Hearing Sense in Smart Devices

Alexander Goldin, PhD
Founder & CEO
Digital Human – Sarah (MWC 2018)
Alango Technologies – human to human

Since 2002

Voice communication
Alango Technologies – human to machine

Since 2014
Our expectations ?
Our expectations: Always ON

Always ON
listening
understanding
answering
Our expectations: Artificial Intelligence
Alan Turing
1912-1954
1950 – Turing test for artificial intelligence

Interrogator
Turing test nowadays
Real test conditions

What AI system hears
Unreal test conditions
Intelligent behavior: wait for request of attention
Device name – the trigger phrase
What was the first documented voice trigger phrase?

Wait

Alexa
OK Google
Hey Siri
Hey Cortana
Anything else?
The first voice trigger phrase

“Open Sesame” - 1704
One Thousand and One Nights
by Antoine Galland
Intelligent behavior: react, listen, execute
Microphone array and beamforming

User

Acoustic Beamforming

Wait

Trigger Phrase Recognizer
Beamforming “time of arrival” principals
Beamforming algorithm

Filter 1
Filter 2
Filter 3

Fixed OR Adaptive
Multiple beamforming
Single Keyword Recognizer architecture
Who is the master?
Multiple Trigger Phrase Recognizers (TPR)
Barge-in problem

Multi-Beamforming

TPR 1

TPR 2

TPR M

Control Logic

ASR & AI
Acoustic echo cancellation principals

Acoustic echo cancellation (AEC) is a technique used to remove echoes from a signal. It involves the use of an adaptive filter to predict the echo signal and subtract it from the original signal. The process is illustrated in the diagram:

- **Loudspeaker signal**
- **Predicted echo**
- **Clean signal**

The adaptive filter is used to estimate the echo signal and subtract it from the loudspeaker signal to provide a clean output.
The front end - acoustic echo cancellation
Voice Enhancement Package (VEP)

- Echo Cancellation
- Multi-Beamforming
- Best-Beam Logic

1. Beam 1
2. Beam 2
N. Beam N

“Best Beam”

Stereo
VEP – “Best Beam” use configuration
VEP – “Best Beam” use configuration

Echo Cancellation

Multi-Beamforming

Best-Beam Logic

Trigger Phrase Recognizer 1

Trigger Phrase Recognizer 2

Trigger Phrase Recognizer N

"Best Beam"

Stereo
VEP configurability - ANY circular array

4 mics

6 mics

8 mics

8 beams

6 beams

8 beams
VEP configurability - ANY (bi) linear array

6 mics

7 beams

2x4 mics

7 beams
VEP performance demo – 6 microphone linear array
VEP performance: case by case comparison

Trigger World Recognition Rate by Noise Environment Type

- Silent room, normal speech volume
- Silent room, weak speech volume
- Surround cafeteria noise
- Uncorrelated white noise (all sides)
- Uncorrelated white noise (from behind)
- Distractor (TV) at 45 degrees
- Distractor (TV) at 135 degrees

Alango Duplexa 4-mic
Amazon Echo 2 7-mic
Synaptics 4-mic
Average “Alexa” Recognition Rate

- Duplexa VEP 4 mic: 61.14%
- Echo 2 7 mic: 53.34%
- Synaptics 4 mic: 36.52%
Live demo: Amazon Echo against Duplexa demo kit
State of the art: How much power is needed?

<table>
<thead>
<tr>
<th>Mode</th>
<th>Amazon Echo (first generation)</th>
<th>Amazon Echo Dot</th>
<th>Google Home</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idle</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Listening</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Playing music (low volume)</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Playing music (high volume)</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Average power draw (watts)
2019 – 120 million smart speakers in US

350 Megawatt just for always listening

3 $\text{km}^2$ of solar panels

400 Always ON

Soccer fields of solar panels in US only

“Always Sleep” solution for Always ON problem
Alango VAD demo (STM Disco board)

2 MHz

VAD flag
Live demo: Temi Robot

Temi Robot at MWC 2019

VEP with 4 mic square array
Year 2021: 1.4 billion smart devices sold

2019-2021 > 3 Billion smart devices
Thank you

Questions, please!