FIATAL MŰSZAKIAK TUDOMÁNYOS ÜLÉSSZAKA XVIII.

Kolozsvár, 2013. március 21–22.

A BESZÉDPRODUKCIÓ AUTOMATIKUS MINŐSÍTÉSE HALLÁSSÉRÜLTEK BESZÉLNI TANÍTÁSÁHOZ

CZAP László, PINTÉR Judit Mária

Abstract

The acoustic speech recognition is used for several reasons and one of them is to teach hearing-impaired people. The evaluation of the pronunciation is based on the spectrum what is compared to the reference sample. The teacher is the one, who does the subjective evaluation. The result of comparison does not equal to the listener's assesment. Therefore the goal is the creation of the quality scale with the involvement of teachers, and to create automatic evaluation as close as possible to the subjective qualification. The examination of the feature extraction methods is necessary for this, because it has to be in harmony with the evaluation. The distance calculation methods, and the dynamic time warping must be examined for the proper fault detection. The investigation of the speech recognition methods also necessary. We will achieve the goal with their combined application.

Key words:

automatic evaluation, speech synthesis, teaching of hearing impaired people, talking head

Összefoglalás

A beszéd az emberek közötti legtermészetesebb információátviteli forma. A beszéd nem csupán fonémák sorozata, hanem fontos a hangsúlyozás a hanglejtés és számos más szupraszegmentális jellemző is. Ezek alapján egyértelmű, hogy a beszéd az emberek legfőbb kommunikációs eszköze, amiért az akusztikus beszédfelismerést igen sok területen és különböző céloknak megfelelően alkalmazzák

Egyik ilyen cél a hallássérültek beszélni tanítását segítő rendszerek fejlesztése. A gyakorlást végző személy aktuális kiejtését általában a spektrum alapján hasonlítják össze a referencia bemondással. Ez az összehasonlítás általában nem egyezik azzal az ítélettel, amit a hallgató az érthetőségről megállapít. A szubjektív értékelést szurdopedagógusok bevonásával a torz beszédre elvégezve megalkotható a minőségi skála. A cél a szubjektív minősítést minél jobban megközelítő automatikus értékelés megalkotása. Ehhez szükség van a lényegkiemelési eljárások elemzésére, a távolság számítás különböző módjainak vizsgálatára, a dinamikus idővetemítés és a gépi beszédfelismerés különböző módszereinek alkotó alkalmazására.

Kulcsszavak:

automatikus minősítés, beszédsszintézis, hallássérültek beszélni tanítása, beszélő fej

1. Bevezetés

Beszédjelek alatt érthetünk akusztikus vagy akár vizuális jeleket is. Akusztikus jelként definiálva a beszéd nem más, mint hangnyomás hullámok keltése, azaz beszédhangok, fonémák kibocsátása. Agyi absztrakció során feleltetjük meg a hangokat fonémáknak (hangok olyan elemi, elvont egysége, amely szavakat különböztet meg egymástól, önálló jelentéssel nem rendelkezik). A beszéd nem csupán fonémák sorozata, hanem fontos a hangsúlyozás a hanglejtés és számos más szupraszegmentális jellemző is. Utóbbi esetben a vizuális jelek mint a gesztikuláció, arcminimika vagy szájmozgás feltérképezésével, megértésével, és rekonstruálásával az audiovizuális beszédfelismerés foglalkozik.

Az elmúlt évtizedekben a gépi beszédfeldolgozás dinamikusan fejlődött, egyre több alkalmazás jelenik meg. A beszédfelismerést igen sok területen és különböző céloknak megfelelően alkalmazzák. A hallássérültek beszélni tanítását és az önálló gyakorlás lehetőségének megteremtését elősegítő szoftverek megalkotásához elengedhetetlen mind az akusztikus mind pedig az audiovizuális beszédfelismerés alkalmazása.

2. A beszédasszisztens rendszer

A „hallássérült” kifejezés gyűjtőfogalom: siketeket, nagyothallókat egyaránt magába foglal, függetlenül hallássérülésük fokától, súlyosságától, hallásállapotuk milyenségétől. Különféle fokozatok vannak a siketség és a nagyothallás között. A hallásmaradványt hallókészülékkel és orvosi rehabilitációval meg lehet őrizni, sőt, bizonyos esetekben fejleszteni is. A teljesen siketek nem rendelkeznek hasznosítható hallásmaradvánnyal, nekik a kommunikáció más formái jelentenek kiutat. A nagyothallókat segíti a beszédben hallásmaradványuk, a siketek – mivel egyáltalán nem hallanak – ebben nem támaszkodhatnak semmiféle önkontrollra. Minden tizedik embernek van valamilyen halláskárosodása Magyarországon. A veleszületett halláskárosodás meglehetősen gyakori, ezer újszülöttről kettőt-hármat érint. A hallássérültek, siketek nehezen, vagy egyáltalán nem értik a helyes beszédet és rosszul beszélnek.

A Debreceni és a Miskolci Egyetem közös projektjének célja a hallássérültek beszédmegértését és beszédtanulását elősegítő olyan összetett rendszer kifejlesztése, amely az eddig ismert módszereknél hatékonyabb támogatást tud nyújtani. A projekt előzménye a Debreceni Egyetemen kidolgozott audiovizuális transzkóder a hang vizualizálására, és a Miskolci Egyetemen kifejlesztett „beszélő fej” [1]. Hallássérültek beszélni tanítását segítheti a helyesen artikuláló virtuális bemondó, amely átlátszó arcával a természetes beszélőnél jobban megmutathatja a hangképzés részleteit (1. ábra). A hangképző szervek jellemző helyzete magyar beszédhangokra megtalálható alapvető munkákban [2, 3].

1. ábra. Fotorealisztikus és transzparens megjelenítés

Hasonló fejlesztések már léteznek a világon, de a projektben elérni kívánt eredménnyel ellentétben nem mutatják meg a szavak helyes artikulációját, így azok nem teszik lehetővé az önálló gyakorlást. A projekt gyakorlatban is hasznosítható célja egy komplex rendszer létrehozása, amely a beszéd folyamatot audiovizuális megjelenítést szolgáltatja, egyrészt a beszéd hangképeinek, másrészt az artikuláció vizuális megjelenítésével, egy oktatási keretrendszerbe foglalva. A technológia részben nyelvfüggetlen, és könnyen adaptálható lesz idegen nyelvekre. A rendszer terjesztése az interneten történik majd, szabad hozzáféréssel alkalmazásban.

3. Automatikus minősítés

A gyakorlást végző személy aktuális kiejtését általában a spektrum alapján hasonlítják össze a referencia bemozdással. Ez az összehasonlítás nem egyezik azzal az ítélettel, amit a hallgató az érthetőségről megállapít. A szubjektív értékelést szurdopedagógusok bevonásával a torz beszédre elvégezve megalkotható a minőségi skála. A szubjektív minősítést minél jobban megközelítő automatikus minősítés megalkotásával lehetőség nyílik az önálló gyakorlásra valamint az előrehaladás nyomon követésére hosszútávon.

A minőségi skála megalkotásához alapmintákat gyűjtünk be a célkorosztályban a beszédminőség felmérésére. A mintákat eltérő fejlettségi fokú személyektől gyűjtjük be és minősítetjük, a szépen beszélő ép hallóktól az alig érthetően beszélő hallássérültekig. Laikus hallgatók (akik ritkán találkoznak hallássérültekkel), valamint a szurdopedagógusok véleménye alapján fogjuk a skála alsó részét meghatározni.

A beszéd felismerőt különböző lényegkiemelési és távolságszámítási módszerekkel tanítjuk, majd elemezzük azokat, és kiválasztjuk a legalkalmasabbakat. Elvégezzük a dinamikus idővetítés és a statisztikai alapú eljárások összehasonlító elemzését is. A helyes és az aktuális kiejtés összehasonlításának eredményét a szubjektív tesztek eredményeivel kell összehangba hozni.

Az automatikus minősítés megalkotása nem új keletű. Hangképzési és beszédzavarok esetén - olyan személyeknél, akiknek gégerák miatt eltávolították a gégéjüket, vagy ajak- és szápadhasadékos gyermekeknél - szignifikáns összefüggések érhetők el a szubjektív és az automatikus minősítés között. A PEAKS (Program for Evaluation and Analysis of all Kinds of Speech Disorders) rögzítő és elemző rendszer hangképzési és beszédzavarok automatikus vagy manuális minősítéséhez. A beteg szabványosított teszteket végezhet el, amiket a rendszer automatikusan minősít. A PEAKS segítséget nyújt a logopédusok munkájához is, az eredményeket, mint második szakvéleményt figyelembe vehetik a diagnózis felállítása során [4,5].

Az automatikus minősítési skála létrehozásához minden egyes beszélő esetén az alábbi műveleteket végezzük el:

- Először a beszéd felismerő és a prozódia modul (a prozódia által hordozott információ felhasználásával javítható a felismerés) alkalmazásával végrehajtjuk a lényegkiemelést;

- az n (összes kiemelt) jellemzők alcsoportját azok a legfontosabb jellemzők fogják alkotni, amelyeknek a legmagasabb a korrelációs célértékük, vagyis a multi-korrelációs és a regresszió analízis alapján adott legjobb lineáris becslések [6];
- a legjobb jellemzők alcsoportjának a felhasználásával pedig az SVN (Support Vector Regression) tanítását végezzük el [7].

A végén a becslült és a célértékek közötti korrelációs érték alapján kiszámítható a becslés pontossága. A kiválasztott jellemzők számát addig növeljük, amíg a becslés pontossága már nem javul tovább.

4. Összefoglaló

A hallássérültek beszédmegértését és beszédtanulását elősegítő összetett rendszer lehetővé teszi az önálló gyakorlást és a hosszú távú fejlődés nyomon követését. Ennek elengedhetetlen feltétele az automatikus minősítés, ami megalkotható beszélő ép hallóktól és az alig érthetően beszélő hallássérültektől származó minták vizsgálata, valamint a szurdopedagógusok szubjektív véleménye alapján. A lényegkiemelési eljárások elemzése és a távolság számítás különböző módjainak kiértékelése elengedhetetlen, hogy az összehasonlítási eredmények összhangba kerüljenek a szubjektív véleményekkel.

5. Köszönetnyilvánítás

A bemutatott kutató munka a TÁMOP-4.2.2.C-11/1/KONV-2012-0002 jelű projekt részeként az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

Irodalomjegyzék

- [1] Czap L.: *'Audio-Visual speech recognition and synthesis'*. PhD Thesis, Budapest University of Technology and Economics. (2004)
- [2] Bolla K.: *'A Phonetic Conspectus of Hungarian'*. Tankönyvkiadó., Budapest. (1995)
- [3] Molnár J.: *'The Map of Hungarian Sounds'*. Tankönyvkiadó, Budapest. (1986)
- [4] Maier A., Haderlein T., Eysholdt U., Rosanowski F., Batliner A., Schuster M., Nöth E.: *'Peaks – A System for the automatic evaluation of voice and speech disorders'*. Speech Communication(2009)
- [5] Maier A., Hönl F., Hacker C., Schuster M., Nöth E.: *'Automatic evaluation of characteristic speech disorders in children with cleft lip and palate'*. Proc. of 11th Int. Conf. on Spoken Language Processing, Brisbane, Australia, 1757-1760 oldal.
- [6] Cohen, J., Cohen, P.: *'Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences'* Lawrence Erlbaum Associates, Hillsdale, New Jersey 1983a.
- [7] Smola, A., Schölkopf, B.: *'A tutorial on support vector regression. Tech.rep.'* Royal Holloway University of London, 1998., nC2-TR-1998-030.

Pintér Judit Mária, PhD hallgató

Munkahely: Miskolci Egyetem

Cím: 3515 Miskolc, Egyetemváros

Telefon: +36 (46) 565-111

E-mail: pinter@mazsola.iit.uni-miskolc.hu

Dr. Czap László, egyetemi docens, dékánhelyettes

Munkahely: Miskolci Egyetem

Cím: 3515 Miskolc, Egyetemváros

Telefon: +36 (46) 565-111

E-mail: czap@uni-miskolc.hu