

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Радіотехнічний факультет
Кафедра прикладної радіоелектроніки**

«На правах рукопису»
УДК 004.05

До захисту допущено:
В.о. зав. кафедри
_____ Михайло СТЕПАНОВ
«__» _____ 2022 р.

**Магістерська дисертація
на здобуття ступеня магістра
за освітньо-професійною програмою «Інтелектуальні технології
радіоелектронної техніки»
за спеціальністю 172 «Телекомунікації та радіотехніка»
на тему: «Аналіз об'єктивної мовної та звукової метрики на основі
ViSQOL»**

Виконав:
студент 2 курсу, групи РЕ-11мп

Ліпський Олег Геннадійович
Прізвище, ім'я, по батькові



Керівник:

ст. викладач Головня Вікторія Мілентіївна
Посада, науковий ступінь, вчене звання,
Прізвище, ім'я, по батькові



Рецензент:

ст. викладач Захарченко Оксана Степанівна,
Посада, науковий ступінь, вчене звання,
Прізвище, ім'я, по батькові

Засвідчую, що у цій магістерській
дисертації немає запозичень з праць
інших авторів без відповідних
посилань.

Студент 
Київ – 2022 року

**Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»**

Радіотехнічний факультет

Кафедра прикладної радіоелектроніки

Рівень вищої освіти – другий (магістерський)

Спеціальність – 172 «Телекомунікації та радіотехніка»

Освітньо-професійна програма «Інтелектуальні технології радіоелектронної техніки»

ЗАТВЕРДЖУЮ

В.о.зав. кафедри

_____ Михайло СТЕПАНОВ

« » _____ 2022 р.

ЗАВДАННЯ

на магістерську дисертацію студента

Ліпському Олегу Геннадійовичу

1. Тема дисертації «Аналіз об'єктивної мовної та звукової метрики на основі ViSQOL»
науковий керівник дисертації Головня Вікторія Мілентіївна, ст. викладач, к.т.н.
затвержені наказом по університету від «09» листопада 2022 р.
№4137-С
2. Термін подання студентом дисертації 11 грудня 2022 року
3. Об'єкт дослідження Оцінки якості записаних аудіокліпів через доступні численні об'єктивні метрики, тобто метрики, отримані шляхом вимірювання аудіосигналу. Приклади фізичних показників включають відношення сигнал/шум (SNR), повне гармонійне спотворення (THD) і спектральне (амплітуда) спотворення.
4. Вихідні дані Мовна база, що включає достатню кількість навчальних і тестових прикладів, бібліотека сигналів, середовище відтворення.
5. Перелік завдань, які потрібно розробити Провести аналіз та порівняння мовних метрик, провести дослідження з готовими бібліотеками, зробити висновки.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу Графіки, таблиці

7. Орієнтовний перелік публікацій Статті, тези, дослідження.

9. Дата видачі завдання 05 вересня 2022 року

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Отримання теми магістерської дисертації	05.09.2022р.	
2	Розробка плану магістерської дисертації	10.09. 2022р.	
3	Початок збору інформації для дослідження	24.09.2022р.	
4	Аналіз метрик	10.10.2022р.	
5	Проведення досліджень	20.10.2022р	
6	Формування висновків	27.10.2022р	
7	Розробка стартап-проекту	05.11.2022р	
	Оформлення магістерської дисертації	09.12.2022р.	

Студент



Олег ЛІПСЬКИЙ

Керівник



Вікторія ГОЛОВНЯ

РЕФЕРАТ

Магістерська дисертація на тему: «Аналіз об'єктивної мовної та звукової метрики на основі ViSQOL» містить 65 сторінок, 10 рисунків, 26 таблиць.

Метою магістерської дисертації є опис та аналіз існуючих мовних метрик, та практичне, ефективне їх використання. Проведення дослідів та формування висновків, на основі ViSQOL.

Актуальність теми: дослідження в данній роботі є актуальними, оскільки правильне сприйняття звуків є одним із ключових способів отримання інформації, її якість потрібна не тільки здоровим людям, але і людям з порушенням слуху, діагностику яких можна покращити.

Наукова новизна: нові сфери застосування та практичне застосування, цих досліджень.

У результаті повного аналізу, досліджень, отримаємо відповіді на перспективність використання мовних метрик та їх практичне застосування.

Ключові слова: звуки, мовлення, мовні метрики, якість звуку та мови, якість досвіду споживачів.

ABSTRACT

Master's thesis on the topic: "Analysis of objective speech and sound metrics based on ViSQOL" contains 65 pages, 10 figures, 26 tables.

The purpose of the master's dissertation is the description and analysis of existing language metrics, and their practical, effective use. Conducting experiments and forming conclusions based on ViSQOL.

Relevance of the topic: the research in this work is relevant, because the correct perception of sounds is one of the key ways of obtaining information, its quality is needed not only by healthy people, but also by people with hearing impairment, whose diagnosis can be improved.

Scientific novelty: new areas of application and practical application of these studies.

As a result of a complete analysis and research, we will receive answers to the prospects of using language metrics and their practical application.

Keywords: sounds, speech, speech metrics, sound and speech quality, quality of consumer experience.

**Пояснювальна записка
до магістерської дисертації
на тему: Аналіз об'єктивної мовної та звукової метрики
на основі ViSQOL**

ВСТУП.....	11
1 ТЕОРЕТИЧНІ ЗАСАДИ МОВНИХ ТА ЗВУКОВИХ МЕТРИК	12
1.1 Основні засади передачі медіа	12
1.2 Оцінки якості	13
1.3 Методи і засоби розпізнавання звуків та мови.....	15
1.4 Особливості розпізнавання звуків та мови при порушенні слуху .	17
1.5 Постановка задачі.....	20
2 ДОСЛІДЖЕННЯ ОБ'ЄКТИВНОЇ МОДЕЛІ ЯКОСТІ МОВЛЕННЯ VISQOL	21
2.1 Основні принципи роботи ViSQOL.....	21
2.2 Суб'єктивні та об'єктивні тестування	22
2.3 Інтеграції ViSQOL3 у різні проекти та їх практичне використання	31
2.3.1 Hangouts Meet.....	31
2.3.2 Кодек Opus	33
3 ВИКОРИСТАННЯ ТА ПОКРАЩЕННЯ	38
3.1 Зміни та покращення.....	38
3.2 Бібліотека та двійковий файл C++.....	39
3.3 Тонкомасштабне вирівнювання часу	40
3.4 Пороги тиші	41
3.5 Модель NSIM до MOS	41
4 СТАРТАП ПРОЕКТ	43
4.1 Опис ідеї проекту (товару, послуги, технології).....	43
4.2 Технологічний аудит ідеї проекту	45

4.3 Аналіз ринкових можливостей запуску стартап-проекту	46
4.4 Розроблення ринкової стратегії проекту.....	53
4.5 Розроблення маркетингової програми стартап-проекту	55
ВИСНОВКИ.....	59
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ	63

ПЕРЕЛІК СКОРОЧЕНЬ

QoE (Quality of experience) — якість досвіду споживачів

SNR — відношення, сигнал/шум

THD — повне гармонійне спотворення

DNN — глибока нейронна мережа

ODG — оцінка якості об'єктивної різниці

MOS — середня оцінка суб'єктивної групи

DSP — процесор цифрового сигналу

CI — кохлеарний імплант

NSIM — індекс подібності нейрограми

WebRTC — інтернет-протокол із відкритим кодом, призначений для організації голосового та відеозв'язку через інтернет у режимі реального часу.

VAD — визначення голосової активності

ERB — еквівалентна прямокутна смуга пропускання

NSIM (ціле) — встановлює кількість діапазонів, які одночасно оптимізуються за допомогою алгоритму RMM-DIIS.

ВСТУП

За останні роки, моделі об'єктивної оцінки якості звуку і мовлення, стали все частіше використовуватися для оцінки або моніторингу якості мови та звуку, для користувачів.

Оцінюючи, сприйману якість, за допомогою суб'єктивної шкали оцінювання, наприклад якість прослуховування (відмінно, добре, погано), ці методи, забезпечують швидкий і повторюваний спосіб оцінки, досвіду клієнтів. Типові програми включають оцінку якості звуку, вибір кодеків або іншого обладнання та вимірювання якості мережі.

В цій роботі представлений аналіз існуючих методів, окреслюючи основні підходи до інтрузивних, неінтрузивних і параметричних моделей, та обговорення деяких з них, обмеження та сфери майбутньої роботи. Терміни показника — якість аудіо, інтрузивне та неінтрузивне тестування, об'єктивні моделі, оцінка якості, якість мовлення.

Оцінка якості звуку була широко досліджена в області обробки сигналів. Повноцінні об'єктивні метрики: POLQA, ViSQOL, були розроблені для оцінки якості аудіо, покладаючись лише на експерименти з рейтингом людини.

Щоб оцінити якість аудіо за допомогою нових методів обробки аудіо, постійно потрібно порівнювати об'єктивні показники якості. Тестування різних реалізацій однієї метрики та оцінка нових наборів даних є фундаментальними і постійними ітеративними діями.

1 ТЕОРЕТИЧНІ ЗАСАДИ МОВНИХ ТА ЗВУКОВИХ МЕТРИК

1.1 Основні засади передачі медіа

Потокове передавання медіа тепер є усталеним способом прослуховування музики та перегляду фільмів і телевізійного вмісту. Обмеження пропускної здатності мережі є змінними для різноманітних пристроїв, на яких споживається вміст: мобільний, комп'ютер, і т.д. Як наслідок, розповсюджувачі контенту, такі як YouTube, Netflix або Spotify, повинні підтримувати низку кодеків і бітрейтів «обробок», щоб оптимізувати якість досвіду споживачів (QoE). Розподіл пропускної здатності для потокового мультимедіа на мобільних пристроях повинен оптимізувати QoE для органів чуття, і в той час як обробки 256 Кбіт/с можуть забезпечити QoE, який неможливо відрізнити від нестисненого аудіо. Будь-яку економію пропускної здатності аудіо можна використати для покращення відеопотоку. Щоб адаптувати до середовища, від смартфона до домашнього кінотеатру, як відео, так і аудіовміст передаються в стисненому вигляді з використанням схем стиснення з втратами. Психоакустичні схеми стиснення мають вихідні сигнали, оптимізовані з точки зору слухової системи людини. Стандартна середньоквадратична похибка або сегментарне співвідношення сигнал/шум не дуже підходять для оцінки якості звуку. Для сучасних кодеків оцінки якості звуку відбуваються завдяки оптимізованому розподілу бітів з врахуванням психоакустики моделі. Відповідні об'єктивні показники можуть допомогти автоматизувати оцінку змін QoE в результаті цього процесу перекодування. Обробки звуку, які зазвичай використовуються для потокового передавання, це кодеки AAC-HE та AAC-LC на чотирьох бітрейтах, на основі MP3 і OPUS. Прослуховування було суб'єктивно оцінено в контексті стереомузики, щоб дослідити, чи сприймали слухачі відмінності в якості звуку, для перевірених кодеків і чи вплинув режим презентації навушники/колонки на результати. Результати показали, що слухачі виявили, що деякі обробки були помітно погіршені, тоді як інші не відрізнялися від вихідних нестиснутих стерео-аудіо

зразків. В роботі, надалі, порівняємо чотири об'єктивні показники з результатами суб'єктивних тестів слухачів, проведених за допомогою навушників, щоб оцінити їх придатність для вимірювання якості звуку для кодеків з низькою швидкістю передачі даних.

1.2 Оцінки якості

Для оцінки якості записаних аудіокліпів доступні численні об'єктивні метрики, тобто метрики, отримані шляхом вимірювання аудіосигналу. Приклади фізичних показників включають відношення сигнал/шум (SNR), повне гармонійне спотворення (THD) і спектральне (амплітуда), спотворення.

При оцінці сприйнятої якості, PESQ [1-2] і POLQA [3-4] стали стандартами для мовлення, а, на практиці, також і для загального аудіо, незважаючи на те, що вони спочатку були розроблені лише для якості мовлення. Є й інші приклади: PEAQ [5] і PEMO-Q [6]. Для більшості цих показників потрібні комерційні ліцензії. ViSQOL [7] і ViSQOLAudio [8] (далі спільно згадуються як ViSQOL) є вільно доступними альтернативами для мовлення та аудіо. Ці показники постійно розширюються, щоб охопити додаткові домени. Наприклад, робота над AMBIQUAL [9] поширює ті самі принципи, що використовуються у ViSQOLAudio, на область амбісоніки.

Удосконалення в обробці мови та аудіо, наприклад усунення шумів і стиснення, викликають потребу в покращенні оцінки якості. Мовні та аудіокодеки досягають все нижчих корисних бітрейтів. Таким чином, можливо, варто проаналізувати продуктивність ViSQOL для цього розширеного домену. Крім того, існує ряд генеративних моделей глибокої нейронної мережі (DNN), які відтворюють форму сигналу шляхом вибірки з розподілу вивчених параметрів. Одним із прикладів є кодер із низьким бітрейтом на основі WaveNet [10], який є генеративним за своєю природою. Існують інші генеративні моделі на основі DNN, включаючи SampleRNN [11] і WaveGlow [12], з позитивними результатами, які припускають, що ця

тенденція триватиме. Ці генеративні моделі зазвичай не піддаються належному аналізу за допомогою існуючих, повних, еталонних показників якості мовлення. Описана данна робота, не пропонує вирішення генеративної проблеми, слід визнати обмеження поточної моделі, щоб заохотити розробку рішень.

Об'єктивні показники

З кінця 1980-х років активно розробляються об'єктивні показники якості звуку. Перцептивна оцінка якості звуку (PEAQ), була стандартизована, як рекомендація ITU-R BS.1387 з 1998 по 2001 рік. Існує дві версії: базова, оптимізована для швидкості, і розширена модель, яка до основної моделі ШПФ, для підвищення точності, додає модель вуха на основі банку фільтрів. Обидві версії виробляють кілька моделей, вихідні змінні яких зіставляються з оцінкою якості об'єктивної різниці (ODG), за допомогою багаторівневої нейронної мережі. Шкала ODG є об'єктивним наближенням оцінки суб'єктивної різниці, яка використовується в рекомендації ITU-R BS.1116, щоб визначити невеликі порушення звуку. Десятиліттям пізніше, рекомендація ITU-T P.863 стандартизувала нову об'єктивну метрику для вимірювання якості мовлення під назвою POLQA. Метрика POLQA була розроблена для оцінки якості мовлення та може працювати у вузькосмуговому режимі, (телефонна якість; 300–3400 Гц) або надширокосмуговому (SWB), режимі (50–14000 Гц). Показано, що POLQA має потенціал якості звуку. Альтернативна модель якості мовлення, під назвою ViSQOL, була адаптована для якості звуку тестування. Ця адаптована метрика називається ViSQOLAudio, описана та оцінена в цій роботі. З метою порівняльного аналізу комерційно випущена версія POLQA, яка відповідає стандарту P.863, тестується разом із базовою та розширеною версіями PEAQ (усі постачаються Opticom, GmbH).

1.3 Методи і засоби розпізнавання звуків та мови

У сучасних системах аудіо-мовленнєвих комунікацій, кодування з низькою швидкістю потоку, стало повсюдним, діапазон від швидкості передачі даних нижче 4 кбіт/с (для деяких військових і програм супутників), приблизно до 128 кбіт/с (для зберігання музики або цифрового радіо). Але кодеки з низькою швидкістю передачі даних і елементи обробки такі, як придушення відлуння та зменшення шуму в телефонних мережах, є дуже нелінійними процесами, залежними від часу та сигналу, і їхній вплив на якість сприйняття є значним. і важко передбачити.

До 1990-х років стандартним способом вимірювання якості цих процесів, було провести суб'єктивний тест, який дає середня оцінка суб'єктивної групи (MOS) якості по кожній умові, що перевіряється. Але використання людей у контрольованому середовищі є дорогим і повільним, тому, поки що, суб'єктивні тести є ідеальним способом прийняття суттєвих системних рішень, таких як вибір кодека для міжнародного стандарту, і вони непридатні для повсякденного оцінювання якості. Метою об'єктивного вимірювання є автоматична оцінка суб'єктивного MOS на основі вимірювань системи.

Завдяки наявній сьогодні обчислювальній потужності, всі методи у поточному використанні можна застосовувати в режимі реального часу — для параметричного голосу, моделі, оцінки через IP (VoIP). Єдиний недорогий, тестовий зонд може оцінювати тисячі одночасних дзвінків, тоді як більш інтенсивні обчислювальні моделі на основі сигналу, все ще можуть підтримувати декілька паралельних каналів з одного процесора цифрового сигналу (DSP). Це означає що доцільно використовувати об'єктивні показники якості для оптимізації мереж, щодо їх якості та пропускної здатності, вартості або моніторингу мережі, на основі досвіду клієнтів.

Важливим фактором у розвитку точних об'єктивних мір, стало використання моделей людського сприйняття. Дослідники в галузі психофізики побудували моделі кількох великомасштабних властивостей периферичної слухової системи, включаючи сприйняття гучності, частоти та

маскування [1]. Способи використання таких прийомів і перцептивна аналогія, значно змінюється, залежно від типу об'єктивної міри. Інтрузивні моделі, обговорювані далі, порівняні з оригінальним, тестовим сигналом із погіршеною версією, яка була оброблена системою. Такі моделі також називають порівняльними, або повними, еталонними. Останні нав'язливі моделі працювали шляхом перетворення обох сигналів за допомогою моделі сприйняття, відтворювали деякі з ключових властивостей слуху, а потім, за допомогою комп'ютера, вимірювали відстані в перетвореному просторі та використовували їх для оцінки MOS [2–7].

Неінтрузивні моделі можна використовувати в кількох конфігураціях. На основі неінтрузивного сигналу моделі (також відомі, як моделі без посилання або односторонні моделі), які знаходяться в зародковому стані порівняно з інтрузивними моделями. В них оцінюють MOS шляхом обробки погіршеного вихідного мовного сигналу [8–11]. Декілька методів на основі сигналу, зосереджені на моделях прямої мови або мовного сигналу, хоча багато хто використовує деякі аспекти сприйняття наприклад, гучність шуму. Навпаки, неінтрузивні параметричні моделі зазвичай мають відсутність звукового сигналу для обробки (і тому обмежене використання методів сприйняття), а натомість оцінюють MOS за виміряними властивостями основного передавання та/або терміналу, наприклад: затримка, рівні мовлення та шум [12], характеристики мережі VoIP [13], [14] або показники прийому стільникового радіо [15].

Параметричні моделі також широко використовуються для мережевого планування побудувати оцінки MOS на основі табличних значень, таких як тип кодека, бітрейт, затримка, статистика втрат пакетів тощо.

Цей останній підхід вимагає повної характеристики системи, що тестується, і тому її іноді називають підходом скляної коробки. Методи, про які немає знання в системі, що тестується, називаються «чорним ящиком». Більшість реальних систем вимірювання якості, використовують змішані

підходи, що вимагають таких припущень, як тип кодека і аудіотермінали. Огляд наведено на рис. 1.

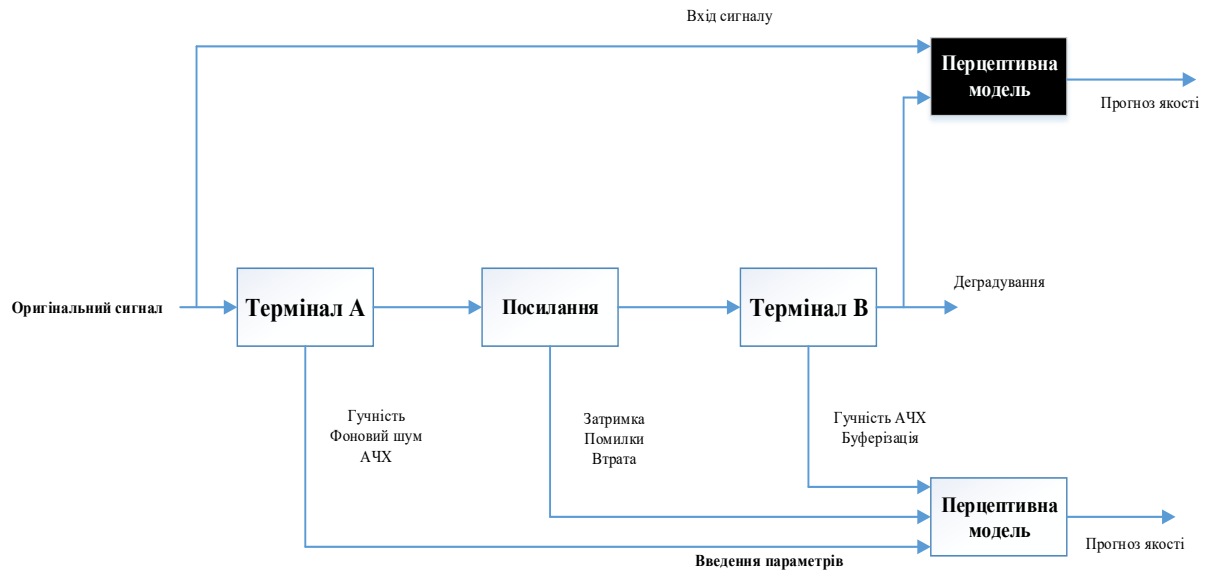


Рисунок 1 — Система вимірювання якості «Black Box»

Важливо відзначити обмеження об'єктивних моделей. Вони навчаються на суб'єктивних тестових даних, які обов'язково охоплюють лише обмежений набір умов і виникають помилкам при голосуванні, а на моделі впливають інші змінні. Це обмежує їх точність [7], [17]. Ці проблеми та інші фактори виникають при застосуванні об'єктивних моделей.

1.4 Особливості розпізнавання звуків та мови при порушенні слуху

Завдяки технологічним досягненням у пристроях кохлеарних імплантатів (СІ) більшість користувачів СІ тепер можуть досягти надійної розбірливості мовлення в контрольованих тихих сценаріях, особливо під час передбачуваних розмов. З іншого боку, відомо, що спотворення навколишнього середовища, такі як додатковий шум (та їх комбінований ефект), значно погіршують розбірливість мовлення.

Наприклад, реверберація та шум, спотворюють важливу інформацію, модуляції мовної оболонки, що робить надзвичайно складним сприйняття для користувачів СІ, наприклад: модуляції висоти тону, переходів формант, тембру та меж слова/складу, вводять небажані ефекти маскування та

спричиняють погану локалізацію звуку. Щоб подолати ці обмеження та покращити розбірливість мовлення в повсякденному середовищі, останні дослідження були зосереджені на розробці алгоритмів покращення мовлення, таких як придушення шуму, вибір каналу та деверберація.

Щоб оцінити вплив умов навколишнього середовища на розбірливість мовлення користувачів СІ, а також підвищення розпізнавання після покращення мовлення, зазвичай використовують два суб'єктивні підходи до тестування. Перший використовує вокодовану мову для імітації обробки СІ та представляє вокодовану мову слухачам із нормальним слухом (NH) для ідентифікації. Другий підхід є більш прямим і надає користувачам СІ мовні стимули, деградовані шумом (та/або підсилені). Проте суб'єктивне тестування є дуже дорогим і трудомістким.

Об'єктивне вимірювання розбірливості мовлення, з іншого боку, замінює слухачів обчислювальним алгоритмом, таким чином дозволяючи автоматизований, повторюваний, швидкий і економічно ефективний моніторинг розбірливості. Крім того, для покращення мовлення об'єктивні показники можуть відігравати важливу роль, оскільки оцінку розбірливості «на місці» можна використовувати для точного налаштування параметрів алгоритму (наприклад, налаштувань банку фільтрів СІ).

Нарешті, об'єктивні показники дозволяють проводити повторюване та недороге кількісне порівняння між кількома пристроями СІ. Показники об'єктивної зрозумілості або якості, можна в цілому класифікувати як інтрузивні (також відомі як двосторонні або повні посилення) або неінтрузивні (односторонні або без посилення) залежно від потреби в еталонному, чистому сигналі, чи ні, відповідно. Інтрузивні показники мають перевагу в тому, що вони здатні безпосередньо оцінити кількість і тип спотворень у спотвореному сигналі. Хоча обидва можуть використовуватися під час розробки алгоритму вдосконалення або для оцінки/порівняння різних пристроїв СІ, інтрузивні метрики не можна використовувати в практичних програмах реального часу, оскільки в цьому випадку еталонний, чистий сигнал недоступний.

Оскільки неінтрузивні метрики не вимагають чистого опорного сигналу, їх можна застосувати для кількісної характеристики підвищення розбірливості, досягнутого за допомогою алгоритму сліпого покращення мови (деверберації), безпосередньо на пристрої. Вони також дозволяють розробляти алгоритми покращення зрозумілості. Зазвичай об'єктивні показники розробляються та оцінюються за допомогою слухачів із нормальним слухом як цільових, з кількома дослідженнями з використанням вокодованого мовлення для імітації слуху СІ. Нещодавно було оцінено кілька об'єктивних показників щодо вокодованого мовлення, погіршеного, а також мовлення, погіршеного шумом і реверберацією, і представленого безпосередньо користувачам СІ.

У цих дослідженнях було помічено, що існуючі нав'язливі метрики не сильно корелюють із розбірливістю СІ користувача, в трьох умовах навколишнього середовища, а саме: лише шум, лише реверберація та шум плюс реверберація. У випадку лише реверберації нещодавно запропонований неінтрузивний показник, названий співвідношенням енергії модуляції мови та реверберації (SRMR), показав позитивний результати.

У цій роботі, буде досліджено та проаналізовано, продуктивність існуючих об'єктивних показників, з подальшим порівнянням їх продуктивності з показниками зрозумілості користувачів СІ.

1.5 Постановка задачі

Завдання роботи складається з аналізу методу розпізнавання звуків та мови, їх обробка, передача та сприйняття, за допомогою мовної та звукової метрики на основі ViSQOL, та проведення дослідів з подальшим висновком, та визначення перспективи використання.

Призначення задачі: проведення дослідів та аналізу, розпізнавання звуків та мови, за допомогою метрики.

Вимоги до програмно-технічного забезпечення:

Апаратне забезпечення:

1. Процесор – Intel Core i3 3110m 2.4 GHz;
2. Пам'ять – від 3 Gb;
3. Жорсткий диск від 500 Gb;
4. Відеокарта;
5. Монітор;
6. Клавіатура та миша;

Програмне забезпечення:

1. Операційна система – Microsoft Windows 10;
2. Бібліотека звуків.
3. Мова програмування – Python;
4. Методи захисту

Висновки

Перший розділ присвячений огляду завдання. У цьому розділі:

1. Проведений первинний аналіз обробки звуку та мови.
2. Проведено аналіз інформації, щоб виділити завдання для розпізнавання мови та описано загальну структуру системи автоматичного розпізнавання мови.
3. Визначені методи дослідження.
4. Визначення необхідної матеріальної та програмної бази для виконання дослідження.

2 ДОСЛІДЖЕННЯ ОБ'ЄКТИВНОЇ МОДЕЛІ ЯКОСТІ МОВЛЕННЯ ViSQOL

2.1 Основні принципи роботи ViSQOL

ViSQOL спочатку був розроблений з поліноміальним відображенням показника індексу подібності нейрограми (NSIM) до MOS [13], а ViSQOLAudio було розширено для використання моделі, навченої для опорної векторної регресії. Відтоді з'явилися моделі глибоких нейронних мереж, які застосовувалися до моделей якості мови [14–15]. Такі підходи є перспективними та потенційно можуть вирішити деякі проблеми, які не можуть існуючі архітектури. Хоча такі нові напрямки явно цікаві та вимагають подальшого дослідження, вони швидко розвиваються. Оскільки ViSQOL було представлено та перевірено у великій кількості експериментів, які підтвердили його застосування до низки випадків використання [7–8], [16–19], вважаю його відносно добре проаналізованим для відомих наборів даних, які мають тенденцію бути меншими та відносно однорідними. Натомість звертаю увагу на дані та типи проблем, які виникають «у дикій природі» в командах Google, які не залежали від розробки ViSQOL, а також на ітераційні вдосконалення, отримані в результаті цього аналізу. Його адаптація до цих випадків призвела до різноманітних покращень у зручності використання та продуктивності, разом із відгуками та уявленнями про дизайн майбутніх систем для оцінки якості сприйняття. Оскільки ці вдосконалення заповнюють «сліпі плями» в наборах даних, очікується, що вони не покращать результати на цих наборах даних. Поки не буде створено більш різноманітні набори даних суб'єктивної оцінки, перевірка в реальному світі здається розумним компромісом.

Крім того, поряд із покращенням якості оцінки MOS на основі даних реального світу, можна бути стурбованим тим, як зробити ViSQOL більш корисним для спільноти з точки зору практичних інструментів. Незважаючи

на те, що ViSQOL був доступний через реалізацію MATLAB, існували непотрібні перешкоди для його використання в деяких випадках, наприклад: виробництво та безперервне інтеграційне тестування g, (яке, можливо, потребує запуску на сервері), або може не мати доступних ліцензій MATLAB. У результаті було вирішено повторно реалізувати його в C++, оскільки це широкодоступна та розширювана мова, яку можна використовувати в інших мовах.

2.2 Суб'єктивні та об'єктивні тестування

Суб'єктивні випробування проводилися з використанням методології тестування MUSHRA, яка визначена в рекомендації ITU-R BS.1534.1. Для тестування якості звуку, існують методології, але для кодека з низькою швидкістю передачі даних тестування, MUSHRA є хорошим компромісом між рейтинговим тестом абсолютної категорії (ITU-R BS.1284-1) і тестом на майже невизначені порушення (ITU-R BS.1116-2). Повідомлялося про помилки в тестах MUSRHA через розподіл стимулів і ефекти вирівнювання діапазону [19], але MUSHRA використовувався в різноманітних тестах, показуючи хорошу здатність ранжувати кодеки з низькою швидкістю передачі даних. [20,21] Дотримуючись методології MUSHRA, слухачам було представлено маркований еталон і ряд непозначених тестових зразків (стимулів). Немічені зразки ранжували за допомогою числової безперервної шкали від 0 до 100 у п'яти описових інтервалах: погано (0–20); бідний (20–40); справедливий (40–60); добре (60–80); і відмінно (80–100). Незмінена версія еталона та два зразки прив'язки (версії еталона з фільтром низьких частот) були приховані серед тестованих обробок. Десять слухачів ранжували 12 музичних зразків для 10 обробок (включаючи приховане посилання та 2 прив'язки). Випробування з використанням Sennheiser HD558, високоякісні з відкритою спинкою навушники, показали найменшу дисперсію між слухачами та використовуються як основна правдива суб'єктивна якість для об'єктивної

метричної оцінки. Суб'єктивні оцінки якості прослуховування відтворені тут, на рис. 2.1. Таблиця 2.1 детально описує тест матеріал, який складався зі стерео музичних зразків тривалістю 7–15 с, що охоплювали різні музичні звуки. Тести матеріалів були отримані з компакт-дисків і музичної бази даних EBU [22] і всі спочатку дискретизувалися на 48 кГц або 44,1 кГц, 16 біт стерео (тому для двоканального аудіо 44,1 кГц швидкість становить 1411,2 кбіт/с). Довідкові файли PCM WAV були створені на 48 кГц для всіх файлів. Потім їх було закодовано та передискретизовано за допомогою `ffmpeg` із кодувальником Fraunhofer AAC для AAC, `libmp3lame` для MP3 та `libopus` для Opus 1.1, щоб створити діапазон обробок, табл. 2.2. Усі зразки були відформатовані як файли WAV PCM для представлення та оцінені авторами, щоб переконатися, що між еталонними зразками та методами прослуховування не було виявлено різниці рівнів. Кожну об'єктивну метрику використовували для порівняння 12 еталонних зразків із 10 їх обробками.

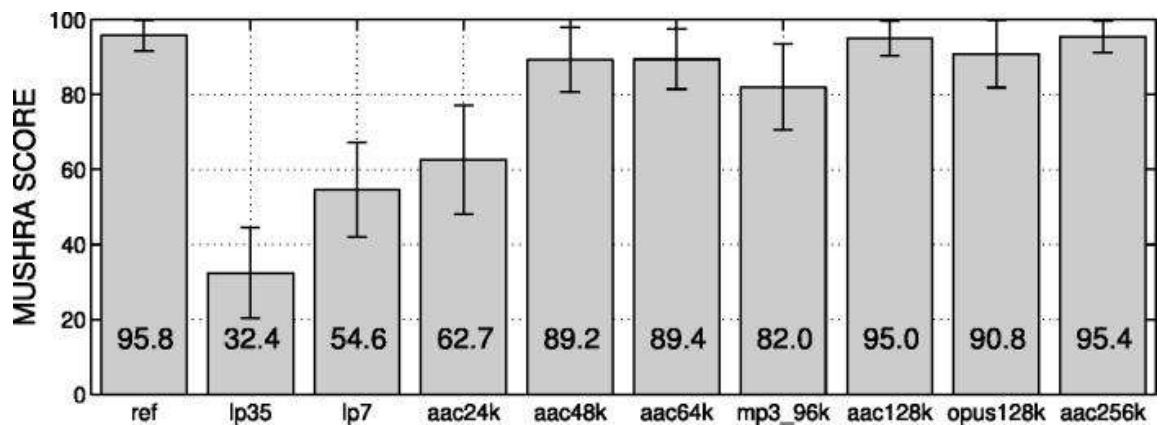


Рисунок 2.1— Середні суб'єктивні результати MUSHRA

Для 12 зразків музики та 10 слухачів у навушниках. Смужки помилок показують 95 довірчих інтервалів. `lp35` і `lp7` — умови прив'язки з фільтром низьких частот.

Таблиця 2.1 – Бібліотека звуків

Мітка	Тип музики	Джерело	Мітка	Тип музики	Джерело
Боз	Рок/R&B (Боз Скегс)	CD	Глок	Глокеншпіль	EBU
Сталевий	М'який рок (Стілі Ден)	CD	Контрафагот	Арпеджіо/мелодійна фраза	EBU
Кастаньети	Кастаньети	EBU	Клавесин	Арпеджіо/мелодійна фраза	EBU
Місячне світло	Фортепіано (Місячна соната)	CD	Сопрано	Співачка сопрано	EBU
Вега	Вокал (Сюзанна Вега)	CD	Гітара	Ларрі Коріел	EBU
Равель	Цигане	EBU	Штрауса	Р. Штраус (оркестр)	EBU

Таблиця 2.2 – Бібліотека рішень

Тип	Пропускна здатність	Бітрейт (кб/с)	Тип	Пропускна здатність	Бітрейт (кб/с)
поширення	22 кГц/Raw-PCM	1536	aac-he	20 кГц (повний діапазон)	48
якір 1	3,5 кГц вузькосмуговий	256	aac-he	20 кГц (повний діапазон)	64
якір 2	Широкосмуговий 7 кГц	512	aac-lc	20 кГц (повний діапазон)	128
mp3	16 кГц (SWB)	96 (CBR)	opus	20 кГц (повний діапазон)	128
aac-he	20 кГц (повний діапазон)	24	aac-lc	20 кГц (повний діапазон)	265

Визначити вплив кодеків і бітрейтів на звук важко. Окрім складного, за своєю суттю, завдання кількісної оцінки суб'єктивних думок слухачів щодо якості, вимірювання якості звуку, має усунути вплив інших факторів на якість досвіду слухача. Вимірювання з гучномовців вказує на те, що акустика приміщення може впливати на результати, тоді як вибір семплів, методологія оцінювання, кількість і досвід слухачів можуть потенційно впливати на результати. Музичні зразки, чутливі до зменшення бітрейту та частотної характеристики, були ретельно відібрані для використання кодеків.

Тести слухача проводилися в звукоізолюваній студії звукозапису та повторювалися з використанням двох типів навушників і знову з гучномовців, оцінити вплив обладнання для прослуховування.

Результати з використанням студійної якості навушників Sennheiser HD558 використовувалися, як суб'єктивна основа правди. Хоча результати іншого обладнання для прослуховування показали ті самі тенденції, були відмінності, які могли бути спричинені маскуванням артефактів стиснення в нижчій якості навушників, через кімнатну реверберацію для гучномовців. Вибір MUSHRA як атестування, методологій дозволила слухачам порівнювати методи прослуховування та ранжувати їх за безперервною шкалою. Суб'єктивні результати прослуховування та параметрів обладнання показали, що слухачі були надзвичайно послідовними у своїх оцінках. Дані та результати дали впевненість, що експерименти з тестуванням прослуховування стануть корисною відправною точкою для оцінки потенціалу об'єктивних показників для прогнозування впливу кодеків з низьким бітрейтом на якість звуку.

Результати дослідження

Рисунок 2.2 відтворює результати суб'єктивних тестів слухача. Кодеки з вищим бітрейтом мали показники якості, які майже не відрізнялися від еталонного нестисненого аудіо. Прив'язки з фільтром низьких частот і AAC-HE 24 кбіт/с отримали значно нижчі оцінки відносно представлених об'єктивних результатів, та згруповані за типом прослуховування. AAC-LC 128 кбіт/с і AAC-LC 256 кбіт/с отримали найвищий рейтинг за всіма об'єктивними показниками з точки зору їхніх середніх значень. Однак смужки помилок показують, що результати для даного прослуховування значно відрізнялися в різних зразках. Наприклад, виміряним PEAQ, стандартне відхилення для AAC-LC 256 кбіт/с.

Аналіз результатів для кожного зразка, що викликав це велике стандартне відхилення, показав, що обидва PEAQ моделі були оцінки, які

групувалися бімодально з чотирма семплами від 0 до -1 (боз, кастаньєти, сталь, вега) та іншими шістьма семплами від -3 до -4 (сопр, равель, гітара, клавесин, контрафагот, штраус). Це показано на рис. 2.2 кружками для кожного зразка на діаграмі PEAQ-Advanced для AAC-LC 256 кбіт/с. Це вказує на те, що для цього конкретного кодека PEAQ демонструє чутливість до типу семплу, чого не відчували суб'єктивні слухачі. Для цих кластерів не спостерігалось очевидної закономірності щодо типу вибірки.

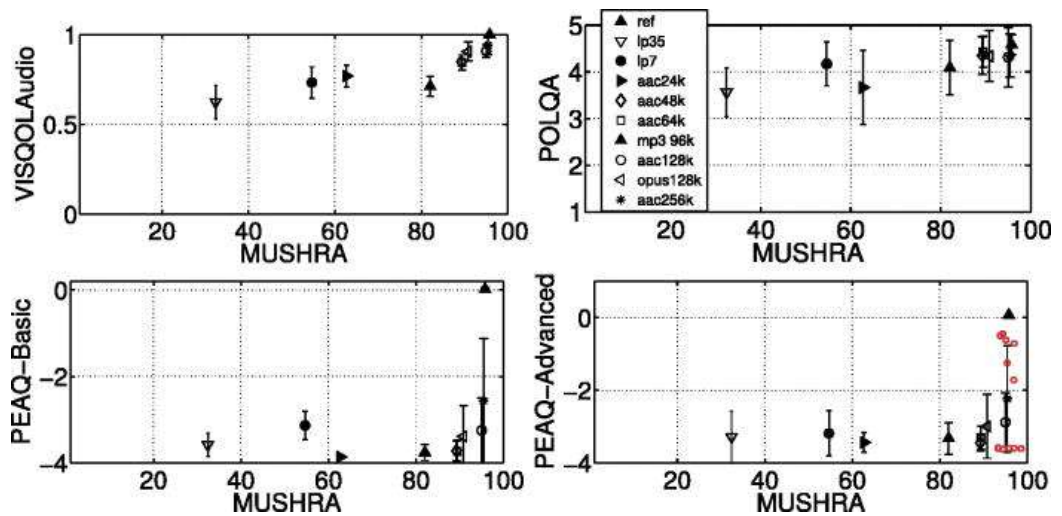


Рисунок 2.2 — Результати середньої об'єктивної метрики.

Нанесені на графік проти суб'єктивних показників якості MUSHRA на прослуховування для 12 зразків. Смуги похибок – це 1 стандартне відхилення. Кружечки (незаповнені) на діаграмі PEAQ-Advanced — результати окремих зразків для AAC 256 k.

Точна оцінка якості анкерів низьких частот була проблемою для всіх показників. Хоча суб'єктивні тести показують, що слухачі дуже погано оцінюють обробку низьких частот навіть порівняно з кодеком AAC-HE 24 кбіт/с із дуже низькою швидкістю передавання даних, це не враховується в прогнозах об'єктивної метрики. Таблиця 2.3 містить коефіцієнти кореляції для всіх показників із середніми суб'єктивними оцінками, згрупованими за прослуховуванням. Рисунок 2.3 дає вказівку на продуктивність, а також

підкреслює погану оцінку прив'язок за всіма перевіреними об'єктивними показниками. Результати кореляції Спірмена та Кендалла були дуже багатообіцяючими для VISQOLAudio, вказуючи на сильну здатність правильно ранжувати прослуховування. Однак, перш ніж надмірно інтерпретувати значення цих статистичних даних, слід взяти до уваги тісне групування ряду методів прослуховування та довірчі інтервали суб'єктивної основної правдивості балів MUSHRA. POLQA і ViSQOLAudio одного порядку за своєю продуктивністю. На рисунку 2.3 представлений результат для балів вище 60 за шкалою MUSHRA для ViSQOLAudio та POLQA, тобто за винятком двох методів прив'язки.

POLQA добре підійшов до всіх обробок, за винятком прив'язок, тоді як VISQOLAudio переоцінює погіршення якості для обробки MP3 96 кбіт/с. Надійність метрик до типу вмісту вибірки була кращою з VISQOLAudio, ніж з POLQA або PEAQ, що підкреслюється розміром смуг помилок.

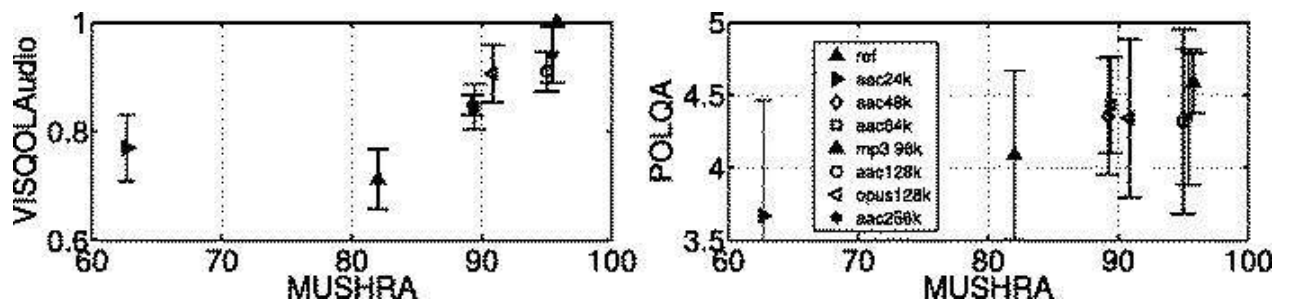


Рисунок 2.3 — Збільшене зображення для ViSQOLAudio та POLQA.

Якірні обробки виключені, а діапазон осей зменшено.

Таблиця 2.3 – Кореляційна статистика для необроблених і регресійно підігнаних середніх оцінок лікування.

Сирий				Припасований			
Виміряти	Пірсо н	Списо- носець	Кендалл	Пірсон	Списо- носець	Кендалл	RMSE
POLQA	0,859	0,77	0,77	0,86	0,77	0,6	14.06
VISQOLA udio	0,866	0,952	0,867	0,895	0,952	0,867	9,94
PEAQ- Advanced	0,673	0,556	0,397	0,4	0,673	0,556	18.93
PEAQ- Базовий	0,337	0,564	0,511	0,293	0,273	0,156	20.08
(Без обробок прив'язки 1 і прив'язки 2)							
POLQA	0,943	0,619	0,619	0,948	0,619	0,5	6,724
VISQOLA udio	0,74	0,952	0,857	0,8	0,929	0,786	6,262
PEAQ- Advanced	0,929	0,857	0,474	0,422	0,833	0,714	9,452
PEAQ- Базовий	0,479	0,976	0,929	0,718	0,976	0,929	7,892

Найвищий рейтинг для кожної статистики виділено жирним шрифтом.

У таблиці 2.3 також представлені статистичні дані за винятком двох опорних точок. ViSQOLAudio є найбільш послідовним в обох сценаріях, хоча показники кореляції впали без обробки прив'язки.

RMSE знизився для всіх заходів після видалення анкерних обробок. Без них кореляційна статистика для PEAQ помітно покращилася, але високий

RMSE все ще підкреслює недостатню надійність між зразками аудіо для даного лікування.

Висновки

Результати для всіх показників є позитивними, незважаючи на те, що PEAQ не розроблявся з урахуванням кодеків з низьким бітрейтом, а POLQA та ViSQOL були задумані як показники якості мови. Адаптація для створення ViSQOLAudio була важливою, оскільки мовна версія не мала надширокосмугового режиму та була оптимізована для оцінки мовлення за допомогою детектора голосової активності. Низьке стандартне відхилення, між зразками для лікування є найбільш вигідною характеристикою при порівнянні результатів з іншими об'єктивними показниками.

POLQA добре працював у режимі мовлення SWB. Статистичні дані на рівні лікування показали потенціал POLQA для цього застосування, але здатність до ранжирування та варіації рівня вибірки вказували на труднощі в розрізненні менших відмінностей у лікуванні. Цю проблему цілком можна вирішити за допомогою ще не опублікованої аудіомодифікованої версії POLQA, де, ймовірно, буде розглянуто зміни вирівнювання, подібні до тих, що застосовуються до ViSQOLAudio. Результати показують, що потрібні лише невеликі адаптації.

PEAQ показав значні варіації в різних зразках для того самого лікування, де слухач оцінив якість, яка була незмінною. У цій рооті, наголошується на необхідності постійного розвитку об'єктивних показників, які можуть мати справу з різними новими кодеками з низьким бітрейтом, які були розроблені. Подальша робота тестування з більш широким діапазоном даних триває. ViSQOLAudio має майбутній потенціал, якості звукуметрики та вигідно порівнює з PEAQ для оцінених кодеків із нижчим бітрейтом.

2.3 Інтеграції ViSQOL3 у різні проекти та їх практичне використання

Ця версія ViSQOL є результатом процесу інтеграції ViSQOL із використанням реальних випадків виробництва та тестування інтеграції в Google. Тематичні дослідження, описані в цьому розділі, були ініційовані окремими командами, які не залежали від попередньої розробки ViSQOL. Зазвичай вони консультувалися з розробником ViSQOL, щоб перевірити правильність використання, або читали документацію та інтегрували ViSQOL самостійно.

2.3.1 *Hangouts Meet*

Команда Meet успішно використовує ViSQOL для оцінки якості звуку в Hangouts Meet. Hangouts Meet – це служба відеозв'язку, яка використовує WebRTC [20] для передачі звуку. Meet використовує тестовий стенд, здатний надійно відтворювати несприятливі умови мережі для оцінки, якості звуку під час виклику. Для цього випадку використання, вони мають еталонні та погіршені зразки аудіо з частотою дискретизації 48 кГц і використовують ViSQOLAudio для обчислення результатів.

Щоб переконатися, що ViSQOL працює надійно для цього випадку використання, його порівняли з внутрішньою метрикою якості аудіо без посилення, яка базується на технічних показниках приймача на основі WebRTC. Показник оцінюється за шкалою від 0 до 1, де нижчі оцінки є кращими. MOS ViSQOL може корелювати з цим показником, як показано на рисунку 2.4

У цьому випадку використання, розробників Meet здебільшого цікавила чутливість ViSQOL до погіршення звуку, внаслідок пошкоджень мережі. На рисунку 2.4 наведено порівняння середніх показників ViSQOL під час виклику, яке показує, що показник чутливий до зміни якості звуку від сценарію хороших умов мережі з оцінками від 4,21 до 4,28 до сценарій із середніми порушеннями з оцінками від 4,04 до 4,16 і, нарешті, надзвичайно складний сценарій мережі з оцінками від 3,72 до 3,94. Хоча точні умови мережі

не можна повідомити, тут хороші умови мережі вказують на те, що з'єднання має забезпечувати майже ідеальні відео та аудіо під час дзвінка, середні умови вказують на те, що дзвінок може мати проблеми, але звук має продовжуватись

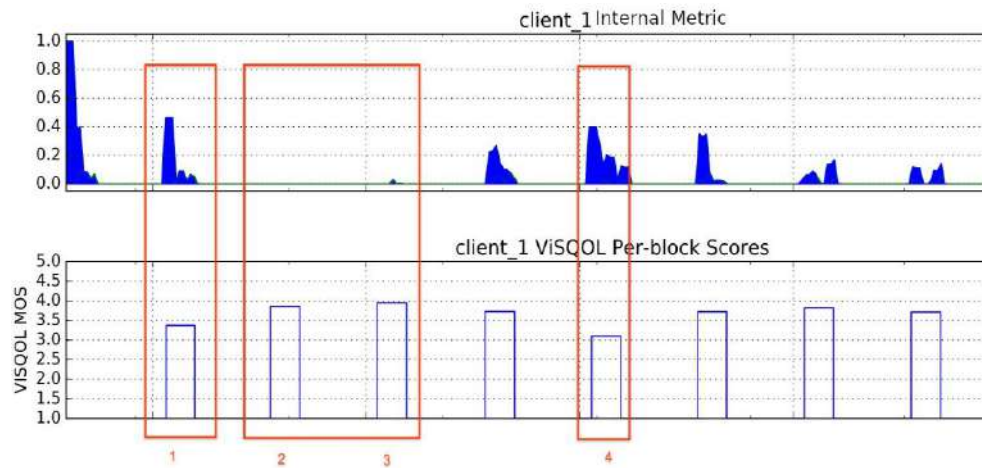


Рисунок 2.4 — Внутрішня метрика без посилань Hangouts Meet.

ViSQOL успішно виявив ці погіршення в аудіо, яке їх містило (блоки 1 і 4), тоді як в аудіоблоках, на які не вплинуло, оцінки ViSQOL були вищими (блоки 2 і 3), тоді як у надзвичайно складних умовах, очікується погіршення сприйняття як відео, так і аудіо, але виклик все одно пройде.

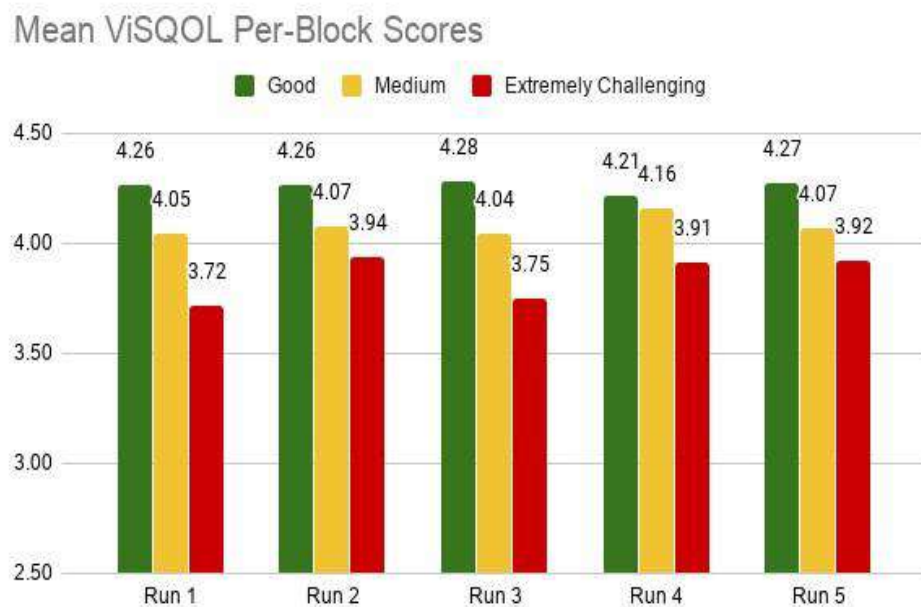


Рисунок 2.5 — Порівняння між середнім ViSQOL MOS і погіршенням мережі.

Деякі дзвінки проводилися за хороших умов мережі (зелений), деякі імітували середні умови мережі, де продукт все ще повинен працювати добре (жовтий), тоді як інші моделювали надзвичайно складні умови мережі, де очікувалося виникнення проблем (червоний).

Щоб переконатися, що ViSQOL працює надійно, було зібрано кілька сотень дзвінків із тестового стенда. Середні значення, отримані від ViSQOL, і внутрішня метрика цих викликів були нанесені на рисунок 2.5 Результати були достовірно відтворені. Після позитивних результатів цього дослідження, ViSQOL наразі є одним із основних об'єктивних показників якості аудіо, який застосовує команда продукту Hangouts Meet у Google.

2.3.2 Кодек Opus

Google бере участь у розробці кодека Opus. ViSQOL і POLQA використовувалися для порівняння якості кодера Opus як для мови, так і для музики на різних бітрейтах і обчислювальних складностях. У попередніх дослідженнях було показано, що ViSQOLAudio добре працює на аудіо з низьким бітрейтом [8]. Однак мовний режим ViSQOL не націлений на низький бітрейт.

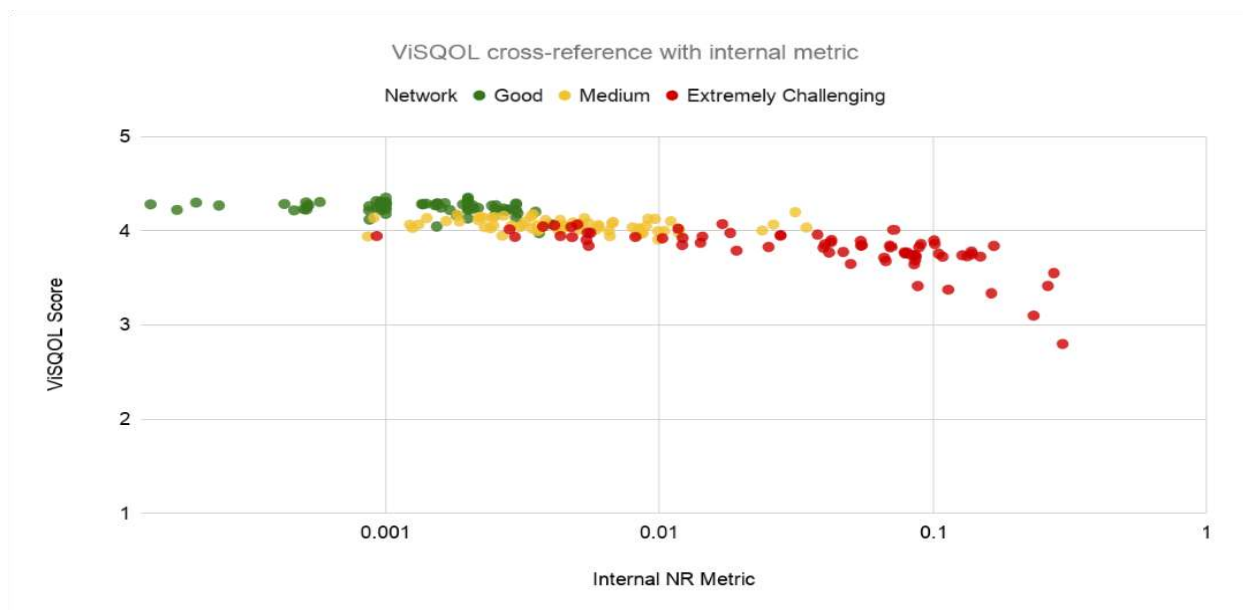


Рисунок 2.6 — Діаграма розсіювання ViSQOL MOS проти внутрішньої метрики без посилання.

Кожна точка означає виклик.

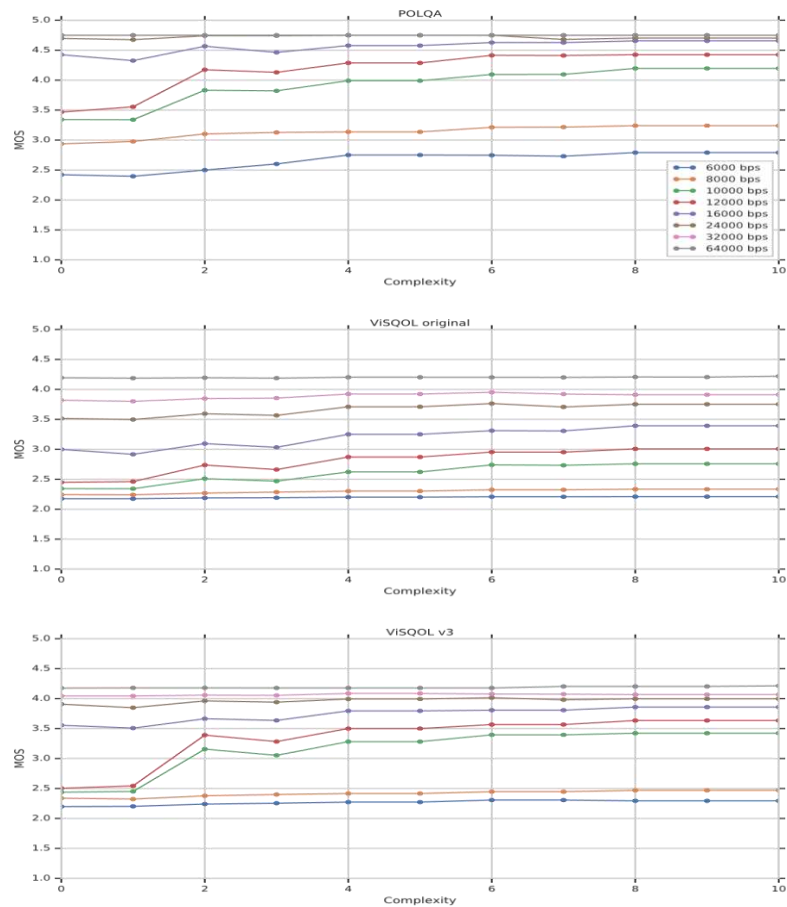


Рисунок 2.7 — Оцінка MOS для різних бітрейтів (6-64 кбіт/с) і складності налаштування мовлення, закодованого Opus.

Крім того, нещодавні досягнення в Opus, ще більше знизили нижню межу діапазону бітрейтів для заданої пропускної здатності, з моменту появи ViSQOL. Наприклад, Opus 1.3 може виробляти широкосмуговий сигнал зі швидкістю 9 Кбіт/с, тоді як набори даних TCDAudio14 [21], CoreSV14 [22] і AACvOpus15 [19], на яких навчалася векторна регресія підтримки ViSQOL, мають бітрейт лише до 24 Кбіт/с. .

POLQA та оригінальна версія ViSQOL у мовному режимі відображають подібні тенденції, які відповідають очікуванням щодо параметрів бітрейту та складності. Різниця в нижчих бітрейтах є більш вираженою, відповідно до POLQA. Відмінності у вищих бітрейтах є більш вираженими, відповідно до ViSQOL. Хоча суб'єктивні оцінки були недоступні, розробники очікували, що

MOS буде менш чутливим до змін вищих бітрейтів, забезпечуючи POLQA кращу відповідність.

Для музичних прикладів розробники виявили, що обидва показники відображають схожі тенденції щодо бітрейтів. Але POLQA показує більшу дискримінацію між 6-8 кбіт/с, 1012 кбіт/с і 16-24 кбіт/с. ViSQOL здатний розрізняти різні бітрейти з монотонною поведінкою, але одним із моментів, що викликає занепокоєння, є те, що це призводить до того, що ViSQOLAudio є відносно нечутливим до відмінностей у налаштуваннях складності. У світлі цього не рекомендується використовувати ViSQOL для автоматизованих регресійних тестів без повторного навчання моделі. Покращення, зроблені в розділі 3, дещо вирішують ці проблеми, як можна побачити на рисунку 2.7 З іншого боку, ViSQOL визначив хибну пропускну здатність на 12 Кбіт/с для параметрів складності 5 і 6 (що сприймалося як вища якість у неформальне прослуховування), де POLQA цього не робив.

Нарешті, ViSQOL використовувався для аналізу результатів як для чистих, так і для зашумлених посилань. ViSQOL не був розроблений для цього випадку, оскільки він передбачає чисте посилання, подібне до PESQ [2] і POLQA [4]. Однак було виявлено, що він працює подібно до чистих випадків як для мовлення, так і для звуку в шумних випадках.

Було зроблено висновок, що ViSQOL можна використовувати для регресії тестування на мовлення. Проте офіційні тести на прослуховування були б бажаними з двох причин: для кращої інтерпретації відмінностей між POLQA та ViSQOLAudio та для навчання моделі, яка представляє діапазони низьких бітрейтів.

Висновки

Багато різних, інших команд також адаптували ViSQOL для своїх продуктів. У більшості випадків їх варіант використання, дещо нагадує навчальні дані (погіршення якості ширококутної мовної мережі або кодування музики), але часто має помітні відмінності. Наприклад, одна команда вирішила проаналізувати мережевий цикл за допомогою цифрового

та аналогового інтерфейсів, вимагаючи створення обладнання для безперервного автоматизованого тестування. Зазвичай ці групи також мали доступ до PESQ, POLQA або суб'єктивних балів для своїх випадків і хотіли оцінити точність вимірювань ViSQOL, а також визначити обмеження. Часта проблема була пов'язана з тривалістю та сегментацією аудіо, яке використовуватиметься з ViSQOL за використання в автоматизованій системі. Хоча ViSQOL у мовному режимі має детектор голосової активності, який виявив, що ViSQOLAudio працюватиме погано для сегментів, де еталонний сигнал був мовчазним, або через ефект усереднення, або через відсутність логарифмічного порогового значення, яке було надто чутливим до невеликих абсолютних різниць у рівнях навколишнього шуму. Щоб усунути ефекти усереднення, було рекомендовано витягти сегменти аудіо від 3 до 10 секунд, де була відома активність. У наступному розділі розглядається вирішення проблем із пороговими значеннями.

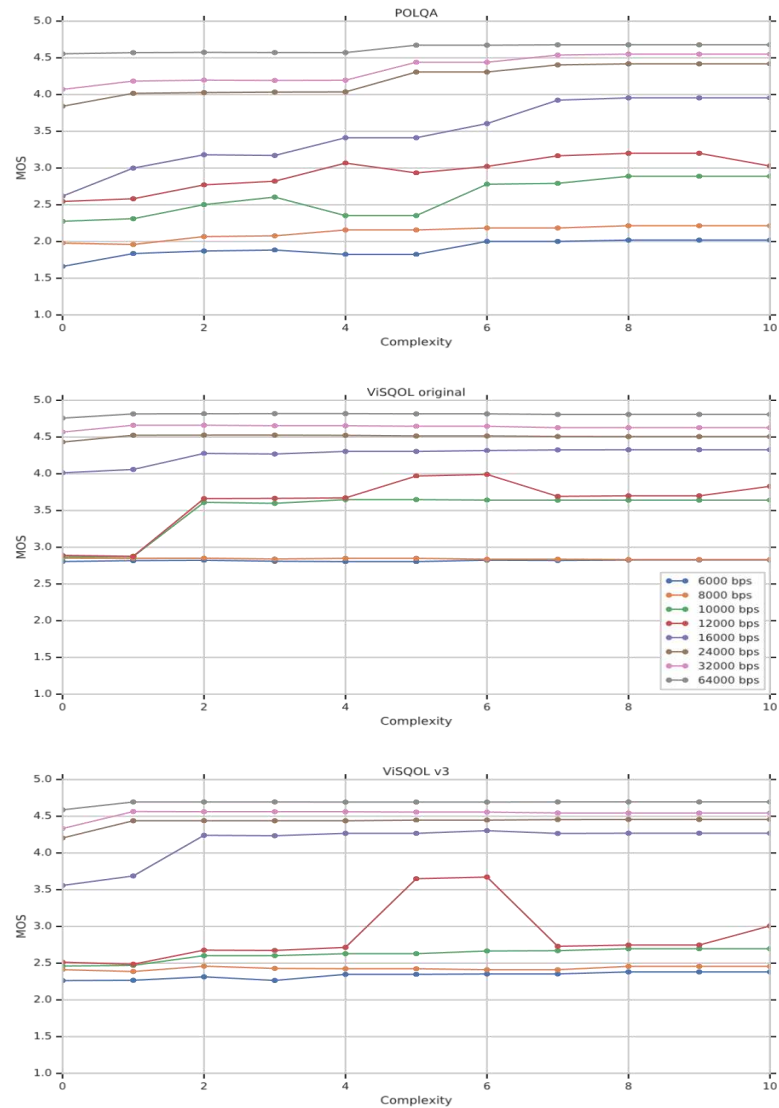


Рисунок 2.8 — Розрахункова MOS для різних бітрейтів.

Бітрейт відповідає тому самому ключу, що й на Рисунок 2.8 Підйом складності 5, 6 і 10 для 12 кбіт/с пов'язаний з тим, що Орус вирішив використовувати смугу пропускання 12 кГц для деякої частини файлів замість смуги пропускання 8 кГц, яку він використовував для складності. 2-4 і 7-9.

3 ВИКОРИСТАННЯ ТА ПОКРАЩЕННЯ

3.1 Зміни та покращення

У цьому розділі коротко викладено попередню версію та описано зміни, внесені до нової версії. На рисунку 3.1 показано загальний потік програми та виділено нові компоненти системи, про які йдеться в підрозділах.

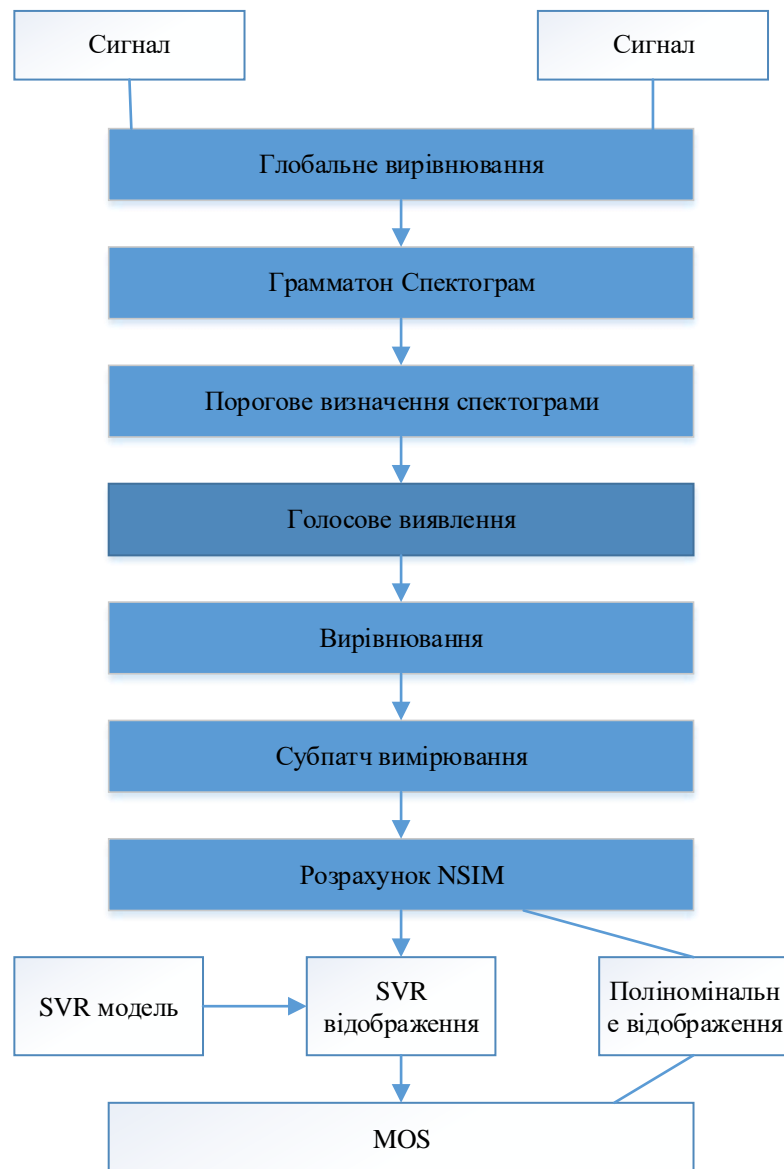


Рисунок 3.1 — Схема системи. Входи та виходи мають білу заливку, а компоненти обробки – синю

Нові компоненти мають товсті краї. Заливка представляє лише мовні та звукові компоненти відповідно.

Алгоритми ViSQOL, описані в [7] і [8], мають багато спільних компонентів, таких як гамматонна спектрограма та обчислення NSIM. Тоді видається доцільним, що загальні компоненти спільно використовуються та розробляються разом. Відмінності між двома алгоритмами пов'язані з відмінностями в характеристиках мови та музики. Наприклад, використання визначення голосової активності (VAD) для мовлення та аналіз вищих діапазонів (до 24 кГц) для загального аудіо/музики. Загальні компоненти як мовної, так і аудіосистеми включають створення гамматонної спектрограми з використанням фільтрів еквівалентної прямокутної смуги пропускання (ERB), створення патчів порядку півсекунди, їх вирівнювання, обчислення NSIM з вирівняних патчів, а потім відображення значення NSIM до MOS.

До деяких із цих компонентів було внесено незначні зміни з практичних міркувань, як от зміна залежностей або виправлення проблем, виявлених у прикладах чи невдалих тестах. Реалізація VAD використовує простий VAD на основі енергії, якого має бути достатньо, враховуючи вимогу чистих посилянь. Як інший приклад, розмір вікна було оновлено до 80 мс зі стрибком у 20 мс після виявлення проблеми з вікнами попередніх версій.

3.2 Бібліотека та двійковий файл C++

Щоб зробити ViSQOL більш доступним, потрібно роз'єднати залежність від MATLAB, реалізувавши версію C++ із залежностями лише з відкритим кодом. Нова версія, v3, доступна як двійковий файл або бібліотека. Кодову базу було опубліковано на GitHub, тому що, щоб нею було легко користуватися для громадськості та запрошувати зовнішні внески.

Більшість користувачів були бінарними користувачами, але деякі мали вимоги до більш точного контролю. Для цього треба розробити бібліотеку з підтримкою protobuf і перевіркою помилок, від яких залежить двійковий файл.

Ця бібліотека також буде корисною для користувача, який бажає обгорнути функції іншою мовою, наприклад, за допомогою прив'язок python.

Було кілька змін у вхідних і вихідних даних. Детальний вихід також змінився, щоб включити середні значення NSIM на смугу частот і середнє значення NSIM на кадр. Оскільки ViSQOL постійно змінюється, щоб адаптуватися до нових проблем, номер версії відповідності включений у вихідні дані. Щоразу, коли MOS змінюється для відомих файлів, номер відповідності буде збільшуватися. Нарешті, також підтримується пакетна обробка за допомогою файлів зі значеннями, розділеними комами (csv).

Для створення цієї версії було використано кілька проектів, пов'язаних із Google. Двійковий файл програми було реалізовано за допомогою Framework додатків Abseil C++ [23]. Інтегровано структуру тестування Google Test C++ [24] і реалізовано різні тести для забезпечення коректності, виявлення регресій і підвищення стабільності для граничних випадків. Реалізовано 23 тестових заняття з кількома тестами. До них входять не лише модульні тести, а й тест для перевірки відповідності поточної версії відомим балам. Фреймворк Bazel [25] використовувався для побудови та вибірки залежностей, а також для розробки тестів.

3.3 Тонкомасштабне вирівнювання часу

Незважаючи на те, що попередні версії ViSQOL виконували два рівні вирівнювання (глобальний і патч), усе ще виникали проблеми з вирівнюванням патчів через зміщення кадрів спектрограм у дрібному масштабі. Щоб вирішити цю проблему, треба запровадити додатковий крок вирівнювання, який компенсує затримку, знайдену в кроці крос-кореляції на областях часової області, що відповідає вирівняним ділянкам, як описано в [26]. Далі гамматонна спектрограма перераховується для вирівняного аудіо патча за зразком і береться оцінка NSIM.

3.4 Пороги тиші

Щоб вирішити проблему логарифмічних амплітуд, розглянутих раніше, потрібно вводити порогові значення мовчання на гамматонній спектрограмі. Оскільки NSIM обчислюється на основі логарифмічних амплітуд, виявилось, що він надто чутливий до різних рівнів навколишнього шуму. Наприклад, майже цифрове посилення на тишу порівняно з дуже низьким рівнем навколишнього шуму все одно матиме дуже низьку оцінку NSIM, незважаючи на те, що воно прозоре для сприйняття. Поріг мовчання вводить як абсолютний, так і відносний рівень, який може бути вищим для кадрів з високою амплітудою.

Порогова амплітуда $u_{t,f}(x)$ для часу t і смуги частот f на вхідній спектрограмі x залежить від:

$$u_{t,f}(x) = \max(U_{\min}, U_{f_{\min}}(t), x(t,f)) \quad (1)$$

де:

$$U_{f_{\min}}(t) = \max(r_f(t), d_f(t)) - U_{f_{\min}} \quad (2)$$

задані опорні та деградовані логарифмічні амплітуди $r_{t,f}$ і $d_{t,f}$, а також глобальний абсолютний поріг U_{\min} і відносний поріг $U_{f_{\min}}$ для кадру.

3.5 Модель NSIM до MOS

Зміни вище в кінцевому підсумку впливають на бали NSIM. Це вимагає, щоб нова модель SVR була навчена відображати діапазон частот NSIM на MOS за допомогою libsvm [27]. Провів пошук у сітці, щоб мінімізувати втрати 4-сторонньої перехресної перевірки, на тому самому навчальному наборі (TCDAudio14, CoreSV14, AACvOpus15). Однак було помічено, що ця модель була надто специфічною для навчальних даних і погано працювала на аудіо з дуже низькою швидкістю потоку (6-18 кбіт/с). Здається, це пов'язано з тим, що у моделі SVR, що використовується ViSQOL, немає обмежень щодо монотонності (строго вищий NSIM для даних поза розподілом, які дають нижчий MOS). Щоб вирішити цю проблему для моделі за замовчуванням, треба пом'якшити параметри SVR, знизивши вартість і гамма-параметри, щоб

мати дещо вищу помилку перехресної перевірки, забезпечуючи при цьому поведінку, наближену до монотонної.

Крім того, ця версія містить деякі інструменти та документацію, які дозволяють користувачам тренувати власну модель SVR за допомогою вхідних файлів CSV, якщо користувач може надати суб'єктивні оцінки для пар погіршених/еталонних. Дотримуючись методів пошуку в сітці, описаних `libsvm`, користувачі повинні мати можливість адаптувати модель, яка здатна представляти їхні дані.

4 СТАРТАП ПРОЕКТ

Розділ має на меті проведення маркетингового аналізу стартап проекту задля визначення принципової можливості його ринкового впровадження та можливих напрямів реалізації цього впровадження. Проведення маркетингового аналізу передбачає виконання нижченаведених кроків.

Етапи маркетингового аналізу стартап-проекту:

4.1 Опис ідеї проекту (товару, послуги, технології)

В межах підпункту слід послідовно проаналізувати та подати у вигляді таблиць:

- зміст ідеї (що пропонується);
- можливі напрямки застосування;
- основні вигоди, що може отримати користувач товару (за кожним напрямком застосування);
- чим відрізняється від існуючих аналогів та замінників;

1.1) Перші три пункти подаються у вигляді таблиці 4.1 і дають цілісне уявлення про зміст ідеї та можливі базові потенційні ринки, в межах яких потрібно шукати групи потенційних клієнтів.

Таблиця 4.1 Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Виготовлення пристроїв діагностики вушної раковини людини «IT Hearing ®»	1.Цивільне застосування у медицині	Діагностика у державних медичних закладах
	2.Частна діагностика слуху	Використання в приватних клініках
	...	

Як видно з таблиці, «IT Hearing ®», можна використовувати у вузькому направленні галузей.

1.2) Аналіз потенційних техніко-економічних переваг ідеї (чим відрізняється від існуючих аналогів та замінників) порівняно із пропозиціями конкурентів передбачає:

- визначення переліку техніко-економічних властивостей та характеристик ідеї (орієнтований можливий перелік властивостей та характеристик подано у додатку А);
- визначення попереднього кола конкурентів (проектів-конкурентів) або товарів-замінників чи товарів-аналогів, що вже існують на ринку, та проводиться збір інформації щодо значень техніко-економічних показників для ідеї власного проекту та проектів-конкурентів відповідно до визначеного вище переліку;
- проводиться порівняльний аналіз показників: для власної ідеї визначаються показники, що мають а) гірші значення (W, слабкі); б) аналогічні (N, нейтральні) значення; в) кращі значення (S, сильні) (табл. 4.2).

Таблиця 4.2 Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п/п	Техніко-економічні характеристики ідеї	(потенційні) товари/концепції конкурентів		W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проект	Конкурент 1			
1.	Продаж пристрою для державних закладів	Пристрій, гарантія, тех. Підтримка та обслуговування	«Aurora®»	Корупційні схеми держави.	Також є можливість подальшого розвитку технології	Робиться на замовлення держави, через державні закупівлі.
2.	Продаж пристрою для приватних закладів.	Пристрій, гарантія, тех. Підтримка та обслуговування	«Aurora®»	Розповсюджена консервативність цільової аудиторії	З часом, подолавши консервативність клієнтів продажі будуть рости.	Короткі терміни виконання замовлення. Великі обсяги.

Визначений перелік слабких, сильних та нейтральних характеристик та властивостей ідеї «IT Hearing ®» є підґрунтям для формування його конкурентоспроможності.

4.2 Технологічний аудит ідеї проекту

В межах даного підрозділу необхідно провести аудит технології, за допомогою якої можна реалізувати ідею проекту (технології створення товару).

Визначення технологічної здійсненності ідеї проекту передбачає аналіз таких складових (табл. 4.3):

- за якою технологією буде виготовлено товар згідно ідеї проекту?
- чи існують такі технології, чи їх потрібно розробити/доробити?
- чи доступні такі технології авторам проекту?

Таблиця 4.3 – Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
	Виготовлення пристроїв діагностики вушної раковини людини «IT Hearing ®» для державних закладів	Програмна складова і апаратна складова. Подальша реалізація продукції за допомогою державних закупівель Контракти по ліцензування технології «IT Hearing ®»	Програмна складова в наявності. Можливість апаратної складової проекту відсутня.	Програмне забезпечення реалізоване, апаратне буде розроблене в подальшому
	Виготовлення пристроїв діагностики вушної раковини людини «IT Hearing ®» для приватних закладів.	Програмна складова і апаратна складова. Подальша реалізація продукції за допомогою контрактів з приватними компаніями або продаж пристроїв фізичним особам.	Програмна складова в наявності. Можливість апаратної складової проекту відсутня.	Програмне забезпечення реалізоване, апаратне буде розроблене в подальшому

Всі технології для реалізації проекту існують і цілком доступні, що, забезпечить цілковито безпроблемне виготовлення та розробку нашої ідеї.

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Визначення ринкових можливостей, які можна використати під час ринкового впровадження проекту, та ринкових загроз, які можуть перешкодити реалізації проекту, дозволяє спланувати напрями розвитку проекту із урахуванням стану ринкового середовища, потреб потенційних клієнтів та пропозицій проектів-конкурентів.

3.1) Спочатку проводиться аналіз попиту: наявність попиту, обсяг, динаміка розвитку ринку (табл. 4.4).

Таблиця 4.4 Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	1
2	Загальний обсяг продаж, грн/ум.од	100000 \$ (або екв)
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Консервативність покупців
5	Специфічні вимоги до стандартизації та сертифікації	ROHS-compliant, ISO/IEC 17021
6	Середня норма рентабельності в галузі (або по ринку), %	70%

Середня норма рентабельності в галузі (або по ринку), 70%, що є пристойною, та задовольняє вимоги проекту.

3.2) Надалі визначаються потенційні групи клієнтів, їх характеристики, та формується орієнтовний перелік вимог до товару для кожної групи (табл. 4.5).

Таблиця 4.5 Характеристика потенційних клієнтів стартап проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
	Вчасна діагностика, та попередження втрати слуху.	Люди з вадами слуху	Однакові вимогу, надійність і точність, оскільки йдеться про медичне обладнання.	До продукції: Якість, надійність, дуже висока точність, можливість роботи при будь-яких погодних умовах та в умовах бойових дій. До компанії постачальника: Прозорість, молода команда вітчизняних розробників, патріотичність команди.

Розраховувати на державні підряди досить успішна практика у нашій країні, але приватний ринок, також є перспективним.

3.3) Після визначення потенційних груп клієнтів проводиться аналіз ринкового середовища: складаються таблиці факторів, що сприяють ринковому впровадженню проекту, та факторів, що йому перешкоджають. Фактори в таблиці подавати в порядку зменшення значущості.

Таблиця 4.6 Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Загроза інтелектуальній власності (копіювання, шпигунство)	Втрата оригінальності ідеї	Судові тяжби до компаній-крадіїв. Паралельно з цим буде робота по модернізації

Ці слабкі й малозначущі загрози не здатні зупинити реалізацію грандіозного проекту.

Таблиця 4.7 – Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Розвиток та тестування ідеї	Можливість подальшого створення нового продукту, використовуючи досвід, отриманий при реалізації попередніх проектів.	Розширення підприємства та збільшення обсягу інтелектуальної власності, масштабування.
2	Плідна співпраця	Заклучення контрактів з державними та приватними мед. закладами	Допомога державній медицині, та заробіток на приватних мед закладах

Можливості даного проекту є досить перспективними та амбітними.

3.4) Надалі проводиться аналіз пропозиції: визначаються загальні риси конкуренції на ринку.

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції монополія	Монополія кардинально відрізняється від досконалого виду конкуренції. У цій системі лише один виробник випускає товари та послуги, на які спостерігається високий попит серед цільової аудиторії. Якісних альтернатив немає, тож компанія-монополіст фактично захоплює ринок. Водночас не завжди споживачі готові будуть заплатити вибрану монополістом суму.	Заздалегідь відоме конкурентне середовище
2. За рівнем конкурентної боротьби Національний, міжнародний	До державних закупівель допускається лише (або пріоритетно) українські компанії.	Середнє коло конкурентів.
3. За галузевою ознакою - внутрішньогалузева	Унікальний продукт з унікальними функціями.	Так як цей продукт є унікальним, передбачаються дуже великі продажі. Відкриває можливість потужних продажів.
4. Конкуренція за видами товарів: - товарно-видова	Конкуренти, які продають товар під конкретну мету.	Наша компанія єдиний постачальник такого роду продукції, тому проблем виникнути не повинно.
5. За характером конкурентних переваг - нецінова	Покращення якісних характеристик продуктів і проведенні активної комунікаційної політики.	Також є ряд критеріїв для покращення: Швидкість обслуговування; Термін доставки продукту; Якість і характер консультацій; Характер особистих відносин; Час обслуговування; Помилки і упущення.

Продовження Таблиці 4.8

6. За інтенсивністю - Марочна, параметрична, товарна (при наявності конкурентів)	Конкуренція торгових марок виробників однорідної продукції (при наявності) Конкуренція серед виробників аналогічної продукції (при наявності) Конкуренція за зовнішнім виглядом і функціями (при наявності)	При виникненні проблем зі збутом товару або навпаки при великому рості продажів ми будемо регулювати ціну. Надання знижок за заявленими тарифами; Залучення більш дешевих фінансових ресурсів; Надання більшого обсягу послуг за діючими тарифами;
---	--	---

3.5) Після аналізу конкуренції проводиться більш детальний аналіз умов конкуренції в галузі.

Таблиця 4.9 Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
Складові аналізу	Прямі конкуренти в галузі відсутні.	Конкуренти, які вкрадуть або зплагіатять.	Всі Постачальники зарубіжні, й наші	Мед. заклади	Заміни відсутні
Висновки:	Інтенсивність відсутня	Є можливість входу на ринок з можливістю порушення закону про інтелектуальну власність ККУ. Строки виходу: від одного року	В даний час можливі проблеми з постачальниками через кризу напівпровідників	Не диктують	Обмеження для роботи на ринку відсутні.

За результатами аналізу таблиці робиться висновок щодо принципової можливості роботи на ринку з огляду на конкурентну ситуацію. Також робиться висновок щодо характеристик (сильних сторін), які повинен мати

проект, щоб бути конкурентоспроможним на ринку. Другий висновок враховується при формулюванні переліку факторів конкурентоспроможності.

3.6) На основі аналізу конкуренції, проведеного, а також із урахуванням характеристик ідеї проекту, вимог споживачів до товару та факторів маркетингового середовища визначається та обґрунтовується перелік факторів конкурентоспроможності. Аналіз оформлюється за табл. 4.10

Таблиця 4.10 – Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1	Унікальність продукту	Відсутність аналогів на ринку.
2	Актуальність	По статистиці в світі близько 466 млн. людей з вадами слуху
3	Новизка	Нові інновації
4	Доступність	Доступна ціна

3.7) За визначеними факторами конкурентоспроможності проводиться аналіз сильних та слабких сторін стартап-проекту (табл. 4.11).

Таблиця 4.11 Порівняльний аналіз сильних та слабких сторін «назва проекту»

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з ... (назва підприємства)						
			-3	-2	-1	0	+1	+2	+3
1	Унікальність продукту	20	-2	-1	-1	0	+1	+2	+3
2	Актуальність	20	-2	-1	-1	0	+1	+2	+3
3	Новизка	20	-2	-1	-1	0	+1	+2	+3
4	Доступність	20	-2	-1	-1	0	+1	+2	+3

3.8) Фінальним етапом ринкового аналізу можливостей впровадження проекту є складання SWOT-аналізу (матриці аналізу сильних (Strength) та слабких (Weak) сторін, загроз (Troubles) та можливостей (Opportunities) на основі виділених ринкових загроз та можливостей, та сильних і слабких сторін (табл. 4.11).

Перелік ринкових загроз та ринкових можливостей складається на основі аналізу факторів загроз та факторів можливостей маркетингового середовища.

Ринкові загрози та ринкові можливості є наслідками (прогнозованими результатами) впливу факторів, і, на відміну від них, ще не є реалізованими на ринку та мають певну ймовірність здійснення. Наприклад: зниження доходів потенційних споживачів – фактор загрози, на основі якого можна зробити прогноз щодо посилення значущості цінового фактору при виборі товару та відповідно, – цінової конкуренції (а це вже – ринкова загроза).

Таблиця 4.12 SWOT- аналіз стартап-проекту

<p>Сильні сторони: Можливість використання широкого спектру інвестицій. Мала кількість конкурентів у сегменті ринку з легким входом. Варіативність реалізації ідеї</p>	<p>Слабкі сторони: Розповсюджена консервативність цільової аудиторії.</p>
<p>Можливості: Вірогідність швидкого розвитку при успішному старті. Можливість подальшого створення нового продукту, використовуючи досвід, отриманий при реалізації попередніх проектів.</p>	<p>Загрози: Втрата оригінальності ідеї. Втрата людських ресурсів з можливим витоком конфіденційної інформації разом з ними</p>

3.9) На основі SWOT-аналізу розробляються альтернативи ринкової поведінки (перелік заходів) для виведення стартап-проекту на ринок та орієнтовний оптимальний час їх ринкової реалізації з огляду на потенційні проекти конкурентів, що можуть бути виведені на ринок. Визначені альтернативи аналізуються з точки зору строків та ймовірності отримання ресурсів (табл. 4.13).

Таблиця 4.13. Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Відкритий краудфандинг	70%	1 рік

З означених альтернатив обирається та, для якої: а) отримання ресурсів є більш простим та ймовірним; б) строки реалізації – більш стислими.

4.4 Розроблення ринкової стратегії проекту

4.1) Розроблення ринкової стратегії першим кроком передбачає визначення стратегії охоплення ринку: опис цільових груп потенційних споживачів (табл. 4.14).

Таблиця 4.14. Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Державні заклади	Середня готовність	Середній-високий попит	Низька-відсутня конкуренція	Середня простота входу
2	Приватні заклади	Середня готовність	Середній-високий попит	Низька-відсутня конкуренція	Середня простота входу

Таблиця відображає високу вірогідність безпроблемного завоювання споживачів.

За результатами аналізу потенційних груп споживачів (сегментів) автори ідеї обирають цільові групи, для яких вони пропонуватимуть свій товар, та визначають стратегію охоплення ринку:

- якщо компанія зосереджується на одному сегменті – вона обирає стратегію концентрованого маркетингу;

- якщо працює із кількома сегментами, розробляючи для них окремо програми ринкового впливу – вона використовує стратегію диференційованого маркетингу;
- якщо компанія працює із всім ринком, пропонуючи стандартизовану програму (включно із характеристиками товару/послуги) – вона використовує масовий маркетинг.

4.2) Для роботи в обраних сегментах ринку необхідно сформуванати базову стратегію розвитку (табл. 4.15).

Таблиця 4.15. Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку*
1	-	Стратегія концентрованого маркетингу	-	-

* – опис базових стратегій розвитку див. у дод. Б.

4.3) Наступним кроком є вибір стратегії конкурентної поведінки (табл. 4.16).

Таблиця 4.16 Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
1	Так	Ні	Ні	Стратегія лідера

* – опис стратегій конкурентної поведінки див. у дод. В.

4.4) На основі вимог споживачів з обраних сегментів до постачальника (стартап-компанії) та до продукту (див. табл. 4.5), а також в залежності від

обраної базової стратегії розвитку (табл. 4.15) та стратегії конкурентної поведінки (табл. 4.16) розробляється стратегія позиціонування (табл. 4.17), що полягає у формуванні ринкової позиції (комплексу асоціацій), за яким споживачі мають ідентифікувати торгівельну марку/проект.

Таблиця 4.17 Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
1	Висока надійність, можливість виконання продукту з широким спектром поставлених вимог, висока чесність та прозорість компанії, доступність	Стратегія спеціалізації	Унікальність ідеї, низька собівартість, актуальність.	Унікальність, компактність, простота, доступність, надійність,

Передбачається великий успіх стратегії описаної у попередній таблиці, що пов'язано з потужною ідеєю стартапу

4.5 Розроблення маркетингової програми стартап-проекту

5.1) Першим кроком є формування маркетингової концепції товару, який отримає споживач. Для цього у табл. 4.18 потрібно підсумувати результати попереднього аналізу конкурентоспроможності товару.

Таблиця 4.18 Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Діагностика слуху.	Точне діагностування	Унікальність, надійність, вартість, функціонал
2	Покращення підбор слухового апарату.	Можливість покращення свого слуху	Унікальність, надійність, вартість, функціонал

Є загальний попит в цій галузі.

5.2) Надалі розробляється трирівнева маркетингова модель товару: уточнюється ідея продукту та/або послуги, його фізичні складові, особливості процесу його надання (табл. 4.19). Орієнтовний перелік можливих характеристик товару наведено у додатку С.

Таблиця 4.19 Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Потреба нормального функціонування слуху.		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. Програмне забезпечення		
	2. Апаратне забезпечення		
	Якість: ROHS-compliant, ISO/IEC 17021		
	Пакування: Картонна коробка з наповнювачем.		
	Марка: «IT Hearing ®»		
III. Товар із підкріпленням	До продажу «IT Hearing ®»+		
	Після продажу «IT Hearing ®»+		
За рахунок чого потенційний товар буде захищено від копіювання: Патентування інтелектуальної властності.			

Після формування маркетингової моделі товару слід особливо відмітити – чим саме проект буде захищено від копіювання. Захист може бути

організовано за рахунок захисту ідеї товару (захист інтелектуальної власності), або ноу-хау, чи комплексне поєднання властивостей і характеристик, закладене на другому та третьому рівнях товару.

5.3) Наступним кроком є визначення цінових меж, якими необхідно керуватись при встановленні ціни на потенційний товар (остаточне визначення ціни відбувається під час фінансово-економічного аналізу проекту), яке передбачає аналіз ціни на товари-аналоги або товари субститути, а також аналіз рівня доходів цільової групи споживачів (табл. 4.20). Аналіз проводиться експертним методом.

Таблиця 4.20 Визначення меж встановлення ціни

№ п/п	Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1	>10000 грн	>10000 грн	Юридичні особи	5000-10000 грн

5.4) Наступним кроком є визначення оптимальної системи збуту, в межах якого приймається рішення (табл. 4.21):

- проводити збут власними силами або залучати сторонніх посередників (власна або залучена система збуту);
- вибір та обґрунтування оптимальної глибини каналу збуту;
- вибір та обґрунтування виду посередників.

Таблиця 4.21 Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1	Невисока вартість, та великий функціонал, не залишить рівнодушним.	Організація сервісу, впровадження цілеспрямованої товарної політики, організація системи товароруху, впровадження цілеспрямованої збутової політики	Канал збуту нульового рівня	Власна система збуту

Проблем, як можна бачити, зі збутом не передбачається

5.5) Останньою складовою маркетингової програми є розроблення концепції маркетингових комунікацій, що спирається на попередньо обрану основу для позиціонування, визначену специфіку поведінки клієнтів (табл. 4.22).

Таблиця 4.22 Концепція маркетингових комунікацій

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Висока вимогливість, широкий спектр запитів	Прямі канали комунікації	Унікальність, надійність, доступність	Впевнити в надійності товару	Роказати, які переваги дає пристрій.

Дана концепція є основною ланкою взаємовигідних відносин між клієнтом та нашою компанією.

Результатом пункту 5 має стати ринкова (маркетингова) програма, що включає в себе концепції товару, збуту, просування та попередній аналіз можливостей ціноутворення, спирається на цінності та потреби потенційних клієнтів, конкурентні переваги ідеї, стан та динаміку ринкового середовища, в межах якого буде впроваджено проект, та відповідну обрану альтернативу ринкової поведінки.[29]

Висновки

У висновках узагальнюється проведений аналіз та зазначається:

- є можливість ринкової комерціалізації проекту;
- є перспективи впровадження з огляду на потенційні групи клієнтів, бар'єри входження, стан конкуренції, конкурентоспроможність проекту;
- є доцільною подальша імплементація проекту.

ВИСНОВКИ

В цій роботі проаналізовано варіанти використання різних метрик для оцінки якості звуку та відео, які надходять до споживачів у світлі постійного вдосконалення. Після цього розглянуто перспективні тенденції.

Тематичні дослідження, висвітлюють проблеми, пов'язані з реальним застосуванням ViSQOL. Отже ViSQOL можна використовувати для різних програм, але для будь-якого випадку використання потрібне ретельне дослідження. Система може бути налаштована під кожного користувача, який методом оцінки метрик її буде вдосконалювати і навіть, є можливість перенавчити модель відповідно до нових даних.

Я виявив, що розробники скептично ставляться до того, наскільки добре ViSQOL застосовуватиметься до їхньої проблеми, оскільки має унікальні характеристики. ViSQOL не гарантує надання значущого абсолютного MOS для випадків, які суттєво відрізняються від початкових умов, але є певна кореляція, яка корисна для їхнього випадку використання. Але використання додаткових показників, які можна використовувати для перевірки застосування ViSQOL, часто полегшує впровадження системи таких метрик.

В інших випадках, наприклад, у генеративному, це вимагає переробки алгоритму на фундаментальному рівні, який може включати різні представлення спектрограм або DNN. Досі не існує стандартного та широкодоступного набору даних суб'єктивної оцінки подібного масштабу.

Отже, запропонована нова версія ViSQOL, яка дає можливість інтегрувати різні команди у Google до реальних проблем. Існує ряд перспективних напрямків для майбутньої роботи, включаючи підходи на основі DNN, більш загальну модель і врахування нових підходів до генеративного аудіо.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. A. W. Rix, J. G. Beerends, M. P. Hollier та A. P. Niekstra, «Перцептивна оцінка якості мовлення (PESQ) — новий метод оцінки якості мовлення в телефонних мережах і кодах», у 2001 році IEEE International Conference on Acoustics, Speech, і обробка сигналів. Proceedings (Cat. No 01CH37221), vol. 2. IEEE, 2001, стор. 749–752.
2. ITU, «Перцептивна оцінка якості мовлення (PESQ): об'єктивний метод наскрізної оцінки якості мовлення вузькосмугових телефонних мереж і мовних кодеків», Int. телеком. Union, Женева, Швейцарія, ITU-T Rec. S.862, 2001.
3. J. G. Beerends, C. Schmidmer, J. Berger, M. Obermann, R. Ullmann, J. Pomy та M. Keyhl, «Перцептуальна об'єктивна оцінка якості прослуховування (POLQA), третє покоління стандарту ITU-T для кінця -до кінця вимірювання якості мовлення, частина I,» Journal of the Audio Engineering Society, том. 61, вип. 6, стор. 366–384, 2013.
4. ITU, «Оцінка якості сприйняття об'єктивного прослуховування», Int. телеком. Union, Женева, Швейцарія, ITU-T Rec. S.863, 2018.
5. T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, and C. Colomes, “PEAQ - The ITU standard for objective measurement of perceived audio quality,» Journal of Audio Engineering Товариства, вип. 48, вип. 1/2, стор. 3–29, 2000.
6. R. Huber і B. Kollmeier, «PEMO-Q-A new method for objective audio assessment using a model of auditor perception», IEEE Transactions on audio, voice, and language processing, том. 14, вип. 6, стор. 1902–1911, 2006.
7. А. Хайнс, Дж. Скоглунд, А. Кокарам і Н. Харт, «ViSQOL: об'єктивний слухач якості віртуальної мови», в IWAENC 2012; Міжнародний семінар з покращення акустичного сигналу. VDE, 2012, стор. 1–4.

8 А. Хайнс, Е. Гіллен, Д. Келлі, Дж. Скоглунд, А. Кокарам і Н. Харт, «ViSQOLAudio: об'єктивна метрика якості звуку для кодеків із низьким бітрейтом», Журнал Акустичної спільноти Америки, вип. 137, вип. 6, стор. EL449–EL455, 2015.

9. М. Нарбутт, А. Аллен, Дж. Скоглунд, М. Чінен та А. Хайнс, «AMBIQUAL – повна еталонна об'єктивна метрика якості для амбізонічного просторового аудіо», у 2018 році на Десятій міжнародній конференції з якості мультимедійного досвіду (QoMEX). IEEE, 2018, стор. 1–6.

10. W. B. Kleijn, F. S. Lim, A. Luebs, J. Skoglund, F. Stimberg, Q. Wang і T. C. Walters, «Низькошвидкісне кодування мови на основі WaveNet», у 2018 р. Міжнародна конференція IEEE з акустики, мови та обробки сигналів (ICASSP). IEEE, 2018 р., стор. 676–680.

11. J. Klejsa, P. Hedelin, C. Zhou, R. Fejgin та L. Villemoes, «Високоякісне кодування мови за допомогою зразка RNN», на Міжнародній конференції ICASSP 2019-2019 IEEE з акустики, мови та обробки сигналів (ICASSP). IEEE, 2019 р., стор. 7155–7159.

12. Р. Пренгер, Р. Валле та Б. Катанцаро, «Waveglow: потокова генеративна мережа для синтезу мовлення», на Міжнародній конференції IEEE з акустики, обробки мови та сигналів ICASSP 2019-2019 (ICASSP). IEEE, 2019 р., стор. 3617–3621.

13. А. Хайнс і Н. Харт, «Передбачення розбірливості мовлення за допомогою показника індексу подібності нейрограми», Мовна комунікація, том. 54, вип. 2012. – С. 306 – 320.

14. А. R. Avila, Н. Gamper, С. Reddy, R. Cutler, I. Tashev та J. Gehrke, «Ненав'язлива оцінка якості мовлення за допомогою нейронних мереж», у ICASSP 2019-2019 IEEE International Conference on Acoustics, Обробка мови та сигналів (ICASSP). IEEE, 2019 р., стор. 631–635.

15. Н. Gamper, С. K. Reddy, R. Cutler, I. J. Tashev та J. Gehrke, «Нав'язлива та ненав'язлива перцептивна оцінка якості мовлення за допомогою згорткової

нейронної мережі», у 2019 році IEEE Workshop on Applications of Signal Processing to Audio та акустика (WASPAA). IEEE, 2019 р., стор. 85–89.

16. А. Хайнс, Дж. Скоглунд, А. Кокарам і Н. Харт, «Стійкість показників якості мовлення до фонового шуму та погіршення мережі: порівняння ViSQOL, PESQ і POLQA», у 2013 р. Міжнародна конференція IEEE з акустики, Мовлення і обробка сигналів. IEEE, 2013, стор. 3697–3701.

17. А. Хайнс, Е. Гіллен і Н. Харт, «Вимірювання та моніторинг якості мовлення для передачі голосу через IP за допомогою POLQA, ViSQOL і P. 563», на шістнадцятій щорічній конференції Міжнародної асоціації мовлення, 2015 р.

18. С. Sloan, N. Harte, D. Kelly, A. C. Kokaram та А. Hines, «Класифікація бітрейту двічі закодованого аудіо з використанням об'єктивних характеристик якості», у 2016 р. Восьма міжнародна конференція з якості мультимедійного досвіду (QoMEX) . IEEE, 2016, стор. 1–6.

19. «Об'єктивна оцінка перцептивної якості звуку за допомогою ViSQOLAudio,» IEEE Transactions on Broadcasting, том. 63, вип. 4, стор. 693–705, 2017.

20. (2011) WebRTC. Онлайн. Доступно: webrtc.org

21. А. Хайнс, Е. Гіллен, Д. Келлі, Дж. Скоглунд, А. Кокарам і Н. Харт, «Відчувана якість звуку для потокової стереомузики», у матеріалах 22-ї міжнародної конференції АСМ з мультимедіа. АСМ, 2014, стор. 1173–1176.

22. Команда CoreSV. (2014) Тест прослуховування CoreSV. Онлайн. Доступно: <http://listening-test.coresv.net/results.htm>

23. Google. (2017) Спуск. Онлайн. Доступно: <https://abseil.io/>

24. (2008) Тестовий фреймворк Google. Онлайн. Доступно: <https://github.com/google/googletest>

25. (2015) Система збірки Bazel. Онлайн. Доступно: <https://bazel.build/>

26. Т. Мо та А. Хайнс, «Компенсація буфера джиттера в оцінці якості передачі голосу через IP», на 30-й конференції Irish Signals and Systems Conference (ISSC) у 2019 році. IEEE, 2019 р., стор. 1–6.

27. К.-К. Чанг і К.-Ж. Лін, “Libsvm: Бібліотека для опорних векторних машин”, транзакції АСМ для інтелектуальних систем і технологій (TIST), том. 2, № 3, С. 1–27, 2011.

28. Н. Zen, V. Dang, R. Clark, Y. Zhang, R. J. Weiss, Y. Jia, Z. Chen, and Y. Wu, “LibriTTS: A corpus derived from librispeech for text-to-speech, ” Препринт arXiv arXiv:1904.02882, 2019. Назва з екрану.

29. chromeextension://efaidnbmnnnibpcajpcglclefindmkaj/https://ep.kpi.ua/files/navchannia/mag/roz_startap_proektiv_met_vk.pdf – [Електронний ресурс]

