UDC: 004.434:004.8 INFO M: str. 19-22

EMOCIJE U ARAPSKOM PRIRODNOM I SINTETIZOVANOM GOVORU
EMOTIONS IN ARABIC NATURAL AND SYNTHESISED SPEECH

Mohammad Al-Abbushi, Vlado Delić

REZIME: Opšti cilj ovog rada je istraživanje mogućnosti uključivanja emocija u sintetizovan govor. Konkretan cilj bio je da se uporedi snimljeni glas govornika za nekoliko određenih rečenica na arapskom jeziku, sa i bez emocija, i da se uporedi sa sintetizovanim govorom pomoću TTS softvera za arapski. U eksperimentalnom delu rada snimljene su odabrane rečenice sa neutralnim govorom i sa odgovarajućim emocijama. Potom su one poređene međusobno i u odnosu na sintetizovani govor istih rečenica. Pomoću PRAAT-a analizirana su obeležja govora kao što su osnovna frekvencija, brzina govora i intenzitet. Izvršena je audio-vizuelna analiza snimljenih rečenica sa i bez emocija. Prezentovana je analiza pet emocija u prirodnom i sintetizovanom govoru: bes, radost, tuga, strah i iznenađenje. Rad pokazuje razlike u emotivnom i neutralnom govoru koje bi trebalo da se izraze i u sintetizovanom govoru. U radu su takođe predstavljene određene specifičnosti tekstova na arapskom jeziku koje su bitne za proces pretvaranja teksta u govor.
KLJUČNE REČI: Emocije u glasu, intenzitet, pitch, F0, trajanje.
ABSTRACT: The general aim of this paper is the research of posibilities of including emotions into synthesized speech. The goal was to compare the recorded voice of human speakers for several selected utterances in Arabic language, either with or without emotions, as well as to compare human utterances to synthesized speech obtained from an Arabic TTS system. In the experimental part of the paper several sentences are recorded with neutral utterances as well as with corresponding emotions. Then they were compared with each other and with synthesized speech of the same sentences. Speech features such as F0, duration and intensity were analyzed using PRAAT. Audio-visual analysis of recorded sentences with and without emotions has been conducted. The analysis of five emotions in natural and synthesized speech was presented: anger, joy, sadness, fear and surprise. The paper shows the differences in emotional and neutral speech that should be expressed in the synthesized speech as well. Moreover, some peculiarities of Arabic texts that are significant in the TTS process are also presented in the paper.
KEY WORDS: Emotions in voice, pitch F0, intensity, duration.

1. UVOD

Govor omogućava najprirodniju i najpotpuniju komunikaciju među ljudima, putem koga čovek izražava svoje misli, ideje, poruke i stavove, namere, kao i emocionalna stanja. Govornu poruku i njen pun smisao ne čini samo ono "šta" je rečeno već i "kako" je rečeno [1]. Više autora implicira da su prozodijska obeležja kao što su osnovna frekvencija (eng. *pitch*), intenzitet i brzina govora usko povezane sa emocijama u govoru [2-4]. Neki autori uzimaju u obzir kvalitet glasa i kratkotrajna spektralna obeležja u svojim studijama emocionalnog govora [5].

Poređenje ljudskog glasa sa sintetizovanim govorom nije uobičajeno i generalno je teže razumljivo [6]. Poznati su nedostaci računarskih sistema za konverziju teksta u govor, odnosno mašina za čitanje. Uloga emocija u govoru je da se obezbedi intonacija govora tako da se može tumačiti namera govornika, a to je bitno i u sintetizovanom govoru. TTS sistemi za sintetizovanje govora iz teksta vrše simulaciju emocija, ako su tako projektovani. Postoje dva ugla da se sagledaju emocije u govoru: (1) Generativni model (govornik) koji zavisi od mentalnog i fizičkog stanja govornika, sintakse i semantike izgovora, i (2) Akustični model (slušalac) koji opisuje parametre akustičkih signala sa strane slušaoca [6], [7].

Cilj ovog rada je da se uporedi snimljeni glas govornika za nekoliko određenih rečenica na arapskom jeziku, sa i bez emocija, i da se uporedi sa sintetizovanim govorom pomoću TTS softvera za arapski. U 2. poglavlju će biti predstavljene određene specifičnosti tekstova na arapskom jeziku koje su bitne za proces pretvaranja teksta u govor. U narednim poglavljima opisan je eksperimentalni deo rada u kom su snimljene određene rečenice sa neutralnim govorom i sa odgovarajućim emocijama. Potom su one poređene međusobno i u odnosu na sintetizovani govor istih rečenica. Na kraju su izvedeni zaključci kao smernice za korake u pravcu sintetizovanja govora.

2. SPECIFIČNOSTI TTS ZA ARAPSKI JEZIK

U opštem slučaju, sistem za pretvaranje teksta u govor se sastoji od dva glavna dela: jezičkog procesora i procesora za sintezu govora. Glavne funkcije jezičkog procesora su [8]:

1) Pretvaranje pisanih slova (grafema) u izgovorene (foneme)
2) Razdvajanje ulaznog teksta na reči
3) Pretvaranje reči u slogove
4) Generisanje prozodijskih simbola koji su potrebni da se sintetizuju prozodijske karakteristike govora
5) Analiza ulaznog teksta s obzirom na sintaksu, semantiku i diskurs.

Svaki jezik ima svoje sopstvene karakteristike, pa mu je potreban poseban jezički procesor. U nekim jezicima, kao što su arapski i španski, postoji bliska veza između pravopisa i izgovora, pa je stoga implementacija ovog procesora lakša nego u engleskom ili francuskom. Ovde ćemo objasniti neke karakteristike arapskog jezika i istaći specifičnosti TTS sistema za arapski jezik.

Arapsko pismo ima 28 slova plus "Hamza" (koji se javlja kao posebno slovo ili dijakritički znak). Arapski jezik ima 6 samoglasnika (3 kratka i 3 duga) i 2 poluvokala koji su diftonzi. Kratki samoglasnici se uglavnom izostavljaju pri svakodnevnom pisanju jer se očekuje od poznavaoca arapskog jezika da zaključi značenje reči odnosno rečenica

od konteksta. U svetim knjigama kao što su Kuran (Qur'an) i Biblija, zatim u knjigama gramatike i knjigama za decu tekstovi obavezno uključuju kratke samoglasnike napisane sa dijakritičkim znacima koji se stavljaju iznad ili ispod suglasnika koji im prethodi u slogu. Dijakritizacija određuje način na koji će reč biti izgovorena i daje jednosmislenost značenja reči. Dijakritički znaci su posebni pisani znaci koji se stavljaju iznad ili ispod slova, a izdvojićemo Fatha, Kasra, Damma (َ , ِ, ُ) koji služe za oznaku kratkih samoglasnika (kratko /a/, kratko /i/ , kratko /u/, respektivno), i znak Sukun (ْ). Znak Sukun iznad suglasnika označava da iza njega ne sledi samoglasnik (neophodan simbol za označavanje CVC slogova koji su veoma česti u arapskom) i još služi za označavanje diftonga. Neki znaci se koriste da bi se označili pojedini morfo-fonemski procesi, kao što je znak za udvajanje znakova Shadda, Tanween i El-madd [9].

Prvi korak u pretvaranju teksta u govor je transkripcija ulaznog teksta u foneme. Načinjen je poseban program za uređivanje teksta za arapski jezik koji prihvata ulazni tekst [10]. Ovaj ulazni tekst je niz karaktera koji mogu biti slova, dijakritički znaci ili ograničavači. Neki ograničavači se koriste da bi se označio kraj rečenice „.“, ili da se istakne povezivanje dveju susednih reči „-“. Dijakritizacija je vrlo važna za arapski jezik jer olakšava čitanje teksta na arapskom jeziku kako čoveku tako i računaru. Navešćemo primer slova 'm' i njegova četiri zapisa na arapskom sa četiri različita dijakritička znaka:

1. مَ se izgovara **‚ma'**
2. مِ se izgovara **‚mi'**
3. مُ se izgovara **‚mu'**
4. مْ se izgovara kao slovo ‚m'

Jasno je da tekst sa dijakritizacijom značajno olakšava proces pretvaranja ulaznog teksta u foneme i pojednostavljuje njegovu implementaciju na računaru jer jednozačno određuje preslikavanje grafem→fonem nezavisno od konteksta.

Formiranje fonetske transkripcije rečenice u arapskom jeziku mnogo je komplikovanije nego prosto nadovezivanje transkripcije reči koje tu rečenicu čine [10]. Problem sa povezivanjem se pojavljuje na granici između dve uzastopne reči ako druga počinje sa određenim članom *Alef lam* ili slovom *Alef* koje predstavlja *Hamzet el Wasl*. Odlučivanje da li je *Alef lam* određeni član ili ne je teško za računar, pa se, ukoliko to treba automatski raditi, mora obaviti semantička analiza. Stoga program za uređivanje teksta koristi poseban ograničavač „-“, da bi prevazišao ovaj problem.

Što se tiče procesa rasčlanjavanja reči na slogove ono je pojednostavljeno činjenicom da se slog u arapskom jeziku lako detektuje (svaki slog počinje suglasnikom iza koga sledi samoglasnik koji se naziva jezgro sloga) i da su brojno ograničeni na 6 tipova slogova (CV, CVC, CVCC plus ove tri varijante kada je vokal dug V:). Različiti tipovi akcentovanja na nivou reči u arapskom jeziku zavise od tipa sloga, njihovog broja i raspodele u reči.

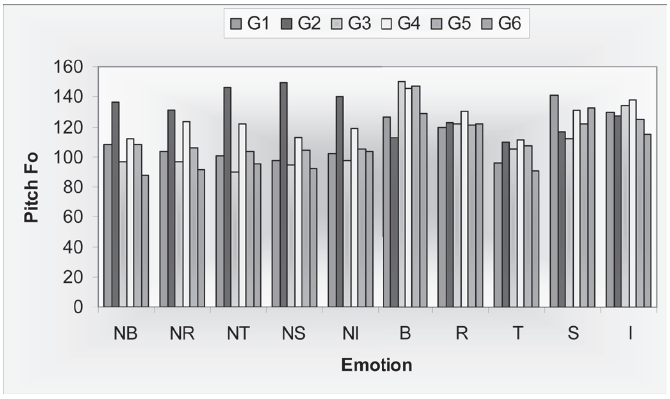
3. SNIMANJE GOVORNE BAZE

Za potrebe analize snimljena je baza od 60 govornih signala na arapskom jeziku. Izabrana je po jedna kratka rečenica za svaku od pet izabranih emocija: bes, radost, tuga, strah i iznenađenje. Ukupno šest govornika muškog pola je izgovaralo izabrane rečenice na dva načina: (1) na neutralni način i (2) sa emocijom, u skladu sa semantičkim sadržajem. Te rečenice su prikazane u tabeli 1. Snimanje govorne baze je urađeno u Laboratoriji za akustiku i govorne tehnologije na Fakultetu tehničkih nauka u Novom Sadu.

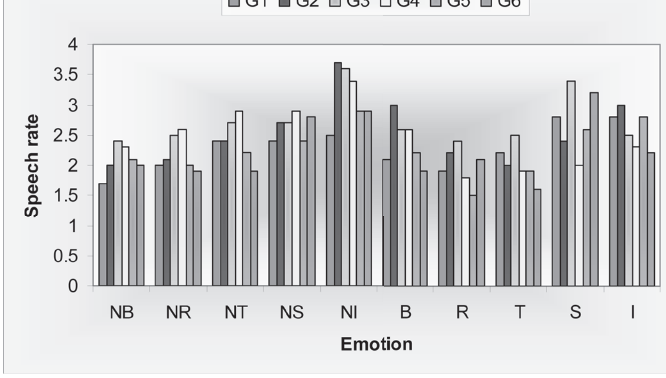
Tabela 1. – *Izabrane rečenice za odgovarajuće emocije na arapskom jeziku*

Emocija	Rečenice na arapskom bez dijakritizacije	Rečenice na arapskom sa dijakritizacijom	Prevod na srpski jezik
Bes	من تظن نفسك ؟	مَنْ تَظُنُّ نَفْسَكَ ؟	Šta ti misliš, ko si ti?
Radost	زالت الغيوم من السماء!	زَالَتْ الغُيُّوْمُ مِنَ السَّمَاءِ!	Nema više oblaka na nebu!
Tuga	انا حزين جدا اليوم!	أَنَا حَزِيْن جِدًّا اليَوْم!	Danas sam tako tužan!
Strah	اي الهي ما هذا المنظر المخيف!	اَي إِلَهِي مَا هَذَا المنظرِ المُخِيفِ!	O, Bože! Kakav zastrašujući prizor!
Iznenađenje	اي له من منظر جميل!	اَي لَهُ مِنْ مَنْظَر جَمِيل!	Kakav divan prizor!

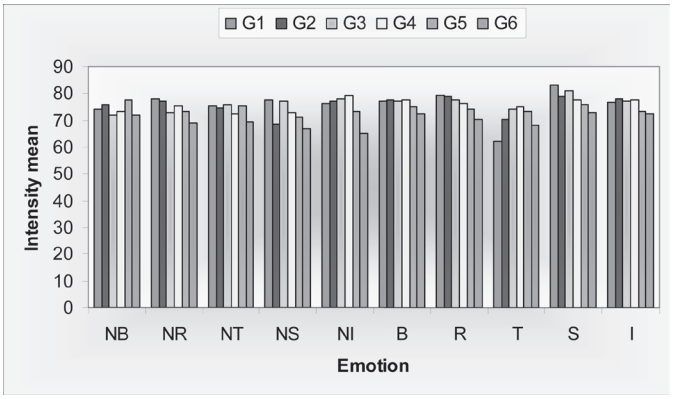
Sve snimljene rečenice analizirane su pomoću programa *PRAAT* da bi se utvrdili njihovi prozodijski parametri za analizu prisustva emocija u govoru: (1) osnovna frekvencija (*pitch*, tj. F0), (2) brzina govora i (3) intenzitet [6], [7], [11], [12] i [13]. Dobijeni rezultati su prikazani na sledeća tri dijagrama ili grafikona (1-3): osnovne frekvencije F0, brzine govora i prosečnog inteziteta svih govornika neutralnog govora uključujući i sve emocije, respektivno.



Dijagram 1. – *Osnovna frekvencija (F0) svih govornika sa i bez emocija*



Dijagram 2. – *Brzina govora svih govornika sa i bez emocija*



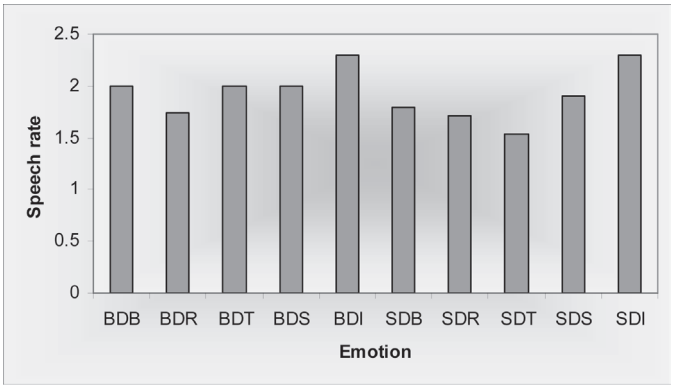
Dijagram 3. – *Prosečan intenzitet svih govornika sa i bez emocija*

Pri čemu su: NB, NR, NT, NS i NI neutralan govor za rečenice (besa, radosti, tuge, straha i iznenađenja) respektivno koje su izgovorene bez emocija (tj. neutralan govor). B-emocija besa, R-emocija radosti, T-emocija tuge, S-emocija straha, I-emocija iznenađenja, koje su izgovorene sa emocijom. G1-prvi govornik, G2-drugi govornik, G3-treći govornik, G4-četvrti govornik, G5-peti govornik i G6-šesti govornik.

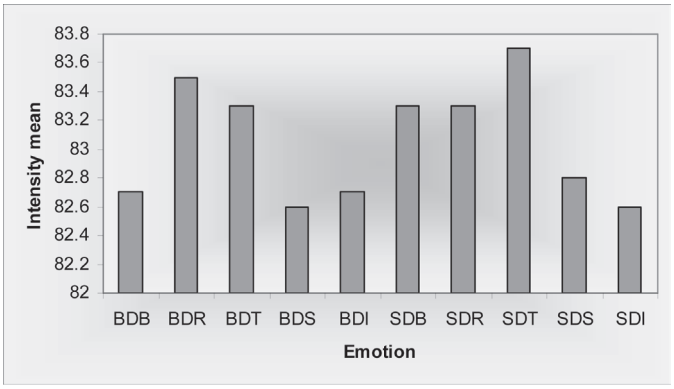
4. SNIMANJE POMOĆU ARAPSKOG TTS SOFTVERA (SINTETIZATORA)

Pomoću arapskog TTS sintetizatora izvršeno je sintetizovanje rečenica (sa i bez dijakritizacije u sintetizatoru) koje su navedene u poglavlju 3, a prikazane u tabeli 1, a zatim su te sintetizovane rečenice snimljene i analizirane. Sve snimljene rečenice analizirane su pomoću programa *PRAAT* da bi se utvrdili njihovi prozodijski parametri: (1) osnovna frekvencija (*pitch*, tj. F0), (2) brzina govora i (3) intenzitet. Što se tiče rečenica u tabeli 1, ne postoji razlika u pisanju arapskih rečenica u poglavlju 3 i 4, nego je razlika da sintetizator izgovara rečenice napisane sa i bez dijakritizacije. Dobijeni rezultati sa i bez dijakritizacije su prikazani na sledeća tri (4-6) dijagrama: osnovne frekvencije F0, brzine govora i prosečnog inteziteta, respektivno.

Dijagram 4. – *Osnovna frekvencija (F0) sintetizatora sa i bez dijakritizacije*



Dijagram 5. – *Brzina govora sintetizatora sa i bez dijakritizacije*



Dijagram 6. – *Prosečan intenzitet sintetizatora sa i bez dijakritizacije*

Pri čemu su: BD-bez dijakritizacije, SD-sa dijakritizacijom, tako da je BDB i SDB rečenica „*Šta ti misliš, ko si ti?*“, izgovorena sa i bez dijakritizacije za emociju besa, BDR i SDR rečenica *“Nema više oblaka na nebu!“,* izgovorena sa i bez dijakritizacije za emociju radosti, BDT i SDT rečenica „*Danas sam tako tužan!*“, izgovorena sa i bez dijakritizacije za emociju tuge, BDS i SDS rečenica *„O, Bože! Kakav zastrašujući prizor!“,* izgovorena sa i bez dijakritizacije za emociju straha i BDI i SDI rečenica „*Kakav divan prizor!*“, izgovorena sa i bez dijakritizacije za emociju iznenađenja.

5. ANALIZA PROZODIJSKIH PARAMETARA

U nastavku je izvršeno poređenje razlika osnovnih prozodijskih parametara u govornoj bazi za šest arapskih govornika. Za ovakve analize trebao bi veći broj govornika, ali je to bilo teško sakupiti u Srbiji za potrebe ovih istraživanja.

Na osnovu istraživanja govorne baze podataka, srednje vrednosti svakog parametra te baze i analize prozodijskih parametara u datim rečenicama na arapskom jeziku koje su izgovorene sa i bez emocija, i koje su isto tako proizvedene mašinskim prevođenjem teksta u sintetizovani govor, zaključuje se sledeće:

Emocija besa:

Postoji mala razlika između srednje vrednosti osnovne frekvencije F0 za prirodan (neutralan) govor bez emocija i osnovne frekvencije sintetizovanog govora sa i bez dijakritizacije, kao i između besa sa emocijom i sintetizovanog govora sa dijakritizacijom, a veća razlika postoji kod besa sa emocijom i sintetizovanog govora bez dijakritizacije.

Emocija radosti:

Kod ove emocije postoji mala razlika između srednje vrednosti osnovne frekvencije F0 za prirodan govor sa emocijom i vrednosti osnovne frekvencije F0 sintetizovanog govora sa i bez dijakritizacije, a malo veća razlika postoji između prirodnog govora bez emocija i sintetizovanog govora sa i bez dijakritizacije.

Emocija tuge:

Kod ove emocije srednja vrednost osnovne frekvencije F0 za prirodan govor sa i bez emocije i vrednost osnovne frekvencije sintetizovanog govora sa i bez dijakritizacije, postoji mala razlika između njih.

Emocija straha:

Kod emocije straha srednja vrednost osnovne frekvencije F0 za prirodan govor bez emocije i vrednost osnovne frekvencije sintetizovanog govora sa i bez dijakritizacije skoro su iste. Isto to važi za srednju vrednost osnovne frekvencije F0 prirodnog govora sa emocijom straha i vrednost osnovne frekvencije F0 sintetizovanog govora sa i bez dijakritizacije.

Emocija iznenađenja:

Kod emocije iznenađenja srednja vrednost osnovne frekvencije F0 za prirodan govor sa emocijom i vrednost osnovne frekvencije F0 sintetizovanog govora sa i bez dijakritizacije, postoji mala razlika između njih, a veća razlika postoji između prirodnog govora bez emocije i sintetizovanog govora sa i bez dijakritizacije.

Srednja vrednost brzine prirodnog govora sa i bez emocija i brzina sintetizovanog govora sa i bez dijakritizacije ne menja se puno za sve emocije.

Srednja vrednost prosečnog intenziteta prirodnog govora sa i bez emocija i prosečnog intenziteta sintetizovanog govora sa i bez dijakritizacije postoje male razlike između njih za sve emocije.

6. ZAKLJUČAK

U radu su predstavljene specifičnosti zapisa tekstova na arapskom jeziku kako bi se ukazalo na probleme u sintetizovanju govora iz teksta na ovom i sličnim jezicima. U većem delu rada analizirani su prozodijski parametri koji se odnose na emocije u prirodnom i sintetizovanom govoru. Snimljena je posebna govorna baza za potrebe istraživanja u ovom radu sa govornicima na arapskom jeziku. Izvršeno je detaljno poređenje razlika govora sa emocijama u odnosu na neutralni izgovor istih rečenica, a poređenja su izvršena i u odnosu na sintetizovani govor tih rečenica u varijantama sa i bez dijakritizacije.

Na osnovu analize ilustracija prozodijskih parametara u datim primerima rečenica na arapskom jeziku koje su izgovorene sa i bez emocija, i koje su isto tako proizvedene mašinskim prevođenjem teksta u sintetizovani govor, zaključuje se da emocije nisu toliko izražene u sintetizovanom govoru kao u prirodnom ljudskom glasu, te da ima još dosta prostora za istraživanja kako na pravi način uključiti emocije u automatski sintetizovani govor. Potrebno je više eksperimenata da bi se postigli konkretniji rezultati, ali je jasno da treba posvetiti više prostora akustičkim modelima koji opisuju parametre akustičkih signala sa strane slušaoca, da bi se dobio bolji kvalitet sintetizovanog glasa. Time bi se dobio adekvatniji pič, a on se pokazao kao važniji parametar od trajanja. Ovaj rad je još jednom potvrdio da su za ocenu emocija najvažnija tri prozodijska parametra govora: F0, trajanje i intenzitet, kao i da se prosečan F0 za neutralni govor najviše razlikovao u odnosu na prosečnu vrednost kod govora sa emocijama.

Dalje istraživanje treba usmeriti ka snimanju veće govorne baze koja bi sadržala obimniji tekstualni korpus izgovaran sa određenim, izabranim emocijama. Baza bi trebala da sadrži i snimke ženskih govornika jer je poznato da postoji razlika u izražavanju određenih emocija između govornika muškog i ženskog pola. Kvalitet snimljene baze treba proveriti auditivnim testom. Tada će analiza izabranih prozodijskih obeležja dati pouzdanije rezultate i pružiti konkretnije zaključke.

LITERATURA

[1] S. T. Jovičić i drugi, "Formiranje korpusa govorne ekspresije emocija i stavova u srpskom jeziku," *11. Telekomunikacioni forum TELFOR,* Beograd, 2003.

[2] M. Rajković i drugi, "Intenzitetske i vremenske karakteristike emotivnih ekspresija u srpskom govornom diskursu," *11. Telekomunikacioni forum TELFOR,* Beograd, 2003.

[3] A. Esposito et al. (Eds.), *Verbal and Nonverbal Communication Behaviours*, LNAI 4775, pp. 74-84, Springer-Verlag Berlin Heidelberg, 2007.

[4] L. ten Bosch, "Emotions, speech and ASR framework," *Speech Communication* 40, pp. 213-225, 2003.

[5] C. Gobl, A. Ni Chasaide, "The role of voice quality in communicating emotion, mood and attitude," *Speech Communication* 40, pp. 189-212, 2003.

[6] J. Cahn, "The generation of affect in synthesized speech," *Journal of the American Voice I/O Society*, Vol. 8. pp. 1-19, Jul. 1990.

[7] O. Pierre-Yves, "The production and recognition of emotions in speech: features and algorithms," *International Journal of Human-Computer Studies Journal of Human-Computer Studies*, vol. 59 no. 1-2, pp. 157-183, Jul. 2003.

[8] J. Allen, "Synthesis of Speech from Unrestricted Text," *Proc. IEEE*, 64 (1976), 433-42.

[9] http://en.wikipedia.org/wiki/Arabic_alphabet

[10] Abu-Elyazeed M. F. , "An Arabic Text-to-Speech System," *Ph.D. Thesis*, Cairo Univ., 1990.

[11] F. Zotter, "Emotional speech," at URL: http://spsc.inw.tugraz.at/courses/asp/ws03/talks/zotter.pdf

[12] I. R. Murray, M. D. Edgington, D. Campion, J. Lynn, "Rule-based emotion synthesis using concatenated speech," *ISCA Workshop on Speech & Emotion*, Northern Ireland, 2000, pp. 173-177.

[13] J. M. Montero, J. Gutiérrez-Arriola, J. Colás, E. Enríquez and J. M. Pardo, "Analysis and modelling of emotional speech in Spanich," at URL: http://lorien.die.upm.es/~juancho/conferences/0237.pdf

[14] M. Al-Abbushi, "Poređenje emocija u arapskom prirodnom i sintetizovanom govoru," 17. Telekomunikacioni forum TELFOR, Beograd, 24-26.11.2009.

[15] M. Al-Abbushi, M. Bojanić, V. Delić, "Prepoznavanje emocija u arapskom prirodnom i sintetizovanom govoru," 9. Infoteh, Jahorina, 17-19.03.2010.

Mohammad Al-Abbushi, PhD student, Univerzitet u Novom Sadu
Oblasti interesovanja: razvoj govornih tehnologija

Prof. dr Vlado Delić, Univerzitet u Novom Sadu, Fakultet tehničkih nauka
Kontakt: vlado.delic@ktios.net
Oblasti interesovanja: razvoj govornih tehnologija