Engineering tinyML models in sound recognition
By Dr Dominic Binks and Arnoladas Jasonas
Audio Analytic, 2 Quayside, Cambridge, CB5 8AB, UK
www.audioanalytic.com
(firstname.surname)@audioanalytic.com

Overview
- Audio Analytic’s edge-based sound recognition software is already available on consumer devices worldwide
- Engineering tinyML models is fundamental as devices push towards cloudless AI, driven by privacy and cost
- To push boundaries of tinyML, Audio Analytic embedded its ai3™ software on an Arm-Cortex M0+ chip
- High-performing tinyML requires a specialist ML pipeline, large dataset, and an end-to-end focus on compactness when it comes to model training and software development
- Our results show model size is significantly reduced with very little change in performance.

ML Pipeline optimized for sound event and acoustic scene recognition

Compactness
Low-level software optimization ensures tinyML models are adequately compact and functioning on small devices.

M0+ implementation
- A Cortex M0+ chip has approx. a fifth of the computational power (for ML-type workloads) of an M4-based chip
- MO+ implementation requires fixed-point implementation. Diagram below shows the relative impact of fixed-point on model performance when compared to floating-point

Comparison of Floating vs Fixed point implementation

Select the best models for tinyML
- The Polyphonic Sound Detection Score (PSDS) is an industry-standard evaluation framework and metric which evaluates the performance of sound event detection models
- The identification of the best performing system is grounded in user experience rather than abstract or impractical statistical definitions.
- Learn more: aud.ai/psds

Four steps to determining PSDS (illustrative purposes only)
- Operating point (trade-off between true positives and false positives) has changed slightly but is still within target performance for application
- Targeted platform optimisation resulted in the reduction of ROM, RAM and MIPS numbers from the baseline version
- In this implementation, our software requires just 116Kb ROM and 65Kb RAM
- Our MO+ implementation is production-ready and meets high accuracy targets of other commercial deployments of ai3™.

An extensive dataset required for tinyML
- Alexandria™ is the world’s largest dataset of sound events and acoustic scenes for machine learning
- It contains 15 million labelled sound events, over 700 label types and over 200 million metadata points
- Foundational part of the machine learning pipeline. Dataset size and quality is key for tinyML
- Sophisticated labelling strategy. Three levels of labels (fine, episodic and weak) support model training and evaluation along the pipeline
- Sounds collected in specialist anechoic chamber, through volunteers and dedicated data collection campaigns
- ‘Baby cry’ model was selected for tinyML research and M0+ implementation.

Visualization of the data stored within Alexandria™