

# Exploring Gender Classification from Speech Signal using Java-DSP Tool

Dr. G. Sasi

Professor, CVR College of Engineering/ECE Department, Hyderabad, India.  
Email: gsasikumar75@gmail.com

**Abstract:** More than a century, speech and audio processing play a crucial role in the entertainment and telecommunication systems. Due to the high demands, expecting more sophisticated features, still audio and speech processing require advanced techniques to serve better. This article focuses mainly on the gender classification from speech signals based on pitch period and pitch frequency analysis. To achieve the required tasks, the online graphical simulation, such as Java-DSP, is used here. J-DSP environment blocks can be dragged and placed in the editor easily, which helps to reduce the time required to complete the design issues. Also, the pitch period estimation is done using autocorrelation and cepstrum methods. The zero-crossing detection is used to focus on the identification of the voiced/unvoiced portion of the speech. The survey results from the participants who attended the training program on speech and audio processing course were obtained as feedback.

**Index Terms:** Autocorrelation, gender classification, J-DSP, on-line DSP simulator, pitch estimation.

## I. INTRODUCTION

The eminent faculty and students developed Java-DSP (J-DSP) at Arizona State University to teach the Digital Signal Processing course and related subjects through Internet-based labs and simulations. Students who are using J-DSP Tools do get benefit to understand, analyze, and realize the DSP concepts via online lab exercises [1–4]. Internet-based simulations, drag-and-drop facility, interactive approach, outreach functions support, interface to DSP hardware, and iPhone/iPad compatibility, basic and advanced functions of DSP are the major key features of the J-DSP tool [5–7].

This article mainly focuses on the usage of J-DSP tool on speech signals, especially with the male and female voice. In the last century and even today, audio signal and speech processing play a vital role in the telecommunication and entertainment systems. The main techniques behind audio and speech processing are as follows [8,9]:

- Quantization
- Entropy coding
- Transform coding
- Filter banks and transforms
- Subband coding
- Linear prediction
- Time-frequency analysis
- Lossless coding

Fig. 1 shows a few primary techniques and methods used in the audio and signal processing domain. Also, the multi-rate

digital signal processing applications are depicted in Fig. 2. Fig. 3 shows a simple J-DSP module on plotting Fast Fourier Transform (FFT) results for the given speech signal.

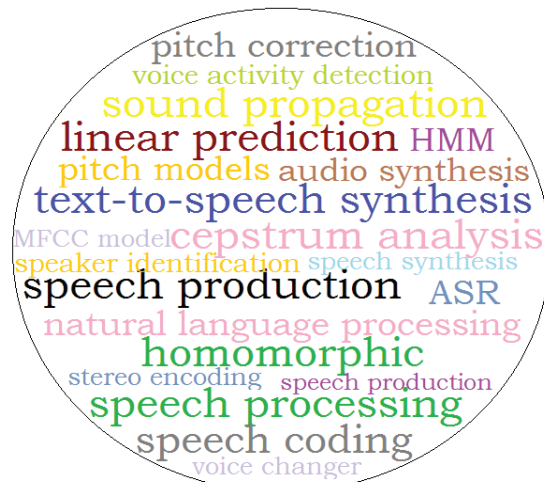


Figure 1. Techniques used in Speech and Audio Processing.

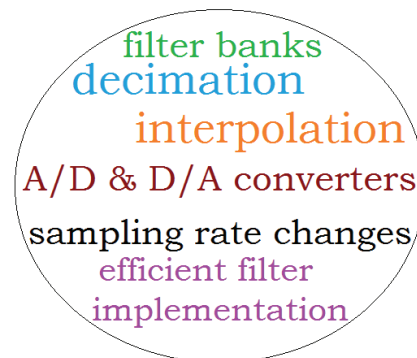


Figure 2. Applications of Multi-rate Digital Signal Processing.

The essential applications related to audio and speech processing, such as gender classification, speech/music classification, voiced/unvoiced classification, text-to-speech synthesis, audio hacking, and MP3 algorithms, can be worked out easily from students' perspectives [10].

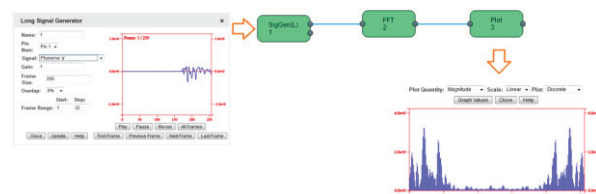


Figure 3. The FFT plot of a speech signal.

Speech is a comfortable medium or communication among people. Obviously, speech is the most powerful and easiest mode to communicate not only among human beings, also among human-animals and human-machines. A speech signal gives lots of information such as emotional status, gender, accent, dialect, etc [11].

Due to the different sizes of vocal cord, gender type, emotional status and dialect, the properties of each speaker vary enormously. Naturally, human beings identify the speaker from speech signals. But, for the automated applications, machines need to recognize the speech or a speaker, which would be the demanding task in modern computer technology. For the past decades, data mining and machine learning models have mostly used gender identification using voice signals. Features of the voice such as the vocal fold length, energy, formant and gait can be used for the machine learning models [12].

This article is organized as follows. Section 2 contains the materials and methods, which include pitch analysis on male and female speakers using both time and frequency domains. Part 3 includes an appraisal result of 30 students who participated in a two-day workshop on speech and audio processing course. Lastly, the conclusions are described in Section 4.

## II. MATERIALS AND METHODS

### A. Classification based on Pitch frequency

In common, the pitch frequency of the female speaker is much higher (almost double) than the male speaker. The vocal tract length is higher for males (approx. 17.5 cm), smaller for females (approx. 13 cm), and much smaller for children (approx. 9 cm). Hence, the pitch frequency of children is much higher than females and males. The variations among different speakers occur due to the age, gender, larynx, and vocal tract properties (physical), articulation habits, and the rate of speaking [13,14]

Fig. 4 shows the male speech signal and a voiced portion of the male speech. The sampling frequency of the speech signal is 8 kHz. The spectral analysis is done here on the 10th frame, which contains 256 samples. Hence, a 256-point Discrete Fourier Transform (DFT) is applied to the frame to view the spectral contents (see Fig. 5). As from Fig. 5, since there are four samples in between the spectral peaks, the peak-to-peak frequency resolution is calculated as follows:

$$\frac{4F_s}{N} = \frac{4 \times 8000}{256} = 125 \text{ Hz} \quad (1)$$

The result confirms gender as a male speaker. Similarly, the female speech signal and a voiced portion of the female speech are shown in Fig. 6. The spectral analysis using DFT and the results are shown in Fig. 7. As from Fig. 7, since there are eight samples in between the spectral peaks, the peak-to-peak frequency resolution is calculated as follows:

$$\frac{8F_s}{N} = \frac{8 \times 8000}{256} = 250 \text{ Hz} \quad (2)$$

The result confirms gender as a female speaker.

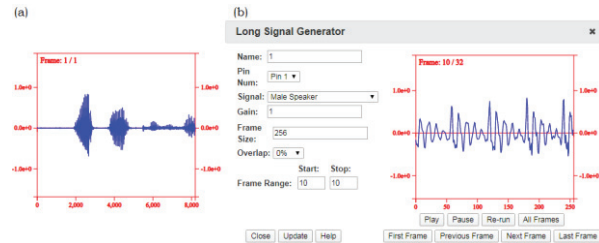


Figure 4. J-DSP environment blocks: (a) male speech signal; (b) the voiced portion of the male speech.

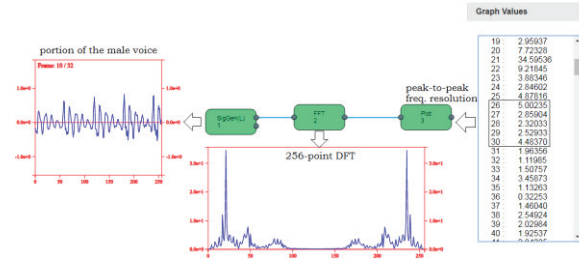


Figure 5. Spectrum analysis on the portion of the male voice.

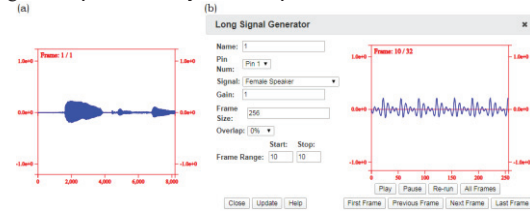


Figure 6. J-DSP environment blocks: (a) female speech signal; (b) the voiced portion of the female speech.

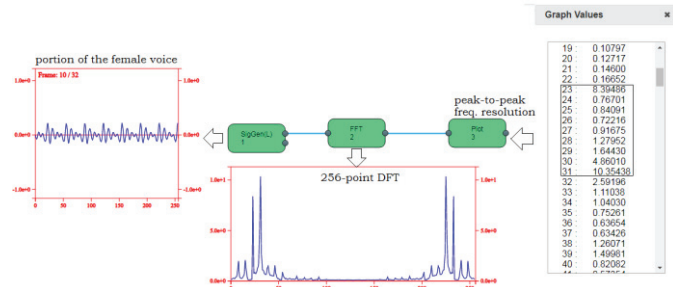


Figure 7. Spectrum analysis on the portion of the female voice.

### B. Classification based on Pitch period

Also, gender classification can be done using a pitch period. Fig. 8 shows the portion of the male speech signal, which contains the peaks in between 58 to 128 (total 62 of samples). Hence, the pitch period can be calculated as,

$$\frac{62 \text{ samples}}{8000 \text{ samples per second}} = 7.7 \text{ ms} \quad (3)$$

In general, the pitch period range for a male speaker is 4 to 20 ms, and for a female speaker is 2 to 8 ms. The above result confirms that the speaker is a male speaker. Similarly, Fig. 9 shows the portion of the female speech signal, which contains the peaks in between 55 to 88 (total 33 samples). Hence, the pitch period can be calculated as,

$$\frac{33 \text{ samples}}{8000 \text{ samples per second}} = 4.1 \text{ ms} \quad (4)$$

The above result confirms that the speaker as a female speaker.

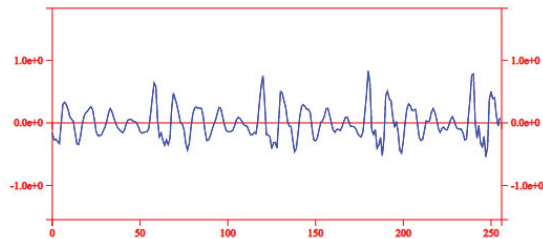


Figure 8. Pitch period analysis on a male speaker.

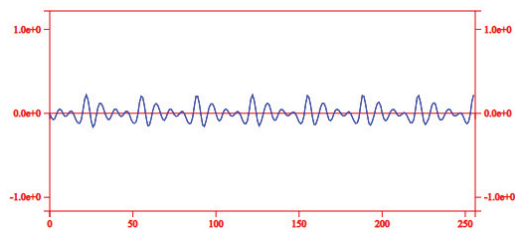


Figure 9. Pitch period analysis on a female speaker.

### C. Pitch period using autocorrelation

Also, the autocorrelation technique can be used to determine the pitch period of a voice. Fig. 10 shows the autocorrelation results of a male speaker (lag with 61 samples), and Fig. 11 shows the results of the autocorrelation of a female speaker (lag with 33 samples). These results match the results obtained using pitch estimation using time-domain analysis for both the male and female speakers.

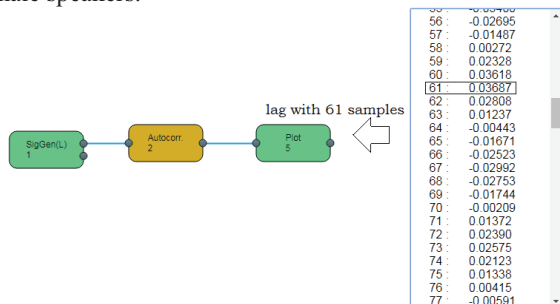


Figure 10. Pitch period estimation using autocorrelation on a male speaker.

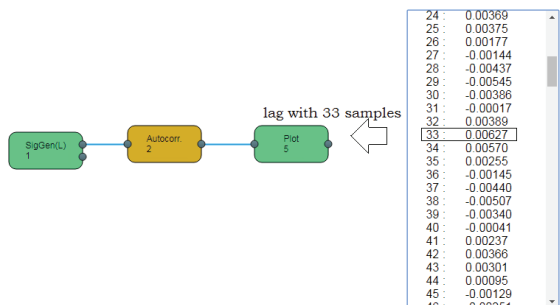


Figure 11. Pitch period estimation by autocorrelation on a female speaker.

### D. Other Experiments

During the two-day workshop on speech and audio processing course, the participants did a few other experiments which are as follows,

- Gender classification using Matlab

- Voice activity detection
- Pitch correction
- Voice detection using zero-crossing rate detection
- Pitch estimation using Cepstrum analysis

## III. RESULTS AND DISCUSSION

The thirty students who participated in a two-day workshop on speech and audio processing course were surveyed to identify the required time (minutes) per task (see Fig. 12). Since the J-DSP tool works fully on the drag-and-drop facility, almost all tasks require less time (except the cepstrum analysis). Also, the students were inspected to specify their interest in which exercise they experimented. The survey had the participants' judge on a scale of 1 to 5, where one was "Boring", two was "Not Interested", three was "Interested", four was "Very Good", and five was "Highly Impressive". Fig. 13 shows the average score (out of 30 scholars) for each task. On average, most of them were inspired by exploring digital filters (systems) using the J-DSP tool.

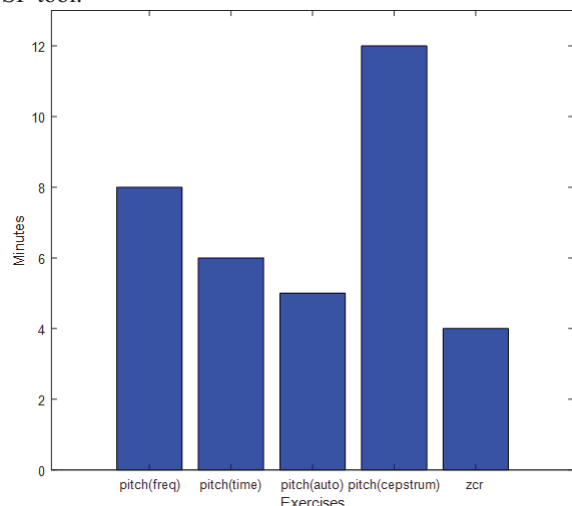


Figure 12. Results on time required per task.

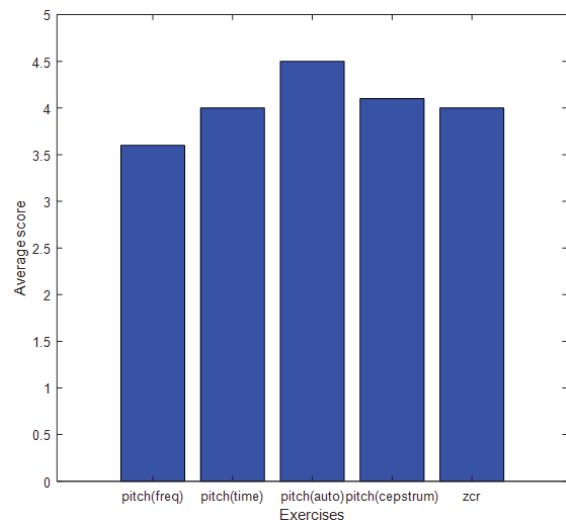


Figure 13. Survey results.

#### IV. CONCLUSIONS

The online based simulation software tools are much handy in teaching subjects such as linear systems, signals and systems, digital signal and image processing, audio and speech processing, control systems, etc. This sort of pedagogical method is comfortable for better analyzing the theoretical concepts with hands-on demonstrations. One such easy tool is J-DSP, the user-friendly tool, is used here to analyze speech signals, especially male and female speaker voices. Moreover, the pitch period is estimated using time domain, frequency domain, autocorrelation, and cepstrum methods. Also, the zero-crossing rate is used to identify the voiced/unvoiced portion of the speech. Conflicts of Interest: The authors declare no conflict of interest.

#### REFERENCES

- [1] Abhinav Dixit, Sameeksha Katoch, Photini Spanias, Mahesh Banavar, Huan Song, and Andreas Spanias, "Development of Signal Processing Online Labs using HTML5 and Mobile platforms", Proceedings of the IEEE Frontiers in Education Conference (FIE), Indianapolis, IN, USA, pp.1–5,2017.
- [2] G.Sasi, P. Thanapal, R.K. Arvind Shriram, V.S. Balaji, V. and V. Elamaran, "Exploring digital signal processing concepts using on-line graphical DSP simulator", International Journal of Advanced Science and Technology, 29: 410–421.
- [3] Andreas Spanias, Ted Painter, and Venkatraman Atti, "Audio Signal Processing and Coding". Wiley India,2013.
- [4] Suhas Ranganath, Jayaraman J. Thiagarajan, Deepta Rajan, Mahesh K. Banavar, Andreas Spanias, Jie Fan, Kristen Jaskie, and Cihan Tepedelenlioglu, "Interactive Signal Processing Education Applications for the Android Platform," Computers in Education Journal, 10: 1–10, 2019. .
- [5] W. Wang, G. Zhang, L. Yang, V.S. Balaji, V. Elamaran, N.Arunkumar, 2019. Revisiting signal processing with spectrogram analysis on EEG, ECG and speech signals. Future Generation Computer Systems, 98: 227–232.
- [6] A. Dixit, U.S. Shanthamallu, A. Spanias, V. Berisha, and M. Banavar, "Online Machine Learning Experiments in HTML5," Proceedings of the IEEE Frontiers in Education (FIE), San Jose, California, pp. 1–5, October 2018.
- [7] Uday S. Shanthamallu, SunilRao, Abhinav Dixit, Vivek S. Narayanaswamy, Jie Fan, and Andreas Spanias, "Introducing Machine Learning in Undergraduate DSP Classes," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, pp. 1–5, May 2019.
- [8] V. Elamaran, A. Aswini, V. Niraimathi, and D. Kokilavani, "FPGA implementation of audio enhancement using adaptive LMS filters", Journal of Artificial Intelligence, 5,pp. 221–226,2012.
- [9] Alison Behrman, "Speech and Voice Science", Third Edition, Plural Publishers,2017.
- [10] B. Sindhura, S. Ashwin, G. Rajkumar, V. Elamaran, and M. Sankar, "Useful Tips and Tricks on Digital Data Processing with Discrete Fourier Transform", Proceedings of the 2nd International Conference on Computing Methodologies and Communication (ICCMC 2018), Erode, India, pp. 738–742, 2018.
- [11] Manish Gupta, Shambhu Shankar Bharti, and Suneeta Agarwal, "Gender-based speaker recognition from speech signals using GMM model," Modern Physics Letters B, 33: 1–23, 2019.
- [12] Ioannis E. Livieris, Emmanuel Pintelas, and Panagiotis Pintelas, "Gender recognition by voice using an improved self-labeled algorithm," Machine Learning & Knowledge Extraction, 1: 492–503, 2019.
- [13] E. Abdulhay, V. Elamaran, M. Chandrasekar, V.S. Balaji, and K. Narasimhan, "Automated diagnosis of epilepsy from EEG signals using ensemble learning approach", Pattern Recognition Letters, 28,pp. 5247–5251, 2017.
- [14] V. Elamaran, A. Praveen, M.S. Reddy, L.V. Aditya, and K.Suman, "FPGA Implementation of Spatial Image Filters using Xilinx System Generator", Proceedings of the International Conference on Modeling Optimization and Computing (ICMOC-12), Procedia Engineering, 38,pp.2244 – 2249, 2012.
- [15] R. Naveena, V. Dorthy Rabecka, G. Rajkumar, and V. Elamaran, "Understanding Digital Filters from Theory to Practice using Matlab and Simulink", International Journal of Pharmacy and Technology, 7,pp.9923–9934,2015.
- [16] V. Elamaran and G. Rajkumar, "FPGA implementation of point processes using Xilinx System Generator". Journal of Theoretical and Applied Information Technology, 41,pp. 201–206,2012
- [17] K. Narasimhan, V. Elamaran, S. Kumar, K. Sharma, and P.R. Abhisek, "Comparison of satellite image enhancement techniques in wavelet domain". Research Journal of Applied Sciences, Engineering, and Technology, 4,pp. 5492 – 5496, 2012.