# USB Synchronous Multichannel Audio Acquisition System

David Abran-Côté, Massinissa Bandou, Alexandre Béland, Gabriel Cayer, Sébastien Choquette, Frédéric Gosselin, Francis Robitaille, Diallo Telly Kizito

University of Sherbrooke

***Abstract* — This document presents a technical description a synchronous multichannel audio acquisition system designed for mobile robotics. This system includes a card of 8 USB synchronous channels and a set of 8 active and differential microphones. The USB card provides also an analog stereo output. It allows a power supply by USB or by an external input. The electrical consumption is specified not to exceed 2.5 W. Design of the different power supplies on the card was done in order to respect those specifications.**

**As the system is designed for intelligent mobile robotics, a considerable quantity of noise can interact with the system. This problem is solved by using differential signals for data transport with a thorough design of power supply and analog parts.**

**The card is compatible with the following operating systems: Windows, Linux and Mac OS. A communication using the Audio Class 2.0 of the USB 2.0 High-Speed norm was used to transfer data to the computer. That will avoid creating a driver for each operating system.**

***Keywords* – Synchronous data acquisition, artificial audition, audio codecs, sound sources localization, open source, mobile robotics, differential signals, USB Audio Class 2.0, USB 2.0 High Speed, XMOS**

## I. CONTEXT

### A. *Introduction and Motivation*

The laboratory of mobile robotics and intelligent systems intRoLab of University of Sherbrooke[1] design mobile robotics adapted to different kinds of environments. For this purpose, the laboratory designed the ManyEars [2] system, an artificial hearing software allowing a mobile robot to locate itself and locate sound provenance in real time.

Intelligent mobile robotics systems will be improved if the robot is able to locate and follow sound provenance in real time. This capacity can help to locate a person or an interesting event in the environment.

To give this capacity to a software system, the challenge is not just to locate simultaneous sound provenance, but to keep track of their evolution with time. The ManyEars project offers a new way to locate the sound sources and a strong following method using 8 microphones.

The tests on a mobile robot show how the algorithm allows to locate and to follow in real time multiple sound sources within 7 m. These new capacities allow the robot to react naturally.

Using such a system, the robot must be equipped with an 8 input synchronous audio acquisition system. The biggest constraints for mobile robotic are the power consumption and the occupied space. The reduction of those aspects will improve the efficiency of mobile robots.

The idea of the project is to develop an audio acquisition card with 8 inputs and one stereo output and 8 preamplifier microphone cards. These two kinds of cards must be cheap to produce, have small dimensions and require less power to operate then the actual professional audio cards used on the market to locate sound provenance for mobile robotics and others applications.

Considering the availability of the USB 2.0 port versus FireWire or Peripheral Component Interconnect (PCI), the project will use USB 2.0 High-Speed for data transfer. Likewise, the transfer rate is at 480 Mbits/s and it is good enough to transfer data for 8 microphones and one stereo output.

This paper presents a description of the developed system. The project specifications, global architecture, materials and software of the system are discussed. This paper ends with a description of tests and results and follows with an analysis and a conclusion.

### B. *State of the Art*

Given that there is no audio acquisition card of 8 inputs intended for robotic applications, the robots are using professional audio cards used for studio recording. These cards provide a large number of functionalities that are not used for the robotic domain (sound effect, integrated mixing, optical inputs/outputs, S/PDIF, MIDI, more than two analogs outputs). Often they are expensive, large and power hungry. With the power consumption constraint and the small sizes that the mobile robotics requires, the actual professional audio cards are not desirable.

Cards like MULTIFACE II from RME Company equipped with PCI interface and the Ultra-Lite-mk3 Hybrid card from MOTU equipped with FireWire interface are compatible with the ManyEars system. Those cards cost around 1000$, measure 20 cm x 12 cm x 6 cm and consume 12 W. In addition, this power doesn't include the 8 microphones' consumption.

Until now, the available USB synchronous audio acquisition interfaces use predominantly the USB Audio Class 1.0 norm that allows only two microphones or a proprietary USB protocol. Those that use the 2.0 version are too voluminous for this application.

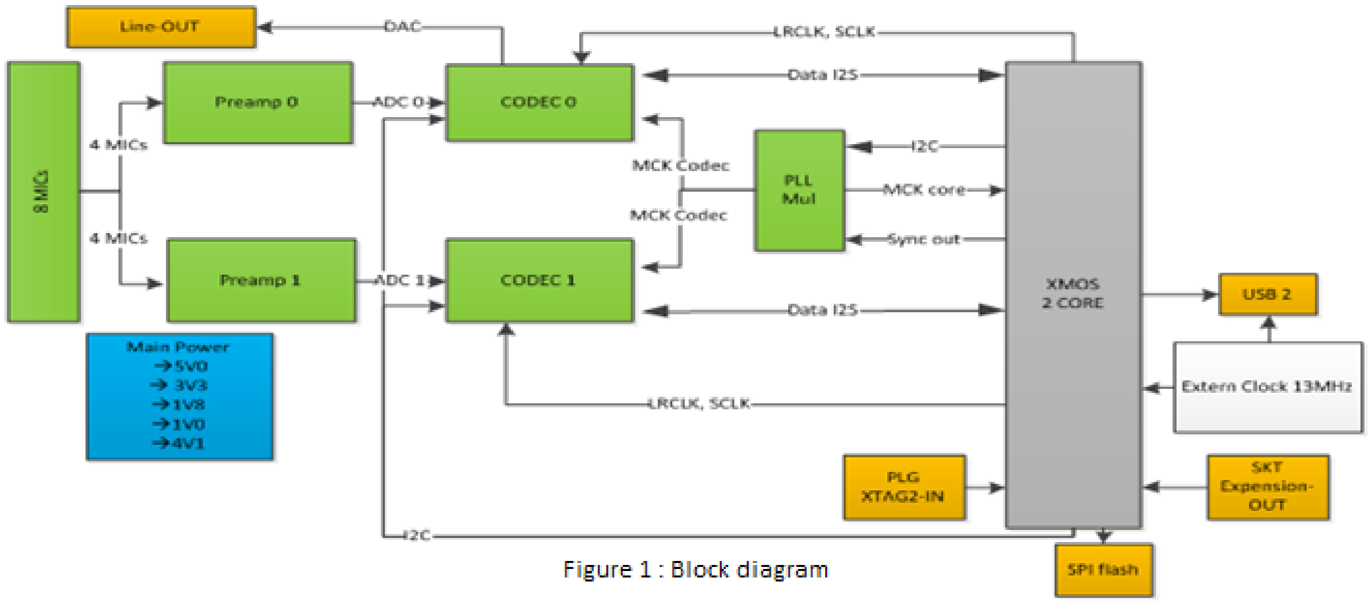

Figure 1 : Block diagram

The definition of the audio class norm is applicable to all devices that have integrated functions for voice and audio manipulating.

Audio acquisition products under the USB Audio Class 2.0 norm are practically new on the market [3]. The improvement compared with Audio Class 1.0 allows increasing the capacity in terms of channels, resolution and sampling frequency. Furthermore, most mobile robots use a personal computer (PC) as central processing system. Linux is the most popular operating system because it is the only one that can allow using Robotic Operating System (ROS) [4].

Until now, a few manufacturers provide devices using USB Audio Class 2.0 but interest seems to appear with reduction in popularity for some products using FireWire and PCI.

## II. Description of the System

### *A. Request for Proposal*

The main objectives are the design of an USB synchronous audio acquisition card with 8 inputs and one analog stereo output and 8 pre-amplification microphone cards. The acquisition card has to allow power supply from the USB port or an external source. The microphones have to be supplied by the main USB card. Total consumption of the system must be under 5 W to allow power supply by USB in overcharge (5 V at 1 A). The card has to measure approximately 7 cm by 10 cm and 3 cm of height.

Regarding sound quality, the system needs to have between 12 to 48 bits of resolution and have a range of sampling frequency between 48 and 192 kHz.

### *B. Global Architecture*

The project presents a simple architecture made up with three principal blocs. The first one is USB communication, the second one is the microcontroller and finally the audio codecs. Those blocs are shown in the Figure 1.

### *C. Important Modules*

#### *1) USB Communication*

The card has to communicate by USB using USB 2.0 protocol. The audio inputs and outputs are supported by USB Audio Class norm. The 8 input channels are supported by version 2.0 and more [3].

#### *2) Microcontroller*

The system is using a XMOS microprocessor with a second generation [5] « L » architecture. XMOS offers a development kit that meets the needs of the project in terms of hardware and software. Indeed, this development kit includes a software implementation of the Audio Class 2.0 norm not available on the market.

The main interest with XMOS is its architecture which is based on events [6] instead of processor cycles like most microcontrollers [7].

The advantage of XMOS event programming is the subsequent improved power management. The processor allows dynamic power management according to the frequency of use.

A software disadvantage of this processor is that the number of available threads and their execution times are fixed by hardware. Consequently, the processor might not be used at its full potential because each defined thread has the same capacity regardless of the quantity of processed data. .

The assembly of the double core processor involves additional costs due to its physical format QFN-124, which requires specialized equipment. Moreover, the assembly implies a validation test using X-rays.

#### *3) Audio acquisition*

To improve noise rejection of the system, the use of differential signals is very advantageous to consider. The codec CS42448 from Cirrus Logic [8] has differential inputs and tools for noise filtering and allow the use of I2S protocol

[9] for audio transfer and I2C protocol [9] for the codec configuration. Furthermore, it was shown that the Cirrus Codec and XMOS processor are compatible by the XMOS audio 2.0 development kit.

To reduce crosstalk in the microphone cables, analog filters proposed by Cirrus Logic for the codec are placed at the differential inputs.

### D. *Hardware*

#### 1) *Power Supply*

The synchronous multichannel audio card was designed to be powered either by USB or an external power supply. To power the card by USB has the benefit of simplifying the integration by using the existing circuit and wiring. However, using an external power supply gives a better control of the power output and avoids the overloading of the USB hub. The power delivered by an external power supply can also be higher (up to 7.5 W) than by USB (up to 2.5 W). Using the USB technology yields stricter power specifications as the maximum current is 500 mA and the voltage can vary between 4.4 V and 5.25 V.

The range of voltage of an external power supply goes from 7 to 36 V to allow a large flexibility in the voltage provided by the robot. To minimize the voltage and current ripple, the input and output of the regulation circuit pass through tight filtering.

When the external power supply is turned on, a power multiplexer (TPS2111) will automatically detect this source and choose it over USB. This design feature allows an easy transfer between the power supplies.

To maximize the signal-to-noise ratio and obtain the resolution of 16 bits required by the specifications, a high analogue reference voltage is needed. It is accomplished using a low-dropout voltage regulator (LDO). Using this method, the analogue reference voltage can be as high as 4.3V even when the USB voltage is at its minimum (4.4 V). Consequently, the ADCs acquire the analogue signal with a resolution of 16 bits and 66 µV per division [11]. The 4.3 V LDO (TPS79501) was also chosen for its small size, good immunity to input power supply noise and rejection ratio (PSRR).

A digital reference voltage of 3.3 V must also be produced on the board to power the codecs, the USB PHY, the phase locked loop (PLL) and the inputs/outputs of the XMOS processor. The conversion to 3.3 V is achieved by a step-down DC-DC converter (NCP1421E).

Two other converters are required. A LDO converter (NCP599SN18) is used to produce a voltage of 1.8 V also to power the USB PHY and a DC-DC converter (FAN2011) is used to power the XMOS cores.

#### 2) *Microphone card*

As mentioned earlier, the sound card is designed to acquire audio signals from 8 microphones placed at predetermined locations (to find the position of sound sources). Each microphone must then have its own preamplifier card which is powered by the main acquisition card. Considering the current conditions of operation, the physical dimension of the microphone cards needs to be minimized. Figure 2 shows a picture of one microphone card. Its dimensions are 2.3 cm by 2.3 cm. The major electronic components are the following; the microphone (of electret type), the preamplifier (TS472) and the RJ11 connector (on the back side). The selection criteria of the preamplifier are a high signal-to-noise ratio, differential input and output and a closed loop gain of approximately 40 dB (to obtain a peak-to-peak amplitude of 4.3V at the output). The RJ11 connector was chosen more importantly because of its parallel insertion to avoid the power line to make contact with the data line. The connector also needed to have a latch mechanism. The RJ11 and XLR mini met those requirements and RJ11 was selected due to its low cost and easy use.

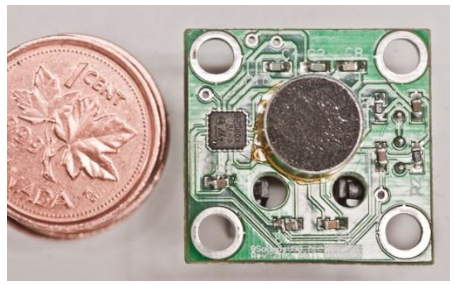

Figure 2: Microphone card

#### 3) *Analogue Signal Conditioning and Integrity*

To preserve the audio signal integrity is one of the main targets of a sound card. In this section, the different measures to reach this goal are described.

Firstly, the analogue signal coming from the microphones is transmitted in differential mode unto the codecs. Differential mode was preferred to single-ended signaling because of its better noise immunity and because it increases the dynamic range of the analogue/digital converter of the codec. Since the acquisition card was designed to be operated under mobile robotics conditions, many sources can induce electromagnetic interference in the transmitted signal. By using twisted pairs for signal transmission (will distribute the interference) and a differential amplifier (will reject the common mode noise), the effect of electromagnetic interference on the audio signal will be minimized and thus its quality increases.

Furthermore a differential signal will produce higher amplitude compared to a single-ended signal taking into account the same reference voltage, which will increase dynamic range. According to the codec's datasheet, dynamic range goes from 102 dB (single-ended signaling) to 105 dB (differential signaling) which yields a 41% increase.

Keeping signal integrity in mind, the microphone preamplifiers were placed the closest possible to the microphones to reduce effects of electromagnetic interference and to preserve a good signal-to-noise ratio.

At the preamplifier's output, the signal goes through a differential audio amplifier (LME49726) that biases the signal for the codec's input and also filters the signal to avoid aliasing. This amplifier was chosen because it is especially intended for audio and it uses a single power supply. The band-pass filter has a flat frequency response in the audio band and has a rejection of 20 dB at the lowest sampling frequency. The configuration of the anti-aliasing filters is the one suggested by the manufacturer of the codecs (Cirrus).

*4) Codecs*

The main purpose of the codecs if the signal conversion from analogue to digital (and the other way around) and I2S (Inter IC Sound) serial protocol management to exchange data with the XMOS processor. To meet the specifications, each codec uses four out of six 24 bits converter inputs and the sampling frequency can go up to 192 kHz.

The various codec settings (converters' operation mode, sampling frequency, power and more) are set by the XMOS processor through a serial communication I2C (Inter Integrated Circuit) link.

The codec was selected for its analogue reference voltage range which allows using 4.3 V versus 5 V for other codecs and also for its dynamic range of 105 dB in differential mode. Moreover, the codec is part of a reference design from XMOS Company. Codecs were used instead of only an ADC with 8 channels because the card needed to have an audio output which required an integrated DAC.

*5) PCB (Printed Circuit Board) Design*

To meet the specifications with regard to size, it is necessary to have a logical component placement.

The PCB has four layers to allow a ground plane and a power plane. A single ground plane is preferred to a separated ground plane to avoid creating voltage between the planes which could produce instability in the analogue power supply. Therefore, the PCB was divided into two sections, a first one intended for the digital domain and a second one for the analogue domain.

The power plane was divided into three sections (as shown by Figure 3). A first one for the digital domain (in red), a second one for the analogue domain (in orange) and a last one to power the XMOS cores (in blue).

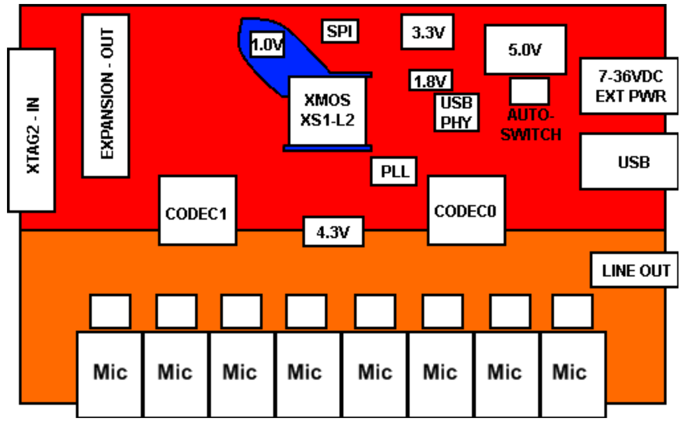


Figure 3: Power Supply Plane

The digital section of the plane is used to power with 3.3 V the digital part of the two codecs, USB PHY, phase locked loop (PLL) and XMOS cores. This section covers approximately half the PCB.

The analogue section of the plane is used to power with 4.3 V the analogue part of the codecs, differential amplifiers, and for the two audio outputs.

The section which powers the XMOS cores with 1 V is located directly under the integrated circuit of the XMOS processor for an easier access to inner pins.

Dividing the power plane in an analogue and digital domain improves the signal-to-noise ratio (SNR) by isolating the high frequency noise generated by digital components from the analogue domain.

To provide the best transmission possible of the input differential signals to the codecs, it is advised to keep a constant length between each differential pair. In order to follow standard engineering practices, there is a large number of decoupling capacitors to transfert power to the different integrated circuits of the card. This will insure proper operation of the card.

*1) Software Components*

*1) USB Audio Class 2.0*

The prototype is compliant with Audio Class 2.0 of the USB 2.0 norm to allow using the card in any operating system without installing a specific driver.

The XMOS reference design provides the implementation of Audio Class 2.0. This implementation is available under a particular license that prevents us from releasing the source code. However, it's possible to take a confidential agreement with XMOS to get access to the source code (NDA). Nevertheless, we decided to use this implementation and adapt it with our design.

To avoid creating a new driver, the generic driver available with Audio Class 2.0 is used. This option also reduces the amount of support to be provided after the project.

This is a class of USB devices supported by Linux (Ubuntu 10.10 and up) and Mac OS. It's not supported by Microsoft (Win7 SP1). However, a driver is available by a third party. It is just a question of time before Microsoft will provide the support.

*2) Event Based Architecture*

The « L » architecture of XMOS microprocessor family is of particular interest for mobile applications. The eight threads for each core are independently activated by events instead of running all the time according to the system clock. This brings a considerable saving of energy because only active threads consume energy. A particular attention is given to the synchronization between the XMOS core and the different threads. An external clock is used to synchronize the codecs and XMOS cores for the data acquisition [12-13].

## III. RESULTS, VALIDATION AND TESTS

Before performing the tests, the various power supplies needed to be validated. With test point support, a reading of tensions produced by power components allowed to ensure that all power nodes were carrying the proper voltage.

A test concerning the power multiplexor allowed confirming that the system can switch without problem from USB power supply to external power supply. It consists in unplugging the external power supply when USB is connected and validating if the card is powered and still working properly.

The first series of tests were followed by the recording of the analog inputs of the system on a computer that was plugged into the USB port.

All tests confirmed that the system meets the goals mentioned in the specifications of the project. The audio acquisition of the 8 analog inputs and the analog stereo output are functional as well as the audio data transfer from the USB port. Those results are valid on Linux, Mac and Windows.

### *A. Validation Method and Selected Tests*

This section presents the key tests made to characterize the functions of the USB synchronous multichannel audio acquisition system.

*1) Electrical Consumption*

The electrical consumption has to be verified for the external power supply and USB when the system is plugged with and without the 8 microphones. The external power supply can change from 7 to 36 V. The consumption will thus be tested for different external voltages (7, 12, 24, and 36 V) and for USB power only. The total consumption of the system must not exceed 5 W.

*2) Latency between Acquisition Channels*

The same signal is transmitted to the 8 microphone inputs installed at the same distance from an omnidirectional sound source (a glass). A recording of the eight tracks was done with Electroacoustic Toolbox [14] software. The integrated oscilloscope was used to confirm that the offset between the inputs is not higher than an acquisition cycle.

*3) Effective Number of Bits (ENOB)*

The ratio between the maximal noise in the inputs of the audio codec versus the reference voltage of the same codec determines the loss in effective bits in the least significant bit (LSB) zone. The effective number of bits desired is 16 bits but the operating threshold for robotics applications is 12 bits [15]. To perform the test, a frequency generator produces a signal with fixed frequency to one of the analog inputs of the system. A slight increase of the signal amplitude is done. We take the smallest amplitude variation (voltage) that changes the digital signal by one bit. Then we divide the range of voltage (4.33V) by the voltage variation to define the effective number of bits (ENOB).

*4) Mean Noise Floor*

This section allows quantifying the mean noise floor of the system. To do this, an analog input of the card will receive a signal from the generator and another input will act as the control group. The difference between the two inputs allows calculating the mean noise floor of the system.

### *B. Quantitative Results*

The following table displays results of important tests performed on the acquisition system.

| **Size** | **Length (mm)** | **Width (mm)** | **Height (mm)** |
|---|---|---|---|
| Sound card | 125 | 74 | 15 |
| Microphone | 23 | 23 | 15 |
| | | | |
| **Power supply** | **Voltage (V)** | **Current (A)** | **Consumption** |
| **External (without microphones)** | **7** | 0.332 | 2.324 W |
| | **12** | 0.198 | 2.376 W |
| | **24** | 0.114 | 2.736 W |
| | **36** | 0.0867 | 3.121 W |
| **External (with microphones)** | **7** | 0.343 | 2.401 W |
| | **12** | 0.208 | 2.496 W |
| | **24** | 0.118 | 2.843 W |
| | **36** | 0.0884 | 3.182 W |
| **USB (without microphones)** | **5** | 0.434 | 2.1675 W |
| **USB (with microphones)** | **5** | 0.449 | 2.246 W |
| **Maximum with external power supply** | | | 3.182 W |
| **Maximum with USB power supply** | | | 2.246 W |
| | | | |
| **Latency** | **Maximum delay between 2 audio channels** | | |
| | 16 µs | | |
| | | | |
| **ENOB** | **Effective number of bits** | | |
| | 22 Effective bits | | |
| | | | |
| **Noise Floor** | | | |
| **Mean** | -132 dBV | | |
| **Maximum** | -112 dBV | | |

The card's size is slightly larger than requested in the initial specifications (i.e. 125 mm × 74 mm against 100 mm × 70 mm) however this small oversize has negligible impacts on the quality of the final product.

As one can see, the maximum power consumption of the sound card is a lot smaller than in the original specifications (i.e. approximately 3.2 W against 5 W). Maximum power consumption occurs while using the maximal external voltage of 36 V. If the external voltage is reduced to the minimal voltage of 7 V, consumption becomes 2.5 W. The variation of consumption is caused by the reduction of power loss in the DC-DC converter at the power supply's input.

When the system is powered through USB, maximum consumption is approximately 2.3 W. Considering this result, the sound card can be powered without the need to use USB in overcharge (a current of 500 mA or less is sufficient).

There is a maximum delay of 16 µs between two acquisition channels. Thus, for a sampling frequency of 96 kHz, the maximum delay between two channels is higher than one acquisition cycle. However, for a target sampling frequency of 48 kHz, the time delay is smaller than one acquisition cycle. That being said, those results may contain an error margin because an ideal test requires placing the microphones at equal distances from the source which is hard to do in practice.

The initial goal of 12 effective bits required from mobile robotics criteria is largely met with 22 bits which means that the sound card can acquire data with excellent precision. The mean value of the noise floor obtained during the tests is located is the same range as the one specified by the codec. Thus, the card has good noise rejection.

## IV. Conclusions and Future Work

### A. Conclusion

The USB synchronous multichannel audio acquisition system meets the objectives on the project specifications. The system allows the acquisition of 8 analog inputs synchronously at the frequency of 48 kHz. The system provides a resolution of 22 effective bits which exceed the specification of 16 bits. The final sound card has a size slightly larger than the targeted size.

The system allows an external or USB power supply and the maximal power consumption using USB is 2.3 W and that meets the requirement on consumption.

Finally, the system is fully operational and it can be used by the IntRoLab mobile robotics and intelligent systems laboratory for the University of Sherbrooke with the ManyEars system.

### B. Future Development

The project allows an opening for possible future developments with XLINK ports [16] having input and output functionality present on the acquisition card. These allow the addition of microcontrollers in a serial topology and thus can increase the number of parallel operation in the system. The actual processor present on the prototype is little used which leaves room for a future development.

One future benefit for IntRoLab could be to include partially or totally the position calculations that are necessary for the operation of ManyEars software on the acquisition card.

## References


[1] IntRoLab. (2009, August 26). [Online]. Available : http://www.introlab.gel.usherbrooke.ca

[2] ManyEars. (2010, November 19). [Online]. Available : http://www.manyears.sourceforge.net

[3] USB. (2006, Mai 31). Audio Devices Rev. 2.0 Spec and Adopters Agreement [Online]. Available : http://www.usb.org/developers/devclass_docs/Audio2.0_final.zip

[4] Robot Operating System (ROS), [Online], Available : http://www.ros.org/wiki/

[5] XMOS. (2011, Mars 17). Description of the "L" architecture. [Online]. Available : http://www.xmos.com/technology/architecture

[6] XMOS. (2011, Mars 17). Event based core. [Online]. Available : http://www.xmos.com/technology/xcore

[7] XMOS. (2011, Mars 17). XMOS vs FPGA. [Online]. Available : https://www.xmos.com/download/public/XMOS-vs-FPGA-Whitepaper%281%29.pdf

[8] Cirrus Logic. (2007, November). CS43448. [Online]. Available : http://www.cirrus.com/en/pubs/proDatasheet/CS42448_F3.pdf

[9] NXP. (1996, June 5). I2S Specifications. [Online]. Available : http://www.nxp.com/acrobat_download2/various/I2SBUS.pdf

[10] NXP. (2000, January). I2C Specifications. [Online]. Available : http://www.nxp.com/documents/other/39340011.pdf

[11] Clayton R. Paul, « Electromagnetics for Engineers » Wiley, 2004.

[12] XMOS (2011, January 14). USB Audio Software Design Guide. [Online]. Available : https://www.xmos.com/published/usb-audio-software-design-guide

[13] XMOS. (2010, June 29). USB Audio 2.0 MC Ref Design Schematics. [Online]. Available : https://www.xmos.com/published/usb-audio-20-mc-ref-design-schematics

[14] Electroacoustic Toolbox 3.0. [Online]. Available : http://www.faberacoustical.com/products/electroacoustics_toolbox/

[15] Texas Instrument. (2009, Mai). The Basics of the Effective Number of Bits. [Online]. Available : http://e2e.ti.com/videos/m/analog/97246.aspx?DCMP=hpa_dc_general&HQS=NotApplicable+OT+bitselection-pr

[16] XMOS. (2011, Mai 7). XS1-L2 124QFN Datasheet. [Online]. Available: https://www.xmos.com/published/xs1-l2-124qfn-datasheet