

SpeechPerfect

The Heckler Challenge

Digital Twin Game for Public Speaking Skill Evaluation

Problem Statement

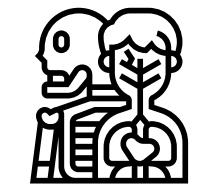


Public speaking is a common challenge face by everyone in various sector.
Many speakers struggle with:



High Anxiety and Stress

Presenter usually experience this during public speaking which can reduce confidence and performance quality.



Lack Awareness of Physiological Responds

Presenter did not aware of how body responds during speaking which limits improvement strategies.



Limited Feedback

systems that focus only on audio metrics (words per minute, filler words) and miss psychological and physiological factors.

Proposed Solution



SpeechPerfect is a **speech assessment system** that uses the **concept of Digital Twin** by creating a **real-time** and use **multimodal** virtual representation of a human speaker that **mirrors speech performance, physiological stress, and behavioral attention**. The **gamified environment** enable simulation, assessment, and later produce a **personalized feedback** of the every speech session.

The system includes:



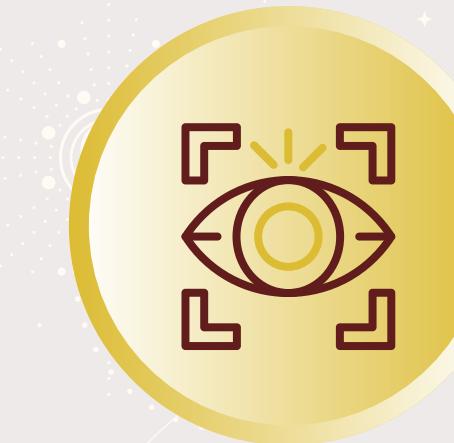
Speech Analysis



AI-Powered
Stress Detection



Physiological
Monitoring



Behavioral
Assessment



Gamified Training
Environment
(Heckler Challenge)

Objective



To develop an **AI-powered and Digital Twin concept speech assessment system** that will evaluate public speaking score and help reducing the Public Speaking Anxiety (PSA).

Project Scope

User Scope



Target User

1. University students
(undergraduate & postgraduate)
2. Person who is involved with :
 - Presentations
 - Oral assessments
 - Public speaking
 - Group discussions

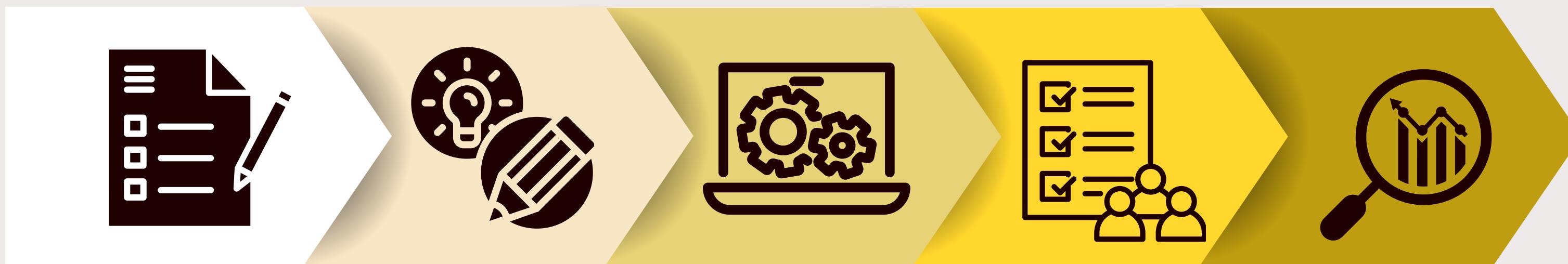
System Scope



Functional Scope

- Multimodal Speech Performance Capture
- Stress Level Detection
- Physiological Monitoring
- Digital Twin Modeling
- Scoring & Feedback System
- Gamified Practice Environment
- Progress Tracking & Persona Classification

Methodology



Planning

Project Design

Development

Testing

Evaluation

Methodology: Planning

Objective: Define requirements, scope, and approach for the multimodal speech assessment system.

1

Problem identification and analysis



Identified gaps in existing system



Analyzed limitations of audio-only approaches



Reviewed literature on public speaking anxiety and multimodal assessment

2

Requirements gathering



Functional:
speech analysis, stress detection, heart rate monitoring, eye tracking, real-time feedback, gamified training



Non-functional:
real-time processing, scalability, accuracy, user-friendly interface



Technical:
AI/ML integration, hardware sensor integration, database storage

3

Project scope definition



Core features:
7 analysis metrics



Integration:
4 modalities



Target users:
Universities students

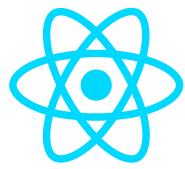
4

Technology selection and research

- AI/ML:** OpenAI Whisper (transcription), Wav2Vec2 (feature extraction), StudentNet (stress detection via knowledge distillation)
- Backend:** Flask (Python), MySQL, Firebase (real-time data)
- Frontend:** React, Vite
- Hardware:** Arduino pulse sensors, webcam for eye tracking
- Audio processing: Librosa, FFmpeg

Methodology: Project Design

Objective: Design system architecture, algorithms, database schema, and user interface.



React

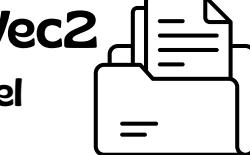
Frontend



Backend API



Databases



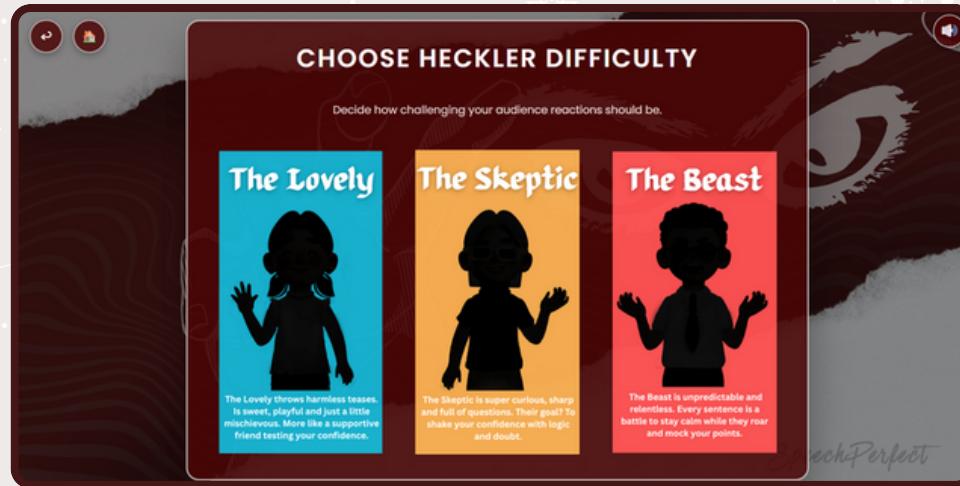
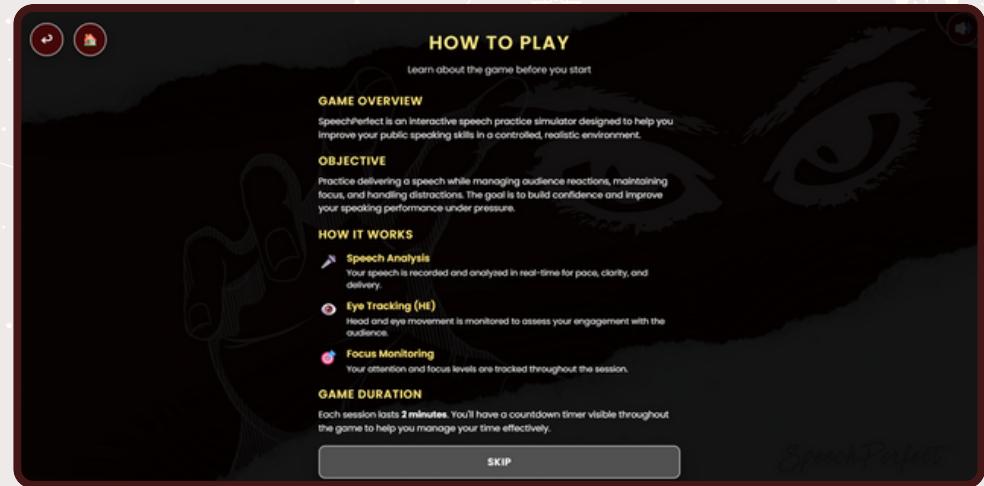
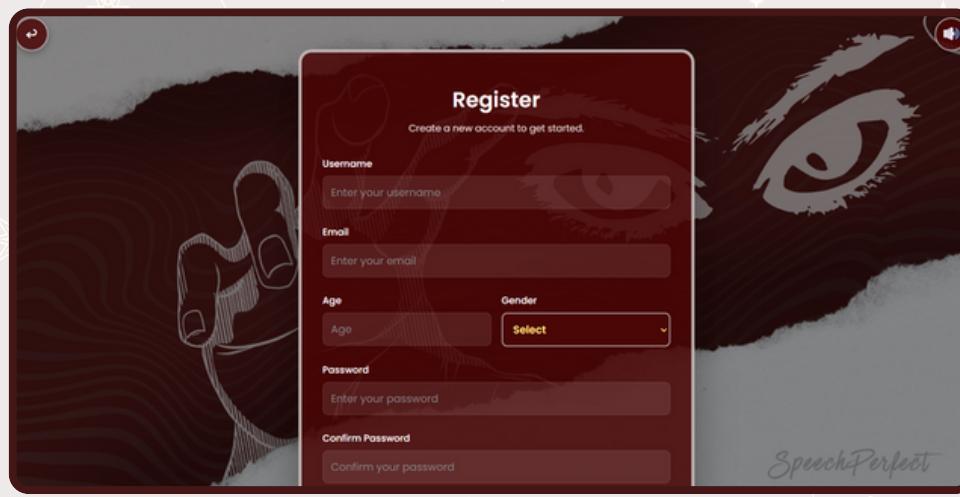
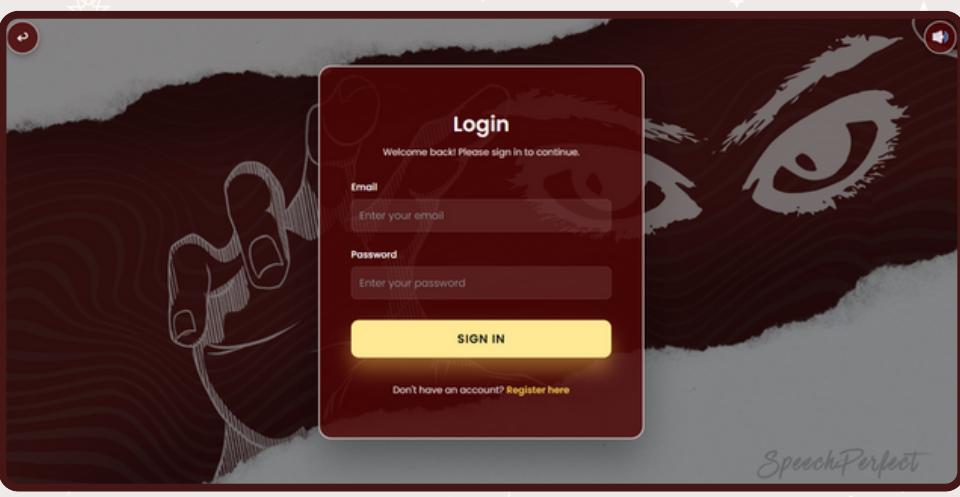
ML Pipeline



Sensors

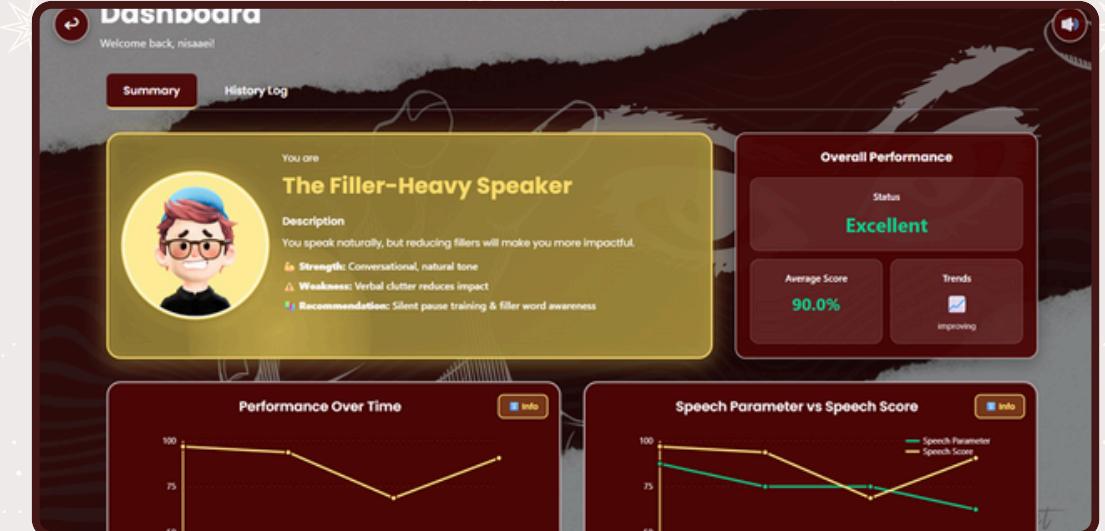
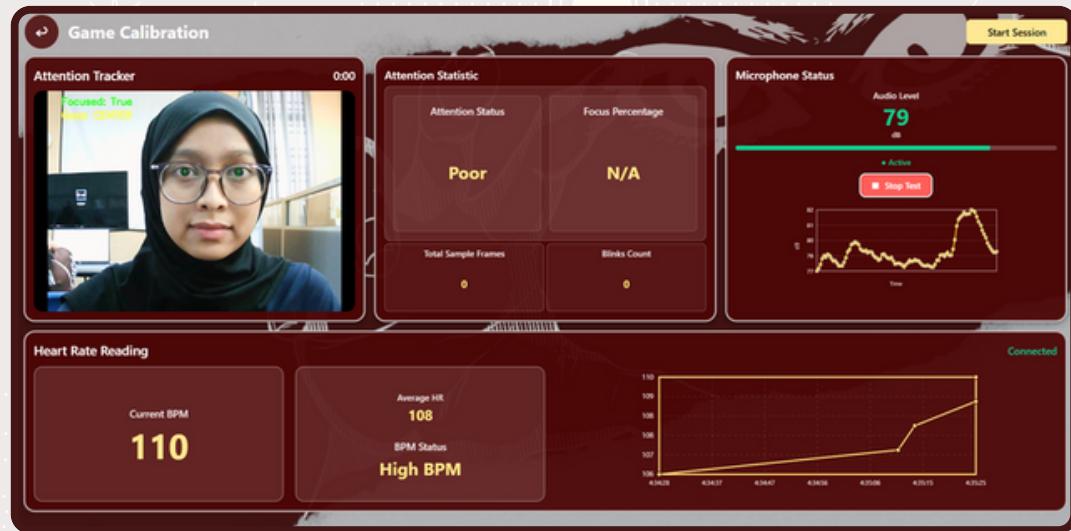
Methodology: Development

User Interfaces



Methodology: Development

User Interfaces



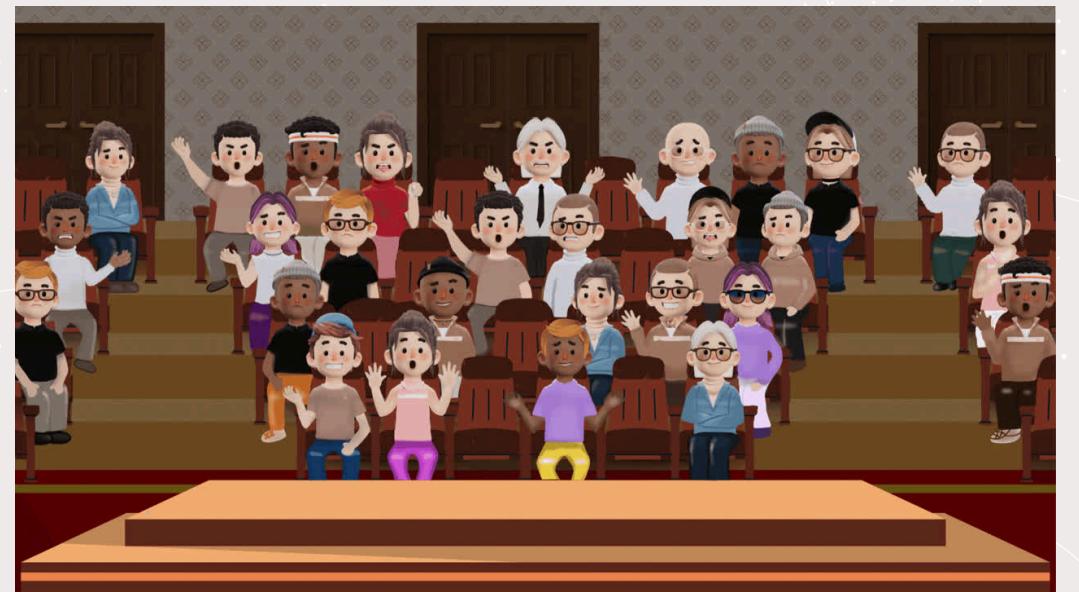
Methodology: Development

Evolution of heckler animation



Methodology: Development

Evolution of game environment



Methodology: Development

Databases



Server: localhost 3306 > Database: humancmt_nd_speechperfect

Table	Action	Rows	Type	Collation	Size	Overhead
tbl_attention		46	InnoDB	latin1_swedish_ci	48.0 Kib	-
tbl_audio		46	InnoDB	latin1_swedish_ci	48.0 Kib	-
tbl_filler		50	InnoDB	latin1_swedish_ci	48.0 Kib	-
tbl_heart_rate		46	InnoDB	latin1_swedish_ci	48.0 Kib	-
tbl_mfcc		50	InnoDB	latin1_swedish_ci	48.0 Kib	-
tbl_pause		50	InnoDB	latin1_swedish_ci	48.0 Kib	-
tbl_session		29	InnoDB	latin1_swedish_ci	64.0 Kib	-
tbl_speech_result		50	InnoDB	latin1_swedish_ci	48.0 Kib	-
tbl_stress		50	InnoDB	latin1_swedish_ci	48.0 Kib	-
tbl_wpm		50	InnoDB	latin1_swedish_ci	128.0 Kib	-
users		30	InnoDB	latin1_swedish_ci	32.0 Kib	-
11 table(s)	Sum	497	InnoDB	latin1_swedish_ci	608 Kib	0 B

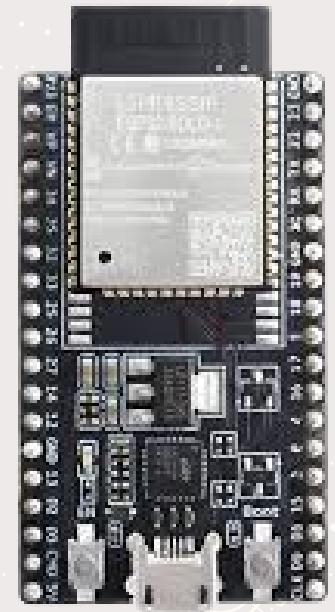
hr_read
bpm: 111

Methodology: Development

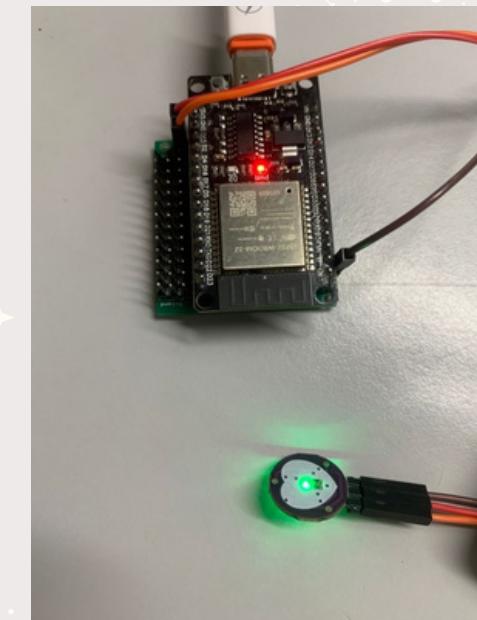
Sensor Connection



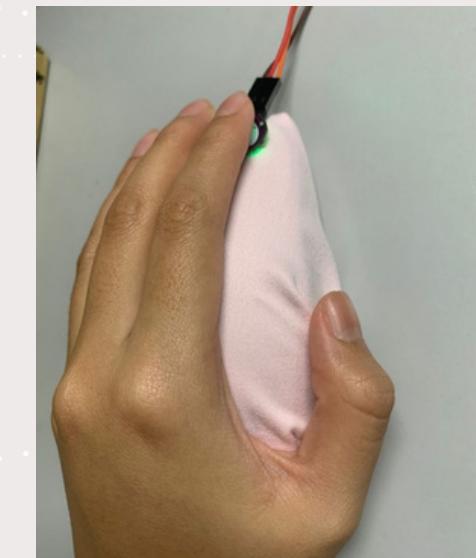
pulse sensor



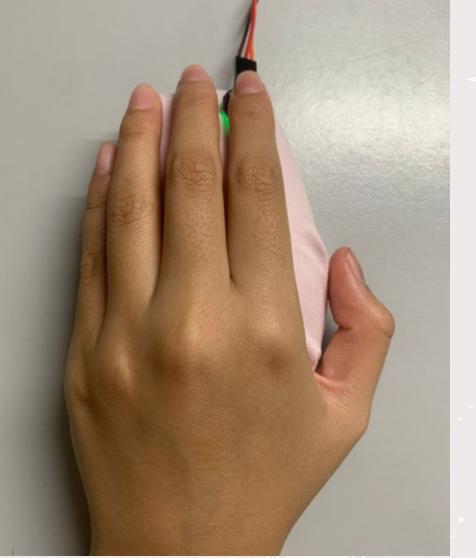
ESP 32



Heart rate
sensor set up

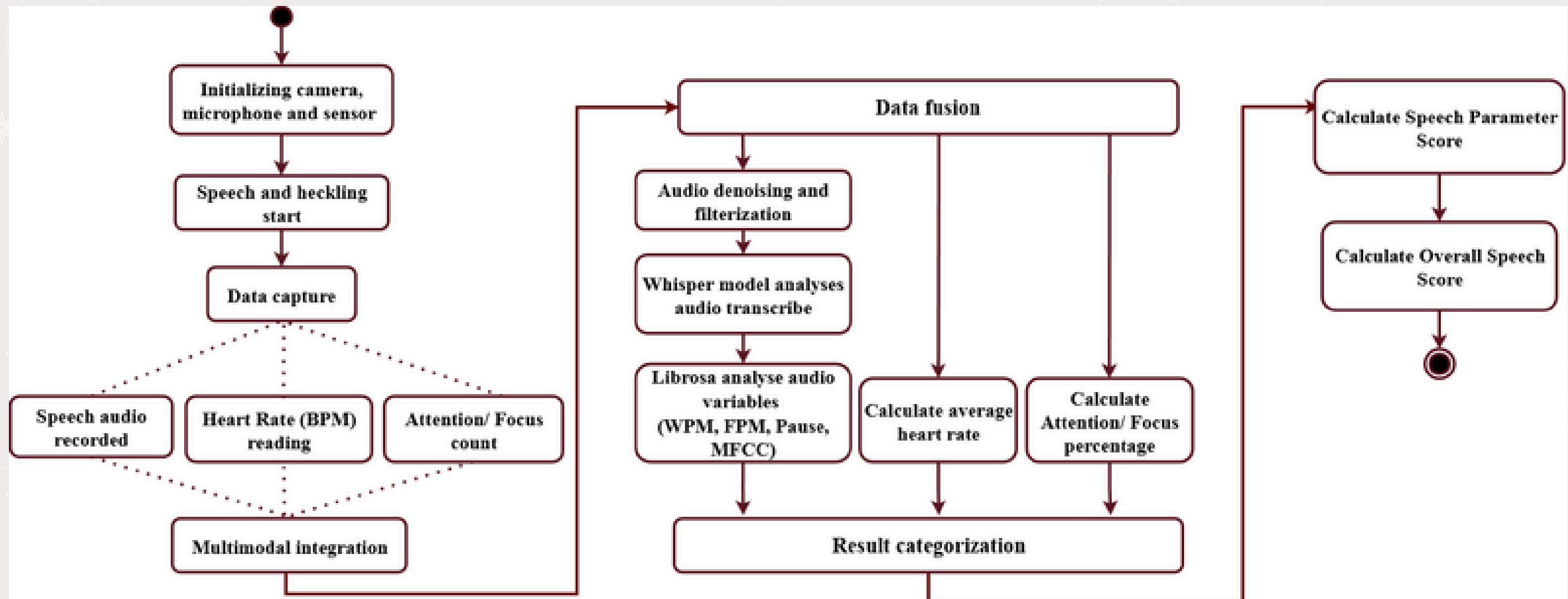


Heart rate sensor with hand rest



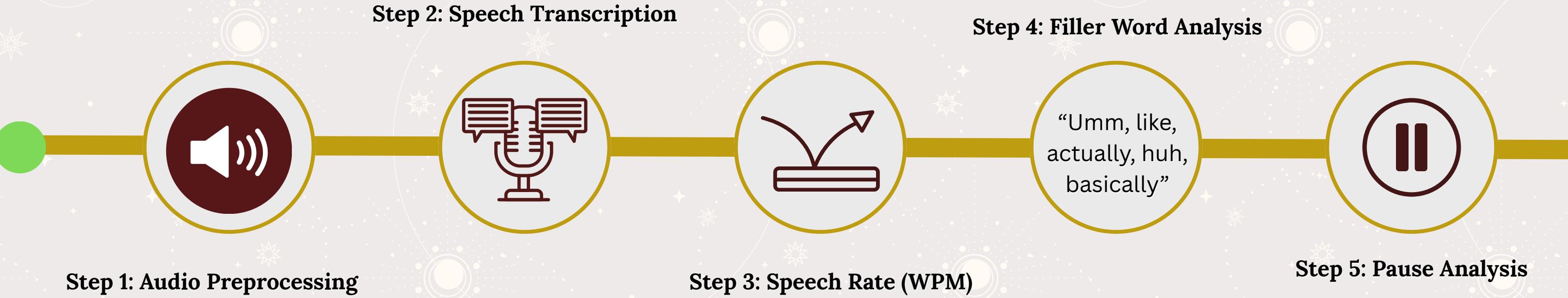
Methodology: Development

Backend Analysis



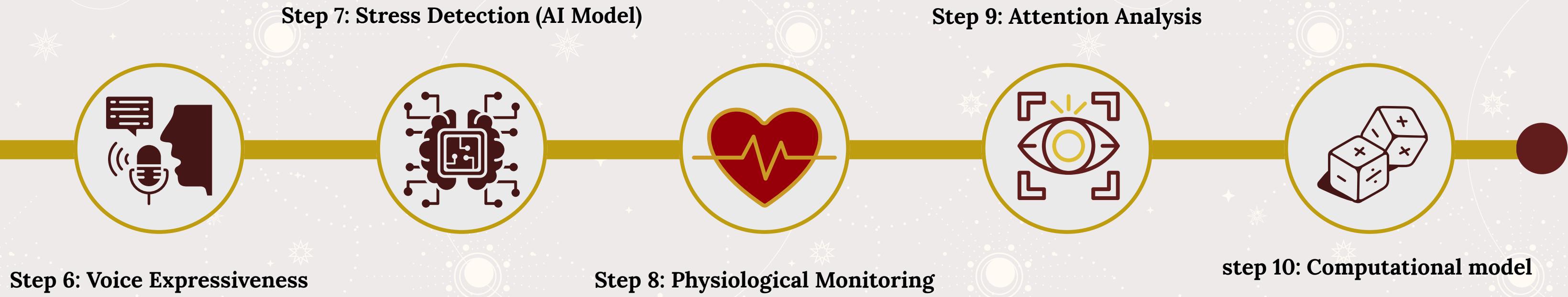
Methodology: Development

Analysis Pipeline



Methodology: Development

Analysis Pipeline



Methodology: Development

- The calculation of speech parameter is based on 4 variables values;
Word Per Minute (wpm), Filler Per Minute (fpm), Pause and MFCC Std. deviation.

Speech Parameter, Sp : $\beta_{wpm} wpm + \beta_{fpm} fpm + \beta_{pause} pause + \beta_{MFCC} MFCC$

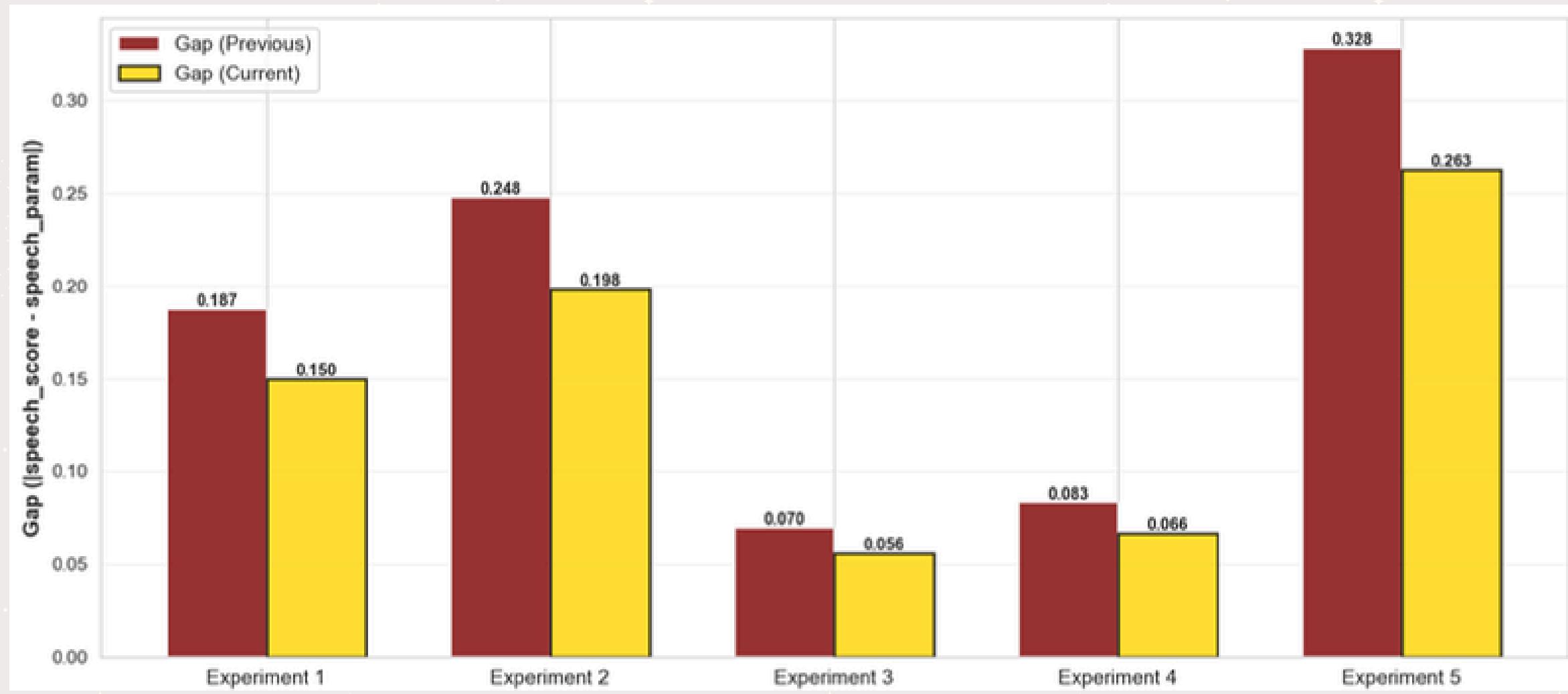
$\beta wpm: 0.25, \beta fpm: 0.25, \beta Pause : 0.25, \beta MFCC : 0.25$

- The calculation of speech score is by combining calculation of **speech parameters (Sp)** and **physiological readings (Heart Rate (HR), Stress and Attention)**.

Speech Score, Sc : $\beta_{Sp} Sp + \beta_{HR} (1 - HR) + \beta_{Stress} (1 - Stress) + \beta_{Attention} Attention$

$\beta Sp: 0.40, \beta HR: 0.20, \beta Stress : 0.20, \beta Attention: 0.20$

Methodology: Development

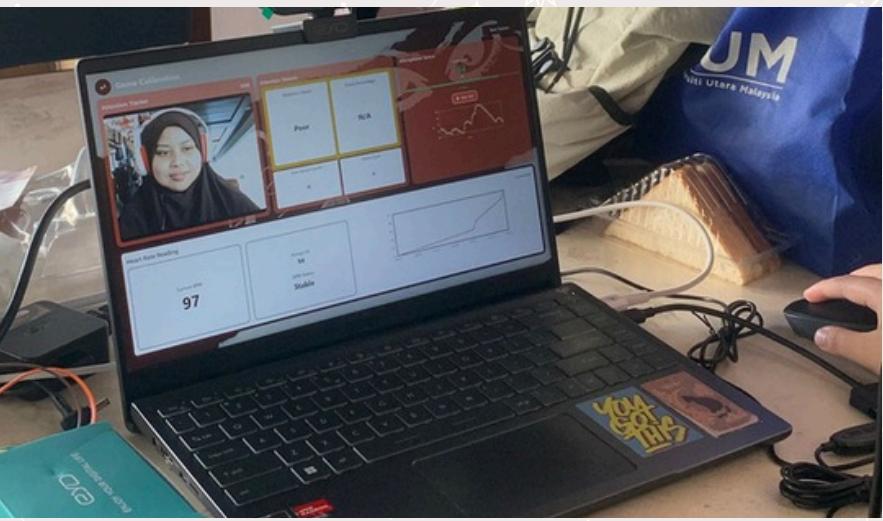


β Sp: 0.25, β HR: 0.25, β Stress : 0.25, β Attention: 0.25

β Sp: 0.40, β HR: 0.20, β Stress : 0.20, β Attention: 0.20

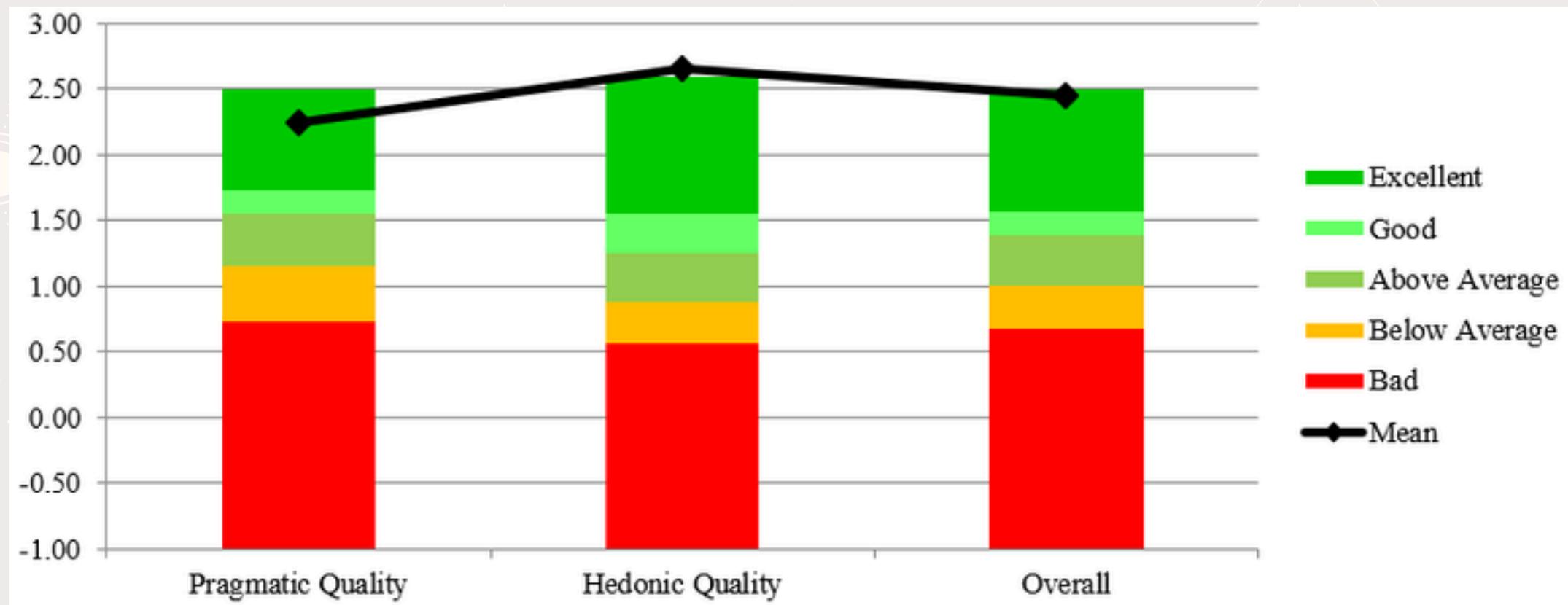
Demonstration

Methodology: Testing



Methodology: Evaluation

User Experience Questionnaire (UEQS) Results



Mean Values

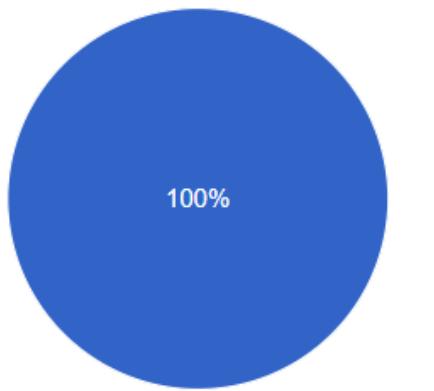
Pragmatic	2.25
Hedonic	2.66
Overall	2.45

- The System is **highly rated for being enjoyable and emotionally satisfying.**
- Its functional performance **effectively supports the engaging experience of the users.**

Methodology: Evaluation

Do you agree with **Speech Parameter Score** that you get ?
Does it correctly represent your public speaking skill ?

16 responses



Yes
No

Do you agree with overall **Speech Score** that you get ?
Does it correctly represent your public speaking skill ?

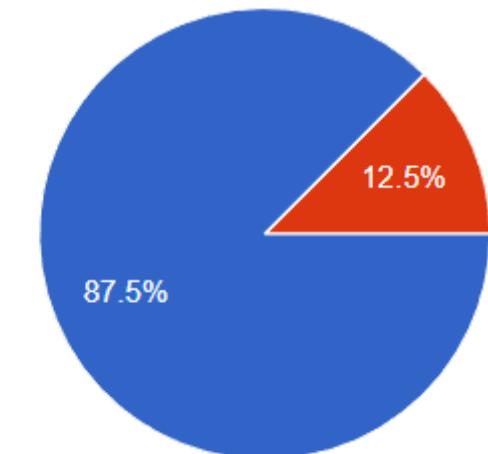
16 responses



Yes
No

Do you feel interrupted by the **Heckler's tricks** and its overall environment ?

16 responses



Yes
No

Novelty



Multimodal Integration

This integration captures both speech and underlying physiological states.

Gamified training environment with contextual stress simulation

Impact



Speech + physiological data

Enable users to understand both speech and how their body responds during presentations.

Leading to better control and startegy for user

Commercialization



Scalable for clubs and Educational purpose

SpeechPerfect can be used by Toastmasters clubs, schools and universities to support speech training.

Future Work



Advanced AI Enhancements & Deployment

Add emotion recognition from audio and facial expressions based on topics

Deployment as a web/ app that can be accessible by everyone

Conclusion

SpeechPerfect demonstrates the feasibility of a **multimodal Digital Twin system for public speaking assessment** targeted at university students. By integrating speech analysis, AI-based stress detection, physiological heart rate monitoring, and attention tracking, the system provides a more holistic and objective evaluation compared to traditional feedback methods.

Plus, the proposed pipeline enables real-time data capturing for analysis of personalized insights and long-term performance tracking within a controlled, gamified environment. Overall, SpeechPerfect has strong potential as an educational tool to help students improve their speaking confidence, delivery and performance under stress.

SpeechPerfect ★
The Heckler Challenge

Thank You So Much