Library Progress International *Print version ISSN 0970 1052*
Vol.44 No. 3, Jul-Dec 2024: P. 19242-19255. *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Efficient Audio Feature Extraction For Iot Devices Using Low-Power Vlsi Architecture

Sathisha S B[1], Venkatesh A[2], Naveen Kumar V[3]

[1] Lecturer, Sound Recording & Engineering, Government Film & Television Institute, Hesaraghatta, Bengaluru, Karnataka, India
[2] Senior Scale Lecturer, Electronics & Communication Engineering, Government Polytechnic Nagamangala, Mandya, Karnataka, India
[3]Senior Scale Lecturer, Electronics & Communication Engineering, Government Polytechnic Arakere, Mandya, Karnataka, India

How to cite this article: Sathisha S B, Venkatesh A, Naveen Kumar V (2024) Efficient Audio Feature Extraction For Iot Devices Using Low-Power Vlsi Architecture. *Library Progress International,* 44(3), 19242-19255.

ABSTRACT

As the use of IoT technology has rapidly advanced, proper algorithms for audio feature extraction that IoT devices can use to process audio data have become critical for the technology's continued progress. This paper reviews the existing developments in IoT audio feature extraction methods with a view of integrating low-power VLSI systems to boost the performance of the feature extraction methods. Since most IoT devices depend on voice recognition and environmental monitoring, there has been a constant search for high-quality real-time audio processing.

Many trends have been outlined in the course of the study which may define the future development of the audio feature extraction process and one of the most important trends is the use of modern machine learning methods, including deep learning and neural networks. Such approaches also enable devices to learn information features from undifferentiated data and enhance probability measures in recognizing intricate sound patterns. Finally, the paper focuses on shouldering around audio processing in edge computing since it can relieve the strain of latency and bandwidth usage by processing audio data on the edge rather than at the central hub. Housing data processing closer to the source it is being collected from is particularly suitable for requests requiring immediate response times such as voice-activated systems as well as real-time watching over huge streams of data.

Another area of interest in this study is the application of context-aware feature extraction techniques where the amount of noise selected is proportional to the environment's context. This capability extends user experience and is effective in ensuring that the target device delivers an appropriate response to intended auditory messages. Furthermore, the integration of multimodal data sets where input from the microphone is merged with visual or motion sensor data is proposed as a way of getting a better understanding of complex surroundings.

The work also contextualizes low-power VLSI design in enhancing feature extraction from audio signals for IoT devices with constrained energy sources. The nature of dedicated architectures that are installed solely with audio processing really plays an important role in ensuring kept-up performances yet energy is considered. Last but not least, the paper underlines the fact that most audio processing techniques should incorporate standards and interfaces to enable the transportation of audio data across different IoT devices.

In conclusion, the findings of this study establish audio feature extraction as a path-breaking technology in IoT and identify machine learning, edge computing, context awareness, and low-power design as core vectors for future progress in the field.

Keywords: Internet of Things, audio feature extraction, low-power VLSI architecture, machine learning, edge computing, context-aware processing, multimodal data fusion, standardization, energy efficiency

I. Introduction to Audio Feature Extraction

In the field of audio signal processing, especially for applications of ML and AI, audio feature extraction is an important step. This is done through with the conversion of raw signal inputs obtained from an audio medium into a set of meaningful descriptors or features that can be used in different applications like speech-to-text conversion, music categorization, and ambient sound identification. Audio feature extraction makes it easy to process and classify complex audio signals to adapt to different applications since they do not require many system resources as is the case with IoT devices.

Acoustic signals are already multivariate by nature, including not only the base pitch of sounds and phenomena but also numerous harmonics, amplitudes, and temporal characteristics of the signals. Nevertheless, raw audio data is frequently unorganized and massive which complicates the pattern of hard-coded systems to understand. Thus, the initial stage in most audio processing systems is the extraction of the feature vector which contains the necessary recognition of the audio signal's attributes while minimizing the dimensionality and the computational complexity.

Spectral features Mel frequency cepstral coefficients and time domain features are common audio features. MFCCs are used routinely in the field of speech processing due mainly to the fact that they can mimic the perceptual property of sound. It is involved in the representation of the short-term power spectrum and is used for the identification of phonetic sounds in the speech systems. Spectral characteristics for instance spectral centrel and spectral flax are features that give detail of the frequency distribution and fluctuations within an audio signal. The set of features based on the time changes in the signal, which may include zero-crossing rate and energy, is valuable for such jobs as music style recognition or speaker recognition.

The essence of audio feature extraction goes beyond what is typically considered relevant; it has the additionally crucial contribution of influencing IoT performance. Audio sensing seems to be a popular feature in IoT systems, including voice-controlled assistants, or environment monitoring; as such, optimization of the audio processing has become necessary. These devices bear limited power capacity and computing power therefore calling for low-power and high-accuracy feature extraction approaches.

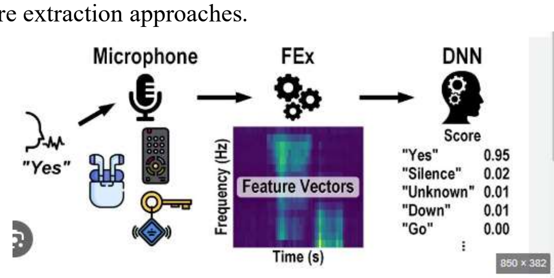

Figure 1: Processing Stages of KWS

In other words, audio feature extraction stands as a rudimentary step of conditioning raw sound signals into valuable representations. In emphasizing some specific characteristics of the audio data, it helps for effective analysis from both single data points and decision-making processes in multidimensional applications; especially in the context of the escalating trends of IoT devices. It can therefore be anticipated that as technology increases, the invention of new approaches to feature extraction will cause enhancements to the audio processing, resulting in intelligent systems.

1. II. Challenges in Audio Processing for IoT Devices

Audio processing in Internet of Things (IoT) devices has two major challenges mainly based on its operating environment. Given that IoT devices may soon be everywhere—voice-controlled speakers such as Amazon's Echo or Alphabet's Nest; health monitors attached to smart clothing and accessories—they require efficient audio processing. Nevertheless, several obstacles need to be overcome to guarantee proper functioning.

1. Finite Computational Power

The IoT devices have many limitations in terms of computational power and the memory and cache that is available for use. In contrast to many conventional computing platforms capable of processing algorithms and big data, IoT devices are required to process a variety of sound analysis tasks reliably within limited resources [1]. This is more so because the extraction of superior audio features may entail certain other prosecution costs that may not be easily justified due to high computational demand.

2. Energy Consumption

I noted that while IoT devices comprise small connected devices, energy consumption is a major concern due to batteries, or energy harvesting mechanisms used to power the devices. Some of the audio processing tasks are computationally and energy demanding; this is especially when in real-time processing or when processing algorithms such as deep learning feature extraction [2]. The requirement to perform accurate analysis of audio information and at the same time consume as little power as possible creates a big problem for developers willing to make the device last as long as possible.

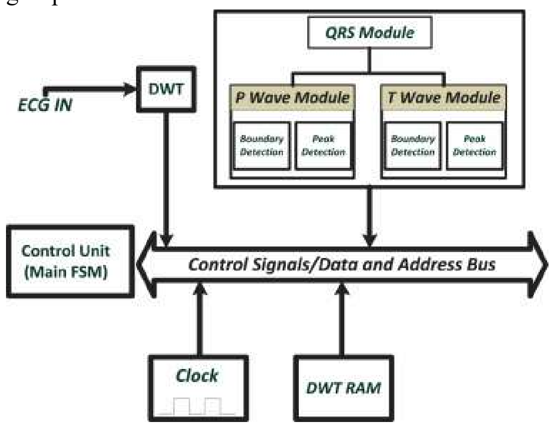

Figure 2: Power Efficient VLSI Structure

3. Environmental Noise and Variability

IoT devices are connected in various and often uncontrollable settings, and their acoustic environment can greatly affect speech quality. Echoes or other interfering sounds included in the captured signal also contribute to temporal variability that hampers direct feature extraction as well as subsequent analysis [3]. The identification of target sounds of interest from the complex acoustic background is a crucial but not trivial task for which sound-matching algorithms are required.

4. Latency and Real-time Processing

Most IoT apps use real-time audio feedback or interaction as many apps utilize voice recognition in smart devices. It is very important but challenging to obtain a low-latency response while keeping high accuracy, primarily because of the restrictions on the computational capabilities and owing to the sophisticated nature of the algorithms used. Longer response times get in the way of usability and the whole purpose of owning and operating the device itself.

5. Data Privacy and Security

With more and more IoT devices working with and analyzing private voice inputs including discussions or individual health data, privacy and security issues are under focus. The protection of the privacy of the users of the audio processing system also brings difficulties in the processing and transmission of the audio data [4].

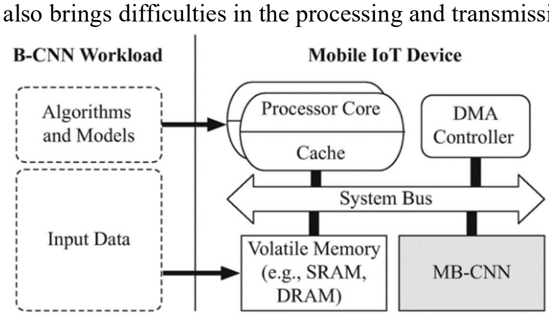

Figure 3: IoT Data Processing for Emerging Technologies

6. Scalability

As the number of IoT devices increases finding ways for designing audio processing solutions that can be scaled is crucial. Systems should be capable of managing their loads, because the number of connected devices and their activity may differ significantly.

Lastly, and importantly, issues concerning audio processing for IoT devices include hardware and compute limitations, energy efficiency and real-time requirements, environmental factors, latency sensitivity, data privacy, and scalability [5]. Meeting these challenges is imperative for designing strategies in audio processing that will improve the functionality and usability of IoT devices.

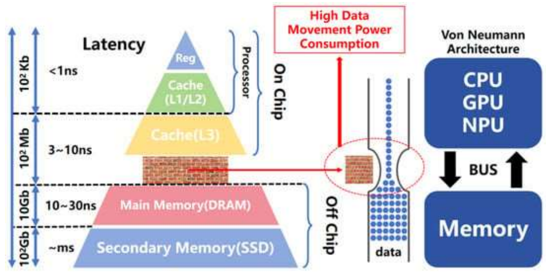

Figure 3: High Speed Low Power CIM SRAM

2. **III. Overview of VLSI Architecture**

Very Large Scale Integration (VLSI) is a technology that allows for the integration of a large combination of transistors and circuits into a single chip. This architecture has transformed the design and fabrication of integrated circuits/circuits–in–package (IC/ CIP) to enable the development of versatile systems on chip (SoC) that propose numerous devices including cellular phones, laptops, IoT, and medical equipment.

VLSI architecture can be classified into two main categories: digital and analog. Discrete signal VLSIs, used mainly in processing digital signals are generally employed in processors like microprocessors and memory like micro memories [6]. Analog VLSI circuits, however, work with analog signals that are created and processed with infinite data points and time frames and are commonly used in audio processing, r/fire signal processing, and sensor interfaces. Most of the present-day systems have both digital and analog elements and as a result, there is a lot of interaction between the two and as such there is a lot of mixed-signal design.

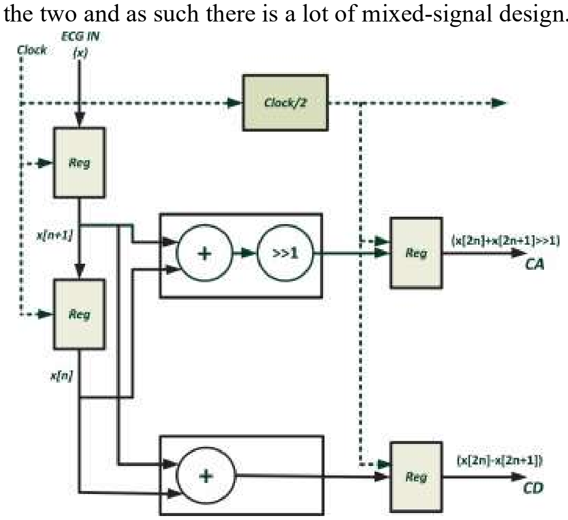

Figure 4: ECG Feature Extraction VLSI Architecture

VLSI is advantageous due to the small scale achieved; the size of devices has been reduced while their efficiency has been increased. This miniaturization is possible with the help of such design strategies as hierarchical design, modularity, and abstraction layers. Hierarchical design enables the engineering process to subdivide a complicated system into sub-systems, while modularity only enhances the reuse of assembly components, hence shorter cycles of design and lower prices.

Another indispensable characteristic of VLSI architecture is the power efficiency. Due to the miniaturization of electronics and continuous integration of more functional capabilities into devices, power consumption is a critical aspect for both enhancing performance and device life. Several architectural energy reduction strategies are used in the VLSI circuits including dynamic voltage and frequency scaling (DVFS) clock gating, and power gating among others [7]. These techniques are particularly important for low-power IoT devices, which are usually equipped with battery power supply, and for which power consumption is an effective and disturbing factor.

The VLSI chips fabrication process comprises designing, laying down, simulating, and manufacturing of the required VLSI chips. EDA tools also come in handy in the current generation, as they allow the design methods to be more refined and are used for the design and verification of the final product.

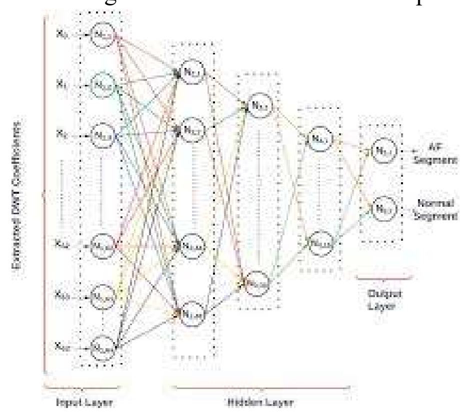

Figure 5: DNN-based Low Power VLSI

Consequently, VLSI architecture is the base or fundamental technology that is changing the entire face of modern electronics. ; Through the integration of a large number of components within a single chip, VLSI makes it easier to design small, fast, and efficient devices. As electronic systems have become more intricate and whilst the requirements for complex functionality and capabilities escalate VLSI technology will indeed respond to such needs, especially in the areas of IoT and audio processing.

3. **IV. Low-Power Design Techniques**

Due to the increasing use of compact devices and systems with a need for energy management, the methods of low-power design have risen to prominence, especially in designing the Internet of Things, mobile and wearable electronics, and communication devices. The use of these techniques is an effort to reduce power consumption just as much as to keep performance, dependability, and service features intact. Here are some key low-power design strategies used in integrated circuits and VLSI architectures:

1. Voltage Scaling

The use of a small current at low voltage is far more efficient than using a large current at high voltages; this is because the power consumption is the square of the voltage, that is power $\propto V^2$. Such methods include Dynamic Voltage Scaling (DVS) where the voltage supply to the circuits is altered corresponding to the workload or other performance correlates [8]. Reducing the voltage during inactive periods results in a significant saving of energy while sacrificing relatively little performance.

2. Frequency Scaling

Like voltage scaling, frequency scaling is another approach to dynamically alter the clock operating frequency of a circuit according to its load. The use of lower clock frequency during the idle or low-demand state is one of the best ways of minimizing dynamic power consumption. This technique is known as dynamic frequency scaling or in conjunction with DVS to achieve improved total power consumption.

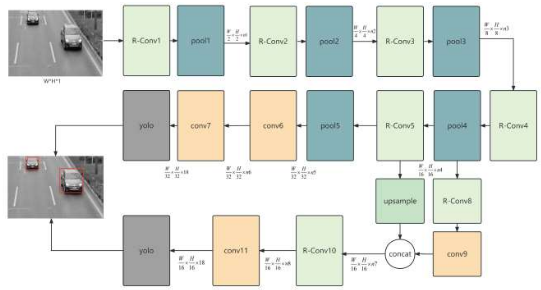

Figure 6: Rotating Kernel CNN Optimization

3. Power Gating

Power gating is a method that is employed to isolate the power supply to various circuit partitions, felt to be idle. By installing switches- transistors of very small size at the node in isolated parts of the circuit, called inactive blocks, leakage power can be constricted, thereby fulfilling power control in advanced technology as devices grow small [9]. It is effective in shutting down more components than are needed hence enabling energy conservation.

4. Clock Gating

Clock gating is the act of turning off the clock to those parts of a circuit that are not needed at the moment. One more advantage is that when the clock is stopped at some block, power consumption for toggling flip-flops and other sequential elements is also stopped [10]. This technique is exercised in both the new generation microprocessors and in mixed-signal designs in particular to reduce power consumption.

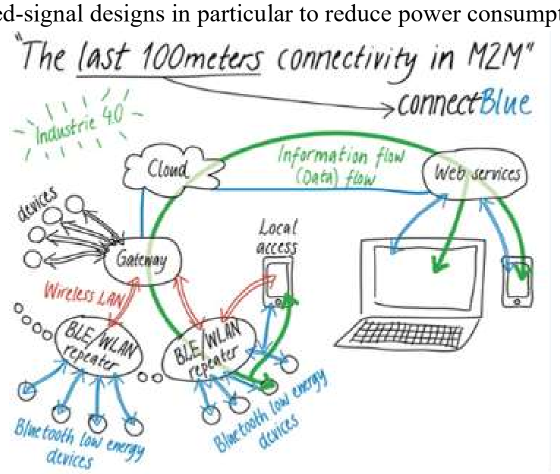

Figure 7: IoT Short Range Key

5. Multi-Threshold CMOS (MTCMOS)

MTCMOS uses several threshold voltages to allow turning "ON" or amplifying various transistors in the design. Power spec is satisfied where for non-critical paths, high-threshold voltage (HVT) transistors are used while critical paths are made of low-threshold voltage (LVT) transistors. This approach further assists in achieving an optimal solution where both the performance and leakage power are considered.

6. Adaptive Techniques

Adaptive techniques use feedback, the means of measuring performance or some other condition, in order to modify the amount of power usage during operation. For example, the self-tuning circuits are capable of changing some parameters in accordance with the temperature or load and thereby can use energy at its best.

Finally, techniques of low-power design play essential roles in the course of designing integrated electronic systems that utilize less power in their operation. Some of these proposed methods include; voltage and frequency scaling, power gating, clock gating, MTCMOS, and adaptive methods which help to lower power consumption but still meet the current performance and functionality. These techniques are most important in situations with IoT devices since the effective usage of energy determines how long and for what the devices will be functional [11].

4. V. Audio Feature Extraction Algorithms

Various audio feature extraction algorithms are indispensable pre-processing techniques in audio signal processing converting original audio data signals to meaningful representations for numerous applications, including speech recognition, music categorization, and environmental sound identification. By finding out parameters that are relevant in audio signals, such algorithms help machines to sort, recognize, and analyze sound suitably.

1. Mel-Frequency Cepstral Coefficients (MFCCs)

The most popular algorithm among a set of audio features is Mel-Frequency Cepstral Coefficients (MFCCs). This algorithm imitates an auditory perception human system that focuses more on specific frequencies where humans can hear. MFCCs are derived by converting the audio signal in the frequency domain by STFT; then each frequency bin is passed through the Mel filter bank that mimics the human ear; DCT is then applied to get a set of coefficients [12]. MFCCs have a special application because of speech and speaker recognition.

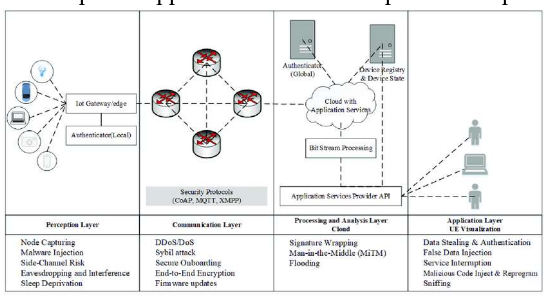

Figure 8: Security Risk Layered View of IoT

2. Spectral Features

Frequency-based features of signals include the spectral analysis that has a variety of features that can be obtained from audio signals. Spectral features that are prominent and generally used for songbird species identification are spectral centroid, spectral flux, and spectral bandwidth. The referenced spectral centroid gives information as to what the perceived brightness of a sound is in this case where it is located [13]. Spectral flux measures the first-order derivative of the spectrum, which, in turn, is useful for measuring timbre changes in music. These features are beneficial for data such as Music and audio data for some of the tasks like Music Genre classification and Music segmentation.

3. Zero-Crossing Rate (ZCR)

The Zero-Crossing Rate (ZCR) is a time-domain feature that quantifies the frequency of an occurrence of zero-crossing in the given signal. It works as a measure of the frequency content of the signal in the form of audio. A higher ZCR means a higher level of noise, and a lower ZCR is associated with smooth noise. This feature is particularly helpful when one wants to check for speech from music or simply check for silence in a recording.

4. Root Mean Square Energy or Root Mean Square Error (RMSE)

RMSE is another time domain feature that measures the energy of the time series signal at different time instances. This figure is obtained by finding the square root of the average of the squares of the amplitude values of the signal in question [14]. RMSE can be used in many applications, as it is especially useful for analyzing the loudness of signals and can be used while detecting actions of speech or while analyzing the dynamics of sound signals.

5. Chroma Features

Chroma features are concerned with how energy is spread across 12 different pitches; these give information on the harmony content of musical audio. These features have proved useful for simple tasks such as genre classification of music, chord recognition, and music similarity analysis.

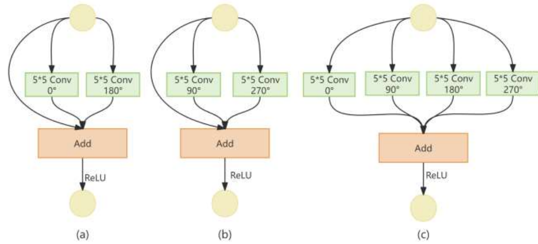

Figure 9: Rotating Kernel CNN Optimization

In sum, algorithms for feature extraction from audio data are of great importance for converting acoustic signals into useful representations. It is possible to analyze and interpret sound samples using MFCCs along with other features such as spectral features, zero-crossing rate, root mean square energy, and chroma features to develop new applications in speech and music processing fields. In the future, the development of better feature extraction techniques will further improve the performance of audio processing.

5. **VI. Case Studies of Existing VLSI Architectures**

VLSI (Very Large Scale Integration) has resulted in the production of large circuits that make up contemporary electronic systems. Several of the outlined cases are exemplary in explaining the diversification and flexibility of the VLSI architectures while presenting its importance in Industries.

1. Intel Core Processors

Intel Core processors are a VLSI breakthrough for personal computing architecture. They are some of the latest processor chips that incorporate one or more processing units in a single chip to optimize throughput and efficiency. The architectural design incorporates features such as Turbo Boost technology meant to monitor the clock frequency depending on the work to be done [15]. Moreover, Intel's 14 nm process design reduces power absorption while increasing the transistor transshipment rate and technical competence. They have brought remarkable effects to laptops, desktops, and servers with enhanced general & specialty operations.

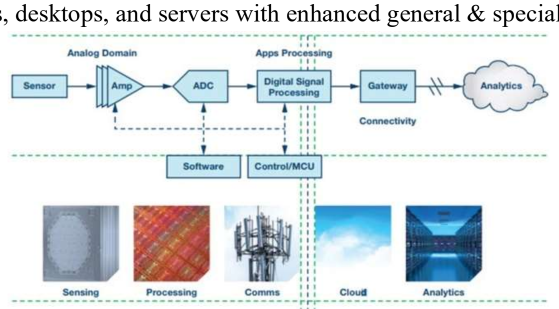

Figure 10: Precision Counts in IoT

2. NVIDIA GPUs

NVIDIA GPUs are the best examples of VLSI architecture perfectly designed to perform parallel computing. Features of the architecture of NVIDIA GPUs are thousands of tiny, power-saving cores performing various operations in parallel. Of the two, this parallel architecture is particularly useful for graphical computations and AI computations. The CUDA programming model which was developed by NVIDIA makes it easier for the developer to harness this parallelism for computations to work faster and faster [16]. The development of a new generation of GPU VLSI architecture has had a great impact on many areas, including but not limited to the game industry, artificial intelligence, and scientific modeling.

3. Integration of System-on-Chip (SoC) in Portable Communication Apparatus

Application processors like Apple A series chips entail the integration of numerous functional units in a VLSI chip including a central processing unit, graphics processing unit memory, and I/O interfaces. These architectures mean that some smartphone architectures are designed for power conservation, as well as performance, to allow smartphones to operate complex functions without high energy consumption. For example, the A14 Bionic chip incorporates a 5 nm process technology and includes 11.8 billion transistors that boost performance while decreasing power use. Other special additions that can be found on this architecture include machine learning and high-performance graphics and therefore it is considered as setting a standard for mobile computing [17].

4. Xilinx FPGAs

Xilinx FPGAs on Field Programmable Gate Arrays are described as versatile VLSI structures, whose functionality can be altered during and after production as required. FPGAs have an architecture of configurable logic blocks and programmable interconnection such that designers can map various functions to develop circuits for use in telecommunication, automotive systems, and many others. Due to the reconfigurability of FPGAs, these devices are widely used in prototyping and developing general-purpose hardware where benefits from fast time-to-market and flexible design changes are crucial.

5. The use of Digital Signal Processors (DSPs)

For these applications, Texas Instruments' DSPs are specifically optimized VLSI architectures capable of digital signal processing. These processors contain additional sets of activate instructions and designs that boost capabilities for iterative computations [18]. For instance, The TMS320 series is designed with parallel processing and dedicated hardware multipliers to be more efficient in the processing of complicated signal processing algorithms in real-time signal processing.

Furthermore, concrete examples of presently used VLSI implementations illustrate the potential and creativity prevailing in today's integrated circuits. They are used in a variety of computations starting from high-end processing to reconfigurable processors known as FPGAs, which help to develop new innovations in computer, graphic, telecommunication, and embedded analysis systems. VLSI architecture can therefore be concluded as being crucial to electronic design and development, especially given the increased technological advancement in future years.

Figure 11: Empowering Low Power Design for Portable Devices

6. **VII. Simulation and Evaluation Metrics**

Simulation and evaluation metrics are important factors that are used in the assessment of systems and processes in engineering, computer science, and telecommunication. These metrics also help researchers and engineers to compare and identify ways in which complex systems can perform to their designed requirements in real-world uses.

1. Purpose of Simulation

An important simulation technique is used to simulate and predict the behavior of a system when it is in a certain state. It enables the modeling of numerous what-if situations without any need to conduct physical tests that are costly and dangerous [19]. By simulation, one can determine how a system will behave under changes in such factors as parameters, system geometry, or input variables which will improve the comprehension of the dynamic and efficient behavior of systems.

2. Types of Simulation

There are several types of simulation methods, each suited for specific applications:

Discrete Event Simulation (DES): This approach is used to model systems whose changes occur in certain time instants, which makes it especially useful to solutions describing processes such as queuing systems, manufacturing, or networks.

Continuous Simulation: Applicable to situations, that are progressive in nature, this approach is typical for the physical systems including the fluid dynamics and thermal analysis.

Monte Carlo Simulation: This is a statistical technique that applies randomness to analyze the societies and their composition and is effective for the estimation of risks and planning, it gives insight into the unknown factors.

3. Evaluation Metrics
To assess the performance of simulated systems, various evaluation metrics are utilized, depending on the context:

Accuracy: This simple metric quantifies how well the simulation outcomes match the actual results or anticipated target numbers essential in model verification that applies in areas such as machine learning and forecasting [20].

Latency: For this reason, in communication and real-time systems, throughput is measured as the time taken for the input to be processed and delivered the output, and as this is important for the responsiveness of applications like online gaming and lives streaming, low latency is preferred.

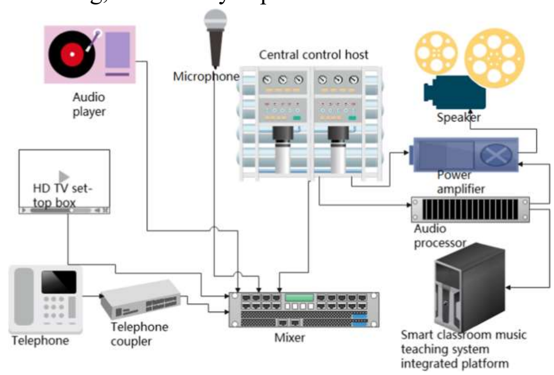

Figure 12: Design and Innovation of Audio IoT

Throughput: This measure specifies the amount of data, which has gone through a network or a data processing system in a given time period This measure is very important in the evaluation of the performance of a network or data processing system.

Energy Consumption: Applied and important in IoT and VLSI designs, it defines the power utilization trend of a system and how best it can be improved.

Resource Utilization: This is a metric that measures the effectiveness of the usage of the system resources for instance CPU, memory, and or bandwidth to establish the potential problems in the utilization of such resources.

One of the powerful activities practiced in the area of simulation is benchmarking, which is the establishment of a standard measure in terms of which system performance can be evaluated. Later, a comparison of simulation results with standard reference data can be made in order to assess the presence of critical points or the identification of their nature, as well as for further optimization.

7. **VIII. Future Trends in Audio Feature Extraction for IoT**
With the development of the concept of the Internet of Things (IoT) active voice issues have started to require more efficient audio feature extraction approaches. These methods are necessary to make IoT devices perform enough computations and understand sounds as to provide capabilities for voice recognition, smart home systems control, environment monitoring, and many other things that are not only potential but already existing. A few directions of future work in the context of audio feature extraction for IoT are evident below.

1. Improved ML algorithms
The future trend predicted about feature extraction from audio is promising, with upcoming developments in machine learning, and deep learning methods planned for breakthroughs in audio feature extraction as well as in various other fields [21]. Future algorithms will use convolutional neural networks (CNNs) and Recurrent Neural Networks (RNNs) for feature learning from the raw data. This approach reduces the level of involvement in the

time-consuming process of feature extraction which increases the efficiency and accuracy of learning the complicated acoustic patterns in speech, music, and other sounds.

2. Edge Computing Integration

People will start using edge computing in conjunction with the audio feature extraction processes more and more. By processing the audio signal at the edge of the network, nearer to the source, less bandwidth is consumed, more real-time analytics can be performed, and less dependency on the cloud is required. This trend is especially important for the networks that need an immediate response, conversing applications like voice-activated assistants and protection systems. Future improvements in the edge processing power will enable better feature extraction for audio on low-power platforms.

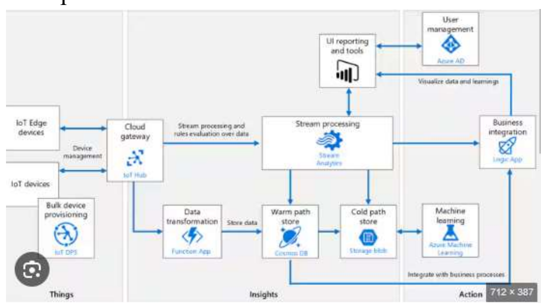

Figure 13: IoT Device Acceleration

3. Context-Aware Feature Extraction

The next generations of audio feature extraction will consist of context-sensitive algorithms that take into account the surroundings when processing an audio signal. This approach will allow IoT devices to filter out sounds that are not relevant due to the environment in which the device finds itself [22]. For instance, information-gathering tools such as smart speakers might look into commands that a user gives out, rather than the noise in the background as a means of enhancing the general experience of voice recognition to the user.

4. Multimodal Data Fusion

With the advancement of IoT devices, the use of audio feature extraction with other sensory features such as visual, temperature, motion, and the like will eventuate. Advanced signal fusion technologies will improve awareness of intricate surroundings; processes will link sound input to visuals or situations. This trend can be greatly beneficial for applications like smart surveillance where the audio signals can also add to the video footage to make the detection and identification.

5. Improved Low-Power VLSI Architectures

There will be an enormous challenge in optimizing audio feature extraction for battery-operated IoT devices through the development of specific low-power VLSI architectures for audio processing. These architectures will also help in computation and energy control so that the devices can easily perform the kind of analysis needed for sound data.

6.Standardization and Interoperability

With the growth in the IoT environment, there will be a lot of pressure towards standardization with concerns about audio feature extraction. It is expected that the creation of a common framework and protocol would enable integration between these devices and applications; in effect, creating awareness and acceptance of audio processing technologies across the digital solutions ecosystem [23].

In conclusion, there are remarkable promising developments in the next few years for audio feature extraction for IoT from the advances in machine learning and deep learning and implementation of edge computing, context awareness, modality fusion at the IoT node, and low-power IoT node architecture. These trends will improve the ability of the IoT devices to work with the audio data and the overall employability of the IoT devices within various applications.

8. **IX. Conclusion**

For many developers working on multiple projects in the burgeoning field of IoT, feature extraction must be performed effectively and on time. Given the fact that IoT devices are increasingly used across industries, extending from houses and health care to manufacturing lines, the capacity to analyze audio signals with a fine measure of efficacy is essential. Various advanced audio processing techniques effective the capability of these devices in terms of intelligently interacting with an expected command, the environment, and the context.

Several trends in the future of audio feature extraction in the Internet of Things IoT are lined up to revolutionize how audio data is managed. One can safely assume that the further development of machine learning algorithms, especially deep learning models, will define this change. Such algorithms can learn when needed to feature extract from raw data and can help improve domain accuracy in recognizing complex audio patterns. If more sophisticated and enhanced Machine Learning algorithms are included in IoT systems, all manners of audio input including human voices and sounds in the environment hence the functionality of IoT devices would improve greatly.

The one big shift in the audio feature extraction domain will be the growing use of edge computing. As compared to the cloud, edge computing minimizes latency and bandwidth demands since audio data will be processed on IoT devices. This local processing ability is particularly important for real-time applications where response time must be fast, for example, voice-activated assistants and surveillance systems. The emergence of the conclusion of edge computing will enable IoT devices to enhance the procedure of analyzing the audio signals while being cost-effective in content bandwidth and data security since the various sound data which is collected in the course of implementing the system do not need to go through the internet server for the analysis to be conducted hence improving on the privacy of data collected.

One of the classic and productive areas that will define the future of audio processing in the IoT environment is context-aware feature extraction. IoT devices can also filter over the contextual info, regarding which info is most relevant for a particular user and what is not, to screen relevant sounds. For instance, a smart speaker might pay attention to the user's directions but not other people talking or singing in the background. This capability is to improve user experience and guarantee that the devices are responding in the right manner to those audio signals. It can be seen that as more and more devices learn more about their environment, they will be able to work in an environment that is more sensitive to the environment and therefore give users an interaction with the devices which is more natural.

Another area of importance in the future of feature extraction for audio features will be the advancement of better low-power VLSI design implementations. Since IoT devices are frequently portable and reliant on batteries, their power utilization must be considered while performance is still successful. Advanced System-on-Chips optimized for audio signal processing will enable efficient computation of various audio analyses to manage the next-generation smart gadgets which are very sensitive and have limited power capacity in their battery compartments. This optimization will be imperative in a way that will likely expand audio-processing capabilities in IoT-enabled apps spanning the mobile and remote realms.

Another important characteristic is also standardization and interoperability as the IoT environment grows day after day. The creation of uniform structures and conventions for audio feature extraction will ease the diffusion of these technologies into a variety of devices and platforms. Standardization will facilitate working in a single environment and interaction among various IoT devices to improve their access to knowledge and operation. Manufacturers are now applying the audio processing of the signal uniformly, and this will translate to aggregated compatibility and interoperability and thus a more seamless and friendly IoT ecosystem.

Therefore, the future of future of audio feature extraction for IoT devices looks bright where the existence of more advanced features that control the reception of their audio features is expected to improve their reliability, efficiency, and compatibility. Machine learning, edge computing, context-aware processing, multimodal data fusion, low-power VLSI architectures, and standardization are the key themes that will stimulate this sector. About these tendencies, we can forecast that future advanced IoT devices will be equipped with high-precision audio identification systems which will considerably diversify the IoT devices' functions, improve user experience, and open new horizons in further IoT technology development. This evolution will not only improve the standard of individual devices but also of smart ecosystems to promote change in the reciprocity between human beings and their environment as well as the surrounding technology. Looking into the future, further research and analysis of advanced solutions for audio feature extraction methods will be crucial in advancing the Internet of Things and achieving an environment in which advancement in technology fits perfectly in our society.

9. **X. Reference**

[1] Fariselli, M., Rusci, M., Cambonie, J. and Flamand, E., 2021, June. Integer-only approximated MFCC for ultra-low power audio NN processing on multi-core MCUs. In *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)* (pp. 1-4). IEEE.

[2] Cerutti, G., Cavigelli, L., Andri, R., Magno, M., Farella, E. and Benini, L., 2022. Sub-mW keyword spotting on an MCU: Analog binary feature extraction and binary neural networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *69*(5), pp.2002-2012.

[3] Mourrane, S., Larras, B., Cathelin, A. and Frappé, A., 2021, June. Event-driven continuous-time feature extraction for ultra low-power audio keyword spotting. In *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)* (pp. 1-4). IEEE.

[4] Profentzas, C., Almgren, M. and Landsiedel, O., 2022, September. MiniLearn: On-Device Learning for Low-Power IoT Devices. In *EWSN* (pp. 1-11).

[5] Li, Q., Yang, Y., Lan, T., Zhu, H., Wei, Q., Qiao, F., Liu, X. and Yang, H., 2020. MSP-MFCC: Energy-efficient MFCC feature extraction method with mixed-signal processing architecture for wearable speech recognition applications. *IEEE Access*, *8*, pp.48720-48730.

[6] Gerum, C., Frischknecht, A., Hald, T., Bernardo, P.P., Lübeck, K. and Bringmann, O., 2022. Hardware accelerator and neural network co-optimization for ultra-low-power audio processing devices. *arXiv preprint arXiv:2209.03807*.

[7] Chundi, P.K., Wang, D., Kim, S.J., Yang, M., Cerqueira, J.P., Kang, J., Jung, S., Kim, S. and Seok, M., 2021. Always-on sub-microwatt spiking neural network based on spike-driven clock-and power-gating for an ultra-low-power intelligent device. *Frontiers in Neuroscience*, *15*, p.684113.

[8] Chen, Q., Chang, Y., Kim, K., Gao, C. and Liu, S.C., 2023, May. An area-efficient ultra-low-power time-domain feature extractor for edge keyword spotting. In *2023 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1-5). IEEE.

[9] Lamrini, M., Chkouri, M.Y. and Touhafi, A., 2023. Evaluating the performance of pre-trained convolutional neural network for audio classification on embedded systems for anomaly detection in smart cities. *Sensors*, *23*(13), p.6227.

[10] Nair, A.R., Chakrabartty, S. and Thakur, C.S., 2021. In-filter computing for designing ultralight acoustic pattern recognizers. *IEEE Internet of Things Journal*, *9*(8), pp.6095-6106.

[11] Wu, C.H. and Chang, T.S., 2024. A Low-Power Streaming Speech Enhancement Accelerator For Edge Devices. *IEEE Open Journal of Circuits and Systems*.

[12] Yang, C., Wei, Z., Gao, H., Heng, C.H. and Zheng, Y., 2024. Towards Ultra-Low Power Transceivers for Pico-IoT. *IEEE Nanotechnology Magazine*.

[13] Nair, A.R., Nath, P.K., Chakrabartty, S. and Thakur, C.S., 2024, January. Multiplierless in-filter computing for tinyML platforms. In *2024 37th International Conference on VLSI Design and 2024 23rd International Conference on Embedded Systems (VLSID)* (pp. 192-197). IEEE.

[14] Schizas, N., Karras, A., Karras, C. and Sioutas, S., 2022. TinyML for ultra-low power AI and large scale IoT deployments: A systematic review. *Future Internet*, *14*(12), p.363.

[15] Chen, F., Un, K.F., Yu, W.H., Mak, P.I. and Martins, R.P., 2022. A 108-nW 0.8-mm 2 Analog Voice Activity Detector Featuring a Time-Domain CNN With Sparsity-Aware Computation and Sparsified Quantization in 28-nm CMOS. *IEEE Journal of Solid-State Circuits*, *57*(11), pp.3288-3297.

[16] Miriyala, V.P.K. and Ishii, M., 2020. Ultra-low power on-chip learning of speech commands with phase-change memories. *arXiv preprint arXiv:2010.11741*.

[17] Chen, Y., 2023. *Low-Power Localization Systems with Hardware-Efficient Deep Neural Networks* (Doctoral dissertation).

[18] Wang, X., Magno, M., Cavigelli, L. and Benini, L., 2020. FANN-on-MCU: An open-source toolkit for

energy-efficient neural network inference at the edge of the Internet of Things. *IEEE Internet of Things Journal*, *7*(5), pp.4403-4417.
[19] Kim, K. and Liu, S.C., 2023. Continuous-time analog filters for audio edge intelligence: Review on circuit designs [feature]. *IEEE Circuits and Systems Magazine*, *23*(2), pp.29-48.
Jokic, P., 2021. *Efficient Neural Network Acceleration for Low Power Edge Processing Devices* (Doctoral dissertation, ETH Zurich).
[20] Haris, G., Shakeel, M., Kenji, N., Katsutoshi, I. and Kazuhiro, N., 2021. Speech-Classification in Edge-Computing for IoT Applications. *人工知能学会第二種研究会資料*, *2021*(Challenge-058), p.06.
[21] Andri, R., 2020. *Machine Learning Acceleration for Tightly Energy-Constrained Devices* (Doctoral dissertation, ETH Zurich).
[22] Elkanishy, A., Furth, P.M., Rivera, D.T. and Badawy, A.H.A., 2021. Low-overhead Hardware Supervision for Securing an IoT Bluetooth-enabled Device: Monitoring Radio Frequency and Supply Voltage. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, *18*(1), pp.1-28.
[23] Baskar, R., Dhanagopal, R., Elangovan, K. and Gunasekaran, K., 2021, July. VLSI based architecture in ECG monitoring for adaptive power management in wireless bio signal acquisition network. In *Journal of Physics: Conference Series* (Vol. 1964, No. 6, p. 062090). IOP Publishing.