

SITUL 'LIMBA ROMÂNĂ VORBITĂ'

H.N. TEODORESCU^{1,2}, M. FERARU², D. TRANDABĂȚ^{1,3}

¹INSTITUTUL DE INFORMATICĂ TEORETICĂ, IASI – ROMÂNIA

²UNIVERSITATEA TEHNICĂ „GH. ASACHI”, IASI – ROMÂNIA

³UNIVERSITATEA “AL.I.CUZA”, FACULTATEA DE INFORMATICĂ,
IASI – ROMÂNIA;

hteodor@etc.tuiasi.ro

Rezumat

Inițiativa construirii unei arhive publice a sunetelor a fost determinată de lipsa unei asemenea resurse pentru limba română, lipsă resimțită atât în cercetare, cât și în învățământ. Situl include peste 600 de înregistrări, în diverse formate de precizie și codare, ale sunetelor limbii române vorbite.

1. Introducere

Situl (http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.htm) a fost creat prin colaborarea dintre Institutul de Informatică Teoretică al Academiei Române - Grupul de Prelucrarea Vorbirii, Universitatea "Al. I. Cuza" Iași - Facultatea de Informatică și Universitatea Tehnică "Gh. Asachi" Iași - Centrul de Excelență în cercetare "CERFS" (coordonat de primul autor) în ideea realizării unui suport pentru un "dicționar al sunetelor" limbii române. Situl cuprinde, pe lângă sunetele propriu-zise (vocale, consoane, diftongi, scurte fraze), informații despre fonetica limbii române, protocoale de documentare și de înregistrare, instrumente de analiză, trimiteri la lucrări referitoare la prelucrarea limbii vorbite etc.

Scopul acestei inițiative a fost realizarea unei arhive pentru sunetele limbii române cu următoarele caracteristici:

- Bază de date cu voci atât profesionale (pronunții "perfecte"), cât și ne-profesionale ("vocea omului de pe stradă"), din zona Iași, iar apoi, pe cât posibil, cu pronunții (sunete) specifice diverselor regiuni.
- Pe baza acestor date, realizarea unui studiu statistic amplu al sunetelor limbii române, care să includă, de exemplu "triunghiul formaților limbii române", "caracteristici statistice ale pronunțiilor regionale" etc.
- Corectarea unor sisteme de sinteză ne-concatenativă, pe baza datelor din arhivă.
- Îmbunătățirea unor sisteme de recunoaștere acustică, pe baza datelor din arhivă. (Arhiva ar putea deveni, sperăm, un banc de probă pentru asemenea sisteme).
- Realizarea unei baze de date (sunete) a limbii vorbite pentru persoane cu diverse patologii (neurologice, laringiene, nazale, bucale, respiratorii) (Teodorescu et al., 2006a). Această bază de sunete produse de voci patologice va fi parțial utilă și în cercetările pentru un grant CEEX.

- Realizarea unei baze de date de tip silabe și cuvinte (sursă pentru sintetizoare concatenative și banc de probă pentru sisteme de recunoaștere de cuvinte).
- Un dicționar electronic al pronunțiilor din limba română. Corelarea cu Atlasele limbii române.
- Pagina de referințe, care să prezinte toate titlurile de volume și lucrări ce au ca obiect sunetele limbii române (fonetică, sinteză, recunoaștere etc.).

S-au efectuat peste 600 de înregistrări, în diverse formate de precizie și codare (Teodorescu et al., 2006b). Fiecare vorbitor a rostit de trei ori fiecare frază, propoziție, cuvânt, vocală, consoană, diftong, triftong, hiat precum și grupuri de sunete specifice limbii române (ex. *ce*, *ci*, *che*, *chi* etc). Înregistrările au fost efectuate folosind programul Goldwave 5.0 la o frecvență de eșantionare de 22050Hz, codate pe 16 și 24 biți, mono.

Fișierele sunt grupate în clase, după cum urmează:

A. Sunete de bază: i) fișiere de vocale; ii) fișiere de consoane, înregistrate conform standardului IPA, în forma VCV, unde V este vocala *a*; iii) fișiere de diftongi, triftongi și hiatusuri; iv) fișiere de sunete specifice, care în scrierea în limba română corespund grupurilor *ce*, *ci*, *che* (ke), *chi* (ki), *ge*, *gi*, *ghe*, *ghi*. Subiecții sunt atât bărbați, cât și femei, persoane cu vârsta cuprinsă între 26-31 ani, proveniți (născuți și educați) din zona Moldovei de mijloc (județele Iași, Vaslui, Bacău), cu educație superioară și fără patologii manifestate. Vocalele sunt înregistrate atât în varianta scurtă (pronunție uzuală), cât și în varianta „susținută”.

B. Scurte propoziții sau segmente de fraze, cu încărcătură emoțională diferită

Pe lângă sunete simple au fost înregistrate și fraze scurte: *Cine a făcut asta*, *Vine mama*, *Aseară*. Fiecărui subiect i s-a cerut să rostească fiecare frază simulând următoarele emoții: fericire, tristețe, bucurie, ură, optimism, pesimism, ton exclamativ, ton interogativ, ton plat și starea de supărare. Ulterior, aceste stări au fost reduse la patru: fericire, supărare, furie și ton neutru.

2. Metodologia de înregistrare

Analiza zgomotului de fond

Zgomotul de fond este un semnal aleator, arareori staționar, mixat cu unele perturbații determinate de tipul perturbațiilor de frecvență ale rețelei. Pentru a asigura calitatea înregistrării, amplitudinea zgomotului trebuie să fie mult mai mică decât amplitudinea semnalului. Înregistrările au fost efectuate într-un laborator cu zgomot redus, dar nu s-a dispus de un spațiu total izolat fonic. O înregistrare de bună calitate are un zgomot în al cărui spectru nu se depășește valoarea de -80dB pentru nici o componentă spectrală, în timp ce formantul F1 are un nivel cu cel puțin 30dB mai mare, iar formanții superiori au tipic amplitudini cu peste 15-20dB peste nivelul zgomotului (Teodorescu, 2006a). Fișierele sunt în curs de verificare, urmând a fi eliminate cele care nu satisfac nivelul de calitate pe care ni l-am impus.

Alegerea microfonului

Microfonul este primul element în convertirea sunetelor și are un rol esențial în calitatea înregistrărilor. Un microfon de bandă limitată, de sensibilitate redusă sau cu zgomot mare poate compromite înregistrarea. Un microfon de calitate are zgomot redus și raportul semnal / zgomot bun. Caracteristica omnidirecțională conduce la o sensibilitate mare la zgomotele ambientale și nu este de dorit. Sunetele au fost înregistrate folosind căști cu microfon SONIC Stereo Dinamic Headphones HP-259 cu caracteristicile: frecvența de răspuns: 20-20.000 Hz; impedanța microfon: $U=3V$, $R=1,5K \Omega$; impedanța căști: 32Ω ; sensibilitate microfon: $-58dB\pm 2$; sensibilitate căști: $100dB/mw$; putere: $100mW$.

Poziționarea microfonului

O atenție specială trebuie acordată poziției microfonului, deoarece apar zgomote sau distorsiuni introduse prin poziționarea deficitară. Ținerea microfonului prea aproape de gură poate duce la efectul de saturare a amplificatorului, cu rezultat de puternică distorsionare a semnalului. Se recomandă menținerea microfonului mai jos de gură, aproximativ în dreptul bărbiei, la câțiva centimetri de aceasta. Distanța de la bărbie trebuie să fie aproximativ egală cu distanța până la buze.

Placa de sunet și driverule corespunzătoare

Majoritatea calculatoarelor actuale conțin pe placa de bază circuite de preluare a semnalelor de la microfon și de generare de semnale audio la căști sau difuzoare (calitatea acestor circuite diferă substanțial de la o placă la alta). Placa de bază a calculatorului pe care au fost efectuate înregistrările este MB FOXCONN 760 GXK8MC-S, având încorporată o placă de sunet Sound MAX Digital Audio produsă de Analog Devices cu caracteristicile: procesor de semnal SiS964, standard AC '97.

3. Adnotări

Sunetele și frazele disponibile pe sit au fost adnotate la diferite niveluri cu scopul de a avea un corpus pentru analiza statistică a datelor. Adnotarea s-a realizat folosind utilitarul Praat™ (www.praat.org), ales datorită eficienței, recunoașterii internaționale și ușurinței în folosire. Primul pas a fost segmentarea la nivel de fonem. Ulterior, au fost grupate fonemele pentru a se realiza o segmentare în silabe, cuvinte și propoziții. Principala problemă a fost faptul că este dificil de stabilit întotdeauna cu exactitate unde se află granița dintre foneme. În figura 1 este prezentat un exemplu de adnotare pentru propoziția *Vine mama!*. În viitor, adnotările vor fi validate prin analiza efectuată de mai mulți adnotatori.

Informația de la nivel fonologic va fi completată cu informații prozodice (ton, intensitate, durată etc.), deoarece unul dintre obiectivele noastre este de a detecta parametri prozodici care fac diferența dintre vorbirea umană și cea sintetică.

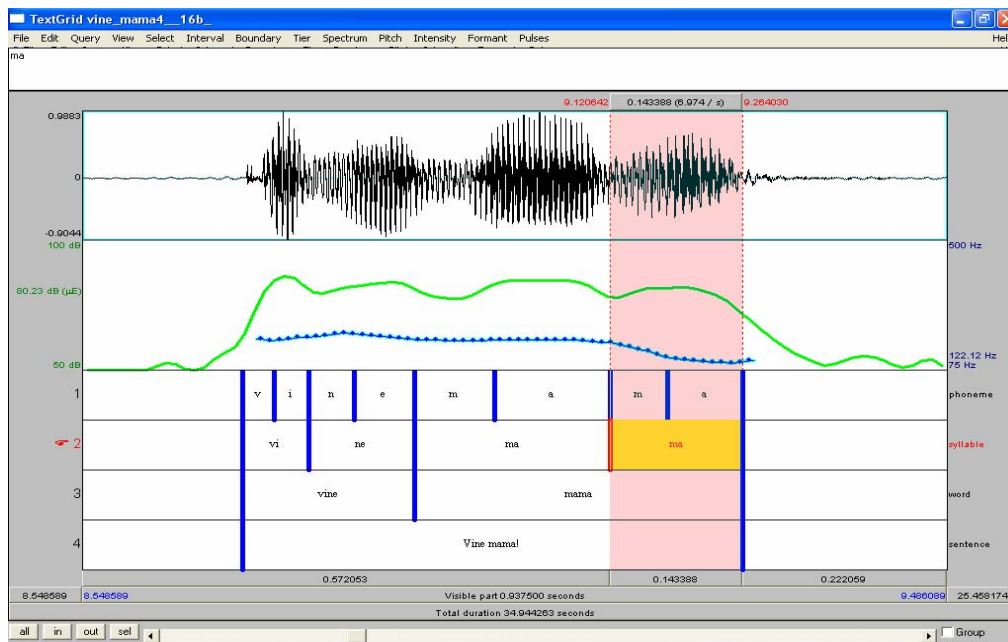


Figura 1: Exemplu de adnotare a propoziției *Vine mama!*

4. Alte elemente ale resursei: documentații, instrumente etc.

Subiecții au fost informați anterior înregistrărilor despre obiectivele proiectului, fiind asigurați de confidențialitatea datelor personale. Subiecții au semnat un consimțământ informat în conformitate cu „Protocolul de Protecție a Subiecților Umani” al *U.S. Food and Drug Administration* (http://www.fda.gov/cdrh/devadvice/ide/informed_consent.shtml) și cu „Principiile etice ale Asociației Acustice Americane privind cercetările care implică ființa umană” (<http://asa.aip.org/ethical.html>). Vorbitorilor li s-au explicat în prealabil condițiile de înregistrare: poziția microfonului, susținerea vocalelor pe o durată cât mai lungă, dar fără a se realiza vreun efort etc. De asemenea, fiecare subiect a completat o fișă personală, care include date despre vârsta, sexul, limba, educația vorbitorului, patologii, precum și evaluarea calității subiective a vocii. Pe lângă sunetele propriu-zise și fișele vorbitorilor, situl mai conține și instrumente de analiză a semnalului vocal.

5. Metode suplimentare de analiză: măsurări accelerometrice

Printre elementele specifice introduse de grupul nostru pentru analiza procesului vorbirii, se afla metoda accelerometrică. Metoda constă în determinarea accelerațiilor mandibulei în timpul vorbirii și corelarea mișcărilor cu tipul sunetelor pronunțate, cu duratele sunetelor și cu energia sonoră. Deși gradul de corelare constatată până în prezent este relativ redus, sperăm că metoda poate ajuta la segmentarea automată a semnalului vocal și la evidențierea unor corelații între caracteristicile pronunției cu mișcările fonatorii. Fără să fie absolut nouă (există un număr mic de lucrări cu abordări oarecum similare, dar nu în scopul segmentării), abordarea sperăm să aducă elemente suplimentare în explicarea proceselor vorbirii. Rezultatele preliminare obținute

(Teodorescu 2006a, b) evidențiază caracteristici ale tranzițiilor sunet nazal - vocală, *o-a* în diftongul *oa*, precum și consoană plozivă-vocală.

6. Concluzii și direcții viitoare

Considerăm că resursa este utilă în prezent ca mijloc educațional, iar în viitorul apropiat ca suport în cercetarea lingvistică și în realizarea de aplicații informatice (sinteză și recunoaștere). Credem că resursa impune și noi standarde de calitate în realizarea unor resurse similare.

Arhiva sunetelor limbii române va fi dezvoltată prin adăugarea de noi înregistrări și de prelucrări statistice ale sunetelor. Ulterior, se va urmări adăugarea unor înregistrări cu ușoare patologii, cum ar fi tremurul vocii (de natura emoțională sau patologică), anotate și prelucrate (Teodorescu et al., 2006c).

Referințe bibliografice

- (Teodorescu, 2006a) Teodorescu H.N. (2006). Gnatofonia și Gnatosonia. Analiza semnalelor vocale, Capitolul 2, *Ed. Performatica*, Iași, România, pag. 29-40.
- (Teodorescu, 2006b) Teodorescu H.N. (2006). Gnatophonetics – A New Discipline Analyzing Relations between Speech and the Stomato-Gnathic System. *Zilele Academice Ieșene, Simp Inventica. Simpozionul național "Bazele performanței și inventică"* organizat în cadrul "Zilelor Academice Ieșene" ISBN 973-730-244-3, 978-973-730-244-1, 9 September 2006.
- (Teodorescu et al., 2006a) Teodorescu H.N., Zbancioc M., Mihăilescu E. (2006), Speech Technology and Bio-Medical Engineering Teaching Based on the Web – A New Tool and Case Study. *Conference ICL 2006*, Villach, 27 -29 September 2006, Proceedings CD 2005 Ambient and Mobile Learning, Kasset University Press, Editors Michael Auer, Ursula Auer and R. Mittermeir, ISBN 3-89958-136-9.
- (Teodorescu et al., 2006b) Teodorescu H.N., Tandabăț D., Feraru M., Zbancioc M., Luca R.(2006). A corpus of the sounds in the Romanian spoken language for language-related education. *International Conference on Human and Material Resources in Foreign Language Learning – RFL 2006*, Murcia, Spania, 12-14 iulie 2006.
- (Teodorescu et al., 2006c) Teodorescu H.N., Feraru M., Tandabat D. (2006), Nonlinear Assessment of Professional Voice 'Pleasantness', Conference *BIOSIGNAL 2006*, ISBN 80-214-3152-0, Brno, 28-30 June 2006, pag. 63-66.
- Voiced Sounds of Romanian Language Project. [http://iit.iit.tuiasi.ro/romanain_spoken_language/index.htm].