
Egy pajzsmirigy scintigráfias leletek diktálására alkalmas rendszer technológiai háttere

Kocsor A., Bánhalmi A., Paczolay D.

ÖSSZEFOGLALÁS

Az automatikus beszéd felismerési technológiák jelentős fejlődésével számos adminisztrációt megkövetelő szakmában megfogalmazódott az igény az ún. beszédalapú dokumentálásra. Különösen igaz ez az orvosi vizsgálati eredmények rögzítésére, amely folyamat felgyorsítása különösen nagy jelentőséggel bír. Kisebb és speciális nyelvi tulajdonságokkal rendelkező nyelvekre egyelőre nagyon kevés orvosi diktáló szoftver látott ezidáig napvilágot, amely többek között a nyelvi sajátosságokon túl a magas fejlesztési költségeknek tudható be. Szegeden kifejlesztettünk egy magyar nyelv automatikus felismerésére alkalmas magmodul, amelyre különböző speciális diktáló rendszer építhető. A magmodul tartalmazza az ún. akusztikai modellt, amely alkalmas a magyar nyelv fonémakészletének felismerésére és reprezentatív módon történő modellezésére. A modell felépítésére két egymástól relevánsan eltérő megközelítést alkalmaztunk. Az egyik a beszéd felismerésben közismert és gyakran alkalmazott Rejtett Markov Modell, a másik pedig a Szegeden kifejlesztett új szerkesztőhasználatú szegmentális megközelítés. Mindkét modell felépítéséhez egy nagyméretű (500 beszélő) tartalmazó beszéd korpuszt használtunk fel, majd tesztadatbázisokon összehasonlítottuk a modulok teljesítményét. A magmodul mellé – a kifejlesztett módszerek alkalmazhatóságát bizonyítandó – kiépítettünk egy Windows-os környezetben használható pajzsmirigy scintigráfias leletek diktálására alkalmas nyelvi modult, amelyet 9231 írott pajzsmirigy lelet és több mint 2500 szóalak alapján építettünk fel. Ismertetjük a kifejlesztett pajzsmirigy diktáló-rendszer felépítését, az épített nyelvi és akusztikai modellek technikáját, a modellek hatékonyságát jellemző teszteredményeket, továbbá kitérünk a program felhasználási lehetőségeinek és technikájának különböző aspektusaira is.

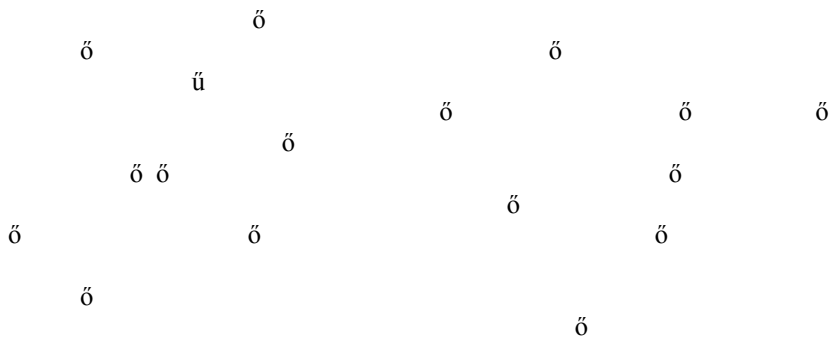
ABSTRACT

Technological background of a speech recognition system for the dictation of thyroid gland medical reports

With the considerable development of speech recognition technologies in several administration-requiring professions the demand for the so called speech-based documentation has grown. This is particularly true in the case of the documentation of medical reports therefore the acceleration of this procedure is of great importance. For smaller languages with special linguistic features few systems for dictating medical reports have been developed so far

which fact can be attributed to linguistic specialties and high development expenses. In Szeged we developed a core module capable of automatic recognition of the Hungarian language on which several domain oriented systems can be built. The core module contains the so called acoustic model, which is suitable for the recognition of the Hungarian phoneme set and representative modeling. For the building of the model we used two significantly different approaches. One is the Hidden Markov Model, well known in speech recognition, the other is the novel stochastic segmental approach developed in Szeged. For the development of both models we used a large speech corpus with 500 speakers, and then the performance of the modules was tested on test databases. To accompany the core module we built a language module (for Windows environment) suitable for the dictation of thyroid gland medical reports in order to justify the applicability of the developed methods. The module was built on 9231 written thyroid medical reports and over 2500 word forms. We present the structure of the developed system for dictating thyroid gland medical reports, the technology of the built language and acoustic models, the test results describing the efficiency of the models, furthermore we mention the different aspects of the application and technology of the software.

BEVEZETÉS



ISMERTEBB DIKTÁLÓ RENDSZEREK

ő ő Smith IBM

ő ő ő

ő ő ő

ő ő ő

Nyers ő ő

ő ő ő
ő ő ő
ő ő ő

KÉT ELTÉRŐ MEGKÖZELÍTÉS A BESZÉDFELISMERÉSBEN

ő

Vapnik

ű

ő

ő

ő

Duda et al.

Huang et al

Becchetti and Ricotti

Kocsor et al.

ő

ő

ű

ő

ő

1. ábra

ő ő

Rabiner and Schafer

Moore

ő

ő

ű

ő

ő

2. ábra

ő ő ű

ő ő

ő

HMM alapú megközelítés

ő

Becchetti and Ricotti

ű

3. ábra

1. ábra

Jellemzőkből álló vektor előállítás a szignálból.

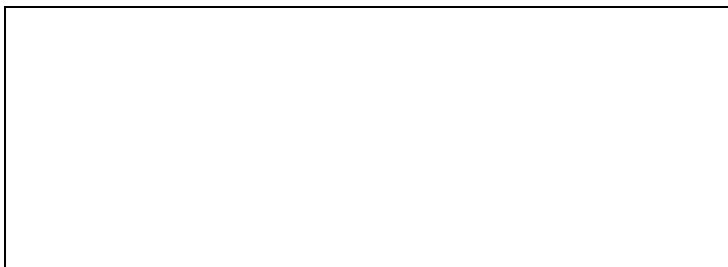
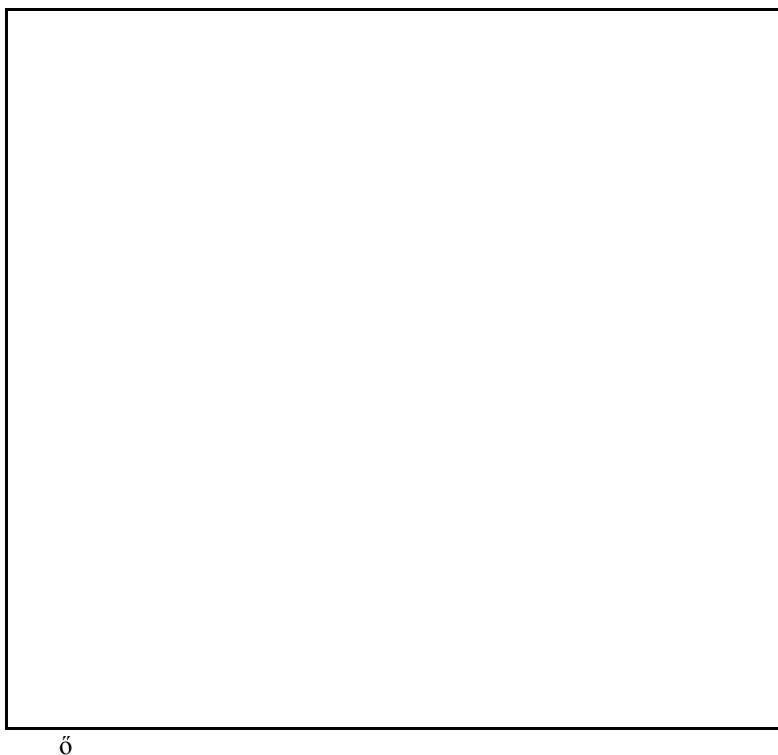


Figure 1: Feature vector extraction from the signal

2. ábra

A szignál fonéma szintű szegmentálása



At the top of the picture the signal is shown while at the bottom the corresponding spectrum can be seen. The vertical lines show the borders of the phoneme segments.

Figure 2: The phoneme level segmentation of the signal

A kétféle irányvonal kísérleti összehasonlítása

ő

ő

Az összehasonlítás metodikája, felhasznált beszédadatbázisok

ő

ő

ő

ő

ű

ő

ő

ő

ő

ő

1. táblázatban

1. táblázat

A tanításhoz és a teszteléshez felhasznált adatbázisok

Tanító adatbázis (1)	Fájlok száma (2)	Fonéma szám (3)	Teszt adatbázis (4)	Fájlok száma	Fonéma szám

Table 1: Database used for training and testing

Database for training(1), The total number of wave files(2), The total number of phonemes(3) Database for testing(4) Numbers(5), MTBA: Hungarian Telephone Speech Database(6) Our special speech database (7) MTBA cities(8), Not segmented(9), MTBA test(8),

ő

ű

ő

ű

Fonéma felismerés hatékonysága

ő

ő

ű ű

ő ő

A PAJZSMIRIGY LELETEK DIKTÁLÁSÁRA ALKALMAS RENDSZER FELÉPÍTÉSE

A nyelvi modell tanításához használt adatbázisok

ő

ő

ő

ő

ő ő

ő

ő

ő

Folyamatos beszédfelismerés gyakorlati megoldása

ő

ű

ő

ő

Rabiner and Juang

ő

ű

ű

5. táblázat

A HMM alapú folyamatos felismerő hatékonysága mondatokon szó N-gram használatakor

Nyelvi modell (1)	correction	accuracy	insertion	deletion	replace
ε					
ε					
ε					
ε					
ε					
ε					
ε					
ε					
ε					

Table 5: The efficiency of the HMM-based continuous speech recognizer on sentences using word N-grams

Language model(1) A and B signs the language model from Table 4.

6. táblázat

A HMM alapú folyamatos felismerő hatékonysága mondatokon szó N-gram és hasonulás használatakor

Nyelvi modell	correction	accuracy	insertion	deletion	replace
ε					
ε					
ε					
ε					
ε					
ε					
ε					
ε					

Table 6: The efficiency of the HMM-based continuous speech recognizer on sentences using word N-grams and assimilation rules

Language model (1) A and B denotes the language model from Table 4, here h means that assimilation rules were used.

7. táblázat

7. táblázat

A HMM alapú felismerő hatékonysága mondatokon különböző nyelvi modellek használata esetén

Nyelvi modell (1)	correction	accuracy	insertion	deletion	replace
ε					
ε					
ε					
ε					
ε					
ε					

Table 7: The efficiency of the HMM-based continous speech recognizer on sentences using different language models

Language model(1) C: denotes the word by word language model(2) D: denotes the model using class N-grams based on MSD codes(3) h means that assimilation rules were used (4) "3 gramm" means that word 3-gram model was used

ő

ő

GYAKORLATI ALKALMAZHATÓSÁG

ő

ő

ő

ő

ő

ő

ő

ő

ő

ő

ú

ő

ő

ú

ő

ő

ő

ő

ő

ő

ő

ő

**EREDMÉNYEK ÖSSZEFOGLALÁSA, TOVÁBBI KUTATÁSI ÉS
FEJLESZTÉSI TERVEK**

ű ő ő ő

Felföldi et al.

ő

ő

ű

ű

ő

ű

ő

ő

ő

ő

ő

ű

ő

IRODALOM

Corresponding author

Kocsor András

*Research Group on Artificial Intelligence of the address: Hungarian Academy of
Sciences and University of Szeged
H-6720 Szeged, Aradi vértanúk tere 1.*