International Journal of Grid and Distributed Computing
Vol. 9, No. 6 (2016), pp.279-288
http://dx.doi.org/10.14257/ijgdc.2016.9.6.26

FPGA Design of Voice Enabled Ignition using G279 for Modal Based Speech Compression

Kelan McLean[1] and Marcus Lloyde George[2]

[1]*Department of Electrical and Computer Engineering*
University of the West Indies
St. Augustine, Trinidad
[2]*Dept. Electrical and Computer Engineering*
University of the West Indies
St. Augustine, Trinidad and Tobago
Kelan.mclean@my.uwi.edu, Marcus.George@sta.uwi.edu

Abstract

Voice enabled ignition combines the speaker recognition and word recognition aspects of speech recognition. It replaces the function of a key in the starting of the ignition system of a car. An Fpga design incorporates the required components of a generic speech recognition system and uses the unique capabilities of hardware in term of parallelism to improve performance. The compression of speech for storage and playback was facilitated by the usage of the G729 standard for compression of speech.

Keywords*: Field Programmable Gate Array, Word Recognition, Speaker Recognition, Modal Based Speech Compression, G729, Voice Recognition, Noise Filtration*

1. Introduction

A. Human Speech

Human speech is governed by both patterns and frequencies that are due to a mixture of the person's background and the person's biological makeup with respect to their mouth and throat. Learned characteristics include the person's accent and manner of speaking whilst the configuration of the person's mouth and throat make up a person's voiceprint or voice biometrics[1].

Voice biometrics would provide the needed authorization levels because they are unique to each individual. This means that they can confirm who a person is and therefore assess their authorization levels. This technology, as stated by [2] is accurate with a 98% confidence interval, leading to a safe accurate and dependable system[3].

B. Types of Voice Recognition and Strategies in voice Recognition

The patterns of human speech may take various forms and the difference can be seen when comparing two person's saying the same word as shown in Figure 1 below.

ISSN: 2005-4262 IJGDC

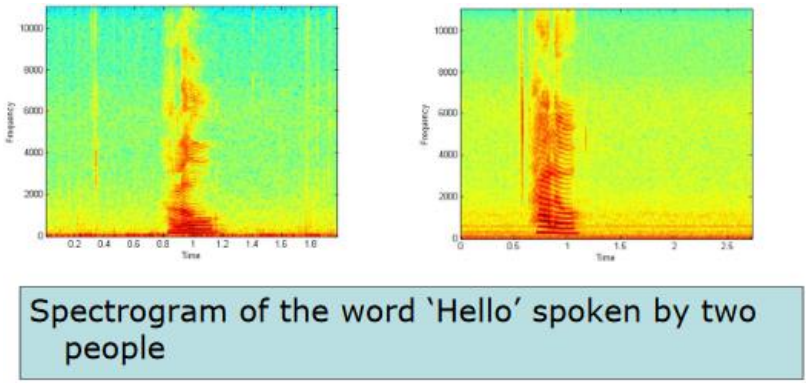

Figure 1. Spectrogram of the Word "Hello"[4]

Therefore the approach taken to recognizing voice would have to account for differences in manner of speaking, including accents, differences in enunciation, speech impediments and tonal variations in voice. There are two primary approaches to recognizing voice which are large vocabulary small user base and small vocabulary large user base. For ignition, a simple start phrase such as "engine start" could be used which would constitute a minimal vocabulary requirement for the system.

As such the system need only compare the spoken and detected language with the small stored vocabulary used to start the engine and determine whether it is in appropriate confidence interval for the phrase used to start the engine. In comparison, if there was a larger vocabulary set, the phrase would have to be broken down and analyzed to determine meaning through the syllabic decomposition of speech. It would then be compared reconstructed to determine meaning before being compared to the stored language templates for speech.

C. Car Security and Ignition

The car ignition system is controlled typically by the car key. A car key therefore provides three main functions. The key allows access to the car itself, access to the ignition mechanism and the ability to start the car. Voice recognition would replace these functions with two capabilities, detection of speaker and detection of speech. The former provides authorization and controls access to the ignition mechanism and the latter controls the actual starting of the car.

Voice authorization is uniquely suited to an ignition system because of the relative cheap cost of the voice input equipment when compared with alternatives involving biological authorization. Additionally, it offers the advantages of non-forgetting implementations, convenience and the ease of accommodating multiple users to account for cars which driven by more than a single person.

A car's ignition works by the engine igniting by a fine spray of petrol and the sparks produced from spark plugs. This allows the engine to start turning due to the small explosion pushing a piston down, spinning the engine and forcing it back up. The spinning motion allows air and fuel to continue flowing into the chamber. The key or voice ignition would be in charge of the initial interactions, which allow the fuel to be lit, that is the fine spray of petrol and the sparking of the spark-plugs.

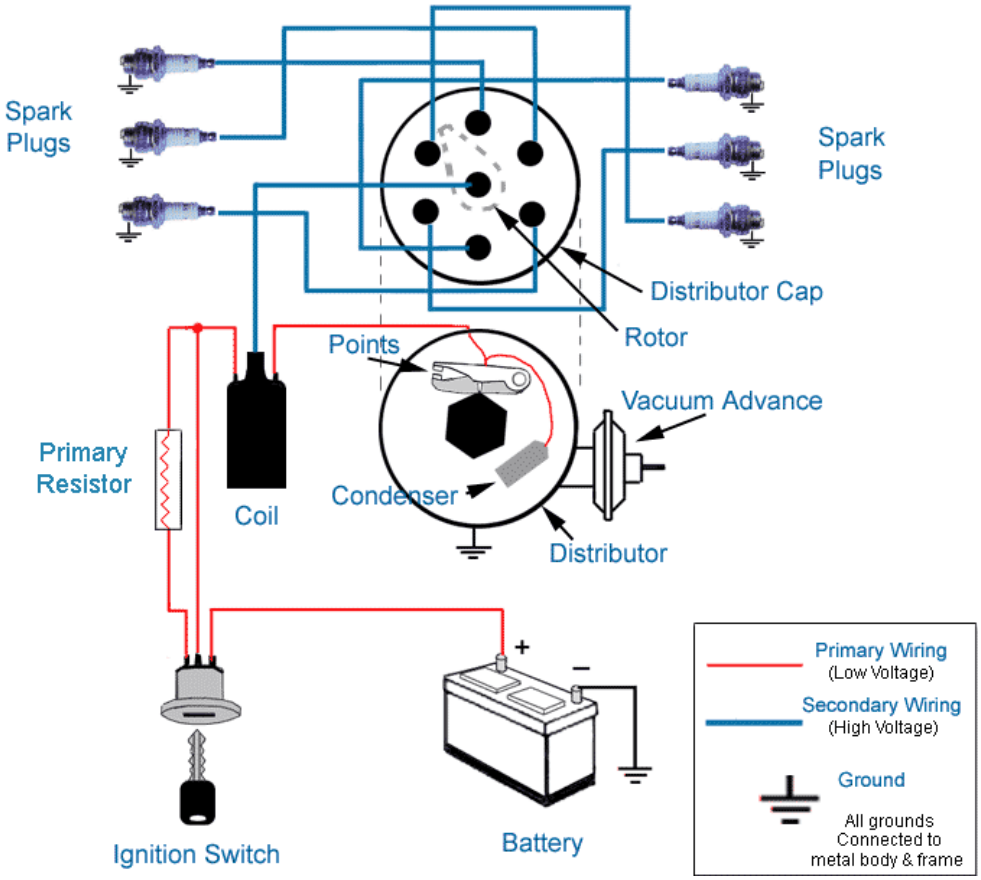

Figure 2. Typical ignition System[5]

D. Modal Based Speech Compression

The system may require the storage and implementation of auditory responses either to confirm the receipt of an input or to prompt the user to try again. This would therefore provide feedback beyond the simple starting of the car to ensure the user is always of aware of what's occurring in the system. Modal based speech compression is used to reduce the storage space of speech. This is especially important for playback but can also be used as a strategy for storing authorization credentials which are in the form of voice.

2. Review of Previous Literature

A. Adapted Strategy for voice recognition

The approach to voice recognition is based on the purpose and the function of the voice recognition. It can be divided into two broad categories of talker recognition and voice recognition [3]. Talker recognition can be further subdivided, such that they are either relevant to text, which builds speech parameters from a read text or not relevant to text which doesn't rule the pronunciation content of the speakers. The purpose would be further subdivided into either for talker recognition or talker confirmation. Talker recognition judges a voice that needs to be identified from several talker and talker confirmation judges whether or not a voice comes from a certain talker, having a yes or no output. Two important parameters of the system would be the capability for error identification and the rejection identification ratio [3]. These refer to the mistake made by the voice signal of non-users accepted by the system and the mistake made by the voice signal of users rejected by the system.

B. Model Based Speech Compression g279

Modal based speech compression which has applications in voice activity detection and includes a linear predictive stage is required by [3]. In that implementation, the PAR-COR function was used to compute the linear prediction coefficients and perform the linear prediction analysis. Additionally there was an included noise filter in the voice analysis component. The output of this analysis component could be directly compared with the voice database to determine and match and perform the voice recognition.

The implementation branches the usage of the received voice input into voice recognition and compression for storage or playback (by decompression). This was done after the digital distillation of the signals in order to extract the needed parameters which would prove most relevant to speech recognition.

The [6] proposes the usage of the LFCC to deal with speech resynthesized using the G729 codec. The codec has applications in speech compression particularly for storage and transmission[7]. The compressed speech can then be resynthesized by applying the G729 resynthesized database. Other strategies propose the usage of logistical regression which may provide a better performance, indicating the suitability of the voice activity detector, despite not providing the optimum performance [8].

C. Previous Implementations

Several different approaches to voice recognition either by hardware or software can be seen. Software implementations such as [6] and [3] typically employ computationally difficult models as seen specifically in [access controls] inclusion of a software component to simplify the computational process. Hardware implementations utilize numerous memory models and parallel processing to perform specifically the digitization stage.

D. Noise

Noise plays an important part in the recognition of speech as it changes the input signal and increases the difficulty of detection. Therefore, after collection of the voice signal, a stage needs to be included which will allow the filtering of noise and the obtaining of the useful voice signals.

According to [8]the usage of a robust voice activity detector is particularly effective for stationary and non-stationary noise environments of which a motor vehicle is one. The reduction in detection error necessitates the usage of a VAD such as the robust VAD or the G729 annex b.

E. Considerations

Some other considerations suggested in the literature are the power consumption, the error rate and error rejection and the speed of processing and memory [9]. Additionally, in voice recognition, the rejection of noise plays an important part especially when environmental variables are beyond control [8] An ignition system, however, allows enough control of the environment to allow filtering stages to remove noise and the speed needs only to be fast enough that the delay does not impact the user's patience. Power consumption can be monitored by the ignition system not needing to be active unless at all times.j

F. Other Applications

Other applications of similar speech or voice recognition technologies discussed include isolated word recognition, connected word recognition, continuous or fluent speech recognition, speech understanding systems and spontaneous conversations systems[10]. The literature states that the general purpose speech recognition systems follows the pattern of the block shown in Figure 3, below.

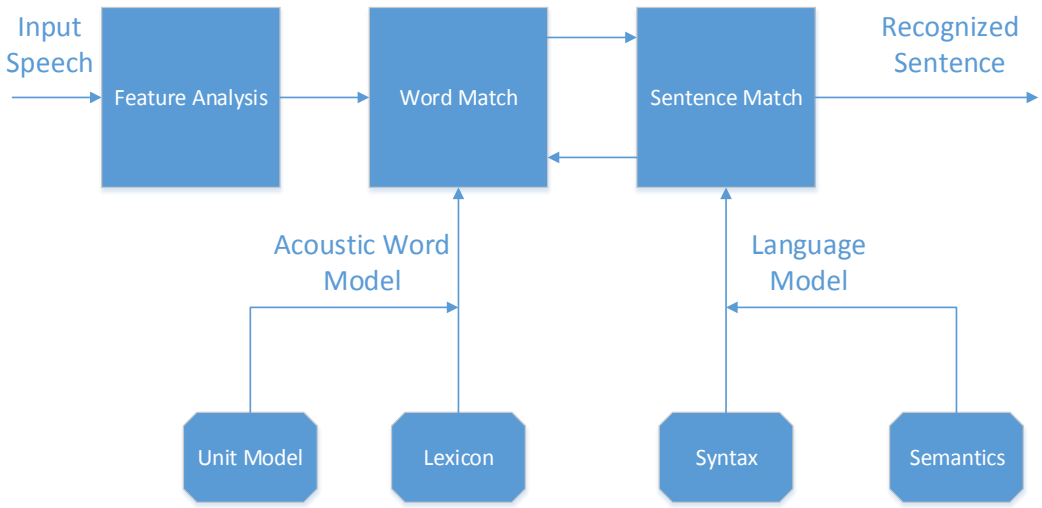

Figure 3. General Purpose Speech Recognition Block Diagram

The general purpose speech recognition format can be modified to suit specific applications. Emphasis may be placed on the function of certain blocks to suit the specific processing requirement.
The same general system is implemented in applications such as

- voice dialing
- automation of operator services
- automation of directory assistance
- voice prompter
- Directory assistance and reverse directory assistance
- Information services
- Voice dictation

G. Other Requirements

Some general requirements of speech or voice recognition applications[10] are:

- Good user interfaces and ease of use
- Robust to the confusion in human-machine communication
- Good models of dialog to move conversation forward
- Task is matched to technology

3. Methodology

The design for the voice enabled ignition would include the following components deemed essential for a minimalistic functioning system design:

- End user input system
- Digital Filtering (inclusive of noise filtering step)
- Characteristic Distilling (Phenomes)
- Module Matching (Training and Recognition) (Database)
- Compression Processing (Decompression-Voice output or Compression using g729) (Database)
- Playback Module

The structure of the system is shown in Figure 4.

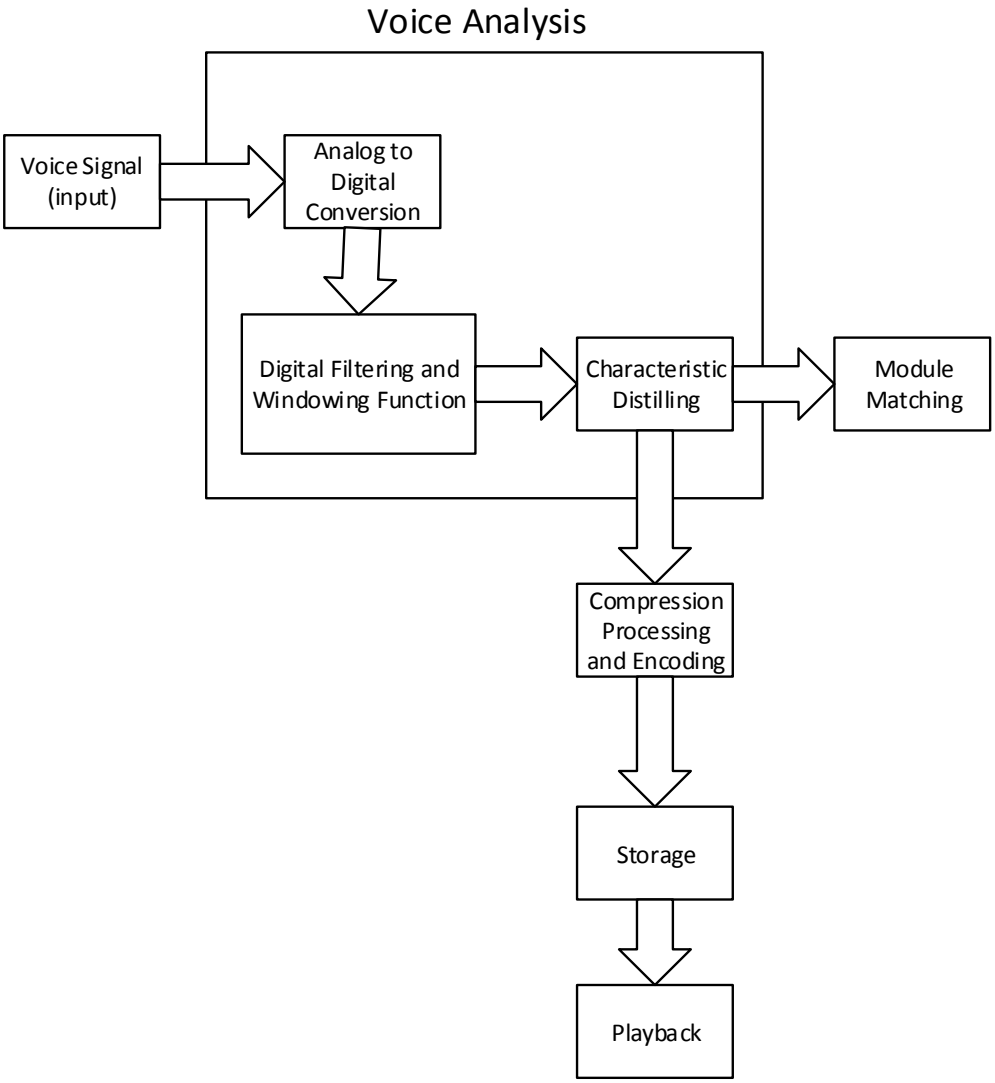

Figure 4 System Structure for Voice Enable Ignition

The end user system would need to consist of a transceiver for speech which would be a microphone to convert the person's voice into an analog electrical signal. This microphone will therefore contain the input signal along with any noise or background sounds within the vocal frequency range as typically microphones filter a component of sound. This filtering performed by the selectivity of the microphone does not remove the necessity for digital filtering. The input will therefore be the user's voice or background noise and the output would be considered as an unfiltered analog voice sample.

The digital filtering component would consist of a digital signal processor which would perform digital filtering on the input signal. Therefore an FPGA with enough processing power for digital signal processing functions would be required for these computations. The processor would continuously perform the short time Fourier transform using a hamming window as suggested[6]. The input signal would need to be digital in nature, therefore the analog speech signal would be converted using an analog to digital converter.

Characteristic distilling would be used to extract some of the vocal parameters from the speech. The approach used would be to perform frequency analysis to separate speech into spaces and phenomes and based on the specific composition of spaces and phenomes

to perform speaker recognition. These characteristics would be used in the compression and matching functions of the system.

Module matching would allow both voice recognition and vocal training to be performed by the vocal ignition system. The training system compares the person's speech to the stored template and based on the differences or similarities to the template adjusts the vocal recognition parameters to allow recognition of the specific person's voice pattern. The voice biometrics would also be included in this step. After the system is trained, it should be able to perform efficient recognition of the speaker's voice.

Compression processing would be facilitated by the usage of the g729 encoder which would compress the speech of the speaker for storage or playback. When compressed for storage, the person's unique voice signature can be stored in the system database for usage in future recognition. Playback is used as a form of feedback to allow the user to know if the data signal was properly captured as the sound being played would be indicative of the sound to be stored. If there is excessive noise in the background or if the sound is otherwise distorted, the user would be able to store or discard the speech sample. The playback module would simply be a speaker which can play the decompressed speech sample.

Hardware to Realize Design

The hardware components needed to implement this design are:

- FPGA
- Microphone
- Analog to Digital Converter
- Sensors and switches
- Speaker
- Ignition System

And have been realized as per the given Figure 5, where various interconnections have been shown according to their input output conditions.

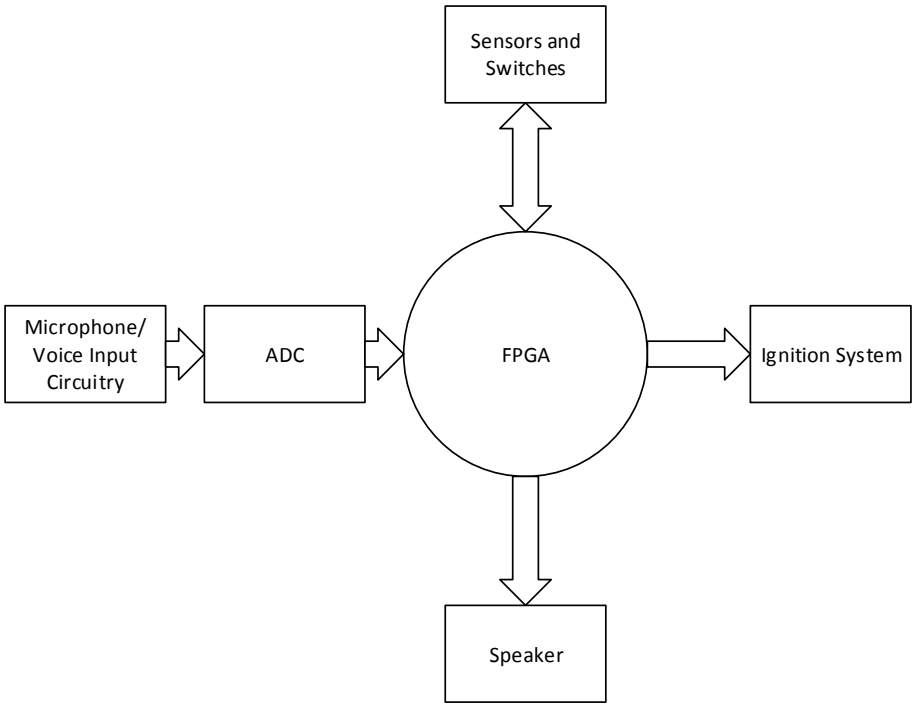

Figure 5. Hardware Realisation of the Design

The FPGA would be the main controlling unit of the ignition system. It would perform the sequencing as well as the discrete signal processing operations of the system. One specific requirement of the FPGA is a sufficient memory space to hold the voice database which would allow the onboard storage of all the parameters and make retrieval and usage of the data quicker and easier. Alternatively, an additional component would be needed for storage, such as an external RAM module.

The FPGA must also be suited for discrete signal processing operations. Whilst theoretically all DSP operations can be implemented on a FPGA, a FPGA with an integrated DSP would prove an asset in ease of system design and move the complexity of the design away from implementing DSP operations on the FPGA and instead to having a functional voice enabled ignitions system.

The FPGA would control the switching on and off of the microphone with a desired stimulus, such as the vehicle door opening and the microphone would continuously accept inputs for a period of approximately 10 minutes. The microphone would be attached to the analog to digital converter to convert the input audio signals into a digital format which would be recognizable by the FPGA. The speaker is for feedback to the user from the system.

FPGA Control Structure

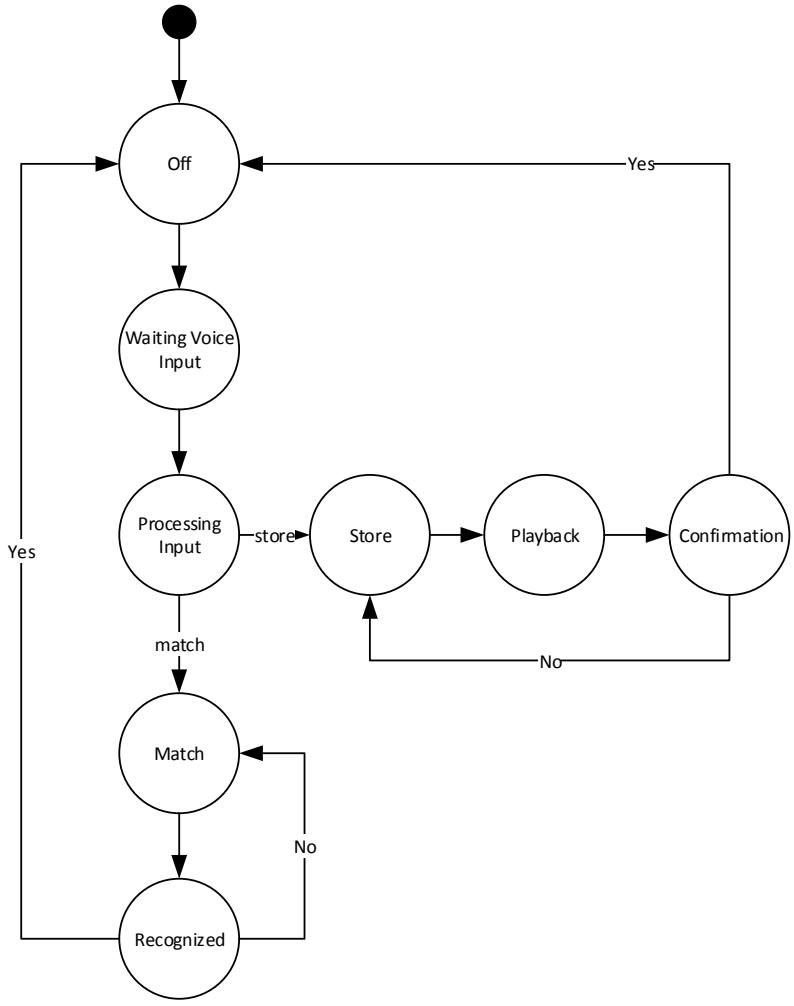

Figure 6. Control Structure of the Design

The control structure of the proposed design has been shown in Figure 6 and the steps have been explained in the previous sections also.

4. System Analysis

The system presented in this paper is a simple voice enabled ignition system. It is meant to replace the security of the key ignition with the ease of voice ignition. Implementation on FPGAs can reduce the overall energy consumption of the vehicle as well as create an inexpensive implementation which can be easily integrated into an already existing vehicle.

The design can be expanded to include configurability of the system, including allowing the user to define what audio signal starts the ignition, adding other authorized users of the car to the voice enabled ignition or defining the start signal for the system, which could be for example the opening of the door, the detection of pressure on the driver's seat or any other suitable stimulus.

Voice controlled systems can be used for other non-critical car operations such as radio control, windshield wiper and sprinkler control, air condition control and window control. Systems within the car that are not safety critical can be converted to voice controls, whilst maintaining traditional controls. Depending on the confidence interval of the voice control used, as the technology develops more operations can be switched. The greatest consideration is if the function is accidentally triggered either by some error in the system or by the user, for example saying a control phrase, that the impact would not endanger lives.

5. Conclusion

The FPGA hardware driven implementation of a voice enabled ignition system would allow a secure implementation of the essential vehicular function. A FPGA can be used as an effective control for ignition, utilizing the superiority of hardware in terms of speed and cycle minimization due to the use of parallelism and pipelining, reliability as there are no abstraction layers which may be corrupted and cost as FPGA design can be implemented as permanent hardware modules.

The usage of g729 for voice parameter extraction allows for compression processing which enables playback and storage as well as offers the possibility for expansion of the system. Expansion can take place by the storage of multiple user voice credentials. Compression also allows the values to be stored in a smaller memory space allowing for scalability of the system.

References

[1] E. Grabianowsky, "How Speech Recognition Works," 10 November 2006. [Online]. Available: http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition.htm. [Accessed 20 March 2016].

[2] J. M. Robin, "Ask an Engineer - Can I Start My Car with a Voice Command," 13 September 2011. [Online]. Available: http://engineering.mit.edu/ask/can-i-start-my-car-voice-command. [Accessed 20 March 2016].

[3] B. Cui and T. Xue, "Design and Realization of an Intelligent Access Control System Based on Voice Recognition," in *ISECS International Colloquium on Computing, Control and Management*, Sanya, 2009.

[4] F. Mudeen, *Lecture 9 - Spectrum Analysis,* UWI, 2016.

[5] Carparts.com, "Ignition Systems a Short Course," [Online]. Available: http://www.carparts.com/classroom/ignition.htm. [Accessed 15 March 2016].

[6] D. Yessad and A. Amrouche , "Fusion Strategies for Distributed Speaker Recognition using Residual

Speaker Based G729 Resynthesized Speech," in *16th International Conference on Information Fusion*, Istanbul, Turkey, 2013.

[7] ITU-T, "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algeabric-Code-Excited Linear-Prediction (CS-ACELP)," 1996.

[8] O. Varela, R. San-Segundo and L. A. Hernandez, "Robust Speech Detection for Noisy Environments," *IEEE A&E Systems Magazine*, pp. 16-23, 2011.

[9] T. Sledevic, G. Tamulevicus and D. Navakauskas, "Upgrading FPGA Implementation of Isolated Word Recognition System for a Real-Time Operation," *ELEKTRONIKA IR ELEKTROTECHNIKA*, vol. 19, no. 10, pp. 123-128, 2013.

[10] L. R. Ragbir, "Applications of Speech Recognition in the area of Telecommunications," in *IEEE Workshop on Automatic Speech Recognition and Understanding*, Santa Barbara, CA, 1997.

[11] D. Chazan, G. Cohen, R. Hoory and M. Zibulski, "Low Bit Rate Speech Compression for Playback in Speech Recognition Systems," in *European Signal Processing Conference*, Tampere, Finland, 2002.

