

Ultra-low-power auditory processing in real-time



CC-BY-SA 2018 aiCTX AG
Weinbergstrasse 161, 8006 Zürich, Switzerland

Consumers expect **high levels of responsiveness** *from their devices. Our* **ultra-low-power neuromorphic auditory processing technology** *enables always-on speech and keyword detection, with very low latencies.*

Provide real-time responsiveness for your devices, by using **intelligent wake-up and command signals** *provided by our neuromorphic processors.*

Ultra-low-power means **long battery life***, without compromising on device intelligence.*



Contents

Real-time auditory processing using ultra-low-power Neuromorphic devices 7

Application: Real-time detection of spoken vowels 15

Technology: Neuromorphic Engineering 17

Technology: "Reservoir computing" 21

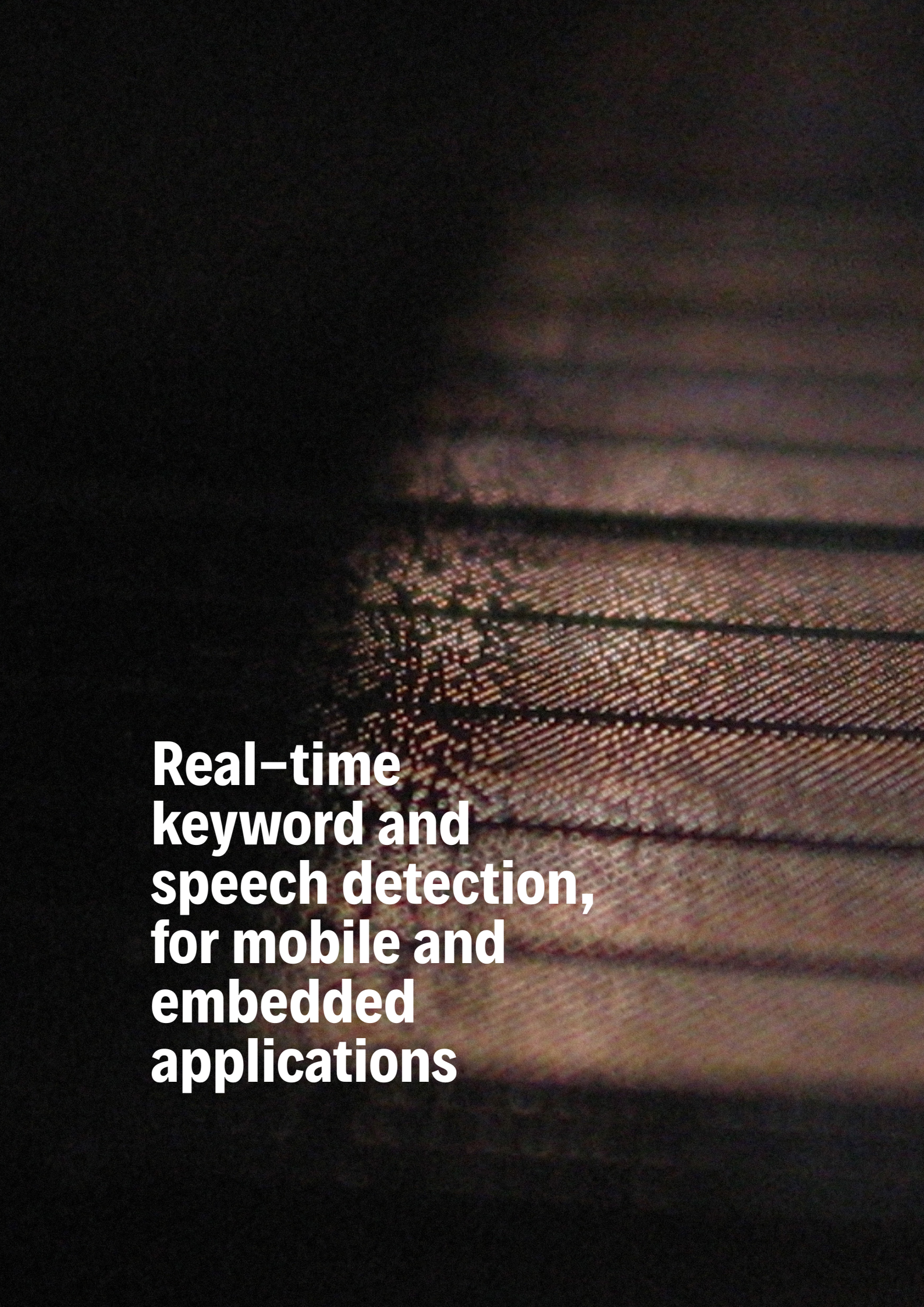
About aiCTX 23

aiCTX technology 24

Key personnel 27

Forward-looking statements disclaimer 30

Image credits 31



Real-time keyword and speech detection, for mobile and embedded applications

Real-time auditory processing using ultra-low-power Neuromorphic devices

aiCTX is building solutions for auditory signal processing based on our neuromorphic processors. We design systems **tuned for human speech**, that operate in **real time with low latencies, and at ultra low power**.

We will provide modules that accept **direct input from analog and digital microphones** (PDM). Simultaneous **multi-channel processing** is supported natively.

Systems built around our solution will provide **always-on, low-latency responsiveness to human speech**. Designed for keyword detection, speaker recognition, and human speech detection, our modules provide **smart wake-up signals** for power-on or to request further processing by higher-powered subsystems.

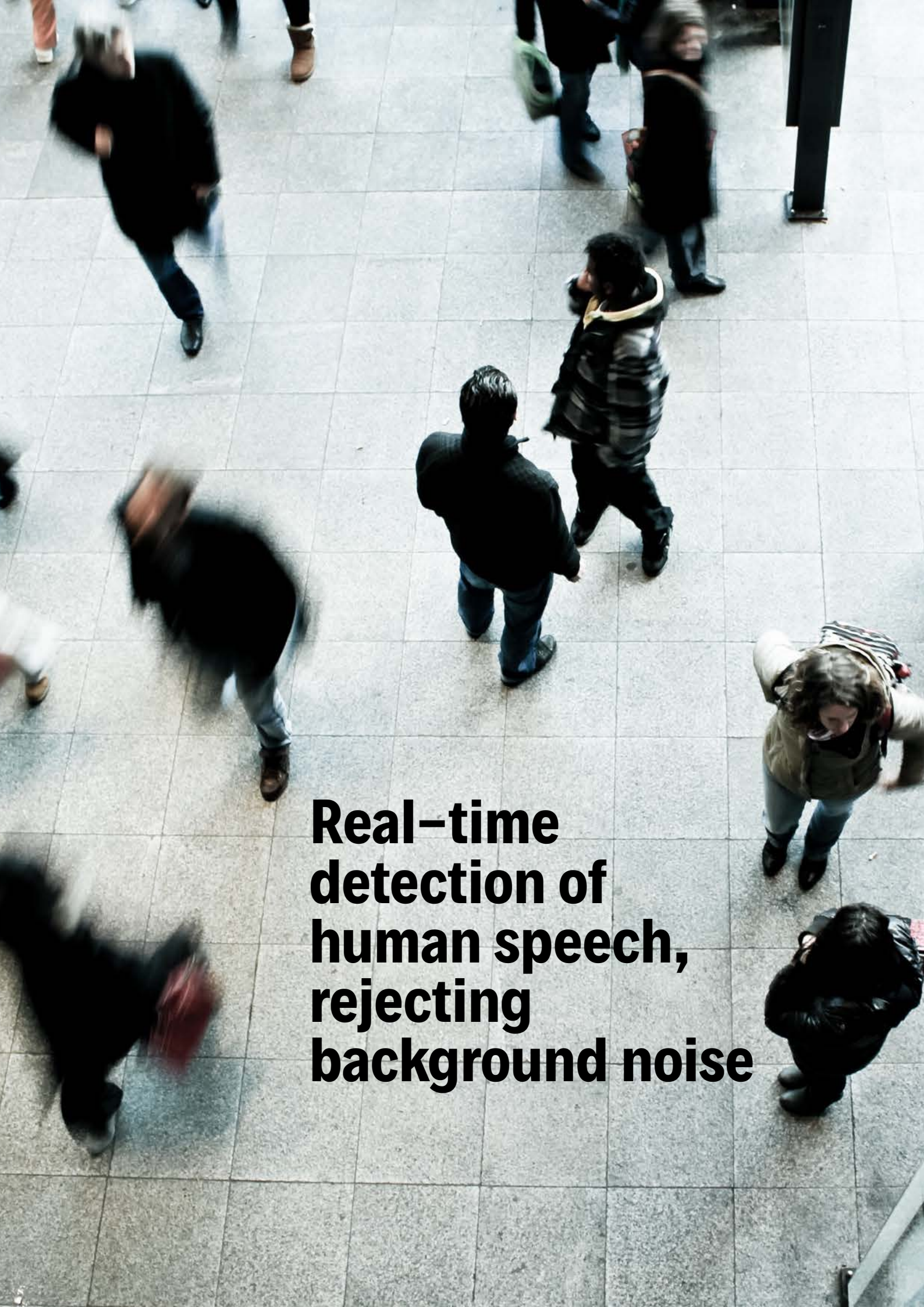
Human speech detection

Build systems that are sensitive to human speech. Powered by our neuromorphic processors, we enable **real-time responsiveness to human speech, at ultra low power**.

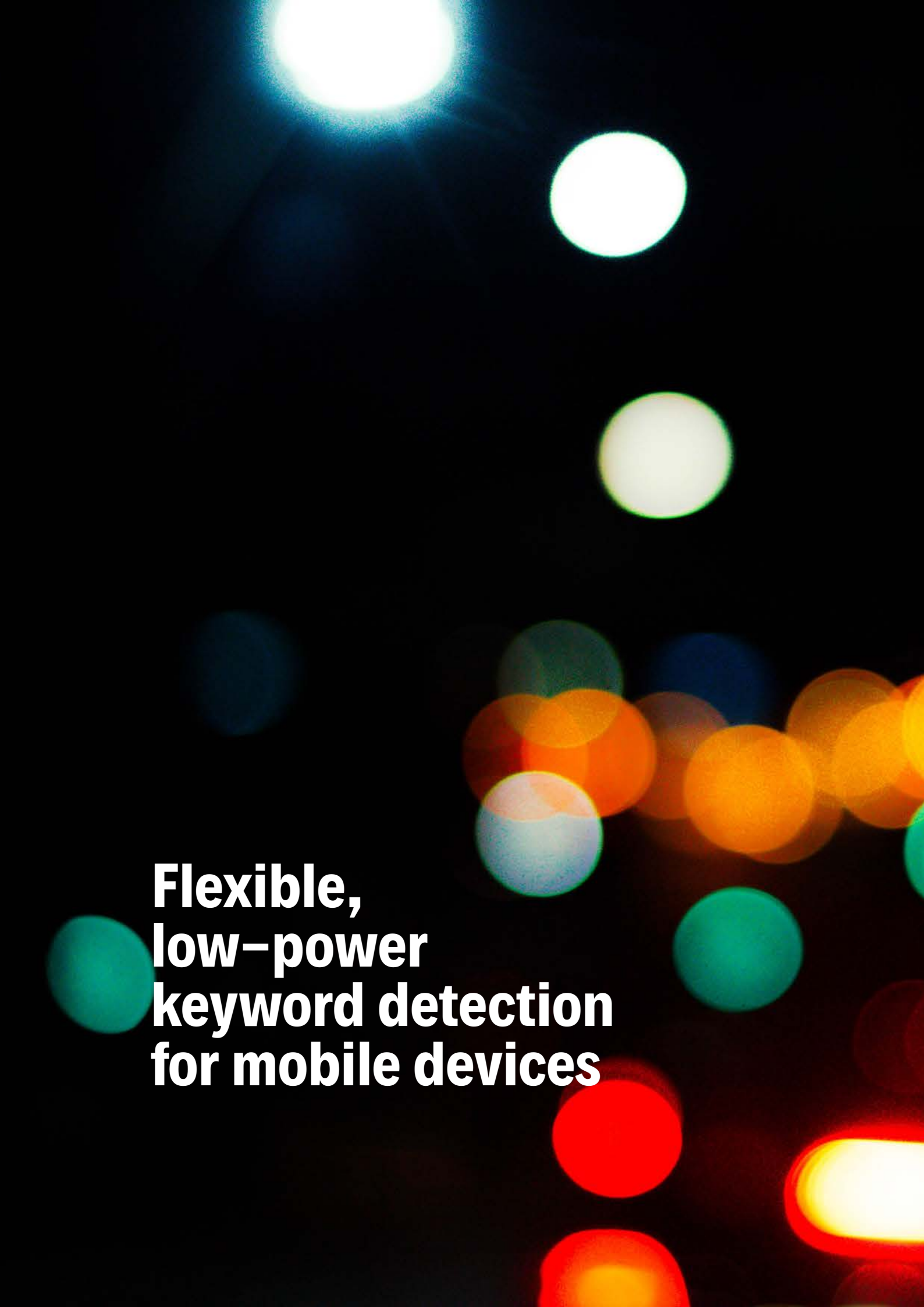
Trained to **reject background noise**, we deliver a smart wake-up signal to power up speech processing hardware, or to send recordings to the cloud for further analysis.

Speaker recognition

Our neuromorphic solution for auditory processing supports sensitivity to speaker identity. Design low power systems that recognise a desired speaker, using always-on neuromorphic technology.



Real-time detection of human speech, rejecting background noise



Flexible, low–power keyword detection for mobile devices

Keyword detection

Build mobile and embedded devices that **respond to spoken commands, without compromising on battery life**. Our systems operate at utra-low-power in active mode, enabling **always-on keyword detection**.

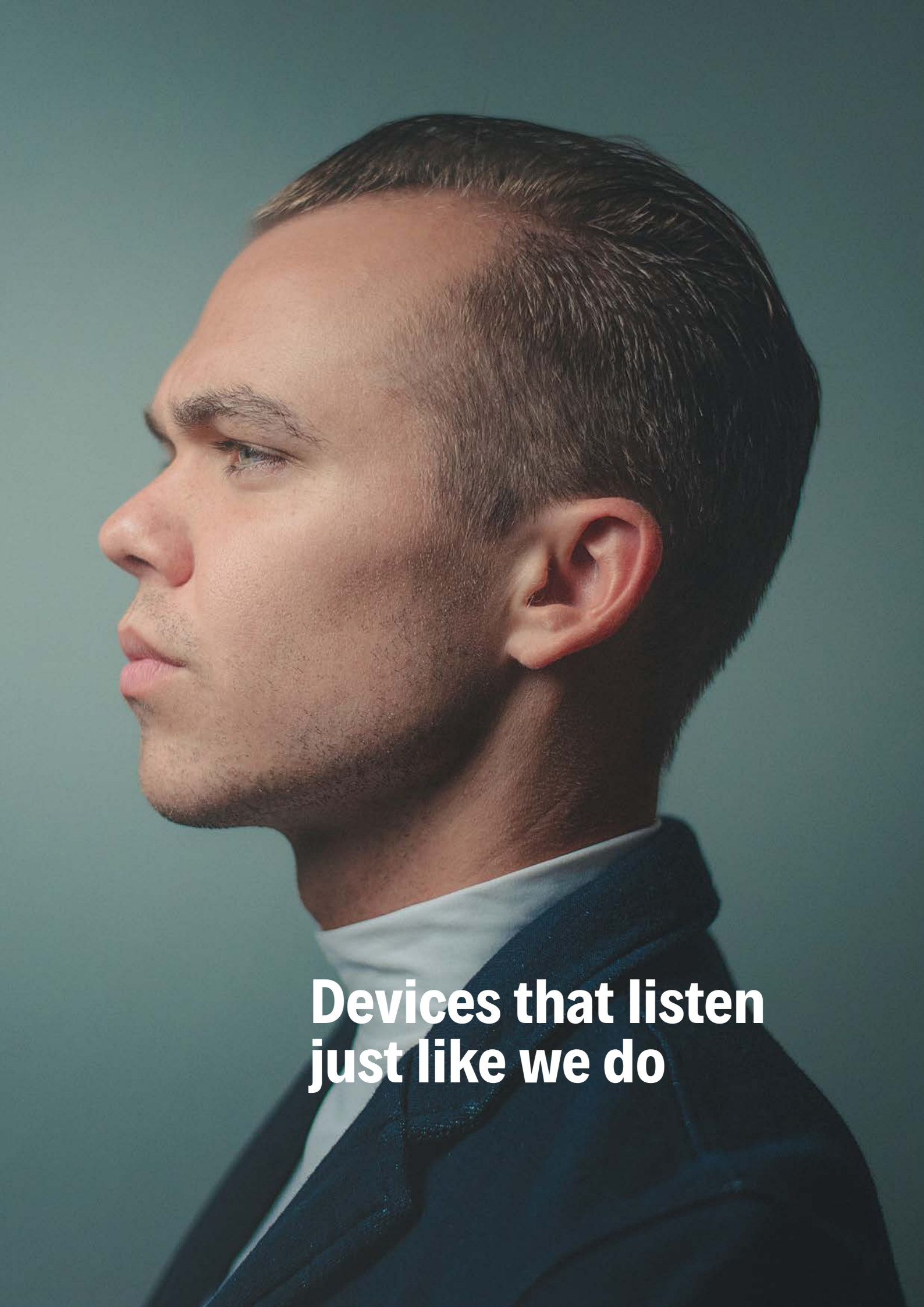
Our solutions are **highly customisable**. You can provide completely pre-trained keyword detection, or permit further customisation by the end user.

Design mobile systems with built in "wake" commands, or recognise a set of defined keywords for flexible spoken commands.

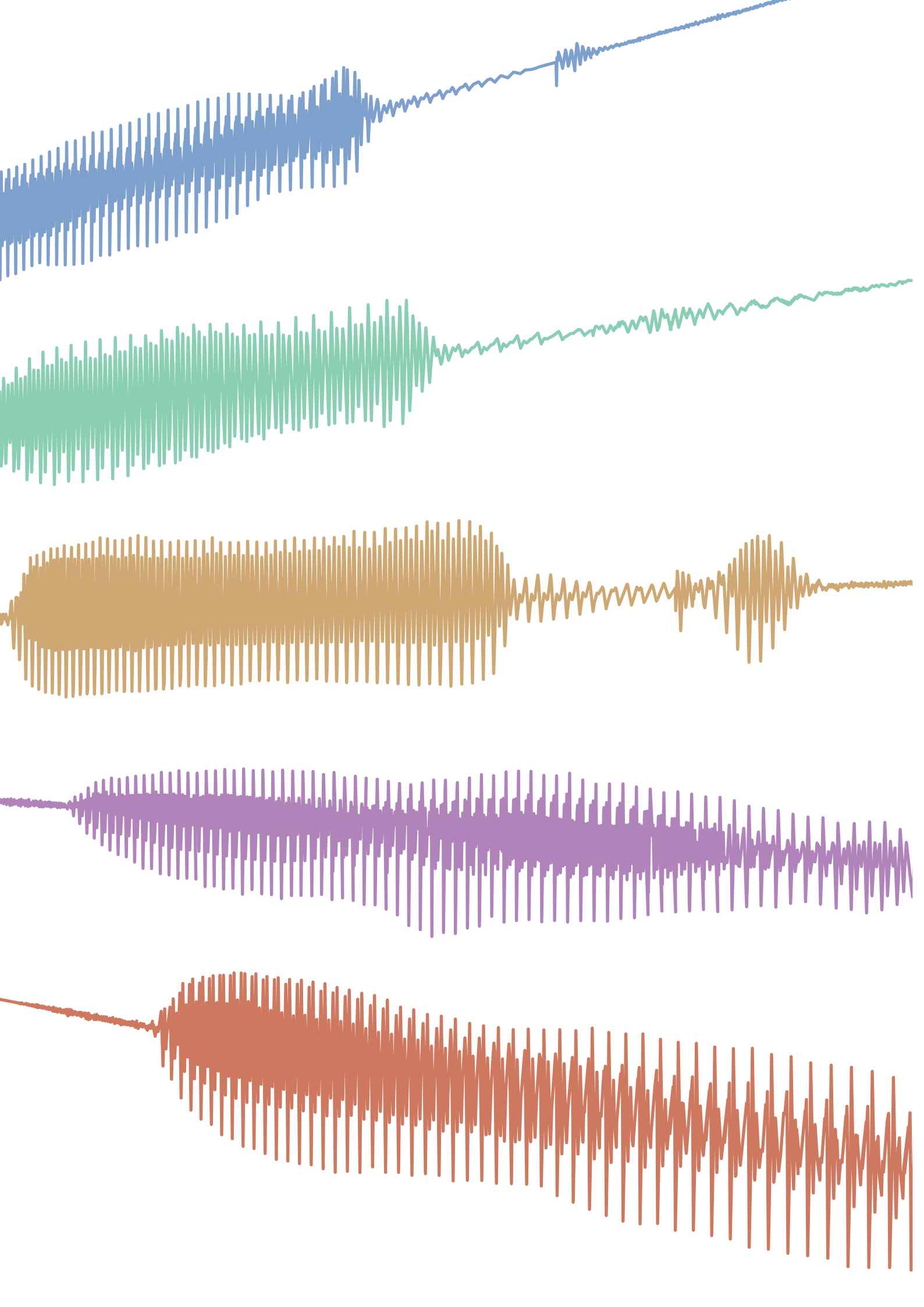
Listening to speech like the human ear

The human cochlea performs a biological Fourier transform of incoming sound waves, by separating out frequency components of sound. Specialised nerve cells transduce the vibrations of the basilar membrane into voltage impulses called spikes. The result is to **encode sounds as multiple parallel streams of events**, each carrying information about a small set of input frequencies.

Our approach is similar — a series of tuned filters span the frequency range of human speech, in the same way as the cochlea. The energy in these frequency bands is converted to trains of events, for further processing by our neuromorphic devices.

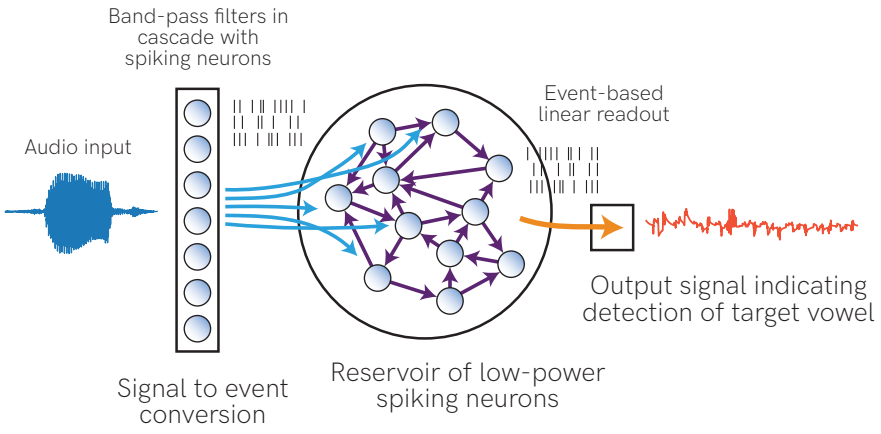


Devices that listen just like we do

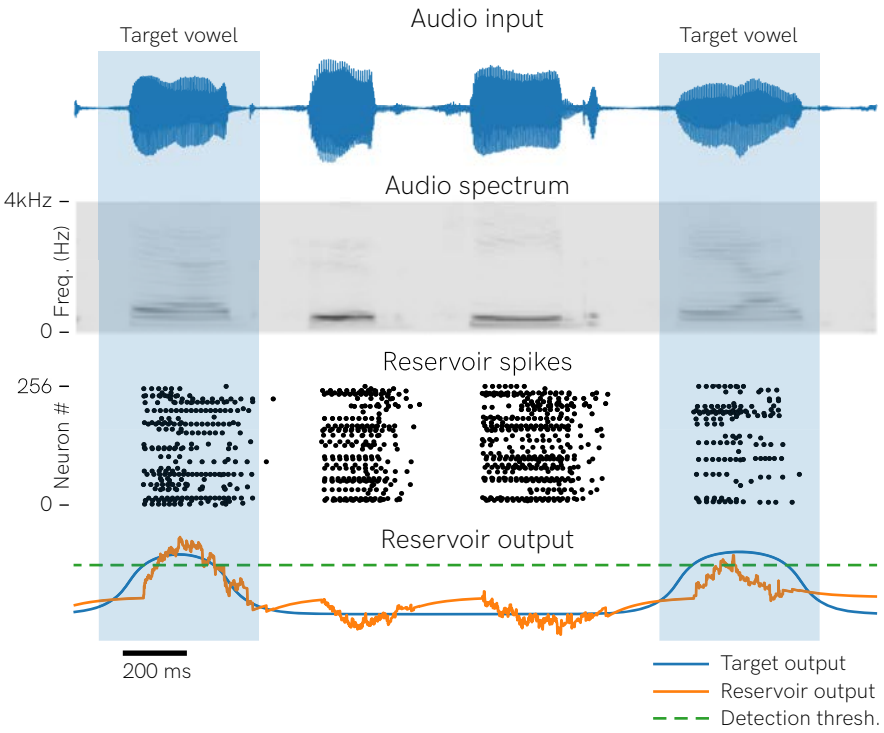


Application: Real-time detection of spoken vowels

aiCTX has developed a proof-of-concept for ultra-low-power analysis of English vowels, using our neuromorphic architectures. Analog audio signals are processed by a network of spiking neurons implemented in custom silicon. One specific vowel is identified in real time. Our system raises a trigger signal when the vowel is detected.



The architecture of our proof-of-concept vowel detection. The audio signal is passed through a bandpass filter bank and single neurons in order to be transformed into spikes. Then the spikes are processed by a reservoir of 256 spiking neurons. Detection is achieved by training outputs to detect a particular vowel.



Vowel detection using auditory-to-spike encoding and reservoir computing, in a spiking network. Target vowels are successfully detected; non-target vowels are ignored.

Implementation

Audio signals are processed through a 256-channel filter bank, where each channel connects to a spiking neuron. If a signal contains sufficient energy in a particular frequency band then the corresponding neuron fires. The spiking events are sent as input to a reservoir of a 256 spiking neurons. Variations in the audio signal are analysed by the network, which collectively produces an output signal. We train the output signal to identify the presence of a vowel by crossing a threshold, and raise a trigger-like event when the target vowel is present in the input signal.



Technology: Neuromorphic Engineering

Neuromorphic means literally "like the brain". aiCTX is developing new classes of computing devices that make use of the **incredible power efficiency and real-time capability** of neural tissue.

We design custom chips that use the physical properties of silicon to efficiently emulate the behaviour of nerve cells. Our neurons receive and transmit signals using brief voltage pulses, **the same communication technique used in mammalian brains**. Silicon synapses (connections between neurons) mimic the chemical transmission of information between neurons in live tissue.

We connect our emulated neurons into communicating networks, that are configured to **process information in real time**.

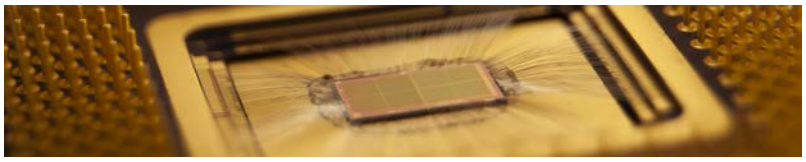
Benefits

Because our neurons are emulated using the physical properties of silicon, they require very little power to operate. Our Dynap-se chips can operate for **several months on a single AA battery**.

We distribute the storage of memory alongside the silicon neurons, with the result that comparatively less data must be shuffled through our chips, compared with a classical CPU or GPU-based implementation. This further reduces the energy requirements of our hardware.

Our hardware communicates through events, in real time. Information is processed immediately when it arrives, using the timing of an input event to directly represent the time at which the event occurs. This provides three benefits:

- When no input is present, no energy-wasting computation occurs
- Processing of information is not delayed. In traditional frame-based image processing, in comparison, events can only be processed once an entire frame has been acquired.
- Our solutions process data in real time — no slower, no faster. As a result, our hardware is designed to interface directly with the real world.



Differences from CPU/GPU simulation

Networks of neurons can be simulated using traditional von Neumann processors — the same CPUs that are in every computer and every

phone. Standard processors are of course highly flexible, general-purpose computing devices, but that flexibility makes them less efficient in simulating networks of neurons. This is especially true when one considers the massive parallelism in networks of neurons; CPUs are available with multiple cores, and usually implement some small-scale parallelism, but for large networks of neurons CPUs must sequentially simulate portions of the network in turn.

“ Our neuromorphic solutions are optimal for mobile and autonomous devices, where energy usage must be kept as low as possible ”

The use of GPUs for simulating neural networks has in recent years led to the commercialisation of deep learning — many-layered neural networks can now be trained and simulated in reasonable times. However, the energy usage required to train these architectures is increasing exponentially (**openai.com**). The massively parallel architecture of GPUs and TPUs (Tensor Processing Units) makes them well matched for simulating feedforward neural networks. Nevertheless, GPU architectures are designed for general-purpose computation, and so cannot reach the efficiency of special-purpose neuomorphic hardware for simulating neurons.

The simulation rate for CPU/GPU solutions usually scales with network complexity, with the result that these architectures are difficult to match to real-time processing. CPU/GPU solutions are a good match for applications where small energy budgets and real-time operation are not required. Our neuromorphic solutions are optimal for mobile and autonomous devices, where energy usage must be kept as low as possible. Our technologies are also ideally suited to real-time signal processing, especially on data that is fundamentally event-based.



Technology: “Reservoir computing”

Reservoir computing is a method for using populations of simulated nerve cells (neurons), to extract desired information from a signal. In our case, the signals are either very simple (e.g. a recording from a single vibration sensor) or slightly more complex (simultaneous recordings from two or more ECG electrodes). Our approach is designed for these spatially “low-dimensional” inputs, rather than highly complex signals such as a video stream — we use other network architectures for these “high-dimensional” data.

A signal is passed to a population of several hundred simulated neurons (the reservoir), which are connected to each other in a complex pattern.

“The input signal passes back and forward between neurons, combining with itself and with the signals already moving through the reservoir”

The input signal passes back and forward between neurons, combining with itself and with the signals already moving through the reservoir. Signals remain in the reservoir for some time, as the input changes, meaning that different temporal components of signals can be examined simultaneously.

The complex patterns of activity that are produced within the reservoir can then be examined and compared to a desired output signal, during the training process. For example, we can inject ECG input signals corresponding to pathological heartbeat conditions, and train the reservoir to recog-

nise the difference between a normal heartbeat, and a pathological heartbeat.

We can also inject vibration signals into the reservoir, corresponding to normal, pre-failure and failure conditions of a mechanical device, and train the reservoir to identify when a component is not operating normally.

Why "reservoir"?

The poetic name for our approach arises in analogy to a pebble tossed into a lake (or reservoir). The pebble causes ripples on the lake, which interact with the complex lake shore by reflecting in complicated ways. By examining the ripples on the lake surface, one can tell when and where the pebble was thrown. This information about the pebble persists for some time, and can interact with ripples from other pebbles tossed later into the same lake.

In this analogy, the pebble is the input signal, the lake is the population of simulated neurons, and the complex lake shore corresponds to the patterns of connections between neurons. The ripples on the lake surface represent the information about the input signal propagating within the population of simulated neurons.

Further reading

- Wikipedia, "Reservoir Computing"
- Maass, Natschläger, and Markram, 2002. "Real-Time Computing Without Stable States." Neural Computation 14 (11) pp 2531–2560. DOI:10.1162/089976602760407955



About aiCTX

aiCTX is a commercial spin-off from the Institute of Neuroinformatics, University of Zurich and ETH Zurich. We specialize in developing mixed-signal neuromorphic silicon hardware for neural simulation and signal processing; develop software for interfacing with and configuring neuromorphic hardware; and develop solutions to analyse and process bio-signals.

“ aiCTX specializes in developing low-power neuromorphic silicon hardware ”

aiCTX is developing approaches for bio- and industrial-signal processing, based on biologically-inspired neuronal architectures emulated on low-power neuromorphic hardware. The neuromorphic industry is growing rapidly, with significant contributions from large players such as Intel and IBM. Our advantages are our long experience in designing low-power neuromorphic circuits, and our experience in theory of neuromorphic and cortical computation.

aiCTX is also developing solutions for rapid event-driven processing of visual information, for application in industrial monitoring and automotive markets. We are building the first commercially available full-stack neuromorphic hardware solutions.

aiCTX technology

DynapSE2

Our new neuromorphic chip combines the **impeccable energy efficiency** of its previous generations with new features for **low latency, real-time end-to-end applications** for general purpose recurrent and reservoir networks. Its integrated biosignal amplifiers and the direct dynamic vision sensor bus make the DynapSE2 a **perfect fit for mobile health and robotic applications**. Each chip features 1k redesigned adaptive exponential integrate-and-fire analog ultra low-power spiking neurons and 65k enhanced synapses with configurable delay, weight and short term plasticity. The DynapSE2 enables **advanced biological processing schemas**, such as ampa, nmda and gaba channels, as well as conductance based dendritic computation to be utilised.

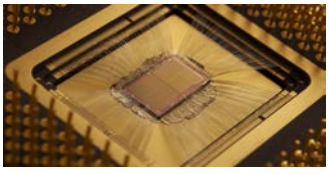


Our DYNAP development kit is available for purchase. Build and test your own neuromorphic solutions using our spiking neuron HW emulators.

The **innovative asynchronous low latency communication infrastructure** (patented) enables each neuron to communicate with up to 230k surrounding neurons and infinite scalability via relay neurons for truly large scale networks.

DynapSEL

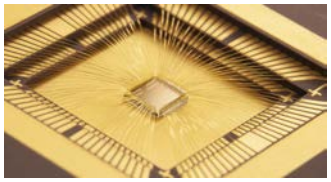
Our next-generation neuromorphic chip, **DynapSEL**, features 1k analog low-power spiking neurons and almost 80k configurable synaptic connections, including 8k synapses with integrated spike-based learning rules. In comparison with our nearest competitor, the TrueNorth



DynapSEL spiking neuron emulator with plastic synaptic connections.

chip from IBM, the DynapSEL features **8x the synaptic fan-in and 32x the fan-out**; these numbers are in the range of connectivity architectures observed in mammalian cortex, and permit networks with biologically realistic connectivity to be emulated using DynapSEL. In addition, the DynapSEL is considerably more energy efficient than TrueNorth, requiring less than a fifth of the energy expenditure per synaptic event and less than half the energy per neuron spike. This is combined with a speed advantage in favour of DynapSEL, resulting in more than **10 times the number of synaptic events per second per watt**.

DynapCNN

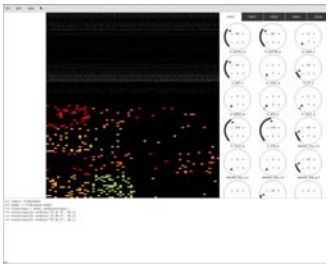


DynapCNN event-based computational fabric for convolutional network architectures, tuned for image processing.

DynapCNN is a scalable, fully-configurable digital event-driven neuromorphic processor with 1M ReLU spiking neurons per chip for implementing **Spiking Convolutional Neural Networks** (sCNNs). This technology is ideal for **always-on, ultra-low power and ultra-low latency** event-driven sensory processing applications. With a dedicated interface for dynamic-vision-sensors, it allows direct input of event streams from most of the advanced dynamic-vision-senors in the world. DynapCNN is fully configurable and supports several types of CNN layers (e.g. ReLU, Cropping, Padding and Pooling) and network components (e.g. LeNet, ResNet and Inception). DynapCNN is highly scalable, enabling implementation of **deep neural networks with unlimited number of layers** over multiple interconnected DynapCNNs.

ctxctl software

ctxctl is the front end interface to the aiCTX toolchain, and cross-platform environment for interacting with our devices. It provides a Python API and data visualization tools for working with spiking neural networks.



ctxctl software visualising the firing activity of a population of spiking analog neurons.

Key personnel



Dr. Ning Qiao, PhD — *CEO, CTO, co-founder*
Dr. Qiao is an expert analog and mixed-signal neuromorphic engineer, and holds the position of Chief Executive Officer at aiCTX. Dr Qiao designs next-generation mixed-signal neuromorphic circuits for rapid event-driven processing. He is the founder, CEO and CTO of aiCTX.



Prof. Dr. Giacomo Indiveri, PhD — *CSO, co-founder*
Prof. Dr. Indiveri is an expert neuromorphic engineer, with more than 20 years' experience in analog and mixed-signal hardware design, and extensive experience managing the development of neuromorphic computing hardware. He holds the positions of Chief Science Officer at aiCTX, as well as Director of the Institute of Neuroinformatics at the University of Zurich and ETH Zurich.



Dr. Kynan Eng — *Chairman, co-founder*
Dr. Eng is a co-founder and chairman of aiCTX. He has led startups to successfully closed funding rounds and product launches (YouRehab, iniVation), and has worked at companies including ABB, Alstom and a network management supplier for Telstra. He holds degrees in computer science and mechanical engineering from Monash University, and a PhD from the ETH Zurich, where he also led a research group investigating the neuroscience of virtual reality interactions.

Francois Stieger — *Board advisor*

Mr Stieger is an experienced executive, investor and board advisor. He has served in senior management roles at Verisign, Oracle, and Broadvision, where he was involved in the IPO and listing on NASDAQ. He is a former partner at Amadeus Capital and is currently active in S-Partners and in other investment vehicles.



Dr. Dylan Muir, PhD — *Senior R&D Engineer*

Dr. Muir is a specialist in architectures for neural computation. He has published extensively in computational and experimental neuroscience. At aiCTX he is responsible for directing development of neural architectures for signal processing. Dr. Muir holds a Doctor of Science (PhD) from ETH Zürich, and undergraduate degrees (Masters) in Electronic Engineering and in Computer Science from QUT, Australia.



Dr. Sadique Sheik — *Senior R&D Engineer*

Dr. Sheik is an expert in neuromorphic computation. He specializes in temporal signal processing in asynchronous event-driven systems and has a decade of experience working with low-power, real-time, mixed-signal neuromorphic hardware. He is passionate about substrate-driven and substrate aware computation. At aiCTX he is responsible for directing development of neuromorphic design efforts for low-power real-time systems. Dr. Sheik holds a Doctor of Science (PhD) from ETH Zürich, and undergraduate degrees (Masters and Bachelors) in Physics from Indian Institute of Technology, Kharagpur.





Christos Sourmpis — *R&D Engineering intern*
Mr Sourmpis is an electrical engineer with a strong math and neuro-science background. He is experienced in working with spiking neural network architectures, neural simulations and online learning.

Forward-looking statements disclaimer

Certain information set forth in this document contains "forward-looking information", including "future oriented financial information" and "financial outlook", under applicable securities laws (collectively referred to herein as forward-looking statements). Except for statements of historical fact, information contained herein constitutes forward-looking statements and includes, but is not limited to

1. the projected financial performance of aiCTX AG (hereafter referred to as "the Company");
2. the expected development of the Company's business, projects and joint ventures;
3. execution of the Company's vision and growth strategy;
4. sources and availability of third-party financing for the Company's projects;
5. completion of the Company's projects that are currently underway, in development or otherwise under consideration;
6. renewal of the Company's current customer, supplier and other material agreements; and
7. future liquidity, working capital, and capital requirements.

Forward-looking statements are provided to allow potential investors the opportunity to understand management's beliefs and opinions in

respect of the future so that they may use such beliefs and opinions as one factor in evaluating an investment. These statements are not guarantees of future performance and undue reliance should not be placed on them. Such forward-looking statements necessarily involve known and unknown risks and uncertainties, which may cause actual performance and financial results in future periods to differ materially from any projections of future performance or result expressed or implied by such forward-looking statements. The Company undertakes no obligation to update forward-looking statements if circumstances or management's estimates or opinions should change except as required by applicable securities laws. 

Image credits

Vinyl record macro grooves Evan Amos. `wiki`

Time-coded vinyl record Jason7825. `wiki`

PIA01817: Space Radar Image of St. Louis, Missouri. Courtesy NASA/JPL-Caltech. `NASA/JPL`