

(19) HU

MAGYAR
NÉPKÖZTÁRSASÁG



ORSZÁGOS
TALÁLMÁNYI
HIVATAL

SZABADALMI LEÍRÁS

SZOLGÁLATI TALÁLMÁNY

A bejelentés napja: (22) 82. 05. 05

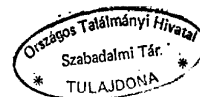
(21) 1410/82

A közzététel napja: (41) (42) 84. 02. 28.

Megjelent: (45) 87.05.13.

(11) 185 527

Nemzetközi
osztályjelzet:
(51) NSZO₄
G 01 L 1/08



Feltaláló(k): (72)

OLASZY Gábor, Kiss Gábor, Budapest okl. vm. programozó mat.

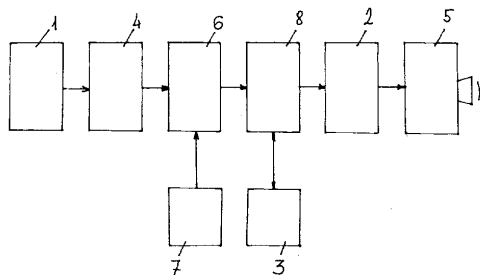
Szabadalmas: (73)

MTA Nyelvtudományi Intézet, Budapest

(54) ELJÁRÁS ÉS BERENDEZÉS SZINTETIZÁTOR(OK) VEZÉRLÉSÉRE, SZÓTÁR NÉLKÜLI SZINTETIZÁLT BESZÉD A VEZÉRLÉSSEL QUASI AZONOS IDŐBEN TÖRTÉNŐ ELŐÁLLTÁSÁRA

(57) KIVONAT

A találmány értelmében akusztikai építőelemeket hozunk létre. Ennek során minden beszédhangra külön-külön előállítjuk a hangsor kezdő elemeket. Ezek időtartamát, frekvencia összetevőit, intenzitás értékeit meghatározzuk. Meghatározzuk továbbá egyrészt az egyszerű vagyis az a, á, o, u, ü, i, é, ö, e, m, n, v, f, sz, s, j, h, l hangok úgynevezett tiszta fázisát reprezentáló elemek időtartamát, frekvencia összetevőit és intenzitás értékeit. Az összetett hangoknál (p, b, t, d, zs, ny, g, gy, ty, k, cs, c, r) is meghatározzuk a kezdő elemén kívüli további építőelemek időtartamát, frekvencia összetevőit, intenzitás értékeit. Az összetett hangok esetében — az egyszerű hangoktól eltérően — a frekvenciaszerkezet, és intenzitás a hang megszólaltatása során nagymértékben, esetenként ugrásszerűen változik. Ezeken túlmenően hangkapcsoló elemeket hozunk létre, ugyancsak meghatározva azok időtartamát, frekvencia összetevőit és intenzitás értékeit, majd minden beszédhangra hangszóró elemet állítunk elő ugyancsak meghatározva azok időtartamát, frekvencia összetevőit és intenzitás értékeit. Fentieket úgy végezzük, hogy 370 akusztikai építőelem álljon rendelkezésre adatbázis képzésére. Ezt az adatbázist minden egyes szintetizált hangsor előállítására ismételten úgy használjuk, hogy a hangsor betűjeleit vagy az azoknak megfelelő kódsorozat útján az adatbázisból a szükséges hangelemeket a betűjelek sorrendjében kiválasztjuk, majd az így előállított monoton hangsorra a hangsorvégi írásjel szerinti önmagában intonációs görbét a monoton hangzásra



1. ábra

megfelelő alaphang frekvencia értékről — és/vagy — irányba megváltoztatva felületjük és az így előállított jellel a szintetizátort vezéreljük.

A találmányunk szerinti vezérlőberendezés azzal jellemezhető, hogy egy beszédszintetizátorhoz oldhatóan vagy fixen önmagában ismert processzor csatlakozik. Ez

utóbbihoz legalább 12 Kbyte kapacitású memóriaegység van kiválasztó áramkör közbeiktatásával kapcsolva. A kiválasztó áramkörhöz összehasonlító egység van kötve, amely egyrészt graféma tárhoz másrészt dekódolóhoz illeszkedik, amely utóbbi adatbefogadó egységhez csatlakozik.

A találmány eljárásra és berendezésre vonatkozik olyan szintetizátor(ok) vezérlésére, amellyel korlátozott szótár nélküli szintetizált beszédet lehet előállítani a vezérléssel quasi azonos időben különösen magyar nyelven.

Ismeretes, hogy a beszéd mesterséges előállításához szükséges szabályok és adatok megállapítása nyelvorientált. Ez azt jelenti, hogy az egyes nyelvek különböző akusztikai szerkezeti sajátosságai miatt külön ki kell dolgozni a szintetizálni kívánt nyelvre használható adatokat. Ezeknek az anyagoknak a birtokában lehet az adott nyelv vonatkozásában beszélszerkezeteket működtetni, így nyelvre jellemző szavakat, mondatokat lehet felépíteni.

Találmányunk elé azt a célt tűztük ki, hogy az MTA Nyelvtudományi Intézetében Bolla Kálmán által létrehozott folyamatos magyar beszédszintetizáló eljárásról eltérően olyan eljárást és berendezést hozzunk létre szintetizátorok vezérlésére, amellyel a folyamatos magyar hangsorokat azonosidejű üzemmódban lehet előállítani.

További célként tűztük ki, hogy a kijelentés, utasítás és kérés hanglejtését a beírt szöveg végén megadott írásjel kód szerint lehessen előállítani.

Célul tűztük ki azt is, hogy a szintetizálandó szavak, mondatok előállításához a magyar helyesírás betűjeleit használjuk, valamint azt, hogy a beszédet ne csak billentyűzet segítségével lehessen megadni a készüléknek, hanem digitális kódformákkal is.

A kitűzött célnak a találmány szerint olyan szintetizátor vezérlő eljárással teszünk eleget, amely azzal jellemezhető, hogy — magyar szakirodalomból önmagában ismert módszerrel — akusztikai építőelemeket hozunk létre. Ennek során minden beszédhangra külön-külön előállítjuk a hangsorkezdő elemeket. Ezek időtartamát, frekvencia összetevőit, intenzitás értékeit meghatározzuk. Meghatározzuk továbbá egyrészt az egyszerű vagyis az a, á, o, u, ü, i, é, ö, e, m, n, v, f, sz, s, j, h, l hangok úgynevezett tiszta fázisát reprezentáló elemek időtartamát, frekvencia összetevőit és intenzitás értékeit. Egyszerű hangoknak tekintjük azokat a beszédhangokat, amelyekben a kezdeti fázistól a befejezésig sem a frekvenciaszerkezetben, sem az intenzitás szerkezetben nagymértékű, illetve ugrásszerű változás nincs.

Erre példaként említjük, hogy amikor az említett egyszerű hangok közül valamelyiket kiejtjük, annak sem frekvenciaszerkezetében, sem intenzitásában változás gyakorlatilag nincs. Az összetett hangoknál (p, b, t, d, zs, ny, g, gy, ty, k, cs, c, r) is meghatározzuk a kezdő elemek kívüli további építőelemek időtartamát, frekvencia összetevőit, intenzitás értékeit. Az összetett hangok esetében — az egyszerű hangoktól eltérően — a frekvenciaszerkezet, és intenzitás a hang megszólaltatása során nagymértékben, esetenként ugrásszerűen változik. Ezekben túlmenően hangkapcsoló elemeket hozunk létre, úgyszintén meghatározva azok időtartamát, frekvencia összetevőit és intenzitás értékeit, majd minden beszédhangra hangsorzáró elemet állítunk elő ugyancsak meghatározva azok

időtartamát, frekvencia összetevőit és intenzitás értékeit.

Találmányunk egyik lényege abban áll, hogy a fent leírt módszerrel 370 elemből álló minimalizált és optimalizált adatbázist hoztunk létre. Azt a továbbiakban minden egyes szintetizált hangsor előállítására ismételtlen úgy használjuk, hogy a hangsor betűjelei vagy az azoknak megfelelő kódsorozat útján az adatbázisból a szükséges hangelemeket a betűjelek sorrendjében kiválasztjuk, majd az így előállított monoton hangsorra a hangsorvégi írásjel szerint intonációs görbét ültetünk rá. Ez képezi találmányunk másik lényegét. Az intonációs görbék és a ráültetés önmagában ismert. Ezt úgy végezzük, hogy a monoton hangzásnak megfelelő alaphang frekvencia értékről az intonációt — és/vagy — irányban megváltoztatva felületjük és az így előállított jellel a szintetizátort vezéreljük.

A találmányunk szerinti vezérlőberendezés azzal jellemezhető, hogy egy beszédszintetizátorhoz oldhatóan vagy fixen önmagában ismert processzor csatlakozik. Ez utóbbihoz legalább 12 Kbyte kapacitású memóriaegység van kiválasztóáramkör közbeiktatásával kapcsolva. A kiválasztóáramkörhöz összehasonlító egység van kötve, amely egyrészt graféma tárhoz, másrészt dekódolóhoz illeszkedik, amely utóbbi adatbefogadó egységhez csatlakozik.

Egy akusztikai építőelemen mint ismeretes a beszédhang azon tört részét értjük, amely meghatározott és a magyar beszédre jellemző

- frekvenciaszerkezettel
- intenzitás szerkezettel
- időtartammal

rendelkezik.

A magyar beszéd építőelemei idegen beszédre (például angol, német, japán) jellemző rendszerek alapján nem határozhatók meg és ahhoz, hogy beszédet az arra jellemző építőelemek alkalmazásával állítsunk elő az építőelemek számát és korábban említett 3 jellemzőjét intuitív módon kell meghatározni.

Úgy találtuk, hogy a magyar beszédet 370, vagy ahhoz közelálló építőelem alkalmazásával elő lehet állítani.

Az akusztikai építőelemek meghatározására úgy találtuk, hogy a magyar nyelv esetében meg kell határozni a beszédhangok és a hangkapcsolódások akusztikai szerkezetére legjellemzőbb műszaki adatokat az idő-, intenzitás-, és a frekvenciaszerkezet síkján, és meg kell határozni azokat az összefüggéseket, amelyek a három sík adatait összefogják. Megállapításunk szerint a magyar beszéd létrehozásához szükséges alapvető hangsorépítő elemek a következők:

- magánhangzó
- magánhangzó-magánhangzó kapcsolódása
- mássalhangzó-magánhangzó kapcsolódása
- magánhangzó-mássalhangzó kapcsolódása
- mássalhangzó-mássalhangzó kapcsolódása

A fenti hangsorépítő elemeket külön kell értelmezni hangsoreleji, hangsorbelseji és abszolút hangsorzáró helyzetre.

A szintetizáláshoz ún. minimalizált és optimalizált adatbázist hoztunk létre, amely a korábban említett 370 hangelemet tartalmazza.

Ezek az építőelemek nem beszédhangok, hanem hangrészek. Rendszerünkben is bizonyos építőelemeket a hangsorépítés során több helyen is fel lehet használni (például a szöveleji „b” hang három elemből áll). Az első a zöngének a csend szintjéről való intenzitásfelvezérlését tartalmazza, a második a fojtott zöngé lényeges szakaszát egészen a zöngés zárfelpattanásig, a harmadik pedig egy rövid időtartamú (kb. 10–12 ms) építőelem, amivel a magyar ejtésű „b” hang zárfelpattanását valósítjuk meg. Ha a „b” hanghoz egy magánhangzót kapcsolunk (például a „ba” szótagban) akkor a zárfelpattanás után képeznünk kell egy olyan részt, amely biztosítja a „b” és a magyar „a” közötti átmenet akusztikai megvalósítását. Ez a „b-a” átmeneti építőelem. Mivel a magánhangzó képzési konfigurációja minden magánhangzótípusnál más (az á, a, o, u, i, é, ü, e hang a magyarban a nyelvre jellemző akusztikai szerkezettel rendelkezik), ezért a „b” magánhangzó átmenetet biztosító építőelemnek is mindig másnak kell lennie attól függően, hogy milyen magánhangzó kapcsolódik a magyar „b” hanghoz. Bonyolítja a szintetizálást, hogy a magyar beszédben a magánhangzó visszahat a „b” hang zárfelpattanásának akusztikai összetevőire. Így a „b” hangnál is annyi (vagy közel annyi) építőelemet kell készíteni ahány magánhangzó van a magyarban. A hangsorépítés során ezután következik a magyar „a” hang ún. tiszta fázisát megvalósító hangszerlet hozzacsatolása. A hangsorépítés hasonlóan folyik tovább. Felhasználás szempontjából a „b” első építőeleme minden hangsorkezdő helyzetben használható (például a ba, bá, bo, ... br, bl, stb. kezdetű szavakban), a második a zöngé középső része minden hangsoreleji helyzetben hozzákapszolható az első születhez, de a magyar beszédben a hangsor j-i helyzetben is ezt használhatjuk (például a baba, bibe, stb. szavak második „b” zöngéjénél). A zárfelpattanás hangszerletét a magyar beszéd felépítésénél nemcsak hangsoreleji, de hangsorbelseji helyzetben is használhatjuk, ha ugyanazon magánhangzóhoz csatlakozunk. Tehát mivel a magyarban 9 magánhangzó van, a „b” zárfelpattanási akusztikai építőelemét (9 db), összesen 18 esetben használhatjuk fel. A magánhangzó tiszta fázisát reprezentáló elemet pedig bárhol felhasználhatjuk a hangsorban. A fenti példából következik, hogy a 370 akusztikai építőelemből vannak olyanok, amelyeket igen sokszor és vannak olyanok, amelyeket csak 1-szer használunk fel a magyar beszéd generálásánál. Az adatbázis minimalizálására azért van szükség, hogy minél kevesebb memóriakapacitást használjunk fel a processzornál. Ez képezi az akusztikai építőelemek jellemzőinek meghatározásán túlmenően találmányunk lényegét.

Ezekkel az építőelemekkel a magyar beszéd bármely szóhangsora létrehozható. Az elemek összes kombinációjából (9 rövid, 5 hosszú magánhangzó, 23 rövid és 23 hosszú mássalhangzó) tehát létrehoztunk egy olyan minimalizált adatbázist, amelyben lévő hangsorépítő elemek kombinációiból bármilyen magyar szóhangsor előállítható. Így bármely szöveg amely a magyar köznyelvnek eleme meghangosítható.

Találmányunk értelmében a fentiekben ismertetett szintetizált beszéd előállítására a következő módon történik:

Először megjelöljük a kérdéses hangsort — esetleg egy adott kódsorozat megadásával — ezt követően értelmezhetjük a betűjeleket, amelyeket magyar beszéd elemeivé alakítunk át.

Ezek helyét a hangsorban is értelmezni kell, azaz meg kell állapítani, hogy az kezdő-, hangsorbelseji-, vagy záróelem-e. Ezt követően kikeressük az adott bázisból azokat az akusztikai építőelemeket, amelyeket a hangsorépítésnél fel kell használni.

További tevékenység a hangsorvégi írásjel értelmezése majd az írásjel szerinti intonációs görbének a hangsorra történő felülletése.

Az így sorbarakott elemeket illetve az azokba inkorporálódó bitsorozatokat önmagában ismert szintetizátorhoz továbbítjuk, amelyek útján a bitsorozatnak megfelelő magyar hangsort megszólaltatjuk.

Az előzőekben ismertetett eljárás fogyanatosítására alkalmas berendezést rajzmellékleten ismertetjük. Az

1. ábra a berendezés vázlatos felépítését mutatja be.

A szintetizálni kívánt szöveg betűsorral, kóddal, illetve kiegészítő jelekkel történő megoldására 1 adatbefogadó egység szolgál, amely billentyűzettel, vagy kódsorozat megadására alkalmas eszközzel rendelkezik. Az említett 1 adatbefogadó egység 4 dekódoló közbeiktatásával 6 összehasonlító egységhez csatlakozik, amely utóbbi egyrészt 7 graféma tárhoz, másrészt 8 kiválasztó áramkörhöz van kötve. A 8 kiválasztó áramkör 3 memóriaegységhez valamint 2 processzorhoz van kapcsolva. Az említett 2 processzor 5 beszéd szintetizátorral van kapcsolatban.

A berendezés működése a következő:

Az 1 adatbefogadó egységet billentyű leütése vagy egy kódsorozat megadása útján aktivizáljuk. Ekkor a 2 processzor értelmezi a betűjeleket, a 4 dekódolótól nyert jelek alapján, amelyeket a 6 összehasonlító egység és 7 graféma tár adataiból képezzük, és meghatározza a magyar hangok helyét a hangsorban. A 2 processzor a 3 memóriaegységben elhelyezkedő adatbázisból a 8 kiválasztó áramkör segítségével kikeresi azoknak az építőelemeknek az azonosító jeleit, amelyeket fel kell használni a hangsorépítésnél, majd azokat sorba helyezi.

Ezt követően az 1 adatbefogadó egység által megadott hangsorvégi írásjelnek megfelelően a hangsorra ráültetésre kerül az intonációs görbe. Az így sorbarakott elemek tartalmának megfelelő bitsorozat alapján, az aktuális hangsort az 5 beszéd szintetizátor közbeiktatásával szólaltatjuk meg. Ezután a 2 processzor és kapcsolt részei eredeti állapotukba térnek vissza, és új hangsorbeütésre kész helyzetbe kerülnek. A berendezés különleges előnye az, hogy a szintetizált magyar beszéd a vezérléssel azonos, illetve quasi azonos időben előállítható.

A 370 elemet tartalmazó adatbázis valamint az átalakítás rendszere lehetőséget nyújt arra, hogy előre betáplált szótár nélkül, tetszőleges magyar hangsorokat akár szótagokat, betűkihagyásos szavakat, logatomokat fordított ejtésű szavakat, stb. generálhassunk. Az általunk kidolgozásra került eljárás és berendezés a szintetizált beszéd korlátozott szótár nélküli és azonosidejű előállítására (realtime) lehetőséget nyújt, a hatékony beszédkutatás kihasználására például a magyar beszéd nyelvészeti-fonetikai vizsgálatára, percepció kísérletekhez, kontrasztív nyelvészeti összehasonlítások elemzésére, szöveg-fonetikai szabályszerűségek meghatározására.

Az előállított mesterséges beszéd iparban történő alkalmazása szinte korlátlan; például „beszélő robotok”, „beszélő” terminálok, riasztórendszerek, gyártásirányítás stb.

A társadalom egyéb területein a mesterséges beszéd felhasználása a mai ismereteket meghaladóan széles körű lehet, például orvosi alkalmazások, beszélő audiométer, vakokat kiszolgáló „beszélő” olvasógép, beszélő hőmérő, mérleg, vérnyomásmérő, némaakat segítő „kézi beszélő gép”; korszerű tömegtájékoztatás, például automatikusan beszélő információs rendszer, pontos idő, időjárásjelentés, meseszolgálat, üzenetfelvevő, tudakozó és más postai alkalmazások: közlekedés, például állomási tájékoztató, megálló bemondó, hibajelző, pilóták oktatása, pilóták tájékoztatása repülés közben: kereskedelem, például áruházi tájékoztató, raktári rendszer tételeinek közlése, hirdetések: játékgyártás, oktatás, például „beszélő tanítógépek”, „beszélő egyszer-egy”, nyelvtanítás, beszélő játékok.

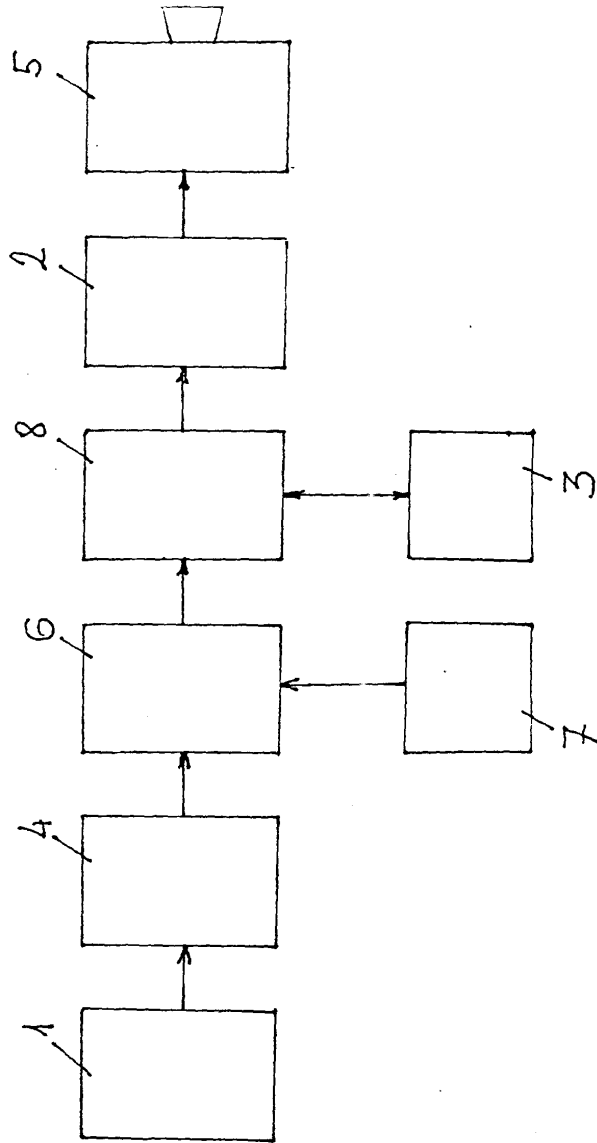
SZABADALMI IGÉNYPONTOK

1. Eljárás szintetizátor(ok) vezérlésére, szótárnélküli szintetizált beszéd vezérléssel quasi időben történő előállítására, azzal *jellemezve*, hogy akusztikai építőelemeket hozunk létre oly módon, hogy minden beszédhangra külön-külön előállítjuk a hangsor kezdő elemeket, meghatározva azok időtartamát, frekvencia összetevőit, intenzitás értékeit, és meghatározzuk egyrészt az egyszerű, vagyis az a, á, o, u, ü, i, é, ö, e, m, n, v, f, sz, s, j, h, l hangok úgynevezett tiszta fázisát reprezentáló elemek

időtartamát, frekvencia összetevőit és intenzitás értékeit, másrészt pedig az összetett hangoknál (p, b, t, d, zs, ny, g, gy, ty, k, cs, c, r) is meghatározzuk a kezdő elemen kívüli további építőelemek időtartamát, frekvencia összetevőit, intenzitás értékeit, ezeken túlmenően pedig hangkapcsoló elemeket hozunk létre, úgyszintén meghatározva azok időtartamát, frekvencia összetevőit és intenzitás értékeit, majd minden beszédhangra hangsorzáró elemet állítunk elő ugyancsak meghatározva azok időtartamát, frekvencia összetevőit és intenzitás értékeit, mindezeket úgy végezzük, hogy 370 akusztikai építőelem álljon rendelkezésünkre adatbázis képzésére, amely adatbázist a továbbiakban minden egyes szintetizált hangsor előállítására úgy használunk, hogy a hangsor betűjelei vagy az azoknak megfelelő kódsorozat útján az adatbázisból a szükséges hangelemeket a betűjelek sorrendjében kiválasztjuk, majd az így előállított monoton hangsorra a hangsorvegi írásjel szerinti önmagában ismert intonációs görbét a monoton hangzásnak megfelelő alaphang frekvencia értékről — és/vagy — irányba megváltoztatva felültetjük és az így előállított jellel a szintetizátort vezéreljük.

2. Vezérlőberendezés az 1. igénypont szerinti eljárás foganatosítására, azzal *jellemezve*, hogy egy beszéd szintetizátorhoz (5) oldhatóan vagy fixen önmagában ismert processzor (2) csatlakozik és az utóbbihoz legalább 12 Kbyte kapacitású memóriaegység (3) van kiválasztó áramkör (8) közbeiktatásával kapcsolva és az említett kiválasztó áramkörhöz összehasonlító egység (6) van kötve, amely egyrészt graféma tárhoz (7), másrészt dekódolóhoz (4) illeszkedik, amely utóbbi adatbefogadó egységhez csatlakozik.

(1 db ábra)



1. ábra

Kiadja az Országos Találmányi Hivatal
A kiadásért felel: Himer Zoltán osztályvezető
Megjelent a Műszaki Könyvkiadó gondozásában

86.0161/3 MSZH Nyomda, Budapest