# Preprints.org

**Article**

# Accompanying Odor Management and Emotion Regulation Fragrant Fumigator Based on Image Recognition and Voice Interaction of IoT

Xintong Wen , Chengtian Liang * , Zhaoyu Zhu

*Article*

# Accompanying Odor Management and Emotion Regulation Fragrant Fumigator Based on Image Recognition and Voice Interaction of IoT

**Xintong Wen, Chengtian Liang * and Zhaoyu Zhu**

School of Physics, Hangzhou Normal University, Hangzhou 311121, China

* Correspondence: lct.lctsoft@hotmail.com

**Abstract:** This research designs and develops an intelligent aromatic fumigator based on Internet of Things, which combines image recognition and voice interaction technology to achieve closed-loop control of odor management and emotion regulation. The system uses Raspberry Pi 4B main control platform, integrates camera, microphone and speaker, uses DeepFace model to analyze user's facial expression in real time, provides personalized emotion companion service through voice interaction technology, uses emotion recognition results to dynamically link micro-hole spray module to release customized aroma. After a series of rigorous testing experiments, the device exhibits high comprehensive accuracy and low power consumption in emotion recognition. The device fills the gap of multimodal emotion regulation, combines home environment health management and mental health intervention functions, and provides an innovative solution for IoT smart home applications.

## 0. Introduction

In recent years, with the rapid development of Internet of Things technology and artificial intelligence, smart home devices have gradually evolved from single function to multi-modal interaction. Mental health issues have attracted increasing social attention. World Health Organization data shows that global depression has exceeded 350 million people [1]. However, the emotion regulation needs in the home environment have not been fully met. Traditional aroma fumigator can only passively release odor through preset program, lack of user emotion perception and dynamic response; on the other hand, existing emotional companion devices (such as smart speakers) mostly focus on voice interaction, making it difficult to achieve closed-loop control of emotion regulation through multi-sensory stimulation. Due to this functional disruption, users often have to rely on multiple independent devices to meet their needs, which not only results in fragmentation of the experience, but also greatly reduces the efficiency of use.

In the current research field, emotion recognition technology has achieved considerable progress. For example, Khan et al. [2] proposed a facial expression recognition model based on convolutional neural networks, but this study did not effectively combine it with odor modulation systems; Zhang et al. [3] developed a smart home control system based on speech emotion analysis, but the fusion application of visual modality is missing. In addition, existing smart aromatherapy machines mostly use all-cloud data processing modes (such as Mika smart aromatherapy machine [4]), which poses a lot of risks of privacy leakage. These studies show that existing technologies still have obvious shortcomings in multimodal emotion perception, localized privacy protection and closed-loop regulation.

To solve the above problems, this study proposes a companion intelligent aromatherapy machine based on raspberry pie [5], which innovatively integrates image recognition and voice interaction technology to achieve the closed-loop system of emotion perception, odor regulation and

emotion companion. Through the Raspberry Pi 4B main control platform and localized data processing, while ensuring privacy security, combined with the DeepFace model to achieve high-precision emotion recognition, and linked with the micro-hole spray module to dynamically release customized aroma. This device not only fills the technical gap of multimodal emotion regulation, but also provides innovative solutions for mental health intervention in the field of smart home.

## 1. System Scheme Design

*1.1. System Scheme*

This system is based on the Raspberry Pi 4B main control platform, and uses layered architecture to achieve multi-modal data acquisition, processing and execution control. The visual perception module uses the ESP32-S3 OV3660 camera to obtain the user's facial image in real time. The voice sensing module collects voice commands through the USB microphone. The collected data was processed synergistically with ESP32-S3 via Raspberry Pi 4B. Among them, the Raspberry Pi runs the DeepFace @ OpenCV model to complete emotion recognition tasks and uses the Vosk engine to parse local speech data; The ESP32-S3 mainly assumes the Internet of Things communication (Wi-Fi) function and is responsible for transmitting image data captured by the directly connected OV3660 camera [6] to the Raspberry Pi. The processing results of the Raspberry Pi dynamically trigger the executive layer response: the micro-hole spray module receives PWM signal to control the atomization tablet to release customized aroma, and the 8002A AMP speaker (I²S interface) plays the voice synthesized by text feedback using Kimi API, forming a "Perception-Analysis-Response" closed loop, realizing real-time linkage adjustment between emotion and environment.
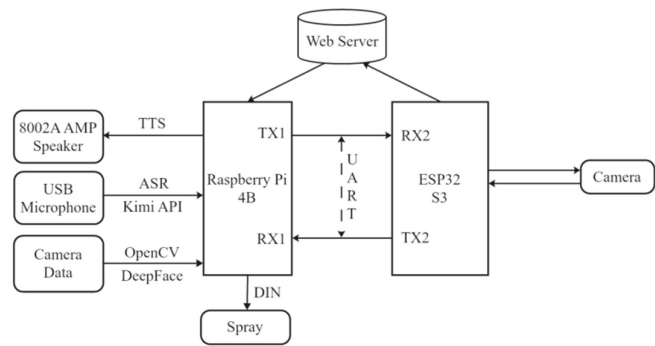


**Figure 1. System Scheme.**

*1.2. Core Functions*

One is emotion perception and odor interaction: capture user's facial expression with camera, then use face recognition [7] and emotion analysis (DeepFace @ OpenCV) to determine user's current emotion state; when negative emotions are recognized, the microporous spray module is triggered through a multi-modal linkage mechanism to release customized aromatherapy with emotion-modulating efficacy (e.g., Bvlgari-Aqva marine aromatherapy when sad/angry). Voice commands can also be used to customize the spray.

Second, emotional exchange and accompanying services: through speech recognition and synthesis, artificial intelligence and other technologies, you can communicate with people to help people alleviate their loneliness and become people's important partners [8]. In addition, it can also provide users with life and spiritual assistance, such as setting reminders, accompanying interactive chats, providing knowledge and telling jokes, etc.

*1.3. Innovation Points*

(1) Odor-emotion closed-loop regulation: This study successfully broke through the limitation of existing aroma fumigators which only have a single diffusion function. By combining emotional feedback from users, the dynamic odor release can be adjusted ingeniously, thus effectively filling the gap in the field of multimodal emotion regulation. At the same time, traditional command-based, on-off passive odor release still supports

(2) Emotional communication and companionship-emotion closed-loop regulation: It has the function of free continuous dialogue, through listening chat interaction, coordinated with scent to regulate users' emotion multimodally.

(3) Private data (expressions, etc.) are processed locally within the LAN to avoid cloud transmission risks.

*1.4. Technical Implementation*

Hardware: (1) Raspberry Pi 4B (Master) (2) Single-channel aromatherapy diffusion module (micro-hole spray) (3) ESP32-S3 + CAM (4) Speaker, microphone

Software: (1) Face-emotion mapping algorithm: DeepFace lightweight model (training set contains 40 million images); (2) Face Recognition Solution: OpenCV (3) Speech Recognition and Synthesis: Multi-language Support, Lightweight Fast Response Edge_tts and Vosk Engine

## 2. Structural Design

The system adopts integrated vertical tower design, the main body is a semi-open cuboid structure with length 20cm, width 12cm, height 27cm (see Figure 2 for effect drawing of physical principle prototype). The system is divided into three functional layers: upper layer, middle layer and lower layer. It supports manual disassembly and maintenance. The top is aromatherapy execution layer, the middle is perception interaction layer, the bottom is core control layer. The rear cover has an open design, allowing the electronic module and the atomization chamber to be removed and cleaned independently.
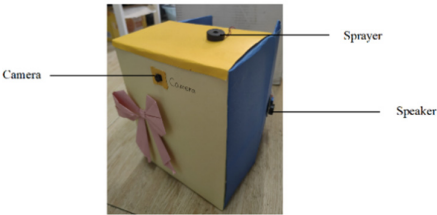


(a) Overall appearance of prototype
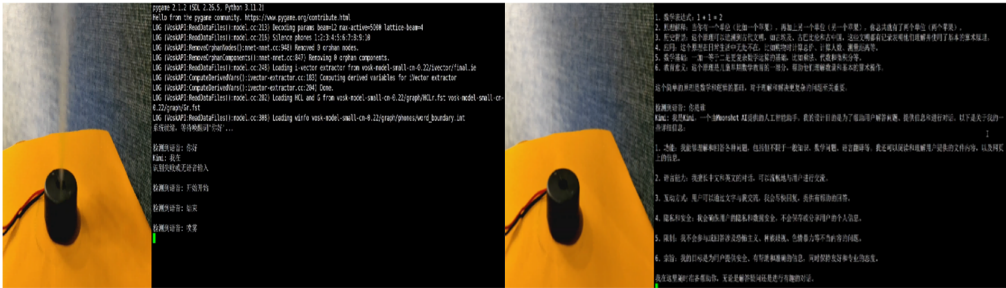


(b) Emotional communication and companionship models (spraying, interactive)
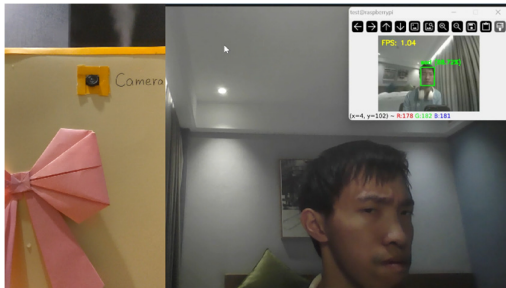
(c) Odor modulation mode of emotion recognition (spraying)

**Figure 2. Effect Drawing of Physical Principle Prototype.**

### 2.1. Design of Perceptual Interaction Layer

The middle module integrates multi-modal sensing devices: OV3660 camera supports 160 ° wide-angle detection and realizes omni-directional face capture; The USB microphone (Figure 3c) is hidden on the inner wall of the middle, and the speaker supporting digital/analog signal control (Figure 3b) is placed on the outer wall of the middle.
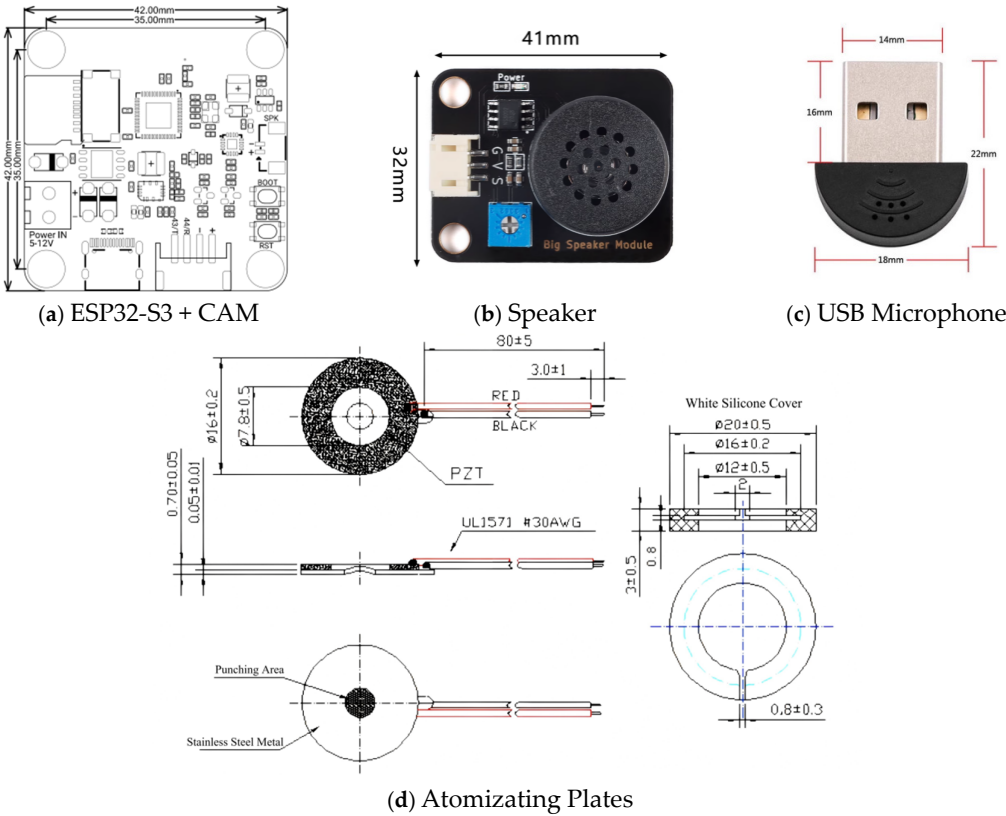


(**a**) ESP32-S3 + CAM      (**b**) Speaker      (**c**) USB Microphone



(**d**) Atomizing Plates

**Figure 3. Module Design (Dimensionally) Drawing.**

### 2.2. Design of Core Control Layer

The bottom module contains the main control unit: Raspberry Pi 4B and ESP32-S3 + CAM (as shown in Figure 3a, see Figure 4 for the schematic diagram of this part) fixed on the left inner wall, and connected through USB, GPIO interface and its built-in WiFi module expansion components; a stand-alone MPU chip runs lightweight DeepFace @ OpenCV, Vosk @ KaldiRecognizer model and Kimi API, dedicated to real-time sentiment analysis, speech recognition and companion interaction, respectively.
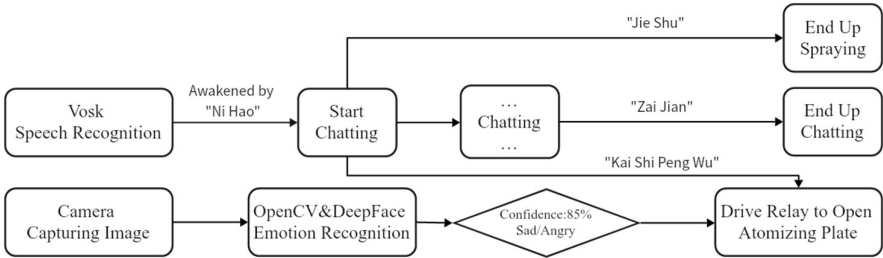
**doi:10.20944/preprints202505.1064.v1**

5 of 7

**Figure 4.** Software program flow chart.

### 2.3. Aromatherapy Executive Layer Design

The top module contains piezoelectric atomizer (Figure 3d) and PWM control drive board, with spray accuracy of ± 1 ~ 5μm. Transparent PET liquid storage bottle (550 mL capacity) can be replaced by quick insertion and removal. Bottle body scale marks are easy to observe remaining amount. When a high frequency voltage is applied to the piezo-atomizer, it begins to vibrate. The vibrations cause small droplets to form on the surface of the water and break them up to form a fine mist, which is more easily evaporated and diffused in the air, thereby providing a more effective atmosphere diffusion effect.

## 3. Structural Design

### 3.1. Perceptual Interaction Layer Design

This section adopts emotion recognition architecture based on deep learning. The visual recognition module is equipped with high-sensitivity OV3660 camera and optimized OpenCV image processing pipeline, which can capture user facial micro-expressions at a steady rate. Using the lightweight DeepFace model trained by large-scale datasets, 85% emotion classification accuracy was achieved on embedded platform.

### 3.2. Closed-Loop Aromatherapy Control Segment

The closed-loop aromatherapy regulation system was innovatively constructed, which was integrated with multi-modal emotion recognition segment, emotional communication and companion segment to achieve adaptive aromatherapy release. The atomization unit is equipped with a high-precision PWM control drive board connected to the Raspberry Pi GPIO port. The Raspberry Pi's GPIO interface [9] is shown in Figure 5. Accurate flow control of ±1 mL can be achieved by programming analog inputs. In emotion regulation mode, according to recognition result dynamically adjust aroma opening and closing; in emotional communication and companion mode, provide preset passive aroma on/off commands. The end-to-end delay of all control commands is controlled within 1 second, ensuring real-time user experience.
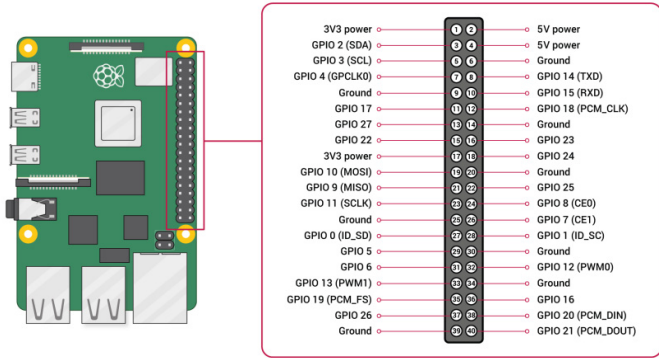
**Figure 5. GPIO interface for Raspberry Pi.**

*3.3. Emotional Communication and Accompaniment*

Speech analysis module integrates Vosk local speech recognition engine, which can keep stable speech quality even in the environment noise; Speech Synthesis Edge _ tts supports Mandarin and multiple dialects, making users more familiar. This paragraph also calls the Kimi API to achieve content-and context-controlled free dialogue [10]. As the number of rounds of communication increases, emotional communication and companionship will also appear more personalized and customized for each user.

*3.4. Emotional Communication and Accompaniment*

The system exhibits excellent performance under standard test conditions (25 ± 2 ℃, 60 ± 5% RH). The system maintained stable operation during a continuous 24-hour pressure test, with the core temperature maintained in the ideal range of 50-55 °C. In terms of power consumption, 4.6 W is typical operation and 8.2 W peak power consumption does not exceed. After optimizing the consumption rate of aromatherapy liquid, the standard 550 mL volume lasted 15-21 days. The system reserves several expansion interfaces by modularization design, which provides sufficient space for subsequent function upgrade.

## 4. Structural Design

*4.1. Experimental Setup*

The test group included 26 students and teachers. Each completed 3 emotion recognition experiments and 3 speech recognition experiments. A total of 156 valid data samples were collected. Hardware configuration adopts Raspberry Pi 4B (4GB RAM) as the main control platform, carries OV3660 camera (1080P), USB microphone and speaker. The software environment is based on Raspbian system, which integrates OpenCV 4.11, DeepFace 0.7 and Vosk 0.3. 45. During the testing process, the system resource usage, including CPU utilization, memory usage and other key indicators were recorded synchronously.

*4.2. Experimental Results and Analysis*

The experimental data show that the system achieves high comprehensive emotion recognition accuracy under standard test environment. Specifically, happy state had the highest recognition accuracy (92.3%), sad state (88.5%) and angry state (80.8%) second.

**Table 1. Comparison of recognition performance of typical emotional states.**

| Emotion Type | Visual Recognition Rate | Speech Recognition Rate | Comprehensive Recognition Rate |
|---|---|---|---|
| Happy | 88.5% | 96.2% | 92.3% |
| Sad | 84.6% | 92.3% | 88.5% |
| Angry | 80.8% | 88.5% | 80.8% |

During the 48-hour stability test, the system memory footprint was maintained within 0.7-1.1 GB, and the CPU temperature was maintained at 51 ± 3 ℃. By analyzing the wrong recognition cases, the main error sources are: extreme profile angle leading to missing features (65%); and color distortion under complex lighting conditions, accounting for 35% of erroneous samples. To solve the above problems, follow-up studies will focus on the following aspects: First, increase the layout of multi-angle cameras to improve the integrity of feature acquisition; secondly, a deep learning noise

reduction algorithm is used to improve the robustness of speech feature extraction; finally, add sensor array to enhance the system-environment linkage control ability.

This experiment verifies the practical value of the system in real scenes. Especially in home office scenarios, the system can effectively identify emotional stress states and trigger corresponding aromatherapy switching and emotional communication and companionship regulation schemes. Tester feedback showed that 81% users considered emotion regulation effect significant. These empirical data provide clear directions for further optimization of the system [11].

## 5. Concluding Remarks

This study successfully developed a companion odor management and emotion regulation aromatherapy machine based on image recognition and voice interaction, and verified its excellent performance in emotion recognition and voice interaction through simple experiments. The system integrates the visual emotion analysis and voice interaction technology innovatively. The hardware design adopts modular architecture, which has both function expansibility and maintenance convenience. This machine effectively realizes emotion recognition and regulation, human-computer voice interaction (emotional communication and companionship) based on Kimi API, which has family practicability and market value.

## References

1. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates [R]. Geneva: WHO, 2017.
2. Khan A, Sohail A, Zahoora U, et al. A survey of the recent architectures of deep convolutional neural networks [J]. Artificial Intelligence Review, 2020, 53 (8): 5455-5516.
3. Yi-Lin Lin and Gang Wei. Speech emotion recognition based on HMM and SVM [C]//2005 International Conference on Machine Learning and Cybernetics, Guangzhou, China, 2005: 4898-4901.
4. Mi family intelligent aromatherapy machine user instruction [EB/OL]. [2025-04-17]. http://home.mi.com/views/introduction.html? model = xiaomi.diffuser.xw002.
5. SMITH S. Raspberry Pi assembly language programming [M]. Appress, 2019.
6. Wang Hong, Wang Xiao-bo, Yao Shuai. Design of pre-inspection and pre-inspection system for agricultural products based on ESP32-CAM [J]. Internet of Things Technology, 2025, 15 (02): 50-53.
7. Liu Chun-xiao, Zhang Wen-hao.Design and implementation of face recognition system based on OpenCV [J]. Modern Information Technology, 2024, 8 (14): 20-25.
8. Li Xuelin.A survey of speech recognition technology based on human-computer interaction [J]. Electronic World, 2018, (21): 105.
9. Tan Rui-H, Liu YJ, Liu Hang, et al.Intelligent garbage sorting system based on raspberry pie [J]. Internet of Things technology.
10. Wang QH, Lin JP. Application of Voice Cloud Platform in Smart Home [J]. Information Computer (Theoretical Edition), 2024, 36 (09): 118-120.
11. LIU Yun, SU Jing-jing, ZOU Fu-min.Intelligent water dispenser design based on Internet of Things cloud platform [J]. Internet of Things Technology, 2025, 15 (07): 94-98 +102