

ЕФЕКТИВНІСТЬ ЗАСТОСУВАННЯ КОМПРЕСІЇ ДАНИХ ПРИ РОЗРОБЦІ ПСИХОАКУСТИЧНОЇ МОДЕЛІ СИГНАЛУ

Дана стаття присвячена розгляду та аналізу психоакустичної моделі сприйняття звуку людиною, її роль та вплив на результати стиснення аудіо даних. Розглянуто перспективність застосування психоакустичної моделі як одного з етапів комплексного підходу до компресії. Наведено приклади реалізації у існуючих рішеннях та виділено їх переваги.

Ключові слова: психоакустика, психоакустична модель, компресія, кодек, спектр, дискретно косинусне перетворення, швидке перетворення Фур'є.

Постановка проблеми

З розвитком телекомунікаційних систем та зростанням швидкості передачі даних питання стиснення аудіо не втрачає своєї актуальності. Оскільки аудіо компресія потребує комплексного підходу для отримання кращих результатів, її можна умовно розбити на три основних етапи: математична модель, кодування та застосування психоакустичної моделі.

Психоакустика — наука, що вивчає психологічні та фізіологічні особливості сприйняття звуку людиною. Розуміння цих процесів дозволяє вченим та інженерам зосередитись на питанні реалізації систем та досліджень за певних умов, що оптимізують цей процес.

Межі сприйняття людиною звуку не регулюються лише фізіологією, не менш важливу роль грає психологічний фактор. Номінально вважається що, діапазон у якому людина сприймає звук від 16 до 2000 Гц. Верхня межа має тенденцію зменшуватись з віком. Більшість людей не сприймає звук за цим діапазоном частот, тому інформацією що міститься за його межами можна знехтувати при реалізації аудіо кодедку з втратами, що дозволяє значною мірою виграти у стисненні. Це не єдиний фактор який можна використовувати для реалізації психоакустичної моделі та завдяки якому можна покращити результат компресії аудіо даних [1, 2].

Аналіз останніх досліджень і публікацій за темою наукового пошуку засвідчує, що проблему дослідження та практичного застосування психоакустичної моделі у стисненні аудіо даних розглядало багато науковців, зокрема, В.М. Шутко, М.О. Шутко, . Питання стиснення аудіо даних є актуальною темою що постійно досліджується, оскільки питання компресії потребує комплексного підходу, потрібно застосовувати знання з багатьох наукових областей. Для розробки психоакустичної моделі потрібно провести також дослідження у фізіології та психології.

Метою статті є огляд та дослідження психоакустичної моделі стиснення аудіо даних, аналіз існуючих рішень які реалізовано у популярних аудіо форматах та загальний вплив на результат стиснення.

Викладення основного матеріалу дослідження

Людина здатна сприймати великі потоки інформації, але фізіологія обумовлює обробку лише 100 біт/с, що говорить про суттєвий надлишок властивий аудіо сигналам. Важливою проблемою для цифрового представлення звукових сигналів є скорочення статистичної та психоакустичної надлишковості [2].

Якщо скорочення статистичної надлишковості полягає лише у математичній обробці дискретного сигналу, то питання психоакустики опирається на слухові властивостях людини. За допомогою усунення психоакустичних надлишковостей можливо забезпечити стиснення в 10-12 разів без суттєвих втрат у якості вихідного сигналу.

Опис характеристик слухової системи здійснюється по частотно-пороговим кривим, котрі відображають певні значення інтенсивності (або сили звукового тиску) в залежності від частоти звукових коливань. Частотно-порогові криві, отримані для слухової системи загалом,

отримали назву кривих чутності. Чуттєвість слуху оцінюється мінімальною інтенсивністю звуку, при якому людина може відрізнити інформацію від постійно існуючого фону.

Крива порогу чутності (рис. 1.1), характеризує найменшу інтенсивність звуку певної частоти, який людина може почути за умов тиші.

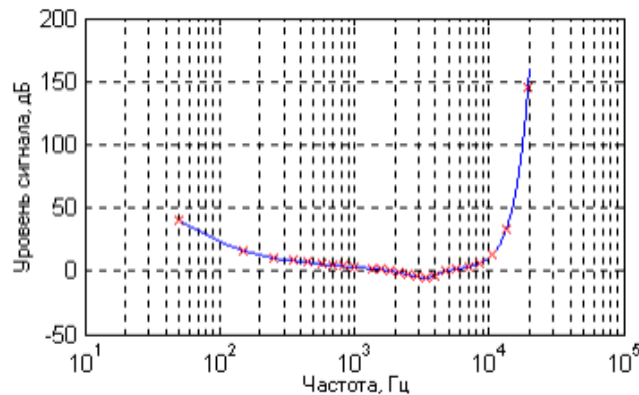


Рис. 1.1. Крива мінімального порогу чутності слухового апарату людини у тиші [3]

Аналіз усередненого графіку порогу чутності показує, що слухова система людини сприймає частоти в діапазоні 0.01 - 20 кГц. Поріг чутності на частоті 3 кГц складає приблизно 0 дБ. Поріг чутності значно збільшується в області низьких (менше 200 Гц) та високих (більше 10 кГц) частот та може досягати значень 40 дБ та вище. Отже для слабких сигналів амплітудно-частотна характеристика слуху є нелінійною, тому слабкі низькочастотні та високочастотні складові сигналу можуть бути не чутні. Завдяки кривим чутності можна оцінити властивості слухової системи по таким ознакам: діапазону частот, полосі частот, у межах якої чуттєвість максимальна, абсолютним значенням звукового тиску та її зміною залежно від частоти.

Однією з важливих властивостей слухового апарату людини є його інтегруюча здатність, тобто групування частотних складових звуку в певні частотні критичні полоси. Слухова система порівнює корисний сигнал та шум по інтенсивності в межах критичних смуг, оцінюючи поріг чутності. Критична полоса слуху є полосою, за межами якої суб'єктивні відчуття звуку різко змінюються.

Для виділення у звуковому сигналі корисної частини інформації і видалення із нього непотрібних деталей слугує ефект маскування. Він пов'язаний з процесом взаємодії сигналів, що приводять до змін слухової чутливості до закодованого сигналу в присутності маскуючого.

Ефект від слухового маскування залежить від спектральних та часових характеристик маскованого та маскуючого сигналу. Також воно безпосередньо пов'язане з не лінійністю слуху. Це проявляється в тому, що при впливі на барабанну перетинку достатньо гучного синусоїдального звуку з частотою f у слуховому апараті формуються гармоніки цього звуку з частотами $2f$, $3f$ і тд. Оскільки в первинному тоні ці гармоніки відсутні, вони отримали назву — суб'єктивних гармонік.

Для маскування більш важлива не абсолютна гучність, а відношення потужності гучного сигналу до потужності тихого. Окрім того, чим ближча частота до маскуючого сигналу, тим вищий ефект маскування. Степінь маскування — це різниця в децибелах між рівнем порогу чутності за умов тиші.

На рисунку 1.2 видно, що зі збільшенням середньої частоти маскуючого сигналу діапазон частот у якому проявляється маскування стає ширшим. Криві маскування є несиметричними, вони мають різкий спад в низьких частотах та пологий в сторону високих.

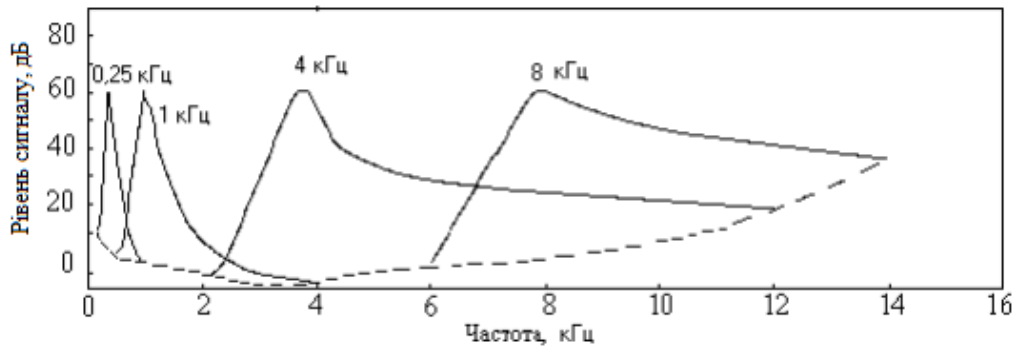


Рис. 1.2. Змінення порогу чутливості слухової системи при маскуванні сигналу [3]

Внаслідок розділу спектру звукового сигналу на частотні полоси слух реагує не на загальну потужність звукового сигналу, а на потужність, зосереджену в окремих частотних групах. При цьому більш інтенсивні частотні групи за певних умов можуть маскувати менш інтенсивні. Щодо стиснення це означає, що в групах меншої інтенсивності припускається більший рівень шуму квантування, тобто можна зменшити число розрядів при кодуванні спектральних складових, що дозволяє виграти у стисненні.

Алгоритм роботи психоакустичної моделі слухової системи призначено для розрахунку глобальних маскуючих порогів та відношення сигнал/маска та складається з п'яти етапів [4]:

1. Спектральний аналіз та нормалізація рівня звукового тиску.
2. Ідентифікація тональних та шумових маскерів.
3. Прорідження та перетворення маскерів.
4. Розрахунок індивідуальних порогів маскуванню.
5. Розрахунок глобальних порогів маскуванню.

Прецептивний кодер, що використовується у стандарті MPEG, є прикладом ефективного стиснення аудіо інформації з урахуванням властивостей слухової системи людини. На рисунку 1.3 представлено структурну схему прецептивного кодера на основі MP3. З рисунку видно, що цифровий 16-бітовий аудіо сигнал розкладається на $M = 32$ частотні полоси з допомогою групи рівномірно розподілених полосних фільтрів. Усі фільтри групи отримуються шляхом модулювання фільтра прототипа, призначеного для збільшення частотної розподільної здатності аудіо сигналу та кращої апроксимації критичних полос слухової системи. Імпульсний відгук фільтра i полоси на одиничний імпульс $h_i(k)$ представляється у вигляді піднесення коефіцієнтів $h(k)$ фільтра прототипа, що є квантованими значеннями імпульсної характеристики цього фільтра, та моделюючої функції, котра зміщує низькочастотний відгук у частотну полосу i .

$$h_i(k) = h(k) \cos\left(\frac{2i-1}{2M} + \phi(i)\right), \tag{1.1}$$

M – кількість полос, на які розділяється звуковий сигнал; $i=1,2,\dots,32$; k – номер коефіцієнту імпульсної характеристики $h(k)$ фільтра прототипу ($k=1,2,\dots,512$), $\phi(i)$ — фазове зміщення для i полоси [5].

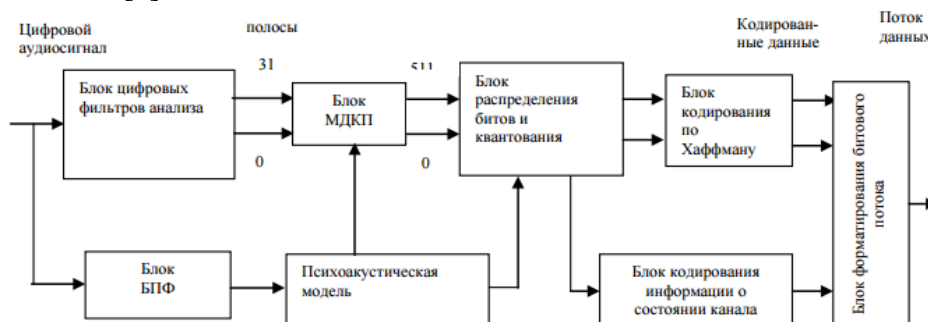


Рис. 1.3. Структурна схема прецептивного кодера MP3

Сигнал на виході i полоси фільтру у дискретний момент часу n представляється у вигляді дискретної згортки:

$$y_i(n) = x(n) * h_i(n) = \sum_{k=0}^{N-1} x(n-k)h_i(k), \quad 1.2$$

* - символ згортки; $h_i(k)$ – масив коефіцієнтів i -го полосового фільтру порядку $N=512$; $x(n-k)$ – вихідний масив цифрових даних фільтру. Число N визначає ефективність фільтру, реалізуючого N операцій множення з накопленням для кожного вихідного відліку $y_i(n)$.

Використання набору цифрових фільтрів при кодуванні, дає високу часову розподільну здатність і таким чином допускається точніше моделювання ефекту маскуванню. Полоси нормалізуються з допомогою масштабуючого фактору так, що максимальне значення амплітуди відліку в кожній полосі рівне одиниці. Збільшення точності розпізнавання маскуючих сигналів досягається з допомогою подальшої обробки 32 полосних сигналів з використанням модифікованого дискретно косинусного перетворення (МДКП), яке використовується для аналізу сусідніх областей частотних полос. МДКП базується на усуненні накладення в часовій області шляхом використання часових вікон, що пересікаються для аналізу полосових сигналів. Накладення в часовій області — усувається при зворотному косинусному перетворенні.

З практичної точки зору сигнали та вузькополосні фільтри не мають обмеженої полоси, тому, присутнє незначне перекриття сусідніх частотних полос, що приводить до втрат при відтворенні сигналу. Для усунення часового ефекту накладення, що має назву *aliasing*, вікна аналізу задаються формулою:

$$h(k) = \pm \sqrt{2} \sin \left[\left(k + \frac{1}{2} \right) \frac{\pi}{N} \right], \quad 1.3$$

Кодер MP3 реалізує метод адаптивної обробки значень МДКП, що дозволяє запобігти появі деяких артефактів, викликаних перекриттям полос групи фільтрів аналізу та більш ефективно використовувати психоакустичні властивості слухової системи. Кодер використовує два блоки МДКП різної довжини.

Для змінення довжини МДКП використовується набір віконних функцій. Рішення у виборі довжини МДКП приймає психоакустична модель кодеру. Вихідні полосові сигнали проходять через МДКП в двох різних блоках: короткому (16 — точкове МДКП) та довгому (36 — точкове). Короткий блок покращує часову розподільну здатність кодеру, що дозволяє ефективно використовувати предмаскування для запобігання появі різних артефактів, що з'являються при кодуванні. Довгий блок дозволяє реалізувати більш високу частотне розширення для квазістаціонарних сигналів, що скорочує додаткову інформацію що до стану каналу та зменшує швидкість аудіо потоку.

Для спектральної декомпозиції аудіо сигналу та досягнення високої спектральної розподільної здатності при обчисленні маскуючих порогів паралельно з розбиттям аудіо сигналу на полоси, обраховується 1024 точкове FFT. Психоакустична модель використовує результати перетворення Фур'є для розділення полосових даних на довгі та короткі блоки в залежності від частотного або часового розширення та виявлення тональних і нетональних компонентів спектру. Значення тональних компонентів знаходиться шляхом знаходження локальних максимумів спектру. Маскуючі ефекти, що зумовлені тональними та нетональними компонентами та порогом чутності, використовуються для знаходження мінімального маскуючого порогу для кожної полоси. Суть розрядного (бітового) розподілу полягає у тому, що часто певні фрагменти аудіо сигналу не можуть бути закодовані в рамках даного розрядного діапазону без втрат якості. В такому випадку використовується невеликий запас розрядів, що зберігаються в буфері, для кодування простіших фрагментів в менший

розрядовий діапазон. Даний режим призначено для кодування блоків сигналу, що містять семпли які легко піддаються стисненню з мінімально необхідною кількістю бітів, а інші розряди додаються в блоки з сигналом що має складну компенсацію.

Висновки

Стиснення аудіо даних є актуальним питанням що постійно досліджується. На сьогоднішній день компресія звуку потребує комплексного підходу, що зачіпає області різних прикладних наук. Якщо математична модель та кодування потребує математичних знань та використання теорії ймовірності, то робота над психоакустичною моделлю здійснюється з допомогою психології та біології. Все це допомагає створити найбільш ефективні методи стиснення аудіо даних, дозволяє зробити їх більш гнучкими та адаптивними. Психоакустична модель є досить важливим етапом стиснення, на якому визначається якою інформацією можна знехтувати при кодуванні без суттєвих втрат, помітних слуху людини. Створення психоакустичної моделі є складним питанням, що потребує математичної реалізації психічних та фізіологічних властивостей які допомагають найбільш ефективно обробити аудіо сигнал та провести компресію.

Список використаної літератури

1. Сэлмон Д. Сжатие данных, изображений и звука.— М.: Техносфера, 2004.
2. Сергиенко В.С., Баранов В.В. Сжатие данных, речи, звука и изображений в телекоммуникационных системах: Учебное пособие – М.: ИП РадиоСофт, 2011. – 360 с.
3. Киркоров С.И. Цифровая обработка речи и изображений: Курс лекций – Минск, 2010.
4. Nikil Jayant, James Johnston, Robert Safranek. (October 1992). «Signal Compression Based on Models of Human Perception». Proceedings of the IEEE 81 (10): 1385—1422.
5. Шутко В.М., Шутко М.О., Методи та засоби стиснення інформації: Навчальний посібник — К.: НАУ, 2012.

Надійшла 02.07.2017 р.

Рецензент: д.т.н., проф. Дружинін В.А.