

한국어 음성으로부터의 조음기관 시각화를 통한 마비말장애 환자의 심각도 측정

주윤지⁰¹, 호드리고², 심윤섭¹, 박운상*²

¹서강대학교 인공지능학과

²서강대학교 컴퓨터공학과

yungie222@sogang.ac.kr, ropimex@sogang.ac.kr, yoonseop@sogang.ac.kr,
unsangpark@sogang.ac.kr

Severity Assessment of Dysarthria of Patients with Speech Disorders through Visualization of Articulatory Mechanisms from Korean Speech

Yunji Chu⁰¹, Rodrigo Picinini Méxas², Yoonseop Shim¹, Unsang Park*²

¹Department of Artificial Intelligence, Sogang University

²Department of Computer Science and Engineering, Sogang University

요약

본 논문은 새로운 접근법을 통해 마비말장애 환자의 발화 심각도를 측정하는 방법을 제안한다. 이를 위해 한국어 음성 데이터를 활용하여 Vocal Tract Lab 라이브러리를 이용하여 조음기관을 시각화하고 분석하였다. 이 분석 결과를 바탕으로, 정상인과 환자의 발화 시 혀 위치를 L2 distance로 측정하여 발화 심각도를 정량적으로 평가하였다. 이 분석 결과를 토대로, 본 연구에서 제시하는 정량화된 데이터가 마비말장애의 진단과 치료 방향성 제시에 중요한 역할을 수행할 것으로 기대해본다.

1. 서론

마비말장애는 언어 및 음성 생산 능력에 영향을 주는 신경근계의 장애로, 말하는 데 필요한 근육의 제어와 조절을 어렵게 만들 수 있다. 이러한 장애는 일상생활에서 의사소통의 중대한 장애물이 되며, 환자들의 생활 품질과 사회적 참여에 부정적인 영향을 줄 수 있다. 따라서 마비말장애의 정확한 평가와 심각도 측정은 적절한 치료와 재활 프로그램 개발에 중요한 역할을 한다.

본 연구에서는 한국어 음성으로부터의 조음기관 시각화를 통해 마비말장애 환자의 심각도를 측정하는 새로운 접근 방식을 제안한다. 조음기관 시각화는 음성 생산에 관련된 근육 및 기관의 동작을 시각적으로 나타내는 도구로, 음성 장애에 대한 평가와 이해에 중요한 역할을 할 수 있다. 한국어로서의 특성을 고려하여, 우리는 한국어 음성 데이터를 활용하여 조음기관의 동작을 시각화하고 이를 통해 마비말장애 환자의 심각도를 측정하고자 한다.

본 논문에서는 한국어 음성을 시각화하기 위해 Vocal Tract Lab[1] (이후 VTL)을 활용하였다. 이를 L2 distance를 이용해 각 환자들의 발화 심각도를 0, 1, 2의 단계로 나누어 정량화하고자 한다.

2. 관련연구

본 연구와 가장 관련 있는 연구는 새로운 언어를 배울 때 화자의 잘못된 발음과 관련된 연구였다. 논문 [2]에서 저자는 언어치료와 외국어 발음을 돕기 위한 연구목적으로 발음할 때의 혀의 위치를 추적하여 이를 초음파 이미지를 통해 시퀀스를 분석하였다.

또한, 논문 [3]에서는, 다층 신경망을 활용하여 Magnetic Resonance Imaging (이후 MRI) 데이터로부터 학습자의 음성을 성대의 좌표로 변환한다. 이후, 이를 활용하여 CG 애니메이션을 만들고, 특정 발음이 실제로 어떻게 발음되어야 하는지와 학습자는 어떻게 발음하고 있는지를 비교 분석한다.

논문 [4]에서 저자는 비원어민의 음성 입력만을 사용하여, 비원어민의 발음에서 정확한 음소와 부정확한 음소를 정확하게 분리하고자 한다. 이를 위해, 정확하게 발음된 음성과 이와 대응되는 성대 이미지와의 관계를 feature inversion을 통해 정립한다.

이러한 선행 연구들은 음성 입력을 발음 기관 이미지로 변환하여 시각화된 정보를 통해 마비말장애 환자의 심각도를 측정하고자 하는 목적에 있어 그 가능성의 근거가 된다.

3. 실험방법

3.1 연구대상

표 1 환자 정보

환자	심각도	나이	성별
No.1	0	70대	F
No.2	1	50대	M
No.3	2	50대	M

정상인과 비정상인의 발화 음성 비교를 위하여 심각도를 기준으로 환자들을 분류하였다. 위의 표 1은 우리가 가진 음성 데이터셋 중에서 시각화를 했을 때 가장 표준적으로 심각도를 대표하는 환자를 뽑아 그 환자들의 정보를 나타낸 것이다. 연구를 진행할 때, 정상인 1명을 기준으로 하여 다음 환자들과 비교를 하는 식으로 연구를 진행하였다.

3.2 데이터셋

본 연구에서 사용한 데이터 셋은 ‘가을(김향희, 1996)’문단 중 일부이다. ‘가을’ 문단은 발화 속도, 발화 명료도 등 음성과 발화의 평가에 널리 사용될 수 있도록 고안된 문단으로, 다양한 음소들이 고르게 분포되어 있는 평가 자료이다. 이는 총 210음절로 구성되어 있으며, 낭독 시 대략 50초 안팎의 시간이 소요된다. 발화 문단의 내용은 아래 표 2에 제시해 두었다.

표 2 발화 문단 내용

‘가을(김향희, 1996)’ 문단의 일부
우리나라의 가을은 참으로 아름답다. 무엇보다도 산에 오를 땐 더욱 더 그 빼어난 아름다움이 느껴진다. 쓰다듬어진 듯한 완만함과 깎아 놓은 듯한 뾰족함이 어우러진 산등성이를 따라 오르다 보면 절로 감탄을 금할 수가 없게 된다. 붉은색, 푸른색, 노란색 등의 여러 가지 색깔이 어우러져 타는 듯한 감동을 주며 나아가 신비롭기까지 하다. 숲 속에 누워서 하늘을 바라보라. 쌍쌍이 짝지어 있는 듯한 흰 구름, 높고 파란 하늘을 쳐다보고 있노라면 과연 옛부터 가을을 천고마비의 계절이라 일컫는 이유를 알게 될 것만 같다.

4. 실험내용

4.1 VTL 유효성 검증

이 실험이 성립하기 위해서는, VTL을 활용하여 만든 이미지의 유효성이 먼저 검증되어야 한다. 이를 검증하기 위해 우리는 기본 모음들(a, e, i, o, u)이 VTL에서 이미지가 어떻게 생성되는지 확인하고, 그 이미지들 위에 모음 사각도를 덧씌웠다.

모음 사각도란 모음이 어떻게 발음되는지, 또한 발음을 할 때 혀가 상대적으로 어디에 위치하는지 이해하기 위해 만들어진 다이어그램을 의미한다[5]. 이 다이어그램은 그림 1에서 확인할 수 있다.

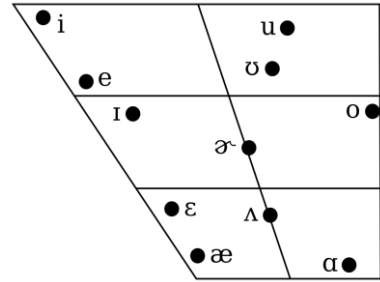


그림 1 모음 사각도 [6]

모음 사각도를 사용하면 발화자가 올바르게 발음하는지에 대한 여부를 확인할 수 있다. 혀가 다이어그램에서 위치해야 할 곳에서 벗어나 있다면, 이는 제대로 발음되지 않고 있다고 간주할 수 있다.

VTL 생성 이미지 위에 모음 사각도를 입히기 위해, 각 이미지의 어느 지점에 기본 모음들이 위치하는지 알아보았다. 4.2에서 제시한 그림 2는 VTL 생성 이미지에 모음 사각도를 덧씌운 그림이고, 그림 3은 각 기본 모음들이 이미지의 어느 위치에 나타내는지 나타낸 그림이다.

4.2 음성의 시각화

4.2.1 한국어 음성의 SAMPA로의 변환

4.1에서 제시한 바와 같이, VTL은 의미 있는 시각화 자료를 제공하는 것을 알 수 있다. 따라서, 우리는 VTL을 사용하여 정상인과 비정상인의 발화 음성을 시각화하여 비교하는 실험을 진행하였다.

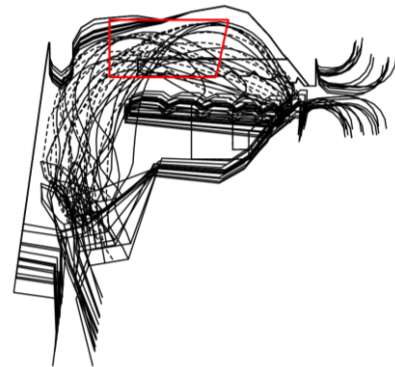


그림 2 VTL에 나타낸 모음사각도

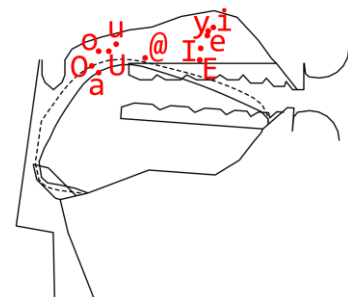


그림 3 모음의 위치를 표기한 VTL 모음사각도

VTL은 'Speech Assessment Methods Phonetic Alphabet (이후 SAMPA)'를 입력값으로 하여 svg 파일을 생성한다. 이에 따라, 발화 음성을 SAMPA로 변환하는 작업을 진행하였다.

3.2절에서 설명한 데이터셋을 확인해 보면, 한국어 음성임을 알 수 있다. 이를 SAMPA로 변환하여 VTL로 시각화하기 위해서는 그림 4와 같은 과정을 거쳐야 했다.

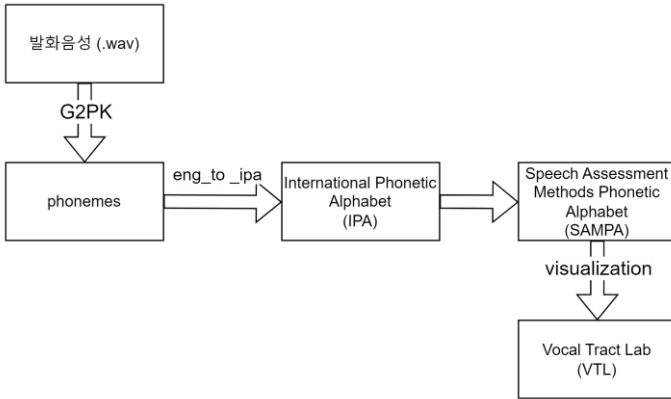


그림 4 발화음성의 시각화 순서도

입력값으로 사용된 가을 문단 발화 음성을 한국어 음소로 먼저 변환하였다. 변환에는 오픈소스 라이브러리인 g2pK 모델을 사용하였다. g2p란 graphemes to phonemes의 약자로, 자소를 음소로 변환하는 모델을 의미한다. 그중 우리가 사용한 g2pK는 한국어의 자소를 음소로 변환하는 모델이다. 이를 활용하여 손쉽게 한국어 문단을 음소로 변환할 수 있었다. 이어서, 음소를 영어로 변환해 주고, 이를 SAMPA로 변경하기 위해 IPA로 먼저 변환을 진행하였다. IPA로의 변환은 제시한 그림 4에서 확인할 수 있듯이, 오픈소스 라이브러리인 eng_to_ipa를 활용하였다. 결과값으로 얻은 IPA는 SAMPA로 변환해 주었다. 이 과정을 보기 쉽게 표로 제시하면 표 3과 같다.

표 3 한국어 문단을 SAMPA로 변환한 예시

한국어 문단	우리나라 우리나라의 가을은 참으로 아름다운 아름답다
한국어 음소	우리나라 우리나라의 가으른 차므로 아름다우 나름답따
영어로 변환	u ri na ra u ri na ra yi ga eu reu n ca meu ro a reum da u na reum dab dda
IPA	['ju'], ['ri*'], ['na'], ['ra'], ['ju'], ['ri*'], ['na'], ['ra'], ['ji'], ['ga', 'dʒi'er', 'dʒordʒə'], ['eu*'], ['reun*'],

	['ka', 'kə', 'si'er'], ['meu*'], ['roo'], ['er', 'ə'], ['riəm'], ['da', 'di'er'], ['ju'], ['na'], ['riəm'], ['dæb'], ['dda*']
SAMPA	['ju', 'ri*', 'nA', 'rA', 'ju', 'ri*', 'nA', 'rA', 'ji', 'gA"dʒi"eI"dʒOrdʒ@', 'eu*', 'reun*', 'kAk@"si"eI', 'meu*', 'roU', 'eI@', "'ri@m', 'dA"di"eI', 'ju', 'nA', "'ri@m', 'd{b', 'dda*']

4.2.2 VTL을 활용한 시각화

4.2.1에서 얻은 SAMPA 정보를 활용하여 VTL에서 시각화를 진행한 결과 그림 5와 같은 결과를 얻을 수 있었다.



그림 5 VTL을 통해 생성한 이미지에 혀를 합성한 그림

4.3 평가지표(L2 distances)

마비말장애 환자의 발화 음성의 심각도를 정량적으로 평가하기 위해, 우리는 이상적인 혀의 위치와 환자의 음성에서 얻은 혀의 위치를 비교하여 거리를 계산하고자 했다. 이를 위해 고안한 방법은 L2 distance이다.

L2 distance[7]란 두 점 사이의 거리를 계산하기 위한 방법 중 하나로 유클리드 거리라고도 알려져 있는 방식이다. 점의 집합이 식 (1)과 같이 주어졌을 때, 식 (2)와 같이 L2 distance를 계산할 수 있다.

$$x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n) \quad (1)$$

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (2)$$



그림 6 VTL에서의 정상 혀와 심각도별 혀 위치의 차이

그림 6은 왼쪽부터 심각도 0, 1, 2로 정상인과 마비말장애를 가진 환자의 음성으로부터 VTL을 통해

얻어낸 발음 기관 이미지를 각각 나타낸 것이다.

각 이미지 내의 혀 부분에 해당하는 두 곡선의 L2 distance를 계산하고, 이를 통해 특정 단어 또는 문장에 대한 시각화 결과에 대한 차이를 활용하여 실험 결과를 정량화하였다.

5. 실험결과

5.1. 실험 목적 및 방식

VTL을 통해 생성된 이미지 내의 혀를 나타내는 곡선 사이의 L2 distance와 마비말장애 환자의 심각도의 상관관계를 확인하고자 한다. 이를 위해 먼저 발화 문장으로부터 발음 기관을 시각화한 후, 각 심각도에 해당하는 마비말장애 환자의 발화 음성으로부터 발음 기관을 시각화한 것과 비교를 진행하였다. 비교를 위하여 시각화 결과 이미지를 이용하여 L2 distance를 계산하여 정상 발화 시의 혀 위치와 환자의 발화 시의 혀 위치의 차이를 측정하였다.

5.2. 실험결과 정량화

정상인과 각 심각도별 환자와의 혀 위치의 L2 distance를 pixel 단위로 구하였다. 혀를 나타내는 곡선이 37개의 점을 잇는 직선으로 이루어져 있기 때문에 이를 통해 곡선 사이의 L2 distance의 값을 구하여 표 4와 같이 제시하였다. 표 5는 L2 distance의 평균과 표준편차를 보인다.

표 4 심각도별 각 12명의 환자에 대한 L2 distance

심각도	0	1	2
No.1	5.84	14.80	22.55
No.2	6.52	17.36	26.72
No.3	13.62	15.96	19.33
No.4	9.96	12.87	18.78
No.5	7.15	16.21	18.78
No.6	13.85	15.08	23.78
No.7	6.07	17.00	16.86
No.8	9.41	14.88	14.92
No.9	8.37	18.29	23.81
No.10	6.51	12.08	27.00
No.11	20.76	21.05	23.51
No.12	22.25	19.00	25.95

표 5 심각도별 L2 distance의 평균과 표준편차

심각도	0	1	2
평균	10.86	16.22	21.83
표준편차	5.67	2.54	4.00

혀를 나타내는 곡선 간 차이의 평균은 심각도가 증가할수록 증가하는 것을 확인하였다. 그러나 표준편차의 경우 각 심각도 내에서의 L2 distance 값의 범위가 얼마나 넓은지를 나타내고 있기 때문에 심각도 1에 비하여 심각도 2, 3의 경우에는 평균 L2 distance가 해당 심각도의 특성을 대표한다고 판단하기 어렵다. 즉, 실제 심각도 0인 환자에 대하여 혀 위치에 대한 L2 distance가 심각도 1의 평균값과 더 가까울 확률이 작지 않다고 해석할 수 있다.

이를 그래프로 나타내면 6에서 제시하는 그림 7과 같으며, 심각도별 정상 혀 곡선과 환자의 혀 곡선 사이의 L2 distance의 평균이 선형 상관관계를 가짐을 확인하였다. 즉, 평균적으로 L2 distance 값을 통하여 발화 음성의 심각도를 측정하는 것이 가능하다는 결론을 얻을 수 있었다.

6. 결론 및 향후계획

본 연구를 통하여 모음 사각도를 이용한 VTL의 정확성을 확인하였으며 L2 distance를 평가 지표로 하여 마비말장애 환자의 심각도에 대한 정량적 측정 가능성에 대하여 확인하였다. 이를 통해, 한국어 음성으로부터 마비말장애 환자의 심각도를 손쉽게 파악 가능한 알고리즘을 설계하기 위한 검증이 이루어졌다. 그러나 실험 결과 심각도 1에 비하여 심각도 0, 심각도 2에 대해서는 상대적으로 L2 distance의 표준편차가 크기 때문에 신뢰도에 있어 개선이 필요할 것으로 판단된다.

따라서, 심각도 평가에 대한 신뢰도 개선을 위해 향후 표준편차에 큰 영향을 주는 데이터에 대한 분석과 또 다른 선형 상관관계를 갖는 정량적 평가 지표를 고안하는 연구를 진행하려고 한다.

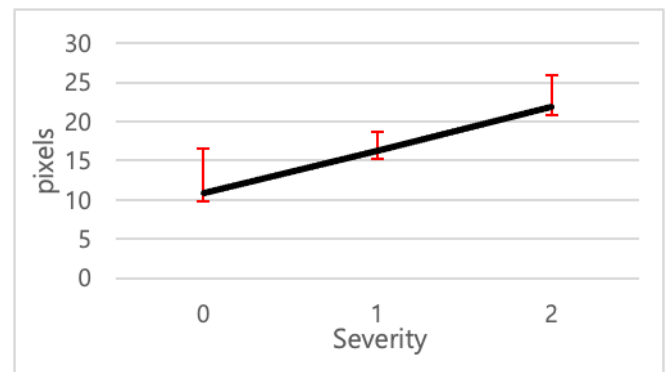


그림 7 Severity에 따른 L2 distance Pixel 그래프

7. 참고문헌

- [1] P. Birkholz, "Modeling consonant-vowel coarticulation for articulatory speech synthesis" in *PLoS one*, Vol, 8, No. 4, 2013, pp. e60603.
- [2] M. H. Mozaffari, S. Guan, S. Wen, N. Wang and W.-S. Lee, "Guided Learning of Pronunciation by Visualizing Tongue Articulation in Ultrasound Image Sequences" in *2018 IEEE CIVEMSA*, 2018, pp. 1-5.
- [3] T. Nitta, S. Manosavan, Y. Iribe, K. Katsurada, R. Hayashi and C. Zhu, "Pronunciation training by extracting articulatory movement from speech" in *Int. Symposium on Automatic Detection of Errors in Pronunciation Training*, 2012, p. 75.
- [4] O. Engwall, "Pronunciation analysis by acoustic-to-articulatory feature inversion." in *Int. Symposium on Automatic Detection of Errors in Pronunciation Training*, 2012, p. 79.
- [5] F. Trujillo, The vowel chart, <https://www.ugr.es/~ftsaez/fonetica/vowels.pdf>, Retrieved May 30, 2023
- [6] P. Ladefoged, "American English", *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 1999, pp. 41-44.
- [7] D. Cohen, "Precalculus: A Problems-Oriented Approach", *Cengage Learning 6th ed*, 2004, p. 698.