Normalizacija dužine vokalnog trakta eksperimenti

Nikša M. Jakovljević, *Member, IEEE*, Marko B. Janev, and Dragiša M. Mišković

***Sadržaj* — U ovom radu je predstavljeno nekoliko modifikacija standardne procedure za normalizaciju dužine vokalnog trakta bazirne na kriterijumu maksimalne verodostojnosti. Modifikacije podrazumevaju izmenjeni set korišćenih obeležja pri estimaciji koeficijenata skaliranja, uvođenje iterativne procedure i implementaciju metoda robustne statistike. Ove modifikacije su rezultovale relativnim unapređenjem performansi od 16-20% na standardnoj Srpskoj telefonskoj bazi u odnosu na performanse sistema koji koristi proceduru normalizacije dužine vokalnog trakta zasnovanu na maksimizaciji verodostojnosti.**

***Ključne reči* — automatsko prepoznavanje govora, normalizacija dužine vokalnog trakta, robustna statistika.**

I. Uvod

Savremeni sistemi za prepoznavanje govora (ASR *Automatic Speech Recognition*) su uglavnom bazirani na skrivenim Markovljevim modelima (HMM *Hidden Markov Models*) i mešavinama Gausovih raspodela (GMM *Gaussian Mixture Models*). Ovi sistemi su izuzetno osetljivi na razlike koje postoje između uslova u kojima su snimani fajlovi za obuku i testiranje, što za rezultat ima degradaciju performansi (tačnosti prepoznavanja) sistema [1], [2]. Jedna od najčešće korišćenih metoda kojom se redukuju varijacije u spektru koje su posledica razlika u obliku vokalnog trakta različitih govornika jeste normalizacija dužine vokalnog trakta (VTN *Vocal Tract length Normalization*) [1], [3]-[7]. VTN podrazumeva skaliranje frekvencijske ose spektra prilikom ekstrakcije vektora obeležja, na osnovu unapred zadate funkcije čiji je parametar korelisan sa dužinom vokalnog trakta pojedinačnog govornika. Uobičajeni naziv parametra funkcije skaliranja je VTN koeficijent.

U ovom radu su navedeni rezultati istraživanja primene VTN tehnika u cilju unapređenja performansi sistema za prepoznavanje govora koji se razvija na Fakultetu tehničkih nauka u Novom Sadu. Inicijalni rezultati ovog istraživanja su izloženi u [8], dok su ovde navedena dodatna poboljšanja koja su posledica izbora drugačijeg skupa obeležja na osnovu kojeg se vrši estimacija VTN koeficijenata i uvođenja iterativne procedure estimacije VTN koeficijenata u fazi obuke.

Izlaganje započinje kratkim opisom osnovnih karakteristika sistema (odeljak II) nakon čega sledi pregled varijanata kriterijuma na osnovu kog se estimira VTN koeficijent (odeljak III). U odeljku IV su dati rezultati eksperimenata u kojima su evaluirani efekti korišćenja novog seta obeležja i iterativne procedure pri estimaciji VTN koeficijenata. Rad se završava zaključkom u kom su istaknuti najvažniji rezultati i navedene potencijalne buduće smernice istraživanja.

II. Opis sistema

Vektor obeležja čini 12 Mel frekvencijskih kepstralnih koeficijenata (MFCC), normalizovana energija i njihovi prvi izvodi u vremenu. Standardna procedura za izdvajanje MFCC koeficijenata je modifikovana uključivanjem segmenta koji vrši skaliranje frekvencijske ose spektra primenom deo po deo linearne funkcije.

Za potrebe obuke i testiranja iskorišćena je srpska telefonska baza koja je formirana u skladu sa SpeechDat(E) standardom [9]. Da bi se obezbedila dovoljna količina podataka za estimaciju VTN koeficijenata u fazi obuke sistema, iz dela baze za obuku izbačeni su iskazi govornika za koje ne postoji bar 30 s snimljenog govora. Skup za obuku sada čine iskazi 699 govornika (376 muškaraca i 323 žene) ukupnog trajanja 11.5 sati (nisu uračunati oštećeni segmenti i segmenti sa tišinom). Pošto u većini realnih telefonskih aplikacija celokupan dijalog sa mašinom traje svega nekoliko sekundi, standardni test skup nije modifikovan pa sadrži i iskaze govornika kod kojih je ukupno trajanje svega nekoliko sekundi. Ovo u određenoj meri predstavlja problem pošto u test skupu nije obezbeđena dovoljna količina podataka za adekvatnu estimaciju. Test skup čine iskazi 184 govornika (107 muškaraca i 77 žena) ukupnog trajanja 18 minuta.

ASR sistem predstavljen u ovom radu je baziran na HMM i GMM. Umesto najčešće korišćenim diagonalnim kovarijansnim matricama, Gausove respodele su opisane punim kovarijansnim matricama. Procedura obuke je bazirana na modifikovanom *k-means* alogoritmu čiji je detaljanan opis dat u [10]. Jedinica modelovanja je trifon (fonem u kontekstu). Broj HMM stanja kojima se modeluje trifon je srazmeran prosečnom trajanju fonema čiju kontekstno zavisnu varijantu predstavlja, te naglašeni vokali imaju 5 stanja, frikativi 4, nazali 3 itd. Broj Gausovih raspodela po jednom HMM stanju prvenstveno

Ovaj rad je podržan od strane Ministarstva za nauku i tehnološki razvoj republike Srbije u okviru projekta „Govorna komunikacija čovek mašina“ (TR11001).

N. M. Jakovljević, Fakultet tehničkih nauka u Novom Sadu, Srbija (telefon: 381-21-4852521; e-mail: jakovnik@uns.ac.rs).

M. B. Janev, Fakultet tehničkih nauka u Novom Sadu, Srbija (telefon: 381-21-4852521; e-mail: marko.janev@alfanum.co.rs).

D. M. Mišković, Fakultet tehničkih nauka u Novom Sadu, Srbija (telefon: 381-21-4852521; e-mail: dragisa.miskovic@alfanum.co.rs).

zavisi od broja opservacija koje su na raspolaganju za obuku datog trifona i varijabilnosti fonema koja je određena heuristički. Maksimalan broj Gausovih raspodela po jednom HMM stanju je ograničen na 6. Da bi se obezbedila koliko toliko pouzdana estimacija parametara Gausove raspodele, minimalan broj opservacija po jednoj raspodeli iznosti 351, tako da se broj od 6 raspodela po jednom stanju retko dostiže.

Pri obuci sistema koji koriste VTN, pored parametara HMM modela estimiraju se i VTN koeficijenti za svakog od govornika koji se nalazi u skupu za obuku. Združena estimacija parametara HMM modela i VTN koeficijenata je prilično komplikovana stoga se obično prvo estimiraju VTN koeficijenti za svakog od govornika, a potom parametri HMM modela koji koriste normalizovane vektore obeležja (vektore obeležja kod kojih je izvršeno skaliranje frekvencijske ose na osnovu prethodno estimiranih VTN koeficijenata).

U slučaju da je VTN primenjen u fazi obuke, tada ga je neophodno primeniti i u fazi testiranja, jer bi u suprotnom došlo do degradacije performansi [1]. Pošto je cilj rada ispitivanje novih načina estimacije VTN koeficijenata, a da bi se eliminsala zavisnost od načina modelovanja u test fazi primenjena je strategija višestrukog prolaza, koja podrazumeva sledeća tri koraka: *i*) Formiranje inicijalnih transkripcija na nenormalizovanoj sekvenci obeležja; *ii*) estimacija VTN koeficijenata korišćenjem inicijalnih transkripcija; *iii*) Finalno prepoznavanje na sekvenci normalizovanih vektora obeležja.

Za potrebe formiranja inicijalnih transkripcija pogodnije je koristiti sistem za prepoznavanje koji je obučen na ne-normalizovanim obeležjima pošto se takva obeležja nalaze i na njegovom ulazu. Pored toga treba imati u vidu da za razliku od estimacije VTN koeficijenata u faze obuke, estimacija u fazi testiranja se vrši na osnovu transkripcija koje ne moraju biti tačne (obično i nisu). Prethodni eksperimenti su pokazali da u slučaju male vrednosti greške na nivou reči (WER Word Error Rate) ovo ne predstavlja ozbiljan problem [8].

III. Estimacija VTN koeficijenata

Za vrednost VTN koeficijenta (α_g) datog govornika (g) uzima se ona vrednost koja maksimizuje verodostojnost svih normalizovanih opservacija (X_g^{α}) koji pripadaju tom govorniku na modelu univerzalnog govornika (λ_u), što se formalno matematički može zapisati u obliku:

$$\alpha_g = \arg\max_{\alpha} P(X_g^{\alpha} \mid \lambda_u) \qquad (1)$$

Da bi se modelovao univerzalni govornik koristi se skup HMM modela sa po jednom Gausovom raspodelom po stanju. Ako bi se koristilo više Gausovih raspodela po HMM stanju postoji mogućnost da λ_u nauči karakteristike i jedne grupe govornika, što bi za posledicu imalo prilagođenje na pojedine grupe govornika umesto na univerzalnog govornika. Ovaj princip estimacije VTN koeficijenata u daljem tekstu ovog rada nosi oznaku M0.

Po našem mišljenju, jedna od mana standardne procedure za estimaciju VTN koeficijentata izložene u [6] je da favorizuje duže i frekventnije foneme, što nas je motivisalo da pokušamo da izmenimo kriterijum definisan izrazom (1). Detaljaniji opis modifikacija kriterijuma estimacije parametara naveden je u [8], stoga će u ovom radu biti izložen u najkraćim crtama.

U tabeli 1 je dat uprošćen prikaz razmotrenih metoda estimacije VTN koeficijenata. Da bi se eliminisao uticaj trajanja fonema umesto prosečne vrednosti verodostojnosti po vektoru obeležja može se uzeti prosečna vrednost verodostojnosti po instanci fonema računata kao uzoračka sredina (metode M1 i M2). Verodostojnost po instanci fonema je definisana kao uzoračka sredina u slučaju metode M1 odnosno uzoračka mediana u slučaju metode M2. Da bi se eliminisao i uticaj frekvencije pojavljivanja fonema može se uzeti prosečna vrednost verodostojnosti po fonemu računata kao uzoračka sredina (metode M3 i M4) odnosno kao uzoračka mediana (metode M5 i M6). Verodostojnost fonema se može definisati kao uzoračka sredina verodostojnosti njemu pripadajućih vektora obeležja, što je slučaj kod metoda označenih sa M3 i M5, odnosno kao uzoračka mediana, što slučaj kod metoda označenih sa M4 i M6. Nedostatak predloženih alternativnih metoda je da ne garantuju povećanje verodostojnosti sekvence reči, što je kriterijum na osnovu kog se donosi odluka pri dekodovanju.

TABELA 1: METODI ESTIMACIJE VTN KOEFICIJENATA.

		Uzoračka sredina	Uzoračka mediana
Vektor obeležja		M0	
Instanca fonema	Uzoračka sredina	M1	
	Uzoračka mediana	M2	
Fonem	Uzoračka sredina	M3	M5
	Uzoračka mediana	M4	M6

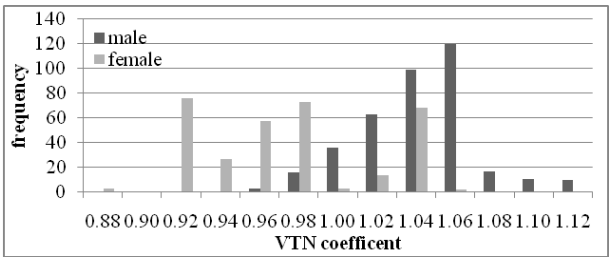
IV. Rezultati

A. *Selekcija obeležja*

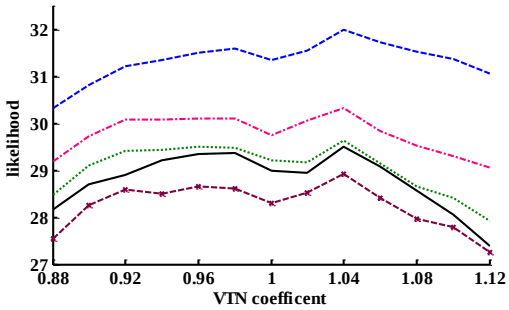
Standardna obeležja koja se koriste pri estimaciji VTN koeficijenata su ista ona obeležja koja se koriste i u samom procesu prepoznavanja. Ovakav pristup je opravdan činjenicom da VTN koeficijenti treba da smanje varijabilnost unutar jednog klastera kako za statička tako i za dinamička obeležja, iako teorijska postavka VTN podrazumeva unifikaciju obvojnice spektra (statička obeležja).

Na Sl 1. je data raspodela VTN koeficijenata dobijenih za govornike koji se nalaze u skupu za obuku u slučaju da se koristi metoda M0. Slične raspodele su dobijene i za druge razmatrane metode estimacije VTN koficijenata, ali zbog prostora koji je na raspolaganju u ovom radu raspodela je prikazana samo za jednu metodu. Uglavnom su dobijeni očekivani rezultati tj. veliki broj ženskih govornika je dobio vrednosti koje su manje od 1, a veliki broj muških govornika vrednosti koje su veće od 1. Ono što je bilo sumnjivo jeste postojanje velikog broja ženskih govornika za koje je estimirana vrednost VTN koeficijenta

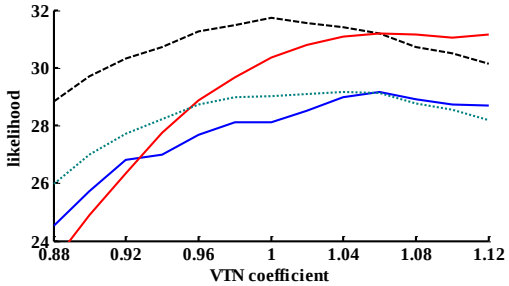
od 1.04. Analiza uzroka koji su doveli do ove pojave je između ostalog (pogrešna oznaka pola govornika, govornica sa nešto tamnijom bojom glasa) uključila i analizu krivih odlučivanja. Kriva odlučivanja je kriva zavisnosti prosečne vrednosti verodostojnosti od VTN koeficijenata, a estimirana vrednost VTN koeficijenta je ona vrednost za koju ova kriva ima maksimalnu vrednost. Kriva odlučivanja kod većine ženskih govornika sa estimiranom vrednošću VTN koeficijenta jednakoj 1.04 su bimodalne (sa dva lokalna maksimuma kao na Sl. 2) umesto očekivanih i dominatnijih unimodalnih (samo jedan lokalni maksimum kao na Sl. 3).



Sl. 1: Histogram VTN koeficijenata za govornike iz skupa za obuku u slučaju da se koristi metod estimacije M0



Sl. 2: Primeri nekoliko bimodalnih krivih odlučivanja, koje su dominantne za ženske govornike kod kojih je estimirana vrednost VTN koeficijenta jednaka 1.04.



Sl. 3: Primeri nekoliko unimodalnih krivih odlučivanja, koje su dominantne za većinu govornika

Izostavljanje dinamičkih obeležja pri estimaciji VTN koeficijenata rezultovala je unimodalnim oblikom krive odlučivanja za sve govornike Vrednosti WER na standardnom test skupu za sve metode estimacije VTN koeficijenata predstavljene su u tabeli 2. Kao što se na osnovu rezultata datih u tabeli 2 može videti izostavljanje dinamičkih obeležja (vrsta sa oznakom S u tabeli 2) rezultovalo je smanjenjem WER za većinu metoda estimacije VTN koeficijenta. U slučaju metode sa oznakom M6 dobijeno je suprotno, što je posledica male efikasnosti uzoračke mediane u test fazi (mala količina podataka koji su na raspolaganju). Efikasnost metode sa oznakom M5 je ista kao i u slučaju M6, ali dobijeni rezultat nije suprotan rezultatima koji su dobijeni kod većine drugih metoda estimacije VTN koeficijenta.

Iako unapređenje performansi kada se pri estimaciji VTN koeficijenata koriste samo statička obeležja nije značajno, dobijena konzistentnost u oblicima krivih odlučivanja (unimodalne krive) je dalje eksperimente ograničila samo na metode u kojima se koriste isključivo statička obeležja.

TABELA 2: VREDNOSTI WER ZA RAZLIČITE METODE ESTIMACIJE VTN KOEFICIJENATA U ZAVISNOSTI DA LI SE KORISTE SAMO STATIČKA (S) ILI I STATIČKA I DINAMIČKA OBLEŽJA (S+D)

	M0	***M1***	***M2***	***M3***	***M4***	***M5***	***M6***
S	4.3	4.5	4.4	4.1	4.4	4.4	4.6
S+D	4.5	4.7	4.8	4.7	4.4	4.9	4.5

B. Broj iteracija u proceduri obuke

Motivacija da se ispita potreba za iterativnom procedurom estimacije VTN koeficijenata u fazi obuke zasniva se na činjenici da su inicijalni testovi pokazali značajne razlike u zavisnosti od toga da li je model univerzalnog govornika obučen na nenormalizovanim ili normalizovanim vektorima obeležja (vidi tabelu 3).

TABELA 3: VREDNOSTI WER ZA RAZLIČITE METODE ESTIMACIJE VTN KOEFICIJENATA U ZAVISNOSTI OD TOGA DA LI JE MODEL UNIVERZALNOG GOVORNIKA OBUČEN NA NORMALIZOVANIM (NORM) ODNOSNO NENORMALIZOVANOM (NENOR) OBELEŽJIMA

	M0	***M1***	***M2***	***M3***	***M4***	***M5***	***M6***
NORM	4.3	4.5	4.4	4.1	4.4	4.4	4.6
NENOR	5.1	5.3	5.1	4.8	4.6	5.4	4.6

Iterativna procedura estimacije VTN koeficijenata u fazi obuke se može svesti na sledeća 3 koraka:

1. Model univerzalnog govornika u *k*-tom koraku iteracije (λ^k_u) se obučava na normalizovanim vektorima obeležja. Normalizacija vektora obeležja se vrši na osnovu vrednosti VTN koeficijenata estimiranih u prethodnom koraku ($\boldsymbol{\alpha}_{k-1}$). U slučaju da je u pitanju prvi korak ($k = 1$) vrednosti VTN koeficijenata su jedinične.
2. Za svakog od govornika u skupu za obuku se estimiraju nove vrednosti VTN koeficijenata ($\boldsymbol{\alpha}_k$) korišćenjem modela univerzalnog govornika λ^k_u.
3. Ponavljati korake 1 i 2 sve dok broj promena vrednosti VTN koeficijenata ne postane dovoljno mala. U ovom radu usvojeno je da se ova procedura završava kada vrednost promene VTN koeficijenta postanje manja od polovine koraka VTN koeficijenta što iznosi 0.01.

Procenat govornika kod kojih dolazi do promene vrednosti VTN koeficijenata opada iz iteracije u iteraciju, bez obzira na metodu estimacije, da bi u 4. iteraciji opao na oko 20%. Prosečna vrednost apsolutne promene vrednosti VTN koeficijenta takođe opada i već nakon 3. iteracije se dostiže postavljeni uslov za prekid predložene iterativne procedure, dok su u 4. iteraciji promene VTN koeficijenata zanemarljive u odnosu na vrednosti dobijene u prethodnom koraku reda $5 \cdot 10^{-3}$.

U [3] je predložena slična iterativna procedura estimacije koja je imala za cilj da onemogući prilagođenje modela na grupu govornika, što je ovde prevaziđeno

usvajanjem modela sa po jednom Gausovom raspodelom po stanju za potrebe estimacije VTN koeficijenata.

TABELA 5: VREDNOST WER ZA RAZLIČITE KORAKE ITERACIJE ZA RAZLIČITE METODE ESTIMACIJE VTN KOEFICIJENATA

	M0	*M1*	*M2*	*M3*	*M4*	*M5*	*M6*
1	4.3	4.5	4.4	4.1	4.4	4.4	4.6
2	4.2	3.8	3.7	4.1	4.1	3.9	4.4
3	4.2	3.8	3.8	4.0	4.0	3.8	4.5
4	4.0	3.8	3.5	4.1	4.1	3.8	4.5

TABELA 6: PERFORMANSE ANALIZIRANIH SISTEMA I NJIHOVO RELATIVNO UNAPREĐENJE U ODNOSU NA REFERENTNE SISTEME

	M0	*M1*	*M2*	*M3*	*M4*	*M5*	*M6*
WER	4.0	3.8	3.5	4.1	4.0	3.8	4.4
RI1	28.0	31.0	36.0	25.0	26.0	31.0	20.0
RI2	25.0	29.0	34.0	22.0	24.0	28.0	16.0
RI3	11.1	16.4	23.1	10.0	11.1	16.0	2.2

Rezultati prepoznavanja nakon svake iteracije (vidi tabelu 5) su potvrdili validnost postavljenog uslova za završetak predložene procedure. Pojedine metode (M1, M2) su dale značajno relativno unapređenje performansi od (17%, 20%) u odnosu na slučaj kad se ne primenjuje iterativna procedura estimacije. Sa druge strane za druge metode estimacije to unapređenje je zanemarljivo, za šta nismo uspeli da nađemo odgovarajuće objašnjenje.

U tabeli 6 je dat uporedni prikaz relativnog unapređenja performansi analiziranih sistema u odnosu na referentne sisteme. Razmotrena su 3 referentna sistema. Prvi referentni sistem je sistem nezavisan od govornika (WER=5.9%), drugi je sistem zavisan od pola govornika (WER=5.3%) i treći je sistem koji koristi nemodifikovanu VTN proceduru opisanu u [6] (WER=4.5%). U tabeli 7 je pored vrednosti relativnog unapređenja performansi navedene i vrednosti za WER za svaku od analiziranih metoda. Kao što se iz priloženog može videti predložene jednostavne modifikacije su doprinele značajnom unapređenju performansi.

V. ZAKLJUČAK

Rezultati eksperimenata prikazanih u ovom radu ukazuju na to da jednostavne modifikacije standardne VTN procedure zasnovane na principu maksimalne verodostojnosti mogu značajno da unaprede performanse sistema, u odnosu na sistem gde se ne primenjuje VTN do 30% odnosno standardnu VTN proceduru do 20%.

Eliminacija uticaja trajanja fonema pri estimaciji VTN koeficijenata se pokazala prilično uspešnom, što je vrlo verovatno posledica smanjenja uticaja grešaka koje postoje u inicijalnim transkripcijama kao i činjenice da se i dalje favorizuju češći fonemi tj. vokali koji nose najviše informacija o dužini vokalnog trakta.

Zanemarivanje učestanosti fonema pri estimaciji VTN koeficijenata se pokazala loše, što objašnjavamo malim brojem instanci koje su na raspolaganju. Formiranje test skupa u kojem bi se obezbedilo bar 30 s govora za svakog od govornika, rezultovalo bi manjim varijacijama tj. većom efikasnošću korišćenih estimatora, ali takvi sistemi ne bi bili funkcionalni u praktičnim aplikacijama gde to nije moguće obezbediti stoga se nije pristupilo proširivanju test skupa.

Eliminacija dinamičkih obeležja pri estimaciji VTN koeficijenata je doprinela konzistentnosti krivih odlučivanja i neznatnom unapređenju performansi. (Pri prepoznavanju se koriste i statička i dinamička obeležja). Ovo objašnjavamo činjenicom da dve različite funkcije mogu da imaju isti prvi izvod. Ovo ne predstavlja toliko radikalan korak pošto VTN metode zasnovane na položajima formanata ne koriste dinamička obeležja.

Dodatno unapređenje performansi koje je za pojedine metode značajno, postignuto je usvajanjem iterativne procedure estimacije VTN koeficijenata. što objašnjavamo dodatnim smanjivanjem varijabilnosti unutar jedne klase.

Budući koraci podrazumevaju implementaciju ovih modifikacija u procedure za brzu estimaciju VTN koeficijenata, kao i eksperimentisanje na nekim od standardnih baza za prepoznavanje govora kao što su TIMIT ili SwitchBoard.

LITERATURA

[1] S. Molau, "Normalization in the Acoustic Feature Space for Improved Speech Recognition," Ph.D. dissertation RWTH Aachen, Germany, 2003

[2] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, C. Wellekens, "Impact of Variabilities on Speech Recognition." in *Proc. SPECOM2006*. Moscow, Russia, 2006.

[3] L. Lee, R. Rose, "Speaker Normalization using Efficient Frequency Warping Procedures", in *Proc. ICASSP-96*, pp. 353-356, Atlanta, GA, 1996.

[4] E. Gouvea, R. Stern, "Speaker Normalization through Formant Based Warping of the Frequency Scale", in *Proc. EUROSPEECH-97*, pp. 1139-1142, Rhodes, Greece, 1997.

[5] L. Uebel, P. Woodland, "An Investigation into Vocal Tract Length Normalization", in *Proc. EUROSPEECH-99*, pp 2527-2530, Budapest, Hungary 1999.

[6] L. Welling, S. Kanthak, H. Ney, "Improved Methods for Vocal Tract Normalization", in *Proc. ICASSP-99*, Phoenix, AZ, 1999.

[7] A. Miguel, E. Lleida, R. Rose, L. Buera, O. Saz, A. Ortega, "Capturing Local Variability for Speaker Normalization in Speech Recognition," *IEEE Trans. on Audio, Speech and Language Processing*, pp: 578-593, March 2008.

[8] N. Jakovljevic, M. Secujski, V. Delic, "Vocal Tract Length Normalization Strategy based on Maximum Likelihood Criterion," in *Proc. EUROCON-09*, St. Petersburg, Russia, 2009.

[9] N. Đurić, D. Pekar, Lj. Jovanov, "Struktura srpske SpeechDat(E) govorne baze snimljene preko fiksne telefonske mreže," in *Proc. DOGS 2002*, str: 57-60, Bečej, Serbia, 2002.

[10] M. Janev, D. Pekar, N. Jakovljević, „Poređenje sistema za prepoznavanje govora na srpskom jeziku baziranih na punim i dijagonalnim kovarijansnim matricam," in *Proc. TELFOR2007*, pp: 342-345, Belgrade, Serbia, 2007.

ABSTRACT

The paper presents several modifications of the standard vocal tract length normalization procedure based on the maximum likelihood criterion. The modifications include: a different set of features, application of an iterative procedure for VTN estimation as well as implementation of robust statistical methods. The result of these modifications is a relative improvement of approximately 16% on a Serbian speech corpus of telephone quality

Experiments in Vocal Tract Length Normalization

N. M. Jakovljević, M. B. Janev. D. M. Mišković