WMM — Dźwięk — Laboratorium 3b:

Marfenko Mykhailo 323558

1 TTS

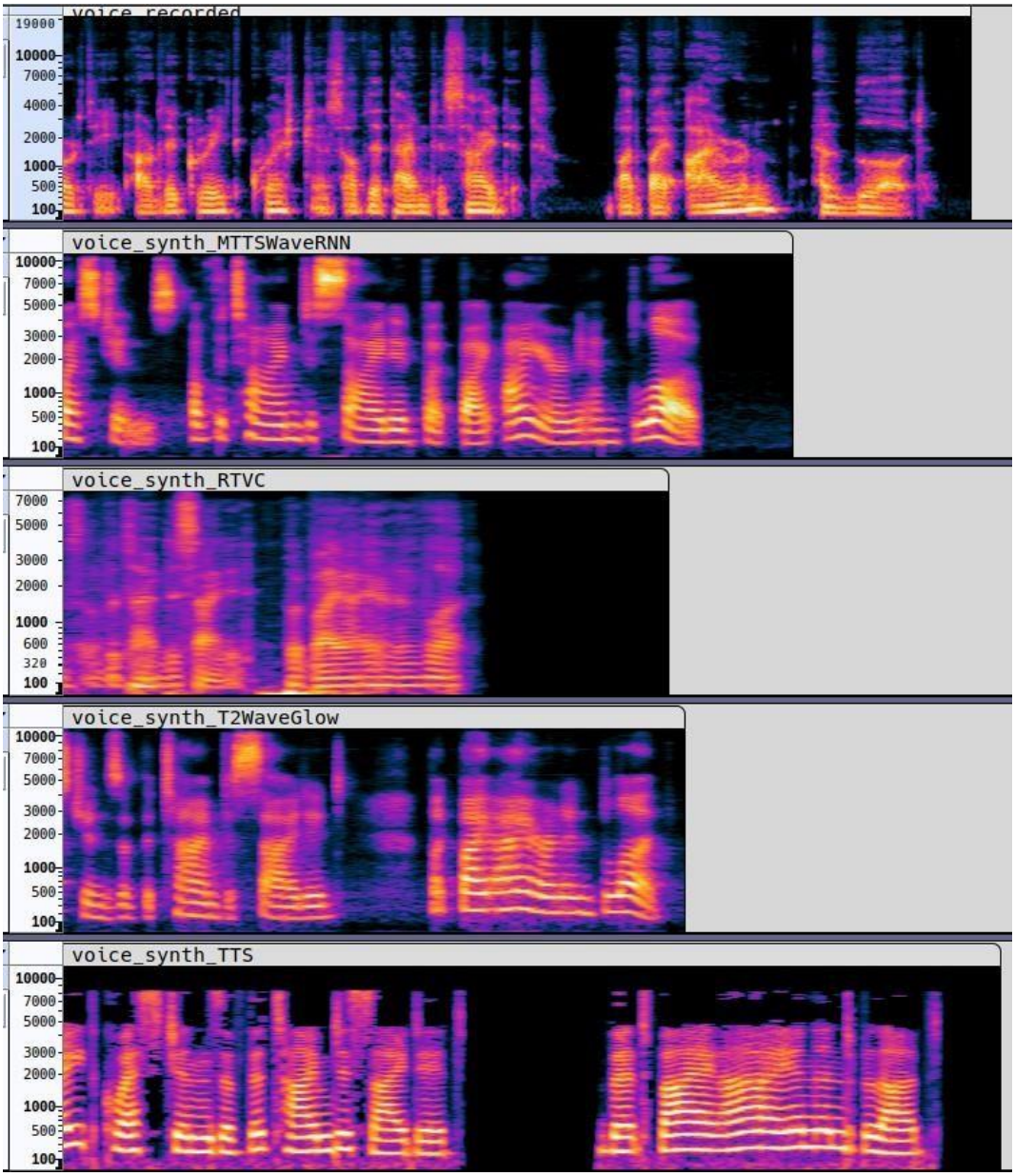
1.1 Porównanie brzmienia

Jako frazę do wygenerowania wybrałem cytat Bismarcka: “Not by speeches and votes of the majority are the great questions of the time decided, but by iron and blood.”

Zdecydowanie najlepiej poradził sobie algorytm od Mozilli. Na drugim miejscu jest algorytm Tacotron2 + Waveglow, chociaż brzmi on trochę sztucznie. Pozostałe dwa chyba nawet nie próbują brzmieć ludzko. Da się usłyszeć, co mówią, ale ciężko jest to klasyfikować jako mowę będącą choćby zbliżoną do ludzkiej.

Oczywiście, wszędzie intonacja jest do niczego.

1.2 Porównanie spektrogramów



(a) Spektrogramy nagrań w Audacity

Widzimy, że ludzka mowa ma dużo szerszy zakres częstotliwości, niż wygenerowane próbki (górna granica 19 kHz vs 9–10 kHz). Ponadto, sztucznie generowany tekst ma lepiej widzoczne pauzy w wymowie, choć może to być

kwestia mojego stylu mówienia. Widać też, że zebrane próbki mają więcej energii w wyższych pasmach od mojego nagrania, ale jest to raczej spowodowane tym, że naturalnie mój głos (męski) ma mniej energii w tamtych pasmach od głosu kobiecego, który był w większości przypadków generowany.

2 STT

2.1 Otrzymane zdania

- Sphinx: “but notably speeches can fault of the majority are degreequestions all the time decided by the iṙand the blood”
- Google: “not by speeches and votes of the majority are the great ques-tions of the time to side it but by iron and blood”
- Mozilla: “another by speeches involved of the majority are degreedquestions of the time decided but by iron and blood”

2.2 WER

silnik	WER
Sphinx	57.1 %
Google	14.3 %
Mozilla	23.8 %