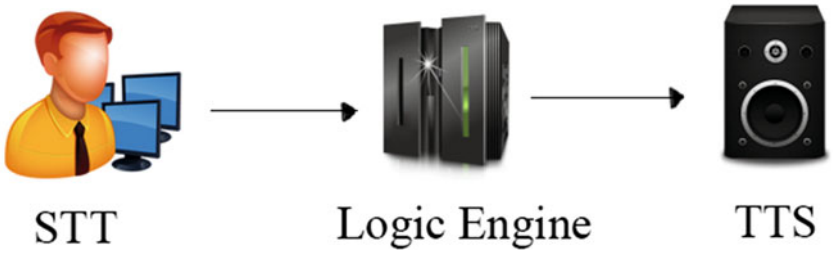

Figure 1-1. *Virtual assistant workflow*

Speech-to-Text Engine

As the name suggests, the STT engine converts the user's speech into a text string that can be processed by the logic engine. This involves recording the user's voice, capturing the words from the recording (cancelling any noise and fixing distortion in the process), and then using natural language processing (NLP) to convert the recording to a text string.

Logic Engine

Melissa's logic engine is the software component that receives the text string from the STT engine and handles the input by processing it and passing the output to the TTS engine. The logic engine can be considered Melissa's brain; it handles user queries via a series of `if-then-else` clauses in the Python programming language. It decides what the output should be in response to specific inputs. You build Melissa's logic engine throughout the book, improving it and adding new functionalities and features as you go.

Text-to-Speech Engine

This component receives the output from Melissa's logic engine and converts the string to speech to complete the interaction with the user. TTS is crucial for making Melissa more humane, compared to giving confirmation via text.

This three-component system removes any physical interaction between the user and the machine; the users can interact with their system the same way they interact with other human beings. You learn more about the STT and TTS engines and how to implement them in Chapter 2.

From a high-level view, these are the three basic components that make up Melissa. This book shows you how to do all the necessary programming to develop them and put them together.