Ignitarium: Real-time Noise Suppression on Edge Devices Using a Practical AI-based Approach

Case Study

Company snapshot

Name: **Ignitarium**

Description: **Ignitarium is a product engineering design company with expertise in multimedia, DSP and deep learning. Combining DSP and deep learning, Ignitarium has been developing software to enable differentiating video and audio-based products.**

Website: https://ignitarium.com/audio-analytics/septra.php

Ignitarium is a member of the Arm AI partner program.

Goal

How many of us have struggled on work calls or when trying to use a virtual assistant because the kids are shouting, the neighbor's dog is barking, or someone is watching TV in the next room? Background noise has always disrupted clear communication and has certainly become one of the top working-from-home challenges in recent times.

Noise is generally assumed to be random and can cover a wide range on the frequency spectrum. If the noise does not fall in the same range as speech, a well-designed digital filter would be sufficient to reduce the background sound. However, it is different when the noise and speech overlap in frequency.

Challenges in Noise Suppression

Suppressing noise by band modification using moving average, inverse filtering, signal averaging, and noise cancellation by subtraction can alter the original signal. These algorithms work well in certain use cases, but they don't scale to the variety and variability of noises that exist in our day-to-day environment.

Traditional DSP algorithms (e.g. adaptive filters) can be effective when filtering stationary noises, which have very low time variant properties – for instance, the sound of an air conditioner running. Non-stationary noises have complicated patterns and have time variant properties. Signals from environmental sounds, such as a dog barking or a door closing, may be a very short burst, which complicates the task of removing those sounds. Some of the commonly used filters to remove noise are LMS (Least Mean Square), RLS (Recursive Least Square) and Kalman filter.

The main challenge of noise suppression techniques is to make it work well consistently for all kinds of noise. To address both stationary and non-stationary noises, designers must look beyond traditional methods.

Solution

Enter, Ignitarium – a product engineering design company based in India, which has implemented in a small-memory footprint a deep learning-based real-time noise suppression software that can run on low-cost micro-controllers.

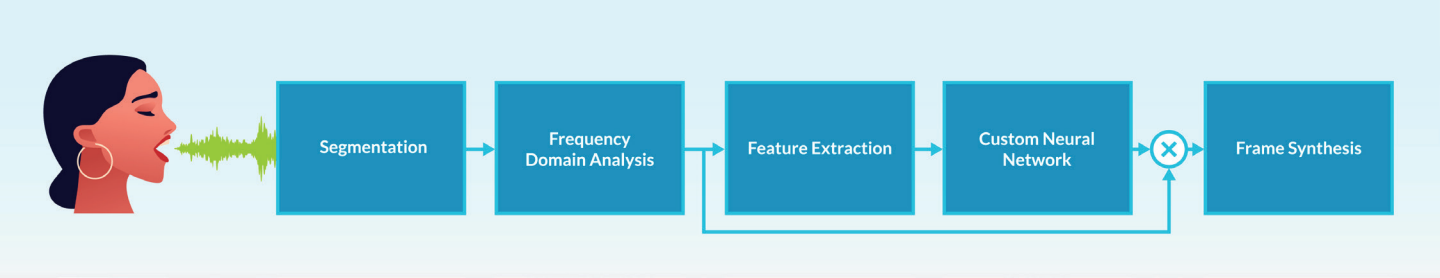
"Drawing on our strength in multimedia, DSP and the many years we spent on video AI, it was natural for us to see what we can do with AI in the field of audio analytics. We understand semiconductors and the MCU/SoC ecosystem very well. This opened up an interesting opportunity to implement real-time audio analytics on embedded devices, paving the way for the first engagement with a lead customer," said Sanjay Jayakumar, CEO of Ignitarium.

To counter stationary and non-stationary noises, a traditional DSP approach is not sufficient. This requires certain hybrid models including a deep learning-based approach. Using deep learning, it is possible to train a model with a variety of conditions and the resulting model can provide good results in these varying environments.

Ignitarium's Real-time Noise Suppression algorithm (IGN-RNS) is one such approach, where a linear regression method is used to denoise noisy speech input in real time.

Key features of IGN-RNS:

- Deep learning-based
- Small memory footprint (few 10s of KB of RAM) and low clock cycles (few 10s of MCPS)
- Low latency (<100 ms)
- Different sampling rates are supported (for e.g., 8KHz/16KHz/48KHz)
- Optimized for high-performance end-to-end voice pipelines



Ignitarium's Real-time Noise Suppression: Functional View

Processing

The input floating point samples are passed through a segmentation and frequency domain analysis process. This includes a high pass filter to avoid low frequency components. The filtered output is then fed to the feature extraction module.

Feature extraction

The noise suppression gain is computed based on 22 bands of the psychoacoustical Bark sound scale. Audio has a huge dynamic range, so we compute the log of the energy. The resulting data is a Cepstrum based on the Bark scale, which is closely related to the Mel-Frequency Cepstral Coefficients (MFCC) that are very commonly used in speech recognition.

The features extracted include:

- BFCC coefficients
- Derivatives of a few coefficients from BFCC
- Pitch related features
- A custom non-stationarity value that is useful for detecting speech

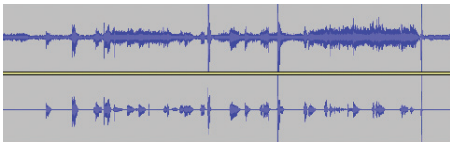


Fig.(a) Before and after noise suppression (time domain view)

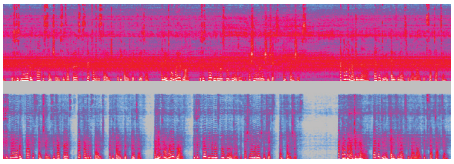


Fig.(b) Before and after noise suppression (spectrum view)

Neural network

The extracted features are fed to a recurrent neural network (RNN). Research and experimentations at Ignitarium labs have shown Gated Recurrent Unit (GRU) to be performing slightly better than LSTM on the same task and requires fewer resources (both CPU and memory for weights). RNN layers are important here because they make it possible to model time sequences instead of just considering input and output frames independently. This is especially important for noise suppression as we need time to get a good estimate of the noise.

Noise profiles

The standard pre-trained model supports the following profiles:

- Urban (motor, traffic)
- Home
- Machinery
- Office

The solution is not limited to the profiles mentioned above. The model can also be customized to a specific noise or a new noise profile can also be added. The NN based models are very good at what they are trained for, hence the quality of the training data is critical in achieving results. The suppression algorithm can be used for stationary as well as for non-stationary noises. As the correlation between the clean audio frequency components and noise frequency components decreases, the performance increases.

Given a new set of noises, a fully functional real-time noise suppression module customized for the new environment can be made ready very quickly.

NN-based implementation on edge devices

The main challenges for edge-level computation are the available processing power and memory. Ignitarium has engineered the right balance by figuring out a suitable NN architecture that can be implemented on microcontroller devices. Most of these devices have RAM under 30 KB. Ignitarium has implemented its solution to suit the limited processing power and memory in these devices.

Cortex-M processors are the computing platform of choice for developers, with Arm partners shipping more than 50 billion chips based on Cortex-M into a vast range of applications. As it is one of the most popular series of processors in this category, Ignitarium chose to implement its solution on the Arm Cortex-M4.

Target Platform	Performance Requirements (MCPS)	Memory Requirements		Target Architecture
		RAM (KB)	ROM (KB)	
LPCXpresso 54628	36	28	30	Arm Cortex - M4F Running at 220 MHz 3.4 CoreMark/MHz

Table 1: Memory and MCPS requirement to process 1 channel of 16KHz 16bit PCM audio

The table below represents noise suppression that can be achieved with the Ignitarium solution with various input signal to noise ratio values:

Input SNR (in dB)	% of Noise reduction
>40	>99%
20	>98%
10	>96%

Table 2: Noise suppression for different SNR levels

Ignitarium is using IGN-RNS (Ignitarium's Real-time Noise Suppression Engine) along with IGN-VCE (Ignitarium's Voice Command Engine) to implement local voice processing solutions on low-power edge devices. This includes smart headphones and equipment that can benefit from voice-activation, such as elevators, coffee vending machines, home appliances, commercial vehicles, medical devices etc. IGN-RNS models trained with noisy conditions remove noisy audio components in the audio spectrum, which produces a cleaner voice signal that IGN-VCE uses for more accurate predictions.

The Arm advantage

Ignitarium is enabling practical audio AI solutions such as noise suppression on low-cost edge devices and is working closely with the vast Arm AI ecosystem to identify scenarios where intelligent combination of deep learning and DSP brings the best of both worlds on low-powered microcontrollers deployable on various edge devices. Ignitarium chose the Arm Cortex-M platform because of its strong ecosystem, power efficiency and a mature software support for advanced math and deep-learning functions (for example, using CMSIS-NN).

Looking ahead

The increase in intelligence on edge devices has led to incredible innovation in the audio and video segments.

Speaking about what lies ahead for Ignitarium in this segment, Sujeeth Joseph, Chief Product Officer, "At Ignitarium, we have been chipping away at some complex video and audio solutions using deep learning as the foundation technology, while at the same time leveraging the best of traditional DSP techniques to enable highly optimized embedded implementations. Our focus on embedded analytics using AI continues across the audio spectrum, as we see traction from global players, who are looking to implement some very innovative products."