

ЗМІСТ

Вступ	4
1 Аналітичний огляд алгоритмів обчислення частоти основного тону та вокальних виконань	6
1.1 Роль частоти основного тону у дослідженні мовлення.....	6
1.2 Алгоритми аналізу частоти основного тону	10
1.3 Дослідження вокальних виконань.....	13
2 Розпізнавання нот у вокальному виконанні	16
2.1 Параметри сегментації та ідентифікації нот	16
2.2 Застосування математичної моделі слухової системи людини.....	20
2.3 Алгоритм розпізнавання нот у вокальному виконанні	27
2.4 Програмний комплекс по визначення нот вокального виконання.....	28
3 Оцінка точності роботи алгоритму розпізнавання нот у вокальному виконанні	33
3.1 Експеримент із розпізнавання нот у вокальному виконанні	33
3.2 Визначення частоти помилок алгоритму розпізнавання нот	37
3.3 Перевірка коректності експертних оцінок	40
4 Аналіз виконання нот із різкою зміною частот основного тону.....	43
4.1 Вплив різкої зміни частот на точність роботи алгоритму	43
4.2 Аналіз вібратороподібних вокальних виконань	47
4.3 Аналіз глісандуючих переходів у співі.....	50
4.4 Порівняння ковзного середнього та лінійного тренду на ділянках з різкою зміною частот основного тону	52
4.5 Експериментальне тестування.....	54
Висновок	62
Перелік джерел посилань	63

ВСТУП

Актуальність теми. Застосування програмних засобів під час навчання поширене у багатьох галузях знань. Однак у сфері навчання вокальній майстерності існуючі програми не завжди точно визначають заспівану виконавцем ноту. Викладач, здійснюючи заняття у класі, по черзі здійснює прослуховування кожного з учнів та дає рекомендації. За такого підходу на початковому етапі навчання проходить тривалий період розвитку музичного слуху. Ефект біологічної зворотний зв'язок корисний у сфері навчання музикантів, що підводить до необхідності розробки системи розпізнавання нот з високою точністю. Існуючі комерційні програми, що реалізують функцію навчання співу, відрізняються низькою точністю розпізнавання заспіваних виконавцем нот.

Дослідженням вокалізованої мови та сигналів, близьким за структурою до співу, займалися такі вчені як М.В. Федотова, Є.В. Жаровська, В.П. Морозов, С.В. Шук, М.С. Агін, С. Лейдон, Дж. Бауер, С. Ларсон, Е.Б. Абдуллін та ін. Як ключова інформація, що використовується в даних дослідженнях, застосовується значення частоти основного тону. Основний тон містить у собі інформацію про інтонаційну структуру вимови, індивідуальність голосу диктора та його емоційний стан, вікові та патологічні зміни голосового апарату. Існуючі алгоритми не дозволяють обчислити значення фундаментальної частоти у вокальному виконанні з високою точністю за рахунок наявності високого відсотка грубих помилок у них та обмежені вузьким спектром частот, що охоплюються. Більшість алгоритмів розроблялися з метою аналізу мовної інформації, що накладає обмеження у вигляді верхньої межі визначення ЧОП, що дорівнює 400 Гц. Однак, під час співу частота звучання мовного сигналу може бути набагато вищою, що робить незастосовними алгоритми, обмежені діапазоном

обробки мови. Також не застосовні алгоритми, що володіють високим відсотком грубих помилок, для ідентифікації ноти, що звучить. Помилка в частоті близько 20% її значення може призвести до промаху більш ніж на 3 ноти. Наявність таких обмежень унеможливорює існуючі рішення щодо ідентифікації нот у задачі навчання вокалу за допомогою програмних засобів.

Метою магістерської роботи є розробка методу ідентифікації вокалізованих ділянок із різкою зміною частоти основного тону сигналу. Ця мета спрямована на підвищення якості роботи програмного комплексу розпізнавання нот у вокальному виконанні.

Для досягнення цієї мети необхідно було вирішити наступне завдання :

- оцінити точність роботи алгоритму розпізнавання нот на аудіозаписах з вокальним виконанням;
- дослідити тимчасові ряди, отримані з аудіозаписів вокального виконання із застосуванням вібрато та глісандо у співі;
- визначити закономірності поведінки оцінок частот основного тону при досліджуваних вокальних техніках;
- вирахувати порогові значення для ідентифікації вібрато та глісандо за оцінками тимчасових рядів.

Об'єктом дослідження даної є мовний сигнал вокального виконання з різкою зміною частот основного тону.

Предметом дослідження є розпізнавання ефектів вібрато та глісандо на основі частоти основного тону вокального виконання.

Методи дослідження . Для вирішення завдань, сформульованих у роботі, використовувалися методи моделювання, системного аналізу, цифрової обробки сигналів, математичної статистики.

1 АНАЛІТИЧНИЙ ОГЛЯД АЛГОРИТМІВ ОБЧИСЛЕННЯ ЧАСТОТИ ОСНОВНОГО ТОНУ ТА ВОКАЛЬНИХ ВИКОНАНЬ

1.1 Роль частоти основного тону у дослідженні мовлення

Знання значення частоти основного тону (ЧОП) сигналу в конкретний час має важливе значення у багатьох сферах мовних технологій. Так, наприклад, визначення основної частоти мовного сигналу з урахуванням особливостей формування мови та сприйняття мови, пов'язаних з анатомією та фізіологією людини, вкрай важливе у сфері реабілітації для онкологічних хворих після резекції гортані [1]. У таких хворих часто виникають побічні ефекти, що ускладнюють розмову та спілкування. Під час логопедичної терапії широко використовують підхід перцептивної оцінки якості голосу. Ряд досліджень спрямовано застосування програмних засобів з оцінки голосу у межах реабілітації [2-4].

Спів може бути розглянутий як особлива форма мови, що створюється таким же чином, але при цьому є додатковий контроль для створення музичного аспекту. Природна мелодія мови (просодія) відрізняється у різних мовах і визначає контур висоти тону, варіації гучності, ритм та темп вираження емоцій. У співі ж висоту, гучність і тембр визначає насамперед композиція за рахунок того, що для відповідності тривалості ноти голосні звучать довше, ніж зазвичай. Мовні голосні характеризуються особливими позиціями формант. У співі положення формант може бути радикально змінено шляхом зміни довжини та форми голосового тракту та положення артикуляторів. Ідентичність голосу багато чому визначається фізичними характеристиками системи виробництва вокалу. Форма голосового тракту визначає форманти, причому дві форманти нижчого порядку (F_1 - F_2) є найважливішими для розбірливості мови, а форманти вищого порядку (F_3 - F_5) сприяють ідентифікації того, хто

говорить. Досвідчені співаки можуть точно контролювати частоти трьох нижніх формантів, варіюючи першу форманту за допомогою ступеня розкриття щелепи, другий - управлінням формою мови і третім положенням кінчика мови. Різні вокалісти налаштовують свої частоти формант по-різному для кожного голосного, причому найбільш яскраві відмінності полягають у жіночих голосах сопрано, де високі значення висоти звуку (1000 Гц) у порівнянні зі звичайним значенням першої форманти (500 Гц) будуть визначати зсувом цієї форманти близько до ЧОТ. Це може призвести до втрати розбірливості, але при класичному співі інтонація та музичні якості голосу є найважливішим аспектом, а розбірливість – другим.

Визначення особливостей у частоті основного тону мовного сигналу є важливим завданням у сфері дослідження особливостей мови. Для деяких мов і акцентів характерна наявність появи висхідного або низхідного тону, який за своїми характеристиками може бути схожим на вокальне виконання. У роботі [5] розглядається можливість застосування виявлення в мові висхідно-низхідного тону як маркер для прийняття рішення про наявність валлійського акценту у диктора. Авторами оцінювалися інтонації на підставі підвищення чи зниження голосу. Отриманий розкид ЧОП був отриманий у діапазоні від 169 до 358 Гц. Дослідження [6] торкнулося питання визначення особливостей у мові молоді. На підставі оцінки ЧОП записів мовних сигналів було визначено, що для молоді характерне навмисне розтягування голосних. Автором [7] проводилося дослідження висотно-мелодійного параметра мови, саме вплив вікових змін на ритмічні характеристики мови. Було визначено, що зі збільшенням віку диктора проявляється зниження максимальних та мінімальних значень ЧОП. Однією з особливостей публікації [8] є визначення на підставі ідентифікації у мовленні характеристик форми мелодії емоційного стану того, хто говорить. Автор звернув увагу на такі характеристики як напрямок, характер та діапазон руху тону. У цьому дослідженні застосовувалися у

вимірі ЧОТ величини, використовувані для дослідження музичні мелодії. Це дозволило наочно уявити особливості змін у мові диктора у разі прояву емоцій чи намірів. Аналогічна задача була поставлена в роботі [9], де за зміною різниці в акустичних значеннях середньої та максимальної ЧОП фрази визначалася оцінність у мові репортерів з метою передачі атмосфери того, що відбувається. Ряд досліджень [10-13], мають на меті створення систем пошуку музичної інформації (Music Information Retrieval – MIR). Вченими переслідуються такі цілі як розробка автоматизованої системи для визначення плагіату між двома уривками музичних відтворень і створення засобів для зручного пошуку музики за критеріями, що задаються. Перед системами, що розробляються, ставляться такі завдання як виявлення співу в музичному творі, класифікація жанру, ідентифікація співака та багато інших. Перетворення композицій на набір векторів ними здійснювався на підставі прийнятої в європейській музичній традиції ідентифікації нот щодо абсолютної початкової частоти. Як зазначено в [14], при обробці музики виникає складність із наявністю гармонік у деяких інструментів. Крім ЧОТ основної ноти, звучать інші частоти, які можуть відповідати іншим нотам. Крім того, у музиці спектр звукового сигналу найчастіше є сумою спектрів окремих інструментів, де кожним із інструментів відтворюються звуки у різних смугах частот.

Огляд досліджень з обробки музики показав, що з погляду обробки сигналу мелодія може бути представлена послідовностями ЧОП, що визначаються в моменти звучання, тобто на ділянках, де активний інструмент, що створює мелодію. Основні алгоритми транскрипції мелодії зазвичай містять два етапи обробки. По-перше, обчислюється уявлення, що підкреслює найімовірніші значення ЧОП у часі, наприклад, у формі матриці видачі, матриці активації джерела голосу або розширеної спектрограми. По-друге, двійкова класифікація обраних ЧОП між мелодичним та фоновим змістом виконується з використанням виявлення мелодійного контуру та

голосу. Наприклад, у [15] підхід до виявлення мелодійних та басових ліній у звуковому сигналі заснований на оцінці відносного домінування кожної можливої ЧОП у вигляді функції щільності ймовірності ЧОП. Проведений вченими експеримент показав, що система, що використовується в реальному часі обробку аудіозапису, змогла виявити необхідну інформацію тільки на 80% ділянках від всієї тривалості протестованих записів.

Таким чином, на основі розуміння того, яку ноту співак співав або музичний інструмент, є знання про те, що на якійсь фундаментальній частоті коливання відбуваються в даний момент часу. Для кожної ноти є дискретне значення частоти (таблиця 1.1) [16-17].

Таблиця 1.1 - Відповідність нот частотам основного тону

Нота	Октава								
	Субконтрокту ва	Контрокта ва	Велик а октава	Мала 1 октава	2 октава	3 октави	4 октави	5 октава	
До	-	32.70	65.41	130.82	261.63	523.25	1046.50	2093.00	4186.00
ДО-гострий	-	34.65	69.30	138.59	277.18	554.36	1108.70	2217.40	4434.80
Ре	-	36.95	73.91	147.83	293.66	587.32	1174.60	2349.20	4698.40
РЕ-гострий	-	38.88	77.78	155.56	311.13	622.26	1244.50	2489.00	4978.00
Ми	20.61	41.21	82.41	164.81	329.63	659.26	1318.50	2637.00	5274.00
Фа	21.82	43.65	87.31	174.62	349.23	698.46	1396.90	2793.80	-
Фа-діз	23.12	46.25	92.50	185.00	369.99	739.98	1480.00	2960.00	-

Сіль	24.50	49.00	98.00	196.0	392.0	784.0	1568.0	3136.0	-
				0	0	0	0	0	
Соль-дієз	25.95	51.90	103.8	207.0	415.3	830.6	1661.2	3332.4	-
			0	0	0	0	0	0	
Ля	27.50	55.00	110.0	220.0	440.0	880.0	1720.0	3440.0	-
			0	0	0	0	0	0	
Сі-бемоль	29.13	58.26	116.5	233.0	466.1	932.3	1864.6	3729.2	-
			4	8	6	2	0	0	
Сі	30.87	61.74	123.4	246.9	493.8	987.7	1975.5	3951.0	-
			8	6	8	5	0	0	

Крім того, в теорії музики [18] відоме поділ співочих голосів в залежності від виконуваних нот. Женские голоса включают в себя диапазон от ноты «фа малой октавы» (174.62 Гц), соответствующий самому низкому женскому певческому голосу – контральто, до ноты «соль-диез третьей октавы» (1661.20 Гц), соответствующий самому высокому женскому певческому голосу – колоратурному сопрано. Чоловічі голоси в діапазоні від ноты «фа contr octave» (43,65 Гц), відповідний нижньої межі бас-профундо - з низькою чоловічого співочого голосу, в записці

«ми другої октави» (659.26 Гц), що відповідає верхній межі найвищого чоловічого співацького голосу – контратенора. В результаті можна виділити діапазон частот, що охоплюються співом професійними оперними співаками – відрізок від 43.65 Гц до 1661.2 Гц. Однак, ґрунтуючись на даних про найпоширеніші чоловіче (баритон) та жіноче (ліричне сопрано) [19] співочі голоси, досліджуваний діапазон можна обмежити відрізком від 110 до 1318.5 Гц.

1.2 Алгоритми аналізу частоти основного тону

Більшість робіт, присвячених алгоритмам обчислення значень ЧОП сигналу у певний момент часу, спрямовані на вивчення мови. Прикладами

таких досліджень можуть бути [20-21]. При цьому переважна більшість алгоритмів визначення частоти основного тону використовують обчислення періоду основного тону з піків мовного сигналу [22-23]. Основним інтересом для роботи [24] є розробка системи корекції слуху, яка може пояснити обмеження, що застосовуються до алгоритму відстеження миттєвих значень Choate (Instantaneous Robust Algorithm for Pitch Tracking).

IRAPT діапазоном від 50 до 500 Гц. В основі запропонованого ними алгоритму є застосування алгоритму RAPT [25]. У ході експерименту цій групі вчених вдалося досягти для розробленого алгоритму відсотка грубих помилок (gross pitch error – GPE) 1.625% для чоловічого голосу та 3.777% для жіночого, а також середнього відсотка дрібних помилок (mean fine pitch error – MFPE) 1.608% для чоловічого голосу та 0.977% для жіночого. В іншій роботі [26] ці ж автори пропонували інший алгоритм (оцінка миттєвої ЧОП на основі багатошвидкісної обробки), де діапазон пошуку частоти основного тону також був обмежений - від 100 до 450 Гц, що позначилося на відсотку помилок. У новій версії алгоритму GPE становив 0.743% для чоловічого голосу та 3.6% для жіночого, а MFPE – 1.268% для чоловічого голосу та 1.039% для жіночого. У дослідженні [27] було запропоновано алгоритм PEFAC (Pitch Frequency Estimation Algorithm Robust to High Levels of Noise), що дозволяє оцінювати значення ЧОТ за умов високого шуму з високою точністю. Цей алгоритм також був вивчений лише на діапазоні до 400 Гц. З урахуванням специфіки даного алгоритму, слід зазначити його показник вокалізованості (Voicing Decision Error, VDE), який визначає кількість правильно розпізнаних вокалізованих кадрів, що дорівнює в середньому 46.45%, що перевищує точність роботи аналогів. При цьому GPE для алгоритму різних типів шумів становить 40% при бульканні, 20% при шумі машин і 24% при білому шумі. Ще одним із алгоритмів, що дозволяє виділити в сигналі значення частоти для основного тону, є SWIPE (Sawtooth Inspired Pitch Estimator) [28]. Як впливає з назви алгоритму,

аналіз відбувається за спектром пилкоподібного сигналу. Алгоритм є інтегральним та показує хороші результати в умовах низького шуму. Верхнє допустиме значення ЧОП для алгоритму SWIPE – 500 Гц. Випробування алгоритму показали величини помилок GPE 32.92% та MFPE 4.45% при частотах близьких до 500 Гц, що може бути спричинено особливістю реакції алгоритму на наявність субгармонік. Помилки 1-го та 2-го роду при цьому становлять 2.77% та 18.41% відповідно. Серед інших варто виділити алгоритм YIN [29], який отримав свою назву від одного з елементів філософії – інь. Алгоритм заснований на використанні автокореляційного методу і, як зазначають автори, не має верхньої межі для визначення ЧОП сигналу та має відсоток грубих помилок, що дорівнює 1.03%. З цієї причини автори запропонувати алгоритм для використання НЕ тільки для мови, але і для музичних сигналів. Проте, показаний [26] порівняльний аналіз перерахованих вище алгоритмів показав, що при обробці сигналів з частотами основного тону в діапазоні від 100 до 350 Гц, GPE і MFPE у алгоритму YIN вище, ніж у аналогів і становлять 3.96% і 1.389% відповідно.

Значення вищевказаних досліджень GPE показує відношення аналізованого кадру з оцінкою відхилення отриманого Choate більше ніж на $\pm 20\%$ від теперішнього значення Choate до загальної кількості озвучених кадрів [30] і оціненого за формулою 1.1.

$$GPE(\%) = \frac{N_{GPE}}{N_v} \cdot 100, \quad (1.1)$$

N_v

де N_{GPE} – кількість кадрів з відхиленням отриманої оцінки більш ніж на $\pm 20\%$ від значення основного тону;

N_v – загальна кількість вокалізованих кадрів.

Як вказується в [31], похибка в межах 20% варіюється в межах октави, а розумність застосування алгоритмів з такою кількістю помилок при оцінюванні ЧОП визначається виключно кінцевою метою роботи. Однак у

питаннях, що стосуються ідентифікації нот, ця помилка призводить до зміщення в результаті не менше ніж на 3 ноти.

Оцінка MFPE, що показує середній відсоток дрібних помилок, що з'являються під час оцінювання ЧОП без урахування грубих помилок.

1.3 Дослідження вокальних виконань

Якщо звернути увагу виключно на співи співаків, то тут також існує низка напрямків, пов'язаних з особливостями виконання. Існує широкий спектр областей оцінки співочих голосів: визначення типу, оцінка дикції та орфоепії, оцінка тембрових якостей та ін. Для робіт [32-36] як область досліджень обрано вивчення вібрато та подібних йому особливостей співу. Авторами приділяється увага частоті зміни ЧОП на секунду. За визначенням ознакою присутності у співі вібрато є коливання в межах півтону (сусідні ноти) [37] із частотою коливань від 5 до 7 за секунду [38]. Крім вібрато, що є показником професіоналізму співака і надає вокальному виконанню унікальність, що дозволяє виділити голос від фонові музики, наявність коливань у ноті, що виконується, може свідчити про такі ефекти, як тремоляція і «хитання звуку» [39]. Вважається, що до наявності вібрато в голосі схильні до високих голосів і найбільш поширений прийом серед власниць сопрано [40-41]. Цей факт ускладнює завдання визначення ноти, що звучить. З урахуванням вище викладених обмежень у діапазоні існуючих алгоритмів похибка визначення ЧОП пропетої ноти може бути високою. У свою чергу, автоматичне визначення частоти зміни ЧОП на секунду з метою виявлення тремоляції в співі також може здійснюватися з великою кількістю помилок 1-го та 2-го роду.

Особливий інтерес для аналізу вокальних виконань часто становлять записи, зроблені професійними оперними співаками. Прикладами можуть бути такі роботи як [42-44]. У цих дослідженнях вивчається вплив емоцій,

що виражаються, на мовленнєвий сигнал. Аналізу емоцій у промови диктора приділяють свою увагу також і в таких роботах, як [45-47], однак у них як матеріал для дослідження використовується розмовна мова. Професійні оперні співаки вибираються через їхню здатність висловлювати автентичну емоцію слухачам. В роботі [48] в якості об'єкта для дослідження також були обрані записи вокалу продуктивність, в яких професійних оперних співаків попросили заспівати послідовності нот, висловлюючи певні емоції. Автори дослідження мали на меті виявлення характеристик у записах, які дозволять однозначно ідентифікувати емоцію диктора. У запропонованому наборі параметрів, що рекомендується для визначення типу емоції, є частота основного тону. Аналіз аудіозаписів показав, що, наприклад, для люті чи гніву характерна низька різноманітність частот основного тону, але висока гучність мови. Незважаючи на те, що фундаментальна частота виявилася включеною в набір параметрів, авторами було визначено, що частота основного тону відіграє другорядну роль. При цьому автори відзначають важливість ЧОП як маркера збудження і дали пояснення отриманому результату. Співаки мали дотримуватися частот, запропонованих партитурою, і було неможливо варіювати її висловлювання емоцій. Слід зазначити, що це обмеження викликане не вимогами дослідження, а співом загалом. У нормальних умовах композитор використовуватиме варіацію частот основного тону під час написання мелодій. Проте, як показали результати дослідження, співаки можуть вносити невелику варіацію в частотах основного тону підтримки емоційної інтерпретації вмісту. Таким чином, наявність і відсутність змін у генерації частоти основного тону здатне служити маркером для визначення емоцій. Аналіз акустичних параметрів, пов'язаних з емоціями, що сприймаються, показав, що гнів був пов'язаний з присутністю вібрато в співі, в той час як сум характеризувався відсутністю вібрато.

Проведений аналіз алгоритмів аналізу частоти основного тону показав, що існуючі алгоритми у поточному вигляді не можуть бути застосовані у задачі обробки вокальних виконань. Насамперед це викликано обмеженням у діапазоні роботи алгоритмів. По-друге, існуючі алгоритми мають високий показник помилок, що унеможливорює точне визначення ноти, що звучить.

2 Розпізнавання нот у вокальному виконанні

2.1 Параметри сегментації та ідентифікації нот

Пропонована у роботі ідея розпізнавання нот полягає в знанні частот основного тону, виконаних диктором. Як описано в розділі 1.1, кожній ноті відповідає частота основного тону. Відповідно, на підставі значення частоти можна визначити, до якої ноти вона відноситься. У зв'язку з тим, що співає ноту в значенні певної є нездійсненним завданням, було прийнято рішення, щоб перейти від дискретних значень в інтервалах частот. На підставі даних таблиці 1.1 було отримано безліч відрізків для кожної ноти

було проведено дослідження коректності застосування середнього арифметичного при обчисленні граничних значень.

Автором проведено порівняння результатів ідентифікації нот за різних підходів до усереднення. До застосовуваного методу були додані: середня гармонійна, середня геометрична, середня арифметична, середня квадратична та середня кубічна. Оскільки при визначенні значення частоти наступної ноти її значення множиться на 2 ступенем 1/12, було вирішено застосувати даний підхід для визначення граничних значень нот. В отриманій методиці усереднення, названої логарифмічної, застосовувалися формули 2.3 та 2.4 для обчислення граничних значень.⁴

$$f_{iH} = f_{i-1} \cdot 2^{24}$$

$$f_{iB} = f_i \cdot 2^{-24}$$

де f_i – частота i -ї ноти;

f_{in} та f_{iv} – значення частот для нижньої та верхньої меж відповідно.

У таблиці 2.1 представлений фрагмент отриманого набору шкал з інтервалами в діапазоні від ноти «ля 1 октави» до ноти «соль-дієз 2- й октави».

Таблиця 2.1 - Шкали ідентифікації нот (фрагмент)

Частота та звуча ння ноти	Середнє											
	Гармонічне		Логарифмічне		Геометричне		Арифметичне		Квадратичне		Кубічне	
	ниж. гр.	Верх. гр.	ниж. гр.	Верх. гр.	ниж. гр.	Верх. гр.	ниж. гр.	Верх. гр.	ниж. гр.	Верх. гр.	ниж. гр.	Верх. гр.
698,456	678,290	718,623	678,571	718,921	678,572	718,923	678,856	719,223	679,138	719,522	679,421	719,822
739,989	718,623	761,355	718,921	761,671	718,923	761,672	719,223	761,990	719,522	762,308	719,822	762,625
783,991	761,355	806,627	761,671	806,962	761,672	806,963	761,990	807,300	762,308	807,636	762,625	807,972
830,609	806,627	854,591	806,962	854,946	806,963	854,948	807,300	855,305	807,636	855,661	807,972	856,017

Порівняння отриманих шкал показало, що відхилення від еталонної ноти значення для кожного з даних усереднення методів є від 2,72% у кубічного методу до 2,88% у гармонійного для нижньої межі та від 2,88% у гармонійного до 3,05 у кубічного для верхнього кордону. У середньому ширина інтервалу щодо значення досліджуваної ноти становить: для гармонійного методу – 5,774%, для логарифмічного та геометричного – 5,777%, для арифметичного – 5,779%, для квадратичного – 5,782%, а для кубічного – 5,784%. У ході обчислювального експерименту було помічено, що за допомогою програми отримано ідентичні результати незалежно від вибору шкали. Причиною для подібного результату може бути те, що серед записів найвищою нотою є «фа-дієз другої октави», що звучить на висоті 739,98 Гц. На даному рівні частот різниця між межами інтервалів різних

шкал досить мала: різниця між середнім гармонійним та середнім кубічним не перевищує 1.5 Гц, що становить 0.17% від значення частоти, на якій прозвучала досліджувана нота.

Кількість значень частот основного тону, що визначаються на 1 секунду досліджуваного аудіозапису, буде дорівнює його частоті дискретизації. Це призводить до того, що в загальному масиві значень траплятимуться послідовності, що відповідають одному значенню ЧОП або як мінімум ноти. Відповідно, ці ділянки можна об'єднати з метою подальшого оцінювання на предмет віднесення до ноти (сегментація). Підставою для сегментації аудіосигналу з вокальним виконанням на вокалізовані та невокалізовані ділянки буде застосування поняття мінімальної тривалості звучання ноти (рисунок 2.1).



Рисунок 2.1 – Сегментація вокалізованих та невокалізованих ділянок з урахуванням мінімальної тривалості звучання ноти

Слід врахувати, що розкид між сусідніми значеннями може бути значно більшим за межі ноти. Ці розриви можуть бути розцінені як індикатори меж вокалізованої ділянки, так і випадковим сплеском у межах однієї ділянки. На рис. 2.1 схематично представлено процес сегментації. У верхній частині схеми представлена мінімальна тривалість ноти, у центральній частині ділянки різної величини з безліччю значень частот основного тону, що належать одному класу (ноті), а в нижній частині малюнка – результат визначення вокалізованості ділянки. У лівій частині

представлена ситуація, коли в множині частот однієї ноти виникають ділянки, віднесені до інших нот. В даному випадку алгоритм сприймає вокалізовану ділянку цілісною, якщо її загальна тривалість з урахуванням присутності сторонніх нот, вища за мінімальну тривалість звучання ноти. Надалі такий лист додатково оцінювали алгоритм для при умови чистоти звуку (відсоток сторонніх частот). У середній частині представлена ситуація, де в отриманій множині частот загальна тривалість менша за мінімальну тривалість звучання ноти. У разі ділянка належить до невокалізованим і сприймається як шум. У правій частині представлена ситуація, де багато частот, віднесених до ноти, становлять мінімальну тривалість звучання. Така ситуація може бути розцінена як чисто заспівана нота.

З урахуванням даних про інтервали звучання нот масив частот основного тону переводиться в масив нот для кожного моменту часу. Послідовності ідентичних нот у даному масиві перетворюються на набір сегментів різної тривалості, всі частоти яких відносяться до обчислених діапазонів звучання нот.

Таким чином, розпізнавання нот у вокальному виконанні вимагатиме виконання наступних операцій:

- обчислення частот основного тону для аудіосигналу з вокальним виконанням кожного моменту часу;
- ідентифікація нот для кожного значення значення частоти основного тону;
- угруповання послідовних моментів часу, що відповідають одній ноті в єдиний сегмент тривалістю, що дорівнює кількості елементів, що увійшли до цієї групи;
- сегментація визначених на попередньому етапі відрізків з урахуванням мінімальної тривалості звучання ноти.

2.2 Застосування математичної моделі слухової системи людини

Крім описаних у розділі 1.2 методів обчислення частоти основного тону розглянуто алгоритм, що ґрунтується на математичній моделі слухової системи людини [51]. Обмеження в частотному діапазоні для нього визначається застосуванням у рамках дослідження параметрів сигналу в злитій мові та при онкологічних захворюваннях. Для визначення ЧОП у досліджуваному алгоритмі використовується ефект одночасного маскування, що полягає у наступному. Кожній точці вздовж основної мембрани внутрішнього вуха, яка перетворює механічні коливання в нервові імпульси, ставиться у відповідність частота звуку, що викликає максимальний відгук у цій точці. Чим більша відстань від цієї точки, тим нижче амплітуда відгуку. Сприйняття сигналів складної форми (зокрема мовного сигналу) характеризується тим, що відгук відбуватиметься попри всі частотні компоненти сигналу. Якщо амплітуда відгуку на компоненту з частотою виявиться нижче, ніж інші, то дана компонента слуховий системою сприйматися нічого очікувати.

Цей принцип застосовується визначення значення частоти основного тону в мовному сигналі. Основну мембрану можна розглядати як набір частотних резонансних фільтрів. При створенні фільтрів використовуються такі характеристики, як частота дискретизації, необхідну кількість каналів фільтрації, значення верхньої та нижньої меж частот проведеного аналізу, а також коефіцієнт точності для створюваних фільтрів.

Модифікація полягає у реалізації можливості вказівки алгоритму граничних частот визначення ЧОП. Ця модифікація була здійснена з метою автоматизації генерації наборів шаблонів досліджуваних частот. Як згадувалося раніше, в вихідній версії моделі, в набір шаблонів для алгоритму для визначення частоти основного тону було отримано

емпірично і може бути застосований лише до діапазону частот від 70 до 400 Гц. У разі застосування шаблонів до сигналів із частотами вище 400 Гц алгоритм ідентифікував значення частот для сторонніх шумів. Застосування емпіричного підходу визначення нового набору шаблонів ускладнюється збільшенням числа каналів визначення частот основного тону.

Кожен із шаблонів у наборі містить у собі послідовності з 0 та 1, що використовуються при маскуванні сигналу. Графічно застосування шаблонів можна наступним чином. Шаблон (рисунок 2.2) є послідовністю з 2 «чорних» і 3 «білих» смуг, де чорні смуги відповідають ділянкам з вокалізацією, а білі – невокалізованим. Цифра P_0 показує результат застосування маскування.

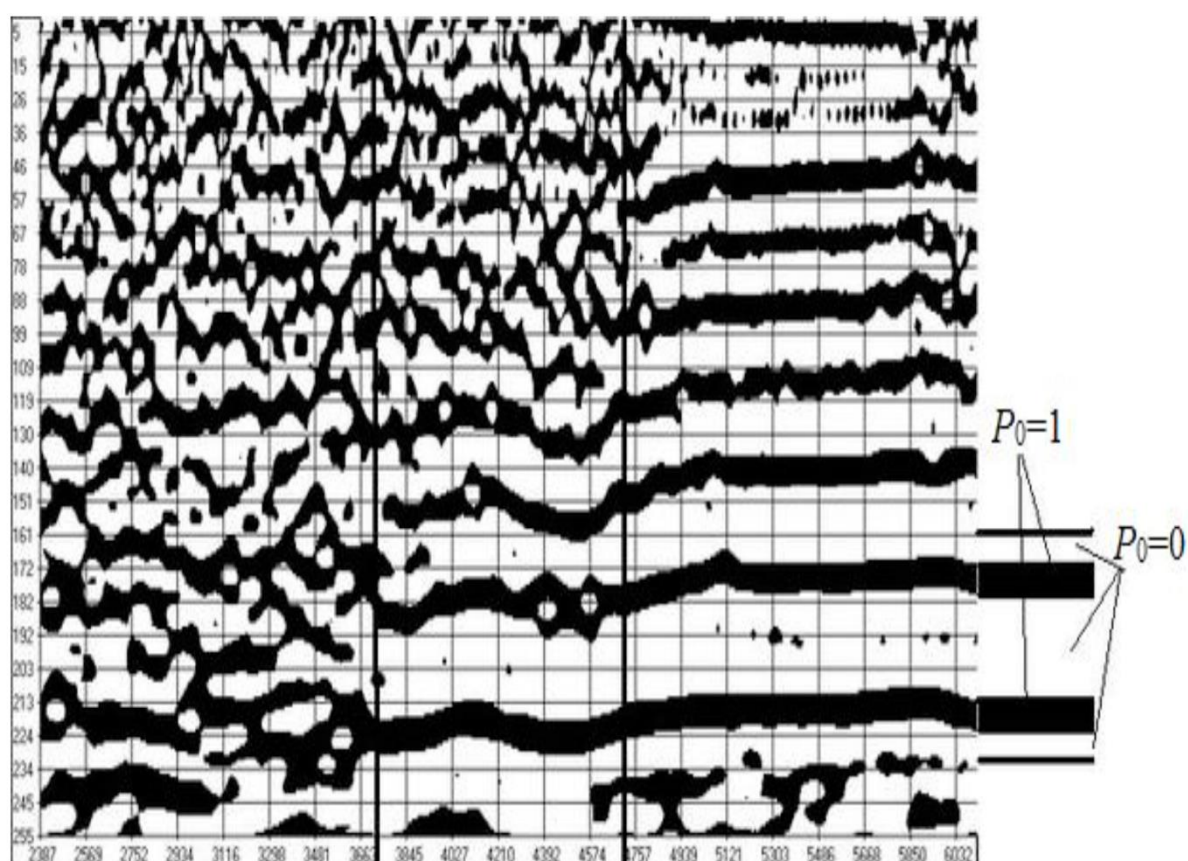


Рисунок 2.2 – Застосування шаблону під час ідентифікації ЧОП

На підставі математичної моделі слухової системи людини було визначено порядок операцій, які потрібно виконати при генерації набору шаблонів. Робота алгоритму генерації шаблонів [52] полягає у виконанні наступних кроків:

- обчислення граничних номерів каналів ЧОП (k_{0n} – нижня та k_{0v} – верхня);
- формування тестових сигналів до створення шаблонів;
- одночасне маскування (обчислення масиву результату маскування $P_0[k_t, k]$);
- обчислення масиву номерів перших каналів згортки $N_1[k_t]$, масиву кількості каналів у шаблоні $N_k[k_t]$, масиву набору шаблонів $Trl[k_t, k]$.

Таким чином, генерація шаблонів відбувається за спрощеною схемою, представленою на рис. 2.3.

Відштовхуючись від формул, отриманих у математичній моделі слуховий системи людини, було визначено, що обчислення значень номерів каналів, що відповідають нижній та верхній межах визначення частоти основного тону, k_{0n} та k_{0v}

На підставі даних про граничних значень номерів каналів, стає можливим визначити кількість масок, необхідних для роботи в моделі генерації алгоритму для до аналізу, що в свою чергу зробить можливим отримання шкала частот для досліджуваного діапазону.

$$k_m = R \cdot (k_{0n} - k_{0v}),$$

де R - коефіцієнт множення (за замовчуванням $R = 1$).

$$F_0 F_{k_1} = k \cdot dx \cdot \exp(-R + k_{ov} \cdot dx)$$

де k_1 — номер маски в діапазоні від 0 до k_m .

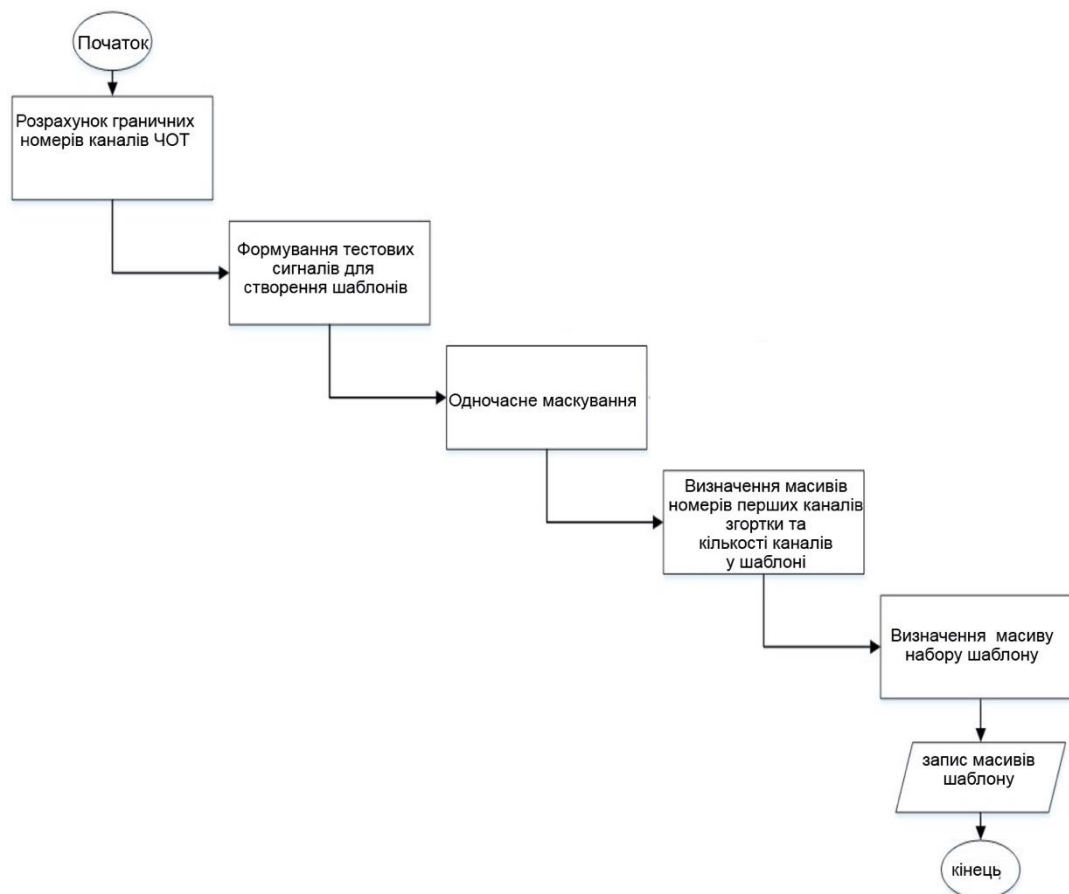


Рисунок 2.3 – Блок-схема алгоритму генерації шаблонів

З цього моменту настає етап формування тестових сигналів. Отримані значення каналів використовуються при обробці тестових сигналів згорткою з фільтрами, що дозволяє визначити амплітуду сигналу. Після цього необхідно визначити реакцію системи фільтрів на сформовані тестові сигнали. Для цього скористаємося формулами щоб отримати їх \sin та \cos складові.

2в якому I - точка часу формування реакційної системи фільтрів на тестових сигналах під час кліщів T_k .

У такому разі реакція системи фільтрів на тестові сигнали визначатиметься формулою 2.11.

$$I_{k, k_1} = (xs(k, k_1))^2 + (xc(k, k_1))^2$$

Після цього здійснюється одночасне маскування за алгоритмом, описаним у [51]. В результаті, функція одночасного маскування набуває вигляду.

На підставі отриманих формул був розроблений алгоритм генерації набору шаблонів (рисунок 2.4), що включає обчислення масиву номерів перших каналів згортки $N_1[k_i]$, при $0 \leq k_i \leq k_{on} - k_{ov}$ і масиву кількості каналів у шаблоні $N_k[k_i]$, при $0 \leq k_i \leq k_{on} - k_{ov}$ з подальшим формуванням масиву набору шаблонів при $0 \leq k_i \leq k_{on} - k_{ov}$; $0 \leq k \leq N_k[k_i] - 1$. Після цього здійснюється запис масиву набору шаблонів Trl . $1Trl = P_{0k, N[k]}$

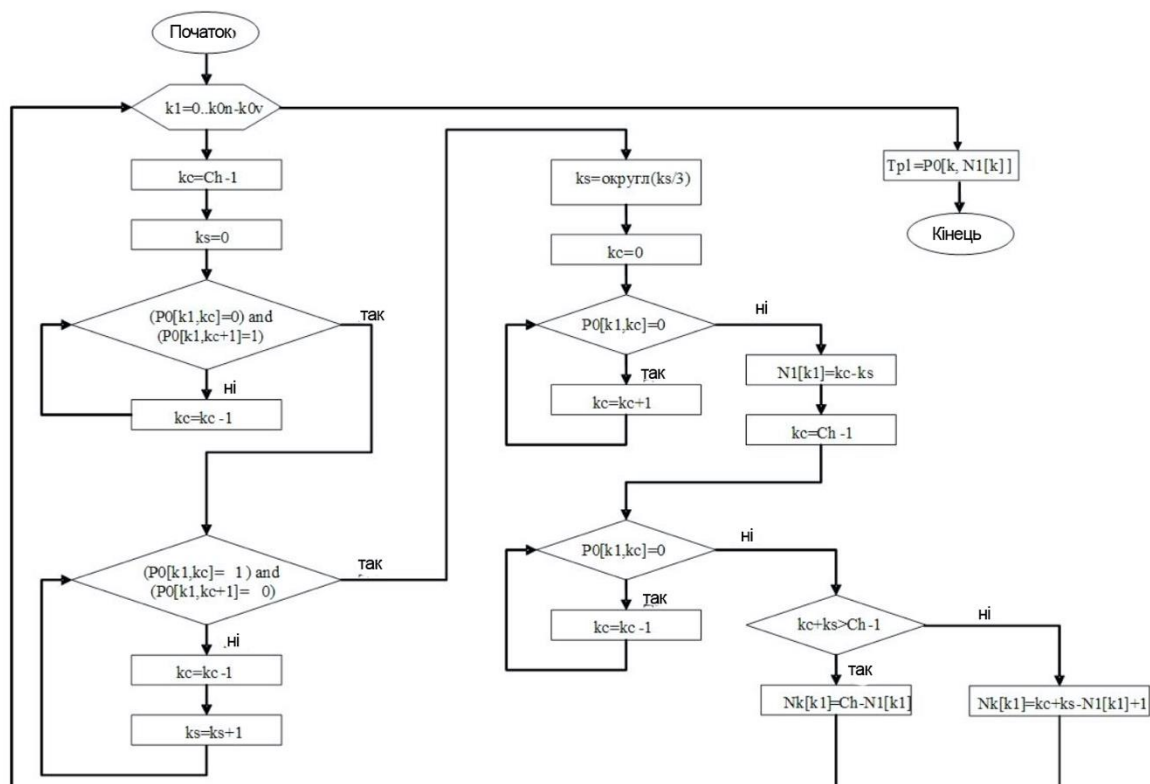


Рисунок 2.4 – Алгоритм генерації набору шаблонів

В результаті настройки значення граничних частот для визначення чо́та, ми вказати алгоритм генерації шаблонів варіюватися канали для якого необхідно сформува́ти набір шаблонів.

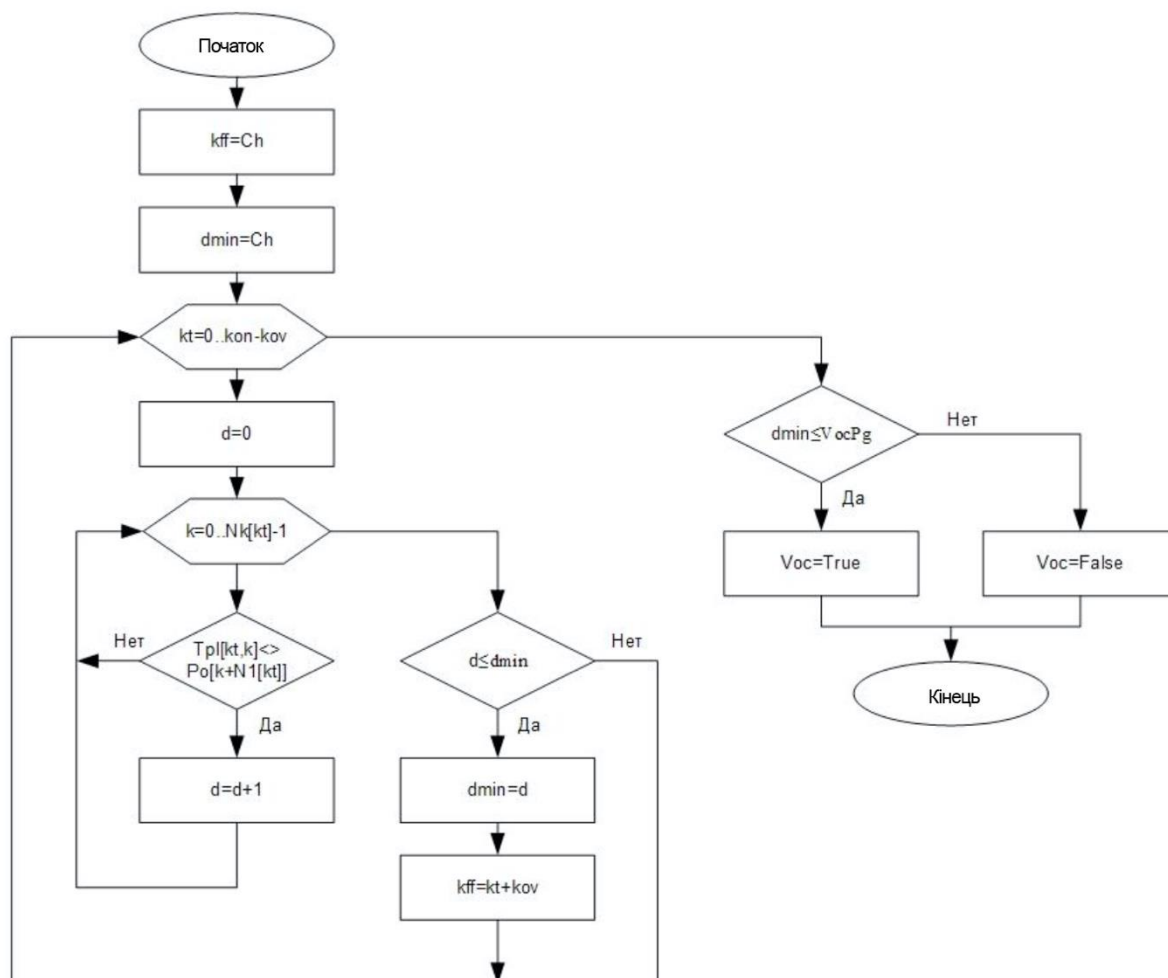


Рисунок 2.5 – Алгоритм визначення номера каналу ЧОП

На підставі отриманих шаблонів відбувається визначення номера каналу ЧОП та визначення значення ЧОП у Гц за алгоритмом, представленим на рис. 2.5. Номер каналу ЧОТ (kff) та мінімальна міра відмінності ($dmin$) прирівнюються до кількості каналів фільтрації (Ch). Цикл порівняння масиву шаблонів з результатами маскуваня здійснюється за $0 \leq k_t \leq k_{on} - k_{ov}$. У разі виявлення збігу відбувається вихід із циклу, а отриманий результат лічильника d порівнюється з мінімальною мірою

відмінності. Якщо мінімальна міра відмінності виявляється більшою або дорівнює d , то номер каналу ЧОТ додається номер верхньої межі. Далі мінімальний захід відмінності порівнюється з порогом вокалізації.

Для проведення експериментальної оцінки коректності роботи алгоритму ідентифікації частот основного тону було згенеровано синусоїдальні сигнали із заданими значеннями ЧОП. Результати роботи алгоритму для синусоїдальних сигналів з різними частотами були проаналізовані щодо помилок. У таблиці 2.2 наведено значення відносних похибок визначення частоти основного тону.

Таблиця 2.2 - Відносна помилка визначення частоти основного тону

Частота основного тону синусоїдального сигналу, Гц	Відносна похибка, %	Частота основного тону синусоїдального сигналу, Гц	Відносна похибка, %
70	<0.01	500	0.60
120	<0.01	550	0.73
200	<0.01	600	0.83
250	0.40	630	0.79
300	0.33	660	0.91
350	0.29	700	0.85
400	0.50	750	0.88
450	0.44	800	0.94

Алгоритм автоматичної генерації шаблонів для зазначеного діапазону частот був перевірений на діапазоні, що обробляється вихідною математичною моделлю. Отриманий набір шаблонів для частот основного

тону в діапазоні від 70 до 400 Гц відповідає набору, отриманому емпірично в [51].

2.3 Алгоритм розпізнавання нот у вокальному виконанні

Ідея алгоритму ось у чому. Для кожного моменту часу визначено значення частоти основного тону, тому можна визначити яка нота «прозвучала» в цей відлік. Послідовно обробляється кожен момент, що дозволяє отримати масив відрізків звучання тієї чи іншої ноти. Отриманий масив оцінюється за алгоритмом.

Якщо в межах мінімальної тривалості ноти виявляється сегмент, що належить тій самій ноті, що і еталонна для даного етапу оцінки, то цей сегмент з усіма сегментами між ним і еталонним складаються в загальну тривалість ноти, до чистого звучання ноти додається тривалість звучання ноти. До тривалості неправильних нот додаються тривалості елементів, що зустрічаються з моменту останньої правильної ноти до поточної оцінки. Дані тривалості в результаті оцінюються щодо чистоти виконання поточної ноти і винесенні рішення при виставленні оцінки студенту. У випадках, якщо в межах мінімальної тривалості ноти не виявлено сегмент тієї ж ноти, що й еталонна, то за еталонну приймається наступна нота, а поточна нота дорівнює тиші.

Оцінка якості співу (рисунок 2.7) здійснюється за двома критеріями: точність попадання в тон P і точність попадання в ритм R . Як було сказано вище, оцінка нот проводиться на основі зіставлення знайдених та еталонних нот. Еталонним нотам ставляться у відповідність знайдені ноти за принципом: якщо $N\%$ тривалості звучання ноти за часом відповідає еталонній ноті, то знайдена нота ставиться їй у відповідність.

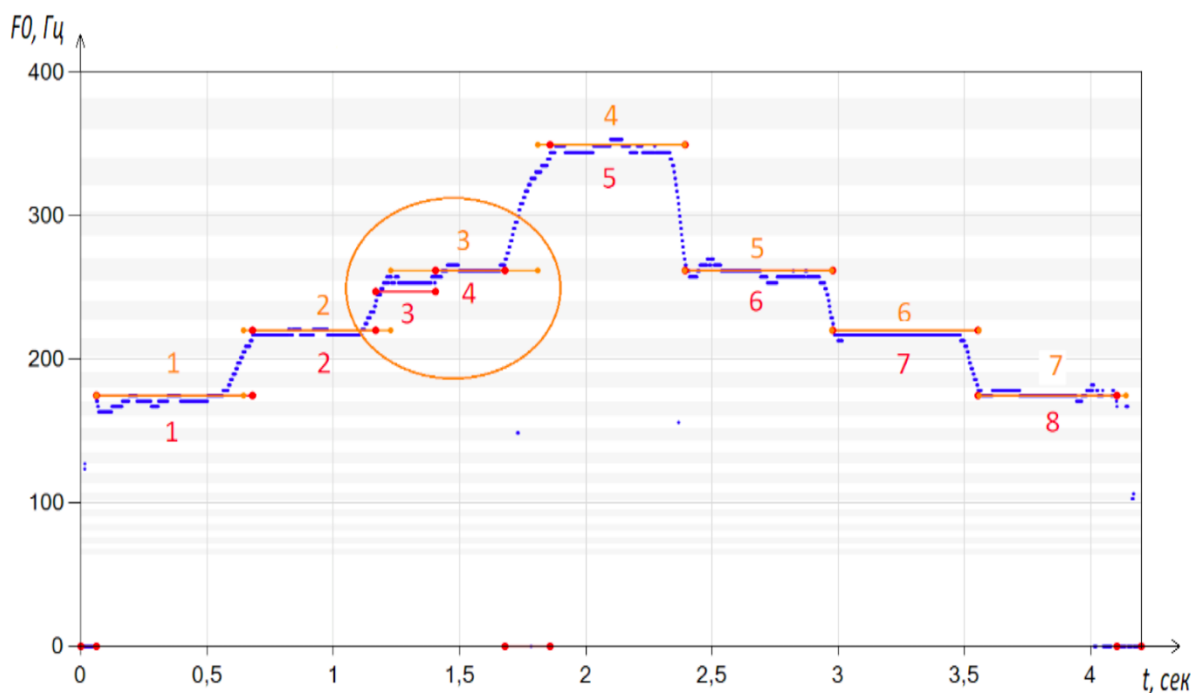


Рисунок 2.7 – Оцінка якості співу

2.4 Програмний комплекс по визначення нот вокального виконання

Структура розробленого програмного комплексу лише на рівні блоків представлена рис. 2.8.

В основу програмного комплексу дослідження вокалізованої мови [53] були покладені алгоритми слухової системи людини, які описані у [54]. Розробка програмного комплексу з вивчення мовних сигналів виконувалася з урахуванням результатів, представлених у [55]. Вирішили виділити в окрему бібліотеку алгоритми виділення частоти основного тону з урахуванням особливостей слухової системи людини, яка сприймає звук. Застосування сформованої бібліотеки дозволило досягти більш точного визначення значення частоти основного тону, порівняно з аналогами, зокрема з піковими методами [56].

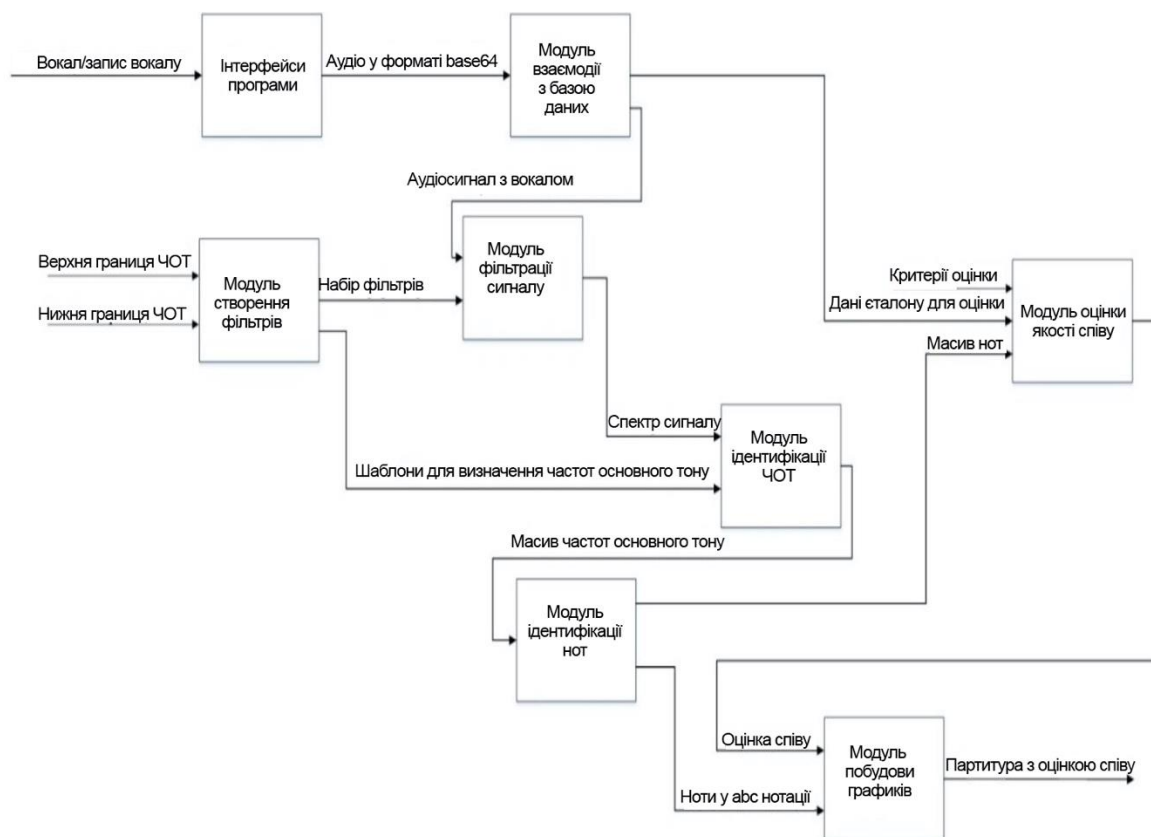


Рисунок 2.8 – Структура програмного комплексу

Програмний комплекс включає в себе базу даних, що містить інформацію про еталонні значення меж сегментів, мовні сигнали та їх описи. Цей принцип було обрано на формування вправ, які потрібно виконати студенту. В основу заносять сигнал, який буде використовуватися як зразок при одному з вправ. Цей сигнал обробляється вручну експертом з метою визначення точних меж звучання нот. Також сегментацію сигналу можна здійснити автоматично, визначивши паралельно коректність знайдених меж за алгоритмом, представленим у попередньому розділі. Модуль, відповідальний оцінку якості сегментації, також здійснює взаємодію Космосу з іншою частиною програмного комплексу з вивчення мовних сигналів. Якщо здійснюється виконання однієї з вправ, внесеної до бази даних, система зіставляє отримані у виконанні нот межі звучання з

еталонними, повідомляючи користувачеві про рівень коректності виконання завдання.

Модуль ручної сегментації було реалізовано доти, як програмний комплекс було доповнено алгоритмом автоматичної сегментації. Згодом цей модуль використовувався для перевірки коректності роботи алгоритму шляхом порівняння результатів ручної обробки експертом із автоматично отриманими результатами.

На етапі (рисунок 2.9) створення фільтрів програма виробляє генерацію вихідних інструментів заявленої області обробки. Для цього в межах від нижньої частоти $F_{он}$ до верхньої $F_{ов}$ за формулами створюється набір фільтрів, необхідний перетворення мовного сигналу в його спектрограму. Крім цього, створюється набір шаблонів, який зіставляється із отриманим після фільтрації сигналу спектром прослуханого сигналу. На основі алгоритмів визначення ЧОП програмою, що надійшли даних, видається масив миттєвих значень частот F_o . По розробленому під час дослідження алгоритму отримання нот з урахуванням даного масиву програмою видається висновок у тому, яка послідовність нот було зафіксовано у мовному сигналі [57].

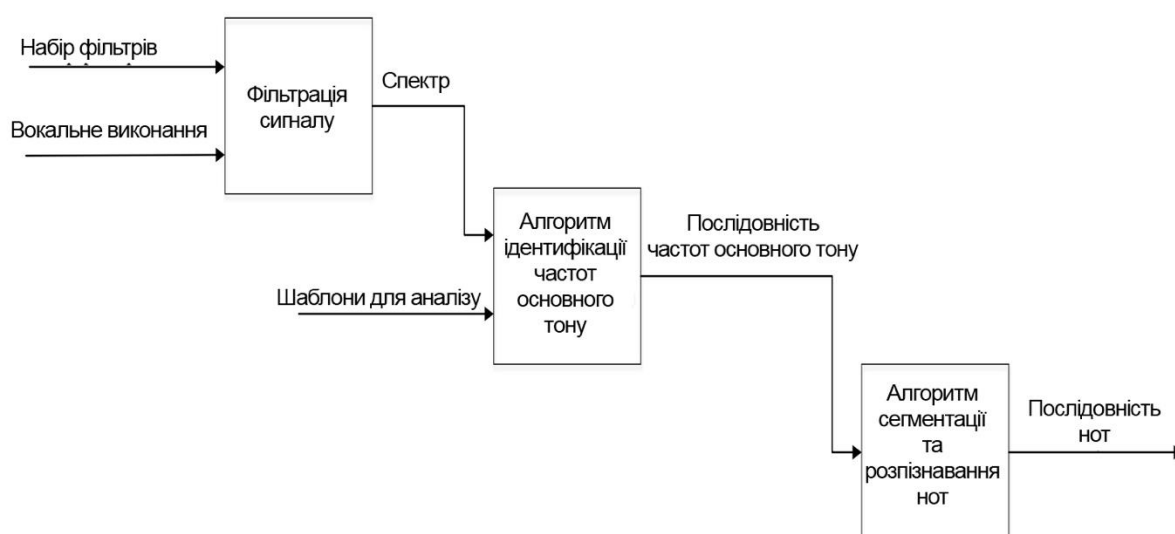


Рисунок 2.9 – Етапи розпізнавання нот у вокальному виконанні

Інтерфейс програми містить такі області: меню аналізу аудіозапису (ліва частина) та область виведення результатів (права частина). У верхній частині меню можна відкрити записи у форматі wav або записати з мікрофона аудіосигнал користувача, використовувати еталонні значення нот вправ в abc-нотації для оцінки якості співу, експортувати отримані частоти основного тону для проаналізованого запису. Після вибору аудіозапису для аналізу програма повідомляє параметри файлу: назву, частоту дискретизації, тривалість тощо. У лівій частині вікна надається графічне відображення результатів аналізу в 3 форматах: графік частот основного тону з маркерами розпізнаних та еталонних нот, партитури розпізнаних нот та еталонних нот.

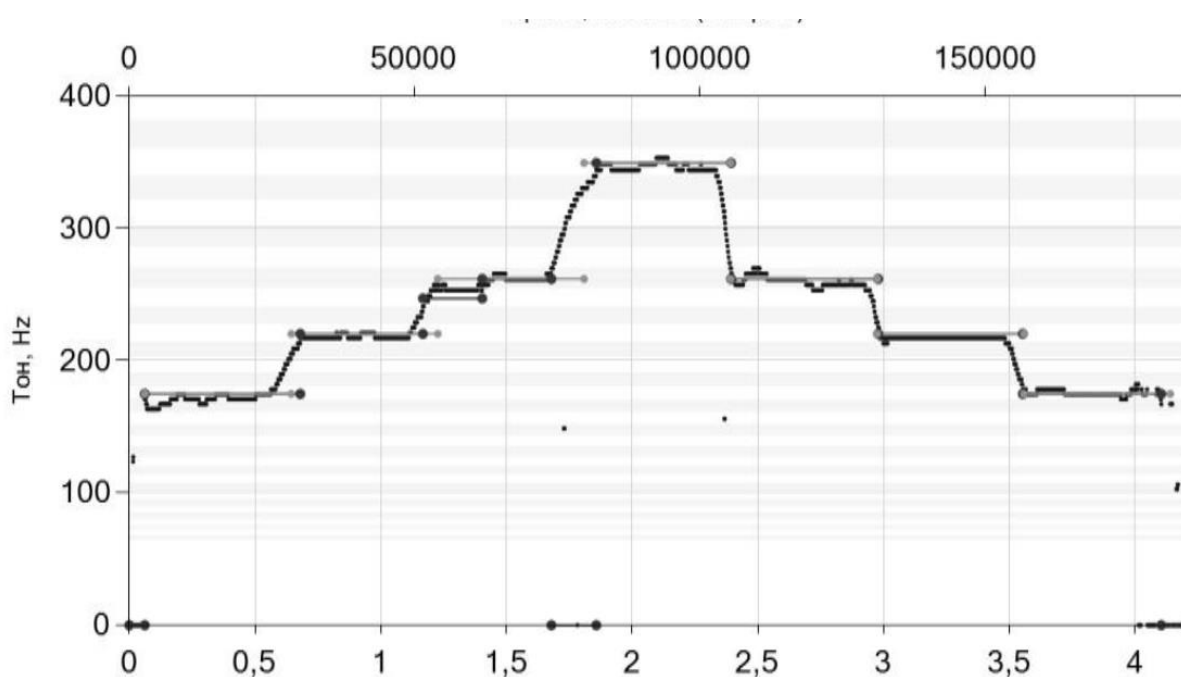


Рисунок 2.10 – Графічне відображення частотної області досліджуваного сигналу зі знайденими та еталонними нотами

Для розділів "Ноти" та "Оцінка" надається abc-нотація, використана для виведення партитури в текстовому форматі, яку можна скопіювати та використовувати у сторонніх програмних засобах, що використовують цю

форму нотації. Також для виконання завдання надається оцінка за критеріями, зазначеними в розділі 2.3.

Представлений результат без урахування еталонних значень. В даному випадку не оцінюється точність влучення в ритм і виводиться відсоток влучення частот основного тону для кожної ідентифікованої ноти в область, що відповідає такій. Для обчислення точності потрапляння в ритм обчислюється різниця між початком еталонної ноти та початком першої знайденої ноти, що відповідає еталонній, а також різниця між кінцем еталонної ноти та кінцем останньої знайденої ноти.

Натискання на кнопку «Налаштування» нижньому лівому куті викликає меню, в якому можна змінити параметри дослідження аудіозаписів. Після натискання на кнопку «Аналізувати» буде запущено алгоритми ідентифікації частот основного тону за заданими в меню параметрами та розпізнавання нот з урахуванням мінімальної тривалості звучання ноти. Мінімальна та максимальна частоти шкали відповідають параметрам F_{0n} та F_{0v} , представленим у формулах 2.5 та 2.6, кількість каналів шкали використовується для розрахунку параметра R у формулі 2.7. Також у цьому вікні можна змінити межі виведення результатів: мінімальна та максимальна частоти основного тону. Конфігурація діапазонів нот

дозволяє вибрати один із методів обчислення граничних значень між нотами з метою визначення ступеня впливу на точність ідентифікації [50]. Інші параметри використовуються для оцінки якості виконання при визначенні мінімальної тривалості звучання ноти, що використовується на етапі сегментації, та обчисленні ступеня відхилень у центах (сота півтони) від еталонної ноти.

3 ОЦІНКА ТОЧНОСТІ РОБОТИ АЛГОРИТМУ РОЗПІЗНАВАННЯ НОТ У ВОКАЛЬНОМУ ВИКОНАННІ

Дослідження алгоритму розпізнавання частот основного тону показало, що діапазоні від 70 до 800 Гц відносна помилка алгоритму ідентифікації ЧОП становила менше 1%. Це робить можливим застосування модифікованої математичної моделі слухової системи людини не тільки для аналізу параметри мовного сигналу, але також для ідентифікації записів. Отже, на точність розпізнавання нот не впливають одержувані значення частот основного тону.

3.1 Експеримент із розпізнавання нот у вокальному виконанні

Для визначення коректності виконання чистої ноти було зібрано аудіозаписи з нотами, зіграними на налаштованому фортепіано, де частота звучання кожної ноти точно відповідала частоті, представлений у таблиці 1.1.

Відсутність сторонніх нот і змін у частоті коливання мало виключити сторонні на роботу алгоритмів. У ході експерименту були протестовані аудіозаписи від нот до-дієз великої октави (69.3 Гц) до ноти сі другої октави (987.75 Гц). Усі ноти до си-бемоль другої октави (932.32 Гц) включно ідентифіковані без помилок.

При дослідженні надійності детектора частот основного тону [58] було визначено кілька стратегій, на основі яких було зроблено 16 записів (8 жіночим голосом та 8 чоловічим голосом):

- тестування коротких звуків (стаккато),
- тестування звуків, що знаходяться на мінімальному інтервалі,
- тестування звуків, що знаходяться в середньому інтервалі,

- тестування звуків з вимовою слів,
- тестування мелодії з вимовою слів.

Загалом у 16 тестах було проаналізовано 114 нот, 58 з яких були заспівані жіночим голосом, а 56 – чоловічим. Для жіночого голосу було отримано результати, подані у таблиці 3.1. Для чоловічого голосу аналогічно представлені в таблиці 3.2.

Таблиця 3.1 – Результат тестування розробленої програми та аналогів

	Praat	Melodyne	Розроблена програма
Правильно розпізнано нот	45	58	58
Помилково нерозпізнаних (Помилки 1 роду)	13	0	0
Помилково розпізнаних (помилки 2 роди)	0	4	0

Таблиця 3.2 – Результат тестування для чоловічого голосу

	Praat	Melodyne	Розроблена програма
Правильно розпізнано нот	41	53	55
Помилково нерозпізнаних (Помилки 1 роду)	15	3	1
Помилково розпізнаних (помилки 2 роди)	0	2	0

Крім аудіозаписів зі співом стакато та легато, були розглянуті такі прийоми, як крещендо та декрещендо, арпеджіо та глісандо (висхідне та

низхідне). Вирішили перевірити наявність впливу різних способів виконання нот на якість їх ідентифікації.

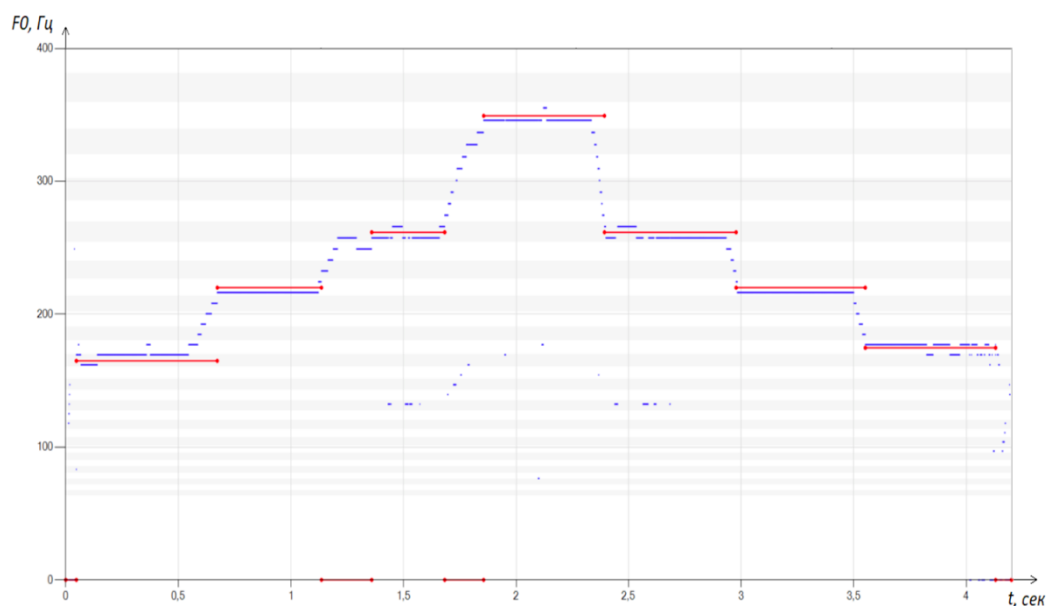


Рисунок 3.1 – Результат обробки аудіозапису з арпеджіо

Як видно з рисунку 3.1, під час співу арпеджіо відсутній вплив на якість роботи алгоритмів у програмному комплексі. Під терміном розуміється послідовне виконання нот без накладання друг на друга.

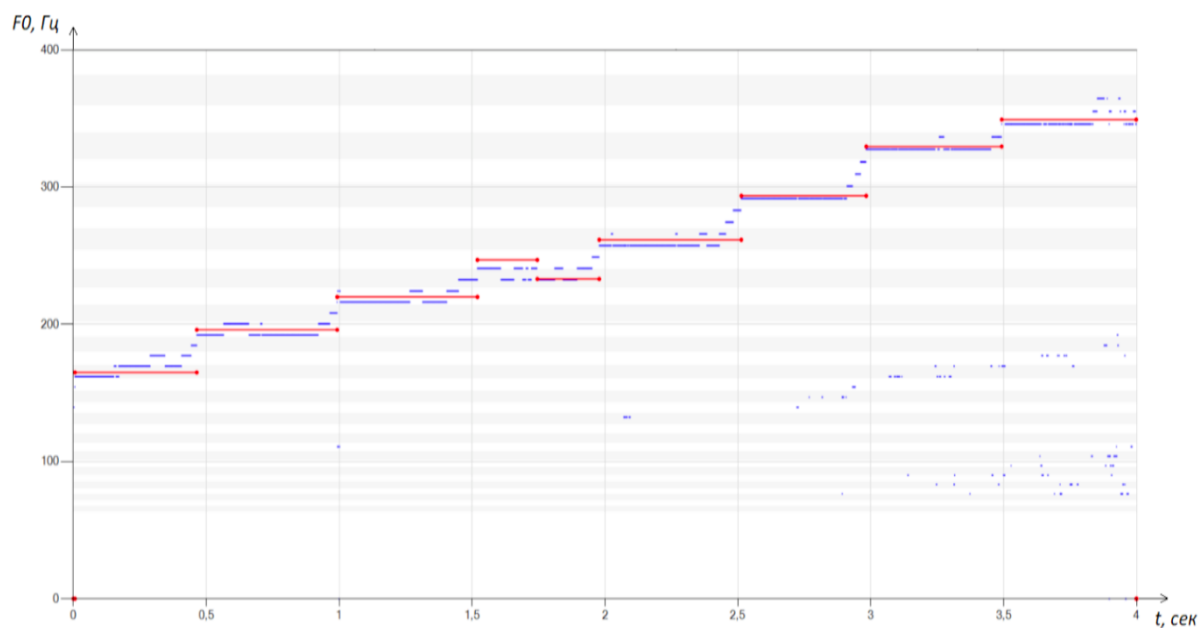


Рисунок 3.2 – Результат обробки аудіозапису з крещендо

Зазвичай, цей термін застосовується до гри на музичних інструментах. Специфічність спів арпеджіо, щоб зберегти однорідність тону голосу по всім нотам смуги співу. Відсутність впливу на якість ідентифікації нот полягає в тому, що під час аналізу алгоритм враховує лише частоту основного тону.

Крецендо та декрецендо – це виконання ноти з поступовим збільшенням чи зменшенням сили звуку. Як можна бачити за результатами аналізу аудіозаписів (рисунки 3.2-3.3), зміни в силі звуку також не впливають на якість розпізнавання заспіваних нот.

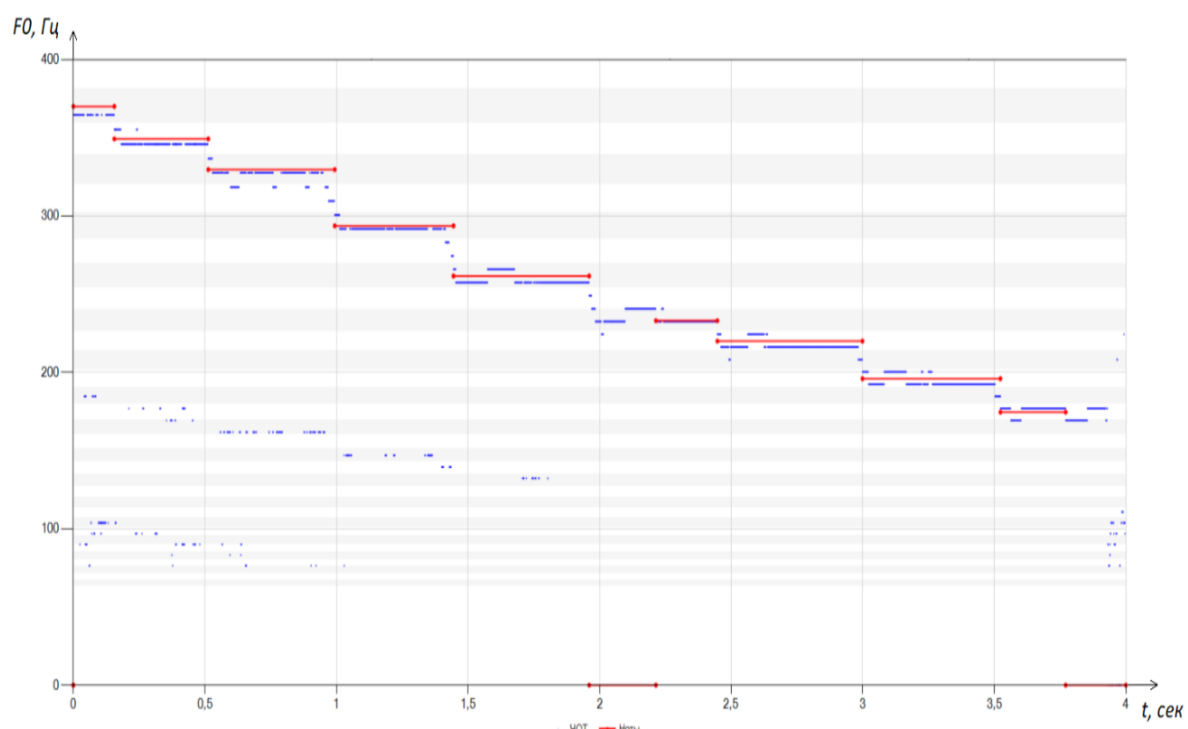


Рисунок 3.3 – Результат обробки аудіозапису з декрецендо

При дослідженні аудіозаписів з вокальним виконанням нот, у яких частота основного тону знаходиться в діапазоні від 400 до 600 Гц, було виявлено ефект вібрато, що впливає на точність розпізнавання нот у зазначеному інтервалі частот.

3.2 Визначення частоти помилок алгоритму розпізнавання нот

В рамках експерименту зі складання вправ зі співом по заданому звуку мелодичних мотивів, що налаштовують слух у відповідній ладо-тональності, було проведено збір записів з учнів музичної студії. Кожному учневі давалося завдання прослухати еталонний запис і зробити 5 записів зі співом прослуханого завдання. Загалом було зібрано 17 еталонних записів і 740 записів учнів, що містить у сукупності 13078 заспіваних нот.

З метою визначення частоти коректної роботи програми була проведена експертна оцінка набір записів для однієї з вправ, в якій здійснювався спів без застосування таких прийомів як вібрато та глісандо. Усього було оцінено 53 записи, що містять 477 нот. У таблиці 3.3 наведено статистику за результатами розпізнавання нот розробленим програмним комплексом та програмою Melodyne.

Таблиця 3.3 - Результати дослідження записів за допомогою програм

Показник	Розроблений програмний комплекс	Melodyne
Розпізнано всі ноти	47 записів	26 записів
Не розпізнано 1 ноти	4 записів	9 записів
Не розпізнано 2 ноти	2 записів	6 записів
Не розпізнано 3 і більше нот	0 записів	12 записів

Загалом, розроблений програмний комплекс розпізнав 469 нот із 477, а програма Melodyne змогла виділити 392 ноти. В якості прикладу, кілька результатів обробки записи різних акустичних систем наведені нижче.

У записі №7 екпертом було виділено такі ноти: «до», «ре»,

"мі", "фа", "сі", "фа", "ре-дієз", "до-дієз" 1 октави і "сі" малої октави.

Програма Melodyne (рисунок 3.4) змогла розпізнати такі ноти: "сі" малої

октави; "до", "ре-дієз", "фа-дієз", "фа", "ре-дієз", "до-дієз" 1 октави і «сі» малої октави, що містить 5 із 9 виділених експертом значень.

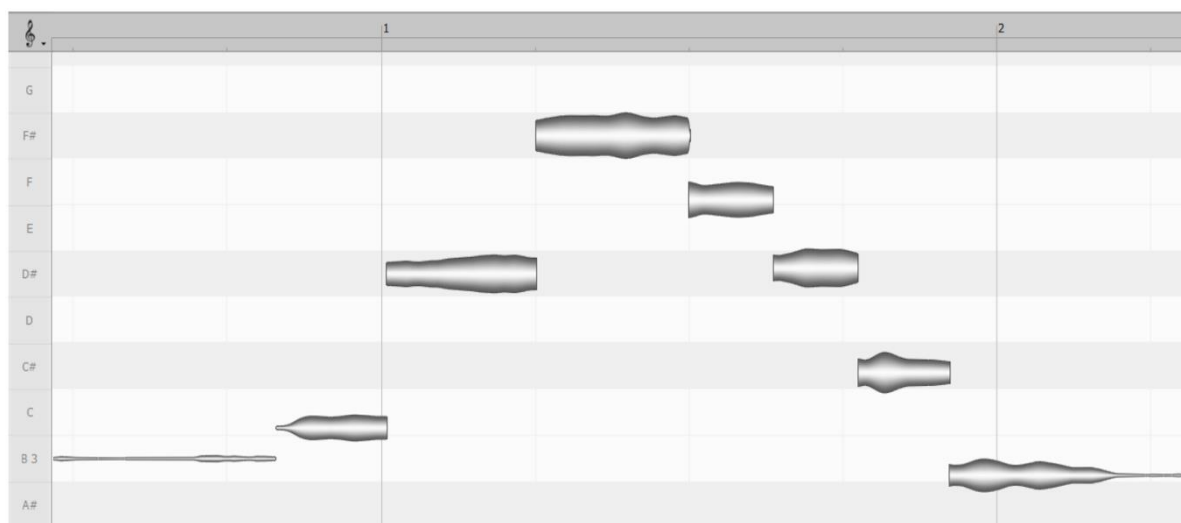


Рисунок 3.4 – Результат обробки запису №7 програмою Melodyne

Розроблений програмний комплекс (рисунок 3.5) розпізнав такі ноти: "сі" малої октави; "до", "ре", "мі", "сінь", "фа", "ре-дієз", "до-дієз" 1 октави; "сі" малої октави, що на 8/9 відповідає експертній оцінці. Нота «фа» на ділянці від 1.75 до 1.9 секунд не була розпізнана програмою через переход на сусідню ноту.

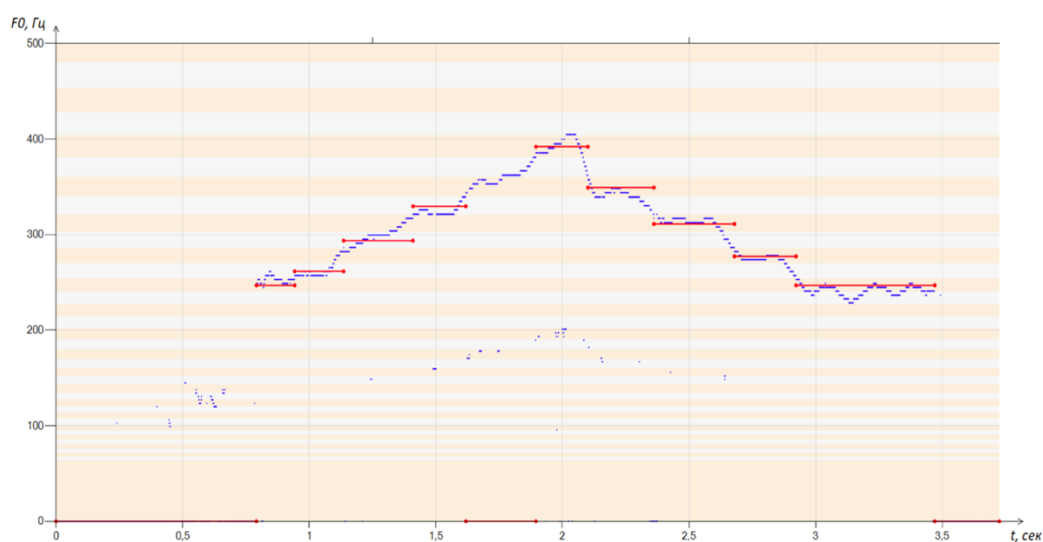


Рисунок 3.5 – Результат обробки запису №7 розробленим програмним комплексом

У записі №28 експертом було виділено такі ноти: «сі» малої октави; "до-дієз", "ре-дієз", "фа", "фа-дієз", "мі", "ре" першої октави;

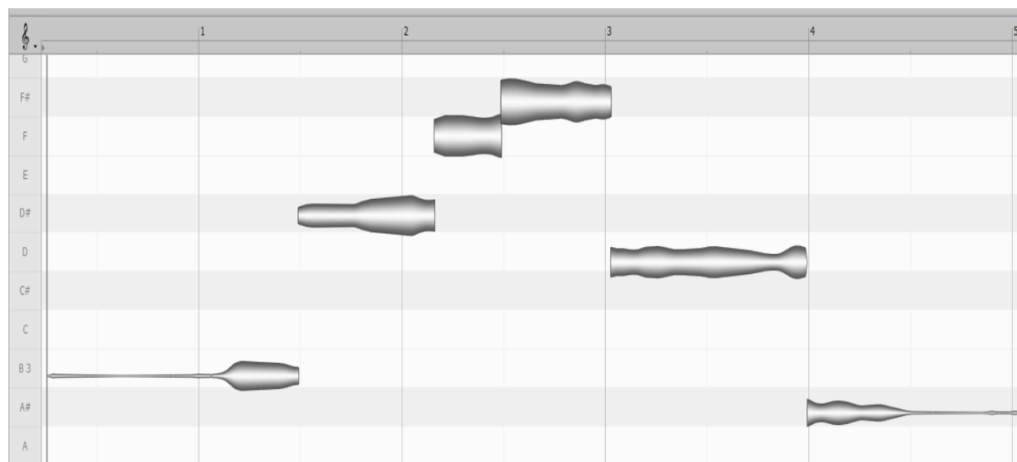


Рисунок 3.6 – Результат обробки запису №28 програмою Melodyne

"сі", "ля-дієз" малої октави. Програма Melodyne (рисунок 3.6) змогла розпізнати наступні ноти: «сі», «ре-дієз», «фа», «фа-дієз», «ре» та «лядієз», що містить 6 із 9 виділених експертом значень. Розроблений програмний комплекс (рисунок 3.7) розпізнав такі ноти: "Сі" малої октави; "до-дієз", "ре-дієз", "фа", "фа-дієз", "мі", "ре" першої октави;

"сі", "ля-дієз" малої октави, що повністю відповідає експертній оцінці.

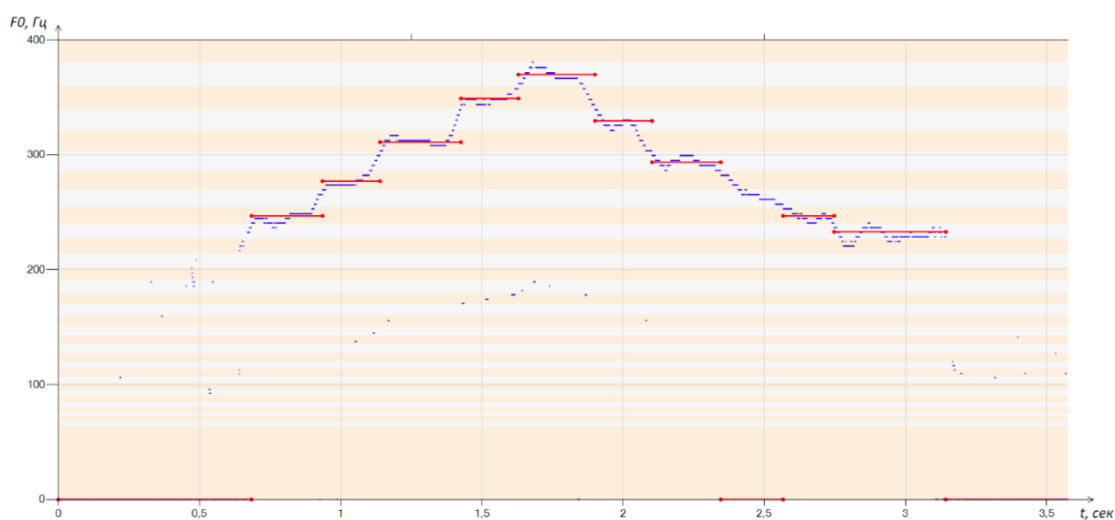


Рисунок 3.7 – Результат обробки запису №28 розробленим програмним комплексом

Нижня та верхня межі довірчого інтервалу для невідомої частоти настання події.

В результаті було визначено, що частота помилок у роботі алгоритму розпізнавання нот у програмному комплексі із ймовірністю 0,95 не перевищує 3.3%

3.3 Перевірка коректності експертних оцінок

З метою визначення коректності оцінки експертами еталонних значень нот був зібраний набір з 29 нот в діапазоні від малої октави до фа-дієз другої октави. У рамках дослідження застосовувалися оцінки 3 експертів, головним критерієм яких було з обов'язковою наявністю музичної освіти за напрямом «вокал». Експертам було видано пронумеровані аудіозаписи та таблиці для запису певних ними нот. Результати представлені у таблиці 3.4.

Таблиця 3.4 – Результати оцінки записів із нотами експертами

Номер запису	Відповідна нота		
	Експерт 1	Експерт 2	Експерт 3
001	До-дієз першої октави	До-дієз першої октави	До-дієз першої октави
002	Ре-дієз малої октави	Ре-дієз малої октави	Ре-дієз малої октави
003	Сі першої октави	Сі першої октави	Сі першої октави
004	Ля-дієз першої октави	Ля-дієз першої октави	Ля першої октави
005	Ре першої октави	Ре першої октави	Ре першої октави
006	Ре-дієз другої октави	Сіль-дієз першої октави	Сіль першої октави
007	Мі першої октави	Мі першої октави	Мі першої октави
008	До-дієз другий октави	До-дієз другий октави	До-дієз другої октави

009	Фа-дієз другий октави	Фа-дієз другий октави	Фа-дієз другої октави
010	Сі малої октави	Сі малої октави	Сі малої октави
011	Ля-дієз малої октави	Ля-дієз малої октави	Ля-дієз малої октави
012	Мі малої октави	Мі малої октави	Мі малої октави
013	Сіль-дієз першої октави	Сіль-дієз малої октави	Сіль-дієз першої октави
014	Фа-дієз малої октави	Фа-дієз малої октави	Фа-дієз малої октави
015	Мі другий октави	Мі другий октави	Мі другий октави
016	Ля першої октави	Ля першої октави	Ля першої октави
017	Фа другої октави	До другої октави	До-дієз другої октави
018	Ля-дієз першої октави	Ре-дієз першої октави	Ре-дієз першої октави
019	Фа другої октави	Фа другої октави	Фа-дієз другої октави
020	Фа великої октави	Фа малої октави	Фа малої октави
021	Сіль першої октави	Сіль першої октави	Сіль першої октави
022	До-дієз другий октави	Фа-дієз першої октави	До другої октави
023	До першої октави	До першої октави	До першої октави
024	Ля малої октави	Ля малої октави	Ля малої октави
025	Ре-дієз другої октави	Ре-дієз другої октави	Ре-дієз другої октави
026	Сі великий октави	Сіль малої октави	Сіль малої октави
027	Фа першої октави	Фа першої октави	Фа першої октави
028	До-дієз великий октави	Ре малої октави	Ре малої октави
029	Ре другий октави	Ре другий октави	Ре-дієз другої октави

Аналіз отриманих результатів показав, що у 62% всі експерти та у 28% випадків 2 експерти з 3 давали однакову оцінку для нот. Далі на основі даних анкетного опитування було складено зведену матрицю рангів (таблиця 3.5).

Таблиця 3.5 - Матриця рангів для експертів

№	Експерт 1	Експерт 2	Експерт 3	Сума рангів
001	11	12	11	34
002	4	2	2	8

003	20	22	21	63
004	18,5	21	19,5	59
005	12	13	12	37
006	24,5	19	16,5	60
007	13	15	14	42
008	21,5	24	23,5	69
009	29	29	28,5	86,5
010	9	10	9	28
011	8	9	8	25
012	5	3	3	11
013	16	7	18	41
014	6	5	5	16
015	26	27	27	80
016	17	20	19,5	56,5
017	27,5	23	23,5	74
018	18,5	14	13	45,5
019	27,5	28	28,5	84
020	2	4	4	10
021	15	18	16,5	49,5
022	21,5	17	22	60,5
023	10	11	10	31
024	7	8	7	22
025	24,5	26	25,5	76
026	3	6	6	15
027	14	16	15	45
028	1	1	1	3
029	23	25	25,5	73,5
Σ	435	435	435	1305

Далі необхідно провести оцінку значення коефіцієнта конкордації за формулою 4.3.

Отримане значення X^2 більше значення таблиці, що вказує на значимість отриманого коефіцієнта узгодженості. Отже, можна дійти невтішного висновку, що узгодженість експертів задовільна.

4 АНАЛІЗ ВИКОНАННЯ НОТ ІЗ РІЗКОЮ ЗМІНОЮ ЧАСТОТ ОСНОВНОГО ТОНУ

Попереднє тестування програмного комплексу показало, що підвищення якості ідентифікації нот можливе при розширенні діапазону частот і визначенні ситуацій, що знижують точність роботи алгоритму.

4.1 Вплив різкої зміни частот на точність роботи алгоритму

Проведені в 3 главі випробування алгоритму показали, що при обробці аудіозаписів, що включають різні підходи до виконання нот у діапазоні від 70 до 800 Гц, програмний комплекс дозволяє розпізнати більше 95% заспіваних диктором нот. Було визначено, що з співи арпеджио, крещендо і декрещендо відсутня впливом геть якість роботи алгоритмів у програмному комплексі. Відсутність впливу на якість ідентифікації нот полягає в тому, що в алгоритмі при аналізі враховується лише частота основного тону та специфікою роботи одного з етапів алгоритму ідентифікації нот, що відповідає за визначення належності вокалізованої ділянки до ноти, полягає в орієнтованості на чисте виконання.

Виконати ноту так, щоб усі частоти в момент співу знаходилися в межах смуги, що відведена під неї, досить складно. З цієї причини враховуються сусідні ноти вище та нижче виконуваної. На рисунках 4.1-4.3 спектр частот поділено на ділянки, що відповідають межам нот. Кожна ділянка виділена горизонтальною смугою на фоні графіка частот. По осі X – час, по осі Y – частоти основного тону.

Як можна бачити на рис. 4.1, сегмент, заспіваний в межах однієї ноти, був розпізнаний програмою. Кількість моментів, що потрапили в сусідні ноти, незначна і не вплинула на результат ідентифікації. Переважна

кількість розпізнаних частот основного тону виявилось в діапазоні заспіваної ноти. З урахуванням прописаних в алгоритмі вимог щодо точності виконання, програма змогла зробити висновок про заспівану ноту. Якщо кількість фрагментів у кожному з трьох ділянок виявилось приблизно однаковим, алгоритм сприймає весь вокалізований сегмент як шум.

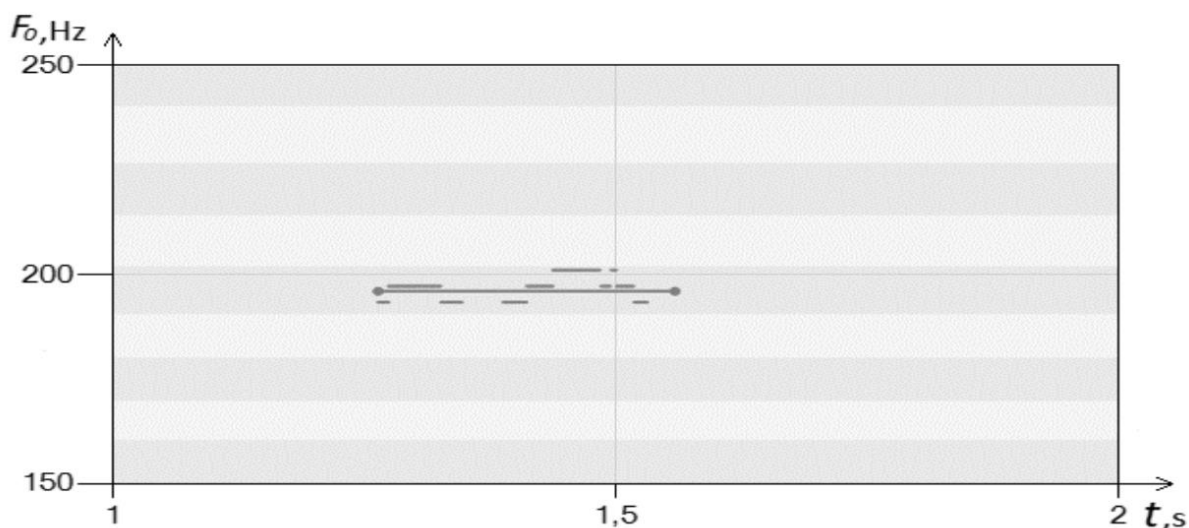


Рисунок 4.1 – Сегмент зі співом у межах однієї ноти

Було визначено, що помилки виникають для ситуацій з підходами, що мають на увазі зміну значення частоти звучання [59] у процесі виконання: вібрато (рисунок 4.2) та глісандо (рисунок 4.3).

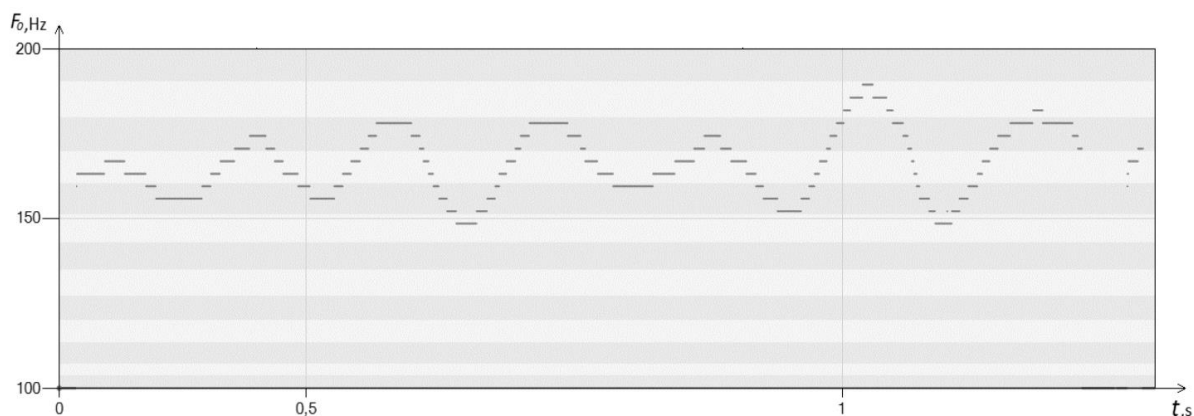


Рисунок 4.2 – Сегмент з нотою, заспіваною з вібрато в голосі

При співі вібрато відбуваються коливання у частоті звучання щодо якоїсь головної ноти у межах півтони. Слід зазначити, що існують різновиди вібратоподібних коливань. Сюди відносять тремоляцію та «хитання голосу». У зв'язку з тим, що алгоритм налаштований на визначення чистих нот, такі сегменти сприймаються як шум. Як можна помітити, коливання під час виконання основної ноти (в даному випадку – нота "мі малої октави") охоплюють одразу 4 ноти. Врахування мінімальної тривалості звучання ноти призводить до того, що для кожної з 4 нот недостатня кількість часу відбувався спів у її межах. При цьому повернення до цієї ноти відбувається після переходу до інших нот, що знижує частку частот, що належать до головної ноти.

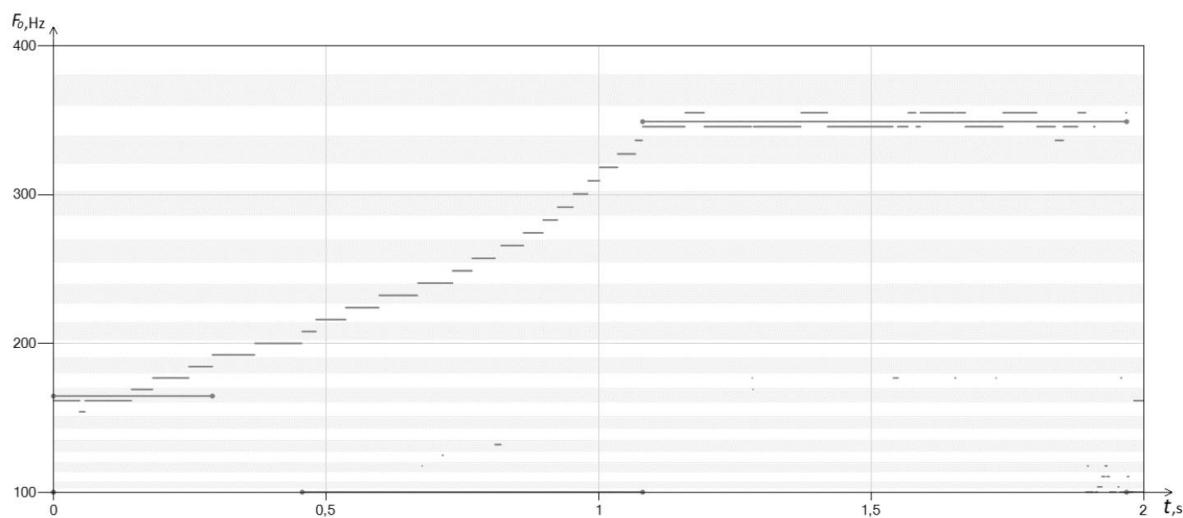


Рисунок 4.3 – Сегмент, заспіваний зі висхідним глісандо

Іншим прикладом внесення колористичного ефекту в спів може бути такий прийом як глісандо. Як відомо з теорії музики, при співі глісандо відбувається плавне ковзання від однієї ноти до іншої. При такому співі перехід відбувається занадто швидко для того, щоб зуміти ідентифікувати кожен окрему ноту в момент ковзання, оскільки на кожен сегмент охоплених нот доводиться відводиться менше 0,1 секунди. Глісандо може

бути як по зростанню (як показано на рис.4.3) і по спадаючій. Як можна помітити на рис., алгоритм здатний визначити початкову та кінцеву ноти, між якими здійснюється ковзання. Однак охоплення 12 нот між ними сприймається як шум.

Таким чином, виникає необхідність ідентифікації ділянок, що сприймаються програмою як шум. Вважатимемо, що спрямований перехід від однієї ноти до іншої – це глісандо, а коливання щодо однієї ноти в рамках єдиного сегмента – це тремоляція. На даному етапі не розглядатиметься визначення типу тремоляції.

Оскільки масив знайдених частот тону, в рамках якого не вдалося визначити чітку ноту, можна розглядати як часовий ряд, було вирішено провести попередній аналіз даних. Серед процедур попереднього аналізу виділяють: пошук аномалій, згладжування часових рядів, перевірку наявності тренду та розрахунок показників динаміки процесів.

З точки зору предметної області, що розглядається, аномальними спостереженнями можна сприймати сплески частот, що визначаються за межами звучання основної мелодії. Сюди відносяться будь-які шуми, що відсіваються алгоритмами на етапі сегментації нот. Відповідно, цю процедуру можна вважати виконаною.

У межах згладжування часового низки відбувається обчислення істинних рівнів ряду розрахунковими значеннями, мають більш гладку динаміку, ніж вихідні дані. Було вирішено сконцентрувати увагу на методах механічного згладжування, оскільки в досліджуваному завданні необхідно оцінювати кожен окремий вирівняти з урахуванням фактичних значень сусідніх з ним рівнів. Метод виваженої ковзної середньої не застосовується для досліджуваного завдання, оскільки в ньому не може утримуватися квадратичний або кубічний тренд. Крім того, немає ніякої необхідності, щоб застосувати метод експоненціального згладжування, так як він використовується у задачах прогнозування розвитку процесу після

досліджуваної галузі. У зв'язку з цим буде використаний метод простий ковзної середньої.

Як кількість спостережень p , що входять до інтервалу згладжування, було вирішено використовувати величину мінімальної тривалості звучання ноти, що застосовується в алгоритмі сегментації та ідентифікації нот. Вибір обумовлений тим, що при співі вібрато і гліссандо тривалість співу в рамках подібних ділянок суттєво вища за цей параметр, але при цьому ідентифікація окремих чистих нот не була здійснена базовим алгоритмом. Недолік методу є, що першими і останніми p спостереження залишаються згладженими, які можна знехтувати з - за загальну тривалість передбачуваної області.

Перевірка наявності тренду за своєю сутністю є перевіркою гіпотези про незмінність середнього значення часового ряду. Для перевірки ряд може бути розбитий на n частин, кожна з яких сприймається як окрема сукупність. Отримані середні значення для кожної сукупності будуть порівнюватися. У разі, якщо середні значення зростають або спадають, будемо вважати, що досліджуваний відрізок може містити гліссандо (висхідне або спадне відповідно).

4.2 Аналіз вібратоподібних вокальних виконань

Усі попередні обчислення здійснювалися у Microsoft Office Excel. Вбудований функціонал містить функцію «Аналіз даних», в якому для вибраного ряду можна побудувати ковзне середнє.

На рис. 4.4 представлений результат застосування цього методу. У вбудованому методі відбувається аналіз лише попередніх значень без урахування наступних після оцінюваного, що дозволяє отримати оригінальну, неінформативну в рамках дослідження, що проводиться.

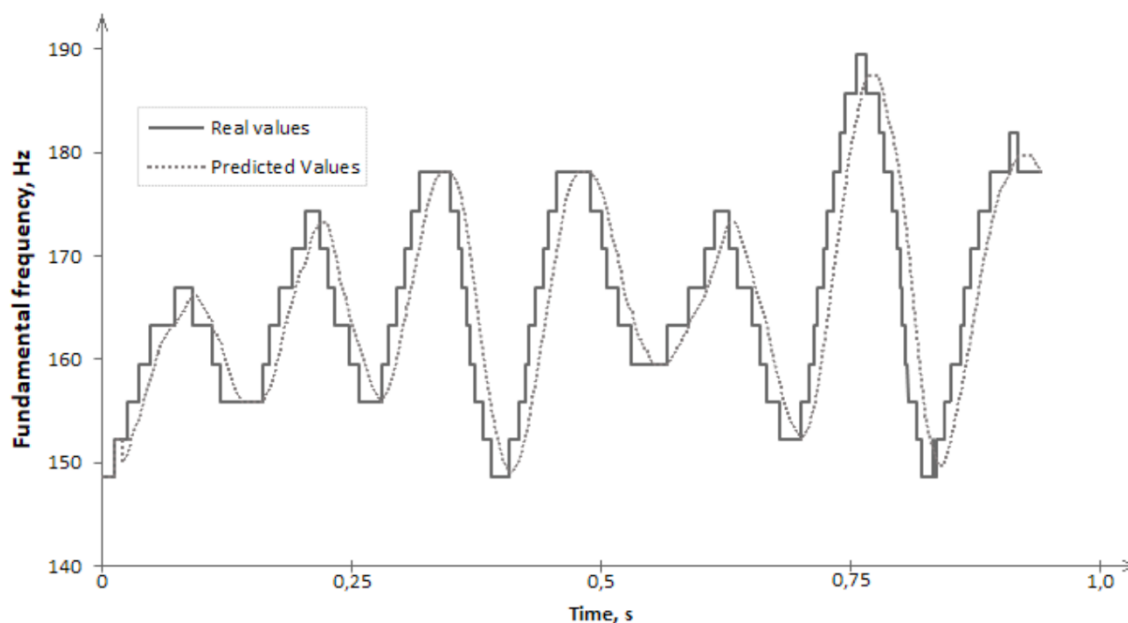


Рисунок 4.4 – Змінна середня для вібраторподібної ділянки в Excel

З метою вирішення даної проблеми була використана класична формула ковзного згладжування з урахуванням вимірювань до і після згладжування, в інтервал згладжування були включені $0,5 * P$ значень, що передують згладжуваного, і стільки ж наступних значень (рисунок 4.5). В результаті на відріжку від 0 до $0,5 * P$ значення рядів збігатимуться, але вже починаючи з наступного значення ми зможемо побачити усереднення. Суцільною лінією позначені значення аналізованих частот основного тону, а переривчастої коливається - згладжені значення.

На представленому рис. для ноти «ми малої октави» паралельними пунктирними лініями вказано нижню (160,121 Гц) та верхню (169,643 Гц) межі звучання, визначені за допомогою середнього логарифмічного [49]. Зазначимо, що в досліджуваному відріжку є 7 коливань у частоті основного тону, якщо судити за графіком. При подальшому дослідженні типу коливань у співі цей аспект може бути застосований як критерій відхилення вібратору від тремоляції.

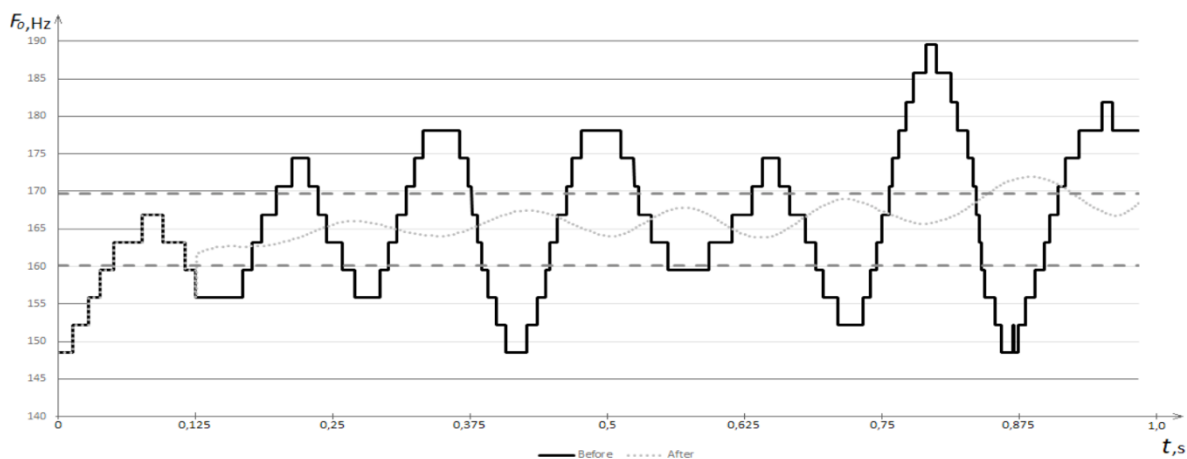


Рисунок 4.5 – Ковзне середнє для вібратоподібної ділянки

Як було зазначено раніше, наявність тренду визначається порівнянням отриманих середніх значень кожної із сукупностей досліджуваного низки. Також засобами Excel на графіці частот основного тону для низки можна встановити побудова лінійного тренда (рисунок 4.6). На етапі аналізу застосовності методу до досліджуваного завдання використовуватимемо вбудовану функцію в Excel, але в етапі реалізації у програмному комплексі – порівняння середніх значень для відрізків досліджуваного діапазону.

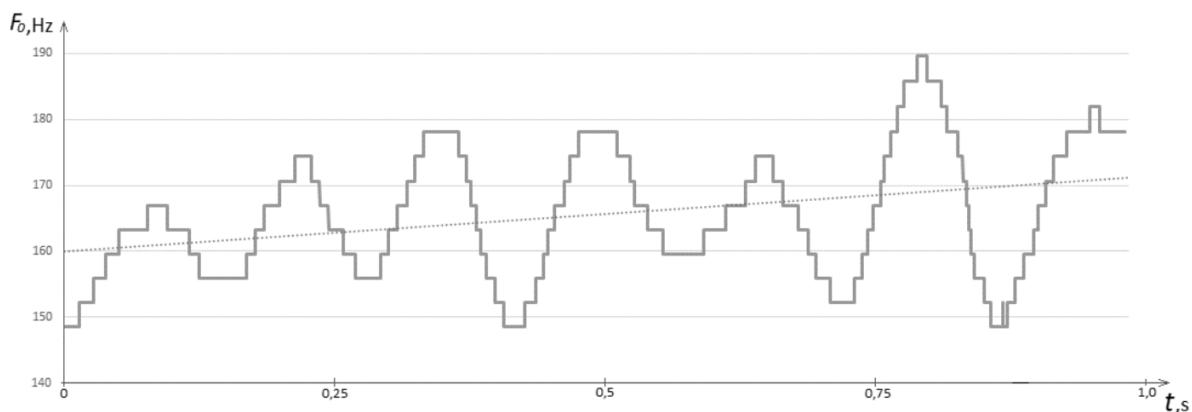


Рисунок 4.6 – Лінійний тренд для вібратоподібної ділянки

Як можна помітити з рисунку, представленого вище, лінійний тренд дозволяє визначити, що на більшій частині досліджуваної ділянки, середні значення частот відносяться до однієї ноти. Однак, за лінійним трендом неможливо оцінити наявність коливань у частотах основного тону. У зв'язку з цим, що це НЕ можливо використовувати тільки лінійний тренд для оцінки областей з tremulation . Тренд дозволяє дати оцінку кількості нот, охоплених на досліджуваному ділянці, але з тип виконання. У свою чергу, кожен середнє одночасно інтерпретує і висвітлює наявність коливань і нот, що робить цей метод кращим для оцінки вібратородібних частин.

4.3 Аналіз глісандуючих переходів у співі

Для визначення глісандо в досліджуваній ділянці необхідно виконати інше завдання. Базовий алгоритм сегментації та ідентифікації нот визначає 2 ноти в подібних ситуаціях: ту, з якої почалося ковзання, та ту, на якій ковзання закінчилося. Відповідно як перша розв'язувана задача для визначення наявності глісандо – порівняння нот на межах досліджуваної ділянки. Якщо ноти виявляться ідентичними, то ділянка між ними не може відповідати глісандо і має бути розглянута на предмет віднесення до вібрато. В інших випадках ми спостерігатимемо лінію тренду, що проходить через кілька нот. Чим більша різниця між нотами і чим швидше був зроблений перехід між ними, тим більше буде кут отриманої лінії тренду щодо осі часу.

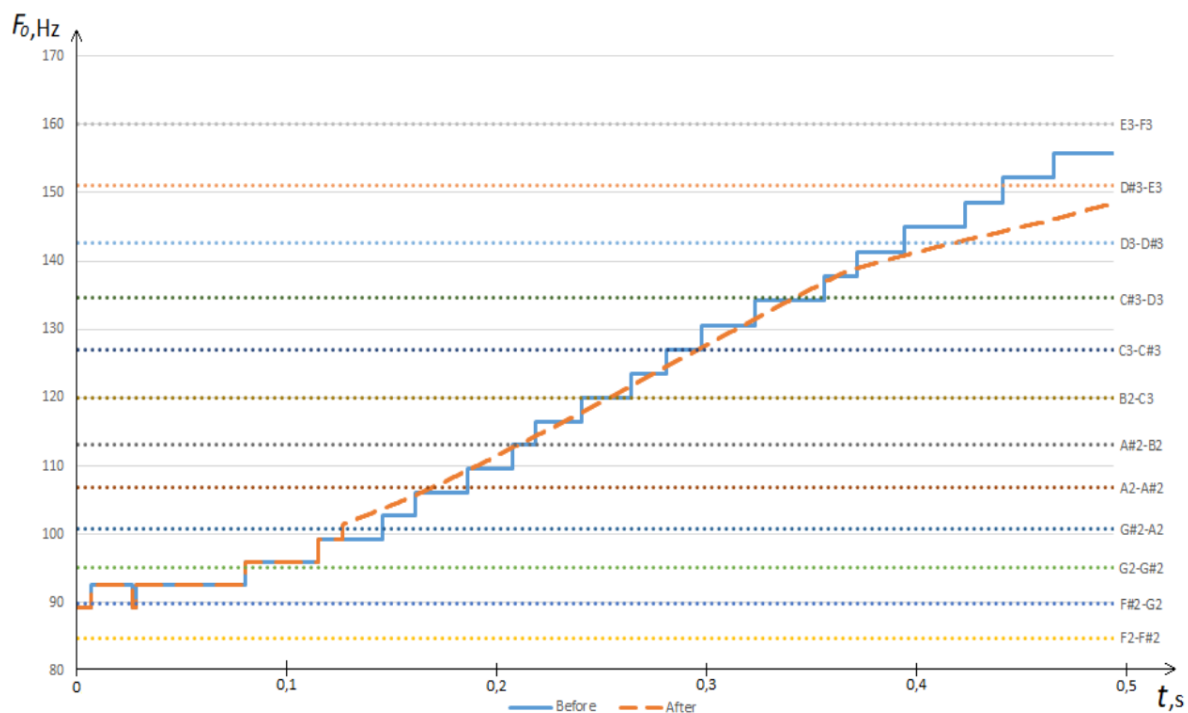


Рисунок 4.7 – Ковзне середнє для ділянки зі висхідним глісандо

Як і ділянки з тремоляцією, розглянемо ковзне середнє для досліджуваного діапазону. Як можна побачити рис. 4.7, у переході від ноти «сінь малої октави» до ноти «ми 1 октави» відсутні явні коливання в частоті основного тону. Суцільною лінією на графіку позначені частоти основного тону, а пунктирною лінією – ковзне середнє для досліджуваного діапазону. Паралельні точкові лінії відповідають межах нот. Отримана лінія ковзного середнього проходить через 9 нот. Таким чином, отримані дані свідчать про висхідне глісандо. Нижче (рис. 4.8) представлений графік для лінійного тренду, отриманого на досліджуваному діапазоні. Фактично, лінійний тренд для досліджуваного діапазону близький до результату, отриманого при побудові ковзного середнього.

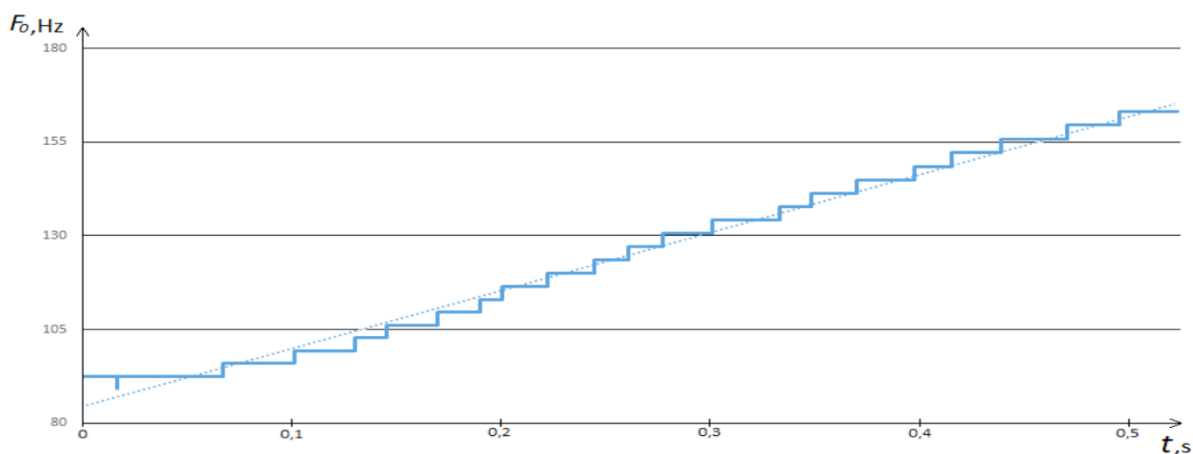


Рисунок 4.8 – Лінійний тренд для ділянки зі висхідним гліссандо

4.4 Порівняння ковзного середнього та лінійного тренду на ділянках з різкою зміною частот основного тону

Як можна помітити з отриманих результатів, при використанні ковзного середнього для ділянок з тремоляцією спостерігається ефект коливання, який відсутній у лінійному тренді для даної ділянки. З іншого боку, для гліссандо рухомих середні і лінійні тренди є самі

Подібним чином. Це дозволяє нам застосовувати обидва методи комплексно для визначення поточної ситуації.

Ідея, що пропонується, полягає в отриманні обох оцінок для досліджуваного відрізка. Отримані набори даних будуть порівнюватися на предмет величини подібності між оцінками деякому параметру N . Розмір величини даного параметра потребує окремого експерименту. В рамках даного експерименту повинні бути розглянуті змішані ситуації, в рамках яких тимчасові ряди поводяться неоднозначно, з метою зниження кількості помилок 1-го та 2-го роду щодо типу досліджуваних діапазонів частот основного тону.

У випадку, якщо в співі було глісандо, результати ковзного середнього та лінійного тренду для даного часового ряду будуть близькі. Для цієї ситуації необхідно буде здійснювати перевірку щодо кількості нот, через які був здійснений перехід, і напрямок переходу (висхідне або низхідне глісандо). В результаті буде можливо відсіювати ділянки з шумом між сусідніми сегментами, що відповідають одній ноті, і ділянки з сплеском шумів, що знаходяться за межами діапазону звучання між нотами.

Для ситуацій, у яких різниця в оцінках часового ряду перевищуватиме величину критерію N , вважатимемо, що відбувався спів з вібрато в голосі. Як було зазначено раніше, ковзне середнє має хвилеподібну структуру у разі наявності коливань в досліджуваній ділянці. Підрахунок кількості коливань в одиницю часу може бути використаний задачі класифікації типу тремоляції. Даний параметр буде корисний під час навчання вокальному виконанню у разі надто частих або рідкісних коливань у виконанні ноти. Крім того, у разі визначення вібрато у співі також буде необхідний контроль кількості нот, через які проходив тренд. Як можна побачити рис. 4.9, коливання для заспіваної ноти проходять зі ковзанням вниз. Фактично, у вказаному прикладі відбувається змішання гліссандо та вібрато. Однак, ні домодному з типів цю ділянку віднести неможливо по двох причин: для гліссандо не характерно наявність коливань у голосі, а вібрато повинно бути ковзань іншу ноту.

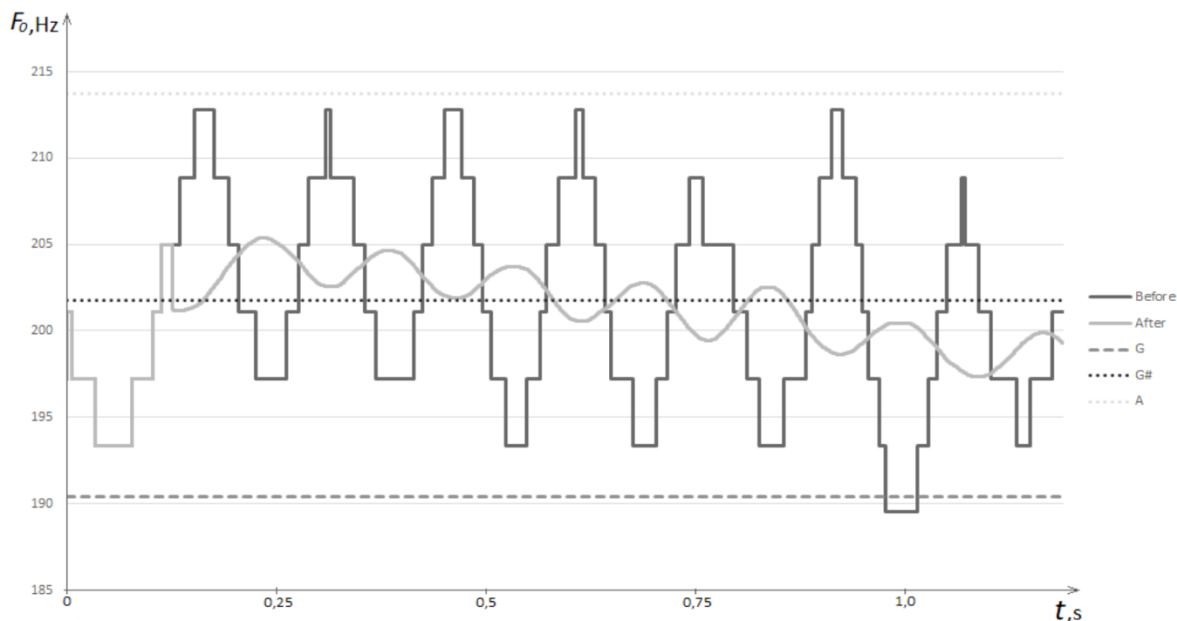


Рисунок 4.9 – Ковзаюча середня для комбінованої ділянки

4.5 Експериментальне тестування

Як в матеріалі для обчислення значення ковзаючого середнього та лінійного тренда, аудіо запис з вокальним виконанням, зазначена в пункті 3.2, були розглянуті. Кожна з аудіозаписів була сегментована на окремі ноти та ділянки, у межах яких не було ідентифіковано ноти. Окремо було виділено ділянки з вібраторподібним або гліссандуючим співом. На рис. 4.10 наведено приклад з 4 виділеними ділянками, що відповідають співу з вібратором в голосі. Як можна помітити, ділянка 1 (рисунок 4.11) коливання відбуваються в межах однієї ноти, що дозволило ідентифікувати ноту. Така ситуація спостерігається на 4 ділянці.

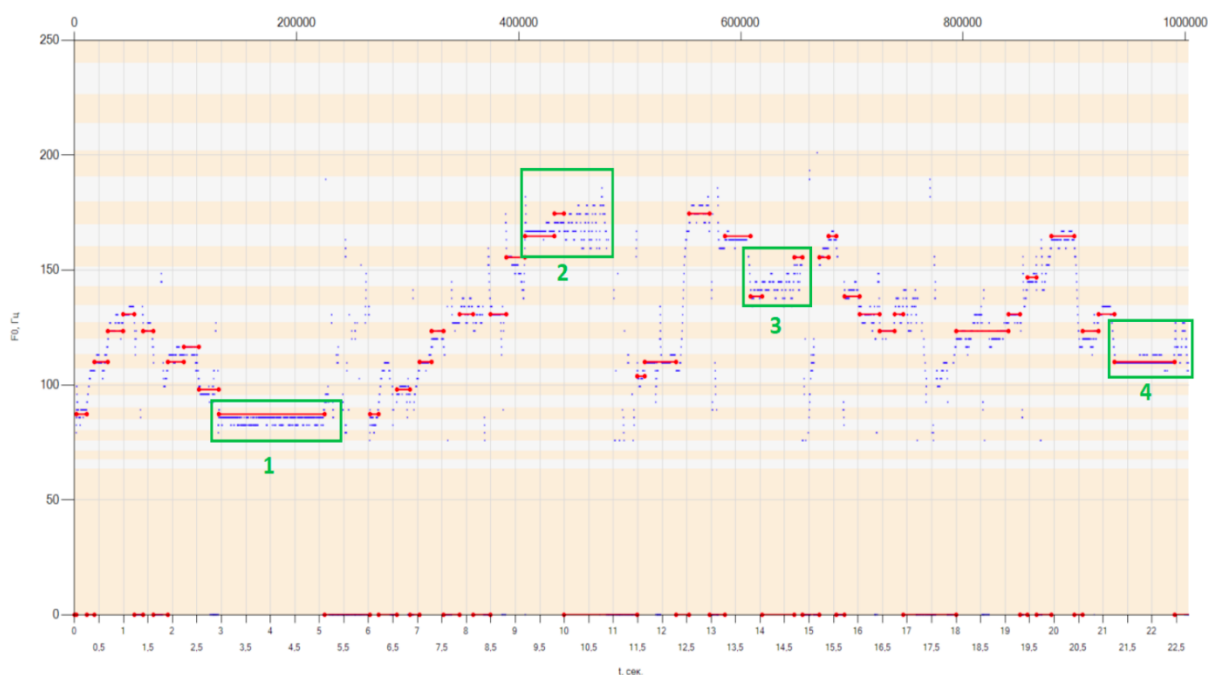


Рисунок 4.10 – Частотна область з розпізнаними нотами для досліджуваного запису «Вийду вночі в поле з конем» заспіваним диктором №9

Незважаючи на те, що наявність коливань у голосі не вплинула на точність розпізнавання нот у розглянутих ситуаціях, було вирішено повторно досліджувати подібні ділянки з метою пошуку можливих відмінностей, що дозволяють класифікувати схожі сегменти.

Це допоможе не тільки точніше розпізнавати ноти з коливаннями, але й врахувати під час виставлення оцінки якості співу навмисність переходів на сусідні ноти. Крім того, схожі сегменти не будуть втрачені на етапі аналізу типу вібраторподібного виконання.

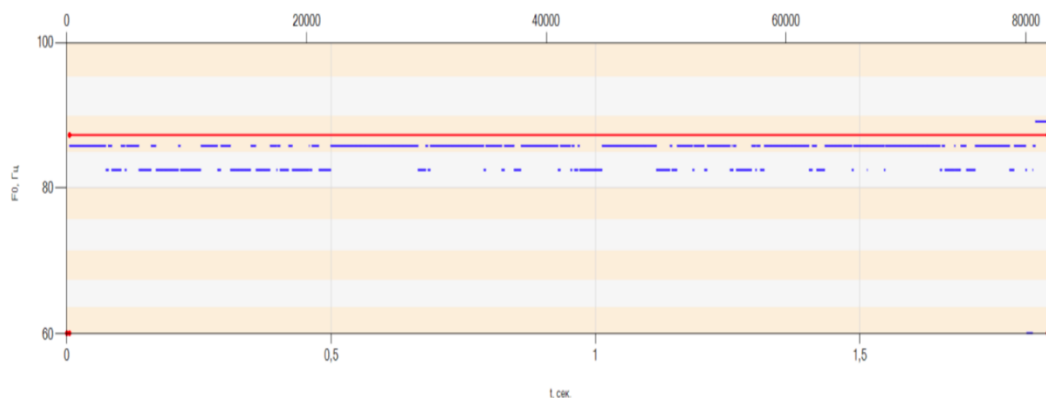


Рисунок 4.11 – Сегмент №1 досліджуваного аудіозапису «Вийду вночі в поле з конем» заспіваним диктором №9

На рис. 4.12 представлений приклад із сегментованими ділянками, що містять висхідне глісандо. Як можна помітити для випадків 1 і 3 алгоритм розпізнавання нот вдалося ідентифікувати початкове і кінцеве значення нот, між якими відбувається ковзання, а частоти між ними сприйняті як шум. Для 2-го випадку початковий момент висхідного глісандо також є кінцевим для того, що сходить від попередньої ноти.

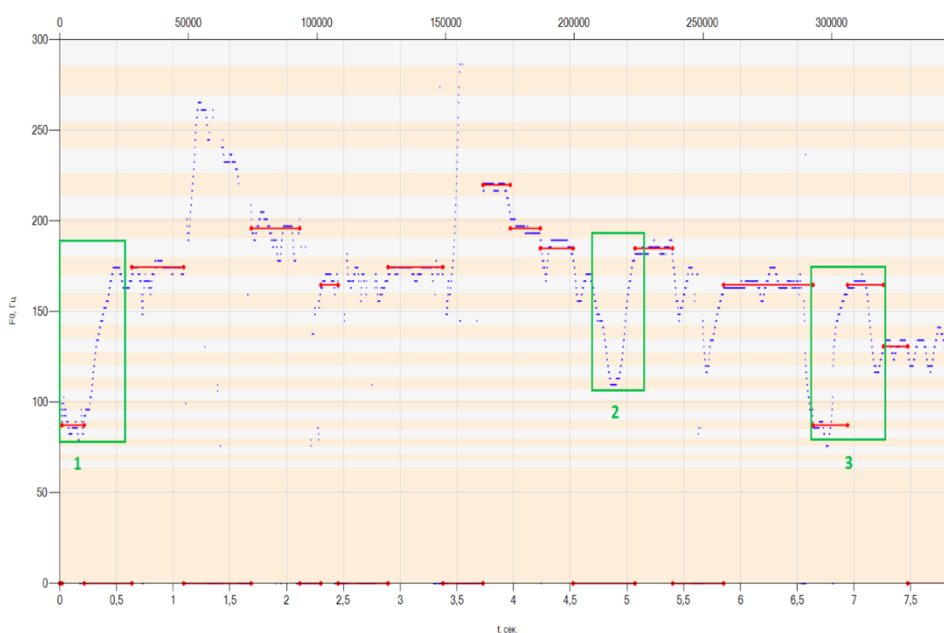


Рисунок 4.12 – Частотна область з розпізнаними нотами для досліджуваного запису №8 заспіваного диктором №5

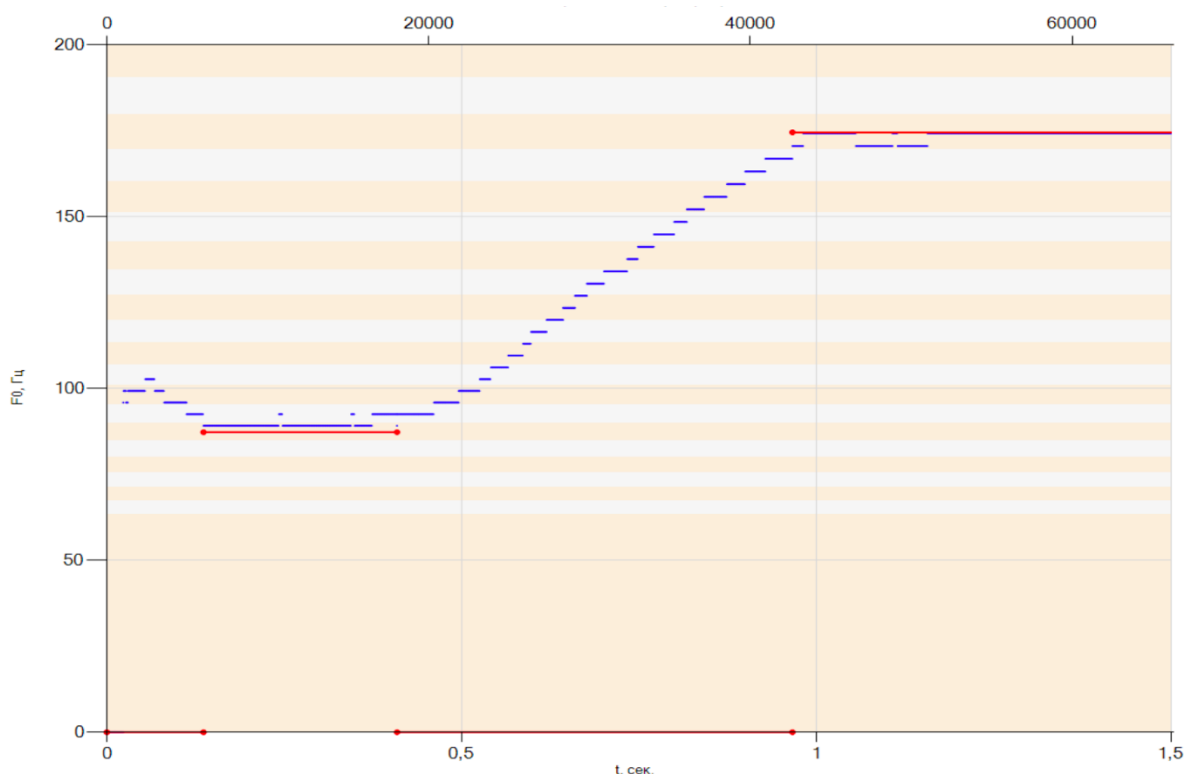


Рисунок 4.13 – Сегмент №3 для досліджуваного запису №8
заспіваного диктором №5

Загалом із 740 аудіозаписів було сегментовано 149 ділянок з вібраторподібним співом та 68 ділянок з глісандо, з яких 54 з висхідним та 14 з низхідним. Для кожного аудіозапису були отримані значення частот основного тону, які були попередньо оцінені вручну Excel з застосуванням описаних вище функцій.

В рамках програмного комплексу [53] було розроблено модуль, який відповідає за аналіз вибраного діапазону частот основного тону. Алгоритм роботи модуля представлений рис. 4.14. На вхід до модуля надходить масив із частотами основного тону сегмента, нерозпізаного алгоритмом розпізнавання нот і аналізованого на предмет наявності тремоляції або переходу глісандо між нотами. Крім того, необхідно повідомити модулю

інформацію про мінімальну тривалість звучання нот, що відповідає аналізованого аудіозапису.

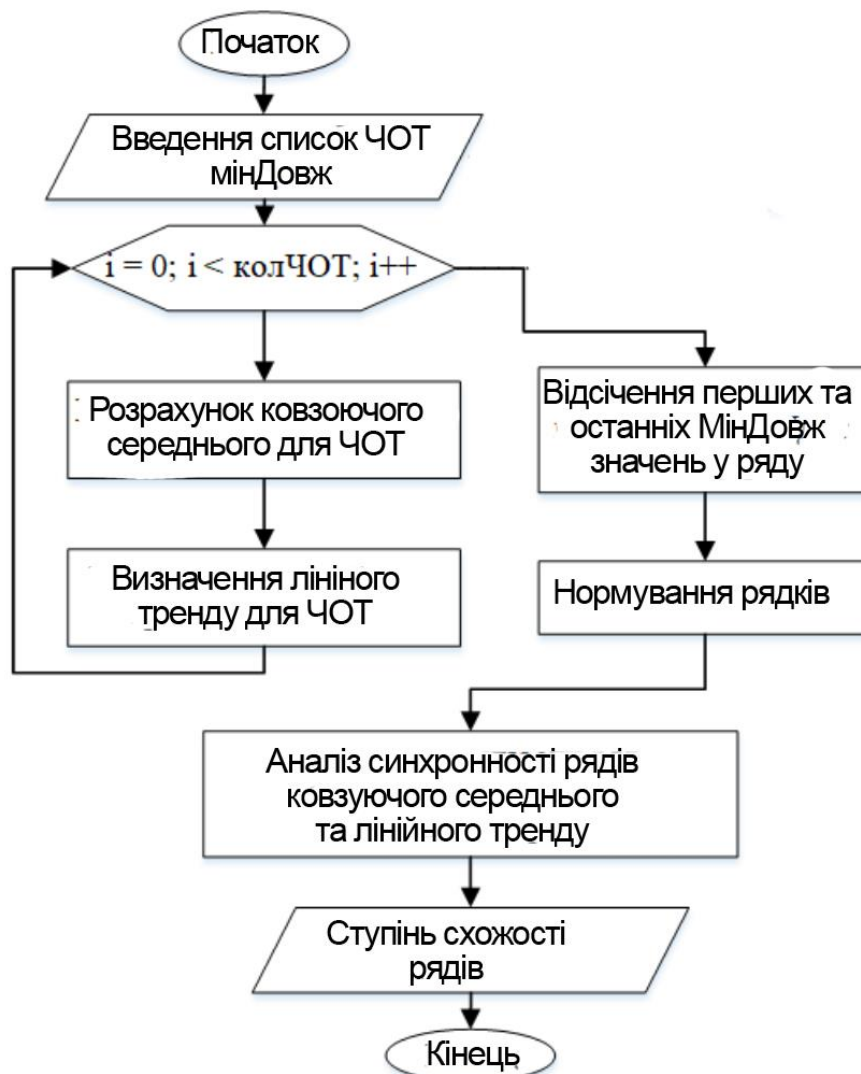


Рисунок 4.14 – Алгоритм роботи модуля аналізу сегментів із різкою зміною частот основного тону

Для досліджуваного масиву частот формується масив зі значеннями ковзного середнього в кожний час досліджуваного діапазону. При обчисленні значень лінійного тренду в кожний момент часу використано метод найменших квадратів, представлений формулою

$$a \cdot \sum x_i^2 + b \cdot \sum x_i = \sum y_i \cdot x_i$$

$$\{$$

$$a \cdot \sum x_i + b \cdot n = \sum y_i$$

Так як вихідні і кінцеві p спостереження змінного середнього залишалися рівними початкових значення частот основного тону, серія змінного середнього і в лінійній тенденції була відрізана від p з перших і останніх значень. На рис. 4.15 наведено приклад графіка з аналізом частот основного тону, на якому позначені точками А та В межі подальшого аналізу даних.

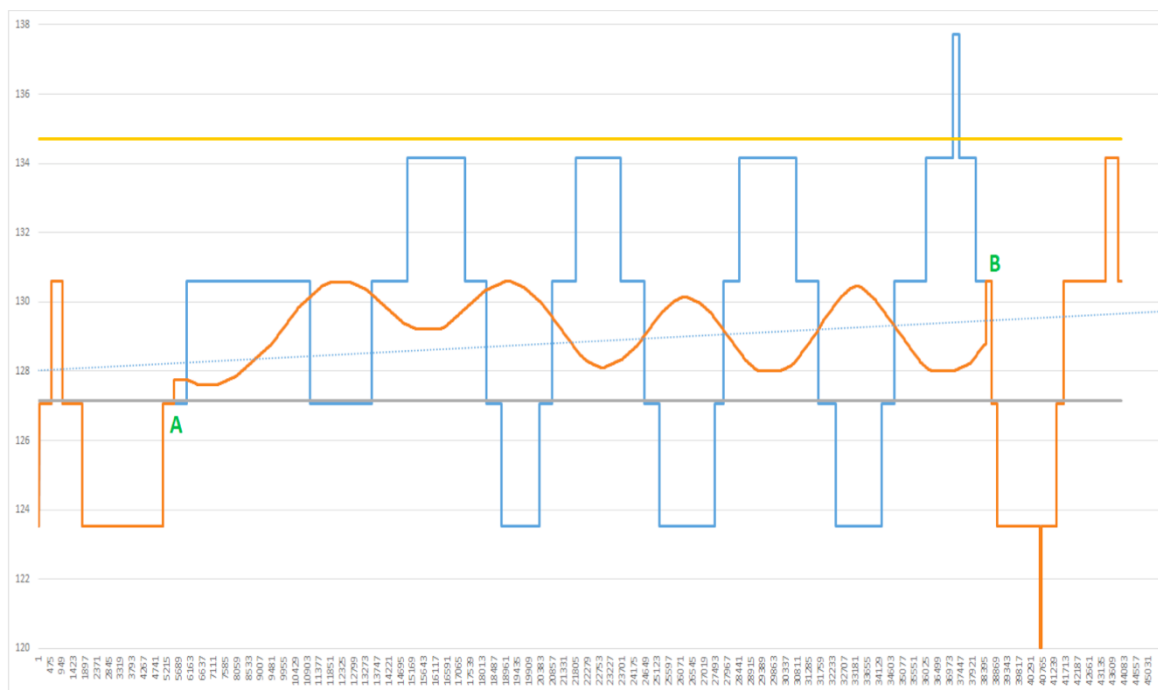


Рисунок 4.15 – Приклад графіків отриманих оцінок

Значення, що залишилися, проходили нормування і оброблялися за допомогою методу виділення синхронності [60]. Зазначений метод дозволяє визначити ступінь подібності в динаміці декількох сигналів, що розглядаються. Раніше цей метод вже застосовувався до вокальних виконань у рамках обчислення подібності між декількома дикторами і

дозволив доповнити базу вправ співом за заданим звуком, де від студента потрібно було повторювати за викладачем мелодійний мотив.

Запуск оцінки подібності часових рядів шляхом виділення синхронності здійснюється викликом підпрограми. На рис. 4.16 наведено приклад результату, що отримується від підпрограми.

Для кожного виділеного аудіозапису було обчислено частоти основного тону та проаналізовано за допомогою розробленого модуля. Результати експерименту (фрагмент) представлені у таблиці 4.1.

Таблиця 4.1 – фрагмент результатів експерименту виділення синхронності

Номер запису	Сегментований з	Тип співу	ЧОТ	Ступінь відмінності
001	Бременські01.wav	вібрато	31985	67,78%
002	Бременські01.wav	вібрато	41473	73,2%
003	Бременські01.wav	вібрато	47265	68,36%
004	Бременські01.wav	змішаний	36071	48,7%
005	Бременські02.wav	змішаний	48217	49,9%
006	Бременські02.wav	вібрато	32969	80,4%
007	Бременські02.wav	вібрато	32170	68,16%
008	Конь01.wav	вібрато	47424	81,44%
009	Конь01.wav	глісандо	59169	15,04%

Аналіз 217 аудіозаписів показав, що для глісандо характерно від 5 до 15% показник синхронності між лінійним трендом і ковзним середнім, а для співу вібрато – діапазон від 65 до 85%. При цьому чим чистішим відбувається спів вібрато – тим вищий відсоток розбіжності між рядами. Для глісандо – навпаки, що менше відсоток, то менше було побічних частот під час переходу в співі. Крім цього було оброблено 67 записів з комбінуванням вібрато та глісандо у співі. Змішані види співу, описані у розділі 4.4, сприймаються методом як на 45-50% різні.

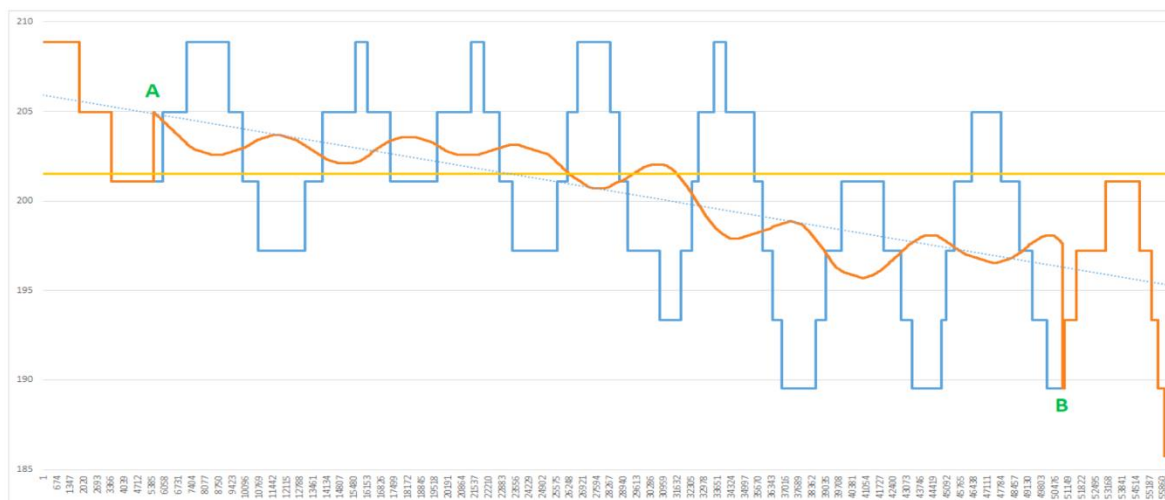


Рисунок 4.17 – Графік з сегментом, що не ідентифікується.

На рис. 4.17 наведено приклад, в якому суміш вібрато в співі з переходом на нижчу ноту призводять до неможливості ідентифікувати окремо кожен із застосованих методів співу. Наявність тренду до зниження або підвищення загального звучання сегмента в частотній області призводить до більш плавної зміни ковзної середньої для досліджуваної ділянки.

Отримані результати дозволяють автоматизувати оцінку ділянок, нерозпізнаних базовим алгоритмом розпізнавання нот у вокальному виконанні. Поділивши діапазон прийнятих результатів від методу поділу синхронності часових рядів на заголовки, можна однозначно класифікувати області з різкою зміною частоти основного тону. У разі, якщо різниця буде становити НЕ більш 15% можна припустити на присутності гліссандо в співі обробленої частини. При оцінках в діапазоні від 65% до 85% буде вважатися, що досліджуваний відрізок був проспіваний з вібратоподібними вібраціями. Проміжні значення (від 16 до 64%) будуть сприйматися як шум.

ВИСНОВОК

У роботі вирішено завдання підвищення якості розпізнавання нот, що звучать у вокальному виконанні за рахунок розробки методу ідентифікації сегментів з різкою зміною частоти основного тону.

Зроблено огляд існуючих методів та алгоритмів розпізнавання нот, у тому числі визначення частот основного тону. Було визначено, що існуючі алгоритми аналізу частоти основного тону не застосовуються до вокальних виконань з 2 причин: високий відсоток грубих помилок та обмеження смуги дослідження діапазоном до 400 Гц.

Описано алгоритм розпізнавання нот, що складається з етапу ідентифікації нот у кожний момент часу з їхньою наступною сегментацією на підставі мінімальної значення тривалості звучання ноти. Як мінімальний розбіжності в алгоритмі був використаний облік мінімальної тривалості звучання ноти.

Програмний комплекс аналізу вокальних виконань протестований на аудіозаписах із різними підходами до вокального виконання. Результати експерименту показали, що при аналізі аудіозаписів вокального виконання, що містять виконання із застосуванням стаккато, легато, арпеджіо, крещендо та декрещендо, алгоритм розпізнав безпомилково не менше 95% нот. Програмний комплекс було оцінено щодо частоти помилок у роботі. Імовірність 0,95 частота виникнення помилок не перевищує 3.3%.

Запропоновано метод ідентифікації сегментів з різкою зміною частоти основного тону, що ґрунтується на застосуванні методу виділення синхронності до оцінок, отриманих на підставі масиву частот основного тону. Експериментально було визначено, що ковзне середнє та лінійний тренд для співу з вібратородібними коливаннями характеризується синхронністю в діапазоні від 65 до 85%, а переходи глісанді від однієї ноти до іншої відрізняються менш ніж на 15%.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Kharchenko S.S. Fundamental frequency evaluation subsystem for natural speech rehabilitation software calculation module for cancer patients after larynx resection /Kharchenko S.S., Mescheryakov R.V., Volf D.A., Balatskaya L.N., Choinzonov E.L.// Proceedings - 2015 International Conference on Biomedical Engineering and Computational Technologies, SIBIRCON 2015. – 2015. – P. 197- 200.
2. Балацкая Л.Н. Речевая реабилитация и качество жизни после хирургического лечения больных раком гортани//Сибирский онкологический журнал. – 2003. – № 2. – С. 54-57.
3. Kostuchenko E., Assessment of Syllable Intelligibility Based on Convolutional Neural Networks for Speech Rehabilitation After Speech Organs Surgical Interventions /Kostuchenko E., Novokhrestova D., Pekarskikh S., Shelupanov A., Nemirovich-Danchenko M., Choynzonov E., Balatskaya L.// SPECOM 2019: Speech and Computer. – 2019. – P. 359-369.
4. Dietz J.H. Adaptive rehabilitation in cancer: A program to improve quality of survival // Postrad. Med. –1980. – Vol. 68. – P. 145-163.
5. Федотова М.В. Мелодическая структура восходяще-нисходящего тона как маркер валлийского акцента в английском языке // Вестник Московского государственного лингвистического университета. Гуманитарные науки. – 2011. – № 607. – С. 233-244.
6. Жаровская Е.В. Просодические особенности речи молодежи // Филологические науки. Вопросы теории и практики. – 2018. – № 8-1 (86). – С. 95-99.
7. Сокорева Т.В. Роль высотно-мелодического компонента в сохранении и развитии ритмических тенденций // Вестник Московского государственного лингвистического университета. Гуманитарные науки. – 2017. – № 771. – С. 105-117.

8. Жаровская Е.В. Характеристика элементов мелодического рисунка речи // Филологические науки. Вопросы теории и практики. – 2017.

–

№ 7-3 (73). – С. 112-114.

9. Шук С.В. Акустические признаки позитивной и негативной оценки в британском радиорепортаже // Вестник Полоцкого государственного университета. Серия А: Гуманитарные науки. – 2011. – № 10. – С. 78-82.

10. Murthy Y.V.S. , Koolagudi S.G. Classification of Vocal and Non-vocal segments in Audio Clips using Genetic Algorithm based Feature Selection (GAFS)

// Expert Systems with Applications. – 2018. – Vol. 106. – P. 77-91.

11. Finley, Michael & Razi, Abolfazl. Musical Key Estimation with Unsupervised Pattern Recognition. // 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). – 2019. – P. 401-408.

12. Bader R. Computational Music Archiving as Physical Culture Theory

// Computational Phonogram Archiving. Current Research in Systematic Musicology. – Vol 5. – Springer, Cham. – 2019. – P. 3-34.

13. McFee B., Wook K.J., Cartwright M., Salamon J. M., Bittner R., Pablo

B. J. Open-Source Practices for Music Signal Processing Research: Recommendations for Transparent, Sustainable, and Reproducible Audio Research

// IEEE Signal Processing Magazine. – 2019. – Vol. 36. – P. 128-137.

14. Глазырин Н.Ю. О задаче распознавания аккордов в цифровых звукозаписях // Известия Иркутского государственного университета. Серия: Математика. – 2013. – Т. 6. – № 2. – С. 2-17.

15. Masataka Goto. A real-time music-scene-description system:

predominant-F0 estimation for detecting melody and bass lines in real-world audio signals // *Speech Communication*. – Vol. 43. – Issue 4. – 2004. – P. 311-329.

16. Способин И.В. Элементарная теория музыки. — М.: Музыка, 1968.

– 204 с.

17. Тюлин Ю.Н. Краткий теоретический курс гармонии. — М.: Музыка, 1978. – 212 с.

18. Искусство пения: Учебное пособие. — 4-е изд., стер. — СПб.: Издательство «Лань»; Издательство «ПЛАНЕТА МУЗЫКИ», 2019. — 212 с.: ил., ноты. — (Учебники для вузов. Специальная литература).

19. Aronson, Arnold Elvin; Bless, Diane M. *Clinical Voice Disorders* (4th ed.). New York, NY: Thieme Medical Publishers. – 2009. – 278 p.

20. Фант Г. Анализ и синтез речи / Г. Фант. – Новосибирск: Наука. – 1970. – 306 с.

21. Шарий, Т. В. О проблеме параметризации речевого сигнала в современных системах распознавания речи / Т. В. Шарий // *Вісник Донецького національного університету*. – Сер. А: Природничі науки. – Вип. 2. – 2008. – С. 536–541.

22. Рабинер Р. Л. Цифровая обработка речевых сигналов / Р. Л. Рабинер, Р. В. Шафер – М.: Радио и связь. – 1981. – 496 с

23. Матвеев Ю.Н., Симончик К.К., Тропченко А.Ю., Хитров М.В. Цифровая обработка сигналов // Учебное пособие: СПбНИУ ИТМО. – 2013. – 166 с.

24. Azarov E. Instantaneous pitch estimation based on RAPT framework /

E. Azarov, M. Vashkevich, A. Petrovsky // *Proceedings of EUSIPCO'12* — European Signal Processing Conference. – Bucharest, Romania – August 27-31, 2012. — P. 2787-2791.

25. Talkin D. A robust algorithm for pitch tracking (RAPT) // *Speech Coding and Synthesis*, W. B. Kleijn and K. Paliwal, Eds. New York: Elsevier. – 1995. — P. 495-518.

26. Вашкевич М.И., Азаров И.С., Петровский А.А. Оценка мгновенной частоты основного тона речевого сигнала на основе многоскоростной обработки // *Речевые технологии*. – 2018. – № 1-2. – С. 12-24.

27. Gonzalez S. PEFAC — A Pitch Estimation Algorithm Robust to High Levels of Noise / S. Gonzalez, M. Brookes // *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. – 2014. – Vol. 22. – No.2. – P. 518-530.

28. Гитлин В.Б. Выделение основного тона речи методом SWIPE из сигнала, ограниченного полосой телефонного канала / В.Б. Гитлин, Д.Ю. Вашурин // *Речевые технологии*. – 2014. – №1. – С. 57-74.

29. De Cheveigné, A., Kawahara, H. YIN, a fundamental frequency estimator for speech and music // *The Journal of the Acoustical Society of America*. 2002. Vol. 111. [Электронный ресурс]. – Режим доступа: <http://asa.scitation.org/doi/abs/10.1121/1.1458024>.

30. Вольф Д.А. Модель, численная и программная реализация оценивания частоты основного тона речевого сигнала с помощью сингулярного спектрального анализа: дис. ... канд. техн. наук. – Томск, 2015.

– 149 с.

31. Морозов В.П. Компьютерная диагностика вокальной одаренности

// *Голос и речь*. – 2010. – № 1. – С. 81-93.

32. Leydon C., Bauer J.J., Larson C.R. The role of auditory feedback in sustaining vocal vibrato // *Acoustical Society of America*. – Vol. 114(3). – 2003. – P. 1575-1581.