An improved method of F0 determination

Grandon Goertz[1], Terese Anderson[2]

[1]Educational Linguistics, University of New Mexico

[2]Chemistry and Chemical Biology, University of New Mexico

https://doi.org/10.36505/ExLing-2021/12/0027/000500

Abstract

This research investigated the structure of the speech Fundamental Frequency (F0). The principal objective of this study was to create an improved method for analyzing, understanding, and plotting the F0 values of spoken words. The F0 is created by a variety of models that produce the F0 values for plotting. These F0 values are a represented average of the lower range frequency of sound.

Keywords: fundamental frequency, sub-harmonic modelling, Chebyshev transform

Introduction

F0 is often understood to be a single, slightly varying waveform resulting from vocal fold vibrations in the throat, which would be consistent with the understanding of a hollow tube model of the throat that is often used to explain the phenomenon.

F0 determination technology

The F0 of speech is an estimated value and is not measured. Tsangas et al. (2014) and Staudacher, et al. (2016) report that there are at least 10 pitch estimation algorithms for use by linguists. F0 estimation algorithms include Nearly Defect-Free (NDF), TEMPO, SWIPE, SHRP, AAC, and YIN. (Pertusa and Iñesta 2012). Roark (2006) mentioned that there are more than 70 algorithms to estimate F0, which illustrates the difficulty in determining fundamental frequencies. Verde, De Pietro and Sannino (2018) describe a personalized methodological F0 determination that provides better than 77% accuracy, 72% sensitivity and 81% specificity.

Speech contains higher formant bands that are integer multiples of a common low frequency, the fundamental frequency. These harmonic sounds evoke a pitch corresponding to their F0 (Micheyl, Ryan, & Oxenham 2012). The fundamental frequency is a source for the harmonics and must be estimated, inferred, or modelled from the harmonic values (Pertusa and Iñesta 2012; Tsanas, Zañartu, Little & McSharry 2012). Staudacher et al. (2016) mention that, “Many pitch detection algorithms (PDAs) analyze a speech signal by partitioning it into segments and calculating the respective fundamental frequencies (short-term analysis). The length of the segments (frames) limits the minimum frequency or the maximum period to be determined.”

Physiology and F0 modeling

Tsanas, Zañartu, Little & McSharry (2012) mention that "The estimation of the fundamental frequency (f_0) is a critical problem in the acoustic characterization of speech signals." They observe that without genuine ground truth information it is impossible to validate many of the F0 determination algorithms. Tsanas et al. (2012) also summarize that the F0 varies in time, the F0 may change between vocal cycles, the sub-harmonics of the actual F0 appear frequently, and the vocal tract resonances affect the vocal folds resulting in harmonics which may be multiples of the actual F0. For their research Tsanas et al. (2012) generated 92 sustained vowel /ɑ/ sounds from physiological examinations of humans and computed the ground truth F0 time series.

Two main variables that affect the range of vocal fold vibration frequency are vocal fold elongation and tissue fiber stress, but other factors come into play. Tsanas et al. (2014) notes that the F0 can change if the F0 is affected by vocal tract resonances that affect the vocal folds in the form of feedback resulting in additional sub-harmonics. These facts indicate that the F0 could contain additional component waveforms. These fractional F0 sub-harmonics appear in the spectrum and modify the waveform values of the F0. Titze et al. (2016) note that the average F0 is predominantly determined by vocal fold length, and other factors include the freedom of movement of the laryngeal muscles that control elongation and collagen density or nonlinearity in tissue fiber tension. The vocal fold tissues consist of three main layers, epithelium, non-muscular lamina propria, and muscle. When the fibers of one layer are under tension, "the layer can be considered a 'thick string vibrating in a viscous soup.'" The string modes of vibration dominate over the gel modes of vibration and so the F0 is largely determined by the fiber component, but the combined properties of the gel and the fibers add to the total range of normal frequencies. Mannell (2007) notes that waves can interfere with themselves, that the supra-glottal vocal tract may attenuate some sound waves, and that waves can be reflected due to radiation impedance when the vocal tract opens into larger space.

Voice production depends on more than one underlying morphologic parameter including laryngeal framework mechanics, the depth of the vocal fold tissue layers, vocal fold boundary geometry and tissue fiber stress resulting in the F0 that is regulated through two distinct mechanisms (Titze et al. 2016).

Procedure

The CHEBYprime computer program written in Matlab was used to evaluate the vowel F0s and provide data on all the formants. Two hundred ten vowel sounds produced by male and female speakers were evaluated. CHEBYprime uses Chebfun routines (Trefethen 2000; Driscoll, Hale, & Trefethen 2014), a Chebyshev Transform (Boyd 2001), and Singular Value Decomposition procedure (Gold and Morgan 2001) to produce verifiable distinct formants, consistent with harmonic principles. The F0 was determined by the pitch determination algorithm, SHRP (Sun 2002), which is based on subharmonic to harmonic ratio.

Results

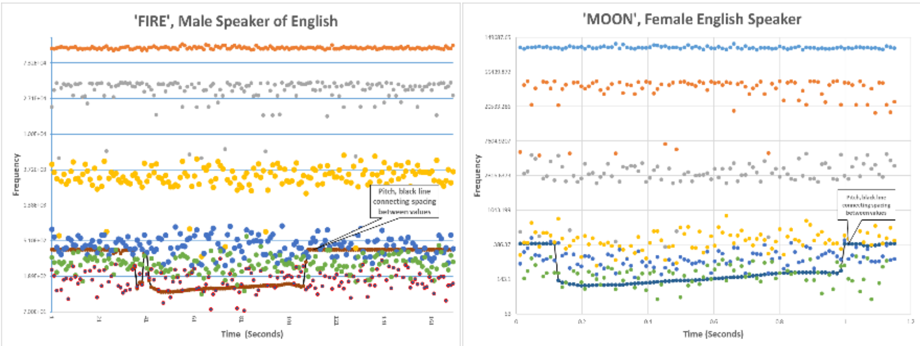

Figure 1a left, 1b right. Formant plots showing clear formant regions as individual points and the F0 linear plot occupying space in the lower formant region. Formants are close to each other, but do not overlap.

The Chebyshev algorithm was able to distinguish vowel formants into harmonic bands of formants, with the uppermost band being out of the hearing range. A Chebyshev transform is suited for speech data analysis because can precisely measure the modulation irregularities of speech data, and does not manipulate output data into the imaginary domain.

Conclusion

F0 was found to occupy the same frequency space as some of the lower formant bands. Data plots indicate that F0 values are a combination of the lowest of the formant frequency bands, as evidenced by formants occurring in the same frequency (Hertz) region as the F0. This complex F0 structure may arise from vocal folds, vocal fold inconsistencies, echo, and irregularities in the throat.

References

Boyd, J. 2001. Chebyshev and Fourier Spectral Methods. Mineola, NY: Dover.

Gold, B., Nelson M. 2000. Speech and Audio Signal Processing. NY: John Wiley & Sons, Inc.

Mannell, R. 2007. Standing Waves and Resonance. Macquarie University. www.ling.mq.edu.au/speech/acoustics/frequency.

Micheyl, C., Ryan, C., Oxenham, A. 2012. Further evidence that fundamental-frequency difference limens measure pitch discrimination. Journal of the Acoustical Society of America. 131(5) May 2012.

Pertusa, A., Iñesta, J. 2012. Efficient methods for joint estimation of multiple fundamental frequencies in music signals. EURASIP Journal on Advances in Signal Processing 2012. 2012:27.

Roark, R. 2006. Frequency and Voice: Perspectives in the Time Domain. Journal of Voice, Volume 20, Issue 3. Pp 325-354

Staudacher, M., Steixner, V., Griessner, A., Zierhofer, C. 2016. Fast fundamental frequency determination via adaptive autocorrelation. EURASIP Journal on Audio, Speech, and music Processing. 2016:17. DOI 10.1186/s13636-016-0095-8.

Sun, X. 2002. The Determination, Analysis, and Synthesis of fundamental Frequency. Dissertation: Northwestern University.

Titze I., Riede, T., Mau, T. 2016. Predicting Achievable Fundamental Frequency Ranges in Vocalization Across Species. PLoS Computational Biology 12 (6): e1004907. doi:10.1371/journal.pcbi.1004907

Trefethen, L. 2000. Spectral methods in Matlab. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Driscoll, T., Hale, N., Trefethen, L. (eds.). (2014). Chebfun Guide.

Tsanas, A., Zañartu, M., Little, M., McSharry, P. 2012. Robust fundamental frequency estimation in sustained vowels using ensembles. IEEE Transactions on Audio Speech Language Process.

Verde, L., De Pietro, G., Sannino, G. 2018. A methodology for voice classification based on the personalized fundamental frequency estimation. Biomedical Signal Processing and Control. Vol. 42(April), pp 134-144.