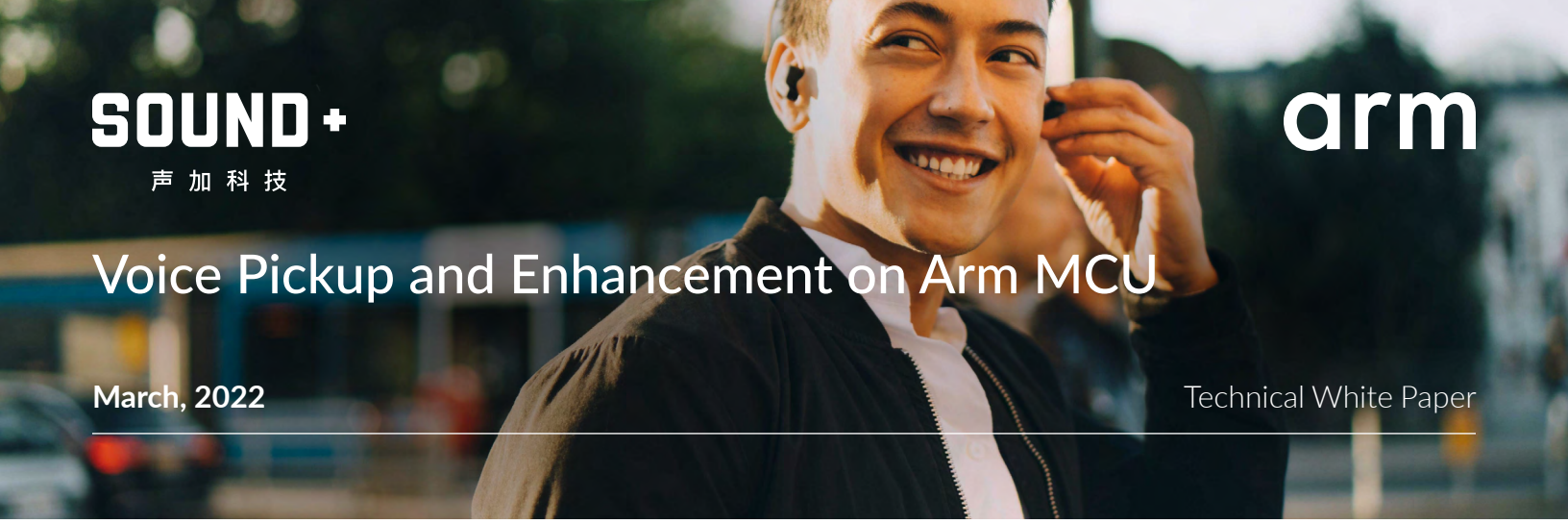


SOUND+
声 加 科 技

arm

Voice Pickup and Enhancement on Arm MCU

March, 2022

Technical White Paper

Company Snapshot

Name: **Beijing SoundPlus Technology Co., Ltd**

Description: **SoundPlus was established by a team of scholars who spent their career lives in the Institute of Acoustic, Chinese Academy of Science (IACAS), focusing on advancing the art of communication acoustic. SoundPlus provides business customers with total solutions to capture voice from near/mid/far field sources, under all perceivable working scenarios. The one stop provision we offer include IP licensing, ICs, PCB module and design service to speed up the integration process.**

Website: soundiot.cn

SoundPlus is a member of the Arm AI partner program.

Introduction

As earphone technology continues to improve, more and more people are enjoying wireless earbuds that are light to wear, free up hands to multitask while listening, and offer excellent true wireless stereo sound. Other new technologies, including active noise control to attend remote conferences, are also becoming increasingly popular.

Currently, most earphones are equipped with a microphone located at the back, at an angle. However, when a person speaks, the sound wave radiates to the front, which means that using the microphone on the earphones is not as effective as a handheld phone, which is much closer to the mouth. Besides this, the limited battery capacity and processor performance of the earphones all make it extremely challenging to provide users with comparable call quality.

AI-based Method Meets Microphone Array

To compensate for the amplitude attenuation caused by the forward movement of the speech signal, especially high-frequency components, clear tones, and so on, the two-microphones array is used to form a spatially directional beam that enhances the gain in a specific direction, while also eliminating interference noise in the other direction.

Considering the inconsistencies in the angle worn by each person, beamforming is an adaptive estimation process. This means that if the surrounding area is noisy, estimation errors likely increase. To avoid accidentally eliminating the user's voice, the residual dynamic noise incident from an approximate direction must be retained.

The advent of machine learning methods has broken through this limitation. Unlike the approach of using statistical signal analysis to distinguish between noise and speech, deep neural networks can use nonlinear modelling and good discriminative performance for dynamic noise interference. However, limited by computing platform resources, the stability of the neural network model itself may be reduced after network pruning and quantization, which may not be enough to cover all usage scenarios.

Combining the adaptive beamforming and machine learning method can make the solution more adaptable to a noisy environment by achieving higher noise reduction. Meanwhile, because beamforming can spatially filter out surrounding interference, the signal-to-noise ratio (SNR) of the neural network input signal is improved. Thus, stability is improved.

See the comparison below:

Diagram 1: Traditional 2-mic array beamforming speech enhancement based on spatial filtering and spectral estimation.

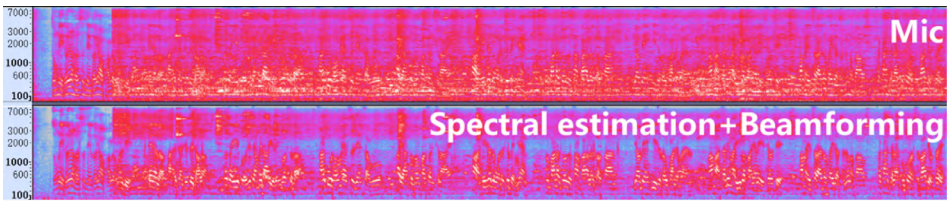
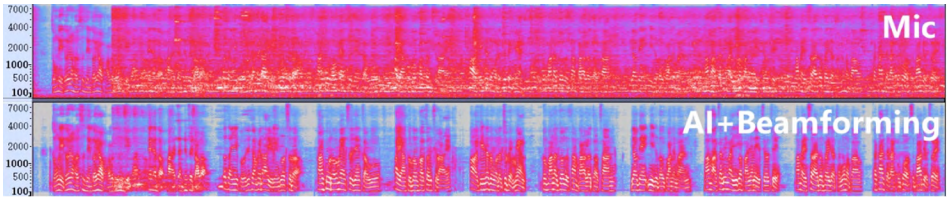


Diagram 2 shows significant differences with diagram 1. By using deep learning methods, we can more accurately identify speech components from noise components and complement high-frequency harmonics to make speech sound fuller and clearer.

Diagram 2: AI augmented 2-mic solution.



Challenges in Deploying AI Speech Enhancement at the Edge

There is a strong requirement for AI voice enhancement among earphone manufacturers. However, several challenges must be overcome before implementing AI voice enhancement on the tiny true wireless stereo earphone.

First, AI algorithms require rather complex calculations, often involving a large number of matrix multiplication operations. Maximizing the efficiency of the supported operators is a key consideration during the architecture design phase.

The second challenge is how to preserve the noise reduction performance and speech quality of the algorithm, while quantizing and pruning the model to make the neural network model fit in the limited RAM.

Thirdly, costs can be high. Some products use a dedicated DSP chip to perform AI-speech enhancement processing. This approach can provide sufficient performance to process user speech in real-time, however, it is costly and takes up valuable space on the earphone due to requiring additional processors. In addition, the introduction of a separate DSP increases the complexity of the system, further driving up the cost of use.

Fourth, deploying an AI speech enhancement solution in runtime on a Bluetooth SoC is much more complex than running the process offline. The system integration requires a high level of effort and commitment effort, as well as close cooperation between algorithm vendor and chip manufacturer.

Lastly, power management on terminal devices, including true wireless stereo earphones is very strict. Adding or enhancing any function must meet power consumption requirements and not significantly affect battery life.

Solution: Arm and SoundPlus Achieve AI Voice Enhancement At The Edge

SoundPlus and Arm have partnered to optimize the SoundPlus AI speech enhancement algorithm (SVE-AI) to run on Arm's Cortex-M4F and Cortex-M33 MCU, achieving the balance between power consumption and market-leading performance.

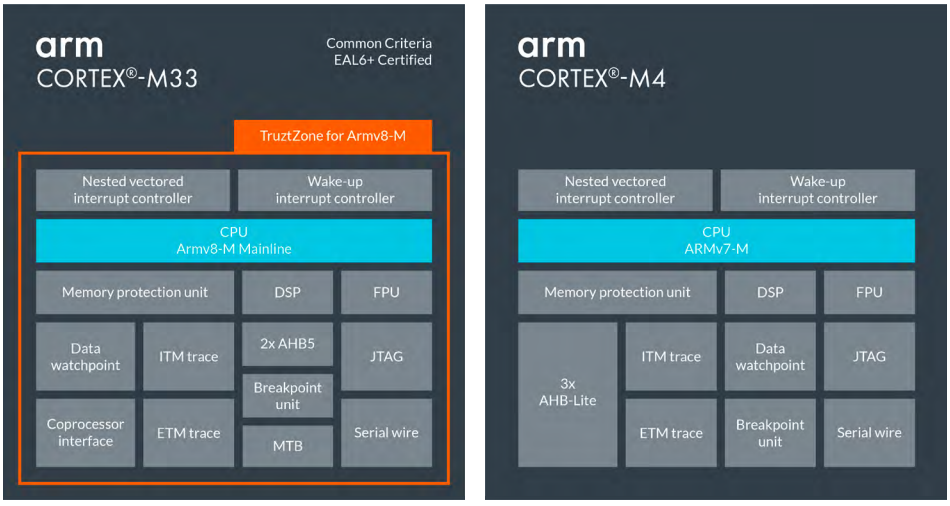
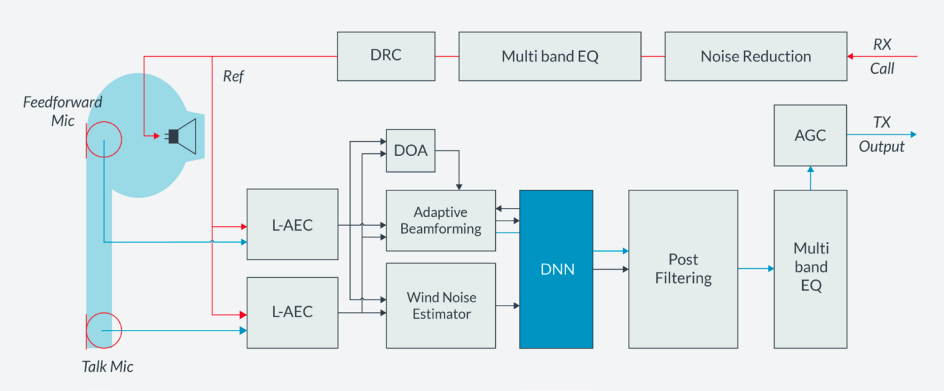


Diagram 3: Arm Cortex-M33 and Arm Cortex-M4.

SoundPlus leveraged the instruction set provided by Arm and added algorithmic features to provide complete compatibility at the IP level. Even if the hardware is changed, the SVE-AI solution provided by SoundPlus can be deployed quickly, as long as the Arm IP and instruction set is used. This helps reduce overall development time and effort and accelerate time to market.

Diagram 4: Block chart of SVE-AI dual mic solution.



As diagram 4 illustrates, the DNN module is embedded in the front-end solution and plays a crucial role in the processing flow, enhancing the performance of each algorithm module:

- It improves the accuracy of estimating the arrival direction and then guides the beamforming.
- It takes the process result of the beamforming and wind noise estimation as input and then separates clear voice components and directional information.
- The DNN module can also indicate how much noise reduction post-filtering needs to achieve.

In this collaboration, we first ensured that all the computation happened on the MCU to keep the system lean.

Secondly, the AI speech enhancement model is run on the MCU in real time. This step includes model building and tuning, inference optimization, data compression (int8) and data-stream management.

Thirdly, by fixing the arithmetic and processing flow on the MCU, SoundPlus Technology provides a range of speech enhancement models with different parameter sizes, leveraging the resources provided by different configurations of the chip.

The SVE-AI solution has been adopted in true wireless stereo earphone products by mainstream mobile brand manufacturers and international audio brands, including earphones equipped with up to four microphones. The AI speech-enhancement model effectively improves the speech quality of true wireless stereo headphones in dynamic interference environments, achieving excellent S-MOS scores in the objective test and receiving positive feedback in subjective tests.

As a result, the ARM MCU on true wireless earphones enables dynamic noise suppression and high-quality speech enhancement at a 16KHz sampling rate in real time.

Looking Ahead

SoundPlus is focused on expanding the leadership in voice enhancement solutions for wearables by consistently innovating and upgrading performance. With the launch of Arm's most AI-capable Cortex-M processor, the Cortex- M55 and a new class of machine learning processor, the Ethos-U55 microNPU, SoundPlus has been working with Arm to explore the benefits and performance uplift of implementing a far larger scale neural network model at the edge.

We're optimistic that the collaboration between Arm and SoundPlus Technology will help define chipset specifications for wearable devices, wireless earphones, and mobile SoCs well into the future.

Learn more on:

- Arm AI Solutions
- Arm AI Technology
- Join an AI Virtual Tech Talk

arm