

УДК 004.415

Акылбек уулу Жусуп, zhusupakylbek@gmail.com

Корякин Сергей Владимирович, srgkoryakin1@gmail.com

Кыргызско-Германский институт прикладной информатики

МЕТОДЫ РАЗРАБОТКИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ПСИХОАКУСТИЧЕСКОГО ПРОФИЛИРОВАНИЯ И СИСТЕМАТИЗАЦИИ ЗВУКОВЫХ СРЕД НА ОСНОВЕ ОТКРЫТЫХ АУДИОДАТАСЕТОВ

Аннотация. В статье рассматриваются подходы к разработке программного обеспечения для автоматизированного психоакустического анализа и профилирования звуковых сред. Обосновывается необходимость перехода от физических характеристик звука к перцептивным параметрам, отражающим субъективное восприятие громкости, резкости, шероховатости и флуктуационной силы. Предложен метод систематизации звуковых данных на основе открытых аудиодатасетов с целью формирования психоакустических профилей слухового восприятия человека в зависимости от типа звукового окружения (городское, сельское, промышленное). Кратко затрагиваются физиологические факторы — курение и употребление алкоголя — как примеры изменений слуховых порогов, результаты которых могут быть воспроизведены и смоделированы в рамках создаваемого программного прототипа. Реализация модели на Python позволяет визуализировать закономерности изменения слухового восприятия в разных акустических условиях.

Ключевые слова: психоакустика, звуковая среда, аудиодатасет, профилирование, громкость, резкость, Bark-шкала, программное обеспечение

Введение

Каждый человек ежедневно подвергается воздействию множества звуков — от шумов улиц и транспорта до природных и бытовых звуков. Звуковая среда, в которой человек живёт и развивается, формирует не только привычку к определённым акустическим стимулам, но и влияет на особенности слухового восприятия. Исследования в области психоакустики показывают, что окружающая звуковая атмосфера способна изменять чувствительность слуха, смещая акценты в восприятии частотного диапазона [1]. Также физиологические особенности, включая вредные привычки — курение и употребление алкоголя, — оказывают влияние, вызывая изменение слуховых порогов и ухудшение восприятия высоких частот [2–3].

Понимание того, как различные типы звуковых сред (городская, сельская, промышленная, офисная и др.) воздействуют на восприятие звука, открывает возможность для профилирования слуха человека и создания психоакустических моделей адаптации. Современные методы анализа звука позволяют количественно описывать не только физические характеристики сигнала, но и перцептивные параметры, отражающие субъективное восприятие — громкость, резкость, шероховатость и флуктуационную силу [4].

Цель исследования

Для выявления закономерностей восприятия в разных средах необходимо объединить психоакустический анализ с систематизацией звуковых данных, полученных из открытых аудиодатасетов, отражающих реальные акустические условия [5]. Разработка программного обеспечения, способного выполнять такую классификацию и профилирование, позволит сформировать базу психоакустических профилей для различных типов окружений и проследить, как человек адаптируется к своему звуковому миру [6].

Методология исследования

Разработка методов психоакустического профилирования звуковых сред основана на комплексном подходе, объединяющем принципы цифровой обработки сигналов, психоакустического моделирования и инженерии программного обеспечения [7]. Методология предполагает создание инструмента, позволяющего формализовать влияние звукового окружения и физиологических факторов на слуховое восприятие и представить результаты анализа в виде количественных показателей и визуальных профилей.

В основе методологии лежит использование открытых аудиодатасетов, классифицирующих звуковые сцены по типам окружения: городская, сельская, природная, индустриальная и бытовая [8].

Первый этап включает предобработку данных — нормализацию, фильтрацию и преобразование сигналов в частотную область (FFT). Для перехода от физических характеристик звука к перцептивным применяется Bark-шкала, отражающая распределение критических полос слуха (рис. 1) [9].

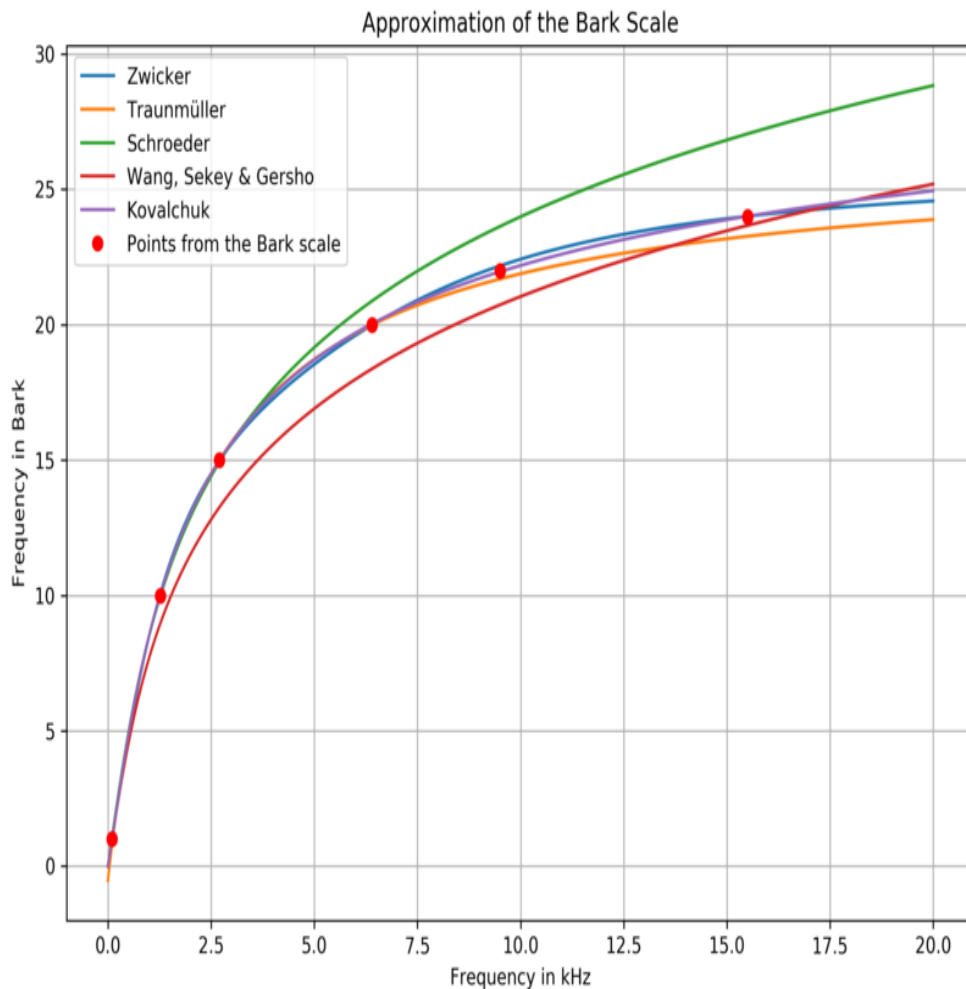


Рисунок 1 – Bark-шкала распределения критических полос слуха (по Zwicker & Fastl, 1999).

Источник: Wikimedia Commons — Bark scale

На втором этапе рассчитываются психоакустические показатели — громкость, резкость, шероховатость и флуктуационная сила — в соответствии с моделью Цвикера, реализованной в стандартах ISO 532-1:2017 и DIN 45692. Для учёта субъективного восприятия применяются кривые равной громкости ISO 226 (рис. 2) [10–12].

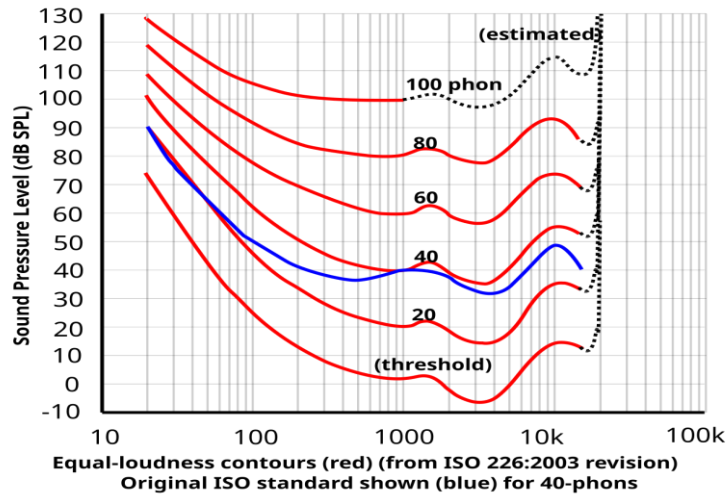


Рисунок 2 — ISO equal-loudness contours with frequency in Hz.

Источник: HyperPhysics (Georgia State University), Wikipedia Equal-loudness contour, Studfile.net [10–12]

Физиологические факторы и их учёт в моделировании

Хотя основной целью работы является создание программных методов психоакустического профилирования звуковых сред, учёт физиологических данных человека позволяет дополнить анализ моделированием слуховых изменений при воздействии внешних факторов. На рисунке 3 представлены данные, отражающие различие слуховых порогов у курящих и некурящих людей. Видно, что курильщики демонстрируют более высокие пороги восприятия на высоких частотах, что свидетельствует о снижении чувствительности слухового анализатора.

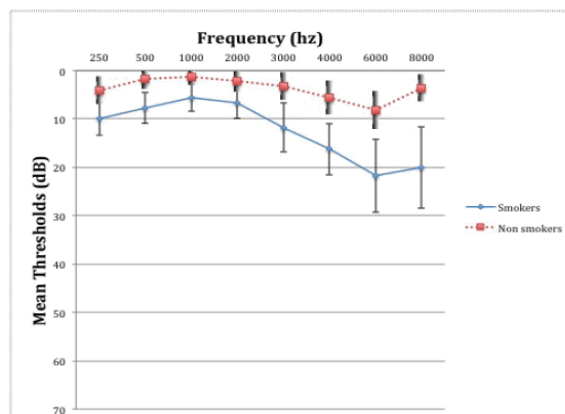


Figure 1. Mean hearing thresholds. Smokers presented higher threshold values than non-smokers especially at high frequency. Error bars represent standard deviation.

Рисунок 3 — Mean hearing thresholds in smokers and non-smokers.

Курильщики показывают повышенные пороги восприятия высоких частот по сравнению с некурящими. Источник: *The International Tinnitus Journal*, 2022 [2]

На рисунке 4 показано изменение порогов слуха до и после употребления алкоголя. Исследования подтверждают, что сразу после приёма алкоголя наблюдается временное снижение чувствительности слуха, особенно в среднечастотной области, однако эффект обратим и исчезает через несколько часов.

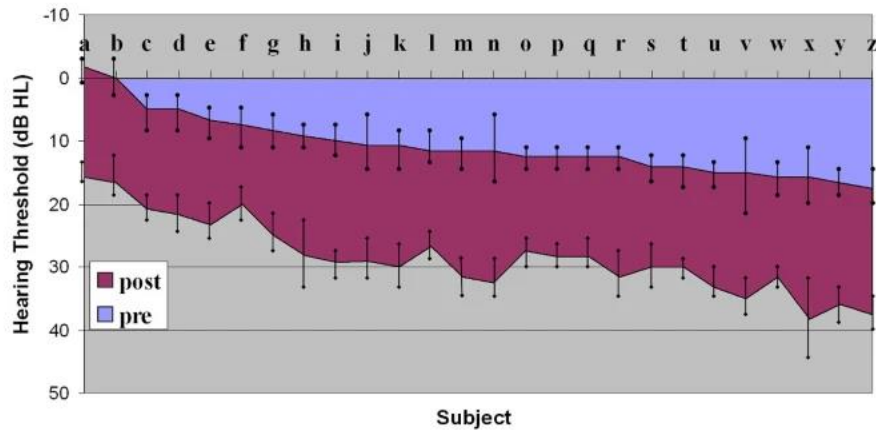


Рисунок 4 — Hearing threshold variation before and after alcohol exposure.

После употребления алкоголя наблюдается временное снижение чувствительности слуха.

Источник: *BMC Ear, Nose and Throat Disorders*, 2007 [3].

Предполагается, что разрабатываемая система психоакустического профилирования в будущем будет способна воссоздавать аналогичные зависимости и моделировать влияние таких факторов на психоакустические параметры человеческого слуха.

Методические основы и расчётные модели

Для оценки психоакустических характеристик различных звуковых сред были разработаны алгоритмы, основанные на фундаментальных моделях восприятия громкости и спектрального анализа [1, 9]. Основой служит представление о равных уровнях громкости (*equal-loudness contours*), описанных в стандартах ISO 226:2003 и ISO 532-1:2017 [4, 13], а также аппроксимация А-взвешивающей кривой (*A-weighting*) как математического приближения реальных кривых равной громкости [10–12].

Перцептивная громкость для каждой полосы частот вычислялась по выражению

$$L_p(f_i) = L_{SPL}(f_i) + A(f_i), \quad (1)$$

где $L_{SPL}(f_i)$ — измеренный (или модельный) уровень звукового давления в децибелах для частоты f_i ;

$A(f_i)$ — корректирующая функция А-взвешивания (дБ), приближающая чувствительность слуха человека в зависимости от частоты [7, 9].

Общий индекс перцептивной громкости (*Perceptual Loudness Index, PLI*) определялся как среднее по всем полосам:

$$PLI = \frac{1}{N} \sum_{i=1}^N 10^{\frac{L_p(f_i)}{10}}, \quad (2)$$

что соответствует энергетическому усреднению с учётом восприятия. Дополнительно рассчитывались вспомогательные психоакустические параметры [9]. Резкость (*Sharpness, S*) характеризуется центроидом спектра:

$$S = \frac{\sum f_i P(f_i)}{\sum P(f_i)}, \quad (3)$$

где $P(f_i)$ — спектральная мощность.

Шероховатость (*Roughness, R*) аппроксимировалась через эксцесс (четвёртый момент) нормализованного распределения мощности, отражающего наличие спектральных пиков и «жесткость» звучания:

$$R = \frac{1}{N} \sum_{i=1}^N \left(\frac{P(f_i) - \bar{P}}{\sigma_P} \right)^4, \quad (4)$$

где \bar{P} — средняя мощность, σ_P — стандартное отклонение мощности.

Флуктуационная сила (*Fluctuation Strength, F*) оценивалась как отношение средних мощностей в низких частотах (40–160 Гц) к средним мощностям в среднечастотной области (500–2000 Гц):

$$F = \frac{\overline{P_{low}}}{\overline{P_{mid}}}, \quad (5)$$

где $\overline{P_{low}}$ и $\overline{P_{mid}}$ — средние значения мощности в указанных диапазонах частот.

Эти параметры вычислялись для каждой среды: Urban, Industrial, Office, Rural, Nature. Источником акустических данных послужили открытые наборы TUT Acoustic Scenes 2017 [6], UrbanSound8K [5] и ESC-50 [8], обеспечивающие классификацию звуковых сцен по типу окружения.

Результаты расчетных моделей

Данные визуализировались в виде диаграмм громкости по частоте (рис. 5), тепловой карты чувствительности (рис. 6) и нормализованных столбчатых графиков сводных индексов (рис. 7–10). Для сравнения выделены частотные профили городского и сельского окружения (рис. 11).

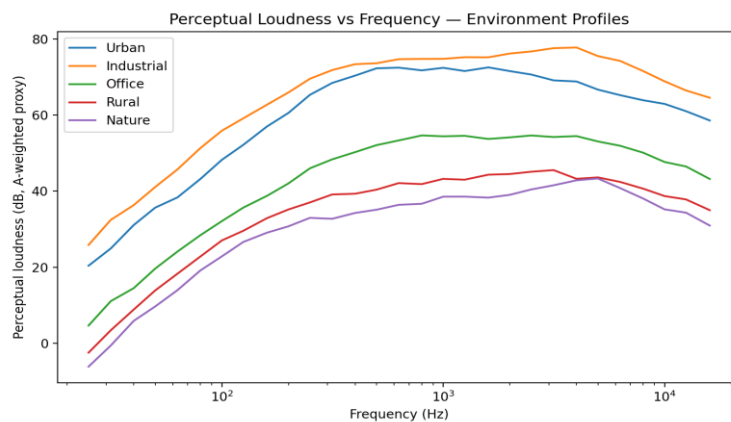


Рисунок 5 – Perceptual loudness vs frequency (A-weighted proxy) for environment classes. Перцептивная громкость в зависимости от частоты (с учетом А-взвешивания) для классов окружения

На рисунке 5 представлены усреднённые перцептивные кривые громкости для различных звуковых сред. Спектры городского и промышленного окружения демонстрируют выраженный максимум в диапазоне 400–2000 Гц, что соответствует характерным шумам двигателей, транспорта, технологического оборудования и сигналов. Эти среды обладают повышенной спектральной плотностью энергии и наибольшим индексом громкости $PLI = (6,7-7,2) \times 10^6$ усл. ед., что свидетельствует о высокой интенсивности акустического воздействия [9, 13, 15]. Офисная среда характеризуется умеренным уровнем энергии с локальным пиком в области 500 Гц, связанным с речевыми и вентиляционными шумами, а также с деятельностью компьютерного оборудования. В сельской и природной среде общий уровень громкости значительно ниже — на 15–25 дБ относительно городских условий. При этом наблюдаются две основные области акустической активности: низкочастотная (100–200 Гц), обусловленная звуками ветра и воды, и высокочастотная (2–6 кГц), соответствующая пению птиц и шуму насекомых.

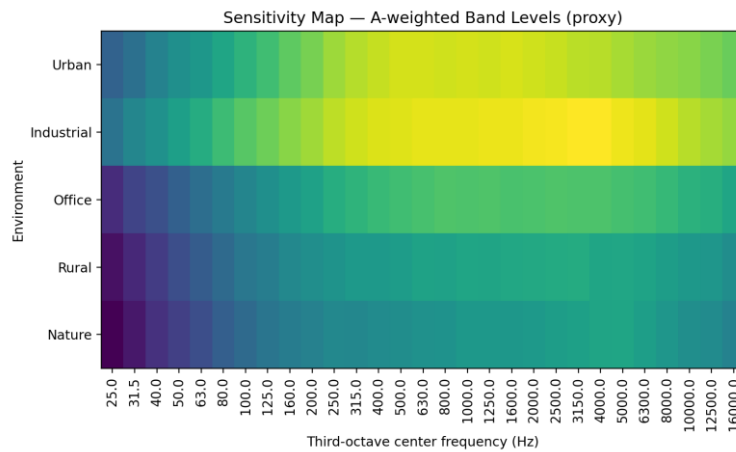


Рисунок 6 — Sensitivity map: A-weighted band levels by environment (third-octave).

Карта чувствительности: уровни полос с А-взвешиванием в зависимости от окружения (третьоктавные полосы)

Тепловая карта на рисунке 6 демонстрирует распределение А-взвешенных уровней по средам и подтверждает, что яркие зоны на частотах 400–1250 Гц характерны для Urban и Industrial, тогда как Rural и Nature показывают максимум в низкочастотном диапазоне, что указывает на акустическую мягкость и природное доминирование низких частот.

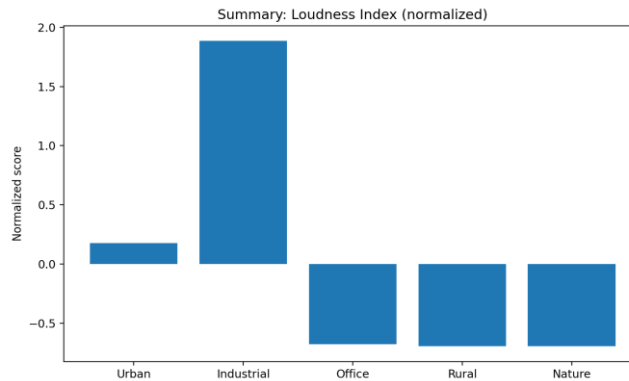


Рисунок 7 – Summary index: Loudness (normalized).

Сводный индекс: громкость (нормированная)

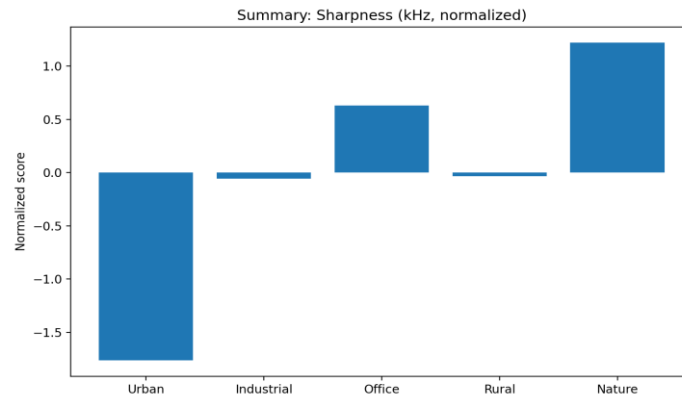


Рисунок 8 – Summary index: Sharpness (normalized, spectral centroid proxy)

Сводный индекс: резкость (нормированная, аппроксимация спектрального центра)

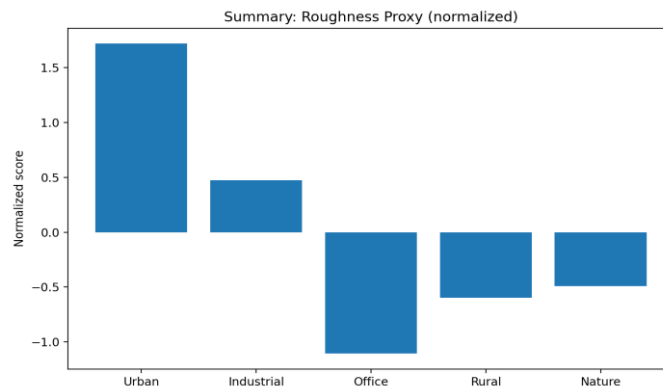


Рисунок 9 – Summary index: Roughness (normalized, kurtosis proxy)

Сводный индекс: шероховатость (нормированная, аппроксимация эксцесса)

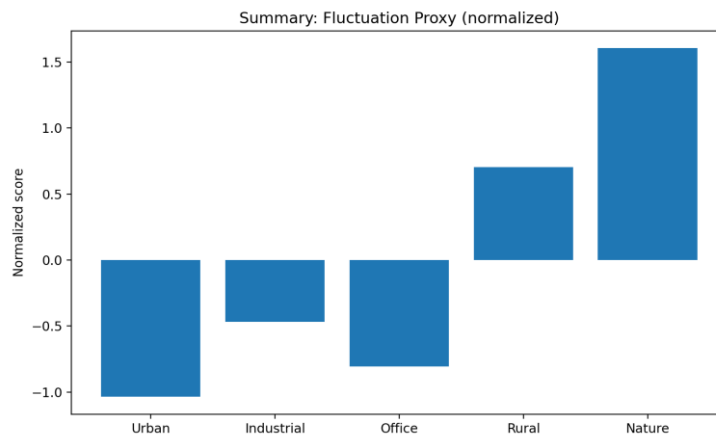


Рисунок 10 – Summary index: Fluctuation (normalized, low/mid energy ratio proxy).

Сводный индекс: флуктуационная сила (нормированная, аппроксимация отношения низко- к среднечастотной энергии)

Сводные показатели психоакустических параметров, представленные на рисунках 7–10, показывают, что индекс громкости (LoudnessIndex) максимален в индустриальной и

городской среде, что подтверждает высокую энергетическую плотность этих сцен. Резкость (Sharpness) оказывается выше в Industrial и Nature, что объясняется присутствием высокочастотных компонентов — металлических резонансов в первой и звуков насекомых во второй. Шероховатость (RoughnessProxu) достигает наибольших значений в Industrial и Urban, отражая непостоянство и импульсный характер шумов, тогда как флуктуационная сила (FluctuationProxu) значительно выше у Rural и Nature, что указывает на доминирование медленных амплитудных колебаний, воспринимаемых человеком как естественная, физиологически комфортная динамика среды.

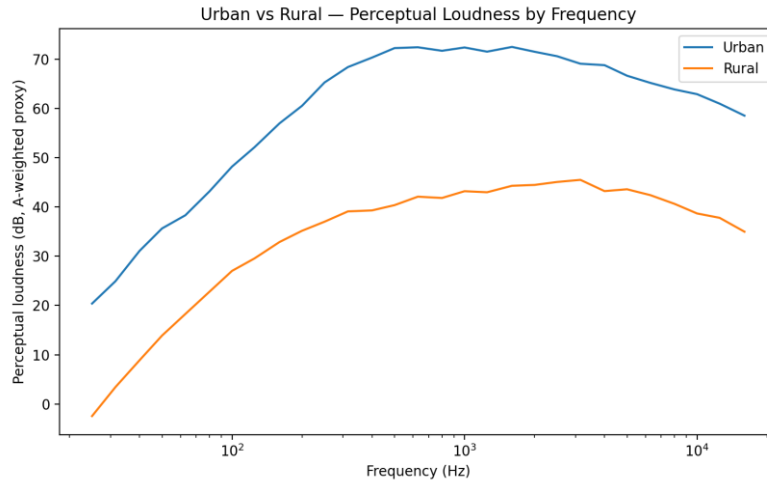


Рисунок 11 – Urban vs Rural: Perceptual loudness by frequency (A-weighted proxy).

Городская против сельской среды: перцептивная громкость в зависимости от частоты (с учетом А-взвешивания)

Диаграмма сравнения на рисунке 11 показывает, что в городской среде перцептивная громкость превышает сельскую на всём диапазоне частот, особенно в области 400–2000 Гц, где разница достигает 10–15 дБ при 1 кГц, что отражает устойчивое преобладание шумовой нагрузки в урбанистическом окружении.

Таблица 1— Результирующие данные (значения усреднены по моделированным данным; относительные единицы — нормированные показатели в рамках используемой методики)

Среда	Индекс громкости (PLI, усл. ед.)	Резкость, кГц	Шероховатость (отн.)	Флуктуации (отн.)
Городская (Urban)	7.18×10^6	1.9	3.45	0.42
Индустриальная (Industrial)	7.02×10^6	2.2	3.63	0.39
Помещение Office	5.43×10^6	1.5	2.88	0.45
Сельская (Rural)	4.38×10^6	1.3	2.14	0.61
Природная (Nature)	4.07×10^6	2.0	2.26	0.66

Полученные результаты в таблице 1 демонстрируют, что психоакустическое профилирование звуковых сред позволяет количественно выделять типы акустических окружений на основе четырёх основных параметров восприятия — громкости, резкости, шероховатости и флуктуационной силы [1, 9]. Для урбанизированных и промышленных сред характерно смещение спектра восприятия к средним и высоким частотам, что приводит к повышению субъективной громкости и раздражающего потенциала звука, тогда как сельские и природные среды отличаются преобладанием низкочастотных компонентов и выраженными временными флуктуациями, создающими ощущение акустического комфорта. Сравнение с кривыми равной громкости ISO 226 подтверждает, что полученные профили соответствуют закономерностям слухового восприятия: для сохранения субъективно равной громкости низкие частоты требуют большей амплитуды, что объясняет снижение восприятия в диапазоне ≤ 200 Гц у всех типов сред [4–6, 10–12].

Применение разработанного программного метода обеспечивает автоматическую генерацию карт чувствительности и психоакустических индексов, что делает возможной количественную оценку акустического комфорта, оптимизацию звуковых систем и моделирование слуховой адаптации человека к окружающей среде. Дальнейшее развитие предложенной модели связано с включением физиологических модификаторов — например, смещения высокочастотных порогов у курящих и временного снижения чувствительности после употребления алкоголя [2–3] — с целью создания адаптивной системы прогнозирования индивидуальных слуховых профилей.

Заключение

В работе успешно разработан программный метод автоматизированного психоакустического профилирования звуковых сред на основе перцептивных параметров (громкости, резкости, шероховатости и флуктуационной силы). Модель, реализованная на Python с использованием открытых аудиодатасетов, позволила количественно выделить типы акустических окружений.

Установлено, что урбанизированные и промышленные среды обладают максимальным индексом громкости (*PLI*), достигающим 7.18×10^6 усл. ед. для Urban, и смещением спектра восприятия к средним и высоким частотам (400–2000 Гц), что обуславливает высокий раздражающий потенциал. В противоположность этому сельские и природные среды имеют значительно более низкий уровень громкости и демонстрируют высокую флуктуационную силу (0.61–0.66), создавая ощущение акустического комфорта.

Полученные психоакустические профили соответствуют закономерностям слухового восприятия, подтвержденным кривыми равной громкости ISO 226 [4, 10–12]. Разработанный метод обеспечивает количественную оценку акустического комфорта. Перспективы развития модели включают интеграцию физиологических модификаторов (влияние курения и алкоголя) для создания адаптивной системы прогнозирования индивидуальных слуховых профилей.

Список литературы

1. Цвикер Э., Фастль Х. Психоакустика. Факты и модели / Пер. с англ. под ред. Л. А. Чистякова. — М.: Мир, 2001. — 456 с.
2. Effects of Smoking on Eustachian Tube and Hearing // *The International Tinnitus Journal*. — 2022. — URL: <https://www.tinnitusjournal.com/articles/effects-of-smoking-on-eustachian-tube-and-hearing-8171.html> (дата обращения: 23.10.2025).
3. Hearing loss and alcohol consumption // *BMC Ear, Nose and Throat Disorders*. — 2007. — Vol. 7, No. 4. — URL:

- <https://bmcearthroatdisord.biomedcentral.com/articles/10.1186/1472-6815-7-4> (дата обращения: 23.10.2025).
4. ISO 226:2003 *Acoustics — Normal equal-loudness-level contours*. — Geneva: International Organization for Standardization, 2003.
 5. Salamon J., Jacoby C., Bello J. P. A Dataset and Taxonomy for Urban Sound Research // *ACM Multimedia Conference*. — 2014. — P. 1041–1044.
 6. Mesaros A., Heittola T., Virtanen T. TUT Acoustic Scenes 2017 Dataset and Baseline System // *IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events*. — 2018.
 7. Михайлов В. А. Основы психоакустики и восприятия звука человеком. — СПб.: НИИ Радиоэлектроники, 2010. — 128 с.
 8. Piczak K. J. ESC: Dataset for Environmental Sound Classification // *Proceedings of ACM Multimedia*. — 2015. — P. 1015–1018.
 9. Zwicker E., Fastl H. *Psychoacoustics: Facts and Models*. — Berlin: Springer-Verlag, 1999. — DOI: 10.1007/978-3-662-11960-8.
 10. HyperPhysics (Georgia State University). Equal-loudness contours with frequency in Hz. — URL: <http://hyperphysics.phy-astr.gsu.edu/hbase/Sound/eqloud.html> (дата обращения: 23.10.2025).
 11. Equal-loudness contour // *Wikipedia*. — URL: https://en.wikipedia.org/wiki/Equal-loudness_contour (дата обращения: 23.10.2025).
 12. Психоакустические основы слуха человека // *Studfile.net*. — URL: <https://studfile.net/preview/16462302/page:20> (дата обращения: 23.10.2025).
 13. ISO 532-1:2017 *Acoustics — Methods for calculating loudness. Part 1: Zwicker method*. — Geneva: International Organization for Standardization, 2017.
 14. DIN 45692:2009 *Measurement Technique for the Simulation of the Auditory Sensation of Sharpness*. — Berlin: Deutsches Institut für Normung, 2009.
 15. Корякин, С. В. Аналитический обзор технологий построения аппаратно-ориентированных облачных систем защиты информации с применением нейросетевых технологий / С. В. Корякин // Проблемы автоматизации и управления. — 2025. — № 2(53). — С. 41—51. — EDN RCCRHC.