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

Jim Steele, VP Technology Strategy, Knowles
Intelligent Audio
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
- Mobile Apps
- IoT
  - Point and Click
  - Touch
  - Voice
Why Voice Interfaces are Becoming Prevalent Now?

ASR = Automatic Speech Recognition

Time variability
Dynamic Time Warping
Spectral variability
Hidden Markov Models
Pronunciation variability
Discriminative training

Acoustic variability
Deep Neural Networks

HUMAN ACCURACY

MACHINE ASR ACCURACY


50% 55% 60% 62% 70% 95%
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

Content Providers and Intent Execution

- Natural Language Understanding (NLU)
- Automatic Speech Recognition (ASR)
- Text to Speech (TTS)
- Acoustic Echo Cancellation (AEC)
- Noise Suppression
- Continuous Voicewake
- Always-on Voicewake

Speech Engines

Hotel Booking

“Book a Hotel in NYC”

Microphones and Signal Processing

Next weekend
New York City
With IoT, distributed computing is on the rise, and will be even larger than mobile

From Peter Levine

Millions
Billions
Trillions

# units worldwide

Main players
HW-Processor-SW

Thousands
Millions
Billions
Trillions

IBM-DEC-DEC
Dell-Intel-Microsoft
Apple-Qualcomm-Google

From Peter Levine
We will see amazing edge devices with new sensors & processors

From Peter Levine

Why edge?
- Real-time
- Always-on
- Proliferated
- Personalized
- Privacy
- Low latency
- Cost
- Low power

Why DSP?
- Data-driven
- Machine learning
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 AIonic™ IA610 Smart Microphone

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

Moving from Cloud to Edge: Real-time \(\rightarrow\) Low latency

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

“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

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

<table>
<thead>
<tr>
<th>Feature</th>
<th>0-60 Minutes</th>
<th>Over 60 Minutes, Up to 1 Million Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech Recognition (all models except video)</td>
<td>Free</td>
<td>$0.006 USD / 15 seconds*</td>
</tr>
<tr>
<td>Video Speech Recognition</td>
<td>$0.006</td>
<td>$0.012 USD / 15 seconds*</td>
</tr>
</tbody>
</table>

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
• **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

Even the RAM industry needs new technology to handle the increasing bandwidth required for machine learning.
Deep Learning Chipsets are becoming more prevalent

Knowles AI Sonic™ Audio Edge Processor: DSP with DNN accelerators

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

- FFT 4096-pt (Complex 32-bit)
- Speex AEC (4-mic, 16kHz)
- 512x512 Matrix-Vector Multiply
- 2-mic Noise Suppression

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
Use Cases Examples for Open Developers with Knowles Smart Microphone

Acoustic Event Detection

Acoustic Processor

Event Detection

Acoustic Sisonic MEMS

Application Processor

Communication

Ideal for the Connected Home IOT

Direct Voice Command

Acoustic Processor

Command Detection

Acoustic Sisonic MEMS

CSR867x or equivalent

Audio Playback

“Previous Song”

Multi-Mic Bridge Processing

Acoustic Processor

MultiMic Bridge Processing

SoundWire

Codec

Audio Processing

Enable 3-Mic audio processing

1-Mic Noise Suppressor

or Active Noise Cancellation

Acoustic Processor

Command Detection

Acoustic Sisonic MEMS

CSR867x or equivalent

Audio Playback

Enable 3-Mic audio processing

Ideal for Wearables or Headsets

Smart Amplifier Control

Acoustic Processor

Speaker Protection Algo

Application Processor

Music Playback

Smart Audio Amplifier

TDM or SDI

IV Sense Protection

Ideal for mid-end Smartphones or IOT platforms

Enable 3-Mic audio processing
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
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