Designing for AI Enabled Audio IoT: A Case for Performing at the Edge

Jim Steele, VP Technology Strategy, Knowles Intelligent Audio

knowles

Enabling Advanced Audio Solutions

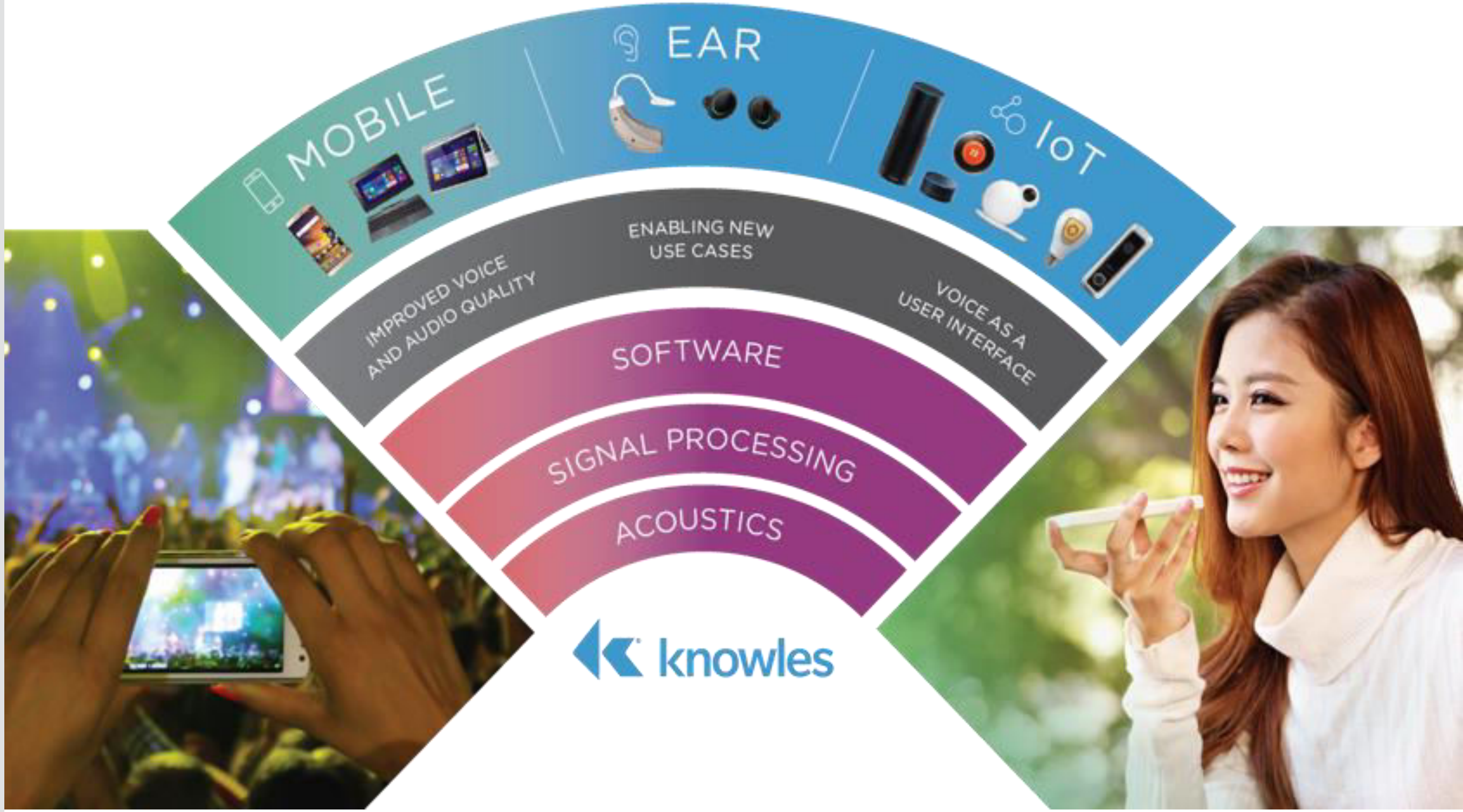
Knowles creates audio solutions that differentiate our customers

OUR OFFERING

- Higher-performance microphones, advanced multi-mic integration
- DSP, smart mics, audio processors, audio software algorithms, tools, and system solutions

KNOWLES ADVANTAGE

- Cross-functional expertise from acoustics to audio algorithms
- Microphone performance leadership and scale
- Over 10B microphones shipped



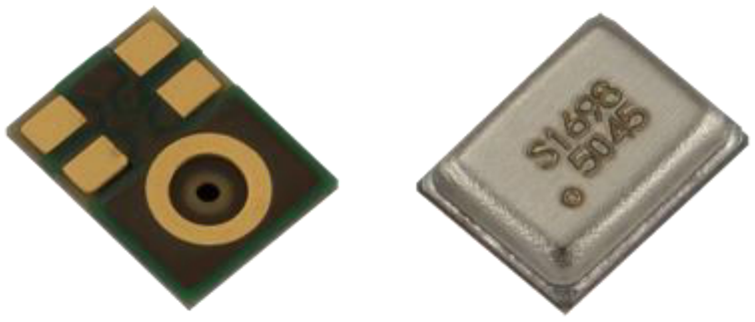
Knowles:
Market Leading Acoustic Supplier

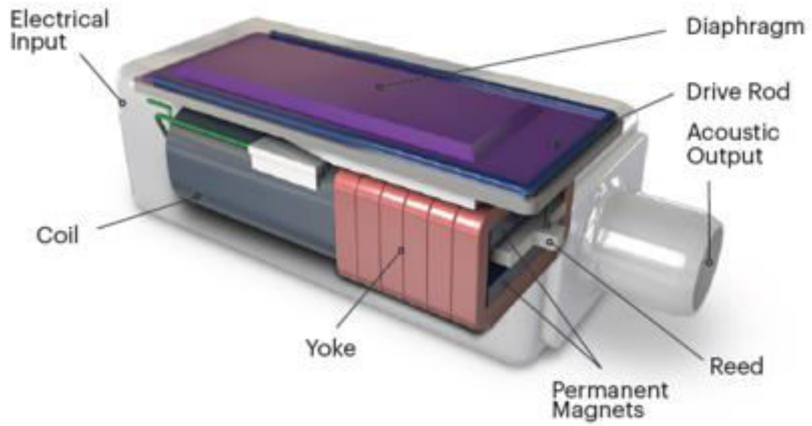
#1 Global Supplier

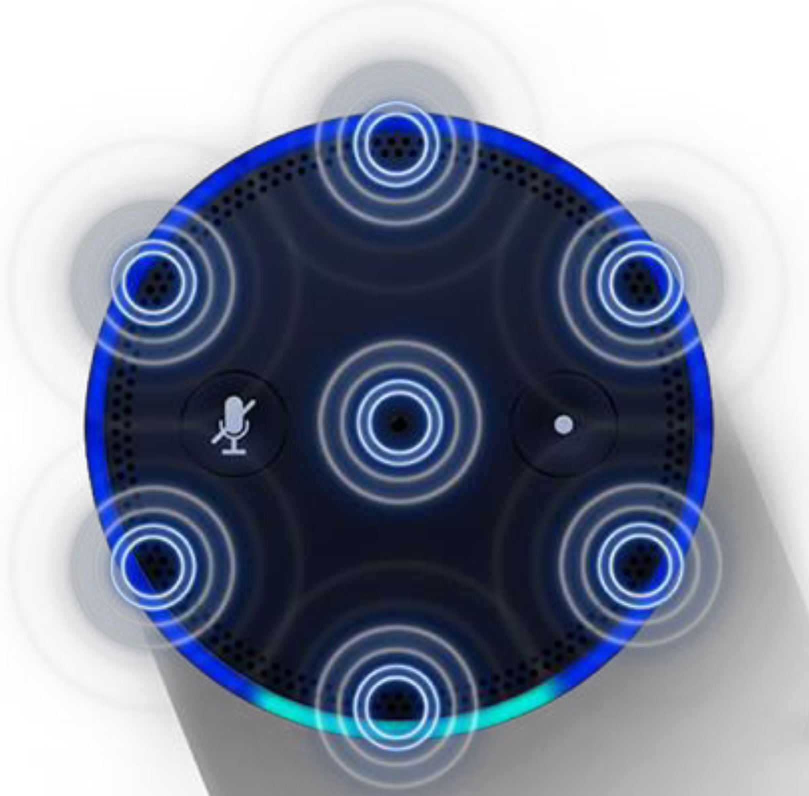
- MEMS microphones
- Hearing aid solutions

Strong Engineering Partnerships

- Smartphones
- Laptops
- Tablets
- Smart home and IoT devices
- Premium earphones
- Smartwatches







Progression of User Interfaces: IoT needs Voice

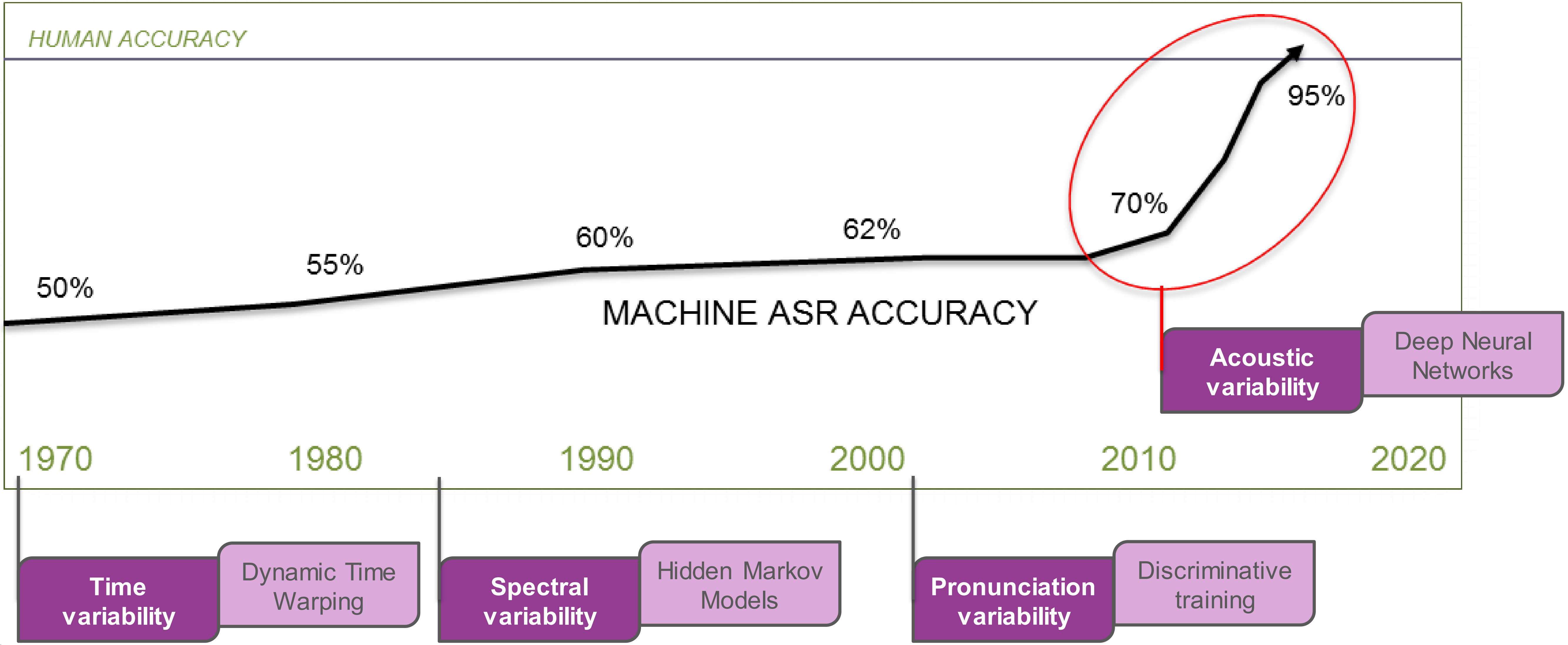
Websites — Point and Click

Mobile Apps — Touch

IoT — Voice

Why Voice Interfaces are Becoming Prevalent Now?

ASR = Automatic Speech Recognitio



knowles

The Future of IoT is Edge Processing

IoT Common Precepts

Principle: extracting and analyzing digital data from the physical world
Characteristics: combination of hardware and software
Opportunities: personalization and intelligence, real-time services
Challenges: connectivity, security, power

Migration to the Edge

"50% of all local data will be processed outside the cloud by 2022"

according to Gartner

> 42 billion
connected devices

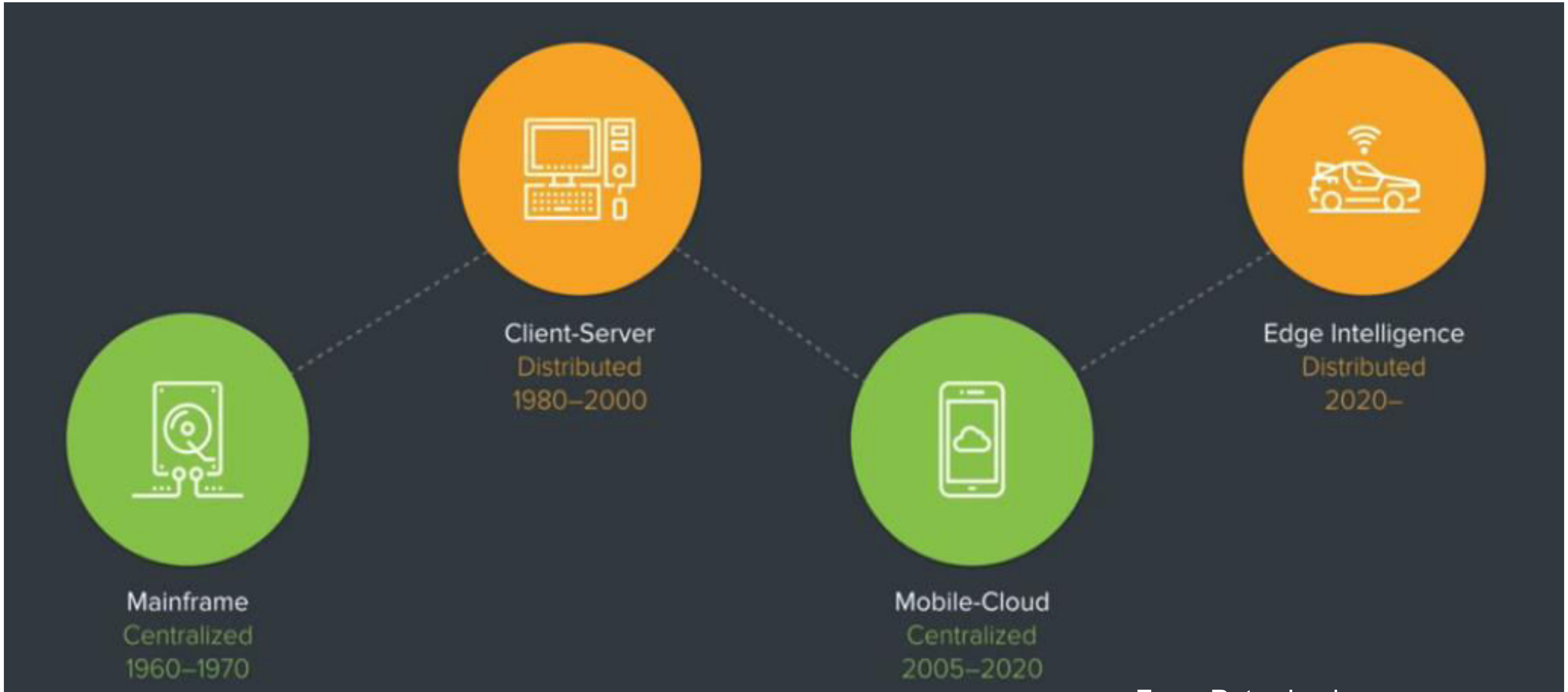
Pieces of a Conversational AI Platform

Next weekend
New York City
Hotel Booking

"Book a Hotel in NYC"

Content Providers and Intent Execution	Intent Engine Natural Language Understanding (NLU)
Speech Engines	Automatic Speech Recognition (ASR) / Text to Speech (TTS)
Microphones and Signal Processing	Acoustic Echo Cancellation (AEC) Noise Suppression Continuous Voicewake Always-on Voicewake

With IoT, distributed computing is on the rise, and will be even larger than mobile

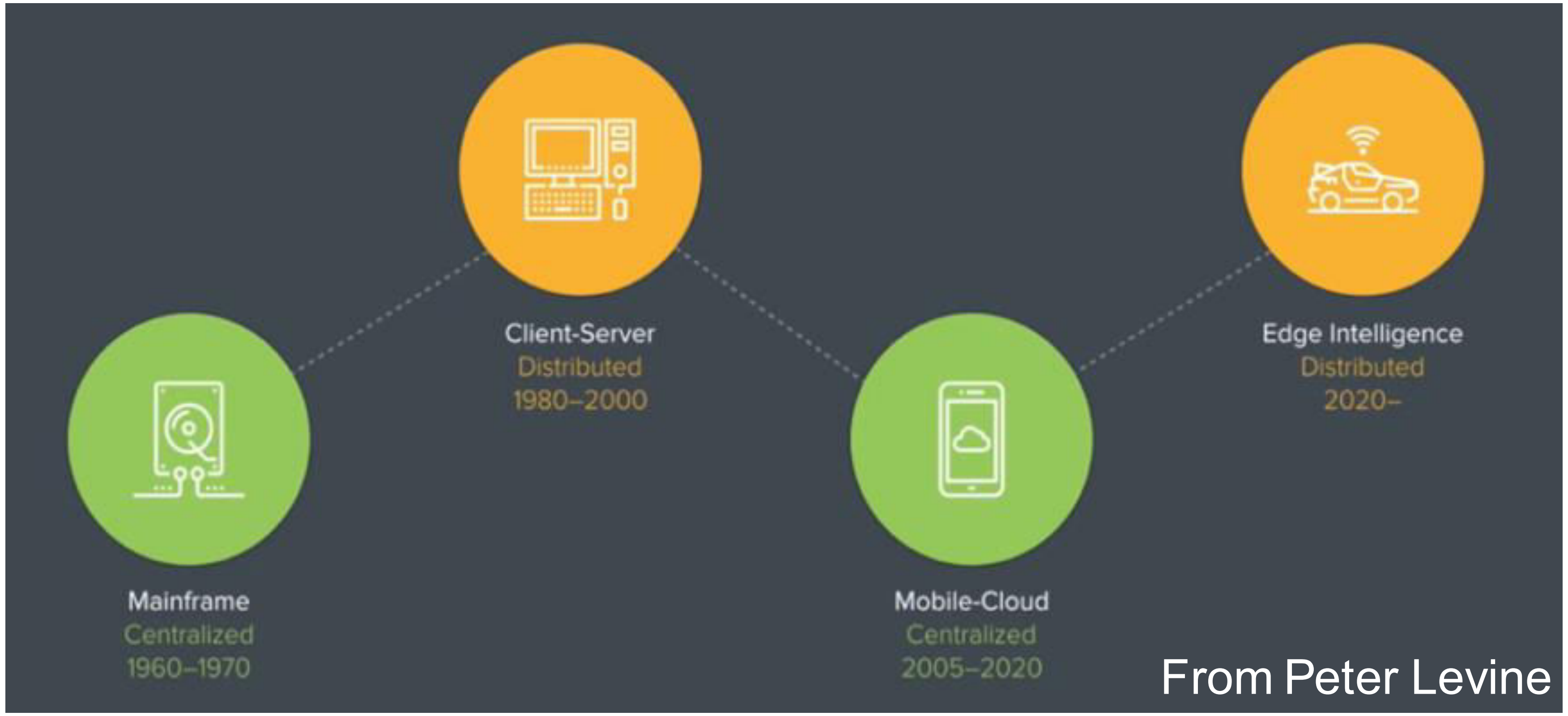


From Peter Levine

# units worldwide	Thousands	Millions	Billions	Trillions
Main players HW-Processor-SW	IBM-DEC-DEC	Dell-Intel-Microsoft	Apple-Qualcomm-Google	?

knowles

We will see amazing edge devices with new sensors & processors



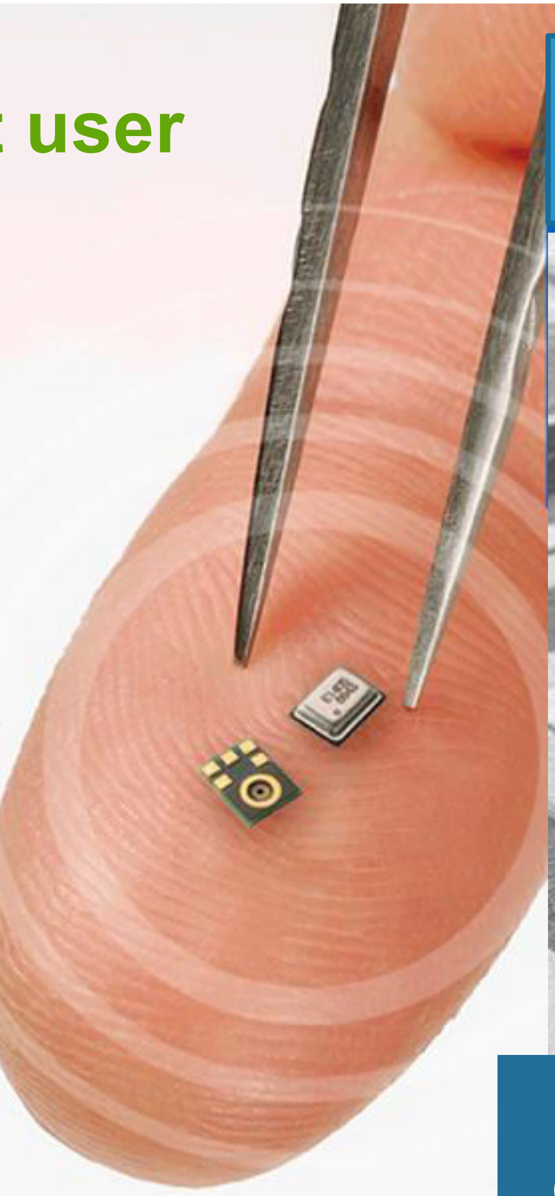
From Peter Levine

Why edge?	Why DSP?
Always-on	Low power
Real-time	Low latency
Personalized	Privacy
Proliferated	Cost
Data-driven	Machine learning

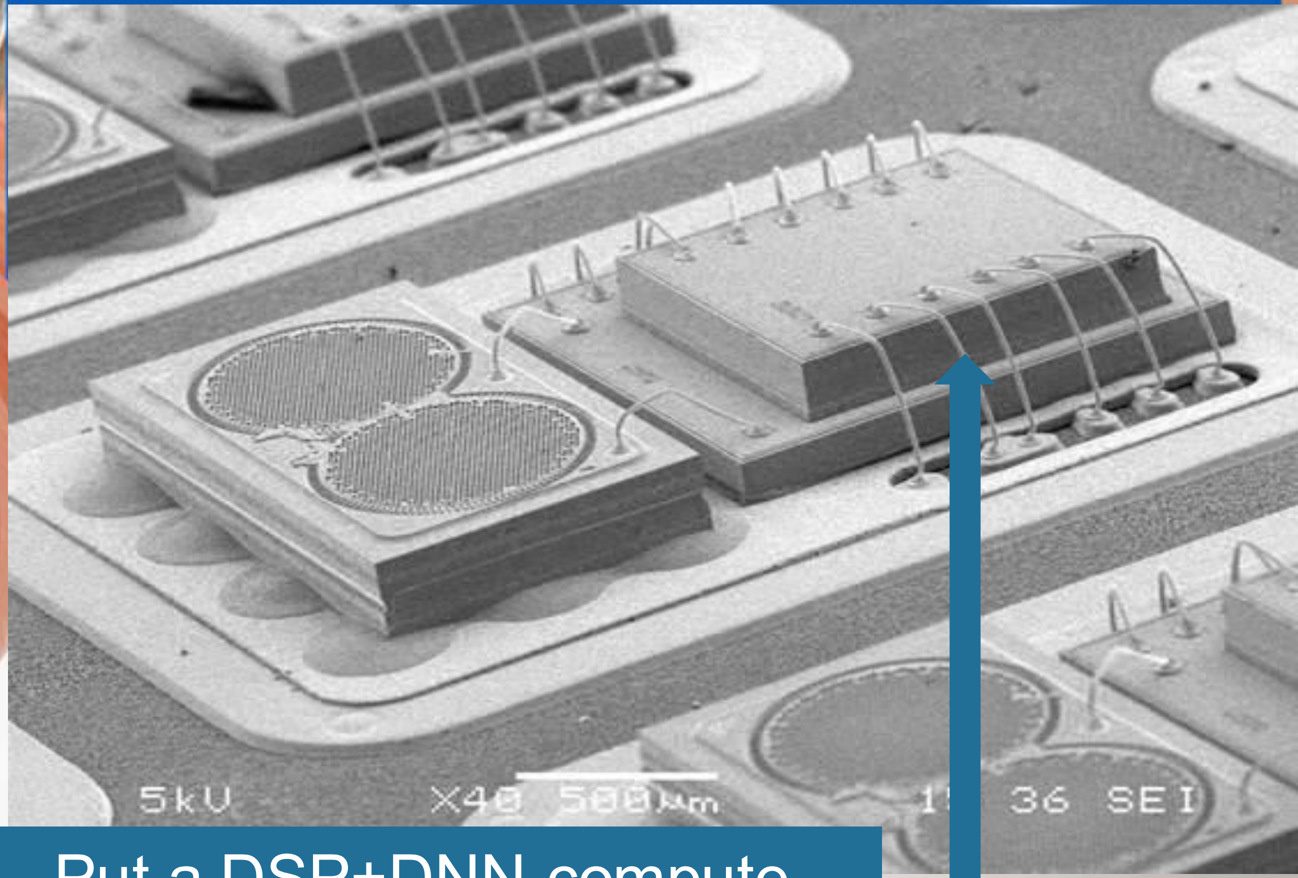
knowles

Moving from Cloud to Edge: Always On → Low power

- **Always-On provides the richest user experience**
  - Optimize for the sensor and data
  - Minimize data transfer
  - Lowest power consumption
  - Minimal hardware used
- **Going to the cloud takes power**
  - Pulling data off sensors
  - Packetizing, sending, retries
  - Waiting for response



Knowles AISonic™
IA610 Smart Microphone



Put a DSP+DNN compute engine right at the sensor for ultra-low power voice wake

Moving from Cloud to Edge: Real-time → Low latency

- **Driving architecture change: real-time data**
  - Sensors (such as microphones) proliferate and provide large amounts of **real-world** data
  - Data needs to be acted on **real-time**
- **Role of Edge Intelligence**
  - Edge provides quick local **Sense-Infer-Act loop**
  - Cloud provides longer global **Learning loop**
- **Network Reliability**
  - Can add unacceptable latency or fail outright

From Peter Levine

Moving from Cloud to Edge: Personalized → Privacy

- **Differential Privacy & Federated Learning**
  - Anonymous data sent to cloud for crowdsourced learning
- **On-device Intelligence for Personalization**
  - Machine learning at the edge required

"Federated Learning allows Google products to work better for everyone without collecting raw data from your devices."
--Sundar Pichai, Google CEO

"We try to keep as much of your information on that device as possible, because we want the device to 'know'…because you count on the device to be smart for you."
-- Tim Cook, Apple CEO

Moving from Cloud to Edge: Proliferation → Cost

Cloud Speech-to-Text API pricing

Powerful speech recognition.

Cloud Speech-to-Text is priced per 15 seconds of audio processed after a 60-minute free tier. For details, please see our pricing guide.

FEATURE	0-60 MINUTES	OVER 60 MINUTES, UP TO 1 MILLION MINUTES
Speech Recognition (all models except video)	Free	$0.006 USD / 15 seconds*
Video Speech Recognition	$0.006	$0.012 USD / 15 seconds*

This pricing is for applications on personal systems (e.g., phones, tablets, laptops, desktops). Please contact us for approval and pricing to use the Speech-to-Text API on embedded devices (e.g., cars, TVs, appliances, or speakers).

Cell service

Internet bandwidth

Compute farm

Speech processing service

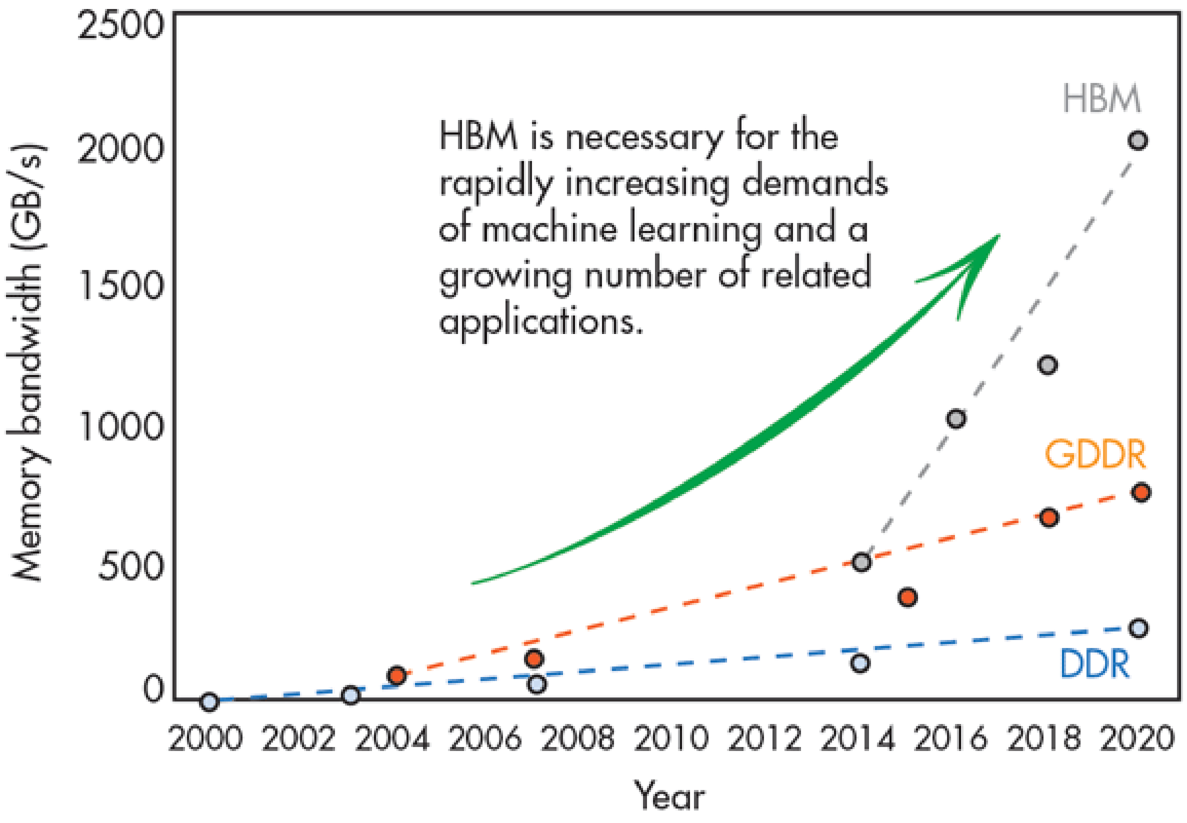
Data center

Moving from Cloud to Edge: Data-driven → Machine learning

- **Advent of Machine Learning**
  - Aggregate as much data as possible from as many sensors as possible to learn "the truth"
  - Deep learning inferences provide better accuracy
  - Continuous training for accuracy and automation

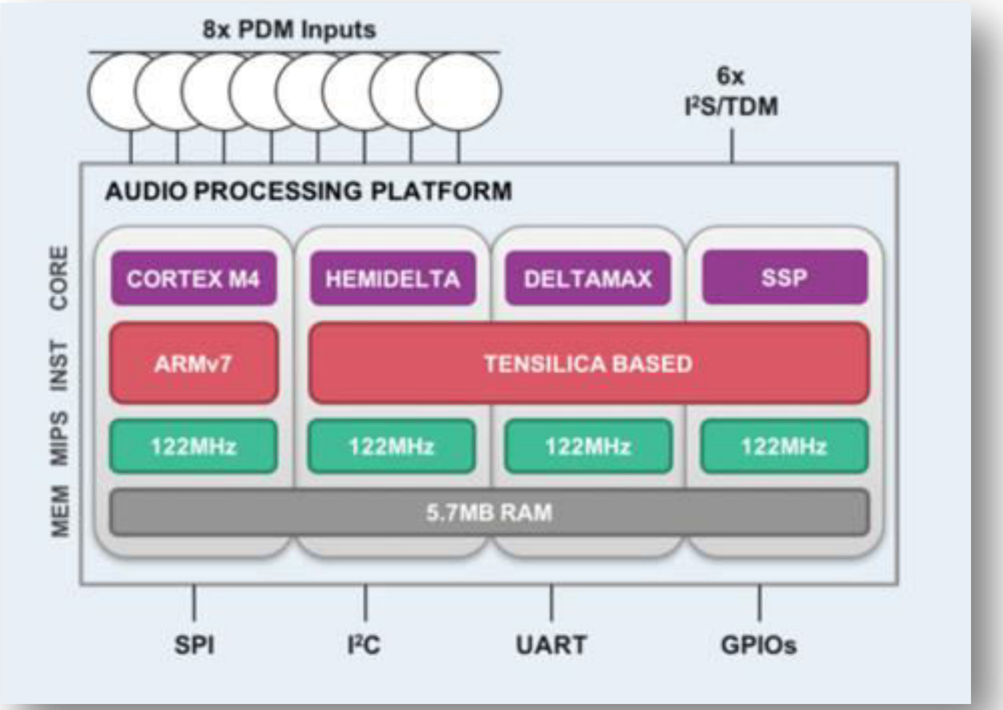
- **Role of hardware**
  - Consume and process as much data as possible as close to the sensors as possible
  - Plethora of use-cases dictates need for OpenDSP with machine learning accelerators

Even the RAM industry needs new technology to handle the increasing bandwidth required for machine learning. HBM – High Bandwidth Memory



Deep Learning Chipsets are becoming more prevalent

Knowles AISonic™ Audio Edge Processor: DSP with DNN accelerators



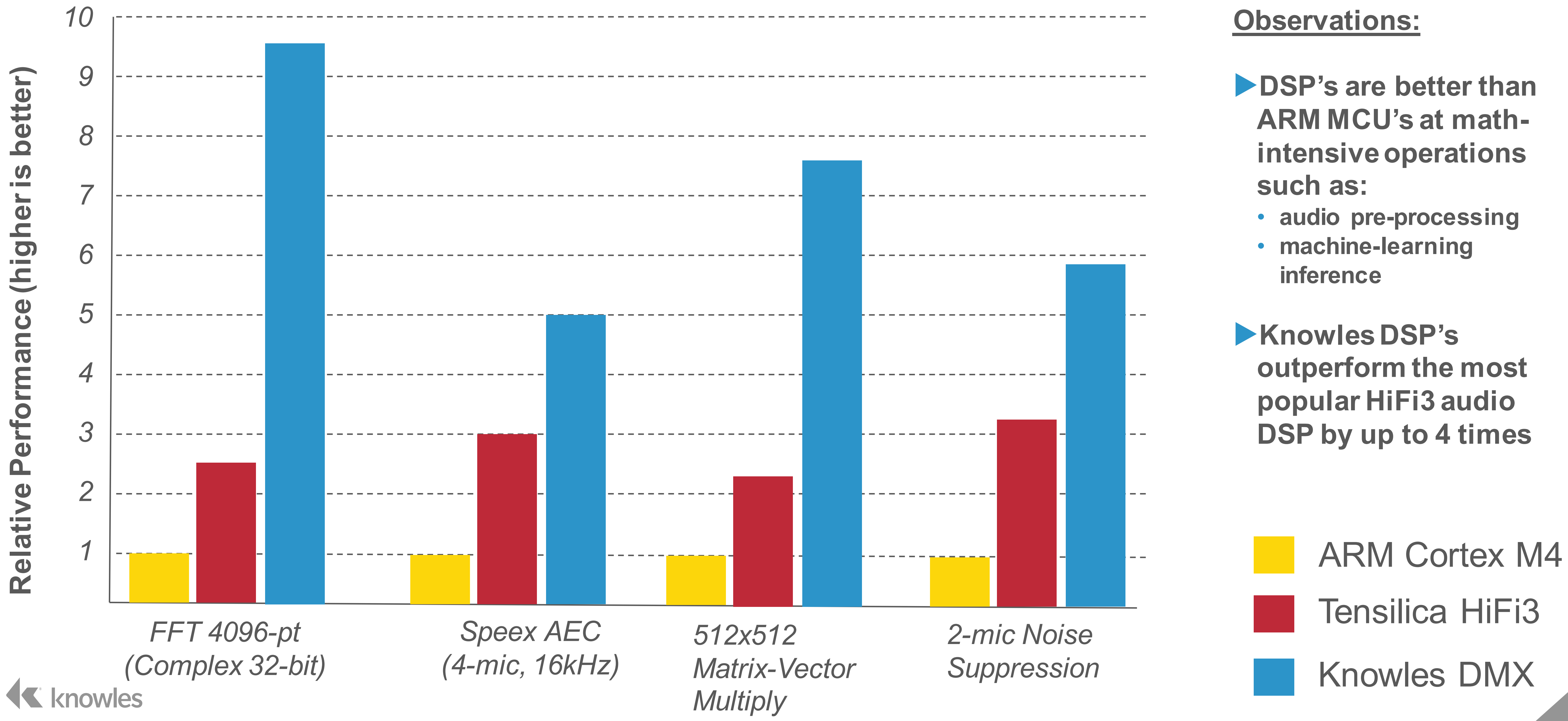
Knowles instruction set & architecture

Machine learning optimized

Feature	Rationale	Impact vs. Competition
Machine learning matrix-vector multipliers	Matrix-vector multiplication used for machine learning (ML) classification	$1/10^{th}$ the energy per multiply
Low-precision operations	ML relies on a huge number of 8-bit operations for inference	Twice the operations per cycle
Large buses, high memory bandwidth	Continuously load data for computing	50% less memory load overhead
Machine learning hardware acceleration	Many non-linear functions needed for machine learning (e.g. sigmoid)	20% less cycles per inference

DSP Efficiency for Audio and Machine Learning

Relative to ARM Cortex-M4 cycle count



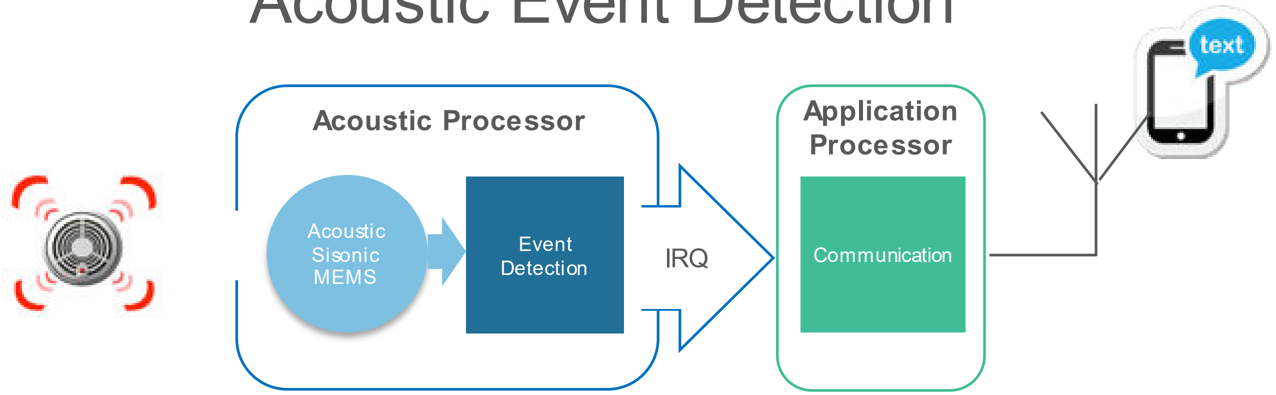
Observations:

- **DSP's are better than ARM MCU's at math-intensive operations such as:**
  - **audio pre-processing**
  - **machine-learning inference**
- **Knowles DSP's outperform the most popular HiFi3 audio DSP by up to 4 times**

knowles

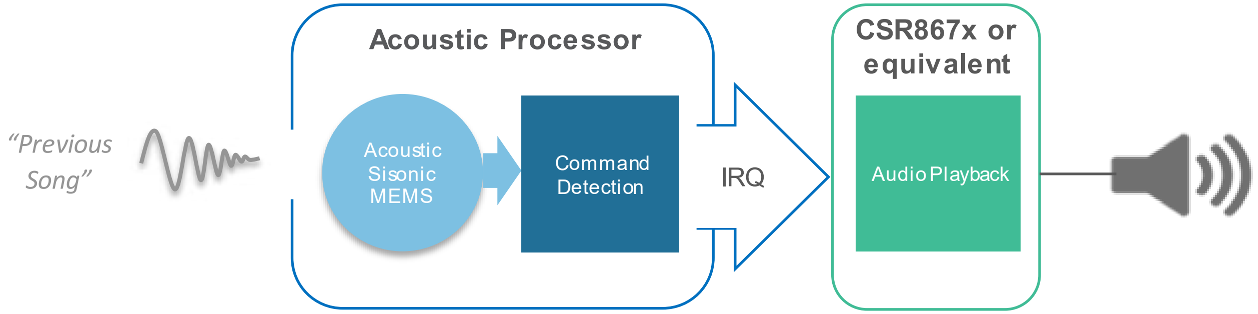
Use Cases Examples for Open Developers with Knowles Smart Microphone

Acoustic Event Detection



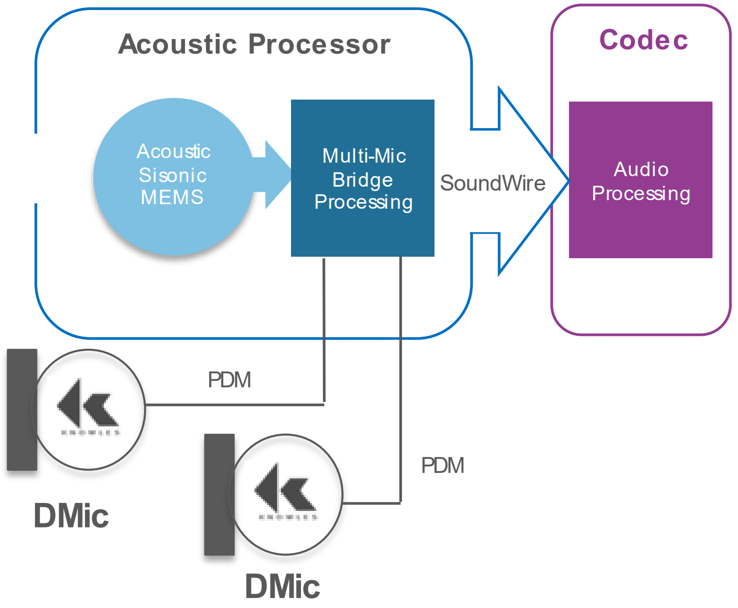
Ideal for the Connected Home IOT

Direct Voice Command



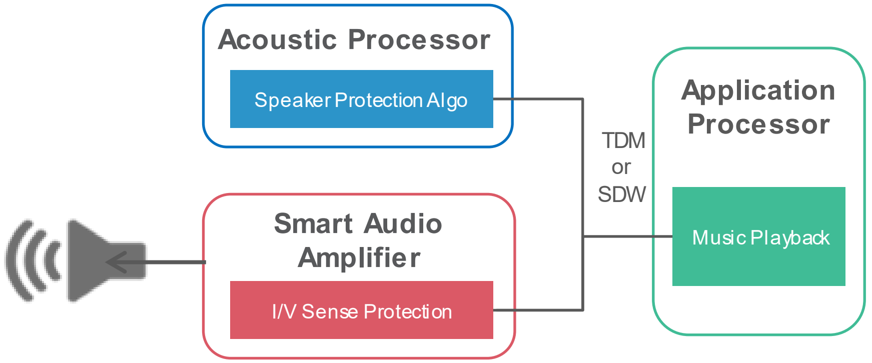
Ideal for music and fitness use cases

Multi-Mic Bridge Processing



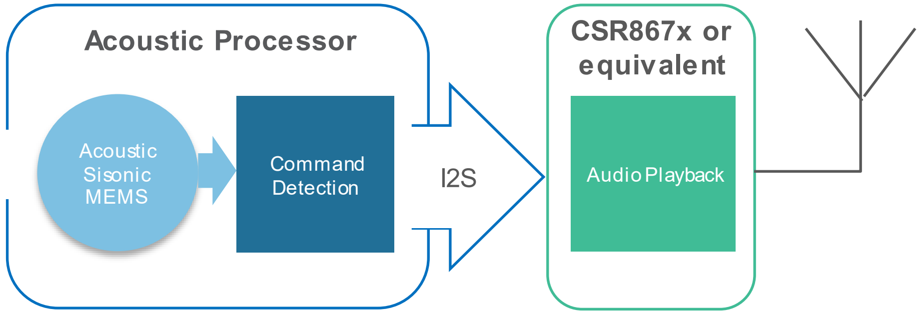
Enable 3-Mic audio processing

Smart Amplifier Control



Ideal for mid-end Smartphones or IOT platforms

1-Mic Noise Suppressor or Active Noise Cancellation



Ideal for Wearables or Headsets

The Future of Consumer Products is Voice

Far-field, Plugged-in, Long life cycle – ***Audio Processor***

Near-field, Low Power, Small Real Estate – ***Smart Microphone***

Knowles High Performance Sisonic Microphone

Enabling Advanced Audio Solutions

Questions?

Knowles creates audio solutions that differentiate our customers

OUR OFFERING

- Higher-performance microphones, advanced multi-mic integration
- DSP, smart mics, audio processors, audio software algorithms, tools, and system solutions

KNOWLES ADVANTAGE

- Cross-functional expertise from acoustics to audio algorithms
- Microphone performance leadership and scale
- Over 10B microphones shipped