

Всеукраїнській конкурс студентських наукових робіт
з природничих, технічних та гуманітарних наук

Спеціальність: Інформаційні системи і технології

Конкурсна робота

Шифр роботи «Малинка»

«Інтонаційна виразність тексту при програмному озвучуванні»

2020-2021 навчальний рік

Зміст

Анотація	3
Вступ.....	5
1. Аналіз стану питання та постановка задачі.....	6
1.1. Аналіз стану в прикладній галузі	6
1.2. Аналіз теоретичних та методичних розробок.....	7
1.3. Постановка задач.....	9
1.4. Математична формалізація	10
2. Побудова інтонації відповідно до пунктуації	13
2.1. Виявлення зв'язку між інтонацією та пунктуацією в реченні	13
2.2. Застосування відмінностей інтонації відповідно до пунктуації ...	15
3. Побудова інтонацій, що виражають почуття мовця.....	18
3.1. Виявлення особливостей інтонації для базових емоцій	18
3.2. Застосування інтонацій для базових емоцій	20
3.3. Лексична емоційність і виразність тексту.....	23
3.4. Покращення обробленого запису	23
4. Опис програмного забезпечення та інформаційних технологій, що були розроблені	26
Висновки	29
Література	30
Додатки	31

Анотація

Актуальність

Зі збільшенням обсягів медіа контенту виникає потреба автоматизованої його обробки, зокрема озвучування.

Практика вимагає, аби результат програмного озвучування був близьким до людського мовлення. Існуючі практичні підходи не спроможні надати механічному голосу потрібного інтонування, тому що на теоретичному рівні не були в достатній мірі досліджені інструменти зміни голосу відповідно до певної емоції. Тому тема дослідження формування інтонації для різних емоцій є актуальною.

Мета: підвищення виразності озвучування текстів за допомогою встановлення закономірностей технічних характеристик звукових сигналів для різних емоцій.

Об'єкт досліджень: звукові характеристики голосу, синтезованого комп'ютером.

Предмет досліджень: інтонаційна зміна голосу при озвучуванні текстів комп'ютером.

Завдання дослідження, які дозволяють досягти поставленої мети

1. Аналіз наявних досліджень і порівняння готових розробок за темою роботи.
2. Дослідження мовленнєвих прийомів і характеристик голосу, які надають емоційного забарвлення прочитаному тексту.
3. Формалізація закономірностей зміни звукових характеристик для передачі різних інтонацій і їх програмна реалізація.

При виконанні досліджень та прийнятті рішень використовувались такі **методи та підходи:** аналіз і узагальнення мовленнєвих засобів вираження голосу, метод ковзного середнього, методи імітаційного моделювання, теорія ймовірностей.

Наукова новизна:

1. Запропонований метод формалізації емоцій при розбитті речень по складах.
2. Вдосконалений метод чисельного аналізу речень за допомогою методу ковзного середнього, ліній апроксимації та перетворення Фур'є.
3. Вдосконалений метод синтезу речень з урахуванням заданих емоцій за допомогою лексичного й синтаксичного аналізу речень.

Практична цінність роботи: В результаті виконання роботи було розроблене програмне забезпечення для емоційного озвучування текстів яке забезпечує сприйняття аудіо-інформації легшим, зрозумілішим і більш комфортним.

Стислий опис результатів дослідження: було проаналізовано фактори, що впливають на формування різних інтонацій, математично описано залежності зміни звукових характеристик відповідно до інтонацій; використано методики різних розборів речення та покращення звуку; було досягнуто надання інтонаційної виразності за типом речень (питальні - стверджувальні, окличні-неокличні), а також підсилення логічного наголошення емоційно забарвлених слів у реченні.

Робота має обсяг 30 аркушів, має 5 малюнків, 15 джерел посилань, додатки на 11 аркушах.

Ключові слова: аналіз тексту, звукові характеристики, інтонаційна виразність.

Вступ

Основні способи подання інформації – графічне та звукове представлення. Але в реальності аудіо формат випереджує візуальний: художні аудіо книги стають популярнішими за друковані; при слідуванні маршруту водії частіше використовують голосового помічника замість карт; для зменшення витрат перекладені фільми озвучують автоматизовано, попри помітну монотонність, «технічність» голосу. Це свідчить про нестачу людського ресурсу для озвучування, потребу в якісному програмному рішенні. Крім того, за твердженням Н. П. Волкової в навчальному посібнику «Педагогіка» [1] близько 30% людей – аудіали, тобто найкраще сприймають інформацію через слух. Виразність і емоційність звучання впливає на утримання уваги слухачів і можливість запам'ятати почуту інформацію.

Покращити інтонаційну передачу тексту можна за рахунок регулювання основних характеристик голосу: гучності, тону, швидкості [7]. Вони є основним інструментом перетворень звуку. Необхідно проаналізувати, як їх зміна пов'язана з емоціями, на що впливає їх підсилення. Важливо відзначити, наскільки різко змінюється певна ознака та які ефекти досягаються при цьому.

Сам текст також містить лінгвістичні ознаки-маркери [8], де мають змінюватися висота звуку, гучність. Це пунктуація, питальні слова (як, чому, звідки, коли...), логічно наголошені слова та сполуки слів (найбільш значущі), вставні та інші емоційно забарвлені слова (вірогідно, малесенький...) тощо.

Приміром, якщо речення містить слова на зразок «страшно», «хвилююсь» (тобто такі, що виражають переживання), то голос має бути менш гучним і високим та дещо «рваним» за рахунок штучного додавання занадто високих або низьких частот. Якщо ж речення містить слова, що вказують на радість і щастя, то гучність і висоту треба збільшити, звучання зробити якомога плавнішим [10].

Тому потрібно навчитися визначати такі маркери в тексті й відповідним чином програмно оброблювати фрагменти аудіо, які їх містять.

Загалом це дозволить зменшити людські ресурси й витрати на них у галузях, які працюють із обробкою звуку та озвучуванням текстів.

1. Аналіз стану питання та постановка задачі

На даний час існуючі засоби дозволяють озвучувати текст механічно, без дотримання інтонацій і виразності. Тому основним завданням є підвищити виразність і підсилити передачу інтонацій при озвучуванні.

1.1. Аналіз стану в прикладній галузі

Теоретичні дослідження інтонації з філологічної точки зору активно досліджувала Багмут Алла Йосипівна - доктор філологічних наук, яка з 1991 до 2002 року була завідувачем відділу фонетики та інтонології й старшим науковим співробітником відділу соціолінгвістики Інституту української мови. У її працях подається дуже детальний розбір інтонацій за синтаксичними й семантичними особливостями речень – деякі монографії містять розбір лише розповідного речення з рівною інтонацією [11], [12]. Але разом із цим інтонації, що надаються мовленню залежно від психоемоційного стану мовця, тобто самих емоцій від прочитаного тексту, досліджені мало.

Формалізувати деякі дослідження А.Й.Багмут вдалося Міннігалімову Руслану Тимуровичу в роботі «Аналіз та синтез українського мовлення» [7], де він математично описав закономірності зміни частот основного тону для розповідних стверджувальних і заперечних, а також питальних речень. Проте його практичні дослідження подекуди суперечать одне одному. Автор пояснює це тим, що різні мовці мають різний стиль прочитання, тому необхідно збільшити статистичну базу дикторів для подальших практичних розробок.

Через відсутність автоматизованих засобів озвучування тексту близьких до мовлення людини, монотонно начитаний текст оброблюється вручну людиною. Проте для вирішення цієї проблеми активно проводяться дослідження й ведуться розробки, переважно з використанням машинного навчання.

Над синтезом голосів для озвучування працювали американські фахівці AT&T Laboratories. Вони створили програму, яка наслідує людський голос після обробки 10-40 годинного справжнього запису. Проте, як відмітили самі розробники, програма ще не здатна повною мірою відтворювати голоси

реальних людей, а звучання синтезованих записів досить технічне й не враховує емоцій мовця при озвучуванні [14].

У минулому році надати емоційності комп'ютерному голосу спробувала британська команда Sonantic. У проекті використовується штучний інтелект, який аналізує великі обсяги людських записів [15]. Синтезований голос справді схожий на природне звучання, але наразі розробка орієнтована лише на негативну емоцію відчаю й плачу.

Спільною рисою описаних вище рішень є використання штучного інтелекту для обробки готових записів. Проте якщо текст містить математичні закономірності при побудові інтонації, то використання їх формалізації було б гарним покращенням і спрощенням у додаток до методів машинного навчання.

Основна термінологія, що стосується характеристик звуку [2]:

гучність звуку - суб'єктивне слухове відчуття інтенсивності (сили) звуку;

висота звуку - суб'єктивна оцінка якості звуку, яка визначається частотою звукової хвилі;

швидкість звуку - швидкість поширення акустичних хвиль.

Основна термінологія, що стосується мовних характеристик [3]:

інтонація - сукупність звукових мовних засобів, завдяки яким передають смисловий, емоційно-експресивний і модальний характер висловлення, комунікативні значення та ситуативну зумовленість, стилістичні забарвлення тексту, індивідуальність виражальних прийомів мовця;

виразність – емоційна насиченість мовлення;

логічний наголос - особливе виділення голосом слова в реченні з метою підкреслити важливість його в смисловій структурі речення;

емоційно забарвлена лексика - слова, які мають негативний або позитивний відтінок у своєму значенні.

1.2. Аналіз теоретичних та методичних розробок

Один із наявних підходів озвучування текстів - монотонне прочитання зі збереженням сталої висоти, гучності звуку тощо та без дотримання пауз. Такий

спосіб передачі аудіо інформації не може бути спрямованим на широкого користувача, оскільки є складним для сприйняття. Він вимагає постійного зосередження уваги й самостійного логічного розбиття прослуховуваного тексту на синтаксично цілісні фрагменти за змістом.

Інший покращений підхід – прочитання усе ще рівним голосом, проте з витримкою пунктуаційних пауз. Так, після коми йде мінімальна пауза, а, наприклад, після тире або крапки – більша. Цей спосіб полегшує сприйняття тексту за рахунок того, що синтаксично цілі одиниці мовлення (цілісні речення чи їх частини) сприймаються окремо завдяки паузам [13]. Недоліком залишається відсутність інтонаційної різниці між фрагментами тексту. Це призводить до того, що слухач швидко втомлюється й втрачає концентрацію. Як наслідок, інформація засвоюється лише частково, а запам'ятовування залишається поверхневим [6], адже більш і менш важливі уривки тексту нічим не відрізняються при прослуховуванні. Такий підхід застосовується у вбудованих бібліотеках мови програмування Python, у додатках для перегляду текстових файлів із функцією озвучування, браузерах тощо.

Один із кращих існуючих підходів – озвучування на базі машинного навчання. При такому способі генерується ряд звукових ефектів для певного набору емоцій. Проте «присвоєння» такої емоції конкретному реченню відбувається вручну, тобто мовні особливості самого тексту не враховуються. Дане рішення застосовується, до прикладу, у британському стартапі Sonantic.

Суперечність між даними підходами й практичними потребами полягає в тому, що принципи побудови інтонації при озвучуванні текстів сформульовані досить нечітко й базуються на людському «відчутті мови», яким часто пояснюють навички виразного читання у філологічних джерелах. Проте якщо людині під силу однозначно визначити настрій того чи іншого речення, то, імовірно, існують певні закономірності в тому, як саме вона це робить і що включає в себе таке «відчуття мови». Із цього можна зробити висновок, що знайдені залежності можна узагальнити й формалізувати, наблизити їх до

поняття правила. Це може стати гарним спрощенням і підґрунтям алгоритмів, які використовуються при машинному навчанні.

Отже, логічно отримуємо мету роботи: встановити закономірності побудови інтонації за ключовими ознаками в тексті та на основі них створити програмне рішення для озвучування текстів.

1.3. Постановка задач

1. Визначити логічні наголоси. Кожне речення (або його частина) має логічний наголос, тобто слово, яке найбільш значуще й сильніше виділяється інтонацією. Проблема в тому, що іноді він може бути неоднозначним, залежно від того, якого відтінку мовець вирішить надати фразі. Але в більшості можна створити чітке правило або узагальнення (подібно до того, як кожен читач має свій стиль і манеру прочитання). Тоді найправильніше робити логічний наголос на слові, що є найбільш значимим у фразі. Це має бути самостійна частина мови, яка найбільше впливає на сенс фрази та є найбільш емоційно забарвленою. Можна упорядкувати слова за пріоритетністю частин мови: іменник, якщо його нема у фразі, то дієслово, інакше – прикметник, далі займенник... Але при накладанні великої кількості інших лексичних «фільтрів» вони починають перекривати ефект одне одного, тому таке впорядкування не дає бажаного результату. До того ж багато слів можуть належати до різних частин мови залежно від контексту. Тоді найбільш дієвий спосіб – вибрати певну скінченну групу слів, які будуть вважатися ключовими й робити акцент на них. Наприклад, це можуть бути прислівники та прикметники, бо саме вони надають як емоційного, так і смислового різноманіття фразам.

2. Надати реченню базових інтонацій. Оскільки теперішні засоби різниці у звучанні речення, коли йде, скажімо, «?» чи «!», практично не дають – те, що речення питальне або окличне, можна зрозуміти хіба завдяки його смислового навантаженню. Проблему доцільно вирішити посимвольно аналізуючи речення з кінця і позначати розділовий знак, яким воно завершується, а потім читати з

тією «інтонацією» (гучністю, висотою тощо), якої вимагає той чи інший розділовий знак.

3. Розробити наближені правила зміни характеристик голосу на основі вирішення перших двох завдань, уникаючи накладань кількох «фільтрів» на один фрагмент речення, у результаті чого звучання цього уривку може бути непередбачувано підсилене (напр.: питальне слово «коли» визначене як прислівник, а відтак, підсилене двічі).

4. Дослідити закономірності побудови базових емоцій. Для можливості оперування різними емоціями при їх мінімальній кількості слід розглянути перш за все протилежні за настроєм. Потрібно визначити, що вирізняє радісний голос від засмученого, агресивний від депресивного тощо. Тобто виділити певну унікальну особливість, за якою можна було б однозначно ідентифікувати той чи інший настрій.

5. Описати математично отримані закономірності зміни голосу для базових емоцій. Узагальнення дасть можливість абстрагування від безпосередньо значень звукових характеристик і спрощеного сприйняття конкретного настрою подібно до функції.

6. Розширення спектру емоцій для подальшого аналізу, враховуючи проміжні (напр.: між агресивним та депресивним настроєм може бути розгубленість або тривожність і т. п.).

7. Оскільки в живій мові люди оперують не лише повнозначними словами для передачі почуттів, то доцільно розглянути різницю між сміхом і плачем, вигуками, що виражають страх і здивування, і т.д. Це допоможе передавати настрої мовця більш точно та яскраво. Особливо корисно на практиці це може бути застосовано саме при озвучуванні кінострічок, де крім реплік акторів використовуються й інші звукові ефекти.

1.4. Математична формалізація

Слово складається зі складів:

$$w = \{n, \dots, n, k, n, \dots, n\}, \text{ де}$$

n – ненаголошений склад,

k – наголошений склад.

Речення складається зі слів:

$$s = \{q, \dots, q, l, q, \dots, q\}, \text{ де}$$

q – логічно ненаголошене слово,

l – логічно наголошене, найбільш значуще слово.

Але речення можна також представляти як послідовність складів, тоді можна записати його так:

$$s = \{n, \dots, n, k, n, \dots, n\}, \text{ де}$$

n – склад логічно ненаголошеного слова або будь-який ненаголошений,

k – наголошений склад найбільш значущого слова.

Поряд із кортежем складів запишемо кортеж, який містить команди для комп'ютера, що містять інформацію про зміну основних технічних параметрів звуку й введемо відповідні позначення для них:

- гучність – v ;

- довжина – t ;

- частота (тон звуку) – r ;

- пауза (до складу або після) – p .

Тоді фраза (речення чи слово) для комп'ютера виглядатиме, наприклад, так:

$s = \{t, v, r, q, \dots, v, r, q, l, p, v, r, q, \dots, v, r, p, q\}$ – на початку задається тривалість усього запису, а потім перед кожним складом вказується потрібна гучність і частота, за потреби ще й пауза, далі йде текст самого складу.

Розглянемо такі базові емоції, що покривають більшість людських відчуттів. Уведемо відповідні їх позначення, враховуючи те, що деякі з них є протилежними до інших:

- радість – j ;

- сум - j' ;
- агресія (реакція нападника) – a ;
- розгубленість/тривожність (захисна реакція жертви) - a' ;
- спокій - c ;
- роздратування/невдоволення - c' .

Кожна емоція має свій характерний рисунок у часі, який можна представити набором функцій основних технічних параметрів звуку, використовуючи попередні позначення:

$v = j(t); r = j(t)$ – для радісної інтонації;

$v = j'(t); r = j'(t)$ – для сумної;

$v = a(t); r = a(t)$ – для сердитої, агресивної;

$v = a'(t); r = a'(t)$ – для тривожної, схвильованої;

$v = c(t); r = c(t)$ – для спокійної;

$v = c'(t); r = c'(t)$ – для роздратованої.

Так само, як речення по-різному закінчуються розділовим знаком, вони різняться інтонаційним завершенням. Загальна характеристика інтонації розповідного речення визначається спадом тону в кінці речення, рухом кривої мелодійної лінії вниз, «інтонацією крапки». Це позначення розповідного речення дане на відміну від інтонації питального і окличного, де в кінці речення тон наростає, рух мелодії висхідний [7].

Усередині фрази теж відрізняються звучанням: різна плавність, різкі чи поступові наростання висоти, міра затихання чи збільшення гучності тощо. Тому треба проаналізувати дані характеристики для різних типів речень і емоцій. Побудувавши графіки залежностей і виконавши необхідні перетворення (напр., згладжування) можна виявити спільні риси, що стануть основою для узагальнення емоцій за певною ознакою. Для роботи з аудіо записом варто перетворити їх у функції для обробки масивів, адже звуковий сигнал при обробці зазвичай подається у вигляді набору значень на зразок масиву.

2. Побудова інтонації відповідно до пунктуації

Основним лексичним засобом позначення емоцій у тексті є пунктуація[9]. Вона дає вказівки, як речення має початися й закінчитися, яка взаємодія зі слухачем передбачена, які почуття треба в нього викликати, наскільки довгі й часті паузи потрібно витримати. Емоції, що формуються завдяки інтонації, не залежать від мовця, його манери оповіді й відчуттів, стилістики тексту й змісту, цільової аудиторії тощо. Тобто вони будуть неясковими, проте завжди однаковими. Якщо кома й тире вказують тільки на потребу в паузі та її довжину, то ключову роль у визначенні інтонації відіграють розділові знаки в кінці речення.

2.1. Виявлення зв'язку між інтонацією та пунктуацією в реченні

За емоційним забарвленням речення бувають окличні та неокличні. Відповідно мають або не мають у кінці знак оклику. Вони відрізняються тим, наскільки важливим є акцент на їх змісті та визначають наскільки сильне враження інформація справить на слухача [9].

Неокличні речення вимовляються спокійним голосом, без екстремального підвищення чи зниження висоти та гучності, не перенасичені акцентами на словах і їх змісті [11].

Окличні вимовляються більш піднесено, гучно, вищим тоном, інтонація може бути менш плавною, з надривами та яскравими акцентами на ключових словах. Такий тип речень забезпечує підсилення емоції й привертає увагу слухача. Але при надто частому використанні окличних речень у тексті він може набувати войовничого, агресивного, спонукального до певних дій характеру, через що його початковий зміст може трактуватися слухачем неправильно чи спотворено [13]. Отож слід уникати окличної інтонації там, де її використання насправді не виправдане. Схематично відмінності представлені в додатку 1 - для гучності та додатку 2 – для висоти.

За метою висловлювання речення бувають розповідні, питальні й спонукальні [3]. Розглянемо відмінності між ними, беручи за приклад неокличну інтонацію, рівний спокійний голос мовця.

Розповідні повідомляють про подію, факт чи явище. Наприкінці таких речень ставиться крапка, іноді - три крапки для позначення незавершеності чи неповноти думки. Якщо вкінці стоїть крапка, то інтонація протягом усього речення залишається рівною, а наприкінці спадає, голос затихає [11]. Гучність рівномірно тримається на одному рівні, досить швидко знижуючись вкінці речення. Це виражає завершеність думки й упевненість у висловлюваному. Таке речення сприймається емоційно нейтрально й не привертає особливої уваги слухачів. Переважно носить інформативний характер і не бере участі у формуванні почуттів людини та її реакції на почуте.

Якщо речення завершується трьома крапками, то спад інтонації більш плавний, помірно згасаючий, голос затихає більше, але поступово. Думка не завершена, залишається простір для власних роздумів слухачів і їхньої особистої оцінки сказаного. Часто саме для підсилення такого ефекту в кінці речення з трьома крапками витримується довша пауза [5]. Це дозволяє слухачам прислухатися до внутрішніх відчуттів і сформуванню своєї ставлення до отриманої інформації, зрозуміти чого не вистачає для довершеності думки, чому мовець залишає деяку невпевненість і, можливо, вагається у власному висловленні. Знак оклику в такому реченні часто виражає тривожність, пов'язану з відчуттям страху.

Оскільки в більшості з підвищенням гучності, підвищується й висота звуку, адже акцент і увиразнення забезпечується обома засобами, надалі графічно зображатимемо їх як одну сукупність звукових характеристик. Схематично це показано в додатку 3. Щоправда, в даному випадку апроксимація відкидає важливий для окличного речення стрибок значень.

Спонукальні речення також мають крапку вкінці. Відрізняються від розповідних змістом: виражають якість прохання чи вимогу, і метою: спонукають до дії. Інтонаційно мають сильний акцент на ключових словах [12],

які, власне, і визначають спонукання до дії (напр., спитай, піди, зачекай, розкажи, принеси і т.п.). Зазвичай такі слова йдуть на початку речення, інверсія трапляється рідко і є радше нетиповим явищем для спонукальних речень. У зв'язку з цим і підвищення голосу, і збільшення гучності характерне для початку речення. Різницю між розповідним і спонукальним реченнями при однаковому емоційному забарвленні показана в додатку 4.

Питальні речення виражають запитання. Мають виражений логічний наголос на найбільш значущому слові та підсилення інтонації вкінці. Якщо таке речення містить питальні слова (як?, звідки? тощо), то логічно наголошуються вони. Тоді в кінці речення інтонація йде на спад, як і в розповідних реченнях. Міра підсилення звукових характеристик визначає, наскільки важливо зацентувати увагу на даному питанні [8]. При занадто сильному інтонаційному вираженні зміст сприймається так, наче мовець здивований або висловлює недовіру до слухачів. Тому, якщо досягнення такого ефекту не стояло на меті, не слід зловживати різким підвищенням тону й перепадами гучності [5]. Голос має бути спокійним і плавним. Схематично інтонація питальних речень з ключовими словами й без відображена в додатку 5.

2.2. Застосування відмінностей інтонації відповідно до пунктуації

Для програмної роботи з текстом його можна зчитувати з файлу. Це зручно - користувач сам задає, із яким файлом працювати, редагує текст у ньому, змінює назву та розміщення на комп'ютері.

Для початку достатньо використати готові засоби озвучування й отримати монотонне прочитання тексту. У даній роботі використовується мова програмування Python, яка містить бібліотеку pyttsx3, що призначена для перетворення тексту в звуковий формат. Голос вбудованого мовця монотонний, технічний, речення однакові за звучанням, незалежно від кінцевого розділового знаку. Дотримано лише короткі паузи після ком і довші після тире.

Працювати з цілісним звуковим файлом незручно – зміна якоїсь характеристики буде накладатися на весь звуковий ряд. Отже, треба розбити

запис на певну кількість однакових фрагментів і надати їм числового представлення. Тому для подальшої обробки звуку застосовується перетворення звукового ряду на масив значень гучності з певною частотою. Стандартно для таких цілей використовується wav-формат. Якщо запис надходить у іншому форматі, він зазвичай попередньо також перетворюється на wav.

Під час озвучування працювати потрібно одночасно й із текстом (оброблювати звучання згідно з пунктуацією), і безпосередньо з його аудіо представленням. Текст із файлу в загальному становить рядок. Так, спершу розбиваємо текст на масив речень, розділювачами слугуватимуть наступні розділові знаки: «.», «!», «?», «...». Потім у циклі оброблюємо кожне з них. Для визначення кінцевого розділового знаку аналізуємо речення посимвольно з кінця.

Основний інтонаційний акцент не надається цілій сполуці слів, а зосереджується на одному складі ключового слова у фразі. Наростання інтонації може бути більш або менш плавним, але не раптовим. Тому склад (або декілька складів) перед найбільш інтонаційно окресленим і після нього теж будуть дещо увиразнені. Це допоможе згладити перепад гучності й висоти тону й зробить звучання більш природним і приємним.

Для підсилення звукових характеристик того чи іншого фрагменту аудіозапису не варто застосовувати множення значень на певний коефіцієнт. Покажемо зміну значень звукових параметрів на прикладі гучності: $v = v_0 * C$, де v – отримана гучність на фрагменті,

v_0 – початкова гучність на фрагменті,

C – сталий коефіцієнт, який додається до значення початкової гучності.

Приклад програмної реалізації наведено в додатку 14. Коефіцієнти тут вибираються із заздалегідь обробленого масиву, в даному випадку `coef_list = [4, 7, 3]`. Для інших звукових характеристик формула й програмна реалізація аналогічна.

Такий підхід може дати неочікувано сильний або занадто малий результат. Крім того, якщо запис має велику амплітуду гучностей або висот, найвищі з них можуть екстремально підсилитися, що призведе до неприємного уривчастого звучання та погіршення якості звуку через появу відчутного шуму. Для регулювання звукових характеристик краще додавати певне число до їхніх значень: $v = v_0 + C$. Приклад програмної реалізації є в додатку 15.

Потрібно зважати на те, наскільки тихим був оригінальний запис і якою була початкова висота звуку, аби в порівнянні з цим коригування не було надто різким при додаванні великого числа або, навпаки, не виявилось, що число було замалим, і зміни не відчутні. Якщо початковий запис уже мав неоднорідну гучність і висоту, треба її узагальнити.

Середнє арифметичне значень однієї з характеристик не завжди дає потрібний результат: приміром, якщо є значення, які кілька разів більші за основну масу [2]. Тому спочатку треба взяти середній діапазон, відкинувши надто великі й малі значення. Це можна зробити за допомогою усереднення або інших згладжувальних методів. Також можна виконати апроксимування масиви значень. Для таких цілей достатньо лінійного, тому що характер самої функції нас не цікавить. Метою є вибір діапазону, у якому зосереджена більшість значень. Схематично принцип такого методу проілюстровано в додатку 12. Прозоро-зеленим кольором позначено діапазон, який використовується для подальших обчислень, червоним – ті значення, що відкидаються.

Інший спосіб відкидання зайвих значень, більш зручний для програмної реалізації – встановлення нижньої і верхньої межі допустимих значень:

```
y = np.where(y > 2000, 2000, y)
y = np.where(y < -2000, -2000, y)
```

Після цього можна взяти середнє значення або моду як базове початкове значення тієї чи іншої характеристики (`np.mean(y)`). Результат встановлення таких порогів побудовано програмними інструментами (додаток 16) і показано в додатку 13, де синім кольором позначено значення частот початкового запису, а зеленим – вибраний діапазон.

3. Побудова інтонацій, що виражають почуття мовця

Іншим інструментом побудови інтонації є власні відчуття та переживання мовця, його особисті враження від тексту. Це є ідентифікатором його психологічного стану, який, безумовно, впливає на голос.

3.1. Виявлення особливостей інтонації для базових емоцій

Емоції є відображенням психічного стану людини й вказують на те, як вона оцінює обставини, факти тощо. Через емоції виявляється її ставлення до об'єкту чи явища. Емоції є віддзеркаленням світосприйняття. Будь-які зовнішні подразники (спілкування з іншими, перегляд фільмів, читання, наукова діяльність і т. д.) та внутрішні переживання впливають на психологічний стан людини й формують її настрій [4]. У загальному всі емоції можна поділити на 3 групи: позитивні, негативні та нейтральні.

Першу група виражає піднесений настрій і задоволення. Такі емоції додають енергії, спонукають людину до діяльності й досягнення результату. Щодо звучання, позитивні емоції високі за тоном голосу, зі звичайною або дещо підвищеною гучністю, незначною ритмічністю, але з плавними перепадами, висхідною інтонацією фраз. У кінці речення голос затихає плавно, але не розтягнуто. Зміна цих характеристик показана в додатку 6. Для визначення загального характеру спадання чи зростання графіків параметрів звуку та їх порівняння з іншими емоціями у роботі застосовується поліноміальна апроксимація. Вона найбільш точно відображає інтонаційні зміни. При цьому апроксимації 2-го порядку недостатньо – перепади гучності чи висоти згладжуються й не беруться до уваги, хоча є важливими для побудови емоцій. Апроксимація 4-го і вище порядків наближується до початкового графіка надто точно, відображаючи навіть незначні коливання голосу. Це робить узагальнення таких графіків складним, бо тоді кожна фраза буде значною мірою відрізнятися за характером. Оптимальний варіант – 3-й порядок: забезпечується достатнє згладжування графіків, зберігаються найбільші перепади значень.

Отож можна математично узагальнити функції позитивних емоцій. Оскільки найкращий результат апроксимації досягається за 3-го порядку, функція є кубічною. Коефіцієнти ліній апроксимації допустимо округлити, адже метою є наближений вигляд функції. Тоді для гучності та висоти позитивних інтонацій (візьмемо радісний голос за основу) відповідно отримується:

$$v = 0.01t^3 - 0.17t^2 + 1.34t + 1.63,$$

$$r = 0.01t^3 - 0.16t^2 + 1.11t + 0.95, \text{ де } t - \text{ час.}$$

Аналогічні узагальнені функції наведено для всіх розглядуваних у подальшому емоцій у підписах ліній апроксимації на рисунках. Округлення для кожної з них окремо не ілюструється, оскільки операція є простою та зрозумілою.

Друга група, навпроти, означає неприємні відчуття, невдоволеність чимось і заперечення. Якщо позитивні емоції, в цілому, дуже споріднені й за своєю природою, і за звучанням, то негативні дають значно ширшу палітру таких собі шаблонів.

Основною відмінністю у звучанні від позитивних є відсутність плавності, ступінчастість сигналу. Також для більшості (не для всіх) негативних емоцій характерне зниження висоти звуку [5]. А от гучність може бути як зменшена, так і збільшена, залежно від гостроти емоції й мети її вираження. Розберемо ці відтінки детальніше.

Як правило, коли людина відчуває роздратування чи невдоволення, її голос стає нижчим (додаток 7). Але відчуття страху й тривоги супроводжується протилежним явищем: голос стає вищим, дуже неоднорідним, неплавним і більш протяжним(додаток 8). Часто голосні звуки видовжуються й підсилюються, у той час як приголосні губляться й заміщуються короткими паузами, у живому мовленні спричиненими незначними раптовими вдихами людини [4].

Збільшення гучності свідчить про «нападаючий» настрій, як показано в додатку 9. Такі емоції виникають при критичному психологічному напруженні [4], швидко розвиваються й інтонаційно наростають. Про це свідчить різке й упевнене закінчення речень, що виражають подібні емоції, при мовленні [13].

Для додаткового увиразнення іноді застосовується пришвидшення або поступове наростання швидкості звуку [13]. До основних таких емоцій належить агресія (різкі акценти й підсилення гучності), роздратування (уривчастість, наростання гучності незначне), невдоволення (звучання плавне, без наростання гучності, але з чітким акцентом на причині), пихатість (схожа з роздратуванням, але без зниження висоти) тощо. Інші емоції, які відносять до цієї групи, є радше похідними від наведених.

Протилежний інструмент (зменшення гучності й розтягування фраз) – захисна реакція, схвильованість, розгубленість і безпомічність. Подібні переживання пригнічують людину, погіршують настрій, самопочуття, зменшують продуктивність, розважливність [4]. Спектр таких м'яких емоцій надзвичайно широкий: туга, сум (додаток 10), розгубленість, відчай, зневіра, почуття провини й багато інших. Відрізняються вони насправді першопричиною, тобто в тексті – за змістом [10]. Інтонаційно практично однакові, отже, програмно вимагають єдиної реалізації.

Нейтральні емоції характеризують спокійний, врівноважений стан мовця (додаток 11). Тут немає безпосереднього зв'язку між ступенем задоволення потреби та реакцією психіки людини. У таких емоціях відсутні стрибки звукових характеристик, голос плавний, звичної висоти й комфортної гучності. Зазвичай відповідають розповідним неокличним реченням. Прикладом таких емоцій є зацікавленість або байдужість.

3.2. Застосування інтонацій для базових емоцій

Використовуючи наведені в попередньому пункті узагальнені схеми емоцій, виконуються необхідні підсилення чи перетворення звукового сигналу. Головною відмінністю від побудови інтонації на основі пунктуації речення є те,

що застосовувати зміни треба до всього речення, а не лише до окремих його частин (слів, складів). Це забезпечить плавний перехід емоції, якість звучання та наближеність до природної мови людини.

Оскільки для надання фразі певної емоції, що базується на відчуттях мовця, немає обов'язкової потреби аналізувати сам текст, варто передбачити можливість обробки запису без початкового текстового представлення. Головною ціллю є розбиття звукового ряду на достатньо малі фонетичні одиниці. Зважаючи на те, що акценти (як сильні, так і слабкі) при вимові припадають на голосні звуки, доцільно ділити запис на склади, бо один склад містить рівно один голосний звук.

Одним із способів є використання готової бібліотеки Python для розпізнавання мови. За допомогою її засобів можна перетворити аудіозапис у текстовий формат, а вже потім його розбити на склади й працювати з ним так само, як у випадку з пунктуацією. Проте цей метод вимагає додаткових технічних засобів, зайвих перетворень, обтяжування програми завантаженням бібліотеки та збільшення самого часу виконання програми. Тому використовуємо ефективніший підхід.

Проаналізуємо можливість роботи лише зі звуком, без використання сторонніх засобів. Припустимо, що можна розбити звукову доріжку на різні фрагменти за часом, наприклад, по 0.5с. Виникає проблема: основною фонетичною одиницею, яка може містити акцент, є склад, але склади різняться за тривалістю звучання. Відповідно після обробки запису акценти можуть бути зміщені неправильно, що відобразиться на якості звучання.

Отож використаємо інший спосіб. Якщо співставити графічне представлення масиву гучностей якогось запису й текстовий варіант фрази, бачимо, що коливання гучності відбуваються на кожному складі. Це проілюстровано на рис.1.

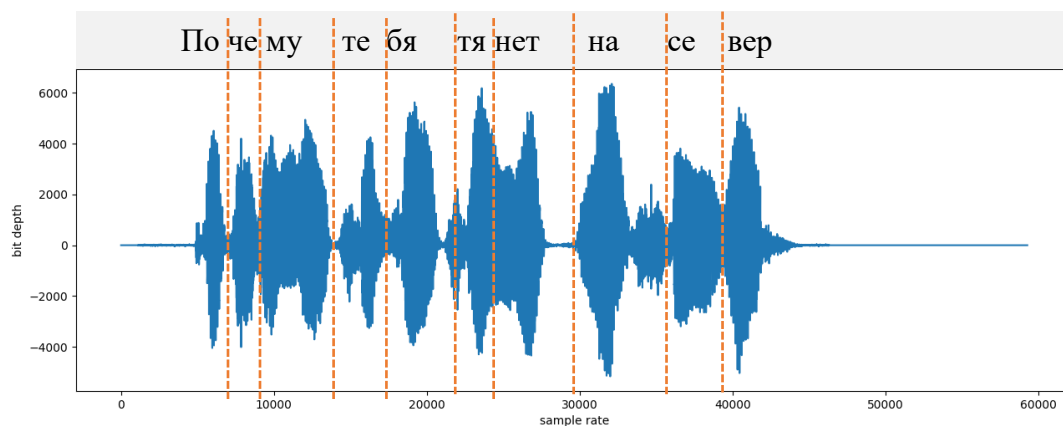


Рис.1. Графічна відповідність тексту й голосу

Так, на голосному звуці гучність зростає, а на приголосному спадає й майже зовсім затихає на шиплячих і свистячих. Із цього слідує висновок, що можна розбити аудіозапис на фрагменти-склади, тобто частинки підсилення-затухання, якщо встановити поріг «тиші». Будемо вважати, що значення, які вищі за цей поріг, відповідають голосним звукам – ключовій складовій складу (звук), а ті, що нижчі – приголосним (тиша). Тоді можна оброблювати звук окремо по частинах (аудіофрагментах фраз тексту або складів слова), а після об'єднувати їх назад у один файл.

Саме ці фрагменти тиші й будуть слугувати розділювачами при розбитті запису на склади. За замовчуванням візьмемо, що читання бібліотекою `pytsx3` відбувається зі швидкістю 100% (значення може бути як меншим, так і більшим) і гучністю 1 (значення від 0 до 1). Оптимальним проміжком «мовчання» для визначення межі складу є 50мс, а порогом тиші – приблизно мінус 30дБ (dBFS). Тоді для регулювання часу тиші, зменшуємо його на стільки ж відсотків, на скільки відсотків збільшуємо швидкість (для 150% швидкості час тиші складатиме приблизно 40мс). Для регулювання порогу тиші, на скільки відсотків зменшуємо гучність, на стільки ж відсотків зменшуємо й поріг, тобто збільшуємо модуль порогового значення (модуль, бо це значення залишається від'ємним). Так, для гучності 0.8 отримаємо поріг тиші мінус 36(приблизно до мінус 40).

3.3. Лексична емоційність і виразність тексту

Для покращення взаємодії з аудиторією мовці широко використовують логічні наголоси [6] як засіб розставлення правильних акцентів у реченні. Таким чином вони підкреслюють найважливіший фрагмент фрази. Зазвичай це найбільш значуще слово чи словосполучення, що має основне смислове навантаження у реченні.

Крім того, завжди виділяються інтонацією емоційно забарвлені слова – такі, що мають позитивний або негативний відтінок у значенні. Вони роблять текст цікавішим, насиченішим і привертають увагу слухача до деталей. Такі слова перш за все вирізняються зменшувально-пестливими суфіксами або суфіксами згрубілості, рідше – префіксами, що підсилюють міру ознаки (пре-, якнай-, най-і т.п.), прислівником «якомога» спереду [8]. Також варто пам'ятати про слова, які вказують на особливі додаткові прикмети [12]. Унікальної ознаки об'єкту чи явищу найчастіше надають прикметники чи прислівники, тому їх слід виділяти голосом.

Для знаходження таких слів у тексті за частиною мови найзручніше використати функцію бібліотеки Python, оскільки перебір усіх слів певної частини мови або створення словника вручну є досить часозатратним, при тому що він буде дуже подібним до вже вбудованого. Емоційно забарвлені слова можна знаходити за типовими суфіксами – перебрати їх неважко й вручну.

3.4. Покращення обробленого запису

Будь-які перетворення оригінального запису частково спотворюють звучання й погіршують якість аудіо (хіба що за винятком ідеальної оберненої функції). Тому повністю їх усунути практично неможливо [2], натомість можна мінімізувати отримані дефекти. Без графічного порівняння з початковим записом для слухача вони будуть непомітні.

Найчастіше причиною погіршення звучання є поява незначного шуму чи рипіння, у гіршому випадку – спотворене неприємно різке звучання в місцях підсилення. Це відбувається, коли з'являються надвисокі чи наднизькі частоти

відповідно, порівняно з переважаючим діапазоном запису, що показано на рис.2 та рис.3. Особливо відчутний такий ефект, якщо при роботі з масивом значень звукових характеристик аудіо застосовувалося множення чи ділення без явної на те потреби.

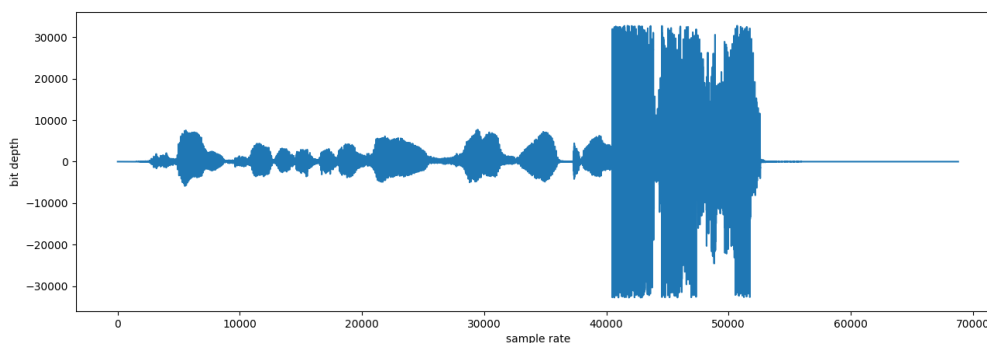


Рис.2. Шум, викликаний занадто високими частотами

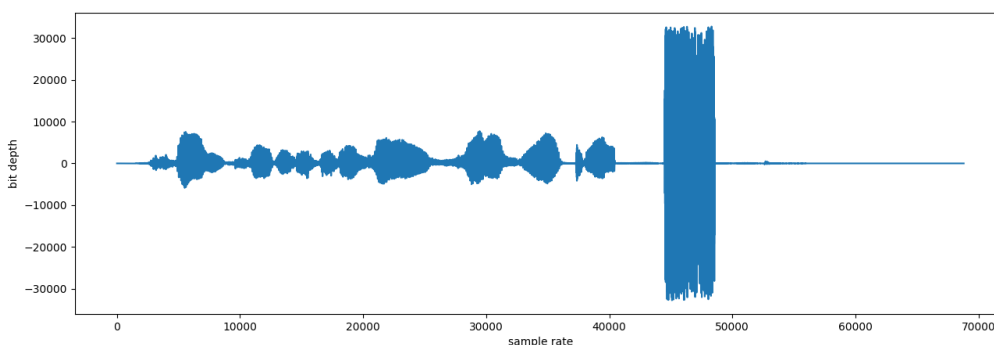


Рис.3. Шум, викликаний як критичним підвищенням, так і критичним зниженням частот

Для усунення цього небажаного явища зазвичай використовуються усереднювальні та згладжувальні фільтри [2]. Зокрема, гарні результати дає ковзне середнє, якщо застосовувати його для кожного значення, а діапазон для усереднення брати в межах 10-20 значень. Тоді звук стає плавнішим і природнішим, а нові незначні спотворення практично не відчутні. Програмна реалізація методу надана в додатку 17. Результат добре видно графічно. Для порівняння на рис.4 зображено початковий запис, а на рис.5 – вже згладжений ковзним середнім.

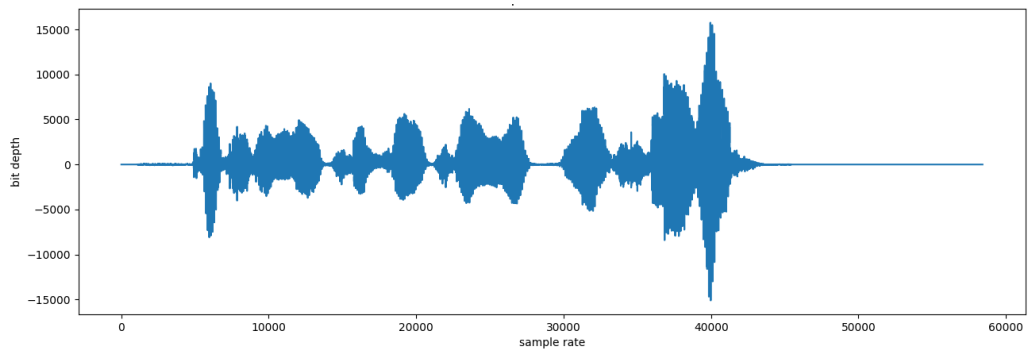


Рис.4. Початковий запис

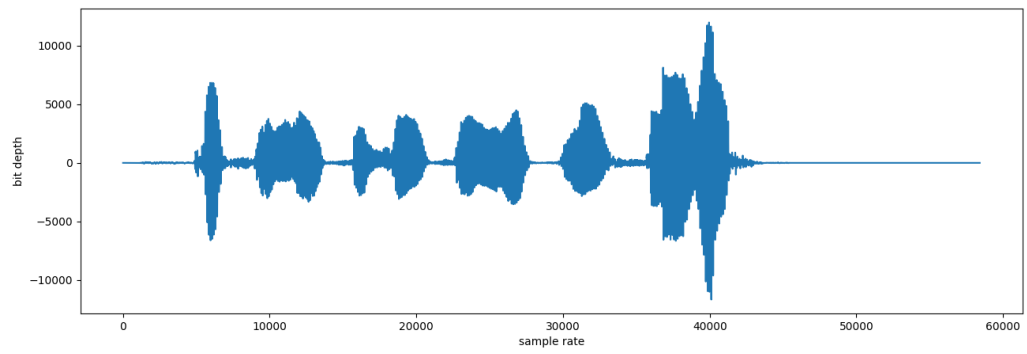


Рис.5. Запис, згладжений ковзним середнім

4. Опис програмного забезпечення та інформаційних технологій, що були розроблені

Для виконання роботи використовується мова програмування Python, оскільки вона містить усі необхідні засоби для роботи зі звуком і графічними представленнями тих або інших явищ. Імпортування готових модулів дещо сповільнює виконання програми, тому слід обходитися мінімально можливою їх кількістю. Але незначне сповільнення в даному випадку не є критичним, адже озвучування тексту в режимі реального часу не є першочерговою задачею.

Бібліотека `pytsx3` містить вбудовані голоси для прочитання тексту та інструменти для збереження озвучування у `wav`-файл. Модуль `scipy.io.wavfile` дає можливість зчитування та перезапису аудіофайлу. Для аналізу використовуваної в тексті лексики добре підходить `nlTK.tokenize` – модуль для обробки природної мови. За допомогою його засобів у роботі визначаються частини мови. Для математичних перетворень використовуємо `numpy`, а для графічної репрезентації записів або певних явищ – `matplotlib`. Цих стандартних бібліотек Python достатньо для повноцінної роботи з озвучуванням текстів і обробкою звуку.

На моменті розробки програма містить функцію побудови графіків з аудіо з указаними параметрами. Вона не бере безпосередньої участі в роботі з текстом, але значною мірою допомагає виявляти зміни й особливості результатів після застосування різних ефектів. Також із цією ж метою використовуються засоби Microsoft Excel.

Для, власне, роботи з тестом і звуком створено функцію, яка спочатку озвучує заданий текст із файлу, а потім перетворює його у масив значень і підсилює відповідні його фрагменти згідно з потрібною умовою. Це є основною функцією, яка потім посилається на інші. Також тут контролюється накладання кількох умов підсилення на один і той самий фрагмент. Наприклад, якщо питальне слово «коли» буде визначено як прислівник, то воно мало б підсилитися двічі – як питальне слово і як прислівник. Це може призвести до небажаного ефекту різкого перепаду значень і занадто сильного підвищення

гучності чи тону звуку. Тому варто перевіряти, чи підсилювалося слово вже за іншою ознакою, і в такому випадку не підсилювати його повторно (або підсилювати значно менше).

Обробка звуку відповідно до пунктуації в реченнях здійснюється за допомогою окремої функції. Вона розбиває текст на речення, визначає кінцевий розділовий знак у них і відповідно до цього задає необхідні параметри підсилення звукових характеристик.

Для знаходження емоційно забарвлених слів у програмі автором створено окрему функцію. Оскільки раніше зазначалося, що найбільш яскравих відтінків словам надають прислівники та прикметники, дана функція знаходить слова, що належать до цих частин мови, й визначає їх порядкові номери в реченні.

Крім розставлення акцентів на словах, які належать до певних частин мови, у природній мові людина інтонаційно увиразнює частину фрази, що містить заперечення. Звичним маркером таких слів у тексті є заперечна частка «не». Отже, слова після неї також варто емоційно підсилювати. Для знаходження таких слів у реченні створено функцію, яка визначає порядкові номери слів із заперечним відтінком значення.

Для пришвидшення запису використовуємо функцію, яка відкидає поодинокі значення з масиву з певним кроком. При не дуже малому кроці спотворення звуку майже не відчутні й виправляються в подальшому згладжуванням. Аналогічна функція для сповільнення запису, навпаки, додає елементи. Тобто так само з певним кроком дублює поодинокі значення.

Для зміщення тону звуку використовується функція, у основі якої лежить перетворення Фур'є. Це дозволяє зробити голос як нижчим, так і вищим при майже повному збереженні початкової тривалості запису. Ступінь погіршення якості звуку пропорційний довжині фрагментів аудіо, якими воно оброблюється. Чим менші шматочки запису зазнають перетворень, тим якіснішим залишається звучання.

Також додано функцію для виділення переважаючого діапазону значень звукової характеристики та знаходження середнього цього діапазону.

Для наочного представлення роботи з текстом та етапів обробки запису у додатках 18-20 наводяться основні блок-схеми даного програмного рішення. Інші функції досить прості й детально описані раніше, тому потреби в побудові блок-схем немає.

Отже, можна підсумувати, що завдяки даній програмі було систематизовано й практично застосовано особливості виникнення різних інтонацій. Рішення враховує як синтаксичні, так і лексичні особливості тексту, а в подальшому буде додано настроєві, тобто, власне, самі емоції, які виникають при прочитанні того чи іншого тексту. Що вдалося зробити краще, ніж було: чітка система технічних параметрів, які дозволяють реалізовувати відомі емоції і навіть моделювати можливі невідомі – вигадувати емоції шляхом перебору можливих комбінацій параметрів.

Висновки

1. Досягнуто підвищення виразності озвучування текстів за допомогою встановлених у ході досліджень залежностей між інтонацією та технічними характеристиками звукових сигналів для різних емоцій.

2. Удосконалено чисельний аналіз речень за допомогою методу ковзного середнього, ліній апроксимації та перетворення Фур'є; застосування виявлених особливостей для синтезу емоцій.

3. Практична значущість полягає в тому, що результати роботи допоможуть мінімізувати використання людського ресурсу при роботі з озвучуванням та надати комп'ютерному звуку більш «людських» інтонацій, що збільшує комфорт при спілкуванні людини із ботом..

4. Було виявлено, що на формування інтонації впливають декілька чинників: пунктуація та зміст тексту, емоційні переживання та психологічний стан мовця, а також цільова аудиторія слухачів, на яких спрямовується інформація.

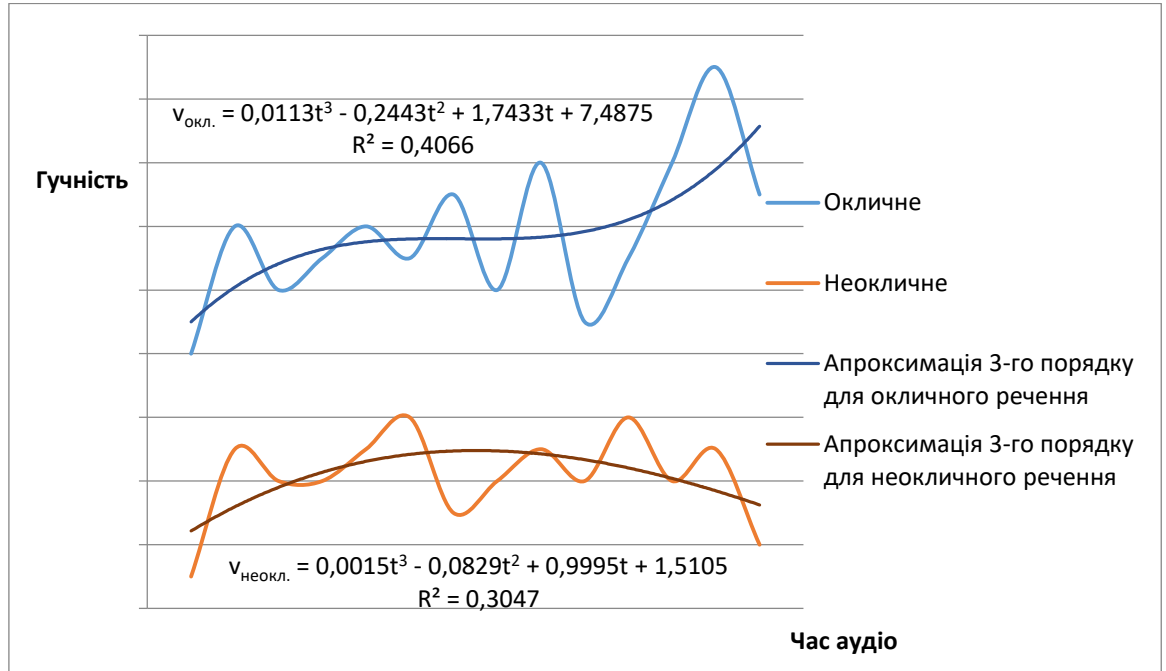
6. **Напрями подальших досліджень:** В подальших дослідженнях планується дослідити, як впливає на зміну інтонації використання пауз – як коротких у словах, так і більших у фразі; врахувати, що кінцевих розділових знаків може бути кілька: «?!», «!!!», «?..» тощо, через що інтонація може мати різні відтінки.

Література

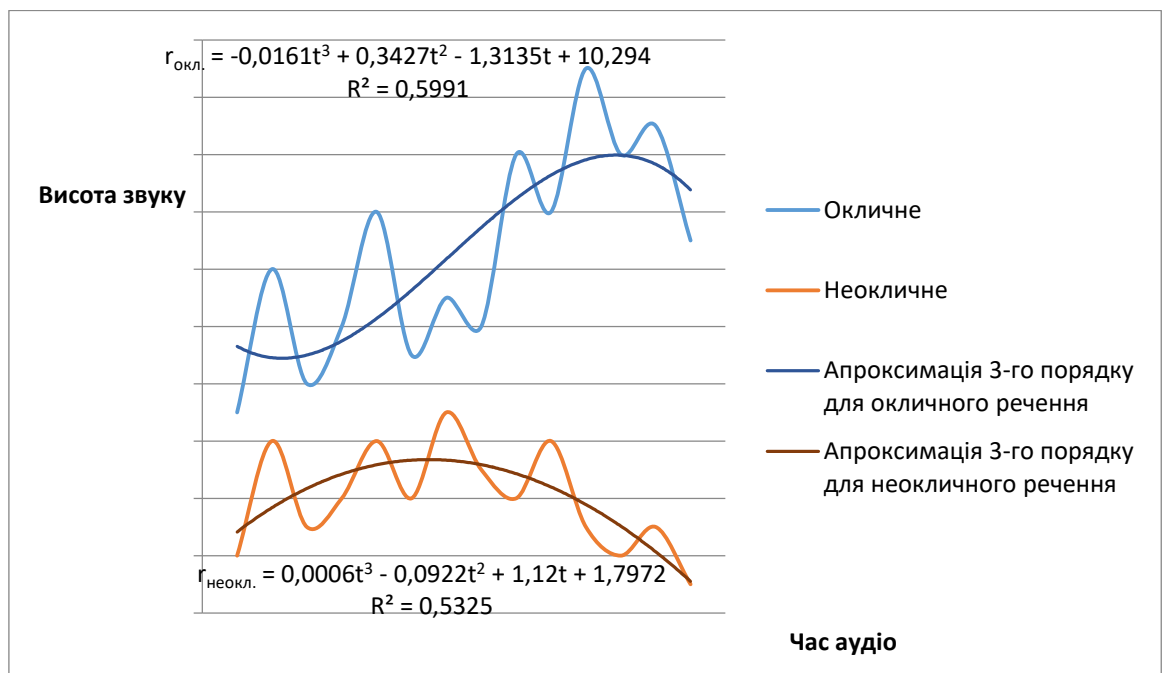
1. Волкова Н. Педагогіка: Навч. посіб. Вид. 2-ге, перероб., доп. — К.: Академвидав, 2007. — 616 с. (Альмаматер). ІЗБК 978-966-8226-47-2
2. Загуменнов, А. П. Компьютерная обработка звука [Электронный ресурс] / А. П. Загуменнов. - Москва : ДМК Пресс, 2006. - 384 с.: ил. - ISBN . - Текст : электронный. - URL: <https://znanium.com/catalog/product/407267>
3. Жовтобрюх, М. А. Курс сучасної української літературної мови : [у 2 ч.] / М. А. Жовтобрюх, Б. М. Кулик. - Київ : Рад. шк., 1965.
4. Feldman Barrett L. How Emotions Are Made: The Secret Life of the Brain/ Lisa Feldman Barrett. – Boston: Houghton Mifflin Harcourt, 2017. – 448p.
5. Сценическая речь: Учебник / Под ред. И. П. Козляниновой и И. Ю. Промптовой. 3-е изд. М.: Изя-во «ГИТИС». 2002. — 511 с.
6. Ораторське мистецтво: Навчальний посібник для студентів вищ. навч. закл. юрид. спец. Видання друге /Н.П. Осипова, В.Д. Воднік, Г.П. Клімова та ін. За ред. професора Н.П. Осипової. — Х.: Одиссей, 2006. — 144 с.
7. Р.Т. Міннігалімов Аналіз та синтез українського мовлення / Р.Т. Міннігалімов, Київ, 2015. – 90 с.
8. А. М. Пешковский Интонация и грамматика. – 1928
9. А. М. Пешковский Знаки препинания и научная грамматика. – 1918
10. А.Й. Багмут Семантика і інтонація в українській мові. – 1991
11. А.Й. Багмут Інтонаційна будова простого розповідного речення в слов'янських мовах. – 1970
12. А.Й. Багмут Структура і функціонально-семантичний аспект інтонації простого розповідного речення у слов'янських мовах. – 1980
13. Л. А. Близниченко Інтонація мовлення. – 1968
14. Офіційний сайт AT&T Laboratories: https://about.att.com/sites/labs_research
15. Офіційний сайт Sonantic: <https://www.sonantic.io/>

Додатки

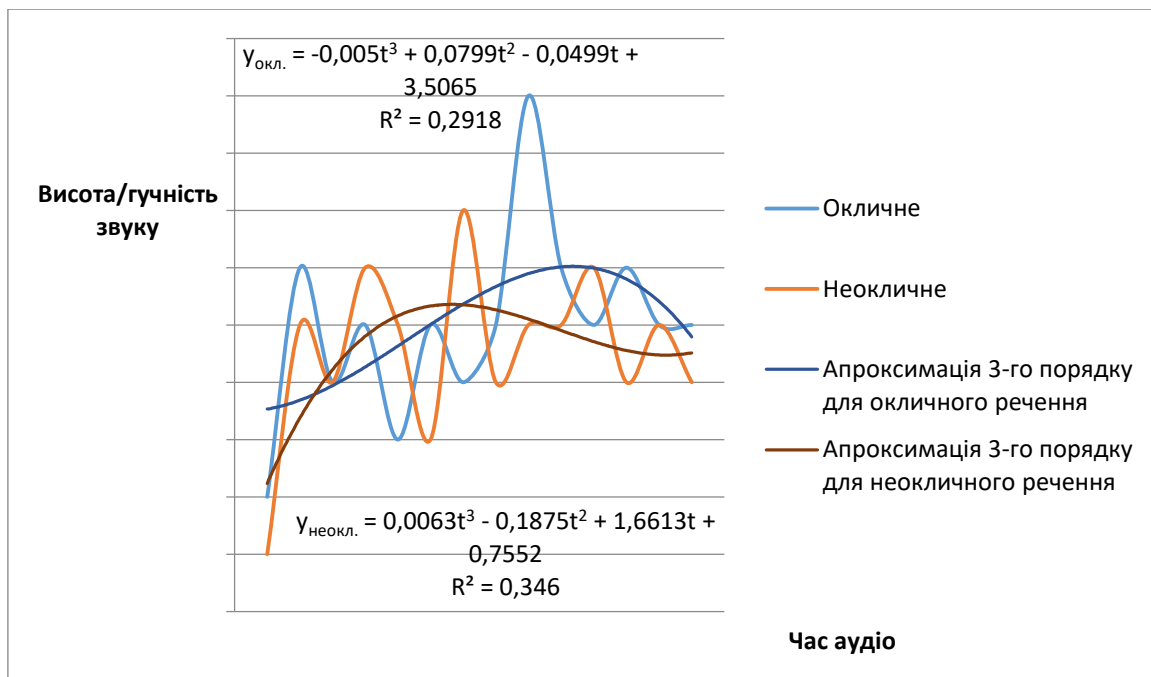
Додаток 1. Зміна гучності окличних і неокличних речень



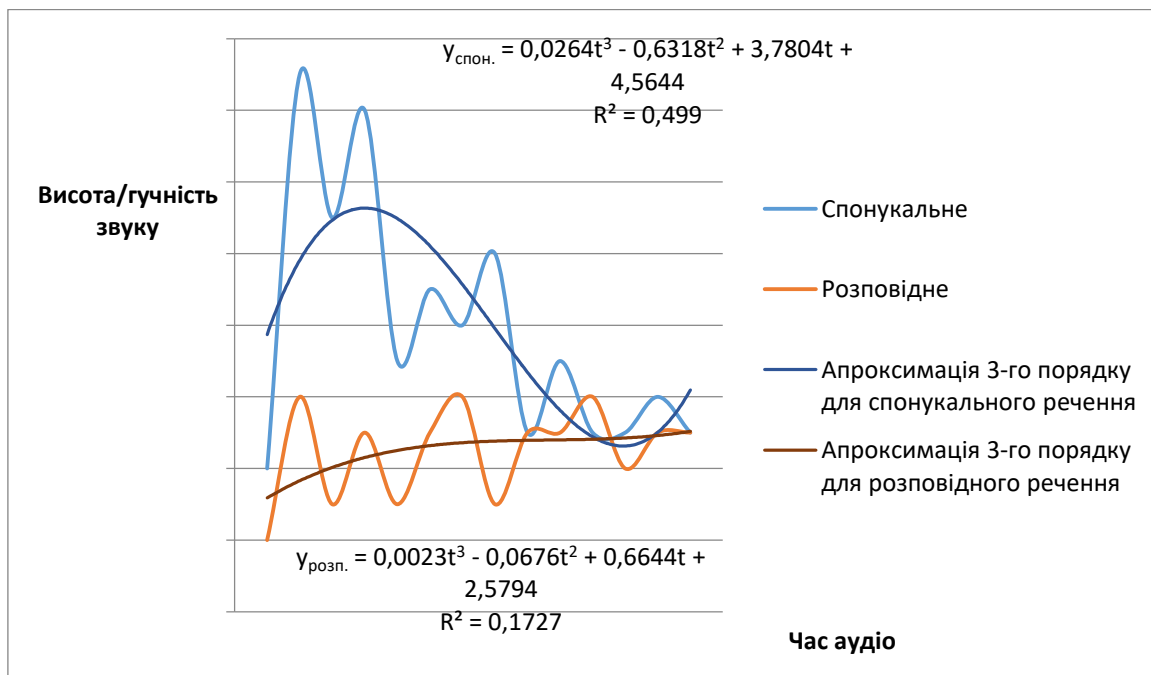
Додаток 2. Зміна висоти звуку окличних і неокличних речень



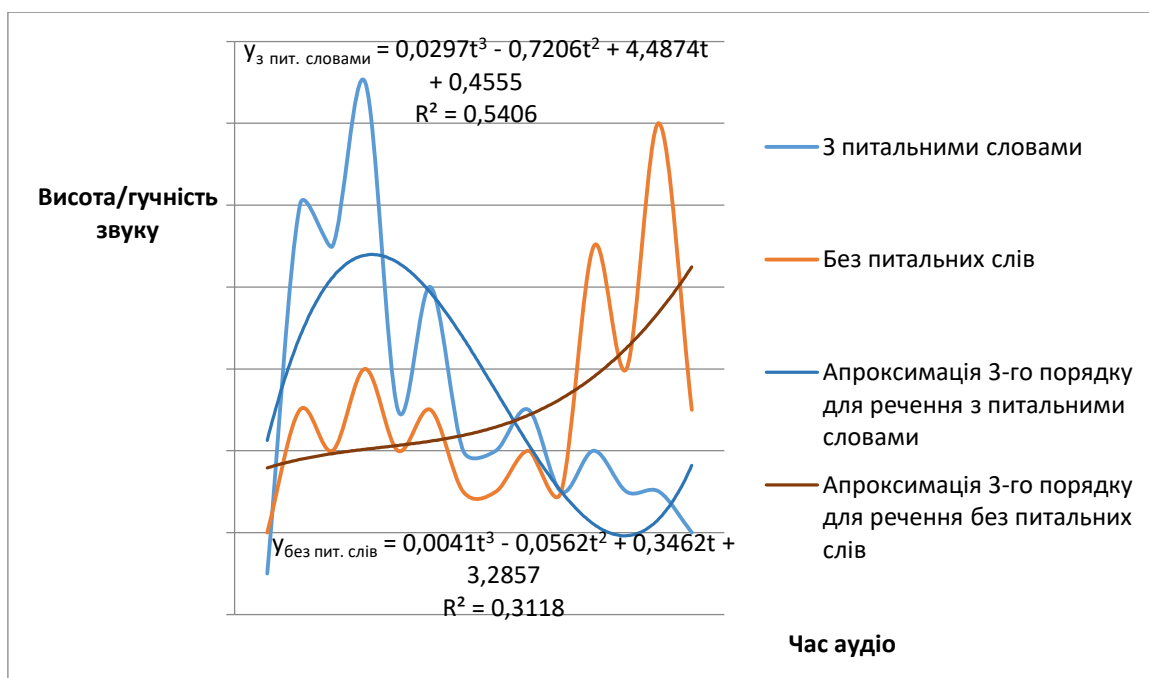
Додаток 3. Зміна звукових характеристик речення з трьома крапками



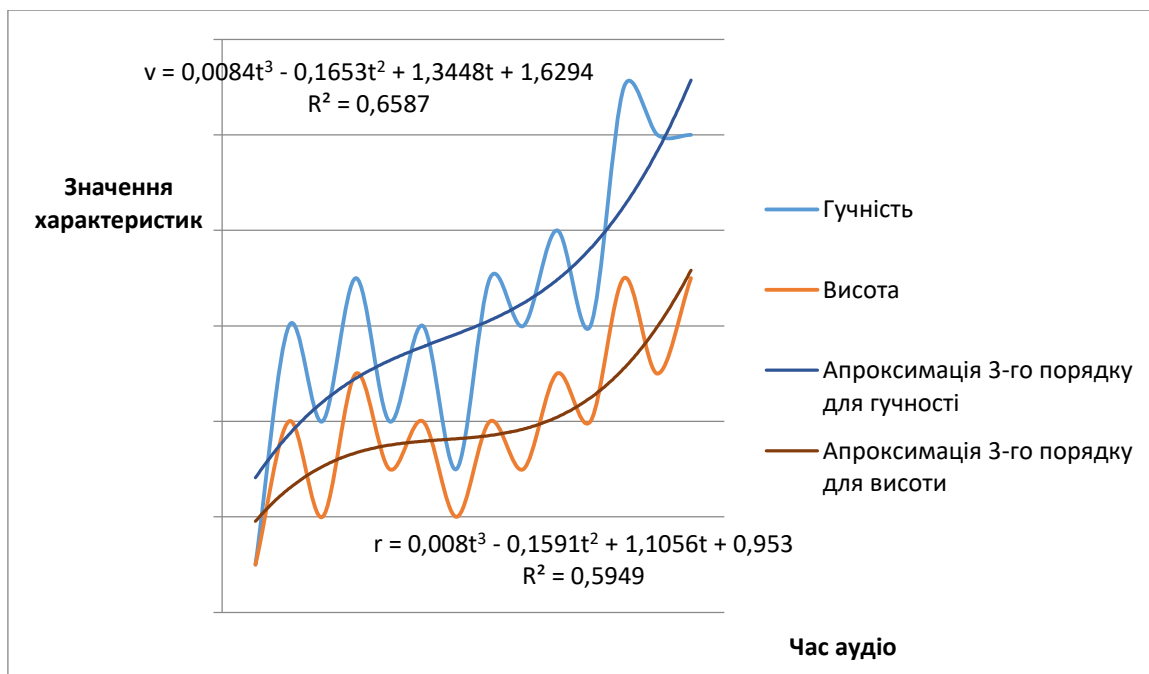
Додаток 4. Зміна звукових характеристик розповідного та спонукального речення



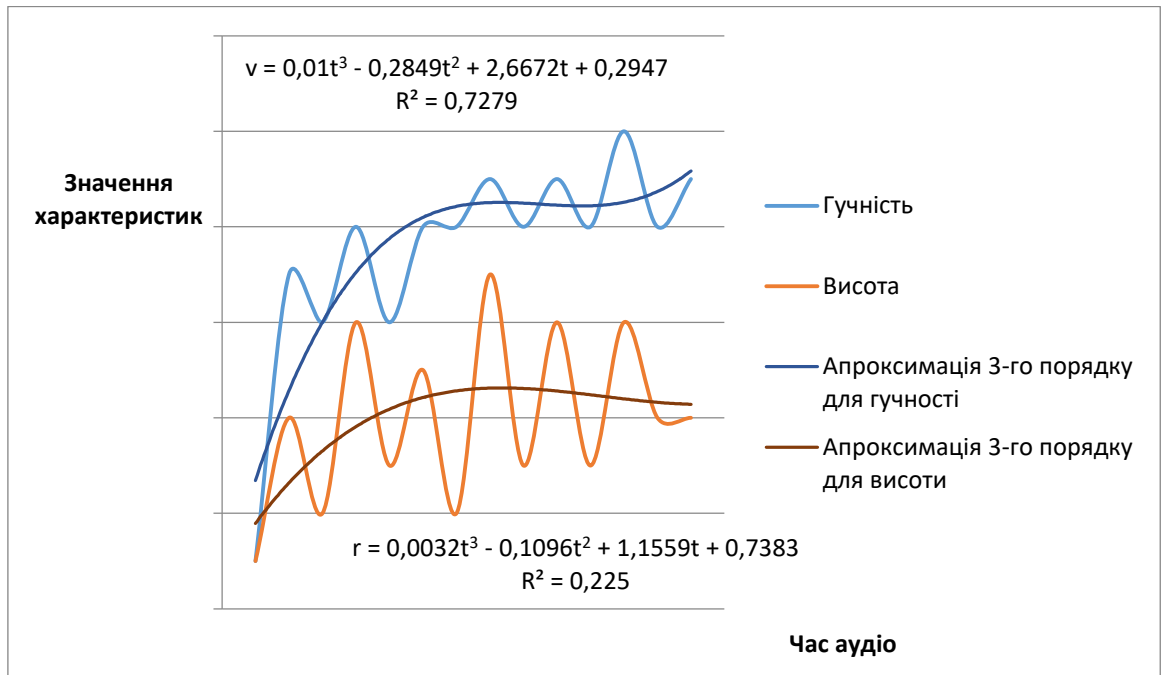
Додаток 5. Зміна звукових характеристик питальних речень



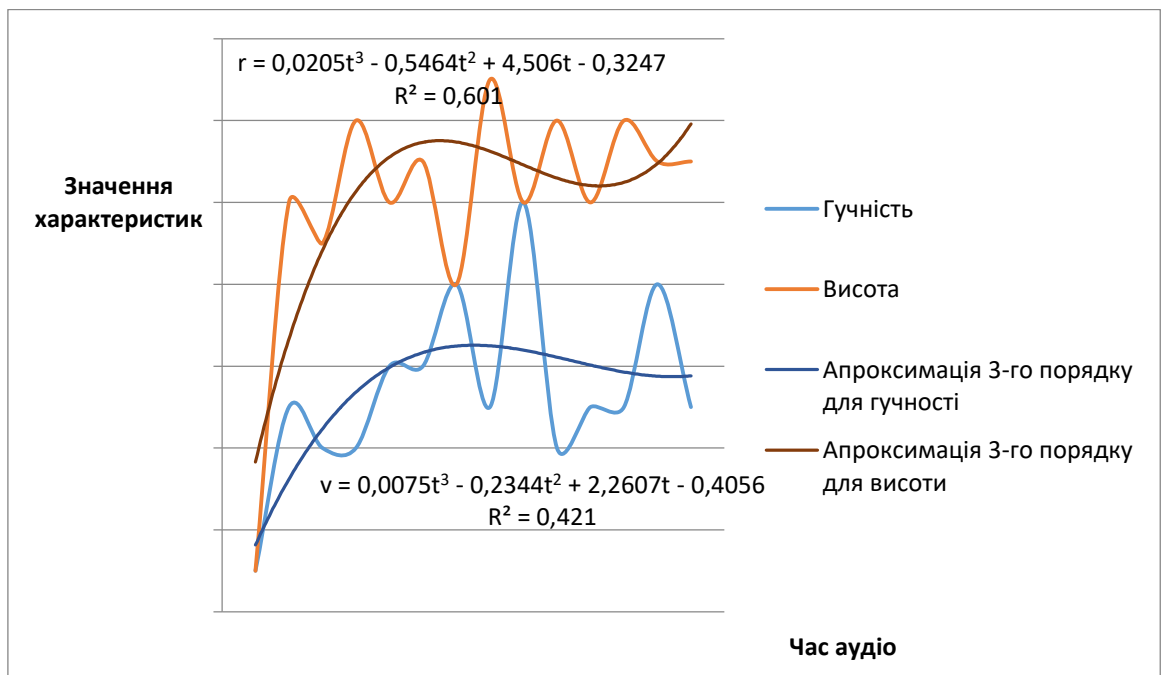
Додаток 6. Зміна звукових характеристик у часі для позитивних емоцій



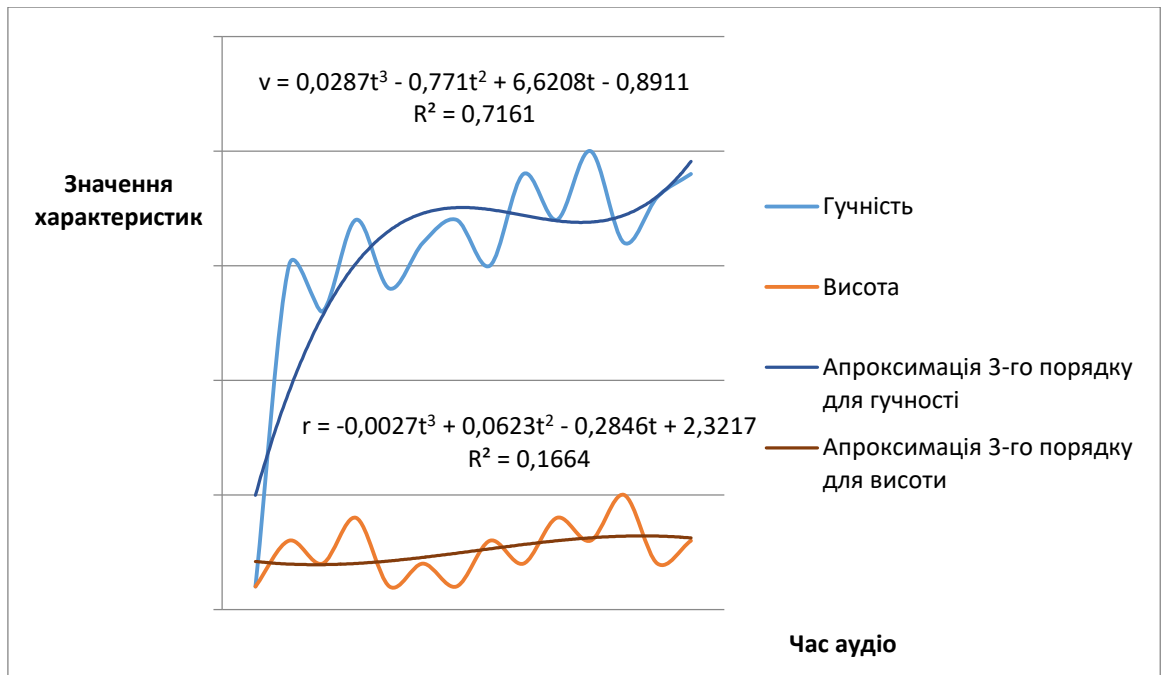
Додаток 7. Зміна звукових характеристик при роздратуванні чи невдоволенні



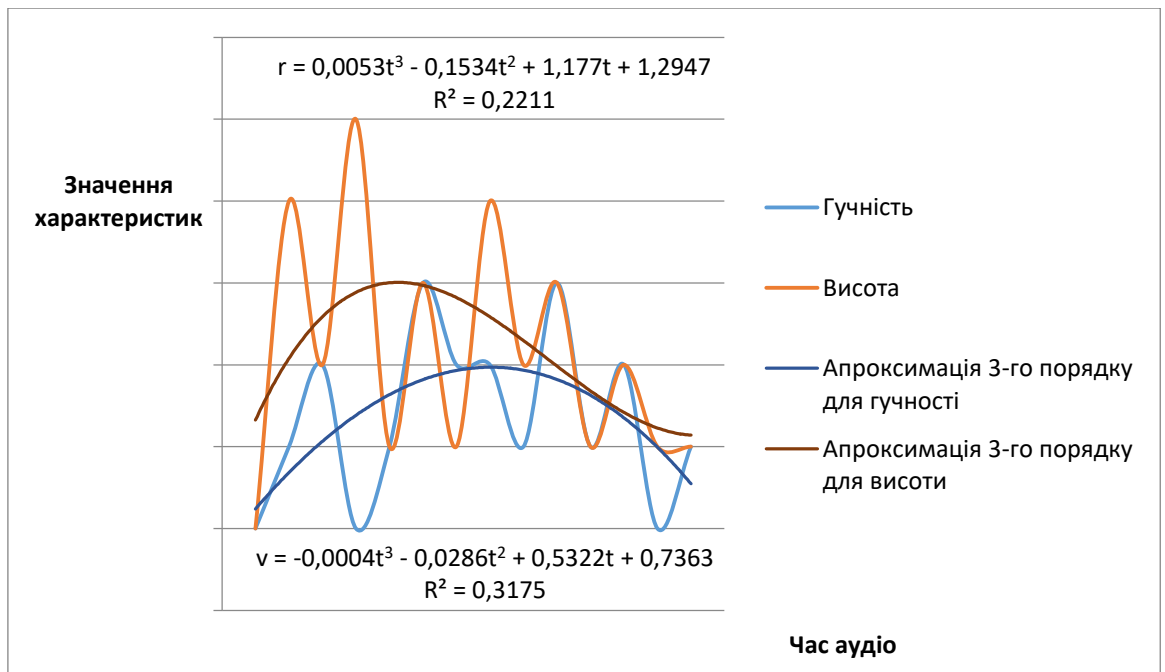
Додаток 8. Зміна звукових характеристик при відчутті страху чи тривоги



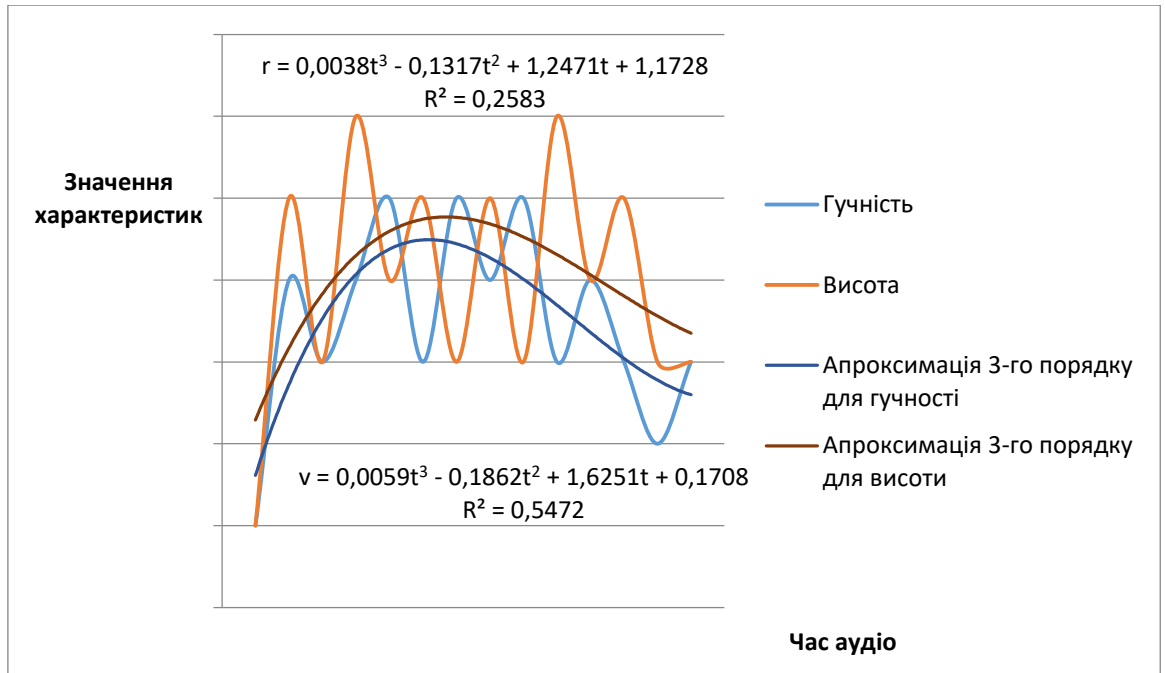
Додаток 9. Зміна звукових характеристик при відчутті агресії



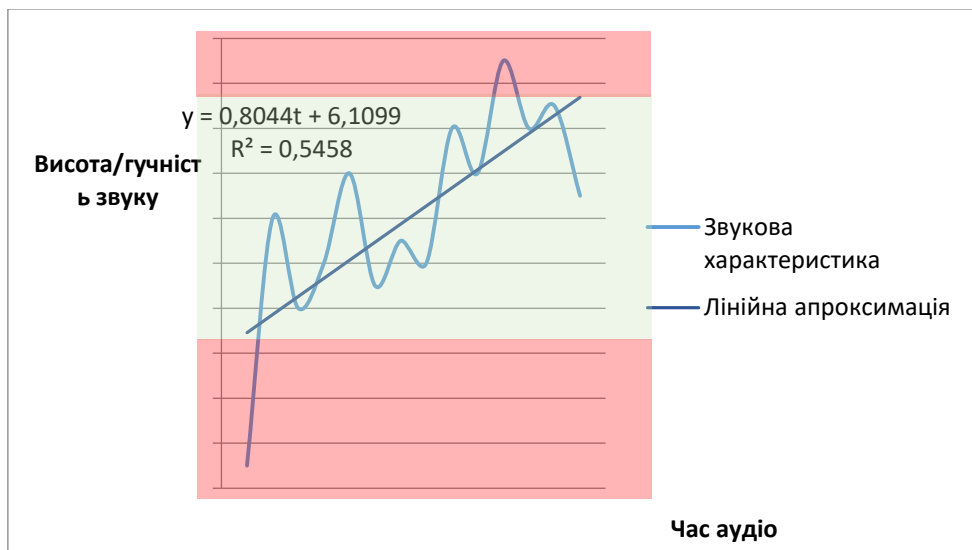
Додаток 10. Зміна звукових характеристик при депресивному стані



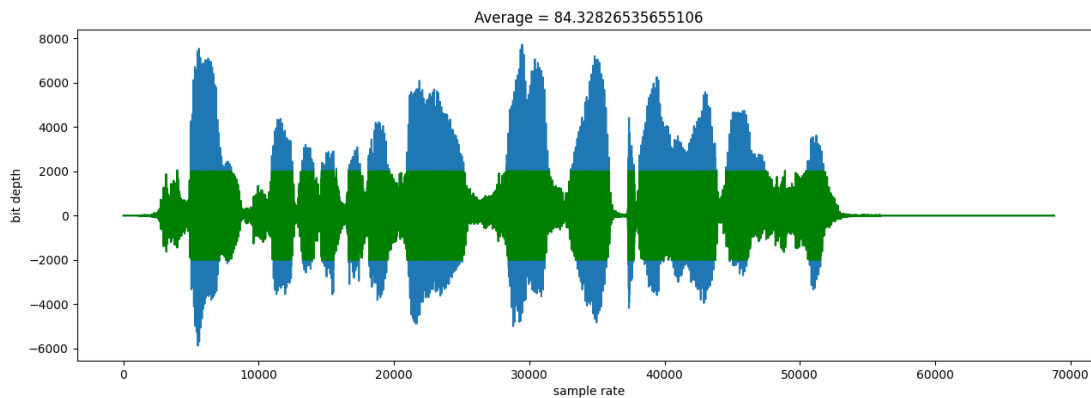
Додаток 11. Зміна звукових характеристик у стані спокою й комфорту



Додаток 12. Вибір середнього діапазону значень за допомогою лінійного апроксимування



Додаток 13. Вибір діапазону значень для усереднення



Додаток 14. Підсилення звукових характеристик із використанням множення на коефіцієнт

```
#first syllable is changed a little just to mark that the difference between volumes goes further
for j in range((q_len - 3)*n, (q_len - 2)*n):
    brr[j] *= coef_list[0] #--> volume increased by 4 times!
#second syllable makes the major difference in generating intonations, so should be emphasised the most
for j in range((q_len - 2)*n, (q_len - 1)*n):
    brr[j] *= coef_list[1] #--> volume increased by 7 times!
#third syllable also takes part in making intonations and in smoothing the volumes
for j in range((q_len - 1)*n, q_len*n):
    brr[j] *= coef_list[2] #--> volume increased by 3 times!
```

Додаток 15. Підсилення звукових характеристик із використанням додавання сталого значення

```
for j in range((q_len - 3)*n, (q_len - 2)*n):
    brr[j] += coef_list[0] #--> volume became a little bit louder

for j in range((q_len - 2)*n, (q_len - 1)*n):
    brr[j] += coef_list[1] #--> volume smoothly became louder a little more

for j in range((q_len - 1)*n, q_len*n):
    brr[j] += coef_list[2] #--> volume smoothly decreased in the end
```

Додаток 16. Вибір середнього діапазону значень звукової характеристики

```
# y - original array
x = np.arange(0, len(y), 1)
plt.figure(figsize=(20, 5))
plt.xlabel("sample rate")
plt.ylabel("bit depth")
plt.plot(x,y)
y = np.where(y > 2000, 2000, y)
y = np.where(y < -2000, -2000, y)
plt.plot(x,y, 'g')
m = "Average = " + str(np.mean(y))
plt.title(m)
plt.show()
return np.average(y)
```

Додаток 17. Згладжування запису метод ковзного середнього

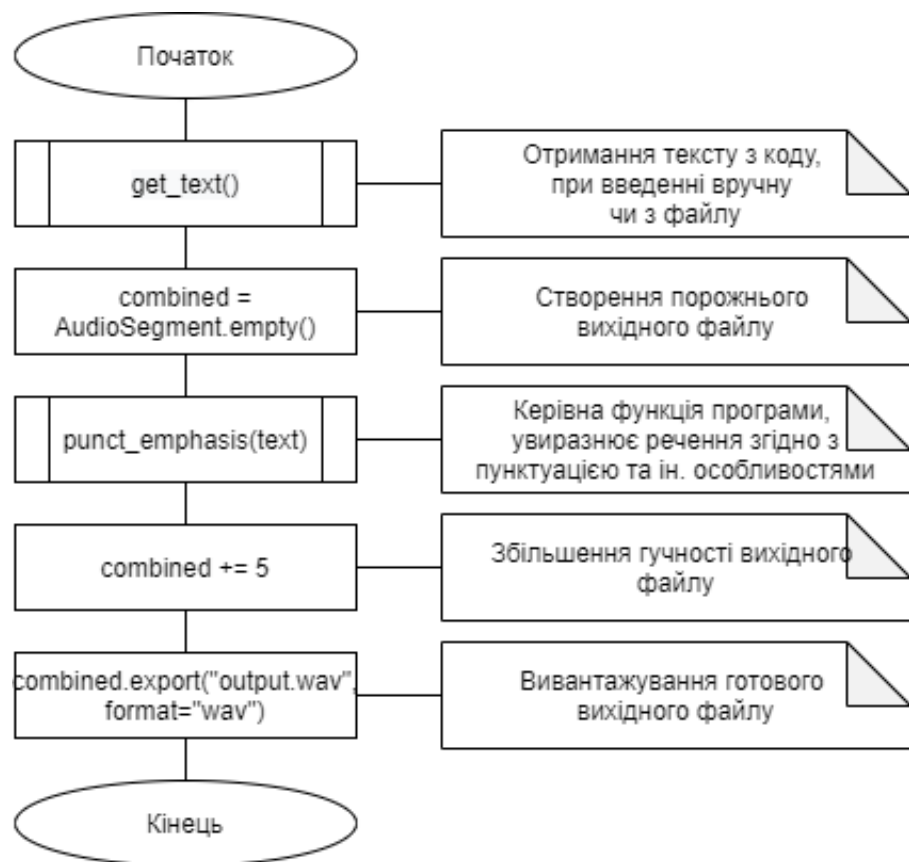
```
def moving_avg(array, window, step):
    new_arr = array.copy()

    for i in range(1 + window//2, len(array) - window//2):
        sum = 0

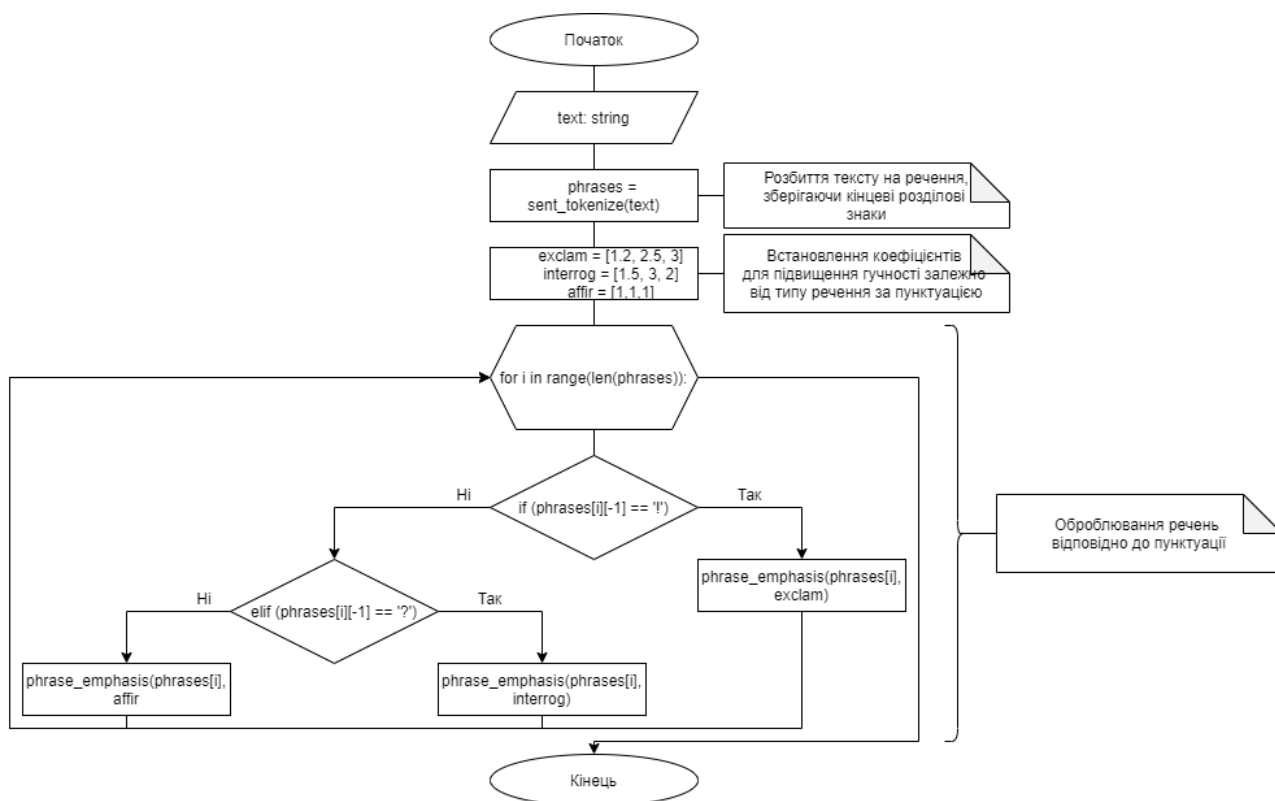
        if (i%step == 0):
            for j in range(i - window//2, i + window//2 + 1):
                sum += array[j]
            avg = sum / ((window//2)*2 + 1)
            new_arr[i] = round(avg)

    return new_arr
```

Додаток 18. Блок-схема основної програми



Додаток 19. Блок-схема функції, що оброблює речення відповідно до пунктуації



Додаток 20. Блок-схема функції, що оброблює текст за лексичними особливостями

