



УКРАЇНА

(19) UA (11) 52351 (13) U
(51) МПК (2009)
G10L 15/00

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ДЕПАРТАМЕНТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ

ОПИС ДО ПАТЕНТУ НА КОРИСНУ МОДЕЛЬ

видається під
відповідальність
власника
патенту

(54) СПОСІБ ГОЛОСОВОГО КЕРУВАННЯ ВІДЕОТЕЛЕФОНОМ-ПРИСТАВКОЮ

1

2

(21) u201001609

(22) 16.02.2010

(24) 25.08.2010

(46) 25.08.2010, Бюл. № 16, 2010 р.

(72) ВІНЦЮК ТАРАС КЛИМОВИЧ, ГРИЦЕНКО ВОЛОДИМИР ІЛЛІЧ, РЯБОВ ОЛЕКСІЙ ПЕТРОВИЧ

(73) МІЖНАРОДНИЙ НАУКОВО-НАВЧАЛЬНИЙ ЦЕНТР ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА СИСТЕМ

(57) Спосіб голосового керування відеотелефоном-приставкою, що передбачає обробку звукового сигналу, який **відрізняється** тим, що звуковий сигнал перетворюють в цифрову форму, розпізнають його за алгоритмом автоматичного розпізнавання звукових сигналів, порівнюючи з еталонними сигналами команд, й в залежності від результату розпізнавання відмовляють в розпізнаванні або генерують сигнал керування функціями та режимами роботи усього пристрою.

Корисна модель відноситься до галузі розпізнавання мовленнєвих сигналів і призначений для створення пристрою „відеотелефон-приставка телефонної мережі загального користування”, що керується голосом.

До аналогів моделі належать способи керування відеотелефоном-приставкою телефонної мережі загального користування, що ґрунтуються на використанні кнопок та екрану. Недоліком такого способу керування є складна структура активації функцій відеотелефона-приставки, що є результатом обмеженості засобів керування в портативних настільних пристроях. Іншим недоліком є необхідність запам'ятовування номера телефону необхідного абонента чи ускладнена процедура його пошуку в звичайному телефонному електронному довідникові.

Відомо декілька видів відеотелефонів-приставок телефонної мережі загального користування, але відеотелефонів-приставок телефонної мережі загального користування із голосовим способом керування функціями та з'єднаннями немає. Тому за найближчий аналог обраний спосіб керування відеотелефоном-приставкою телефонної мережі загального користування BEAMER-3 (www.vialta.com), яким керують за допомогою дєв'яти кнопок керування та рідкокристалічного екрану. Керування пристроєм відбувається за допомогою меню. Відповідно, в кожному з пунктів меню змінюються функції кнопок керування й призначення спеціальних позначок на рідкокристалічному екрані.

Недоліком найближчого аналога є багаторівневий процес керування функціями. Зважаючи на значну кількість пунктів меню, станів рідкокристалічного екрану й різні функції кнопок за різних пунктів меню, керування пристроєм вимагає неабиякої початкової підготовки. Іншим недоліком є ускладненість пошуку записів телефонного електронного довідника. Причиною цих недоліків є стала технологія екранного, візуального та тактильного керування, натомість дана модель пропонує інше, голосове керування функціями відеотелефона-приставки.

В основу моделі покладено задачу вдосконалення процесу керування функціями відеотелефона-приставки шляхом реалізації технології розпізнавання голосових команд, що дозволяє керувати функціями відеотелефона-приставки природною мовою та організувати телефонний електронний довідник з включенням голосових ідентифікаторів.

Поставлена задача вирішується способом голосового керування відеотелефоном - приставкою, що передбачає обробку звукового сигналу, який відрізняється тим, що звуковий сигнал перетворюють в цифрову форму, розпізнають його за алгоритмом автоматичного розпізнавання звукових сигналів, порівнюючи з еталонними сигналами команд, й в залежності від результату розпізнавання відмовляють в розпізнаванні або генерують сигнал керування функціями та режимами роботи усього пристрою.

Модель пропонує принципово новий, відсутній в найближчому аналозі, спосіб керування функціями відеотелефона-приставки. Так, для реалізації

(13) U

(11) 52351

(19) UA

функції фотографування на відсилання знімку кореспондентові в BEAMER-3, за допомогою кнопок та екрану фіксують певний кадр, потім за допомогою кнопок обирають функцію пересилки зображення. У моделі, для фіксації певного кадру та його відправки можна назвати певну голосову команду. Окрім того, назвати таку команду може й кореспондент, якщо дозволяють такий режим роботи. Для здійснення телефонного з'єднання в BEAMER-3, чи вводять телефонний номер на екрані за допомогою клавіш, чи знаходять потрібний номер в телефонному електронному довідникові також за допомогою клавіш, якщо ж надходить виклик від кореспондента, то для встановлення з'єднання також натискають клавішу. У моделі ж для здійснення з'єднання промовляють голосовий ідентифікатор кореспондента чи промовляють команду у випадку надходження виклику від кореспондента, якщо увімкнено режим голосового розпізнавання, також передбачено традиційний спосіб встановлення з'єднання за допомогою клавіш. Загалом, у моделі за голосовими командами можуть виконувати функції встановлення з'єднання, фіксації та передачі зображення, всі функції роботи з телефонним електронним довідником, цифровим автовідповідачем.

Найважливішою новацією моделі є спосіб реалізації керування відеотелефоном-приставкою за допомогою голосових команд як від користувача, так і від кореспондента. Такий спосіб керування гарантує зручність, швидкість та простоту при керуванні пристроєм як для безпосереднього користувача, так і для його кореспондента або для віддаленого користувача через телефонну мережу.

Пристрій під'єднують до проводової телефонної мережі загального користування. Він дозволяє під час встановленого голосового телефонного зв'язку виконувати фіксацію та передачу по цій телефонній лінії сигналу відеозображення з можливістю використання для цієї мети голосових команд користувача, а також реалізовувати приймання відеозображення з відображенням його на дисплеї пристрою. Для одночасної передачі по телефонній мережі голосового та відеосигналу використовується стиснення інформації.

Керування функціями відеотелефона-приставки можна виконувати за допомогою голосових команд. Для цього називають команду та/або голосовий ідентифікатор абонента, виклик якого потрібно здійснити. Результат дії команди, підказки або виявлені помилки можуть озвучуватися у вигляді голосових повідомлень поряд із екранним відображенням. Іменування абонента здійснюють під час формування телефонного електронного довідника. Для цього вимовляють ім'я абонента. Запис, що містить телефонний номер абонента, його текстовий ідентифікатор та може містити голосовий ідентифікатор, зберігається в телефонному електронному довідникові. Якщо запис містить голосовий ідентифікатор, то виклик цього абонента чи редагування запису телефонного електронного довідника можуть здійснювати шляхом промовляння голосового ідентифікатора. Також виклик будь-якого кореспондента можуть здійснювати шляхом промовляння цифрового но-

меру кореспондента в телефонній мережі. Якщо умикають режим розпізнавання мови користувача, то фіксацію та відправлення зображення можуть здійснювати після промовляння команд, наприклад, "СФОТОГРАФУВАТИ" та "ВІДПРАВИТИ". При віддаленому керуванні пристроєм, викликають пристрій з будь-якого іншого терміналу телефонної мережі, попередньо виконують активацію цього режиму безпосередньо на пристрої. При цьому пристрій виконує операцію встановлення зв'язку та очікує певний час від користувача голосового пароллю. Якщо його називають, то пристрій переходить в описуваний режим роботи. При цьому для користувача доступні всі голосові команди. Таким чином, можна отримати інформацію про стан приміщення, в якому знаходиться пристрій, а також отримати записи телефонного електронного довідника та прослухати повідомлення автовідповідача.

На фіг. 1 зображено послідовність дії команди голосового керування. Звуковий сигнал поступає на аналізатор, в якому сигнал перетворюється до форми, прийнятної для розпізнавання. В блоці розпізнавання перетворений сигнал порівнюється з еталонами команд, що зберігаються в енергонезалежній пам'яті відеотелефона-приставки, в результаті чого або генерується сигнал керування функціями відеотелефона-приставки або відбувається відмова від розпізнавання. У випадку, коли голосова команда потребує подальшого уточнення або підтвердження дії, може відбуватися діалог між пристроєм та людиною.

Спосіб голосового керування функціями відеотелефона-приставки, реалізований у моделі, дозволив зменшити кількість кнопок керування до п'яти та реалізувати функцію голосового керування пристроєм через телефонну мережу.

На фіг.2 зображена структурна схема відеотелефона-приставки з голосовим керуванням. Пристрій складається з наступних блоків:

- 1 - кольоровий дисплей
- 2- матриця, здатна до світлосприйняття
- 3- перетворювач інтерфейсу дисплея
- 4- перетворювач інтерфейсу матриці, що здатна до світлосприйняття
- 5- центральний процесор обробки сигналів
- 6- керуємий модуль розпізнавання
- 7- блок аналого-цифрового та цифро-аналогового перетворення
- 8- мікрофонний підсилювач
- 9- підсилювач звукового спектру
- 10- мікрофон
- 11 - звуковий випромінювач
- 12- перетворювач телефонного інтерфейсу
- 13- блок аналого-цифрового та цифро-аналогового перетворення

Вхід дисплею 1 під'єднаний до виходу перетворювача 3 інтерфейсу дисплея, вхід якого під'єднаний до виходу центрального процесора 5 обробки сигналів. Вихід матриці 2 під'єднаний до входу перетворювача 4 інтерфейсу матриці, вихід якого під'єднаний до входу процесора 5. Відповідні входи та виходи процесора 5 також під'єднані до відповідних входів та виходів керуемого модуля 6 розпізнавання, першого блока 7 аналого-

цифрового та цифро-аналогового перетворення, перетворювача 12 телефонного інтерфейсу та другого блока 13 аналого-цифрового та цифро-аналогового перетворення для двостороннього обміну даними. Керуємий модуль 6 також під'єднаний до першого 7 та другого 13 блоків аналого-цифрового та цифро-аналогового перетворення для двостороннього обміну даними. Вихід мікрофона 10 під'єднаний до входу мікрофонного підсилювача 8, вихід якого під'єднаний до входу першого блока 7 аналого-цифрового та цифро-аналогового перетворення. Вхід звукового випромінювача 11 під'єднаний до виходу підсилювача 9 звукового підсилювача, вхід якого під'єднаний до виходу першого блока 7 цифро-аналогового перетворення. Другий блок 13 аналого-цифрового та цифро-аналогового перетворення під'єднаний до перетворювача 12 телефонного інтерфейсу для двостороннього обміну даними.

Після включення пристрою він переходить у початковий режим, при якому відбувається очікування можливих натискань клавіш користувачем.

Пристрій очікує команди від користувача про перехід його або в режим навчання голосовим командам, або в черговий робочий режим. У режимі навчання пристрою називають голосові команди, що необхідні для його коректної роботи в черговому робочому. Після навчання пристрою голосовим командам і запису в енергонезалежну пам'ять необхідної для розпізнавання голосових команд інформації, пристрій переходить у черговий робочий режим, при якому ним керують або від натискання кнопок, або від голосових команд, або від кнопок і голосових команд одночасно, і здатний сприймати вхідні телефонні виклики, ініціювати виклики а також здійснювати передачу голосу і голосу з зображеннями (відповідно, режими "голос" і "голос-зображення").

Пристрій працює в такий спосіб. Центральний процесор 5 обробки сигналів може або сам ініціювати телефонний виклик через перетворювач 12 телефонного інтерфейсу, або одержати від нього сигнал про виклик, що надійшов. У випадку виклику, що надійшов, процесор 5 здійснює операцію встановлення зв'язку відповідною командою перетворювача 12. У випадку, коли ініціюють виклик від пристрою, зв'язок встановлюється після виконання операції зняття трубки віддаленим терміналом. При встановленому зв'язку пристрій спочатку переходить у режим звичайного телефонного голосового зв'язку (режим "голос"), при якому звуковий сигнал прийому відокремлюється від сигналу передачі в перетворювачі 12 телефонного інтерфейсу, надходить на другий блок 13 аналого-цифрового та цифро-аналогового перетворення, де він перетворюється з аналогової форми в цифрову, далі надходить на керуємий модуль 6 розпізнавання. Модуль 6 аналізує сигнал прийому на наявність маркера переведення пристрою в режим "голос-зображення". При його відсутності прийнятий сигнал без перетворень надходить на перший блок 7 аналого-цифрового та цифро-аналогового перетворення, де він перетворюється знову в аналогову форму, і далі поступає на підсилювач 9 звукового сигналу і звуковий випромінювач 11.

Якщо попередньо вмикають режим голосового розпізнавання, то з прийнятого звукового сигналу виділяються команди керування пристроєм. Звуковий сигнал передачі проходить від мікрофону 10 на мікрофонний підсилювач 8, де він підсилюється, далі поступає на перший блок 7 аналого-цифрового та цифро-аналогового перетворення, у якому він перетворюється з аналогової форми в цифрову і далі надходить на керуємий модуль 6. Якщо попередньо включають режим розпізнавання, то з голосового сигналу виділяються команди керування. Отриманий цифровий сигнал передачі надходить далі на другий блок 13 аналого-цифрового та цифро-аналогового перетворення, де він знову перетворюється в аналогову форму, після чого надходить на перетворювач 12 телефонного інтерфейсу, у якому він змішується з сигналом прийому і передається по двопроводовій телефонній лінії.

Переведення пристрою в режим "голос-зображення" відбувається або від віддаленого терміналу шляхом розпізнавання маркера в сигналі прийомного каналу блоком 6, або натискають кнопки переведення пристрою в режим "голос-зображення".

У режимі "голос-зображення" голосова частина сигналу передачі проходить від мікрофона 10 через мікрофонний підсилювач 8, надходить на перший блок 7 аналого-цифрового та цифро-аналогового перетворення, у якому вона перетворюється в цифрову форму, і далі надходить на керуємий модуль 6 розпізнавання і на процесор 5 обробки сигналів. У модулі 6 здійснюється розпізнавання голосової складової сигналу і виділення команд керування від цього сигналу. Відеочастина передавального сигналу надходить від матриці 2, перетворюється в перетворювачі 4 інтерфейсу матриці у цифрову форму і теж надходить на процесор 5. Процесор 5 здійснює стискання і змішування двох складових передавального сигналу, а також керування матрицею 2 і перетворювачем 4. Вхідний вплив для цього керування ініціюється або від голосової частини передавального сигналу шляхом реалізації алгоритму його розпізнавання в керуємому модулі 6 розпізнавання, або від зовнішньої кнопки, що електрично з'єднана з процесором 5. Сумарний передавальний сигнал від процесора 5 надходить без перетворення на другий блок 13 аналого-цифрового та цифро-аналогового перетворення, і далі на перетворювач 12 телефонного інтерфейсу, у якому він змішується з сигналом прийому і передається по двопроводовій телефонній лінії.

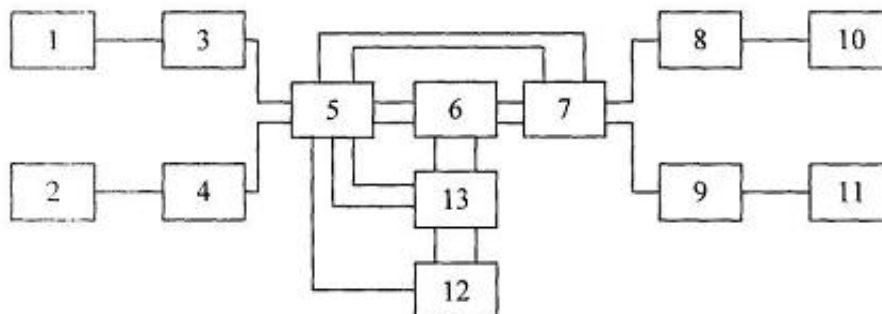
Сумарний сигнал прийому надходить від двопроводової телефонної лінії, відокремлюється від передавального сигналу в перетворювачі 12 телефонного інтерфейсу, перетворюється в цифрову форму в другому блоці 13 аналого-цифрового та цифро-аналогового перетворення, і надходить в процесор 5. Процесор 5 здійснює розпакування прийомного стиснутого сигналу і його поділ на звукову частину і відеочастину. Звукова частина прийомного сигналу надходить далі в керуємий модуль 6 розпізнавання і перший блок 7 аналого-цифрового та цифро-аналогового перетворення, .

У модулі 6 відбувається розпізнавання звукової складової і виділення команд керування, а також маркера переведення пристрою в режим "голос". Одночасно звукова складова надходить із процесора 5 в перший блок 7 аналого-цифрового та цифро-аналогового перетворення, де перетворюється в ньому в аналогову форму, потім надходить в підсилювач 9 і далі в звуковий випромінювач 11. Відеочастина прийомного сигналу після виділення в процесорі 5 змішується з інтерфейсною графічною інформацією, надходить в перетворювач 3

інтерфейсу дисплея і далі на дисплей 1. Можливі різні модифікації і операції розпізнавання мови, розпаковки, заправки та змішування звукового сигналу та відеосигналу в одному модулі, проте це доречно лише при значному збільшенні швидкості обчислювальної здатності цього модуля. Також можливе поєднання в пристрої інших функцій традиційного телефонного апарату зокрема виконання у вигляді пристрою з слухавкою, яка забезпечує персональне звукоприйняття.



Фіг. 1



Фіг. 2