

Київський національний університет імені Тараса Шевченка



Колєнов С. О.

МЕДІЙНІ ТЕХНОЛОГІЇ В РАДІОФІЗИЦІ

Частина 1

Людина і звук

**Навчально-методичний посібник до лекційного курсу
для студентів радіофізичного факультету**

Київ 2013

УДК 0047 (381.3)

Медійні технології в радіофізиці: Навчально-методичний посібник до лекційного курсу для студентів радіофізичного факультету. Частина 1: Людина і звук / С. О. Колєнов. – Київ: Радіофізичний факультет Київського національного університету імені Тараса Шевченка, 2013. – 94 с.

Рецензент: заступник директора Інституту проблем реєстрації інформації, член-кореспондент НАН України, доктор технічних наук, проф. Крючин А. А.

Навчально-методичний посібник до лекційного курсу складений у відповідності з навчальною програмою дисципліни "Медійні технології в радіофізиці", що викладається студентам радіофізичного факультету. Перша частина цього посібника присвячена медійним технологіям в області звуку. Вона містить п'ять глав з основними відомостями про особливості сприйняття, обробки та використання засобів відтворення звукової інформації людиною. Посібник спрямований на допомогу студенту в оволодінні необхідними знаннями та навичками при підготовці до заліку.

Ухвалено Вченою радою радіофізичного факультету
Київського національного університету імені Тараса Шевченка
(протокол № 4 від 11.11.2013)

© Видавництво радіофізичного факультету
Київського національного університету
імені Тараса Шевченка, 2013
© Колєнов С. О., 2013

Вступ

Даний навчально-методичний посібник написаний у відповідності з навчальною програмою дисципліни "Медійні технології в радіофізиці", що викладається у вигляді лекцій для студентів радіофізичного факультету спеціальності 6.040204. Даний курс лекцій розрахований на один семестр, складається з двох модулів та закінчується заліком. Перший модуль присвячений медійним технологіям в області звуку, а другий – медійним технологіям в області візуалізації інформації. Відповідно до модулів посібник також розділений на дві частини. В цій частині розглядаються основні поняття та принципи, пов'язані з особливостями сприйняття, обробки та використання засобів відтворення звукової інформації людиною.

Сьогодні немає потреби пояснювати наскільки важливими є інформаційні технології взагалі і, зокрема, їх частина – медійні технології. З кожним днем кількість інформації, яку людина отримує з використанням цих технологій, постійно зростає. В зв'язку з цим актуальними є задачі оптимізації та удосконалення технологій отримання інформації з метою полегшення її сприйняття людиною з одночасним забезпеченням захисту здоров'я людини від негативних факторів, пов'язаних із зростанням інтенсивності та об'ємів як корисної інформації, так і різноманітних фонових шумів. Особливо це стосується звукової інформації. Однак, розв'язок цих задач є неможливим без вивчення особливостей людського слуху, що стосуються сприйняття звуків.

Посібник не ставить за мету надання вичерпної інформації про всі аспекти функціонування слухового аналізатора людини чи тонкощі технічної реалізації приладів обробки та відтворення звукових сигналів. Ті, хто цікавляться цими питаннями, можуть за додатковою інформацією звернутися до переліку рекомендованої наприкінці літератури. Головна мета – пояснити студенту простими словами, без використання громіздких формул та складних теорій, як працюють ті методи, якими ми користуємося кожен день, коли, наприклад, слухаємо музику, що відтворюється MP3 програвачем, чи розмовляємо по мобільному телефону. Крім того, матеріал, зібраний в даному посібнику, може стати додатковою основою для подальшого вивчення таких дисциплін як "Телекомунікаційні технології" та "Цифровий зв'язок". Посібник можна також рекомендувати і іншим студентам радіофізичних та споріднених з ними факультетів університетів.

Глава 1. Людський слух

Поява персональних комп'ютерів дала значний поштовх розвитку медійних технологій, що призвело до революційних змін в таких областях як освіта, наука, мистецтво, комп'ютерні ігри тощо. Разом з представленням текстової, графічної або відео інформації медійні технології неможливо уявити без використання звуку. Сприйняття людиною звукової інформації є не менш важливим питанням, ніж сприйняття візуальної інформації. Саме завдяки звукам люди можуть спілкуватися один з одним та передавати повідомлення швидким та зручним способом. І хоча теж саме можна робити і використовуючи лише зорове сприйняття, однак, зниження чи втрата слуху у людини призводить до її суттєвого обмеження у сприйнятті навколишнього світу та відчувається так само болісно, як втрата зору. Необхідно зазначити, що сьогодні зниження слуху зустрічається частіше, ніж це може здаватися. Через все більш зростаючі оточуючі нас шуми, зниження слуху спостерігається у все більш молодих груп населення. Сумна статистика останніх років засвідчує:

- близько 500 млн. людей у світі мають порушення слуху;
- у США 1/12 частина населення у віці 30 років та 1/8 – у віці 50 років має зниження слуху;
- більш ніж 1/3 усіх випадків зниження слуху викликано шумом: голосна музика, шум на роботі, голосний звук у місцях відпочинку (дискотека, нічний клуб);
- тільки у 10% випадків зниження слуху може бути усунено хірургічним шляхом;
- рівень шуму у 80 – 90 дБ протягом 8 годин вважається небезпечним для здоров'я. Отже різні винаходи людини, якими ми користуємось та які дають звук за рівнем більший ніж 80 дБ, є потенційно небезпечним при тривалому впливі.

Тому при розробці техніки, яка пов'язана з генеруванням звуку, необхідно чітко усвідомлювати причини, що можуть призвести до зниження слуху. У 1979 році конструктори першого портативного касетного магнітофону Walkman фірми Sony навряд чи задумувалися над тим, що їх винахід через декілька років призведе до появи цілих поколінь людей, що втратили слух змолоду. Наслідком цього стало прийняття закону про обмеження максимального звукового тиску, що генерується в навушниках портативних плеєрів, до рівня 100 дБ. Деякі

компанії навіть вбудовують в свої плеєри системи примусового обмеження гучності. Тим не менш, користуються такими системами далеко не всі.

З цього всього випливає, що людський слуховий апарат – це достатньо вразливий, але водночас і дуже складний орган. Слуховим органом людина сприймає дію зовнішнього середовища, виражену у вигляді звукових коливань, які спричиняють фізичний тиск на барабанну перетинку. Кількість інформації, що людина отримує за допомогою слуху складає біля 10% від тої інформації, що отримується через органи зору. Людина може сприймати до 300000 різних відтінків звуку та шуму в діапазоні від 16 до 20000 Гц. Людське вухо здатне підсилити звук майже у 200 разів. При цьому підсилюються тільки слабкі звуки, сильні – послаблюються. Також людський слух має певні особливості пов'язані з наступними ефектами:

- важко визначити напрямок звуків дуже високої та дуже низької частоти;
- слух дає тільки приблизний напрямок звуку (точність визначення напрямку складає декілька градусів);
- відбиття звуку від навколишніх об'єктів може змінювати сприйняття напрямку звуку;
- багатократні джерела одного і того ж звуку можуть створювати уявлення, що звуки виходять з різних напрямків, або приховувати положення деяких джерел звуку;
- декілька різних звуків, що йдуть з різних боків, при накладанні можуть давати сприйняття зовсім іншого звуку.

Всі ці особливості людського слуху, а також знання будови та принципів функціонування людського вуха на даний час широко використовуються у сучасних медійних технологіях.

1.1. Будова людського вуха

Людське вухо – це орган слуху та рівноваги. Воно складається з трьох частин: зовнішнього, середнього та внутрішнього (рис. 1). Їхній злагоджений ансамбль сприймає та перетворює звукові сигнали на біоелектричні імпульси, що передаються у мозок за допомогою нервових волокон. Головні частини слухового апарату – середнє вухо та внутрішнє вухо – розташовані у вигляді маленьких порожнин у скроневій кістці.

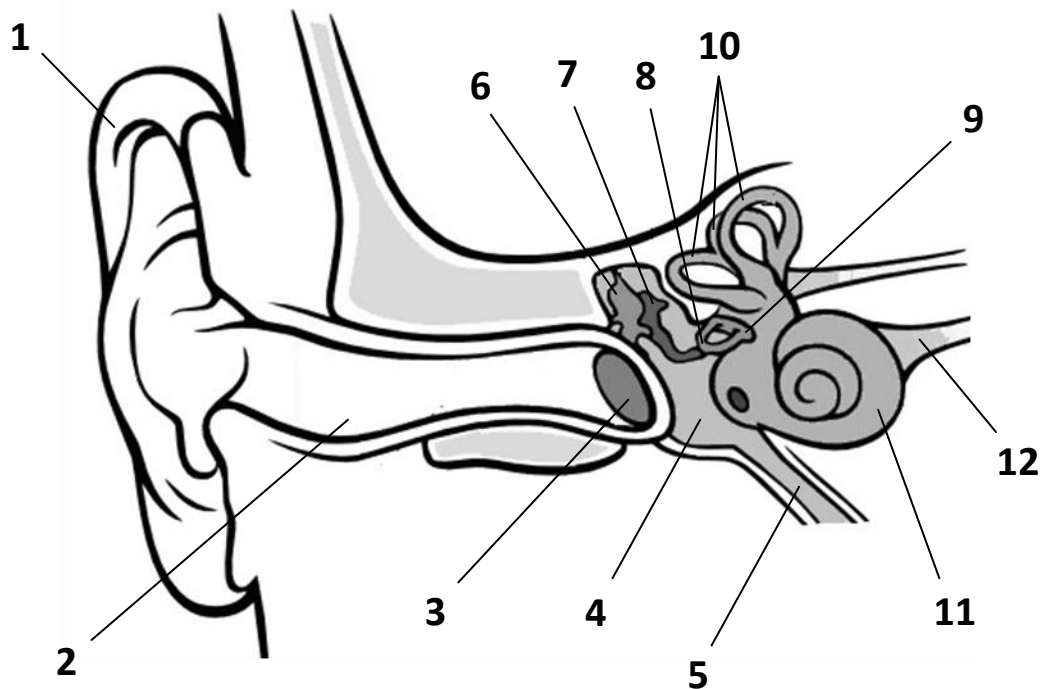


Рис. 1. Будова людського вуха: 1 – вушна раковина, 2 – слуховий прохід, 3 – барабанна перетинка, 4 – порожнина середнього вуха, 5 – євстахієва труба, 6 – молоточок, 7 – ковадло, 8 – стремінце, 9 – овальне вікно, 10 – напівкružні канали, 11 – равлик, 12 – слуховий нерв

Зовнішнє вухо складається з вушної раковини (1) та слухового проходу (2). Функціонально воно призначено для уловлювання та фокусування звукових хвиль (що потрібно для поліпшення слуху), а також для захисту середнього та внутрішнього вуха від механічних пошкоджень. Вушна раковина як радар направляє звукові хвилі до слухового проходу. Крім того, вона дає можливість розрізняти чи надходить звук з переду, чи ззаду. Водночас здатність розрізняти напрямок поширення звуку дуже мало залежить від вушної раковини. Вже у слуховому проході звук суттєво підсилюється (за рахунок резонансних властивостей). Що ж стосується перетворення звукових коливань повітря в електричні імпульси, то зовнішнє вуха не має до цього процесу жодного відношення.

Слуховий прохід закритий зсередини тонкою мембраною – барабанною перетинкою (3), яка відокремлює зовнішнє вухо від порожнини середнього вуха (4) – невеличкої камери об'ємом біля 1 см^3 . Барабанна перетинка дуже тонка, її товщина всього 0,1 мм. Проте вона достатньо міцна для того, щоб витримати навіть дуже гучні звуки, наприклад звуки гарматних вистрілів. Неправильна воронкоподібна форма та нерівномірний натяг барабанної перетинки роблять можливими її співколивання у відповідь на дію різноманітних тонів звукових хвиль. Оптимальною умовою для коливань барабанної перетинки є однаковий тиск повітря з обох її боків. Це забезпечується тим, що порожнина середнього вуха поєднана із зовнішнім середовищем через носоглотку тонким каналом – евстахієвою трубою (5). При ковтанні або позіханні повітря потрапляє до евстахієвої труби, а звідти до порожнини середнього вуха, що дозволяє підтримувати в ній тиск рівний зовнішньому атмосферному тиску. При різкій зміні зовнішнього тиску (наприклад, в літаку) з'являється відчуття закладених вух. Проте, варто зробити декілька ковтків – та проблема зникає, оскільки тиск вирівнюється через евстахієву трубу.

У порожнині середнього вуха звук багатократно підсилюється за рахунок системи так званих слухових кісточок – молоточка (6), ковадла (7) та стремінця (8), які отримали свої назви відповідно до їх форми та призначення. Ці кісточки проходять крізь всю порожнину. Вони тримаються у висячому положенні двома м'язами, які натягуються при занадто гучних звуках та послабляють роботу кісточок, захищаючи слуховий апарат від травми. Молоточок власною ручкою прикріплений до центра барабанної перетинки і таким чином знаходиться у постійному контакті з нею. Ковадло забезпечує передачу рухів від молоточка до стремінця. Інтенсивність звуку в середньому вусі підвищується в 30 разів завдяки різниці між площею барабанної перетинки (90 мм^2) та площею основи стремінця (3 мм^2). Стремінце своєю основою кріпиться до твердої мембрани – овального вікна (9) у верхній частині перегородки між середнім та внутрішнім вухом.

За перетинкою овального вікна розташоване внутрішнє вухо, в якому знаходиться лабіринт. Головними частинами лабіринту є напівкругні канали (10) та равлик (11). Напівкругні канали є органом рівноваги, який визначає орієнтацію організму у просторі. З точки зору функції слуху найважливішим органом є равлик. Равлик – це спіральний орган, наповнений рідиною і обладнаний рецепторами та нервовими закінченнями. Коливання

овального вікна породжують хвилі в рідині, які і сприймаються відповідними рецепторами спірального органу. У людини канал равлика складає біля 35 мм (2,5 оберти спіралі). Анатомія равлика вуха та механізм функціонування внутрішнього вуха дуже складні та описані у [1, 2]. Важливим є те, що всередині равлика знаходиться біля 20000 чутливих волосинок, які "підключені" за допомогою нервів до головного мозку. Ці волосинки і є звуковими рецепторами. Саме вони сприймають звукові хвилі та перетворюють їх в електричні імпульси, які по слуховому нерву (12) передаються у головний мозок. Волосинки мають різну довжину. При цьому низькі звуки впливають на довгі волосинки, а високі – на короткі.

1.2. Розпізнавання звуків

Загалом, слуховий аналізатор людини чутливий до висоти, сили та тембру звука. Висота звука визначається кількістю коливань джерела звуку за 1 секунду (або частотою коливань). Орган слуху людини чутливий до звуків в межах від 16 до 20 тис. коливань за секунду. Найбільша слухова чутливість знаходиться в межах 2000 – 3000 коливань за секунду.

Фізіологічно людське вухо має близько 20 тис. резонаторів, налаштованих на різні частоти. Якщо у вхідному звуковому сигналі втримуються коливання з частотою близькою до частоти настроювання одного з резонаторів, на виході з'являється електрична напруга пропорційна потужності коливання, яка по нервових волокнах надходить у мозок для подальшого аналізу. При цьому потрібно зазначити, що вибірковість слуху є незначною. Смуга пропускання резонаторів слухового аналізатора змінюється від 35 Гц на низьких частотах до 200 Гц на високих частотах. Вхідний сигнал збуджує не тільки сам резонатор, точно налаштований на його частоту, але і декілька сусідніх. Проте, при подальшій обробці в мозку це "розщеплення" усувається. Внаслідок обмеженої кількості нервових закінчень у равлику, людина розрізняє у всьому частотному діапазоні не більше 250 градацій частоти. При цьому кількість цих градацій різко зменшується із зменшенням інтенсивності звуку і в середньому дорівнює 150. Навіть найкращі музиканти не помічають різниці між звуками, частоти яких відрізняються менш ніж на 4%. Крім того, людський слух менш чутливий до

найнижчих і найвищих частот. Вважається, що переважна більшість людей не відчуває частоти нижче 10 Гц і вище 25 кГц.

Слухове сприйняття у людини збільшується стрибками із збільшенням кількості збуджених нервових клітин. Такі стрибки зветься порогом розрізнення інтенсивності. Кількість таких стрибків на середніх частотах не перевищує 250. Таким чином, по амплітуді людський слуховий аналізатор має дискретне сприйняття звуку.

Місцезнаходження джерела звуку визначається за допомогою врахування декількох складних і одночасно тонких звукових параметрів. Наприклад, звук, джерело якого перебуває праворуч, досягає правого вуха на декілька мілісекунд раніше, ніж лівого. Крім того, тембр та гучність звуку, що сприймаються в лівому вусі, будуть іншими, оскільки звукова хвиля огинає голову і лише потім потрапляє у відповідну вушну раковину.

Ще у 1714 році було помічено дуже цікаве явище: коли на скрипці голосно програють дві ноти, можна чітко відчуті третій тон. Це явище викликало великий інтерес серед музикантів та вчених. Було встановлено, що ці фантомні тони виникають безпосередньо в слуховій системі. Взагалі ж процес аналізу звуку вважається ще не до кінця вивченим.

Контрольні запитання

1. Яке місце займає звук у медійних технологіях?
2. Чому знання особливостей людського слуху та принципів функціонування людського вуха є важливим для розвитку та реалізації медійних технологій?
3. З яких основних частин складається орган слуху людини?
4. Що входить до складу внутрішнього вуха?
5. Яка частина вуха сприймає та підсилює звукові сигнали?
6. Що є оптимальною умовою для коливань барабанної перетинки всередині вуха?
7. Яким чином відбувається розпізнавання звуків в слуховому аналізаторі людини?
8. Яким чином людина розрізняє звуки за інтенсивністю?
9. Які найнижчу та найвищу частоти може чути людина?
10. В якому діапазоні частот людина має найбільшу слухову чутливість?

Глава 2. Сприйняття звуку людиною

Вивченням слуху та сприйняттям звуку займається наука під назвою психоакустика. У цій науці на базі суб'єктивних спостережень встановлюються закономірності та взаємозв'язки між об'єктивними характеристиками звуку та відчуттям його сприйняття. Ці взаємозв'язки лише іноді описуються за допомогою емпіричних формул. Частіше вони представляються у вигляді графіків, а іноді мають лише описовий характер, наприклад, для тембру звуку.

Вивчення психоакустики радіоінженерами та особливо спеціалістами в області звукотехніки необхідно тому, що результати досліджень в цій області використовуються в цілій низці розділів звукотехніки.

В області аналогової звукотехніки результати отримані у психоакустиці дозволяють розробляти регулятори рівня тембру, акустичні голівки та акустичні системи, шумопригнічувачі, еквалайзери та підсилювачі потужності з врахуванням слухового сприйняття. Без знань в області психоакустики, наприклад, неможливо зрозуміти, чому ми не чуємо 20%-ві нелінійні спотворення гучномовців на низьких частотах і, з іншого боку, чому нас вкрай дратують нелінійні спотворення лише у 0,1%, що виникають у транзисторних підсилювачах.

У цифровій звукотехніці досі існує потреба зниження шуму квантування, який виникає при аналогово-цифровому перетворенні. Цей шум дратує слух набагато сильніше ніж шум аналогової апаратури. Істотне зниження шуму квантування при розробці цифрової апаратури можливе лише з врахуванням особливостей людського слуху.

У системах зв'язку та радіомовлення вкрай обмежені можливості вибору вільних частотних діапазонів, тому гостро постає задача зниження швидкості цифрових звукових потоків без зниження суб'єктивної якості звучання.

В області цифрового звукозапису для підвищення якості відтворення звуку потрібне збільшення частоти дискретизації та числа розрядів без збільшення розмірів носія і скорочення часу запису. Для цього потрібно здійснювати суттєве стиснення звуку та зменшення швидкості цифрового потоку. За результатами досліджень в області психоакустики було розроблено багато систем стиснення цифрових аудіо даних, що основані на різних моделях слухового сприйняття: MASCAM, MUSICAM, ATRAC, ASPEC, AC-3, MP3 та інші.

Розробка стереофонічних звукових систем відбувається на базі знань про бінауральні особливості слуху людини. Така особливість слуху дозволяє сприймати об'ємне звучання з локалізацією джерел звуку у просторі.

2.1. Сприйняття звуку за частотою

Частотна вибірковість слухового аналізатора представляє значний інтерес, оскільки від цього параметра залежать вимоги до електроакустичної апаратури. Для кількісної оцінки цього параметра вводять поняття висоти звуку. Можливість визначення висоти звуку є найважливішою властивістю слухової системи людини. Вона має велике значення для ідентифікації та класифікації звуків в оточуючому середовищі.

У відповідності з міжнародним стандартом ANSI-994 (American National Standards Institute) висота звуку – це *атрибут слухового сприйняття в термінах, в яких звуки можна розташувати за шкалою від низьких до високих*. Дуже часто висоту звуку ототожнюють з його частотою. І це здебільшого має сенс, тому що висота звуку залежить головним чином від його частоти. Проте, потрібно пам'ятати, що це не одне і те саме, тому що окрім частоти, висота звуку також залежить від звукового тиску та від форми хвилі, хоча і у меншій мірі.

Таким чином, висота звуку задає лінійну класифікація звукових сигналів, на відміну від гучності, яка має лише відносну класифікацію – більше чи менше.

Перш за все необхідно відмітити, що слухова система здатна розрізняти висоту звуку лише для періодичних сигналів. Якщо це складний звук, то його висоту слухова система визначає по основному тону, але тільки якщо він має періодичну структуру, тобто його спектр складається з гармонік. Якщо ця вимога не виконується, то визначити висоту тону слухова система не здатна. Наприклад, звуки таких інструментів як тарілки, гонги та ін. не мають певної висоти.

Висота звуку вимірюється у спеціальних одиницях – *мелах*. Одна тисяча мел дорівнює висоті звуку з частотою 1 кГц, що відчувається при рівні інтенсивності звуку у 40 дБ. У даний час на основі ґрунтовних експериментів, у процесі яких слухачеві пропонувалося розташувати за висотою два звуки різної

частоти, встановлена емпірична залежність висоти тону сигналу від його частоти F , яка показана на рисунку 2.

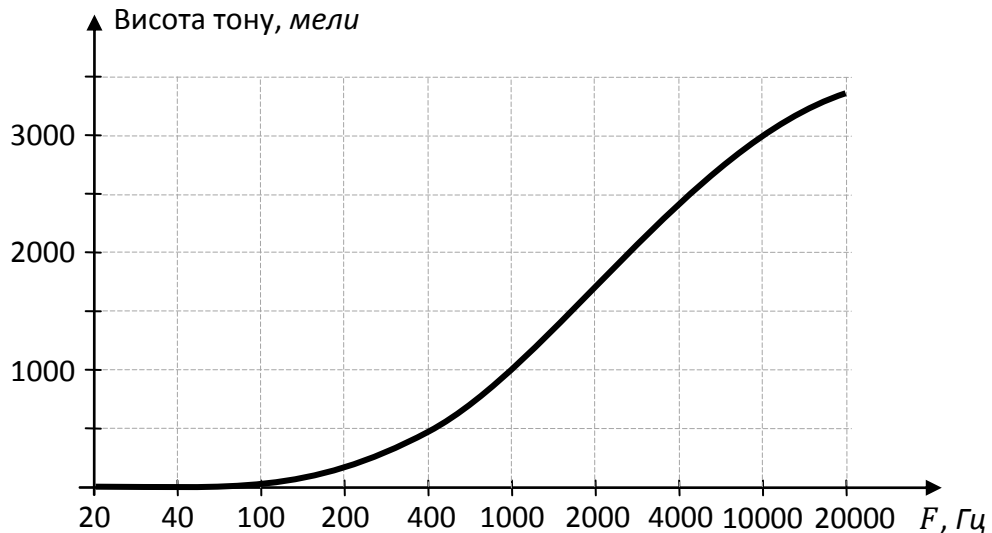


Рис. 2. Залежність висоти звуку від його частоти

Як видно з графіка, зв'язок висоти звуку з частотою є нелінійним. При збільшенні частоти, наприклад, втричі (від 1000 до 3000 Гц) висота тону змінюється лише вдвічі (від 1000 до 2000 мел). Нелінійна залежність особливо виражена на низьких та високих частотах. Всередині діапазону частот, які відчуває людина, зміна висоти тону пропорційна логарифму частоти.

Чутливість до висоти тону пов'язана не тільки з частотою, а також з інтенсивністю звуку та його тривалістю. Як показали дослідження, із збільшенням інтенсивності звуку гучні низькі звуки здаються ще нижчими, а високі звуки здаються трохи вищими. Тільки для середніх частот (1 – 2 кГц) вплив інтенсивності є непомітним. Варто зазначити, що ця залежність дуже незначна і для складних музичних тонів майже непомітна.

З технічної точки зору також необхідно відмітити, що велика точність слуху за частотою висуває доволі жорсткі вимоги до точності ходу стрічкопротяжних механізмів магнітофонів, а також до точності швидкості та стабільності обертання дисків при запису звуку.

2.2. Критичні смуги слуху

При сприйнятті звуку слуховий апарат людини розділяє його на частотні групи, що називаються критичними смугами. Ця особливість слуху є однією з найважливіших його особливостей. Поняття частотна група та критична смуга слуху є тотожними. В діапазоні від 20 до 16000 Гц кількість критичних смуг дорівнює 24. Ширина цих смуг змінюється від низьких до високих частот нерегулярним чином. Графік зміни ширини критичної смуги $\Delta F_{\text{кр}}$ від середньої частоти звуку $F_{\text{сер}}$ показаний на рисунку 3.

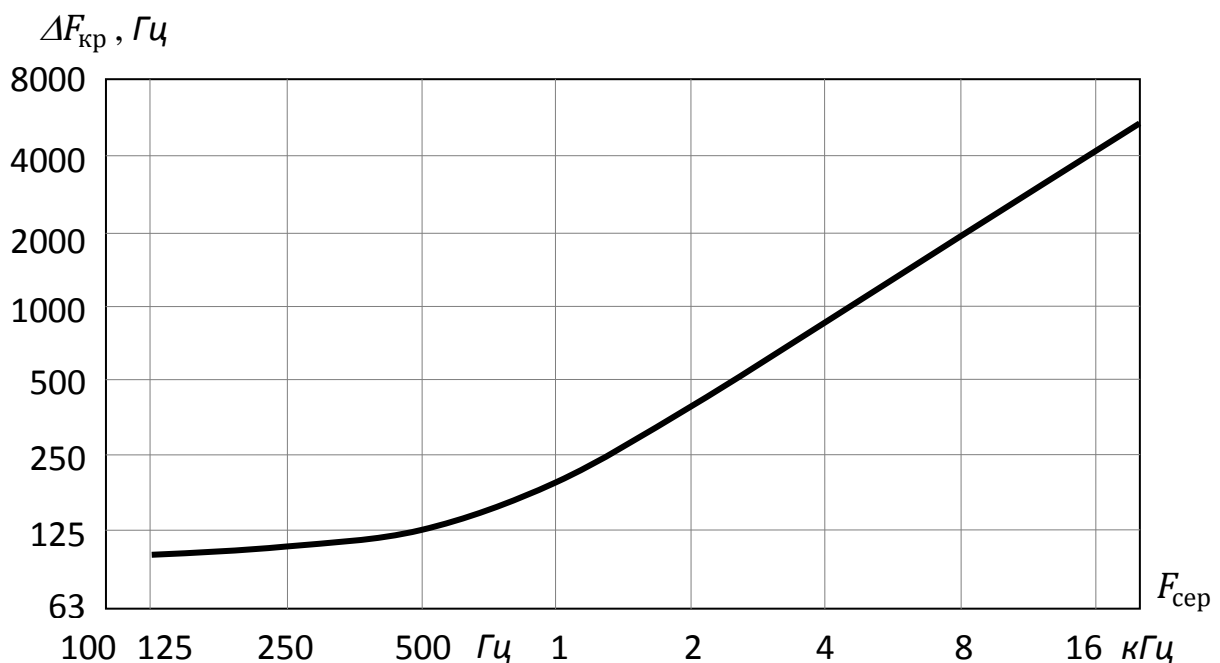


Рис. 3. Графік залежності ширини критичної смуги слуху від її середньої частоти

Стандартом MPEG встановлені нижні межі критичних смуг слуху: 20, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4440, 5300, 5400, 7700, 9500, 12000, 16000 Гц.

Критичні смуги слуху не залежать від рівня інтенсивності звуку. В області до 500 Гц ширина частотних смуг майже не залежить від середньої частоти звуку і дорівнює приблизно 100 Гц. В області вище 500 Гц вона збільшується пропорційно середній частоті (рис. 3). При цьому дотримується відносна

постійність ширини смуги $\Delta F_{кр}/F_{сер} = const$. Ширина критичної смуги при цьому визначається рівністю

$$\Delta F_{кр} = 0,2 F_{сер}.$$

В межах частотних смуг слух інтегрує збудження по частоті та не розрізняє тонкої структури збудження. На базі цього при субсмуговому кодуванні звуку в межах кожної частотної смуги достатньо передавати лише одну максимальну складову спектру. З тієї ж причини слух реагує не на загальну потужність звуку, а на потужність звуку в критичних смугах. Таким чином, при дії широкосмугового звуку слух ніби перетворює суцільний спектр у дискретний. Такий спектр складається з кінцевої кількості складових по кількості критичних смуг слухового апарату.

Якщо ширина спектру вузькосмугового звуку є меншою від ширини відповідної критичної смуги слуху, то рівень гучності в цій смузі визначається лише загальною енергією звуку і зовсім не залежить від характеру розподілу інтенсивності в смузі. Вона може бути розподілена рівномірно або зосереджена в частині смуги, або бути у вигляді одного тону.

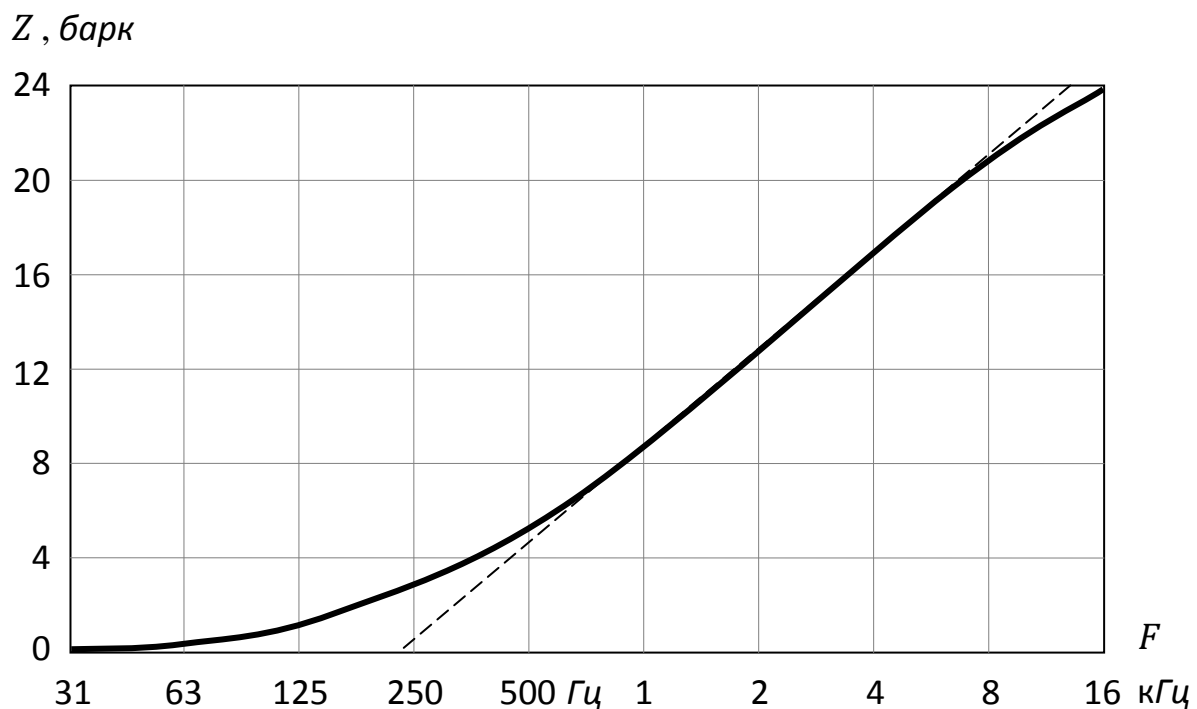


Рис. 4. Графік переведу висоти тону з октавної шкали частот в шкалу висоти тону в барках

Критичні смуги слуху виявилося зручно використовувати як одиницю суб'єктивної висоти тону, яку назвали барк. На даний час у психоакустичних моделях слуху замість частотних шкал використовуються шкали висоти тону звуку у барках. Ті ж самі шкали використовуються при розрахунках коефіцієнтів маскування. На рисунку 4 наведений графік переводу частотної шкали висоти тону в октавах в шкалу висоти тону в барках. По осі абсцис відкладена висота тону в октавній шкалі частот, а по осі ординат – висота тону Z в барках.

2.3. Поріг відчуття звуку та больовий поріг

Для того, щоб звук був почутий, необхідно, щоб амплітуда коливань волокна основної мембрани слухового аналізатора людини була достатньою для подразнення нервового закінчення, яке відразу ж почне надсилати електричні імпульси до слухового центру мозку.

Експериментально було визначено, що для забезпечення відчуття звуку з частотою 1 кГц у повній тиші амплітуда тиску звуку поблизу людського вуха повинна досягати лише $2,84 \cdot 10^{-5}\text{ Н/м}^2$, що складає лише $2 \cdot 10^{-10}$ від атмосферного тиску. Інтенсивність відповідної плоскої хвилі в повітрі при цьому складає 10^{-12} Вт/м^2 . Цікаво відмітити, що при цьому амплітуда зміщення частинок повітря є меншою ніж десята доля радіуса молекули. При цьому величина випадкових флуктуацій сили тиску на барабанну перетинку, що пов'язана з тепловим молекулярним рухом, лише у 5 – 10 разів менше сили тиску звуку, який помітний у повній тиші.

Величина звукового тиску, яка ледве помітна на слух при відсутності усіляких інших заважаючи шумів та звуків, називається пороговою величиною звукового тиску або, скорочено, порогом чутності. Поріг чутності дослідники намагалися визначити неодноразово. Було виявлено, що у різних людей цей поріг може суттєво відрізнятися. Ці відмінності мають випадковий характер для групи осіб однакового віку, що мають нормальний здоровий слуховий орган.

Такі статистичні дослідження проводилися у США (1938-1939 рр.), Англії (1956-1957 рр.) та СРСР (1958 р.). В результаті на основі міжнародної угоди як стандарт прийнята крива залежності порогу чутності від частоти для чистого синусоїдального звуку, що наведена на рисунку 5 [3]. Для визначення цієї

залежності залучалися особи у віці від 18 до 23 років з свідомо здоровими органами слуху.



Рис. 5. Залежності порогу чутності від частоти для чистого синусоїдального звуку

Як видно з рисунку, поріг чутності сильно залежить від частоти. Графік залежності мінімального порогу чутності від частоти отримав назву абсолютний поріг чутності. Цей графік має ділянку найвищої чутливості в області частот від 2 до 4 кГц, де звуки помічаються при звуковому тиску навіть меншому за $2 \cdot 10^{-5}$ Па. У той же час на нижніх та верхніх звукових частотах поріг чутності істотно збільшується. Але, як би ми не збільшували інтенсивність звуку на частотах вище 20 кГц, відчуття звуку не виникає, оскільки звуки з частотою більшою за 20 кГц лежать за межами частот чутних звуків більшості людей. Так само відбувається і зі звуками з частотою нижчою за 16 – 20 Гц. Частоти нижче 16 Гц називають інфразвуковими, а частоти вище 20 кГц – ультразвуковими.

Діапазон гучності сприйняття звуків дуже великий. Для порівняння на рисунку 5 також наведені типові діапазони звукового тиску, характерні для звуків, що передають музику та звичайну людську мову. Верхня межа чутності, яку називали больовий поріг, відноситься скоріше до порогу дискомфорту та далі – до порушення слуху, контузії і т. і. Ця межа залежить від того, як довго за

часом ми слухаємо звук. Вухо здатне переносити короточасні суттєві підвищення гучності звуку (навіть за межі больового порогу) без наслідків. Проте, довготривале сприйняття гучних звуків (гучність яких навіть не досягає больового порогу) може викликати втрату слуху.

2.4. Диференційний поріг сприйняття інтенсивності звуку

Диференційний поріг чутності звуку по інтенсивності визначає різниця інтенсивності двох звуків однієї і тієї ж частоти, що розрізняється на слух. Проте, сила звуку, що виражена у вигляді звукового тиску або потоку звукової енергії, не є мірою величини відчуття або суб'єктивної сили звуку, яку прийнято називати відчуттям гучності.

Оцінити гучність, як міру суб'єктивної сили звуку, дає можливість психофізичний закон, встановлений Вебером у 1846 році, який полягає у тому, що *відношення мінімально сприйнятого приросту деякого стимулу до початкової його величини є постійним*.

Якщо позначити силу звуку (звукового стимулу) через J , то закон Вебера можна записати у наступній формі:

$$\frac{\Delta J}{J} = \text{const}, \quad (1)$$

де ΔJ – приріст інтенсивності, який потрібно надати звуку з інтенсивністю J , щоб стала помітною різниця в гучності між звуками J та $J + \Delta J$. Величина $\Delta J/J$ складає приблизно 10%. Необхідно підкреслити, що це співвідношення пов'язано з властивостями нервової системи і спостерігається не лише при звукових подразненнях, але і при зорових, дотикових та ін. Саме тому воно носить назву загального фізіологічного закону.

Наступний крок в розвитку теорії Вебера був зроблений у 1860 році Фехнером, який застосував до співвідношення (1) математичну обробку. Фехнер прийняв, що приріст ΔJ є нескінченно малою величиною dJ , а праву частину визначив як пропорційну нескінченно малому приросту відчуття dE (в даному випадку відчуття гучності). Вийшло наступне співвідношення:

$$A \frac{dJ}{J} = dE, \quad (2)$$

де A – деяка довільна стала, що залежить від вибору одиниць відчуття.

Закон Вебера у вигляді (2) є простим диференціальним рівнянням, яке можна проінтегрувати. Після інтегрування (2) отримаємо:

$$E = A \cdot \ln J + C, \quad (3)$$

де C – стала інтегрування.

Вважаючи, що на порозі чутності $E = 0$, а $J = J_0$, можна визначити сталу інтегрування як

$$C = -A \cdot \ln J_0. \quad (4)$$

Тоді звідси отримаємо відомий логарифмічний закон, який носить назву *закону Вебера-Фехнера*, згідно якого відчуття E пропорційно логарифму подразнення:

$$E = A \cdot \ln \frac{J}{J_0} = 2,3 A \lg \frac{J}{J_0}. \quad (5)$$

Величина A визначає масштаб відчуття і тому може бути обрана довільно. Наприклад, умовно можна прийняти $2,3A = 10$. Тоді отримаємо:

$$E = 10 \lg \frac{J}{J_0}. \quad (6)$$

Формула (6) визначає величину в ∂B , яку називають рівнем відчуття. Застосування шкали ∂B зручно тому, що при оцінці величини слухового відчуття мінімально відчутні прирости мають величину порядку 0,5 ∂B .

Внаслідок логарифмічного закону сприйняття та широкого діапазону інтенсивності чутних звуків для об'єктивної оцінки введено поняття *рівня інтенсивності*

$$N_I = 10 \lg \frac{I}{I_0}. \quad (7)$$

За нульовий рівень домовились приймати інтенсивність I_0 , близьку до порогової інтенсивності для нормального слуху на частоті 1000 Гц. Ця інтенсивність, як вже було зазначено раніше, приблизно дорівнює 10^{-12} Вт/м^2 . Таким чином, больовий поріг інтенсивності знаходиться біля 120 дБ.

Оскільки звуковий тиск та інтенсивність звуку пов'язані між собою квадратичною залежністю, *рівень звукового тиску* дорівнює:

$$N_P = 20 \lg \frac{P}{P_0}, \quad (8)$$

де P_0 – звуковий тиск на нульовому рівні, який дорівнює $2 \cdot 10^{-5} \text{ Па}$, що відповідає стандартному порогу чутності людського вуха на частоті 1000 Гц. Якщо підставити це у (8), то отримаємо:

$$N_P = 20 \lg P + 94 \text{ дБ}. \quad (9)$$

Визначений з (8) чи (9) рівень тиску в дБ буде мати таке ж числове значення, як і рівень інтенсивності звуку, що визначений з (7), тільки для чистого тону 1000 Гц, оскільки при цьому значення I_0 співпадає з порогом чутності на частоті 1000 Гц.

З формул (7) та (8) виходить, що кожна двократна зміна інтенсивності відповідає зміні рівня на 3 дБ, а кожна двократна зміна звукового тиску приводить до зміни його рівня на 6 дБ.

Рівні інтенсивності і звукового тиску складного звуку підсумовуються алгебраїчно:

$$N_I = 10 \lg \frac{I_1 + I_2 + I_3 + \dots + I_n}{I_0} = N_{I_1} + N_{I_2} + N_{I_3} + \dots + N_{I_n}. \quad (10)$$

2.5. Пороги чутності при маскуванні

При одночасній дії на слух двох звуків один з них може не прослуховуватись на фоні іншого. Цей ефект називається *маскуванням*.

Маскувальна дія різних звуків виявляється шляхом визначення підвищення порогу чутності досліджуваних тонів $N_{\text{ПЧ}}$ відносно рівня абсолютного порогу

чутності $N_{\text{АПЧ}}$. Коефіцієнт маскування K_M розраховується в дБ, як різниця цих двох рівнів:

$$K_M = N_{\text{ПЧ}} - N_{\text{АПЧ}}.$$

Маскування чистим тоном

На рисунку 6 наведені криві маскування тону в певному діапазоні частот, коли звуком, що заважає, є той самий тон, але з фіксованою частотою $F_M = 1$ кГц при різних рівнях звукового тиску N_M . За цими кривими визначається поріг чутності на конкретних частотах тону, що маскується, та розраховується коефіцієнт маскування.

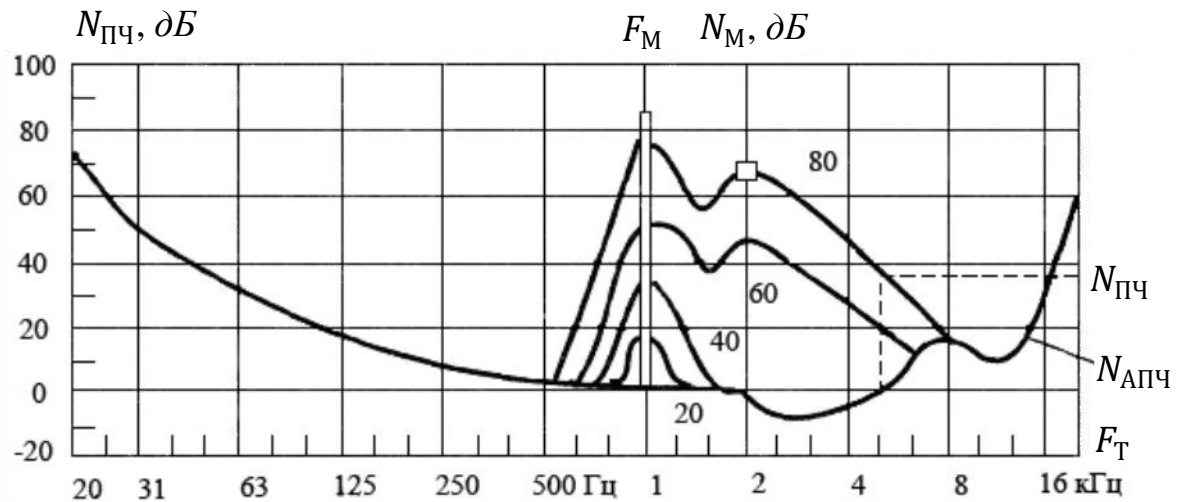


Рис. 6. Криві порогу чутності тону з частотою F_T при маскуванні тоном з частотою $F_M = 1$ кГц та рівнем N_M

Особливості маскування чистим тоном можна сформулювати наступним чином:

- 1) Маскування високе, коли частоти F_T та F_M є близькими. Чим далі по частоті вони розташовані один від одного, тим меншою є величина маскування.
- 2) Величина маскування збільшується по мірі зростання рівня тону N_M , що заважає.
- 3) Криві маскування суттєво несиметричні по шкалі частот. У них різкіший спад у бік низьких частот і плавне спадання у бік високих частот. Тому маскування сильніше проявляється в області високих частот.

4) Зсув звуку, що маскує, по частоті призводить до зміни форми кривої маскування.

Якщо частота тону F_T дорівнює подвійній, потрійній та вище частоті тону F_M , що заважає, то у широкому діапазоні чутності виникають биття, які проявляються у вигляді провалів на графіках кривих маскування. Ці биття виникають через нелінійність слуху і є тим помітніші, чим більшим є рівень тону, що заважає.

Якщо маскування здійснюється відразу декількома тонами (співзвуччям), то крива порогу чутності в усьому частотному діапазоні співзвуччя лежить вище за абсолютний поріг чутності. В цьому випадку криві порогу чутності мають безліч максимумів і зон биття. При малій гучності високий тон добре чутний одночасно з низьким тоном. При високій гучності високі частоти можуть повністю маскуватися низькими.

Маскування вузькосмуговим шумом

При маскуванні тону з частотою F_T вузькосмуговим шумом з центральною частотою F_M і рівнем інтенсивності N_M (Рис. 7) криві маскування приблизно такі ж, як при маскуванні чистим тоном з деякими відмінностями:

- 1) Немає биття і провалу коефіцієнта маскування на середній частоті звуку, що заважає.
- 2) Максимум маскування на 4 дБ нижчий за максимум рівня інтенсивності шуму, що маскує.

Для тональних звуків вище 1 кГц маскуюча дія шуму значна, тоді як частоти на октаву нижче за частоту F_M практично не маскуються. Цей факт є дуже важливим, оскільки при дискретизації часто виникають спотворення нижче спектру основного тону, які не маскуються. Наприклад, на частоті тону 9 кГц 5-та гармоніка помилок квантування з частотою дискретизації 44 кГц утворюють паразитну складову в звуковому спектрі з частотою 1 кГц.

Графіки на рисунку 7 дозволяють розрахувати допустиме відношення сигнал-шум квантування з врахуванням маскування. Річ у тому, що тональний сигнал та вузькосмуговий шум маскують не лише тони, але і широкосмугові шуми квантування з рівномірною спектральною густиною. Так для звуків на частоті 1 кГц, що маскують, з великим рівнем інтенсивності, наприклад, 100 дБ, нижня границя частотної області, де відбувається маскування шумів

квантування, дорівнює приблизно 500 Гц. На цій частоті поріг чутності при маскуванні на 20 дБ вище за поріг чутності у тиші (це видно з графіка на рис. 7).

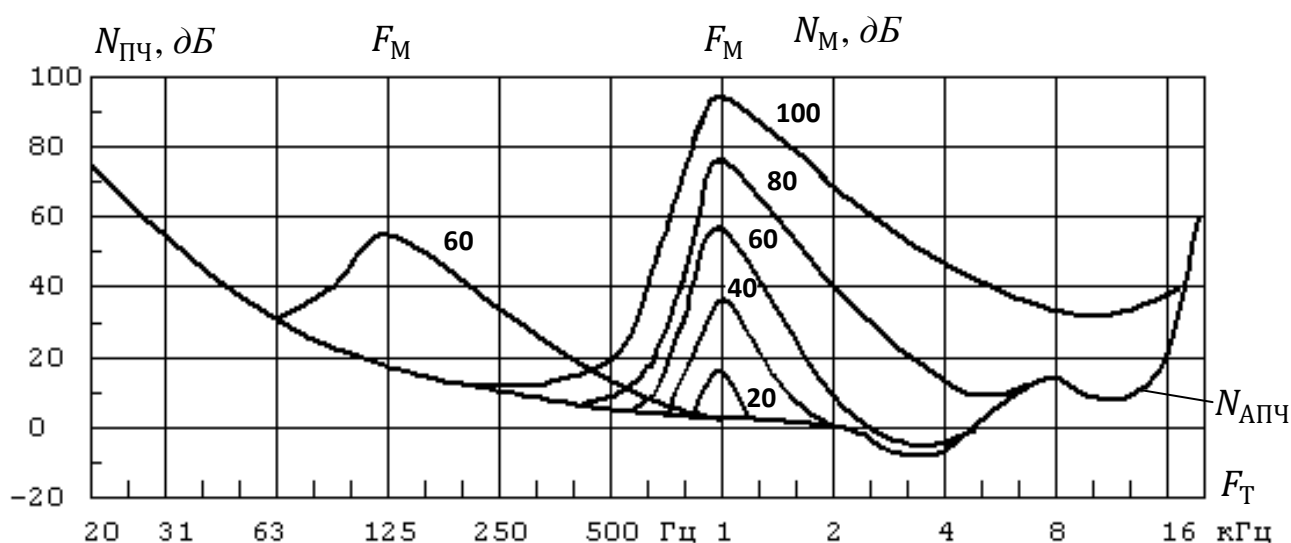


Рис. 7. Криві порогу чутності тону з частотою F_T при маскуванні вузько-смуговим шумом з середньою частотою $F_M = 1$ кГц та рівнем N_M , а також з середньою частотою $F_M = 125$ Гц та рівнем 60 дБ

Маскування шуму квантування тональними звуками та вузькосмуговими шумами має дуже велике значення при кодуванні та передачі звуку, оскільки передавати складові спектру, що маскуються, непотрібно. Це дозволяє суттєво скоротити необхідну кількість розрядів i , відповідно, надлишковість у звуці та швидкості цифрових потоків.

Зсув частоти звуку, що маскує, приводить до зсуву кривої маскування по осі частот (рис. 7). У області вище 1 кГц крива маскування змінюється мало. На частотах нижче 1 кГц ця крива видозмінюється досить сильно. Особливо важливо, що вона розширюється в область низьких частот.

2.6. Рівень гучності та гучність

Рівень гучності

Звуковий тиск та інтенсивність тонального звуку, які можна виміряти приладами, не дають уявлення про відчуття гучності. Між тим, людина може досить точно встановити рівність гучності двох звуків будь-якого спектрального

складу. Ця властивість слуху була використана для введення поняття рівня гучності, що дозволяє визначити гучність одного звуку відносно іншого, прийнятого як опорний.

За одиницю рівня гучності була прийнята величина рівня інтенсивності чистого тону з частотою 1 кГц. Одиниця рівня гучності називається фоном. Кількісно рівень гучності в фонах дорівнює рівню інтенсивності звуку в дБ на частоті 1 кГц.

Визначення рівня гучності тонів інших частот виконується по графіку кривих однакового рівня гучності. Ці графіки називаються ізофонами (рис. 8).

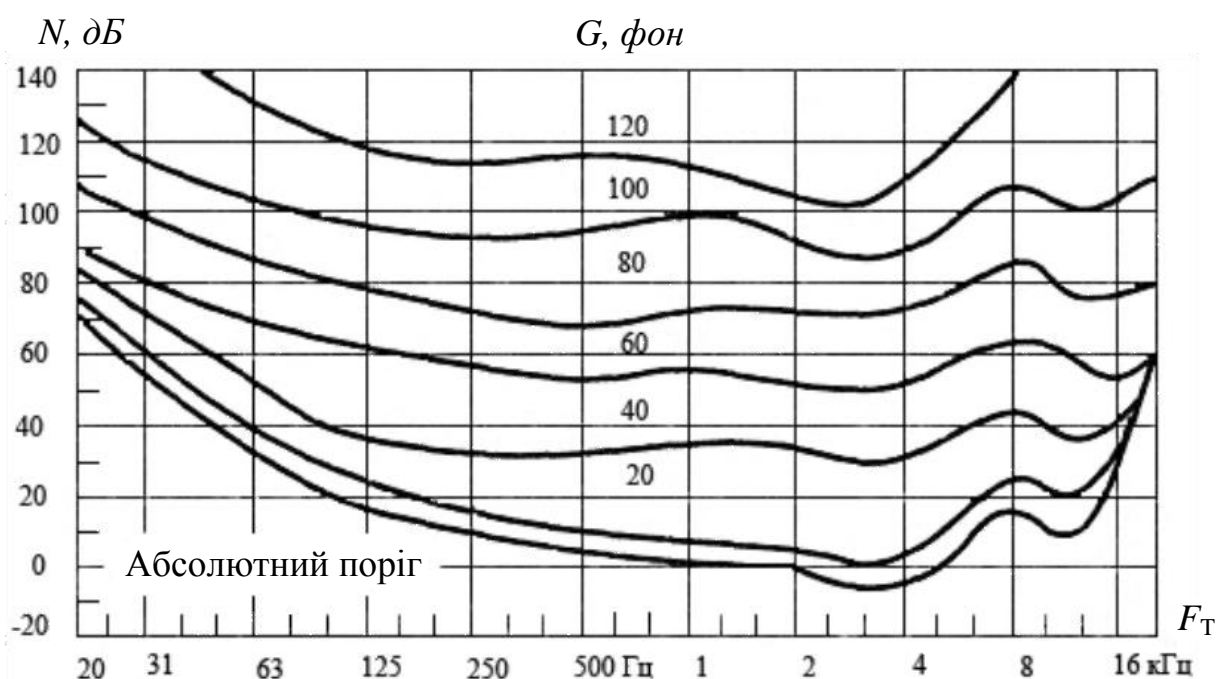


Рис. 8. Криві однакового рівня гучності

Кожна крива отримана шляхом суб'єктивного порівняння гучності синусоїдального звуку довільної частоти з гучністю звуку з частотою 1 кГц. Якщо звук з частотою 100 Гц має рівень 50 дБ, а однаково гучний з ним звук з частотою 1 кГц має рівень 40 дБ, то рівень гучності звуку з частотою 100 Гц дорівнює 40 фон.

Ізофони вперше були отримані у 1933 році та опубліковані в [4]. Пізніше у роботі [5] були представлені точніші вимірювання, які лягли в основу стандарту ISO 226 [6] у 1986 році. У 2003 році стандарт ISO 226 був оновлений з

врахуванням даних, які були отримані в результаті вимірювань при проведенні 12-ти нових міжнародних досліджень.

Для усіх ізофон характерне невелике зниження в області частот 3 — 4 кГц і рівномірний ріст при пониженні частоти нижче 500 Гц. Величина росту на різних рівнях неоднакова. При підвищенні рівня ізофони випрямляються і їх ріст стає менш крутим. Це значить, що частотна характеристика слуху при великих рівнях звукового тиску є більш рівномірною ніж при малих.

При малих рівнях гучності форма ізофон наближається до кривої абсолютного порогу гучності, який відповідає рівню гучності 3 фона.

Щоб отримати уявлення про рівень гучності звуків, з якими нам доводиться повсякденно стикатися, розглянемо декілька прикладів. Гучна розмова на відстані 1 метр відповідає рівню гучності 65 — 70 фон, рівень гучності у вагоні метро під час руху дорівнює 85 — 90 фон, рівень гучності звичайної розмови — 55 — 60 фон, рівень гучності у читальному залі бібліотеки — 25 — 30 фон.

З кривих однакового рівня гучності випливає, що збереження природного звучання, яке відповідає умовам запису звуку, можливе тільки при умові відтворення запису з тим самим рівнем гучності, при якому цей запис здійснювався. Інакше порушуються співвідношення між частотними складовими. При великих рівнях відтворення відбувається суб'єктивне посилення низькочастотних складових. При малих рівнях виникає їх суб'єктивне пригнічення. Для збереження тембру незалежно від рівня в радіотехніці застосовують компенсовані регулятори підсилення, в яких при зменшенні підсилення здійснюється підйом частотної характеристики в області низьких та високих частот. Цей прийом називається тонкомпенсацією.

Гучність

Гучністю називається суб'єктивне відчуття, яке дозволяє слуховій системі розташувати звуки за шкалою від тихих до гучних звуків. Гучність звуку пов'язана перш за все зі звуковим тиском. Зазвичай, чим більше звуковий тиск, тим гучніше звучить акустична система. Проте, це не завжди так. Гучність залежить також від частоти, спектрального складу, тривалості звуку та його локалізації у просторі.

Необхідність введення поняття гучності пов'язана з тим, що на практиці часто цікавить не рівень гучності, виражений у фонах, а величина, яка показує у

скільки разів даний звук гучніше за інший. Також цікавим є питання про те, як додаються гучності двох різних тонів. Наприклад, якщо є два тони різних частот з рівнем гучності 70 *фон*, то це не означає, що сумарний рівень гучності буде дорівнювати 140 *фон*.

Для оцінки гучності прийнята спеціальна одиниця, яка називається сон. Гучність 1 *сон* – це гучність тонального звуку з частотою 1 *кГц* з рівнем 40 *дБ*.

Залежність гучності від рівня звукового тиску є суттєво нелінійною, у неї логарифмічний характер. При збільшенні рівня тиску звуку на 10 *дБ* його гучність зростає у 2 рази. Це означає, що рівням гучності 40, 50 та 60 *фон* відповідають гучності 1, 2 та 4 *сони*.

Розглянемо, як можна розрахувати рівень гучності складного звуку з дискретним спектром. Це дозволяє зробити графік гучності Q тонального звуку в залежності від його рівня гучності G (рис. 9).

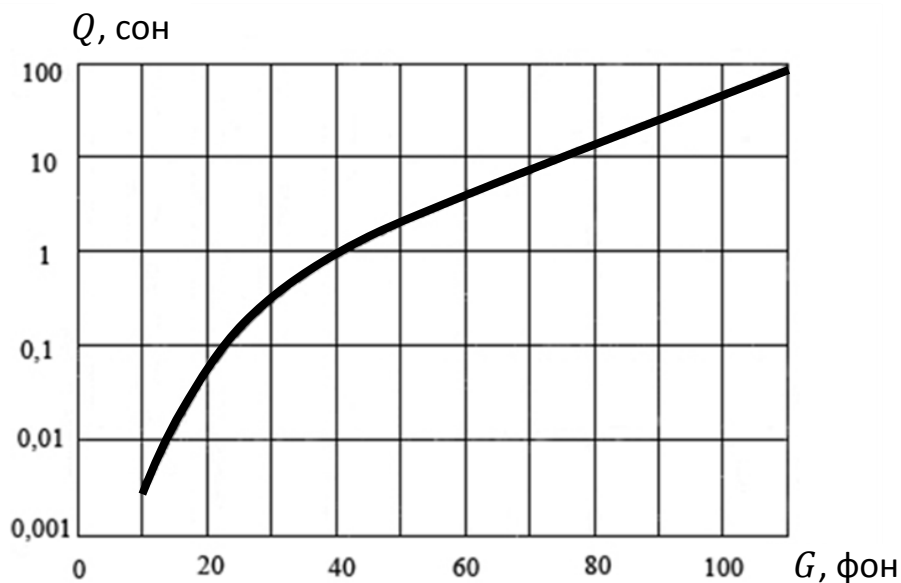


Рис. 9. Графік залежності гучності тонального звуку від його рівня гучності

Наприклад, нехай у нас є три тони: 100, 500 та 3500 *Гц* з рівнями звукового тиску відповідно 84, 77 та 71 *дБ*. Потрібно визначити рівень гучності складного звуку. По кривих однакового рівня гучності (див. рис. 8) визначаємо рівень гучності кожної складової. Всі вони потрапляють на одну криву, тому у всіх один рівень гучності – 80 *фон*. По наведеному на рисунку 9 графіку визначаємо

гучність кожної. Вона відповідає 21,5 *сон*. Сумарна гучність буде дорівнювати відповідно сумі усіх складових – 64,5 *сон*. По цьому ж графіку знаходимо загальний рівень гучності, який складатиме 93 *фона*.

Приведений на рис. 9 графік гучності S для рівнів гучності G вище 40 *фон* добре апроксимується формулою

$$Q = 2^{(G-40)/10} \text{ сон}, \quad G > 40 \text{ фон.} \quad (11)$$

Для менших значень рівнів гучності можна скористатися формулою Стівенса для частоти 1 *кГц* та формулою (9):

$$Q(1\text{кГц}) = 42 P^{0,6} = 4,2 \cdot 10^{0,03G-1,82} \text{ сон}, \quad G < 40 \text{ фон.} \quad (12)$$

Щоб уявити гучність звуків, з якими ми постійно стикаємося у житті, розглянемо таблицю 1 рівнів гучності та гучність найтипівіших звуків.

Таблиця 1. Середній рівень гучності та гучність деяких звуків та шумів, що найчастіше трапляються.

Джерело шуму та місце його вимірювання	Рівень гучності, фон	Гучність, сон
1. Авіаційний мотор на відстані 5 м	116 – 120	346 – 556
2. Гучний автомобільний сигнал на відстані 8 м	95 – 100	57 – 88
3. Електропотяг на естакаді на відстані 6 м	90	38
4. Шум у потязі метро під час руху	85 – 90	25 – 38
5. Трамвай на відстані 10 – 20 м	80 – 85	17,1 – 25
6. Вантажівка у місті на відстані 5 – 20 м	60 – 75	4,35 – 11,4
7. Легковий автомобіль у місті на відстані 5 – 20 м	50 – 65	2,2 – 5,87
8. Свисток міліціонера на відстані 20 м	70	7,95
9. Тиха вулиця (без руху транспорту)	30 – 35	0,36 – 0,62
10. Тишина у саду	20	0,097
Промислові шуми		
1. Удари молота по сталі на відстані 2 – 4 м	110 – 113	215 – 288
2. Загальний шум у ткацькому цеху	96 – 100	62 – 88
3. Деревообробна фабрика	96 – 98	62 – 74

Театри, школи, лікарні		
1. Симфонічний оркестр	80 – 100	17,1 – 88
2. Аплодисменти	60 – 75	4,4 – 11,4
3. Зали в школах під час перерви	55 – 62	3,1 – 4,7
4. Бібліотеки	25 – 30	0,2 – 0,36
5. Лікарні	20 – 30	0,1 – 0,36
6. Гучна розмова на відстані 1 м	65 – 70	5,87 – 7,95
7. Звичайна розмова на відстані 1 м	55 – 60	3,08 – 4,35
8. Шепіт середньої гучності на відстані 1 м	20	0,1

2.7. Адаптація слуху

При дії на барабанну перетинку вуха достатньо тривалого звуку великої інтенсивності гучність, що сприймається, поступово зменшується. Це означає, що під час дії тривалого гучного звуку падає чутливість вуха. Після припинення дії звуку чутливість слуху поступово відновлюється. Це явище називають адаптацією слуху. Фактично адаптація слуху – це здатність слуху підсвідомо пристосовуватися до сприйняття звуків різної гучності. Якщо нижній поріг адаптації слуху практично є необмеженим, то верхнім порогом є звуковий тиск на рівні 140–150 дБ, перевищення якого може викликати у слуховому апараті незворотні зміни, аж до часткової чи повної глухоти.

Явище адаптації слуху пояснюється рисунками 10 та 11. На цих рисунках як приклад приведена дія звуку у вигляді імпульсів тривалістю по 2 хвилини на чутливість слуху. При рівні звукового тиску 94 дБ (рис. 10, а) відбувається плавне пониження рівня гучності, що сприймається, з 94 до 85 фон (рис. 10, б). Швидкість падіння рівня гучності за часом від 0 до 40 секунд висока, потім вона зменшується майже до нуля. При подальшому стрибкоподібному зростанні звукового тиску на 6 дБ (рис. 10, а) рівень гучності спочатку зростає на 9 фон, а потім знов починає зменшуватися (рис. 10, б). Проте, зменшення рівня гучності в цьому випадку помітно менше з 95 до 88 фон. Це означає, що міра адаптації тим вище, чим голосніше звук, що подразнює.

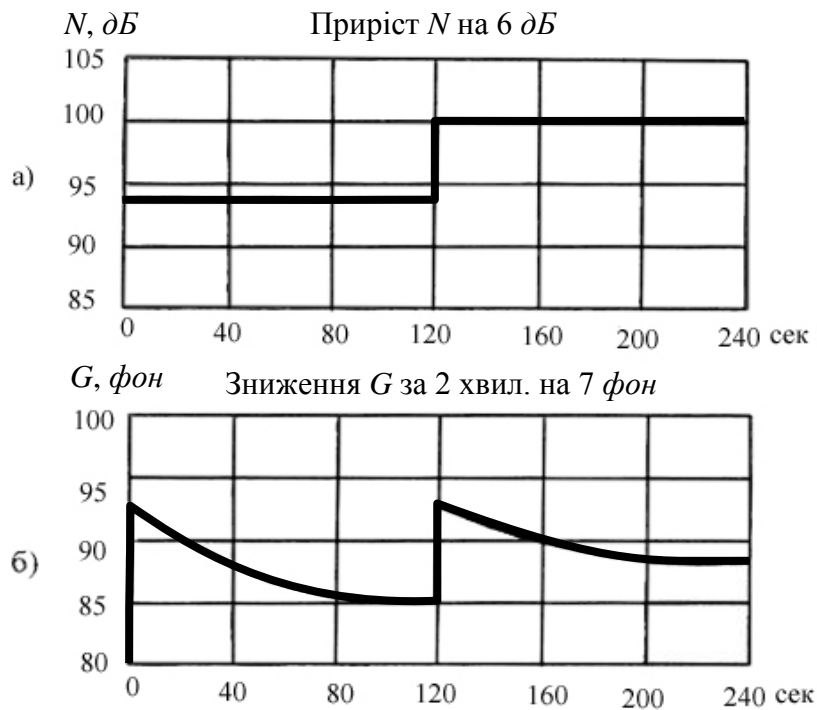


Рис. 10. Графіки адаптації слуху людини при збільшенні звукового тиску на 6 дБ

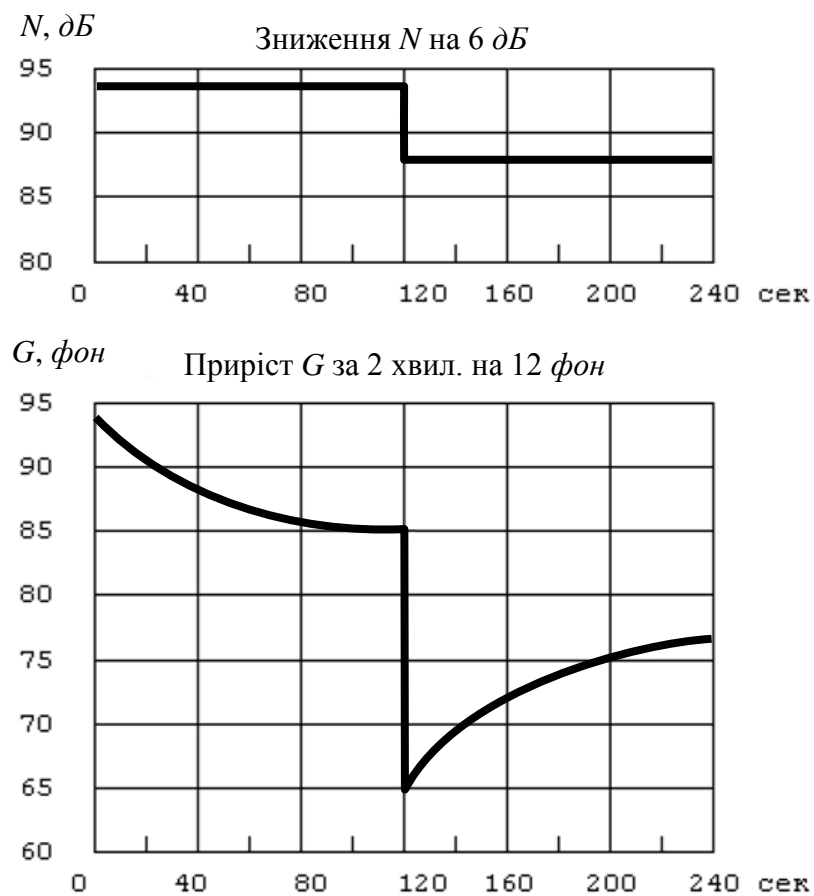


Рис. 11. Графіки адаптації слуху людини при зменшенні звукового тиску на 6 дБ

На рисунку 11 показана зміна сприйманого рівня гучності при стрибкоподібному зменшенні звукового тиску з 94 до 88 дБ. При цьому стрибок зменшення рівня гучності складає майже 20 фон. Потім відбувається адаптація слуху до тиші і чутливість вуха частково відновлюється, а сприйнятий рівень гучності через 2 хвилини зростає на 12 фон.

Дослідження показують, що адаптаційні процеси, що відбуваються у слуховому аналізаторі людини, мають централізовану природу. Про це свідчить зниження слуху не тільки в тому вусі, яке піддається звуковому подразненню, але й у іншому вусі.

Загалом здатність до адаптації є позитивною властивістю слухового апарата, оскільки поліпшується прийом слабких звуків та слуховий орган захищається від перевантажень занадто сильними звуками. Захисна здатність вуха володіє інерцією – цей механізм починає працювати через 30 – 40 мс після початку звуку. Повний захист слуху від дії високих рівнів звуку не досягається навіть за 150 мс. Тому для слуху найбільш небезпечними є дуже короткі гучні звуки.

Проте, інерційність адаптації може відігравати і негативну роль. Так при передачі голосових повідомлень в каналах зв'язку, якщо інтенсивність звуку в каналі буде різко змінюватись, то деякі частини повідомлення можуть бути незрозумілими чи неправильно інтерпретованими.

2.8. Часові характеристики слуху

Слуховий апарат є інерційним, тому при зникненні звуку слухове відчуття зникає не відразу, а поступово зменшується до нуля. Час, за який відчуття по рівню гучності зменшується на 9 – 10 фон, називається сталюю часу слуху. У середньому вона дорівнює 30 – 50 мс.

Якщо до слухача приходять два короткі звукові імпульси, однакові по складу і рівню, але один з них запізнюється, то вони сприйматимуться зливо, коли запізнення не перевищує 50 мс. При великих інтервалах запізнення обидва імпульси сприймаються окремо. Це явище називається відлунням, воно проявляється, коли різниця ходу прямого та відбитого звуків більше 18 м.

Якщо рівень звуку, що запізнюється, набагато менше рівня першого, то він не буде прийнятий окремо, навіть якщо час запізнення більше 50 мс. Це обумовлено ефектом маскування – відчуття від першого звуку маскує другий.

До часових характеристик слуху відноситься час встановлення висоти тону звуку. Для того, щоб слух міг наближено оцінити висоту тону звуку потрібно щонайменше 2 – 3 періоди коливань. На низьких частотах час встановлення складає біля 30 мс. На високих частотах цей час дещо менший.

Також до часових характеристик слуху відноситься таке явище як часове маскування.

До цього коефіцієнт маскування визначався лише у припущенні, що звуки, які маскують, та звуки, які маскуються, присутні одночасно. Проте, через явище адаптації слуху виникають ситуації, коли досить гучні звуки маскують, роблять практично нечутними, звуки, що сліднують за ними. В деяких випадках маскуються також й попередні звуки.

Такий вид маскування, коли звуки не перекриваються у часі і називається часовим маскуванням. Він розділяється на передмаскування та післямаскування. Післямаскування проявляється на інтервалі часу 100 – 200 мс після закінчення звуку, що маскує. Передмаскування проявляється на значно коротших часових інтервалах – близько 10 мс. Тривалість передмаскування в значній мірі залежить від особливостей слуху конкретних людей. З цієї причини механізми часового маскування при цифровому кодуванні звуку доки не використовуються.

Основні особливості часового маскування:

- післямаскування є ефективнішим, ніж передмаскування;
- вищий рівень маскування спостерігається під час надходження звуку, що маскує, через короткий часовий інтервал слідом за звуком, що маскується;
- маскування виражене сильніше, коли звуки, що маскують, та звуки, що маскуються, потрапляють в одне вухо;
- рівень маскування зазвичай різко падає при збільшенні часового інтервалу понад 15 мс;
- збільшення рівня інтенсивності звуку, що маскує, на 10 дБ викликає зміщення порогу маскування на 3 дБ;
- тривалість звуку, що маскує, впливає на міру передмаскування, але не на післямаскування;

- часове маскування залежить від частотного співвідношення звуків, що маскують, та звуків, що маскуються, маскування проявляється сильніше, коли частоти цих звуків є близькими.

2.9. Бінауральний слух

Бінауральним слухом називається його здатність визначати напрямок приходу звукової хвилі, тобто здатність локалізувати положення джерела звуку у просторі. Ця здатність досягається завдяки просторовій несуміщеності двох вух у поєднанні з екрануючим впливом голови. Це призводить до того, що завжди має місце неідентичність збудження правого та лівого вуха. Цей факт забезпечує людині можливість сприймати просторовий звуковий світ та оцінювати переміщення джерел звуку у просторі.

До основних властивостей бінаурального слуху можна віднести:

- просторову локалізацію;
- ефект передування (полягає у тому, що звуки, які надійшли раніше в межах певного проміжку часу, домінують над звуками, що надійшли пізніше; локалізація при цьому відбувається тільки по сигналу, що випереджає);
- бінауральне сумування гучності (звук, який відчувається двома вухами, має удвічі більшу гучність, ніж для одного вуха; відповідно для двох вух абсолютний звуковий поріг зменшується на 3 дБ);
- бінауральне маскування та демаскування.

Такі можливості слуху досягаються завдяки трьом факторам:

- 1) часовому – виникає через неспівпадіння моментів дії однакових фаз звуку на ліве та праве вухо;
- 2) амплітудному – виникає через неоднаковість величини звукового тиску на праве та ліве вухо внаслідок дифракції звукової хвилі навколо голови;
- 3) спектральному – виникає через різницю у спектральному складі звуків, що сприймаються правим та лівим вухом внаслідок неоднакового екрануючого впливу голови та вушних раковин на низьких та високих частотах.

Локалізація джерел звуку

Зазвичай людина дуже добре розрізняє розташування звуків у горизонтальній площині та за глибиною, а також відчуває просторові

характеристики звукового образу. Внаслідок дифракції звукової хвилі навколо голови слухача та частотно-залежного затухання цієї хвилі з відстанню, звуки до правого та лівого вух слухача надходять неоднаковими. Вони відрізняються за рівнем інтенсивності, за часом надходження, а також є функціями азимута та частоти. Всі ці параметри є носіями інформації про локалізацію джерела звуку. Проте, механізми локалізації джерел звуку по глибині, в горизонтальній та вертикальній площинах розрізняються. Розглянемо ці механізми окремо.

1. Локалізація джерел звуку у горизонтальній площині.

Якщо під деяким кутом φ до площини 1-1 голови слухача знаходиться джерело звуку (рис. 12), то фронтальна локалізація цього джерела включає визначення азимуту (кута φ) та відстані l до джерела звуку.

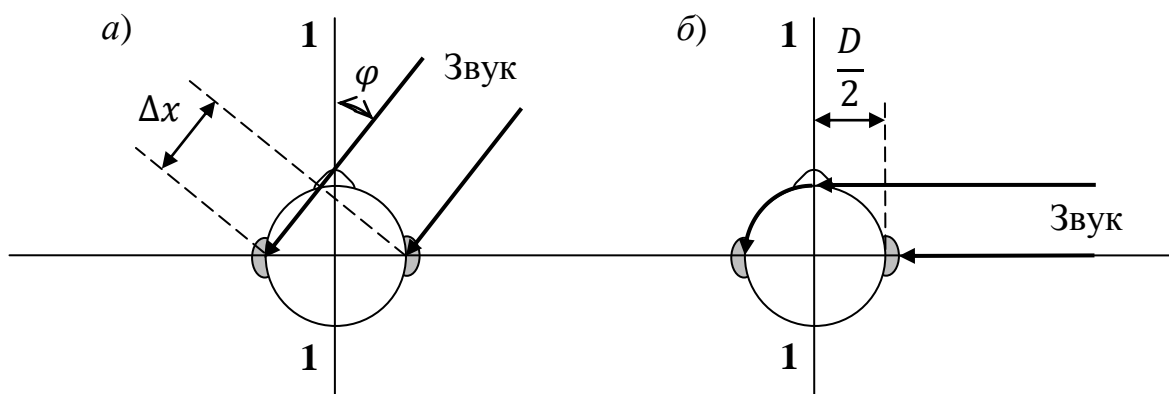


Рис. 12. Різниця у надходженні звукової хвилі до правого та лівого вуха: а) звукова хвиля йде під кутом φ до площини 1-1; б) звукова хвиля йде під кутом 90° до площини 1-1

Різниця часу Δt надходження однакових фаз звуку до вух визначається рівністю:

$$\Delta t = \frac{\Delta x}{v_0},$$

де Δx – різниця ходу звукової хвилі до лівого та правого вуха, v_0 – швидкість розповсюдження звуку у повітрі (340 м/с). У свою чергу Δx залежить від

розмірів голови людини, точніше від відстані між вухами D , яка в середньому дорівнює 18 см. Максимальну різницю ходу, яка виникає при розташуванні джерела під кутом 90° (збоку), можна розрахувати за формулою (рис. 12, б):

$$\Delta x = \frac{D}{2} + \frac{D \cdot \pi}{4}.$$

Виходячи з цієї формули максимальна різниця ходу досягає 23 см. В даному випадку вона більше за середню відстань між вухами ($D = 18$ см) і є меншою за найкоротшу відстань між вухами навколо голови, що дорівнює 28 см.

При такій різниці ходу максимальний час затримки $\Delta t = 0,63$ мс. Проте, це є справедливим тільки для синусоїдальних звуків з частотою не вище 800 Гц. На більших частотах час запізнення стає більшим за половину періоду коливання, тому поняття фази втрачає сенс.

Звуки високої частоти мають довжину хвилі, що є меншою за діаметр голови слухача, тому вони не проходять до вуха, яке знаходиться з протилежної сторони голови по відношенню до джерела звуку. При цьому, внаслідок явища дифракції, виникає, так звана, акустична тінь, яка зменшує інтенсивність звуку, що потрапляє у вуха (рис. 13, а). Низькі частоти мають довжину хвилі більшу, ніж діаметр голови слухача, тому вони огинають голову і не дають акустичної тіні (рис. 13, б).

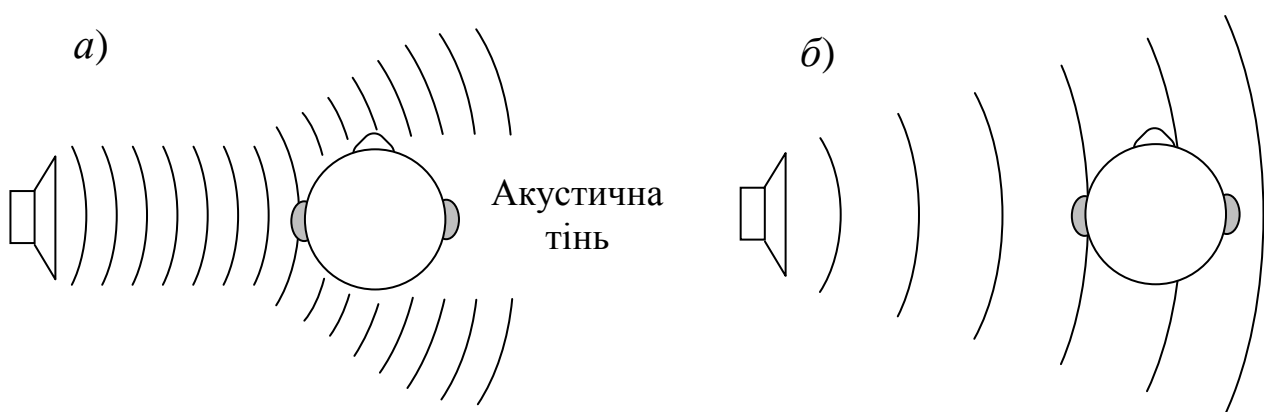


Рис. 13. Утворення акустичної тіні: а) на високих частотах; б) на низьких частотах

З досвіду випливає, що чисті тони високих частот (більш ніж 8 кГц) майже не піддаються локалізації. Також є слабо вираженою здатність до локалізації на частотах нижче за 300 Гц. На частотах, що є меншими за 150 Гц, локалізація взагалі відсутня. З цього випливає, що не важливо де розміщувати низькочастотні акустичні системи при стерео відтворенні звуку.

Найкраща точність локалізації досягається при сприйнятті складних звуків та звукових імпульсів. При цьому важливим фактором є спектральний склад звуків. Так, якщо звук містить низькочастотні та високочастотні складові та діє під кутом 90° на праве вухо, то у лівому вусі високочастотних складових не буде через акустичну тінь голови.

Найменший кут сприйняття відхилення джерела звукових сигналів, що відчувається, дорівнює 3° . Цю величину варто вважати бінауральною роздільною здатністю слуху для фронтального напрямку. Точність локалізації джерел звуку, розташованих ліворуч та праворуч, є значно меншою та складає біля 12° . Для тилового напрямку ця величина дорівнює біля 6° .

Додатковим механізмом при оцінці азимуту джерела звуку у просторі є, так званий механізм "фронт-тил", який дозволяє визначити чи знаходиться джерело звуку попереду, чи воно знаходиться позаду слухача. Ключовим елементом цього механізму є вушні раковини, які виконують роль просторового фільтру. При цьому потрібно відмітити, що цей механізм взагалі не працює, коли кут φ дорівнює 90 або 270 градусів.

2. Локалізація джерел звуку у вертикальній площині.

Здатність визначати напрямок надходження звуку у вертикальній площині у людини розвинута дуже слабо і роздільна здатність складає всього $10 \dots 15$ градусів. Ця здатність пов'язана з орієнтацією та формою вух.

3. Локалізація джерел звуку по глибині.

Поряд з визначенням напрямку на джерело звуку слухач впевнено оцінює відстань до джерела звуку. При цьому для різних відстаней до джерела така оцінка має певні особливості:

- При середніх відстанях до джерела звуку ($3 \dots 15$ м) наближення та віддалення джерела звуку супроводжується зміною його інтенсивності. Так, на низьких частотах збільшення відстані удвічі супроводжується зменшенням рівня звукового тиску на 6 дБ. На слух відстань завжди

визначається меншою, ніж вона є насправді. Помилка збільшується з ростом відстані.

- На близьких відстанях (менш ніж 3 м) на глибинну локалізацію здійснює вплив дифракція на вушній раковині та голові, тобто дається в знаки різниця рівнів інтенсивності та часової затримки. Точність глибинної локалізації при зміщенні широкосмугового джерела від 50 до 150 см не перевищує 15 – 30%.
- При великих відстанях (більш ніж 15 м) починає даватися в знаки згасання, що залежить від відстані, яку проходить звукова хвиля. При цьому високочастотні складові згасають швидше, змінюється тембр. Також на розповсюдження звуку впливає вологість повітря та напрямок вітру. Збільшення амплітуди низькочастотних складових спектру звуку викликає відчуття наближення джерела звуку. Штучне зменшення низькочастотних складових спектру сприймається як віддалення джерела звуку.

На відміну від азимутальної локалізації глибинна локалізація можлива при моноуральному прослуховуванні, проте бінауральне сприйняття істотно підвищує точність оцінки відстані. На відстанях більших за 10 м наявність бінаурального сприйняття вже не надає таких переваг.

Бінауральне маскування та демаскування

Коли звук, що маскує, та звук, що маскується, поступають в різні вуха, то слід говорити про бінауральне маскування. Величина зсуву слухового порогу при бінауральному маскуванні завжди значно менше, ніж при моноуральному маскуванні і проявляється вона більшою мірою на високих частотах.

Бінауральне маскування виникає лише за умови, якщо час дії звуку, що маскує, є не меншим 250 мс. Найбільш вираженим є маскування, коли звук, що маскується, та звук, що маскує, є близькими по частоті. При цьому на відміну від моноурального маскування відсутні биття і маскування найбільш проявляється в дуже вузькій смузі частот, яка збігається з шириною критичних смуг слуху. Міра бінаурального маскування наростає, коли обидва звуки мають пульсуючий характер, що характерний для мови та музики.

Бінауральним демаскуванням називають зниження порогу маскування при виділенні окремих звуків з їх сукупності, що одночасно діє на слухача. Це досягається шляхом додаткової обробки вхідної інформації як в периферійному, так і центральному відділах слухової системи та мозку.

Бінауральне демаскування передбачає використання слуховою системою часових та інтенсивністних відмінностей пар бінауральних звуків, що відповідають окремим джерелам, для їх розділення один від одного. Деякі дослідники вважають, що в основі цього явища лежить чутливість до зсуву фаз між звуками на частотах нижче 1500 Гц.

Якщо в стереотелефони подати одночасно один і той самий шум, та один і той самий звук, що маскується, то окрім шуму нічого не почуємо. Якщо ж звук, що маскується, подати лише в одне вухо, то ми його чутиемо, його рівень суб'єктивно підвищується. Цей ефект і називається бінауральним демаскуванням. При цьому шум і звук локалізуються в різних місцях голови. Шум в середині голови, а звук ближче до вуха, на яке він подається. Таким чином, демаскування здійснюється через різну суб'єктивну локалізацію шуму та звуку.

Контрольні запитання

1. Що вивчає наука "психоакустика"?
2. Яке визначення поняття "висота звуку"?
3. Для яких сигналів слухова система людини здатна визначити висоту звуку?
4. Як слухова система людини визначає висоту звуку складного звукового сигналу?
5. В яких одиницях вимірюється висота звуку?
6. Як висота тону залежить від частоти сигналу?
7. Від яких ще параметрів залежить висота тону звукового сигналу?
8. Що називають критичними смугами слуху?
9. Як ширина критичних смуг слуху залежить від частоти?
10. Яким чином можна охарактеризувати тембр звуку?
11. Який рівень звукового тиску відповідає больовому порогу для людини?
12. Що називають порогом чутності?
13. Як називаються частоти нижчі за 16 Гц та вищі за 20 кГц?
14. Як формулюється психофізичний закон Вебера?
15. За якою формулою визначається рівень інтенсивності звуку згідно закону Вебера-Фехнера?
16. Що називають маскуванням звукових сигналів?

17. Яким чином визначається маскувальна дія різних звуків?
18. Які є особливості маскування чистим тоном?
19. В яких одиницях вимірюється рівень гучності?
20. Що прийнято за одиницю рівня гучності?
21. Як визначається рівень гучності тонів на певних частотах?
22. Як називаються графіки однакового рівня гучності?
23. Що називається гучністю?
24. В яких одиницях вимірюється гучність?
25. Який характер має залежність гучності від рівня звукового тиску?
26. Як визначається гучність складного сигналу?
27. Що таке стала часу слуху?
28. Що розуміють під адаптацією слуху?
29. Що називають часовим маскуванням?
30. Що називається бінауральним слухом?
31. В чому полягає ефект передування, який відноситься до основних властивостей бінаурального слуху?
32. Як визначається загальна гучність звуку, який відчувається правим та лівим вухами людини?
33. Які фактори забезпечують можливості бінаурального сприйняття звуків?
34. Для яких частот звукових хвиль з діапазону, що може чути людина, можливе виникнення явища акустичної тіні?
35. Що називають бінауральним маскуванням?

Глава 3. Обробка звуку. Звукові сигнали

3.1. Первинні та вторинні звукові сигнали

Розрізняють первинні та вторинні звукові сигнали. До первинних відносять сигнали, що створюються музичними інструментами, співом, мовленням, або шумові сигнали, що створюються для супроводження інших сигналів (шум поїзда, рокіт моря, свист вітру і т.п.).

До вторинних звукових сигналів відносять сигнали, що відтворюються електроакустичними пристроями, тобто первинні сигнали, що пройшли

електроакустичними трактами зв'язку або каналами звукопередачі та відповідно видозмінені за своїми параметрами.

Сигнали, сформовані в каналах звукопередачі, служать, в більшості своїй, для поширення певних повідомлень в людському суспільстві. Будь-які перетворення звуків необхідні для спілкування (у широкому сенсі) людей. Тому обґрунтовано можна говорити про те, що звуки, відтворні на виході акустичного тракту (тобто вторинні звукові сигнали), первинно містять деяку комунікативну інформацію. А це означає, що при аналізі образу сприйняття таких звуків необхідно враховувати вплив цієї комунікативної складової на характеристики слухового образу.

Виходячи з таких міркувань, мова йде про тракти перетворення первинних звукових сигналів у вторинні як про канали комунікації. При цьому самі первинні сигнали можуть і не містити явної комунікативної складової. Проте, така складова незмінно з'являється в процесі формування вторинного сигналу. Дійсно, сам факт перетворення деякого первинного звуку у вторинний вже говорить про необхідність передачі (кимось, наприклад звукорежисером) інформації про конкретну акустичну ситуацію слухачеві (можливо, умовному). Звукорежисер, будучи посередником при передачі звуку, своєю роботою вносить комунікативну складову до вторинного сигналу, що їм продукується.

Останнє можна пояснити на наступному прикладі. Так шум листя в лісі дає нам певну акустичну інформацію про стан довкілля. Якщо ми самі знаходимося в лісі, то цей шум не містить для нас жодної комунікативної інформації. Він направлений в основному на забезпечення когнітивної функції слухового сприйняття. Звуки природи самі по собі не можуть бути комунікативними сигналами. Інша справа, якщо той самий шум ми прослухуватимемо у себе вдома з магнітофонного запису. У записаному звуці вже буде закладена інформація, призначена для повідомлення про те, що хтось зробив даний запис з цілком певною метою: для прослухування в інший час і у іншому місці. Звукозапис фактично є "матеріалізацію" звукової інформації про акустичну ситуацію, яка спостерігалася в лісі. Відтворення звуку у вторинному сигналі означає передачу цієї матеріалізованої інформації слухачеві. А це вже є етапом комунікації.

В ідеальному випадку вторинний сигнал повинен точно відповідати первинному. Але це не завжди потрібно, оскільки слух людини може і не помітити, що вони не збігаються. До того ж на практиці точної відповідності

первинного та вторинного сигналів буває дуже важко досягти, а іноді і просто неможливо. При цьому намагаються зробити так, щоб слухове відчуття, яке виникає у слухача, було як найближче до того відчуття, яке він отримував би, якщо б знаходився у місці виникнення первинного сигналу. Найголовніше ж полягає в тому, що у вторинному сигналі відбивається, через його параметри, той образ, яким керувалася людина, що здійснювала запис. Іншими словами, запис характеризує в певному значенні, як ця людина уявляла собі подальше звучання вторинного сигналу (у простому випадку параметри запису визначатимуться уявленнями розробника апаратури про його необхідну якість).

Те ж саме можна сказати і про первинні звукові сигнали, що призначені безпосередньо для спілкування, наприклад про звуки мови. Мовні звуки, у своїй основі будучи комунікативними сигналами, при перетворенні в акустичному каналі отримують деяку додаткову комунікативну якість, що визначається фактом передачі.

В рамках таких уявлень необхідно говорити про трансформацію звуків при формуванні вторинних акустичних сигналів. При цьому для забезпечення заданих характеристик сприйняття передбачається спеціальне спотворення фізичної структури вторинного сигналу по відношенню до первинного. Таке спотворення компенсує відмінності в акустичних умовах звучання первинного і вторинного сигналів.

Загалом усі невідповідності первинного та вторинного сигналів є спотвореннями у широкому сенсі цього поняття. Але зазвичай під цим терміном розуміють вузький тип спотворень. До них відносяться лінійні, нелінійні, параметричні та перехідні (часові) спотворення. Формуючи у слухача за допомогою вторинного сигналу слуховий образ, що є адекватним образу, який виникає при прослуховуванні первинного сигналу, необхідно передбачати не лише спотворення акустичного сигналу, але і оцінювати баланс цих спотворень із спотворенням комунікативної складової.

Характер спотворень акустичного сигналу, внесених каналом звукопередачі, а також допустимі межі і форма цих спотворень залежать в першу чергу від призначення самого каналу, тобто від того, з якою метою створюється вторинний сигнал. У цьому сенсі спотворення, що вносяться каналом, матимуть різну суб'єктивну значущість залежно від призначення звукового каналу. Тут йдеться про значущість спотворень з точки зору зовнішнього спостерігача, а не слухача. Таким зовнішнім спостерігачем є в

першу чергу розробник апаратного забезпечення каналу передачі звуку. Якщо звуковий канал призначений лише для передачі повідомлення, тобто лише для комунікації (у вузькому сенсі), то жоден розробник не прагнучиме забезпечувати такі вимоги до параметрів тракту, які дозволили б створити у вторинному сигналі слуховий образ, адекватний сприйняттю звучання в первинному сигналі. Його будуть цікавити перш за все можливі спотворення, які будуть впливати на змістовну частину звукового сигналу, тобто на якість комунікативної інформації, що передається. Сформований за таким принципом вторинний сигнал буде цілком адекватним первинному сигналу з точки зору сприйняття змісту повідомлення. При цьому значна частина когнітивної складової первинного сигналу може бути втрачена без суттєвої шкоди для слухача.

При оцінці тракту передачі вважають, що кожний акустичний сигнал майже завжди є випадковим у ймовірнісному розумінні і несе у собі певну інформацію, що відповідає його об'єму. Строго кажучи, звукові сигнали художніх програм, не можуть розглядатися як випадкові, оскільки вони можуть бути добре відомими слухачам і тому не несуть у собі інформацію, тобто їх інформаційний об'єм дорівнює нулеві. Для слухача ці сигнали є засобом задоволення власних естетичних потреб, а не отримання інформації. Проте, при розрахунках каналів звукопередачі ці сигнали вважають такими, що несуть інформацію, яка відповідає певному об'єму. Цей об'єм визначається виходячи з параметрів звукових сигналів, які характеризують основні властивості сигналів з точки зору їх передачі. Відповідні параметри визначаються за рівнем, за частотним діапазоном та у часі.

3.2. Динамічний діапазон

У процесі будь-якої передачі рівень акустичного сигналу постійно та неперервно змінюється. При цьому діапазон зміни сигналу може мати доволі широкі межі. Залежність рівня сигналу від часу називається рівнеграмою. Приклад такої залежності наведений на рисунку 14. Зазвичай рівнеграму наводять для рівня, визначеного при постійній часу вимірювача, що дорівнює або 150 – 200 мс (суб'єктивна рівнеграма), або 20 – 30 мс (об'єктивна рівнеграма).

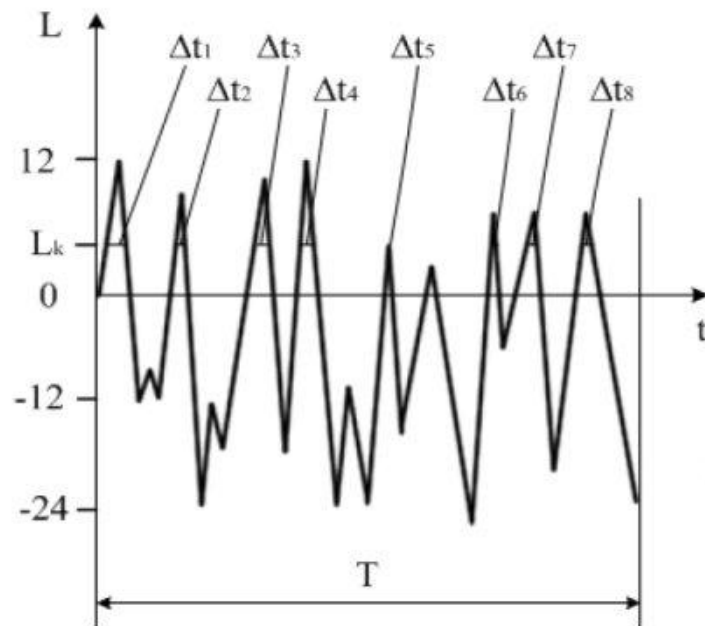


Рис. 14. Рівнеграма акустичного сигналу

Як показано на рисунку 14, максимальному рівню сигналу L_k , що визначається з рівнеграми, відповідатиме рівень, для якого виконується наступна умова:

$$\sum_i \frac{\Delta t_i}{T} \leq 0,01.$$

Рівнеграма подібна до осцилограми. Зазвичай осцилограми застосовуються для вивчення короточасних процесів (соті долі секунд). Проте, осцилограми дуже детально описують звукові коливання і тому вони незручні для вивчення порівняно великих (за часом) фрагментів. На рисунку 15 зображено фрагмент звучання тривалістю всього в одну соту секунди. Як можна побачити ця осцилограма відображає коливання звукового тиску у часі, що йдуть одне за одним, та мають схожість форми та величини. При цьому амплітуда таких коливань може змінюватись у надзвичайно широких межах – у сотні та тисячі разів. В той же час, враховуючі інерційність людського слуху, який реагує не на миттєві зміни звукового тиску, а оцінює звукові коливання у сумі (з постійною часу біля $1/5$ секунди), детально зображати коливання звуку є недоцільним. Достатньо враховувати лише зміну амплітуди цих коливань, тобто обвідну, яка змінюється набагато повільніше ніж миттєві значення (рис. 16).

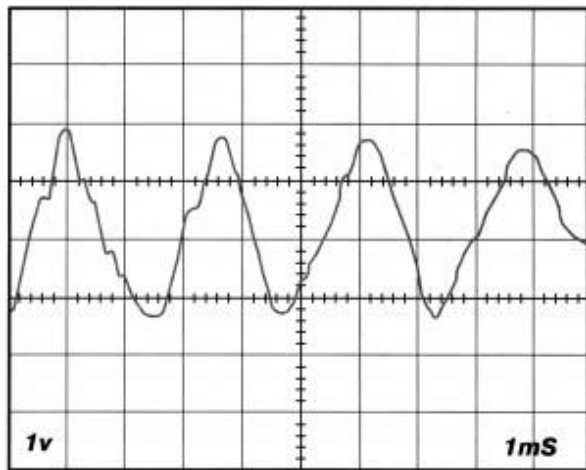


Рис. 15. Осцилограма звукового сигналу (за час 10 мс)

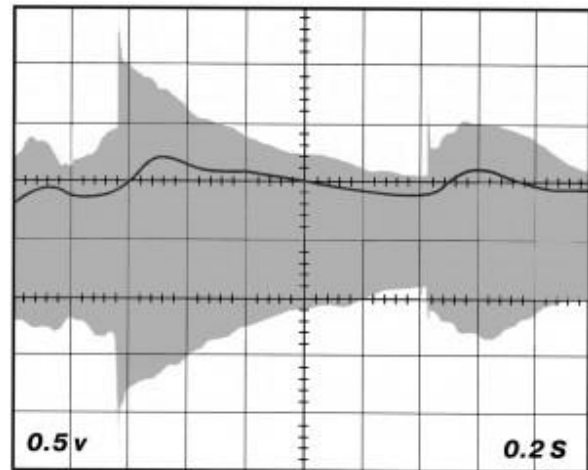


Рис. 16. Осцилограма звукового сигналу (за час 2 с) та його усереднена обвідна (чорна лінія)

Оскільки музичні та мовні сигнали є сигналами квазівипадковими, тобто передбачити їх значення можна лише з певною вірогідністю, то для аналізу їх акустичних характеристик застосовуються методи статистичного, спектрального та кореляційного аналізу. Статистичний аналіз рівнеграм звукових сигналів має дуже важливе значення при монтажі та редагуванні звукових програм, узгодженні їх з параметрами апаратури та звукових трактів, при виборі різних способів обробки сигналів, при оптимальній організації систем звукового мовлення, звукозапису і т. д.

Одним з найбільш важливих параметрів, що отримуються із статистичного аналізу рівнеграм, є динамічний діапазон сигналу. Динамічний діапазон сигналу завжди потрібно порівнювати з динамічним діапазоном каналу передачі:

$$D_k = 20 \lg \frac{U_{\text{ном}}}{U_{\text{ш}}} - (\Delta N_1 + \Delta N_2), \quad (13)$$

де $U_{\text{ш}}$ – рівень шуму в каналі, $U_{\text{ном}}$ – номінальна напруга, ΔN_1 – рівень перекриття завад та шумів γ (не менш ніж 10 дБ), ΔN_2 – допуск на перевантаження (3 – 6 дБ).

Динамічний діапазон для деяких видів звукових сигналів наведений у таблиці 2.

Таблиця 2. Динамічний діапазон для деяких видів звукових сигналів.

Вид сигналу	Динамічний діапазон, дБ
Мова диктора	25 ÷ 35
Художнє читання	35 ÷ 45
Телефонні розмови	35 ÷ 45
Симфонічний оркестр	65 ÷ 75
Рок музика	до 118

Як видно з таблиці 2, для передачі натурального динамічного діапазону потрібна високоякісна апаратура (з великим запасом лінійної частини амплітудної характеристики). В більшості випадків динамічний діапазон первинних звукових сигналів перевищує можливості аналогових засобів зв'язку та мовлення. Тому при їх використанні доводиться заздалегідь стискати динамічний діапазон або ж миритися з появою значних спотворень у тракці передачі. Для стискання динамічного діапазону використовується попередня динамічна обробка сигналу.

Проте, необхідність стискання динамічного діапазону може бути викликана не тільки технічними можливостями апаратури. Розглянемо конкретний приклад. Нехай у нас є деякий звуковий сигнал, який має діапазон зміни рівня біля 70 дБ. На рисунку 17 представлений фрагмент рівнеграми такого сигналу, де одночасно відображені як піковий, так і усереднений за деякий час рівень сигналу. На даній рівнеграмі можна помітити, що на графіку пікових рівнів є багато дуже короткочасних піків, які майже не впливають на усереднений рівень та гучність.

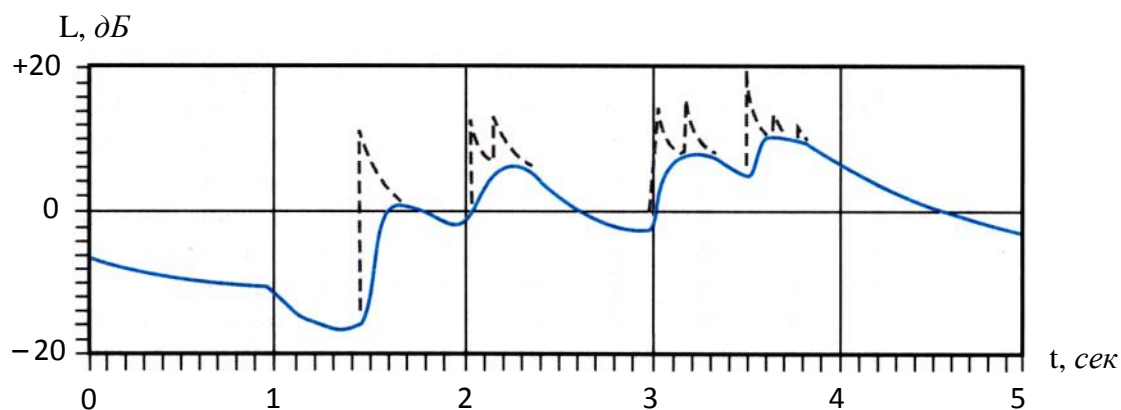


Рис. 17. Фрагмент рівнеграми звукового сигналу

Таким чином, якщо ми хочемо без спотворень передати такий динамічний діапазон, то, щоб найслабші звуки не "потонули" в шумах, а піки сигналу не викликали спотворень, ми повинні мати відношення максимального звукового тиску, що відтворюється у слухача, до рівня шуму на місці прослуховування по меншій мірі на 8 – 12 дБ більше ніж різниця між мінімальним та піковим рівнем на рівнеграмі. З іншого боку, рівень шуму у міській квартирі зазвичай набагато вище, ніж у концертному залі, а тим більше у студії, складаючи у середньому 30 – 35 дБ. Якщо при цьому не піднімати звукового тиску до рівня, коли звук вже буде дратувати сусідів (це близько 80 дБ), то тихі місця у звуковому сигналі будуть повністю заглушатися шумом у кімнаті. При відсутності сусідів з'являється інше обмеження – якщо підняти гучність настільки, щоб навіть найтихіші місця у звуковому сигналі перекривали рівень кімнатного шуму, то гучні акорди та тони будуть ураганом вириватися з динаміків (при умові, звичайно, що вони здатні це витримати), і 120 дБ звукового тиску будуть "спалювати" наш слух.

Що ж робити, коли динамічний діапазон звукового сигналу значно перевищує максимально допустимі межі? Вихід з такого положення знову підказує рівнеграма. Якщо уважно подивитися на будь-яку рівнеграму "живого" звуку, то можна помітити, що дуже гучні та занадто тихі звуки складають лише малу частину від загального часу звучання. Отже, якщо зменшити рівень найгучніших сигналів і підняти рівень найтихіших, то перепад рівнів сигналу стане набагато меншим. Таке регулювання і зветься стисканням (компресією) динамічного діапазону і може бути зроблене як вручну, так і автоматично, за допомогою спеціального пристрою, який називається компресором. Якщо звук, що відтворюється, є знайомим звукорежисеру, то результат ручного стискання зазвичай звучить природніше, ніж при повністю автоматичній компресії, оскільки компресор реагує із запізненням, тоді як людина може завчасно вжити необхідних заходів. З іншого боку, ручне регулювання – це процедура, що дуже втомлює.

Є ще один резерв зменшення динамічного діапазону. Звернемося знову до рисунку 17. Як вже було зазначено раніше, короточасні піки рівня практично не впливають на загальну усереднену гучність, і якщо б їх можна було звести до середнього рівня, то необхідний динамічний діапазон звукового тиску у слухача можна було б скоротити на ті самі 8 – 12 дБ, на які пікові рівні відрізняються від середніх (а це порядку 10 разів по потужності). Вручну відслідковувати такі піки

неможливо. Тому їх "зрізання" повинно виконуватись лише автоматично. Найпростішим чином це можна зробити за допомогою безінерційного обмежувача – кліпера, який зрізує верхівки хвилі сигналу на певному рівні (рис. 18, верхня крива). Подібні пристрої сильно спотворюють звук, тому їх використовують тільки для передачі мови.

Для музикальних сигналів існує більш складний пристрій, який швидко змінює свій коефіцієнт підсилення при перевищенні сигналом деякого рівня (порогу), не спотворюючи при цьому форму хвилі. Такий пристрій називається лімітером (або жорстким компресором, або інерційним обмежувачем). Результат його роботи показаний на нижній кривій рисунку 18.

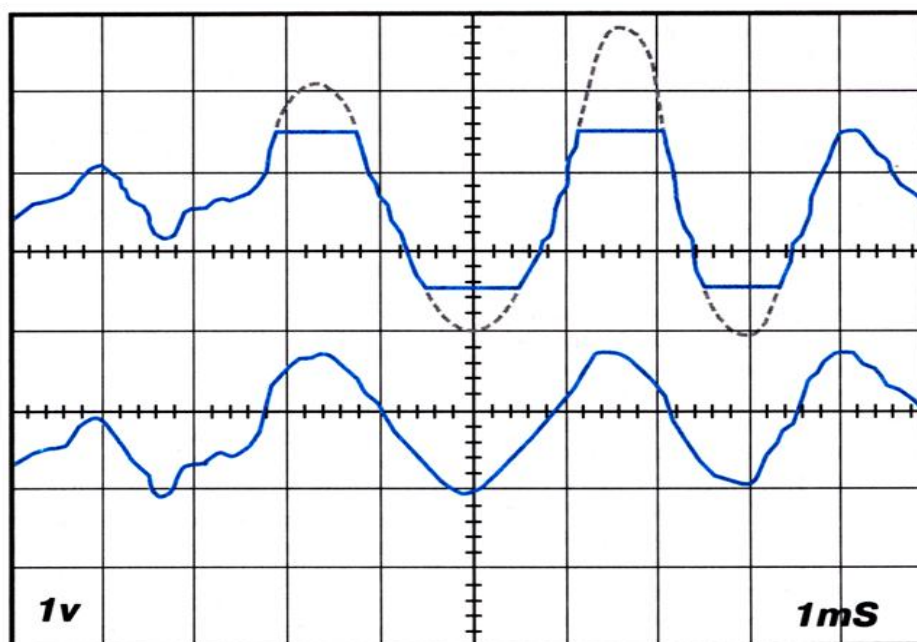


Рис. 18. Осцилограма, що ілюструє роботу кліпера (верхня крива) та лімітера (нижня крива)

Якщо лімітер працює тільки на піках рівня, то його робота практично не помітна на слух: гучність залишається попередньою, а пікові рівні помітно знижуються. При частішому спрацюванні лімітера звучання музики стає монотонним та "жорсткішим", оскільки усі гучні місця починають мати один і той самий рівень – спотворюється баланс рівнів, проте загальна гучність (при незмінному порозі лімітування) помітно зростає. Цим ефектом часто

користуються для підвищення суб'єктивної гучності передач у радіомовленні та при запису танцювальної поп-музики.

Стискання динамічного діапазону – це історично перше і по цей час найпоширеніше застосування приладів динамічної обробки. Проте, на компресії світ клином не зійшовся, і обробка динаміки сигналу не зводиться лише до його стискання.

За видом обробки, що виконується, динамічні процесори можна розділити на 6 основних класів:

- 1) лімітери (інерційні обмежувачі);
- 2) компресори (стискателі);
- 3) гейти (порогові шумопригнічувачі – відсікають слабкі сигнали);
- 4) експандери (розширювачі – використовуються для відновлення динамічного діапазону після його стискання компресором);
- 5) динамічні еквайзери та мікшери;
- 6) психоакустичні процесори.

Останні два класи процесорів є доволі складними системами и опис принципів їх роботи потребує докладнішого розгляду. Почнемо з динамічних еквайзерів.

Загалом еквайзери бувають графічними та параметричними. Графічний еквайзер – це набір паралельно з'єднаних смугових фільтрів, кожен з яких відповідає за пропускання сигналів у своїй власній смузі частот. Рівень пропускання сигналу в кожній смузі задається регулятором-повзунком. Набір регуляторів-повзунків нагадує графічне завдання амплітудно-частотної характеристики (АЧХ) сигналу, що обробляється. Тому цей еквайзер і назвали графічним.

Параметричний еквайзер дозволяє не тільки задавати коефіцієнт передачі, але й частоту, що обробляється, та ширину смуги обробки. В деяких моделях існує також можливість задавати для кожної зі смуг тип фільтра. Таким чином, параметричні еквайзери мають по три органи керування для кожної смуги регулювання. Причому діапазони зміни параметрів можуть сягати вельми істотних величин.

Динамічний еквайзер є різновидом односмугового параметричного еквайзера з динамічним керуванням частотою настройки, підсиленням та смугою пропускання.

Психоакустичні процесори

Психоакустичні процесори – це особливі прилади, що поєднують у собі відразу декілька способів обробки звукового сигналу. Робота психоакустичного процесора базується на використанні дуже тонких, скритих або маловідомих особливостей слуху людини, тому вона сприймається різними слухачами по-різному. На практиці як психоакустичний процесор використовують декілька типів приладів, таких як енхансери, ексайтери, максимайзери, віталайзери та спектральні процесори.

Енхансер (enhancer) – це один з найперших психоакустичних процесорів. Цей процесор випускався (і випускається по цей час) багатьма фірмами. У нашій країні цей клас пристроїв найбільш відомий по апаратурі фірми Alesis. Цей процесор дозволяє у більшості випадків зробити звучання дещо чіткішим та дзвінким. Особливо доцільним є його використання для обробки окремих звуків, в яких переважають різкі атаки (ударні, "залізо" і т. і.).

Фактично енхансер – це експандер, але такий, що працює лише у високочастотній області спектру звукових сигналів. Узагальнена структурна схема представлена на рисунку 19. Вхідний сигнал потрапляє на фільтр (1), що виділяє з усього спектру тільки високочастотні (ВЧ) складові. На елементі VCA (2) здійснюється керування амплітудою відфільтрованого сигналу, після чого в суматорі (3) результуючий сигнал додається до початкового сигналу. Керуюча напруга для VCA виробляється блоком керування (4) на основі аналізу ВЧ складових вихідного сигналу.

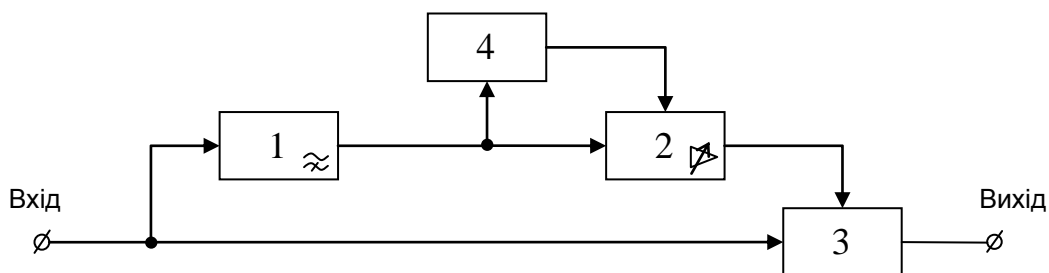


Рис. 19. Узагальнена структурна схема енхансера. На схемі: 1 – фільтр ВЧ, 2 – керуючий елемент (VCA – Voltage Control Amplifier), 3 – суматор, 4 – блок керування.

Різні моделі енхансерів відрізняються між собою, головним чином, характеристиками фільтрів ВЧ та алгоритмом роботи й керування. Проте, не дивлячись на відмінності, всі енхансери можуть тільки збільшувати долю ВЧ складових в сумарному вихідному сигналі. Це дозволяє зробити роботу енхансера менш помітною на слух та більш живою. При цьому загострюються, стають чіткішими тільки моменти атаки ударних інструментів, а на загальний сигнал робота енхансера практично не впливає. Завдяки цьому краще пророблюються дрібні деталі звукової картини, звучання стає акцентованішим.

Ексайтер з моменту своєї появи наприкінці 70-х років минулого століття є найпопулярнішим у світі психоакустичних процесорів. Можна вважати, що саме з нього почалася ера психоакустичних процесорів. Серед фірм, які займаються виробництвом психоакустичних процесорів немає жодної, яка б не випускала хоча б однієї моделі ексайтера. Родоначальником випуску цих процесорів була фірма APHEX.

Принцип роботи ексайтера полягає у тому, що до вихідного сигналу підмішуються додаткові гармоніки. В залежності від того, які гармоніки додаються – парні чи непарні – звучання робиться гострим або різким.

Всі ексайтери, що випускаються різними фірмами мають схожі структури. Типова схема ексайтера показана на рисунку 20. Сигнал, що потрапляє на вхід ексайтера, розгалужується на два: один з них потрапляє безпосередньо на вихідний суматор, а інший спрямовується в ланцюги обробки, після яких він додається до прямого необробленого сигналу.

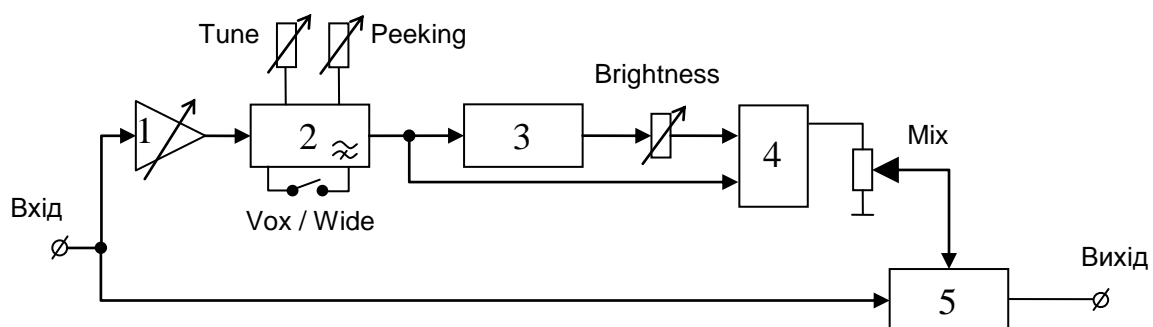


Рис. 20. Типова схема ексайтера. На схемі: 1 – вхідний підсилювач, що регулюється, 2 – спеціальний фільтр ВЧ, 3 – генератор гармонік, 4 – суматор сигналу ефекту, 5 – вихідний суматор.

В ланцюгу обробки ексайтера сигнал спочатку потрапляє на вхідний підсилювач (1), що регулюється та за допомогою якого можна підібрати необхідну величину завантаження (рівень збудження) генератора гармонік (3), що знаходиться після спеціального фільтру ВЧ (2). Цей фільтр має особливі АЧХ та ФЧХ, які дозволяють при подальшому сумуванні обробленого та прямого сигналів отримати "розширення" коротких імпульсів та, як наслідок, дещо збільшити їх суб'єктивну гучність, що сприймається. У спеціальному фільтрі (2) є регулятор настройки частоти (Tune), що дозволяє обрати для обробки бажану частину спектру вхідного звукового сигналу, а також регулятор добротності (Peeking), що дозволяє створити додатковий акцент у звучанні. Крім цього в фільтрі є перемикач (Vox/Wide), що кардинальним чином змінює характер роботи і, відповідно, звучання ексайтера, особливо в області середніх частот. Після фільтрації сигнал знову розгалужується на два. Один з них потрапляє прямо на суматор сигналу ефекту (4), а інший – подається на керування генератором гармонік. В генераторі (3) на основі інформації про вхідний сигнал здійснюється генерація вищих гармонік. При цьому синтезується, головним чином, друга гармоніка (як благозвучніша), а також ще декілька, але значно меншої амплітуди. Синтезовані гармоніки подаються на суматор сигналу ефекту (4) через регулятор (Brightness), який дозволяє встановити їх бажану величину у загальному сигналі. Потім сформований сигнал за допомогою регулятора (Mix) змішується з початковим сигналом у вихідному суматорі (5). Ручка (Mix) фактично регулює величину ефекту дії ексайтера.

По звуку ексайтер відноситься до тої групи пристроїв, робота яких непомітна доки їх не вимкнути. При застосуванні ексайтера із звуку зникає "ватність". Звук стає чітким та прозорим. Дія ексайтера основана на складному процесі, який враховує комплексний характер сприйняття звуків людським вухом. Ефект, який вносить ексайтер неможливо охарактеризувати за допомогою цифр. Цей ефект носить лише описовий характер. Можна сказати, що при дії ексайтера звук стає наче "живим" та починає "дихати".

Максимайзер є дещо схожим на енайсер. Але головна його відмінність полягає у тому, що максимайзер здатний працювати як "в плюс", так і "в мінус". Основу схеми максимайзера складають два регулятори тембру по низьким частотам (НЧ) та по ВЧ. При цьому регулятором НЧ користувач може керувати сам, а регулятором ВЧ керує сама схема. Приблизна спрощена схема

максимайзера показана на рисунку 21. Сигнал з входу максимайзера потрапляє на регулятори тембру (1) та (2), а також одночасно на два фільтри ВЧ (3) та смуговий (4). При цьому ФВЧ, відповідно до своєї назви, виділяє тільки високочастотні складові, а смуговий фільтр (СФ) – тільки середні частоти, що лежать нижче смуги пропускання ФВЧ. Сигнали цих двох смуг звукових частот потрапляють в блок керування (5), який порівнює їх величини і на підставі цього порівняння вирішує, що робити з високими частотами – піднімати чи ослаблювати. Відповідний сигнал подається на регулятор тембру ВЧ (2). Яким чином здійснюється регулювання вирішує максимайзер. Користувач може лише встановлювати межі глибини цього регулювання за допомогою спеціального регулятора (Definition) – чіткість.

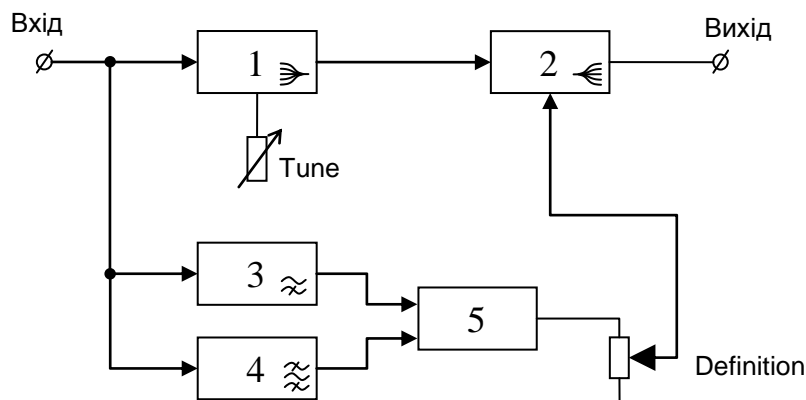


Рис. 21. Спрощена структурна схема максимайзера. На схемі:
1 – регулятор тембру НЧ, 2 – регулятор тембру ВЧ, 3 – фільтр ВЧ,
4 – смуговий фільтр, 5 – блок керування.

Вочевидь, найкраще застосування максимайзера – це обробка широкосмугових сигналів. Раніше ці пристрої були дуже популярними. Проте, сьогодні максимайзер вже майже не використовується.

Віталайзер є звичайним розширювачем стереобаз, проте з оригінальними еквалайзерами. Цей прилад є розробкою німецької фірми SPL. Віталайзер містить у собі доволі своєрідний "психоакустичний еквалайзер", а також, так званий, Surround-Processor. Останній є тривіальним розширювачем стереобаз.

Розширювач стереобазис або стереокартини дозволяє нібито розсунути звук у боки, звук при цьому стає об'ємніше, а в деяких випадках і прозоріше. Досягається це завдяки збільшенню різниці між правим та лівим каналами. Тобто усі джерела звуку, які знаходилися не в центрі панорами, розсуваються в боки на відстань, що прямо пропорційна значенню розширення. Детальніше з технологією розширення стерео можна ознайомитись в роботах [7, 8].

Що до еквалайзера, то він складається з двох частин, що діють практично незалежно одна від одної. Загальне у них тільки те, що, крім окремих регуляторів на різні смуги спектру, є й загальний регулятор, що встановлює глибину впливу відразу усіх темброутворюючих ланцюгів на сигнал, що обробляється. Це досягається завдяки тому, що у віталайзері застосовується так званий "паралельний" принцип побудови тракту обробки. При цьому різні частотні складові сигналу ефектів спочатку підсумовуються між собою, і тільки потім додаються до первинного сигналу.

Ще один прилад, який відноситься до психоакустичних процесорів – це спектральний процесор фірми Долбі (Dolby Spectral Processor). Цей процесор по суті є одним з різновидів багатосмугових компресорів. Проте, від компресорів його відрізняє те, що він обробляє тільки сигнали низького рівня і абсолютно не зачіпає сильних сигналів. Прилад містить два ідентичних канали з можливістю їх об'єднання у стереопару. Кожний канал містить трьохсмуговий кросовер з частотами розділу, що регулюються від 75 Гц до 1 кГц для розділу НЧ/СЧ та від 500 Гц до 8 кГц – для розділу СЧ/ВЧ. У кожній смузі увімкнений особливий компресор. Подробиці його конструкції та роботи, нажаль, фірмою не розкриваються, так само як і його характеристики. При роботі компресора сигнали з рівнями нижчими за поріг (поріг спрацювання виставляється користувачем від -60 дБ до -40 дБ) можуть підсилюватися на величину аж до 20 дБ, яка виставляється за допомогою регуляторів для кожної смуги окремо. При цьому усі сигналу, що мають рівень вищий за порогів, абсолютно не зачіпаються. Це дозволяє ефективно "витягати" навіть найдрібніші деталі звукової картини, які часто маскуються сильнішими звуками. Крім того, потрібно враховувати також той факт, що для слабких сигналів АЧХ слуху є суттєво нелінійною, тобто слабкі низько- та високочастотні складові звуку можна просто не почути. Приклад цього можна спостерігати на кінцівках пісень, що затухають. Спектральний процесор може зробити такі місця рівнішими,

такими, що не змінюють свого тембрального забарвлення навіть у найтихіших ділянках.

Після обробки цим процесором звук стає конкретнішим, таким, що розрізняється навіть у найдрібніших деталях. При цьому загальна динаміка сигналу практично не змінюється, так само як не змінюється і характер перехідних процесів на атаках музичних інструментів. Само звучання набуває щільності та "соковитості".

Загалом, характеризуючи психоакустичні процесори за їх впливом на звуковий сигнал, який обробляється, можна зазначити, що практично всі вони не мають якихось кількісних характеристик, які б мали відношення до параметрів, що описують звук. Здебільшого ефект від їх роботи можна оцінити лише якісно.

3.3. Частотний діапазон та спектри

Звуковий сигнал від кожного з первинних джерел звуку, як правило, має форму та склад спектру, що безперервно змінюються. Спектри можуть бути високо- та низькочастотними, дискретними та неперервними. У кожного джерела звуку, навіть однакового типу (наприклад, скрипка у оркестрі), спектри мають індивідуальні особливості, що надає звучанню характерне забарвлення. Це забарвлення називають тембром. Існує поняття тембру музичних інструментів, а також тембру голосу: дзвінкий, коли підкреслені високочастотні складові, або глухий, коли вони пригнічені. В першу чергу цікавлять: середній спектр для джерел звуку кожного типу, а для оцінки спотворень сигналу – спектр, усереднений за тривалий інтервал часу (15 сек. Для інформаційних сигналів та 1 хв. – для художніх). Усереднений спектр є, як правило, суцільним і достатньо згладженим за формою.

Суцільні спектри характеризуються залежністю спектральної густини від частоти [9]. Цю залежність називають енергетичним спектром. Спектральною густиною називають інтенсивність звуку у смузі частот з шириною, що дорівнює одиниці частоти.

$$S = \frac{I_{\Delta F}}{\Delta F}, \quad (14)$$

де $I_{\Delta F}$ – інтенсивність, що виміряна за допомогою вузько смугових фільтрів. Для акустики ширину смуги ΔF беруть рівною 1 Гц.

Для зручності оцінки вводиться логарифмічна міра густини спектру аналогічно рівню інтенсивності:

$$B = 10 \lg \frac{S}{I_0}, \quad (15)$$

де $I_0 = 10^{-12}$ Вт/м² – інтенсивність, що відповідає нульовому рівню. Цю міру називають рівнем спектральної густини або спектральним рівнем.

Дуже часто для характеристики спектру замість спектральної густини використовують інтенсивності та рівні інтенсивності, виміряні у октавній, напівоктавній або трьохоктавній смузі частот. Неважко встановити зв'язок між спектральним рівнем та рівнем в октавній смузі.

Спектральний рівень в октавній смузі визначається виразом:

$$B = 10 \lg \frac{I_{\Delta F_{\text{окт}}}}{\Delta F_{\text{окт}} \cdot I_0}, \quad (16)$$

де $\Delta F_{\text{окт}}$ – ширина відповідної октавної смуги.

Рівень в октавній смузі:

$$L_{\text{окт}} = 10 \lg \frac{I_{\Delta F_{\text{окт}}}}{I_0}. \quad (17)$$

Віднімаючи з виразу (17) вираз (16), отримаємо шуканий зв'язок:

$$L_{\text{окт}} - B = 10 \lg \Delta F_{\text{окт}}. \quad (18)$$

При відомому спектрі сигналу можна визначити його сумарну інтенсивність. Так, якщо спектр заданий в рівнях інтенсивності для трьохоктавних смуг, то достатньо перевести ці рівні (в кожній смузі) в інтенсивності:

$$I_{\text{окт}} = I_0 \cdot 10^{0,1 \cdot L_{\text{окт}}}, \quad (19)$$

а потім просумувати всі інтенсивності. Сума всіх $I_{\text{окт}}$ дасть сумарну інтенсивність $I_{\text{сум}}$ для всього спектру. Тоді сумарний рівень буде визначатися виразом:

$$L_{\text{сум}} = 10 \lg \frac{I_{\text{сум}}}{I_0}. \quad (20)$$

Якщо ж спектр заданий у спектральних рівнях, то для всього спектру точний сумарний рівень

$$L_{\text{сум}} = 10 \lg \int_{F_H}^{F_B} 10^{0,1 \cdot B} dF, \quad (21)$$

де F_H та F_B – нижня та верхня межі частотного діапазону. Якщо ж частотний діапазон розділити на n смужок шириною ΔF_k , в межах яких спектральний рівень B_k є приблизно постійним, то загальний сумарний рівень буде дорівнювати:

$$L_{\text{сум}} \approx 10 \lg \sum_{k=1}^n 10^{0,1 \cdot B_k} \Delta F_k. \quad (22)$$

Частотний діапазон акустичного сигналу визначають з частотної залежності спектральних рівнів. Це визначення можна зробити або по спаду спектральних рівнів, або наближено, на слух. Суб'єктивними границями вважають помітність обмеження діапазону для 75% слухачів. В таблиці 3 наведено приклади частотних діапазонів для ряду первинних джерел акустичного сигналу.

Таблиця 3. Частотні діапазони деяких джерел акустичного сигналу

Джерело	Частотний діапазон, Гц
Мова	70 – 7000
Скрипка	250 – 15000
Орган	20 – 15000
Симфонічний оркестр	30 – 15000

Якщо спектри мають плавний спад в той чи інший бік, то їх ще оцінюють тенденцією, тобто середнім нахилом спектральних рівнів у бік низьких або високих частот. Наприклад, мовний спектр має тенденцію спаду у бік високих частот, що дорівнює 6 дБ/окт .

До акустичних сигналів відносять у деяких випадках і шуми (акустичні). На рисунку 22 наведені спектри трьох типів акустичних шумів: білого, рожевого та мовного.

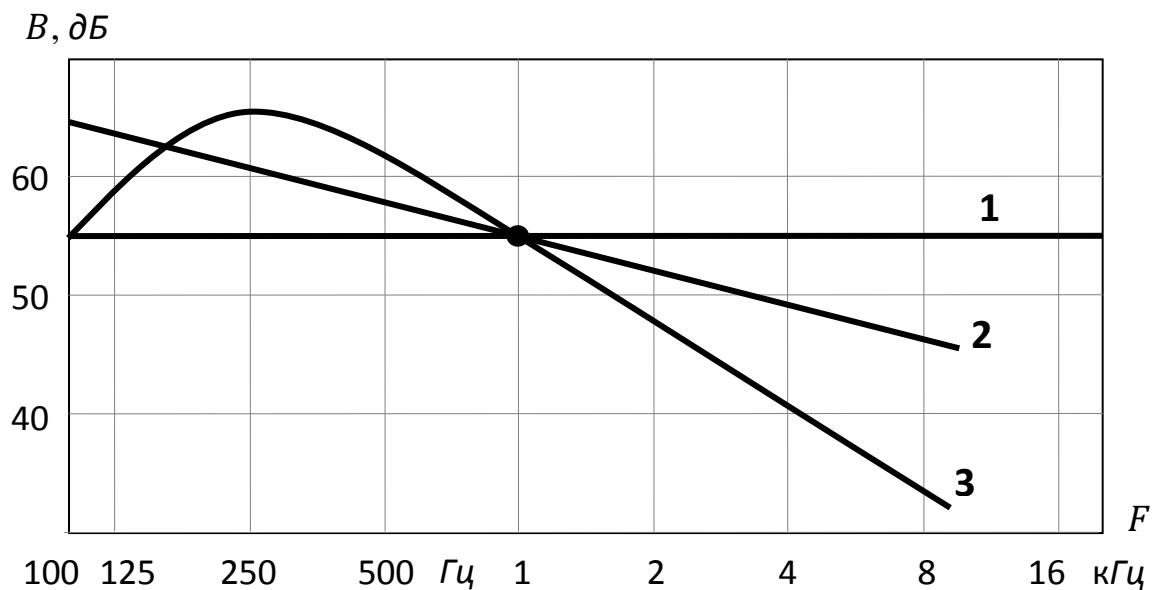


Рис. 22. Спектри трьох типів шумів: 1 – білий, 2 – рожевий, 3 – мовний

Термін "білий" відноситься до шумів, що мають однакову спектральну густину у всьому частотному діапазоні, "рожевий" – до шумів з тенденцією спаду густини на 3 дБ/окт у бік високих частот. Мовні шуми – це шуми, які створюються одночасно розмовою декількох людей.

3.4. Первинний мовний сигнал

Кожній людині властива власна манера вимовляти звуки (свого роду усний почерк). Вимова звуків залежить, наприклад, від наголосу, сусідніх звуків та інших факторів. Фізичні реалізації обмеженої кількості узагальнених звуків мови називають фонемами. Фонема – це те, що людина хоче вимовити, а звук мови –

це те, що людина фактично вимовляє. Фонема по відношенню до звуку мови грає ту ж саму роль, що і зразкова буква, що називається графемою (наприклад, друкована курсивом) по відношенню до її рукописної форми в конкретному написанні.

Імпульси потоку повітря, що створюються голосовими зв'язками з достатньою точністю, можуть вважатися періодичними. Відповідний період повторення імпульсів називають періодом основного тону голосу T_0 . Обернену величину $F_0 = 1/T_0$ називають частотою основного тону. Якщо зв'язки тонкі і сильно напружені, то період виходить коротким і частота основного тону є високою; для товстих, слабконапружених зв'язок частота основного тону низька. Ця частота для всіх голосів лежить в межах від 70 до 450 Гц. При виголошуванні промови вона безперервно змінюється відповідно до наголосу і підкреслення звуків і слів, а також при прояві емоцій (питання, вигук, здивування і т. і.). Зміну частоти основного тону називають інтонацією. У кожної людини свій діапазон зміни частоти основного тону (зазвичай він не перевищує октави) і своя інтонація. Остання має велике значення для впізнання того, хто говорить. Основний тон, інтонація, усний "почерк" і тембр (забарвлення) голосу можуть слугувати для впізнання людини. При цьому міра достовірності впізнання буде вищою, ніж за відбитками пальців. Цю властивість використовують в розробленій останнім часом апаратурі, що реагує лише на певні голоси.

Відомо, що для передачі одного і того ж повідомлення по телеграфу і по мовному тракту потрібна різна пропускна здатність: для телеграфного повідомлення не більше 100 *біт/с*, а для мовного – близько 100000 *біт/с* (смуга 7000 Гц та динамічний діапазон 42 дБ, тобто потрібний семизначний код, звідки маємо: $2 \cdot 7000 \cdot 7 = 98000$ *біт/с*), тобто в 1000 разів більше.

Може здатися, що мовний сигнал має величезну надлишковість. Це не зовсім так. Утворення звуків мови відбувається шляхом подачі команд до м'язів артикуляційних органів мови від мовного центру мозку. Цих сигналів не більше десяти, при цьому вони змінюються повільно (у темпі зміни звуків мови, тобто не більше десяти разів за секунду). Тому загальний потік повідомлень для них складає біля 100 *біт/с*.

Різке збільшення об'єму мовного сигналу пояснюється наступним. Мовний сигнал є спектрально-модульованою несучою $\xi = E(\omega) \cdot S(\omega)$, де $E(\omega)$ –

спектр імпульсів основного тону або шуму, $S(\omega)$ – модуляційна (фільтруюча) функція мовного тракту.

В результаті спектральної модуляції змінюється співвідношення між частотними складовими несучої, тобто змінюється форма обвідної її спектру. Майже вся інформація про звуки мови знаходиться у цій спектральній обвідній та її часовій зміні. Ці зміни відбуваються повільно (у темпі вимовляння звуків), тому передача даних про обвідну та її зміну не потребує пропускну здатності тракту більше 100 біт/с . Але для передачі широкосмисловою несучої з її широким динамічним діапазоном потрібна дуже велика пропускну здатність. До того ж, мовний сигнал при утворенні в мовному тракті набуває багато інформації, що не відноситься до змісту мови, що передається (наприклад, фазова інформація). Ця інформація називається супутньою. Для її передачі також витрачається пропускну здатність тракту. З цього випливає, що інформаційна ємність мовного сигналу лише трохи перевищує інформаційну ємність телеграфного сигналу з таким самим повідомленням (мовний сигнал відрізняється від телеграфного лише інформацією про емоції та особу того, хто говорить). Для передачі суті повідомлення досить передавати відомості лише про форму обвідної спектру мови, а також про зміну основного тону мови та переходів тон-шум.

Контрольні запитання

1. Що називається рівнеграмою?
2. Яке значення повинна мати постійна часу вимірювання для отримання суб'єктивної рівнеграми?
3. Яке значення повинна мати постійна часу вимірювання для отримання об'єктивної рівнеграми?
4. Що таке динамічний діапазон сигналу?
5. Для чого потрібно стискати динамічний діапазон сигналу?
6. Які пристрої використовують для стискання динамічного діапазону сигналів?
7. Що таке безінерційний обмежувач, як він працює?
8. Яку функцію виконує лімітер при обробці звукових сигналів?
9. Як називається пристрій, який використовується для відновлення динамічного діапазону після його стискання?

10. Який пристрій використовується для відсікання слабких сигналів в процесі стискання динамічного діапазону?
11. Які психоакустичні процесори Ви знаєте? Як вони працюють?
12. Що називають спектральною щільністю сигналу?
13. Що називають спектральним рівнем?
14. Як визначається сумарний рівень звукового сигналу?
15. Як визначають частотні діапазони акустичних сигналів?
16. Як називають зміну частоти основного тону мовного сигналу?

Глава 4. Обробка звуку.

Цифрове представлення звукових сигналів

4.1. Процес аналогово-цифрового та цифро-аналогового перетворення звукових сигналів

Цифрове представлення звуку хоч і є складнішим, ніж аналогове, проте, це дозволяє значно спростити роботу з ним. До того ж, параметри цифрового звуку не змінюються у часі та суттєво у меншій мірі піддаються спотворенням, що характерні для аналогових сигналів, перш за все за рахунок технічних характеристик носіїв інформації. Цифровий код у будь-якій ситуації завжди сприймається однаково.

Перетворення аналогового звукового сигналу в цифровій включає декілька етапів. Спочатку аналоговий звуковий сигнал подається на аналоговий фільтр, який обмежує смугу частот сигналу і усуває перешкоди і шуми. Потім з аналогового сигналу за допомогою схеми вибірки/зберігання виділяються відліки: з певною періодичністю здійснюється запам'ятовування миттєвого рівня аналогового сигналу. Далі відліки поступають в аналого-цифровий перетворювач (АЦП), який перетворює миттєві значення кожного відліку в цифровий код або числа. Отримана послідовність біт цифрового коду, власне, і є звуковим сигналом в цифровій формі. В результаті перетворення неперервний аналоговий звуковий сигнал перетворюється на цифровий – дискретний як за часом, так і за величиною.

Структурна схема каналу цифрового запису (передачі) та відтворення звуку показана на рисунку 23.

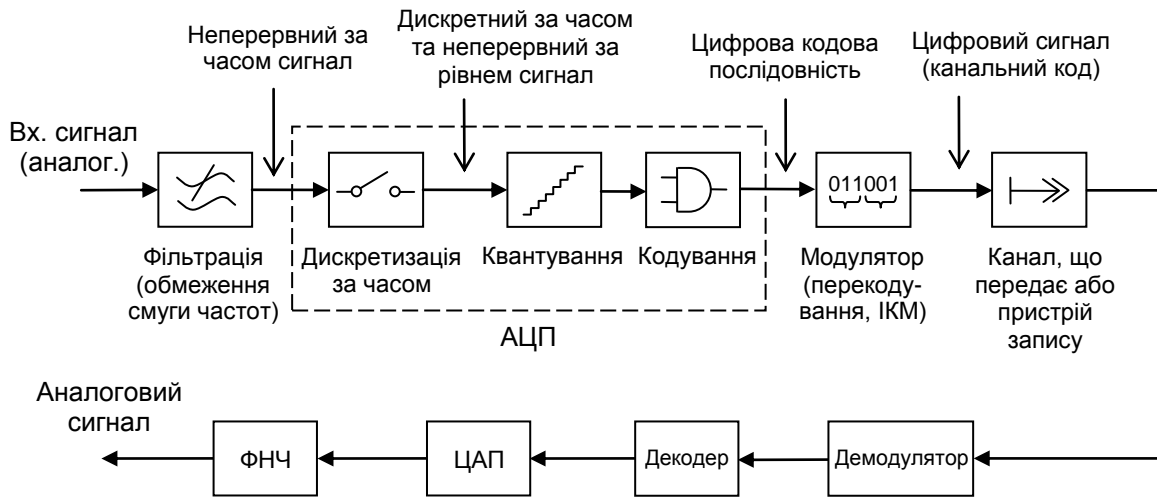


Рис. 23. Узагальнена структурна схема каналу цифрового запису (передачі) та відтворення звуку

Головну роль в процесі перетворення сигналу з аналогової форми в цифрову грає АЦП (Analog/Digital Converter – ADC), який поєднує процес дискретизації вхідного аналогового сигналу з його подальшим квантуванням та первинним кодуванням. Зворотний процес перетворення цифрового звукового сигналу в аналоговий реалізується за допомогою цифро-аналогового перетворювача – ЦАП (Digital/Analog Converter – DAC). Структурна схема ЦАП показана на рисунку 24.

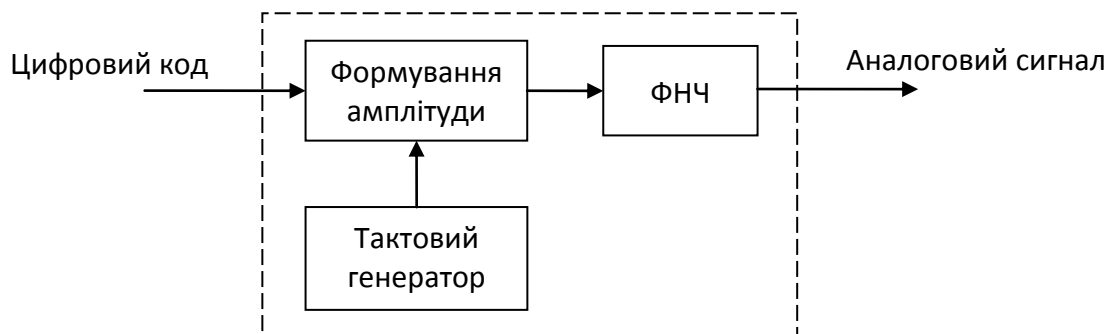


Рис. 24. Блок-схема цифро-аналогового перетворювача

Фактично робота ЦАП полягає у тому, щоб у певні моменти часу, що відповідають імпульсам тактового сигналу, перетворювати цифрові коди, які потрапляють на вхід ЦАП, у імпульси відповідної амплітуди. Сформовані таким чином імпульси амплітуди пропускаються через фільтр низької частоти (ФНЧ) для згладжування імпульсних сходинок та пригнічення періодичних складових спектру дискретного сигналу, що в результаті і призводить до появи вихідного аналогового сигналу. Таким чином відбувається відновлення початкового аналогового сигналу з його цифрового коду.

4.2. Дискретизація

Найважливішим етапом аналого-цифрового перетворення є дискретизація аналогового сигналу. Замість терміну "дискретизація" в технічній літературі інколи використовують термін "вибірка", а в літературі, присвяченій обробці звуку, використовується поняття – "семплінг" (від англійського слова "sampling").

За визначенням, дискретизація – це процес взяття відліків неперервного у часі сигналу у рівновіддалених одна від одної за часом точках (у даному випадку йдеться про рівномірну дискретизацію, оскільки ще буває й нерівномірна дискретизація, коли відліки беруться у точках, що неоднаково віддалені одна від одної у часі). Іншими словами, в процесі дискретизації вимірюється і запам'ятовується рівень аналогового сигналу. Через заданий інтервал часу, який називається інтервалом дискретизації, процедура повторюється.

Результатом процесу дискретизації є сигнал з амплітудно-імпульсною модуляцією РАМ (Pulse-Amplitude Modulation). Технічно ця процедура реалізується за допомогою пристрою вибірки/зберігання. Як елемент, що запам'ятовує, зазвичай використовують конденсатор, що заряджається до рівня напруги вхідного сигналу. При цьому потенціал заряду конденсатора відповідає миттєвому значенню напруги сигналу. Напруга на конденсаторі зберігається незмінною протягом деякого відрізка часу, що зветься часом зберігання. У ідеальному випадку взяття відліку повинне відбуватися миттєво, реально ж тривалість цього процесу складає приблизно 1 мкс.

Аналоговий сигнал можна відтворити з РАМ сигналу пропустив останній через ФНЧ. Важливо знати наскільки точно відфільтрований сигнал співпадатиме з початковим аналоговим сигналом. Відповідь на це питання дає теорема відліків Котельникова-Найквіста, яка формулюється наступним чином: *сигнал з обмеженою смугою, що не має спектральних компонентів з частотами, які перевищують частоту F_m , однозначно визначається значеннями, які обрані через рівні проміжки часу T_d , такими що:*

$$T_d \leq \frac{1}{2F_m}. \quad (23)$$

Це твердження також відоме як теорема про рівномірну дискретизацію. Частота $F_d = 1/T_d$ називається частотою дискретизації. Для заданої частоти дискретизації з умови $F_d = 2F_m$ визначається відповідна максимально можлива частота в спектрі сигналу – F_m , яка називається частотою Найквіста. Критерій Найквіста $F_d \geq 2F_m$ – це достатня умова, яка дає можливість повного відтворення аналогового сигналу з послідовності дискретних відліків.

Суть умови (23) полягає у наступному. В межах смуги частот вхідного сигналу спектр дискретного сигналу з точністю до множника повторює спектр аналогового сигналу. Крім того, цей спектр для дискретного сигналу періодично повторюється з частотою F_d (рис. 25). Тобто спектр сигналу, перетвореного за допомогою АЦП у цифрову форму, буде мати періодичний характер.

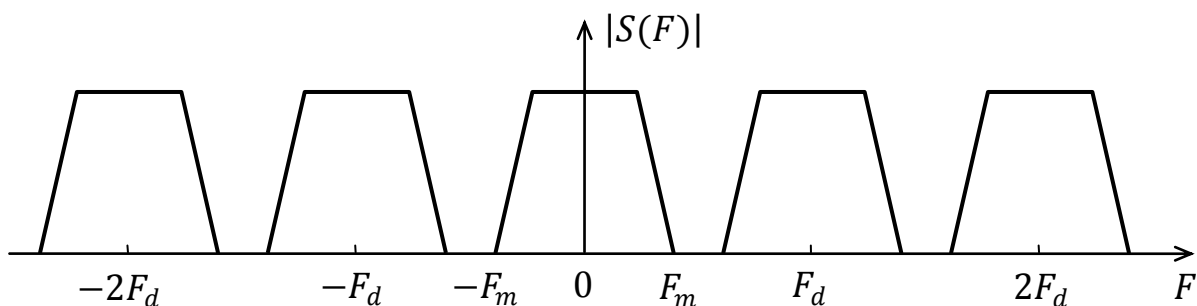


Рис. 25. Спектр дискретного сигналу

Якщо $F_d \gg 2F_m$, то копії спектру легко видалити за допомогою ФНЧ і таким чином легко відновити аналоговий сигнал. Проте, чим ближче F_d до $2F_m$, тим

важче виділяється аналоговий сигнал, оскільки це потребує збільшення крутизни спаду АЧХ фільтра, а це у свою чергу призводить до збільшення порядку фільтра і, як наслідок, до збільшення його складності.

Якщо $F_d < 2F_m$, то копії спектрів почнуть перекриватися (рис. 26). Явище, що є наслідком недостатньої дискретизації називається накладанням (aliasing). Накладення спектрів завжди приводить до появи нових спектральних складових в сигналі і, отже, до неможливості його правильного відновлення. Класичним прикладом накладення спектрів є випадок, коли при перегляді кінофільму здається, що колесо рухомої карети крутиться зі швидкістю, що не відповідає швидкості руху карети, або навіть у зворотний бік. Виникнення цього ефекту обумовлене тим, що швидкість зміни кадрів (частота дискретизації зображення) мала у порівнянні з кутовою швидкістю обертання колеса.

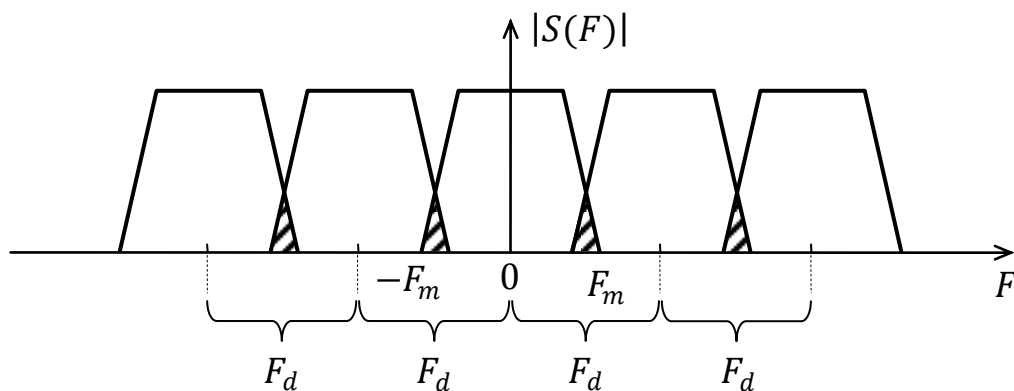


Рис. 26. Перекриття спектрів сигналу при недостатній частоті дискретизації

З практичної точки зору ні сигнали, ні спектри, ні фільтри не мають обмеженої смуги частот, тому певне накладання спектрів при дискретизації аналогових сигналів є завжди. Для того, щоб мінімізувати наслідки накладання застосовують попередню фільтрацію аналогового сигналу фільтром, який має граничну частоту меншу за $F_d/2$. Отже, після фільтру сигнал буде мати таку частоту F_m , яка точно буде задовольняти вимозі (23). У такому випадку, однак, це може призвести до втрати деякої частини інформації аналогового сигналу. Крім того, фільтри, що реалізуються на практиці, потребують певної ширини смуги, що є смугою переходу між областю пропускання та областю затухання. Ця смуга називається смугою переходу. Бажано, щоб смуга переходу була

якомога меншою, але при зменшенні цієї смуги різко підвищується складність фільтру та його вартість. В багатьох системах оптимальною шириною смуги переходу вважається 10 – 20% від ширини смуги сигналу. Отже, для випадку фільтрів, що фізично реалізуються критерій Найквіста буде мати наступний вигляд:

$$F_d \geq 2,2F_m . \quad (24)$$

4.3. Передискретизація (оверсемплінг)

З усього вищевикладеного випливає, що для здійснення аналого-цифрового перетворення з високою якістю необхідно виконати низку вимог.

По-перше, при дискретизації звукового сигналу слід використовувати частоту дискретизації, яка відповідає критерію Найквіста. При цьому, чим вищою буде частота дискретизації, тим більшу смугу частот зможе мати вхідний сигнал, при якій буде забезпечуватись висока якість його відтворення. На жаль, пропорційно збільшенню частоти дискретизації зростає потік цифрових даних в каналі звукозапису, а також об'єм пам'яті, необхідної для зберігання звукового сигналу в цифровій формі.

По-друге, перед дискретизацією необхідно обмежити спектр вхідного сигналу за допомогою фільтру низької частоти (ФНЧ). Він повинен видалити всі гармоніки з частотами, що є вищими за частоту дискретизації, і тим самим запобігти накладанню спектрів. При цьому смуга переходу такого фільтра повинна складати 20% від ширини смуги сигналу. Забезпечити останню умову можна, наприклад, за допомогою еліптичного фільтру високого порядку. Проте, такий фільтр може бути доволі складним у реалізації та мати високу вартість.

В сучасних АЦП проблема фільтрації з метою усунення високочастотних компонентів спектру вирішується за допомогою передискретизації – дискретизації на підвищеній частоті. Це найбільш економічний розв'язок задачі перетворення аналогового сигналу в цифровий та цифрового в аналоговий. Терміну "передискретизація" в англійській технічній літературі відповідає термін "оверсемплінг" (від англійського "oversampling").

При передискретизації діапазон частот вхідного аналогового звукового сигналу обмежується за допомогою дешевого ФНЧ низького порядку (зазвичай 3 – 5-го), що має лінійну фазову характеристику, яка практично не спотворює імпульсний сигнал, а також доволі широку смугу переходу. Відфільтрований сигнал, що має обмежений по частоті спектр, піддається дискретизації на підвищеній частоті, яка виключає накладення і спотворення спектру. Далі дискретні відліки сигналу перетворюються в послідовність чисел за допомогою АЦП, причому потік цифрових даних включає і небажані високочастотні компоненти спектру. Отримані цифрові дані піддаються цифровій фільтрації. Цифровий фільтр високого порядку з крутим зрізом частотної характеристики виготовити набагато простіше. Причому, завдяки лінійності фазової характеристики цифрового фільтру, фазові спотворення сигналу будуть відсутні. Більш того, якщо попередня фільтрація аналоговим фільтром вносить певні спотворення, то такі спотворення можуть бути враховані та компенсовані при проектуванні цифрового фільтру. Після цифрового фільтру сигнал матиме спектр, коректно обмежений по частоті і частота дискретизації сигналу знижується до подвоєного значення найвищої корисної частотної складової шляхом видалення "надлишкових" відліків.

Зазвичай починають з вибору частоти дискретизації, що в 4 рази перевищує потрібну. Після цього розробляють аналоговий фільтр, ширина смуги якого відповідає збільшеній частоті дискретизації.

Наприклад, для звукового сигналу з максимальною частотою у спектрі 20 кГц потрібна частота дискретизації, розрахована у відповідності з (24), що дорівнює 44,1 кГц. При цьому смуга переходу, що складає 4,1 кГц, може бути реалізована за допомогою найскладнішого еліптичного фільтру 10 – 20 порядку. Натомість, використовуючи оверсемплінг, ми збільшуємо частоту дискретизації до 176,4 кГц. В результаті смуга переходу збільшується до 136,4 кГц, що реалізується за допомогою простого еліптичного фільтру 4-го порядку. Після дискретизації із збільшеною частотою цифрові дані з виходу АЦП пропускаються через недорогий високопродуктивний цифровий фільтр, який реалізує смугу переходу у 4,1 кГц без спотворень. Після цього проводиться повторна вибірка і частота дискретизації знижується до 44,1 кГц.

Так само передискретизація використовується і в процесі перетворення цифрового сигналу в аналоговий.

4.4. Квантування

Після дискретизації відбувається другий етап аналого-цифрового перетворення – квантування відліків. В процесі квантування відбувається заміна неперервної функції сигналу її окремими значеннями, що віддалені одне від одного на кінцевий інтервал (рівень). Відповідно такий процес називається квантуванням за рівнем. При квантуванні значення функції сигналу в довільний момент часу замінюється найближчим значенням рівня, що називається рівнем квантування. Точність, з якою відбувається така заміна, визначається кількістю розрядів n , які використовуються для запису значення рівня. Кількість рівнів при цьому дорівнює 2^n . Інтервал між двома сусідніми дискретними значеннями рівнів називається кроком квантування. Якщо рівні квантування рівномірно розподілені по всьому діапазону зміни функції вхідного сигналу, то таке квантування буде називатися рівномірним, і крок квантування при цьому буде постійним.

Перетворення сигналу з аналогової форми в цифрову можливе лише з певною мірою точності. При цьому чим вища частота дискретизації та розрядність АЦП, тим точніше відбувається перетворення.

Спотворення сигналу, що виникає в процесі квантування, викликає втрати інформації, які при оберненому цифро-аналоговому перетворенні ліквідувати або зменшити практично неможливо. Наприклад, при запису звуку спектр помилки при квантуванні подібний до спектру білого шуму і займає частотний діапазон від нуля до частоти дискретизації. Тобто поява помилок квантування сигналу еквівалентна додаванню до відновлювального сигналу деякого шуму. Тому похибки квантування називають шумом квантування. Величина цього шуму обернено пропорційна кількості рівнів квантування. Рівень шуму квантування зазвичай вимірюється у присутності сигналу як рівень в дБ відносно максимального значення сигналу. Чим меншим є цей рівень, тим вища якість звуку. Детальніше про шум квантування та методи боротьби з ним можна прочитати в [10].

Шум квантування особливо помітний при малих рівнях сигналу. Отже, якщо у сигналі присутні ділянки, де рівень сигналу є малим, і є необхідність зменшити на цих ділянках рівень шуму квантування, то в цьому випадку може застосовуватись нерівномірне квантування за рівнем. В процесі нерівномірного квантування крок квантування є змінною величиною.

Окрім помилок, пов'язаних з шумом квантування, при перетворенні аналогового сигналу у цифровий можливі також і інші помилки, пов'язані з насиченням пристрою квантування. Кожний АЦП має певний робочий діапазон значень вхідного сигналу, в межах якого різниця між входом та виходом є незначною. Але, якщо вхідне значення не належить цьому діапазону, то значення на вході та виході будуть відрізнятися сильніше. В такому випадку кажуть, що пристрій працює у режимі насичення. Помилки насичення суттєвіші та менш бажані, ніж шум квантування. В загальному випадку насичення можна уникнути за допомогою автоматичного регулювання підсилення, яке ефективно розширює робочий діапазон АЦП. Детальніше про це можна прочитати в [11].

Знаючи розрядність АЦП можна отримати деякі цікаві цифри. Наприклад, у випадку двійкового кодування даних, якщо помножити кількість розрядів на частоту дискретизації сигналу, що виражена в *Гц*, то отримаємо швидкість передачі даних, яку повинен забезпечувати цифровий канал запису/відтворення звуку. Якщо отриману швидкість передачі даних помножити на загальну тривалість звукового сигналу в секундах, то отримаємо об'єм пам'яті, який потрібен для зберігання звукових даних. Якщо мова йде про стерео-сигнал, то швидкість передачі даних та загальний об'єм даних подвоюються.

4.5. Цифрове кодування

Кодування інформації, тобто перетворення значень вхідного сигналу у двійковий код, відбувається за допомогою кодерів, а декодування – за допомогою декодерів. Цифрова передача інформаційного сигналу стає можливою тільки після того, як дискретні та квантовані значення цього сигналу будуть закодовані у послідовність бітів, яка зветься потоком бітів або відеосигналом. Після цього формуються групи з k -біт, які зветься символами. При цьому кількість усіх символів скінчена ($M = 2^k$), а їх сукупність зветься алфавітом. Система, що використовує символний набір розміру M зветься M -арною. При $k = 1$ система зветься бінарною. Розмір набору символів дорівнює $M = 2$. При цьому модулятор використовує один з двох різних сигналів для представлення двійкового значення "1" та інший – для представлення двійкового значення "0". В цьому випадку символ і біт – це одне

і те саме. Якщо $k = 2$, то система є четвірковою або 4-рівневою. Модулятор при цьому використовується для представлення символу одним з 4 можливих сигналів.

Приклад 32-рівневої системи:

Числова послідовність:	10,	4,	36,	28,	52	
Знакове кодування (6-бітний код):	$\overbrace{001010}$	$\overbrace{000100}$	$\overbrace{100100}$	$\overbrace{011100}$	$\overbrace{110100}$	
32-рівневі символи (цифри):	5,	1,	4,	17,	25,	20
32-рівневі сигнали:	$u_5(t)$,	$u_1(t)$,	$u_5(t)$,	$u_{17}(t)$,	$u_{25}(t)$,	$u_{20}(t)$

У наведеному прикладі видно, що границі символів та знаків у загальному випадку можуть не співпадати.

Імпульсно-кодова модуляція (ІКМ)

Імпульсно-кодова модуляція (англ. PCM – pulse-code modulation) – це назва класу низькочастотних сигналів, отриманих з сигналів імпульсно-амплітудної модуляції (англ. PAM – pulse-amplitude modulation) шляхом кодування кожної квантової вибірки цифровим словом. Вихідна інформація дискретизується та квантується в один з M рівнів. Після цього кожна квантова вибірка проходить цифрове кодування для перетворення в k -бітне ($k = \log_2 M$) кодове слово. Далі біти кодового слова перетворюються в імпульсні сигнали. Відповідний процес бінарної імпульсно-кодової модуляції показаний на рисунку 27. На представленому рисунку амплітуди аналогового сигналу $A(t)$ обмежені діапазоном від 0 до 8 В. Крок між рівнями квантування складає 1 В. Відповідно, використовується 8 квантових рівнів, які мають значення: 0,5; 1,5;...; 7,5. Розташування рівнів обрано так, щоб вони рівномірно розташовувались в межах заданого діапазону значень сигналу. Кожному рівню присвоєний кодовий номер від 0 до 7. Отже кожний з цих кодових номерів можна представити 3-розрядним двійковим числом. Таким чином, на рисунку 27 кожна вибірка сигналу співвіднесена з одним з восьми рівнів або 3-бітовою послідовністю ІКМ.

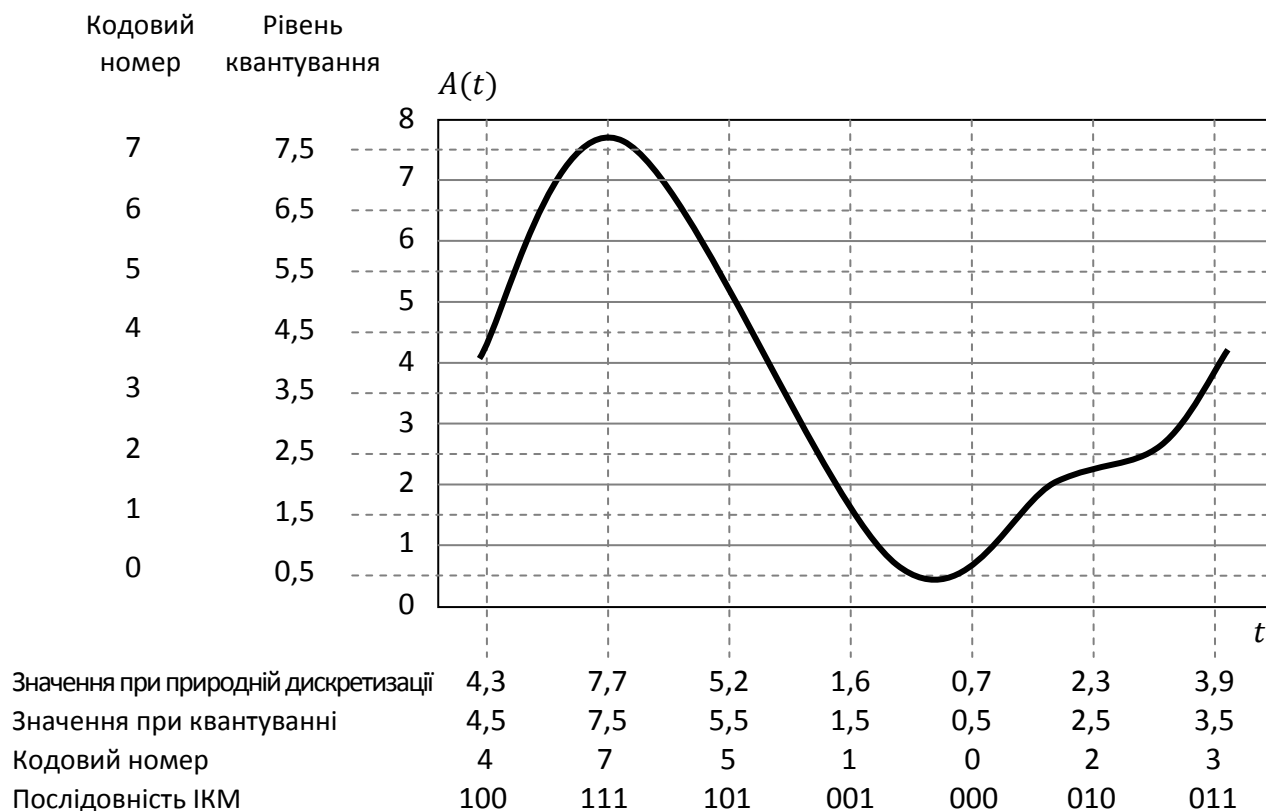


Рис. 27. Процес перетворення неперервного аналогового сигналу у послідовність ІКМ

Припустимо, що аналоговий сигнал є музичним фрагментом і при прослуховуванні музики у цифровій формі якість є поганою. Передусім це може бути пов'язано з високим рівнем шуму квантування. Для того, щоб зменшити шум квантування, потрібно збільшити кількість рівнів квантування. Це, у свою чергу, призведе до збільшення розрядності ІКМ послідовності. Як це позначиться на параметрах системи? Якщо параметри імпульсів, що реалізують задану ІКМ послідовність залишити без змін, то вочевидь збільшення розрядності призведе до збільшення часу передачі сигналу або, інакше кажучи, до зменшення швидкості передачі. Такий варіант є прийнятним тільки в тому випадку, якщо для системи критичним параметром є ширина смуги передачі, тобто вона повинна залишатися постійною незалежно від того, скільки бітів припадає на вибірку. Але в системах реального часу будь-які сигнали повинні передаватися в мінімальними затримками. Отже, якщо час передачі не повинен змінюватися при збільшенні розрядності ІКМ послідовності, то за це доведеться розраховуватися збільшенням ширини смуги передачі сигналу, оскільки потрібно буде збільшувати швидкість передачі даних.

Таким чином, якість перетворення аналогового сигналу у цифрову форму залежить не тільки від розрядності АЦП, а і визначається тими обмеженнями, які накладаються на ширину смуги сигналу. Але як визначити, яка ширина смуги сигналу буде у ІКМ послідовності певної розрядності? Само по собі використання ІКМ не дає нам на виході нічого "фізичного", тільки цифри. Цифри – це тільки абстракція, яка дозволяє описати інформацію, що міститься у повідомленні. Отже необхідно мати щось фізичне, що буде представляти цифри або "бути носієм" цифр. Так двійкові числа в лінії передачі можна представляти електричними імпульсами. Існує багато реалізацій представлення сигналів ІКМ електричними імпульсами. На рисунку 28 показаний приклад представлення двійкового числа "1011100101" деякими типами сигналів ІКМ.

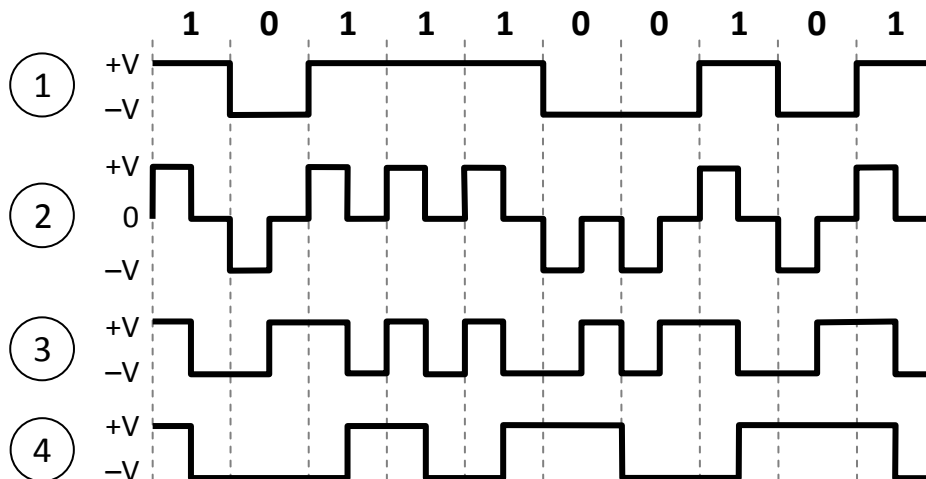


Рис. 28. Приклади реалізацій сигналів ІКМ: 1 – кодування NRZ-L, 2 – біполярне RZ кодування, 3 – манчестерське кодування, 4 – модуляція затримки

Причина існування великої кількості різних реалізацій сигналів ІКМ полягає у відмінностях, що пов'язані з продуктивністю, яка характеризує кожний тип ІКМ кодування. Зокрема, різні типи ІКМ сигналів вирізняються такими параметрами як ширина смуги частот та ефективність її використання, можливість забезпечення надійної синхронізації, наявність вбудованих засобів виявлення помилок, завадостійкість та інше. Вибір типу ІКМ сигналу, що буде використовуватись, залежить від особливостей конкретної системи та вимог, які до неї висуваються. Детальніше з різними типами сигналів ІКМ та їх характеристиками можна ознайомитись в [11].

Контрольні запитання

1. Що за визначенням називають дискретизацією сигналу?
2. Що є результатом дискретизації сигналу?
3. Яку схему використовують зазвичай при технічній реалізації процесу дискретизації?
4. Якою повинна бути частота дискретизації сигналу згідно теореми Котельникова-Найквіста?
5. До яких наслідків призводить неправильний вибір частоти дискретизації?
6. Яку частоту називають частотою Найквіста в цифровій обробці сигналів?
7. Який пристрій відіграє головну роль при перетворенні аналогового сигналу у цифрову форму?
8. Що таке передискретизація і для чого вона потрібна?
9. Для чого при дискретизації використовують попередню низькочастотну фільтрацію сигналу?
10. Що таке квантування сигналу?
11. Що таке шум квантування і як він впливає на якість цифрового звуку?
12. Що таке помилки насичення?
13. Як відбувається перетворення значень сигналу у двійковий код?
14. За допомогою яких сигналів закодовані двійковим кодом дані передаються цифровим каналом зв'язку?
15. Які чинники впливають на якість отриманого звукового сигналу при його перетворенні з аналогової форми у цифрову?

Глава 5. Обробка звуку. Стиснення звукової інформації

5.1. Алгоритми компресії цифрових аудіоданих

Представлення високоякісних звукових сигналів у цифровій формі широко використовується в системах радіомовлення та телебачення, в системах запису, зберігання та відтворення інформації, а також при обробці, монтажу та

редагуванні фонограм, при створенні нових незвичних видів звучання, в мультимедійних технологіях тощо.

При первинному кодуванні в ході перетворення аналогового звукового сигналу в цифрову форму зазвичай використовується рівномірне квантування відліків цього сигналу з роздільною здатністю $\Delta A = 16 - 24 \text{ біт/відлік}$ при частоті дискретизації $F_d = 44,1 - 96 \text{ кГц}$. У каналах студійної якості зазвичай $\Delta A = 16 \text{ біт/відлік}$, $F_d = 48 \text{ кГц}$, смуга частот кодованого звукового сигналу $\Delta F = 20 - 20000 \text{ Гц}$. Динамічний діапазон такого цифрового каналу складає близько 54 дБ. Якщо $F_d = 48 \text{ кГц}$ та $\Delta A = 16 \text{ біт/відлік}$, то швидкість цифрового потоку при передачі одного такого сигналу дорівнює $48 \times 16 = 768 \text{ кбіт/с}$. Це вимагає сумарної пропускної здатності каналу зв'язку при передачі звукового сигналу форматів 5.1 (Dolby Digital) або 3/2 плюс канал наднизьких частот (Dolby Surround, Dolby-Pro-Logic, Dolby THX) більше 3,840 Мбіт/с. Але людина здатна своїми органами відчуття свідомо обробляти лише близько 100 біт/с інформації. Тому можна говорити про властиву первинним цифровим звуковим сигналам значну надлишковість.

Статистична надлишковість обумовлена наявністю кореляційного зв'язку між сусідніми відліками часової функції звукового сигналу при його дискретизації. Для її зменшення застосовують досить складні алгоритми обробки. При їх використанні втрати інформації немає, проте вихідний сигнал виявляється представленим в компактнішій формі, що вимагає меншої кількості біт при його кодуванні. Важливо, щоб всі ці алгоритми дозволяли б при зворотному перетворенні відновлювати вихідні сигнали без спотворень. Проте, навіть при використанні досить складних процедур обробки усунення статистичної надлишковості звукових сигналів дозволяє зрештою зменшити необхідну пропускну здатність каналу зв'язку лише на 15–25% в порівнянні з її початковою величиною, що ніяк не можна вважати значним досягненням.

Після усунення статистичної надлишковості швидкість цифрового потоку при передачі високоякісних звукових сигналів і можливості людини по їх обробці відрізняються, принаймні, на декілька порядків. Це свідчить також про істотну психоакустичну надлишковість первинних цифрових звукових сигналів і, отже, про можливість її зменшення. Найбільш перспективними з цієї точки зору виявилися методи, що враховують такі властивості слуху, як маскування, передмаскування та післямаскування. Якщо відомо, які долі (частини) звукового сигналу вухо сприймає, а які ні внаслідок маскування, то можна вичленувати і

потім передати по каналу зв'язки лише ті частини сигналу, які вухо здатне сприйняти, а нечутні долі (складові вихідного сигналу) можна відкинути (не передавати по каналу зв'язку). Крім того, сигнали можна квантувати з якомога меншою роздільною здатністю по рівню так, щоб спотворення квантування, змінюючись по величині із зміною рівня самого сигналу, ще залишалися б нечутними, тобто маскувалися б вихідним сигналом. Проте, після усунення психоакустичної надлишковості точне відтворення форми часової функції звукового сигналу при декодуванні виявляється вже неможливим.

Найновіші дослідження, виконані в області компресії цифрових аудіоданих наприкінці XX та початку XXI століть, призвели до появи нових методів кодування:

- MUSICAM (Masking Pattern Adapted Universal Subband Integrated Coding and Multiplexing), розроблений спільно Інститутом техніки радіомовлення (IRT, Німеччина), фірмою Philips (Нідерланди) та Міжнародним консультативним комітетом по телеграфії та телефонії (CCITT, Франція);
- OCF (Optimum Coding in the Frequency Domain), розроблений в університеті Эрлангена-Нюрнберга (Німеччина);
- PEXFM (Perceptual Transform Coder), розроблений у AT&T Bell Labs (США);
- ASPEC (Adaptive Spectral Perceptual Entropy Coding), розроблений спільно фірмами AT&T Bell Labs (США), Deutsche Thomson Brandt (Франція), університетами Eriangen та Hannover (Німеччина), а також фірмою Fraunhofer Gesellschaft (Німеччина);
- ATRAC (Adaptive Transform Acoustic Coding), розроблений фірмою Sony.

На базі цих методів були розроблені європейські стандарти MPEG-1 ISO/IEC 11172-3 [12], MPEG-2 ISO/IEC 13818-3, 13818-7 [13, 14], MPEG-4 ISO/IEC 14496-3 [15], а також стандарт ATSC Dolby AC-3 (розроблений у США, як альтернатива стандартам MPEG) [16].

Не дивлячись на значне різноманіття алгоритмів компресії цифрових аудіоданих, структура кодера, що реалізує такий алгоритм обробки сигналів, може бути представлена у вигляді узагальненої схеми, що показана на рисунку 29. В основі цієї схеми лежить психоакустична модель, яка будується на основі використання феномену сприйняття людиною звуку для стиснення та передачі звукової інформації з втратами при одночасному збереженні її якості на достатньому рівні.

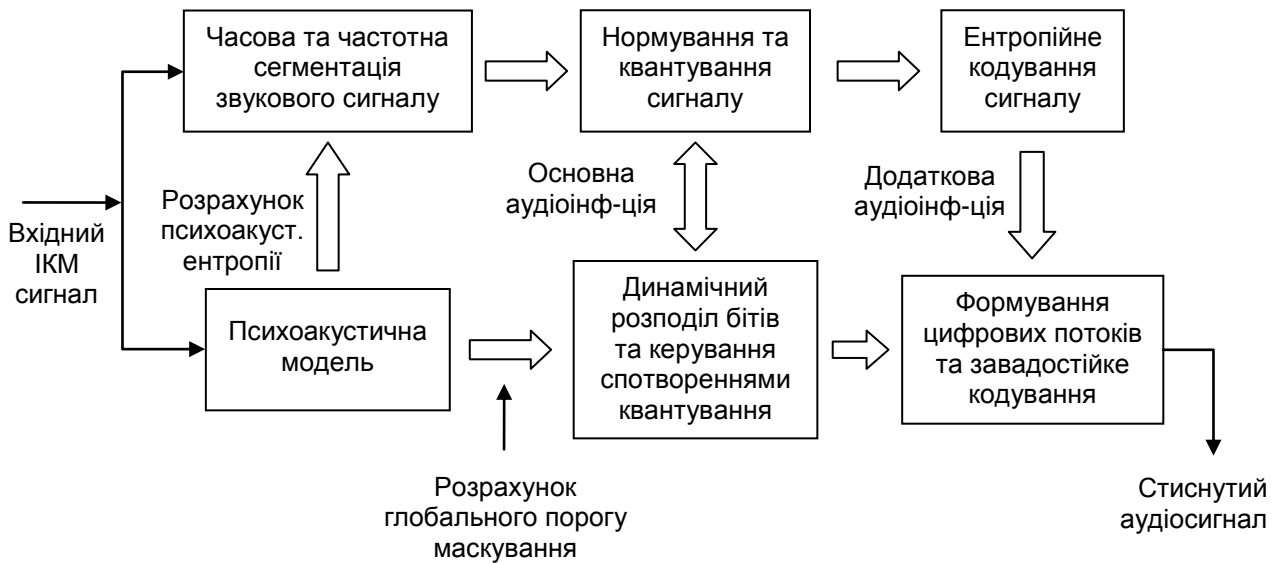


Рис. 29. Узагальнена структурна схема кодера з компресією цифрових аудіоданих

Початковий цифровий звуковий сигнал, що є нестиснутим ІКМ цифровим потоком даних, розділяється на частотні піддіпазони та сегментується за часом у блоці часової та частотної сегментації. Довжина сегменту, що кодується, залежить від форми часової функції звукового сигналу. При відсутності різких викидів по амплітуді використовується так звана довга вибірка, що забезпечує високу роздільну здатність за частотою. У випадку різкої зміни амплітуди сигналу довжина сегменту, що кодується, різко зменшується, що дає більшу роздільну здатність за часом. Рішення про зміну довжини сегмента, що кодується, приймає блок психоакустичної моделі, обчислюючи значення психоакустичної ентропії сигналу.

Психоакустична ентропія – це мінімальна можлива кількість бітів, яка необхідна для кодування даного звуку без відчутного погіршення якості. Для визначення цього параметру сигнал спочатку переводиться в частотно-часову область, а потім будуються глобальні пороги психоакустичного маскування, що є розподілом енергії сигналу за частотами та за часом, при яких досліджуваний сигнал буде маскуватися заданим маскуючим сигналом. Цей розподіл будується як суперпозиція порогів маскування від окремих гармонік сигналу, що маскує. Саме тому такий отриманий при цьому поріг маскування називається глобальним, оскільки він характеризує не окремі гармоніки, а весь

звук. Виходячи з глобальних порогів маскування, для кожного частотного піддіапазону розраховується максимально допустимий рівень спотворень (шумів) квантування, при якому вони ще маскуються корисним сигналом даного піддіапазону. Після цього оцінюється кількість бітів, якими необхідно квантувати спектр, щоб шум квантування не перевищував порогів маскування. Вочевидь така оцінка залежить від способу кодування звуку.

Після сегментації сигнали частотних піддіапазонів нормуються, квантуються та кодуються. В найефективніших алгоритмах компресії кодуванню піддаються не самі відліки сегментів звукового сигналу, а коефіцієнти, що їм відповідають.

Блок динамічного розподілу біт у відповідності з вимогами психоакустичної моделі для кожного піддіапазону кодування виділяє таку мінімально можливу їх кількість, при якій рівень спотворень, що викликаний квантуванням, не перевищує порогу їх чутності, що розрахований психоакустичною моделлю.

Для додаткового стиснення інформації виконується *ентропійне кодування* вихідного цифрового потоку. Ентропійним кодуванням називається кодування значень цифрової послідовності (з можливістю однозначного її відтворення) з метою зменшення об'єму інформації (довжини послідовності) за допомогою усереднення ймовірності появи символів послідовності. Суть цього кодування – зробити ентропію максимальною (ймовірності появи окремих символів послідовності є практично однаковими). Перед кодуванням окремі символи послідовності зазвичай мають істотно різну ймовірність появи. Після ентропійного кодування утворюються нові символи послідовності, ймовірність появи яких стає однаковою. Згідно теореми Шеннона, існує межа стиснення без втрат, яка залежить від ентропії джерела. Чим більш передбачуваною є інформація, тим краще її можна стиснути. Відповідно випадкова послідовність стисненню без втрат не піддається.

5.2. Сімейство стандартів MPEG

MPEG розшифровується як "Moving Picture Coding Experts Group", дослівно – група експертів по кодуванню рухливих зображень. Ця група була сформована для розробки стандартів стиснення та передачі цифрової відео та

аудіо інформації. MPEG веде свою історію з січня 1988 року. Починаючи з перших зборів в травні 1988 року, група почала зростати, і виросла до дуже великого колективу фахівців. Зазвичай, в зборах MPEG беруть участь близько 350 фахівців з понад 200 компаній. Велика частина учасників MPEG – це фахівці, зайняті в тих або інших наукових та академічних установах.

Детальний огляд стандартів MPEG представлений у [17]. Тут будуть розглянуті лише основні технології та принципи, покладені в основу цих стандартів, для кодування та стиснення звукової інформації.

5.2.1. Стандарт MPEG-1

Стандарт MPEG-1 (ISO/IEC 11172-3) включає три алгоритми різних рівнів складності: Layer (рівень) I, Layer II та Layer III. Прообразом для Layer I та Layer II став стандарт MUSICAM. Цим ім'ям іноді називають Layer II. Layer III досягає найбільшого стиснення, але, відповідно, потребує більше ресурсів на кодування. Загальна структура процесу кодування однакова для всіх рівнів. Проте, не дивлячись на схожість рівнів в загальному підході до кодування, рівні розрізняються по цільовому використанню та внутрішнім механізмам. Для кожного рівня визначений свій цифровий потік (загальна ширина потоку) та свій алгоритм декодування.

MPEG-1 призначений для кодування сигналів, оцифрованих з частотою дискретизації 32, 44.1 або 48 кГц. Кожен рівень, що використовується в MPEG-1 забезпечує різні коефіцієнти стиснення та різну якість звучання потоків, що отримуються. Відповідні коефіцієнти стиснення та швидкість передачі даних для одного каналу з частотою дискретизації 32 кГц представлені у таблиці 4.

Таблиця 4. Параметри MPEG-1 (затверджені у 1992 році)

Спосіб кодування	Швидкість передачі, kbps (кілобіт за сек.)	Коефіцієнт стиснення
Layer I	192	1:4
Layer II	128..96	1:6..8
Layer III	64..56	1:10..12

MPEG-1 нормує для всіх трьох рівнів наступні номінали швидкостей цифрового потоку: 32, 48, 56, 64, 96, 112, 192, 256, 384 і 448 *кбіт/с*, кількість рівнів квантування вхідного сигналу – від 16 до 24. Стандартним вхідним сигналом для кодера MPEG-1 прийнятий цифровий сигнал AES/EBU (двоканальний цифровий звуковий сигнал з розрядністю квантування 20 – 24 біта на відлік). Передбачаються наступні режими роботи звукового кодера:

- поодинокий канал (моно);
- подвійний канал (стерео або два моноканали);
- joint stereo (у даному режимі кодуються не лівий і правий канал, а їх сумарна складова та різницеві).

Найважливішою властивістю MPEG-1 є повна зворотна сумісність усіх трьох рівнів. Це означає, що кожен декодер може декодувати сигнали не лише свого, але й рівнів, що знаходяться нижче. MPEG-1 виявився першим міжнародним стандартом цифрового стиснення звукових сигналів і це зумовило його широке застосування в багатьох областях: мовленні, звукозаписі, зв'язку та у мультимедійних застосуваннях. Найширше використовується Layer II, який увійшов складовою частиною в європейські стандарти супутникового, кабельного і наземного мовлення цифрового ТБ, в стандарти звукового мовлення, записи на DVD та ін. Рівень III (його ще називають MP3) знайшов широке застосування в цифрових мережах з інтегральним обслуговуванням (ISDN) та в мережі Інтернет. Переважна більшість музичних файлів в мережі Інтернет записані саме в цьому стандарті.

Формат MP3

Формат стискування аудіоданих MP3 (скорочення від MPEG Layer III) — один з перших популярних способів стиснення звуку. Розроблений німецькою компанією Fraunhofer IIS за підтримки фірми THOMSON. Забезпечує високу якість звуку при порівняно невеликих розмірах файлу.

Високий ступінь стиснення в MP3 досягається за рахунок достатньо складного алгоритму кодування. Використовуються як математичні методи компресії, так і особливості людського слуху (психоакустична модель): ефект маскування слабкого звуку однієї частоти гучнішим звуком тієї ж або сусідньої частоти, пониження чутливості вуха до тихого звуку відразу після гучного, а також несприйнятність до звуків нижче певного рівня гучності. Потік звуку при кодуванні розбивається на рівні ділянки (фрейми). Кожен з фреймів кодується

окремо зі своїми параметрами і містить заголовок, в якому ці параметри вказані. Файли MP3 зазвичай закодовані з бітрейтом від 64 до 320 *кбіт/с*, а також із змінним бітрейтом (VBR) — коли для кожного фрейма використовується свій, оптимальний для даної ділянки, бітрейт.

Принцип кодування полягає у наступному. Початковий сигнал за допомогою фільтрів розділяється на декілька частотних діапазонів, для кожного діапазону визначається величина маскуючого ефекту від сусідніх діапазонів і попереднього фрейма, неістотні сигнали ігноруються. Для даних, що залишилися, для кожного діапазону визначається, скількома бітами можна пожертвувати, щоб втрати були нижчі за величину маскуючого ефекту. На цьому робота психоакустичної моделі завершується, а підсумковий потік додатково стискається по алгоритму Хаффмана (аналогічно RAR архіватору). При бітрейті 320 *кбіт/с* застосовується лише підсумкове стискування, без психоакустичного моделювання.

Перевагами стандарту MP3 є:

- високий ступінь стиснення при прийнятній якості звуку;
- ступінь стиснення і якість може регулюватися користувачем;
- фреймова структура зручна для передачі по мережі, дозволяє перехід до будь-якого місця файлу;
- широке поширення апаратури та програм.

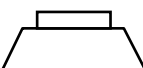
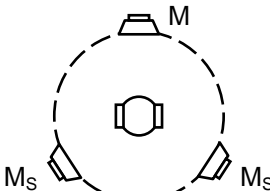
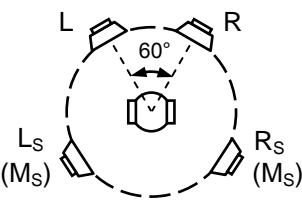
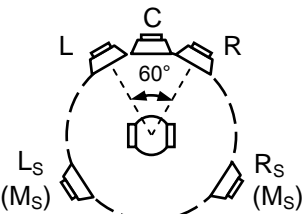
5.2.2. Стандарт MPEG-2

MPEG-2 це розширення MPEG-1 у бік багатоканального звуку. Наслідком сумісності MPEG-2 з MPEG-1 в частині кодування звуку стало повне використання трирівневої системи, розробленої в MPEG-1 для обробки звукових даних кодерами стандарту MPEG-2. Відмінності між стандартами починаються при переході від двоканального звуку, прийнятого за основу в MPEG-1, до багатоканального звуку, що підтримується в MPEG-2.

MPEG-2 специфікує відмінності режиму передачі багатоканального звуку, у тому числі п'ятиканальний формат, семиканальний звук з двома додатковими гучномовцями, що використовуються в кінотеатрах з дуже широким екраном, а також розширення цих форматів, що містять окремий низькочастотний канал.

Відповідне розташування гучномовців показано в таблиці 5. В даній таблиці в чисельнику дробу вказується число фронтальних каналів, в знаменнику – число каналів, що використовуються позаду.

Таблиця 5. Ієрархія багатоканальних звукових систем згідно стандартів MPEG-2

Система	Канали	Позначення	Розташування гучномовців
Моно	M	1/0	
Моно + моно	M	1/1	
Двоканальна стерео	L/R	2/0	
Двоканальна стерео + 1 оточуючий	L/R/M _s	2/1	
Двоканальна стерео + 2 оточуючих	L/R/L _s /R _s	2/2	
Триканальна стерео	L/C/R	3/0	
Триканальна стерео + 1 оточуючий	L/C/R/M _s	3/1	
Триканальна стерео + 2 оточуючих	L/C/R/L _s /R _s	3/2	

Одним з різновидів багатоканального звуку є багатомовний звуковий супровід. Він може здійснюватися або передачею окремого цифрового потоку для кожної мови, або додаванням декількох (до 7) мовних каналів із швидкістю 64 кбіт/с до багатоканального потоку 384 кбіт/с. Можлива передача додаткових звукових каналів для людей з погіршенням зору та слуху.

Система поліпшеного кодування звуку AAC

Однією з кращих сучасних систем стискування звуку визнана система AAC (Advanced Audio Coding – вдосконалена система кодування звуку), специфікована в сьомій частині стандарту ISO/IEC 13818. Формат AAC є складовою частиною стандарту MPEG-Audio. Специфікація формату була

розроблена тією ж фірмою, що розробила стандарт MP3 (Fraunhofer IIS). Технічні характеристики формату: забезпечення підтримки від 1 до 48 каналів з частотою дискретизації від 8 до 96 кГц, робота з бітрейтами від 8 кбіт/с для моно сигналу до 160 кбіт/с на канал для високоякісного кодування.

На даний момент найвідомішими є чотири різновиди формату AAC, що були створені в рамках стандарту MPEG-2:

1. Homeboy AAC;
2. AT&T a2b AAC;
3. LiquifierPRO AAC;
4. Astrid/Quartex AAC.

Всі ці модифікації несумісні між собою, мають власні кодери/декодери та є неоднаковими за якістю. Розглянемо зазначені формати детальніше.

Кодер Homeboy AAC став першим загальнодоступним кодером для першої версії AAC. Проте, його якість залишала бажати кращого. Єдиним плюсом цього кодера було досягнення якості MP3 з бітрейтом 128 кбіт/с на нижчих бітрейтах. На сьогодні цей формат є практично мертвим і майже не зустрічається.

В спеціалізованих лабораторіях потужної компанії AT&T був створений інший формат стиснення аудіо даних, що отримав назву a2b. Цей формат практично повністю наслідує ті ж самі принципи компресії, що і формат Homeboy AAC, проте в ньому були зроблені значні зусилля по підвищенню якості звучання. Складовою частиною a2b стали можливість включення зображення виконавця та тексту пісні у звуковий файл. По якості стиснення цей формат переважає MP3. Якість звучання a2b з бітрейтом 96 кбіт/с порівняна з якістю MP3, що має бітрейт 128 кбіт/с. Не дивлячись на очевидну перевагу формату, він не знайшов широкого вжитку передусім за рахунок того, що загальнодоступного кодера для нього не існувало.

Фірма Liquid Audio у боротьбі за першість в області високоякісної компресії звуку завдала нищівний удар по амбіціях інших конкурентів. Чітко дотримуючись усіх технічних тонкощів вона розробила версію формату AAC, яка по ступеню стиснення та якості звучання переважає усі відомі на сьогодні формати MPEG-2. Нажаль, цей формат так само, як і формат від компанії AT&T, не має загальнодоступного кодера.

Формат Astrid/Quartex AAC був розроблений одним нікому невідомим програмістом. На основі опублікованих фірмою NTT вихідних кодів MP4, він створив свою власну систему компресії. За ступенем стискування та якістю звучання формат перевершує всі інші окрім LiquifierPro AAC і є об'єктивно кращим серед тих, що мають загальнодоступні кодери та доступні для поширення в Інтернет звукові файли. У порівнянні з MP3 забезпечується аналогічна якість при потоці на 30% менше. Також можливе потокове відтворення.

5.2.3. Стандарт MPEG-4

Основи розробки стандарту MPEG-4 закладені групою учених з MPEG ще у 1993 році, і вже до кінця 1998 року відбулося затвердження першого стандарту. Згодом стандарт неодноразово допрацьовувався. У 1999 році він отримав офіційний статус, а потім був стандартизований з боку ISO/IEC.

Метою створення MPEG-4 була розробка стандарту кодування, який забезпечив би розробників універсальним засобом для стискування медіа даних як природного (записаного за допомогою мікрофону), так і штучного (синтезованого або генерованого на комп'ютері) походження. Ця обставина кардинальним чином відрізняє MPEG-4 від його попередників MPEG-1 та MPEG-2, в яких ефективне стискування даних досягається лише стосовно природного сигналу.

Як засоби компресії звуку в MPEG-4 (ISO/IEC 14496-3) використовується комплекс декількох стандартів кодування звуку: покращений алгоритм MPEG-2 AAC (AACPlus), алгоритм TwinVQ, а також алгоритми кодування мови HVXC та CELP. Крім того, MPEG-4 передбачає безліч механізмів забезпечення масштабованості та передбачення.

Стандарт AACPlus в цілому є продовженням стандарту MPEG-2 AAC. Завдяки своїй ефективності в передачі якісного звуку формат відразу отримав досить широке визнання: він забезпечує якість близьку до компакт-диска при стереопотоці 48 *кбіт/с*, високоякісне стерео при потоці 32 *кбіт/с* та якісний моносигнал при 20 *кбіт/с*. А при потоці 128 *кбіт/с* AACPlus забезпечує передачу багатоканального звуку формату 5.1.

TwinVQ (Transform-domain Weighted Interleave Vector Quantization – трансформаційно-областне зважувань багатoshарове векторне квантування) – це ще один зі способів стиснення аудіо даних, який був розроблений NTT Human Interface Laboratories в Японії. Він оптимізований для наднизьких бітрейтів (біля 8 кбіт/с). Заснований на звичайному методі кодування, який подібний до методів, що використовуються в MP3 та AAC. Проте, у цьому методі, індивідуальні біти аудіо даних безпосередньо не кодуються, а об'єднуються в сегменти (вектора). Ці вектора порівнюються із стандартними зразками, які підготовлені заздалегідь. Вибирається стандартний вектор, який забезпечує найближчу відповідність, і число пов'язане з цим зразком передається як код стиснення. Метод TwinVQ має дуже високу ефективність для низьких бітрейтів, а також володіє стійкістю до помилок та втрати пакетів. Водночас, швидкість стискання порівняно з MP3 є дуже низькою, а висока якість досягається великими затратами процесорних ресурсів. Кодек TwinVQ підтримує наступні постійні бітрейти: 80, 96, 112, 128, 160 та 192 кбіт/с .

HVXC (Harmonic Vector Excitation Codec – кодування шляхом використання збудження вектора гармонік) – це мовний кодек, який призначений для кодування та декодування первинного мовного сигналу з цифровою швидкістю від 2 до 4 кбіт/с . HVXC кодер здійснює кодування дзвінких звуків мови методом лінійного передбачення та кодування глухих звуків мови методом векторного збудження. Цей кодек забезпечує стандартну якість мови (близьку до якості міжнародного телефонного зв'язку) із смугою частот від 100 до 3800 Гц та частотою дискретизації 8 кГц . Алгоритм кодування HVXC характеризується високою завадостійкістю, що дозволяє застосовувати його в каналах з високим рівнем завад. Для цього кодер має нескладний алгоритм маскування помилок.

CELP (Codebook Excited Linear Prediction – лінійне передбачення із збудженням на основі кодової книги) – це стандарт кодування мови, який включає принципи стиснення та розшифровки природного звуку мови з цифровими швидкостями передачі даних від 4 до 24 кбіт/с . Пристрій кодування працює з частотами дискретизації 8 або 16 кГц , забезпечуючи смуги кодованого сигналу відповідно $100 - 3800 \text{ Гц}$ та $50 - 7000 \text{ Гц}$. В основі алгоритму CELP лежить лінійне передбачення з кодовим збудженням. Спочатку в CELP кодері за допомогою процедури "аналіз через синтез" з адаптивної кодової книги обирається вектор (сигнал) збудження, що є найбільш

наближеним до вхідного сигналу. Далі обраний сигнал фільтрується за допомогою фільтра-синтезатора (предиктора) та порівнюється з оригінальним сигналом. Ця процедура повторюється з метою мінімізації похибки передбачення. Сигнал збудження, який дасть мінімум зваженої похибки між оригіналом та синтезованою мовою, обирається як результат кодування. Параметри збудження разом з параметрами предиктора складають вихідну інформацію, за допомогою якої після декодування в приймачі синтезується мовний сигнал.

Для перетворення мовних сигналів на даний час широко застосовуються три основні типи пристроїв: кодери форми, вокодери та гібридні кодери [18].

Кодери форми характеризуються здатністю зберігати основну форму мовного сигналу. До кодерів форми відносяться кодери з ІКМ, кодери з диференційною ІКМ (ДІКМ), кодери з адаптивною диференційною ІКМ (АДІКМ) та інші.

При використанні ІКМ дискретне повідомлення піддається рівномірному квантуванню, в результаті якого отримується двійковий код розрядності n . Для досягнення прийнятної якості відновлення мовного повідомлення необхідно $n \geq 12$. Далі відбувається цифрова компресія, в результаті якої довжина кодової комбінації зменшується до $n = 8$ розрядів. Для передачі такого цифрового потоку, що отриманий за допомогою ІКМ, необхідне виділення смуги пропускання 64 кбіт/с (рекомендація ITU G.771).

Метод ДІКМ є ефективнішим за ІКМ. В цьому методі для зниження вимог до пропускної здатності каналу використовується наявність кореляції між значеннями відліків повідомлення, що передається. Це дозволяє, зберігаючи ту ж саму якість сигналу, що і в ІКМ, суттєво зменшити кількість рівнів квантування і, як наслідок, використати меншу кількість розрядів в кодових комбінаціях.

Адаптивна диференційна ІКМ (АДІКМ) – один з найпоширеніших та давно використовуваних алгоритмів стиснення мови. Цей алгоритм дає практично таку ж саму якість відтворення мови, як і ІКМ, проте для передачі інформації при його використанні потрібна пропускна здатність лише 32 кбіт/с . Метод заснований на тому, що в аналоговому сигналі, що передає мову, неможливі різкі стрибки інтенсивності. Тому, якщо кодувати не саму амплітуду сигналу, а її зміну порівняно з попереднім значенням, то можна обійтися меншою кількістю розрядів. В АДІКМ зміна рівня сигналу кодується 4-розрядним двійковим числом.

Кодери форми забезпечують гарну якість відтворення мовних сигналів (стандартна смуга частот 300 – 3400 Гц), а також і більш широкосмугових звукових сигналів. Проте, ці кодери є малоефективними з точки зору зниження швидкості передачі цифрових сигналів.

Вокодери (від англ. Voice – голос; Coder – кодуєчий пристрій) пристрої, що працюють за принципом кодування спектру мовного сигналу та відновлення аналогового сигналу, що звучить подібно до оригіналу, але не повторює його форми. Вокодери забезпечують значно більше зниження потрібної для передачі пропускної здатності каналу. Кодування мови відбувається на швидкостях від 1,2 до 9,6 кбіт/с. Стиснення на тому боці, що передає, відбувається в аналізаторі, який виділяє з мовного сигналу складові, що повільно змінюються, які потім передаються по каналу зв'язку у вигляді кодових комбінацій. На боці, що приймає, з цих складових синтезується мовний сигнал.

Гібридні кодери використовують кодування як форми, так і спектру мовного сигналу і базуються на методах вкладеного кодування [19].

5.2.4. Стандарт MPEG-7

Аудіовізуальна інформація відіграє важливу роль в житті суспільства – незалежно від того, записана вона на магнітну чи фотоплівку або потрапляє до нас у реальному часі, чи є вона аналоговою, чи цифровою. При цьому інформація може бути представлена у самих різноманітних медіа-форматах. Зазвичай аудіовізуальна інформація завжди була призначена для безпосереднього сприйняття людиною. Проте, з кожним роком збільшується кількість випадків, коли ця інформація створюється, передається, дістається та знову використовується обчислювальними системами: при розпізнанні голосу чи зображень, при медійних перетвореннях (мови в текст, зображення в мову, мови в зображення и т.п.), при виділенні та фільтрації потоку описів інформації (для швидкого та ефективного пошуку медійних документів та відбору з них тільки тих, які відповідають уподобанням користувача). Сьогодні є доступною велика кількість аудіовізуальної інформації. Проте, цінність інформації, що отримується, дуже часто залежить від того, наскільки просто її можна знайти, дістати, відфільтрувати або керувати нею.

У 1996 році група MPEG приступила до створення нового формату MPEG-7, який повинен був розв'язати перераховані вище задачі та визначити універсальні механізми опису візуальної та аудіо інформації. Цей формат отримав назву "Мультимедіа інтерфейс для опису вмісту" (англ. Multimedia Content Description Interface). На відміну від попередніх форматів стиснення сімейства MPEG, формат MPEG-7 описує інформацію у будь-якій формі (в тому числі і аналоговій) та не залежить від середовища передачі даних.

Стандарт аудіо MPEG-7 (ISO/IEC 15938), описаний в його четвертій частині [20], включає п'ять технологій:

- 1) структура опису звуку, яка містить у собі масштабовані послідовності, дескриптори нижнього рівня та уніфіковані сегменти тиші;
- 2) засоби опису тембру музичного інструменту;
- 3) засоби розпізнавання звуку;
- 4) засоби опису голосового матеріалу;
- 5) засоби опису мелодії.

Аудіоструктура містить широкий набір засобів нижнього рівня, які забезпечують основи для формування звукових прикладень високого рівня, що потребують ефективного доступу до медійного контенту. Такими засобами можуть бути елементи метаданих, їх структура та взаємозв'язки, що визначаються стандартом в формі дескрипторів та схем опису. Через надання загальної платформи структури описів MPEG-7 встановлює базис для сумісності усіх прикладень, які можуть бути створені в рамках даної системи.

В область дії MPEG-7 входять як користувачі-люди, так і автоматизовані системи, які виконують обробку аудіовізуального матеріалу. Цей стандарт надає високоякісний доступ до медіа матеріалу, що передбачає рішення для запису, ідентифікації матеріалу, забезпечення прав власності, швидкого, зручного та точного фільтрування або пошуку. MPEG-7 робить можливим пошук медійного контенту в Web так само простим, як і пошук текстової інформації. Наприклад, програвте декілька нот на клавіатурі і отримайте перелік музичних уривків, які схожі з мелодією, що програвалась. Або, використовуючи фрагмент голосу Паваротті, отримайте список його записів, відеокліпів або інший матеріал, що до нього відноситься.

Додаткову інформацію щодо стандарту MPEG-7 можна знайти в [17].

5.3. Альтернативні формати кодування звуку

Метод стиснення звуку Ogg Vorbis

Відразу після своєї появи формат MP3 отримав величезну популярність у користувачів персональних комп'ютерів. Причиною цього стало те, що на компакт-диск можна було помістити в 10 разів більше музики, ніж на звичайний аудіо-диск, зберігаючи при цьому достатню якість звучання. Створені таким чином файли можна легко пересилати через Інтернет, користуватися ними у переносних пристроях, збирати музичні колекції.

Ogg Vorbis належить до того ж типу форматів стиснення аудіо, тобто до форматів стиснення з втратами. Психоакустична модель, що використовується в Ogg Vorbis, за принципами дії близька до MP3, але математична обробка та практична реалізація докорінно відрізняються.

У формату є велика кількість переваг. Наприклад, формат Ogg Vorbis не обмежує користувача тільки двома аудіо каналами (стерео лівий та правий). Він підтримує до 225 окремих каналів з частотою дискретизації до 192 кГц та розрядністю до 32 біт (що не дозволяє жоден формат стиснення з втратами). Тому Ogg Vorbis чудово підходить до кодування 6-канального звуку DVD-Audio. До того ж, формат Ogg Vorbis має таку властивість як "sample accurate", яка гарантує, що звукові дані перед кодуванням та після декодування не будуть мати зміщень або додаткових/втрачених відліків відносно один одного.

Ogg Vorbis з самого початку розроблявся як мережевий потоковий формат. Це дає формату корисний побічний ефект – в одному файлі можна зберігати декілька композицій з власними тегами. При завантаженні такого файлу у програвач будуть відображатися усі композиції так, начебто їх завантажили з різних файлів.

Окремо варто додати, що теги в цьому форматі мають доволі гнучку систему. Заголовок тегів може містити у собі тексти будь-якої довжини та складності, що може перемежатися зображеннями (наприклад, це може бути текст пісні та фотографія з обкладинки). Текстові теги зберігаються у форматі UTF-8, що виключає проблеми з відображенням тексту з різними кодуваннями на різних мовах.

За умовчанням Ogg Vorbis використовує змінний бітрейт, причому його значення не мають обмежень (він може варіюватися від 1 кбіт/с до 700 кбіт/с). Такою ж гнучкістю володіє і частота дискретизації, яка може

змінюватися від 2000 Гц до 192 кГц. Це дозволяє суттєво зменшити розмір композиції при незначній втраті якості. Що ж стосується швидкості кодування, то тут у цього формату порівняно з MP3 немає ніякої різниці.

Але найголовніша перевага формату – це його повна відкритість та безкоштовність. І хоча існує багато інших безкоштовних форматів, але автори цих форматів, на відміну від Ogg Vorbis, не надають їх вихідних кодів. Основною метою розробки Ogg Vorbis було замінити усі платні запатентовані аудіо формати. Таким чином, формат Ogg Vorbis є повністю відкритим для комерційного та некомерційного використання. Його коди можна модифікувати без обмежень. Група розробників (співтовариство Xiphophorus) залишає за собою лише право затверджувати нові специфікації формату.

Метод стиснення звуку MusePack

Стиснення звуку, що реалізує Ogg Vorbis, не єдина некомерційна розробка такого роду. Кодек MusePack (MPC) – це ще один з різновидів стиснення звуку з втратами (як і MP3). Формат був створений студентом Андре Бушманом в домашніх умовах (фактично "на колінках") у вільний від навчання час. Мотивацією цієї розробки стало незадоволення якістю перших MP3-кодерів (1997-1998 рр.). Спочатку цей кодек був названий MPEG-plus (MPEG+), але його довелося переіменувати через проблеми, які виникли у автора у зв'язку з тим, що назва кодека містила у собі абревіатуру "MPEG".

MPC не є продовженням MPEG-1 Layer III, а лише бере свій початок в MPEG-1 Layer II. Його направленість – це кодування на високих бітрейтах. В цьому кодеку передбачено кодування тільки в режимі із змінним бітрейтом.

У середньому якість кодування MPC на високих бітрейтах (160 кбіт/с та вище) помітно вища за якість, що забезпечує MP3. За цим параметром він випереджає всі існуючі сьогодні кодекі. Це пов'язано з відмінностями в механізмах кодування. MPC після розбиття сигналу на частотні підсмуги просто робить переквантування сигналу (спираючись на психоакустичну модель) в кожній підсмугі і отримані значення записує у вихідний потік. Цим також пояснюється велика швидкість компресії та декомпресії MPC. Але, разом з тим, на низьких бітрейтах цей кодек має низьку якість.

Переваги формату:

- оскільки формат не робить повторне dct-перетворення (дискретне косинус-перетворення), то він фактично не страждає артефактами "pre-echo" (коли

окремі тони інструменту починають звучати раніше, ніж починає звучати сам інструмент), на відміну від таких форматів, як MP3, Ogg Vorbis, AAC та WMA;

- ефективніші алгоритми змінного бітрейту;
- потужна та гнучка психоакустична модель;
- дуже швидке кодування та декодування;
- відкритий формат;
- властивість "sample accurate";
- підтримка потокової передачі даних;
- підтримка частот дискретизації більших за 48 кГц.

Недоліки формату:

- низька розповсюдженість та повна відсутність апаратної підтримки;
- низька якість на низьких та середніх бітрейтах;
- незрозуміла підтримка більш ніж 2-х каналів;
- відсутність кодування з розрядністю 24 біта та вище.

Сам кодек сьогодні переживає не кращі часи. Він залишається маловідомим, хоча здебільшого і не заслуговує на це.

Формат Windows Media Audio (WMA)

WMA — ліцензований формат файлу, розроблений компанією Microsoft для зберігання і трансляції аудіо-інформації. З самого початку формат WMA позиціонувався як альтернатива MP3. Не дивлячись на те, що цей формат з'явився найостаннішим з усіх конкурентів MP3, його історія почалася значно раніше.

По-перше, в цьому кодеку була вперше досягнута якість MP3 128 *кбіт/с* на швидкості всього 64 *кбіт/с*.

По-друге, на цей формат відразу звернули увагу декілька крупних телефонних компаній, оскільки стиснена цим кодеком голосова інформація навіть при швидкості 64 *кбіт/с* володіла дуже високою розбірливістю (до 90%). Для досягнення цього в інших форматах була необхідна швидкість у 2 — 2,5 рази більша.

Новий формат, який назвали Voxware, як виявилось, ідеально підходить саме для стиснення оцифрованої людської мови. Підтримка цього кодека була включена у безкоштовний Media Player. В ньому він так і називається "Voxware Audio Codec". Цей кодек дозволяє легко перекодувати MP3 в WMA. Сьогодні вже існують і апаратні плеєри з підтримкою цього формату.

З точки зору якості звучання формат WMA практично є ідентичним формату MP3. В чомусь він переважає MP3, а в чомусь поступається. Якщо на низьких бітрейтах цей формат явно переважає конкурентів, то на бітрейті вже 128 *кбіт/с* він уступає по якості не тільки компакт-дискам, але і MP3.

Цікавою особливістю формату є захист авторських прав, який реалізований за допомогою спеціальної цифрової системи керування авторськими правами – DRM (Digital Rights Management). Якщо не відключити при кодуванні захист DRM, то не можливо буде відтворити створений файл на будь-якому іншому комп'ютері. Для цього потрібно завантажити спеціальний сертифікат, який дозволяє відтворювати файл на протязі обмеженого часу, після якого необхідно буде купити цей файл або цілий диск.

В останніх версіях формату передбачено кодування без втрати якості, а також багатоканальне кодування об'ємного звуку та голосу.

Формат стиснення звуку QDesign AIF

Цей формат був розроблений компанією QDesign при підтримці компанії Apple/Macintosh. QDesign AIF є доробкою сімейства стандартів AIFF, яке є різновидом мультимедійних стандартів, що використовуються на платформі Apple. Пара AIF-AIFF є повним аналогом пари WAV-MP3 за виключенням ступеня стиснення. Багато хто вважає, що формат AIF є революційним. Було виявлено, що при кодуванні відповідний кодек практично не вносить жодних лінійних спотворень в сигнал, що кодується. В цьому він значно переважає такі формати як MP3, AAC та інші на бітрейті 24 *кбіт/с*. Для будь-яких видів кодування з стисненням відмінність від оригінального WAV файлу дуже добре помітна. Але для кодера від QDesign з бітрейтом 24 *кбіт/с* на канал в частотному діапазоні від 12 Гц до 24 кГц максимальне відхилення від оригіналу не перевищує 2 дБ! Але ці властивості AIFF безпосередньо проявляються не при прослуховуванні даних, записаних цим форматом, а при спробі перетворити AIFF назад у WAV. При реальному прослуховуванні 2-х каналного звуку, закодованого за допомогою кодера AIF з бітрейтом 48 *кбіт/с*, якість звучання можна порівняти хіба що з якістю MP3 на 64 *кбіт/с*. Крім того, швидкість кодування є дуже низькою. В даний час єдиною програмою, яка забезпечує кодування та відтворення цього формату на персональному комп'ютері, є QuickTime Player.

Формат стиснення звуку РАС

Назва цього формату розшифровується як perceptual audio coding, що важко перекладається з англійської (perceptual у перекладі з англійської – сприйняття). Найбільш благозвучний варіант – це "аудіо кодування, засноване на сприйнятті". Даний формат був розроблений фірмою Lucent Technologies при підтримці Bell Labs, яку цікавили системи стиснення мовної інформації, що передавалася цифровими телефонними мережами. Перша загальнодоступна версія цього кодека вийшла під назвою Audio Library 1.0.

До недоліків формату РАС можна віднести те, що його не можна використовувати як мережевий формат. Цей формат не підтримує потокову передачу даних, тобто одночасне відтворення та отримання аудіо даних. Основне місце застосування формату РАС – це домашня аудіотека та компакт-диски. Особливістю формату є потужний захист аудіо запису від нелегального копіювання. Тому формат РАС дуже погано працює з компакт дисками типу CD-R або CD-RW, якщо відтворення з цих дисків відбувається не на тому пристрої, де відбулося кодування. Кодекі РАС є швидкими та якісними, з гарним співвідношенням розмір/якість. Вони ідеально підходять для створення домашніх аудіотек. Проте, поділитися цими аудіотеками з іншими не вийде.

Контрольні запитання

1. Які чинники обумовлюють можливість стиснення звукових сигналів без виникнення спотворень при їх сприйнятті людиною?
2. Чому не можливо точно відновити форму часової функції звукового сигналу після усунення його психоакустичної надлишковості?
3. Що таке психоакустична ентропія?
4. Як і для чого визначаються глобальні пороги психоакустичного маскування?
5. В чому полягає суть ентропійного кодування при стисненні інформації без втрат?
6. Що каже теорема Шеннона щодо можливості стиснення інформації без втрат?
7. Які основні характеристики та властивості стандарту MPEG-1.
8. Які принципи кодування застосовуються у форматі MP3?

9. Які додаткові можливості порівняно із стандартом MPEG-1 з'явилися у стандарті MPEG-2?
10. З якою метою створювався стандарт MPEG-4?
11. Які алгоритми кодування мови використовуються в стандарті MPEG-4?
12. Які основні класи кодерів використовуються для кодування мови?
13. Які можливості надає користувачам стандарт MPEG-7?
14. Які альтернативні методи стиснення звуку існують наряду зі стандартами MPEG?
15. Звукові кодекі яких форматів підтримують потокову передачу даних?

Література

1. Бекеши Г. Механические свойства уха / Г. Бекеши, В. А. Розенблит. В кн. Экспериментальная психология (том 2). – М.: Изд-во иностр. лит, 1963. – 1035 с.
2. Цвикер Э. Ухо как приемник информации / Э. Цвикер, Р. Фельдкеллер. — М.: Связь, 1971. — 256 с.
3. Алдошина И. Основы психоакустики. – М.: Оборонгиз., 2000. – 155 с.
4. Fletcher H., Munson W. A. Loudness, its definition, measurement and calculation // J. Acoust. Soc Am. – № 5, 1933. – P. 82-108.
5. Robinson D. W., Dadson R. S. A re-determination of the equal-loudness relations for pure tones // Br. J. Appl. Phys. – № 7, 1956. – P. 166—181.
6. International Standard ISO 226:2003. Acoustics – Normal equal-loudness-level contours. – http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=34222
7. Thomas D. Rossing, F. Richard Moore, Paul A. Wheeler. The Science of Sound (3rd Edition). – Addison-Wesley Pub. Co., 2001. – 783 p.
8. Maher Robert C. Old and New Techniques for Artificial Stereophonic Image Enhancement. – Proc. of the 101st Audio Engineering. Soc. Convention, Preprint 4371, November, 1996. – 17 p.
9. Сергиенко А. Б. Цифровая обработка сигналов. – Спб.: БХВ-Петербург, 2011. – 758 с.
10. Радзишевский А. Ю. Основы аналогового и цифрового звука. – М.: Издательский дом "Вильямс", 2006. – 288 с.
11. Скляр Б. Цифровая связь. Теоретические основы и практическое применение. – М.: Издательский дом "Вильямс", 2003. – 1104 с.
12. International Standard ISO/IEC 11172-3. Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio, 1993-08-01. – http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=22412
13. International Standard ISO/IEC 13818-3. Information technology – Coding of moving pictures and associated audio information. – Part 3: Audio, 1998-05-15. – http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=26797
14. International Standard ISO/IEC 13818-7. Information technology – Generic coding of moving pictures and associated audio information. – Part 7: Advanced Audio Coding (AAC), 2006. – http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=43345

15. ISO/IEC FCD 14496-3 Subpart 1. Information Technology – Very Low Bitrate Audio-Visual Coding. Part 3: Audio (ISO/JTC 1/SC 29. – № 2203). – http://www.mp3-tech.org/programmer/docs/ISO_14496-3.pdf
16. Digital Audio Compression Standard (AC-3). Doc. A/52. – [http://www.atsc.org/cms/standards/A52-2012\(12-17\).pdf](http://www.atsc.org/cms/standards/A52-2012(12-17).pdf)
17. Артюшенко В. М., Шелухин О. И., Афонин М. Ю. Цифровое сжатие видеоинформации и звука: Учебное пособие / Под ред. В. М. Артюшенко. – М.: Издательско-торговая корпорация "Дашков и К°", 2003. – 426 с.
18. Шелухин О. И., Лукьянцев Н. Ф. Цифровая обработка и передача речи / Под ред. О. И. Шелухина – М.: Радио и связь, 2000. – 456 с.
19. Советов Б. Я., Яковлев С. А. Построение сетей интегрального обслуживания. – Л.: Машиностроение. Лен. отд., 1990. – 332 с.
20. International Standard ISO/IEC 15938. Information technology – Multimedia content description interface. – Part 4: Audio, 2002. – http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=34231

ЗМІСТ

Стор.

Вступ.....	3
Глава 1. Людський слух.....	4
1.1. Будова людського вуха.....	5
1.2. Розпізнавання звуків	8
Контрольні запитання	9
Глава 2. Сприйняття звуку людиною.....	10
2.1. Сприйняття звуку за частотою.....	11
2.2. Критичні смуги слуху	13
2.3. Поріг відчуття звуку та больовий поріг.....	15
2.4. Диференційний поріг сприйняття інтенсивності звуку	17
2.5. Пороги чутності при маскуванні	19
2.6. Рівень гучності та гучність	22
2.7. Адаптація слуху.....	27
2.8. Часові характеристики слуху	29
2.9. Бінауральний слух	31
Контрольні запитання	36
Глава 3. Обробка звуку. Звукові сигнали	37
3.1. Первинні та вторинні звукові сигнали.....	37
3.2. Динамічний діапазон	40
3.3. Частотний діапазон та спектри	52
3.4. Первинний мовний сигнал.....	55
Контрольні запитання	57
Глава 4. Обробка звуку. Цифрове представлення звукових сигналів	58
4.1. Процес аналого-цифрового та цифро-аналогового перетворення звукових сигналів	58
4.2. Дискретизація.....	60
4.3. Передискретизація (оверсемплінг)	63
4.4. Квантування	65
4.5. Цифрове кодування.....	66
Контрольні запитання	70

Глава 5. Обробка звуку. Стиснення звукової інформації.....	70
5.1. Алгоритми компресії цифрових аудіо даних	70
5.2. Сімейство стандартів MPEG	74
5.2.1. Стандарт MPEG-1	75
5.2.2. Стандарт MPEG-2	77
5.2.3. Стандарт MPEG-4	80
5.2.4. Стандарт MPEG-7	83
5.3. Альтернативні формати кодування звуку	85
Контрольні запитання	89
Література.....	91

Навчальне видання

КОЛЕНОВ *Сергій Олександрович*

МЕДІЙНІ ТЕХНОЛОГІЇ В РАДІОФІЗИЦІ

Частина 1

Людина і звук

Навчально-методичний посібник до лекційного курсу
для студентів радіофізичного факультету

Підписано до друку _____ . Формат 60x84¹⁶.

Гарнітура Calibri. Папір офсетний. Друк офсетний.

Наклад 100 примірників. Ум. друк. арк. 2.7.

Видавнича лабораторія радіофізичного факультету
Київського національного університету імені Тараса Шевченка