



УКРАЇНА

(19) UA (11) 67702 (13) C2
(51) МПК (2006)
G10L 15/00
H04N 7/00

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ДЕПАРТАМЕНТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ

ОПИС ДО ПАТЕНТУ НА ВІНАХІД

(54) СПОСІБ ГОЛОСОВОГО КЕРУВАННЯ ВІДЕОТЕЛЕФОНОМ-ПРИСТАВКОЮ ТА ПРИСТРІЙ ДЛЯ ЙОГО ЗДІЙСНЕННЯ

1

(21) 20031213080
(22) 30.12.2003
(24) 25.09.2007
(46) 25.09.2007, Бюл. №15, 2007р.
(72) Вінцюк Тарас Климович, Гриценко Володимир Ільїч, Федорин Ярослав Володимирович, Рябов Олексій Петрович, Босенко Ростислав Володимирович
(73) МІЖНАРОДНИЙ НАУКОВО-НАВЧАЛЬНИЙ ЦЕНТР ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА СИСТЕМ, ЗАКРИТЕ АКЦІОНЕРНЕ ТОВАРИСТВО "ЕЛЕКТРОНМАШ-ОБРАЗ"
(56) US 5923363, 13.07.1999
RU 2000101320, 27.01.2002
JP 3032148, 12.02.1991
EP 0194387, 17.09.1986
US 2002133621, 03.07.2003
US 6470076, 24.09.1998
(57) 1. Спосіб голосового керування відеотелефоном-приставкою, який заснований на генеруванні сигналу керування функцією набору телефонного номера, сигналу керування функцією відповіді на телефонний виклик, сигналу керування функцією вибору абонента, сигналу керування функцією прийому-передачі зображень та сигналу керування функціями налаштування сервісних функцій, який відрізняється тим, що звуковий сигнал, що несе інформацію про одну з команд керування відеотелефоном-приставкою, перетворюють у електричний аналоговий сигнал, підсилюють його та перетворюють у інформативний цифровий код, останній запам'ятовують і послідовно порівнюють з набором еталонних цифрових кодів, кожному з яких відповідає одна з команд керування відеотелефоном-приставкою, за результатом порівняння інформативного і еталонних цифрових кодів встановлюють той еталонний цифровий код, який із

2

наперед заданою похибкою дорівнює інформативному цифровому коду, і генерують команду керування, що відповідає даному цифровому коду.
2. Пристрій за п. 1, що містить центральний процесор обробки сигналів, дисплей, перетворювач інтерфейсу дисплея, матрицю, що здатна до світлоприйняття, перетворювач інтерфейсу матриці, перший та другий блоки аналого-цифрового та цифро-аналогового перетворення, мікрофонний підсилювач, підсилювач звукового спектра, мікрофон, звуковий випромінювач і перетворювач телефонного інтерфейсу, причому вхід дисплея під'єднаний до виходу перетворювача інтерфейсу дисплея, вхід якого підключений до одного із виходів центрального процесора обробки сигналів, вихід матриці з'єднаний із входом перетворювача інтерфейсу матриці, вихід якого підключений до одного із входів центрального процесора обробки сигналів, який своїми входами та виходами під'єднаний також до відповідних входів та виходів першого і другого блоків аналого-цифрового та цифро-аналогового перетворення і перетворювача телефонного інтерфейсу, при цьому вихід мікрофона з'єднаний із входом мікрофонного підсилювача, вихід якого підключений до входу першого блока аналого-цифрового та цифро-аналогового перетворення, відповідні входи і виходи другого блока аналого-цифрового та цифро-аналогового перетворення підключені до входів і виходів перетворювача телефонного інтерфейсу, який відрізняється тим, що додатково введено керуючий модуль розпізнавання мови, до якого під'єднані відповідні входи та виходи центрального процесора обробки сигналів, а також входи і виходи першого та другого блоків аналого-цифрового та цифро-аналогового перетворення.

Винахід відноситься до галузі розпізнавання мовленнєвих сигналів і призначений для створення пристрою „відеотелефон-приставка телефонної

мережі загального користування", що керується голосом.

До аналогів винаходу належать способи управління відеотелефоном-приставкою телефон-

(19) UA (11) 67702 (13) C2

ної 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 прототип обраний спосіб керування відеотелефоном-приставкою телефонної мережі загального користування BEAMER-3 (www.vialta.com), яким керують за допомогою дев'яти кнопок керування та рідкокристалічного екрану. Управління пристроєм відбувається за допомогою меню. Відповідно, в кожному з пунктів меню змінюються функції кнопок керування й призначення спеціальних позначок на рідкокристалічному екрані.

Недоліком прототипу 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 даний винахід пропонує інше, голосове керування функціями відеотелефона-приставки.

В основу винаходу 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.

Винахід пропонує принципово новий, відсутній в прототипі, спосіб керування функціями відеотелефона-приставки. Так, для реалізації функції фотографування на відсилення знімку кореспондентів в BEAMER-3, за допомогою кнопок та екрану фіксують певний кадр, потім за допомогою кнопок обирають функцію пересилки зображення. У винаході, 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 винаході 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 винаході за голосовими командами можуть виконувати функції встановлення з'єднання, фіксації та передачі зображення, всі функції роботи з телефонним електронним довідником, цифровим автовідповідачем.

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

Пристрій під'єднують до дротової 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 використовуються 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 номеру 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 управлінні 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 паролю. 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 управління. 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 управління функціями відеотелефона-приставки, реалізований у винаході, дозволив зменшити кількість кнопок керування до п'яти та реалізувати функцію голосового управління пристроєм через телефонну мережу.

На Фіг.2 зображена структурна схема відеотелефона-приставки з голосовим управлінням. 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 - керуючий 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 інтерфейсу дисплея 3, вхід якого підключений до одного із виходів центрального процесора обробки сигналів 5. Вихід матриці 2 з'єднаний із входом перетворювача інтерфейсу матриці 4, вихід якого підключений до одного із входів центрального процесора обробки сигналів 5. Останній своїми входами та виходами під'єднаний також до відповідних входів і виходів першого 7 і другого 13 блоків аналого-цифрового та цифро-аналогового перетворення і перетворювача телефонного інтерфейсу 12. Вихід мікрофона 10 з'єднаний із входом мікрофонного підсилювача 8, вихід якого підключений до входу першого блоку аналого-цифрового та цифро-аналогового перетворення 7. Відповідні входи і виходи другого блоку аналого-цифрового та цифро-аналогового перетворення 13 підключені до входів і виходів перетворювача телефонного інтерфейсу 12. У пристрій додатково введено керуючий модуль розпізнавання мови 6, до якого під'єднані відповідні входи та виходи центрального процесора обробки сигналів 5, а також входи і виходи першого 7 та другого 13 блоків аналого-цифрового та цифро-аналогового перетворення.

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 режимі. 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 шляхом натискання кнопок користувачем, або шляхом подачі голосових команд користувача, або за допомогою 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 обробки сигналів 5 може або сам ініціювати телефонний виклик через перетворювач телефонного інтерфейсу 12, або одержати від нього сигнал про виклик, що надійшов. У випадку надходження виклику, центральний процесор обробки сигналів 5 здійснює операцію встановлення зв'язку відповідною командою перетворювача телефонного інтерфейсу 12. У випадку ініціювання виклику 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 телефонного інтерфейсу 12, надходить на другий блок аналого-цифрового та цифро-аналогового перетворення 13, де він перетворюється з аналогової форми в цифрову, далі надходить на керуючий модуль розпізнавання мови 6. Останній аналізує сигнал прийому на наявність маркера переведення пристрою в режим "голос-зображення". При його відсутності прийнятий сигнал без перетворень надходить на перший блок аналого-цифрового та цифро-аналогового перетворення 7, де він перетворюється знову в аналогову форму, і далі на підсилювач звукового спектру 9 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 на перший блок аналого-цифрового та цифро-аналогового перетворення 7, у якому він перетворюється з аналогової форми в цифрову і далі надходить на керуючий модуль розпізнавання мови 6. Якщо попередньо вмикають режим розпізнавання, то з голосового сигналу виділяються команди керування. Отриманий цифровий сигнал передачі надходить далі на другий блок аналого-цифрового та цифро-аналогового перетворення 13, де він знову перетворюється в аналогову форму, після чого надходить на перетворювач телефонного інтерфейсу 12, у якому він

змішується з сигналом прийому і передається по дводротовій телефонній лінії.

Переведення пристрою в режим "голос-зображення" здійснюється або віддаленим терміналом шляхом розпізнавання керуючим модулем розпізнавання мови 6 відповідного маркера в сигналі прийомного каналу, або за допомогою 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 аналого-цифрового та цифро-аналогового перетворення 7, у якому вона перетворюється в цифрову форму, і далі надходить на керуючий модуль розпізнавання мови 6 і на центральний процесор обробки сигналів 5. У керуючому модулі розпізнавання мови 6 здійснюється розпізнавання голосової складової сигналу і виділення команд керування від цього сигналу. Відеочастина сигналу, що передається, надходить від матриці 2, перетворюється в перетворювачі інтерфейсу матриці 4 у цифрову форму і теж надходить до центрального процесору обробки сигналів 5. Останній здійснює стискання і змішування двох складових сигналу, що передається. Крім того від здійснює керування матрицею 2 і перетворювачем інтерфейсу матриці 4. Керування здійснюється або шляхом реалізації алгоритму розпізнавання голосової частини сигналу, що передається, в керуючому модулі розпізнавання мови 6, або шляхом натискання зовнішньої кнопки, що електрично з'єднана з центральним процесором обробки сигналів 5. Сумарний сигнал, що передається, від центрального процесору обробки сигналів 5 надходить без перетворення на другий блок аналого-цифрового та цифро-аналогового перетворення 13 і далі на перетворювач телефонного інтерфейсу 12, у якому він змішується з сиг-

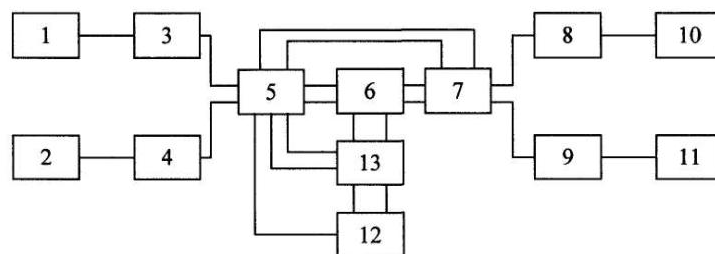
налом прийому і передається по дводротовій телефонній лінії.

Сумарний сигнал прийому надходить від дводротової телефонної лінії, відокремлюється від передавального сигналу в перетворювачі телефонного інтерфейсу 12, перетворюється в цифрову форму в другому блоці аналого-цифрового та цифро-аналогового перетворення 13 і надходить до центрального процесору обробки сигналів 5. Останній здійснює розпакування прийомного стиснутого сигналу і його поділ на звукову частину і відеочастину. Звукова частина прийомного сигналу надходить далі в керуючий модуль розпізнавання мови 6 і перший блок аналого-цифрового та цифро-аналогового перетворення 7. У керуючому модулі розпізнавання мови 6 відбувається розпізнавання звукової складової і виділення команд керування, а також маркера переведення пристрою в режим "голос". Одночасно звукова складова надходить із центрального процесору обробки сигналів 5 у перший блок аналого-цифрового та цифро-аналогового перетворення 7, перетворюється в ньому в аналогову форму, потім надходить у підсилювач звукового спектру 9 і далі на звуковий випромінювач 11. Відеочастина прийомного сигналу після виділення в центральному процесорі обробки сигналів 5 змішується з інтерфейсною графічною інформацією, надходить у перетворювач інтерфейсу дисплея 3 і далі на кольоровий дисплей 1.

Можливі різні модифікації винаходу, зокрема в поєднанні операцій розпізнавання мови, розпакування, запакування 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 його у вигляді пристрою з слухавкою, яка забезпечує персональне звукоприйняття.



Фіг. 1



Фіг. 2