

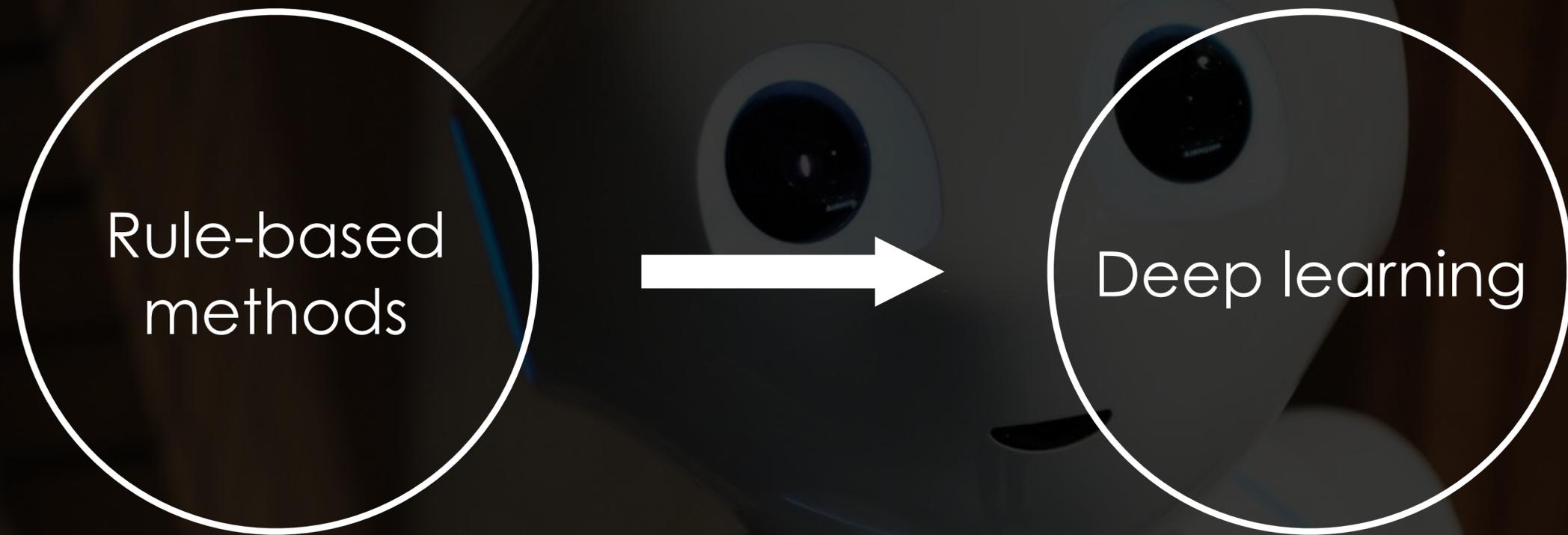
cochlear.ai

Audio recognition, **context-awareness**,
and its applications

Yoonchang Han

Co-founder & CEO, Cochlear.ai

26 March, 2018



Rule-based
methods

Deep learning



See



Computer vision



Understand
language



Natural language
processing



Listen



Speech recognition



Taking an umbrella

Closing the window



Foot step sound

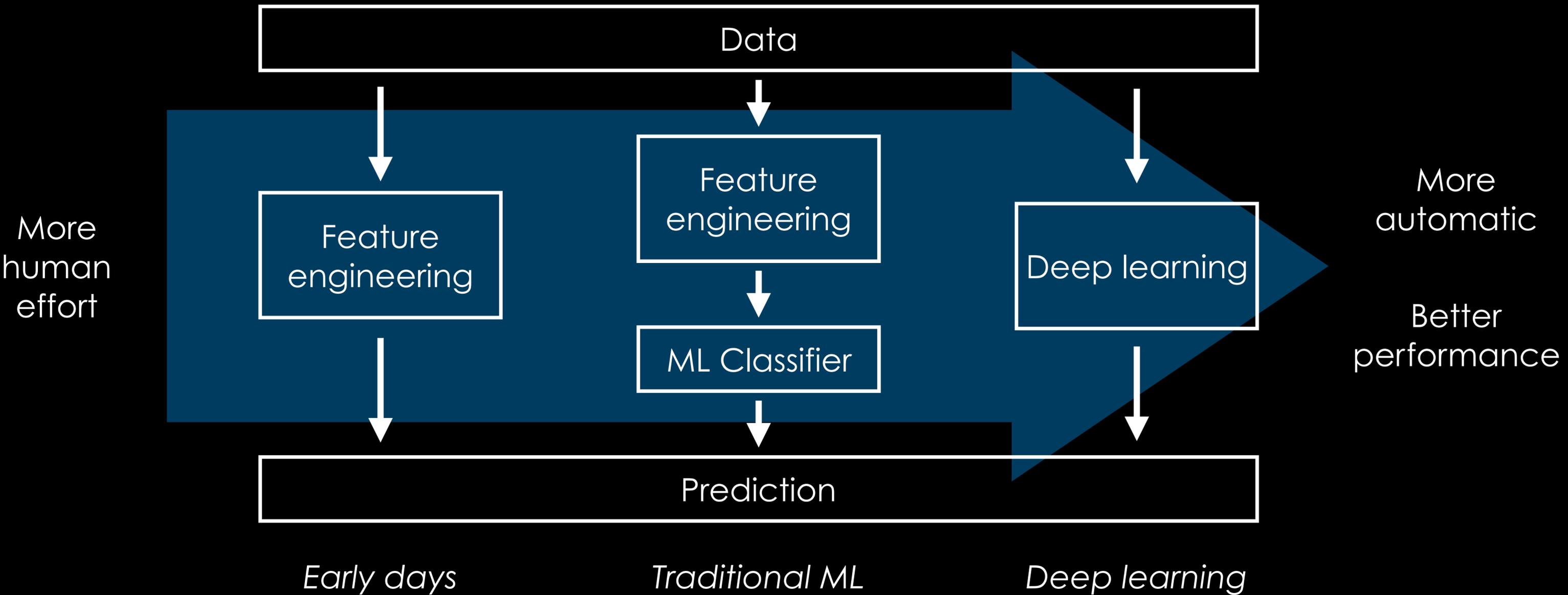
High heels



Easy for Humans

Hard for Machines

Evolution of data processing technique



Domain knowledge

To tackle each topic
(make some “rules”)



To simulate how
human understand the sound
(and prepare data)

Required domain knowledge

Signal
Processing

Cognitive
Sciences

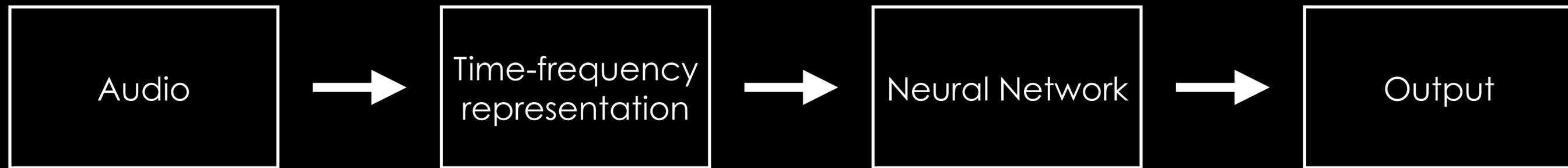
Music

Psychoacoustics

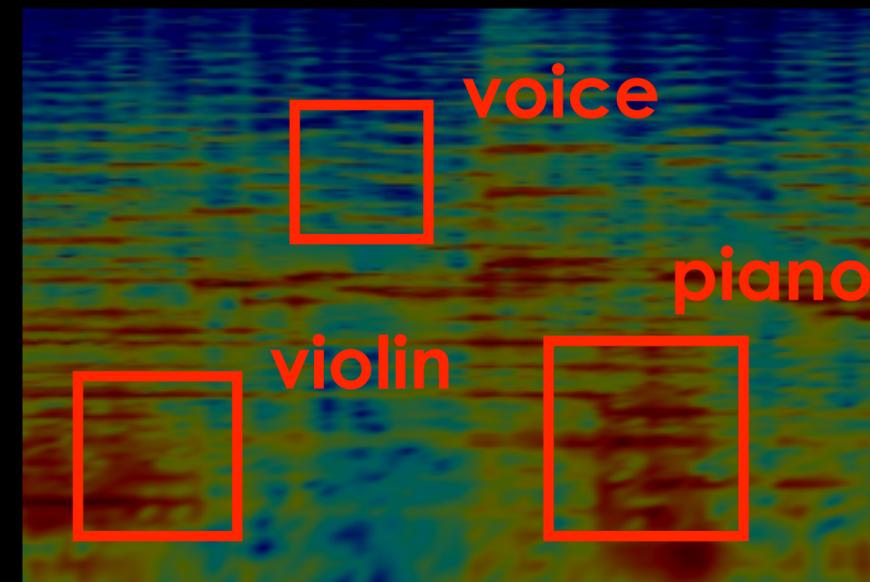
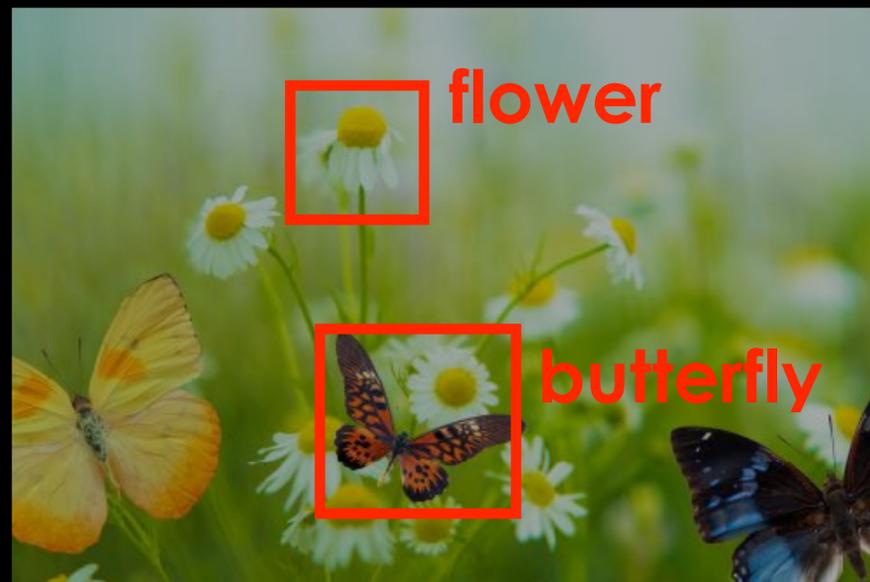
Acoustics

Machine
Learning

“Modern” audio identification pipeline



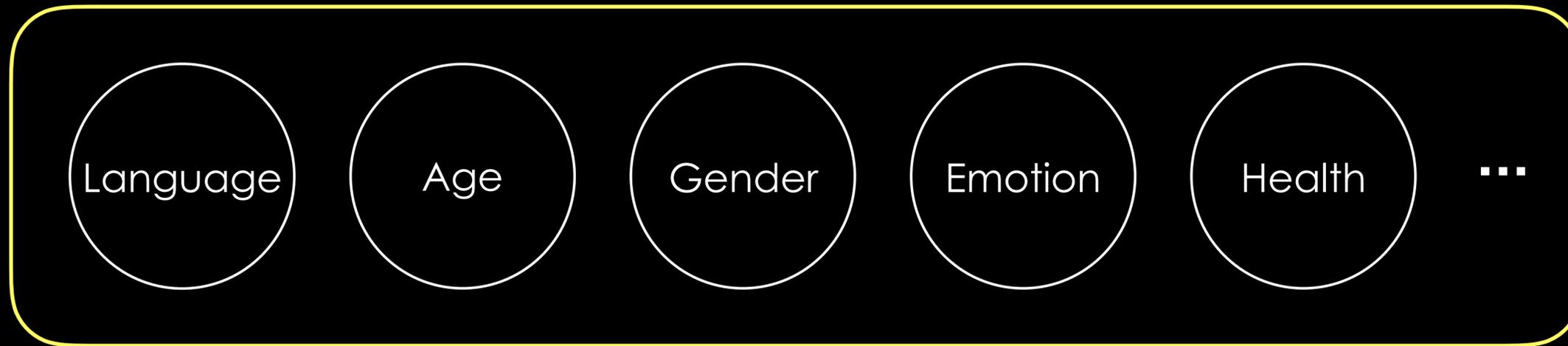
objects in an image \approx instruments in a spectrogram



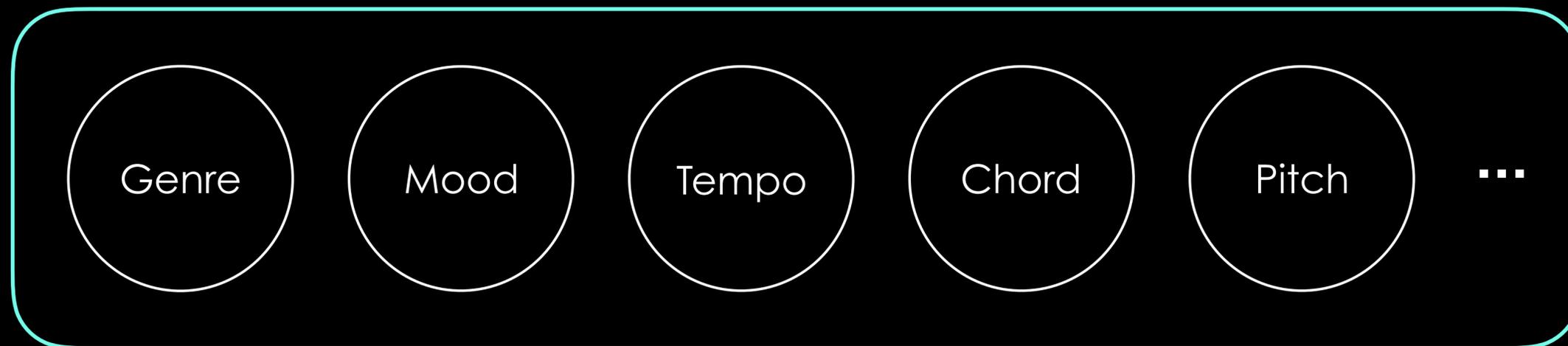
“**Machine listening**” is the use of signal processing and machine learning for making sense of natural / everyday sounds, and recorded music.

- Machine listening lab, Queen Mary, Univ. of London

Voice



Music



Machine listening

Acoustic scenes

bus

park

library

city centre

train

driving

home

market

cafe

...

Music

Voice

Acoustic events

glass break

knock

car horn

dog bark

footstep

water boil

gun shot

snoring

bird chirping

crying

sneeze

...

"Any" sound we hear everyday



Computer vision



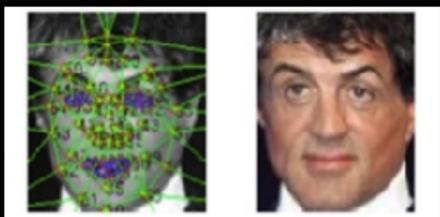
Machine listening



Optical Character Recognition (OCR)



Voice recognition



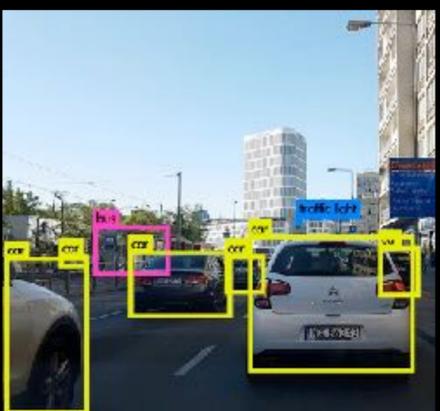
Facial recognition



Music search



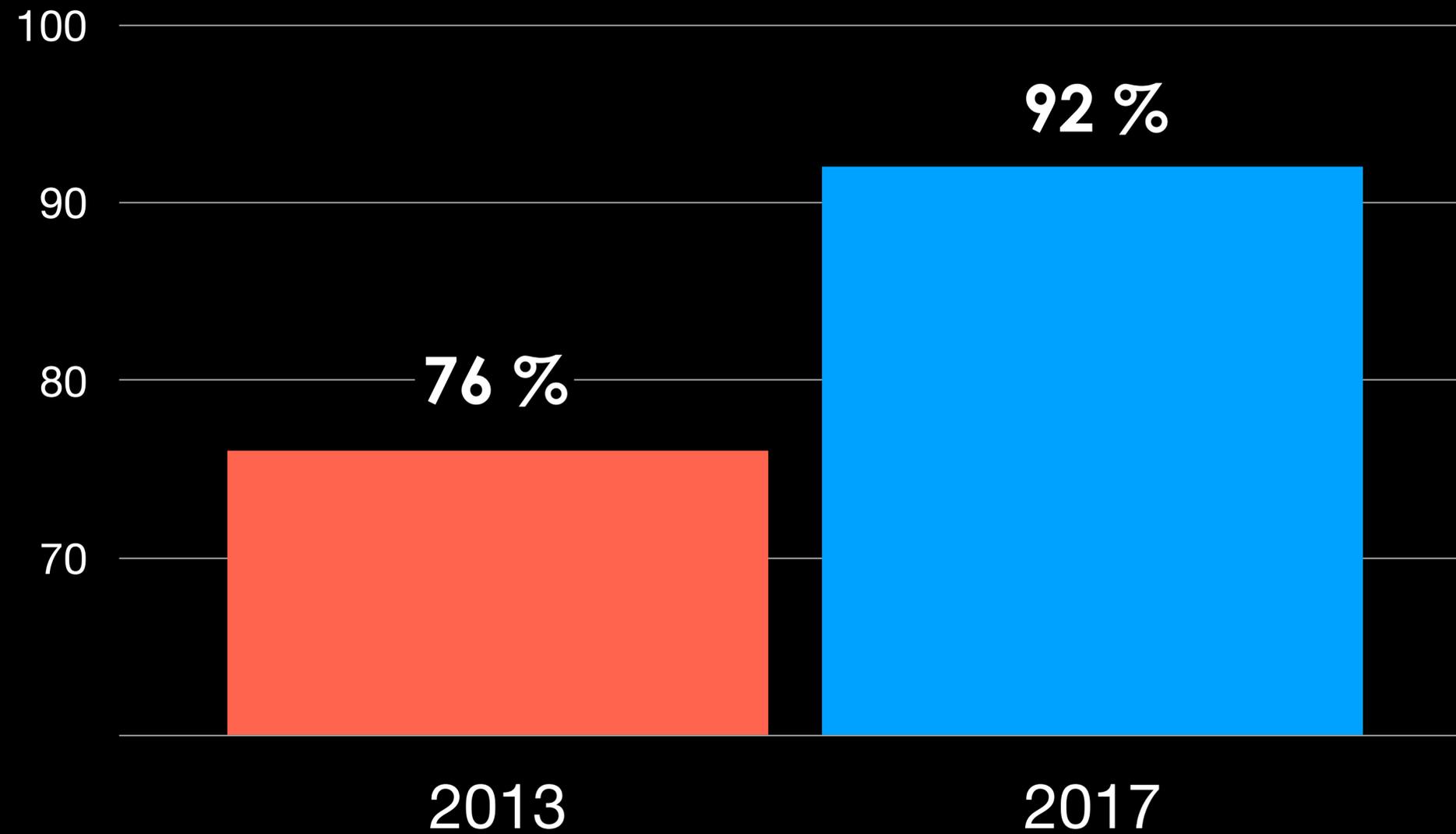
Speaker identification



Object detection



Acoustic scene/event detection



Scene classification accuracy (IEEE DCASE)

A Venn diagram illustrating the relationship between three fields: Deep Learning, Machine Learning, and Artificial Intelligence. It consists of three overlapping circles. The leftmost circle is labeled 'Deep Learning'. The middle circle is labeled 'Machine Learning'. The rightmost circle is labeled 'Artificial Intelligence'. The circles overlap in various combinations, with the largest circle encompassing all three.

Deep
Learning

Machine
Learning

Artificial
Intelligence

A young girl with dark hair, wearing a white shirt with a pink pattern and a pink backpack, is holding hands with a white robot. The robot has a round head with large eyes and is holding a tablet. They are surrounded by pink cherry blossoms and colorful festival lights in the background.

Perceive

Think

Act



Five, Zero



Cat

Know what it is (with input restriction)

Know what it is

Know what/where it is

Know what/where it is + why

Simple
Identification



Closer to human

cochlear.ai **Sense** (closed alpha release in April)



Music, Speech, Others



Genre / Mood
/ Key / Tempo



Age / Gender
/ Emotion



Indoor / Outdoor
/ Vehicles



Dog bark / Baby cry
Car horn / Snoring ...

Why do we need...

Activity detection

Unified model

It is really challenging because...

Recording environment

Recording device

Noises

Local characteristics

Overlapped / Polyphonic

Probability or Saliency ?

Example: AI speakers

Simple voice control

“Alexa, turn on the light”

“Alexa, play dance music”

“Alexa, turn on TV”



IoT control-tower
with context-awareness

(footstep sound, door slam, cough,
Someone got back home, got a bad cold)

turn on light / TV

play suitable music

adjust room temperature warmer
(not just a pattern, there is a “reason”)

ask to take cold medicine before sleep

Example: Humanoid robots

See things

Understand speech

+

Listen things other than voice

Know who they talk to



(Source: Atlas, Boston Dynamics)

Example: Autonomous car



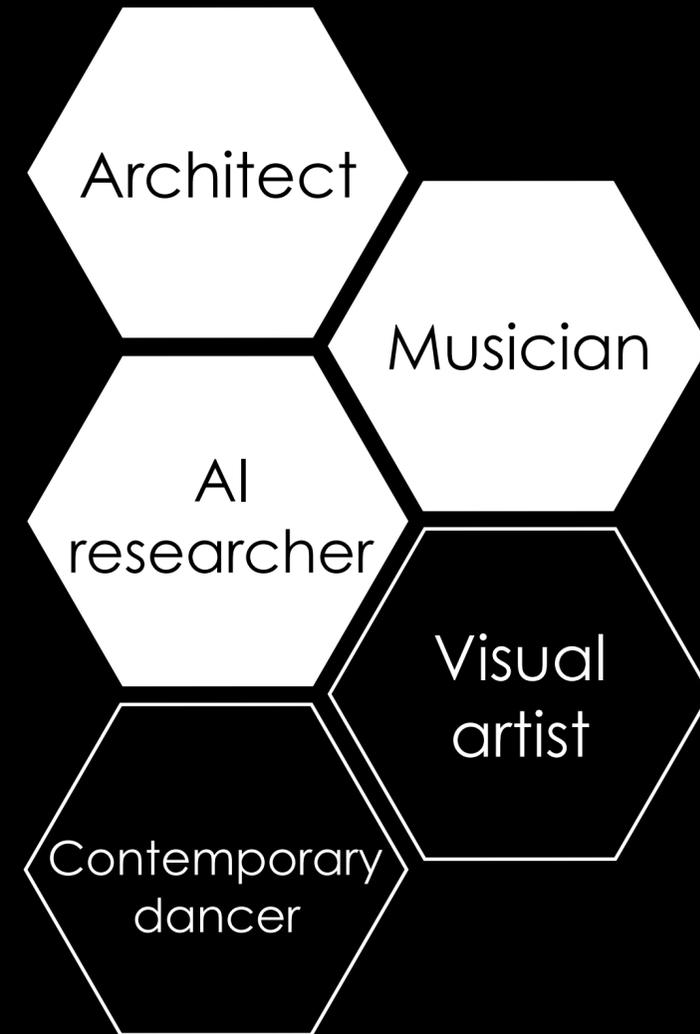
Outside - Car horn (normal, air horn), Siren (fire truck, police, ambulance)

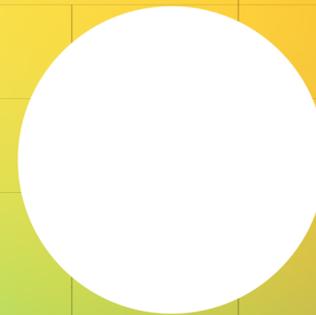
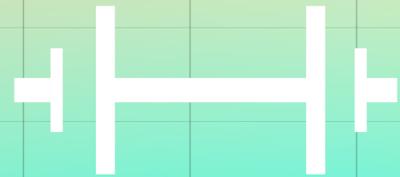
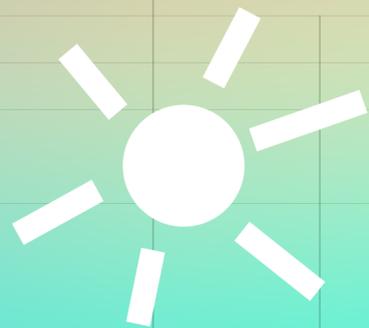
Inside - Music mood, snoring, baby, anomaly detection (malfunction warning)

ATMO: Generative music for spatial atmo-sphere

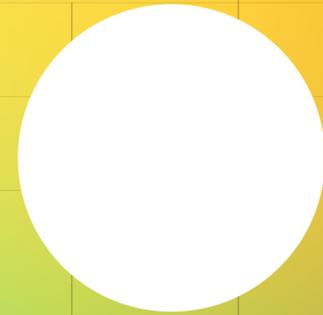
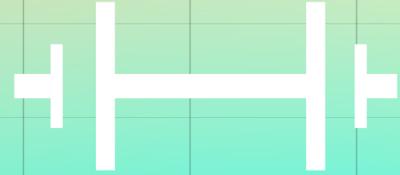
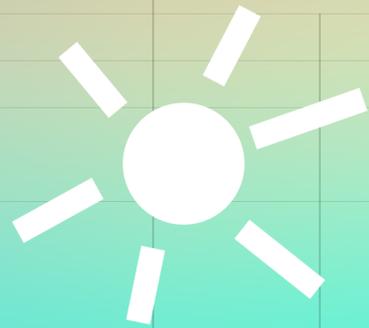
Create NeWave
KoCCA
한국콘텐츠진흥원

+





Generative Music with contextual information

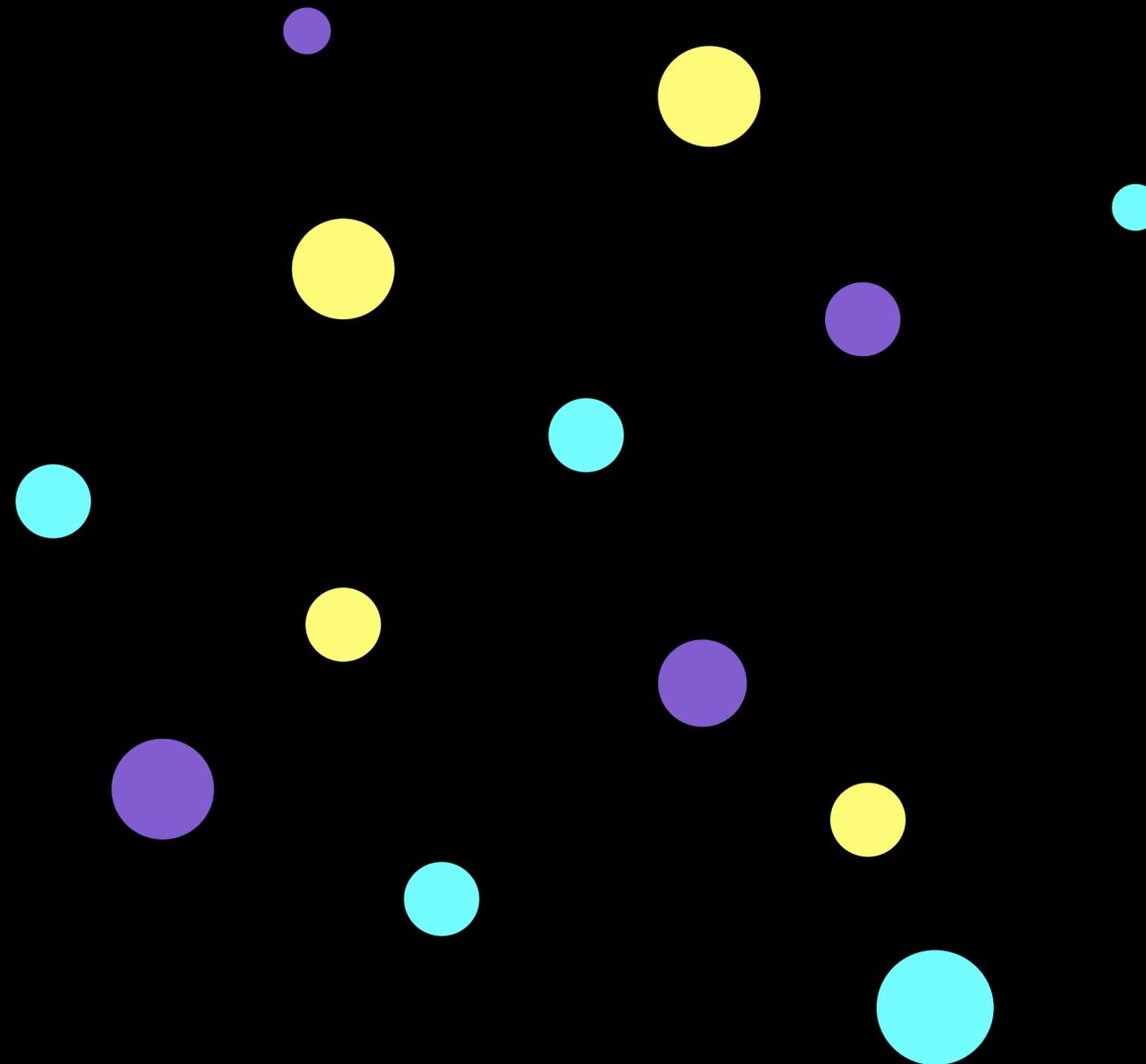


Generative Music with contextual information

Analysis Result : Typing in a rainy day...



Typing...
Reading a book...
Raining outside...



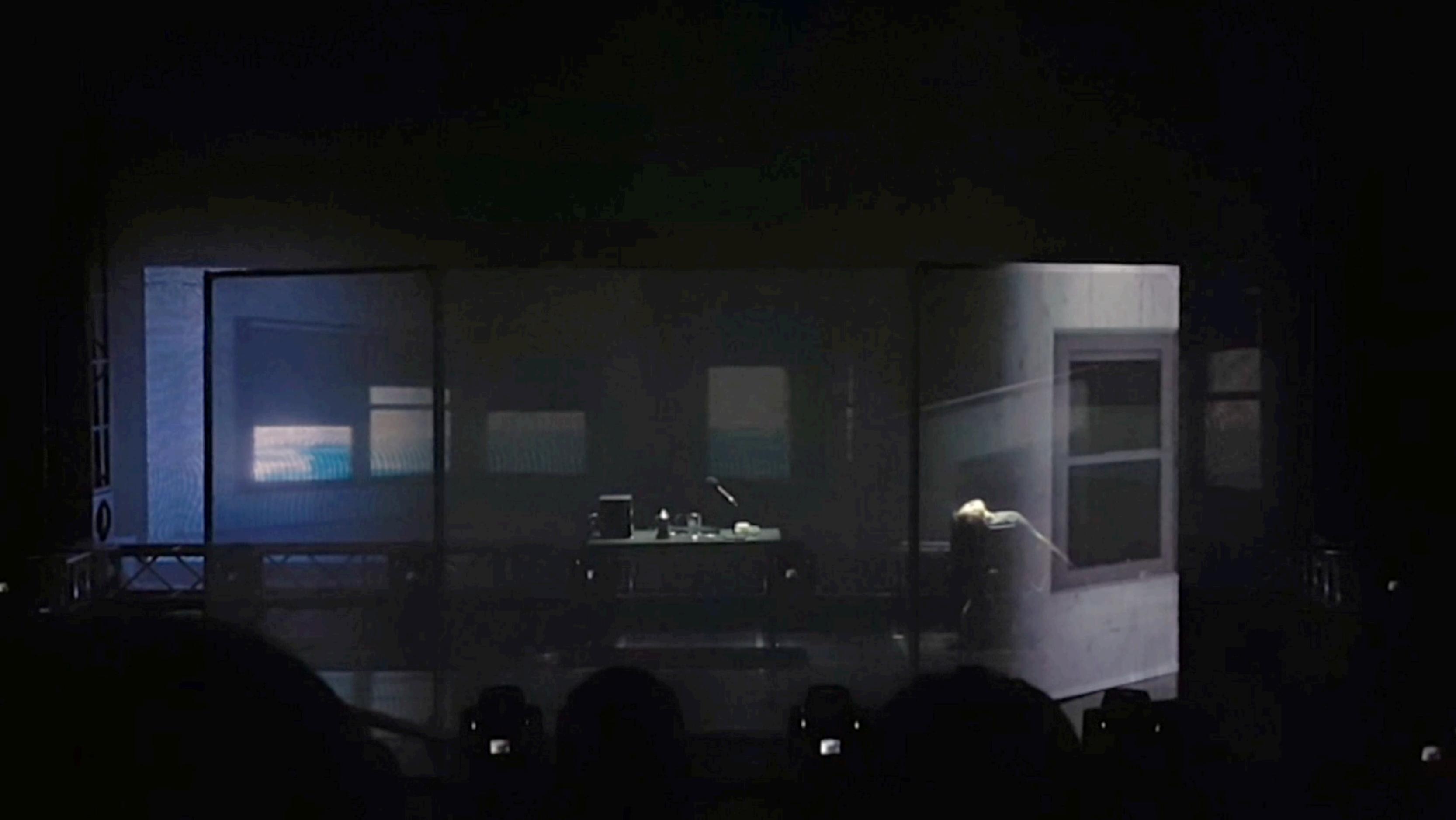
A photograph of a recording studio workstation. A black desk holds a microphone on a stand, a glass of water, a silver teapot, a keyboard, and a mouse. A black studio monitor is positioned to the right. The scene is lit with blue light. Two dashed yellow circles highlight the microphone and the speaker. Labels 'Microphone' and 'Speaker' are placed over the respective items.

Microphone

Speaker

detection : detected
class : /Lighter
amp : 0.632432
wave : noisy
centroid : 3318.813639





cochlear.ai

contact@cochlear.ai