Chapter No. 1
"Speech on Android Devices"
In this package, you will find:
A Biography of the authors of the book
A preview chapter from the book, Chapter NO.1 "Speech on Android Devices"
A synopsis of the book’s content
Information on where to buy this book

About the Authors

Michael McTeer is Emeritus Professor of Knowledge Engineering at the University of Ulster with a special research interest in spoken language technologies. He graduated in German Language and Literature from Queens University Belfast in 1965, was awarded MA in Linguistics at University of Essex in 1975, and a PhD at the University of Ulster in 1981. He has been Visiting Professor at the University of Hawaii (1986-87), the University of Koblenz, Germany (1994-95), and University of Granada, Spain (2006-2010). He has been researching in the field of spoken dialogue systems for more than 15 years and is the author of the widely used text book Spoken Dialogue Technology: Toward the Conversational User Interface (Springer Verlag, 2004). He also is a co-author of the book Spoken Dialogue Systems (Morgan and Claypool, 2010).

Michael has delivered keynote addresses at many conferences and workshops, including the EU funded DUMAS Workshop, Geneva, 2004, the SIGDial workshop, Lisbon, 2005, the Spanish Conference on Natural Language Processing (SEPLN), Granada, 2005, and has delivered invited tutorials at IEEE/ACL Conference on Spoken Language Technologies, Aruba, 2006, and ACL 2007, Prague. He has presented on several occasions at SpeechTEK, a conference for speech technology professionals, in New York and London. He is a certified VoiceXML developer and has taught VoiceXML at training courses to professionals from companies including Genesys, Oracle, Orange, 3, Fujitsu, and Santander. He was the main developer of the VoiceXML-based home monitoring system for patients with type-2 diabetes, currently in use at the Ulster Hospital, Northern Ireland.

For More Information:
Zoraida Callejas is Assistant Professor at the University of Granada, Spain, where she has been teaching several subjects related to Oral and Multimodal Interfaces, Object Oriented Programming, and Software Engineering for the last eight years. She graduated in Computer Science in 2005, and was awarded a PhD in 2008 from the University of Granada. She has been Visiting Professor in Technical University of Liberec, Czech Republic (2007-13), University of Trento, Italy (2008), University of Ulster, Northern Ireland (2009), Technical University of Berlin, Germany (2010), University of Ulm, Germany (2012), and Telecom ParisTech, France (2013).

Zoraida focuses her research on speech technology and in particular, on spoken and multimodal dialogue systems. Zoraida has made presentations at the main conferences in the area of dialogue systems, and has published her research in several international journals and books. She has also coordinated training courses in the development of interactive speech processing systems, and has regularly taught object-oriented software development in Java in different graduate courses for nine years. Currently, she leads a local project for the development of Android speech applications for intellectually disabled users.

For More Information:
Voice Application Development for Android

The idea of being able to talk with a computer has fascinated many people for a long time. However, until recently, this has seemed to be the stuff of science fiction. Now things have changed so that people who own a smartphone or tablet can perform many tasks on their device using voice—you can send a text message, update your calendar, set an alarm, and ask the sorts of queries that you would previously have typed into your search box. Often voice input is more convenient, especially on small devices where physical limitations make typing and tapping more difficult.

This book provides a practical guide to the development of voice apps for Android devices, using the Google Speech APIs for text-to-speech (TTS) and automated speech recognition (ASR) as well as other open source software. Although there are many books that cover Android programming in general, there is no single source that deals comprehensively with the development of voice-based applications for Android.

Developing for a voice user interface shares many of the characteristics of developing for more traditional interfaces, but there are also ways in which voice application development has its own specific requirements and it is important that developers coming to this area are aware of common pitfalls and difficulties. This book provides some introductory material to cover those aspects that may not be familiar to professionals from a mainstream computing background. It then goes on to show in detail how to put together complete apps, beginning with simple programs and progressing to more sophisticated applications. By building on the examples in the book and experimenting with the techniques described, you will be able to bring the power of voice to your Android apps, making them smarter and more intuitive, and boosting your users' mobile experience.

What This Book Covers

Chapter 1, Speech on Android Devices, discusses how speech can be used on Android devices and outlines the technologies involved.

Chapter 2, Text-to-Speech Synthesis, covers the technology of text-to-speech synthesis and how to use the Google TTS engine.

Chapter 3, Speech Recognition, provides an overview of the technology of speech recognition and how to use the Google Speech to Text engine.

Chapter 4, Simple Voice Interactions, shows how to build simple interactions in which the user and app can talk to each other to retrieve some information or perform an action.

For More Information:  
Chapter 5, *Form-filling Dialogs*, illustrates how to create voice-enabled dialogs that are similar to form-filling in a traditional web application.

Chapter 6, *Grammars for Dialog*, introduces the use of grammars to interpret inputs from the user that go beyond single words and phrases.

Chapter 7, *Multilingual and Multimodal Dialogs*, looks at how to build apps that use different languages and modalities.

Chapter 8, *Dialogs with Virtual Personal Assistants*, shows how to build a speech-enabled personal assistant.

For More Information:  
Speech on Android Devices

Have you ever wanted to create voice-based apps that you could run on your own Android device; apps that you could talk to and that could talk back to you? This chapter provides an introduction to the use of speech on Android devices, using open-source APIs from Google for text-to-speech synthesis and speech recognition. Following a brief overview of the world of Voice User Interfaces (VUIs), the chapter outlines the components of an interactive voice application (or virtual personal assistant).

By the end of this chapter you should have a good understanding of what is required to create a voice-based app using freely available resources from Google.

Using speech on an Android device

Android devices provide built-in speech-to-text and text-to-speech capabilities. The following are some examples of speech-based apps on Android:

Speech-to-text

With speech-to-text users of Android devices can dictate into any text box on the device where textual input is required, for example, e-mail, text messaging, and search. The keyboard control contains a button with a microphone symbol and two letters indicating the language input settings, which can be changed by the user. On pressing the microphone button a window pops up asking the user to Speak Now. The spoken input is automatically transcribed into written text. The user can then decide what to do with the transcribed text.

For More Information:
Accuracy rates have improved considerably for dictation on small devices, on one hand due to the use of large-scale cloud-based resources for speech recognition, and on the other, to the fact that the device is usually held close to the user's mouth so that a more reliable acoustic signal can be obtained. One of the main challenges for voice dictation is that the input is unpredictable—users can say literally anything—and so a large general vocabulary is required to cover all possible inputs. Other challenges include dealing with background noise, sloppy speech, and unfamiliar accents.

**Text-to-speech**

Text-to-speech (TTS) is used to convert text to speech. Various applications can take advantage of TTS. For example, TalkBack, which is available through the Accessibility option, uses TTS to help blind and visually impaired users by describing what items are touched, selected and activated. TalkBack can also be used to read a book in the Google Play Books app. The TTS function is also available on Android Kindle as well as on Google Maps for giving step-by-step driving instructions. There is a wide range of third-party apps that make use of TTS, and alternative TTS engines are also available.

**Voice Search**

Voice Search provides the same functionality on Android devices as the traditional Google Search except that instead of typing a query the user speaks it. Voice Search is available using the microphone in the Google Search widget. In Voice Search the recognized text is passed to the search engine and executed in the same way that a typed query is executed.

A new feature of Voice Search is that, in addition to returning a list of links, a spoken response to the query is returned. For example, in response to the question "How tall is the Eiffel tower?", the app replies, "The Eiffel tower is 324 meters tall." It is also possible to ask follow-up questions using pronouns, for example, "When was it built?". This additional functionality is made possible by combining Google's Knowledge Graph—a knowledge base used by Google—with its conversational search technology to provide a more conversational style of interaction.

**Android Voice Actions**

Android Voice Actions can also be accessed using the microphone in the Google Search widget. Voice Actions allow the user to control their device using voice commands. Voice Actions require input that matches a particular structure, as shown in the following list from Google's webpage: [http://www.google.co.uk/intl/en_uk/mobile/voice-actions/](http://www.google.co.uk/intl/en_uk/mobile/voice-actions/). Note: items with * are optional. Italicized items are the words to be spoken.

---

For More Information:

<table>
<thead>
<tr>
<th>Voice Action</th>
<th>Structure</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Send text messages</td>
<td>send text to [recipient] [message]*</td>
<td>send text to Allison Miller Running late. I will be home around 9</td>
</tr>
<tr>
<td>Call businesses</td>
<td>call [business name] [location]*</td>
<td>call Soho Pizzeria London</td>
</tr>
<tr>
<td>View a map</td>
<td>map of [address/city]</td>
<td>map of London</td>
</tr>
<tr>
<td>Search Google</td>
<td>[your query]</td>
<td>pictures of Stonehenge at sunset</td>
</tr>
<tr>
<td>Get directions</td>
<td>navigate to [address/city/ business name]</td>
<td>navigate to British Museum London or navigate to 24 Mill Street</td>
</tr>
<tr>
<td>Call contacts</td>
<td>call [contact name] [phone type]*</td>
<td>call Allison Miller home</td>
</tr>
<tr>
<td>Go to websites</td>
<td>go to [website]</td>
<td>go to Wikipedia</td>
</tr>
</tbody>
</table>

The structures in Voice Actions allow them to be mapped on to actions that are available on the device. For example, the keyword call indicates a phone call while the key phrase go to indicates a website to be launched. Additional processing is required to extract the parameters of the actions, such as contact name and website.

**Virtual Personal Assistants**

One of the most exciting speech-based apps is the Virtual Personal Assistant (VPA), which acts like a personal assistant, performing a range of tasks such as finding information about local restaurants; carrying out commands involving apps on the device, for example, using speech to set the alarm or update the calendar; and engaging in general conversation. There are at least 20 VPAs available for Android devices (see the web page for this book) although the best-known VPA is Siri, which has been available on the iPhone iOS since 2011. You can find examples of interactions with Siri that are similar to those performed by Android VPAs on Apple's website [http://www.apple.com/uk/ios/siri/](http://www.apple.com/uk/ios/siri/). Many VPAs, including Siri, have been created with a personality and an ability to respond in a humorous way to trick questions and dubious input, thus adding to their entertainment value. See examples at [http://www.sirifunny.com](http://www.sirifunny.com) as well as numerous video clips on YouTube.
It is worth mentioning that a number of technologies share some of the characteristics of VPAs as explained in the following:

**Dialog systems**, which have a long tradition in academic research, are based on the vision of developing systems that can communicate with humans in natural language (initially written text but more recently speech). The first systems were concerned with obtaining information, for example, flight times or stock quotes. The next generation enabled users to engage in some form of transaction, in banking or making a travel reservation, while more recent systems are being developed to assist in troubleshooting, for example, guiding a user who is having difficulty setting up some item of equipment. A wide range of techniques have been used to implement dialog systems, including rule-based and statistically-based dialog processing.

**Voice User Interfaces (VUIs)**, which are similar to dialog systems but with the emphasis on commercial deployment. Here the focus has tended to be on systems for specific purposes, such as call routing, directory assistance, and transactional dialogs for example, travel, hotel, flight, car rental, or bank balance. Many current VUIs have been designed using VoiceXML, a markup language based on XML. The VoiceXML scripts are then interpreted on a voice browser that also provides the required speech and telephony functions.

**Chatbots**, which have been used traditionally to simulate human conversation. The earliest chatbots go back to the 1960s with the famous ELIZA program written by Joseph Weizenbaum that simulated a Rogerian psychotherapist—often in a convincing way. More recently chatbots have been used in education, information retrieval, business, e-commerce, and in automated help desks. Chatbots use a sophisticated pattern-matching algorithm to match the user's input and to retrieve appropriate responses. Most chatbots have been text-based although increasingly speech-based chatbots are beginning to emerge (see further in Chapter 8, Dialogs with Virtual Personal Assistants).

**Embodied conversational agents (ECAs)**, are computer-generated animated characters that combine facial expression, body stance, hand gestures, and speech to provide an enriched channel of communication. By enhancing the visual dimensions of face-to-face interaction embodied conversational agents can appear more trustworthy and believable, and also more interesting and entertaining. Embodied conversational agents have been used in applications such as interactive language learning, virtual training environments, virtual reality game shows, and interactive fiction and storytelling systems. Increasingly they are being used in e-commerce and e-banking to provide friendly and helpful automated help. See, for example, the agent Anna at the IKEA website http://www.ikea.com/gb/en/.

For More Information:
Virtual Personal Assistants differ from these technologies in that they allow users to use speech to perform many of the functions that are available on mobile devices, such as sending a text message, consulting and updating the calendar, or setting an alarm. They also provide access to web services, such as finding a restaurant, tracking a delivery, booking a flight, or using information services such as Knowledge Graph, Wolfram Alpha, or Wikipedia. Because they have access to contextual information on the device such as the user's location, time and date, contacts, and calendar, the VPA can provide information such as restaurant recommendations relevant to the user's location and preferences.

Designing and developing a speech app
Speech app design shares many of the characteristics of software design in general, but there are also some aspects unique to voice interfaces—for example, dealing with the issue that speech recognition is always going to be less than 100 percent accurate, and so is less reliable compared with input when using a GUI. Another issue is that, since speech is transient, especially on devices with no visual display, greater demands are put on the user's memory compared with a GUI app.

There are many factors that contribute to the usability of a speech-based app. It is important to perform extensive use case analysis in order to determine the requirements of the system, looking at issues such as whether the app is to replace or complement an existing app; whether speech is appropriate as a medium for input/output; the type of service to be provided by the app; the types of user who will make use of the app; and the general deployment environment for the app.

Why Google speech?
The following are our reasons for using Google speech:

- **The proliferation of Android devices**: Recent information on Android states that "Android had a worldwide smartphone market share of 75% during the third quarter of 2012, with 750 million devices activated in total and 1.5 million activations per day." (From [http://www.idc.com/getdoc.jsp?containerId=prUS23771812](http://www.idc.com/getdoc.jsp?containerId=prUS23771812) Retrieved 09/07/2013).

- **The Android SDK is open source**: The fact that the Android SDK is open source makes it more easily available for developers and enthusiasts to create apps, compared with some other operating systems. Anyone can develop their own apps using a free development environment such as Eclipse and then upload it to their Android device for their own personal use and enjoyment.

---

For More Information:  
• **The Google Speech APIs**: The Google Speech APIs are available for free for use on Android devices. This means that the Speech APIs are useful for developers wishing to try out speech without investing in expensive commercially available alternatives. As Google employs many of the top speech scientists, their speech APIs are comparable in performance to those on offer commercially.

**You may also try...**

Nuance NDEV Mobile, which supports a number of languages for text-to-speech synthesis and speech recognition as well as providing a PhoneGap plug-in to enable developers to implement their apps on different platforms (http://dragonmobile.nuancemobiledeveloper.com).

The AT&T Speech Mashup (http://www.research.att.com/projects/SpeechMashup/), which supports the development of speech-based apps and the use of W3C standard speech recognition grammars.

**What is needed to create a Virtual Personal Assistant?**

The following figure shows the various components required to build a speech-enabled VPA.

---

**For More Information:**

A basic requirement for a VPA is that it should be able to speak and to understand speech. Text to speech synthesis, which provides the ability to speak, is discussed in Chapter 2, Text To Speech Synthesis, while speech recognition is covered in Chapter 3, Speech Recognition. However, while these capabilities are fundamental for a voice-enabled assistant, they are not sufficient. The ability to engage in dialog and connect to web services and device functions is also required as the basis of personal assistance. To do these things a VPA requires the following:

- A method for controlling the dialog, determining who should take the dialog initiative and what topics they should cover. In practice this can be simplified by having one-shot interactions in which the user simply speaks their query and the app responds. One-shot interactions are covered in Chapter 4, Simple Voice Interactions. System-directed dialogs, in which the app asks a series of questions—as in web-based form-filling (for example, to book a hotel or rent a car), are covered in Chapter 5, Form-filling Dialogs.

- A method for interpreting the user’s input once it has been recognized. This is the task of the Spoken Language Understanding component which, among other things, provides a semantic interpretation representing the meaning of what the user said. Since in many commercial systems input is restricted to single words or phrases, the interpretation is relatively straightforward. Two different approaches will be illustrated in Chapter 6, Grammars for Dialog: how to create a hand-crafted grammar that covers the words and phrases that the user might say; and how to use statistical grammars to cover a wider range of inputs and to provide a more robust interpretation. It also provides different modalities if speech input and output is not possible or performance is poor. A VPA should also have the ability to use different languages, if required. These topics are covered in Chapter 7, Multilingual and Multimodal Dialogs.

- Determining relevant actions and generating appropriate responses. These aspects of dialog management and response generation are described in Chapter 7, Multilingual and Multimodal Dialogs, and in Chapter 8, Dialogs with Personal Virtual Assistants.

Building on the basic technologies of text-to-speech synthesis and speech recognition, as presented in Chapter 2 and Chapter 3, Chapters 4-8 cover a range of techniques that will enable developers to take the basic technologies further and create speech-based apps using the Google speech APIs.
Summary

This chapter has provided an introduction to speech technology on Android devices. We examined various types of speech app that are currently available on Android devices. We also looked at why we decided to focus on Google Speech APIs as tools for the developer. Finally we introduced the main technologies required to create a Virtual Personal Assistant. These technologies will be covered in the remaining chapters of this book.

We will introduce you to text-to-speech synthesis (TTS) and show how to use the Google TTS API to develop applications that speak in the next chapter.
Where to buy this book


Free shipping to the US, UK, Europe and selected Asian countries. For more information, please read our shipping policy.

Alternatively, you can buy the book from Amazon, BN.com, Computer Manuals and most internet book retailers.