International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Automatic Gender Identification Based on Voice Recognition

Rakesh Mali

Department of Computer Science, Rani Channamma University, Belagavi, Karnataka, India

A B S T R A C T

Gender detection based on speech signal is one of the most crucial challenges in the context of content-based multimedia indexing. This study explores many classifiers as well as a set of auditory and pitch features to address the gender identification problem. We demonstrate that the combined performance of attributes and classifiers outperforms that of each classifier working independently. We developed a system for gender identification in multimedia applications based on these findings. Numerous neural networks containing pitch- and sound-related properties are used in the system.90% classification accuracy is attained for 1-second segments for male voice, irrespective of language or sound channel. It has been demonstrated that utilizing several experts as opposed to only one expert and voice continuity may raise classification accuracy to 93%. When applied to a portion of the Switchboard database at 5-second intervals, the categorization accuracy is close to 98.5%, for female voice segment.

Keywords: Support Vector Machines (SVM), content-based audio indexing, identification, numerous neural networks

Introduction

The gender of a speaker may be automatically determined in a number of situations. Gender-dependent models are more accurate than gender-independent ones for automated speech recognition. Thus, using a gender dependent model requires gender identification [1-2], by limiting the search area to sounds of the same gender, accurate gender identification can increase the recognition accuracy of speakers. The speaker's gender is a cue used in the annotation for content-based multimedia indexing. Marston [3] and Potamitis et al. [4] assert that gender-dependent speech coders are also more accurate than gender-neutral ones. Software for gender detection can therefore be useful in multimodal signal processing systems.

Literature Survey

All Gender categorization is used in contemporary multimedia information retrieval systems for a variety of possible purposes, including voice recognition, speaker diarization, intelligent human-computer interaction, biometric social robots, audio or video content indexing, etc. [5, 6]. Additionally helpful in some circumstances is automatic gender recognition in mobile healthcare Page | 2 program to put it another way, there are some disorders, such vocal fold cyst. One of the most important issues in the developing world is gender identity, which is also a rapidly increasing computerized environment [7]. Gender information is used to normalize voice characteristics and reduce word errors in speech recognition. The speaker's gender identity is frequently important for more realistic and customized conversation systems. Comparing speech signals in order to create a gender classifier \ based on auto-correlation approach for pitch analysis of voice sounds [8]. Gender classifications based on power spectrum estimates have emerged [9], with an average recognition accuracy of 80%. On discriminative weight training, a support vector machine (SVM) based gender identification is used. To represent the differences between male and female speech, a unique gender categorization system based on GMM has been developed [10–12]. When evaluating the gender classification system in clean speech conditions, gender classification accuracy was 98%, and it remained at 95% under most loud speech situations. Experimental comparison of the different classifiers led to the identification of the best classifier for gender and age categorization when voice signals were processed [13]. The method for automatically determining gender in a brief passage of continuously uttered speech [14].

Proposed Methodology

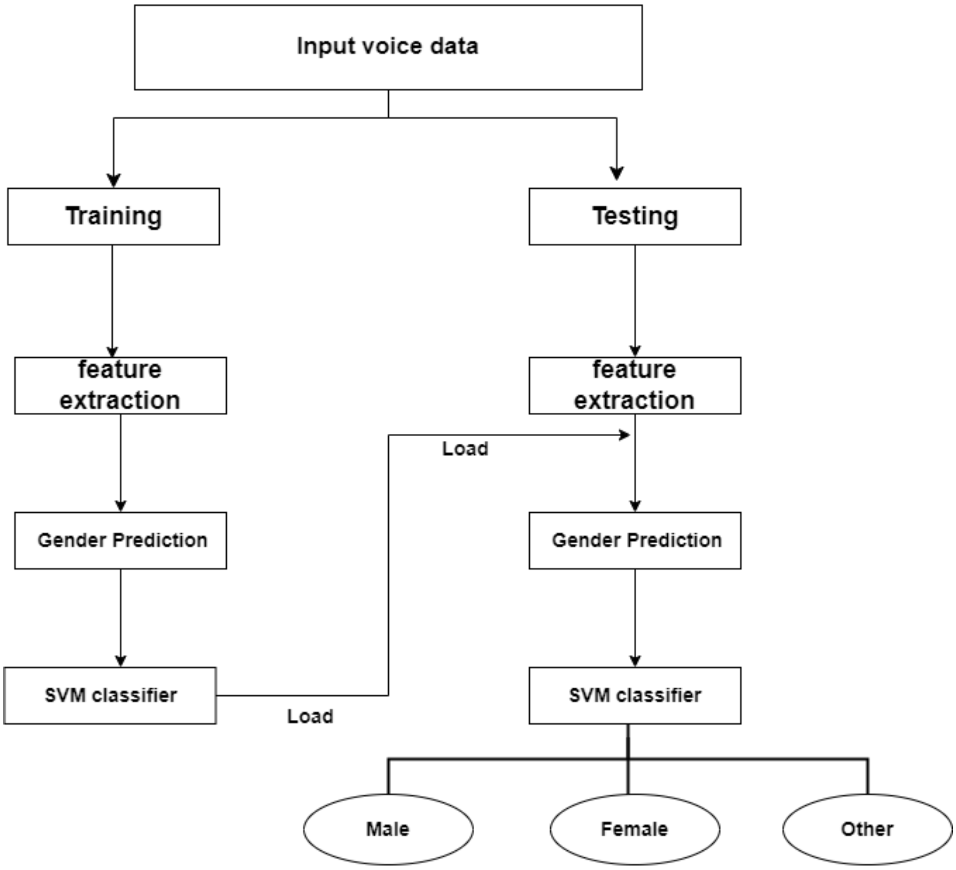

Figure 1. General architecture of the proposed work

The suggested process for speech recognition-based automatic gender identification entails the following steps:

- Data collection: Compile a variety of voice samples from people of various genders. To accommodate for variances and guarantee inclusion, the dataset should include a broad variety of voices, dialects, languages, and speech patterns.
- Pre-processing: consists of cleaning and removing noise from the speech data, adjusting audio levels, and extracting pertinent acoustic properties. Pitch, formant frequencies, measurements of the voice's quality, and length are often utilized characteristics for gender identification.
- Feature Extraction: Take the pre-processed voice samples and extract the acoustic characteristics. In order to analyse the voice signals and determine the necessary qualities, this stage entails using signal processing techniques and algorithms.
- Feature Selection/Dimensionality Reduction: If the feature set has a high degree of dimension, dimensionality reduction methods such as Principal Component Analysis (PCA) or feature selection techniques can be used to condense the feature space while preserving the most discriminative data.
- Model Training: To train a gender categorization model, use a machine learning or statistical modelling technique. Support Vector Machines (SVM), Gaussian Mixture Models (GMM) are examples of frequently used approaches. The gender labels are linked to the relevant auditory characteristics in the labelled dataset, which is used to train the algorithm.
- Model Evaluation: Evaluate the trained model's performance using evaluation measures including accuracy, precision, recall, and F1-score. A second validation or test dataset is often used to evaluate the model's performance in order to determine its generalizability.
- Fine-tuning and Optimization: To boost performance, tweak the model's hyper parameters and optimize it. In order to improve the accuracy of gender recognition, this stage involves iteratively modifying the model settings, feature selection, or pre-processing methods.
- Validation & Testing: Verify the robustness and generalizability of the trained model using an independent dataset. Different voice samples from the training and validation phases should be utilized for testing.

- Bias Analysis and Mitigation: Analyse the model's predictions carefully for any potential biases. Examine the model for any gender-based or demographic biases, and then take action to reduce or eliminate them.
- Deployment and Integration: After the model exhibits acceptable performance, incorporate it into the intended application or system. Make sure the implementation is done correctly and take into account things like computing efficiency, the need for real-time processing, and scalability.

Experimental results and discussion

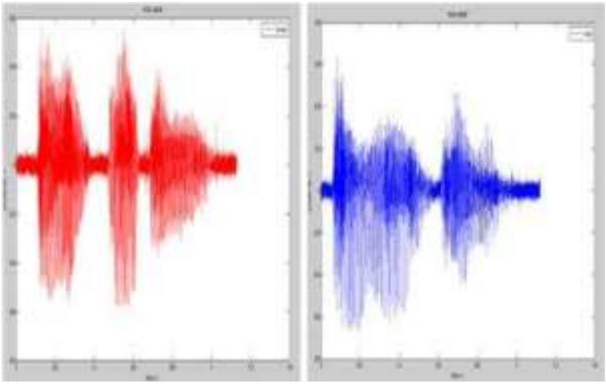

Fig. 1 Domain graph of male and female voice samples

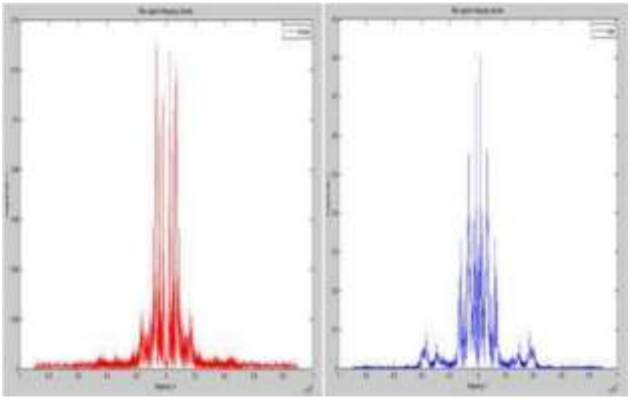

Fig. 2 Domain graph of male and female voice samples After applying Butterworth low-pass filter

Time domain graphs of a male and female voice sample are shown in Figure 1.and Figure 2 shows the frequency domain plots of a sample of male and female voice after a filter has been applied.

To improve analysis and remove noise, time domain data were transformed to frequency domain. Applying a low-pass filter—named, Butterworth low-pass filter is used to remove the noise.

In order to show how the pitch fluctuates throughout the voice signal, pitch analysis was done, pertinent parameters were collected, and a graph of pitch contour vs. time frame was also presented. It was determined what the voice signal's average pitch was.

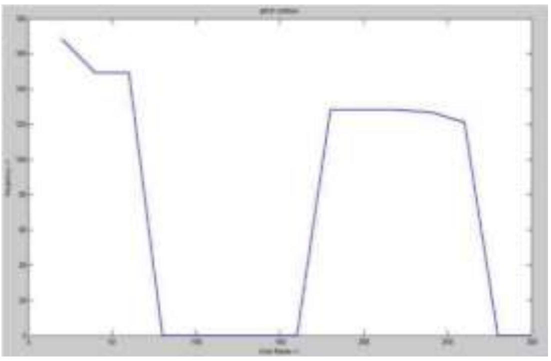

Fig. 3(a): Pitch contour plot of a male voice sample.

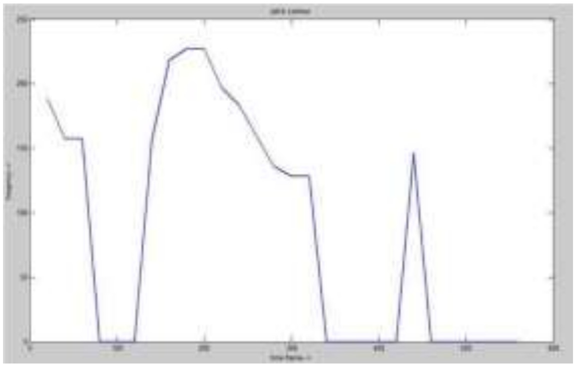

Fig. 3(b): Pitch contour plot of a male voice sample

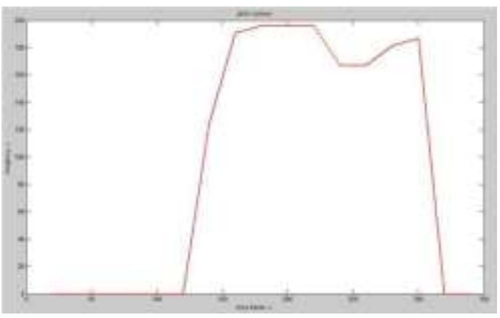

Fig. 4(a): Pitch contour plot of a female voice sample.

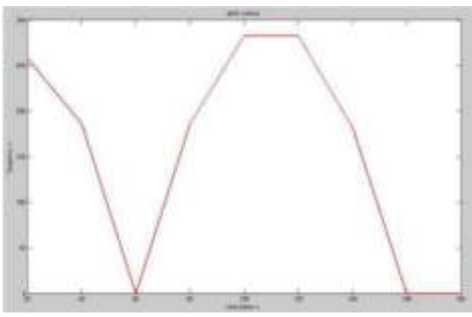

Fig. 4(b): Pitch contour plot of a female voice sample

It is evident from the aforementioned facts that female pitch contour has a greater average magnitude than male pitch counter.

The average pitch of each signal was determined and summarized for better examination. Table 1 shows male and female voice pitch comparison.

Table 1. Male and female voice sample pitch comparison

Female voice	**Average pitch (in Hz)**	**Male voice**	**Average pitch (in Hz)**
Female 1	235.8848	Male 1	153.6777
Female 2	271.6006	Male 2	185.2170
Female 3	311.1007	Male 3	191.9196
Female 4	258.8448	Male 4	190.7082
Female 5	179.9638	Male 5	136.3465
Female 6	230.6140	Male 6	171.2671
Female 7	251.0028	Male 7	143.6738
Female 8	283.8033	Male 8	145.9776
Female 9	301.8447	Male 9	200.9593
Female 10	240.7985	Male 10	179.1027

Table 2: The classification accuracy of the proposed method using SVM classifier

Classification Method	**Samples**	**Segment**	**Classification result Accuracy %**
SVM	Male	1 sec	91.2%
	Male	1 sec	92.6%
	Female	1 sec	94.9%
	Female	1 sec	93.5%

Table 3: The classification accuracy of the proposed method using SVM classifier

Classification Method	**Samples**	**Segment**	**Classification result Accuracy %**
SVM	Male	5 sec	90.2%
	Male	5 sec	93.6%
	Female	5 sec	95.9%
	Female	5 sec	98.5%

Table 2 and Table 3 both represent the classification accuracy of the proposed method utilizing an SVM classifier Due to their distinct numbers, these tables most likely contain data from various studies, datasets, or parameter settings. These tables' accuracy numbers serve as a gauge of how well the model performed in accurately classifying data points.

As shown in Figure 5(a), choose the "Audio for Training" option to start the execution procedure. In this phase, a male voice must be entered, and the system correctly recognizes it as a man, as seen in Figure 5(b). Input a female voice after that, and the system will recognize it as such as seen in Figure 5(c). Once constructed, this model may be used to analyses all user input. Following that, the module will try to identify the exact person's voice using the trained model. Following speech recognition, the outcome will be indicated as "Male" or "Female." Users have two choices: they may select a prerecorded audio file by selecting "Select File" or they can click the "Record" button to record an audio file in real-time. The system will output the expected gender result after the process is complete.

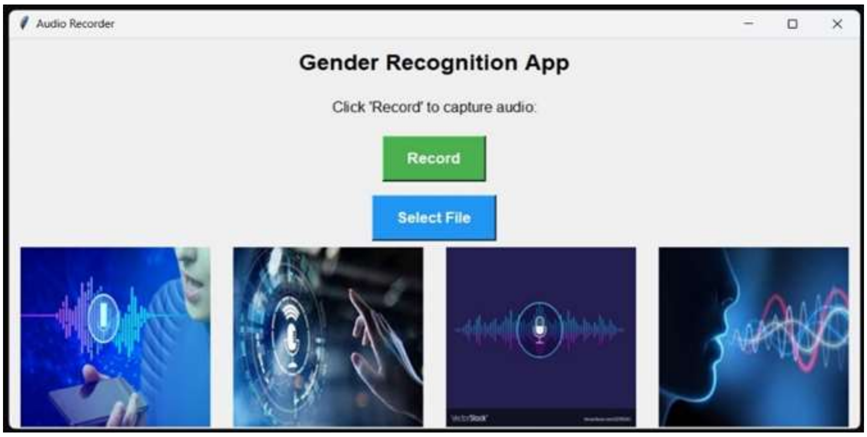

Fig 5(a) choose the Audio for Training

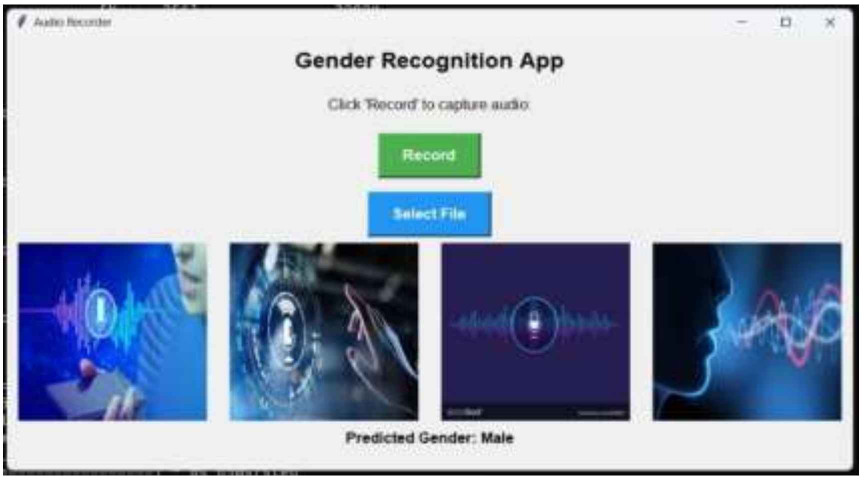

Fig 5(b) system correctly recognizes it as a Male

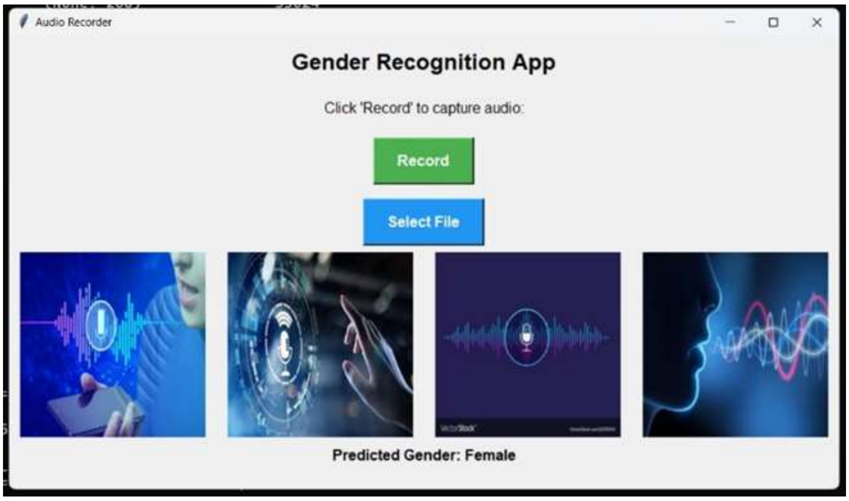

Fig 5(c) system correctly recognizes it as a Female

Conclusion

The goal of speech recognition-based automatic gender identification is to correctly identify a person's gender from their voice. It may be advantageous for forensics, accessibility, marketing, customer service, and user interfaces. Important considerations include correcting biases, protecting privacy, and ethical issues. The present technique involves gathering data, pre-processing, feature extraction, training models, evaluation, optimization and, deployment. The key to successful tool implementation is continuous improvement. Responsible usage requires striking a balance between advantages and moral issues. 90% classification accuracy is attained for 1-second segments for male voice, irrespective of language or sound channel. When applied to a portion of the Switchboard database at 5-second intervals, the categorization accuracy is close to 98.5%, for female voice segment.

Acknowledgements

I am grateful to Dr. Parashuram Bannigidad, Chairman Department of Computer Science, Rani Channamma University, Belagavi for his valuable guidance for completion of this work

References

1) A Acero and X Huang Speaker and gender normalization for continuous-density hidden markov models, 09-09 May 1996

2) C Neti and S Roukos Towards a universal speech recognizer for multiple languages, 17-17 December 1997

3) AL Marston polar localization of the mind protein of Becillus subtilis and its role in selection of the mid-cell division site, September 4, 1998

4) Potamitis and N Fakotakis, gender-dependent and speaker-dependent speech enhancement, 13-17 May 2002

5) H. Harb and L. Chen, Voice-based gender identification in multimedia applications, Journal of Intelligent Information Systems, 24(2), 179-198 (2005)

6) D. Ververidis and C. Kotropoulos. Automatic speech classification to five emotional states based on gender information. In Proc. XII European Signal Processing Conf., volume 1, pages 341– 344. Vienna, Austria,(2004).

7) Musaed Alhussein, Zulfiqar Ali, Muhammad Imran,and Wadood Abdul, Automatic Gender Detection Based on Characteristics of Vocal Folds for Mobile Healthcare System, Mobile Information Systems, Article ID 7805217, 1–12 (2016).

8) Bhagyalaxmi Jena and Beda Prakash Panigrahi, Gender Classification by Pitch Analysis,Electronics & Telecommunication Dept., Silicon Institute of Technology, 1(1), 2012.

9) Md. Sadek Ali1, Md. Shariful Isla1 and Md. AlamgirHossain, Gender recognition of speech signal, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), 2(1), 1–9 (2012).

10) S.-I. Kang and J.-H. Chang, Discriminative weight trainingbased optimally weighted MFCC for gender identification, IEICE Electronics Express 6(19), 1374–1379 (2009).

11) Y. Hu, D. Wu, and A. Nucci, Pitch-based gender identification withtwo-stage classification, Security and Communication Networks, 5(2), 211–225 (2012).

12) S. Gaikwad, B. Gawali, and S.C. Mehrotra, Gender identification using SVM with combination of MFCC, Advances in Computational Research, 4(1), 69-73, 2012.

13) M. Sedaaghi, A Comparative Study of Gender and Age Classification in Speech Signals, Iranian Journal of Electrical & Electronic Engineering, 5(1), 1–12 (2009).

14) M. Sigmund, Gender Distinction using Short Segments of Speech Signal, International Journal of Computer Science and Network Security, 8(10), 159–162 (2008).

15) M. C. Whitton, “Making virtual environments compelling,” Communications of the ACM, vol. 46, no. 7, pp. 40-46, July 2003.

16) D. Ververidis and C. Kotropoulos, “A State of the Art Review on Emotional Speech Databases”, in Proc. 1st Richmedia Conference, Lausanne, Switzerland, pp. 109-119, October 2003.

17) S. McGilloway, R. Cowie, E. Douglas-Cowie et al., “Approaching automatic recognition of emotion from voice: A rough benchmark”, in Proc. ISCA Workshop Speech and Emotion, pp. 207-212, Newcastle, 2000.

18) P. Loizou, “Colea: A MATLAB software-tool for Speech Analysis”, University of Arkansas, May 2003, http://www.utdallas.edu/ loizou/speech/colea.htm

19) S. Malcolm, “Auditory toolbox in MATLAB”, version 2, University of Purdue, http://rvl4.ecn.purdue.edu/ malcolm/interval/1998- 010/*236*, 54.