Chain architecture: An efficient hardware solution for a large microphone array system

Dmitry N. Zotkin, Ross Adelman, Adam E. O'Donovan and Ramani Duraiswami*

*Corresponding author's address: Institute for Advanced Computer Studies, University of Maryland College Park, A. V. Williams Building, College Park, Maryland 20740, ramani@umiacs.umd.edu

A typical microphone array system consists of a number of microphones connected to the digitization hardware and central processing unit in a parallel fashion. Such radial, hub-and-spoke architecture has multiple points of failure, suffers from electromagnetic interference, and does not scale well. In this paper, an alternative, chain-like architecture is described. In such setup, the microphones in a system are organized in a single chain. Each individual microphone board has an ADC chip and is connected to the previous and to the next microphones in the chain with short multi-wire cables carrying digital signals. A buffer board at the end of the chain converts the digital data stream into the industry-standard USB 2.0 format. In this way, the individual microphone board becomes the building block for quick and easy arbitrary-configuration microphone array assembly with minimal amount of wiring involved. A hardware implementation of the chain architecture was developed and is described. Accompanying drivers and software allow the user to perform on-the-fly data acquisition and processing in C and in MATLAB. As an example, a 64-microphone array was built, and several source localization and beamforming algorithms were implemented in MATLAB. Experimental results using the data gathered from the array are presented.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

At the current state of the art in technology, multimedia processing is starting to become ubiquitous, and smart multimodal data capture and processing systems are at the edge of becoming universally used by individual society members in the households. However, the human-computer interaction is still somewhat limited to modes that are natural for computers rather than for humans. While large advances have been made in video processing, the audio counterpart is comparatively underdeveloped. The main reason for that is that the untethered (distant) acquisition of high-quality audio signals (e.g., of a human speech) requires a microphone array—a number of spatially separated microphones whose signals are processed in a way as to enhance the desired and to suppress the undesired audio input.

Multichannel signal processing and microphone array research have very rich history [1]. A common task in setting up a microphone array is to physically and electronically connect all microphones to the digitization hardware and further to the processing unit. Typically, a separate amplifier is used for each microphone, and each amplifier’s output is fed to a separate channel of a analog-to-digital conversion (ADC) board. Such architecture is heavily parallel with one individual cable per microphone and all cables converging at the central hub. When the number of microphones becomes large, the amount of wiring involved makes the resulting system quite cumbersome.

In the current paper, an alternative architecture is described. The architecture is based on building a basic “microphone unit” by combining a microphone and an ADC chip on a small printed circuit board (PCB) and on connecting these microphone units sequentially into the chains of substantial length. The paper is organized as follows. In Section 2, requirements for microphones / microphone arrays and advantages and disadvantages of traditional and proposed architectures are discussed. In Section 3 and 4, the developed hardware solution and supporting software are described in details. Section 5 presents the experimental results obtained with a 64-microphone array assembled using the described technology. Finally, Section 6 concludes the paper.

MICROPHONES AND MICROPHONE ARRAYS

In this section, generic requirements for individual microphones and microphone arrays are discussed, and a brief overview of the state of the art technology is provided.

General Requirements

Microphones: Individual microphones come in a variety of shapes and sizes. By far, the most common type currently in use is electret. A higher signal quality is associated with condenser microphone; however, these require phantom power for operation. Dynamic microphones operate using a principle reverse to that of the loudspeaker and tend to have relatively narrow operational bandwidth. Ribbon microphones operate similarly to the dynamic ones but respond to the pressure gradient (as opposed to the pressure itself). A relatively new development is a MEMS microphone, where the mechanical membrane is carved out directly from the silicone substrate; these are extremely small but about 10-15 dB noisier than conventional electret ones. There is also a host of other, more exotic microphone varieties. Almost all microphones require a pre-amp to be used with them (often built-in in the microphone cartridge). For digital processing of the recorded signals, ADC hardware is necessary. General characteristics of the audio processing chain (microphone, pre-amp, and ADC) are sensitivity, frequency response, noise floor, signal-to-noise ratio, sampling frequency, and sampling bit depth.

Microphone Arrays: In the most general form, a microphone array is a collection of microphones located at known, spatially distinct locations. Differences in the acoustic signals recorded allow one to infer the spatial structure of the acoustic field and to obtain related information such as sound source(s) position(s) [2]. Conversely, if the field structure is known, one can apply spatial filtering so as to amplify or suppress certain parts of the audio scene [3]. Various array configurations are possible, including e.g. linear, planar, and spherical arrays. Each of these has certain advantages and disadvantages and is suitable for specific types of applications; for example, the spherical array has fully symmetrical coverage of the
three-dimensional space surrounding the array, which can be used to provide co-registered multimodal (video and audio) images [4]. There are additional requirements for the microphone array system, such as the need for the “common clock” to synchronize data capture and the calibration required to have identical magnitude/phase response across all microphones (or to compensate for inter-microphone differences). Also, for arrays larger than a few microphones, engineering issues such as power consumption, heat dissipation, physical array size, cabling, electromagnetic interference (EMI), and space requirements often pose additional challenges.

**Hub-and-Spoke Architecture**

Traditionally, a microphone array is built in a parallelized fashion. Each microphone has a (possibly built-in) pre-amplifier, and the audio signal travels over the microphone cable to another amplifier (possibly combined with signal conditioner) and then to the multichannel ADC board typically installed in a desktop computer. The array built in this fashion has a number of weak points:

- Involvement of bulky hardware and lack of portability;
- Excessively large amount of cabling required;
- The need for sturdy mounting hardware and acoustic interference from it;
- Presence of multiple points of failures at cables’ connectors;
- EMI susceptibility of analog signals in transit; and
- Non-simultaneous sampling due to sequential operation of the ADC board.

An example of a relatively large array built by the authors of the current paper is a 128-element array used for reciprocal HRTF measurement at the University of Maryland [5]. The array was set up using four 32-channel NI PCI-6071E data acquisition cards, four 32-channel custom amplifier boxes, and 128 Knowles FG-3629 microphones. Another example of a large array is a 512-element Huge Microphone Array (HMA) project [6] at Brown University. The latter was built using specialized DSP hardware and avoids some of the aforementioned problems; however, the equipment, cabling, and space requirements obviously make it non-portable.

**Chain Architecture**

As an alternative architecture, a chain design is considered in this paper. In the chain architecture, individual microphone units feature analog-to-digital converter located immediately next to the microphone to reduce EMI, and the circuitry is designed for connecting these units in a serial fashion so that each unit sends the ADC conversion results to the output data connector and then relays whatever data is presented at the input data connector to the output. The unit located at the far end of chain has its input data connector grounded and its output data connector connected to the input data connector of the next unit. The units are interconnected in a similar fashion through the whole chain, and the first unit (the one at the near end of the chain) relays the data from the whole chain to the data consumer.

The chain architecture avoids all the above-mentioned problems associated with hub-and-spoke architecture. In particular, a bulky set of long cables (one per each microphone) is replaced with short links connecting individual microphone units together. However, a single cable or board failure in chain architecture would obviously render the rest of the chain disconnected, which is a substantial weakness. To minimize a possibility of such event, one can encase a microphone array (if permitted by application) into a physical “black box” so that all inter-unit cables are securely mounted inside and interfacing with the array is done via a single cable.

---

1A provisional patent application on the described technique was filed with U. S. Patent and Trademark Office (No. 61/545,180).
FIGURE 1: A microphone unit designed by VisiSonics Corporation (front and reverse PCB views). The microphone is attached at the top. An 18-bin ribbon flat cable connector and a micro-BNC connector are seen at the bottom of the PCB. Two potentiometers are seen on reverse view for gain calibration and ADC reference voltage fine tuning. The board width is 25 mm.

HARDWARE DESCRIPTION

In order to advance the state of the art in the microphone array research, a chain-architecture hardware solution was developed, designed, and implemented. A sample microphone unit is shown in Figure 1. In this section, details of implementation are described.

The microphone used in the design is Panasonic WM-61A (electret type). The microphone amplifier is built on an LTC 6912 chip, which is a low-noise AC amplifier with the gain and bandwidth programmable using 3-wire SPI interface. The single-ended output signal is converted to fully differential using THS 4521 chip with unity gain and is fed to the AD 7767 24-bit ADC operating in daisy-chained mode, which is the key of the described architecture.

There are two separate clocks provided to AD 7767. The first clock (MCLK) serves as a master clock for performing the analog-to-digital conversion. The sampling frequency is equal to the MCLK frequency divided by 8. The second clock (SCLK) controls the output data transfer; at each SCLK pulse, the next bit of data is output on SDO pin. There is also an SDI pin, which is used for daisy-chaining and should be connected to the SDO pin on the next board in chain. Input on SDI is shifted into an internal register on each SCLK pulse and then shifted out onto SDO after the ADC has finished outputting its own conversion result. In this way, the data words propagate through the whole chain driven by SCLK. The number of boards in the chain is limited by how fast the read operation can proceed so that all data is consumed before the next conversion starts. In the current setup, the MCLK is 352.8 kHz and the SCLK is 22.5 MHz so that approximately 20 boards can be chained. The maximum SCLK frequency as per ADC 7767 spec sheet is about 42 MHz, allowing one to accommodate more than 32 boards on one chain.

Most of the signals traveling between microphone units are of relatively low frequency, except the data bus (SCLK and SDO/SDI). The inter-unit link is provided by two cables – an 18-pin flat ribbon and a micro-BNC coaxial cable. SCLK is produced by the interface board and is connected in parallel to all microphone units; therefore, it is routed on the coax to minimize distortion and interference. On the other
hand, SDO/SDI signal is re-generated at every board and is therefore placed on the flex-ribbon cable. Also the following signals are present on the ribbon cable: SPI CLK, SPI DATA, SPI CS (for LTC 6912 programming); MCLK; ADC CS; ADC RESET; and ADC DRDY. The DRDY line stands for “data ready” and is set by ADC to indicate the end of the conversion. The power is also supplied via the ribbon cable, and each microphone unit has several high-precision voltage regulators for main power, ADC voltage reference, and microphone power.

The chain of microphones is connected via the same dual-cable link to the buffer board containing drivers for high-load signals and further to the interface board. The buffer board is designed for connecting up to four chains at the same time to the same interface board. The interface board is an Opal Kelly XEM3010-1500P product based on Xilinx XC5S1500-4FG320 Spartan-3 FPGA featuring USB 2.0 interface. A firmware written in Verilog handles the interfacing details such as MCLK/SCLK production; synchronous ADC reset; gain/bandwidth settings transmission over SPI bus; and data acquisition, buffering, and USB transmission when triggered by DRDY. Seven gain settings are possible. The ADC output saturates at 94 dB SPL and at 128 dB SPL for the highest and the lowest possible gain settings respectively.

**Software Description**

The Opal Kelly interface board used in the project is bundled with a software package called FrontPanel. The software provides a convenient API for interfacing between C / C++ / MATLAB / Java / Python code running on the host PC and the FPGA firmware. From the software engineer point of view, the interface is defined via communication endpoints (pipes). An endpoint is established in firmware, an identifier is assigned to it, and data is streamed into the endpoint controlled by a user-defined clock. On the host computer, the endpoint is then opened in a way similar to opening a file or socket and a read operation is issued to obtain the data submitted to the endpoint by firmware. Data buffering and USB transfer negotiations are done automatically and seamlessly by Opal Kelly provided drivers operating on the host PC and by a firmware module that must be instantiated in user's FPGA design.

A simple software development kit was developed; it is implemented as a dynamic link library compiled from ANSI-compliant C source code. It is intended for use with arbitrary C / C++ applications that have a need to consume the audio stream for online data processing. It has been used to perform source localization
and beamforming, to implement a remote audio telepresence application, and to visualize the spatial
distribution of the acoustic energy, all in real time. The SDK has the ability to change the microphone gain
setting and the acquisition precision (number of bits per sample) dynamically. The SDK can also be used
directly from MATLAB via “loadlibrary” and “calllib” functions; simple real-time “oscilloscope” / “auditory
spectrogram” MATLAB software module was implemented as a proof of concept.

A typical modern off-the-shelf office PC is easily able to handle computational and data transfer loads
involved. For reference, the USB bandwidth consumed by a 64-microphone array operating at 44.1 kHz at 24
bits per sample is about 11.2 megabytes per second.

**Experimental Results**

In this section, selected performance results obtained with the chain-architecture microphone array are
presented. Figure 2 shows the spectrum of the sound recorded in silence at the lowest gain setting. A
microphone array was placed in a large foam-insulated case, which was shut tight around chain interface
cables. Air conditioning and lighting in the room was turned off, and the recording computer was located in
an adjacent room. As seen in the Figure 2, the noise floor of the array is about 20 dB SPL. A formal
computation of the A-weighted equivalent background noise in accordance with IEC 60268-1 (section 6.2.1)
gives the value of 23.2 dB SPL. The output saturation point at the same gain is about 128 dB SPL; therefore,
the useful dynamic range of the array microphone is about 105 dB.

A signal-to-noise ratio was measured using a PCB Piezotronics CAL 200 pistonphone producing a 94 dB
SPL 1000 Hz acoustic signal output. Spectrum of the recorded signal is shown in Figure 3. The SNR
computed in accordance with [7] is 61.4 dB, which agrees with the SNR specified by the microphone
manufacturer. A measurement of the response flatness was also performed per IEC 61094-5; it was found
that the free-field response deviates from that of the reference microphone used (Earthworks M23
measurement microphone on M-Audio Fast Track Pro USB interface) by no more than 2.1 dB over the entire
frequency range (0 through 20 kHz).

A data acquisition experiment was also undertaken. The 64-microphone array was placed in the room in
the vertical plane with two (horizontal and vertical) linear 32-microphone chains (each consisting of

**Figure 3:** Spectrogram of the 94 dB SPL 1000 Hz calibration signal recorded by a microphone unit.

![Spectrogram](image)
FIGURE 4: A RealSpace Audio Camera – a commercial product developed by VisiSonics Corporation that utilizes the described microphone array technology.

equispaced microphones and spanning approximately 1400 mm) intersecting in the middle. Two persons were speaking at the same time at fixed known locations. A simple delay-and-sum beamformer [8] was implemented in MATLAB for data processing. The expected beamforming gain was 18 dB (each doubling of the number of microphones increases the gain by 3 dB). The spatial aliasing limit of the array is approximately 4 kHz. In the useful frequency range, the beamforming gains obtained when steering to the first and to the second speaker were 15.6 and 14.3 dB, respectively. The discrepancy with the theoretical predictions is likely due to the room reverberation and to lesser extent to the inaccuracies in microphone and source position measurements.

Additional data was collected with a third source (a portable music player, also in a fixed known location) to test the interference suppression (null steering) ability. An MVDR beamformer [8] was implemented, and data processing was done assuming known interference signal direction of arrival. The unwanted source suppression level achieved was 12.89 dB while maintaining distortionless response in the main look direction; the interference signal is leaking into the output for the same reasons as in the first experiment above.

CONCLUSIONS

A portable, low-power, robust microphone array system was designed. The microphones in the array are digitally connected in a chain-like fashion to dramatically reduce the amount of wiring required and to eliminate electromagnetic interference possibilities. An interface board was also developed streaming the audio data over an industry-standard USB 2.0 interface. As such, the array is hot-pluggable into any common desktop / laptop computer with no additional hardware necessary. An accompanying SDK is available for data capture and live data streaming. The audio characteristics of the array microphones are on par with the microphones sold commercially as calibration or reference microphones. The developed hardware can be used to quickly assemble large arbitrary-shaped microphone array and comprises a flexible tool for research and industrial applications; for example, the VisiSonics' Audio Camera product (Figure 4) employs the described circuitry for audio data acquisition. Individual microphone units and interface boards are also commercially available at VisiSonics for end-users wishing to built their own array configuration.
Future research work includes increasing the number of microphones, migrating to faster host interfaces (PCI Express or USB 3.0), and applying the array in HRTF measurement studies.

REFERENCES


