CASE STUDY

Hello Future: AI-Enabled Voice Control For Consumer Devices

- Sensory Inc.
- AI Technologies
- 200+ employees
- California, U.S.
- Founded 1994

OVERVIEW

Sensory Inc. is a leading provider of embedded AI technologies including voice control in automobiles and consumer electronic products. The company recently demonstrated its TrulyNatural and TrulyHandsfree speech recognition and natural language understanding (NLU) platforms running on the Arm Cortex-M55 processor and Arm Ethos-U55 machine learning accelerator. This enables advanced wake words, voice-control, and NLU capabilities on low-power microcontrollers designed for consumer appliances.

Challenge

Creating an embedded voice assistant that understands natural language commands, while being extremely efficient in memory and power usage, is a major challenge. Sensory aimed to leverage Arm's latest ML accelerator in Alif's Ensemble SoC to run its TrulyHandsfree and TrulyNatural platform for voice control of appliances. Key requirements were:

- Run on Arm Cortex-M55 and Ethos-U55
- Easy to customize vocabulary and languages
- Small memory footprint
- Low power consumption
- High accuracy in real-world conditions

Solution

Sensory used its VoiceHub tool to rapidly build a compact wake word and large vocabulary language model tailored for appliance voice control. VoiceHub, which is used by more than 1,000 companies, allows rapid creation of custom wake words, commands, and grammars using generated synthetic data, which eliminates the time and cost of collecting custom data. This enabled Sensory to easily define the required voice commands like “set microwave power to 50%” without a lengthy development process.

The TrulyNatural speech recognition and NLU engine were optimized to run efficiently on the Alif Semiconductor Ensemble E3 SoC, which is powered by Cortex-M55 and Ethos-U55, and all data stays on device, ensuring absolute data privacy for users of the system. The model can run effectively on the Cortex-M55 alone, but Sensory wanted to achieve greater efficiencies by using the Cortex-M55 and Ethos-U55. Key optimizations included:

- Partitioning workloads across the Cortex-M55 and Ethos-U55 to maximize throughput. Speech recognition runs primarily on the Ethos-U55, with the Cortex-M55 handling text processing and intent determination.
- Using Helium for accelerating machine learning functions within the Cortex-M55 to reduce power consumption for always-on wake word detection.
- Optimizing memory usage and network architectures to minimize bandwidth needs for the Cortex-M55 core.
- Using the Ethos-U55’s native support for INT8 quantization and sparsity to reduce model size. This was critical for fitting into the memory constrained environment of a microcontroller.

SOLUTION

— Leveraging the Ethos-U55 to run the acoustic model, which was 300% more efficient than running everything only on the Cortex-M55. Sensory was able to run complex deep learning models quickly and efficiently for speech recognition by using the TensorFlow Lite framework on Ethos-U55 AI accelerator.

These software optimizations allowed the advanced TrulyNatural engine to run smoothly on Arm's Cortex-M55 and Ethos-U55 accelerator on Alif's Ensemble E3 platform. The solution met the goals of providing customizable, accurate speech recognition and understanding while adhering to the tight performance constraints of consumer devices.

Results

Sensory showcased its TrulyNatural platform running on an Alif Ensemble E3 evaluation board with the Cortex-M55 and Ethos-U55. It handled open-ended voice commands with a broad range of user-centric cooking terminology for tasks like setting microwave cook times and power levels.

RESULTS

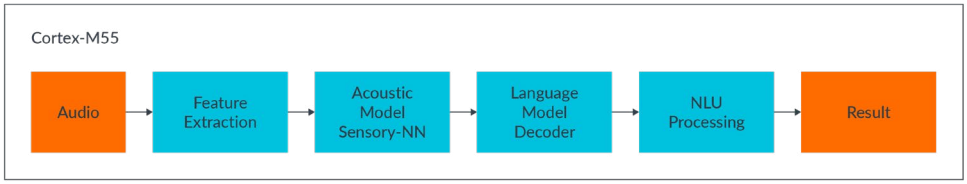
By leveraging Arm's AI-capable technologies, Sensory delivered an embedded speech assistant that provides an intuitive user experience through natural voice controls, while meeting the tight memory and power limits of consumer devices.

By collaborating with Arm on AI at the edge, Sensory is poised to enable a new generation of specialized on-device experiences leveraging these processors. For instance, Sensory reduced the memory footprint of a microwave language model from 17MB to 1.6MB, partly attributed to efficiencies from the Cortex-M55 and Ethos-U55.

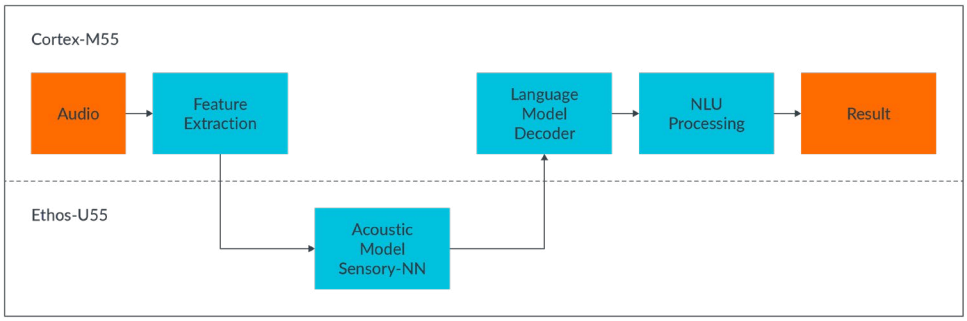
Conclusion

Sensory's use of Arm's Cortex-M55 and Ethos-U55 processors demonstrates that natural language voice interfaces can now be embedded cost-effectively into appliances and other products. Companies can use Sensory's optimized speech recognition solutions together with Arm's ultra-efficient AI technology to rapidly deploy customized voice control experiences in their consumer devices.

Baseline Architecture



Hybrid Architecture



arm