SPEECH EMOTIONAL RECOGNITION ON LIVE CALL

Problem Statement

- Cybercrime police receive thousands of calls daily.

- Difficulties in identifying genuine distress calls due to lack of emotion detection.

- Victims may hesitate to express distress, while fraudsters may manipulate emotions.

- Need for an AI-based Speech Emotion Recognition (SER) system to analyze emotions in real-time and prioritize emergency responses.

Objectives

1.Real-Time Emotion Analysis – Detect emotions like happy, sad, fear, panic, angry, neutral, nervous, etc.

2.Prioritize Critical Cases – Identify urgency or distress to trigger immediate action.

3.Support Investigations – Provide emotional context to assist cybercrime officers.

4.Improve Victim Assistance – Ensure better support based on emotional state.

System Workflow

1. **Victim Dials Emergency Number** – Calls are received on a cybercrime helpline.

2. **Call Forwarding** – Calls are routed to an SIP(Session Initiation Protocol) number.

3. **Asterisk VoIP Server** – Captures and records calls in WAV format and converts GSM(Global System for Mobile Communications) calls into VoIP calls.

4. **Audio Processing** – FFmpeg enhances and normalizes speech signals (FFmpeg is a software that can be used to enhance audio quality).

5. **Emotion Recognition Model** – Analyzes emotions from speech using CNN & NLP.

6. **Real-Time Dashboard** – Displays detected emotions for immediate action.

FLOW DIAGRAM:

SPEECH EMOTION RECOGNITION ON LIVE CALLS
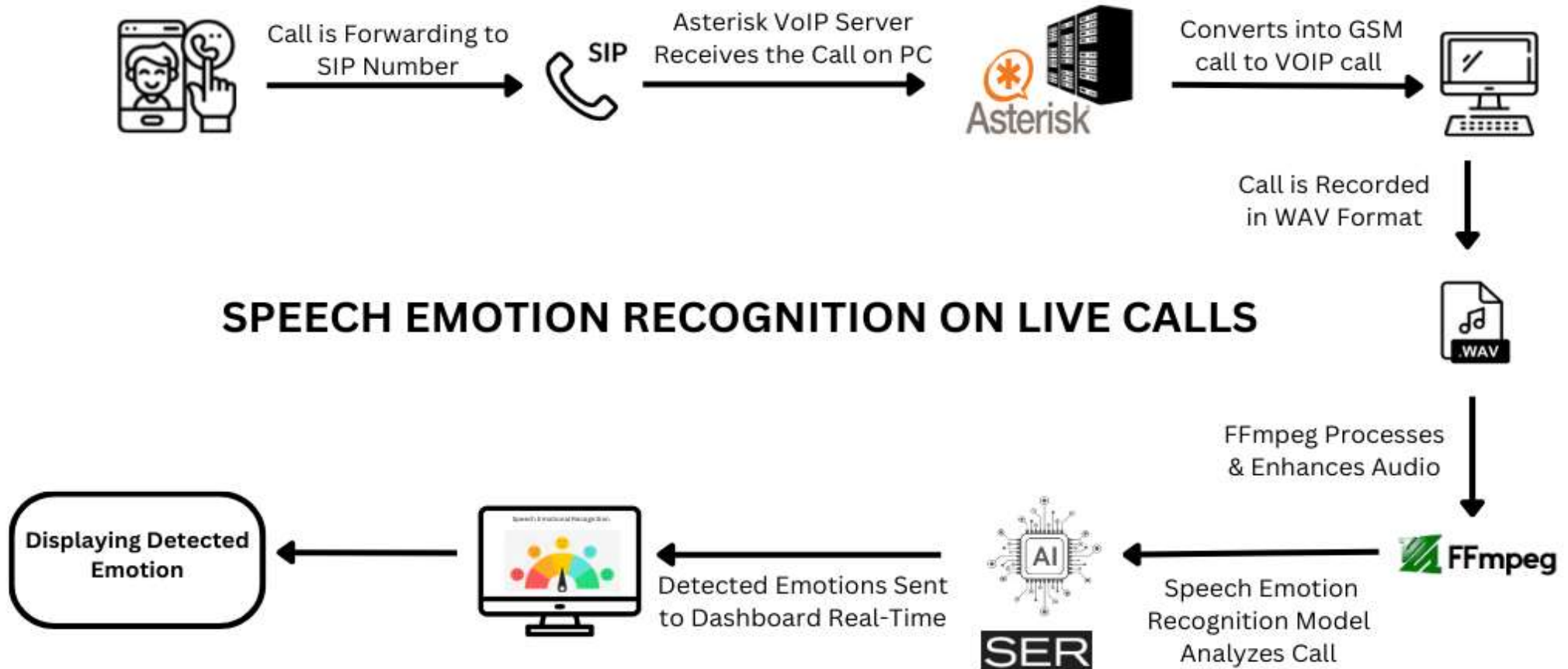
AI Model for Speech Emotion Recognition

1. CNN (Convolutional Neural Network) – Extracting Features from Audio

- **Purpose:** CNN processes speech by converting it into a Mel spectrogram (a visual representation of audio frequencies over time).

- **How it Works:**

1. **Convert WAV to Mel Spectrogram:** Represents frequency and intensity of speech over time.

2. **Apply CNN Layers:** CNN scans the spectrogram like an image to identify emotion-related patterns.

3. **Feature Extraction:** CNN learns key frequency variations (e.g., anger has a higher pitch, sadness has lower energy).

4. **Classification:** Outputs probability scores for emotions (Happy, Sad, Angry, etc.).

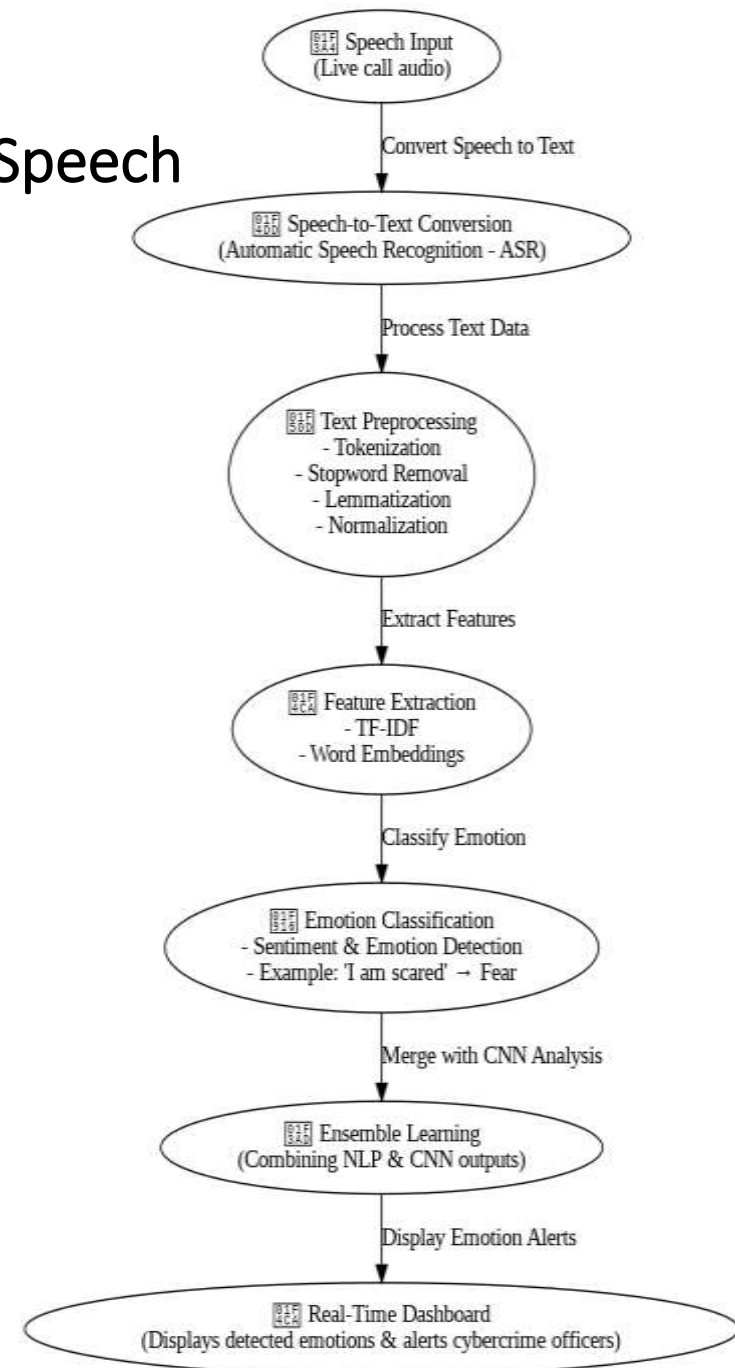# 2. NLP (Natural Language Processing) – Understanding Text from Speech

- Purpose: NLP analyzes spoken words and determines emotional intent from text.

- How it Works:

1. Speech-to-Text Conversion: Converts speech to text using Automatic Speech Recognition (ASR).

2. Text Preprocessing: Tokenizing words, removing stop words, and normalizing text.
3. Sentiment & Emotion Analysis:

    NLP models analyze word meanings.
    Example: "I am scared" → Detected as Fear.
    Example: "I am so happy today" → Detected as Happiness.

4. Emotion Classification: The NLP model assigns an emotion label based on word meanings.

Speech Input
(Live call audio)

Convert Speech to Text

Speech-to-Text Conversion
(Automatic Speech Recognition - ASR)

Process Text Data

Text Preprocessing
- Tokenization
- Stopword Removal
- Lemmatization
- Normalization

Extract Features

Feature Extraction
- TF-IDF
- Word Embeddings

Classify Emotion

Emotion Classification
- Sentiment & Emotion Detection
- Example: 'I am scared' → Fear

Merge with CNN Analysis

Ensemble Learning
(Combining NLP & CNN outputs)

Display Emotion Alerts

Real-Time Dashboard
(Displays detected emotions & alerts cybercrime officers)

AI Model for Speech Emotion Recognition

3. Ensemble Learning – Combining CNN & NLP for Better Accuracy

- **Purpose:** Since emotions can be detected through both voice tone (CNN) and words (NLP), ensemble learning combines both models for higher accuracy.

- **How it Works:**

1. CNN Prediction: CNN model gives emotion labels based on Mel spectrograms.

2. NLP Prediction: NLP model gives emotion labels based on text analysis.

3. **Final Decision Making:**

    Weighted Averaging: Assigns different importance to CNN and NLP predictions.
    Voting Mechanism: If both CNN and NLP agree on an emotion, it is selected.
    Neural Fusion: A separate model learns to combine CNN and NLP outputs.

Real-Time Dashboard & Alert System

- **Purpose:** Displays live emotions detected in calls, enabling quick decision-making.

- **Features:**
1. Live Call Feed: Shows ongoing calls in real time.
2. Emotion Alerts: Displays detected emotions on a dashboard.

- **Color-Coded Alerts:**

- ❖ Fear/Panic → Immediate emergency response.
- ❖ Anger   → Potential escalation, needs monitoring.
- ❖ Sadness   → Distressed caller, requires attention.
- ❖ Happiness → No urgent action required.

- **Impact:**

1. Helps prioritize emergency calls.
2. Reduces response time for high-risk cases.
3. Provides data-driven insights for cybercrime investigations.

Future Work & Enhancements

Future Enhancements:

- Multi-Language Support: Expand to **Tamil, Telugu, Hindi**, etc.

- Larger Dataset Training: Improve model accuracy with diverse speech samples.

- More Advanced Ensemble Learning: Implement meta-classifiers for better fusion.

- Integration with Call History: Use past records to refine emotion analysis.

THANK YOU