creative commons

COMMONS DEED

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

dCollection

문학박사 학위논문

한국어 자동 음성 인식의 오류 유형에 대한 음운론적 연구

2022년 2월

서울대학교 대학원

국어국문학과 국어학 전공

김 아 름

한국어 자동 음성 인식의 오류 유형에 대한 음운론적 연구

지도교수 김 성 규

이 논문을 문학박사 학위논문으로 제출함.
2021년 10월

서울대학교 대학원
국어국문학과 국어학 전공
김 아 름

김아름의 문학박사 학위논문을 인준함.
2022년 1월

위 원 장 박진호 (인)

부위원장 김현 (인)

위 원 이진호 (인)

위 원 이찬규 (인)

위 원 김성규 (인)

국문초록

본 연구의 목적은 한국어 자동 음성 인식의 인식 오류에 영향을 주는 음운론적 요인들을 밝히고, 음성 인식 오류가 주로 발생하는 음운론적 환경과 오인식되는 음소의 특성 분석을 통해, 음성 인식 오류의 유발 요인에 따라 오류를 유형화하는 것이다. 본고에서는 이를 위해 발음 가능한 한국어의 모든 단음절을 대상으로 하여 음성 인식 실험을 진행하였으며, 인식 오류가 발생하는 환경에서 한 음소를 특정 음소로 오인식하는 경향이 나타날 때, 이를 해당 음소의 '유사 음소'로 간주하고, 초성·중성·종성 순으로 각 음소별 유사 음소를 제시하였다.

이 중 초성의 인식 오류 양상은 3장에서 기술된다. 초성의 경우 주파수 대역이 가장 극단에 위치한 마찰음과 공명음의 인식률이 높게 나타났으며, 조음 과정 중 묵음에 가까운 폐쇄 단계를 가지며 추가적인 인식 단서가 적은 평파열음의 인식률이 가장 낮은 것으로 확인되었다. 초성 자음들은 음성적 변이가 발생하는 후행 모음 환경에 놓이거나, 자음들 간의 음향 특성이 혼동되는 환경에서 오류가 발생하는 경향을 보인다. 초성 파열음 중 평음은 주로 격음과 혼동되는데, 이러한 오류는 성별에 따른 F0 특성과의 관계 속에서 이해될 수 있다. 남성의 낮은 F0값은 격음에서 평음으로의 오인식을, 여성의 높은 F0값은 평음에서 격음으로의 오인식을 유발하는 것으로 추정된다. 경음과 격음은 경음 간, 격음 간의 오인식이 주로 나타난다. 이는 해당 음소가 경음과 격음이라는 점은 비교적 정확히 인식된다는 의미이며, 경음과 격음은 '긴장'이나 '기식'과 같은 유표적인 특성을 가지고 있어 평음보다 인식이 용이한 듯 보인다. 마찰음과 파찰음도 경음·격음과 마찬가지로 마찰음이라는 범주와 파찰음이라는 범주는 비교적 명확히 지각된다. 단 이들은 고모음 환경에서 마찰 구간의 변별이 약해지거나, 마찰과 기식이 적절히 변별되지 않을 경우 서로 간의 오류가 발생한다.

중성의 인식 오류 양상은 4장에서 기술된다. 단모음 중 인식률이 가장 높은 모음은 'ㅣ', 가장 낮은 모음은 'ㅜ'이며 대체로 후설모음의 인식률이 낮게 나타난다. 단모음은 선행 자음의 조음 위치의 영향으로 인식 오류가 유발되거나, 성별에 따른 모음의 음성적·조음 공간적 차이에 의해 오류가 발생한다. 조음 위치와 관련된 인식

오류는 주로 모음 'ㅡ'에서 관찰된다. 'ㅗ'와 'ㅜ'는 현대 한국어에서 세대·성별에 따른 조음 공간의 차이가 큰 모음으로 간주되는데, 이 중 'ㅗ'의 'ㅜ'로의 오인식은 여성에게서 유의미하게 나타난다. 이는 성별에 따른 원순모음의 조음 공간 차이에 의한 오류로 판단된다. 'ㅓ'의 'ㅏ'로의 오인식도 여성에게서 유의미하게 관찰되는 오류이며, 이는 남성과 여성의 F1 특성 차이에서 기인한 것으로 판단된다. 이중모음의 인식에 있어서 j계 이중모음은 자연 발화에서의 발음 특성이 인식에 영향을 미치며, w계 이중모음은 초성의 존재 여부와 함께, 사용 빈도에 근거한 이중모음의 출현 확률이 인식 오류에 영향을 주었다.

종성의 인식 오류 양상은 5장에서 기술된다. 종성의 경우 유음, 비음, 파열음 순으로 인식률이 낮아지며, 파열음 상호 간, 비음 상호 간의 인식 오류가 주로 발생한다. 종성의 경우 음소의 존재 자체를 인식하지 못하는 오류도 다수 관찰되는데, 이는 포먼트 전이 양상이나 선행 모음과 종성 간의 경계가 명확히 인식되지 못한 데에서 기인한 오류이다.

이상의 논의를 바탕으로 6장에서는 음성 인식 오류의 음운론적 유형에 대해 기술한다. 음성 인식 오류는 음성 인식 알고리즘이 정해둔 각 음소들의 전형적인 음향 특징 범위를 벗어나거나, 그 경계에 있는 음이 감지되었을 때 발생하는 것으로 보인다. 그리고 이러한 경우 음소를 정의하는 여러 음향 특성 중 어떠한 특성이 임계점을 벗어났는지에 따라 오인식되는 음소가 결정된다. 이때 발화음의 특성이 해당 음소의 임계점을 벗어나게 만드는 데에는 특정 음운론적 환경이 관여할 수도 있고, 특정 조건 없이 음향적 특성이 유사한 음소들 간의 오류가 인식을 유발할 수도 있다. 본고에서는 이 중 전자와 같은 요인에 의해 발생하는 음성 인식 오류를 '음성적 변이에 의한 인식 오류' 유형으로, 후자를 '음향적 요인에 의한 인식 오류' 유형으로 분류하였다. 이외에 남성과 여성의 발화가 가진 음성적·음운론적 특성도 음성 인식 오류에 영향을 주는 것으로 판단하였으며, 이는 '성별 요인에 의한 인식 오류' 유형으로 설정하였다.

주요어: 한국어 음성 인식, 음성 인식 오류, 유사 음소, 성별 음성 인식, 변이음 음성 인식

학 번: 2011-30001

차 례

1. 서론 ····· 1

1.1. 연구 목적 ····· 1

1.2. 연구의 필요성 ····· 3

1.3. 기존 논의 검토 ····· 5

1.4. 논의의 구성 ····· 12

2. 음성 인식 오류 확인 실험 ····· 13

2.1. 실험 절차 ····· 13

2.2. 전반적인 인식률 ····· 19

3. 초성의 음성 인식 오류 양상 및 유사 음소 ····· 27

3.1. 오류 양상 및 유사 음소 ····· 27

3.1.1. 파열음 ····· 31

3.1.1.1. 평음 ····· 31

3.1.1.2. 경음 ····· 37

3.1.1.3. 격음 ····· 42

3.1.2. 파찰음 ····· 48

3.1.3. 마찰음 ····· 51

3.1.4. 공명음 ····· 54

3.2. 종합적 검토 ····· 58

4. 중성의 음성 인식 오류 양상 및 유사 음소 ····· 62
4.1. 오류 양상 및 유사 음소 ····· 62
4.1.1. 단모음 ····· 62
4.1.2. 이중모음 ····· 83
4.2. 종합적 검토 ····· 88

5. 종성의 음성 인식 오류 양상 및 유사 음소 ····· 89
5.1. 오류 양상 및 유사 음소 ····· 89
5.1.1. 장애음 ····· 91
5.1.2. 공명음 ····· 94
5.2. 종합적 검토 ····· 97

6. 음성 인식 오류의 음운론적 유발 요인 및 유형 ····· 99
6.1. 음성적 변이에 의한 인식 오류 ····· 99
6.2. 음향적 요인에 의한 인식 오류 ····· 106
6.3. 성별 요인에 의한 인식 오류 ····· 111

7. 결론 ····· 117

참고 논저 ····· 123
Abstract ····· 132

<표 차례>

표 1 박준호(2014: 18)의 음소 유사도 점수 ····· 10
표 2 음성 제공자 정보 ····· 14
표 3 실험 대상 음소 목록 ····· 15
표 4 실험 대상 음절 목록 ····· 18
표 5 실험 대상 음절 '만'의 음성 제공자별 인식 결과 ····· 22
표 6 실험 대상 음절 '깁'의 음성 제공자별 인식 결과 ····· 23
표 7 음성 제공자들의 음절 인식률 평균 ····· 23
표 8 음절 구조 및 종성 자음별 음절 인식률 평균 ····· 25
표 9 음절 구조·자음 유형별 인식률 ····· 28
표 10 음절 구조·자음 유형별 인식률 순 ····· 28
표 11 초성의 전체 인식률(파열음) ····· 30
표 12 초성의 전체 인식률(파열음 외) ····· 31
표 13 초성 파열음 중 평음의 인식률 ····· 32
표 14 초성 'ㄱ'의 후행 모음별 인식률 ····· 33
표 15 초성 'ㄷ'의 후행 모음별 인식률 ····· 35
표 16 초성 'ㅂ'의 후행 모음별 인식률 ····· 36
표 17 초성 파열음 중 경음의 인식률 ····· 37
표 18 초성 'ㄲ'의 후행 모음별 인식률 ····· 38
표 19 초성 'ㄸ'의 후행 모음별 인식률 ····· 39
표 20 초성 'ㅃ'의 후행 모음별 인식률 ····· 41

표 21 초성 파열음 중 격음의 인식률 ····· 42
표 22 초성 'ㅋ'의 후행 모음별 인식률 ····· 43
표 23 연구개음·양순음 계열의 오인식 환경 ····· 44
표 24 성별에 따른 평음·격음 간 오인식 빈도 ····· 45
표 25 초성 'ㅌ'의 후행 모음별 인식률 ····· 46
표 26 초성 'ㅍ'의 후행 모음별 인식률 ····· 47
표 27 초성 파찰음의 인식률 ····· 48
표 28 초성 'ㅈ'의 후행 모음별 인식률 ····· 49
표 29 초성 'ㅉ'의 후행 모음별 인식률 ····· 49
표 30 초성 'ㅊ'의 후행 모음별 인식률 ····· 50
표 31 초성 마찰음의 인식률 ····· 51
표 32 초성 'ㅅ'의 후행 모음별 인식률 ····· 52
표 33 초성 'ㅆ'의 후행 모음별 인식률 ····· 53
표 34 초성 'ㅎ'의 후행 모음별 인식률 ····· 54
표 35 초성 공명음의 인식률 ····· 55
표 36 초성 'ㄴ'의 후행 모음별 인식률 ····· 56
표 37 초성 'ㅁ'의 후행 모음별 인식률 ····· 57
표 38 초성 'ㄹ'의 후행 모음별 인식률 ····· 58
표 39 후행 모음별 '평음 ⇒ 경음' 인식 오류 ····· 59
표 40 후행 모음별 '경음 ⇒ 평음' 인식 오류 ····· 60
표 41 음절 구조별 단모음 인식률 ····· 62
표 42 음절 구조별 단모음 인식률 순 ····· 63

표 43 초성 자음별 중성 인식률 ····· 64
표 44 V음절의 세부 환경별 인식률 ····· 65
표 45 VC음절의 종성별 인식률 ····· 66
표 46 CV, CVC음절 초성별 모음 ‘ㅏ’ 인식률 ····· 66
표 47 CV, CVC음절 초성별 모음 ‘ㅐ’ 인식률 ····· 66
표 48 CV, CVC음절 초성별 모음 ‘ㅓ’ 인식률 ····· 67
표 49 CV, CVC음절 초성별 모음 ‘ㅗ’ 인식률 ····· 67
표 50 CV, CVC음절 초성별 모음 ‘ㅜ’ 인식률 ····· 67
표 51 CV, CVC음절 초성별 모음 ‘ㅡ’ 인식률 ····· 67
표 52 CV, CVC음절 초성별 모음 ‘ㅣ’ 인식률 ····· 68
표 53 V음절의 중성 인식률 ····· 68
표 54 VC 음절의 중성 인식률 ····· 69
표 55 VC 음절의 종성별 ‘ㅗ’ 인식 오류 ····· 70
표 56 VC 음절의 종성별 ‘ㅜ’ 인식 오류 ····· 71
표 57 중성 ‘ㅏ’의 선행 자음 환경별 유사 음소(CV) ····· 72
표 58 중성 ‘ㅏ’의 선행 자음 환경별 유사 음소(CVC) ····· 72
표 59 중성 ‘ㅐ’의 선행 자음 환경별 유사 음소(CV) ····· 73
표 60 중성 ‘ㅐ’의 선행 자음 환경별 유사 음소(CVC) ····· 73
표 61 중성 ‘ㅓ’의 선행 자음 환경별 유사 음소(CV) ····· 74
표 62 중성 ‘ㅓ’의 선행 자음 환경별 유사 음소(CVC) ····· 75
표 63 중성 ‘ㅗ’의 선행 자음 환경별 유사 음소(CV) ····· 76
표 64 중성 ‘ㅗ’의 선행 자음 환경별 유사 음소(CVC) ····· 77

표 65 남녀 ‘ㅗ-ㅜ’ 간 오인식률 ····· 78
표 66 중성 ‘ㅜ’의 선행 자음 환경별 유사 음소(CV) ····· 79
표 67 중성 ‘ㅜ’의 선행 자음 환경별 유사 음소(CVC) ····· 79
표 68 중성 ‘ㅡ’의 선행 자음 환경별 유사 음소(CV) ····· 81
표 69 중성 ‘ㅡ’의 선행 자음 환경별 유사 음소(CVC) ····· 81
표 70 중성 ‘ㅣ’의 선행 자음 환경별 유사 음소(CV) ····· 83
표 71 중성 ‘ㅣ’의 선행 자음 환경별 유사 음소(CVC) ····· 83
표 72 음절 구조별 이중모음의 인식률 ····· 84
표 73 j계 이중모음과 w계 이중모음의 음절 구조별 인식률 평균 ····· 84
표 74 음절 구조별 이중모음 인식률 순 ····· 84
표 75 CGV 구조 w계 이중모음 ‘ㅝ’의 초성별 인식률 ····· 86
표 76 종성의 음절 구조별 인식률 ····· 89
표 77 종성의 음절 구조별 인식률 순 ····· 90
표 78 종성 파열음 인식률 ····· 91
표 79 종성 ‘ㄱ’의 선행 모음별 인식률 ····· 92
표 80 종성 ‘ㅂ’의 선행 모음별 인식률 ····· 93
표 81 종성 공명음 인식률 ····· 94
표 82 종성 ‘ㄴ’의 선행 모음별 인식률 ····· 95
표 83 종성 ‘ㅁ’의 선행 모음별 인식률 ····· 95
표 84 종성 ‘ㅇ’의 선행 모음별 인식률 ····· 96
표 85 종성 ‘ㄹ’의 선행 모음별 인식률 ····· 97
표 86 ‘ㄴ, ㄹ, ㅅ, ㅆ, ㅎ’의 후행 모음별 오류 양상 ····· 102

표 87 CVC음절 ‘ㄷ, ㅌ’의 ‘ㄱ, ㅋ’로의 인식 오류 ····· 103
표 88 양순음 뒤 ‘ㅡ ⇒ ㅜ’ 인식률 및 오류 발생 확률 ····· 104
표 89 치조 마찰음·파찰음 뒤 ‘ㅡ ⇒ ㅣ’ 인식률 및 오류 발생 확률 ····· 104
표 90 양순음-연구개음 간 음성인식 오류 양상 ····· 108
표 91 소신애(2019: 56)의 음절 말 비음 간 대치 현상의 유형 ····· 109
표 92 성별에 따른 평음·격음 간 오인식 빈도 ····· 112
표 93 성별 평음과 격음의 F0 평균(Hz) ····· 113
표 94 중간형과 최종형의 ‘격음 ⇒ 평음’ 오인식 빈도 ····· 113
표 95 성별 ‘ㅓ ⇒ ㅏ’ 오인식률과 빈도 ····· 114
표 96 모음별 남녀 F1, F2 평균(이주현 2005: 10) ····· 115
표 97 파열음의 유사 음소 ····· 119
표 98 파열음 외 자음의 유사 음소 ····· 120
표 99 단모음의 유사 음소 ····· 121

1. 서론

1.1. 연구 목적

한국어 음성언어의 발화자와 수신자가 모두 인간이었던 시기를 지나, 이제는 AI 음성대화형 어시스턴트도 발화자가 되어 한국어를 구사하고, 또 수신자가 되어 한국어를 인식하는 모습을 흔히 볼 수 있게 되었다. 음성대화형 어시스턴트가 한국어 음성을 수신하는 '음성 인식' 과정에서 발생하는 오류는, 기본적으로는 인간의 음성 지각 오류와 유사한 요인에 의해 발생할 것으로 예상된다. 인간의 음성 지각 오류가 음성적 유사성을 가지는 음소들 간의 혼란 및 발화 속도, 소음 환경, 발화 맥락 등과 관련되어 발생한다면, 음성 인식 상황에서도 이와 같은 요인들이 오류 발생에 영향을 줄 수 있을 것이다. 본 연구는 음성 인식 알고리즘의 음성 인식 과정에서 인식 오류를 유발하는 음운론적 요인이 무엇인지, 반대로 인식을 용이하게 만드는 한국어의 음운론적 특성이 있다면 그것이 무엇인지, 그리고 이러한 요인들이 인간의 음성 지각 특성과 어떠한 점이 같고 다른지 살피는 데에 연구의 목적을 두고 있다.

본고에서 연구하고자 하는 '자동 음성 인식(Automatic Speech Recognition)'이란 인간의 음성언어와 문자언어를 컴퓨터가 이해할 수 있도록 프로그래밍하는 '자연언어 처리 기술(NLP, Natural Language Processing)'의 하나로, NLP의 음성 인식, 자연언어 이해, 자연언어 생성, 음성 합성 영역 중 인간의 '음성'을 '텍스트'로 변환하는 기술 영역을 나타낸다. 일상생활에서 쉽게 접할 수 있는 모바일 기기에 탑재된 음성 인식 비서나 인공지능 스피커, 음성을 통해 조작하는 내비게이션 및 통·번역기에 이러한 자동 음성 인식 원리가 적용되어 있다.

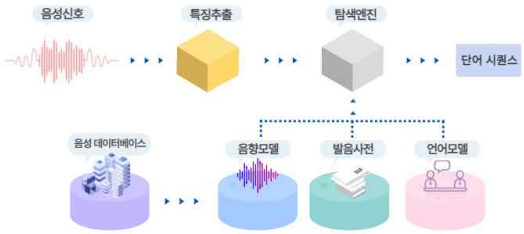
음성 인식 프로그램의 음성 인식 원리는 기기에 적용되어 있는 알고리즘마다 차이가 있으며 또 지속적으로 발전하고 있으나, 대략적인 인식 원리는 다음과 같이 요약될 수 있다. 먼저 감지된 인간의 음성 신호에서 음향 특질들을 추출해 낸 뒤 음성

인식 시스템 내 '탐색 엔진'의 첫 모듈인 '음향 모델'에서 음성 신호를 음소 기호로 변환하고, '발음 사전'에서 음소 기호를 단어 기호로 변환한다. 마지막으로 '언어 모델'에서 단어와 단어 사이의 통계적 관계 모델링을 거치며 감지된 음성을 최종 텍스트로 도출한다(방정욱 2020: 8~9).[1] 이때 음성 인식의 첫 단계라 할 수 있는 '음향 모델'에서는 음성 신호에서 추출한 음향 특질을 스펙트로그램이나 멜(MF: Mel Frequency), 멜 켑스트럼(MC: Mel Cepstrum)으로 변환하여 음소 인식의 기본 단위로 사용한다.

이와 같은 자동 음성 인식 프로그램의 여러 모듈 중 본 연구에서 살피고자 하는 음운론적 차원의 인식 오류는 주로 '음향 모델'의 처리 과정과 관련되어 있을 듯하다. 근래의 음성 인식 알고리즘은 딥러닝 학습을 바탕으로 각 음소들의 전형적인 음향 특징을 수집하고, 음성 신호가 주어지면 감지된 음성 신호가 데이터베이스 내의 어떤 음소와 가장 가까운지에 대한 확률적 판단을 통해 음소를 할당하는데, 이러한 알고리즘의 처리 과정에서 인간의 음성 지각 상황과는 다른 종류의 오류가 발생할 수 있을 것으로 예상된다. 또한 음소 층위에서 포착되지 않는 음성 층위의 변이 현상들도 음성 인식 오류에 영향을 줄 수 있는바, 음소들의 다양한 변이 현상들 중 무엇이 음성 인식에 관여하는지도 음운론적 관심의 대상이 될 수 있다.

여기에 더하여 인간의 음성 지각과 달리 AI의 음성 인식 상황에서는, 인공지능에 탑재된 음성 인식 알고리즘의 잘못된 추론 과정이 인식 오류를 유발하는 하나의 요인으로 작용할 수 있을 듯하다. 뿐만 아니라 발화자의 세대·성별과 같은 사회언어학적 정보와 그에 따른 발화상의 특성도 음성 인식 오류를 유발하는 하나의 변수가 될 수 있으며, 인공지능과의 대화 상황이라는 발화 양식이 가진 특수성도 음성 인식

1) 일반적인 음성 인식 시스템의 구조도를 제시하면 아래와 같다(방정욱 2020: 8에서 참조).



알고리즘의 판단에 영향을 주는 요인이 될 가능성이 있다.

본 연구는 이러한 관점 하에 자동 음성 인식 프로그램이 인식해야 하는 한국어 음성을 대상으로 하여, 음성 인식 환경에서 나타나는 한국어의 발화상·지각상 특성에 관해 음운론적 관점에서 논의해보고자 하는 것이다. 구체적으로는 어떠한 음운론적 환경에서 음성 인식 오류가 주로 발생하는지, 또 이때 오인식되는 음소는 어떤 특성을 가지는지 분석하는 것이 본고의 첫 번째 연구 목적이며, 이를 통해 음성 인식 오류의 발생 유형을 음운론적 관점으로 분류하는 것이 본 연구의 최종 목적이다.

1.2. 연구의 필요성

한국어 자동 음성 인식 프로그램의 인식 대상이 '한국어 음성'이기에, 이 과정에서 인식 오류가 발생한다면 일차적으로는 한국어 음소와 음성들이 가진 음운론적 요인이 오류에 영향을 주었을 것으로 예상해 볼 수 있다. 실제 음성 인식 처리 기술을 개발하는 공학 연구들에서 제시된 음성 인식 오류 개선의 당면 과제들도, 궁극적으로는 이와 같은 한국어 음소의 음운론적 특성들과 관련되어 있다. 기존의 논의에서 언급된 음성 인식 오류의 문제점과 몇몇 개선 과제들에 대해 소개하면 다음과 같다.

(1) 가. 자유발화형 음성 인식에서의 비문법적 발성(이경님 2017)
나. 음향 모델 학습 데이터와 인식 데이터의 불일치(방정욱 2020)
다. 유사한 특성을 가진 음소들 간의 변별 문제(오동훈 2019)

주지하다시피 자연언어에서는 다양한 음운적·음성적 차원의 변이가 발생하고, 소위 이러한 '비문법적 발성'과 같은 비정형 언어처리 문제는 이경님(2017: 114)에서 지적한 바와 같이 음성 인식 기술이 풀어나가야 할 문제로 언급된다. 이를 해결하기 위해 대규모 말뭉치를 통한 딥러닝 학습들이 현재 진행되고 있으나, 이경님(2017)은 이와 같은 '학습'을 통한 모델링으로는 비정형 언어처리 문제를 해결하는 데 한계가 있음을 강조하기도 하였다. 비문법적 발성 및 비정형 언어처리 문제는 자연언어의

다채로움에서 기인하는 것이라 판단되는데, 자연언어가 가진 이러한 특성은 비교적 정제된 음성 인식 학습 데이터로 훈련된 음성 인식 알고리즘이, 실제 음성 인식 상황에서 접하는 자연 발화 데이터들을 처리하는 데 어려움을 겪는 문제와도 연결된다(1나).

그리고 방정욱(2020)에서는 실제 (1나)와 같은 학습 데이터와 인식 데이터의 불일치 문제가 자유 발화 음성 인식에서 발생하는 성능 저하의 주요 요인이 됨을 지적한 바 있다. 오동훈(2019)에서는 (1다)와 같은 음성 인식 처리상의 문제를 지적하며, 음소의 서로 다른 특성을 고려하지 않은 채 음성 특징을 추출하고 심층신경망 모델을 통해 음성 인식을 수행하는 모델들의 경우, 유사한 특성을 지닌 음소들을 변별하지 못 하게 되고, 이로 인해 발생한 음성 인식 성능 저하 문제를 해결하기 위해서는 음소 인식의 정확도를 높여야 함을 강조하였다.

(1)에 제시된 음성 인식 처리 기술의 당면 과제들은 결국 음운 층위, 더 나아가 음성 층위에서 발생하는 다양한 변이 현상들의 처리 문제로 귀결되는 듯하다. 그리고 이를 개선하기 위해 공학적 관점에서 음소들의 발음 변이 정보를 분석하고, '음소 유사성' 개념을 바탕으로 인식 오류가 발생하는 음소들의 관계를 정립하고자 하는 여러 시도들이 있었으나,[2] 한국어를 대상으로 한 연구가 희박하다는 점과, 연구 과정에서 언어학적 정밀함이 부족하다는 점은 다소간의 문제라 여겨진다. 따라서 한국어 음성 인식 과정에서 발생하는 인식 오류에 관한 음운론적 관점의 논의를 통해, 음성 인식 오류를 하나의 언어 현상으로서 정밀하게 분석해볼 필요가 있을 것으로 보인다.

음성 인식 오류 문제를 음운론적 관점에서 살핀다면, 인간의 음성 지각 상황에 비추어 예측될 수 있는 지점들이 있다. 만약 음성 인식 알고리즘이 어두의 'ㅂ'를 잘못 인식하여 오류가 발생한다면, 이를 'ㅅ'나 'ㄴ'로 오인식할 가능성보다는 'ㄱ'로 오인식할 가능성이 높을 것이다. 'ㅂ'와 'ㄱ'는 음향적 유사성으로 인해 통시적으로나 공

2) 이와 관련하여 '음소 유사율'을 이용한 오류 보정을 통해 실제 인식률 향상을 유도한 연구 결과도 보고되어 있다. 한동조·최기호(2007: 84)에 따르면 음소 유사율을 통한 오류 보정을 통해 MFCC 특성 기반 음성 인식 프로그램의 인식률은 7.5%, LPC 기반의 인식률은 5.3% 향상되었다고 한다.

시적으로 곧잘 오지각되어 온 관계의 자음들이기 때문이다. 또 만약 모음 'ㅣ' 앞의 'ㅎ'를 잘못 인식한다면 이때의 'ㅎ'를 'ㅂ'나 'ㄹ'로 인식할 가능성보다는 'ㅅ'로 인식할 가능성이 높을 것이다. 'ㅎ'는 경구개 모음 'ㅣ' 앞에서 경구개 변이음으로 실현되고 그것이 다른 자음들보다는 'ㅅ'와 유사하기 때문이다. 자동 음성 인식 처리 기술 안에 이미 음소 차원의 발음 변이들은 '발음 사전' 형태로 데이터베이스화되어 있으나, 이보다 다양한 실제 음성적 차원에서의 발음 변이들과 음성 인식 오류와의 관계가 본 연구와 같은 음운론적 관점의 논의에서는 보다 선명하게 분석될 수 있을 것이다.[3]

1.3. 기존 논의 검토

자동 음성 인식의 오류 문제를 해결하기 위한 기존의 논의들은 대량의 데이터 공급으로 알고리즘의 학습 역량을 높이고자 하는 관점과, 음성 인식의 여러 하위 모듈의 처리 방식을 개선하고자 하는 관점, 오류 발생 뒤의 후처리 과정을 통해 오류를 해결하고자 하는 관점으로 나뉘어 주로 공학 분야에서 연구되어 왔다.

음성 인식이 확률 기반 통계 방식인 은닉 마르코프 모델(Hidden Markov Model, HMM)과 가우시안 믹스처 모델(Gaussian Mixture Model) 등에서 심층 신경망 DNN(Deep Neural Network)을 이용한 딥러닝 방식으로 전환되어 가고 있는 현 상황에서는, 대규모 데이터베이스를 통한 음성 인식기의 성능 향상이 더욱 강조되고 있는 듯하다.[4] 실제 방정욱(2020)에서는 딥러닝 기반의 학습이 대두되면서 음향 모

3) 음성 인식 시스템에 대한 관심은 이밖에 다양한 한국어학적 논의의 화두를 제시해줄 수도 있을 것이다. 인공지능과의 대화 상황에서 인간의 '발화' 측면에 주목해 본다면, 인공지능과 대화 시에는 자연 발화와는 다른 새로운 발화 스타일이 관찰되는데, 이러한 발화 스타일로 인해 유발되는 음성 특성이 무엇인지, 또 이때 나타나는 발화 스타일의 특성이 무엇인지에 대해서도 언어학적 관점에서 논의될 수 있다.

4) 이러한 음성 인식의 처리 방법은 크게 두 가지로 나누어 볼 수 있는데, 확률을 기반으로 하는 전통적인 음성 인식 모델과 종단 간 음성 인식 모델(End-to-End automatic speech

델의 학습에 사용되는 음성 데이터베이스의 보강이 성능에 큰 영향을 미치고 있다는 점과 함께, 학습 데이터의 양이 특정 크기를 초과하여 증가함에 따라 기계학습 접근 방식의 성능은 일정 값으로 수렴하는 것에 반하여, 딥러닝 접근 방식의 성능은 계속 향상된다는 점을 통해 대규모 데이터베이스의 중요성을 강조하였다. 부정확한 전사 자료를 이용한 학습도 음성 인식 시스템의 성능을 향상시킬 수 있음을 주장한 G. Evermann et al(2005)의 실험 결과를[5] 바탕으로, 대량의 데이터 공급을 통한 학습의 필요성을 주장한 오준석(2019: 9)도 이와 같은 관점의 논의이다.

음성 인식의 과정이 서버 기반에서 on-drive 기반으로, 전통적 모델에서 종단 간 모델로 전환되며 처리 효율이 높아진 만큼 대용량 데이터베이스의 처리도 이전보다 용이해졌기에, 이러한 흐름에서 대용량 데이터를 통한 학습은 음성 인식 성능 향상에 긍정적인 영향을 줄 수 있을 것으로 보인다. 하지만 홍윤석(2019)에서 지적한 바와 같이, 현재 상용화 단계에 있는 자동 음성 인식 시스템들이 여전히 전통적인 확률 모델 방식을 바탕으로 구현되어 있다는 점과, 음성 인식에 사용되는 데이터베이스가 강제 정렬 데이터 형식을 띠고 있어 전문가의 수작업으로 제작되어야 한다는 점에서는 한계를 가진다.[6][7] 이와 같은 문제로 인해 대용량 데이터를 통한 학습 기술의 발전과 더불어, 음성 인식 처리 모델의 정교화 내지는 정형 데이터의 구축을

---

recognition algorithms)이 그것이다. 음성 인식 처리 기술은 HMM(Hidden Markov Model), GMM(Gaussian Mixture Model)과 같은 확률 기반 모델에서 CTC(Connectionist Temporal Classification)와 LAS(Listen, Attend Spell) 등의 딥러닝에 기반한 종단 간 음성 인식 모델을 거쳐, 최근에는 셀프-어텐션 메커니즘(Self-Attention Mechanism)에 중점을 둔 트랜스포머(Transformer) 네트워크가 개발되어 여러 음성대화형 어시스턴트에 적용되고 있다.

5) 이 논의에 따르면 같은 수의 파라미터를 이용하여 360시간의 정확한 전사 자료와 400시간의 부정확한 전사 자료로 학습하였을 때 WER(Word Error Rate)의 차이가 1% 내외였다고 한다. 이와 더불어 이 논의에서는 부정확한 전사 자료의 양을 1800시간 추가한 2200 시간의 자료를 Maximum Likelihood(ML) 방식으로 학습하였을 때 WER이 1.5%, Minimum Phone Error(MPE) 방식으로 학습하였을 때 3.1% 향상되었음을 보고하고 있다.

6) 강제정렬 데이터는 음소의 시작과 끝을 음성 파일에 표시함으로써 음소 인식 모델이 각각의 음소에 해당하는 음성 파형을 학습할 수 있도록 만들어진 데이터 형식을 의미한다(홍윤석 2019: 4).

7) 이러한 문제를 해결하고자 강제 정렬 데이터를 사용하지 않는 CTC 기반 음소 인식 모델의 연구나(홍윤석 2019), 원시(raw) 데이터에서 음향 모델 학습용 데이터베이스를 자동으로 추출하는 등의 연구가 진행되고 있다.

강조하는 논의도 꾸준히 진행되고 있다.

다음으로 살펴볼 음성 인식 모듈 개선에 관한 연구는 음성 신호를 일차적으로 받아들여 음소로 매핑하는 음향 모델 개선에 관한 논의와, 이외의 모델 개선에 관한 논의로 나누어진다. 이 중 전자의 논의들은 음소 인식 자체의 정확도를 높이는 데에 집중하여, 적절한 음소 인식 '단위'에 대해 논하거나, 음소의 효율적인 분류를 통한 인식률 상승 방안 등을 연구해 왔다. 최근에는 각 음소마다의 전형적인 파형을 추출하여 이를 통해 음소 인식 학습을 제안하는 연구도 진행되고 있다.[8] 한편 음향 모델 이외의 모듈 개선을 통해 음성 인식 오류 문제를 해결하고자 했던 논의로는 음성 인식 오류를 유발하는 요인 중 하나인 '미등록어' 처리를 담당하는 언어 모델을 개선하고자 한 이창수(2014) 등을 들 수 있다.[9]

한편 음성 인식 알고리즘 자체에 대한 개선 연구와 더불어 음성 인식 단계를 거쳐 변환된 문자열에 대한 후처리 과정을 통해 오류를 수정하고자 한 연구도 다수 관찰된다. 여기에는 박준호(2014)와 같이 유사 음소 데이터베이스 구축을 시도하거나, 김상훈(2016)과 같이 오류 발생 시 재발화 알고리즘을 추가하거나, 음소 유사율을 이용하여 오류 보정을 제안한 연구 등이 해당된다. 오류 패턴 정보 후처리는 문자를 오인식하는 과정에는 일정한 패턴이 있다는 전제 하에 오인식이 발생하는 단어와 단어의 오인식 형태를 한 쌍으로 묶어 오류 패턴 사전을 구축하여 사용하는 방법으로 Satoshi Kaki(1998)에서 제안되었다.[10] 본고의 주 논의 대상인 음성 인식 오류의 유형에 관한 논의는 대개 이 같은 후처리 연구에서 다루어지고 있으며, 오류 유형을 파악함에 있어서 언어적 특성 분석을 중시하는 논의와 프로그래밍 보완을 중시하는 논의로 양분된다.

---

8) 이지현·이재원·이종우(2013), 이재원(2015), 이재원(2016), 노원빈·이종우·이재원(2016) 등이 이에 해당된다.

9) 한편 장한솔(2020)과 같이 언어 모델의 보완을 위한 음소 모델의 연구 방법도 제안된 바 있다. 이 논의에서는 문장 단위 인식을 위한 문법 오류를 교정하는 데에 음소 모델 차원의 수정이 도움이 될 수 있다는 점과, 자질 정보 임베딩을 통해 언어 모델을 음소 차원에서 보완할 수 있음을 증명하고 있는데, 이 논의는 언어학이 자연어 처리 기술의 과제를 해결하는 데 기여할 수 있음을 보여준 실증 연구라는 점에서 의의를 가진다.

10) 이 논의에서 오인식을 수정하기 위한 후처리 방법으로는 EPC(error-pattern correction)와 SSC(similar-string correction)이 제안되었다(박준호 2014: 3).

이 중 전자와 관련하여 음성 인식 오류를 해결하기 위해서는 더욱 정밀한 언어학적 토대가 필요하다는 관점의 논의로는 이수길(2005), 김민제·이정철·김종진(2006), 김진형·박소형(2012) 등이 있고, 이를 위해 본고와 같은 음성 인식 오류 확인 실험을 진행한 연구로는 이창영(2010), 양병곤(2017), 노희경·이강희(2017), 최승주·김종배(2017), 유현재·김명화·박상길·김광용(2020) 등을 들 수 있다.[11][12] 그 중 최승주·김종배(2017)은 한국어가 지원되는 음성 인식 오픈 API들의 특징을 분석하여 각 오픈 API의 장단점을 비교한 논의로, 실험에 사용한 음성대화형 어시스턴트는 Google Assistant, Naver Clova, KaKao Newtone이다. 이 논의에서 활용된 실험 방법과 대상을 정리해 보이면 다음과 같다.

(2) 가. 실험 방법

- 각 API의 샘플 APK를 설치하거나 샘플코드를 Android Studio를 이용하여 모바일 어플리케이션으로 컴파일한 후 음성을 입력하여 음성 인식 정확도를 비교함
- KaKao의 Newtone은 모바일 어플리케이션 개발 환경만 지원하기 때문에 입력 장치는 스마트폰을 사용하여 실험을 진행함
- 실험은 3단계로 진행. 잡음이 적은 공간에서 API 종류별로 각각 10회씩 발성

나. 실험 대상자

- 음성 제공자는 1명

다. 실험 대상 어휘 및 문장

- ① 0부터 19까지의 한국어 숫자 20단어
- ② 가나다 14개 음절
- ③ 모바일에서 많이 쓰이는 명령 문장을 통한 문장 인식 실험

  문장1) 최근 찍은 사진 친구한테 보내줘.

  문장2) 화요일 오후 3시 30분에 알람해 줘.

---

11) 음소 '분할'에 관한 오류 측정 실험이기는 하나, 김민제·이정철·김종진(2006)의 실험도 유사한 범주의 실험에 속한다.

12) 양병곤(2017)은 한국어 화자의 영어 발음을 대상으로 한 논의이고, 이창영(2010)에는 '한글의 50여 음소들의 고른 분포를 고려한 300단어'가 실험에 사용되었다는 점만 제시되어 있어서 구체적인 실험 어휘 목록을 확인할 수 없었다.

이 논의는 사용자의 자연스러운 음성대화형 어시스턴트 사용 환경에서 각 API별 인식률을 측정하고 있다는 점에서 의의를 가지나, 실험 대상 어휘 및 문장이 언어학적으로 통제되어 있지 않다는 점은 다소간의 문제라 여겨진다. 특히 (2다②)의 '가나다 14개 음절'에는 경음이 포함되어 있지 않아 초성에 출현 가능한 자음들의 인식 양상을 전반적으로 살펴볼 수 없다. 또한 대상 음절이 모두 개음절이라는 점과 모음이 단모음 'ㅏ'로 고정되어 있다는 점, (2다②)에 포함된 음절 '아'의 경우 모음으로만 구성되어 있으나 여타의 CV음절과 함께 분석되고 있다는 점이 실험 설계상의 문제로 지적될 수 있다. (2다③)의 '모바일에서 많이 쓰이는 명령 문장'도 실제 사용 빈도에 근거하고 있지 않으며, 음운론적으로나 형태·통사론적으로 통제되지 않은 단어들의 조합인 것으로 판단된다.[13)]

다음으로 본고에서 다루고 있는 음소 유사도와 관련된 선행 연구로는 한동조·유제만·최기호(2006), 한동조·최기호(2007), 오상엽(2013), 박준호(2014) 등을 들 수 있다. 이 중 박준호(2014)에서는 음소 유사율 기준에 따른 후처리 오류 보정 알고리즘을 제안하였는데, 발음상 유사도 및 형태상 유사도를 평가하여 이를 초성·중성·종성별 1~100점 구간의 유사도 점수로 척도화하였다. 그 구체적인 내용을 인용해 보이면 다음과 같다.

(3) 가. 초성 'ㄱ'와 'ㄲ'는 형태 및 발음 유사도 측면에서 50점 부여
나. 중성 'ㅐ'와 'ㅔ'는 형태 및 발음 유사도가 높으므로 90점 부여
다. 종성 'ㅅ'과 'ㅆ'은 형태 및 발음 유사도가 높으므로 90점 부여

13) 본고의 관찰 결과 음성 인식 오류 분석을 위한 소수의 실증 연구들 중, 실험 대상 어휘 및 문장 목록이 공개된 논의는 최승주·김종배(2017)가 유일한 것으로 보인다. 이지현·이재원·이종우(2013), 이재원(2015), 이재원(2016), 노원빈·이종우·이재원(2016) 등에서도 음성 인식 실험이 진행된 바 있으나, 이는 오류율이나 오류 환경 분석을 위한 것이 아니라 개별 음소들의 파형 특성을 추출하기 위함이었다.

초성 유사도 점수	
ㄱ/ㄲ	50
ㄲ/ㄱ	50
ㄷ/ㄸ	50
ㄸ/ㄷ	50
ㅅ/ㅆ	50
ㅆ/ㅅ	50
ㅈ/ㅉ	50
ㅉ/ㅈ	50
ㅂ/ㅃ	50
ㅃ/ㅂ	50

중성 유사도 점수	
ㅏ/ㅑ	30
ㅑ/ㅏ	30
ㅓ/ㅕ	30
ㅕ/ㅓ	30
ㅗ/ㅛ	30
ㅛ/ㅗ	30
ㅜ/ㅠ	30
ㅠ/ㅜ	30
ㅐ/ㅒ	50
ㅔ/ㅖ	50
ㅚ/ㅞ	70
ㅞ/ㅚ	70
ㅙ/ㅞ	70
ㅞ/ㅙ	70
ㅚ/ㅙ	90
ㅙ/ㅚ	90
ㅔ/ㅒ	90
ㅒ/ㅔ	90
ㅐ/ㅖ	90
ㅖ/ㅐ	90

종성 유사도 점수	
ㄱ/ㄹㄱ	70
ㄹㄱ/ㄱ	70
ㅎ/ㅅ	70
ㅅ/ㅎ	70
ㄱ/ㄲ	90
ㄲ/ㄱ	90
ㄷ/ㅅ	90
ㅅ/ㄷ	90
ㄹ/ㄹㄱ	90
ㄹㄱ/ㄹ	90
ㅅ/ㅆ	90
ㅆ/ㅅ	90
ㅈ/ㅊ	90
ㅊ/ㅈ	90
ㅍ/ㅂ	90
ㅂ/ㅍ	90
ㅅ/ㅈ	90
ㅈ/ㅅ	90
ㄷ/ㅈ	90
ㅈ/ㄷ	90
ㄱ/ㅋ	90
ㅋ/ㄱ	90

<표 1> 박준호(2014: 18)의 음소 유사도 점수

위의 (3)과 <표 1>은 개별 음소들의 유사도가 전체적으로 반영되어 있다는 점에서 나름의 의의를 가지나, 형태 및 발음 '유사도'의 기준이 명확하지 않다는 점은 문제로 지적될 수 있다. 특히 <표 1>에는 초성 자음들 중 평음과 경음만의 유사도가 마련되어 있는바 나머지 자음들의 유사도는 파악할 수 없고, 음성학적으로 평음과 보다 더 가까운 것으로 알려져 있는 격음 대신 경음을 유사 음소로 설정한 까닭도 확인되지 않는다. 또한 청·장년층에서는 합류되어 변별이 가능하지 않은 'ㅔ'와 'ㅐ'가 각각 반영되어 있다는 점, 종성에서 발음이 가능하지 않은 음소들 간의 유사도가 설정되어 있다는 점 등은 재고의 여지가 있다.

김상훈(2016)을 비롯한 대다수의 논의에서 지적하고 있는 바와 같이 음성 인식 알고리즘의 첫 단계인 음성 인식기에서 발생하는 오류는 다음 단계의 오류로 전달될

가능성이 높다는 점에서 매우 치명적이라 할 수 있다. 주지하다시피 음소는 환경에 따라 각기 다른 음성적 특성을 가진 변이음으로 실현되고, 음절 내 위치나 음절 유형, 선후행 음소의 영향 등에 의해 발음상의 변이를 보인다. 오동훈(2019)는 현재의 음성 인식 프로그램들이 이러한 음소의 특성을 고려하지 않은 상태로 음성 특징을 추출하고 이를 바탕으로 심층신경망 모델을 만들어 음성 인식을 수행함으로써, 유사한 특성을 지닌 음소들 간의 구별이 어려워졌고 이는 곧 전체적인 음성 인식 성능을 떨어뜨리는 요인이 되고 있음을 지적한다.

본고에서는 이러한 맥락에서 한국어 음소가 가진 음운론적 특성이 음성 인식 오류에 어떠한 영향을 미치는지 한국어학의 관점에서 분석될 필요가 있다고 판단한다. 이는 이수길(2005)을[14] 비롯한 2000년대 초반 논의에서부터 방정욱(2020)을 포함한 최근의 연구에서 지속적으로 요구되어온 바이기도 하다. 본 논의는 자동 음성 인식 처리 기술의 오류 유형을 한국어 음운론적 관점에서 분석하고자 하며, 이를 궁극적 목적으로 삼고 있지는 않으나 그간의 선행연구 흐름에 비추어 볼 때 오류 패턴 매칭 기반 후처리 오류 보정 분야와 관련지을 수 있을 듯하다.

14) 이 논의에서는 '음성의 변화에 대응할 수 있는 조음음성학과 음향음성학 그리고 청음음성학을 연계한 음소 분리 체계의 확립과 모음과 모음의 연결, 자음과 모음의 연결 시 자음이 모음에 주는 영향 등에서 나타나는 음성의 변화를 체계화하여 효과적으로 음성을 음소단위로 분할할 수 있는' 한국어학적 연구가 필요함을 역설한다.

1.4. 논의의 구성

본 연구는 한국어 자동 음성 인식의 인식 오류에 영향을 주는 음운론적 요인들을 밝히고, 음성 인식 오류가 주로 발생하는 음운론적 환경과 오인식되는 음소의 특성 분석을 통해, 음성 인식 오류의 유발 요인에 따라 오류를 유형화하는 것을 목적으로 한다. 이에 1장에서는 자동 음성 인식 처리 기술의 음성 인식 원리에 대해 소개하고, 현재까지 진행되어 온 한국어 음성 인식에 관한 선행 연구들의 논의 방향을 정리하며 본 연구의 필요성을 밝혔다.

2장에서는 음성대화형 어시스턴트의 인식 오류가 어떠한 음운론적 요인에 의해 발생하는지 분석하기 위한 음성 인식 실험 절차와 대상을 살핀 뒤, 실제 실험 결과의 전반적인 인식률을 제시한다. 3장·4장·5장은 2장의 실험 결과를 바탕으로 음성 인식 오류 양상을 면밀히 분석하는 장이다. 실험 결과를 초성·중성·종성으로 나눈 뒤, 초성은 다시 파열음, 파찰음, 마찰음, 공명음 순으로 분류하여 3장에서 각 자음별 인식 오류 양상을 기술하였고, 중성은 단모음과 이중모음으로 나누어 4장에서 오류 양상을 분석하였다. 종성은 장애음 계열과 공명음 계열로 나누어 인식 오류상의 특징을 살펴보았으며 이는 5장에서 기술된다.

이후 6장에서는 5장까지의 논의 결과를 바탕으로 음성 인식 오류를 음운론적 유발 요인에 따라 ‘음성적 변이에 의한 인식 오류’, ‘음향적 요인에 의한 인식 오류’, ‘성별 요인에 의한 인식 오류’로 유형화하였으며, 7장에서는 이상의 내용을 요약하고 남은 문제를 제시하는 것으로 논의를 마무리한다.

2. 음성 인식 오류 확인 실험

2.1. 실험 절차

본 절에서는 음성대화형 어시스턴트의 음성 인식 오류에 어떠한 음운론적 요인들이 관여하는지 확인하기 위한 실험 절차에 대해 소개하도록 한다. 본 실험은 실제 음성대화형 어시스턴트의 사용 상황과 유사한 환경에서 음성 인식 양상을 확인하기 위하여, 음성대화형 어시스턴트에 음성 제공자가 실험 대상 음절을 직접 발화하는 형식으로 진행하였다.[15] 실험에 사용한 어시스턴트는 '삼성' 모바일 기기에 탑재되어 있는 인공지능 가상비서 '빅스비(Bixby)'이며[16] 실험 절차는 다음과 같다. 소음과 잡음이 통제된 환경에서 실험 대상 기기와 음성 제공자의 거리를 15~25㎝ 간격으로 유지한 뒤, 빅스비를 활성화하여 음성 제공자는 실험 대상 목록을 한 음절당 다섯 번씩 발화하고 본 연구자는 그 결과 값을 기록한다.

본 실험에 참여한 음성 제공자들은 서울·경기 지역에 거주하는 20대부터 40대까지의 남녀 화자 10명이며, 이들의 정보는 아래의 <표 2>에 제시하였다 본고에서는

15) 음성대화형 어시스턴트의 음성 인식 실험은 자동 음성 인식 처리 모듈 중 '음소 인식기' 혹은 '음성 인식기' 모듈을 따로 활용하여 진행하기도 하고(한동조·최기호 2007), 본고와 같이 실제 음성대화형 어시스턴트의 사용 환경과 동일한 환경을 설정하여 실험이 진행되기도 한다[이지현·이재원·이종우(2013), 이재원(2015), 노원빈·이종우·이재원(2016), 이재원(2016), 최승주·김종배(2017), 이재원(2018) 등]. 실험 목적에 따라 대상 음성으로 ETRI(한국전자통신연구원)에서 구축한 정제된 음성 데이터베이스를 사용하기도 하며, 본고에서처럼 일반 사용자들의 발화를 통해 음성 인식 오류를 확인하는 경우도 있다. 본고의 경우 모든 단음절 단독 발화를 실험 대상으로 삼고 있어, ETRI를 비롯한 기존에 구축된 음성 데이터베이스를 사용할 수 없었기에, 실제 음성 제공자들을 대상으로 한 인식 실험을 따로 진행하였다.

16) 실험 대상 음성대화형 어시스턴트로 삼성 '빅스비'를 선택한 이유는 다음과 같다. 음성 인식 처리 기술은 개별 언어에 따라 특화되어 개발되기에, 한국에서 개발한 '네이버'의 'CLOVA'와 , '카카오'의 '헤이 카카오', '삼성'의 '빅스비'를 실험 대상으로 우선 고려하였다. 그리고 이 중 2021년 2분기 기준 국내 스마트폰 시장 점유율의 73%를 차지한 삼성의 빅스비를 최종 대상으로 선택하였다.

음성 제공자들의 출신 지역이나 지역방언적 요소는 특별히 고려하지 않았는데, 이는 음성대화형 어시스턴트와의 대화 시에는 지역방언적 특성이 크게 작용하지 않을 것이라 전제하였기 때문이다. 일상생활에서 음성대화형 어시스턴트를 사용하는 상황은 TV프로그램명이나 일부 키워드들을 검색해야 하는 경우이거나, 날씨나 시간, 경로 등을 찾아주기를 명령하는 경우가 대부분이다. 이러한 상황에서는 지역방언적 요소들 중 어휘적 차원의 방언 특성은 사용될 여지가 적다. 마치 노래를 부를 때 굳이 지역방언 어휘를 사용하지 않는 것과 같이 음성 인식 호출 시에도 어휘적 차원의 방언은 사용하지 않을 가능성이 높다는 것이다.[17][18]

	F01	F02	F03	F04	F05	M01	M02	M03	M04	M05
성별	여성	여성	여성	여성	여성	남성	남성	남성	남성	남성
연령	20대	30대	30대	30대	40대	20대	30대	30대	30대	40대
출신지역	포항	서울	경기	서울	서울	경기	경기	서울	경기	전주
거주지역	서울	경기	경기	경기	서울	경기	경기	서울	경기	경기

<표 2> 음성 제공자 정보

위의 음성 제공자들이 발화한 실험 대상 음절은 발음 가능한 한국어의 모든 단음절이며, 실험 대상 음소는 초성에 사용되는 자음 18개, 중성 단모음 7개, 이중모음 10개, 종성 자음 7개이다.[19] 실험에 사용된 음소 목록을 표로 정리하여 보이면 다음과 같다.

17) 실제 실험 결과에서도 음성 제공자마다 인식률 수치에서는 차이가 있었으나, 인식 오류의 경향성 면에서는 음성 제공자들의 개별 특성이 크게 관여하지 않은 것으로 보인다.

18) 이러한 맥락에서 본 논의에서는 지역방언의 개입 가능성을 배제한 채 실험을 설계하였으나, 어휘적 차원의 방언 특성은 나타나지 않는다 하여도 음운·음성 층위의 방언 요소는 개입될 수 있을 듯하다. 뿐만 아니라 단순한 검색이나 명령 작업 외에, 문자 입력의 번거로움을 줄이고자 발화문 전체를 음성대화형 어시스턴트를 통해 텍스트로 전사하는 경우라면 이때는 발화자의 지역방언적 특성이 드러날 가능성이 높다. 따라서 추후에는 지역방언 화자가 보이는 음운·음성적 특성이 음성 인식에 미치는 영향 관계에 대해서도 연구될 필요가 있다.

19) 이 중 종성 'ㄷ'의 경우는, 'ㄷ'를 종성으로 가지는 단음절 어형이 극히 드물어 주로 음절말 평파열음화를 겪은 음소들이 종성에서 [ㄷ]로 실현된다. 이로 인해 종성 'ㄷ'의 음성 인식 실험 결과, 종성 'ㄷ'를 'ㄷ'로 정지각한 인식률은 평균 0.9%에 불과한 것으로 나타나, 여타의 음성 인식 결과들과 크게 차이를 보였다. 종성 'ㄷ'의 인식 결과는 발화된 음성을 어떤 음소

초성	파열음	평음	ㄱ, ㄷ, ㅂ
		경음	ㄲ, ㄸ, ㅃ
		격음	ㅋ, ㅌ, ㅍ
	파찰음	평음	ㅈ
		경음	ㅉ
		격음	ㅊ
	마찰음	평음	ㅅ, ㅎ[20]
		경음	ㅆ
	공명음	비음	ㄴ, ㅁ
		유음	ㄹ

중성	단모음		ㅏ, ㅐ, ㅓ, ㅗ, ㅜ, ㅡ, ㅣ
	이중모음	j계	ㅑ, ㅒ, ㅕ, ㅛ, ㅠ, ㅢ
		w계	ㅘ, ㅝ, ㅙ, ㅟ

종성	장애음	ㄱ, (ㄷ), ㅂ
	공명음	ㄴ, ㄹ, ㅁ, ㅇ

<표 3> 실험 대상 음소 목록

주지하다시피 현대 한국어의 단모음 목록은 'ㅣ, ㅔ, ㅐ, ㅚ, ㅟ, ㅡ, ㅓ, ㅏ, ㅜ, ㅗ'와 같은 10모음으로 구성되어 있으나, 이 중 'ㅔ'와 'ㅐ'는 대다수의 청·장년층에서 합류를 겪어 변별되지 않으므로, 본 실험에서도 이를 구별하지 않고 'ㅐ'로 표기하여 실험을 진행하였다.[21] 이는 단모음 'ㅔ'와 'ㅐ'가 포함된 이중모음에서도 동일하다. 'ㅚ'와 'ㅟ'는 자연발화에서 단모음으로 발음되는 경우가 극히 드물고, 이중모음으로 발음될 때 'ㅚ'는 주로 [we], 'ㅟ'는 [wi]로 발음되는바, 'ㅚ, ㅟ, ㅙ, ㅞ' 중 'ㅙ'와 'ㅟ'만 실험 목록에 포함하였다.

로 인식하는지, 특정 음성을 특정 음소로 잘못 인식하는 경향이 있다면 거기에는 어떠한 음운론적 요인이 관여하는지 파악하고자 하는 본고의 연구 목적에서 벗어나는 경우라 판단하여, 종성 장애음을 다루는 5.1.1.에서 대략적인 인식 오류 양상을 제시하고, 그 외의 전체 인식 통계에는 종성 'ㄷ'의 인식 결과를 반영하지 않았다.

20) /ㅎ/의 음운론적 지위 문제는 한국어학에서 논쟁의 대상이나, 본 논의에서는 편의상 평음으로 간주한다.

21) 실제 실험 참가자들 중 'ㅔ'와 'ㅐ'를 변별하여 발음할 수 있는 음성 제공자는 없었다.

발화된 음성을 어떤 음소로 인식하는지, 특정 음성을 특정 음소로 잘못 인식하는 경향이 있다면 거기에는 어떠한 음운론적 요인이 관여하는지 파악하는 것이 본고의 실험 목적인바, 음절말 평파열음화나 자음군 단순화를 겪어서 'ㄱ, ㄷ, ㅂ'가 되는 어휘는 실험 대상 목록에 포함하지 않았다. 예를 들어 종성에서 [ㄱ]로 발음될 수 있는 음절말 자음으로는 'ㄱ'뿐만 아니라 'ㅋ, ㄹㄱ, ㄱㅅ'도 있지만, 본 논의에서는 음절말 자음 'ㄱ'만을 실험 대상으로 선정하였다는 것이다.

<표 3>의 음소들을 바탕으로 구성한 실험 대상 음절은 (C)V 133음절, (C)VC 798음절, (C)GV 57음절, (C)GVC 99음절로 총 1,087음절이었다. 이 중 (C)V와 (C)VC형은 조합 가능한 모든 음절을 대상으로 선정하였고, (C)GV, (C)GVC의 경우에는 자연 발화에서 발음 불가능한 음절 구성이 많아(예: '폑', '푠', '뒐', '뤌' 등) 실현 가능한 음절(예: '견', '뼘', '켠', '옙' 등)들을 선별하여 실험을 진행하였다. 2음절 이상의 어휘를 실험에서 제외한 것은 가급적 음운 현상의 적용을 받지 않은 상태에서의 인식 오류 관찰을 위해서였으나, 추후에는 다음절을 대상으로 한 음운론적 연구도 필요할 것으로 판단된다. 실험 대상 음절 목록은 아래의 <표 4>와 같다.

V 음절						
아	애	어	오	우	으	이

CV 음절													
ㅏ		ㅐ		ㅓ		ㅗ		ㅜ		ㅡ		ㅣ	
가	사	개	새	거	서	고	소	구	수	그	스	기	시
까	싸	깨	쌔	꺼	써	꼬	쏘	꾸	쑤	끄	쓰	끼	씨
나	자	내	재	너	저	노	조	누	주	느	즈	니	지
다	짜	대	째	더	쩌	도	쪼	두	쭈	드	쯔	디	찌
따	차	때	채	떠	처	또	초	뚜	추	뜨	츠	띠	치
라	카	래	캐	러	커	로	코	루	쿠	르	크	리	키
마	타	매	태	머	터	모	토	무	투	므	트	미	티
바	파	배	패	버	퍼	보	포	부	푸	브	프	비	피
빠	하	빼	해	뻐	허	뽀	호	뿌	후	쁘	흐	삐	히

CVC 음절																							
/ㄱ/ 종성							/ㄴ/ 종성							/ㄹ/ 종성						
ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
각	객	걱	곡	국	극	긱	간	갠	건	곤	군	근	긴	갈	갤	걸	골	굴	글	길
깍	깩	꺽	꼭	꾹	끅	끽	깐	깬	껀	꼰	꾼	끈	낀	깔	깰	껄	꼴	꿀	끌	낄
낙	낵	넉	녹	눅	늑	닉	난	낸	넌	논	눈	는	닌	날	낼	널	놀	눌	늘	닐
닥	댁	덕	독	둑	득	딕	단	댄	던	돈	둔	든	딘	달	댈	덜	돌	둘	들	딜
딱	땍	떡	똑	뚝	뜩	띡	딴	땐	떤	똔	뚠	뜬	띤	딸	땔	떨	똘	뚤	뜰	띨
락	랙	럭	록	룩	륵	릭	란	랜	런	론	룬	른	린	랄	랠	럴	롤	룰	를	릴
막	맥	먹	목	묵	믁	믹	만	맨	먼	몬	문	믄	민	말	맬	멀	몰	물	믈	밀
박	백	벅	복	북	븍	빅	반	밴	번	본	분	븐	빈	발	밸	벌	볼	불	블	빌
빡	빽	뻑	뽁	뿍	쁙	삑	빤	뺀	뻔	뽄	뿐	쁜	삔	빨	뺄	뻘	뽈	뿔	쁠	삘
삭	색	석	속	숙	슥	식	산	샌	선	손	순	슨	신	살	샐	설	솔	술	슬	실
싹	쌕	썩	쏙	쑥	쓱	씩	싼	쌘	썬	쏜	쑨	쓴	씬	쌀	쌜	썰	쏠	쑬	쓸	씰
작	잭	적	족	죽	즉	직	잔	잰	전	존	준	즌	진	잘	잴	절	졸	줄	즐	질
짝	짹	쩍	쪽	쭉	쯕	찍	짠	짼	쩐	쫀	쭌	쯘	찐	짤	짿	쩔	쫄	쭐	쯜	찔
착	책	척	촉	축	측	칙	찬	챈	천	촌	춘	츤	친	찰	챌	철	촐	출	츨	칠
칵	캑	컥	콕	쿡	큭	킥	칸	캔	컨	콘	쿤	큰	킨	칼	캘	컬	콜	쿨	클	킬
탁	택	턱	톡	툭	특	틱	탄	탠	턴	톤	툰	튼	틴	탈	탤	털	톨	툴	틀	틸
팍	팩	퍽	폭	푹	픅	픽	판	팬	펀	폰	푼	픈	핀	팔	팰	펄	폴	풀	플	필
학	핵	헉	혹	훅	흑	힉	한	핸	헌	혼	훈	흔	힌	할	핼	헐	홀	훌	흘	힐

/ㅁ/ 종성							/ㅂ/ 종성							/ㅇ/ 종성						
ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
감	갬	검	곰	굼	금	김	갑	갭	겁	곱	굽	급	깁	강	갱	겅	공	궁	긍	깅
깜	깸	껌	꼼	꿈	끔	낌	깝	깹	껍	꼽	꿉	끕	낍	깡	깽	껑	꽁	꿍	끙	낑
남	냄	넘	놈	눔	늠	님	납	냅	넙	놉	눕	늡	닙	낭	냉	넝	농	눙	능	닝
담	댐	덤	돔	둠	듬	딤	답	댑	덥	돕	둡	듭	딥	당	댕	덩	동	둥	등	딩
땀	땜	떰	똠	뚬	뜸	띰	땁	땝	떱	똡	뚭	뜹	띱	땅	땡	떵	똥	뚱	뜽	띵
람	램	럼	롬	룸	름	림	랍	랩	럽	롭	룹	릅	립	랑	랭	렁	롱	룽	릉	링
맘	맴	멈	몸	묨	믐	밈	맙	맵	멉	몹	묩	믑	밉	망	맹	멍	몽	뭉	믕	밍
밤	뱀	범	봄	붐	븜	빔	밥	뱁	법	봅	붑	븝	빕	방	뱅	벙	봉	붕	븡	빙
빰	뺌	뻠	뽐	뿜	쁨	삠	빱	뺍	뻡	뽑	뿝	쁩	삡	빵	뺑	뻥	뽕	뿡	쁭	삥
삼	샘	섬	솜	숨	슴	심	삽	샙	섭	솝	숩	습	십	상	생	성	송	숭	승	싱
쌈	쌤	썸	쏨	쑴	쓤	씸	쌉	쌥	썹	쏩	쑵	쓥	씹	쌍	쌩	썽	쏭	쑹	쓩	씽
잠	잼	점	좀	줌	즘	짐	잡	잽	접	좁	줍	즙	집	장	쟁	정	종	중	증	징
짬	쨈	쩜	쫌	쭘	쯤	찜	짭	짶	쩝	쫍	쭙	쯥	찝	짱	쨍	쩡	쫑	쭝	쯩	찡
참	챔	첨	촘	춤	츰	침	찹	챕	첩	촙	춥	츱	칩	창	챙	청	총	충	층	칭
캄	캠	컴	콤	쿰	큼	킴	캅	캡	컵	콥	쿱	큽	킵	캉	캥	컹	콩	쿵	킁	킹
탐	탬	텀	톰	툼	틈	팀	탑	탭	텁	톱	툽	틉	팁	탕	탱	텅	통	퉁	틍	팅
팜	팸	펌	폼	품	픔	핌	팝	팹	펍	폽	풉	픕	핍	팡	팽	펑	퐁	풍	픙	핑
함	햄	험	홈	훔	흠	힘	합	햅	헙	홉	훕	흡	힙	항	행	헝	홍	훙	흥	힝

VC 음절																					
/ㄱ/ 종성							/ㄴ/ 종성							/ㄹ/ 종성						
ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
악	액	억	옥	욱	윽	익	안	앤	언	온	운	은	인	알	앨	얼	올	울	을	일
/ㅁ/ 종성							/ㅂ/ 종성							/ㅇ/ 종성						
ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
암	앰	엄	옴	움	음	임	압	앱	업	옵	웁	읍	입	압	앱	업	옵	웁	읍	입

(C)GV 음절									
j계						w계			
ㅑ	ㅕ	ㅛ	ㅠ	ㅒ	ㅢ	ㅘ	ㅝ	ㅙ	ㅟ
야	여	요	유	예	의	와	워	왜	위
	겨	교	규	계	희	과	궈	궤	귀
	껴	묘	류	폐		봐	꿔	꽤	뉘
	며	표	큐	혜		좌	눠	퉤	쉬
	벼	효	퓨			콰	둬	훼	쥐
	뼈		휴			화	뭐		취
	켜						쒀		튀
	펴						줘		퀴
	혀						춰		휘
							훠		

(C)GVC 음절								
j계						w계		
ㅑ	ㅕ		ㅛ	ㅠ	ㅒ	ㅘ	ㅝ	ㅙ
약	역	멸	욕	육	옉	곽	권	왠
얀	연	명	욘	윤	옌	관	궐	왱
얄	열	역	욜	율	옐	괄	뭔	휑
얌	염	연	욤	윰	옘	괌	뭘	
얍	엽	열	욥	윱	옙	광	웍	
양	영	염	용	융	옝	좔	원	
	격	엽		균		확	월	
	견	영		귤		환	웜	
	결	혁		륜		활	훤	
	겸	현		률		황	훨	
	겹	혈		륭		콱	꿩	
	경	혐		휼		콴		
	벽	협		흉		콸		
	변	형		큘		쾅		
	별	뼘						
	볍	켠						
	병	켤						
	멱	편						
	면	펼						
		평						

<표 4> 실험 대상 음절 목록

<표 4>의 1,087개의 음절을 한 음절당 5번씩 발화하여 음성 제공자 한 사람당 일차적으로 5,435음절의 인식형이 수집되었고, 여기에 음성 인식 과정에서 디스플레이에 노출되는 중간 단계의 인식형을[22] 추가로 수집하여 최종적으로는 한 사람의 음

22) 중간 단계의 인식형에 대해서는 2.2.에서 상술된다.

성 제공자마다 10,870음절의 데이터가 생성되었다. 한편 빅스비는 인식된 결과가 사용자가 의도한 결과와 일치하는지 여부를 확인하여 학습에 활용하도록 설계되었는데, 만약 본 실험에서 음성 제공자가 인식 결과의 일치 여부에 반응을 해준다면, 다섯 번의 발화 중 후반부로 갈수록 발화의 인식 정확도가 높아져 실험상 하나의 변수가 될 수 있기에, 이를 방지하고자 인식 결과의 일치 여부에는 반응하지 않았다.

2.2. 전반적인 인식률

본 절에서는 2.1에서 소개한 음성 인식 오류 확인 실험의 전반적인 인식률 결과를 종합적으로 살펴보도록 한다. 실험 결과의 분석 절차는 다음과 같다. 먼저 음성대화형 어시스턴트의 인식 결과 값을 데이터베이스화하기 위해 필요한 부분들을 가공하고, 그 뒤 실험 대상 음절의 인식률을 '음절 인식률', '초성'·'중성'·'종성'별 인식률로 나누어 측정한다. 초성·중성·종성별 인식률은 인접 음소나, 발화자의 성별과 같은 여러 기준들을 설정하여 통계 처리하였다. 실험 결과를 계량화하기 위한 데이터 정리 기준을 제시하면 다음과 같다.

(4) 가. 음성 인식 결과가 숫자로 제시된 경우

① 해당 숫자가 하나의 음절형으로 대응되는 경우에는 해당 음절로 단순 변환함.

	실험 음절	인식 결과	수정
예시	'일'	'1'	'일'

② 해당 숫자가 두 가지의 음절형으로 대응되는 경우에는 상황에 따라 변환함.

	실험 음절	인식 결과	수정
예시	'영' 또는 '공'	'0'	'영' 또는 '공'
	'이' 또는 '투'	'2'	'이' 또는 '투'
	'사' 또는 '포'	'4'	'사' 또는 '포'
	'십' 또는 '텐'	'10'	'십' 또는 '텐'

나. 음성 인식 결과가 알파벳으로 제시된 경우: 해당 알파벳의 한글 표기에 따라 변환함.

	실험 음절	인식 결과	수정
예시	‘티’	‘T’	‘티’
	‘알’	‘R’	‘알’
	‘원’	‘ONE’	‘원’

다. 음성 인식 결과가 2음절 이상으로 제시된 경우

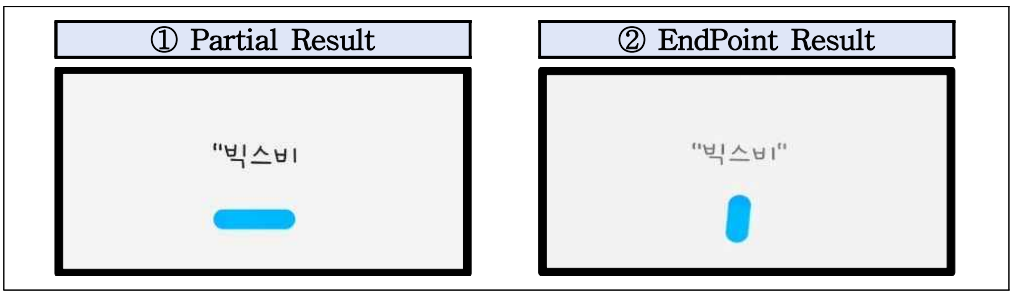
① 한 음절이 길게 발음된 것으로 인식한 경우라면 한 음절로 변환함

	실험 음절	인식 결과	수정
예시	‘젤’	‘제일’	‘젤’
	‘맬’	‘매일’	‘맬’

② 이외의 경우에는 변환 없이 오류로 처리함.

	실험 음절	인식 결과	수정
예시	‘깨’	‘게임’	×
	‘래’	‘그래’	×

한편 대부분의 음성대화형 어시스턴트들은 음성이 감지되면 음성 인식이 진행되는 동안 감지된 음성의 인식 결과를 즉각적으로 디스플레이에 표시한다. 그 뒤 발화가 종료되었다고 판단되면 음성 인식의 최종 결과형을 다시 제시하는데, 이때 중간 과정으로서의 인식 결과, 즉 ‘Partial Result’와 최종 판단 결과는 차이가 날 수 있다. 노희경·이강희(2017)에 따르면 ‘들려오는 중의 음성의 실시간 결과를 계속해서 보여주는 것’을 의미하는 이 Partial Result는, 서버를 거쳐 데이터를 보정하는 과정을 거치지 않으므로 정확한 결과값을 보여주지는 못한다고 한다. 본고에서 실험에 사용한 ‘Bixby 3.1.20.2’의 경우에는 음성이 감지되면 왼쪽 따옴표만 노출한 형태로 Partial Result를 제시하고, 그 뒤 발화가 종료되었다고 판단되면 양쪽 따옴표가 갖춰진 형태로 최종 인식형을 제시하였다. 실제 빅스비 인식 화면을 보이면 다음과 같다.



<그림> 삼성 빅스비의 음성 인식 화면

음성 인식 과정에서 접하게 되는 Partial Result와 EndPoint Result는 음향 모델뿐만 아니라 발음 사전과 언어 모델, 즉 탐색 엔진의 판단을 모두 거친 결과 값을 나타내는 것이지만, 본고와 같은 단음절 단독 발화의 실험 환경에서는 ①에서 참고할 만한 정보가 대체로 음향 정보에 한정되어 있을 것으로 예상된다.[23] 단음절 발화의 특성상 음성 인식 결과를 도출할 때 참고하는 문맥이나 선·후행 요소와의 관계에 따른 출현 확률 등의 정보가 제공되지 않기 때문이다. 실제 빅스비의 실행 과정에서 ①의 결과는 인식이 진행되는 동안 여러 차례 변경되기도 하나 ②는 최종적으로 하나의 형태로만 제시된다. 본고에서는 이러한 Partial Result와 EndPoint Result를 따로 기록하고 이들 각각을 통계 처리하였으며, 편의상 전자를 '중간형', 후자를 '최종형'으로 표현하기로 한다.

중간형과 최종형의 음성 인식 결과 중 본고에서는 중간형을 주 연구 대상으로 삼고 있다. 본고의 연구 목적이 발화된 음성을 어떤 음소로 인식하는지, 이때 관여하는 음운론적 요소가 있다면 무엇인지 확인하고자 하는 것이기에, 연구 목적에 따라 언어 모델의 영향이 강한 최종형이 아닌 중간형을 대상으로 삼은 것이다. 언어 모델의 영향 여부는 실제 음성 인식 결과형들의 관찰을 통해 확인할 수 있는데, 아래의

23) 물론 단음절 발화에서도 Partial Result에서부터 음향 정보 이외의 정보가 적극적으로 반영된 결과값이 제시되는 경우가 있다. 예를 들어 실험 대상 음절 '머'는 모든 음성 제공자의 발화가 Partial Result에서부터 예외 없이 '뭐'로 인식되었는데, '뭐'와 같이 단음절 구성이 동시에 하나의 문장 역할을 담당할 수 있는 경우에는, Partial Result에서부터 음향 모델 이외의 판단이 주효하게 작용하는 듯하다.

<표 5>에 제시한 실험 대상 음절 ‘만’의 인식 결과를 바탕으로 이를 자세히 살펴보도록 한다.

‘만’	중간형	최종형
F01	만,만,만,만,만	그만,그만,그만,그만,그만
F02	빤,만,만,만,만	만,그만,그만,그만,그만
F03	만,만,만,만,만	그만,그만,그만,그만,그만
F04	만,만,만,만,만	그만,그만,그만,그만,그만
F05	만,맘,맘,난,만	그만,램,그만,그만,맘
M01	만,만,만,만,만	그만,그만,그만,그만,그만
M02	만,만,만,만,만	그만,그만,그만,그만,그만
M03	만,만,만,만,만	그만,그만,그만,그만,그만
M04	말,만,만,만,만	그만,그만,그만,맘,그만
M05	만,만,그만,만,맘	그만,말,난,난,그만

<표 5> 실험 대상 음절 ‘만’의 음성 제공자별 인식 결과

실험 대상 음절 ‘만’은 <표 5>를 통해 확인할 수 있듯이 중간형에서는 ‘만’의 음향적 특성에 기반한 인식 결과가 나타난다. 인식 오류형 ‘맘’, ‘난’ 등은 실험 대상 음절 ‘만’과 음성적으로 혼동될 만한 형태로 보인다. 하지만 이와 같은 중간형의 인식 결과가 최종형에서는 대부분 ‘그만’으로 정정된다. 중간형과 최종형 사이에는 발화가 종료되었다는 정보 이외에는 추가된 정보가 없었으나, 중간형의 86%(43음절)가 ‘그만’으로 바뀐 것이다. 이는 음성비서와의 대화 상황에서 ‘만’이라는 단독 음절이 발화될 확률보다는 ‘그만’이 발화될 확률이 높을 것이라는 언어 모델의 판단이 개입된 결과인 듯하다. 이러한 예는 실험 대상 음절 ‘깁’의 인식 결과에서도 발견된다.

‘깁’	중간형	최종형
F01	힙,힙,킵,킵,킵	**집,집,집,집,집**
F02	키,킵,킵,키,킵	**집,집,집,집,**김
F03	깁,깁,기,깁,깁	**집,집,집,집,집**
F04	팁,킵,깁,기,킵	**집,집,집,집,집**
F05	급,깁,김,키,**집**	**집,집,집,집,집**
M01	킵,킵,킵,킵,킵	**집,집,집,집,집**
M02	깁,깁,깁,깁,깁	**집,집,집,집,집**
M03	팀,깁,깁,깁,깁	**집,집,집,집,집**
M04	깁,깁,깁,깁,깁	**집,집,집,집,집**
M05	킵,쉽,킵,킵,킵	**집**,김,**집,집,집**

<표 6> 실험 대상 음절 ‘깁’의 음성 제공자별 인식 결과

<표 6>에서 중간형의 인식 결과에는 실험 대상 음절 ‘깁’과 음성적으로 유사한 다양한 인식형이 등장한다(‘킵, 팁, 집, 힙’ 등). 그러나 최종형에서는 이 인식형들이 대부분 ‘집’으로 정정된다. 중간형에서는 두 차례(4%) 등장했던 오인식형 ‘집’이 최종형에서는 48차례(96%)로 증가하게 된 것이다. 이는 감지된 음성의 음향적 특성에 근거한 인식 결과라기보다는 사용 빈도나 출현 확률과 같은 음향 단서 이외의 정보들이 영향을 준 결과일 듯하다. 이처럼 최종형에는 발화음의 음향적 특성 이외의 추가적인 정보들이 개입될 여지가 높아, 제공된 음성을 어떤 음소로 모델링하는지 확인하기 어렵다고 판단하여 중간형을 주 연구 대상으로 삼은 것이다. 이러한 이유로 음성 인식 오류 양상을 구체적으로 살피는 3장~5장에서는 중간형의 인식 결과가 논의의 대상이 되나, 음성 인식 실험의 전반적인 인식률을 제시하는 본 절에서는 중간형과 최종형의 인식 결과를 모두 제시하도록 한다. 가장 먼저 제시할 음성 인식 결과는 10명의 음성 제공자 각각의 음절 인식률과 성별 인식률 평균이다.

인식률	F01	F02	F03	F04	F05
중간형	56.6	57.5	58.0	65.0	50.2
최종형	47.5	50.2	51.0	58.4	50.6

여성 평균
57.5
51.5

인식률	M01	M02	M03	M04	M05
중간형	40.9	70.1	59.2	58.7	58.3
최종형	37.4	60.3	41.8	49.3	43.8

남성 평균
57.4
46.5

<표 7> 음성 제공자들의 음절 인식률 평균

위의 <표 7>을 통해 확인할 수 있는 점은 다음과 같다. 주지하다시피 음성대화형 어시스턴트 사용자에게 제공되는 인식 결과 값은 최종형이다. 즉 이 최종형의 평균이 음성대화형 어시스턴트의 최종 인식률인 셈인데 여성은 51.5%, 남성은 46.5%로 측정되어 전체 평균은 과반에서 다소 모자란 49%로 도출되었다. 이는 곧 단음절 단독 발화의 경우 절반은 사용자의 의도대로 인식되는 반면, 절반은 의도와 다르게 오류가 발생한다는 것을 의미한다.[24] 한편 본고에서 논의 대상으로 삼고 있는 중간형의 경우에는 평균 인식률이 57.5%로 최종형보다 다소 높게 나타난다.[25]

실험 결과 여성과 남성 모두 중간형보다 최종형의 인식률이 낮게 측정되었는데, 단음절 단독 발화의 특성상 Partial Result와 EndPoint Result 사이에 추가된 정보가 없는 상황에서 탐색 엔진이 어떤 점을 고려하여 최종형을 교체한 것인지는 의문이다. 추정하기에는 Partial Result에서는 실험 대상 음절이 발화된 후 어떠한 발화가 추가될지 알지 못하는 상황에서 음성대화형 어시스턴트가 감지된 음향 정보를 인식의 주된 단서로 삼았을 듯하고, EndPoint Result에서는 이미 발화가 종료된 상황 즉 단음절 단독 발화라는 정보가 추가된 상태에서, 인식한 음성형이 단음절 단독

24) 음성대화형 어시스턴트를 일상에서 사용할 때 본고의 실험 상황과 같이 단음절만을 단독으로 발화하는 경우는 드물 것이다. 그러나 음절 경계를 명확히 하여 발음하는 음성 인식 발화의 음운론적 특성과, 조사가 비실현되는 문법적 특성상, '문장 안'에서는 단음절 구성이 발화되는 일이 생각보다 적지 않다. 예를 들어 "가을에 가기 좋은 산 찾아줘.", "○○ 가는 길 안내해 줘.", "'정인'(가수)의 '그 뻔한 말'(노래 제목) 들려 줘." 등과 같은 일상적인 발화 안에서도 단음절 구성이 포함된 경우를 쉽게 찾아볼 수 있다.

25) 본 연구의 음성 인식 실험 결과 도출된 평균 인식률은, 선행 연구들의 음성 인식 정확도에 비해 낮게 나타난다. 클라우드 기반 음성 인식 오픈 API의 연속 음성 인식 정확도를 비교 분석한 유현재·김명화·박상길·김광용(2020)의 실험 결과에서는, 오픈 API마다 차이는 있으나 평균 85.8%의 인식률을 보여서 본 연구 결과와 큰 차이를 보였다. 이는 본 연구의 실험 대상이 음성 인식 에이전트가 참고할 만한 정보가 가장 적은 단음절 발화였기 때문이며, 본고의 실험 상황과 달리 음성 인식 프로그램이 참고할 수 있는 문맥 및 의미에 관한 정보량이 늘어날 경우에는 인식 오류가 더 적어질 것으로 예상된다. 참고로 유현재·김명화·박상길·김광용(2020)에 제시된 오픈 API들의 인식률은 아래와 같다.

종류 / 정확도	Google	Amazon	IBM	MS	Naver	Kakao	ETRI
평균	76.76	91.31	70.94	91.27	85.99	94.00	90.27

발화로 출현할 확률과 기타 정보들을 고려하여 최종 인식형을 정정하였을 것으로 보인다. 즉 중간형에는 여러 단서들 중 음향 정보가, 최종형에는 음향 정보와 더불어 그 외의 정보들이 판단의 주요 근거가 되었을 것이라 추정하는 것이다. 이러한 관점을 받아들인다면 남녀 모두 중간형보다 최종형의 인식률이 낮게 측정된 것은, 음향 정보를 토대로 한 인식 결과를 언어 모델이 교정하는 가운데 오류가 추가 발생한 것으로 해석해 볼 수 있을 듯하다. 다음으로 살펴볼 것은 음절 구조 및 종성 자음들을 기준으로 측정한 음절 인식률 평균이다.

음절 인식률			여성 평균		남성 평균		전체 평균	
			중간형	최종형	중간형	최종형	중간형	최종형
V			63.4	74.9	61.1	70.3	62.3	72.6
VC	종성	ㄱ	30.9	37.7	37.7	26.9	34.3	32.3
		ㄴ	81.7	73.1	83.4	73.1	82.6	73.1
		ㄹ	89.1	75.4	92.0	80.0	90.6	77.7
		ㅁ	69.7	63.4	77.1	66.3	73.4	64.9
		ㅂ	64.6	30.3	56.6	21.1	60.6	25.7
		ㅇ	77.7	60.0	73.7	65.7	75.7	62.9
	평균		69.0	56.7	70.1	55.5	69.5	56.1
CV			71.9	59.2	66.2	51.5	69.1	55.3
CVC	종성	ㄱ	43.0	44.9	46.4	37.2	44.7	41.1
		ㄴ	62.0	54.5	59.0	50.9	60.5	52.7
		ㄹ	67.1	60.9	65.7	52.5	66.4	56.7
		ㅁ	57.7	56.1	65.3	50.5	61.5	53.3
		ㅂ	36.2	26.5	35.0	22.0	35.5	24.2
		ㅇ	61.5	56.7	60.1	52.3	60.8	54.5
	평균		54.6	49.9	55.3	44.2	54.9	47.1
CGV			56.5	47.0	55.3	45.0	56.0	46.0
CGVC			61.1	57.7	63.8	56.7	62.4	57.2

<표 8> 음절 구조 및 종성 자음별 음절 인식률 평균

<표 8>의 인식률 자료에서 주목해야 할 점도 대부분의 경우 중간형보다 최종형의 인식률이 낮다는 점이다. 여성 평균 중 'ㄱ' 종성을 가진 CVC음절만이 중간형보다 최종형의 인식률이 높았고, 나머지는 모두 '중간형 > 최종형'의 모습을 보인다. 최종형을 기준으로 볼 때 음절 구조별 인식률 순은 V > CGVC > VC > CV > CVC

> CGV로 나타났다. 중간형을 기준으로 볼 때는 이와 달리 VC > CV > CGVC > V > CGV > CVC인 것으로 확인된다. 본고에서는 음절 구조가 단순할수록 인식률이 높아질 것이라 예상하였으나('CV > CGV > CVC > CGVC' 혹은 'CV > CVC > CGV > CGVC'), 최종형에서는 음절 구조가 가장 복잡한 CGVC의 인식률이 비교적 높게 측정되었다. 이는 음절 구조가 복잡하다는 것이 음성 인식에 있어서 일종의 유표적 장치로 작용하였을 가능성과[26], 단순히 CGVC 구조를 가지는 실험 대상 목록이 가장 적은 데서 기인한 결과일 가능성이 있다. 음절 인식률에 있어서 성별에 따른 인식률 차이는 통계적 유의성을 가지지 않는 것으로 확인되었다(Chi-square test, p-value > 0.05). 하지만 음절 인식률이 아닌 초·중·종성별 인식률에서는 성별에 따른 유의미한 인식률 차가 나타나는 경우도 있다.

26) 예를 들어 음절 구조가 가장 단순한 '가'와 '바'의 경우에는, 이 둘을 구별할 단서가 초성의 음향 특질밖에 없지만, 이보다 음절 구조가 복잡한 '권, 뷘'의 경우에는 초성 자음의 음향 특질 외에도 이들이 실제 발화에서 사용되는 음절인지 여부도 인식에 참고할 만한 정보가 된다. 또 다른 예로 감지된 음성의 음절 구조가 CGV인 것으로 추정되고 이때 중성이 'we'라고 판단되었다면, 이에 따라 초성에 올 수 있는 자음의 수도 제한되므로 음성 인식 결과로 도출될 후보가 줄어들어 결과적으로 인식 정확도가 상승될 만한 환경이 조성될 수 있다.

3. 초성의 음성 인식 오류 양상 및 유사 음소

본 장에서는 음성 인식 오류의 확인 실험 결과 관찰된 음성 인식 오류 양상에 대해 살피고 각 환경별 유사 음소를 제시하도록 한다. 단음절 단독 발화를 대상으로 한 본고의 실험 결과 음성대화형 어시스턴트의 전체 인식률은 앞서 2.2.에서 제시한 바와 같이 57.5%로 도출되었다. 전체적인 인식률은 비교적 낮게 나타났으나, 환경을 나누어 미시적으로 살피면 그 중에서도 음성 인식 오류가 거의 발생하지 않는 환경과 빈번하게 발생하는 환경이 나누어질 수 있을 것이다. 본 장에서는 이 중 인식 오류가 빈번히 발생하는 환경에서 한 음소를 특정 음소로 오인식하는 경향이 나타날 때 이를 해당 음소의 '유사 음소'로 간주하고, 이러한 오류가 음운론적 요인에 의한 것인지 비음운론적 요인에 의해 유발된 것인지 판단한 뒤, 그 중 전자를 중심으로 하여 어떠한 음운론적 요인이 음성 인식 오류에 영향을 미치는지 알아보고자 한다.

3.1. 오류 양상 및 유사 음소

본 절에서는 음성 인식 오류 확인 실험의 초성 인식 결과를 전반적으로 살펴보도록 한다. 초성의 인식 오류 양상은 후행 모음을 기준으로 분석하였으며, 이에 모든 후행 모음 환경에서의 인식 오류 실험이 진행된 CV음절과 CVC음절을 중심으로 인식률을 제시하고자 한다. CV음절 구조의 초성 인식률 평균은 81.4%, CVC음절의 초성 인식률 평균은 78.0%였으며, 이 둘을 합한 전체 평균은 79.7%이다. 음절 구조에 따라 차이를 보이는 인식률 값이 음절 구조에 의한 차이인지 확인해 보고자 Chi-square test를 진행하였고, 그 결과 p-value가 0.05 이하로 도출되어 초성의 인식률 차이는 음절 구조와 관련된 것으로 판단하였다($p\text{-value} < .001$). 이제 음절 구조와 더불어 조음 방법에 따른 자음 유형을 기준으로 초성의 전체 인식률을 정리한

<표 9>와, 이를 인식률 순서대로 간단히 다시 정렬한 <표 10>을 살펴보도록 한다.

유형 / 음절구조	파열음			마찰음	파찰음	공명음
	평음	격음	경음			
CV	59.2	82.9	81.9	90.5	84.3	88.0
CVC	45.0	84.4	82.6	87.7	81.0	87.0

<표 9> 음절 구조·자음 유형별 인식률(%)

CV	마찰음	공명음	파찰음	격음	경음	평음
CVC	마찰음	공명음	격음	경음	파찰음	평음

<표 10> 음절 구조·자음 유형별 인식률 순

<표 9>를 통해 CV음절 구조에서는 마찰음 > 공명음 > 파찰음 > 격음 > 경음 > 평음 순으로 인식률이 낮아지고, CVC음절 구조에서는 마찰음 > 공명음 > 격음 > 경음 > 파찰음 > 평음 순으로 인식률이 낮아진다는 점이 확인된다. <표 10>을 통해서는 마찰음과 공명음의 인식률이 상대적으로 높고, 파열음 중 평음의 인식률이 가장 낮다는 점과 함께, 파찰음과 격음, 경음이 중간에 위치해 있다는 점을 알 수 있다. 이때 파찰음·격음·경음 간의 인식률은 크게 차이가 나지 않는다.

인식률이 높은 마찰음과 공명음은 주파수 대역이 가장 극단에 위치해 있다. 마찰음은 고주파수 대역에, 공명음은 저주파수 대역에 에너지가 분포되어 있어 음성 인식 상황에서 '마찰음'과 '공명음'이라는 범주적 특성이 비교적 쉽게 확인될 수 있는 것으로 보인다. 특히 마찰음은 스펙트로그램상에 나타나는 마찰로 인한 고주파수 대역의 소음부가[27] 가장 유표적인 변별 표지가 된 듯하다.[28] 공명음도 이와 마찬가지

27) 신지영(2014: 238, 244)에서 참고하였다.

28) Google의 Speech Recognition Soundwriter을 통해 영어 음성 인식 실험을 진행한 양병곤(2017)에서는 이와 달리 마찰음이 음성 인식 오류의 주요 원인으로 분석되기도 하였다. 이 논의에서는 마찰음의 공명도가 음향적으로 다른 자음에 비해 상대적으로 낮다는 점과, 스펙트럼상에서 4000~8000Hz 사이에 나타나는 절단주파수(cutoff frequency) 부분이 적절히 입력이 되지 않거나 활용되지 못하는 점이 인식 오류를 유발하는 것으로 추정하였다. 본고와 양병곤(2017)의 상반된 인식 결과는 두 논의가 서로 다른 언어를 대상으로 하고 있다는 점과,

로 비음은 모음에 비해 음향 에너지가 매우 약하고 저주파수 쪽에 에너지가 몰려 나타나는 특성이,[29] 유음은 모음에 비해 고주파수 대역의 음향 에너지가 약하게 나타나는 특성이 있어,[30][31] 전체 음소들 중 공명음을 변별해 내는 것을 용이하게 만들어 결과적으로 높은 인식률을 나타내게 하였다. 파찰음도 마찰음과 마찬가지로 구강 폐쇄의 개방으로 인해 만들어지는 수직의 스파이크를 가지고 있으나, 파찰음의 조음 과정 중 '폐쇄' 단계의 묵음에 가까운 음향 특성이 음성 인식의 방해 요소로 작용하여 마찰음에 비해 낮은 인식률을 보이는 것으로 판단된다.

파열음의 인식률이 낮게 나타나는 원인은, 파열음의 발화 과정에서 나타나는 '폐쇄-지속-파열' 단계 중 '폐쇄' 단계의 음향적 특성이 스펙트로그램상 거의 드러나지 않는 데서 찾을 수 있다. 배문정·김정오(2002)에서는 '기식성, 긴장성/이완성, 공명성, 치찰성, 설정성'이 자음의 지각에서 가장 중요한 지각적 특질임을 지적한 바 있는데, 파열음 중 경음과 격음은 '긴장'이나 '기식'과 같은 유표적인 특성을 가지고 있어 평음보다는 인식이 용이하나, 주요 지각적 특질을 가지지 않은 평음은 초성 자음의 음성 인식 상황에서 가장 많은 오류를 보인다.

다음으로는 개별 초성들의 인식률을 음절 구조와 후행 모음을 기준으로 정리한 결과를 살펴보도록 한다. 개별 초성들은 파열음, 파찰음, 마찰음, 공명음 순으로 배열하였으며, 파열음은 다시 평음·경음·격음 순으로 인식률을 제시하였다.

두 음성대화형 어시스턴트에 적용된 음성 인식 알고리즘의 차이에서 기인한 것으로 판단된다.

29) Kim(2011)에 따르면 특히 어두의 비음은 nasal formant가 나타나지 않고, waveform에서 눈에 띄는 진폭의 감소가 관찰된다는 점에서 어중의 비음과 차별화된 특징을 보인다고 한다.

30) 이들은 본고의 실험 상황과 같은 '어두' 환경에서 두드러지게 나타나는 특징이다.

31) 신지영(2014: 247, 252)에서 참고하였다.

후행 모음 \ 초성		인식률								
		ㄱ	ㄷ	ㅂ	ㄲ	ㄸ	ㅃ	ㅋ	ㅌ	ㅍ
_V	ㅏ	82.0	84.0	52.0	94.0	70.0	86.0	90.0	78.0	64.0
	ㅐ	88.0	66.0	70.0	96.0	62.0	82.0	72.0	82.0	68.0
	ㅓ	42.0	92.0	12.0	98.0	76.0	40.0	86.0	64.0	96.0
	ㅗ	36.0	44.0	26.0	94.0	100.0	76.0	96.0	94.0	86.0
	ㅜ	84.0	70.0	70.0	92.0	94.0	74.0	96.0	96.0	82.0
	ㅡ	88.0	36.0	12.0	78.0	90.0	54.0	94.0	90.0	40.0
	ㅣ	66.0	58.0	66.0	90.0	86.0	88.0	100.0	70.0	96.0
	평균	69.4	64.3	44.0	91.7	82.6	71.4	90.6	82.0	76.0
_VC	ㅏ	63.3	53.3	65.7	81.3	81.0	70.7	75.0	76.0	82.3
	ㅐ	57.7	40.0	61.3	74.3	73.3	67.3	94.0	87.7	81.0
	ㅓ	39.7	47.0	44.7	95.0	88.7	79.7	90.7	84.0	87.7
	ㅗ	22.7	38.7	23.7	94.7	95.0	83.0	94.0	90.3	78.7
	ㅜ	38.0	46.0	37.7	96.0	86.7	90.0	97.3	86.7	93.7
	ㅡ	56.7	47.7	15.7	96.7	96.0	50.3	97.0	84.7	44.7
	ㅣ	57.3	24.0	64.0	68.3	84.7	82.3	90.0	64.7	92.0
	평균	47.9	42.4	44.7	86.6	86.5	74.8	91.1	82.0	80.0
_GV		77.8	56.0	74.0	71.1		86.0	50.4	32.0	58.3
총 평균		65.0	54.2	54.2	83.1	84.6	77.4	77.4	65.3	71.4

<표 11> 초성의 전체 인식률(파열음)

후행 모음 \ 초성		인식률								
		ㅈ	ㅉ	ㅊ	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
_V	ㅏ	86.0	98.0	98.0	94.0	98.0	86.0	100.0	96.0	88.0
	ㅐ	82.0	86.0	98.0	98.0	100.0	98.0	100.0	56.0	66.0
	ㅓ	90.0	72.0	94.0	100.0	94.0	86.0	98.0	100.0	60.0
	ㅗ	74.0	80.0	100.0	100.0	98.0	88.0	96.0	98.0	90.0
	ㅜ	54.0	88.0	100.0	90.0	80.0	90.0	98.0	96.0	84.0
	ㅡ	74.0	82.0	80.0	90.0	82.0	90.0	92.0	74.0	84.0
	ㅣ	68.0	76.0	90.0	64.0	88.0	86.0	82.0	98.0	92.0
	평균	75.4	83.1	94.3	90.9	91.4	89.1	95.1	88.3	80.6
_VC	ㅏ	62.3	98.3	100.0	92.7	92.0	93.3	88.7	91.0	92.0
	ㅐ	69.7	92.0	95.0	95.0	53.3	92.7	86.7	88.0	92.7
	ㅓ	60.7	97.0	82.7	100.0	98.0	93.3	88.3	91.7	87.3
	ㅗ	62.0	95.3	88.5	95.0	94.3	90.0	87.3	97.3	94.0
	ㅜ	43.3	91.7	100.0	91.3	76.3	93.3	85.0	93.3	87.0
	ㅡ	56.3	92.3	88.5	78.0	91.7	96.3	91.0	67.3	68.7
	ㅣ	58.7	79.3	82.7	91.7	43.3	90.0	80.0	85.0	85.7
	평균	59.0	92.3	91.1	92.0	78.4	92.7	86.7	87.7	86.8
_GV		82.7				56.0	83.9	88.0	77.5	78.7
총 평균		72.4	87.7	92.7	91.5	75.3	88.6	89.9	84.5	82.0

<표 12> 초성의 전체 인식률(파열음 외)

이후에는 위의 초성 자음들의 음성 인식 오류 양상을 음절 구조와 후행 모음을 기준으로 구체적으로 살피고, 유사 음소의 설정 과정을 통해 각 초성별로 가장 빈번하게 오인식되는 음소가 무엇인지 확인해보고자 한다.

3.1.1. 파열음

3.1.1.1. 평음

본 절에서는 음성 인식 오류의 확인 실험 결과를 통해 분석된 초성 파열음의 인식 오류 양상을 살피고, 후행 모음별 유사 음소에 대해 알아보도록 한다. 먼저 제시

할 것은 파열음 중 평음 ‘ㄱ, ㄷ, ㅂ’의 전체 인식 결과이다.

음절구조	후행모음 \ 초성	ㄱ	ㄷ	ㅂ	평균
CV	ㅏ	82.0(41)	84.0(42)	52.0(26)	72.7
	ㅐ	88.0(44)	66.0(33)	70.0(35)	74.7
	ㅓ	42.0(21)	92.0(46)	12.0(6)	48.7
	ㅗ	36.0(18)	44.0(22)	26.0(13)	35.3
	ㅜ	84.0(42)	70.0(35)	70.0(35)	74.7
	ㅡ	88.0(44)	36.0(18)	12.0(6)	45.3
	ㅣ	66.0(33)	58.0(29)	66.0(33)	63.3
인식률 총 평균		69.4	64.3	44.0	
		59.2%			
CVC	ㅏ	63.3(190)	53.3(160)	65.7(197)	60.8
	ㅐ	57.7(173)	40.0(120)	61.3(184)	53.0
	ㅓ	39.7(119)	47.0(141)	44.7(134)	43.8
	ㅗ	22.7(68)	38.7(116)	23.7(71)	28.4
	ㅜ	38.0(114)	46.0(138)	37.7(113)	40.6
	ㅡ	56.7(170)	47.7(143)	15.7(47)	40.0
	ㅣ	57.3(172)	24.0(72)	64.0(192)	48.4
인식률 총 평균		47.9	42.4	44.7	
		45.0%			

<표 13> 초성 파열음 중 평음의 인식률(%)

위의 표는 CV음절과 CVC음절에서의 초성의 인식률을 후행 모음을 기준으로 나누어 제시한 것이며, 해당 환경에서의 인식률은 소수점으로, 출현 빈도는 괄호 안의 숫자로 나타냈다. 각 초성별 평균 인식률보다 인식률이 낮거나 같은 경우에는 표 안에 음영을 넣어 표시하였다.

<표 13>을 통해 알 수 있는 바와 같이 CV음절의 초성 평파열음 인식률 평균은 59.2%이며, CVC음절의 인식률 평균은 45.0%, 이들을 합한 초성 평파열음 전체 인식률은 52.1%로 과반 정도의 수치로 나타난다. 각 자음별 인식률 평균은 ㄱ > ㄷ > ㅂ 순으로 낮아진다. 초성 파열음의 인식률이 이처럼 낮게 나타나는 이유는, 전술한 바와 같이 스펙트로그램을 기반으로 추출한 음향 특질들을 통해 음소를 모델링하는 현재의 음성 인식 알고리즘의 특성 때문인 것으로 보인다. 주지하다시피 파열

음의 발화 과정에서 나타나는 '폐쇄-지속-파열' 단계 중 '폐쇄' 단계의 음향적 특성은 스펙트로그램상 거의 드러나지 않는다. 이로 인해 그리 길지 않은 '파열' 단계의 음향 특성만으로 초성에 사용되는 아홉 종류의 파열음을 변별하는 과정에서 음성 인식 오류가 유발되는 것으로 판단된다.[32]

다음으로는 각 평파열음들의 개별적 인식 오류 양상과 유사 음소에 대해 구체적으로 살펴보도록 한다. 먼저 'ㄱ'의 경우이다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㄱ⇒ㄱ	82.0(41)	88.0(44)	42.0(21)	36.0(18)	84.0(42)	88.0(44)	66.0(33)
	ㄱ⇒ㅋ	16.0(8)	12.0(6)	34.0(17)	56.0(28)	14.0(7)	12.0(6)	32.0(16)
	ㄱ⇒ㄲ	-	-	14.0(7)	-	-	-	-
CVC	ㄱ⇒ㄱ	63.3(190)	57.7(173)	39.7(119)	22.7(68)	38.0(114)	56.7(170)	57.3(172)
	ㄱ⇒ㅋ	19.7(59)	36.0(108)	51.7(155)	60.0(180)	50.3(151)	32.3(97)	30.0(90)
	ㄱ⇒ㄲ	0.3(1)	0.3(1)	3.3(10)	9.0(27)	7.3(22)	7.3(22)	7.0(21)
	ㄱ⇒ㅌ	6.3(19)	1.3(4)	-	-	-	0.3(1)	1.0(3)
	ㄱ⇒ㅎ	1.7(5)	0.7(2)	3.3(10)	6.7(20)	1.0(3)	0.3(1)	0.7(2)

<표 14> 초성 'ㄱ'의 후행 모음별 인식률[33]

CV음절 초성 'ㄱ'의 인식률 평균은 69.4%, CVC음절 평균은 47.9%이며, CV음절에

32) 사람을 대상으로 본 연구와 같은 '단음절' 발화의 지각 실험을 진행한 김연화(2015: 63, 65)에서는 CV구조 평파열음의 인식률 평균이 92.4%로 도출되었다. 인간의 지각 결과와 비교할 때에도 음성 인식 에이전트의 평파열음 인식률은 매우 낮은 수준이다. 김연화(2015)에 제시된 평파열음의 개별 인식률은 각각 ㄱ(97.9%), ㄷ(91.7%), ㅂ(87.5%)로 나타난다.

33) 실제 실험 결과에서는 <표 14> 제시된 자음 'ㄱ, ㅋ, ㄲ, ㅌ, ㅎ' 외에도 'ㄷ, ㅂ, ㅅ, ㅈ, ㅊ, ㅌ, ㅍ'가 오인식 자음으로 나타났으나(CVC음절), 해당 자음들은 모두 10회 이하의 빈도를 보였다(ㄷ(2.7%, 8회), ㅂ(1.7%, 5회), ㅅ(1.0%, 3회), ㅈ(0.3%, 1회), ㅊ(0.3%, 1회), ㅍ(1.3%, 4회)). 이처럼 빈도 수 10회 이하로 나타나는 오인식형들은 소수의 인식 오류인 것으로 판단하여 편의상 표에 담지 않았다. 단 특정 후행 모음 환경에서의 오인식률이 높아 인식 결과를 표시한 음소의 경우에는, 나머지 모음 환경에서 10회 이하의 빈도로 출현하였어도 그 결과를 모두 제시하였다. 또한 오인식 실험 데이터 수 자체가 적거나, 오류 발생 빈도가 매우 적은 경우에는 소수의 출현 빈도를 가지는 예들을 함께 제시하기도 하였다.

서는 후행 모음이 'ㅓ, ㅗ, ㅣ'일 때, CVC음절에서는 후행 모음이 'ㅓ, ㅗ, ㅜ'일 때 평균 이하의 인식률을 보인다. 초성 'ㄱ'를 주로 어떤 자음으로 오인식하는지 살펴본 결과, CV음절과 CVC음절에서 모두 같은 연구개음 계열인 'ㅋ'와 'ㄲ'로 오인식되는 양상이 나타나며, CVC음절의 경우에는 여기에 더하여 'ㅎ'도 오인식 자음으로 나타난다. 본고에서는 여러 오인식 자음들 중에서 가장 높은 빈도로 오인식된 자음을 음영으로 표시하고, 이를 해당 환경에서의 '유사 음소'로 간주하고자 하는데, 'ㄱ'의 경우 모든 후행 모음 환경에서 가장 높은 빈도로 오인식된 자음이 'ㅋ'이므로 초성 'ㄱ'의 유사 음소는 'ㅋ'가 된다. 특정 환경(CVC구조 'ㅓ, ㅗ, ㅜ' 앞)에서는 'ㄱ'로 인식되는 비율보다 'ㅋ'로 오인식되는 비율이 더 높게 나타나기도 한다.

위의 음성 인식 오류 결과를 보며 이제 관심이 가는 것은 초성 'ㄱ'를 다른 평음이나 경음이 아닌 격음으로 오인식한 이유이다. 이와 관련하여 신지영(2014: 226)에서는 어두 위치의 평음과 격음은 모두 기식성을 가지고 산출된다는 점에서 경음과 구분되는 특성을 가지는 것으로 보았는데, 이는 곧 어두 위치의 평음은 경음보다는 격음과 유사하다는 의미로, 본 연구의 실험 결과 초성의 평음 'ㄱ'가 'ㅋ'로 오인식된 데에는 이와 같은 평음과 격음의 공통된 음향 특성이 기본적으로 관여했을 것으로 보인다. 이와 더불어 평음과 격음의 변별에 이용되는 F0값도 인식 오류의 주요 영향 요인으로 작용한 듯하나,[34] 이는 격음의 인식 오류를 다루는 3.1.1.3.에서 후술한다.

평파열음 'ㄱ'가 마찰음인 'ㅎ'로 인식되는 현상은, 후술할 평파열음 'ㅂ'와, 격음 'ㅋ, ㅍ'의 음성 인식 상황에서도 발견된다. 파열음의 조음 단계 중 '파열' 과정이 제대로 인식되지 못할 경우에는 '기식' 특성만 남아 이것이 'ㅎ'로 오인식될 수 있는데, 이와 같은 현상이 기식이 없는 경음에서는 나타나지 않는다는 점에 주목해 볼 때, 'ㄱ, ㅂ'계열의 'ㅎ'로의 오인식은 기식성을 가지는 파열음들의 '파열' 단계의 부정확

34) 추가적으로 초성 평음 'ㄱ'를 수차례 발화하여 이를 'ㄱ'로 인식할 때의 발화음의 음향적 특성과, 'ㅋ'로 인식할 때의 음향적 특성을 각각 녹음하여 VOT와 F0값(모음 시작 시점 및 안정 구간 측정)을 측정해본 결과, VOT 및 안정 구간의 F0값은 평음을 격음으로 오인식하는 데 영향을 주지 않는 것으로 나타났으나, 시작 지점의 F0값은 평음/격음 인식에 통계적으로 유의미한 차이를 나타내는 것으로 확인되었다. 연구자 한 명의 발화 실험 결과를 일반화할 수는 없지만 이를 바탕으로 평음/격음의 인식 오류와 F0값을 연결 지어 생각해 볼 수는 있을 듯하다.

한 인식에 의해 발생한 오류로 이해할 수 있다.

한편 후술할 'ㄷ'나 'ㅂ'와 달리 'ㄱ'의 경우에는 치조음 계열로 오인식되는 소수의 예들을 제외하면 조음 위치상의 오류는 거의 나타나지 않는다. 또한 연구개음 계열의 파열음이 여타의 파열음보다 평음·경음·격음 모두 높은 인식률을 보이는데, 이는 연구개음이 가진 [compactness] 자질과 관련된 것으로 보인다. 연구개음은 양순음이나 치조음과 달리 추가적인 음향 단서인 [compactness]를 가진다. 이 [compactness] 자질은 Jakobson·Fant·Halle(1951)에서 '중앙에 위치한 하나의 포먼트 영역이 상대적으로 돋들리는 특징'으로 정의된 바 있으며, 연구개 혹은 경구개에서 조음되는 자음들이 [+compactness] 해당된다. 연구개 파열음은 이와 같은 특성으로 인해 치조 파열음과 양순 파열음에 비해 조음 위치상의 오류가 적고, 높은 인식률을 보이게 된 듯하다.

다음으로는 'ㄷ'의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㄷ⇒ㄷ	84.0(42)	66.0(33)	92.0(46)	44.0(22)	70.0(35)	36.0(18)	58.0(29)
	ㄷ⇒ㅌ	12.0(6)	16.0(8)	6.0(3)	30.0(15)	28.0(14)	30.0(15)	20.0(10)
	ㄷ⇒ㅃ	-	-	-	18.0(9)	-	2.0(1)	6.0(3)
	ㄷ⇒ㄱ	-	2.0(1)	-	-	-	28.0(14)	6.0(3)
CVC	ㄷ⇒ㄷ	53.3(160)	40.0(120)	47.0(141)	38.7(116)	46.0(138)	47.7(143)	24.0(72)
	ㄷ⇒ㅌ	33.7(101)	34.3(103)	32.0(96)	47.0(141)	40.0(120)	23.3(70)	32.0(96)
	ㄷ⇒ㅃ	-	7.0(21)	2.3(7)	4.3(13)	6.7(20)	17.3(52)	19.7(59)
	ㄷ⇒ㄱ	-	0.3(1)	-	-	-	2.3(7)	11.0(33)

<표 15> 초성 'ㄷ'의 후행 모음별 인식률

CV음절 초성 'ㄷ'의 인식률 평균은 64.3%, CVC음절 평균은 42.4%이며, CV음절에서는 후행 모음이 'ㅡ, ㅣ, ㅗ'일 때, CVC음절에서는 후행 모음이 'ㅐ, ㅣ, ㅗ'일 때 평균 이하의 인식률을 보인다. 초성 'ㄷ'를 주로 어떤 자음으로 오인식하는지 살펴본 결과, CV음절과 CVC음절에서 모두 같은 치조음 계열인 'ㅌ'로 오인식되는 비율이

높게 나타나, 초성 ‘ㄷ’의 유사음소는 ‘ㅌ’인 것으로 판단된다. CVC음절의 경우 후행 모음이 ‘ㅣ’일 때 ‘ㄷ’의 인식률이 24.0%에 그치는데, 이때 오인식되는 자음으로 같은 치조 파열음 계열의 ‘ㅌ’, ‘ㄸ’ 외에 ‘ㄱ’도 나타난다. 자연 발화에서는 일반적으로 ‘ㅣ’ 모음 앞의 ‘ㄷ’는 ‘ㅈ’와 유사한 것으로 인식되나, 음성 인식 상황에서는 ‘ㄱ’와 더 가까운 것으로 판단되는 듯하다.

다음으로는 ‘ㅂ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅂ⇒ㅂ	52.0(26)	70.0(35)	12.0(6)	26.0(13)	70.0(35)	12.0(6)	66.0(33)
	ㅂ⇒ㅍ	12.0(6)	2.0(1)	38.0(19)	42.0(21)	14.0(7)	8.0(4)	30.0(15)
	ㅂ⇒ㄱ	-	-	-	-	-	64.0(32)	2.0(1)
	ㅂ⇒ㄷ	18.0(9)	-	30.0(15)	-	2.0(1)	2.0(1)	-
	ㅂ⇒ㅌ	6.0(3)	2.0(1)	2.0(1)	2.0(1)	4.0(2)	10.0(5)	-
	ㅂ⇒ㅎ	6.0(3)	14.0(7)	2.0(1)	2.0(1)	6.0(3)	4.0(2)	-
	ㅂ⇒∅	6.0(3)	6.0(3)	12.0(6)	14.0(7)	-	-	-
CVC	ㅂ⇒ㅂ	65.7(197)	61.3(184)	44.7(134)	23.7(71)	37.7(113)	15.7(47)	64.0(192)
	ㅂ⇒ㅍ	27.3(82)	15.7(47)	43.0(129)	41.7(125)	42.7(128)	22.0(66)	24.0(72)
	ㅂ⇒ㅃ	-	-	2.0(6)	15.7(47)	10.0(30)	8.0(24)	4.3(13)
	ㅂ⇒ㄱ	-	-	-	-	0.3(1)	14.0(42)	2.0(6)
	ㅂ⇒ㅎ	1.0(3)	2.7(8)	3.0(9)	8.3(25)	1.7(5)	3.3(10)	0.7(2)
	ㅂ⇒∅	2.0(6)	12.3(37)	3.0(9)	3.3(10)	0.3(1)	8.7(26)	2.3(7)

<표 16> 초성 ‘ㅂ’의 후행 모음별 인식률

CV음절 초성 ‘ㅂ’의 인식률 평균은 44.0%, CVC음절 평균은 44.7%이며, CV음절에서는 후행 모음이 ‘ㅓ, ㅗ, ㅡ’일 때, CVC음절에서는 후행 모음이 ‘ㅓ, ㅗ, ㅡ, ㅜ’일 때 평균 이하의 인식률을 보인다. 초성 ‘ㅂ’를 주로 어떤 자음으로 오인식하는지 살펴본 결과, CV음절과 CVC음절에서 모두 같은 양순음 계열의 격음 ‘ㅍ’로 오인식되는 비율이 높게 나타나, 초성 ‘ㅂ’의 유사음소는 ‘ㅍ’인 것으로 판단된다. CV음절 ‘ㅂ’의 경우에는 ‘ㅍ’ 이외에 같은 평음 계열에 해당되는 ‘ㄱ, ㄷ’로의 오인식 예도 관찰된다. 이 중 ‘ㅂ’의 ‘ㄱ’로의 오인식은 양순음과 연구개음이 가진 음향적 유사성에 의

해 발생한 오류이며, 특히 파열음의 조음 위치에 관한 음향적 단서를 가진 인접 모음으로부터의, 혹은 인접한 모음으로의 F2 전이 양상과 관련지어 이해할 수 있다.[35] 이에 대해서는 음향적 특성에 의한 인식 오류 유형을 다루는 6.2.에서 후술된다.

3.1.1.2. 경음

본 절에서는 파열음에 해당되는 초성 자음 중 경음 'ㄲ, ㄸ, ㅃ'의 인식률과 오류 환경에 따른 유사 음소에 대해 알아보도록 한다. 가장 먼저 제시할 것은 'ㄲ, ㄸ, ㅃ'의 전체 인식률 결과이다.

음절구조	후행모음 \ 초성	ㄲ	ㄸ	ㅃ	평균
CV	ㅏ	94.0(47)	70.0(35)	86.0(43)	83.3
	ㅐ	96.0(48)	62.0(31)	82.0(41)	80.0
	ㅓ	98.0(49)	76.0(38)	40.0(20)	71.3
	ㅗ	94.0(47)	100.0(50)	76.0(38)	90.0
	ㅜ	92.0(46)	94.0(47)	74.0(37)	86.7
	ㅡ	78.0(39)	90.0(45)	54.0(27)	74.0
	ㅣ	90.0(45)	86.0(43)	88.0(44)	88.0
인식률 총 평균		91.7	82.6	71.4	
		81.9%			
CVC	ㅏ	81.3(244)	81.0(243)	70.7(212)	77.7
	ㅐ	74.3(223)	73.3(220)	67.3(202)	71.6
	ㅓ	95.0(285)	88.7(266)	79.7(239)	87.8
	ㅗ	94.7(284)	95.0(285)	83.0(249)	90.9
	ㅜ	96.0(288)	86.7(260)	90.0(270)	90.9
	ㅡ	96.7(290)	96.0(288)	50.3(151)	81.0
	ㅣ	68.3(205)	84.7(254)	82.3(247)	78.4
인식률 총 평균		86.6	86.5	74.8	
		82.6%			

<표 17> 초성 파열음 중 경음의 인식률

35) Raphael et al.(1980/2011: 214-216). 소신애(2019: 62)에서 재인용함.

<표 17>을 통해 확인할 수 있는 CV음절 초성 경음의 인식률 평균은 81.9%이며, CVC음절의 인식률 평균은 82.6%, 이들을 합한 초성 경음 전체의 인식률은 82.3%로 확인된다. 각 자음별 인식률은 ㄲ > ㄸ > ㅃ 순으로 낮아진다. CVC음절의 경우 후행 모음이 'ㅏ, ㅐ'일 때 공통적으로 인식률이 낮으며, 'ㄲ, ㄸ'는 'ㅣ' 모음이 후행할 때, 'ㅃ'는 'ㅡ' 모음이 후행할 때 개별적으로 낮은 인식률을 보인다.

다음으로는 각 경음들의 개별적 인식 오류 양상과 유사 음소에 대해 구체적으로 살펴보도록 한다. 먼저 'ㄲ'의 경우이다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㄲ⇒ㄲ	94.0(47)	96.0(48)	98.0(49)	94.0(47)	92.0(46)	78.0(39)	90.0(45)
	ㄲ⇒ㄱ	-	-	-	-	-	22.0(11)	6.0(3)
CVC	ㄲ⇒ㄲ	81.3(244)	74.3(223)	95.0(285)	94.7(284)	96.0(288)	96.7(290)	68.3(205)
	ㄲ⇒ㄱ	-	19.3(58)	2.0(6)	1.0(3)	-	0.3(1)	20.3(61)
	ㄲ⇒ㄸ	16.3(49)	2.3(7)	0.7(2)	-	-	0.7(2)	4.3(13)

<표 18> 초성 'ㄲ'의 후행 모음별 인식률

CV음절 초성 'ㄲ'의 인식률 평균은 91.7%, CVC음절 평균은 86.6%로 앞서 살펴본 평파열음에 비해 인식률이 높은 편이다. CV음절에서는 후행 모음이 'ㅡ, ㅣ'일 때, CVC음절에서는 후행 모음이 'ㅏ, ㅐ, ㅣ'일 때 평균 이하의 인식률을 보인다. 초성 'ㄲ'를 주로 어떤 자음으로 오인식하는지 살펴본 결과, CV음절과 CVC음절에서 모두 같은 연구개음 계열의 평음 'ㄱ'로 오인식되는 비율이 비교적 높게 나타났으나, CVC음절에서는 같은 경음에 해당되는 'ㄸ'로의 오인식 예도 발견된다. 따라서 초성 'ㄲ'의 유사음소는 'ㄱ' 또는 'ㄸ'인 것으로 간주한다.

연구개음 계열의 자음은 여타의 자음들에 비해 조음 위치상의 오류는 적은 편이지만, 'ㄲ ⇒ ㄸ'뿐 아니라 'ㄱ', 'ㅋ'에서도 'ㄷ'나 'ㅌ'로의 조음 위치의 오인식 예들이 다소 관찰되기도 한다. 이러한 오류는 주파수 필터에 의해 특정 주파수가 차단될 때

나타나는 현상과 동일해 보이는데, 이와 관련하여 주파수 차단 필터를 통해 특정 주파수를 차단한 상태에서의 한국어 단음절 음성 지각 실험을 진행한 김연화(2015: 63)에서도, 초성 연구개 파열음들이 치조 파열음 계열로 오인식되는 현상이 나타났다는 점이 주목된다. 이 논의에서는 300Hz~5000Hz 사이의 차단 주파수를 각각 LPF(Low Pass Filter, 저역 통과 필터)와 HPF(High Pass Filter, 고역 통과 필터)를 통과시켜 만든 음성 자료를 바탕으로 6명의 피험자들을 통해 음성 지각 실험을 진행하였는데, 그 결과 연구개 파열음들은 HPF가 적용된 경우, 차단 주파수가 2500Hz 이상일 때 치조 파열음 계열로 오지각되었다. 연구개 파열음의 평균 주파수는 600~3300Hz 사이에 분포되어 있어[36] 파열음들 중 가장 낮은 주파수 대역이 포함되는데, 여기에 차단 주파수 2500Hz 이상의 고역 통과 필터가 적용된다면, 연구개음이 가진 주파수 정보가 대부분 소실될 가능성이 높다. 이는 평균 주파수가 1200~2500Hz 사이에 분포한 양순음의 경우도 마찬가지이다. 이로 인해 연구개음은 2500Hz 이상의 고역 필터 적용 시, 양순음은 3000Hz까지의 고역 필터 적용 시 모두 주파수 대역이 가장 높은 'ㄷ'계열로 오인식되는 모습을 보인다. 이와 같은 김연화(2015)의 실험 결과를 토대로 볼 때, 연구개음의 치조음으로의 오인식은 음성 인식 상황에서 연구개 파열음의 저역 주파수가 정확히 인식되지 않은 데에서 기인하였을 가능성도 있을 듯하다.

다음으로는 'ㄸ'의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㄸ⇒ㄸ	70.0(35)	62.0(31)	76.0(38)	100.0(50)	94.0(47)	90.0(45)	86.0(43)
	ㄸ⇒ㄴ	14.0(7)	24.0(12)	16.0(8)	-	-	-	-
	ㄸ⇒ㅃ	14.0(7)	-	-	-	-	-	-
CVC	ㄸ⇒ㄸ	81.0(243)	73.3(220)	88.7(266)	95.0(285)	86.7(260)	96.0(288)	84.7(254)
	ㄸ⇒ㅃ	7.0(21)	10.7(32)	4.3(13)	-	1.0(3)	0.3(1)	4.3(13)

<표 19> 초성 'ㄸ'의 후행 모음별 인식률

36) 이주현(2005: 44)에서 참고함.

CV음절 초성 'ㄸ'의 인식률 평균은 82.6%, CVC음절 평균은 86.5%이며, CV음절에서는 후행 모음이 'ㅏ, ㅐ, ㅓ'일 때, CVC음절에서는 후행 모음이 'ㅏ, ㅐ, ㅣ'일 때 평균 이하의 인식률을 보인다. 초성 'ㄸ'를 주로 어떤 자음으로 오인식하는지 살펴본 결과, CV음절에서는 주로 비음 'ㄴ'로, CVC음절에서는 같은 경음 계열의 'ㅃ'로 오인식되는 경향이 나타났으며, 이들 각각이 음절 구조에 따른 'ㄸ'의 유사 음소인 것으로 판단된다. CVC음절에서 'ㄸ'가 'ㅃ'로 오인식되는 경우는 모두 종성이 'ㄱ, ㅇ'와 같은 연구개음 계열이거나 양순음 'ㅂ'일 때이다.

이때 CV음절 초성 'ㄸ'의 유사 음소가 비음 'ㄴ'로 나타나는 것은 다소 의외의 결과이다. 경음 'ㄸ'는 파열이 시작되는 시점에 스펙트로그램상 수직 스파이크 형태의 에너지가 관찰되지만, 국어의 어두 비음은 신지영(2014: 247)에서 지적된 바와 같이 스펙트로그램상의 음향 에너지가 약하고 어중이나 어말 비음에 비해 그 길이도 짧은바, 'ㄸ'와 'ㄴ' 간의 오지각이 음향적 유사성에서 기인한 것으로 단정하기는 어려워 보인다. 그러나 후술할 초성 공명음의 인식 오류에서도 'ㄴ'가 'ㄸ'로 오인식되는 양상이 나타나고, 김연화(2015)의 사람을 대상으로 한 단음절 음성 지각 실험 결과에서도 아홉 종류의 초성 파열음 중 'ㄸ'만이 'ㄴ'로 오인식되었다는 점을 토대로 본다면,[37] 'ㄸ'의 'ㄴ'로의 오인식은 이들 간의 음향적 유사성에 의한 인식 오류일 가능성이 있을 듯하다. 파열음이 비음으로 오인식되는 예는 'ㄸ'와 같은 경음에 한하여 발견되는데, 경음은 평음·격음과 달리 기식이 없다는 점에서 비음과 유사성을 가지며, 이러한 유사성이 'ㄸ, ㄴ' 상호간의 오인식에 영향을 주었을 수 있을 것이다. 이와 더불어 'ㄴ'의 자유 변이음 중 하나인 [n^d]의 실현과 관련되었을 가능성도 있으나, 이는 추후 더 정확한 음향 분석을 통해 확인되어야 할 것이다.[38][39]

---

37) 이 논의에서는 차단 주파수가 2000Hz인 고역 통과 필터가 적용됐을 때 'ㄸ'가 'ㄴ'로 오인식되었다.

38) 이와 관련하여 이진호(2013)에서는 평파열음과 비음 간의 대응 현상에 대해 면밀하게 고찰하며, 국어와 외국어의 상호 수용 과정에서 발견되는 평파열음과 비음 간의 음성적 유사성에 대해 논의하였는데, 이와 같이 주로 평음과 비음 간에 나타나는 인식상의 대응이 본 연구에서는 경음과 비음 사이에서 나타나고 있다는 점이 다소 특이한 양상으로 보인다.

39) 추가 실험으로 '따, 때, 떠'를 수차례 발화하여 초성 자음이 'ㄴ'로 인식될 때의 발화음을 녹음한 뒤 음향 분석을 시도해 본 결과, 위에서 언급한 요소들 외에 '음절의 길이'가 'ㄸ'를 'ㄴ'

다음으로는 ‘ㅃ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅃ⇒ㅃ	86.0(43)	82.0(41)	40.0(20)	76.0(38)	74.0(37)	54.0(27)	88.0(44)
	ㅃ⇒ㄲ	2.0(1)	-	48.0(24)	-	-	2.0(1)	-
	ㅃ⇒ㄴ	-	14.0(7)	-	-	-	-	-
	ㅃ⇒ㄸ	4.0(2)	-	-	-	16.0(8)	16.0(8)	2.0(1)
	ㅃ⇒∅	8.0(4)	-	10.0(5)	16.0(8)	10.0(5)	20.0(10)	10.0(5)
CVC	ㅃ⇒ㅃ	70.7(212)	67.3(202)	79.7(239)	83.0(249)	90.0(270)	50.3(151)	82.3(247)
	ㅃ⇒ㅂ	13.0(39)	3.0(9)	5.3(16)	2.7(8)	0.7(2)	-	6.3(19)
	ㅃ⇒ㄲ	0.7(2)	-	6.7(20)	4.3(13)	4.0(12)	13.0(39)	-
	ㅃ⇒ㄸ	8.3(25)	10.7(32)	3.0(9)	0.3(1)	3.0(9)	9.7(29)	1.3(4)
	ㅃ⇒∅	3.7(11)	14.7(44)	4.0(12)	4.7(14)	1.7(5)	25.0(75)	10.0(30)

<표 20> 초성 ‘ㅃ’의 후행 모음별 인식률

CV음절 초성 ‘ㅃ’의 인식률 평균은 71.4%, CVC음절 평균은 74.8%이며, CV음절에서는 후행 모음이 ‘ㅓ, ㅡ’일 때, CVC음절에서는 후행 모음이 ‘ㅏ, ㅐ, ㅡ’일 때 평균 이하의 인식률을 보인다. 초성 ‘ㅃ’의 오인식 양상을 세부적으로 살펴본 결과, CV음절과 CVC음절에서 모두 초성 자음이 인식되지 않는 오류가 가장 빈번하게 발생하였다. 그 외에는 같은 양순음 계열의 평음 ‘ㅂ’나, 경음 계열의 ‘ㄲ, ㄸ’ 등으로 오인식되는 경향이 산발적으로 나타난다.

로 오인식하게 하는 유의미한 변수로 확인되었다. 음성 인식 상황에서 ‘ㄸ’를 정확히 발음하면서 ‘ㄴ’로 인식되도록 유도하기가 쉽지는 않았는데, 목소리를 크게 하거나 음절의 길이를 길게 하거나, 음높이를 높게 하거나 낮게 하는 등의 여러 시도들 중, 음절의 길이를 길게 하였을 때 ‘ㄸ’로 발화한 음이 ‘ㄴ’로 인식되기 시작하였다. 다만 본 연구자 한 명의 발화를 대상으로 진행한 이 실험 결과를 일반화할 수는 없으므로 ‘ㄸ ⇒ ㄴ’의 오인식에 음절 길이가 관련되어 있을 가능성 정도만 언급해 두고자 한다.

3.1.1.3. 격음

본 절에서는 파열음에 해당되는 초성 자음 중 격음 ‘ㅋ, ㅌ, ㅍ’의 인식률과 오류 환경에 따른 유사 음소에 대해 알아보도록 한다. 가장 먼저 제시할 것은 ‘ㅋ, ㅌ, ㅍ’의 전체 인식률이다.

음절구조	후행모음 \ 초성	ㅋ	ㅌ	ㅍ	평균
CV	ㅏ	90.0(45)	78.0(39)	64.0(32)	77.3
	ㅐ	72.0(36)	82.0(41)	68.0(34)	74.0
	ㅓ	86.0(43)	64.0(32)	96.0(48)	82.0
	ㅗ	96.0(48)	94.0(47)	86.0(43)	92.0
	ㅜ	96.0(48)	96.0(48)	82.0(41)	91.3
	ㅡ	94.0(47)	90.0(45)	40.0(20)	74.7
	ㅣ	100.0(50)	70.0(35)	96.0(48)	88.7
인식률 총 평균		90.6	82.0	76.0	
		82.9%			
CVC	ㅏ	75.0(225)	76.0(228)	82.3(247)	77.8
	ㅐ	94.0(282)	87.7(263)	81.0(243)	87.6
	ㅓ	90.7(272)	84.0(252)	87.7(263)	87.5
	ㅗ	94.0(282)	90.3(271)	78.7(236)	87.7
	ㅜ	97.3(292)	86.7(260)	93.7(281)	92.6
	ㅡ	97.0(291)	84.7(254)	44.7(134)	75.5
	ㅣ	90.0(270)	64.7(194)	92.0(276)	82.2
인식률 총 평균		91.1	82.0	80.0	
		84.4%			

<표 21> 초성 파열음 중 격음의 인식률

<표 21>을 통해 확인할 수 있는 CV음절 초성 격음의 인식률 평균은 82.9%이며, CVC음절의 인식률 평균은 84.4%, 이들을 합한 초성 격음 전체의 인식률은 83.7%로 확인된다. 각 자음별 인식률은 평음·경음과 마찬가지로 ㅋ > ㅌ > ㅍ 순으로 낮아진다. 다음으로는 각 음소들의 개별적 인식 오류 양상과 유사 음소에 대해 구체적으로 살펴보도록 한다. 먼저 ‘ㅋ’의 경우이다.

음절 구조	인식 자음	후행 모음						
		ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅋ⇒ㅋ	90.0(45)	72.0(36)	86.0(43)	96.0(48)	96.0(48)	94.0(47)	100.0(50)
	ㅋ⇒ㄱ	8.0(4)	24.0(12)	4.0(2)	-	-	6.0(3)	-
CVC	ㅋ⇒ㅋ	75.0(225)	94.0(282)	90.7(272)	94.0(282)	97.3(292)	97.0(291)	90.0(270)
	ㅋ⇒ㄱ	8.3(25)	0.7(2)	1.3(4)	-	1.0(3)	1.0(3)	3.7(11)
	ㅋ⇒ㅌ	6.3(19)	1.3(4)	1.0(3)	-	0.3(1)	-	2.0(6)
	ㅋ⇒ㅎ	2.3(7)	1.3(4)	4.0(12)	5.0(15)	0.7(2)	2.0(6)	0.7(2)

<표 22> 초성 /ㅋ/의 후행 모음별 인식률

CV음절 초성 'ㅋ'의 인식률 평균은 90.6%, CVC음절 평균은 91.1%로 앞서 살펴본 평음과 경음에 비해 인식률이 높은 편이다. CV음절에서는 후행 모음이 'ㅏ, ㅓ, ㅐ'일 때, CVC음절에서는 후행 모음이 'ㅏ, ㅓ, ㅣ'일 때[40] 평균 이하의 인식률을 보인다. 초성 'ㅋ'를 주로 어떤 자음으로 오인식하는지 살펴본 결과, CV음절과 CVC음절에서 모두 같은 연구개음 계열의 평음 'ㄱ'로 오인식되는 비율이 비교적 높게 나타났으나, CVC음절에서는 같은 격음에 해당되는 'ㅌ'나 마찰음 'ㅎ'로의 오인식 예도 발견되었다. 'ㅋ'는 인식률 자체가 높아 오류 발생 빈도가 적어 유사 음소를 따로 설정하는 것이 다소 무리스럽기는 하나, 만약 'ㅋ'에서 인식 오류가 발생한다면 이때의 유사 음소는 'ㄱ' 또는 'ㅌ, ㅎ'일 가능성이 높다.

한편 'ㅋ'는 후행 모음이 'ㅓ'나 'ㅗ'일 때 'ㅎ'로 오인식되는 모습이 발견되는데, 이는 'ㅋ'뿐만 아니라 전술한 바와 같이 연구개음 계열의 평음 'ㄱ'에서도 나타나고, 양순음 계열의 자음 'ㅂ, ㅍ'와 'ㅓ, ㅗ'가 연쇄되는 환경에서도 관찰된다. 음성 인식 상황에서 연구개음은 양순음과 유사한 특성을 보이기도 하는데, 연구개음 계열과 양순음 계열 자음들이 'ㅎ'로 오인식되는 이유에 대해서는 여러 가지 추론이 가능할 것이나 후행 모음의 특성과 관련지어 생각해 본다면, 초·중·종성이 모두 [+grave]인

40) 이 중 후행 모음 'ㅓ, ㅣ'의 인식률은 각각 90.7%, 90.0%로 상대적으로 높은 편이기도 하고, 평균과도 근소한 차이를 보이므로 이를 'ㅋ'의 인식 오류를 유발하는 환경이라 하기는 어려울 듯하다. 'ㅋ'의 인식 오류가 주로 발생하는 환경은 후행 모음이 'ㅏ'일 때이다.

환경에서 음향 특색이 모두 겹쳐 기식만 유의미하게 남아 ‘ㅎ’로 오인식되었을 가능성을 우선 생각해 볼 수 있다. 더불어 이 현상이 기식음인 평음과 격음에서는 발견되나 무기음인 경음에서는 나타나지 않는다는 점에 주목해 본다면, 파열음의 조음 단계 중 ‘파열’ 과정이 정확히 인식되지 못하고 기식만 남아 있는 경우, 이 특성이 ‘ㅎ’에 해당되는 음향 특성으로 판단되어 발생한 오류로도 해석할 수 있다. 참고로 연구개음과 양순음 계열 자음들이 ‘ㅎ’로 오인식되는 양상을 종합하여 제시하면 아래의 <표 23>과 같다.

		후행모음 / 인식자음	/ㅏ/	/ㅐ/	/ㅓ/	/ㅗ/	/ㅜ/	/ㅡ/	/ㅣ/
연구개음	CV	ㄱ⇒ㅎ	-	-	-	-	-	-	-
		ㄲ⇒ㅎ	-	-	-	-	-	-	-
		ㅋ⇒ㅎ	-	-	-	-	-	-	-
	CVC	ㄱ⇒ㅎ	1.7(5)	0.7(2)	**3.3(10)**	**6.7(20)**	1.0(3)	0.3(1)	0.7(2)
		ㄲ⇒ㅎ	-	-	-	-	-	-	-
		ㅋ⇒ㅎ	2.3(7)	1.3(4)	**4.0(12)**	**5.0(15)**	0.7(2)	2.0(6)	0.7(2)
양순음	CV	ㅂ⇒ㅎ	6.0(3)	14.0(7)	2.0(1)	2.0(1)	6.0(3)	4.0(2)	-
		ㅃ⇒ㅎ	-	-	-	-	-	-	-
		ㅍ⇒ㅎ	-	**22.0(11)**	-	6.0(3)	4.0(2)	-	-
	CVC	ㅂ⇒ㅎ	1.0(3)	2.7(8)	3.0(9)	**8.3(25)**	1.7(5)	3.3(10)	0.7(2)
		ㅃ⇒ㅎ	-	-	-	-	-	-	-
		ㅍ⇒ㅎ	4.3(13)	**8.3(25)**	**5.7(17)**	**9.7(29)**	1.3(4)	**10.3(31)**	1.7(5)

<표 23> 연구개음·양순음 계열의 오인식 환경

한편 ‘ㅋ’를 ‘ㄱ’로 오인식하는 것과 같이 격음을 평음으로 인식하는 오류는 음성 인식 상황에서 자주 발견되는 오류 유형은 아니다. 그런데 이러한 오류가 대부분 남성의 발화에서 나타난다는 점은 다소 흥미로운 점이다. 아래의 <표 24>를 통해 격음을 평음으로 오인식한 전체 빈도와, 그 중 남성과 여성의 오류 빈도를 따로 정리하여 살펴보도록 한다.

		오인식 빈도		
		전체	남성	여성
격음⇒평음	ㅋ⇒ㄱ	48	45	3
	ㅌ⇒ㄷ	42	40	2
	ㅍ⇒ㅂ	35	33	2
평음⇒격음	ㄱ⇒ㅋ	840	376	464
	ㄷ⇒ㅌ	727	319	408
	ㅂ⇒ㅍ	649	284	365

<표 24> 성별에 따른 평음·격음 간 오인식 빈도

<표 24>를 통해 확인할 수 있는 것은, 격음을 평음으로 인식하는 오류는 발생 빈도는 적으나 오류의 대부분이 남성의 발화에서 나타난다는 점과, 평음을 격음으로 인식하는 오류는 성별에 따른 오류 빈도 차가 크지는 않지만 여성에게서 상대적으로 높게 나타난다는 점이다. 장혜진(2011)에 의하면 평음과 격음의 지각 변별에는 후행 모음의 F0값이[41] 영향을 미쳐, F0값이 높으면 격음으로, 낮으면 평음으로 인식되는 경향이 관찰된다고 한다.[42] 그리고 주지하다시피 여성의 F0값은 남성에 비해 높게 나타난다.[43] 이를 종합적으로 고려한다면 <표 24>와 같은 인식 결과는, 음성 인식 알고리즘이 F0값을 기준으로 하여 상대적으로 낮은 F0값을 가지는 남성의 격음은 평음으로 판단하고, F0값이 상대적으로 높은 여성의 평음은 격음으로 판단하는 오

41) F0는 모음의 스펙트럼에서 첫 번째 구성 성분(첫 번째 '배음(harmonic')인 '기본 주파수(fundamental frequency)'를 의미하며, 지각적으로는 '음높이(pitch)'를 나타낸다(신지영 2014: 183).

42) 이경희·정명숙(2000)의 실제 음성 지각 실험 결과에 따르면, /다/의 모음 'ㅏ'를 후행 모음으로 둔 음절은 '다'로, /타/의 모음 'ㅏ'를 후행 모음으로 둔 음절은 '타'로, /따/의 모음 'ㅏ'를 후행 모음으로 둔 음절은 '따'로 지각되었다고 한다. 이때 각각의 후행 모음의 음높이는 202Hz, 259Hz, 224Hz로 뚜렷한 차이를 보인바, 이 결과를 바탕으로 이경희·정명숙(2000: 148)에서는 후행 모음의 음높이가 어두 위치 파열음의 지각 단서로 고려될 수 있음을 주장하였다. 이에 의하면 후행 모음의 음높이가 높을 때는 선행하는 파열음이 격음으로, 음높이가 낮을 때는 평음으로 지각된다.

43) 양병곤(2003: 65)에 따르면 피치(기본 주파수 F0의 역수)는 성대의 크기와 길이, 질량 등에 영향을 받는 요소이며, 성대의 크기가 클수록 진동하는 속도가 느리기 때문에 여성보다 상대적으로 큰 성대를 가진 남성의 피치 값은 약 120Hz, 여성은 약 230Hz, 어린아이는 300Hz 전후로 나타난다고 한다.

류를 보인 것으로 추정해 볼 수 있다.

다음으로는 'ㅌ'의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅌ⇒ㅌ	78.0(39)	82.0(41)	64.0(32)	94.0(47)	96.0(48)	90.0(45)	70.0(35)
	ㅌ⇒ㄷ	12.0(6)	8.0(4)	18.0(9)	-	-	-	-
	ㅌ⇒ㅍ			10.0(5)				8.0(4)
	ㅌ⇒ㅋ	2.0(1)		2.0(1)			2.0(1)	18.0(9)
CVC	ㅌ⇒ㅌ	76.0(228)	87.7(263)	84.0(252)	90.3(271)	86.7(260)	84.7(254)	64.7(194)
	ㅌ⇒ㄷ	8.0(24)	0.7(2)	1.7(5)	1.0(3)	-	1.3(4)	-
	ㅌ⇒ㅍ	11.0(33)	3.7(11)	10.0(30)	3.7(11)	6.7(20)	4.7(14)	1.3(4)
	ㅌ⇒ㅋ	1.0(3)	2.0(6)	1.7(5)	1.3(4)	-	1.3(4)	24.7(74)

<표 25> 초성 'ㅌ'의 후행 모음별 인식률

초성 'ㅌ'의 인식률 평균은 CV음절과 CVC음절 모두 82.0%이다. CV음절에서는 후행 모음이 'ㅏ, ㅣ, ㅐ, ㅓ'일 때, CVC음절에서는 후행 모음이 'ㅏ, ㅣ'일 때 평균 이하의 인식률을 보인다. 초성 'ㅌ'를 주로 어떤 자음으로 오인식하는지 살펴본 결과, CV음절과 CVC음절에서 모두 같은 격음에 해당되는 'ㅍ, ㅋ'로 오인식되는 경우와, 치조음 계열의 평음 'ㄷ'로 오인식되는 경우가 관찰되었다. 후행 모음이 'ㅣ'일 때는 'ㅌ'를 'ㅋ'로 오인식하는 모습이 나타난다. 경구개 모음 'ㅣ' 앞에서 'ㄷ, ㅌ'는 곧잘 경구개음화된 변이음으로 실현되는데, 음성 인식 상황에서는 이를 'ㅈ, ㅊ'보다는 'ㄱ, ㅋ'와 유사한 것으로 판단하는 것으로 보인다. 이러한 오류는 평음('ㄷ')에서보다 격음('ㅌ')에서 빈번하게 관찰되는데, 이는 이호영·지민제·김영송(1993: 133)에 기술된 바와 같이 파열음의 구개음화된 변이음의 음향적 특성이 격음에서 잘 나타나는 데에서 기인된 것인 듯하다.

'ㅌ'는 오인식 자음들이 산발적으로 분포되어 있어 유사 음소를 설정하는 것이 다소 무리스럽다. 다만 'ㅌ'가 같은 치조음 계열의 평음 'ㄷ'로 오인식되는 경우는 소수이고, 'ㅋ'로 오인식되는 것은 주로 'ㅣ' 앞에서만 나타나는 현상이라는 점을 토대로

한다면, ‘ㅌ’의 유사 음소는 ‘ㅍ’로 설정하는 것이 적절할 듯 보인다.

마지막으로 ‘ㅍ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅍ⇒ㅍ	64.0(32)	68.0(34)	96.0(48)	86.0(43)	82.0(41)	40.0(20)	96.0(48)
	ㅍ⇒ㅋ	2.0(1)	-	-	-	-	40.0(20)	-
	ㅍ⇒ㅌ	20.0(10)	6.0(3)	-	6.0(3)	4.0(2)	20.0(10)	-
	ㅍ⇒ㅎ	-	22.0(11)	-	6.0(3)	4.0(2)	-	-
CVC	ㅍ⇒ㅍ	82.3(247)	81.0(243)	87.7(263)	78.7(236)	93.7(281)	44.7(134)	92.0(276)
	ㅍ⇒ㅋ	-	-	1.0(3)	6.3(19)	2.0(6)	25.0(75)	2.3(7)
	ㅍ⇒ㅌ	8.3(25)	6.7(20)	4.0(12)	3.3(10)	0.7(2)	17.3(52)	0.7(2)
	ㅍ⇒ㅎ	4.3(13)	8.3(25)	5.7(17)	9.7(29)	1.3(4)	10.3(31)	1.7(5)

<표 26> 초성 ‘ㅍ’의 후행 모음별 인식률

CV음절 초성 ‘ㅍ’의 인식률 평균은 76.0%, CVC음절은 80.0%이다. CV음절에서는 후행 모음이 ‘ㅏ, ㅐ, ㅡ’일 때, CVC음절에서는 후행 모음이 ‘ㅗ, ㅡ’일 때 평균 이하의 인식률을 보인다. 초성 ‘ㅍ’를 주로 어떤 자음으로 오인식하는지 살펴본 결과, 주로 같은 격음에 해당되는 ‘ㅌ, ㅋ’나 마찰음 ‘ㅎ’로 오인식되는 모습이 관찰되었다. 이는 격음임을 판단할 수 있는 기식 특성은 감지되었으나, 어떠한 격음인지 분명치 않아 같은 격음 계열의 음소 중 하나로 오인식되거나, 같은 기식 특성을 가지는 마찰음으로 오인식된 결과로 보인다.

지금까지 살펴본 초성 파열음들의 음성 인식 오류 양상을 종합하면 다음과 같다. 먼저 평음의 유사 음소는 후행 모음 환경에 따라 차이는 있으나 대개 격음으로 나타난다. 경음은 평음 또는 같은 경음 계열의 유사 음소를 가지며, 격음 역시 평음 또는 같은 격음 계열의 유사 음소를 가지는 것으로 확인된다. 평음과 격음 간의 오인식은 평음과 격음이 가진 음향적 유사성과 더불어, 성별에 따른 파열음의 발화 산출 차이가 영향을 주었을 것으로 판단된다.

3.1.2. 파찰음

본 절에서는 파찰음에 해당되는 초성 자음들의 인식률과 오류 환경에 따른 유사 음소에 대해 알아보도록 한다. 가장 먼저 제시할 것은 'ㅈ, ㅉ, ㅊ'의 초성 인식률이다.

음절구조	후행모음 \ 초성	ㅈ	ㅉ	ㅊ	평균
CV	ㅏ	86.0(43)	98.0(49)	98.0(49)	94.0
	ㅐ	82.0(41)	86.0(43)	98.0(49)	88.7
	ㅓ	90.0(45)	72.0(36)	94.0(47)	85.3
	ㅗ	74.0(37)	80.0(40)	100.0(50)	84.7
	ㅜ	54.0(27)	88.0(44)	100.0(50)	80.7
	ㅡ	74.0(37)	82.0(41)	80.0(40)	78.7
	ㅣ	68.0(34)	76.0(38)	90.0(45)	78.0
인식률 총 평균		75.4	83.1	94.3	
		84.3%			
CVC	ㅏ	62.3(187)	98.3(295)	100.0(300)	86.9
	ㅐ	69.7(209)	92.0(276)	95.0(285)	85.6
	ㅓ	60.7(182)	97.0(291)	82.7(248)	80.1
	ㅗ	62.0(186)	95.3(286)	88.5(265)	81.9
	ㅜ	43.3(130)	91.7(275)	100.0(300)	78.3
	ㅡ	56.3(169)	92.3(277)	88.5(265)	79.0
	ㅣ	58.7(176)	79.3(238)	82.7(248)	73.6
인식률 총 평균		59.0	92.3	91.1	
		80.8%			

<표 27> 초성 파찰음의 인식률

<표 26>에 제시된 CV음절 초성 파찰음의 인식률 평균은 84.3%이며, CVC음절의 인식률 평균은 80.8%, 이들을 합한 초성 파찰음 전체의 인식률은 83.1%로 확인된다. 경음, 격음에 비해 평음의 인식률이 낮게 나타나며, 이는 파열음의 경우와 동일하다. 다음으로는 각 음소들의 개별적 인식 오류 양상과 유사 음소에 대해 구체적으로 살

펴보도록 한다. 먼저 'ㅈ'의 경우이다.

음절 구조	인식 자음	후행 모음						
		ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅈ⇒ㅈ	86.0(43)	82.0(41)	90.0(45)	74.0(37)	54.0(27)	74.0(37)	68.0(34)
	ㅈ⇒ㅊ	6.0(3)	8.0(4)	8.0(4)	24.0(12)	18.0(9)	22.0(11)	18.0(9)
	ㅈ⇒ㅆ	-	-	-	-	-	-	10.0(5)
	ㅈ⇒ㅉ	-	-	-	2.0(1)	16.0(8)	2.0(1)	-
CVC	ㅈ⇒ㅈ	62.3(187)	69.7(209)	60.7(182)	62.0(186)	43.3(130)	56.3(169)	58.7(176)
	ㅈ⇒ㅊ	33.3(100)	20.7(62)	34.0(102)	19.7(59)	35.0(105)	26.0(78)	19.7(59)
	ㅈ⇒ㅉ	0.7(2)	2.0(6)	-	14.3(43)	16.3(49)	12.7(38)	13.0(39)

<표 28> 초성 'ㅈ'의 후행 모음별 인식률

CV음절 초성 'ㅈ'의 인식률 평균은 75.4%, CVC음절 평균은 59.0%이며, CV음절에서는 후행 모음이 'ㅗ, ㅜ, ㅡ, ㅣ'일 때, CVC음절에서는 후행 모음이 'ㅜ, ㅡ, ㅣ'일 때 평균 이하의 인식률을 보인다. 초성 'ㅈ'를 주로 어떤 자음으로 오인식하는지 살펴본 결과, CV음절과 CVC음절에서 모두 같은 파찰음 계열의 'ㅊ'가 유사 음소로 나타났다. CVC음절에서는 'ㅉ'로 오인식되는 경우도 확인된다.

다음으로는 'ㅉ'의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

음절 구조	인식 자음	후행 모음						
		ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅉ⇒ㅉ	98.0(49)	86.0(43)	72.0(36)	80.0(40)	88.0(44)	82.0(41)	76.0(38)
	ㅉ⇒ㅈ	2.0(1)	-	16.0(8)	16.0(8)	6.0(3)	8.0(4)	16.0(8)
CVC	ㅉ⇒ㅉ	98.3(295)	92.0(276)	97.0(291)	95.3(286)	91.7(275)	92.3(277)	79.3(238)
	ㅉ⇒ㅈ	0.3(1)	4.7(14)	2.0(6)	4.7(14)	7.0(21)	6.3(19)	19.7(59)

<표 29> 초성 'ㅉ'의 후행 모음별 인식률

CV음절 초성 'ㅉ'의 인식률 평균은 83.1%, CVC음절 평균은 80.0%이며, 유사 음소

는 ‘ㅈ’로 확인된다. 음절 구조에 따라 인식률이 낮은 모음 환경은 다르나, 모든 경우에 오인식되는 자음은 ㅈ이다.

다음으로는 ‘ㅊ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅊ⇒ㅊ	98.0(49)	98.0(49)	94.0(47)	100.0(50)	100.0(50)	80.0(40)	90.0(45)
	ㅊ⇒ㅅ	2.0(1)	2.0(1)	6.0(3)	-	-	20.0(10)	-
	ㅊ⇒ㅋ							10.0(5)
CVC	ㅊ⇒ㅊ	100.0(300)	95.0(285)	82.7(248)	88.5(265)	100.0(300)	88.5(265)	82.7(248)
	ㅊ⇒ㅅ	-	5.0(15)	2.9(9)	11.4(34)	-	11.4(34)	17.3(52)
	ㅊ⇒ㅈ	-	-	14.3(43)	-	-	-	-

<표 30> 초성 ‘ㅊ’의 후행 모음별 인식률

CV음절 초성 ‘ㅊ’의 인식률 평균은 94.3%, CVC음절 평균은 91.1%이며, CV음절에서는 후행 모음이 ‘ㅓ, ㅡ, ㅣ’일 때, CVC음절에서는 ‘ㅏ, ㅓ, ㅡ, ㅣ’일 때의 인식률이 평균보다 낮다. 초성 ‘ㅊ’를 주로 어떤 자음으로 오인식하는지 살펴본 결과, CV음절과 CVC음절에서 모두 마찰음 ‘ㅅ’로 오인식되는 경우가 가장 빈번하게 관찰되었다. 이러한 오류는 ‘ㅊ’의 폐쇄 과정이 정확히 인식되지 못하고, ‘ㅊ’가 가진 ‘기식’ 특성이 ‘마찰’과 혼동되어 발생한 결과로 보인다. 그 외에 CV음절에서는 같은 격음 계열인 ‘ㅋ’로 오인식되는 경우가, CVC음절에서는 파찰음에 해당되는 ‘ㅈ’로 오인식되는 경우가 발견된다. 유사 음소는 ‘ㅅ’로 판단된다.

이상으로 살펴본 파찰음의 인식 오류에서는 주로 파찰음이라는 범주 안에서 음소들이 서로 오인식되는 경향이 관찰되었다. 단 해당 음소가 파찰음이라는 것은 분명히 인식이 되나, 파찰음의 조음 과정 중 폐쇄 특성이 정확히 인식되지 못하고, 마찰 특성이 강하게 나타나는 경우에는 이를 ‘ㅅ’나 ‘ㅆ’로 인식하기도 한다. 파찰음은 CV음절과 CVC음절 모두 공통적으로 후행 모음이 ‘ㅡ, ㅣ’일 때 평균 이하의 인식률을 나타내는데, 이는 이들 환경에서 전형적인 ‘ㅈ, ㅉ, ㅊ’의 음이 아닌 후행 모음에 영향을 받은 변이음으로 실현되기 때문으로 보인다.

3.1.3. 마찰음

본 절에서는 마찰음에 해당되는 초성 자음들의 인식률과 오류 환경에 따른 유사 음소에 대해 알아보도록 한다. 가장 먼저 제시할 것은 'ㅅ, ㅆ, ㅎ'의 초성 인식률이다.

음절구조	후행모음 \ 초성	ㅅ	ㅆ	ㅎ	평균
CV	ㅏ	94.0(47)	98.0(49)	86.0(43)	92.7
	ㅐ	98.0(49)	100.0(50)	98.0(49)	98.7
	ㅓ	100.0(50)	94.0(47)	86.0(43)	93.3
	ㅗ	100.0(50)	98.0(49)	88.0(44)	95.3
	ㅜ	90.0(45)	80.0(40)	90.0(45)	86.7
	ㅡ	90.0(45)	82.0(41)	90.0(45)	87.3
	ㅣ	64.0(32)	88.0(44)	86.0(43)	79.3
인식률 총 평균		90.9	91.4	89.1	
		90.5%			
CVC	ㅏ	92.7(278)	92.0(276)	93.3(280)	92.7
	ㅐ	95.0(285)	53.3(160)	92.7(278)	80.3
	ㅓ	100.0(300)	98.0(294)	93.3(280)	97.1
	ㅗ	95.0(285)	94.3(283)	90.0(270)	93.1
	ㅜ	91.3(274)	76.3(229)	93.3(280)	87.0
	ㅡ	78.0(234)	91.7(275)	96.3(289)	88.7
	ㅣ	91.7(275)	43.3(130)	90.0(270)	75.0
인식률 총 평균		92.0	78.4	92.7	
		87.7%			

<표 31> 초성 마찰음의 인식률

<표 31>을 통해 확인할 수 있는 CV음절 초성 마찰음의 인식률 평균은 90.5%이며, CVC음절의 인식률 평균은 87.7%, 이들을 합한 초성 격음 전체의 인식률은 89.1%이다. 마찰음이 가진 고주파대역의 마찰 소음이 음성 인식에 있어서는 하나의 유표적 인식 단서가 되어, 여타의 자음들에 비해 높은 인식률을 보이는 것으로 여겨진다.

<표 30>의 마찰음들은 후행 모음이 'ㅣ'일 때 공통적으로 인식률이 낮아진다. 다음으로는 각 음소들의 개별적 인식 오류 양상과 유사 음소에 대해 구체적으로 살펴보도록 한다. 먼저 'ㅅ'의 경우이다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅅ⇒ㅅ	94.0(47)	98.0(49)	100.0(50)	100.0(50)	90.0(45)	90.0(45)	64.0(32)
	ㅅ⇒ㅆ	-	-	-	-	-	4.0(2)	36.0(18)
CVC	ㅅ⇒ㅅ	92.7(278)	95.0(285)	100.0(300)	95.0(285)	91.3(274)	78.0(234)	91.7(275)
	ㅅ⇒ㅆ	-	2.0(6)	-	-	6.3(19)	19.7(59)	2.0(6)

<표 32> 초성 'ㅅ'의 후행 모음별 인식률

CV음절 초성 'ㅅ'의 인식률 평균은 90.9%, CVC음절 평균은 92.0%이며, 모든 초성 자음들 중 인식률이 가장 높게 나타난다. 인식률이 높은 만큼 오류 발생 빈도도 적으며, 모든 후행 모음 환경에서 'ㅆ'를 유사 음소로 가진다. 인식률이 가장 낮은 경우는 'ㅡ, ㅣ'가 후행할 때이다. 'ㅡ'와 'ㅣ'는 모음체계상 고모음이라는 점에서 공통점을 가지는데, 이경희·이봉원(1999)에 따르면 'ㅆ'는 'ㅅ'보다 마찰 구간이 길지만, 고모음 앞에서는 이러한 차이가 드러나지 않는 특성을 가진다고 한다. 즉 비고모음 앞에서는 'ㅅ'와 'ㅆ'를 마찰 구간의 길이를 바탕으로 변별할 수 있으나, 고모음 앞에서는 이러한 변별이 나타나지 않는다는 것이다. 이와 같은 특징에 주목해 본다면 'ㅅ'가 'ㅆ'로 오인식되는 오류가 고모음 'ㅡ, ㅣ' 앞에서 발생하는 것은, 고모음 앞에서 'ㅅ'와 'ㅆ'의 마찰 구간에 따른 변별이 어려워지며 나타난 오류라 추정해 볼 수 있다. 한국인 화자의 외래어 발음을 대상으로 음성 인식 결과를 분석한 류혁수·나민수·정민화(2015)에서 's'의 경음화, 즉 'ㅅ'를 'ㅆ'로 오인식하는 예가 'ㅡ'(59.44%), 'ㅣ'(12.11%) 앞에서 두드러지게 나타나는 것도 이와 동궤의 현상으로 보인다.

한편 이와 더불어 'ㅅ'는 경구개 모음 'ㅣ' 앞에서 경구개 마찰음을 변이음으로 가지는데, 음성 인식 상황에서 이 변이음을 전형적인 'ㅅ'의 음향 특질에서 벗어난 음으로 판단하여, 같은 마찰음 범주 안의 'ㅆ'로 오인식한 것으로 추정해 볼 수도 있

다. 이호영·지민제·김영송(1993: 132)에 따르면 ‘ㅅ, ㅆ’는 모음 ‘ㅣ’ 앞에서 조음점이 뒤로 이동하여 좁힘점과 입술과의 길이가 길어지기 때문에 다른 모음 앞에서보다 중심 공명 주파수대가 낮아져 스펙트로그램상 낮은 주파수 대역에도 에너지가 분포하여 중심 에너지의 주파수 대역이 약간 하강하는 특징을 가진다고 한다. ‘ㅣ’에 선행하는 ‘ㅅ’의 오인식은 이러한 비전형적인 음향 특징을 ‘ㅆ’의 범주로 판단한 데에서 기인하였을 가능성도 있을 듯하다.

다음으로는 ‘ㅆ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅆ⇒ㅆ	98.0(49)	100.0(50)	94.0(47)	98.0(49)	80.0(40)	82.0(41)	88.0(44)
	ㅆ⇒ㅅ	-	-	-	-	14.0(7)	18.0(9)	6.0(3)
CVC	ㅆ⇒ㅆ	92.0(276)	53.3(160)	98.0(294)	94.3(283)	76.3(229)	91.7(275)	43.3(130)
	ㅆ⇒ㅅ	2.0(6)	43.7(131)	0.7(2)	5.3(16)	21.7(65)	7.0(21)	54.0(162)

<표 33> 초성 ‘ㅆ’의 후행 모음별 인식률

CV음절 초성 ‘ㅆ’의 인식률 평균은 91.4%, CVC음절 평균은 78.4%이며, 모든 후행 모음 환경에서 ‘ㅅ’가 유사 음소로 확인된다. 후행 모음이 ‘ㅜ, ㅡ, ㅣ’와 같은 고모음일 때 평균보다 낮은 인식률을 보이며, 이는 전술한 바와 같이 고모음 앞에서 ‘ㅅ’와 ‘ㅆ’의 마찰 구간의 길이가 유사한 데에서 기인한 것으로 보인다. 이 외에 CVC음절의 ‘ㅐ’ 앞에서도 ‘ㅆ’가 ‘ㅅ’로 오인식되는 비율이 높게 나타나는데, 이는 실험 대상 음절 ‘쌕’을 주로 ‘섹스’로 오인식한 결과로, 음향적 유사성에 의한 오류라기보다는 언어 모델의 교정에 따른 것으로 판단된다.

다음으로는 ‘ㅎ’의 인식 오류 양상과 유사 음소에 대해 알아보도록 한다.

음절 구조	인식 자음	후행 모음						
		ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅎ⇒ㅎ	86.0(43)	98.0(49)	86.0(43)	88.0(44)	90.0(45)	90.0(45)	86.0(43)
	ㅎ⇒∅	-	-	10.0(5)	8.0(4)	4.0(2)	-	2.0(1)
CVC	ㅎ⇒ㅎ	93.3(280)	92.7(278)	93.3(280)	90.0(270)	93.3(280)	96.3(289)	90.0(270)
	ㅎ⇒∅	1.3(4)	6.7(20)	1.7(5)	1.0(3)	-	1.0(3)	0.3(1)
	ㅎ⇒ㅅ	2.7(8)	-	1.0(3)	-	-	1.0(3)	4.3(13)

<표 34> 초성 'ㅎ'의 후행 모음별 인식률

CV음절 초성 'ㅎ'의 인식률 평균은 89.1%, CVC음절 평균은 92.7%로 다른 자음들에 비해 상대적으로 인식률이 높게 나타난다. CV음절의 경우 오류 발생 빈도가 적고, 'ㅎ'의 존재를 인식하지 못하는 오류가 가장 빈번하게 나타나, 유사 '음소'를 따로 설정하는 것은 어려울 듯하다. CVC음절에서도 'ㅎ'를 인식하지 못하는 오류가 가장 빈번히 발생하나, 'ㅅ'로 오인식되는 경우도 관찰된다. 'ㅎ'는 후행 모음이 'ㅣ'일 때 CV음절, CVC음절 모두 가장 낮은 인식률을 보이는데, 이는 이호영·지민제·김영송(1993: 132)의 기술과 같이 'ㅎ'가 여타의 모음 앞에서는 조음 시 이렇다 할 좁힘이 일어나지 않아 약한 에너지만이 분포하지만, 'ㅣ'가 뒤이어 올 때는 전설이 경구개 쪽으로 접근하며 강한 마찰을 초래하여 비교적 넓은 주파수대에 걸쳐 강한 에너지가 분포하는 특징을 가지기 때문인 것으로 판단된다. 즉 이는 'ㅣ' 앞의 'ㅎ'가 가진 비전형적인 음향 특징을 'ㅅ'의 범주로 판단한 데에서 기인한 오류로 보이며, 'ㅎ'의 유사 '음소'로는 'ㅅ'를 들 수 있을 듯하다.

3.1.4. 공명음

본 절에서는 음성 인식 오류의 확인 실험 결과를 통해 분석된 초성 공명음의 인식 오류 양상을 살피고, 후행 모음 환경별 유사 음소에 대해 알아보도록 한다. 먼저 제시할 것은 공명음 'ㄴ, ㅁ, ㄹ'의 전체 인식 결과이다.

음절구조	후행모음 \ 초성	ㄴ	ㅁ	ㄹ	평균
CV	ㅏ	100.0(50)	96.0(48)	88.0(44)	94.7
	ㅐ	100.0(50)	**56.0(28)**	**66.0(33)**	74.0
	ㅓ	98.0(49)	100.0(50)	**60.0(30)**	86.0
	ㅗ	96.0(48)	98.0(49)	90.0(45)	94.7
	ㅜ	98.0(49)	96.0(48)	84.0(42)	92.7
	ㅡ	**92.0(46)**	**74.0(37)**	84.0(42)	83.3
	ㅣ	**82.0(41)**	98.0(49)	92.0(46)	90.7
인식률 총 평균		95.1	88.3	80.6	
		88.0%			
CVC	ㅏ	88.7(266)	91.0(273)	92.0(276)	90.6
	ㅐ	**86.7(260)**	88.0(264)	92.7(278)	89.1
	ㅓ	88.3(265)	91.7(275)	87.3(262)	89.1
	ㅗ	87.3(262)	97.3(292)	94.0(282)	92.9
	ㅜ	**85.0(255)**	93.3(280)	87.0(261)	88.4
	ㅡ	91.0(273)	**67.3(202)**	**68.7(206)**	75.7
	ㅣ	**80.0(240)**	**85.0(255)**	**85.7(257)**	83.6
인식률 총 평균		86.7	87.7	86.8	
		87.0%			

<표 35> 초성 공명음의 인식률

<표 35>에 제시된 CV음절 초성 공명음의 인식률 평균은 88.0%이며, CVC음절의 인식률 평균은 87.0%, 이들을 합한 초성 공명음 전체의 인식률은 87.5%로 확인된다. 다음으로는 각 공명음들의 개별적 인식 오류 양상과 유사 음소에 대해 구체적으로 살펴보도록 한다. 먼저 'ㄴ'의 경우이다.

음절 구조	인식 자음	후행 모음						
		ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㄴ⇒ㄴ	100.0(50)	100.0(50)	98.0(49)	96.0(48)	98.0(49)	92.0(46)	82.0(41)
	ㄴ⇒ㅁ	-	-	-	-	-	2.0(1)	8.0(4)
CVC	ㄴ⇒ㄴ	88.7(266)	86.7(260)	88.3(265)	87.3(262)	85.0(255)	91.0(273)	80.0(240)
	ㄴ⇒ㄸ	3.0(9)	2.3(7)	3.7(11)	8.3(25)	9.3(28)	3.3(10)	1.3(4)
	ㄴ⇒ㅁ	3.7(11)	3.3(10)	6.0(18)	1.3(4)	1.0(3)	-	11.7(35)

<표 36> 초성 ‘ㄴ’의 후행 모음별 인식률

CV음절 초성 ‘ㄴ’의 인식률 평균은 95.1%, CVC음절 평균은 86.7%이고, CV음절은 전체 오류 발생 빈도가 전체 350회의 발화 중 17회에 그치며 매우 높은 인식률을 나타냈다. 오류가 적은 만큼 유사 음소를 설정하는 것은 불필요한 일일 듯하나, 이때 오인식된 자음은 대체로 ‘ㅁ’이나 ‘ㄹ’와 같은 공명음 계열이었다. CVC음절의 초성 ‘ㄴ’는 CV음절보다는 낮은 인식률을 보이며, 주로 후행 모음이 ‘ㅣ, ㅗ, ㅣ’일 때 평균 이하의 인식률이 나타난다.

경음 ‘ㄸ’를 ‘ㄴ’로 오인식했던 것과 마찬가지로 ‘ㄴ’도 ‘ㄸ’로 오인식되는 경우가 관찰된다. 앞서 ‘ㄸ’가 ‘ㄴ’로 오인식되었던 환경은 CV음절에서 모음 ‘ㅏ, ㅐ, ㅓ’가 후행할 때였으나, ‘ㄴ’의 ‘ㄸ’로의 오인식은 CVC음절에서만 나타나며, 빈도의 차이는 있으나 모든 모음들 앞에서 두루 발견된다. 여기에는 앞서 3.1.1.2.에서 언급한 바와 같은 음절의 길이, 경음과 비음 간의 무기음적 특성, ‘ㄴ’의 자유 변이음 중 하나인 [n^d]의 실현 등이 관련되어 있을 것으로 보인다. 그 외의 경우에는 ‘ㄴ’는 대개 같은 비음 계열인 ‘ㅁ’로 오인식되며, 따라서 초성 ‘ㄴ’의 유사 음소는 ‘ㅁ’와 ‘ㄸ’로 볼 수 있을 듯하다.

다음으로는 ‘ㅁ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

음절 구조	인식 자음	후행 모음						
		ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㅁ⇒ㅁ	96.0(48)	56.0(28)	100.0(50)	98.0(49)	96.0(48)	74.0(37)	98.0(49)
	ㅁ⇒ㄴ	4.0(2)	44.0(22)	-	-	-	14.0(7)	-
CVC	ㅁ⇒ㅁ	91.0(273)	88.0(264)	91.7(275)	97.3(292)	93.3(280)	67.3(202)	85.0(255)
	ㅁ⇒ㄴ	1.3(4)	7.7(23)	1.0(3)	-	2.7(8)	11.0(33)	12.0(36)
	ㅁ⇒∅	2.3(7)	1.0(3)	2.0(6)	0.3(1)	1.3(4)	19.7(59)	0.7(2)

<표 37> 초성 ‘ㅁ’의 후행 모음별 인식률

CV음절 초성 ‘ㅁ’의 인식률 평균은 88.3%, CVC음절 평균은 87.0%이며, 두 음절 구조 모두 후행 모음이 ‘ㅡ’일 때 공통적으로 인식률이 낮게 나타난다. 양순음과 ‘ㅡ’가 연쇄하는 경우, 자연 발화에서는 선행하는 양순음으로 인해 평순모음 ‘ㅡ’가 곧잘 원순모음 ‘ㅜ’로 발음된다.[44] 이로 인해 ‘ㅁ’가 선행할 때의 중성 ‘ㅡ’의 인식률도 28.3%로 매우 낮게 나타나며, 이때의 ‘ㅡ’는 대부분 ‘ㅜ’로 오인식된다(45.0%). 음성 인식 상황에서는 이처럼 음절 내에 불안정한 음소가 존재할 때 인접한 요소들의 인식률도 함께 낮아지는 모습이 관찰된다. ‘ㅡ’가 후행할 때 초성 ‘ㅁ’의 인식률이 낮게 나타나는 것도 동궤의 현상이라 할 수 있다. 그 외에 CV음절의 ‘ㅁ’는 후행 모음이 ‘ㅐ’인 경우에, CVC음절의 ‘ㅁ’는 후행 모음이 ‘ㅣ’인 경우에 인식률이 유독 낮아지는데, 후행 모음 환경은 다르나 모든 경우 오인식 자음은 ‘ㄴ’로 확인된다. CVC음절에서는 초성 ㅁ의 존재가 인식되지 않는 경우도 관찰된다. 유사 음소는 ‘ㄴ’이다.

다음으로는 ‘ㄹ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

44) 흔히 국어의 음소배열제약으로 인해 양순음 뒤의 ‘ㅡ’는 발음되지 못하는 것으로 여겨지기도 하나, 양순음 뒤의 ‘ㅡ’와 기타 자음 뒤의 ‘ㅡ’의 산출 실험을 통해 양순음 뒤 ‘ㅡ’의 음가를 분석한 신우봉(2010)에 따르면, 양순음 뒤의 ‘ㅡ’의 실제 발음은 ‘ㅜ’보다는 ‘ㅡ’에 가까운 것으로 확인된다고 한다. 이는 곧 양순음 뒤의 ‘ㅡ’라 하더라도 실제 음소 ‘ㅜ’로 발음되지는 않음을 의미한다. 하지만 이 논의의 ‘ㅡ+C+ㅡ’ 연쇄의 지각 실험에서는 기타 자음 뒤의 ‘ㅡ’를 ‘ㅜ’로 지각하는 경우는 0.002%에 불과하지만, 양순음 뒤의 ‘ㅡ’를 ‘ㅜ’로 지각하는 경우는 21.7%에 달하는 것을 통해 볼 때, 양순음 뒤의 ‘ㅡ’를 전형적인 ‘ㅜ’의 음가로 발음하지는 않으나, 지각 시에는 ‘ㅜ’와 혼동될 수 있다는 점이 확인된다.

음절 구조	인식 자음	후행 모음						
		ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㄹ⇒ㄹ	88.0(44)	66.0(33)	60.0(30)	90.0(45)	84.0(42)	84.0(42)	92.0(46)
	ㄹ⇒ㄴ	10.0(5)	18.0(9)	24.0(12)	-	-	6.0(3)	2.0(1)
	ㄹ⇒ㄱ	-	14.0(7)	-	-	-	4.0(2)	-
CVC	ㄹ⇒ㄹ	92.0(276)	92.7(278)	87.3(262)	94.0(282)	87.0(261)	68.7(206)	85.7(257)
	ㄹ⇒ㄴ	3.0(9)	4.7(14)	1.0(3)	1.3(4)	1.7(5)	6.0(18)	6.3(19)
	ㄹ⇒∅	1.7(5)	1.7(5)	3.7(11)	0.3(1)	5.3(16)	17.3(52)	5.0(15)

<표 38> 초성 ‘ㄹ’의 후행 모음별 인식률

CV음절 초성 ‘ㄹ’의 인식률 평균은 80.6%, CVC음절 평균은 86.8%이며, CV음절에서는 후행 모음이 ‘ㅐ, ㅓ’일 때, CVC음절에서는 후행 모음이 ‘ㅡ, ㅣ’일 때 평균보다 인식률이 낮게 나타난다. 후행 모음 환경은 다르나 모든 경우 오인식 자음은 ‘ㄴ’인 것으로 확인된다. ‘ㄹ’는 어두 초성 위치에서는 주로 탄설음 [ɾ]로 실현되고, 신지영(2014: 251)에 서술된 바와 같이 탄설음은 폐쇄음과 마찬가지로 스펙트로그램에서 음향 에너지가 거의 관찰되지 않는 짧은 흰 공간으로 나타나는 특성을 가진다. 이러한 ‘ㄹ’의 음향 특성으로 인해 CVC음절에서는 초성 ㄹ의 존재가 인식되지 않는 오류가 관찰되기도 한다. 초성 ‘ㄹ’의 음소로의 오인식은 대개 ‘ㄴ’로 나타나, 이에 ‘ㄹ’의 유사 음소는 ‘ㄴ’인 것으로 간주한다.

3.2. 종합적 검토

지금까지 살펴본 초성의 음성 인식 오류를 종합적으로 검토해 보면, 평음과 경음 간, 평음과 격음 간, 경음의 평음으로의 오인식과 같은 발성 유형과 관련된 오류가 눈에 띈다. 이러한 오류들 중 ‘평음 ⇒ 경음’으로의 인식 오류 양상에 대해 먼저 정리해 보이면 아래의 <표 39>와 같다.

음절 구조	인식 자음 \ 후행모음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㄱ⇒ㄲ	-	-	14.0(7)	-	-	-	-
	ㄷ⇒ㄸ	-	-	-	18.0(9)	-	2.0(1)	6.0(3)
	ㅂ⇒ㅃ	-	-	-	-	-	-	-
	ㅈ⇒ㅉ	-	-	-	2.0(1)	16.0(8)	2.0(1)	-
	ㅅ⇒ㅆ	-	-	-	-	-	4.0(2)	36.0(18)
CVC	ㄱ⇒ㄲ	0.3(1)	0.3(1)	3.3(10)	9.0(27)	7.3(22)	7.3(22)	7.0(21)
	ㄷ⇒ㄸ	-	7.0(21)	2.3(7)	4.3(13)	6.7(20)	17.3(52)	19.7(59)
	ㅂ⇒ㅃ	-	-	2.0(6)	15.7(47)	10.0(30)	8.0(24)	4.3(13)
	ㅈ⇒ㅉ	0.7(2)	2.0(6)	-	14.3(43)	16.3(49)	12.7(38)	13.0(39)
	ㅅ⇒ㅆ	-	2.0(6)	-	-	6.3(19)	19.7(59)	2.0(6)

<표 39> 후행 모음별 '평음 ⇒ 경음' 인식 오류

<표 39>를 통해 확인할 수 있듯이 평음에서 경음으로의 오인식은 CV음절보다는 CVC음절에서 다수 관찰되고, 'ㅏ, ㅐ, ㅓ'와 같이 개구도가 큰 모음이 후행할 때보다는 비교적 개구도가 작은 'ㅗ, ㅜ, ㅡ, ㅣ'가 후행하는 환경에서 더 많은 오류가 발생한다. 또한 'ㄷ, ㅅ, ㅈ'와 같은 [+coronal] 자질을 가지는 자음들이 [-coronal]인 'ㄱ, ㅂ'에 비해 경음으로 오인식되는 비율이 높은데,[45] 이를 종합하면 비교적 입을 적게 벌리는 환경에서 치조·경구개 부근에 장애가 발생하는 경우에, 평음에서 경음으로의 인식 오류가 높게 나타나는 것으로 정리해 볼 수 있을 듯하다.

다음으로 살펴볼 '경음 ⇒ 평음'으로의 인식 오류는 경음과 평음 간의 음향적 유사성과 함께 어두 경음화를 고려한 일부 과도 교정이 적용된 오류 유형으로 보인다.

45) 참고로 각 초성별 전체 오류 빈도와 평음에서 경음으로의 오류 빈도를 백분율로 계산한 결과는 다음과 같다.

인식 자음	평음⇒경음 오류율
ㄱ⇒ㄲ	9.5
ㄷ⇒ㄸ	14.2
ㅂ⇒ㅃ	10.3
ㅈ⇒ㅉ	20.6
ㅅ⇒ㅆ	53.3

음절 구조	후행모음 / 인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
CV	ㄲ⇒ㄱ	-	-	-	-	-	22.0(11)	6.0(3)
	ㄸ⇒ㄷ	-	-	-	-	-	-	-
	ㅃ⇒ㅂ	-	-	-	-	-	-	-
	ㅉ⇒ㅈ	2.0(1)	-	16.0(8)	16.0(8)	6.0(3)	8.0(4)	16.0(8)
	ㅆ⇒ㅅ	-	-	-	-	14.0(7)	18.0(9)	6.0(3)
CVC	ㄲ⇒ㄱ	-	19.3(58)	2.0(6)	1.0(3)	-	0.3(1)	20.3(61)
	ㄸ⇒ㄷ	-	-	-	-	-	-	-
	ㅃ⇒ㅂ	13.0(39)	3.0(9)	5.3(16)	2.7(8)	0.7(2)	-	6.3(19)
	ㅉ⇒ㅈ	0.3(1)	4.7(14)	2.0(6)	4.7(14)	7.0(21)	6.3(19)	**19.7(59)**
	ㅆ⇒ㅅ	2.0(6)	**43.7(131)**	0.7(2)	5.3(16)	21.7(65)	7.0(21)	54.0(162)

<표 40> 후행 모음별 '경음 ⇒ 평음' 인식 오류

위의 <표 40>에서 굵게 표시한 예들은 대부분 언어 모델에 의해 과도 교정된 오류가 반영된 것으로 추정되는 경우이다. 구체적으로는 CVC음절의 실험 대상 음절 '찝'이 '집'으로, '쌕'이 '섹스'로 오인식된 예들이 이에 해당되는데, 이들은 사용 빈도나 어두 경음화와 같은 자연 발화의 발음 특성을 고려한 교정형으로 판단된다. 이와 같은 교정형들을 제외한 나머지 예들은, 경음과 평음의 VOT 특성과 관련지어 이해해 볼 수 있을 듯하다. 어두의 평음과 경음은 VOT 값에 있어서 가장 큰 차이를 보이는데, 어두에서 짧은 VOT 값을 가지는 경음의 VOT가 길게 발음될 경우 평음으로 인식되기도 한다는 점에 주목해 본다면, 위와 같은 경음에서 평음으로의 인식 오류는 VOT의 길이와 관련되어 나타난 오류인 듯 보인다.

마지막으로 평음과 격음 간의 인식 오류는 이들이 가진 음향 특성의 유사성과 함께, 성별에 따라 다른 F0값이 영향을 준 현상으로 간주된다. 이에 대해서는 성별 요인에 의한 음성 인식 오류 유형을 다루는 6.3.에서 후술된다.

다음으로는 자음의 부류에 따라 초성의 인식 오류 양상을 정리해 보도록 한다. 먼저 파열음 중 평음은 주로 후행 모음과 조음 위치가 유사하거나 음향적 특성이 겹칠 때 인식률이 낮아지는 경향을 보인다. 이때 평음은 주로 격음과 혼동되는데, 이러한 오류는 전술한 바와 같이 성별에 따라 다른 F0값과의 관계 속에서 이해될 수 있을 듯하다. 남성의 낮은 F0값은 격음에서 평음으로의 오인식을, 여성의 높은 F0

값은 평음에서 격음으로의 오인식을 유발하는 것으로 추정된다. 경음과 격음은 경음 간, 격음 간의 오인식이 주로 나타난다. 이는 해당 음소가 경음과 격음이라는 점은 비교적 정확히 인식된다는 의미이다. 경음과 격음은 '긴장'이나 '기식'과 같은 유표적인 특성을 가지고 있어 평음보다 인식이 용이한 듯 보인다.

마찰음과 파찰음도 경음·격음과 마찬가지로 마찰음이라는 범주와 파찰음이라는 범주는 비교적 명확히 지각되는 듯 보인다. 단 이들은 마찰 정도에 따라 'ㅅ'와 'ㅆ', 'ㅊ'와 'ㅅ'가 혼동되는 경향이 있으므로, 마찰 정도 이외의 변별 기준과 마찰음 및 파찰음 내에서의 인식 기준이 정밀화될 필요가 있다.

후술할 종성에서는 음소의 존재 자체가 인식되지 않는 오류가 빈번하게 발생하지만, 초성에서는 이와 같은 오류가 드물게 나타난다. 공명음은 인식률이 마찰음 다음으로 높으나, 공명음들 간의 인식 오류와 더불어 음소의 존재가 인식되지 않는 오류가 발생하기도 하는데, 이는 공명음의 음성 신호가 장애음들에 비해 미약한 데서 기인된 오류로 추정된다.

4. 중성의 음성 인식 오류 양상 및 유사 음소

본 장에서는 중성 모음들의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다. 실험에 사용된 중성 모음 중 단모음은 /ㅏ, ㅐ, ㅓ, ㅗ, ㅜ, ㅡ, ㅣ/이며, 이중모음은 j계 /ㅑ, ㅒ, ㅕ, ㅛ, ㅠ, ㅢ/와 w계 /ㅘ, ㅝ, ㅙ, ㅟ/이다. 중성의 인식률은 초성 자음을 기준으로 나누어 제시하였으며, 필요에 따라 초성뿐만 아니라 종성도 함께 고려하여 인식 양상을 확인하였다.

4.1. 오류 양상 및 유사 음소

4.1.1. 단모음

개별 단모음들의 인식 오류 결과를 구체적으로 살피기에 앞서 먼저 보일 것은 음절 구조별 단모음들의 전반적인 인식 결과이다.

	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	계
V	94.0	44.0	96.0	96.0	12.0	70.0	100.0	73.1
VC	94.7	98.3	96.0	82.7	56.3	95.7	99.0	89.0
CV	96.4	88.3	88.0	85.7	85.0	82.1	98.0	89.1
CVC	95.8	96.4	89.3	89.2	82.6	70.7	96.2	88.6
계	95.2	81.8	92.3	88.4	59.0	79.6	98.3	

<표 41> 음절 구조별 단모음 인식률

실험에 사용된 모든 단모음들의 인식률 평균은 85.0%이며, CV > VC > CVC > V 순으로 낮아지지만, V구조를 제외한 나머지 음절 구조별 인식률은 거의 유사하다. 모음별 인식률은 ㅣ > ㅏ > ㅓ > ㅗ> ㅐ > ㅡ > ㅜ 순으로 낮아져, 인식률이 가

장 높은 모음은 ‘ㅣ’, 가장 낮은 모음은 ‘ㅜ’인 것으로 확인된다. 이수길(2005)에서는 음성 파형이 약하게 나타나는 ‘ㅣ, ㅡ, ㅜ’ 등이 음성 인식 상황에서 잘 구별되지 않음을 지적한 바 있는데, 본고의 실험 결과 이 중 ‘ㅡ, ㅜ’는 이수길(2005)의 논의와 같이 인식률이 저조하게 나타났으나, /ㅣ/는 가장 높은 인식률을 보였다. 위의 <표 41>의 인식률을 음절 구조별로 인식률 순서에 따라 정리한 것이 아래의 <표 42>이다.

음절구조	인식률						
V	ㅣ	ㅓ, ㅗ		ㅏ	ㅡ	ㅐ	ㅜ
VC	ㅣ	ㅐ	ㅓ	ㅡ	ㅏ	ㅗ	ㅜ
CV	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ
CVC	ㅐ	ㅣ	ㅏ	ㅓ	ㅗ	ㅜ	ㅡ

<표 42> 음절 구조별 단모음 인식률 순

<표 42>에서 음영으로 표시한 모음들은, 각 음절 구조별 평균 인식률보다 낮거나, 평균과 소수점 이하만의 차이를 보이는 경우를 나타낸 것이다. 그 중 빗금으로 표시한 V음절 구조의 모음 ‘ㅐ’는 실험 대상 음절 ‘애’가 대부분 ‘얘’로 오인식되어 전체 인식률이 낮아진 경우인데, 이를 제외하면 인식률이 낮은 모음들은 모두 후설모음이다. 이제 좀 더 구체적으로 초성이 있는 CV음절과 CVC음절을 대상으로 하여, 초성 자음별 단모음들의 인식률을 정리해 보이도록 한다.

초성 자음별 중성 인식률(CV)													
ㅏ		ㅐ		ㅓ		ㅗ		ㅜ		ㅡ		ㅣ	
ㅉ	100.0	ㄴ	100.0	ㄲ	100.0	ㄱ	100.0	ㅋ	100.0	ㄱ	100.0	ㄱ	100.0
ㅎ	100.0	ㄸ	100.0	ㄴ	100.0	ㄷ	100.0	ㅌ	100.0	ㅋ	100.0	ㄸ	100.0
ㅍ	100.0	ㅁ	100.0	ㄷ	100.0	ㄸ	100.0	ㅍ	100.0	ㄷ	98.0	ㅁ	100.0
ㅊ	100.0	ㅅ	100.0	ㅅ	100.0	ㅅ	100.0	ㄱ	98.0	ㅆ	98.0	ㅂ	100.0
ㄴ	98.0	ㅆ	100.0	ㅊ	100.0	ㅆ	98.0	ㄴ	98.0	ㅅ	96.0	ㅃ	100.0
ㄷ	98.0	ㅊ	100.0	ㄱ	98.0	ㅋ	98.0	ㅂ	98.0	ㅌ	96.0	ㅉ	100.0
ㅈ	98.0	ㅎ	100.0	ㅆ	98.0	ㅌ	98.0	ㄷ	96.0	ㅎ	96.0	ㅋ	100.0
ㅁ	98.0	ㅃ	98.0	ㅋ	98.0	ㅍ	98.0	ㅅ	94.0	ㄲ	92.0	ㅍ	100.0
ㄹ	98.0	ㅈ	98.0	ㅍ	98.0	ㄲ	94.0	ㅉ	94.0	ㄹ	92.0	ㄴ	98.0
ㄱ	96.0	ㅂ	96.0	ㅎ	98.0	ㅃ	94.0	ㅎ	92.0	ㄸ	90.0	ㄷ	98.0
ㅃ	96.0	ㅉ	94.0	ㅈ	96.0	ㅎ	92.0	ㅊ	90.0	ㅉ	90.0	ㄹ	98.0
ㅅ	96.0	ㅋ	90.0	ㅌ	96.0	ㄹ	90.0	ㄹ	84.0	ㅂ	80.0	ㅈ	98.0
ㅋ	96.0	ㅌ	90.0	ㅂ	94.0	ㅂ	90.0	ㅈ	82.0	ㅍ	72.0	ㅌ	98.0
ㅌ	96.0	ㄹ	84.0	ㅃ	92.0	ㄴ	84.0	ㅁ	76.0	ㅈ	70.0	ㅆ	96.0
ㄸ	94.0	ㄲ	72.0	ㄸ	84.0	ㅉ	80.0	ㅆ	76.0	ㅊ	70.0	ㅊ	96.0
ㅂ	92.0	ㄷ	72.0	ㄹ	70.0	ㅊ	80.0	ㄲ	62.0	ㄴ	68.0	ㅎ	96.0
ㅆ	90.0	ㅍ	54.0	ㅉ	62.0	ㅈ	44.0	ㅃ	50.0	ㅃ	42.0	ㄲ	94.0
ㄲ	90.0	ㄱ	42.0	ㅁ	0.0	ㅁ	2.0	ㄸ	40.0	ㅁ	28.0	ㅅ	92.0
평균	96.4	평균	88.3	평균	88.0	평균	85.7	평균	85.0	평균	82.1	평균	98.0

초성 자음별 중성 인식률(CVC)													
ㅏ		ㅐ		ㅓ		ㅗ		ㅜ		ㅡ		ㅣ	
ㄷ	99.7	ㅊ	100.0	ㄲ	98.3	ㄸ	98.3	ㅋ	99.7	ㄲ	98.0	ㅊ	100.0
ㅂ	99.7	ㅍ	99.3	ㅊ	97.0	ㄷ	97.7	ㅌ	94.3	ㅋ	96.3	ㅈ	99.3
ㄴ	99.0	ㅅ	99.3	ㄱ	96.7	ㅌ	97.0	ㄱ	94.0	ㄱ	93.0	ㅁ	99.3
ㅌ	98.0	ㅁ	99.0	ㅋ	96.0	ㄱ	95.7	ㅍ	92.7	ㄷ	91.0	ㅂ	98.7
ㅈ	97.3	ㅃ	99.0	ㅅ	95.0	ㅎ	94.3	ㅈ	92.3	ㄸ	89.3	ㄴ	98.3
ㄱ	97.3	ㄴ	98.7	ㅈ	93.7	ㅂ	94.0	ㅂ	92.0	ㅆ	89.0	ㄷ	98.0
ㅁ	96.7	ㅂ	98.7	ㄴ	93.7	ㄲ	93.0	ㅅ	91.7	ㅎ	88.3	ㅋ	97.7
ㄹ	96.3	ㄸ	98.6	ㄷ	93.3	ㅉ	92.7	ㅊ	90.0	ㅌ	87.3	ㄸ	97.0
ㅍ	96.3	ㅌ	98.6	ㅆ	92.3	ㅍ	92.0	ㄷ	88.0	ㄴ	84.3	ㅍ	97.0
ㄸ	95.3	ㅎ	98.4	ㅌ	92.3	ㅋ	91.7	ㄴ	82.7	ㅅ	83.3	ㅌ	96.7
ㅉ	95.3	ㅈ	98.0	ㅉ	91.0	ㄹ	90.3	ㅎ	80.3	ㅉ	68.7	ㅃ	96.3
ㄲ	94.7	ㄹ	98.0	ㅂ	90.7	ㅈ	90.0	ㅉ	79.3	ㅈ	54.3	ㄱ	95.7
ㅃ	94.3	ㅋ	97.7	ㅃ	90.3	ㅅ	89.0	ㅃ	75.3	ㄹ	52.0	ㅉ	94.7
ㅆ	94.3	ㅆ	97.3	ㅎ	89.7	ㅆ	89.0	ㅆ	75.3	ㅂ	50.3	ㅆ	94.3
ㅊ	94.3	ㄷ	96.4	ㅍ	89.3	ㅊ	88.7	ㅁ	73.3	ㅍ	45.3	ㄹ	93.3
ㅎ	94.0	ㅉ	94.4	ㄹ	81.3	ㄴ	84.0	ㄲ	64.0	ㅃ	42.7	ㄲ	93.0
ㅋ	93.0	ㄱ	92.6	ㄸ	80.3	ㅃ	77.3	ㄹ	61.7	ㅊ	31.7	ㅅ	91.7
ㅅ	88.7	ㄲ	70.7	ㅁ	47.3	ㅁ	51.3	ㄸ	60.0	ㅁ	28.3	ㅎ	91.3
평균	95.8	평균	96.4	평균	89.3	평균	89.2	평균	82.6	평균	70.7	평균	96.2

<표 43> 초성 자음별 중성 인식률

위의 <표 43>에서 음영으로 표시한 부분은 각 모음별로 평균 이하의 인식률을 보이는 경우를 나타낸 것이다. 빗금친 부분은 CV음절과 CVC음절 구조에서 공통적으로 평균 이하의 인식률을 보이는 경우이다. 음성 인식 오류에는 매우 다양한 요인들이 영향을 미치는 까닭에, 단순히 순서대로 선행 자음별 인식률을 나열하는 것만으로는 특정 환경에서 인식률이 낮게 나타나는 요인을 파악하기 어려울 것이다. 또한 인식률 자체가 높은 모음은 오류 발생 빈도 자체가 적어 인식률 순서를 나누어 오류 발생의 원인을 살피는 일이 의미가 없을 수도 있다. 그럼에도 CV음절과 CVC음절 구조에서 공통적으로 낮은 인식률을 보이는 경우, 즉 빗금친 자음들을 중심으로 눈에 띄는 점을 찾아보자면 다음과 같다.

모음 'ㅡ'는 양순음과 일부 치조·경구개 자음이 선행할 때 인식률이 낮아지며, 'ㅣ'는 'ㅣ' 모음 앞에서 구개음을 변이음으로 가지는 자음들이 선행하는 경우에 낮은 인식률을 보이는 경향이 있다. 'ㅜ'는 주로 경음이 선행할 때의 인식률이 낮은 편이며, 'ㅗ'와 'ㅓ'는 선행 자음이 'ㅁ'일 때 거의 인식이 되지 못하는 모습을 보인다. 이는 특히 CV음절 구조에서 두드러지게 나타나, 이때의 'ㅓ', 'ㅗ' 인식률은 각각 0%, 2.0%에 그친다. 'ㅓ'는 'ㅁ' 외에는 'ㄹ'와 'ㄸ'가 선행할 때 인식률이 낮아진다. 'ㅐ'는 선행 자음이 'ㄱ, ㄲ, ㄷ'일 때, 'ㅏ'는 선행 자음이 경음일 때 인식률이 낮아지는 듯하나 매우 산발적으로 나타난다. CV음절과 CVC음절의 전반적인 인식률은 이상과 같은 양상을 가지고 있으며, 다음으로는 V, VC, CV, CVC 음절의 세부 환경별 인식률을 초성 자음이 있는 경우에는 초성 자음을 기준으로(CV, CVC), 종성이 있는 경우에는 종성 자음을 기준으로(VC) 나누어 살펴보도록 한다.

V 음절						
ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
94.0	44.0	96.0	96.0	12.0	70.0	100.0

<표 44> V음절의 세부 환경별 인식률

VC 음절													
ㄱ 종성							ㄴ 종성						
ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
90.0	98.0	100.0	32.0	2.0	94.0	98.0	96.0	98.0	100.0	100.0	86.0	100.0	100.0
ㄹ 종성							ㅁ 종성						
ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
100.0	96.0	86.0	94.0	98.0	90.0	100.0	98.0	100.0	96.0	90.0	50.0	100.0	98.0
ㅂ 종성							ㅇ 종성						
ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
90.0	98.0	98.0	88.0	56.0	94.0	100.0	94.0	100.0	96.0	92.0	46.0	96.0	98.0

<표 45> VC음절의 종성별 인식률

ㅏ	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅂ	ㅃ	ㅍ
CV	82.0	94.0	90.0	84.0	70.0	78.0	52.0	86.0	64.0
CVC	97.3	94.7	93.0	99.7	95.3	98.0	99.7	94.3	96.3
	ㅈ	ㅉ	ㅊ	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
CV	86.0	98.0	100.0	94.0	98.0	86.0	100.0	88.0	96.0
CVC	97.3	95.3	94.3	78.7	94.3	94.0	99.0	96.3	96.7

<표 46> CV, CVC음절 초성별 모음 'ㅏ' 인식률

ㅐ	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅂ	ㅃ	ㅍ
CV	88.0	96.0	72.0	66.0	62.0	82.0	70.0	82.0	68.0
CVC	92.3	70.7	97.7	96.3	98.7	98.7	98.7	99.0	99.3
	ㅈ	ㅉ	ㅊ	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
CV	82.0	86.0	100.0	98.0	100.0	98.0	100.0	66.0	56.0
CVC	98.0	94.3	100.0	89.3	97.3	98.3	98.7	98.0	99.0

<표 47> CV, CVC음절 초성별 모음 'ㅐ' 인식률

ㅓ	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅂ	ㅃ	ㅍ
CV	42.0	98.0	86.0	92.0	76.0	64.0	12.0	40.0	68.0
CVC	96.7	98.3	96.0	93.3	80.3	92.3	90.7	90.3	89.3
	ㅈ	ㅉ	ㅊ	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
CV	90.0	72.0	100.0	100.0	94.0	86.0	98.0	60.0	100.0
CVC	93.7	91.0	97.1	85.0	92.3	89.7	93.7	81.3	47.3

<표 48> CV, CVC음절 초성별 모음 ‘ㅓ’ 인식률

ㅗ	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅂ	ㅃ	ㅍ
CV	36.0	94.0	96.0	44.0	100.0	94.0	26.0	76.0	86.0
CVC	95.7	93.0	91.7	97.7	98.3	97.0	94.0	77.3	92.0
	ㅈ	ㅉ	ㅊ	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
CV	74.0	80.0	80.0	100.0	98.0	88.0	96.0	90.0	98.0
CVC	90.0	92.7	88.6	79.0	89.0	94.3	84.0	90.3	51.3

<표 49> CV, CVC음절 초성별 모음 ‘ㅗ’ 인식률

ㅜ	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅂ	ㅃ	ㅍ
CV	84.0	78.0	96.0	36.0	90.0	96.0	70.0	74.0	82.0
CVC	94.0	64.0	99.7	88.0	60.0	94.3	92.0	75.3	92.7
	ㅈ	ㅉ	ㅊ	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
CV	54.0	88.0	90.0	90.0	80.0	90.0	92.0	84.0	96.0
CVC	92.3	79.3	100.0	81.7	75.3	80.3	82.7	61.7	73.3

<표 50> CV, CVC음절 초성별 모음 ‘ㅜ’ 인식률

ㅡ	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅂ	ㅃ	ㅍ
CV	88.0	78.0	94.0	36.0	90.0	90.0	12.0	54.0	40.0
CVC	93.0	98.0	96.3	91.0	89.3	87.3	50.3	42.7	45.3
	ㅈ	ㅉ	ㅊ	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
CV	76.0	86.0	70.0	90.0	82.0	90.0	92.0	84.0	74.0
CVC	54.3	68.7	31.4	73.3	89.0	88.3	84.3	52.0	28.3

<표 51> CV, CVC음절 초성별 모음 ‘ㅡ’ 인식률

ㅣ	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅂ	ㅃ	ㅍ
CV	66.0	90.0	100.0	58.0	86.0	70.0	66.0	88.0	96.0
CVC	95.7	93.0	97.7	98.0	97.0	96.7	98.7	96.3	97.0
	ㅈ	ㅉ	ㅊ	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
CV	68.0	76.0	96.0	64.0	88.0	86.0	82.0	92.0	98.0
CVC	99.3	94.7	100.0	83.3	94.3	91.3	98.3	93.3	99.3

<표 52> CV, CVC음절 초성별 모음 ‘ㅣ’ 인식률

음절 구조마다 달리 나타나는 중성의 인식률 차이가 음절 구조와의 관련성을 의미하는 것인지 알아보고자 Chi-square test를 진행한 결과 p-value가 0.05 이하로 도출되어 중성의 인식률 차이는 음절 구조와 관련된 것으로 판단하였다(p-value < .001). 뿐만 아니라 초성 자음의 종류와 종성의 유무도 인식률과 유의미한 관계를 가지는 것으로 확인된다(p-value < .001). 이를 바탕으로 중성의 인식률에는 음절 구조의 차이와 초성 자음의 종류 및 개폐음절 여부가 영향을 주는 것으로 간주할 수 있을 듯하다.

이제 각 단모음들의 음성 인식 오류 양상과 유사 음소에 대해 하나씩 구체적으로 살펴보도록 한다. 선행 자음이 후행 모음의 인식에 미치는 영향 관계를 파악하기 위해 초성이 없는 V, VC음절과, 초성이 존재하는 CV, CVC음절을 나누어 살펴보고자 한다. 먼저 제시할 것은 V, VC음절의 경우이다.

V 음절							
ㅏ		ㅐ		ㅓ		ㅗ	
ㅏ	94.0(47)	ㅐ	44.0(22)	ㅓ	96.0(48)	ㅗ	96.0(48)
ㅓ	6.03(3)	ㅒ	28.0(14)	ㅗ	4.0(2)	ㅣ	4.0(2)
		ㅣ	22.0(11)				

ㅜ		ㅡ		ㅣ	
ㅜ	12.0(6)	ㅡ	70.0(35)	ㅣ	100.0(50)
ㅗ	52.0(26)	ㅣ	26.0(13)		
ㅣ	26.0(13)				

<표 53> V음절의 중성 인식률

<표 53>에 제시된 V음절 구조의 전체 인식률 평균은 73.1%이고, 이보다 낮은 인식률을 보이는 모음은 'ㅐ, ㅜ, ㅡ'이다. 이 중 'ㅐ'는 'ㅒ'나 'ㅣ'로 오인식되는데, 'ㅒ'는 모두 남성에게서, 'ㅣ'는 여성에게서만 나타나는 오류라는 점이 인상적이다. 선행 연구에 따르면 20-30대 남녀 화자 사이에서 'ㅐ'와 'ㅣ' 간의 유의미한 조음 위치 차이는 나타나지 않는 것으로 간주되므로,[46] 본 실험에서는 'ㅐ'의 'ㅣ'로의 오인식이 여성 화자에게서만 나타났다는 점을 언급해 두기로만 한다.

한편 'ㅜ'는 인식률이 12%에 그쳐 가장 낮은 인식률을 기록하였는데, 이때 오인식되는 모음은 'ㅗ'와 'ㅣ'이다. 'ㅜ'는 후술할 음절 구조들에서도 낮은 인식률은 보인다. 'ㅜ'는 발화음에서 원순성이 감지된 경우라면 같은 원순모음 계열의 'ㅗ'와 혼동될 가능성이 높고, 원순성이 미약한 경우라면 'ㅡ'를 비롯한 평순모음으로 혼동될 가능성이 높은 듯하다. V음절 구조에서는 'ㅜ'로 인식되는 경우보다 'ㅗ'로 오인식되는 예가 더 빈번하게 나타난다.

다음으로는 VC음절 구조의 예를 살핀다.

VC 음절						
ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
94.7	98.3	96.0	82.6	56.3	95.7	99.0

<표 54> VC 음절의 중성 인식률

<표 54>에 제시된 VC구조의 각 모음별 인식률을 합한 전체 인식률 평균은 89.0%이고, 이보다 낮은 인식률을 보이는 모음은 'ㅗ, ㅜ'이다. 이들을 제외한 나머지 모음들은 오인식 빈도 자체가 적어, 여기에서는 'ㅗ'와 'ㅜ'의 인식 오류 양상만을 살펴보도록 한다.

46) 채서영(2001: 658)의 논의에 51세 이상 여성들이 남성에 비해 'i' 변이형을 보다 자주 사용한다는 /ㅐ/와 /ㅣ/ 간의 성별에 관한 언급이 있었으나, 통계적으로 유의미한 차이를 보이지는 않았다고 한다. 유클리디안 거리 공식에 기반하여 세대별 모음 사이의 거리를 측정한 장혜진·신지영·남호성(2015: 352)의 분석 결과에서는 20-30대 남녀 화자 간에 /ㅔ/와 /ㅣ/ 간의 유의미한 차이는 나타나지 않았다.

종성 / 인식모음	ㄱ	ㄴ	ㄹ	ㅁ	ㅂ	ㅇ
ㅗ ⇒ ㅗ	32.0(16)	100.0(50)	94.0(47)	90.0(45)	88.0(44)	92.0(46)
ㅗ ⇒ ㅜ	42.0(21)	-	-	-	-	2.0(1)
ㅗ ⇒ ㅓ	20.0(10)	-	6.0(3)	4.0(2)	12.0(6)	6.0(3)

<표 55> VC 음절의 종성별 'ㅗ' 인식 오류

VC음절의 'ㅗ'는 종성이 'ㄱ'인 경우에 한하여 인식률이 평균 이하로 나타나며, 이때 'ㅗ'는 'ㅜ'나 'ㅓ'로 오인식된다. 그리고 중성의 인식률이 아닌 실험 대상 음절 '옥'의 음절 자체의 인식률은 0%이다. 즉 '옥'을 '옥'으로 정확히 인식한 경우는 없었다는 것이다. 실험 대상 음절 '옥'의 'ㅗ'를 'ㅜ'로 오인식한 경우에는, 대상 음절의 초·중·종성 중 중성만 'ㅗ'에서 'ㅜ'로 오인식되었는데(인식 결과: '욱'), 'ㅓ'로 오인식한 경우에는 종성의 존재 또한 인식되지 않아, '옥'이 '어'로 인식되는 모습을 보인다. 이와 같은 현상은 'ㅗ'를 'ㅗ'로 바르게 인식한 예에서도 발견되는데, 실험 대상 음절 '옥'의 경우, 중성은 'ㅗ'로 정확히 인식되나, 초성에 자음이 첨가된 형태나(인식 결과: '꼭'), 종성이 탈락된 형태(인식 결과: '오')로 인식되는 모습이 보인다. 즉 실험 대상 음절 '옥'은 종성의 존재가 인식된 경우에는 '욱'이나 '꼭'으로 오인식되고, 종성의 존재가 인식되지 않은 경우에는 '오'나 '어'로 인식되며, 종성의 존재와 중성이 'ㅗ'라는 점이 감지된 경우에는 초성에 자음이 첨가된 형태로 인식된다. 초성이 없는 VC음절의 음성 인식 실험 시, 초성이 있는 형태로 인식형이 제시되는 경우가 있는데, 이때 첨가되는 초성은 대개 'ㅎ'이지만,[47] 중성이 'ㅗ, ㅜ'일 때에는 'ㄲ'와 'ㅃ'가 첨가된다.

다음으로는 VC 음절의 종성별 'ㅜ'의 인식 오류 양상에 대해 살펴보도록 한다.

47) 발화 시 약한 정도의 기식이나 마찰 소음이 포함되더라도 이를 초성 'ㅎ'으로 감지하는 것으로 보인다.

종성 / 인식모음	ㄱ	ㄴ	ㄹ	ㅁ	ㅂ	ㅇ
ㅜ ⇒ ㅜ	2.0(1)	86.0(43)	98.0(49)	50.0(25)	56.0(28)	46.0(23)
ㅜ ⇒ ㅗ	86.0(43)	4.0(2)	2.0(1)	18.0(9)	30.0(16)	28.0(14)
ㅜ ⇒ ㅓ	12.0(6)	-	-	-	4.0(2)	-
ㅜ ⇒ ㅡ	-	10.0(5)	-	32.0(16)	6.0(3)	26.0(13)

<표 56> VC 음절의 종성별 'ㅜ' 인식 오류

<표 56>을 통해 확인할 수 있는 바와 같이 VC음절의 'ㅜ'는 종성이 연구개음이나 양순음인 경우에 인식률이 평균 이하로 나타나며, 이때 'ㅜ'는 'ㅗ, ㅓ, ㅡ'로 오인식된다. 그 중에서도 종성이 'ㄱ'일 때의 인식률이 가장 낮게 도출되었는데, 이에 해당되는 실험 대상 음절 '욱'은 실험 대상 음절 자체의 인식률이 2%에 불과하다. 즉 '욱'을 '욱'으로 정확히 인식한 경우는 한 차례뿐이었던 것이다. 실험 대상 음절 '욱'은 종성이 인지된 경우에는 '옥'으로, 종성이 인지되지 않은 경우에는 '오'나 '어'로 오인식되었다.

이러한 오인식 양상은 종성이 'ㅂ'인 경우에도 유사하게 관찰되는데, 이에 해당되는 실험 대상 음절 '웁'의 음절 인식률은 0%였다. 종성의 'ㅂ'는 'ㄱ'로 오인식되는 경우가 많은데, 여기에서도 '웁'의 종성 'ㅂ'가 'ㄱ'로 인식되면 최종 인식형이 '욱'으로 도출되고, 종성이 인식이 안 될 경우에는 '우'나 '오'로 오인식되었다.

종합하자면 VC 음절 구조 중 인식률이 낮은 중성은 'ㅗ, ㅜ'이며, 이들은 주로 종성이 연구개음 계열이거나 양순음 계열일 때 혼란을 겪는다. 음성 인식 상황에서 연구개음은 양순음과 유사한 특성을 보이는 경우가 있고, 종성 'ㅂ'는 'ㄱ'로 오인식되는 경우가 빈번한데, 이는 이러한 'ㄱ'와 'ㅂ' 간의 혼동과 'ㅗ'와 'ㅜ' 간의 혼동이 상호 영향을 주어 나타난 오류인 듯하다.

다음으로는 CV, CVC음절의 각 단모음들의 음성 인식 오류 양상과 유사 음소에 대해 하나씩 구체적으로 살펴보도록 한다. 처음 제시할 모음은 'ㅏ'이다.

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅏ⇒ㅏ	96.0(48)	90.0(45)	96.0(48)	98.0(49)	94.0(47)	96.0(48)
ㅏ⇒ㅓ	4.0(2)	10.0(5)	2.0(1)	2.0(1)	4.0(2)	2.0(1)

	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅏ⇒ㅏ	92.0(46)	96.0(48)	100.0(50)	98.0(49)	100.0(50)	100.0(50)
ㅏ⇒ㅓ	2.0(1)	2.0(1)	-	2.0(1)	-	

	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅏ⇒ㅏ	96.0(48)	90.0(45)	100.0(50)	98.0(49)	98.0(49)	98.0(49)
ㅏ⇒ㅓ	2.0(1)	10.0(5)	-	2.0(1)	2.0(1)	-

<표 57> 중성 'ㅏ'의 선행 자음 환경별 유사 음소(CV)

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅏ⇒ㅏ	97.3(292)	94.7(284)	93.0(279)	99.7(299)	95.3(286)	98.0(294)
ㅏ⇒ㅓ	2.7(8)	4.3(13)	4.3(13)	0.3(1)	3.7(11)	1.0(3)

	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅏ⇒ㅏ	99.7(299)	94.3(283)	96.3(289)	97.3(292)	95.3(286)	94.3(283)
ㅏ⇒ㅓ	0.3(1)	4.0(12)	1.0(3)	0.7(2)	2.0(6)	5.7(17)

	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅏ⇒ㅏ	88.7(266)	94.3(283)	94.0(282)	99.0(297)	96.3(289)	96.7(290)
ㅏ⇒ㅓ	4.3(13)	3.3(10)	1.3(4)	1.0(3)	1.3(4)	1.0(3)

<표 58> 중성 'ㅏ'의 선행 자음 환경별 유사 음소(CVC)

위의 표는 CV음절과 CVC음절에서의 중성 단모음의 인식률을 선행 자음을 기준으로 나누어 제시한 것이며, 각 인식률 옆의 괄호 안의 숫자는 출현 빈도를 나타낸다. CV음절 중성 'ㅏ'의 인식률 평균은 96.4%, CVC음절 평균은 95.8%이며, 'ㅏ'를 주로 어떤 모음으로 오인식하는지 살펴본 결과, 오류 자체가 많지는 않으나 발생한 오류 형들은 대부분 CV음절과 CVC음절 모두 모음 'ㅓ'로 나타났다. 이에 'ㅏ'의 유사 음소는 'ㅓ'로 판단된다.

다음으로는 중성 /ㅐ/의 인식 오류 양상과 유사 음소에 대해 알아보도록 한다.

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅐ⇒ㅐ	42.0(21)	72.0(36)	90.0(45)	72.0(36)	100.0(50)	90.0(45)
ㅐ⇒ㅒ	54.0(27)	2.0(1)	10.0(5)	2.0(1)	-	2.0(1)
ㅐ⇒ㅙ	-	24.0(12)	-	26.0(13)	-	8.0(4)

	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅐ⇒ㅐ	96.0(48)	98.0(49)	54.0(27)	98.0(49)	94.0(47)	100.0(50)
ㅐ⇒ㅒ	-	-	44.0(22)	-	-	-
ㅐ⇒ㅙ	-	-	-	-	-	-

	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅐ⇒ㅐ	100.0(50)	100.0(50)	100.0(50)	100.0(50)	84.0(42)	100.0(50)
ㅐ⇒ㅒ	-	-	-	-	-	-
ㅐ⇒ㅙ	-	-	-	-	-	-

<표 59> 중성 'ㅐ'의 선행 자음 환경별 유사 음소(CV)

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅐ⇒ㅐ	92.6(277)	70.7(212)	97.7(293)	96.4(289)	98.6(296)	98.6(296)
ㅐ⇒ㅒ	4.3(13)	3.7(11)	-	-	0.3(1)	-
ㅐ⇒ㅙ	0.7(2)	23.3(70)	1.0(3)	0.7(2)	0.3(1)	-

	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅐ⇒ㅐ	98.7(296)	99.0(297)	99.3(298)	98.0(294)	94.4(283)	100.0(300)
ㅐ⇒ㅒ	-	-	-	-	-	-
ㅐ⇒ㅙ	-	-	-	-	-	-

	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅐ⇒ㅐ	99.3(298)	97.3(292)	98.4(295)	98.7(296)	98.0(294)	99.0(297)
ㅐ⇒ㅒ	-	-	-	0.3(1)	-	-
ㅐ⇒ㅙ	-	-	-	-	0.3(1)	0.3(1)

<표 60> 중성 'ㅐ'의 선행 자음 환경별 유사 음소(CVC)

CV음절 중성 'ㅐ'의 인식률 평균은 88.3%, CVC음절 평균은 96.4%로 음절 구조에 따른 인식률 차이가 큰 편이다. 'ㅐ'는 전체적으로 인식률이 높은 편이나, CV음절 구조

에서는 선행 자음이 ‘ㄱ, ㄲ, ㄷ, ㅍ’일 때, CVC음절 구조에서는 선행 자음이 ‘ㄲ’일 때 인식률이 낮게 나타난다. 이때 오인식되는 모음은 주로 ‘ㅐ’나 ‘ㅒ’인데, ‘깨, 대’는 ‘꽤, 돼’로 오인식되고, ‘개, 패’는 ‘걔, 폐’로 오인식된다. 모두 활음이 첨가된 형태로 오인식된 것인데, 이는 이들 환경에서의 ‘ㅐ’와 ‘ㅙ’, ‘ㅒ’ 간의 음성적 유사성에 의해 유발된 오류일 수도 있고, 자연 발화에서의 활음 탈락 현상을 바탕으로 적용된 언어 모델의 과도 교정의 결과일 수도 있을 듯하다. 이러한 현상이 CV음절에서 더 강하게 작용되어 CVC음절과의 인식률 차이가 벌어지게 된 것으로 보이며, 어떤 요인에 의한 것이든 실험 결과 관찰된 ‘ㅐ’의 유사 음소는 ‘ㅙ’와 ‘ㅒ’가 된다.

CVC음절 초성 ‘ㄲ’에서 나타나는 오류는 이 환경에 해당되는 실험 대상 음절 ‘깩, 깬, 깰, 깸, 깹, 깽’ 중 ‘깩, 깹’에 한정되어 나타난다. 종성 ‘ㄱ’와 ‘ㅂ’는 음성 인식 과정에서 그 존재 자체가 인식되지 못하는 경우가 빈번하게 관찰되는데, 여기에서도 ‘깩, 깹’의 종성 존재가 인식되지 않아 개음절 ‘깨’로 받아들여진 뒤, ‘꽤’로 교정된 것으로 보인다.

다음으로는 중성 ‘ㅓ’의 인식 오류 양상과 유사 음소에 대해 알아보도록 한다.

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅓ⇒ㅓ	98.0(49)	100.0(50)	98.0(49)	100.0(50)	84.0(42)	96.0(48)
ㅓ⇒ㅗ	-	-	-	-	14.0(7)	-
ㅓ⇒ㅜ	-	-	-	-	-	-
ㅓ⇒ㅕ	-	-	-	-	-	-

인식 모음	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅓ⇒ㅓ	94.0(47)	92.0(46)	98.0(49)	96.0(48)	62.0(31)	100.0(50)
ㅓ⇒ㅗ	4.0(2)	2.0(1)	-	-	-	-
ㅓ⇒ㅜ	-	-	-	4.0(2)	2.0(1)	-
ㅓ⇒ㅕ	-	-	-	-	32.0(16)	-

인식 모음	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅓ⇒ㅓ	100.0(50)	98.0(49)	98.0(49)	100.0(50)	70.0(35)	0.0(0)
ㅓ⇒ㅗ	-	-	-	-	14.0(7)	-
ㅓ⇒ㅜ	-	-	-	-	-	100.0(50)
ㅓ⇒ㅕ	-	-	-	-	-	-

<표 61> 중성 ‘ㅓ’의 선행 자음 환경별 유사 음소(CV)

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅓ⇒ㅓ	96.7(290)	98.3(295)	96.0(288)	93.3(280)	80.3(241)	92.3(277)
ㅓ⇒ㅏ	1.0(3)	0.7(2)	2.3(7)	4.7(14)	10.0(30)	4.7(14)
ㅓ⇒ㅡ	-	-	1.0(3)	-	-	-
ㅓ⇒ㅜ	-	-	-	-	-	-

인식 모음	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅓ⇒ㅓ	90.7(272)	90.3(271)	89.3(268)	93.7(281)	91.0(273)	97.0(291)
ㅓ⇒ㅏ	8.0(24)	7.0(21)	7.3(22)	4.7(14)	5.7(17)	2.3(7)
ㅓ⇒ㅡ	-	0.3(1)	-	-	-	-
ㅓ⇒ㅜ	-	-	-	-	-	-

인식 모음	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅓ⇒ㅓ	95.0(285)	92.3(277)	89.7(269)	93.7(281)	81.3(244)	47.3(142)
ㅓ⇒ㅏ	4.0(12)	5.0(15)	8.0(24)	3.7(11)	4.0(12)	9.3(28)
ㅓ⇒ㅡ	-	-	-	-	6.0(18)	-
ㅓ⇒ㅜ	-	-	-	-	-	38.3(115)

<표 62> 중성 ‘ㅓ’의 선행 자음 환경별 유사 음소(CVC)

CV음절 중성 ‘ㅓ’의 인식률 평균은 88.0%, CVC음절 평균은 89.3%이다. CV음절 구조에서는 선행 자음이 ‘ㅉ, ㅁ’일 때, CVC음절 구조에서는 선행 자음이 ‘ㄸ, ㄹ, ㅁ’일 때 인식률이 낮게 나타난다. 이때 오인식되는 모음은 음절 구조에 따라 차이를 보이는데, CV음절에서는 ‘ㅗ, ㅕ, ㅜ’가 오인식 모음으로, CVC음절에서는 ‘ㅏ, ㅡ, ㅜ’가 오인식 모음으로 나타난다. CV음절에서 ‘ㅓ’가 ‘ㅜ’로 오인식되는 경우는 선행 자음이 ‘ㅁ’일 때뿐인데, 이 환경에서는 실험 대상 음절 ‘머’의 ‘ㅓ’가 모두 ‘ㅜ’로 오인식되었다(100%). 중간형은 음성 인식 모델이 판단에 참고할 만한 정보가 없어, 대체로 감지된 음성의 음향적 특성을 주요 근거로 인식형을 도출해낼 것으로 추측되나, ‘뭐’와 같이 단음절 구성이 동시에 하나의 문장 역할을 담당할 수 있는 경우에는 중간형에서도 음향 모델 이외 모듈들의 판단이 주효하게 작용하는 듯하다.

CVC음절에서도 선행 자음이 ‘ㅁ’일 때의 인식률이 가장 낮게 나타나는데, 이때는 실험 대상 음절 ‘먹, 먼, 멀, 멈, 멉, 멍’ 중에서 ‘먹, 멉’은 종성의 존재가 인식되지

않아 개음절 형태의 '뭐'로 오인식되고, '멀'은 폐음절 '뭘'로 오인식된다. CV음절과 CVC음절에서의 'ㅓ'의 인식률 차이는, 선행 자음이 'ㅁ'인 CV음절의 'ㅓ'의 인식률이 0%인 데서 기인한 것으로, 이는 'ㅓ'의 음향적 요인에 의해 유발된 오류가 아닌, 언어모델에 의한 오류로 분류될 수 있을 것이다. CVC음절 구조에서는 'ㅓ'의 유사 음소는 'ㅏ'로 판단된다. 그런데 이때 'ㅓ'를 'ㅏ'로 오인식하는 오류는 여성에게서 두드러지게 관찰된다는 점이 특징적인데, 이는 성별 요인에 의한 인식 오류를 다루는 6.3.에서 후술된다.

다음으로는 중성 'ㅗ'의 인식 오류 양상과 유사 음소에 대해 알아보도록 한다.

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅗ⇒ㅗ	100.0(50)	94.0(47)	98.0(49)	100.0(50)	100.0(50)	98.0(49)
ㅗ⇒ㅓ	-	-	-	-	-	-
ㅗ⇒ㅜ	-	2.0(1)	-	-	-	2.0(1)
ㅗ⇒ㅓ	-	-	-	-	-	-

인식 모음	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅗ⇒ㅗ	90.0(45)	94.0(47)	98.0(49)	44.0(22)	80.0(40)	80.0(40)
ㅗ⇒ㅓ	2.0(1)	2.0(1)	-	-	-	-
ㅗ⇒ㅜ	-	-	-	2.0(1)	10.0(5)	-
ㅗ⇒ㅓ	8.0(4)	4.0(2)	-	54.0(27)	8.0(4)	20.0(10)

인식 모음	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅗ⇒ㅗ	100.0(50)	98.0(49)	92.0(46)	84.0(42)	90.0(45)	2.0(1)
ㅗ⇒ㅓ	-	2.0(1)	-	10.0(5)	-	-
ㅗ⇒ㅜ	-	-	4.0(2)	4.0(2)	-	-
ㅗ⇒ㅓ	-	-	-	-	-	98.0(49)

<표 63> 중성 'ㅗ'의 선행 자음 환경별 유사 음소(CV)

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅗ⇒ㅗ	95.7(287)	93.0(279)	91.7(275)	97.7(293)	98.3(295)	97.0(291)
ㅗ⇒ㅓ	2.0(6)	0.7(2)	3.0(9)	1.0(3)	0.7(2)	1.3(4)
ㅗ⇒ㅜ	0.7(2)	6.0(18)	5.3(16)	1.0(3)	0.3(1)	1.7(5)
ㅗ⇒ㅝ	-	-	-	-	-	-

인식 모음	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅗ⇒ㅗ	94.0(282)	77.3(232)	92.0(276)	90.0(270)	92.7(278)	88.7(266)
ㅗ⇒ㅓ	0.3(1)	1.0(3)	4.7(14)	1.7(5)	1.3(4)	-
ㅗ⇒ㅜ	2.7(8)	18.3(55)	3.3(10)	1.0(3)	5.0(15)	3.0(9)
ㅗ⇒ㅝ	-	1.0(3)	-	3.7(11)	0.7(2)	-

인식 모음	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅗ⇒ㅗ	89.0(267)	89.0(267)	94.3(283)	84.0(252)	90.3(271)	51.3(154)
ㅗ⇒ㅓ	2.3(7)	7.7(23)	5.7(17)	13.7(41)	5.0(15)	4.0(12)
ㅗ⇒ㅜ	7.3(22)	2.7(8)	-	1.7(5)	0.3(1)	3.7(11)
ㅗ⇒ㅝ	-	-	-	0.3(1)	0.3(1)	40.7(122)

<표 64> 중성 'ㅗ'의 선행 자음 환경별 유사 음소(CVC)

CV음절 중성 'ㅗ'의 인식률 평균은 85.7%, CVC음절 평균은 89.2%이며, 두 음절 구조에서 모두 'ㅓ, ㅜ, ㅝ'가 오인식 모음으로 확인된다. CV음절 구조에서는 선행 자음이 'ㅈ, ㅊ, ㅉ, ㅁ'일 때, CVC음절 구조에서는 선행 자음이 'ㅊ, ㅁ, ㅃ'일 때 인식률이 낮게 나타난다. 즉 평균보다 낮은 인식률을 보이는 선행 자음은 양순음 계열과 파찰음 계열의 자음인 것인데, 이 중 'ㅈ'와 'ㅁ'에 후행하는 'ㅗ'는 'ㅝ'로 오인식되는 빈도가 높다. 이는 앞서 살펴본 'ㅓ'의 경우와 마찬가지로 자연 발화에서 흔히 '줘', '뭐'가 [조], [모]로 발음된다는 점을 근거로 언어 모델에 의해 과도 교정된 형태로 보인다. 이를 제외하고 나면 음향적 유사성에 의한 인식 오류는 'ㅗ'를 'ㅓ'나 'ㅜ'로 오인식하는 경우에 한정된다고 볼 수 있을 듯하다.[48] 그런데 이때 'ㅗ'를 'ㅜ'로 오인

48) 황미경·조남민(2020)에서는 마찰음이 가진 소음성이 후행 모음의 F1과 F2값을 높이며, 그 영향은 [i] 모음보다는 후설모음 [u]의 환경에서 효과적으로 작용한다는 Kent and Read(2002)의 논의에 근거하여, 현대 한국어 모음체계에서 가장 후설에서 실현되는 모음인 'ㅗ'가 마찰음의 소음 특성에 가장 큰 영향을 받을 수 있음을 논의한 바 있다. 본 실험에서는 중성 'ㅗ'의 인식률이 주로 'ㅗ'와 원순성을 공유하는 '양순음' 또는 파찰음이 선행할 때

식하는 경우는 대부분 여성의 발화에서 관찰되는 오류이다. ‘ㅗ’와 ‘ㅜ’ 간의 오인식 예 중 남성과 여성의 오인식 비율과 빈도를 따로 정리하여 보이면 아래의 <표 65>와 같다.

	남성	여성
ㅗ ⇒ ㅜ	10.9%(15)	**73.2%(123)**
ㅜ ⇒ ㅗ	**55.9%(372)**	44.1%(294)

<표 65> 남녀 ‘ㅗ-ㅜ’ 간 오인식률

주지하다시피 현대 한국어 화자들의 후설원순모음의 조음 위치는 세대나 성별과 같은 사회언어학적 조건에 따라 상이하게 나타난다. 서울 출신의 20대 남녀들을 대상으로 단모음 포먼트를 분석한 조성문(2003) 논의에 따르면, 20대 여성 화자들의 경우 ‘ㅗ’와 ‘ㅜ’는 고저 대립이 없어지고, ‘ㅜ’는 원순성이 없는 [ɯ]로, ‘ㅗ’는 원래 ‘ㅜ’ 발음이라고 할 수 있는 [u]에 가깝게 대치되어 있는 것으로 보고된다. 후설원순모음들의 조음 위치에 관한 음향음성학적 연구인 윤규철·김순옥(2015)에서도 특히 20-30대 여성의 경우 ‘ㅗ’와 ‘ㅜ’가 고저대립이 아닌 전후대립의 양상을 보이는 것으로 간주한 바 있다. 두 논의 모두 20-30대 여성 화자들의 후설원순모음의 조음 위치 변화를 지적하고 있는 것인데, 본 연구의 실험 결과에서도 여성의 음성 인식 상황에서 ‘ㅗ’와 ‘ㅜ’ 간의 인식 오류가 높게 나타나는 것으로 볼 때, 이와 같은 후설원순모음에 관한 여성의 발화 특성이 음성 인식 오류를 유발하는 요인으로 작용한 듯하다. 이러한 관점을 받아들인다면 앞서 살핀 ‘ㅓ’의 ‘ㅏ’로의 오인식, ‘ㅗ’의 ‘ㅜ’로의 오인식은 모두 발화자의 성별에 따른 모음들의 음향적·조음적 특성과 관련된 음성 인식 오류로 분류된다.

다음으로는 중성 ‘ㅜ’의 인식 오류 양상과 유사 음소에 대해 알아보도록 한다.

낮게 관찰되는데, 추후 이를 파찰음의 소음 특성과 후행 모음의 포먼트 변화와의 관련성 속에서 이해해볼 수도 있을 듯하다.

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅜ⇒ㅜ	98.0(49)	62.0(31)	100.0(50)	96.0(48)	40.0(20)	100.0(50)
ㅜ⇒ㅗ	2.0(1)	34.0(17)	-	-	58.0(29)	-
ㅜ⇒ㅓ	-	-	-	-	-	-

	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅜ⇒ㅜ	98.0(49)	50.0(25)	100.0(50)	82.0(41)	94.0(47)	90.0(45)
ㅜ⇒ㅗ	2.0(1)	48.0(24)	-	-	-	-
ㅜ⇒ㅓ	-	-	-	2.0(1)	2.0(1)	-

	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅜ⇒ㅜ	94.0(47)	76.0(38)	92.0(46)	98.0(49)	84.0(42)	76.0(38)
ㅜ⇒ㅗ	2.0(1)	18.0(9)	2.0(1)	2.0(1)	8.0(4)	2.0(1)
ㅜ⇒ㅓ	-	-	-	-	-	22.0(11)

<표 66> 중성 'ㅜ'의 선행 자음 환경별 유사 음소(CV)

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅜ⇒ㅜ	94.0(282)	64.0(192)	99.7(299)	88.0(264)	60.0(180)	94.3(283)
ㅜ⇒ㅗ	-	34.3(103)	0.3(1)	8.7(26)	37.7(113)	5.0(15)
ㅜ⇒ㅡ	-	-	-	-	-	0.3(1)

	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅜ⇒ㅜ	92.0(276)	75.3(226)	92.7(278)	92.3(277)	79.3(238)	90.0(270)
ㅜ⇒ㅗ	6.0(18)	22.7(68)	4.7(14)	2.7(8)	14.7(44)	10.0(30)
ㅜ⇒ㅡ	-	-	1.0(3)	1.3(4)	4.3(13)	-

	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅜ⇒ㅜ	91.7(245)	75.3(226)	80.3(241)	82.7(248)	61.7(185)	73.3(220)
ㅜ⇒ㅗ	5.3(16)	14.0(42)	8.0(24)	9.7(29)	29.3(88)	14.7(44)
ㅜ⇒ㅡ	2.3(7)	8.0(24)	8.7(26)	-	-	-

<표 67> 중성 'ㅜ'의 선행 자음 환경별 유사 음소(CVC)

CV음절 중성 'ㅜ'의 인식률 평균은 85.0%, CVC음절 평균은 82.6%이며, 'ㅗ, ㅓ, ㅡ'가 오인식 모음으로 확인된다. CV구조에서는 선행 자음이 경음 'ㄲ, ㄸ, ㅃ, ㅆ' 또는 'ㅁ'일 때, CVC 구조에서 역시 경음 'ㄲ, ㄸ, ㅃ, ㅉ, ㅆ' 또는 'ㅎ, ㄹ, ㅁ'일 때 평균

이하의 인식률을 보인다. CV음절에서의 오인식 모음 ‘ㅓ’는 선행 자음이 주로 ‘ㅁ’인 경우에 한하여 관찰된다. 이를 제외하면 CV음절의 ‘ㅜ’는 대부분 ‘ㅗ’로 오인식되어 이를 유사 음소로 가진다. ‘ㅁ’에 후행하는 ‘ㅜ’가 ‘ㅓ’로 오인식된 경우 역시, 앞서 살펴본 예들과 같이 실험 대상 음절 ‘무’가 ‘모’로 지각된 뒤 이것이 언어 모델에 의해 ‘뭐’로 과도 교정된 결과로 볼 수 있을 듯하다.

한편 중성 ‘ㅜ’의 ‘ㅗ’와 ‘ㅡ’로의 오인식은 이들 모음 간의 음향적 유사성에서 기인한 오류인 것으로 추정되는데, 현대 한국어 ‘ㅜ’의 조음 위치에 관한 기존의 논의들은 ‘ㅜ’의 조음 공간에 대해 서로 다른 관점들을 보여주고 있다. 조성문(2005)에는 ‘ㅜ’가 ‘ㅗ’의 상승으로 인해 점차 전설화되며 ‘ㅡ’와 조음 위치가 유사해졌다는 견해가, 장혜진 외(2015), 이향원 외(2017)에서는 역시 ‘ㅗ’의 상승으로 ‘ㅗ’와 ‘ㅜ’가 세대에 따라 합류되고 있는 것으로 보는 견해가 제시되었으며, 성철재(2004)에서는 여기에 성별 요소를 더하여 여성 화자의 경우 ‘ㅜ’가 오히려 ‘ㅗ’보다 조음 자리가 더 낮고 후설화되는 경향을 보임을 밝힌 바 있다. 황미경·조남민(2020)은 ‘ㅡ’와 ‘ㅜ’의 중화가 의심된다는 점을 지적하였는데, 이를 종합하면 ‘ㅜ’의 조음 공간은 ‘ㅗ’, ‘ㅡ’와 일부 겹치는 방향으로 이동하고 있는 듯하다. ‘ㅜ’의 이러한 조음상의 특성은 음성 인식 오류에도 그대로 반영되어 나타난다. ‘ㅗ’의 ‘ㅜ’로의 오인식에는 남녀 성차가 관여하였으나, ‘ㅜ’의 ‘ㅗ’로의 오인식에는 성별 요인은 관여하지 않는 것으로 보인다.

다음으로는 중성 ‘ㅡ’의 인식 오류 양상과 유사 음소에 대해 알아보도록 한다.

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅡ⇒ㅡ	100.0(50)	92.0(46)	100.0(50)	98.0(49)	90.0(45)	96.0(48)
ㅡ⇒ㅐ	-	-	-	-	-	-
ㅡ⇒ㅜ	-	-	-	-	-	-
ㅡ⇒ㅓ	-	-	-	-	-	-
ㅡ⇒ㅣ	-	-	-	2.0(1)	-	2.0(1)

인식 모음	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅡ⇒ㅡ	80.0(40)	42.0(21)	72.0(36)	70.0(35)	78.0(39)	70.0(35)
ㅡ⇒ㅐ	-	-	-	-	-	-
ㅡ⇒ㅜ	20.0(10)	42.0(21)	26.0(13)	4.0(2)	2.0(1)	-
ㅡ⇒ㅓ	-	-	-	-	-	-
ㅡ⇒ㅣ	-	6.0(3)	-	22.0(11)	20.0(10)	28.0(14)

인식 모음	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅡ⇒ㅡ	96.0(48)	98.0(49)	96.0(48)	68.0(34)	92.0(46)	28.0(14)
ㅡ⇒ㅐ	-	-	-	30.0(15)	-	
ㅡ⇒ㅜ	2.0(1)	-	-	-	-	28.0(14)
ㅡ⇒ㅓ	-	-	-	-	-	28.0(14)
ㅡ⇒ㅣ	2.0(1)	-	-	2.0(1)	4.0(2)	6.0(3)

<표 68> 중성 'ㅡ'의 선행 자음 환경별 유사 음소(CV)

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅡ⇒ㅡ	93.0(279)	98.0(294)	96.3(289)	91.0(273)	89.3(268)	87.3(262)
ㅡ⇒ㅐ	0.3(1)	-	-	2.3(7)	5.3(16)	2.0(6)
ㅡ⇒ㅜ	4.3(13)	0.3(1)	2.0(6)	2.3(7)	2.7(8)	7.0(21)
ㅡ⇒ㅣ	-	0.7(2)	-	2.3(7)	0.7(2)	0.7(2)

인식 모음	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅡ⇒ㅡ	50.3(151)	42.7(128)	45.3(136)	54.3(163)	68.7(206)	31.7(95)
ㅡ⇒ㅐ	0.3(1)	6.0(18)	0.7(2)	2.6(8)	8.3(25)	-
ㅡ⇒ㅜ	44.7(134)	38.0(114)	45.3(136)	7.3(22)	8.0(24)	6.0(18)
ㅡ⇒ㅣ	0.3(1)	1.0(3)	0.7(2)	35.0(105)	15.0(45)	48.0(187)

인식 모음	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅡ⇒ㅡ	83.3(249)	89.0(267)	88.3(265)	84.3(253)	52.0(156)	28.3(85)
ㅡ⇒ㅐ	2.0(6)	7.7(23)	2.3(7)	10.7(32)	20.0(60)	5.4(16)
ㅡ⇒ㅜ	6.0(18)	1.3(4)	0.3(1)	2.0(6)	13.7(41)	45.0(135)
ㅡ⇒ㅣ	5.3(16)	1.7(5)	-	1.3(4)	7.3(22)	0.3(1)

<표 69> 중성 'ㅡ'의 선행 자음 환경별 유사 음소(CVC)

CV음절 중성 ‘ㅡ’의 인식률 평균은 81.4%, CVC음절 평균은 70.7%로, 음절 구조에 따른 인식률 차이가 큰 편이다. <표 68>의 관찰 결과 가장 눈에 띄는 점은 중성 ‘ㅡ’의 유사 음소가 매우 산발적으로 분포되어 있다는 것이다. 대체적으로 ‘ㅜ’나 ‘ㅣ’로 오인식되는 경우가 많은 듯하나, ‘ㅐ’나 ‘ㅓ’가 최빈도 오인식 모음인 경우도 존재한다. CV음절에서는 선행 자음이 ‘ㅁ, ㅃ, ㅍ / ㅈ, ㅊ, ㄹ’일 때, CVC음절에서는 ‘ㅁ, ㅂ, ㅃ, ㅍ / ㅈ, ㅉ, ㅊ, ㄹ’일 때 평균 이하의 인식률을 보인다.

‘ㅡ’의 오인식 모음으로는 ‘ㅣ, ㅜ, ㅐ, ㅓ’가 관찰되는데, 이 중 ‘ㅓ’로의 오인식은 CV음절에서 선행 자음이 ‘ㅁ’인 경우에 한하여 발생하는 오류이다. 이는 앞서 살핀 ‘ㅓ, ㅗ’의 경우와 마찬가지로 언어 모델의 과도 교정에서 기인된 오류로 보인다. ‘ㅡ’와의 음향적 유사성에 의해 오인식된 모음은 ‘ㅣ, ㅜ, ㅐ’인 듯하며, 선행 자음이 양순음일 때는 ‘ㅜ’로,[49] 경구개음일 때는 ‘ㅣ’로, 치조음일 때는 ‘ㅐ’로 오인식되는 경향이 나타난다. 황미경·조남민(2020)에서는 마찰음이나 파찰음이 선행할 때 ‘ㅡ’의 F2값이 단독 모음과 비교해서 상대적으로 높아지며 전설 방향으로 이동한다는 점을 밝힌 바 있는데, ‘ㅡ’의 이와 같은 특성으로 인해 마찰음·파찰음과 같은 치조·경구개 음소들이 선행할 때 전설모음인 ‘ㅣ, ㅐ’로 오인식된 것으로 보인다. 음성 인식 상황에서 연구개음은 양순음과 유사한 인식 양상을 보이기도 하는데, 선행 자음이 연구개음인 경우에는 양순음과 마찬가지로 ‘ㅡ’가 ‘ㅜ’로 오인식되는 모습이 관찰된다.

다음으로는 중성 ‘ㅣ’의 인식 오류 양상과 유사 음소에 대해 알아보도록 한다.

---

49) 양순음에 의한 모음 ‘ㅡ’의 원순화 현상에 대한 논의인 신우봉(2010)에서는 개재자음이 평음인 경우가 경음이나 격음인 경우보다 원순화가 더 잘 일어나는 환경임을 분석하고 있으나, 본 실험에서는 이와 같은 경향성은 발견되지 않았다.

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅣ⇒ㅣ	100.0(50)	94.0(47)	100.0(50)	98.0(49)	100.0(50)	98.0(49)

인식 모음	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅣ⇒ㅣ	100.0(50)	100.0(50)	100.0(50)	98.0(49)	100.0(50)	96.0(48)

인식 모음	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅣ⇒ㅣ	92.0(46)	96.0(48)	96.0(48)	98.0(49)	98.0(49)	100.0(50)

<표 70> 중성 'ㅣ'의 선행 자음 환경별 유사 음소(CV)

인식 모음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ
ㅣ⇒ㅣ	95.7(287)	93.0(279)	97.7(293)	98.0(294)	97.0(291)	96.7(290)
ㅣ⇒ㅏ	-	0.3(1)	0.3(1)	-	0.3(1)	0.3(1)

	ㅂ	ㅃ	ㅍ	ㅈ	ㅉ	ㅊ
ㅣ⇒ㅣ	98.7(296)	96.3(289)	97.0(291)	99.3(298)	94.7(284)	100.0(300)
ㅣ⇒ㅏ	-	-	0.3(1)	-	1.7(5)	

	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ	ㅁ
ㅣ⇒ㅣ	91.7(270)	94.3(283)	91.3(274)	98.3(295)	93.3(280)	99.3(298)
ㅣ⇒ㅏ	0.3(1)	-	-	-	0.7(2)	0.3(1)

<표 71> 중성 'ㅣ'의 선행 자음 환경별 유사 음소(CVC)

CV음절 중성 'ㅣ'의 인식률 평균은 98.0%, CVC음절 평균은 96.1%로 모든 단모음 중 가장 높은 인식률을 보인다. 유사 음소를 따로 설정하기에는 오류 발생 빈도가 적어서, 오인식 모음이 주로 'ㅏ'로 나타난다는 점만을 언급해 두도록 한다.

4.1.2. 이중모음

실험 대상 음절 중 이중모음을 가지는 GV, GVC, CGV, CGVC 구성의 경우에는 발음이 불가능한 음절들이 다수 존재하는바, 단모음 구성과 달리 조합 가능한 모든 음절들을 전수 조사하지는 않았다. 이로 인해 선행 자음 환경별 인식률을 단모음과

같이 정리해 보일 수는 없기에, 본 절에서는 이중모음형들을 j계 이중모음 구성과 w계 이중모음 구성으로 나누어 음절 구조별 인식률을 통해 종합적으로 제시하도록 한다. 아래의 <표 72>는 개별 이중모음들의 인식률을 음절 구조에 따라 나타낸 것이고, <표 73>은 <표 72>의 결과를 종합하여 j계 이중모음과 w계 이중모음의 음절 구조별 인식률 평균을 정리한 것이다. 마지막으로 <표 74>는 j계 이중모음과 w계 이중모음의 음절 구조별 인식률을 순서대로 나타낸 것이다.

j계 / 음절구조	ㅑ	ㅕ	ㅛ	ㅠ	ㅒ	ㅢ
GV	94.0	94.0	90.0	100.0	98.0	86.0
GVC	96.0	99.0	79.2	68.8	79.3	-
CGV	-	95.5	83.3	95.2	**30.7**	**4.0**
CGVC	-	95.0	-	88.1	-	-

w계 / 음절구조	ㅘ	ㅝ	ㅙ	ㅟ
GV	100.0	82.0	99.0	100.0
GVC	94.7	83.0	97.5	-
CGV	91.2	**19.6**	**48.5**	**52.0**
CGVC	81.2	**31.7**	**2.0**	-

<표 72> 음절 구조별 이중모음의 인식률

음절구조 / 종류	GV	GVC	CGV	CGVC	평균
j계	93.7	84.5	**61.7**	92.5	83.1
w계	95.3	91.7	**52.8**	**38.3**	69.5
평균	94.5	88.1	57.3	65.4	76.3

<표 73> j계 이중모음과 w계 이중모음의 음절 구조별 인식률 평균

j계 이중모음	GV	CGVC	GVC	CGV
w계 이중모음	GV	GVC	CGV	CGVC

<표 74> 음절 구조별 이중모음 인식률 순

<표 73>에 제시된 바와 같이 j계 이중모음의 인식률 평균은 83.1%, w계 이중모음 평균은 69.5%로, 이들을 합한 이중모음 전체의 인식률은 76.3%로 나타난다. 앞서 살펴본 전체 단모음의 인식률이 85.0%였던 것에 비하면 이중모음은 이보다 다소 낮은 인식률을 보인다. <표 74>에 따르면 j계 이중모음과 w계 이중모음 모두 음절 구조가 가장 단순한 GV구조일 때 인식률이 높게 나타났다.

이제 각 음절 구조별 인식 오류 양상을 살펴보도록 한다. 먼저 GV 구조의 j계 이중모음의 인식률 평균은 93.7%, w계 인식률 평균은 95.0%로 오류가 거의 발생하지 않았다. j계 이중모음 중에서는 'ㅢ', w계 이중모음 중에서는 'ㅝ'의 인식률이 비교적 낮은 편이며, 이 중 'ㅢ'는 'ㅓ, ㅔ, ㅜ'와 같은 단모음으로 산발적으로 오인식되고, 'ㅝ'는 주로 'ㅜ'로 오인식된다.

다음으로 살펴볼 GVC 구조의 j계 이중모음의 인식률 평균은 84.5%, w계 이중모음의 평균은 91.7%로 나타난다. j계 이중모음 중 인식률이 가장 낮은 모음은 'ㅠ'이며, 종성이 'ㅁ'(60.0%)나 'ㅇ'(24.0%)일 때 유독 낮은 인식률을 보인다. 두 경우 모두 오인식 모음은 'ㅛ'이다. 'ㅠ' 다음으로는 'ㅛ'와 'ㅐ'의 인식률이 낮게 나타나는데, 이 중 'ㅛ'는 종성이 'ㄹ'인 경우에 한하여 낮은 인식률(22.0%)을 보이며, 이때 오인식되는 모음은 'ㅠ'(22.0%)나 'ㅕ'(38.0%)이다. GVC구조의 'ㅛ', 'ㅠ'는 이처럼 특정 종성과 연쇄될 때에 인식률이 낮아지는 경향이 있으나, 'ㅐ'는 종성의 종류와 상관없이 두루 낮은 인식률을 보인다. 이때 오인식되는 모음은 'ㅠ'(10%)나 'ㅣ'(10%)이다. 한편 GVC 구조 w계 이중모음의 경우에는 'ㅝ'의 인식률이 상대적으로 낮게 나타나며 이때 'ㅝ'는 주로 'ㅘ'로 오인식된다.

다음으로 살펴볼 CGV 구조의 j계 이중모음의 인식률 평균은 61.7%, w계 이중모음의 평균은 52.8%, 이들의 합한 CGV 이중모음의 전체 평균은 57.3%로 전체 음절 구조 중 가장 낮은 인식률을 보인다. CGV 구조의 j계 이중모음은 개별 모음별 인식률 편차가 큰 편으로, 'ㅕ', 'ㅠ'의 인식률은 각각 95.5%, 95.2%로 높게 나타나나, 'ㅐ', 'ㅢ'의 인식률은 30.7%, 4.0%로 매우 낮게 나타난다. 'ㅐ'는 j가 탈락된 형태인 'ㅔ'나 'ㅐ'로 오인식되고, 'ㅢ'도 단모음 'ㅣ'로 오인식되는 경향을 보인다. 'ㅛ'도 'ㅕ'나 'ㅠ'에 비해 인식률이 낮은 편이며, 이때의 'ㅛ'는 'ㅕ'나 'ㅠ'로 오인식된다. 이는

GVC 구조 'ㅛ'의 인식 오류 양상과 동일하다.

CGV 구조 w계 이중모음의 인식률 평균은 52.8%로 낮은 편이나, 그 중 'ㅘ'의 인식률은 91.2%로 매우 높게 나타난다. 반면 'ㅝ, ㅙ, ㅚ'는 인식률이 각각 19.6%, 48.5%, 52.0%로 낮게 도출되었다. 이 중 'ㅝ'의 인식 양상을 구체적으로 살펴보면, 아래의 <표 75>와 같이 선행하는 초성 자음에 따라 인식률이 크게 차이가 난다는 점이 주목된다.

ㄱ	ㄴ	ㄷ	ㅁ	ㅈ	ㅊ	ㅎ	ㄲ	ㅆ
0	6.0	6.0	**96.0**	**56.0**	10.0	0	2.0	0

<표 75> CGV 구조 w계 이중모음 'ㅝ'의 초성별 인식률

<표 75>를 통해 확인할 수 있듯이 초성 'ㅁ'가 선행할 때(실험 음절: '뭐')의 'ㅝ'의 인식률은 96%로 높게 나타나고, 이보다는 낮지만 초성 'ㅈ'가 선행할 때(실험 음절: '줘')의 인식률도 56.0%로 여타의 경우보다 높은 모습을 보인다. 나머지 '궈, 눠, 둬, 춰, 훠, 꿔, 쒀'와 같은 실험 음절에서의 'ㅝ'의 인식률은 0%에서 최대 10%로 매우 낮게 관찰되는데, 이 음절들은 모두 단독형에서는 사용되지 않고 용언 활용형에서 출현하는 발음형이라는 공통점을 가진다. '뭐'나 '줘'도 이와 마찬가지로 활용형에서 출현하는 형태이기는 하나, '뭐'는 대명사나 감탄사로 문장 안에서 단독으로 사용되기도 하고, '줘'와 더불어 하나의 문장으로도 사용될 수도 있으므로, 단독형으로서의 출현 확률이 적은 '궈, 눠……'에 비해 인식률이 높았던 것으로 추정된다. 이렇게 본다면 'ㅝ'의 인식 오류에는 음향적 요인 외에 CGV 음절 w계 이중모음의 출현 확률과 같은 비음운론적 정보가 관여한 것으로 해석해 볼 수 있다. 같은 맥락에서 'ㅙ' 역시 초성이 'ㄲ'인 경우(실험 음절: '꽤')에는 98.0%의 높은 인식률을 보이나, 초성이 'ㅎ'인 경우(실험 음절: '홰')에는 인식률이 8%에 그친다. 이때 실험 음절 '홰'는 70%가 '회'로 오인식되는데, 이는 'ㅙ'와 'ㅚ'의 음향적 유사성과 함께 '홰'와 '회'의 사용 빈도에 따른 출현 확률이 영향을 준 결과로 판단된다. 'ㅝ'는 대체로 'ㅗ, ㅜ, ㅓ'로 오인식되고, 'ㅙ'는 'ㅚ'나 'ㅐ', 'ㅟ'는 주로 'ㅣ'로 오인식된다.

다음으로 살펴볼 CGVC 구조의 경우, j계 이중모음의 인식률 평균은 92.5%, w계

이중모음의 평균은 38.3%로 j계와 w계의 인식률 차이가 가장 크게 나타난다. j계 이중모음 중 'ㅠ'(88.1%)는 'ㅕ(95.0%)'에 비해 인식률이 낮은 편이며 이때의 'ㅠ'는 주로 'ㅛ'로 오인식된다. 한편 w계 이중모음은 'ㅘ'(81.2%)에 비해 'ㅝ'(31.7)와 'ㅙ'(2.0)의 인식률이 매우 낮게 나타나는데, 앞서 CGV 구조에서는 중성이 'ㅝ'와 'ㅙ'인 실험 음절들이 활용형에서만 출현하는 형태이거나 사용 빈도가 적어 인식률이 낮은 것으로 판단되었으나, CGVC 구조의 실험 음절들은 실제 발화에서 사용될 만한 대상이었으나 인식률이 낮다는 점에서 차이를 가진다. CGVC 구조의 실험 대상 음절 '권, 궐, 뭔, 뭘, 훤, 훨, 꿩' 중 특히 '꿩'은 98%가 '꽝'으로 오인식되었는데, 이는 '꿩'과 '꽝'의 사용 빈도나 출현 확률만으로는 설명되기 어려운 오류로 보인다. CGVC 구조의 'ㅝ'는 초성에 따라 'ㅜ, ㅗ, ㅓ, ㅘ'로 산발적으로 오인식된다.

지금까지 살펴본 이중모음의 인식 오류 양상을 종합해 본다면, 이중모음은 단모음에 비해 낮은 인식률을 보였으며, 구조가 가장 단순한 GV음절의 인식률이 가장 높게 나타났다. j계 이중모음은 전반적으로 초성의 존재 여부가 인식 오류에 크게 영향을 주지 않는 듯하나, 'ㅐ'와 'ㅢ'는 초성이 있을 때의 인식률(각 30.7%, 4.0%)과 초성이 없을 때의 인식률(88.7%, 86.0%)이 큰 차이를 보인다. 이러한 오류는 'ㅐ'와 'ㅢ'가 자연 발화에서도 초성이 있는 경우에는 흔히 단모음으로 발음된다는 점을 고려한다면(실험 음절: '계, 폐, 혜, 희')[50], 'ㅐ'와 'ㅢ'의 실제 발화 특성으로 인해 발생한 오류로 간주할 수 있을 듯하다.

한편 w계 이중모음은 초성의 존재 여부가 음성 인식에 전반적으로 큰 영향을 미치는 것으로 보인다. 특히 'ㅝ, ㅙ, ㅟ'는 초성이 있을 때의 인식률(각 25.7%, 25.3%, 52.0%)과, 초성이 없을 때의 인식률(각 82.5%, 98.3%, 100.0%)이 크게 차이가 난다. 자연 발화에서는 w계 이중모음의 활음 w가 곧잘 탈락하여 발음되기도 하지만, 단음절 단독 발화를 대상으로 하는 본고의 음성 인식 실험에서는 w를 탈락시켜 발음하는 경우는 관찰되지 않았다. 다만 w가 약화된 발화음은 다소 관찰되었는데 초성 뒤에서 비교적 약화된 w를 정확히 인식하지 못하고 단모음으로 오인식하거나, w계

50) 본 연구에서는 'ㅐ'와 'ㅔ', 'ㅒ'와 'ㅖ'를 변별하지 않고 모두 'ㅐ', 'ㅒ'로 표기하여 실험을 진행하였으나, 이 경우에는 '개, 패, 해'보다 표기상 익숙한 '계, 폐, 혜'를 실험 대상 음절로 사용하였다.

이중모음들의 사용 빈도에 근거한 단음절 발화 출현 확률을 토대로, 정확히 인식된 w계 이중모음을 단모음이나 여타의 이중모음으로 교정한 데서 기인한 것으로 추정된다.

4.2. 종합적 검토

본 절에서는 지금까지 살펴본 중성의 음성 인식 오류 양상에 대해 종합적으로 정리해 보고자 한다. 먼저 중성 단모음에서는 선행 자음의 조음 위치의 영향으로 발생하는 인식 오류와, 성별에 따른 모음의 음성적·조음 공간적 차이에 의해 발생하는 인식 오류가 발견되었다. 조음 위치와 관련된 인식 오류는 주로 모음 'ㅡ'에서 나타나는데, 'ㅡ'는 선행 자음이 양순음인 경우에는 원순모음 'ㅜ'로, 마찰음·파찰음인 경우에는 경구개 모음 'ㅣ'로, 치조음인 경우에는 전설모음 'ㅐ'로 오인식된다. 'ㅗ'와 'ㅜ'는 현대 한국어에서 세대·성별에 따른 조음 공간의 차이가 큰 모음으로 간주되며, 이 중 'ㅗ'의 'ㅜ'로의 오인식은 여성에게서 두드러지게 나타나, 성별에 따른 원순모음의 조음 차이에 따른 오류로 판단된다. 'ㅓ'의 'ㅏ'로의 오인식도 여성에게서 더욱 빈번하게 관찰되는 오류이며, 이는 여성과 남성이 가진 'ㅓ'와 'ㅏ'의 F1값의 차이가 음성 인식 상황에 반영되지 못한 결과로 보인다.

단모음 중 'ㅏ'와 'ㅓ', 'ㅗ'와 'ㅜ'는 서로를 유사 음소로 가지는데, 이러한 양상은 이중모음의 인식 상황에서도 관찰되어 'ㅘ'와 'ㅝ', 'ㅛ'와 'ㅠ'가 서로 오인식되는 모습을 보인다. 단모음 'ㅐ'는 활음이 첨가된 'ㅒ'나 'ㅙ'로 오인식되고, 이중모음 'ㅒ, ㅙ'는 반대로 활음이 탈락한 'ㅐ'로 오인식된다. 이중모음의 인식에 있어서 j계 이중모음은 자연 발화에서의 발음 특성이, w계 이중모음은 초성의 존재 여부와 함께 사용 빈도에 근거한 출현 확률이 인식에 영향을 준다.

5. 종성의 음성 인식 오류 양상 및 유사 음소

5.1. 오류 양상 및 유사 음소

본 장에서는 종성 자음들의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다. 종성의 인식률은 선행 모음을 기준으로 나누어 측정하였으며, 개별 종성들의 인식 오류 결과를 구체적으로 살피기에 앞서 먼저 보일 것은 음절 구조별 종성의 전반적인 인식 결과이다.[51]

종성 / 음절구조	파열음		비음			유음	평균
	ㄱ	ㅂ	ㄴ	ㅁ	ㅇ	ㄹ	
VC	41.1	66.0	84.6	88.6	88.0	96.3	77.4
CVC	60.4	52.0	79.9	83.2	86.4	93.9	76.0
총합	50.8	59.0	82.3	85.9	87.2	95.1	

<표 76> 종성의 음절 구조별 인식률

<표 76>에 제시된 바와 같이 VC음절의 종성 인식률은 평균 77.4%, CVC음절의 종성 인식률 평균은 76.0%로 도출되어, 종성의 인식률은 음절 구조와 상관없이 거의 유사하게 나타난다. <표 76>의 인식률을 음절 구조별로 인식률 순서에 따라 다시

51) 종성 'ㄷ'의 경우, 'ㄷ'를 종성으로 가지는 단음절 어형이 극히 드물어 주로 음절말 평파열음화를 겪은 음소들이 종성에서 [ㄷ]로 실현된다. 이로 인해 종성 'ㄷ'의 음성 인식 실험 결과, 종성 'ㄷ'를 'ㄷ'로 인식한 CVC음절 및 VC음절 전체 인식률 평균이 0.9%에 그치는 것으로 나타났다. 대부분의 'ㄷ' 종성 실험 대상 음절들은 'ㄷ'가 아닌 'ㅅ, ㅈ, ㅊ, ㅌ……' 등으로 인식되었는데, 이들은 애초에 종성에서 발음될 수 없는 음소들이므로 종성 [ㄷ]와의 음향적 유사성에 의해 오인된 것으로 볼 수 없다. 따라서 이는 발화된 음성을 어떤 음소로 인식하는지, 특정 음성을 특정 음소로 잘못 인식하는 경향이 있다면 거기에는 어떠한 음운론적 요인이 관여하는지 파악하고자 하는 본고의 연구 목적에서 벗어나는 경우라 판단하여, 'ㄷ' 종성의 인식 결과를 <표 76>에 함께 제시하는 대신, 아래의 표를 통해 대략적인 인식 경향을 정리해 보이고자 한다.

정리해 보이면 아래의 <표 77>과 같다.

음절구조	인식률					
VC	ㄹ	ㅁ	ㅇ	ㄴ	ㅂ	ㄱ
VCV	ㄹ	ㅇ	ㅁ	ㄴ	ㄱ	ㅂ

<표 77> 종성의 음절 구조별 인식률 순

종성의 인식률은 음절 구조와 상관없이 '유음 > 비음 > 파열음' 순으로 낮아진다. 비음들 간의 인식률 차는 크지 않으나, 'ㄴ'의 인식률이 'ㅁ, ㅇ'에 비해 낮은 편이다. 파열음 'ㄱ'와 'ㅂ'의 인식률은 50%대로 공명음들에 비해 인식률이 떨어진다. 초성 및 중성에서와 달리 종성 자음은 다른 자음으로 오인식되는 오류와 함께 자음의 존재 자체가 인식되지 않는 오류가 빈번하게 관찰된다. 이하에서는 종성 자음을 장애음과 공명음으로 나누어 구체적인 인식 오류 양상을 서술한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
VC	ㄷ⇒ㄷ	-	-	-	-	-	-	-
	ㄷ⇒ㅅ	100.0(50)	-	90.0(45)	80.0(40)	100.0(50)	40.0(20)	90.0(45)
	ㄷ⇒ㅂ	-	100.0(50)	-	-	-	-	-
	ㄷ⇒ㅌ	-	-	-	-	-	60.0(30)	-
	ㄷ⇒ㄱ	-	-	-	-	-	-	10.0(5)
	ㄷ⇒∅	-	-	10.0(5)	20.0(10)	-	-	-
CVC	ㄷ⇒ㄷ	-	-	-	1.8(16)	-	-	-
	ㄷ⇒ㅅ	83.0(747)	76.6(689)	67.1(604)	64.8(583)	63.6(572)	40.7(366)	14.8(133)
	ㄷ⇒ㅂ	6.5(58)	11.2(101)	4.1(37)	4.1(37)	12.3(111)	6.5(58)	36.4(328)
	ㄷ⇒ㅌ	3.5(32)	1.2(11)	4.1(37)	-	4.7(42)	31.7(286)	0.6(5)
	ㄷ⇒ㄱ	1.8(16)	7.1(63)	4.7(42)	3.5(32)	7.6(69)	12.9(116)	42.9(386)
	ㄷ⇒ㅊ	-	-	0.6(5)	14.1(127)	5.9(53)	-	0.6(5)
	ㄷ⇒∅	2.9(26)	4.2(38)	18.2(164)	10.0(90)	3.5(32)	3.5(32)	4.6(41)

VC음절의 'ㄷ'의 인식률은 0%, CVC음절의 인식률은 1.8%이며, 표에 제시된 바와 같이 두 음절 구조 모두 'ㅅ'로 오인식되는 비율이 높게 나타난다.

5.1.1. 장애음

본 절에서는 음성 인식 오류 실험을 통해 분석된 종성 파열음의 음성 인식 오류 양상을 살피고, 선행 모음 환경별 유사 음소에 대해 알아보도록 한다. 먼저 제시할 것은 종성 파열음 'ㄱ, ㅂ'의 전체 인식 결과이다.

음절구조	후행모음 \ 종성	ㄱ	ㅂ	평균
VC	ㅏ	36.0(18)	92.0(46)	64.0
	ㅐ	58.0(29)	96.0(48)	77.0
	ㅓ	12.0(6)	86.0(43)	49.0
	ㅗ	60.0(30)	18.0(9)	39.0
	ㅜ	56.0(28)	10.0(5)	33.0
	ㅡ	26.0(13)	76.0(38)	51.0
	ㅣ	40.0(20)	84.0(42)	62.0
인식률 총 평균		41.1	66.0	
		53.6%		
CVC	ㅏ	53.1(478)	76.7(690)	64.9
	ㅐ	67.7(609)	77.9(701)	72.8
	ㅓ	52.4(472)	60.9(548)	56.7
	ㅗ	67.3(606)	11.9(107)	39.6
	ㅜ	75.0(675)	12.7(114)	43.9
	ㅡ	41.3(372)	40.4(364)	40.9
	ㅣ	66.2(596)	83.2(749)	74.7
인식률 총 평균		60.4	52.0	
		56.2%		

<표 78> 종성 파열음 인식률

<표 78>을 바탕으로 도출한 VC음절과 CVC음절 종성 파열음의 전체 인식률은 54.9%, 그 중 'ㄱ' 종성의 총 인식률은 53.6%, 'ㅂ' 종성의 총 인식률은 56.2%로 계산된다. <표 78>에서 눈에 띄는 점은 원순모음 'ㅗ, ㅜ' 뒤에서의 종성 'ㅂ'의 인식률이다. 'ㅗ, ㅜ'에 후행하는 'ㅂ'는 10%대의 인식률을 보이는데, 음성대화형 어시스턴트가 이 환경에서는 종성 'ㅂ'를 거의 인식해 내지 못하는 것으로 보아도 될 듯하다. 이제 각각의 종성 파열음들을 어떤 자음으로 오인식하는지 살펴보도록 한다. 아

래의 <표 79>는 종성 ‘ㄱ’의 선행 모음별 인식 오류 양상과 유사 음소를 정리한 것이다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
VC	ㄱ⇒ㄱ	36.0(18)	58.0(29)	12.0(6)	60.0(30)	56.0(28)	26.0(13)	40.0(20)
	ㄱ⇒ㅂ	30.0(15)	36.0(18)	18.0(9)	-	2.0(1)	10.0(5)	-
	ㄱ⇒ㅇ	-	-	-	4.0(2)	-	52.0(26)	52.0(26)
	ㄱ⇒∅	28.0(14)	6.0(3)	70.0(35)	36.0(18)	40.0(20)	8.0(4)	-
CVC	ㄱ⇒ㄱ	53.1(478)	67.7(609)	52.4(472)	67.3(606)	75.0(675)	41.3(372)	66.2(596)
	ㄱ⇒ㅂ	19.0(171)	11.7(105)	8.4(76)	1.1(10)	0.7(6)	8.1(73)	7.2(65)
	ㄱ⇒∅	19.4(175)	13.9(125)	38.0(342)	30.6(275)	22.9(206)	21.4(193)	21.0(189)

<표 79> 종성 ‘ㄱ’의 선행 모음별 인식률

종성 ‘ㄱ’에서 가장 빈번하게 발생하는 인식 오류는 종성의 존재가 인식되지 않는 오류이다. VC음절과 CVC음절에서 모두 종성이 없는 개음절 구조로 인식하는 빈도가 가장 높게 나타난다. 파열음은 종성에서 불파음으로 실현되어 어말 평파열음의 스펙트로그램은 묵음에 가까운 형태로 나타난다. 초성 파열음의 경우에는 스펙트로그램상 파형이 갑자기 시작되어 나타나기 때문에 파열음 자체가 없는 것으로 인식되기는 어려울 것이나, 종성 파열음과 같이 파형이 갑작스레 끝날 경우에는 종성 없이 모음만으로 발화가 종료된 것으로 오인될 가능성이 있다. 이와 같은 이유로 <표 79>에 제시된 바와 같이 종성 ‘ㄱ’의 존재 자체가 인식되지 않는 오류가 발생한 것으로 판단된다.

한편 종성 ‘ㄱ’를 다른 음소로 인식하는 경우, 이때 오인식되는 자음은 ‘ㅂ’와 ‘ㅇ’이다. 앞서 살펴본 초성의 인식 결과에서는 연구개음 계열의 ‘ㄱ, ㄲ, ㅋ’가 양순음 계열로 오인식되는 오류가 거의 발생하지 않았으나, 종성에서는 ‘ㄱ’가 ‘ㅂ’로 오인식되는 경우가 다수 관찰된다. 단 이러한 오인식이 원순모음 ‘ㅗ, ㅜ’가 선행할 때는 거의 나타나지 않는다. 종성 ‘ㄱ’를 ‘ㅂ’로 오인식하는 비율보다 ‘음소가 없는 것’으로 오인식하는 비율이 더 높으나, 유사 ‘음소’로서의 오인식률은 ‘ㅂ’가 가장 높으므로

종성 ‘ㄱ’의 유사 음소는 ‘ㅂ’로 설정한다.

다음으로는 종성 ‘ㅂ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
VC	ㅂ⇒ㅂ	92.0(46)	96.0(48)	86.0(43)	18.0(9)	10.0(5)	76.0(38)	84.0(42)
	ㅂ⇒ㄱ	-	-	-	32.0(16)	46.0(23)	-	-
	ㅂ⇒ㅇ	-	-	-	-	6.0(3)	20.0(10)	-
	ㅂ⇒∅	8.0(4)	2.0(1)	14.0(7)	36.0(18)	32.0(16)	-	16.0(8)
CVC	ㅂ⇒ㅂ	76.7(690)	77.9(701)	60.9(548)	11.9(107)	12.7(114)	40.4(364)	83.2(749)
	ㅂ⇒ㄱ	9.3(84)	5.3(48)	17.0(153)	43.3(390)	52.8(475)	18.1(163)	1.7(15)
	ㅂ⇒∅	7.6(68)	10.2(92)	21.2(191)	40.8(367)	25.1(226)	9.3(84)	11.8(106)

<표 80> 종성 ‘ㅂ’의 선행 모음별 인식률

종성 ‘ㅂ’의 VC음절의 인식률 평균은 66.0%, CVC음절의 인식률 평균은 52.0%, 이 둘을 합한 전체 인식률 평균은 59.0%로 종성 자음들 중 가장 낮은 인식률을 보인다. ‘ㅂ’는 선행 모음이 ‘ㅗ, ㅜ, ㅡ’일 때 인식률이 평균 이하로 나타나며, 이때는 ‘ㅂ’가 ‘ㄱ’로 오인식되거나 종성의 존재 자체가 인식되지 않는 오류가 발생한다. ‘ㅂ’는 후설원순모음이 인접하여 원순성(혹은 순음성)이 중복될 때나, 양순음과의 연쇄가 제한된 ‘ㅡ’가 선행할 때 인식률이 낮아지며, 유사 음소는 ‘ㄱ’이다.

지금까지 살펴본 종성 파열음들의 음성 인식 오류 양상을 종합하여 정리하면 다음과 같다. 종성 ‘ㄱ’의 유사 음소는 ‘ㅂ’, 종성 ‘ㅂ’의 유사 음소는 ‘ㄱ’인 것으로 판단되며, 초성에서와 달리 종성 파열음은 음소의 존재 자체가 인식되지 않는 경우가 나타난다는 점이 특징적이다. 선행 모음이 원순모음일 때 ‘ㄱ’가 ‘ㅂ’로 오인식되는 비율은 0~2%대에 그칠 뿐이지만, ‘ㅂ’는 선행 모음이 원순모음일 때 가장 높은 비율로(30~50%) ‘ㄱ’로 오인식된다.

5.1.2. 공명음

본 절에서는 음성 인식 오류 확인 실험 결과를 통해 분석된 종성 공명음들의 음성 인식 오류 양상을 살피고, 선행 모음 환경별 유사 음소에 대해 알아보도록 한다. 먼저 제시할 것은 종성 공명음 'ㄴ, ㅁ, ㅇ, ㄹ'의 전체 인식 결과이다.

음절구조	선행모음 \ 종성	ㄴ	ㅁ	ㅇ	ㄹ	평균
VC	ㅏ	98.0(49)	94.0(47)	94.0(47)	100.0(50)	96.5
	ㅐ	80.0(40)	96.0(48)	98.0(49)	90.0(45)	91.0
	ㅓ	96.0(48)	90.0(45)	86.0(43)	98.0(49)	92.5
	ㅗ	100.0(50)	92.0(46)	68.0(34)	100.0(50)	90.0
	ㅜ	90.0(45)	66.0(33)	88.0(44)	100.0(50)	86.0
	ㅡ	70.0(35)	84.0(42)	96.0(48)	86.0(43)	84.0
	ㅣ	58.0(29)	98.0(49)	86.0(43)	100.0(50)	85.5
인식률 총 평균		84.6	88.6	88.0	96.3	
		89.4%				
CVC	ㅏ	88.6(797)	89.2(803)	92.0(828)	99.6(896)	92.4
	ㅐ	84.7(762)	86.8(781)	93.8(844)	81.2(731)	86.6
	ㅓ	88.6(797)	88.1(793)	78.1(703)	99.1(892)	88.5
	ㅗ	79.4(715)	76.0(684)	90.4(814)	96.9(872)	85.7
	ㅜ	88.6(797)	72.7(654)	87.9(791)	97.6(878)	86.7
	ㅡ	75.8(682)	80.3(723)	72.7(654)	90.7(816)	79.9
	ㅣ	53.7(483)	89.6(806)	90.0(810)	92.1(829)	81.4
인식률 총 평균		79.9	83.2	86.4	93.9	
		85.9%				

<표 81> 종성 공명음 인식률

<표 81>을 바탕으로 도출한 VC음절과 CVC음절 종성 공명음의 전체 인식률은 87.6%로 계산되며, 유음 'ㄹ'의 인식률이 가장 높고, 비음들 중에서는 ㅇ > ㅁ > ㄴ 순으로 인식률이 낮아진다. 종성 파열음은 선행 모음이 'ㅗ, ㅜ'일 때 인식 오류가 가장 많이 발생하였으나, 공명음은 선행 모음이 주로 'ㅡ, ㅣ'일 때 낮은 인식률을 보인다. 이제 각각의 종성 공명음들을 어떤 자음으로 오인식하는지 살펴보도록 한다.

먼저 살필 대상은 'ㄴ'이며, 아래의 <표 82>는 종성 'ㄴ'의 선행 모음별 인식 오류 양상과 유사 음소를 정리한 것이다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
VC	ㄴ⇒ㄴ	98.0(49)	80.0(40)	96.0(48)	100.0(50)	90.0(45)	70.0(35)	58.0(29)
	ㄴ⇒ㅁ	2.0(1)	18.0(9)	4.0(2)	-	-	2.0(1)	26.0(13)
	ㄴ⇒ㅇ	-	2.0(1)	-	-	10.0(5)	28.0(14)	14.0(7)
CVC	ㄴ⇒ㄴ	88.6(797)	84.7(762)	88.6(797)	79.4(715)	88.6(797)	75.8(682)	53.7(483)
	ㄴ⇒ㅁ	6.4(58)	10.0(90)	8.8(79)	10.2(92)	6.0(54)	14.2(128)	23.9(215)
	ㄴ⇒ㅇ	3.1(28)	3.7(33)	0.6(5)	3.8(34)	3.4(31)	7.3(66)	19.6(173)

<표 82> 종성 'ㄴ'의 선행 모음별 인식률

종성 'ㄴ'의 VC음절의 인식률 평균은 84.6%, CVC음절의 인식률 평균은 79.9%이며, 이 둘을 합한 전체 인식률 평균은 82.3%로 종성 공명음 중에서는 가장 낮은 인식률을 보인다. 'ㄴ'는 선행 모음이 주로 'ㅡ, ㅣ'일 때 인식률이 평균 이하로 나타나며 같은 비음 계열인 'ㅁ'나 'ㅇ'로 오인식된다.

다음으로 종성 'ㅁ'의 인식 오류 양상과 유사 음소에 대해 알아보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
VC	ㅁ⇒ㅁ	94.0(47)	96.0(48)	90.0(45)	92.0(46)	66.0(33)	84.0(42)	98.0(49)
	ㅁ⇒ㅇ	2.0(1)	-	2.0(1)	4.0(2)	32.0(16)	16.0(8)	2.0(1)
	ㅁ⇒ㄴ	4.0(2)	4.0(2)	2.0(1)	2.0(1)	-	-	-
CVC	ㅁ⇒ㅁ	89.2(803)	86.8(781)	88.1(793)	76.0(684)	72.7(654)	80.3(723)	89.6(806)
	ㅁ⇒ㅇ	3.8(34)	1.4(13)	3.8(34)	15.6(140)	19.0(171)	6.6(59)	4.3(39)

<표 83> 종성 'ㅁ'의 선행 모음별 인식률

종성 'ㅁ'의 VC음절의 인식률 평균은 88.6%, CVC음절의 인식률 평균은 83.2%이며,

이 둘을 합한 인식률 평균은 86.0%로 계산된다. ‘ㅁ’는 후설 원순모음이 인접하여 원순성(혹은 순음성)이 중복될 때나, 양순음과의 연쇄가 제한된 ‘ㅡ’가 선행할 때 인식률이 낮아지며, 유사 음소는 ‘ㅇ’이다. ‘ㅁ’의 인식 오류는 앞서 살핀 같은 양순음 계열의 종성인 ‘ㅂ’의 오류 양상과 동일한 경향을 보인다.[52]

다음으로는 종성 ‘ㅇ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

		후행 모음						
음절 구조	인식 자음	ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
VC	ㅇ⇒ㅇ	94.0(47)	98.0(49)	86.0(43)	68.0(34)	88.0(44)	96.0(48)	86.0(43)
	ㅇ⇒ㅁ	4.0(2)	-	2.0(1)	28.0(14)	12.0(6)	4.0(2)	4.0(2)
	ㅇ⇒∅	-	2.0(1)	12.0(6)	4.0(2)	-	-	2.0(1)
CVC	ㅇ⇒ㅇ	92.0(828)	93.8(844)	78.1(703)	90.4(814)	87.9(791)	72.7(654)	90.0(810)
	ㅇ⇒ㅁ	2.2(20)	-	14.0(126)	7.7(69)	9.2(83)	11.3(102)	4.2(38)

<표 84> 종성 ‘ㅇ’의 선행 모음별 인식률

종성 ‘ㅇ’의 VC음절의 인식률 평균은 88.0%, CVC음절의 인식률 평균은 86.4%이며, 모음 ‘ㅓ, ㅗ, ㅡ’가 선행할 때의 인식률이 평균보다 낮게 나타난다. <표 84>의 관찰 결과 ‘ㅇ’의 유사 음소는 ‘ㅁ’인 것으로 보인다.

다음으로는 종성 ‘ㄹ’의 인식 오류 양상과 유사 음소에 대해 살펴보도록 한다.

52) 한편 양순음이 초성에 위치한 경우에는 후행하는 모음 ‘ㅡ’가 정확히 인식되지 못하고(‘ㅁ’ 뒤 ‘ㅡ’의 인식률 28.2%) ‘ㅜ’로 오인식되는 경우가 많았으나(45%), 종성 위치의 양순음은 선행 모음 ‘ㅡ’ 인식에는 영향을 미치지 않는 것으로 보인다. 이 경우 종성 ‘ㅁ’만 주로 ‘ㅇ’로 오인식될 뿐, 모음 ‘ㅡ’는 그대로 인식되었다. 이는 종성 ‘ㅂ’의 경우도 마찬가지이다.

음절 구조	인식 자음	후행 모음						
		ㅏ	ㅐ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ
VC	ㄹ⇒ㄹ	100.0(50)	90.0(45)	98.0(49)	100.0(50)	100.0(50)	86.0(43)	100.0(50)
	ㄹ⇒ㅇ	-	4.0(2)	-	-	-	10.0(5)	-
CVC	ㄹ⇒ㄹ	99.6(896)	81.2(731)	99.1(892)	96.9(872)	97.6(878)	90.7(816)	92.1(829)
	ㄹ⇒∅	0.2(2)	13.7(123)	0.8(7)	2.4(22)	0.7(6)	6.7(60)	5.9(53)

<표 85> 종성 ‘ㄹ’의 선행 모음별 인식률

종성 ‘ㄹ’의 VC음절의 인식률 평균은 96.3%, CVC음절의 인식률 평균은 93.9%로, 종성 중 가장 높은 인식률을 보인다. ‘ㄹ’는 발생하는 오류 빈도가 높지 않으나, 모음 ‘ㅐ, ㅡ, ㅣ’가 선행할 때의 인식률은 평균보다 낮게 나타난다. 이때 종성의 존재가 인지된 경우에는 ‘ㅇ’로 오인식되기도 하나, 대부분의 경우에는 종성이 없는 것으로 오인식된다. 종성 ‘ㄹ’는 음소의 존재 자체가 인식되지 않는 경우가 다른 공명자음들에 비해 빈번하게 관찰된다.

5.2. 종합적 검토

본 절에서는 종성의 음성 인식 오류 양상에 대해 종합 정리하며, 종성의 인식 결과를 초성에서의 인식 결과와 함께 대비해 보고자 한다. 종성 파열음은 초성 파열음들과 마찬가지로 공명음에 비해 인식률이 낮게 나타났다. 초성의 경우 연구개음 > 치조음 > 양순음 순으로 인식률이 낮아졌는데, 종성에서도 양순음의 인식률이 연구개음에 비해 낮은 것으로 확인된다. 초성에서 ‘ㄱ, ㅂ’의 유사 음소는 같은 계열의 격음인 ‘ㅋ, ㅍ’로 나타났으나, 종성의 ‘ㄱ, ㅂ’는 서로를 유사 음소로 가진다.

공명음은 초성에서는 비음 ‘ㄴ’의 인식률이 가장 높고, 유음 ‘ㄹ’의 인식률이 가장 낮은 것으로 관찰되었으나, 종성에서는 이와 반대로 ‘ㄹ’의 인식률이 가장 높고 ‘ㄴ’의 인식률이 가장 낮게 도출되는 결과를 보였다. 이는 ‘ㄴ’보다는 ‘ㄹ’의 인식 양상과 관련된 결과로 여겨지는데, 전술한 바와 같이 어두에서 주로 탄설음으로 실현되는

‘ㄹ’는 스펙트로그램상 음향 에너지가 거의 관찰되지 않는 짧은 흰 공간으로 나타나, 음성 인식 상황에서 참고할 만한 음향 단서가 매우 적다. 이와 더불어 ‘ㄹ’는 어두에서 분포가 제한되는 특성을 가지는바, 어두 ‘ㄹ’의 이러한 특성들로 인해 초성에서의 ‘ㄹ’의 인식 오류가 종성 ‘ㄹ’에 비해 다수 발생한 것으로 이해할 수 있을 듯하다.

6. 음성 인식 오류의 음운론적 유발 요인 및 유형

근래의 음성 인식 알고리즘은 딥러닝 학습을 바탕으로 각 음소들의 전형적인 음향 특징을 수집하고, 음성 신호가 주어지면 감지된 음성 신호가 데이터베이스 내의 어떤 음소와 가장 가까운지에 대한 확률적 판단을 통해 음소를 할당한다. 음성 인식 오류는 알고리즘이 정해둔 각 음소들의 전형적인 음향 특징 범위를 벗어나거나, 그 경계에 있는 음이 감지되었을 때 발생하는 듯하다. 그리고 이러한 경우 음소를 정의하는 여러 음향 특성 중 어떠한 특성이 임계점을 벗어났는지에 따라 오인식되는 음소가 결정될 것이다. 이때 발화음의 특성이 해당 음소의 임계점을 벗어나게 만드는 데에는 특정 음운론적 환경이 관여할 수도 있고, 특정 조건 없이 음향적 특성이 유사한 음소들 간의 오류가 인식을 유발할 수도 있다. 본고에서는 이 중 전자와 같은 요인에 의해 발생하는 음성 인식 오류를 '음성적 변이에 의한 인식 오류' 유형으로, 후자를 '음향적 특성에 의한 인식 오류' 유형으로 나누고자 한다. 이와 같은 요인 외에 남성과 여성의 발화가 가진 음성적·음운론적 특성이 음성 인식 오류에 영향을 주는 경우는, '성별 요인에 의한 인식 오류' 유형으로 설정하였다.

6.1. 음성적 변이에 의한 인식 오류

'음성적 변이에 의한 인식 오류'는 특정 음운론적 환경에서 음성적 변이가 실현되고, 음성 인식 알고리즘이 이를 전형적인 음소의 인식 범위를 벗어난다고 판단함으로써 발생하는 오류 유형이다. 주지하다시피 한국어 자음 중에는 'ㅣ' 모음 앞에서 구개음을 변이음으로 가지는 음소들이 존재한다. 이러한 자음들이 'ㅣ' 모음과 연쇄할 때에 여타의 모음들에 비해 음성 인식 오류가 증가하는 경향이 발견되는데, 이와 같은 변이음의 실현 환경에서 보이는 음성적 변이가 음성 인식 오류에 영향을 미치는 경우를 '음성적 변이에 의한 인식 오류' 유형으로 설정한다. 경구개 모음 앞에서

경구개 변이음이 실현되는 경우 외에도, 치조 마찰음·파찰음과 ‘ㅡ’가 연쇄될 때나, 양순음과 ‘ㅡ’가 연쇄하는 경우에도 선행 자음의 음향적·조음적 특성으로 인해 ‘ㅡ’의 음성 실현형이 여타의 경우와 달라지는데, 이러한 ‘ㅡ’의 비전형적인 음성 실현이 인식 오류를 유발하는 예들도, 음성적 변이에 의한 인식 오류 유형으로 간주하였다.

1) 경구개 모음 앞 경구개 변이음의 실현

현대 한국어의 ‘ㄴ, ㄹ, ㅅ(ㅆ), ㅎ’은 ‘ㅣ’나 ‘j’ 앞에서 구개음을 변이음으로 가진다. 이를 음운 현상으로는 ‘음성적 구개음화’라 할 수 있는데, 음성적 구개음화는 비구개음 자음과 구개음 모음이 만났을 때 구개 모음에 동화되어 비구개 자음이 구개성 자음으로 실현되는 결과를 가져온다. 이들 자음의 음성 실현형과 음향적 특징에 대해 하나씩 살펴보면, 우선 ‘ㄹ’는 ‘ㅣ’ 앞에서 경구개 유음 [ɾʲ]로 실현되는데, 어두 유음에 관한 실험 음성학적 논의인 조남민(2015)에 의하면, 경구개 부근에서 발음되는 [ɾʲ]는 후행하는 ‘ㅣ’ 모음과 유사한 포먼트 유형을 보이며, 치조에서 발음되는 [l]에 비해 포먼트가 덜 뚜렷하게 나타나는 특징을 가진다고 한다. 이 논의에서는 ‘[ɾʲ]’의 조음 시 중앙이 폐쇄되고 혀의 양 측면으로 기류가 통과하면서 공명을 일으키는 설측음의 특성상 폐쇄가 구강 깊숙한 지점에서 일어남으로써 공명도가 낮아져 상대적으로 포먼트가 덜 뚜렷하게 나타나는 것으로 해석한다. ‘ㄹ+ㅣ’ 연쇄에서 나타나는 ‘ㄹ’의 음성 인식 오류는 이와 같은 ‘ㅣ’ 앞의 ‘ㄹ’의 경구개 변이음 [ɾʲ]이 가진 음향 특성을 ‘ㄹ’의 범주로 판단하지 못한 데에서 기인한 것으로 보인다.

‘ㅅ’ 역시 모음 ‘ㅣ’ 앞에서 경구개 마찰음으로 실현되는데, 신지영(2014: 242)에서는 이와 관련하여 ‘ㅅ’는 후행 모음에 따라 스펙트로그램에서 관찰되는 소음의 에너지 분포가 확연히 차이가 나며, 전강의 크기가 큰 ‘ㅣ’ 앞의 ‘ㅅ’는 여타의 모음에 비해 소음부 에너지가 시작하는 주파수 대역 및 가장 강한 에너지를 보이는 주파수 대역이 훨씬 낮게 나타난다고 기술한 바 있다. 이호영·지민제·김영송(1993)에서도 이와 마찬가지로 ‘ㅣ’ 앞의 ‘ㅅ, ㅆ’가 다른 모음 앞에서보다 조음점이 뒤로 이동하여 좁힘점과 입술과의 길이가 길어져, 다른 모음 앞에서보다 중심 공명 주파수대가 낮

아지므로 스펙트로그램상 낮은 주파수대역에도 에너지가 분포하여 중심에너지의 주파수대역이 약간 하강하게 됨을 지적하였다. 두 논의에서 공통적으로 지적한 'ㅣ' 앞에서 하강한 'ㅅ(ㅆ)'의 주파수 대역을 'ㅅ'의 음향 특성 범주로 판단하지 않을 경우 음성 인식 오류가 유발될 수 있다.

'ㅎ'도 이와 마찬가지로 'ㅣ'가 후행할 때 여타의 모음과 다른 음향 특성을 보인다. 이호영·지민제·김영송(1993)에서는 이와 관련하여 'ㅎ'와 'ㅣ'의 연쇄 시 전설이 경구개 쪽으로 접근하여 강한 마찰을 초래하기 때문에 비교적 넓은 주파수대에 걸쳐 강한 에너지가 분포하고, 좁힘점과 입술 사이의 길이가 다른 모음 앞에서보다 훨씬 짧아 고주파수대에도 강한 에너지가 분포하게 됨을 지적한 바 있다.

한편 'ㅣ' 모음 앞에서의 'ㄴ'의 변이음에 대해서는 이를 진정한 경구개 변이음으로 판단할 수 있는지에 대한 의문이 제기되어 왔다. 신지영(2014)에서는 'ㄴ'가 모음 'ㅣ'나 반모음 'ㅣ'에 선행하는 경우, 후행 모음의 영향으로 조음점이 약간 뒤로 이동하기는 하지만, 이 소리의 조음에는 혓날과 치경(혹은 후치경)이 관계하고 있으므로, 혓몸이 조음체로 작용하여 입천장의 오목한 경구개부에 닿아야 하는 진정한 의미의 경구개음은 아닌 것으로 판단하였다.[53] 치조 비음과 경구개 비음의 실현에 관한 음향음성학적 논의인 김현(2018나)에서도 실험을 통해 'ㅣ'에 앞서는 'ㄴ'의 약 76%가 치조 환경의 그것과 별 차이 없는 포먼트 구조를 보인다는 점을 분석하였고, 이 결과는 'ㅣ' 앞의 'ㄴ'가 경구개 비음으로 실현되지 않는 경우가 많다는 것을 의미하는 것이라 할 수 있다. 하지만 김현(2018나)에서는 나머지 23%의 경우에는 'ㅣ' 앞의 'ㄴ'가 '경구개 비음'으로 실현되었을 가능성과, 전설이 상승하는 이차 조음을 동반한 '치조 비음'으로 실현되었을 가능성을 모두 가지나, 이 중 경구개 비음의 실현 가능성이 다소 커 보임을 밝힌 바 있다. 선행 연구에 따라 이견의 여지는 있으나[54] 음향음성학적으로 'ㅣ' 앞에서 'ㄴ'이 경구개음으로 실현될 수 있으며, 이처럼 'ㄴ+ㅣ' 연쇄에서 경구개음으로 실현된 'ㄴ'를 음소 'ㄴ'의 범주로 판단하지 못할 경우, 인식 오류를 유발하는 하나의 요인이 될 수 있다.

---

53) 이 논의에서는 'ㄴ'뿐만 아니라 'ㅣ'나 반모음 'ㅣ'에 선행하는 'ㄹ'의 이음도 경구개음이 아님을 강조하였다.

54) 이호영(1996: 96)과 같이 'ㄴ'가 전설 고모음 'ㅣ', 경구개 반모음 'j', 경구개 파찰음 'ㅈ, ㅊ, ㅉ' 앞에서 구개음화되어 [nʲ]나 [ɲ]로 발음되는 것으로 간주하는 논의들도 다수 관찰된다.

이와 같은 ‘ㄴ, ㄹ, ㅅ, ㅆ, ㅎ’의 경구개 모음 앞에서의 음성적 변이로 인해 이 환경에서는 음성 인식 오류가 여타 모음 환경에 비해 빈번히 발생하는데, 구체적인 오류 양상을 종합하여 제시하면 아래의 <표 86>과 같다.

	후행모음 / 음절구조	/ㅏ/	/ㅐ/	/ㅓ/	/ㅗ/	/ㅜ/	/ㅡ/	/ㅣ/
ㄴ⇒ㅁ	CV	-	-	-	-	-	2.0(1)	**8.0(4)**
	CVC	3.7(11)	3.3(10)	6.0(18)	1.3(4)	1.0(3)	-	**11.7(35)**
ㄹ⇒ㄴ	CVC	3.0(9)	4.7(14)	1.0(3)	1.3(4)	1.7(5)	6.0(18)	**6.3(19)**
ㅅ⇒ㅆ	CV	-	-	-	-	-	4.0(2)	**36.0(18)**
	CVC	-	2.0(6)	-	-	6.3(19)	***19.7(59)***	2.0(6)
ㅆ⇒ㅅ	CV	-	-	-	-	14.0(7)	***18.0(9)***	6.0(3)
	CVC	2.0(6)	*43.7(131)*	0.7(2)	5.3(16)	21.7(65)	7.0(21)	**54.0(162)**
ㅎ⇒ㅅ	CVC	2.7(8)	-	1.0(3)	-	-	1.0(3)	**4.3(13)**

<표 86> ‘ㄴ, ㄹ, ㅅ, ㅆ, ㅎ’의 후행 모음별 오류 양상

<표 86>에서 음영으로 표시한 부분은 각 음소별 인식 오류가 가장 높은 경우를 나타낸 것이며, 대체로 후행 모음이 ‘ㅣ’일 때 높은 오류율을 보인다. CVC음절의 ‘ㅅ’나 CV음절의 ‘ㅆ’는 ‘ㅣ’가 아닌 ‘ㅡ’ 앞에서의 오류가 보다 많이 관찰되는데, 이는 치조 마찰음·파찰음 뒤 ‘ㅡ’의 음성 실현과 관련하여 후술된다.

한편 ‘ㅣ’ 앞에서 구개음을 변이음으로 가지는 음소는 아니지만, 자연 발화에서는 치조음인 ‘ㄷ’가 경구개 모음 ‘ㅣ’의 조음 위치에 이끌려 경구개음인 ‘ㅈ’로 발음되기도 한다. 음성 인식 상황에서는 자연 발화에서와 같이 실험 대상 음절 ‘디’나 ‘ㄷ-ㅣ-C’를 ‘지’ 계열로 바꾸어 발음하지는 않을 것이나, 이 환경에서의 ‘ㄷ’가 전형적인 ‘ㄷ’의 특성과는 다르게 발음되기도 하는 듯하다. 이러한 비전형적인 ‘ㅣ’ 모음 앞의 ‘ㄷ’를 음성 인식 알고리즘이 ‘ㄱ’와 가깝게 받아들이는 경향이 관찰되는데, CVC음절의 ‘ㄷ’와 ‘ㅌ’가 ‘ㅣ’ 모음 앞에서 ‘ㄱ’나 ‘ㅋ’로 오인식되는 예가 이에 해당된다.

	후행모음 / 음절구조	/ㅏ/	/ㅐ/	/ㅓ/	/ㅗ/	/ㅜ/	/ㅡ/	/ㅣ/
ㄷ⇒ㄱ	CV	-	2.0(1)	-	-	-	*28.0(14)*	6.0(3)
	CVC	-	0.3(1)	-	-	-	2.3(7)	11.0(33)
ㅌ⇒ㅋ	CV	2.0(1)	-	2.0(1)	-	-	2.0(1)	18.0(9)
	CVC	1.0(3)	2.0(6)	1.7(5)	1.3(4)	-	1.3(4)	24.7(74)

<표 87> CVC음절 ‘ㄷ, ㅌ’의 ‘ㄱ, ㅋ’로의 인식 오류

음성 인식 알고리즘이 ‘ㅣ’ 앞의 ‘ㄷ’를 ‘ㅈ’가 아닌 ‘ㄱ’로 인식하는 이유는 앞서 3.1.1.3.에서 기술한 바와 같이 음성 인식 알고리즘이 초성 인식에서 비교적 자음의 범주 지각을 명확히 한다는 점과 관련지어 생각해 볼 수 있을 듯하다. 초성의 음성 인식 오류형들은 ‘폐쇄음-마찰음·파찰음-공명음’과 같은 자음의 유형을 벗어나며 혼동되지는 않는 경향이 나타나는데, ‘ㅣ’ 앞의 ‘ㄷ’의 음성에 해당되는 음소를 파열음 범주 안에서 우선순위를 두고 탐색하고자 한다면, 범주가 다른 ‘ㅈ’가 아닌 ‘ㄱ’이 선택될 가능성이 높다.

2) 양순음, 치조 마찰음·파찰음 뒤 ‘ㅡ’의 음성 실현

중성 ‘ㅡ’는 초성 자음이 양순음이거나 치조 마찰음·파찰음일 때 평균 이하의 인식률을 보인다. 전자는 양순음 뒤 ‘ㅡ’의 음성 특성이 ‘ㅜ’와 가까운 데서 기인한 오류로 보이며, 후자는 마찰음과 파찰음에 후행하는 ‘ㅡ’의 F2값 상승과 관련된 오류로 파악된다. ‘ㅡ’는 김현(2021나)에서 언급된 바와 같이 ‘조음 영역이 한 자리에 고정되어 있지 않고 연구개를 중심으로 하여 다소 앞쪽과 뒤쪽에 걸쳐 발음되는 모음’으로, 선·후행 자음에 따라 조음 위치가 이동하기도 하고, 정일진(1997), 장혜진·신지영·남호성(2015)의 논의와 같이 50~60대에 비해 20대의 발화에서 보다 더 전설화되기도 하는 특성을 가진다. 이러한 ‘ㅡ’의 특성으로 인해 본고의 음성 인식 실험 결과 선행 자음이 양순음일 때의 ‘ㅡ’는 원순모음 ‘ㅜ’로, 치조 마찰음·파찰음일 때의 ‘ㅡ’는 ‘ㅣ’와 같은 전설모음으로 오인식되는 경향이 관찰되었다. 논의에 앞서 4.1.1.에서

제시하였던 중성 ‘ㅡ’의 초성별 인식 오류 양상을 양순음과 치조 마찰음·파찰음을 중심으로 정리하여 보이면 아래의 <표 88>, <표 89>와 같다.

초성 / 음절구조	ㅁ	ㅂ	ㅃ	ㅍ
CV	28.0	20.0	42.0	26.0
CVC	45.0	44.7	38.0	45.3

➡

ㅁ	ㅂ	ㅃ	ㅍ	
38.9	100.0	72.4	92.9	
62.8	89.9	66.3	82.9	
50.9	95.0	69.3	87.9	평균

<표 88> 양순음 뒤 ‘ㅡ ⇒ ㅜ’ 인식률 및 오류 발생 확률[55]

초성 / 음절구조	ㅅ	ㅆ	ㅈ	ㅉ	ㅊ
CV	2.0	-	22.0	20.0	28.0
CVC	5.3	1.7	35.0	15.0	48.0

➡

ㅅ	ㅆ	ㅈ	ㅉ	ㅊ	
11.8	-	73.3	90.9	93.3	
31.4	15.2	76.6	47.9	91.2	
21.6	7.6	74.9	69.4	92.3	평균

<표 89> 치조 마찰음·파찰음 뒤 ‘ㅡ ⇒ ㅣ’ 인식률 및 오류 발생 확률[56]

위의 표에 제시된 ‘오류 발생 확률’이란 단어 간의 유사율을 정의하기 위해 한동조·유제만·최기호(2006)에서 제안한 것으로, 이 논의에서는 오류 발생 확률 모형을[57]

55) 참고로 양순음 외의 자음들이 선행하는 경우의 ‘ㅡ’의 ‘ㅜ’로의 인식률을 정리해 보이면 아래와 같다.

초성 / 음절구조	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅈ	ㅉ	ㅊ	ㅅ	ㅆ	ㅎ	ㄴ	ㄹ
CV	-	-	-	-	-	-	4.0	2.0	-	2.0	-	-	-	-
CVC	4.3	0.3	2.0	2.3	2.7	7.0	7.3	8.0	6.0	6.0	1.3	0.3	2.0	13.7

56) 참고로 치조 마찰음·파찰음 외의 자음들이 선행하는 경우의 ‘ㅡ’의 ‘ㅣ’로의 인식률을 정리해 보이면 아래와 같다.

초성 / 음절구조	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅂ	ㅃ	ㅍ	ㅎ	ㄴ	ㅁ	ㄹ
CV	-	-	-	2.0	-	2.0	-	-	-	-	2.0	6.3	4.0
CVC	-	0.7	-	2.3	0.7	0.7	0.3	1.0	0.7	-	1.3	0.3	7.3

통해 각 음소들이 어떤 음소로 오류를 겪는지를 확률적으로 구성하고자 하였다. <표 88>, <표 89>에 제시한 오류 발생 확률은 이와 같은 한동조·유제만·최기호(2006)의 오류 발생 확률 모델에 따라 해당 음소의 전체 오류 발생 빈도 중 특정 음소로의 오류 발생 빈도를 계산하여 도출한 것이다.[58]

<표 88>의 오류 발생 확률을 중심으로 먼저 양순음과 'ㅡ'의 연쇄에 대해 살펴보면, 이 환경에서 'ㅡ'에 인식 오류가 발생할 때, 'ㅁ'가 선행할 때의 'ㅡ'는 평균 50.9%, 'ㅂ'는 95.0%, 'ㅃ'는 69.3%, 'ㅍ'는 87.9%가 'ㅜ'로 오인식됨을 알 수 있다. 양순음 뒤의 'ㅡ'는 자연 발화에서 'ㅜ'로 발음되기도 하나, 본고와 같은 단음절 음성 인식 실험 상황에서는 'ㅡ'를 음소 'ㅜ'로 바꾸어 발음하는 경우는 관찰되지 않았다. 하지만 이 환경에서의 'ㅡ'가 음성적으로는 여타의 환경에서와 달리 실현되기도 하는데, 음성 인식 알고리즘은 양순음이 선행할 때의 이와 같은 'ㅡ'의 음향 특성을 전형적인 'ㅡ'보다는 'ㅜ'에 가깝다고 판단하는 듯하다.[59] 양순음과 'ㅡ'가 연쇄된 환경

---

57) **<음소 오류 발생 확률 모델 도출 식>**

$$p(m \mid n) = \frac{E_{m \mid n}}{E_n} \times 100$$

E_n: n음소의 오류 발생 빈도
$E_{m \mid n}$: n음소의 오류 발생 빈도 중 음소 m으로의 오류 발생 빈도

58) 인식률과 함께 오류 발생 확률을 따로 제시한 이유는 다음과 같다. 주지하다시피 모음의 인식률은 초성 자음에 따라 차이를 보이는바, CV음절의 경우 'ㅁ'가 선행할 때의 'ㅡ'의 인식률은 28%에 불과하지만, 'ㄱ, ㅋ'가 선행할 때의 인식률은 100%로 초성에 따른 중성의 인식률 차이가 매우 큰 편이다. 이로 인해 단순한 오인식률의 비교만으로는 음성 인식 오류상의 특징을 파악하기 어렵다고 판단하여 오류 발생 확률 값을 함께 제시하였다. <표 89>의 'ㅊ'를 예로 들어 볼 때, CV구조에서 'ㅊ'에 후행하는 'ㅡ'가 'ㅣ'로 오인식된 비율은 28.0%이고, CVC구조에서는 48.0%로 나타나 총 20%의 인식률 차를 보였지만, 이를 오류 발생 확률로 계산하면 CV구조의 경우 93.3%, CVC구조에서는 91.2%로 도출되어 둘 간에 2.1%의 차이만을 보이게 된다. 이때의 오류 발생 확률 값은, CV음절의 경우 'ㅊ'에 후행하는 'ㅡ'의 인식률이 높아 오류 발생률 자체는 적었으나(28%), 발생한 오류 중 93.3%가 'ㅡ'를 'ㅣ'로 오인식하였음을 나타내고, CVC음절에서 'ㅊ'에 후행하는 'ㅡ'는 인식률이 낮아 오류 발생률은 비교적 높았으나(48%), 발생한 오류 중 91.2%가 'ㅡ'를 'ㅣ'로 오인식하였음을 의미한다.

59) 김현(2021나)에 따르면 양순음 뒤의 'ㅡ'를 발음할 때는 성도의 전강이 넓어 'ㅜ'에 보다 가까운 것으로 볼 수 있다고 한다.

에서 나타나는 음성 인식 오류는 이와 같은 양순음 뒤의 ‘ㅡ’의 음성 실현상의 특성에 의해 유발된 것으로 간주할 수 있을 듯하다.

한편 치조 마찰음·파찰음에 후행하는 ‘ㅡ’의 인식 오류는 <표 89>에 제시된 오류 발생 확률을 통해 볼 때, 이 환경에서 ‘ㅡ’에 인식 오류가 발생한다면, ‘ㅅ’가 선행할 때의 ‘ㅡ’는 평균 21.6%, ‘ㅆ’는 7.6%, ‘ㅈ’는 74.9%, ‘ㅉ’는 69.4%, ‘ㅊ’는 92.3%가 ‘ㅣ’로 오인식됨을 알 수 있다. 황미경·조남민(2020)에 따르면 ‘ㅡ’는 마찰음·파찰음이 선행할 때의 F2값이 단독 모음일 때와 비교하여 상대적으로 높아지며 전설 방향으로 이동하는 경향이 나타난다고 하는데, 이와 같은 마찰음·파찰음 뒤의 ‘ㅡ’의 음성 실현상의 특성이 <표 89>와 같은 인식 오류를 유발한 것으로 보인다. 단 이러한 경향은 치조 마찰음보다는 파찰음 계열에서 더 강하게 나타나는데, 이는 선행 자음이 ‘ㅅ’계열일 때의 ‘ㅡ’의 인식률이 여타의 환경에 비해 높은 데서 기인한 것으로 보이기도 하며, 홍은영(2012)의 논의와 같이 ‘ㅅ’보다 ‘ㅈ’ 앞에서 더 강하게 적용되는 전설고모음화 현상의[60] 영향으로 간주할 수도 있을 듯하다.

6.2. 음향적 요인에 의한 인식 오류

전술한 바와 같이 음성 인식 오류는 음성 인식 알고리즘이 범주화한 각 음소들의 전형적인 음향 특징 범위를 벗어나거나 그 경계에 있는 음이 감지되었을 때 주로 발생한다. 6.1.에서 살핀 음성적 변이에 의한 인식 오류 유형이 주로 전자와 관련되어 발생한다면, 본 절에서 살필 음향적 요인에 의한 인식 오류는 서로 닮은 음소들의 음향적 특성이 음소 범주의 경계를 넘나들며 유발되는 것으로 보인다. 이러한 인식 오류는 음향적으로 유사한 음소들 사이에서 나타나거나, 특정 환경에서 음향적 대립이 없어지는 경우, 또는 음향 특성 중 일부가 명확히 인식되지 못하는 경우에 주로 발생한다. 본 연구의 단음절 음성 인식 실험 상황에서 음향적 유사성에 의해

60) 치조 마찰음·파찰음 뒤에서 ‘ㅡ’가 ‘ㅣ’로 인식되는 오류 현상은, 음운 현상으로는 주로 /ㅅ, ㅆ, ㅈ, ㅊ/ 뒤의 후설모음 /ㅡ/가 전설모음 /ㅣ/로 변하는 ‘전설고모음화’와 관련된다.

유발된 인식 오류로는 크게 양순음과 연구개음 간의 오인식과, 종성 비음들 간 오인식이 관찰되었다.

1) 양순음과 연구개음 간의 오류

양순음과 연구개음은 음향적 유사성으로 인해 자연 발화의 음성 지각에서도 흔히 혼동되는 음소이다. 양순음과 연구개음 간의 음향적 유사성은 Jakobson·Fant·Halle(1951) 이래로 [grave] 자질과 관련되어 설명되어 왔다. Jakobson·Fant·Halle(1951: 29-30)에서는 음향적 측면에서 [grave] 자질을 스펙트럼의 아랫부분이 강하게 나타나는 특성을 가진 것으로 정의하고, '후설모음', '양순음', '연구개음'을 [grave] 자질을 가지는 음소로 분류하였다. Jakobson·Fant·Halle(1951)에서는 이와 더불어 [grave] 자질을 가지는 음소들은 인두를 좁힘으로써 구강 공명기의 후강을 수축시키는 형성 과정을 가진다는 점과, [grave] 자질의 자음들의 경우 인접한 모음의 제2포먼트가 내려가는 음향적 특징을 보인다는 점을 지적하였다. 이진호(2012: 260)에서도 같은 맥락에서 양순음과 연구개음의 음향적 유사성이 인접 모음의 F2 전이의 유사성에서 확인된다고 서술한 바 있다. 즉 양순음과 연구개음의 음향적 유사성은 스펙트럼의 아랫부분에 강한 주파수 대역이 형성되는 [grave] 자질의 공유와, 인접 모음의 F2 전이 양상의 유사성에 토대를 두고 있다는 것이다.

양순음과 연구개음 간의 음향적 유사성은 'pk-교체 현상'을 이해하는 데에도 적용된다. pk-교체에 관한 종합적 고찰을 담고 있는 이진호(2012)에서는 pk-교체는 p와 k를 정확히 구분하여 듣지 못하기 때문에 일어나며, 이와 같은 지각의 혼동이 발생한다는 것은 p와 k가 잘 구별되지 않음을 의미하는 것으로 간주하였다. 그리고 이러한 'ㅂ'와 'ㄱ' 간의 음향적 유사성은 음성 인식 상황에서도 인식 오류를 유발하는 요인이 된다. 아래의 <표 90>은 양순음과 연구개음 간의 음성 인식 실험 결과를 pk-교체와 관련지어 정리한 것이다.

'p⇒k'		후행모음 / 인식자음	/ㅏ/	/ㅐ/	/ㅓ/	/ㅗ/	/ㅜ/	/ㅡ/	/ㅣ/
초성	CV	ㅂ⇒ㄱ	-	-	-	-	-	64.0(32)	2.0(1)
		ㅃ⇒ㄲ	2.0(1)	-	48.0(24)	-	-	2.0(1)	-
		ㅍ⇒ㅋ	2.0(1)	-	-	-	-	40.0(20)	-
	CVC	ㅂ⇒ㄱ	-	-	-	-	0.3(1)	14.0(42)	2.0(6)
		ㅃ⇒ㄲ	0.7(2)	-	6.7(20)	4.3(13)	4.0(12)	13.0(39)	-
		ㅍ⇒ㅋ	-	-	1.0(3)	6.3(19)	2.0(6)	25.0(75)	2.3(7)
종성	VC	ㅂ⇒ㄱ	-	-	-	32.0(16)	46.0(23)	-	-
	CVC	ㅂ⇒ㄱ	9.3(84)	5.3(48)	17.0(153)	43.3(390)	52.8(475)	18.1(163)	1.7(15)

'k⇒p'		후행모음 / 인식자음	/ㅏ/	/ㅐ/	/ㅓ/	/ㅗ/	/ㅜ/	/ㅡ/	/ㅣ/
종성	VC	ㄱ⇒ㅂ	30.0(15)	36.0(18)	18.0(9)	-	2.0(1)	10.0(5)	-
	CVC	ㄱ⇒ㅂ	19.0(171)	11.7(105)	8.4(76)	1.1(10)	0.7(6)	8.1(73)	7.2(65)

<표 90> 양순음-연구개음 간 음성인식 오류 양상

위의 <표 90>에서 특징적인 것은 'p ⇒ k'로의 인식 오류는 초성과 종성에서도 모두 나타나지만, 'k ⇒ p'는 종성에서만 나타난다는 점이다. 그리고 종성의 'p ⇒ k'는 원순모음이 선행할 때 가장 활발히 적용되지만, 'k ⇒ p'는 반대로 선행 모음이 원순모음일 때의 'k'에서 'p'로의 오인식이 저조하다. 이러한 인식 오류 양상은 'pk-교체' 현상과 유사한 면이 있다. 실제 'pk-교체'의 두 유형 즉 'k ⇒ p' 방향으로의 교체와 'p ⇒ k' 방향으로의 교체 중에서는 'p ⇒ k' 방향의 변화가 통시적으로나 공시적으로 더 빈번하게 적용되어 왔고, 이와 같은 'p ⇒ k' 방향의 교체 현상은 소신애(2011), 이진호(2012)에서 지적된 바와 같이 p와 원순모음이 인접한 환경에서 두드러지게 나타난다. 'p ⇒ k' 방향의 교체가 더 활발한 원인을 이진호(2012)에서는 'p'와 'k'의 음성학적 성격과 관련지어 파악하고 있는데, 조음 시 혀가 관여하지 않는 양순음은 후행 모음과 상관없이 조음 위치가 일정하고, 그렇지 못한 연구개음에 비해 'pk-교체'의 적용을 받을 가능성이 높아지는 것으로 간주한다. 'pk-교체'가 'p'와 원순모음이 인접한 환경에서 나타나는 이유와 관련하여 소신애(2011)은 순음성 자질을 공유한 분절음의 연쇄를 회피하는 경향에서 기인한 것으로 판단하였다.[61]

이러한 'pk-교체'의 방향성으로 인해 음성 인식 상황에서도 'k ⇒ p'로의 오인식은 종성에만 국한되어 나타나는 듯하다. 그리고 순음성 자질을 공유한 분절음 연쇄를 회피하려는 경향으로 인해 'k ⇒ p'로의 오인식이 원순모음이 선행할 때는 저조하게 나타나는 것으로 판단된다. 양순음과 연구개음 간의 오류가 종성에서 유독 활발한 이유는 'ㅂ'와 'ㄱ' 모두 종성 위치에서는 불파화되기 때문이며, 이로 인해 종성 위치의 'ㅂ'와 'ㄱ'는 상호 혼란을 겪을 뿐만 아니라 종성의 존재 자체가 인식되지 않기도 한다.

2) 종성 비음 간의 오류

앞서 5.1.2.에서 살핀 종성 공명음들의 음성 인식 오류에서는 종성 비음들의 상호 오인식 양상이 주로 관찰되었다. 이때 종성 'ㄴ'는 'ㅁ, ㅇ'로, 'ㅁ'는 'ㅇ'로, 'ㅇ'는 'ㅁ'로 오인식되었는데, 이와 같은 종성 비음 간의 혼란은 자연 발화에서도 나타나는 현상이다. 소신애(2019)에서는 이를 '음절 말 비음 간 대치 현상'으로 간주하고, 현상의 유형을 네 가지로 나누어 제시하였는데, 이를 인용해 보이면 아래의 <표 92>와 같다.

유형	대치의 방향
유형1	ㄴ → ㅁ → ㅇ
유형2	ㄴ → ㅁ
유형3	ㅁ → ㅇ
유형4	ㄴ → ㅇ

<표 91> 소신애(2019: 56)의 음절 말 비음 간 대치 현상의 유형

소신애(2019)에서는 위와 같은 유형의 음절말 비음 간 대치 현상들이 본질적으로는 지각적 변별력을 강화시키고자 하는 동인에 의해 발생하는 것으로 전제한다. 그

61) pk-교체와 관련된 이른 시기의 논의인 주왕산(1948)에서도 'ㅂ+모음+ㅂ' 연쇄가 'ㅂ+모음+ㄱ'가 되는 '붑 → 북', '거붑 → 거북'과 같은 예들을 자음의 '이화' 작용에 의한 것으로 전제하고, 이러한 이화가 일어나는 원인으로 순음의 연속, 말음에 오는 순음의 음향적 효과 등을 제시한 바 있다(이진호 2009: 153에서 재인용).

리고 이와 같은 지각적 변별력을 강화하는 방식으로는, 발화 산출 과정에서 화자에 의해 직접적으로 비음 대치가 수행되는 방법과, 발화 해석 과정에서의 청자의 오지각으로 인해 비음 간 대치가 촉발될 가능성을 제시하였다. 이 중 전자와 관련하여 소신애(2019)는 음절 말 비음 간 대치 현상과 같은 동화와 무관한 조건 환경에서 'ㄴ→ㅁ이나 ㅁ→ㅇ, ㄴ→ㅇ'과 같은 방향의 음운 대치가 일어나는 것은 조음적으로 자연스러운 현상이 아니라는 점과, 대부분의 조음상의 강화 현상이 어두 음절 초에서 일어난다는 점을 근거로 하여, 발화 산출 과정에서의 비음 대치 가능성을 배제하였다. 그 뒤 음절 말 비음 간 대치 현상이 특정 분절음들의 음향적 유사성으로 인해 발생한 음향적 혼동이, 발화 해석 과정에서 청자의 오지각을 야기하여 촉발된 것으로 간주한다. 즉 소신애(2019)에서는 음절 말에서 실현된 특정 비음을 청자가 음향적 유사성을 가진 다른 비음으로 지각함으로써 이와 같은 비음 대치가 일어나는 것으로 판단한 것이다.

나아가 소신애(2019)에서는 종성 비음들의 음향적 유사성이, 조음 방법에 대한 지각 단서가 풍부한 데 반해 조음 위치에 대한 지각 단서가 다른 조음 방법의 자음들에 비해 빈약한 비음의 특성에서 기인한 것으로 설명한다. 이 논의에 따르면 비음은 다른 조음 방법을 가진 자음에 비하여 조음 위치 변별에 필요한 음향적 단서의 실현 강도가 약하여 조음 위치를 변별적으로 지각하기 어려운 면이 있으며, 또한 비음이 가진 음향적으로 미분화된 스펙트럼 형태 특성으로 인해 조음 위치 단서의 지각적 가치가 감소된다고 한다. 이와 같은 선행 연구의 논의 결과를 토대로 할 때, 조음 방법이 동일한 종성의 비음은, 이들을 변별할 수 있는 조음 위치에 관한 음향적 단서들이 적어 음향적 유사성을 가지게 된 것으로 볼 수 있을 듯하다.

본 연구의 실험 결과 관찰된 종성 비음 간의 오류 현상들도, 이와 같은 종성 위치 비음들이 가진 음향적 유사성에서 기인한 것으로 판단된다. 본 실험의 종성 비음 간 오인식 양상은 앞서 제시한 <표 91>의 네 가지 음절 말 비음 간 대치 유형과 동일하게 관찰되었으며, 이 외에 <표 91>에 포함되지 않은 종성 'ㅇ'의 'ㅁ'로의 오류가 추가적으로 확인되었다.

그밖에 마찰 구간의 길이 및 기식 길이 등으로 변별되는 'ㅅ'와 'ㅆ'가 고모음 앞에서 음향적 대립이 약해져 상호 오인식되는 경우도, 음향적 유사성에 의한 인식 오

류 유형에 포함될 수 있다. 또한 앞서 살핀 'ㄱ/ㅋ, ㅂ/ㅍ ⇒ ㅎ'와 같은 인식 오류는 파열음의 조음 과정에서 나타나는 파열을 인식하지 못하여 유발된 것으로, 이 역시 음향적 요인과 관련된 오류 유형으로 판단된다. 'ㅊ ⇒ ㅅ'의 오류와 같이 파찰음의 조음 과정 중 폐쇄 단계가 인식되지 못하여 발생한 오류도, 음향적 특성이 소실되어 발생한 오류 즉 음향적 요인에 의해 유발된 오류 유형으로 간주한다.

6.3. 성별 요인에 의한 인식 오류

본고에서는 앞서 음성대화형 어시스턴트와의 대화 시 발화자의 어휘적 차원의 지역방언적 특성은 강하게 드러나지 않을 것으로 간주하였으나, 이 밖의 음성·음운적 층위의 지역방언 혹은 사회언어학적 특성은 음성 인식 상황에서도 표현될 가능성이 높을 것으로 보인다. 이와 관련하여 이창영(2010)은 I. Kudo, T. Nakama, & T. Watanabe(1996), S. Yildirim & S. S. Narayanan(2003) 등의 논의를 통해 화자의 나이에 따른 음향학적 정보 및 화자의 세대 정보가 음성 인식에 영향을 미친다는 연구 결과가 다수 보고된 바 있음을 언급하기도 하였다. 이 논의에서는 이와 더불어 남녀를 구분하여 성별마다 다른 음성 인식을 수행한다면, 인식률의 향상을 꾀할 수 있을 것으로 기대된다는 서술과 함께, 이러한 연유로 실제 음성 인식 개발 시 성별 음성 인식이 종종 실행되고 있음을 언급하기도 하였다(R. Muralishankar & D. O'Shaughnessy 2008, I. Gavat & C. O. Dumitru 2007).

위와 같은 세대·연령·성별과 같은 다양한 사회언어학적 요소들 중 본고의 음성 인식 실험에서는 성별에 따른 인식 오류 차이만이 관찰되었는데,[62] 자음의 인식과 관련하여서는 '평음-격음' 간의 인식 오류가, 모음의 인식과 관련하여서는 'ㅓ'의 'ㅏ'로의 인식 오류, 'ㅗ'의 'ㅜ'로의 인식 오류가 성별 요인에 의해 유발된 것으로 판단된다.

이 중 먼저 '평음-격음' 간 인식 오류에 대해 구체적으로 살펴보면, 격음을 평음으

62) 본고의 음성 인식 실험은 20-40대 음성 제공자들을 대상으로 하고 있어, 세대나 연령에 따른 음성 인식 오류 양상은 확인할 수는 없었다.

로 인식하는 오류는 발생 빈도는 적으나 오류의 대부분이 남성의 발화에서 나타나고, 평음을 격음으로 인식하는 오류는 성별에 따른 오류 빈도 차가 크지는 않지만 여성에게서 상대적으로 높게 나타난다. 편의상 3.1.1.3.에서 제시한 성별에 따른 평음·격음 간 오인식 빈도 표를 다시 가져와 보이면 아래의 <표 93>과 같다.

		오인식 빈도		
		전체	남성	여성
격음⇒평음	ㅋ⇒ㄱ	48	45	3
	ㅌ⇒ㄷ	42	40	2
	ㅍ⇒ㅂ	35	33	2
평음⇒격음	ㄱ⇒ㅋ	840	376	464
	ㄷ⇒ㅌ	727	319	408
	ㅂ⇒ㅍ	649	284	365

<표 92> 성별에 따른 평음·격음 간 오인식 빈도

주지하다시피 평음과 격음의 변별에는 자음의 VOT값과 후행 모음의 F0값이 결정적인 영향을 미친다. 이 중 VOT는 젊은 세대로 갈수록 변별에 관여하지 않는다는 논의들도 있으나(Kim, Mi-Ryoung 2008, 김지은 2017), F0값은 평음과 격음 변별의 유의미한 변수로 다루어진다. 대체로 F0가 높으면 격음으로 낮으면 평음으로 인식되므로, 음성 인식 오류형들 중 평음을 격음으로, 격음을 평음으로 오인식하는 경우는 우선 이 F0값과 관련되어 발생한 오류로 생각해 볼 수 있다.

그런데 이러한 오류가 <표 92>와 같이 성별에 따라 달리 나타난다는 점에 주목해 본다면, 이는 평음과 격음 간 F0값의 차이로 인해 발생한 오류라기보다는, 성별에 따른 F0의 특성이 고려되지 않은 음향 모델링에 의해 유발된 오류일 가능성이 높아 보인다. 남성과 여성은 성대의 크기와 길이, 질량 등의 차이로 인해 F0값이 차이를 보이는데, 이러한 남성과 여성의 평음과 격음의 평균 F0값을 정리해 보이면 아래의 <표 94>와 같다.

	남성	여성
평음	119	233
격음	128	256

<표 93> 성별 평음과 격음의 F0 평균(Hz)

위의 <표 93>은 박한상(2007: 60)에 제시된 남녀 평음·격음의 F0 평균값으로, 이를 통해 여성의 F0값이 남성에 비해 거의 두 배가 높다는 점과, 남성 격음의 F0값이 여성의 평음의 F0값보다 낮게 나타난다는 점을 확인할 수 있다. 그런데 이와 같은 F0값의 성차가 음성 인식 알고리즘의 모델링 과정에는 적절히 반영되어 있지 않은 듯하다. 평음과 격음을 변별하는 F0값이 성별에 관계없이 설정되었을 경우, 남성 격음의 낮은 F0값은 평음으로 판단될 가능성이 높다. 실제 남성의 격음은 아래의 <표 95>에서 확인할 수 있듯이 중간형에서 최종형으로 넘어가며 더욱 많은 예들이 평음으로 오인식되는데, 이는 낮은 F0값을 가지는 남성의 격음이 언어 모델에 의해 일률적으로 평음으로 교정된 결과인 듯하다.[63]

격음⇒평음	오인식 빈도					
	전체		남성		여성	
	중간	최종	중간	최종	중간	최종
ㅋ⇒ㄱ	48 → 301		45 → 278		3 → 23	
ㅌ⇒ㄷ	42 → 175		40 → 144		2 → 31	
ㅍ⇒ㅂ	35 → 188		33 → 163		2 → 25	

<표 94> 중간형과 최종형의 '격음 ⇒ 평음' 오인식 빈도

전술한 바와 같이 본고의 단음절 음성 인식 실험 결과 중 '중간형'에는 발화된 음성의 음향 특성에 근거한 음향 모델의 판단이 작용했을 가능성이 높고, '최종형'에는

63) 음성 인식 상황에서 언어모델은 대개 부정확한 음향 정보를 보완해주는 역할을 하지만, 본 실험에서는 사용 빈도가 적은 음절을 음성적으로 유사한 고빈도 음절로 잘못 정정하거나, 비규범적 음운 현상(주로 '어두 경음화'나 '활음 탈락')의 적용형으로 오인하여 과도 교정하는 등의 처리 결과로 인식 오류를 유발하기도 하였다.

해당 음절의 사용 빈도나 출현 확률과 같은 음향 단서 이외의 정보들이 반영되었을 가능성이 높다. 위의 <표 94>의 중간형에서 발생한 '격음 ⇒ 평음' 오류는 그 빈도가 다소 적은 편이었으나, 최종형 단계에서는 크게 늘어나는 모습이 확인된다. 이는 음향적 근거에 따라 격음으로 올바르게 인식했던 예들을 이후의 추론 과정에서 평음으로 잘못 정정한 것으로 판단되는데, 이와 같은 중간형과 최종형 사이의 인식 변화는 음성 인식 알고리즘이 상대적으로 낮은 F0값을 가지는 남성의 격음을 평음으로 판단하고, F0값이 상대적으로 높은 여성의 평음은 격음으로 판단한 데에서 기인한 오류로 추정된다.

한편 모음의 인식 오류와 관련하여 'ㅓ'를 'ㅏ'로 인식하는 오류 역시 이와 비슷한 요인에 의해 유발된 현상인 듯하다. 앞서 4.1.1.에서 단모음 'ㅓ'의 유사 음소가 'ㅗ'와 'ㅏ'라는 점을 확인하였는데, 이때 'ㅓ'를 'ㅏ'로 오인식하는 예들은 아래의 <표 95>에 제시된 바와 같이 대부분 여성에게서 두드러지게 관찰된다.

	남성	여성
'ㅓ ⇒ ㅏ'	9.4%(26)	90.6%(251)

<표 95> 성별 'ㅓ ⇒ ㅏ' 오인식률과 빈도

<표 95>를 통해 알 수 있듯이 본 실험에서 'ㅓ'를 'ㅏ'로 오인식한 빈도는 전체 277회였고, 그 중 90.6%는 여성의 발화에서 나타났다. 모음의 조음 위치는 주로 F1값과 F2값을 통해 확인되는데, 이주현(2005: 10)에 제시된 모음별 남녀 포먼트 데이터에 의하면, 'ㅓ'의 F2값은 다른 모음들이 비해 남녀 차가 그리 크지 않으나, F1값은 크게 차이가 나는 것으로 관찰된다.

F1	남	여	여성F1-남성F1	
ㅓ	445	576	+131	↑
ㅏ	651	945	+294	↑↑
ㅐ	415	545	+130	↑
ㅗ	320	371	+51	↓
ㅜ	324	346	+22	↓
ㅡ	317	390	+73	↓
ㅣ	236	273	+37	↓

F2	남	여	여성F2-남성F2	
ㅓ	845	961	+116	↓
ㅏ	1156	1582	+426	↑
ㅐ	1848	2436	+113	↓
ㅗ	587	700	+215	↓
ㅜ	595	810	+588	↑
ㅡ	1218	1752	+534	↑
ㅣ	2183	2864	+681	↑

<표 96> 모음별 남녀 F1, F2 평균(이주현 2005: 10)

<표 96>의 남녀 F1, F2 차이를 토대로 본다면, ‘ㅓ’의 인식 오류에는 F2보다는 남성과 여성의 F1의 차이가 관여했을 가능성이 높을 듯하다. 여성의 ‘ㅓ’의 F1값은 남성의 ‘ㅓ’보다는 ‘ㅏ’에 더 가까울 만큼 높게 나타나는데, ‘ㅓ ⇒ ㅏ’의 인식 오류는 이와 같이 높은 F1값을 가지는 여성의 ‘ㅓ’를 모음들 중 가장 F1값이 큰 ‘ㅏ’로 오인식한 결과로 추정된다.

다음으로 살펴볼 ‘ㅗ ⇒ ㅜ’ 오인식 역시 주로 여성에게서 나타나는 현상이기는 하나, 앞서 다룬 평음-격음 간 오류와 ‘ㅓ ⇒ ㅏ’ 오인식은 성별에 따라 다른 음소들의 음향 특성이 고려되지 못한 결과 발생한 오류인 반면, ‘ㅗ ⇒ ㅜ’ 오인식은 성별에 따라 다른 ‘ㅗ’와 ‘ㅜ’의 조음상의 특징에 의해 발생한 오류로 판단된다. 주지하다시피 모음의 조음 공간은 지역과 세대에 따라 차이가 있으며, 특히 많은 실험음성학 및 사회언어학 연구에서 지적하고 있는 바는 20~30대 여성에게서 주로 나타나는 후설 원순모음 ‘ㅗ’의 상승과 이로 인한 ‘ㅜ’의 전설화 문제이다. 논의들마다 ‘ㅗ’와 ‘ㅜ’의 조음 공간상의 위치에 대한 관점 차는 있으나, 젊은 여성들을 중심으로 후설 원순모음의 조음 공간 이동이 존재한다는 점은 인정되는 듯하다. 이와 같은 조음 공간

의 변이로 인해 음성 인식 실험 결과 실제 여성이 발화한 'ㅗ'가 'ㅜ'로 오인식되는 오류가 남성의 그것에 비해 높게 나타났다.

이로 인해 'ㅜ'의 인식률도 더불어 낮게 나타나는데, 음성 인식 실험 결과 감지된 음성 'ㅜ'에서 원순성이 인식되면 'ㅗ'로, 원순성이 인식되지 않으면 'ㅡ'로 오인식되는 경향이 관찰된다. 이와 같은 오류는 'ㅗ'와 'ㅜ'가 안정적으로 발음되는 세대의 음성 인식 상황에서는 발생 가능성이 낮아질 것이라 예상된다. 후설 원순모음에서 나타나는 인식 오류는 성별 조음에 따른 조음상의 특성에서 유발된 오류 유형으로 분류할 수 있다.

7. 결론

이상으로 자동 음성 인식의 인식 오류에 영향을 주는 음운론적 요인들과, 음운 환경에 따른 유사 음소 및 음성 인식 오류의 유발 요인에 근거한 오류 유형에 대해 살펴보았다. 본고에서는 이를 위해 먼저 자동 음성 인식 기술의 음성 인식 원리와 현재까지 진행되어 온 한국어 음성 인식에 관한 선행 연구들의 연구 방향을 살펴보았으며, 음성 인식 오류가 어떠한 음운론적 요인에 의해 발생하는지 확인하기 위해 한국어의 모든 단음절을 대상으로 한 음성 인식 실험을 진행한 뒤, 실험 결과를 바탕으로 음성 인식 오류 양상을 분석하였다. 이후에는 실험 결과를 초성·중성·종성으로 나눈 뒤, 초성은 다시 파열음, 파찰음, 마찰음, 공명음 순으로 분류하여 각 자음별 인식 오류 양상을 기술하였고, 중성은 단모음과 이중모음으로 나누어 오류 양상을 분석하였다. 중성은 장애음 계열과 공명음 계열로 나누어 인식 오류상의 특징을 살펴보았으며, 모든 음소마다 오류 빈도가 가장 높은 음소를 특정하여 환경별 유사 음소를 제시하였다.

초성의 음성 인식 오류 양상은 다음과 같이 요약된다. 초성의 경우 주파수 대역이 가장 극단에 위치한 마찰음과 공명음의 인식률이 높게 나타났으며, 조음 과정 중 묵음에 가까운 폐쇄 단계를 가지며 추가적인 인식 단서가 적은 평파열음의 인식률이 가장 낮은 것으로 확인되었다. 초성 자음들은 음성적 변이가 발생하는 후행 모음 환경에 놓이거나, 자음들 간의 음향 특성이 혼동되는 환경에서 오류가 발생하는 경향을 보인다. 초성 파열음 중 평음은 주로 격음과 혼동되는데, 이러한 오류는 성별에 따른 F0 특성과의 관계 속에서 이해될 수 있다. 남성의 낮은 F0값은 격음에서 평음으로의 오인식을, 여성의 높은 F0값은 평음에서 격음으로의 오인식을 유발하는 것으로 추정된다. 경음과 격음은 경음 간, 격음 간의 오인식이 주로 나타난다. 이는 해당 음소가 경음과 격음이라는 점은 비교적 정확히 인식된다는 의미이며, 경음과 격음은 '긴장'이나 '기식'과 같은 유표적인 특성을 가지고 있어 평음보다 인식이 용이한 듯 보인다. 마찰음과 파찰음도 경음·격음과 마찬가지로 마찰음이라는 범주와 파찰음이라는 범주는 비교적 명확히 지각된다. 단 이들은 고모음 환경에서 마찰 구간

의 변별이 약해지거나, 마찰과 기식이 적절히 변별되지 않을 경우 서로 간의 오류가 발생한다.

중성의 음성 인식 오류 양상은 다음과 같이 요약된다. 단모음 중 인식률이 가장 높은 모음은 'ㅣ', 가장 낮은 모음은 'ㅜ'이며 대체로 후설모음의 인식률이 낮게 나타난다. 단모음은 선행 자음의 조음 위치의 영향으로 인식 오류가 유발되거나, 성별에 따른 모음의 음성적·조음 공간적 차이에 의해 오류가 발생한다. 조음 위치와 관련된 인식 오류는 주로 모음 'ㅡ'에서 관찰된다. 'ㅗ'와 'ㅜ'는 현대 한국어에서 세대·성별에 따른 조음 공간의 차이가 큰 모음으로 간주되는데, 이 중 'ㅗ'의 'ㅜ'로의 오인식은 여성에게서 유의미하게 나타난다. 이는 성별에 따른 원순모음의 조음 공간 차이에 의한 오류로 판단된다. 'ㅓ'의 'ㅏ'로의 오인식도 여성에게서 유의미하게 관찰되는 오류이며, 이는 남성과 여성의 F1 특성 차이에서 기인된 것으로 판단된다. 이중모음의 인식에 있어서 j계 이중모음은 자연 발화에서의 발음 특성이 인식에 영향을 미치며, w계 이중모음은 초성의 존재 여부와 함께, 사용 빈도에 근거한 이중모음의 출현 확률이 인식 오류에 영향을 주었다.

종성의 음성 인식 오류 양상은 다음과 같이 요약된다. 종성의 경우 유음, 비음, 파열음 순으로 인식률이 낮아지며, 파열음 상호 간, 비음 상호 간의 인식 오류가 주로 발생한다. 종성의 경우 음소의 존재 자체를 인식하지 못하는 오류도 다수 관찰되는데, 이는 포먼트 전이 양상이나 선행 모음과 종성 간의 경계가 명확히 인식되지 못한 데에서 기인한 오류이다.

음성 인식 오류는 음성 인식 알고리즘이 정해둔 각 음소들의 전형적인 음향 특징 범위를 벗어나거나, 그 경계에 있는 음이 감지되었을 때 발생하는 것으로 보인다. 그리고 이러한 경우 음소를 정의하는 여러 음향 특성 중 어떠한 특성이 임계점을 벗어났는지에 따라 오인식되는 음소가 결정된다. 이때 발화음의 특성이 해당 음소의 임계점을 벗어나게 만드는 데에는 특정 음운론적 환경이 관여할 수도 있고, 특정 조건 없이 음향적 특성이 유사한 음소들 간의 오류가 인식을 유발할 수도 있다. 본고에서는 이 중 전자와 같은 요인에 의해 발생하는 음성 인식 오류를 '음성적 변이에 의한 인식 오류' 유형으로, 후자를 '음향적 요인에 의한 인식 오류' 유형으로 분류하였다. 이외에 남성과 여성의 발화가 가진 음성학적·음운론적 특성도 음성 인식 오류

에 영향을 주는 것으로 판단하였으며, 이는 '성별 요인에 의한 인식 오류' 유형으로 설정하였다.

본 연구의 단음절 음성 인식 실험 결과 도출된 각 음소의 유사 음소를 정리해 보이면 아래의 <표 97>, <표 98>, <표 99>와 같다.

		발화음	음절 구조	유사음소	비고
초성	평음	ㄱ	CV	ㅋ	
			CVC		
		ㄷ	CV	ㅌ	
			CVC		
		ㅂ	CV	ㅍ	'ㅡ' 앞에서는 'ㄱ'도 고려
			CVC		
	경음	ㄲ	CV	ㄱ	
			CVC		'ㅏ' 앞에서는 'ㄸ'도 고려
		ㄸ	CV	ㄴ	
			CVC	ㅃ	
		ㅃ	CV	∅	산발적이나 'ㄲ'가 고려될 수 있음
			CVC		
	격음	ㅋ	CV	ㄱ	
			CVC		'ㅓ, ㅗ' 앞에서는 'ㅎ'도 고려
		ㅌ	CV	ㄷ	'ㅣ' 앞에서는 'ㅋ'가 유사음소
			CVC	ㅍ	
		ㅍ	CV	ㅋ, ㅎ	
			CVC		
종성		ㄱ	VC	∅, ㅂ	'ㅡ, ㅣ' 뒤에서는 'ㅇ'도 고려
			CVC		
		ㅂ	VC	∅, ㄱ	'ㅡ' 뒤에서는 'ㅇ'도 고려
			CVC		

<표 97> 파열음의 유사 음소

		발화음	음절 구조	유사음소	비고
초성	파찰음	ㅈ	CV	ㅊ	
			CVC		
		ㅉ	CV	ㅈ	
			CVC		
		ㅊ	CV	ㅅ	
			CVC		
	마찰음	ㅅ	CV	ㅆ	
			CVC		
		ㅆ	CV	ㅅ	
			CVC		
		ㅎ	CV	∅	
			CVC	∅, ㅅ	‘ㅣ’ 앞에서는 ‘ㅅ’가 유사음소
	공명음	ㄴ	CV	ㅁ	
			CVC	ㅁ, ㄸ	후설모음 앞에서는 ‘ㄸ’가 유사음소
		ㅁ	CV	ㄴ	
			CVC	∅, ㄴ	
		ㄹ	CV	ㄴ	
			CVC	∅, ㄴ	
종성		ㄴ	VC	ㅁ, ㅇ	
			CVC		
		ㅁ	VC	ㅇ	
			CVC		
		ㅇ	VC	ㅁ	
			CVC		
		ㄹ	VC	ㅇ	
			CVC	∅	

<표 98> 파열음 외 자음의 유사 음소

발화음	음절 구조	유사음소	비고
ㅏ	CV CVC	ㅓ	
ㅐ	CV CVC	ㅐ, ㅙ	활음 첨가
ㅓ	CV	ㅗ	
	CVC	ㅏ	
ㅗ	CV CVC	ㅜ, ㅓ	‘연구개음, 양순음’ 뒤에서는 ‘ㅜ’가 유사음소
ㅜ	CV CVC	ㅗ	
ㅡ	CV CVC	ㅜ, ㅣ, ㅐ	‘연구개음, 양순음’ 뒤에서는 ‘ㅜ’가 유사 음소 ‘경구개음’ 뒤에서는 ‘ㅣ’가 유사 음소 ‘치조음’ 뒤에서는 ‘ㅐ’가 유사 음소
ㅣ	CV CVC		관찰되지 않음.

<표 99> 단모음의 유사 음소

일상생활에서 쉽게 접할 수 있는 모바일 기기에 탑재된 음성 인식 비서나 인공지능 스피커, 음성을 통해 조작하는 내비게이션 및 통·번역기에는 인간의 음성을 텍스트로 변환하는 자동 음성 인식 처리 기술이 적용되어 있다. 본고에서는 이러한 음성 인식 처리 기술이 인식해야 하는 대상이 한국어 음성이고, 또 우리의 음성 특성이 인간을 대상으로 할 때와 인공지능을 대상으로 할 때 차이를 보일 수 있기에, 음성 인식 환경에서 나타나는 한국어의 발화상·지각상 특성에 관한 음운론적 관점의 연구가 필요하다고 판단하였다. 본고는 이러한 관점 하에 한국어에서 실현 가능한 모든 단음절 목록을 대상으로 하여 어떠한 음운론적 환경에서 음성 인식 오류가 주로 발생하는지, 또 이때 오인식되는 음소는 어떤 특성을 가지는지 관찰해보고자 하였으며 유사 음소의 개념을 통해 이를 정리하였다.

이상의 연구는 자동 음성 인식의 한국어 음성 인식 오류 문제를 음운론적 관점에서 분석한 첫 시도라는 점에서 다소간의 의의를 가질 수 있으나, 무엇보다도 중요한

음성 인식 오류의 발생 원인을 설명하지 못한 부분이 많다는 점에서 한계를 가진다. 본고에서 실험 대상을 단음절 구성으로 제한한 것은, 음성 인식 알고리즘이 발화된 음성의 음향 특성을 얼마나 정확하게 지각해낼 수 있는지 확인하고자 함이었으나, 이로 인해 탐색 엔진이 참고할 수 있는 맥락 정보가 차단되어 전체 인식률을 떨어뜨리는 요인이 되었다는 점도 본고가 가지고 있는 문제라 할 수 있다. 또한 본고의 실험 상황에서 음성 인식 알고리즘의 '인식' 과정뿐만 아니라, 음성 제공자들의 발화 '산출' 과정도 인식 오류에 영향을 주었을 가능성이 있으나, 이 부분을 면밀히 고찰하지 못한 것도 본 논의가 가진 한계이다. 추후에는 이와 같은 한계들을 바탕으로, 맥락 정보가 주어진 단음절 음성 인식 실험을 통해 탐색 엔진의 적극적인 개입으로도 해결되지 않는 음성 인식 오류 유형을 특정하는 연구와, 다음절 또는 문장을 대상으로 한 음성 인식 실험 연구들을 통해 한국어 음성 인식 오류에 관여하는 음운론적 요인들을 찾아나가야 할 것으로 보인다.

참고 논저

강요셉(2018), 확장 음소정보기반 Lattice-free Maximun Mutual Information 음향모델, 서강대학교 컴퓨터공학과 대학원 석사학위 논문.

고도흥(2003), 개별화자의 음성파라미터 추출에 관한 연구: 음성파라미터의 상관관계를 중심으로, 《음성과학》 10-2, pp. 129-143.

곽충구(2003), 현대국어의 모음체계와 그 변화의 방향, ≪국어학≫ 41, pp. 59-91.

김교진(2021), 능동 학습과 소형 플랫폼을 이용한 영역별 한국어 자동 음성 인식, Domain-specific Korean automatic speech recognition using active learning on a small generic platform, 충북대학교 정보통신공학 박사학위 논문.

김민제·이정철·김종진(2006), HMM기반 자동음소분할기의 음소분할 오류 유형 분석, 《한국음향학회지》 25-5, pp. 213-221.

김범승(2012), 한국의 기차역명 음성 인식을 위한 효율적 PLU와 상태 군집화에 관한 연구, 광운대학교 컴퓨터공학과 대학원 박사학위 논문.

김상훈(2016), 딥 러닝을 이용한 음성인식 오류 판별 방법, 과학기술연합대학원대학교 컴퓨터소프트웨어전공 석사학위 논문.

김선희(2003), 한국어 자동 발음열 생성 시스템을 위한 예외 발음 연구, 《말소리》 48, pp. 57-67.

김성규(1999), 빠른 발화에서 음절 수 줄이기, 《애산학보》 23, pp. 109-137.

김성규(2001), 음성언어층위와 문자언어층위의 위상에 대한 연구, 《언어학》 30, pp. 65-88.

김성규(2004), '워>오'의 통시적 고찰, 《국제어문》 30, pp. 5-26.

김성규(2015), 15세기 국어와 표준어의 음운론적 유사도 측정 시론, 《국어학》 75, pp. 3-37.

김성빈(2021), 음성-문자 간 교차-모드 지식 증류를 이용한 종단 간 방식의 음성언어이해, 인하대학교 전기컴퓨터공학과 전자공학 대학원 석사학위 논문.

김성아·육동석·권오일(2004), 음소 질의어 집합 생성 알고리즘, 《한국음향학회지》

23-2, pp. 173-179.
김성환·조환규(2012), PAM 행렬 모델을 이용한 음소 간 유사도 자동 계산 기법, 《한국콘텐츠학회논문지》12-3, pp. 34-43.
김아름(2021), 자동 음성인식 처리 기술의 한국어 파열음 인식 오류에 관한 연구, 《영주어문》 49, pp. 5-33.
김연화(2015), 다양한 전송 왜곡에 의한 한국어 음성의 지각특성의 변화에 관한 연구, 원광대학교 컴퓨터공학과 박사학위 논문.
김의수·정은주(2017), 음운론적 복잡성과 다양성 측정 기제 연구: 음소 및 음절 차원에서, 《우리말연구》 49, PP. 5-36,
김지은(2015), 노년층 파열음 발음의 재해석, 《말소리와 음성과학》 7-2, pp. 139-145.
김지은(2017), 한국어 폐쇄음 발음과 최근의 발음 변이: 발화 형태별 VOT와 f0를 중심으로, 《말소리와 음성과학》 9-3, pp. 41-47.
김진형·박소형(2012), 모바일 환경을 고려한 규칙기반 음성인식 오류교정, 《한국컴퓨터정보학회지》 17-10, pp. 24-33.
김현(2008), /ㅓ/의 음성 실현과 그 실현 조건, 《국어학》 52, pp. 3-25.
김현(2009), 음소의 변별적 자질과 음성 실현, 《한중인문학연구》26, pp. 187-208.
김현(2014), 어두와 비어두 'ㄷ, ㅌ, ㄸ'의 폐쇄 구간, 《국어학》72, pp. 183-203.
김현(2018가), 공시 음운론과 음성학, 《국어학》85, pp. 369-398.
김현(2018나), 국어의 치조 비음과 경구개 비음, 《송철의 선생 퇴임 기념 국어학 논총》, 태학사
김현(2019), 비음화의 발생: 모음 끝 전이와 비강의 선개방, 《국어학》92, pp. 37-62.
김현(2021가), 모음 연쇄의 음성적 실현, 《어문학》151, pp. 37-62.
김현(2021나), 양순음 뒤 'ㅡ'의 음성학: 원순모음화는 동화인가, 《국어학》 98, pp. 83-111.
김현·배윤정(2018), 단일 비음과 중첩 비음에 대한 음향적 및 지각적 연구, 《어문학》137, pp. 1-20.

김현·이영환(2015), 모음 간 /j/의 음성 실현과 지각 판단, 《국어학》 75, pp. 39-57.
김현우(2004), 음성 특징벡터 추출을 위한 적응적인 주파수특성의 필터뱅크, 경북대학교 전자공학과 석사학위 논문.
김현호(2016), 딥 러닝을 이용한 음성인식 오류 판별 방법, 과학기술연합대학원대학교 컴퓨터 소프트웨어전공 석사학위 논문.
김형주·이찬규(2019), 포스트휴머니즘의 저편 인공지능인문학 개념 정립을 위한 시론, 《철학탐구》 53, pp. 51-80.
노원빈(2015), 공통 특징 추출 기반 한글 모음 'ㅏ', 'ㅓ', 'ㅣ' 음성 인식 기법, 숙명여자대학교 대학원 석사학위 논문.
노원빈·이종우·이재원(2016), 파형 특징 추출과 신경망 학습 기반 모음 음성 인식, 《정보과학회 컴퓨팅의 실제 논문지》 22-2, 69-76.
노희경·이강희(2017), 구글, 네이버, 다음 카카오 API 활용앱의 표준어 및 방언 음성 인식 기초 성능평가, 《예술인문사회융합멀티미디어논문지》 7-12, pp. 819-829.
류혁수·나민수·정민화(2015), 한국인 화자의 외래어 발음 변이 양상과 음절 기반 외래어 자소-음소 변환, 《말소리와 음성과학》 7-3, pp. 139-149.
명주현·정민화(2000), 기능어용 음소 모델을 적용한 한국어 연속음성 인식, 《한국정보과학회 학술발표논문집》 27-1B, pp. 354-356.
박권식·김성태·송상헌(2021), 한국어 최소대립 문장쌍에 대한 딥러닝 실험, 《제81차 한국어학회·한국어문교육연구소 공동 전국학술대회 발표자료집》, pp. 7-21.
박준호(2014), 음소 유사도 평가 기반의 문자 오류교정 후처리 알고리즘, 한양대학교 전기및전자공학과 석사학위 논문.
박진호(2006), 자연언어처리와 국어 연구, 《국어국문학》 143, pp. 97-122.
박진호(2019), 딥러닝 기반 자연어 처리에서 도메인 지식의 역할, 《인공지능인문학연구》 4, pp. 135-166.
박한상(2007), 한국어 자음군의 후행모음에 나타난 발성유형의 음향음성학적 연구, 《말소리》 64, pp. 53-76.
방만식·장두봉·김영일(1995), 한국어 음성 신호에서의 음소의 유사도 측정, 《생산기술연구소논문집》 11, pp. 253-262.

방정욱(2020), 자유발화 음성인식을 위한 방송 데이터 정제 및 음소 집합 확장, 충북대학교 전기·전자·정보·컴퓨터학부 제어로봇공학전공 대학원 박사학위 논문.

배문정·김정오(2002) 한국어 자음의 지각 구조, 《인지 및 생물》 14-4, pp. 375-408.

백운일(2001), 한국어 자음의 긴장도 증가에 따른 조음점 이동에 관한 연구, 《음성과학》 8-3, pp. 115-120.

복거철(2020), 음성특징의 거리에 기반한 한국어 발음의 시각화, 《한국정보전자통신기술학회 논문지》 13-3, pp. 197-205.

성철재(2004), 한국어 단모음 8개에 대한 음향 분석: F1/F2 모음공간에서의 음향변수를 중심으로, 《한국음향학회지》 23-6, pp. 454-461.

소신애(2011), 국어의 'p/k 교체'에 대하여,《국어국문학》158, pp. 101-134쪽.

소신애(2019), 음절말 비음 간 대치 현상에 대하여, 《국어국문학》 188, pp. 33-70.

신우봉(2010), 순행동시조음에 의한 원순화 현상 연구, 고려대학교 대학원 석사학위 논문.

신우봉(2018), 한국어 /ㅓ/의 음성 실현 양상 연구: 서울코퍼스를 중심으로, 《겨레어문학》 60, pp. 233-258.

신지영(2008), 성인 자유 발화 자료 분석을 바탕으로 한 한국어의 음소 전이 빈도, 《언어청각장애연구》 13, pp. 477-502.

신지영(2014), 《말소리의 이해》(개정판), 한국문화사.

양병곤(2017), 대학생들이 또렷한 음성과 대화체로 발화한 영어문단의 구글음성인식, 《말소리와 음성과학》 9-4, pp. 43-50.

양순임(2008), 전설모음과 경구개 자음의 음운론적 공모성, 《인문학논총》 13-1, pp. 23-41.

오동훈(2019), Hierarchical Phoneme Classification for Improved Speech Recognition: 향상된 음성 인식을 위한 계층적 음소 분류, 경북대학교 전자공학부 대학원 석사학위 논문.

오상엽(2011), 바타챠랴 거리 측정 기법을 사용한 가우시안 모델 기반 음소 인식 향상, 《멀티미디어학회 논문지》 14-1, pp. 85-93.

오상엽(2013), MFCC와 LPC 특징 추출 방법을 이용한 음성 인식 오류 보정, 《디지털융복합연구》 11-6, pp. 137-142.

오상엽(2021), 음성 인식을 위한 개선된 평균 예측 LMS 필터를 이용한 DNN 기반의 강인한 음성 특징 추출 및 신호 잡음 제거 기법, 《융합정보논문지》11-6, pp. 1-6.

오준석(2019), Weakly Labeled Corpus를 이용한 한국어 실시간 자동 전사 시스템, 서강대학교 컴퓨터공학과 석사학위 논문.

유재원·이경미(2013), 한국어 유아 음성인식을 위한 수정된 Mel 주파수 캡스트럼, 《한국콘텐츠학회논문지》 13-3, pp. 1-8.

유현재·김명화·박상길·김광용(2020), 클라우드 기반의 음성인식 오픈 API의 응용 분야별 한국어 연속음성인식 정확도 비교 분석, 《한국통신학회논문지》 45-10, pp. 1793-1803.

윤규철(2019), 영어와 한국어 자연발화 음성 코퍼스에서의 무성 파열음 연구, 《말소리와 음성과학》 11-4, pp. 45-53.

윤영선·강점자(2009), 레벤스타인 거리에 기초한 위치 정확도를 이용한 고립 단어 인식 결과의 비유사 후보 단어 제외, 《말소리와 음성과학》 1-3, pp. 109-115.

이경님(2017), 음성 언어 처리 기술, 어디까지 왔나, 《새국어생활》 27-4, pp. 99-116.

이경희 이봉원(1999), 한국어 평마찰음과 경마찰음의 음향적 특성: 길이를 중심으로, 《한국어학》 10, pp. 47-66.

이경희·정명숙(2000), 한국어 파열음의 음향적 특성과 지각 단서, 《음성과학》 7-2, pp. 139-155.

이수길(2005), 한국어 음성에 대한 음소분석에 관한 연구, 배재대학교 컴퓨터공학과 박사학위 논문.

이재강(1998), 한국인 화자의 한국어 모음에 관한 실험음성학적 분석, ≪언어연구≫ 17, pp. 41-56.

이재원(2015), 한국어 음성 인식에서 변동성과 벌크 지표에 기반한 음소 경계 검출,

《정보과학회 컴퓨팅의 실제 논문지》 21-10, pp. 631-638.
이재원(2017), 벌크 지표의 신경망 학습에 기반한 한국어 모음의 음성 인식, 《정보과학회 컴퓨팅의 실제 논문지》 23-11, pp. 617-624.
이재원(2018), 시간 영역 음소 특징 지표를 이용한 유성 자음 인식, 《정보과학회 컴퓨팅의 실제 논문지》 24-10, pp. 554-562.
이주현(2005), 한국어 음소의 주파수 특성에 관한 연구, 한림대학교 학과간협동과정 언어청각학전공 석사학위논문.
이지현·이재원·이종우(2013), 부호 분포 변동성에 기반한 한국어 음소, ‘ㅅ’, ‘ㅈ’, ‘ㅊ’ 음성 인식, 《정보과학회 논문지: 컴퓨팅의 실제 및 레터》 19-7, pp. 377-382.
이진호(2004), 국어 활용형에 적용되는 모음 축약, 《어학연구》40-1, pp. 177-194.
이진호(2005), 國語의 音韻論的 制約 體系, 《어문연구》33-2, pp. 59-82.
이진호(2006), 음운 규칙의 공시성을 바라보는 시각, 《국어학》47, pp. 39-64.
이진호(2009), 《국어 음운 교육 변천사》, 박이정.
이진호(2012), 국어 PK-교체에 대한 종합적 고찰, 《국어학》63, pp. 247-273.
이진호(2013), 국어 평파열음과 비음의 상호 대응에 대한 통시적 고찰, 《진단학보》 117, pp. 235-257.
이진호(2017), 언어 유형론 연구에서의 평균값 도출 방안: 음운 목록을 중심으로, 《건지인문학》19, pp. 221-240.
이진호(2020), 한국어 파열음 체계의 언어 유형론적 고찰, 《어문연구》48-2, pp. 5-29.
이창수(2014), 음성 대화 인식 모듈을 위한 미등록어 처리 및 동시 인식 기술, 동아대학교 컴퓨터공학과 대학원 석사학위 논문.
이창영(2010), 남성과 여성의 음성 특징 비교 및 성별 음성인식에 의한 인식 성능의 향상, 《한국전자통신학회논문지》 5-6, pp. 568-574.
이현복(1977), 서울말과 표준말의 음성학적 비교연구, ≪언어학≫ 2, pp. 167-184.
이호영(1996), 국어음성학, 태학사.
이호영·지민제·김영송(1993), 동시조음에 의한 변이음들의 음향적 특징, 《한글》 220.

임영춘·오세진·김광동·노덕규·송민규·정현열(2003), 음성인식에서 문맥의존 음향모델의 성능향상을 위한 유사음소단위에 관한 연구, 《한국음향학회지》 22-5, pp. 388-402.

장한솔(2020), 자질 정보 임베딩을 결합한 신경망 음소 모델, 서울대학교 언어학과 언어학전공 석사학위 논문.

장혜진·신지영(2006), 대구 방언 단모음의 세대 간 차이에 대한 음향・음성학적 연구, ≪말소리≫ 57.

장혜진·신지영·남호성(2015), 서울 방언 단모음의 연령별 실현 양상, 《음성·음운·형태론 연구》 21-2, pp. 341-258.

전상범(2005), 《영어음성학개론》, 을유문화사.

전진호(2021), 한국어 모음-자음 연쇄에서의 F2 전이 연구, 서울대학교 대학원 석사학위 논문.

정유남·강아름(2021), 인공지능 번역기에 나타난 의미 오류 고찰, 《제81차 한국어학회·한국어문교육연구소 공동 전국학술대회 발표자료집》, pp. 61-83.

정윤자(2010), 후행모음의 피치를 고려한 한국어 평음, 경음, 격음의 인지 실험, 《우리말연구》 27, pp. 73-94.

정일진(1997), 표준어 단순 모음의 세대간 차이에 대한 실험음성학적 분석 연구, ≪말소리≫ 33·34, pp. 111-125.

조남민(2015), 어두 유음의 음성적 실현에 관한 연구: 인접 모음과의 상관성을 중심으로, 《언어과학연구》 75, pp. 289-310.

조성문(2003), 현대 국어의 모음 체계에 대한 음향음성학적인 연구, ≪한국언어문화≫ 24, pp. 427-441.

조희련·이유미·임현열·차준우·이찬규(2021), 딥러닝 기반 언어모델을 이용한 한국어 학습자 쓰기 평가의 자동 점수 구간 분류: KoBERT와 KoGPT2를 중심으로, 《한국언어문화학》 18-1, pp. 217-241.

주왕산(1948), 《말의 소리》, 동양사.

지승은(2017), 효과적인 음성 인식 평가를 위한 심층 신경망 기반의 음성 인식 성능 지표에 관한 연구, 인천대학교 컴퓨터공학과 석사학위 논문.

진양·이찬규(2021), 중국 감 방언권 학습자의 한국어 어두 파열음 음성적 특징 연구, 《어문론집》 85, pp. 475-500.

채서영(2000), 서울말의 비어두 모음 /오/의 상승현상: 모음조화의 이화나 중화 혹은 독립적인 변화, ≪서울말 연구≫ 1, 박이정.

채서영(2001), 서울말 전설모음 /e/의 변이현상: 사회언어학적 조사와 기능적 분석, 《어학연구》 37-3, pp. 655-674.

최승주·김종배(2017), 음성 인식 오픈 API의 음성 인식 정확도 비교 분석, 《예술인문사회융합멀티미디어논문지》 7-8, pp. 411-418.

한동조·유제만·최기호(2006), 음소 오류확률모델을 이용한 핵심어 오류 검출 및 보정에 관한 연구, 《한국멀티미디어학회 학술발표논문집》 2006-1, pp. 25-28.

한동조·최기호(2007), 음성인식 후처리에서 음소 유사율을 이용한 오류보정에 관한 연구, 《한국ITS학회논문지》 6-3, pp. 77-86.

허성필·정현열·김경태(1995), 음소판별필터를 이용한 한국어 단음절 음성인식, 《한국음향학회지》 14-1, pp. 31-39.

홍윤석(2019), 한국어 발화에서 사용되는 음소 인식을 위한 CTC 알고리즘 기반의 음소 인식 모델 연구, 서울대학교 융합과학기술대학원 석사학위 논문.

홍은영(2012), 한국어 전설고모음화 현상 연구, 서울대학교 대학원 석사학위 논문.

홍혜진·김선희·정민화(2008), 한국어 음성인식을 위한 음성학 기반의 유사음소단위 집합 설계, 《말소리》 65, pp. 105-124.

황미경·조남민(2020), 한국어 후설모음에 대한 음향음성학적 연구: 선행 인접 자음과의 상관성을 중심으로, 《언어과학》 27-4, pp. 122-145.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N.Gomez, Ł. Kaiser, and I. Polosukhin(2017), “Attention is all you need,” Advances in neural information processing systems, vol. 30, pp. 5998 - 6008.

Daeun Lee · Joon S. Lim(2020), ADHD Classification Based on Deep Learning Using MFCC Coefficients, 《한국인공지능학회 2020 하계 및 추계 학술대회 논문집》, pp. 105-108.

G. Evermann et al.(2005), “Training LVCSR systems on thousands of hours of

data," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing vol. 1, pp. I/209-I/212.

Hyunsook Kang · Jeong-Im Han(2018), Perceptual Cues for the Word-Initial Nasel Stops in Korean, 《인문사회21》 10-1, pp. 33-43.

Jeong-Young Song · Kill-Soo Lee(2017), A Phonemic Classification for Speech Recognition, 《The Journal of Engineering》 19-1, pp. 16-2

Kang, Yoonjung(2014), Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study, Journal of Phonetics 45, 76~90.

Kim, Mi-Ryoung(2008), "Lax" stops in Korean revisited: VOT neutralization, 《음성음운형태론연구》 14-2, pp. 3-21.

Oh, Mira(2002), Place perception in Korean consonants, 《음성과학》9-4, pp. 131-142.

R. Jakobson · C. Fant · M. Halle(1951), "Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates", The M.I.T PRESS, Boston.

Satoshi Kaki, Eiichiro Sumita, and Hitoshi Iida(1998), "A method for correcting speech recognition using the statistical features of character co-occurrence", International Conference On Com-putational Linguistics, vol.1, pp. 653-657.

Yoon, Kyuchul(2002), A production and perception experiment of Korean alveolar fricatives, 《음성과학》 9-3, pp. 169~184.

<Abstract>

A Phonological Study on the Error Types of Korean Automatic Speech Recognition

Kim Areum

Department of Korean language and literature

The Graduate School

Seoul National University

The purpose of this dissertation is to explain the phonological factors that affect errors of Korean automatic speech recognition and to classify the patterns of speech recognition errors through characteristics of phoneme analysis. A speech recognition experiment was conducted with all pronounced monosyllables as the subject of the study. When there is a tendency to misrecognize one phoneme as a specific phoneme in an environment where sound recognition errors occur, I suggested the notion of 'similar phoneme', then I presented similar phonemes at each phoneme.

I described recognition errors of initial consonants in chapter 3. Recognition rate of fricatives and sonorant, which the frequency band is located in the extreme, is high. However, recognition rate of plosive is very low because they are close to silence and there are few additional clues. Recognition of initial consonant tend to make errors in the environment that phonetic variants where vowel follows, and in which sound characteristics are confused between consonants. Plain sound is mainly confused with aspirated sound, and this error

can be understood in the relationship with the F0 characteristic according to gender. It is estimated that the low F0 value of men causes misrecognition from aspirated to plain sound, and the high F0 value of women causes misrecognition from plain to aspirated sound. For fortis and aspirated, misrecognition between fortis and aspirates mainly appears. This means that the phoneme is relatively accurate in that it is fortis and aspirated. Fortis and aspirated sounds have distinctive characteristics such as 'tension' and 'aspiration', making it easier to recognize than plain sounds. Like fortis and aspirated sounds, fricatives and affricatives are perceived relatively clearly, however, in a high vowel environment, when the discrimination of the friction section is weakened, or friction and aspiration are not properly discriminated, an error occurs between themselves.

I described recognition errors of vowels in chapter 4. Among the monophthongs, the vowel with the highest recognition rate is 'ㅣ', and the vowel with the lowest recognition rate is 'ㅜ', and in general, the recognition rate of the back vowels appears to be low. Monophthongs cause recognition errors due to the influence of the articulation position of the preceding consonants, or errors due to phonetic and articulation spatial differences in vowels according to gender. Recognition errors related to articulation positions are mainly observed in vowels 'ㅡ'. In modern Korean, 'ㅗ' and 'ㅜ' are considered to be vowels with a large difference in articulation space according to generation and gender, of which the misrecognition of 'ㅗ' as 'ㅜ' is significant in women. This is regarded as an error due to the difference in articulation space of the round vowel according to gender. The misrecognition of 'ㅓ' as 'ㅏ' is also a significant error in women, which is concerned with the difference in F1 characteristics between men and women.

In the recognition of diphthongs, the pronunciation characteristics in natural speech affect the recognition of j-combined diphthong. In addition to the

existence of initial consonants, the probability of the appearance of w-combined diphthong based on the frequency of use influenced the recognition error.

I described recognition errors of final consonants in chapter 5. In the case of final sound, the recognition rate is lowered in the order of lateral, nasal, and plosive and recognition errors between plosives and nasals mainly occur. In the case of final consonants, many errors in not recognizing the existence of phonemes themselves are observed. This is an error due to the lack of clear recognition of the boundary between the formant trasition or preceding vowels and final consonants.

Based on the above discussion, I describes the phonological types of speech recognition errors in chapter 6. Speech recognition errors appear to occur when they deviate from the typical range of acoustic features of each phoneme set by the speech recognition algorithm, or when a sound at its boundary is detected. In this case, the misrecognized phoneme is determined by which of the various acoustic characteristics defining phoneme is beyond the limit. At this time, a specific phonological environment may be involved in causing the characteristics of the spoken sound to deviate from the limitations of the corresponding phoneme, or errors between phonemes with similar acoustic characteristics may cause recognition without specific conditions. I classified speech recognition errors caused by factors such as the former as 'recognition errors due to phonetic variation' type, and the latter as 'recognition errors due to acoustic factors'. In addition, I supposed that the phoetic and phonological characteristics of male and female speech also affected speech recognition errors, which was set as the type of 'recognition errors due to gender factors'.

Keywords: Korean speech recognition, speech recognition errors, similar phonemes, speech recognition based on gender, speech recognition based on variants