

Szymon Dyszewski

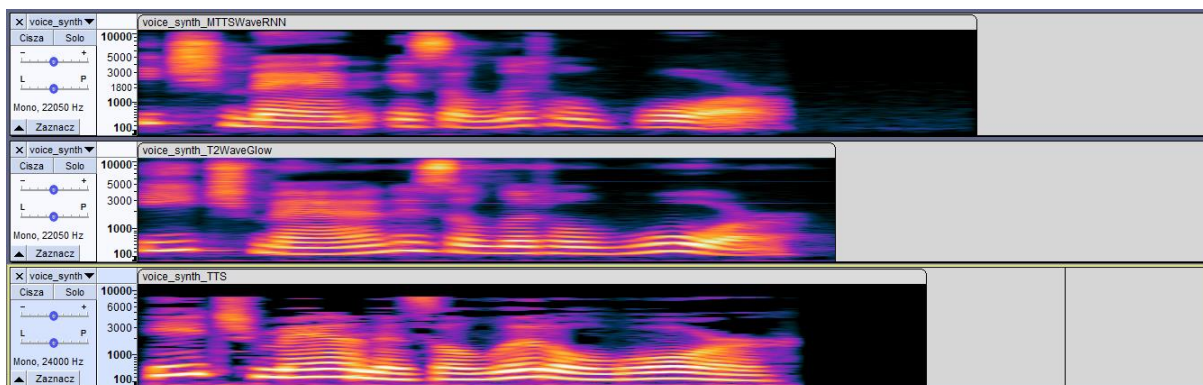
Numer albumu: 310625

## Sprawozdanie z laboratorium nr 3 z przedmiotu WMM

- *na podstawie własnych wrażeń słuchowych opisać różnice między algorytmami (Real-Time Voice Cloning, Google TTS (może działać po polsku), Tacotron2 + Waveglow (angielski) i Mozilla TTS (angielski)), który z nich jest w stanie bardziej naturalnie syntetyzować ludzką mowę?*

Najgorzej wypadł algorytm RTVC, nie dość, że brzmi, jak bot to dźwięk jest niewyraźny. Google TTS brzmi jak bot, ale przynajmniej nie słychać, żadnych zakłóceń. T2WG mówi jednostajnie (słabe akcentowanie), w tle nagrania słyszalny jest delikatny pisk, ale dźwięk można nazwać ludzkim głosem. MTTS sprawia wrażenie akcentującego odpowiednie głoski i pozbawiony jest jakichkolwiek zakłóceń, zdecydowanie najlepszy z dostępnych algorytmów.

- *dokonać syntezy tego samego zdania wpisując je do algorytmu Google TTS (potrzebna będzie zmiana w kodzie z języka "pl" na "en"), Tacotron2 + Waveglow oraz Mozilla TTS i zapisać wynikowe pliki dźwiękowe na dysk*
- *dokonać analizy porównawczej spektrogramów (w oprogramowaniu Audacity) i zapisać wniosek*



Ścieżki dźwiękowe są różnej długości, przez różne tempa wymowy algorytmów. Spektrogram T2WG posiada dwa cienkie pasma odpowiadające częstotliwościom 3 i 6k HZ, najprawdopodobniej to one są odpowiedzialne za słyszalny pisk. Spektrogram MTTS jest najbardziej „spójny” tj. wszystkie dźwięki się ze sobą stykają i nie posiadają czarnych szpar dla określonych częstotliwości między sobą, co więcej ścieżka nie kończy się po zakończeniu wymowy, słychać tam szmery i szumy. Spektrogram TTS ucina się powyżej 9k HZ. Na spektrogramach widać wydajność akcentowania algorytmów, intensywność barwy przemieszcza się po spektrogramie do góry i na dół w zależności od akcentów. Najlepiej wypada MTTS, TTS natomiast praktycznie wcale nie prezentuje akcentów.

- *zapisać nagrany tekst w postaci pliku dźwiękowego, a w sprawozdaniu zapisać teksty, które udało się rozpoznać każdemu z algorytmów (CMUSphinx, Google, Mozilla)*

sphinx: "each man de soleil war"

google: "each man is a laser war"

Mozilla DeepSpeech: "each man is a little war"