

A **SPEECH TECHNOLOGY Fejlesztő és Kereskedelmi Korlátolt Felelősségű Társaság** (Hu- 1133 Budapest Véső u. 7. Cg.: 01-09-700204) a melléklet szabadalmi bejelentési kérelem alapján a szabadalmi oltalommal védeni kívánt eljárás kivonatát alábbiak szerint terjesztjük elő:

(A találmány Címe:) **Beszélőorientált hangtani alapegységek létrehozása nyelv-független beszédszintézishez.**

A találmány, egy olyan eljárás, amely alkalmas írott szövegnek, előzetesen megadott emberi beszédhanggá átalakítására, függetlenül a szöveg nyelvétől, illetve a beszélő személyétől.

Az eljárás sematikus áttekintése:

1) Az eljárás alapján a célnyelven alapszöveg választása. Alapszöveg az a szöveg, amelynek felolvasásáról hangfelvétel készül, és a hangfelvétel és az alapszöveg együttesen a beszédszintézis rendszer előállításának kiinduló adata.

2) A beszédjel automatikus egységekre bontása (szegmentálás) a gördülő jellemzők alapján. Valamennyi gördülő jellemzőre el kell végezni a következő vizsgálatot. Szélső értékek keresése a jellemző időfüggvényekben. A szélső értékhelyek közül meg kell jelölni azokat, amelyek adott p t időtartamon belül (25-100 msec) x dB-es pozitív vagy negatív csúcsot képeznek (x értéke 3-6dB között). Szegmens határt ott keletkezik, ahol a gördülő jellemzőknek – adott időintervallumban - közös szélsőérték helye van.

3) Az előző lépések alapján ^{λ} Definiálható egy költségfüggvény, amely arányos a beszédjel két szegmenshatár környezetében vett jellemzőinek távolságával. A költségfüggvényben szerepel a jel logspektruma illetve lineáris predikciós (LPC) együttthatók, továbbá a prozódiaira vonatkozó értékek: hangmagasság (pitch) és időtartam (duration). Ennek a költségfüggvények az értéke a szegmenshatárok pozíciójának ismeretében automatikusan (emberi közreműködés nélkül) meghatározható.

4) A költség mátrix alapján a szegmenshatárokat csoportokra osztjuk. Az egyes elemek alcsoportokbeli gyakoriságai alapján határozzuk meg az ún. szupercsoportokat. A szupercsoportok elemei nem alcsoportok, hanem az egyes elemek. Ilyen módon megengedett, hogy két elem, amelyek elemei egy alcsoportnak más-más szupercsoportba tartoznak. Minden elemet besorolunk egy vagy több alcsoportba és egyetlen szupercsoportba. Az alcsoportok és a szupercsoportok valódi szemléletes fonológiai jelentéssel bírnak. Egy alcsoportba tartozó elemek jelentik az azonos prozódiai kontextusból származó hasonló hangtani elemeket.

5) A gördülő energia függvény alapján a hangfelvétel darabjaihoz a felolvasott szöveg mondatai automatikusan hozzárendelhetők. Az előző lépésekben meghatározott szegmenshatárok típusait szupercsoport és alcsoport indexeit véve egy újabb fonetikai tartalommal bíró diszkrét szimbólumsorozatot kapunk. Ez a szimbólumsorozat felfogható egy csak a beszélőre jellemző hangtani abc-vel írott szövegnek. A tetszőleges szöveg hanggá alakításához meg kell adni azokat a szabályokat, amelyek alkalmazásával egy szöveg ilyen szimbólumsorozatra átírható.

~~Budapest, 2002. július 25.~~



SPEECH TECHNOLOGY KFT.

1139 Bp., Váci út 95.

Asz.: 12757682-2-41

Cgj.: 01-09-702004

Rszsz.: 10300002-20138372

SPEECH TECHNOLOGY KFT

Bejelentő



KÖZZÉTÉTELI PÉLDÁNY

A₁

3 1 7 2 7

P0202518

Szabadalmi Leírás, igénypontokkal

A **SPEECH TECHNOLOGY Fejlesztő és Kereskedelmi Korlátolt Felelősségű Társaság** (Hu- 1133 Budapest Véső u. 7. Cg.: 01-09-700204) a melléklet szabadalmi bejelentési kérelem alapján a szabadalmi oltalommal védeni kívánt eljárás leírását, és igénypontjait az alábbiak szerint terjesztjük elő:

(A találmány címe:) **Beszélőorientált hangtani alapegységek létrehozása nyelv független beszédszintézishez**

A találmány, egy olyan eljárás amely alkalmas írott szövegnek, előzetesen megadott emberi beszédhanggá átalakítására, függetlenül a szöveg nyelvétől, illetve a beszélő személyétől.

I. Az eljárás leírása

I. a. Alapszöveg választás és hangfelvétel készítés

Az eljárás alapján a célnyelven alapszöveg választása. Alapszöveg az a szöveg, amelynek felolvasásáról hangfelvétel készül, és a hangfelvétel és az alapszöveg együttesen a beszédszintézis rendszer előállításának kiinduló adata.

Az alapszöveg választásának szempontjai:

Fonetikailag teljes korpusz szükséges, ehhez előzetes statisztikai vizsgálatok szükségesek. Célszerű különböző intonációjú, és összetettségű mondatokat választani a szövegben. Mivel a beszédszintetizátor az alapszöveg stílusában generál beszédet, célszerű az elérni kívánt stílusú hangmintát alkalmazni (pl.: hírolvasás, elbeszélő stílus stb...).

I. b. Beszédhang jellemzőinek számítása

A hangfelvétel alapján gördülő jellemzők kiszámítása (features). Ezek a következők:

a) gördülő energia, b) turbulencia; c) spektrális állandóság (spectral stationarity).

I.c. Szegmensek automatikus meghatározása

A beszédjel automatikus egységekre bontása (szegmentálás) a gördülő jellemzők alapján. Valamennyi gördülő jellemzőre el kell végezni a következő vizsgálatot. Szélső értékek keresése a jellemző időfüggvényekben. A szélső értékhelyek közül meg kell jelölni azokat, amelyek adott Δt időtartamon belül (25-100 msec) x dB-es pozitív vagy negatív csúcsot képeznek (x értéke 3-6dB között). Szegmens határt ott keletkezik, ahol a gördülő jellemzőknek – adott időintervallumban - közös szélsőérték helye van. A spektrális állandóság kitüntetett jelentőséggel bír a jellemzők között: az intervallumon belül a szegmenshatár pontos helyét ennek a jellemzőnek a szélső értéke jelöli ki. Ahol a spektrális állandóság függvénynek maximum helye van ott hangtani egység közép (A típusú) szegmenshatár, ahol pedig minimum helye van ott hangtani egységeket elválasztó (B típusú) szegmenshatár van. Ezen belül a mindkét típusnak további altípusai vannak attól függően, hogy ott a többi gördülő jellemzőnek maximum vagy minimum helye található. Az automatikus szegmentálás azt jelenti, hogy meghatározzuk az egyes szegmenshatárok helyét és típusát.

I.d. Konkatenációs költség mátrix előállítása

Az előző lépések alapján Definiálható egy költség függvény, amely arányos a beszédjel két szegmenshatár környezetében vett jellemzőinek távolságával. A költségfüggvényben szerepel a jel logspektruma illetve lineáris predikciós (LPC) együtthatók, továbbá a prozódia-ra vonatkozó értékek: hangmagasság (pitch) és időtartam (duration). Ennek a költségfüggvények az értéke a szegmenshatárok pozíciójának ismeretében automatikusan (emberi közreműködés nélkül) meghatározható. Ilyen módon létrehozható egy mátrix amelynek soraiban és oszlopaiban az egyes szegmenshatárok szerepelnek, a mátrix elemei a megfelelő sor és oszlop által kijelölt szegmenshatárookra számított költséget tartalmazzák. Így főatlóban minden elem nulla értékű.

I.e. Hangtani csoportok képzése



A költség mátrix alapján a szegmenshatárokat csoportokra osztjuk, olyan módon, hogy az egy csoportba tartozó elemek közötti konkatenációs költség minimális legyen. Ez végezhető több ismert cluster-képző eljárással is pl.: Kohonen féle Self Organization Map vagy K-mean clustering. Lényeges eltérés azonban az általában végzett csoportképző eljárásoktól, hogy jelen esetben minden elemhez képzünk egy alcsoportot. Az egyes elemek alcsoportokbeli gyakoriságai alapján határozzuk meg az ún. szupercsoportokat. A szupercsoportok elemei nem alcsoportok, hanem az egyes elemek. Ilyen módon megengedett, hogy két elem, amelyek elemei egy alcsoportnak más-más szupercsoportba tartoznak. Minden elemet besorolunk egy vagy több alcsoportba és egyetlen szupercsoportba. Az alcsoportok és a szupercsoportok valódi szemléletes fonológiai jelentéssel bírnak. Egy alcsoportba tartozó elemek jelentik az azonos prozódiai kontextusból származó hasonló hangtani elemeket. Az egy szupercsoportba tartozó elemek pedig azonos hangtani jellemzőkkel (ezáltal jelentéssel) bírnak, de különböző prozódiai kontextusból származhatnak.

I.f. Transzkripció szabályok meghatározása

A gördülő energia függvény alapján a hangfelvétel darabjaihoz a felolvasott szöveg mondatai automatikusan hozzárendelhetők. Az előző lépésekben meghatározott szegmenshatárok típusait szupercsoport és alcsoport indexeit véve egy újabb fonetikai tartalommal bíró diszkrét szimbólumsorozatot kapunk. Ez a szimbólumsorozat felfogható egy csak a beszélőre jellemző hangtani abc-vel írott szövegnek. A tetszőleges szöveg hanggá alakításához meg kell adni azokat a szabályokat, amelyek alkalmazásával egy szöveg ilyen szimbólumsorozatra átirható. Ennek a szabályrendszernek a megadása időigényes feladat, viszont kellő mennyiségű hangminta és megfelelő méretű szegmentumok esetén megadható ilyen szabályrendszer. A szabályok ismeretében a beszédszintézis legbonyolultabb része egyszerű helyettesítési feladattá alakul, ami tetszőleges szövegre egyszerűen végezhető.

II. igénypontok

A fenti szabadalmi leírás alapján a védeni kívánt eljárás fő igénypontjai a következők:



- 1) Az elkészített konkatenációs költség mátrix alapján a szegmenshatárokat csoportokra osztása, olyan módon, hogy az egy csoportba tartozó elemek közötti konkatenációs költség minimális legyen, amelyre jellemző, hogy végezhető több ismert cluster-képző eljárással is pl.: Kohonen féle Self Organization Map vagy K-mean clustering, azonban jelen eljárás lényeges eltérése, hogy az általában végzett csoportképző eljárásokkal szemben, hogy minden elemhez képzünk egy alcsoportot, és az egyes elemek alcsoportokbeli gyakoriságai alapján határozzuk meg az ún. szupercsoportokat.
- 2) Az előzetesen beszerzett hangmintából gördülő jellemzők kiszámítása (features), amelyek: a) gördülő energia, b) turbulencia; c) spektrális állandóság (spectral stationarity), amelyre jellemző, hogy a beszédszintézis alapját képezi.
- 3) A beszédjel automatikus egységekre bontása (segmentálás), amelyet valamennyi gördülő jellemzőre el kell végezni, amelyre jellemző, hogy szélső értékeket keresünk a jellemző időfüggvényekben., és a szélső értékhelyek közül megjelöljük azokat, amelyek adott Δt időtartamon belül (25-100 msec) x dB-es pozitív vagy negatív csúcsot képeznek (x értéke 3-6dB között), ezután megállapítható, hogy szegmens határt ott keletkezik, ahol a gördülő jellemzőknek – adott időintervallumban - közös szélsőérték helye van.
- 4) A konkatenációs költség mátrix előállítás, amellyel definiálható egy költség függvény, amely arányos a beszédjel két szegmenshatár környezetében vett jellemzőinek távolságával, amelyre jellemző, hogy a költségfüggvényben szerepel a jel logspektruma illetve lineáris predikciós (LPC) együtthatók, továbbá a prozódiaáram vonatkozó értékek: hangmagasság (pitch) és időtartam (duration), így a költségfüggvények az értéke a szegmenshatárok pozíciójának ismeretében automatikusan (emberi közreműködés nélkül) meghatározott, amely alapján létrehozható a konkatenációs költség mátrix.
- 5) A transzkipción szabályok meghatározása, amely alapján az előzetesen meghatározott szegmenshatárok típusait szupercsoport és alcsoport indexeit véve egy újabb fonetikai tartalommal bíró diszkrét szimbólumsorozat keletkezik, és megállapítható a hangtani szabályrendszer, amelyre jellemző, hogy a beszédszintézis

legbonyolultabb része egyszerű helyettesítési feladattá alakul, ami tetszőleges szövegre egyszerűen végezhető.

III. A találmány tárgya, és alkalmazási területe:

1) A találmány, lényegét tekintve egy olyan eljárás amely alkalmas írott szövegnek, előzetesen megadott emberi beszédhanggá átalakítására, függetlenül a szöveg nyelvétől, illetve a beszélő személyétől .

A találmány alkalmazható, minden olyan területen, ahol szükség van az írott formában rögzített információk gyors ismertetésére, illetve minden olyan területen ahol nincs mód, és lehetőség az írott formájú információk, szűkebb, illetve tágabb körű ismertetésére.

2) A technika állása, a találmányhoz közel álló megoldások.

A technika állása szerint a védeni kívánt találmány által alkalmazott technikai megoldás nem ismert. A találmányhoz közel álló megoldás egyes bankokban, illetve call centerekben alkalmazott hangközlő technika, azonban az fejlettségében jelentősen elmaradnak a most védeni kívánt találmánytól.

3) A találmánnyal megoldandó feladat

A találmány következtében lehetőség nyílik előre elkészített, bármely nyelven írott szöveg, adat, információnak az előzetesen rögzített beszélő hangján történő reprodukálására. Ennek értelmében alkalmazható minden olyan területen, ahol szükség van gyors információközlésre, illetve minden olyan területen, ahol nincs lehetőség az információ írásos továbbítására, terjesztésére. Alkalmazható a találmány különösen a különböző hívasközpontokban és egyéb területeken, ahol írott, vagy elektronikusan rögzített szövegek emberi beszédhanggá alakítása szükséges.

4) Az oltalom terjedelmét alátámasztó példák.

Az eljárás alkalmazható például a vakok és gyengén látók számára, mivel az alkalmazott rendszer segítségével bármely szöveget meghallgathatnak, és ennek

következtében bármely információ elérhetővé válik számukra (újság, könyv) sőt egyszerű módszerekkel nyelvtanulási lehetőség is nyílik számukra

5) A találmányhoz fűződő előnyös hatások.

A találmány révén gyorsabb, pontosabb információáramlásra nyílik lehetőség, amely a gazdasági élet, és az élet egyéb területein számos, eddig nem ismert előnyös hatást vélt ki. Így lehetőség nyílik az írott információk megismertetésére olyan személyekkel, akik eddig bármely okból el voltak zárva az írott információk megismerésétől.

Igy például a vakok, és gyengén látók számára új, eddig nem ismert lehetőség nyílik bármely írott formában készült szöveg gyors, és pontos megismerésére.

Lehetőség nyílik továbbá a találmány révén központi információ feldolgozó, illetve információ továbbító egységek részére, a rögzített információk gyors, pontos továbbítására, nagy számú felhasználó részére.

Budapest, 2002. július 25.



SPEECH TECHNOLOGY KFT.

1139 Bp., Váci út 95.

Asz.: 12757682-2-41

Cgj.: 01-09-702004

Rszsz.: 10300002-20138372

SPEECH TECHNOLOGY KFT

Bejelentő

