



УКРАЇНА

(19) UA (11) 93243 (13) C2
(51) МПК (2011.01)
H03G 3/00
H03G 7/00

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ДЕПАРТАМЕНТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ

ОПИС ДО ПАТЕНТУ НА ВИНАХІД

(54) РЕГУЛЮВАННЯ ПОСИЛЕННЯ ЗВУКУ З ВИКОРИСТАННЯМ ОСНОВАНОВОГО НА КОНКРЕТНІЙ ГУЧНОСТІ ВІЯВЛЕННЯ АКУСТИЧНИХ ПОДІЙ

1

(21) а200813668
(22) 30.03.2007
(24) 25.01.2011
(86) РСТ/US2007/008313, 30.03.2007
(31) 60/795,808
(32) 27.04.2006
(33) US
(46) 25.01.2011, Бюл.№ 2, 2011 р.
(72) КРОКЕТТ БРЕТТ ГРЕХЕМ, US, СІФЕЛЬДТ АЛАН ДЖЕФФРІ, US
(73) ДОЛБІ ЛЕБОРЕТЕРІЗ ЛАЙСЕНСІНГ КОРПОРЕЙШН, US
(56) US 2004165730 A1; 26.08.2004
(57) 1. Спосіб зміни параметра динамічної обробки звукового сигналу, який включає етапи, на яких: виявляють зміни спектральних характеристик з часом в звуковому сигналі, ідентифікують, що границі акустичної події змінюються сильніше, ніж порогова величина в спектральних характеристиках з часом у згаданому звуковому сигналі, причому звуковий сегмент між послідовними границями становить акустичну подію, генерують керуючий сигнал, який змінює параметр, на основі згаданих ідентифікованих границь події і змінюють параметр динамічної обробки звукового сигналу залежно від керуючого сигналу.
2. Спосіб за п. 1, в якому параметром є одне з часу наростання, часу повернення і відношення.
3. Спосіб за п. 1, в якому змінним параметром є згладжуюча посилення стала часу.
4. Спосіб за п. 3, в якому згладжуючою посилення сталою часу є згладжуюча посилення стала часу наростання.
5. Спосіб за п. 3, в якому згладжуючою посилення сталою часу є згладжуюча посилення стала часу повернення.
6. Спосіб за будь-яким з пп. 1-5, в якому згаданий керуючий сигнал, який змінює параметр, оснований на положенні згаданих ідентифікованих границь акустичної події і ступені зміни спектральних характеристик, зв'язаних з кожною зі згаданих границь акустичної події.
7. Спосіб за п. 6, в якому при генерації керуючого сигналу, який змінює параметр:

2

забезпечують імпульс на кожній границі акустичної події, причому кожний такий імпульс має амплітуду, пропорційну ступеню згаданих змін спектральних характеристик, і згладжують за часом кожний такий імпульс, щоб його амплітуда плавно спадала до нуля.

8. Спосіб за будь-яким з пп. 1-7, в якому зміни спектральних характеристик з часом виявляють, порівнюючи різниці в конкретній гучності.

9. Спосіб за п. 8, в якому згаданий звуковий сигнал представляють дискретною часовою послідовністю $x[n]$, яка одержана взяттям вибірок відносно джерела звукового сигналу при частоті f_s дискретизації, і зміни спектральних характеристик з часом обчислюють, порівнюючи різницю в конкретній гучності $N[b, t]$ по смугах b частот між послідовними блоками t часу.

10. Спосіб за п. 9, в якому різницю в спектральному складі між послідовними блоками часу звукового сигналу обчислюють відповідно до

$$D[t] = \sum_b |N_{\text{NORM}}[b, t] - N_{\text{NORM}}[b, t-1]|,$$

де

$$N_{\text{NORM}}[b, t] = \frac{N[b, t]}{\max_b \{N[b, t]\}}.$$

11. Спосіб за п. 9, в якому різницю в спектральному складі між послідовними блоками часу звукового сигналу обчислюють відповідно до

$$D[t] = \sum_b |N_{\text{NORM}}[b, t] - N_{\text{NORM}}[b, t-1]|,$$

де

$$N_{\text{NORM}}[b, t] = \frac{N[b, t]}{\text{avg}_b \{N[b, t]\}}.$$

12. Апаратура для зміни параметра динамічної обробки звукового сигналу, яка містить:

засіб для виявлення змін спектральних характеристик з часом у звуковому сигналі, засіб для ідентифікації того, що границі акустичної події змінюються сильніше, ніж порогова величина в спектральних характеристиках з часом у згаданому звуковому сигналі, причому звуковий сегмент між послідовними границями становить акустичну подію,

(13) C2

(11) 93243

(19) UA

засіб для генерації керуючого сигналу, що змінює параметр, на основі згаданих ідентифікованих границь події і

засіб для зміни параметра динамічної обробки звукового сигналу залежно від керуючого сигналу.

Галузь техніки, якої стосується винахід

Даний винахід стосується способів регулювання динамічного діапазону звуку та апаратури, в якій пристрій обробки звуку аналізує звуковий сигнал і змінює рівень, коефіцієнт посилення або динамічний діапазон звуку, і всі або деякі параметри посилення звуку і динамічної обробки генеруються у вигляді функції акустичних подій. Винахід також стосується комп'ютерних програм для здійснення на практиці таких способів або керування такою апаратурою.

Даний винахід також стосується способів та апаратури, що використовують основане на конкретній гучності виявлення акустичних подій. Винахід також стосується комп'ютерних програм для здійснення на практиці таких способів або керування такою апаратурою.

Попередній рівень техніки винаходу

Динамічна обробка звуку

Технології автоматичного регулювання посилення (AGC) і регулювання динамічного діапазону (DRC) добре відомі і є звичайним елементом багатьох шляхів звукового сигналу. В абстрактному значенні обидві технології деяким чином вимірюють рівень звукового сигналу і потім змінюють коефіцієнт посилення сигналу на величину, яка є функцією вимірюваного рівня. У лінійній динамічній системі зі ступенем компресії (стиснення) 1:1 вхідний сигнал не обробляється, і вихідний сигнал ідеально співпадає з вхідним звуковим сигналом. Додатково, якщо має місце динамічна система обробки звуку, яка автоматично вимірює характеристики вхідного сигналу і використовує це вимірювання для регулювання вихідного сигналу, якщо рівень вхідного сигналу збільшується на 6 дБ, а вихідний сигнал обробляється таким чином, що його рівень збільшується тільки на 3 дБ, ступінь компресії для вихідного сигналу становить 2:1 відносно вхідного сигналу. Міжнародна публікація під номером WO 2006/047600 A1 («Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal», Alan Jeffrey Seefeldt («Обчислення та настройка сприйманої гучності і/або сприйманий спектральний баланс звукового сигналу», Алан Джеффрі Сіфелдт)) забезпечує докладний огляд п'яти основних типів динамічної обробки звуку: компресія, обмеження, автоматичне регулювання посилення (AGC), розширення і пропускання.

Акустичні події і виявлення акустичних подій

Розділення звуків на блоки або сегменти, які сприймаються як окремі і різні, іноді називають «аналізом акустичних подій» або «аналізом акустичних сцен» (ASA), і сегменти іноді називаються «акустичними подіями» або «звуковими подіями». Докладне обговорення аналізу акустичних сцен

13. Зчитуваний комп'ютером носій, на якому збережена комп'ютерна програма для вказівки комп'ютеру виконувати спосіб за будь-яким з пп. 1-11.

викладене Альбертом С. Брегманом в його книзі Auditory Scene Analysis-The Perceptual Organization of Sound (Аналіз акустичних сцен - сприймана організація звуку), Інститут технології Массачусетса, 1991, четвертий друк, 2001 друге видання в паперовій палітурці MIT Press. Додатково, патент США № 6002776, виданий Bhadkamkar (Бхадкамкар) та ін. 14 грудня 1999, посилається на публікації, що належать до 1976, як на «роботу попереднього рівня техніки, що стосується виділення звуку за допомогою аналізу акустичних сцен». Проте, патент Бхадкамкар та ін. не схвалює практичне використання аналізу акустичних сцен, роблячи висновок що «технології, які включають аналіз акустичних сцен, хоча і цікаві з наукової точки зору як моделі обробки людського слуху, в наш час вимагають дуже багато обчислень і можуть розглядатися як практичні технології для виділення звуку до досягнення фундаментального прогресу».

Корисний спосіб ідентифікації акустичних подій викладений Crockett і Crockett та ін. (Крокетт і Крокет та ін.) в різних патентних заявках і статтях, перелічених нижче під заголовком «Включення за допомогою посилання». Згідно з цими документами звуковий сигнал поділяється на акустичні події, кожна з яких прагне до того, що вона сприймається як окрема і відмінна від інших подія, за допомогою виявлення змін спектрального складу (амплітуди як функції частоти) з часом. Це можна зробити, наприклад, обчислюючи спектральний склад послідовних часових блоків звукового сигналу та ідентифікуючи межу акустичної події як межу між послідовними часовими блоками, якщо різниця в спектральному складі між такими послідовними часовими блоками перевищує порогову величину. Альтернативно, замість змін спектрального складу з часом або додатково до них можуть бути обчислені зміни амплітуди з часом.

При реалізації, яка вимагає мінімальних обчислень, процес поділяє звуковий сигнал на часові сегменти, аналізуючи всю смугу частот (повну смугу пропускання звукового сигналу) або, по суті, цілу смугу частот (в конкретних реалізаціях часто використовується фільтр обмеження смуги частот на кінцях спектра) і додаючи найбільше значення найгучнішим компонентам звукового сигналу. Цей підхід використовує перевагу явища, яке належить до психології слухового сприйняття, при якому при меншій шкалі часу (20 мілісекунд (мсек.) і менше) вухо може мати схильність до фокусування на одній акустичній події за заданий час. Мається на увазі те, що хоча одночасно може відбуватися декілька подій, одна компонента прагне до більшого виділення при сприйнятті і може бути оброблена індивідуально, як якби вона була єдиною подією, яка мала місце. Використання переваги цього

ефекту також допускає виявлення акустичної події, порівнянне зі складністю обробленого звукового сигналу. Наприклад, якщо обробленим вхідним звуковим сигналом є звук сольного інструмента, ідентифікованими акустичними подіями ймовірноше усього будуть окремі зіграні ноти. Подібним чином, для вхідного мовного сигналу, швидше за все, будуть ідентифіковані окремі компоненти мови, голосні звуки і приголосні звуки як окремі звукові елементи. Із збільшенням складності звукового сигналу, наприклад музики з барабанним боєм або декількома інструментами і голосом, виявлення акустичної події ідентифікує «той, що найбільше виділяється» (тобто найгучніший) звуковий елемент в будь-який заданий момент.

За рахунок більшої обчислювальної складності процес може також врахувати зміни спектрального складу з часом в дискретних частотних піддіапазонах (постійних або динамічно визначених піддіапазонах або постійних і динамічно визначених піддіапазонах), а не на повній смузі пропускання. Цей альтернативний підхід враховує більше одного звукового потоку в різних частотних піддіапазонах, а не припускає, що тільки один потік може бути сприйнятий в конкретний момент часу.

Виявлення акустичної події може бути реалізоване розділенням звукового сигналу часової області на часові інтервали або блоки і потім перетворенням даних з кожного блока на частотну область, використовуючи або гребінку фільтрів, або частотно-часове перетворення, наприклад, FFT. Амплітуда спектральних складових кожного блока може бути нормована для виключення або зменшення ефекту амплітудних змін. Кожне результуюче представлення частотної області забезпечує індикацію спектрального складу звуку в конкретному блоці. Спектральний склад послідовних блоків порівнюється, і зміни, які перевищують порогове значення, можуть бути взяті для індикації часового початку або часового кінця акустичної події.

Переважно, дані частотної області нормуються, як описано нижче. Ступінь, до якого дані частотної області повинні бути нормовані, дає індикацію амплітуди. Отже, якщо зміна в цій мірі перевищує заздалегідь задане порогове значення, це також може бути прийняте за індикацію меж події. Початкові і кінцеві точки події, які є результатом спектральних змін і амплітудних змін, можуть бути логічно складені разом, щоб межі події, одержані в результаті змін обох типів, були ідентифіковані.

Хоча технології, описані в згаданих публікаціях і статтях Крокетт і Крокет та ін., є зокрема корисними разом з аспектами даного винаходу, інші технології для ідентифікації акустичних подій і меж події можуть бути використані в аспектах даного винаходу.

Розкриття винаходу

Традиційна динамічна обробка звукового сигналу попереднього рівня техніки задіює множення звукового сигналу на керуючий сигнал, що залежить від часу, який настраює посилення звукового сигналу, одержуючи необхідний результат. «Посиленням» є масштабний коефіцієнт, який змінює масштаб амплітуди звукового сигналу. Цей керую-

чий сигнал може генеруватися на постійній основі або з блоків звукових даних, але звичайно його виробляє деяка форма вимірювання обробленого звукового сигналу, і швидкість його зміни визначають згладжуючі фільтри, іноді з постійними характеристиками, а іноді з характеристиками, які змінюються з динамікою звукового сигналу. Наприклад, час відгуку можна настроїти відповідно до змін магнітуди або потужності звукового сигналу. Способи попереднього рівня техніки, такі як автоматичне регулювання посилення (AGC) і компресія динамічного діапазону (DRC), не визначають яким-небудь способом, що належить до психології слухового сприйняття, часові інтервали протягом яких зміни посилення можуть бути сприйняті як погіршення і коли вони можуть бути застосовані без передачі чутних артефактів. Отже, традиційні звукові динамічні процеси можуть часто вносити чутні артефакти, тобто ефекти динамічної обробки можуть внести небажані сприймані зміни в звуковий сигнал.

Аналіз акустичних сцен ідентифікує дискретні сприймані акустичні події, причому кожна подія відбувається між двома послідовними межами акустичних подій. Акустичні погіршення, викликані зміною посилення, можуть бути значно зменшені забезпеченням того, що в межах акустичної події посилення залишається майже постійним, і обмеженням більшої частини зміни по сусідству з межею події. У контексті компресорів (засобів стиснення) і розширювачів (засобів розширення) відгук на збільшення рівня звукового сигналу (яке часто називається атакою) може бути швидким, порівнянним з мінімальною тривалістю акустичних подій або коротше, але відгук на зменшення (відпускання або повернення) може бути повільнішим, так що звуки, які повинні здаватися постійними або повинні поступово ослаблятися, можуть бути явно порушені. За таких обставин дуже вигідно затримати повернення посилення до наступної межі або сповільнити швидкість зміни посилення під час події. Для додатків автоматичного регулювання посилення, коли рівень гучності звуку з тривалістю від середньої до тривалої нормується і обидва часи атаки і повернення можуть, отже, бути тривалими в порівнянні з мінімальною тривалістю акустичної події, вигідно під час події затримати зміни або сповільнити швидкості зміни посилення до межі наступної події для збільшуючого посилення і для зменшуючого посилення.

Згідно з одним аспектом даного винаходу система обробки звукового сигналу приймає звуковий сигнал і аналізує і змінює характеристики посилення і/або динамічного діапазону звукового сигналу. Зміною динамічного діапазону звукового сигналу часто керують параметри динамічної системи обробки (час атаки і повернення, ступінь компресії тощо), які значно впливають на сприймані артефакти, які вносяться динамічною обробкою. Зміни характеристик сигналу з часом в звуковому сигналі виявляються та ідентифікуються як межі акустичної події, так що звуковий сегмент між послідовними межами становить акустичну подію в звуковому сигналі. Характеристики акустичних подій, що представляють інтерес, можуть включа-

ти в себе такі характеристики подій, як сила і тривалість сприйняття. Деякі із згаданих одного або більше параметрів динамічної обробки генеруються, щонайменше частково, у відповідь на акустичні події і/або ступінь зміни характеристик сигналу, пов'язаних із згаданими межами акустичної події.

Звичайно, акустичною подією є сегмент звукового сигналу, який має схильність до того, що він сприймається як окремий і відмінний від інших сегментів. Одне використовуване вимірювання характеристик сигналу включає в себе вимірювання спектрального складу звукового сигналу, наприклад, як описано в згаданих документах Крокетт і Крокет та ін. Всі або деякі з одного або більше параметрів динамічної обробки звукового сигналу можуть генеруватися щонайменше частково, у відповідь на наявність або відсутність і характеристики однієї або більше акустичних подій. Межа акустичної події може бути ідентифікована як зміна характеристик сигналу з часом, який перевищує порогове значення. Альтернативно, всі або деякі з одного або більше параметрів можуть генеруватися, щонайменше частково, у відповідь продовжуваного вимірювання ступеня зміни характеристик сигналу, пов'язаних із згаданими межами акустичної події. Хоча, в принципі, аспекти винаходу можуть бути реалізовані в аналогових і/або цифрових областях, практичні реалізації ймовірніше усього повинні бути реалізовані в цифровій області, в якій кожний із звукових сигналів представлений окремими вибірками або вибірками в межах блоків даних. У цьому випадку характеристиками сигналу може бути спектральний склад звукового сигналу в межах блока, виявленням змін в характеристиках сигналу з часом може бути виявлення змін спектрального складу звукового сигналу від блока до блока, і початкова і кінцева межі акустичної події кожна співпадає з межами блока даних. Потрібно зазначити, що для більш традиційного випадку виконання динамічних змін посилення на основі вибірок описаний аналіз акустичних сцен може бути виконаний на основі блока і результуючої інформації про акустичну подію, що використовується для виконання динамічних змін посилення, які були застосовані до вибірок.

Регулюючи ключові параметри динамічної обробки звукового сигналу з використанням результатів аналізу акустичних сцен, можна одержати раптові зменшення чутних артефактів, що вносяться динамічною обробкою.

Даний винахід представляє два способи виконання аналізу акустичних сцен. Перший спосіб виконує спектральний аналіз та ідентифікує положення сприйманих звукових подій, які використовуються для регулювання параметрів динамічного посилення, ідентифікуючи зміни в спектральному складі. Другий спосіб перетворює звуковий сигнал в область сприйманої гучності (яка може забезпечити більш релевантну інформацію, що належить до психології слухового сприйняття, ніж перший спосіб) та ідентифікує положення акустичних подій, які потім використовуються для регулювання параметрів динамічного посилення. Потрібно зазначити, що другий спосіб вимагає, щоб обробка звукового сигналу знала про рівні абсолютного

відтворення звуку, які можуть бути неможливі при деяких реалізаціях. Представлення обох способів аналізу акустичних сцен допускає реалізації зміни динамічного посилення, регульованого на основі ASA, з використанням процесів або пристроїв, які можуть або не можуть бути відкалібровані для врахування рівнів абсолютного відтворення.

Аспекти даного винаходу описані тут в умовах динамічної обробки звукового сигналу, яка включає аспекти інших винаходів. Такі інші винаходи описані в різних патентних заявках США і міжнародних патентних заявках Dolby Laboratories Licensing Corporation, що розглядаються, власником даної заявки, причому ці заявки встановлені тут.

Перелік фігур креслень

Фіг. 1 є логічною блок-схемою, що показує приклад етапів обробки для виконання аналізу акустичних сцен.

Фіг. 2 показує приклад обробки блока, застосування вікна і виконання DFT відносно звукового сигналу під час виконання аналізу акустичних сцен.

Фіг. 3 є свого роду логічною блок-схемою або функціональною блок-схемою, що показує паралельну обробку, при якій звуковий сигнал використовують для ідентифікації акустичних подій і для ідентифікації характеристик акустичних подій з умови, щоб події та їх характеристики були використані для зміни параметрів динамічної обробки.

Фіг. 4 є свого роду логічною блок-схемою або функціональною блок-схемою, яка показує обробку, при якій звуковий сигнал використовують тільки для ідентифікації акустичних подій, а характеристики події визначають з виявлення акустичної події з умови, щоб події та їх характеристики були використані для зміни параметрів динамічної обробки.

Фіг. 5 є свого роду логічною блок-схемою або функціональною блок-схемою, яка показує обробку, при якій звуковий сигнал використовують тільки для ідентифікації акустичних подій, а характеристики події визначають з виявлення акустичної події з умови, щоб тільки характеристики акустичних подій були використані для зміни параметрів динамічної обробки.

Фіг. 6 показує набір ідеалізованих характеристик характеристик звукового фільтра, який апроксимує критичну смуговість шкали ERB. По горизонтальній шкалі відмічена частота в герцах, а по вертикальній шкалі - рівень в децибелах.

Фіг. 7 показує контури однакової гучності ISO 226. По горизонтальній шкалі відмічена частота в герцах (шкала десятикових логарифмів), а по вертикальній шкалі відмічений рівень звукового тиску в децибелах.

Фіг. 8a-c показують ідеалізовані вхідні/вихідні характеристики і вхідні характеристики посилення компресора динамічного діапазону звуку.

Фіг. 9a-f показують приклад використання акустичних подій для регулювання часу повернення при цифровій реалізації традиційного контролера динамічного діапазону (DRC), при якій регулювання посилення витягують з середньоквадратичного (RMS) значення потужності сигналу.

Фіг. 10a-f показують приклад використання акустичних подій для регулювання часу повернення при цифровій реалізації традиційного контролера динамічного діапазону (DRC), при якій регулювання посилення витягують з середньоквадратичного (RMS) значення потужності сигналу для сигналу, альтернативного використовуваному на Фіг. 9 сигналу.

Фіг. 11 зображує відповідний набір ідеалізованих кривих AGC та DRC для застосування AGC, за якою іде DRC, в системі динамічної обробки області гучності. Задачею комбінації є одержання приблизно однакової сприйманої гучності для всіх оброблених звукових сигналів, зберігаючи при цьому щонайменше деякі з вихідних динамічних характеристик звукового сигналу.

Кращий режим для виконання винаходу

Аналіз акустичних сцен (первинний спосіб, при якому відсутня область гучності)

Згідно з втіленням одного аспекту даного винаходу аналіз акустичних сцен може складатися з чотирьох загальних етапів обробки, як показано на Фіг. 1. На першому етапі 1-1 («Виконати спектральний аналіз») беруть звуковий сигнал часової області, ділять його на блоки та обчислюють спектральний профіль або спектральний склад для кожного з блоків. Спектральним аналізом перетворюють звуковий сигнал на короткоперіодну частотну область. Це можна виконати, використовуючи будь-яку гребінку фільтрів, основу або на перетвореннях, або на блоках смугових фільтрів, і або в лінійному, або викривленому частотному просторі (наприклад, шкалі Барк (Bark) або критичній смузі, які краще апроксимують характеристики людського вуха). За наявності будь-якої гребінки фільтрів має місце компроміс між часом і частотою. Більше розрізнення за часом і, отже, більш короткі інтервали часу приводять до меншого розрізнення за частотою. Більше розрізнення за частотою і, отже, більш вузькі піддіапазони приводять до більш тривалих інтервалів часу.

На першому етапі, проілюстрованому по суті на Фіг. 1, обчислюють спектральний склад послідовних сегментів часу звукового сигналу. У практичному втіленні розмір блока ASA може складатися з будь-якої кількості вибірок вхідного звукового сигналу, хоча 512 вибірок забезпечують хороший компроміс між розрізненнями за часом і частотою. На другому етапі 1-2 визначають різницю між спектральними складами від блока до блока («Виконати вимірювання різниці спектральних профілів»). Таким чином, на другому етапі обчислюють різницю спектральних складів між послідовними сегментами часу звукового сигналу. Як обговорювалося вище, зміна спектрального складу вважається ефективним індикатором початку або кінця сприйнятої акустичної події. На третьому етапі 1-3 («Ідентифікувати положення меж акустичної події»), коли різниця спектральних складів між одним блоком спектрального профілю і наступним перевищує порогову величину, за межу акустичної події приймають межу блока. Звуковий сегмент між послідовними межами становить акустичну подію. Таким чином, на третьому етапі встановлюють межі акустичної події між послідовними

сегментами часу, коли різниця спектральних профілів між такими послідовними сегментами часу перевищує порогову величину, визначаючи, таким чином, акустичні події. У цьому втіленні межі акустичної події визначають акустичні події, що мають довжину, яка є цілим кратним довжині блоків спектрального профілю при мінімальній довжині одного блока спектрального профілю (512 вибірок у даному прикладі). У принципі, межі події необов'язково повинні бути так обмежені. Як альтернатива обговореним тут практичним втіленням, розмір вхідного блока може змінюватися, наприклад, щоб бути рівним, по суті, розміру акустичної події.

Після ідентифікації меж події визначають ключові характеристики акустичної події, як показано на етапі 1-4.

Сегменти звукового сигналу, що перекриваються або не перекриваються, можуть бути реалізовані за допомогою застосування вікна і використані для обчислення спектральних профілів вхідного звукового сигналу. Перекриття приводить до кращого розрізнення відносно положення акустичних подій, і зменшує імовірність пропуску події, такої як короткий одиночний імпульс. Проте, перекриття також збільшує обчислювальну складність. Таким чином, перекриття можна виключити. Фіг. 2 показує принципове представлення N блоків вибірок, що не перекриваються, реалізованих за допомогою застосування вікна і перетворених на частотну область за допомогою дискретного перетворення Фур'є (DFT). Кожний блок може бути реалізований за допомогою застосування вікна і перетворений на частотну область, наприклад за допомогою DFT, переважно реалізованим як швидке перетворення Фур'є (FFT) для швидкості.

Наступні змінні можуть бути використані для обчислення спектрального профілю вхідного блока:

M=кількість вибірок, охоплених вікном, в блоці, використовуваному для обчислення спектрального профілю,

R=кількість вибірок, що перекриваються, при спектральному обчисленні.

Загалом, для вищезазначених змінних можуть бути використані будь-які цілі числа. Проте, реалізація буде більш ефективною, якщо встановити M рівним степені числа 2, щоб можна було використовувати стандартні перетворення FFT для обчислень спектрального профілю. У практичному втіленні процесу аналізу акустичних сцен перелічені параметри можна встановити рівними:

M=512 вибірок (або 11,6 мсек. при 44,1 кГц),

R=0 вибірок (без перекриття).

Вищеперелічені значення були визначені експериментально, і було виявлено, що вони з достатньою точністю ідентифікують положення і тривалість акустичних подій. Проте, було виявлено, що встановлення значення R, рівного 256 вибірок (50% перекриття), а не нулю вибірок (без перекриття), корисне при ідентифікації деяких подій, які складно виявити. Хоча для мінімізації спектральних артефактів внаслідок застосування вікон можуть бути використані багато різних типів вікон, використовуваним при обчисленнях спектрального профілю вікном є вікно типу M-точкове Hanning,

Kaiser-Bessel або інше відповідне, переважно, непрямокутне вікно. Вищезгадані значення і вікно типу Hanning були вибрані після тривалого аналізу експериментів, оскільки було показано, що вони забезпечують чудові результати для широкого діапазону звукових матеріалів. Застосування непрямокутних вікон переважно для обробки звукових сигналів з переважно низькочастотним вмістом. Застосування прямокутних вікон виробляє спектральні артефакти, які можуть стати причиною некоректного виявлення подій. На відміну від визначених кодуєчих/декодуєчих додатків (кодеків), де процес повного перекриття/підсумовування повинен забезпечувати постійний рівень, таке обмеження непридатне тут, і вікно може бути вибрано через його характеристики, такі як розрізнення за часом/частотою і заглушення за допомогою смуги затримання.

На етапі 1-1 (Фіг. 1) спектр кожного блока М-вибірки може бути обчислений одержанням даних за допомогою застосування вікна, такого як М-точкове Hanning, Kaiser-Bessel або іншого придатного вікна, перетворенням на частотну область, використовуючи М-точкове швидке перетворення Фур'є, і обчисленням величини комплексних коефіцієнтів FFT. Результуючі дані нормують, приймаючи найбільшу величину за одиницю, і нормований масив з М чисел перетворюють на логарифмічну область. Дані можуть бути також нормовані на деякі інші показники, такі як значення середньої величини або значення середньої потужності даних. Перетворення масиву на логарифмічну область необов'язкове, але воно спрощує обчислення значення різниці на етапі 1-2. Крім того, логарифмічна область ближче відповідає природі людської зорової системи. Значення результуючої логарифмічної області мають діапазон від мінус нескінченності до нуля. У практичному втіленні на діапазон значень може накладатися нижня межа; обмеження може бути постійним, наприклад, -60 дБ, або може залежати від частоти для відображення більш низької чутності тихих звуків при низьких і дуже високих частотах (зазначимо, що можна було б зменшити розмір масиву до $M/2$ з тієї причини, що FFT представляє негативні, так само як і позитивні частоти).

На етапі 1-2 обчислюють міру різниці між спектрами сусідніх блоків. Для кожного блока кожний з М (логарифмічних) спектральних коефіцієнтів, обчислених на етапі 1-1, віднімають з відповідного коефіцієнта для попереднього блока, і обчислюють значення різниці (не враховуючи знак). Ці М різниці потім складають в одне число. Цю міру значення різниці можна також виразити як середню різницю, що припадає на спектральний коефіцієнт, діленням міри різниці на кількість спектральних коефіцієнтів, використовуваних в додаванні (в даному випадку - М коефіцієнтів).

На етапі 1-3 ідентифікують положення меж акустичних подій застосуванням порогової величини до масиву мір різниці, обчислених на етапі 1-2. Коли міра різниці перевищує порогову величину, зміну спектра вважають достатньою для сигналізації про нову подію, і номер блока із зміною реструктурують як межу події. Для значень М та Р, нада-

них вище, і для значень логарифмічної області (одержаних на етапі 1-1), виражених в одиницях дБ, порогова величина може бути встановлена рівною 2500, якщо порівнюється вся величина FFT (включаючи негативну частину), або 1250, якщо порівнюється половина FFT (як зазначено вище, FFT представляє негативні, так само як і позитивні частоти, для величини FFT одна частота є дзеркальним відображенням іншої). Це значення може бути вибрано експериментально, і воно забезпечує хороше виявлення меж акустичної події. Це значення параметра може бути змінено для зменшення (збільшивши порогову величину) або збільшення (зменшивши порогову величину) виявлення подій.

Процес на Фіг. 1 може бути представлений в більш звичайному вигляді еквівалентними виконаннями, представленими на Фіг. 3, 4 та 5. На Фіг. 3 до звукового сигналу паралельно застосовують функцію «Ідентифікувати акустичні події» або етап 3-1, який ділить звуковий сигнал на акустичні події, кожна з яких прагне до того, щоб вона сприймалася як окрема і відмінна від інших подія, і необов'язкову функцію «Ідентифікувати характеристики акустичних подій» або етап 3-2. Процес за Фіг. 1 може бути використаний для розділення звукового сигналу на акустичні події та ідентифікації їх характеристик, або може бути використаний деякий інший процес. Інформацію про акустичну подію, якою може бути ідентифікація меж акустичної події, що визначається функцією або етапом 3-1, потім використовують для зміни параметрів динамічної обробки звукового сигналу (таких як атака, повернення, ступінь тощо), як потрібно, за допомогою функції «Змінити динамічні параметри» або етапу 3-3. Необов'язкова функція «Ідентифікувати характеристики» або етап 3-3 також приймає інформацію про акустичну подію. Функція «Ідентифікувати характеристики» або етап 3-3 може характеризувати деякі або всі акустичні події однієї або більше характеристик. Такі характеристики можуть включати ідентифікацію переважного піддіапазону акустичної події, як описано в зв'язку з процесом на Фіг. 1. Характеристики можуть також включати одну або більше звукових характеристик, включаючи, наприклад, величину потужності акустичної події, величину амплітуди акустичної події, величину спектральної нерівномірності акустичної події та інформацію про те, чи є акустична подія, по суті, безшумною, або інші характеристики, які допомагають змінити динамічні параметри, щоб зменшити або видалити чутні артефакти обробки. Характеристики можуть також включати інші характеристики, наприклад, інформацію про те, чи включає акустична подія короткий одиночний імпульс.

Альтернативи виконання за Фіг. 3 показані на Фіг. 4 та 5. На Фіг. 4 до вхідного звукового сигналу не застосовують безпосередньо функцію «Ідентифікувати характеристики» або етап 4-3, але він приймає інформацію від функції «Ідентифікувати акустичні події» або етапу 4-1. Виконання за Фіг. 1 є конкретним прикладом такого виконання. На Фіг. 5 функції або етапи 5-1, 5-2 та 5-3 виконані послідовно.

Деталі цього конкретного втілення не є критичними. Можуть бути використані інші способи обчислення спектрального складу послідовних сегментів часу звукового сигналу, обчислення різниці між послідовними сегментами часу і встановлення меж акустичної події на відповідних межах між послідовними сегментами часу, коли різниця спектральних профілів між такими послідовними сегментами часу перевищує порогову величину.

Аналіз акустичних сцен (новий спосіб з областю гучності)

Міжнародна заявка згідно з Договором про патентну кооперацію з номером PCT/US2005/038579, подана 25 жовтня 2005, опублікована як Міжнародна публікація під номером WO 2006/047600 A1, озаглавлена «Обчислення і настройка сприйманої гучності і/або сприйманий спектральний баланс звукового сигналу», Алан Джеффері Сифелдт, розкриває крім всього іншого об'єктивну міру сприйманої гучності, основану на моделі, що належить до психології слухового сприйняття. Згадана заявка за допомогою посилання включена в дану заявку у всій повноті. Як описано в згаданій заявці, із звукового сигналу $x[n]$ обчислюють сигнал $E[b,t]$ збудження, який апроксимує розподіл енергії вздовж базиллярної мембрани внутрішнього вуха в критичній смузі b протягом часового блока t . Це збудження може бути обчислене з короткочасного дискретного перетворення Фур'є (STDFT) звукового сигналу таким чином:

$$E[b,t] = \lambda_b E[b,t-1] + (1 - \lambda_b) \sum_k |T[k]|^2 |C_b[k]|^2 |X[k,t]|^2 \quad (1)$$

де $X[k,t]$ представляє STDFT сигналу $x[n]$ у часовому блоці t та елементі дискретизації k . Зазначимо, що в рівнянні 1 t представляє час в дискретних одиницях блоків перетворення на відміну від безперервної міри, такої як секунди. $T[k]$ представляє частотну характеристику фільтра, що імітує передачу звукового сигналу через зовнішнє і середнє вухо, і $C_b[k]$ представляє частотну характеристику базиллярної мембрани в положенні, що відповідає критичній смузі b . Фіг. 6 зображує відповідний набір характеристик фільтра критичної смуги, в якому 40 смуг рівномірно рознесені вздовж шкали еквівалентних прямокутних смуг частот (ERB), визначеної Моор і Glasberg (Муром і Глазбергом). Форма кожного фільтра описується закругленою експоненціальною функцією, і смуги розподілені з використанням відстані 1 ERB. Нарешті, згладжуюча часова константа λ_b , в рівнянні 1 може бути переважно вибрана пропорційною часу інтегрування людського сприйняття гучності в межах смуги b .

Використовуючи контури однакової гучності, такі як контури, зображені на Фіг. 7, збудження кожної смуги перетворюється в рівень збудження, який генерував би таку саму сприйману гучність при 1 кГц. Конкретну гучність, міру сприйманої гучності, розподіленої за частотою і часом, потім обчислюють з перетвореного збудження $E_{1kHz}[b,t]$ за допомогою компресійної нелінійності. Одна така придатна функція для обчислення конкретної гучності $N[b,t]$ задається таким чином:

$$N[b,t] = \beta \left(\left(\frac{E_{1kHz}[b,t]}{TQ_{1kHz}} \right)^\alpha - 1 \right) \quad (2)$$

де T_{1kHz} є пороговою величиною в тиші при 1 кГц, а константи β та α вибирають для відповідності зростанню даних про гучність, накопичуваних при експериментах з прослуховування. В абстрактному значенні це перетворення із збудження в конкретну гучність може бути представлене функцією $\Psi\{\}$, такою як:

$$N[b,t] = \Psi\{E[b,t]\}.$$

В результаті загальну гучність $L[t]$, представлену в сонах, обчислюють підсумовуванням конкретної гучності по смугах:

$$L[t] = \sum_b N[b,t] \quad (3)$$

Конкретна гучність $N[b,t]$ є спектральним представленням, призначеним для імітації способу сприйняття людиною звукового сигналу як функції частоти і часу. Вона вловлює зміни чутливості до різних частот, зміни чутливості до рівня і зміни розрізнення за частотою. З цієї причини вона є спектральним представленням, яке добре відповідає виявленню акустичних подій. Порівняння різниці $N[b,t]$ по смугах між послідовними блоками часу може в багатьох випадках привести до більш точно сприйманого виявлення акустичних подій в порівнянні з описаним вище безпосереднім використанням послідовних спектрів FFT, хоча воно більш складне для обчислення.

У згаданій патентній заявці розкрито декілька варіантів застосування для зміни звукового сигналу на основі цієї моделі гучності, що належить до психології слухового сприйняття (психоакустичної моделі). Серед цих застосувань розкриті декілька алгоритмів динамічної обробки, наприклад, AGC та DRC. Ці розкриті алгоритми можуть використовувати переваги використання акустичних подій для регулювання різних пов'язаних з ними параметрів. Оскільки конкретна гучність вже обчислена, вона легкодоступна для цілей виявлення згаданих подій. Подробиці переважного втілення обговорюються нижче.

Регулювання параметра динамічної обробки звукового сигналу за наявності акустичних подій

Далі представлені два приклади втілень винаходу. Перший приклад описує використання акустичних подій для регулювання часу повернення при цифровій реалізації регулювання динамічного діапазону (DRC), при якій регулювання посилення витягують з середньоквадратичного (RMS) значення потужності сигналу. Друге втілення описує використання акустичних подій для керування визначеними аспектами більш складної комбінації AGC та DRC, реалізованої в контексті моделі гучності, що належить до психології слухового сприйняття, описаної вище. Ці два втілення служать тільки прикладами винаходу, і потрібно розуміти,

що використання акустичних подій для параметрів регулювання алгоритму динамічної обробки не обмежене описаними нижче конкретними прикладами.

Регулювання динамічного діапазону

Описана цифрова реалізація DRC сегментує звуковий сигнал $x[n]$ на реалізовані за допомогою застосування вікна блоки, що наполовину перекриваються, і для кожного блока обчислюється посилення модифікування на основі міри локальної потужності сигналу і вибраної кривої компресії. Посилення згладжують по блоках і потім множать з кожним блоком. До змінених блоків в результаті додають перекриття для генерації зміненого звукового сигналу $y[n]$.

Потрібно зазначити, що, хоча аналіз акустичних сцен і цифрова реалізація DRC, описувані тут, поділяють звуковий сигнал часової області на блоки для виконання аналізу та обробки, обробку DRC необов'язково виконувати з використанням сегментації на блоки. Наприклад, аналіз акустичних сцен може бути виконаний з використанням сегментації на блоки, а спектральний аналіз, описаний вище, і результуючі положення і характеристики акустичних подій можуть бути використані для забезпечення регулюючої інформації для цифрової реалізації традиційної реалізації DRC, яка звичайно працює на основі послідовних вибірок. Тут, проте, для DRC застосовується така сама структура розділення на блоки, використовується для аналізу акустичних сцен, щоб спростити опис їх комбінації.

Пристаючи до опису, основаної на блоках реалізації DRC, блоки, що перекриваються, звукового сигналу можуть бути представлені у вигляді:

$$x[n, t] = w[n]x[n + tM/2], \text{ для } 0 < n < M-1 \quad (4)$$

де M є довжиною блока, розмір стрибка дорівнює $M/2$, $w[n]$ є вікном, n є індексом вибірки в межах блока, а t є індексом блока (зазначимо, що тут t використовується таким саме чином, як і для STDFТ в рівнянні 1; тобто представляє час в дискретних одиницях блоків, а не в секундах, наприклад). В ідеальному варіанті, вікно $w[n]$ звужується до нуля на обох кінцях і дорівнює в сумі одиниці, коли наполовину перекриває саме себе; звичайно використовуване синусоїдальне вікно, наприклад, відповідає цим критеріям.

Зазначимо, що внаслідок того, що блоки помножені на звужуване вікно, як показано в рівнянні 4, синтез з підсумовування з перекриттям, показаний вище, ефективно згладжує посилення по вибірках обробленого сигналу $y[n]$. Таким чином, керуючий сигнал посилення одержує згладжування додатково до показаного в рівнянні 7а згладжування. При більш традиційній реалізації DRC, який працює по послідовних вибірках, а не по послідовних блоках, може бути потрібне більш складне

Для кожного блока можна потім обчислити RMS потужності для генерації значення $P[t]$ потужності в дБ, що припадає на кожний блок:

$$P[t] = 10 * \log_{10} \left(\frac{1}{M} \sum_{n=1}^M x^2[n, t] \right) \quad (5)$$

Як згадувалося раніше, можна згладити це значення потужності швидкою атакою і повільним поверненням до обробки за допомогою кривої компресії, але як альтернатива обробляється миттєва потужність $P[t]$ і згладжується результуюче посилення. Цей альтернативний підхід має перевагу в тому, що може бути використана проста крива компресії з гострими точками загину, але результуючі посилення залишаються як і раніше гладкими при проходженні потужності через точки загину. Представляючи криву компресії так, як показано на Фіг. 8с у вигляді функції F рівня сигналу, який генерує посилення, посилення $G[t]$ блока задається таким чином:

$$G[t] = F\{p[t]\} \quad (6)$$

Припускаючи, що крива компресії накладає більше ослаблення із збільшенням рівня сигналу, посилення буде зменшуватися, коли сигнал знаходиться в «режимі атаки», і збільшуватися, коли сигнал знаходиться в «режимі повернення». Отже, згладжене посилення $G[t]$ може бути обчислене відповідно до:

$$\bar{G}[t] = \alpha[t] \cdot \bar{G}[t-1] + (1 - \alpha[t])G[t] \quad (7a)$$

де

$$\alpha[t] = \begin{cases} \alpha_{\text{attach}} & G[t] < \bar{G}[t-1] \\ \alpha_{\text{release}} & G[t] \geq \bar{G}[t-1] \end{cases} \quad (7b)$$

і

$$\alpha_{\text{release}} \gg \alpha_{\text{attach}} \quad (7c)$$

Нарешті, згладжене посилення $\bar{G}[t]$, виражене в дБ, застосовують до кожного блока сигналу, і змінені блоки підсумовуються з перекриттям для одержання зміненого звукового сигналу:

$$y[n + tM/2] = \left(10^{\bar{G}[t]/20} \right) x[n, t] + \left(10^{\bar{G}[t-1]/20} \right) x[n + M/2, t-1] \text{ для } 0 < n < M/2 \quad (8)$$

згладжування посилення, ніж простий однополюсний фільтр, показаний в рівнянні 7а, для запобігання чутному спотворенню обробленого сигналу. Також, використання основаної на блоках обробки вносить в систему характерну затримку в $M/2$ вибірок, і доти, доки час спаду, пов'язаний з α_{attach} , близький до цієї затримки, сигнал $x[n]$ не потрібно додатково затримувати до застосування посилення з метою запобігання викиду.

Фіг. 9а-9с зображують результат застосування описаної обробки DRC до звукового сигналу. Для цієї конкретної реалізації використовується довжина блока $M=512$ при частоті дискретизації, яка дорівнює 44,1 кГц. Використовується крива компресії, аналогічна показаній на Фіг. 8b:

вище -20 дБ відносно повномасштабної цифри сигнал ослабляється із співвідношенням 5:1, а нижче

-30 дБ сигнал посилюється із співвідношенням 5:1.

Посилення згладжується з коефіцієнтом a_{attack} атаки, відповідним часу напівспаду в 500 мсек. Вихідний звуковий сигнал, зображений на Фіг. 9а, складається з шести послідовних акордів для фортепіано, причому кінцевий акорд, розташований навколо вибірки $1,75 \times 10^5$, затухає до тиші. Вивчаючи графік посилення $G[t]$ на Фіг. 9b, потрібно зазначити, що посилення залишається близьким до 0 дБ під час програвання шести акордів. Це результат збереження більшої частини енергії в діапазоні від -30 дБ до -20 дБ, області, в межах якої крива DRC не вимагає змін. Проте, після досягнення останнього акорду енергія сигналу падає нижче -30 дБ, і посилення починає рости, в кінцевому результаті до значення понад 15 дБ, по мірі загасання акордів. Фіг. 9с зображує результуючий змінений звуковий сигнал, і можна бачити, що шлейф кінцевого акорду значно посилений. На слух, це посилення природного низькорівневого затухаючого звуку акорду дає надзвичайно неприродний результат. Метою даного винаходу є запобігання проблемам такого типу, пов'язаним з традиційною динамічною обробкою.

Фіг. 10а-10с зображують результати застосування точно такої самої системи DRC до іншого звукового сигналу. У цьому випадку перша половина сигналу складається з музичного відрізка швидкого темпу при високому рівні, а потім при приблизній вибірці 10×10^4 сигнал перемикається на другий музичний відрізок швидкого темпу, але при значно більш низькому рівні. Вивчаючи посилення за Фіг. 10b, можна бачити, що сигнал ослабляється приблизно на 10 дБ протягом першої половини, а потім посилення зворотно зростає аж до 0 дБ протягом другої половини, коли програвється більш м'який відрізок. У цьому випадку посилення працює так, як потрібно. Бажано посилити другий відрізок відносно першого, і посилення повинно швидко зрости після переходу на другий відрізок, щоб бути малопомітним на слух. Видно, що поведінка посилення аналогічно обговореній поведінці у випадку першого сигналу, але в цьому випадку ця поведінка є бажаною. Отже, бажано виправити перший випадок, не торкаючись другого. Використання акустичних подій для регулювання часу повернення цієї системи DRC забезпечує таке рішення.

У першому сигналі, який був досліджений на Фіг. 9, посилення загасання останнього акорду здається неприродним, оскільки акорд та його загасання сприймаються як одна акустична подія, збереження цілісності якої чекають. У другому випадку, проте, під час збільшення посилення відбувається багато акустичних подій, означаючи, що

кожній окремій події надається маленька зміна. Отже, загальна зміна посилення не є такою небажаною. Отже, можна міркувати про те, що зміну посилення потрібно допускати тільки в найближчому часовому околі межі акустичної події. Можна застосувати цей принцип до посилення, коли воно знаходиться або в режимі атаки, або в режимі повернення, але для більшості практичних реалізацій DRC посилення так швидко переходить в режим атаки в порівнянні з розрізненням за часом сприйняття людиною події, що регулювання не потрібне. Отже, можна використовувати події для регулювання згладжування посилення DRC тільки коли воно знаходиться в режимі повернення.

Далі буде описана відповідна поведінка регулювання повернення. В якісних показниках, якщо виявлена подія, посилення згладжують залежно від константи часу повернення, визначеної вище в рівнянні 7а. Протягом часу після виявленої події, і якщо не виявлені наступні події, константа часу повернення безперервно збільшується, так що в кінцевому результаті згладжене посилення «заморожується» на місці. Якщо виявлена інша подія, згладжуючу константу часу повторно встановлюють на вихідне значення, і процес повторюється. Для модуляції часу повернення можна спочатку згенерувати керуючий сигнал на основі меж виявленої події.

Як обговорювалося раніше, межі події можуть бути виявлені пошуком змін в послідовних спектрах звукового сигналу. При цій конкретній реалізації DFT кожного блока, що перекривається, $x[n, t]$ може бути обчислене для генерації STDFT звукового сигналу $x[n]$:

$$X[x, t] = \sum_{n=0}^{M-1} x[n, t] e^{-j \frac{2\pi kn}{M}} \quad (9)$$

Далі, різниця між нормованими логарифмічними величинами спектрів послідовних блоків може бути обчислена відповідно до:

$$D[t] = \sum_k X_{\text{NORM}}[k, t] - X_{\text{NORM}}[k, t-1] \quad (10a)$$

де

$$X_{\text{NORM}}[k, t] = \log \left(\frac{|X[k, t]|}{\max_k |X[k, t]|} \right) \quad (10b)$$

Тут максимальне значення $|X[k, t]|$ по елементах дискретизації k використовується для нормування, хоча можна використовувати інші нормувальні множники; наприклад, середнє значення $|X[k, t]|$ по елементах дискретизації. Якщо різниця $D[t]$ перевищує порогову величину D_{min} , вважають, що подія відбулася. Додатково, даній події можна надати силу, яка лежить в інтервалі від нуля до одиниці, на основі розміру $D[t]$ в порівнянні з максимальною пороговою величиною D_{max} . Результуючий сигнал $A[t]$ сили акустичної події може бути обчислений як:

$$A[t] = \begin{cases} 0 & D[t] \leq D_{\min} \\ \frac{D[t] - D_{\min}}{D_{\max} - D_{\min}} & D_{\min} < D[t] < D_{\max} \\ 1 & D[t] \geq D_{\max} \end{cases} \quad (11)$$

Надаючи силу акустичній події, пропорційну величині зміни спектра, пов'язаного з цією подією, досягають кращого регулювання динамічної обробки в порівнянні з прийняттям рішення про бінарну подію. Автори винаходу виявили, що більш значні зміни посилення допустимі під час більш сильних подій, і сигнал в рівнянні 11 допускає таке змінне регулювання.

Сигнал $A[t]$ є імпульсним сигналом, причому імпульси відбуваються в положеннях межі подій. З метою регулювання часу повернення можна додатково згладжувати сигнал $A[t]$, так щоб він плавно спадав до нуля після виявлення межі події. Згладжений керуючий сигнал $\bar{A}[t]$ події може бути обчислений з $A[t]$ відповідно до:

$$\bar{A}[t] = \begin{cases} A[t] & A[t] > \alpha_{\text{event}} \bar{A}[t-1] \\ \alpha_{\text{event}} \bar{A}[t-1] & \text{в іншому випадку} \end{cases} \quad (12)$$

Тут α_{event} регулює час спадання керуючого сигналу події. Фіг. 9d та 10d зображують керуючий сигнал $\bar{A}[t]$ події для двох відповідних звукових сигналів, причому час напівспаду більш гладкої з них встановлений на 250 мсек. У першому випадку видно, що межа події виявлена для кожного з шести акордів для фортепіано і керуючий сигнал події плавно спадав до нуля після кожної події. Для другого сигналу виявлено багато подій, дуже близьких одна до одної за часом, і, отже, керуючий сигнал події ніколи не спадає повністю до нуля.

Тепер можна використовувати керуючий сигнал $\bar{A}[t]$ події для зміни константи часу повернення, використовуючої для згладжування посилення. Коли керуючий сигнал дорівнює одиниці, згладжуючий коефіцієнт $\alpha[t]$ з рівняння 7a дорівнює α_{release} , як і раніше, а коли керуючий сигнал дорівнює нулю, коефіцієнт дорівнює одиниці, так що запобігається зміна згладженого посилення. Згладжуючий коефіцієнт інтерполюють між цими двома межами, використовуючи керуючий сигнал, відповідно до:

$$\alpha[t] = \begin{cases} \alpha_{\text{attach}} & G[t] < \bar{G}[t-1] \\ \bar{A}[t] \alpha_{\text{release}} + (1 - \bar{A}[t]) & G[t] \geq \bar{G}[t-1] \end{cases} \quad (13)$$

При безперервній інтерполяції згладжуючого коефіцієнта як функції керуючого сигналу події час повернення повторно встановлюється на значення, пропорційне силі події на початку події, і потім плавно збільшується до нескінченності після випадку події. Швидкість цього збільшення диктує коефіцієнт α_{event} , використовуваний для генерації згладженого керуючого сигналу події.

Фіг. 9e та 10e показують результат згладжування посилення за допомогою регульованого подією коефіцієнта з рівняння 13, на протилежність не регульованому подією коефіцієнту з рівняння 7b. У першому випадку керуючий сигнал події падає до нуля після останнього акорду для фортепіано, запобігаючи тим самим руху посилення вгору. У результаті, відповідний змінений звуковий сигнал, зображений на Фіг. 9f, не страждає від неприродного посилення загасання акорду. У другому випадку керуючий сигнал події ніколи не досягає нуля, і, отже, згладжений сигнал посилення дуже мало затримується внаслідок застосування регулювання події. Траєкторія згладженого посилення майже ідентична траєкторії не регульованого подією посилення на Фіг. 10b. Це в точності відповідає необхідному ефекту.

Основа на гучності AGC та DRC

Як альтернатива традиційній технології динамічної обробки, при якій зміни сигналу є прямою функцією простих вимірювань сигналу, наприклад, амплітуди або RMS потужності, Міжнародна патентна заявка з номером PCT/US2005/038579 розкриває використання моделі гучності, що належить до психології слухового сприйняття, описаної раніше у вигляді оточення, в якому повинна виконуватися динамічна обробка. Наведено декілька переваг. По-перше, вимірювання і зміни визначені в тонах, що є більш точною мірою сприйняття гучності, ніж більш фундаментальні значення, такі як амплітуда або RMS потужності. По-друге, звуковий сигнал може бути змінений з умови, щоб сприйманий спектральний баланс вихідного звукового сигналу був збережений при зміні загальної гучності. Таким чином, зміни загальної гучності стають менш очевидними на слух в порівнянні з динамічною обробкою, яка використовує широкосмугове посилення, наприклад, для зміни звукового сигналу. І, нарешті, модель, що належить до психології слухового сприйняття, є, по суті, багатосмуговою, і, отже, система легко конфігурується для виконання багатосмугової динамічної обробки для зменшення добре відомих проблем міжспектральної підкачки, пов'язаних з багатосмуговою динамічною обробкою.

Хоча виконання динамічної обробки в даній області гучності вже має декілька переваг перед традиційною динамічною обробкою, технологія може бути додатково поліпшена за допомогою використання акустичних подій для регулювання різних параметрів. Розглянемо звуковий сегмент, що містить акорди для фортепіано, як зображено на Фіг. 10a, і відповідну DRC, показану на Фіг. 10b та с. Можна виконати подібну DRC в області гучності, і в цьому випадку, коли загасання гучності кінцевого акорду для фортепіано посилюється, посилення буде менш очевидним, оскільки спектральний баланс затухаючої ноти буде збережений при застосуванні посилення. Проте, кращим рішенням буде не посилювати загасання зовсім, і, отже, можна вигідно застосувати той самий принцип регулювання часу атаки і повернення при наявності акустичних подій в області гучності, як описувалося раніше, для традиційної DRC.

Система динамічної обробки з областю гучності, яка буде зараз описана, складається з AGC, за якою іде DRC. Задачею цієї комбінації є одержання приблизно однакової сприйманої гучності для всіх оброблених звукових сигналів, зберігаючи при цьому щонайменше деякі з вихідних динамічних характеристик звукового сигналу. Фіг. 11 зображує відповідний набір кривих AGC та DRC для цього додатку. Зазначимо, що вхідні і вихідні дані обох кривих представлені в сонах, оскільки обробка виконується в області гучності. Крива AGC прагне донести вихідний звуковий сигнал ближче до визначеного цільового рівня і, як згадувалося раніше, це відбувається при відносно повільних константах часу. Можна подумати, що AGC створює довгострокову гучність звукового сигналу, яка дорівнює цільовій, але протягом коротких періодів часу гучність може сильно коливатися навколо цієї цільової гучності. Отже, можна використовувати більш швидкодіючу DRC для обмеження цих коливань в деякому діапазоні, який вважається прийнятним для конкретного додатку. Фіг. 11 показує таку криву DRC, де мета AGC попадає в межі «нульової смуги» DRC, області кривої, яка не вимагає змін. При такій комбінації кривих AGC встановлює довгострокову гучність звукового сигналу в межах нульової смуги кривої DRC, так що необхідне застосування мінімальних змін швидкодіючої DRC. Якщо короткострокова гучність все ще коливається поза нульовою смугою, DRC працює таким чином, щоб змстити гучність звукового сигналу до цієї нульової смуги. Як останнє загальне зауваження, можна застосувати повільнодіючу AGC з умови, щоб на кожну смугу моделі гучності припала однакова кількість змін гучності, за допомогою чого зберігається сприйманий спектральний баланс, і можна застосувати швидкодіючу DRC таким чином, щоб допустити зміну модифікації гучності по смугах для зменшення міжспектральної підкачки, яка в іншому випадку може з'явитися внаслідок швидкодіючої модифікації гучності, що не залежить від смуг.

Акустичні події можуть бути використані для регулювання атаки і повернення обох регулювань AGC та DRC. У випадку AGC обидва часи атаки і повернення є великими в порівнянні з розрізненням за часом сприйняття події, і, отже, регулювання події можна вигідно застосовувати в обох випадках. При DRC атака є відносно короткою, і, отже, регулювання події може потребуватися тільки для повернення, як при DRC, описаної вище.

Як обговорювалося раніше, можна використовувати спектр конкретної гучності, пов'язаний із застосовуваною моделлю гучності, з метою виявлення події. Різницевий сигнал $D[t]$, аналогічний сигналу в рівняннях 10a та b, можна обчислити з

$$\alpha_{AGC}[t] = \begin{cases} \bar{A}[t]\alpha_{AGCattack} + (1 - \bar{A}[t]) & L[t] > L_{AGC}[t-1] \\ \bar{A}[t]\alpha_{AGCrelease} + (1 - \bar{A}[t]) & L[t] \leq L_{AGC}[t-1] \end{cases} \quad (16b)$$

Додатково, можна обчислити відповідний спектр довгострокової конкретної гучності, який пізніше буде використаний для багатосмугової DRC:

конкретної гучності $N[b,t]$, визначеної в рівнянні 2, таким чином:

$$D[t] = \sum_b |N_{NORM}[b,t] - N_{NORM}[b,t-1]| \quad (14a)$$

де

$$N_{NORM}[b,t] = \frac{N[b,t]}{\max_b \{X[b,t]\}} \quad (14b)$$

Тут максимальне значення $|N[b,t]|$ по смугах b частот використовується для нормування, хоча можна використовувати інші нормувальні множники; наприклад, середнє значення $|N[b,t]|$ по смугах частот. Якщо різниця $D[t]$ перевищує порогову величину D_{min} , вважають, що подія відбулася. Різницевий сигнал може бути потім оброблений способом, аналогічним показаному в рівняннях 11 та 12, для генерації згладженого керуючого сигналу $\bar{A}[t]$ події, що використовується для регулювання часу атаки і повернення.

Крива AGC, зображена на Фіг. 11, може бути представлена у вигляді функції, яка як своє вхідне значення бере значення гучності і генерує необхідну вихідну гучність:

$$L_o = F_{AGC}\{L_i\} \quad (15a)$$

Крива DRC може бути представлена аналогічним чином:

$$L_o = F_{DRC}\{L_i\} \quad (15b)$$

Для AGC вхідна гучність є мірою довгострокової гучності звукового сигналу. Можна обчислити таку міру згладжуванням миттєвої гучності $L[t]$, визначеної в рівнянні 3, використовуючи відносно довгі константи часу (порядку декількох секунд). Показано, що при оцінці довгострокової гучності звукового сегмента людям важче аналізувати більш гучні області, ніж більш тихі, і можна використовувати більш швидку атаку замість повернення при згладжуванні, щоб імітувати цей ефект. При включенні регулювання події в атаку і повернення довгострокову гучність, використовувану для визначення модифікації AGC, можна, отже, обчислити відповідно до:

$$L_{AGC}[t] = \alpha_{AGC}[t]L_{AGC}[t-1] + (1 - \alpha_{AGC}[t])L[t] \quad (16a)$$

де

$$N_{AGC}[b,t] = \alpha_{AGC}[t]N_{AGC}[b,t-1] + (1 - \alpha_{AGC}[t])N[b,t] \quad (16c)$$

На практиці можна вибрати згладжуючі коефіцієнти з умови, щоб час атаки дорівнював приблизно половині часу повернення. Задаючи значення довгострокової гучності, можна потім обчислити масштабування модифікації гучності, пов'язане з AGC, у вигляді відношення вихідної гучності до вхідної гучності:

$$S_{AGC}[t] = \frac{F_{AGC} \{L_{AGC}[t]\}}{L_{AGC}[t]} \quad (17)$$

Тепер можна обчислити модифікацію DRC з гучності після застосування зміни масштабу AGC. Замість згладжування значення гучності до застосування кривої DRC можна, як альтернативу, застосувати криву DRC до миттєвої гучності і потім з часом згладити результуючу модифікацію. Це аналогічно технології, описаній раніше для згладжування посилення традиційної DRC. Додатково,

$$S_{DRC}[b, t] = \frac{N_{AGC}[b, t]}{S_{AGC}[t] L_{AGC}[t]} F_{DRC} \left\{ \frac{S_{AGC}[t] L_{AGC}[t]}{N_{AGC}[b, t]} N[b, t] \right\} \quad (18)$$

Модифікації AGC та DRC можна потім об'єднати для одержання загального масштабування, що припадає на кожну смугу:

$$S_{TOT}[b, t] = S_{AGC}[t] S_{DRC}[t] \quad (19)$$

Це загальне масштабування потім може бути згладжене за часом незалежно для кожної смуги швидкою атакою і повільним поверненням і регулюванням події, застосованим тільки до повер-

DRC можна застосувати в багатосмуговому режимі, що означає, що модифікація DRC є функцією конкретної гучності $N[b, t]$ в кожній смузі b , а не загальної гучності $L[t]$. Проте, для збереження середнього спектрального балансу вихідного звукового сигналу можна застосувати DRC до кожної смуги з умови, щоб результуючі модифікації мали такий самий середній ефект, який був би в результаті застосування DRC до загальної гучності. Цього можна досягнути, масштабуючи кожну смугу відношенням довгострокової загальної гучності (після застосування масштабування AGC) до довгострокової конкретної гучності і використовуючи це значення як аргумент функції DRC. Результат потім повторно масштабують за допомогою зворотної величини згаданого відношення для одержання вихідної конкретної гучності. Таким чином, масштабування DRC в кожній смузі може бути обчислене відповідно до:

нення. В ідеальному варіанті згладжування виконують за логарифмом масштабування аналогічно з посиленнями традиційної DRC при згладжуванні їх представлення в децибел, хоча це не суттєво. Щоб гарантувати, що згладжене загальне масштабування рухається синхронно з конкретною гучністю в кожній смузі, режими атаки і повернення можуть бути визначені одночасним згладжуванням найбільш конкретної гучності:

$$\bar{S}_{TOT}[b, t] = \exp(\alpha_{TOT}[b, t] \log(\bar{S}_{TOT}[b, t-1]) + (1 - \alpha_{TOT}[b, t]) \log(\bar{S}_{TOT}[b, t])) \quad (20a)$$

$$\bar{N}[b, t] = \alpha_{TOT}[b, t] \bar{N}[b, t-1] + (1 - \alpha_{TOT}[b, t]) N[b, t] \quad (20b)$$

де

$$\alpha_{TOT}[b, t] = \begin{cases} \alpha_{TOTAttack} & N[b, t] > \bar{N}[t-1] \\ \bar{A}[t] \alpha_{TOTRelease} + (1 - \bar{A}[t]) & N[b, t] \leq \bar{N}[t-1] \end{cases} \quad (20c)$$

В результаті можна обчислити цільову конкретну гучність, основуючись на згладженій зміні масштабу, який був застосований до вихідної конкретної гучності

$$\hat{N}[b, t] = \bar{S}_{TOT}[b, t] N[b, t] \quad (21)$$

і потім знайти посилення $G[b, t]$, які при застосуванні до вихідного збудження приводять до конкретної гучності, яка дорівнює цільовій:

$$\hat{N}[b, t] = \Psi \{G^2[b, t] E[b, t]\} \quad (22)$$

Посилення можна застосувати до кожної смуги гребінки фільтрів, використовуюваної для обчислення збудження, і потім змінений звуковий сигнал може бути згенерований інвертуванням гребінки

фільтрів для одержання зміненого звукового сигналу часової області.

Регулювання додаткових параметрів

Хоча вищенаведене обговорення сфокусоване на регулюванні параметрів атаки і повернення AGC та DRC за допомогою аналізу акустичних сцен оброблюваного звукового сигналу, перевагою регулювання за допомогою результатів ASA можуть бути інші важливі параметри. Наприклад, керуючий сигнал $\bar{A}[t]$ події з рівняння 12 може бути використаний для зміни значення параметра відношення DRC, який використовується для динамічної настройки посилення звукового сигналу. Параметр відношення, аналогічний параметрам часу атаки і повернення, може додати значний внесок в чутні артефакти, що виробляються динамічними настройками посилення.

Реалізація

Винахід може бути реалізований за допомогою апаратних або програмних засобів або їх комбінації (наприклад, програмованої логічної матриці). Якщо інше не встановлене, алгоритми, включені як частина винаходу, по суті не зв'язані з яким-небудь конкретним комп'ютером або іншою апаратурою. Зокрема, різні машини загального призначення можуть бути використані з програмами, написаними відповідно до викладених тут ідей, або може бути більш зручним конструювання більш спеціалізованого приладу (наприклад, інтегральних схем) для виконання необхідних етапів способу. Таким чином, винахід може бути реалізований в одній або більше комп'ютерних програмах, що виконуються в одній або більше програмованих комп'ютерних системах, кожна з яких містить щонайменше один процесор, щонайменше одну систему для зберігання даних (включаючи енергозалежну та енергонезалежну пам'ять і/або елементи зберігання), щонайменше один пристрій вводу або порт і щонайменше один пристрій виводу або порт. Програмний код застосовують до вхідних даних для виконання функцій, описуваних тут, і генерації вихідної інформації. Вихідну інформацію застосовують до одного або більше вихідних пристроїв відомим способом.

Кожна така програма може бути реалізована на будь-якій необхідній мові програмування (включаючи машинну мову, мову асемблера або високорівневі процедурні, логічні або об'єктно-орієнтовані мови програмування) для взаємодії з комп'ютерною системою. У будь-якому випадку мова може бути трансльованою або інтерпретованою мовою.

Кожну таку комп'ютерну програму переважно зберігають або завантажують на запам'ятовуючий носій або пристрій (наприклад, твердотільну пам'ять або носій, або магнітний або оптичний носій), зчитуваний програмованим комп'ютером загального або спеціального призначення для конфігурування і роботи комп'ютера при зчитуванні комп'ютерною системою запам'ятовуючого носія або пристрою для виконання описаних тут процедур. Систему згідно з винаходом можна також розглядати як систему, що реалізовується у вигляді зчитуваного комп'ютером запам'ятовуючого носія, на якому записана комп'ютерна програма, причому сконфігуроване таким чином запам'ятовуюче середовище приписує комп'ютерній системі працювати конкретним і заданим чином для виконання описаних тут функцій.

Описана визначена кількість втілень винаходу. Проте, потрібно розуміти, що можуть бути зроблені різні модифікації без відхилення від суті та обсягу винаходу. Наприклад, деякі з описаних тут етапів можуть не залежати від порядку виконання і, таким чином, можуть бути виконані в порядку, відмінному від описаного.

Потрібно розуміти, що реалізація інших варіантів і модифікацій винаходу та його різних аспектів буде очевидна фахівцям в даній галузі техніки і що винахід не обмежений цими конкретними описаними втіленнями. Отже, передбачене охоплення даним винаходом будь-яких і всіх модифікацій, варіантів або еквівалентів, які підпадають під дані

суті та обсяг розкритих і заявлених тут основних переважних принципів.

Включення за допомогою посилання

Наступні патенти, патентні заявки і публікації включені в дану заявку за допомогою посилання у всій повноті.

Динамічна обробка звукового сигналу

Audio Engineer's Reference Book, під редакцією Michael Talbot-Smith, дьуре видання. Limiters and Compressors, Alan Tutton, 2-1492-165. Focal Press, Reed Educational and Professional Publishing, Ltd., 1999.

Виявлення та використання акустичних подій

Патентна заявка США 10/474,387, «High Quality Time-Scaling and Pitch-Scaling of Audio Signals», Brett Graham Crockett, опублікована 24 червня 2004 як US 2004/0122662 A1.

Патентна заявка США 10/478,398, «Method for Time Aligning Audio Signals Using Characterizations Based on Auditory Events», Brett G. Crockett та ін., опублікована 29 липня 2004 як US 2004/0148159 A1.

Патентна заявка США 10/478,538, «Segmenting Audio Signals Into Auditory Events», Brett G. Crockett, опублікована 26 серпня 2004 як US 2004/0165730 A1. Аспекти даного винаходу забезпечують спосіб виявлення акустичних подій додатково до способів, розкритих в згаданій заявці Crockett.

Патентна заявка США 10/478,397, «Comparing Audio Using Characterizations Based on Auditory Events», Brett G. Crockett та ін., опублікована 2 вересня 2004 як US 2004/0172240 A1.

Міжнародна заявка згідно з Договором про патентну кооперацію PCT/US 05/24630, подана 13 липня 2005, озаглавлена «Method for Combining Audio Signals Using Auditory Scene Analysis», Michael John Smithers, опублікована 9 березня 2006 як WO 2006/026161.

Міжнародна заявка згідно з Договором про патентну кооперацію PCT/US2004/016964, подана 27 травня 2004, озаглавлена «Method, Apparatus and Computer Program for Calculating and Adjusting the Perceived Loudness of an Audio Signal», Alan Jeffrey Seefeldt та ін., опублікована 23 грудня 2004 як WO 2004/111994 A2.

Міжнародна заявка згідно з Договором про патентну кооперацію PCT/US2005/038579, подана 25 жовтня 2005, озаглавлена «Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal», Alan Jeffrey Seefeldt, та опублікована як Міжнародна публікація під номером WO 2006/047600.

«A Method for Characterizing and Identifying Audio Based on Auditory Scene Analysis», Brett Crockett та Michael Smithers, Audio Engineering Society Convention Paper 6416, 118th З'їзд, Барселона, 28-31 травня 2005.

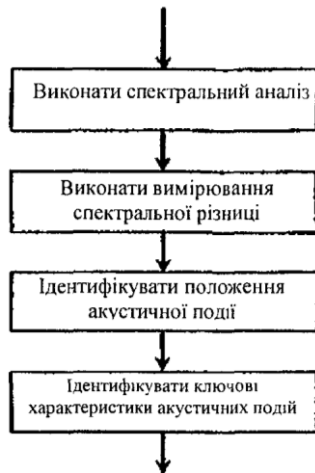
«High Quality Multichannel Time Scaling and Pitch-Shifting using Auditory Scene Analysis», Brett Crockett, Audio Engineering Society Convention Paper 5948, Нью-Йорк, жовтень 2003.

«A New Objective Measure of Perceived Loudness», Alan Seefeldt та ін., Audio Engineering

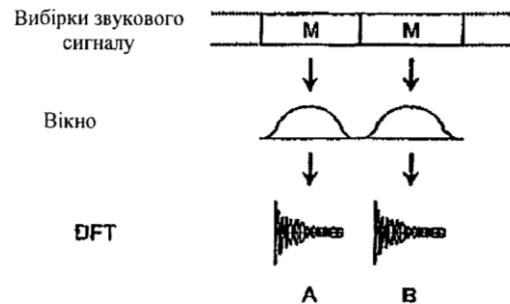
Society Convention Paper 6236, Сан-Франциско, 28 жовтня 2004.

Handbook for Sound Engineers, The New Audio Cyclopedia, під редакцією Glen M. Ballou, друге видання. Dynamics, 850-851. Focal Press an imprint of Butterworth-Heinemann, 1998.

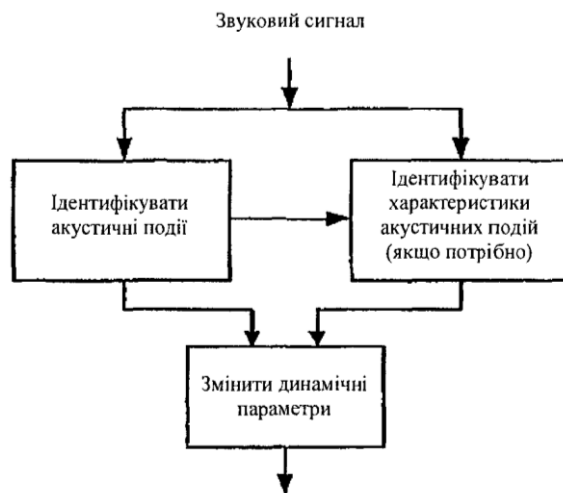
Audio Engineer's Reference Book, під редакцією Michael Talbot-Smith, друге видання, розділ 2.9 («Limiters and Compressors», Alan Tutton), сс. 2-1492-165. Focal Press, Reed Educational and Professional Publishing, Ltd., 1999.



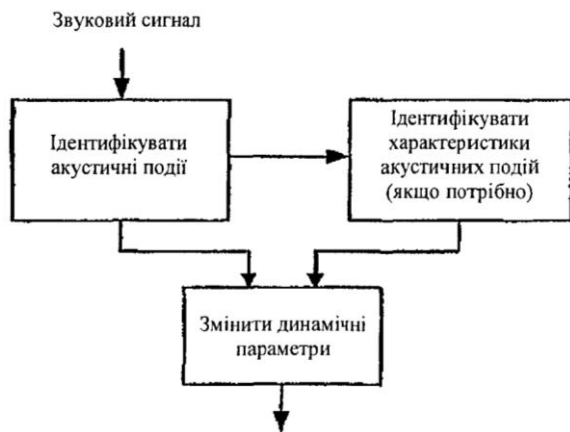
Фіг. 1



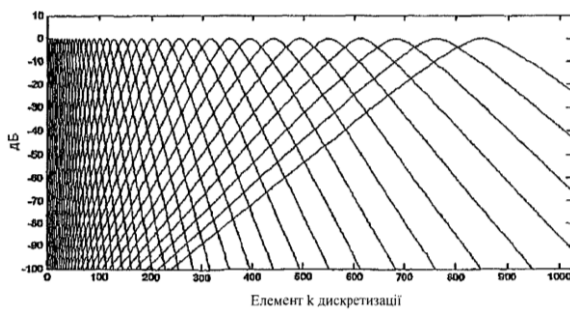
Фіг. 2



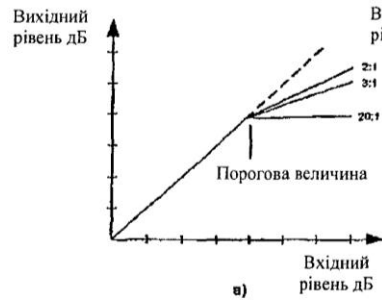
Фіг. 3



Фіг. 4



Фіг. 6

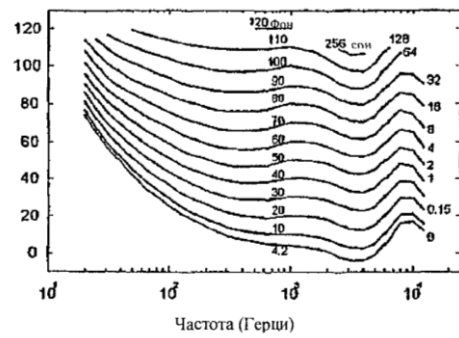


Звуковий сигнал



Фіг. 5

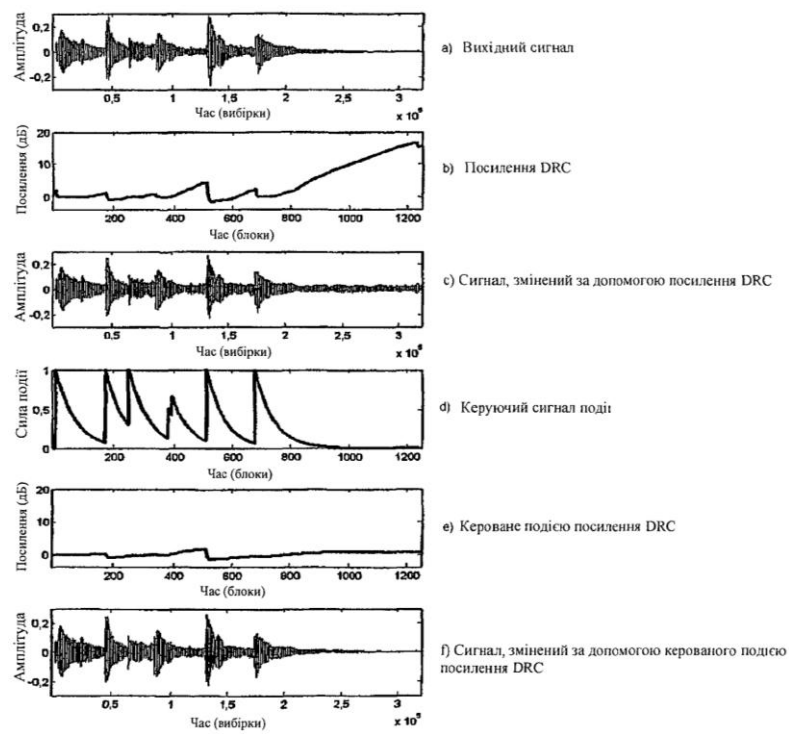
«Контури однакової гучності» ISO226 Annex A



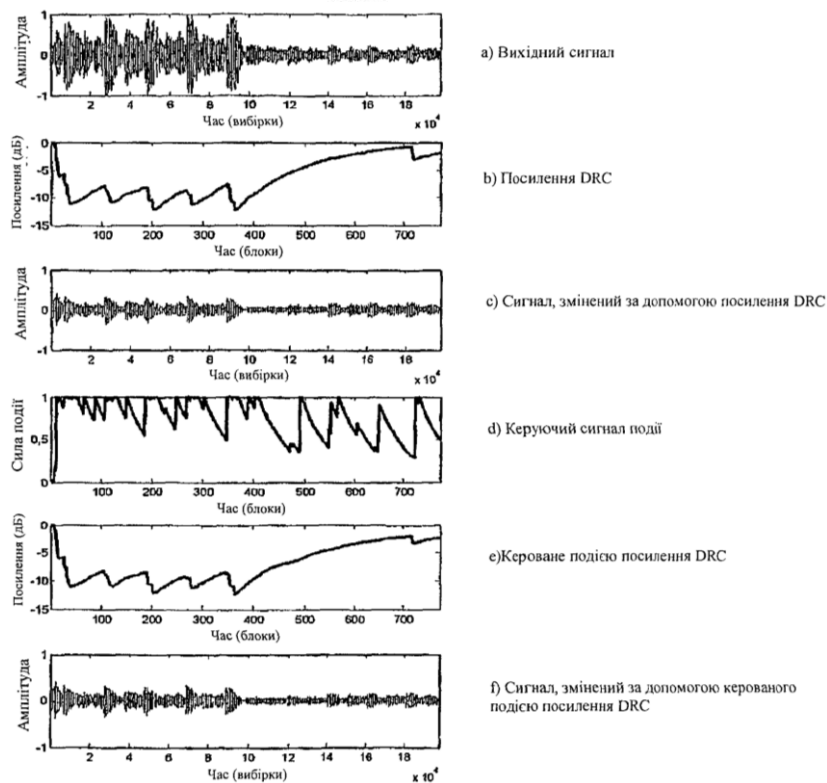
Фіг. 7



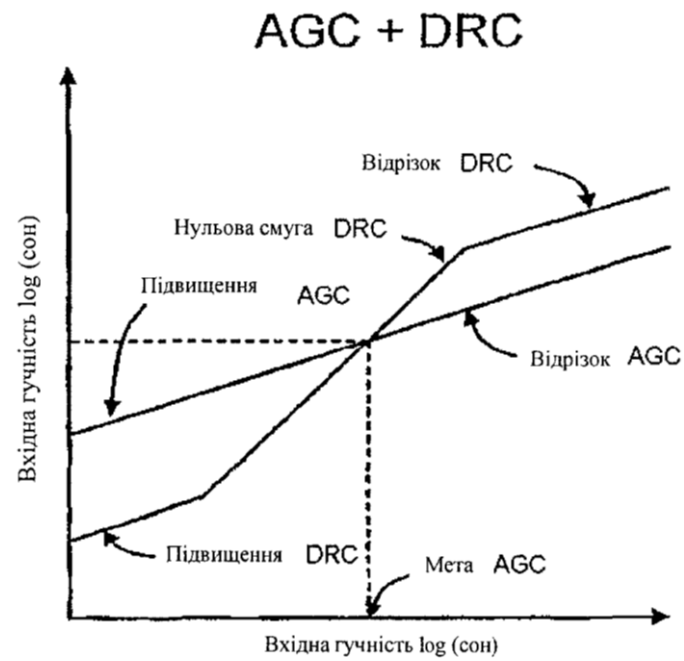
Фіг. 8



Фіг. 9



Фіг. 10



Фіг. 11