

# **Синхронізація анімації обличчя персонажа у віртуальній реальності із голосом в режимі реального часу**

Андрій Малишев, ДА-52м

Науковий керівник: к. т. н., доц. Харченко К. В.

## **Мета**

Дослідити принцип роботи систем розпізнавання мовлення, проаналізувати та обрати ту, яка найкраще підходить для розпізнавання в режимі реального часу із можливістю виділення фонем та їх тривалостей. Проаналізувати принцип роботи сучасних додатків віртуальної реальності та обрати графічний рушій, який надає можливості їх розробки. На основі отриманих даних розробити алгоритм синхронізації лицьової анімації віртуального персонажа із голосом користувача.

## **Об'єкт дослідження**

Обробка та синхронізація голосу людини із анімацією обличчя віртуального персонажа

## **Предмет дослідження**

Алгоритм синхронізації лицьової анімації із голосом у реальному часі

## Завдання

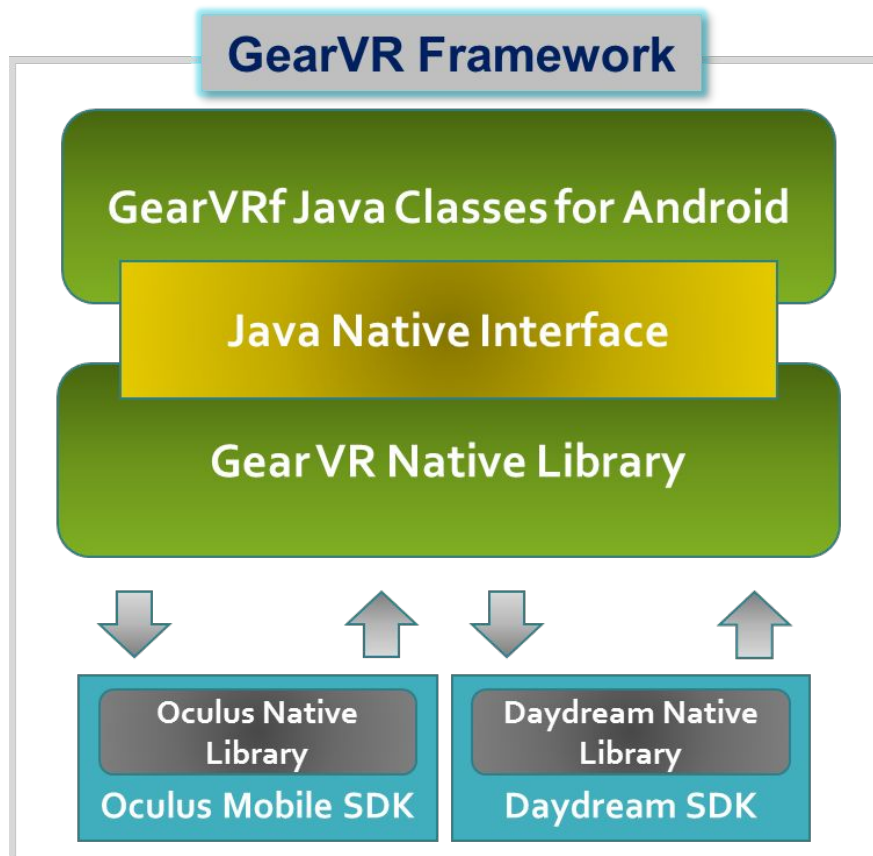
- Розглянути актуальні можливості віртуальної реальності, принципи побудови додатків під неї та графічні рушії, які дозволяють створювати такі додатки.
- Дослідити алгоритми розпізнавання голосу
- Розглянути та проаналізувати наявні відкриті системи розпізнавання голосу та обрати найоптимальнішу для досягнення мети.
- Розробити прототипу алгоритму та застосувати його у додатку віртуальної реальності, реалізованому на обраних системах.
- Проаналізувати отримані результати

## Віртуальна реальність. Графічний рушій.

Контент віртуальної реальності повинен відповідати усім стандартам індустрії, слідувати всім рекомендаціям з безпеки і комфорту. Було проведено аналіз необхідної наукової літератури по темі віртуальної реальності. Обрано графічний рушій Gear VR Framework, який задовільняє всім вимогам створення віртуального контенту.



# Gear VR Framework



## **Розпізнавання мовлення. Вибір системи розпізнавання.**

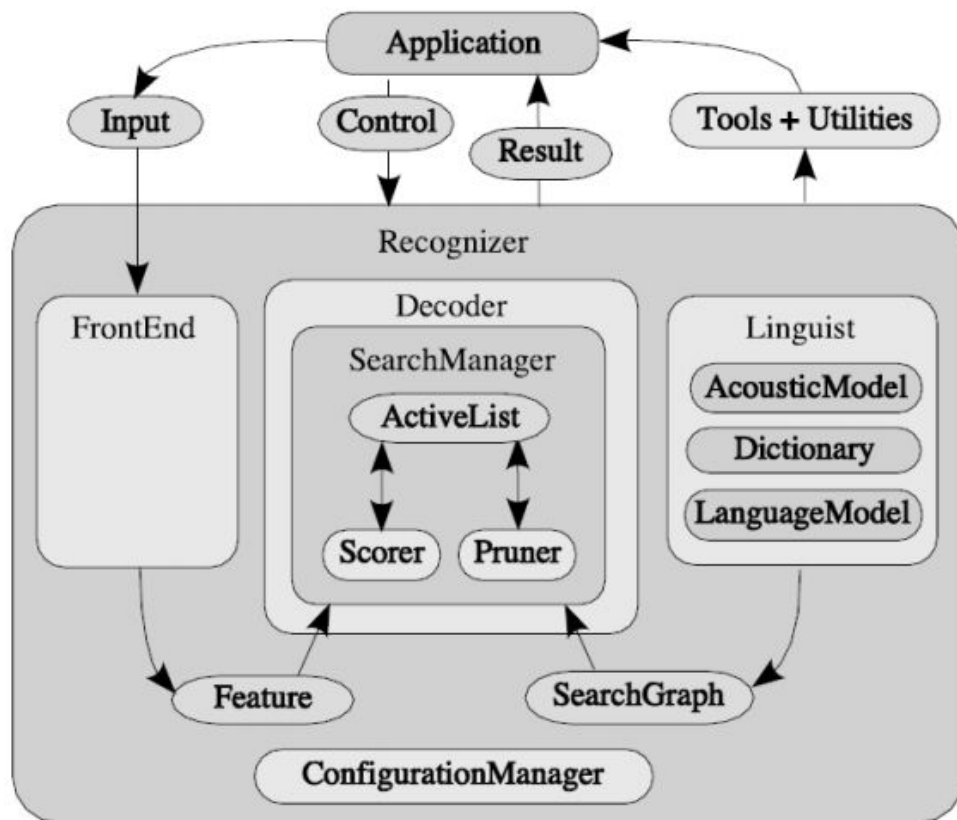
Для розпізнавання мовлення було обрано відкриту систему CMU Sphinx. Вона повністю безкоштовна, є бібліотека для застосування на мобільній платформі Android та надає можливість розпізнавання в режимі реального часу.

Sphinx - дикторонезалежна система розпізнавання безперервної мови, яка використовує приховану Марківську акустичну модель і n-грамну статичну модель.

The logo for CMU Sphinx, featuring the text "CMU Sphinx" in a blue, sans-serif font. The "CMU" is in all caps and a larger font size than "Sphinx", which is in title case. The text is centered on a light gray rectangular background.

CMU Sphinx

# CMU Sphinx

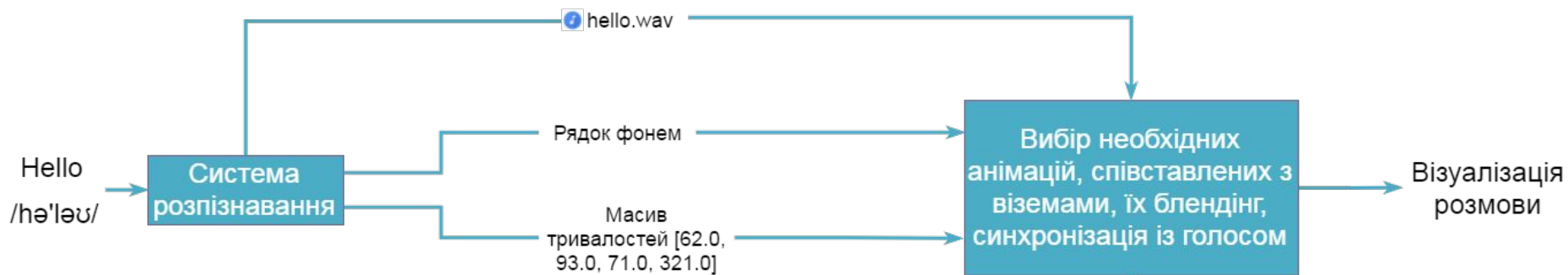


## Схема алгоритму














# Схематичний приклад роботи



Карта відповідності візем і фонем

 A Phonemes AH, AW, AY	 O Phonemes AA, AH, ER, OY	 E Phonemes AE, EH, EY, IH, IY, OW
 W, R Phonemes R, W	 T, S Phonemes D, DH, G, JH, S, T, TH, Z, ZH	 L, N Phonemes L, N, NG
 U, Q Phonemes AW, OW, UH, UW	 M, B, P Phonemes B, M, P	 F, V Phonemes AY, EY, F, OY, V, Y

## Карта співставлення візем до фонем

 <p>A</p> <table border="1"><tr><td>Phonemes</td></tr><tr><td>AH, AW, AY</td></tr></table>	Phonemes	AH, AW, AY	 <p>O</p> <table border="1"><tr><td>Phonemes</td></tr><tr><td>AA, AH, ER, OY</td></tr></table>	Phonemes	AA, AH, ER, OY	 <p>E</p> <table border="1"><tr><td>Phonemes</td></tr><tr><td>AE, EH, EY, IH, IY, OW</td></tr></table>	Phonemes	AE, EH, EY, IH, IY, OW
Phonemes								
AH, AW, AY								
Phonemes								
AA, AH, ER, OY								
Phonemes								
AE, EH, EY, IH, IY, OW								
 <p>W, R</p> <table border="1"><tr><td>Phonemes</td></tr><tr><td>R, W</td></tr></table>	Phonemes	R, W	 <p>T, S</p> <table border="1"><tr><td>Phonemes</td></tr><tr><td>D, DH, G, JH, S, T, TH, Z, ZH</td></tr></table>	Phonemes	D, DH, G, JH, S, T, TH, Z, ZH	 <p>L, N</p> <table border="1"><tr><td>Phonemes</td></tr><tr><td>L, N, NG</td></tr></table>	Phonemes	L, N, NG
Phonemes								
R, W								
Phonemes								
D, DH, G, JH, S, T, TH, Z, ZH								
Phonemes								
L, N, NG								
 <p>U, Q</p> <table border="1"><tr><td>Phonemes</td></tr><tr><td>AW, OW, UH, UW</td></tr></table>	Phonemes	AW, OW, UH, UW	 <p>M, B, P</p> <table border="1"><tr><td>Phonemes</td></tr><tr><td>B, M, P</td></tr></table>	Phonemes	B, M, P	 <p>F, V</p> <table border="1"><tr><td>Phonemes</td></tr><tr><td>AY, EY, F, OY, V, Y</td></tr></table>	Phonemes	AY, EY, F, OY, V, Y
Phonemes								
AW, OW, UH, UW								
Phonemes								
B, M, P								
Phonemes								
AY, EY, F, OY, V, Y								

# Порівняння із рішенням розпізнавання обличчя камерою

Алгоритм з розпізнаванням голосу

Переваги:

- Користувач не прив'язаний до місця, де знаходиться камера.
- Легка і продуктивна 3D-модель
- Можливість не передавати голос, а синтезувати його на стороні клієнта

Недоліки:

- Затримка при розпізнаванні
- Точність розпізнавання приблизно 70%
- Не зовсім повноцінна передача емоцій

# Порівняння із рішенням розпізнавання обличчя камерою

Алгоритм з розпізнаванням обличчя камерою

Переваги:

- Швидкодія
- Передача повного спектру емоцій

Недоліки:

- Недостатньо датасетів
- Важка 3D-модель персонажа з великою кількістю кісток
- Прив'язаність користувача до робочого місця
- Необхідність передачі звукового потоку

# Новизна

Алгоритм синхронізації звукової доріжки та анімації обличчя персонажа доволі давно застосовується у мультиплікації. Але у такому випадку в наявності є запис звуку, текст, та фонемі, тому синхронізація проблем не викликає. У даній роботі було створено схожий алгоритм та застосовано його у додатку віртуальної реальності для синхронізації лицьової анімації віртуального аватара з голосом користувача у режимі реального часу.

# Висновки

- Розглянуто основні принципи роботи додатків віртуальної реальності. Сформовано рекомендації для правильної побудови такого типу програм.
- Розглянуто графічні рушії. Проаналізовано принципи їх роботи та обрано оптимальний для простої розробки якісного контенту віртуальної реальності. Описано основні можливості та переваги даного рушія.
- Розглянуто Oculus Mobile SDK – набір інструментів для розробки додатків віртуальної реальності для мобільних платформ.

# Висновки

- Досліджено та описано принципи роботи систем розпізнавання мовлення людини. Проаналізовано та описано основні алгоритми розпізнавання, такі як приховані моделі Маркова та нейронні мережі.
- Виходячи із поставлених задач, обрано найбільш оптимальну систему розпізнавання мовлення та детально розглянуто та описано її архітектуру.
- На основі описаних вище результатів розроблено та протестовано прототип алгоритму синхронізації лицьової анімації віртуального персонажа із голосом користувача в режимі реального часу.

## **Іноваційність та можливість створення стартап проекту**

Даний алгоритм може бути застосовано у додатках віртуальної реальності у соціальному контексті. Тобто на основі нього можна розробляти ігрові системи, віртуальні світи або системи для навчання чи демонстрації рідного роду контенту. Алгоритм дозволить користувачам краще зануритись у віртуальний світ, з більшою довірою відноситись до своїх співрозмовників та передавати емоції за допомогою розмов. Також він надасть будь-якому віртуальному персонажу ще більш природного вигляду.