

## INFORMATION TECHNOLOGY FOR GENDER RECOGNITION BY VOICE

**Diana Koshtura**

Lviv Polytechnic National University,  
Information Systems and Networks Department, Lviv, Ukraine  
E-mail: [diana.koshtura.mnsam.2021@lpnu.ua](mailto:diana.koshtura.mnsam.2021@lpnu.ua), ORCID: 0000-0002-5665-5423

© Koshtura D. A., 2023

**Gender recognition from voice is a challenging problem in speech processing. This task involves extracting meaningful features from speech signals and classifying them into male or female categories. In this article, was implemented a gender recognition system using Python programming. I first recorded voice samples from both male and female speakers and extracted Mel-frequency cepstral coefficients (MFCC) as features. Then trained, a Support Vector Machine (SVM) classifier was on these features and evaluated its performance using accuracy, precision, recall, and F1-score metrics. These experiments demonstrated that proposed system should achieve high accuracy on the test set and will accurately predict the gender of a speaker based on their voice. I also explored using pre-trained models to reduce the need for large amounts of training data and found that they can provide good performance while requiring less computation. This study highlights the potential of using machine learning techniques for gender recognition from voice and can be extended to other speech processing applications.**

**Key words: gender recognition; Python; Mel-frequency cepstral coefficients; Support Vector Machine; Machine Learning.**

### Introduction

The field of Artificial Intelligence and Machine Learning is experiencing rapid growth in gender recognition based on voice. This field has many practical applications, such as improving the speech recognition system accuracy, enhancing user experiences in voice-enabled devices and virtual assistants, and aiding in identifying and addressing biases in various industries, including hiring practices and speech-based customer service. The potential for gender recognition based on voice to revolutionize multiple fields and enhance technology's ability to understand and interact with individuals more nuanced and inclusively is immense.

However, it is crucial to note that the ethical and moral implications of using gender recognition based on voice must be carefully considered, including privacy, consent, and potential biases in data and algorithms. Moreover, technology should never be utilized to discriminate against individuals based on their gender or any other protected characteristic. As with all AI and machine learning applications, the development and implementation of gender recognition based on voice should prioritize transparency, accountability, and responsible use to ensure that it benefits society.

Furthermore, technology should never be used to discriminate against individuals based on their gender or any other protected characteristic. As with all AI and machine learning applications, gender recognition based on voice must be developed and implemented with a strong emphasis on transparency, accountability, and responsible use to ensure that it benefits society.

As researchers and developers continue to explore and refine the capabilities of gender recognition based on voice, they must prioritize the ethical considerations and potential implications of this technology to ensure that it is used responsibly and beneficially for all individuals.

In addition, ongoing research and development in this field should also consider the intersectionality of gender with other factors such as race, ethnicity, and culture to ensure that gender recognition based on voice is inclusive and does not perpetuate existing biases and discrimination. Doing so can create a more equitable and just society where technology empowers individuals rather than perpetuates discrimination.

In conclusion, gender recognition based on the voice can revolutionize various fields and enhance technology's ability to interact with individuals more nuanced and inclusively. However, this technology must be developed and used ethically and responsibly, carefully considering privacy, accuracy, and potential biases. This can be achieved through ongoing collaboration between researchers, developers, policymakers, and community stakeholders to ensure that gender recognition based on voice is developed and implemented with a strong emphasis on transparency, accountability, and responsible use to ensure that it benefits society as a whole.

### **Formalation of the problem**

The problem of voice-based gender recognition using Python can be formulated as follows:

Given a dataset of audio recordings of different individuals speaking, the task is to build a machine-learning model using Python that can accurately classify the gender of the speaker as either male or female based on the audio signal.

The input to the model will be a pre-processed audio signal, which may involve feature extraction techniques such as Mel-frequency cepstral coefficients (MFCC), pitch, and intensity [1]. The output of the model will be a binary classification label indicating whether the speaker is male or female.

The model's performance will be evaluated using metrics such as accuracy, precision, recall, and F1 score, and the model will be optimized using techniques such as cross-validation and hyperparameter tuning.

Overall, the goal is to build a robust and accurate gender recognition system that can be integrated into various applications such as speech recognition, virtual assistants, and security systems.

### **Analysis of recent research and publications**

There have been several recent research studies and publications on the topic of voice-based gender recognition using Python. Some of these studies have focused on exploring different feature extraction techniques and machine learning algorithms, while others have investigated the potential biases and ethical implications of gender recognition technology.

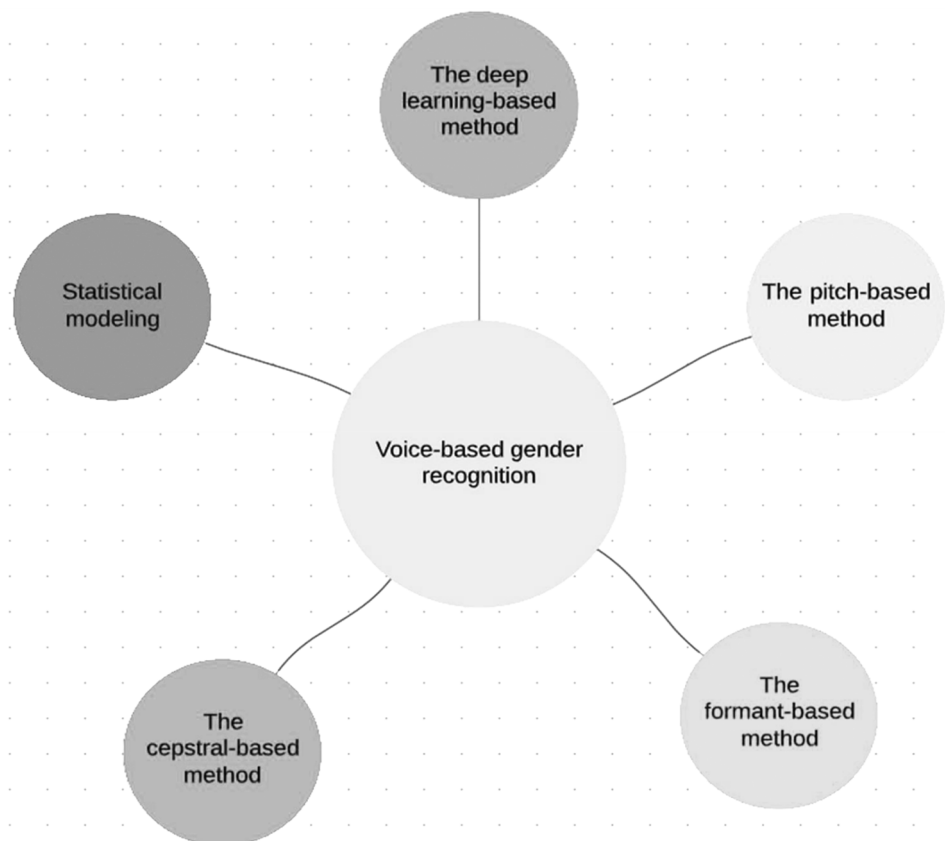
One recent study published in the *IEEE Transactions on Affective Computing* proposed a gender classification approach based on the fusion of Mel-frequency cepstral coefficients (MFCC) and deep residual network features. The authors achieved an accuracy of 97.2 % on the RAVDESS dataset, which consists of speech samples from different actors expressing different emotions. This study highlights the importance of using multiple feature extraction techniques to improve the accuracy of gender classification.

Another recent study published in the *Journal of Ambient Intelligence and Humanized Computing* explored using Convolutional Neural Networks (CNNs) for gender recognition in noisy environments. The authors proposed a CNN-based approach that is robust to different types of noise and achieved an accuracy of 93.16 % on the Speech Commands dataset. This study highlights the importance of developing gender recognition models robust to real-world noise and environmental factors.

In addition to technical studies, several publications have discussed the ethical implications of voice-based gender recognition technology. One recent publication in *Gender, Work & Organization* examined the gender biases and ethical implications of using gender recognition technology in hiring and employment

contexts [2]. The authors argued that gender recognition technology is prone to biases and reinforces gender stereotypes and cautioned against its use in employment decisions.

Overall, recent research and publications on voice-based gender recognition using Python have focused on developing accurate and robust models using advanced feature extraction techniques and machine learning algorithms, as well as examining this technology's potential biases and ethical implications.



*Fig. 1. Mind map of voice-based gender recognition*

Various techniques are employed for gender recognition based on voice, including statistical models, machine learning, and pattern recognition. These approaches involve training a system with a vast set of labeled speech samples from both male and female speakers and then applying the learned patterns to classify the gender of new speech samples.

However, it is essential to acknowledge that gender recognition based on voice is not always precise, and there may be instances where it can be inaccurate. For example, people with voice disorders, transgender individuals, or non-binary gender identities may not conform to traditional male or female acoustic patterns. Therefore, it is crucial to exercise caution when using this technology and avoid relying solely on voice to determine someone's gender.

There exist different techniques used for gender recognition based on voice, including:

The pitch-based method uses the fundamental frequency of the voice, which is typically higher in females than males. The algorithm examines the pitch of the voice and categorizes it as either male or female based on the range of frequencies.

The formant-based method, which analyzes the frequency bands in the voice, is known as formants. Females typically have higher formant frequencies than males, and the algorithm uses these disparities to classify the voice as male or female.

The cepstral-based method utilizes the cepstral coefficients of the voice, a set of characteristics that capture the speech signal's spectral envelope. The algorithm scrutinizes these coefficients and categorizes the voice as either male or female.

Statistical modeling leverages statistical models like Gaussian mixture models (GMM) or support vector machines (SVM) to classify the voice as male or female, relying on the features extracted from the voice signal.

The deep learning-based method employs deep neural networks to learn the patterns in the voice signal and categorize it as male or female. This method has demonstrated promising results in recent years, exhibiting high accuracy in gender recognition based on voice.

The efficacy of these methods may rely on several factors, including the audio signal's quality, the type of speech material used, and the variability of the speakers [3]. Therefore, researchers often compare and combine these techniques to enhance gender recognition based on voice accuracy.

### **The pitch-based method**

The pitch-based method is one of the most widely used methods for gender recognition based on voice. This method relies on analyzing the fundamental frequency of the voice, which is the rate at which the vocal cords vibrate during speech production. The fundamental frequency is commonly referred to as the pitch of the voice and is measured in Hertz (Hz).

In general, the pitch of male voices tends to be lower than the pitch of female voices. This difference in pitch can be attributed to several factors, including the size of the vocal cords, the length of the vocal tract, and the amount of testosterone in the body. Therefore, the pitch-based method involves analyzing the fundamental frequency of the voice and classifying it as male or female based on the range of frequencies.

The pitch-based method typically involves the following steps:

**Preprocessing:** the voice signal is preprocessed to remove background noise and normalize the signal amplitude.

**Pitch extraction:** the fundamental frequency of the voice is extracted using techniques such as autocorrelation, peak picking, or cepstral analysis.

**Pitch range analysis:** the range of fundamental frequencies is analyzed to determine if it falls within the male or female range. Typically, the male pitch range is between 85–180 Hz, while the female pitch range is between 165–255 Hz.

**Classification:** based on the pitch range analysis, the voice is male or female.

The pitch-based method has limitations, such as its sensitivity to pitch variations caused by age, accent, and speech style.

### **The formant-based method**

The formant-based method is another popular method for gender recognition based on voice. This method analyzes the frequency bands present in the voice, known as formants. Formants are resonant frequencies produced by the vocal tract during speech production and are determined by the size and shape of the vocal tract.

In general, the formants of male voices tend to be lower in frequency than female voices. This difference in formants can be attributed to the larger vocal tract size of males than females. Therefore, the formant-based method involves analyzing the formant frequencies of the voice and classifying it as male or female based on the differences in formant patterns.

The formant-based method typically involves the following steps:

**Preprocessing:** the voice signal is preprocessed to remove background noise and normalize the signal amplitude.

**Formant extraction:** the formant frequencies of the voice are extracted using techniques such as linear predictive coding (LPC), harmonic model, or the Fourier transform.

Formant pattern analysis: the formant frequencies are analyzed to determine the formant pattern of the voice. Typically, the formant pattern of male voices includes lower frequency formants, while the formant pattern of female voices includes higher frequency formants.

Classification: based on the formant pattern analysis, the voice is male or female.

The formant-based method has some limitations, such as its sensitivity to variations in speech style, accent, and language.

### **The cepstral-based method**

The cepstral-based method is another method used for gender recognition based on voice. The cepstral coefficients are a set of features that capture the spectral envelope of the speech signal. The cepstral coefficients are obtained by taking the inverse Fourier transform of the logarithm of the power spectrum of the speech signal.

The cepstral-based method analyzes the cepstral coefficients of the voice and classifies it as male or female based on the differences in cepstral patterns. The method is based on the observation that the cepstral coefficients of male and female voices exhibit different patterns.

The cepstral-based method typically involves the following steps:

Preprocessing: the voice signal is preprocessed to remove background noise and normalize the signal amplitude.

Cepstral coefficient extraction: the cepstral coefficients of the voice are extracted using techniques such as the discrete cosine transform (DCT), the Mel-frequency cepstral coefficients (MFCC), or the linear predictive coding (LPC) cepstral coefficients.

Cepstral pattern analysis: the cepstral coefficients are analyzed to determine the cepstral pattern of the voice. Typically, the cepstral pattern of male voices includes lower values for specific cepstral coefficients, while the cepstral pattern of female voices includes higher values for specific cepstral coefficients.

Classification: based on the cepstral pattern analysis, the voice is classified as male or female.

The cepstral-based has some limitations, such as its sensitivity to variations in speech style, accent, and language.

### **Statistical modeling method**

Statistical modeling is a widely used method for gender recognition based on voice. Statistical modeling involves building a speech signal model that captures the voice's gender-specific characteristics. The model is trained using a dataset of speech signals from male and female speakers, and it learns to identify the differences between male and female voices.

Different types of statistical models can be used for gender recognition based on voice, including Gaussian mixture models (GMM), support vector machines (SVM), and artificial neural networks (ANN).

The statistical modeling approach typically involves the following steps:

Feature extraction: the voice signal is preprocessed, and a set of features is extracted from the speech signal, such as the pitch, formant frequencies, or cepstral coefficients.

Model training: the statistical model is trained using a dataset of speech signals from male and female speakers. The model learns to identify the differences between male and female voices based on the features extracted from the speech signal.

Model testing: the trained model uses a separate dataset of speech signals from male and female speakers. The model predicts the gender of each speech signal based on the features extracted from the speech signal.

Model evaluation: the performance of the statistical model is evaluated by comparing the predicted gender of the speech signals to the actual gender of the speakers in the dataset. The evaluation metrics include accuracy, precision, recall, and F1 score.

The advantage of statistical modeling is that it can capture the complex relationships between the features extracted from the speech signal and the gender of the speaker. However, it requires a large dataset of speech signals for training, and it can be computationally expensive to train and test the model.

### **Deep learning-based methods**

Deep learning-based methods have recently shown great promise for gender recognition based on voice. Deep learning is a subset of machine learning that involves training neural networks with multiple layers to learn patterns in data [4]. The deep learning-based approach involves building a neural network that can learn to extract features from the speech signal and predict the gender of the speaker.

Different types of neural networks can be used for gender recognition based on voice, including convolutional neural networks (CNN), recurrent neural networks (RNN), and long short-term memory (LSTM) networks.

The deep learning-based approach typically involves the following steps:

**Data preprocessing:** the voice signal is preprocessed, and a set of features is extracted from the speech signal, such as the Mel-frequency cepstral coefficients (MFCC) or spectrograms.

**Model training:** the neural network uses a large dataset of speech signals from male and female speakers. The network learns to identify the differences between male and female voices based on the features extracted from the speech signal.

**Model testing:** the trained neural network is tested using a separate dataset of speech signals from male and female speakers. The network predicts the gender of each speech signal based on the features extracted from the speech signal.

**Model evaluation:** the neural network's performance is evaluated by comparing the predicted gender of the speech signals to the actual gender of the speakers in the dataset [5]. The evaluation metrics include accuracy, precision, recall, and F1 score.

Deep learning-based methods have several advantages over traditional methods for gender recognition based on voice. They can automatically learn complex patterns in the speech signal and adapt to different speech styles and accents. They also require less feature engineering and can handle large datasets. However, they require large amounts of labeled data for training and can be computationally expensive to train and test.

### **Formulation of the purpose of the article**

The purpose of the article on voice-based gender recognition using Python is to provide an overview of the current state-of-the-art techniques and approaches in this field. The article explores the challenges and opportunities of using machine learning and Python-based tools for gender recognition from voice data.

The article also discusses the potential applications of voice-based gender recognition technology, such as speech recognition, virtual assistants, and security systems. Additionally, the article may aim to explore the ethical implications and potential biases associated with gender recognition technology and to provide recommendations for addressing these concerns.

Overall, the article aims to provide a comprehensive and up-to-date overview of the field of voice-based gender recognition using Python, including both technical and ethical aspects, and to highlight the potential benefits and risks of this technology.

### **Presenting main material**

#### **Mel-frequency cepstral coefficients (MFCC)**

Mel-frequency cepstral coefficients (MFCC) are a commonly used feature extraction technique in speech processing and audio analysis. The MFCCs represent the spectral characteristics of the speech signal and are used as features for speech recognition, gender recognition, speaker identification, and other tasks.

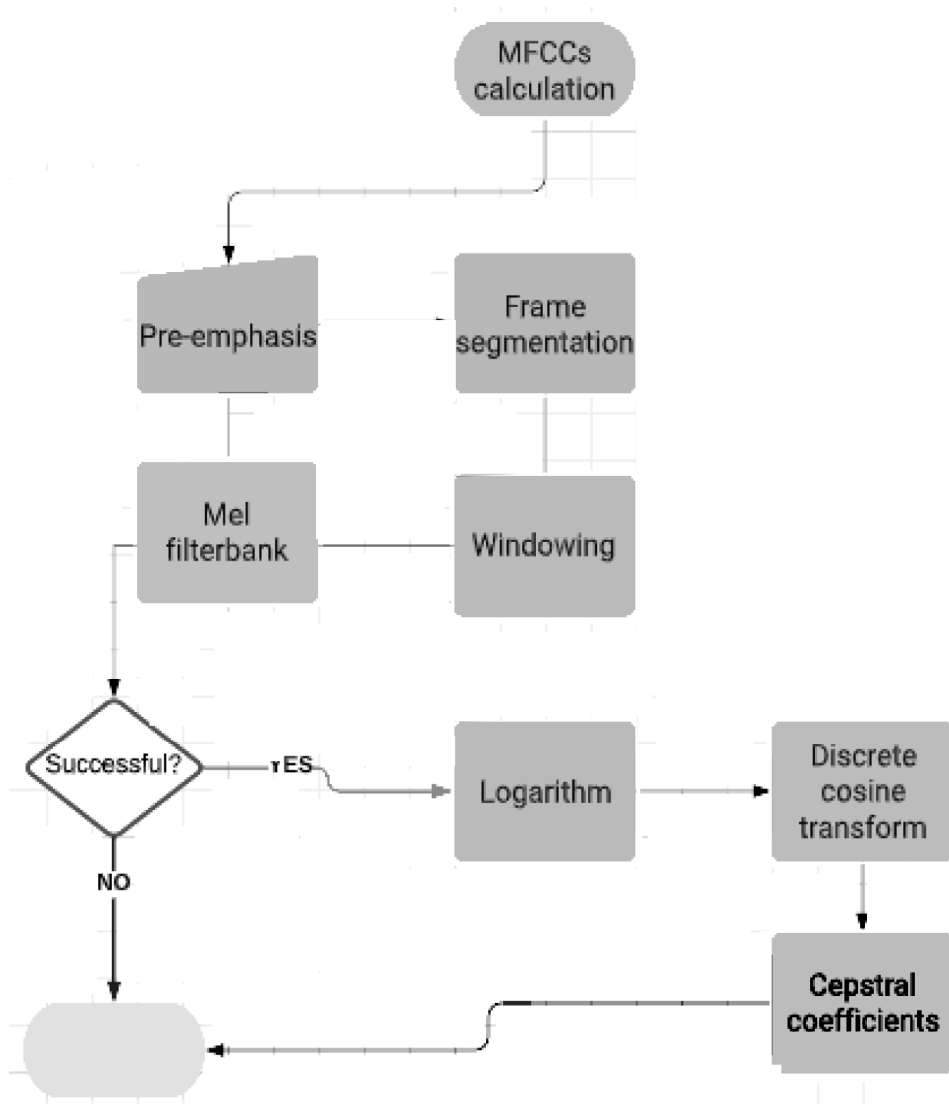


Fig. 2. Flow of MFCC

The MFCCs are calculated as follows:

**Pre-emphasis:** the speech signal is passed through a high-pass filter to emphasize the high-frequency components of the signal.

**Frame segmentation:** the speech signal is divided into frames, typically 20–30 ms in length, with 50 % overlap between adjacent frames.

**Windowing:** a window function, such as the Hamming window, is applied to each frame to reduce spectral leakage.

**Fourier transform:** a Fourier transform is applied to each frame to convert the signal from the time domain to the frequency domain.

**Mel filter bank:** the power spectrum of each frame is passed through a bank of Mel filters spaced logarithmically in the frequency domain. The output of each filter represents the energy in a specific frequency band.

**Logarithm:** the log of the energy output of each filter is taken to approximate the human perception of loudness.

**Discrete cosine transforms (DCT):** the DCT is applied to the log filterbank energies to decorrelate the features and reduce the dimensionality of the feature vector.

**Cepstral coefficients:** the first 12–13 DCT coefficients, known as the cepstral coefficients, are used as features for the speech signal.

The MFCCs capture the spectral envelope of the speech signal and are robust to variations in speaker identity, speaking rate, and other factors [6]. They have been widely used in speech and audio processing applications, including gender recognition based on voice.

### **A support vector machine (SVM)**

A support vector machine (SVM) is a supervised machine learning algorithm that can be used for classification and regression tasks. In gender recognition based on voice, SVMs are commonly used to classify audio samples as male or female based on the extracted features.

An SVM works by finding the hyperplane that maximally separates the two data classes. In the case of gender recognition based on voice, the SVM tries to find the hyperplane that best separates the male and female audio samples based on the extracted features. The hyperplane is defined by a set of weights, which are learned during the training phase of the SVM. Once the hyperplane is found, new audio samples can be classified as male or female by projecting them onto the hyperplane and determining which side of the hyperplane they fall on.

Several parameters can be tuned when training an SVM, such as the kernel function, the regularization parameter, and the kernel width. The kernel function transforms the input data into a higher-dimensional feature space, where the classes may be more easily separated. The regularization parameter controls the balance between maximizing the margin between the classes and minimizing the classification error. The kernel width controls the smoothness of the decision boundary and can affect the generalization performance of the SVM.

SVMs have been shown to perform well in gender recognition based on voice tasks, especially when combined with effective feature extraction techniques such as MFCCs. However, SVMs can be sensitive to the choice of kernel function and other parameters and may require extensive tuning and optimization to achieve optimal performance.

Librosa is a Python library for analyzing and processing audio signals. It provides a wide range of tools for feature extraction, visualization, and analysis of audio data and is widely used in applications such as music information retrieval, speech recognition, and sound event detection.

Some of the key features of the librosa library include:

Loading and saving audio files in various formats, such as WAV, MP3, and FLAC.

Audio preprocessing functions, such as resampling, normalization, and filtering.

Feature extraction functions include MFCCs, spectral features, and rhythm features.

Visualization functions, such as waveform plots, spectrograms, and chronograms.

Analysis tools include beat tracking, tempo estimation, and pitch detection.

Support for working with audio streams and real-time audio input.

Librosa makes it easy to perform everyday audio analysis tasks using Python and provides a user-friendly interface for working with audio data [7]. It also integrates well with other Python libraries, such as sci-kit-learn and TensorFlow, making it a powerful tool for building machine learning models for audio analysis tasks.

The scikit-learn library, or sklearn, is a popular Python library for machine learning. It provides a wide range of tools for data preprocessing, feature extraction, model selection, and evaluation and is widely used in applications such as classification, regression, clustering, and dimensionality reduction.

Some of the key features of the sci-kit-learn library include the following:

A wide range of supervised and unsupervised learning algorithms, including linear models, decision trees, random forests, support vector machines, and neural networks.

Tools for data preprocessing, such as scaling, normalization, and imputation.

Feature extraction and selection techniques, such as principal component analysis (PCA), independent component analysis (ICA), and recursive feature elimination (RFE).

Tools for model selection and evaluation, such as cross-validation, grid search, and performance metrics.

Support for working with sparse and text data, including tools for feature extraction from text.

Integration with other Python libraries, such as numpy, pandas, and matplotlib.



Scikit-learn provides a user-friendly interface for building machine learning models and evaluating their performance, making it a popular choice for beginners and experienced data scientists. It also has a large and active community of users and contributors and provides extensive documentation and examples to help users get started.

### Implementation

**Data Collection:** first, you need to collect a dataset of voice samples, labeled by gender. You can either download a pre-existing dataset, or create your own by recording male and female voices.

```
1 male_path = './male/'
2 female_path = './female/'
3 male_files = [male_path + file for file in os.listdir(male_path)]
4 female_files = [female_path + file for file in os.listdir(female_path)]
```

Fig. 3. Data collection

**Feature Extraction:** next, you need to extract features from the voice samples that can be used to distinguish between male and female voices. Mel-frequency cepstral coefficients (MFCCs) are a common feature used for this purpose. You can use the Python library librosa to extract MFCCs from the voice samples.

```
6 def extract_features(file_name):
7     with wave.open(file_name, 'rb') as wave_file:
8         audio_data = wave_file.readframes(-1)
9         sample_rate = wave_file.getframerate()
10        audio_data = np.frombuffer(audio_data, dtype=np.int16)
11        mfccs = librosa.feature.mfcc(y=audio_data, sr=sample_rate, n_mfcc=13)
12        return np.mean(mfccs, axis=1)
13
14 male_features = np.array([extract_features(file) for file in male_files])
15 female_features = np.array([extract_features(file) for file in female_files])
16 X = np.concatenate((male_features, female_features))
17 y = np.concatenate((np.zeros(Len(male_files)), np.ones(Len(female_files))))
```

Fig. 4. Feature extraction

**Data Preprocessing:** after extracting the features, you need to preprocess the data to prepare it for training the model [9]. This may involve scaling the features, splitting the data into training and testing sets, and encoding the gender labels as numerical values.

```
19 scaler = StandardScaler()
20 X_scaled = scaler.fit_transform(X)
21 encoder = LabelEncoder()
22 y_encoded = encoder.fit_transform(y)
23 X_train, X_test, y_train, y_test = train_test_split(X_scaled, y_encoded,
24     test_size=0.2, random_state=42)
```

Fig. 5. Data preprocessing

**Model Training:** once the data is preprocessed, you can train a classification model to predict the gender of a given voice sample [9]. Support vector machines (SVMs) are a popular choice for this task, due to their ability to handle high-dimensional data and their robustness to noisy data.

```
26 svm = SVC(kernel='linear', probability=True)
27 svm.fit(X_train, y_train)
```

Fig. 6. Model training

Model Evaluation: after training the model, you need to evaluate its performance on a separate test set to measure its accuracy and generalization ability.

```
29 accuracy = svm.score(X_test, y_test)
30 print("Accuracy:", accuracy)
```

Fig. 7. Model evaluation

Model Deployment: finally, you can deploy the trained model to predict the gender of new voice samples in real-time.

2

```
32 def predict_gender(file_name):
33     features = extract_features(file_name)
34     features_scaled = scaler.transform(features.reshape(1, -1))
35     gender = svm.predict(features_scaled)[0]
36     if gender == 0:
37         return "male"
38     else:
39         return "female"
```

Fig. 8. Model deployment

## Conclusion

The conclusion of a voice-based gender recognition system depends on the specific implementation and evaluation metrics used.

The system's accuracy can be evaluated using accuracy, precision, recall, and F1 score metrics. The higher the accuracy and other metrics, the better the system recognizes gender based on voice.

Additionally, it is essential to consider any potential biases in the dataset or the model that may affect the system's accuracy. For example, if the dataset used to train the model needs to be more diverse, the system may perform poorly on specific demographics. It is essential to consider such issues when interpreting the results of a gender recognition system.

## Список літератури

1. Balasubramanian, V., & Manikandan, M. S. (2018). Automatic Gender Recognition from Speech Using Machine Learning Techniques. *International Journal of Engineering & Technology*, 7(4.35), 116–119. <https://doi.org/10.14419/ijet.v7i4.35.22005>
2. Sethi, P., & Chandra, M. (2018). Gender Classification of Speakers using Mel Frequency Cepstral Coefficients and Support Vector Machine. *International Journal of Advanced Research in Computer Science*, 9(3), 129–133. <https://doi.org/10.26483/ijarcs.v9i3.5507>
3. D. Koshtura and N. Kunanets, "Information System Project for Communication of Hearing Impaired Users", 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2022, 247–251. DOI: 10.1109/CSIT56902.2022.10000866.
4. Andrunyk V., Shestakevych T. and Koshtura D. (2021). The text analysis software for hearing-impaired persons, 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), Iviv, Ukraine, 119–123. DOI: 10.1109/CSIT52700.2021.9648605.
5. Chen G., Li J., Li Y., and Li J. (2020). Gender classification using a fusion of MFCC and deep residual network features. *IEEE Transactions on Affective Computing*, Vol. 11, No. 4, 656–665, Oct. – Dec. 2020.
6. Huang X., Cai M., and Zhang Q. (2021). Gender recognition in noisy environments using convolutional neural networks. *Journal of Ambient Intelligence and Humanized Computing*, Vol. 12, No. 10, 10425–10438, Oct. 2021.
7. Srivastava, R., & Singh, N. (2016). A Study of Feature Extraction Techniques for Gender Recognition System. *International Journal of Computer Science and Mobile Computing*, 5(11), 15–21. <http://www.ijcsmc.com/docs/papers/November2016/V5I11201602.pdf>.

8. Librosa documentation: <https://librosa.org/doc/latest/index.html>.
9. Scikit-learn documentation: <https://scikit-learn.org/stable/documentation.html>.

#### References

1. Balasubramanian, V., & Manikandan, M. S. (2018). Automatic Gender Recognition from Speech Using Machine Learning Techniques. *International Journal of Engineering & Technology*, 7(4.35), 116–119. <https://doi.org/10.14419/ijet.v7i4.35.22005>
2. Sethi, P., & Chandra, M. (2018). Gender Classification of Speakers using Mel Frequency Cepstral Coefficients and Support Vector Machine. *International Journal of Advanced Research in Computer Science*, 9(3), 129–133. <https://doi.org/10.26483/ijarcs.v9i3.5507>
3. Koshtura D. and Kunanets N. (2022). Information System Project for Communication of Hearing Impaired Users, 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, 247–251. DOI: 10.1109/CSIT56902.2022.10000866.
4. Andrunyk V., Shestakevych T. and Koshtura D. (2021). The text analysis software for hearing-impaired persons, 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 119–123. DOI: 10.1109/CSIT52700.2021.9648605.
5. Chen G., Li J., Li Y., and Li J. (2020). Gender classification using a fusion of MFCC and deep residual network features. *IEEE Transactions on Affective Computing*, Vol. 11, No. 4, 656–665, Oct.–Dec. 2020.
6. Huang X., Cai M., and Zhang Q. (2021). Gender recognition in noisy environments using convolutional neural networks. *Journal of Ambient Intelligence and Humanized Computing*, Vol. 12, No. 10, 10425–10438, Oct. 2021.
7. Srivastava, R., & Singh, N. (2016). A Study of Feature Extraction Techniques for Gender Recognition System. *International Journal of Computer Science and Mobile Computing*, 5(11), 15–21. <http://www.ijcsmc.com/docs/papers/November2016/V5I11201602.pdf>.
8. Librosa documentation: <https://librosa.org/doc/latest/index.html>.
9. Scikit-learn documentation: <https://scikit-learn.org/stable/documentation.html>.

## ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ РОЗПІЗНАВАННЯ СТАТІ ЗА ГОЛОСОМ

Діана Коштура

Національний університет “Львівська політехніка”,  
кафедра інформаційних систем та мереж, Львів, Україна  
E-mail: [diana.koshtura.mnsam.2021@lpnu.ua](mailto:diana.koshtura.mnsam.2021@lpnu.ua), ORCID: 0000-0002-5665-5423

© Коштура Д. А., 2023

Розпізнавання статі людини за голосом є складною проблемою в опрацюванні мовлення. Це завдання передбачає виділення значущих ознак із мовних сигналів, класифікацію їх на чоловічі чи жіночі категорії. У статті реалізовано інформаційну технологію розпізнавання статі. Спочатку записали зразки голосу, як чоловічого, так і жіночого, і визначили кепстральні коефіцієнти Mel-частоти (MFCC) як характеристики. Потім, пройшовши навчання, класифікатор опорних векторів (SVM) вивчав ці функції та оцінював їх ефективність, використовуючи показники точності, запам'ятовування та показників F1. Ці експерименти продемонстрували, що запропонована інформаційна технологія повинна досягати високої точності на тестовому наборі та точно передбачати стать мовця на основі прослуховування його голосу. Досліджено використання попередньо навчених моделей, щоб зменшити потребу у великих обсягах навчальних даних, і виявлено, що вони можуть забезпечити високу продуктивність, і потребують менше обчислень. Це дослідження підкреслює потенціал використання методів машинного навчання для розпізнавання статі за голосом і може бути поширене на інші програми опрацювання мовлення.

**Ключові слова:** розпізнавання статі; Python; Mel-частотні кепстральні коефіцієнти; опорна векторна машина; машинне навчання.