# RESEARCH AND SELECTION

## 1. A study on Data Augmentation in Voice Anti- Spoofing

### 1.1 ABSTRACT

In this paper, the author investigates how data augmentation techniques can enhance the detection of spoofed audio. It also talks about how challenges like audio compressions, channel variability and different bandwiths in the same audio sample can degrade the performance of anti spoofing system.

### 1.2 KEY TECHNICAL INNOVATIONS

- **Compression and Channel Augmentation:** This increases the robustness of the system against the compressed and transmitted audio data. This helps the model to generalize the real-world scenarios better when the audio quality maybe compromised.
- **Double sided log spectrogram:** Proposed a unique design that centers the sub bands of interest in the log spectrogram. This enhances the model's ability to detect discriminatory features associated with spoofing artifacts.
- **SpecAverage Augmentation:** This new technique masks the audio features with their average values. This improves the generalization capabilities and prevents overfitting.

### 1.3 REPORTED PERFORMANCE METRICS

- **Deepfake:** Their best single system achieved Equal Error Rate of 15.46%.
- **Logical access:** Their approach reduced the EER by 50%.

### 1.4 RELEVANCE TO SPECIFIC NEEDS

- **Detecting AI Generated Human Speech:** The proposed technique enhances the model’s ability to detect synthetic speech which is crucial in identifying spoofed or generated speech.
- **Real time or near real time detection:** The paper calls for improvements in the model’s robustness which if integrated with the system, it can be possible
- **Analysis of real conversations:** It simulates real world audio degenerations through compression and channel augmentation which ensures that the model remains effective when analyzing authentic conversations

### 1.5 POSSIBLE LIMITATIONS

- **Computational costs:** In augmentation techniques, a lot of complex computation requirements are necessary which often at times are not fulfilled This may impact the feasibility of deployment
- **Generalization:** Even though there are number of methods of improving generalization, the ever evolving nature of spoofing attacks means there is a need of continuous updates and evolving strategies to track them.

## 2. <u>**A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions**</u>

### 2.1 ABSTRACT

In this paper the author stresses on the current techniques for detecting deepfakes and does in depth analysis on both imitation based and synthetic based methods. They discuss the various pros and cons of different strategy used and provide us with a good comparative study to understand the stark differences of the techniques based on the different situations

### 2.2 KEY TECHNICAL INSIGHTS

- **Categorization of detection methods:** The paper classifies the existing audio deepfake techniques into imitation based which focuses on human imitating audio and synthetic based which focuses on audio aletred by AI.

- **Evaluation of datasets:** The authors review the existing datasets utilized for training and testing and come with a conclusion that since most of it was in English, it doesnt substantiate its credibility as for such models diverse languages must be used for robustness.

### 2.3 CHALLENGES

- **Language limitations:** As discussed earlier, majority of the datasets were in English and limited attention was given to other languages.
- **Accent variability:** The existing models overlooked the impact of different accents which affected the detection accuracy.
- **Noise robustness:** There is a lack of assessment regarding how real world noise hinders detection performance.

### 2.4 Potential Limitations or Challenges

- **Language and Accent Diversity:** The dataset was found to be too english centric along with very less diverse accents which makes it vulnerable to other language and other accent detection.
- **Scalability Concerns:** Since there are so many preprocessing steps, it is difficult to scale such a model.
- **Robustness to Environmental Noise:** There were no concrete methods that could tell how to deal with noise and how the models would combat them.

## 3. Voice Spoofing Attacks and Countermeasures: A Systematic Review, Analysis, and Experimental Evaluation

### 3.1 ABSTRACT

In this paper the authors did a comprehensive study of voice spoofing attacks and the countermeasures designed to protect the ASV (Automatic Speaker Verification) systems. The paper evaluates the effective existing detection measures by categorizing them and discusses the future challenges and directions in this field.

### 3.2 KEY TECHNICAL INSIGHTS

- **Spoofing attacks classification:** The authors classified the spoofing attacks into Logical Access which involves the synthetic speech and voice conversion techniques and Physical Access attacks wherein a clip of pre-recorded audio would be played.
- **Countermeasures:** They also categorized countermeasures which revolved around deep learning, hand crafted features and end to end system approaches for the same.
- **Experimental analysis:** Their study conducts extensive experiments with various heavy datasets like ASVspoof2019, 2021, etc using various classifiers like CNN, SVM, Gaussian distribution. All this was done to assess the generalizability of countermeasures across different datasets.

### 3.3 CHALLENGES

- **Dataset variability:** The difference in datasets due to their recording equipment, environment and other measures.
- **Generalizability:** Many countermeasures perform well on specific datasets but fail to generalize across others. This leads to vulnerability in the ASV systems as there is a lack of adversarial aware countermeasures

### 3.4 RELEVANCE TO SPECIFIC NEEDS

- **Detecting AI generated human speech:** The research aligns with the need to identify AI-generated human speech effectively.
- **Real time detection:** The computational complexity of various countermeasures provides insights into their real time application.
- **Analysis of real conversations:** The evaluating the countermeasures across all the different datasets that simulate real world scenarios, and they address the applicability of these methods to authenticate conversational data.