

ИНТЕГРАЦИЯ МОДЕЛИ ВРЕМЕННОЙ МАСКИРОВКИ В АСИММЕТРИЧНУЮ ПСИХОАКУСТИЧЕСКУЮ СИСТЕМУ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ СЖАТИЯ ЗВУКА

Д.С. Кирпичев, аспирант

Е.И. Маркин, канд. техн. наук, доцент

Пензенский государственный технологический университет
(Россия, г. Пенза)

DOI:10.24412/2500-1000-2026-5-1-332-337

Аннотация. В этой работе рассматривается психоакустическая модель, которая объединяет как частотную, так и временную маскировку для повышения эффективности сжатия звука в реальном времени. Выявлено, что предложенная модель основана на сочетании двух подходов: частотной маскировки с адаптивным коэффициентом асимметрии и функцией плавного перехода и временной маскировки, включающей предварительную и последующую маскировку. Модель была реализована и протестирована на трех типах аудиосигналов. Сделан вывод о том, что модель удаляла от 2,32% до 3,33% энергии сигнала в зависимости от типа сигнала. Максимальное удаление энергии было достигнуто для сигнала, похожего на речь. Степень сжатия варьировалась от 1,02x до 1,03x, было рассчитано соотношение сигнал/шум.

Ключевые слова: психоакустическая модель; частотная маскировка; временная маскировка; сжатие звука; адаптивная асимметрия; обработка в реальном времени; Python.

В настоящее время возможно уменьшить аудиофайлы без значительной потери качества, отбрасывая частоты, не слышимые для среднего слушателя. Это привело к созданию кодеков для различных целей, например сжатия для показа фильмов или сжатия речевого сигнала при общении в Интернете [9]. Для каждого конкретного случая подбирается собственный формат сжатия, который востребован в определенных обстоятельствах, существует возможность выбора между методами с потерями и без потерь. При сжатии без потерь получается тот же файл с невысокой степенью сжатия, которая лишь незначительно отличается от оригинала [14]. Выбор варианта сжатия определяется характером избыточности, имеющейся в данных. Процесс создания алгоритмов сжатия для различных данных разделяется на два этапа. Первый этап носит название моделирования. В результате этого модель описывает любую избыточность, встречающуюся в данных. Следующий этап – это кодирование полученной модели, которое отражает различия в данных с помощью конкретной модели. Данная процедура выполняется с применением двоичного алфавита. Разница между данными и моделью обозначается как отклонение [15].

Алгоритмы сжатия аудио с потерями используют особенности человеческого восприятия звука, а именно его неспособность четко различать все звуковые детали. Ограниченность слуха, проявляющаяся в том, что тихие звуки становятся незаметными на фоне более громких, известна как эффект маскировки. Данное свойство легло в основу методик сокращения психоакустической избыточности, когда часть информации, не воспринимаемой ухом, отбрасывается без ущерба для качества восприятия [10].

Характеристики слуховой маскировки зависят от спектрального состава и временных параметров как маскирующего, так и маскируемого сигналов. По этим признакам эффекты делятся на два основных типа:

- 1) частотная (одновременная) маскировка;
- 2) временная (неодновременная) маскировка.

В частотной области маскировка возникает из-за того, что при наличии мощного звука ухо теряет чувствительность к звукам малой интенсивности на близких частотах. Иными словами, если два сигнала звучат одновременно в одной полосе частот, более слабый из них становится неразличимым на фоне более сильного.

Временная маскировка отражает динамику работы слуха, демонстрируя, как меняется порог слышимости одного сигнала в присутствии другого, когда эти сигналы не звучат одновременно. Здесь выделяют два явления: постмаскировку – изменение порога слышимости после завершения мощного сигнала, и предварительную маскировку – изменение порога слышимости до начала мощного сигнала. Например, слабый звук может быть не слышен за 5-20 миллисекунд до появления маскирующего сигнала и становится различимым только через 50-200 миллисекунд после его начала.

Новые системы сжатия звука во многих приложениях, работающих на основе психоакустических моделей, эти модели имитируют слуховую систему человека, которая может маскировать неслышимые звуки или фильтровать диапазон частот [1]. В патенте [2] изучалась только частотная маскировка, игнорируя временную маскировку (предварительную маскировку и последующую маскировку).

Научная новизна заключается в коэффициенте адаптивной асимметрии, который изменяется в зависимости от частоты и уровня маскировки, а также при предварительной маскировке (на определенный период времени) и после маскировки (на определенный период времени).

В области интеграции модели временной маскировки в асимметричную психоакустическую систему для повышения эффективности сжатия звука предлагается метод неравномерного (асимметричного) сжатия, вдохновлённый свойством слуховой системы, где динамическое сжатие выше в базальных отделах базилярной мембраны по сравнению с апикальными [13]. Данный метод предусматривает более сильное сжатие в области высоких частот по сравнению с низкими частотами, что позволяет улучшить разборчивость речи в шумных условиях за счёт спектрального усиления низких частот, особенно формант гласных [3]. В [4] анализируется статическое логарифмическое сжатие, используемое в MFCC, и на основе свойств слуховой системы предлагается метод асимметричного сжатия, при котором высокие частоты сжимаются сильнее, чем низкие. Предложенный метод позволяет улучшить производительность автоматического распознавания речи и разборчивость в шумных условиях за счёт спектрального усиления низкочастотных областей, в частности формант гласных.

Математическая модель временной маскировки [5-7].

Первый случай: Предмаскировка, для времен $t < t_m$ (до момента маскера):

$$M_{\text{Пред}}(t, L_m) = L_m - 25 + 0.2L_m - \frac{20}{\Delta t + 1}, \quad \Delta t \leq 20 \text{ мс} \quad (1)$$

где: t_m — момент времени возникновения маскера, $\Delta t = |t_m - t|$ временное расстояние в миллисекундах, L_m — уровень маскера в дБ и M_{pre} — порог маскировки в дБ.

Второй случай: Послемаскировка, для времен $t > t_m$ (после момента маскера):

$$M_{\text{После}}(t, L_m) = L_m - 25 + 0.2L_m - \frac{30}{\Delta t + 1}, \quad \Delta t \leq 200 \text{ мс} \quad (2)$$

Объединённая временная маскировка. Итоговый порог временной маскировки:

$$T_{temp}(t) = \max(M_{\text{Пред}}(t, L_m), M_{\text{После}}(t, L_m), АНТ) \quad (3)$$

Где АНТ — абсолютный порог слышимости [6].

Интеграция частотной и временной маскировки. Объединённый порог

$$T_{\text{общ}}(f, t) = 10 \log_{10} \left(10^{\frac{M_{\text{част}}(f)}{10}} + 10^{\frac{T_{\text{врем}}(t)}{10}} \right) \quad (4)$$

Альтернативный метод

$$T_{\text{общ}}(f, t) = \max(M_{\text{част}}(f), T_{\text{врем}}(t)) \quad (5)$$

Правило принятия решения: Звуковой компонент на частоте f и в момент времени t является неслышимым и может быть удалён или грубо проквантован [7], если:

$$L_s(f, t) < T_{\text{общ}}(f, t) + \beta \quad (6)$$

$L_s(f, t)$ — уровень сигнала в дБ и $\beta = 5$ дБ — запас надёжности.

Результаты. Предложенная психоакустическая модель была протестирована на трёх типах звуковых сигналов:

1. Импульс + тихий тон (1000 Гц)
2. Удар барабана + послезвучие
3. Речеподобный сигнал (взрывной + гласный)

Модель удалила от 2.32% до 3.33% энергии сигнала. Наибольшее удаление: 3.33% для ре-

чеподобный сигнал и 2.32% для импульс + тихий тоню.

Коэффициент сжатия составил 1.02–1.03х. Модель смогла сжать сигнал на 2-3% от исходного размера. Для коротких тестовых сигналов (0.5 секунды) это хороший результат. Для более длинных реальных сигналов (музыка, речь) ожидается большее сжатие.

Сравнение результатов показано в таблице 1.

Таблица 1. Сравнение результатов типов сигналов [8; 11]

Тип сигнала	Удалённая энергия	SNR (дБ)	Коэффициент сжатия
Импульс + тихий тон	2.32%	16.35	1.02
Удар барабана + послезвучие	2.41%	16.17	1.02
Речеподобный сигнал	3.33%	14.77	1.03

Адаптивный коэффициент асимметрии, временная маскировка: предмаскировка и послемаскировка и другие результаты показаны на рисунках 1-3.

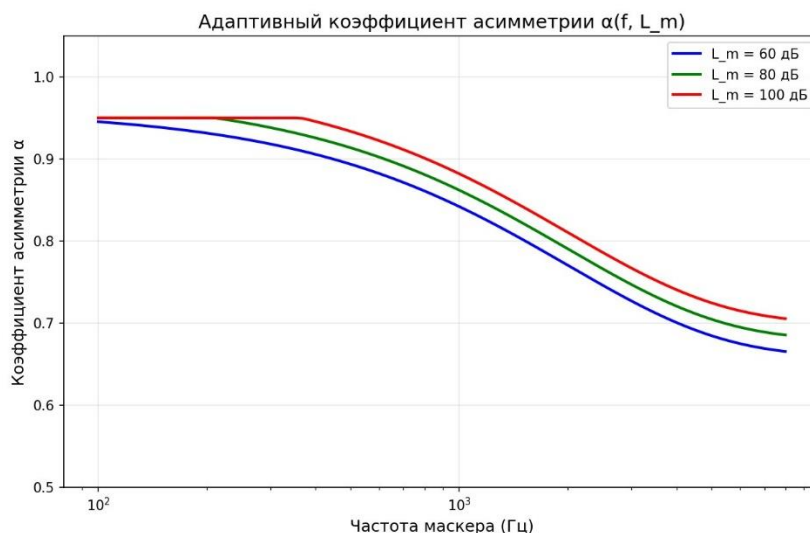


Рис. 1. Адаптивный коэффициент асимметрии

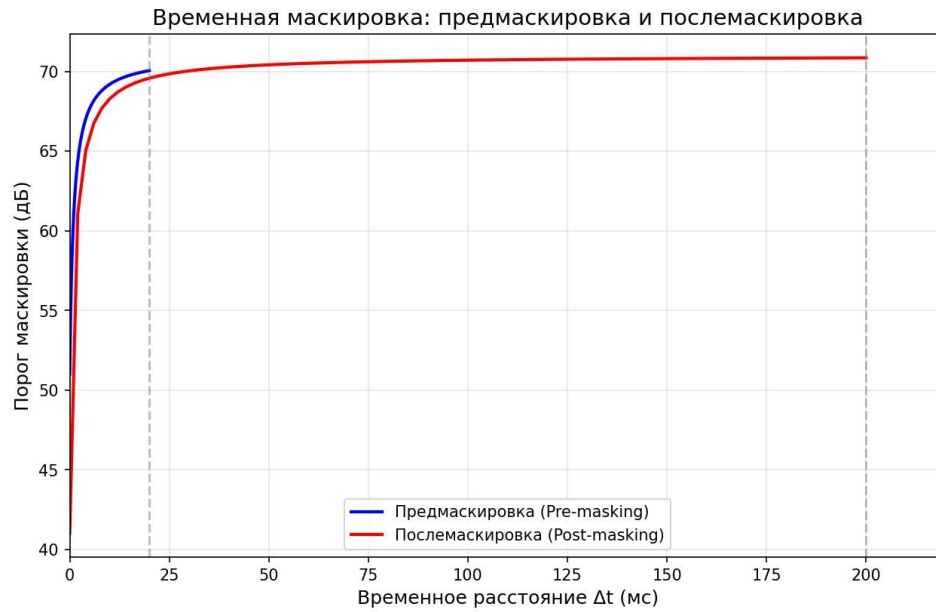


Рис. 2. Временная маскировка: предмаскировка и послемаскировка

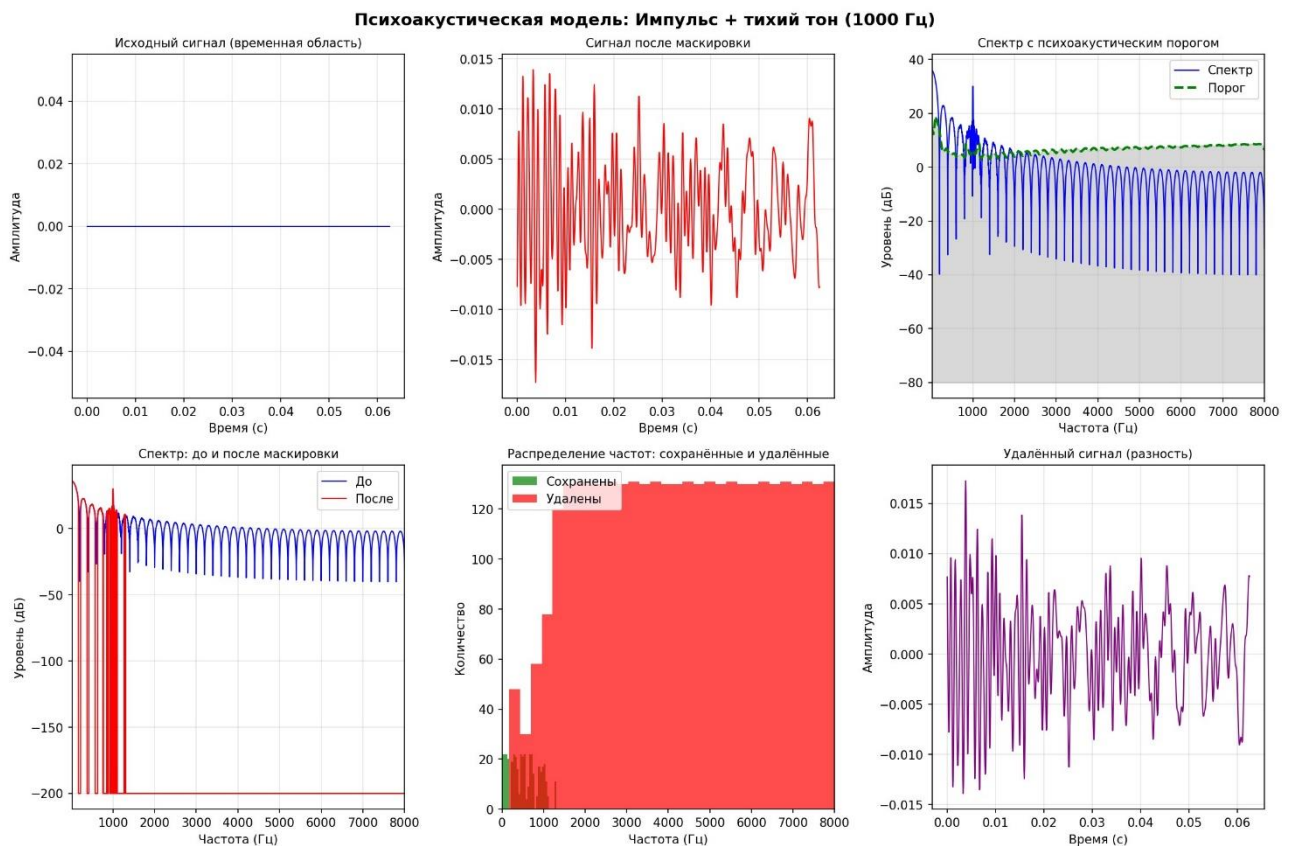


Рис. 3. Другие результаты

Предложенная модель успешно удаляет неслышимые частоты из аудиосигналов, сохраняя при этом качество звука. Разработанная

модель подходит для приложений сжатия звука в реальном времени на различных устройствах.

Библиографический список

1. Заболотов В.А., Стефанова И.А. Сжатие аудиоданных на основе психоакустических свойств слуха человека // Естественные и математические науки в современном мире. 2016. №6 (41). С. 43-51.

2. Ковалгин Ю.А., Фадеев Д.Р. Исследование психоакустических моделей кодеков с компрессией цифровых аудиоданных // Современная наука: актуальные проблемы теории и практики, 2016. №7. 2016. С. 29-39.

3. Патент № 2801156 С2 Российская Федерация, МПК G10L 21/0232. Система компандирования и способ для снижения шума квантования с использованием усовершенствованного спектрального расширения: № 2020100016: заявл. 08.01.2020: опубл. 02.08.2023 / П. Хеделин, А. Бисвас, М. Шуг, В. Мелкоте; заявитель ДОЛБИ ЛАБОРАТОРИС ЛАЙСЭНЗИН КОРПОРЕЙШН, ДОЛБИ ИНТЕРНЕСНЛ АБ. – EDN ERECLK.

4. Фадеев Д.Р. Исследование эффективности психоакустических моделей в алгоритмах перцепционного сжатия звуковых сигналов / Д.Р. Фадеев // Приоритетные направления развития науки и технологий: Тезисы докладов XVIII Международной научно-технической конференции, Тула, 20 ноября 2015 года / Под общей редакцией В.М. Панарина. – Тула: Издательство «Инновационные технологии», 2015. – С. 70-72. – EDN VCBTQP.

5. Deshotels L. Inaudible sound as a covert channel in mobile devices // 8th USENIX Workshop on Offensive Technologies (WOOT 14). – 2014. – 9 с.

6. Fereczkowski M. et al. Amplitude compression for preventing rollover at above-conversational speech levels // Trends in Hearing. – 2024. – Т. 28. – С. 23312165231224597.

7. Haque S., Togneri R., Zaknich A. An auditory motivated asymmetric compression technique for speech recognition // IEEE transactions on audio, speech, and language processing. – 2011. – Т. 19. – № 7. – С. 2111-2124.

8. Herre J., Dick S. Psychoacoustic Models for Perceptual Audio Coding—A Tutorial Review // Appl. Sci. 2019, 9, 2854. – [Электронный ресурс]. – Режим доступа: <https://doi.org/10.3390/app9142854>.

9. Jesteadt W., Bacon S. P., Lehman J. R. Forward masking as a function of frequency, masker level, and signal delay // The Journal of the Acoustical Society of America. – 1982. – Т. 71. № 4. – С. 950-962.

10. Kohlrausch A., van de Par S. and Their Role in Hearing Research // Acoustics, Information, and Communication: Memorial Volume in Honor of Manfred R. Schroeder. – 2014. – С. 113.

11. Luthfi F., Erwin S. Data audio compression lossless FLAC format to lossy audio MP3 format with Huffman Shift Coding algorithm // 2016 4th Intern. Conf. on Information and Communication Technology (ICoICT). – Bandung, 2016. – P. 1-5.

12. Moore B.C., Glasberg B.R., Baer T. A model for the prediction of thresholds, loudness, and partial loudness // Journal of the Audio Engineering Society. – 1997. – Vol. 45, № 4. – Pp. 224-240.

13. Nowak N., Zabierowski W. Methods of Sound Data Compression – Comparison of Different Standards // Радиоэлектроника и информатика. – 2011. – № 4.

14. Rupali B., Patil Dr., Kulat K.D. Audio compression using dynamic Huffman and RLE coding // 2017 2nd Intern. Conf. on Communication and Electronics Systems (ICCES). – Coimbatore, 2017. – P. 160-162.

15. Yang W. Enhanced modified bark spectral distortion (EMBSD): An objective speech quality measure based on audible distortion and cognition model. – Temple University, 1999.

**INTEGRATION OF THE TIME MASKING MODEL INTO AN ASYMMETRIC
PSYCHOACOUSTIC SYSTEM TO INCREASE THE EFFICIENCY
OF SOUND COMPRESSION**

D.S. Kirpichev, *Postgraduate Student*

E.I. Markin, *Candidate of Technical Sciences, Associate Professor, Department of Programming
Penza State Technological University, Graduate Student
(Russia, Penza)*

***Abstract.** This paper presents a psychoacoustic model that combines both frequency and time masking to improve the efficiency of sound compression in real time applications. The proposed model is based on a combination of two approaches: frequency masking with an adaptive coefficient of asymmetry and a smooth masking transition function and temporary masking, including preliminary and subsequent masking. The model was implemented and tested on three types of audio signals. It was concluded that the model removed from 2.32% to 3.33% of the signal energy, depending on the type of signal. The maximum energy removal was achieved for a speech-like signal. The compression ratio ranged from 1.02x to 1.03x, and the signal-to-noise ratio was calculated.*

***Keywords:** psychoacoustic model; frequency masking; temporal masking; audio compression; adaptive asymmetry; real-time processing; Python.*