Neural Knights

Domain 3 Task 1

Table Of Content

01. Our Team

02. Problem statement

03. Developing and Training a Speech to Text Model.

04. User Workflow

05. Solution Architecture

06. Demo

Meet Our Team

Yap Jack

Ng Jie Ru

Phoo Cheng Yang

Pang YIk Neng

Ian Tong Yuan Jun

Problem Statement

To build a **robust voice interaction system** that enables reliable driver–assistant communication in **challenging audio environments**

What to ACHIEVE

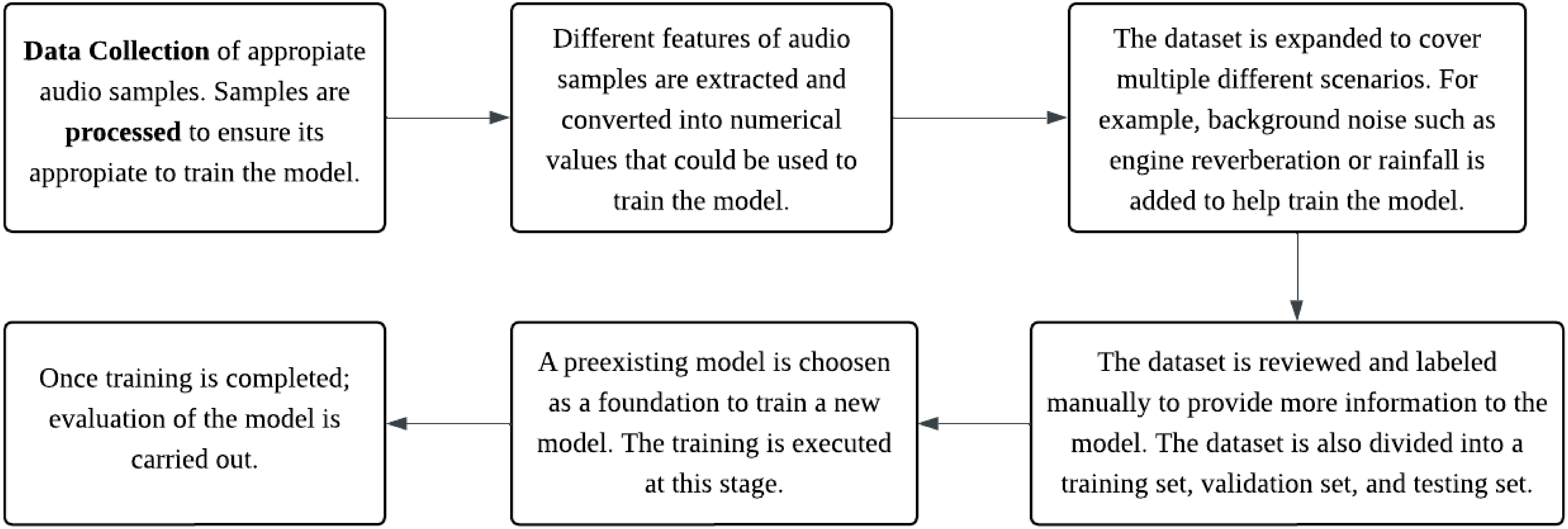
01 MAINTAIN **HIGH ACCURACY** IN NOISY CONDITIONS

02 ADAPT TO **DIVERSE** SPEECH PATTERNS

03 **CLEAR & RELIABLE FUNCTIONALITY** WITH PARTIAL AUDIO CLARITY

04 **RESILIENCE** ACROSS VARIOUS ENVIRONMENTAL CHALLENGES

Developing and Training a Speech-to-Text Model



Data Collection of appropiate audio samples. Samples are **processed** to ensure its appropiate to train the model.

Different features of audio samples are extracted and converted into numerical values that could be used to train the model.

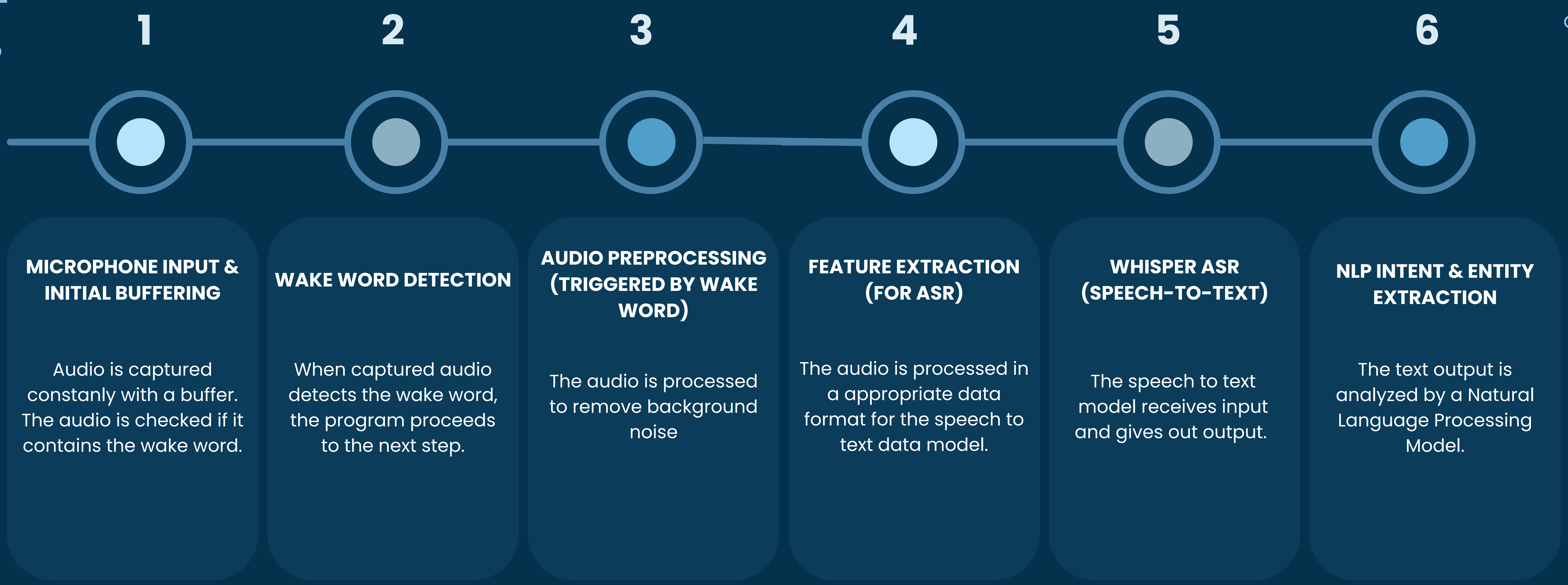
The dataset is expanded to cover multiple different scenarios. For example, background noise such as engine reverberation or rainfall is added to help train the model.

The dataset is reviewed and labeled manually to provide more information to the model. The dataset is also divided into a training set, validation set, and testing set.

A preexisting model is choosen as a foundation to train a new model. The training is executed at this stage.

Once training is completed; evaluation of the model is carried out.

User Workflow



1

MICROPHONE INPUT & INITIAL BUFFERING

Audio is captured constanly with a buffer. The audio is checked if it contains the wake word.

2

WAKE WORD DETECTION

When captured audio detects the wake word, the program proceeds to the next step.

3

AUDIO PREPROCESSING (TRIGGERED BY WAKE WORD)

The audio is processed to remove background noise

4

FEATURE EXTRACTION (FOR ASR)

The audio is processed in a appropriate data format for the speech to text data model.

5

WHISPER ASR (SPEECH-TO-TEXT)

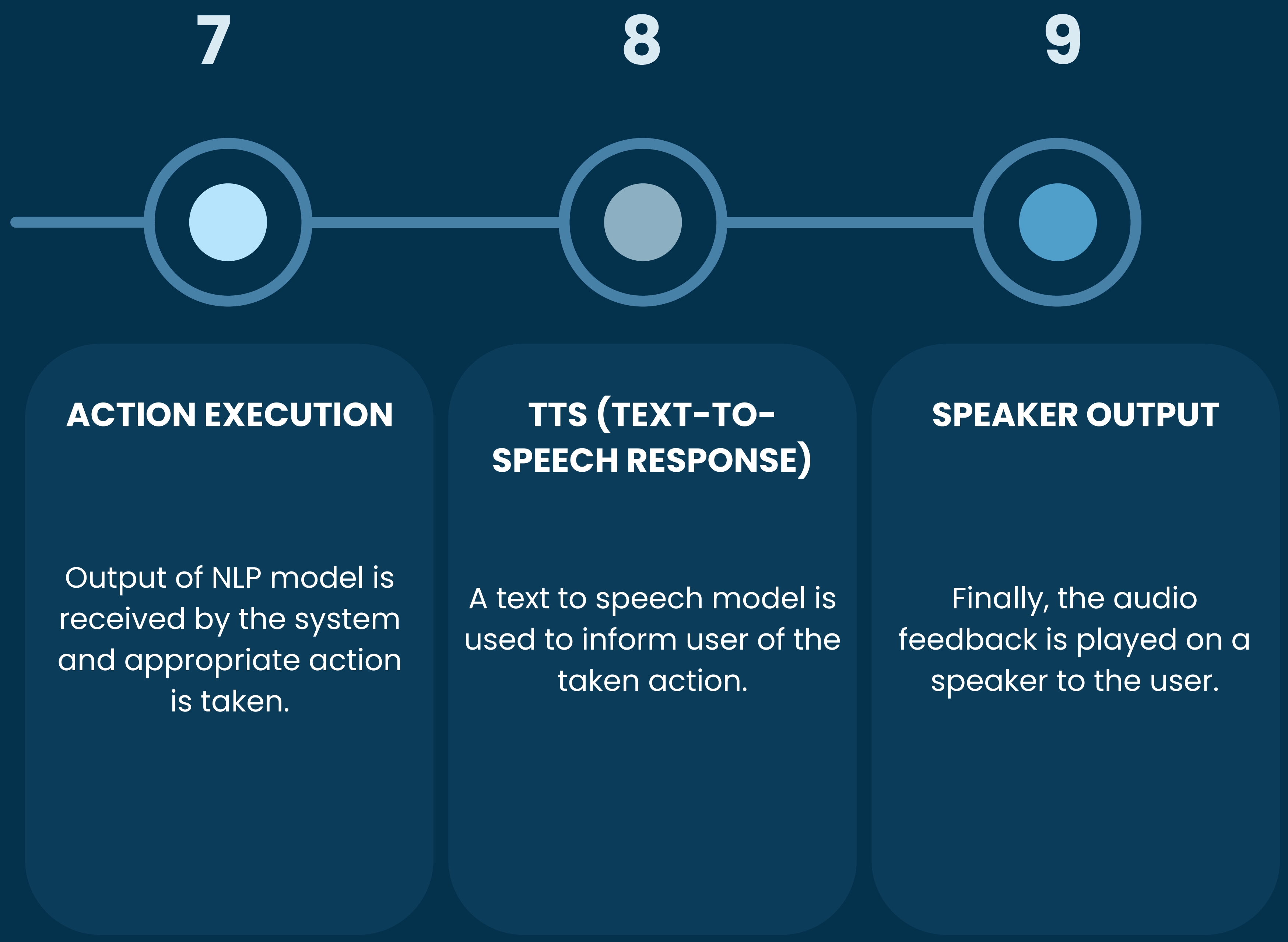
The speech to text model receives input and gives out output.

6

NLP INTENT & ENTITY EXTRACTION

The text output is analyzed by a Natural Language Processing Model.

User Workflow



7

ACTION EXECUTION

Output of NLP model is received by the system and appropriate action is taken.

8

TTS (TEXT-TO-SPEECH RESPONSE)

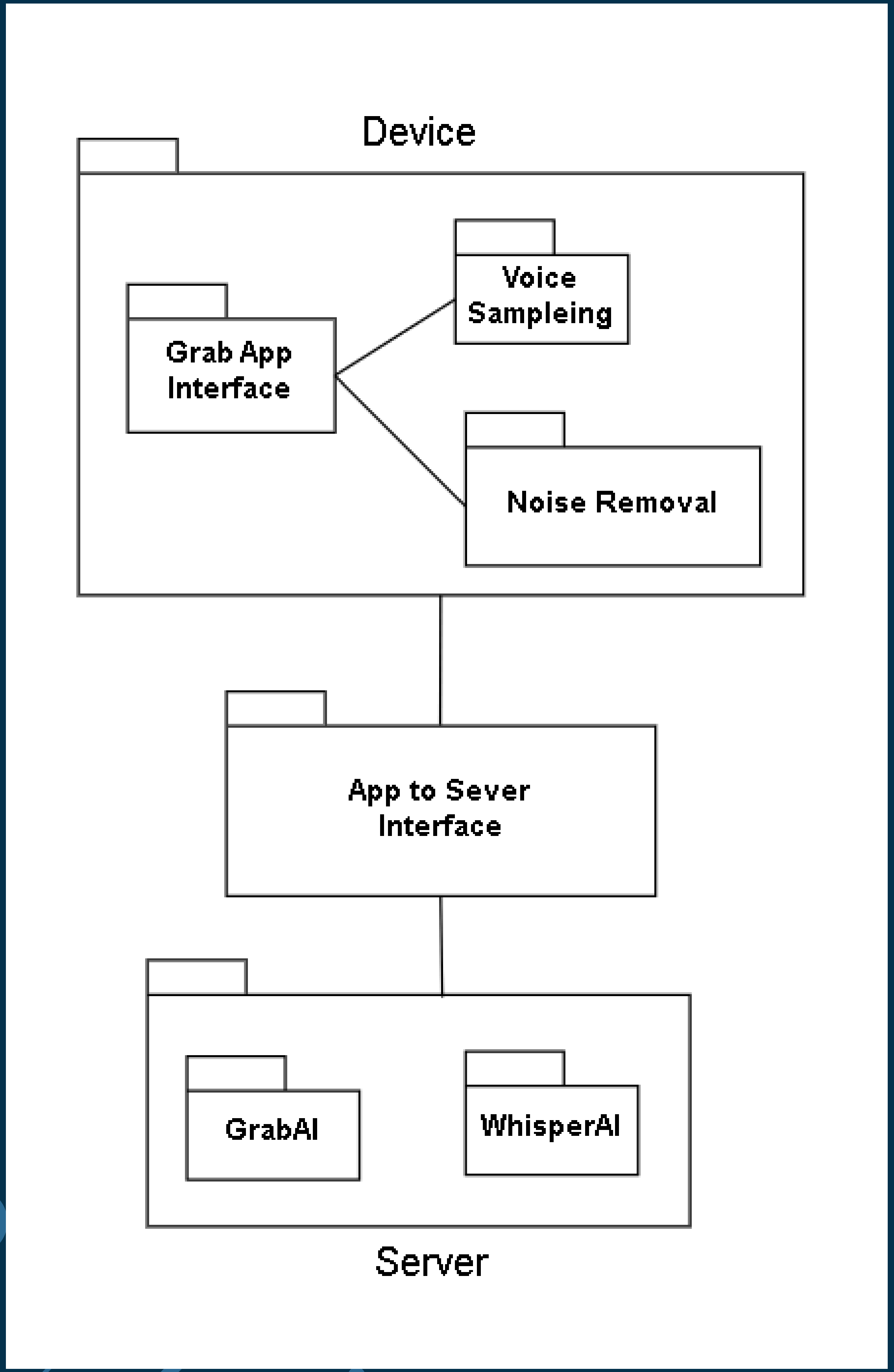
A text to speech model is used to inform user of the taken action.

9

SPEAKER OUTPUT

Finally, the audio feedback is played on a speaker to the user.

Deployment Architecture



Device Side

Audio Detection, Recording & Preprocessing

- **Grab App Interface -** Integrates the voice processing programs
- **Voice Sampling -** Audio capture & Wake Word Detection
- **Noise Removal -** Noise Cleaning & Suppresion

Service Side

Speech-To-Text, Taking Actions

- **Whisper AI -** Performs speech to text
- **Grab AI - T**aking actions after receiving inputs from WhisperAI

Bonus Feature

SPEECH EMOTION RECOGNITION

Enhances the DAX Assistant by detecting the driver’s emotional state through speech through safer and more context-aware interactions.

01 FEATURE EXTRACTION (SER-SPECIFIC)

- Prosodic features: Pitch, Energy ,Speaking rate, Pauses and silences
- Spectral features: MFCCs, Spectral centroid, Spectral bandwidth

02 EMOTION CLASSIFICATION

- The extracted SER features are fed into a trained emotion classification model.

03 OUTPUT

- Detected emotion category or a probability distribution over emotion categories.
- Confidence score indicating the model's certainty in its prediction.

WHY IS IT IMPORTANT?

DRIVER FATIGUE DETECTION

Prompt the assistant to **suggest taking a break** by detecting signs of drowsiness or exhaustion in the driver's voice (e.g., monotonous tone, slow speech rate)

STRESS MONITORING

Identify stress or agitation (e.g., raised pitch, rapid speech) & **offer calming suggestions** or **adjust its communication style** to be more supportive.

WHY IS IT IMPORTANT?

EMERGENCY ALERT

In extreme cases, **trigger automatic alert** to emergency contacts or a dispatch center by detecting panic or distress

PERSONALIZED ASSISTANCE

Provide more **personalized and contextually appropriate responses** by understanding the driver's emotional state

DEMO

Q&A

Thank You