IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Accentify

[1]Ms. Siddu Tushara M S , [2]Bindu K Reddy, [3]Nandalur Bhagyasri, [4]Preethi P, [5]Yuvaraj K

[1]Assistant Proffesor, [2]Student, [3]Student, [4]Student, [5]Student

[1]Department of Information Science and Engineering,

[1]T. John Institute of Technology, Bangalore, India

Abstract: Accentify is a browser-based platform that transforms communication and learning by combining advanced visual data extraction with accentual speech synthesis. It converts text from images into speech, offering diverse accents and gendered voice options for a personalized and inclusive experience. Designed to bridge linguistic barriers, Accentify promotes cultural understanding and accessibility, making it ideal for education, global communication, and fostering engagement in multicultural settings.

KeyWords: Accentify, text-to-speech, visual data extraction, diverse accents, accessibility.

I. INTRODUCTION

Accentify is a browser-based platform revolutionizing communication and accessibility by combining visual data extraction with accentual speech synthesis. It extracts text from images and converts it into speech, offering diverse accents and gendered voice options for personalized, inclusive user experiences. Designed to bridge linguistic and cultural divides, Accentify empowers individuals with varying literacy levels or visual impairments and enhances engagement across educational, professional, and multicultural settings. By prioritizing accessibility, diversity, and user-centric design, Accentify fosters global understanding, redefining how we connect, learn, and interact with information.

II. OBJECTIVES

- **Efficient Visual Data Extraction**: Uses advanced OCR technology to extract text from images with accuracy, enhancing accessibility for users with visual impairments, learning disabilities, or language barriers.

- **Multi-Accent Speech Synthesis**: Converts text to speech in various accents, promoting cultural inclusivity and improving comprehension and engagement for global users.

- **Gendered Voice Options**: Offers male, female, and other voice profiles for personalized experiences, fostering representation and comfort.

- **Optimized User Experience**: Ensures speed, accuracy, and ease of use through a user-friendly interface, enabling effortless interaction with content.

III. EXISTING SYSTEM

The field of text extraction and speech synthesis is led by tools like Google Lens, OCR platforms (e.g., Adobe Acrobat, ABBYY FineReader), and TTS engines (Amazon Polly, Google Text-to-Speech). While these solutions excel in specific areas—Google Lens for text extraction, OCR tools for document digitization, and TTS engines for lifelike speech—they often operate in isolation without holistic integration or advanced personalization.Accentify bridges this gap by combining visual data extraction, multi-accent speech synthesis, and gendered voice options in a single platform. Unlike traditional tools, Accentify emphasizes inclusivity, cultural diversity, and personalized user experiences, making it a transformative solution for accessibility, communication, and engagement across linguistic and cultural boundaries.

IV. PROPOSED SYSTEM

Efficient Visual Data Extraction: Accentify uses advanced OCR technology to convert images, documents, and handwritten notes into accurate, machine-readable text. This feature enhances accessibility for students, professionals, and visually impaired individuals, automating data extraction and saving time.

Multi-Accent Speech Synthesis: Accentify converts text into lifelike speech with diverse accents like American, British, and Australian, improving comprehension and engagement for global users. It bridges linguistic barriers and fosters inclusivity by delivering content in culturally familiar voices.

Gendered Voice Options: Users can choose male, female, or other voice profiles for a personalized experience, enhancing accessibility, representation, and comfort in education, professional settings, and casual use.

Optimized User Experience: Accentify offers a user-friendly interface with fast processing, high-fidelity audio, and seamless navigation, ensuring accessibility for users with varying technical skills.

V. HARDWARE AND SOFTWARE REQUIREMENTS

HARDWARE COMPONENTS:

1. **PC**: Central system for developing, testing, and deploying Accentify. Provides computing power for OCR, speech synthesis, and API integration.

SOFTWARE COMPONENTS:

1. **HTML:** Structures UI elements (buttons, input fields).
2. **CSS:** Creates responsive, user-friendly designs.
3. **JAVASCRIPT:** Enables interactivity (text-to-speech, OCR triggering).
4. **IDE:** Visual Studio Code

VI. LITERATURE SURVEY

1. **Speech-to-Text (STT) and Text-to-Speech (TTS) Conversion**

- Techniques: MFCCs, LPC, deep learning models like HMMs, RNNs, and Transformers.
- Applications: Voice assistants, educational tools, healthcare accessibility.

2. **Speech Recognition and Synthesis**

- TTS: Tacotron, WaveNet; challenges include accented speech and background noise.
- Applications: Human-computer interaction and smart devices.

3. **TTS Implementation**

- Methods: Concatenative, Parametric, and Neural-based synthesis (Tacotron, WaveGlow).
- Challenges: Pronunciation errors, unnatural intonation, resource constraints.

4. Textless Speech-to-Speech Translation

- End-to-end neural models bypass text conversion, reducing transcription errors.
- Applications: Real-time multilingual communication, especially for low-resource languages.

5. **jsPsych for Behavioral Experiments**

- Evaluates TTS/STT systems for clarity, naturalness, and user interaction feedback.

VII. IMPLEMENTATION AND DESIGN

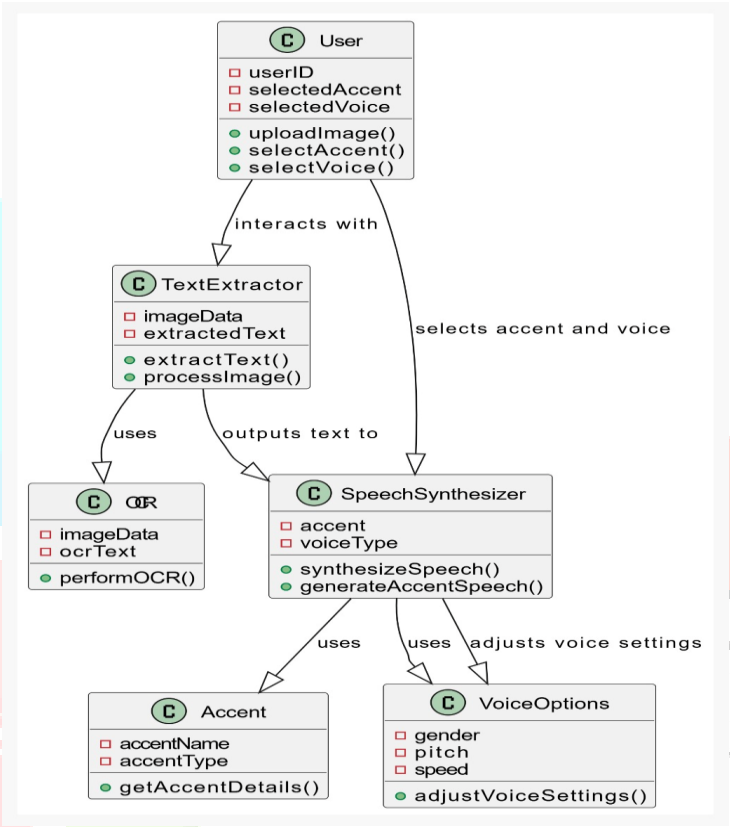

Fig 7.1 UML DIAGRAM

Key Classes in UML Diagram:

1. User Class
2. Accent Class
3. Speech Synthesis Class
4. Voice Option Class
5. Text Extractor Class
6. OCR Class

VIII. CODE

```
<div id="cameraContainer">
  <h2>Capture Image from Camera:</h2>
  <video id="cameraPreview" autoplay playsinline></video>
  <button id="captureButton">Capture Image</button>
  <canvas id="snapshotCanvas" style="display: none"></canvas>
</div>

<div id="previewContainer">
  <h2>Captured Image Preview:</h2>
  <img id="capturedImagePreview" style="display: none" />
</div>

<div class="upload-section">
  <h2>Or</h2>
  <h2>Upload an Image:</h2>
  <label for="imageUpload">Upload an Image:</label>
  <input type="file" id="imageUpload" accept="image/*" />
  <button id="extractButton">Extract Text</button>
</div>
```

Fig 8.1 CAMERA CAPTURE AND IMAGE UPLOAD SECTION

```
// Populate Available Voices Dynamically
function populateVoices() {
  const voices = window.speechSynthesis.getVoices();

  accentSelect.innerHTML = ""; // Clear previous options

  voices.forEach((voice) => {
    const option = document.createElement("option");
    option.value = voice.name;
    option.textContent = `${voice.name} (${voice.lang})`;
    accentSelect.appendChild(option);
  });

  if (voices.length === 0) {
    const option = document.createElement("option");
    option.textContent = "No voices available. Please reload.";
    accentSelect.appendChild(option);
  }
}

if (window.speechSynthesis.onvoiceschanged !== undefined) {
  window.speechSynthesis.onvoiceschanged = populateVoices;
} else {
  populateVoices();
}
```

Fig 8.2 POPULATE AVAILABLE VOICES DYNAMICALLY

IX. RESULT

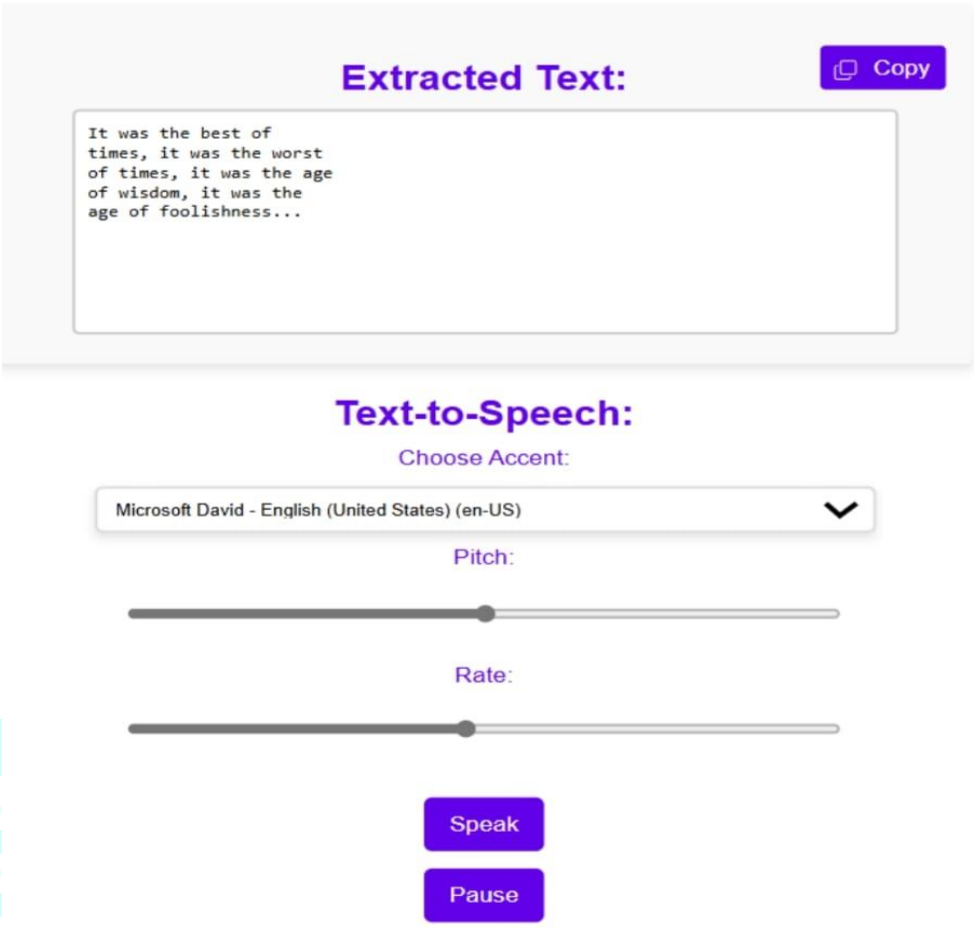

Fig 9.1 EXTRACTED TEXT

X. PERFORMANCE ANALYSIS

The application integrates OCR, text-to-speech, and camera functionalities seamlessly for robust performance. It efficiently extracts text from images using Tesseract.js, with minor limitations for handwritten text. The app features smooth live image capture, copy-to-clipboard functionality, and customizable text-to-speech options. Its lightweight design and intuitive interface enable efficient performance across devices. Overall, the application delivers accurate text processing, seamless interactivity, and strong usability.

XI. CONCLUSION

This project seamlessly integrates modern web technologies to provide an accessibility-focused solution for converting visual information into auditory feedback. It employs Tesseract.js for OCR and the SpeechSynthesis API for text-to-speech conversion, bridging the gap between visual and auditory communication. The application features a clean interface, intuitive controls, and accessibility options, making it a valuable tool for individuals with visual impairments. While it excels in its primary functionality, areas for improvement include OCR accuracy for handwritten images and additional features like multilingual text extraction. Overall, the project demonstrates a well-executed implementation of emerging technologies to address practical challenges in accessibility.

XII. REFERENCES

[1] Poonam S. Shetake, S. A. Patil, P. M. Jadhav [2014] Review of text to speech conversion methods.

[2] S. Venkateswarlu, D. B. K. Kamesh, J. K. R. Sastry, Radhika Rani [2016] Text to Speech Conversion. Indian Journal of Science and Technology, Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/102967, October 2016.

[3] Zhang, Y., & Lu, Y. (2023). Image-to-Text and Text-to-Speech Systems: Bridging the Gap Between Visual and Auditory Modalities.

[4] Smith, A., & Jones, P. (2022). A Comprehensive Approach to Text Recognition from Images and Its Integration into Speech Synthesis Systems.

[5] Smith, A., & Jones, P. (2022). A Comprehensive Approach to Text Recognition from Images and Its Integration into Speech Synthesis Systems.

[6] Kumar, S., & Bhattacharya, S. (2020). OCR and Speech Synthesis for Multi-lingual and Multiaccent Speech Generation. International Journal of Speech Technology, 29(5), 47-61.

[7] A REVIEW ON METHODS FOR SPEECH-TO-TEXT AND TEXT-TO-SPEECH CONVERSION. Shivangi Nagdewani, Ashika Jain

[8] Speech to text and text to speech recognition systems-Areview. Ayushi Trivedi,Navya Pant, Pinal Shah,Simran Sonik and Supriya Agrawal

Implementation of Text to Speech Conversion. Chaw Su Thu Thu, Theingi Zin

[9] Textless Speech-to-Speech Translation on Real Data. Ann Lee, Hongyu Gong, Paul-Ambroise Duquenne, Holger Schwenk, Peng-Jen Chen, Changhan Wang, Sravya Popuri, Yossi Adi, Juan Pino, Jiatao Gu, Wei-Ning Hsu

[10] jsPsych: A JavaScript library for creating behavioral experiments in a Web browser Joshua R. de Leeuw

[11] Text Information Extraction in Images and Video: A Survey Keechul Jung, Kwang In Kim, Anil K. Jain

[12] Accented Text-to-Speech Synthesis With Limited Data Xuehao Zhou , Mingyang Zhang , Yi Zhou , Zhizheng Wu and Haizhou Li

[13] Accented Text-to-Speech Synthesis With Limited Data Xuehao Zhou , Mingyang Zhang , Yi Zhou , Zhizheng Wu and Haizhou Li

[14] INVESTIGATION OF ENHANCED TACOTRON TEXT-TO-SPEECH SYNTHESIS SYSTEMS WITH SELF-ATTENTION FOR PITCH ACCENT LANGUAGE. Yusuke Yasuda, Xin Wang, Shinji Takaki1, Junichi Yamagish