USB Synchronous Multichannel Audio Acquisition System

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

University of Sherbrooke

***Abstract* — This paper is a technical description of a XMOS XS1-L2 based synchronous multichannel audio acquisition system designed for mobile applications, or more specifically mobile robots. The design is made available through an Open Source Hardware license, allowing easy modifications and sharing. The system includes an 8 channels USB sound card coupled with 8 active and differential microphones. The sound card also provides an analog stereo output. It can be powered by either USB or an external wide range (7-36V) DC power supply. The electrical consumption over USB does not exceed 2.5 W, conforming to the USB 2.0 power specifications enabling the system to be solely powered by the USB bus. Differential codecs and amplification circuitry are used, allowing operation in noisy setups. Finally the sound card is compatible with the following operating systems: Windows, Linux and Mac OS. Audio data is transferred to the computer using the USB 2.0 High-Speed Audio Class 2.0 standard avoiding designing a different 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*

IntRoLab [1] is a research laboratory based in Sherbrooke, Québec, Canada. IntRoLab is pursuing the goal of studying, developing and integrating technologies for the design of autonomous and intelligent systems. Research activities involve software and hardware design of mobile robots, embedded systems and autonomous agents. To enhance the robot perception, an artificial audition system named ManyEars [2] has been developed to locate, track and separate sound sources in real time with an array of eight microphones. This capability allows the robot to locate a person or an interesting event in its environment. Tracking the sound sources over time also lets the robot follow different speakers in motion. ManyEars is also able to separate the individual speech of multiple simultaneous speakers. Tests show that the system performs well with up to four sound sources located within seven meters.

Since the algorithm uses an array of eight microphones, an eight channel input synchronous audio acquisition system is needed. Two major challenges are present in mobile robotics: 1) power consumption must be low and 2) physical dimensions of the system must be optimized. The goal of this project is thus to design a small, low-cost and low-power audio acquisition card with microphones. To accommodate the needs of the ManyEars algorithm, the card has eight input channels and one stereo output channel and is connected to eight preamplifier microphone cards, all powered by the main audio card.

The USB 2.0 High-Speed interface is used for data transfer as it is more commonly available compared to other interfaces such as FireWire or Peripheral Component Interconnect (PCI) and more portable. The USB 2.0 transfer rate reaches 480 Mbits/sec, which is sufficient to transfer the raw (uncompressed) data of 8 microphones and one stereo output.

This paper presents a description of the designed system. The project specifications, global architecture, materials and software, tests results and analysis are presented.

B. *State of the Art*

Robot artificial audition systems use studio sound card since there is no audio acquisition card with eight input channels specifically designed for robotic applications. These professional cards often provide unnecessary functionalities (sound effects, integrated mixing, optical inputs/outputs, S/PDIF, MIDI, numerous analogs outputs, etc.). Moreover, they are large, expensive, and require a significant amount of power. Clearly, these devices do not meet the requirements for the current application.

Previously, sound cards like the RME MULTIFACE II (PCI) and the MOTU Ultra-Lite-mk3 Hybrid (FireWire) were used at IntRoLab with the ManyEars system. These cards cost around US $500-$1000, consume too much power (10W+, without the power consumption of the eight microphones), provide inadequate or complicated connectivity (power supply + PC interface) and dimensions are not suitable for robotic applications.

So far, the USB synchronous audio acquisition interfaces use mostly either the USB Audio Class 1.0 standard, which is limited to two input channels, or a proprietary USB protocol. A few other systems use the USB Audio Class 2.0 standard but these are not well suited for this application.

The definition of the audio class norm is applicable to all USB computer compatible devices that have integrated functions for voice or audio manipulation.

Recently, the USB Audio Class 2.0 standard was introduced in few audio acquisition devices [3]. The improvements over the Audio Class 1.0 standard include more channels and better sampling resolutions and frequencies. Most mobile robots use a personal computer (PC) as their central processing system. Linux is the most popular operating system as it exclusively supports the Robotic Operating System (ROS) [4].

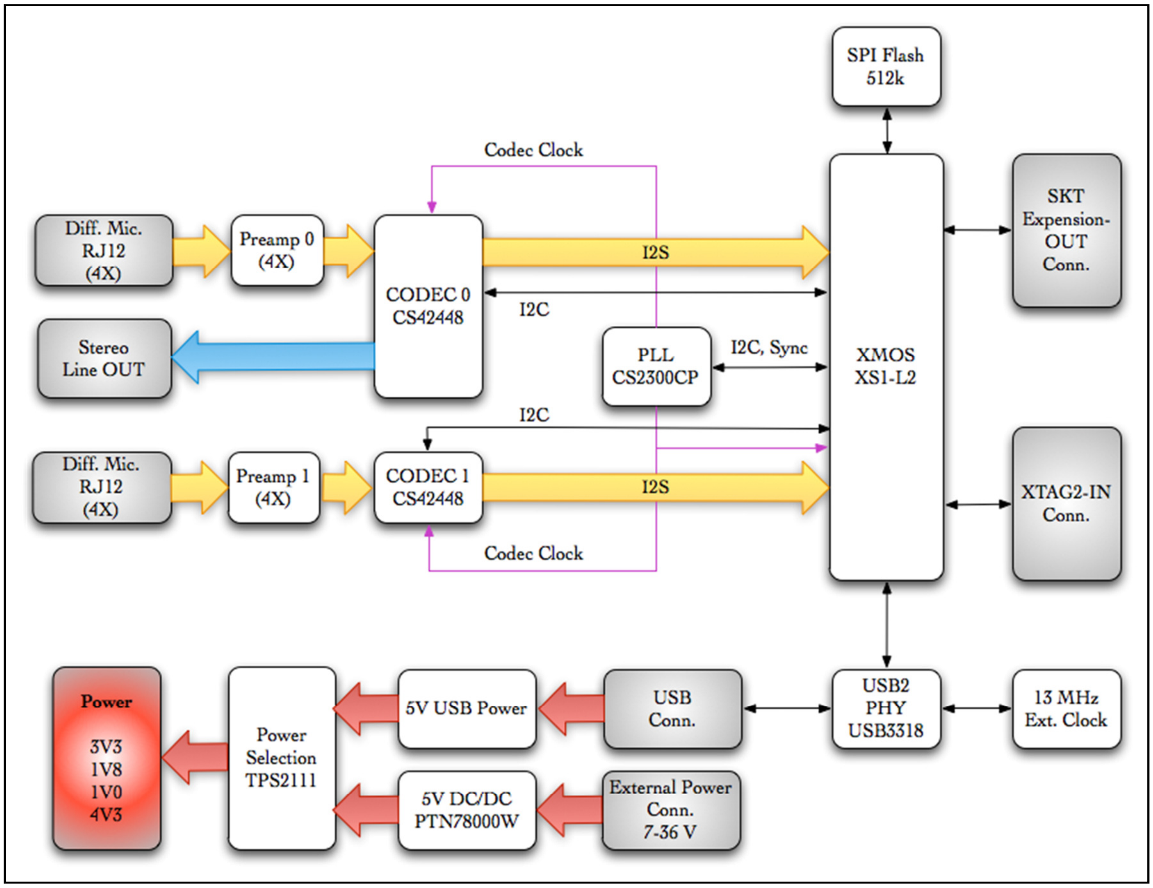

Figure 1: System diagram

Until now, a few manufacturers provide devices using USB Audio Class 2.0 but interest is growing with the development of systems with multiples channels that needs more resolution and flexibility.

II. Description of the System

A. *Request for Proposal*

The main objective is to design a USB synchronous audio acquisition card, which has eight input channels and one analog stereo output channel, and eight pre-amplification microphone cards. The acquisition card needs to be powered by the USB port or an external source. The microphones have to be powered by the main USB card. The total consumption of the system must not exceed 5W when powered by the USB interface in overcharge (5V at 1A). The card dimensions should be approximately 7cm x 10cm x 3cm. Signal resolution should be a minimum of 12 effective bits and the sampling frequency should range from 48 to 192 kHz.

B. *Global Hardware Architecture*

The hardware architecture, shown at Figure 1, is divided in three main components: the USB communication, the microcontroller and the audio codecs. These components are shown in Figure 1.

C. *Components*

1) *USB Communication*

The card has to communicate by USB using the USB 2.0 protocol. The audio input and output channels are driven by the USB Audio Class standard, which supports eight input channels since version 2.0 [3].

2) *Microcontroller*

The system uses a XMOS microprocessor with second generation « L » architecture [5]. A reference design is available and meets the needs of the project in terms of hardware and software. This reference design also includes a software implementation of the Audio Class 2.0 standard, which is not available with other processor manufacturer than XMOS.

The XMOS architecture is promising as it is based on events [6] instead of processor cycles like most microcontrollers [7]. Event programming improves power management as the processor dynamically manages power according to the frequency of the events. On the other hand, this architecture limits the number of available threads and their execution times are fixed by the 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 programmer must then take the XMOS architecture into account to optimize power consumption and performances.

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

3) Audio acquisition

Differential signals are used to reduce noise. The codec CS42448 from Cirrus Logic [8] has differential inputs, provides tools to filter noise and supports both the I2S protocol [9] for audio transfer and the I2C protocol [9] for the codec configuration. Furthermore, experiments with the XMOS audio 2.0 development kit confirmed that the Cirrus Logic Codec and XMOS processor are compatible. Some analog filters proposed by Cirrus Logic are added to the differential inputs to reduce crosstalk in the microphone cables.

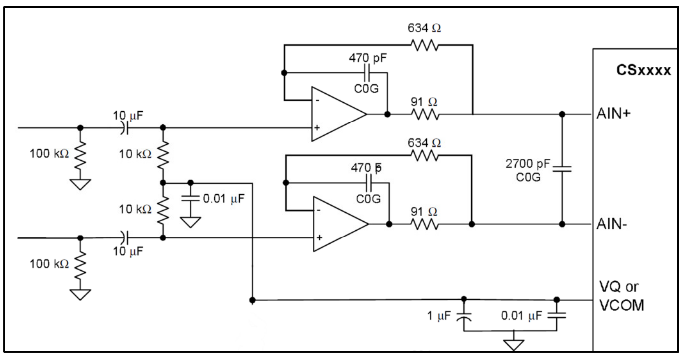

Figure 2: Differential input buffer

D. Hardware

1) Power Supply

The synchronous multichannel audio card is designed to be powered either by USB or an external power supply. The first scenario simplifies the integration as the existing circuitry and wiring are used. However, an external power supply improves the controllability of the power output and avoids overloading the USB hub. Powering the device from USB is limited to 2.5W and yields to stricter power specifications as the maximum allowed current is 500 mA and the voltage randomly ranges from 4.4V to 5.25V.

The voltage of an external power supply ranges from 7V to 36V, which allows different input voltages provided by the robot batteries. To minimize the voltage and current ripples, the input and output signals of the regulation circuit are filtered.

When the external power supply is turned on, a power multiplexer (TPS2111) automatically detects this new source and disables the USB power source. This allows hot plugging of power sources.

To maximize the signal-to-noise ratio and to obtain the 12 effective bits resolution previously specified, a high analog reference voltage is needed. This is done with a low-dropout voltage regulator (LDO) that ensures that the analog reference voltage stays at 4.3V when the USB voltage is at its minimum voltage (4.4V). Consequently, when the ADCs acquire the analog signal with a resolution of 16 bits, we have 66 µV per division [11]. The 4.3V LDO (TPS79501) is also chosen for its small size, good immunity to input power supply noise and rejection ratio (PSRR).

A digital reference voltage of 3.3V is also required to power the codecs, the USB PHY, the phase locked loop (PLL) and the input/output ports of the XMOS processor. The conversion to 3.3V 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 1.8V 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 eight microphones placed at fixed known locations. Each microphone needs its own preamplifier circuit board, which is powered by the main acquisition card. The physical dimension of these cards needs to be minimized. Figure 2 shows a picture of one 2.3cm x 2.3cm microphone card. The major electronic components include the microphone (of electret type), the preamplifier (TS472) and the RJ11 connector (on the back side). The preamplifier is selected since it has a high signal-to-noise ratio, differential input and output channels and a closed loop gain of approximately 40dB (to obtain a peak-to-peak amplitude of 4.3V at the output). The RJ11 connector is proposed since parallel insertion prevents the power line from making contact with the data line. The connector should also have a latch mechanism. The RJ11 and XLR mini both meet these requirements and RJ11 is selected due to its low cost and easy use.

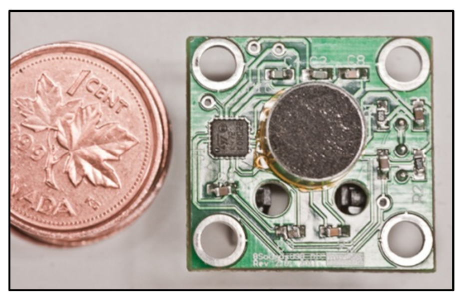

Figure 3: Microphone card

3) Analog Signal Conditioning and Integrity

Strategies to preserve the audio signal integrity are described this section.

The analog signal coming from the microphones is transmitted in differential mode to the codecs. Differential mode is preferred to single-ended signaling because of its better noise immunity and because it increases the dynamic range of the analog/digital converter of the codec. Since the acquisition card is designed to operate on a robot, many external devices can induce electromagnetic interference in the transmitted signal. Twisted pairs for signal transmission distribute the interference and a differential amplifier rejects the common mode noise, which minimizes the effect of overall electromagnetic interference.

Moreover, the amplitude of a differential signal is higher than the amplitude of a single-ended signal for the same reference voltage, which increases the dynamic range. According to the datasheet of the codec, the dynamic range goes from 102dB (single-ended signaling) to 105dB (differential signaling), which yields to a 41% increase.

The microphone preamplifiers are positioned as close as possible to the electret in order to reduce the 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 is 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 20dB at the lowest sampling frequency. The configuration of the anti-aliasing filters is the one suggested by the manufacturer.

4) Codecs

Cirrus Logic CS42448 codecs are used to convert the signals from the analog domain to the digital domain in both directions and the I2S (Inter IC Sound) serial protocol is used to exchange data with the XMOS processor. To meet the specifications, each codec uses four out of the six 24 bits converter input channels. The sampling frequency is configurable and 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 needs to operate with an analog reference voltage of 4.3V (as opposed to 5V for other codecs). It has a dynamic range of 105 dB in differential mode. Moreover, the codec is part of a reference design from XMOS Company. Codecs are preferred over ADC with eight channels because the card needs to have an audio output channel, which would require an additional DAC. However, the current design somewhat underuse the codecs, because each codec could support up to 6 analog inputs and 6 analog outputs. A future version of the sound card could then support up to 12 inputs and 12 outputs. We prioritized size over number of inputs / outputs or this iteration of the project.

5) PCB (Printed Circuit Board) Design

The PCB has four layers: a ground plane, a power plane and two planes for routing. A single ground plane is preferred over separated ground planes to avoid ground loops. The PCB is divided into two sections, one for the digital parts and another one for analog parts.

The power plane is divided in three sections, as shown in figure 3. The first section is used for the digital parts (in red), the second one is used for the analog parts (in orange) and the third one powers the XMOS cores (in blue).

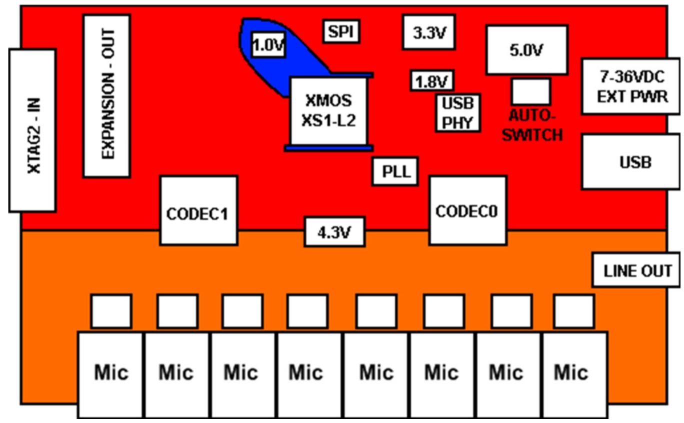

Figure 4: Power Supply Plane

The digital section of the power plane provides 3.3V to the digital parts of the two codecs, the USB PHY, the phase locked loop (PLL) and the XMOS digital I/O. This section covers approximately half the PCB.

The analog section of the power plane provides 4.3V to the analog parts of the codecs, the differential amplifiers, and the two audio outputs.

The XMOS section of the power plane provides 1V to the XMOS cores and is located directly under the integrated circuit of the XMOS processor for an easier access to inner pins.

This division improves the signal-to-noise ratio (SNR) since the high frequency noise generated by the digital parts does not interfere to the analog components.

A constant length between each differential pair is ensured to minimize the signal distortion. Decoupling capacitors are also used to ensure proper operation of the card.

1) Software Components

1) USB Audio Class 2.0

The prototype uses Audio Class 2.0 of the USB 2.0 standard and is already compatible with Linux (Ubuntu 10.10 and up) and Mac OS. It is not supported by Windows (Win7 SP1) but a driver for this operating system is provided by a third party. Hoping that Microsoft provides support soon due to the rising popularity of Audio Class 2.0.

The XMOS reference design provides 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, this implementation is used with the current design.

The generic driver available with Audio Class 2.0 is used. There is thus no need to design a custom driver. This option also reduces the amount of support to be provided after the project.

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 and thus do not run continuously with the system clock. This reduces power consumption since only active threads require power. An external clock is used to synchronize the codecs and XMOS cores for the data acquisition [12-13].

III. Results, validation and tests

The first test consists in measuring the voltage of each power line via test points.

The external power supply is then unplugged when the USB interface is connected. This ensures the power multiplexor works properly and that the device supports hot plugging.

Sound is then played in a loudspeaker and recorded with the audio card connected via USB to a PC. The recorded signals are then listened back.

All tests are passed, which confirms that the system meets the project requirements. The audio acquisition of the eight analog inputs and the analog stereo output are functional as well as the audio data transfer from the USB port. The system is functional on Linux, Mac and Windows.

A. Validation Method and Selected Tests

This section presents the key tests performed to characterize the USB synchronous multichannel audio acquisition system.

1) Electrical Consumption

The electrical consumption is verified for the external power supply and USB when the system is plugged with and without the eight microphones. The external power supply can range from 7 to 36 V. The consumption is thus 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 5W.

2) Latency between Acquisition Channels

The same signal is transmitted to the eight microphone inputs installed at the same distance from an omnidirectional sound source. Recording is performed on all channels with the Electroacoustic Toolbox [14] software. The integrated oscilloscope is used to confirm that the offset between the inputs is not higher than one acquisition cycle.

3) Effective Number of Bits (ENOB)

The ratio between the maximal noise in the input signals 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 desired effective number must be higher than the 12 bits operating threshold for ManyEars[15]. A fixed-frequency signal produced by a frequency generator is used as the analog input signal of the system. A slight increase of the signal amplitude is done. The smallest amplitude variation (voltage) that changes the digital signal by one bit is measured. The range of voltage (4.33V) is then divided by this voltage variation in order to obtain the effective number of bits (ENOB).

4) Mean Noise Floor

To quantify the mean noise floor of the system, an analog input channel of the card receives a signal from the generator and another input channel acts as the control signal. The difference between the two inputs is used to calculate the mean noise floor of the system.

B. Quantitative Results

The following table displays the results of these 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		

Figure 5: Results table

The card's dimensions are slightly larger than the ones requested in the initial specifications (i.e. 125mm × 74mm vs 100mm × 70mm). However, this small oversize has a negligible impact on the quality of the final product. We have chosen bigger and easy to use RJ-12 connectors that could be replaced by something smaller to gain PCB space easily.

On the other hand, the maximum power consumption of the sound card is less than the one in the original specifications (i.e. approximately 3.2W against 5W). Maximum power consumption occurs when using the maximal external voltage of 36V. If the external voltage is reduced to the minimal voltage of 7V, consumption drops to 2.5W. The variation of consumption is caused by the reduction of power loss in the DC-DC converter at the input of the power supply.

When the system is powered through USB, the maximum consumption is approximately 2.3W. This implies that the sound card can be powered without using USB in overcharge (a current of 500 mA or less is sufficient).

There is a maximum delay of 16µs between two acquisition channels. Thus, for a sampling frequency of 96kHz, the maximum delay between two channels is higher than one sample. However, for a target sampling frequency of 48kHz, the time delay is smaller than one sample. This test was done with a common signal for two distinct inputs on each physical codec. However, we could have pushed the test further by testing delay between each of the 8 inputs simultaneously.

The initial goal of 12 effective bits is met as the resolution goes up to 22 bits. The mean value of the noise floor obtained during the tests is similar to the one specified by the codec. This means that 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 eight analog input channels synchronously at the frequency of 48kHz. The system provides a resolution of 22 effective bits which exceeds easily the specification of 12 bits required by the ManyEars algorithm. The final sound card has a size slightly larger than the targeted size.

The system can be powered by an external or USB power supply. The maximal power consumption with USB is 2.3W, which meets the requirements.

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

B. *Future Development*

This project could lead to future developments with the XLINK ports [16]. These ports allow the addition of microcontrollers in a serial topology and thus can increase the number of parallel operations in the system. Moreover, the current threads do not use all the resources on the processor. These resources could be used to perform some computations for sound source localization directly 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