

УДК 004

Сергій ЛУКЬЯНЧИКОВ

lsd57@ukr.net

ORCID: 0000-0002-6837-2930

м. Миколаїв

ДОСЛІДЖЕННЯ СУЧАСНИХ МЕТОДІВ СТИСНЕННЯ ЗВУКОВОЇ ІНФОРМАЦІЇ З ВТРАТАМИ ТА БЕЗ ВТРАТ

У роботі досліджені методи архівації які можна застосувати для стиснення звукової інформації з втратами та без втрат. Зроблений порівняльний аналіз викладених методів.

Ключові слова: звукова інформація, імпульсно-кодова модуляція, семпли, RIFF WAVE, стиснення без втрат, MP3, бітрейт, вейвлет, фрактальне стиснення.

Постановка проблеми

Звукова інформація, яка зберігається в пам'яті ЕОМ, це набір частотно-времених відліків звукового сигналу. Іншими словами, звук зберігається у вигляді фрагментів за допомогою пар значень "частота-тривалість" і відтворюється за допомогою стандартного динаміка ПЕОМ, керованого найпростішої схемою. Схема управління динаміком відтворює звуковий сигнал за допомогою тактового генератора, який генерує звуковий сигнал заданої частоти протягом заданого відрізка часу [1]. Цей механізм досі використовується в BIOS ПЕОМ для генерації звукових повідомлень про помилки.

Аналіз останніх досліджень і публікацій

Природно, що так можна було зберігати синтезований звук, але ніяк не складні музичні композиції. Тому був розроблений формат цифрового представлення звуку RIFF WAVE, в основі якого лежить принцип імпульсно-кодової модуляції звукового сигналу.

Імпульсно-кодова модуляція звукового сигналу має на увазі заміну безперервного звукового сигналу набором дискретних значень амплітуди звуку, вимірених із заданою частотою дискретизації (рис. 1). Кожен вимір амплітуди (так званий "семпл") зберігається у вигляді дискретного коду певної розрядності.



Рис. 1. Імпульсно-кодова модуляція звуку

Формулювання цілей

Метою роботи є дослідження сучасних методів стиснення звукової інформації з втратами та без втрат. Результати дослідження будуть враховані при вдосконаленні методу запропонованому автором в роботі [2].

Основний матеріал

Звукова інформація в форматі RIFF WAVE крім, власне, значень семплів містить також спеціальний заголовок, в якому вказані параметри звукової інформації (частота дискретизації, розрядність семплів, кількість звукових каналів).

Формат RIFF WAVE передбачає наступні значення для параметрів звукової інформації:

- 1) частота дискретизації: 8000..48000 Гц;

- 2) розрядність семплів: 8 або 16 біт;
- 3) кількість звукових каналів: 1 (моно) або 2 (стерео).

Формат RIFF WAVE дозволяє зберігати абсолютно будь-який звуковий сигнал: мова, музичні композиції і т.п.

Залежно від типу звукового сигналу можна вибрати відповідні величини для параметрів звукової інформації. Тут необхідно брати до уваги теорему Котельникова, яка показує, що частота дискретизації звукового сигналу повинна як мінімум вдвічі перевищувати частоту найбільш високочастотної гармоніки сигналу. Таким чином, для мови можна вибрати формат 8000Гц, моно, 8 біт на семпл, тому що спектр мови лежить в межах 300..4000 Гц. Для зберігання високоякісних музичних композицій необхідно 44-48кГц, 16 біт/семпл, стерео.

Безсумнівною перевагою формату RIFF WAVE є практично аналогова якість звуку при максимальних частотах дискретизації. Але у цього методу є і істотний недолік: значний об'єм звукової інформації.

Неважно підрахувати, що при параметрах 48000Гц, 16біт (2 байта) на семпл, стерео, для зберігання однієї секунди звучання необхідно виділити $48000 \times 2 \times 2 = 192000$ байт інформації. Очевидно, що при часі звучання 5 хвилин (середня тривалість музичної композиції) звукова інформація в форматі RIFF WAVE займе $5 \times 60 \times 192000 = 57600000$ байт, або майже 58 мегабайт (!). Тобто, формат RIFF WAVE є, м'яко кажучи, надмірним. Все вищесказане наводить на думку про стиснення звукової інформації в форматі RIFF WAVE з метою зменшення її об'єму.

Класичні схеми архівації (LZW, метод Хаффмена) не підходять для стиснення звукової інформації, оскільки не враховують її особливостей і, як наслідок, дають низькі ступені стиснення (не більше, ніж в 1,2..1,5 рази).

Однією із спроб стиснення звукової інформації була розробка формату збері-

гання звукової інформації, яка носить назву ADPCM (Adaptive Differential Pulse Code Modulation – Адаптивна Диференціальна Імпульсно-Кодова Модуляція). Цей формат передбачає зберігання в кожен дискретний момент часу не амплітуди звукового сигналу, а відхилення від попередньої амплітуди. Відхилення амплітуд менше самих амплітуд, тому, виділивши менше біт під зберігання відхилень можна стиснути звукові дані.

У ADPCM під відхилення виділяється кількість біт в 2 рази менше, ніж під значення семплів. При цьому звукові дані зменшуються в об'ємі в 2 рази. Однак, найчастіше стиснути звукову інформацію в 2 рази без втрат неможливо, тому ADPCM неточно передає деякі відхилення через недостатню розрядність. Це призводить до деяких втрат якості звуку через що виникає шум квантування: відновлені при відтворенні семпли дещо відрізняються від тих, які були закодовані. В цілому, формат ADPCM малоефективний для стиснення звукової інформації, тому що дає стиснення всього в 2 рази, при цьому з деякими втратами.

Дослідження, що проводяться асоціацією MPEG (Motion Picture Experts Group) в області психоакустики дали свої плоди: був розроблений метод стиснення звуку MPEG Layer-3 (або MP3).

Психоакустика вивчає особливості сприйняття звуку людиною. З'ясувалося, що в спектрі чутних людиною частот (умовно від 20Гц до 20кГц) можна виділити частотні смуги, звані критичними смугами (в англійській термінології – critical bands). Хоча роздільна здатність людського вуха нерівномірна і різниться в залежності від частоти, всередині критичних смуг вона приблизно однакова. Це дає можливість квантування звуку по амплітуді всередині критичних смуг у відповідність з роздільною здатністю вуха.

На цьому принципі і працює MP3. Спочатку звукова інформація фільтрується по критичним смугам, потім усередині ко-

жної смуги звук квантується. При цьому "несуттєві" семпли, які незначно впливають на звучання, просто видаляються, а що залишилися після фільтрації стискаються за методом Хаффмена. Фактично, MP3 звужує спектр сигналу, виключаючи з нього "несуттєві моменти", за рахунок чого і відбувається стиснення звукової інформації.

Коли мова йде про психоакустичне кодування, часто застосовується поняття бітрейт (від англ. Bitrate – бітова швидкість). Цей термін позначає кількість біт інформації, яка передає секунду звучання. Бітрейт, як правило, вимірюється в кілобітах в секунду (кбіт/с). Так, в MP3 застосовується кодування звуку з постійним бітрейтом в межах 32..320 кбіт/с.

MP3 дає досить високі ступеня стиснення звуку (в залежності від бітрейту – до 12:1) і придатний для потокового стиснення звуку (в реальному часі). Також, цей метод стиснення є найбільш популярним для зберігання музичних композицій.

Однак, психоакустична модель, що застосовується в MP3, є досить грубою, і при відтворенні стислій звукової інформації неминуче виникають звукові артефакти (кляцання і "булькання"), особливо на низьких бітрейтах. Тому останнім часом з'явилося кілька альтернатив MP3.

Одним з альтернативних психоакустических методів стиснення звуку є OggVorbis. Цей метод стиснення за своїми принципами нагадує MP3, однак має істотну відмінність – змінний бітрейт, варійований в заданих межах.

Іншими словами, в OggVorbis звукова інформація стискається адаптивно, в залежності від складності звукового фрагмента: в ті моменти, коли складність звуку вище, відбувається кодування з великим бітрейтом; у відносно нескладних ділянках використовується менший бітрейт. Цей механізм дає кращу якість при тому ж рівні ступеня компресії. Особливо ефективний OggVorbis при низьких бітрейтах.

Ще одним психоакустичним методом, альтернативним MP3, є метод AAC (Advanced Audio Coding – Покращене Кодування Аудіо). AAC стискає якісніше, ніж MP3, і дозволяє стискати багатоканальний звук, що особливо важливо сьогодні, коли широко використовуються акустичні системи стандарту 5.1 (5 динаміків-сателітів + басовий динамік-сабвуфер). В даний час AAC широко використовується для кодування звукових доріжок DVD-фільмів.

На даний момент існує вже досить багато AAC-кодеків, серед яких найбільш відомим є AAC LC, застосовуваний в технології Apple QuickTime. Також поширений Nero AAC, використовуваний в пакеті Nero Burning Rom. В даний час зростає число апаратних плеєрів, що підтримують формат AAC, найбільш популярним з яких є Apple iPod.

До психоакустичних методів стиснення звуку також відноситься метод MusePack. Цей метод дає найкращу серед усіх психоакустичних алгоритмів якість звучання відтвореного звукового сигналу. MusePack дозволяє стискати звук зі змінним бітрейтом, що значно підвищує ступінь компресії.

Алгоритм MusePack дозволяє налаштувати якість кодування звуку за допомогою декількох типових психоакустичних пресетів (попередніх установок), які задають межі зміни бітрейта. Найкращої якості звуку, при використанні пресету "braindead" (бітрейт в межах 232..278 кбіт/с).

Останнім часом, коли значно зросла ємкість носіїв інформації, розроблено кілька методів безвратного стиснення звукової інформації. Такі методи стиснення характерні тим, що забезпечують відновлення закодованої звукової інформації без будь-яких втрат або спотворень.

Методи безвратного стиснення звуку часто застосовуються для стиснення музичних композицій, коли якість відіграє основну роль. Такі методи стиснення не забезпечують високих ступенів компресії,

як, наприклад, психоакустичні методи, але зате зберігають первинну якість оригіналу. До методів безвтратного стиснення звуку відносяться FLAC, WavPack і Monkey's Audio.

FLAC (Free Lossless Audio Codec – Безкоштовний Безвтратний Аудіо-Кодек) є одним з найбільш популярних форматів стиснення звуку без втрат. Кодек є безкоштовним і набір інструментів, написаний (існує в MacOS X, Linux, Unix). Існують плагіни для використання FLAC в популярних музичних програвачах WinAmp і FooBar2000.

WavPack стискає інформацію з деякими втратами, але створює два результуючих файлу – перший містить звук, стиснений з втратами, а другий є коригувальним, за допомогою якого з першого файлу можна відновити оригінал. Файл, що містить звук, стиснений з втратами, може також прослуховуватися окремо від частини, яка коректує. Це робить WavPack гібридним методом стиснення (дає стиснення як з втратами, так і без). Для WavPack також існують плагіни до WinAmp і FooBar2000.

Monkey's Audio є, фактично, найбільш популярним серед методів безвтратного стиснення звуку. Дозволяє стискати звукову інформацію в одному з п'яти режимів – від найбільш швидкого (дає меншу ступінь стиснення) до найбільш повільного (ступінь стиснення найбільша).

Monkey's Audio забезпечує найбільш високі ступені стиснення звуку без втрат, що перевищують ступеня стиснення за допомогою архіватора WinRAR з максимальними настройками стиснення. Єдиним недоліком формату Monkey's Audio є його фактична одноплатформеність (присутній тільки для платформи Windows).

Одним з напрямків розвитку методів стиснення звуку останнім часом є стиснення звуку на основі математичних методів. До таких методів стиснення відносяться:

- 1) методи стиснення на основі вейвлетів;
- 2) фрактальні методи стиснення;

3) методи стиснення на основі стохастичних диференціальних рівнянь.

Методи стиснення на основі вейвлетів (від англ. Wavelet – хвильової сплеск) використовують параметризовану апроксимацію звукового сигналу в форматі RIFF WAVE за допомогою деякого набору вже вивчених стандартних вейвлетів.

Фрактальні методи стиснення звуку використовують математичний апарат фракталів. Найчастіше апарат фракталів застосовується до стиснення зображень. Стосовно до звукової інформації, фрактали застосовуються для стиснення не самого звукового сигналу, а його зображення (реалізації). Суть фрактального стиснення полягає в знаходженні самоподібних областей в зображенні, за допомогою яких можна представити інші області, застосувавши афінні перетворення на площині. Фрактальні методи стиснення звуку є дуже перспективними з точки зору ступеня стиснення (вдається стиснути звук в 1000 і більше разів), але поки досить сильні втрати в якості звуку і дуже великий час стиснення: він може в сотні разів перевершувати тривалість стискання звукового сигналу.

Ще однією групою математичних методів стиснення звуку з втратами є група методів стиснення на основі стохастичних диференціальних рівнянь (СДУ). В основі СДУ-методів лежить ідея апроксимації звукового сигналу за допомогою СДУ [3]. Замість апроксимованих ділянок звукового сигналу зберігаються параметри СДУ, за рахунок чого і відбувається стиснення.

Як і фрактальні методи, СДУ-методи стиснення дають дуже високі ступеня компресії (1:1000 і вище), але великі втрати в якості звуку. Також дуже великий час стиснення з наступних причин:

1) складно підібрати конкретну модель СДУ, як правило, застосовується декілька типових моделей, між якими здійснюється вибір найбільш підходящої;

2) для кожної з моделей проводиться параметрична ідентифікація звукового сигналу (визначаються параметри СДУ);

3) відбувається підбір датчика випадкових чисел для передачі випадкових відхилень сигналу від СДУ.

Проблема полягає в тому, що на сьогоднішній день не існує стрункого математичного апарату для розв'язання задачі підбору датчика випадкових чисел. Як правило, використовуються конгруентні генератори випадкових чисел і проводиться перебір їх параметрів в заданих межах. Через неточності підбору параметрів датчиків і виникають сильні відхилення амплітуд звукового сигналу від оригіналу (сигнал містить багато шуму). Тому використовують різні методи фільтрації звукового сигналу для часткового усунення шуму.

На сьогоднішній день СДУ-методи стиснення звуку знаходяться на експериментальній стадії; ще не існує серйозних алгоритмів для стиснення звуку на основі СДУ, але цей напрямок є дуже перспективним.

Спільною особливістю як фрактальних, так і СДУ-методів є великий час стиснення і дуже малий час відновлення сигналу. Такі методи стиснення можна застосовувати в тих областях, де найчастіше потрібно багаторазове відтворення звукового сигналу. Стиснення при цьому може бути зроблено лише одноразово.

У підсумку, можна відзначити наступне. Методи представлення та зберігання звукової інформації в ЕОМ за останні 20-30 років зазнали значних змін – від простого частотно-часового подання та імпульсно-кодової модуляції до складних форматів стиснення.

Розглянуті методи стиснення звуку можна класифікувати по забезпеченню цілісності звукової інформації на:

1) безтратні методи стиснення (FLAC, Monkey's Audio);

2) методи стиснення з втратами (ADPCM, MP3, AAC і ін.);

3) гібридні (WavPack)

Також методи стиснення з втратами можна класифікувати за принципами стиснення на:

1) психоакустичні (MP3, AAC, OggVorbis, MusePack і ін.);

2) математичні.

Математичні методи також розподіляються за групами:

1) вейвлет-методи;

2) фрактальні методи;

3) СДУ-методи стиснення.

Висновки і перспективи досліджень

Для зберігання музичних композицій найчастіше застосовуються формати психоакустичних методів стиснення. Ці методи дозволяють здійснювати швидко стиснення в реальному часі.

Високоякісні аудіозаписи найбільш доцільно зберігати або в форматі RIFF WAVE, або в одному з форматів безтратного стиснення, які забезпечують стовідсоткову відповідність оригіналу.

Фрактальні і СДУ-методи стиснення звуку поки мало використовуються через недосконалість математичного апарату і високих вимог до обчислювальних ресурсів ЕОМ. Можливо, в майбутньому, коли стануть доступними більш досконалі технології, і обчислювальна потужність ЕОМ значно зросте, методи стиснення звукової інформації, заснованих на фракталах і СДУ знайдуть своє застосування.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Скэнлон, Л. Персональные ЭВМ IBM PC и XT. Программирование на языке ассемблера [Текст] / Скэнлон Л. // Пер. с англ. – 2-е изд., стереотип. – М.: Радио и связь. 1991. – 336 с.
2. Ustenko, S. A software for data compression using double differential transformation of sound signal [Text] / S. Ustenko, S. Lukuanchikov, I. Lukuanchikov // Геометричне моделювання та інформаційні технології: науковий журнал. – №1 (3), квітень 2017. – Миколаїв: МНУ імені В.О. Сухомлинського, 2017. – С. 128-132.

3. Приходько, С.Б. Сжатие звука на основе стохастических дифференциальных уравнений второго порядка [Текст] / С.Б. Приходько // Вестник Херсонского государственного технического университета. – Херсон: ХГТУ. – 2002. – Вып. 2(15).

Serhiy LUKYANCHIKOV

Mykolayiv

**RESEARCH OF MODERN METHODS OF AUDIO CONTENT STORAGE
WITH LOSSES AND WITHOUT LOSS**

The paper investigates the archiving methods that can be used to compress audio information with losses and without loss. A comparative analysis of the described methods is made.

Keywords: audio information, pulse-code modulation, samples, RIFF WAVE, lossless compression, MP3, bitrate, wavelet, fractal compression.

Сергей ЛУКЪЯНЧИКОВ

Николаев

**ИССЛЕДОВАНИЕ СОВРЕМЕННЫХ МЕТОДОВ СЖАТИЯ
ЗВУКОВОЙ ИНФОРМАЦИИ С ПОТЕРЯМИ И БЕЗ ПОТЕРЬ**

В работе исследованы методы архивации которые можно применить для сжатия звуковой информации с потерями и без потерь. Произведен сравнительный анализ изложенных методов.

Ключевые слова: звуковая информация, импульсно-кодовая модуляция, сэмплы, RIFF WAVE, сжатие без потерь, MP3, битрейт, вейвлеты, фрактальное сжатие.

Стаття надійшла до редколегії 23.10.2017