TIPS FOR TEXT-TO-SPEECH PRONUNCIATION

The digital voices available for text-to-speech are excellent but not perfect. Some ways to correct common mispronunciation issues are provided in this article.

By John Krout, Potomac Area Technology And Computer Society (www.patacs.org)

Introduction

In 2025, I have tested and used text-to-speech capabilities on Windows, Android and Apple iOS. All provide excellent quality English voices and some features for controlling the speed of pronunciation. The speed control is very useful for making the pace of the spoken words understandable for you.

All provide a text entry method, and some way to hear the entered text read aloud by the text-to-speech capability. In addition, all provide a way to record the text read aloud into a sound file.

However, I found some cases of text-to-speech mispronunciations, as well as workarounds that correct those cases. Here you will learn the fairly frequent problems and the workarounds that have worked for me.

I tested three text-to-speech capabilities: Windows Narrator on a Windows 10 computer, a text-to-speech app using the Google digital voice engine on my Samsung Galaxy S20 FE smartphone running Android 13, and a text-to-speech app on my retired iPhone X using Apple digital voices and running iOS 16.7.

None of these text-to-speech capabilities require an Internet connection to work. I believe none are artificial intelligence (AI) clients because they work without an Internet connection. However, it seems possible that the digital voices used have been tweaked over time by AI engines.

Acronyms

For decades, acronyms have found wide use in corporations and government agencies including the military. The federal agency known as NASA is actually named the National Aeronautics and Space Administration.

Acronyms have mushroomed as texts and tweets have become popular, providing a quick way to communicate common phrases. Examples are IDK for I don’t know and LOL for lots of laughs.

Even in user groups, acronyms are common. Special Interest Groups (SIGs), describing sub-groups for devotees of a particular application, a particular use such as digital photography, or particular operating systems, has been widely used for decades.

Some examples are now widely recognized in the 21st century automobile business: ICE for Internal Combustion Engine, for cars using gasoline a fuel, EV for Electric Vehicles, and PHEV for Pluggable Hybrid Electric Vehicles. EVs get far greater coverage in general media, but PHEVs have consistently outsold EVs in the United States for good reasons. I started driving a PHEV in 2015, and I am now driving my second one.

Acronyms are a persistent challenge for text-to-speech. I have found two workarounds.

Acronyms pronounced as individual letters

One easy workaround is for acronyms that are pronounced as individual letters, such as APCUG, USAF, IRS, ASAP, EV and ISO. For digital photographers, the ISO is a scale that specifies how the digital image sensor sensitivity can be adjusted. ISO stands for the International Standards Organization that adopted the linear sensitivity scale originally defined decades ago in the film era by the American Standards Association (ASA).

Many corporation acronyms for senior executives are pronounced as individual letters, such as CEO (chief executive officer), CFO (chief financial officer), and CTO (chief technology officer).

The workaround for text-to-speech is to type a space between each acronym letter. That tactic forces text-to-speech to pronounce each letter by itself. So, for instance, In the text entry field, APCUG becomes A P C U G, and USAF becomes U S A F, IRS become I R S, ASAP becomes A S A P, EV becomes E V, and ISO becomes I S O.

Most of the acronyms used in texts and tweets are also pronounced as individual letters.

Windows Narrator pronounced all workarounds as desired. Windows Narrators also pronounced ASAP without spaces as the desired A S A P and ISO without spaces as I S O.

The Google Engine US English voices for Android pronounced all these workarounds as desired.

The Apple US English voices did fine with two of the four acronym workarounds. It did not pronounce the A as a long A in A P C U G and U S A F and A S A P. When I changed A to Aa, the Apple US English voices did pronounce a long A for each workaround: Aa P C UG and U S Aa F and Aa S Aa P

Both the Google Engine US English voices and the Apple US English voices pronounced ASAP as A SAP, which I think is recognized as an alternative pronunciation.

Acronyms pronounced as words

Many acronyms are pronounced like words, instead of individual letters.

One of the most interesting examples is SCUBA, meaning Self-Contained Underwater Breathing Apparatus. The phrase is so long that few use it and some might even not know it.

Others include digital image file types, such as GIF (graphics image format), JPEG (joint photo engineering group), and MPEG (motion picture engineering group).

I tested GIF, MPEG and JPEG. All three text-to speech systems pronounced those correctly. The iPhone Apple system pronounced GIF with a hard G sound, which I think is an acceptable pronunciation, though the other two pronounced with a soft G sound, which I think is the majority pronunciation.

I tested SCUBA and NASA. All three text-to-speech capabilities tested pronounced both correctly. I believe it is reasonable for US English voices to have been taught the pronunciation of such acronyms that have become widely used words.

Another more recent but very widely used acronym is **CAPTCHA**, which stands for Completely Automated Public Turing Test to tell Computers and Humans Apart. Anyone using a Web site that involves a login may have been required to endure a CAPTCHA test, to identify crosswalks or other features in photo thumbnails.

All three text-to-speech systems pronounced CAPTCHA correctly.

Other examples involve acronyms that are far less widely known. We cannot expect the text-to-speech systems to have been taught the correct pronunciation of these.

One is the acronym for my local user group: PATACS. It is pronounced as two words. The group has not adopted an official pronunciation.

Most members pronounce the group acronym as PAT ACES, but a few pronounce it PAY TAXES. The group is located in the Virginia suburbs of Washington DC. I suppose it is possible the second pronunciation was adopted by members who were IRS or US Treasury employees.

As you can guess, the workaround for text-to-speech is to type the words. Then the text-to-speech accurately pronounces the acronym. That worked for Narrator and the Android and Apple text-to-speech apps.

Proper names

I have found some proper names are not pronounced correctly by text-to-speech. Here are a couple of examples.

One of my long-term acquaintances in my local user group is a man who introduces himself using the short form of his first name: Geof. I think that spelling is more common in the United Kingdom than in the United States.

Windows Narrator pronounced Geof as Gee Off.

On the other hand, the smartphone text-to-speech had no trouble. The Google Engine English US voices on my Android smart phone pronounced Geof as Jeff. The Apple US English voices on my iPhone X pronounced Geof as Jeff.

As you might expect, in Windows Narrator, the workaround is to spell the first name as Jeff.

I found a much tougher issue with the family name of one of my high school classmates. He lives relatively near me in Northern Virginia. Decades after high school, the two of us learned through online family genealogy research that we are distant cousins due to an 1810 marriage of a woman from his family to my great-great-grandfather.

His family name is **Anstine**. The proper pronunciation is two syllables, sounding like this: Ann Stein.

The Google Engine English US voices on my Android smartphone pronounced that family name as three syllables: **On Stin Ee**. Those voices pronounced the workaround Anne Stein correctly. Those same voices also pronounced the spelling Anstein the same way.

The Apple US English voices got closer, pronouncing the proper Anstine spelling as Ann steen. For the Apple voices, I used the spelling Ann Stein to obtain correct pronunciation.

Windows Narrator pronounced Anstine correctly. This is an example why having more than one text-to-speech capability to try is useful.

In general, phonetic workarounds are often successful.

For proper names that originate from other parts of the world, considering a voice from that other place. Those other voices can easily be used on smartphones but might possibly use diacritical marks, meaning added symbols to alter pronunciation of a letter, or different alphabets.

ABOUT THE AUTHOR: John Krout is a retired software engineer. He has been writing about and delivering presentations on interesting uses of personal computers since the early 1980s. In the 21st century, as digital tech became more powerful and widespread, he has also been writing and delivering presentations about interesting uses of

smartphones, tablets, digital music, digital photography, Electric Vehicles (EVs) and Pluggable Hybrid Electric Vehicles (PHEVs). He lives in Arlington Virginia.