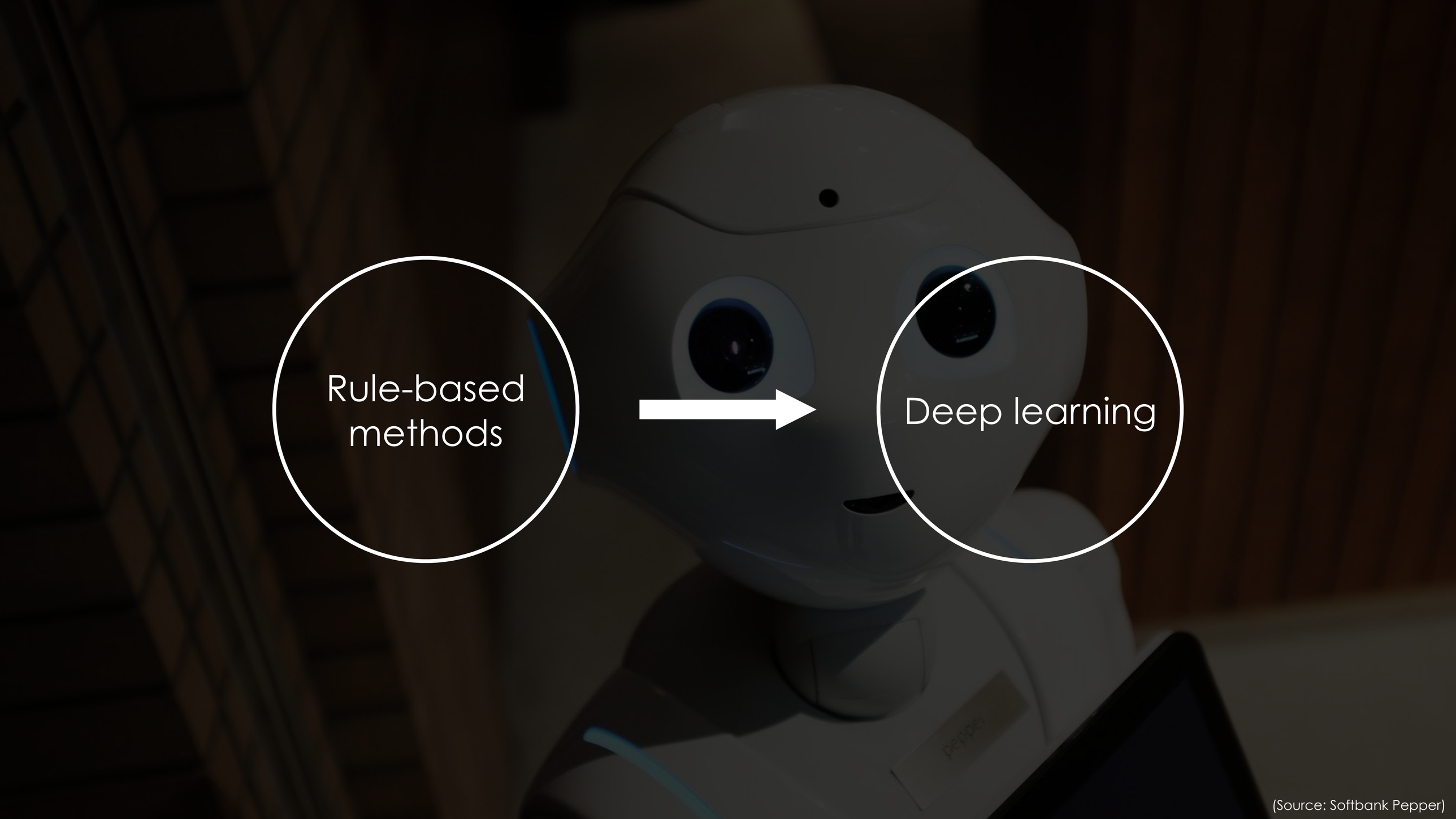
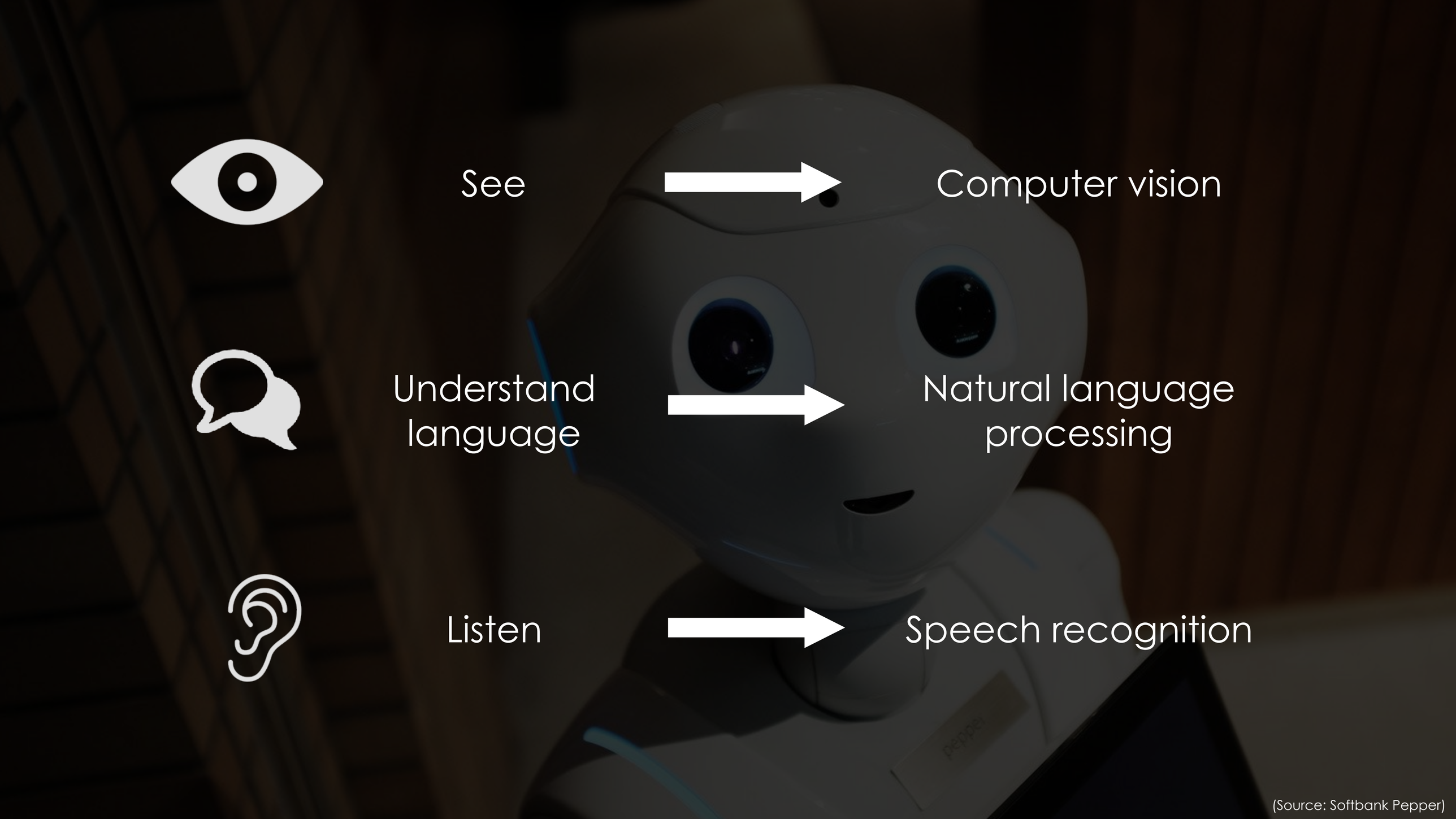
cochlear.ai

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

Yoonchang Han
Co-founder & CEO, Cochlear.ai
26 March, 2018



(Source: Softbank Pepper)



(Source: Softbank Pepper)


Taking an umbrella
Closing the window

Foot step sound

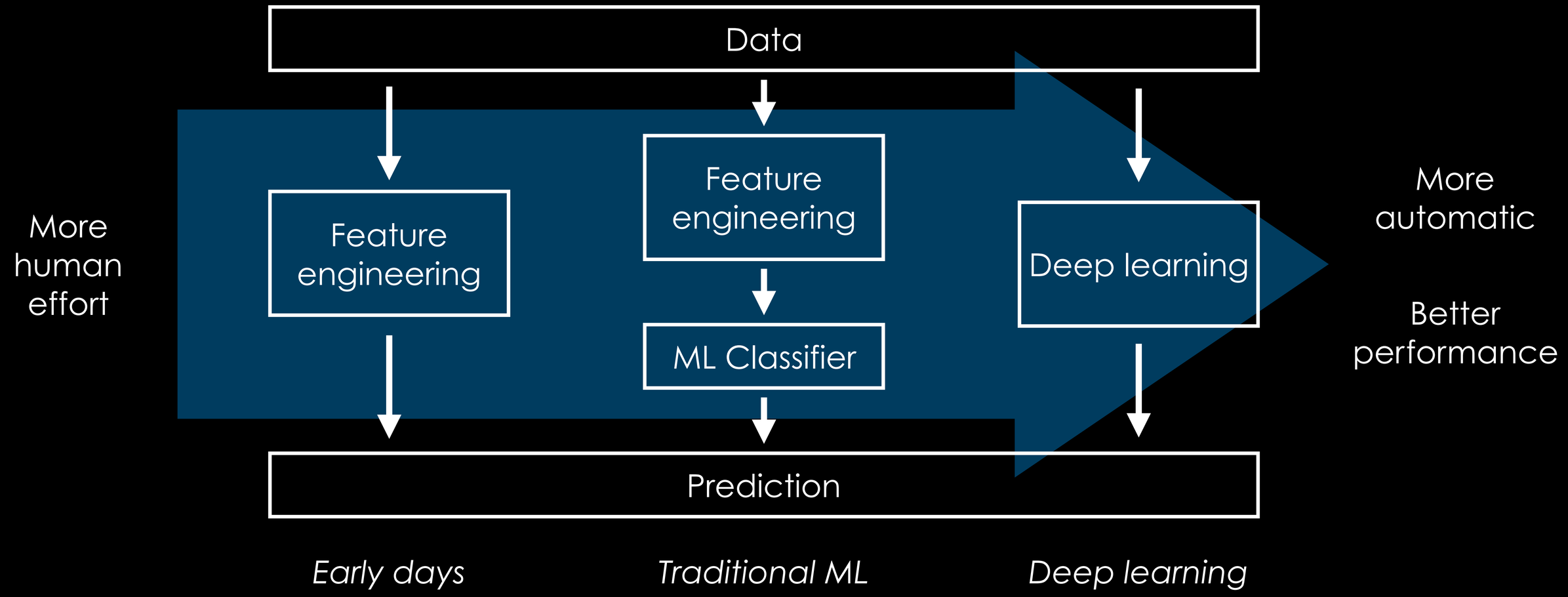
High heels

(Source: BBC)

Easy for Humans

Hard for Machines

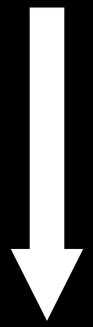
Evolution of data processing technique



Domain knowledge

To tackle each topic

(make some “rules”)

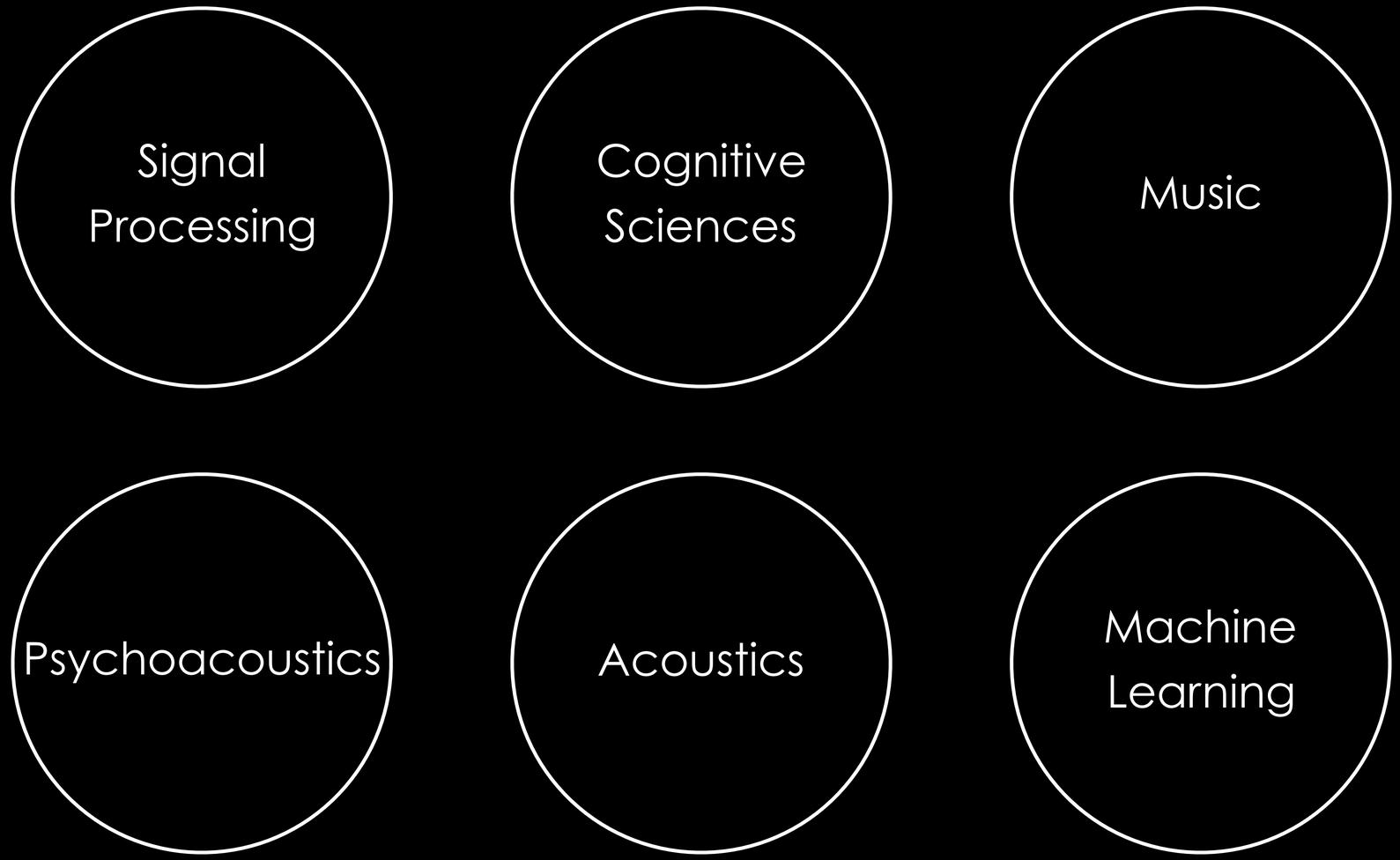


To simulate how

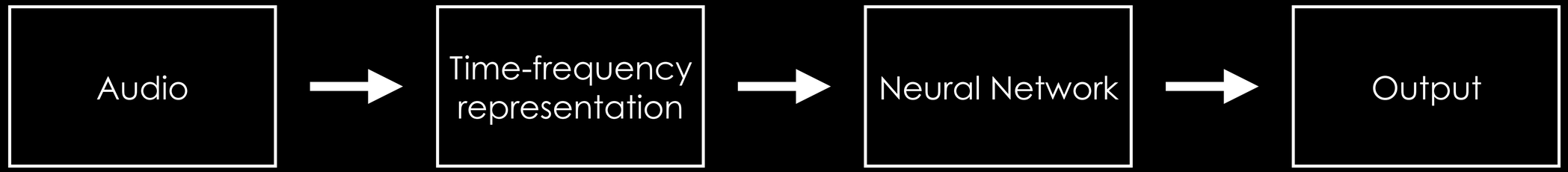
human understand the sound

(and prepare data)

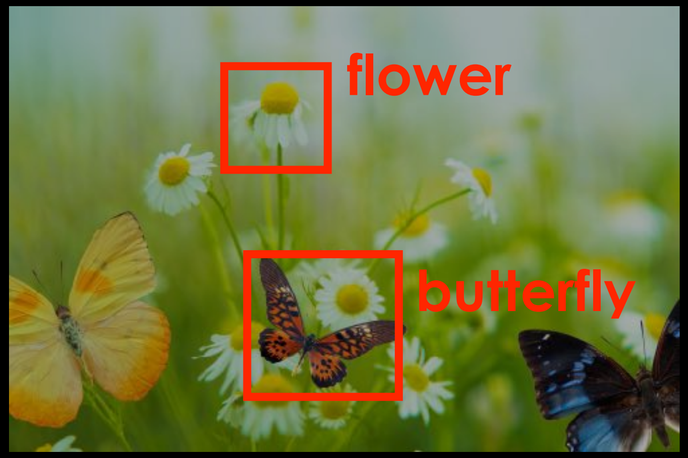
Required domain knowledge

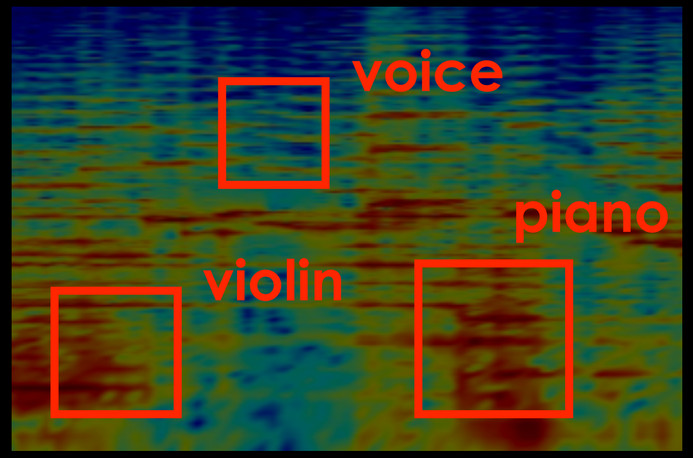


“Modern” audio identification pipeline



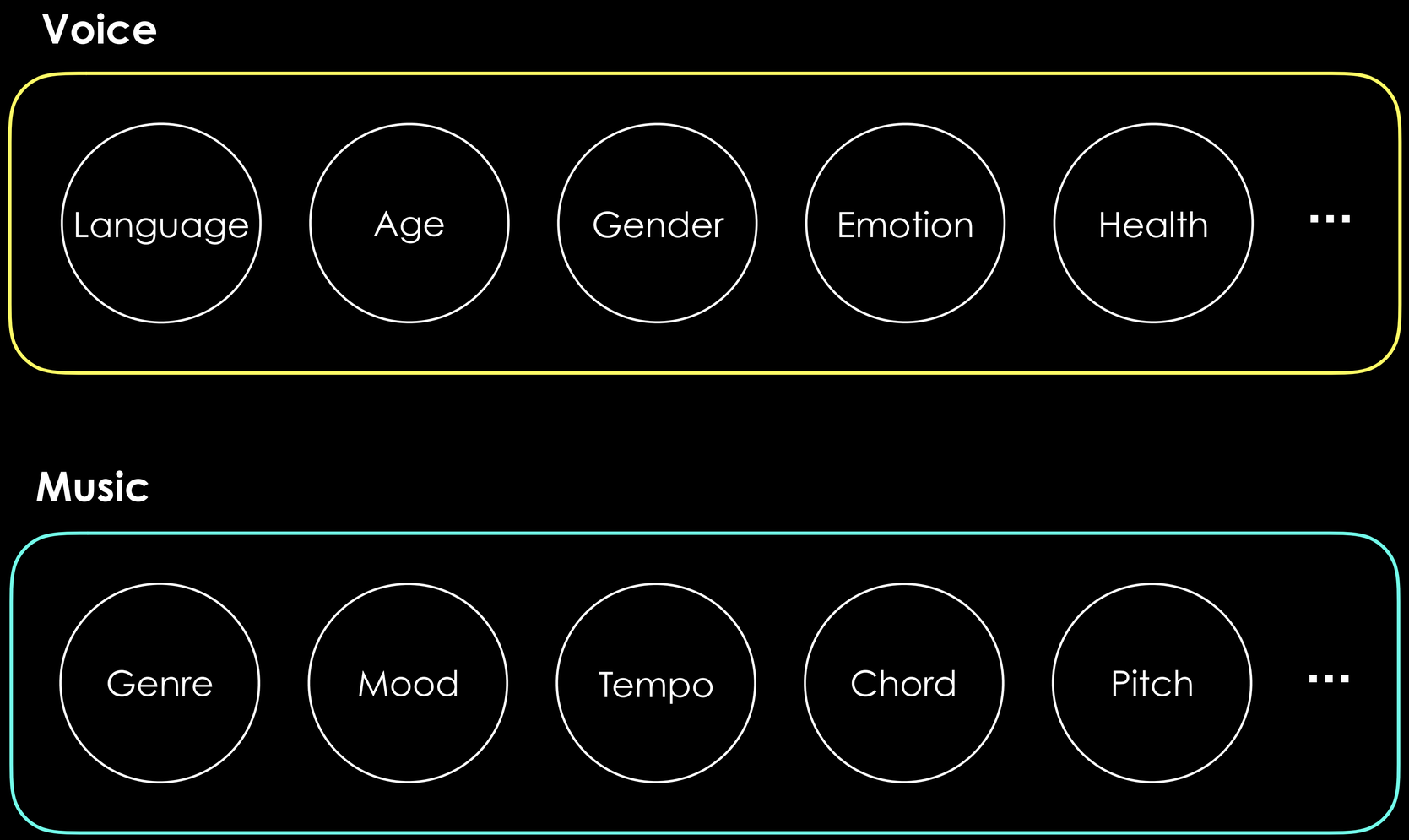
objects in an image ≈ instruments in a spectrogram





“**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
Language
Age
Gender
Emotion
Health
...
Music
Genre
Mood
Tempo
Chord
Pitch
...

Machine listening

Acoustic scenes

bus	park
library	city centre
train	driving
home	market
cafe	…

Music

Voice

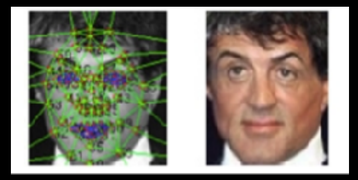
Acoustic events

glass break	knock
car horn	dog bark
footstep	water boil
gun shot	snoring
bird chirping	crying
sneeze	…

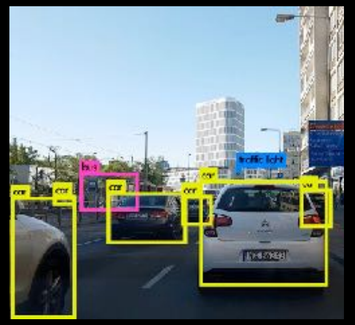
“Any” sound we hear everyday

Computer vision

Optical Character Recognition (OCR)



Facial recognition



Object detection

Machine listening

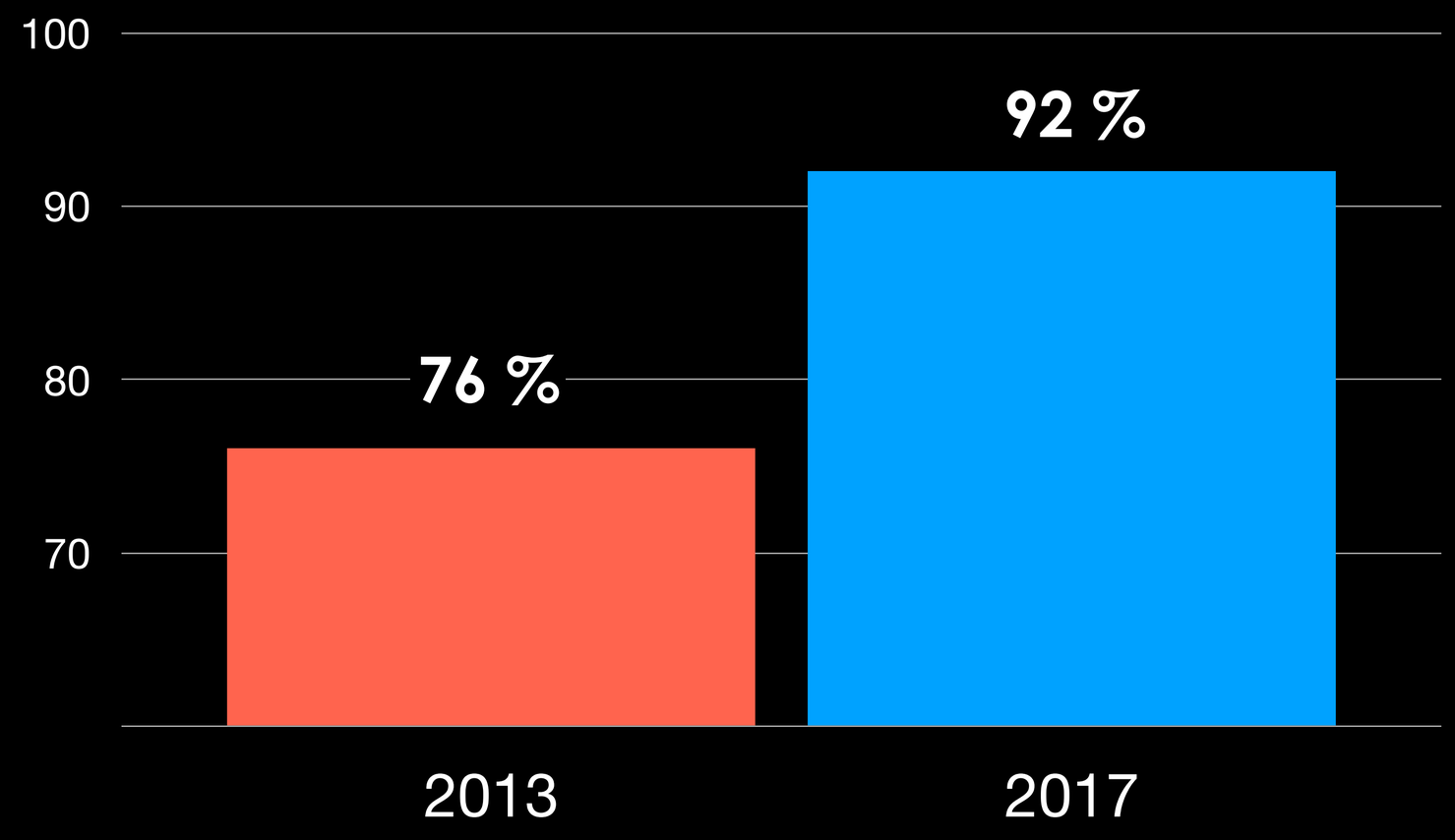
Voice recognition

Music search

Speaker identification

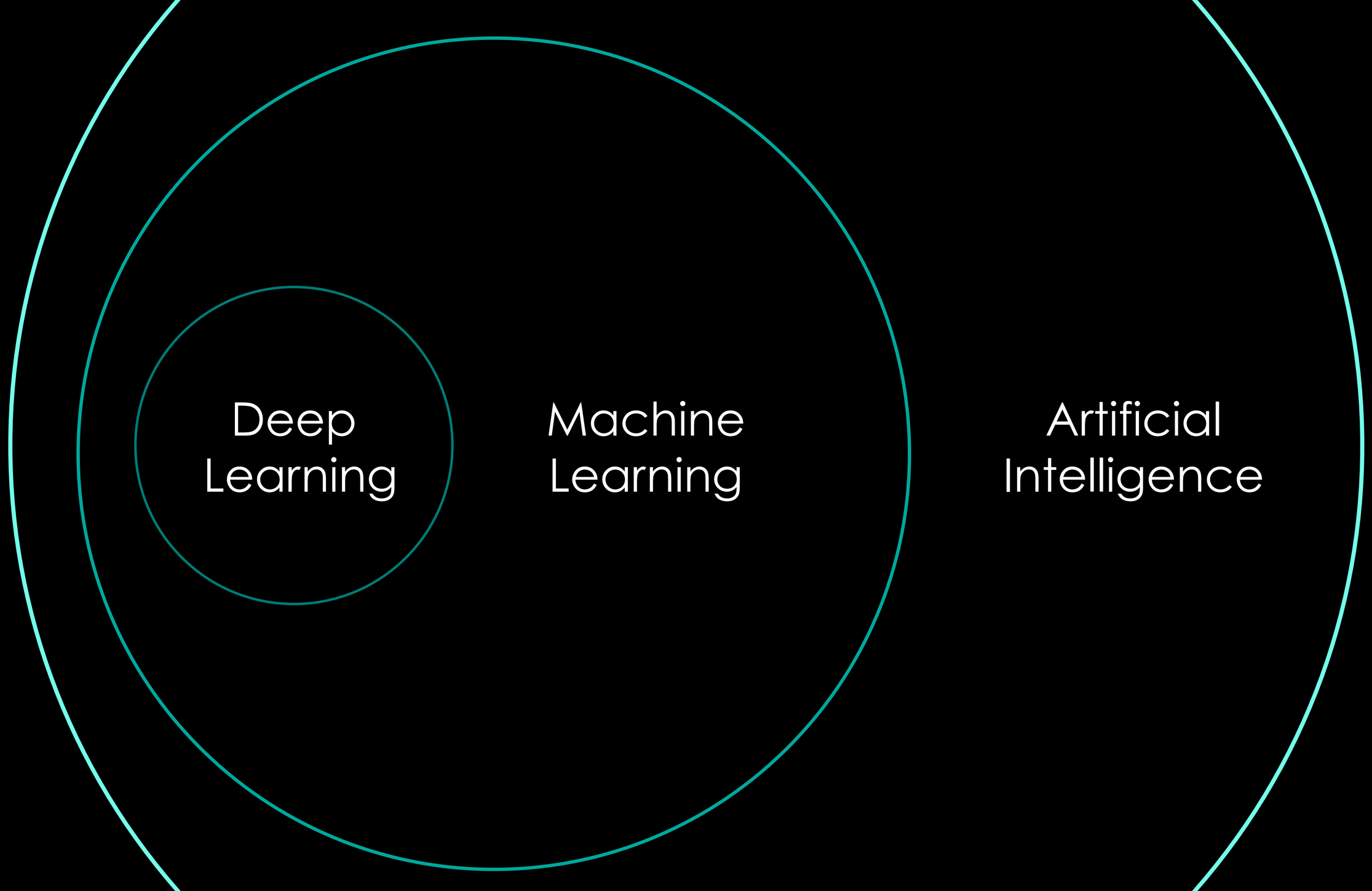
Acoustic scene/event detection

(Sources: Tensorflow, Facebook , Microsoft, Apple, Shazam)



Scene classification accuracy (IEEE DCASE)

(Source: http://www.cs.tut.fi/sgn/arg/dcase2017/, http://c4dm.eecs.qmul.ac.uk/sceneseventschallenge/resultsSC.html)


Deep
Learning
Machine
Learning
Artificial
Intelligence


Perceive
Think
Act

Five, Zero

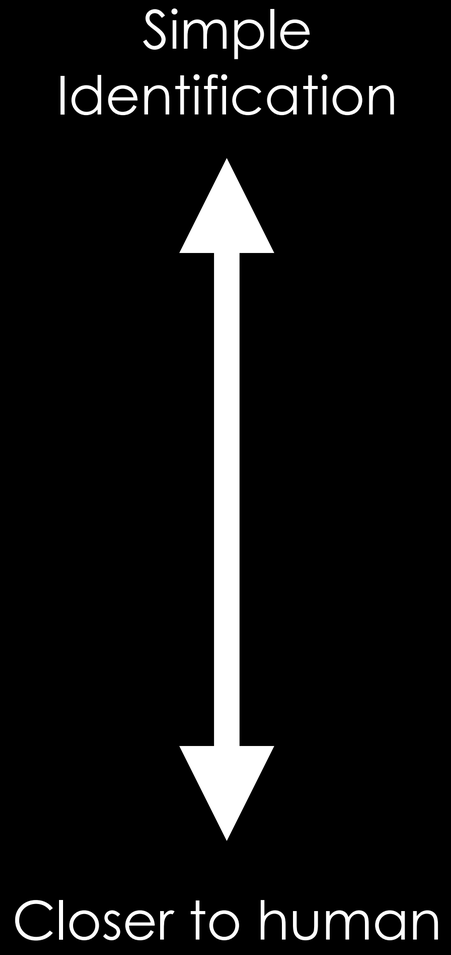
Cat

Know what it is (with input restriction)

Know what it is

Know what/where it is

Know what/where it is + why



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

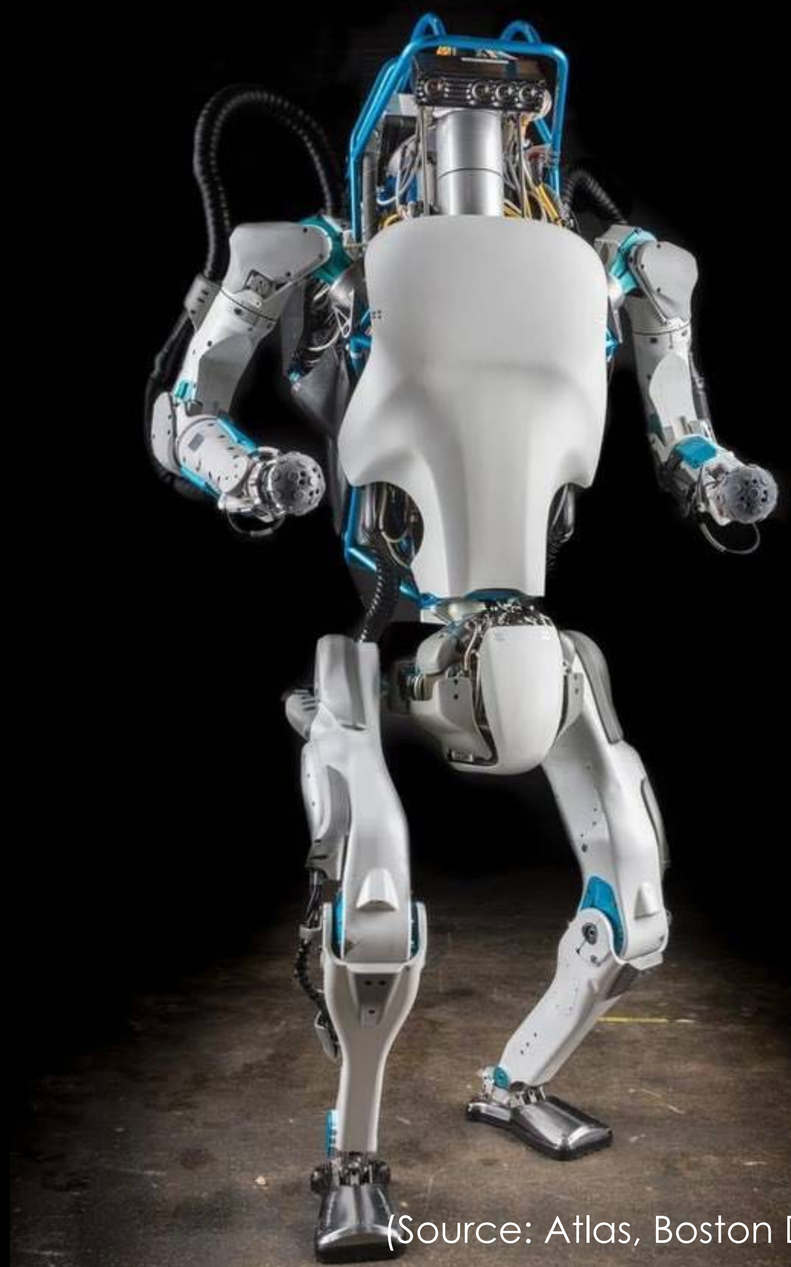
See things

Understand speech

+

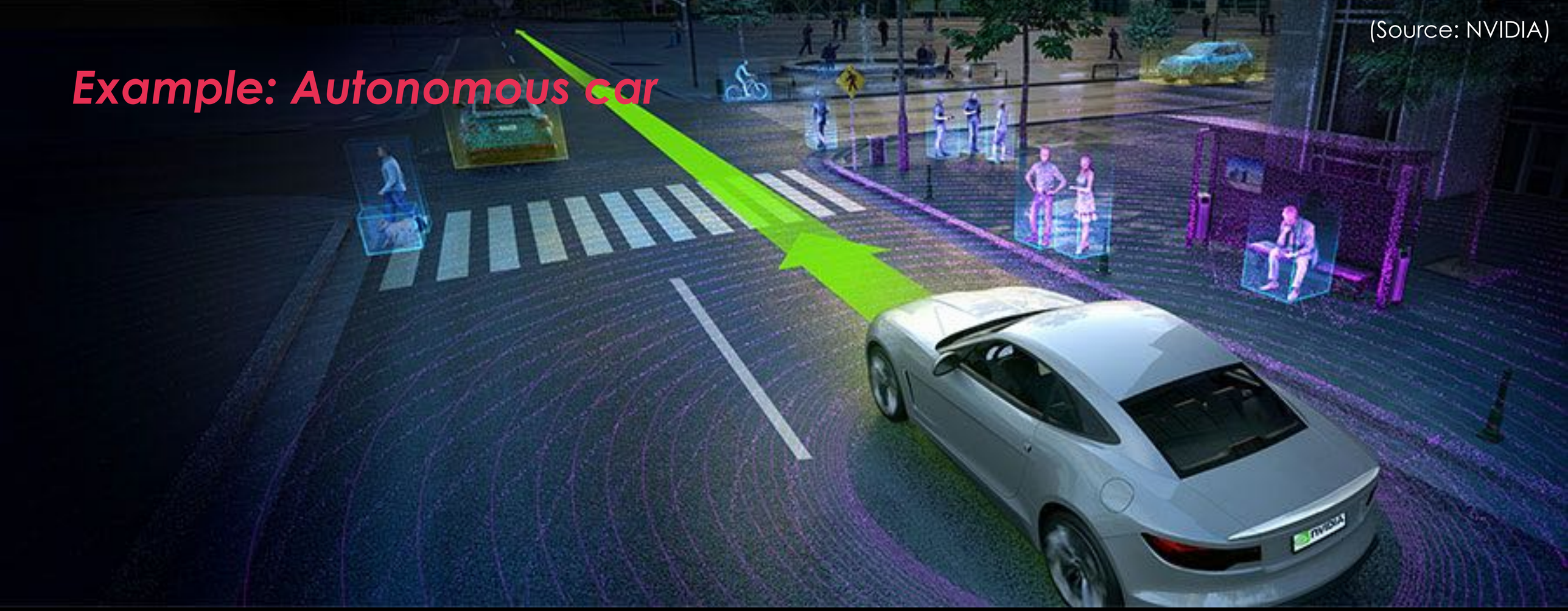
Listen things other than voice

Know who they talk to



(Source: Atlas, Boston Dynamics)

Example: Autonomous car



(Source: NVIDIA)

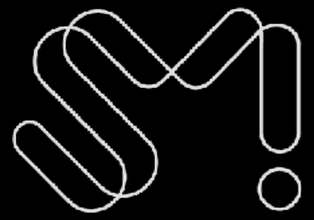
Outside - Car horn (normal, air horn), Siren (fire truck, police, ambulance)

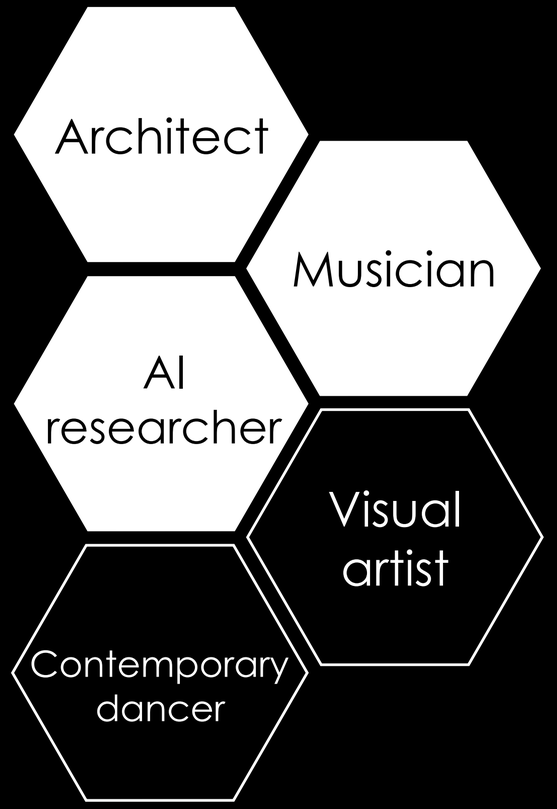
Inside - Music mood, snoring, baby, anomaly detection (malfunction warning)

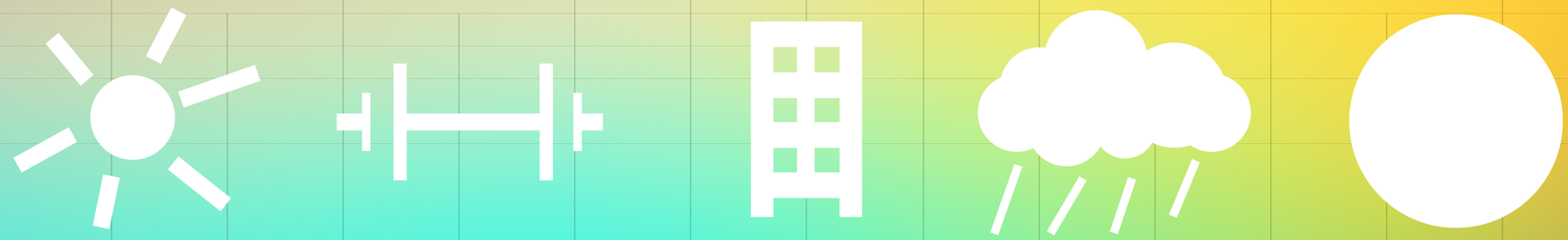
ATMO: Generative music for spatial atmo-sphere

Create NeWave
kocca
한국콘텐츠진흥원

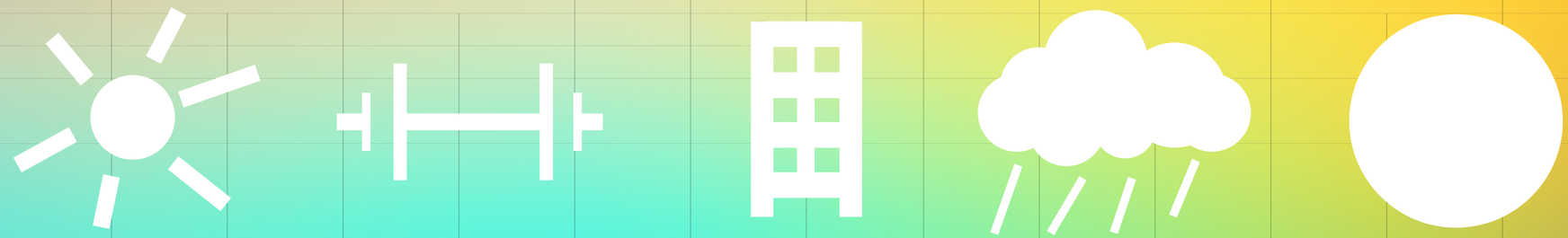
+





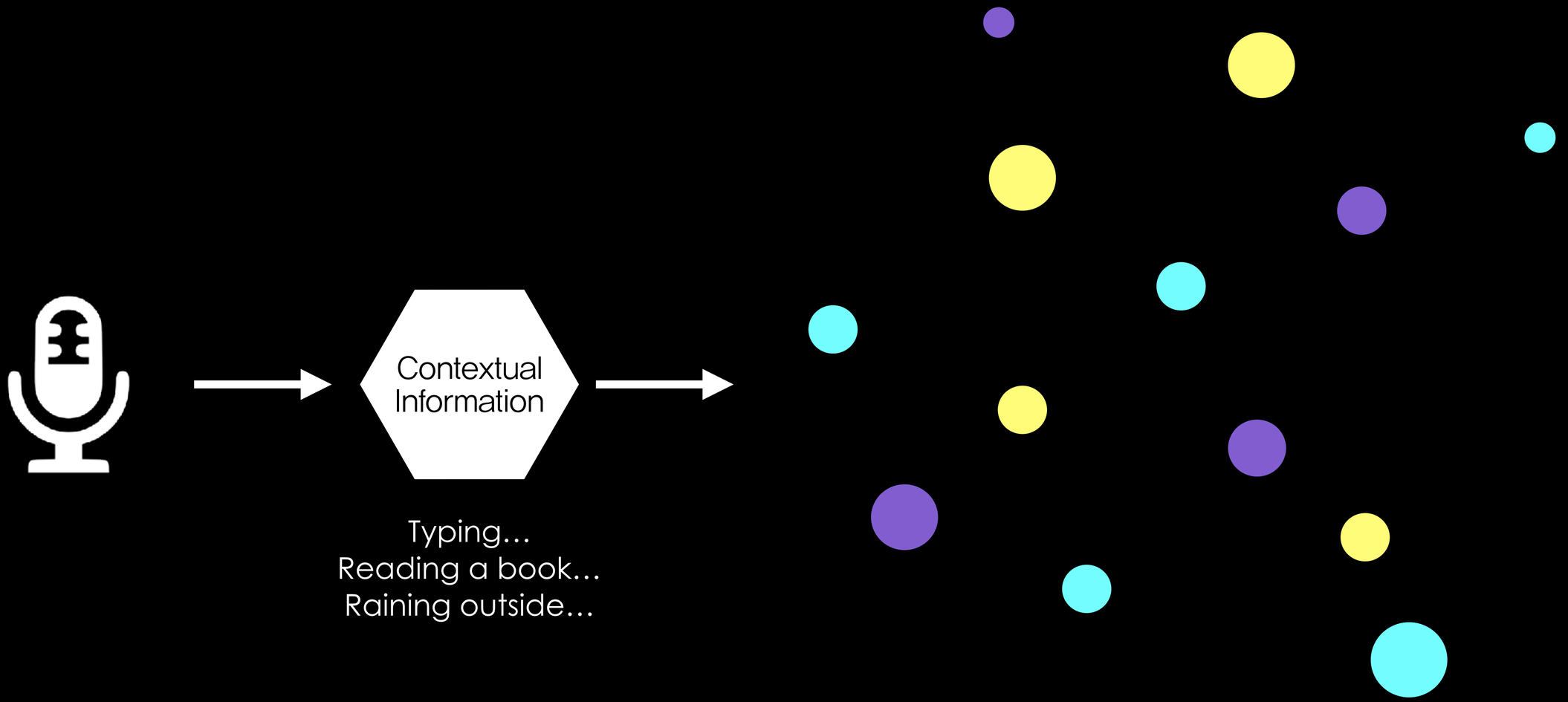


Generative Music with contextual information



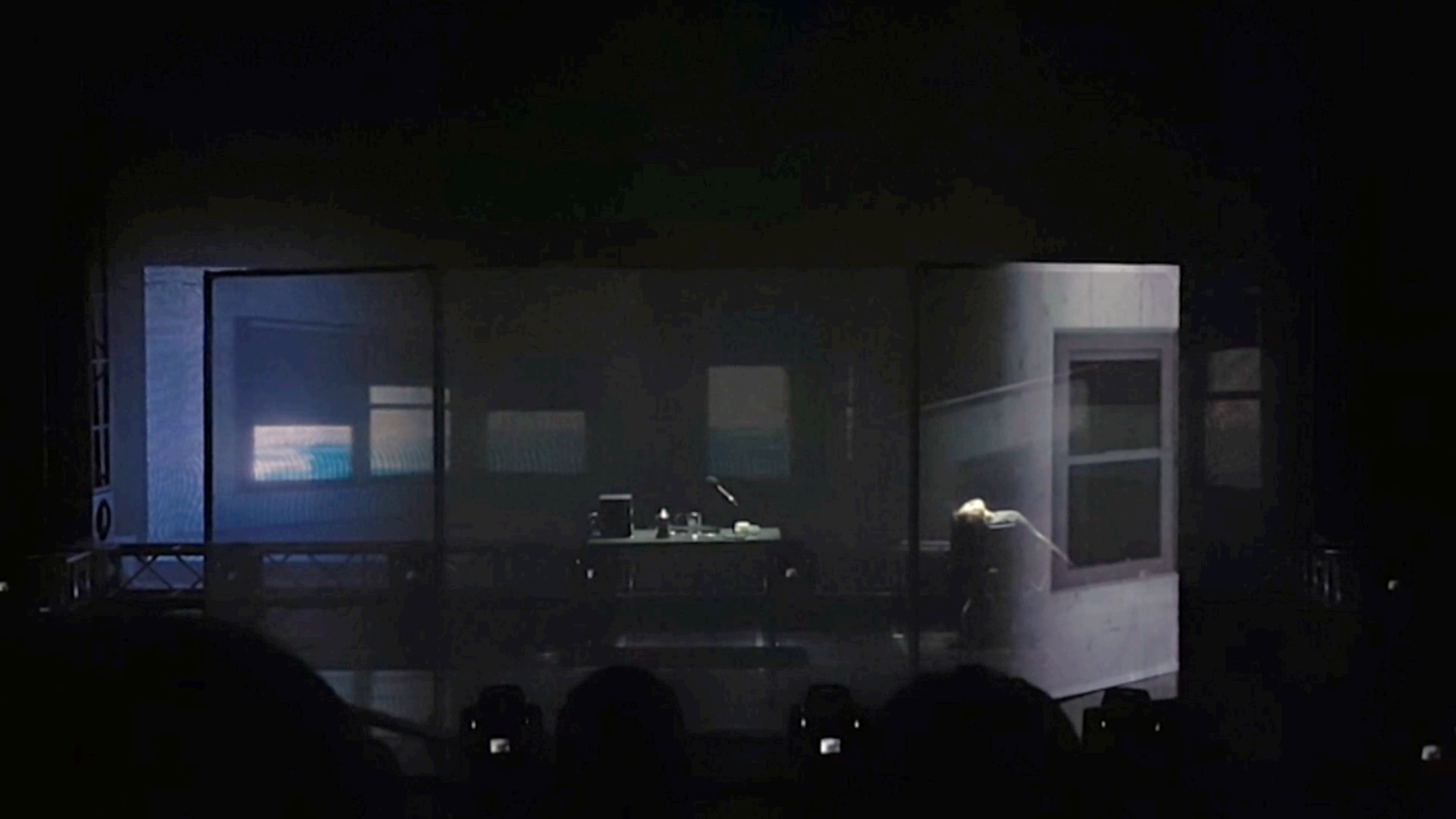
Generative Music with contextual information

Analysis Result : Typing in a rainy day…




Microphone
Speaker


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



cochlear.ai

contact@cochlear.ai