Building Windows 10 Universal Apps with Speech and Cortana

Nick Landry
Senior Technical Evangelist – NY Metro
AgeofMobility.com
github.com/ActiveNick
@ActiveNick
Agenda

Why Integrate Speech in Apps?
Speech Comparison: Windows vs. iOS & Android
Cortana Extensibility & Voice Commands
Speech Synthesis
In-App Speech Recognition
Demos
Join the Discussion!
One Store + One Dev Center

Reuse Existing Code

One SDK + Tooling

Adaptive User Interface

Natural User Inputs

One Universal Windows Platform

Multiple Device Families

Devices + IoT

Mobile

PC

XBox

Surface Hub

HoloLens
• Differentiate your app with voice commands
• Delight your users with natural interactions
• Increase user engagement
• Increase productivity
• Increase discovery
# Speech SDK Features by Mobile OS

<table>
<thead>
<tr>
<th>Speech SDK Features for Devs</th>
<th>Windows 10</th>
<th>Windows/WP 8.1</th>
<th>iOS (iPhone/iPad)</th>
<th>Android (Phone/Tablet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-in Personal Assistant with Speech</td>
<td>Cortana</td>
<td>Cortana (WP8.1 only)</td>
<td>Siri</td>
<td>Google Now Launcher</td>
</tr>
<tr>
<td>Personal Assistant Extensions in third-party apps using Voice Commands</td>
<td>Yes</td>
<td>Yes (WP8.0+ only)</td>
<td>No (first-party apps only)</td>
<td>No (invite-only third-party apps)</td>
</tr>
<tr>
<td>Speech Synthesis SDK for Devs (Text-to-Speech)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>In-app Speech Recognition SDK for Devs</td>
<td>Yes</td>
<td>Yes (WP8.0+ only)</td>
<td>No (entry field dictation only)</td>
<td>Yes</td>
</tr>
</tbody>
</table>
What about iOS and Android?

- Cortana is available as an app on iOS and Android
- No integration available to launch your own apps via Voice Commands
- No developer story (yet)
- Microsoft Cognitive Services provide many services:
  - Vision, Speech, Language
  - Available to iOS, Android & Windows developers via REST services & client libs
- For more information, see: https://www.microsoft.com/cognitive-services/ (aka http://projectoxford.ai)
Cortana

Cortana is available on Windows 10 desktops, tablets, and phones – as well as iOS, Android and Cyanogen phones.
Cortana Countries

- Windows 10 Launch: Cortana available in the following countries/regions:
  - United States, United Kingdom, China, France, Italy, Germany, Spain, Japan, Australia, English Canada and India (English only)

- Windows Insiders also get extra Cortana countries:
  - Brazil, Mexico, and French Canada

- For more info on Cortana’s worldwide expansion:
Speech Recognition Options

- If the app is not running...
  - Use Cortana integration
  - Voice Commands via .vcd file

- If the app is running...
  - Use in-app speech recognition
  - Work with default open dictation grammar or custom grammars (phrase lists, GRXML)
  - You can still use activation via voice commands, but your app will get re-launched
Cortana Extensibility

From Cortana to your apps: Voice Commands
Voice Commands for Universal Windows Apps

Deep linking with Foreground Voice Commands

- Universal Apps can register to handle Cortana voice commands directed at their app
- When users speak these commands, Cortana activates the app with the spoken context
- Apps trigger scenarios based on this context
- Available for Universal Windows Apps built with XAML and JavaScript
- Feature originally introduced in Windows Phone 8.0 SDK (i.e. Silverlight app model)
Adventure Works, show my trip to London
DEMO: Cortana Extensibility

Deep Linking to App Scenarios with Third-Party Voice Commands
<?xml version="1.0" encoding="utf-8"?>
<VoiceCommands xmlns="http://schemas.microsoft.com/voicecommands/1.2">
  <CommandSet xml:lang="en-us" Name="en-us-CommandSet">
    <AppName>Adventure Works</AppName>
    <Example> Show my trip to San Francisco </Example>
    <Command Name="showTripToDestination">
      <Example>Show my trip to San Francisco</Example>
      <ListenFor>Show [my] trip to {destination}</ListenFor>
      <Feedback>Looking for trip to {destination}</Feedback>
      <Navigate/>
    </Command>
  </CommandSet>
</VoiceCommands>
<xml version="1.0" encoding="utf-8"?>
<VoiceCommands xmlns="http://schemas.microsoft.com/voicecommands/1.2">
  <CommandSet xml:lang="en-us" Name="en-us-CommandSet">
    <AppName>Quick Start</AppName>
    <Example>Ask me something</Example>
    <Command Name="showTripToDestination">
      <Example>Show my trip to San Francisco</Example>
      <ListenFor>show [my] trip to {destination}</ListenFor>
      <Feedback>Looking for trip to {destination}</Feedback>
      <Navigate/>
    </Command>
  </CommandSet>
</VoiceCommands>
<xml version="1.0" encoding="utf-8"/>
<VoiceCommands xmlns="http://schemas.microsoft.com/voicecommands/1.2">
    <CommandSet xml:lang="en-us" Name="en-us-CommandSet">
        <AppName> Adventure Works </AppName>
        <Example> Ask me something </Example>
        <Command Name="showTripToDestination">
            <Example> Show my trip to San Francisco </Example>
            <ListenFor> show [my] trip to {destination} </ListenFor>
            <Feedback> Looking for trip to {destination} </Feedback>
            <Navigate/>
        </Command>
        <AppName> Adventure Works </AppName>
    </CommandSet>
</VoiceCommands>
<?xml version="1.0" encoding="utf-8"?>
<VoiceCommands xmlns="http://schemas.microsoft.com/voicecommands/1.2">
  <CommandSet xml:lang="en-us" Name="en-us-CommandSet">
    <AppName>Adventure Works</AppName>
    <Example>Show my trips</Example>
    <Example>Show my trip to San Francisco</Example>
    <ListenFor>show [my] trip to {destination}</ListenFor>
    <Feedback>Looking for trip to {destination}</Feedback>
    <Navigate/>
  </CommandSet>
</VoiceCommands>
Voice Command Definition

```xml
<Command Name="showTripToDestination">
  <Example>Show my trip to San Francisco</Example>
  <ListenFor>show [my] trip to {destination}</ListenFor>
  <Feedback>Looking for trip to {destination}</Feedback>
  <Navigate/>
</Command>
```

**Command**
- Logical unit of user “intent”. Contains ...
  - ... What the user says to Cortana
  - ... What Cortana says in response
  - ... What action Cortana will perform, i.e.
    - <Navigate/> element indicates it is a foreground voice command

**Next Steps:**
- Register VCD with `Windows.ApplicationModel.VoiceCommands.VoiceCommandDefinitionManager`
- Override `App.OnActivated()` and launch your scenario based on context passed as name/value pairs
Cortana Extensibility

Completing Tasks in Cortana with Background Voice Commands
Completing tasks in Cortana

- Introducing new capabilities that enable apps to complete tasks within Cortana.
- Capabilities include:
  - APIs for task workflow
  - UI templates
  - Speech interactivity model
Background Voice Commands Architecture

Go to Adventure Works

Here is your trip to London

Microsoft Speech Platform Service (cloud)

Isolated Storage

VCD

IBackgroundTask

Implement Windows, AppModel.Background

XAML

Render

Application Code

Go to Adventure Works

Here's your trip to London

Adventure Works, show my trip to London

Which 'Vegas' trip do you wanna cancel?
Cancel this trip?

Vegas Tech Conference
May 14th 2015

Yes

Cancel this trip...

Cancelling this trip...

Disambiguation

Confirmation

Progress

Completion

Cancel my Adventure Works Trip

Vegas Tech Conference

Party in Phoenix

May 14th 2015

Yes

Cancel

Go to Adventure Works

I've cancelled it. [success-earcon]
Cortana Interaction APIs


**To:**

- Request a **disambiguation** screen in Cortana
- Request a **confirmation** screen in Cortana
- Request an **app launch** to complete the task
- Inform Cortana on status of an **in-progress** task
- Inform Cortana the task completed **successfully**
- Inform Cortana the task **failed** to complete

**Use this API:**

- `RequestDisambiguationAsync(VoiceCommandResponse)`
- `RequestConfirmationAsync(VoiceCommandResponse)`
- `RequestAppLaunchAsync(VoiceCommandResponse)`
- `ReportProgressAsync(VoiceCommandResponse)`
- `ReportSuccessAsync(VoiceCommandResponse)`
- `ReportFailureAsync(VoiceCommandResponse)`
Which one do you want to add to your trip?

VoiceCommandResponse
{
  VoiceCommandUserMessage Message;
  VoiceCommandUserMessage RepeatMessage;
  List<VoiceCommandContentTile> VoiceCommandContentTiles;
  string AppLaunchArgument;
}
Which one do you want to add to your trip?

Title Name

Text Line One

Text Line Two

Text Line Three

A
Change budget

Looking for groceries budget...

Change groceries budget to $200?

Your budget was updated. (success-earcon)
DEMO: Completing Tasks in Cortana

Background Voice Commands
In-App Speech Integration

Speech Synthesis (aka Text-to-Speech)
Speech Synthesis (aka Text-to-Speech)

- Speak with default speech settings
- Speak with any installed speech language
- Customize TTS voice with Speech Synthesis Markup Language (SSML)
- New since Windows & WP 8.1: Audio streams!

- Supported Languages in WP 8.1 & Windows 10
  - English (US, UK, India)
  - German
  - Spanish (Spain, Mexico)
  - French
  - Italian
  - Polish
  - Portuguese
  - Russian
  - Japanese
  - Chinese (Traditional, Simplified)
In Application Speech Synthesis

// Universal Windows 10, Windows & Windows Phone 8.1 Store App (WinRT)

// Speech Synthesis
<!-MediaElement in xaml file-->
<MediaElement Name="audioPlayer" AutoPlay="True" />

// C# code behind
// Function to speak a text string
private async void SpeakText(MediaElement audioPlayer, string textToSpeak)
{
    SpeechSynthesizer synthesizer = new SpeechSynthesizer();

    SpeechSynthesisStream ttsStream = await synthesizer.SynthesizeTextToStreamAsync(textToSpeak);

    audioPlayer.SetSource(ttsStream, ""); // This starts the player because AutoPlay="True"
In Application Speech Synthesis

// Universal Windows 10, Windows & Windows Phone 8.1 Store App (WinRT)

// Speech Synthesis
<!-[--MediaElement in xaml file-->
<MediaElement Name="audioPlayer" AutoPlay="True" .../>
</MediaElement>

// C# code behind
// Function to speak a text string
private async void SpeakText(MediaElement audioPlayer, string textToSpeak)
{
SpeechSynthesizer synthesizer = new SpeechSynthesizer();
SpeechSynthesisStream ttsStream = await synthesizer.SynthesizeTextToStreamAsync(textToSpeak);
audioPlayer.SetSource(ttsStream, ""); // This starts the player because AutoPlay="True"
}

// Synthesis
<!-[--MediaElement in xaml file-->
<MediaElement Name="audioPlayer" AutoPlay="True" .../>
</MediaElement>

Speech synthesis uses a MediaElement control to speak text
private async void SpeakText(MediaElement audioPlayer, string textToSpeak) {
    SpeechSynthesizer synthesizer = new SpeechSynthesizer();
    SpeechSynthesisStream ttsStream = await synthesizer.SynthesizeTextToStreamAsync(textToSpeak);
    audioPlayer.SetSource(ttsStream, "");
    // This starts the player because AutoPlay="True"
}
DEMO: Text-to-Speech

Speech Synthesis in Universal Windows Apps
In-App Speech Integration

Speech Recognition & Grammars
In-App Speech Recognition

• Check if the interaction was input through voice or text
• For voice, use Speech to continue the user interaction in the app
• Use Windows.Media.SpeechRecognition for speech recognition
• Use Windows.Media.SpeechSynthesis APIs to reply/talk to the user
On-Demand Speech Recognition

- Use `SpeechRecognizer.RecognizeWithUIAsync` to present the user with System listening GUI that can be customized.
- Use `SpeechRecognizer.RecognizeAsync` for on-demand recognition and provide a listening UI experience.
- System predefined grammars: Dictation, Web Search, Form Filling.
- App grammars: a list of strings or SRGS file.
Continuous Speech Recognition

- `SpeechRecognizer.ContinuousRecognitionSession` enables listening continuously

- Can be used with Dictation grammar for scenarios where users speak for long periods

- Can be used with app-provided grammars for always-listening in-app commands
DEMO: Speech Interactions

In-App Speech Recognition & Grammars
Conversational Writing Guidance

• **Be efficient.** Less is more!
• **Be relevant.** Provide information that’s relevant to the given task, content, and context.
• **Be clear.** Avoid ambiguity. Use everyday language.
• **Be trustworthy.** Present accurate information. Be transparent. Respect privacy.
Localization

- For each supported language, localize the CommandSet in the Voice Command Definition
- When launched, inspect the value of Language property of the VoiceCommandServiceConnection
- Localize the dialog on the app provided screens to match Cortana’s language
Voice Command Handling Guidance

**Foreground**
- for tasks that require complex interactions
- for tasks that demand the user’s attention for a long time

**Background**
- for simple tasks that don’t require additional input beyond Confirmation and Disambiguation
- for tasks that complete within seconds
- to *initiate* longer tasks that execute remotely
- may use `RequestAppLaunchAsync` method to launch the app and complete remainder of the flow
New: Proactive Actions

- **Proactively suggests actions** to users, at just the right time, to drive usage of your apps and websites

- **Easy to register actions** in Cortana portal

- **Works everywhere Cortana is** starting with Windows 10 and Android
Microsoft Cognitive Services: Speech

**Speech API**
Communicate with your users with speech recognition and synthesis powered by Bing

**Speaker Recognition API**
Recognize your users from their voice

**Custom Recognition Intelligent Service**
Customize both language and acoustic models for better speech recognition tailored to your application

https://www.microsoft.com/cognitive-services/
Microsoft Cognitive Services: Language

Spell Check API
Detect and correct common and uncommon spelling errors, via the Bing document index

Language Understanding Intelligent Service
Understand natural language commands tailored to your application

Web Language Model API
Leverage the power of language models trained on web-scale data

https://www.microsoft.com/cognitive-services/
Call to Action: Speech-enable your Apps!

• Speech is a convenience, it’s fun, and it’s also a responsibility!
• Speech synthesis is the low-hanging fruit
• Deep linking to frequently accessed scenarios via Voice Commands
• Let users complete high-value tasks in Cortana using background Voice Commands
• Continue the conversation in the app with Speech Recognition
• Augment the value of Voice Commands with backend services
Online Microsoft training delivered by experts to help technologists continually learn.

Hundreds of courses for developers, IT Pros, students, entrepreneurs and enthusiasts.

11 different languages

3M+ students registered

Build your own Learning Plan

All free!

http://www.microsoftvirtualacademy.com
My MVA Course Online

• Universal Windows App Development with Cortana and the Speech SDK
  • 01 | Introducing Cortana & Getting Started with Speech
  • 02 | Using Speech Synthesis in Mobile Apps
  • 03 | Cortana Integration & Voice Commands
  • 04 | Speech Recognition in Apps
  • 05 | Designing Apps with Speech
  • 06 | Advanced Speech Topics

• Available for on-demand viewing now:
  • Microsoft Virtual Academy: http://aka.ms/CortanaMVA
More Cortana Dev on Windows 10

• Channel 9 Show
• Visual Studio Toolbox with Robert Green
  • New Voice Commands
  • Integration with Cortana’s canvas
  • Background Voice Commands
  • Continuous dictation
  • Poutine in Montreal!

Thank You!

Slides are in SlideShare. Demos are on GitHub.
Contact me and let me know what you build, I will be happy to help promote your apps.

Blog: www.AgeofMobility.com
Twitter: @ActiveNick
Email: nick.landry@microsoft.com
Mobile Apps: www.bigbaldapps.com
LinkedIn: www.linkedin.com/in/activenick
GitHub: github.com/ActiveNick
Slideshare: www.slideshare.net/ActiveNick