D1.15. Identificarea pattern-urilor prozodice și evidențierea corelațiilor între text și semnal vocal

Aceste rezultate au fost obținute prin finanțare in cadrul Programului PN-III Proiecte complexe realizate în consorții CDI, derulat cu sprijinul MEN – UEFISCDI,
Cod: PN-III-P1-1.2-PCCDI-2017-0818, Contract Nr. 73 PCCDI/2018:
“SINTERO: Tehnologii de realizare a interfețelor om-mașină pentru sinteza text-vorbire cu expresivitate”

Denumirea organizației participante in proiect	Acronim organizație	Tip organizație	Rolul organizației în proiect (Coordonator/partener)
Institutul de Cercetări Pentru Inteligență Artificială “Mihai Drăgănescu”	ICIA	UNI	CO
Universitatea Tehnică din Cluj-Napoca	UTCN	UNI	P1
Universitatea Politehnica din București	UPB	UNI	P2
Universitatea "Alexandru Ioan Cuza" din Iași	UAIC	UNI	P3

MINISTERUL EDUCAȚIEI NAȚIONALE
UEFISCDI

Date de identificare proiect

Număr contract:	PN-III-P1-1.2-PCCDI-2017-0818, Nr. 73 PCCDI/2018
Acronim / titlu:	**„SINTERO: Tehnologii de realizare a interfețelor om-mașină pentru sinteza text-vorbire cu expresivitate”**
Titlu livrabil:	**D1.15. Identificarea pattern-urilor prozodice și evidențierea corelațiilor între text și semnal vocal**
Termen:	**Mai 2018**
Editor:	**Mircea Giurgiu (Universitatea Tehnică din Cluj-Napoca)**
Adresa de eMail editor:	**Mircea.Giurgiu@com.utcluj.ro**
Autori, in ordine alfabetică:	**Mircea Giurgiu, Adriana Stan**
Ofițer de proiect:	**Cristian STROE**

Rezumat:

Ca fundament pentru cercetările raportate în acest livrabil sunt rezultatele anterioare obținute de partenerii CO-ICIA (procesarea limbajului natural) și P1-UTCN (analiza unităților acustice din semnalul vocal), care pun în evidență principalii factori de natură linvistică prin care se manifestă modificările prozodice in forma de undă: accentul, intonația în vorbire, silabificarea, pauzele, ritmul vorbirii, respectiv elemente de morfologie și sintaxă în interacțiune. Pornind de aici s-au ramificat două direcții de cercetare: identificarea modului de manifestare a prozodiei în parametrii semnalului vocal, respectiv corelația parametrilor prozodici cu caracteristici extrase din text.

În primul rând sunt prezentate rezultatele experimentale privind variația parametrilor prozodici frecvență fundamentală pentru vocale, frecvența fundamentală în funcție de accent, frecvență fundamentală în funcție de intonația din propoziție, variația frecvenței formanților pentru diferiți vorbitori, respectiv rolul duratei și a pauzelor în modelarea pattern-urilor prozodice. Analiza s-a realizat pe un corpus de semnal vocal înregistrat în acest scop. De exemplu, pentru unitățile acustice diftongi, pattern-urile prozodice indica faptul ca frecvenţele fundamentale suferă variaţii atunci când diftongii (respectiv vocalele) sunt încadraţi în cuvinte; F0 maxim scade atunci când avem grupuri de vocale încadrate împreună în cuvânt, iar energia acestor diftongi încadraţi în cuvinte este sensibil mai mică decât cea a diftongilor, triftongilor izolaţi. Similar s-au obținut rezultate pentru diferite categorii de unitîți acustice. Un alt exemplu este pentru accent. Una din concluziile importante ale studiului se referă la o creştere a frecventei fundamentale pentru silabele (sau vocalele) accentuate, fata de cele neaccentuate in medie cu 5%..20% (în 90% din cazuri creşterea s-a plasat in intervalul 9%..12%). Merita făcuta si observaţia ca au existat si câteva cazuri in care accentuarea unei silabe nu a adus nici un fel de diferenţiere din punctul de vedere al valorii F0. Similar sunt prezentate rezultate pentru formanți, respectiv evaluarea duratei unităților acustice în funcție de accent.

În al doilea rând sunt prezentate rezultate privind analiza caracteristicilor de natura lingvistică ce afecteaza prozodia, in special la nivel de intonație de propoziție. Sunt identificate un set de 7 pattern-uri intonaționale la nivel de propoziție, dar și efcetul prozodic al semnelor de punctuație.

Cercetările demonstrează faptul că pattern-urile prozodice manifestate la nivelul semnalului vocal au legătură directă și prezintă strînse corelații pe termen scurt sau pe termen lung cu atribute de morfologie și sintaxă aferente textului. Principalele atribute se refera la poziționare accent în cuvinte, silabificare, părțile de vorbire, sintaxa, respectiv punctuație. Aceste rezultate prezintă fundamentul pentru dezvoltarea unor noi metode de sinteză expresiva a vorbirii prin intermediul unor module de analiza a expresivității textului (în componenta software de procesare de text), respectiv de modificare automată a prozodiei (în componenta software de sinteză de semnal).

Cuprins

1. Introducere

Acest livrabil (D1.15 „Identificarea pattern-urilor prozodice și evidențierea corelațiilor între text și semnal vocal") prezintă rezultatele obținute în activitatea A1.15 din planul de realizare a proiectelor componente, in mod specific din cadrul sub-proiectului P4 (SINTERO).

Raportul demonstrează faptul că pattern-urile prozodice manifestate la nivelul semnalului vocal au legătură directă și prezintă strînse corelații pe termen scurt sau pe termen lung, în interacțiune, cu atribute de morfologie și sintaxă aferente textului. Principalele atribute se refera la poziționare accent în cuvinte, silabificare, părțile de vorbire, sintaxa, respectiv punctuație.

Pentru a pune în evidența aceste pattern-uri prozodice s-a achiziționat semnal vocal și din acesta au fost extrași parametri acustici precum frecvența fundamentală, formanții, s-a estimat durata unităților acustice, toate în corelație cu caracteristcile lingvistice prezentate mai sus.

Aceste rezultate prezintă fundamentul pentru dezvoltarea unor noi metode de sinteză expresiva a vorbirii prin intermediul unor module de analiza a expresivității textului (în componenta software de procesare de text), respectiv de modificare automată a prozodiei (în componenta software de sinteză de semnal).

2. Identificarea pattern-urilor prozodice, o problemă deschisă

Impresia naturaleţii semnalului sintetizat de un sistem de Sinteză din Text a Semnalului Vocal (STSV) depinde de bogăţia de contururi intonative şi de calitatea pattern-urilor prozodice. Generatorul prozodic este responsabil pentru aceste două aspecte supra-segmentale. Cele trei elemente aferente prozodiei sunt:

- accentul (cu efect asupra amplitudinii si duratei fonemului);
- intonaţia (variaţia in timp a frecvenţei fundamentale – F0);
- ritmul (durata fonemelor si viteza cu care sunt acestea sintetizate),

Până nu demult, majoritatea cercetărilor tratau separat aceste trei elemente. În realitate ele sunt în strânsă interdependenţă. De exemplu, în limba română accentul pe cuvânt este liber, variind între ultimele două silabe ale cuvântului. Exista multe excepţii de la această regulă. Cuvinte cu aceeaşi ortografie au semantică diferită în funcţie de locul accentului. De exemplu: *vesélă-véselă* sau *curéle - cúrele*. Ca atare, trebuie găsite reguli care să transforme vorbirea sintetic monotonă în una naturală. Pentru aceasta este necesar un studiu experimental amănunţit asupra conturului frecvenţei fundamentale pentru diferite tipuri de propoziţii (declarative, întrebări, exclamaţii, etc).

Pentru propoziţii declarative frecvenţa fundamentală creşte pe primul cuvânt (de la 100% la 140% din valoarea sa, apoi coboară la 125% pe ultima parte a cuvântului) şi descreşte până la sfârşitul propoziţiei, cu excepţia ultimului cuvânt. Aici scade la 70% şi rămâne constantă. Propoziţiile interogative pot avea un cuvânt specific de interogare (*cine, unde, când*) sau pot să nu aibă. În primul caz frecvenţa fundamentală creşte pe acel cuvânt de la 100% la 160% şi revine la 100%. Pentru al doilea tip de întrebări s-a adoptat un contur convenţional, dar efectele subtile de intonaţie nu pot fi rezolvate. În esenţă, pentru a reprezenta conturul intonaţional, frecvenţa fundamentală (pitch) este supusă unei operaţii de "stilizare", cu scopul de a aproxima F0 printr-o secvenţă de segmente de dreaptă, rezultatul fiind o reprezentare foarte apropiată de variaţiile din intonaţia vorbitorului. De exemplu, Figura 2.1. arată câţiva dintre parametrii unei propoziţii interogative în care conturul intonaţional scade lin în primul cuvânt, urmând apoi o porţiune constantă în al doilea, urmată de o creştere lină finală. Realizarea acustică a semnelor ca virgula, două puncte şi punct şi virgulă contribuie la îmbunătăţirea naturaleţii vorbirii.

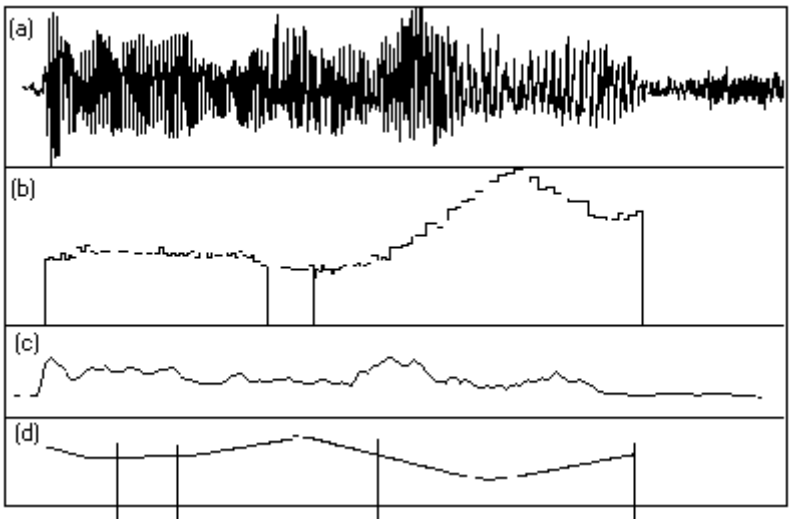

Figura 2.1. Variaţia intonaţiei (F0).
(a) - semnalul original, (b) – variaţia F0, (c) - amplitudinea, (d) - melodia prozodicã

2.1. Accentul

O anume silabă poate fi rostită cu mai mare sau mai mică intensitate la fel ca în exemplul "*ca*sa" sau "ca*sa* ". Aceste fenomen este descris prin accent. În funcţie de domeniile sale diferite de acţiune, se pot distinge trei tipuri de accent :

- *accentul la nivel de silabă* sau de cuvânt
- accentul pe *propoziţie*
- accentul pe *frază*

O altă caracteristică acustică importantă este durata relativă a silabei. Accentul silabei precum şi silaba de la sfârşitul propoziţiei sunt în principiu lungi. Durata pauzei şi a silabei se obţin din segmentarea fonetică. Realizarea acustică a accentului va afecta întotdeauna cel puţin doi dintre următorii parametrii prozodici: frecvenţa fundamentală, intensitatea sau durata, iar în unele cazuri pe toţi trei. Astfel, se constată o creştere a frecvenţei fundamentale şi a amplitudinii pentru silabele accentuate comparativ cu echivalentele lor neaccentuate.

Din punctul de vedere al poziţiei în cuvânt, accentul în limba română este liber (nu cade în toate cuvintele pe aceeaşi silabă) şi mobil (îşi schimbă locul în cursul flexiuni), mai ales la verb, dar şi la unele substantive sau pronume. În consecinţă, accentuarea cuvintelor trebuie marcată pentru fiecare cuvânt, iar în cursul flexiunii trebuie cunoscute regulile după care acesta îşi schimbă locul. De obicei fiecare cuvânt polisilabic are un singur accent (o singură culme dinamică), întrucât cuvintele româneşti nu sunt excesiv de lungi. Un accent secundar apare în împrumuturi (recente), în derivate şi compuse după model străin (de ex. *autocisternă, interdependenţă, supraaglomerat)*.

Accentul în propoziţie / frază are rolul de a reliefa aspecte semantice relevante. Accentul mai puternic aparţine cuvântului celui mai important:

Ion îmi aduce astăzi cartea.
Ion îmi aduce astăzi cartea.
Ion îmi aduce astăzi cartea.
Ion îmi aduce astăzi cartea.

2.2. Intonaţia

Al doilea fenomen prozodic ca importanţă îl reprezintă percepţia melodică a vorbirii. Aceasta este denumită intonaţie. Acest aspect e uşor de observat în variaţiile ascendente ale curbei F0 din propoziţiile interogative, comparativ cu variaţia descendentă din propoziţiile

exclamative. Intonaţia o folosim atunci când dorim să transmitem comentarii paralingvistice: îndoială, ironie, sarcasm.

În manifestare acustică, intonaţia este strâns legată de frecvenţa fundamentală si chiar de accent, într-o manieră greu de descifrat. Ascendenţa şi descendenţa din rostire sunt urmărite cu fidelitate de inflexiunile frecvenţei fundamentale.

Adeseori în sinteza de voce se pune problema interacţiunii frecvenţei fundamentale relativ la accent şi intonaţie. O ipoteză simplistă ar putea sugera că cele două efecte sunt cumulative. Măsurătorile au arătat insă, că situaţia reală este mult mai complexă: acestea au contrazis ipoteza, frecvenţa fundamentală prezentând o valoare medie mai redusă pentru varianta accentuată şi chiar o diferenţa mai mare, în acelaşi sens (de exemplu scădere) pentru varianta interogativă comparativ cu cea afirmativă.

Există multiple astfel de consecinţe intrigante referitor la evoluţia frecvenţei fundamentale, iar cercetările din ultimii ani au arătat că structura F0 este de o complexitate ridicată şi prezintă o variabilitate intra-vorbitor şi inter-vorbitor foarte ridicată. Mai mult, specialiştii se pot baza pe un număr relativ redus de certitudini; este adeseori neclar care segment anume al vorbirii ar trebui analizat şi care sunt premizele acceptabile pentru intonaţie.

Producerea intonaţiei este un proces descris, in principiu, din doi paşi:

- accentul sau tonul este prezis din informaţiile extrase din procesarea textului;
- tonul este folosit la generarea curbei frecvenţei fundamentale.

Intonaţia a devenit stereotipă mai ales în cazul anumitor tipuri de interogative (o categorie redusă de propoziţii şi fraze), când este urcătoare. Alte tipuri de pattern-uri intonaționale:

- silaba accentuată este rostită de obicei pe un ton mai înalt decât cea neaccentuată.
- chiar când cuvântul iniţial este accentuat pe prima silabă, tonul se menţine egal pe următoarele silabe accentuate, pentru a urca în continuare.
- când silaba iniţială poartă accentul propoziţiei / frazei, primul ton este înalt, iar următoarele sunt coborâtoare.
- intonaţia descendentă este stereotipă la sintagmele negative (enunţiative, exclamative sau interogative) care încep cu un pronume interogativ, adverb relativ, interjecţie: *Ce faci astăzi? Unde te duci la vară? Vai, ce bine-mi pare!*
- vorbirea în limba română are două tipuri principale de intonaţie iniţială: a) ascendentă, care începe cu ton relativ coborât (frecventă); b) descendentă – porneşte de la un ton relativ înalt.
- modificări ale intonaţiei pot să apară numai din motive afective:

  Nu vrei să ştii!

  Ce albastru-i cerul!

  Ce prostie!

2.3. Ritmul

Un al treilea tip de pattern prozodic se concretizează în variaţiile de viteză ale vorbirii, variaţii care lasă loc la diferite interpretări perceptuale în funcţie de mărimea ariei afectate de variaţie. Astfel, dacă întreaga rostire se face cu o viteză redusă sau la viteză ridicată, asta va corespunde unei modificări de ritm sau de rată de vorbire.

Dacă însă, variaţiile au un aspect local, efectul de durată se corelează cel mai probabil cu accentuarea. Deci creşterea duratei unui cuvânt sau a unor silabe ale acestuia, precum şi diminuarea puterii de rostire a unui cuvânt dintr-o frază, va fi un indiciu pentru fenomenul prozodic *accent*.

În ceea ce priveşte sinteza automată a vorbirii, trebuie notat faptul că nici ratele de variaţie locale, nici cele globale nu vor avea o variaţie liniară, iar de aici vor rezulta anumite limitări în posibilitatea de producere a sunetelor naturale printr-un mecanism de accelerare sau încetinire a ratei vorbirii pe baza unui parametru fix. O observaţie importantă ar fi aceea că modificarea ratei de vorbire va afecta mai mult segmentele vocale decât consoanele.

Durata fonetică este generată în cele mai multe tipuri de sisteme de sinteză din reprezentarea simbolică bazată pe arbori de clasificare si regresie. Construcţia acestor arbori este realizată pe baza următoarelor principii:

- la nivelul fonemei:
  - fonemul curent;
  - clasa acestuia;
  - poziţia acestuia în silabă,
- la nivelul silabei:
  - tipul silabei (de exemplu: CV, CVC, unde C = consoană, V = vocală);
  - tipul accentului acesteia;
  - dimensiunea în foneme.
- la nivelul contextual al fonemelor:
  - clasa fonetică a următorului fonem;
- la nivelul ritmului:
  - poziţia în silabe a ultimului accent.

Sarcina componentei pentru controlul duratei este de a realiza structura temporală care cuprinde accentuările, in corelaţie cu intonaţia. Această sarcină este foarte grea din motiv că durata este afectată de mai mulţi factori, iar efectul integral al acestora este foarte complex. Pe de altă parte, varietatea cazurilor de modificare a duratei într-o limbă este vastă. Din acest ultim motiv se cunoaşte foarte puţin despre procesele responsabile in controlul duratei durată.

Componenta de durată a unui sistem TTS este convenţională. De exemplu, durata segmentală. Se pot alege însă şi durate sub-segmentale, cum ar fi o perioadă de semnal, în cazul fonemelor sonore. La nivel sub-segmental se foloseşte modificarea neliniară în timp a vocalelor şi modificarea neuniformă a consoanelor. Mai apare încă o problemă suplimentară: aceea că în vorbirea naturală vocalele nu sunt întotdeauna prelungite la mijloc. Pentru modificarea duratei se apelează, în multe cazuri, la efecte privind modificarea ratei vorbirii.

2.4. Alte aspecte ale prozodiei

Tonul: În limbile tonale anumite cuvinte se vor distinge de altele printr-o diferenţiere a direcţiei de variaţie şi a conturului frecvenţei fundamentale. În aceste cazuri, semnificaţiile cuvintelor se stabilesc în funcţie de intonaţie.

Momente de legătură: La tranziţiile între cuvinte există adeseori reguli precise de poziţionare a accentului şi a pauzelor. S-a constatat că pauzele dintre cuvintele unei propoziţii sau fraze au tendinţa de a se lungi spre sfârşitul acestora.

Punctul final: O modalitate de estimare doar din semnalul vocal a sfârșitului unei propoziții se bazează pe pauza dintre pitch-uri. Se poate face o combinare între acest algoritm şi un model lingvistic prin stabilirea unui prag pentru probabilitatea de apariţie a sfârşitului de propoziţie şi prin stabilirea, prin antrenare, a unor valori ale pauzei.

Exista pattern-uri prozodice universale? O întrebare care revine mereu în contextul procesărilor de semnal vocal este măsura în care anumite componente de sinteză dintr-o anumită limba pot fi aplicate unei alte limbi. Desigur, unele fenomene prozodice operează aproximativ la fel în toate limbile fiind probabil înrădăcinate în psihicul uman , însă altele se comportă total diferit de la o limbă la alta. În consecinţa, nu există un model universal valabil pentru sinteza din text a vorbirii pentru diferite limbi. Sinteza trebuie adaptată şi implementată in conformitate cu sistemul fonetic şi cu regulile lingvistice ale limbii în cauză.

3. Manifestarea prozodiei în parametrii semnalului vocal

Această secțiune prezintă rezultatele experimentale privind variația parametrilor prozodici frecvență fundamentală pentru vocale, frecvența fundamentală în funcție de accent, frecvență fundamentală în funcție de intonația din propoziție, variația frecvenței formanților pentru diferiți vorbitori, respectiv rolul duratei și a pauzelor în modelarea pattern-urilor prozodice.

Analiza s-a realizat pe un corpus de semnal vocal înregistrat în acest scop. Experimentele efectuate pun în evidență dificultăţile în încercarea de a stabili modele de variaţie pentru parametrul frecvenţă fundamentală a semnalului vocal funcţie de parametrii prozodici.

3.1. Frecvența fundamentală a unităților sonore din semnalul vocal

Prin acest experiment s-a urmărit obţinerea de informaţii referitoare la frecvenţele fundamentale ale vocalelor şi grupurilor primare de vocale diftongi şi triftongi din limba română (Tabel 3.1 / 3.2). Pentru aceasta se vor efectua o serie de măsurători pe înregistrările făcute pentru parametrii: pitch (frecvenţa fundamentală F0) minim, maxim, mediu şi intensitate. Se urmăreşte apoi determinarea unui model de variaţie a acestor mărimi atunci când unităţile fonetice în discuţie nu mai sunt considerate izolate, ci sunt încadrate în cuvinte, respectiv fraze.

Tabelul 3.1. Variaţia F_0 pentru vocalele unui vorbitor feminin

Parametru	Vocala						
	/ a /	/ e /	/ i /	/ o /	/ u /	/ ă /	/ î /
F_0 minim [Hz]	*161*	*175*	*207*	*103*	*100*	*203*	*220*
F_0 mediu [Hz]	*186*	*191*	*214*	*197*	*135*	*224*	*276*
F_0 maxim [Hz]	*261*	*217*	*229*	*232*	*235*	*273*	*289*
Intensitate [dB]	*73*	*60*	*68*	*72*	*80*	*80*	*67*

Se expun în continuare rezultatele măsurătorilor efectuate pentru o serie de diftongi rostiţi ca atare, sau izolaţi în paralel cu măsurătorile efectuate pe cuvinte care conţin aceşti diftongi.

Tabelul 3.2. Variaţia F_0 pentru diftongi / triftongi în pronunție izolată, respectiv în context

Parametru	Diftong / Triftong						
	/au/	/ae/	/ai/	/oi/	/ie/	/aie/	/iau/
F_0 minim [Hz] - izolat	*129*	*157*	*168*	*93*	*128*	*168*	*203*
F_0 minim [Hz] - context	*176*	*174*	*167*	*170*	*137*	*163*	*175*
F_0 maxim [Hz] - izolat	*237*	*272*	*228*	*202*	*226*	*236*	*217*
F_0 maxim [Hz] – context	*217*	*227*	*229*	*222*	*202*	*258*	*243*
Intensitate prima vocală	*73*	*60*	*73*	*72*	*68*	*65*	*59*
Int a doua vocală [dB]	*80*	*73*	*68*	*68*	*60*	*67*	*67*
Intensitate context [dB]	*55*	*59*	*59*	*63*	*53*	*52*	*53*

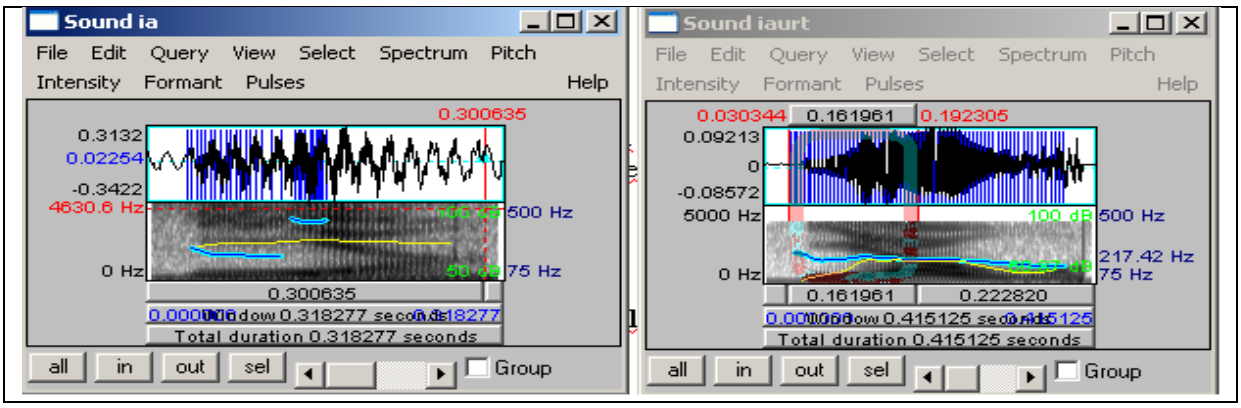

Figura 3.1. Diftongul /au/ izolat, respectiv în context

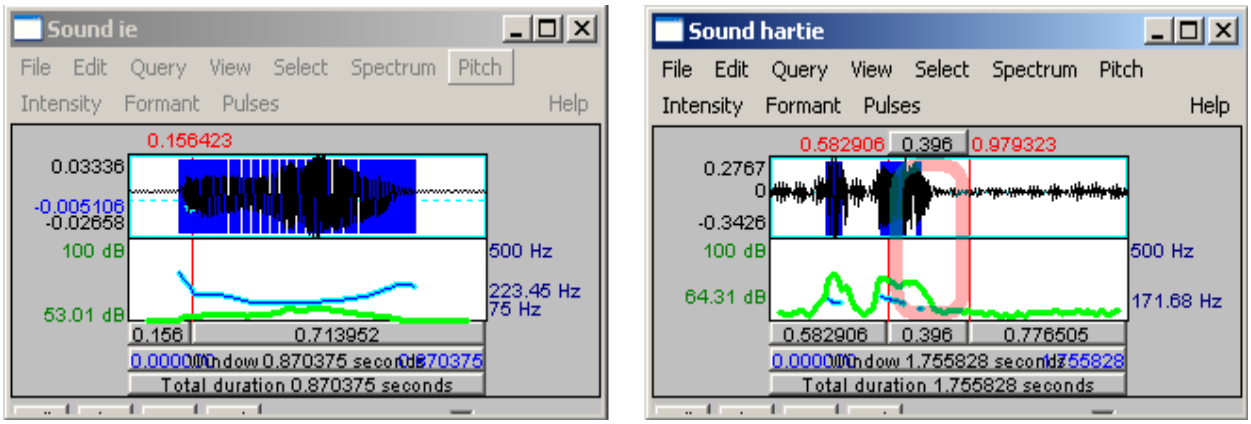

Figura 3.2. Variaţia parametrilor prozodici pentru diftongul /ie/

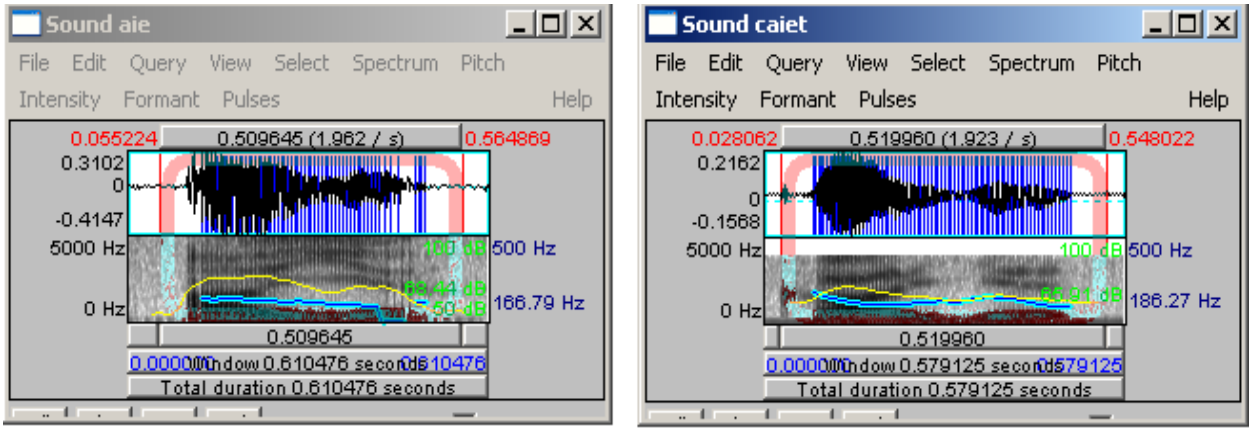

Figura 3.3. Variaţia parametrilor prozodici pentru triftongul /aie/

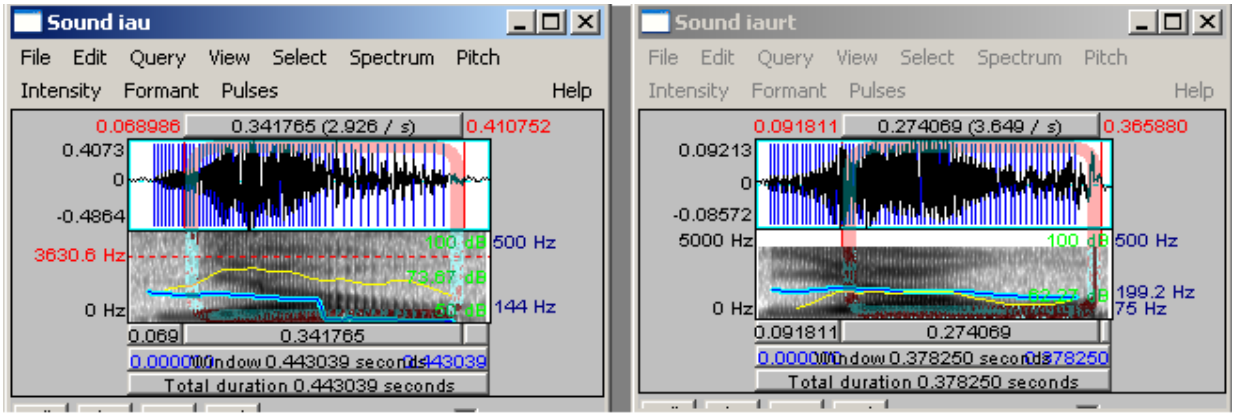

Figura 3.4. Variaţia parametrilor prozodici pentru triftongul /iau/

Concluzii:

a) Diftongi: Aşa după cum arată măsurătorile efectuate, frecvenţele fundamentale suferă variaţii atunci când diftongii (respectiv vocalele) sunt încadraţi în cuvinte; F0 maxim scade atunci când avem grupuri de vocale încadrate împreună în cuvânt, iar energia acestor diftongi încadraţi în cuvinte este sensibil mai mică decât cea a diftongilor, triftongilor izolaţi. Contează aici, desigur, şi poziţia diftongului în cadrul cuvântului, numărul de silabe ale acestuia precum şi aportul accentului. E un fapt dovedit, şi se poate uşor observa din figurile anterior amintite, că energia cuvintelor din finalul propoziţiilor este mai mică decât cele situate undeva la început sau pe poziţii centrale.

b) Triftongi: Să observăm că măsurătorile nu respectă modelul de la diftongi: se înregistrează frecvenţe sensibil mai mari pentru triftongii încadraţi în cuvinte decât cei izolaţi. În ceea ce priveşte intensitatea, aceasta păstrează tendinţa de a se diminua odată cu încadrarea în cuvinte. Ca o menţiune ce vine în suportul modelului de variaţie iniţial propus, să menţionăm că situaţiile în care grupurile de vocale diftongi şi triftongi apar ca atare sunt foarte rare, aşadar relevanţa acestor excepţii va fi minimă.

Odată cu încadrarea vocalelor, diftongilor, triftongilor etc. în cuvinte şi fraze, în general va scădea frecvenţa fundamentală şi intensitatea va creşte. Aşadar, putem afirma că domeniul de variaţie a frecvenţei fundamentale scade când avem de-a face cu grupuri de vocale încadrate comparativ cu varianta izolată.

3.2. Variația frecvenței fundamentale în funcție de accent

Se doreşte măsurarea şi determinarea frecvenţei fundamentale şi a intensităţii atunci când unităţile lingvistice sunt *afectate de accent*. În acest scop s-au efectuat înregistrări pentru o serie de cuvinte, pentru varianta accentuată şi neaccentuată a acestora şi s-au înregistrat de asemenea propoziţii cu aceste cuvinte. În continuare sunt prezentate în paralel rezultatele măsurătorilor pentru cele patru grupe de cuvinte propuse:

- factura-facturá;
- haină-haină`;
- veselă-vèsèlă;
- lumina-luminá

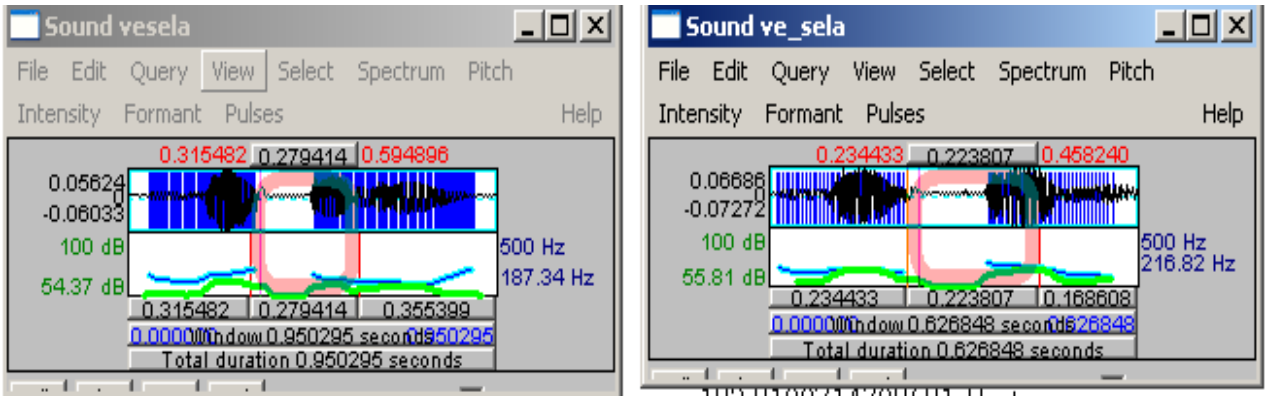

Figura 3.5. Comparaţie grafică între variaţia parametrilor prozodici /vesela/

Notă: În Figura 3.5. s-a identificat şi selectat silaba afectată de accent pentru cuvântul accentuat şi pentru cel neaccentuat. Se observă cum creste şi aici F0 (valoarea afişata in partea dreapta a capturilor reprezintă pitch-ul mediu al undei sonore pe secţiunea selectata), iar intensitatea are o tendinţă ascendentă (pentru silaba afectată de accent).

Tabelul 3.3. Tabel sintetic privind variaţia parametrilor prozodici în funcţie de accent

Cuvânt	Silabe purtătoare de accent	Valoarea de referinţă a parametrului F0 (neaccentuat) [Hz]	Valoarea lui F0 pentru silaba accentuata [Hz]	Valoarea de referinţă a parametrului intensitate[dB] (neaccentuat)	Valoarea intensităţii [dB] pe silaba accentuată
Fac**tu’**ra	tu	*172*	241	*55*	57
Factu**ra**’	ra	*195*	231	*54*	60
Hai**’**na	i	*193*	205	*54*	53
Lu’mina	lu	*199*	245	*57*	59
Lu**mi’**na	mi	*177*	257	*55*	59
Ve**se**’la	se	*183*	201	*57*	59

Conform studiilor realizate pană acum s-a arătat *ca in general silabele accentuate au tendinţa de a avea frecventa fundamentala, durata si amplitudine mai ridicata decât silabele neaccentuate*. Insa, exista si cazuri in care numai unul sau doi din aceşti parametri este mai ridicat, precum si situaţii in care tendinţa silabelor accentuate este de a-si reduce fundamentala sau ceilalţi parametrii.

Pentru fiecare din înregistrările făcute s-a realizat analiza si prelucrare separata. Modul de lucru a presupus mai mulţi paşi. S-a luat o înregistrare a unui cuvânt si s-a determinat pentru aceasta forma de unda, frecventa fundamentala si anvelopa. Mai apoi s-a luat fiecare cuvânt in parte si din acesta au fost extrase silabele accentuate, respectiv neaccentuate. Pentru ca rezultatele sa fie mai clare au fost analizate vocalele din cadrul silabelor extrase.

Pentru a fi mai clar vom ilustra cu un exemplu modul de lucru. Vom alege cuvântul *gazele* care poate avea accent pe prima silaba (când se refera la un element chimic - gaz), respectiv pe cea de a doua silaba (and este pluralul cuvântului *gazela* - animal). Pentru început vom prezenta formele de unda pentru cele doua cazuri amintite (Figura 3.5 / 3.6).

Pe Figura 3.5. se pot distinge patru grafice. Primul, respectiv al treilea sunt formele de unda ale celor doua cazuri luate in discuţie, iar al doilea, respectiv al patrulea reprezintă anvelopele de energie corespunzătoare undelor. In prima imagine accentul cade pe prima silaba ga. In imaginea a treia se poate distinge aceeaşi silaba ga, dar neaccentuata. Chiar si la o analiza superficiala se pot distinge anumite caracteristici pentru cele doua cazuri distincte. Amplitudinea si durata in primul caz sunt sensibil mai ridicate decât in cazul cu aceeaşi silaba neaccentuata. Daca discutam aceeaşi situaţie in cazul silabei ze vom observa ca situaţiile se inversează pentru ca de aceasta data in prima imagine silaba nu e accentuata, respectiv în imaginea a treia accentul cade pe aceasta silaba. Si de aceasta data se pot observa relativ uşor o amplitudine mai ridicata si o durata mai mare in cazul silabei accentuate.

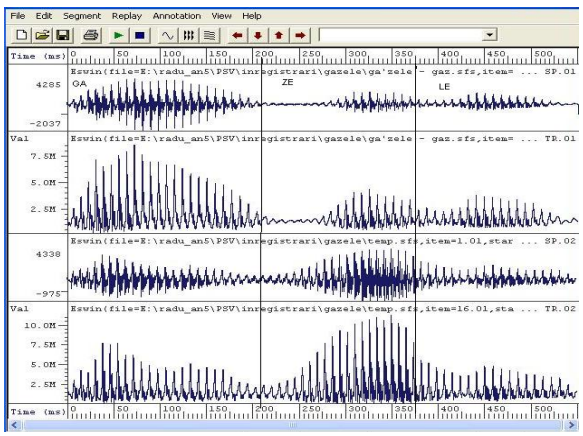

Figura 3.5. Formele de unda şi anvelopa pentru cele două cuvinte.

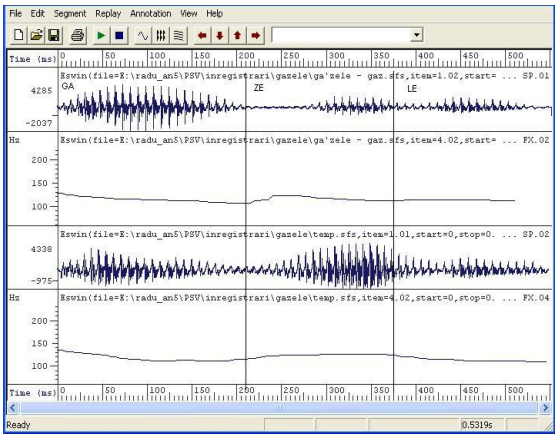

Figura 3.6. Formele de undă şi variaţia frecvenţei fundamentale

Pentru a putea vedea si evoluţia frecventei fundamentale se va mai adăuga o imagine cu cele doua forme de unda pentru cele doua cazuri distincte (imagine dreapta). Se poate observa cum frecventa fundamentala este mai ridica pe silaba accentuata. O analiza mai detaliata a adus mai multe informaţii care vor fi amintite ceva mai târziu. Daca ne concentram si asupra ultimei silabe, rezultatele sunt conform aşteptărilor, in sensul in care intre cele doua cazuri exista foarte mici deosebiri legate de durata, amplitudine sau frecventa fundamentala.

Pentru determinarea mai exacta a parametrilor importanţi (frecventa fundamentala, durata, amplitudine) fiecare cuvânt a fost segmentat in silabe. Din fiecare cuvânt au fost reţinute silabele (sau vocalele din cadrul silabelor) care aveau relevanta in analiza, adică acele silabe care intr-un caz sunt cu accent, respectiv fără accent (vezi Figura 3.7).

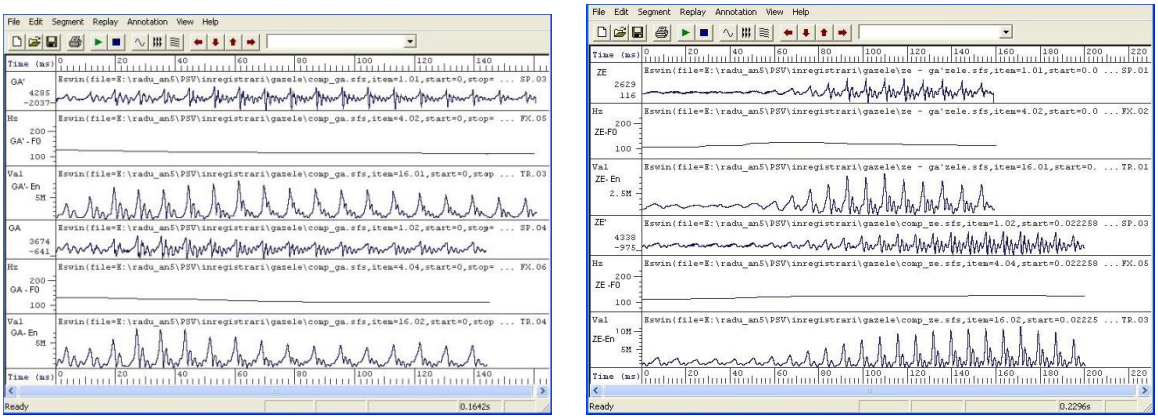

Figura 3.7. Analiza prozodica a silabelor /ga/ (stânga), respectiv /ze/ (dreapta), cu accent (sus), respectiv fără accent (jos)

Si in acest caz putem distinge o durata mai ridicata a silabei accentuate (‘ze’). In acest caz la un studiu mai detaliat s-a determinat si o creştere a frecventei purtătoare si a amplitudinii. Daca privim din nou formele de unda prezentate, vom putea observa încă un lucru care merita a fi subliniat, si anume faptul ca in cadrul aceluiaşi cuvânt putem distinge din forma de unda care silaba este accentuata, datorita amplitudinii mai ridicate a acesteia fata de silabele neaccentuate.

În experimentele anterioare s-a urmărit determinarea modelului de variaţie a frecvenţei fundamentale în funcţie de accent. Astfel s-a ajuns la un acord comun în ceea ce priveşte câteva reguli privind distribuţia si sistematica a accentului, reguli confirmate de rezultatele experimentelor de mai sus :

- deşi domeniul de influenţa al accentului poate să acopere cuvinte în întregime, fraze sau propoziţii, va exista întotdeauna o singură silabă care va susţine accentul.
- în procedura de identificare a silabei afectate de accent dintr-un cuvânt, frază sau propoziţie, accentul principal se va determina detaşat de celelalte silabe cu accent.
- accentul unei fraze sau al unei propoziţii va coincide în general cu cel al unui cuvânt; diferitele niveluri de accent se vor exclude şi susţine reciproc.

Din analiza înregistrărilor din baza de date se pot desprinde aspecte calitative legate de elementele prozodice analizate. O parte dintre acestea sunt cele legate de variaţia frecventei fundamentale in cazul unei vocale sau difonem. Astfel in cadrul unor structuri analizate (vocala, silaba, cuvânt) am întâlnit situaţii când frecventa fundamentala avea valori cuprinse intre 128 si 156 Hz si este clar ca simpla mediere a unui număr de 10 valori din acest interval nu garantează un punct de pornire de buna calitate pentru sinteza vocii.

In cazul vocalelor accentuate (a – in exemplul următor) se constata o creştere pentru Fo cum ar fi in cazul cuvântului “*a*cele”. Daca este rostit cu sensul de “acele de cusut” atunci Fo=137 Hz, iar in cazul rostirii ca si pronume (“acele persoane”) Fo va scădea la 125 Hz. Am

putea continua enumerarea de astfel de situaţii, creşterea medie a frecventei fundamentale fiind de 5- 20% in funcţie de vorbitor, cuvânt etc.

Un alt aspect observat a fost acela ca prezenta pronunţata a accentului poate modifica evoluţia conturului frecventei fundamentale a întregului cuvânt. Cazurile elocvente au fost acelea când vocala sau silaba accentuata era situata la mijlocul sau in a doua parte a cuvântului. In acest caz s-a observat o evoluţie crescătoare a valorilor frecventei fundamentale cu puţin înainte de începutul zonei accentuate si sfârşind odată cu începerea unui nou cuvânt sau silaba. O tendinţă asemănătoare, dar in sens descrescător, se constata in cazul prezentei accentului la începutul cuvântului sau in prima jumătate a acestuia.

Una din concluziile importante ale acestui studiu se referă la o creştere a frecventei fundamentale pentru silabele (sau vocalele) accentuate, fata de cele neaccentuate in medie cu 5%..20% (în 90% din cazuri creşterea s-a plasat in intervalul 9%..12%). Merita făcuta si observaţia ca au existat si câteva cazuri in care accentuarea unei silabe nu a adus nici un fel de diferenţiere din punctul de vedere al valorii F0.

Un aspect frecvent in studierea formelor de unda a fost creşterea amplitudinii semnalului pe silaba accentuata in cadrul cuvântului. In acest fel, de cele mai multe ori *silaba accentuata se distinge intr-un cuvânt prin nivelul cel mai ridicat al amplitudinii*.

3.3. Frecvența fundamentală în funcție de intonația propoziției

În acest experiment s-a pus problema modului de variaţie a parametrilor semnalului vocal în funcţie de intonaţie. Pentru a păstra o continuitate fată de celelalte experimente, s-au efectuat înregistrări pentru propoziţii construite în jurul celor 4 grupe de cuvinte folosite la experimentele anterioare, de data aceasta fiind afectate de intonaţie. Se studiază astfel, în paralel, influenţa accentului şi a intonaţiei, prin comparaţii ale mărimilor obţinute pentru cuvintele neaccentuate, accentuate respectiv accentuate şi intonate. Acest aspect are o importanţa majoră dacă ne gândim la implicaţiile sale în procesul de sinteză de voce, unde pentru a forma spre exemplu o frază, nu va fi suficient să asamblăm cuvinte izolate.

In acest scop, s-au separat cuvintele accentuate: *lumina, factura, haină şi veselă* din propoziţii şi s-au comparat cu cele înregistrate în afara vreunui context (Tabel 3.4).

Tabelul 3.4. Variaţia frecventei fundamentale în funcţie de contextul cuvântului

Cuvântul	Măsurători pentru cuvintele rostite izolat [Hz]			Măsurători pentru cuvintele selectate din propoziţii [Hz]		
	Pitch min.	Pitch.max.	variaţia	Pitch min	Pitch max.	variaţia
Lumina	*157*	*226*	*+ 68*	*181*	*236*	*+ 55*
Factura	*155*	*253*	*+ 97*	*142*	*231*	*+ 88*
Haină	*144*	*255*	*+ 110*	*151*	*230*	*+ 78*
Veselă	*152*	*227*	*+ 74*	*131*	*227*	*+ 96*

În urma măsurătorilor anterioare, putem observa diferenţe minore in ceea ce priveşte valorile maxime ale lui F0 in cadrul cuvintelor rostite separat comparativ cu cele selectate din cadrul unor fraze. În acest context e important sa observam ca gama de variaţie a parametrului F0 pentru cuvintele izolate va fi sensibil mai mare decât pentru cele extrase din context.

În manifestarea sa acustică intonaţia este în strânsă legătură cu frecvenţa fundamentală F0. Astfel, intonaţia ascendentă caracteristică secvenţei finale a interogaţiilor, la fel ca cea descendentă ce caracterizează exclamaţiile, sunt foarte bine surprinse de inflexiunile frecvenţei fundamentale F0.

Având în vedere rezultatele experimentelor anterioare, ne punem problema determinării modului de interacţiune între variaţiile lui F0 datorate accentului şi cele datorate intonaţiei. Se doreşte stabilirea unei reguli de variaţie a frecvenţei fundamentale în cadrul unui cuvânt afectat,

respectiv neafectat de accent şi afectat sau neafectat de intonaţie. O ipoteză simplistă ar fi aceea de-a considera cele două efecte cumulative. Măsurătorile arată însă că lucrurile au o natură mai complexă. Redăm mai jos 3 scenarii de experiementare.

Cazul 1: Să considerăm de pildă cuvântul *lumina*, pe care l-am abordat deja şi pe care-l vom privi integrat în propoziţiile:

La ora 23 felinarul nu mai lumina.
La ora 23 felinarul nu mai lumina!
La ora 23 felinarul nu mai lumina?

Vom studia tendinţele de evoluţie a pitch-ului în cadrul cuvântului lumina, pentru fiecare intonaţie în parte (forma afirmativă coincide cu cea accentuată) şi îl vom compara cu cel al cuvântului neaccentuat lumina.

Rezultate experimentale:

-forma exclamativă e descrisă de-o caracteristică descendentă a intonaţiei
-forma interogativă e descrisă de-o caracteristică ascendentă a intonaţiei
-forma afirmativă va avea o caracteristică relativ liniară pentru intonaţie
-forma accentuată a cuvintelor e caracterizată de-o scădere a frecvenţei fundamentale F0.

Notă: Se observă că intonaţia interogativă e caracterizată şi de o creştere a frecvenţei fundamentale F0.

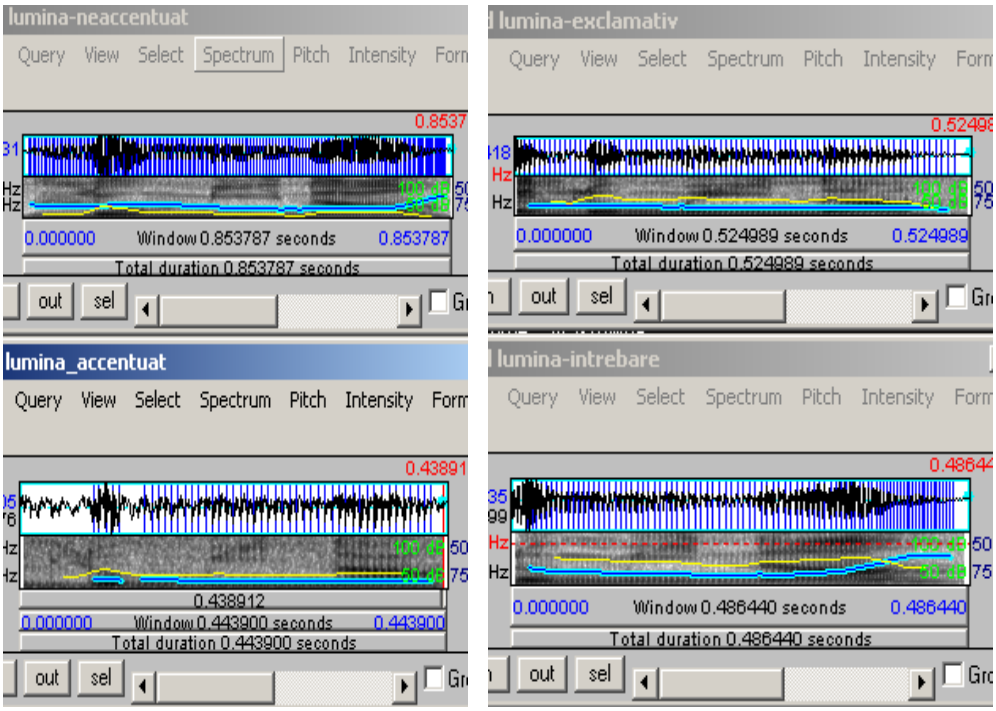

Figura 3.8. Cuvântul /lumina/ in diferite contexte intonaţionale

Măsurătorile pentru:

-cuvânt lumina neaccentuat: *186 Hz*
- cuvânt lumina accentuat(intonaţie afirmativă): *136 Hz*
- cuvânt lumina intonat interogativ: *244 Hz*
- cuvânt lumina intonat exclamativ: *150 Hz*

Cazul 2: Analiza cuvântului *haină* integrat în propoziţiile :

Nu ştia cât e de haină.
Nu ştia cât e de haină!
Nu ştia cât e de haină?

Rezultatele măsurătorilor:

- cuvânt haină neaccentuat : *179 Hz*
- cuvânt haină accentuat: *135 Hz*

- cuvânt haină accentuat cu intonaţie exclamativă: 150 Hz
- cuvânt haină accentuat cu intonaţie interogativă: 194 Hz

La fel ca în cazul anterior, frecvenţa fundamentală F0:
- scade pentru varianta accentuată comparativ cu cea neaccentuată
- creşte pentru cuvântul accentuat cu intonaţie interogativă
- creşte pentru cuvântul accentuat cu intonaţie exclamativă (deşi nu la fel de mult ca în cazul intonaţiei interogative), comparativ cu varianta afirmativă sau fără intonaţie, dar scade comparativ cu varianta accentuată.

Cazul 3: Analiza cuvântului *veselă* integrat în propoziţiile:

Am uitat de veselă.
Am uitat de veselă?
Am uitat de veselă!

Rezultatele măsurătorilor:
- cuvânt veselă neaccentuat: 189 Hz
- cuvânt veselă accentuat si cu intonaţie afirmativă: 183 Hz
- cuvânt veselă accentuat si cu intonaţie interogativă: 224 Hz
- cuvânt veselă accentuat si cu intonaţie exclamativă: 154 Hz .

Şi pentru acest caz rămân valabile observaţiile făcute la punctele anterioare, cu singura diferenţă, că frecvenţa fundamentală a cuvântului accentuat şi cu intonaţie exclamativă e mai mică decât cea a cuvântului neaccentuat.

Tabelul 3.5. Frecvenţa fundamentală în funcţie de accent şi intonaţie la nivel de cuvânt

Cuvânt analizat	F0 pentru cuvânt neaccentuat [Hz]	F0 pentru cuvânt accentuat [Hz]	F0 pentru cuvânt cu intonaţie **?** [Hz]	F0 pentru cuvânt cu intonaţie **!** [Hz]
lumina	183	*136*	*244*	*150*
haină	179	*135*	*194*	*150*
veselă	189	*183*	*224*	*154*

Un alt tip de analiza care a fost făcuta, este cea referitoare la intonaţia la nivel de propoziţie. S-a urmărit modul in care evoluează frecventa fundamentala de-a lungul unei propoziţii in doua situaţii. Prima situaţie presupunea exprimarea unei propoziţii in forma interogativa, iar apoi exprimarea sub forma exclamativa sau imperativa a aceleiasi propoziţii. S-a observat ca in cazul propoziţiilor interogative tendinţa fundamentalei este aceea de creste spre sfârşitul propoziţiei, pe când in cazul celălalt tendinţa este inversa (scădere sfârşitul propoziţiei).

In primele doua imagini sunt prezentate forma de unda, respectiv, evoluţia fundamentalei pentru enunţul “Mergeţi la şcoală?”. În imaginile trei, respectiv patru sunt reprezentate formele de unda pentru acelaşi enunţ sub forma imperativa “(Mergeţi la şcoală!”).

Analizând pe rând cele doua situaţii se poate observa cum propoziţia interogativa are o frecventa fundamentala mai ridica la sfârşit (cu tendinţa de creştere pentru F0), in timp ce propoziţia imperativa are o frecventa fundamentala ridicata la început, iar tendinţa fundamentalei este de descreştere spre finalul propoziţiei. In urma analizelor s-a observat ca diferenţa intre valorile fundamentalei la sfârşitul frazei este de aproximativ 40..50Hz (cca.25..30%), diferenţa in care fundamentala este mai mare in cazul propoziţiilor interogative.

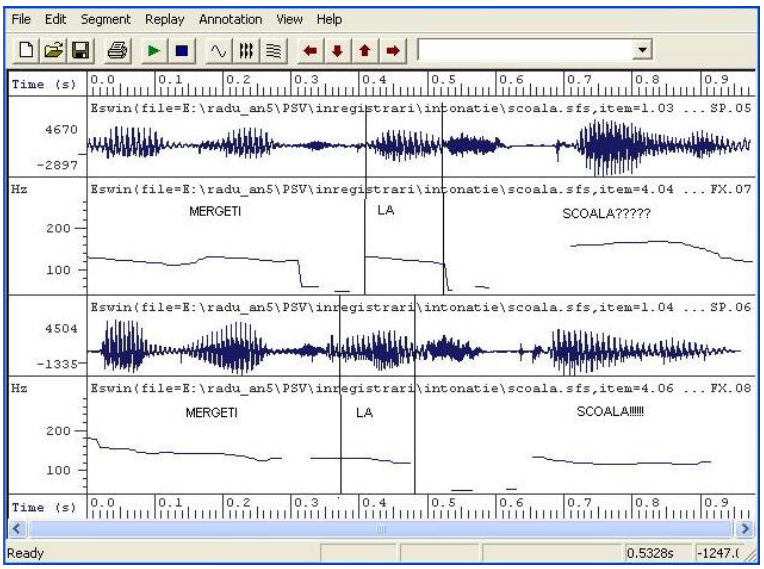

Figura 3.8. Variaţia frecvenţei fundamentale pentru propoziţii interogative (sus), respectiv exclamative(jos).

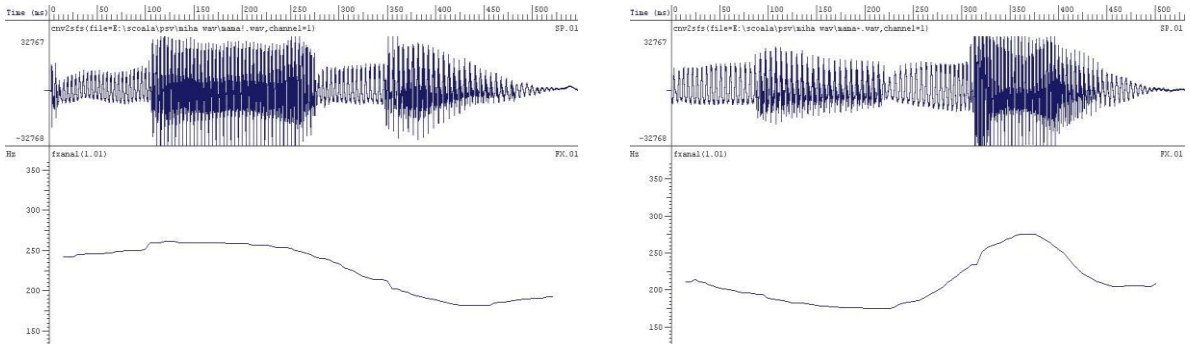

Figura 3.9. Rostire exclamativă (stânga: mama!), respectiv interogativă (dreapta: mama?)

Un alt exemplu de interpretare a măsurătorilor privind variaţia frecvenţei fundamentale este pentru cuvântul “mama”, rostit cu diferite intonaţii.

- în propoziţie declarativa frecventa fundamentala este aproape liniara de la 100% la început de cuvânt, pana la 101% si 103% spre sfârşit.
- în propoziţie exclamativa, creste la început pe primul “a” pana la 106% fata de începutul cuvântului si apoi descreşte pana la 75% la sfârşit.
- într-o propoziţie interogativa, daca la început scade de la 100% spre 85%, in a doua jumătate creste tare pana la 130% pe al doilea “a”, iar spre sfârşit scade înapoi spre 100%.

Co’pii: a) afirmativ – frecventa: scade de la 265 Hz (pe “co”) la 191 Hz pe primul “i” si isi revine pana la 202 Hz pe al doilea “i”, b) interogativ – frecventa : urca de la 223 Hz pe “co” pana la 230 Hz pe primul “i” si scade apoi pe al doilea.

Copii’: a) afirmativ – frecventa: scade de la 260 Hz pe “co” pana la 240 Hz pe primul “i” si chiar 230 Hz pe al doilea, b) interogativ - frecventa: urca de la 230 Hz pe “co”, se păstrează la 230 pe primul “i” dar urca pana la 250 Hz pe al doilea.

Copiii: a) afirmativ – frecventa: scade de la 227 Hz treptat pana la 198 Hz pana la ultimul “i” , b) interogativ – frecventa: urca de la 200 Hz pana la 275 Hz pe al doilea “i” dar coboară puţin pe al treilea.

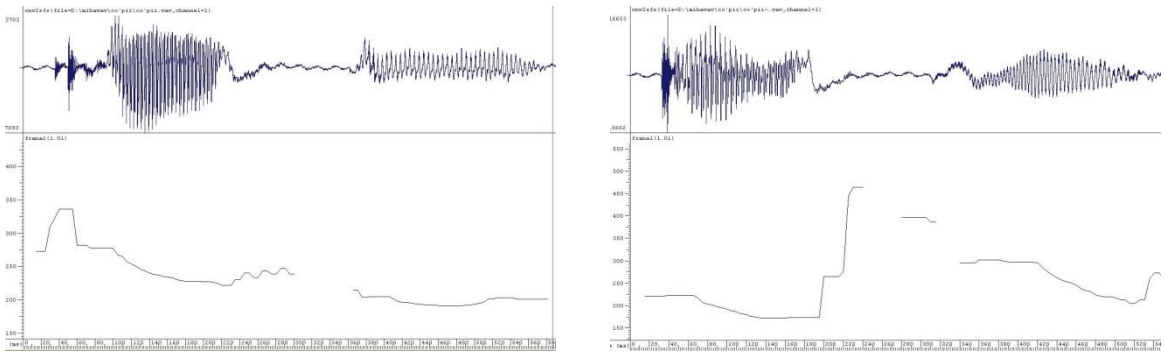

Figura 3.10. Fundamentala pentru ‘co’pii’ (afirmativ), respectiv ‘co’pii?’ (interogativ)

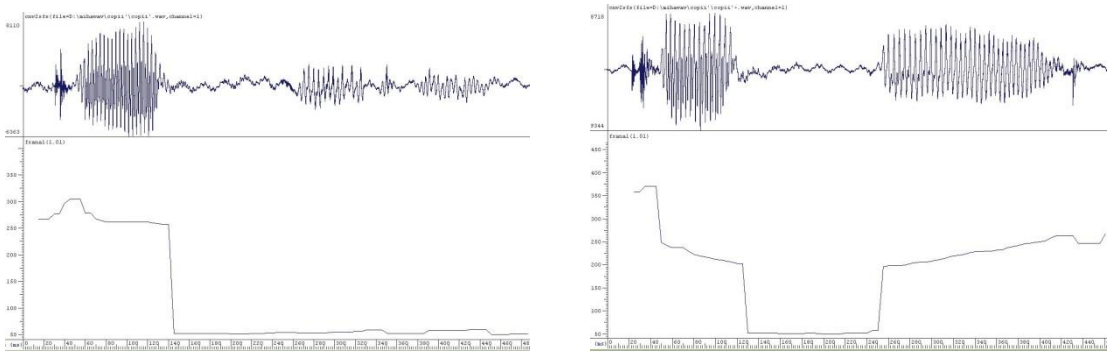

Figura 3.11. Fundamentala pentru ‘Copii’’ (afirmativ), respectiv ‘Copii’?’ (interogativ)

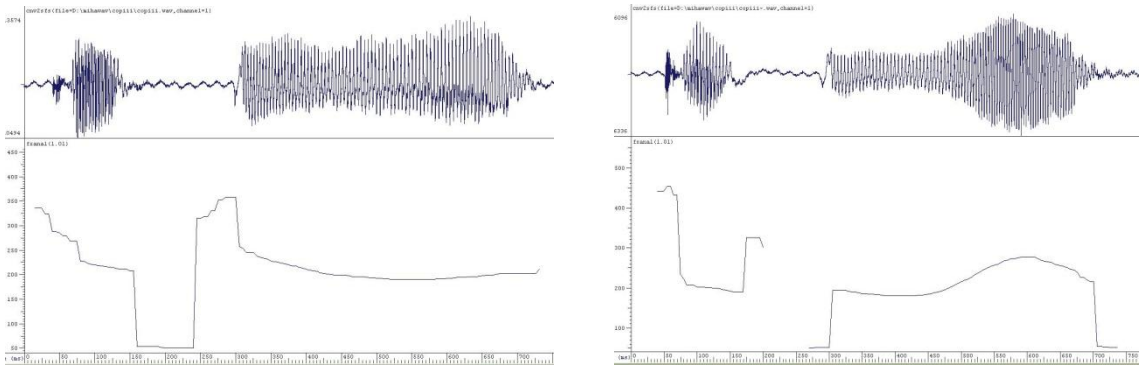

Figura 3.12. Fundamentala pentru ‘Copiii’’ (afirmativ), respectiv ‘Copiii?’’ (interogativ)

In cazul celor trei cuvinte, identice ortografic, se observa următoarele diferenţe:

- accentul in “co’pii” este clar pus pe difonemul “co” care are durata , dar si frecventa cea mai mare din cadrul întregului cuvânt, precum si fata de celelalte doua cuvinte. (fata de “copii’ ” frecventa este doar cu 0.02% insa durata este cu 50% mai mare; fata de “copiii”, care pune clar accentul pe cei trei “i”, creşterea frecventei e de 17% iar a duratei de 20%).
- durata si frecventa celor 2, respectiv 3 de “i” : la “copii” si “co’pii” diferenţa este la frecventa – crescuta cu 26%, in timp nu este diferenţă mare, dar este o mare deosebire când este rostit cu 3 ”i” când durata este mult mai lunga, iar frecventele sunt relativ înalte.
- tendinţa este la interogativ de a urca frecventa si durata pe penultima silaba sau difonema si a cobori puţin chiar la sfârşit, decât daca avem accentul pus expres pe ultima silaba (copii’) si atunci avem o linie ascendenta a frecventei pana la sfârşit.

3.4. Analiza formanților în funcție de vorbitori pe tot corpusul

Interfaţa de dialog a instrumentului software utilizat pentru analiza formanţilor este prezentată in Figura 3.13. Structura de organizare a probelor de voce: în directorul 'Date' se găsesc doua subdirectoare: 'm' si 'f' (male/female, corespunzător achiziţiilor de la vorbitorii masculini, respectiv feminini). In fiecare dintre acestea, se găsesc 7 subdirectoare, câte unul pentru fiecare vocala. Pe interfaţă trebuie sa precizam calea către date, respectiv daca e vorba de directorul 'm' sau 'f'. Se alege vocala de prelucrat, si câte probe din aceasta vocala sa fie analizate(in cazul de fata maximum 21 de probe, deoarece in directorul corespunzător fiecărei vocale sunt 21 fişiere de sunet).

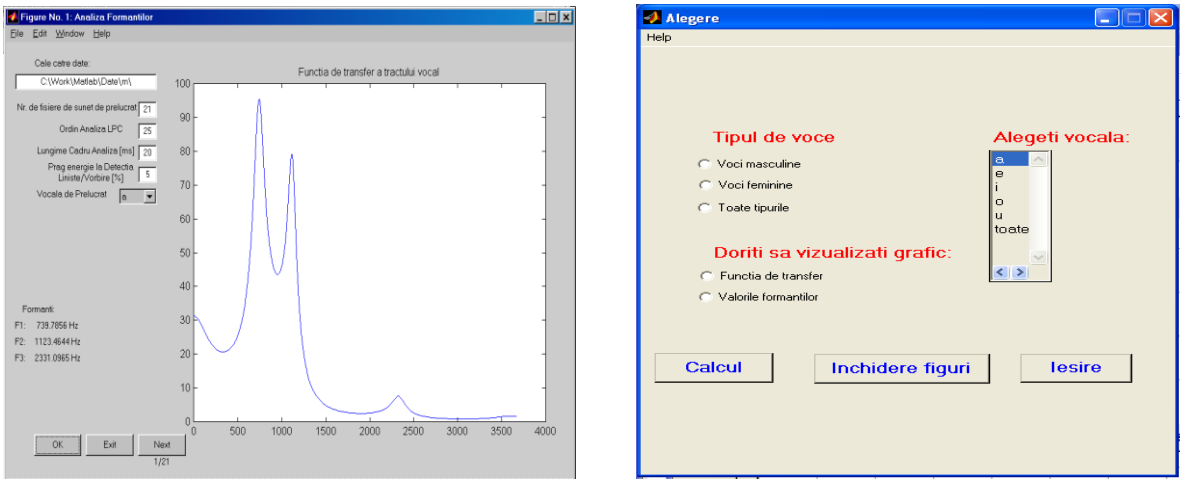

Figura 3.13. Interfeţele pentru realizarea studiului privind analiza formantică

Tabelul 3.5. Pattern-uri pentru formanții vocalelor pentru vorbitorii feminini din corpus

Vocala	F1 [Hz]			F2 [Hz]			F3 [Hz]		
	minim	mediu	maxim	minim	mediu	maxim	minim	mediu	maxim
/a/	156	827	2182	1257	1792	3135	1825	2487	4357
/e/	81	867	2247	838	1706	3032	1928	2853	4208
/i/	251	634	2158	653	2392	3227	2388	3259	4030
/o/	178	966	1937	940	1518	3328	1407	2981	3971
/u/	169	647	1942	635	1245	3293	1360	2986	4107

Tabelul 3.6 Pattern-uri pentru formanții vocalelor pentru vorbitorii masculini din corpus

Vocala	F1 [Hz]			F2 [Hz]			F3 [Hz]		
	minim	mediu	maxim	minim	mediu	maxim	minim	mediu	maxim
/a/	92	837	1979	676	1735	3043	1621	2523	4106
/e/	87	791	2366	441	1951	3217	1914	1951	3217
/i/	92	634	2158	653	2392	3227	2388	3259	4030
/o/	84	761	1718	526	1321	2924	1028	2408	2845
/u/	83	574	1738	572	1227	2916	1021	2420	3980

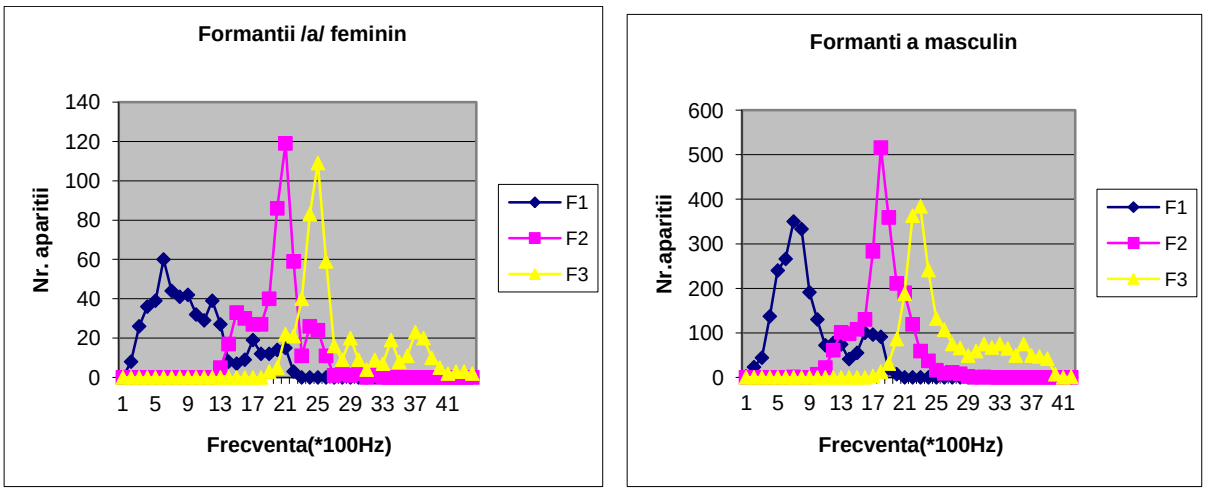

Figura 3.14. Distribuţia formanţilor pentru vocala /a/ feminin (stânga) şi masculin (dreapta)

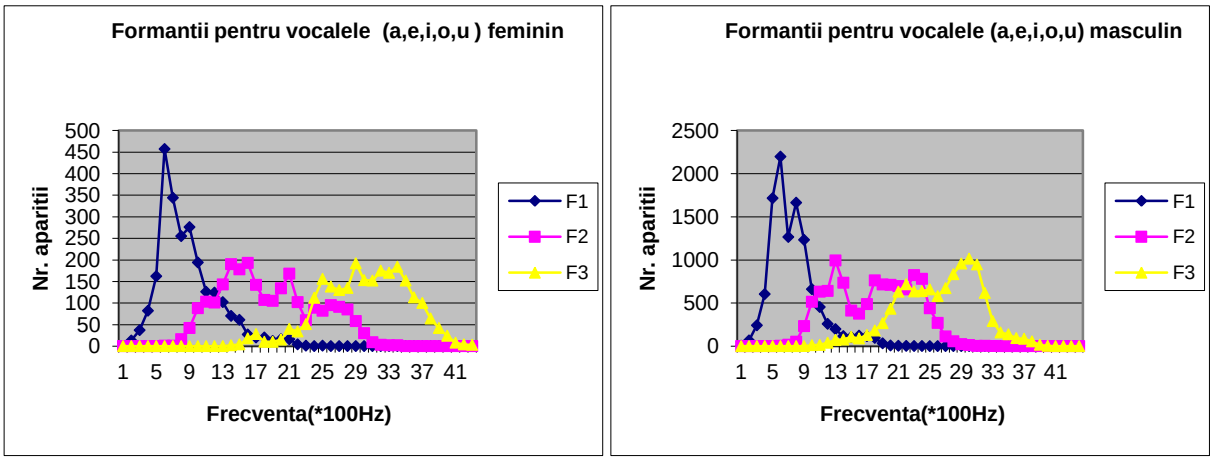

Figura 3.15. Distribuţia formanţilor pentru toate vocalele feminin (stânga) şi masculin

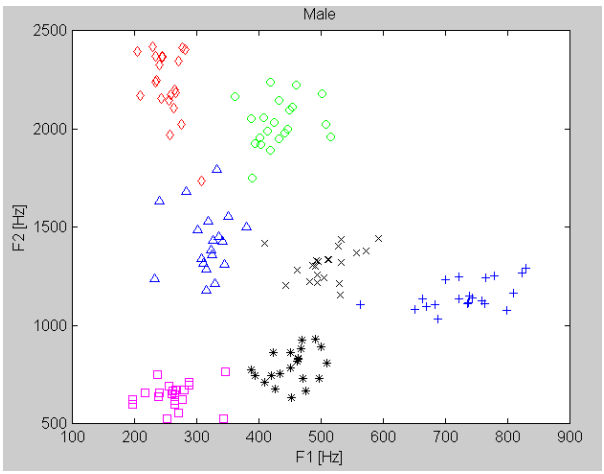

Fig. 3.16. F1-F2 pentru vorbitorii masculini

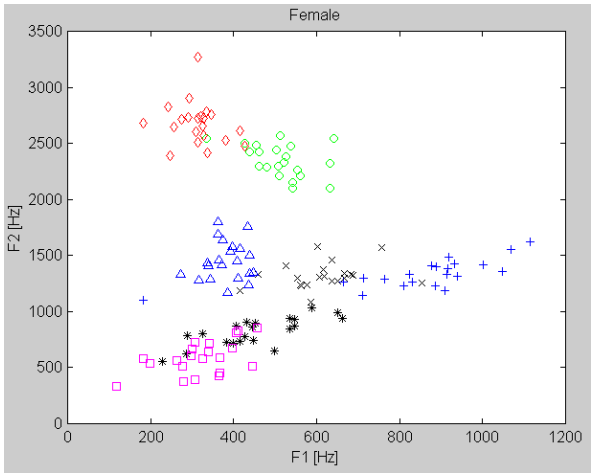

Fig. 3.17. F1-F2 pentru vorbitorii feminini

Pe lângă diferenţele de frecventa fundamentala intre o voce masculina si una feminina s-a încercat sa se observe care sunt diferenţele si la nivel de formanţi. Aceste diferenţe se pot observa in figurile de mai sus, pentru fiecare vocală în parte, dar şi pentru toate vocale (Fig. 3.14 – 3.17) pentru vorbitorii masculini, respectiv feminini. Ceea ce am observat este faptul ca distribuţia formanţilor la bărbaţi este mai compacta decât la femei, ‘norii’ sunt mai concentraţi.

3.5. Ritmul vorbirii și durata unităților acustice

Un alt parametru analizat in detaliu a fost durata diferitelor structuri fonetice. Aceasta se modifica in funcţie de poziţia accentului, in funcţie de intonaţia cuvintelor astfel ca de exemplu in cazul unei vocale durata poate avea valori cuprinse intre 40 si 140 ms. Din păcate aici intervine si o latura subiectiva in stabilirea acestei durate. Se ştie ca vocalele in general au forma de unda periodica si frecventa lor fundamentală se poate determina având un contur liniar – orizontal.

Având o astfel de forma de unda de este greu de sesizat din punct de vedere auditiv vreo diferenţa intre o durata de 85 ms si una de 120 ms. In schimb consoanele au o durata foarte mica de aprox. 30 ms (depinde de la caz la caz). Faptul ca exista aceste mari diferenţe de durata intre vocale si consoane poate crea unele probleme in interpretarea rezultatelor. Un lucru important si care trebuie reţinut este acela ca in cazul vocalelor accentuate durata este sensibil mai mare decât in cazul celor neaccentuate.

Un alt lucru observat in urma studierii in amănunt a înregistrărilor a fost o creştere a duratei silabelor accentuate, fata de cazul in care acestea nu erau cu accent. Creşterea duratelor se datorează in principal creşterii duratei vocalei accentuate. Aceste creşteri se situează cel mai adesea in intervalul 27%..33% pentru silabele accentuate.

Durata: propoziţiile declarative fiind „mai liniare”, timpul de pronunţie este mai mic. Explicaţie posibila: deschiderea gurii pentru vocale este mai mica intr-o propoziţie afirmativa fata de o propoziţie exclamativa sau interogativa – 260ms pentru a rosti simplu “mama” si aproximativ 500ms atunci când exclamam sau întrebăm. Vezi tabelul de mai jos.

Tabelul 3.7 Legătura între durata cuvintelor / difonemelor şi intonaţie (frecvența fundamentală)

Unitate acustică	*Mama.*		*Mama!*		*Mama?*	
	Fo (Hz)	durata(ms)	Fo (Hz)	durata(ms)	Fo (Hz)	durata(ms)
_m	221	53	245	100	208	81
ma	222	119	260	250	180	200
am	225	113	255	235	174	215
ma	229	109	189	177	274	210
a_	229	85	185	124	205	180

Cuvântul *“aleea”* – intr-o propoziţie afirmativa are primul “e” accentuat, fapt ce reiese din durata mai lunga a difonemelor ce conţin primul “e” si a frecventei, care ajunge pana la 110% din frecventa fundamentala de la începutul cuvântului. Într-o propoziţie interogativa are acelaşi comportament ca si la “mama” – frecventa creste aproape la 157% fata de începutul cuvântului la al doilea “e”.

La cuvintele *“co’pii”* (ca si in “copii xerox”) si “copii’ ”(in sensul de “prunci”, “fii”) cei doi “i” au aceeaşi durata in difoneme. Totuşi, frecventa este crescuta cu 126% in “copii’ ” fata de “co’pii” tocmai datorita accentului care este pus diferit in cele doua cuvinte: pe “i” respectiv pe “o”.

Diferenţa dintre *“copii’ “* si *“copiii”* se observa prin creşterea duratei vocalei “i” pana aproape de doua ori atunci and e cu trei de ‘i’. Care este diferenţa intre accentul pus pentru a deosebi ‘co’pii” de “copii’ “ sau “copiii” si a desemna o afirmaţie sau o întrebare?

Un lucru interesant, a fost acela al influentei accentului nu numai asupra vocalelor ci asupra consoanelor. Este vorba de difoneme formate dintr-o o consoana si o vocala si care sunt accentuate. Durata consoanelor creste intr-o mai mica măsura decât cea a vocalelor. Daca analizam, in schimb, valorile amplitudinilor maxime, se poate constata ca valorile relative si creşterile procentuale sunt de valori apropiate, indiferent de vorbitor.

Studiind problema literei “C” am înregistrat mai multe cuvinte care conţin litera urmata de o vocala mai puţin sonora (“câini”), o vocala puternic sonora (“convex”), “c” urmat de consoana (“crâmpeie”) si cazurile speciale din grupurile “che”, “chi”, “ce”, “ci”.

Se observa ca fonemul “c” împreuna cu ‘o” are o durata relativ lunga si frecventa fundamentala mare datorita sonorităţii vocalei “o”, deşi pentru consoana “c” singura nu se poate stabili o frecventa fundamentala.

Acelaşi fenomen se înregistrează si la grupul “câ” si la “che”. Insa la grupurile “ce”, “ci” dau o alta sonoritate literei “c” tinzând sa fie mai scurte, spre un singur sunet. Acelaşi lucru se poate spune si despre litera “g”. “C” urmat de “r” nu este destul de sonor, fiind foarte scurt si fără rezonanta. Aceasta observaţie este valabila pentru toate combinaţiile de consoane nesonore (“c”,“p”,“d”, ”t”,etc).

4. Manifestarea prozodiei în caracteristici de natură lingvistică

4.1. Aspecte de natură lingvistică

Determinarea predictivă a prozodiei în sistemele TTS trebuie să se bazeze pe o analiză mai mult sau mai puţin complexă. În cazul cel mai simplu, intonarea poate consta numai din rostirea cuvintelor disparate care constituie enunţurile simple declarative neutre, unde curba intonaţională ţine cont doar de poziţionarea accentului din fiecare cuvânt, iar cuvintele sunt rostite cu intervale constante. O analiză mai evoluată (gramaticală) pune în evidenţă grupurile de cuvinte ce reprezintă diferite locuţiuni: substantive calificate prin adjective, substantive compuse realizate prin prepoziţii, locuţiuni verbale, timpuri compuse etc. La acestea, prin reducerea intervalelor dintre cuvintele constitutive, prin modificarea accentelor originale etc., se poate obţine o curbă intonaţională mai aproape de cea naturală.

Până în acest moment, predicţia prozodică se poate realiza bazându-se exclusiv pe informaţiile codificate în textul pur. Primul nivel de specificare a prozodiei în scris se realizează prin semnele de punctuaţie care au influenţă specifică asupra liniei melodice la rostirea textului scris. De aici încolo, orice modificare dorită a prozodiei trebuie marcată efectiv şi specific – vezi marcajele subiective ale recitatorilor concreţi pe textele dramatice sau poetice, respectiv notele muzicale recitative.

Chiar şi din această prezentare succintă rezultă că prozodia este un fenomen deosebit de complex: are aspecte general valabile comunicaţiei umane, aspecte dependente de limba folosită, respectiv aspecte strict legate de subiecţii vorbitori şi de intenţiile comunicării unor informaţii suplimentare necodificate neapărat în textele scrise. Din acest motiv la realizarea sistemelor TTS nu se mai ţinteşte realizarea unei prozodii stricte, ci a uneia care să *semene* într-o măsură acceptabilă melodiei comunicaţiei umane într-o limbă concretă, ţinându-se cont de aspectele dependente de limbă, dar nicidecum de subiecţii vorbitori.

4.2. Pattern-uri intonaționale la nivel de propoziție

Tiparul descendent declarativ neutru. Tiparul intonaţional al unei fraze declarative scurte are un aspect ascendent-descendent, indiferent de ordinea elementelor:

ma *ne*
Ma vi **Vi ma**
ne. *ma.*

În realitate cele două mişcări ale tonului sunt asimetrice: coborârea ajunge la un punct mai coborât al vocii decât cel de la care a început urcarea, aspect şi mai evident la amplificarea tiparului:

ma
Ma vi
ne
re
pe
de.

Tiparul declarativ enumerativ. Construcţiile simetrice (enumerative, repetitive) se caracterizează prin ridicări sau coborâri de intonaţie care revin aproape regulat. Intonaţia enumerativă este, în limba română ca şi în alte limbi, cea *ascendentă de continuitate*:

la *ri*
lele, *cele,*
Cumpără *flo* *bombo*
nele.

Tiparul declarativ negativ. Frazele negative sunt cazuri tipice de enunţuri cu emfază, deoarece negaţia nu poartă totdeauna accentul frazei. De obicei sunt caracterizate printr-un tipar descendent cu proeminenţă pozitivă, având un vârf melodic pe *nu*:

nu
larna *era*
apa cal
dă.

Tiparul interogativ ascendent. Acest tipar este considerat ca fiind caracteristic pentru întrebările care presupun răspunsul “da” sau “nu, unde întrebările respective pot avea o structură segmentală identică cu cea a unor enunţuri declarative, singura diferenţă faţă de acestea fiind intonaţia urcătoare a întrebărilor. Aceste întrebări se numesc *întrebări totale*.

noi?
ne la
Mama vi

Tiparul interogativ descendent. Faţă de întrebările totale, un al doilea tip interogativ cuprinde aşa numitele *întrebări parţiale*, introduse printr-un cuvânt interogativ (*care?, ce?, cine?, cum?, când?, de ce?, unde?,* etc.), caracterizate printr-un tipar melodic coborâtor.

Cine *Când*
vine *ajungeţi*
la *acasă?*
noi?

Tiparul interogativ, total negativ. În limba română adverbul negativ "*nu*" poartă, în general, accentul frazei. În acest tipar este menţinut un ton coborât până la silaba accentuată a ultimului cuvânt, unde începe o urcare abruptă. Acest platou coborât se menţine în toate întrebările totale negative, indiferent de lungimea lor

i?
o
Nu vine la n

În limba română sunt frecvente fenomenele de dublă negaţie, exprimate prin pronume ca *nimeni, nimic, nici*, sau prin adverbe ca *niciodată, nicăieri, nicicum* etc. În aceste tipare accentul frazei se mută pe acest al doilea cuvânt. Dacă acesta este plasat al doilea element negativ, tonul cel mai coborât din enunţ este atins abia pe silaba accentuată a acestuia, continuă ca atare până la ultima silabă accentuată din frază, unde se produce urcarea finală abruptă.

oi?
Nu vine
nimeni la n

Când un alt cuvânt negativ este plasat în frază înaintea lui nu, după punctul minim atins în silaba sa accentuată, tonul urcă în silaba următore, apoi se prelungeşte la acelaşi nivel până la urcarea finală, luând naştere un contur melodic în două trepte ascendente:

oi?
meni nu vine la n
Ni

Tiparul ascendent-descendent interogativ. Acest tipar reprezintă intonaţia cu care sunt rostite, în general, enunţurile care exprimă o alternativă, construite cu conjuncţii disjunctive ca *sau* şi *ori*. Aceste fraze pot avea aceeaşi structură segmentală, singurul element distinctiv fiind intonaţia.

apa
Doriți ori
suc?

băm
Ne plim sau cumpărăm ca
do
uri?

4.3. Rolul accentului în prozodie

Accentul este unul dintre cei mai importanţi parametri prozodici prin care o limbă particulară se poate distinge de alte limbi. Acest aspect se evidenţiază prin ceea ce se numeşte "*accentul străin*", când o limbă străină este rostită cu intonaţia limbii materne (de exemplu la utilizarea accentelor gramaticale). După cum se cunoaşte, accentul românesc este *expiratoric* sau *dinamic* şi are un caracter liber (adică locul său nu este fixat pe o anumită silabă a cuvintelor) şi mobil. Diferenţa de *intensitate* dintre silabele accentuate şi neaccentuate nu este prea mare şi se complementează prin diferenţă de *durată* şi de *înălţime a tonului*. Ultimele totuşi sunt aproape irelevante, deci factorul esenţial distinctiv rămâne intensitatea.

Accentul la nivel de cuvânt. Libertatea accentuării nu presupune, totuşi, posibilitatea de a deplasa în mod arbitrar accentul de pe o silabă pe alta. Mai mult, în unele cazuri de omografe, numai accentul (de altfel fixat) poate deosebi între ele cuvintele (de ex. cop**i**i – c**o**pii, m**â**nji – mânj**i**, v**e**selă – ves**e**lă etc.) sau chiar formele flexionare (s**u**nă - sun**ă** etc.). Abaterile de la

accentuarea naturală vor trăda imediat vorbitorii străini, care au tendinţa de a aplica regulile de accentuare specifice limbii lor materne. În mod obişnuit, fiecare cuvânt polisilabic din limba română are o singură culme dinamică. Această regulă poate fi respectată uşor dacă ţinem seama că majoritatea cuvintelor româneşti au un număr redus de silabe. Totuşi, la unele împrumuturi recente sau derivate şi compuse din elemente străine, apare un *accent secundar* plasat pe prima silabă a cuvântului (de ex. **a**utoapăr**a**re, c**o**ntrasemnăt**u**ră, r**e**constitu**i**re, s**u**prasolicit**a**t, dar găsim exemple şi în fondul autohton cum ar fi **u**ntdel**e**mn, b**u**năvo**i**nţă etc.). Accentul secundar nu posedă valori distinctive. Fiind însoţit însă şi de o uşoară ridicare a tonului, joacă foarte adesea în frază rolul unui accent de insistenţă. Rar, limba română mai conţine şi cuvinte cu *două accente principale* (echivalente), în compusele nesudate: d**u**pă-m**a**să etc.

Accentul în propoziție și frază. Principiul expresivităţii recomandă ca fiecare propoziţie şi fiecare frază să aibă o singură culme dinamică, cu rolul de a scoate în relief ceea ce este nou, necunoscut sau important faţă de restul enunţului. În limba română (asemeni multor alte limbi) acest accent poate sta pe oricare membru al propoziţiei sau al frazei. Dacă în cadrul frazei o propoziţie este accentuată, accentul cel mai puternic cade pe predicatul propoziţiei, dar pot fi accentuate şi alte cuvinte din propoziţie. Pentru accentuarea propoziţiei, una dintre silabele accentuate ale cuvintelor frazei primeşte o forţă expiratorie mărită.

De exemplu, în propoziţia *Mama vine mâine la mine* fiecare cuvânt polisilabic îşi păstrează accentul propriu, dar unul este supus accentuării mai puternice faţă de celelalte (depinzând de tipul de întrebare la care trebuie să răspundă propoziţia. După cum se vede, accentul frazei se opune celorlalte accente, indicând adevăratul sens al enunţului.

1. (cine vine?) **Mama** vine mâine la mine.
2. (ce face?) Mama **vine** mâine la mine.
3. (când vine?) Mama vine **mâine** la mine.
4. (unde vine?) Mama vine mâine **la mine**.

sau

1. (cine a făcut?) **Noi** am făcut toate temele.
2. (ce se întâmplă?) Noi **am făcut** toate temele. ***
3. (ce am făcut?) Noi am făcut toate **temele**.
4. (care teme?) Noi am făcut **toate** temele.

O sinteză a modalităților de accentuare.

1) *Propoziţiile enunţiative* (excepţie fac exclamativele) nu au în mod obligatoriu un cuvânt mai accentuat faţă de celelalte.

2) *Propoziţiile interogative* au de obicei un cuvânt mai accentuat, prin care se arată la ce segment se aşteaptă răspuns, iar în enunţiativele ce răspund sau aprobă spusele interlocutorului, accentul cade pe adverbul de întărire. De ex. **Aşa** cred / **Sigur** e un băiat.

3) *Cuvintele de interogaţie* se accentuează totdeauna:

a) pronume interogative: ***Cine** trece pe stradă?**, De ce** nu vii?, **Cui** să-i spun?*

b) adverbe interogative: ***Cum** să-i răspund?, **Unde** să plec?, **Când** te-ntorci?*

4) *În propoziţiile negative: **Nu** călcaţi iarba!, El **nu** bea, **nu** fumează, **nu** e risipitor.*

5) *Conjuncţiile**, Ori** tăia lemne, **ori** căra apă, mereu lucra. **Acum** râde, **acum** plânge.*

6) *Prepoziţiile s*unt neaccentuate, excepţie făcând în anumite situaţii: *Mergea **după** el, niciodată înaintea lui. **Pentru** numele Domnului!*

7) *Pronumele:* Sunt de obicei neaccentuate în propoziţie.

8) *Propoziţiile exclamative: **Cât de** bine-mi pare! **Of,** ce necaz!*

9) *Accentul în frază: P*ropoziţia accentuată poate fi principală sau secundară. Propoziţiile coordonate de obicei nu diferă prin accentuare, fiind puse de vorbitor pe acelaşi plan, mai ales

la coordonarea copulativă (propoziţii legate prin conj. ***şi***). Propoziţiile disjunctive sunt de obicei ambele accentuate, fiind puse de vorbitor pe acelaşi plan: *Hotărăşte-te ce faci,* ***sau pleci, sau rămâi. L****a* adversative poate fi accentuată propoziţia a doua: *Nu plânge,* ***ci se preface.*** La fel la coordonarea conclusivă - *E vreme foarte urâtă,* ***deci rămân acasă.*** În cazul frazelor formate prin subordonare, în principiu orice propoziţie, regentă sau subordonată, poate fi accentuată. (Cf. *Gramatica limbii române,* ed. a II-a revăzută şi adăugită, Bucureşti, 1963, vol. II, p. 462-478)

4.4. Rolul semnelor de punctuație în prozodie

Valorile semnelor de punctuaţie sunt în primul rând semantice (aşa cum o dovedesc şi denumirile unora dintre ele: “semn de întrebare”, “semn de exclamare”, “puncte de suspensie” etc.); numai în măsura în care aceste categorii se manifestă în plan prozodic, semnele de punctuaţie redau şi prozodia.

Punctul: Aserţiunile neutre, care îmbracă forma unor fraze enunţiative, sunt rostite, în general, cu o intonaţie continuu descendentă, reprezentată prin tiparul descendent declarativ. Acest tipar şi pauza care urmează sunt indicate prin punct.

Semnul întrebării: Este folosit în scriere pentru a marca intonaţia frazelor interogative, dar se aminteşte că în limba română există cel puţin trei tipare fundamentale cu valoare interogativă: tiparul ascendent interogativ, tiparul descendent interogativ, respectiv tiparul ascendent-descendent interogativ.

Semnul exclamării: Folosit pentru propoziţii enunţiative afective, acest semn de punctuaţie are influenţa poate cea mai vagă asupra prozodiei.

Punctele de suspensie: Acest semn redă, în general întreruperea vorbirii: dislocarea frazei, pauzele, ezitările, vorbire sacadată etc.

Virgula. Marchează pauzele semnificative din interiorul enunţurilor, având o valoare cuprinsă între valorile pauzelor reprezentând pauza de sfârşit de enunţ şi pauza normală dintre cuvinte sau sintagme. În majoritatea cazurilor reprezintă delimitatorul între componentele unei înşiruiri, respectiv delimitează o subordonată în cadrul unei fraze.

Semnele două puncte respectiv punct şi virgulă. Prozodic vor fi echivalate virgulei, deci vor marca o pauză semnificativă, dar mai mică decât cea produsă de punct.

Parantezele. La prima aproximaţie le vom echivala cu efectul virgulei, urmând să revenim asupra definiţiei în urma rezultatelor experienţelor acustice.

Cratima: Din punct de vedere prozodic, cratima redă pronunţarea “legată” a unor cuvinte care pot avea sau nu şi existenţă independentă, notând o realitate *fonetică permanentă* sau *accidentală*. Această legare este însoţită uneori şi de anumite modificări fonetice: *sinereza* şi/sau *eliziunea*. *Sinereza* înseamnă pronunţarea accidentală într-o singură silabă a vocalei finale a unui cuvânt şi a vocalei iniţiale a cuvântului următor, deci transformarea unui hiat în diftong. Fenomenul acesta poate fi obligatoriu (de ex. *mi-a spus*, *ne-am dat* etc.) sau facultativă, redând în forma scrisă tempoul rapid al rostirii (de ex. *de-abia*) faţă de tempoul lent (*de abia*). *Eliziunea* înseamnă căderea accidentală a vocalei neaccentuate de la finala unui cuvânt în contact cu vocala iniţială a cuvântului următor. Şi aceasta poate fi obligatorie (de ex. *dintr-un*, *printr-o*, *m-a văzut* etc.) sau facultativă, diferenţiind în scris tempoul rapid faţă de cel lent al rostirii (de ex. *c-a văzut -> că a văzut*, *c-un copil -> cu un copil* etc.). Fără funcţie gramaticală cratima se poate utiliza şi la redarea în scris rostirea în tempo rapid a derivatelor cu prefixele *ne-* şi *re-* de la temele care încep cu *îm-*, şi *în-***,** notând afereza vocalei **î-** la începutul cuvintelor de bază (de ex. *ne-ncetat -> neîncetat*, *re-ncălzi -> reîncălzi* etc.).

5. Concluzii

Rezultatele prezentate în acest livrabil corespund activității A1.15 din planul de implementare și se referă a identificarea pattern-urilor prozodice din semnalul vocal și corelația acestora cu textul.

Cercetările demonstrează faptul că pattern-urile prozodice manifestate la nivelul semnalului vocal au legătură directă și prezintă strînse corelații pe termen scurt sau pe termen lung cu atribute de morfologie și sintaxă aferente textului. Principalele atribute se refera la poziționare accent în cuvinte, silabificare, părțile de vorbire, sintaxa, respectiv punctuație. Aceste rezultate prezintă fundamentul pentru dezvoltarea unor noi metode de sinteză expresiva a vorbirii prin intermediul unor module de analiza a expresivității textului (în componenta software de procesare de text), respectiv de modificare automată a prozodiei (în componenta software de sinteză de semnal).

6. Bibliografie

[Fer14] Raul Fernandez, Asaf Rendel, Bhuvana Ramabhadran, Ron Hoory, “Prosody Contour Prediction with Long Short-Term Memory, Bi-Directional, Deep Recurrent Neural Networks”, Interspeech 2014

[Giu07] Giurgiu M, Peev L, “Sinteza din text a semnalului vocal. Vol I”, Ed Risoprint 2007.

[Han15] Han Yang, et al, “Integrating Prosodic Information into Recurrent Neural Network Language Model For Speech Recognition”, Proceedings of APSIPA Annual Summit and Conference 2015.

[Ngu15] Hy Quy Nguyen, Siu Wa Lee, Xiaohai Tian, Minghui Don, „High quality voice conversion using prosodic and high-resolution spectral features”, Multimedia Tools and Applications: 7 June 2015.

[Rab93] Rabiner, Juang, Fundamentals of Speech Recognition, Prentice Hall, ISBN 0-13-015157-2, 1993.

[Rud94] Rudnick,A., Hauptmann,A., Lee,K., "Survey of Current Speech Technology", Communications of the ACM, Vol.37, No.3, March 1994.

[Sak03] Sakai S. and J. Glass, “Fundamental frequency modeling for corpus-based speech synthesis based on a statistical learning technique,” in Proc. ASRU 2003, 2003, pp. 712–717.

[Sta96] Stan, Ioan Tedor, „Fonetică”, Ed. Presa Universităţii Clujene, Cluj-Napoca 1996.

[YiJ02] J. Yi and J. Glass, “Information-theoretic criteria for unit selection synthesis,” in Proc. ICSLP 2002, Denver, 2002, pp. 2617–2620.

[Tay91] Taylor P.A., I.A. Nairn, A.M. Sutherland & M.A. Jack (1991) "A real time speech synthesis system", EUROSPEECH, 1991, pp. 341-344.

[Tay91] Taylor P.A. and S.D. Isard, „Automatic phone segmetation” in Proceedings of Eurospeech, September, 1991, pp. 709.711, Genova, Italy.

[Tok00] Tokuda, Yoshimura, Masuko, Kobayashi, Kitamura, Speech Parameter Generation Algorithms for HMM-based speech synthesis, 2000.