Summary of Automatic Audio Recognition Papers

Abdulkader Sardini   Sohaib Belaroussi   Henry Odongo   Syuja Akmal

April 22, 2025

1 Kader's part

1.1 Arabic Paper

try

Key findings: The study reveals that...

2 Henry's part [replace text with content]

2.1 Test

3 Suhaib's part [replace text with content]

3.1 Test

Speech Recognition by Machine: A Review

This article offers a comprehensive survey of the technological developments and foundational concepts in Automatic Speech Recognition (ASR) spanning six decades of research. It explores three major approaches in the field: the Acoustic-Phonetic approach, which decodes speech based on phonetic units; the Pattern Recognition approach, which applies statistical models such as Hidden Markov Models (HMMs) and techniques like Dynamic Time Warping (DTW); and the Artificial Intelligence approach, which involves expert systems and neural networks for modeling and adapting to speech patterns.

The article categorizes ASR systems by the types of speech they process—such as isolated words, connected speech, continuous, and spontaneous speech—and outlines practical applications in telecommunications, education, healthcare, and military domains. It highlights performance factors like vocabulary size, noise handling, speaker variability, and system adaptability. Essential techniques such as Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC), and Support Vector Machines (SVMs) are also emphasized for their importance in feature extraction and classification.

The paper also includes a historical overview, tracing progress from early analog devices like "Radio Rex" in the 1920s to modern pattern-based and neural systems. It emphasizes how innovations in hardware and algorithms have driven advances in ASR. The review concludes by identifying current challenges in spontaneous speech recognition, robust performance in diverse environments, and the need for system personalization.

The History of Speech Recognition to the Year 2030

This article by Awni Hannun reflects on the evolution of ASR technology from 2010 to 2020 and anticipates future developments through 2030. Major breakthroughs during the last decade—including deep learning, large-scale annotated datasets, and GPU acceleration—have enabled dramatic reductions in Word Error Rates (WER), surpassing human transcription performance in benchmark tasks. Innovations like Kaldi, LibriSpeech, Deep Speech models, and streaming on-device systems have redefined the ASR landscape.

Looking forward, Hannun predicts a shift in research focus from reducing WER to enhancing system usability and integration with downstream applications. He emphasizes the growing importance of self- and semi-supervised learning, lightweight model design, and on-device inference. These approaches offer benefits such as improved privacy, lower latency, and consistent performance in offline settings.

The article concludes by highlighting the need for personalized ASR systems that adapt to individual users' accents, speech patterns, and environments. Hannun also warns that increasing centralization of ASR research in large technology companies may hinder academic progress. Nevertheless, he remains optimistic about ASR's future in enabling accessible, intelligent, and context-aware speech technologies across various industries.