

ВИКОРИСТАННЯ GOOGLE API ДЛЯ РЕАЛІЗАЦІЇ ТЕХТ-ТО-SPEECH ТЕЛЕГРАМ БОТА

Панченко В.І., Перець К.Г.

*Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків*

Однією з проблем сучасного суспільства є мовний бар'єр між людьми різних національностей або людьми з обмеженими можливостями, яким необхідно донести свою думку до співрозмовника. Далеко не кожна людина знає мову країни, в якій він перебуває з туристичними або іншими намірами, або мову жестів, яку не розуміють більшість людей без обмежених можливостей. Для вирішення цієї проблеми пропонується розробка телеграм боту, який дозволить переводити мову користувача на мову співрозмовника, а також озвучувати її. Використання цієї програми допоможе розширити межі спілкування людей і надасть можливість поліпшити взаємодію між громадянами різних країн з відмінними мовами спілкування.

Для розробки даного бота використовувалася хмарна технологія Google Cloud Text-to-Speech. Ця хмара перетворює текст в мову, схожу на людську, в більш ніж 180 голосах на більш ніж 30 мовах і в різних варіантах. Вона застосовує новаторські дослідження в області синтезу мови (WaveNet) і потужних нейронних мереж Google для надання високоякісного звуку [1].

Технологія заснована на оновленій версії WaveNet, тому можна мати певну впевненість в правильному звучанні навіть складного тексту. Завдяки хмарному процесору Google TPU [2], штучна мова генерується в 1000 разів швидше: одна секунда відтвореного тексту створюється за 50 мілісекунд. Для більш природного звучання якість звукових фрагментів підвищено з 8 до 16 біт.

WaveNet навчається відповідності тексту певним формам коливань хвилі і потім на підставі цієї бази знань формує з текстових уривків окремі звукові хвилі.

Для оцінки якості мови були залучені добровольці. Створені системою WaveNet аудіозаписи отримали в середньому 4,1 бала. Для порівняння, голос реальної людини був оцінений максимум на 4,59 бала з 5.

Література:

1. *Николенко С., Кадурич А., Архангельская Е.* Глубокое обучение / *С. Николенко, А. Кадурич, Е. Архангельская.* – Издательский дом Питер, 2017. – 480 с.
2. *Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu.* WaveNet: A Generative Model for Raw Audio– Режим доступу: <https://arxiv.org/pdf/1609.03499.pdf> – Дата доступу: 03.03.2020.