

RZECZPOSPOLITA
POLSKA



Urząd Patentowy
Rzeczypospolitej Polskiej

(12) **OPIS PATENTOWY** (19) **PL** (11) **215208**

(13) **B1**

(21) Numer zgłoszenia: **388382**

(51) Int.Cl.

G09B 21/04 (2006.01)

G10L 13/00 (2006.01)

G10L 21/00 (2006.01)

(22) Data zgłoszenia: **25.06.2009**

(54) **Układ do komunikowania się osób głuchoniemych z osobami słyszącymi**

(43) Zgłoszenie ogłoszono:
03.01.2011 BUP 01/11

(45) O udzieleniu patentu ogłoszono:
29.11.2013 WUP 11/13

(73) Uprawniony z patentu:

**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA, Kraków, PL**

(72) Twórca(y) wynalazku:

WIESŁAW WAJS, Kraków, PL

(74) Pełnomocnik:

rzec. pat. Janina Biernat

PL 215208 B1

Opis wynalazku

Przedmiotem wynalazku jest układ do komunikowania się osób głuchoniemych z osobami słyszającymi, znajdujący zastosowanie do wspomagania wzajemnej komunikacji między tymi osobami, zwłaszcza w różnych obiektach administracji oraz obiektach użyteczności publicznej.

Znany z patentu nr US 6377925 układ elektronicznego urządzenia tłumaczącego przetwarza wejściowy sygnał mowy i tekstu na wielorakie strumienie danych i dostarcza je symultanicznie do użytkownika z dysfunkcją słuchu. Układ takiego urządzenia w preferowanej postaci zawiera komputer PC z podłączonymi odpowiednio urządzeniami peryferyjnymi w postaci kamery video, monitora, klawiatury, myszki i urządzeniami audio w postaci mikrofonu, telefonu oraz z podłączonym łączem internetowym. Do komputera podłączone jest również jedno lub wiele urządzeń wyjściowych typu audio w postaci głośnika i/albo urządzeń wspomagających słyszalność w postaci implantu albo aparatu słuchowego. Komputer wyposażony jest ponadto w odpowiednie oprogramowanie oraz moduły programowe zapisane w pamięci komputera, które realizują różne tłumaczenia tekstu na mowę. W szczególności mowa otrzymywana za pomocą wejściowego urządzenia audio na przykład mikrofonu przekształcana jest na postać tekstu, znaki języka migowego, generowaną komputerowo mowę na wejście urządzenia audio w postaci głośnika lub implantu oraz animację albo dodatkowo na znaki na przykład obrazy obiektów, które są stowarzyszone ze słowami. Znaki języka migowego są tłumaczone za pomocą implementowanego medium cyfrowego ruchomych obrazów, dostarczanych selektywnie z bazy danych i wyświetlanych na monitorze, na których postać ukazuje słowa, albo frazy w języku migowym albo słowa literowane za pomocą układu palców w języku migowym.

Niedogodnością znanego rozwiązania jest to, że jest dostosowane do nauki mowy przez osobę głuchoniemą albo z dysfunkcją słuchu i uniemożliwia komunikowanie się tych osób z osobami słyszającymi.

Układ, według wynalazku, wyposażony w kamerę wizyjną mikrofon oraz co najmniej jeden monitor i klawiaturę, a także elektroniczny układ przetwarzania oraz odpowiednie bazy danych charakteryzuje się tym, że zawiera dwa tora komunikacji, odpowiednio pierwszy tor do komunikowania się osoby głuchoniemiej z osobą słyszającą i drugi tor do komunikowania się osoby słyszającej z osobą głuchoniemą a pierwszy tor komunikacji zawiera klawiaturę i /lub kamerę video, przy czym klawiatura bezpośrednio, a kamera video poprzez blok rozpoznawania gestów połączone są z oddzielnymi wejściami bloku translacji gramatycznej, którego wyjście połączone jest poprzez blok wizualizacji opisów tekstowych z monitorem osoby słyszającej, a wyjście bloku wizualizacji opisów tekstowych połączone jest również poprzez syntezator mowy z głośnikiem osoby słyszającej. Blok rozpoznawania gestów połączony jest z bazą gestów wzorcowych i bazą opisów tekstowych gestów wzorcowych, a blok translacji gramatycznych połączony jest z bazą odmian form gramatycznych i końcówek fleksyjnych, zaś blok wizualizacji opisów tekstowych połączony jest z bazą opisów tekstowych gestów wzorcowych. Natomiast drugi tor zawiera klawiaturę i/lub mikrofon, przy czym klawiatura bezpośrednio, a mikrofon poprzez blok rozpoznawania mowy połączone są z kolejnym wejściem bloku translacji gramatycznej. Wyjście bloku translacji gramatycznej połączone jest również poprzez drugi blok wizualizacji opisów tekstowych gestów wzorcowych z blokiem identyfikacji rodzaju języka głuchoniemych, którego jedno wyjście języka migowego poprzez blok wizualizacji grafiką 3D, zaś drugie wyjście języka migowego poprzez blok wizualizacji sekwencjami filmowymi i równocześnie poprzez blok wizualizacji grafiką 3D połączone są z monitorem osoby głuchoniemiej, przy czym drugi blok wizualizacji opisów tekstowych połączony jest z bazą gestów wzorcowych i bazą opisów tekstowych gestów wzorcowych, a blok wizualizacji grafiką 3D i blok wizualizacji sekwencjami filmowymi połączone są odpowiednio z bazą animacji 3D i bazą sekwencji filmowych.

Układ do komunikowania się osób głuchoniemych z osobami słyszającymi, według wynalazku, umożliwia porównanie wypowiedzi w formie sekwencji gestów osoby niesłyszającej z zapisaną w bazie sekwencją gestów wzorcowych i przekształcenie istoty wypowiedzi gestów/znaków języka migowego na sekwencje słów w formie zdania języka naturalnego, a także przedstawienie w postaci sekwencji gestów/znaków języka migowego i lub miganego istoty zdania, złożonego ze słów języka naturalnego wypowiedzianego przez osobę słyszającą.

Rozwiązanie, według wynalazku, przedstawione jest w przykładzie wykonania na rysunku, który przedstawia schemat ideowo-blokowy układu.

Układ, według wynalazku, zawiera dwa tora komunikacji A, B, odpowiednio pierwszy tor A do komunikowania się osoby głuchoniemiej ON z osobą słyszającą OS i drugi tor B do komunikowania się osoby słyszającej OS z osobą głuchoniemą ON. Pierwszy tor komunikacji A zawiera klawiaturę 1

połączoną bezpośrednio z wejściem bloku translacji gramatycznej 4, z którego kolejnym wejściem poprzez blok rozpoznawania gestów 3 sprzężona jest również kamera video 2. Wyjście bloku translacji gramatycznej 4 połączone jest poprzez blok wizualizacji opisów tekstowych 5 z monitorem 6 osoby słyszającej OS, a wyjście bloku wizualizacji opisów tekstowych 5 połączone jest również poprzez syntezator mowy 7 z głośnikiem 8 tej osoby słyszającej OS. Blok rozpoznawania gestów 3 połączony jest z bazą gestów wzorcowych 9 i bazą opisów tekstowych gestów wzorcowych 13, a blok translacji gramatycznej 4 połączony jest z bazą odmian form gramatycznych i końcówek fleksyjnych 12, zaś blok wizualizacji opisów tekstowych 5 połączony jest z bazą opisów tekstowych gestów wzorcowych 13.

Drugi tor komunikacji B zawiera klawiaturę 14 i mikrofon 15, które są połączone z kolejnym wejściem bloku translacji gramatycznej 4, przy czym klawiatura 14 podłączona jest bezpośrednio, a mikrofon 15 sprzężony jest poprzez blok rozpoznawania mowy 16. Natomiast wyjście bloku translacji gramatycznej 4 połączone jest również poprzez drugi blok wizualizacji opisów tekstowych 17 z blokiem identyfikacji języka 18 głuchoniemych. Jedno wyjście bloku identyfikacji języka 18 głuchoniemych, stanowiące wyjście języka miganego poprzez blok wizualizacji grafiką 3D 19, a drugie wyjście bloku identyfikacji języka 18, będące wyjściem języka migowego, poprzez blok wizualizacji sekwencjami filmowymi 20 i równocześnie poprzez blok wizualizacji grafiką 3D 19 połączone są z monitorem 21 osoby głuchoniemej ON. Ponadto drugi blok wizualizacji opisów tekstowych 17 połączony jest z bazą gestów wzorcowych 9 i bazą opisów tekstowych gestów wzorcowych 13, zaś blok wizualizacji grafiką 3D 19 połączony jest z bazą animacji 3D 10, a blok wizualizacji sekwencjami filmowymi 20 połączony jest z bazą sekwencji filmowych 11.

Układ zrealizowany został w oparciu o komputer PC, który wyposażony jest w odpowiednie oprogramowanie zapisane w jego pamięci, nie uwidocznionej na rysunku, a bloki funkcjonalne wyposażone są w odpowiednie podprogramy, które realizują odpowiednio tłumaczenia tekstu albo mowy języka naturalnego na gesty języka migowego, udostępnianego na monitorze 21 osoby głuchoniemej ON oraz tłumaczenie gestów języka głuchoniemych albo tekstu odpowiednio na tekst albo sygnał akustyczny, udostępniany za pomocą monitora 6 albo głośnika 8 osoby słyszającej OS.

Działanie układu jest następujące. Wzajemne komunikowanie się osoby głuchoniemej ON z osobą słyszącą OS realizowane jest w układzie według wynalazku, w którym pierwszy tor A służy do komunikacji osoby głuchoniemej ON z osobą słyszącą OS, a drugi tor B służy do komunikacji w kierunku przeciwnym. Osoba głuchoniema ON jako użytkownik toru A przedstawia swoją wypowiedź w postaci tekstu korzystając z klawiatury 1 albo w postaci sekwencji gestów języka głuchoniemych, które są rejestrowane w postaci cyfrowego sygnału video przez kamerę video 2. Sygnał z kamery video 2 jest następnie porównywany w bloku rozpoznawania gestów 3 z sygnałem zapisanym w bazie gestów wzorcowych 9. Sekwencja gestów wzorcowych rozpoznana z wysokim prawdopodobieństwem w stosunku do przedstawionej wypowiedzi osoby głuchoniemej ON jest następnie konwertowana na tekst, po czym jest poddawana w bloku translacji gramatycznej 4 gramatycznej analizie porównawczej polegającej na rozpoznawaniu części zdania jakim jest orzeczenie, a następnie w oparciu o konotację czasownika będącego orzeczeniem, sukcesywnym logicznym rozpoznawaniu kolejnych części zdania, czyli podmiotu, dopełnienia bliższego, dopełnienia dalszego, a także okoliczników, z równoczesnym wykorzystaniem bazy odmian form gramatycznych i końcówek fleksyjnych 12 oraz jednoczesnym tworzeniu struktur zdaniowych, które sukcesywnie do ich powstawania porównywane są kolejno w bloku wizualizacji opisów tekstowych 5 z danymi zgromadzonymi w bazie opisów tekstowych gestów wzorcowych 13 języka naturalnego. Gdy na którymkolwiek etapie tworzenia zdania wynik porównania z zapisem zdania w bazie 13 jest jednoznaczny lub wysoce prawdopodobny, wówczas rozpoznane w bazie opisów tekstowych 13 zdanie zostaje wyświetlone w postaci tekstu na monitorze 6 osoby słyszającej OS albo zostaje przetworzone w syntezatorze mowy 7 do postaci akustycznej i wyemitowane przez jej głośnik 8. Gdy natomiast, na którymkolwiek etapie generowania zdania wynik porównywania nie jest jednoznaczny, wówczas realizowany jest etap rozpoznawania następnej w kolejności, według przyjętych reguł gramatycznych danego języka naturalnego, logicznej części zdania, aż do wyczerpania wszystkich możliwości, po czym na monitorze 6 osoby słyszającej OS wyświetlony zostaje komunikat o braku w bazie danych 13 opisu tekstowego, będącego ekwiwalentem gestów wzorcowych jednoznacznych z sekwencją danych gestów osoby głuchoniemej ON.

Osoba słyszająca OS, będąca użytkownikiem toru B przedstawia zaś swoją wypowiedź w postaci tekstu języka naturalnego za pomocą klawiatury 14 lub mówi do mikrofonu 15, w którym sygnał akustyczny przetworzony zostaje na elektryczny sygnał cyfrowy. Sygnał uzyskany z mikrofonu 15 poddawany jest procesowi rozpoznawania mowy w bloku rozpoznawania mowy 16, polegającym na prze-

tworzeniu tego sygnału na zapis tekstowy w postaci zdania. Uzyskany sygnał z bloku rozpoznawania mowy 16 albo sygnał z klawiatury 14 poddawane są następnie procesowi translacji gramatycznej w bloku translacji gramatycznej 4, który polega na analizie zdania poprzez identyfikację obligatoryjnych części zdania począwszy od predykatu czasownikowego wraz z przypisanym schematem zdaniowym, w oparciu o jednostki słownikowe wraz z ich zestawem właściwości morfologicznych (flexsji), składniowych i semantycznych oraz reguły klasyfikacji gramatycznej. Do określonych schematów zdaniowych przypisane są reguły ekwiwalencji języka naturalnego i języka migowego.

W strukturze analizowanego zdania języka naturalnego, przykładowo języka polskiego identyfikowane są następnie sukcesywnie obligatoryjne części zdania w kolejności: podmiot, dopełnienie bliższe, dopełnienie dalsze oraz okolicznik, a następnie nieobligatoryjne części zdania w postaci przydawek. Następnie, uporządkowane fakty informacji źródłowej podlegają w drugim bloku wizualizacji opisów tekstowych 17 porównywaniu z wzorcami opisów tekstowych zgromadzonymi w bazie 13 oraz równocześnie procesowi ustalania ekwiwalentów języka migowego i miganego przy wykorzystaniu danych z bazy gestów wzorcowych 9.

Gdy wynik porównania jest jednoznaczny albo wysoce prawdopodobny, wówczas generowany jest sygnał opisujący zdanie języka naturalnego ekwiwalentami języka migowego i/ lub miganego, a następnie w zależności od rodzaju języka głuchoniemych identyfikowanego w bloku identyfikacji 18 jest on porównywany z wzorcowymi sekwencjami z bazy animacji 3D 10 albo z bazy sekwencji filmowych 11 odpowiednio w bloku wizualizacji grafiką 3D 19 albo w bloku wizualizacji sekwencjami filmowymi 20.

Jeśli wynik porównania w blokach 19 i/lub 20 jest jednoznaczny lub wysoce prawdopodobny, wówczas odpowiednia sekwencja gestów przedstawiana jest na ekranie monitora 21 osoby głuchoniemej ON w języku migowym, albo miganym albo równocześnie w obu językach.

Wzorcowy opis tekstowy zapisany w bazie danych 13 jest zdaniem, dla którego prawdopodobieństwo wystąpienia w konkretnej sytuacji rozmowy osoby głuchoniemej ON z osobą słyszącą OS jest wysokie.

W układzie, według wynalazku, do realizacji przetwarzania sygnałów tekstu, mowy na gesty/znaki języka migowego oraz tekstu, gestu na sygnał akustyczny wykorzystywany jest język programowania PROLOG.

Wykaz oznaczeń na rysunku

A - tor komunikacji osoby głuchoniemej z osobą słyszącą

B - tor komunikacji osoby słyszącej z osobą głuchoniemą

ON - osoba głuchoniema

OS - osoba słyszająca

1 - klawiatura

2 - kamera video

3 - blok rozpoznawania gestów

4 - blok translacji gramatycznej

5 - blok wizualizacji opisów tekstowych

6 - monitor

7 - syntezytor mowy

8 - głośnik

9 - baza gestów wzorcowych

10 - baza animacji 3D

11 - baza sekwencji filmowych

12 - baza odmian form gramatycznych i końcówek fleksyjnych

13 - baza opisów tekstowych gestów wzorcowych

14 - druga klawiatura

15 - mikrofon

16 - blok rozpoznawania mowy

17 - drugi blok wizualizacji opisów tekstowych gestów wzorcowych

18 - blok identyfikacji rodzaju języka głuchoniemych

19 - blok wizualizacji grafiką 3D

20 - blok wizualizacji sekwencjami filmowymi

21 - drugi monitor

PC - komputer

Zastrzeżenie patentowe

Układ do komunikowania się osób głuchoniemych z osobami słyszącymi wyposażony w kamerę wizyjną, mikrofon oraz co najmniej jeden monitor i klawiaturę, a także elektroniczny układ przetwarzania oraz odpowiednie bazy danych, **znamienny tym**, że zawiera dwa tory komunikacji (A, B), odpowiednio pierwszy tor (A) do komunikowania się osoby głuchoniemej (ON) z osobą słyszącą (OS) i drugi tor (B) do komunikowania się osoby słyszącej (OS) z osobą głuchoniemą (ON), a pierwszy tor komunikacji (A) zawiera klawiaturę (1) i /lub kamerę video (2), przy czym klawiatura (1) bezpośrednio, a kamera video (2) poprzez blok rozpoznawania gestów (3) połączone są z oddzielnymi wejściami bloku translacji gramatycznej (4), którego wyjście połączone jest poprzez blok wizualizacji opisów tekstowych (5) z monitorem (6) osoby słyszącej (OS), a wyjście bloku wizualizacji opisów tekstowych (5) połączone jest również poprzez syntezytor mowy (7) z głośnikiem (8) osoby słyszącej (OS), przy czym blok rozpoznawania gestów (3) połączony jest z bazą gestów wzorcowych (9) i bazą opisów tekstowych gestów wzorcowych (13), a blok translacji gramatycznych (4) połączony jest z bazą odmian form gramatycznych i końcówek fleksyjnych (12), zaś blok wizualizacji opisów tekstowych (5) połączony jest z bazą opisów tekstowych gestów wzorcowych (13), natomiast drugi tor (B) zawiera klawiaturę (14) i/lub mikrofon (15), przy czym klawiatura (14) bezpośrednio, a mikrofon (15) poprzez blok rozpoznawania mowy (16) połączone są z kolejnym wejściem bloku translacji gramatycznej (4), a wyjście bloku translacji gramatycznej (4) połączone jest również poprzez drugi blok wizualizacji opisów tekstowych gestów wzorcowych (17) z blokiem identyfikacji rodzaju języka głuchoniemych (18), którego jedno wyjście języka miganego poprzez blok wizualizacji grafiką 3D (19), zaś drugie wyjście języka migowego poprzez blok wizualizacji sekwencjami filmowymi (20) i równocześnie poprzez blok wizualizacji grafiką 3D (19) połączone są z monitorem (21) osoby głuchoniemej (ON), przy czym drugi blok wizualizacji opisów tekstowych (17) połączony jest z bazą gestów wzorcowych (9) i bazą opisów tekstowych gestów wzorcowych (13), a blok wizualizacji grafiką 3D (19) i blok wizualizacji sekwencjami filmowymi (20) połączone są odpowiednio z bazą animacji 3D (10) i bazą sekwencji filmowych (11).

Rysunek

