The f0 distribution of Korean speakers in a spontaneous speech corpus*

Byunggon Yang**

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

Abstract

The fundamental frequency, or f0, is an important acoustic measure in the prosody of human speech. The current study examined the f0 distribution of a corpus of spontaneous speech in order to provide normative data for Korean speakers. The corpus consists of 40 speakers talking freely about their daily activities and their personal views. Praat scripts were created to collect f0 values, and a majority of obvious errors were corrected manually by watching and listening to the f0 contour on a narrow-band spectrogram. Statistical analyses of the f0 distribution were conducted using R. The results showed that the f0 values of all the Korean speakers were right-skewed, with a pointy distribution. The speakers produced spontaneous speech within a frequency range of 274 Hz (from 65 Hz to 339 Hz), excluding statistical outliers. The mode of the total f0 data was 102 Hz. The female f0 range, with a bimodal distribution, appeared wider than that of the male group. Regression analyses based on age and f0 values yielded negligible R-squared values. As the mode of an individual speaker could be predicted from the median, either the median or mode could serve as a good reference for the individual f0 range. Finally, an analysis of the continuous f0 points of intonational phrases revealed that the initial and final segments of the phrases yielded several f0 measurement errors. From these results, we conclude that an examination of a spontaneous speech corpus can provide linguists with useful measures to generalize acoustic properties of f0 variability in a language by an individual or groups. Further studies would be desirable of the use of statistical measures to secure reliable f0 values of individual speakers.

Keywords: f0, distribution, statistics, variability, Korean corpus, spontaneous speech

1. Introduction

A speaker’s vibrations of vocal folds are acoustically measured by the fundamental frequency or f0 (Fant, 1973). The fundamental frequency is also referred to as pitch, emphasizing the perceptual dimension of a sound property (Zheng & Brette, 2017). People can roughly identify a speaker’s sex or age group by listening to the range of pitch from a brief excerpt of his/her speech. f0 tends to vary depending on intrinsic anatomical factors and extrinsic psychological factors (Yang, 1990). f0 can be regarded as inversely proportional to the mass and length of the vocal fold and proportional to the tension (Boothroyd, 1986). Thus, the vocal fold vibration of women, who have smaller vocal cords, is predictably faster than that of men. It also varies as the speaker changes the tension of the laryngeal

* This work was supported by a 2-Year Research Grant of Pusan National University.
** byang@pusan.ac.kr, Corresponding author
Received 27 July 2021; Revised 7 September 2021; Accepted 7 September 2021
© Copyright 2021 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.
muscles and, to some extent, subglottal pressure (Lieberman, 1967). Anatomically, the vocal folds can be lengthened with increased tension when the cricothyroid muscle contracts, causing the cricoid to tilt back and thereby stretching the vocal folds. In this way, speakers can express their emotional state, controlling the speed of vocal fold vibration. Boothroyd (1986) reported f0 ranges from 70 to 200 Hz in men, 140 to 400 Hz in women, and 180 to 500 Hz in children. In sum, f0 may vary greatly both by a speaker’s anatomy and emotional state during speech communication.

Few studies have dealt with f0 measurements in spontaneous speech (Lennes et al., 2016). Lennes et al. compared f0 distributions of two conversational Finnish speech corpora. One contained ten conversations between young, native Finnish-speaking adults; the other included shorter dialogues among eight adult men and women. They reported that the distribution of probability density curves for all speakers was right-skewed, which may be related to human vocal fold properties that can be stretched to produce higher f0 values, but not below the anatomical contraction limit. This study attempted to apply some of their procedures and methods to a Korean corpus.

Linguists and phoneticians have attempted to explore the subtle but complicated usage of f0 by people in daily conversations. Couper-Kuhlen (1996) noticed that the functional significance of f0 in conversation lies mainly in relation to the speaker-specific f0 range. Ladd (1996) observed that the variability of f0 and relative f0 levels are necessary to establish theories of intonational phonology. Morrill (2012) investigated the phonetic implementation of stress in American English adjective-noun compounds in different prosodic positions and sentence types. He measured f0, vowel duration, and intensity of compounds and phrases; he found some distinctive f0 patterns along with interactions with other phonetic cues, depending on the intonational and prosodic environments. In his study, Morrill converted all f0 values from linear Hertz to semitone values. This transformation might approximate the auditory scale of the acoustic measurements. In his Figure 5, the sex difference is still observable in the utterance of “blue Book/Blue Book.” The patterns look quite similar, but the contour might not converge when they collapsed the data into one figure. He showed that the statistical comparison of f0 values in both Hertz and semitone yielded nearly identical patterns, with small differences in F values (see Table 5 and Appendix A in his paper). Similarly, Lennes et al. (2016) transformed the f0 values into a semitone scale to normalize physiological variability. The transformation did not fully standardize the distribution, as depicted in their Figure 3. Even after the transformation, the difference between men and women still prevailed. In other words, the auditory scale should incorporate perceptual results from speech-like stimuli into its revision. In addition, Kunter (2011) found that listeners relied more on f0 in the perception of prominence in the left-prominent noun-noun compounds extracted from a speech corpus. He also reported high rates of variability in the perception of prominence in compounds with right prominence. Since a speaker controls his/her own f0 by listening to the produced sound, the auditory aspect of the speech should also be considered to explain factors related to acoustic variability. Medeiros et al. (2021) compared measures of f0 variability between singing and speech after building a database with parallel singing and speech recordings. They confirmed the hypothesis of higher f0 stability in singing than in speaking, and applied a machine learning method to classify them from f0 values at the syllable level with an accuracy of 80%. f0 variability can be an interesting measure to describe individual or group characteristics of speech.

Through bootstrapping using random resampling of the collected data (Efron, 2003), Lennes et al. (2016) found that at least 34 seconds of speaking time would yield the f0 mode of many speakers, with a standard deviation (SD) of 1 semitone. For some speakers in the second set of the two corpora, they found unstable bimodal distributions after three minutes of speaking time. They attributed the bimodal distribution to the background noise and short recording. Nolan (1983) suggested that one minute of speech would reliably capture individuals’ within-speaker variation in f0. Moreover, Catford (1977) showed that the voiced-to-voiceless ratio in the utterance differs from language to language. For example, 78% of utterances in French are reported to have voiced segments, while 41% of Cantonese segments are voiced, from which we can posit that an appropriate duration to specify individual f0 range may vary depending on a language.

To date, few studies have explored an f0 distribution of Korean speakers, despite the prosodic importance of f0 in daily conversations. The main purpose of this study was to provide normative data on Korean speakers’ f0 values in spontaneous speech for phoneticians and linguists, as well as for a cross-linguistic comparison. Specifically, the current study was designed to investigate (1) descriptive statistics on the collected f0 data; (2) group and individual f0 variability; (3) relationship between f0 and age; and finally (4) median f0 contours of intonational phrases.

2. Method

2.1. The speech corpus and participants

The Korean Corpus of Spontaneous Speech consists of 240 sound files (Yun et al., 2015). Forty speakers of Seoul Korean participated in the recordings. They were divided into four age groups in their 10s, 20s, 30s, and 40s and consisted of five males and five females in each age group. The participants talked freely about general topics, including their personal backgrounds, and expressed views on miscellaneous issues. The Korean corpus also provides transcript files of each speaker’s production in detail. An hour-long recording per participant may provide sufficient data to establish individual or group f0 distribution.

2.2. Data measurements and analyses

f0 values were collected using Praat (v.6.0.46, Boersma & Weenink, 2019), and their distribution was analyzed through statistical analyses on the data using R (R Core Team, 2021).

Several Praat scripts were created to edit the sound files and to obtain valid and reliable f0 values. The sound files were preprocessed using a script to silence sound segments of noise or laughter to zero amplitude, referring to the temporal information of those segments in each transcript file. Another script measured f0 values of all the files every 20 ms, with a range of 75 to 600 Hz, and appended them to the data on a notebook computer. A third script allowed the author to jump around the sound segments of each speaker and to visually cross-check the f0 values in the files. Every 2 seconds of a given sound segment was displayed on the computer window, as shown in Figure 1. Major errors of f0 measurements were corrected while watching the sound wave, narrow-band spectrogram, and pitch object in parallel. The pitch contour on a narrow-band spectrogram, with a window size setting of 0.029 ms,
was quite useful to find a majority of measurement errors, such as \( f_0 \) halving and doubling (Murray, 2001). For example, the final five \( f_0 \) points of the second pitch contour or an intonational phrase, and the initial six \( f_0 \) points of the fourth pitch contour, recorded \( f_0 \) halvings in the edit window and were corrected later, as seen in the final pitch object in the figure. A majority of errors were octave jumps or drops found at initial and final measurement points of a continuous pitch contour. In the analysis, the Korean stops, fricatives, and affricates seemed to produce measurement errors.

![Figure 1](https://example.com/figure1.png)

**Figure 1.** A screen shot of \( f_0 \) correction. The \( f_0 \) halving errors in dotted circles were adjusted upward to form contours straight to the previous \( f_0 \) points in the lower \( f_0 \) object window below the filled arrows

The \( f_0 \) and speech wave windows were often zoomed in on to verify the value directly from the duration of each vocal fold vibratory cycle (Yang, 1998; 2018; 2021). The overall \( f_0 \) range of each speaker was also considered to remove any separate island or protruding \( f_0 \) values that were visually deviant from adjacent \( f_0 \) dots. The final decision on the valid measures was made after listening to synthesized pitch humming of an expanded speech segment around a selected segment for analysis.

Statistical analyses on the distribution of \( f_0 \) values were conducted using R. An R script proposed by Lennes et al. (2016) was modified to draw probability density curves of the \( f_0 \) distribution of age and sex groups, and to collect modes at the highest peak. The mode, median, and mean values, along with the SD, were obtained to discuss group and individual variability. Additional measures, such as skewness and kurtosis, were also collected to describe group characteristics. Moreover, an \( f_0 \) range or bandwidth of each individual speaker between the lower/upper bounds and his/her median or mode, as a reference point, was examined after collecting corresponding values through a boxplot function of R. The \( f_0 \) bandwidth might be a useful criterion in a cross-linguistic comparison. Finally, each voiced point of an \( f_0 \) contour was numbered from the start to the end. Then, the median values at each voiced point of the intonational phrases were plotted to display a general pitch distribution of consecutive \( f_0 \) contours to identify which part of the contour had the most pitch errors. In Figure 1, there are 27 voiced points in the first intonational phrase, followed by 20 voiced points in the second intonational phrase.

3. Results and Discussion

3.1. \( f_0 \) distribution of the Korean corpus

Table 1 lists a statistical summary of all the \( f_0 \) values of the Korean speakers. The final dataset included 2,996 million \( f_0 \) values.

| n      | Mode | Median | Mean | SD | Min | Max |
|--------|------|--------|------|----|-----|-----|
| 2,996,164 | 102 | 148    | 160  | 62 | 65  | 624 |

SD, standard deviation.

As shown in the table, the mode or most frequent value was 102 Hz. The median or middle number in the ordered \( f_0 \) data was 148 Hz. The mean value of 160 Hz, with an SD of 62 Hz, was highest among the three statistical measurements. The skewness was 0.92, which implies a positive skew in which the frequent \( f_0 \) values are at the lower end, and the tail points toward the higher end of the scale (Field, 2013). The kurtosis was 0.94, which yielded a leptokurtic or pointy distribution. The difference between the mean and the median was 12 Hz, while that between the mode and the median was 46 Hz, which may be related to the right-skewed extreme \( f_0 \) values. The outliers in the higher frequency region must have pulled the mean upward. In that sense, the median seems to do a better job of summarizing the \( f_0 \) values, avoiding any biased statistics caused by extremely low or high \( f_0 \) values. The mode, as a representative value for all Korean speakers, may leave room for discussion when considering the group differences in male and female \( f_0 \) values.

Figure 2 illustrates a box plot of the \( f_0 \) values collected from all 40 Korean speakers every 20 ms from the nearly 40 hours of the Korean corpus.

![Figure 2](https://example.com/figure2.png)

**Figure 2.** Box plot of \( f_0 \) values (Hz) of forty speakers in the Korean corpus

As seen in the figure, the lower bound was at 65 Hz, and the upper bound was at 339 Hz. Over or under these bounds are generally considered outliers in statistics. Thus, the frequency band of 274 Hz, between the lower and upper bound, could be used to describe the general \( f_0 \) range of Korean speakers. Many outliers form a long tail above the upper bound, but no outliers are seen below the lower bound. Here, we would like to remind the reader of the procedure of \( f_0 \) correction, through which a majority of extreme values must have been screened from the final dataset first. Otherwise, the statistics, such as the three major statistical measurements, may have been biased by those values. In addition, the first or lower quartile amounted to 107 Hz, while the third or upper quartile was at 200 Hz, from which we can calculate the interquartile range: 93 Hz.

An upper bound for outliers can be roughly guessed by multiplying the interquartile range by 1.5 and adding the third quartile value, which yields 339.5 Hz, which is 0.5 Hz above the obtained value from the R boxplot statistics. As presented in Figure 2, the bandwidth
between the median and the lower bound (83 Hz) seemed narrower than that between the median and the upper bound (191 Hz), which clearly defines the right skewness in specific digits.

### 3.2. f0 statistics of groups by sex and individual speakers

Table 2 lists basic statistical measures of the data grouped by sex. Figure 3 shows the probability density curves of the f0 values of each individual speaker in the Korean corpus. The f0 values of male and female speakers are drawn in blue and red, respectively.

**Table 2.** Statistics of f0 values grouped by sex in the Korean corpus. n denotes the total number of the measured f0 values. Min indicates the minimum f0 value while Max represents the maximum one (unit: Hz)

| Group   | n   | Mode | Median | Mean | SD   | Min | Max |
|---------|-----|------|--------|------|------|-----|-----|
| Male    | 1,540,262 | 103 | 111   | 116  | 26   | 62  | 465 |
| Female  | 1,455,902 | 186 | 200   | 206  | 54   | 69  | 630 |

SD, standard deviation.

**Figure 3.** The probability density curves of the individual f0 values grouped and colored by the male (M) and female (F) speakers in the Korean corpus

The mode of the male speakers was 103 Hz, while that of the female speakers was 186 Hz. There was an 83 Hz difference between the two groups. The f0 values of the male and female speakers were clustered around the modes. The mode seemed to be a good indicator of the group characteristics. The median of the male group was 111 Hz, while that of the female group was 200 Hz. The means of the male and female groups were 116 and 206 Hz, respectively. The differences between the two measures of the male and female groups were 5 and 6 Hz, respectively. The SD, 26.4 Hz in the male group, amounted to almost double the deviation of the female group, 53.7 Hz. A wider dispersion of the female group is apparent in the figure. The anatomical sex difference in the length and mass of the vocal folds may account for the separation of the two groups, as described in the introduction. Moreover, the distributions in both groups appear right-skewed. The skew for the male and female data were 1.34 and 0.66, respectively, while the kurtosis for the data amounted to 2.88 and 2.37. The skewed distribution may be partly due to the length of the vocal folds, which sets a natural lower limit to glottal frequency, whereas those speakers can stretch their vocal folds to produce higher f0 (Lennes et al., 2016). Thus, we can see a long tail over the median of each group. Lennes et al. also reported the overall mean pitch and SDs of Finish male speakers as 117.5 Hz (SD: 33 Hz) and for female speakers as 191.3 Hz (SD: 44.6 Hz). These values are slightly higher than those of the Korean group means. The Finnish participants were mostly young adults. Lennes et al. displayed similar f0 density curves of the two groups (as seen in their Figure 3), but the skewness had a tail shorter than that in Figure 3 because they transformed the acoustic frequency values on an auditory scale or semitone, which matches a logarithmic scale of base 2. Here, clipping of potential f0 values below the threshold might have occurred because of the lower f0 limit to 75 Hz in the Praat software itself. The f0 range of the female group seemed wider than that of the male group. In addition, the female group had a bimodal distribution. The lower valley of the female group went below 100 Hz; a majority of the female speakers showed a similar bimodal distribution. The lower peak in the female speakers seems to have influenced the mode, mean, and median.

Table 3 summarizes the statistics of each individual speaker. The mode ranged from 81 (s23) to 215 Hz (s20). The median spread from 93 (s02) to 227 Hz (s20). The mean ranged from 101 (s02) to 235 Hz (s20), while the SD ranged from 16 (s11) to 71 Hz (s39). As portrayed in Figure 3, individual values of the male group generally converged around the median, and those in the female group converged around the median. The spread might be related to the higher f0 values of the female participants as well as the bimodal distribution. Since the mode indicates the most frequent value or the maximum value in the probability density graph, this value would be a good measure for describing the individual characteristics of each speaker. Instead, we delve into a detailed analysis of the characteristics in the following section after checking for a relationship between f0 and age. Here, we may have to think about the mode as a valid statistical measure for all speakers or each group by sex. When the two groups were pooled together, the mode was 102 Hz (as depicted in Table 1), which reflects more of the pointy distribution of the male speakers, as well as the lower peak of the female group. In that sense, the medians might better represent their characteristics after removing all outliers.

**Table 3.** Statistical summary of f0 values of each individual speaker of the Korean corpus

| Subj | Mode | Median | Mean | SD | Subj | Mode | Median | Mean | SD |
|------|------|--------|------|----|------|------|--------|------|----|
| s01  | 106  | 115    | 122  | 25 | s21  | 104  | 115    | 123  | 26 |
| s02  | 85   | 93     | 101  | 24 | s22  | 94   | 106    | 113  | 28 |
| s03  | 106  | 110    | 114  | 19 | s23  | 81   | 96     | 105  | 28 |
| s04  | 93   | 102    | 108  | 23 | s24  | 104  | 112    | 119  | 27 |
| s05  | 114  | 121    | 126  | 22 | s25  | 122  | 126    | 130  | 17 |
| s06  | 165  | 177    | 183  | 52 | s26  | 190  | 203    | 208  | 63 |
| s07  | 182  | 200    | 209  | 41 | s27  | 186  | 196    | 200  | 49 |
| s08  | 195  | 209    | 214  | 41 | s28  | 184  | 195    | 203  | 44 |
| s09  | 191  | 202    | 207  | 41 | s29  | 182  | 187    | 191  | 43 |
| s10  | 189  | 198    | 202  | 43 | s30  | 157  | 168    | 177  | 42 |
| s11  | 100  | 105    | 108  | 16 | s31  | 105  | 118    | 127  | 33 |
| s12  | 89   | 96     | 102  | 21 | s32  | 106  | 111    | 114  | 17 |
| s13  | 92   | 100    | 106  | 23 | s33  | 97   | 106    | 111  | 23 |
| s14  | 100  | 113    | 120  | 27 | s34  | 119  | 133    | 139  | 29 |
| s15  | 103  | 112    | 119  | 25 | s35  | 103  | 109    | 112  | 19 |
| s16  | 202  | 214    | 219  | 40 | s36  | 176  | 205    | 219  | 65 |
| s17  | 207  | 218    | 230  | 47 | s37  | 174  | 184    | 186  | 57 |
| s18  | 195  | 206    | 209  | 49 | s38  | 190  | 212    | 223  | 68 |
| s19  | 158  | 171    | 182  | 43 | s39  | 200  | 214    | 220  | 71 |
| s20  | 215  | 227    | 235  | 38 | s40  | 185  | 201    | 213  | 53 |

SD, standard deviation.
3.3. Individual f0 variability

The kurtosis and skewness measures may offer a general picture of the overall or group distribution. Instead, we wanted to explore some more detailed characteristics of the individual f0 variability. Specifically, this section focuses on individual f0 bandwidths between a reference point and either the lower bound or the upper bound. Table 4 lists the f0 median, mode, lower bound, and upper bound, as well as the bandwidth from the reference median and mode of the individual Korean male and female speaker.

Table 4. f0 mode (Mo), median (Me), lower bound (LB) and upper bound (UB) and bandwidth from the lower/upper bound and mode/median of the individual male and female speaker in the Korean corpus to the upper and lower bound.

| Subject | LB | Mo | Me | UB | UB-LB | UB-Me | Me-LB | UB-Mo | Mo-LB |
|---------|----|----|----|----|-------|-------|-------|-------|-------|
| s01     | 75 | 106| 115| 170| 95    | 55    | 40    | 64    | 31    |
| s02     | 72 | 85 | 93 | 145| 73    | 52    | 21    | 60    | 13    |
| s03     | 75 | 106| 110| 156| 81    | 46    | 35    | 50    | 31    |
| s04     | 75 | 93 | 102| 159| 84    | 57    | 27    | 66    | 18    |
| s05     | 75 | 114| 121| 173| 98    | 52    | 46    | 59    | 39    |
| s06     | 89 | 165| 177| 273| 184   | 96    | 88    | 109   | 76    |
| s07     | 110| 182| 200| 302| 192   | 102   | 90    | 120   | 72    |
| s08     | 116| 195| 209| 311| 195   | 102   | 136   | 116   | 79    |
| s09     | 126| 191| 202| 286| 160   | 84    | 76    | 95    | 65    |
| s10     | 130| 189| 198| 274| 144   | 76    | 68    | 85    | 59    |
| s11     | 75 | 100| 105| 147| 72    | 42    | 30    | 47    | 25    |
| s12     | 75 | 89 | 96 | 147| 72    | 51    | 21    | 58    | 14    |
| s13     | 75 | 92 | 100| 150| 75    | 50    | 25    | 58    | 17    |
| s14     | 75 | 100| 113| 187| 112   | 74    | 38    | 87    | 25    |
| s15     | 75 | 103| 112| 172| 97    | 60    | 37    | 69    | 28    |
| s16     | 144| 202| 214| 291| 147   | 77    | 70    | 89    | 58    |
| s17     | 123| 207| 218| 326| 203   | 108   | 95    | 119   | 84    |
| s18     | 118| 195| 206| 302| 184   | 96    | 88    | 107   | 77    |
| s19     | 85 | 158| 171| 272| 187   | 101   | 86    | 114   | 73    |
| s20     | 155| 215| 227| 307| 152   | 80    | 72    | 92    | 60    |
| s21     | 75 | 104| 115| 181| 106   | 66    | 40    | 77    | 29    |
| s22     | 65 | 94 | 106| 178| 113   | 72    | 41    | 84    | 29    |
| s23     | 70 | 81 | 96 | 173| 103   | 77    | 26    | 92    | 11    |
| s24     | 70 | 104| 112| 173| 103   | 61    | 42    | 69    | 34    |
| s25     | 88 | 122| 126| 168| 80    | 42    | 38    | 46    | 34    |
| s26     | 90 | 190| 203| 325| 235   | 122   | 113   | 135   | 100   |
| s27     | 112| 186| 196| 291| 179   | 95    | 84    | 105   | 74    |
| s28     | 117| 184| 195| 280| 165   | 83    | 78    | 96    | 67    |
| s29     | 126| 182| 207| 357| 137   | 115   | 60    | 188   | 56    |
| s30     | 79 | 157| 168| 266| 187   | 98    | 89    | 109   | 78    |
| s31     | 75 | 105| 118| 200| 125   | 82    | 43    | 95    | 30    |
| s32     | 75 | 106| 111| 150| 75    | 39    | 36    | 44    | 31    |
| s33     | 75 | 97 | 106| 160| 85    | 54    | 31    | 63    | 22    |
| s34     | 75 | 119| 133| 212| 137   | 79    | 58    | 93    | 44    |
| s35     | 75 | 103| 109| 159| 84    | 50    | 34    | 56    | 28    |
| s36     | 75 | 176| 205| 565| 290   | 160   | 130   | 189   | 101   |
| s37     | 75 | 174| 184| 302| 227   | 118   | 109   | 128   | 99    |
| s38     | 86 | 190| 212| 350| 264   | 138   | 126   | 160   | 104   |
| s39     | 94 | 200| 214| 350| 256   | 136   | 120   | 150   | 106   |
| s40     | 98 | 185| 201| 317| 219   | 116   | 103   | 132   | 87    |

The mean bandwidth of all the speakers between the upper bound and the lower bound in this study was 144 Hz (SD: 61 Hz, range (max-min): 72–290 Hz). The mean bandwidth between the upper bound and the median was 81 Hz (SD: 29 Hz, range: 39–160 Hz), while that between the upper bound and the median was 64 Hz (SD: 32 Hz, range: 21–130 Hz). The upper or right side of the distribution had a longer tail by a difference of 17 Hz. The deviation of the mean bandwidth was comparable by a 3 Hz difference. On the other hand, those equivalent taken from the mode were as follows: the bandwidth between the upper bound and the mode was 92 Hz (SD: 34 Hz, range: 44–189 Hz), while that between the upper bound and the mode was 53 Hz (SD: 29 Hz, range: 11–106 Hz). The difference between the two bandwidth measurements amounted to 39 Hz, and the SDs of the measurements were switched. The reason why we had a smaller median bandwidth between the upper bound and the mode comes from the fact that the total mode (102 Hz) recorded a lower value away from the total median (148 Hz). The mode may better define the f0 characteristics of an individual speaker. However, the model needs an additional R script to make determinations from a large set of data, while the median can be obtained immediately along with the lower and upper bounds from a boxplot analysis. Figure 4 illustrates the individual bandwidth of the f0 distribution. If we look at the figure, the relationship is quite consistent, with both the median and mode as reference points. The following regression analyses also support the idea of safely using one of the measures.

![Image](Byunggon_Yang_Phonetics_and_Speech_Sciences_Vol.13_No.3_(2021)_31-37)
3.4. Relationship between age and \( f_0 \)

As people get older, their vocal cords tend to grow and produce lower \( f_0 \) values. We would like to examine any potential relationship between age and \( f_0 \) values in the Korean corpus. Table 5 lists the median \( f_0 \) values of male and female speakers grouped by age.

**Table 5.** Median \( f_0 \) values of male and female speakers by age in the Korean corpus

| Male age | Male \( f_0 \) median | Female age | Female \( f_0 \) median |
|----------|-----------------------|------------|------------------------|
| 15       | 106                   | 16         | 199                    |
| 16       | 110                   | 17         | 199                    |
| 22       | 112                   | 18         | 177                    |
| 23       | 105                   | 22         | 214                    |
| 25       | 105                   | 24         | 214                    |
| 26       | 100                   | 27         | 206                    |
| 31       | 115                   | 32         | 199                    |
| 32       | 126                   | 34         | 195                    |
| 36       | 106                   | 37         | 187                    |
| 37       | 106                   | 38         | 168                    |
| 43       | 113                   | 43         | 207                    |
| 44       | 106                   | 46         | 199                    |
| 47       | 133                   |            |                        |

The median age of all speakers was 32 years old ranging from 15 to 47 years. Ten groups of two or three male and female speakers were of the same age. Neither group in the table seemed to show any systematic decrease in \( f_0 \) by age. We conducted regression analyses on the age and \( f_0 \) values of the male (\( f_{0m} \)) and female (\( f_0 \)) speakers separately, considering a large \( f_0 \) difference between them, and obtained the following results:

\[
\text{lm(formula = } f_{0m} \sim \text{Age})
\]

| Coefficients: Estimate Std. Error t value Pr(>|t|) |
|-------------|-------------|-------------|---------------|
| (Intercept) | 0.01        | 0.067       | 1.579,5      | \(<2e-16\) |
| Age         | 0.34        | 0.002       | 166.9        | \(<2e-16\) |
| Residual standard error: 26.12 on 1,540,260 degrees of freedom |
| Adjusted R-squared: 0.018 (\(p<.05\)) |

\[
\text{lm(formula = } f_0 \sim \text{Age})
\]

| Coefficients: Estimate Std. Error t value Pr(>|t|) |
|-------------|-------------|-------------|---------------|
| (Intercept) | 206.1       | 0.135       | 1,524.3      | \(<2e-16\) |
| Age         | -0.002      | 0.004       | -0.577       | 0.564       |
| Residual standard error: 53.66 on 1,455,900 degrees of freedom |
| Adjusted R-squared: <0.001 (\(p<.05\)) |

From the two regression analyses above, both male and female intercepts were significant. Age in the male \( f_0 \) data, as a predictor, was significant, but not in the female \( f_0 \) data. Despite the significance of both factors, the adjusted R-squared values of both groups were very low: 0.018 and <0.001, respectively. These results imply that the age effect on \( f_0 \) is significant, but very small. The strong significance may be related to the enormous size of the data. Further research with much more participants under or over the age range may be desirable to define the relationship appropriately.

3.5. Median \( f_0 \) contours of intonational phrases

This section addresses median \( f_0 \) values at numbered points of measurement of intonational phrases or continuous voiced points in the utterance to examine any error pattern in the data measurements. Figure 5 illustrates the median \( f_0 \) values at each voiced segment grouped by sex. A horizontal line of the male group was drawn at 111 Hz, which was the median of all male speakers. Another horizontal line was drawn at 200 Hz, which was the median of all female speakers.

![Figure 5. Median \( f_0 \) values of the male (M) and female (F) speakers at the continuous voiced points in the Korean corpus. Two horizontal lines of the medians of male and female groups are drawn at 111 Hz and 200 Hz, respectively.](image)

4. Summary and Conclusion

The fundamental frequency, or \( f_0 \), plays an important role in the exploration of human speech. The current study examined \( f_0 \)
distributions of a Korean corpus of spontaneous speech to provide normative data for the speakers. The sound files of the corpus were analyzed using the speech analysis software Praat to collect f0 values, and to correct a majority of obvious errors manually by both watching and listening to the f0 contour on a narrow-band spectrogram. Descriptive statistical analyses on the distribution of collected f0 values were conducted using MATLAB. The f0 values of Korean speakers were right-skewed with a pointy distribution. They produced a frequency bandwidth of 274 Hz, from 65 to 339 Hz in spontaneous speech, excluding statistical outliers. The mode of the total f0 data was 102 Hz, which may represent the most frequent f0. From the probability density function of the f0 values of each individual speaker, we found that the female f0 bandwidth, with a bimodal distribution, seemed wider than that of the male group. Regression analyses based on age and f0 showed negligible R-squared values. Individual f0 analyses revealed the right skewness in specific digits. The mean bandwidth between the upper bound and the median was 81 Hz, while that between the upper bound and the median was 64 Hz. The upper or right side of the distribution had a longer tail by a difference of 17 Hz. The reference point of the mode can be predicted from the median so that either the median or mode can be used as a good reference point of the f0 bandwidth of an individual speaker. Finally, an analysis of the f0 pattern of international phrases indicated that the initial points and final sections of the pitch contours produced several f0 errors, which require attention by researchers. From these results, we conclude that an analysis of a spontaneous speech corpus can provide researchers with useful measures to describe acoustic properties of f0 variability in a given language by an individual or a group of speakers.

We would like to mention that there are issues related to the f0 measurement errors and incomplete corrections in the data. In addition, we did not discuss the distribution after transforming the acoustic f0 values into a semitone scale, considering that the auditory scale still needs revision.

References

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (version 6.0.46) [Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/

Boothroyd, A. (1986). Speech acoustics and perception. Austin, TX: Pro-Ed.

Catford, J. C. (1977). Fundamental problems in phonetics. Edinburgh, UK: Edinburgh University Press.

Couper-Kuhlen, E. (1996). The prosody of repetition: On quoting and mimicry. In E. Couper-Kuhlen & M. Selting (Eds.), Prosody in conversation (pp. 366-405). Cambridge, UK: Cambridge University Press.

Efрон, B. (2003). Second thoughts on the bootstrap. Statistical, 18(2), 135-140.

Fant, G. (1973). Speech sounds and features. Cambridge, MA: MIT Press.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. London, UK: Sage.

Kunter, G. (2011). Compound stress in English. The phonetics and phonology of prosodic prominence. Berlin, Germany: De Gruyter.

Ladd, D. (1996). Intonational phonology. (Cambridge Studies in Linguistics 79). Cambridge, UK: Cambridge University Press.

Lennes, M., Stevanovic, M., Aalto, D., & Palo, P. (2016). Comparing pitch distributions using Praat and R. Phonetica, 111(2), 35-53.

Lieberman, P. (1967). Intonation perception and language. Cambridge, MA: MIT Press.

Medeiros, B. R., Cabral, J. P., Meireles, A. R., & Baceti, A. A. (2021). A comparative study of fundamental frequency stability between speech and singing. Speech Communication, 128, 15-23.

Morrill, T. (2012). Acoustic correlates of stress in English adjectival compounds. Language and Speech, 55(2), 167-201.

Murray, K. (2001). A study of automatic pitch tracker doubling/halving “Errors”. Proceedings of the Second SIGdial Workshop on Discourse and Dialogue. Philadelphia, PA.

Nolan, F. J. (1983). The phonetic bases of speaker recognition. Cambridge, UK: Cambridge University Press.

R Core Team. (2021). R: A language and environment for statistical computing (version 4.1.0) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Yang, B. (1990). Development of vowel normalization procedures: English and Korean (Doctoral dissertation). The University of Texas, Arlington, TX.

Yang, B. (1998). A study of pitch analysis by Signalize. Dongeui 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. (2021). Measuring vowels. In R. A. Knight, & J. Setter (Eds.), The Cambridge handbook of phonetics (pp. 261-284). Cambridge, UK: Cambridge University Press.

Yun, W., Yoon, K., Park, S., Lee, J., Cho, S., Byun, K., Hahn, H., & Kim, J. (2015). The Korean corpus of spontaneous speech. Phonetics and Speech Sciences, 7(2), 103-109.

Zheng, Y., & Brette, R. (2017). On the relation between pitch and level. Hearing Research, 348, 63-69.

• Byunggon Yang, Corresponding author

Professor, Dept. of English Education

Pusan National University

30 Changiundong, Keumjunggu, Pusan, 46261 Korea

Tel: +82-51-510-2619

Email: byang@pusan.ac.kr

Homepage: http://fonetiks.info/byang

Fields of interest: Phonetics, Phonology