ELEARNINGNEWS ARTICLE

Year 6 - number 192 Wednesday 2 March 2022

The evolution of synthetic voices: opportunities and risks

Advances in the application of deep learning algorithms make it possible to produce artificial voices that are increasingly more lifelike than human language. An interesting development also for the creation of eLearning courses.

Synthetic voices, also called "digital voices" or "artificial voices", are voices created by specific software called "speech engines" that have the characteristic of "sounding human".

We know how crucial voice and voice-over in online courses are (and that there are many choices to be made in this respect, as explained in the past in the article **How to choose the voice of your online course**).

Synthesized voices, which are based on systems capable of emitting a sequence of pre-recorded words (with often unnatural results) and usually employed in all IT tools (think Siri or Alexa), are on the rise. That's because, in addition to being cheap and scalable, they're evolving rapidly.

In the past, adding naturalness to the artificial voice required manual work to harmonize timbre, intensity and intonation. Today, innovation brought by **deep learning and AI** allows you to configure the exact rhythm, pronunciation or intonation automatically. How? By using the recording of a real voice to train an algorithm: in this way, the algorithm will learn to read a written text in a more natural way based on reading rules applied by a real human being.

The business opportunities of synthetic voice

In addition to concerns about the possible fraudulent use of artificial voices, for example in telephone calls (a phenomenon known as deepfake), the interest of companies in the opportunities for legitimate use of synthetic voices is growing, also thanks to the improvement in the quality of voice-over achievable.

Many subtleties of human speech (pauses, breath, intonation, expressed emotion) are now automatically reproducible by a digital voice, especially when employed in short audio clips, are very lifelike. In general, the matter changes when they speak for too long: in that case it is easier to distinguish them from real human voices.

Other "plusses" of digital voices are certainly their **affordability, scalability and ease of use**. Note that updating a synthetic voiceover is much easier than updating a recording produced with a human voice actor: in the field of **online training**, this is a key point.

In addition to the eLearning industry, digital voices are increasingly popular among companies that have a strong customer interaction, because they allow them to present themselves with a voice that is credible, consistent, and in line with the corporate message.

Creating a synthetic voice: risks and limitations

As we mentioned, producing a convincing (i.e. expressive and contextualized) synthetic voice is automatic, but not simple: you have to record at least one or two hours of human audio with the chosen voice actors to provide AI with training data and refine deep learning models (a process that takes weeks of work to arrive at a realistic synthetic replica).

This is not to say that synthetic voices are **taking the place of voice actors** who, in fact, may benefit from the situation. In fact, if a company uses a voice actor's voice as the basis for generating new audio, it pays the human owner of the voice a royalty and, then, in the event of successful sales of the digital voice made from its own, the voice actor receives economic rewards tied to the volume of business generated.

In addition, synthetic voices still have **some non-negligible limitations**

- difficulty in maintaining voice realism over long periods of time (required for the creation of an audiobook or podcast, for example);
- limited ability to control voice characteristics (as, instead, a director could do with a human actor).

This puts human voice actors in the position of being the only ones who can interpret expressive, creative and long-lasting projects.

Translated with www.DeepL.com/Translator