

# ОБЪЕКТИВНАЯ ОЦЕНКА КАЧЕСТВА ВОССТАНОВЛЕННОГО АУДИО СИГНАЛА ПЕРЦЕПТУАЛЬНЫМ ПДВП-КОДЕРОМ НА БАЗЕ ПЕРИФЕРИЙНОЙ МОДЕЛИ УХА ЧЕЛОВЕКА

Петровский А.

Белорусский государственный университет информатики и радиоэлектроники,  
ул. П.Бровки 6, Минск, Беларусь, e-mail: [apetrows@bas-net.by](mailto:apetrows@bas-net.by)

**Реферат:** В данной статье показан метод объективной оценки качества восстановленного аудио сигнала на основе периферийной модели уха человека. Применение банка вэйвлет-фильтров, адаптированного к шкале критических частот, и использования отношения шума к порогу маскирования в качестве оценки качества, рассчитанной в вэйвлет-области, позволяет получить оценки, которые в большей степени коррелированы с результатами субъективных тестов, нежели традиционные (отношение сигнала к порогу маскирования или к мощности вносимых искажений).

## Введение

Традиционная оценка качества сигнала – отношение сигнала к шуму  $SNR$  (рис.1) не отражает объективной картины качества восстановленного сигнала аудио кодера [1]. Тесты для проверки качества аудио кодера регламентируются по ITU-R Recommendation BS.1116, 1997. Стандарт для обеспечения оценки качества перцептуального аудио кодера PEAQ – Perceptual Evaluation of Audio Quality, схема которого основывается на периферийной модели уха человека [2], может быть применён для перцептуального аудио кодера на основе адаптивного пакета дискретного вэйвлет-преобразования (ПДВП) [3,4].

## Оценки качества реконструированного аудио сигнала

Большинство оценок качества восстановленного сигнала базируется на отношении энергии шума (энергии вносимых искажений) к порогу маскирования  $NMR$  (см. рис.1), которая оценивает расстояние между фактическими и максимально неслышимыми искажениями. Исследования последних лет [2] показывают высокую коррелированность  $NMR$  с субъективными тестами, основанными на прослушивании аудио сигналов. Другим важным параметром объективной оценки качества реконструированного аудио сигнала является относительное число искаженных фреймов к общему числу фреймов ( $MF$ - masking flag) [1] кодируемого аудио сигнала, которая даёт оценку вероятности того, что восстановленный аудио сигнал содержит слышимые искажения. Фрейм считается искаженным тогда, когда  $NMR$  хотя бы в одной анализируемой частотной полосе превышает величину 1,5 дБ. Третьим параметром, характеризующим искажение тембра, притупление или заглушение аудио сигнала, является параметр - частотная полоса сигнала ( $SB$  – signal bandwidth). Данные искажения возникают, когда аудио кодеры ограничивают частотную полосу сигнала для увеличения степени компрессии. За оценку  $SB$  для данного фрейма принимается ближайшая слева от частоты 21,5 кГц частота, спектральная энергия которой на 10 дБ превышает оценку минимального уровня шума, вычисленную на интервале [21,5-22,05 кГц] для этого фрейма. Средняя величина  $SB$  на некотором количестве фреймов оценивается независимо для оригинального и восстановленного сигналов [2].

## Периферийная модель уха – ПДВП-дерево критических частот

Периферийная модель уха базируется на модели дерева ПДВП  $(l,n) \in E_{CB}$  (см. рис. 2), которое делит весь частотный диапазон [0 – 22,05 кГц] на 25 неравномерных полос, согласованных со шкалой критических частот, т.е. на 25 барков. Корневой узел  $(l,n) = (0,0)$  данного дерева соответствует всему частотному диапазону сигнала. Каждый внутренний узел  $n$  уровня  $l$  дерева  $E_{CB}$  (узел предок), делится на два потомка ассоциируемые с высокочастотной и низкочастотной фильтрацией с последующей децимацией выходных сигналов (вэйвлет-коэффициентов). Банк вэйвлет-фильтров, построенный в соответствии со структурой  $E_{CB}$ , также используется при перцептуальном кодировании аудио сигналов.

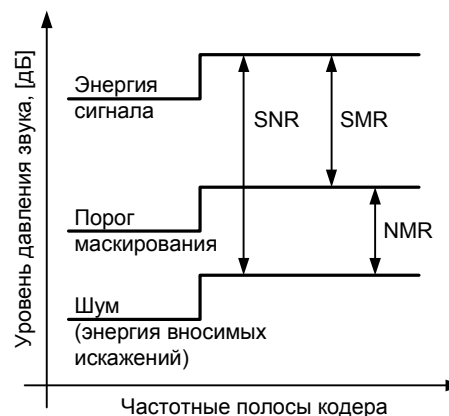


Рис.1. Объективные оценки качества сигнала

**Схема вычисления оценок**

Оценка  $NMR$   $i$ -го фрейма сигнала  $x_i(n/f_s)$  вычисляется по следующей схеме (рис. 3). Определяются ошибка восстановления сигнала  $e_i(n/f_s) = \hat{x}_i(n/f_s) - x_i(n/f_s)$  и её дисперсия в каждой критической частотной полосе  $\sigma_{e_{l,n}}^2(i)$ ,  $(l,n) \in E_{CB}$ . По параллельному каналу для оригинального аудио сигнала вычисляются пороги маскирования  $T_{l,n}(i)$ ,  $(l,n) \in E_{CB}$  в соответствующих критических полосах на основании процедуры расчета порогов маскирования в взйвлет-области [3].

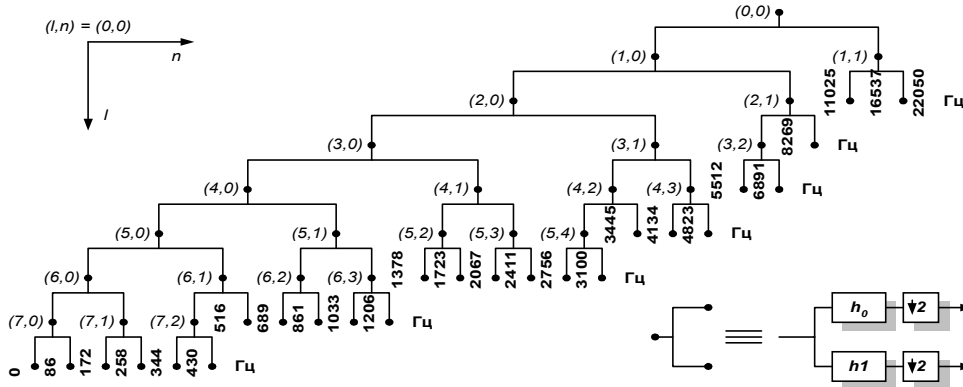


Рис. 2 Дерево ПДВП  $(l,n) \in E_{CB}$

На базе данных оценок для  $i$ -го фрейма находится соотношение

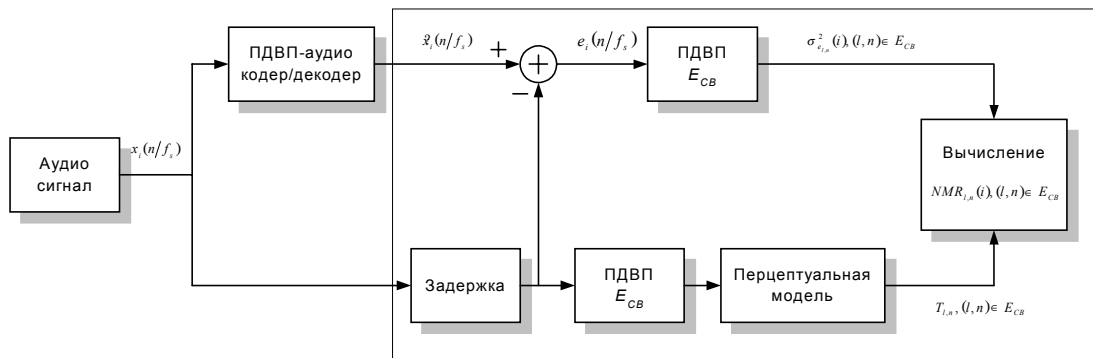


Рис. 3 Блок-схема вычисления  $NMR_{l,n}(i)$ ,  $(l,n) \in E_{CB}$

$$NMR_{l,n}(i) = \frac{\sigma_{e_{l,n}}^2(i)}{T_{l,n}(i)}, (l,n) \in E_{CB}$$

и среднее арифметическое по всем критическим частотным полосам для  $i$ -го фрейма:

$$NMR_{loc}(i) = 10 \cdot \log_{10} \left( 1/25 \cdot \sum_{\forall (l,n) \in CB-WPD} NMR_{l,n}(i) \right) \text{ [дБ]}. \quad (2)$$

Объективные показатели, характеризующие качество кодирования аудио сигнала данным перцептуальным кодером, определяются средним арифметическим  $NMR_{loc}(i)$  и средним геометрическим  $NMR_{loc}(i)$ :

$$NMR_{total} = 10 \cdot \log_{10} \left( \frac{1}{N} \sum_{i=1}^N 10^{(NMR_{loc}(i)/10)} \right) \text{ [дБ]}, \quad NMR_{SEG} = \frac{1}{N} \sum_i NMR_{loc}(i), \text{ [дБ]}. \quad (3)$$

Негативная величина оценок  $NMR_{total}$  или  $NMR_{SEG}$  показывает оценку нижней границы порога восприятия, а позитивная величина данных значений является оценкой энергии воспринимаемых искажений.

## Заключение

Результаты экспериментальных исследований (возможная минимальная скорость передачи (кбит/с), максимальная степень компрессии, среднее количество бит на отсчет входного аудио сигнала, объективные оценки качества:  $NMR_{total}$ ,  $NMR_{SEG}$ ,  $SB$ ,  $MF$ ) перцептуального ПДВП-кодера аудио сигналов для разных аудио сигналов (шумовых, тональных, ритмичных) в сравнении с популярным аудио кодером MPEG-1 уровень III, минимальная скорость передачи которого 64 кбит/с, представлены в табл. 1. Анализ результатов показал, что минимальная скорость передачи для ПДВП-кодера аудио сигналов меняется от 36 кбит/с для тональных, ритмичных аудио сигналов до 45 кбит/с для шумовых сигналов. При этом оценка  $NMR_{total}$  меняется от -6.495 дБ до -8,857 дБ, т.е. граница порога восприятия минимум в два раза ниже порога маскирования. Для сравнения у MPEG-1 данная оценка позитивная, т.е. энергия воспринимаемых шумов не менее чем в два раза больше порога маскирования и, соответственно, практически каждый фрейм реконструированного сигнала содержит искажения,  $MF_{max} = 99,78 \%$ . Это обусловлено тем, что ширина частотной полосы  $SB$  восстановленного сигнала достаточно мала  $SB_{min} = 11,89$  кГц, против 22,05 кГц в оригинальном сигнале.

Таблица 1

Объективные оценки перцептуального ПДВП-аудио кодера и MPEG-1 уровень III

Тестовый материал $f_s = 44,1$ кГц, 16 бит РСМ	Тип кодера	Мин. скорость передачи, [кбит/с]	Макс. степень компрессии	Среднее количество, [бит/отсчет]	$SB$ , [кГц]	$NMR_{total}$ , [дБ]	$NMR_{SEG}$ , [дБ]	$MF$ , %
"Boney-M" 15 сек.	ПДВП-кодер	45	15.68	1.02	18.4 9	-8.297	-0.8699	27.0 4
	MPEG-1	64	11.02	1.45	15.1 3	4.430	0.4076	99.8 5
"Classic" 10 сек.	ПДВП-кодер	42	16.8	0.95	17.6 2	-6.495	-0.7987	43.4 5
	MPEG-1	64	11.02	1.45	12.5 8	3.284	0.3145	99.7 8
"Queen" 10 сек	ПДВП-кодер	41	17.2	0.92	16.9 3	-7.597	-0.7765	40.3 9
	MPEG-1	64	11.02	1.45	11.8 9	4.236	0.3994	97.8 0
"ABBA" 6 сек.	ПДВП-кодер	36	19.5	0.81	18.7 2	-8.857	-0.9026	20.6 5
	MPEG-1	64	11.02	1.45	15.2 0	3.740	0.3585	99.6 3

## Литература

1. Brandenburg K. and Sporer T. "NMR" and "masking flag": Evaluation of quality using perceptual criteria // The Proc. of the 11<sup>th</sup> Int. Conf. AES, "Test and measurement". – Portland, USA, May 1992. – pp. 169-179.
2. Colomes C., Schmidmer C., et.al. Perceptual quality assessment for digital audio: PEAQ-the new standard for objective measurement of the perceptual audio quality // The Proc. of the AES 17<sup>th</sup> Intern. Conference. – Florence, Italy, 1999. – pp. 337-351.
3. Petrovsky Al., Petrovsky A.A., Audio coding with a masking threshold adapted wavelet packet based on run-time reconfigurable processor architecture // AES-110<sup>th</sup> Convention; № 5298 – Amsterdam, Netherlands, May 2001. – 8 p.
4. Петровский Ал.А. Процессор на базе ЦПОС/ПЛИС для кодирования аудио сигналов в реальном времени // Труды 4-й Международной конференции "Цифровая обработка сигналов и её применение (DSPA'2002)". – Москва, Россия, 2002. – т.2. – с.453-456.

**OBJECTIVE QUALITY EVALUATION OF RECONSTRUCTED AUDIO SIGNAL OF PERCEPTUAL WP-ENCODER BASED ON THE PERIPHERAL MODEL OF HUMAN EAR**

Petrovsky A.

Belorussian state University of Informatics and Radioelectronics  
 6, P.Brovky st., Minsk, Belarus  
 e-mail: [apetrows@bas-net.by](mailto:apetrows@bas-net.by)

**Introduction**

Traditionally, the audio signals objective quality tests are based on the relation of signal-to-noise ration, but they do not show a real objective situation of the reconstructed audio signal quality. The standard of the quality measurements assurance for perceptual audio coder is PEAQ – Perceptual Evaluation of Audio Quality, its schema is based on peripheral model of human ear. It is used for the estimation the signal quality in the perceptual audio coder based on adaptive wavelet packet (WP).

**Objectives quality of the reconstructed audio signals**

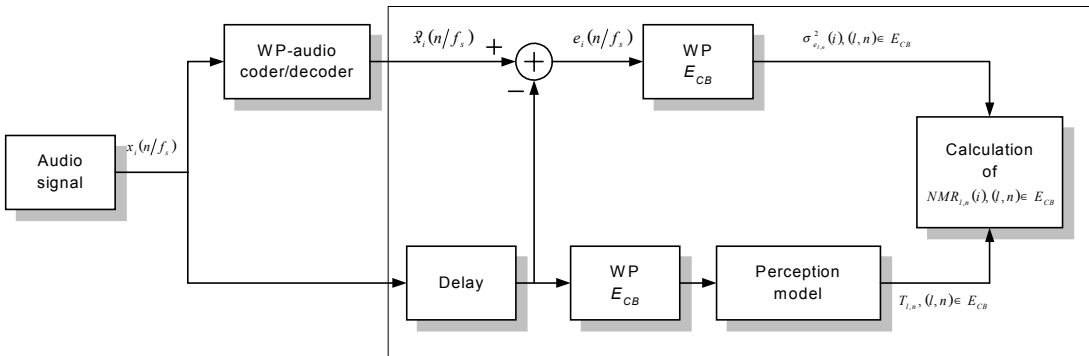


Fig. 4  $NMR_{l,n}(i), (l,n) \in E_{CB}$  calculation

The reconstructed signal quality estimations are based on several parameters: noise-to-mask ratio which estimates the distance between an audible and non-audible distortions; the relative number of the distorted frames to the number of all frames, which gives the probability estimation of the distortions being kept in the reconstructed signal; signal frequency bandwidth.

The  $NMR$  estimation for  $i$ -th frame of audio signal  $x_i(n/f_s)$  is calculated as shown on (рис. 3). The reconstructed signal error  $e_i(n/f_s) = \hat{x}_i(n/f_s) - x_i(n/f_s)$  and its dispersion in each critical subband  $\sigma^2_{e_{l,n}}(i), (l,n) \in E_{CB}$  are defined. In parallel with the original signal the masking thresholds  $T_{l,n}(i), (l,n) \in E_{CB}$  are calculated by procedure of the masking thresholds computation in wavelet-domain. The  $i$ -th frame estimations could be found from ratios:

$$NMR_{l,n}(i) = \frac{\sigma^2_{e_{l,n}}(i)}{T_{l,n}(i)}, (l,n) \in E_{CB}, NMR_{loc}(i) = 10 \cdot \log_{10} \left( 1/25 \cdot \sum_{\forall (l,n) \in CB-WPD} NMR_{l,n}(i) \right) \text{ [dB]}. \quad (1)$$

The objective indices which characterize the audio signal coding quality for given perceptual audio coder schema are defined by arithmetic mean  $NMR_{loc}(i)$  and geometric mean  $NMR_{SEG}(i)$ :

$$NMR_{total} = 10 \cdot \log_{10} \left( \frac{1}{N} \sum_{i=1}^N 10^{(NMR_{loc}(i)/10)} \right) \text{ [dB]}, NMR_{SEG} = \frac{1}{N} \sum_i NMR_{loc}(i), \text{ [dB]}. \quad (2)$$

The negative value of  $NMR_{total}$  or  $NMR_{SEG}$  shows the lower boundary of perception threshold and positive value of it - the energy estimation of the audible distortions.

**Conclusion**

According to the experimental results the minimal bit-rate for the WP-coder lies between 36-45 kbps with  $NMR_{total}$  changed from  $\approx -6.5$  dB to  $-8,9$  dB. Comparisons with 64 kbps MPEG-1 layer III and its positive  $NMR_{total}$  show that the inserted noise energy is two times more than masking threshold. It means that reconstructed signal still keeps audible distortion. It determines signal bandwidth limitation by 11.89 kHz against 17.53 kHz in WP-coder and 22.05 kHz in the original source.