

# A new model to synthesize emotional speech for companion robots

October 6 2021, by Ingrid Fadelli

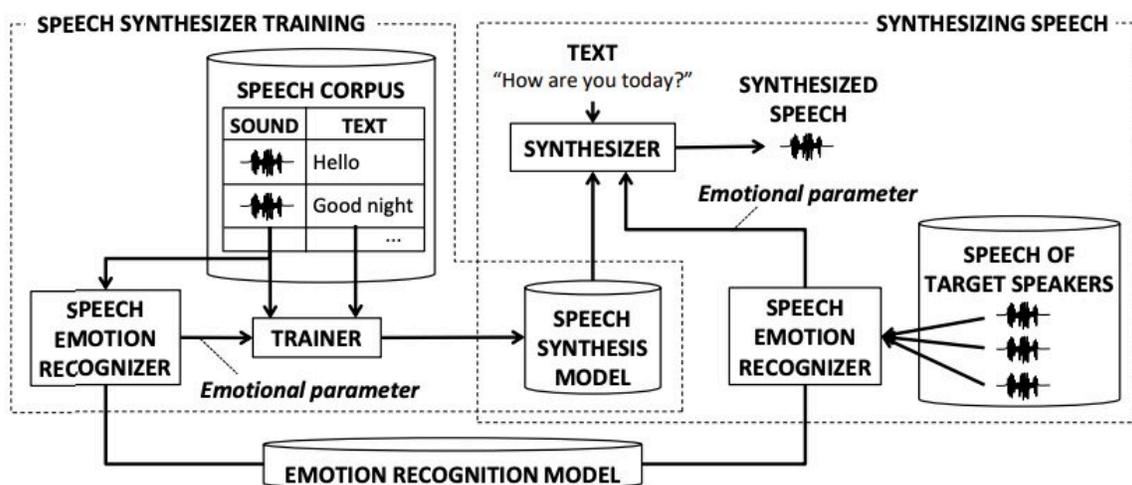


Image summarizing the pipeline of the emotional speech synthesizer. Credit: Homma et al.

Over the past few decades, roboticists have designed a variety of robots to assist humans. These include robots that could assist the elderly and serve as companions to improve their wellbeing and quality of life.

Companion robots and other social robots should ideally have human-like qualities or be perceived as discrete, empathic and supportive by users. In recent years, many computer scientists have thus been trying to

give these robots qualities that are typically observed in human care givers or health professionals.

Researchers at Hitachi R&D Group and University of Tsukuba in Japan have developed a new method to synthesize emotional speech that could allow companion robots to imitate the ways in which caregivers communicate with older adults or vulnerable patients. This method, presented in a paper pre-published on arXiv, can produce emotional speech that is also aligned with a user's circadian rhythm, the internal process that regulates sleeping and waking patterns in humans.

"When people try to influence others to do something, they subconsciously adjust their speech to include appropriate emotional information," Takeshi Homma et al explained in their paper. "For a [robot](#) to influence people in the same way, it should be able to imitate the range of human emotions when speaking. To achieve this, we propose a speech synthesis method for imitating the [emotional states](#) in human speech."

The method combines speech synthesis with emotional speech recognition methods. Initially, the researchers trained a [machine-learning model](#) on a dataset of human voice recordings gathered at different points during the day. During training, the emotion recognition component of the [model](#) learned to recognize emotions in [human speech](#).

Subsequently, the speech synthesis component of the model synthesized speech aligned with a given emotion. In addition, their model can recognize emotions in the speech of human target speakers (i.e., caregivers) and produce speech that is aligned with these emotions. Contrarily to other emotional speech synthesis techniques developed in the past, the team's approach requires less manual work aimed at adjusting emotions expressed in the synthesized speech.

"Our synthesizer receives an emotion vector to characterize the emotion of synthesized speech," the researchers wrote in their paper. "The vector is automatically obtained from human utterances using a speech emotion recognizer."

To evaluate their model's effectiveness in producing appropriate emotional speech, the researchers carried out a series of experiments. In these experiments, a robot communicated with elderly users and tried to influence their mood and arousal levels by adapting the emotion expressed in its speech.

After the participants had listened to samples produced by the model and other emotionally neutral speech samples, they provided their feedback on how they felt. They were also asked whether the synthetic speech had influenced their arousal levels (i.e., whether they felt more awake or sleepy after listening to the recordings).

"We conducted a subjective evaluation where the elderly participants listened to the speech samples generated by our method," he researchers wrote in their paper. "The results showed that listening to the samples made the participants feel more active in the early morning and calmer in the middle of the night."

The results are highly promising, as they suggest that their [emotional](#) speech synthesizer that can effectively produce caregiver-like speech that is aligned with the circadian rhythms of most elderly users. In the future, the new model presented on arXiv could thus allow roboticists to develop more advanced companion robots that can adapt the emotion in their [speech](#) based on the time of the day at which they are interacting with users, to match their levels of wakefulness and arousal.

**More information:** Takeshi Homma et al, Emotional speech synthesis for companion robot to imitate professional caregiver speech.

arXiv:2109.12787v1 [cs.RO], [arxiv.org/abs/2109.12787](https://arxiv.org/abs/2109.12787)

© 2021 Science X Network

Citation: A new model to synthesize emotional speech for companion robots (2021, October 6)  
retrieved 25 April 2024 from

<https://techxplore.com/news/2021-10-emotional-speech-companion-robots.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.