



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

문학석사 학위논문

발음 정확도 기반의  
마비말장애 중증도 자동 분류

Automatic Severity Classification of Dysarthric  
Speech based on Pronunciation Accuracy

2021년 2월

서울대학교 대학원  
언어학과  
여 은 정

# 발음 정확도 기반의 마비말장애 중증도 자동 분류

지도교수 정 민 화

이 논문을 문학석사 학위논문으로 제출함

2020년 11월

서울대학교 대학원

언어학과 언어학 전공

여 은 정

여은정의 석사 학위논문을 인준함

2021년 1월

위 원 장

이 호 성



부 위 원 장

정 민 화



위 원

김 선 희

(인)



## 국문 초록

마비말장애는 중추 신경계 및 자율 신경계의 손상으로 말소리 산출과 관련된 근육이 마비되거나 약해지면서 생기는 말운동장애이다. 언어재활사는 알맞은 중재방안을 모색하기 위해 마비말장애의 중증도를 판단한다. 그러나 일반적으로 장애 중증도 분류에 사용되는 청지각적 평가는 많은 시간과 노력이 소요될 뿐만 아니라 평가자 간 및 평가자 내 신뢰도를 확보하기 어렵다는 단점이 있다. 마비말장애 중증도 자동 분류 기술은 이러한 단점들을 보완하며 언어재활사의 업무를 보조할 수 있다. 분류에 소요되는 시간과 노력을 절약하고, 객관적이고 일관된 결과를 제공하기 때문이다.

선행연구에서는 마비말장애 중증도 자동 분류를 위한 다양한 특징 셋을 제안하였다. 음성의 전반적 특징을 반영하는 스펙트럼 특징만을 사용하거나, 음성적 특징을 세분화하여 음질 특징, 운율 특징 등으로 특징 셋을 구성하였다. 그러나 선행 연구의 특징 셋은 음소 단위의 특징인 발음 정확도 특징을 반영하고 있지 않다. 발음 정확도 특징은 언어치료 분야에서 마비말장애의 중증도를 구분할 때 가장 일반적으로 사용되는 반면, 장애 중증도 자동 분류 연구에서는 상대적으로 거의 사용되지 않았다.

본 논문은 발음 정확도 특징을 마비말장애 중증도 자동 분류에 사용하는 것을 제안한다. 발음 정확도 특징은 음소정확도 특징과 모음왜곡도 특징을 포함하는 개념으로 음소정확도 특징은 자음정확도, 모음정확도, 음소정확도(자음+모음), 모음왜곡도 특징은 모음사각도 면적, VAI, FCR, F2-Ratio로 구성된다. 본 논문에서는 발음 정확도 특징의 유용성을 확인하기 위해 앞서 언급한 스펙트럼 특징(MFCCs), 음질 특징, 운율 특징을 베이스라인 특징으로 사용하였다. 추출된 특징 셋은 Recursive Feature Elimination(RFE)과 Extra Trees Classifier(ETC) 두 개의 특징 선택 알고리즘을 통해 최적화되었다. 선택된 특징들은 SVM(Support Vector Machine)과 MLP(Multiple Layer Perceptron) 분류기의 입력값으로 사용되었고, 분류기는 각 음성의 장애 중증도(비장애/경도/경도-중등도/중등

도-중도/중도)를 분류하도록 학습되었다. 분류기의 성능은 정확도, 정밀도, 재현율, F1-점수로 평가되었다.

발음 정확도 특징 추가 전후의 실험 결과를 살펴본 결과, 특징 선택 전, RFE 적용 후, ETC 적용 후 SVM의 분류 정확도 상대적 증가율은 각각 15.83%, 25.42%, 23.39%였고, MLP의 분류 정확도 상대적 증가율은 각각 28.97%, 21.19%, 22.95%로 나타났다. 최고의 성능을 보인 실험은 ETC 특징 선택 알고리즘-SVM 조합 실험으로, 77.5%의 분류 정확도를 보였다. 더 나아가, 각 실험에서 특징 선택 알고리즘이 선택한 특징과 특징 별 개별 기여도를 살펴보았다. 그 결과, 발음 정확도 특징을 추가했을 때 베이스라인에서 선택되었던 음질, 운율 특징 다수가 발음 정확도 특징에 의해 대체되었으며, 발음 정확도 특징은 음질, 운율 특징보다 더 높은 기여가중치를 가졌다.

실험 결과는 다음을 시사한다. 첫째, 발음 정확도 특징은 마비말장애 중증도 자동 분류에 도움이 된다. 발음 정확도 특징 추가 전후 분류 정확도를 비교했을 때, 발음 정확도 특징을 추가했을 때 더 높은 분류 정확도를 보였다. 발음 정확도 특징은 언어치료 분야에서 일반적으로 사용되어왔지만, 자동 분류 연구에서는 명시적으로 사용된 경우가 적다. 실험 결과는 발음 정확도 특징이 자동 분류에서도 사용되어야 함을 시사한다. 둘째, 발음 정확도 특징은 마비말장애 중증도 자동 분류에서 음질, 운율 특징보다 더 큰 영향력을 행사한다. 실험 별 선택된 특징을 살펴본 결과, 발음 정확도 특징이 음질, 운율 특징에 비해 대체되었다. 특징 별 개별 기여도를 살펴본 결과, 모든 발음 정확도 특징의 개별 기여도가 음질, 운율 특징의 개별 기여도가 높았다. 이는 발음 정확도 특징이 다른 특징보다 마비말장애 중증도와 높은 상관관계를 보인다는 선행연구와 일맥상통한 결과이다.

**주요어 :** 마비말장애, 장애 중증도 자동 분류, 발음 정확도, 머신러닝, 특징 선택

**학번 :** 2019-24939

# 목 차

1. 서론 .....	1
2. 관련 연구 .....	4
2.1 장애 중증도 분류 기준 .....	4
2.1.1 말명료도 .....	4
2.1.2 자음정확도 .....	5
2.2 마비말장애 음성의 발음 정확도 특징 .....	6
2.3 마비말장애 음성 장애 중증도 자동 분류 .....	7
3. 실험 방법론 .....	11
3.1 실험 설계 .....	11
3.2 특징 정의 및 특징 추출 방법 .....	11
3.2.1 Mel Frequency Cepstral Coefficients (MFCCs) .....	12
3.2.2 음질 특징 .....	14
3.2.3 운율 특징 .....	15
3.2.3.1 발화 속도 .....	15
3.2.3.2 음높이 .....	15
3.2.3.3 리듬 .....	16
3.2.4 발음 정확도 특징 .....	17
3.4.1 음소정확도 .....	17
3.4.2 모음왜곡도 .....	18
3.3 특징 선택 알고리즘 .....	19
3.2.3.1 Recursive Feature Elimination(RFE) .....	20
3.2.3.2 Extra Trees Classifier(ETC) .....	20
3.4 머신러닝 분류기 .....	21
3.4.1 Support Vector Machine(SVM) .....	21
3.4.2 Multiple Layer Perceptron(MLP) .....	22
4. 데이터베이스 .....	24
4.1 QoLT(Quality of Life Technology) 코퍼스 .....	24

4.2	자음정확도 평가 .....	24
4.3	말명료도 평가 .....	25
4.4	자음정확도 평가 결과와 말명료도 평가 결과 비교 .....	26
4.5	데이터베이스 통계 분석 .....	27
4.5.1	음절 특징 .....	27
4.5.2	운율 특징 .....	28
4.5.2.1	발화 속도 .....	28
4.5.2.2	음높이 .....	29
4.5.2.3	리듬 .....	30
4.5.3	발음 정확도 특징 .....	31
4.5.3.1	음소정확도 .....	31
4.5.3.2	모음왜곡도 .....	32
<b>5.</b>	<b>실험 .....</b>	<b>34</b>
5.1	실험 데이터 구성 .....	34
5.2	분류기 성능 척도 .....	34
5.3	실험 결과 .....	35
5.3.1	특징 선택 적용 전 .....	35
5.3.2	RFE 적용 후 .....	35
5.3.3	ETC 적용 후 .....	37
5.4	실험 결과 정리 .....	39
5.4.1	분류 정확도 비교 .....	39
5.4.2	특징 선택 .....	41
5.5	토의.....	43
<b>6.</b>	<b>결론 .....</b>	<b>45</b>
	<b>참 고 문 헌 .....</b>	<b>47</b>
	<b>Abstract .....</b>	<b>52</b>

## 표 목차

표 2.1	말명료도 기반 장애 중증도 기준(철자 전사)	4
표 2.2	자음정확도 기반 장애 중증도 기준	6
표 4.1	문장 자료	24
표 4.2	언어재활사 정보	25
표 4.3	자음 정확도 평가 결과 별 말명료도 평가 결과 분포	26
표 4.4	음질 측정치 평균	28
표 4.5	운율 - 발화 속도 측정치 평균	29
표 4.6	운율 - 음높이 측정치 평균	30
표 4.7	운율 - 리듬 측정치 평균	31
표 4.8	발음 정확도 - 음소정확도 측정치 평균	31
표 4.9	발음 정확도 - 모음왜곡도 측정치 평균	33
표 5.1	훈련/테스트 화자 및 문장 수	34
표 5.2	특징 선택 알고리즘 적용 전 실험 결과	35
표 5.3	발음 정확도 특징 추가에 따른 정확도 상대적 증가 비율	35
표 5.4	RFE 적용 후 실험 결과	36
표 5.5	RFE 실험의 상대적 증가 비율	36
표 5.6	실험 별 선택된 최적 특징 비교(RFE)	37
표 5.7	ETC 적용 후 실험 결과	38
표 5.8	ETC 실험의 상대적 증가 비율	38
표 5.9	실험 별 선택된 최적 특징 비교(ETC)	39
표 5.10	발음 정확도 특징 추가에 따른 상대적 증가 비율(종합)	39
표 5.11	발음 정확도 특징 추가 실험 세부 결과	40
표 5.12	장애 중증도 별 분류 정확도 비교	41
표 5.13	최종 특징 선택 결과 비교	42
표 5.14	특징 별 기여가중치 평균 순서	43



## 그림 목차

그림 2.1	말명료도 평가 기반 장애 중증도 기준(5점 척도) .....	5
그림 3.1	실험 설계도 .....	11
그림 3.2	MFCCs 추출 과정 .....	12
그림 3.3	해밍 윈도우의 그래프 및 수식.....	13
그림 3.4	mel scale filter .....	13
그림 3.5	자음정확도, 모음정확도, 음소정확도 산출 예시 .....	17
그림 3.6	특징 개수에 따른 정확도(RFE) .....	20
그림 3.7	서포트 벡터 머신(SVM) .....	22
그림 3.8	다층 퍼셉트론(MLP) .....	23
그림 4.1	음질 측정치 평균 .....	28
그림 4.2	발화 속도 측정치 평균 .....	29
그림 4.3	음높이 측정치 평균 .....	30
그림 4.4	리듬 측정치 평균 .....	31
그림 4.5	음소정확도 측정치 평균 .....	31
그림 4.6	모음왜곡도 측정치 평균 .....	33
그림 5.1	SVM의 발음 정확도 특징 추가 전후 혼동 행렬(RFE) .....	36
그림 5.2	SVM의 발음 정확도 추가 전후 혼동 행렬(ETC) .....	38
그림 5.3	SVM의 발음 정확도 추가 전후 혼동 행렬 .....	41

## 수식 목차

수식 3.1	Varco-V와 Varco-C 수식 .....	16
수식 3.2	rPVI와 nPVI 수식 .....	16
수식 3.3	모음 공간 면적 수식 .....	18
수식 3.4	FCR, VAI, F2-Ratio 수식 .....	19

# 1. 서론

마비말장애는 중추 신경계 및 자율 신경계 손상에 기인하며, 말소리 산출과 관련된 근육이 마비되거나 약해지면서 나타나는 말운동장애이다. 마비말장애는 뇌손상, 뇌졸중, 신경 퇴행성 질환 등의 결과로 나타날 수 있다. 마비말장애인의 약 20~30%가 뇌손상 또는 뇌졸중에 기인하며(홍새미 외, 2018; Theodoros, Murdoch, & Goozee, 2001; Warlow et al., 2001), 신경 퇴행성 질환의 경우 파킨슨병 환자의 90% 이상(Muller et al., 2001)이, 루게릭병 환자의 약 25%가 마비말장애를 동반한다(Traynor et al., 2000). 이처럼 마비말장애는 선천적 및 후천적 뇌 손상, 노화에 따른 퇴행성 신경질환 등 넓은 범위에서 나타나는 장애이다.

언어재활사는 마비말장애 중증도 분류를 통해 화자가 의사소통 상황에서 어느 정도로 어려움을 겪고 있는지 확인하고, 화자의 상황에 알맞은 중재 방안을 계획한다. 일반적으로 마비말장애 음성 중증도 분류는 언어재활사의 청지각적 평가로 진행되는데, 이는 많은 시간과 노력이 소요될 뿐만 아니라 주관적인 평가로 평가자 간 및 평가자 내 신뢰도 확보가 어려운 주관적 평가라는 단점이 있다(Narenda & Alku 2021; Hernandez et al., 2020; Kim, Kim & Kim, 2015). 반면 장애 중증도 자동 분류 기술은 평가에 소요되는 시간과 노력을 절약하고, 객관적이고 일관된 평가 결과를 제공함으로써 언어재활사의 업무를 효율적으로 보조할 수 있다.

최근 딥러닝 기술이 발전함에 따라 마비말장애 중증도 자동 분류 연구에서도 딥러닝 기술을 사용하는 연구들이 늘어났다. 다양한 음성 특징들을 추출하여 머신러닝 분류기의 입력값으로 사용하거나(Narendra & Alku, 2020, Hernandez et al., 2020; Bhat, Vachhani, & Kopparapu., 2017), 심층 신경망에 음성을 통째로 입력값으로 넣는 시도들이 있었다(Janbakhshi, Kodrasi, & Bourlard, 2019; Bhat & Strik, 2020). 후자의 심층 신경망을 제안하는 방법론은 특징 엔지니어링 과정이 간단하다는 장점이 있지만, 많은 양의 데이터가 필요하다는 단점이 있다. 그러나 장

에 발화의 경우 대규모의 코퍼스를 수집하기란 쉽지 않다. 심층 신경망을 제안한 선행연구들 역시 마비말장애 중증도 자동 분류를 비장애 음성과 장애 음성으로 구분하는 정도에 그쳤다. 따라서 본 논문에서는 전자의 방법론인 다양한 음성 특징을 사용한 머신러닝 분류기 실험에 집중하였다.

선행연구에서는 마비말장애 중증도 자동 분류를 위한 다양한 특징 셋을 제안하였다. 음성의 전반적인 특징을 반영하는 스펙트럼 특징만을 사용하거나, 음성적 특징을 세분화하여 음질 특징, 운율 특징 등으로 특징 셋을 구성하였다. 최신 연구인 Hernandez et al.(2020)은 베이스라인으로 스펙트럼 특징인 MFCCs를 사용하고, 운율 특징을 추가했을 때 분류 정확도가 향상되었음을 보고하였다. Narendra & Alku(2020)은 성문 특징이 마비말장애 검출에 도움이 되는지 분석하였다. OpenSMILE 툴킷(Eyben et al., 2013)으로 추출한 스펙트럼 특징을 베이스라인으로 삼고 성문 특징을 추가한 결과, 성문 특징을 추가했을 때 더 높은 성능을 보였다.

이처럼 선행연구는 음성의 전반적인 특징 정보를 반영하고자 했지만, 음소 단위의 특징인 발음 정확도 특징을 반영하지 않았다. 발음 정확도 특징은 언어치료 분야에서 마비말장애 중증도를 구분할 때 가장 일반적으로 사용되는 특징으로, 장애 중증도를 자동 분류할 때에도 도움이 될 수 있다. 따라서 본 논문이 탐색하고자 하는 질문은 다음과 같다:

**1. 발음 정확도 특징은 마비말장애 중증도 자동 분류에 도움이 되는가?**

**2. 마비말장애 중증도 자동 분류에서 발음 정확도 특징은 음질 특징, 운율 특징과 어떤 관계에 있는가?**

본 논문의 구성은 다음과 같다. 2장에서는 마비말장애 중증도 분류 기준인 말명료도와 자음정확도를 살펴본 후, 본 논문에서 제안하는 발음 정확도 특징과 마비말장애 중증도 자동 분류와 관련된 선행연구를 요약한다. 3장에서는 개괄적인 실험 구조와 실험 방법론에 대해 기술한다. 4

장에서는 실험에 사용된 데이터베이스를 소개하고 장애 중증도 별 통계 분석을 실시한다. 5장에서는 실험 환경과 함께 실험 결과를 요약하고, 마지막으로 6장에서는 실험 결과가 시사하는 바를 서술하고 향후 보완할 사항에 대해 언급한다.

## 2. 관련 연구

### 2.1 장애 중증도 분류 기준

언어치료 분야에서 사용되는 장애 중증도 분류 기준은 말명료도와 자음정확도이다. 2.1장에서 각 기준에 대해 알아본다.

#### 2.1.1 말명료도

말명료도란 화자가 의도한 메시지가 청자에게 성공적으로 전달되는 정도이다. 말명료도를 평가하는 방법에는 여러 가지가 있다.

첫째, 철자 전사에 의한 평가로 Frenchay Dysarthric Assessment(FDA-2) 표준화 도구에서 사용하는 방식이다(Enderby, 1980). 이때 말명료도는 화자가 발화한 전체 단어 중 화자의 의도와 동일하게 철자 전사된 단어의 비율로 산출된다. 철자 전사를 이용한 말명료도 기반 장애 중증도 판단 기준은 표 2.1과 같다.

표 2.1 말명료도 기반 장애 중증도 기준(철자 전사)

기준	장애 중증도
76~100%	경도
51~75%	경도-중등도
26~50%	중등도-중도
25% 이하	중도

둘째, 등간 척도에 의한 평가로 Mayo Clinic Rating System 표준화 도구(Darley, Aronson & Brown, 1969)에서 사용하는 방식이다. Mayo Clinic Rating System 개정판(Gerald, Murdoch, & Chenery, 1987)은 총 30개의 평가 항목을 포함하는 세부평가와 호흡, 발성, 공명, 조음, 운율, 전반적 중증도의 범주로 평가하는 범주평가로 구성된 4점 등간 척도 평가이다. 이때 전반적 중증도는 말명료도를 의미한다. 홍새미, 정필연, 심현섭(2018)은 세부평가와 범주평가를 비교하여 각 평가 별 타당도와 신뢰도를 분석하였다. 분석 결과, 두 평가의 평균 평정치는 큰 차이를 보이

지 않았지만, 범주평가가 세부평가보다 더 높은 평가자 간 신뢰도와 평가자 내 신뢰도를 보였다. 범주평가의 호흡, 발성, 공명, 조음, 운율의 중증도와 전반적 중증도 간 상관관계를 분석한 결과, 공명을 제외하고 모두 유의미한 상관관계를 보였으며, 조음, 운율, 발성, 호흡 순으로 높은 상관성이 나타났다. 이는 말의 전반적 중증도, 즉 말명료도를 결정하기 위해서는 말 산출과 관련된 다양한 요소의 특징이 반영되어야 함을 의미한다.

또한, 등간 척도에는 다양한 종류가 있다. 박혜진(2007)은 화자의 말명료도를 3점, 5점, 7점의 세 가지 등간 척도법으로 평가하고 척도법 별 타당도 및 신뢰도를 평가하였다. 실험 결과, 치료 경험의 여부와 상관없이 타당도와 신뢰도가 높은 평가방법은 5점 척도였다. 5점 척도 기준의 말명료도 평가 결과 화자는 정상, 경도, 경도-중등도, 중등도-중도, 중도로 분류된다. 5점 척도는 그림 2.1에 제시하였다.



그림 2.1 말명료도 평가 기반 장애 중증도 기준(5점 척도)

### 2.1.2 자음정확도

자음정확도란 전체 목표 자음 중 정조음한 자음의 비율을 의미한다. 자음정확도는 말명료도와 강한 상관관계를 가지고 있을 뿐만 아니라(홍새미, 정필연, 심현섭, 2018; 장선정 & 최현주, 2016; 송한내 외., 2013), 객관적인 수치로 표현하기에 용이하다는 장점이 있다. 국내에서는 Assessment of Phonology and Articulation for Children:APAC(Kim, Pae, & Park, 2007) 또는 Urimal Test of Articulation and Phonation:U-Tap(Shin & Kim, 2007) 표준화 도구를 사용하여 자음정확도를 산출하고 장애 중증도를 판단한다. 구체적으로, 화자에게 다양한 음소 환경의 자음이 포함된 단어 및 문장을 발화하게 한 다음 정조음한 자음의 개수를 세는 방식으로 진행된다. 그러나 자음정확도는 화자의 조음

능력, 그 중에서도 자음 산출 능력만을 반영한다는 단점이 있다. 자음정확도 기반의 장애 중증도 판단 기준은 표 2.2과 같다.

표 2.2 자음정확도 기반 장애 중증도 기준

자음정확도	장애 중증도
85~100%	경도
65~84.9%	경도-중등도
50~64.9%	중등도-중도
50% 이하	중도

## 2.2 마비말장애 음성의 발음 정확도 특징

마비말장애 음성의 발음 정확도 특징은 부정확한 음소 산출이다. Clarke & Hoops(1980)는 뇌성마비로 인한 마비말장애 아동 50명의 전체 음소 오류의 출현 빈도와 말명료도를 분석하고, 두 변수 간 높은 상관관계를 보고하였다. Whitehill & Ciocca(2000)는 마비말장애인 22명의 발화에 나타난 개별 조음오류 유형의 출현 비율을 분석하였다. 분석 결과, 음소 대치 오류가 70.29%로 가장 빈번하게 출현하였으며, 생략, 왜곡, 축약, 첨가 순으로 오류 빈도가 높았다.

자음정확도는 특히 마비말장애인의 장애 중증도 예측에 대한 강력한 변수이다. 이영미 외(2012)는 장애 중증도 고집단 44명, 중집단 33명, 저집단 23명을 대상으로 집단 간 자음정확도를 비교분석하였다. 분석 결과, 장애중증도가 심한 집단일수록 자음정확도가 낮게 나타났으며, 자음의 대치와 생략은 말명료도를 유의하게 예측하는 변수였다. 또한 자음은 조음위치 및 조음방법에 따라서도 정확도에 차이가 난다. Plaat, Andrews & Howie(1980)은 뇌성마비로 인한 마비말장애인 50명에 대한 자음 오류 빈도를 확인했을 때 조음 위치에서는 순치음과 치조음에서, 조음 방법에서는 마찰음과 파찰음에서 높은 오류 빈도가 나타남을 보고하였다. 이영미, 성지은, 심현섭(2013)은 뇌성마비로 인한 마비말장애 성인 100명의 자음 오류 양상을 분석하였는데, 자음 /ㄴ/가 99.3%로 정확도가 가장 높았던 반면 초성 자음 /ㅍ/의 정확도는 52.3%로 가장 낮았다. 더 나아가,

중성 자음의 생략 오류가 초성 자음에 비해 4배 정도 높게 출현하였다. 초성 자음은 다른 자음으로 대치하는 반면 중성 자음은 생략하는 양상을 보였기 때문이다. 이는 자음의 대치와 생략이 말명료도를 유의미하게 예측할 수 있다는 이영미 외(2012)와도 일맥상통한 결과이다.

마비말장애 음성의 또 다른 발음 정확도 특징으로는 모음왜곡이 있다. 자음에 비해 상대적으로 조음이 쉬운 모음은 다른 모음으로 대치되기보다 왜곡되는 경향을 보인다. 구체적으로, 마비말장애 화자는 조음기관의 운동력이 줄어들기 때문에 중증도가 심해질수록 모음 공간 면적 축소, 모음 포먼트 중앙화, 모음 대비 감소가 일어난다. Kim, Kent, & Weismer(2011)은 마비말장애 화자 107명의 모음 산출을 분석하고, 모음 공간 축소가 말명료도 예측에 대한 유의한 변수라고 보고하였다. 김성윤, 김정환, & 고도홍(2014)은 마비말장애 화자 15명과 비장애 화자 15명을 대상으로 모음 공간 면적, 모음 포먼트 중앙화, 모음 대비를 고립 환경과 문맥 환경에서 비교분석하였다. 맨-휘트니 U 검정(Mann-Whitney U test) 결과, 마비말장애 화자의 모음공간 면적은 고립 환경과 문맥 환경 모두에서 유의미하게 작게 나타났고, 마비말장애인의 모음 포먼트 중앙화는 고립 환경과 문맥 환경 모두에서 나타났으며, 마비말장애인의 모음간 대비는 비장애인보다 유의하게 작았다. 모음왜곡 특징과 관련된 구체적인 내용은 3.4.2에서 확인할 수 있다.

## 2.3 마비말장애 중증도 자동 분류

마비말장애 중증도 자동 분류 연구는 머신러닝 방법론을 중심으로 진행되었다. 머신러닝 방법론에서는 집단 간 차이를 반영하는 최적의 특징셋을 구성하는 것이 매우 중요하다. 선행연구 역시 장애 중증도의 기준인 말명료도에 영향을 미치는 요소를 중심으로 특징 셋을 제안하였다.

Narendra & Alku(2020)은 openSMILE에서 추출한 스펙트럼 특징과 성문 특징의 조합으로 마비말장애 자동 검출을 제안하였다. 스펙트럼 특징으로 구성된 베이스라인은 74.63%의 분류 정확도를 보인 반면, 베이스



라인에 성문 특징을 추가했을 때 91.38%의 정확도를 보였다. 실험에서 사용된 스펙트럼 특징은 MFCCs, 영점교차율(zero-crossing rate), 로그 에너지, 멜 스펙트럼(mel spectrum), 스펙트럴 플럭스(spectral flux), 스펙트럴 센트로이드(spectral centroid), 스펙트럴 롤오프(spectral rolloff) 등이 있다. 각 스펙트럼 특징은 25ms의 프레임 단위로 추출된 후 프레임 간 평균, 표준편차, 왜도, 첨도 등으로 계산되어 사용되었다. 프레임 특징은 각 음성의 길이에 따라 추출되는 개수가 달라지는데, 머신러닝 분류기는 일정한 길이의 특징만을 입력값으로 사용할 수 있기 때문이다. 이처럼 스펙트럼 특징은 음성의 전반적인 특징을 반영할 수는 있지만, 음소 단위의 특징이 명시적으로 반영하지 못할 것을 예상할 수 있다.

Hernandez, Kim, & Chung(2020)은 스펙트럼 특징(MFCCs), 음질 특징, 운율 특징의 조합을 사용하였다. 운율은 발화속도, 음높이, 리듬 특징을 포함하였다. MFCCs를 베이스라인으로 삼고, 음질 특징과 운율 특징을 추가했을 때의 성능을 비교하였다. MFCCs를 단독으로 사용했을 때 60%, 음질 특징과 운율 특징을 각각 추가했을 때는 64.11%, 66.67%, 음질 특징과 운율 특징을 추가했을 때 70.1%로 가장 높은 정확도를 보였다. 이때, 특징 선택 알고리즘을 적용하여 분류에 중요한 특징만을 사용하였는데, 운율의 발화속도, 음높이, 리듬 모두 중요한 특징값으로 포함되었다. 이 연구 역시 앞의 연구와 마찬가지로 발음 정확도 특징을 반영하지 않았다는 한계가 있다.

반면, 발음 정확도 특징을 장애 중증도 자동 분류에 사용한 연구들도 있다. 발음 정확도 특징은 음소정확도와 모음왜곡도로 나누어 살펴보았다. 우선, Kim, Kim & Kim(2015)는 음소정확도를 특징으로 사용하였다. 먼저, 음성인식기가 예측한 발음열과 실제 발음열이 음소 단위로 강제정렬되었다. 만약 강제정렬된 두 음소가 다를 경우, 해당 오조음 사례는 히스토그램에 추가되었다. 마지막으로 모델은 히스토그램을 바탕으로 화자별 음소 오조음 패턴을 학습하여 말명료도를 예측하였다. 이 연구는 음소정확도를 음성인식기의 결과값을 기반으로 자동 산출하였다는 의의가 있지만, 발음 외의 특징인 음질이나 운율 등의 특징을 고려하지 않았다

는 단점이 있다.

스펙트럼 특징, 음질 특징, 운율 특징과 음소정확도 특징을 함께 사용한 연구도 있었다. Tripathi, Bhosale, & Kopparapu(2020)는 openSMILE 툴킷으로 추출한 스펙트럼 특징들과 end-to-end 음성인식기 모델인 DeepSpeech의 음소 단위 출력값을 사용하였다. 이 실험에서 사용된 openSMILE 특징은 스펙트럼 특징 외에도 F0, HNR 등 운율과 음질 특징을 포함하였다. 실험은 화자 중속 실험과 화자 독립 실험으로 진행되었다. 화자 중속 실험은 모델 훈련과 테스트에 같은 화자가 포함되는 반면, 화자 독립 실험은 모델 훈련과 테스트에 다른 화자가 포함된다. 실험 결과, openSMILE 특징 또는 DeepSpeech 출력값을 단독으로 사용했을 때보다 두 특징 모두를 사용했을 때 가장 좋은 성능을 보였다. 화자 중속 실험은 openSMILE, DeepSpeech, openSMILE+DeepSpeech 특징을 사용했을 때 각각 37.2%, 42.5%, 53.9% 정확도를 보였다. 화자 독립 실험은 94.9%, 95.6%, 97.4%의 정확도를 보였다. 이 결과는 음성인식기 출력값으로 산출된 음소정확도가 화자의 전반적 정보가 아닌, 화자의 장애 중증도에 대한 정보를 제공함을 시사한다. 이 연구는 장애 중증도 자동 분류에서 음성인식 기반의 음소정확도 특징뿐만 아니라 스펙트럼 특징, 음질 특징, 운율 특징을 함께 고려했다는 의의가 있다. 그러나 음소정확도 특징이 다른 특징과 어떤 관계에 있는지, 즉 어떤 특징이 특히 자동 분류에 도움이 되는지에 대한 분석이 이루어지지 않았다는 한계가 있다.

두 번째 발음 정확도 특징인 모음왜곡도를 사용한 연구도 있다. Lansford & Liss(2014)는 모음왜곡도 특징을 바탕으로 마비말장애 자동 검출을 시도하였다. 이때 특징을 모음공간 면적(Vowel Space Area; VSA), Formant Centralized Ratio(FCR), Vowel Articulatory Index(VAI), F2-slope이 속한다. 먼저, 다양한 모음이 포함된 구를 낭독한 자료에서 숙련된 언어재활사들이 분석할 모음 구간을 태깅하였다. 그 다음 태깅된 모음 구간의 F1, F2를 바탕으로 모음왜곡도 특징이 추출되었다. 추출된 모음왜곡도는 t-검정 분석을 통해 비장애 발화와 장애 발화 간 차이의 유의미성이 판단되었다. 통계 분석 결과, 모음 공간 면적,

F<sub>1</sub>-Score, VAI는 유의미한 차이를 보였지만, F2-Ratio는 유의미한 차이를 보이지 않았다. 마지막으로 판별함수분석(DFA)를 바탕으로 각 특징들이 마비말장애 음성 자동 검출에 얼마나 효과적인지 살펴보았다. 그 결과, 모음공간 면적이 80.7%의 정확도로 가장 높은 성능을 보였고, 다른 특징치들은 66.7%에서 73.7% 사이의 정확도를 보였다. 이 연구는 모음왜곡도 특징 단독으로 사용한 연구로, 음소정확도, 음질, 운율 등 다양한 특징들의 조합을 고려하지 않았다는 한계가 있다.

본 논문에서는 Hernandez, Kim, & Chung(2020)에서 사용한 특징 셋(스펙트럼 특징, 음질 특징, 운율 특징)을 베이스라인으로 삼고, 발음 정확도 특징-음소정확도와 모음왜곡도-을 추가하여 장애 중증도 자동 분류에 대한 발음 정확도 특징의 유용성에 대해 알아본다. 구체적으로, 발음 정확도 특징이 마비말장애 중증도 자동 분류에 도움이 되는 특징인지 확인하고, 발음 정확도 특징과 음질, 운율 특징이 장애 중증도 자동 분류 과제에서 어떤 관계에 있는지 분석한다.

### 3. 실험 방법론

#### 3.1 실험 설계

실험의 전체적 설계도는 그림 3.1과 같다. 먼저, 13차원의 MFCCs와 음질 특징, 운율 특징, 발음 정확도 특징을 추출한다. MFCCs 특징을 제외한 음질 특징, 운율 특징, 발음 정확도 특징은 특징 선택 알고리즘 RFE(Recursive Feature Elimination) 또는 ETC(Extra Trees Classifier)에 의해 추려진다. 선택된 특징들은 MFCCs와 함께 SVM, MLP 분류기의 입력값으로 사용된다. 분류기는 주어진 특징들을 바탕으로 음성을 비장애, 경도, 경도-중등도, 중등도-중도, 중도로 분류하도록 학습된다.

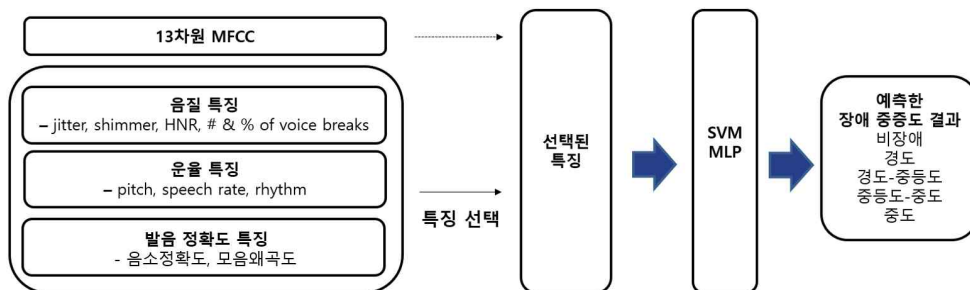


그림 3.1 실험 설계도

#### 3.2 특징 정의 및 특징 추출방법

3.2장에서는 본 논문에서 사용한 특징들을 정의하고, 각 특징의 추출 방법을 제시한다. MFCCs, 음질 특징, 운율 특징, 발음 정확도 특징으로 구분되며, 운율 특징은 발화 속도 특징, 음높이 특징, 리듬 특징을, 발음 정확도 특징은 음소정확도 특징과 모음왜곡도 특징을 포함한다.

### 3.2.1 Mel Frequency Cepstral Coefficients(MFCCs)

본 연구에서는 13차원의 MFCCs를 베이스라인 특징 중 하나로 사용하였다. MFCCs는 소리의 고유한 특징을 나타내는 수치로, 음성인식, 음성합성 등 오디오 도메인에서 기본적으로 사용하는 음성 특징이다. 그림 3.2는 MFCCs 추출과정에 대한 도식이다.



그림 3.2 MFCCs 추출 과정

먼저, 입력 음성 신호의 고주파 영역을 pre-emphasis 필터를 통해 증폭시킨다. 이를 통해 상대적으로 에너지 크기가 작은 고주파 영역과 저주파 영역 간 에너지 균형을 맞출 수 있다.

다음으로 framing 과정을 거친다. Framing 과정은 음성 신호를 일정한 길이의 짧은 구간으로 분절하는 과정으로, 시간에 따라 변화하는 음성 신호의 특징을 추출하는 단계이다. 이때 프레임의 길이가 지나치게 길 경우 변화하는 신호의 주파수 정보를 추출하기 어렵다. 따라서 적절한 길이의 프레임을 사용하는 것이 중요하다. 일반적으로 특징은 시간 단위 25ms, 이동 구간을 10ms의 연속된 프레임을 통해 추출된다.

Framing 과정을 통해 분절된 구간은 windowing 과정을 거친다. Windowing이란 각 프레임에 윈도우 함수를 적용해 각 프레임의 경계를 스무딩하는 방법이다. 스무딩이란 각 프레임에서 추출한 정보가 프레임의 양끝에서 갑자기 0이 되어 끊기지 않고 서서히 줄어들도록 처리하는 과정을 의미한다. 윈도우 함수의 종류는 사각형 윈도우, 해밍 윈도우, 해닝 윈도우 등 다양하지만 가장 일반적으로 해밍 윈도우가 사용된다. 해밍 윈도우는 프레임 중간의 값들은 그대로 사용하는 반면 양끝 부분은 0에 가까운 값으로 사용하는 함수이다. 해밍 윈도우의 그래프와 수식은 그림 3.3과 같다.

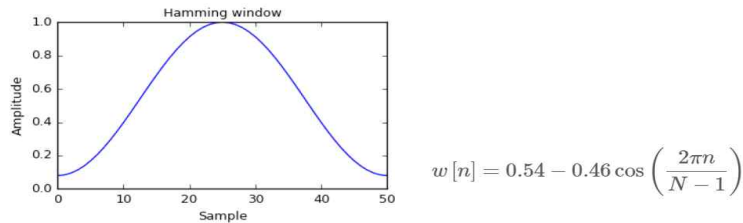


그림 3.3 해밍 윈도우의 그래프 및 수식

그 다음 FFT(Fast Fourier Transform-고속 푸리에 변환)를 적용한다. 푸리에 변환이란 시간 도메인의 음성 신호를 주파수 도메인으로 바꾸는 과정이다. FFT를 적용한 결과물을 spectrum이라고 부른다.

다음으로 spectrum에 mel filter bank를 적용한다. 구체적으로, 삼각형 모양의 mel scale filter 40개를 적용하는 과정이다. Mel scale filter는 인간이 저주파수 영역대 음성은 촘촘하게 지각하는 반면 고주파수 영역대 음성은 대략적으로 지각하는 것을 반영한 필터이다. 그림 3.4는 mel scale filter 그림이다. Mel filter bank을 적용한 결과물은 Mel spectrum이라고 부른다.

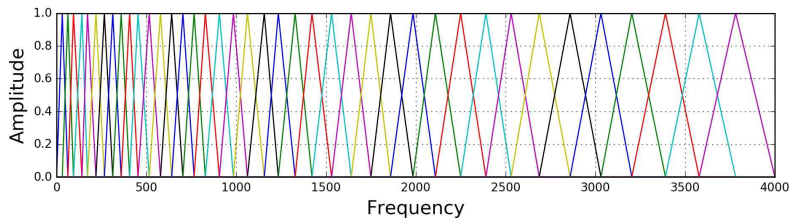


그림 3.4 mel scale filter<sup>1)</sup>

Mel spectrum에는 log scale을 적용한다. Log scale은 인간의 말소리 인식에 중요한 특징을 강조하여 추출하게 한다. 이에 대한 결과물을 Log-Mel spectrum이라고 한다.

마지막으로 Log-Mel spectrum에 DCT(Discrete Cosine Transform)를 적용한다. Log-Mel spectrum의 특징 간에는 상관관계가 있다. 이러한 상관관계는 변수 간 독립을 가정하고 모델링하는 이후 단계에서 문제

1) <https://ratsgo.github.io/speechbook/docs/fe/mfcc>

로 작용할 수 있다. 따라서 Log-Mel spectrum은 DCT 과정을 통해 특징 간 상관관계를 해소하는 과정을 거친다. 이렇게 추출된 특징이 바로 MFCCs 계수 특징이다. 일반적으로 MFCCs 계수 중 앞 12개와 로그 에너지 1개를 합한 13개가 사용되며, 연구에 따라 13개 값에 대한 delta 값과 delta-delta 값을 합한 39개가 사용된다.

본 연구에서는 librosa 툴킷을 이용하여 13차원의 MFCCs를 추출하였다. 이렇게 추출된 MFCCs 값은 발화 별로 프레임의 평균값으로 사용되었다. 그 결과, 본 연구에서는 발화 별 13개의 MFCCs 특징이 사용되었다.

### 3.2.2 음질

음질은 성도의 크기나 모양 및 신경학적 요인과 관련된 말의 특성으로, 성대 진동의 규칙성으로 판단할 수 있다. 마비말장애 화자의 경우 후두계 조절에 장애가 발생하여 음질과 같은 발성에서도 부정적인 영향을 받는다(서인호와 성철재, 2013). 따라서 본 연구에서는 음질 특징으로 jitter, shimmer, Harmonic to Noise Ratio(HNR), 끊어진 유성음 구간 개수(# of voice breaks), 끊어진 유성음 정도(% of voice breaks)를 사용하였다.

Jitter, shimmer, HNR은 언어 병리학 분야에서 음성 장애를 진단하는 척도로 사용된다. Jitter는 주파수 변동률로 단위 시간 내 F0의 변화가 얼마나 규칙적인 주기를 보이는가를, shimmer는 진폭 변동률로 음성 강도가 얼마나 규칙적인 주기를 보이는가를 평가하는 지표이다. 마지막으로 HNR은 소음 대 배음 비율로 소음 에너지에 비교했을 때 배음의 에너지가 차지하는 비율을 의미한다.

유성음의 끊어진 구간(voice break) 특징은 마비말장애 화자가 비장애 화자에 비해 유성음 구간을 유지하는 것에 어려움을 겪는 것을 반영한다. 본 연구에서는 끊어진 유성음 구간의 개수(# of voice breaks)와 끊어진 유성음 구간의 비율(% of voice breaks)을 사용하였다.

본 연구에서는 praat 스크립트의 voice report 기능을 사용하여 음질

특징들을 추출하였다. 구체적으로, 끊어진 유성음 구간 개수는 연속된 펄스 사이에 1.25ms 이상의 거리가 있었던 개수로 추출하였고, 끊어진 유성음 구간의 비율은 (끊어진 유성음 구간 길이)/(휴지 구간을 제외한 전체 길이)로 계산하여 사용하였다.

### 3.2.3 운율 특징

마비말장애 화자 중 다수는 다양한 운율 장애를 동반한다(남현욱 외., 2009). 본 연구에서는 다양한 운율 특징을 반영하기 위해 발화 속도, 음높이, 리듬으로 구분하여 사용하였다.

#### 3.2.3.1 발화 속도

마비말장애 화자는 조음기관 사용에 어려움을 겪기 때문에 구어 속도가 전반적으로 느리며, 쉼의 빈도 및 시간이 비정상적인 경향을 보인다(남현욱 외., 2009). 따라서 본 연구에서는 말속도(speaking rate), 조음속도(articulation rate), 말길이(speaking duration), 전체길이(total duration)를 발화 속도 특징으로 사용하였다. 말속도는 전체 음절 개수를 전체 말 길이로 나눈 값을, 조음 속도는 발화 음절 수(전체 음절 수 - 휴지 수)를 말길이(전체길이 - 휴지길이)로 나눈 값을 의미한다. 이때 말길이는 전체 길이 중 휴지 구간을 제외한 구간을, 전체길이는 휴지 구간을 포함한 말 길이를 의미한다.

본 연구에서는 리듬 특징 추출 소프트웨어인 Correlatore를 사용하여 전체 음절 수와 휴지 수를, 음성인식 툴킷인 kaldi를 사용하여 말길리와 전체길이를 추출하였다.

#### 3.3.3.2 음높이

마비말장애는 비장애 화자보다 더 넓은 F0 범위를 가지는 경향을 보인다(Nam & Kwon, 2005). 따라서 본 연구에서는 음높이의 측정치로 F0의 평균값, 표준편차, 최소값, 최대값, 중앙값, 25 분위수, 75 분위수를 사용하였다. 해당 특징들은 praat 스크립트를 사용하여 추출하였다.



### 3.3.3.3 리듬

마비말장애 화자는 비장애 화자와 비교했을 때 상이한 리듬 특징을 보일 뿐만 아니라 장애 중증도를 비교할 때에도 유의미한 도움이 된다 (Hernandez et al., 2020). 본 연구에서는 리듬 특징으로 %V, deltas(deltaV, deltaC), Varcos(Varco-V, Varco-C), rPVI(VrPVI, CrPVI), nPVI(VnPVI, CnPVI)를 사용하였다.

%V는 음성 전체 중 모음이 차지하는 비율을 의미하며, deltaV와 deltaC는 델타값으로 각각 모음과 자음 길이의 표준편차로 산출된다. Varco-V와 Varco-C는 발화 속도에 영향을 받는 델타값들을 정규화하기 위해 제안된 값으로, 각각 델타값을 모음 또는 자음의 평균 길이로 나눈 값이다. 수식은 수식 3.1과 같다.

$$\text{Varco-V} = \frac{\Delta V * 100}{\text{mean}(V)} \quad \text{Varco-C} = \frac{\Delta C * 100}{\text{mean}(C)}$$

수식 3.1 Varco-V와 Varco-C 수식

다음으로 PVI(Pairwise Variability Index) 측정치이다. 우선 rPVI는 raw PVI를 뜻하는 용어로, VrPVI는 연속된 모음 간의 간격을, CrPVI는 연속된 자음 간의 간격을 의미한다. nPVI는 normalized PVI를 뜻하며, 각 rPVI를 정규화한 값이다. rPVI와 nPVI를 산출하는 수식은 수식 3.2에 제시하였다.

$$\text{rPVI} = \sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m-1)$$

$$\text{nPVI} = 100 * \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right| / (m-1)$$

수식 3.2 rPVI와 nPVI 수식

리듬 특징들은 모두 리듬 특징 추출 소프트웨어인 Correlatore를 이용하여 추출하였다.

### 3.2.4 발음 정확도 특징

마비말장애 화자는 비장애 화자와 비교했을 때 부정확한 음소를 산출하는 경향을 보이며, 장애 중증도가 심해질수록 그 산출 정확도는 떨어진다. 발음 정확도 특징은 음소정확도와 모음왜곡도로 구분된다.

#### 3.2.4.1 음소정확도

마비말장애 화자는 부정확한 자음과 부정확한 모음을 산출하는 경향이 있다. 따라서 본 연구에서는 자음정확도(Percentage of Correct Phonemes; PCC), 모음정확도(Percentage of Correct Vowels; PCV), 그리고 자음과 모음 모두를 고려한 음소정확도(Percentage of Correct Phonemes; PCP)를 사용하였다.

자음정확도는 전체 목표 자음 개수 중에서 정조음한 자음 개수의 비율을, 모음정확도는 전체 목표 모음 개수 중에서 정조음한 모음 개수의 비율을, 음소정확도는 전체 목표 음소 개수 중에서 정조음한 음소 개수의 비율을 의미한다. 그림 3.5는 자음정확도, 모음정확도, 음소정확도를 산출하는 예시를 제시한다.

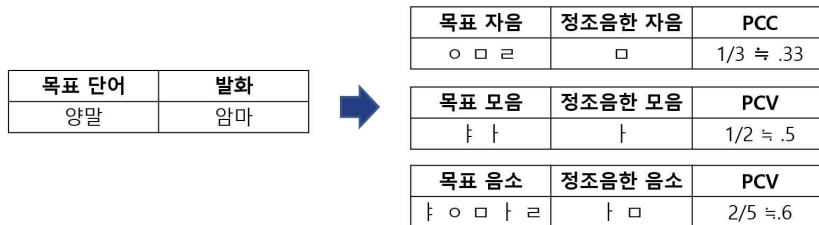


그림 3.5 자음정확도, 모음정확도, 음소정확도 산출 예시

자음정확도, 모음정확도, 음소정확도는 비장애인 대상 대어휘 연속음 성인식기를 사용하여 산출되었다. 음소 단위로 디코딩한 음성인식기의 결과를 바탕으로 전체 목표 음소 개수 중에서 정확하게 인식된 음소의 개수를 세었다. 음성인식기 모델은 비장애 음성 6,000시간의 음향모델과 약 370만개의 어휘 발음사전으로 구성하였으며, 발음 사전에 사용된 음소 개수는 40개이다. 언어 모델은 다음 음운에 대한 예측을 방지하기 위

해 monogram을 사용하였다.

음성인식기 사용의 타당성을 살펴보기 위해 전문가에 의해 산출된 자음 정확도와 음성인식기를 사용하여 산출한 자음 정확도의 상관관계를 살펴본 결과, 두 측정치는 0.58로 유의미한 상관관계를 보였다.

### 3.2.4.2 모음 왜곡도

마비말장애 화자는 조음기관의 운동력이 줄어들기 때문에 중증도가 심해질수록 모음 공간 면적 축소, 모음 포먼트 중앙화, 모음 대비 감소가 일어나는 경향을 보인다. 모음 공간은 성도(vocal tract)를 포먼트의 2차원적인 그림으로 나타낸 것으로, 혀의 높낮이를 나타내는 F1과 혀의 전후 위치를 나타내는 F2의 수치를 이용하여 나타낸다. 모음 포먼트 중앙화는 모든 포먼트의 수치가 슈와(schwa)를 목표점으로 모이는 현상을 의미한다. 마지막으로 모음 대비 감소는 개별 모음의 포먼트 간 거리가 줄어들러는 현상을 뜻한다. 본 논문에서는 모음 공간 면적(Vowel Space Area; VSA), Formant Centralized Ratio(FCR), Vowel Articulatory Index(VAI), F2-Ratio 네 개의 특징을 사용하였다(김성운, 김정환, & 고도홍, 2014; Kang, Yoon, Lee, & Seong, 2010). 단, 코퍼스 내 문장의 한계로 모든 특징의 수식에서 /우/가 /오/로 대체되어 사용되었다. 한국어의 /오/는 /우/와 동일하게 원순후설모음으로서 한국어의 꼭지모음으로 기능할 수 있다.

모음 공간 면적 수식은 수식 3.3과 같다. F1-F2 모음 사각형의 면적을 구하는 공식으로, 불규칙 사각형의 면적을 구한다. 사각형의 두 꼭지점에서 두 수선이 교차하는 점을 기준으로 네 개의 직각 삼각형으로 구역을 나눈 다음, 각 구역의 면적 공식을 합하는 공식이다.

#### 모음 공간 면적 (VSA)

$$=1/2*[(F2/이/*F1/애/+F2/애/*F1/아/+F2/아/*F1/오/+F2/오/*F1/이/)- (F1/이/*F2/애/+F1/애/*F2/아/+F1/아/*F2/오/+F1/오/*F2/이/)]$$

수식 3.3 모음 공간 면적 수식

FCR과 VAI는 모음 포먼트 중앙화를 반영하는 특징이다. 모음이 중앙화되면 FCR 수치가 증가하고 VAI가 감소하는 반면, 모음이 바깥으로 확장되면 FCR 수치가 감소하며 VAI는 증가한다.

F2-ratio는 모음 대비 감소에 대한 특징이다. 본 연구에서는 평순 모음의 대표인 /이/와 원순모음의 대표인 /오/의 F2 값의 비율을 사용하였다. F2-ratio는 조음의 움직임이 부정확할수록 감소하고, 조음 움직임이 개선될수록 증가한다(강영애 외., 2010). FCR, VAI, F2 ratio는 각각 수식 3.4에 제시하였다.

$$\text{FCR} = (F2/\text{오}/ + F2/\text{아}/ + F1/\text{이}/ + F1/\text{오}/) / (F2/\text{이}/ + F1/\text{아}/)$$

$$\text{VAI} = (F2/\text{이}/ + F1/\text{아}/) / (F1/\text{이}/ + F1/\text{오}/ + F2/\text{오}/ + F2/\text{아}/)$$

$$\text{F2 ratio} = (F2/\text{이}/) / (F2/\text{오}/)$$

#### 수식 3.4 FCR, VAI, F2-Ratio 수식

모음 왜곡도 특징은 음성인식기와 praat 스크립트를 통해서 추출되었다. 음성인식기를 이용해 음성을 음소 단위로 강제정렬하고, praat 스크립트를 이용해 필요한 모음 구간의 중앙에서 F1, F2 포먼트를 추출하였다. 한 문장에 해당 모음이 여러 개 있을 경우, 모음들의 평균값을 사용하였다. 이때, 여성 음성은 최대 포먼트 주파수를 5500Hz, 남성 음성은 5000Hz로 설정하여 추출하였다.

### 3.3 특징 선택 알고리즘

특징 선택의 목적은 추출된 특징 중 학습에 불필요한 특징을 제거하는 것에 있다. 특징 선택은 사람이 배경 지식을 바탕으로 이루어질 수도 있지만, 특징 선택 알고리즘을 사용할 수도 있다. 본 연구에서는 scikit-learn 라이브러리에서 제공하는 Recursive Feature Elimination(RFE)과 Extra Trees Classifier(ETC)를 사용하였다.

### 3.3.1 Recursive Feature Elimination(RFE)

Recursive Feature Elimination(RFE)은 우선 특징들을 모두 포함시킨 다음 중요도가 낮은 특징들을 하나씩 제거하며 반복 학습하면서 중요한 특징들을 골라내는 알고리즘이다. 중요도는 Support Vector Regression(SVR) 또는 로지스틱 회귀(Logistic Regression)를 바탕으로 산출된다. SVR는 서포트 벡터 머신(Support Vector Machine; SVM)의 손실함수를 사용하여 회귀식을 구성하는 방식으로, SVM에 대한 설명은 3.4.1에서 설명한다. 로지스틱 회귀는 독립 변수와 종속 변수 간의 관계를 예측 모델에 사용하는 방법이다.

본 논문에서는 실험 결과 두 기법 중 더 높은 성능을 보인 로지스틱 회귀를 사용하였다. 반복 학습이 진행될수록 특징 셋의 크기는 점점 작아지며, 학습은 하이퍼파라미터로 설정한 특징 개수에 도달할 때까지 진행된다. 본 논문은 RFECV를 사용하여 최적의 특징 개수를 결정하였다. 그림 3.6은 특징 개수에 따른 정확도를 나타낸 그래프이다.

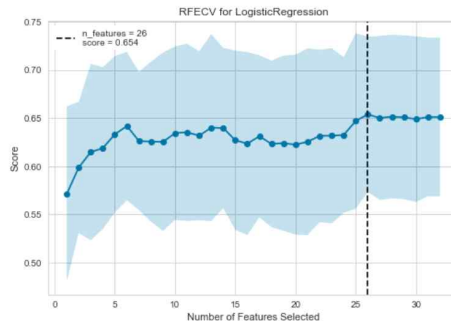


그림 3.6 특징 개수에 따른 정확도(RFE)

### 3.3.2 Extra Trees Classifier(ETC)

Extra Trees Classifier(ETC)는 앙상블 머신러닝 모델로, 분류 모델인 랜덤 포레스트(Random Forest; RF)의 변형이다. RF는 여러 개의 의사 결정 트리(decision tree)를 형성하고, 각 의사 결정 트리로부터 분류 결과를 취합하여 결론을 내리는 모델이다. 단, RF는 최적의 임계값을 바탕으로 데이터를 나누는 반면 ETC는 후보 변수를 임계값으로 사용하여 데이터를 무작위로 나눈 후 최상의 분할을 선택한다는 차이가 있다. 이

처럼 ETC 모델은 불순도(impurity)를 낮추는 방향으로 학습이 진행된다. 이때 불순도란 각 범주 안에 다른 범주의 데이터가 섞여 있는 정도를 의미한다. 정리하자면, ETC 알고리즘은 불순도를 낮출 수 있는 특징 순서대로 높은 가중치를 부여하여 중요한 특징들을 추려내는 알고리즘이다.

## 3.4 머신러닝 분류기

### 3.4.1 Support Vector Machine(SVM)

서포트 벡터 머신(Support Vector Machine; SVM)은 분류 과제에 사용되는 머신러닝 지도학습 모델로, 장애 발화처럼 소규모의 데이터 분류 과제에서 자주 사용된다(Hernandez, 2020; Narendra & Alku; 2018; Kadi, Boudraa, & Boudraa, 2013). SVM은 구분하고자 하는 두 집단의 거리인 마진(margin)을 최대화하는 결정 경계(decision boundary)를 찾는 모델이다. 마진은 결정 경계에서 가장 가까운 데이터인 서포트 벡터(support vector) 간의 거리로 산출된다. 마진을 최대화하는 최적의 결정 경계를 학습한 SVM은 이 결정 경계를 기준으로 새로운 입력값이 어느 집단에 속하는지 분류한다. 그림 3.7은 SVM의 한 예시이다.

SVM의 중요한 하이퍼파라미터는 C와 gamma가 있다. C는 마진의 손실함수와 관련된 값으로, 결정 경계 결정 시 잘못 분류된 데이터에 대한 페널티를 얼마나 줄 것인지를 의미한다. C가 작을수록 잘못 분류된 데이터에 대해 작은 페널티가 부여되며, 그에 따라 모델은 유연한 결정 경계를 찾는다. 유연한 결정 경계를 찾는 이유는 학습 데이터에만 적합하고 새로운 데이터에는 대처하지 못하는 과적합(overfitting) 현상을 방지하기 위함이다. 반면 C가 너무 작을 경우 분류 정확도가 떨어진다. 따라서 다양한 실험을 통해 적절한 C의 값을 찾는 것이 중요하다.

Gamma는 커널(kernel) 함수와 관련된 하이퍼파라미터이다. 커널 함수란 저차원의 데이터를 고차원으로 매핑하여 비선형 데이터를 선형 분류할 수 있도록 하는 함수이다. 단 이 함수는 실제로 데이터를 새로운 고차원으로 변형하는 대신, 데이터 간 유사도를 바탕으로 결정 경계를

산출한다. Gamma는 커널 함수의 표준편차를 결정하는데, 이때 gamma가 클수록 커널 함수의 표준편차는 작아진다. 즉, gamma가 커질수록 한 훈련 데이터 샘플이 미치는 영향력은 작아지고, 데이터들 간의 거리가 매우 가까워야만 하나의 집단으로 묶일 수 있다. Gamma의 값이 너무 크다면 과적합을 방지할 수 없지만, 반대로 gamma의 값이 너무 작다면 복잡한 데이터의 분포를 학습할 수 없다. 따라서 gamma 역시 다양한 실험을 통해 최적화하여야 한다.

본 논문에서는 SVM의 C와 gamma를 그리드 서치(grid search)를 이용하여 최적화하였다. 파라미터의 범위는  $10^{-4}$ 에서  $10^4$  사이였다. 가능한 커널 함수인 poly, rbf(radial basis function), sigmoid를 모두 실험한 결과, 가장 좋은 성능을 보인 커널함수는 rbf였다. 마지막으로, 다분류(multiclass classification) 실험을 위해 one-versus-one 옵션을 사용하였다. One-versus-one 옵션이란  $n*(n-1)/2$  개의 분류기를 형성하여 각 집단마다 다른 집단과 한 번씩 모두 비교하는 방식이다.

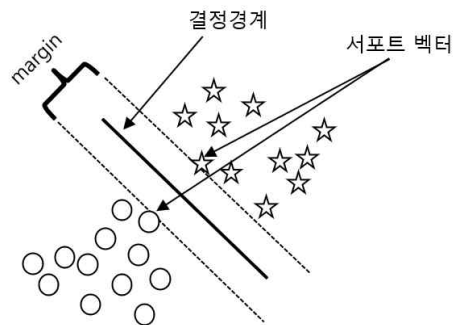


그림 3.7 서포트 벡터 머신(SVM)

### 3.4.2 Multiple Layer Perceptron(MLP)

다층 퍼셉트론(Multiple Layer Perceptron; MLP)은 인공신경망에 해당하는 분류 모델이다. MLP는 입력층, 은닉층, 출력층으로 구성되며, 입력층에서 전달되는 값이 은닉층의 모든 노드로 전달되고, 또 은닉층의 모든 노드의 출력값이 출력층의 모든 노드로 전달된다. MLP의 목적은 입력값을 사용하여 정답 레이블을 가장 잘 예측할 수 있는 weight를 구

하는 것이다. Weight 값은 역전파(back-propagation) 과정을 거쳐 결정되는데, 과정은 다음과 같다. 먼저, weight 값을 랜덤으로 정한 상태에서 학습 데이터를 이용하여 출력값을 구한다. 다음으로 모델의 출력값과 정답 레이블에 대한 차이를 계산한다. 마지막으로 모델이 예측한 출력값과 정답 레이블 간의 차이가 감소하는 방향으로 weight 값을 조정한다. 이 세 단계를 반복함으로써 모델은 weight 값을 최적화할 수 있다. 그림 3.6은 MLP의 예시를 나타낸 것이다.

MLP의 최적화에 중요한 하이퍼파라미터는 은닉층 개수, 학습률(learning rate), 옵티마이저(optimizer), 활성화 함수(activation function)가 있다. 옵티마이저는 학습 방법을 결정하며, 종류로는 SGD(Stochastic Gradient Descent), lbfgs(limited-memory quasi-Newton methods), adam 등이 있다. 활성화 함수는 신경망의 노드에 적용하는 비선형 함수로, 종류로는 logistic, tanh, relu 함수 등이 있다.

본 논문에서는 MLP의 은닉층, 학습율, 옵티마이저, 활성화 함수를 그리드 서치를 이용하여 최적화하였다. 그 결과, 은닉층 3개, 학습률 0.001, adam 옵티마이저, ReLu 활성화 함수를 사용한 모델이 가장 높은 성능을 보였다.

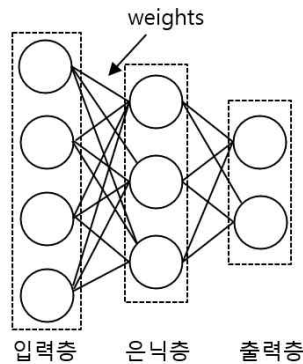


그림 3.8 다층 퍼셉트론(MLP)



## 4. 데이터베이스

### 4.1 QoLT(Quality of Life Technology) 코퍼스

본 연구에서 사용한 음성 코퍼스는 QoLT 코퍼스로, 2011년부터 2014년까지 구축된 마비말장애인 대상의 음성인식기 개발 목적 코퍼스이다 (Choi et al., 2012). 단어 목록은 APAC 단어 37개, VOT 평가 단어 9개, 한국어 자음 및 모음 조사를 위한 목록인 문장 5개와 단어 112개로 구분된다. APAC과 VOT 평가용은 각 1회, 자음 및 모음 조사를 위한 목록은 각 2회 반복하였다.

본 논문에서 사용한 데이터베이스는 QoLT\_2012로, 비장애인 10명과 마비말장애인 70명의 음성으로 구성되어 있다. QoLT\_2012 데이터베이스를 선택한 이유는 QoLT 코퍼스 내 데이터베이스 중 가장 다수의 마비말장애 화자의 연속음성을 포함하고 있기 때문이다.

본 논문에서는 연속음성의 특징 분석을 위해 여러 발화 목록 중 문장 자료만을 사용하였다. 또, 모음왜곡도 계산을 위해 모음 /ㄴ/를 포함하고 있지 않은 문장은 본 실험에서 제외하였다. 이에 따라 화자 별 4개의 문장을 2회 반복한 자료가 실험에서 사용되었다. 표 4.1은 데이터베이스의 문장 5개를 제시한다.

표 4.1 문장 자료

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

### 4.2 자음정확도 평가

언어치료 경험이 5년 이상이며, 언어치료사 1급 자격을 소지한 언어재활사 2명의 APAC 평가, 즉 자음정확도를 기반으로 각 화자의 장애

중증도를 분류하였다. APAC 평가는 전체 목표 자음 90개 중 화자가 정조음한 자음의 개수의 비율로 자음정확도를 계산한다. 평가 결과, 경도 화자, 경도-중등도 화자, 중등도-중도 화자, 중도 화자는 각각 51명, 13명, 3명, 3명이었다. 장애 중증도의 기준은 2.1.2장에 제시한 기준이 사용되었다.

### 4.3 말명료도 평가

마비말장애 중증도의 또 다른 기준인 말명료도에 기반하여 각 화자의 장애 중증도를 분류하였다. 말명료도 평가는 신뢰도 및 타당도가 가장 높은 5점 척도로 진행되었으며(박혜진, 2007), 2년 이상의 임상 경험이 있는 언어재활사 5명이 실시하였다. 평가를 진행한 언어재활사의 정보는 표 4.2와 같다.

표 4.2 언어재활사 정보

#	성별	연령(세)	언어 치료사 자격증	치료 경험	근무지
1	여	26	2급	2년	서울 신경과
2	여	30	1급	2년	서울 듣기뇌발달센터
3	여	31	1급	6년	서울 재활의학과
4	여	30	1급	2년	서울 재활의학과
5	여	27	2급	2년	서울 재활의학과

언어재활사는 화자가 발화한 APAC 단어 37개와 문장 8개를 모두 청취한 다음 화자의 전반적 중증도, 즉 말명료도를 판단하였다. 평가는 신뢰도 분석을 위해 일주일 간격으로 두 차례 진행되었으며, 각 평가마다 화자의 순서는 무선배치되었다.

평가자 신뢰도를 급내상관계수(ICC)로 분석한 결과, 평가자 간 신뢰도는 .95, 평가자 내 신뢰도는 각각 .92, .98, .84, .93, .93로 나타났다. 화자의 중증도에 대한 평가자 간 불일치가 있을 경우, 화자에 대한 10개의 중증도 평가 결과(5명\*2회) 중 최빈치의 점수로 화자의 중증도가 결정되었다. 평가 결과, 경도 화자, 경도-중등도 화자, 중등도-중도 화자, 중도

화자가 각각 24명, 24명, 15명, 7명이었다. 장애 중증도의 기준은 2.1.3장에 제시한 기준을 사용하였다.

#### 4.4 자음정확도 평가와 말명료도 평가 비교

표 4.3은 자음정확도 평가 결과 별 말명료도 평가 결과를 제시한 것이다. 화자 별로 각 평가 결과를 비교한 결과, 전체 화자의 50% 이상이 자음 정확도 평가보다 말명료도 평가에서 더 낮게 평가되었다. 이는 높은 자음정확도로 인해 말명료도가 낮음에도 장애 정도가 과소평가될 수 있음을 시사한다. 또한 이는 말명료도가 분절음의 부정확한 산출뿐만 아니라, 음질과 운율의 영향으로도 저하된다는 선행연구와도 일맥상통한 결과이다(이은경 & 김지채, 2012; 최여진 & 심현섭, 2013; 이옥분, 박상희, & 남현욱, 2012). 따라서 본 논문에서는 자음정확도가 아닌 말명료도를 장애 중증도 자동 분류 기준으로 삼아 실험을 진행하였다.

표 4.3 자음정확도 평가 결과 별 말명료도 평가 결과 분포

자음정확도 평가	말명료도 평가	빈도	비율(%)
경도(1)	1	24	47
	2	18	35
	3	9	18
	4	0	0
경도-중등도(2)	1	0	0
	2	6	46
	3	4	31
	4	3	23
중등도-중도(3)	1	0	0
	2	0	0
	3	1	33
	4	2	67
중도(4)	1	0	0
	2	0	0
	3	1	33
	4	2	67

## 4.5 데이터베이스 통계 분석

4.5장에서는 집단 간 음질 특징, 운율 특징, 발음 정확도 특징에 차이가 있는지 분석하였다. 먼저, 장애 중증도 별 평균을 살펴보았다. 그 다음, 각 특징이 장애 중증도 별 유의미한 차이를 보이는지 확인하기 위해 양방향 혼합 ANOVA(two-way mixed ANOVA) 분석을 실시하였다. 이때 개체 간 요인은 장애 중증도 집단으로, 개체 내 요인은 특징으로 설정하였다. 마지막으로, 각 특징이 장애 중증도 집단 간 유의미한 차이를 보이는지 확인하기 위해 Dunnett T3 사후분석을 실시하였다.

### 4.5.1. 음질 특징

음질 특징인 jitter, shimmer, HNR, # of voice breaks, % of voice breaks를 장애 중증도 집단 별로 비교하였다. 집단 별 음질 특징의 평균은 표 4.4, 그림 4.1에 제시하였다. 모든 음질 특징은 마비말장애 화자 집단에서 비장애 화자 집단보다 높게 나타나 선행연구의 결과와 동일한 양상을 보였다(정필연 & 심현섭, 2017; 서인호, 2014; 서인호와 성철재, 2013).

양방향 혼합 ANOVA 분석 결과, 장애 중증도 집단 별 음질 특징 간에는 유의미한 차이가 있었다,  $F(4, 630) = 121.87, p < 0.05$ . 집단 별 차이를 확인하기 위해 Dunnett T3 사후분석을 실시한 결과, 음질 특징 모두 유의미한 효과가 있었지만 특히 voice breaks 관련 측정치들이 뚜렷한 차이를 보였다: # of voice breaks는 (경도, 경도-중등도), (중등도-중등도, 중등도)를 제외한 모든 집단에서 유의한 차이를 보였고, % of voice breaks는 모든 집단에서 유의한 차이를 보였다( $p < 0.05$ ).

표 4.4 음질 측정치 평균

집단	jitter	shimmer	HNR	# voice breaks	% voice breaks
비장애	1.87	9.33	15.99	7.3	15.95
경도	1.84	8.48	16.61	9.92	23.93
경도-중등도	2.00	9.45	16.32	10.37	29.34
중등도-중도	2.55	10.86	14.71	13.75	41.06
중도	2.34	9.32	16.30	14.47	54.55

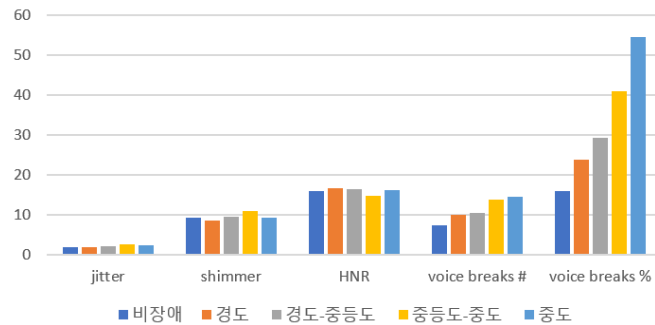


그림 4.1 음질 측정치 평균

## 4.5.2. 운율 특징

### 4.5.2.1 발화 속도

발화 속도 특징인 말속도, 조음속도, 말길이, 전체길이를 장애 중증도 집단 별로 비교하였다. 표 4.5와 그림 4.2는 화자 집단 별 발화 속도 특징의 평균값을 정리한 것이다. 마비말장애 화자는 비장애 화자보다 더 느린 말속도, 조음속도, 그리고 더 긴 말길이와 전체길이를 보여 장애 중증도가 심할수록 발화 속도가 더 느리다는 선행 연구의 보고와 일치하였다(홍새미 & 변해원, 2014; 이은경 & 김지채, 2012).

양방향 혼합 ANOVA 분석 결과, 장애 중증도 집단 별 발화 속도 특징 간에는 유의미한 차이가 있었다,  $F(4, 630) = 126.89, p < 0.05$ . Dunnett T3 사후검정 결과, 말속도, 조음속도, 말길이, 전체길이 모두 모든 집단 간 유의미한 차이가 나타났다( $p < 0.05$ ).

표 4.5 운율-발화 속도 측정치 평균

집단	말속도	조음속도	말 길이	전체 길이
비장애	3.77	4.02	3.54	4.06
경도	2.56	2.72	5.91	6.73
경도-중등도	2.11	2.24	7.54	8.56
중등도-중도	1.63	1.73	9.70	11.0
중도	1.14	1.21	13.09	16.93

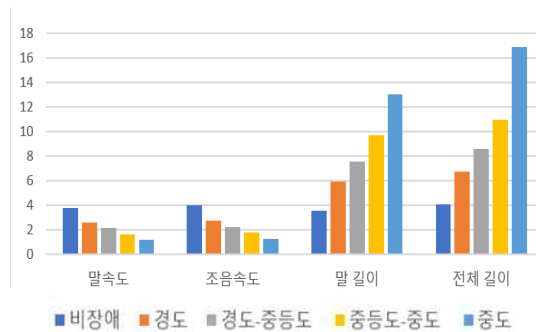


그림 4.2 발화 속도 측정치 평균

#### 4.5.2.1 음높이

음높이 특징인 F0의 평균, 표준편차, 중앙값, 최솟값, 최댓값, 25 분위수, 75 분위수를 장애 중증도 집단 별로 살펴보았다. 화자 집단 별 음높이 특징의 평균은 표 4.6와 그림 4.3과 같았다. 비장애 화자와 마비말장애 화자를 비교했을 때, 마비말장애 화자가 비장애 화자보다 더 넓은 F0 범위를 가지는 경향을 보여 선행연구의 결과를 검증하였다(Nam & Kwon, 2005).

양방향 혼합 ANOVA 분석 결과, 장애 중증도 집단 별 음높이 특징 간에는 유의미한 차이가 있었다,  $F(4, 630) = 16.37, p < 0.05$ . Dunnett T3 사후검정 결과, F0 최댓값은 (경도/경도-중등도), (경도-중등도/중등도), (중등도-중도/중도)를 제외한 모든 집단에서 유의미한 차이가 있었고, F0 75 분위수는 (비장애/경도), (경도/경도-중등도), (중등도-중도/중등도)를 제외한 모든 집단에서 유의미한 차이를 보였다( $p < 0.05$ ). 반면

F0 평균, 표준편차, 중앙값, 25 분위수는 중도 중증도 집단을 제외한 집단에서는 유의미한 차이가 나타나지 않았으며, F0 최솟값은 모든 집단에서 차이를 보이지 않았다.

표 4.6 운율-음높이 측정치 평균

집단	평균	표준 편차	중앙값	최솟값	최댓값	25 분위수	75 분위수
비장애	185.27	50.83	187.28	92.09	369.82	139.68	278.54
경도	187.28	47.28	187.34	91.35	428.75	139.35	308.04
경도-중등도	190.01	46.09	187.68	88.90	462.77	138.29	325.23
중등도-중도	201.10	57.61	191.32	87.00	529.93	139.16	360.62
중도	239.44	47.66	232.13	82.23	495.72	157.18	363.92

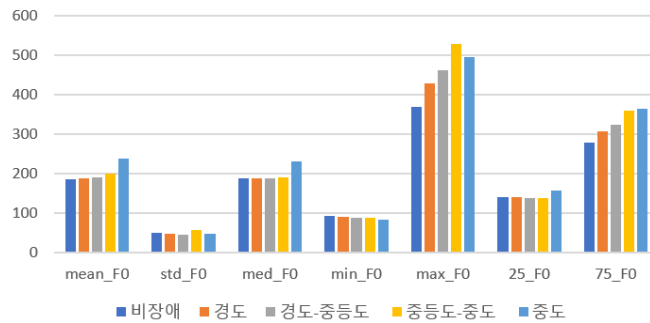


그림 4.3 음높이 측정치 평균

#### 4.5.2.1 리듬

리듬 특징인 %V, deltas, Varcos, rPVI, nPVI를 장애 중증도 별로 비교하였다. 표 4.7과 그림 4.4는 화자 집단 별 리듬 특징의 평균을 제시한다. 마비말장애 집단의 리듬 특징은 비장애 집단보다 모두 높은 경향을 보여 선행연구의 결과를 지지하였다(Hernandez, Kim, & Chung, 2020).

양방향 혼합 ANOVA 분석 결과, 장애 중증도 집단 별 리듬 특징 간에는 유의미한 차이가 있었다,  $F(4, 630) = 133.72, p < 0.05$ . Dunnett T3

사후검정 결과, deltaV, deltaC, Varco-C, rPVI-V, rPVI-C, nPVI-C에서 모든 집단 간 유의미한 차이가 나타났다( $p < 0.05$ ).

표 4.7 운율-리듬 측정치 평균

집단	%V	deltaV	deltaC
비장애	79.93	317.45	49.26
경도	83.25	556.84	93.39
경도-중등도	84.54	717.61	138.17
중등도-중도	86.55	953.10	180.45
중도	87.23	1858.60	333.76

집단	Varco-V	Varco-C	rPVI-V	rPVI-C	nPVI-V	nPVI-C
비장애	79.01	43.19	288.97	54.25	74.28	46.20
경도	82.66	61.20	526.69	97.75	80.58	60.88
경도-중등도	83.98	80.64	716.67	152.15	89.83	78.85
중등도-중도	87.47	92.53	971.39	192.46	95.16	88.55
중도	102.35	102.35	117.42	1608.09	374.22	98.72

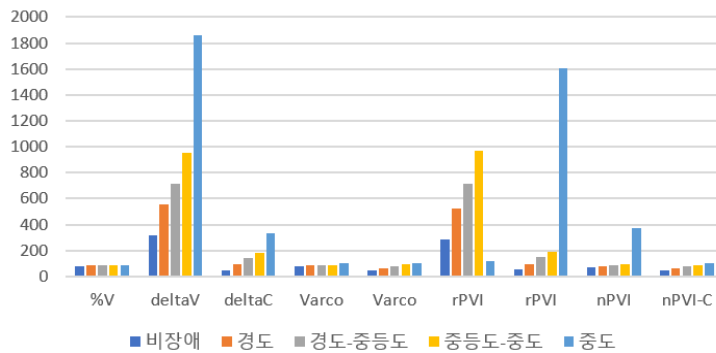


그림 4.4 리듬 측정치 평균

### 4.3.3 발음 정확도 특징

#### 4.3.3.1 음소정확도

음소정확도 특징인 자음정확도, 모음정확도, 음소정확도를 장애 중증도 별로 살펴보았다. 화자 집단 별 음소정확도 특징의 평균은 표 4.8, 그림 4.5와 같았다. 마비말장애 집단의 음소정확도 특징 모두 비장애 집단

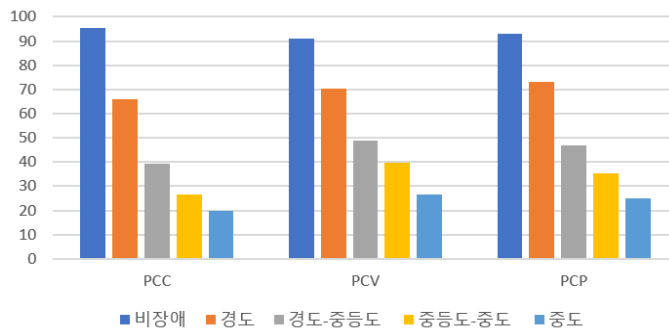


보다 낮게 나타났으며, 장애 중증도가 심해질수록 더 낮아지는 경향이 나타났다. 이는 분절음 산출 능력, 즉 음소정확도가 말명료도와 강력한 상관관계를 보인다는 선행 연구와 일치하는 결과이다(이영미 외, 2012; Whitehill & Ciocca, 2000; Clarke & Hoops, 1980). 또한, 모음 산출이 자음 산출보다 상대적으로 더 용이하기 때문에 모음정확도가 자음정확도보다 더 높다는 선행 연구와도 유사한 결과를 보였다(Clarke & Hoops, 1980).

양방향 혼합 ANOVA 분석 결과, 장애 중증도 집단 별 음소정확도 특징 간에는 유의미한 차이가 있었다,  $F(4, 630) = 8101.36, p < 0.05$ . Dunnett T3 사후검정 결과, 자음정확도, 모음정확도, 음소정확도 모두 모든 집단 간 유의미한 차이가 나타났다( $p < 0.05$ ).

표 4.8 발음정확도 - 음소정확도 측정치 평균

집단	자음정확도	모음정확도	음소정확도
비장애	95.20	90.83	93.13
경도	65.99	70.21	73.13
경도-중등도	39.37	48.86	46.78
중등도-중도	26.58	39.78	35.40
중도	19.70	26.78	25.16



#### 4.3.3.2 모음왜곡도

모음 왜곡도 특징 모음공간 면적(VSA; Vowel Space Area), FCR(Formant Centralized Ratio), VAI(Vowel Articulatory Index),

F2-ratio를 중증도 집단 별 비교하였다. 비장애 집단과 마비말장애 집단을 비교했을 때, 마비말장애 화자가 비장애 화자보다 더 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 선행연구 결과를 지지하는 결과이다(김성운, 김정환, 고도홍, 2014; 강영애 외., 2010). 모음왜곡도 특징은 화자 집단 별 모음왜곡도 특징 측정치의 평균은 표 4.9, 그림 4.6에 제시하였다.

양방향 혼합 ANOVA 분석 결과, 장애 중증도 집단 별 모음왜곡도 특징 간에는 유의미한 차이가 있었다,  $F(4, 630) = 840.31, p < 0.05$ . Dunnett T3 사후검정 결과, VSA, FCR, VAI, F2-Ratio 모두 비장애 집단과 경도 집단은 다른 집단과 유의미한 차이를 보였지만( $p < 0.05$ ), 경도-중등도, 중등도-중도, 중도 간에는 유의미한 차이가 나타나지 않았다.

표 4.9 음운 - 모음왜곡도 측정치 평균

집단	VSA	FCR	VAI	F2-Ratio
비장애	280428.48	1.02	0.98	2.61
경도	125018.71	1.27	0.79	1.94
경도-중등도	120809.81	1.27	0.79	1.93
중등도-중도	117465.69	1.27	0.79	1.93
중도	124252.46	1.27	0.79	1.93

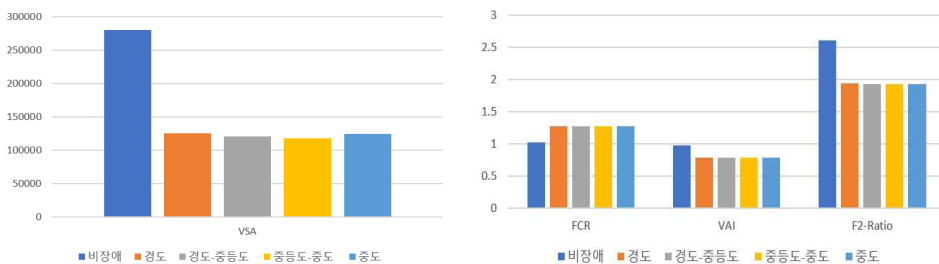


그림 4.6 모음왜곡도 측정치 평균

## 5. 실험

5장에서는 실험을 통해 마비말장애 중증도 자동 분류에 대한 발음 정확도 특징의 유용성에 대해 알아본다. MFCCs, 음질 특징, 운율 특징을 베이스라인으로 사용하고, 베이스라인에 발음 정확도 특징을 추가했을 때 분류 정확도를 확인한다. 실험은 특징 선택 알고리즘 적용 전, Recursive Feature Elimination(RFE) 적용 후, Extra Trees Classifier(ETC) 적용 후로 구분되며, 분류기는 Support Vector Machine(SVM)과 Multiple Layer Perceptron(MLP)가 사용된다. 각 분류기의 성능은 정확도(accuracy), 정밀도(precision), 재현율(recall), f1-점수로 평가되었다.

### 5.1 실험 데이터 구성

분류 실험에 사용된 훈련과 테스트 화자는 장애 중증도 집단 별 7:3으로 구성하였으며, 화자별 8개의 문장을 사용하였다. 훈련과 테스트에서 사용된 화자는 겹치지 않도록 구성하였다. 훈련과 테스트에서 사용된 화자 수와 문장 수는 표 5.1에 제시하였다.

표 5.1 훈련/테스트 화자 수 및 문장 수

	훈련	테스트
비장애	7명/56개	3명/24개
경도	16명/128개	8명/48개
경도-중등도	16명/128개	8명/48개
중등도-중도	10명/80개	5명/40개
중도	5명/40개	2명/16개

### 5.2 분류기 성능 척도

분류기의 성능은 정확도(accuracy), 정밀도(precision), 재현율(recall), f1-점수로 평가되었다. 정확도는 분류기의 전반적인 성능을 평가하는 반면, 정밀도, 재현율, f1-점수는 집단 별 정확도를 반영한다. 정확도란 전

체 샘플 중 맞게 예측한 샘플 수의 비율을 의미한다. 정밀도는 해당 집단에 속한다고 출력한 샘플 중 실제 그 집단에 속하는 수의 비율을, 재현율은 실제 해당 집단에 속하는 샘플 중 그 집단에 속한다고 출력한 수의 비율을 의미한다. 마지막으로 f1-점수는 정밀도와 재현율의 조화평균으로, 정밀도와 재현율을 모두 반영한 점수이다.

### 5.3 실험 결과

#### 5.3.1 특징 선택 알고리즘 적용 전

먼저, 모든 특징을 사용하여 실험을 진행하였다. 발음 정확도 특징을 추가했을 때 SVM 분류기는 분류 정확도 60%에서 69.5%로 상대적 증가 비율 15.83%, MLP 분류기는 분류 정확도 57%에서 69%로 상대적 증가 비율 28.97%를 보였다. 즉, 발음 정확도 특징을 추가했을 때 SVM, MLP 분류기 모두 더 좋은 성능을 보였다. 표 5.2는 특징 선택 알고리즘 적용 전의 실험 결과를 제시한다.

표 5.2 특징 선택 알고리즘 적용 전 실험 결과

특징 셋	분류기	정확도	정밀도	재현율	f1-점수
MFCCs +음질+운율	SVM	60	62	60	60.09
	MLP	57	47.88	57	59.96
MFCCs +음질+운율 +발음정확도	SVM	69.5	72.32	69.5	69.41
	MLP	69	69.31	69	69.13

표 5.3 발음 정확도 특징 추가에 따른 정확도 상대적 증가 비율

분류기	상대적 증가 비율(%)
SVM	15.83
MLP	28.97

#### 5.3.2 RFE(Recursive Feature Elimination) 적용 후

추출된 특징 중 분류에 필요한 특징만을 사용하기 위해 RFE 특징 선택 알고리즘을 적용한 후 분류 실험을 진행하였다. 실험 결과는 표 5.4와 그림 5.1에 제시하였다. 전체 특징 #는 특징 선택 전의 특징 개수를,

최적 특징 #는 특정 선택 후 남은 특징 개수를 의미한다. 분류 실험 결과, SVM은 정확도 59%에서 74%로 상대적 증가 비율 25.42%, MLP는 정확도 59%에서 71.5%로 상대적 증가 비율 21.19%를 보여 SVM, MLP 분류기 모두 발음 정확도 특징을 추가했을 때 더 높은 성능을 보였다.

표 5.4 RFE 적용 후 실험 결과

특징 셋	전체 특징 #	최적 특징 #	분류기	정확도	정밀도	재현율	f1-점수
MFCC+ 음길+ 운율	13 +25	13 +13	SVM	59	60.63	59	58.97
			MLP	59	55.96	53.5	53.6
MFCCs+ 음길+ 운율+ 발음정확도	13 +32	13 +16	SVM	74	76.09	74	74.4
			MLP	71.5	72.15	71.5	71.74

표 5.5 RFE 실험의 상대적 증가 비율

분류기	상대적 증가 비율(%)
SVM	25.42
MLP	21.19

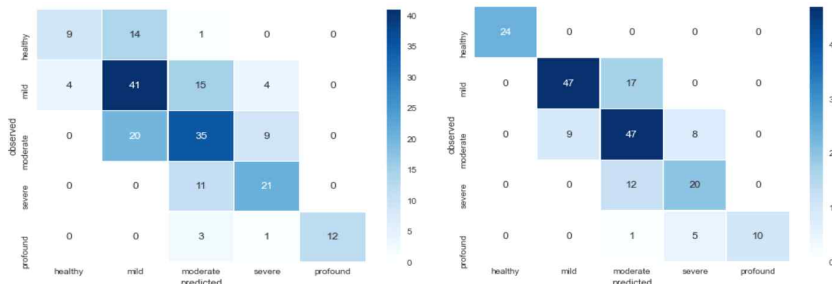


그림 5.1 SVM의 발음 정확도 특징 추가 전후 혼동 행렬(RFE)

더 나아가, 실험 별로 RFE 알고리즘에 의해 선택된 특징 셋을 비교해보았다. 베이스라인 특징 셋에서는 음질 특징 4개(jitter, HNR, # of voice breaks, % of voice breaks), 운율-발화 속도 특징 4개(말속도, 조음속도, 말길이, 전체길이), 운율-음높이 특징 1개(F0 평균), 운율-리듬 특징 4개(Varco-V, Varco-C, rPVI-V, nPVI-C)를 포함한 운율 특징 8개가 선택되었다. 이는 마비말장애 장애 중증도 자동 평가에 발화 속도,

음높이, 리듬 특징이 모두 도움이 된다는 Hernandez et al.(2020)의 결과와 일맥상통하다. 그러나 발음 정확도 특징을 추가한 후 선택된 특징 셋을 살펴보았을 때, 기존에 선택되었던 음질 특징(HNR), 운율-발화 속도 특징(말속도), 운율-리듬 특징(Varco-V, rPVI-V, nPVI-C)이 제외된 것이 확인되었다. 반면 운율-음높이 특징 F0 중앙값이 추가되었으며, 발음 정확도 특징 7개는 모두 선택되었다. 실험 별 선택된 특징 셋은 표 5.6과 같으며, 두 실험에서 공통적으로 선택된 특징은 볼드체로 표시하였다.

표 5.6 실험 별 선택된 최적 특징 비교(RFE)

특징 셋		베이스라인		발음 정확도 특징 추가	
		특징 #	특징	특징 #	특징
음질		4	jitter, HNR, #ofvoice breaks, %ofvoice breaks	3	jitter, #of voice breaks, %of voice breaks
운율	발화 속도	4	조음 속도, 말속도, 말길이, 전체길이	3	조음속도, 말길이, 전체길이
	음높이	1	<b>F0 평균</b>	2	<b>F0 평균, F0 중앙값</b>
	리듬	4	Varco-V, <b>Varco-C</b> , rPVI-V, nPVI-C	1	<b>Varco-C</b>
발음 정확도	음운 정확도	/		3	PCC, PCV, PCP
	모음 왜곡도	/		4	VSA, VAI, FCR, F2-Ratio

### 5.3.3 ETC(Extra Trees Classifier) 적용 후

또 다른 특징 선택 알고리즘인 ETC 특징 선택 알고리즘을 적용한 후 분류 실험을 진행하였다. 실험 결과는 표 5.7과 그림 5.2에 제시하였다. 분류 실험 결과, SVM은 정확도 59%에서 77.5%로 상대적 증가 비율 31.36%, MLP는 정확도 61%에서 77%로 상대적 증가 비율 22.95%를 보여 두 분류기 모두 발음 정확도 특징을 추가했을 때 더 높은 성능을 보였다.

표 5.7 ETC 적용 후 실험 결과

특징 셋	전체 특징 #	최적 특징 #	분류기	정확도	정밀도	재현율	f1-점수
MFCCs+ 음질+ 운율	13+ 25	13+ 11	SVM	59	61.51	59	58.44
			MLP	61	60.96	61	60.7
MFCCs+ 음질+ 운율+ 발음정확도	13+ 32	13+ 9	SVM	77.5	77.82	77.5	77.38
			MLP	77	77.54	77	76.76

표 5.8 ETC 실험의 상대적 증가 비율

분류기	상대적 증가 비율(%)
SVM	31.36
MLP	26.23

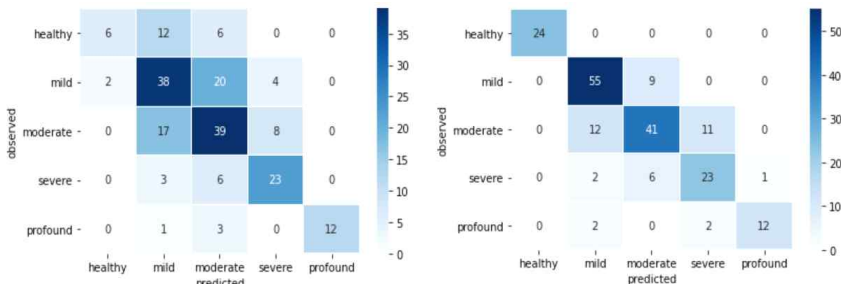


그림 5.2 SVM의 발음 정확도 특징 추가 전후 혼동 행렬(ETC)

더 나아가, ETC 알고리즘이 선택한 특징 셋을 실험 별로 비교하였다. 베이스라인 특징 셋에서 선택된 특징은 음질 특징 2개(% of voice breaks, shimmer), 운율-발화 속도 특징 4개(말속도, 조음속도, 말길이, 전체길이), 운율-음높이 특징 2개(F0 평균, F0 중앙값), 운율-리듬 특징 3개(deltaV, Varco-C, nPVI-C)를 포함한 운율 특징 9개였다. 그러나 발음 정확도 특징을 추가한 후 선택된 특징 셋을 살펴본 결과, 기존에 선택되었던 음질 특징(shimmer)과 운율-발화 속도(말속도, 조음속도, 전체 길이), 운율-음높이(mean\_F0, med\_F0), 운율-리듬 특징(deltaV, Varco-C, nPVI-C)은 제외된 반면, 발음 정확도 특징 7개 모두가 선택되었다. 각 실험에서 선택된 특징 셋은 표 5.9와 같으며, 두 실험에서 공통적으로 선택된 특징은 볼드체로 표시하였다.

표 5.9 실험 별 선택된 최적 특징 비교(ETC)

특징 셋		베이스라인		발음 정확도 특징 추가	
		특징#	특징	특징#	특징
음질		2	%ofvoice breaks, shimmer	1	%ofvoice breaks
운율	발화 속도	4	말속도, 조음속도, 말길이, 전체길이	1	말길이
	음높이	2	F0 평균, F0 중앙값	0	-
	리듬	3	deltaV, Varco-C, nPVI-C	0	-
발음 정확도	음운 정확도			3	PCP, PCV, PCC
	모음 왜곡도			4	VSA, FCR, VAI, F2-Ratio

## 5.4 실험 결과 정리

### 5.4.1 분류 정확도 비교

특징 선택 알고리즘 적용 전, RFE 특징 선택 알고리즘 적용 후, ETC 특징 선택 알고리즘 적용 후의 실험 결과를 정리하면 표 5.10과 같다. 모든 실험에서 발음 정확도 특징을 추가했을 때 더 좋은 성능을 보였으며, 실험에 따라 15%에서 31% 사이의 상대적 증가율을 보였다. 그 중 ETC 특징 선택 알고리즘과 SVM 분류기 조합의 실험이 분류 정확도 77.5%로 가장 좋은 성능을 보였으며, 상대적 증가 비율도 31.36%로 가장 높았다.

표 5.10 발음 정확도 특징 추가에 따른 상대적 증가 비율(종합)

특징 선택	분류기	MFCC +음질+운율	발음 정확도 추가	상대적 증가 비율
-	SVM	60	69.5	15.83
	MLP	53.5	69	28.97
RFE	SVM	59	74	25.42
	MLP	59	71.5	21.19
ETC	SVM	59	<b>77.5</b>	<b>31.36</b>
	MLP	61	77	26.23



가장 좋은 성능을 보인 ETC 특징 선택 알고리즘-SVM 조합 모델에서 발음 정확도 특징이 미치는 영향을 세분화하여 살펴보기 위해 발음 정확도의 특징을 음소정확도 특징과 모음왜곡도 특징으로 나누어 분석하였다. 이때 베이스라인인 음질, 운율은 각각 ETC 특징 셋에 포함된 % of voice breaks와 말길이를 의미한다. 실험 결과, 베이스라인 실험의 분류 정확도는 59%, 음소정확도 특징과 모음왜곡도 특징을 각각 추가했을 때 74.5%, 69.5% 분류 정확도가 나타났다. 즉, 음소정확도 특징과 모음왜곡도 특징을 베이스라인에 각각 추가하더라도 베이스라인보다 향상된 성능을 보였으며, 음소정확도 특징을 추가했을 때 모음왜곡도를 추가했을 때보다 더 좋은 성능을 보였다. 그러나 두 실험 모두 음소정확도 특징과 모음왜곡도 특징을 함께 추가했을 때(77.5%)보다는 낮은 분류 정확도를 보였다.

표 5.11 발음 정확도 특징 추가 실험 세부 결과

특징 셋	정확도	정밀도	재현율	f1-점수
MFCCs+음질+운율	59	61.51	59	58.44
MFCCs+음질+운율 +발음정확도 (음소정확도)	74.5	76.27	74.5	74.53
MFCCs+음질+운율 +발음정확도 (모음왜곡도)	69.5	70.16	69.5	69.66
MFCCs+음질+운율 +발음정확도	<b>77.5</b>	77.82	77.5	77.38

더 나아가, ETC 특징 선택 알고리즘-SVM 분류기의 분류 결과를 장애 중증도 별로 분석해 보았다. 장애 중증도 별 정확도는 표 5.12에 제시하였다. 분석 결과, 발음 정확도 특징을 추가했을 때 모든 집단 측면에서 분류 정확도가 상승했다. 구체적으로, 음소정확도를 추가했을 때 경도-중등도, 중등도-중도 집단의 분류 정확도는 73.34%, 75%로 가장 높은 성능을 보였고, 모음왜곡도를 추가했을 때에는 비장애 집단 분류 정확도 100%로 가장 높은 성능을 보였다. 이는 음소정확도는 경도-중등도, 중등도-중도 화자를 분류하는 것에, 모음왜곡도는 비장애화자와 장애화자를

구분하는 것에 도움이 되었음을 의미한다. 반면 중도 집단에 대한 분류 정확도는 모든 실험에서 75%로 나타나 발음 정확도 특징이 중도 집단을 분류할 때에는 음질, 운율 특징에 추가적인 도움을 주지 못하였다.

표 5.12 장애 중증도 별 분류 정확도 비교

장애 중증도	MFCC +음질 +운율	MFCC +음질+운율+ 발음정확도 (음소정확도)	MFCC +음질+운율+ 발음정확도 (모음왜곡도)	MFCC +음질+운율 +발음정확도
비장애	25	58.33	100	100
경도	59.38	81.25	67.19	85.94
경도-중등도	60.94	73.34	57.81	64.06
중등도-중도	71.88	75	71.88	71.88
중도	75	75	75	75

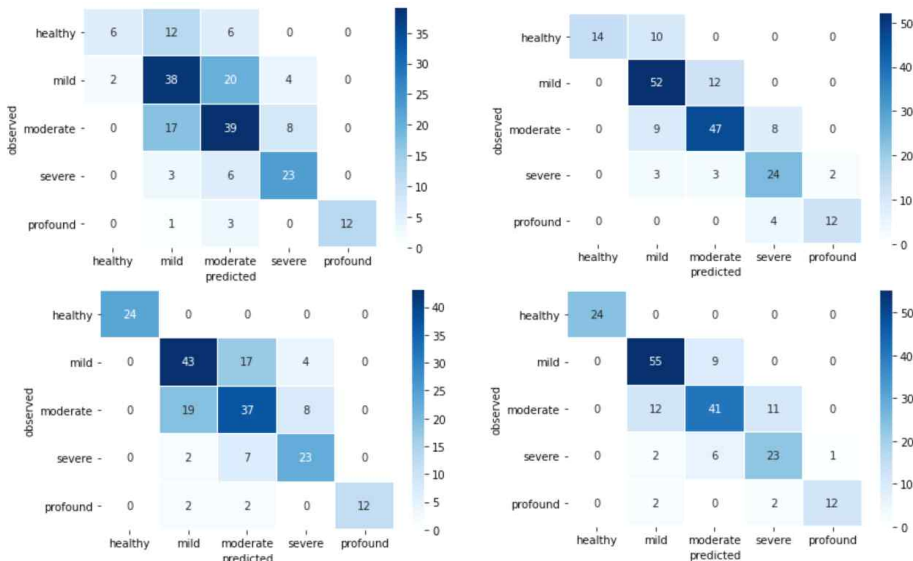


그림 5.3 SVM의 발음 정확도 특징 추가 전후 혼동 행렬  
(왼쪽 위부터 시계방향으로 음질+운율, 음질+운율+음소정확도,  
음질+운율+모음왜곡도, 음질+운율+발음정확도)

#### 5.4.2 특징 선택

마비말장애 중증도 자동 분류에 중요한 특징을 알아보기 위해 RFE와 ETC 두 특징 선택 알고리즘이 선택한 특징을 살펴보았다. 표 5.13은

각 특징 선택 알고리즘에 의해 추려진 특징들을 제시한 표이며, 두 특징 셋에서 공통적으로 선택된 특징들은 볼드체로 표시하였다. 두 특징 선택 알고리즘이 모두 선택한 특징은 다음과 같았다: 음질 특징 1개(% of voice breaks), 운율 특징 1개(발화 속도-말길이), 발음 정확도 특징 7개(PCP, PCV, PCC, VSA, FCR, VAI, F2-Ratio).

더 나아가 RFE 특징 셋과 ETC 특징 셋을 비교한 결과, RFE 특징 셋이 ETC 특징 셋을 포함하였다. ETC 특징 셋이 RFE 특징 셋보다 더 좋은 성능을 보인 것을 고려했을 때, 공통적으로 선택된 특징들은 마비말장애 음성의 장애 중증도 자동 분류에 도움이 되는 반면, RFE의 나머지 특징들(jitter, voice break 개수, 조음속도, 전체길이, F0 평균, F0 중앙값, Varco-C)은 분류 과제에 다소 방해가 되었음을 예상할 수 있다.

표 5.13 최종 특징 선택 결과 비교

Feature set		RFE		ETC	
		최적 특징 #	features	최적 특징 #	features
음질		3	jitter, voice breaks 개수, <b>voice breaks 비율</b>	1	<b>voice breaks비율</b>
운율	발화 속도	3	조음속도, <b>말길이</b> , 전체길이	1	<b>말길이</b>
	음높이	2	F0 평균, F0 중앙값	0	-
	리듬	1	Varco-C	0	-
음운	음운 정확도	3	<b>자음정확도</b> , <b>모음정확도</b> , <b>음소정확도</b>	3	<b>자음정확도</b> , <b>모음정확도</b> , <b>음소정확도</b>
	모음 왜곡도	4	<b>VSA, VAI, FCR, F2-Ratio</b>	4	<b>VSA, VAI, FCR, F2-Ratio</b>

따라서 본 논문에서 제시하는 마비말장애 중증도 자동 분류의 최적 셋은 다음과 같다: % of voice breaks, 말길이, 자음정확도, 모음정확도, 음소정확도, VSA, VAI, FCR, F2-Ratio. 우선, % of voice breaks, 말길

이, 자음정확도, 모음정확도는 4.5장 데이터베이스 통계 분석에서 모든 장애 중증도 집단 간 유의미한 차이가 나타난 특징들로 최적의 특징 셋에 포함된 근거가 충분하다. VSA, VAI, FCR, F2-Ratio는 장애 중증도 집단 간 유의미한 차이는 없었지만 비장애 집단과 경도 집단 간에는 유의미한 차이가 있었다. 이는 5.4.1장에서 모음왜곡도 특징이 비장애 집단과 경도 집단을 구분하는 것에 도움이 된 이유를 뒷받침한다.

마지막으로 최적의 특징 셋 특징들에 대한 ETC 알고리즘의 특징 별 기여가중치를 살펴보았다. 특징 별 기여가중치 평균 순서는 표 5.14과 같았다. 발음 정확도 특징에 모두 음질 및 운율 특징보다 더 높은 기여가중치가 부여된 것을 확인할 수 있다.

표 5.14 특징 별 기여가중치 평균 순서

순위	특징	특징구분
1	음소정확도	발음정확도
2	FCR	발음정확도
3	VAI	발음정확도
4	F2-Ratio	발음정확도
5	자음정확도	발음정확도
6	모음정확도	발음정확도
7	VSA	발음정확도
8	%of voice breaks	음질
9	말길이	운율-발화 속도

## 5.5. 토의

한국어 마비말장애 중증도 자동 분류를 진행한 연구는 Hernandez, Kim, & Chung(2020)이 전부이다. 해당 연구는 본 논문과 동일한 QoLT\_2012의 연속음성을 사용했으나, 비장애(healthy), 경도(mild), 중도(moderate), 심도(severe)로 나누어 4분류 과제로 시행했다는 것에 차이가 있다. 이 연구의 최고 성능을 보인 모델은 MFCCs, 음질, 운율(발화 속도, 음높이, 리듬)을 사용한 모델로, 70.10%의 정확도를 보였다. 반면 본 논문은 비장애, 경도, 경도-중증도, 중증도-중도, 중도로 나누어 5분

류 과제를 시행하였다. 이처럼 본 논문에서는 선행연구보다 더 어려운 과제를 수행했음에도 발음 정확도 특징을 추가함으로써 77.5%라는 더 높은 분류 정확도를 달성하였다. 더 나아가, 실험 별 특징 선택 알고리즘의 결과를 살펴보았을 때, 흥미로운 결과를 확인할 수 있었다. 베이스라인 실험에서는 Hernandez, Kim, & Chung(2020)에서 주장한 것과 같이 운율 특징에서 발화 속도, 음높이, 리듬 특징이 모두 주요한 특징으로 선택되었지만, 발음 정확도 특징을 추가했을 때에는 운율 특징 중 발화속도의 말길이만이 주요한 특징으로 선택되었다. 이는 마비말장애 중증도를 자동 분류할 때, 운율의 음높이나 리듬 특징은 발음 정확도 특징에 비해 주요한 정보를 주지 못한다는 것을 의미하는 동시에, 말길이 특징이 마비말장애의 중증도 분류에서 발음 정확도 특징과 상호보완적인 관계에 있다는 것을 시사한다. 단, 본 연구는 화자 70명의 소규모 데이터에 대한 분석으로 해석에 유의해야 한다.

더 나아가, 영어 마비말장애 중증도 자동 분류 연구와도 성능을 비교해 보았다. 비장애 음성과 장애 음성을 구분한 Narendra & Alku(2020)는 스펙트럼 특징과 성문 특징을 사용하여 91.38%의 분류 정확도를, Lansford & Liss(2014)는 모음왜곡도 특징을 사용하여 80.7%의 분류 정확도를 보였다. 본 논문에서는 이진 분류 실험을 진행하지 않았지만, 장애 중증도 별 분류 정확도로 비교 가능하다. 비교 결과, 비장애 음성 분류 정확도 100%로 본 논문에서 제안한 모델이 기존 선행 연구보다 더 우수한 성능으로 비장애 음성과 장애 음성을 구분한 것을 확인할 수 있다. Tripathi, Bhosale, & Kopparapu(2020)은 4분류 과제를 시행했으며, 스펙트럼 특징과 발음 정확도 특징 중 음소정확도를 사용하였다. 화자 종속 실험 결과 최고 성능 53.9%로, 본 연구의 최고 성능인 77.5%보다 더 낮은 성능을 보였다. 이처럼 본 논문에서는 다른 언어의 코퍼스 결과와 비교했을 때에도 더 높은 성능을 보고하였다. 이때, 해당 연구들은 본 논문에서 사용한 코퍼스와 다를 뿐만 아니라, 언어도 다르기 때문에 해석에 유의해야 할 것이다.

## 6. 결론

본 논문은 발음 정확도 특징이 마비말장애 중증도 자동 분류에 도움이 되는지, 그리고 선행연구에서 사용되어 온 음질, 운율 특징과 어떤 관계에 있는지 살펴보았다.

실험 결과, 발음 정확도 특징을 추가했을 때 베이스라인 실험보다 더 높은 분류 정확도를 보였다. MFCCs, 음질, 운율 특징 조합을 베이스라인으로 설정하고 베이스라인에 발음 정확도 특징을 추가하였을 때, 분류 정확도는 실험에 따라 15%에서 31% 사이의 상대적 증가 비율을 보였다. 최고의 성능을 보인 실험은 ETC 특징 선택 알고리즘-SVM 조합 실험이었으며, 분류 정확도 77.5%로 나타났다. ETC가 선택한 최적의 특징 셋은 다음과 같다 : % of voice breaks, 말 길이(speech duration), 자음 정확도, 모음정확도, 음소정확도(자음+모음), 모음 사각형 면적, VAI, FCR, F2-Ratio.

더 나아가, 베이스라인 실험과 발음 정확도 특징 추가 실험에서 선택된 특징을 비교하였다. 발음 정확도 특징을 추가했을 때, 베이스라인 특징 셋에서 선택되었던 음질 특징과 운율 특징 대다수는 더 이상 선택되지 않았다. 반면 발음 정확도 특징은 모두 선택되었다. 또한, 최적의 특징 셋에 대한 특징 별 기여가중치를 살펴본 결과, 모든 발음 정확도 특징에 음질, 운율 특징보다 높은 기여가중치가 부여되었다.

실험 결과는 다음 두 가지를 시사한다. 첫째, 발음 정확도 특징을 사용하는 것은 마비말장애 중증도 자동 분류에 도움이 된다. 발음 정확도 특징은 언어치료 분야에서 일반적으로 사용되어왔지만, 자동 분류 과제에서는 명시적으로 사용된 경우가 거의 없다. 본 논문은 발음 정확도 특징이 임상 현장과 마찬가지로 장애 중증도 자동 분류 과제에도 도움이 되는 것을 입증하였다.

둘째, 발음 정확도 특징은 마비말장애 중증도 자동 분류 과제에서 음질 특징, 운율 특징보다 더 큰 영향력을 보인다. 이는 말 산출과 관련된 다양한 범주의 특징 중 조음 범주 특징이 말명료도와 가장 높은 상관관

계를 보인다는 언어치료 분야의 선행 연구와 동일한 결과이다.

본 연구의 한계점은 다음과 같다. 첫째, 적은 화자 수이다. 본 연구에서 사용한 화자의 수는 70명으로, 장애 발화 연구 분야에서는 많은 편이지만, 머신러닝 기술을 사용하기에는 과적합의 우려가 있다. 향후 연구에서는 더 많은 마비말장애 화자의 음성을 사용할 필요가 있겠다. 둘째, 발음정확도 특징 추출의 낮은 정확도이다. 수동 자음정확도 결과와 자동 자음정확도 결과의 상관관계를 살펴보았을 때, 상관계수는 .58로 자동으로 추출한 결과로 수동 자음정확도를 대체하기에는 어려움이 있다. 향후 연구에서는 발음정확도 특징의 자동 추출 정확도를 향상시킬 수 있는 방법을 강구해야 할 것이다.

마비말장애 중증도 자동 분류 기술은 평가에 소요되는 시간과 노력을 절약하고, 객관적, 일관적 평가를 제공함으로써 언어재활사의 업무를 보조할 수 있다. 또한, 마비말장애 중증도 자동 분류 기술은 마비말장애 음성인식기의 성능 향상에도 도움이 될 수 있다. 마비말장애 음성은 비장애 음성과 다른 특성을 가지기 때문에 장애 음성의 음향적 특징들을 음향모델에 반영하는 것이 중요하다. 이때 마비말장애 음성은 장애 중증도 별로 음향적 특징이 구분되기 때문에, 장애 중증도 별로 음향모델을 학습하는 것이 바람직하다. 마비말장애 중증도 자동 분류 모델은 각 음성의 중증도에 따라 최적화된 음성인식기를 자동 선택하는 것에 사용될 수 있다. 향후 연구에서는 마비말장애 중증도 자동 분류 모델과 음성인식기를 결합한 모델을 통해 마비말장애 음성인식기의 성능 향상을 꾀하고자 한다.

## 참 고 문 헌

- 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. *Communication Sciences & Disorders*, 19(3), 352-360.
- 나민수. (2016). 마비말장애 화자의 음성 인터페이스 활용을 위한 어휘모델링 최적화 (Doctoral dissertation, 서울대학교 대학원).
- 박혜진. (2007). 마비말장애 환자의 말명료도 측정을 위한 등간척도법 비교 (Doctoral dissertation, 연세대학교 대학원).
- 서인호, & 성철재. (2013). 연결발화에서 마비말화자의 음질 특성. *말소리와 음성과학*, 5(4), 33-41.
- 서인호 (2014). 마비말장애의 발화조건에 따른 음질 및 발성유형의 음향음성학적 접근. 충남대학교 박사학위논문.
- 송한내, 이영미, 심현섭, & 성지은. (2013). 조음복잡성 및 조음중증도에 따른 마비말장애인의 자음정확도와 말명료도. *말소리와 음성과학*, 5(1), 39-46.
- 이영미, 성지은, & 심현섭. (2013). 뇌성마비로 인한 마비말장애 성인의 자음 오류 분석. *말소리와 음성과학*, 5(1), 47-54.
- 이영미, 성지은, 심현섭, 한지후, & 송한내. (2012). 마비말장애인의 조음 오류 유형에 따른 말명료도 분석. *언어척각장애연구*, 17(1), 130-142.
- 이영미, 성지은, & 심현섭. (2013). 뇌성마비로 인한 마비말장애 성인의 자음 오류 분석. *말소리와 음성과학*, 5(1), 47-54.
- 이옥분, 박상희, & 남현욱. (2012). 뇌성마비 화자의 말명료도 매개변수간의 상관성 연구: 예비연구. *언어치료연구*, 21(3), 115-126.
- 이은경, & 김지채. (2012). 마비말장애 화자의 말 속도에 따른 말 명료도 및 말 용인도. *언어치료연구*, 21(3), 127-144.
- 장선정, & 최현주. (2016). 마비말장애 성인의 자음정확도와 비음치가 말명료도 및 말용인도에 미치는 영향. *재활복지*, 20, 67-82.
- 정필연, & 심현섭. (2017). 경직형 뇌성마비 아동의 음질이 말명료도에



- 미치는 영향. *말소리와 음성과학*, 9(4), 129-136.
- 최여진, & 심현섭. (2013). 뇌성마비로 인한 마비말장애 성인의 최대 혀 및 입술 강도와 자음정확도 및 말명료도의 관계. *말소리와 음성과학*, 5(2), 11-22.
- 홍새미, & 변해원. (2014). 이완형 마비말장애 화자의 말속도와 씹 특성. *한국산학기술학회 논문지*, 15(5), 2930-2936.
- 홍새미, 정필연, & 심현섭. (2018). 마비말장애 발화의 청지각적 평가방법 비교: 세부평가와 범주평가. *Communication Sciences & Disorders*, 23(1), 242-253.
- 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.
- Bhat, C., Vachhani, B., & Kopparapu, S. K. (2017, March). Automatic assessment of dysarthria severity level using audio descriptors. *In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5070-5074.
- 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. *International Conference on Language Resources and Evaluation*, 3378-3381.
- 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.
- Enderby, P. (1980). Frenchay dysarthria assessment. *British Journal of Disorders of Communication*, 15(3), 165-173.
- Eyben, F., Wöllmer, M., & Schuller, B. (2010, October). Opensmile: the munich versatile and fast open-source audio feature extractor.

- In Proceedings of the 18th ACM international conference on Multimedia*, 1459–1462.
- Gerald, F.J.F., Murdoch, B.E., & Chenery, H.J. (1987). Multiple sclerosis: associated speech and language disorders. *Australian Journal of Human Communication Disorders*. 15, 15 - 35.
- 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, October). Dysarthria Detection and Severity Assessment using Rhythm-Based Metrics. *In Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Shanghai, China, 25–29.
- Kang, Y., Yoon, K., Lee, H., & Seong, C. (2010). A comparison of parameters of acoustic vowel space in patients with Parkinson's disease. *Journal of the Korean Society of Speech Sciences*. 2, 185 - 192.
- Kim, M. J., Pae, S., & Park, C. I. (2007). Assessment of phonology and articulation for children (APAC). Seoul: *Human Brain Research & Consulting*.
- Kim, Y., Kent, R. D., & Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *Journal of speech, language, and hearing research*.
- 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, M. J., Yoo, J., & Kim, H. (2013, August). Dysarthric speech recognition using dysarthria-severity-dependent and speaker-adaptive models. In *Interspeech* (pp. 3622–3626).

- Lansford, K. L., & Liss, J. M. (2014). Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. *Journal of Speech, Language, and Hearing Research*.
- Manochiopinig, S., Thubthong, N., & Kayasith, P. (2008). Dysarthric speech characteristics of Thai stroke patients. *Disability and Rehabilitation: Assistive Technology*, 3(6), 332-338.
- Müller, J., Wenning, G. K., Verny, M., McKee, A., Chaudhuri, K. R., Jellinger, K., ... & Litvan, I. (2001). Progression of dysarthria and dysphagia in postmortem-confirmed parkinsonian disorders. *Archives of neurology*, 58(2), 259-264.
- Mustafa, M. B., Salim, S. S., Mohamed, N., Al-Qatab, B., & Siong, C. E. (2014). Severity-based adaptation with limited data for ASR to aid dysarthric speakers. *PLOS ONE*, 9(1), e86285.
- Nam, H., & Kwon, D. (2005). Prosodic Characteristics in the Persons with Spastic and Athetoid Cerebral Palsy. *Journal of Speech and Hearing Disorders*, 14, 111-127.
- Narendra, N. P., & Alku, P. (2020). Automatic assessment of intelligibility in speakers with dysarthria from coded telephone speech using glottal features. *Computer Speech & Language*, 101117.
- Platt, L. J., Andrews, G., & Howie, P. M. (1980). Dysarthria of adult cerebral palsy: II. Phonemic analysis of articulation errors. *Journal of Speech, Language, and Hearing Research*, 23(1), 41-55.
- Shin, M., & Kim, Y. T. (2007). *Urimal test of articulation and phonation (U-TAP)*. Seoul: *Hakjisa*.
- Shor, J., Lang, O., Tuval, O., Brenner, M., Cattiau, J., Viera, F., McNally, M., Charbonneau, T., Nollstadt, M., Hassidim, A., Matias, Y. (2019). Personalizing ASR for dysarthric and accented speech with limited data. arXiv preprint arXiv:1907.13511.
- Theodoros, DG., Murdoch, BE., & Goozee, JV. (2001). Dysarthria

- following traumatic brain injury: incidence, recovery and perceptual features. Murdoch, BE., & Theodoros, DG. *Traumatic brain injury: associated speech, language, and swallowing disorders*. 27 - 51.
- Traynor, B. J., Codd, M. B., Corr, B., Forde, C., Frost, E., & Hardiman, O. M. (2000). Clinical features of amyotrophic lateral sclerosis according to the El Escorial and Airlie House diagnostic criteria: A population-based study. *Archives of neurology*, 57(8), 1171-1176.
- Tripathi, A., Bhosale, S., & Kopparapu, S. K. (2020, May). Improved Speaker Independent Dysarthria Intelligibility Classification Using DeepSpeech Posteriors. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6114-6118.
- Warlow, C. P., Dennis, M. S., Gijn, J. V., Hankey, G. J., Sandercock, P. A. G., Bamford, J. M., ... & Brown, M. M. (1997). Stroke: a practical guide to management. *BMJ-British Medical Journal-International Edition*, 314(7097), 1840.
- Whitehill, T. L., & Ciocca, V. (2000). Speech errors in Cantonese speaking adults with cerebral palsy. *Clinical linguistics & phonetics*, 14(2), 111-130.

## Abstract

# Automatic Severity Classification of Dysarthric Speech based on Pronunciation Accuracy

Yeo, Eun Jung

Department of Linguistics

The Graduate School

Seoul National University

Dysarthria is a motor speech disorder that occurs when muscles related to speech production are paralyzed or weakened. Dysarthria is diagnosed of its severity levels by trained speech therapists, who use perceptual evaluation on the purpose of providing appropriate treatments to each patient. While the professional diagnosis is important, perceptual evaluation not only takes a lot of time and effort but can also be biased and subjective. The automatic severity classification of dysarthric speech could compensate for these shortcomings and aid the therapists.

Pronunciation accuracy, consisted of the percentage of correct phonemes and the degree of vowel distortion, is one of the most commonly used features in a clinical setting to classify the severity levels of dysarthria. However, few previous studies have considered pronunciation accuracy as a feature for automatic severity classification. In this paper, we propose pronunciation accuracy to be beneficial in automatically classifying the severity levels for dysarthric speech.

Experiments were designed to confirm the usefulness of these features in contrast to the features used in previous studies: spectral features(MFCCs), voice quality features, and prosody features. Two feature selection methods–Recursive Feature Elimination(RFE) and Extra Trees Classifier(ETC) were used to determine the optimal feature set. Each optimal feature set was used as the input to two classifiers–Support Vector Machine(SVM) and Multiple Layer Perceptron(MLP). The classifiers were trained to determine the severity levels of each utterance into five categories – healthy, mild, mild-to-moderate, moderate-to-severe, severe. The performance of the classifier was evaluated using accuracy, precision, recall, and F1-score metrics.

Results from the experiments before and after adding pronunciation accuracy features were compared. For the SVM classifier, the classification accuracy showed a relative increase of 15.83%, 25.42%, 23.39% before feature selection, after applying the RFE algorithm, and applying the ETC algorithm, respectively. For the MLP classifier, the relative increase accuracy of 28.97%, 21.19%, 22.95% were seen. ETC algorithm–SVM classifier experiment showed the best performance with 77.5% accuracy. The optimal feature set included % of voice breaks, speech duration, Percentage of Correct Consonants, Percentage of Correct Vowels, Percentage of Correct Phonemes, Vowel Space Area(VSA), Vowel Articulatory Index(VAI), Formant Centralized Ratio(FCR), and F2-ratio. Furthermore, the selected feature sets from each experiment were compared. When the pronunciation accuracy features were included, many voice quality features and prosody features that were selected in the baseline experiment were replaced by the pronunciation accuracy features. The contribution weight of the features from the optimal feature set

showed that all pronunciation accuracy features have higher contribution weight compared to voice quality and prosody features.

The results suggest two ways. First, the pronunciation accuracy features are helpful for the automatic severity classification of dysarthria. While pronunciation accuracy features have been generally used by speech pathologists, few studies related to automatic severity classification have looked into their effect. This study proves that the pronunciation accuracy features are useful for automatic severity classification as for a clinical setting.

Second, the pronunciation accuracy features play a more important role than voice quality features or prosody features. Features related to articulation are proven to have the highest correlation with the speech intelligibility score of dysarthric speech among several features related to speech production. This study indicates that this fact holds the same for automatic severity classification.

**Keyword : dysarthria, automatic severity classification, pronunciation accuracy, machine learning, feature selection**

**Student number : 2019-24939**