# 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.

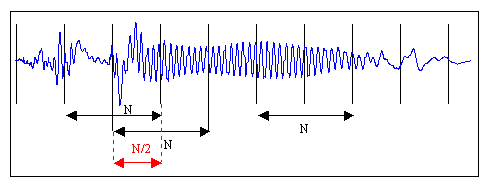


Рисунок 2.1 – розбиття довготривалого сигналу на «вікна»

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.

До голосового апарату людини входять ротова і носова порожнини з додатковими порожнинами, пелька, гортань з голосовими зв'язками, трахея, бронхи, легені, грудна клітка з дихальними м'язами і діафрагмою, м'язи черевної порожнини. Центральна нервова система організовує їх функції в єдиний, цілісний процес звукотворення, що є складним психофізичним актом. Існує безпосередній взаємозв'язок між голосом і слухом: голос не може розвиватися без участі слуху, слух не може розвиватися без участі голосових органів. У глухих голос не функціонує, оскільки немає слухових сприйнять, і, відповідно, стимуляції мовно-рухових центрів. Голосом людина здатна відтворювати багато різноманітних звуків, часто досить складних. У голосі виражається емоційний стан людини: сердитість, здивування, радість і тому подібне. Люди мають голосові зв'язки, які можуть розтягуватися і стискатися, міняти свою товщину; людина може довільно міняти тиск повітря, інтенсивність повітряного потоку, що подається на голосові зв'язки. Форма грудної клітки, гортані, положення язика, ступінь натягнення інших м'язів може змінюватися. Результатом будь-якої з цих дій будуть зміни в висоті, силі, тембрі, чіткості, емоційному забарвленню утвореного звуку. На якість голосу впливають загальне фізичне самопочуття людини, постава, настрій, емоційний стан та ін.

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

Для усунення небажаних 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.

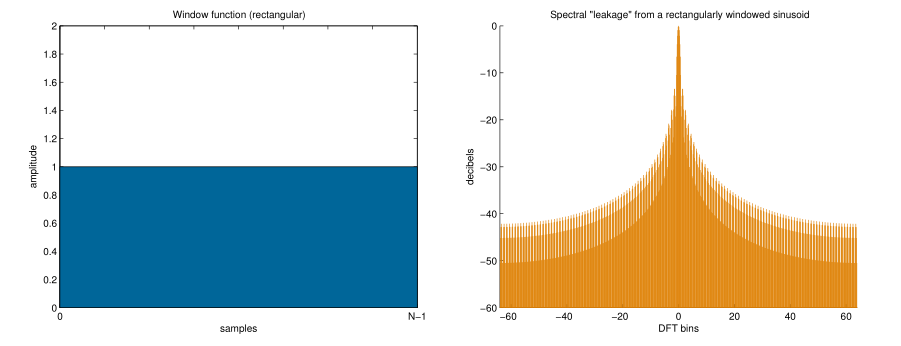


Рисунок 2.2 – прямокутне вікно та спектр перетворення Фур’є для гармонічного сигналу

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)

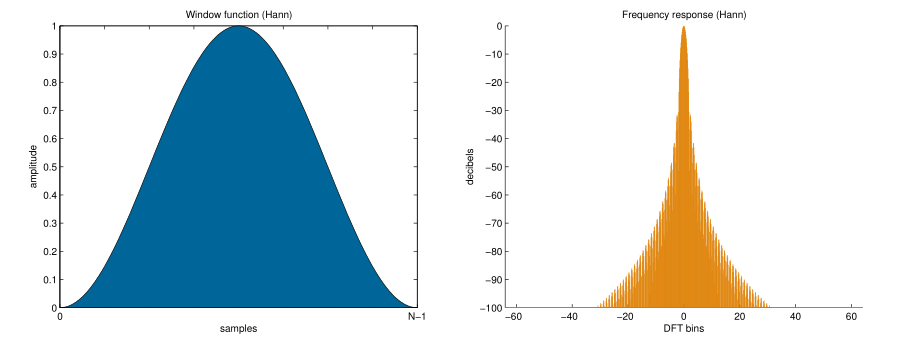


Рисунок 2.3 – вікно Ханнінга та спектр перетворення Фур’є для гармонічного сигналу

Вікно Хемінга – рівень бокових пелюсток – -42 дБ.

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

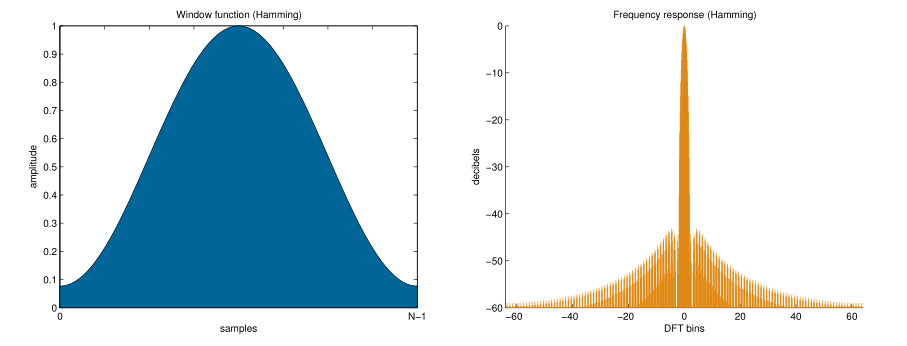


Рисунок 2.4 – вікно Хеммінга та спектр перетворення Фур’є для гармонічного сигналу

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}\,

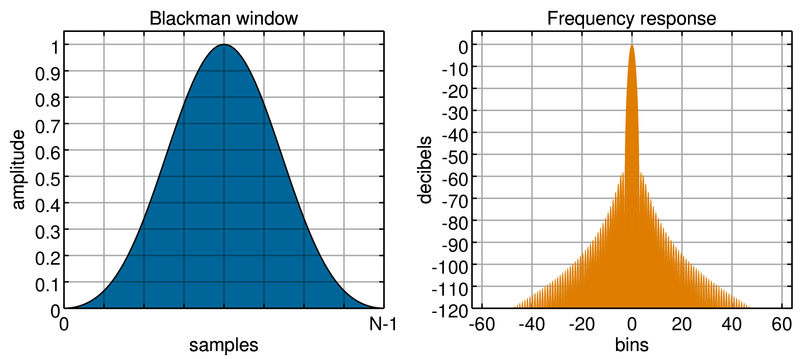


Рисунок 2.5 – вікно Блекмана та спектр перетворення Фур’є для гармонічного сигналу

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

Арк.

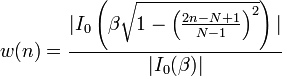
№ документа

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.

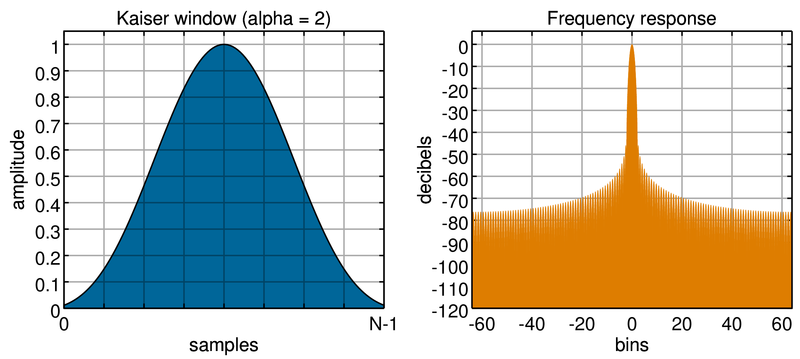


Рисунок 2.6 – вікно Кайзера та спектр перетворення Фур’є для гармонічного сигналу

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».

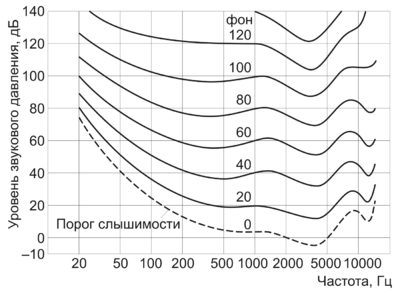


Рисунок 2.7 – чутливість людського вуха до звуків

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

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

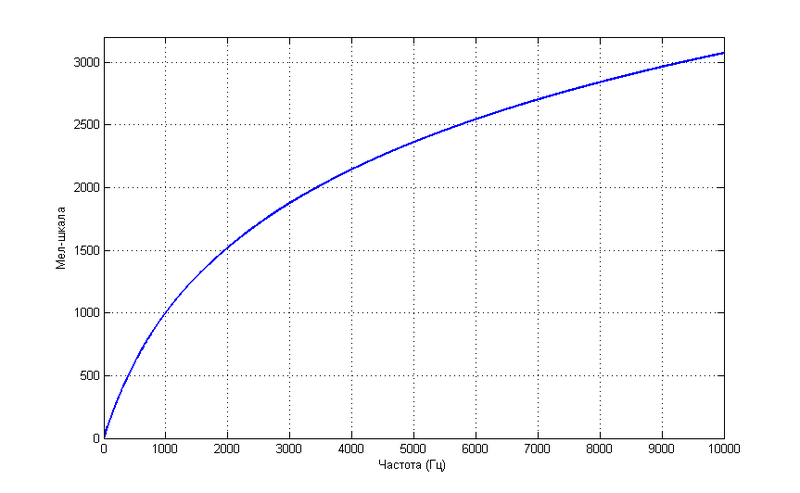


Рисунок 2.8 – графік відповідності мел-шкали до лінійної

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.

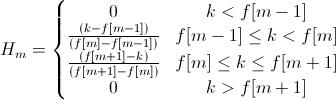
1. Вихідний мовний сигнал запишемо в дискретному вигляді як

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

## 2.2 Створення системи запису голосу

Для запису звуку з мікрофона використаємо бібліотеку NAudio.

WAV — формат аудіофайла розроблений компаніями Microsoft та IBM. WAVE базується на форматі RIFF, поширюючи його на інформацію про такі параметри аудіо, як застосований кодек, частота дискретизації та кількість каналів. WAV як і RIFF передбачався для комп'ютерів IBM PC, тому всі змінні записані у форматі little endian. Відповідником WAV для комп'ютерів PowerPC є AIFF.

Хоча файли WAVE можуть бути записані за допомогою будь-яких кодеків аудіо, зазвичай використовується нестиснений PCM, який призводить до великих обсягів файлу (близько 172 кБ на секунду для CD-якості). Іншим недоліком файлу є обмеження обсягу до 4 ГБ, через 32-бітну змінну. Формат WAV був частково витіснений стисненими форматами, проте, завдяки своїй простоті, надалі знаходить широке використання в процесі редагування звуку та на переносних аудіопристроях, як програвачі та цифрові диктофони.

WAVE файл складається з двох частин: заголовку файлу і області даних. Канонічний формат WAVE файлу починається з RIFF(таблиця 2.1) заголовку і двох підсекцій: "fmt "(таблиця 2.2) і "data"(таблиця 2.3). Підсекція "fmt " описує параметри даних звукозапису. У підсекції даних "data" міститься розмір даних і фактичні дані звукозапису.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Таблиця 2.1 - RIFF Заголовок WAV-файлу | | | | |
| Позиція (HEX) | Позиція | Розмір | Назва | Пояснення |
| 0000 | 0 | 4 байти | ChunkID | Містить літери "RIFF" в ASCII кодуванні  (0x52494646 у порядку байт big-endian) |
| 0004 | 4 | 4 байти | ChunkSize | 36 + SubChunk2Size, або більш точно: 4 + (8 + SubChunk1Size) + (8 + SubChunk2Size)  Це розмір всього файлу в байтах, не враховуючи 8 байтів перших двох полів: ChunkID і ChunkSize. |
| 0008 | 8 | 4 байти | Format | Містить літери "WAVE"  (0x57415645 у порядку байт big-endian). |

Таблиця 2.2 - Підсекція "fmt" WAV-файлу

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Позиція (HEX) | Позиція | Розмір | Назва | Пояснення |
| 000C | 12 | 4 байти | Subchunk1ID | Містить літери "fmt "  (0x666d7420 у порядку байт big-endian). |
| 0010 | 16 | 4 байти | Subchunk1Size | Приймає значення 16 для формату кодування PCM.  Це розмір області цієї підсекції, що слідує після цього поля. |
| 0014 | 20 | 2 байта | AudioFormat | PCM = 1 (Лінійне квантування)  Значення відмнінні від 1 вказують на наявність конкретного  кодування аудіо даних. |
| 0016 | 22 | 2 байта | NumChannels | Кількість звукових каналів. Моно = 1, Стерео = 2, і т.д. |
| 0018 | 24 | 4 байти | SampleRate | Частота дискретизації у Герцах. Наприклад, 8000, 44100, інші. |
| 001С | 28 | 4 байти | ByteRate | = SampleRate \* NumChannels \* BitsPerSample/8 |
| 0020 | 32 | 2 байта | BlockAlign | = NumChannels \* BitsPerSample/8  Кількість байт, яка міститься в одному семплі враховуючи кількість каналів. |
| 0022 | 34 | 2 байта | BitsPerSample | Кількість біт в одному семплі. Так звана "глибина" чи точність звучання.  8, 16, 32, і так далі. |
|  |  | 2 байта | ExtraParamSize | Розмір поля з параметрами. Якщо PCM, ці два поля ExtraParamSize і ExtraParams не записуються. |
|  |  | X | ExtraParams | Місце для запису додаткових параметрів |

Таблиця 2.3 - Підсекція "data" WAV-файлу

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Позиція (HEX) | Позиція | Розмір | Назва | Пояснення |
| 0024 | 36 | 4 байти | Subchunk2ID | Містить літери "data"  (0x64617461 у порядку байт big-endian). |
| 0028 | 40 | 4 байти | Subchunk2Size | Розмір даних звукозапису в байтах, які містяться в наступному полі.  Кількість семплів \* NumChannels \* BitsPerSample/8 |
| 002С | 44 | X | Data | Фактичні дані звукозапису. |

Створимо об’єкт звукозапису який буде містити потрібні методи:

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

Таблиця 2.4 – 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 2.5 – 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 |

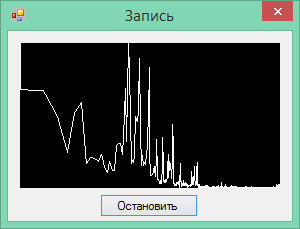


Рисунок 2.9 – вікно запису голосу.

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());

}

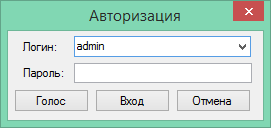


Рисунок 2.10 – діалог авторизації.



Рисунок 2.11 – структурна схема системи авторизації

MD5 — 128-бітний алгоритм хешування, розроблений професором Рональдом Л. Рівестом в 1991 році. Призначений для створення «відбитків» або «дайджестів» повідомлень довільної довжини. Прийшов на зміну MD4, що був недосконалим. Описаний в RFC 1321. У 2004 році китайські дослідники Сяоюнь Ван (Xiaoyun Wang), Денгуо Фен (Dengguo Feng), Сюецзя Лай (Xuejia Lai) і Хонбо Ю (Hongbo Yu) повідомили про знаходження ними вразливості в алгоритмі, що дозволяє за невеликий час (1 годину на кластері IBM p690) знаходити колізії хеш-функцій. На жаль, автори так і не відкрили свій секрет широкій публіці. У 2006 році чеський дослідник Властимил Клима опублікував алгоритм, що дозволяє знаходити колізії на звичайному комп'ютері з довільним початковим вектором (A,B,C,D), за допомогою методу, що був названий: "тунелювання".

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

Вхідні дані вирівнюються так, щоб їхній розмір можна було порівняти з 448 по модулю з 512. Спочатку дописують одиничний біт (навіть якщо довжина порівняна з 448), далі необхідна кількість нульових бітів. Дописування 64-бітного представлення довжини даних по вирівнюванню. Якщо довжина перевищує 264-1, то дописують молодші біти.

Ініціалізуть 4 змінних розміром по 32 біта:

А = 01 23 45 67;

В = 89 AB CD EF;

С = FE DC BA 98;

D = 76 54 32 10.

Вирівнювані дані розбиваються на блоки по 32 біта, і кожен проходить 4 раунда з 16 операторів. Всі оператори однотипні і мають вигляд: [abcd k s i], визначений як

a = b + ((a+ Fun(b, c, d) + X[k] + T<i>) <<< s)

де X - блок даних, а T[1..64] – 64-ох елементна таблиця побудована наступним чином:

T[i]=int(4294967296*|sin(i)|)

де s - циклічний зсув вліво на s біт отриманого 32-бітного аргументу.

В першому раунді Fun F(X, Y, Z) = XY v (not X)Z

В другому раунді Fun G(X, Y, Z) = XZ v (not Z)Y.

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

В третьому раунді Fun Н(Х, Y, Z) = Х xor Y xor Z.

В четвертому раунді Fun I(Х, Y, Z) = Y xor (X v (not Z)).

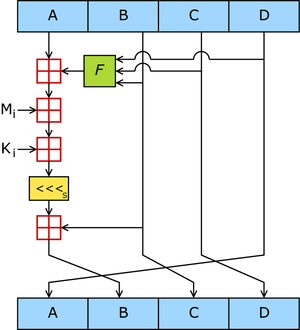
[](https://uk.wikipedia.org/wiki/%D0%A4%D0%B0%D0%B9%D0%BB:MD5.png)

Рисунок 2.12 – Алгоритм MD5 хешування

2.4 Створення системи налаштувань

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

Основні функції системи налаштувань:

* збереження та відтворення списку користувачів;
* збереження та відтворення хешів до паролів користувачів;
* шифрування даних про користувачів;
* збереження та відтворення налаштувань аудіоінтерфейсу;
* збереження та відтворення налаштувань мел-кепстрального перетворення;
* збереження та відтворення налаштувань нейромережі для розпізнавання голосу.
* відображення спектру останнього голосового входу користувача.
* додавання нових користувачів;
* видалення або редагування даних існуючих користувачів.



Рисунок 2.13 – струтктурна схема системи налаштувань

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

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

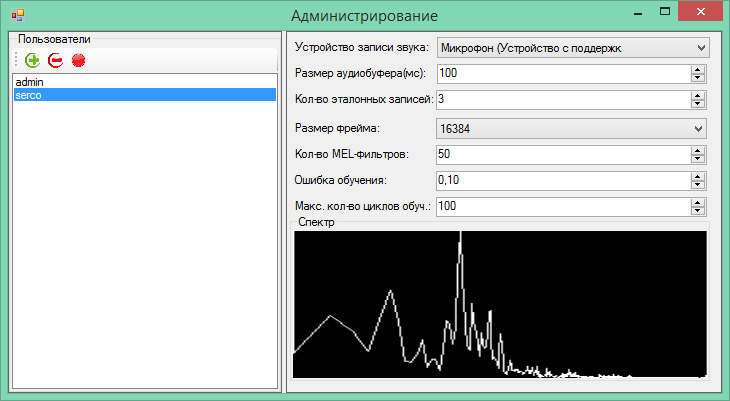


Рисунок 2.14 – вікно налаштувань

Арк.

№ документа

Підпис

Дата

Арк.

КНУ.РБ.6050102.15.33.ВС

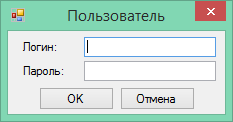


Рисунок 2.15 – діалог створення/редагування користувача

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