VOICE

DESIGNING AND DEVELOPING
FOR THE FUTURE OF USER EXPERIENCE
WHY ARE AMAZON TALKING ABOUT IOT AND VOICE?

The IoT is growing rapidly with some research estimating that there will be 50Bn connected devices by 2020.

- If you aren't interested in making your own connected devices, you can use ASK to get your application to millions of users.
- If you are making IoT devices you can voice enable it with AVS.
- If you're making smart home products, you can control those products via voice with Alexa.
VOICE WILL BE EVERYWHERE
The age of touch could soon come to an end. From smartphones and smartwatches, to home devices, to in-car infotainment systems, *touch is no longer the primary user interface.*

Source: Design News
Although voice technology is still in its relative infancy, the future is not as far off as you think.
Rapid Advancement

**ASR**
Automated Speech Recognition

What is the user actually saying?

**NLP**
Natural Language Processing

What is the intent of the user?

**ACCESS**
Democratizing Voice Technology

Better ASR and NLP has lead to better access and adoption

**CLOUD**
Scalable, Reliable & Secure

Ability to introduce features at scale, that continuously add value over time

**AI**
Machine Learning and Intelligence

Provide extraordinary and advanced customer experiences
ASR accuracy has dramatically increased in the last 4-5 years. This inflection point has created **sustained momentum** in consumer adoption of voice technology.
As speech recognition accuracy goes from 95% to 99%, all of us in the room will go from barely using it today to using it all the time. Most people underestimate the difference between 95% and 99% accuracy.

99% is a game changer.

Andrew Ng, Chief Scientist at Baidu
MACHINE LEARNING & INTELLIGENCE

WE ARE HERE

HUMAN PERFORMANCE

COMPUTER PERFORMANCE & MACHINE LEARNING
Soon it will seem almost quaint there was a time we looked at voice assistants as virtual friends who lived in our pockets and answered our questions.

It took generations and several major technological advancements for touchscreens, GUI and VUI to achieve critical adoption.

Following non-commercial GUI milestones, the advances of the early 80s and 90s (Windows 95, Apple’s OS, the Internet) changed trajectory of the GUI.

It wasn’t until the Palm Pilot of the late ‘90s and smartphones of the mid-2000s that allowed Touch to emerge as a key interaction modality.

Human-To-Human VUI was brought in with the dawn of the telephone, but Human-To-Machine VUIs have just recently become viable.
TOUCH vs. VOICE

- Can handle more info
- More familiar
- Harder to get lost
- Provides flexibility

- Faster
- Less cumbersome
- Universal
- Removes noise
VOICE AS A KEY MODALITY

Where it’s going is limited by one’s imagination but voice WILL play a key role in how we control our homes, our outdoor spaces and access information...Why?

ACCESS
Across billions of devices between phones, watches cars, Alexa-powered devices such as the Amazon Echo, Echo Dot, Amazon Tap, Amazon Fire TV, is making access to voice ubiquitous.

ACCURACY
Advances in the ability to understand user intention is changing the game in adoption as interactions become faster and more reliable.

EFFICIENCY
Ease of use will make it a powerful choice for quick access to anything inside and outside our environment.

SECURITY
Voice is a highly unique signature, as advances in biometrics are integrated with voice, our individuality will become a key to further personalization and security.
DEVELOPING
FOR VOICE
Create Great Content: ASK is how you connect to your consumer

Alexa Skills Kit

Lives In The Cloud
Automated Speech Recognition (ASR)
Natural Language Understanding (NLU)
Always Learning

Alexa Voice Service

Unparalleled Distribution: AVS allows your content to be everywhere
ALEXA

SKILLS KIT
UNDER THE HOOD OF THE ALEXA SKILLS KIT
A closer look at how the Alexa Skills Kit processes a request and returns an appropriate response

User Makes a Request

Audio Stream is sent up to Alexa

Alexa Identifies Skill & Recognizes Intent Through ASR & NLU

Alexa Converts Text-to-Speech (TTS) & Renders Graphical Component

Alexa sends Customer Intent to Your Service

Your Service processes Request

Respond to Intent through Text & Visual

You Pass Back a Textual or Audio Response

You Pass Back a Graphical Response

Your Service processess Request
Alexa, what’s the weather?

Alexa Voice Service

Speech Platform

ASR

NLU

Skills

TTS

user’s utterance

recognize

recognition result

intent

“speak” directive

text/SSML

recognition result

intent

Alexa’s voice

Alexa’s voice

SPEECH CAPTURE

TEXT TO SPEECH OUTPUT

WAKE WORD DETECTION
WHAT COMPONENTS MAKE UP A SKILL

Skills are made up of two components

Skill configuration in the Amazon Developer Portal

and

Your skill code, hosted in AWS Lambda or your own HTTPS endpoint
INVOCATION NAMES

Invocation names are how we know to route traffic to your particular skill.

Interactions can be either:

One Shot – open your skill and perform an action such as ‘Alexa, ask National Rail for my commute’

Conversational – Alexa, ask National Rail to set up my commute’ - ‘OK, what is your regular departure station’ – ‘Birmingham New Street’

Open Only – Alexa, open National Rail

Your skill can support all of these, it’s not one or the other.

‘Alexa, ask National Rail for my commute’

Alexa, open Just Eat

Alexa tell Uber to get me a ride

Alexa, launch Cat Facts

Alexa, play Reindeer Trivia
You define interactions for your voice app through intent schemas.

Each intent consists of two fields. The intent field gives the name of the intent. The slots field lists the slots associated with that intent.

Slots can also include types such as LITERAL, NUMBER, DATE, etc.

Intent schemas are uploaded to your skill in the Amazon Developer Portal.
Custom Slots increase the accuracy of Alexa when identifying an argument within an intent.

They are created as a line separated list of values.

It is recommended to have as many possible slots as possible.

There are some built in slots for things such as GB.City and GB.FirstName.

bakerloo
central
circle
district
hammersmith and city
jubilee
metropolitan
northern
piccadilly
victoria
waterloo and city
london overground
tfl rail
DLR
The mappings between intents and the typical utterances that invoke those intents are provided in a tab-separated text document of sample utterances.

Each possible phrase is assigned to one of the defined intents.

tubeinfo are there any disruptions on the {LINENAME} line

tubeinfo {LINENAME} line
Putting It All Together

tubeinfo are there any delays on the {LINENAME} line

```json
{
   "intent": "tubeinfo",
   "slots": [
      {
         "name": "LINENAME",
         "type": "LINENAMES"
      }
   ]
}
```

Utterance

**Utterance**
tubeinfo are there any delays on the {LINENAME} line

**Intent**

```json
{
   "intent": "tubeinfo",
   "slots": [ ]
}
```

**Slots**
bakerloo
central
...

REQUEST TYPES

LaunchRequest
Occurs when the users launch the app without specifying what they want

IntentRequest
Occurs when the user specifies an intent

SessionEndedRequest
Occurs when the user ends the session
AN EXAMPLE REQUEST

If hosting your own service, you will need to handle POST requests to your service over port 443 and parse the JSON:

With AWS Lambda, the event object that is passed when invoking your function is equal to the request JSON:

Requests always include a type, requestId, and timestamp:

If an IntentRequest they will include the intent and its slots:

type maps directly to LaunchRequest, IntentRequest, and SessionEndedRequest:

```
"request": {
  "type": "IntentRequest",
  "requestId": "string",
  "intent": {
    "name": "tubeinfo",
    "slots": {
      "LINENAME": {
        "name": "LINENAME",
        "value": "circle"
      }
    }
  },
  "locale": "en-GB"
}
```
Your app will need to build a response object that includes the relevant keys and values.

The alexa-sdk for Node.js makes this super simple.

`outputSpeech`, `card` and `reprompt` are the supported response objects.

`shouldEndSession` is a boolean value that determines whether the conversation is complete or not.

You can also store session data in the Alexa Voice Service. These are in the `sessionAttributes` object.
THE CODE
The promise for voice is great... but so is the potential for failure
DESIGNING FOR VOICE
WHAT EXACTLY IS VOICE USER INTERFACE (VUI) DESIGN?
Voice User Interface (VUI) design creates delightful experiences using voice and natural language, by designing voice interactions that fulfill a user's request, engages them in conversation, and makes the technology they're using, seem totally invisible.
DO
I can give you disruption information for the London Underground. What line would you like to check?

DONT
I can give you disruption information for the London Underground
**SOME DO’s & DONT’s of VUI**

**DO**
I can give you disruption information for the London Underground. What line would you like to check?

**DO**
Welcome to the The Underground skill. You can get disruption information by saying a London Underground line name. What line would you like to check?

**DONT**
I can give you disruption information for the London Underground

**DONT**
Welcome to The Underground
SOME DO’s & DONT’s of VUI

**DO**
I can give you disruption information for the London Underground. What line would you like to check?

**DO**
Welcome to the The Underground skill. You can get disruption information by saying a London Underground line name. What line would you like to check?

**DO**
Which line would you like disruption information for?

**DONT**
I can give you disruption information for the London Underground

**DONT**
Welcome to The Underground

**DONT**
I can give disruption info for all of the London Underground lines. Which one would you like....
**SOME DO’s & DONT's of VUI**

**DO**
I can give you disruption information for the London Underground. What line would you like to check?

**DO**
Welcome to the The Underground skill. You can get disruption information by saying a London Underground line name. What line would you like to check?

**DO**
Which line would you like disruption information for?

**DO**
There are currently no delays on the circle line

**DONT**
I can give you disruption information for the London Underground

**DONT**
Welcome to The Underground

**DONT**
I can give disruption info for all of the London Underground lines. Which one would you like....

**DONT**
You would like disruption information for the circle line right?
WHAT NEXT?
ENTER ONE OF OUR PROMOS

Hackster.io – API Mashup Contest

Attend a MLH Hackathon

Publish a Skill – Get a Free t-shirt
PLENTY MORE RESOURCES OVER AT DEVELOPER.AMAZON.COM/ASK
THANK YOU QUESTIONS?