

노인 음성인식을 위한 전처리에 관한 연구

박지웅, 이승준, 권순일¹⁾
세종대학교 디지털콘텐츠학과
e-mail:skwon@sejong.edu

A Study on Preprocessing for Elderly Voice Recognition

Ji-Woong Park, Seung-Jun Lee, Soonil Kwon
Dept. of Digital Contents, Sejong University

요 약

고령화 되어 가는 현대 사회에서 노인들이 일반 성인과 동등한 수준에서 정보를 접근 가능하도록 스마트기기의 손쉬운 인터페이스 방법이 요구된다. 음성 인터페이스는 노인들의 스마트기기 활용도를 높여 줄 수 있지만, 성능이 평균적 성인연령 대의 발성행태에 최적화되어 있어, 노인들이 사용할 경우 음성인식을 저하를 초래한다. 그래서 노인 친화형 음성 인터페이스를 개발하기 위한 일환으로 노인음성에 대한 인식률을 향상시켜 줄 수 있는 전처리 알고리즘을 개발하고자 한다. 이를 위해 노인층과 청년층을 대상으로 음성샘플을 수집하여 분석하였고, 그 결과 노인이 청년에 비해 발성속도가 느리며 이는 스마트기기의 음성인식 기능저하로 이어진다는 것을 확인할 수 있었다.

1. 서론

전 세계 선진국들의 고령 인구는 지속적으로 비중을 높여가고 있으며, 특히 우리나라는 역사적으로 전례 없이 빠른 속도로 증가하고 있다. 이러한 고령화 시대에 급격한 IT 기술의 발전은 세대 간의 접근성에 있어 오히려 심각한 불균형을 초래하여 사회적 소통부재로 인한 노인의 상실감과 우울감을 심화시키는 요인이 되고 있다. 고독사를 비롯한 노인자살률은 2007년 기준 OECD국가 중 한국이 1위라는 불명예를 안게 되었다. 이러한 문제점을 해결하기 위해 지역사회 복지시설 등을 통해 전산교육 및 스마트기기 교육 등이 활발하게 이루어지고 있지만 아직까지 노인들이 스마트기기를 활용하기가 쉽지 않다.

최근 스마트 기기가 지원해 주고 있는 인터페이스에는 터치/음성/동작 등이 있다. 이런 다양한 인터페이스 방법들 중에서 실버세대가 스마트 기기를 사용하는데 가장 용이한 것은 음성인터페이스, 즉 음성인식일 것이다. 하지만 음성인식이 평균적인 연령대의 사용자들에게 맞추어져 성

능이 평가 및 지원되고 있어 실버세대가 이용하는데 불편을 초래하고 있다.

현재 스마트 기기가 지원해 주고 있는 음성 인터페이스에는 청·장년층의 평균적인 발화속도를 기준으로 하여 이에 최적화된 음성인식 방법을 채택하고 있기 때문에 발성속도의 편차가 조금이라도 커지기 시작하면 음성인식 성능이 저하되는 현상을 유발할 수 있다. 노년기에는 혀 움직임의 범위(range)나 지속성(duration), 두께(thickness) 등이 감소되며[1], 이로 인해 조음을 하는 동안 구강의 피드백이 어느 정도 감소되게 되므로 조음기능에 일부 영향을 미칠 수 있다[2]. 즉 노인들의 말속도가 느려지고, 묵음 구간이 증가하며, 말소리를 산출하는데 있어 부정확성(imprecision)이 증가한다[3].

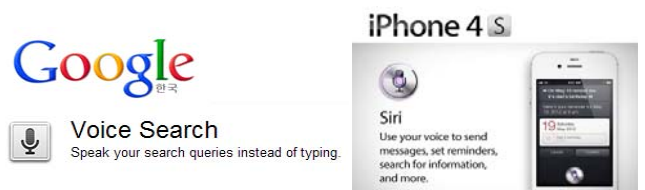
이번 연구에서는 노인들의 발성능력의 저하를 감안하여 이를 보정하기 위한 전처리를 수행함으로써 기존 스마트기기의 노인에 대한 음성인식 성능을 향상시켜주고자 한다. 이는 급속한 정보사회화에서 소외되었던 노인/장애인 계층에게 스마트기기의 손쉬운 접근 수단을 제공하여 정보소외계층이 일반인과 거의 동등한 수준에서 정보에 접근 가능하도록 하는 매개체가 될 것이다.



(그림 1) 스마트기기를 이용한 노인의 사회적 상호작용의 예시

1) 교신저자

주소: 143-747 서울시 광진구 군자동 98
연락처: +82-2-3408-3847



(그림 2) 대표적인 스마트기기의 음성인터페이스 시스템

2. 기존 연구

성대나 후두의 문제가 있는 노인에 대한 연구에 비해 정상 노인에 대한 연구는 부족한 실정이다. 일반적으로 노년기에 접어들면 호흡기의 구조와 폐 용량, 통풍, 가스 교환 등 기능에 변화가 발생하게 되며, 이런 이유로 75세 이상 노인들은 숨을 쉴 때 마다 말할 수 있는 음절수가 청년층에 비해 현저하게 적어진다[4].

노인 남성의 경우 40-50세까지는 Vocal Pitch가 낮아지고 그 이후에 다시 상승하나, 여성은 나이가 들에 따라 Vocal Pitch가 떨어지는 경향을 보인다[5]. 반면에 노년기 남성의 Pitch 경향은 같으나, 여성의 Pitch는 하강하다가 80대에 다소 상승한다는 연구도 있다[6]. 또한 /b/와 /w/ 발음을 표현하는데 있어서 나이가 들어감에 따라서 조음 시간이 눈에 띄게 길다. 그러나 전달이나 지속시간에서는 나이에 따른 영향은 적은편이다[7]. 다른 연구에서는 65세 이상의 성인은 입술과 혀의 움직임에 대한 교대운동속도의 감소, 연구개음 제어의 가벼운 손실이 나타난다[8][9].

노인층이 청년층에 비해서 정상적으로 유창 휴식(Fluency Breaks)이 증가하는 추세가 나타나며, 특히 형식적인 유창 휴식과 삽입어, 감탄사가 더 많이 사용된다[10]. 또한 청년층이 노년층에 비하여 더 긴 MPT(Maximum Phonation Time)를 보였고, 이는 정상 성인이라도 연령이 증가함에 따른 모음의 연장 수행 능력이 낮아진다는 것을 나타낸다[11].

노년기에 해부학적인 변화로 jitter와 shimmer를 측정 한 결과, 노인남성은 jitter와 shimmer가 증가하였고, 노인 여성의 경우 jitter만 유의하게 증가하였다. 이는 후두조절 기능의 감소나 후두조직의 퇴행성 변화를 나타내며, 연령 증가에 따른 발성의 불안정성이 증가된다[6].

위의 연구들로부터 우리는 노인의 음성 특징으로 유창 휴식과 조음시간의 증가와 성대와 후두부의 퇴화되는 것을 알게 되었으며, 이는 발성에 영향을 미쳐 일정 시간 내에 답해야하는 음성인식의 정확성을 저해하는 요소가 된다. 따라서 우리는 음성인식의 전처리 과정에서 청·장년층과의 다른 요인들의 보정을 통해 청·장년층과 노인층의 음성인식 성공률을 비슷한 수준까지 향상시키고자 한다.

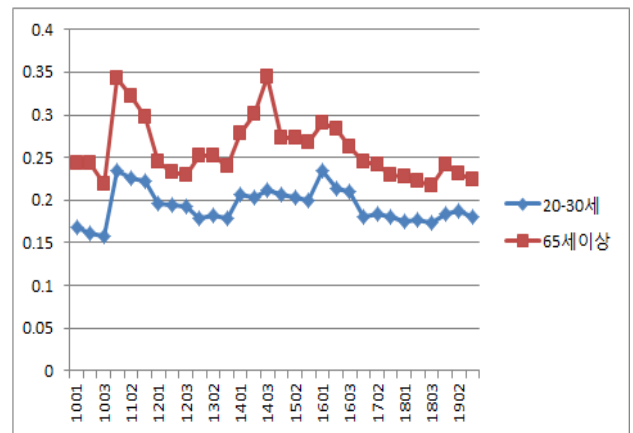
3. 실험 결과

DB녹음을 위한 대상자는 시각장애, 언어장애 신경 운동 질환 및 성대 질환 경험이 없는 20-30세(10명), 65세 이상(10명) 남녀를 대상으로 하였다. 비유창성을 유발할 가능성을 최소화시키기 위하여 실험 전 1분여 동안 단어들을 대상자에게 노출시켜 연습할 수 있도록 한 후 녹음을 실시하였고, 실험 수행 시 특정 부분을 강조하거나 과장해서 읽지 않도록 지시하였다. 단어는 기존의 언어학적 논문을 참고하여, 노인이 상대적으로 발음하기 힘든 ‘ㅍ’, ‘ㅌ’, ‘ㄱ’ 발음이 함유된 단어를 선정하였고, 입에서 20cm 간격으로 SAMSON Condenser Microphone이 위치하도록

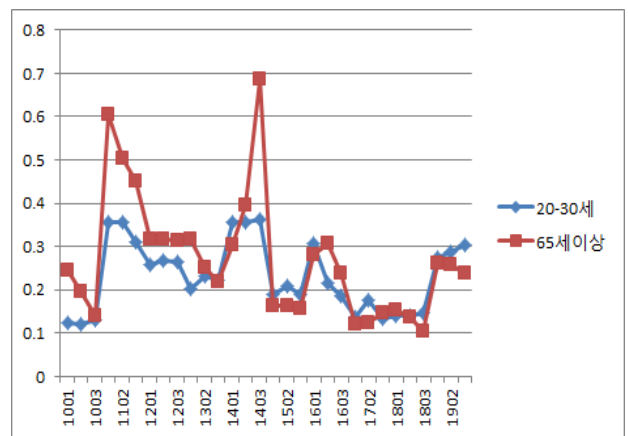
하였다.

발성속도 측정결과, 그림 3에서 볼 수 있듯이 음절 당 발성속도가 거의 일정한 비율로 청년층이 노인층보다 빠른 것으로 나타났다. 이는 나이가 들수록 폐 용량의 감소, 통풍과 가스 교환에 있어서 폐의 기능이 퇴화되어 숨 쉴 때 마다 말할 수 있는 음절수가 적어지는 면모를 단적으로 보여주는 실험결과이다[4]. 또한 발성 중 묵음구간 길이 측정결과, 그림 4와 같이 노인의 발성속도가 느린 것에 비해 묵음구간이 차지하는 비율이 항상 높기 때문만은 아니라는 것이 증명되었다. 그러므로 단지 음소 별 발성속도가 느린 것인지 아니면 다른 요소가 있는지에 대한 추가적인 실험 및 분석이 필요하다.

추가적으로 이번 실험에서 나온 발성속도의 차이가 실제 스마트기기의 음성인식에 어떠한 영향을 주는지를 실험해 보았다. 안드로이드 스마트기기의 구글보이스 키워드 검색 앱을 기반으로 청년층과 노인층 음성에 대해 각각 50회 시도한 결과 청년층의 음성은 인식률이 70% 이었고, 노인층은 52%를 보였다.



(그림 3) 음절 당 발성속도(단위: 초)



(그림 4) 음절 당 묵음구간 길이 (단위: 초)

이번 실험결과를 바탕으로 좀 더 객관적인 결론을 도출할 수 있도록 추가적인 DB녹음이 진행될 예정으로 총 40명에 이르는 노인들의 음성샘플을 수집할 것이다. 또한 이번 실험에서 10가지의 4-8음절로 구성된 단어 또는 어구들을 사용하였지만, 앞으로는 한국어에 있어서 사용빈도가 높은 단어들을 중심으로 선택하여 DB를 구축할 예정이다. 그 밖에 조음기능의 저하에 의한 요인을 발생속도와 묵음구간의 길이에 한정짓지 않고, 음성신호 분석을 통해 발음의 명료도와 포먼트의 변이, 그리고 jitter와 shimmer의 영향을 추가로 연구해 볼 것이다.

4. 결론

65세 이상의 노인은 점막, 후두, 폐 기능 등의 변화로 인해 전체말속도와 조음 속도가 느려진다고 연구되어왔다. 이는 발성속도의 변화로 이어지고 기존의 스마트 기기 음성인터페이스의 오류를 유발하게 되어 활용에 어려움을 준다. 이러한 문제점을 해결하기 위해서 노인들이 청년에 비해 발성속도가 어느 정도 저하되며 이것이 단순히 조음 속도만의 문제인지를 실험적으로 분석해 보았다. 그 결과 발성 중에 끼어있는 묵음구간도 상당한 영향을 주지만, 이것이 전체적인 발성속도에 절대적인 영향을 준다고는 볼 수 없다는 결론이 이르렀다. 이 후 실험에서는 발성속도 뿐만 아닌, 발음명료도와 음향적 요소를 고려하여 보다 더 복합적인 상관관계를 구축하고, 이를 바탕으로 노인 친화형 음성인터페이스를 만드는 데 주력할 예정이다. 이는 실버세대들에게 있어서 기존의 검색 서비스나 Social Networking Service의 지능화와 사용성 및 효용성 증대에 기여할 것으로 예상된다.

5. ACKNOWLEDGMENTS

이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2013R1A1A2008554).

참고문헌

- [1] Bennett, J. W., Van Lieshout, P. H., & Steele, C. M. (2007). Tongue control for speech and swallowing in healthy younger and older subjects. *International Journal of Orofacial Myology*, 33, 5-18.
- [2] Sonies, B. C. (1987). Oral-motor problems. In H. G. Mueller & V. C. Geoffrey (Eds.), *Communication disorders in aging* (pp. 185-213). Washington, DC: Gallaudet University Press.
- [3] Kahane, J. C. (1981). Anatomic and physiologic changes in the aging peripheral speech mechanism. In D. S. Beasley & G. A. Davis (Eds.), *Aging: Communication processes and disorders* (pp. 21-45).

New York: Grune & Stratton.

- [4] Hooper, Celia R., and Ann Cralidis. "Normal Changes in the Speech of Older Adults: You've still got what it takes; it just takes a little longer!." *Perspectives on Gerontology* 14.2 (2009): 47-56.
- [5] 김영호. 노인성음성. Plenary Session IV. (pp. 205-207). 연세대학교 의과대학 이비인후과학교실.
- [6] 고도홍. "한국 정상 노인 음성의 기본주파수." *음성과학* 15.3 (2008): 95-102.
- [7] Slawinski, Elzbieta B. "Age related changes in articulation of/b/and/w/." *The Journal of the Acoustical Society of America* 90 (1991): 2363.
- [8] P. H. Ptacek and E. D. Sander, "Age recognition from voice," *J. Speech Hear, Res.* 9:273-277, 1966.
- [9] P. H. Ptacek, E. K. Sander, W. H. Maloney, and C. C. Roe Jackson, "Phonatory and related changes with advanced age," *J. Speech Hear, Res.* 9:353-360, 1966.
- [10] Walter H. Manning and Karen L. Monte, "Fluency Breaks in Older Speakers: Implications for a Model of Stuttering Throughout the Life Cycle," *J. Fluency Disorders*, 6:35-48, 1981.
- [11] 김은정, "정상 청년층과 노년층의 최대발성시간 및 조음교대운동속도 비교," 연세대학교 대학원, 언어병리학 협동과정 석사논문, 2003.
- [12] 진성민, 권기환, and 강현국. "연령증가에 따른 정상 노인의 음향 분석학적 특징." *대한음성언어의학회지* 8.1 (1997): 44-48.