

# **Метод за определяне периода на основния тон на вокализиран речеви сигнал**

доц. д-р инж. Йордан Николов Колев  
инж. Тодор Димитров Ганчев  
Драгомир Димитров Николов  
Технически университет - Варна

## **A Pitch Determination Method**

Jordan N. Kolev - Associated Professor,  
Todor D. Ganchev - Ph.D. student,  
Dragomir D. Nikolov - student

The pitch determination in the real time tasks such as low-bit-rate speech coding requires a considerable part of the processor time. This is the reason the problem to be still actual. The research performed by the authors shows that the precision of the pitch determination does not reflect substantially on the comprehension of the resynthesized speech, for the reason that quantization of the pitch is made.

The method is based on autocorrelation of ultra-short energy of the speech. As a result a substantial decrease of computation is achieved, having an error not exceeding 5% for male voices and 10% for female ones.

The investigations using different speakers and different SNRs have been done. The results demonstrate a good reliability. The estimation of amount of operations needed when using TMS320C50 DSP is presented in comparison to other widely used methods.

The method have been proposed in the frame of investigations on low-bit-rate speech coding performed in the Microprocessor Systems and DSP Laboratory in the Department of Electronics at the TU of Varna.

### **Анотация**

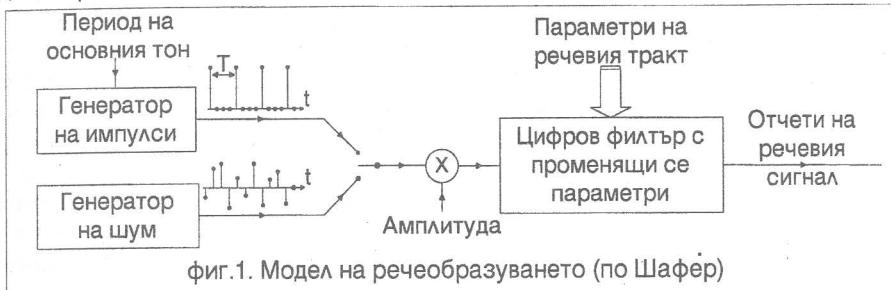
Като правило определянето в реално време на периода на основния тон (ПОТ) за вокализираните участъци от речта в системите за ниско-скростно предаване на реч е трудоемка задача, отнемаща значителна част от ресурсите на процесора. Това е една от причините търсенето на нови алгоритми за определяне на ПОТ да е актуално и в настоящия момент, въпреки постигнатите вече резултати в тази област.

Въз основа на направени изследвания за влиянието на точността на определяне на ПОТ при анализа върху качеството на синтезираната реч, авторите предлагат нов изчислително ефективен алгоритъм за определяне ПОТ.

В доклада се представят резултати от експерименталните тестове на предлагания метод и сравнение с други традиционни методи.

## I. Въведение в проблема

Съгласно линейния модел на речеобразуването (фиг.1.), предложен от Фант [2], речевият тракт на човека може да се представи като линейна система, на входа на която действа или поредица от единични импулси с период  $P$  (в случая на вокализиран сегмент) или бял шум за невокализираните сегменти реч. Периодът на повторение на импулсната поредица представлява ПОТ, който е основен параметър на речта.



фиг.1. Модел на речеобразуването (по Шафер)

В действителност ПОТ носи информация за честотата на трептене на гласните струни на човека. Това е честотата, с която импулсите въздушен поток постъпват в речевия тракт, което добре съответства на модела на Фант.

ПОТ е параметър, носещ информация както за индивидуалността на диктора, така и за емоционалното състояние, в което той се намира. То-ва обуславя значимостта му за естественото звучене на речта. Поради тази причина в системите за нискоскоростното предаване на реч ПОТ е един от параметрите, които задължително се предават по канала за връзка. Най-обща структурна схема на вокодерна система [2] е показана на фиг.2:



фиг.2. Структурна схема на вокодерна система

Където: ПО- Предварителна обработка; ОФФ- Определяне на филтрова-та функция; Т-Ш- Решение за типа на сегмента (вокализиран/невокали-зиран); ООТ- Определяне на периода на основния тон; ГОТ- Генератор на основния тон; ГШ- Генератор на бял шум; СФ- Синтезиращ филтър; К- Комутатор тон-шум; Ф- Филтър за подобряване звученето на речта.

След сегментация и анализ на речевия сигнал по канала за връзка се предават само параметрите на речта, по които в приемната страна речевият сигнал отново се възстановява. Освен ПОТ се предават типът на текущия сегмент (вокализиран или невокализиран) и параметрите на речевия тракт.

Във вокодерните системи, които не използват векторно квантуване на параметрите на речта, е прието за ПОТ по канала за връзка да се предават по шест бита за сегмент вокализирана реч. Авторите на настоящия доклад са извършили експерименти, потвърждаващи, че шест бита са достатъчни за постигане на добро качество на синтезираната реч. На практика точността на ПОТ, използван при синтеза на реч, зависи от броя битове, с които се квантува ПОТ при предаването по канала, а не от точността, с която се определя при анализа, дори и ако тя е много по-висока. От тук следва, че във вокодерните системи, при които анализът не се извършва синхронно с ПОТ, не е оправдано да се изразходва значително количество от времето на процесора за свръткочно определяне на ПОТ, ако след това се извършва квантуване и точността се понижава.

## II. Описание на предлагания метод за определяне ПОТ

Въз основа на направени изследвания за влиянието на точността на определяне на ПОТ, при анализа върху разбираемостта и качеството на синтезираната реч, и съобразявайки се с точността, постигана при квантуването, авторите предлагат изчислително ефективен метод за определяне ПОТ. Методът се състои от две стъпки: определяне типа на сегмента и определяне стойността на ПОТ.

### 2.1 Определяне типа на сегмента - вокализиран/невокализиран

По време на първата стъпка (фиг.3.) се определя типът на текущия сегмент от речевия сигнал - вокализиран или невокализиран.



- Това става по стандартна процедура, като се използват параметрите:
- брой преминавания на сигнала през нулата (изчислени с преварително централно ограничение на речевия сигнал)
  - енергията на сигнала в анализирания сегмент, изчислена като сума от свръхкратковременните енергии (СКЕ):

$$E = \sum_{n=0}^{\frac{N}{K}-1} E_{us}(n), \quad K = 1 \div 10; N = Fs \cdot 32ms \quad (1)$$

- коефициент на линейно предсказване A(1) (получен по процедурата на Левинсън)

Тук трябва да се отбележи, че само броят преминавания на сигнала през нулата се изчисляват специално за определяне типа на сегмента. Пълната енергия по принцип е необходима за съгласуване на енергията на синтезирания речеви сигнал с тази на входния. В предлагания от авторите алгоритъм пълната енергия се изчислява по модифицирана процедура, позволяваща едновременно получаване и на СКЕ на сигнала, без значително да се увеличи обема на изчисленията. СКЕ представя енергията на сигнала, изчислена последователно за неприпокриващи се групи от по K отчета от входния сигнал:

$$E_{us}(n) = \sum_{i=0}^{K-1} x(n.K + i)^2, \quad n = 0, 1, \dots, \frac{N}{K} - 1; K = 1 \div 10 \quad (2)$$

Когато размерът на групата K=1, имаме традиционния начин за определяне на енергията на сигнала. Интерес представляват стойностите за размер на групата K=2...10, при които се получава значително съкращаване на обема на изчисленията при допустимо ниво на грешката в ПОТ. Естествено стойностите за параметрите размер на групата K и броят отчети N се подбират в зависимост от честотата на дискретизация на речевия сигнал. Броят на отчетите N трябва да е такъв, че анализираният сегмент от речевия сигнал да бъде с продължителност, близка до 32ms (два периода на максимално възможния ПОТ  $MaxT \approx 16ms$ ), а размерът на групата отчети K, за които става определянето на СКЕ, се подбира така, че да се получи желаната разделителна способност за ПОТ.

Коефициентът на линейно предсказване A(1) се получава като страничен продукт от рекурсивната процедура на Левинсън за определяне коефициентите на отражение на речевия тракт и на практика не изиска допълнителни изчисления.

## 2.2 Определяне периода на основния тон

След определяне типа на сегмента се преминава към втората стъпка от алгоритъма: определяне ПОТ. Ако сегментът реч е невокализиран, ПОТ T се приема за равен на нула и не се изискват допълнителни изчисления. Ако обаче сегментът е вокализиран, се извършват следните операции (фиг.4.):



фиг.4. Стъпка 2 на алгоритъма:  
Изчисляване периода Т на основния тон

За отстраняване на грешки при оценката на ПОТ е въведено изглеждане, чрез използване на стойностите на периода за съседните два сегмента. Допълнително се правят проверки за удвояване или разполовяване на ПОТ. Чрез използване на параболична интерполяция (също както при SIFT метода [1]) се подобрява разрешаващата способност по време на алгоритъма.

### III. Експериментални резултати

#### 3.1 Симулация в среда на програмния продукт MATLAB

На фиг.5. е показана диаграма на речевия сигнал при диктор мъж и съотношение сигнал/шум -20dB за израза: 'Определяне периода на основния тон' и изчислените стойности за ПОТ по предлагания от авторите метод.



фиг.5. Диаграма на речевия сигнал и определения ПОТ

Алгоритъма е тестван с общо 1302 сегмента вокализирана реч при съотношение сигнал/шум от -20dB до 0dB. Процентното съотношение на

- центриране на СКЕ на сигнала спрямо нулата:

$$E_{us}(n) = E_{us}(n) - \frac{K}{N} \cdot \sum_{i=0}^{N-1} E_{us}(i),$$

$$n = 0 \dots \frac{N}{K} - 1; K = 1 \div 10 \quad (3)$$

- автокорелационен (АК) анализ на СКЕ на сигнала, като се изчисляват само АК коефициенти в допустимия за ПОТ интервал:

$$R(i) = \sum_{n=0}^{\frac{N}{K}-i-2} E_{us}(n) \cdot E_{us}(n+i),$$

$$i = MinT \dots MaxT \quad (4)$$

- откриване на максимума в АК функция на СКЕ на сигнала.
- изглеждане на получения ПОТ.

получените погрешни оценки на ПОТ е съизмеримо с това на традиционния автокорелационен метод.

За проверка работоспособността на предлагания алгоритм, в среда на MATLAB е симулирана вокодерна система LPC-15 (4800b/s). Речевият сигнал се дискретизира с честота 11025 Hz, размерът на анализирания сегмент N е 360 отчета, а размерът на групата K е от 4 до 6 отчета.

При слушателските тестове е установена пълна разбираемост на речта, като качеството се оценява като добро. При преминаване от размер на групата K=4 към K=5 и K=6 отчета, деградацията на качеството на синтезирания сигнал е слабо изразена поради използването на параболична интерполяция на ПОТ.

### **3.2 Тестване на алгоритъма в реално време с помощта на развойна система TMS320C50 DSK на фирмата Texas Instruments**

За сравнение обемът на изчисленията на предлагания метод с други традиционни методи е извършена оценка на времето, необходимо за определяне ПОТ с използване на специализиран процесор за цифрова обработка на сигнали TMS320C50 с цикъл на инструкцията 50ns.

Изследваният от авторите метод за определяне на ПОТ при размер на сегмента N=256 отчета (честота на дискретизация  $F_s=8064$  Hz) и размер на групата K=4, за която се изчислява СКЕ, изисква по-малко от 9000 цикъла на процесора TMS320C50. При същия размер на сегмента N=256 отчета традиционният автокорелационен метод изисква повече от 17000 цикъла на процесора и то при условие, че за изчисляване на АКФ се използва БПФ. Ако за сравнение се вземе модифицирания автокорелационен метод (автокорелационна оценка на грешката от линейно предсказване [1]), за който броят цикли на процесора надхвърля 21000, може да се заключи, че снижаването обема на изчисленията за предлагания от авторите метод е значително.

## **IV. Обобщение и заключение**

Предлаганият от авторите метод за определяне на ПОТ е съобразен с точността, постигана при квантуването и поради това значително намалява обема на изчисленията. Практическите предимства от това са, че алгоритъмът позволява използването на по-евтин процесор, или освободеното процесорно време да се използва за вмъкване на допълнителни процедури за подобряване модела на анализа на речта. Съществено предимство на предлагания алгоритъм е, че не изисква аритметика с плаваща запетая.

## **V. Литература**

1. Дж. Д. Маркел, А. Х. Грей, „Линейное предсказание речи“, Москва, Связь, 1980г.
2. Л. Р. Рабинер, Р. В. Шафер, „Цифровая обработка речевых сигналов“, Москва, Радио и связь, 1981г.
3. Wolfgang Hess, „Pitch Determination of Speech Signals“, Springer-Verlag, Berlin, 1983.