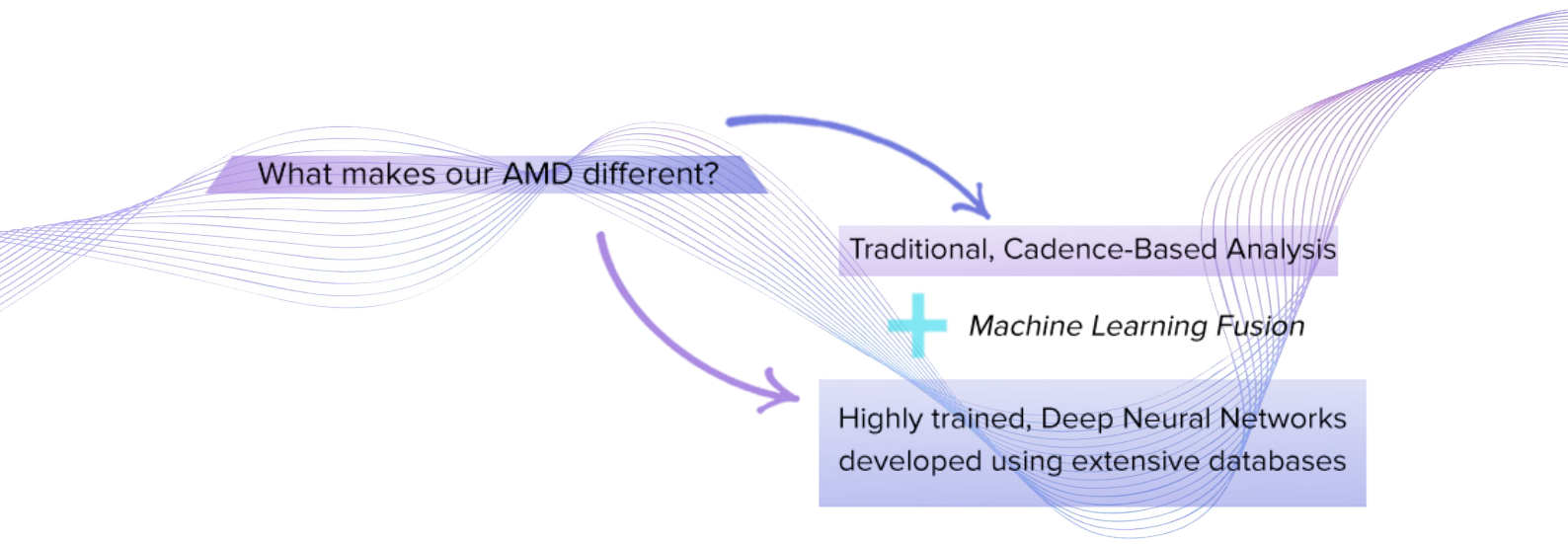
Aculab AI-Driven Answering Machine Detection

AI-Driven Answering Machine Detection

Introduction

Aculab's new Answering Machine Detection (AMD) algorithm, also known as Live Speaker Detection (LSD), exploits the company's expertise in both traditional 'cadence-based' AMD and the latest Artificial Intelligence (AI) technologies, as already used in VoiSentry, its speaker identification and verification system.



It's this combination which ensures that Aculab's AMD embodies the key benefits of both approaches.

Aculab's AI Development

Artificial Intelligence (AI) in the form of Deep Neural Network (DNN) embeddings allows the system to capture pertinent information regarding the speech, together with any other signal components, but without explicitly separating them. **This provides simultaneous robustness and discriminative power.**

The combination of these disparate forms of data is itself performed using a Machine Learning (ML) process, and thus benefits from automatic data-driven fine-tuning, independent of any human bias or preconceptions. To date, **the learning process of the whole system has made use of hundreds of thousands of calls.** Now, as development proceeds, data already collected from a variety of locations around the world will be incorporated to ensure the highest performance is maintained in an international context.

Testing of Aculab's AMD system has shown that it is **able to discriminate between live speakers and answering machines with an accuracy in excess of 99%** (using calls within the USA, partitioned into distinct training and test sets). What's more, the result is produced a mere 1.5 seconds after the far-end speech has been detected.

Aculab's AMD has algorithmically learned subtle cues that may not be apparent to the human ear, and balances this against an immunity to factors that are unimportant. For example, background noise that - without this hybrid approach - could result in misclassifications.

How does it work?

The Nature of the Problem

The figures here show examples of typical far-end signals that may be presented to AMD systems. *Fig. 1* shows a recording of a live speaker answering the phone with a typical (short) salutation, while *Fig. 2* is more typical of an answering machine; a much longer message with no significant pauses.

A human listening to these audio files could categorise a live speaker or an answering machine, but how do we create a program do this?

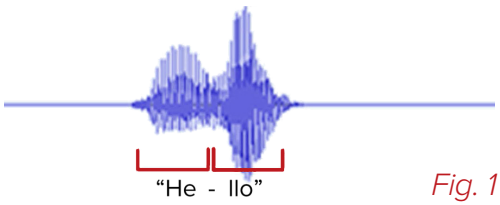


Fig. 1

Fig. 2

Traditional (Cadence-Based) AMD

The simplest method to classify such signals is to use the 'cadence' of the speech (the variations in the intensity of the speech over time - Fig. 3). A basic 'cadence-based' AMD could classify them merely by detecting the duration of any detected speech. If the first long pause starts very soon after speech is first detected, the call will be classified as a live speaker; otherwise as a machine.

In reality, more sophisticated analysis is needed to cope with the signal in *Fig. 4*. This is from a live speaker who follows up their initial salutation with a brief pause and a further greeting. Although less common, this kind of 'live speaker' response is certainly not rare.

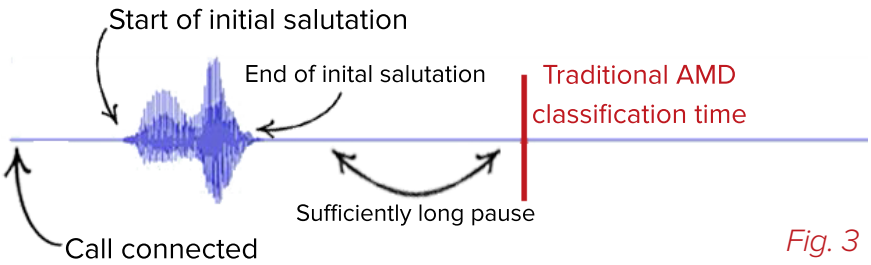


Fig. 3

Fig. 4

Traditional AMDs have difficulty with signals such as *Fig 4*. Such an AMD may classify this as a machine, or a human, but the decision is critically dependent on the duration of the pause. If it's short, the system will consider the whole greeting and (wrongly) assume that it is a recorded message. However if the pause is just slightly longer, the typical AMD will (still wrongly) assume the response has ended. This time, because the speech was short, it will end up being (correctly) classified as human, even though the reasoning behind that conclusion was flawed. This demonstrates the fragility of cadence-based systems.

Typical AMD systems such as these can correctly identify around 94% of real-world calls within an acceptably short time. The remaining 6% of errors are due to a wide range of factors. These include live speakers announcing themselves with formal messages (as in *Fig. 4* above), and recorded greetings being unusually short or containing lengthy pauses.

Aculab's AI-Driven AMD

Aculab's new AMD uses a hybrid formed from novel speech analysis algorithms and a state-of-the-art deep neural network. **When analysing a call, the system only uses up to 0.5 seconds of audio before the start of the initial salutation, and only up to 1.5 seconds after it.** *(Fig. 5)*.

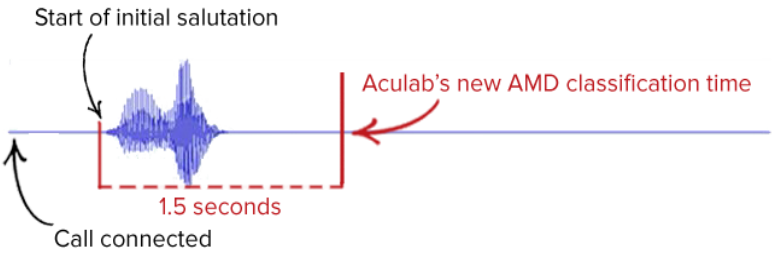


Fig. 5

It achieves better results, and classifies more quickly than other AMD systems in the market. This is because its neural network takes advantage of a considerably more complete characterisation of the far-end audio. The AMD analyses the fine details of background noises and signal distortions, as well as the speech itself, using thousands of statistical measurements.

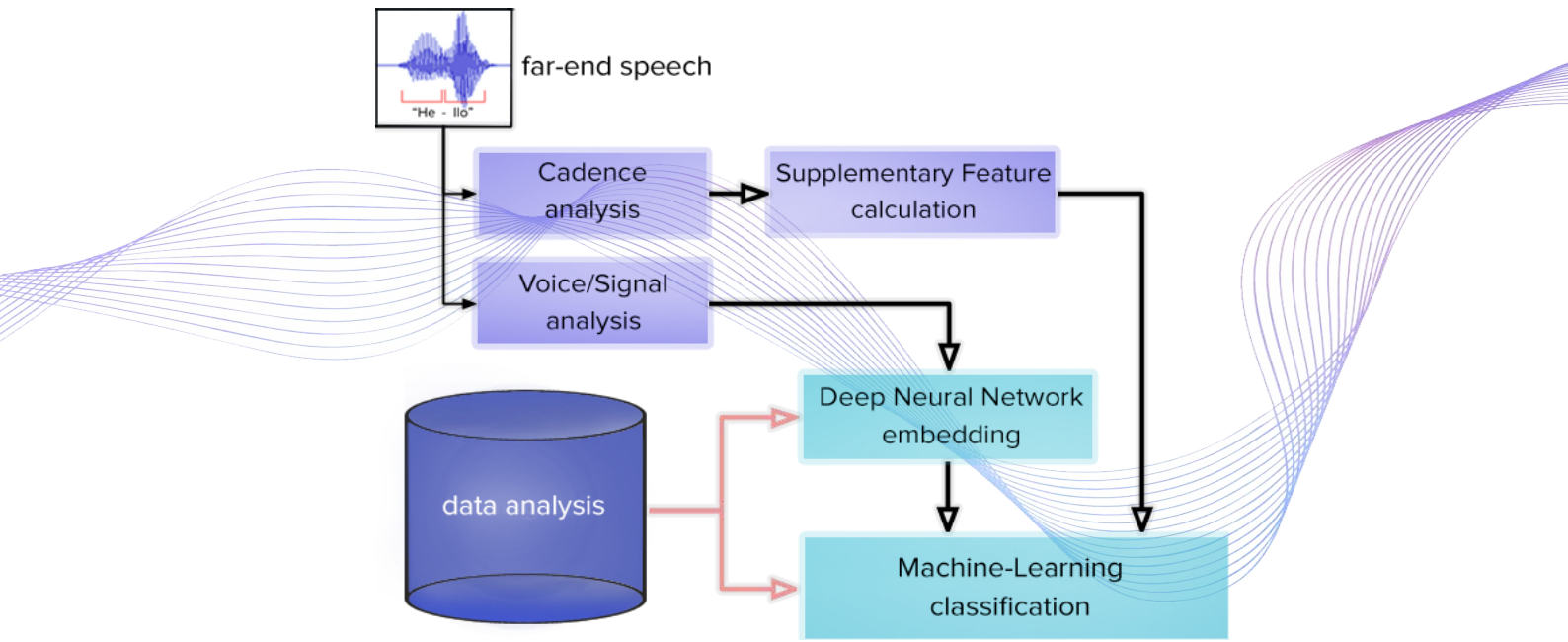
The Benefits of a Hybrid System

Dealing with such a wide range of complex and atypical responses clearly requires some intelligence, and thats where AI comes to the fore. To ensure optimal performance, Aculab's AMD utilises extensive in-house AI expertise and resources, including recordings chosen to be:

- Accurately labeled
- Representative of the whole gamut of signals likely to be observed in the environment where the AMD will be deployed
- Plentiful, with over 200,000 call recording samples

If the decision-making and classification process was completely devolved to the Machine Learning algorithm, it would prevent us from taking advantage of any of the beneficial aspects of human experience, intelligence and intuition.

Accordingly, Aculab has enhanced and extended its earlier AMD characteristic parameters, and incorporated them into a hybrid AI system with the results of the DNN algorithm. This takes the achievable accuracy to in excess of 99%.



Regulatory Compliance

Regulators around the World are attempting to reduce the scourge of silent calls, thus improving the called party's experience. Different regulators are working to achieve this in different ways. For example, some specify the maximum length of time it takes for the far-end to hear a response to their initial salutation, and set an upper limit on the percentage of calls perceived as silent. In the UK, Ofcom specifies these figures as 2 seconds from start of initial salutation and 3% of calls, respectively. These regulations are generally becoming tighter.

Silent calls happen for a number of reasons, and a failure to correctly distinguish live humans from answering machines is one of the most common.

Thus, compliance with statutory regulations requires an AMD which classifies both quickly and accurately. Looking again at the UK example, Aculab's new AMD's 1.5 second classification time, coupled with its high accuracy, allows these requirements to be not just met, but comfortably exceeded.

Conclusion

Aculab's new AMD uses a hybrid approach incorporating established human expertise within a cutting-edge, bespoke AI framework. This allows it to classify only 1.5 seconds after detection of far-end speech, and to do so with an accuracy in excess of 99%.

This best-in-class performance enables companies to achieve compliance with the limits set by various regulators around the world, reducing the percentage of silent calls and improving the far-end user experience.

To find out how you can integrate Aculab's new AMD within your system, as a standalone product, or as part of our highly customisable CPaaS, Aculab Cloud, contact us with the details below.

About Aculab

Voice-Tech Specialists

Aculab is an innovative company that offers deployment proven technology for any telecoms related application. With over 43 years in the telecoms industry, and with a specialism in Voice-related equipment, Aculab's enabling technology serves the evolving needs of automated and interactive systems, whether on-premise, data centre hosted, or cloud-based.

Over 1000 customers in more than 80 countries worldwide, including developers, integrators, and solutions and service providers, have adopted Aculab's technology for a wide variety of business critical services and solutions.

Aculab offers bespoke development APIs for voice, data, fax and SMS, on hardware, software and cloud- based platforms, giving a choice between capital investment and cost-effective, 'pay as you go' alternatives.

For more information

To learn more about Aculab Cloud and Aculab's extensive telephony solutions visit:

www.aculab.com

Contact us

Phone

+44 (0) 1908 273800 (UK)
+1 (781) 352 3550 (USA)

Email

info@aculab.com
sales@aculab.com
support@aculab.com

Social

@aculab

aculab