Frequency analyses of human voice using fast Fourier transform

Jinan F. Mahdi

Department of Medical instruments, College of Engineering Electrical and Electronic Engineering Techniques, Middle Technical University

E-mail: jinan f 2008@yahoo.com

Abstract

Quantitative analysis of human voice has been subject of interest and the subject gained momentum when human voice was identified as a modality for human authentication and identification. The main organ responsible for production of sound is larynx and the structure of larynx along with its physical properties and modes of vibration determine the nature and quality of sound produced. There has been lot of work from the point of view of fundamental frequency of sound and its characteristics. With the introduction of additional applications of human voice interest grew in other characteristics of sound and possibility of extracting useful features from human voice. We conducted a study using Fast Fourier Transform (FFT) technique to analyze human voice to identify different frequencies present in the voice with their relative proportion while pronouncing selected words like numbers. Details of findings are presented

Key words

Human voice, larynx, fundamental frequency, Fast Fourier Transform, Frequency spectrum.

Article info.
Received: Jun. 2015
Accepted: Sep. 2015
Published: Sep. 2015

التحليل الترددي للصوت البشري بأستخدام تحويل فورير

جنان فاضل مهدي

قسم الاجهزة الطبية، كلية التقنيات الهندسية الكهربائية و الالكترونية، الجامعة التقنية الوسطى

الخلاصة

التحليل الكمي للصوت البشري موضوع مهم واكتسب اهمية عندما تم تحديد الصوت البشري كنموذج لمعرفة صوت الإنسان وهويتة. الهيئة الرئيسية المسؤولة عن إنتاج الصوت هي الحنجرة وهيكل الحنجرة جنبا إلى جنب مع خصائصه الفيزيائية ووسائط الاهتزاز يتم تحديد طبيعة ونوعية الصوت المنتج. كان هناك الكثير من العمل من وجهة نظر تردد الصوت الاساسي وخصائصه. مع إدخال تطبيقات إضافية لصوت الإنسان نمت خصائص اخرى للصوت وإمكن استخراج الخصائص المفيدة من الصوت البشري. أجرينا دراسة باستخدام تحويل تقنية فوريير السريع Fast Fourier Transform لتحليل الصوت البشري لتحديد ترددات مختلفة موجودة في الصوت مع بنسبة نسبيا في حين لفظ الكلمات المختارة مثل أرقام. وسيتم في هذا البحث عرض تفاصيل النتائج.

Introduction

The acoustic structure of human voice contains a wealth of information related to the speaker's identity and state of emotional which can easily be retrieved with accuracy [1–3]. Human voice related study in its initial phases concentrated mainly on the quality of sound from the point of view of audible quality and phonetic significance [2–4]. Deeper investigations led to quantitative analysis of voice and it gained momentum when human voice was identified as a tool for human authentication and identification that resulted in brisk activity in this area[4]. The main organ responsible for sound production is larynx and the structure of larynx along with its physical properties and modes of vibration determine the nature and quality of sound produced. There has been lot of work from the point of view of

fundamental frequency of sound. Minoru Hirano et al.[5] showed that three laryngeal muscles and their activity governs intensity, fundamental frequency and Phonation of sound. Role of frequency distribution [6] is investigation on sound produced by birds is reported by birds by Mohammad Moaviyah Moghal and interesting results are presented in relation to syrinxp[7].

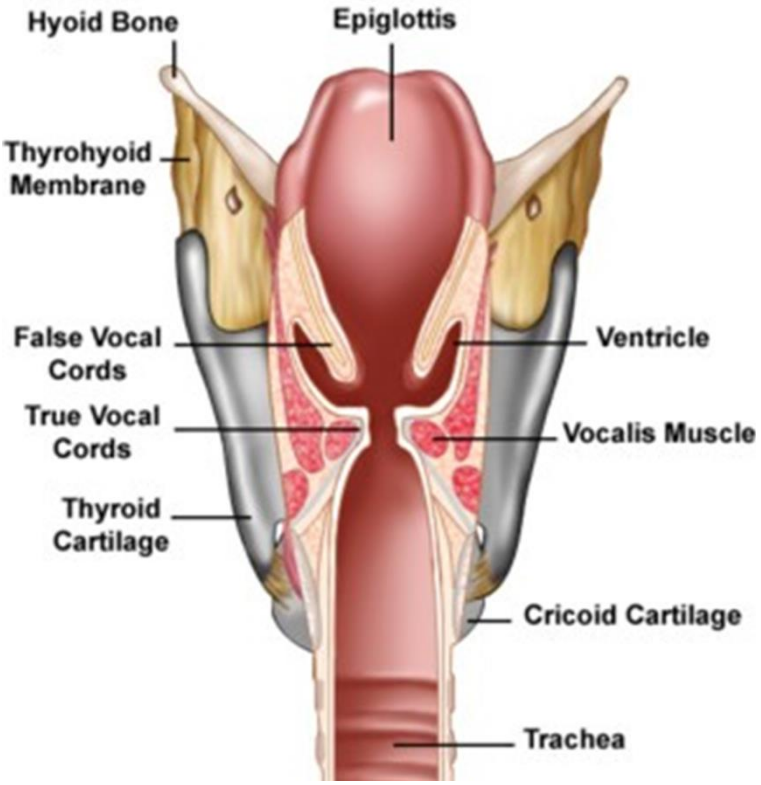

Fig. 1: Cavity of the Larynx showing true and false vocal cords.

Most significant changes in the voice, during childhood, are due to the rapid growth of the larynx, the vocal folds and the surrounding support structures. At birth, the membranous length of the vocal folds i.e. the part that actually vibrates is around 2 mm both for males and females [8]. For infants the voice is with a higher fundamental frequency F_0 due to this shorter length and smaller size of the vibrating components. With age the size, structure and local properties conge that result in lower frequencies. If Lm is the membranous length, Lc is the cartilaginous length then the total vocal fold length is L = Lm + Lc [8].

During initial stages the growth rate of length of vocal fold for males is approximately 0.7 mm per year and for females it is 0.4 mm per year. As the size of vocal fold increases faster in males, the voice tend to be lower frequency with lower values of F_0 as compared to females where the growth of vocal fold is very slow. For grownups the maximum length is about 16 mm for men and 10 mm for women. The relation between the length of vocal fold and the fundamental

$$F_0 = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}}$$

where F_0 is the fundamental frequency, L is the vocal fold length, σ the longitudinal stress and ρ is the tissue density. This relationship can be used to estimate the vocal fold length L from F_0.

With the introduction of additional applications of human voice interest grew in other characteristics of sound and possibility of extracting useful features from human voice[9]. We conducted a study using Fast Fourier Transform (FFT) technique using mathematical software MathCAD to analyze human voice to identify different frequencies present in the voice with their relative proportion while pronouncing selected words; we used numbers from one to ten. Selected subjects were asked to pronounce these numbers one by one under normal relaxed condition and recorded their voices for analysis. Details of findings are presented.

Methodology

Two males 10 and 12 year of age were selected two females with ages 6 and 9 years and one female 35 year of age. The subjects were asked to record their voice under normal relaxed conditions. The voice was recorded using standard audio recording equipment's with flat frequency response over the entire audible frequency range. The sampling rate of the digital audio recording device was 44.1kHz, 2 channels at 16 bits per channel or 2×16=32 bits per sample. As the same voice is recorded in the two channels, we used mean of the two channels as the actual data for analysis.

The sound files were converted to wave format at the same sampling rate and the files were read in MathCAD. It is the sampling rate and the Nyquist criterion that provides the frequency steps at which the amplitudes are found on the implementation of the Fast Fourier Transform (FFT). The data read from the voice files in the wave format was subjected to FFT in MathCAD. These returns the Fourier transform of the input data which is amplitude of the sound at different frequencies. The amplitudes so found are complex quantities with both real and imaginary part. The sound power at different frequencies can be found from the square of the amplitude as A·A*, where A* is the complex conjugate of A. Typical screenshot of MathCad – 14 software while processing one of the sound segments is shown in Fig. 2

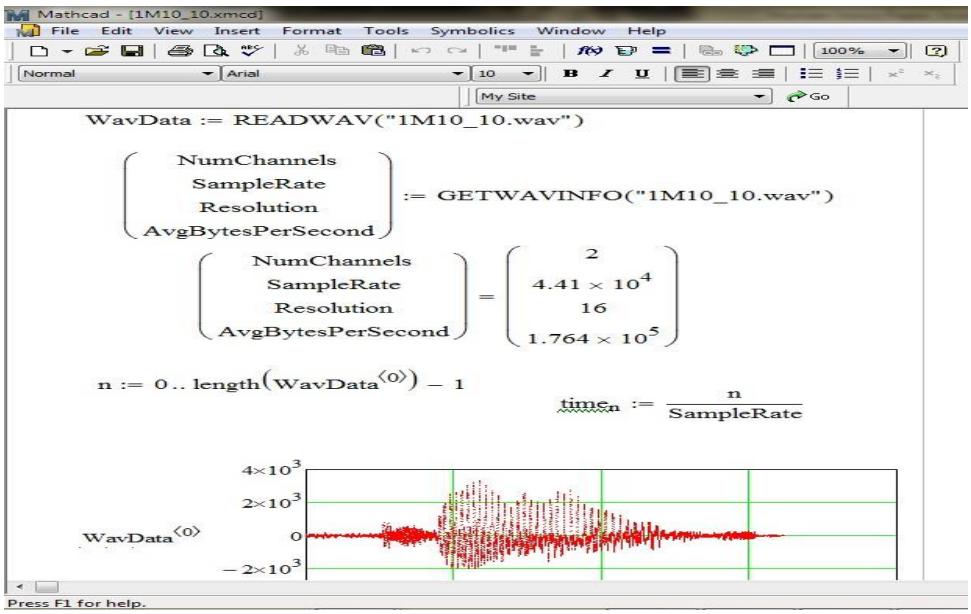

Fig. 2: A typical screenshot of the software Math Cad – 14 while processing a sound segment to find the FFT.

Experimental work

A typical sound recording is shown in Fig. 3 the wave file contains intensities of sound recorded at different points of time (the sampling points). The figure shows the sound amplitude plot as a function of time when different numbers are pronounced in English i.e. one, two three and so on. Ten different bunches of sound amplitude recorded while pronouncing those numbers are shown with labeling.

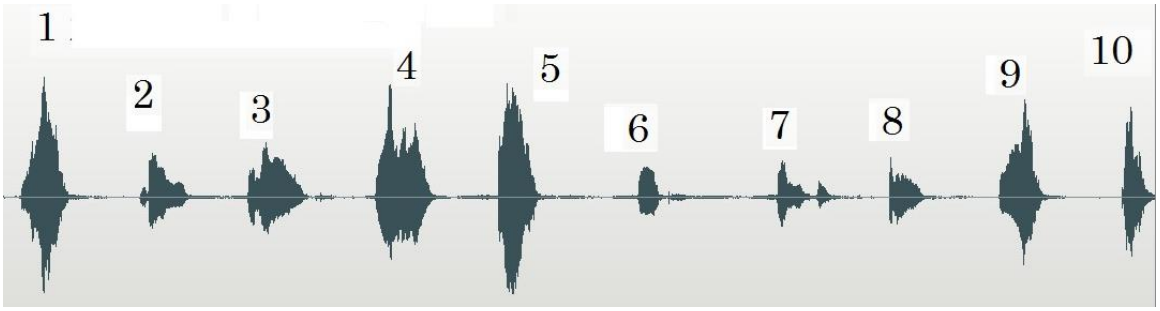

Fig. 3: A typical sound recording while pronouncing one, two, three etc.

The sound segments for different numbers (from one to ten) are separated using audio editor software and all the ten segments were separately saved in files for further processing. A typical segment for word 'three' is shown in Fig. 4.

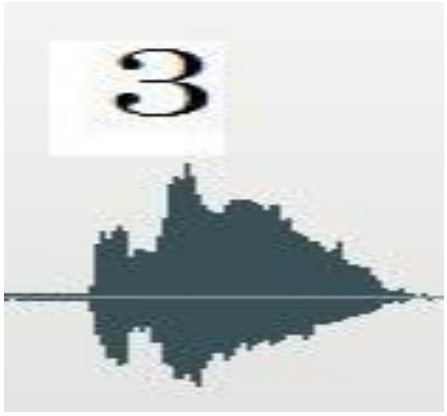

Fig. 4: Typical segment for word 'three'.

After collecting all the ten segment of sound, they were subjected to FFT and the resulting transform i.e. the amplitude and time data was saved in a text file for further processing and plotting. It was observed the main fundamental frequency of different subjects is different and while pronouncing different words, a certain pattern is produced. Almost entire sound was limited to a frequency range up to 200 Hz, beyond which very limited or no appreciable sound was present. A comparison of same word pronounced by four different subjects is shown in Fig. 5.

The amplitude frequency spectra shown in Fig. 5 are for two male subjects with age 12 and 10 years shown in A and B respectively and for two females with age of 9 and 6 years shown in C and D respectively. Typically the voiced of adult male has a fundamental frequency from 85 to 180 Hz, and that for a typical female adult is from 165 to 255 Hz [9,10].

Table 1 shows the fundamental frequency F_0 and other prominent frequencies like F1, F2 etc that happen to be the harmonics having frequencies as integral multiple of F_0 and combination thereof. M10, M12, F9 and F6 are Male subjects with age of 12 and 10 and female subjects with age of 9 and 6 years respectively.

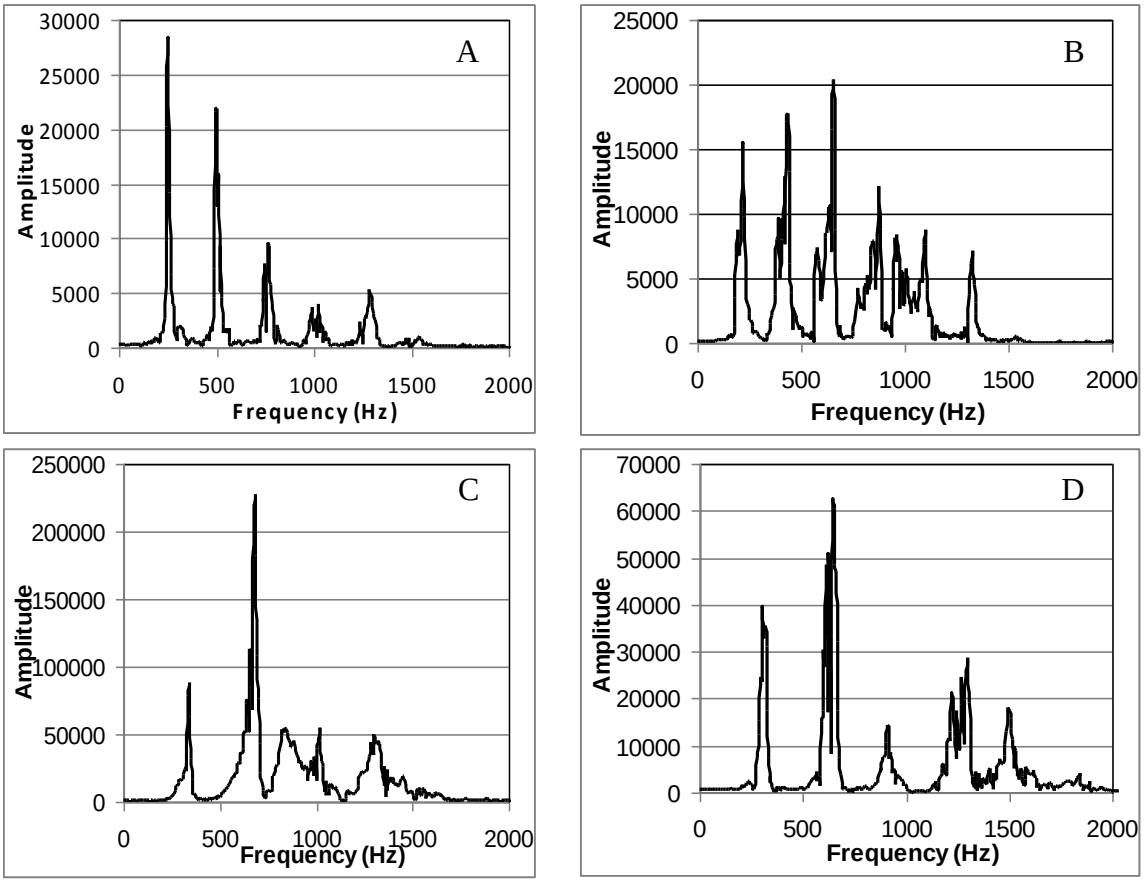

Fig. 5: Amplitude versus frequency plot for four voice samples from M12, M10, F9 and F6 (A, B, C, D) respectively while pronouncing the word 'Four'.

Table 1: Prominent frequencies produced by four subjects.

Frequency	Prominent Frequencies of Subjects			
	M12	M10	F9	F6
F0	253	215	339	306
F1	501	431	683	614
F2	769	640	1017	915
F3	1022	872	1330	1248
F4	1275	1050	–	1502

The fundamental frequency for the male with age 12 is 253 Hz and that for male with age of 10 is 215 which are slightly higher than those with the age and fundamental frequencies discussed above. Reason being that the ranges of frequencies mention are for adults and the subjects in the present case are children from lower age group. This indicates that the fundamental frequency for children is on the higher side because of the fact that the larynx of children is smaller than that of adults and natural frequencies of smaller objects is higher. Same thing applies to the two female subjects listed in Table 1 and shown in Fig. 5.

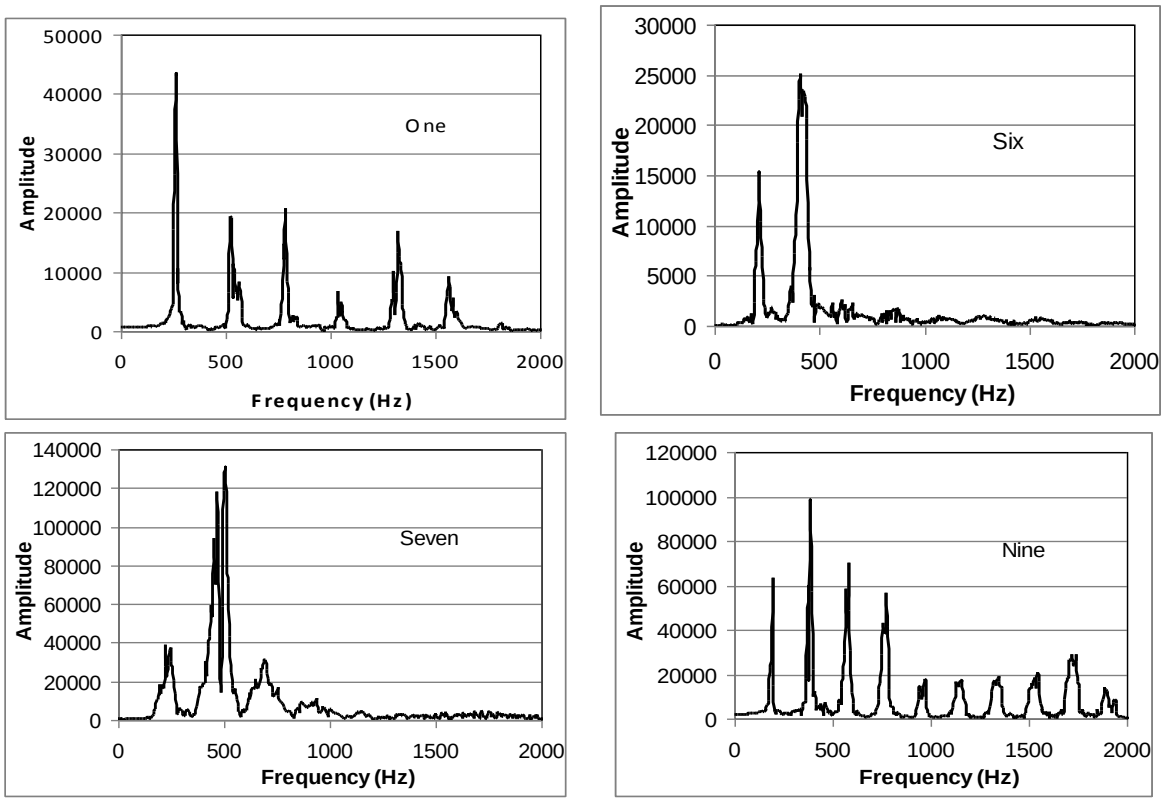

Fig. 6: Amplitude frequency spectra for voice samples from M10 while pronouncing the four words 'One', 'Six', 'Seven' and 'Nine'.

The number of peaks present in Fig.5 A, B, C and D are different because of the fact that the way of pronouncing differs from individual to individual and the components of the laryngeal cavity participating could differ. The Amplitude frequency spectrum presented in Fig. 6 is for the same person while pronouncing different numbers, the characteristic frequency and its harmonics are distinctly seen, the patterns are drastically different and correspond to the sound produced.

For the purpose of comparison of the amplitude frequency spectrum of voice pronouncing different words Fig.5 presents four selected spectra for Male subject with age of 10 years. The four spectra very much differ from each other as they represent a different sound. In the first plot there are six peaks where as in the second one there are only two prominent peaks followed by some wavy pattern. Third one (for 'Seven') has one middle broad peak and two relatively smaller peaks followed by some background. It is to be noted that the middle tall peak is not the fundamental one, the fundamental one is the short first peak having a frequency of 215 Hz, and is in accordance with the earlier plot and Table 1. The last one corresponding to 'Nine' shows about ten peaks that are harmonics of the fundamental as the pronunciation results from several overtones of the fundamental frequency F_0. Careful inspection of plots for 'Six' and 'Seven' brings out the fact that all the peaks are not pure fundamental or their harmonics but in fact are a superimposition of several components from modes of vibration of larynx.

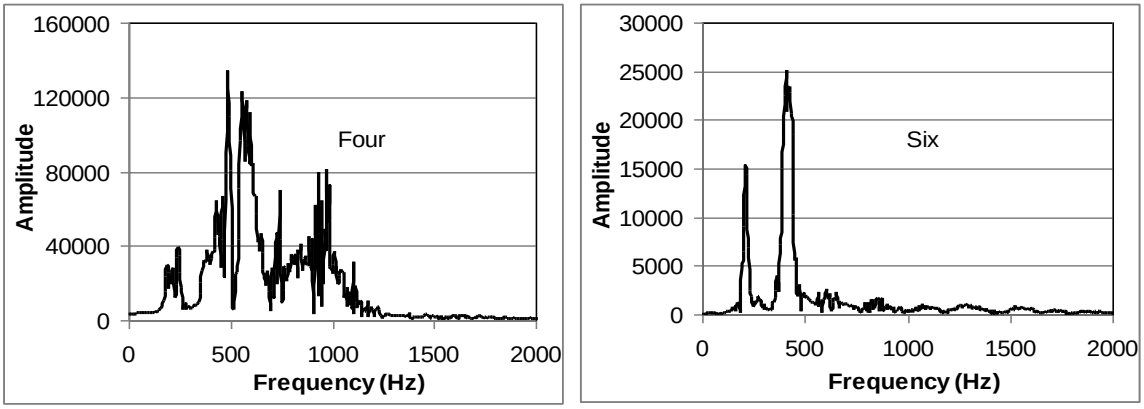

Fig. 7: Two amplitude frequency spectra for voice of female with age of 30 year while pronouncing the two words 'Four' and 'Six'.

Fig. 7 is the amplitude frequency spectrum for a female subject with age of 30 years and presented for the purpose of demonstrating that in certain voice types many of the frequencies and their superimposition is present as is seen from the amplitude frequency spectrum for 'four'. Practically it is difficult to figure out the fundamental frequency from the plot. The fundamental frequency is clearly seen in the next plot for 'Six'. The plot is very much identical for the plot for 'Six' shown in Fig. 6 for male subject having 10 years of age. The major difference comes from the fundamental frequency and the harmonics present and their composition.

Results and discussion

This work demonstrated that the FFT is successfully useful in finding out the fundamental frequency of voice and obtaining the amplitude frequency spectrum of human voice from which power spectrum can be obtained. It is also seen that children from lower age group have higher values of fundamental frequency F_0 as compared to adults. Also it is seen that with age the fundamental frequency F_0 decreases as size of larynx grows with age that in turn results in reduction of the natural resonant frequency. It is also shown that the voice is composed of fundamental frequencies and superimposition of its harmonics. As extension of this work we plan to design technique for estimation of the size of active components of larynx using the existing information and estimation of size using fundamental frequency.

Conclusions

It is demonstrated that the technique is capable of identifying sound produced by different people and while pronouncing different words. It is also shown that the dependence of sound characteristics and fundamental frequency F_0 based on age and sex can also be quantified. Such type of study in humans is presented for the first time and lot more work can be done to systematically correlate these features. It is useful if estimating the larynx characteristics like size and structural properties, similar work has been done by others in relation to complexity of syrinx[7,12]. Comparison of the frequencies shows that for birds the sizes being small, the frequencies are much on the higher side as compared to humans.

References

[1] G.Papcun, Kreiman, J. and Davis, A. J. Acoust. Soc. Am. 85 (1989) 913–925.

[2] D. G. Doehring, and Bartholomeus, B. N. Neuropsychologia, 9 (1971) 425 - 430.
[3] W. A.Van Dommelen, Acoustic Speech 33, 3 (1990) 259 – 272.
[4] Pascal Belin, Nature, 403 (2000) 309-312.
[5] Minoru Hirano, John Ohala, William Vennard, Journal of Speech, Language, and Hearing Research, 12 (September 1969) 616-628.
[6] S. Manikandan, J Pharmacol Pharmacother, Jan-Mar, 2, 1 (2011) 54–56.
[7] Mohammad Moaviyah Moghal, Vidya S. Pradhan A. R. Khan Mazahar Farooqui, Journal of Medicinal Chemistry and Drug Discovery, Special issue, National Convention (January 2015) 583 – 595.
[8] H. Brad Story, Eric A. Hoffman, Ingo R. Titze, NCVS Status and Progress Report - I I (May 1997) 153-161.
[9] D. R. Van Lancker, Canter, G. J. Brain Cogn. 1 (1982) 185– 195.
[10] I.R. Titze, Principles of Voice Production, Prentice Hall (currently published by NCVS.org), pp. 188 (1994)
[11] R. J. Baken, Clinical Measurement of Speech and Voice. London: Taylor and Francis Ltd. pp. 177, (1987).
[12] Moavia Moghal, Vidya S Pradhan, A. R. Khan and Mazhar Farooqui, 4, 6 (2015) 2486 – 2495.
[13] M. Kob and C.N Euschaefer-Rube, 3rd International workshop on Model and Analysis of Vocal Emissions for Biomedical Applications, 187 (2003) 187.
[14] R. Foresman, Bryant, "Acoustical Measurement of the Human Vocal Tract Quantifying Speech and Throat-Singing". Ph.D Thesis Pomona Senio (2008).