

УДК 004.048

Андрій Луцків канд. техн. наук, доцент, Сергій Макогон  
Тернопільський національний технічний університет імені Івана Пулюя

## ТИПИ АРХІТЕКТУР НЕЙРОННИХ МЕРЕЖ ДЛЯ ПЕРЕТВОРЕННЯ ТЕКСТОВИХ ПОВІДОМЛЕНЬ У ЗВУКОВИЙ ПОТІК

Andriy Lutskiv PhD., Assoc. Prof., Serhii Makohon  
TYPES OF NEURAL NETWORK ARCHITECTURES FOR TEXT TO SPEECH

Сьогодні існують три основні типи архітектури, які використовують акустичні алгоритми. До них належать:

- алгоритми на основі рекурентних нейронних мереж;
- алгоритми на основі згорткових нейронних мереж;
- алгоритми на основі трансформерів.

Рекурентна нейронна мережа може використовуватися для представлення такої структури акустичної моделі та таких алгоритмів, як неавто регресійний Tacotron 2, структуру якого показано на рис. 1.

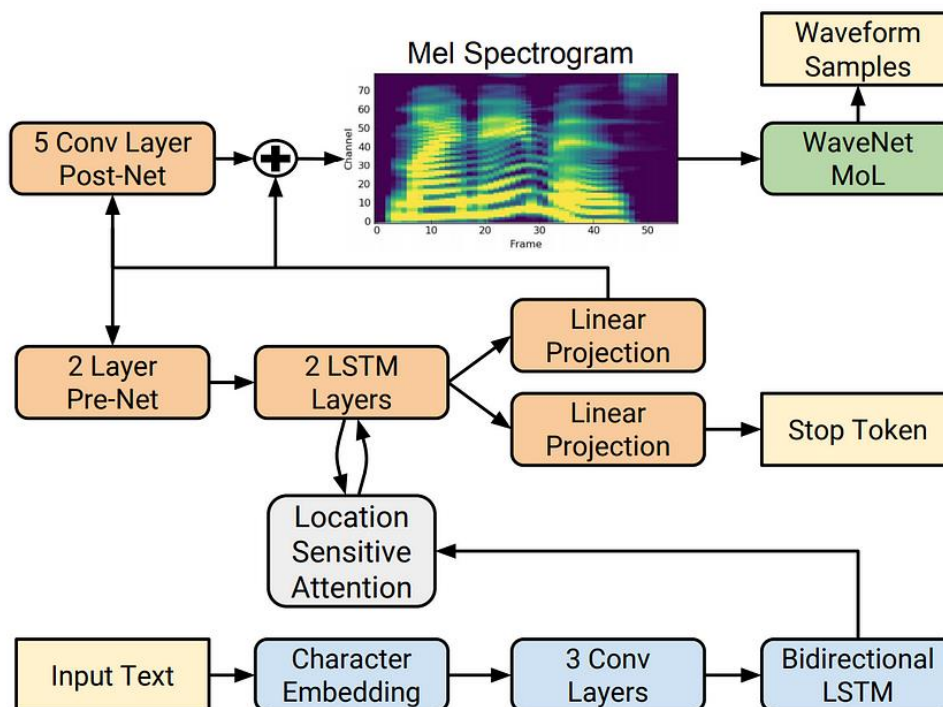


Рис. 1. Архітектура Tacotron 2

По факту, RNN мережа виконує прогнозування ознак «sequence-to-sequence», яка в свою чергу забезпечує прогнозування послідовності кадрів Mel-спектрограми з послідовності вхідних символів.

Модифікована версія WaveNet генерує зразки хвилі у часовій області, обумовлені прогнозованими фреймами Mel-спектрограми. Архітектура Tacotron 2 стала величезним кроком вперед у покращенні якості голосу порівняно з іншими методами, такими як конкатенативний, параметричний і авторегресійний Tacotron 1.