A New AI Platform Architecture for the Smart Toys of the Future

Gabriel Costache
Senior R&D Director
XPERI

Portfolio of Trusted Brands

dts	HD Radio	IMAX ENHANCED	invensas	TESSERA	TiVo
Sound, Connected Media, Imaging, and Computer Vision Solutions	Digital Broadcast Radio and Connected Car Services	Creation and Delivery of Immersive Cinematic Entertainment	Semiconductor and Interconnect Packaging Technology and Solutions	Semiconductor Intellectual Property Licensing	Entertainment Software, Consumer Hardware, Data and Advertising Solutions

40+ offices worldwide headquarters in San Jose, CA

$1.5B + market cap public company, trading under XPER

1,600 + employees worldwide

1,500 + engineers

11,000 + patent assets

100B+ devices worldwide empowered by technologies delivered via Xperi brands

Ideal Smart Toy

- Safe
- Secure
- Private
- Enhances child development
- Uses natural interaction
- Monitors child cognitive load
- Develops with the child
- Long battery life
- Re-usable

Smart Toy Examples

Privacy Issues

German parents told to destroy Cayla dolls over hacking fears

17 February 2017 Share

The My Friend Cayla doll has been shown in the past to be hackable

An official watchdog in Germany has told parents to destroy a talking doll called Cayla because its smart technology can reveal personal data.

The New York Times

Mattel Pulls Aristotle Children's Device After Privacy Concerns

aristotle

c|net BEST PRODUCTS REVIEWS NEWS VIDEO HOW TO SMART HOME CARS DEALS JOIN

SECURITY | LEER EN ESPAÑOL

Hello Barbie: She's just insecure

Researchers revealed new flaws in the doll Friday, adding to problems publicized last week. Hello Barbie's software maker is racing to patch security bugs during the holiday shopping season.

BY LAURA HAUTALA | DECEMBER 4, 2015 3:00 AM PST

Play Sound

XPERI

Smart Toy Challenges

- Data privacy
- Safety
- Battery life
- Fast response
- AI technologies for children
- Data bias in AI
- Natural interaction with children
  - Multimodal: audio, imaging, sensing

DTIF (Disruptive Technology Innovation Fund)

XPERI

SoapBox

NUI Galway
OÉ Gaillimh

Project Ireland 2040
Building Ireland's Future
Disruptive Technologies Innovation Fund

D.A.V.I.D

DAVID – Data-center Audio/Video Intelligence on Device

DAVID will develop a "privacy by design" AI platform, capable of multi-modal, ultra-low power consumption, "data center" level processing of audio and vision data on-device, without the need to transmit any personal data to the cloud.

What DAVID will deliver to the smart toy market:

- A platform for a wide range of learning and interactive applications in the toy market
- A smart, trusted proof-of-concept toy using this platform that helps children learn and develop, using XPERI imaging technology, Perceive® Ergo® chip and SoapBox Labs speech technology capabilities in collaboration with the National University of Ireland, Galway.
- Cloud-free capabilities to ensure privacy and wonderfully immersive user experiences for children of all abilities.

All-in-one Chip/Platform

Designed for Privacy

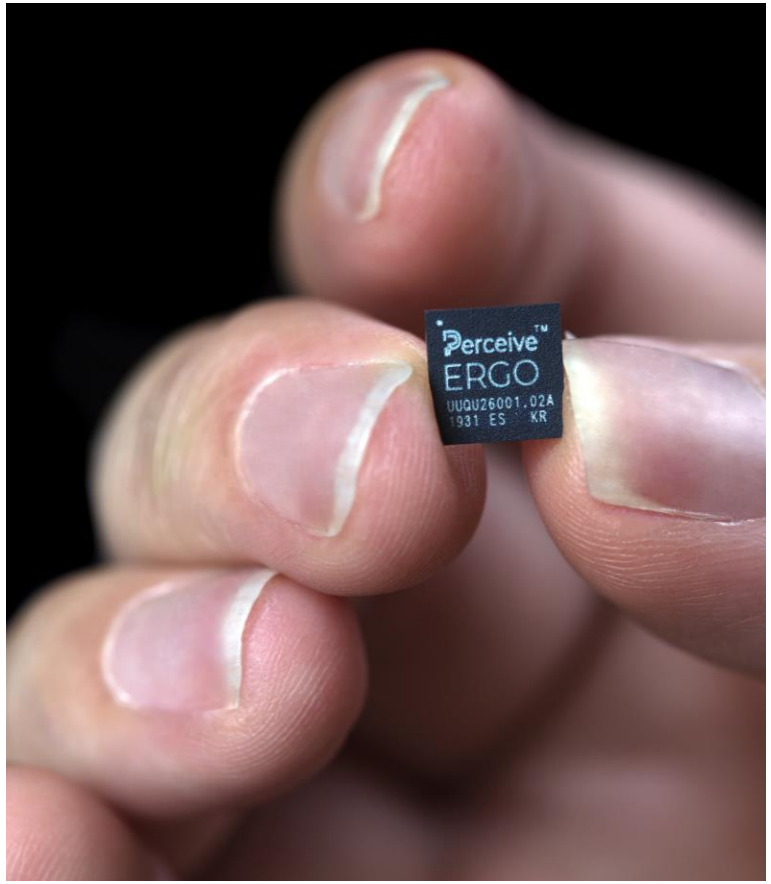
Multi-modal Platform Communication

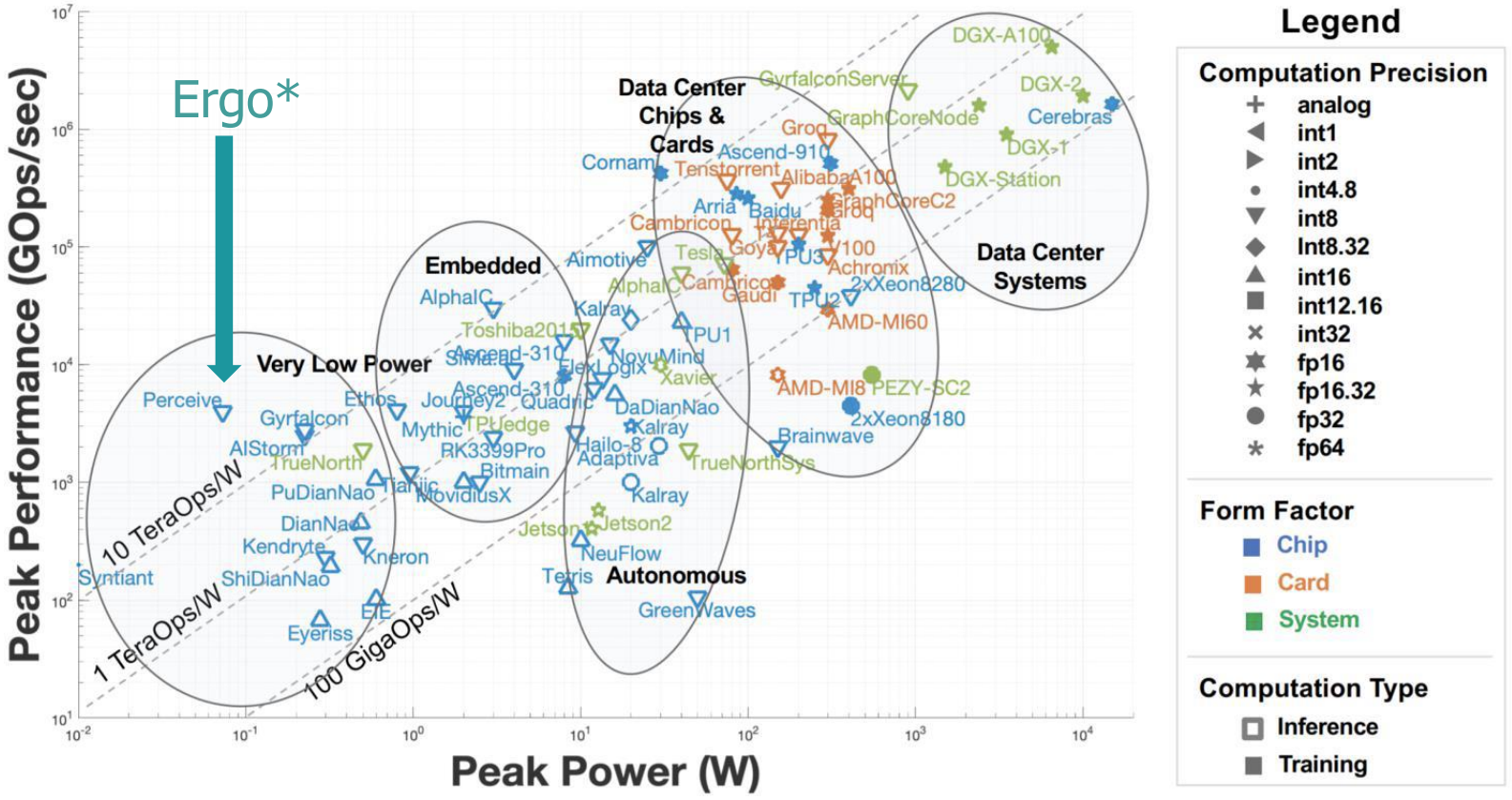
Speech, Expressions, Emotions, Gesture, Context and others..

AI Technologies to be Considered

- Perception
  - Imaging/Vision
    - Face Analytics
    - Body Analytics
    - Hand Analytics
    - Video Compression
    - Thermal Imaging
  - Audio
    - Wake Words / VAD
    - Speech2Text / ASR
    - Voice Analytics / Biometrics
  - Sensing
- Interaction
  - Visual
  - Audio
    - Text2Speech
    - Sound Generation
- Others
  - Language Models / Conversational Models
  - Multi Modal Intent
  - Cognitive and Behaviour Analysis
  - Personalization
  - Interactive Games

Perceive® Ergo® AI Processor





Legend

Computation Precision
- analog
- int1
- int2
- int4.8
- int8
- Int8.32
- int16
- int12.16
- int32
- fp16
- fp16.32
- fp32
- fp64

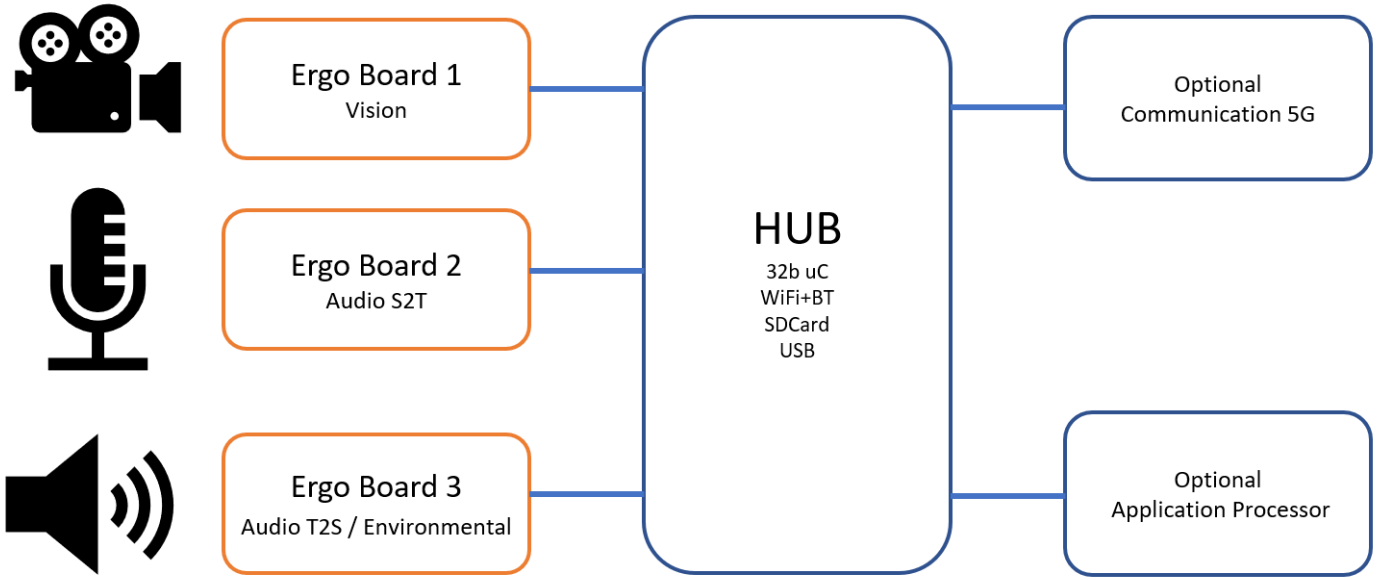
Form Factor
- Chip
- Card
- System

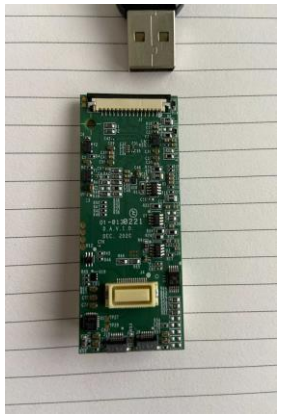
Computation Type
- Inference
- Training

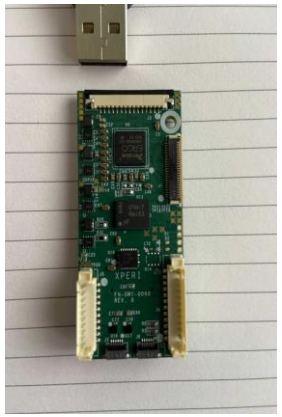
*Note: Ergo uses a proprietary representation. Ergo is not INT8.

Source: A Reuther et al. MIT Lincoln Laboratory Supercomputing Center-arXiv:2009.00993

DAVID Platform Design







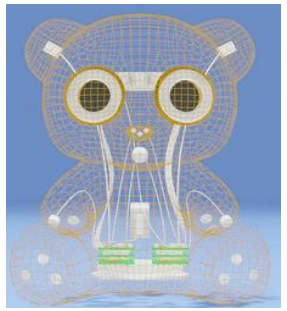
DAVID Platform Specifications

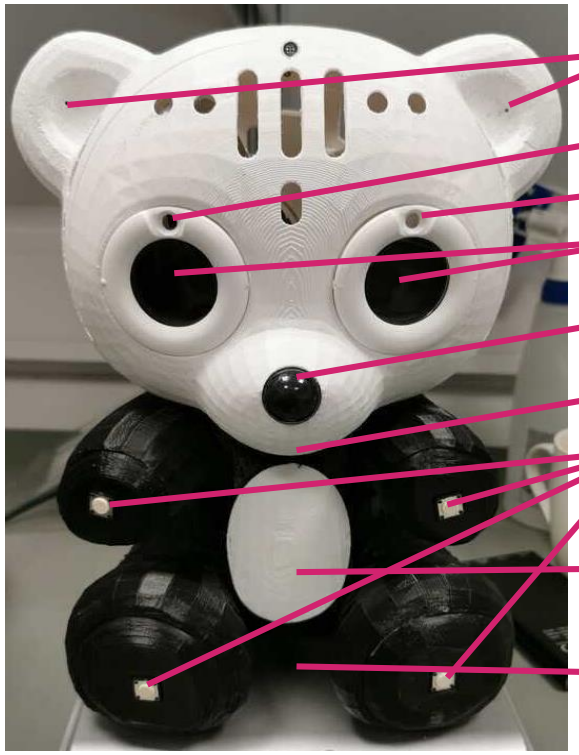
- **Interfaces:**
  - I2S (Tx, Rx), I2C (Tx, Rx) – (HUB and Ergo)
  - MIPI and Parallel (Ergo)
  - SPI & QSPI (HUB & Ergo)
  - GPIO (HUB and Ergo)
  - FTDI (JTAG, UART) (HUB)
  - WiFi/BT (HUB)
  - USB OTG (HUB)

- **Computation Units:**
  - 3 x Ergo (55 TOPS/Watt + Arc CPU\DSP)
  - HUB STM32 MCU (Arm M7)
  - ESP32 (2x Xtensa LX6)

- **Memory:**
  - 16MB QSPI Flash (Ergo)
  - 128MB QSPI Flash + 32MB SRAM (HUB)
  - 448 KB ROM + 520 KB SRAM (ESP32)
  - SDCard (HUB)

DAVID Toy PoC

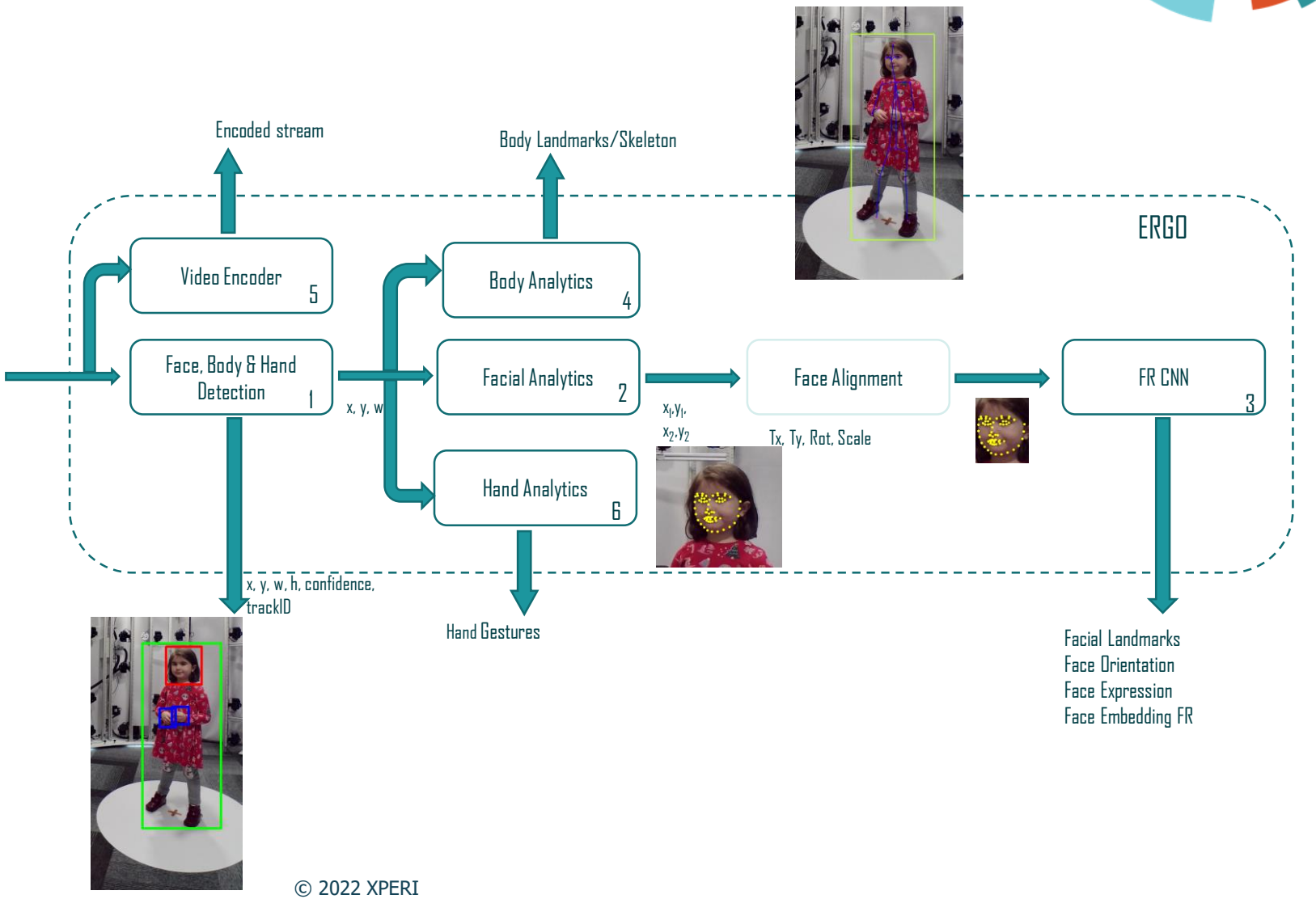



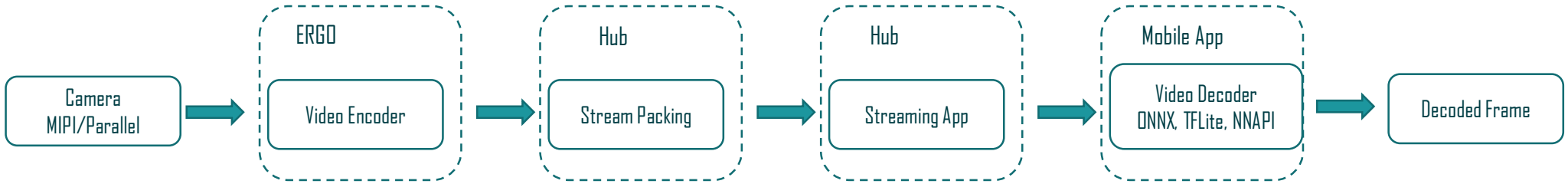
Current Ergo Vision Application

Example Ergo Application

- Frame rate 30 fps
- Resolution 320x320
- Power ~100 mW



Video Encoding



Fully neural video encoder (Ergo) and decoder (generic)

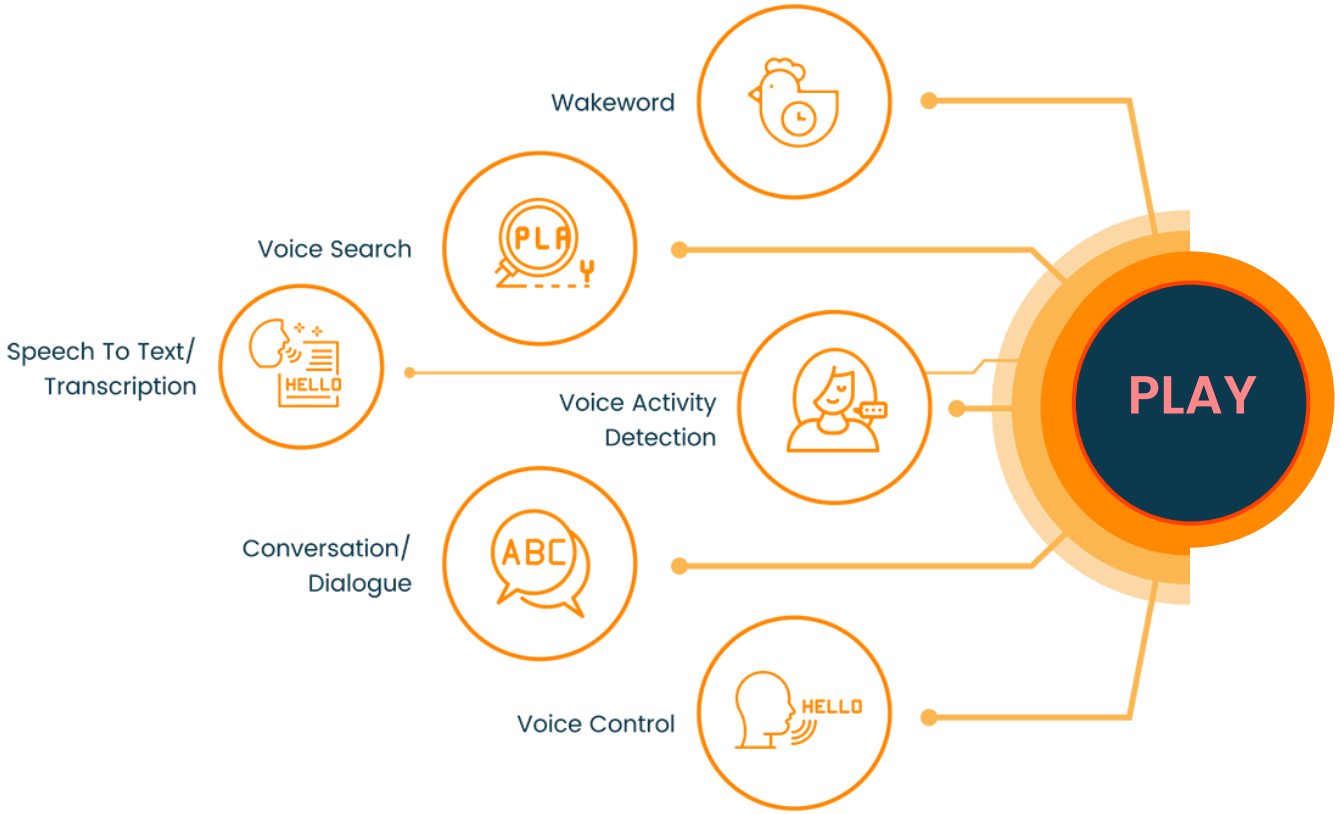
- Trained end-to-end
- Custom stream – data privacy
- Extra security can be added
- Y only currently but can be easily extended to color
- Enabler of other image enhancement technologies: colorization, super resolution
- Can enable smart monitoring

Speech/Audio Neural Synthesis

- Current Ergo board 3 application Text2Speech -> spectrogram generation + vocoder
  - Focus on comprehension, less on naturalness
- Next focus on: voice adaptation, voice cloning
- Extend to sound/music generation

DAVID Partners

SoapBox powers magical and joyful experiences for kids using speech technology that is engaging, fun, and frictionless.



NUIG C3I - Center for Computational, Cognitive & Connected Imaging

Conclusions

- Smart Toy requirements:
  - Privacy
  - Battery life
  - Multimodal interaction
- Platform requirements:
  - Dedicated NN unit with very high OPs/W
  - Communication unit
  - Multiple sensor support
  - Generic processing unit
- DAVID platform and toy PoC
  - Available Q3/Q4 2022 for selected partners

Resources

XPERI®

Resources

- Xperi – www.Xperi.com
- Perceive, Ergo – www.perceive.io
- SoapBox Labs – www.soapboxlabs.com
- C3I, National University of Ireland, Galway - www.nuigalway.ie/c3i
- Disruptive Technologies Innovation Fund – DTIF
- STMicroelectronics STM32 MCU
- Espressif Systems ESP32