

What can we do to properly convey human expressiveness with synthetic speech?

Expressive speech synthesis: From emotions to speaking styles

どんな研究？

- Computers can generate high quality synthetic speech.
- But they sound very boring!
- Human voices are much more expressive (happy, sad, excited...)
- We want to teach computers to talk like us!

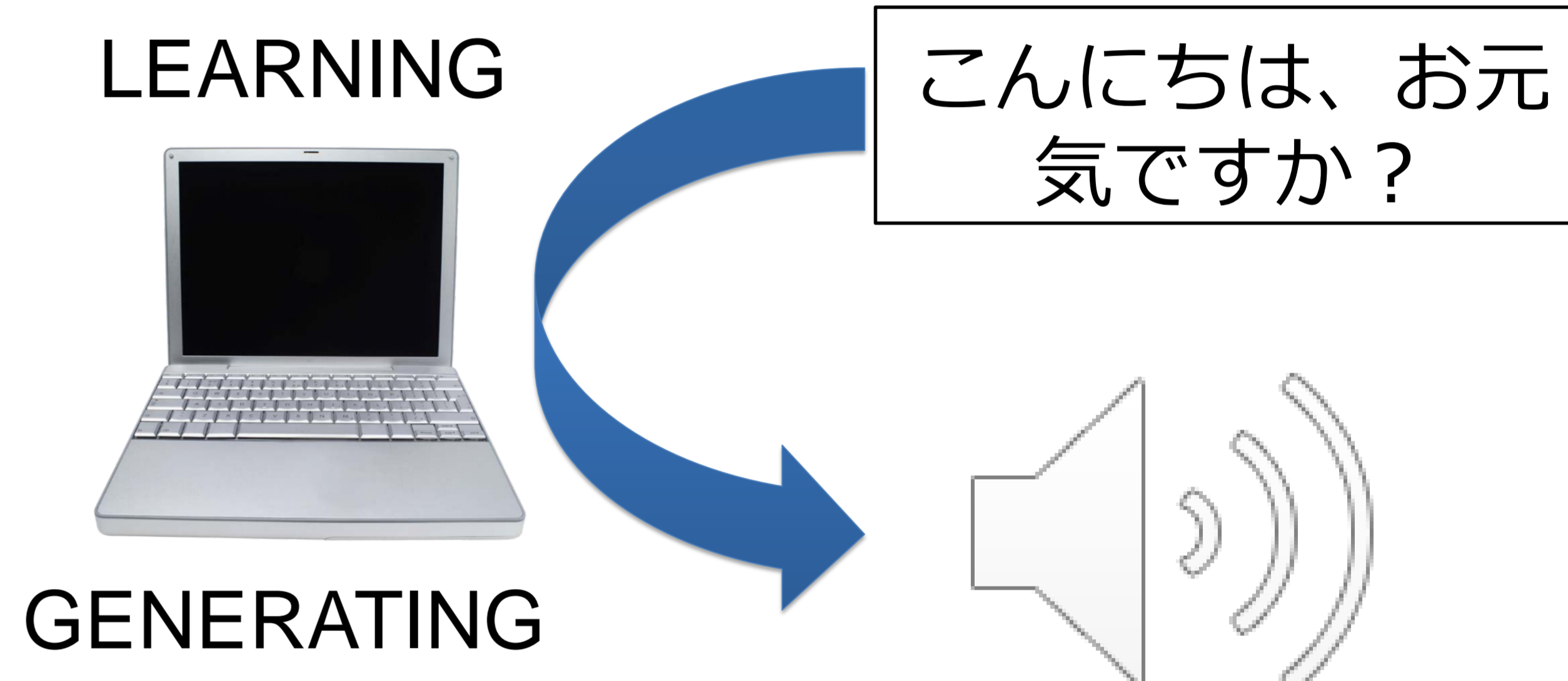
何ができる？

- Expressive text-to-speech
- More human-like robots
- Automatic audiobook narrations, sports commentaries...

状況設定

- In **expressive text-to-speech system** we want to:

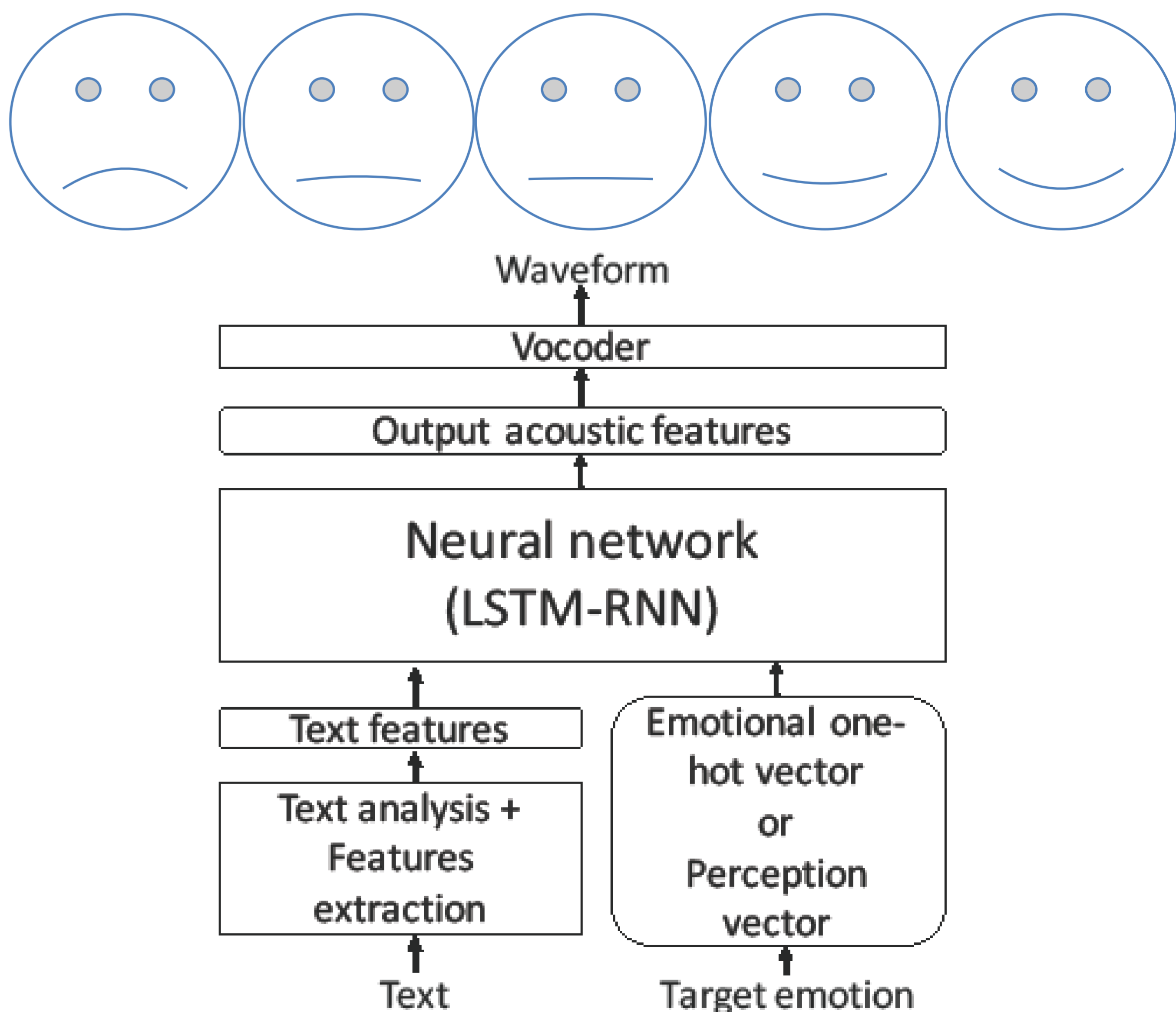
- Make our computer say some text **using the way of speaking we want!**
- Not only computers, but robots, car navigation systems, audiobooks, videogames...



研究内容

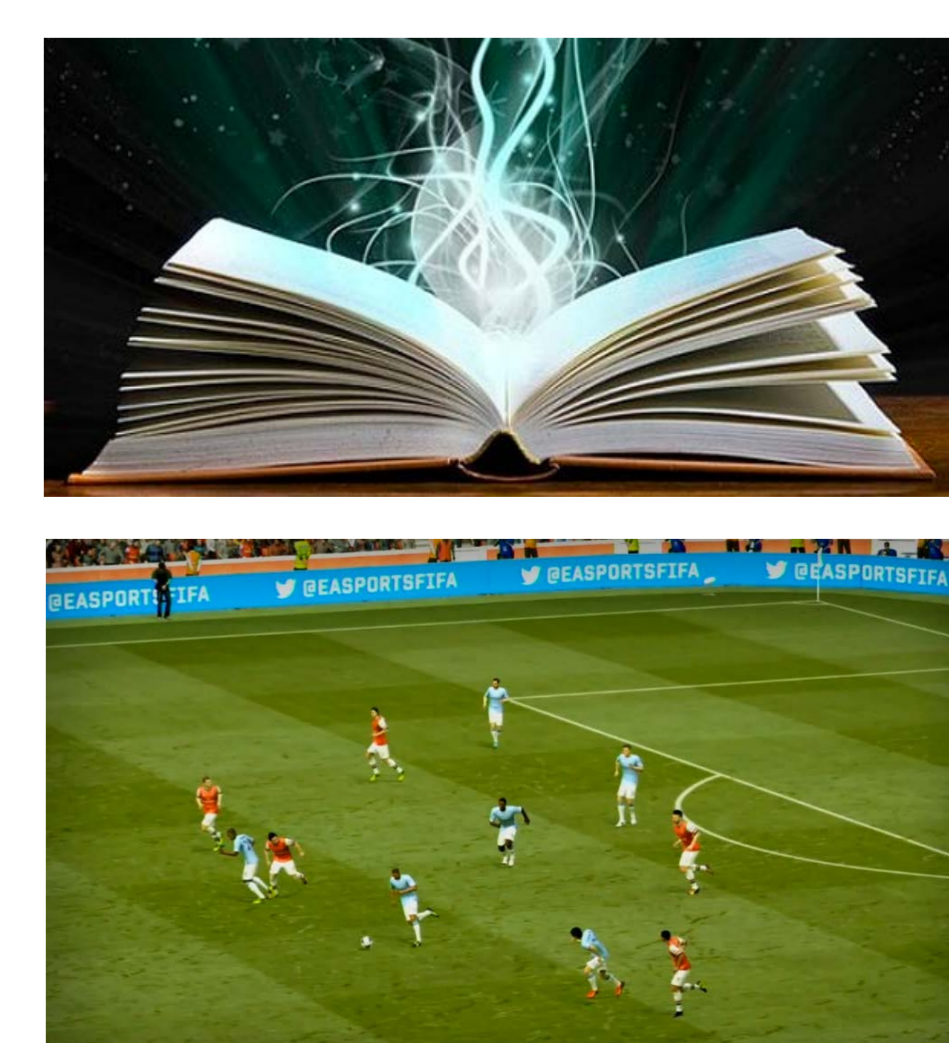
Our emotional text-to-speech system:

1. We convert the text and the desired emotion to something that the computer can understand.
2. By using neural networks we teach the computer how the different emotions and letters of the language are spoken.
3. Then, we transform the output of the neural network back into speech!
4. If we have a robot or an avatar, we can make it look as if it was also emotional.



Other uses we have experimented with:

- Audiobooks (for example: Harry Potter!)
- Car navigation systems
- Dynamic videogame commentaries...



We have a lot of samples to show, just ask and listen!