

## Изследвания по синтез на българска реч на базата на срички.

инж. Юлиан Христов Великов, ТУ-Варна  
доц. д-р. Йордан Николов Колев, ТУ-Варна

### Abstract

The speech synthesis problem becomes more and more important for machine-human interaction, particularly for helping blind people to use computers and modern informational systems like Internet. The speech synthesis is a process of converting written text, usually available as ASCII strings, to natural speech. The paper presents some results of the author's study of syllables based speech synthesis for the Bulgarian language.

The syllables based speech synthesis is chosen as a compromise between the quality of the synthesized speech and the memory required, as compared to other approaches, those based on phonemes, diphones, words or expressions.

The study is based on a created by the authors data base which contains description of the syllables in the Bulgarian language. The data base is established by processing large amount of different texts using a specially developed algorithm for converting texts to phonetic description following the 'write as you say' rule. After that the text is divided into syllables.

The authors have investigated experimentally the influence of syllable location within the word on some syllable parameters. The results arrived at are used as an element of the syllable descriptor in the experimental data base. The other elements of the syllable descriptor are the phonetic description and the digitized syllable. The established experimental data base is used for speech synthesis and is a versatile tool for future experiments.

**Keywords:** text-to-speech. syllable synthesis, speech synthesis.

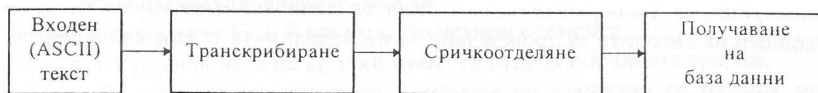
### Увод

В настоящия доклад се разглеждат изследванията по синтез на българска реч. Проблемът за синтез на реч е актуален от гледна точка улесняване комуникацията машина - човек. И в частност за улесняване на хората с нарушено зрение при работата им с информационните системи. Синтезът е процес на преобразуване постъпващият текст под формата на 'ASCII' символи до естествена реч.

Обемът на съхраняваната информация и качеството на синтеза зависи от това чрез какви елементи на речта ще се синтезира: фонемите, дипфони, срички, думи или изрази. Синтезът посредством срички е един компромис между качеството на синтезираната реч и обема на използваната памет.

Изследванията проведени от авторите се базират на създадената от тях база данни съдържаща текстовото описание на сричките в българския език. За получаването на необходимата за синтеза база е разработен алгоритъм, чрез който входният текст се преобразува до транскрибиран [1], т.е. записва се според принципа "пиши както говориш", след което се разделя на срички. На фигура 1 е показана блоковата схема описваща алгоритъма за получаването на базата съдържаща срички.

## Получаване на сричкова база



фиг.1 Създаване на сричкова база данни

Началният етап е прочитане на постъпващият правописен текст под формата на ASCII символи и разделянето му на думи. Следващата стъпка е преобразуването на постъпилия правописен текст в правоговорен. Този процес на преобразуване се нарича транскрибиране. Правоговорния текст получен след транскрибирането се разделя на срички. Отделните срички получени след сричкоразделянето [2] образуват сричковата база използвана при синтеза. По направени от авторите изследвания [3] сричките в българския език са около 3500.

*Сали беше се издигнал над всички по божа дарба*

*сали беше се издигнал нат фсички по божа дарба*

*са-ли бе-ше се из-ди-гнал нат фсич-ки по бо-жа дар-ба*

ба	бе	бо	гнал	дар	ди	жа	из	ки	ли	нат	по	са	се	фсич	ше
----	----	----	------	-----	----	----	----	----	----	-----	----	----	----	------	----

фиг. 2 Пример за създаване на сричкова база

С помощта на посоченият на фиг.2 пример се онагледяват отделните етапи при получаването на сричковата база. Първото изречение *‘Сали беше се издигнал над всички по божа дарба’* представлява постъпващият правописен текст. Второто изречение се получава след процеса транскрибиране. То е правоговорния еквивалент на първото изречение. Нека се обърне внимание на думите *‘над’* и *‘всички’*, вижда се че в транскрибираното изречение те са описани както се изговарят, а не както се пишат. Преобразуването на постъпващия текст до правоговорен се налага за да може синтезираната реч да бъде по близка до естествената. Това е така понеже получената база за синтез е от срички получени от правоговорен текст, а не от правописен. Получената база както се вижда от примера съдържа текстовото описание на сричките.

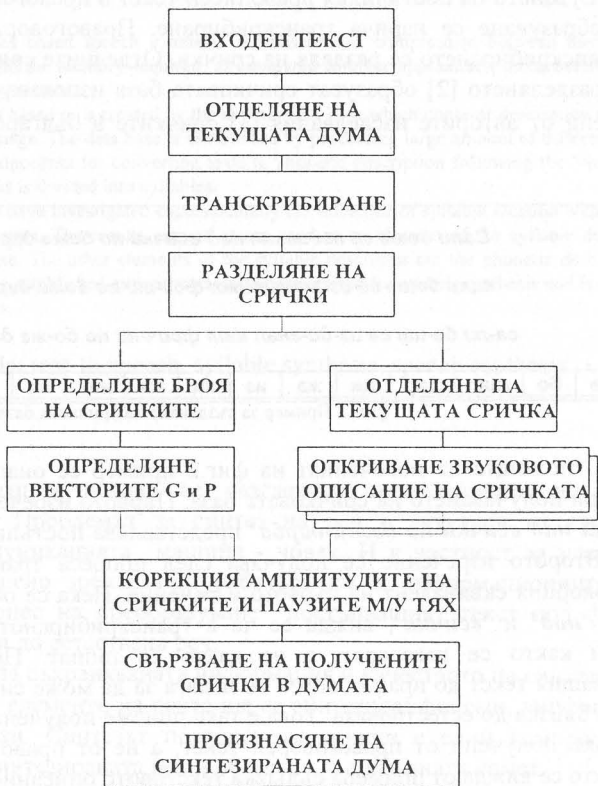
## Създаване на звукова база

Следващият етап е записването на звуковия еквивалент на получените срички в звукова база. Звуковата база може да съдържа както параметри описващи звуковата информация на отделните срички, така може да съдържа и

времевите отчети на сигнала. За създаването на експерименталната база авторите работиха с времевите отчети на сигнала, като всички записани срички са нормирани. В експерименталната база са записани минимума от срички необходими на авторите за провеждане на изследванията за синтез на реч.

### Синтез на реч посредством сричкова база

Посоченият на фиг. 3 алгоритъм за получаване на синтезирана реч посредством предварително записана звукова база е следният:



фиг.3 Синтез на реч посредством сричкова база

1. Постъпващият на входа текст се разделя на думи, като по нататък се работи с отделната дума.
2. Отделената дума се транскрибира и разделя на срички.
3. Определя се броя на сричките в думата.

4. В зависимост от броя на сричките се определят векторите  $G$  и  $P$ . Във вектора  $G$  се записват стойностите на амплитудите на отделните срички, а във вектора  $P$  се записват паузите между тях. Ако думата се състои от четири срички, то векторите  $G$  и  $P$  ще имат по четири елемента.

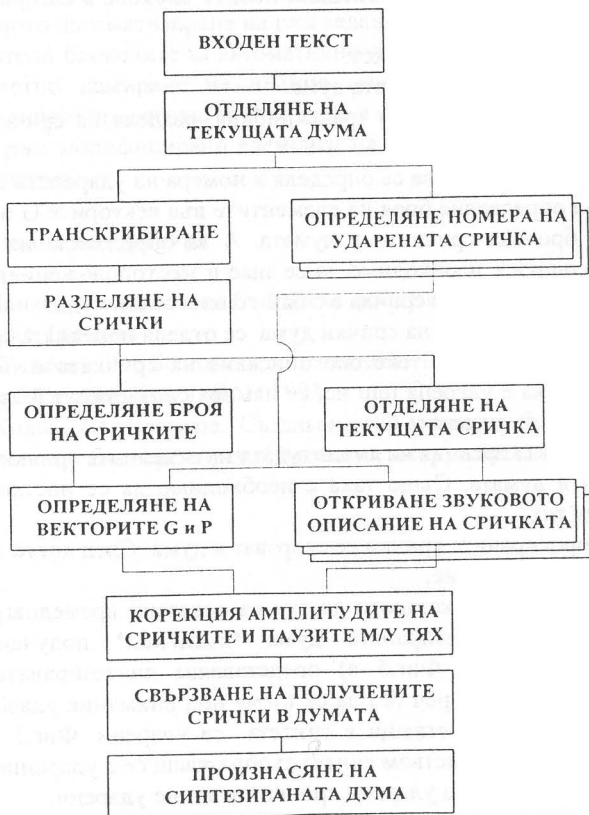
5. От разделената на срички дума се отделя поредната сричка.

6. По полученото текстово описание на сричката се намира нейният звуков еквивалент стоящ във вече създадената звукова база.

7. Понеже сричките стоящи в базата са нормирани е необходима корекция на амплитудата на отделните срички в зависимост от мястото им в думата. Също така е необходимо да се постави необходимата пауза между тях.

8. След направената корекция на отделните срички те се свързват в дума. Посредством тази корекция се получава динамиката на получената реч.

9. Последният етап от алгоритъма е произнасянето на синтезираната дума и свързването и в изречение.



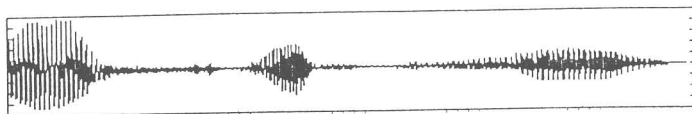
фиг. 4 Синтез на реч посредством сричкова база отчитайки ударението в думата

Получената посредством алгоритъма на фиг.3 реч е разбираема, но неестествена. Това се дължи на факта, че в посочения алгоритъм не се взема под внимание ударението в думата. Вземайки в предвид факта, че в българският език ударението е подвижно и не съществуват правила за неговото локализиране, авторите разработиха модифициран алгоритъм фиг.4 за синтез съобразяващ се с ударението в думата.

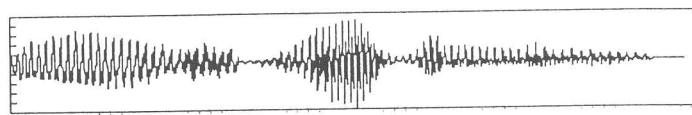
Този алгоритъм наподобява предишния, но в него е добавен блок определящ номера на ударената сричка. Определяне местоположението на ударената сричка в думата може да се извърши по два начина: чрез създаване на база съдържаща думите в българския език и указване позицията на ударената сричка или чрез указването на позицията директно в постъпващият текст. За нуждите на експеримента авторите използват втория начин. Новото в този алгоритъм е че в звуковата база се пазят както звуковите еквиваленти на ударените срички така и на неударените, т.е. на една сричка отговарят две описания. След като бяха разгледани новите блокове в алгоритъма принципът му на действие е :

1. Постъпва входен текст.
2. Отделяне на текущата дума.
3. Отделената дума се транскрибира, разделя на срички и се определя броят им.
4. Успоредно с това се определя и номера на ударената сричка.
5. За определяне броя на елементите във векторите  $G$  и  $P$  е необходимо да се знае броя на сричките в думата. А за определяне на амплитудата на отделните срички е необходимо да се знае и местоположението на ударението в думата. Като ударената сричка е с най-голяма амплитуда.
6. От разделената на срички дума се отделя поредната сричка.
7. По полученото текстово описание на сричката и по информацията дали тази сричка е ударена или не, се намира съответният ѝ звук еквивалент стоящ във вече създадената база.
8. Следва корекция на амплитудата на отделните срички в зависимост от мястото им в думата. Също така е необходимо да се постави необходимата пауза между тях.
9. Коририраните срички се свързват в дума. След което получената дума се произнася в изречение.

В подкрепа на казаното до тук са показани времедиаграмите на фиг.5 представляващи синтезираната дума 'издигнал' получена посредством описаните алгоритми. Фиг.5 а) представлява синтезираната реч получена посредством синтез на реч без да се взема под внимание ударението в думата, т.е. всички срички участващи в синтеза са ударени. Фиг.5 б) представлява сигнала получен посредством синтез съобразяващ се с ударението в думата, т.е. в синтеза участват както ударени срички така и не ударени.



а) и з д и г н а л



б) и з д и г н а л

фиг.5 Синтезиран речеви сигнал на думата 'издигнал'.  
а) без отчитане на ударението в думата, б) с отчитане на ударението в думата.

### Заклучение

За пълното автоматизиране на разгледаните алгоритми е необходимо да бъдат разработени блоковете за автоматично откриване на ударената сричка в думата, съответно намиране на нейния номер. Друг проблем свързан с качеството на синтез предстои за решаване е създаване на такава звукова база която да съдържа модифицирани елементи на звуковата сричка, т.е. да бъде дикторo независима.

### Използвана литература

1. К. Цанков, Помагало по българска фонетика (транскрипция). Знак '94 Велико Търново
2. Д. Тилков, Т. Бояджиев. Граматика на съвременния български книжовен език. Том 1. БАН 1993.
3. Й. Колев, Ю. Великов. Създаване на сричкова база за нуждите на синтеза на българска реч. Статията е изпратена за публикуване в сп. Е+Е