



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

Aceste rezultate au fost obținute prin finanțare în 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”

© 2018-2020 – SINTERO

Acest document este proprietatea organizațiilor participante în proiect și nu poate fi reprodus, distribuit sau diseminat către terți, fără acordul prealabil al autorilor.

Denumirea organizației participante în 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 Politehnică din București	UPB	UNI	P2
Universitatea "Alexandru Ioan Cuza" din Iași	UAIC	UNI	P3

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, în 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 lingvistică prin care se manifestă modificările prozodice în 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 formanțelor 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 frecvențele fundamentale suferă variații atunci când diftongii (respectiv vocalele) sunt încadrați în cuvinte; F0 maxim scade atunci când avem grupuri de vocale încadrate împreună în cuvânt, iar energia acestor diftongi încadrați în cuvinte este sensibil mai mică decât cea a diftongilor, triftongilor izolați. Similar s-au obținut rezultate pentru diferite categorii de unități acustice. Un alt exemplu este pentru accent. Una din concluziile importante ale studiului se referă la o creștere a frecvenței fundamentale pentru silabele (sau vocalele) accentuate, față de cele neaccentuate în medie cu 5%..20% (în 90% din cazuri creșterea s-a plasat în intervalul 9%..12%). Merită făcută și observația ca au existat și câteva cazuri în care accentuarea unei silabe nu a adus nici un fel de diferențiere 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 afectează prozodia, în special la nivel de intonație de propoziție. Sunt identificate un set de 7 pattern-uri intonaționale la nivel de propoziție, dar și efectul prozodic al semnelor de punctuație.

Cercetările demonstrează faptul că pattern-urile prozodice manifestate la nivelul semnalului vocal au legătură directă și prezintă strânse corelații pe termen scurt sau pe termen lung cu attribute de morfologie și sintaxă aferente textului. Principalele attribute se referă la poziționare accent în cuvinte, silabificare, părți de vorbire, sintaxă, respectiv punctuație. Aceste rezultate prezintă fundamentul pentru dezvoltarea unor noi metode de sinteză expresivă a vorbirii prin intermediul unor module de analiză 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	4
2. Identificarea pattern-urilor prozodice, o problemă deschisă	4
2.1. Accentul	5
2.2. Intonația	5
2.3. Ritmul.....	6
2.4. Alte aspecte ale prozodiei	7
3. Manifestarea prozodiei în parametrii semnalului vocal	8
3.1. Frecvența fundamentală a unităților sonore din semnalul vocal	8
3.2. Variația frecvenței fundamentale în funcție de accent	10
3.3. Frecvența fundamentală în funcție de intonația propoziției	13
3.4. Analiza formanților în funcție de vorbitori pe tot corpusul	18
3.5. Ritmul vorbirii și durata unităților acustice	20
4. Manifestarea prozodiei în caracteristici de natură lingvistică	22
4.1. Aspecte de natură lingvistică	22
4.2. Pattern-uri intonaționale la nivel de propoziție	22
4.3. Rolul accentului în prozodie	24
4.4. Rolul semnelor de punctuație în prozodie	25
5. Concluzii	26
6. Bibliografie	26

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, în 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 strânse corelații pe termen scurt sau lung, în interacțiune, cu atribute de morfologie și sintaxă aferente textului. Principalele atribute se referă la poziționare accent în cuvinte, silabificare, părți de vorbire, sintaxă, respectiv punctuație.

Pentru a pune în evidență 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 caracteristicile 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 naturaleței semnalului sintetizat de un sistem de Sinteza din Text a Semnalului Vocal (STSV) depinde de bogăția de contururi intonative și 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 și duratei fonemului);
- intonația (variația în timp a frecvenței fundamentale – F0);
- ritmul (durata fonemelor și viteza cu care sunt acestea sintetizate),

Până nu demult, majoritatea cercetărilor tratau separat aceste trei elemente. În realitate ele sunt în strânsă interdependență. De exemplu, în limba română accentul pe cuvânt este liber, variind între ultimele două silabe ale cuvântului. Exista multe excepții de la această regulă. Cuvinte cu aceeași ortografie au semantică diferită în funcție 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 amănunțit asupra conturului frecvenței fundamentale pentru diferite tipuri de propoziții (declarative, întrebări, exclamații, etc).

Pentru propoziții declarative frecvența fundamentală crește pe primul cuvânt (de la 100% la 140% din valoarea sa, apoi coboară la 125% pe ultima parte a cuvântului) și descrește până la sfârșitul propoziției, cu excepția ultimului cuvânt. Aici scade la 70% și rămâne constantă. Propozițiile interogative pot avea un cuvânt specific de interogare (cine, unde, când) sau pot să nu aibă. În primul caz frecvența fundamentală crește pe acel cuvânt de la 100% la 160% și revine la 100%. Pentru al doilea tip de întrebări s-a adoptat un contur convențional, dar efectele subtile de intonație nu pot fi rezolvate. În esență, pentru a reprezenta conturul intonațional, frecvența fundamentală (pitch) este supusă unei operații de “stilizare”, cu scopul de a aproxima F0 printr-o secvență de segmente de dreaptă, rezultatul fiind o reprezentare foarte apropiată de variațiile din intonația vorbitorului. De exemplu, Figura 2.1. arată câțiva dintre parametrii unei propoziții interogative în care conturul intonațional scade lin în primul cuvânt, urmând apoi o porțiune constantă în al doilea, urmată de o creștere lină finală. Realizarea acustică a semnelor ca virgula, două puncte și punct și virgulă contribuie la îmbunătățirea naturaleței vorbirii.

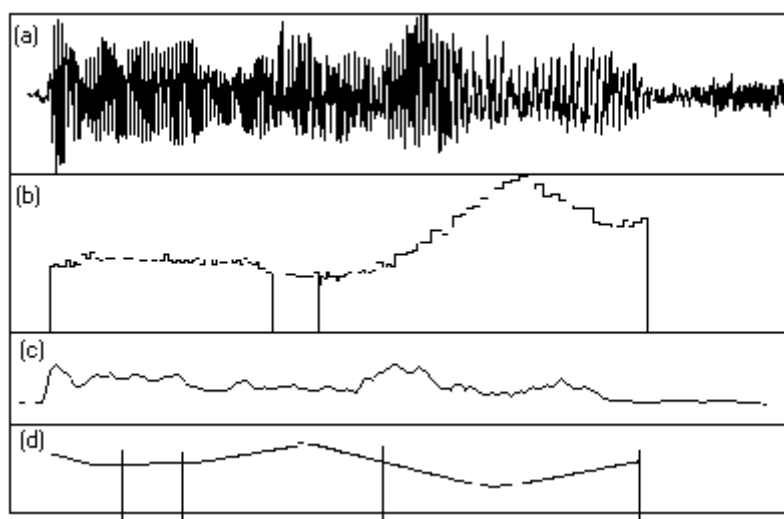


Figura 2.1. Variația intonației (F_0).

(a) - semnalul original, (b) – variația F_0 , (c) - amplitudinea, (d) - melodia prozodică

2.1. Accentul

O anumită silabă poate fi rostită cu mai mare sau mai mică intensitate la fel ca în exemplul “casa” sau “casa”. Acest fenomen este descris prin accent. În funcție de domeniile sale diferite de acțiune, se pot distinge trei tipuri de accent :

accentul la nivel de silabă sau de cuvânt

accentul pe propoziție

accentul pe frază

O altă caracteristică acustică importantă este durata relativă a silabei. Accentul silabei precum și silaba de la sfârșitul propoziției sunt în principiu lungi. Durata pauzei și a silabei se obțin din segmentarea fonetică. Realizarea acustică a accentului va afecta întotdeauna cel puțin doi dintre următorii parametri prozodici: frecvența fundamentală, intensitatea sau durata, iar în unele cazuri pe toți trei. Astfel, se constată o creștere a frecvenței fundamentale și a amplitudinii pentru silabele accentuate comparativ cu echivalentele lor neaccentuate.

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

Accentul în propoziție / frază are rolul de a reliefa aspecte semantice relevante. Accentul mai puternic aparține 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. Intonația

Al doilea fenomen prozodic ca importanță îl reprezintă percepția melodică a vorbirii. Aceasta este denumită intonație. Acest aspect e ușor de observat în variațiile ascendente ale curbei F_0 din propozițiile interogative, comparativ cu variația descendentă din propozițiile

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

În manifestare acustică, intonația este strâns legată de frecvența fundamentală și chiar de accent, într-o manieră greu de descifrat. Ascendența și descendența din rostire sunt urmărite cu fidelitate de inflexiunile frecvenței fundamentale.

Adeseori în sinteza de voce se pune problema interacțiunii frecvenței fundamentale relativ la accent și intonație. O ipoteză simplistă ar putea sugera că cele două efecte sunt cumulative. Măsurătorile au arătat însă, că situația reală este mult mai complexă: acestea au contrazis ipoteza, frecvența fundamentală prezentând o valoare medie mai redusă pentru varianta accentuată și chiar o diferență mai mare, în același sens (de exemplu scădere) pentru varianta interogativă comparativ cu cea afirmativă.

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

Producerea intonației este un proces descris, în principiu, din doi pași:

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

Intonația a devenit stereotipă mai ales în cazul anumitor tipuri de interogative (o categorie redusă de propoziții și 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 inițial este accentuat pe prima silabă, tonul se menține egal pe următoarele silabe accentuate, pentru a urca în continuare.
- când silaba inițială poartă accentul propoziției / frazei, primul ton este înalt, iar următoarele sunt coborâtoare.
- intonația descendentă este stereotipă la sintagmele negative (enunțative, exclamative sau interogative) care încep cu un pronume interogativ, adverb relativ, interjecție: *Ce faci astăzi? Unde te duci la vară? Vai, ce bine-mi pare!*
- vorbirea în limba română are două tipuri principale de intonație inițială: a) ascendentă, care începe cu ton relativ coborât (frecvență); b) descendentă – pornește de la un ton relativ înalt.
- modificări ale intonației pot să apară numai din motive afective:

Nu vrei să știi!

Ce albastru-i cerul!

Ce prostie!

2.3. Ritmul

Un al treilea tip de pattern prozodic se concretizează în variațiile de viteză ale vorbirii, variații care lasă loc la diferite interpretări perceptuale în funcție de mărimea ariei afectate de variație. 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, variațiile au un aspect local, efectul de durată se corelează cel mai probabil cu accentuarea. Deci creșterea duratei unui cuvânt sau a unor silabe ale acestuia, precum și diminuarea puterii de rostire a unui cuvânt dintr-o frază, va fi un indiciu pentru fenomenul prozodic *accent*.

În ceea ce privește sinteza automată a vorbirii, trebuie notat faptul că nici ratele de variație locale, nici cele globale nu vor avea o variație 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 observație 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 și regresie. Construcția acestor arbori este realizată pe baza următoarelor principii:

- la nivelul fonemului:
 - fonemul curent;
 - clasa acestuia;
 - poziția 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:
 - poziția în silabe a ultimului accent.

Sarcina componentei pentru controlul duratei este de a realiza structura temporală care cuprinde accentuările, în corelație cu intonația. Această sarcină este foarte grea din motiv că durata este afectată de mai mulți 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 cunoaște foarte puțin despre procesele responsabile în controlul duratei.

Componenta de durată a unui sistem TTS este convențională. De exemplu, durata segmentală. Se pot alege însă și durate sub-segmentale, cum ar fi o perioadă de semnal, în cazul fonemelor sonore. La nivel sub-segmental se folosește modificarea neliniară în timp a vocalelor și 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 diferențiere a direcției de variație și a conturului frecvenței fundamentale. În aceste cazuri, semnificațiile cuvintelor se stabilesc în funcție de intonație.

Momente de legătură: La tranzițiile între cuvinte există adeseori reguli precise de poziționare a accentului și a pauzelor. S-a constatat că pauzele dintre cuvintele unei propoziții sau fraze au tendința de a se lungi spre sfârșitul 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 combinație între acest algoritm și un model lingvistic prin stabilirea unui prag pentru probabilitatea de apariție a sfârșitului de propoziție și 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 consecință, nu există un model universal valabil pentru sinteza din text a vorbirii pentru diferite limbi. Sinteza trebuie adaptată și implementată în conformitate cu sistemul fonetic și cu regulile lingvistice ale limbii în cauză.

3. Manifestarea prozodiei în parametrii semnalului vocal

Această secțiune prezintă rezultatele experimentale privind variația parametrilor prozodici frecvența fundamentală pentru vocale, frecvența fundamentală în funcție de accent, frecvența fundamentală în funcție de intonația din propoziție, variația frecvenței formantelor 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 dificultățile în încercarea de a stabili modele de variație pentru parametrul frecvență fundamentală a semnalului vocal funcție de parametrii prozodici.

3.1. Frecvența fundamentală a unităților sonore din semnalul vocal

Prin acest experiment s-a urmărit obținerea de informații referitoare la frecvențele fundamentale ale vocalelor și grupurilor primare de vocale diftongi și triftongi din limba română (Tabel 3.1 / 3.2). Pentru aceasta se vor efectua o serie de măsurători pe înregistrările făcute pentru parametri: pitch (frecvența fundamentală F_0) minim, maxim, mediu și intensitate. Se urmărește apoi determinarea unui model de variație a acestor mărimi atunci când unitățile fonetice în discuție nu mai sunt considerate izolate, ci sunt încadrate în cuvinte, respectiv fraze.

Tabelul 3.1. Variația 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 rostiți ca atare, sau izolați în paralel cu măsurătorile efectuate pe cuvinte care conțin acești diftongi.

Tabelul 3.2. Variația 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

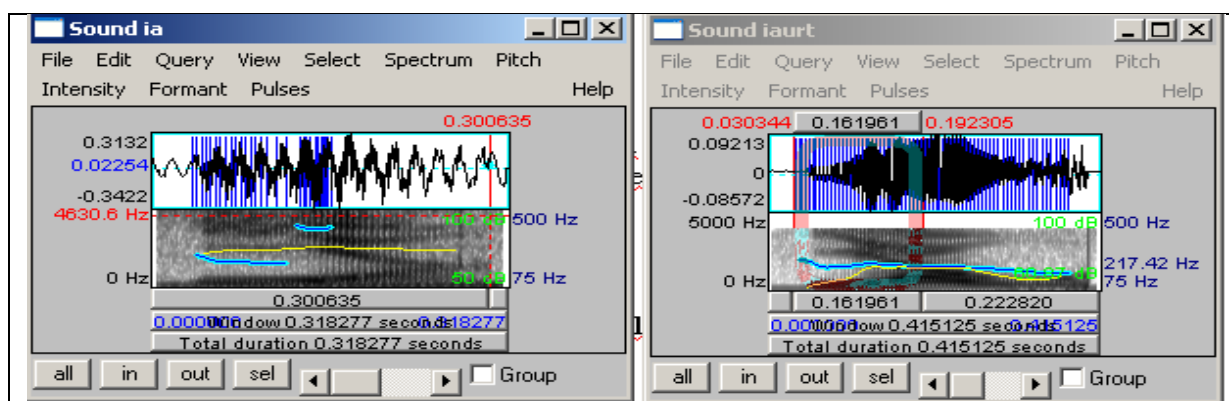


Figura 3.1. Diftongul /au/ izolat, respectiv în context

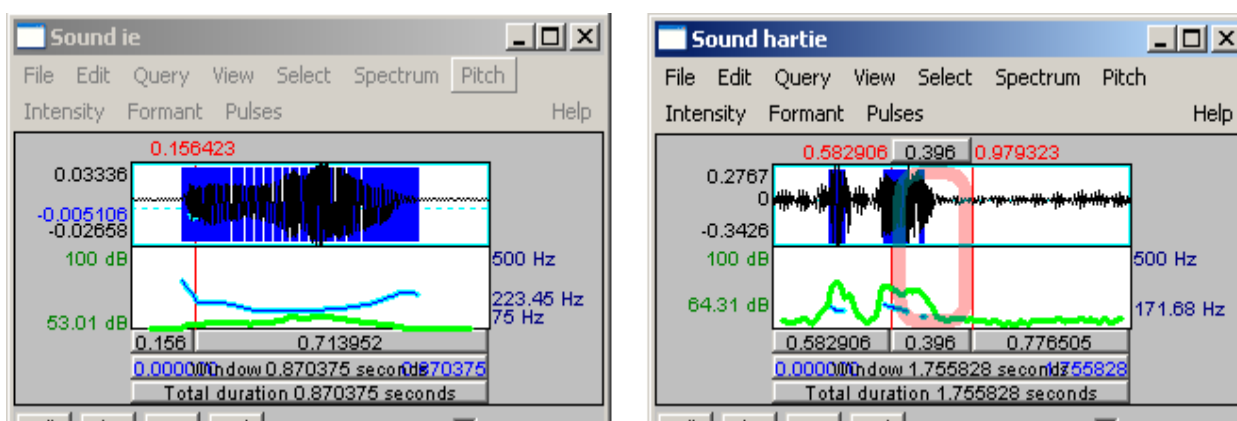


Figura 3.2. Variația parametrilor prozodici pentru diftongul /ie/

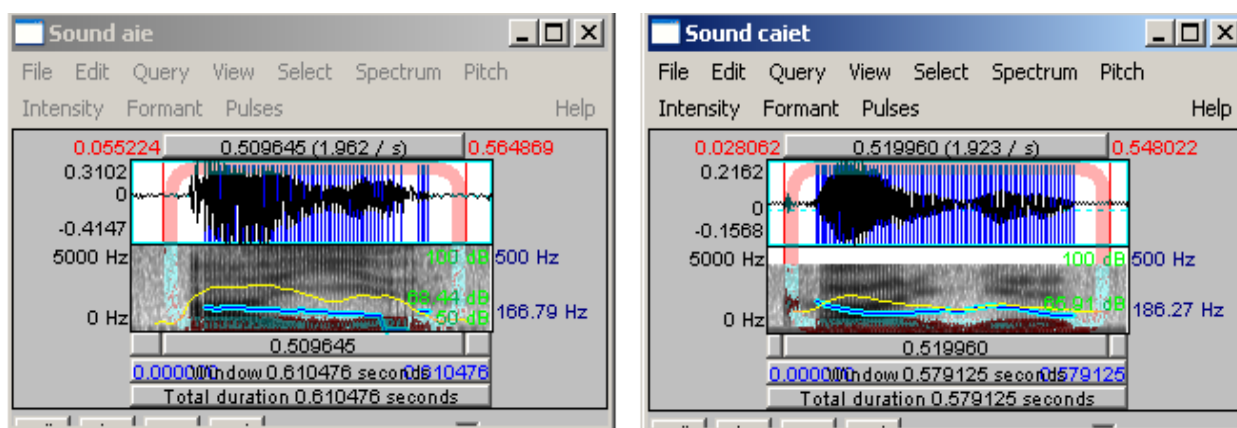


Figura 3.3. Variația parametrilor prozodici pentru triftongul /aie/

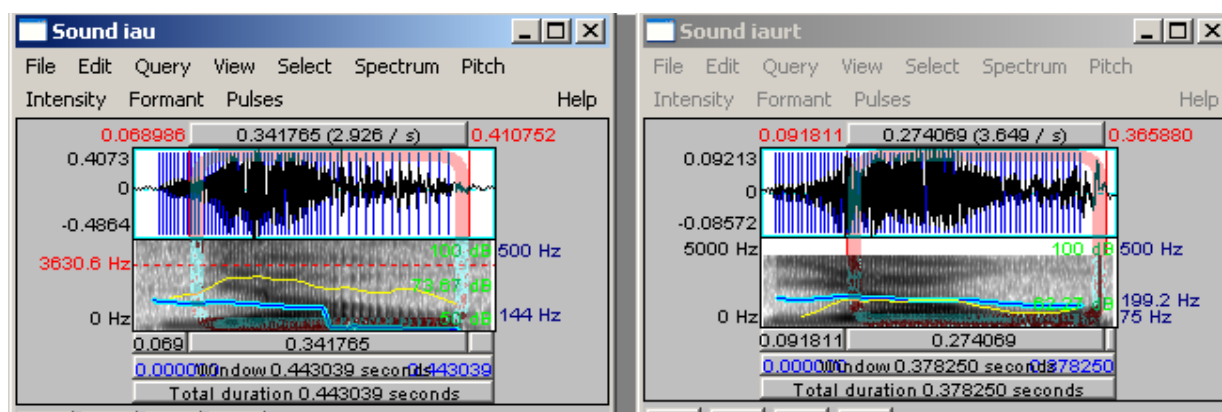


Figura 3.4. Variația parametrilor prozodici pentru triftongul /iau/

Concluzii:

a) *Diftongi*: Așa după cum arată măsurătorile efectuate, frecvențele fundamentale suferă variații atunci când diftongii (respectiv vocalele) sunt încadrați în cuvinte; F0 maxim scade atunci când avem grupuri de vocale încadrate împreună în cuvânt, iar energia acestor diftongi încadrați în cuvinte este sensibil mai mică decât cea a diftongilor, triftongilor izolați. Contează aici, desigur, și poziția diftongului în cadrul cuvântului, numărul de silabe ale acestuia precum și aportul accentului. E un fapt dovedit, și se poate ușor observa din figurile anterior amintite, că energia cuvintelor din finalul propozițiilor este mai mică decât cele situate undeva la început sau pe poziții centrale.

b) *Triftongi*: Să observăm că măsurătorile nu respectă modelul de la diftongi: se înregistrează frecvențe sensibil mai mari pentru triftongii încadrați în cuvinte decât cei izolați. În ceea ce privește intensitatea, aceasta păstrează tendința de a se diminua odată cu încadrarea în cuvinte. Ca o mențiune ce vine în suportul modelului de variație inițial propus, să menționăm că situațiile în care grupurile de vocale diftongi și triftongi apar ca atare sunt foarte rare, așadar relevanța acestor excepții va fi minimă.

Odată cu încadrarea vocalelor, diftongilor, triftongilor etc. în cuvinte și fraze, în general va scădea frecvența fundamentală și intensitatea va crește. Așadar, putem afirma că domeniul de variație a frecvenței 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 dorește măsurarea și determinarea frecvenței fundamentale și a intensității atunci când unitățile lingvistice sunt *afectate de accent*. În acest scop s-au efectuat înregistrări pentru o serie de cuvinte, pentru varianta accentuată și neaccentuată a acestora și s-au înregistrat de asemenea propoziții cu aceste cuvinte. În continuare sunt prezentate în paralel rezultatele măsurătorilor pentru cele patru grupe de cuvinte propuse:

- factura-factură;
- haină-haină`;
- veselă-vèselă;
- lumina-lumină

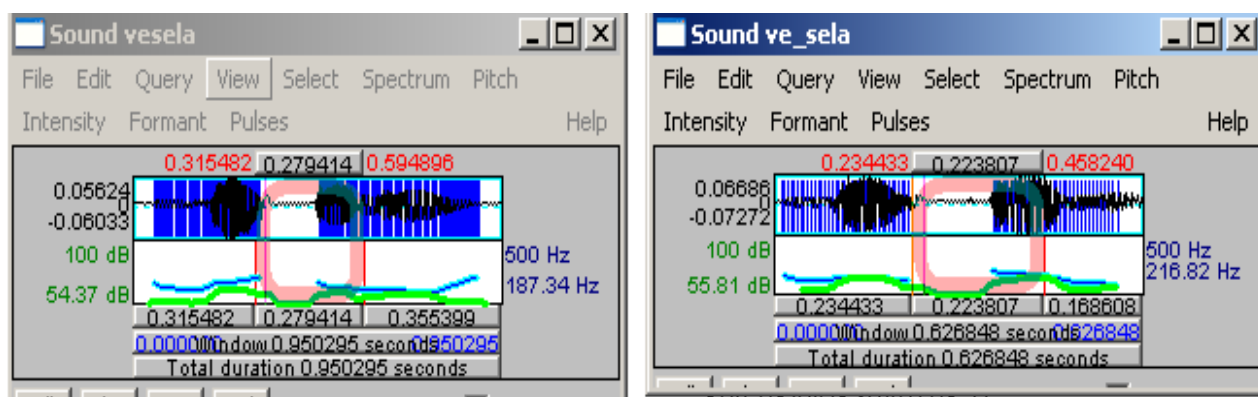


Figura 3.5. Comparație grafică între variația parametrilor prozodici /vesela/

Notă: În Figura 3.5. s-a identificat și selectat silaba afectată de accent pentru cuvântul accentuat și pentru cel neaccentuat. Se observă cum crește și aici F0 (valoarea afișată în partea dreapta a capturilor reprezintă pitch-ul mediu al undei sonore pe secțiunea selectată), iar intensitatea are o tendință ascendentă (pentru silaba afectată de accent).

Tabelul 3.3. Tabel sintetic privind variația parametrilor prozodici în funcție de accent

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

Conform studiilor realizate până acum s-a arătat că în general silabele accentuate au tendința de a avea frecvența fundamentală, durata și amplitudine mai ridicată decât silabele neaccentuate. Însă, există și cazuri în care numai unul sau doi din acești parametri este mai ridicat, precum și situații în care tendința silabelor accentuate este de a-și reduce fundamentală sau ceilalți parametri.

Pentru fiecare din înregistrările făcute s-a realizat analiza și prelucrare separată. Modul de lucru a presupus mai mulți pași. S-a luat o înregistrare a unui cuvânt și s-a determinat pentru aceasta forma de undă, frecvența fundamentală și anvelopa. Mai apoi s-a luat fiecare cuvânt în parte și din acesta au fost extrase silabele accentuate, respectiv neaccentuate. Pentru ca rezultatele să 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 silabă (când se referă la un element chimic - gaz), respectiv pe cea de a doua silabă (când este pluralul cuvântului gazela - animal). Pentru început vom prezenta formele de undă pentru cele două cazuri amintite (Figura 3.5 / 3.6).

Pe Figura 3.5. se pot distinge patru grafice. Primul, respectiv al treilea sunt formele de undă ale celor două cazuri luate în discuție, iar al doilea, respectiv al patrulea reprezintă anvelopele de energie corespunzătoare undelor. În prima imagine accentul cade pe prima silabă ga. În imaginea a treia se poate distinge aceeași silabă ga, dar neaccentuată. Chiar și la o analiză superficială se pot distinge anumite caracteristici pentru cele două cazuri distincte. Amplitudinea și durata în primul caz sunt sensibil mai ridicate decât în cazul cu aceeași silabă neaccentuată. Dacă discutăm aceeași situație în cazul silabei ze vom observa că situațiile se

inversează pentru ca de aceasta data în prima imagine silaba nu e accentuata, respectiv în imaginea a treia accentul cade pe aceasta silaba. Si de aceasta data se pot observa relativ ușor o amplitudine mai ridicată și o durată mai mare în cazul silabei accentuate.

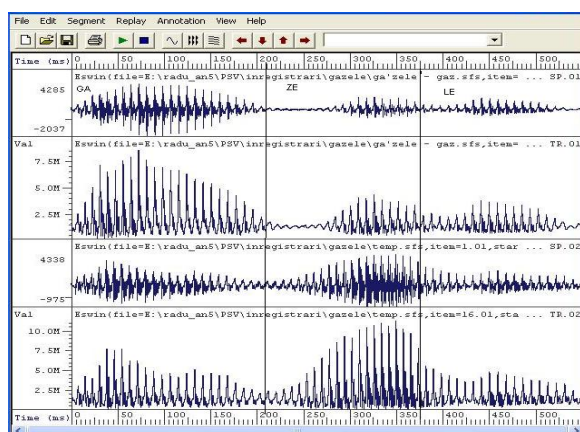


Figura 3.5. Formele de undă și anvelopa pentru cele două cuvinte.

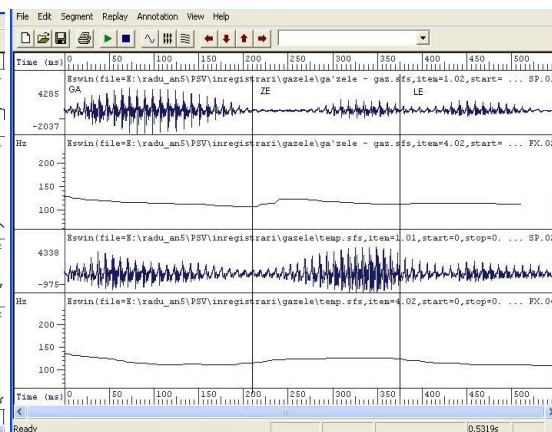


Figura 3.6. Formele de undă și variația frecvenței fundamentale

Pentru a putea vedea și evoluția frecvenței fundamentale se va mai adăuga o imagine cu cele două forme de undă pentru cele două cazuri distincte (imagine dreapta). Se poate observa cum frecvența fundamentală este mai ridicată pe silaba accentuată. O analiză mai detaliată a adus mai multe informații care vor fi amintite ceva mai târziu. Dacă ne concentrăm și asupra ultimei silabe, rezultatele sunt conform așteptărilor, în sensul în care între cele două cazuri există foarte mici deosebiri legate de durată, amplitudine sau frecvența fundamentală.

Pentru determinarea mai exactă a parametrilor importanți (frecvența fundamentală, durată, amplitudine) fiecare cuvânt a fost segmentat în silabe. Din fiecare cuvânt au fost reținute silabele (sau vocalele din cadrul silabelor) care aveau relevanța în analiză, adică acele silabe care într-un caz sunt cu accent, respectiv fără accent (vezi Figura 3.7).

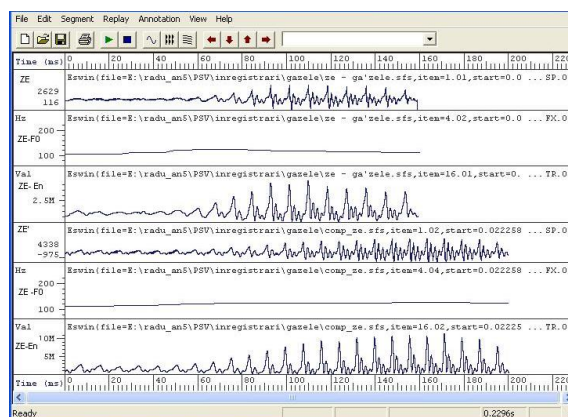
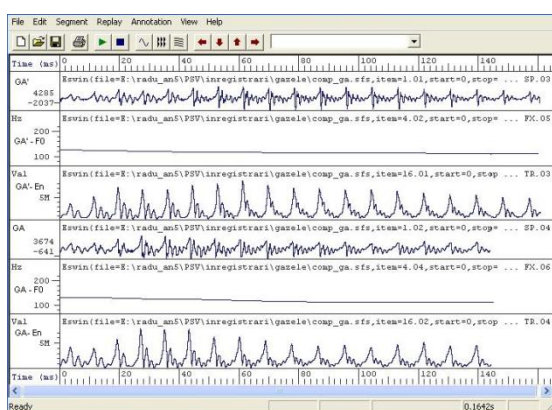


Figura 3.7. Analiza prozodică a silabelor /ga/ (stânga), respectiv /ze/ (dreapta), cu accent (sus), respectiv fără accent (jos)

Si în acest caz putem distinge o durată mai ridicată a silabei accentuate ('ze'). În acest caz la un studiu mai detaliat s-a determinat și o creștere a frecvenței purtătoare și a amplitudinii. Dacă privim din nou formele de undă prezentate, vom putea observa încă un lucru care merita a fi subliniat, și anume faptul că în cadrul aceluiași cuvânt putem distinge din forma de undă

care silaba este accentuata, datorită amplitudinii mai ridicate a acesteia față de silabele neaccentuate.

În experimentele anterioare s-a urmărit determinarea modelului de variație a frecvenței fundamentale în funcție de accent. Astfel s-a ajuns la un acord comun în ceea ce privește câteva reguli privind distribuția și sistematică a accentului, reguli confirmate de rezultatele experimentelor de mai sus :

- deși domeniul de influență al accentului poate să acopere cuvinte în întregime, fraze sau propoziții, va exista întotdeauna o singură silabă care va susține accentul.
- în procedura de identificare a silabei afectate de accent dintr-un cuvânt, frază sau propoziție, accentul principal se va determina detașat de celelalte silabe cu accent.
- accentul unei fraze sau al unei propoziții va coincide în general cu cel al unui cuvânt; diferitele niveluri de accent se vor exclude și susține 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 variația frecvenței fundamentale în cazul unei vocale sau difonem. Astfel în cadrul unor structuri analizate (vocala, silaba, cuvânt) am întâlnit situații când frecvența fundamentală avea valori cuprinse între 128 și 156 Hz și 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.

În cazul vocalelor accentuate (a – în exemplul următor) se constata o creștere pentru F_0 cum ar fi în cazul cuvântului “acele”. Dacă este rostit cu sensul de “acele de cusut” atunci $F_0=137$ Hz, iar în cazul rostirii ca și pronume (“acele persoane”) F_0 va scădea la 125 Hz. Am putea continua enumerarea de astfel de situații, creșterea medie a frecvenței fundamentale fiind de 5- 20% în funcție de vorbitor, cuvânt etc.

Un alt aspect observat a fost acela ca prezenta pronunțată a accentului poate modifica evoluția conturului frecvenței fundamentale a întregului cuvânt. Cazurile elocvente au fost acelea când vocala sau silaba accentuata era situata la mijlocul sau în a doua parte a cuvântului. În acest caz s-a observat o evoluție crescătoare a valorilor frecvenței fundamentale cu puțin înainte de începutul zonei accentuate și sfârșind odată cu începerea unui nou cuvânt sau silaba. O tendință asemănătoare, dar în sens descrescător, se constata în cazul prezentei accentului la începutul cuvântului sau în prima jumătate a acestuia.

Una din concluziile importante ale acestui studiu se referă la o creștere a frecvenței fundamentale pentru silabele (sau vocalele) accentuate, față de cele neaccentuate în medie cu 5%..20% (în 90% din cazuri creșterea s-a plasat în intervalul 9%..12%). Merita făcută și observația ca au existat și câteva cazuri în care accentuarea unei silabe nu a adus nici un fel de diferențiere din punctul de vedere al valorii F_0 .

Un aspect frecvent în studierea formelor de unda a fost creșterea amplitudinii semnalului pe silaba accentuata în cadrul cuvântului. În acest fel, de cele mai multe ori *silaba accentuata se distinge într-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 variație a parametrilor semnalului vocal în funcție de intonație. Pentru a păstra o continuitate față de celelalte experimente, s-au efectuat înregistrări pentru propoziții construite în jurul celor 4 grupe de cuvinte folosite la experimentele anterioare, de data aceasta fiind afectate de intonație. Se studiază astfel, în paralel, influența accentului și a intonației, prin comparații ale mărimilor obținute pentru cuvintele neaccentuate, accentuate respectiv accentuate și intonate. Acest aspect are o importanță majoră dacă ne gândim la implicațiile sale în procesul de sinteză de voce, unde pentru a forma spre exemplu o frază, nu va fi suficient să asamblăm cuvinte izolate.

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

Tabelul 3.4. Variația frecvenței fundamentale în funcție de contextul cuvântului

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

În urma măsurărilor anterioare, putem observa diferențe minore în ceea ce privește valorile maxime ale lui F0 în cadrul cuvintelor rostite separat comparativ cu cele selectate din cadrul unor fraze. În acest context e important să observăm că gama de variație a parametrului F0 pentru cuvintele izolate va fi sensibil mai mare decât pentru cele extrase din context.

În manifestarea sa acustică intonația este în strânsă legătură cu frecvența fundamentală F0. Astfel, intonația ascendentă caracteristică secvenței finale a interogațiilor, la fel ca cea descendentă ce caracterizează exclamațiile, sunt foarte bine surprinse de inflexiunile frecvenței fundamentale F0.

Având în vedere rezultatele experimentelor anterioare, ne punem problema determinării modului de interacțiune între variațiile lui F0 datorate accentului și cele datorate intonației. Se dorește stabilirea unei reguli de variație a frecvenței fundamentale în cadrul unui cuvânt afectat, respectiv neafectat de accent și afectat sau neafectat de intonație. 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 experimentare.

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

La ora 23 felinarul nu mai lumina.

La ora 23 felinarul nu mai lumina!

La ora 23 felinarul nu mai lumina?

Vom studia tendințele de evoluție a pitch-ului în cadrul cuvântului *lumina*, pentru fiecare intonație în parte (forma afirmativă coincide cu cea accentuată) și îl vom compara cu cel al cuvântului neaccentuat *lumina*.

Rezultate experimentale:

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

Notă: Se observă că intonația interogativă e caracterizată și de o creștere a frecvenței fundamentale F0.

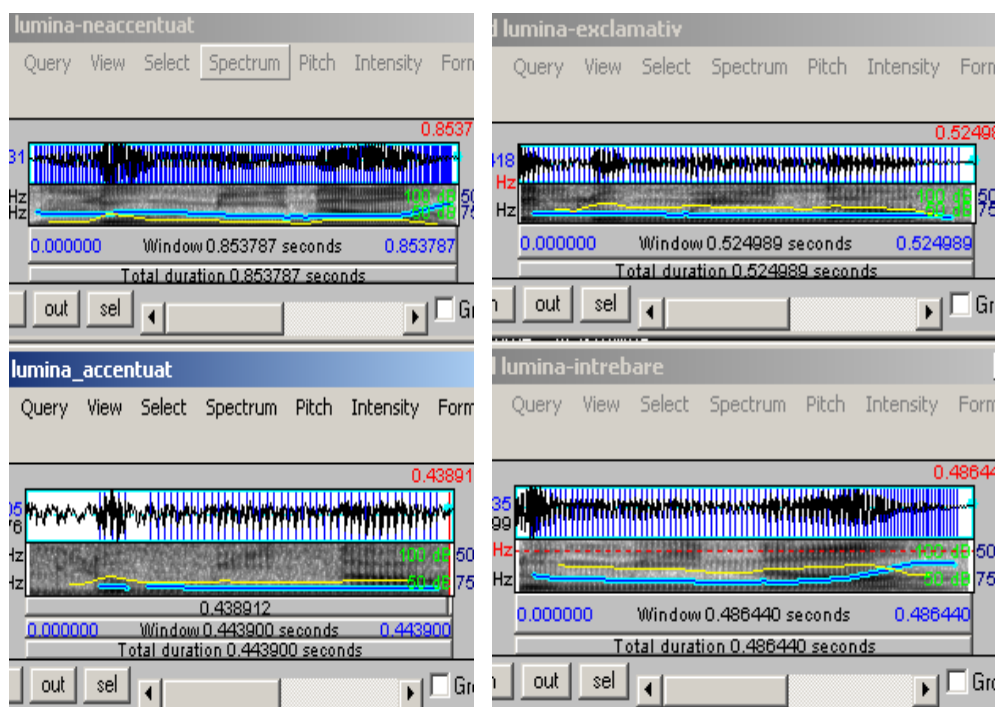


Figura 3.8. Cuvântul /lumina/ în diferite contexte intonaționale

Măsurătorile pentru:

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

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

Nu știa cât e de haină.

Nu știa cât e de haină!

Nu știa cât e de haină?

Rezultatele măsurătorilor:

- cuvânt haină neaccentuat : 179 Hz
- cuvânt haină accentuat: 135 Hz
- cuvânt haină accentuat cu intonație exclamativă: 150 Hz
- cuvânt haină accentuat cu intonație interogativă: 194 Hz

La fel ca în cazul anterior, frecvența fundamentală F0:

- scade pentru varianta accentuată comparativ cu cea neaccentuată
- crește pentru cuvântul accentuat cu intonație interogativă
- crește pentru cuvântul accentuat cu intonație exclamativă (deși nu la fel de mult ca în cazul intonației interogative), comparativ cu varianta afirmativă sau fără intonație, dar scade comparativ cu varianta accentuată.

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

Am uitat de veselă.

Am uitat de veselă?

Am uitat de veselă!

Rezultatele măsurătorilor:

- cuvânt veselă neaccentuat: 189 Hz
- cuvânt veselă accentuat și cu intonație afirmativă: 183 Hz
- cuvânt veselă accentuat și cu intonație interogativă: 224 Hz

- cuvânt veselă accentuat și cu intonație exclamativă: 154 Hz .

Și pentru acest caz rămân valabile observațiile făcute la punctele anterioare, cu singura diferență, că frecvența fundamentală a cuvântului accentuat și cu intonație exclamativă e mai mică decât cea a cuvântului neaccentuat.

Tabelul 3.5. Frecvența fundamentală în funcție de accent și intonație la nivel de cuvânt

Cuvânt analizat	F0 pentru cuvânt neaccentuat [Hz]	F0 pentru cuvânt accentuat [Hz]	F0 pentru cuvânt cu intonație ? [Hz]	F0 pentru cuvânt cu intonație ! [Hz]
<i>lumina</i>	183	136	244	150
<i>haină</i>	179	135	194	150
<i>veselă</i>	189	183	224	154

Un alt tip de analiza care a fost făcută, este cea referitoare la intonația la nivel de propoziție. S-a urmărit modul în care evoluează frecvența fundamentală de-a lungul unei propoziții în doua situații. Prima situație presupunea exprimarea unei propoziții în forma interogativă, iar apoi exprimarea sub forma exclamativă sau imperativă a aceleiași propoziții. S-a observat că în cazul propozițiilor interogative tendința fundamentalei este aceea de creștere spre sfârșitul propoziției, pe când în cazul celălalt tendința este inversă (scădere sfârșitul propoziției).

În primele doua imagini sunt prezentate forma de unda, respectiv, evoluția fundamentalei pentru enunțul “Mergeți la școală?”. În imaginile trei, respectiv patru sunt reprezentate formele de unda pentru același enunț sub forma imperativă “(Mergeți la școală!)”.

Analizând pe rând cele doua situații se poate observa cum propoziția interogativă are o frecvența fundamentală mai ridicată la sfârșit (cu tendința de creștere pentru F0), în timp ce propoziția imperativă are o frecvența fundamentală ridicată la început, iar tendința fundamentalei este de descreștere spre finalul propoziției. În urma analizelor s-a observat că diferența între valorile fundamentalei la sfârșitul frazei este de aproximativ 40..50Hz (cca.25..30%), diferența în care fundamentală este mai mare în cazul propozițiilor interogative.

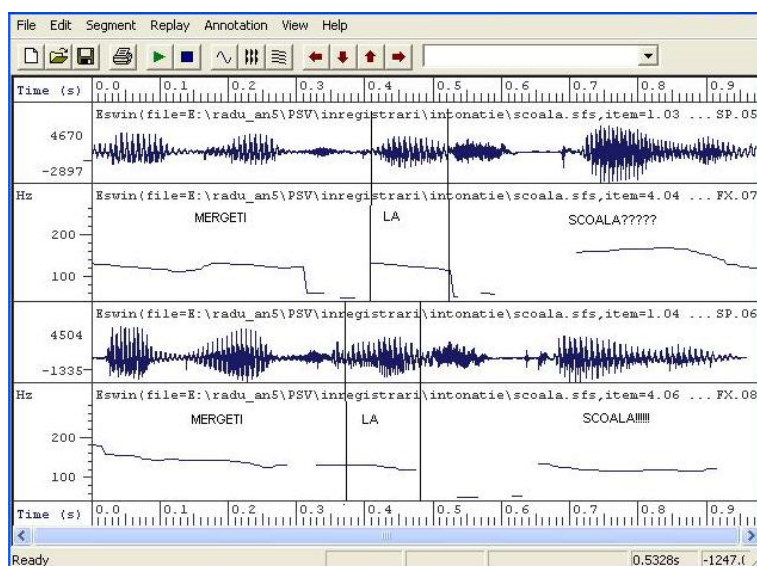


Figura 3.8. Variația frecvenței fundamentale pentru propoziții interogative (sus), respectiv exclamative(jos).

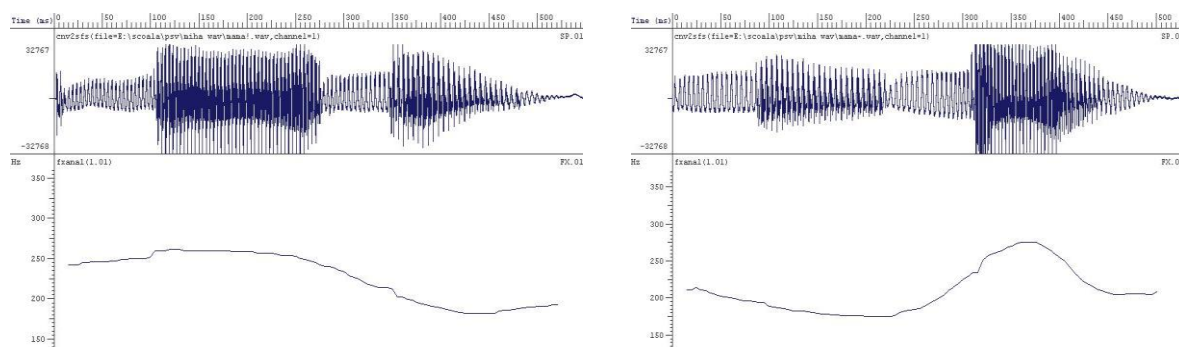


Figura 3.9. Rostire exclamativă (stânga: mama!), respectiv interogativă (dreapta: mama?)

Un alt exemplu de interpretare a măsurătorilor privind variația frecvenței fundamentale este pentru cuvântul “mama”, rostit cu diferite intonații.

- în propoziție declarativa frecvența fundamentală este aproape liniară de la 100% la început de cuvânt, până la 101% și 103% spre sfârșit.
- în propoziție exclamativă, crește la început pe primul “a” până la 106% față de începutul cuvântului și apoi descrește până la 75% la sfârșit.
- într-o propoziție interogativă, dacă la început scade de la 100% spre 85%, în a doua jumătate crește tare până la 130% pe al doilea “a”, iar spre sfârșit scade înapoi spre 100%.

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

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

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

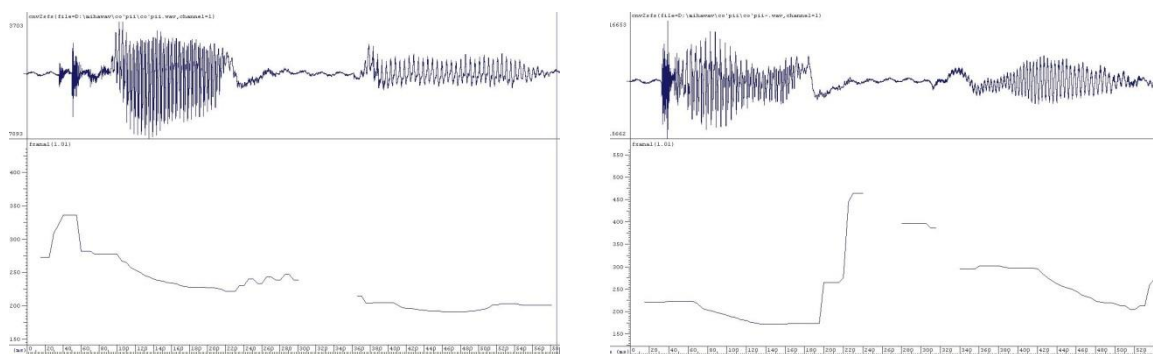


Figura 3.10. Fundamentală pentru ‘co’pii’ (afirmativ), respectiv ‘co’pii?’ (interogativ)

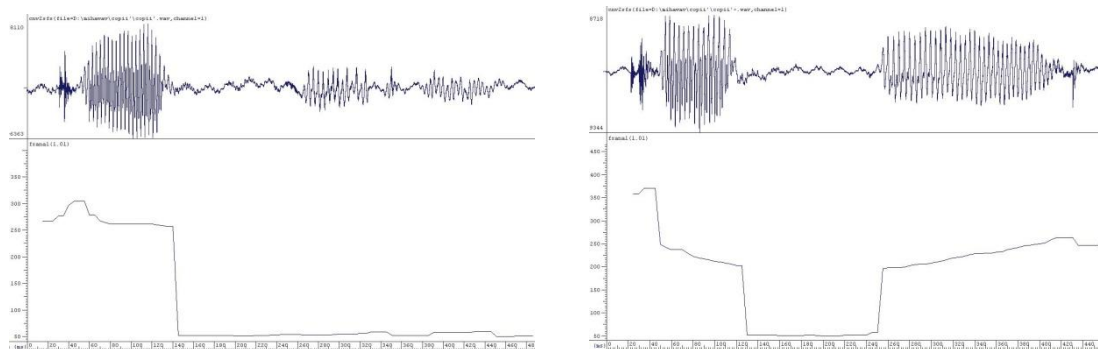


Figura 3.11. Fundamentală pentru 'Copii' (afirmativ), respectiv 'Copii?' (interogativ)

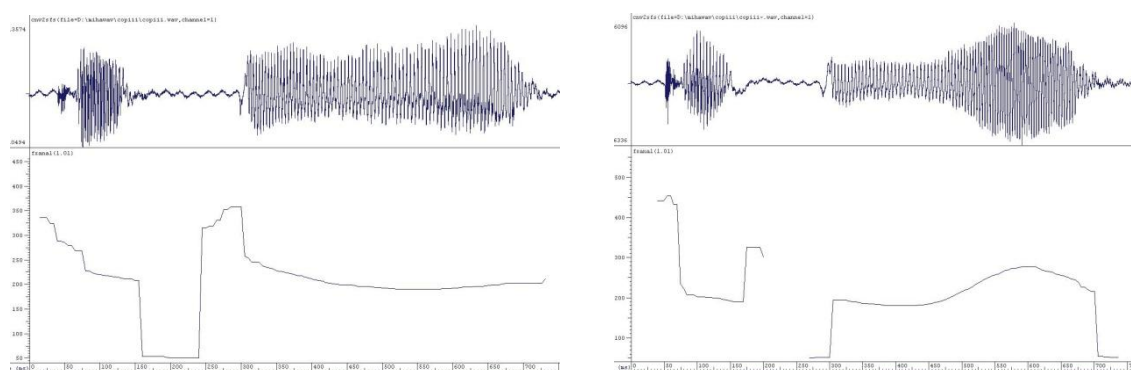


Figura 3.12. Fundamentală pentru 'Copiii' (afirmativ), respectiv 'Copiii?' (interogativ)

În cazul celor trei cuvinte, identice ortografic, se observa următoarele diferențe:

- accentul în "co'pii" este clar pus pe difonemul "co" care are durată, dar și frecvența cea mai mare din cadrul întregului cuvânt, precum și față de celelalte două cuvinte. (față de "copii" frecvența este doar cu 0.02% însă durată este cu 50% mai mare; față de "copiii", care pune clar accentul pe cei trei "i", creșterea frecvenței e de 17% iar a duratei de 20%).
- durată și frecvența celor 2, respectiv 3 de "i" : la "copii" și "co'pii" diferența este la frecvența – crescută cu 26%, în timp nu este diferență mare, dar este o mare deosebire când este rostit cu 3 "i" când durată este mult mai lungă, iar frecvențele sunt relativ înalte.
- tendința este la interogativ de a urca frecvența și durată pe penultima silabă sau difonemă și a cobori puțin chiar la sfârșit, decât dacă avem accentul pus expres pe ultima silabă (copii) și atunci avem o linie ascendentă a frecvenței până la sfârșit.

3.4. Analiza formanților în funcție de vorbitori pe tot corpusul

Interfața de dialog a instrumentului software utilizat pentru analiza formanților este prezentată în Figura 3.13. Structura de organizare a probelor de voce: în directorul 'Date' se găsesc două subdirectoare: 'm' și 'f' (male/female, corespunzător achizițiilor de la vorbitorii masculini, respectiv feminini). În fiecare dintre acestea, se găsesc 7 subdirectoare, câte unul pentru fiecare vocală. Pe interfață trebuie să precizăm calea către date, respectiv dacă e vorba de directorul 'm' sau 'f'. Se alege vocala de prelucrat, și câte probe din această vocală să fie analizate (în cazul de față maximum 21 de probe, deoarece în directorul corespunzător fiecărei vocale sunt 21 fișiere de sunet).

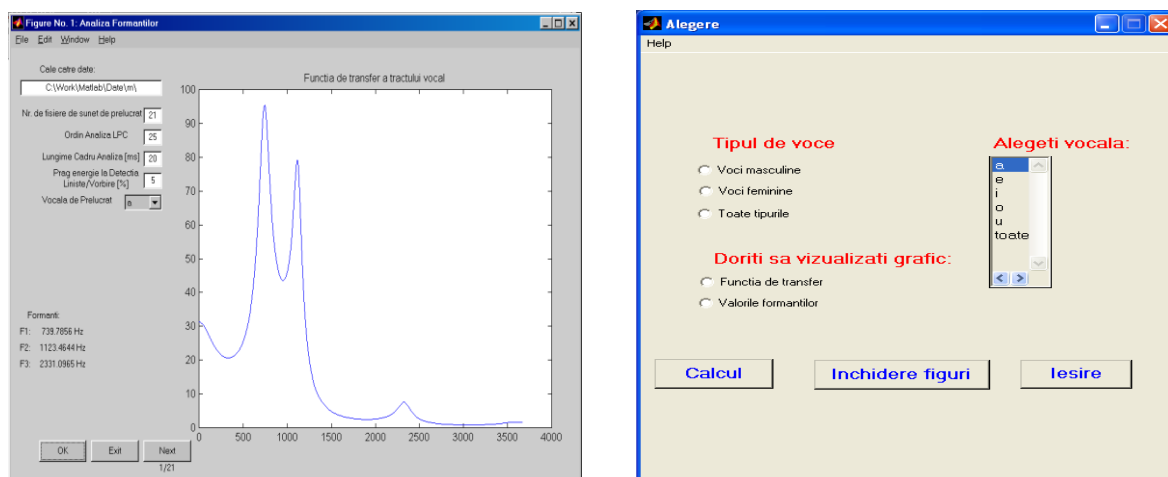


Figura 3.13. Interfețele pentru realizarea studiului privind analiza formantică

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

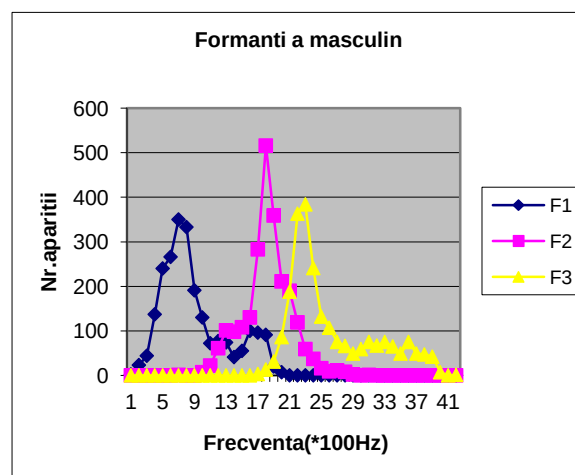
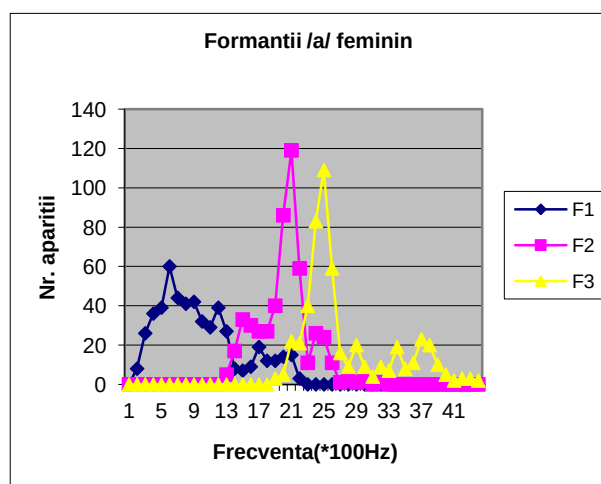


Figura 3.14. Distribuția formanților pentru vocala /a/ feminin (stânga) și masculin (dreapta)

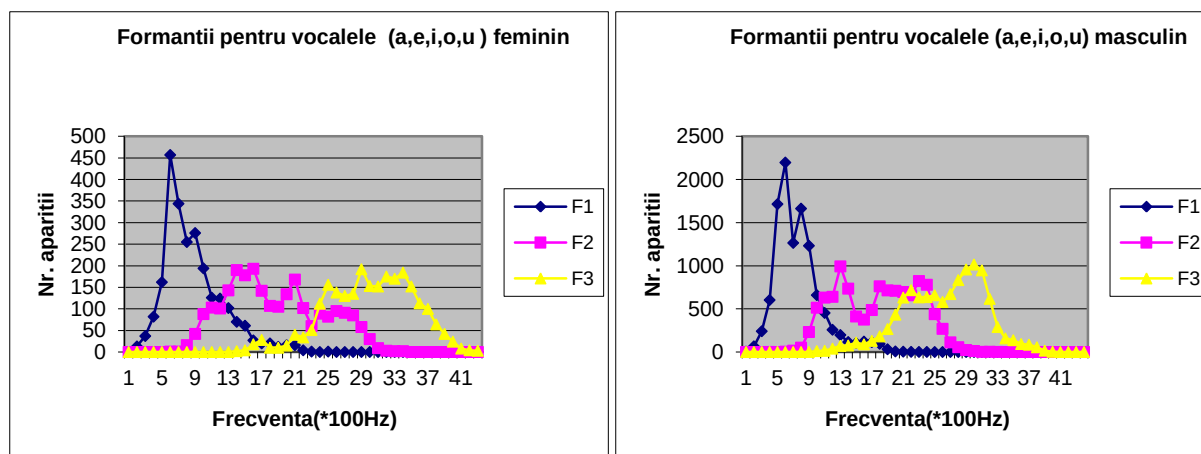


Figura 3.15. Distribuția formanților pentru toate vocalele feminin (stânga) și masculin

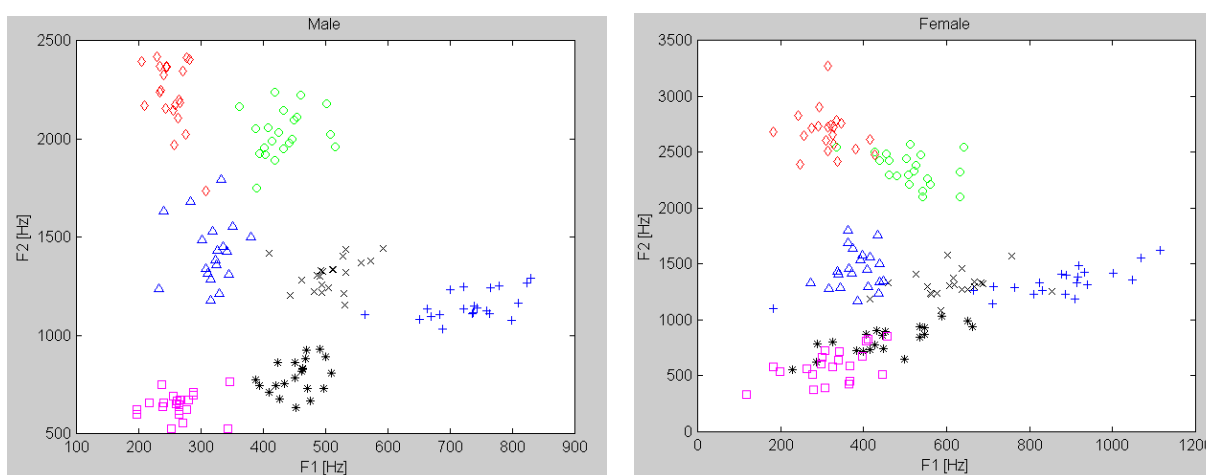


Fig. 3.16. F1-F2 pentru vorbitorii masculini

Fig. 3.17. F1-F2 pentru vorbitorii feminini

Pe lângă diferențele de frecvență fundamentală între o voce masculină și una feminină s-a încercat să se observe care sunt diferențele și la nivel de formanți. Aceste diferențe se pot observa în figurile de mai sus, pentru fiecare vocală în parte, dar și pentru toate vocale (Fig. 3.14 – 3.17) pentru vorbitorii masculini, respectiv feminini. Ceea ce am observat este faptul că distribuția formanților la bărbați este mai compactă decât la femei, ‘norii’ sunt mai concentrați.

3.5. Ritmul vorbirii și durata unităților acustice

Un alt parametru analizat în detaliu a fost durata diferitelor structuri fonetice. Aceasta se modifică în funcție de poziția accentului, în funcție de intonația cuvintelor astfel că de exemplu în cazul unei vocale durata poate avea valori cuprinse între 40 și 140 ms. Din păcate aici intervine și o latură subiectivă în stabilirea acestei durate. Se știe că vocalele în general au forma de undă periodică și frecvența lor fundamentală se poate determina având un contur liniar – orizontal.

Având o astfel de formă de undă este greu de sesizat din punct de vedere auditiv vreo diferență între o durată de 85 ms și una de 120 ms. În schimb consoanele au o durată foarte mică de aprox. 30 ms (depinde de la caz la caz). Faptul că există aceste mari diferențe de durată între vocale și consoane poate crea unele probleme în interpretarea rezultatelor. Un lucru important și care trebuie reținut este acela că în cazul vocalelor accentuate durată este sensibil mai mare decât în cazul celor neaccentuate.

Un alt lucru observat în urma studierii în amănunt a înregistrărilor a fost o creștere a duratei silabelor accentuate, față de cazul în care acestea nu erau cu accent. Creșterea

duratelor se datorează în principal creșterii duratei vocalei accentuate. Aceste creșteri se situează cel mai adesea în intervalul 27%..33% pentru silabele accentuate.

Durata: propozițiile declarative fiind „mai liniare”, timpul de pronunție este mai mic. Explicație posibilă: deschiderea gurii pentru vocale este mai mică într-o propoziție afirmativă față de o propoziție exclamativă sau interogativă – 260ms pentru a rosti simplu „mama” și aproximativ 500ms atunci când exclamam sau întrebăm. Vezi tabelul de mai jos.

Tabelul 3.7 Legătura între durata cuvintelor / difonemelor și intonație (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” – într-o propoziție afirmativă are primul „e” accentuat, fapt ce reiese din durata mai lungă a difonemelor ce conțin primul „e” și a frecvenței, care ajunge până la 110% din frecvența fundamentală de la începutul cuvântului. Într-o propoziție interogativă are același comportament ca și la „mama” – frecvența crește aproape la 157% față de începutul cuvântului la al doilea „e”.

La cuvintele „co’pii” (ca și în „copii xerox”) și „copii” (în sensul de „prunci”, „fii”) cei doi „i” au aceeași durată în difoneme. Totuși, frecvența este crescută cu 126% în „copii” față de „co’pii” tocmai datorită accentului care este pus diferit în cele două cuvinte: pe „i” respectiv pe „o”.

Diferența dintre „copii” și „copiii” se observă prin creșterea duratei vocalei „i” până aproape de două ori atunci când e cu trei de „i”. Care este diferența între accentul pus pentru a deosebi „co’pii” de „copii” sau „copiii” și a desemna o afirmație sau o întrebare?

Un lucru interesant, a fost acela al influenței accentului nu doar asupra vocalelor, ci și asupra consoanelor. Este vorba de difoneme formate dintr-o consoană și o vocală și care sunt accentuate. Durata consoanelor crește într-o măsură mai mică decât cea a vocalelor. Dacă analizăm, în schimb, valorile amplitudinilor maxime, se poate constata că valorile relative și creșterile procentuale sunt de valori apropiate, indiferent de vorbitor.

Studiind problema literei „C” am înregistrat mai multe cuvinte care conțin litera urmată de o vocală mai puțin sonoră („câini”), o vocală puternic sonoră („convex”), „c” urmat de consoană („crâmpie”) și cazurile speciale din grupurile „che”, „chi”, „ce”, „ci”.

Se observă că fonemul „c” împreună cu „o” are o durată relativ lungă și frecvența fundamentală mare datorită sonorității vocalei „o”, deși pentru consoana „c” singură nu se poate stabili o frecvență fundamentală.

Același fenomen se înregistrează și la grupul „că” și la „che”. Insa la grupurile „ce”, „ci” dau o altă sonoritate literei „c” tinzând să fie mai scurte, spre un singur sunet. Același lucru se poate spune și despre litera „g”. „C” urmat de „r” nu este destul de sonor, fiind foarte scurt și fără rezonanță. Această observație este valabilă pentru toate combinațiile de consoane nesonore („c”, „p”, „d”, „t”, etc.).

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

nu
Iarna era
apa caldă.

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 enunțuri declarative, singura diferență față de acestea fiind intonația urcătoare a întrebărilor. Aceste întrebări se numesc *întrebări totale*.

noi?
ne la

Mama vi

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

Cine Când
vine ajungeți
la acasă?
noi?

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

î?
o
Nu vine la n

În limba română sunt frecvente fenomenele de dublă negație, 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 pe al doilea element negativ, tonul cel mai coborât din enunț 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 următoare, apoi se prelungește la același nivel până la urcarea finală, luând naștere un contur melodic în două trepte ascendente:

oi?
meni nu vine la n
Ni

Tiparul ascendent-descendent interogativ. Acest tipar reprezintă intonația cu care sunt rostite, în general, enunțurile care exprimă o alternativă, construite cu conjuncții disjunctive ca *sau* și *ori*. Aceste fraze pot avea aceeași structură segmentală, singurul element distinctiv fiind intonația.

apa

Doriți ori
 suc?

 băm
Ne plim sau cumpărăm ca
 do
 uri?

4.3. Rolul accentului în prozodie

Accentul este unul dintre cei mai importanți parametri prozodici prin care o limbă particulară se poate distinge de alte limbi. Acest aspect se evidențiază prin ceea ce se numește “*accentul străin*”, când o limbă străină este rostită cu intonația limbii materne (de exemplu la utilizarea accentelor gramaticale). După cum se cunoaște, accentul românesc este *expiratoric* sau *dinamic* și are un caracter liber (adică locul său nu este fixat pe o anumită silabă a cuvintelor) și mobil. Diferența de *intensitate* dintre silabele accentuate și neaccentuate nu este prea mare și se completează prin diferență de *durată* și de *înălțime a tonului*. Ultimele totuși sunt aproape irelevante, deci factorul esențial distinctiv rămâne intensitatea.

Accentul la nivel de cuvânt. Libertatea accentuării nu presupune, totuși, 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. copii – copii, mânji – mânji, veselă – veselă etc.) sau chiar formele flexionare (sună - sună etc.). Abaterile de la accentuarea naturală vor trăda imediat vorbitorii străini, care au tendința de a aplica regulile de accentuare specifice limbii lor materne. În mod obișnuit, fiecare cuvânt polisilabic din limba română are o singură culme dinamică. Această regulă poate fi respectată ușor dacă ținem seama că majoritatea cuvintelor românești au un număr redus de silabe. Totuși, la unele împrumuturi recente sau derivate și compuse din elemente străine, apare un *accent secundar* plasat pe prima silabă a cuvântului (de ex. autoapărare, contrasemnătură, reconstituire, suprasolicitat, dar găsim exemple și în fondul autohton cum ar fi untdelemn, bunăvoință etc.). Accentul secundar nu posedă valori distinctive. Fiind însoțit însă și de o ușoară ridicare a tonului, joacă foarte adesea în frază rolul unui accent de insistență. Rar, limba română mai conține și cuvinte cu *două accente principale* (echivalente), în compusele nesudate: după-masă etc.

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

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

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) *Propozițiile enunțiative* (excepție fac exclamativele) nu au în mod obligatoriu un cuvânt mai accentuat față de celelalte.
- 2) *Propozițiile interogative* au de obicei un cuvânt mai accentuat, prin care se arată la ce segment se așteaptă răspuns, iar în enunțiativele ce răspund sau aprobă spusele interlocutorului, accentul cade pe adverbul de întărire. De ex. **Așa** cred / **Sigur** e un băiat.
- 3) *Cuvintele de interogație* se accentuează totdeauna:
 - a) pronume interogative: **Cine** trece pe stradă?, **De ce** nu vii?, **Cui** să-i spun?
 - b) adverbale interogative: **Cum** să-i răspund?, **Unde** să plec?, **Când** te-ntorci?
- 4) *În propozițiile negative*: **Nu** călcați iarba!, El **nu** bea, **nu** fumează, **nu** e risipitor.
- 5) *Conjuncțiile*, **Ori** tăia lemne, **ori** căra apă, mereu lucra. **Acum** râde, **acum** plânge.
- 6) *Prepozițiile* sunt neaccentuate, excepție făcând în anumite situații: Mergea **după** el, **niciodată** înaintea lui. **Pentru** numele Domnului!
- 7) *Pronumele*: Sunt de obicei neaccentuate în propoziție.
- 8) *Propozițiile exclamative*: **Cât de bine-mi pare!** **Of**, ce necaz!
- 9) *Accentul în frază*: Propoziția accentuată poate fi principală sau secundară. Propozițiile coordonate de obicei nu diferă prin accentuare, fiind puse de vorbitor pe același plan, mai ales la coordonarea copulativă (propoziții legate prin conj. **și**). Propozițiile disjunctive sunt de obicei ambele accentuate, fiind puse de vorbitor pe același plan: **Hotărăște-te ce faci, sau pleci, sau rămâi.** La adversative poate fi accentuată propoziția a doua: **Nu plânge, ci se preface.** La fel la coordonarea concluzivă - **E vreme foarte urâtă, deci rămân acasă.** În cazul frazelor formate prin subordonare, în principiu orice propoziție, regentă sau subordonată, poate fi accentuată. (Cf. *Gramatica limbii române*, ed. a II-a revăzută și adăugită, București, 1963, vol. II, p. 462-478)

4.4. Rolul semnelor de punctuație în prozodie

Valorile semnelor de punctuație sunt în primul rând semantice (așa cum o dovedesc și 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 punctuație redau și prozodia.

Punctul: Aserțiunile neutre, care îmbracă forma unor fraze enunțiative, sunt rostite, în general, cu o intonație continuu descendentă, reprezentată prin tiparul descendent declarativ. Acest tipar și pauza care urmează sunt indicate prin punct.

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

Semnul exclamării: Folosit pentru propoziții enunțiative afective, acest semn de punctuație are influența poate cea mai vagă asupra prozodiei.

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

Virgula. Marchează pauzele semnificative din interiorul enunțurilor, având o valoare cuprinsă între valorile pauzelor reprezentând pauza de sfârșit de enunț și pauza normală dintre cuvinte sau sintagme. În majoritatea cazurilor reprezintă delimitatorul între componentele unei înșirui, respectiv delimitează o subordonată în cadrul unei fraze.

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

Parantezele. La prima aproximație le vom echivala cu efectul virgulei, urmând să revenim asupra definiției în urma rezultatelor experiențelor acustice.

Cratima: Din punct de vedere prozodic, cratima redă pronunțarea “legată” a unor cuvinte care pot avea sau nu și existență independentă, notând o realitate *fonetică permanentă* sau *accidentală*. Această legare este însoțită uneori și de anumite modificări fonetice: *sinereza* și/sau *eliziunea*. *Sinereza* înseamnă pronunțarea accidentală într-o singură silabă a vocalei finale a unui cuvânt și a vocalei inițiale 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*) față de tempoul lent (*de abia*). *Eliziunea* înseamnă căderea accidentală a vocalei neaccentuate de la finala unui cuvânt în contact cu vocala inițială a cuvântului următor. Și aceasta poate fi obligatorie (de ex. *dintr-un*, *printr-o*, *m-a văzut* etc.) sau facultativă, diferențiind în scris tempoul rapid față de cel lent al rostirii (de ex. *c-a văzut* -> *că a văzut*, *c-un copil* -> *cu un copil* etc.). Fără funcție gramaticală cratima se poate utiliza și la redarea în scris rostirea în tempo rapid a derivatelor cu prefixele *ne-* și *re-* de la temele care încep cu *îm-*, și *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 la 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 strânse corelații pe termen scurt sau pe termen lung cu atribute de morfologie și sintaxă aferente textului. Principalele atribute se referă la poziționare accent în cuvinte, silabificare, părțile de vorbire, sintaxa, respectiv punctuație. Aceste rezultate prezintă fundamentul pentru dezvoltarea unor noi metode de sinteză expresivă 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 Todor, „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 segmentation” 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.