

Національний Університет
“Києво-Могилянська Академія”

Генерація мови з використанням нейронних мереж

Підготувала:

Пархоменко Анастасія Олександрівна, студентка БП КН

Наукова керівниця: Бучко Олена Андріївна, доцент, к.н.

Постановка задачі

- Дослідити наявні можливості нейронних мереж до генерації мовлення.
- На основі зібраних та проаналізованих знань обрати найоптимальнішу модель та досягти якісного результату в генерації української мови.
- Залучити, знайти, дописати інструменти для досягнення кращої вимови.

Хід роботи

Аналіз тексту

1. Токенізація
2. Синтаксичний і морфологічний аналіз
3. Фонеми і просодії

Моделі для генерації

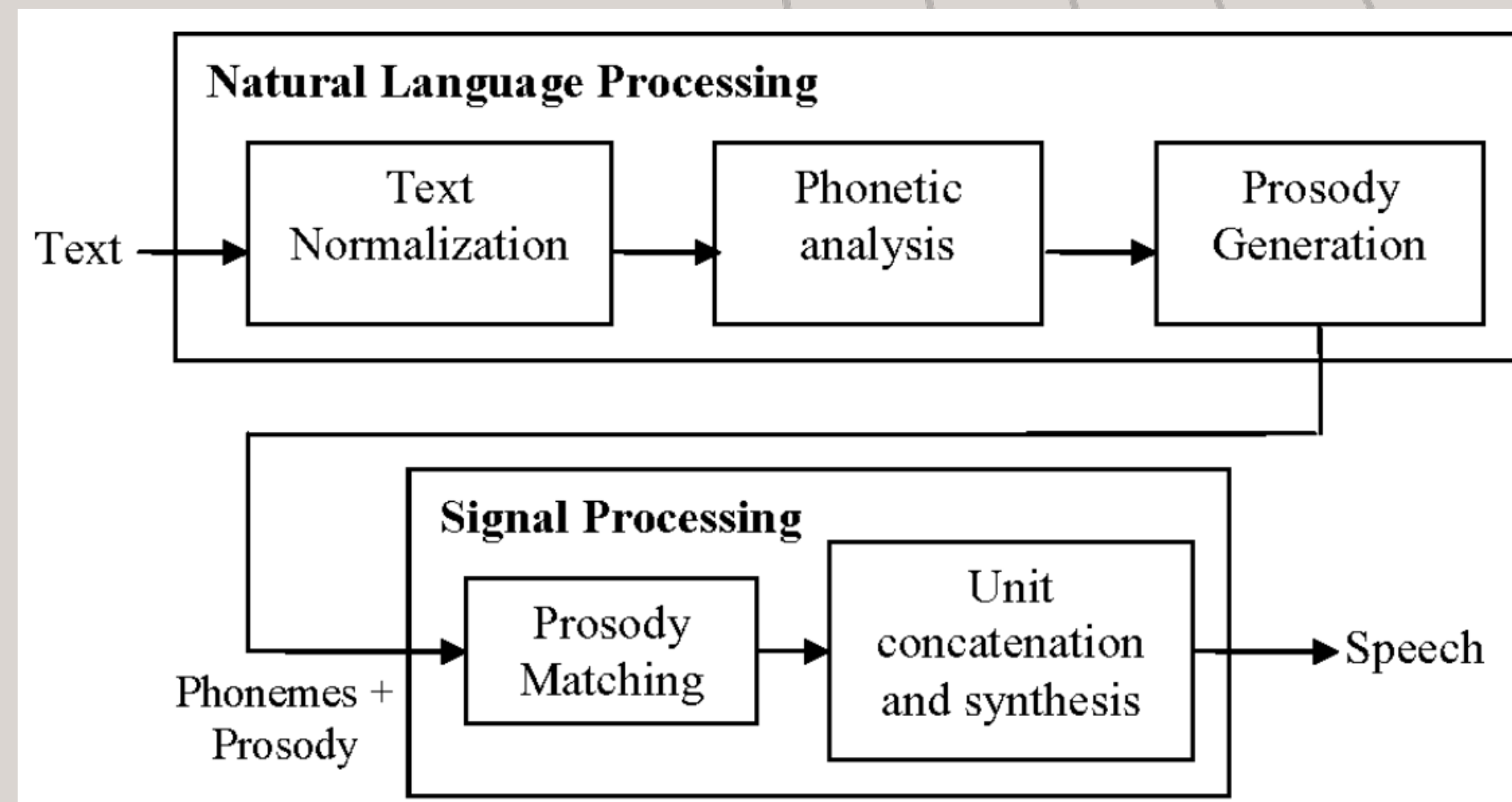
1. Традиційні методи:
 - Формантний синтез
 - Конкатенативний синтез
2. Нейронні мережі:
 - Tacotron 2
 - FastSpeech
 - WaveNet
 - Glow-TTS

Власне навчання

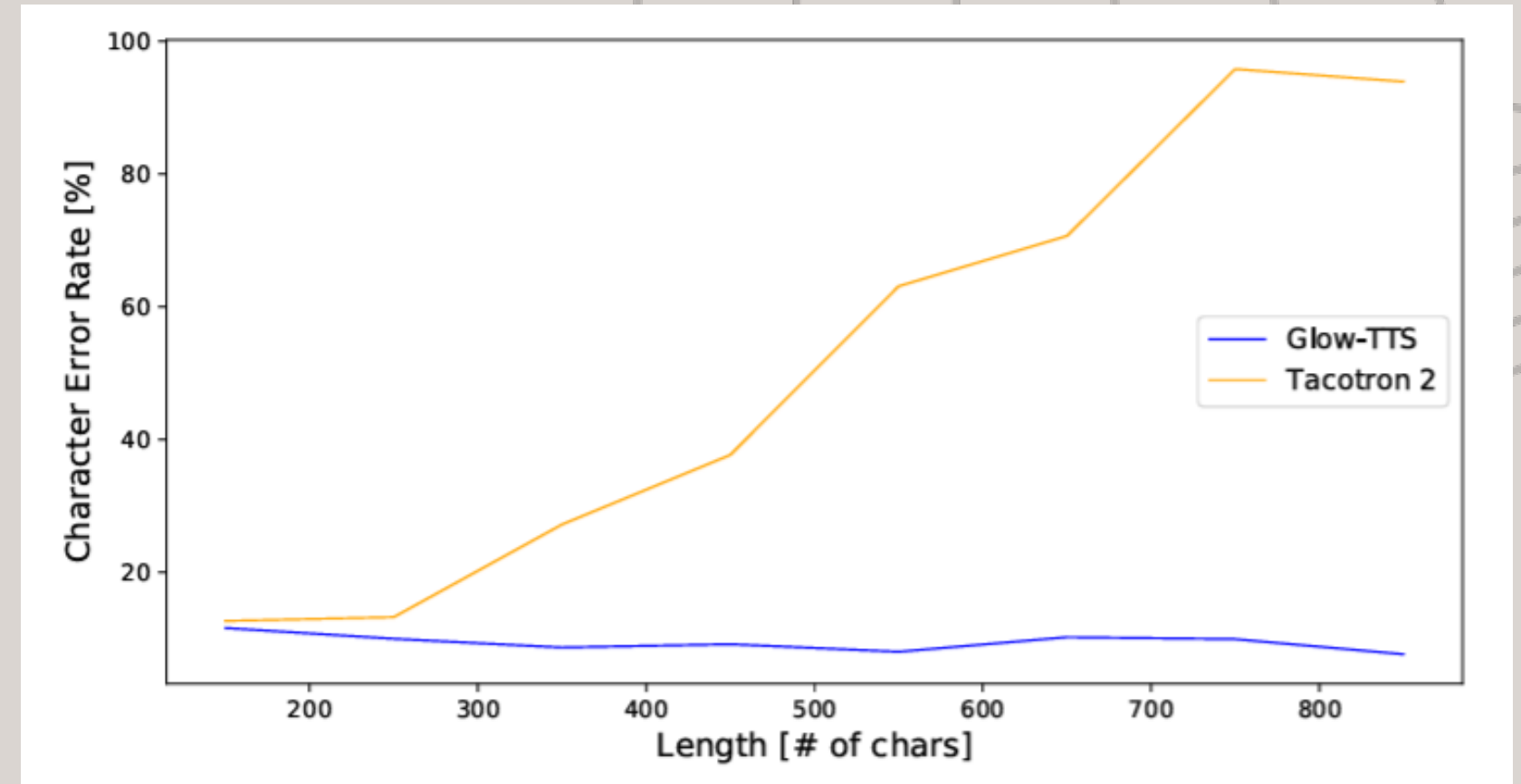
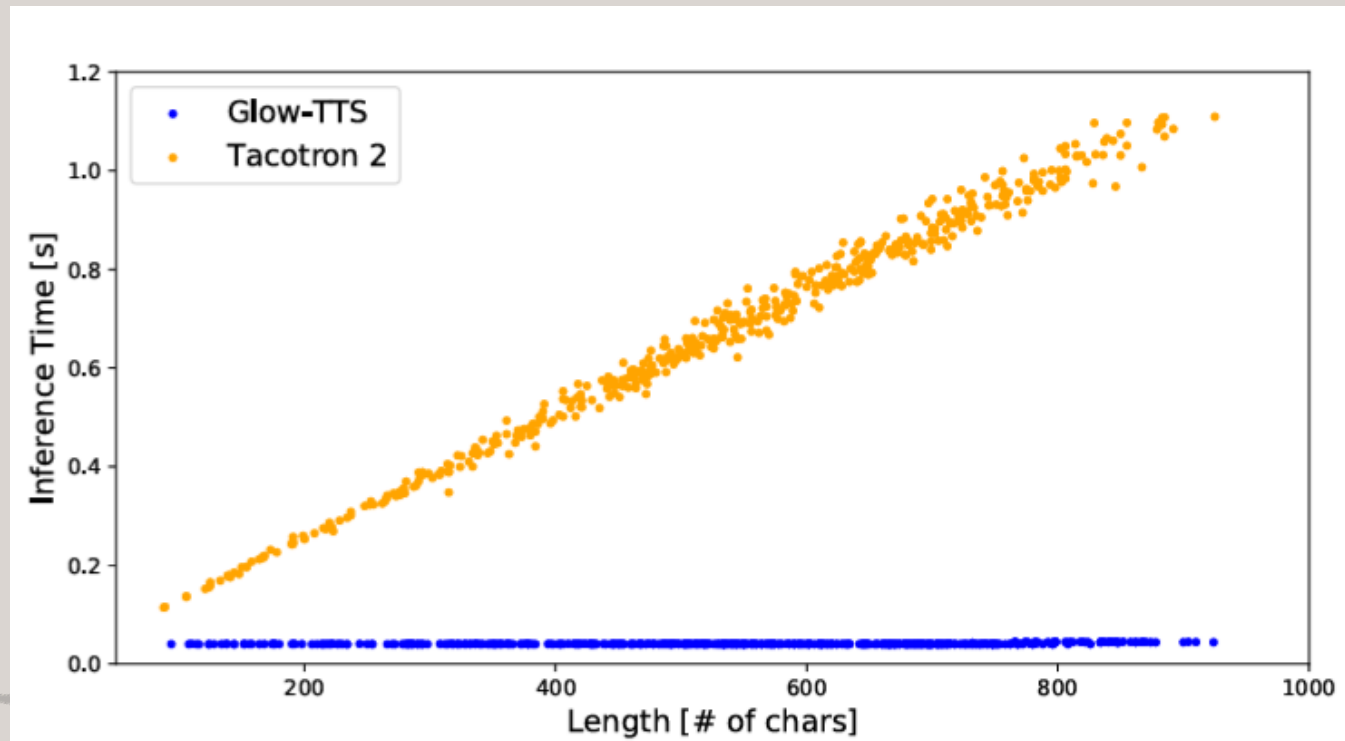
1. Порівняльний аналіз
2. Тестування базової моделі на базовому наборі даних
3. Навчання на українському датасеті
4. Файн-тюнінг + залучення додаткових інструментів

СИНТЕЗ МОВЛЕННЯ

Синтез мовлення – це процес автоматичного перетворення тексту в мовлення за допомогою комп'ютерних систем. Мета синтезу мовлення полягає в генерації зрозумілого та природного звучання аудіосигналу, який відповідає введеному тексту.



Порівняння моделей



METHOD	9-SCALE MOS
GT	4.54 ± 0.06
GT (MEL + WAVEGLOW)	4.19 ± 0.07
TACOTRON2 (MEL + WAVEGLOW)	3.88 ± 0.08
GLOW-TTS ($T = 0.333$, MEL + WAVEGLOW)	4.01 ± 0.08
GLOW-TTS ($T = 0.500$, MEL + WAVEGLOW)	3.96 ± 0.08
GLOW-TTS ($T = 0.667$, MEL + WAVEGLOW)	3.97 ± 0.08

Glow TTS

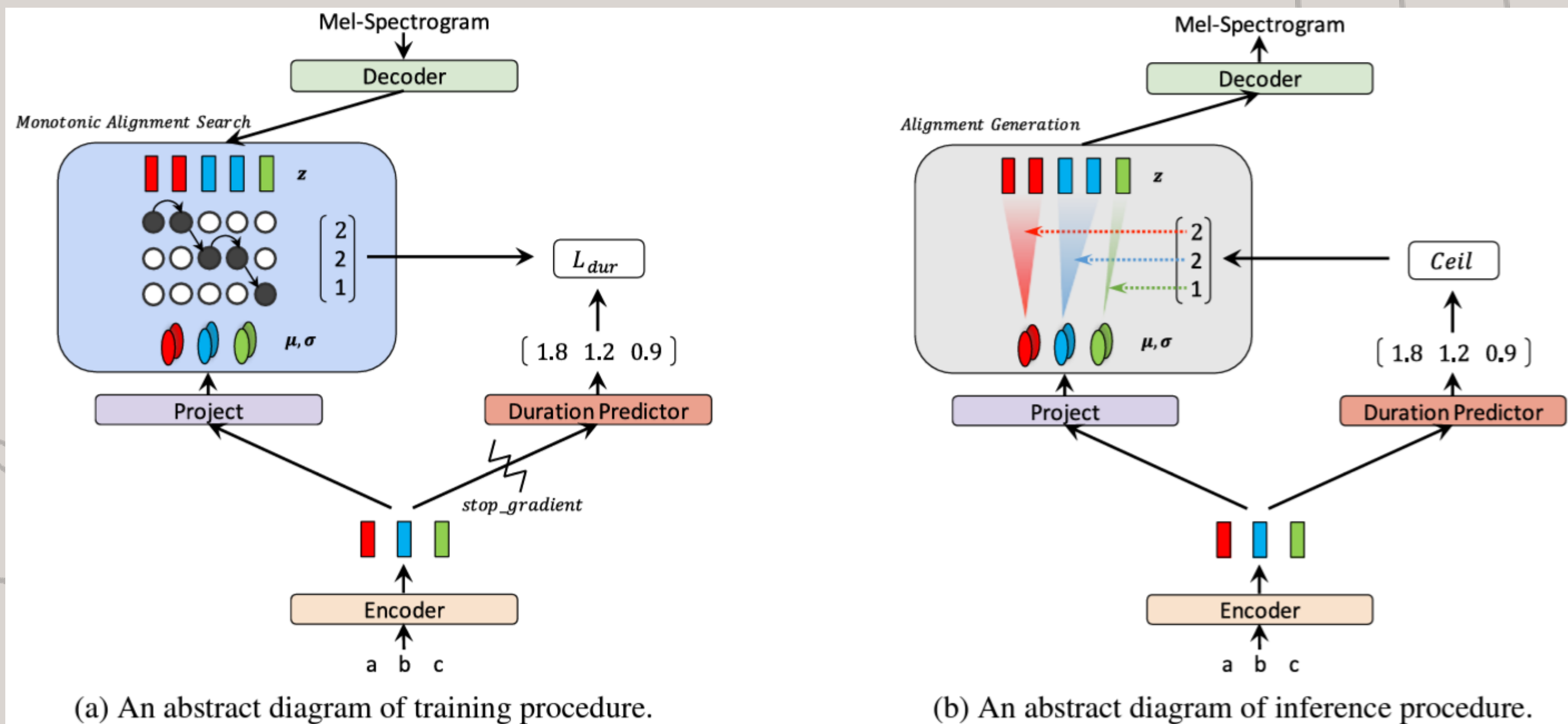


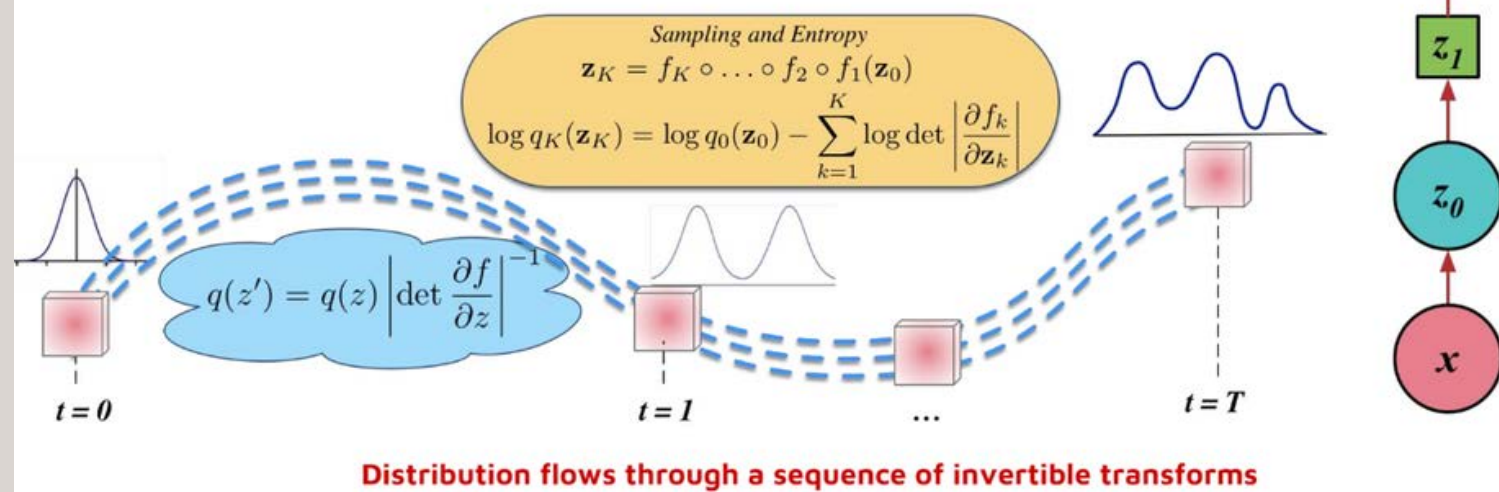
Figure 1. Training and inference procedures of Glow-TTS.

Суть методу

Normalising Flows

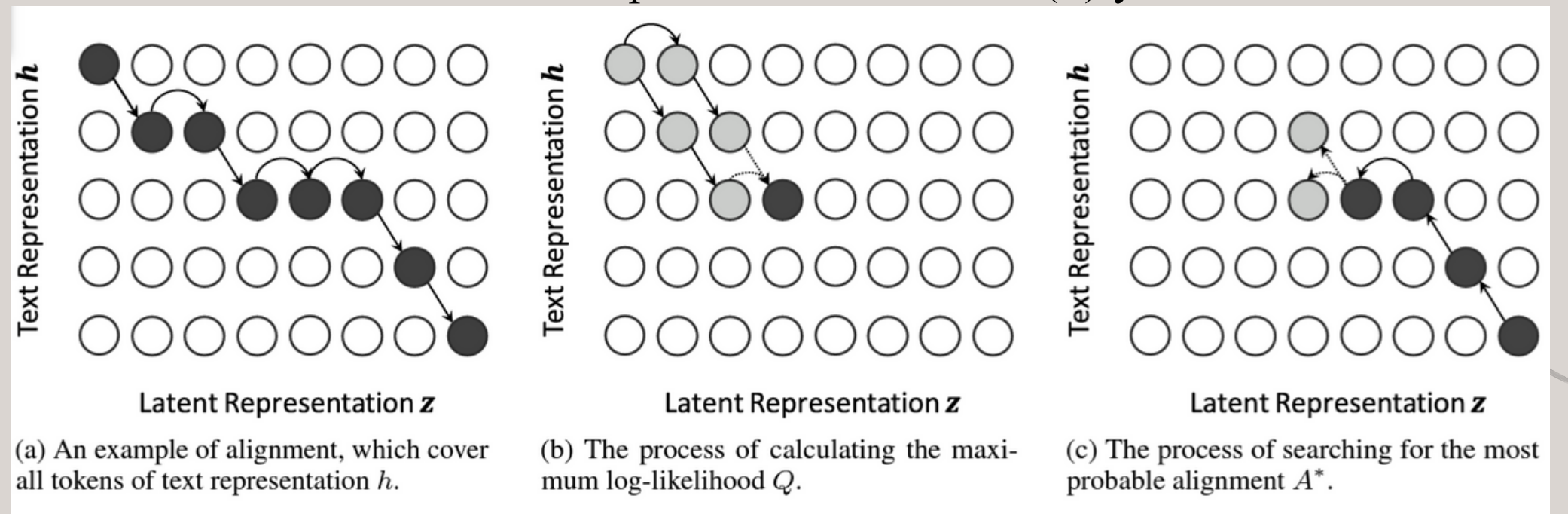
Exploit the rule for change of variables:

- Begin with an initial distribution
- Apply a sequence of K invertible transforms



концепція нормалізаційних потоків

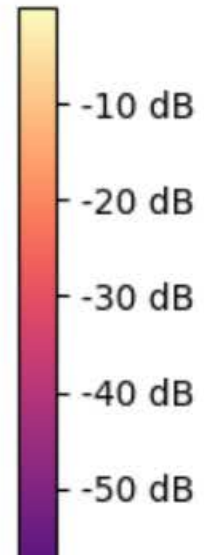
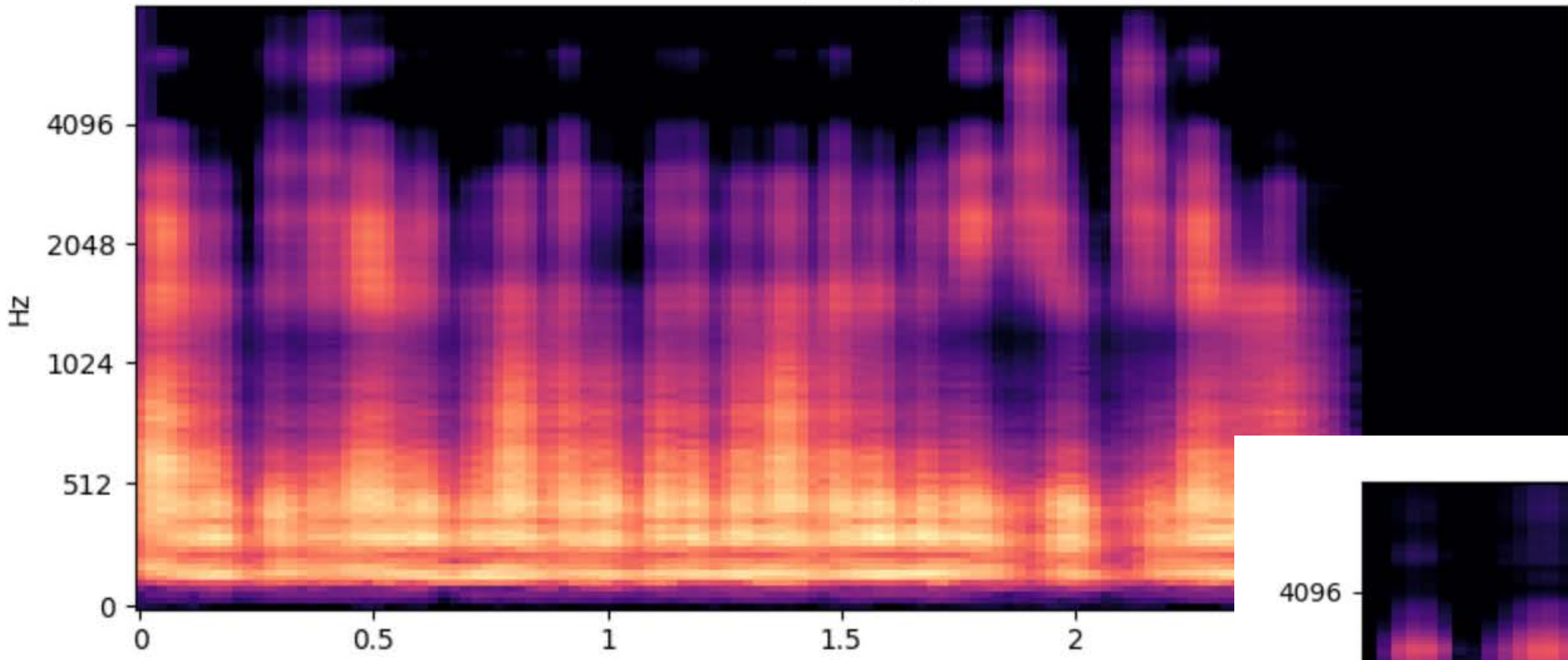
процес вирівнювання між текстовими представленнями (h) та латентними представленнями (z) у моделі Glow-TTS



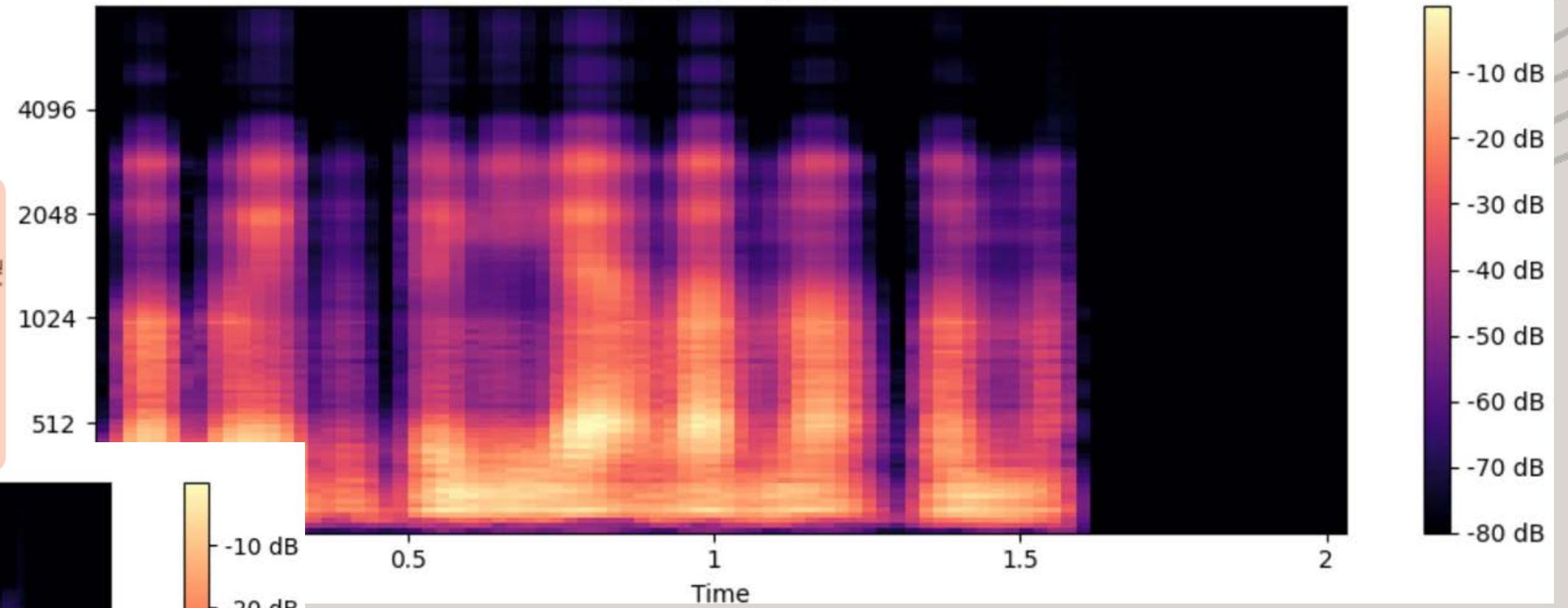
Експериментальна частина

1. стандартне на ljspeech
2. на україномовному датасеті з графемами
3. на україномовному датасеті з фонемами

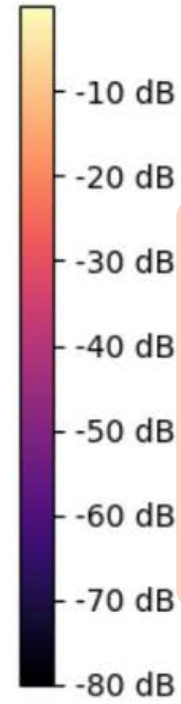
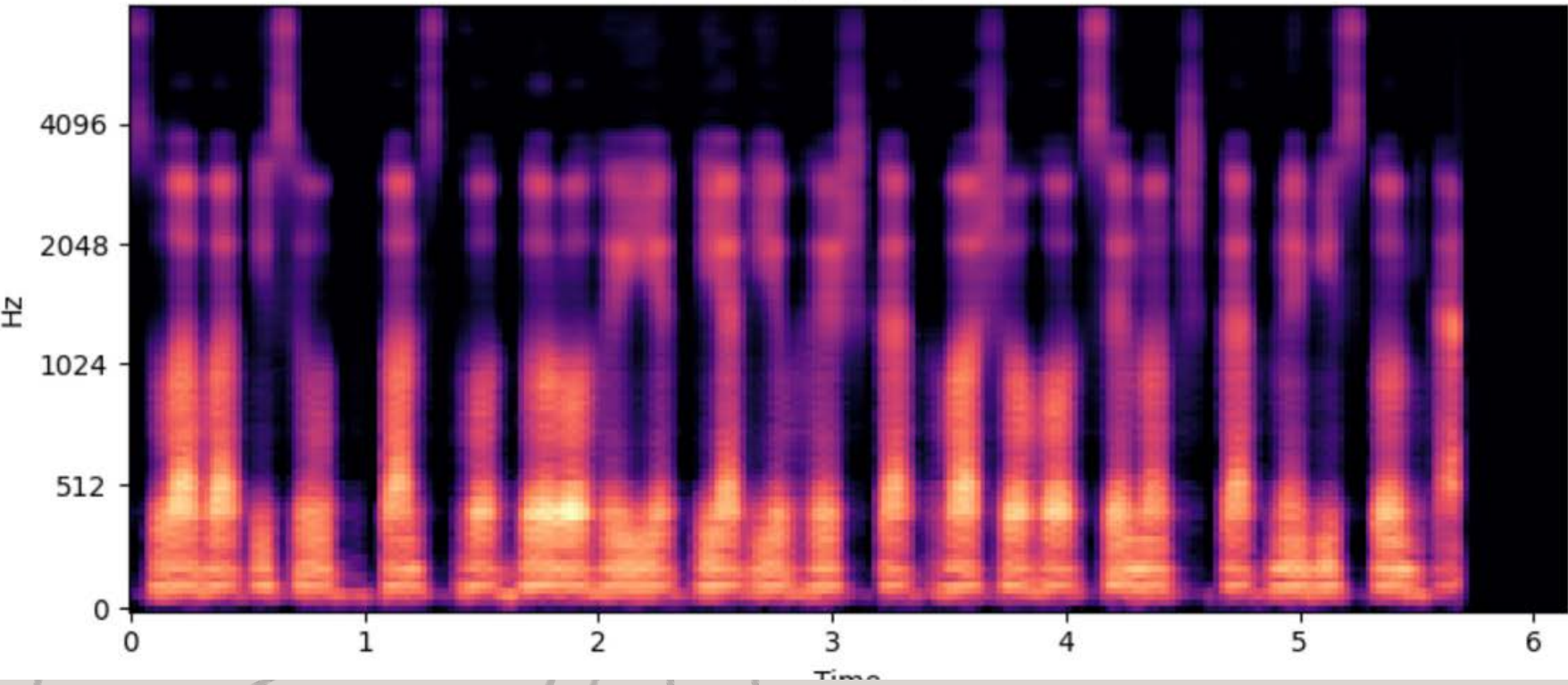
Mel spectrogram



Mel spectrogram



Mel spectrogram





**Дякую за
увагу**