



# The fundamental frequency (f0) distribution of Korean speakers in a dialogue corpus using Praat and R\*

Byunggon Yang\*\*

*Department of English Education, Pusan National University, Busan, Korea*

## Abstract

This study examines the fundamental frequency(f0) distribution of 2,740 Korean speakers in a dialogue speech corpus. Praat and R were used for the collection and analysis of acoustical f0 data after removing extreme values considering the interquartile f0 range of the intonational phrases produced by each individual speaker. Results showed that the average f0 value of all speakers was 185 Hz and the median value was 187 Hz. The f0 data showed a positively skewed distribution of 0.11, and the kurtosis was -0.09, which is close to the normal distribution. The pitch values of daily conversations varied in the range of 238 Hz. Further examination of the male and female groups showed distinct median f0 values: 114 Hz for males and 199 Hz for females. A *t*-test between the two groups yielded a significant difference. The skewness representing the distribution shape was 1.24 for the male group and 0.58 for the female group. The kurtosis was 5.21 and 3.88 for the male and female groups, and the male group values appeared leptokurtic. A regression analysis between the median f0 and age yielded a slope of 0.15 for the male group and -0.586 for the female group, which indicated a divergent relationship. In conclusion, a normative f0 distribution of different Korean age and sex groups can be examined in the conversational speech corpus recorded by a massive number of participants. However, more rigorous data might be required to define a relation between age and f0 values.

**Keywords:** fundamental frequency (f0) distribution, Korean dialogue speech corpus, age and sex groups, statistics

## 1. 서론

사람의 음성은 성대와 성도의 공명으로 만들어진다(Fant, 1973). 성대의 진동은 음향적으로는 fundamental frequency(f0)로 측정되고, 귀로 들려지는 높낮이를 말할 때 피치라는 용어를 사용한다. f0값은 사람의 발성기관인 성대의 크기와 무게 같은 해

부화적인 특징에 기본적인 영향을 받으며, 화자가 자신의 감정을 표현하는 과정에서 성대의 긴장도를 조절하게 되어 변화범위가 확장되기도 한다(Yang, 1990, 1998). 말소리의 날카로움과 둔탁한 정도를 나타내는 음색은 성대의 진동방식에 따라 다른데 음성병리학자들은 성대 근접 촬영을 하거나 갑상연골의 양쪽 피부에 한 쌍의 전극을 부착하여 정밀한 측정을 하기도 한다

\* This work was supported by a 2-Year Research Grant of Pusan National University.

\*\* bgyang@pusan.ac.kr, Corresponding author

Received 31 July 2023; Revised 6 September 2023; Accepted 9 September 2023

© Copyright 2023 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

(Yang, 1996). 덧붙여, 실제 발화를 분석한 피치값의 궤적을 추적해보면 수시로 변하는 것을 알 수 있다(Yang, 2018, 2021a). 일반적으로 감정이 매우 쌓인 경우의 발화에서는 변화범위가 매우 넓을 수도 있지만, 자연스런 일상생활의 대화에서는 이 범위가 좁으며, 사람마다 고유한 성대의 변화 특징을 반영하게 되는 경향이 많다. 따라서, 한 언어의 화자가 대규모로 참여한 대화 음성 코퍼스를 분석하면 그 언어의 기준이 될 만한  $f_0$ 범위를 파악할 수 있을 것으로 기대된다.

$f_0$  관련 연구를 살펴보면, Kinoshita et al.(2009)은 201명의 일본인 남성들을 대상으로 장기적인 발화에서의  $f_0$ 값의 분포를 통계적으로 분석하여 여섯 개의 변수를 기준으로 화자확인에서 10.7%의 오인식률과 오거부율이 같아지는 equal error rate(EER)를 구했다고 한다. 그들은  $f_0$ 값이 화자의 성도의 모양을 나타내는 포먼트값보다 특정한 단어나 음소에 영향을 받지 않아 측정이 쉽고 소음에 강하기 때문에 초기의 과학적 범죄수사에서 화자확인의 매력적인 도구였지만, 건강이나 정서 상태, 주변 잡음 등에 영향을 받아 화자 내에서도 많이 변하기 때문에 과학적 범죄수사에 제한적으로 이용되었음을 지적했다. 그들은 서로 다른 시간에 발화한 개인별 장기적인  $f_0$ 평균값(Long-term  $F_0$ : LTF0)이나 표준편차(SD)에서는 차이가 있을 수도 있지만,  $f_0$ 분포는 거의 비슷한 모양의 분포를 보인 점에 주목하고 이러한 분포의 특성을 나타내는 첨도(kurtosis), 왜도(skewness), 최빈값(mode), 최빈값의 확률밀도값(probability density at the mode)을 LTF0나 SD에 변수로 추가해서 화자식별을 처리해본 결과 10.7%의 EER을 보였고, 발화시간은 2분 이상일 때 참여자들을 적절하게 구별할 수 있으며 짧아질수록 EER이 증가하여 구별이 어렵다고 보고했다. 이와 비슷한 주제의 연구로 Hudson et al.(2007)은 영국 남부표준영어(standard Southern British English)를 구사하는 100명의 18-25세의 젊은 남성들이 경찰과 모의 대화를 나눈 20여분의 자료에서 3-5분의 음성자료를 Praat 스크립트로 추출한 최빈값, 평균값, 중앙값이 각각 102 Hz, 106 Hz, 105 Hz로 대체로 정상분포를 나타냈으며, 60%의 화자들의 장기간  $f_0$ 값이 20 Hz 범위 내에 중첩되어 있어서 화자식별의 한계점을 보였음을 보고했다. Lindh(2006)는 스웨덴의 방언이 포함된 20-30세에 해당하는 109명의 젊은 남성화자들이 17초에서 2분에 걸쳐서 녹음한 자연발화의  $f_0$ 값을 측정하여 개인별 평균값과 중앙값, 표준편차의 분포를 살펴보았다. 그 결과 전체평균은 120.8 Hz로 나왔고 중앙값은 평균보다 약간 내려간 115.8 Hz이었으며, 68%의 화자들이 100-130 Hz 범위에 분포했다고 한다. 자료분포에서 구한 왜도는 0.6으로 5명의 극단값이 영향을 미치기 때문에 평균값보다는 중앙값을 활용하라고 제안했다. 덧붙여, 표준편차 값은 10-55 Hz에 걸쳐 나타났고 가장 많은 27명이 20 Hz를 보였다. 이러한 연구를 보면 화자별  $f_0$ 분포가 다소 안정적임을 알 수 있으며 대용량 말뭉치에서 구한 통계치들은 화자의 특징을 잘 요약해 줄 것으로 기대된다.

한국인의 자연발화를 대상으로 Yang(2021b)은 40명의 참여자가 한 명당 1시간 동안 일상적인 주제에 대한 생각을 자연스럽게 말하게 하여 녹음한 250개의 음성파일로 구성된 서울코퍼

스  $f_0$ 값의 분포를 분석했다.  $f_0$ 값의 분석은 Praat 스크립트를 이용하여 20 ms마다 한 개의 값을 구했고 0.029초의 윈도우 크기로 만든 좁은대역 스펙트로그램과  $f_0$ 곡선을 동시에 보면서 한 옥타브 위아래값으로 갑자기 떨어진 경우나, 초성이나 종성의 파열음과 마찰음, 파찰음 등에서 에러가 난 값을 수작업으로 걸러서 처리했다. 그 결과 약 3백만 개의 데이터에 대한 중앙값이 148 Hz로 나타났고, R의 상자도로 보았을 때 일상발화에서의 한국인 남녀별  $f_0$ 분포는 최소 65 Hz에서 최대 274 Hz까지의 범위에서 변한다고 보고했다. 또한 남녀집단별로 나뉘었을 때는 남성집단에서는  $f_0$ 중앙값이 111 Hz인데 비해 여성집단에서는 남성집단의 거의 두 배에 해당하는 200 Hz의 중앙값을 기록했다. 그 외에도 10대에서 40대까지 다양한 나이층으로 구성된 참여자들의 나이와  $f_0$ 값에 대한 회귀분석을 실시해본 결과, 나이가  $f_0$ 에 미치는 영향은 유의미하지만, R제곱값으로 본 예측정확도에서는 매우 낮은 값을 보여서 나이층이 다양한 더 많은 참여자로 이뤄진 코퍼스의 자료분석이 필요하다고 제안한 바 있다. 이 연구는 국립국어원이 공개한 대규모의 일상 대화 음성 말뭉치에는 나이와  $f_0$  사이의 관계가 어떤지 살펴보기 위해 시작하게 되었다.

구체적으로 이 연구에서는 국립국어원의 ‘일상 대화 음성 말뭉치 2020’(National Institute of Korean Language, 2023)을 음성분석기 Praat(v.6.2.14, Boersma & Weenink, 2022)의 스크립트를 이용하여  $f_0$ 값을 구하고 통계패키지 R(v.4.3.1, R Core Team, 2023)을 이용하여, 성별, 나이별 집단 및 개인별 분포와 변화범위를 살펴보고,  $f_0$  중앙값과 나이에 대한 회귀분석을 실시하여 상호관계를 규명하고자 한다.

## 2. 연구 방법

### 2.1. 한국어 말뭉치의 참여자와 대화 주제 및 파일 구성

국립국어원의 ‘일상 대화 음성 말뭉치 2020’ 녹음에 참여한 인원수는 총 2,740명이다. 이 가운데 남성화자는 677명이고 여성은 2,063명이다. 여성화자의 수가 거의 남성화자의 세 배를 넘는다. 표 1은 말뭉치 녹음에 참여한 대상자들의 정보를 담은 JSON 텍스트 파일에서 추출한 나이대별, 성별 분포를 보여준다.

표 1. 한국어 ‘일상 대화 음성 말뭉치 2020’의 나이대별 참여자 인원수

Table 1. Number of participants by age groups in the Korean ‘Dialogue Corpus (audio) 2020’

Age group	10s	20s	30s	40s	50s	60s	70
n	327	874	314	641	437	146	1
Males	45	236	115	160	84	36	1
Females	282	638	199	481	353	110	0

참여자들 중 20대가 874명으로 가장 많고, 이 가운데 여성이 638명이나 된다. 이어서 40대가 641명이며 이 가운데 여성이 481명이다. 70대는 1명이 참여했다. 남녀로 구분해보면 이런 나이 분포는 연령대를 균등하게 표집한 것으로 보긴 어렵지만, 다

수의 화자가 참여해서 나이별  $f_0$ 특성을 어느 정도 분석해볼 수 있는 있을 것으로 기대된다.

표 2는 말뭉치 녹음에 참여한 사람들의 지역별 분포를 인원 수가 많은 것부터 정렬하여 보여준다.

**표 2.** 한국어 ‘일상 대화 음성 말뭉치 2020’의 지역별 참여자 인원수

**Table 2.** Number of regional participants in the Korean ‘Dialogue Corpus (audio) 2020’

Region	서울	부산	경기	광주	대구	대전
n	676	380	314	285	201	155
Region	강원	제주	인천	전남	경북	경남
n	151	108	90	77	61	60
Region	충남	전북	충북	울산	세종	
n	56	54	43	24	5	

표 2에서 17개 주요 도시와 도별로 구분한 참여자 인원수를 보면, 서울 지역 참여자가 676명으로 가장 많으며, 이어서 부산 지역에서 380명, 경기 지역에서 314명이 참여했다. 울산 24명에 이어 세종시 5명이 가장 적다.

참여자들은 가족, 영화, 먹거리, 반려동물 등, 총 15개 주제에 대해 2명이 한조를 이루어 1,818건의 일상적인 대화 자료와, 13개의 신문 기사에 대해 의견을 주고받는 414건의 대화 자료로 되어 있다. 파일 수는 긴 휴지, 경계 억양, 경계말 장음화 등을 특징으로 하는 억양구 단위로 분할한 음성 파일 870,675개, 텍스트 파일 2,232개로 약 54.3 GB를 차지한다. 음성 파일은 16비트 양자화 선형 PCM파일로 16 kHz 표본화율로 녹음되었고, 대화문과 참여자들에 대한 정보를 담은 JSON 텍스트 파일은 UTF-8로 저장되었다.

## 2.2. Fundamental Frequency( $f_0$ )값과 참여자 정보 수집

먼저  $f_0$ 값을 구하는 과정은 폴더의 음성파일을 Praat에 불러와  $f_0$ 분석을 하고 결과값을 컴퓨터에 직접 저장하는 Praat스크립트를 이용했고, 참여자 개인별 정보 수집은 R을 이용했다(Yang, 1998, 2021a, 2021b 참고). Praat스크립트는 국립국어원에서 제공한 음성 파일 폴더의 모든 목록 이름을 리스트 형태로 구한 다음, 개인별로 녹음한 음성파일이 들어 있는 폴더에 들어가 PCM형태로 저장된 파일을 16비트 Little Endian 형태로 Praat 개체창에 불러왔다. 이어서 개체창의 파일을 처음부터 차례로 선택하고  $f_0$ 값을 다음과 같은 Praat 스크립트로 구하고 그 파일 구간 내에서 20 ms마다 75 Hz에서 600 Hz 범위로  $f_0$ 개체를 만들고  $f_0$ 개체창의 값들을 제1사분위( $q_{25}$ )와 제3사분위( $q_{75}$ )를 구하고 사분위 간 범위인 interquartile range( $iqr$ )를 구하여 이 값의 1.5배 아래나 위에 해당하는 상한값( $highcut$ ), 하한값( $lowcut$ )을 구해 이 범위 안의  $f_0$ 값만 하드디스크에 파일로 저장했다. 이런 상하한값을 이용하면 억양구 단위로 구분된 파일 내에서의 피치 측정에러에 의한 극단값을 어느 정도 제거할 수 있다.

```
To Pitch: 0.02, 75, 600
```

```
q25=Get quantile: 0, 0, 0.25, "Hertz"
```

```
q75=Get quantile: 0, 0, 0.75, "Hertz"
```

```
iqr=q75-q25
```

```
lowcut=q25-1.5*iqr
```

```
highcut=q75+1.5*iqr
```

```
if f0<lowcut or f0>highcut
```

```
f0=undefined
```

```
endif
```

이 과정에서 주로 어절 초나 말에서  $f_0$ 값이 구해지지 않아 undefined로 나타난 경우와, 앞의 단락에서 언급한 사분위 밖의 값을 undefined로 정의하여 출력파일에 덧붙여 저장하지 않고 건너뛰도록 했다. 이렇게 제외된 값의 수는 모든 화자들의 전체  $f_0$ 값에서 차지하는 비율이 아주 낮아서 전반적인 분포를 파악하기 위한 통계분석에 영향을 끼치지 않을 것이다. 참고로 일부 연구에서는 40 dB 이하의 신호에서 구한 값을 제거하거나, 남성의 범위는 70–350 Hz로 여성은 100–500 Hz 범위로 지정하여 처리한 경우도 있다(Shi et al., 2014). 이렇게 특정범위의 값만 수용할 경우에는 다양한 감정을 표현하는 실제 대화에서의 타당한  $f_0$ 값을 놓칠 수도 있다. Boothroyd(1986)는 성인남성의  $f_0$ 값의 범위가 70–200 Hz를 보였고, 여성의 경우에는 140–400 Hz, 아이들에서는 180–500 Hz로 나타났다고 한다.

마지막으로 집단별 특성을 파악하기 위해서 참여자들에 대한 정보가 있는 JSON 텍스트 파일에서 R스크립트를 이용해서 화자 고유번호, 연령대, 성별, 지역 정보를 추출한 다음, 앞에서 구한  $f_0$ 자료와 합쳐서 최종 자료파일로 만들었다.

## 2.3. Fundamental Frequency( $f_0$ )자료 분석 및 통계 처리 과정

$f_0$ 값에 대한 자료 분석과 통계는 Lennes et al.(2015)의 R스크립트의 일부 코드를 활용하여 상자도와 밀도함수로 나타내었다. 이어서 중앙값, 평균값과 표준편차값 등의 기초통계자료를 구하여, 성별, 연령대에 따른 집단별 자료 분포를 살펴보고, 나이와  $f_0$ 값의 관계는 연령대별과 성별로 구한 중앙값과 나이에 대한 회귀분석으로 대체적인 관계를 살펴보았다.

참고로 개인별 억양구를 중심으로 극단값을 제거한 후에 저장한 자료를 한꺼번에 상자도로 그려보면 전체값을 반영하여 추정된 극단값들이 여전히 나타났다. 이 문제를 해결하기 위해서는 수많은 원자료의 극단값을 계속 되풀이해서 제거하여 지각적으로 더 이상 진행할 필요가 없는 단계를 찾는 추가 연구를 하거나, 화자개인별  $f_0$ 범위나 억양구 내 앞 뒤 구간의 자료를 반영하지 않고 20 ms마다 구분된 분석 창에서 구해 주는 Praat의  $f_0$ 분석 에러를 제거하는 과정이 필요하다.

## 3. 분석 결과와 논의

### 3.1. 한국어 말뭉치의 Fundamental Frequency( $f_0$ ) 분포

표 3은 ‘일상 대화 음성 말뭉치 2020’에 녹음된 2,740명 화자

들의 f0값의 통계를 보여준다.

**표 3.** 한국어 ‘일상 대화 음성 말뭉치 2020’의 2,740명 화자에 대한 f0값의 통계(Hz)

**Table 3.** Statistics of all the f0 values of 2,740 speakers in the ‘Dialogue Corpus (audio) 2020’

n	Mode	Mean	SD	Median	Min	Max
88,643,412	186	185	52	187	75	503

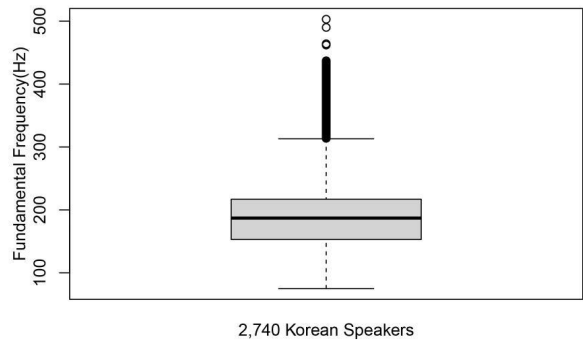
n denotes the number of the f0 values. SD denotes standard deviation. Min indicates the minimum f0 value while Max represents the maximum one (Hz).

f0, fundamental frequency.

표 3에 나타나 있듯이 모든 한국인 화자의 f0값의 최빈값은 186 Hz이고 총 816,798개로 모든 f0값의 약 1%를 차지했다. 평균값과 중앙값은 이보다 1 Hz 적거나 많은 185 Hz, 187 Hz로 각각 측정되었고, 표준편차는 52 Hz였다. 이 중앙값은 Yang(2021b)이 서울코퍼스에서 측정한 f0 평균값 160 Hz보다 25 Hz나 높아진 값인데, 이 연구에서는 여성화자의 수가 많아서 전체 평균값이 다소 높아진 것으로 보인다. 덧붙여, Yang의 연구에서는 중앙값이 148 Hz로 평균값과 12 Hz나 차이가 났었는데, 이 연구에서는 최빈값, 평균값, 중앙값의 차이가 거의 없다. 이러한 경향은 개인별 사분위값을 반영해 극단값을 통제된 결과로 여겨진다. 자료의 분포 모양을 보여주는 왜도는 0.11의 정적 편포를 나타내어 f0값이 높은 부분이 점점 작아지는 경사 모양을 보였고, 첨도값은 -0.09로 정상분포에 비해 아주 약간 납작하게 펼쳐진 모양을 보였다(Field, 2013). Yang의 서울코퍼스 연구에서는 왜도가 0.92로 정적 편포를 보였고, 첨도도 0.94로 뾰족한 모양을 보였다.

그림 1은 한국어 대화 음성 말뭉치의 상자도를 보여준다. 상자도 안의 굵은 선으로 표시된 중앙값은 187 Hz이고 아랫수염(lower whisker)은 75 Hz이고, 윗수염(higher whisker)은 313 Hz이다. 이 값들은 표준편차의 제1·제3 사분위값에  $\pm 2.7\sigma$ 로 구한 값으로 99.3%의 자료가 포함되는 구간이다. 상자도에서는 보통 이 기준 이상이나 이하는 작은 동그라미로 나타난다. 실제 이 값은 앞 절의 상한선 제1사분위값이 153 Hz이고, 제3사분위값은 217 Hz여서 그 범위는 64 Hz로 이를 1.5배하여 제3사분위값에 더하면 313 Hz가 되어 윗수염의 값과 같고, 이 값 위의 자료는 통계적으로는 집단에서 벗어난 극단값으로 처리될 수 있다. 하지만, 실제 대화 음성에서 이런 값들이 나올 수도 있기 때문에 전체 음성에서 무조건 제거하는 것은 바람직하지 않을 것으로 여겨져서 이 논문에서는 추가 제거를 실시하지 않았다. 앞으로 이런 부분에 대한 통계처리 과정과 실제 음성자료에 대한 서술이 적정한 지는 추가 논의나 검증이 필요하다. 여기서 한국인의 일상 대화에서의 윗수염값에서 아랫수염값을 뺀 99.3%에 해당하는 f0자료의 변화 범위는 238 Hz라고 할 수 있다. Yang (2021b)의 연구에서는 274 Hz의 범위를 보였는데 이는 옥타브 점프로 눈에 띄게 에러 난 부분을 수작업으로 처리하는 방법을 적용했지만, 모든 자료를 다 꼼꼼하게 확인하고 수정하기는 어

려웠을 것이다. 실제 수천 명이 참여한 한국어 말뭉치와 같은 대규모 코퍼스에서는 이렇게 확인하고 제거하는 작업이 거의 불가능하기 때문에 본 연구에서 채택한 개인별 억양구의 사분위 범위를 우선 활용하는 방법을 권한다. 또한 억양구별로 구한 값보다는 개인별로 모든 피치값을 다 구한 뒤 사분위를 구하여 한 번 더 걸렀을 때 얼마나 극단값을 정밀하게 제거할 수 있을 지에 대한 추가연구도 필요하다.



**그림 1.** 한국어 ‘일상 대화 음성 말뭉치 2020’의 2,740명에 대한 f0값(Hz)의 상자도

**Figure 1.** Boxplot of f0 values (Hz) of 2,740 speakers in the Korean ‘Dialogue Corpus (audio) 2020’

### 3.2. 한국어 말뭉치의 남녀별 Fundamental Frequency(f0) 분포

표 4는 남녀별 f0값의 통계치를 보여준다. f0값의 표본 수에서는 여성화자의 f0값의 표본수가 3.4배나 더 많다. 이러한 결과는 표 1에서와 같이 여성참여자도 남성참여자보다 3배나 많았고, 대화에서도 좀 더 길게 말한 것으로 보인다. 남성의 f0 중앙값은 114 Hz이고, 여성의 중앙값은 199 Hz로 여성화자의 f0값의 표본수가 1.75배나 된다. 남녀 간의 차이를 t검정으로 실시해 본 결과( $t=-10,362$ ,  $df=45,088,097$ ,  $p<.05$ )로 유의미한 차이를 보였다. 남성에 비해 여성의 표준편차가 높게 나온 것은 남성보다 높은 f0값과 많은 참여자 때문일 것으로 추정된다. 많은 자료임에도 개인별 억양구에서 구한 f0극단값을 제거했기 때문에 중앙값을 기준으로 보았을 때 평균값과 최빈값 등이 10 Hz의 작은 차이를 보였다. 서울코퍼스에서 측정한 남녀별 f0의 중앙값은 각각 111 Hz와 200 Hz를 기록했다. 이 연구의 결과와 비교해보면, 각 남녀 집단에서 1-3 Hz밖에 차이가 없다. 이런 결과를 보면 극단치를 적절하게 제거하면 인원수가 많은 적은 안정적인 f0의 중앙값을 구할 수 있음을 알 수 있다.

**표 4.** 한국어 ‘일상 대화 음성 말뭉치 2020’의 남녀별 f0값의 통계(Hz)

**Table 4.** Statistics of f0 values in Hz of male and female speakers in the Korean ‘Dialogue Corpus (audio) 2020’

Group	n	Mode	Mean	SD	Median	Min	Max
Male	19,963,978	104	120	29	114	75	321
Female	68,679,434	189	204	40	199	75	503

f0, fundamental frequency.

그림 2는 남녀별  $f_0$ 값의 밀도그래프를 보여준다. 일반적으로 밀도그래프는 막대그래프의 단점을 보완하여 각 데이터를 전체면적 1 내에서 밀도분포라는 개념을 사용하여 추정하여 보여준다. 이 그림에서 보면 남녀 모두 오른쪽으로 긴 꼬리가 내려지는 모양을 보이고 있고, 남성의  $f_0$ 값이 여성의 값보다 상대적으로 뾰족한 모양을 보이고 있다. 분산을 보여주는 왜도 통계치로는 남성의  $f_0$ 값이 1.24이고, 여성의 값은 0.58로 양수값을 나타내어 남성의 반에 해당하며 넓게 퍼져있다. 뾰족한 정도를 보여주는 첨도값은 각각 5.21과 3.88로 남성의 값이 34% 정도 상대적으로 더 뾰족한 모양으로 중앙값 주변에 몰려있다. Yang(2021b)의 서울코퍼스 연구에서도 남녀집단의 왜도가 각각 1.34와 0.66으로 나타났고 첨도는 2.88과 2.37로 거의 비슷한 유형의 그림을 보였는데, 이 연구의 결과는 Yang(2021b)에 비해 남성의 첨도값이 55% 더 뾰족한 모양을 보였다.

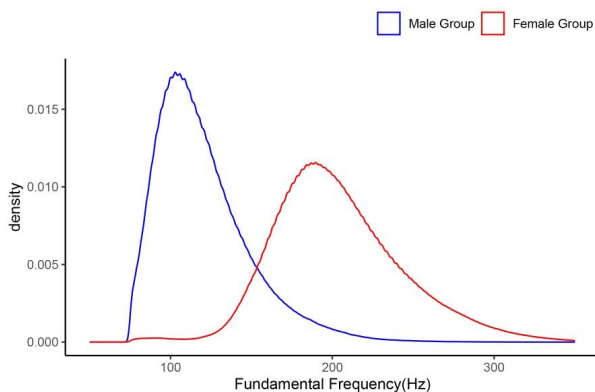


그림 2. 한국어 ‘일상 대화 음성 말뭉치 2020’의 남녀별  $f_0$ 값 밀도그래프

Figure 2.  $f_0$  density plot of male and female speakers in the Korean ‘Dialogue Corpus (audio) 2020’

### 3.3. 한국어 말뭉치의 연령별 Fundamental Frequency( $f_0$ ) 분포

표 5는 연령대별  $f_0$ 값의 최빈값, 평균, 표준편차, 중앙값 등을 보여주고 있다.

표 5. 한국어 ‘일상 대화 음성 말뭉치 2020’의 연령대별  $f_0$ 값의 통계(Hz)

Table 5. Statistics of  $f_0$  values by age groups in the Korean ‘Dialogue Corpus (audio) 2020’

Age group	n	Mode	Mean	SD	Median
10s	10,421,346	186	201	48	204
20s	27,128,866	193	187	53	191
30s	10,019,774	182	173	53	176
40s	20,723,334	180	182	52	181
50s	15,380,422	175	185	50	183
60s	4,922,674	168	175	47	173
70	46,996	153	160	37	155

n indicates the number of the  $f_0$  values. SD denotes standard deviation (Hz).

$f_0$ , fundamental frequency.

연령대별 자료의 수를 보면 20대가 가장 많고 70대가 가장 적다. 구체적으로 앞에서 제시한 표 1에서 20대의 참여자수가 874명으로 가장 많았고 70대는 1명에 불과했다. 표준편차의 평균은 48.6 Hz인데, 20대와 30대의 표준편차가 53 Hz로 가장 높고, 70대는 37 Hz로 가장 낮다. 개인별 억양구마다 사분위범위에 해당하는  $f_0$ 값을 이용해서 극단값을 어느 정도 걸러주었기 때문에 전체적으로 비슷한 표준편차값을 보였고, 70대의 참여자는 한 명이어서 편차가 가장 적었다.

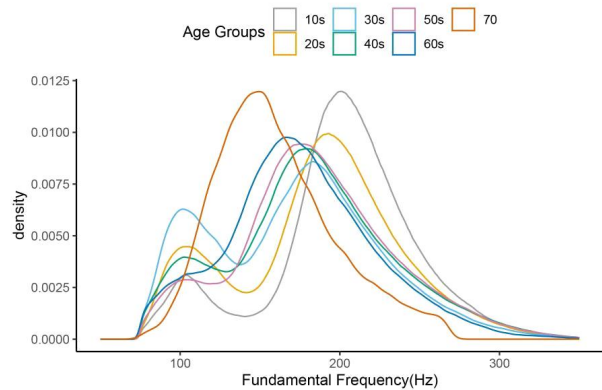


그림 3. 한국어 ‘일상 대화 음성 말뭉치 2020’의 연령대별  $f_0$ 값 밀도그래프

Figure 3.  $f_0$  density plot of age groups in the Korean ‘Dialogue Corpus (audio) 2020’

그림 3은 한국어 말뭉치의 연령별  $f_0$ 값의 밀도그래프를 보여준다. 그림은 연령별로 쌍봉과 단봉의 형태를 보여주고 있다. 70대는 한 명이라 단봉으로 나타났고, 나머지 나이대는 거의 쌍봉에 가까우며 쌍봉의 위치 차이는 나이가 적을수록 넓어져 있음을 볼 수 있다. 이러한 모양은 두 가지 원인이 있을 것으로 추정된다. 첫째, 남녀 참여자를 함께 그렸기 때문에 여성화자수가 많은 연령대인 경우에는  $f_0$ 값이 높은 쪽의 봉우리가 높게 나타나고, 여성보다 적은 수이지만 남성 참여자의  $f_0$ 값이 집단별  $f_0$ 값을 낮은 쪽으로 끌어 내리는 작용을 한 것으로 보인다. 둘째, 여성화자에게서도 Praat의  $f_0$ 측정 알고리즘에 일부 에러가 있어서, 화자의 평균  $f_0$ 에 비해 반이나 두 배가 되는 값들이 여전히 측정되고 있기 때문이다. 대규모의 음성파일에서 Yang(2021b)의 연구에서와 같이 측정된  $f_0$ 값을 좁은대역 스펙트로그램에서 대조해 보면서 일일이 점검하지 못하는 문제가 발생하기 때문에 이 두 번째 문제는 Praat의 측정 방법이 개선되면 해결될 수 있을 것이다.

그림 4는 한국어 말뭉치의 연령대별  $f_0$ 값의 상자도를 보여준다. 이 그림에서 보면 나이가 들수록  $f_0$ 값이 낮아지는 경향을 보인다. 나이와  $f_0$ 의 관계를 추정하려면 회귀분석을 실시하면 된다. 그런데 여기서 한 가지 지적할 것은 한국어 말뭉치를 수집할 당시에 참여자들의 실제 나이를 입력하지 않고, 10년 단위로 체크하여 표시하였기 때문에 두 변수 사이의 관계를 예측하기 위한 기울기나 절편을 계산하는 데 정확도가 떨어질 우려가 있다. 덧붙여, 연령별로 남녀 간의 참여자 비율이 비슷하지 않고

여성의 자료가 훨씬 더 많이 차지하고 있었다. 마지막으로, 최신의 16 GB메모리를 갖춘 iMac 컴퓨터로 남녀 집단의 모든 f0 데이터를 한꺼번에 회귀분석을 실행해보았지만 오류가 나서, 이 논문에서는 다음과 같이 첫 세 줄의 R코드를 입력하여 표 5의 남녀별 f0중앙값과 연령대별 회귀분석결과인 점선 아래 결과값의 예측 계수로 대체적인 관계를 살펴보았다. 남성 70대는 한 명이라 대푯값으로 볼 수 없어서 제외했다.

```
f0median<-c(204,191,176,181,183,173)
agegroup<-c(10,20,30,40,50,60)
summary(lm(f0median~agegroup))
----- Coefficients:
            Estimate Std. Error t-value Pr(>|t|)
(Intercept) 202.067    6.726   30.041 7.31e-06*
agegroup   -0.497     0.173   -2.878 0.045*
Adjusted R-squared: 0.593          *p<.05
```

모든 참여자들에 대한 회귀분석 결과를 보면 기울기가 -0.497로 절편값인 202 Hz에서 서서히 내려가는 경향을 보여준다. 이러한 경향은 그림 4에서와 같이 나이가 들수록 f0중앙값이 점점 낮아지는 모습과 일치한다. 이 회귀공식대로라면 수정된 R제곱값에 100을 곱한 약 59.3%의 정확도로, 실제값을 예측할 수 있다는 것인데 실제 데이터에서는 그림 4의 윗수염과 아랫수염의 범위나 극단치값들에서 볼 수 있듯이 자연스러운 감정표현이 들어간 대화체 음성이라 연령대별로 편차가 많아서 예측값의 정확도는 이보다 훨씬 낮아질 것이다.

이번에는 남녀별로 f0값이 다르기 때문에 각 집단(맨 뒷글자 f, m으로 남녀 구분)의 나이대별 중앙값을 구해서 각각 3줄의 R 코드로 전체적인 경향을 파악하기 위해 처리한 뒤 결과값의 계수만 제시해 보았다.

```
f0medianf<-c(202,193,183,180,179,170)
agegroupf<-c(10,20,30,40,50,60)
summary(lm(f0medianf~agegroupf))
----- Coefficients:
            Estimate Std. Error t-value Pr(>|t|)
(Intercept) 205.00    2.99   68.66 2.7e-07*
agegroupf   -0.586    0.077  -7.64 0.002*
Adjusted R-squared: 0.92          *p<.05
```

```
f0medianm<-c(104,105,101,104,106,113)
agegroupm<-c(10,20,30,40,50,60)
summary(lm(f0medianm~agegroupm))
----- Coefficients:
            Estimate Std. Error t-value Pr(>|t|)
(Intercept) 100.40    3.10   32.39 5.42e-06*
agegroupm    0.15     0.08    1.83 0.141
Adjusted R-squared: 0.32          *p<.05
```

여성집단에서는 기울기가 -0.586이고 절편이 205 Hz로 두 계수 모두 유의미한 관계를 보였고, 예측도를 보이는 수정된 R제곱값도 0.92로 매우 높게 나타났다. 이에 비해 남성집단에서는 기울기가 0.15, 절편이 100.4 Hz이지만 기울기는 유의미하지 않고 절편만 유의미한 관계를 보였고 수정된 R제곱값도 0.32로 매우 낮게 나타났다. Yang(2021b)의 남녀 집단별 회귀분석에서 남성집단의 기울기가 0.34이고 절편이 0.01로 나타났고, 여성집단의 기울기가 -0.002이고 절편은 206.1 Hz로 나타났는데, 남성집단은 유의미한 결과이지만 수정된 R제곱값은 0.018로 매우 낮았고, 여성집단의 기울기값은 통계적으로 유의미하지 않는 결과로 나타났다. 이러한 관계를 시각적으로 확인하기 위해 그림 5와 같이 남녀집단별로 회귀분석한 자료를 나타내 보았다.

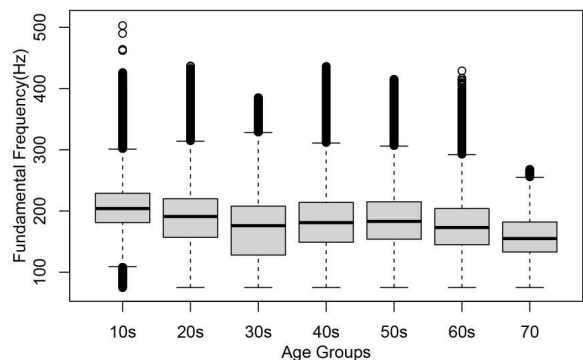


그림 4. 한국어 ‘일상 대화 음성 말뭉치 2020’의 연령대별 f0값의 상자도

Figure 4. Boxplots of f0 distribution by age groups in the Korean ‘Dialogue Corpus (audio) 2020’

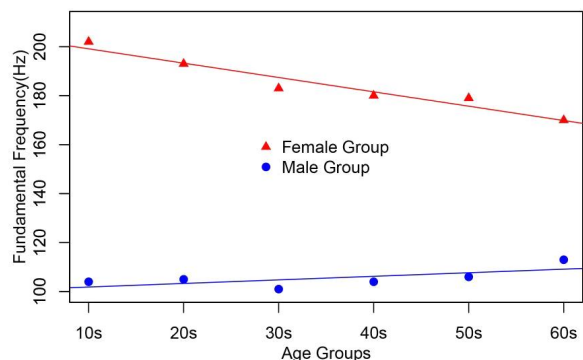


그림 5. 한국어 ‘일상 대화 음성 말뭉치 2020’의 남녀집단별 f0중앙값과 나이 간의 회귀분석 예측선

Figure 5. Ablines of the regression analyses between median f0 values and ages by male and female groups in the Korean ‘Dialogue Corpus (audio) 2020’

그림 5의 남성집단의 중앙값의 분포를 보면 여성집단의 음의 기울기와는 다소 다른 양의 기울기를 보이고 있는데 Yang(2021b)의 연구결과에서도 비슷한 모양을 나타냈다. 이러한 관계는 나이가 적은 참여자들의 신체가 60대 참여자들보다 더 크고, 이들의 성대의 크기와 무게도 상대적으로 더 커서, 피치값이 낮아졌

을 수도 있다(Yang, 1998, 2021b). 결국 이 남녀 집단의 회귀분석을 본다면, 앞의 전체집단의  $f_0$ 와 나이에 대한 관계는 참여자가 훨씬 많은 여성집단의 기여도가 주를 이루고 있음을 알 수 있다. 특히 참여자의 실제 나이를 사용하기보다는 10년 간격의 나이를 사용하는 데서  $f_0$  측정값들이 뒤섞여서 정밀한 예측값을 구할 수 없게 된 것이 이들 사이의 관계를 확정적으로 정의할 수 없는 근본적인 문제점이라 여겨진다. 덧붙여, 이 말뭉치에서 60대 화자가 146명이었고 70대가 1명으로 60대 이후의 참여자들이 많이 확보되지 않아 노년 화자들의 나이와  $f_0$ 의 관계를 보여주는 데 한계가 있다고 본다. 예를 들어, Hollien & Shipp(1972)의 연구에서는 20세에서 89세까지 남성 175명의 나이와  $f_0$  사이의 관계를 조사했는데, 20세에서 40세까지는 점점 내려가다가 60세에서 80세로 갈수록 올라가는 경향을 보였다고 한다. 이에 반해 여성의  $f_0$ 는 중년 이후에는 낮아지게 되어 남녀의  $f_0$ 가 서로 수렴하는 경향을 보였다(Baken, 2005, Figure 2 참조). 나이와  $f_0$  사이의 타당하고 신뢰할만한 관계를 정립하기 위해서는 앞으로 측정이 가능한 어린 연령대의 남녀집단과 노인 남녀집단에 대한 대규모의 연구가 필요할 것으로 여겨진다. 특히, 두 변수 사이의 근본적인 관계규명을 위해서는 참여자들의 해부학적인 성대모양과 조직의 변화에 대한 측정을 반영할 필요가 있다.

#### 4. 요약 및 결론

이 연구는 국립국어원에서 배포한 한국인 대화 음성 말뭉치에서 화자의 성대의 진동을 나타내는  $f_0$ 값을 측정해서 한국인이 일상 대화를 할 때  $f_0$ 값의 기초적인 통계자료를 구하고, 연령대와  $f_0$ 값의 관계를 조사했다. 연구자료 수집과 분석은 Praat과 R을 이용했으며, 개인별 억양구 단위 내에서의 상자도를 구해 극단값을 제거하는 방법으로 최종  $f_0$ 값자료를 구했다. 연구 결과는 다음과 같다.

전체 한국인들의  $f_0$ 값의 평균값은 185 Hz이고 중앙값은 187 Hz로 나왔다. 자료의 분포모양을 나타내는 왜도는 0.11의 정적분포를 보였고, 첨도는 -0.09로 정상분포에 거의 가까운 모양을 보였다. 일상대화의 피치값의 변화범위로는 상자도에서 윗수염 313 Hz와 아랫수염 75 Hz 사이의 238 Hz 범위로 나타났다. 남녀 간의 성별  $f_0$ 값의 차이는 남성의 중앙값 114 Hz의 거의 두 배에 해당하는 199 Hz가 여성의 중앙값으로 나타났고  $t$ 검증결과 유의미한 차이를 보였다. 분포모양을 나타내는 왜도는 남성이 1.24이었고, 여성은 그것의 반에 해당하는 0.58이었다. 첨도는 남녀집단 각각 5.21과 3.88로 나타나 남성의 값이 34% 정도 더 뾰족한 모양을 보였다. 연령대별로는 남녀집단을 합하여 보았을 때, 나이가 들수록  $f_0$ 값이 서서히 내려가는 경향을 보였는데, 상대적으로 많은 참여자가 녹음한 여성집단이 이런 관계를 주도하였지만, 남성화자에서는 이와는 다른 경향을 보여서, 남녀의 나이와  $f_0$ 값이 한 방향으로 수렴하지는 않았다.

결론적으로, 대규모 참여자가 녹음한 대화 음성 말뭉치에서 한국인의 집단별 연령별 기본적인  $f_0$ 분포를 살펴볼 수 있지만, 이 자료로  $f_0$ 값과 나이의 관계를 규명하기에는 한계가 있음

을 알 수 있었다.

이 연구에서는 개인별로 발화한 억양구별  $f_0$ 분포에서 사분위값의 범위를 이용해서 극단값을 제거했지만, 전체 집단에서도 여전히 극단값을 볼 수 있어서 앞으로 실제 데이터를 가장 잘 요약할 수 있는 통계 처리방법에 대한 연구가 더 필요하고,  $f_0$ 값과 나이의 관계를 규명하기 위해서는 성대의 변화가 많은 생후에서 청소년기로 진행되는 나이대별로 촘촘한 자료와 이를 보완할 수 있는 해부학적인 측정치를 확보하는 추가 연구를 제안한다.

#### 감사의 글

이 연구는 국립국어원의 ‘대화 음성 말뭉치 2020’을 내려받아  $f_0$ 값을 구했으며, 대규모 음성 자료를 만든 국립국어원에 감사드립니다.

#### References

- Baken, R. J. (2005). The aged voice: A new hypothesis. *Journal of Voice*, 19(3), 317-325.
- Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer (version 6.2.14). [Computer software]. Retrieved from <http://www.praat.org/>
- Boothroyd, A. (1986). *Speech acoustics and perception*. Austin, TX: Pro-ED.
- Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London, UK: Sage Publications.
- Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research*, 15(1), 155-159.
- Hudson, T., De Jong, G., McDougall, K., Harrison, P., & Nolan, F. (2007, August). F0 statistics for 100 young male speakers of Standard Southern British English. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany.
- Kinoshita, Y., Ishihara, S., & Rose, P. (2009). Exploring the discriminatory potential of  $f_0$  distribution parameters in traditional forensic speaker recognition. *Journal of Speech, Language and the Law*, 16(1), 91-111.
- Lennes, M., Stevanovic, M., Aalto, D., & Palo, P. (2015). Comparing pitch distributions using Praat and R. *Phonetician*, 111(2), 35-53.
- Lindh, J. (2006). *Preliminary descriptive  $f_0$ -statistics for young male speakers* (Research Report of Centre for Languages & Literature, Department of Linguistics & Phonetics, Lund University, pp. 89-92). Lund, Sweden.
- National Institute of Korean Language. (2023). Dialogue Corpus (audio) 2020. Retrieved from <https://corpus.korean.go.kr/>

- R Core Team. (2023). R: A language and environment for statistical computing (version 4.3.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Shi, S. J., Zhang, J., & Xie, Y. (2014, September). Cross-language comparison of F0 range in speakers of native Chinese, native Japanese and Chinese L2 of Japanese: Preliminary results of a corpus-based analysis. *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing* (pp. 241-244). Singapore, Singapore.
- Yang, B. (1990). *Development of vowel normalization procedures: English and Korean* (Doctoral dissertation). The University of Texas at Austin, Austin, TX.
- Yang, B. (1996). An analysis of Korean glottal waves by the laryngograph and a perceptual study of synthesized vowels. *Korean Journal of Linguistics*, 21(4), 1025-1040.
- Yang, B. (1998). A study of pitch analysis by Signalize. *Dongguk Nonjip*, 28, 68-79.
- Yang, B. (2018). Pitch trajectories of English vowels produced by American men, women, and children. *Phonetics and Speech Sciences*, 10(4), 31-37.
- Yang, B. (2021a). Measuring vowels. In R. A. Knight, & J. Setter (Eds.), *The Cambridge handbook of phonetics* (pp. 261-284). Cambridge, UK: Cambridge University Press.
- Yang, B. (2021b). The f0 distribution of Korean speakers in a spontaneous speech corpus. *Phonetics and Speech Sciences*, 13(3), 31-37.

• 양병곤 (Byunggon Yang) 교신저자  
 부산대학교 영어교육과 교수  
 부산시 금정구 부산대학로 63번길 2  
 Tel: 051-510-2619  
 Email: bgyang@pusan.ac.kr  
 관심분야: 음성학, 음운론

# Praat과 R로 분석한 한국인 대화 음성 말뭉치의 fundamental frequency(f0)값 분포\*

양 병 곤

부산대학교 영어교육과

## 국문초록

이 논문은 국립국어원에서 배포한 한국인 대화 음성 말뭉치에서 화자의 성대의 진동을 나타내는 fundamental frequency(f0)값을 측정해서 한국인이 일상 대화를 할 때 f0값의 기초적인 통계자료를 살펴보고, 나이와 f0값의 분포는 어떤 관계를 보이는지를 조사했다. 연구자료 수집과 분석은 Praat과 R을 이용했고, 개인별 억양구마다 상자도를 구하고 사분위값을 활용하여 극단값을 제거하는 방법으로 최종 f0값 자료를 구했다. 그 결과 전체 한국인들의 f0값의 평균값은 185 Hz이고 중앙값은 187 Hz로 나왔다. 자료의 분포모양을 나타내는 왜도는 0.11의 정적분포를 보였고, 첨도는 -0.09로 정상분포에 거의 가까운 모양을 보였다. 일상대화의 피치값의 변화범위로는 238 Hz로 나타났다. 남녀 간의 f0값의 차이는 남성의 중앙값 114 Hz의 거의 두 배에 해당하는 199 Hz가 여성의 중앙값으로 나타났고 t검증결과 유의미한 차이를 보였다. 분포모양을 나타내는 왜도는 남성이 1.24이었고, 여성은 그것의 반에 해당하는 0.58이었다. 첨도는 남녀집단 각각 5.21과 3.88로 나타나 남성의 값이 34% 정도 더 뽀족한 모양을 보였다. 연령대별로는 남녀집단을 합하여 볼 때, 나이가 들수록 f0값이 서서히 내려가는 경향을 보였다. 연령대별 f0중앙값과 나이 간의 회귀분석을 실행한 결과 기울기가 남성집단에서는 0.15, 여성집단에서는 -0.586으로 서로 반대되는 경향을 기록했다. 결론적으로, 대규모 참여자가 녹음한 대화 음성에서 한국인의 집단별 연령별 다양한 f0분포를 규명할 수 있지만, 나이와 f0관계는 더 정밀한 자료수집이 필요함을 알 수 있었다.

**핵심어:** fundamental frequency (f0), 한국인, 대화 음성 말뭉치, 성별, 나이별 통계값 분포

## 참고문헌

- 국립국어원(2023). 일상 대화 음성 말뭉치. 2020. <https://corpus.korean.go.kr/>
- 양병곤(1996). 라리로그그래프에 의한 한국인의 성문파형 분석 및 합성모음의 청각실험. *언어*, 21(4), 1025-1040.
- 양병곤(1998). Signalize 에 의한 피치 분석방법 고찰. *동의논집*, 28, 68-79.

\* 이 과제는 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음.