IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, December 2024

A Review of AI-Enhanced Speech Recognition Systems

Prajwal Andanur[1], Suraj S Airani[2], Chandan N R[3], Kiran M S[4], Dr. Harish Bhat[5]
Under Graduate Students, Department of Electronics and Communication Engineering[1-4]
Assistant Professor, Department of Electronics and Communication Engineering[5]
Alvas Institute of Engineering and Technology, Mijar, Mangalore, India

Abstract: *Human-computer interaction has been revolutionized by voice recognition, which enables robots to comprehend and accurately record human speech. Accent, sound, and contextual nuances were too much for early systems, which depended on statistical models and rule-based algorithms. The field has reached previously unheard-of levels of precision thanks to AI, especially neural networks like transformers. From the beginning of speech recognition systems to the most recent developments in artificial intelligence (AI), such as deep machine learning, end-to-end training, and transformer-based architecture, this overview focuses on these themes. It also emphasizes how crucial various datasets like Svarah are for resolving issues with accents and other multilingual situations.Additionally, it explores how Retrieval-Augmented Generation (RAG) might be integrated for contextual understanding and talks about applications in client service, healthcare, education, and accessibility. Data biases, computing needs, and ethical issues still pose significant challenges in spite of these advancements. By working together to innovate solutions, we can fully utilize AI-powered voice recognition to build an interconnected inclusive society.*

Keywords: Speech Recognition, Artificial Intelligence (AI), Deep Learning, Transformer Architectures, Retrieval-Augmented Generation (RAG), Multilingual Contexts, Svarah Dataset, Human-Computer Interaction, End-to-End Learning, Accessibility Technologies

DOI: 10.48175/568