

EVALUATION OF F0 STABILITY IN SPEECH AND SINGING USING LARYNGEAL BIOIMPEDANCE MEASUREMENTS.

E Donati University of West London, London, England
JZ Tomaszewska University of West London, London, England
C Chousidis University of Surrey, Guildford, England

Abstract

Within human interaction, speech and singing represent essential and primordial forms of communication. The differentiation between these acts is hypothesised to be highly dependent on the stability of their fundamental frequency. Voice analysis is inherently dependent on the use of microphone which can make an accurate analysis of the fundamental frequency computationally dispendious and susceptible to noisy environments. In turn, current studies on singing–speaking distinction mainly rely on acoustical evaluation. The presented research seeks to provide a statistical analysis of the fundamental frequency variations in speech and singing by means of laryngeal bioimpedance measurements. These signals allow an accurate evaluation of the fundamental frequency whilst reducing the errors arising from noise and acoustic interferences. A dataset was created containing 2400 laryngeal bioimpedance measurements with a 50:50 ratio of speech and singing. An evaluation of the fundamental frequency variability was conducted using the YIN pitch estimation method on batches of 200 randomly selected samples (100 of speech and 100 of singing). A statistical evaluation was then conducted to compute the variability of the fundamental frequency across the batches, The results support the hypothesis for which fundamental frequency stability is highly impactful in the distinction of speech and singing with speech displaying a substantially higher variability across all batches.

1 INTRODUCTION

In human interaction, voice plays a pivotal role in day-to-day communication as well as being employed in other aspects of human society such as, for instance, arts. Whilst speech sits as the primary form of communication, voice is also vastly employed as a musical instrument through the act of singing. Despite the distinction between singing and speaking is perceptually clear for humans, the characterising elements of such distinction are not always clearly defined. Fundamental frequency (f_0) is regarded to be the main distinguishing factor between the two voice acts [1, 2]. In singing, the tuning of the pitch and the control over the duration of the phonation is dependent on the tension of the vocal folds which causes f_0 to remain more stable when compared to speaking voices [3]. Almost the entirety of the literature, however, analyses such classification more qualitatively through acoustic and perceptual evaluation rather than providing a mathematical and statistical evaluation [4]. In our previous research [5, 6] we proposed using laryngeal bioimpedance for f_0 extraction by measuring directly at the vocal folds level. In [7] we also demonstrated the effectiveness and efficiency of laryngeal bioimpedance in the classification of singing and speaking voices. The use of this type of signal in fact allows to achieve very high levels of classification accuracy whilst using much lighter and computationally non-taxing machine learning algorithms when compared to the methods commonly used in the literature for phonation classification. Indeed, given the functioning of the vocal apparatus, this technique allows to measure directly f_0 and deliver a more effective and efficient analysis of the differences between the two types of phonation. The experiment discussed in this paper uses laryngeal bioimpedance measurements to perform an evaluation of f_0 stability in singing and speech. This processs derive a statistical proof of the role of fundamental frequency in the distinction of speaking and singing voices. The rest of the paper will be organised as follows: chapter 2 will present a background of human phonation and laryngeal bioimpedance measurements, chapter 3 will discuss the methodology employed, chapter 4 will show the statistical results of the evaluation, and chapter 5 will present conclusions and discussion.

2 BACKGROUND ON PHONATION AND LARYNGEAL BIOIMPEDANCE

The functioning of laryngeal bioimpedance measurements and the information that this approach can deliver, is closely dependant on the anatomy and the behaviour of human phonatory system. Effectively, the phonatory apparatus acts as a converter of kinetic to acoustic energy. First, the lungs expand and contract to generate a steady airflow that is then passed through the trachea until it impacts the vocal folds within the larynx. When the airflow applies pressure on the vocal folds, a vibration is created, and the folds convert the flow of air into acoustic energy. The periodic oscillation of the vocal folds acts as the sound source of voice. The frequency of oscillation effectively represents the fundamental frequency of voice in any given phonation. Following the vocal folds oscillation, the generated acoustic wave passes through the vocal tract. Here, the elements of the vocal tract cause a series of resonances, through reflection and vibration, that define the distinguishable voice sound by adding harmonic content.

Using a pair of electrodes positioned across the neck, it is possible to measure the oscillation of the vocal folds bypassing the vocal tract and, in turn, any added harmonic content [8]. This is achieved by applying an alternating current through the larynx; the impedance changes of the larynx caused by the vibrational cycle of the vocal folds results in an amplitude modulation of the excitation signal [9]. This measurement technique delivers a very simple signal that shows the position alternation of the vocal folds and describes the fundamental frequency of voice. Figure 1 shows the comparison between a laryngeal bioimpedance measurement and its concurrent microphone recording. As shown in both time and frequency domain, the former delivers a much simpler signal that allows for fast and precise evaluation of the fundamental frequency.

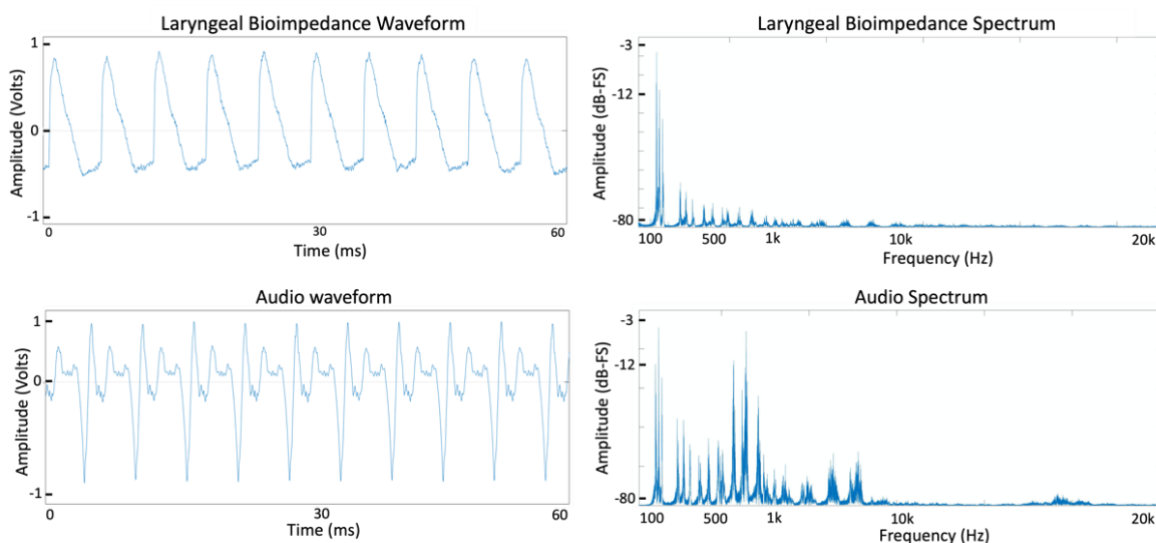


Figure 1 – Laryngeal Bioimpedance and respective Audio comparison

3 METHODOLOGY

The experiment presented in this paper uses recordings of laryngeal bioimpedance to evaluate the difference in stability between speech and singing. The analysis was conducted using a dataset of 2400 measurements with a 50:50 ratio of speech and singing. The dataset was developed for our previous research as discussed in [7]. The dataset contains recordings of laryngeal bioimpedance from a total of 12 participants each performing word utterances, sustained sung phonation, and brief note progressions. For this implementation, 200 samples, each with a duration of 200ms, were randomly selected from the dataset and divided into two batches of 100 singing samples and 100

speech samples. To evaluate the variability of f_0 , multiple pitch measurements were required across each sample to gauge how much the fundamental frequency changes through the duration of the sample. To measure f_0 values, the YIN pitch tracking algorithm [10] was used with windows of 1024 samples and a hop length of 512. This allows to perform multiple measurements across each sample and to use the resulting data to derive how much f_0 varies over time. As shown in equation 1, for 200ms samples, the YIN yields 19 f_0 values for each voice recording. Figure 2 shows a graphical representation of the windowed YIN algorithm.

Sampling Frequency: 44.1kHz
 Sample duration: 200ms
 Window Length: 1024
 Hop Length: 512

$$\frac{\text{duration in samples}}{\text{window} - \text{hop}} + 1 = \frac{44.1\text{kHz} \cdot 200\text{ms}}{1024 - 512} + 1 \cong 19 \quad (1)$$

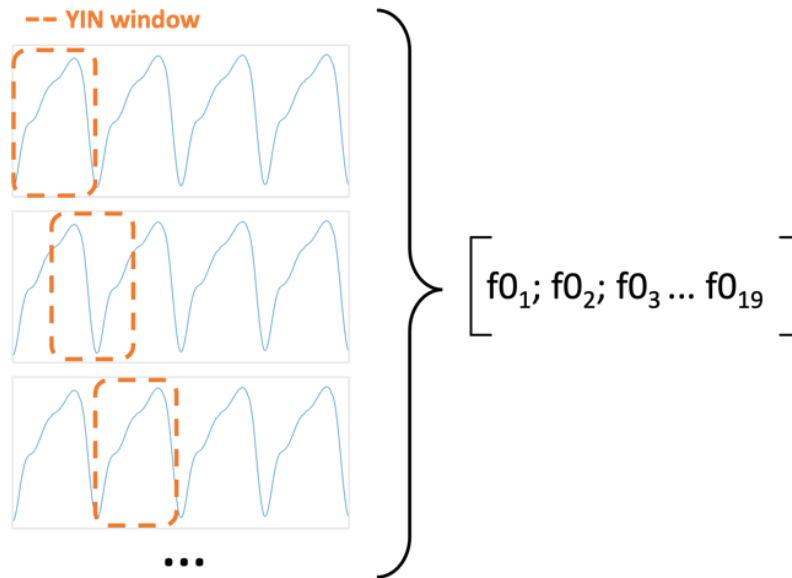


Figure 2 – Windowed YIN

For each resulting f_0 array, a statistical analysis was performed to extract the variance and the coefficient of variation (CV) of f_0 for each sample in the batches. Both measures were used to evaluate the variability of f_0 as the main distinguishing element between speaking and singing voices. Once computed for each sample, the average values of variance and CV were calculated for each class (singing and speech). The coefficient of variation was calculated as shown in equation 2.

Standard deviation: σ
 Mean: μ
 Coefficient of Variation: CV

$$\frac{\sigma}{\mu} = CV \quad (2)$$

4 RESULTS

The results of the experiment were analysed in terms of f_0 values for each sample, and in terms of variance and CV for each class. The first analysis was conducted by plotting the values of f_0 for each sample in a box plot and observing the distribution of the fundamental frequency values. Figures 3 and 4 show the box plots for the first 25 samples of each class. The observation conducted on the box plots demonstrate a much higher variability of f_0 in speech samples when compared to singing samples. The box plots were limited to 25 entries for improved visibility.

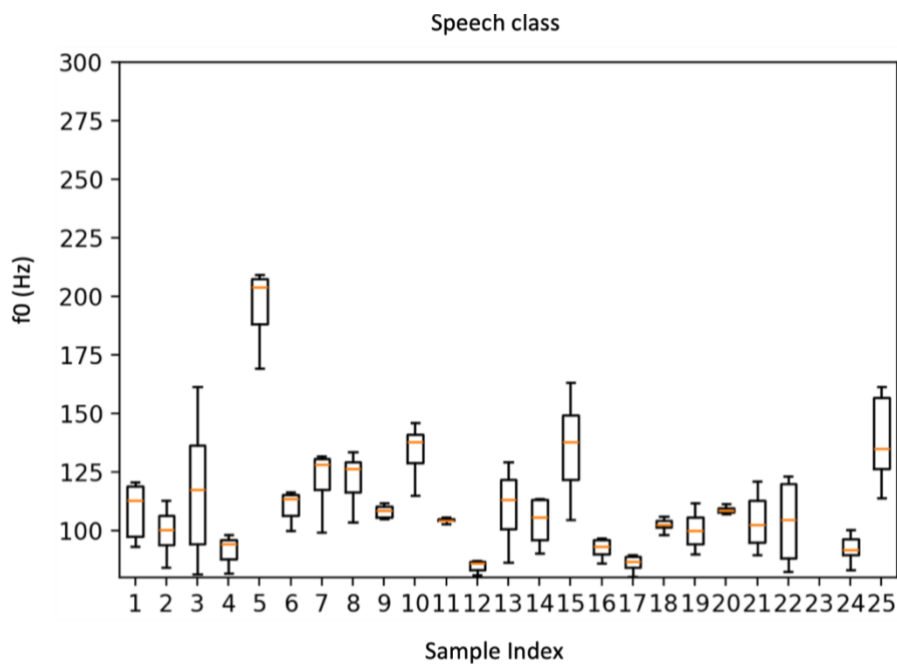


Figure 3 – Fundamental Frequency distribution per sample in speech

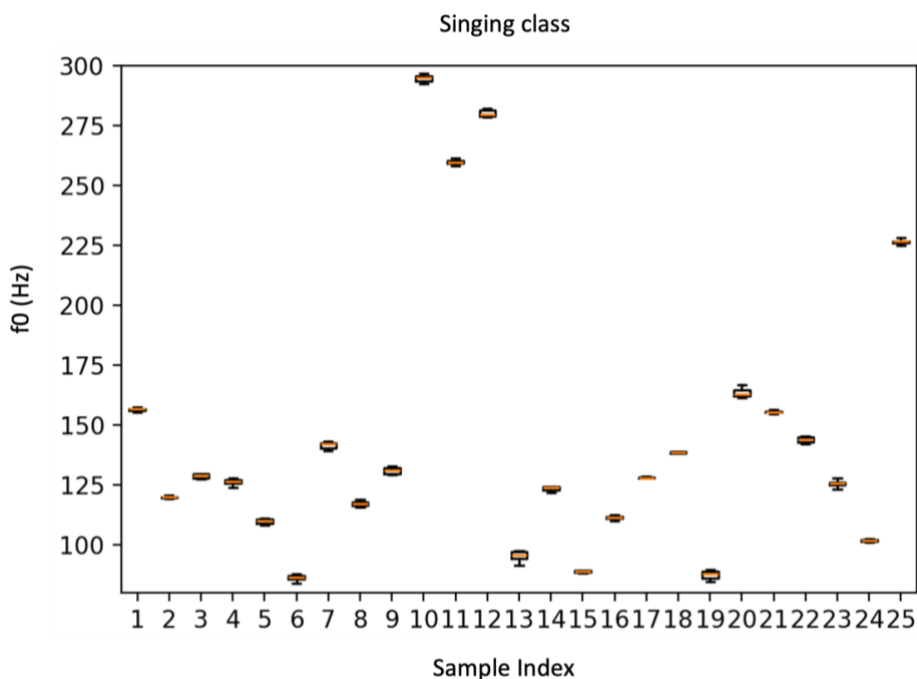


Figure 4 – Fundamental Frequency distribution per sample in singing

The second stage of the analysis involved the computing of the average variance and mean CV across each class. Both variance and CV were calculated for individual samples and the average was subsequently calculated across each batch. The observation of the results matches the results observed in the box plots, demonstrating a much higher variance and CV in speech and, in turn, a higher f0 stability in singing. Table 1 shows the results obtained for the full dataset.

Table 1 – Mean variance and mean CV for speech and singing

Class	Mean Variance	Mean Coefficient of Variation
Speech	4.45	1.6%
Singing	2600.57	16.0%

5 CONCLUSIONS

This paper presented an experiment for the evaluation of f0 variability as the main characteristic in the distinction of speaking and singing voices. This was approached by using laryngeal bioimpedance measurements which allows to analyse voice directly at the vocal folds level. The analysis was performed on a dataset of 200 samples divided with equal ratio into two classes, speech and singing. The f0 of the samples was measured using of a windowed YIN algorithm to derive multiple f0 values throughout the duration of the individual samples. The resulting f0 arrays were then analysed visually through boxplots, and statistically interpreted by computing the mean variance and the coefficient of variation (CV) across each class. The results of the experiment demonstrate a much higher f0 variability in speech then in singing displaying a CV ten times higher in speech. The proposed experiment strongly supports the assumption for which the stability of the fundamental frequency in voice represents a key characteristic in the distinction of speech and singing. The results, moreover, support the use of laryngeal bioimpedance as an efficient and effective method for the analysis of the fundamental frequency of voice which can be employed in several areas such as human-machine interaction, education, music technologies, and medical diagnosis. Beside contributing to the general knowledge of vocal characteristics, the presented research represents a contribution towards efficient systems for the real-time distinction of speaking and singing voices. The results suggests that by using bioimpedance, simple algorithmic approaches can be implemented to achieve fast and efficient voice classification systems through the f0 stability data.

6 REFERENCES

1. Fujisaki, H.I.R.O.Y.A., 1981. Dynamic characteristics of voice fundamental frequency in speech and singing. acoustical analysis and physiological interpretations. Dept. for Speech, Music and Hearing, Tech. Rep, pp.43-47.
2. de Medeiros, B.R., Cabral, J.P., Meireles, A.R. and Baceti, A.A., (2021). A comparative study of fundamental frequency stability between speech and singing. Speech Communication, 128, pp.15-2.
3. Vijayan, K., Li, H. and Toda, T. (2018) 'Speech-to-singing voice conversion: The challenges and strategies for improving vocal conversion processes', IEEE Signal Processing Magazine, 36(1), pp. 95-102
4. de Medeiros, B.R. and Cabral, J.P., 2018, June. Acoustic distinctions between speech and singing: Is singing acoustically more stable than speech?. In Proc. 9th Int. Conf. Speech Prosody (pp. 542-546).
5. Donati, E. and Chousidis, C. (2022). Electroglottography based real-time voice-to-MIDI controller. Neuroscience Informatics, p.100041.

6. Donati E., Chousidis C. (2022). Electroglottography based voice-to-MIDI real-time converter with AI voice act classification. 17th IEEE Medical Measurement & Application.
7. Donati, E., Chousidis, C., Ribeiro, H.D.M. and Russo, N., 2023. Classification of Speaking and Singing Voices Using Bioimpedance Measurements and Deep Learning. *Journal of Voice*.
8. Drugman, T., Bozkurt, B. and Dutoit, T., (2012). A comparative study of glottal source estimation techniques. *Computer Speech & Language*, pp.20.
9. Fabre, P. (1959). La glottographie electrique en haute frequence, particularites de l'appareillage. *comptes rendus des seances de la societe de biologie et de ses filiales*, 153(8-9), pp.1361-1364. [Publication in French].
10. De Cheveigné, A. and Kawahara, H., 2002. YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4), pp.1917-1930.