Design and Clinical Co-Validation of an AI-Assisted Open-Source Rehabilitation Interface for Assistive Device Interaction

Adwait Shinde, Faizan Mir, Yassin Elsalamouny, Nic Brink, Ghaly Mikhail, Appolinaire C. Etoundi

UWE Bristol University of the West of England

NHS Bristol Centre for Enablement

BAPO

Introduction

Problem
- Costly, proprietary systems with limited customisation
- Connectivity dependence → reliability & privacy issues

Why it matters
- Patients: simple, independent control
- Clinicians: safe, repeatable, easy setup

Aim
- Develop an open-source, offline AI interface for personalised, voice-driven assistive control suitable for clinical use

Key contributions
- Offline speech recognition → command mapping → real-time actuation
- Modular system: Pi 5 (AI inference) + Pico (deterministic control)
- Clinical co-validation: (usability, feasibility, performance)

Method

System Architecture:
- Pi 5: offline speech recognition & logic
- Pico: deterministic motor/IO control (I2C)

AI / Software pipeline (Vosk):
- Vosk ASR → command match → hex ID → Pico

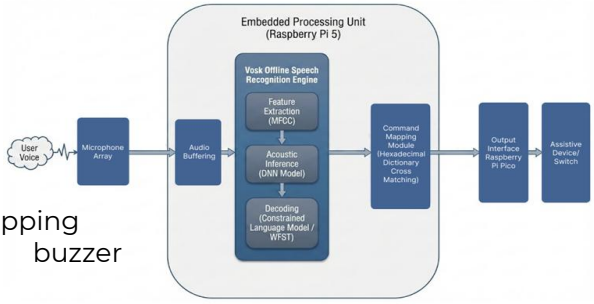
Command mapping + safety logic:
- Whitelisted commands only
- "Stop" → safe state
- Buzzer gated by hardware switch

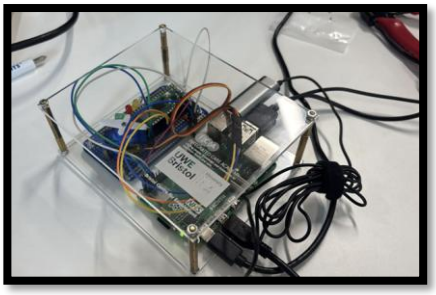
Co-validated approach (clinical-in-the-loop):
- Prototype to be reviewed and tested with rehabilitation stakeholders
- Evaluated for response time, command reliability, usability in realistic use

Results

Performance:
- Response time: ~120–130 ms
- Accuracy: >90%
- Stable across repeated tests

Functional outcomes:
- Reliable voice → actuation mapping
- Multimodal control: LED, buzzer (gated), servo, stepper





Discussion

Key Insights:
- Offline ASR with constrained commands enables low latency and reliable performance
- Split architecture (Pi for AI, Pico for control) ensures safe, deterministic real-time operation

Impact:
- Patients: greater independence with personalised, low-effort control
- Clinicians: easier setup, customisation, and integration into workflows

Limitations:
- Early-stage validation
- Requires testing across diverse users, speech impairments, and noisy environments

Conclusion

- Demonstrated an offline, open-source voice-to-actuation interface for rehabilitation
- Achieved ~120–130 ms latency with stable, reliable control
- Clinical co-validation shows feasibility and readiness for patient trials

Future work:
- Larger patient studies and noise-robust testing
- Expanded multimodal sensing
- PCB-based productisation for wider deployment

References

Alphacep, "vosk-api: Offline speech recognition API," GitHub. [On line]. Available: https://github.com/alphacep/vosk-api

Alphacephei, "VOSK Offline Speech Recognition API," Alphacephei. [Online]. Available: https://alphacephei.com/vosk

J. Lin et al., "AWQ: Activation-aware weight quantization for on-device LLM compression and acceleration," in Proc. Machine Learning and Systems (MLSys), 2024. [Online]. Available: https://proceedings.mlsys.org/paper files/paper/2024/hash/ 42a452cbafa9dd64e9ba4aa95cc1ef21-Abstract-Conference.html

M5Stack, "Module LLM Kit documentation," M5Stack Docs. [Online]. Available: https://docs.m5stack.com/en/module/Module%20LLM% 20Kit

M5Stack, "M5Module-LLM Arduino API," M5Stack Docs. [Online]. Available: https://docs.m5stack.com/en/stackflow/module llm/arduino api

M5Stack, "LLM630 Compute Kit documentation," M5Stack Docs. [Online]. Available: https://docs.m5stack.com/en/core/LLM630% 20Compute%20Kit

M5Stack, "StackFlow framework repository," GitHub. [Online]. Avail able: https://github.com/m5stack/StackFlow

M5Stack, "M5Stack LLM (AX630C) module kit," M5Stack Shop. [Online]. Available: https://shop.m5stack.com/products/ m5stack-llm-large-language-model-module-kit-ax630c

D. Povey et al., "The Kaldi speech recognition toolkit," in Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2011. [Online]. Available: https://www.danielpovey.com/files/ 2011 asru kaldi.pdf

Qwen Team, "Qwen2.5 technical report," arXiv preprint arXiv:2412.15115, 2024. [Online]. Available: https://arxiv.org/abs/ 2412.15115

H. Wang et al., "lm-Meter: Unveiling runtime inference latency for on-device language models," arXiv preprint arXiv:2510.06126, 2025. [Online]. Available: https://arxiv.org/abs/2510.06126

X. Wang and W. Jia, "Optimizing edge AI: A comprehensive survey on data, model, and system strategies," arXiv preprint arXiv:2501.03265, 2025. [Online]. Available: https://arxiv.org/abs/2501.03265

P. Warden, "Speech commands: A dataset for limited-vocabulary speech recognition," arXiv preprint arXiv:1804.03209, 2018. [Online]. Avail able: https://arxiv.org/abs/1804.03209

J. Xu et al., "On-device language models: A comprehensive review," arXiv preprint arXiv:2409.00088, 2024. [Online]. Available: https:// arxiv.org/abs/2409.0008