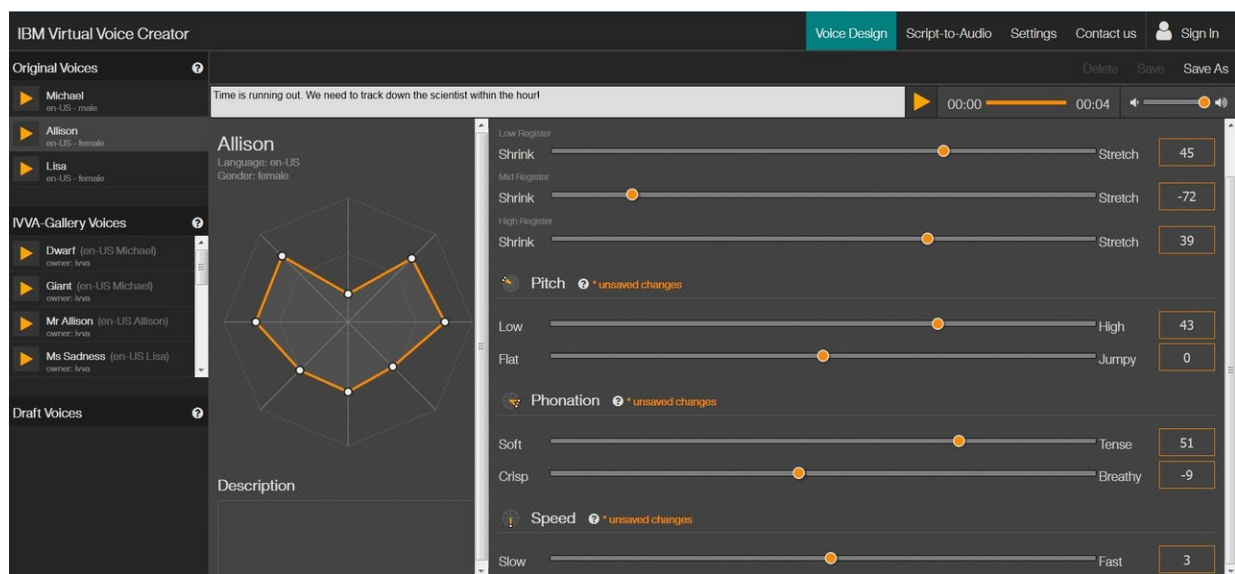


New text-to-speech tool for DIY voiceovers—from soft, sad and sultry to scary

November 1 2017, by Ron Hoory And Alex Sorin



IBM Virtual Voice Creator. Credit: IBM

The animation world is rich in lovable and memorable characters, each with its own unique voice and personality—and animators, writers, and designers keep coming up with even more new games, film ideas, villains, and heroes. Creating voiceovers for these characters is a time-consuming and expensive process that often involves holding auditions for voice actors, and studio time to record.

What if you could take a text, generate speech, and go on from there to create a whole bunch of new voices – just by changing different vocal aspects such as pitch, rhythm, timbre, etc.? My team at IBM Research-Haifa is building on top of Watson text-to-speech technology to create customizable voices. Our vision for a solution to easily create new, distinct, expressive voices, led us to develop an [automated voice creation process](#) that is fast and flexible.

Our vision comes to life

Our vision has already come to life in cooperation with Sesame Street and the IBM Research Education team. We participated in an IBM-Sesame Street pilot at Georgia's Gwinnett County Public Schools in April-May 2017. Sesame Workshop content and Watson Education technology were introduced into classrooms for the first time, using an app for learning new vocabulary. Our challenge was to synthesize voices for new Sesame characters that will make kids smile, similar to familiar characters like Ernie, Big Bird, and Elmo.

Using voice as a tool

The IBM Virtual Voice Creator is a web-based tool that starts with three standard text-to-speech voices available for American English at WDC TTS service. Using the tool, we can change different parameters and transform these standard voices into new virtual voices.

Think of it as a kind of a mixing console like sound engineers use – but for [voice](#) manipulation. Sliders control and change each different vocal aspect, such as pitch, speed, timbre, and breathiness, and can apply them in endless combinations. The GUI also creates a "visual signature" of the parameter controls, with a visual representation that changes its shape as the parameters are manipulated. It's easy to play around and create new

voices. Happy Lisa can quickly turn into a wicked witch, and sullen Michael can be recreated as a cheerful little boy.

Animators can use this tool to create new voices for game characters or cartoon heroes. All they have to do is choose a standard voice, and play around with the sliders, using their imagination to shape the voice persona of the new character. Then, they can add in the text to get the audio output in their "new" voice for the soundtrack without any need for voice actors and recording studios.

The technology enables emerging games where the scripts are generated on-line, and the audio cannot be recorded in advance.

The entertainment field is just one example. We're very excited about the possibilities of applying this text-to-speech technology in any context that needs multiple distinct voices created on-demand, such as education, advertising, and more.

Provided by IBM

Citation: New text-to-speech tool for DIY voiceovers—from soft, sad and sultry to scary (2017, November 1) retrieved 25 April 2024 from <https://phys.org/news/2017-11-text-to-speech-tool-diy-voiceoversfrom-soft.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.