

ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

Харківський національний університет радіоелектроніки
Кафедра ЕОМ

Аналіз впливу SSI-підходу на точність розпізнавання голосових команд

Кваліфікаційна робота
Другий рівень (магістр)

Автор:

Терещенко О.В.,
студ. гр. СПм-21-1

Керівник:

Барковська О.Ю.,
доц. каф. ЕОМ

Актуальність



Розпізнавання у шумних місцях



Пілоти



Водії



Глухі та слабочуючі

Мета і задачі роботи

Мета: дослідження впливу використання інтерфейсу безмовного доступу (SSI), що забезпечує визначення початкової фази звукового ряду, асоційованого із початком мовлення, на точність розпізнавання голосових команд у різних звукових оточеннях.

Задачі:

- аналіз методів попередньої обробки та розпізнавання мови;
- аналіз методів розпізнавання візем людини, що розмовляє;
- створення моделі розпізнавання голосових команд, вдосконаленої завдяки аналізу віземи людини, що розмовляє;
- реалізація моделі розпізнавання голосових команд на основі комбінації аналізу звукового ряду та зображення губ диктора;
- виконання експериментів із зашумленим мовним корпусом;
- аналіз отриманих результатів.

3

Системи розпізнавання мови

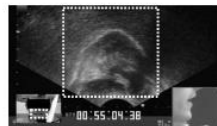


Microsoft Cortana



LipNet

Система	SSI	Технології	Dataset	Тип команди	Точність
Kaldi	-	GMM/HMM	Switchboard	Речення	80%
Microsoft	-	RNNLM	Switchboard	Речення	93,8%
Oisper	+	УЗД	Власний	Слово	70%
SpeakUP	+	ЕМГ	Власний	Словосполучення	80,1%
Alter Ego	+	ЕМГ	Власний	Словосполучення	91,2%
Gergen et al (2016)	+	ALR	GRID	Слово	86,4%
LipNet	+	ALR	GRID	Речення	95,2%



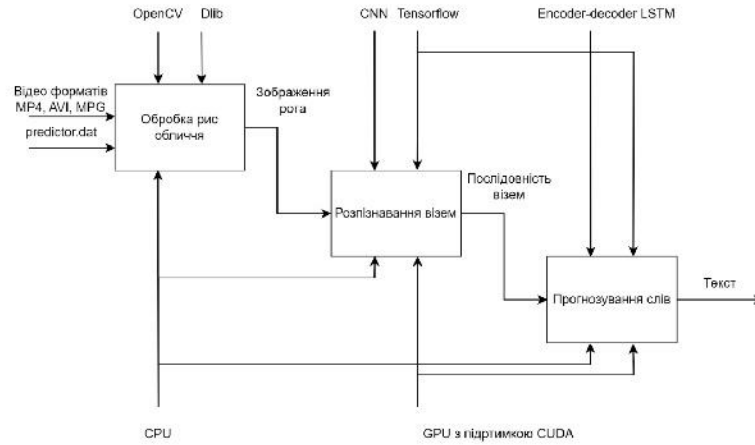
Oisper



Alter Ego

4

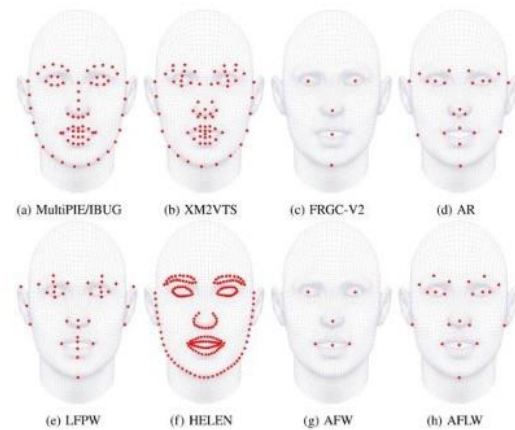
Запропонована модель з SSI-підходом



5

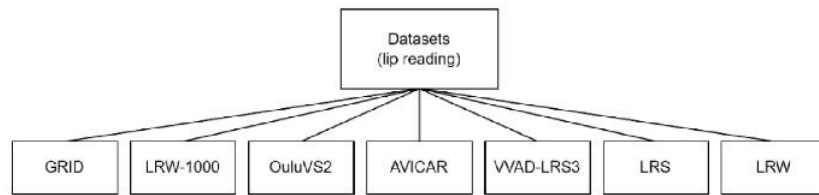
Розпізнавання візем

Приголосці			Голосці		
Візема	Фоонема	Приклад	Візема	Фоонема	Приклад
$V_{f,cl}$	/dʒ/ /tʃ/ /ʒ/ /ʒʲ/		V_a	/a:/ /ao/ /au/ /a/	
$V_{r,bl}$	/p/ /b/ /m/		V_e	/e/ /eu/ /ee/	
$V_{r,v}$	/f/ /v/		V_i	/i:/ /i/	
$V_{d,t,s}$	/d/ /t/ /z/ /ʒ/ /b/		V_o	/o:/ /ou/ /oo/	
$V_{r,w}$	/r/ /w/		V_u	/u/ /u:/	
$V_{b,kl}$	/g/ /k/ /a/ /y/ /l/ /y/ /b/		Silent		



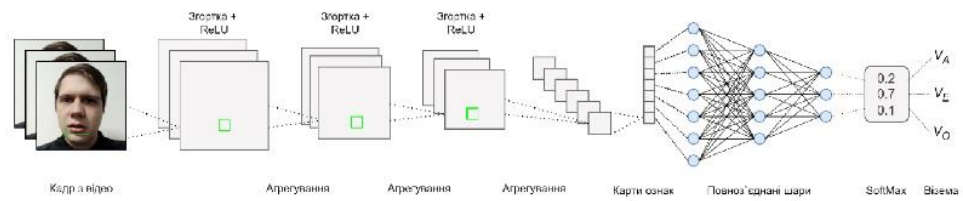
6

Набір даних для ALR



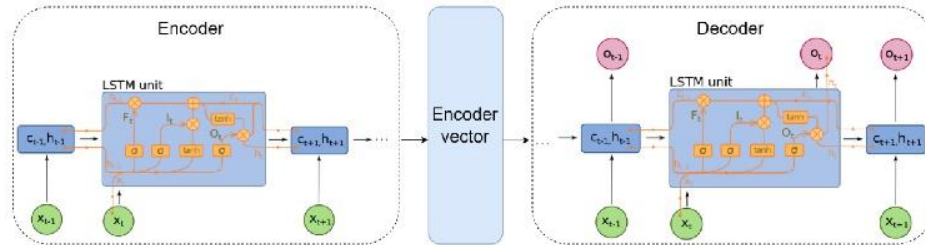
7

CNN



8

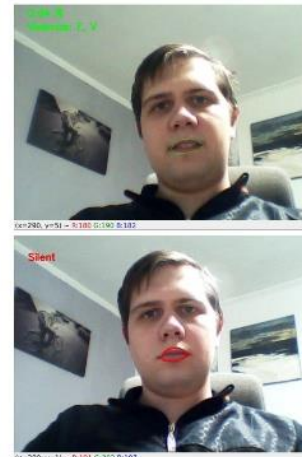
Encoder-decoder LSTM



9

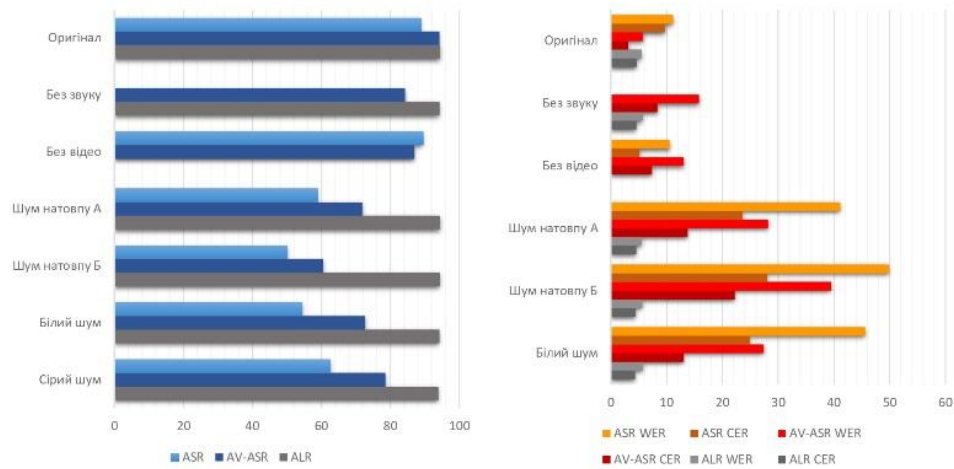
Результати

Умови	AIR			AV-ASR			ASR		
	Власна система			LipNet			CMU Sphinx		
	CER	WER	WRR	CER	WER	WRR	CER	WER	WRR
Білий шум	4,34	5,79	94,21	13,04	27,4	72,6	24,9	45,59	54,41
Сірий шум	5,12	6,06	93,94	11,38	21,46	78,54	22,17	37,42	62,58
Шум наговну А	4,52	5,52	94,48	13,74	28,18	71,82	23,6	41,11	58,89
Шум наговну Б	4,43	5,65	94,35	22,2	39,52	60,48	28,04	49,86	50,14
Без звуку	4,58	5,73	94,27	8,34	15,8	84,2	-	-	-
Без відео	-	-	-	7,35	13,1	86,9	5,1	10,51	89,49
Оригінал	4,6	5,54	94,46	3,1	5,8	94,2	9,67	11,14	88,86



10

Результати



11

Висновки

- розглянуто проблему розпізнавання мови у шумному середовищі
- запропоновано модель розпізнавання з SSI-підходом на основі автоматизованого читання по губах з використанням encoder-decoder LSTM
- створено програмне забезпечення з SSI-підходом та проаналізовано розпізнавання команд у різних звукових оточеннях
- зроблено порівняння з іншими системами розпізнавання мови

12