좋은 성능으로 귀결되는 것은 아님을 시사한다. 또한 RFE보다 ETC를 사용했을 때 더 높은 성능을 보였다. 그림 8은 특징 선택 전, RFE 특징 선택 후, ETC 특징 선택 후의 실험 결과(F1-score)를 제시한 그래프이다.



SVM, support vector machine; MLP, multiple layer perceptron.

그림 8. 특징 선택 과정 유무에 따른 실험 결과(SVM, MLP)

Figure 8. Experiment result without and with feature selection(SVM, MLP)

표 4는 각 분류 실험별 F1-score를 제시한다. #F는 분류에 사용된 특징 개수를 의미하며, 가장 좋은 성능을 보인 실험은 볼

드체로 표시하였다. MFCCs만을 사용한 베이스라인 분류 실험 결과, SVM은 53.69%, MLP는 42.52%로 나타났다. MFCCs와 함께 다양한 특징 조합을 실험한 결과, 두 가지 양상이 확인되었다. 첫째, SVM 실험의 MFCCs+음질 특징 조합을 제외한 모든 특징 조합 실험이 베이스라인 실험보다 높은 성능을 보였다. 이는 사용된 말 기능 특징들이 음성의 장애 특성을 반영함을 시사한다. 둘째, 말 기능 특징을 다양하게 사용할수록 분류 성능이 높아졌다. 말 기능 특징을 한 개만 사용했을 때보다 두 개를 사용했을 때 성능이 더 높았으며, 음질, 운율, 발음 세 가지 말 기능 특징을 모두 사용했을 때 SVM 80.15%, MLP 78.09% F1-score로 가장 높은 성능이 나타났다. 이는 다양한 말 기능 특징을 반영했을 때 음성의 장애 특성을 충분히 반영됨을 의미한다. 표 5는 각 실험별 베이스라인과의 상대적 증가율을 제시하며, 그림 9는 baseline_SVM 실험(왼쪽)과 본 연구에서 가장 높은 성능을 보인 ETC_SVM 실험(오른쪽)의 혼동 행렬이다.

표 4. 실험 결과(F1-score)

Table 4. Experimental result (F1-score)

특징 조합	FS	#F	SVM	MLP
MFCCs (baseline)	-	13	53.69	46.52
	-	13+5	52.73	51.10
MFCCs+음질	RFE	13+2	53.31	48.29
	ETC	13+2	53.31	48.29
MFCCs+운율	-	13+20	59.80	56.14
	RFE	13+8	58.8	51.11
	ETC	13+7	57.51	50.18
MFCCs+발음	-	13+7	74.45	64.45
	RFE	13+3	73.23	66.56
	ETC	13+5	72.03	67.37
MFCCs+음질+운율	-	13+25	60.09	59.96
	RFE	13+11	58.97	48.19
MFCCs+음질+발음	-	13+12	72.89	68.79
	RFE	13+6	76.87	73.85
MFCCs+운율+발음	-	13+27	74.20	66.17
	RFE	13+12	76.47	74.91
MFCCs+음질+운율+발음	-	13+32	69.41	69.13
	RFE	13+14	77.28	65.51
	ETC	13+10	80.15	78.09

SVM, support vector machine; MLP, multiple layer perceptron; MFCCs, mel frequency cepstral coefficients; RFE, recursive feature elimination; ETC, extra trees classifier.

표 5. 특징 조합별 상대적 증가율

Table 5. Relative increase of different feature combinations

특징 조합	SVM	MLP
MFCCs+음질	-0.71	3.80
MFCCs+운율	7.11	7.87
MFCCs+발음	34.16	44.82
MFCCs+음질+운율	18.12	18.10
MFCCs+음질+발음	42.99	56.96
MFCCs+운율+발음	45.52	46.65
MFCCs+음질+운율+발음	49.28	67.86

MFCCs, mel frequency cepstral coefficients; SVM, support vector machine; MLP, multiple layer perceptron.

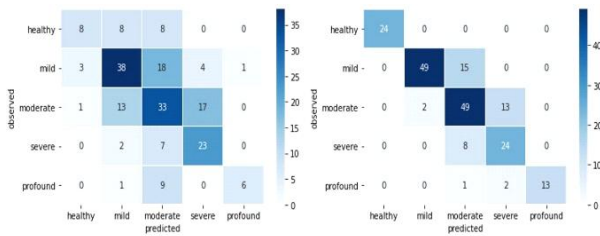


그림 9. 실험 결과 혼동 행렬
Figure 9. Confusion matrix of experimental result

4.2. 특징 선택 결과

ETC 특징 선택 알고리즘이 특징 조합별로 선택한 특징들을 살펴보았다. 특징 조합 중 음질, 운율, 발음을 단독으로 사용했을 때에는 각 말 기능 특징별 세분화 특징이 모두 선택되었다. 그러나 두 개 이상의 말 기능 특징을 사용했을 때에는 선택에서 제외되는 세부 특징들도 있었다. 예를 들어, 발음 특징이 특징 조합에 포함된 경우 리듬 특징과 음높이 특징은 선택에서 제외되는 경향이 나타났다. 이는 한국어의 경우 발음 특징과 리듬 특징, 음높이 특징이 마비말장애 중증도 자동 분류에서 상호보완적 관계가 아니라고 해석할 수 있다. 표 6은 특징 조합별 ETC 알고리즘이 선택한 특징 리스트이다.

표 6. ETC 알고리즘 적용 결과
Table 6. Features selected by ETC algorithm

특징 조합	#F	특징
음질	2	HNR / voice breaks 정도
운율	7	(발화 속도) 전체 길이, 말 길이, 말 속도, 조음 속도 (음높이) F0 중앙값 (리듬) Varco-C, deltaV
		(음소 정확도) PCC, PCT (모음 왜곡도) VSA, VAI, F2-ratio
음질+운율	1	Voice breaks 정도
	7	(발화 속도) 전체 길이, 말 길이, 말 속도, 조음 속도 (리듬) deltaV, Varco-C, nPVI-C
운율+발음	2	(발화 속도) 전체 길이, 말 길이
	7	(음소 정확도) PCC, PCV, PCT (모음 왜곡도) VSA, FCR, VAI, F2-ratio
음질+발음	1	Voice breaks 정도
	5	(음소 정확도) PCC, PCT (모음 왜곡도) VSA, VAI, FCR
음질+운율+발음	1	Voice breaks 정도
	2	(발화 속도) 전체 길이, 말 길이
	7	(음소 정확도) PCC, PCV, PCT (모음 왜곡도) VSA, VAI, FCR, F2-ratio

ETC, extra trees classifier; HNR, harmonic to noise ratio; PCC, percentage of correct consonants; PCT, percentage of total correct phonemes; PVI, pairwise variability index; PCV, percentage of correct vowels; VSA, vowel space area; VAI, vowel articulatory index; FCR, formant centralized ratio.

표 6에 제시된 것처럼, 가장 높은 성능을 보인 음질+운율+발음 조합의 특징 선택 결과는 다음과 같았다: Voice breaks 정도, 전체 길이, 말 길이, PCC, PCV, PCT, VSA, VAI, FCR, F2-Ratio. 즉, 최적의 특징 셋은 음질 특징 1개(voice breaks 1개), 운율 특

징 2개(발화 속도 2개), 발음 특징 7개(음소 정확도 3개, 모음 왜곡도 4개)로 음질, 운율, 발음 특징이 모두 포함되었다. 더 나아가, 세 가지 말 기능 특징 중 발음 특징이 가장 높은 비중을 차지하였다. 이는 발음 특징이 마비말장애 중증도 자동 분류에서 가장 큰 영향력을 가지고 있음을 의미한다.

4.3. 논의

본 연구는 마비말장애 중증도 자동 분류에서 음질, 운율, 발음의 다양한 말 기능 특징 조합을 사용할 것을 제안하였다. 본 연구는 두 가지 측면에서 의의가 있다. 첫째, 분류 정확도이다. 본 연구의 최고 성능인 80.15% F1-score는 한국어 발화를 대상으로 진행된 연구 중 최고 성능으로, 선행연구 중 가장 높게 보고된 70.10%보다 14.34% 상대적 증가율을 보였다(Hernandez et al., 2020). 이는 운율 특징에만 초점을 맞추어 연구를 진행한 선행연구에 비해 본 연구는 음질, 운율, 발음 특징을 모두 중증도 자동 분류에 반영하였기 때문이다. 둘째, 마비말장애 음성 특징 탐색이다. 본 연구는 마비말장애 중증도 자동 분류에서 음질, 운율, 발음 특징이 종합적으로 고려되어야 함을 실험으로써 증명하였다. 이는 말 명료도가 호흡, 발성, 조음, 운율의 복합적인 영향을 받는다는 언어 병리학과의 선행연구를 지지하는 결과이기도 하다. 더 나아가 발음 특징은 최적의 특징 셋에서 가장 큰 비중을 차지하였다. 이는 중증도 자동 평가에서 음성의 질, 운율, 음질 특징 중 음성의 질 특징이 가장 많이 선택된 Kim & Kim(2012)과 일맥상통한 결과이다. 또한 청지각적 평가의 결과와도 유사한데, Hong et al.(2018)은 청지각적 평가 시 말 명료도와 부정확한 자음, 모음 왜곡 간 가장 높은 상관성이 나타난다고 보고하였다. 반면, 본 연구 결과와 달리, Hong et al.(2018)에서는 말 명료도와 음질 특징 중 음성 끊김, 그리고 운율 특징 중 발화 속도 간에는 낮은 상관도가 나타났다. 두 연구 간 다른 양상을 보인 이유는 선행연구에서는 세부평가 요소 각각의 상관 정도를 살펴본 반면, 본 연구에서는 평가 요소 간 상호작용 요소가 반영되었기 때문으로 해석할 수 있다. 단, 본 연구와 선행연구에서 사용된 데이터가 다르므로 해석에 있어서 주의할 기를 기울여야 할 것이다. 추후 연구에서는 본 연구에서 사용한 QoLT 데이터에 대한 전문가의 청지각적 평가 결과를 수집하여 청지각적 평가 결과와 중증도 자동 분류 결과의 관계를 보다 긴밀하게 살펴볼 필요가 있겠다.

5. 결론

본 논문은 마비말장애 중증도 자동 분류에서 사용되는 다양한 특징 조합을 실험해보았다. 선행연구에서는 하나의 말 기능 특징에 집중하여 그 유용성을 밝혔다면, 본 연구에서는 음성의 장애 특징을 다양한 말 기능의 측면에서 종합적으로 반영하고자 했다. 실험 결과, SVM과 MLP 분류기에서 음질, 운율, 조음 특징 모두를 사용하였을 때 가장 높은 성능인 80.15 F1-score와 78.09% F1-score를 달성하였다. 이는 말 명료도가 한 가지 말 기능 특징만으로 결정되는 것이 아닌, 다양한 말 기능 특징의 영

향을 다면적으로 받는다는 것을 의미한다. 또한, 본 연구에서 사용한 특징들은 언어 병리학 분야에서 말 명료도를 평가할 때 주로 사용하는 지표들로, 실험 결과를 보다 직관적으로 해석할 수 있도록 하였다. 가장 좋은 성능을 보인 ETC 특징 선택 알고리즘의 특징 셋을 살펴보면, 발음이 말 명료도에 가장 큰 영향을 미치며, 운율, 음질 순으로 영향력이 있음을 확인할 수 있다. 이는 청지각적 평가에서 여러 말 기능 특징 중 조음, 운율, 발성, 호흡 순으로 말 명료도와 강한 상관을 보인다고 보고한 Hong et al.(2018) 결과와 유사한 결과이다. 본 연구의 한계점은 머신러닝에 적용하기에는 다소 작은 규모의 데이터베이스를 사용했다는 것이다. 본 연구는 화자 80명(비장애 화자 10명, 장애 화자 70명), 총 640개의 문장으로 진행된 연구로, 결과의 일반화를 위해서는 더 규모 있는 데이터베이스로 연구를 진행할 필요가 있다. 또한, 청지각적 평가와 자동 분류 결과 간의 관계를 보다 긴밀하게 분석하는 것이 필요하다. 이러한 분석은 한국어를 사용하는 마비말장애 화자의 말 특성을 연구하는 것에 중요한 통찰력을 제공할 수 있을 것으로 기대된다.

References

- Bhat, C., & Strik, H. (2020). Automatic assessment of sentence-level dysarthria intelligibility using BLSTM. *IEEE Journal of Selected Topics in Signal Processing*, 14(2), 322-330.
- Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341-345.
- Choi, D. L., Kim, B. W., Kim, Y. W., Lee, Y. J., Um, Y., & Chung, M. (2012, May). Dysarthric speech database for development of QoLT software technology. *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 3378-3381). Istanbul, Turkey.
- Clarke, W. M., & Hoops, H. R. (1980). Predictive measures of speech proficiency in cerebral palsied speakers. *Journal of Communication Disorders*, 13(5), 385-394.
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 12(2), 246-269.
- Dellwo, V., & Wagner, P. (2003, August). Relationships between speech rate and rhythm. *Proceedings of the 15th International Congress of the Phonetic Sciences*. Barcelona, Spain.
- Hernandez, A., Kim, S., & Chung, M. (2020). Prosody-based measures for automatic severity assessment of dysarthric speech. *Applied Sciences*, 10(19), 6999.
- Hernandez, A., Yeo, E. J., Kim, S., & Chung, M. (2020). Dysarthria detection and severity assessment using rhythm-based metrics. *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH 2020)* (pp. 2897-2901). Shanghai, China.
- Hong, S., & Byeon, H. (2014). Speech rate and pause characteristics in speaker with flaccid dysarthria. *The Korea Academia-Industrial Cooperation Society*, 15(1), 2930-2936.
- Hong, S. M., Jeong, P. Y., & Sim, H. S. (2018). Comparison of perceptual assessment for dysarthric speech: The detailed and general assessments. *Communication Sciences & Disorders*, 23(1), 242-253.
- Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1-15.
- Janbakhshi, P., Kodrasi, I., & Bourlard, H. (2019, May). Pathological speech intelligibility assessment based on the short-time objective intelligibility measure. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton, UK.
- Kadi, K. L., Selouani, S. A., Boudraa, B., & Boudraa, M. (2013, October). Discriminative prosodic features to assess the dysarthria severity levels. *Proceedings of the World Congress on Engineering*. London, UK.
- Kang, Y. A., Yoon, K. C., Lee, H. S., & Seong, C. J. (2010). A comparison of parameters of acoustic vowel space in patients with Parkinson's disease. *Phonetics and Speech Sciences*, 2(4), 185-192.
- Kim, M. J., & Kim, H. (2012, September). Combination of multiple speech dimensions for automatic assessment of dysarthric speech intelligibility. *Proceedings of the 13th Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Portland, OR.
- Kim, M. J., Kim, Y., & Kim, H. (2015). Automatic intelligibility assessment of dysarthric speech using phonologically-structured sparse linear model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(4), 694-704.
- Kim, S., Kim, J. H., & Ko, D. H. (2014). Characteristics of vowel space and speech intelligibility in patients with spastic dysarthria. *Communication Sciences & Disorders*, 19(3), 352-360.
- Lansford, K. L., & Liss, J. M. (2014). Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. *Journal of Speech, Language, and Hearing Research*, 57(1), 57-67.
- Lee, Y. M., Sung, J. E., Sim, H. S., Han, J. H., & Song, H. N. (2012). Analysis of articulation error patterns depending on the level of speech intelligibility in adults with dysarthria. *The Korean Academy of Speech-Language Pathology and Audiology*, 17(1), 130-142.
- Lee, E., & Kim, J. (2012). Correlation of speech rate changes on intelligibility and acceptability in dysarthric speakers. *Journal of Speech-language & Hearing Disorders*, 21(3), 127-144.
- Mairano, P., & Romano, A. (2010). Un confronto tra diverse metriche ritmiche usando Correlatore. In S. Schmid, M. Schwarzenbach, & D. Studer (Eds.), *La dimensione temporale del parlato* (pp. 79-100). Torriana, Italy: EDK.

- McFee, B., Colin, R., Dawen, L., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto, O. (2015, July). Librosa: Audio and music signal analysis in Python. *Proceedings of the 14th Python in Science Conference* (pp. 18-25). Austin, TX.
- Narendra, N. P., & Alku, P. (2018, September). Dysarthric Speech Classification Using Glottal Features Computed from Non-words, Words and Sentences. *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH 2020)* (pp. 3403-3407). Hyderabad, India.
- Narendra, N. P., & Alku, P. (2021). Automatic assessment of intelligibility in speakers with dysarthria from coded telephone speech using glottal features. *Computer Speech & Language*, 65, 101117.
- Seo, I., & Seong, C. (2013). Voice quality of dysarthric speakers in connected speech. *Phonetics and Speech Sciences*, 5(4), 33-41.
- Seo I. H. (2014). *Acoustic measures of voice quality and phonation types across speech conditions in dysarthria* (Doctoral dissertation). Chungnam National University, Daejeon, Korea.
- Whitehill, T. L., & Ciocca, V. (2000). Speech errors in Cantonese speaking adults with cerebral palsy. *Clinical Linguistics & Phonetics*, 14(2), 111-130.

• **여은정 (Eun Jung Yeo)**

서울대학교 언어학과 박사과정
 서울시 관악구 관악로 1
 Tel: 02-880-6162
 Email: ej.yeo@snu.ac.kr
 관심분야: 장애음성인식, 음성언어처리, 딥러닝

• **김선희 (Sunhee Kim)**

서울대학교 불어교육과 교수
 서울시 관악구 관악로 1
 Tel: 02-880-7693
 Email: sunhim@snu.ac.kr
 관심분야: 프랑스어 음성학, 음성언어처리

• **정민화 (Minhwa Chung)** 교신저자

서울대학교 언어학과 교수
 서울시 관악구 관악로 1
 Tel: 02-880-9195
 Email: mchung@snu.ac.kr
 관심분야: 음성인식, 음성언어처리, 컴퓨터 기반 언어교육

음질, 운율, 발음 특징을 이용한 마비말장애 중증도 자동 분류*

여 은 정¹ · 김 선 희² · 정 민 화¹¹서울대학교 언어학과, ²서울대학교 불어교육과

국문초록

본 논문은 말 명료도 기준의 마비말장애 중증도 자동 분류 문제에 초점을 둔다. 말 명료도는 호흡, 발성, 공명, 조음, 운율 등 다양한 말 기능 특징의 영향을 받는다. 그러나 대부분의 선행연구는 한 개의 말 기능 특징만을 중증도 자동 분류에 사용하였다. 본 논문에서는 음성의 장애 특성을 효과적으로 포착하기 위해 마비말장애 중증도 자동 분류에서 음질, 운율, 발음의 다양한 말 기능 특징을 반영하고자 하였다. 음질은 jitter, shimmer, HNR, voice breaks 개수, voice breaks 정도로 구성된다. 운율은 발화 속도(전체 길이, 말 길이, 말 속도, 조음 속도), 음높이(F0 평균, 표준편차, 최솟값, 최댓값, 중간값, 25 사분위값, 75 사분위값), 그리고 리듬(% V, deltas, Varcos, rPVI, nPVI)을 포함한다. 발음에는 음소 정확도(자음 정확도, 모음 정확도, 전체 음소 정확도)와 모음 왜곡도[VSA(vowel space area), FCR (formant centralized ratio), VAI(vowel articulatory index), F2 비율]가 있다. 본 논문에서는 다양한 특징 조합을 사용하여 중증도 자동 분류를 시행하였다. 실험 결과, 음질, 운율, 발음 특징 세 가지 말 기능 특징 모두를 분류에 사용했을 때 F1-score 80.15%로 가장 높은 성능이 나타났다. 이는 마비말장애 중증도 자동 분류에는 음질, 운율, 발음 특징이 모두 함께 고려되어야 함을 시사한다.

핵심어: 마비말장애, 장애 중증도 자동 분류, 발음 정확도, 머신러닝, 특징선택

참고문헌

- 강영애, 윤규철, 이학승, 성철재 (2010). 파킨슨병 환자의 음향 모음 공간 파라미터 비교. *말소리와 음성과학*, 2(4), 185-192.
- 김성윤, 김정환, 고도홍 (2014). 경직형 마비말장애의 말 명료도와 모음공간 특성. *언어청각장애연구*, 19(3), 352-360.
- 서인효, 성철재. (2013). 연결발화에서 마비말장애자의 음질 특성. *말소리와 음성과학*, 5(4), 33-41.
- 서인효 (2014). *마비말장애의 발화조건에 따른 음질 및 발생유형의 음향음성학적 접근*. 충남대학교 박사학위논문.
- 이영미, 성지은, 심현섭, 한지후, 송한내 (2012). 마비말장애인의 조음오류 유형에 따른 말명료도 분석. *언어청각장애연구*, 17(1), 130-142.
- 이은경, 김지채 (2012). 마비말장애 화자의 말 속도에 따른 말 명료도 및 말 용인도. *언어치료연구*, 21(3), 127-144.
- 홍새미, 변해원 (2014). 이완형 마비말장애 화자의 말속도와 습 특성. *한국산학기술학회 논문지*, 15(5), 2930-2936.
- 홍새미, 정필연, 심현섭 (2018). 마비말장애 발화의 청지각적 평가방법 비교: 세부평가와 범주평가. *언어청각장애연구*, 23(1), 242-253.

* 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 연구개발지원사업으로 수행되었음(과제번호: R2019080018).