

POLITENICO DI MILANO

DIPARTIMENTO DI ELETTRONICA, INFORMAZIONE
E BIOINGEGNERIA

Advanced Operating Systems: PROJECT REPORT

Embedded Neural Network

Authors:

Marina NIKOLIC
Michele SCUTTARI

Supervisor:

Dr. Federico TERRANEO

September 2019



POLITECNICO
MILANO 1863

1 Introduction

1.1 The problem to solve

The goal of this project is the recognition of specific sounds (whistles and hands claps) through the usage of a STM32 board and a neural network, deployed on the aforementioned board and acting as sound classifier.

1.2 Why neural networks?

For recognition and analysis of sound, AI techniques are used because of the complexity of the computations and the amount of noise present in the environment. Another important issue is that instances of the same sound have high variability due to different (yet homogeneous) sources. For example, think about words recognition: an effective application should recognise a word even if spoken by different people.

1.2.1 Which kind of NN?

The neural network used in this project is a sequential feed-forward one. Since the aim is to distinguish two different sounds (actually three, considering also the silence one), the NN is a classifier which input is the FFT of a time window and the output has one-hot encoding.

This simple model is expected to work because of the simplicity of the problem and the characterization of the two sounds.

1.2.2 How to implement NN on a board?

The **STM32cube.ai** package allows to compile a pre-trained neural network into a library to be called in the code.

1.3 Acronyms and Definitions

- **AI**: Artificial Intelligence
- **NN**: Neural Network
- **FFT**: Fast Fourier Transformation

2 Design and Implementation

2.1 Board Programming

2.1.1 Board description

The board in use is a **STM32F4 Discovery** (STM32F407VGT6) and is equipped with the open source operating system Miosix.

The peripherals used in this project are the microphone, the user button and the USART, although also the CRC module is enabled because needed by the ST's AI library.

2.1.2 Microphone

The board is equipped with a **MP45DT02** MEMS microphone, which produces a stream of 1-bit **Digital Pulse Modulation** (PDM) samples. The PDM samples are received using DMA and then converted to 16-bit signed **Pulse Code Modulation** (PCM) samples through the usage of a software low pass filter.

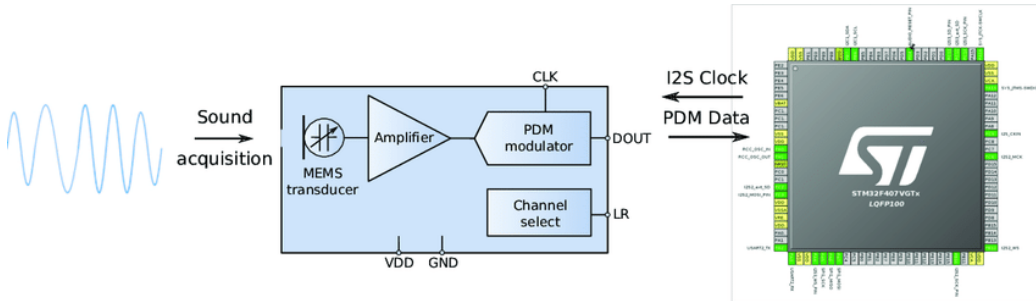
The PCM data is sampled at a frequency of 32 kHz, which is obtained by setting the *PLLI2SN*, *PLLI2SR* and *I2SDIV* registers to appropriate values (see chapters 7.3.23 and 28.4.4 of the datasheet for further details):

$$f_{(\text{VCO clock})} = f_{(\text{PLLI2S clock input})} * \frac{PLLI2SN}{PLLM} = 8 \text{ MHz} \frac{213}{8} = 213 \text{ MHz}$$

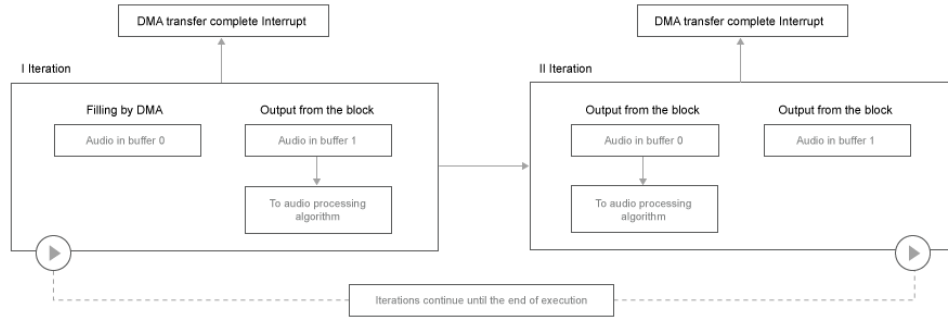
$$f_{(\text{PLL I2S clock output})} = \frac{f_{(\text{VCO clock})}}{PLLI2SR} = \frac{213 \text{ MHz}}{2} = 106,5 \text{ MHz}$$

$$f_{(\text{PDM})} = \frac{f_{(\text{PLL I2S clock output})}}{2 * I2SDIV * 8} = \frac{106,5 \text{ MHz}}{2 * 13 * 8} = 512 \text{ kHz}$$

$$f_{(\text{PCM})} = \frac{f_{(\text{PDM})}}{16} = \frac{512 \text{ kHz}}{16} = 32 \text{ kHz}$$



The driver also uses a double buffering strategy: one buffer (the “ready buffer”) contains the stable data that are sent to the network, while the other one (the “processing buffer”) is populated by the new incoming samples. When the processing buffer is full, they are swapped and thus the meanwhile collected data becomes ready.



2.1.3 FFT

The PCM samples are considered in blocks of 1024 values, called “windows”. Before the values are processed, they are normalized and passed through a **Hann function**, which makes sure that the discontinuity at the windows borders are eliminated¹. Then the FFT is computed and, being the output symmetric (because all the input has zero imaginary part), only half of the 1024 output bins are taken². Finally, the magnitude is computed on each single value.

Being the Cortex M4 equipped with the **Digital Signal Processor** (DSP), the FFTs are determined using the dedicated hardware.

2.1.4 User button

The user button press is handled by enabling the **EXTIO** interrupt.

The calling thread is paused while waiting for the press event, in order to allow other tasks to be executed meanwhile.

Moreover, a software debouncing feature is implemented in order to avoid

¹Shlomo Engelberg. *Digital signal processing: an experimental approach*. Springer Science & Business Media, 2008.

²Michael Parker. *Digital Signal Processing 101: Everything you need to know to get started*. Newnes, 2017.

multiple clicks due to the mechanical structure of the button. This is achieved by recognizing the first button press and ignoring further clicks for a pre-defined amount of time (200 ms), whose length is sufficiently long to discard false clicks and small enough to improbably ignore real clicks.

2.1.5 USART

The communication between the board and the client is done through USART connection ($PA2 = TX$, $PA3 = RX$), configured with a speed of 115200 bps. The communication is set in raw mode by disabling the canonical mode and not converting controls (i.e. \hat{C} , \hat{Z}) into signals.

2.1.6 Neural network library

The neural network library is generated by the **X-CUBE-AI** expansion pack for STM32CubeMX. This tool automatically converts the pretrained Keras model into a C library that can be called by the main program.

Instead, the core of the network runtime is provided by ST as a static closed source library (*neural_network.a*).

2.1.7 Issues

- The suggested IDE Netbeans gives a lot of troubles in the development, such as missing code completion and unrecognized definitions. So the Miosix build system has been converted to CMake, in order to allow the usage of modern IDEs such as CLion.
- The Miosix toolchain has a bug in the linking phase, resulting in the network runtime library to block in an infinite loop when its functions are called. To solve this issue, the linking process has been executed using the standard ARM linker.

2.2 Network Training

The software can be compiled in two different ways, by enabling or disabling the *TRAINING* variable definition. When the variable is defined, the neural network is not executed and the FFT samples are directly transferred to the client. This way, it is possible to obtain the data to separately train the network. When the network is trained and uploaded, the variable definition

can be removed in order to enable the normal operating mode, in which the neural network runs and outputs the recognized sounds.

2.3 Testing

3 Experimental Results

4 Conclusions

4.1 ...

4.2 Possible Use Cases

4.3 Future Improvements

5 Some L^AT_EX Examples

5.1 Sections

Use section and subsection commands to organize your document. L^AT_EX handles all the formatting and numbering automatically. Use ref and label commands for cross-references.

5.2 Comments

Comments can be added to the margins of the document using the `todo` command, as shown in the example on the right. You can also add inline comments too:

This is an inline comment.

Here's
a com-
ment
in the
mar-
gin!

5.3 Tables and Figures

Use the `table` and `tabular` commands for basic tables — see Table 1, for example. You can upload a figure (JPEG, PNG or PDF) using the files menu. To include it in your document, use the `includegraphics` command as in the code for Figure 1 below.

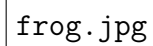
A square box with a thin black border. Inside the box, the text "frog.jpg" is written in a monospaced font, positioned towards the left side of the box.

Figure 1: This is a figure caption.

Item	Quantity
Widgets	42
Gadgets	13

Table 1: An example table.

5.4 Mathematics

L^AT_EX is great at typesetting mathematics. Let X_1, X_2, \dots, X_n be a sequence of independent and identically distributed random variables with $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 < \infty$, and let

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_i^n X_i$$

denote their mean. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$.

5.5 Lists

You can make lists with automatic numbering ...

1. Like this,
2. and like this.

...or bullet points ...

- Like this,
- and like this.

We hope you find writeL^AT_EX useful, and please let us know if you have any feedback using the help menu above.