Towards an FPGA implementation of IOT-based multi-modal Hearing AID System

Enemali, Godwin; Bishnu, Abhijeet; Ratnarajah, Tharmalingam; Arslan, Tughrul

Published in:
2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW): AMHAT 2023: Advances in Multimodal Hearing Assistive Technologies

DOI:
10.1109/ICASSPW59220.2023.10192936

Publication date:
2023

Document Version
Author accepted manuscript

Citation for published version (Harvard):
Enemali, G, Bishnu, A, Ratnarajah, T & Arslan, T 2023, Towards an FPGA implementation of IOT-based multi-modal Hearing AID System. in *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW): AMHAT 2023: Advances in Multimodal Hearing Assistive Technologies.* IEEE. https://doi.org/10.1109/ICASSPW59220.2023.10192936

TOWARDS AN FPGA IMPLEMENTATION OF AN IOT-BASED MULTIMODAL HEARING AID PROTOTYPE

Godwin Enemali, Abhijeet Bishnu† Tharmalingam Ratnarajah† and Tughrul Arslan†*

†School of Engineering, University of Edinburgh, Edinburgh, United Kingdom
Email: abishnu@ed.ac.uk,t.ratnarajah@ed.ac.uk, tughrul.arslan@ed.ac.uk;
*Department of Electrical and Electronic Engineering, Glasgow Caledonian University, Glasgow, United Kingdom; Email: godwin.enemali@gcu.ac.uk

ABSTRACT

This paper presents work in progress with aims to develop a first of its kind IoT Cloud-based multimodal hearing aid (HA) prototype. Specifically, it identifies the major processing blocks for implementation on an embedded Field Programmable Gate Array (FPGA) and focuses on aspects of a custom block that deals with the integration of multiple input sources of the HA. In particular, the integration block deals with the combination of audio and video (AV) data at different sampling rates, with each having a different number of bits to encode their information. This realises a seamless combination of AV data by assigning a scaled number of bits to each component in the combined signal to achieve real-time data transmission of all data sources. Finally, we outline ongoing work aiming to address a range of key challenges including required optimisation of tradeoffs between HA latency, streaming bandwidth, security, privacy, power consumption and performance.

***Index Terms*—** Audio-visual speech enhancement, embedded platform, multimodal hearing aid

1. INTRODUCTION

The capacity to hear is a major aspect of our social and cultural life as humans. Consequently, loss of hearing can severely affect, not only our ability to function effectively, but also lead to social alienation. Loss of hearing affects about a fifth of the human population [1, 2, 3], and can be attributed to to a plethora of reasons including congenital factors, disease, accidents or simply aging [2]. Hearing aids (HAs) are the most widely used devices to improve sound perception in those with hearing difficulty. They are designed to discriminate the target sound from background noise. However, traditional hearing aids are often based on achieving this task primarily by sound amplification alone [4]. Researchers are now exploring the use of additional information, inspired by the way humans perceive speech, to improve the performance of future hearing aids. Examples of these additional information include visual information (e.g. lip reading), facial expression, etc. The COG-MHEAR project [5] is a cutting edge multi-million-pound research project that aims not only to address foundational research issues in the development of multimodal HAs, but also address practical issues in the implementation of multimodal HAs.

In addition to the requirement of high sound quality outputs, HAs are embedded systems that need to satisfy several other requirements such as user convenience, cost, size and low power overhead. To address some of these requirements, several machine learning (ML) algorithms have been developed that offer great functionality in modern multimodal HAs [6, 7, 8, 9]. However, the implementation of these proposed algorithms in way that meets the constraints of efficient multimodal HA remains an open challenge.

IoT cloud-based HA architectures have also been proposed [1]. In this architecture, a transceiver IoT device (HA device) samples raw source information (e.g. audio and video data), pre-processes the information, then transmit data in real-time to a cloud for further processing. The clean audio information is then sent back to the HA device. An overview of this architecture is illustrated in Fig. 1. As shown in the figure, multimodal information are acquired, digitized and processed on an embedded device before being sent to a cloud for further processing using 5G networks. ML algorithms are implemented in the cloud, where the constraints of power and cost are not as stringent as on the embedded device. The ML algorithms produce enhanced audio output. Enhanced audio information are sent back in real-time to the HA device for final processing and outputting to a speaker. There are several advantages for this architecture. For example, the edge device could be optimized for size, power and cost while offloading significant aspects of the processing to the cloud.

For effective functionality, and to guarantee secure, real-time operation of multimodal HAs in a cloud scenario, several

This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) programme grant: COGMHEAR (Grant reference EP/T021063/1).

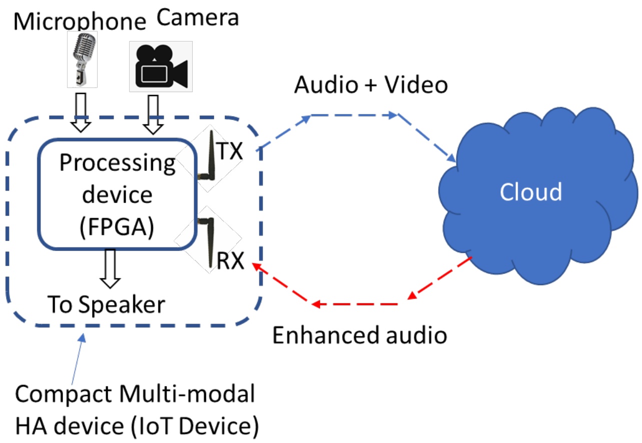

Fig. 1. Overview of a typical IoT-based Multimodal HA system.

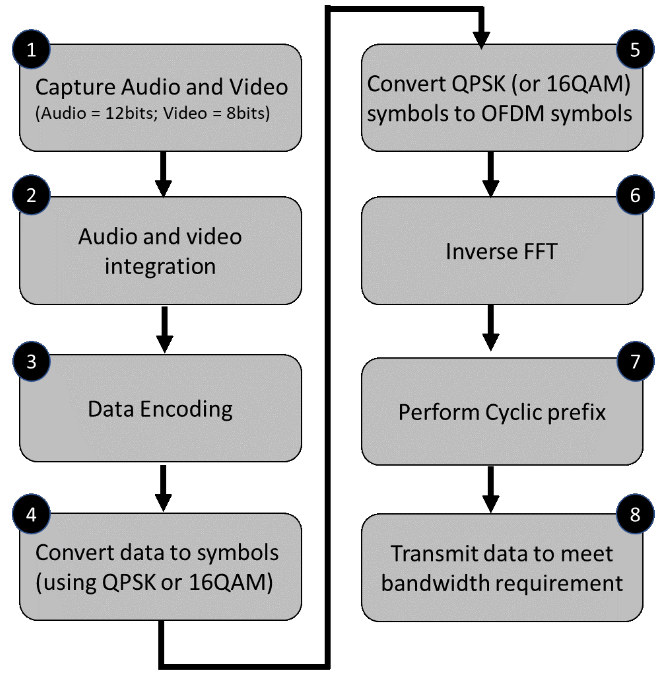

Fig. 2. Major processing blocks in the uplink section of a typical IoT-based HA system.

issues need to be addressed. In our previous work [1], a novel frame architecture that supports the stringent requirement of cloud-based multimodal HAs was proposed. The proposed frame structure addresses the differences in the data rate between the uplink (i.e. from the HA device to the cloud) and the downlink (i.e. cloud to HA device), with a focus on the physical layer (PHY layer).

In particular, the proposed frame architecture addresses the unique requirement of data transmission in HAs. Unlike in conventional cloud-based scenarios where the downlink section has a higher data-rate than the uplink section, in HAs the uplink (carrying multimodal information) has a higher data-rate than the downlink (carrying only clean audio information). In addition, the paper also provides details of the source encoding, data transmission as well as the ML algorithms for cloud-based multimodal HA architectures. However, that work utilized personal computers and large processing equipment for validation of the proposed cloud-based architecture and the developed communication frame structure. In this work we report our initial efforts towards customization of the algorithms developed in [1] for embedded reconfigurable platforms to address the challenges of real-time, device size, low cost and power overhead.

Although cloud-based architectures have many promising outcomes and several use cases, there are several very important practical implementation challenges that need to be addressed to produce a prototype. As a starting point for our implementation, we adopt the framework in [1]. Fig. 2 shows major pre-processing blocks in the uplink section of a typical cloud-based HA system. These blocks work together to upload audio and video data to the cloud. For HA system, the process in Fig. 2 must be accomplished in real-time. As shown in the figure, in addition to the core audio and video data pre-processed on the chip, a number of communication protocol processing steps are carried out (blocks 3 – 7) before data is transmitted to the cloud. The raw audio-visual data is encapsulated according to these communication protocols to produce transmission-ready frames. The details of the frame structures for both uplink and downlink are described in [1]. As described, the uplink frame structure (UFS) comprises of a synchronization signal, USS, a physical uplink control channel, PUCCH, and a physical uplink shared channel, PUSCH. It also contains two reference signals: an uplink reference signal, URS, and a physical uplink control channel reference signal, PUCCH-RS. Fig. 3 summarizes the PUCCH symbol generation process, where UCI is the uplink control information and the CRC is the cyclic redundancy check on the UCI bits.

To enable real-time operation and to meet some of the challenges of low power, size and adaptability, the processing blocks in Fig. 2 are being implemented on a 5G enabled, commercial-off-the-self, FPGA-platform, namely the Xilinx's Zynq UltraScale+ RFSoC. FPGA platforms are popular for the implementation of several components of hearing aids and cochlear sub-systems [10, 11, 12, 13, 14]. In addition to providing a high-performance platform for accelerating custom HA algorithms to meet strict timing requirements, they also provide the flexibility to be easily re-configured even while in operation. In addition, The Xilinx's RFSoC being used in this work provides multiple RF channels for expansion and adaptability. Some of the blocks in Fig. 2 are realized by adapting existing Xilinx IPs, while others require custom Register Transfer Level (RTL) implementation. One of the blocks in the flow that require custom implementation

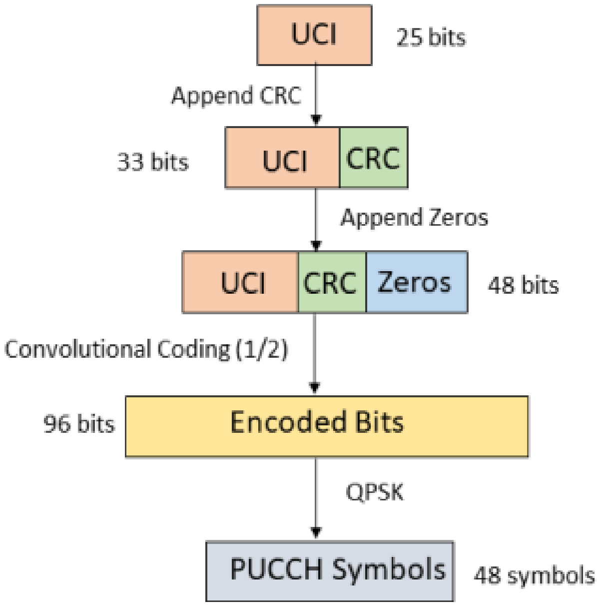

Fig. 3. PUCCH symbol generation [1].

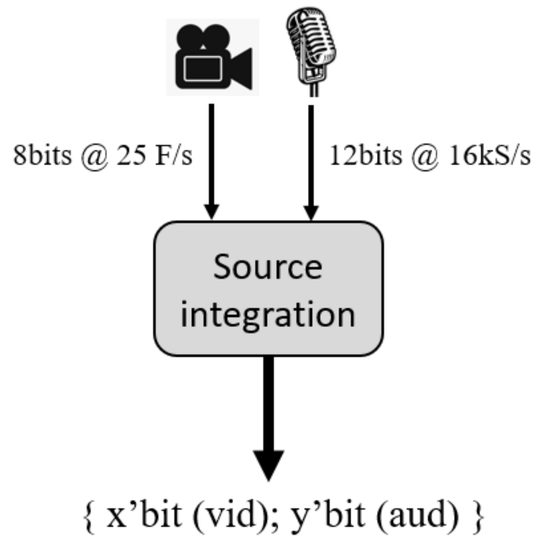

Fig. 4. Integration of multiple sources in Multimodal HA system.

is the integration of the audio and video data (i.e. the second block in Fig. 2). Its aim is to seamlessly combine the audio and video data, dealing with their different source properties such as sampling rates and number of bits to encode source information.

2. AUDIO AND VIDEO DATA INTEGRATION

To illustrate the operation of the proposed source integration of the audio and video data, consider a case where audio is sampled at 16 kSPs and video is sampled at 25 frames per second (FPS). As an example, a one second duration of audio data sampled at 16 kSPs with 12-bits for each sample will produce 196 608 bits that needs to be processed, while a 1 second video frame with 60-by-80 pixels per frame, sampled at 25 FPS and 8-bits per pixel will produce nearly 5-times as much data as in the audio case.

The block handles the combination of these data into a single signal for onward processing. To deal with the different data rates, data from the multiple sources are combined by slicing bits from 'samples' from each source and concatenating them to form new composite samples. Fig. 4 gives a simple high-level illustration of the concept. In essence, different number of bits are assigned to each source in the combined signal as shown in Fig. 4. The number of bits assigned to each source is proportional to its source load. Using this strategy, the data rate for each source is scaled to achieve real-time data transmission of all sources.

However, in the context of hearing-aids, a continuous real-time integration of AV data is required. Therefore, the proposed firmware integration of AV data is parallelised and pipelined to deal with streaming data rate. In particular, each data processing stage in the data integration pipeline handles a fixed number of bits and completes its task in a fixed number of clock cycles, immediately after which a new set of bits are accepted for processing. To increase the bandwidth and meet timing requirements (e.g. processing every frame within 10ms), the level of parallelism is increased. For example, the bandwidth of the process is doubled by duplicating the number of the AV source-integration blocks. However, this will increase the chip-area and power overhead. This performance trade-off is being characterised to ensure that both timing and power requirements are met.

3. SECURITY AND PRIVACY CHALLENGES

One of the challenges of the proposed cloud-based architecture is to assure the security of sensitive user data [15, 16]. Wireless signal transmission could be vulnerable to attacks that may not only put user information at risk, but could also be harnessed to drain battery power of device, confuse users, etc. In our proposed system, we are minimizing these risks by implementing privacy-preserving mechanisms on-chip by evaluating a number of state-of-the-art techniques [16]. However, it is important to carefully balance their effectiveness against other specific HA requirements such as strict latency, power and performance metrics including user preferences.

4. CONCLUSION AND FUTURE PLANS

In this paper, we discussed initial steps towards an FPGA implementation of an IoT-based multimodal hearing aid system prototype. We identified key processing blocks being implemented on a commercial off-the-self FPGA board. In ongoing work, we are addressing key implementation challenges including optimising a range of trade-offs between streaming bandwidth, latency, privacy, security, device size, power consumption, cost and performance. We aim to complete the implementation of all processing blocks and produce a demo in collaboration with our industrial partners.

5. REFERENCES

[1] A. Bishnu, A. Gupta, M. Gogate, K. Dashtipour, A. Adeel, A. Hussain, M. Sellathurai, and T. Ratnarajah, “A novel frame structure for cloud-based audio-visual speech enhancement in multimodal hearing-aids,” in *2022 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*. IEEE, 2022, pp. 75–80.

[2] G. Stevens, S. Flaxman, E. Brunskill, M. Mascarenhas, C. D. Mathers, and M. Finucane, “Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries,” *The European Journal of Public Health*, vol. 23, no. 1, pp. 146–152, 2013.

[3] A. Shrivastav, M. Kolte, S. Bhandari, and V. Harpale, “An efficient reconfigurable frm filterbank for digital hearing aid,” in *2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA*. IEEE, 2022, pp. 1–4.

[4] N. A Lesica, “Why do hearing aids fail to restore normal auditory perception?,” *Trends in neurosciences*, vol. 41, no. 4, pp. 174–185, 2018.

[5] “Cog-mhear,” https://cogmhear.org/, [Accessed]: 2023-03-30.

[6] A. Adeel, M. Gogate, A. Hussain, and W. M. Whitmer, “Lip-reading driven deep learning approach for speech enhancement,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 3, pp. 481–490, 2019.

[7] T. Zhang, F. Mustiere, and C. Micheyl, “Intelligent hearing aids: The next revolution,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 72–76.

[8] H. Schröter, T. Rosenkranz, A. Escalante-B, and A. Maier, “Low latency speech enhancement for hearing aids using deep filtering,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2716–2728, 2022.

[9] H. Salehi, D. Suelzle, P. Folkeard, and V. Parsa, “Learning-based reference-free speech quality measures for hearing aid applications,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 12, pp. 2277–2288, 2018.

[10] V. Hanson and K. Odame, “Real-time embedded implementation of the binary mask algorithm for hearing prosthetics,” *IEEE transactions on biomedical circuits and systems*, vol. 8, no. 4, pp. 465–473, 2013.

[11] L. Bendaouia, H. Salhi, Si M. Karabernou, L. Kessal, and F. Ykhlef, “Fpga-implementation of a bio-inspired medical hearing aid based dwt-ola,” in *2014 International Conference on Audio, Language and Image Processing*. IEEE, 2014, pp. 806–811.

[12] G. Enemali, S. Wang, and A. Hamilton, “An fpga-based adaptive tuning control for a cochlea-mimicking filter channel,” in *2017 15th IEEE International New Circuits and Systems Conference (NEWCAS)*. IEEE, 2017, pp. 189–192.

[13] K. Samtani, J. Thomas, G. A. Varma, and SP Deepu, “Fpga implementation of adaptive beamforming in hearing aids,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 2239–2242.

[14] A. Mishra and A. E. Hubbard, “A cochlear filter implemented with a field-programmable gate array,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 49, no. 1, pp. 54–60, 2002.

[15] R. L. Brennan, S. Steffler, J. Dods, and J. He, “Hearing aid and extreme edge iot acceleration,” in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2021, pp. 684–687.

[16] P. Katrakazas and D. Koutsouris, “A (lack of) review on cyber-security and privacy concerns in hearing aids,” in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2018, pp. 223–226.