

YouTube - Fusion

Fusion Interaction

David Palricas nº 108780
Henrique Coelho, nº 108342

Multimodal Interaction - TP1
Master's in Digital Game Development
Master's in Informatics Engineering
17/01/2025



universidade
de aveiro




Introduction

- **Fusion-Controlled YouTube Assistant** combines **Gesture** and **Voice** control to transform the user experience on YouTube. Leveraging the **Kinect Sensor & Kinect Studio** for capturing gestures and refining them with **Visual Gesture Builder**, alongside **Rasa** for natural language understanding and the **IEETA Multimodal Framework** for speech processing, we enable intuitive **multimodal interaction**. By integrating **Selenium/WebDriver**, the assistant automates various YouTube functionalities, providing seamless hands-free control of the platform.
- Through this presentation, we will talk about:
 - **Gestures & Voice Functionalities**
 - **Fusion - Redundancy & Complementary Interactions**
 - **Architecture**
 - **Demo - Fusion**
 - **Improvements**
 - **Demo - Improvements**
 - **Conclusion**



Gestures & Voice Functionalities

Gestures	Voice
<ol style="list-style-type: none">1. Like2. Dislike	<ol style="list-style-type: none">1. Subscribe Channels2. Save Videos in playlists3. Share Video (Whatsapp)4. Activate/Deactivate Subtitles5. Play/Stop Videos6. Activate/Deactivate Sound7. Seek Forward and Seek Backward8. Search Videos9. Video Speed Adjustment10. Write Comments11. Help Command (Fusion, Gestures & Voice)12. Shutdown the Assistant



Fusion - Redundancy & Complementary Interactions

Redundancy

FullScreen

Join your hands in front of your chest and stretch your arms diagonally keeping your hands closed, when you can't stretch them anymore open your hands.

Or say “Ecrã cheio por favor”

NormalScreen

Stretch your arms diagonally until you can't anymore, with your hands open and slowly bring your arms to your chest, when your hands are in front of your chest join them.

Or say “Ecrã normal por favor”



Fusion - Redundancy & Complementary Interactions

Complementary

Volume UP

Extend one of your arms to the side at your waist level and raise it up to your head.
And say something like “Aumenta um pouco o volume”

Scroll UP

Stretch one of your arms forward and extend the other downward, with your palm open.
Slowly raise it to your chest.
And say “Dá scroll um pouco para cima”

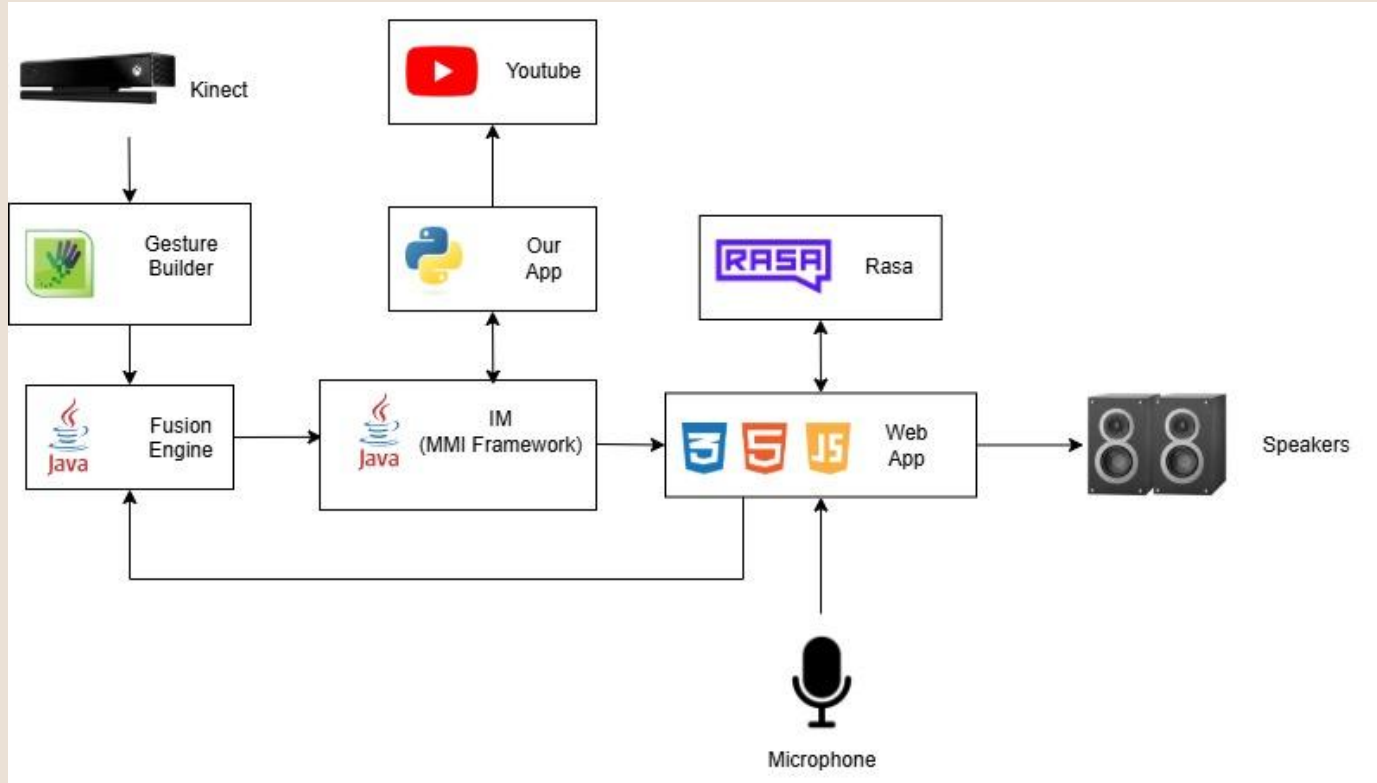
Volume DOWN

Stretch your arm up to your head and lower it to the side, slowly until it reaches your waist level.
And say something like “Diminui o volume do vídeo”.

Scroll DOWN

Stretch one of your arms forward and extend the other upward, with your palm open. Slowly lower it to your chest.
And say “Dá scroll para baixo”

Architecture





Demo

Fusion

Improvements

Voice

1. Share Video (Whatsapp) - **Now w/ Message Confirmation**
2. Help Command - **(Fusion, Gestures & Voice) w/ speech examples | gestures description**
3. Shutdown the Assistant - **Now w/ Confirmation**
4. **Excessive** Feedback Removal



Demo

Improvements



Conclusion

In conclusion, while the Fusion project successfully integrated both gestures and voice for a hands-free YouTube experience, challenges persisted with gesture recognition. Combining voice control significantly improved the overall usability and reduced reliance on gestures alone. For instance, redundancy was implemented for functionalities such as FullScreen and NormalScreen, allowing users to perform the same actions using either gestures or voice. Complementary interactions, such as Volume Up, Volume Down, Scroll Down, and Scroll UP, leveraged both modalities to support each other, enhancing user control and adaptability.