МНОГОУРОВНЕВОЕ ВЕКТОРНОЕ КВАНТОВАНИЕ С ПЕРЕМЕННОЙ ГЛУБИНОЙ ПОИСКА В ПЕРЦЕПТУАЛЬНЫХ КОДЕРАХ РЕЧИ С ПСИХОАКУСТИЧЕСКОЙ МОДЕЛЬЮ НА ОСНОВЕ СЖАТОГО ДПФ

Лившиц М.З., Парфенюк М., Петровский А.А.

Белорусский государственный университет информатики и радиоэлектроники
ул. П. Бровки, 6, БГУИР, каф. ЭВС, 220013, Минск Беларусь, E-mail: mlivshitz@tut.by, palex@it.org.by

Аннотация

Для кодеров с многополосным возбуждением количество бит, необходимых для представления информации, возрастает пропорционально количеству субполос, на которые разбивается частотный диапазон кодируемого сигнала. Следовательно, необходим компромисс между количеством субполос, величиной (глубиной) субполосных кодовых книг и качеством реконструированной речи. В настоящей работе предлагается эффективная схема оптимизации глубины поиска в многоуровневом векторном квантовании по мультиполосной кодовой книге [1], использующая психоакустическую модель на базе сжатого ДПФ (WDFT).

1. Основные принципы сжатого ДПФ

В известной работе Джонстона [2] психоакустическая модель основана на ДПФ: расчет ДПФ взвешенного сегмента сигнала, группировка коэффициентов преобразования в группы, соответствующие критическим полосам восприятия и расчет энергии в критических полосах. Достижение приемлемого спектрального разрешения в наименьших критических полосах требует использования ДПФ с достаточно длинным временным окном. Поэтому концептуальная простота и эффективность нивелируются недостаточным временным разрешением, неприемлемым для анализа более тонкого феномена, такого как маскирование назад (pre-masking) [3]. Второй класс, использующий неравномерные банки фильтров для декомпозиции сигнала, исключает этот недостаток, однако имеет достаточную вычислительную сложность, особенно если необходима хорошая аппроксимация критических полос. Ни одно из решений не превосходит другое, оба из подходов находят свое применение.

Исследования, проведенные в работах [4],[5], показали, что малоразмерное сжатое ДПФ, может успешно заменить ДПФ с большой длиной выборки. Это оказалось возможным благодаря тому, что сжатое преобразование позволяет разместить частотные компоненты в соответствии с распределением критических полос, поэтому в психоакустической модели на базе WDFT могут быть уравновешены как хорошее частотное, так и временное разрешение.

Сжатое дискретное преобразование Фурье (WDFT) последовательности $x[n]$ из N точек определяется по формуле [4]

$$X(z_k) = X(\mathcal{E}_k) = \sum_{n=0}^{N-1} x[n]\mathcal{E}_k^{-n} \qquad k = 0,\ldots,N-1, \qquad (1)$$

где $\mathcal{E}_k$ -изображения равноотстоящих точек на единичной окружности в z -плоскости, получаемые из преобразования

$$z_k^{-1} = e^{-j\frac{2\pi k}{N}} \quad \rightarrow \quad \mathcal{E}_k^{-1} = A(z_k) \quad k = 0,\ldots,N-1 \qquad (2)$$

с произвольным порядком всепропускающей функции $A(z)$.

Простейший вариант WDFT основан на всепропускающем звене первого порядка с действительным коэффициентом [4]

$$z^{-1} \quad \rightarrow \quad A(z) = \frac{-a + z^{-1}}{1 - az^{-1}}. \qquad (3)$$

Условием стабильности является $|a| < 1$. В зависимости от знака a, растягивается низкочастотный или высокочастотный диапазон, таким образом, что оставшаяся часть единичной окружности становится сжатой. Формально это может быть выражено [5]

$$\theta = \omega + 2\arctan\left(\frac{a\sin\omega}{1 - a\cos\omega}\right) \quad for \quad \begin{cases} z = e^{j\omega} \\ \mathcal{E} = e^{j\theta} \end{cases}. \qquad (4)$$

Первый шаг при использовании WDFT в психоакустической модели – проектирование соответствующего всепропускающего преобразования. Частотные коэффициенты z -преобразования должны быть представлены равномерно в перцептуальной области. В работе [6] было показано, что всепропускающее звено первого порядка достаточно хорошо аппроксимирует перцептуальную шкалу барков, при этом значение коэффициента всепропускающего фильтра для заданной частоты дискретизации определяется по следующему выражению:

$$a_{Bark} = 0.1957 - 1.048\left[\frac{2}{\pi}\arctan\left(0.07212\frac{f_s}{1000}\right)\right]^{\frac{1}{2}}. \qquad (5)$$

Для нашего случая $a = -0.57827$ (Fs=16 kHz). Вопросы, связанные с эффективным вычислением WDFT подробно обсуждаются в [5].

2. Формирование критических частотных полос на основе WDFT

Таблица 1 – Отображение коэффициентов ДПФ и сжатого ДПФ на критические полосы

		Part A (FFT size = 512, Fs=16 kHz)				Part B (WDFT size =512, Fs=16 kHz)		
CB		Bin range	No.	Freq. [Hz]		Bin range	No.	Freq. [Hz]
1		1 – 3	3	31 - 94		1 - 12	12	8 – 100
2		4 – 6	3	125 - 188		13 - 24	12	109 – 202
3		7 – 9	3	219 - 281		25 - 36	12	210 – 305
4		10 – 13	4	313 - 406		37 - 48	12	314 - 412
5		14 – 16	3	438 - 500		49 - 60	12	421 - 523
6		17 – 20	4	531 - 625		61 - 73	13	533 - 650
7		21 – 24	4	656 - 750		74 - 85	12	660 - 776
8		25 – 29	5	781 - 906		86 - 97	12	787 - 912
9		30 – 34	5	938 - 1063		98 - 110	13	923 - 1073
10		35 – 40	6	1094 - 1250		111 - 123	13	1086 - 1254
11		41 – 46	6	1281 - 1438		124 - 135	12	1269 - 1443
12		47 – 54	8	1469 - 1688		136 - 148	13	1460 - 1680
13		55 – 62	8	1719 - 1938		149 - 161	13	1700 - 1961
14		63 – 73	11	1969 - 2281		162 - 174	13	1985 - 2302
15		74 – 86	13	2313 - 2688		175 - 186	12	2331 - 2690
16		87 – 102	16	2719 - 3188		187 - 198	12	2726 - 3174
17		103 – 122	20	3219 - 3813		199 - 210	12	3220 - 3792
18		123 – 145	23	3844 - 4531		211 - 221	11	3851 - 4513
19		146 – 173	28	4563 - 5406		222 - 231	10	4588 - 5328
20		174 – 205	32	5438 - 6406		232 - 242	11	5419 - 6412
21		206 – 243	38	6438 - 7594		243 - 252	10	6520 - 7533
22		244 – 256	13	7625 - 8000		253 - 256	4	7650 - 8000

Таблица 2 – Характеристика субполос кодера

Subband	Freq.range,Hz	Barks
1	100-510	4
2	510-1080	4
3	1080-1720	3
4	1720-2320	2
5	2320-3150	2
6	3150-4100	1.5
7	4100-5300	1.5
8	5300-8000	3

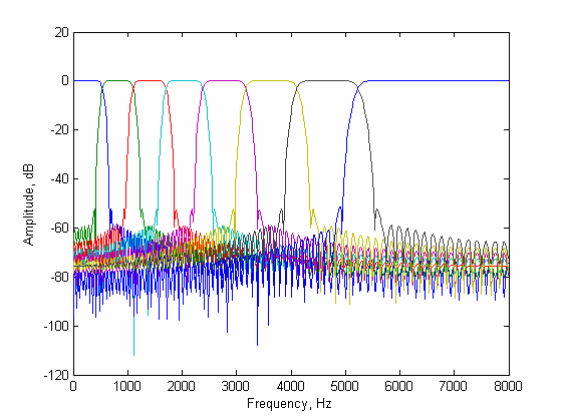


Рис.1. АЧХ банка фильтров

Так как ДПФ имеет равномерное частотное разрешение, в то время как ширина критических полос строго изменяются с их местоположением на частотной шкале, то различное количество коэффициентов

преобразования ассоциируется с конкретной критической полосой. В части А таблицы 1 количество коэффициентов в группах варьируется от 3 до 38, в то время как для WDFT той же размерности не отдается предпочтения ни одной из полос, все коэффициенты преобразования распределены практически равномерно (часть В).

В предлагаемом широкополосном перцептуальном CELP-кодере принята шкала барков и используется схема разбиения на субполосы, представленная в таблице 2, а АЧХ банка фильтров показана на рис.1.

При фиксированной схеме многоуровневого квантования [1], кодер обеспечивает высокое качество реконструированной речи, при этом банк фильтров используется лишь при составлении кодовых книг. Упрощенная схема перцептульного кодера представлена на рис.2 (слева), где часть, отвечающая за психоакустический критерий глубины поиска, обведена жирным контуром. Стратегия поиска в субполосной книге определяется на основе перцептуальной важности полосы, оценка которой осуществляется на основе порога маскирования и перцептульной энтропии [2] (рис.2, справа).

Как показали практические исследования, согласованность нелинейного банка фильтров, используемого для обучения кодовых книг, и психоакустической модели на базе WDFT, аппроксимирующей шкалу барков, обеспечивает значительное снижение перцептуальной избыточности, в то время как модель многоуровневого векторного квантования [1] снижает статистическую избыточность сигнала.

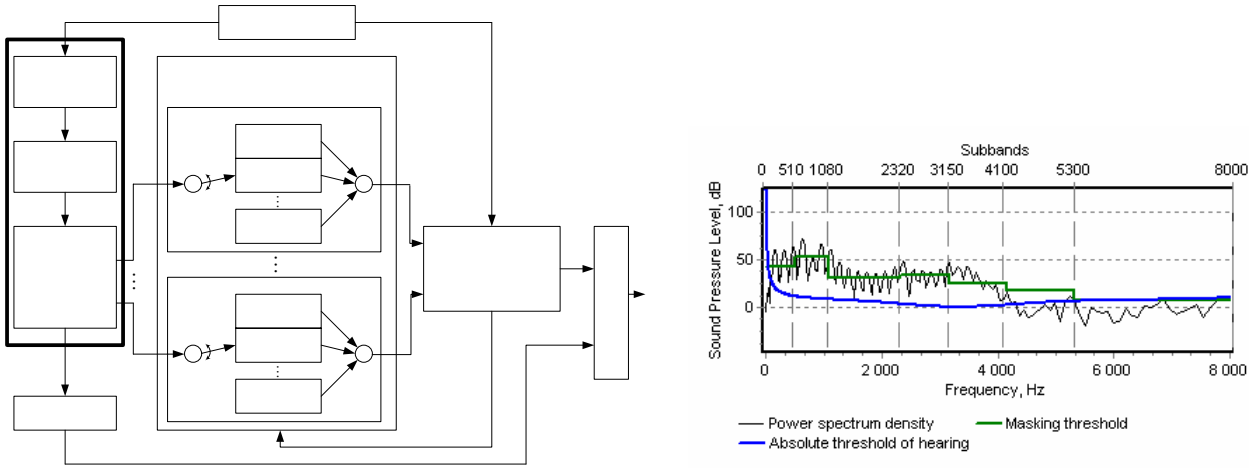


Рис.2. Перцептуальный кодер с многоуровневым векторным квантованием на базе WDFT-психоакустической модели (слева) и оценка порогов маскирования в субполосах (справа)

3. Экспериментальные результаты

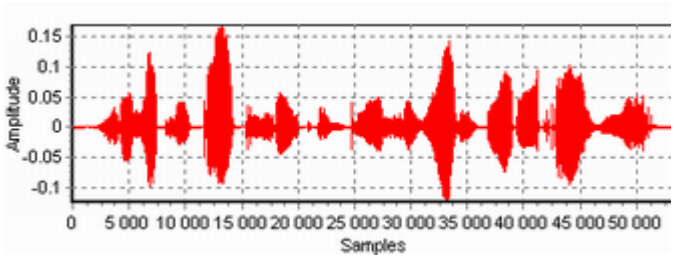
Задана кодовая книга с L=5 уровнями (16;32;64;128;256 векторов) и N=8 субполосами.

Анализ данных таблицы 3 и рис.3 показывает, что эффективность схемы с переменной глубиной поиска на базе предлагаемой модели превосходит MPEG Psychoacoustic Model 1 [7] по пиковым значениям потока данных приблизительно в $256/150 \approx 1.7$ раза, в то время как по среднему потоку данных показатель составляет приблизительно $12105/2811 \approx 4.31$ раза при отсутствии перцептуальных различий в реконструированной речи. В обеих схемах поиска по мультиполосной кодовой книге использовалось квантование параметров последнего по 10-битной кодовой книге, что, для принятой схемы кодера, вносит вклад в суммарный поток данных 500 бит/с. При этом требуемое количество информации для кодирования коэффициентов усиления векторов возбуждения, также снижается.

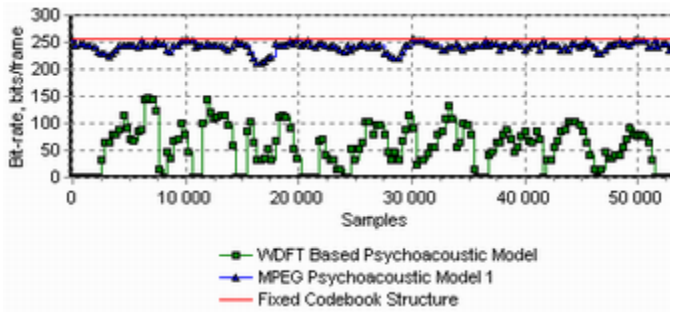
Таблица 3 – Оценка качества оптимизации

Codebook Search Optimization Model	Bit-rate peak value, bits/frame	Codebook Information Burden, bits/s			segSNR, dB	Perceptual Quality
		Codebook Indexes	Bit Allocation Strategy	Overall bit-rate		
Fixed Structure	256	12800	0	12800	13,59	Absence of perceptual differences
MPEG Model 1	256	12105	500	12605	13,21	
WDFT Model	**150**	**2811**	**500**	**3311**	**12,15**	

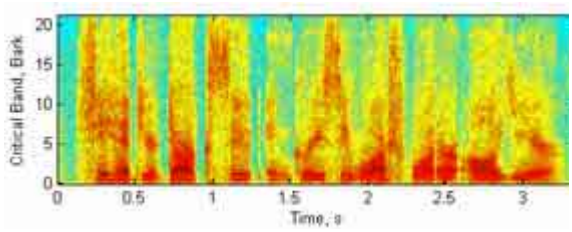
Эффективность предлагаемой схемы векторного квантования по мультиполосной многоуровневой кодовой книге с переменной глубиной поиска может быть оценена как отношение потока данных, представляющего информацию в кодовой книге при фиксированной схеме поиска, к потоку данных при динамической схеме поиска на основе WDFT-психоакустической модели. Выигрыш составляет примерно $12800/3311 = 3.86$ раза по среднему значению потока и $256/160 = 1.6$ раза – по пиковому.



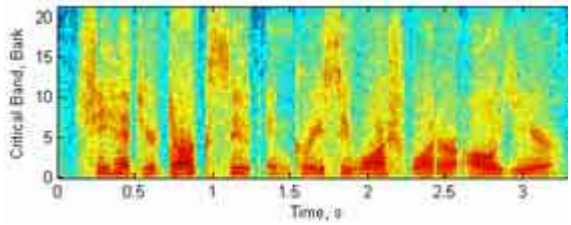
а)



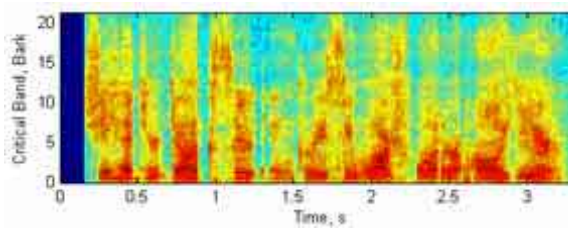
б)



в)



г)



д)

а) оригинальный сигнал во времени; б) сравнение потоков данных для различных схем поиска; в) спектрограмма речевого сигнала с фиксированной структурой поиска; г) спектрограмма сигнала с MPEG-оптимизацией поиска; д) спектрограмма сигнала с WDFT-оптимизацией поиска

Рис.3. Экспериментальные данные

4. Выводы

Экспериментальные данные показывают эффективность применения психоакустической модели на базе WDFT в перцептуальных кодерах речи. При этом ее сложность в вычислительном плане сопоставима с другими моделями, существующими на сегодняшний день. Показано, что количество бит для кодирования информации в кодовых книгах может быть снижено в несколько раз, при незначительном ухудшении соотношения сигнал-шум и без перцептуального различия реконструированной речи для сравниваемых схем поиска.

Литература

1. М.З. Лившиц, А.А. Петровский, “Многоуровневое векторное квантование речевого сигнала по мультиполосной кодовой книге в широкополосном CELP-кодере с психоакустической мотивацией”, *IV Международная конференция “Цифровая обработка сигналов и ее применение”*, труды РНТОРЭС им. А.С. Попова, выпуск VI-1, стр.119-123, Москва, 2004.

2. J.D. Johnston, “Transform coding of audio signals using perceptual noise criteria,” *IEEE J. Selected Areas in Comm.*, vol.6, pp.314-323, Feb. 1988.

3. T. Thiede, E. Kabot, “A New Perceptual Quality Measure for Bit Rate Reduced Audio,” *Proc. 100th AES Convention*, Copenhagen, Preprint 4280, 1996.

4. A. Makur, S.K. Mitra, “Warped Discrete-Fourier Transform: Theory and Applications,” *IEEE Trans. Circuits Systems I*, vol.48, pp.1086-1093, Sept. 2001.

5. M. Parfieniuk, A. Petrovsky, “Warped DFT as the basis for psychoacoustic model,” *Proc. ICASSP*, vol. IV, pp.185-188, May 2004, Montreal, Canada.

6. J.O. Smith III, J.S. Abel, “Bark and ERB Bilinear Transforms,” *IEEE Trans. Speech, Audio Processing*, vol.7, pp.697-708, June 1999.

7. ISO/IEC JTC1/SC29/WG11 NO803, MPEG, International Standard IS 13818-3 Information Technology – Generic Coding of Moving Pictures and Associated Audio: Audio, 11th November 1994.