Изследвания по усъвършенстване и развитие на системите за ниско-скоростно пренасяне на речеви сигнали

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

Annotation

This paper presents an author's results in the field of low-bit-rate speech coding received during the past few years. These results are core points of a Ph.D. thesis. Total assessment of a new fast and cost-effective pitch estimation method is presented. A subjective Diagnostic Rhyme Test for the Bulgarian Language (DRTBL) is proposed for correct and reliable quality assessment of low-bit-rate vocoders adopted for Bulgarian language. According special features of the Bulgarian language, an appropriate vector quantization (VQ) scheme is investigated

I. Въведение.

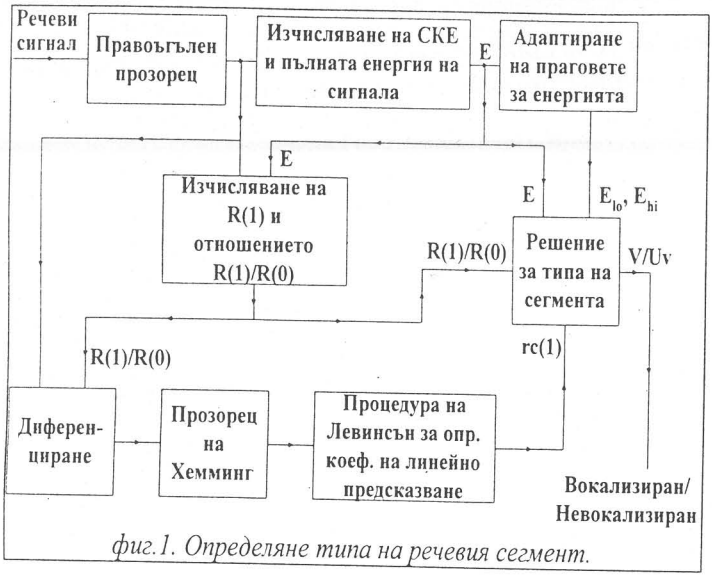
Настоящият доклад представя основните резултати от разработките на авторите в областта на ниско скоростното пренасяне на речеви сигнали. Тези резултати се явяват ядро на дисертационен труд, с тема "Усъвършенстване и развитие на системите за ниско скоростно пренасяне на речеви сигнали".

В представените три разработки, усилията на авторите са насочени към: подобряване качеството на звучене на синтезираната реч; оптимизация на известни и търсене на нови алгоритми за намаляване обема на изчисленията при реализация на вокодерни системи. В стремежа към адаптиране на LPC (Linear Predictive Coding) вокодерите с векторно квантуване (VQ) на параметрите за нуждите на българския език, авторите решават съпътстващи проблеми, като създаването на Субективен Диагностичен Римов Тест за български език (СДРТБЕ). Работи се по изследване на разпределението на различни параметри на речта в многомерното пространство, с цел създаване на структура на векторно квантуващо устройство, оптимизирана за български език. Създаден и изследван е нов бърз алгоритъм за класифициране типа на речевия сегмент и определяне периода на основния тон.

II. Бърз метод за определяне периода на основния тон на вокализираната реч.

Методът се състои от две стъпки: определяне типа на речевия сегмент и определяне стойността на периода на основния тон (ПОТ).

По време на първата стъпка (фиг.1.) се определя вокализиран или невокализиран е текущият речеви сегмент. Параметърът **rc(1)** се получава от рекурсивната процедура на Левинсън [Л.4], отношението **R(1)/R(0)** е необходимо при диференциране на речевия сигнал, а пълната енергия **E** се оценява за съгласуване на енергията на синтезирания речеви сигнал с тази на входния, така че за определяне на входните за класификатора параметри не се изискват допълнителни изчисления. В предложеният от авторите алгоритъм за определяне на ПОТ [Л.1], пълната енергия **E** се изчислява по модифицирана процедура, поз-



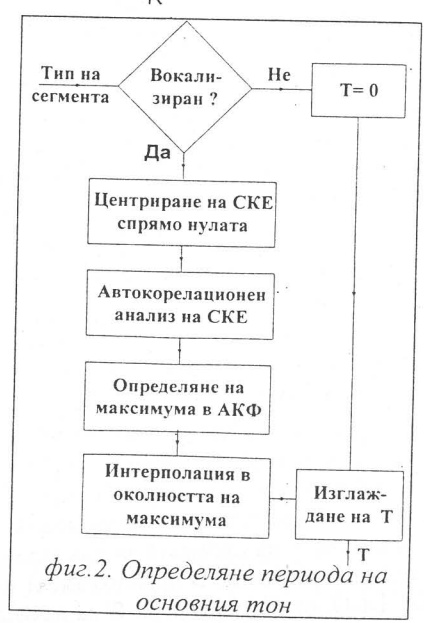
фиг.1. Определяне типа на речевия сегмент.

воляваща едновременно получаване и на свръх кратковременната енергия (СКЕ), при незначително увеличаване на обема на изчисленията. СКЕ представлява енергията на сигнала, изчислена последователно за не припокриващи се групи от по **K** отчета от входния речеви сигнал:

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

При размер на групата отчети **K**=1, имаме традиционния начин за определяне енергията на сигнала. Интерес представляват стойностите за размер на групата **K**=2...10, при които се получава значително съкращаване на обема на изчисленията при приемливо ниво на грешката в ПОТ. Стойностите за параметрите размер на групата **K** и броя отчети **N** се подбират в зависимост от честотата на дискретизация на речевия сигнал и желаната разделителна способност за ПОТ.

След определяне типа на речевия сегмент, се преминава към стъпка 2 от алгоритъма: определяне ПОТ (фиг.2). Ако сегментът е невокализиран, ПОТ **T** се приема за равен на нула и не се изискват допъл-

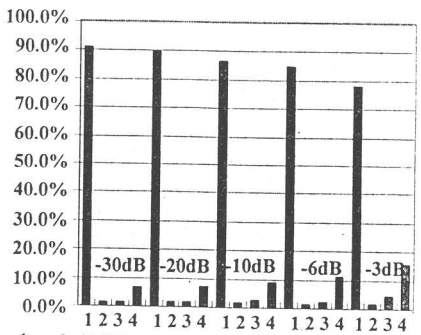


фиг.2. Определяне периода на основния тон

нителни изчисления. За отстраняване на случайни грешки при оценка на ПОТ е въведено изглаждане (по система от правила + медианна филтрация), чрез използване на стойностите на периода **T** за два съседни сегмента - предишен и следващ.Чрез параболична интерполация, аналогично на **SIFT** метода [Л.4], се подобрява разрешаващата способност по време на получения ПОТ.

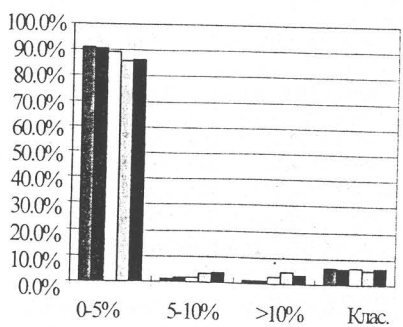
Така описаният метод за определяне ПОТ е изследван за група от шест диктора - трима мъже и три жени (общо 6000 сегмента), при нива на околния шум от -30dB до -3dB. Изследванията са извършени при размери на речевия сегмент 22.5ms, 30ms и 40ms. На фиг.3. е показано съвпадението на изчисления с истинския ПОТ (100% на графиката), получен чрез полуавтоматично визуално определяне в отсъствие на шум. За всяко ниво на околния шум е показано разпределението на относителната грешка в четири групи: '<5%', '5%÷10%', '>10%' и грешка от класификация на сегмента. При увеличаване нивото на околния шум се увеличава предимно процента на грешките от класификация. Причината е, че в участъците, където речевия сигнал е с ниска енергия, добавения шум поради повисоката си енергия, разрушава периодичната структура на речта и я превръща в невокализирана, а истинския ПОТ се определя при липса на шум и тези участъци се възприемат като вокализирани.

Легенда:
1) <5%; 2) 5%-10%; 3) >10%; 4) Класификация



фиг.3. ПОТ при адитивен околен шум (за fs=8kHz, K=2, 40ms сегмент)

На фиг.4 е показано разпределението на относителната грешка на изчисления спрямо истинския ПОТ за размер на групата отчети **K**=2÷6. Графиката е построена за 1000 речеви сегмента при ниво на околния шум -30dB, честота на дискретизация **fs**=8kHz и размер на сегмента 40ms. (Аналогично изследване е извършено за **K**=2÷9 при **fs**=11kHz.) В около 90% от случаите отклонението от истинския ПОТ е по-малко от 5%, което не влошава осезателно качеството и разбираемостта на речта. При увеличаване на размера на групата отчети **K** се наблюдава известно увеличение на отклонението от истинския период, поради намаляване на разрешаващата способност по време. Изборът на честота на дискретизация **fs**, размер на групата от-



фиг.4. Разпределение на грешката за размер на групата K=2,3,4,5,6

чети **К**, големина на речевия сегмент, се прави в зависимост от конкретните изисквания и наличните изчислителни ресурси. При фиксирана честота на дискретизация и размер на речевия сегмент, за оразмеряването на параметъра **К** се търси оптимум на критерия 'близост до истинския ПОТ/обем на изчисленията'. За честоти на дискретизация 8 - 12kHz и размер на сегмента 22.5 - 40ms, добри резултати се получават при **К**=3 и **К**=4.

При сравнение на шумоустойчивостта на предложения от авторите метод с тази на традиционният автокорелационен метод [Л.5] и **SIFT** метода [Л.4] е установено, че новият метод има малко по-ниска шумоустойчивост от автокорелационния и значително по-висока шумоустойчивост от SIFT метода, при значително намаляване обема на изчисленията спрямо всеки от двата метода.

III. Субективен диагностичен римов тест за български език за оценка на разбираемостта на речта.

Субективните тестове за оценка на разбираемостта на речта, се основават на изпитване способността на слушателите да различават фонеми, притежаващи общи признаци. Характерна особеност за субективните тестове е специфичността им за конкретният език, което и налага разработването на СДРТБЕ.

Предложеният [Л.2] от авторите субективен римов тест за български език съдържа 200 думи, групирани в 100 римувани двойки :

СБОР - СПОР	БИТИЕ - ПИТИЕ	БОР - ПОР	БАРА - ПАРА
БИЯ - ПИЯ	БРИМКА - ПРИМКА	БЛАТО - ПЛАТО	БОЛКА - ПОЛКА
БРАВО - ПРАВО	ВАЛ - БАЛ	ВАР - БАР	ВИНТ - БИНТ
ДРЕВЕН - ДРЕБЕН	ВОЛТ - БОЛТ	ЗАВИЯ - ЗАБИЯ	ОТВОР - ОТБОР
ИЗВИРАМ - ИЗБИРАМ	ВАЗА - БАЗА	СВИТ - СБИТ	ВИЯ - БИЯ
ГОРА - КОРА	ГРАЧА - КРАЧА	ГОСТ - КОСТ	ГАЛЕН - КАЛЕН
ГЛАС - КЛАС	ГРИВА - КРИВА	ГОРЕН - КОРЕН	ГОЛА - КОЛА
ГОРЯ - КОРЯ	ГРАХ - КРАХ	ГЛАСИ - КЛАСИ	ДЕН - ТЕН
МОДЕЛ - МОТЕЛ	ДУШ - ТУШ	ДОМ - ТОМ	ДИНЯ - ТИНЯ
ДВОРЕЦ - ТВОРЕЦ	ДОК - ТОК	ДЪГА - ТЪГА	ДРЕВЕН - ТРЕВЕН
ПЯЛ - БЯЛ	ДЕЛА - ТЕЛА	ПИЯ - БИЯ	ПАЛКА - БАЛКА
ТЯСНО - ДЯСНО	СЕЯ - ЗЕЯ	ТАМ - ДАМ	ТОМ - ДОМ
КОСТ - ГОСТ	ЖАРЯ - ШАРЯ	ЖИЛО - ШИЛО	ЖЕСТ - ШЕСТ
ОПЕРА - ОБЕРА	КУКА - ГУКА	ЧАС - ДЖАС	ЧОП - ДЖОБ
ПРАВО - БРАВО	ТРЕПЯ - ТРЕБЯ	КОСА - КОЗА	НИСЪК - НИЗЪК
ТРЪПНА - ДРЪПНА	ПРЯК - БРЯГ	ПРАЛ - БРАЛ	ФРАК - ВРАГ
СВИНЕЦ - ЗВЪНЕЦ	СМЕЙ - ЗМЕЙ	РОТНА - РОДНА	ТВОРЕЦ - ДВОРЕЦ
ДАЛ - ДЯЛ	ПАСВА - ПАЗВА	БАЛ - БЯЛ	ВАЛ - ВЯЛ
ЛАВА - ЛЯВА	САЛ - СЯЛ	МАРКА - МЯРКА	ЦАР - ЦЯР
РОДЪТ - РОДЯТ	ЧЕТАЛ - ЧЕТЯЛ	ГРАХ - ГРЯХ	НАМ - НЯМ
ГОЛ - ГЬОЛ	КУПЪТ - КУПЯТ	ГЛАСЪТ - ГЛАСЯТ	БРОДЪТ - БРОДЯТ
ЛУД - ЛЮТ	ПОЗОР - ПОЗЬОР	КУП - КЮП	ЛУК - ЛЮК
ЛОМ - РОМ	МИЕ - НИЕ	МАС - БАС	ЛЕВ - РЕВ
РАЙ - ДАЙ	ЛЕДНИК - РЕДНИК	РОМ - СОМ	ЛОСТ - МОСТ
КАСА - КАЦА	ЛЕД - ЗЕТ	ТИП - ЦИП	КАША - КАЧА

Таблица 1. Двойки думи, използвани в субективния диагностичен римов тест за оценка разбираемостта на речта.

При съставянето на теста са подбрани двойки думи, имащи различен смисъл и различаващи се само по една съгласна фонема, която обикновено се

намира в началото или средата на думата. Спазено е правилото различаващите се фонеми да образуват двойки на противопоставяне по някой от корелативните признаци: звучност-беззвучност или мекост-твърдост.

По време на изпитанията пред слушателите се произнася по една дума от всяка римувана двойка. Слушателите трябва да определят коя дума е произнесена. При коректен отговор се приема, че изпитваната система запазва конкретния признак. При грешен отговор, се получава информация, точно кой признак на речта не се представя добре и какви мерки да се предприемат за отстраняване на недостатъка.

Количествена оценка на разбираемостта на речта [Л.2] за тестваната система е процентното отношение от разликата между верни и грешни отговори към общия брой на двойките думи, включени в теста:

$$DRT_{BGL} = \frac{N_{right} - N_{wrong}}{N_{total}}.100[\%],$$

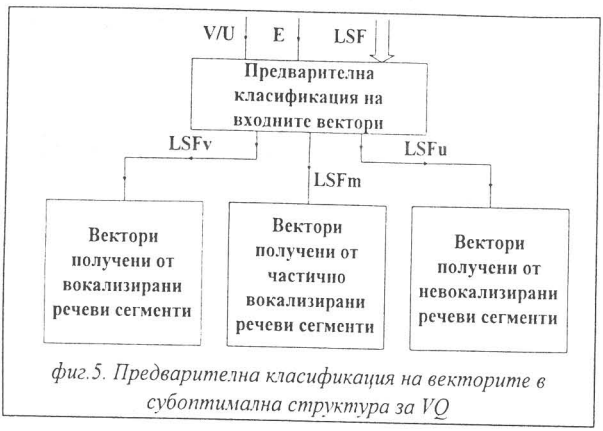
където **Nright** е броят на верните отговори, **Nwrong** е броят на грешните отговори, **Ntotal** е общият брой на включените в теста двойки думи. Типични стойности за $\mathbf{DRT_{BGL}}$ са в обхвата от 76 до 98, като за "добра" система $\mathbf{DRT_{BGL}}$ е около 90.

Резултатите от теста до известна степен зависят от броя и пола на диктторите, тъй като използваният в изследваната система метод за обработка на речевия сигнал може да има различен ефект върху различните характеристики на речевия апарат на отделните личности. Експериментално е доказано, че три мъжки и три женски гласа са достатъчни за постигане на достоверна оценка на изпитваната система.

При съпоставяне на резултатите от изпитанията на LPC-10 вокодерна система със субективния диагностичен римов тест за български език и Diagnostic Rhyme Test (DRT) за английски език се получават аналогични резултати, но не и пълно съвпадение. Това се дължи на различната специфика на двата езика, начина по който вокодерната система деформира синтезираната реч и степента на значимост на тези деформации за съответния език, в зависимост от позицията им в думата, фразата или изречението.

IV. Изследване разпределението на LSF (Line Spectrum Frequency) параметрите в многомерното пространство.

Използването на векторното квантуване на параметрите на речта в съвременните ниско скоростни вокодерни системи е традиционно средство за снижаване обема на пренасяната информация, със запазване на относително по-високо качество спрямо стандартните LPC вокодери. При адаптиране на вокодерите за нуждите на българския език е необходима съответна оптимизация на структурата на устройството за векторно квантуване, съгласно особеностите на езика. Изследва се разпределението на LSF параметрите на речта в многомерното пространство, за откриване оптималното групиране на елементите принадлежащи на един входен вектор. За постигане на оптимално съотношение



фиг.5. Предварителна класификация на векторите в субоптимална структура за VQ

'качество на реч-та/ обем на изчисленията' при зададена скорост на обмен по канала за връзка, авторите провеждат поредица от изследвания, целящи създаване на подходяща субоптимална многослойна структура на векторната таблица и съответен бърз алгоритъм за търсене на най-близкия до текущият вектор (за сметка на увеличен обем на паметта, необходима за съхраняване на кодовата таблица). Предварителното разделение на структурата на поднива (фиг.5.) се основава на базовите признаци: вокализиран/частично вокализиран/ невокализиран сегмент, относително ниво на енергията и параметрите на входния вектор. Енергията на речевия сегмент и ПОТ се кодират самостоятелно.

Получените първоначални резултати са обнадеждаващи. Последващите задълбочени изследвания ще бъдат обобщени и представени в самостоятелна публикация.

V.Обобщение и заключение.

В доклада са представени разработки на авторите в областта на нискоскоростно пренасяне на речеви сигнали. Предложен е бърз метод за класифициране типа на речевия сегмент и определяне периода на основния тон. Направени са изследвания за устойчивостта на метода при различни отношения сигнал/шум. Предложен е Диагностичен Римов Тест за Български Език за изпитване на вокодерни системи. Анонсират се изследвания по създаване на субоптимална схема за векторно квантуване на LSF параметрите на речта, предназначена за оптимизиране качеството на пренасяната реч при определена скорост на обмен по канала за връзка.

Библиография :

1. Колев Й.Н.,Т.Д. Ганчев, Др.Д. Николов, "Метод за определяне периода на основния тон на вокализиран речеви сигнал", Електронна Техника '96, Созопол, 1996г.
2. Колев Й.Н., Т.Д. Ганчев, Хр.И.Сверчкова, К.П.Грибнева, "Субективни методи за оценка на качествата на вокодерни системи за български език", Юбилейна научна сесия '97, Технически Университет-Варна, 16-18 Октомври, 1997г.
3. Party, I.S. Burnett, J.F. Chicharo, "The Consequences of Linguistic Perception on Low-Rate Speech Coding", Proceedings ICASSP-97, Munich, Germany, April 1997, Vol.2, pp.1383-1386
4. Дж. Д. Маркел, А. Х. Грей, "Линейное предсказание речи", Москва, Связь, 1980г.
5. Wolfgang Hess, "Pitch Determination of Speech Signals", Springer-Verlag, Berlin, 1983г.