

각종 음성분석 상용 프로그램의 모음 기본주기 분석 결과 비교

Comparison of vowel pitch results among several commercial voice analysis programs

남기창*, 이승훈**, 최재남***, 최홍식***, 남도현***, 김덕원*

Ki-Chang Nam*, Seung-Hoon Lee**, Jai-Nam Choi***, Hong-Shik Choi***, Do-Hyun Nam***, Deok-Won Kim*

Abstract - Analysis of the voice and its corresponding studies are examined from the recording of the voice through microphone and various calculation processes of the signals by using computer. Voice analyser include data acquisition and analyzing program. Since each program uses different voice signal processing algorithm, thorough understanding of the operation is essential. In this study, analysis result of patient voice were compared by using four different voice analysis programs such as MDVP, Praat, TF32, and the program developed in this study. Pitch, jitter and shimmer were selected as comparison analysis factors. As a result, pitch, jitter and shimmer showed different result since each program uses different pitch computation algorithm.

Key Words : voice analyser, pitch, jitter, shimmer.

1. 서 론

음성을 평가하는 방법에는 청각적 검사, 음향음성학적 검사, 공기역학적 검사, 발성기관의 검사, 근육과 신경 기능 검사 등이 있다. 이 중 음향음성학적 검사는 조음 및 공명기관을 거쳐 최종적으로 발생된 음성에 대한 분석으로 크게 두 가지 과정을 거치게 된다. 첫 번째는 마이크를 통해 아날로그 신호인 음성신호를 표본화하고 양자화 하여 디지털 신호로 변환한 후 저장하는 과정이다. 두 번째는 저장된 음성신호를 컴퓨터를 통해 다양한 신호처리 방법에 의해 분석하는 과정이다. 이 두 가지 과정을 포함하는 다양한 음성 분석 프로그램이 이미 많이 발표되어 있다.

각각의 음성 분석 프로그램은 음성신호처리에 대한 다양한 알고리즘을 사용하고 있으므로 같은 음성에 대해서 각각 다른 분석 결과를 나타낼 수 있다. 따라서 본 연구에서는 병적 음성을 대상으로 음성의 기본 특징이 되는 기본주기(pitch), 주파수 변동율(jitter), 진폭 변동율(shimmer)의 결과를 비교하고자 하였다. 비교대상으로는 본 연구에서 개발된 분석프로그램과 MDVP(multi-dimensional voice program) (Kay Elemetrics, USA), TF32(University of Winconsin-Madison, USA) 및 Praat를 사용하였다.

2. 실험방법

2.1. 분석 데이터의 구성

분석 데이터는 6 명의 파킨슨 병 환자의 모음 /아/ 를 발생했을 때, 샘플링 레이트 44,100Hz로 녹음한 웨이브 파일이다. 녹음된 웨이브 파일에서 전체 구간 중 중앙의 1초간을 선택하여 총 15개의 분석 자료를 구성했다.

2.2. 기준 값(gold standard) 분석

분석 자료의 기준 값 분석을 위하여 그림 1과 같이 분석 자료인 음성신호를 확대하여 관찰하였다. 확대된 음성신호의 피크와 다음 피크까지를 기본주기로 하여 피치를 계산하였다. 그림에서 분석된 피치는 153.7Hz이다. 주파수 변동율과 진폭 변동율은 각각 식 (1), 식 (2)에 의해 계산하였다 [1,2].

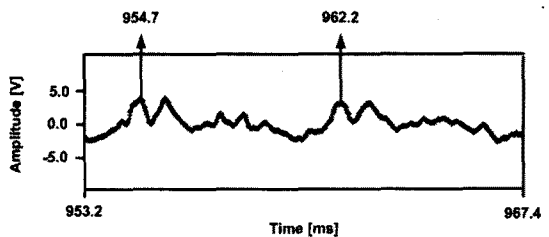


그림 1. 음성신호의 확대 파형

저자 소개

- * 연세대학교 의과대학 의학공학교실
- ** 연세대학교 대학원 생체공학 협동과정
- *** 연세대학교 의과대학 이비인후과학교실, 음성언어의학연구실

$$Jitter = \frac{\frac{1}{N-1} \sum_{i=0}^{N-1} |T_0(i) - T_0(i+1)|}{\frac{1}{N} \sum_{i=0}^N T_0(i)} \quad (1)$$

여기서 $T_0(i)$ 는 음성의 피치, N 은 분석구간의 피치 프레임 수이다.

$$Shimmer = \frac{\frac{1}{N-1} \sum_{i=0}^{N-1} |P_0(i) - P_0(i+1)|}{\frac{1}{N} \sum_{i=0}^N P_0(i)} \quad (2)$$

여기서 $P_0(i)$ 는 음성의 진폭, N 은 분석구간의 피치 프레임 수이다.

2.3 음성분석 프로그램에 의한 분석

2.3.1 개발된 음성분석 프로그램

개발된 음성 분석 프로그램에서는 자기상관함수법 (autocorrelation)을 기본으로 피치를 계산한다. 그러나 정해진 윈도우 크기마다 피치를 구하지 않고, 이전 피치 프레임에서의 피치를 참조하여 다음 구간의 피치 분석을 위한 윈도우 크기를 조절하도록 하였다. 그림 2는 본 연구에서 개발된 음성분석 프로그램으로 각 환자의 음성파일을 분석하는 화면이다. 분석 프로그램은 Visual C++ 6.0 (Microsoft, USA)와 Measurement Studio (National Instruments, USA)를 이용하였다.

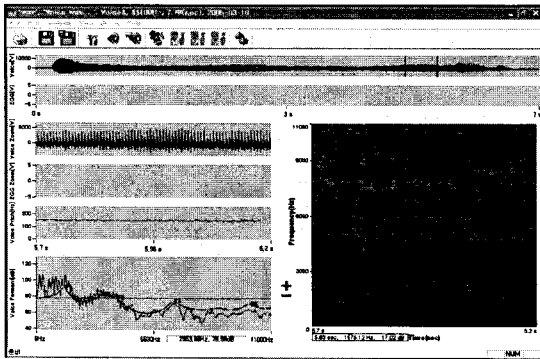


그림 2 개발된 음성 분석 프로그램

2.3.2 음성분석 프로그램

15개의 환자 음성 분석 자료를 MDVP, TF32 및 Praat의 각 프로그램의 사용법에 따라 분석하였다[3]. Praat의 경우는 피치를 구하는 방법이 여러 가지 있는데, 이 중 단순자기상관함수법(simple autocorrelation), 자기상관함수법(autocorrelation), 교차상관함수법(cross correlation), STFT(short time fourier transform)를 사용하여 분석하였다[4]. MDVP와 TF32는 자기상관함수법과 교차상관함수법을 이용하여 피치를 계산한다.

3. 결과

분석 자료의 기준 값과 각 프로그램으로 분석된 결과를 SPSS 8.0 (SPSS Inc., USA)를 이용하여 유의수준 $p = 0.01$ 로 검정하였다. 각 분석결과 간의 비교는 비모수적 일원분산분석(Kruskal-Wallis Test)을 이용하여 검정하였다. 기준 값과 각 프로그램에 의한 분석 결과는 비모수적 쌍체 t 검정(wilcoxon signed rank test)을 하였다. 또한 기준 값과의 오차율도 계산하였다.

3.1 피치 결과 비교

표 1은 15개의 분석 자료에 대한 기준 값과 각 프로그램이 분석한 피치의 평균이다. 피치는 주파수로 환산하여 단위는 Hz이다. 비모수 일원분산분석 검정결과 분석방법에 따른 통계적 유의차는 없었다($p = 0.982$). 기준 값에 대한 각 프로그램에 의한 분석 결과와의 쌍체 t 검정결과, 개발된 프로그램에 의한 결과만 기준 값과의 차이를 보였다. 그러나 개발된 프로그램에 의한 오차율은 0.3 %로 작았다.

표 1. 피치 분석 결과 (n = 15)

| 분석방법 | 피치 (평균±표준편차) [Hz] | 오차율(평균 ±표준편차) [%] | 쌍체 t 검정 결과 (p 값) |
|--------------------------------------|-------------------|-------------------|------------------|
| 기준 값 | 152.7 ± 33.7 | 0.0 | - |
| 개발 s/w | 152.3 ± 33.4 | 0.3 ± 0.2 | 0.005 |
| MDVP | 152.7 ± 33.7 | 0.1 ± 0.1 | 0.140 |
| TF32 | 152.8 ± 34.0 | 0.3 ± 0.3 | 0.031 |
| Praat (simple) | 149.4 ± 35.3 | 2.1 ± 7.9 | 0.011 |
| Praat (autocorrelation) | 149.4 ± 35.3 | 2.1 ± 7.9 | 0.011 |
| Praat (cross correlation) | 147.8 ± 37.4 | 3.0 ± 11.7 | 0.015 |
| Praat (short time fourier transform) | 151.3 ± 33.4 | 0.9 ± 0.3 | 0.001 |

3.2 주파수 변동율 결과 비교

표 2는 15개의 분석 자료에 대한 기준 값과 각 프로그램이 분석한 주파수 변동율의 평균이다. 평균 주파수 변동율의 단위는 %이다. Praat의 경우 주파수 변동율이 자기상관함수법에 의한 결과만 산출된다. 기준 값과 각 프로그램으로 분석된 주파수 변동율의 그룹에 따른 결과를 비모수 일원분산분석 검정한 결과, 분석방법에 따른 통계적 유의차가 있었다($p = 0.002$). 기준 값에 대한 각 프로그램에 의한 분석 결과와의 쌍체 t 검정결과, 개발된 프로그램과 MDVP를 제외하고 다른 두 방법이 차이를 보였다. 또한 개발된 프로그램에 의한 결과는 오차율이 6.6 %인 반면에 다른 세 방법은 33 % 이상의 오차율을 보였다.

표 2. 주파수 변동을 분석결과 (n = 15)

| 분석방법 | 기준 값 | 개발 s/w | MDVP | TF32 | Praat |
|-----------------------|-----------|------------|-------------|-------------|-------------|
| 주파수 변동율 (평균±표준편차) [%] | 1.2 ± 0.7 | 1.3 ± 0.8 | 1.5 ± 1.5 | 0.7 ± 4.0 | 0.5 ± 0.4 |
| 오차율(평균±표준편차) [%] | 0.0 | 6.6 ± 12.6 | 33.6 ± 50.1 | 59.8 ± 31.9 | 56.0 ± 20.1 |
| 쌍체 t 검정 결과 (p 값) | - | 0.017 | 0.820 | 0.005 | 0.001 |

3.1.3 진폭 변동율 결과 비교

표 3은 15개의 분석 자료에 대한 기준 값과 각 프로그램이 분석한 진폭 변동율의 평균이다. 평균 진폭 변동율의 단위는 %이다. Praat의 경우 진폭 변동율이 자기상관함수법에 의한 결과만 산출된다. 기준 값과 각 프로그램으로 분석된 진폭 변동율의 그룹에 따른 결과를 비모수적 검정을 한 결과, 분석방법에 따른 통계적 유의차는 없었다(p = 0.360). 기준 값에 대한 각 프로그램에 의한 분석 결과와의 쌍체 t 검정 결과, 개발된 프로그램과 TF32를 제외하고 다른 두 방법은 기준 값과의 차이를 보였다. 또한 개발된 프로그램의 오차율이 가장 작았다.

표 3. 진폭 변동율 분석결과 (n = 15)

| 분석방법 | 표준 데이터 | 개발 s/w | MDVP | TF32 | Praat |
|----------------------|-----------|-----------|-------------|-------------|-------------|
| 진폭 변동율 (평균±표준편차) [%] | 5.8 ± 2.9 | 5.8 ± 3.0 | 7.5 ± 4.9 | 6.2 ± 5.0 | 7.2 ± 3.9 |
| 오차율(평균±표준편차) [%] | 0.0 | 2.9 ± 2.4 | 24.9 ± 16.7 | 14.9 ± 19.5 | 25.1 ± 19.5 |
| 쌍체 t 검정 결과 (p 값) | - | 0.570 | 0.001 | 0.140 | 0.002 |

4. 고찰 및 결론

본 연구에서는 15개의 환자 음성 분석 자료를 이용하여 4가지 음성 분석 프로그램으로 분석한 결과를 비교하였다. 음성 분석의 기본이 되는 피치, 주파수 변동율과 진폭 변동율에 대한 결과를 비교하여 그 차이를 볼 수 있었다.

피치 분석 결과, 통계적으로 기준 값과의 유의차를 보이지 않은 것은 MDVP와 TF32였다. 정확도는 MDVP-개발된 프로그램-TF32-Praat 순이었다. 주파수 변동율 분석 결과, 통계적으로 기준 값과의 유의차를 보이지 않은 것은 MDVP와 개발된 프로그램 이었다. 정확도는 개발된 프로그램

-MDVP-TF32-Praat 순이었다. 진폭 변동율 분석 결과, 통계적으로 기준 값과의 유의차를 보이지 않은 것은 개발된 프로그램 TF32 이었다. 정확도는 개발된 프로그램 -TF32-MDVP-Praat 순이었다. 따라서, 정확도에 대한 전체적인 순위는 개발된 프로그램-MDVP-TF32-Praat 로 볼 수 있었다.

환자의 음성은 일반인과 다르게 진폭이 현저히 낮은 경우도 있다. 이 같은 경우 분석프로그램인 MDVP의 경우 배율을 3배로 증가시켜야만 분석이 가능하다. 본 연구에서 MDVP 적용 시 이러한 과정을 거쳤기 때문에 MDVP에 의한 결과는 더 안 좋을 수 있다. 또한 실제 사용자들이 이러한 사실을 간과하면 나쁜 결과를 초래할 수 있다.

향후, 분석자료의 수를 추가하고, 포먼트(formant)와 신호의 배율과 잡음의 비율(HNR : harmonics to noise ratio)등의 분석 요인들을 추가하여 각 음성 분석기 들의 특성을 비교하고자 한다. 이러한 결과를 바탕으로 임상에서 음성 자료 수집 및 분석에 참고할 만한 기본 자료가 될 수 있으리라 기대한다.

본 연구는 2005년 중소기업청 중소기업기술혁신개발사업 지원에 의해 수행되었음.

참 고 문 헌

- [1] 안희영 역, "음성검사법 - 기초편", 제2판, 군자출판사, 1996
- [2] 안희영 역, "음성검사법 - 임상편", 제2판, 군자출판사, 1996
- [3] 고도홍 외, "음성 및 언어 분석기기 활용법", 한국문화사, 2001
- [4] 양병곤, "프라트(Praat)를 이용한 음성분석의 이론과 실제", 만수출판사, 2003