FIATAL MŰSZAKIAK TUDOMÁNYOS ÜLÉSSZAKA XIX.

Kolozsvár, 2014. március 20–21.

HANGSÚLYDETEKTÁLÁS RELATÍV INTENZITÁS ALAPJÁN

STRESS DETECTION BY RELATIVE INTENSITY

CZAP László⁽¹⁾, PINTÉR Judit Mária⁽²⁾, VICSI Klára⁽³⁾

(1) (2) Miskolci Egyetem, Gépészmérnöki és Informatikai kar, Automatizálási és Infokommunikációs Intézeti Tanszék. Cím: 3515, Magyarország, Miskolc-Egyetemváros; Telefon: +36-46-565-140, czap@uni-miskolc.hu, pinterjm@uni-miskolc.hu

(3) Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszék. Cím: 1117, Magyarország, Budapest; Magyar tudósok körútja 2.; Telefon: +36-1-463-1940, vicsi@tmit.bme.hu

Abstract

Suprasegmental features are fundamental properties of speech. They can improve not only the naturalness of synthesized speech, but the performance of machine speech recognition as well. Sentence and word stress are crucial prosodic features. They usually estimated from the energy of a syllable, but the average energy of vowels is different. Energy of a stressed weak vowel can be lower than that of an unstressed strong voice. We compare the current amplitude of a vowel to its average one to show the stressed or unstressed quality of a syllable. Average energies of vowels are obtained from a speech recognizer trained with voices of hundreds of speakers.

Keywords: stress detection, speech recognition, acoustic feature extraction methods.

Összefoglalás

A szupraszegmentális jellemzők alapvető tulajdonságai a beszédnek. Ezek a tulajdonságok nemcsak a mesterséges beszéd természetességét javítják, hanem a gépi beszédfelismerés hatékonyságát is. A mondathangsúly és a szóhangsúly egyaránt fontos prosódiai jellemző. Ezek a tulajdonságok általában a magánhangzók energiái alapján megbecsülhetők, de az egyes magánhangzók átlagenergiái eltérőek. Egy hangsúlyos gyengébb magánhangzó energiája kisebb lehet, mint egy hangsúlytalan erősebb magánhangzóé. Az aktuális magánhangzó amplitúdóját összehasonlítjuk az átlagával, hogy megtudjuk: az adott szótag hangsúlyos vagy hangsúlytalan-e. A magánhangzók átlagenergiáit egy több száz beszélővel betanított beszédfelismerőből nyertük ki.

Kulcsszavak: hangsúlydetektálás, beszédfelismerés, lényegkiemelési módszerek.

1. Bevezetés

A hagyományos, statisztikai alapú folyamatos beszédfelismerés –, mint ahogyan a rejtett Markov modelleken alapuló beszédfelismerés is – a beszédet valamilyen adott felismerési egység sorozatának tekinti (pl. fonémák, diádok, szótagok, szavak stb.) Statisztikai alapú felismerésnél a lényegkiemelés folyamata után csak a lényeges, csökkentett redundanciájú jellemzők képezik a felismerő bemenetét. A beszéd azonban nem csupán ezen felismerési egységek sorozata, hanem fontos a hangsúlyozás a hanglejtés és számos más szupraszegmentális beszédjellemző is.

2. Szupraszegmentális jellemzők

A szupraszegmentális jellemzők többletinformációval bírnak, amelyek hozzájárulnak a beszédmegértéshez. Segítségével a beszélő érzelmeit, szintaktikai és pragmatikai információt stb. fejezhet ki [1]. Szupraszegmentális jellemzők a beszédtempó, szünet, ritmus és hangszínezet, hangerő, hanglejtés, intonáció és a hangsúly.

Egységhez kapcsolva beszélhetünk szóhangsúlyról, szakaszhangsúlyról, (szószerkezetek esetében) és mondathangsúlyról. A szóhangsúly valamely szó egy szótagjának a kiemelése, megkülönböztetése a többi szótagtól. Hangsúlyozás szempontjából a nyelvek két csoportba sorolhatók, a kötött (a hangsúly mindig a szó egyértelműen azonosított szótagjára esik) és a kötetlen. A magyar nyelv kötött, mivel mindig az első szótagon realizálódik (szerepe tisztán a közlés lényeges elemeinek kiemelésére és a közlés logikai tagolására szorítkozik). Erős érzelmei kifejezésekor a hangsúly a kötött hangsúlyozású nyelvekben is eltolódhat, illetve akár egy szó minden szótagján is megjelenhet. A hangsúly létrehozásában három fő tényező együttesen vagy egyedileg játszik szerepet adott nyelvtől függően, ahol szabályszerűségek figyelhetők meg [2]. Ezek a tényezők:

- az alaphangfrekvencia kiemelkedése a hangsúlyos szótagon;
- a hangsúlyos szótag nagyobb intenzitással való kiejtése;
- és a hangsúlyos szótag magánhangzójának időtartambeli hosszabbodása.

3. Hangsúlydetektálás

3.1 Hangsúlydetektálási módszerek

A hangsúly detektálását általában az energia alapján végzik. Megvizsgálták a hossz, az amplitúdó és a spektrális változások különböző módokon normalizált értékeit [3]. Több esetben mély neurális hálók betanításával és alkalmazásával valósították meg az automatikus hangsúlydetektálást angol nyelvre [4,5]. Angol nyelvet tanító szoftver fejlesztése során, amely a nyelvtanulók által produkált hangsúlymintázatokat vizsgálja, és adott esetben kijavítja azokat, egyesek arra az eredményre jutottak, hogy a hangsúly legmegbízhatóbb jelzése a hossz és az amplitúdó információ kombinációja [5]. A magyar nyelvben a hangsúly, illetve áttételesen a hangsúlyt meghatározó akusztikai-prozódiai jellemzők, az alaphangfrekvencia és az energia alapján lehetséges a szóhatárok detektálása is [6]. Megfelelő lényegkiemelési módszert választva az egyes hangok átlagenergiája nagy különbségeket mutat, ezért a kis energiájú hangsúlyos magánhangzók energiája nem éri el a nagyobb energiájú hangsúlytalan magánhangzókét. Módszerünkben a magánhangzók pillanatnyi energiáját az adott magánhangzók átlagenergiájához viszonyítjuk így mutatjuk ki hangsúlytalan vagy hangsúlyos jellegét.

3.2 Magyar referencia beszédadatbázis

Az magyar referencia beszédadatbázis (MRBA) hanganyagaival tanított és tesztelt diád alapú beszédfelismerőn végeztünk hangsúlydetektálási vizsgálatokat. Az adatbázis olvasott szövegeket tartalmaz, speciális fonetikai elvárásoknak megfelelően került összeállításra és általános felhasználói környezetben (irodák, laboratóriumok, lakások) rögzítették, ezért megfelelően alkalmazható volt a célzott vizsgálatok elvégzéséhez.

Az MRBA összefoglaló műszaki adatai [7]:

- magyar nyelvű, olvasott szövegű, személyi számítógépes környezetben felvett adatbázis,
- 16 bites, 16 kHz-es mintavételezéssel;
- 332 beszélő közvetlenül a számítógépbe rögzített hanganyaga;

- beszélőnként 12 mondat és 12 szó;
- a felvételek többféle mikrofonnal, hangkártyával, PC-kel készültek;
- környezet változó zajosságú irodahelyiség, laboratórium, otthoni környezet;
- az adatbázis teljes anyaga annotált;
- az adatbázis harmada (100 beszélő) kézzel szegmentált és címkézett.

3.3 Különböző lényegkiemelési módszerek vizsgálata

A gépi beszéd felismerésnél akusztikai szinten működik az akusztikai előfeldolgozó egység, amely a beszédjel elemzését, a lényegkiemelést, tömörítést végzi el. Kimenetén az időkeretenkénti lényegi paraméterek (jellemző vektorok) jelennek meg, azaz a digitalizált (beszéd) jelből egy diszkrét idejű, adott dimenziójú lényegvektor-sorozatot alkot. Ezek leggyakrabban Mel-frekvencia kepsztrális együtthatók (MFCC) [8].

1. ábra. A magánhangzók magnitúdója MFCC lényegkiemelési módszer esetén

Az MFCC jellemzőkhöz hozzáadtuk a logaritmikus energia komponenst is, ami révén megvizsgálhattuk az egyes magánhangzók átlagenergiáját. Az 1. ábra diagramja szemlélteti az MRBA beszédatadabázison betanított diád alapú beszéd felismerő rejtett Markov- modelljeiből kinyert egyes magánhangzókra vonatkozó átlagenergiákat. Az energiák alapján látható, hogy egy hangsúlytalan *á* magánhangzó energiája nagyobb lehet, mint egy hangsúlyos *u*, *ü* vagy *i* magánhangzó energiája. Ezért érdemes összehasonlítani az aktuális magánhangzó energiáját a megfelelő diád magánhangzójának átlagenergiájával, amit a betanított HMM állapotaiból nyerünk. A 2. ábra egy, az MRBA adatbázisból kiválasztott mondat hangsúlyozását szemlélteti: *Csaposnak mondja magát, s nem tudja, hogy micsoda a pint.* A HMM modellekből kinyert átlagolt energiát és a példamondat lényegkiemelt fájljából kinyert aktuális energia értékeket normáltuk és az alapján végeztük el az összehasonlítást. A *csapos*, *mondja* és *nem* szavak hangsúlyos volta megmutatkozik a kiértékelésben, de a *pint* hangsúlya elvész az ereszkedő hangerőben. A könnyebb érthetőség érdekében a példamondat hangfájlját elérhetővé tettük az alábbi linken:

http://mzsola.iit.uni-miskolc.hu/DATA/research/audio/pelda_mondat.wav.

2. ábra. A szótagok energiájának relatív intenzitása

4. Összefoglaló

Az eddig elért eredmények azt mutatják, hogy az MFCC energia komponense és a módszerünk további vizsgálatokat igényel hangsúlydetektálási feladatok intenzitás komponensének vizsgálatára. A példamondat kiértékelésében megmutatkoznak a hangsúlyos szótagok. A továbbiakban több lényegkiemelési módszert is tervezünk megvizsgálni és összehasonlítani az eddig elért eredményekkel, majd automatikussá tenni a kiértékelés folyamatát.

5. Köszönetnyilvánítás

A kutató munka a Miskolci Egyetem stratégiai kutatási területén működő Mechatronikai és Logisztikai Kiválósági Központ keretében és 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.

Irodalom

- [1] GÓSY M.: *Fonetika, a beszéd tudománya*, Osiris, Budapest, 2004. pp.182–243.
- [2] KASSAI I.: *Fonetika*, Nemzeti Tankönyvkiadó, Budapest, 1998.
- [3] VAN KUIJK, D., BOVES, L. *Acoustic characteristics of lexical stress in continuous telephone speech*. *Speech Communication*, 27(2), 1999, pp.95–112.
- [4] KUN L., XIAOJUN Q., SHIYIN K., HELEN M.: *Lexical Stress Detection for L2 English Speech Using Deep Belief Networks*, INTERSPEECH, 2013, pp. 1811–1815.
- [5] XIE, H., ANDREAE, P., ZHANG, M., WARREN, P.: *Detecting stress in spoken English using decision trees and support vector machines*. In: *Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation*, Australian Computer Society, Inc. 2004, 145–150.
- [6] VICSI, K., SZASZÁK, GY.: *Automatic Segmentation of Continuous Speech on Word Level Based on Supra-segmental Features*. *International Journal of Speech Technology*, Vol. 8, Num. 4, 2005, pp. 363–70.
- [7] VICSI K., KOCSOR A., TELEKI CS., TÓTH L.: *Hungarian Speech Database for Computer-using Environment in Offices*, II. Magyar Számítógépes Nyelvészet Konferencia kötet, 2004, pp. 315–318.
- [8] S. YOUNG.: *The HTK Book (for HTK Version 3.3)*, 2005.