Audio Analytic

2019 Intelligent Sensing Summer school
CIS - Queen Mary University of London

The challenges and benefits of sound sensing

Dr Sacha Krstulović

Director of Research – Audio Analytic

September 4th, 2019

"Like a Shazam for real-world sounds"

Bloomberg

as featured in

THE TIMES

techradar.

WIRED

engadget

BBC

The Economist

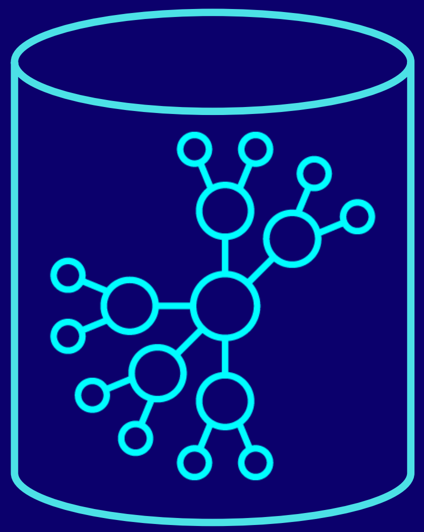
FAST COMPANY

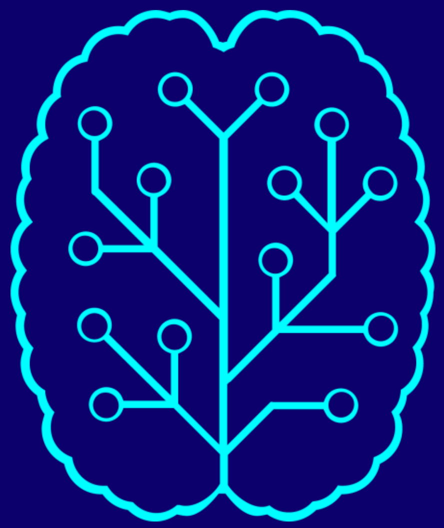
IEEE SPECTRUM

Forbes

Gartner

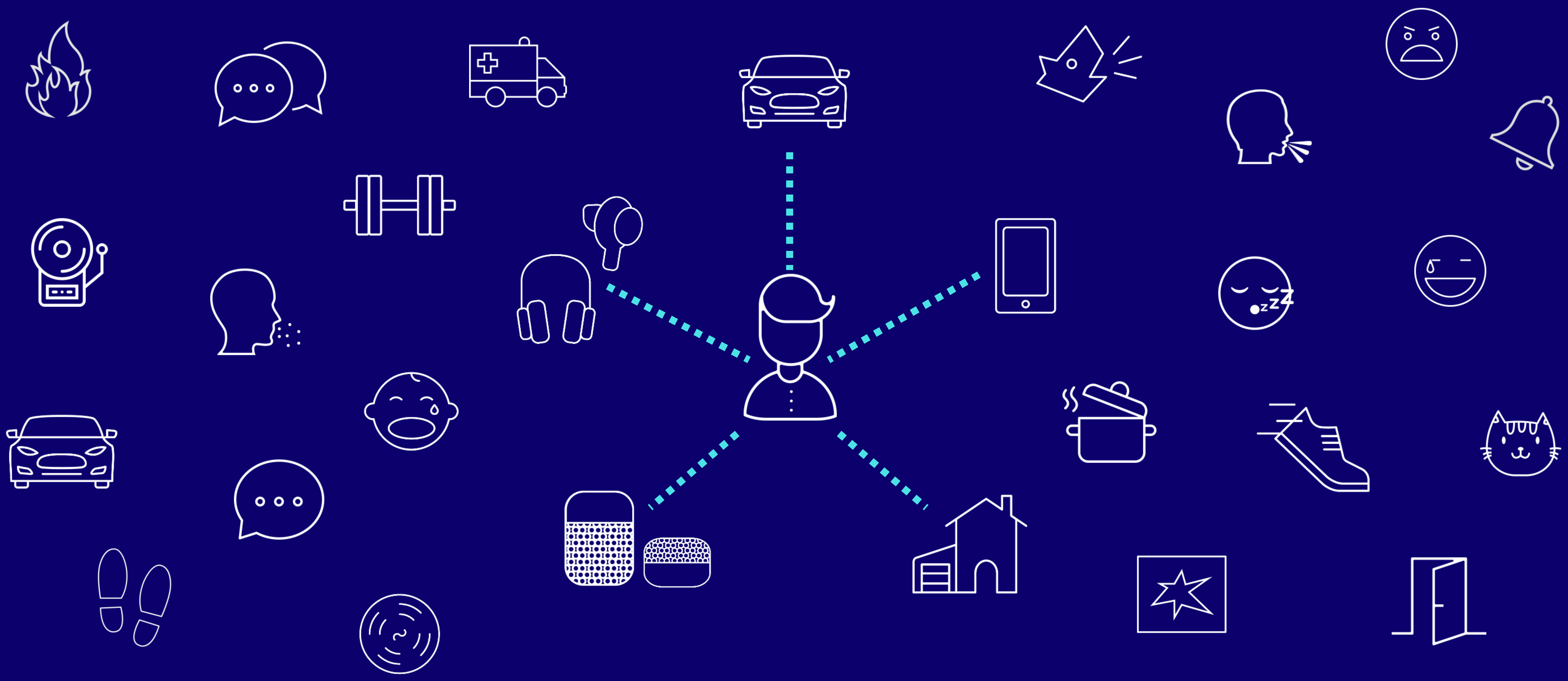
Audio Analytic’s software and technology give machines the broader sense of hearing.





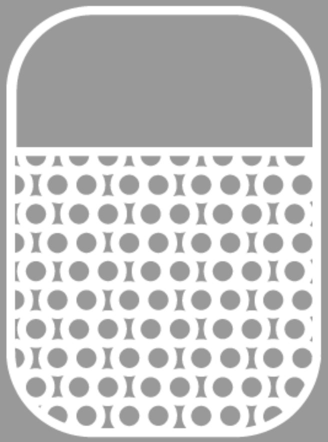
The benefits

Evolving from ‘connected devices’ to ‘intelligent experiences’



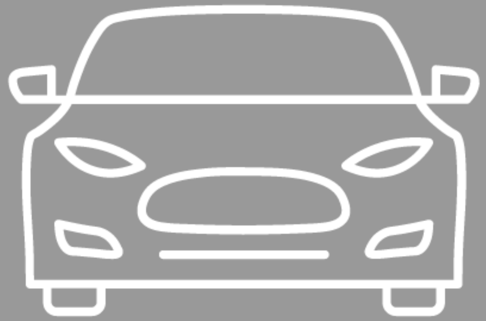
Audio Analytic

230m smart home devices in 2021*

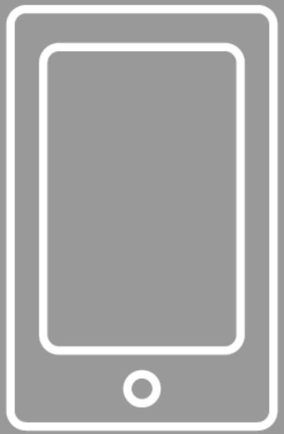


206m smart speakers in 2021*

82m hearables in 2021^



92m connected cars in 2021*



1.7bn smartphones in 2021*

"Sound recognition is a key strategic technology that should be made available in all connected devices."

Francisco Jeronimo, IDC

*Source: *IDC forecasts, ^SAR Insight/Audio Analytic forecasts*

Audio Analytic

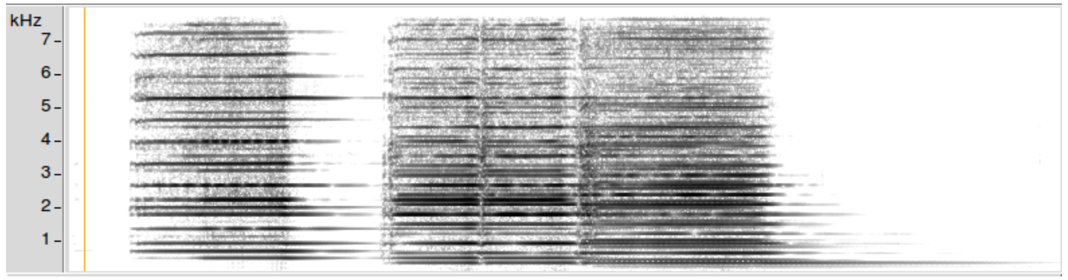
The challenges

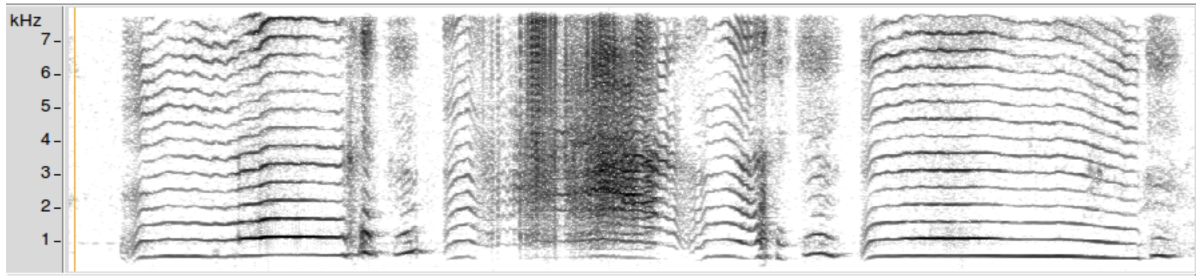
The core challenge: acoustic variability

Variety of production processes = variety of acoustic features.

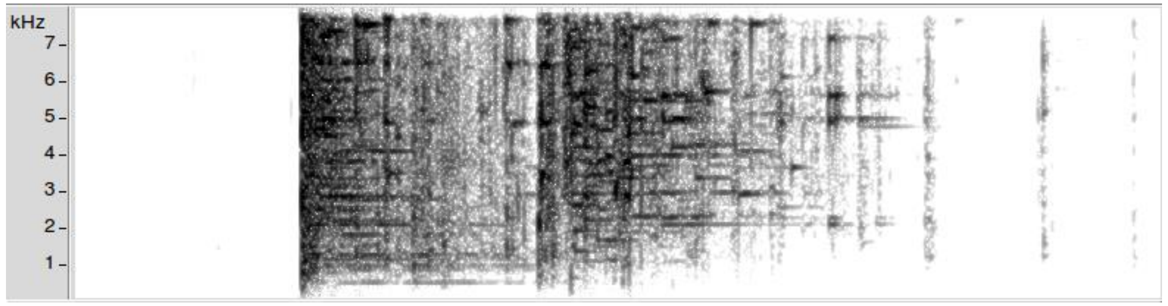
Beeps

Harmonic Sounds

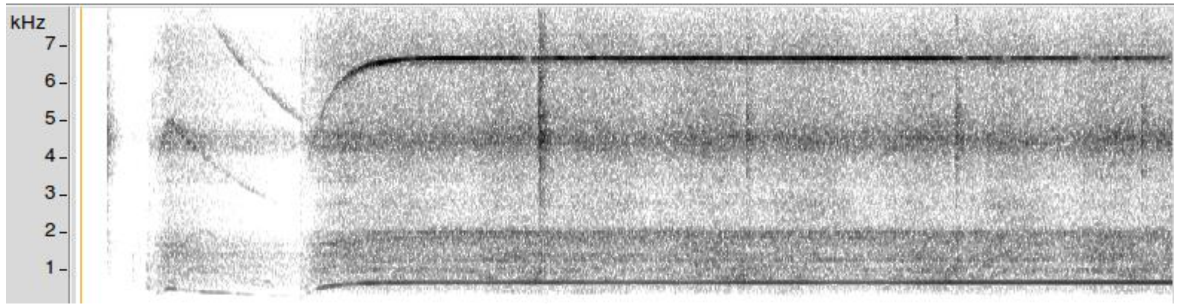




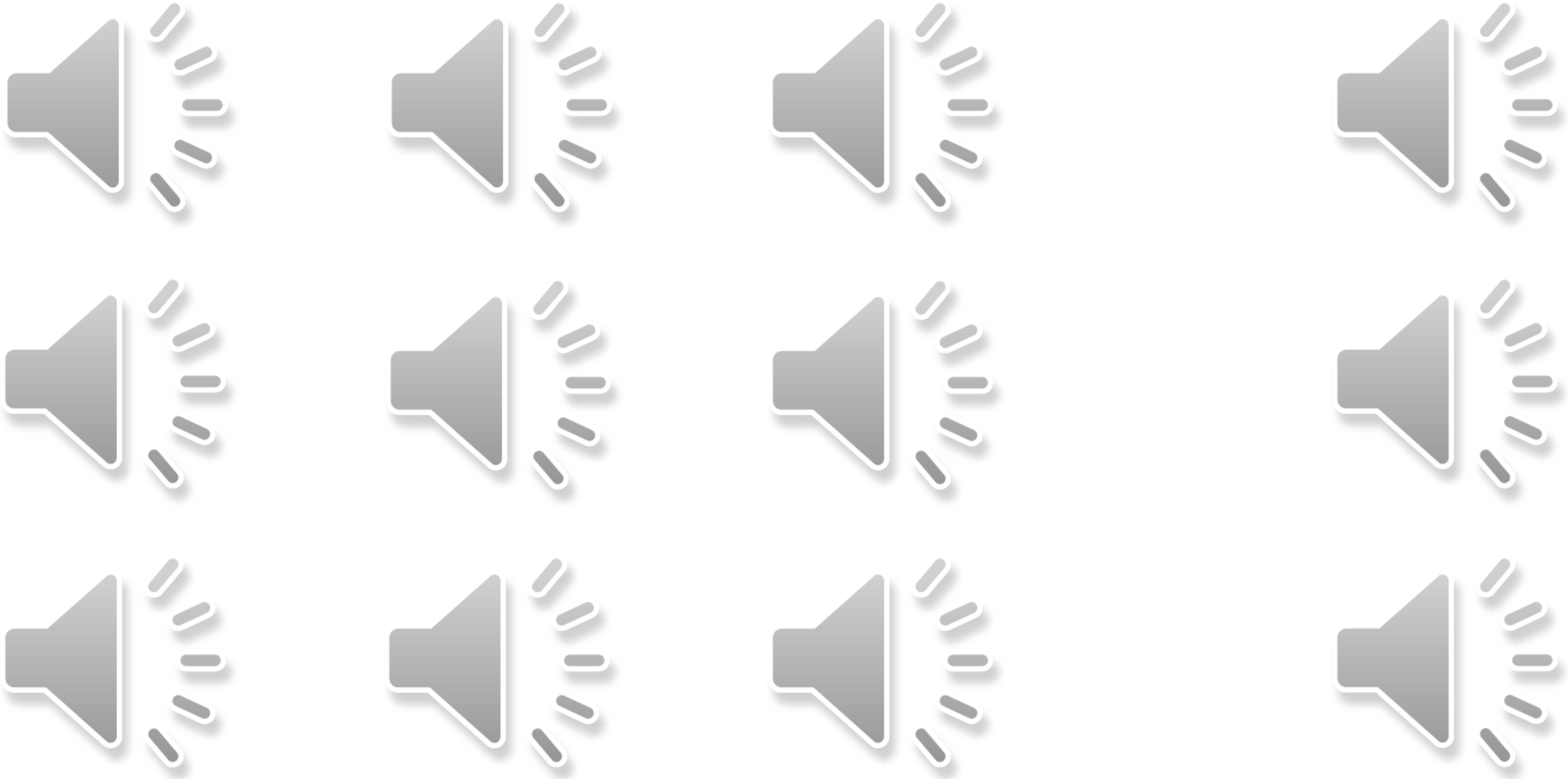
Crash/Bangs



Shaped noise

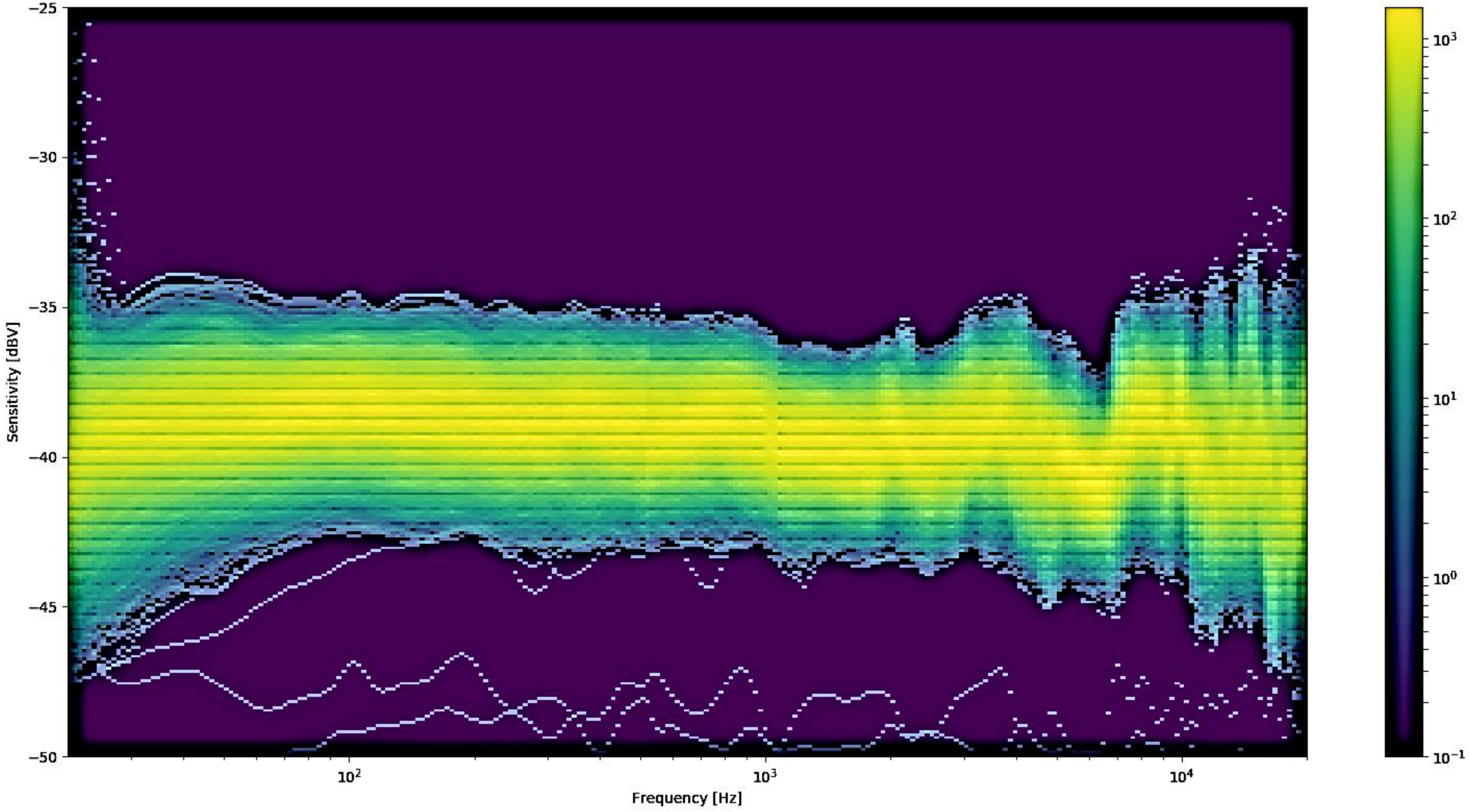


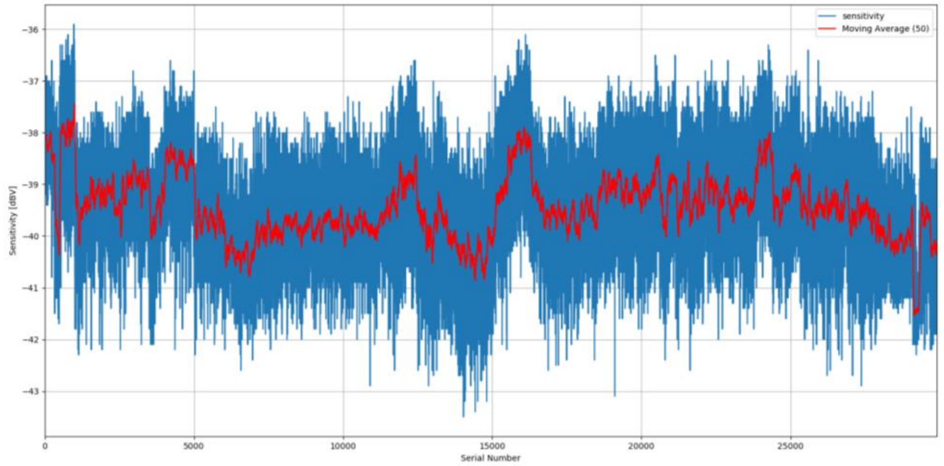
Effects of channel and room variety

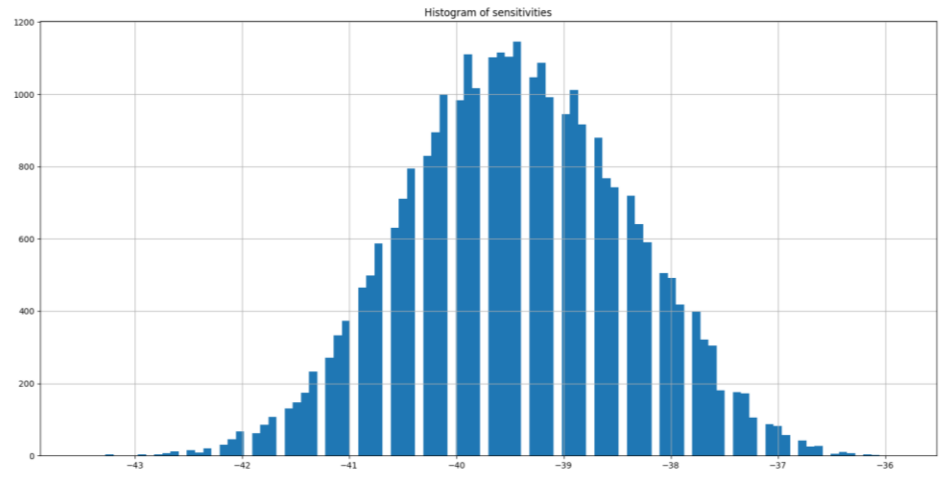


Microphone variability

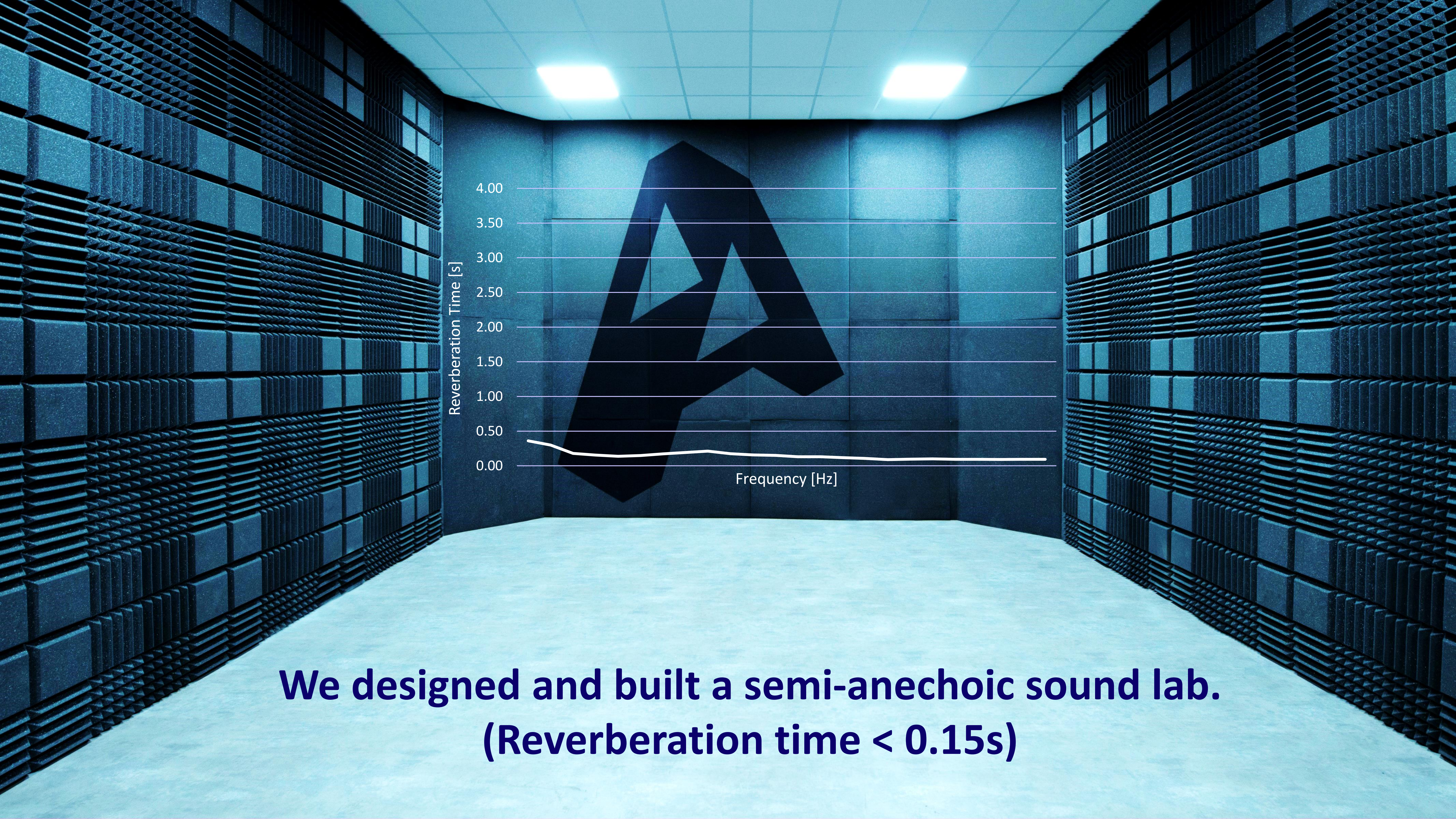
Analysis of 29,900 microphone response curves from the same manufacturer and model:







Audio Analytic



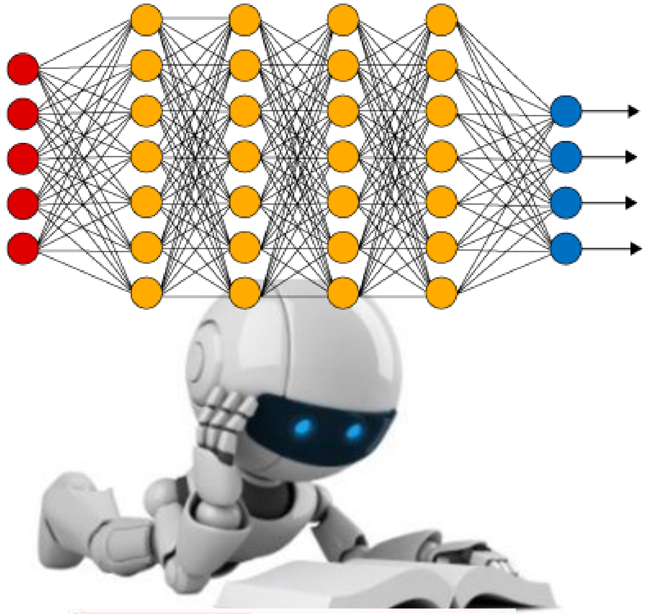
We designed and built a semi-anechoic sound lab.
(Reverberation time < 0.15s)

Data augmentation with realistic conditions

Training Data

Convolution with real or simulated room and microphone responses

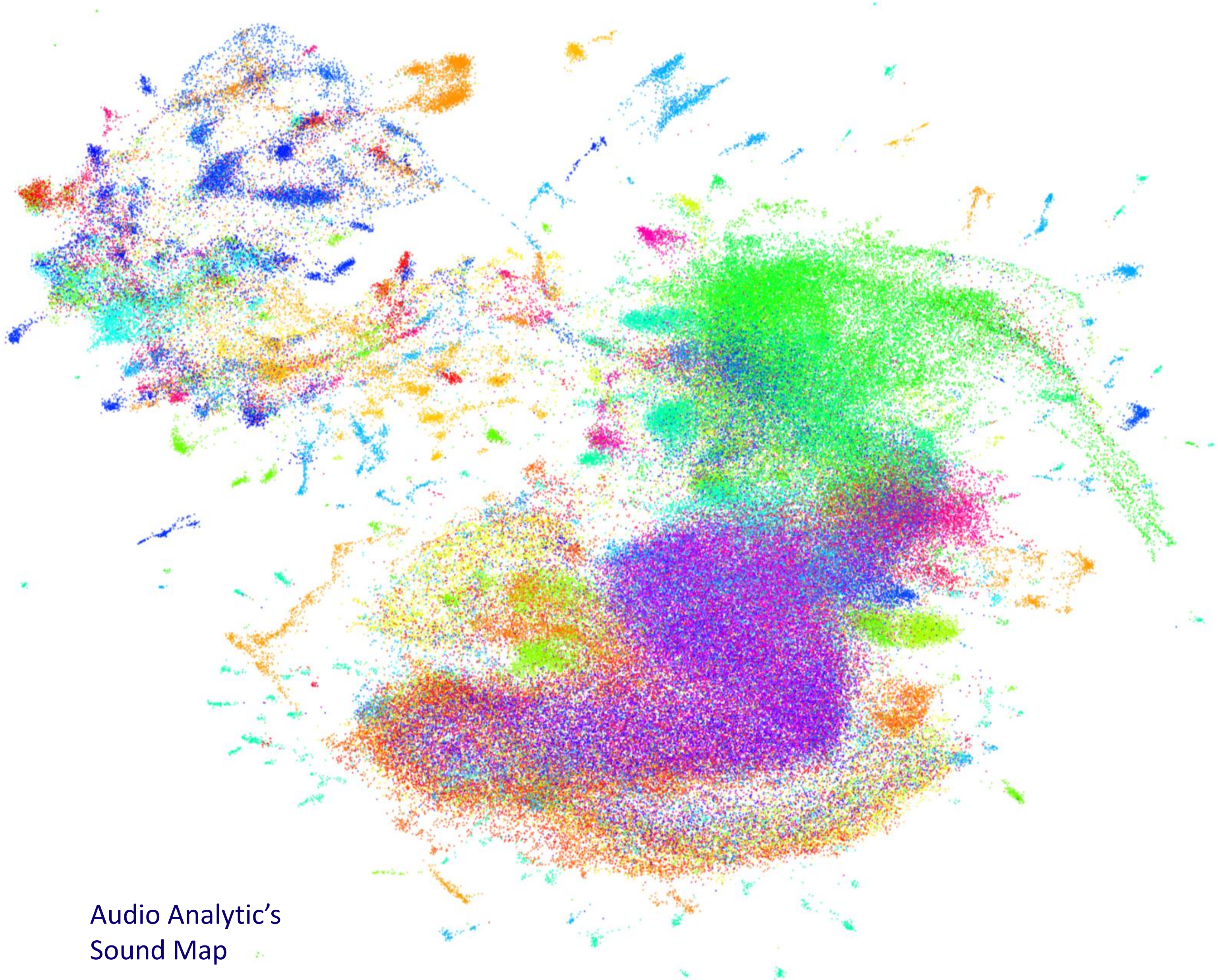
Augmented Training Data



Not quite zero data anymore…

- AA's Alexandria™ database contains 13,508,727 labelled sound events across ~600 label types.



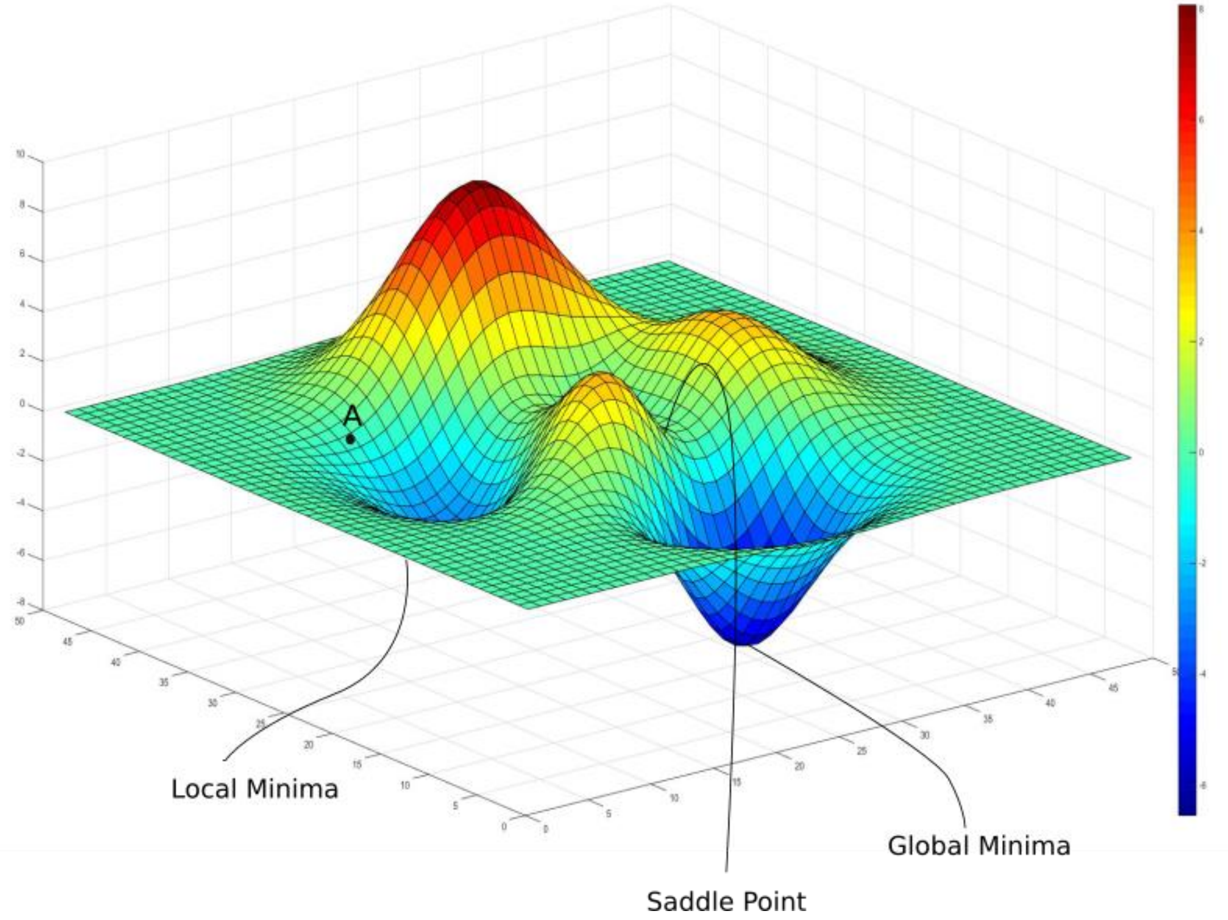


Audio Analytic's
Sound Map

What is the best sound recognition model?

- Think of it as an optimisation problem:
  - Given a realistic data set
  - And a meaningful metric for sound classification
  - And a search space of DNN architectures

  then find the best performing architecture
- Meaningful metric:
  - Is cross-entropy on audio frames meaningful?
  - True positives are discrete, but what about false positives in continuous audio?
  - Prior probabilities: empty home or busy home? What is the system really exposed to?
- Model search space: Software 2.0 approach where guided search yields AuditoryNet™



Differentiable programming / Software 2.0

"Deep Learning est mort. Vive Differentiable Programming!
People are now building a new kind of software by assembling networks of parameterized functional blocks and by training them from examples using some form of gradient-based optimization.
It's really very much like a regular program, except it's parameterized, automatically differentiated, and trainable/optimizable."
- Yann LeCun, Director of Facebook AI Research

The "classical stack" of Software 1.0 is what we're all familiar with — it is written in languages such as Python, C++, etc. […] By writing each line of code, the programmer identifies a specific point in program space with some desirable behavior.
In contrast, Software 2.0 […] approach is to specify some goal on the behavior of a desirable program (e.g., "satisfy a dataset of input output pairs of examples", or "win a game of Go"), write a rough skeleton of the code (e.g. a neural net architecture) that identifies a subset of program space to search, and use the computational resources at our disposal to search this space for a program that works […] the search process can be made efficient with backpropagation and stochastic gradient descent.

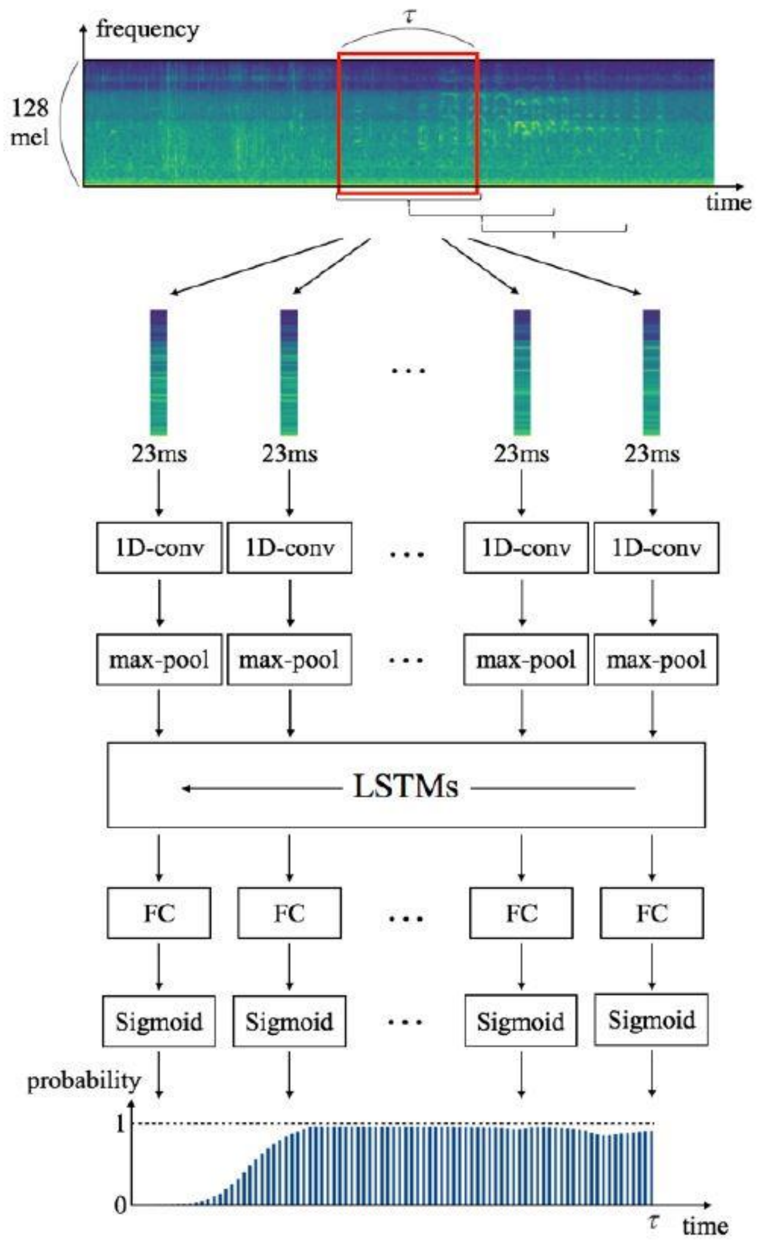
- Andrej Karpathy, Senior Director of AI at Tesla
  https://vimeo.com/274274744

Software 2.0 search space: e.g. recent DCASE challenges

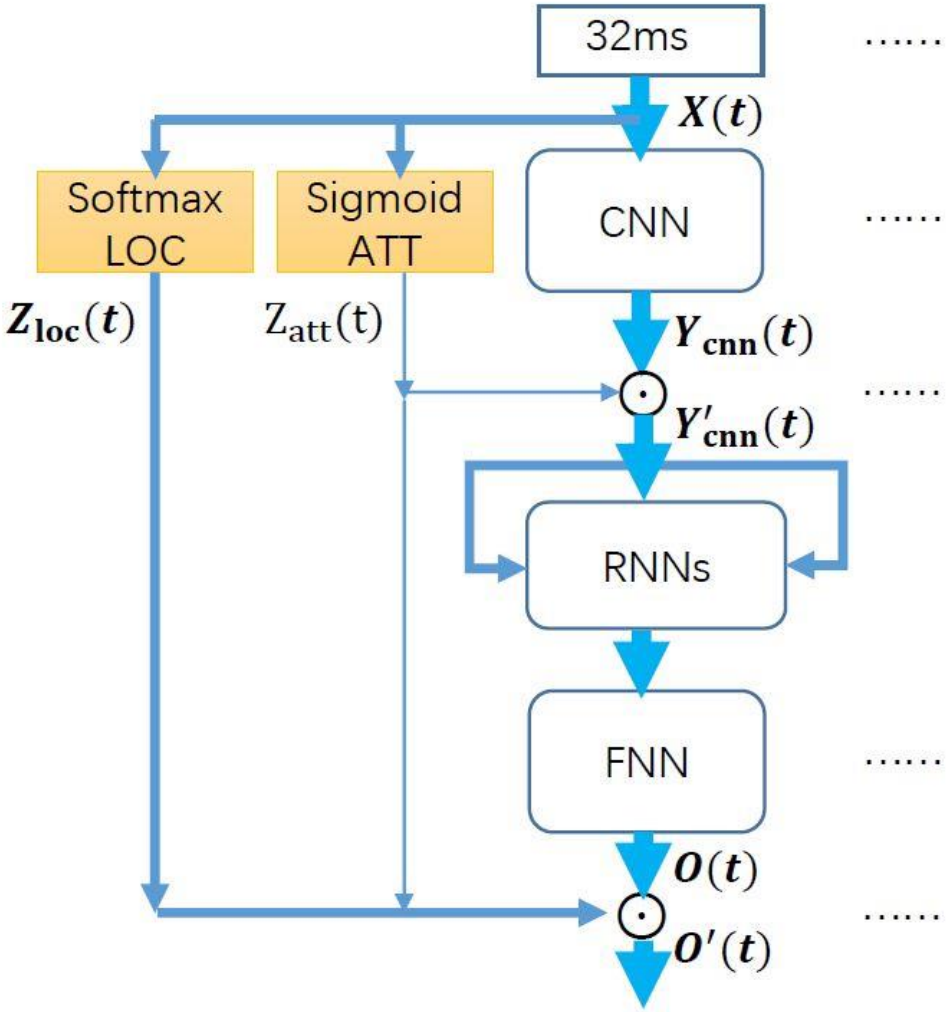
acoustic analysis (CNN)

sequence modelling (LSTM-RNN)

decision layer (FFDNN)



H. Lim, J. Park and Y. Han, DCASE 2017 T2



"Attention and Localization Based on a Deep Convolutional Recurrent Model for Weakly Supervised Audio Tagging"
Y. Xu, Q. Kong, Q. Huang, W. Wang and M. D. Plumbley
Interspeech 2017 - DCASE 2016 Task4

System evaluation matters

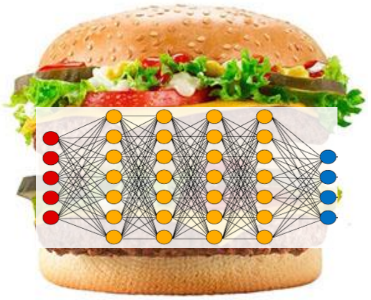
- Open evaluations can be useful
  - Supply reference data, foster development and exchange of novel ideas
  - Sound scenes and events recognition: DCASE
- Open evaluations can also bias the problem significantly
  - Is the database realistic or whatever was affordable?
  - Solving the problem or winning the competitive eval?
  - Are comparative eval metrics relevant to user experience?
- Evaluation in real conditions is hard and costly, but necessary for industrial success
  - Real time operation on a consumer chip?
  - Required audio and computation hardware?
  - Were significant variability effects neglected?
- Is your system really doing what you think it is doing?
  - Clever Hans was a horse who could do mathematics…

Privacy

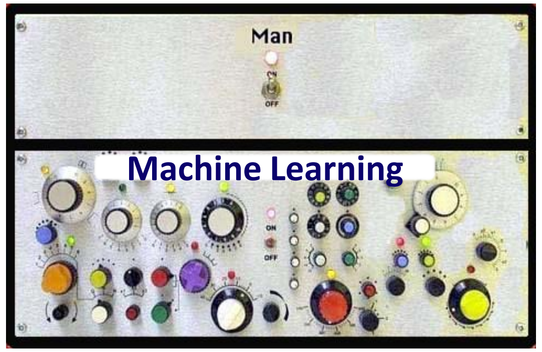
- Privacy concerns can be traced back to WWII
  - Privacy laws largely rely on definitions of *sensitive* and *identifiable* information
  - Consent matters
- Public perception matters and impacts commercial success.
  - Limited faith in how data is being protected and processed in the cloud
  - Limited faith in T&Cs
- People would like AI devices to be more like humans: all the knowledge is locked into the head
  - Running onboard the device matters
  - Avoid data leaving the device
- New technical solutions are emerging
  - E.g., federated learning

Take aways

- Data is a parameter: must capture full real-world variability
  - Part of industrial expertise is about judging coverage, relevance and variability in order to build high performance data sets



- DNNs are a given, but what is the best model?
  - It’s an optimisation problem: Software 2.0, expert-guided search into the model space given realistic data



- Meaningful metrics are crucial
  - Comparative metrics may not be relevant to user experience
  - The system is a horse until proven otherwise

- AI is expected to remain as private as the brain

Many thanks!

We are hiring!
http://www.audioanalytic.com/careers

UK headquarters
2 Quayside
Cambridge
CB5 8AB, UK

Sound Labs
Unit 11b, Nuffield Road
Cambridge
CB4 1TF, UK

US office
44 Montgomery Street
San Francisco
CA 94104, USA

info@audioanalytic.com
audioanalytic.com