ARM

audio analytic

Whitepaper

Machine learning in low-power devices brings sound recognition to the smart home market

Tamara Sword – CMO, Audio Analytic

Dr. Dominic Binks – VP of Technology, Audio Analytic

Dr. Sacha Krstulovic – Director of AA Labs, Audio Analytic

Thomas Lorenser – Product marketing manager, ARM

January 2017

Table of Contents

Introduction

The Smart Home market is now at an inflection point. Early devices in the market were connected to the internet but were typically single-function, often lacking connectivity to other devices and with closed APIs, denying the user the ability to design multi-device applications around smarter living use-cases.

The first wave of Smart Home assistants such as Amazon Echo and Google Home are now in the market. They combine artificial intelligence and speech recognition to deliver new services and manage other devices within the home. Their arrival and accessible price points are rapidly moving the Smart Home experience from early adopter to mainstream adoption.

The popularity and familiarity of voice assistants on mobile has transferred to the Smart Home ecosystem and voice has quickly become a standard user-interface for Smart Home assistants and devices. Once a Smart Home assistant is adopted within the home, homeowners are more inclined to purchase other smart devices they can incorporate into their home and manage through their assistant.

Increasingly, Original Equipment Manufacturers (OEMs) are responding to the growing market opportunity by seeking ways to transform low power single-purpose devices into low power-multi-purpose devices, connected to the wider Smart Home ecosystem. Consumers are also looking for ways to make traditionally "dumb" unconnected devices such as smoke alarms smarter, but without incurring the cost of replacing existing units with more expensive, connected versions. Hence industry analysis firm IHS predicts a significant growth of the Home consumer IoT device market over the next years and expects above 1 billion in unit shipment worldwide by 2025.

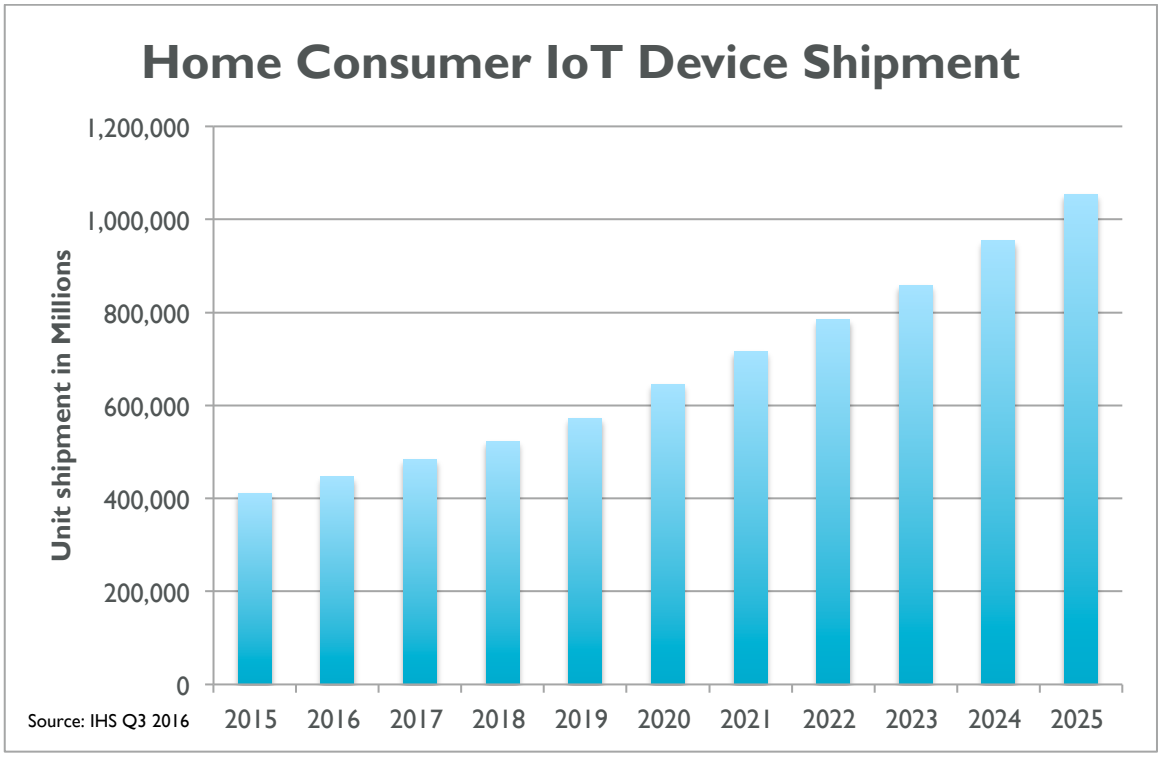

Fig 1. Global Home Consumer IoT Device shipment forecast (Source: IHS Q3 2016)

This white paper addresses how artificial intelligence and machine learning are used in the Smart Home market focusing on sound recognition. We will describe use cases and requirements across several end-device categories and how Audio Analytic's embedded software platform Artificial Audio Intelligence (ai3™) can be used on low-power ARM Cortex processors.

Artificial intelligence's missing link

Artificial intelligence (AI) is increasingly being deployed by OEMs addressing the smart home. The technology enables devices to make executive decisions on the management of the smart home based on contextual information and voice commands – an interface increasingly popular with consumers.

AI also enables smart assistants to deliver a 'conversational' user experience, understanding and responding to the user's natural language commands.

AI provides smart home devices with the capability to recognize and respond to user voice commands. While capable of recognizing speech, it often comes as a surprise to even those working in the technology industry that such AI solutions are as yet unable to recognize or respond to other sounds.

Every sound tells a story – especially in the home where even mundane sounds can be significant. The human brain intuitively detects, recognizes and processes sound in order to understand its context and environment. If the smart home is to be truly intelligent, it too must be able to recognize sound. Doing so unlocks compelling new use cases, both inside the smart home and beyond.

Sound recognition in the smart home

A baby stirs and begins to cry in the night. A smart home device in the nursery recognizes the sound and automatically plays a lullaby to sooth baby back to sleep. If the baby continues to cry, the device Event Management System (EMS) sends an alert to one parent's wearable device to gently wake them and turns on night lights so they can make their way safely to the nursery. Over time, the data collected and provided as a service gives the parents new insight into when their baby cries, helping them establish a better sleep routine.

An intruder breaks a window of the home while the owners are away. A smart home device recognizes the sound and its EMS turns on the lights and plays loud music from a speaker to actively deter the intruder. Simultaneously, an alert is sent to the home owner's mobile device so they can alert the authorities and their neighbours.

An electric appliance is accidentally left on at home after the home owners leave for work. It overheats and starts a fire. The home smoke alarm begins to sound, but no one is at home to hear it. The sound of the alarm is recognized by a smart home device. The device EMS sends an alert to the home owner's mobile device together with and a clip of the sound so the home owner can take appropriate action, such as call the emergency services.

A dog spends a lot of time alone at home while his owner is at work. When the dog barks, a smart home device recognizes the sound and sends an alert to the owner's mobile device. She can choose to talk to her dog over the intercom, view a video feed and even dispense treats from a smart feeder. Again, the data collected on when the dog barks is provided as a service to help the owner manage and improve her pet's wellbeing.

How sound recognition differs from voice recognition

While voice recognition has been commercially available for more than twenty-five years, sound recognition is an emerging field, requiring a different technological approach. Let us first consider voice recognition. In scientific terms, voice recognition is a discrete problem. There are a limited number of sounds the human voice is capable of making and a finite number of phonemes that appear within human language.

Sound recognition – from a technical stand point – is a far more challenging task. Sounds come in a far greater variety than human speech. In addition, any sound recognition solution must be able to recognize a given sound among a continually shifting background of other sounds. The structure of an environmental "soundscape" is often different from speech or music. Sentences are structured by grammar, and musical pieces by a score, but in a soundscape there is a random element: any sound can follow any other sound, with loose sequential constraints. Environmental sounds can also overlap, with the sound of interest occurring on top of a noisy background. While the sounds of speech are restricted to those that the human vocal tract can produce and the sounds of music are often structured as tuned notes, environmental sounds can sound like anything and be produced by any means: crashing objects, explosions, beeping circuits, animal sounds, machines humming - each results from a different physical process.

The quality of audio captured by devices can also be a challenge. Speech is most often spoken close to the microphone of a mobile phone or home assistant, or at the sweet spot of some in-car audio capture system - situations known in the literature as 'close talk' or 'near field audio capture'. Environmental sounds, on the other hand, can happen anywhere, at any given distance from the microphone - a situation known as "far field audio capture".

Enabling smart home devices with sound recognition

Environmental sound recognition must work on any embedded device, including devices whose audio circuitry was designed with cost reduction in mind, rather than the quality of audio capture and transmission. For example, while Amazon Echo, Google Home and most mobile phones are now using microphone arrays to "target" people's speech through a technique called beamforming, the vast majority of Consumer Electronics and Smart Home devices are still natively using mono audio capture, in which case standard sound enhancement techniques cannot be taken for granted. Dealing with such practical constraints and suboptimal audio in order to fit into existing hardware designs is thus part of the art of environmental sound recognition.

In addition, while mainstream speech and music recognition services are typically reliant on great computational power hosted in enormous data centres, sound recognition software is most often required to run 'on the edge', i.e. it must work within the limited computational power available on its device host. For example, directly within a smart lock, smart light or other common place household device.

In order to solve the challenges of a complex soundscape, variable audio quality and low computational power, a sound recognition system must be very good at modelling a wide range of acoustical phenomena, whilst also being able to cope with a multiplicity of noise conditions and microphone types. In the same way that speech recognition must be tolerant to variations in people's voices when they have a cold, a sound recognition system must also be able to generalise between different instances of a sound, e.g. window glass broken at different thicknesses and sizes, different models of smoke alarms, etc. All of this intelligence must be able to run at a very small computational cost.

Key end-user requirements

End-users have a low tolerance for inaccurate or missed alerts, especially when it comes to applications in the home related to security, safety and wellbeing. Creating a precise sound recognition solution for the home that accurately recognizes given sounds and meets end-user expectations requires:

- Machine learning specifically designed for the unique demands of sound recognition
- A high-quality and high-volume data sets gathered from real home environments and lab tests
- Complex feature extraction of more than 500 individual sound features
- An intelligent decision engine within the software that goes beyond traditional score-based approaches
- On-device sound recognition, so home audio is not continually streamed to the cloud, ensuring user privacy

Only a system with all the attributes above can deliver the accurate and dependable experience required to meet end-user expectations.

Key OEM requirements

Original Equipment Manufacturers (OEMs) addressing the smart home are increasingly looking to deliver new joined up smart home use cases to their customers and make their products and services "smarter" through the application of artificial intelligence.

Sound recognition offers OEMs a way to create smarter products, opening up entirely new connected use cases and services. OEMs seeking to embed sound recognition may have some or all of the following requirements:

- Enabling low-power devices with sound recognition
- Deploying sound recognition on devices with low memory capacity
- Easily retrofitting with new sounds to existing devices via software updates
- Assuaging consumer concerns about privacy and security by executing all sound recognition on the device with no cloud streaming required

Introducing Artificial Audio Intelligence

Artificial Audio Intelligence (ai3™) is an embedded software platform that recognizes and responds to sounds within the home – a feature that enables smart home devices to better respond to their environment when occupants are out or asleep.

The ai3™ platform is capable of recognizing a range of significant sounds including the sound of window glass breaking, smoke and CO alarms, dog barks - even a baby cry. The platform can also be trained by the end-user to recognize custom sounds, such as a door bell and white goods alerts. In addition, ai3™ is capable of understanding the normative pattern of sounds in a given home and creating an alert whenever anomalous sound events occur.

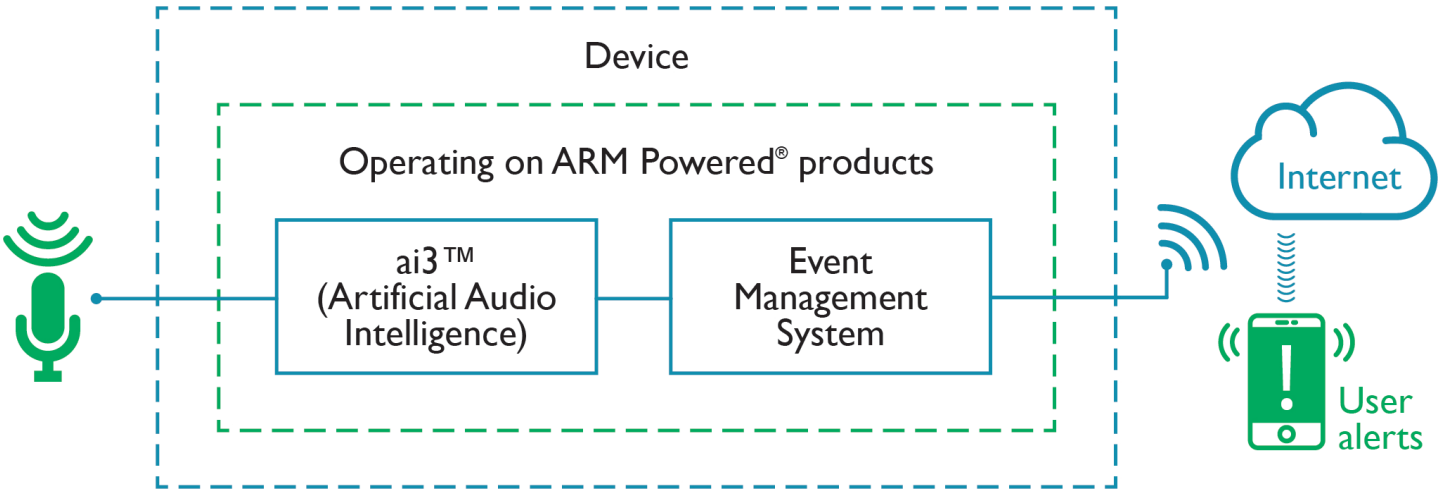

Fig 2. System overview of sound recognition operating on ARM Powered® products

When ai3™ recognizes a given sound, it connects to the device Event Management Service (EMS) to respond as required by the use case, i.e. sending a notification to a mobile device, connecting to other devices, etc. ai3™ offers an audio clipping service where a short clip of the sound that triggered an ai3™ alert is sent to the absent user for validation and executive action, i.e. sound of a window glass break or smoke alarm. Data of ai3™ alerts over time can be served to the end-user by the OEM or service provider as an additional tracking service, for example baby cry or dog bark logging to gain meta-insights and manage wellbeing.

Low-power ARM IP for machine learning applications

ARM's energy-efficient Cortex-M processors are used in a wide range of machine learning applications, from Cortex-M0 and Cortex-M0+ for activity detection, to Cortex-M4 in keyword-spotting algorithms and Cortex-M7 in cough detection.

The instructions found in Cortex-M4 and Cortex-M7 provide a range of of operations suitable to accelerate signal processing and machine learning algorithms. For example, these processors offer a large variety of SIMD instructions, each of which executes in a single cycle operating on 8- or 16-bit data. Cortex-M4 and Cortex-M7 provide the opportunity to perform signal processing operations directly on a microcontroller without the need for additional hardware, such as an external digital signal processor.

Within the Cortex Microcontroller Software Interface Standard, CMSIS package (which is available for free download), there is a library written entirely in C which implements a number of optimised functions using the SIMD, saturating arithmetic and single-cycle MAC instructions of the Cortex-M4 and Cortex-M7. This library contains optimized functions for transforms, such as FFT, and filters such as Biquad, FIR, and IIR, as well as a wide range of other mathematical functions. The CMSIS-DSP library includes optimized functions, such as matrix multiplication, to accelerate a wide range of machine learning algorithms and neural networks.

Cortex-M processors bring sound recognition to deeply embedded devices

Cortex-M4 and Cortex-M7 deliver a high level of DSP and integer performance, while maintaining the energy-efficiency and ease-of-use hallmarks of the Cortex-M processor family. Cortex-M7 achieves twice the DSP performance of Cortex-M4 and offers a flexible memory system consisting of caches (instruction and data) up to 64KB each and TCMs (tightly coupled memories) up to 16MB. Both processors open up new opportunities for machine learning in a wide range of applications. For example, it allows the deployment of sound recognition into deeply embedded low-power devices such as thermostats, smart locks, smart home assistants and wireless speakers.

	Typical Cortex-M4 based MCU	Typical Cortex-M7 based MCU	Typical MCUs based on Cortex-A7, Cortex-A8 or Cortex-A9
RAM usage (peak) (kbytes) - 3 sound profiles	26	97	110
MIPS for 3 sound profiles on single chip	26.6	36	36
Sufficient RAM for recording audio clips	No*	Yes	Yes
Capable of supporting custom sounds	No*	Yes	Yes
Max number of sound profiles supported on single chip (MIPS & RAM related constraint)	3	6	>10
Example Devices	Smart thermostats, intellgent sensors, smart locks	Set-top boxes,audio/video processors, low cost hubs, IP cameras, smart speakers	Smart home assistants, higher resolution IP cameras, PVRs/set-top boxes, routers, tablets / smartphones

* depends on system RAM availability

Table 1: Comparing ai3 on several ARM based MCUs

Typical Cortex-M4 based MCUs can support sound recognition for low-power devices but are often limited in memory size for the full deployment of ai3™ with clip service or Custom Sound Recognition capability. Essentially, Cortex-M4 based MCUs are compatible with ai3™ "lite": that is sound recognition of up to three sound profiles, but with no sound clipping service unless the minimum RAM requirement of 512KB is met. This combination allows deployment in very constrained applications, such as smart thermostats and smart locks.

With the introduction of Cortex-M7, sound recognition is available for high performance, low-power devices with full deployment of ai3™ with clip service. Cortex-M7 based MCUs then open up new opportunities for sound recognition of new devices, providing enough RAM as standard for up to six sound profiles and a full sound clipping service.

Systems based on the Cortex-A series , such as Cortex-A7, Cortex-A8 and Cortex-A9, provide OEMs with sufficient MIPS and RAM to support ai3™ capable of recognizing up to ten sound profiles with a full sound clipping service.

The recognition of new sounds can be retrofitted to existing devices via software updates. New sounds are regularly being added to ai3™, enabling OEMs to bring additional value to users and open up opportunities for related income streams from subscription services.

With ai3™, all sound recognition is executed on device, meaning there is no requirement for wi-fi or internet access. This removes the need for audio from the home to be streamed continually to the cloud, avoiding potential consumer concerns about security and privacy.

Sound recognition - software architecture

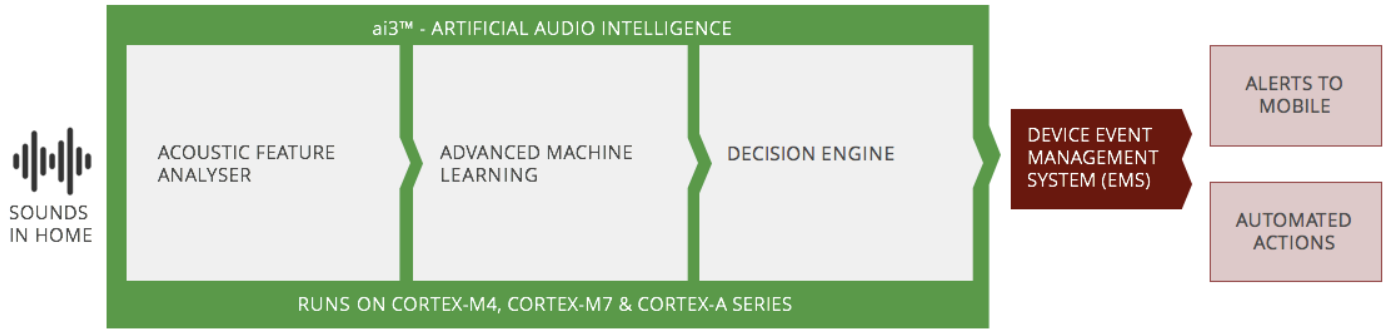

Fig 3. Software architecture overview

Summary

In this whitepaper, we introduced sound recognition for low-power devices. ai3™'s novel sound recognition software has a wide range of applications in the smart home, enabling OEMs to equip their devices and services with Artificial Audio Intelligence. It also enables OEMs to turn single-use smart home devices into multi-functional devices connected into the wider smart home ecosystem.

ai3™ is a fully scalable platform which, once trained, has the capability to recognize any sound. New sounds and features are continually being added to the platform, offering OEMs an ongoing roadmap for product and service innovation.

In addition, the ARM Cortex-M7 processor opens up new design freedom for OEMs looking to embed sound recognition software in high performance, low-power devices. The Cortex-M7 offers a flexible memory system and allows energy-efficient processing of demanding signal-processing applications, combined with the CMSIS-DSP library.

Enabling devices to recognize significant sounds is required if artificial intelligence is to fully enable devices to accurately understand and respond to their context. With Audio Analytic's ai3™ technology available on ARM Powered products, such powerful Artificial Audio Intelligence is no longer restricted to cloud applications, and can run directly on the edge. As such, sound recognition has applications not only in the smart home, but can also be harnessed in the wider IoT market, impacting healthcare, elderly care, automotive and industrial applications and smart buildings.

Trademarks