

# Bilibili Video Playback Assistant

---

- 2152598 郑皓予
- 2151136 朱开来
- 2153174 陈华机

## 0 Content

---

### Bilibili Video Playback Assistant

#### 0 Content

#### 1 Project Description

- 1.1 Project Background
- 1.2 Function Description
  - 1.2.1 Speech Recognition System
  - 1.2.2 Gesture Control Mouse
  - 1.2.3 Gesture-based Shortcuts Implementation
- 1.3 Structures and Modules
  - 1.3.1 User Interface Layer
  - 1.3.2 Control Logic Layer
  - 1.3.3 Business Logic Layer
  - 1.3.4 Integration Layer

#### 2 The Implemented Requirements

#### 3 Advantages and Disadvantages

- 3.1 Advantages
- 3.2 Disadvantages

#### 4 How to Improve

#### 5 Contribution

## 1 Project Description

---

### 1.1 Project Background

---

In today's world, it has become common for young people to multitask by watching videos while eating. However, this activity can pose challenges when it comes to operating a computer or navigating video websites. Factors such as wearing gloves or having greasy hands make it inconvenient to use a traditional mouse or keyboard for video playback.

In response to this practical scenario, our team has developed an innovative solution. We have designed an application that leverages a combination of speech recognition and gesture recognition technologies. This application enables users to control video playback without the need for direct interaction with a mouse or keyboard.

### 1.2 Function Description

---

Our program is an innovative video playback assistive system that combines speech recognition and gesture recognition technologies. It aims to provide a seamless and convenient experience for users to interact with Bilibili videos without the need for traditional mouse and keyboard inputs.

## 1.2.1 Speech Recognition System

In the speech recognition part, the program allows users to control various functions through voice commands. Users can wake up the program and open the Bilibili website by simply speaking. They can also navigate to their own dynamic section or terminate the program at any stage using voice input.

## 1.2.2 Gesture Control Mouse

Users can control the mouse cursor by moving their index finger, and perform actions such as clicking the left mouse button by merging their index and middle fingers. This allows them to adjust the video's progress, pause or resume playback, switch videos, like content, and follow other users.

## 1.2.3 Gesture-based Shortcuts Implementation

Additionally, the gesture recognition part enhances the program's functionality by enabling users to control video playback through intuitive hand movements. For instance, users can adjust the volume by changing the distance between their thumb and index finger. Increasing the distance raises the volume, while decreasing it lowers the volume.

The program also offers the ability to scroll through web pages using specific gestures. Users can swipe upwards with their thumb extended to scroll up, while extending all fingers except the thumb will scroll down, providing a convenient way to browse through content.

Furthermore, the program enables users to capture screenshots by performing a specific gesture using both hands.

Overall, our program revolutionizes the way users interact with Bilibili videos by integrating speech recognition and gesture recognition technologies. It offers hands-free and intuitive controls, enhancing accessibility and convenience while watching videos and engaging with the Bilibili platform.

# 1.3 Structures and Modules

---

Our project consists of the following structures:

## 1.3.1 User Interface Layer

Responsible for interacting with the user. It includes our User Interface Module, which consists of the program startup interface, main interface, and help interface.

## 1.3.2 Control Logic Layer

Includes the Speech Recognition Module and Gesture Recognition Module. The Speech Recognition Module converts user's voice commands into executable instructions, while the Gesture Recognition Module identifies user's hand gestures and translates them into corresponding control commands.

## 1.3.3 Business Logic Layer

Processes instructions from the Control Logic Layer and performs corresponding operations. It includes functionalities such as opening websites, navigating to specific sections, and adjusting volume.

### 1.3.4 Integration Layer

- **Speech Recognition Integration:** Connects to the Baidu Speech Recognition API to convert voice commands into text.
- **Gesture Recognition Integration:** Integrates with gesture recognition libraries or frameworks to interpret hand gestures.
- **Website Interaction:** Enables the program to interact with the Bilibili website, such as opening the Bilibili main page and dynamic page.

These structures ensure a systematic organization of our project, allowing for clear separation of responsibilities and smooth communication between different components.

## 2 The Implemented Requirements

---

This project is targeted towards Bilibili video platform users, especially young users who enjoy watching videos while multitasking. The aim of this project is to provide a convenient way for users to control Bilibili videos without the need to directly interact with a mouse or keyboard in specific scenarios.

The specific **requirements from users** for this project are as follows:

- Ability to open the Bilibili web version using speech recognition technology and perform related operations.
- Capability to control videos through gesture recognition, including volume adjustment, playback control, and page scrolling.
- Implementation of a screenshot feature through gesture recognition.
- User-friendly interface and simple operation, allowing for hands-free interactions.

To meet these user requirements, the project **implements the following functionalities:**

- Opening the Bilibili web version through speech recognition, enabling users to directly access the platform.
- Using gestures to adjust the volume, allowing users to control the video's audio level based on gesture movements.
- Controlling the video playback, progress bar, liking, and following features through gesture-based mouse control.
- Enabling page scrolling through specific gestures, facilitating convenient browsing of additional content.
- Providing a screenshot feature that allows users to capture their screen through hands-on interaction.

The speech recognition and gesture recognition components highlight the user interaction and user-friendly aspects of the project. Users can control video playback through voice commands and natural hand gestures, eliminating the need for direct keyboard or mouse interactions and providing a more intuitive, convenient, and comfortable experience. Additionally, the design of the user interface and implementation of feedback systems aim to enhance user-friendliness, ensuring that users have a clear understanding of their actions' outcomes and allowing for adjustments and control as needed.

## 3 Advantages and Disadvantages

---

### 3.1 Advantages

---

- **Convenience:** The program provides a convenient way for users to interact with Bilibili videos without the need to directly touch a mouse or keyboard. Users can control video playback, adjust volume, and perform other actions using voice commands and gestures, enabling hands-free operation.
- **Multitasking:** Our program caters to the needs of users who enjoy multitasking while watching videos. Users can simultaneously eat, wear gloves, or have oily hands while still being able to control the video playback easily through voice and gesture recognition.
- **Enhanced User Experience:** By utilizing speech and gesture recognition technologies, our program offers an interactive and user-friendly experience. Users can enjoy a seamless and intuitive interaction with Bilibili videos, making it more engaging and enjoyable.
- **Accessibility:** The program accommodates users with different abilities or physical limitations. It eliminates the need for precise mouse movements or complex keyboard commands, making it accessible to a wider range of users.
- 

## 3.2 Disadvantages

---

- **Dependency on Recognition Accuracy:** The effectiveness and usability of our program heavily rely on the accuracy of speech and gesture recognition. Inaccurate recognition or misinterpretation of user commands may lead to undesired actions or frustrations for the users.
- **Learning Curve:** Users need to become familiar with the specific voice commands and gestures required to operate the program effectively. This may involve a learning curve and initial adjustments for users who are new to the program.
- **Hardware and Software Requirements:** The program may require specific hardware components, such as cameras or sensors for gesture recognition, and compatible software libraries or APIs. Users without the necessary hardware or software support may not be able to utilize the program fully.
- **Limited Functionality:** While our program offers various functionalities such as video playback control and volume adjustment, it may not cover all the features and capabilities available on the Bilibili platform. Some advanced or specific functions of Bilibili may not be accessible through our program.

Overall, the advantages of convenience, multitasking support, enhanced user experience, and accessibility make our program beneficial for users. However, challenges related to recognition accuracy, learning curve, hardware requirements, and limited functionality should be considered when using the program.

## 4 How to Improve

---

- **Expand Supported Commands and Gestures:** Increase the range of recognized voice commands and gestures to provide users with more control options. Conduct user research and gather feedback to identify commonly desired commands and gestures and incorporate them into the program. This will make the program more versatile and adaptable to individual user preferences.
- **Provide Customization Options:** Allow users to customize their own voice commands and gestures. This feature can enhance the user experience by accommodating personal preferences and making the program more personalized.
- **Implement User Training and Onboarding:** Provide a guided training process or onboarding tutorial to familiarize users with the program's functionalities and how to effectively use voice commands and gestures. Clear instructions, tooltips, or interactive demos can help users quickly grasp the program's capabilities and improve their proficiency.

- **Support Multiple Platforms and Devices:** Extend the program's compatibility to work across different platforms and devices, such as mobile devices, smart TVs, or virtual reality headsets. This will enable users to access and control Bilibili videos on their preferred devices, expanding the program's reach and usability.
- **Improving recognition accuracy:** Adopting more precise algorithms for speech recognition and gesture recognition. Conducting targeted optimizations in the existing codebase to enhance performance for the specific application of this project.
- **Gather User Feedback:** Actively collect user feedback and suggestions to understand their needs and expectations. Feedback can be collected through surveys, user forums, or in-app feedback mechanisms. Incorporate valuable user insights into future updates and feature enhancements.

By implementing these improvements, our program can become more accurate, versatile, user-friendly, and adaptable to different user preferences and scenarios.

## 5 Contribution

---

**The team members contribute equally.**

**陈华机:** Collaboratively establish project outline; search for relevant project materials; implement gesture recognition module.

**朱开来:** Collaboratively establish project outline; implement speech recognition module; UI interface design; modify PPT.

**郑皓予:** Integrate code from different modules; design UI interface; write REPORT document; create PPT.