



ДЕРЖАВНА СЛУЖБА
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ
УКРАЇНИ

УКРАЇНА

(19) **UA** (11) **101974** (13) **C2**
(51) МПК (2013.01)
G10L 21/00

(12) ОПИС ДО ПАТЕНТУ НА ВИНАХІД

(21) Номер заявки:	а 2010 13673	(72) Винахідник(и):	Мюш Ханнес (US)
(22) Дата подання заявки:	17.04.2009	(73) Власник(и):	ДОЛБІ ЛЕБОРЕТЕРІЗ ЛАЙСЕНСІНГ КОРПОРЕЙШН, 100 Potrero Avenue, San Francisco, CA 94103-4813, United States of America (US)
(24) Дата, з якої є чинними права на винахід:	27.05.2013	(74) Представник:	Мошинська Ніна Миколаївна, реєстр. №115
(31) Номер попередньої заявки відповідно до Паризької конвенції:	61/046,271	(56) Перелік документів, взятих до уваги експертизою:	WO 03/022003 A2; 13.03.2003 EP 0637011 A1; 01.02.1995 EP 0517233 A1; 09.12.1992
(32) Дата подання попередньої заявки відповідно до Паризької конвенції:	18.04.2008		
(33) Код держави-учасниці Паризької конвенції, до якої подано попередню заявку:	US		
(41) Публікація відомостей про заявку:	10.02.2011, Бюл.№ 3		
(46) Публікація відомостей про видачу патенту:	27.05.2013, Бюл.№ 10		
(86) Номер та дата подання міжнародної заявки, поданої відповідно до Договору РСТ	PCT/US2009/040900, 17.04.2009		

(54) СПОСІБ І ПРИСТРІЙ ДЛЯ ПІДТРИМКИ СПРИЙМАНОСТІ МОВИ В БАГАТОКАНАЛЬНОМУ ЗВУКОВОМУ СУПРОВОДІ З МІНІМАЛЬНИМ ВПЛИВОМ НА СИСТЕМУ ОБ'ЄМНОГО ЗВУЧАННЯ

(57) Реферат:

У одному втіленні даний винахід включає в себе спосіб поліпшення сприйманості мови в багатоканальному звуковому сигналі. Цей спосіб включає в себе порівняння першої характеристики і другої характеристики багатоканального звукового сигналу для генерації коефіцієнта ослаблення. Ця перша характеристика відповідає першому каналу цього багатоканального звукового сигналу, який містить мовні і немовні звукові сигнали, а друга характеристика відповідає другому каналу цього багатоканального звукового сигналу, який переважно містить немовні звукові сигнали. Цей спосіб додатково включає в себе коректування цього послаблюючого коефіцієнта, згідно з оцінкою імовірності мови, для генерації скоректованого послаблюючого коефіцієнта. Цей спосіб додатково включає в себе ослаблення другого каналу з використанням цього скоректованого послаблюючого коефіцієнта.

UA 101974 C2

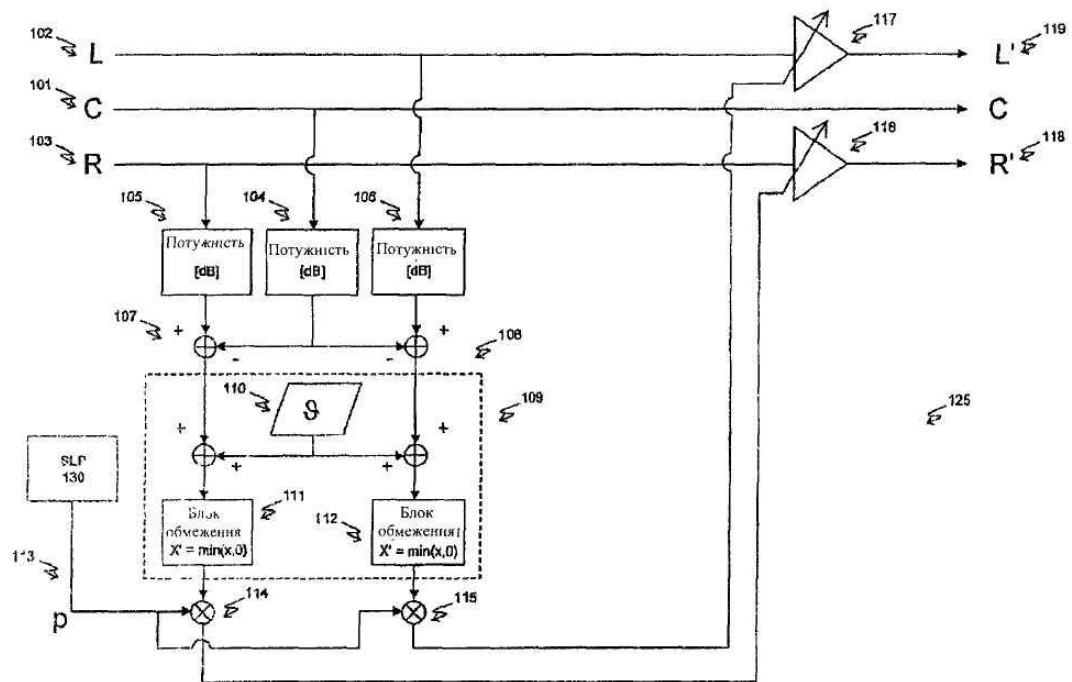


Fig. 1

Дана заявка на винахід заявляє пріоритет попередньої заявки на патент США № 61/046,271, поданої 18 квітня 2008 року, яка шляхом посилання включається в даний документ у всій своїй повноті.

Рівень техніки

Даний винахід, загалом, належить до обробки звукових сигналів, а більш конкретно, до поліпшення чіткості діалогу і усного мовлення, зокрема, в об'ємному розважальному звуковому супроводі.

Підходи, описані в даному розділі документа, не являють собою попередній рівень техніки по відношенню до формули винаходу в даній заявці і не можуть бути визнані як попередній рівень техніки через включення в даний розділ, якщо тільки не вказане зворотне.

Сучасний розважальний звуковий супровід з численними одночасними звуковими каналами (система об'ємного звуку) надає слухачам реалістичні звукові оточення з ефектом занурення, що мають колосальне розважальне значення. У таких оточеннях багато які звукові елементи, такі як діалог, музика і звукові ефекти, представлені одночасно і конкурують, відволікаючи увагу слухача. Для деяких членів аудиторії - особливо зі зниженими слуховими рецепторами або з уповільненим когнітивним сприйняттям - діалог і усне мовлення можуть бути важкі для розуміння протягом деяких частин програми, в яких представлені гучні конкуруючі звукові елементи. Протягом таких епізодів для цих слухачів було б корисно, якби рівень конкуруючих звуків знизився.

Усвідомлення того, що музика і ефекти можуть придушувати діалог, не нове, і було запропоновано декілька способів для виправлення цієї ситуації. Однак, як буде стисло викладено далі, ці запропоновані способи або несумісні з сучасною практикою широкомовних передач, накладають зайво високу плату на всю індустрію розваг, або і те, і інше.

У виробництві об'ємного звукового супроводу в кіно і на телебаченні загальноприйнятою практикою є розміщення більшої частини діалогу і усного мовлення тільки в один канал (центральный канал, його називають також мовним каналом). Звичайно, музика, звуки навколишнього середовища і звукові ефекти мікшуються, як в мовному, так і у всіх інших каналах (наприклад, в Лівому [L], Правому [R], Лівому об'ємному [ls] і в Правому об'ємному [rs] каналах, їх називають також немовними каналами). В результаті цього, мовний канал переносить велику частину мовного і значну кількість немовного звукового супроводу, що міститься в звуковій програмі, тоді як немовні канали переносять, переважно, немовний звуковий супровід, але також можуть переносити невелику кількість мови. Один простий підхід до полегшення сприйманості діалогу або усного мовлення в цих уживаних музичних сумішах полягає в постійному зниженні рівня гучності всіх немовних каналів, відносно рівня гучності мовного каналу, наприклад, на 6 dB. Цей підхід простий і ефективний, і він практикується в наші дні (наприклад, система відновлення звуку SRS [Sound Retrieval System] для чистоти діалогу (Dialog Clarity) або модифіковані рівняння понижувального мікшування в об'ємних декодерах). Однак він страждає щонайменше на один недолік: постійне ослаблення немовних каналів може до такої міри знизити рівень гучності спокійних звуків навколишнього середовища, які не заважають сприйняттю мови, що їх неможливо буде почути. При ослабленні не заважаючих звуків навколишнього середовища порушується естетичний баланс передачі без якої-небудь користі для розуміння мови слухачами.

Альтернативне рішення описане в серії патентів авторів Vaudrey і Saunders (U.S. Patent No. 7,266,501, U.S. Patent No. 6,772,127, U.S. Patent No. 6,912,501, і U.S. Patent No. 6,650,755). Наскільки зрозуміло, їх підхід має на увазі модифікацію змісту і розподілу продукції. Згідно з цією конфігурацією, споживач отримує два різних звукових сигнали. Перший з цих сигналів містить "Головний зміст" звукового супроводу. У багатьох випадках цей сигнал повністю поглинається мовою, але, за бажанням продюсера продукції, він може містити також і інші типи сигналів. Другий сигнал містить "Повторний зміст" звукового супроводу, який складений з всіх звукових елементів, що залишилися. Користувачеві надане керування відносними рівнями гучності цих двох сигналів або за допомогою ручного налаштування рівня гучності кожного з сигналів або за допомогою автоматичної підтримки відношення потужностей, вибраного користувачем. Хоч ця конфігурація допомагає обмежити зайве ослаблення не заважаючих звуків навколишнього середовища, її широкому поширенню заважає несумісність з усталеними способами виробництва і розподілу продукції.

Інший приклад способу керування відносними рівнями гучності мовного і немовного звукового супроводу був запропонований автором Bennett в U.S. Application Publication No. 20070027682.

Всі приклади на попередньому рівні техніки розділяють один спільний недолік: вони не надають ніяких технічних засобів мінімізації впливу, який надає підвищення чіткості діалогу на

систему звучання, що мається на увазі творцем програми, крім інших вад. Отже, об'єктом даного винаходу є надання технічних засобів для обмеження рівня гучності немовних каналів в традиційно мікшованій багатоканальній розважальній програмі таким чином, щоб мова залишалася зрозумілою, в той час як підтримувалася б також сприйманість немовних звукових компонентів.

Таким чином, є потреба в поліпшених методиках підтримки сприйманості мови. Даний винахід вирішує ці та інші проблеми за допомогою надання пристрою і способу поліпшення сприйманості мови в багатоканальному звуковому сигналі.

Суть винаходу

Втілення даного винаходу поліпшують сприйманість мови. У одному втіленні даний винахід включає в себе спосіб поліпшення сприйманості мови в багатоканальному звуковому сигналі. Цей спосіб включає в себе порівняння першої характеристики і другої характеристики багатоканального звукового сигналу для генерації коефіцієнта ослаблення. Ця перша характеристика відповідає першому каналу цього багатоканального звукового сигналу, який містить мовні і немовні звукові сигнали, а друга характеристика відповідає другому каналу цього багатоканального звукового сигналу, який, переважно, містить немовні звукові сигнали. Цей спосіб додатково включає в себе коректування цього послаблюючого коефіцієнта, згідно з оцінкою імовірності мови, для генерації скоректованого послаблюючого коефіцієнта. Цей спосіб додатково включає в себе ослаблення другого каналу з використанням цього скоректованого послаблюючого коефіцієнта.

Перший аспект цього винаходу оснований на спостереженні, що мовний канал типової розважальної програми протягом значної частини цієї програми переносить немовний сигнал. Тому, згідно з цим першим аспектом винаходу, маскування мовного звукового супроводу немовним звуковим супроводом може керуватися за допомогою: (а) визначення ослаблення сигналу в немовному каналі, необхідного для того, щоб межа відношення потужності сигналу в немовному каналі до потужності сигналу в мовному каналі не перевершувала заздалегідь визначеної порогової величини, і (b) градування цього ослаблення за допомогою коефіцієнта, який монотонно пов'язаний з оцінкою імовірності того, що сигнал в мовному каналі є мовою, і (c) застосування цього градуйованого ослаблення.

Другий аспект цього винаходу оснований на спостереженні, що відношення потужності мовного сигналу до потужності маскуючого сигналу є поганим показником для прогнозу сприйманості мови. Тому, згідно з цим другим аспектом винаходу, ослаблення сигналу в немовному каналі, яке необхідне для підтримки заздалегідь визначеного рівня сприйманості мови, обчислюється за допомогою прогнозування сприйманості мовного сигналу в присутності немовних сигналів за допомогою прогнозуючої моделі сприйманості мови, основаної на психоакустиці.

Третій аспект цього винаходу оснований на спостереженнях, що, якщо ослабленню дозволити змінюватися в залежності від частоти, то (а) заданий рівень сприйманості мови може бути досягнутий за допомогою багатьох схем ослаблення, і (b) різні схеми ослаблення можуть виробляти різні рівні інтенсивності або виразності немовного звукового супроводу. Тому, згідно з цим третім аспектом винаходу, маскування мовного звукового супроводу немовним звуковим супроводом керує за допомогою знаходження схеми ослаблення, яка максимізує інтенсивність або деякі інші показники виразності немовного звукового супроводу при обмеженні, що досягається заздалегідь визначеним рівнем прогнозованої сприйманості мови.

Втілення даного винаходу можуть бути здійснені як способи або технологічний процес. Ці способи можуть бути реалізовані як електронна схема, як обладнання або програмне забезпечення супроводу або як комбінація вищезазначеного. Електронна схема, що звичайно використовується для реалізації цього технологічного процесу, може являти собою спеціалізовану електронну схему (що виконує тільки специфічні завдання) або загальну електронну схему (запрограмовану для здійснення одного або декількох конкретних завдань).

Наступний докладний опис і супутні креслення забезпечують краще розуміння суті і переваг даного винаходу.

Короткий опис креслень

Фіг. 1 демонструє процесор сигналів, згідно з одним втіленням даного винаходу.

Фіг. 2 демонструє процесор сигналів, згідно з іншим втіленням даного винаходу.

Фіг. 3 демонструє процесор сигналів, згідно з іншим втіленням даного винаходу.

Фіг. 4А і Фіг. 4В являють собою структурні діаграми, які демонструють додаткові варіації втілень по кресленнях 1-3.

Докладний опис суті винаходу

Тут описані технічні прийоми для підтримки сприйманості мови. У подальшому описі, з метою пояснення, приведені численні приклади і конкретні технічні подробиці для надання повного розуміння даного винаходу. Однак фахівцям в даній галузі техніки буде зрозуміло, що даний винахід, як це визначено в формулі винаходу, може включати в себе деякі або всі ознаки тільки цих прикладів або в комбінації з іншими ознаками, описаними нижче, і може додатково включати в себе модифікації або еквіваленти ознак і концепцій, описаних в даному документі.

Різні способи і технологічні процеси описуються нижче. Те, що вони описуються в певному порядку, зроблено в основному для полегшення викладу. Потрібно розуміти, що конкретні етапи при бажанні можуть бути здійснені в іншому порядку або паралельно, в залежності від різних реалізацій. Якщо деякий конкретний етап повинен передувати або слідувати за іншим етапом, це буде точно вказано, якщо тільки це не зрозуміло з контексту.

Принцип першого втілення винаходу демонструє Фіг. 1. Посилаючись тепер на Фіг. 1, приймається багатоканальний сигнал, що складається з мовного каналу (101) і двох немовних каналів (102 і 103). Потужності сигналів в кожному з цих каналів вимірюються групою блоків оцінки потужності (104, 105 і 106) і виражаються в логарифмічній шкалі [dB]. Ці блоки оцінки потужності можуть мати механізм згладжування, такий як інтегратор витоків, з тим, щоб результат вимірювання рівня потужності відображав рівень потужності, усереднений по тривалості пропозиції або усього мовного епізоду. Цей рівень потужності в мовному каналі віднімається з рівня потужності в кожному з немовних каналів (за допомогою блоків підсумовування 107 і 108), щоб отримати показник різниці рівнів потужності між цими двома типами сигналів. Контур порівняння 109 визначає для кожного немовного каналу кількість dB, на яку цей немовний канал повинен бути ослаблений для того, щоб його рівень потужності залишався щонайменше на 9 dB нижче за рівень потужності сигналу в мовному каналі. (Символ "9" означає змінну і на нього також можна посилатися як на букву тета рукописного шрифту). Згідно з одним втіленням, однією з реалізацій цього є додавання цієї порогової величини (яка зберігається в електронному контурі 110) до різниці рівнів потужності (цей проміжний результат називають допуском) з обмеженням, щоб цей результат дорівнював або був меншим ніж нуль (за допомогою блоків обмеження 111 і 112). Цей результат є приростом (або інвертованим ослабленням) в dB, який повинен бути застосований до немовних каналів для того, щоб підтримувати рівень їх потужності на 9 dB нижче рівня потужності мовного каналу. Прийнятне значення величини 9 становить 15 dB. Це значення величини 9 при бажанні може бути скоректоване в інших втіленнях.

Оскільки має місце однозначна відповідність між показником, вираженим в логарифмічній шкалі (dB), і тим же самим показником, вираженим в лінійній шкалі, може бути виготовлений електронний контур, який еквівалентний Фіг. 1, в якому потужність, приріст і порогова величина виражаються в лінійній шкалі. У цій реалізації всі різниці рівнів замінюються відношеннями лінійних оцінок. У альтернативній реалізації можна замінити цей показник потужності показником, який пов'язаний з силою сигналу, таким як абсолютна величина сигналу.

Потрібно згадати, що однією з важливих ознак цього першого аспекту винаходу є градування отриманим таким чином приросту за допомогою оцінки, монотонно пов'язаної з імовірністю того, що сигнал в мовному каналі дійсно є мовою. Все ще посилаючись на Фіг. 1, приймається керуючий сигнал (113) і множиться за приростами (за допомогою блоків множення 114 і 115). Ці градуйовані прирости потім застосовуються до відповідних немовних каналів (за допомогою підсилювачів 116 і 117) для вироблення модифікованих сигналів L' і R' (118 і 119). Керуючий сигнал (113) звичайно є автоматично отриманим показником імовірності того, що сигнал в мовному каналі є мовою. Можуть використовуватися різні способи автоматичного визначення імовірності того, що сигнал є мовою. Згідно з одним втіленням, процесор імовірності мови генерує значення імовірності мови p (113) з інформації в C каналі 101. Один з прикладів такого механізму описується авторами Robinson і Vinton в "Automated Speech/Other Discrimination for Loudness Monitoring" (Audio Engineering Society, Preprint number 6437 °F Convention 118, May 2005). Як альтернатива, цей керуючий сигнал (113) може бути створений вручну, наприклад, творцем програми, і переданий разом із звуковим сигналом кінцевому користувачеві.

Фахівці в даній галузі техніки без великих зусиль зрозуміють, як ця конфігурація може бути поширена на будь-яку кількість вхідних каналів.

Фіг. 2 демонструє принцип другого аспекту винаходу. Посилаючись тепер на Фіг. 2, приймається багатоканальний сигнал, що складається з мовного каналу (101) і двох немовних каналів (102 і 103). Потужності сигналів в кожному з цих каналів вимірюються групою блоків оцінки потужності (201, 202 і 203). На відміну від відповідної групи блоків на Фіг. 1, ці блоки оцінки потужності вимірюють розподіл потужності сигналу відносно частоти, що в результаті дає

спектр потужності, а не єдине число. Цей спектральний дозвіл спектра потужності ідеально відповідає спектральному дозволу моделі прогнозування сприйманості мови (205 і 206, це поки ще не обговорювалося).

Ці два спектри потужності завантажуються в контур 204 порівняння. Цей блок призначений для визначення ослаблення, яке потрібно застосувати до кожного з немовних каналів для забезпечення того, щоб немовний канал не зменшив сприйманість мови сигналу в мовному каналі до величини, яка менша ніж заздалегідь визначений критерій. Цей функціональний засіб здійснюється за допомогою використання контурів прогнозування сприйманості мови (205 і 206), які прогнозують сприйманість мови на основі спектрів потужності мовного сигналу (201) і немовних сигналів (202 і 203). Контури 205 і 206 прогнозуючі сприйманості мови можуть реалізувати відповідну модель прогнозування сприйманості мови, в залежності від вибраної архітектури і вибору оптимальних співвідношень. Прикладом цього є індекс сприйманості мови (Speech Intelligibility Index), детально описаний в ANSI S3.5-1997 ("Methods for Calculation of the Speech Intelligibility Index"), і модель чутливості розпізнавання мови (Speech Recognition Sensitivity model) авторів Muesch і "Buus (Using statistical decision theory to predict speech intelligibility. I. Model structure" Journal of the Acoustical Society of America, 2001, Vol 109, p 2896-2909). Зрозуміло, що вихідні дані моделі прогнозування сприйманості мови не мають ніякого значення у випадку, коли сигнал в мовних каналах є чимось іншим, відмінним від мови. Незважаючи на це, в подальшому цей вихідний результат моделі прогнозування сприйманості мови буде називатися як прогнозна сприйманість мови. Відмічена помилка буде врахована в подальшій обробці за допомогою градування оцінок приросту на виході з контуру 204 порівняння з параметром, який пов'язаний з імовірністю того, сигнал є мовою (113, це поки ще не обговорювалося).

Спільна риса моделей прогнозування сприйманості мови полягає в тому, що вони дають прогноз або на поліпшення, або на незмінність сприйманості мови в результаті зниження рівня гучності немовного сигналу. Просуваючись по структурній схемі етапів технологічного процесу по Фіг. 2, контури 207 і 208 порівняння порівнюють прогнозу сприйманість мови з оцінкою критерію. Якщо оцінка рівня немовного сигналу низька, так що прогнозна сприйманість мови перевершує критерій, параметр приросту, який початково встановлений на 0 dB, витягується з контурів 209 або 210 і надається на контури 211 і 212 як вихідний результат контуру 204 порівняння. Якщо критерій не досягнутий, параметр приросту меншає на фіксовану величину і прогнозування сприйманості мови повторюється. Прийнятний розмір кроку для зменшення приросту дорівнює 1 dB. Описаний тут ітеративний процес продовжується доти, доки прогнозна сприйманість мови не досягне або перевершить величину критерію. Звичайно, можливо таке, що сигнал в мовному каналі такий, що критерій сприйманості мови не може бути досягнутий навіть при відсутності сигналу в немовному каналі. Прикладом такої ситуації служить мовний сигнал дуже низького рівня або з надзвичайно обмеженою смугою частот. Якщо таке трапилось, наступить момент, коли ніяке додаткове скорочення приросту, що застосовується до немовного каналу, не надає ефекту на прогнозу сприйманість мови, і критерій ніколи не може бути досягнутий. У таких умовах, петля, утворена з (205, 206), (207, 208) і (209, 210), продовжується нескінченно, і може бути застосований додатковий логічний блок для розриву цієї петлі. Одним з особливо простих прикладів такого логічного блоку може служити підрахунок числа ітерацій і вихід з петлі, як тільки буде перевершено заздалегідь визначена кількість ітерацій.

Просуваючись по структурній схемі етапів технологічного процесу по Фіг. 2, керуючий сигнал р (113) приймається і множиться на приріст (за допомогою блоків множення 114 і 115). Керуючий сигнал (113) звичайно буде являти собою автоматично зроблений показник імовірності того, що сигнал в мовному каналі є мовою. Способи автоматичного визначення імовірності того, що сигнал є мовою, відомі per se і обговорювалися в контексті Фіг. 1 (див. процесор 130 імовірності мови). Ці скоректовані прирости потім застосовуються до своїх відповідних немовних каналів (за допомогою блоків посилення 116 і 117) для вироблення модифікованих сигналів R' і L' (118 і 119).

Фіг. 3 демонструє принцип третього аспекту винаходу. З посиланням тепер на Фіг. 3, приймається багатоканальний сигнал, що складається з мовного каналу (101) і двох немовних каналів (102 і 103). Кожен з цих трьох немовних каналів розбивається на свої спектральні компоненти (за допомогою групи блоків 301, 302 і 303 фільтрації). Цей спектральний аналіз може бути отриманий за допомогою N-канальної групи блоків фільтрації у часовій області. Згідно з одним втіленням, це розбиття діапазону частот групою блоків фільтрації на смуги частот в 1/3 октави нагадує фільтрацію, яка, як передбачають, здійснюється всередині людського вуха. Той факт, що тепер сигнал складається з N підсигналів, продемонстрований за допомогою використання жирних ліній. Процес по Фіг. 3 може бути ідентифікований як

розгалужений процес (sidebranch process). Слідуючи по шляху сигналу, кожен з цих N підсигналів, які утворюють немовні канали, градується за допомогою одного з членів множини з N оцінок приростів (блоками посилення 116 і 117). Виробництво цих оцінок приростів буде описане пізніше. Далі, ці градуйовані підсигнали возз'єднуються в єдиний звуковий канал, це може бути зроблено через просте підсумовування (за допомогою контурів 313 і 314 підсумовуючі). Як альтернатива, може бути використана група фільтруючих блоків синтезу, яка сполучена з групою фільтруючих блоків аналізу. Результатом цього процесу є модифіковані сигнали R' і L' (118 і 119).

Описуючи тепер шлях розгалуженого процесу по Фіг. 3, кожне з вихідних даних групи фільтруючих блоків віддається в розпорядження відповідної групи з N блоків оцінки (304, 305 і 306) потужності. Отримані в результаті цього спектри служать як вхідні дані для контурів (307 і 308) оптимізації, які видають як вихідні дані N -вимірний вектор приростів. Ця оптимізація використовує як контур (309 і 310) прогнозу сприйманості мови, так і контур (311 і 312) обчислення інтенсивності звуку для знаходження вектора приростів, який максимізує інтенсивність звуку в немовному каналі, при цьому підтримує заздалегідь визначену оцінку прогнозованої сприйманості мови мовного сигналу. Відповідні моделі для прогнозування сприйманості мови обговорювалися в зв'язку з Фіг. 2. Контур 311 і 312 обчислення інтенсивності звуку можуть реалізувати відповідну модель прогнозування інтенсивності звуку, в залежності від вибраної архітектури і вибору оптимальних співвідношень. Прикладами відповідних моделей є американський національний стандарт (American National Standard) "ANSI S3.4-2007 Procedure for the Computation of Loudness of Steady Sounds" і німецький стандарт (German standard) "DIN 45631 Berechnung des Lautstarkepegels und der Lautheit aus dem Gerauschspektrum".

У залежності від обчислювальних ресурсів і накладених обмежень, вид і складність цих контурів (307, 308) оптимізації можуть надзвичайно сильно відрізнятися. Згідно з одним втіленням використовується ітераційна багатомірна оптимізація з обмеженнями N вільних параметрів. Кожен параметр представляє приріст, що застосовується до кожної з смуг частот в немовному каналі. Для знаходження максимуму можуть бути застосовані стандартні технічні засоби, такі як рух по шляху найбільшого градієнта в N -вимірному просторі. У іншому втіленні, обчислювально менш вимогливий підхід обмежує функціональні засоби прирости-частота, як ті, що лежать в малій множині можливих функціональних засобів прирости-частота, такої як множина різних спектральних градієнтів або shelf-фільтрів (super-hard extremely-low frequency). З такими додатковими обмеженнями задача оптимізації може бути зведена до малої кількості одномірних оптимізацій. Ще в одному втіленні здійснюється вичерпний пошук в дуже малій множині можливих функцій приросту. Цей останній підхід може виявитися особливо бажаним в додатках в реальному часі, в яких потребується постійне завантаження і швидкість пошуку.

Фахівці в даній галузі техніки легко розпізнають додаткові обмеження, які можуть бути накладені на оптимізацію, відповідно до додаткових втілень даного винаходу. Одним з прикладів є обмеження, щоб інтенсивність звуку модифікованого немовного каналу була не більшою, ніж інтенсивність звуку до модифікації. Інший приклад являє собою обмеження на різниці приростів між прилягаючими смугами частот для того, щоб обмежити можливості для часового спотворення реконструюючою групою фільтруючих блоків (313, 314) або скоротити можливості для небажаних модифікацій тембру. Бажані обмеження залежать як від технічної реалізації групи блоків фільтрації, так і від вибору оптимальних співвідношень між поліпшенням сприйманості мови і модифікацією тембру. Для зрозумілості демонстрації на Фіг. 3 ці обмеження опущені.

Просуваючись по структурній схемі технологічного процесу по Фіг. 3, приймається керуючий сигнал p (113) і множиться на прирости (за допомогою блоків множення 114 і 115). Керуючий сигнал (113) звичайно буде являти собою автоматично зроблений показник імовірності того, що сигнал в мовному каналі є мовою. Способи автоматичного визначення імовірності того, що сигнал є мовою, обговорювалися в зв'язку з Фіг. 1 (див. процесор 130 імовірності мови). Ці скоректовані прирости потім застосовуються до своїх відповідних немовних каналів (за допомогою блоків посилення 116 і 117), як це описано раніше.

Фіг. 4А і Фіг. 4В являють собою структурні діаграми, що демонструють варіації аспектів, показаних на Фіг. 1-3. Додатково, фахівці в даній галузі техніки розпізнають декілька шляхів комбінування елементів винаходу, описаних на кресленнях 1-3.

Фіг. 4А показує, що конфігурація на Фіг. 1 також може бути застосована до однієї або декількох підсмуг частот сигналів L , C , і R . Більш конкретно, кожен з цих сигналів L , C і R може бути пропущений через групу фільтруючих блоків (441, 442 і 443) для вироблення трьох множин з n підсмуг смуги частот: $\{L_1, L_2, \dots, L_n\}$, $\{C_1, C_2, \dots, C_n\}$ і $\{R_1, R_2, \dots, R_n\}$. Підсмуги, які підходять в

пару, пропускаються в n екземплярів контуру 125, продемонстрованого на Фіг. 1, і оброблені підсигнали рекомбінуються (за допомогою контурів підсумовування 451 і 452). Для кожної із субсмуг можуть бути вибрані окремі порогові величини Θ_n . Гарним вибором є множина, в якій Θ_n пропорційні середньому числу мовних тональних міток, які переносяться у відповідній області частот; тобто смугам на краях спектра частот приписуються менші порогові величини, ніж смугам, відповідним домінуючим частотам мови. Ця реалізація винаходу пропонує дуже хороший вибір оптимальних співвідношень між складністю обчислень і продуктивністю системи.

Фіг. 4В показує інший варіант. Наприклад, для зниження обчислювального навантаження може бути поліпшений типовий об'ємний звуковий сигнал з п'ятьма каналами (C, L, R, ls і rs) за допомогою обробки сигналів L і R відповідно до контуру 325, показаного на Фіг. 3, і сигналів ls і rs, які звичайно менш потужні, ніж сигнали L і R, відповідно до контуру 125, показаного на Фіг. 1.

У описах, приведених вище, використовуються терміни "мова" (або мовний звуковий супровід або мовний канал або мовний сигнал) і "не мова" (або немовний звуковий супровід або немовний канал або немовний сигнал). Кваліфікований фахівець в даній галузі техніки зрозуміє, що ці терміни в більшій мірі використовуються для того, щоб встановити відмінність, а в меншій мірі для того, щоб абсолютно описати зміст цих каналів. Наприклад, в сцені фільму в ресторані, мовний канал переважно може нести в собі діалог за одним столом, а немовні канали можуть нести в собі діалоги за іншими столами (таким чином, обидва канали несуть "мову", як використав би цей термін не професіонал). Проте, визначені втілення даного винаходу направлені на ослаблення саме діалогів за іншими столами.

Реалізації винаходу

Цей винахід може бути реалізований у вигляді обладнання або програмного забезпечення супроводу, або у вигляді комбінації і того, і іншого (наприклад, матриці програмованих логічних елементів). Якщо точно не вказано, алгоритми, включені до складу винаходу, по суті не належать до якого-небудь конкретного комп'ютера або іншого пристрою. Зокрема, можуть бути використані різні комп'ютери загального користування з програмами, написаними відповідно до того, що пояснено в даному документі, або може виявитися більш зручним сконструювати спеціалізований пристрій (наприклад, інтегральну схему) для здійснення необхідних етапів способу.

Отже, цей винахід може бути реалізований у вигляді однієї або декількох комп'ютерних програм, що виконуються на одній або декількох комп'ютерних системах, що програмуються, кожна з яких містить щонайменше один процесор, щонайменше одну систему зберігання даних (включаючи довготривалу і не довготривалу пам'ять і/або елементи зберігання даних), щонайменше один пристрій введення або порт введення і щонайменше один пристрій виведення або порт виведення. Програмний код застосовує вхідні дані для здійснення функціональних засобів, описаних тут, і генерує вихідну інформацію. Ця вихідна інформація, відомим чином, прямує до одного або декількох пристроїв виходу.

Кожна така програма може бути реалізована на будь-якій бажаній комп'ютерній мові (включаючи машинні, асемблерні або процедурні, логічні або об'єктно-орієнтовані мови програмування) для роботи з комп'ютерною системою. У будь-якому випадку, мова може бути мовою програмування, що інтерпретується або транслюється.

Кожна така комп'ютерна програма переважно зберігається в середовищі або пристрої зберігання інформації або завантажується туди (наприклад, твердотільна пам'ять або середовище, або магнітне або оптичне середовище), зчитувана програмованим комп'ютером (спеціалізованим або загального користування), для налаштування і функціонування цього комп'ютера після того, як комп'ютерна програма звернеться до середовища або пристрою зберігання інформації для здійснення описаних тут процедур. Може також бути розглянута реалізація цієї системи винаходу як середовище зберігання зчитуваної комп'ютером інформації, оснащена комп'ютерною програмою, при цьому середовище зберігання інформації, настроєне таким чином, примушує цю комп'ютерну систему функціонувати спеціальним і заздалегідь визначеним чином для здійснення функціональних засобів, описаних тут.

Опис, приведений вище, демонструє різні втілення даного винаходу разом з прикладами того, як може бути реалізоване даний винахід. Приклади і втілення, приведені вище, не треба сприймати як єдино можливі втілення, і вони представлені для демонстрації гнучкості і переваги даного винаходу, як це визначено в подальшій формулі винаходу. На основі розкриття суті винаходу, приведенного вище, і наступної формули винаходу, фахівцям в даній галузі техніки будуть зрозумілі інші конфігурації, втілення, реалізації винаходу і їх еквіваленти, які можуть бути використані без відходу від духу і букви цього винаходу, як це визначено в формулі винаходу.

ФОРМУЛА ВИНАХОДУ

1. Спосіб поліпшення чутності мови в багатоканальному звуковому сигналі, який включає етапи, на яких:

5 порівнюють першу характеристику і другу характеристику багатоканального звукового сигналу для формування коефіцієнта ослаблення, причому перша характеристика відповідає першому каналу багатоканального звукового сигналу, який містить мовний звук і немовний звук, причому перша характеристика відповідає першому виміряному показнику, який належить до рівня сигналу в першому каналі, причому друга характеристика відповідає другому каналу багатоканального звукового сигналу, який містить переважно немовний звук, і причому друга характеристика відповідає другому виміряному показнику, який належить до рівня сигналу у другому каналі, причому згадане порівняння містить наступні операції: визначають різницю між першим вимірним показником і другим вимірним показником, і обчислюють коефіцієнт ослаблення на основі різниці між першим вимірним показником і другим вимірним показником, і пороговою величиною; коректують коефіцієнт ослаблення відповідно до значення імовірності мови для формування скоригованого коефіцієнта ослаблення; і ослаблюють другий канал з використанням скоригованого коефіцієнта ослаблення.

2. Спосіб за п. 1, який додатково включає наступний етап:

20 обробляють багатоканальний звуковий сигнал для формування першої характеристики і другої характеристики.

3. Спосіб за п. 1, який додатково включає наступний етап:

обробляють перший канал для формування значення імовірності мови.

25 4. Спосіб за п. 1, в якому другий канал є одним з множини других каналів, при цьому друга характеристика є однією з множини других характеристик, причому коефіцієнт ослаблення є одним з множини коефіцієнтів ослаблення, і причому скоригований коефіцієнт ослаблення є одним з множини скоригованих коефіцієнтів ослаблення, причому спосіб додатково включає наступні етапи:

30 порівнюють першу характеристику і множину других характеристик для формування множини коефіцієнтів ослаблення; корегують множину коефіцієнтів ослаблення відповідно до значення імовірності мови для формування множини скоригованих коефіцієнтів ослаблення; і ослаблюють множину других каналів з використанням множини скоригованих коефіцієнтів ослаблення.

35 5. Спосіб за п. 1, в якому багатоканальний звуковий сигнал містить третій канал, який містить переважно немовний звук, при цьому спосіб додатково включає наступні етапи:

40 порівнюють першу характеристику і третю характеристику для формування додаткового коефіцієнта ослаблення, причому третя характеристика відповідає третьому каналу; коректують додатковий коефіцієнт ослаблення відповідно до значення імовірності мови для формування скоригованого додаткового коефіцієнта ослаблення; і ослаблюють третій канал з використанням скоригованого коефіцієнта ослаблення.

45 6. Спосіб за п. 1, в якому перший вимірний показник є першим рівнем потужності сигналу в першому каналі, при цьому другий вимірний показник є другим рівнем потужності сигналу у другому каналі, і причому різниця є різницею між першим рівнем потужності і другим рівнем потужності.

7. Спосіб за пп. 1-5, в якому перший вимірний показник є першою потужністю сигналу в першому каналі, при цьому другий вимірний показник є другою потужністю сигналу у другому каналі, і причому різниця є відношенням між першою потужністю і другою потужністю.

50 8. Пристрій, який містить схему для поліпшення чутності мови в багатоканальному звуковому сигналі, при цьому пристрій містить:

55 схему порівняння, яка виконана з можливістю порівняння першої характеристики і другої характеристики багатоканального звукового сигналу для формування коефіцієнта ослаблення, причому перша характеристика відповідає першому каналу багатоканального звукового сигналу, який містить мовний звук і немовний звук, причому перша характеристика відповідає першому виміряному показнику, який належить до рівня сигналу в першому каналі, причому друга характеристика відповідає другому каналу багатоканального звукового сигналу, який містить переважно немовний звук, і причому друга характеристика відповідає другому виміряному показнику, який належить до рівня сигналу у другому каналі, причому схема порівняння виконана з можливістю:

60 визначення різниці між першим вимірним показником і другим вимірним показником, і

обчислення коефіцієнта ослаблення на основі різниці між першим вимірним показником і другим вимірним показником, і пороговою величиною;

помножувач, який виконаний з можливістю корекції коефіцієнта ослаблення відповідно до значення імовірності мови, для формування скоригованого коефіцієнта ослаблення; і

5 підсилювач, який виконаний з можливістю ослаблення другого каналу з використанням скоригованого коефіцієнта ослаблення.

9. Пристрій за п. 8, в якому перша характеристика відповідає першому рівню потужності, і при цьому друга характеристика відповідає другому рівню потужності, і причому схема порівняння містить:

10 перший суматор, який виконаний з можливістю віднімання першого рівня потужності з другого рівня потужності, для формування різниці рівнів потужності;

другий суматор, який виконаний з можливістю підсумовування різниці рівнів потужності і порогової величини, для формування границі робочого режиму;

15 схему обмеження, яка виконана з можливістю обчислення коефіцієнта ослаблення як більшої величини з границі робочого режиму і нуля.

10. Пристрій за п. 8, в якому перша характеристика відповідає першому рівню потужності, і при цьому друга характеристика відповідає другому рівню потужності, причому пристрій додатково містить:

20 перший блок оцінки потужності, який виконаний з можливістю обчислення першого рівня потужності першого каналу; і

другий блок оцінки потужності, який виконаний з можливістю обчислення другого рівня потужності другого каналу.

11. Пристрій за п. 8, який додатково містить:

25 процесор визначення мови, який виконаний з можливістю обробки першого каналу, для формування значення імовірності мови.

12. Машиночитаний носій інформації, який містить комп'ютерну програму для поліпшення чутності мови в багатоканальному звуковому сигналі, при цьому комп'ютерна програма керує пристроєм для виконання обробки, яка включає:

30 порівняння першої характеристики і другої характеристики багатоканального звукового сигналу для формування коефіцієнта ослаблення, причому перша характеристика відповідає першому каналу багатоканального звукового сигналу, який містить мовний звук і немовний звук, причому перша характеристика відповідає першому вимірному показнику, який належить до рівня сигналу в першому каналі, причому друга характеристика відповідає другому каналу багатоканального звукового сигналу, який містить переважно немовний звук, і причому друга

35 характеристика відповідає другому вимірному показнику, який належить до рівня сигналу у другому каналі, причому порівняння включає:

визначення різниці між першим вимірним показником і другим вимірним показником, і обчислення коефіцієнта ослаблення на основі різниці між першим вимірним показником і

40 другим вимірним показником, і пороговою величиною;

корекцію коефіцієнта ослаблення відповідно до значення імовірності мови для формування скоригованого коефіцієнта ослаблення; і

ослаблення другого каналу з використанням скоригованого коефіцієнта ослаблення.

13. Пристрій для поліпшення чутності мови в багатоканальному звуковому сигналі, при цьому пристрій містить:

45 засіб для порівняння першої характеристики і другої характеристики багатоканального звукового сигналу, для формування коефіцієнта ослаблення, причому перша характеристика відповідає першому каналу багатоканального звукового сигналу, який містить мовний звук і немовний звук, причому перша характеристика відповідає першому вимірному показнику, який належить до рівня сигналу в першому каналі, причому друга характеристика відповідає другому

50 каналу багатоканального звукового сигналу, який містить переважно немовний звук, і причому друга характеристика відповідає другому вимірному показнику, який належить до рівня сигналу у другому каналі, причому засіб для порівняння містить:

засіб для визначення різниці між першим вимірним показником і другим вимірним показником, і

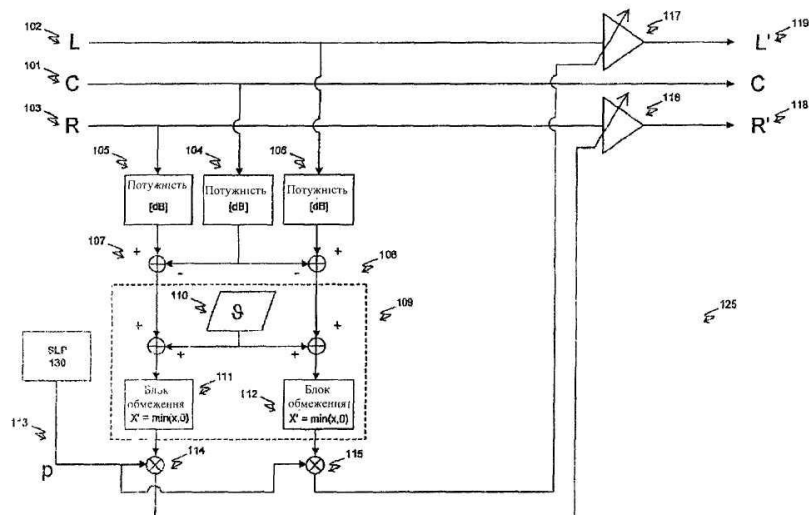
55 засіб для обчислення коефіцієнта ослаблення на основі різниці між першим вимірним показником і другим вимірним показником, і пороговою величиною;

засіб для корекції коефіцієнта ослаблення відповідно до значення імовірності мови, для формування скоригованого коефіцієнта ослаблення; і

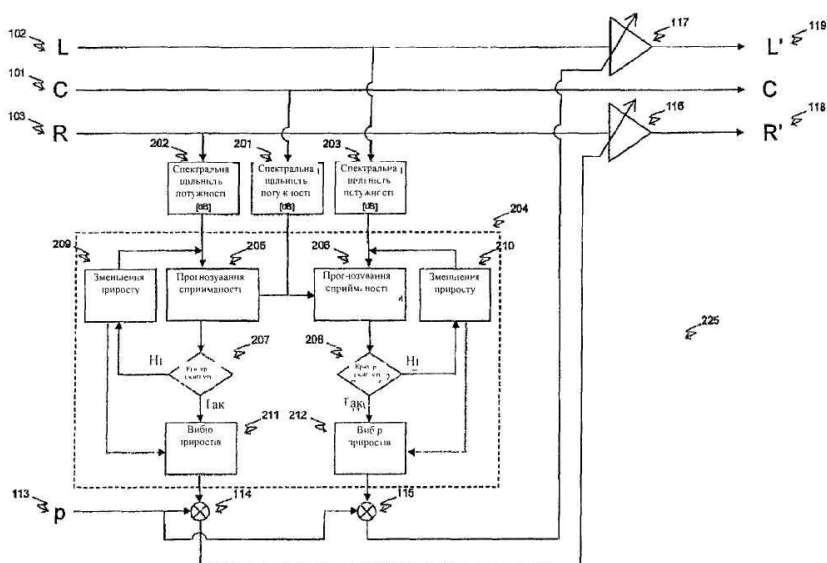
засіб для ослаблення другого каналу з використанням скоригованого коефіцієнта ослаблення.

14. Пристрій за п. 13, в якому перша характеристика відповідає першому рівню потужності, і при цьому друга характеристика відповідає другому рівню потужності, і причому засіб для порівняння містить:

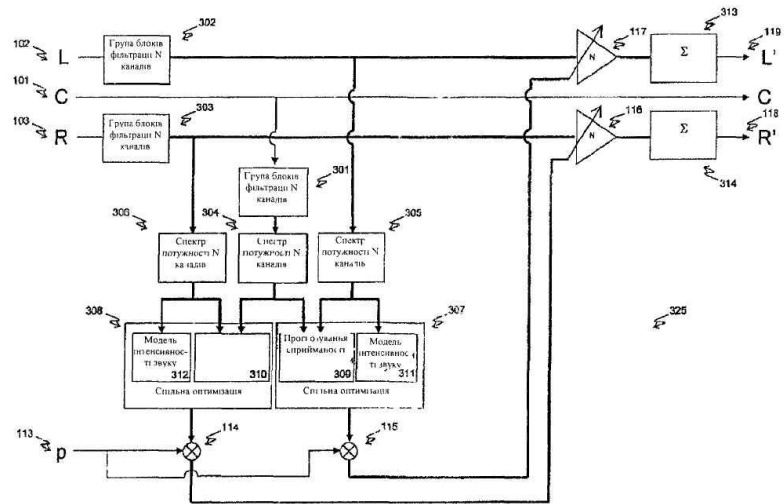
засіб для віднімання першого рівня потужності з другого рівня потужності, для формування різниці рівнів потужності.



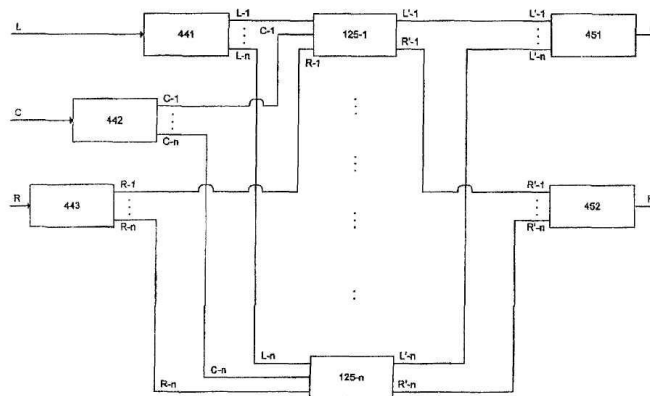
Фиг. 1



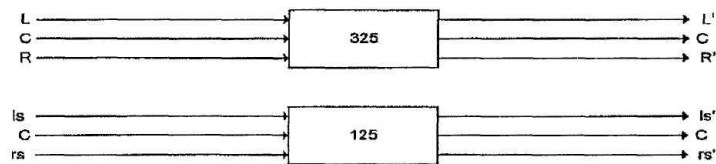
Фиг. 2



Фиг. 3



Фиг. 4A



Фиг. 4B

Комп'ютерна верстка Л. Бурлак

Державна служба інтелектуальної власності України, вул. Урицького, 45, м. Київ, МСП, 03680, Україна

ДП "Український інститут промислової власності", вул. Глазунова, 1, м. Київ – 42, 01601