Smart Baby Monitor

xcore.ai Baby Monitor with Integrated White Noise Machine

ELECOMP Capstone Design Project 2021-2022

Sponsoring Company:

XMOS Ltd.
Queens Quay
33 – 35 Queen Square
Bristol, BS1 4LU
United Kingdom
http://www.xmos.com

Company Overview:

Our relationship and reliance on technology is changing rapidly. In this increasingly complex, connected world, we want a more seamless connection with the technology around us. We want technology that can talk, empathize, relate, and react – an interface between human and machine that feels … human. One that comes with no instructions needed, delivered through artificial intelligence and connectivity.

XMOS is at the forefront of the far-field voice interface market, with the most qualified, most comprehensive range. Adopted by leading brands, our high-performance, easily integrated solutions are transforming consumers' interactions with everyday devices.

With more than 78 hardware and algorithm patents, XMOS silicon and software solutions deliver best in class voice capture for the smart environment.

We are headquartered in the UK, with offices across Asia and the United States.

Technical Directors:

Dr. Andrew Cavanaugh (URI CoE '08)
Algorithm Engineer
andrewc@xmos.com
https://www.linkedin.com/in/andrew-c-7805ba8/

Steven Anzivino (URI CoE '21)
ELECOMP Capstone Graduate 2021
Junior Software Engineer
stevenanzivino@xmos.com
https://www.linkedin.com/in/steven-anzivino-2619aa1a1

Project Motivation:

XMOS voice frontend technology is currently underutilized in the non-internet connected device market. The xcore.ai chip is very well suited to be the only processor in the box for many smart appliances, and baby monitors are a perfect blend of classical audio processing and lightweight on-device AI. The market for baby monitors is large, and there is specifically a demand for units that are not connected to the internet for privacy reasons. Most monitors will also be placed in a bedroom that includes a white noise machine or music player. By combining these with the baby monitor we reduce cost, clutter, and complexity, while providing a better experience on the remote unit(s) which will have the noise/music removed from the stream so that parents only hear the sounds being made in the room. This noise/echo reduction also allows for more advanced processing such as sound even detection (cry, fall, glass break, etc.), which can provide enhanced alerts to parents.

Anticipated Best Outcome:

The Anticipated Best Outcome (ABO) of this project is a fully functional prototype that can utilize a wide array of peripheral devices to function as a baby monitor. Ideally this system would produce audio through a speaker, record audio through a microphone, and perform audio processing to reduce echoes or perform analysis. This device should be able to be housed within a housing no larger than existing baby monitoring devices on the market.

This project should attempt to use:

- XMOS Osprey development kit (includes microphones)
- XMOS Audio Frontend libraries
- Team-designed 3rd party DAC (or design a class D amplifier if EE resources permit)
- As many off-the-shelf parts as possible

Additional features that go beyond the scope of the ABO would be...

- Audio processing to discern crying or other events
- WIFI connectivity between two devices to transmit audio
- Single board version merging DAC board and Osprey board
- Design and build the remote (parent) unit

Project Details:

The end goal of this project is to have a representative prototype for a device that listens to room audio and can play noise/music into the room. The composition of the team will determine how refined the hardware design is vs. how feature rich the firmware on the device will be. In all cases an xcore.ai development board (with built-in microphones) will need to be augmented with a Digital to Analog Converter (DAC) and an audio amplifier to power a small loudspeaker, and that audio stream will also have to be fed into an audio front end that should allow for simultaneous transmit and receive of audio.

For a device with a microphone to transmit sound while listening on the mic (full duplex) that device needs to be able to ignore its own sound. The attenuation of this ("far-end") audio is called Automatic Echo Cancelation (AEC) and was first developed for telecommunications (hence the terms like "near end" and "far end"), to allow full duplex conference phones, and dialup modems. XMOS will provide AEC and other audio processing components, but it will be up to the team to integrate these and test the overall performance of the prototype.

Processed audio will be sent to a remote unit (initially just a PC) that will represent the parent unit(s) and this can be done via Wi-Fi. XMOS will provide working examples of audio over Wi-Fi, but there will be significant work required to move these into the prototype project. This will involve a lot of networking knowledge as well as embedded programming, and the technical directors will offer guidance on avoiding pitfalls.

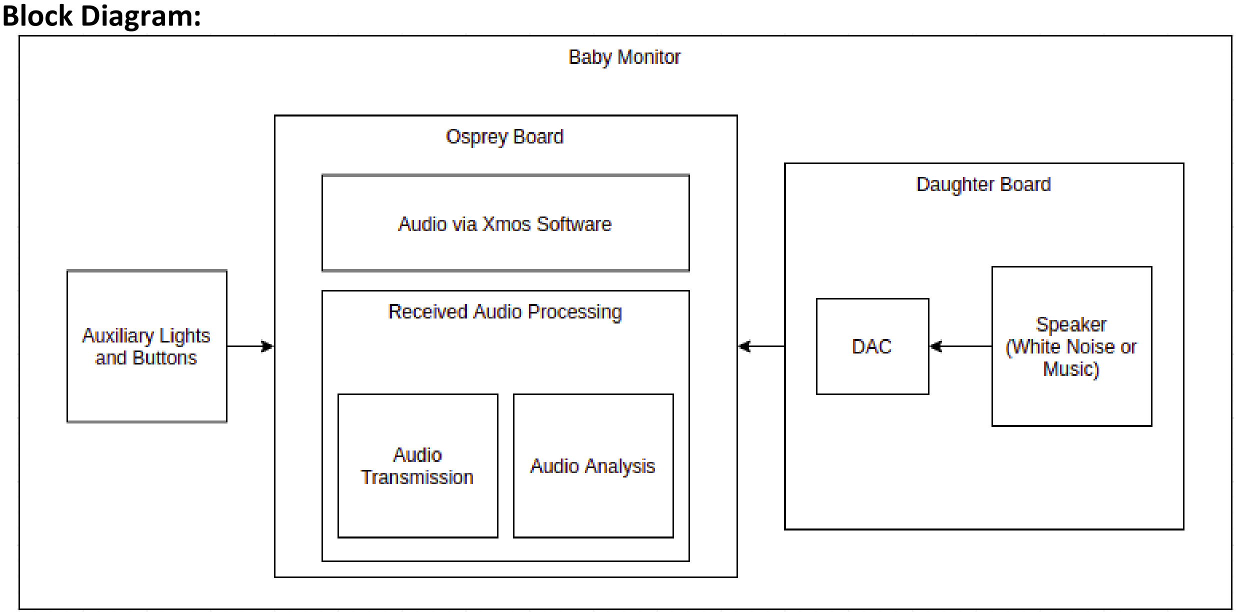
In addition to sending processed audio, the monitor should also be able to send metadata which could be as simple as track/volume for the playback audio (which won't really be audible on the remote unit) and as complex as events such as crying determined by a neural network-based classifier that the team can either find and port or train and port. The latter option should only be taken if the team is computer engineers, or if one or more members have a hobbyist-level familiarity with machine learning.

The playback/far-end/music/noise audio will need to be played over a small loudspeaker that, when combined with a DAC and amplifier, should have a very linear frequency response. Since the algorithms that are used for AEC model this chain with a linear adaptive filter there will be a larger system mismatch if nonlinearities are present on the playback audio (clipping, hard limiting, distortion, etc.). Any signal processing that must be done on the playback audio should be done in the digital domain before the signal is sent to the DAC and the AEC.

In addition to a daughterboard for the playback audio subsystem there will be some circuit design and fabrication involved in hooking up the loudspeaker, powering everything, and ensuring that the final product is compact. If time / resources permit it would also be useful to have the entire prototype as a single board design (baby / parent units can be constructed to differ only in firmware and possibly the labeling of external buttons).

These devices should be low power, with the baby unit being plugged into the wall, and the parent units being plugged in, but with an optional battery to extend range. While the final product likely wouldn't use Wi-Fi for its RF communication, the existence of Wi-Fi modules on the Osprey development kits, and the ubiquity of Wi-Fi make it a sensible choice for demo purposes.

Block Diagram:



Hardware/Electrical Tasks:

- Design DAC -> amplifier -> speaker sub system
- Create daughterboard for playback audio sub system
- Design power supply for the entire system
- Create power supply (maybe part of daughterboard)
- Understand linearity for playback audio system

Firmware/Software/Computer Tasks:

- Create overall software design for the monitor unit
- Interface wirelessly with the unit on a host PC
- Write any configuration parameters to the DAC
- Implement any machine learning solutions (time permitting)
- Understand I2S and other audio encoding / transport

Joint Tasks:

- Understand the basics of adaptive filtering
- Test the performance of XMOS echo cancellation with your hardware
- Find and port interesting sound event detection neural network (time permitting)

Composition of Team:

1 Electrical Engineer & 2 Computer Engineer (more AI focused route)

or

2 Electrical Engineer & 1 Computer Engineer (more hardware focused route)

Skills Required:

Electrical Engineering Skills Required:

- Familiarity with embedded hardware
- Familiarity with digital signal processing for audio devices (ELE314)
- Hardware communication protocols
- Hardware design

Computer Engineering Skills Required:

- C or C++ programming experience
- Software engineering classes preferable (CSC305 or similar)
- Wi-Fi communication protocols (ELE437)

Mutual Skills

- Familiarity with Linux operating systems
- Familiarity with version control systems (GitHub)
- Read documentation and ask questions / submit bugs

Anticipated Best Outcome's Impact on Company's Business, and Economic Impact

Achieving the ABO would give XMOS a very compelling demo to share with customers who either produce baby monitors or noise machines on the market today. This market should be much larger than the smart speaker market (when you remove first-party Echo/Home devices) and have steady growth. The enhanced AI functions will also allow companies to add compelling functionality just by updating firmware.

Broader Implications of the Best Outcome on the Company's Industry:

Smart speakers have been incredibly popular, but they are just the first product utilizing voice recognition to have major commercial success. Many devices could be voice enabled, or have artificial intelligence based on other audio events, but the trend has been to simply have devices that can interreact with an existing smart speaker. For example, the plug socket that lets you turn the power on or off with Alexa. The problems with this approach are:

1. Many of these devices have been rendered "dumb" by breaking API "updates"
2. Sending audio to the cloud to infer a few simple commands is a bad idea for both energy efficiency and network congestion
3. Having to "enroll" smart devices is a major barrier to adoption for customers

We hope that having a simple device that does ad job, does it well, and does not need the internet to do it will set a much-needed precedent in the consumer electronics industry.