# 2. Розробка програмного забезпечення

Змн.

Арк.

№ документа

Підпис

Дата

Аркуш

КНУ.РБ.6050102.03.33.ВС

Розробив

Перевірив

Н.контроль

Музика І.О.

Затвердив

Купін А.І.

Літера

Аркушів

КСМ-10с

## 2.1 Алгоритм розпізнавання голосу

Навіть при роботі зі звуком з мінімально достатньою частотою дискретизації (16 кГц) розмір унікальних характеристик для секундного зразка звуку просто величезний - 16000 значень амплітуд. Виробляти скільки-небудь складні операції над такими об'ємами даних не представляється можливим. Крім того, не зовсім зрозуміло, як порівнювати об'єкти з різною кількістю унікальних рис.

Для початку знизимо обчислювальну складність завдання, розбивши її на менші за складністю підзадачі. Цим ходом вбиваємо відразу двох зайців, адже встановивши фіксований розмір підзадачі і усереднивши результати обчислень по всіх завдань, отримаємо наперед задану кількість ознак для класифікації.

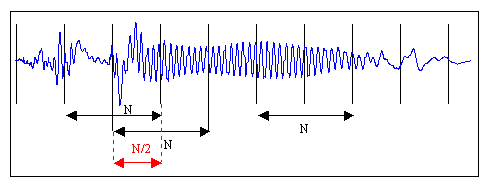


Рисунок 2.1 – розбиття довготривалого сигналу на «вікна»

На рисунку 2.1 зображена «порізка» звукового сигналу на кадри довжини N з половинною перекриттям. Необхідність у перекритті викликана спотворенням звуку у випадку, якби кадри були розташовані поруч. Хоча на практиці цим прийомом часто прінебрегают для економії обчислювальних ресурсів. Слідуючи рекомендаціям, виберемо довжину кадру рівній у налаштуваннях, як компроміс між точністю (довгі кадри) і швидкістю (короткі кадри). Залишок мови, який не займає повний кадр, можна заповнити нулями до бажаного розміру або просто відкинути.

Для усунення нежелаетльних ефектів при подальшій обробці кадрів, помножимо кожен елемент кадру на особливу вагову функцію («вікно»). Результатом стане виділення центральній частині кадру і плавне загасання амплітуд на його краях. Це необхідно для досягнення кращих результатів при прогоні перетворення Фур'є, оскільки воно орієнтоване на нескінченно повторюється сигнал. Відповідно, наш кадр повинен стикуватися сам з собою і як можна більш плавно. Вікон існує безліч.

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

Існує безліч математичних формул візуально поліпшують частотний спектр на розриві меж вікна. Для цього застосовуються перетворення: прямокутне (ніяке), трикутне, звужується косинус, фрагмент синусоїди, синус в кубі, синус 4-го ступеня, перетворення Парзена, Велша, Гаусса, Хеннінга, Хеммінга, Чебишева, з пульсаціями, Розенфілд, Блекмана-Харріса, горизонтальне і з плоскою вершиною. Також існує методика за взаємною перекриттю вікон, при цьому зазвичай можна вибрати скільки семплів з попереднього вікна буде усереднено з поточним вікном.

Прямокутне вікно виходить автоматично при обмеженні вибірки N відліками. Максимальний рівень бічних пелюсток частотної характеристики: -13 дБ

w(n)=\left\{ \begin{matrix}
   1 , & n\in[0,N-1] \\
   0, & n\notin[0,N-1]\\
\end{matrix} \right.

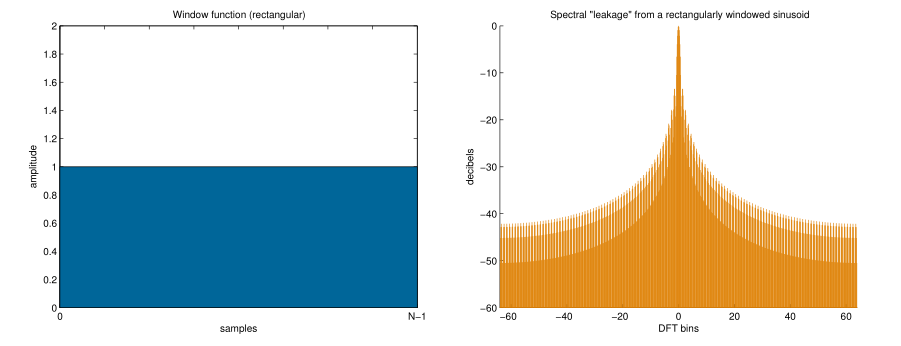


Рисунок 2.2 – прямокутне вікно та спектр перетворення Фур’є для гармонічного сигналу

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

Вікно Ханна(Ханнінга) – рівень бокових пелюсток – -31,5 дБ.

w(n) = 0.5\; \left(1 - \cos \left ( \frac{2 \pi n}{N-1} \right) \right)

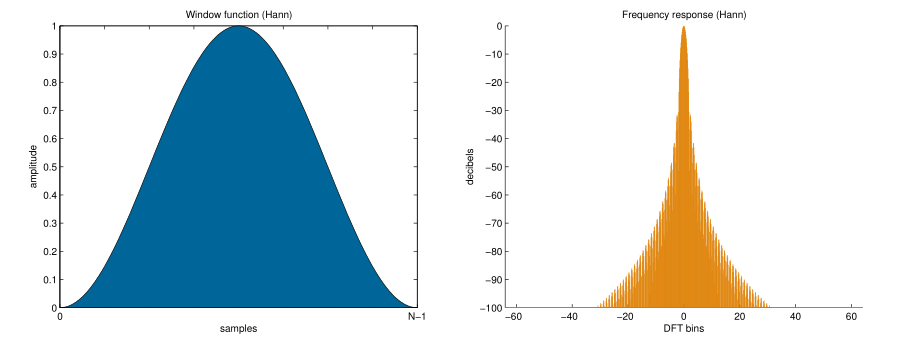


Рисунок 2.3 – вікно Ханнінга та спектр перетворення Фур’є для гармонічного сигналу

Вікно Хемінга – рівень бокових пелюсток – -42 дБ.

w(n)=0.53836 - 0.46164\; \cos \left ( \frac{2\pi n}{N-1} \right)

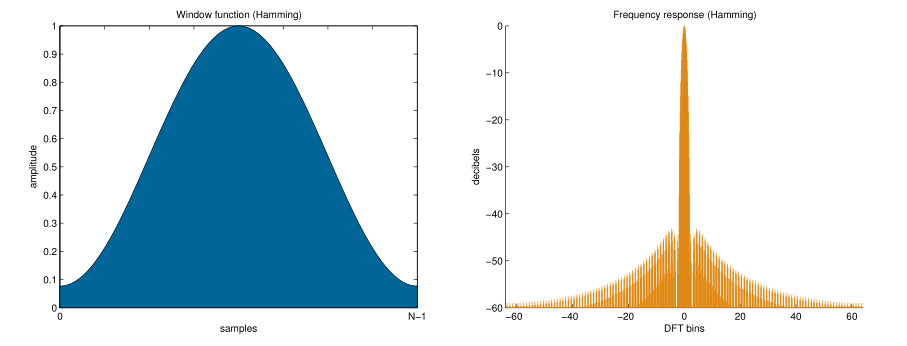


Рисунок 2.4 – вікно Хеммінга та спектр перетворення Фур’є для гармонічного сигналу

Вікно Блекмана – рівень бокових пелюсток – -58 дБ(α = 0,16).

w(n)=a_0 - a_1 \cos \left ( \frac{2 \pi n}{N-1} \right) + a_2 \cos \left ( \frac{4 \pi n}{N-1} \right)

a_0=\frac{1-\alpha}{2};\quad a_1=\frac{1}{2};\quad a_2=\frac{\alpha}{2}\,

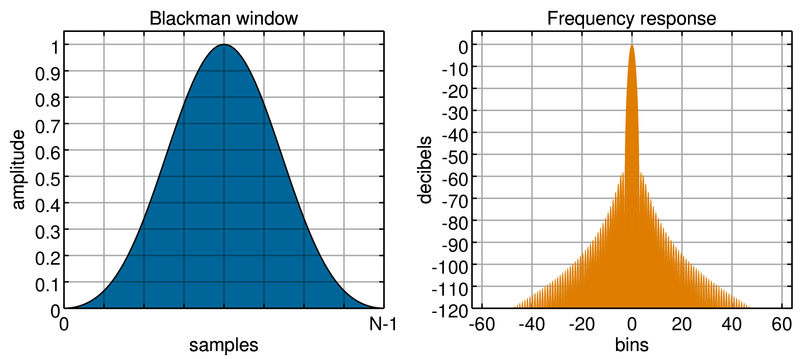


Рисунок 2.5 – вікно Блекмана та спектр перетворення Фур’є для гармонічного сигналу

Вікно Кайзера – рівень бокових пелюсток – -31,5 дБ.

Арк.

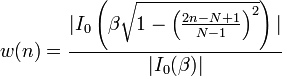
№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС



де I0 - модифікована функція Бесселя першого роду нульового порядку; β - коефіцієнт визначає частку енергії, зосередженої в головному пелюстці спектра віконної функції. Чим більше β тим більше частка енергії, і ширше головний пелюсток, і менше рівень бічних пелюсток. На практиці використовуються значення від 4 до 9.

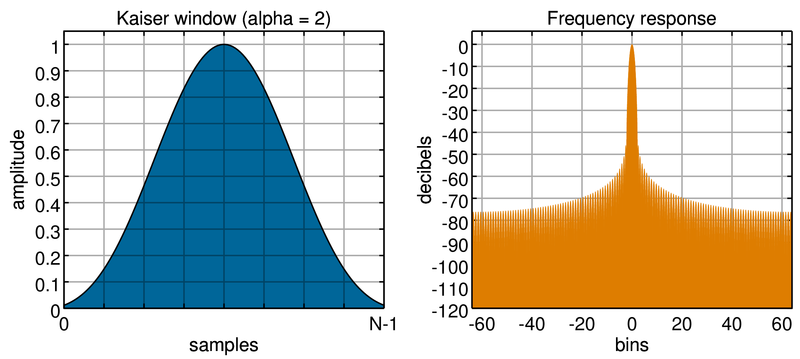


Рисунок 2.6 – вікно Кайзера та спектр перетворення Фур’є для гармонічного сигналу

Ми ж будемо використовувати вікно Хеммінга.

Наступним кроком буде отримання короткочасної спектрограми кожного кадру окремо. Для цих цілей використовуємо дискретне перетворення Фур'є.

http://habrastorage.org/storage2/fd8/9fa/dad/fd89fadad17dab81b8507f289a5f8a52.png

На сьогоднішній день найбільш успішними є системи розпізнавання голосу, що використовують знання про пристрій слухового апарату. Якщо говорити коротко, то вухо інтерпретує звуки нелінійно, а в логарифмічному масштабі. До цих пір всі операції ми проробляли над «Герцами», тепер перейдемо до «мелів».

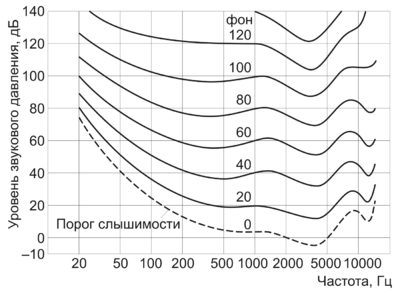


Рисунок 2.7 – чутливість людського вуха до звуків

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

Аналогічно, сприйнята людським слухом висота звуку не зовсім лінійно залежить від його частоти.

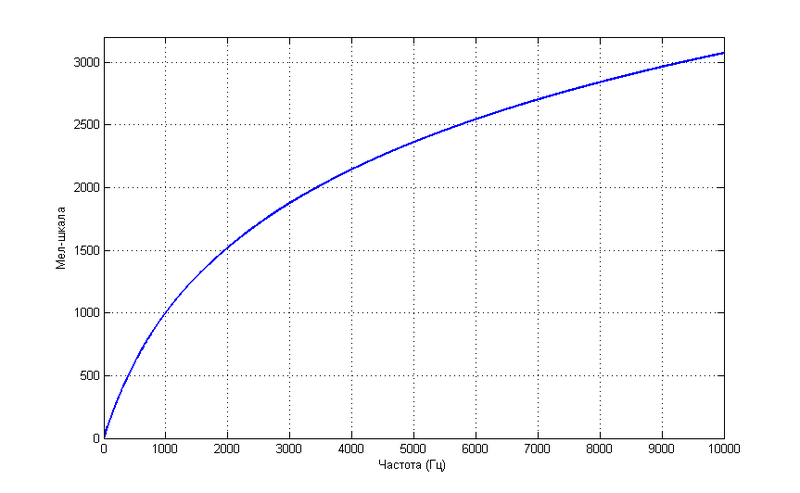


Рисунок 2.8 – графік відповідності мел-шкали до лінійної

Така залежність не претендує на більшу точність, але зате описується простою формулою:

http://habrastorage.org/storage2/60f/d6d/42a/60fd6d42aa74673955637ae5cbf5271f.gif

Подібні одиниці виміру часто використовують при вирішенні задач розпізнавання, так як вони дозволяють наблизитися до механізмів людського сприйняття, яке поки що лідирує серед відомих систем розпізнавання мови.

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

Отже складемо алгоритм розпізнавання голосу.

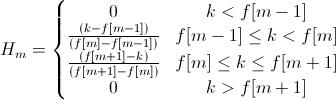
1. Вихідний мовний сигнал запишемо в дискретному вигляді як

http://habrastorage.org/storage2/d94/3c9/aaa/d943c9aaa0edb6e9f628fb607a8388eb.gif

2. Застосовуємо до нього перетворення Фур'є

http://habrastorage.org/storage2/f4f/0bc/556/f4f0bc556ff9fff89ade82d5412067d1.gif

 3. Складаємо гребінку фільтрів, використовуючи віконну функцію



  4. Для якої частоти f [m] отримуємо з рівності

http://habrastorage.org/storage2/e50/0f9/489/e500f9489367addaa6d0ed4b482bcbd9.gif

B (b) - перетворення значення частоти в мел-шкалу, відповідно,

http://habrastorage.org/storage2/3ae/bba/1b3/3aebba1b385e3fb449405829afb33889.gif

5. Обчислюємо енергію для кожного вікна

http://habrastorage.org/storage2/71c/b7e/adb/71cb7eadb7a222e6b65409227d65bcb1.gif

  6. Застосовуємо ДКП

http://habrastorage.org/storage2/082/bbc/14c/082bbc14c78e0847e9080a7d30763e14.gif

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

  7. Отримуємо набір MFCC

Источник:

Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, Spoken Language Processing: A Guide to Theory, Algorithm, and System Development, Prentice Hall, 2001, ISBN:0130226165

## 2.2 Створення системи запису голосу

Для запису звуку з мікрофона використаємо бібліотеку NAudio. Створимо об’єкт звукозапису який буде містити потрібні методи:

Таблиця 2.1 – Опис класу об’єктів аудіозапису

|  |  |  |  |
| --- | --- | --- | --- |
| № | Назва методу | Повертаючий тип | Агрументи |
| 1 | StartRecord | void | - |
| 2 | StopRecord | void | - |
| 3 | GetMels | Double[] | - |
| 4 | Visulization | Bitmap | Розмір зображення |

Метод StartRecord починає запис, та записує звукові данні до єдиного буферу для подальшого аналізу.

Метод StopRecord закінчує запис звуку й починає його обробку для подальшого зберігання оброблених данних.

Метод GetMels повертає набір потужностей для вказаного у налаштуваннях набору мел-фільтрів.

Метод Visulization повертає зображення спактру останнього записаного вікна у вигляді матриці пікселів.

До налаштувань програми входять такі параметри:

Таблиця 2.2 – Опис класу налаштувань

|  |  |  |  |
| --- | --- | --- | --- |
| № | Назва | Тип | Опис |
| 1 | BufferMilliseconds | int | Довжина аудіобуферу |
| 2 | DeviceNumber | int | Номер пристрою звукозапису |
| 3 | WaveFormat | WaveFormat | Формат даних для запису |
| 4 | FrameSize | int | Розмір вікна для обробки |
| 5 | MelsNumber | int | Кількість мел-фільтрів |
| 6 | RecCount | int | Кількість еталонних записів |

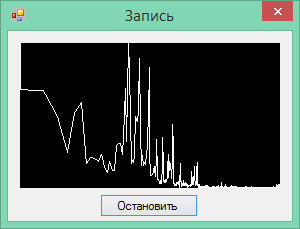


Рисунок 2.9 – вікно запису голосу.

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

## 2.3 Створення системи автентифікації

Як основу даного ПЗ було створено типову систему аутентифікації з використанням текстового паролю. Вона полягла основою для подальшого вдосконалення та перероблення.

Отже система аутентифікації має наступні вимоги:

* можливість створення/редагування користувача
* можливість видалення користувача
* стійкість до можливих зламів

Створена система аутентифікації зберігає шифровану базу даних з логінами та хешами паролів для зменшення вирогідності можливого зламу.

У якості хешу було обрано MD5 хеш функцію.

Додаток виконано у вигляді бібліотеки, що дозволяє використовувати дане ПЗ у будь-якому проекті, що сумісний з .NET. Достатньо додати посилання на бібліотеку та кілька рядків:

Лістинг 2.1 – приклад викристання бібліотеки.

[STAThread]

static void Main()

{

Application.EnableVisualStyles();

Application.SetCompatibleTextRenderingDefault(false);

if (!VoiceAuth.VoiceAuth.Login())

{

MessageBox.Show("Вы не авторизованы!", "Ошибка авторизации");

return;

}

Application.Run(new Form1());

}

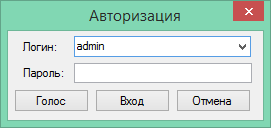


Рисунок 2.10 – діалог авторизації.

## 2.4 Створення системи налаштувань

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

Для доступу до панелі налаштувань необхідно залогінитись з логіном «admin» у системі авторизації.

Панель налаштувань містить список користувачів, а також засоби його редагування. Також є налаштування для запису аудіо, та системи розпізнавання.

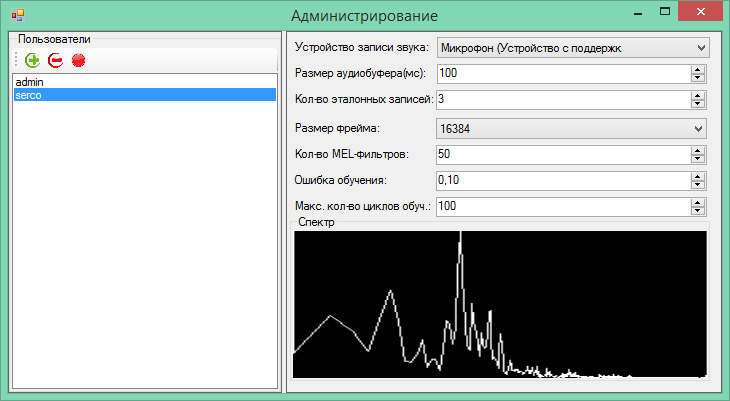


Рисунок 2.11 – вікно налаштувань

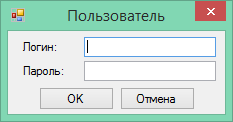


Рисунок 2.12 – діалог створення/редагування користувача

Всі налаштування зберігаються у спеціальному файлі, якщо ж файл відсутній, то установлюються налаштування за замовчанням.

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС