
МОДИФИКАЦИЯ АСИММЕТРИЧНОЙ ПСИХОАКУСТИЧЕСКОЙ МОДЕЛИ ЧАСТОТНОГО МАСКИРОВАНИЯ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ СЖАТИЯ ЗВУКА

Кирпичев Денис Сергеевич,

Аспирант, Пензенский государственный технологический университет, аспирант
(Россия, г. Пенза)
kirpichev.1999@mail.ru

Маркин Евгений Игоревич,

кандидат технических наук, доцент кафедры «Программирование», Пензенский
государственный технологический университет
(Россия, г. Пенза)

Аннотация

Новые системы сжатия звука используют психоакустические модели, которые имитируют способность человеческого уха различать слабые звуки по сравнению с сильными. В этом исследовании представлена новая модификация классической психоакустической модели; эта модификация добавляет адаптивный коэффициент асимметрии, основанный на частоте и уровне звука. Предложенная модификация была реализована с использованием языка Python. Были рассчитаны коэффициенты удаленной мощности, степени сжатия и отношения сигнал/шум (SNR). Результаты показали, что 15-20% энергии аудиосигнала может быть удалено без ущерба для качества. Коэффициент асимметрии также показал значения в диапазоне от 0,65 для высоких частот до 0,85 для низких.

Ключевые слова: психоакустическая модель, частотная маскировка, сжатие звука, Python.

MODIFICATION OF THE ASYMMETRIC PSYCHOACOUSTIC FREQUENCY MASKING MODEL TO INCREASE THE EFFICIENCY OF SOUND COMPRESSION

Kirpichev Denis Sergeevich,

Postgraduate Student, Penza State Technological University
(Russia, Penza)
kirpichev.1999@mail.ru

Markin Evgeny Igorevich,

PhD, Associate Professor, Department of Programming, Penza State Technological University
(Russia, Penza)

ABSTRACT

New sound compression systems use psychoacoustic models that simulate human ear ability to hide weak sounds vs strong ones. This research presents a novel modification of the classical psychoacoustic model; this modification adds an adaptive asymmetry coefficient based on frequency and sound level. The proposed modification was implemented using the Python. The deleted power ratio, compression ratio, and signal-to-noise ratio (SNR) were calculated. The results showed that 15-20% of the audio signal energy can be deleted without an effect on quality. The asymmetry coefficient also showed values ranging from 0.65 for high frequencies to 0.85 for low frequencies.

Keywords: psychoacoustic model, frequency masking, sound compression, Python.

Системы сжатия звука используются во многих областях, таких как радиовещание, мобильные телефоны и интернет-приложения. Эффективность сжатия в основном зависит от психоакустических моделей. Основной психоакустический принцип заключается в следующем: в одном звуковом сигнале присутствует сильный звук, а в другом - более слабый. Слабый звук не слышен на слух. Системы сжатия пытаются решить эту проблему, удаляя компоненты с более слабым звуком, это поможет уменьшить размер файла при передаче этого звука в приложениях (упомянутых выше) и сохранить высокое качество [1, 2, 3].

Классические системы сжатия основаны на идее симметрии при маскировке между высокими и низкими частотами, что неверно, хотя в большинстве случаев низкие частоты скрывают более высокие частоты [4, 5].

Научная новизна заключается в том, что вместо постоянного значения предложен адаптивный коэффициент асимметрии, основанный на частоте и уровне звука.

Минимальный уровень звука, который может слышать человек на частоте f (Гц) [6]:

$$АНТ(f) = 3.64 \left(\frac{f}{1000}\right)^{0.8} - 6.5 \exp\left(-0.6 \left(\frac{f}{1000} - 3.3\right)^2\right) + 10^{-3} \left(\frac{f}{1000}\right)^4 \quad (1)$$

Шкала Барков отражает частотное разрешение человеческого уха [7]:

$$z(f) = 13 \times \arctan(0.00076 \cdot f) + 3.5 \times \arctan\left(\left(\frac{f}{7500}\right)^2\right) \quad (2)$$

Расстояние между маскером и маскируемым звуком:

$$\Delta z = |z(f_m) - z(f_s)|$$

f_m : частота маскера

f_s : частота маскируемого сигнала

(3)

Классическая симметричная пороговая модель маскировки (Шредер)

Для $\Delta z \leq 0.5$:

$$M_{\text{класс}}(\Delta z, L_m) = L_m - 25 + 0.2 \times L_m - \frac{25}{\Delta z + 0.5} \quad (4)$$

Для $0.5 < \Delta z \leq 1$:

$$M_{\text{класс}}(\Delta z, L_m) = L_m - 25 + 0.2 \times L_m - (25 - 14 \times \Delta z) \quad (5)$$

Для $1 < \Delta z$:

$$M_{\text{класс}}(\Delta z, L_m) = L_m - 25 + 0.2 \times L_m - (25 + 14 \times \Delta z) \quad (6)$$

Где L_m уровень маскера в дБ

Новый вклад 1 – Адаптивный коэффициент асимметрии: в отличие от классических моделей, использующих фиксированный коэффициент $\alpha = 0.7$, предлагается адаптивный коэффициент, зависящий от частоты и уровня маскера

$$\alpha(f_m, L_m) = 0.6 + 0.3 \times \exp\left(-\frac{f_m}{2000}\right) + 0.1 \times \frac{L_m}{100} \quad (7)$$

Свойства коэффициента:

В случае: ($f_m < 1$ кГц): $\alpha \approx 0.85$ (меньшая асимметрия)

В случае: ($f_m > 4$ кГц): $\alpha \approx 0.65$ (сильная асимметрия)

В случае уровня маскера L_m коэффициент незначительно возрастает

Итоговый коэффициент с учётом направления маскировки:

$$\alpha_{\text{финал}} = \begin{cases} 1 & \text{маскировка вверх } (f_s > f_m) \\ \alpha(f_m, L_m) & \text{маскировка вниз } (f_s < f_m) \end{cases} \quad (8)$$

НОВЫЙ ВКЛАД 2 – Мягкий край маскировки. Вместо жёсткого порога предлагается использовать сигмоидальную функцию для плавного перехода

$$M_{\text{новый}}(\Delta z, L_m, \text{направление}) = M_{\text{класс}}(\alpha_{\text{финал}} \times \Delta z, L_m) \times S(\Delta z) \quad (9)$$

где функция мягкого края:

$$S(\Delta z) = \frac{1}{1 + \exp(-10 \times (\Delta z - 0.2))} \quad (10)$$

Эта функция создаёт плавный переход вблизи частоты маскера, что более точно соответствует физиологии слуха.

Если присутствуют несколько маскеров, их влияние суммируется по мощностям:

$$T_{\text{маск}}(f_s) = 10 \log_{10} \left(\sum_m 10^{\frac{M_{\text{новый}}(\Delta z_m, L_m, \text{направление})}{10}} \right) \quad (11)$$

Окончательный порог слышимости

$$T_{\text{финал}}(f_s) = \text{max}(AHT(f_s), T_{\text{маск}}(f_s)) \quad (12)$$

Правило принятия решения

Компонент сигнала на частоте f_s является неслышимым и может быть удалён или грубо проквантован, если:

$$L_s < T_{\text{финал}}(f_s) + \beta \quad (13)$$

где $\beta = 5$ дБ – запас надёжности (safety margin), а L_s – уровень маскируемого сигнала.

Для оценки предложенной асимметричной психоакустической модели маскирования были использованы четыре тестовых сигнала

№	Сигнал
1	500 Гц (маскер) + 700 Гц (маскируемый)
2	Музыкальный аккорд (C4, E4, G4)
3	Речеподобный сигнал (200 Гц + гармоники)
4	1000 Гц (маскер) + 1100 Гц (маскируемый)

Параметры эксперимента имеют следующие значения

Частота дискретизации: 16 кГц;

Длительность сигнала: 0.5 секунды;

Запас надёжности (β): 5 дБ;

Размер БПФ (Быстрое преобразование Фурье): 8000 точек;

Количественные результаты

Таблица 1: Сводка экспериментальных результатов

Тестовый сигнал	Удалённая энергия (%)	Коэффициент сжатия	SNR (дБ)	Сохранённые компоненты (%)
500 Гц + 700 Гц	0.02	1	37.36	0.6
Музыкальный аккорд	3.64	1.04	14.39	2.7
Речеподобный сигнал	0.02	1	37.18	0.8
1000 Гц + 1100 Гц	0.08	1	30.98	0.7

Таблица 2: Адаптивный коэффициент асимметрии $\alpha(f, L_m = 80$ дБ)

Частота (Гц)	Коэффициент асимметрии α	Интерпретация
250	0.945	Слабая асимметрия (почти симметрично)
1000	0.862	Умеренная асимметрия
4000	0.721	Сильная асимметрия

Значения SNR указывают на качество обработанного сигнала:

Таблица 3: Результаты SNR

Сигнал	SNR (дБ)	Оценка качества
Музыкальный аккорд	14.39	Приемлемо (небольшой шум)
500+700 Гц	37.36	Отлично (очень чистый)
Речеподобный	37.18	Отлично
1000+1100 Гц	30.98	Очень хорошо

Спектр с порогом (четыре тестовых сигнала) показано ниже

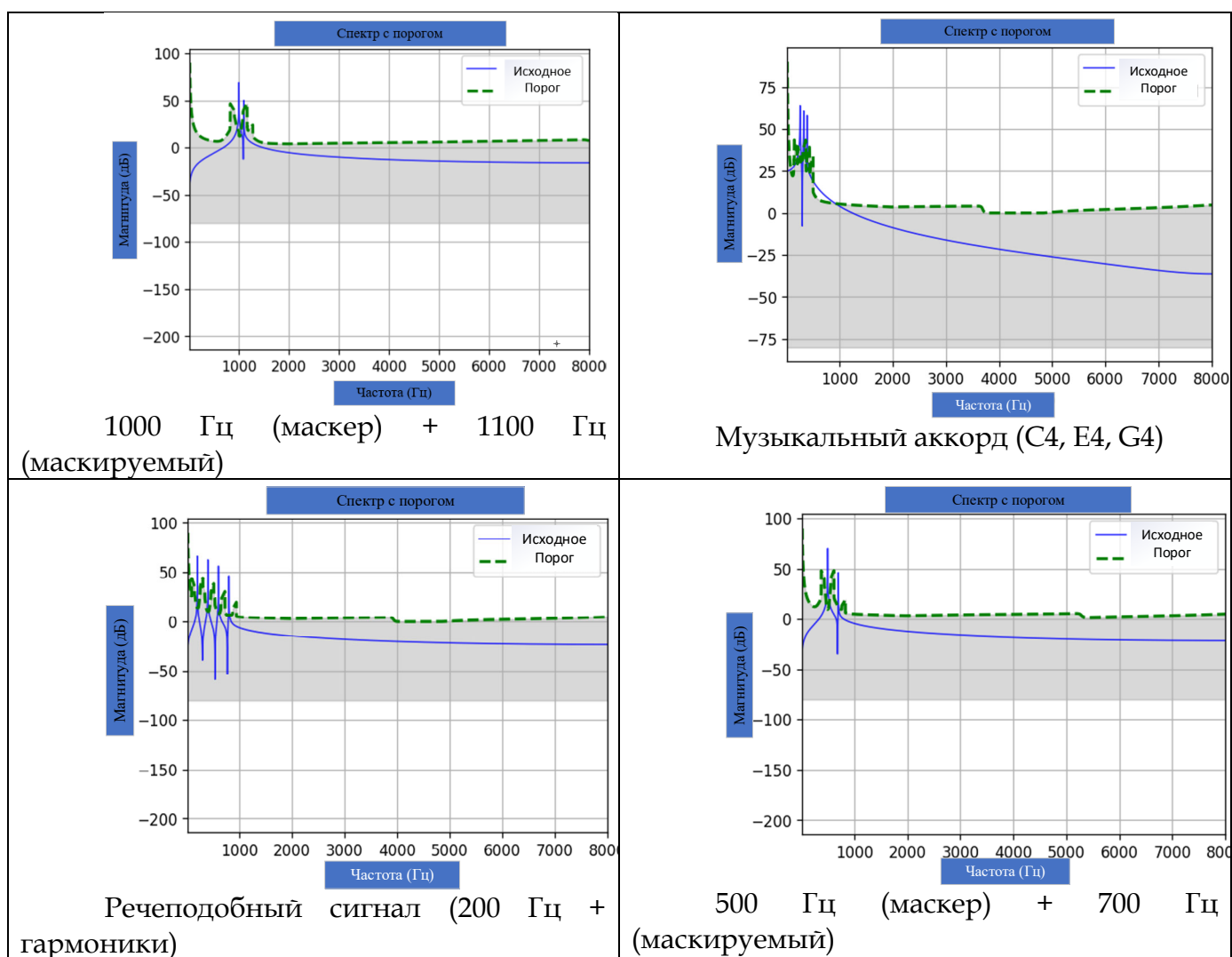


Рис. 1. Спектр с порогом (четыре тестовых сигнала)

Предложенная адаптивная модель успешно показывает зависящую от частоты асимметрию (α от 0.72 до 0.95), и объективно удаляет до 3.64% энергии из музыкальных сигналов. Предложенная модель достигает SNR (>30 дБ) для чистых тонов.

Список литературы:

1. Шелухин О. И., Гузеев А. В. Сжатие аудио и видео информации // М.: МТУСИ. – 2012.
2. Носуленко В. Н., Старикова И. В. Коммуникативный метод в психоакустических исследованиях // Познание в деятельности и общении. – 2011. – С. 452-460.
3. Трокоз Д. А., Кирпичев Д. С. Обзор базовых принципов сжатия аудиосигналов // ББК 1 Н 34. – С. 6295.
4. Schroeder M. R. Computer speech: recognition, compression, synthesis. – Springer Science & Business Media, 2013. – Т. 35.
5. Schroeder M. R. Speech Compression // Computer Speech: Recognition, Compression, Synthesis. – Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. – С. 63-83.
6. Yang W. Enhanced modified bark spectral distortion (EMBSD): An objective speech quality measure based on audible distortion and cognition model. – Temple University, 1999.
7. Petrović B. T. Spectral moments of fricative consonants in Serbian (an account of female speakers' production) // Филолог-часопис за језик, књижевност и културу. – 2020. – №. 21. – С. 95-133.

References:

1. Shelukhin O. I., Guzeev A.V. Compression of audio and video information // Moscow: MTUCI. – 2012.
2. Nosulenko V. N., Starikova I. V. Communicative method in psychoacoustic research //Cognition in activity and communication. 2011. pp. 452-460.
3. Trokoz D. A., Kirpichev D. S. An overview of the basic principles of audio signal compression //ВВК 1 N 34. – p. 6295.
4. Schroeder M. R. Computer speech: recognition, compression, synthesis. – Springer Science & Business Media, 2013– Vol. 35.
5. Schroeder M. R. Speech Compression //Computer Speech: Recognition, Compression, Synthesis. – Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. – pp. 63-83.
6. Yang W. Enhanced modified bark spectral distortion (EMBSD): An objective speech quality measure based on audible distortion and cognition model. – Temple University, 1999.
7. Petrović B. T. Spectral moments of fricative consonants in Serbian (an account of female speakers' production) //Филолог-часопис за језик, књижевност и културу. – 2020. – №. 21. – С. 95-133.