A STUDY OF THE RESOURCE EFFICIENCY OF VOICE COMMAND RECOGNITION METHODS IN PORTABLE COMPUTING SYSTEMS

Barkovska O., Romanenkov Yu., Trunov V.
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

The relevance of research on the development of inclusive voice-controlled systems for people with visual impairments, elderly users, and individuals with temporary or permanent mobility limitations is beyond doubt today [1, 2]. This research topic takes on particular importance in the context of implementing solutions on portable computing devices capable of operating in real time without relying on a powerful server infrastructure.

A practical application of the aforementioned systems could be guiding people through hospitals, shopping malls, government offices, universities, and other large facilities where speech signals are distorted by background noise, reverberation, and overlapping voices.

This leads to a decrease in the accuracy of voice command recognition related to finding a destination or clarifying a route.

Thus, there is a need to develop robust algorithms for processing and recognizing voice commands that ensure resource efficiency for implementation in modern assistive navigation systems.

The aim of this study is a comparative analysis of the resource efficiency of two proposed voice signal processing pipelines under acoustic noise conditions. The conditions of the study are as follows: a single computing device, identical conditions, and a duration of three minutes, with metrics recorded every second.

The algorithm of the first pipeline under study is shown in Fig. 1.

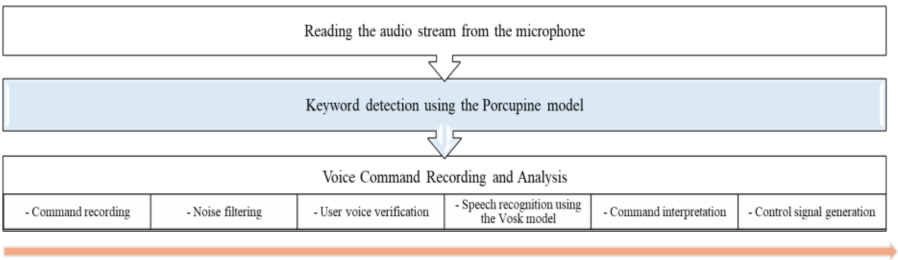

Figure 1 – Pipeline for the detection, processing, and analysis of voice commands

The combination of the Porcupine and Vosk neural network models—where a separate wakeword module continuously monitors the channel and only transfers control to the main command processing loop upon detection of a wakeword—allows the main STT model to remain in standby mode, thereby ensuring greater resource efficiency.

The second approach under investigation relies solely on the Vosk speech recognition model, which is constantly involved in both keyword detection attempts and subsequent processing. This increases the duration and intensity of the computational load.

For the Porcupine-based approach, typical CPU utilization after the initial phase is around 10% and increases primarily only during keyword triggering and the transition to command analysis.

For the Vosk-only approach, the load is more often in the 30–50% range and periodically reaches values close to 100% (Fig. 2).

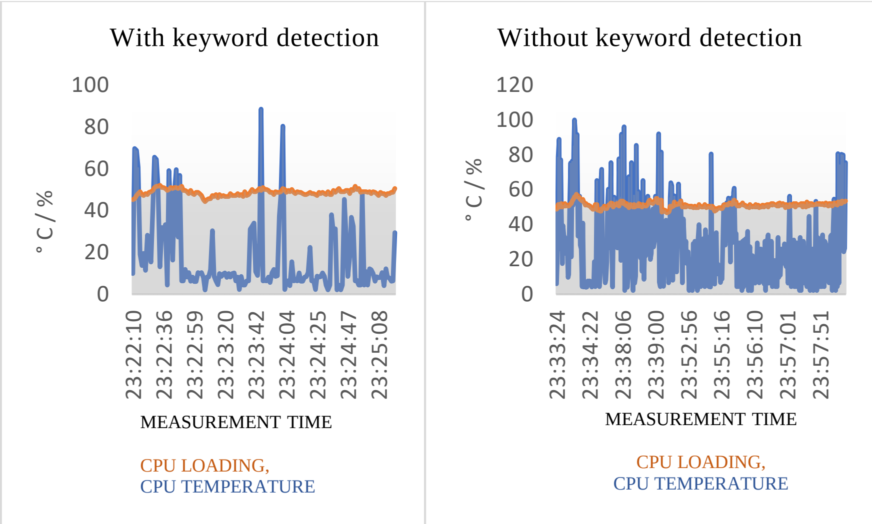

Figure 2 – Graph showing CPU temperature and load
with and without a dedicated wakeword processing module

Analysis of the results shows that extending the standard voice command processing pipeline with a separate keyword detection module based on the Porcupine model results in lower CPU load and fewer peak loads, demonstrating the higher resource efficiency of the proposed approach.

References

1. Hariharan, S., Abinaya, A., Anjuga, V., & Bhuvaneshwari, V. (2025). Voice Controlled Wheelchair for Physically Disabled People and Blind People. *Asian Journal of Applied Science and Technology (AJAST)*, *9*(1), 21-28. : https://doi.org/10.38177/ajast.2025.9103

2. Barkovska, O. and Serdechnyi, V. (2024) "Intelligent assistance system for people with visual impairments", *INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES*, (2(28), pp. 6–16. https://doi.org/10.30837/2522-9818.2024.28.006 .