

음성 인식 인식을 향상을 위한 단어 평가에 관한 연구

박규영, 김기돈*, 강애띠*, 김경배

서원대학교, 인포씨드*

pkyong0419q@naver.com, kdkim@infoseed.kr, dlsdnr64@naver.com, gbkim@seowon.ac.kr

A Study on Word Evaluation to Improve Speech Recognition Recognition Rate

Park Gyu Yeong, Kim Ki don, Kang Ae Tti, Kim Gyoung bae

요 약

본 논문은 음성 인식 인식을 향상을 위한 단어 선별을 통해 단어의 인식성, 발음의 난이도 등 단어에 대한 전반적인 평가를 진행한 후 실제 음성 인식을 수행하고 음성 인식의 정확성 및 성공률을 집계하여 격자 기반 주소체계 이외에도 음성 인식을 이용하는 기술에서 사용하기 쉬운 단어에 대한 설명, 발음 난이도와 실제 인식률의 일치 여부가 다른 경우를 분석하였다.

I. 서론

컴퓨터가 사람의 말을 인식하는 과정은 일종의 패턴인식 과정으로 볼 수 있다. 즉, 사람이 발성하는 특정 단어의 신호를 패턴화하여 컴퓨터 메모리 안에 기억시킨 다음 새로운 음성이 입력되면 입력된 음성이 저장되어 있는 패턴들 중 어느 것과 가장 유사한 지를 판단해 내면 되는 것이다[1].

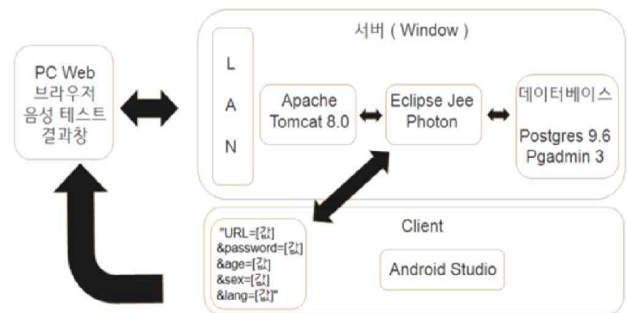
알고리즘, 혹은 기계가 인식한 발음을 분석하여 유사한 패턴을 판단하는 방법은 유사 단어의 존재로 인해 인식률이 떨어지는 원인이 되고 있으며, 이는 계속되어 온 알고리즘의 수정 및 보완으로 인해 인식률을 높여주고 있지만, 그것이 모든 단어의 완벽한 인식을 의미하지는 않는다.

본 논문에서는 음성 인식 인식률의 향상을 위해 알고리즘의 개선이 아닌 실제 단어들의 테스트 결과를 이용하여 정확하게 음성 인식이 된 단어들을 선별하여 이용함으로써 격자 기반 주소체계에서 사용될 단어뿐 아니라 음성 인식 기술을 사용하는 다른 분야에서도 접목하여 사용할 수 있도록 그 결과를 분석하였다.

II. 본론

음성 인식 과정에 있어 실험자들이 발음 난이도를 쉽고 책정하지 않은 경우에도 인식률이 높거나 보편적으로 발음의 난이도가 쉬운 편에 속한다고 생각됨에도 인식 정확도가 떨어지는 경우가 발생하여 테스트 내용을 종합하여 분석하게 되었다.

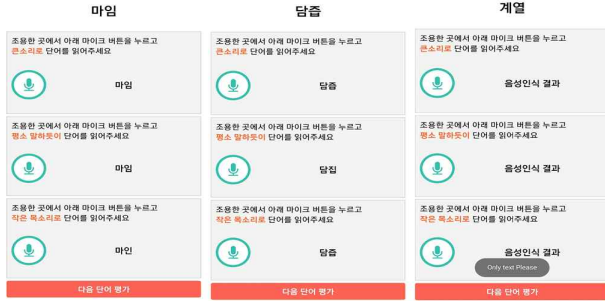
음성 인식 평가에 사용된 개발 도구는 Tomcat 8.0, Eclipse Jee Photon, Postgres 96, Pgadmin 3, Android Studio 3.5.0, SDK 19~28, JDK - 1.8이며, 데이터베이스에서 순서에 해당하는 단어를 발음하면 현재 읽어야 할 단어와 인식된 단어를 비교하여 값을 저장한 뒤 해당 결과를 서버에 저장하는 방식으로 진행되었다. 그 구조는 다음과 같다.



[그림 1] 개발 툴, 설계도

음성 인식 프로그램으로는 안드로이드 음성 인식 기술인 Google API를 사용하였는데, 더 좋은 성능과 알고리즘을 지닌 다른 음성 인식 프로그램이 존재함에도 해당 프로그램을 사용한 이유는 조금 모자란 성능을 가지고 있는 음성 인식 프로그램에서 단어를 선별해야 추후 더 좋은 프로그램을 장착한 완성품에서의 성능이 보장되기 때문이다.

단어 평가 내용으로는 표준어인지를 묻는 내용부터 사용빈도, 발음의 난이도, 복합어인지 등 16개의 문항이 있었으며, 그중 주목할 내용으로는 사용빈도와 발음의 난이도 두 문항이 가장 주목할 내용인데 특히 발음의 난이도 부분에 주목하였다. 이는 실험자가 생각하는 발음의 난이도와 실제 음성 인식을 진행했을 때 제대로 인식이 되지 않는 경우가 생각보다 잦아 주목하게 되었다.



[그림 2] 단어들의 음성 인식 테스트

‘계열’과 같은 모음이 복합적으로 작용하는 단어는 누구나 발음이 어려울 것으로 예상하고, 실제로도 발음에 꽤 집중하여야 하는 단어이다. 그러나 ‘마임’과 같은 단어는 모음이 복합적으로 작용하는 것도 아니며 받침도 유성음이라 부드럽게 읽혀 발음도 쉬울 것으로 예상하지만 주의를 기울여 발음하지 않으면 실제 인식되는 과정에서 ‘마인’처럼 받침이 다른 자음으로 바뀌어 인식되는 경우가 생긴다. 예를 들어 격자 기반 주소체계의 경우를 살펴본다면 장점인 ‘지도를 보고 위치를 쉽고 정확하게 말하고 공유할 수 있다’라는 점에서 ‘외진 곳에서 조난을 당하게 됐을 때도 빠르고 정확하게 위치를 전달해 구조 요청을 할 수 있다’라는 항목을 볼 때, 보편적으로 생각하는 발음의 난이도가 낮고, 단어 구조상 간단한 단어임에도 위급 상황에서 오인식(misrecognized, 誤認識)되어 잘못된 좌표가 표기된다면 사고로 번질 위험이 생긴다. 위급 상황이 아니더라도 오인식되어 반복하여 발음해야 인식되는 경우가 빈번하다면 기술과 취지가 뛰어나고 적용 범위가 넓더라도 불편함이 생길 수 있다.

| No | 단어 | 평가점수 | 음성점수 | 음성평가 | No | 단어 | 평가점수 | 음성점수 | 음성평가 |
|----|-----|------|---------|-------|------|-----|------|---------|-------|
| 1 | 뽕뽕 | 5494 | 100.00% | 30/30 | 4126 | 압식 | 8318 | 100.00% | 30/30 |
| 2 | 인디언 | 5496 | 100.00% | 30/30 | 4127 | 원액 | 8318 | 100.00% | 30/30 |
| 2 | 동원 | 5688 | 100.00% | 30/30 | 4128 | 판독 | 8318 | 100.00% | 30/30 |
| 2 | 적주 | 5848 | 100.00% | 30/30 | 4129 | 심신 | 8318 | 93.00% | 28/30 |
| 2 | 꼭지 | 5916 | 100.00% | 30/30 | 4130 | 침투 | 8318 | 100.00% | 30/30 |
| 6 | 끓주 | 5940 | 93.00% | 28/30 | 4131 | 수혈 | 8318 | 100.00% | 30/30 |
| 7 | 팔도 | 6014 | 100.00% | 30/30 | 4132 | 체질 | 8318 | 93.00% | 28/30 |
| 9 | 하프 | 6034 | 100.00% | 30/30 | 4133 | 차관 | 8318 | 100.00% | 30/30 |
| 9 | 야옹 | 6044 | 100.00% | 30/30 | 4134 | 풍족 | 8318 | 100.00% | 30/30 |
| 10 | 야옌 | 6056 | 100.00% | 30/30 | 4135 | 담변 | 8318 | 100.00% | 30/30 |
| 11 | 눈 | 6130 | 100.00% | 30/30 | 4136 | 군마 | 8318 | 93.00% | 28/30 |
| 12 | 능금 | 6160 | 100.00% | 30/30 | 4137 | 판자 | 8318 | 93.00% | 28/30 |
| 13 | 콩밭 | 6180 | 100.00% | 30/30 | 4138 | 상승세 | 8318 | 100.00% | 30/30 |
| 14 | 마운드 | 6194 | 93.00% | 28/30 | 4139 | 영점 | 8318 | 93.00% | 28/30 |
| 15 | 키다리 | 6206 | 100.00% | 30/30 | 4140 | 직물 | 8318 | 93.00% | 28/30 |
| 16 | 스윙 | 6208 | 100.00% | 30/30 | 4141 | 관문 | 8318 | 100.00% | 30/30 |

[사진 3] 단어 평가 결과 일부

8192개의 단어를 10명의 테스터를 통하여 단어의 이미지, 음성 인식 인식률을 분석하였다. 이에 따른 결과로 8192 단어 중 2739개의 단어가 음성 인식을 부분에서 100% 인식되지 않는 것으로 분석되었다. 이는 발음이 실패한 단어들로서, 이 중 ‘사사’ 등 발음에 불편이 없는 단어가 포함된 경우는 실험자 실수로 넘어간 경우이기에 완벽히 신뢰할 수 있는 자료가 아닐 수 있으나, 전체적으로 볼 때 ‘ㄱ’, ‘ㄴ’, ‘ㄷ’, ‘ㄹ’ 등 둘 이상의 모음이 복합적으로 작용하여 소리를 내야 하는 단어의 경우 특히 ‘only one text’라는 실패 문구가 자주 발생하는 경향을 보여 음성 인식 도중 둘 이상의 단어로 분할하여 오인식하여 음성 인식 자체가 실패한 경우가 상당

부분 차지하는 것으로 보인다. 또한, [그림 3]에서는 나타나지 않는 지표로서 음성 인식은 되었으나 오인식된 경우들이 존재하는데 이는 [그림 2]에 나타난 것처럼 ‘마임’이 ‘마인’으로 인식되거나 ‘심령’이 ‘신명’으로 인식되는 등 유성음 부분에서 주의를 기울여야 하는 부분들이 계속해서 있었고, ‘각막’ 등의 단어에서 일어나는 자음 동화, 그리고 유사어의 존재 역시도 단어 평가를 지속하는 과정에서 오인식의 가능성을 계속해서 내포하고 있어 발음에 유의하게 되었다.

모음이 복합적으로 작용하는 경우와 변하기 쉬운 유성음을 제외하면 ‘자두’와 같이 받침이 없는 경우, 혹은 ‘응급’처럼 자음 동화가 일어나지 않으면서 음절의 끝소리 규칙에 걸맞아 표기와 발음이 일치하는 경우, ‘눈썹’과 같이 된소리로 인해 발음에 자연스레 유의하게 되는 경우 정확한 음성 인식이 될 확률이 높아지는 경향을 보였다.

그러므로 단어를 선별함에 있어 모음이 복합적으로 작용하는 경우, 자음동화, 유성음 부분을 유의하여 선별한다면 알고리즘의 개선 없이 유의미한 인식을 상승을 기대할 수 있다.

III. 결론

단어의 인식률을 높이기 위해 본 논문에서는 Google API를 이용해 총 10명의 실험자를 통해 8192개의 단어를 평가 및 음성 인식 테스트하였고, 분석 결과로서 유사어 이외에도 ‘ㄱ’, ‘ㄴ’ 등 발음 자체가 어렵고 중첩되어 하나의 단어로써 인식이 실패하는 경우와 단어 자체를 발음하며 생기는 음운 변동 현상에 의해 발음이 변질되며 오인식되는 경우를 발견할 수 있었다. 특히 음운 변동 현상은 실험자가 발음의 난이도가 높다고 생각되는 단어보다 발음이 쉽다고 느끼는 ‘ㄱ’, ‘ㄴ’, ‘ㅇ’, ‘ㄹ’ 같은 유성음에서 자주 발생하는 경향을 보여 실험자가 생각하는 발음의 난이도와 실제 음성 인식 결과가 다를 수 있음을 시사하였다. 오인식 될 가능성이 큰 단어 이외에 사용에 적합한 단어를 찾는 과정에서 받침이 없거나, 음절의 끝소리 규칙에 맞는 경우 인식이 비교적 정확하게 이루어지는 경향을 보였으며, 또한 된소리로 인해 자연스레 발음에 유의하는 경우에도 정확한 인식이 되는 경우가 많았다. 해당 분석 결과는 음성 인식 기술을 포함하는 사업에서 활용하여 알고리즘 개선 없이도 좋은 효율을 얻을 수 있는 단어를 선별하여 인식률의 향상을 꾀할 수 있는 효과를 기대한다.

ACKNOWLEDGMENT

본 연구는 2019년도 중소벤처기업부의 기술개발사업 지원에 의한 연구임 [S2768125]

참 고 문 헌

- [1] 김화린. (2003). 음성인식 기술. 한국멀티미디어학회지. pp. 17-17.
- [2] 김동영, 김재민, 이민영, 김기돈, 이덕규, 김경배. (2020). 격자 기반 주소체계를 활용한 주소 정밀성 향상에 관한 연구. 한국통신학회 학술대회논문집, (), 333-33
- [3] Infoseed <https://cobak.co.kr/community/17/post/1298954>
- [4] 국립국어원 https://opendict.korean.go.kr/dictionary/view?sense_no=711577