

Проблема узгодження експертних і алгоритмічних оцінок якості мастерингу аудіо

УДК 004.04

Євген Лабун¹, Богдан Худік²

*Державний Університет Інформаційно Комунікаційних Технологій,
¹e.labun@duikt.edu.ua, ²b.khudik@duikt.edu.ua*

Автоматизоване оцінювання якості мастерингу аудіо потребує узгодження технічних параметрів сигналу з тим, як результат сприймається експертами та слухачами. Мастеринг є завершальним етапом підготовки композиції до публікації, де приймаються рішення щодо гучності, динаміки, спектрального балансу, просторовості та цілісності звучання [6, 7]. Водночас відповідність окремим нормам не гарантує високої перцептивної якості: трек може мати коректні LUFS і true peak, але сприйматися як мутний, різкий, надмірно стиснений або неприродний [1, 4, 7, 8].

Актуальність роботи зумовлена тим, що в задачах машинного навчання експертні судження потрібно перетворювати на формалізовані цільові оцінки, а алгоритмічні метрики — на інформативні ознаки. ITU-R BS.1770-5 формалізує вимірювання гучності та true peak [1], ITU-R BS.1116-3 і BS.1534-3 задають основу контрольованих слухових тестів [2, 3], ITU-R BS.1387-2 описує об'єктивне оцінювання сприйманої якості аудіо [4], а ITU-T P.1401 визначає статистичні підходи до перевірки objective quality prediction models відносно суб'єктивних результатів [5].

Метою роботи є обґрунтування проблеми узгодження експертних і алгоритмічних оцінок якості мастерингу аудіо та визначення підходу, який може бути використаний на підготовчому етапі створення датасету для моделей машинного навчання. На відміну від простого переліку критеріїв, акцент зроблено на зв'язку між об'єктивними дескрипторами сигналу та суб'єктивними оцінками якості звучання.

Проблема узгодження проявляється у трьох аспектах. По-перше, технічні метрики мають різну перцептивну значущість залежно від жанру, матеріалу та контексту відтворення [6, 7]. По-друге, якість є багатовимірним сприйняттям: чистота, прозорість, тембральна природність, просторовість, відсутність артефактів і загальне враження не зводяться до одного числового показника [2–4, 8]. По-третє, об'єктивні моделі залежать від домену застосування, тому метрики, валідовані для кодування або інших задач, не можна автоматично переносити на оцінювання мастерингу без додаткової перевірки [4, 9].

Запропонований підхід передбачає парне представлення кожного аудіофрагмента: з одного боку обчислюються алгоритмічні ознаки (LUFS, true peak, RMS, crest factor, динамічний діапазон, спектральні характеристики, міжканальна кореляція, ознаки кліпінгу або надмірної компресії), з іншого — збираються експертні оцінки за перцептивними шкалами [1–3, 6, 8]. Такі оцінки доцільно зберігати не лише як інтегральний бал, а як набір пов'язаних міток: загальна якість, прозорість, тембральний баланс, просторовість, виразність динаміки та наявність артефактів.

Для організації слухового оцінювання доцільно використовувати логіку ITU-R BS.1116-3 у випадку малих відмінностей між варіантами мастерингу та ITU-R BS.1534-3 / MUSHRA-подібні процедури для проміжних рівнів якості [2, 3]. Для перевірки узгодженості прогнозів моделі з експертними оцінками варто враховувати розбіжності між оцінками слухачів, можливі нетипові судження та невизначеність суб'єктивних результатів [5].

У задачах машинного навчання об'єктивні метрики можуть виступати вхідними ознаками, експертні оцінки — цільовими значеннями, а випадки розбіжності між ними — матеріалом для аналізу обмежень моделі. Робота Shtern, Casas і Tzerpos демонструє релевантність ML саме для оцінювання якості музичного мастерингу відносно професійного еталона [6]. Дослідження якості популярної музики показує, що суб'єктивні оцінки якості пов'язані з об'єктивними ознаками сигналу, зокрема гучністю та динамічним стисненням, тоді як оцінка «подобається» має іншу природу [8]. Очікуване значення підходу полягає в тому, що він створює основу для датасету, у якому якість мастерингу описується не лише технічними параметрами, а й експертною перцептивною розміткою. Це дає змогу досліджувати, які об'єктивні ознаки корелюють із людським сприйняттям, у яких випадках алгоритмічна оцінка суперечить експертній і які фактори — жанр, стиль, тип обробки або особливості слухача — впливають на результат [5, 8, 9].

Отже, проблема узгодження експертних і алгоритмічних оцінок є ключовою для розробки методів автоматизованого оцінювання якості мастерингу аудіо. Запропонований підхід не передбачає готової навченої моделі, а визначає методологічну основу для подальшого формування датасету, проведення слухових тестів і побудови моделей, здатних пояснювати розбіжності між технічними метриками та людським сприйняттям.

1. ITU-R BS.1770-5. Algorithms to measure audio programme loudness and true-peak audio level. Geneva: ITU, 2023.
2. ITU-R BS.1116-3. Methods for the subjective assessment of small impairments in audio systems. Geneva: ITU, 2015.
3. ITU-R BS.1534-3. Method for the subjective assessment of intermediate quality level of audio systems. Geneva: ITU, 2015.
4. ITU-R BS.1387-2. Method for objective measurements of perceived audio quality. Geneva: ITU, 2023.
5. ITU-T P.1401. Methods, metrics and procedures for statistical evaluation of objective quality prediction models. Geneva: ITU, 2020.
6. Shtern M., Casas P., Tzerpos V. Evaluating Music Mastering Quality Using Machine Learning. CASCON, 2018. P. 126–135.
7. Katz B. Mastering Audio: The Art and the Science. 3rd ed. New York: Routledge / Focal Press, 2015.
8. Wilson A., Fazenda B. M. Perception of Audio Quality in Productions of Popular Music. J. Audio Eng. Soc. 2016. Vol. 64, No. 1/2. P. 23–34. DOI:10.17743/jaes.2015.0090.
9. Torcoli M., Kastner T., Herre J. Objective Measures of Perceptual Audio Quality Reviewed. arXiv:2110.11438, 2021.