



winhec

Workshop Taipei 2016

Any questions please contact winhectpe@microsoft.com



Natural User Interface Technologies for IoT

Tony Lin
Software Engineer

Partner Enablement Team

Speech Capabilities Available

Microsoft Speech Offering

- Win 10 Speech APIs
 - Local Commands with constrained grammar
 - E.g. Turn on, turn off
 - Cloud dictation
 - Typing a message, Web search, complex phrases
- Azure marketplace
 - Oxford APIs
 - Language Understanding Intelligent Service (LUIS) – For enabling rich natural language
 - Speech Recognition
 - Similar to Cloud Dictation of Speech APIs
 - Bing Translate
- Cortana

Microsoft Speech APIs

- Win 10 Speech APIs
 - Local commands with constrained grammar
 - Higher recognition rate for local with constrained grammar
 - E.g. Turn on, turn off
 - Cloud dictation
 - Typing a message, Web search, Complex phrases
- If on Windows 10, use Speech APIs – its free and available on the platform
- For non-Windows platform, use Azure marketplace solutions
- On IoT Core, if using Speech APIs cloud dictation is auto-enabled. If needed, please disable it explicitly

Using Microsoft Speech Platform

- Using a combination of recognition and synthesis capabilities listed below, you could build a complete speech interface for your device
 - Synthesizing text to speech (TTS)
 - Synthesizing Speech Synthesis Markup Language (SSML)
 - One-shot recognition using the
 - predefined dictation grammar, predefined web search grammar, custom list-based grammar, custom SRGS/GRXML grammar
 - Continuous dictation
 - Continuous recognition using a
 - custom list-based grammar, custom SRGS/GRXML grammar
 - Pausing and resuming continuous recognition

Demo

<https://github.com/Microsoft/Windows-universal-samples/tree/master/Samples/SpeechRecognitionAndSynthesis>

Oxford APIs

- Speech Recognition
- Text to Speech Conversion or Speech Synthesis
- Speech Intent Recognition
 - Convert spoken audio to intent
 - Speech Intent Recognition: In addition to returning recognized text from audio input, the server returns structured information about the incoming speech so that apps can easily parse the intent of the speaker, and subsequently drive further action
 - Models trained by the Project Oxford LUIS service are used to generate the intent

Project Oxford - LUIS

Create language understanding models

Create models for your application to better understand intents like "turn on the lights", or entities such as "start a new jog/walk/hike/bikeride". Tune your model with in-depth performance visualizations.

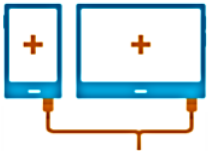


Use pre-built, world-class models from Bing and Cortana

Use the pre-built, world-class models to recognize entities like places, times, numbers, temperatures, and to also handle common requests like "set an alarm for 8 AM". Immediately enable personal assistant functionalities by using a selection of Cortana understanding models.

Deploy your models to an HTTP endpoint

Deploy models to an HTTP endpoint with one click. LUIS returns easy-to-use JSON.



Activate models on any device

Activate your language understanding models from your application on your phone, tablet, or any other device that has access to the Internet.

Maintain models with ease

Review commands spoken to your application to spot and correct errors. LUIS automatically suggests the most important commands to label.



LUIS – Contd..

- LUIS endpoints work seamlessly with Project Oxford's speech recognition service. In the C# SDK for the Project Oxford Speech API, you can simply add the LUIS application ID and LUIS subscription key, and the speech recognition result will be sent for interpretation.
- *Currently available for only English and Chinese
- **Use it only if full natural language capabilities are needed and if you are willing to invest a developer to create, train and improve models.

Demo

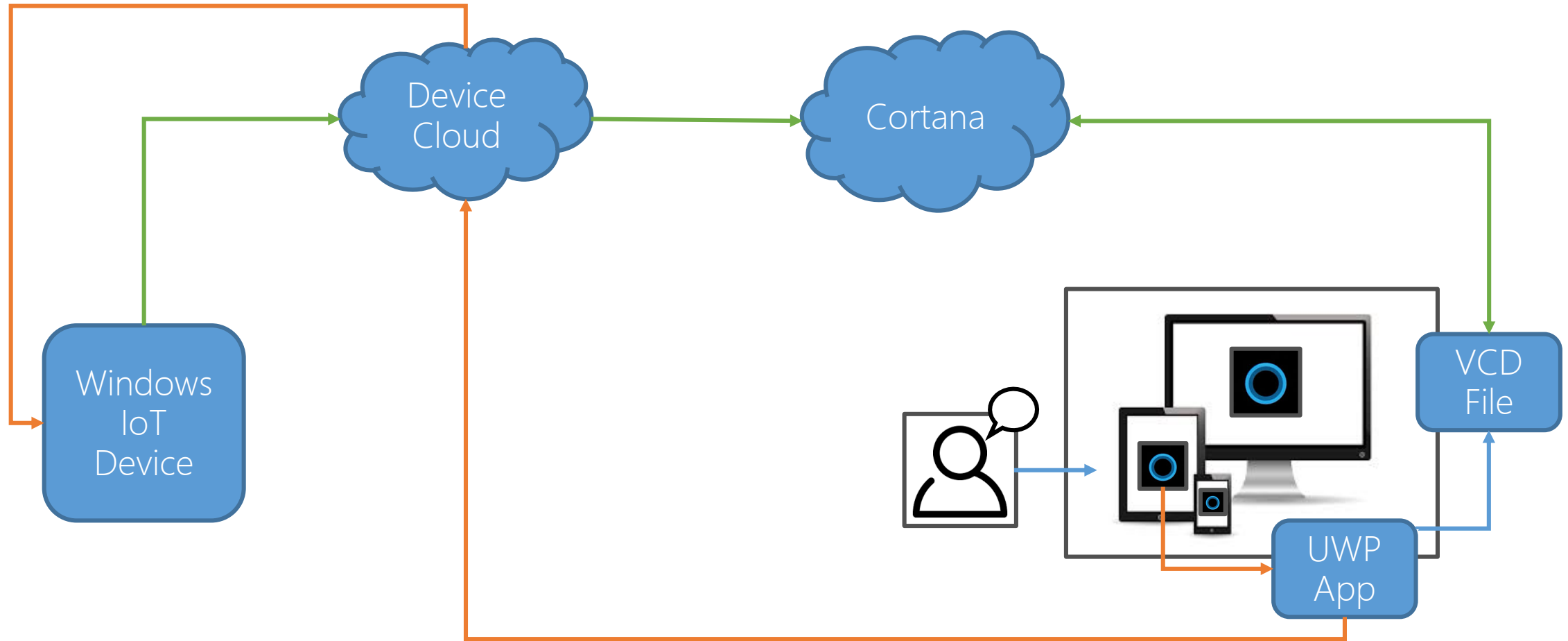
<https://www.projectoxford.ai/demo/speech>

Cortana Core Capabilities

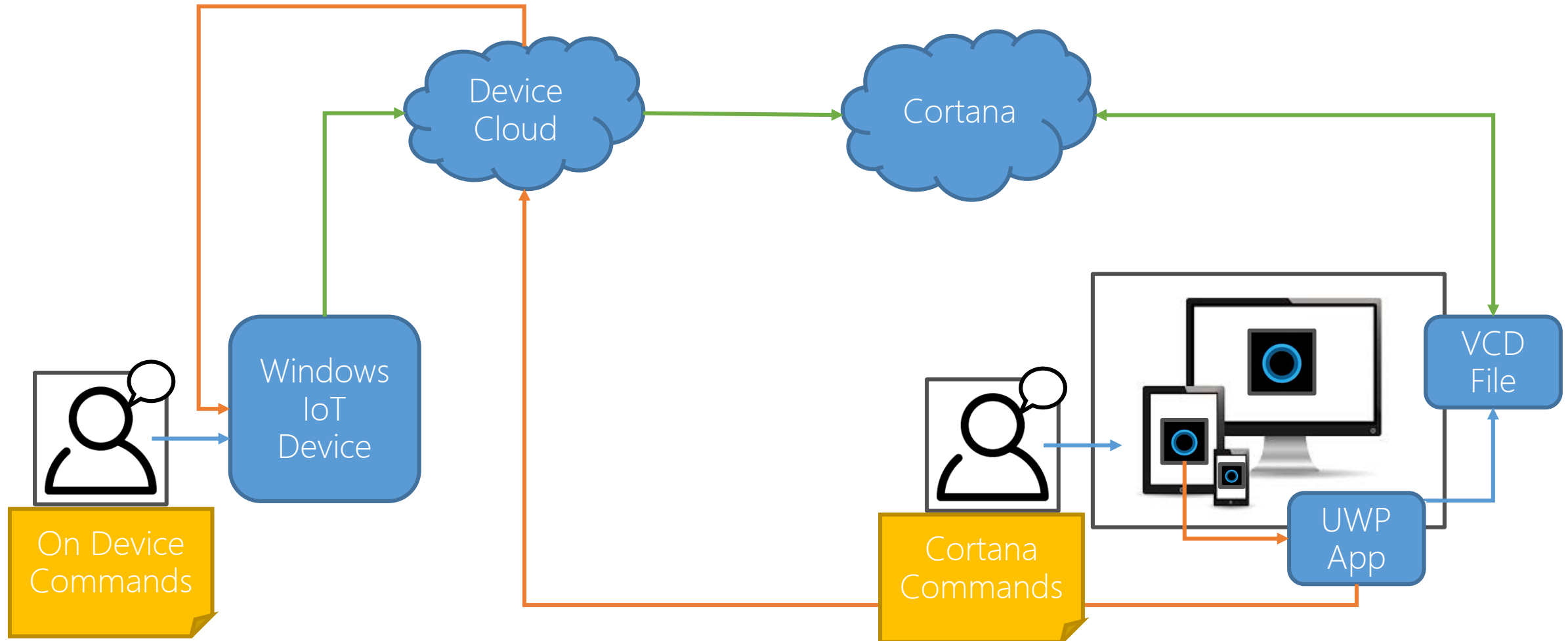
- Cortana is primarily a clever personal assistant (with language capabilities)
- Cortana can search the web, find things on your PC, and keep track of your calendar, even tell you jokes
- Key features
 - Setting appointments and reminders
 - Finding stuff – Search
 - Managing tasks
 - Support for text and speech input

Area	Microsoft Speech Platform	Azure Speech – Oxford	Cortana
Local/ Cloud	Local AND/OR Cloud	Cloud only	Cloud only
Languages supported		English(U.S.), English(U.K.), German, Spanish, French, Italian, Mandarin	Chinese (Simplified), English (U.K.), English (U.S.), French, Italian, German, and Spanish.
End-user MSA Needed	No	No	Yes
Azure subscription	No	Yes	No
Cost	Free	Paid (metered based on number of REST calls)	Free
User Experience and Branding	Non-branded speech platform	Non-branded speech platform	Cortana brand For use in personal assistance scenarios
Devices	First party tight integration with Windows Devices	For use on any device (REST based)	Windows including IoT Mobile and Industry (not available on IoT Core Devices) Coming soon on Android and iOS
LUIS Integration (Available in English and Chinese only)	Manual, but possible	Tight out-of-the-box integration	No

Companion App Model (for IoT Devices)



Speech on Device + Cortana Companion App



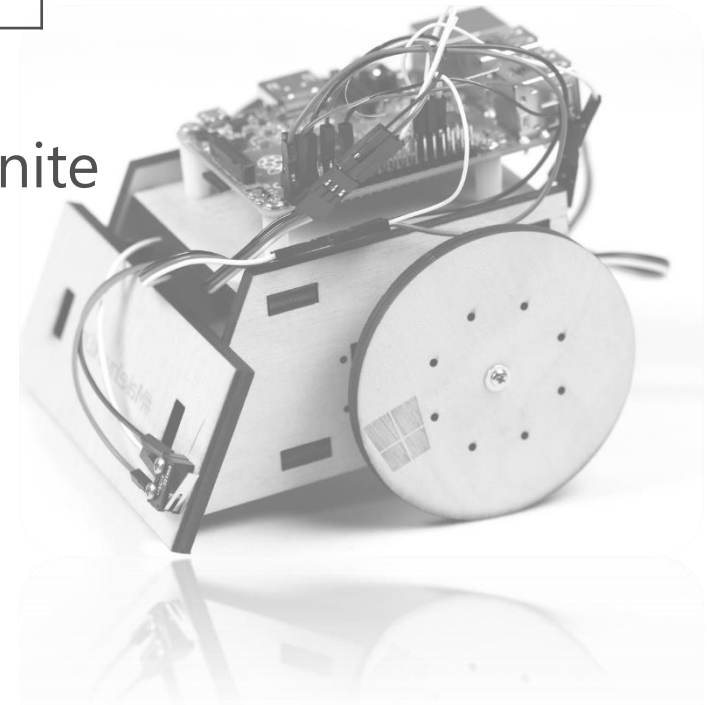
Voice Scenarios

Scenario: Speech Controlled Robot

- Ram wants to build a robot with IoT Core. He wants to create the following interactions

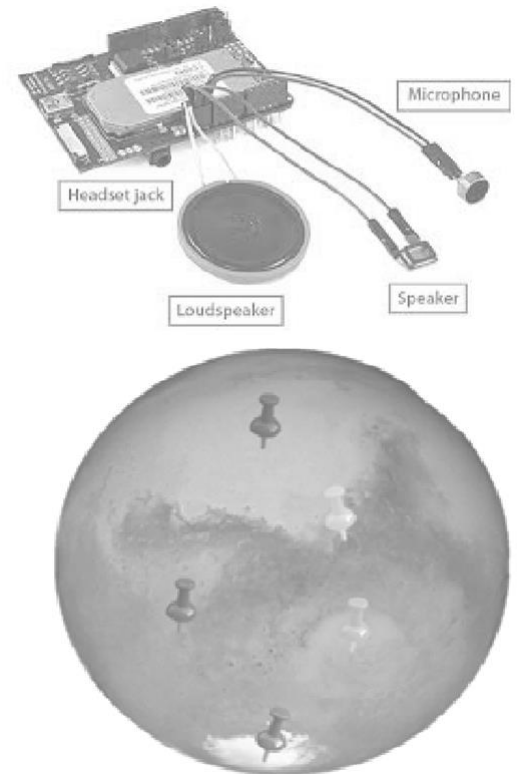
Come forward	Go back	Spin around
Go faster	Go slower	How far did you go?

- Can't afford latency in speech processing, so it is local
- The set of commands that the device can respond to are finite



Scenario: Speech-enabled Mars Exhibit

- Nicki is making an interactive Mars Exhibit for her science class. She wants her classmates to ask questions about Mars, explore various 'pins' or points of interest she has picked out
 - Tell me about the red pin
 - What is at the blue pin?
- Attributes of this scenario
 - It is a public device, everyone at the science fair has access
 - The device doesn't need to know any personal data to enable E2E speech
 - The set of commands that the device can respond to are finite



Scenario: Front Door Messages

- Jaden wants to automate her front door. She wants her door to announce when someone is at the door. If she is not at home, visitor can leave a message which is transcribed and sent to her
- Attributes of this scenario
 - Long form dictation
 - Internet connection is available and accessible



Scenario: Home Automation – Device Command and Control

- Jaden is on her way to work, but wants to check if her garage door is closed with queries
 - Is the garage door is closed
 - Close the garage door
 - Show me the camera feed from garage
- Attributes of this scenario
 - Speech enquiry / control is done away from the phone or tablet



Scenario: Speech Controlled Farming System

- Carly, a small scale farmer wants to harvest rain water for use in her farm. However, she wants to be able to control the water based on inspection of each plant
 - "Give plant 4 30 gms of magnesium with water for next 5 days"
 - Water plant 7 tomorrow morning and evening for 5 mins at 200 ml/sec
 - Don't water any plant if it rains tomorrow morning
 - Remind me to check on the pests in potatoes next time I am here
- Attributes of this scenario
 - Solution needs rich grammars with advanced language models
 - Carly, a maker pro, is willing to create, maintain models and refine models but wants to also save on additional expenses



Scenario: Sous Vide Machine

- It has a food scale, timer & integrated temperature gauge & enables the users to get perfect results without doing more than just indicating what foods they are trying to cook.
 - Turn on/off the machine – “turn it on”, “switch it on”, “is it on?”
 - Set the alarm/ timer – “Set the timer for 15 minutes”, “Remind me to check it in a 3 minutes”
 - Configure the machine and ask for the current status of the cook
- Attributes of this scenario
 - It is a specific built device
 - No user MSA is needed for enabling speech
 - Natural language is used to interact with the machine.
 - Models for “time”, “device on/off”
 - Can fold in the cost of maintaining speech back-end into the cost of the device



Summary

Top priority	Use	Example	Also consider
Latency	Windows 10 Local Speech API	Voice controlled robot	Are the commands complex? Do you need to use LUIS for intent recognition?
Cost	Windows 10 Speech API – Free; Available on all Windows Devices	Mars Exhibit	Does a version of the solution need to work on non-Windows Device? For example, an Android component? In that case, use Bing Speech APIs, so that the speech code can be reused across both Windows and non-Windows solution.
Dictation (Long-form)	Windows 10 Speech APIs	Front door messages	Although dictation is available in Oxford APIs, long form dictation is available only locally.

Top priority	Use	Example	Also consider
Speech synthesis	Windows 10 Speech APIs	All	Speech synthesis is also possible via. Oxford, but comes for free with Speech APIs on Windows devices. Except if you want to speak in a language different than what is default on OS. In that case, use Oxford APIs.
Remote control of IoT Devices	Cortana Extensions on the companion app	Home automation	Developers need to write connectivity layer
Natural Language with themes such as 'time'	LUIS with either Windows 10 Speech APIs or Oxford Speech APIs	Sous Vide machine	Use Windows 10 speech APIs if cost is a factor with LUIS (Need additional steps to integrate with LUIS) If using Oxford Speech, cost is incurred on both Speech API and LUIS
Natural language for highly custom domain	LUIS with either Windows 10 Speech APIs or Oxford Speech APIs	Speech controlled farming system	Custom models need to be built, trained and refined. Needs a dedicated speech developer.

Facial Analytics

Facial Analytics using Project Oxford

- Face Detection
 - To detect human faces in image with face rectangles and face attributes including face landmarks, pose, gender and age
- Face Identification
 - To search which specific person entity a query face belongs to, from user-provided person-face data
- Face Verification
 - To check two faces belonging to same person or not, with confidence score
- Similar Face Searching
 - To find similar-looking faces from many faces by a query face.
- Face Grouping
 - To organize many faces into face groups based on their visual similarity.

Demo

<https://www.projectoxford.ai/demo/face>

Recap

- Microsoft speech offering
 - Win 10 speech APIs
 - Azure marketplace
 - Cortana
- Facial analytics
 - Face Detection
 - Face Identification
 - Face Verification
 - Face Searching, grouping

Reference

- Windows.Media.SpeechRecognition
 - <https://msdn.microsoft.com/en-us/library/windows.media.speechrecognition.aspx>
- LUIS
 - <https://www.projectoxford.ai/luis>
- Project Oxford Face APIs
 - <https://www.projectoxford.ai/face>

Calls to Action

Join WinHEC LINE Community
[@winhec](#)



We want to hear from you!

Please Complete the Evaluation
Form and return it to our reception.

Your input is highly important to us!
Thank you!! 😊



