Пловдивски университет „Паисий Хилендарски“

Факултет по математика и информатика

Катедра “Компютърни информатика”

ДИПЛОМНА РАБОТА

Глас 2 - синтезатор на българска реч

Дипломант: Тодор Илиев Арнаудов

Фак.№ 0701407025

Научен ръководител: доц. д-р Христо Крушков

Пловдив
юни 2008 г.

Съдържание

Благодарности

На научния ми ръководител доц. Христо Крушков за напътствията и подкрепата по време на създаването на тази дипломна работа и преди това. На първия ми научен ръководител в Университета - доц. д.м.н. Георги Тотков - за това, че желанието ми да направя синтезатор, който говори по-гладко от неговия, ме стимулира да превърна страстта си по звука в синтезатора на реч „Глас”. На проф. Руслан Митков, защото по време на специализацията си при него научих как се пишат дипломни работи; и много други неща...

Искам да благодаря и на доц. Димитър Мекеров и на преподавателите и колегите за хубавата атмосфера във ФМИ, докато все още бяхме студенти.

Въведение

Синтезатори на реч от текст

Синтезаторите на реч от текст са софтуерни и хардуерни системи, които генерират звуци на човешка реч въз основа на анализ на текст на естествен език и/или команди за управление на модел на говорен апарат. За разработката им се прилагат методи от Компютърната лингвистика, Фонетиката, Акустиката, Психоакустиката, Цифровата обработка на сигнали и др.

Синтезаторите на реч намират редица практически приложения - екранни четци за незрящи, говорещи машини за хора с проблеми с говора, за информиране за случване на определени събития без да се ангажира зрително потребителят; в мултимедийни диалогови системи, мултимедийни играчки и др.

Съществуват няколко основни типа синтезатори на реч: със слепване, формантни и артикулационни, като са възможни и хибриди между тях, които използват комбинирани методи.

Глас

“Глас” е хибриден синтезатор на българска реч, разработен първоначално за да подобри говорещата част на Системата за Лингвистично Осигуряване на Говора (СЛОГ). „Глас” използва микрофонеми – откъси от реална реч с продължителността на основната формантна честота, - които след това манипулира. За синтезиране на шумови съгласни обаче се използва слепване на цялостни записи на фонеми.

Основните възможности на „Глас” са:

- Озвучаване на цифри, съкращения и произволни специални низове от речник
- Различаване на меко и твърдо „л”, обеззвучаване на съгласни, удължаване на думи в края на изговор.
- Възможност да изговаря ударени и редуцирани гласни, но ако са изрично отбелязани във входния текст, защото липсва анализатор.

- Плавни звукови преходи между гласни и звучни съгласни чрез формантна интерполация, и възможности за "извиване" на гласа чрез променлива продължителност на преходите между съседни гласни и звучни съгласни.
- Променлива височина на гласните и звучните съгласни, чрез промяна на формантните честоти.
- Променлива скорост, чрез промяна на броя на периодите на тоналните звукове и дължината на преходите.

Цел на дипломната работа

Общите цели на настоящата дипломна работа са:

1. Да се разработи подобрен синтезатор на реч от текст – *Глас 2*, - който да се опре на разработките, изследванията и идеите за бъдещи усъвършенствания от създаването на Глас.
2. Да се представят методи за обработка на речеви сигнали, променящи продължителността и височината на звука, и опити за автоматично извличане на форманти.

3. Да се представят идеи и направления за изследвания и разработки, свързани с подобряване на синтезатора, предмет на тази дипломна работа, и със създаване на по-съвършен синтезатор и анализатор на реч и музика – *Глас 3* – по който да се работи в бъдеще.

Целите на *Глас 2* могат да се конкретизират като:

1. Внедряване на модула libmorf за определяне на ударения, който да позволи да се определят и редукциите на гласните и по този начин да се подобри фонетичната коректност на синтезатора. Модулът е разработен в катедрата по Компютърна информатика.
2. Включване на интонационни контури.
3. Синтез на пеене.
4. Специални звукови ефекти и промяна на тембъра.
5. Опити с темброви контури.
6. Озвучаване не емотикони и междуметия.
7. Включване на динамика на амплитудата на звука.

8. Други по-дребни подобрения и разширяване на възможностите за параметризация на изговора.

Структура на работата

Първа глава разглежда накратко звука, музиката и човешката реч като сигнали, и представя методи за анализ и синтез на звук, реч и музика. Втора глава представя основните методи за синтез на реч – формантен, със слепване, артикулационен. В трета глава се обсъждат приложенията на синтеза на реч. В четвърта глава са описани особеностите на някои български синтезатори на реч. Пета глава представя методи за промяна на височината и продължителността на звука в запис, разработени от автора, и експерименти по автоматично извличане на форманти.

В шеста глава са описани устройството и възможностите на хибридния синтезатор на реч от текст „Глас 2”.

Седма глава дава множество идеи и насоки за изследвания и разработка, които биха били продължение на дипломната работа –

В осма глава е заключението на дипломната работа. Списъкът с използвана литература е в девета глава.

1. Звук, музика и реч

1.1. Звук

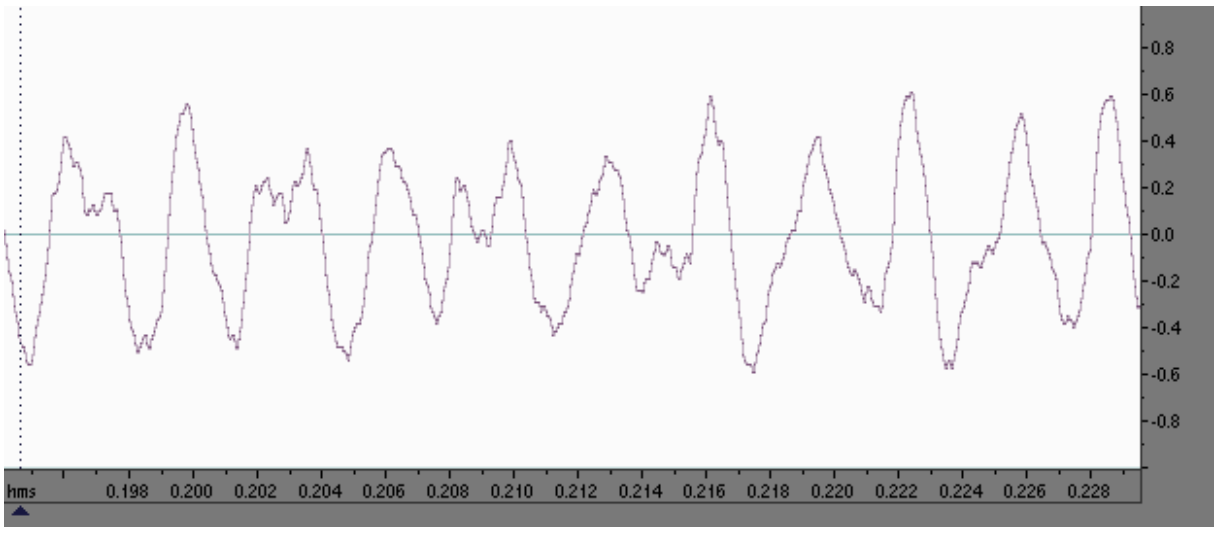
Във физически смисъл звукът представлява трептение на среда с определена честота и амплитуда. В паметта на изчислителна машина, звукът представлява цифров сигнал – поредица от отчети, описващи моментния интензитет на трептението през зададен интервал, определен от честотата на дискретизация. По-висока честота на дискретизация определя по-висока честотна вярност на записания звук спрямо аналоговия сигнал.

Човешкият слух може да долавя звуци между 16 и 20 кХц, като горната стойност е достижима само от малки деца. Максималната доловима от слуха честота спада с възрастта и слиза под 10 кХц при хората над 60 г. (Medscape) При цифров запис на звук, е необходима два пъти по-висока честота на дискретизация от аналоговата честотата на звука. Така запис при честота на дискретизация 44.1 кХц (качество на аудио-CD) може да запише звук с честота 22.1 кХц..

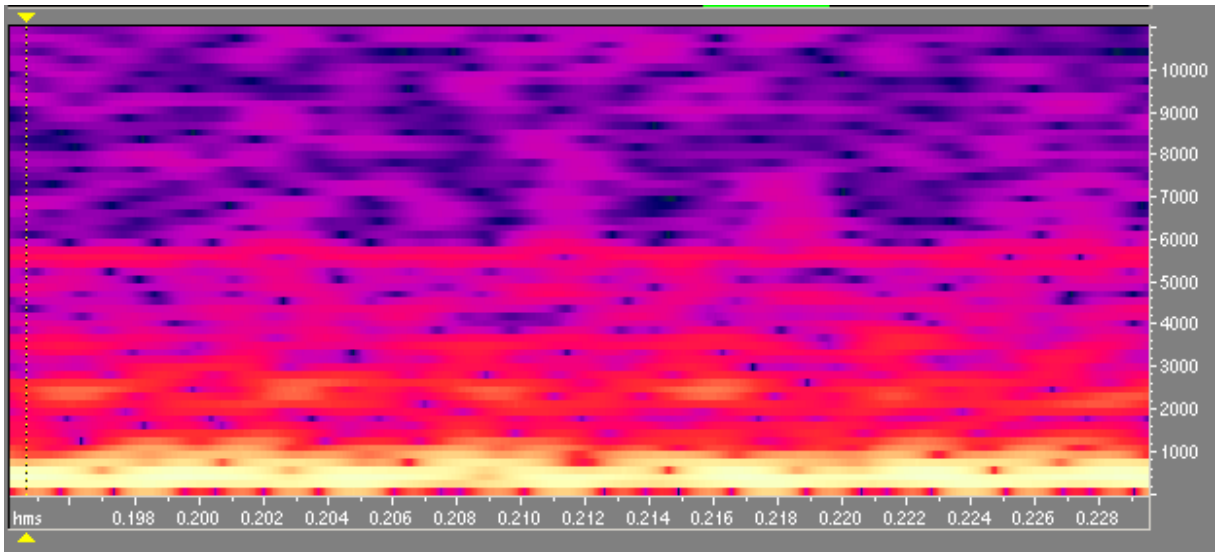
Субективното възприятие за звук представлява усещане за различна височина, гръмкост, както и усещане за шум (непериодично трептение).

1.1.1 Изобразяване на звука

Звукът може да се представи чрез вълни и чрез спектограми.



Фиг. 1 - Вълнова времедиаграма на запис на реч



Фиг. 2 - Спектрограма на записа на реч от фиг.

Вълновите представяния описват амплитудата на звуковата вълна през определен интервал, зависещ от честотата на дискретизация на звука. Стойностите в междинните моменти се интерполират от механичните трептящите системи - високоговорителите. Графичното изобразяване на вълнови представяния е тривиално.

Спектралните представяния се по-сложен вид изобразяване на звука. Те се изграждат чрез серия от т.нар. бързи преобразувания на Фурие (БПФ или FFT, Fast Fourier Transform) върху последователни кратки откъси от звукозаписа. По абсцисата на спектрограмата отново е времето, както при вълновата времедиаграма, но по ординатата е нанесена честотата, а не амплитудата. Различните цветове показват амплитудата на хипотетични съставящи периодични трептения със съответна честота: от бяло-жълто (най-голяма енергия) до тъмносиньо-черно (най-слаба енергия). На примерната диаграма на запис на реч може да се забележи например, че най-голяма амплитуда имат хармониците с честота под 1000 Хц.

1.1.2. Основни методи за анализ на звук

Основният метод за анализ и синтез на звук е бързото преобразование на Фурие (БПФ, FFT), който се основава на т.нар. ред на Фурие.

Ред на Фурие (Fourier) – чрез преобразование на Фурие всеки сигнал може да бъде разложен до ред она Фурие, състоящ се от множество синусоидални трептения с определени съотношения на фазата и амплитудата. Чрез наслагване на множество от функции от ред на Фурие може да се генерира сигнал с произволни характеристики. Трептенията от ред на Фурие се наричат още хармоници. Честотата на хармоника с най-голяма амплитуда е основната честота на дадено периодично трептене, и височината на тона, която се възприема субективно.

Бързо преобразование на Фурие – набор ефективни алгоритми за разлагане на цифров сигнал до ред на Фурие. БПФ е мощен метод, който е в основата на цифровата обработка на сигнали, но притежава някои недостатъци, защото е оптимизиран за скорост, за сметка на точността. Например, при работа с БПФ, входният сигнал се разделя на прозорци с дължина (брой отчети), която е степен на 2: 128, 256, 512, 1024... Дори за минимално увеличение на честотната точност е необходимо двукратно удължаване на прозореца. Това прави употребата на други алгоритми по-подходяща в някои случаи.

Алгоритъм на Гьорцел (Goertzel) – ефективен алгоритъм за откриване на хармоници с честоти в много тесни диапазони и на точно определени честоти (напр. разпознаване на сигналите от телефон при тонално избиране).

Алгоритъм на Чирп-Зед (Chirp-Z) – ефективен алгоритъм за разлагане на сигнал на хармоници, който при работата си използва БПФ, но постига по-висока честотна точност в рамките на по-къс прозорец.

Метод за откриване на почтипериодични трептения в запис на реч – ще бъде разгледан в глава 5.

1.2. Музика

Музиката представлява трептение, в което съществуват времеви и честотни зависимости, които субективно се възприемат като ритъм, мелодия, хармония.

Ритъмът представлява възприемане на определени събития през определено, предсказуемо от слушателя въз основа на предходната част от музиката, време.

От математическа гледна точка тоновете представляват периодични функции с определени честоти на основния тон; мелодиите са поредици от периодични трептения, които се вместват в определена тоналност, което на практика означава че между честотите на основните хармоници на тоновете от мелодията винаги се запазват съотношения в определени граници.

1.3. Реч

Речта е сигнал, в който субективно могат да бъдат уловени членоразделни звукови *сегменти* и да се разпознаят *суперсегментни* единици.

Сегментните единици са звуковете на речта (гласни, съгласни). Те могат да се отделят една от друга и да се изговарят самостоятелно. Суперсегментните единици са ударенията и различните видове интонация. Те влияят върху начина на реализация на сегментните единици, но освен това носят по-абстрактна информация. (Бояджиев и Тилков, 1999). Например в общия случай ударението удължава гласната и увеличава височината на основния й тон (влияе на сегментно ниво), а интонацията представлява промяната на честота на основния тон; смисловото ударение обаче – подчертаването на ударението в някоя определена дума и интонацията носят прагматична информация за отношението или емоцията на говорещия. Суперсегментните единици се наричат още *прозодия*.

1.4. Пеене

Пеенето може да се класифицира като говор с някои особености:

- Дължината на гласните зависи от ритъма и темпото на песента.
- Основната честота на гласните се настройва с честотата на изпятия тон.
- Гласните се орнаментират с вибрато (люлеене на основния тон) и други „извивки”.
- Силата на звука се променя по-динамично.

2. Принципи на синтез на реч от текст

2.1. Обща схема

Входните данни на синтезаторите на реч – текстът – минават през редица обработки, преди да прозвучат вокализирани. Най-общо те биха могли да се обособят по следния начин.

1. Предварителна обработка на текста.
2. Преобразуване на текста във фонеми.
3. Генерация.

В литературата (Lemmetty, 1999) се среща и по-грубо разделение на етапите:

1. Синтез на високо ниво
2. Синтез на ниско ниво.

Синтезът на високо ниво включва обработката до свеждането до фонеми, преди генерацията на звук, а синтезът на ниско ниво е звуковият синтез.

Етапите на обработка могат да се разбият на подетапи:

1. Предварителна обработка на текста

1.1. Сегментация – първично разделяне на думи.

1.2. Сканиране - разпознаване на числа, съкращения, съставни думи, специални думи (команди за настройки на синтезатора и пр.).

1.3. Нормализация – преобразуване на съкращенията и числата в текст.

1.4. Сегментиране на изречения.

1.5. Маркиране на частите на речта, морфологичен и синтактичен разбор.

1.6. Маркиране на прозодия – откриване на интонационни контури, планиране на темпото на говор и паузите; планиране на тембъра.

2. Преобразуване на текста във фонеми

2.1. Използване на речник с ударения

2.2. Прилагане на правила за редукция, потъмнявания, обеззвучаване, сливане и др. фонетични промени.

3. Звуков синтез

а) Формантен синтез – синтезиране на звук по параметри за промяна на основния тон и хармониците му и др., синтез на шум; интерполации при преходите между форманти и др.

б) Синтез със слепване – извличане на подходящите речеви фрагменти от базата данни; слепване; анализ на звука и обработка на местата на слепване за създаване на по-плавно свързване; откриване на носещи честоти и премодулация за промяна на интонацията...

в) Артикулационен синтез – симулиране на работата на говорен апарат.

2.2. Предварителна обработка на текста

.

1. Сегментиране текста до думи.

Текстът се разделя на думи, или по-точно на низове, ограничени с разделителни символи, тъй като част от думите могат да са съкращения и числа.

2. Сканиране

Разпознаване на съкращения, числа, имена със специално произношение и др. Разпознаване на команди за синтезатора (промяна на скоростта, на тембъра и пр.)

3. Сегментиране на изречения и фрази (изкази)

Това разделяне подпомага планиране на изкази между своеобразни „вдишвания”, както и обособяването на интонационни контури в рамките на изреченията и фразата.

4. Нормализация

Откритите съкращения и числа трябва да се преобразуват в текст, за да могат да бъдат изговорени. Този етап може да съдържа голяма многозначност, и задачата не е тривиална. Например как да се реши дали „г.” да се прочете като година, или като грама? Могат да се използват се речници със съкращения, евристики спрямо контекста като например корпуси с колокации, където е маркирано кое съкращение е по-вероятно в даден контекст.

5. Маркиране на частите на речта, морфологичен и синтактичен разбор.

Тази обработка подпомага правилното откриване на ударенията и изчисляване

на фонетичните промени. Тези задачи също не са тривиални, и съществуващите методи допускат грешки.

6. Маркиране на прозодия.

Прозодията в речта включва интонацията, темпото на говор, паузите, тембъра. Тя носи информация за целта на изказването (въпрос, молба, съобщение и пр.), настроението, отношението, емоцията и др. състояния на говорещия.

За да се синтезира немонотонен говор, който е неприятен, е необходимо поне в някаква степен да се моделира интонация. Най-често това се прави въз основа на препинателните знаци в текста, като се работи с т.нар. интонационни контури. Разпознават се типове на изреченията – разказно, въпросително, възклицателно; откриват се вметнати изрази, изреждания и пр. Въз основа на препинателните знаци се избира подходящ интонационен контур – как се променя интонацията по време на изречението и/или фразата.

Съществуват синтезатори, които могат да моделират и емоции на говорещия (гняв, тъга, радост, умора, страх; възраст, пол и др.). (Burkhardt F. 2008). Добро качество се постигат някои системи, които модулират цялостен запис на реална реч, записан с неутрална емоция, напр. L2F INESC-ID.

2.3. Преобразуване на текста във фонеми

Трудността на тази задача зависи от особеностите на различните езици. Българският или италианският например има сравнително прости и еднозначни правила за преобразуване на текста във фонеми, докато при английският съществуват огромен брой изключения, които трябва да се имат предвид за да се постигне правилно произношение.

За езици с променливо ударение и редукция на неударените гласн като българският, е необходимо да се използва и речник с ударения.

2.4. Звуков синтез

Методите за звуков синтез на реч се разпределят в три категории:

- Синтез със слепване – използват се голям брой къси откъси от записи на реална реч (алофони, дифони; дори цели думи), които се слепват един след друг
- Формантен синтез – моделират се честотите на звуците на речта.
- Артикулационен синтез – симулира се самият говорен апарат в най-големи подробности.

Най-разпространеният метод днес е синтезът за слепване. Засега чрез него се постига най-високо качество на звука при изискване за генериране в реално време. Артикулационният синтез обещава напълно реалистичен говор, но все още е твърде сложен за реализация. Сегашните реализации не могат да достигнат достатъчно детайлно симулиране на вокалния тракт и не могат да работят в реално време.

Форманти се наричат усилените честоти на гласовия канал (Бояджиев и Тилков, 1999). Формантните синтезатори се опитват да моделират говор чрез генериране на сигнали с формантни честоти. Този тип синтез позволява малко изчислителни ресурси да се постигне разбираем плавен говор, на който с лекота могат да се параметризират скоростта, височината и тембъра, но гласът на формантните синтезатори често звучи „металически" и „нечовешки".

Съществуват и хибридни синтезатори, които комбинират формантен синтез и синтез със слепване. При микрофонемния метод например, се работи с фрагменти от реална реч, които обаче са малък брой и съдържат спектрална информация за формантните честоти, а не са записи на алофони, дифони и пр. Някои формантни синтезатори синтезират тоналните звуци, но използват цели откъси за беззвучните съгласни, които са много по-сложни за формантен синтез.

Освен това, съвременните синтезатори със слепване използват средства на цифровата обработка на сигнали за спектрален анализ, откриване на форманти и „повдигане" и „сваляне" на основния тон.

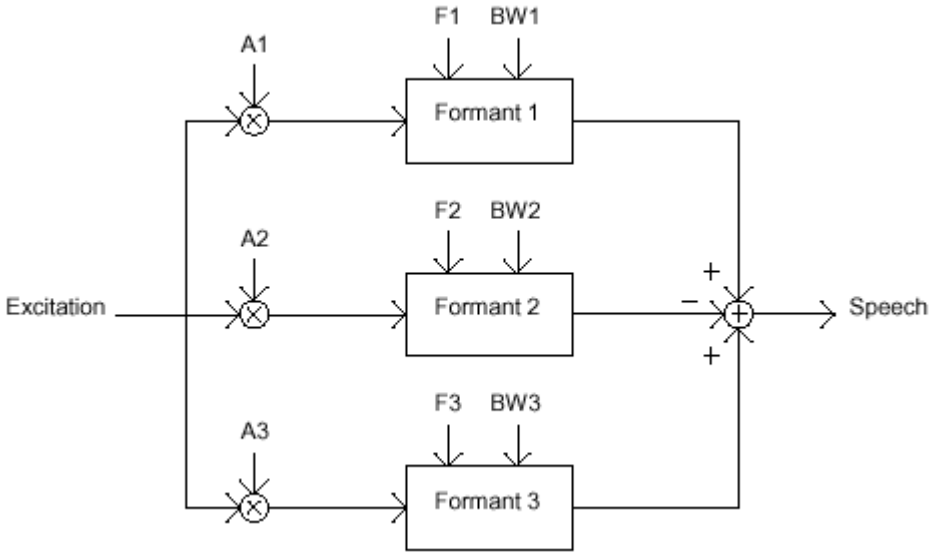
2.4.1. Формантен синтез

При формантният синтез звуците на речта се генерират чрез симулация на трептящи системи, които генерират трептения със спектрална структура сходна на структурата на речта. В основния метод – LPC , линейно прогнозиране – се използва

множество от предварително изчислени коефициенти (параметри) с които по време на синтеза се контролират:

- Основната честота и възбуждащ сигнал (F0)
- Амплитуда на възбуждащ сигнал (VO, Voice Excitation)
- Формантни честоти и амплитуди (F1…Fn и A1…An)
- Честоти на доплнителни нискочестотни трептящи системи (FN)
- Интензитета във високите и в ниските честоти (ALF, AHF)
- Честотна лента (BW)
- Съотношение сигнал/шум (Lemmetty, 1999)

Чрез тези параметри, възбуждащият сигнал, имитиращ въздушна струя в човешкия говорен апарат, се филтрира за да получи формата която би получил ако преминава през вокалния тракт



Фиг. 3 - Схема на паралелен формантен синтезатор (По Lemmetty, 1999)
Възбуждащият сигнал се филтрира и накрая се смесва.

Формантните синтезатори са най-ефективни откъм изчислителни ресурси и памет - българският синтезатор на реч за Правец-82 на Борислав Захариев например се е събирал в няколко КБ.

Обикновено говорът е с „металическо” или „изкуствено” звучене, но е плавен,

без накъсване. Силна страна на формантните синтезатори е възможността да се генерира разбираема реч с много висока скорост.

2.4.2. Синтез със слепване

На пръв поглед това е най-очевидният начин за създаване за синтезатор на реч. Той се състои се в извличане на голям брой фрагменти от записи на реална реч – алофони, дифони, срички, думи – които след това се слепват един след друг.

Основните трудности при този метод са големият брой фрагменти, които трябва да бъдат систематично извлечени и каталогизирани (над 1000 при използване на дифони), и изглаждането на местата на слепване между два фрагмента, без получаване на неприятни доловими от слушателя изкривявания.

Ако не се извърши обработка на местата на слепване, се получава характерно накъсване на речта, подобно на сричане, тъй като съседните фрагменти са от различни записи и от различно фонетично обкръжение.

Този проблем се решава като се извършват различни видове спектрални трансформации и интерполации, чрез варианти на PSOLA (Pitch Synchronous Overlap Add) - FD-PSOLA и TD-PSOLA. (Lemmetty, 1999). FD-PSOLA се отнася за промяна на височината (Frequency Domain – честотна област), а TD-PSOLA за промяна на продължителността на звука (Time Domain). Чрез тези методи се излглаждат местата на слепване и се променя бързината на говора. Тези спектрални преобразувания обаче често водят до появяване на паразитни хармоници и не винаги успяват да се преборят с неестественост на преходите, което понижава качеството на звука.

Принципът на PSOLA е разгледан накратко в глава 5, заедно с опростени методи за промяна на височината и продължителността на произволен звук, които постигат сходен ефект.

Някои изследователи причисляват към синтеза със слепване и синтеза с «микрофонеми» (Lemmetty, 1999). Подобен метод се използва в “Глас” за синтез на тонални звуци, и той ще бъде разгледан в глава 6.

2.4.3. Артикулационен синтез

Артикулационните синтезатори на реч моделират физически акустичните ефекти на модел на човешки говорен апарат. Теоретично, този метод би трябвало да

доведе до напълно реалистичен говор. Моделирането на говорния апарат на толкова ниско ниво обаче е изключително тежко, и все още няма резултати, които да позволяват използването на този метод в реални приложения, които да конкурират другите методи. Интересни са експериментите, които могат да се провеждат чрез приложението Praat (Boersma, Weenink).

2.5. Синтез на пеене

Пеещите синтезатори са по-малко познати и разработвани от синтезаторите на обикновена реч, донякъде заради по-малкото практическо приложение на този тип синтез. Вокодерите и други техники за цифрова обработка на сигнали позволяват звуци на реално пеене да се преобразува така че да звучат подобно на звука от съвременните синтезатори на пеене, което донякъде ги обезсмисля за комерсиални приложения; от друга страна, обикновените потребители имат по-малка нужда от синтезатори на пеене, отколкото от синтезатори на обикновена реч.

Съществуват формантни (синусоидални) и слепващи пеещи синтезатори, като засега чрез слепващите се постига по-високо качество.

При формантния синтез, принципът на работа е подобен на синтеза на реч, като се спазват особеностите на пеенето: говорът е съобразен със зададен ритъм, а гласните са удължени и са модулирани върху основна честоти, съвпадащи с тоновете на изпълняваната мелодия. От синусоидалните синтезатори може да се спомене (LYRICOS).

Пеещите синтезатори със слепване използват предварително извлечени фрагменти, най-често дифони, от обикновена реч или от пеене. При синтеза основният тон се променя до желаната октава чрез спектрални преобразувания. Примери за слепващи пеещи синтезатори са *Whistler* на Microsoft Research и *Vocaloid* на Yamaha.

Whistler е самообучаващ се синтезатор на реч със слепване, който извлича дифони от маркиран от речев корпус. За да се синтезират звуци на пеене, системата извършва спектрални преобразувания.

Vocaloid 2 (Kenmochi 2007) е комерсиален продукт на Yamaha.. Той използва предварително записани кратки откъси от реално пеене – дифони и др. Откъсите след

това се обработват с методи на цифровата обработка на сигнали и се постига много високо качество.

3. Приложения на синтеза на реч

Синтезаторите на реч намират приложение в множество области и са незаменими инструменти за хора с увреждания.

3.1. Помощни средства за незрящи

Това е сферата, в която синтезаторите на реч са най-полезни за потребителите, а в много случаи – незаменими. Първата система текст-реч, подпомагаща незрящи хора, е разработена от Реймънд Кърцуейл (Raymond Kurzweil) през 1976 г. и включва хардуерна многошрифтова система за оптическо разпознаване на символи, която изговаря по относително разбираем начин прочетения текст. Системата обаче е изключително скъпа и навлиза само в някои библиотеки. (Lemmetty, 1999)

В края на 70-те започват да се създават комерсиални синтезатори на реч от текст за персонални компютри – Apple][, Commodore 64 и др. Във втората половина на 80-те години се появяват т.нар. екранни четци - IBM Screen Reader(1986) и JAWS (1989). Функцията им е да прочитат съдържанието на екрана, с което позволяват на незрящи да работят пълноценно с компютър.

3.2. Помощни средства за хора със слухови и говорни увреждания

Децата, родени глухи, не могат да се научат да говорят нормално. Чрез синтезаторите на реч обаче, те могат да общуват наживо малко по-лесно с хора, които не разбират езика на жестовете. Синтезаторите на реч, синхронно с изображения на движения на устните, също са полезни за хора със слухови увреждания.

По подобен начин синтезаторите на реч са полезни и за хора, които по някакви причини са загубили гласа си - увреждане на гласните струни или езика, мускулна дистрофия и др.

Проблем при тези приложения на синтезаторите на реч представлява реалистичното изразяване на емоции, тъй като става въпрос за общуване лице в

лице.

3.3. Приложения в образованието

В образованието, синтезаторите на реч са може би най-полезни за обучение на деца със затруднения в четенето (дислексия) с това, че могат да позволят на ученика да прослушва думите самостоятелно. А играчки като Speak & Spell (Texas Instruments, 1980) и различни игрови софтуерни продукти за деца в предучилищна възраст показват, че синтезаторите на реч могат да направят обучението по четене по-забавно за всички деца.

Синтезаторите на реч могат да се използват и като помощно средство при учене на чужд език.

3.4. Мултимедийни приложения

Синтез на реч се използва в четци на електронна поща - потребителят може да прослуша съобщенията си, без да отваря приложение или дори без да е седнал на компютъра - например докато закусва. Могат да бъдат прослушвани произволни текстове, докато зрително се четат други текстове или пък се върши работа, без да се гледа в екран. В някои случаи човек няма възможност или не бива да разсейва вниманието си, например ако получи текстово съобщение, докато шофира. Единственият начин да го възприеме тогава е като го прослуша.

Съществуват синтезатори на реч, които генерират и движения на устните и лицеви изражения. Те се използват в интерфейса на различни видове мултимедийни агенти, които дават напътствия на потребителя при работа с определено софтуерно приложение, или пък симулират водене на свободен разговор с него („чат-ботове”).

Синтезът на реч се използват за менюта на телефонни оператори и за автоматично съобщаване на разписания на гари и летища, като в тези случаи най-често се използват синтезатори с предварително записани думи и фрази, които са частен случай на синтезаторите със слепване.

Обещаващи са и приложенията на говорещите интерфейси в съчетание с разпознаване на реч в т.нар. „вездесъща компютризация” (ubiquitous computing) на бъдещето, когато разпознаване на реч и речеви интерфейси ще се използват дори за управление на домакинските уреди.

4. Български синтезатори на реч

4.1. Формантни синтезатори

4.1.1. Синтезаторът на Борислав Захариев

За пионер в синтеза на реч в България се смята инж. Борислав Захариев, който създава формантни синтезатори на реч за Правец-82 и Правец-8Д* през втората половина на 80-те години. Синтезът с Правец-82 е специфичен заради ограниченията на машината. Звукова система на компютъра е двоична, и може да се кара да трепти единствено между две крайни положения, като липсва контрол на амплитудата. По тази причина генерираният звук от Правец-82 (Apple][) и съвместимите с него представлява поредица от правоъгълни импулси, макар че несъвършенствата на говорителя и инерцията на мембраната разбира се не позволяват реалната амплитудна-характеристика на звука да бъде както на фигурата по-долу.

. Честотата на звука се контролира чрез широчинно-импулсна модулация, като ширината на импулсите се задава чрез изпълнение на подходящи запълващи инструкции, през които сигналът се задържа на ниво „0” или „1”.

Фиг. 4 - Широчинно-импулсна модулация при синтезиране на звук с Правец-82

Според откъс от статия в сп. „Компютър за вас” (КВ, бр. 7-8, 1990) съществува и синтезатор на реч за Правец-8Д от Б. Захариев, създаден през 1987. Правец-8Д притежава много по-съвършена звукова система – вграден честотно-модулиращ синтезатор, с който могат да се генерира многоканален звук с различна честота и амплитуда. Информацията за синтезатора за Правец-8Д обаче не е потвърдена от други източници.

Съществува версия на синтезатора за Правец-82 за Правец-16 (IBM-PC), който също има двоичен говорител. Тази версия се използва в първия български екранен четец за ДОС.

4.1.2. Txt2Speech Demo 1.12

Това е експериментална разработка на формантен синтезатор от анонимен автор, който предлага да разработи по-добър синтезатор в помощ на незрящите. Системата използва класическия метод с линейно прогнозиране (LPC). (Txt2Spc, 2003). Синтезаторът не претендира за завършеност и е характерен с неприятно металическо звучене.

4.2. Синтезатори със слепване

4.2.1. SpeechLab 2

“SpeechLab 2” (2005) е разработка на Българската Асоциация за Компютърна Лингвистика. Системата предлага 3 гласа (два женски и един мъжки), моделирани с по 1200 дифона, извлечени от студийни записи с ларингофон. SpeechLab извършва голям брой обработки на естествен език преди озвучаването – разпознават се дати, използва разширяем речник за нормализация на съкращения и специални думи, и претендира че спазва всички фонетични правила за редукция, загуба на ударения на клитики, междусловна асимилация и др. Могат да се изговарят и някои английски думи. (БАКЛ, 2005)

Синтезаторът генерира интонационни контури, може да променя скоростта и височината на гласа и извършва спектрална интерполация между дифоните. Въпреки това обаче се усеща характерното за този тип синтезатори “сричане”.

Авторите на системата са представили специално демо на синтез на пеене чрез моделиране на интонацията.

Системата е съвместима с MS SAPI 5.1 и може да се използва и като екранен четец.

4.2.2. СЛОГ

СЛОГ (2003) - Система за лингвистично осигуряване на говора – е разработен в катедра Компютърна информатика на Пловдивския Университет (Тотков, Ангелова,

Благоев, 2003). Тя използва моделиране чрез 495 алофона, извлечени ръчно. Предварителната обработка на текста е силно развита: извършва се морфологичен анализ, правилно се открива ударението, изчисляват се редукции и др. фонетични промени. Разпознават се дати, числа в различни формати и съкращения. СЛОГ обаче е характерен със силно насеченото си звучене, заради използването на много кратки фрагменти (алофони) и липсата на интерполация между тях.

4.2.3. Други синтезатори със слепване

New voice (2003) е разработка на българската фирма Manro, и може да работи заедно с екранен четец.

Reader (2003) е програма за експериментиране със синтез със слепване, създадена от инж. Иван Иванов. (Иванов, 2003) Тя позволява на потребителя да създава библиотеки от фрагменти чрез които да се изгради гласов модел с гласа на потребителя. Накъсаността на синтеза е голяма, защото не се извършва спектрално изглаждане.

4.3. Хибридни синтезатори

“Глас” е хибриден синтезатор, разработен през 2004 г. и е първата версия на синтезатора, разглеждан в тази дипломна работа. Една от целите на създаването му е да накара СЛОГ да говори по-плавно.

Устройството на втората версия на „Глас” ще бъде разгледано в глава 6.

4.4. Български екранни четци

Първият екранен четец на български - за Правец-82 – е създаден от Борислав Захариев през 1990 г. (Стойчева). През 1991 г. той го преработва за Правец-16. През 1993-94 г. Торос Хованесян подобрява произношението и предлага нови програмни продукти като «Електронна бланка» и англо-български и българо-английски речници. Незрящият програмист Хюсеин Исмаил адаптира екранния четец Vocal eyes за български през 1995 и създава «Ехо» за ДОС, който позволява пълноценна работа с много приложни програми. През 2003 г. фирмата „Manro” създава екранен четец за

Windows. Две години по-късно Българската Асоциация за Компютърна Лингвистика (БАКЛ) с подкрепата на Microsoft разработва „Speech Lab 2”, който се използва от незрящите в момента.

5. Методи за обработка на речеви сигнали

В тази глава са представени няколко метода, чрез които може да се променя височината и темпото на произволни записи, в частност на реч, открити от автора на дипломната работа при негови експерименти през 2000 г. и 2004 г. Резултатът от действието на ефекта е подобен на ефекта от методите PSOLA, използвани в синтезаторите със слепване.

Представени са и експерименти и идеи за автоматично извличане на форманти, които да подпомогнат автоматизираното създаване на гласови модели на потребители.

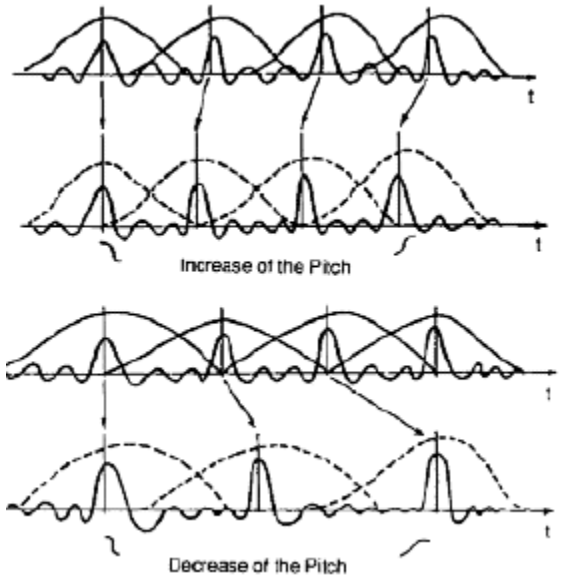
Звукова тъкан – *звукова тъкан* в работата се нарича свръхкратък откъс (прозорец) от запис на звук с продължителност 0.01 - 0.02 сек. Това е по-дълго време от един формант (около 0.005 сек при честота 200 Хц), но и много по-кратко от фонема при нормален говор (0.05 – 0.1 сек). Времето за един прозорец („късче”) звукова тъкан е толкова кратко, че слухът не може да улови отделен звук, макар че записът като цяло се състои от огромен брой „късчета” звукова тъкан. Тази особеност на слуха се използва в методите, описани по-долу.

Някои синтезатори, като Mikropuhe за финландски (Lemmetty 1999, с. 78) използва подобно на понятие - „микрофонема”, като в него така се наричат кратки откъси от реална реч, с дължина около 0.01 сек, които се слепват. „Глас” също използва подбрани свръх кратки откъси от реална реч (0.006 – 0.009 с) като модели на формантите на тоналните звуци.

5.1. PSOLA

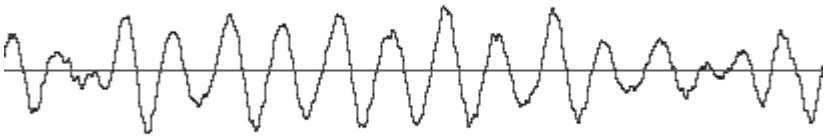
PSOLA – Pitch Synchronous Overlap Add – е множество от методи за промяна на височината и продължителността на звучене на говорни сигнали, и за изглаждане на прехода между записи на фрагменти от реч. Този метод е разработен първоначално във France Telecom в края на 80-те години (Lemmety, 1999), и бързо намира приложение в синтезаторите със слепване.

Същността на метода е в атоматичното откриване или предварително ръчно маркиране на формантни маркери и след това “разтягане” или “свиване” на отрязъците между два маркера, което звучи като промяна на височината на тона.



Фиг. 5 - Илюстрация на метода PSOLA (по Lemmetty, 1999)

Трябва да се отбележи, че прецизното откриване на формантните маркери и отделяне на формантните честоти не е толкова тривиална задача, както изглежда на фиг.5, която показва сигнал със силно изразена основна честота.



Фиг. 6 - Сигнал, при който откриването на формантни маркери изисква по-сложен анализ

С приближение, задачата е решима например като се използва БПФ или метод на Чирп-Зед и се открие честотата на основния тон.

В т. 5.4. са разгледани опити за автоматично откриване на микрофонеми, чиито основен тон съвпада с основната формантната честота. В тези експерименти се използват други методи - евристики за търсене на почтипериодични функции с размито сравнение, и анализ на статични и динамични характеристики на сигнала като: брой смени на знака, стойности на минимуми и максимуми на съседни периоди на почтипериодични функции, размах на сигнала в рамките на период и др.

5.2. Методи за промяна на височината на звука в запис, без промяна на продължителността

Предложеният от автора метод не използва формантни маркери, а извършва преобразования, подобни на преобразованията при PSOLA върху сегменти с дължина в рамките на „късче звукова тъкан” и/или микрофонема.

5.2.1. Метод за повишаване на височината, без промяна на продължителността

1. Сигналът се разделя на много кратки откъси, „прозорци”, подобно на процедурата при работа с БПФ. За разлика от него обаче прозорците при този метод са много по-кратки, отколкото при БПФ, и са с продължителност в рамките на времето на „микрофонема” или „късче звукова тъкан” – около и под 0.01 s. За толкова кратко време човешкият слух не може да различи отделен звук.

2. Сигналът във всяко парченце от звуковата тъкан се компресира, така че да се вмести в по-кратко време.

3. Новополучената звукова тъкан се копира два (или цяло число пъти). За по-добра спектрална гладкост вторият може да се извършва амплитудна интерполация на няколко отчета в областта на свързване.

5.2.2. Метод за понижаване на височината, без промяна на продължителността

Методът за понижаване на височината е аналогичен. Разликата е в това, че сигналът от извлечената звукова тъкан се ”разтяга” така че да се вмести в повече време. Това налага при извличане на звукова тъкан или да се прескача по един прозорец, или два съседни прозореца да се интерполират в един и след това интерполираната звукова тъкан да се разтегне във време за два прозореца.

За изглаждане на връзката между късчетата звукова тъкан и в двата случая може да се използва проста интерполация на амплитудата в много малка област на свързването.

5.3. Метод за промяна на продължителността на звука, без промяна на честотата

В този случай решението е дори по-просто:

За удължаване на продължителността на звука цяло число пъти, късчетата звукова тъкан се копират два, три и т.н. пъти. Така спектралната картина по време на един прозорец е запазена и тонът звучи удължен.

За съкращаване на продължителността на звука цяло число пъти, късчетата звукова тъкан се копират в изходния запис като се прескачат n-1 прозореца, където n е кратността на удължаването; или n прозореца се усредняват до един, който се копира в изходния запис.

Възможни са прости изглаждания, подобни на тези при предходните методи.

5.4. Опити за автоматично извличане на тонални форманти от записи на реч

В тази точка са описани някои експерименти и насоки за изследвания и експерименти за автоматично разпознаване и извличане на тонални форманти от записи на реч.

5.4.1. Класификация на звуците на речта от гледна точка на анализа и генерирането

Първата стъпка на анализа представлява разделяне на записа на множество кратки „прозорци” и бърза класификация на амплитудно-честотните им характеристики на:

- тишина (ниска амплитуда)
- шум (висока честота)
- тон или преход (средна честота)

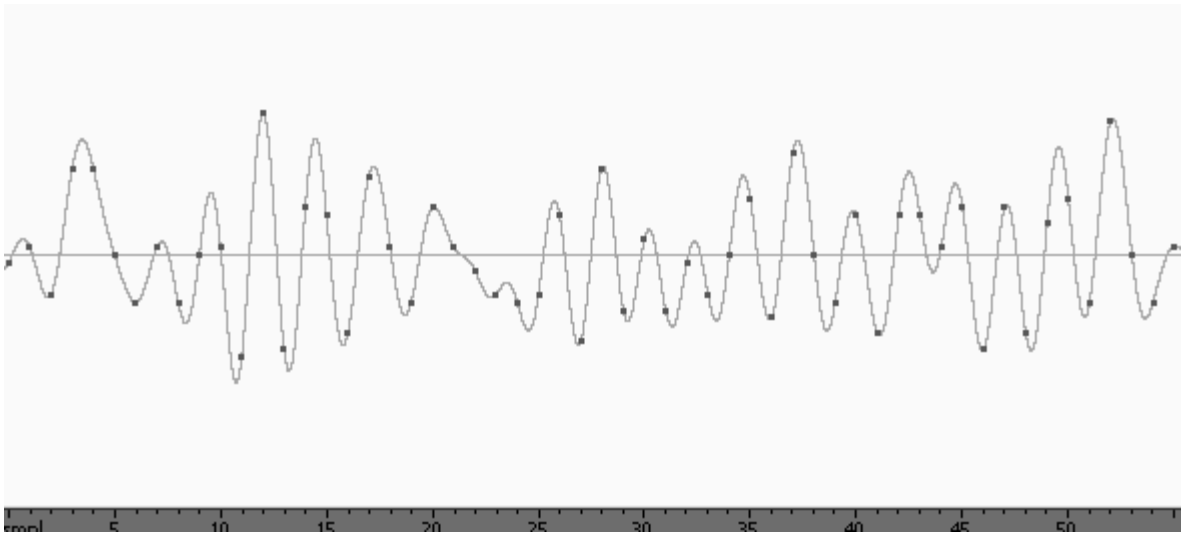
Търсенето на форманти става в областите със средна честота.

5.4.1.1. Тишина

Тишина е област, в която сигналът има енергия, която е под дадено прагово ниво, без значение от спектъра му. Енергията на сигнала е размахът, разстоянието между два съседни пикови момента с различни знаци. В речта тези моменти са паузаи или част от момента на изговор на беззвучна съгласна. Откриването на тишина става, като се измери средната амплитуда в даден прозорец. Усредняването помага да се елиминират кратки пикове. Прагът на тишина може да се извлича автоматично от запис, в който звукът е класифициран и маркиран от човек.

5.4.1.2. Шум

Формално, шум е непериодичен сигнал. На практика, при речеви сигнали, като шум може да се класифицира област, в която не само липсва периодичност с избрана точност на измерването й, но и когато е отчетена честотата на сигнала над 5-6 кХц с амплитуда над определена прагова стойност, защото основната енергия на говора е съсредоточена в много по-ниски честоти - до около 3-4 кХц.



Фиг. 7 - Шум

Ефективни методи за откриване на наличието на високочестотен сигнал:

1. Преброяване на честотата на смените на знака на сигнала в кратък интервал - ако сигналът е нормализиран спрямо нулата, както на фиг. 7.
2. Преброяване на честотата на смяна на знака на нарастване на амплитудата на сигнала.

5.4.1.3. Преход

Преходи са области, в които се отчита основен тон в в зададени честотни граници (предполагаем диапазон на основни честоти на форманти), но основният тон на търпи изменения и параметрите му са неустойчиви.

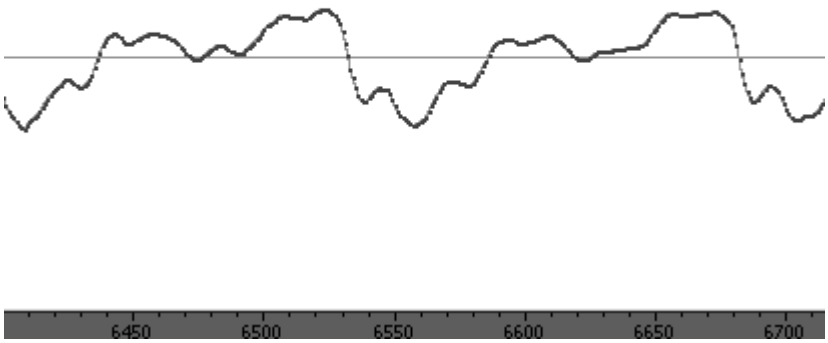
5.4.2. Характеристики на сигнала, които могат да се използват за разпознаване на форманти и откриване на преходи

1. Периодичност при определена точност на сравнение (почти периодични функции)
2. Брой устойчиви периоди на хипотетичните почтипериодични функции - поне 2, за да бъде потвърдена периодичност; поредица от повече устойчиви периоди показва че вероятно е открит чист формант.
3. Абсолютни стойности на локални минимуми и максимуми в рамките на прозорец и/или период от почтипериодична функция.
4. Размах: abs(loc_max - loc_min) в рамките на период
5. Енергия в един период от периодичната функция – Sum(abs(sample[i])), където sample[i] е амплитудата на отчет i.
6. Устойчивост на енергията/размаха/абсолютните стойности от период на период
7. Устойчивост на изменението на енергията от период на период (само растене или само намаляване може да показва преход)
8. Изменение на дължината на периода от период на период.
9. И мн. др. параметри, извлечени чрез по-сложни методи за обработка на сигнали

5.4.3. Метод за автоматично откриване на тонални форманти чрез размито сравнение и търсене на почтипериодични функции

Методът действа върху немаркиран запис и открива почтипериодични функции с дадена точност на сравнение. Входните данни в ескперимента са запис на дума, в която тоналните (периодични) звукове - гласни или носови съгласни - са между

беззвучни съгласни. При първичният анализ се открива шумът и се ограничават областите с речеви честоти. След това в локализираните области от записа се търсят устойчиви периоди на почтипериодични функции. Достоверността, че е открит формант, се потвърждава от броя на устойчивите периоди на откритата почтипериодична функция.



Фиг. 8 – два устойчиви периода при изговор на “Е”

Методът е подобен на начина по който работи човек, който ръчно извлича форманти и зрително разпознава устойчиви периоди.

1. Задава се честотна лента на търсене

Тя трябва да е в рамките на основната честота на формантите (напр. 100-500 Хц). Първите периодични функции, които ще се открият, ще са тези с по-висока честота, защото са по-кратки и по-бързо връщат знак за успех (виж по-долу). Затова се извършват повторни търсения, така че ако се открие периодична функция с по-ниска честота, тя отменя по-високата. Тук също обаче е възможно объркване, защото ако е зададена много ниска честота, е възможно да открием два (или повече) съседни периода на тона, които да бъдат разпознати като един период на функция с два пъти по-ниска честота. В такъв случай може да се проверява дали в откритата периодична функция няма периодична функция с цяло число периоди.

2. Задава се функция за "размито сравнение", която при най-простия вид да връща „еднакви", ако две стойности се различават не повече от някаква относителна или

абсолютна граница. Размитото сравняване е необходимо, защото е почти невероятно в реален запис на звук да открием "съвършена" периодична функция, в която стойностите съвпадат с максимална точност, но в същото време в запис на реч реално съществуват съседни периоди, които имат почти еднакъв цифров запис.

Например графиката на фиг. не е на периодична функция в строг смисъл, но наличието на два периода и приблизителната им еднаквост се виждат с просто око. Ако се намали разделителната способност, което се извършва чрез размито сравнение, в даден момент ще се получат и формално два еднакви периода.

След това методът следва определението за периодична функция: необходимо е стойностите й да се повтарят на разстояние един период.

1. Поставят се указател на място, което се избира за начало на функцията. (УКАЗ-0) и се създава друг, който ще обхожда (УКАЗ-1).

2. Поставя се друг указател на място, отдалечено от първия на разстояние минимална дължина на периода (УКАЗ-2), минималната честота; след това се създава трети указател, който да обхожда успоредно на първия (УКАЗ-3).

3. Сравняват се с размито сравнение моментните стойности на функцията на местата УКАЗ-3 и УКАЗ-1. „Момента стойност” могат да бъдат буквално моментните стойности на отчетите, усреднените стойност на амплитудата на няколко съседни отчети и др. по-сложни алгоритми за търсене на шаблони.

4. Ако размитотото сравнение върне еднаквост, хипотезата за перидочиност се запазва. Двата указателя се преместват с един отчет напред. Ако функцията за сравнение връща съвпадение до момента, в който УКАЗ-1 съвпадне с УКАЗ-3, то отбелязваме, че е открита периодична функция с два устойчиви периода. След това се проверява и за следващи устойчиви периоди по същия алгоритъм, като се приеме УКАЗ-3 за начален указател (УКАЗ-0), или като УКАЗ-0 си остане начален указател и се търсят нови съвпадения на моментните стойности в продължение на един период на новооткритата предполагаема почтипериодична функция.

5. Ако размитото сравнение върне "неравни", тогава не е открита периодична функция с начало УКАЗ-0; той се премества с избран брой отчета напред и се преминава към т. 1.

Отчети	Устойчиви периоди/ Период(отчета)	Смени на знака/Средна енергия на период	Размах в рамките на период
5173 - 5401 :	2 228	5 451453	7941
5912 - 6057 :	2 145	3 499961	13467
6860 - 7008 :	2 148	9 487657	14798
8030 - 8174 :	2 144	3 432548	12766
8393 - 8546 :	2 153	3 462276	11045
9349 - 9512 :	**3 163**	**3 522278**	**10741**
9512 - 9676 :	2 164	3 551688	11026
9676 - 9842 :	2 166	3 562663	10975
9842 - 10008 :	2 166	3 559344	11025
10008 - 10175 :	2 167	3 556995	10626
10175 - 10341 :	2 166	3 540333	10544
10341 - 10509 :	2 168	3 536316	10376
10582 - 10755 :	**3 173**	**3 552075**	**10465**
10755 - 10926 :	2 171	3 530208	10253
10926 - 11101 :	2 175	3 552146	10393
11101 - 11278 :	**3 177**	**3 552370**	**10198**
11278 - 11455 :	2 177	3 545550	10396
11455 - 11635 :	3 180	3 552083	10335
11635 - 11816 :	**3 181**	**3 567661**	**10390**
11816 - 11996 :	2 180	3 572385	10567
11996 - 12178 :	2 182	3 564907	10554
12178 - 12347 :	2 169	3 597736	11755
12347 - 12531 :	2 184	3 615298	10609
12604 - 12784 :	2 180	3 435160	8369
13368 - 13568 :	2 200	7 291158	9010

Табл. 1. – Резултати от експерименти с автоматично откриване на почтипериодични функции. Записът е на думата „цена” при 44 кХц/16-бита и е от базата данни на СЛОГ (Тотков и др. 2003)

5.4.3.1. Насоки за изследване и опити

- **Размито сравняване с натрупване на инерция** - всяко сравнение, при което разликата между сравняваните моментни стойности на функцията е в границите

на допустимото отклонение, се натрупва "запас", който да се използва за компресиране в случай на по-големи от допустимото, но краткотрайни отклонения вследствие на шум или включване на честоти, които са липсвали в предходния период. Ако е било натрупано достатъчно количество "запас" в предходни сравнения, то функцията връща знак "равни" дори ако разликата между текущо сравняваните стойности е над допустимото.

- **Автоматично извличане на динамиката на изменение на форманти по време на преходите от маркирани записи –** получените статистически данни могат да бъдат използвани за моделиране на трансформацията между форманти при синтеза на реч.

- **Опити с маркирани записи**, в които ръчно са отбелязани местата на преход и на устойчиви периоди. Чрез статистически методи да бъде намерена връзката между получените от анализа данни и маркираната информация – в данните по-горе например може да се забележат поредици от почтипериодични функции, които започват в момента в който свършва предната.

6. Устройство на Глас 2

6.1. Въведение

„Глас 2” е хибриден синтезатор на българска реч. Системата използва микрофонемен метод за синтез на тонални звуци, който някои изследователи класифицират като „слепващ” метод (Lemmetty, 1999), защото формантите са извлечени от реални записи и след това действително се слепват. Записите обаче са изключително кратки – в „Глас 2” те са с продължителността на един формант и след това се манпилуриат на ниво отделен период на основната честота на трептение, което е подобно на модулация на формантни честоти. Затова „Глас 2” звучи по-скоро като формантен синтезатор.

От друга страна, за синтеза на беззвучни съгласни действително се използва слепване на цялостни записи. Така е постигнат реалистичен звук на беззвучните съгласни, като в същото време са запазени особеностите на микрофонемния синтез, които позволяват гъвкави промени на време-честотните характеристики на тоналните звукове, без появата на паразитни хармоници. Още един елемент на слепване в синтезатора е използването на специални звуци за озвучаване на емотикони – хихикане, злобен смях, изненада, уплаха и др.

Системата е разработена на C++ и е интегрирана в средата и графичния интерфейс на Microsoft Win32 (Windows-95 и следващи версии). Използвани са средствата на Windows за многонишково програмиране, за да може интерфейсът да работи успоредно със синтезатора. Това позволява параметрите на говора да бъдат променяни без прекъсване на синтеза, както и преждевременно прекратяване на изговора.

6.2. Предварителна обработка на текста

Текстът минава редица преобразувания, преди да бъде вокализиран:

6.2.1. Разделяне на думи

Входният текст се разделя на думи с прост алгоритъм, който отделя низовете, намиращи се между интервали и знаци за пунктуация. Самите пунктуационни знаци също се запомнят като „думи” (tokens), за да могат да се използват в следващите фази – откриване на изречения и интонационни контури.

Откриването на изречения става въз основа на отделените препинателни знаци,

като се разпознават и пряка реч и съчетанията от символи „?!”, „!?”, „...”, „!...”, „?...”, за да може изреченията които ограничават да бъдат маркирани с подходящи интонационни контури. Разпознават се и емотикони.

6.2.2. Нормализация

В този етап съкращения и низове, за които важат специални фонетични правила се преобразуват до низове готови за преобразуване до фонеми, или се маркират съответните специални звуци, които трябва да прозвучат. Нормализирането обхваща цифрите, които се изговарят една по една, за да се избегне двусмислица при неправилно интерпретиране; съкращения на български и често срещани съкращения на английски – PC, CD и др; както и емотикони. Използват се дефиниции в разширяем речник.

6.2.3. Маркиране на изречения, интонационни и темброви контури

Въз основа на откриването на препинателните знаци и въпросителни думи се извършва маркиране на типове изречения – повествователно, възклицателно, въпросително.

Маркирането на типовете изречения представлява първия етап на маркирането на интонационните контури. Вторият етап е търсене на шаблони в рамките на изречението – изброявания (поредица от думи, разделени със запетайка или „;”) и вметнати изрази (текст между две запетайки или отваряща и затваряща скоба или между две тирета). След това се изчисляват коефициенти за промяна на височината на тона, които се разпределят върху гласните на сегментираните думи в изречението.

Използвани са основни мелодични контури за завършени, незавършени и въпросителни синтагми; еднородни части на речта. (Бояджиев и Тилков, 1999)

Експериментално са разработени прости контури за заповедна интонация и за емоционална реч, но поради многозначността на емоциите в текст, грешно предположение за емоцията може да объркa слушателя.

Пряката реч, както и текстът в кавички и скоби могат освен да бъдат включени в интонационни контури, да се използват и за маркиране на темброви контури, при което части от текста се изговарят с различен тембър, за да се открояят. Тембровите контури могат да бъдат много полезни при озвучаване на диалози и мултилози, например при синтезиране на говореща книга. За тази цел трябва да се разработят методи за

разпознаване на идентичности (Named Entity Recognition) и да се правят хипотези за това кой говори. Например, разпознават се говорител мъж и жена. Тогава в пасажа ще се използват два гласа - един с мъжки, и един глас с женски.

Разпознаването на говорителите, работещо с висока точност в произволен текст е много сложна задача, за която е необходим дълбок семантичен анализ, а може би и система, която да може да разтълкува текста като човек, който чете. В много случаи обаче разпознаването е възможно и с използване на лесно осъществими методи:
- търсене на шаблони в авторовите обяснения: «каза/рече [тя/името_на_жена]»;
- в места където няма авторова реч, може да се правят предположения за говорителя по съседните реплики, които са известни и ритъма на диалога (говорител А/ говорител Б/говорител А/);
- разпознаването е тривиално, ако говорителите и разказвачът са винаги изрично отбелязани, както е например в текстовете на пиеси или във филмовите сценарии

6.2.4. Преобразуване на текста във фонеми

Тоези етап се разделя на няколко подетапа:

6.2.4.1. Маркиране на ударенията

Използва се функцията Stress97 от библиотеката libmorf.dll, разработена в катедрата по Компютърна информатика. Функцията приема следните аргументи: (char word, short s1, short s2, short s3). Първият е входната дума; в s1 се връща номерът на знака в низа, върху който пада ударението, ако е разпознато. В s2 и s3 се записват алтернативните места на ударението при многозначност. „Глас 2” използва само първото предложено ударение, защото при многозначност няма средства, с които да реши кое е правилното.В случай че не е разпознато мястото на ударението, функцията връща 0, и ударение не се маркира.

Ударените гласни се маркират като удължени, като удължаването им спрямо неударените може да се настройва.

6.2.4.2. Маркиране на редукции на гласните

В българския език, гласните, които не са под ударение търпят определени промени. Използват се правила, дефинирани в СЛОГ (Тотков и др., 2003) и подходящи форманти. Виж т. 6.4.

6.2.4.3. Маркиране на меко „л”

Тази обработка се грижи за правилното озвучаване на „л”, в зависимост от това дали е следвано от мека или твърда гласна (леля – две меки Л-та, и „лодка” – твърдо Л).

6.2.4.4. Маркиране на обеззвучавания и крайни дълги съгласни

Звучните съгласни се обеззвучават, когато са в края на думата. По подобен начин беззвучните съгласни в края на думата и когато се следвани от беззвучна съгласна са по-дълги и звучат по различен начин отколкото в останалите случаи.

6.3. Форманти, шумове и междуметия

Формантите на тоналните звукове в „Глас” представляват един период от хипотетична почтипериодична функция, която произвежда трептението на гласна или звучна съгласна. Поради това че са извлечени от реални записи, се използва и терминът «микрофонеми».

Микрофонемите са извлечени ръчноот амплитудни времедиаграми на подбрани реални записи на реч, като основната евристика при извличането е трептението да изглежда като най-устойчив период от няколко периода на почтипериодична функция. Този период се намира приблизително в момента на най-устойчиво подаване на въздух и най-устойчива конфигурация на говорния апарат. Също така, началото и краят на условната почтипериодична функция се взимат в момента в който сигналът има амплитуда 0. Това позволява дори при най-просият синтез с този формант, дори без да се извършва интерполация на преходи, да се получи плавен звук през накъсване.

Шумовете - записите на беззвучните съгласни - са извлечени като е спазвано правилото да не започват и да не завършват с преход, тъй като той би предизвикал неестествено звучене. Всъщност осезаемата част на преходите между тонални звукове и шумове е от страната на тоналния звук, затова не е трудно да се изпълни това изискване. Също така, амплитудата на сигнала в началото и в края на записа трябва бъде нула.

Статичната информация, описваща гласовия модел, се съхранява в множество от RIFF wav файлове. Файловете се пакетират и се зареждат наведнъж при стартиране, след което достъп до тях се получава като до виртуална файлова система. Трябва да се отбележи, че макар и съхранени в обикновени Wav-файлове, записите на формантите не са членоразделни звуци – дължината им е около 0.005 сек (при честота 200 Хц), което звучи като пукане, ако се възпроизведе директно.

На най-ниско ниво, формантите и шумовете са дефинирани в конфигурационния

Файл vhod.dll

1 1.wav N

2 2.wav T 5 0.4

3 3.wav N

4 4.wav N

5 5.wav T 7 0.45

...

Команда	Име на звуков файл	Тип	Периоди	Преход
1	1.wav	N		
2	2.wav	T	5	0.4
3	3.wav	N		
,,,				
65	62.wav	T	4	0.3

Табл. 2. – Формат на дефиниране на форманти и шумове

Първото число е номер на командата – използва вътрешно от синтезатора. Следва име на съответен звуков файл. След това се задава типа на звука - (N, Noise) са шумовете и беззвучните съгласни, а (T, Tone) – тоналните звуци - гласните и звучните съгласни. Четвъртата стойност е цяло число, което определя броя периоди на форманта, които се генерират по подразбиране. Коефициентът на прехода е реално число между 0 и 1 и определя начина на интерполация при срещане на два съседни тонални звука. Интерполацията ще бъде разгледана по-долу.

За шумовете не се задават брой периоди – те се копират направо в звуковия буфер и между тях и съседните звукове не се генерират преходи.

Звуковите файлове трябва да са във формат RIFF Windows PCM, 16-bit mono, и да са записани на една и съща честота. За гласовия модел в „Глас” е използвана честота 22050 Хц.

Списък с форманти, шумове и специални звуци:

vhod.dll

Форманти и шумове	
1 1.wav N	32 32.wav N
2 2.wav T 5 0.4	33 33.wav T 2 0.3
3 3.wav N	34 34.wav N
4 4.wav N	35 35.wav T 3 0.4
5 5.wav T 7 0.45	36 36.wav T 4 0.4
6 6.wav N	37 37.wav T 6 0.4
7 7.wav T 9 0.3	38 38.wav T 6 0.4
8 8.wav T 6 0.4	39 39.wav T 6 0.4
10 10.wav T 6 0.4	40 40.wav T 4 0.4
11 11.wav T 4 0.9	41 41.wav N
12 12.wav N	42 42.wav N
13 13.wav N	43 43.wav N
14 14.wav N	44 44.wav N
15 15.wav N	45 45.wav N
	46 46.wav N

16 16.wav T 4 0.4	47 47.wav N
17 17.wav T 6 0.4	48 48.wav N
18 18.wav N	49 49.wav N
19 19.wav T 2 0.3	50 5.wav T 2 0.3
20 20.wav N	51 51.wav T 6 0.3
21 21.wav T 5 0.4	52 52.wav T 6 0.3
22 22.wav N	53 53.wav T 6 0.3
23 23.wav N	54 54.wav T 6 0.3
24 24.wav N	55 55.wav T 6 0.3
25 25.wav T 3 0.4	56 56.wav T 6 0.3
26 26.wav N	58 5.wav T 2 0.2
27 27.wav N	60 41.wav T 1 0.1
28 28.wav T 5 0.3	61 61.wav T 1 0.1
29 29.wav N	62 58.wav T 4 0.3
30 30.wav N	63 59.wav T 4 0.3
31 31.wav N	64 60.wav T 4 0.3
32 32.wav N	65 62.wav T 4 0.3
33 33.wav T 2 0.3	***(Емотикони)***
34 34.wav N	66 66.wav N – *хахаха :-D*
35 35.wav T 3 0.4	67 67.wav N – *муахаха :-] >-]*
36 36.wav T 4 0.4	68 68.wav N – *хихихи ;-)*
37 37.wav T 6 0.4	69 69.wav N – *ммм :- (*
38 38.wav T 6 0.4	70 70.wav N – *хаааа :-o*
	-1

Табл. 3 – Формат на дефиниране на фонеми и междуметия

На практика около 40 записа на форманти и шумове достатъчни, за да се поражда разбираема реч от произволни текстове на български език. В настоящата версия на „Глас” има много експериментални варианти на форманти, които са използвани по време на разработката, но след това не са изтрити. Например първият прототип, който не използва ударения и редукции, работеше с 38 записа на форманти и шумове.

В „Глас 2” са включени още звуци за озвучаване на емотикони, които биха били полезни за озвучаване на електронна поща. Междуметията за емотикони са класифицирани като „шум”, защото с тях се работи със слепване. Разпознават се различни варианти на изписване на емотиконите:

Хахаха - :-) :) ^ ^

Муахахаха :-] :]

Хихихи ;-)

Придихание при голяма изненада :-o :o

Смущение, въпросителна интонация :-S :S

и др.

Възможно е потребителят сам да запише произволни звуци и да дефинира съответствие графема-звук. След това синтезаторът ще възпроизведе съответния звук при среща на дадената графема.

6.4. Фонеми

Фонемите са дефинирани в друг конфигурационен файл, като низове от «команди на говорния апарат», които определят микрофонеми, шумове или междуметия, които да бъдат синтезирани.

Знак на фонема	Брой команди	Команди
а	1	37
б	1	24
в	1	19
г	2	43 26
д	2	14 27
..	..	..

Табл. 4 – формат на конфиурационния файл, дефиниращ фонемите

Чрез първият низ се задава обръщение към съответните команди към синтезатора. Това позволява по-голяма гъвкавост и лекота при експериментиране с нови фонеми.

fonemi0.dll

а 1 37	п 1 22	е2 1 52
б 1 24	р 2 12 33	ю2 2 50 55
в 1 19	с 1 3	я2 2 50 51
г 2 43 26	т 1 6	к2 1 29
д 2 14 27	у 1 8	ъ2 1 56
д2 2 14 28	ф 1 47	и2 1 54
е 1 38	х 1 15	о2 1 53
ж 1 35	ц 1 45	у2 1 55
з 1 34	ч 1 41	ц2 2 31 45
и 1 5	ш 1 13	ф2 1 44
й 1 58	щ 2 13 6	ч2 1 42
к 1 23	ъ 1 21	п2 1 30
л 1 40	ь 1 50	ш2 1 1
м 1 17	ю 2 50 8	с2 1 3
н 1 36	я 2 50 37	щ2 2 13 4
о 1 39	дж 1 46	:) 1 66
п 1 22	т2 1 4	:-] 1 67
р 2 12 33	л2 1 16	;-) 1 68

с 1 3	Л 1 16	(...)
т 1 6	а2 1 51	END
	а3 1 62	
	а4 1 63	
	ъ3 1 62	
	ъ4 1 63	
	о3 1 64	
	о4 1 65	
	у3 1 64	

Табл. 5 – Дефиниране на фонеми

6.4.1. Ударени гласни

а2, е2, и2, о2, у2, ъ2; ю2, я2

ю2 == й + у2

я2 == й + а2

6.4.2. Дълги беззвучни съгласни

к2, с2, т2, ф2, х2, ц2, ч2, ш2, щ2, п2 са "дълги" беззвучни съгласни - те звучат тогава, когато говорният апарат спира да работи, преди пауза - в края на изкуствено ограничен изказ или в края на изречение, или когато са следвани от други беззвучни съгласни.

6.4.3. Редуцирани гласни

Форманти за редуцирани гласни. Правилата за редукция са по работата за СЛОГ (Тотков и др., 2003):

а3 - "а" на позиция втора пред ударението; и първа след ударението

а4 - "а" на позиция първа пред ударението

ъ3 - "ъ" на позиция втора пред, и първа след ударението (съвпада с а3)

ъ4 - "ъ" на позиция първа пред ударението (съвп. с а4)

о3 - "о" на позиция втора пред, и първа след ударението

о4 - "о" на позиция първа пред ударението
у3 - "о" на позиция първа пред ударението (съвпада с о3)

Входният текст се приравнява към малки букви при предварителната обработка. Мекият звук "л" е зададен два пъти: веднъж като "л2" и веднъж с главна буква. С „л2" във входния текст се задава задължително меко "л". Главно „Л" се слага служебно след обработка на текста там, където е необхоимо да има меко Л според фонетичните правила на българския език.

6.4.4. Междуметия

Разпознават се някои емотикони и междуметия (хахаха, муахахаха, :-), :-о и др.), за които са подготвени подходящи записи.

6.5. Синтез на звук

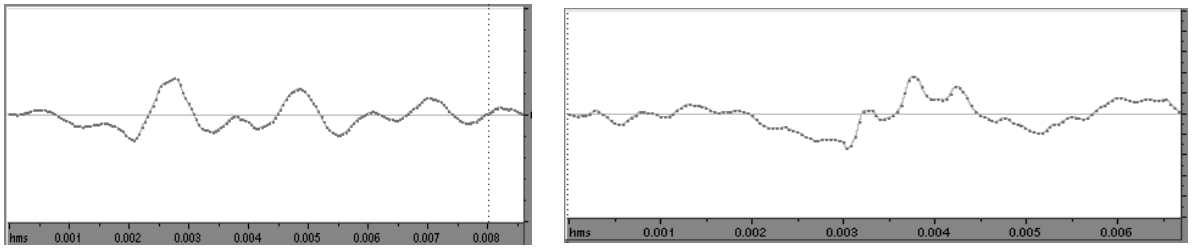
6.5.1. Беззвучни съгласни и междуметия

Звуците за беззвучни съгласни и за озвучаване на емотикони са цели записи, извлечени от реална реч. Те просто се копират в изходния запис.

6.5.2. Тонални фонеми - гласни, носови съгласни и звучни съгласни

Тонални звукове в "Глас" са фонемите за гласни (всички форми), за носови съгласни ("м" и "н") и за някои звучни съгласни (б, в). В първата версия на синтезатора, те се генерират като техният формант, или „микрофонема", се слепва многократно един след друг, чрез което се симулира периодично трептение.

При извличането на микрофонемите зрително се разпознават устойчиви периоди на почтипериодични функции от записи на думи.



Фиг. 9 - Микрофонема за„Е” (вляво, 6.7 ms) и „О” (вдясно – 8.6 ms)

6.5.3. Съставни фонеми

Съставни фонеми са „г”, „д” и „р”. При синтеза „г” и „д” се съставят от два шумови фрагмента. ”Р” е особен с това, че започва с шумова съставка, но продължава с тонална, която може да бъде удължавана.

6.5.4. Моделиране на динамиката на завука на гласните струни

В рамките на отделните фонеми може да се променя амплитудата на синтезираните тонални звуци, така че да се симулира увеличаването и намаляването на енергията на звука, който се генерира от гласните струни.

6.5.5. Преходи

Както беше споменато в 5. глава, преход представлява период от запис на реч, през който честотата на основния тон е в процес на бърза трансформация. „Глас 2” генерира преходи между два тонални звука.

6.5.5.1. Синтез на преходи

Преходът в „Глас” е реализиран като трансформация на една периодична функция в друга. При трансформацията се генерира трета периодична функция, която притежава междинни амплитудно-честотни характеристики. Преходът представлява поредица от амплитудно-честотно интерполирани периоди, при които коефициентът на участие на първата функция намалява от 1 до 0, и съответно участието на втората се повишава от 0 до 1; т.е. в началото на прехода в синтезираният звук участва период на първата функция, а в края на прехода – период на втората.

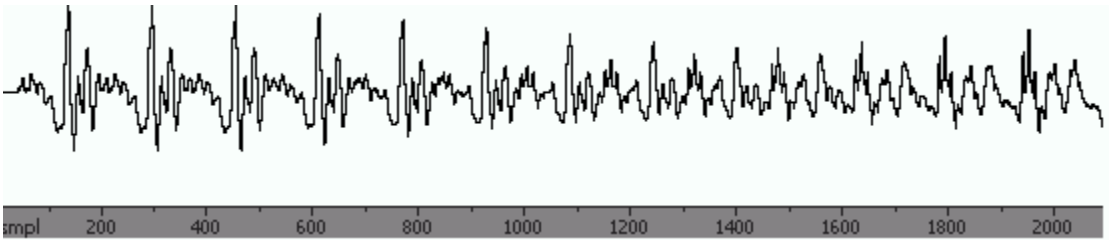
Алгоритъмът е следният:

1. Задават се две периодични функции (като записи на един период от тях), и броя периоди за извършване на трансформация на първата във втората.

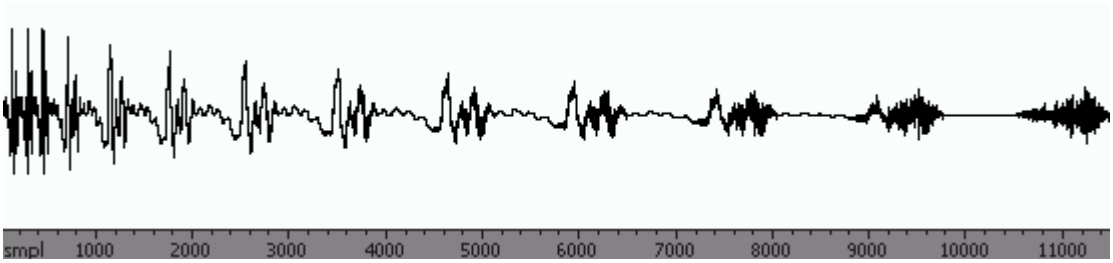
2. На всяка стъпка се изчисляват "разтегнати" или "свити" форми на двата форманта с еднаква, междинна основна честота. По този начин спектърът на по-нискочестотната функция се повдига, а на по- високочестотната - се понижава.

3. Така получените трансформирани форманти с еднаква междинна честота се интерполират помежду си, като коефициентът на влияние на първия формант намалява на всяка стъпка, а коефициентът на втория – се увеличава.

Чрез този метод се постига плавно спектрално свързване. А когато преходът нарочно се настрои да се извършва за много голям брой стъпки, се получава интересен звуков ефект и говорът звучи с особени извивки.



Фиг. 10 – Нереален, силно удължен синтезиран преход между ударено "а" и ударено "е"; в случая двете функции имат еднакъв период - 158 отчета.

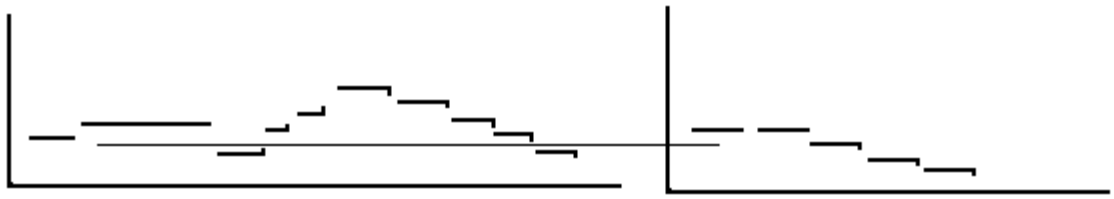


Фиг. 11 – невъзможен за изговаряне “преход” между микрофонема на ударено "а" (период 158 отчета,) и запис на "ч" с дължина 1724 отчета, зададен като периодична функция. Този преход звучи като забавящ скоростта си парен двигател.

6.6. Промяна на височината и интонационни контури

Промяната на основната височина на синтезирания глас се осъществява чрез трансформации, подобни на ефектите за промяна на височината на звука на произволен запис, описани в глава 5. Преди синтеза формантът се “разтяга” или „свива” така че да се побере в по-дълго време (понижаване на височината) или по-късо време (повишаване на височината). За разлика от методите за работа с произволни записи обаче, при синтеза се работи с предварително обособени микрофонеми, което позволява спектралните промени да се извършват с по-високо качество.

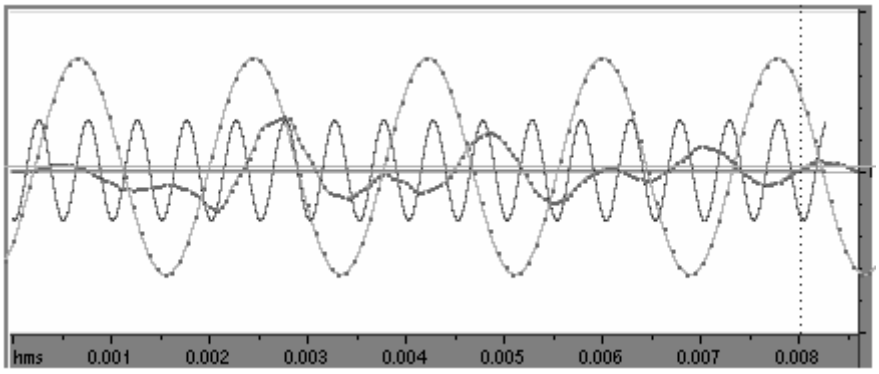
За да се симулира интонация, височината при синтеза допълнително се изменя с коефициентите, предварително изчислени при маркирането на интонационните контури. Използвани са контури по (Бояджиев и Тилков, 1999).



Фиг. 12 - Интонационни контури за съобщителни изречения.

6.7. Промяна на тембъра и темброви контури

Промяна на тембъра се извършва като се манпулират микрофонемите. Синтезират се сигнали с форма, честота, фаза и амплитуда, които могат да се настройват (синусоидални, триъгълни, трионообразни, експоненциални, правоъгълни, дефинирани от потребителя...). Синтезираните сигнали модулират микрофонемите и се получава тон с променена формантна структура.



Фиг. 13 – Промяна на тембъра чрез модулиране със синусоидални трептения

6.8. Пеене

Синтезаторът може да работи в експериментален режим на пеене, при който входният текст представлява специален нотопис, подобен на нотописа за звънене на някои стари модели мобилни телефони. Предварително се настройва продължителността на един такт. Текстът на песента се записва на срички, разделени с тире. На същия ред следват нотите и продължителността им в тактове. Текстът на един един ред се изпява на един "дъх". В края на нотописа се отбелязва броят тактове пауза, които се оставят между два реда (възможно е да са нула).

Нотите се записват като *буква*[#]*цифра*. Буквата означава нотата: C, D, E, F, A, B. Това съответства на до, ре, ми, фа, сол, ла, си. Цифрата показва номера на октавата. С *@N* се означава пауза в брой тактове, като N е цифра.

Ми-ла мо-я ма-мо | B4 1 C5 1 B4 1 C5 1 D5 2 D5 2 @1

При синтеза, продължителността на съответните микрофонеми се променя, така че да съвпада с честотата на съответната нота и се добавят хармонични трептения с честоти кратни на основната и с по-ниска амплитуда, като може да се настройва формата им, съотношението на амплитудите и коефициентът на вибратото, който кара честотата на синтезирания звук да се люлее около честотата на тонът, който се пее в момента.

Продължителността на звучене се синхронизира спрямо записаните тактове.

Могат да се включват в действие и модулациите на тембъра и специалните електронни ефекти.

6.9. Електронни ефекти

При обикновена реч или при пеене може да се включва режим на "електронни ефекти", при който се извършват допълнителни манипулации върху микрофонемите. Тези ефекти са създадени с експериментална цел като основа за бъдещи разработки на музикален синтез.

1. **Намаляване на честотната разделителна способност и преобразуване на нелинейните преходи в линейни** – всяка микрофонема се разделя на определен

брой ключови точки (2, 5, 10, ...), които обособяват *микропрозорци* с равна или с различна дължина, която е предварително дефинирана. Амплитудата на отчета на мястото на ключовата точка се определя по два начина:

- [x]*A + [x-1]*B + [x+1],*C където [x] е стойност на отчета върху ключовата точка, а [x-1] и [x+1] са съответно с една стъпка назад и напред, а параметрите A, B, C са тегла, за които трябва да се запази съотношението: A + B + C = 1. Значението на разпределението на теглата зависи от броя на ключовите точки, и от честотата на сигнала в микрофонемата. Ако ключовите точки са малко, и B и C имат малки тегла, графиката може много да се изкриви от оригиналната, но това може да е желан резултат.

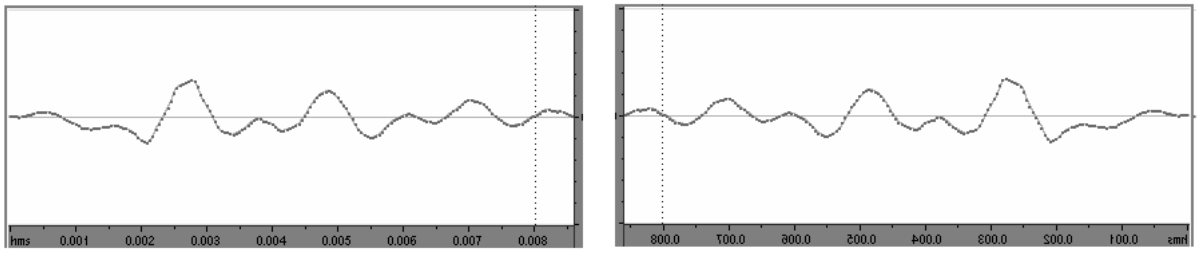
2. **Положителна/отрицателна средна амплитуда** - в рамките на всеки микропрозорец се изчисляват две стойности - средна амплитуда на отчетите положителните и отрицателни стойности.

   2.1. По-голямата абсолютна стойност се взима за амплитуда на ключовата точка.

   2.2. По-голямата абсолютна стойност се поставя като стойност на всички отчети в микропрозореца.

3. **Максимална абсолютна стойност в рамките на микропрозореца** – на мястото на ключовия отчет се поставя отчетът с най-голяма абсолютна стойност (с положителен или отрицателен знак).

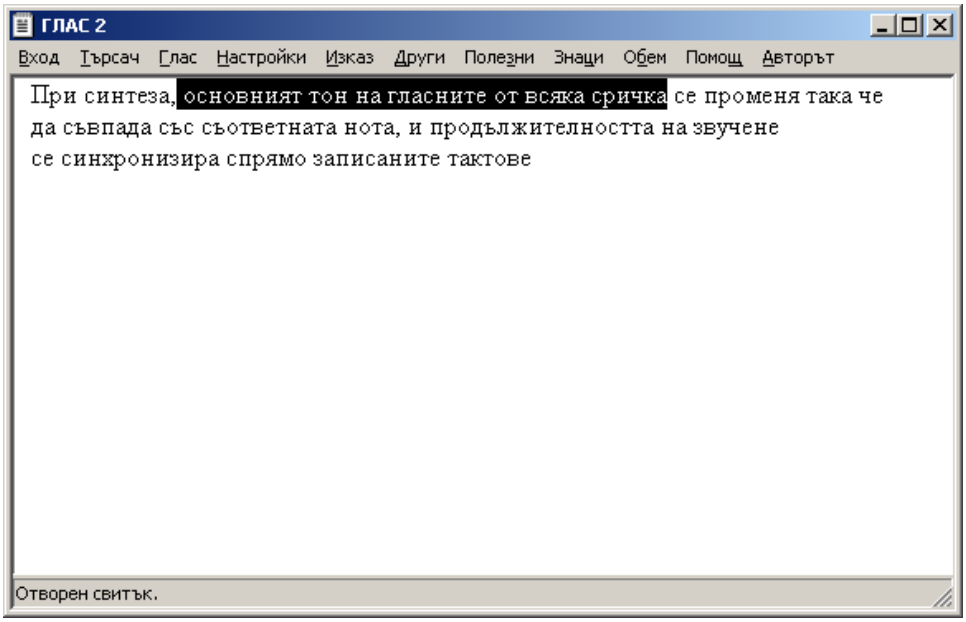
4. **Огледална микрофонема**



Фиг. 14 – Обръщане на микрофонема във времето

6.10. Ръководство за потребителя

6.10.1. Графичен интерфейс



Фиг. 15 – Графичен интерфейс на „Глас 2”

Графичният интерфейс на „Глас” е като на прост текстов редактор от типа на Notepad.

6.10.2. Клавишни комбинации

Важат стандартните комбинации за редактиране:

Ctrl+V – paste (лепи)

Ctrl+C – copy (запомни)

Ctrl+X – cut (режи)

Ctrl+Z – undo (отмени)

Има някои разлики в клавишните комбинации:

Ctrl+Q -- Търсач->Опростен – търсене на низ (пълно съвпадение)

Ctrl+W -- Търсач->Пак_търси – ново търсене

Ctrl+S -- Вход->Запиши – запис на файл

Ctrl+1 -- Вход->Изход – изход от програмата

Ctrl+N – изчистване на текста

Ctrl+O – диалог за отваряне на нов файл

Ctrl+P – презареждане на последния отворен файл

Ctrl+S - минимизиране

Ctrl+G – започва да изговаря маркирания в прозореца текст, или целият текст, ако няма маркирани пасажи

Ctrl+H – прекъсва започнатият говор след края на текущия изказ

Разработени са диалози с надписи на български, в които изписването на текста е анимирано.

6.10.3. Настройки на параметрите на синтезатора.

През съответните подменюта от менюто «Настройки» и „Глас” могат динамично да се променят:

- Бързината на говора.
- Височината на звука.
- Наличието и вида на звуковите ефекти и ефектите за промяна на тембъра.
- Включване/изключване на интонационни и темброви контури.
- Настройка на коефициент на удължаване на ударението.
- Включване на режим на пеене.

Параметрите на синтезатора могат да се задават и в конфигурационен файл – *osnova.txt*, - който може да се зарежда автоматично при стартиране, или да бъде избран и зареден по време на работа.

6.10.4. Работа с различни кодови таблици

„Глас 2” може да работи с текстови файлове в кодировка МИК (кирилица под ДОС), UTF-8 и Windows-1251, до която се преобразуват текстовете на първите две кодировки.

7. Глас 3 - направления за изследвания и разработка

В тази глава са представени идеи и насоки за изследвания и разработка за създаване на подобрена версия на синтезатора с настоящата архитектура, както и да се създаде по-развит, квазиартикулационен синтезатор.

7.1. Подобрения на нормализацията

Ще се състои в обогатяване на речника със съкращения с малка многозначност и думи на английски. Експерименти с корпусни методи за разпознаване на съкращения.

7.2. Подобрения на прозодията

Те ще включват работа върху синтеза на емоционално натоварена реч и по-прецизна прозодия, включваща по-точен контрол върху продължителността на ударените и неударените гласни при изразяване на различни емоции. Ще съдържа методи за разпознаване на идентичности (Named Entity Resolution) и развитие и реализация на идеите за темброви контури, предложени в дипломната работа.

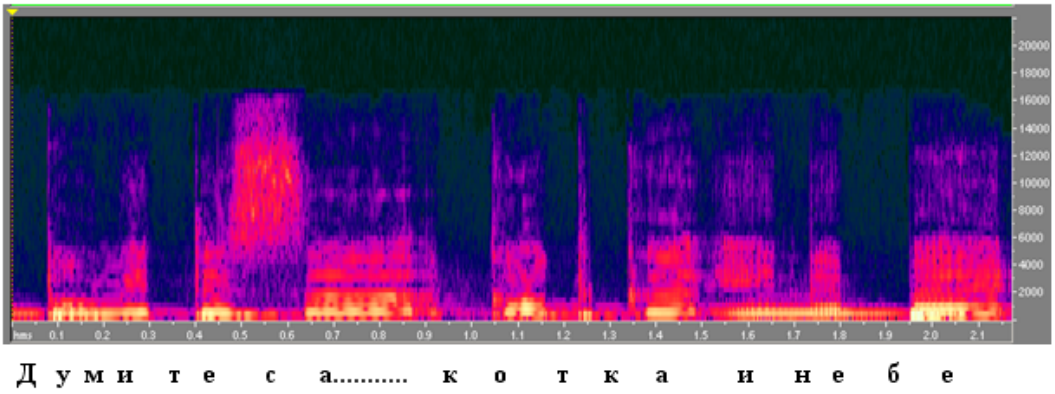
Ще бъдат необходими изследвания върху средства за автоматично разпознаване на емоциите и по-дълбок анализ за откриване на интонационни контури.

Интерес представлява синтезът на реч, по която да може да се разпознае дали говорителят се усмихва, намръщен е, уморен е и др.

7.3. Звукови ефекти

Включване на ефекти като реверберация, закъснителна линия, дисторжън и филтриране на избрани честоти.

7.4. Среда за маркиране на корпуси с реч



Фиг. 16 – Маркиране на съответствието фонема – графема

Разработка на графична среда, в която да се отбелязва изображението графеми-фонеми, ударения, прозодия, логически ударения, говорител (при записи на диалози и мултилози) и др. Такава среда би подпомогнала автоматичното извличане на форманти и преходи, създаване на схеми за промяна на тембъра (от съгласувани корпуси, в които двама или повече говорители говорят синхронно); разпознаване на гласове; обучение на адаптивни квазиартикулационни синтезатори и др.

7.4. Среда за интерактивно моделиране на звуци

Би съдържала развит графичен интерфейс подпомагащ:

- извличането на микрофонеми
- „рисуването” и графично редактирането на нови микрофонеми
- редактиране на спектри на шум
- създаването на модели на преходи между микрофонеми, подпомогнато от методи за автоматично и полуавтоматично извличане на форманти
- моделирането на параметри за формантен синтез
- моделиране на инерционни трептящи системи за синтез на музикални звуци
- „рисуване” на параметрите на гласа по време на пеене

7.5. Адаптивен квазиартикулационен синтез и анализ с обратна връзка

Нормално чуващите деца започват да възпроизвеждат звуците на речта чрез многобройни опити да ги налучкат чрез своя говорен апарат - процесът преминава през търсене и откриване на зависимостта между двигателните команди, звуците които те възпроизвеждат, и звуците които са били чути и трябва да бъдат повторени с някаква степен на подобие. Освен това, проговарящото дете получава обратна връзка за успеха си чрез поощренията и реакциите на околните.

Подобна би била идеята на квазиартикулационния адаптивен синтезатор, който би бил едно от продълженията на изследванията. Система би съдържала модел на говорен апарат, който ще се състои от трептящи системи и резонатори, но няма да е задължително моделът на говорния апарат, начинът на действие и законите за взаимодействие между компонентите му да бъдат точна и пълна физическа симулация на човешки говорен апарат. Също така, изображението фонеми-команди на говорния апарат няма да бъде дефинирано в детайли, а ще трябва да бъде извлечено от маркирани и/или немаркирани корпуси с реч, музика и чисти звуци, и от диалог с учител, който също ще може да произвежда и звуци от музикални инструменти.

В началото системата ще започва обхождане на пространството от възможни команди на говорния апарат като ще бъде грубо насочена да търси подобия между синтезирания звук и чутите звуци. За изчисляване на степента на подобие между чутото и изговореното ще бъдат използвани методи от разпознаването на реч. Би било възможно учителят ръчно да донастройва някои параметри и да „подсказва” на системата как да движи говорния си апарат, като се използва графична среда и развити входни устройства, напр. графични таблети с които да се редактират формата, честотите, динамиката и др. параметри на говорния апарат.

Успешен синтез с обратна връзка може би би имал и приложение в разпознаването на реч. Ако системата успее да се настрои да синтезира звуци с параметри близки до говорни, то при чуване на звука, тя би могла да търси възможна комбинация от команди на говорния си апарат, която би породила съответните звуци. Така може да се предположи възможен текст, който е породил звука.

Разбира се, тези идеи имат нужда от много допълнителни експерименти, изследвания и доуточнения, за да придобият убедителна форма.

7.6. Разпознаване на наличие на говор и опити за автоматично разпознаване на говорителя

- Извличане на реплките на участници в диалог или мултилог и записване отделно и/или маркиране.
- Търсене на изказванията на даден говорител в голям запис или корпус.

7.7. Преобразуване на гласове и синтез на емоционално модулирана реч

Това направление би включвало опити за манипулиране на запис на реч на един говорител да звучат с тембъра, скоростта или с други особености на гласа на друг говорител, и експерименти за синтез на реч емоции, чрез контролиране на темпото, динамиката и интонацията на речта.

7.8. Пресъздаване на гласове

Това направление е свързана със средите за графично моделиране на форманти и интонационни контури. Две примерни задачи в тази насока са:

- **Пресъздаване на гласове по памет** - потребителят иска да пресъздаде звученето на глас на някого, на когото няма наличен запис, сравнявайки синтезиран звук със спомените си. За целта чрез средата за моделиране се манипулират формантите, интонационните контури, динамиката на скоростта на говор и др. и се търсят параметрите на синтезирания глас, които да съвпаднат с това, което си спомня. За улеснение е възможно предварително да се създаде голяма библиотека от параметризирани тембри на гласове.

Друга идея в тази насока е създаването на библиотека с много голям брой записи на различни говорители, като могат да бъдат отбелязани общи характеристики – пол, възраст, пушач/непушач, липсващи зъби и др. особености. Потребителят може да ги прослушва записите и да търси глас, който прилича на гласа, който си спомня.

- **Пресъздаване на гласов модел от непълни данни** – системата получава кратки записи на реч, в които липсват част от звуковете. Липсващото се възстановява чрез предсказване на липсващите форманти въз основа на наличните и на зависимости, открити в други гласове, за които има пълен модел.

8. Заключение

В тази дипломна работа беше представен хибридният синтезатор на реч от текст „Глас 2”, който подобрява синтезатора на реч „Глас”, разработен от автора на дипломната работа, и включва следните нови възможности:

- Правилно разпознаване и озвучаване на ударени и редуцирани гласни.
- Амплитудна динамика на звука по време на синтеза.
- Интонационни контури.
- Пеене.
- Настройки на тембъра.
- Специални звукови ефекти върху микрофонемите.
- Темброви контури.
- Озвучаване на емотикони.

В работата още бяха дадени насоки за изследвания и разработки за автоматично извличане на форманти, които да помогнат за ускоряване на процеса на създаване на нови гласови модели за синтезатора, и бяха представени методи за промяна на продължителността и височината на записи на звук.

Дипломната работа завърши с редица идеи и направления за изследвания и разработки. Някои от тях биха подпомогнали усъвършенстването на настоящия синтезатор и разширение на музикалните му възможности, а други са насочени към създаването на нов, квазиартикулационен синтезатор на реч

9. Библиография

БАКЛ 2005 – SpeechLab 2.0 - http://www.bacl.org/speechlabbg.html

Бояджиев Т., Тилков, Д. – Фонетика на българския книжовен език, 1999

Иванов, И. – Reader - http://tts.data.bg/

Сп. „Компютър за Вас" бр. 7-8/1990, с. 23

Стойчева, С. - Ролята на Националния център за рехабилитация на слепи за квалификацията на учителите по компютърна и брайлова грамотност.

Тотков Г., Д. Благоев, В. Ангелова, *За озвучаването на компютърен български текст,* Национална научна конференция „10 години катедра *Компютърни системи* към ТУ – София, филиал Пловдив", ноември 2003, 119-131.

Banks, Kevin (2002) – "The Goertzel Algorithm", Embedded Systems Design, http://www.embedded.com

Boersma P., Weenink D. – Praat – doing phonetics by computer - http://www.fon.hum.uva.nl/praat/

Burkhardt, F. - Examples of synthesized emotional speech, http://emosamples.syntheticspeech.de/ - 2008

Kenochi, H.; Oshita, H. - Vocaloid – "VOCALOID – Commercial singing synthesizer based on sample concatenation", Yamaha Corporation, Japan

Lemmetty, S. (1999) - "Review of Speech Synthesis Technology" - Helsinki University of Technology, Department of Electrical and Communications Engineering, 1999.

LYRICOS Project - http://www.lyricos.org/

Medscape Today - Hearing Loss: Does Gender Play a Role? - http://www.medscape.com/viewarticle/408872_5

Microsoft Research - Whistler - http://research.microsoft.com/srg/whistmusic/

Sirin, Simon (2004) - "A DSP algorithm for frequency analysis", Embedded Systems Design, http://www.embedded.com

Txt2Spc, Demo 1.12 - http://bezmonitor.info/articles/txt2spc.htm