ЗМІСТ

ВСТУП

1. МЕТОДИ СИНТЕЗУ МОВИ

1.1 Повний синтез мови за правилами

1.2 Синтез фонемами

2. АНАЛІЗ ЗАВДАННЯ

2.1 Якість звуку

2.2 Підбір фонем

3. СИНТЕЗУВАННЯ МОВИ

3.1 Обробка тексту

3.2 Генерація проміжків між фонемами

3.3 Згладжування звуку

3.4 Обробка вихідного звуку

3.5 Аналіз результатів синтезування

4. ІНСТРУКЦІЯ КОРИСТУВАЧА

4.1 Опис головного вікна програми

4.2 Відтворення мови

4.3 Модифікація мови

4.4 Збереження мови

СИСТЕМНІ ВИМОГИ

ВИСНОВОК

СПИСОК ЛІТЕРАТУРИ

ВСТУП

У теперішній час в багатьох сферах застосовуються синтезатори мови, зокрема й російської теж. Найчастіше такі синтезатори обмежуються набором певних слів, враз чи пропозиції. Але іноді потрібно синтезувати мову по довільному, заздалегідь не відомому тексту. Останніх налічується не так вже й багато і голоси, якими вони говорять дуже сильно обмежені за різноманітністю. Створення такого синтезатора досить складне завдання, притому складність особливо не залежить від мови, на якому повинна говорити програма. Найпростішим із способів синтезі мови по довільному текстом - є синтез з фонем. У ході виконання роботи буде використаний саме цей метод. На жаль він не дає високої якості мови, але менш витративши за часом реалізації, ніж інші. Так само він видає мову, яка задовольняє вимогам чіткості.

1. МЕТОДИ СИНТЕЗУ МОВИ

1.1 Повний синтез мови за правилами

Повний синтез мови за правилами (або синтез по друкарському тексту) забезпечує управління всіма параметрами мовного сигналу і, таким чином, може генерувати мову по заздалегідь невідомому тексту. В цьому випадку параметри, отримані при аналізі мовного сигналу, зберігаються в пам'яті так само, як і правила з'єднання звуків в слова і фрази. Синтез реалізується шляхом моделювання мовного тракту, застосування аналогової або цифрової техніки. Причому в процесі синтезування значення параметрів і правила з'єднання фонем вводять послідовно через певний часовий інтервал, наприклад 5-10 мс. Метод синтезу мови по друкарському тексту (синтез по правилах) базується на запрограмованому знанні акустичних і лінгвістичних обмежень і не використовує безпосередньо елементів людської мови. У системах, заснованих на цьому способі синтезу, виділяється два підходи. Перший підхід направлений на побудову моделі мовотворчої системи людини, він відомий під назвою артикуляторного синтеза. Другий підхід — формантний синтез за правилами. Розбірливість і натуральність таких синтезаторів може бути доведена до величин, порівнянних з характеристиками природної мови.

1.2 Синтез фонемами

Синтез мови по правилах з використанням попередніх відрізань природної мови, що запам'ятали, — це різновид синтезу мови по правилах, яка набула поширення у зв'язку з появою можливостей маніпулювання мовним сигналом в оцифрованій формі. Залежно від розміру початкових елементів синтезу виділяються наступні види синтезу:

* мікросегментний (мікрохвильовий);
* аллофонічний;
* дифонний;
* напівскладовий;
* складовий;

синтез з одиниць довільного розміру.

Зазвичай як такі елементи використовуються напівсклади — сегменти, що містять половину згідного і половину що примикає до нього голосного. При цьому можна синтезувати мову по заздалегідь не заданому тексту, але важко управляти інтонаційними характеристиками. Якість такого синтезу не відповідає якості природної мови, оскільки на границях зшивання дифонів часто виникають спотворення. Компіляція мови із заздалегідь записаних словоформ також не вирішує проблеми високоякісного синтезу довільних повідомлень, оскільки акустичні і просодичні (тривалість і інтонація) характеристики слів змінюються залежно від типу фрази і місця слова у фразі. Це положення не міняється навіть при використанні великих об'ємів пам'яті для зберігання словоформ.

2. АНАЛІЗ ЗАВДАННЯ

2.1 Якість звуку

Так як використовуваний метод не дає високої якості мови, то й сенсу використовувати високоякісний звук сенсу немає. Тому для запису фонем буде використано нижче якість звуку. Амплітуда - 8біт, а частота 8кГц. Це якість звуку для телефонної розмови.

2.2 Підбір фонем

Для російської мови виділяють наступні фонеми а э и о у ы п п' б б' м м' ф ф' в в' т т' д д' н н' с с' з з' р р' л л' ш ж щ җ ц ч й к к' г г' х х'. Але виділені фонеми є деякою абстракцією і не можуть в даному виді застосуються для синтезу мови.

У ході експериментів було встановлено, що звучання відрізків слів, де після голосної стоїть приголосна, досить задовільно. Однак при зворотному порядку всі інакше - одна приголосна може звучати як інша, а може бути взагалі не чутно. У деяких випадках допомагає метод з'єднання звуків, де між стиками вставляється згенерований звук, що враховує закінчення попереднього і початку наступного. Про цьому методі буде докладно розказано в пункті 3.2. Для інших сполучень такий підхід не дав бажаного результату і було ухвалено рішення записати всі можливі поєднання. У результаті ми отримали фонеми зазначені в таблиці 1. Фонеми закінчуються на знак "!" Знаходяться в кінці слів і є більш дзвінкими, ніж їх звичайні аналоги (без закінчень). Знак "\_" в кінці назви фонеми означає, що вона м'якша, ніж аналог без закінчення і йде переважно перед голосними.

Таблиця 1 – Фонеми

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| А | - | - | - | - | - | - | - | - | - | - | - |
| Е | - | - | - | - | - | - | - | - | - | - | - |
| Ё | - | - | - | - | - | - | - | - | - | - | - |
| И | - | - | - | - | - | - | - | - | - | - | - |
| О | - | - | - | - | - | - | - | - | - | - | - |
| У | - | - | - | - | - | - | - | - | - | - | - |
| Ы | - | - | - | - | - | - | - | - | - | - | - |
| Э | - | - | - | - | - | - | - | - | - | - | - |
| Ю | - | - | - | - | - | - | - | - | - | - | - |
| Я | - | - | - | - | - | - | - | - | - | - | - |
| Б | - | ба | бе | бё | би | бо | бу | бы | бэ | бю | бя |
| В | - | ва | ве | вё | ви | во | ву | вы | вэ | вю | вя |
| Г | Г! | - | - | - | - | - | - | - | - | - | - |
| Д | Д! | - | - | - | - | - | - | - | - | - | - |
| Ж | - | - | - | - | - | - | - | - | - | - | - |
| З | З\_ | - | - | - | - | - | - | - | - | - | - |
| Й | - | - | - | - | - | - | - | - | - | - | - |
| К | К\_ | - | - | - | - | - | - | - | - | - | - |
| Л | Л\_ | ла | ле | лё | ли | ло | лу | лы | лэ | лю | ля |
| М | - | ма | ме | мё | ми | мо | му | мы | мэ | мю | мя |
| Н | - | - | - | - | - | - | - | - | - | - | - |
| П | П\_ | - | - | - | - | - | - | - | - | - | - |
| Р | Р\_ | - | - | - | - | - | - | - | - | - | - |
| С | С\_ | - | - | - | - | - | су | - | - | - | - |
| Т | Т\_ | - | - | - | - | - | - | - | - | - | - |
| Ф | Ф\_ | фа | фе | фё | фи | фо | фу | фы | фэ | фю | Фя |
| Х | Х\_ | - | - | - | - | - | - | - | - | - | - |
| Ц | - | - | - | - | - | - | - | - | - | - | - |
| Ч | - | - | - | - | - | - | - | - | - | - | - |
| Ш | - | - | - | - | - | - | - | - | - | - | - |

3. СИНТЕЗУВАННЯ МОВИ

3.1 Обробка тексту

Перш ніж приступити до синтезування мови, необхідно визначити які фонеми, де і в якому порядку повинні відтворюється. Для цього спочатку текст розбивається на окремі слова, а слова потім на фонеми. Коли ми маємо список фонем, між фонемами приголосна-голосна вставляється згенерованих послідовність. Метод генерації описаний в пункті 3.2. Потім все слово згладжується, для придушення клацань, які виникають при різких перепадах значення сигналу. Метод згладжування описаний в пункті 3.3. Повна схема алгоритму наведена у додатку А.

3.2 Генерація проміжків між фонемами

Для генерації послідовності береться остання частина лівої фонеми довго рівний періоду повторення схожих ділянок у звуці. А так само початок правою за тим же принципом. Потім генеруються п'ять ділянок, які плавно перетікають з лівої фонеми в праву. Тобто перший ділянка дуже схожий на ліву фонему, третій щось середнє між лівою і правою, а п'ятий дуже схожий на праву. Фактично фонеми частково накладаються один на одного. На малюнках 3.1, 3.2, 3.3 відображує графічне представлення цих звуків. Як видно з малюнка 3.3, згенерованих частина сильно виділяється різкими переходами, що призводить до клацанням під час відтворення звуку.

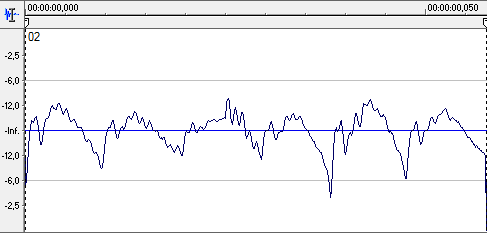


Рисунок 3.1 – Фонема “Р\_”

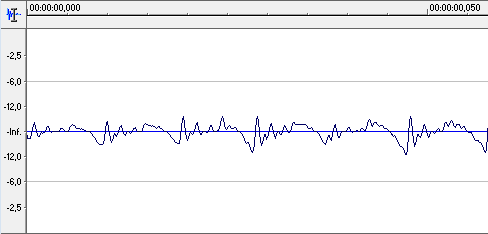


Рисунок 3.2 – Початок фонеми “А”

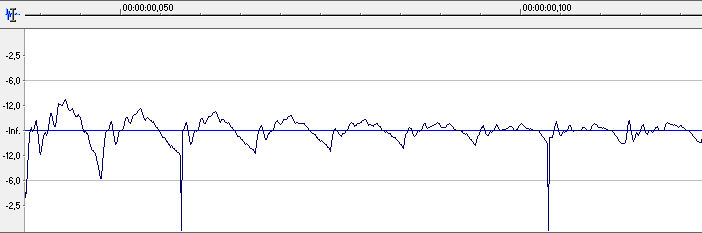


Рисунок 3.3 – Згенерований проміжок між фонемами “Р\_” і “А”

3.3 Згладжування звуку

Після злиття фонем та згенерованих ділянок, може утворитися місця різкого перепаду сигналу, що веде за собою клацання, на кожному такому перепаді. Що в середньому складає по півтора клацання на кожну букву в слові. Це призводить до постійного клацання, для усунення якого був використаний усереднюються фільтр за трьома точками.

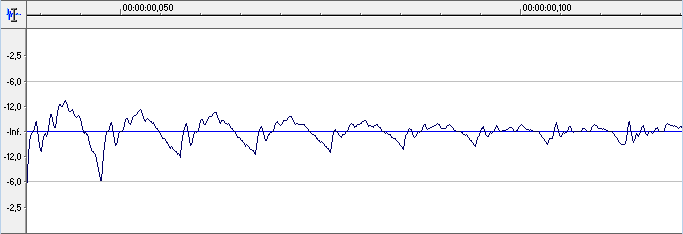


Рисунок 3.4 - Згенерований проміжок між фонемами "Р\_" і "А" із застосуванням усереднює фільтра.

3.4 Обробка вихідного звуку

Перед відтворенням звук ще два рази зазнає обробці. При першій його гучність коригується залежно від установок користувача. Для чого значення сигналу множиться на коефіцієнт гучності, який варіюється від нуля і до 1.5. На малюнках 3.5 - 3.8 зображена фонема "Р" при різних рівнях гучності.

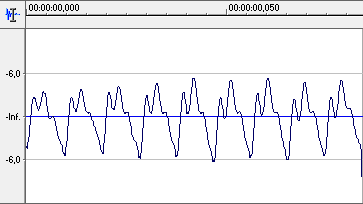


Рисунок 3.5 – Фонема "Г" при гучності 100%

Як видно з малюнка 3.8, при гучності 250% фонема "Г" значно спотворюється і перетворюється на шум. Для різних фонем максимальна гучність різна, тому було прийнято рішення обмежити користувача максимальним рівнем 150%. Даних рівень підходить для всіх використовуваних фонем.

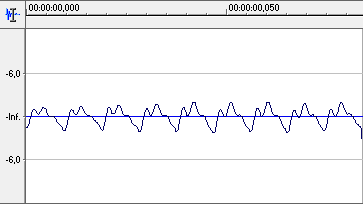


Рисунок 3.6 – Фонема “Г” при гучності 40%

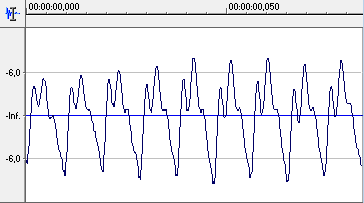


Рисунок 3.7 – Фонема “Г” при гучності 150%

Потім відрегульований по гучності звук приводиться до необхідної швидкості. Для цього він розтягується на заданий користувачем коефіцієнт. Фактично для звуку змінюється частота дискретизації за часом. Коефіцієнт варіюється від 0.5 до 5. Але при підвищенні швидкості за позначку 2.0, мова перестає бути розбірливим. На малюнку 3.5 зображена фонема "Г" при швидкості 100%. На малюнках 3.9 - 3.11 наведена вона ж, але на різних швидкостях. За зовнішнім виглядом вони здаються однаковими, але з тимчасової шкалою помітно відмінність. На малюнку 3.11 також помітно що графік стає грубим через дуже малої частоти дискретизації.

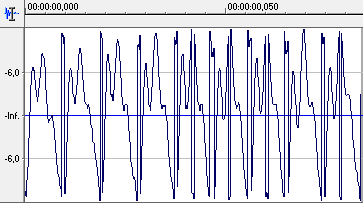


Рисунок 3.8 – Фонема “Г” при гучності 250%

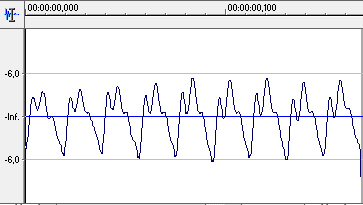


Рисунок 3.9 – Фонема “Г” при швидкості 50%

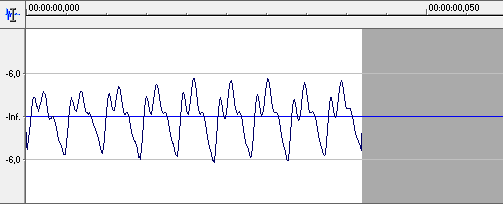


Рисунок 3.10 – Фонема “Г” при швидкості 200%

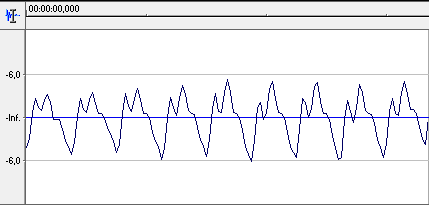


Рисунок 3.11 – Фонема “Г” при швидкості 500%

3.5 Аналіз результатів синтезування

Для аналізу якості синтезування звуків, найкращим методом буде порівняння синтезованої мови з оригінальною. Для цього візьмемо пару слів, запишемо їх і синтезуємо. Потім порівняємо графіки. В якості тестових слів візьмемо "Мурка" і "кактуси". Їхні зображення наведені на малюнках 3.12 - 3.15.

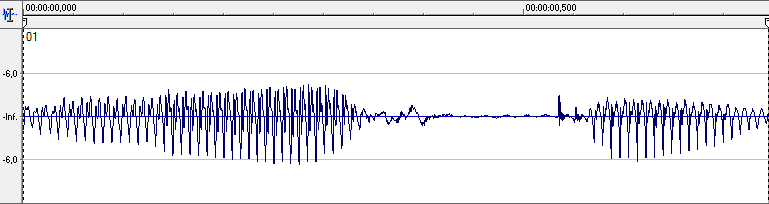


Рисунок 3.12 – Слово "мурка" (оригінал)

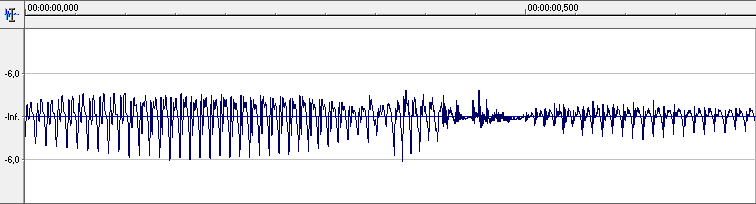


Рисунок 3.13 – Слово "мурка" (синтезоване)

Як видно з малюнків 3.12 і 3.13, графіки досить схожі, за винятком того, що в оригіналі, пере звуком "до", знаходиться пауза. Тривалість паузи приблизно 50 мкс. З'єднання "до" і "а" практично ідентичні, лише за винятком того, що звук "а" в оригіналі голосніше. Звуки "ма" в синтезованому слові є монолітними, тобто записаними відразу разом.

Щодо слова "кактуси" (Мал. 3.14 і 3.15) простежуються ті ж тенденції. Помічені 50 мілісекундного проміжки перед і після звуку "т". Отже між вибуховими звуками можна спробувати вставляти пази, що можливо призведе до поліпшення звучання.

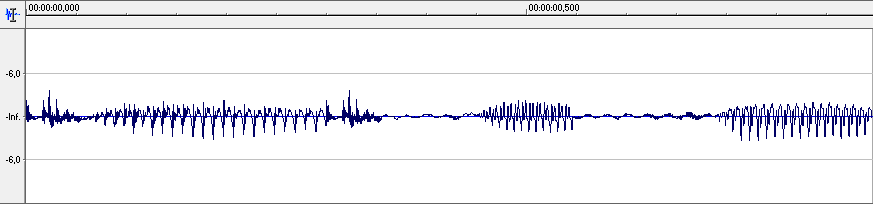


Рисунок 3.14 – Слово "кактусы" (синтезоване)

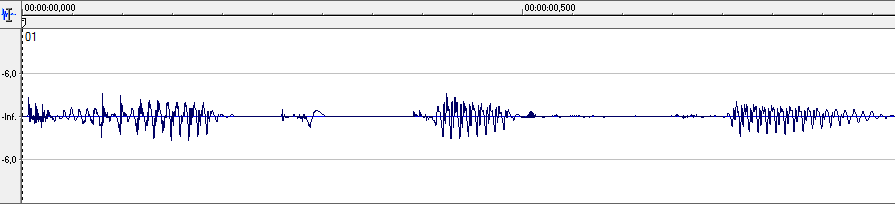


Рисунок 3.15 – Слово "кактусы" (оригінал)

4. ІНСТРУКЦІЯ КОРИСТУВАЧА

4.1 Опис головного вікна програми

Інтерфейс користувача являє собою вікно програми зображене на малюнку 4.1. Вся робота ведеться саме в ньому. У верхній частині вікна знаходиться головне меню (1), а під ним панель інструментів (2), яка повністю дублює меню. Ще нижче знаходиться панель модифікаторів (3). Частину, що залишилася вікна займає текстове поле, для введення тексту, який буде проговорюватися (4).

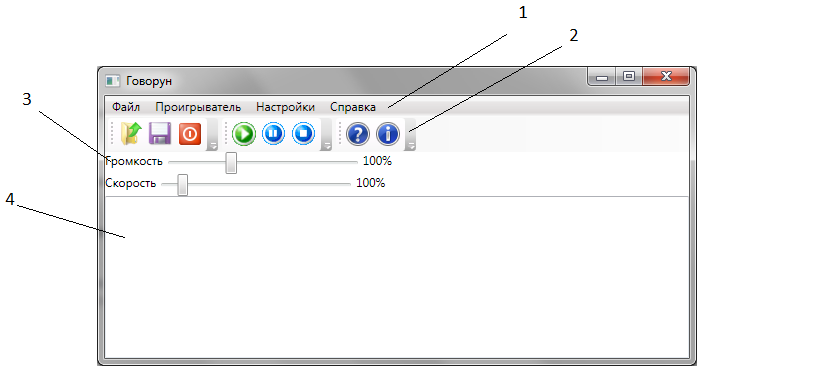


Рисунок 4.1 – Головне вікно програми

4.2 Відтворення мови

Для відтворення мови, потрібно в текстове поле ввести текст, або відкрити файл, вибравши в меню Файл-> Відкрити. У вікні вказати потрібний файл. На поточний момент підтримуються формати. Txt і. Rtf. Для відтворення звуку необхідно вибрати пункт меню Програвач-> Говорити. Для припинення, необхідно натиснути в тому ж пункті вибрати пункт Пауза, а для зупинки Стоп. Усі ці пункти продубльовані в панелі інструментів. Якщо необхідно відтворити не весь текст, а його частина, то можна виділити частину тексту запустити програвання, як вказано на малюнку 4.2.

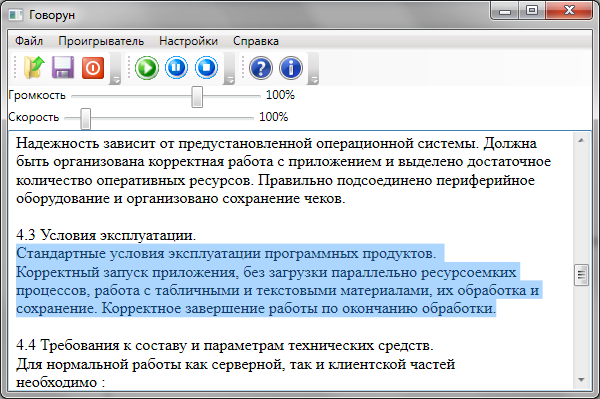


Рисунок 4.2 – Відтворення виділеного тексту

4.3 Модифікація мови

На панелі модифікаторів знаходяться повзунки швидкості і гучності. При їх пересування змінюються відповідні параметри мови. Всі зміни застосовуються динамічно, тобто не непотрібно перезапускати відтворення.

4.4 Збереження мови

Для збереження мови в файл, необхідно вибрати пункт меню Файл-> Зберегти. У вікні, вказати місце збереження і зберегти. На даний момент доступний тільки формат. Wav. Звук зберігається з урахуванням положення повзунків.

СИСТЕМНІ ВИМОГИ

Операційна система Windows XP/Vista/7.

Необхідна версія Framework 3.5/4.0

Пристрій виведення звуку.

Для установки програми необхідно 1Мб вільного місця на жорсткому диску.

ВИСНОВОК

Створення синтезатора мови є досить трудомісткою роботою. А створення синтезатора розбірливим мови, ще більш складним завданням. Створити ж синтезатор, голос якого можна було б відрізнити від справжнього, вимагає великої кількості досліджень. Так само його доцільно створювати з допомогою синтезу за правилами, а не готовими фонемами. У даній же курсової роботі була здійснена спроба створення синтезатора, який хоча би задовольняв вимогам чіткість. Для створення повноцінного синтезатора необхідна величезна база слів, хоча б для того, що б у словах відтворювати наголосу, а також врахування особливостей вимови деяких слів. Але деякі особливості мови не піддаються алгоритмізації ні якими шляхами. Приміром наголосу деяких слів залежить від контексту їх використання. Одним з таких слів є замок.

Створений у цій курсової роботі синтезатор, є дуже наближеною моделлю того, яким повинен бути справжній синтезатор мови. Але для розробки дійсно стоїть синтезатора потрібно значний проміжок часу і велика кількість досліджень.

СПИСОК ЛІТЕРАТУРИ

1. Синтез мови - http://ru.wikipedia.org/wiki/
2. Винцюк Т.К. - Анализ, распознавание и интерпретация речевых сигналов, 1987, Киев - Наукова Думка
3. Фланаган Дж.Л. - Анализ, синтез и восприятие речи, 1968, Связь – Москва
4. Сорокин - Синтез речи, 1992, Связь – Москва
5. Кучеров, Лобанов -Синтезированная речь в СМО, 1983, Киев - Наукова Думка.