

Предаване на реч в цифров вид по нискоскоростни канали

доц.ктн инж.Йордан Николов Колев - ТУ Варна

инж.Тодор Димитров Ганчев - ТУ Варна

Настоящият доклад има за цел да представи първите резултати постигнати от колектива в изследванията по нискоскоростно предаване на реч. Тези резултати не се възприемат от авторите като окончателни, а само като начало на серия от по-задълбочени изследвания, свързани с изучаване особеностите на българския език и използването им за постигане на по-високо качество на речта и по-ниски скорости на обмен.

Първите изследвания бяха свързани с експериментирание на някои базови алгоритми реализиращи А-закон и μ -закон за компресия и АДИКМ при различни скорости на обмен. На следващият етап бе поставена целта да се реализира вокодерна система съгласно известни от литературата принципи, но адаптирани за конкретните нужди. Преди представянето на реализирания от авторите проект, ще се спрем на кратко на механизма на речеобразуването и теоретичните граници, към които се стремят разработките в тази област.

Анализът на работата на центровете на речта и системите за управление на артикулационния апарат на човека е показал, че на участъка от централната нервна система до речевия апарат, скоростта на предаване на сигналите за управление по средна стойност не превишава 100 bit/s. Всъщност това са десетина сигнала изработвани от центъра на речта в мозъка, движещи се към артикулационните органи със скорост не превишаваща 10 звука за секунда. В същото време за качествено предаване на речеви сигнали без предварителна обработка е необходима скорост около 100 kbit/s (честотна лента на речевия сигнал 7kHz, динамичен диапазон 42dB и честота на дискретизация 14kHz). Причината за това несъответствие е, че в артикулационния апарат постъпват управляващи сигнали с ограничен обем, а се създава речеви сигнал с много по-голям обем. Същността на това явление се състои в факта, че с помощта на артикулационния апарат се извършва спектрално-временна модулация на широколентов спектър, генериран от източниците на речеви колебания, който играе ролята на носещо колебание и почти не съдържа полезна информация отнасяща се към разбираемостта на речта. Очевидно е, че за икономично предаване на речевия сигнал е необходимо да се отдели и предава по канала само полезната информация, а съпътстващата да се генерира в приемната страна. В идеалният случай, за това би бил достатъчен канал с пропускателна способност около 100 bit/s, по който да се предават само артикулационните

сигнали. Предаването на индивидуалните особености на произнасянето на речта би изисквало още известно количество битове, но общото количество на информацията остава не повече от 200-300 bit/s.

Опитите за практическа реализация на такова устройство, наречено вокодер, са започнати от H.Dudley [5] още през 30-те години. Открити са различни алгоритми, регистрирани са няколко стандарта (VSELP, CELP, RELP, MLPC и др. [5]), предложени са разработки доближаващи се до теоретичните граници, но изследванията за подобряване качеството на синтезираната реч и снижаване обема на предаваната информация продължават.

От авторите е реализиран параметричен спектрално-временен вокодер с линейно предсказване на речта - ВОКОДЕР LPC-10. От отчетите на речевия сигнал се определят две групи параметри: параметрите характеризиращи спектрално-временната обвивка на речевия сигнал (филтровата функция) и параметрите характеризиращи източниците на речеви колебания (генераторна функция). Към параметрите на генераторната функция се отнасят честотата на основния тон и информация за типа на сегмента - вокализиран/невокализиран. Реализираният ВОКОДЕР LPC-10 е предназначен за нискоскоростно предаване на реч по телефонна линия и скорост 3600 bit/s.

Работата на ВОКОДЕР LPC-10, както и на повечето известни вокодерни системи, се базира на разделяне на речевия сигнал на сегменти и анализ на параметрите на речта в рамките на един сегмент. При анализа се извършва оценка на следните параметри: коефициенти на отражение, средна стойност на енергията на сигнала на възбуждане и период на основния тон. Тези параметри се предават по канала за връзка и по тях в приемника се синтезира речевия сигнал. При анализа се използва сегмент с фиксирана дължината от 31.74ms, еквивалентен на 256 отчета от входния сигнал при честота на дискретизация 8.064kHz. Сегментите се презастъпват с 50%. Анализират се 63 сегмента за секунда, като за всеки сегмент по канала за връзка се предават по 57 бита, или общо 3591 bit/s. Разпределението на тези 57 бита за сегмент е следното:

- | | |
|---|--------------------|
| 1) Коефициенти на отражение RC(0) ... RC(9) | |
| RC(0)...RC(6) | по 5 бита за всеки |
| RC(7) | 4 бита |
| RC(8) и RC(9) | по 3 бита |
| 2) Енергия на с-ла (σ) | 6 бита |
| (по логаритмичен закон) | |
| 3) Период на ОТ (P) | 6 бита |

Общото закъснение на речевият сигнал във ВОКОДЕР LPC-10 е около 80ms и се формира като сума от две закъснения: 64ms при анализа, поради въведената интерполация на периода на ОТ; 16ms при синтеза поради интерполацията на параметрите за два съседни сегмента.

Вокодерът е експериментиран в среда на MATLAB, а в последствие реализиран на език С. При изследванията е използвана развойна система DSP Starter Kit с процесор TMS320C50, на Texas Instruments и разработени към нея разширителни модули.

Принципът на действие на вокодера се пояснява с блоковата схема показана на фиг.1. Оцифрованият с помощта на АЦП входен сигнал се подлага на предварителна филтрация с цифров филтър от първи ред $x(n)=s(n)-s(n-1)$ за подчертаване на ВЧ съставляващи. Така полученият сигнал се притегля с прозорец на Hamming. Процедурата за изчисляване на коефициентите на обратния филтър е реализация на рекурсивния алгоритъм на Levinson [3], който ще изложим тук накратко. Алгоритъма се състои от две части: задаване на начални условия и рекурсивно изчисляване с $M-1$ на брой стъпки на коефициентите на обратния филтър от M -ти ред. На всяка m -та стъпка от рекурсивната процедура се изчислява израза:

$$1) \quad A_{m+1}(z) = A_m(z) + k_{m+1} \cdot B_m(z), \quad \text{за } m=2,3,\dots,M.$$

Коефициентите на отражение се определят от израза:

$$2) \quad k_{m+1} = \frac{-1}{\alpha_m} \sum_{i=0}^m r(m+1-i) \cdot a_{m,i}$$

След като коефициентите на обратния филтър са изчислени входният речеви сигнал $S(z)$ се подлага на филтрация с обратния филтър $A(z)$:

$$3) \quad E(z) = S(z) \cdot A(z)$$

Където $E(z)$ е функцията на възбуждане на речевия тракт. Аналог на 3) във временната област е разликовото уравнение:

$$4) \quad e(n) = \sum_{i=0}^{M-1} a_i \cdot x(n-i), \quad \text{където } a_0 = 1 \text{ и } M=10$$

По зависимостта:

$$5) \quad \sigma = \frac{1}{N} \sum_{n=0}^{N-1} |e(n)|, \quad \text{при } N=256$$

се изчислява средната стойност σ на сигнала $e(n)$ и се предава по канала за връзка за съгласуване на енергиите на синтезирания сигнал с оригиналния речеви сигнал. Решение за типа (вокализиран/невокализиран) на сегмента, се взема след изчисляване на автокорелационната функция на сигнала $e(n)$

$$6) \quad r_e(k) = \sum_{n=0}^{N-1-k} e(n) \cdot e(n-k), \text{ където } N=256, \text{ а } k=0,1,\dots,128$$

и сравнение на максималният коефициент $r_e(k)$ с променлив праг зависещ от позицията на $r_e(k)$. Ако амплитудата на максималният коефициент $r_e(k)$ превишава прага, то сегмента се счита за вокализиран и периода на основния тон се получава чрез параболична интерполация на позицията на максималния $r_e(k)$ и съседните му два коефициента. Максимумът се търси в интервала от 1.5ms до 15.9ms, което отговаря на честота на основния тон от 63Hz до около 690Hz и покрива целия диапазон от ниските мъжки до високите детски гласове.

Възстановяването на речевия сигнал в приемника се осъществява чрез синтезиращ филтър с предавателна характеристика:

$$7) \quad S(z) = \frac{E(z)}{A(z)},$$

където $S(z)$ е синтезираният сигнал, $E(z)$ е възбуждането на филтъра, а $A(z)$ е предавателната характеристика на обратния филтър. Използуван е синтезиращ филтър със решетъчна структура с два множителя [3].

Синтезиращият филтър, коефициентите на който се предават по канала, формира сигнал $s(n)$, при входен сигнал $e(n)$ и като се отчита и средната енергия σ . Изходът от филтъра $s(n)$, се подлага на нискочестотна филтрация с филтър $x(n) = s(n) + x(n-1)$, възстановяващ речевия сигнал. За подобряване качеството на звучене на речта, допълнително се извършва нискочестотна филтрация с филтър с предавателна характеристика $\sin(x)/x$. ЦАП преобразува сигнала в аналогов, който след това се изглажда с аналогов филтър от втори ред и се излъчва от високоговорител.

Реализираният по тази схема вокодер синтезира напълно разбираема реч със запазване специфичната индивидуалност на диктора, но естествеността и натуралността на звучене е ниска. Усилията за усъвършенстване на тази разработка ще бъдат насочени главно в две направления: подобряване качеството на синтезираната реч и постигане на по-ниски скорости на обмен.

ЛИТЕРАТУРА:

- 1) Л.Р.Рабинер, Р.В.Шафер, Цифровая обработка речевых сигналов, Радио и связь, Москва, 1981г.
- 2) М.В.Назаров, Ю.Н.Прохоров, Методы цифровой обработки и передачи речевых сигналов, Радио и связь, Москва, 1985г.
- 3) Дж.Д.Маркел, А.Х. Грей, Линейное предсказание речи, Связь, Москва, 1980г.
- 4) Y.M.Cheng, D.O'Shaughnessy, On 450-600 b/s Natural Sound Speech Coding, IEEE Trans. on Speech and Audio Processing, vol.1, 1993
- 5) Gr. Marven, G. Ewers, A simple approach to Digital Signal Processing, Texas Instruments, 1993

