

INSTITUTUL POLITEHNIC BUCUREȘTI
FACULTATEA DE ELECTRONICĂ ȘI TELECOMUNICAȚII

TEZA DE DOCTORAT

SISTEME CU VOEBIRE ARTIFICIALĂ PENTRU
TELECOMUNICAȚII

Ing. LORIN FORTUNA

BIBLIOTECA CENTRALĂ
UNIVERSITATEA "POLITEHNICA"
TIMIȘOARA

557408
X
MEXICO

Conducător științific:
Prof.dr.ing. ADALIDA MATEASCU

INSTITUTUL POLITEHNIC TIMIȘOARA	
BIBLIOTECA	
CENTRALĂ	
Volanul nr.	557408
Data	779

1991

1. METODA DE SINTEZA A VORBIRII

1.1. Introducere. Clasificări

În general, un sintetizor de vorbire poate fi definit ca un aparat capabil, să genereze mesaje vorbite, pe baza unor înregistrări înregistrate în prealabil, dar care diferă din punct de vedere semantic de mesajul generat.

Informația mesajului se constituie ca o mulțime de parametri asupra cărora trebuie efectuată o anumită prelucrare în vederea obținerii mesajului ce trebuie generat. Această prelucrare este definită ca fiind operația de sinteză a vorbirii artificiale, vorbire care, în acest caz, se mai numește și vorbire sintetizată.

Caracteristică sintezei vorbirii mai este și faptul că trebuie să se realizeze în timp real, existând în acest caz o asemănare între sinteza vorbirii și pronunția vorbirii umane, unde, de asemenea, trecerea de la stadiul de concept la cel de undă acustică trebuie să se realizeze în timp real, pentru a se asigura continuitatea vorbirii și perceperea mesajului.

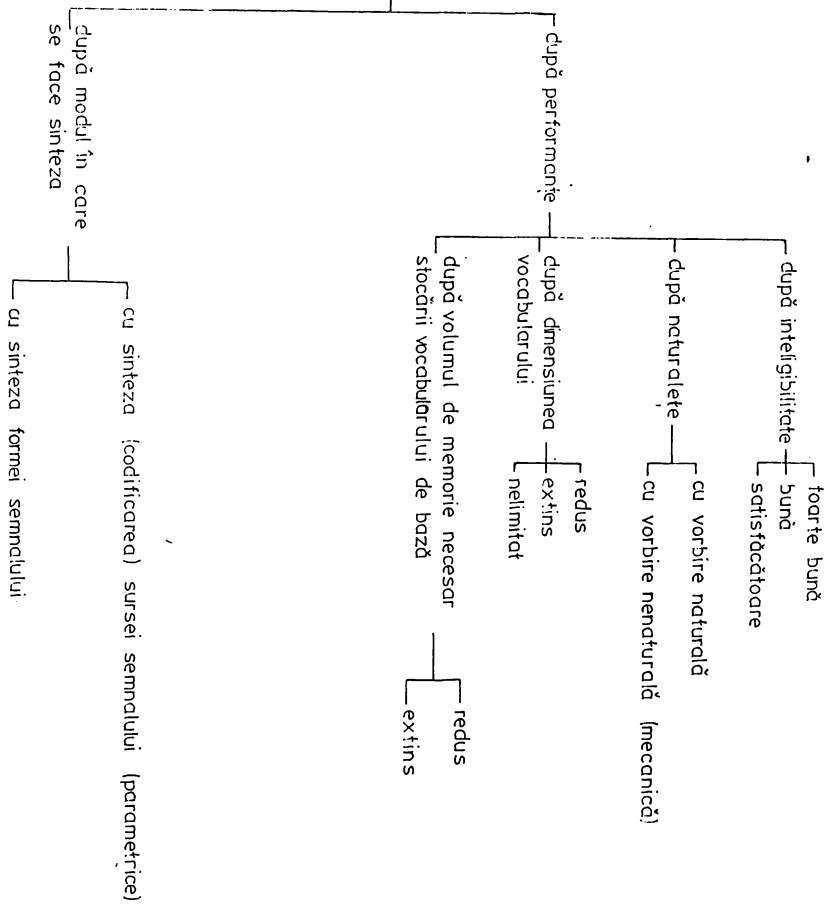
O altă caracteristică importantă a sintezei vorbirii este aceea că pentru a selecta informația de bază, ce urmează a fi înregistrată în sintetizor, este necesară o operație prealabilă, numită analiză, prin care se extrag din vorbirea naturală un set de parametri ce se constituie apoi ca informație de bază în vederea sintezei. Funcție de felul în care se face analiza, depinde și tipul acestor parametri, precum și modul în care se realizează sinteza, rezultând astfel mai multe metode de sinteză a vorbirii.

Există două criterii generale de clasificare a sintetizorilor de vorbire:

- după performanțele sintetizorului;
- după felul în care se realizează sinteza.

Principali parametri ce definesc performanțele unui sintetizor de vorbire sînt: inteligibilitatea, naturalitatea, dimon-

CRITERII DE CLASIFICARE
A SINTETIZAZOARELOR DE
VORBIRE



TABEL 1.2. CRITERII GENERALE DE CLASIFICARE A SINTETIZAZOARELOR DE VORBIRE

siunea vocabularului precum și volumul de memorie necesar stocării informației de bază.

Ca și în alte cazuri, îmbunătățirea tuturor acestor parametri introduce condiții contradictorii, ceea ce implică alegerea unor soluții de compromis pentru a obține o situație optimă din punct de vedere al unuia sau mai mulți parametri. Conform acestor parametri se pot face următoarele clasificări (Tab.1.0.).

Mai este de remarcat și faptul că aceleași metode de sinteză pot fi realizate fie în tehnică analogică, fie în tehnică digitală, ultima impunându-se în exclusivitate odată cu răspîndirea tehnicilor de prelucrare digitală a semnalelor, cu ajutorul sistemelor cu logică programată /2/.

1.2. Metoda de sinteză a vorbirii cu codificarea sursei de semnal (Metode parametrice de sinteză) /1/, /2/

1.2.1. Introducere. Clasificări

Aceste metode sintetizează de fapt sursa generatoare de vorbire, adică simulează funcționarea tractului vocal și a semnalului de excitație a aceluia.

Ele au avantajul că necesită un volum mai redus de informații de bază și, ca atare, un volum mai redus de memorie pentru stocarea lor, dar prezintă și dezavantajul că impun prelucrări mai complexe asupra acestor informații, pentru a realiza sinteza. Totodată, etapa de analiză a vorbirii, efectuată în scopul selectării informației de bază, pentru sinteză, este mai complexă și deoarece această informație diferă semnificativ de vorbirea propriu-zisă, ea se constituie ca un set de parametri necesari sintezei, motiv pentru care aceste metode au mai fost denumite și metode parametrice de sinteză a vorbirii /2/.

Setul de parametri ce alcătuiesc informația de bază în vederea sintezei poate fi separat în două categorii și anume:

- parametri aferenți tractului vocal (t_1), și
- parametri aferenți surselor de excitație a tractului vocal (a_1).

Astfel, schema bloc generală a unui sintetizor parametric de vorbire este cea reprezentată în figura 1.1.

Această schemă bloc generală mai poate fi detaliată funcție de tehnica, analogică sau digitală, în care se realizează sinteza.

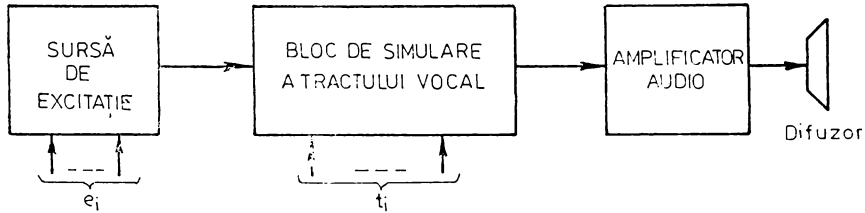


Fig.1.1. Schema bloc generală a unui sintetizor parametric de vorbire

În figurile 1.2 și 1.3 sînt reprezentate schemele bloc corespunzătoare, celor două cazuri, scheme ce simulează modelele mecanic și respectiv electric ale aparatului fonator-uman /3/,/4/.

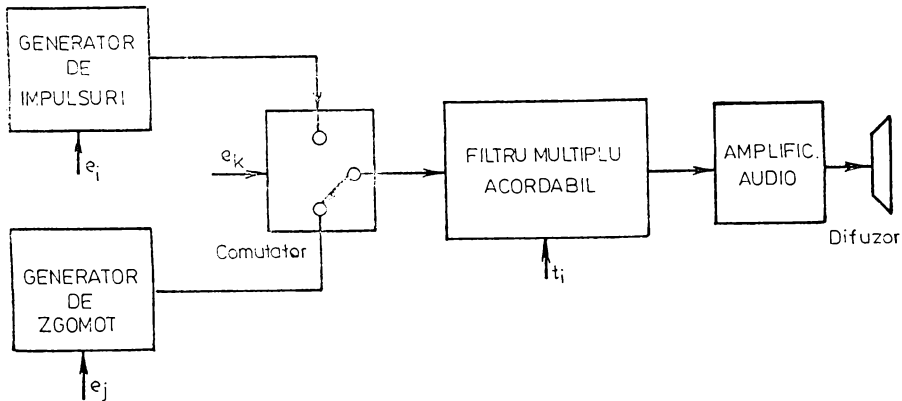


Fig.1.2. Schema bloc a unui sintetizor parametric analogic

Prin intermediul celor două generatoare se furnizează semnalul de excitație, echivalent fluxului de aer, modulat sau nu de corzile vocale.

Semnalul de excitație a tractului vocal poate fi simulat din punct de vedere electric cu un semnal periodic, de impulsuri aproximativ triunghiulare, în cazul producerii sunetelor sonore, și cu un semnal de zgomot, în cazul producerii sunetelor nesonore.

Blocul comutator are rolul de a comuta unul din cele două generatoare de semnal de excitație, în funcție de tipul

sunetului ce urmează a fi sintetizat (sonor sau nesonor), comanda comutării efectuându-se prin intermediul unui parametru binar (a_k).

Filtrul multiplu acordabil are rolul de a simula comportarea în frecvență a tractului vocal, el fiind realizat, de regulă, prin conectarea în serie, sau paralel, a mai multor filtre simple, acordabile, a căror acord se realizează prin intermediul parametrilor t_j . Se modelează astfel comportarea tractului vocal, pentru producerea unui anumit sunet, utilizând un vector de comandă specific aceluia sunet, a cărui componente sînt tocmai valorile parametrilor de sinteză corespunzătoare producerii sunetului respectiv:

$$V = V(a_i; a_j; a_k; t_i) \quad (1.1)$$

Prin intermediul parametrilor a_j și a_i se comandă amplitudinea semnalului furnizat de generatorul de zgomot și respectiv frecvența semnalului dat de generatorul de impulsuri, echivalentă frecvenței fundamentale de vibrație a corzilor vocale.

Deci prin intermediul parametrului a_i se poate regla înălțimea "voicii" sintetizatorului, putîndu-se astfel simula o voce masculină, feminină, sau mecanică. Parametrii de comandă a filtrului multiplu sînt specializați, de regulă, pentru:

- comanda frecvenței de rezonanță;
- comanda lățimii de bandă;
- comanda amplitudinii semnalului.

Filtrul multiplu, ce simulează funcționarea tractului vocal, trebuie să fie format dintr-un număr suficient de elemente componente, astfel încît să permită realizarea unei bune aproximări a înfășurătorii spectrale corespunzătoare fiecărui sunet de vorbire, fără de care sinteza ar conduce la un sunet neinteligibil. De regulă, în sintetizoarele analogice se utilizează structuri simplificate ale filtrului multiplu, care au doar rolul de a reface înfășurătoarea spectrală corespunzătoare zonelor în care apar formanții specifici sunetului respectiv, conform proprietăților acestor formanți de a fi determinați (în special primii trei) pentru recunoașterea sunetelor vorbirii.

În cazul sintetizatorului parametric digital, locul generatorului de zgomot este luat de un generator de impulsuri aleatoare, iar în locul generatorului de impulsuri periodice

triunghiulare, se utilizează un generator de impulsuri periodice dreptunghiulare. Filtrul analogic este de asemenea înlocuit cu un filtru digital comandabil, urmat de un convertor numeric-analogic.

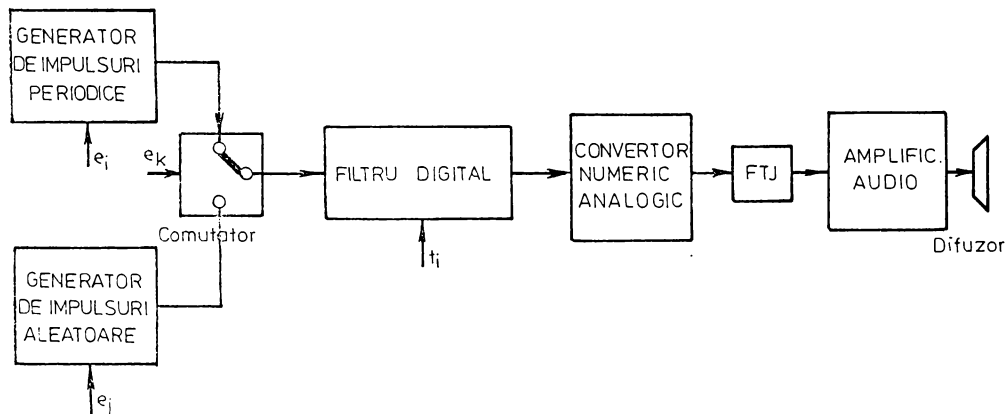


Fig.1.3. Schema bloc a unui sintetizor parametric digital

Deși, de-a lungul timpului, s-au dezvoltat mai multe metode de sinteză parametrică a vorbirii ca :

- sinteza fonemică ;
- sinteza prin refacerea înfășurătorii spectrale ;
- sinteza prin formanți ;
- sinteza homomorfică ;
- sinteza prin predicție liniară, etc.,

totuși, doar două din aceste metode s-au impus în ultima vreme, ajungînd să fie implementate și sub forma unor circuite integrate specializate, de largă circulație /1/, și anume : sinteza prin formanți și sinteza prin predicție liniară.

Mai trebuie remarcat că, în ordinea apariției, primele sintetizoare au fost cele parametrică, în timp ce sintetizoarele bazate pe codificarea formei semnalului au apărut abia mai târziu, odată cu dezvoltarea metodelor de codificare și prelucrare numerică a semnalelor și cu apariția sistemelor cu logică programată pentru aplicarea acestor metode.

Sintetizoarele parametrică au fost denumite la început și "vocodere" /3/ fiind considerate ca niște sisteme de codificare a vocii și utilizate atît pentru sinteza vorbirii, cît

și în telecomunicații, pentru reducerea redundanței semnalului vorbit, în scopul creșterii gradului de multiplexare a canalului de transmisie /4/. Trebuie specificat însă faptul că un vocoder includea, pe lângă sintetizorul propriuzis, și un analizor, cu rolul de a obține informația de bază necesară sintezei, conform figurii 1.4.

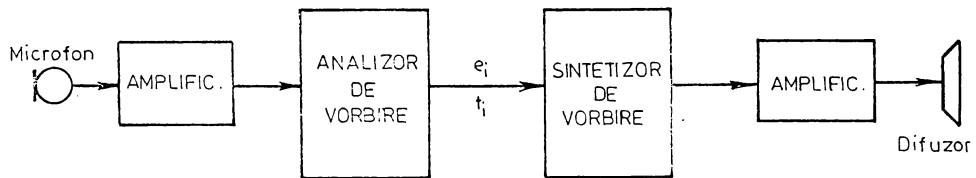


Fig.1.4. Schema bloc a unui vocoder

Odată cu dezvoltarea sistemelor cu logică programată, analiza a început să fie efectuată doar ca o etapă de obținere a parametrilor pentru sinteză, care au fost apoi memorati în memoria sistemului, urmînd a fi extrași și folosiți, pentru sinteză, atunci cînd este nevoie.

În felul acesta, rolul vocoderului a fost preluat doar de sintetizor, care s-a impus ca un sistem de sine stătător.

1.2.2. Sinteza fonemică /4/

S-a constatat că vorbirea normală produce, în medie, aproximativ 10 foneme/secundă, care ar putea fi codificate și transmise cu un debit de numai 60 biți/secundă. Comparînd acest debit cu cel corespunzător codificării undei acustice prin metoda MIC standard (64 kbiți/secundă) rezultă că dacă s-ar reuși codificarea vorbirii direct la nivel fonemic, s-ar putea reduce debitul de peste 1000 ori, ceea ce ar conduce la o economie spectaculoasă de memorie, la înregistrare, sau de volum de canal pentru transmisie.

Pentru a obține însă o compresie atît de mare și a extrage numărul minim de parametri necesari sintezei fonemelor, se impun condiții foarte dificile operației de analiză, care devine foarte complexă, iar vorbirea sintetizată, în acest mod, nu va păstra decît inteligibilitatea, din totalitatea caracte-

răsticilor vorbirii umane, fiind neplăcută și obositoare la ascultat și interpretat.

Deși s-au făcut multe încercări de realizare a unor sintetizoare fonemice, limita maximă de compresie nu a fost încă atinsă. Sinteza vorbirii, pe baza fonemelor, presupune stabilirea unui mare număr de reguli, referitoare nu numai la modul de producere al unui fonem, dar și la alăturarea lui cu alte foneme, la accentuarea sau neaccentuarea lui funcție de cuvântul în care intervine, etc.

Pe de altă parte, se pune problema stabilirii unui număr minim de reguli care să permită sinteza fonemică.

Cercetări fonetice, întreprinse în acest domeniu /4/, au evidențiat că există un număr de aproximativ 12 opțiuni binare, de tip "da" sau "nu", de care dispune omul pentru a produce vorbirea, fenomen valabil pentru toate limbile, și care sînt expuse și comentate în tabelul din figura 1.5.

Modelarea electrică eficientă a acestor condiții rămîne însă deocamdată o problemă de viitor.

Caracterizarea sintezei fonemice printr-un mare număr de reguli a făcut ca această metodă să fie inclusă în așa numita categorie de metode de "sinteză prin reguli" /4/, apreciindu-se că reprezintă o cale pentru realizarea sintetizoarelor de vorbire cu vocabular nelimitat, în condițiile cele mai avantajoase, dar, deocamdată, greu de abordat /1/.

În figura 1.6 este indicată schema bloc a unui sintetizor fonemic prin reguli /6/.

Sinteza prin reguli poate fi considerată ca o metodă de transformare a unei reprezentări discrete a vorbirii, într-o formă acustică continuă, echivalentă. Transformarea se realizează în două etape. În prima se transformă parametrii discreti, ce caracterizează fonemele, în parametri continui, pentru controlul funcționării unui sintetizor, iar în etapa a doua se transformă parametrii de control ai sintetizorului în vorbire sintetică, sintetizorul fiind de tipul celui reprezentat în figurile 1.2 sau 1.3.

Dezvoltarea metodelor de analiză a vorbirii, utilizînd logica programată, a condus și la elaborarea unor algoritmi complexi de sinteză prin reguli, specializați pentru diferite limbi, și aplicați mai ales în realizarea unor sisteme de conversie directă text-vorbire /26, 27/.

Nr. crt.	Caracterul	Caracteristici acustice
1.	Vocal/nevocal	Existența/absența unei structuri formantice evidente.
2.	Constant/neconstant	Nivel mare/nivel mic al energiei totale.
3.	Compact/difuz	Concentrație mare/concentrație mică a energiei (a intensității) într-un domeniu relativ îngust al spectrului, însoțită de o creștere/scădere a energiei totale.
4.	Incordat/neîncordat	Energie totală mai mare/energie totală mai mică, împreună cu o dispersare mai mare-mai mică a energiei în spectru și în timp.
5.	Sonor/surd	Existența/absența excitației periodice de frecvență joasă.
6.	Nazal/oral	Dispersarea energiei într-o bandă de frecvențe mai largă/mai îngustă datorită scăderii intensității unor formanți (în special a primului formant) și introducerea formanților suplimentari (nazali).
7.	Exploziv/neexploziv	Pauza după care urmează și/sau care precede o dispersare a energiei într-o gamă largă de frecvențe (explozia sau tranziția rapidă a formanților vocalelor, sau absența tranzițiilor bruște între sunet și o asemenea pauză).
8.	Acut/plat	Intensitatea zgomotului mai mare/mai mică
9.	Oclusiv/neoclusiv	Viteză mai mare de scădere a energiei în limitele unui interval mai mic de timp/viteză mai mică de scădere a energiei în limitele unui interval de timp mai mare.
10.	Tonalitate joasă/tonalitate înaltă	Concentrația energiei în porțiunea inferioară/superioară a spectrului.
11.	Bemol/simplu	Unele componente de frecvență se deplasează în jos/nu se deplasează, sau sînt/nu sînt atenuate.
12.	Diez/simplu	Unele componente de frecvență se deplasează/nu se deplasează în sus, sau sînt/nu sînt amplificate.

Fig.1.5. Tabel cu caracteristici discriminatorii binare ale fonemelor

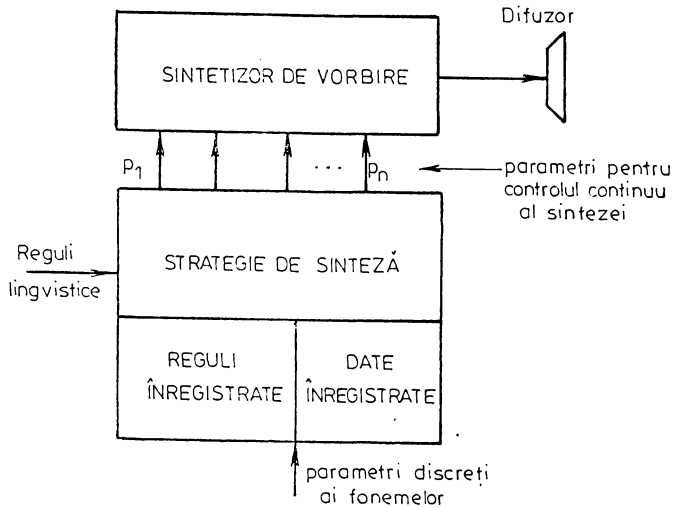


Fig.1.6. Schema bloc a unui sintetizor fonemic prin reguli

1.2.3. Sinteza prin refacerea înfășurătorii spectrale /2/

Această metodă utilizează ca parametri pentru sinteză o serie de valori ale înfășurătorii spectrale a semnalului de vorbire, determinați printr-o analiză prealabilă. De regulă este utilizată pentru analiză banda vocală telefonică (300-3.400 Hz), sau banda 200-4.000 Hz, avînd în vedere că frecvențele ce depășesc valoarea de 4.000 Hz au o importanță redusă pentru inteligibilitate, contribuind mai ales la asigurarea calității vorbirii, atît din punct de vedere a caracterului natural uman al vocii, cît și a recunoașterii vorbitorului.

Calea cea mai simplă de analiză și sinteză, prin această metodă, constă în eșantionarea, în frecvență, cu un set de filtre adiacente trece bandă, a semnalului de vorbire, conform figurii 1.7.

Ieșirile acestor filtre sînt prelucrate (de regulă prin redresare și integrare), obținîndu-se astfel un set de parametri de bază, sub forma unor tensiuni lent variabile, cu care se comandă apoi realizarea sintezei.

Calitatea sintezei va depinde, pentru o bandă de frecvență fixată pentru analiză, de intervalul $\Delta\omega$ dintre două eșantioane. Cu cât acesta va fi mai mic, -cu atât aproximarea înfășurătorii va fi mai bună, dar cu atât și numărul parametrilor va fi mai mare, iar sinteza va fi mai complicată.

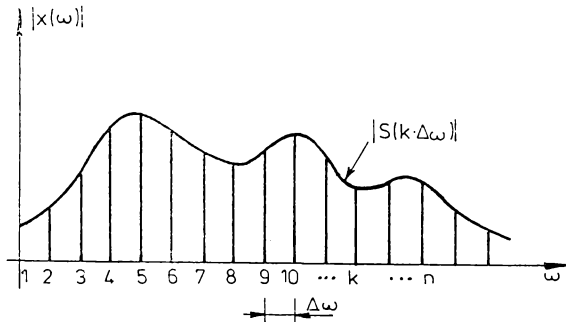


Fig.1.7. Infășurătoare spectrală a unui semnal de vorbire la un moment de timp dat

Sintetizoarele care funcționează pe acest principiu se numesc sintetizoare de bandă, iar schema bloc de analiză a unui canal este indicată în figura 1.8.

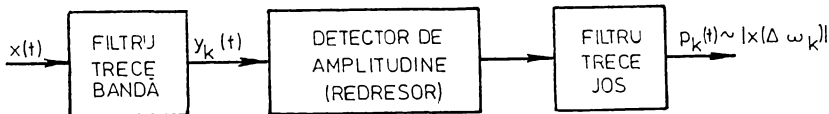


Fig.1.8. Schema bloc a analizorului de canal corespunzător sintetizorului de bandă

Pentru a modela matematic obținerea unui astfel de parametru /4/ se ține cont că semnalul de la ieșirea filtrului trece bandă se obține prin convoluția :

$$y_k(t) = \int_0^t x(\tau) \cdot g_k(t-\tau) d\tau \quad (1.2)$$

în care

$$g_k(t) = \frac{\Delta\omega}{\pi} \frac{\sin \frac{\Delta\omega}{2}(t-\tau)}{\frac{\Delta\omega}{2}(t-\tau)} \cdot \cos \omega_k t \quad (1.3)$$

este funcția de răspuns a filtrului considerat ideal avînd caracteristica de fază liniară în banda de trecere și caracteristica de frecvență dată de relația :

$$G_k(\omega) = \begin{cases} 1 & \text{pentru } \omega_1 \leq |\omega| \leq \omega_2 \\ 0 & \text{pentru } \omega_2 \leq |\omega| \leq \omega_1 \end{cases} \quad (1.4)$$

$\Delta\omega = \omega_2 - \omega_1$ (1.5) fiind banda de trecere, iar ω_k frecvența centrală a acestei benzi. Notînd cu $h(t)$ înfășurătoarea funcției de răspuns a filtrului, relația (1.3) devine :

$$g_k(t) = \frac{\Delta\omega}{\pi} \cdot h(t-\bar{t}) \cdot \cos \omega_k t \quad (1.6)$$

iar relația (1.2) :

$$y_k(t) = \frac{\Delta\omega}{\pi} \int_{-\infty}^t x(\bar{t}) \cdot h(t-\bar{t}) \cdot \cos \omega_k (t-\bar{t}) d\bar{t} \quad (1.7)$$

sau :

$$y_k(t) = \frac{\Delta\omega}{\pi} \operatorname{Re} e^{j\omega_k t} \int_{-\infty}^t x(\bar{t}) \cdot h(t-\bar{t}) \cdot e^{-j\omega_k \bar{t}} d\bar{t} \quad (1.8)$$

$$\text{Notînd } X(\omega_k, t) = \int_{-\infty}^t x(\bar{t}) \cdot h(t-\bar{t}) \cdot e^{-j\omega_k \bar{t}} d\bar{t} \quad (1.9)$$

relația (1.8) devine :

$$y_k(t) = \frac{\Delta\omega}{\pi} \operatorname{Re} X(\omega_k, t) \cdot e^{j\omega_k t} \quad (1.10)$$

și pune în evidență faptul că tensiunea de la ieșirea filtrului trece jos, este proporțională cu spectrul semnalului $x(t)$ și reprezintă un semnal modulat, semnificînd faptul că după detecție și filtrare se va obține o tensiune proporțională cu înfășurătoarea spectrală la frecvența de analiză ω_k :

$$p_k(t) = \frac{\Delta\omega}{\pi} \int_{-\infty}^{\infty} x(\bar{t}) \cdot h(t-\bar{t}) d\bar{t} \quad (1.11)$$

care va reprezenta chiar parametrul corespunzător canalului respectiv. Efectuînd astfel o analiză pe mai multe canale se obține setul de parametri utilizați la sinteză. Se constată că variația acestor tensiuni este lentă, astfel că în cazul în care sinteza se face pe cale numerică, conversia analog-numerică a acestor parametri va conduce la un debit total binar

mult mai scăzut comparativ cu cel rezultat în urma conversiei semnalului de vorbire propriu-zis.

La sintetizorul de bandă, lăţimea canalelor spectrale de analiză nu este aceeaşi pe toate canalele, ci se stabileşte în funcţie de caracteristicile analizorului auditiv uman. Astfel, deoarece s-a constatat că rolul cel mai important pentru percepţia vorbirii, corespunzătoare zonei primului formant, este cuprinsă în zona de pînă la 1000 Hz, în timp ce, peste această valoare, aportul la inteligibilitate al componentelor spectrale scade, aproximativ logaritmice, cu creşterea frecvenţei, eşantionarea înfăşurătorii se va realiza conform acestor proprietăţi, ajungîndu-se la repartizarea canalelor spectrale în bandă conform aşa numitei "scări de egală articulaţie" a lui Koenig /4/. Canalele vor fi astfel mai dese şi mai înguste pînă la 1000 Hz şi mai rare şi mai largi după această valoare.

În tabelul din figura 1.9 este indicată, spre exemplificare, distribuţia filtrelor corespunzătoare unui sintetizor de bandă realizat /7/. Sintetizorul permite sinteza oricărui cuvînt din limba română, cu o inteligibilitate satisfăcătoare, dar cu o voce nenaturală, mecanică.

Canal	Bandă de trecere	Frecvenţa centrală	Canal	Bandă de trecere	Frecvenţa centrală
1	300 Hz	275 Hz	7	300 Hz	1663 Hz
2	150 Hz	540 Hz	8	330 Hz	1980 Hz
3	180 Hz	704 Hz	9	360 Hz	2320 Hz
4	210 Hz	893 Hz	10	380 Hz	2690 Hz
5	240 Hz	1120 Hz	11	400 Hz	3080 Hz
6	270 Hz	1379 Hz	12	600 Hz	3650 Hz

Fig.1.9. Tabel cu distribuţia canalelor spectrale pentru un sintetizor de bandă.

Schema bloc a unui sintetizor de bandă este, în principiu, similară celei din figura 1.2, cu menţiunea că în locul filtrului multiplu se utilizează un bloc de filtre trece bandă, a căror caracteristici sînt echivalente cu cele utilizate în analizorul din care se obţin parametrii ce comandă sinteza, prin reglarea nivelului semnalului aplicat la intrarea fiecărui filtru, astfel încît suma semnalelor de ieşire să refacă înfăşurătoarea semnalului de vorbire. În figura 1.10 este reprezentată schema bloc a unui vocoder de bandă realizat /7/, /8/, /9/.

compus dintr-un analizor și un sintetizor de vorbire.

În cazul variantei digitale, la sintetizorul de bandă prezentat în figura 1.10, se înlocuiesc filtrele de bandă cu filtre digitale, iar parametrii de comandă se convertesc analog-numeric și comandă nivelul semnalului de la ieșirea fiecărui filtru.

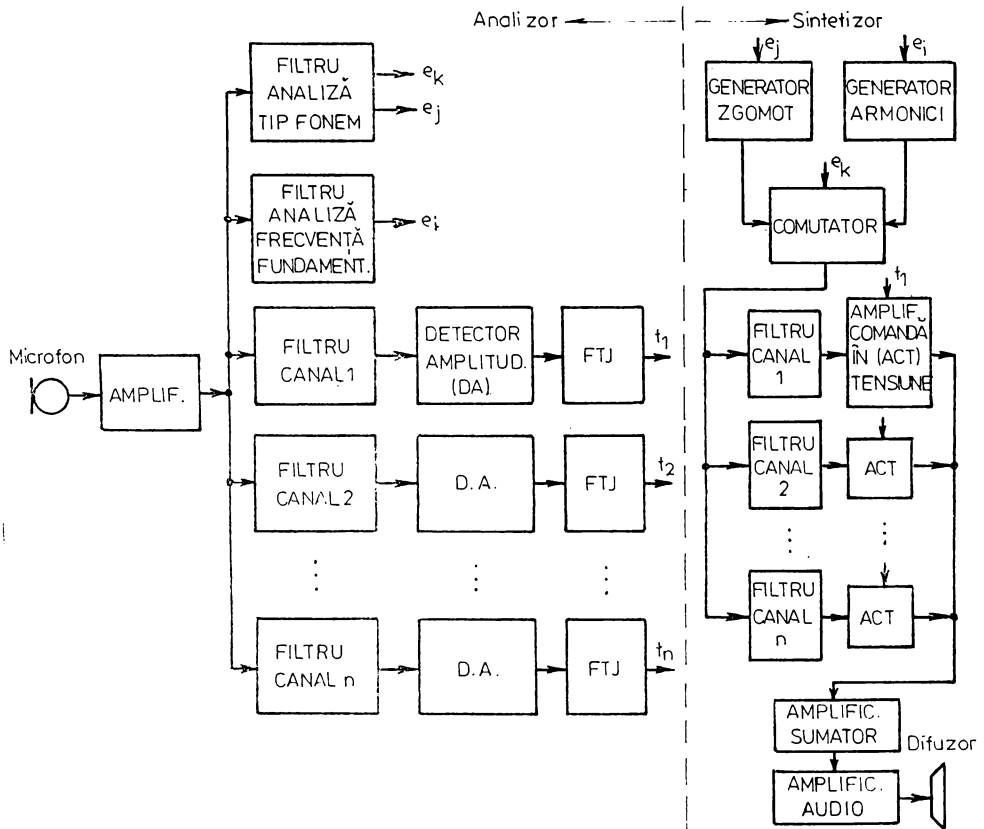


Fig.1.10. Schema bloc a unui vocoder de bandă

O variantă digitală mai complexă se poate utiliza cu ajutorul unui sistem de achiziție și prelucrări digitale a semnalelor analogice, pe care să fie implementat un analizor Fourier și prin intermediul căruia să se realizeze eșantionarea spectrului semnalului de vorbire pentru a obține setul de parametri, cu care să se realizeze apoi, tot pe cale digitală, și sinteza.

Viteza de calcul necesară unui astfel de sistem este însă mare, datorită necesității de a funcționa în timp real și

ea nu poate fi realizată prin implementarea microprocesoarelor de uz general, ci doar cu procesoare de semnal, sau bit-slice.

Cercetări actuale, în domeniul vocoderilor, au ca scop îmbunătățirea performanțelor, iar ca obiect perfecționarea modului de selectare a parametrilor utilizați pentru sinteză.

Dintre acești parametri, frecvența tonului fundamental și timpul fonemului (sonor/nesonor) sînt luați, îndeosebi, în considerare, avînd în vedere contribuția lor decisivă la calitatea vorbirii sintetizate /29,30/.

În acest context, o serie de cercetări recente au evidențiat faptul că utilizarea unei decizii unice asupra tipului fonemului, la sinteză, este o cauză care introduce distorsiuni importante /10/.

A rezultat astfel că limitarea sursei de excitație a modelului tractului vocal numai la un semnal periodic, sau la un semnal de zgomot, constituie o limitare care, la sinteză, face ca energia spectrală a anumitor benzi de frecvență, din spectrul fonemului sintetizat, să difere substanțial de cea a fonemului original și să constituie una din cele mai importante surse de distorsiuni, specifică vocodereilor de bandă.

O altă sursă de perturbații este legată de determinarea parametrilor ce caracterizează modelul tractului vocal.

Pentru a obține o calitate îmbunătățită a performanțelor vocoderilor, cercetările actuale au în vedere atît un model îmbunătățit al tractului vocal, cît și metode mă perfecționate pentru estimarea parametrilor acestui model și a sursei de excitație. Una dintre cele mai interesante realizări în acest domeniu o constituie modelul excitației multibandă /10/, în care, pentru banda de frecvență din jurul fiecărei armonici a frecvenței fundamentale, se ia o decizie de a fi considerată drept sonoră (vocală), sau nesonoră (nevocală).

Se utilizează, de asemenea, și o metodă mai perfecționată de determinare a parametrilor tractului vocal. Astfel, parametrul clasic, specific sursei de excitație, ce indică tipul fonemului, la nivel global, cu o funcție (sonor/nesonor) dependentă de frecvență. Spectrul fonemului analizat este divizat, în acest scop, în mai multe benzi (peste 20), grupate în jurul armonicilor frecvenței fundamentale. Se analizează spectrul fiecăreia din aceste benzi și, în func-

ție de componența acestuia, se stabilește dacă banda respectivă este o bandă sonoră, sau nesonoră.

Datorită naturii quasistaționare a semnalului vorbit, după eșantionare, $[x(n)]$, se aplică mai întâi o funcție fereastră $w(n)$, pentru a separa astfel un interval de 10-40 ms, obținându-se segmentul:

$$x_w(n) = w(n) \cdot x(n) \quad (1.12)$$

Fereastra $w(n)$ poate fi deplasată, în timp, pentru a selecta orice segment din reprezentarea fonemului respectiv. Pe un interval de timp scurt, transformata Fourier $X_w(\omega)$ a acestui segment poate fi modelată ca un produs :

$$X_w(\omega) = H_w(\omega) \cdot |E_w(\omega)| \quad (1.13)$$

în care $H_w(\omega)$ reprezintă înfășurătoarea spectrală a segmentului analizat, iar $E_w(\omega)$ semnalul de excitație.

Înfășurătoarea spectrală poate fi reprezentată, eventual, prin coeficienți liniari de predicție, coeficienți cepstrali și lățimile de bandă respective, sau, efectiv, prin benzi filtrate din spectrul semnalului original. De asemenea, parametrii sursei de excitație și cei ai anvelopei spectrale, se pot determina nu doar în două procese diferite, ci într-unul singur, metodă cunoscută și sub denumirea de "analiză prin sinteză".

Estimarea parametrilor menționați se face însă în două etape. În prima etapă, perioada și parametrii înfășurătorii spectrale sînt estimați prin minimizarea erorii între spectrul original $X_w(\omega)$ și cel sintetic $\tilde{X}_w(\omega)$. În etapa a doua se estimează decizia sonor /nesonor, în funcție de apropierea existentă între cele două spectre, corespunzătoare benzii din jurul fiecărei armonici a frecvenței fundamentale.

Parametrii acestui model de vorbire se determină minimizînd eroarea dată de relația :

$$\xi = \frac{1}{2\pi} \int_{-\pi}^{\pi} [|X_w(\omega)| - |\tilde{X}_w(\omega)|]^2 d\omega \quad (1.14)$$

Considerînd criteriul de eroare propus, în intervalele spectrale situate în jurul armonicilor frecvenței fundamentale, pentru armonica a m-a, criteriul de eroare va fi aplicat intervalului $[a_m, b_m]$, avînd lățimea egală cu cea a frecvenței fundamentale și centrat pe a m-a armonică a acestei frecvențe :

$$\xi_m = \frac{1}{2\pi} \int_{a_m}^{b_m} [|X_w(\omega)| - |A_m| \cdot |E_w(\omega)|]^2 d\omega \quad (1.15)$$

Pentru simplificare, în formula precedentă, s-a presupus că înfășurătoarea spectrală a fiecărui interval considerat $[a_m, b_m]$ este de amplitudine constantă: A_m .

Din condiția de minimizare a erorii ξ_m rezultă :

$$|A_m| = \frac{\int_{a_m}^{b_m} |X_w(\omega)| \cdot |E_w(\omega)| d\omega}{\int_{a_m}^{b_m} |E_w(\omega)|^2 d\omega} \quad (1.16)$$

Se pot astfel obține parametrii corespunzători înfășurătorii spectrale, presupusă de amplitudine constantă, pentru banda din jurul fiecărei armonici a fundamentalei, care se pot utiliza apoi la sinteză.

Eroarea totală minimă, pentru toate intervalele adiacente, adică pentru întregul semnal de excitație, corespunzător unei perioade a tonului fundamental, se calculează prin însumare :

$$\bar{\xi} \approx \sum_m \bar{\xi}_m \quad (1.17)$$

în care $\bar{\xi}_m$ este ξ_m în formula căreia (1.15) s-a înlocuit $|A_m|$ cu valoarea din relația (1.16).

Se pot determina astfel parametrii anvelopei spectrale care să minimizeze eroarea pentru o perioadă întreagă a frecvenței tonului fundamental. Experimental, s-a observat că eroarea tinde să varieze cu perioada T a frecvenței tonului fundamental. Se poate obține astfel o estimare inițială a perioadei, apropiată de minimul global $\bar{\xi}$, urmată de o determinare mai precisă, corespunzătoare și minimizării fiecărei componente $\bar{\xi}_m$ a erorii globale. În practică, în locul integralelor din relațiile (1.14, 1.15, 1.16) vor fi utilizate, bineînțeles, aproximările numerice ale funcțiilor respective.

Pentru o determinare mai exactă a valorii frecvenței tonului fundamental, se poate efectua și o analiză bazată pe programare dinamică, procedeu specific mai ales recunoașterii automate a vorbirii.

Estimarea deciziei sonor/nesonor pentru fiecare armonică a frecvenței tonului fundamental se face printr-o comparare, cu o valoare de prag a erorii normalizate, specifice zonei acelei armonici, definită prin relația :

557400 1770

$$\sigma_m^2 = \frac{\varepsilon_m}{\frac{1}{2\pi} \int_{a_m}^{b_m} |X_w(\omega)|^2 d\omega} \quad (1.18)$$

Dacă valoarea erorii normalizate este sub valoarea de prag, se consideră că pe $[a_m, b_m]$ spectrul semnalului va fi de tip sonor și de tip nesonor în caz că este depășită valoarea de prag. Apoi sînt determinați și parametrii ce caracterizează înfășurătoarea spectrală pe intervalele respective.

Sinteza se poate apoi obține prin efectuarea unei sumări de semnale sinusoidale, avînd frecvențele egale cu cele ale armonicilor fundamentalei și amplitudinile determinate de parametrii ce definesc valoarea înfășurătorii spectrale, pentru fiecare interval $[a_m, b_m]$ determinat ca fiind de tip sonor, precum și prin sumarea unor semnale de zgomot alb, filtrate conform intervalelor de tip nesonor, de amplitudine corespunzătoare coeficienților anvelopei spectrale în intervalele respective.

Performanțele vocoderului cu excitație multibandă rezultă printr-o comparație cu cele ale unui vocoder cu excitație simplăbandă, fiind, în general, mai bune și chiar mult mai bune în cazul în care vorbirea supusă analizei și apoi sintezei este însoțită de zgomot de fond /10/.

1.2.4. Sinteza pe bază de formanti /4,1 /

Această metodă de sinteză reprezintă, de fapt, un caz particular al metodei precedente, fiind considerată ca o metodă distinctă doar datorită faptului că este una din cele mai frecvent utilizate, făcînd obiectul mai multor implementări sub forma unor circuite integrate specializate /1, 11/.

Spre deosebire de metoda precedentă, ce avea ca obiect refacerea înfășurătorii spectrale în toată banda de frecvență stabilită pentru sinteză, sinteza pe baza de formanți își propune să refacă înfășurătoarea spectrală doar în acele regiuni care prezintă o importanță deosebită pentru inteligibilitatea vorbirii, adică în zona formanților /3,4,9/. Experimentele efectuate în acest domeniu /4/ au pus în evidență faptul că pentru a asigura o inteligibilitate foarte bună a vorbirii sintetizate este suficientă refacerea înfășurătorii spectrale corespunzătoare primilor trei formanți, importanța celorlalți doi fiind destul de redusă din acest punct de vedere.

Analiza va avea, în acest caz, rolul de a extrage din vorbirea naturală parametrii caracteristici pentru definirea formanților, precum și cei corespunzători sursei de excitație, pentru a comanda un sintetizor care, în principal, este de tipul celui reprezentat în figura 1.1. Parametrii formanților se referă la frecvența centrală, amplitudine și bandă corespunzătoare, măsurată, de regulă, la 3 dB față de nivelul componentei centrale.

Frecvențele formanților corespund unor frecvențe de rezonanță a tractului vocal, ce determină apariția unor maxime în spectrul semnalului de vorbire și ele se modifică permanent în timpul vorbirii, funcție de poziția organelor de articulație. Variații similare suferă și amplitudinile, precum și benzile acestor formanți, ultimele însă în destul de mică măsură, deoarece lățimea formanților depinde mai ales de pierderile ce au loc la aceste frecvențe în tractul vocal, ele putând fi considerate constante într-o primă aproximație /4/.

Importanța, de aceea, pentru asigurarea inteligibilității vorbirii cursive, este dinamica formanților, adică cunoaș-

terea modului de variație a frecvențelor centrale și a amplitudinilor lor.

Datorită inerției organelor de articulație, aceste variații sînt însă destul de lente, ceea ce face ca parametrii respectivi să ocupe un volum redus de semnal, iar conversia lor analog-numerică să conducă, în cazul utilizării tehnicii digitale pentru realizarea sintetizorului, la un volum foarte redus de memorie necesar stocării informației de bază, comparativ cu metoda precedentă.

Considerînd că se urmărește doar sinteza primilor trei formanți, se poate determina foarte exact numărul de parametri corespunzători pentru comanda filtrului multiplu, ca fiind 6 (trei referitori la frecvența centrală și trei la amplitudinea ei), la care se mai adaugă alți 3 corespunzători sursei de excitație (frecvența generatorului de armonici, amplitudinea și tipul generatorului de excitație, care va fi folosit pentru sinteza unui fonem). Calitatea vorbirii sintetizate prin metoda formanților depinde în cea mai mare măsură de corectitudinea cu care se realizează analize, referitor, mai ales, la determinarea corectă a frecvențelor centrale și a amplitudinii formanților, precum și a variațiilor lor în timp. Dacă determinarea acestor parametri pentru sunete izolate este o problemă destul de simplă, determinarea lor în cursul vorbirii cursive este destul de complicată, datorită atât variațiilor, cît și interferențelor care apar cu sunetele vecine. Domeniul de variație a primilor trei formanți pentru cele mai importante vocale ale limbii române este prezentat în tabelul din figura 1.11.a, pe baza căruia s-a realizat și graficul din figura 1.11.b, în care se observă că apar suprapuneri între domeniile de apariție a formanților pentru vocale diferite, ceea ce explică complexitatea problemei. De asemenea, problema sintezei prezintă complicații prin necesitatea de a realiza filtre comandabile, atât ca bandă, cît mai ales ca frecvență și amplitudine, corespunzătoare formanților. Tehnica analogică a fost în acest caz depășită complet de cea digitală, unde filtrele digitale prezintă facilități mult mai importante pentru generarea formanților.

Sintetizoarele analogice formentice au filtrul multiplu format dintr-un ansamblu de filtre, egal cu numărul formanților propuși a fi sintetizați și care pot fi conectate în serie, sau în paralel.

Astfel, schema bloc generală a unui sintetizor formantic poate avea una din cele două forme reprezentate în figura 1.12,

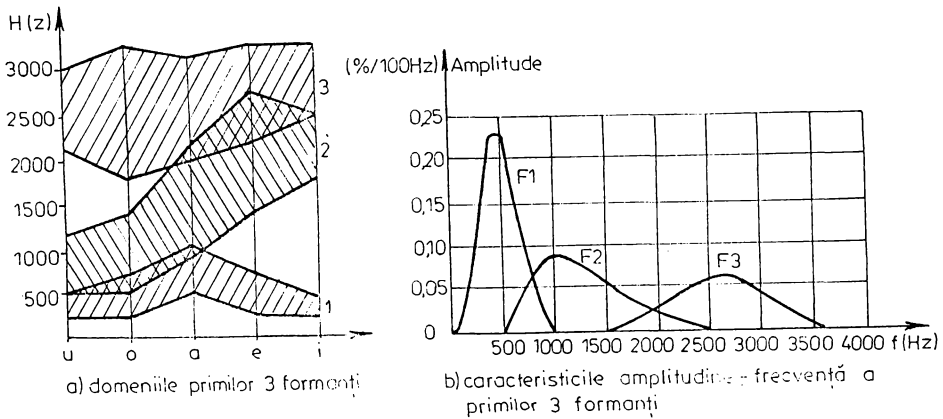


Fig.1.11. Date caracteristice formanților unor vocale ale limbii române

În varianta analogică de realizare, varianta digitală utilizând un singur filtru digital, datorită posibilităților oferite de filtrele digitale de a realiza structuri de filtre multiple /12/.

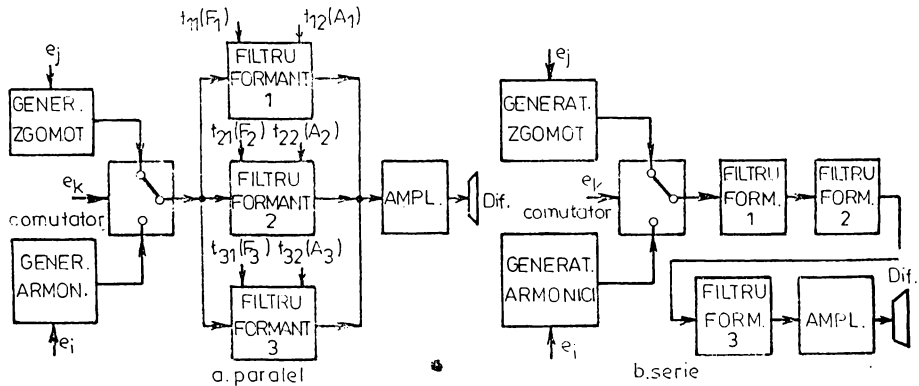


Fig.1.12. Schema bloc a unui sintetizor formantic pentru 3 formanți

Determinarea parametrilor formanților se poate realiza și în tehnică analogică, în special prin numărarea trecerilor prin zero a semnalului rezultat printr-o filtrare în zona de apariție a fiecărui formant (figura 1.13), dar soluția cea mai avantajoasă o oferă totuși analiza spectrală, prin transformata Fourier, realizată pe cale digitală, cu un sistem de achizi-

ție și prelucrării de date analogice prin metode digitale.

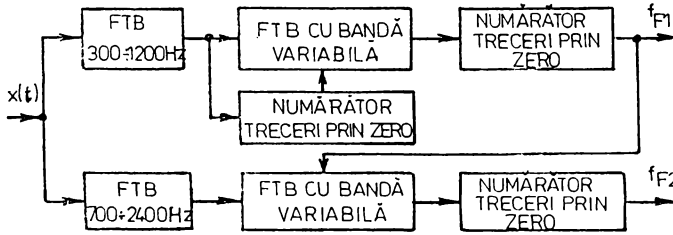


Fig.1.13. Schema bloc a unui sistem de determinare pe cale analogică a frecvențelor primilor doi formați.

Problema determinării formațiilor se complică în cazul consoanelor, comparativ cu vocalele, deoarece, în acest caz, formațiile nu sînt la fel de conturate ca la vocale, iar variația lor în timp este mai rapidă. Din aceste motive, la consoane, se recomandă ca fiind mai indicat să se măsoare momentele spectrale în care apar maxime de energie: $M_k(f_k; A_k)$ utilizîndu-se în acest scop niște valori medii și medii pătratice de frecvență:

$$f_I = \frac{\sum_k f_k \cdot A_k}{\sum_k A_k} \quad (1.19)$$

$$f_{II} = \frac{\sum_k f_k^2 \cdot A_k}{\sum_k A_k} \quad (1.20)$$

definiindu-se și o lățime (dispersie) a spectrului prin relația :

$$\Delta f = f_{II}^2 - f_I^2 \quad (1.21)$$

Un sintetizor formantic mai performant va conține astfel încă două filtre, unul special pentru consoane, comandat prin intermediul parametrilor ce definesc momentele spectrale ale acestora, un altul specific consoanelor nazale, conform schemei bloc din figura 1.14.

Un sintetizor formantic digital, destul de performant, implementat într-un circuit integrat specializat și foarte răspîndit la ora actuală este circuitul EBA 8000, realizat de firma Philips /1/, iar cercetări recente în domeniul sintezei prin formați urmăresc și obțin o vorbire sintetizată de bună calitate /24, 28/.

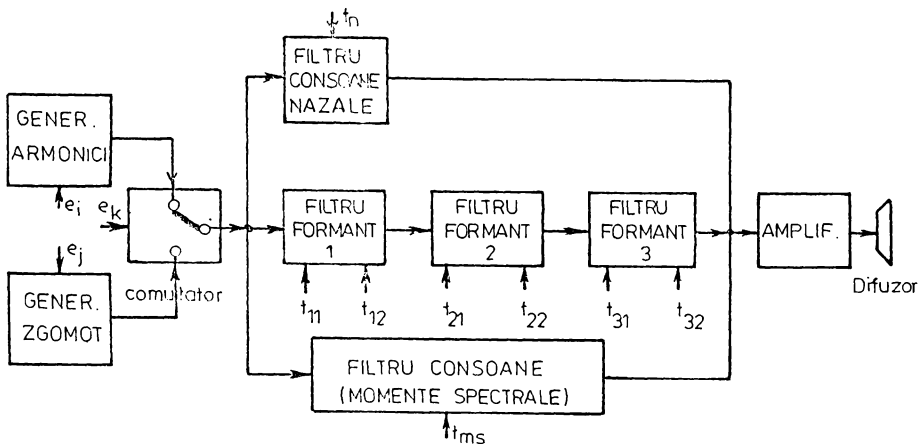


Fig.1.14. Scheme bloc a unui sintetizor formantic adaptat și pentru consoane

1.2.5. Sinteza prin predicție liniară /13/,/14/,/15/

1.2.5.1. Principiul predicției liniare /16/

Metoda predicției liniare este o metodă de analiză și sinteză a semnalelor reprezentate numeric, care au proprietatea că între eșantioanele succesive există o dependență specifică autocorelației.

Diferența dintre eșantioanele adiacente va avea în acest caz o gamă de variație mai redusă decât a semnalului însuși, proprietate utilizată și în cadrul metodelor diferențiale de codare și care favorizează codarea acestei diferențe cu un număr mai redus de biți, comparativ cu semnalul propriuzis.

Existența corelației între eșantioanele unui semnal va permite însă și o reprezentare a acestuia printr-o altă metodă, ce exploatează existența unei dependențe nu numai între două eșantioane adiacente, ci și în cadrul unei secvențe formate din mai multe eșantioane consecutive.

Predicția liniară constă în aceea că un eșantion al unei secvențe numerice corelate poate fi aproximat printr-o combinație liniară a mai multor eșantioane precedente conform relației:

$$x(n) = \sum_{k=1}^p a_k \cdot x(n-k) + G \cdot s(n) \quad (1.22)$$

în care a_k sînt niște coeficienți de ponderare, G un factor de câștig, iar $s(n)$ un semnal numit semnal de excitație.

Semnalul de vorbire fiind un semnal corelat, se pretează bine acestui mod de tratare. Similar metodelor de codare diferențială, se poate defini un semnal de aproximare (predicție) :

$$\tilde{x}(n) = \sum_{k=1}^p \alpha_k \cdot x(n-k) \quad (1.23)$$

numit predicție liniară, cu coeficienții $\{\alpha_k\}$ numiți coeficienți de predicție și cu sistemul care-l generează numit predictor liniar.

Intre relațiile (1.22) și (1.23) se poate defini eroarea de predicție $e(n)$ ca fiind diferența dintre cele două semnale :

$$e(n) = x(n) - \tilde{x}(n) = x(n) - \sum_{k=1}^p \alpha_k \cdot x(n-k) \quad (1.24)$$

pentru care se poate obține o reprezentare echivalentă, cu ajutorul transformatei z :

$$E(z) = A(z) \cdot X(z) \quad (1.25)$$

în care

$$A(z) = 1 - \sum_{k=1}^p \alpha_k \cdot z^{-k} \quad (1.26)$$

poate fi considerată ca funcție de transfer a unui sistem liniar numit filtru al erorii de predicție, conform figurii 1.15.

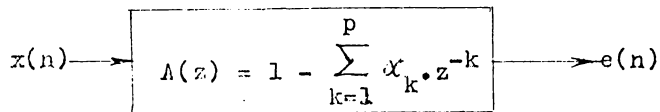


Fig.1.15. Simbolul filtrului erorii de predicție

Conform relațiilor (1.25) și (1.26) se poate obține :

$$X(z) = \frac{1}{A(z)} \cdot E(z) = \frac{E(z)}{1 - \sum_{k=1}^p \alpha_k \cdot z^{-k}} \quad (1.27)$$

Relația precedentă indică posibilitatea de a realiza aproximarea funcției $X(z)$ cu o funcție de transfer de forma :

$$H(z) = \frac{G}{1 - \sum_{k=1}^p a_k \cdot z^{-k}} \quad (1.28)$$

ce corespunde funcției de transfer a unui filtru numeric de tip "numai poli" și în care G reprezintă un factor de câștig, ce urmează a fi determinat din condiția ca să se poată scrie :

$$\tilde{K}(z) = G \quad (1.29)$$

în care $\tilde{K}(z)$ este o aproximare a funcției $K(z)$. În domeniul timp relația precedentă este echivalentă, cu :

$$\tilde{e}(n) = \begin{cases} G & \text{pentru } n=0 \\ 0 & \text{în rest} \end{cases} \quad (1.30)$$

Valoarea factorului G se determină aplicând condiția de conservare a energiei între funcția $e(n)$ și eroarea de predicție minimă a secvenței $\{e_n\}$:

$$\bar{e}_n = \sum_m \tilde{e}_n^2 = \sum_m e_n^2 \quad (1.31)$$

în care \bar{e}_n este eroarea pătratică minimă totală, definită și ca energie a erorii de predicție. Dacă se consideră :

$$a_k = \alpha_k \text{ și se ține cont de rel. (1.22),} \quad (1.32)$$

relația (1.24) devine :

$$e(n) = G \cdot s(n) \quad (1.33)$$

ceea ce pune în evidență un aspect foarte important și anume că semnalul erorii de predicție este proporțional cu semnalul de excitație, prin intermediul factorului G , ceea ce înseamnă că semnalul de eroare poate fi utilizat și ca semnal de excitație

$$s(n) = \frac{1}{G} \cdot e(n) \quad (1.34)$$

1.2.5.2. Aplicarea predicției liniare la semnalul de vorbire

Având în vedere considerentele precedente, rezultă că tractul vocal poate fi modelat din punct de vedere al predicției liniare cu modelul din figura 1.16.

Datorită naturii nestaționare a semnalului vocal, estimarea coeficienților de predicție liniară trebuie făcută pe segmente scurte de timp (10-20 ms), pe care se constată o comportare a tractului vocal ce poate fi considerată staționară și corespunzătoare generării unui anumit sunet de vorbire. Pentru a determina coeficienții de predicție se utilizează tot metoda mi-

minimizării energiei erorii de predicție medie pătratică totale:

$$\bar{e}_n = \sum_m e_n^2(m) = \sum_m \left[x_n(m) - \sum_{k=1}^p \alpha_k \cdot x_n(m-k) \right]^2 \quad (1.35)$$

în care $x_n(m)$ reprezintă un segment vocal din vecinătatea eşan-
tionului n , pe care se consideră că s-a realizat analiza.

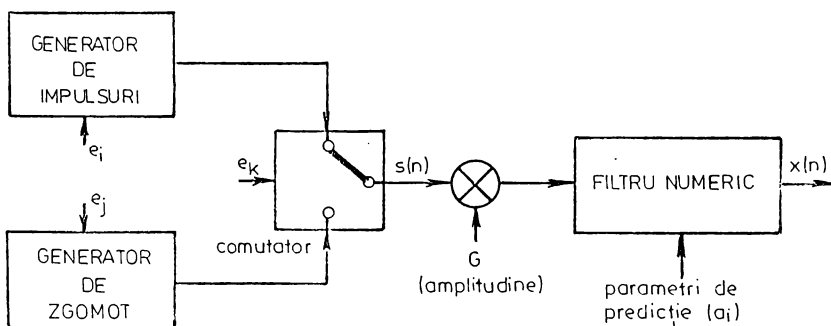


Fig.1.16. Model de aproximare a tractului vocal prin predicție liniară

Pentru minimizare, din derivarea relației precedente rezultă un sistem de ecuații diferențiale liniare :

$$\frac{\partial \bar{e}_n}{\partial \alpha_i} = 0 \quad i = 1 \dots p \quad (1.36)$$

care conduce la sistemul de ecuații :

$$\sum_m x_n(m-i) \cdot x_n(m) = \sum_{k=1}^p \alpha_k \cdot \sum_m x_n(m-i) \cdot x_n(m-k) \quad (1.37)$$

Introducând notația :

$$\phi_n(i, k) = \sum_m x_n(m-i) \cdot x_n(m-k) \quad (1.38)$$

sistemul (1.37) se mai poate scrie :

$$\sum_{k=1}^p \alpha_k \phi_n(i, k) = \phi_n(i, 0) \quad \text{cu : } 1 \leq i \leq p \quad (1.39)$$

Rezolvarea prin metode adecvate a acestui sistem de ecuații pune în evidență /2/ faptul că eroarea minimă totală e_T conține o componentă fixă și una variabilă, care depinde de coeficienții de predicție conform relației :

$$e_T = \phi_n(0,0) - \sum_{k=1}^p \alpha_k \phi_n(0,k) \quad (1.40)$$

Pentru rezolvarea sistemului de ecuații (1.39) s-au dezvoltat mai multe metode, care depind de modul de alegere a limitelor de sumare, precum și a segmentului vocal m . Dintre aceste metode, cele mai utilizate sînt metoda autocorelației și metoda rețelelor cu celule în X .

1.2.5.3. Determinarea coeficienților de predicție prin metoda autocorelației

Metoda autocorelației consideră semnalul de vorbire, supus analizei definit în cadrul unui segment

$$0 \leq m \leq N-1 \quad (1.41)$$

și nul în afara acestui interval.

Pentru a obține un astfel de efect, semnalul de vorbire este înmulțit cu un semnal de tip "fereastră" $w(m)$ conform relației :

$$x_n(m) = x(n+m) \cdot w(m) \quad (1.42)$$

ceea ce corespunde în calculul erorii medii pătrătice de predicție \bar{e}_n (relația 1.35) la o sumare pînă la $N-1+p$ conform relației :

$$\bar{e}_n = \sum_{m=0}^{N-1+p} e_n^2(m) \quad (1.43)$$

Eroarea de predicție va fi în acest caz mai mare la capetele intervalului segmentului considerat deoarece, la început, semnalul va fi forțat din eșantioane cărora li s-a atribuit forțat valoarea zero, iar la sfîrșit deoarece semnalul va fi forțat să ia valori nule din eșantioane nenule.

Pentru a atenua acest efect, funcția fereastră cu care se face ponderarea va trebui astfel încît să acționeze la extremitățile segmentului $x_n(m)$. Considerînd noile limite de sumare precum și faptul că $x_n(m)$ este considerat nul în afara intervalului $[0, N-1]$, relația (1.38) devine :

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i+k)} x_n(m) \cdot x_n(m+i+k) \quad \text{cu } 1 \leq i \leq p \text{ și } 0 \leq k \leq p \quad (1.44)$$

și pune în evidență funcția de autocorelație pe timp scurt, adică pe intervalul $i-k$:

$$\Phi_N(i,k) = R_N(i-k) \quad \text{cu } \begin{matrix} 1 \leq i \leq p \\ 0 \leq k \leq p \end{matrix} \quad (1.45)$$

unde

$$R_N(k) = \sum_{m=0}^{N-1-k} x_N(m)x_N(m+k) \quad (1.46)$$

Relațiile (1.45) și (1.46) se utilizează pentru calculul coeficienților α_k din relația (1.38), care se mai poate scrie sub formă matricială :

$$\sum_{k=1}^p R_N(i-k) \cdot \alpha_k = R_N(i) \quad \text{cu } 1 \leq i \leq p \quad (1.47)$$

În care membrul stâng reprezintă o matrice cu proprietăți speciale, fiind simetrică și cu elementele de pe diagonale egale (matrice Toeplitz) :

$$\begin{bmatrix} R(0) & R(1) & R(2) & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & R(p-2) \\ R(2) & R(1) & R(0) & \dots & R(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & R(p-3) & \dots & R(0) \end{bmatrix} \cdot \begin{bmatrix} \alpha(1) \\ \alpha(2) \\ \alpha(3) \\ \vdots \\ \alpha(p) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ \vdots \\ R(p) \end{bmatrix} \quad (1.48)$$

Aceste proprietăți au permis elaborarea unor algoritmi eficienți de rezolvare a sistemului (1.39). Cel mai utilizat și eficient algoritm este algoritmul Durbin, care se bazează pe existența următoarelor relații de recurență :

$$e^{(0)} = R(0) \quad (1.49)$$

$$K_i = \left[R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right] / e^{(i-1)} \quad \text{cu } 1 \leq i \leq p \quad (1.50)$$

$$\alpha_i^{(i)} = K_i \quad (1.51)$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - K_i \alpha_{i-j}^{(i-1)} \quad \text{cu } 1 \leq j \leq i-1 \quad (1.52)$$

$$e^{(i)} = (1 - K_i^2) \cdot e^{(i-1)} \quad (1.53)$$

În care $e^{(i)}$ este eroarea de predicție pentru un predictor de ordin i , cu $i=1 \dots p$.

Rezolvând succesiv ecuațiile (1.50)...(1.53) se obține în final soluția :

$$\alpha_j = \alpha_j^{(p)} \quad \text{cu } 1 \leq j \leq p \quad (1.54)$$

Această metodă de rezolvare permite atât determinarea succesivă a coeficienților de predicție, pentru ordine mai mici decât p , cât și erorile de predicție de diferite ordine $e^{(i)}$.

Acest avantaj este important, deoarece atunci când nu se cunoaște dinainte ordinul modelului de predicție liniară, necesar pentru o bună aproximare, se poate continua algoritmul, cu valori crescătoare pentru i , pînă ce eroarea de predicție $e^{(i)}$ scade sub o anumită valoare. Se mai poate arăta că utilizarea metodei autocorelației, pentru determinarea coeficienților de predicție, are avantajul că asigură și condiția de stabilitate a filtrului ce reprezintă modelul liniar predictiv, dacă este verificată relația :

$$|K_i| < 1 \quad \text{pentru } 1 \leq i \leq p \quad (1.55)$$

și care impune ca rădăcinile polinomului ce reprezintă funcția de transfer $A(z)$ a acestui model să se găsească, în planul z , în interiorul cercului unitate /17/, /18/.

1.2.5.4. Determinarea coeficienților de predicție prin metoda rețelelor cu celule în x (lattice)

Considerînd, conform formulei (1.26) și figurii (1.15), funcția de transfer a unui filtru al erorii de predicție de ordinul (i) ca fiind :

$$A^{(i)}(z) = 1 - \sum_{k=1}^i \alpha_k^{(i)} \cdot z^{-k} \quad (1.56)$$

dacă acestui filtru i se aplică semnalul $x_n(n)$ la intrare, la ieșire se va obține eroarea de predicție de ordinul (i) :

$$e_n^{(i)}(m) = e^{(i)}(nm) \quad (1.57)$$

în care, conform relației (1.24) :

$$e^{(i)}(m) = x(m) - \sum_{k=1}^i \alpha_k^{(i)} \cdot x(m-k) \quad (1.58)$$

și ținând cont de relația (1.55) relația precedentă se mai poate scrie cu ajutorul transformatei z :

$$E^{(i)}(z) = A^{(i)}(z) \cdot X(z) \quad (1.59)$$

Ținând cont de modul de definire a funcției $A^{(i)}(z)$ conform relației (1.56), precum și relației de recurență (1.52), ce exprimă pe $\alpha_k^{(i)}$ în funcție de $\alpha_k^{(i-1)}$, relația (1.56) devine :

$$A^{(i)}(z) = A^{(i-1)}(z) - k_1 z^{-1} A^{(i-1)}(z^{-1}) \quad (1.60)$$

care introdusă în (1.59) conduce la relația :

$$E^{(i)}(z) = A^{(i-1)}(z) \cdot X(z) - k_1 z^{-1} A^{(i-1)}(z^{-1}) \cdot X(z) \quad (1.61)$$

În care primul termen din partea dreaptă reprezintă chiar transformata z a erorii de predicție de ordinul $i-1$, adică :

$$E^{(i)}(z) = E^{(i-1)}(z) - k_1 z^{-1} A^{(i-1)}(z^{-1}) \cdot X(z) \quad (1.62)$$

Pentru interpretarea celui de-al doilea termen din partea dreaptă a relației precedente, se face notația :

$$B^{(i)}(z) = z^{-1} \cdot A^{(i)}(z^{-1}) \cdot X(z) \quad (1.63)$$

Similar vom avea pentru ordinul $(i-1)$:

$$B^{(i-1)}(z) = z^{-(i-1)} \cdot A^{(i-1)}(z^{-1}) \cdot X(z) \quad (1.64)$$

Aplicând transformata z inversă relației (1.63) se obține:

$$b^{(i)}(m) = x_{n(m-i)} - \sum_{k=1}^i \alpha_k^{(i)} \cdot x_{n[m-(i-k)]} \quad (1.65)$$

Analizând relațiile (1.58) și (1.65) se poate remarca faptul că $e^{(i)}(m)$ din relația (1.58) reprezintă o eroare de predicție de tip înainte (progresivă) a eșantionului $x_n(m)$, exprimată în funcție de precedentele i eșantioane $\{x_n(m-k), k=1, 2, \dots, i\}$, iar $b^{(i)}(m)$ poate fi considerată că reprezintă o eroare de predicție de tip înapoi (regresivă) a eșantionului $x_n(m-i)$, în funcție de următoarele i eșantioane care vor urma $\{x_n(m-i+k), k=1, 2, \dots, i\}$.

Aplicând transformata z inversă relației (1.62) se obține :

$$e^{(i)}(m) = e^{(i-1)}(m) - k_1 b^{(i-1)}(m-1) \quad (1.66)$$

Înlocuind pe $A^{(i)}$ în relația (1.63) conform relației (1.66) se obține relația :

$$B^{(i)}(z) = z^{-i} \cdot A^{(i-1)}(z^{-1}) \cdot X(z) - k_1 \cdot A^{(i-1)}(z) \cdot X(z) \quad (1.67)$$

și ținând cont de relația (1.64) se obține :

$$B^{(i)}(z) = z^{-1} \cdot B^{(i-1)}(z) - k_1 R^{(i-1)}(z) \quad (1.68)$$

Aplicând transformata z inversă relației precedente se obține :

$$b^{(i)}(m) = b^{(i-1)}(m-1) - k_1 e^{(i-1)}(m) \quad (1.69)$$

Relațiile (1.66) și (1.69) exprimă erorile de predicție progresive și regresive de ordinul (i) în funcție de aceleași erori de ordinul (i-1).

Dacă se consideră și secvența originală $x_n(m)$ ca o eroare de predicție de ordinul zero, astfel încât să se poată scrie :

$$e^{(0)}(m) = b^{(0)}(m) = x_n(m) \quad (1.70)$$

atunci, cu ajutorul relațiilor (1.69), (1.66) și (1.70) se poate reprezenta o rețea ca cea din figura 1.17, formată din p etaje identice conectate în cascadă, care reprezintă de fapt chiar implementarea numerică a filtrului eroare de predicție, ce are funcția de transfer $A(z)$.

Deoarece coeficienții de predicție liniară nu apar în mod explicit în structura rețelei din figura 1.17, înseamnă că minimizarea energiei erorii de predicție se poate efectua în funcție de parametrii $\{k_i, i=1, 2, \dots, p\}$, denumiți coeficienți de reflexie. Itakura /17/, /19/, /20/ a arătat că parametrii $\{k_i\}$ pot fi determinați pe baza erorilor de predicție progresive și regresive, fără a mai fi nevoie de calculul coeficienților de autocorelație, cu ajutorul relației /21/ :

$$k_i = \frac{\sum_{m=0}^{N-1} e^{(i-1)}(m) \cdot b^{(i-1)}(m-1)}{\sqrt{\sum_{m=0}^{N-1} [e^{(i-1)}(m)]^2 \sum_{m=0}^{N-1} [b^{(i-1)}(m-1)]^2}} \quad (1.71)$$

Datorită faptului că relația (1.71) este de forma unei funcții de corelație, coeficientul k_i apare ca un coeficient de

corelație între secvențele erorilor de predicție $\{e^{(i+1)}(n)\}$ și $\{e^{(i-1)}(n)\}$ de la intrarea etajului din structura rețelei din figura 1.17.

Din acest motiv, parametrii k_i au mai fost denumiți și coeficienți de corelație parțială, sau coeficienți PARCOR, și

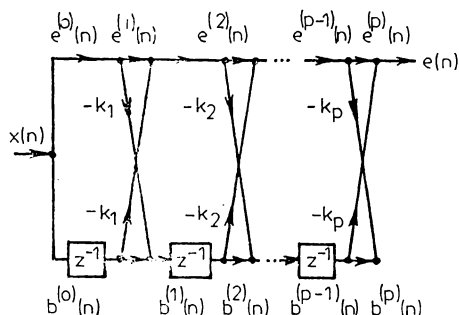


Fig.1.17. Rețeaua filtrului numeric al erorii de predicție (lattice)

s-au elaborat programe de calcul pe calculator, care permit determinarea rapidă a acestor coeficienți.

În literatură se indică faptul că metoda celulelor în X, ca și cea a autocorelației este stabilă și, în plus, în cazul metodei rețelelor, utilizarea funcției fereastră nu este necesară /21/.

1.2.5.5. Considerații cu privire la sinteza vorbirii prin predicție liniară

O problemă importantă, în cadrul sintezei prin predicție liniară, constă în determinarea numărului optim de coeficienți de predicție ce trebuie determinați pentru a se asigura o bună aproximare a semnalului vorbit. Alegerea acestui număr depinde, în esență, de mărimea frecvenței de eșantionare f_0 .

Măsurări experimentale au pus în evidență faptul că pentru modelarea corectă a comportării tractului vocal sînt necesari aproximativ f_0 poli, pentru o frecvență de eșantionare de f_0 kHz /21/. De acest număr mai trebuie adăugați încă aproximativ 3-4 poli pentru a reprezenta spectrul sursei de excitație și interfața tractului vocal cu exteriorul (radiația), ajungîndu-se astfel, pentru o frecvență de eșantionare de 8 kHz, la un număr de 11-12 poli, iar pentru frecvența de eșantionare de

10 kHz la 13-14 poli, adică tot atîția coeficienți de predicție.

Această determinare este exemplificată cu reprezentarea în figura 1.18.a erorii de predicție normalizate $/13/$ în funcție de numărul de coeficienți de predicție (p), putîndu-se remarca faptul că scăderea erorii de predicție, pentru $p > 14$, devine tot mai lentă, astfel încît creșterea în continuare a numărului de coeficienți de predicție nu mai este justificată de rezultatele care se obțin.

Din aceeași figură se mai poate remarca și faptul că eroarea de predicție este mai mare pentru sunetele sonore decît pentru cele nesonore, deoarece și energia acestor sunete este, în general, cu aproximativ un ordin de mărime mai mare ca cea a sunetelor nesonore.

De o mare importanță în realizarea unei bune sinteze prin predicție liniară este și alegerea segmentului de analiză (N). Deoarece numărul total de calcule depinde de N , este de dorit ca acesta să fie cît mai mic, dar, pe de altă parte, datorită periodicității sunetelor sonore, pentru a se reflecta acest aspect în sinteză, este necesar ca N să conțină cîteva perioade ale frecvenței fundamentale.

Utilizarea unei ferestre de ponderare impune, de asemenea, lungirea secvenței de analiză, astfel încît erorile de început și sfîrșit, mai mari, ce se obțin în acest caz, să poată fi neglijate.

Practic, pentru o frecvență de eșantionare cuprinsă între 8-10 kHz, se recomandă o durată de analiză corespunzătoare la 100-400 eșantioane. Acest număr mai poate fi redus în cazul în care se realizează o analiză sincronă cu frecvența fundamentală a vorbirii, în care caz se pot folosi segmente de analiză ce conțin numai două perioade ale fundamentalei.

Realizarea unui sintetizor de vorbire prin predicție liniară presupune simularea modelului reprezentat în figura 1.16 și comanda sa cu un set de vectori de forma :

$$V_n = V_n(e_1; e_k, G, a_1, a_2 \dots a_i) \quad (1.72)$$

în care e_1 - reprezintă parametrul ce comandă frecvența tonului fundamental pentru generatorul de armonici ;

e_k - reprezintă parametrul ce comandă comutatorul sonor/nesonor ;

G - factorul de câștig al semnalului de excitație;

a_i - parametri de predicție.

Implementarea sintezei prin predicție liniară, pe sisteme cu logică programată, impune condiții de viteză de calcul, în vederea realizării sintezei în timp real, motiv pentru care realizarea unui sintetizor eficient, prin predicție liniară, pentru limba română, rămâne încă o problemă deschisă, neputând fi rezolvată cu

microprocesoare uzuale pe 8, sau 16 biți, ci doar cu microprocesoare de tip bit-slice, sau cu procesarea de semnal.

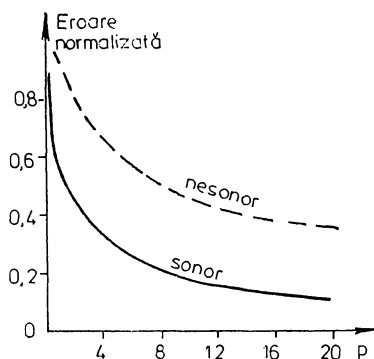


Fig.1.13. Variația erorii de predicție normalizate în funcție de numărul coeficienților de predicție

1.3. Metode de sinteză a vorbirii cu codificarea formei de semnal

În cadrul acestor metode, informația de bază este formată din segmente ale formei acustice a vorbirii, putându-se opera la unul din următoarele nivele :

- la nivel de cuvânt;
- la nivel de morfem;
- la nivel de diftong;
- la nivel de fonem (alofon).

Structura sintetizorului este asemănătoare celei a unui sistem de achiziție și prelucrare de date analogice prin metode numerice, asistat de un microcalculator, sau un sistem similar dedicat, realizat pe baze implementării unui microprocesor.

Schema bloc a sistemului este reprezentată în figura 1.19.

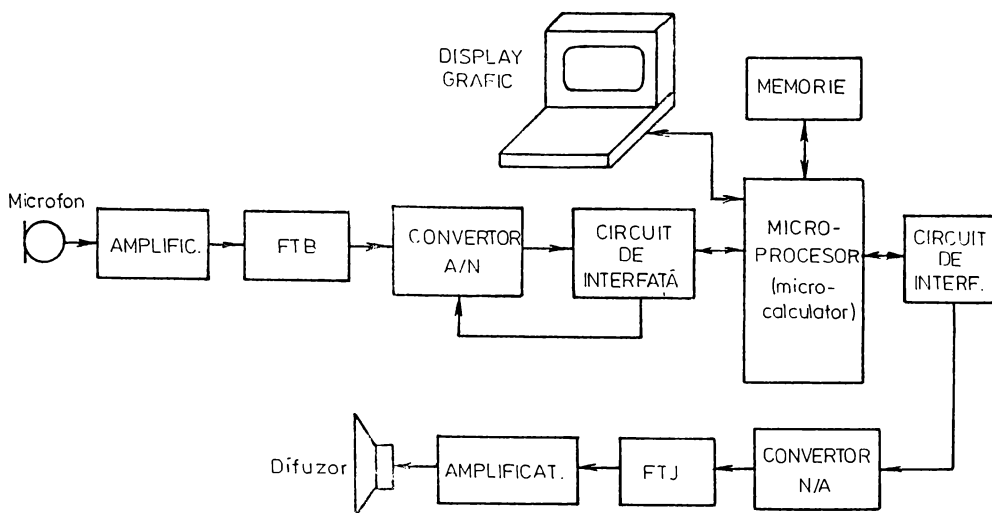


Fig.1.19. Schema bloc a unui sintetizor cu codificarea formei de semnal.

În cazul variantei în care segmentele ce constituie informația de bază sînt chiar cuvintele, acestea se înregistrează sub formă numerică în memoria calculatorului, la adrese precise, fără pauze, de unde se extrag în ordinea alcătuirii textului ce urmează a fi generat. Pentru a genera pauzele, se înregistrează o zonă de memorie cu zgomot de fond, care se ciclează apoi de un număr de ori pentru a obține o pauză de o anumită durată. Prin tehnici de programare, se realizează apoi un tabel de adrese corespunzătoare începutului fiecărui cuvînt și sfîrșitului său, în care sînt intercalate și adrese ale zonei de zgomot de fond, pentru a marca pauzele dintre cuvinte.

Conform acestui tabel, datele sînt extrase din memorie cu o viteză corespunzătoare realizării unei vorbiri cursive (viteză ce poate fi programată) și aplicate la intrarea convertorului numeric analogic, la a cărui ieșire, după filtrare, se obține mesajul ce trebuie generat.

Pauzele se introduc sub forma unor zone cu zgomot de fond, deoarece acest fapt corespunde și vorbirii reale. Absența oricărui semnal între cuvinte dă aspectul unei vorbiri nenaturale, în care începutul și sfîrșitul cuvîntului se percep foarte brusc, ceea ce deranjează la audiție.

Refacearea formei semnalului, plecând de la înregistrarea unor cuvinte, asigură calitatea cea mai bună a vorbirii sintetizate, permițând recunoașterea vocii celui ce a pronunțat cuvintele, deci asigurând o naturalețe și o calitate foarte bune. Este necesar însă, în acest caz, un volum de memorie pentru înregistrare, motiv pentru care sintetizorul respectiv dispune de un vocabular destul de limitat. Înregistrarea cuvintelor în memorie se face printr-una din tehnicile obișnuite de conversie analog-numerică (MIC, delta, sau variante ale acestora).

În cazul în care informația de bază este înregistrată la nivel de morfeme, acestea se aleg, de regulă, ca fiind silabele cuvintelor care trebuie să alcătuiască vocabularul sintetizorului.

Se procedează deci, mai întâi, la o analiză a semnalului de vorbire, care se realizează cu același sistem de achiziție și prelucrări de date (prezentat în figura 1.19), analiza și segmentarea efectuându-se cu ajutorul unui dispozitiv de tip display grafic.

Dacă informația de bază este selectată sub forma difonilor, se obține o sinteză prin difoneme, iar dacă se înregistrează sub formă de foneme, sau alofoni, se obține o sinteză fonemică. Acest sintetizor fonemic, deși poartă același nume cu cel specific codificării sursei de vorbire, se deosebește fundamental de acela, deoarece nu-și propune să modeleze tractul vocal, corespunzător producerii unui anumit fonem, ci intercalează doar funcția de timp corespunzătoare semnalului de vorbire sub formă acustică, pe care o are memorată.

Pentru a reda fidel și tranzițiile de la un fonem la altul, practic, același fonem se înregistrează în mai multe variante, adăugându-i-se și un rest din fonemul cu care se leagă în cadrul cuvântului respectiv,⁸ iar sintetizorul este cunoscut frecvent sub denumirea de sintetizor cu alofoni.

Trebuie făcută o distincție însă între acest tip de sintetizor și sintetizorul cu difoneme, deoarece primul înregistrează doar un rest din fonemul de legătură, pe când cel de-al doilea înregistrează ambele foneme, inclusiv tranziția dintre ele.

În cazul sintetizorului cu alofoni, s-a calculat că pentru fiecare limbă există un număr de câteva mii de astfel de combinații între foneme diferite, ce se pot cupla, și care ar trebui memorate pentru a permite o sinteză cu o bună inteligibilitate.

Pentru a se asigura și o calitate corespunzătoare este necesar să se mai prevadă și înregistrarea fonemelor în două variante : accentuate și neaccentuate, Astfel alofonul "o" din cuvântul "doi" este diferit de cel din cuvântul "două", ultimul fiind accentuat.

Cuplarea a doi alofoni, în cadrul sintezei, trebuie să se realizeze printr-o tehnică de interpolare, care, în cel mai simplu caz, constă în a delimita fiecare astfel de fonem cu un început și un sfârșit de nivel aproximativ constant. Această măsură de precauție înlătură tranzițiile bruște dintre două foneme de nivele mult diferite, diferența de nivel fiind percepută ca o pocnitură, ce deranjează auditiia.

Necesarul de memorie pentru memorarea tuturor alofonilor necesari sintezei unui vocabular nelimitat este destul de mare, dar se constată că, pentru un vocabular uzual, acesta este în jur de 2.000, putând fi astfel implementat și pe un microcalculator cu o memorie disponibilă de ordinul megaoctetilor, sau chiar sutelor de kiloocteti. Pentru ca această memorie să fie cât mai ușor de realizat, se combină avantajele memoriilor RAM dinamice, de a realiza capacități mari de memorie, cu capsule puține, cu memoriile externe, de tip disc flexibil, pentru care au fost elaborate programe de utilizare foarte eficiente și care permit schimbarea rapidă a suportului de memorare (dischete).

Se ajunge astfel la așa numitul "disc virtual" sau "RAM-disc" reprezentat de o memorie de câteva sute de kiloocteti, sau câțiva megaocteti, care se încarcă prin intermediul unei unități de disc flexibil, și care, în continuare, este exploatată prin intermediul programelor utilitare aferente acestui periferic, dar în condițiile unei viteze de acces și a unei fiabilități superioare (figura 1.20) /22/.

Un sintetizor pe bază de segmente de vorbire înregistrate și realizat cu logică programată va avea astfel schema bloc generală conform celei reprezentate în figura 1.20, în care a fost inclusă și partea de achiziție a segmentelor de vorbire.

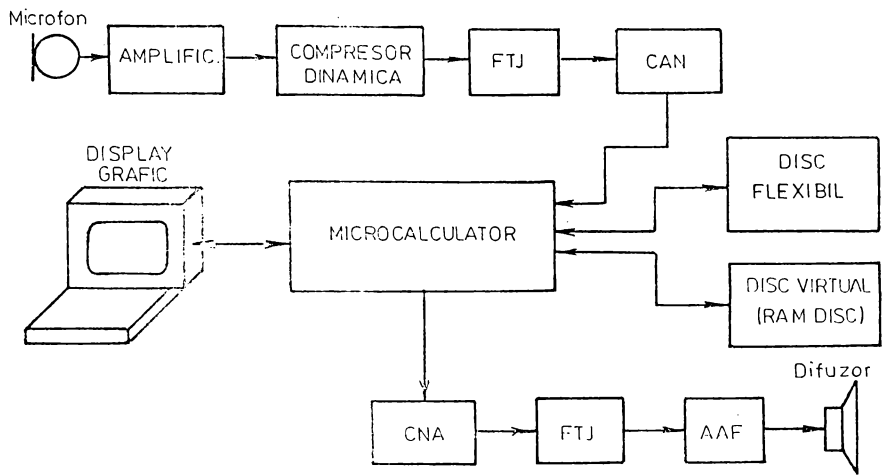


Fig.1.20. Utilizarea discului virtual în configurația unui microcalculator

2. SISTEME CU LOGICA PROGRAMATA PENTRU ANALIZA VORBIRII

2.1. Obiective impuse de sinteză analizei vorbirii

Sinteza vorbirii implică realizarea unei operații prealabile de analiză, în scopul determinării parametrilor pe baza cărora se efectuează sinteza.

Funcția de metoda de sinteză abordată, va depinde și modul în care trebuie să se efectueze analiza.

În general însă, pot fi deosebite trei domenii principale de efectuare a analizei vorbirii, valabile pentru oricare din metodele de sinteză utilizate și anume :

- analiza în scopul selectării unor parametri ce caracterizează vorbitorul ;
- analiza în scopul selectării unor parametri ce definesc semnificația semantică a vorbirii ;
- analiza în scopul realizării unei compresii a vorbirii.

În primul caz, parametrul esențial este frecvența fundamentală a vorbitorului, funcție de care depinde naturațea vorbirii sintetizate. Se poate astfel genera, prin sinteză, o vorbire cu o voce aparent naturală, ce sugerează prezența unei persoane, sau cu o voce impersonală (mecanică), ce sugerează prezența unui robot, sau, în general, a unui autorat care vorbește. În primul caz, se poate genera un ton fundamental prelevat de la un anumit vorbitor uman, sau unul generat artificial, dar asemănător unei anumite voci, astfel încât ascultătorul să aibă senzația că identifică sexul, vârsta, sau chiar persoana care vorbește.

Funcție de metoda de sinteză utilizată, acest parametru poate fi individual (cazul metodelor cu codificarea sursei), sau inclus în restul informației (cazul metodelor cu codificarea formei).

Analiza vorbirii efectuată în scopul identificării unor parametri caracteristici din punct de vedere semantic implică :

- stabilirea unor parametri care să permită identificarea tipului fundamental de fonem (sonor, nesonor) ;

- stabilirea unor parametri care să permită identificarea fonemului propriuzis (formanți, coeficienți de predicție, etc).

Analiza în scopul realizării unei compresii este specifică sintetizoarelor bazate pe logică programată și urmărește reducerea volumului de memorie necesar stocării informației de bază.

Indiferent care din obiectivele anterioare este urmărit, analiza semnalului vorbit se realizează în domeniile : amplitudine, timp și frecvență, ceea ce impune analizoarelor de vorbire capacitatea de a efectua o analiză temporală și una spectrală.

Având în vedere posibilitățile de a rezolva aceste probleme prin metode numerice, soluția optimă pentru realizarea analizei vorbirii constă în implementarea logicii programate.

Pentru aceasta, se pot realiza sisteme dedicate analizei vorbirii, sau se pot adapta în acest scop calculatoare de uz general.

Din punct de vedere soft, pe lângă softul curent, specific sistemelor de calcul (soft de operare, de programare, etc), sistemele de analiză a vorbirii se pot considera un caz particular al sistemelor de analiză prin metode digitale a semnalelor analogice, cărora le este necesar și un soft dedicat, destinat prelucrărilor propriuzise, în care trebuie să existe rutine specializate pentru realizarea următoarelor funcții principale :

- comanda și controlul sistemelor de achiziție și restituire a semnalelor analogice (eșantionare, menținere, conversie A/N și N/A) ;

- afișarea pe un terminal grafic a semnalelor achiziționate, sau restituite ;

- posibilitatea de a izola un anumit segment de semnal ;

- posibilitatea de a indica pe terminalul grafic un anumit moment al semnalului, prin intermediul unui cursor, cu specificarea valorilor ce caracterizează dimensional acel moment ;

- efectuarea transformărilor Fourier, directă și inversă ;

- posibilitatea de a indica, printr-un cursor, o anumită linie spectrală din spectrul semnalului analizat, cu specificarea amplitudinii și frecvenței ;

- efectuarea unei analize de corelație și autocorelație ;

- implementarea unor structuri de filtre digitale ;

- posibilitatea de a efectua segmentări ale semnalului analizat (în domeniile timp sau frecvență) și de a le combina în

diferite moduri (concatenare).

Această gamă largă de cerințe impune un grad de dificultate destul de ridicat în realizarea unui analizor digital de semnale analogice performant. Astfel de sisteme sînt produse, în serie, doar în țările cu tehnologie avansată, iar pretul lor este destul de ridicat /23/.

Mai trebuie remarcat și faptul că analiza vorbirii, efectuată în scopul determinării informației de bază pentru sinteză, este mai simplă de realizat, comparativ cu analiza aceleiași vorbiri efectuată în scopul recunoașterii automate a vorbirii, sau a vorbitorului, care introduc dificultăți suplimentare, impuse atît de viteza de lucru, care trebuie să fie în timp real, cît și de faptul că analiza și interpretarea rezultatelor trebuie să se facă automat, fără intervenția operatorului uman.

Astfel, de exemplu, tehnici speciale de determinare automată a frecvenței fundamentale, a tipului de fonem, a formanților, sau compararea unor tipare fonetice, cum sînt numărarea trecerilor prin zero, determinarea spectrului, sau programarea dinamică /13/25/46/, nu sînt necesare.

Deoarece etapele de analiză și sinteză a vorbirii sînt distincte, analiza efectuîndu-se doar la început, pentru definitivarea informației de bază, operatorul uman poate determina cu suficientă precizie parametrii importanți, ce definesc vorbirea, pe baza formelor de undă corespunzătoare în domeniile timp și frecvență. În cazul sintezei prin predicție liniară, se pot determina, de asemenea, prin calcul, coeficienții de predicție corespunzatori, utilizînd algoritmi de calcul ce implementează una din metodele elaborate pentru calculul acestor coeficienți /2/. Alte metode de analiză au în vedere doar o codificare optimă a vorbirii discretizate, funcție de anumite criterii urmărite /52/53/.

2.2. Metode numerice de prelucrare a semnalelor analogice eșantionate, utilizate frecvent în analiza vorbirii

Eșantionarea și codarea semnalelor analogice deschid calea prelucrării lor prin metode numerice, cu ajutorul sistemelor bazate pe logică programată.

În cadrul acestor metode, din punct de vedere al analizei vorbirii, efectuată în vederea sintezei, interesează mai ales transformata Fourier discretă (directă și inversă) și analiza prin predicție liniară.

Transformata Fourier discretă directă permite obținerea spectrului de frecvență corespunzător sunetelor vorbirii, din care se pot determina :

- frecvența fundamentală și amplitudinea ei ;
- tipul fonemului (sonor, nesonor) ;
- frecvențele de apariție a formațiilor și amplitudinile lor ;
- spectral semnificativ al unui anumit sunet de vorbire ;
- modificările de spectru survenite în cazul sunetelor accentuate, față de cele neaccentuate , etc.

Transformata Fourier discretă inversă permite refacerea formei de variație în timp a unui sunet, plecând de la spectrul său de frecvență și este frecvent utilizată în determinarea zonelor spectrale semnificative ale aceluși sunet, din punct de vedere a inteligibilității. Se pot astfel elimina anumite zone ale spectrului și se poate reveni la forma în timp a semnalului, ce poate fi ascultată prin intermediul unui traductor electroacustic, pentru a se aprecia în ce măsură au fost afectate inteligibilitatea, naturalitatea, etc. Se pot astfel depista zone spectrale redundante din punct de vedere a inteligibilității, ce pot constitui o cale de realizare a compresiei informației de bază pentru sinteză, cu avantaje în reducerea volumului de memorie necesar stocării acestei informații în sintetizoarele bazate pe logică programată.

Analiza prin predicție liniară este necesară determinării coeficienților de predicție liniară pentru sunetele vorbirii, coeficienți ce vor constitui informația de bază în vederea sintezei prin această metodă, conform modelului predicției liniare, prezentat în capitolul precedent.

2.2.1. Transformata Fourier rapidă /13/, /18/

2.2.1.1. Clasificare a tipurilor de algoritmi TFR

Plecând de la transformata Fourier normală, algoritmi de TFR s-au dezvoltat, în general, pe trei direcții :

- algoritmi în care secvența de N eșantioane este împărțită succesiv în subsecvențe din ce în ce mai mici (algoritmi cu decimare) ;
- algoritmi pentru un număr N impus ;

- algoritmi cu decompunerea matricii T_N într-un produs de mai multe matrici (algoritmi Winograd).

Primii (printre care și algoritmul Cooley-Tukey) se utilizează în cazul în care numărul de eșantioane (N), cărora li se aplică transformata, verifică relația:

$$N = 2^M \quad (2.1)$$

Algoritmi pentru N impus se utilizează în cazul în care N poate fi reprezentat ca un produs de factori :

$$N = p_1 \cdot p_2 \cdots p_n \quad (2.2)$$

Notînd

$$q_1 = p_2 \cdots p_n \quad (2.3)$$

rezultă

$$N = p_1 \cdot q_1 \quad (2.4)$$

ceea ce conduce la ideea de a împărți secvența de intrare în p_1 secvențe de câte q_1 eșantioane, asociind fiecărui al p_1 -lea eșantion cu o secvență dată. Relația

$$\text{TFD}\{x_n\} = X_k = \sum_{n=0}^{N-1} x(n) \cdot W_N^{nk} \quad \text{cu } k=0,1,2,\dots,N-1, \quad (2.5)$$

corespunzătoare TFR normale /18/ devine în acest caz :

$$\begin{aligned} X_k = \sum_{n=0}^{N-1} x(n) \cdot W_N^{nk} &= \sum_{r=0}^{q_1-1} x(p_1 \cdot r) \cdot W_N^{p_1 \cdot r \cdot k} + \sum_{r=0}^{q_1-1} x(p_1 \cdot r + 1) W_N^k \cdot W_N^{p_1 \cdot r \cdot k} + \\ &+ \dots + \sum_{r=0}^{q_1-1} x(p_1 \cdot r + p_1 - 1) \cdot W_N^{(p_1-1)k} \cdot W_N^{p_1 \cdot r \cdot k} \end{aligned} \quad (2.6)$$

sau :

$$X_k = \sum_{\ell=0}^{p_1-1} W_N^{\ell \cdot k} \sum_{r=0}^{q_1-1} x(p_1 \cdot r + \ell) \cdot W_N^{p_1 \cdot r \cdot k} \quad (2.7)$$

în care

$$G_k = \sum_{r=0}^{q_1-1} x(p_1 \cdot r + \ell) \cdot W_N^{p_1 \cdot r \cdot k} = \sum_{r=0}^{q_1-1} x(p_1 \cdot r + \ell) \cdot W_{q_1}^{rk} \quad (2.8)$$

deoarece :

$$\text{pentru } N = p_1 q_1 \text{ rezultă } W_N^{p_1 \cdot r \cdot k} = W_{q_1}^{rk} \quad (2.9)$$

și se poate interpreta că X_k , dat de relația (2.7), este expri-

mat în funcție de p_1 TFD a unor secvențe de lungime q_1 eganticoane.

Se poate arăta că în acest caz sînt necesare $N(p_1-1)+p_1 \cdot q_1^2$ înmulțiri și sumări complexe. În algoritmi de tip Winograd, se caută pentru matricea T din ecuația

$$[X_k] = T[x_n] \quad (2.10)$$

ce caracterizează TFR normală, unde

$$T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & W^1 & \dots & W^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & W^{N-1} & \dots & W^{(N-1)^2} \end{bmatrix} \quad (2.11)$$

o descompunere de formă :

$$T_N = S_N \cdot C_N \cdot V_N, \quad (2.12)$$

în care V_N este o matrice incidentă, de dimensiuni $J \times N$ (cu elemente de 0, 1 și -1), C_N este o matrice diagonală, de dimensiuni $J \times J$, iar S_N este tot o matrice incidentă, dar de dimensiuni $N \times J$.

Se poate demonstra existența descompunerii din relația (2.12) pentru valori suficient de mari ale lui J (exemplu : $J=N^2$). Winograd a elaborat și metoda de combinare a acestui algoritm în structuri înălțate, pentru valori mai mari ale numărului de eşantioane considerate (N).

În tabelul din figura 2.1 sînt reprezentate cifre comparative privind numărul de operații necesare pentru TFD și TFR de tip Winograd.

Implementarea acestei proceduri implică însă atît reordonarea datelor de intrare, înainte de prelucrare, cît și a celor de ieșire, după prelucrare.

N	TFD		N	Winograd	
	(x)	(+)		(x)	(+)
32	192	416	30	72	384
48	352	784	48	108	636
256	2304	5248	240	648	5016
512	6144	12288	504	1872	14796
1024	12288	26624	1008	4212	35244
2520	44032	92072	2520	11232	102348

Fig.2.1. Tabel comparativ cu numărul de operații necesare pentru realizarea TFD și a TFR tip Winograd

2.2.1.2. Algoritmi cu decimare pentru TFR

Sub această denumire sînt cunoscute variantele clasice și totodată mai frecvent utilizate ale algoritmilor pentru TFR. Se poate descompune în secvențe, în acest caz, atât secvența în timp a semnalului $\{x_n\}$, cît și cea în frecvență $\{X_k\}$, funcție de care algoritmi respectivi se vor numi cu "decimare în timp", sau cu "decimare în frecvență".

2.2.1.2.1. Algoritmi TFR cu decimare în timp

Algoritmi cu decimare în timp presupun divizarea în sub-secvențe mai scurte a secvenței $\{x_n\}$, iar secvența $\{X_k\}$ va rezulta din combinarea transformărilor corespunzătoare. Acești algoritmi se utilizează, de regulă, atunci cînd numărul eșantioane-lor (N) luate în considerare se poate scrie ca o putere a lui 2:

$$N = 2^M \quad (2.13)$$

În acest caz secvența $\{x_n\}$ se poate împărți în două sub-secvențe :

- pară $\{x_{2n}\}$ și
- impară $\{x_{2n+1}\}$,

separînd eșantioanele cu indice par de cele cu indice impar.

Avem :

$$\begin{cases} g_n = x_{2n} \\ h_n = x_{2n+1} \end{cases} \quad \text{cu } n \in [0, \frac{N}{2} - 1] \cap \mathbb{Z} \quad (2.14)$$

La fel se poate separa în două componente și X_k :

$$X_k = \sum_{n=0}^{N-1} x_n w_N^{nk} = \sum_{n=0}^{\frac{N}{2}-1} [x_{2n} w_N^{2nk} + x_{2n+1} w_N^{(2n+1)k}] \quad (2.15)$$

Dar datorită modului de definire a funcției $w_N / 18/$

$$w_N = e^{-j \cdot \frac{2\pi}{N}} \quad (2.16)$$

avem :

$$w_N^{2nk} = (e^{-j \frac{2\pi}{N}})^{2nk} = (e^{-j \frac{2\pi}{N}})^{nk} = (e^{-j \frac{2\pi}{N/2}})^{nk} = w_{N/2}^{nk} \quad (2.17)$$

Tinînd cont de proprietatea precedentă, X_k devine :

$$X_k = \sum_{n=0}^{N/2-1} x_{2n} w_{N/2}^{nk} + \sum_{n=0}^{N/2-1} x_{2n+1} w_N^k \cdot w_{N/2}^{2nk} \quad (2.18)$$

sau :

$$X_k = \sum_{n=0}^{N/2-1} x_{2n} w_N^{nk} + w_N^k \sum_{n=0}^{N/2-1} x_{2n+1} w_N^{nk} \text{ cu } k \in [0, N-1] \cap \mathbb{Z} \quad (2.19)$$

Introducînd notațiile g_n și h_n în relația precedentă obținem :

$$G_k = \sum_{n=0}^{N/2-1} g_n w_N^{nk} \quad (2.20)$$

și

$$H_k = \sum_{n=0}^{N/2-1} h_n w_N^{nk} \text{ cu } k \in [0, \frac{N}{2} - 1] \cap \mathbb{Z} \quad (2.21)$$

Secvențele $\{G_k\}$ și $\{H_k\}$ vor avea perioada $\frac{N}{2}$ și conform proprietăților de periodicitate vom avea :

$$G_k = G_{k+N/2} \quad (2.22)$$

și

$$H_k = H_{k+N/2} \quad (2.23)$$

deoarece

$$w_N^{(k+N/2)} = -w_N^k \quad (2.24)$$

Rezultă că X_k se obține din combinațiile date de relația :

$$X_k = G_k + w_N^k \cdot H_k \quad (2.25)$$

și

$$\begin{cases} X_{k+N/2} = G_{k+N/2} \cdot H_k = G_k - w_N^{-k} \cdot H_k \\ \text{cu } k \in [0; \frac{N}{2} - 1] \cap \mathbb{Z} \end{cases} \quad (2.26)$$

Cu relațiile (2.25), (2.26) se determină N ecuații pentru X_k combinînd $\frac{N}{2}$ ecuații ale funcției G_k și respectiv H_k , a căror calcul necesită $(\frac{N}{2})^2$ înmulțiri în complex.

Numărul total de înmulțiri în complex pentru calculul funcției X_k în cele N puncte va fi

$$2 \cdot (\frac{N}{2})^2 + N = N^2 (\frac{1}{2} + \frac{1}{N}) \simeq \frac{N^2}{2} \text{ (dacă } N \text{ crește)} \quad (2.27)$$

Relațiile (2.25) și (2.26) reprezintă principiul de realizare al algoritmului de TFR prin decimare în timp, constituind un modul de combinații elementare care permite reducerea numărului de înmulțiri, necesare unei etape de calcul, la aproximativ jumătate,

față de această etapă corespunzătoare algoritmului TFD. Un astfel de modul constituie o etapă de divizare, numită și "fluture", după forma sa, și este reprezentată în figura 2.2.

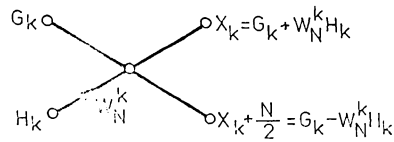


Fig.2.2. Reprezentarea grafică a unei etape de divizare (fluture) în cadrul algoritmului TFD cu decimare în timp

Cercul de la intersecția celor două ramuri semnifică realizarea unei subsecvențe TFD cu $N=2$ iar factorul w_N^k se numește factor de rotație.

În continuare, diviziunea cu 2 a secvențelor în timp este repetată, astfel încât G_k și H_k se obțin, la rândul lor, din subsecvențe de lungime $\frac{N}{4}$, prin relații analoge celor ce definesc etapa de divizare.

În cazul în care are loc relația (2.1), decimarea în timp poate fi repetată de m ori, iar ultima divizare va corespunde unui factor de rotație:

$$w_N^k \Big|_{k=\frac{N}{2}} = w_N^{N/2} = e^{-j\pi} = -1 \quad (2.28)$$

ceea ce înseamnă că înmulțirea corespunzătoare cu acest factor va fi de fapt o scădere. În total, deci, pentru determinarea secvenței $\{X_k\}$ vor fi necesare:

$$\frac{N}{2} \times m = \frac{N}{2} \log_2 N \text{ înmulțiri} \quad (2.29)$$

și

$$N \times m = N \log_2 N \text{ adunări} \quad (2.30)$$

iar reducerea volumului de calcul devine cu atât mai spectaculoasă, cu cât N este mai mare. De exemplu, pentru:

$$N = 2^{10} = 1024 \quad (2.31)$$

rezultă o reducere de aproximativ 100 ori.

În figura 2.3 este reprezentată, sub forma unei scheme bloc, desfășurarea succesivă a etapelor de calcul pentru cazul particular $N = 2^3$

$$(2.32)$$

iar în figura 2.4 graful corespunzător.

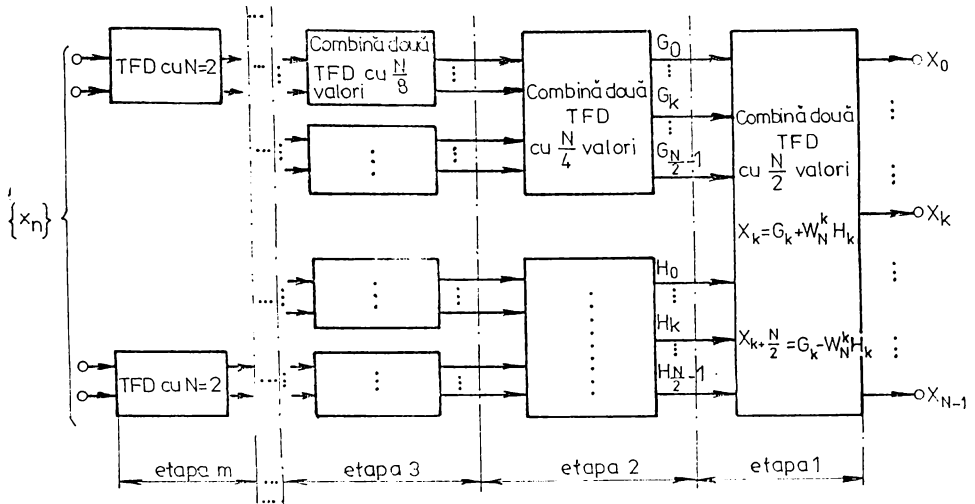


Fig.2.3. Împărțirea succesivă a etapelor de calcul a TFD prin împănțărirea secvențelor de valori

Analizând figura 2.4 se constată că, dacă secvența $\{X_k\}$ apare în ordinea naturală, secvența de intrare $\{x_n\}$ nu apare în aceeași ordine.

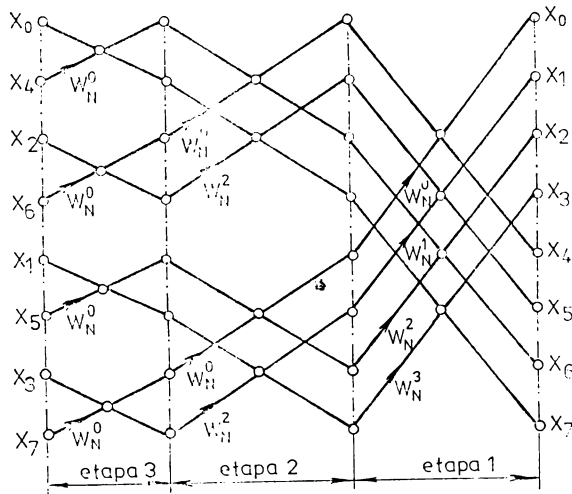


Fig.2.4. Graful TFD cu decimare în timp pentru $N=8$

Ordinea de intrare corespunde unei așa numite legi de "inversiune binară", în care ordonarea se face cu inversarea biților, astfel că poziția de intrare a eșantionului n va fi cea corespunzătoare citirii inverse a codului binar corespunzător acestei cifre, conform tabelului din figura 2.5.

Ordine de apariție	Cod binar	Cod cu inversarea biților	Ordine de intrare (n)
0	000	000	0
1	001	100	4
2	010	010	2
3	011	110	6
4	100	001	1
5	101	101	5
6	110	011	3
7	111	111	7

Fig.2.5. Tabel cu pozițiile de intrare în algoritmul TFR cu decimare în timp, corespunzătoare legii de inversare a biților pentru cazul $N=8$

Acest aspect introduce o serie de complicații în implementarea algoritmului pe calculator, necesitând mai întâi memorarea întregii secvențe de intrare, într-o zonă de memorie, deci un consum de memorie suplimentar.

Trebuie remarcat însă că s-a reușit și elaborarea unor algoritmi TFR cu decimare în timp, în care ordinea de intrare a secvenței $\{x_n\}$ este cea naturală de apariție.

2.2.1.2.2 Algoritmi de TFR cu decimare în frecvență

Acești algoritmi sînt echivalenți din punct de vedere al volumului de calcul cu cei precedenți, diferența constînd în faptul că secvența care se împarte în două este $\{x_n\}$:

$$\begin{cases} x_n^1 = x_n \\ x_n^2 = x_{n+N/2} \end{cases} \quad \text{cu } n \in [0; \frac{N}{2} - 1] \cap \mathbb{Z} \quad (2.33)$$

Se va proceda în acest caz la o decimare în frecvență, adică în sens invers cazului precedent, conform relațiilor :

$$X_k = \sum_{n=0}^{N/2-1} x_n w_N^{nk} + \sum_{n=N/2}^{N-1} x_n w_N^{nk} = \sum_{n=0}^{N/2-1} x_n^1 w_{N/2}^{nk} + \sum_{n=0}^{N/2-1} x_n^2 w_{N/2}^{(n+N/2)k}$$

cu $k \in [0; N-1] \cap \mathbb{Z}$ (2.34)

În relația precedentă se separă secvențele pare și impare :

$$X_{2r} = \sum_{n=0}^{N/2-1} [x_n^1 w_{N/2}^{nr} + x_n^2 w_{N/2}^{nr}]$$

(2.35)

și

$$X_{2r+1} = \sum_{n=0}^{N/2-1} [x_n^1 w_{N/2}^{nr} + x_n^2 w_{N/2}^{nr} \cdot w_N^{(rn+N/2)}] w_N^n$$

(2.36)

cu $r \in [0; \frac{N}{2} - 1] \cap \mathbb{Z}$

Avînd în vedere că :

$$w_N^{rN} = 1$$

(2.37)

și

$$w_N^{N/2} = -1$$

(2.38)

relațiile precedente se simplifică și devin :

$$X_{2r} = \sum_{n=0}^{N/2-1} (x_n^1 - x_n^2) \cdot w_{N/2}^{nr}$$

(2.39)

și respectiv :

$$X_{2r+1} = \sum_{n=0}^{N/2-1} (x_n^1 + x_n^2) \cdot w_N^n \cdot w_{N/2}^{nr}$$

(2.40)

cu $r \in [0; \frac{N}{2} - 1] \cap \mathbb{Z}$

Relațiile precedente sugerează împărțirea semnalului de intrare $\{x_n\}$ în :

$$\begin{cases} g_n = x_n^1 + x_n^2 & \text{și} \\ h_n = (x_n^1 - x_n^2) \cdot w_N^n \end{cases}$$

(2.42)

care reprezintă împreună etapa de divizare, sau "fluturile" cu decimare în frecvență.

În figura 2.6 este reprezentat graful corespunzător, tot în cazul exemplului $N=8$.

Se remarcă faptul că în acest caz intrările apar în ordine naturală, în timp ce ieșirile sînt ordonate conform inversării biților.

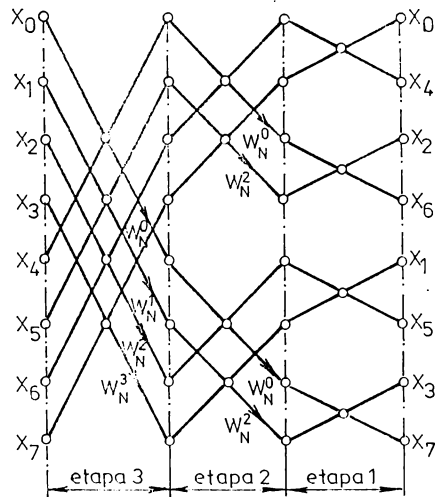


Fig.2.6. Graful corespunzător TFR cu decimare în frecvență în cazul $N=8$

2.2.2. Tehnica ferestruirii în analiza semnalului vocal /13/

Caracterizându-se prin faptul că atât semnalul original $\{x_n\}$, cât și spectrul său $\{X_k\}$, sînt secvențe periodice finite, iar perioadele lor conțin același număr de eșantioane, se poate arăta că TFD înlocuiește o secvență periodică în domeniul timp, cu perioade de eșantionare T_0 și perioada de repetiție $\frac{T_0}{2}$, cu o secvență periodică în domeniul frecvență, cu perioada $\frac{1}{T_0}$ și cu perioada de repetiție $\frac{1}{T_0}$.

TFD are un rol esențial în prelucrarea prin metode numerice a semnalelor analogice, cu ajutorul tehnicii de calcul, permițînd nu numai determinarea spectrului secvenței analizate, dar și a altor funcțiuni importante, specifice prelucrării numerice a semnalelor, ca transformata Z, produsul de convoluție și parametrării unor filtre digitale.

Se poate arăta că datorită limitării secvenței analizate la un număr finit (N) de eșantioane, TFD conduce la apariția

unui reziduu spectral.

Acest reziduu va fi cu atât mai important, cu cât secvența analizată nu este corelată cu perioada naturală a semnalului analizat și prin prelungirea periodică a segmentului supus analizei apar discontinuități la limitele secvenței analizate. Efectul acestui reziduu spectral poate deforma spectrul real al semnalului analizat până într-atât încît anumite linii spectrale, de amplitudine mai mică, sau chiar zone întregi din spectru, pot fi mascate prin "îngroparea" lor în acest reziduu.

Pe de altă parte, proprietățile temporale ale semnalului vorbit, care sînt caracterizate de modificări permanente, necesită ca analiza acestui semnal să se facă pe timp scurt, deoarece s-a observat experimental că pe intervale de timp de ordinul $10 - 30$ ms, proprietățile temporale ale semnalului vocal (energie, număr de treceri prin zero, corelație) pot fi considerate invariabile.

Cum corelarea exactă a domeniului de analiză cu perioada naturală a semnalului vorbit (corespunzătoare frecvenței fundamentale) nu este întotdeauna posibilă (mai ales pentru fonemele nesonore, la producerea cărora corzile vocale nu participă, deci nu apare fenomenul pseudoperiodic caracterizat prin frecvența fundamentală), analiza semnalului vorbit esanționat prin TFD va fi însoțită de prezența reziduuului spectral.

Se poate arăta că micșorarea acestui reziduu este posibilă în cazul în care pentru segmentarea sa temporală nu se utilizează un simplu interval dreptunghiular (fereastră dreptunghiulară, definită prin relația (2.43),

$$w(h) = \begin{cases} 1 & \text{pentru } n \in \left[-\frac{N}{2}; \frac{N}{2}\right] \cap \mathbb{Z} \\ 0 & \text{în rest} \end{cases} \quad (2.43)$$

ci ferestre speciale, caracterizate printr-o anumită funcție de timp în intervalul analizat.

Această tehnică de analiză pe timp scurt se numește tehnica ferestruirii și s-au definit diferite tipuri de funcții fereastră (Hanning, Hamming, Blackman, Gauss, etc) fiecare cu avantaje și dezavantaje din anumite puncte de vedere. Cunoscînd aproximativ forma spectrului semnalului analizat, se poate alege și fereastra temporală corespunzătoare pentru analiză, iar dacă forma spectrului nu e cunoscută, se recomandă efectuarea analizei cu ajutorul mai multor tipuri de ferestre, pentru a observa deo-

scbirile care apar și a determina astfel spectrul real.

În acest scop se poate defini și o TF dependentă de timp prin relația :

$$X_k(j\omega) = \sum_{n=-\infty}^{\infty} w(k-n) \cdot x(n) \cdot e^{-j\omega k} \quad (2.44)$$

în care $x(n)$ reprezintă semnalul vorbit egantionat, iar $w(k-n)$ este o fereastră care determină segmentul analizat la momentul fixat de indicele n . Schimbând în relația precedentă indicii de sumare prin relația :

$$(k-n) \rightarrow n \quad (2.45)$$

aceasta se mai poate pune sub forma :

$$X_k(j\omega) = e^{-j\omega k} \cdot X_k^1(j\omega) \quad (2.46)$$

în care

$$X_k^1(j\omega) = \sum_{n=-\infty}^{\infty} x(k-n) \cdot w(n) \cdot e^{j\omega n} \quad (2.47)$$

Analizînd relația (2.46), TF dependentă de timp poate fi considerată ca o reprezentare bidimensională a unui semnal unidimensional $x(n)$, cele două coordonate fiind : timpul (marcat de indicele n) și frecvența (marcată prin ω , în radiani).

Dacă K se consideră fixat, $X_k(j\omega)$ constituie TF a secvenței $w(k-n) \cdot x(n)$ cu $n \in (-\infty; \infty) \cap \mathbb{Z}$. În acest caz se poate obține prin transformare inversă :

$$w(k-n) \cdot x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X_k(j\omega) \cdot e^{j\omega n} d\omega \quad (2.48)$$

relație care pentru $k=n$ devine :

$$x(n) = \frac{1}{2\pi w(0)} \int_{-\pi}^{\pi} X_n(j\omega) \cdot e^{j\omega n} d\omega \quad (2.49)$$

cu condiția

$$w(0) \neq 0 \quad (2.50)$$

se poate astfel determina secvența $\{x_n\}$ dacă se cunoaște $\{X_n(j\omega)\}$ pentru toate valorile lui ω dintr-o perioadă.

Prin variația lui n în acest caz, fereastra poate fi deplasată de-a lungul formei de undă a semnalului, iar aceste considerații pot fi aplicate și în cazul TFD.

2.2.3. Spectrul de putere al TFD

TFD a unei secvențe temporale periodice $\{x(n)\}$, de perioadă N , este exprimată prin funcția de densitate spectrală corespunzătoare $\{X(k)\}$, care este însă o funcție complexă, cu parte reală și imaginară.

Din punct de vedere metrologic, exprimarea în unități de măsură a amplitudinii liniilor spectrale prezintă dificultăți în acest caz, deoarece spectrul rezultat este format de fapt din două spectre (unul real și altul imaginar), simetrice față de origine.

Este mai convenabilă, de aceea, măsurarea spectrului de putere $\{P(k)\}$ corespunzător semnalului $\{x(n)\}$, definit ca fiind spectrul cu aceleași număr de linii spectrale ca și $\{X(k)\}$, dar în care amplitudinea fiecărei linii spectrale este egală cu modulul numărului complex format din cele două componente (reală : $X_R(k)$ și imaginară $X_I(k)$) corespunzătoare acelei linii, conform relației :

$$P(k) = X_R(k) \cdot X_I(k) = |X(k)|^2 \quad (2.51)$$

Din punct de vedere dimensional, această mărime este doar proporțională cu puterea disipată și nu identică cu ea. Similar, pentru a permite exprimarea în decibeli, poate fi considerat și logaritmul spectrului de putere : $\lg P(k) = \lg |X(k)|^2$ (2.52)

2.3. Contribuții la realizarea sistemelor cu logică programată pentru analiza vorbirii

2.3.1. Prezentarea generală a sistemului de analiză realizat

În concepția generală care a stat la baza proiectării și realizării sistemului de analiză a vorbirii, ce face obiectul acestei prezentări, s-a avut în vedere obiectivul ca acest sistem să permită efectuarea unei analize corespunzătoare obținerii informației de bază pentru principalele metode de sinteză a vorbirii, care s-au impus pe plan mondial și anume : metoda predicției liniare și a formanților (din categoria metodelor cu codificarea sursei) și metoda concatenării unor segmente de vorbire preînregistrate (din categoria metodelor cu codificarea formei semnalului).

Totodată, s-au prezentat, atât la nivel hard, cât și soft, resursele necesare în vederea realizării unei sinteze de probă, de mici dimensiuni, pentru metodele de sinteză avute în vedere, în scopul verificării prin sinteză a corectitudinii analizei efectuate, precum și pentru a putea pune în evidență anumite aspecte particulare, nesesizabile decât în contextul efectuării sintezei.

Schema bloc a sistemului de analiză realizat este reprezentat în figura 2.7. Sistemul este alcătuit, în mare, prin interconectarea a două microcalculatoare, ambele bazate pe implementarea microprocesorului Z80, la care s-a adăugat și un sistem de achiziție și restituire pentru semnale analogice. Resursele hardware ale sistemului conțin :

- un microcalculator de uz general, de tip Junior ,
cu următoarea configurație :
 - unitate centrală ;
 - display (dispozitiv de afișare alfanumerică plus tastatură) ;
 - două unități de disc flexibil, lucrând în regim de lucru cu dublă densitate ;
 - imprimantă ;
 - extensie memorie (256)octeți) de tip RAM-DISC ;
 - înregistrator x-y ;
- un microcalculator modular, realizat cu module de tip MADS-80 (Microelectronica) , incluzând :
 - un modul unitate centrală (MADS80-UC) ;
 - un modul extensie memorie (MADS80-EM) ;
 - un modul programator EPROM (MADS80-PP) ;
- un sistem de afișare grafică, incluzând :
 - un display grafic (DAF2020) ;
 - o imprimantă grafică (CDC 93335) ;
- un sistem de achiziție și restituire pentru semnale analogice, incluzând :
 - un preamplificator de microfon ;
 - un filtru trece jos, cu frecvența de tăiere regalabilă ;
 - un amplificator compresor de dinamică ;
 - un subsistem hibrid de achiziție și conversie date analogice pe 12 biți (DAS 1123) , incluzând :

- un multiplexor cu 16 canale de intrare ;
- un circuit de egantionare - semorare ;
- un convertor A/N ;
- o sursă de referință de precizie ;
- o logică de control ;
- un subsistem de restituire și conversie date digitale, pe 12 biți, incluzînd :
 - un convertor N/A ;
 - un filtru trece jos ;
 - un amplificator audio.

Resursele software ale sistemului includ :

- sistemul de operare, de tip CP/M30, al microcalculatorului Junior ;
- sistemul de operare al microcalculatorului modular MADS-80 ;
- sistemul de operare al terminalului grafic DAF2020 ;
- softul special de interconectare a microcalculatorului Junior cu microcalculatorul modular MADS-80 ;
- softul special de interfațare a extensiei de memorie "RAM-DISC" cu calculatorul Junior și interpretarea acestei extensii ca o unitate suplimentară de disc flexibil, tratabilă prin setul de rutine specifice terminalelor cu disc flexibil ;
- softul special de interfațare a unui caracterograf ;
- softul special de gestionare a sistemului de achiziție și restituire a datelor analogice ;
- sistemul de programe utilizate, destinate prelucrării prin metode digitale a semnalelor analogice convertite digital și înregistrate în memoria unuia din cele două microcalculatoare ale sistemului.

Utilizarea microcalculatorului Junior a fost impusă de necesitatea de a dispune de un sistem de calcul care să includă terminale de tip disc flexibil, aflate sub comanda unui sistem de operare performant, cum este sistemul CP/M30, pentru a permite stît realizarea unui sistem de dezvoltare puternic, cît și cuplarea unei extensii de memorie de tip RAM-DISC (de capacitatea unei fețe de dischetă, dar realizată cu circuite integrate de memorie RAM dinamice) și utilizarea acestuia ca pe o unitate de disc flexibil.

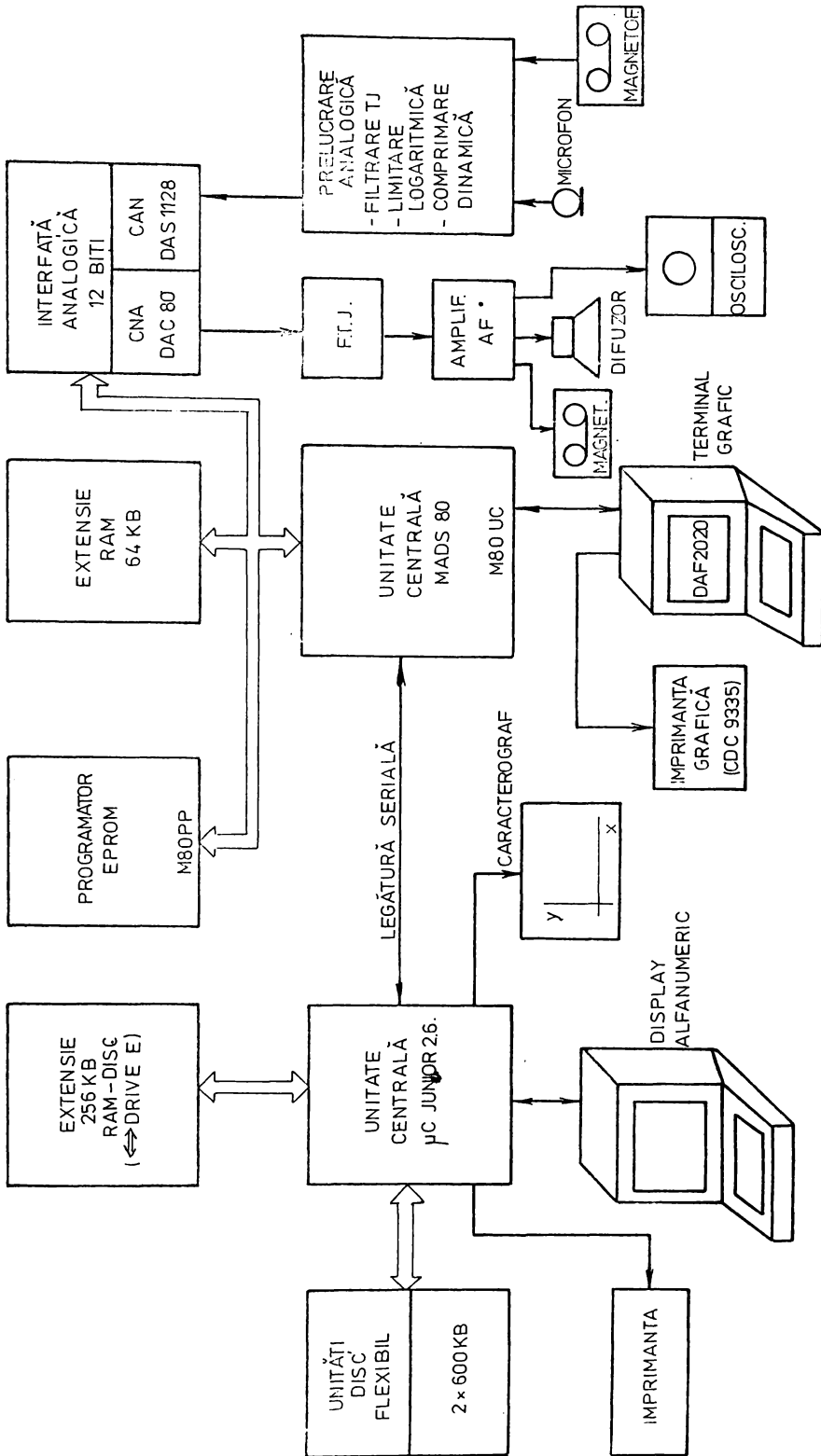


Figura 2.7. Schema bloc a sistemului de analiză realizat.

Utilizarea microcalculatorului modular MADS-30 este motivată de existența unor module funcționale, cu care unitatea centrală se poate cupla (extensie de memorie, programator EPROM), precum și de numărul relativ mare de porturi disponibile ale unității centrale, care permit conectarea unei game variate de terminale suplimentare (DAF grafic, sistem de achiziție și restituire date analogice, etc), posibilități de care microcalculatorul Junior nu dispune.

Utilizarea terminalului grafic DAF2020 este motivată de faptul că analiza vorbirii presupune și o vizualizare grafică a unor forme de undă, ce necesită resurse hardware și software speciale, neimplementate pe displayurile obișnuite. În plus, terminalul DAF2020 permite și conectarea directă a unei imprimante grafice, incluzând și softul adecvat de interfațare.

Din punct de vedere al funcțiunilor pe care le poate realiza, sistemul de analiză descris anterior răspunde cerințelor generale ale unui astfel de sistem, enumerate în paragraful introductiv (2.1) al acestui capitol.

Punerea la punct și extensia acestui sistem de analiză s-a făcut în mai multe etape /31/, /32/, /33/, iar utilizarea sa și o parte din soluțiile elaborate au fost extinse și în alte domenii, ca recunoașterea automată a vorbirii /31/, /34/, analiza semnalelor biologice /35/, realizarea de sisteme de achiziție standard pentru calculatoare autohtone /36/, /37/, etc.

2.3.2. Extensia de memorie RAM-DISC

Implementarea unei extensii de memorie în cadrul microcalculatorului Junior este posibilă datorită faptului că, prin construcție, această posibilitate a fost prevăzută, facilitată existentă și la alte tipuri de microcalculatoare.

În acest scop, la microcalculatorul Junior, blocul de memorie internă cuprins între adresele 4000H și 0C000H și având dimensiunea de 32 KO, poate fi deselectat, creîndu-se astfel un spațiu de adresare disponibil pentru extensii de memorie.

În cazul de față, s-a realizat o extensie de memorie de 256 KO, organizată în 4 blocuri de câte 64 KO și 8 pagini, de câte 32 KO fiecare. Dimensiunea extensiei s-a ales astfel încît să corespundă capacității unei fețe de dischetă, în simplă densitate, în ideea că această extensie să fie implementată astfel

încît să fie considerată de către sistemul de operare al microcalculatorului ca aparținînd unei unități suplimentare de disc flexibil.

În acest scop a fost modificat sistemul de operare pentru ca extensia de memorie să fie admisă ca o nouă unitate de disc flexibil, notată "E", adăugîndu-se nucleului BIOS al sistemului de operare /38/ rutinele necesare.

S-a creat astfel o structură nouă de disc, mai simplă ca cele reale, cu 8 piste, fiecare pistă cuprinzînd cîte o pagină de 32 KO din memoria suplimentară, renunțîndu-se la pistele alocate sistemului și existente la discurile reale. La rîndul ei, fiecare pistă a fost împărțită în 256 sectoare, a cîte 128 octeți fiecare, corespunzător organizării unei dischete obișnuite, în simplă densitate.

Extensia de memorie a fost realizată cu 4x8 capsule de memorie RAM dinamică, de tip 4164 /39/, fiecare circuit avînd o capacitate de 64 kbiți, motiv pentru care această extensie a fost denumită și "disc virtual" sau "RAM-DISC".

O schemă bloc detaliată a acestei memorii suplimentare este reprezentată în figura 2.8. Extensia de memorie este conectată la magistrala de date a microcalculatorului prin intermediul unei memorii tampon (BUFF MEM). Comanda memoriei tampon se face printr-un decodificator (DCD) care permite accesul la extensia de memorie doar în cazul adreselor alocate acestei extensii.

Conectarea extensiei la magistrala de adrese a microsistemului se face prin intermediul unui multiplexor de date (MUX-ADR), dar numai pentru liniile $A_0 - A_{14}$. În locul liniei A_{15} a magistralei de adrese se conectează la intrarea MUX.ADR ieșirea E_0 a unui alt circuit multiplexor (MUX.PAG), care are rolul de a face selecția între paginile (pare, sau impare) corespunzătoare unui bloc.

Selecția paginii se face printr-un port, realizat cu un registru (LATCH CPU), iar accesul procesorului la acest registru se face printr-o memorie tampon de intrare ieșire (BUFF I/O), comandată de un circuit logic combinațional (CLC), prin intermediul căruia se realizează semnalul de comandă din semnalele corespunzătoare magistralei de comandă-control și a adresei respective.

Comanda efectivă a accesului la extensie se realizează prin ieșirile multiplexorului MUX.PAG, care au următoarea semnificație:

E_0 - este folosită direct la selecția unei pagini (pare, sau impare) de 32 KO, din cele două disponibile într-un bloc de 64 KO ;

E_0, E_1, E_2 - sînt folosite pentru a realiza codul binar al paginii (1 din 8) ;

E_3 - este folosită pentru a permite, sau a inhiba, accesul microsistemului la extensia de memorie

Biții E_1 și E_2 sînt decodificați prin intermediul decodificatorului DCD RAS, împreună cu semnalele de comandă pentru citire (MRD) și scriere (MMR) emise de microsistem.

Se obțin astfel, la ieșire, *patru* semnale binare de selecție pentru cele patru blocuri de 74 KO, în care este împărțită extensia de 256 KO, considerate corependente semnalelor de selecție a rîndului (Row Address Strobe - RAS) în unitățile de memorie obișnuite ale microsistemelor.

Cele patru porți SI au rolul de a corela generarea semnalelor RAS cu durata cererii de reîmprospătare (refresh) \overline{RFSH} primită de la microsistem. Din cele 4 semnale de RAS, din care doar unul este activ la un moment dat, se generează și un semnal necesar multiplexării adreselor (SEL), precum și semnalul de selecție a coloanei blocurilor de memorie (CAS- Column Address Strobe) în care este împărțită extensia, prin realizarea unor întîrzieri corespunzătoare.

Pentru a permite accesul la extensia de memorie și în regim de acces direct la memorie (DMA), s-a prevăzut registrul LATCH DMA, prin intermediul căruia este comandat multiplexorul MUX PAG în acest caz.

LATCH DMA este văzut de microsistem tot ca un port, dar numai pentru scriere, avînd aceeași structură ca și LATCH CPU.

Multiplexarea acestor două registre se face în MUX PAG, prin intermediul semnalului de comandă DACKO, emis de microsistem, care semnalează o cerere de transfer de tip DMA pe canalul 0 al circuitului DMA din microsistem. Utilizarea extensiei de memorie s-a dovedit a fi foarte avantajoasă în cadrul sistemului de analiză realizat, mai ales la depozitarea semnalelor analogice achiziționate, asigurînd o fiabilitate în funcționare superioară unităților de disc flexibil, precum și timpi de acces de 6-8 ori mai mici (funcție de instrucțiunea utilizată) față de același periferic.

2.3.3. Rezultate experimentale obținute

Cercetările efectuate cu sistemul de analiză prezentat anterior, asupra semnalului vorbit, au urmărit mai multe obiective,

fundamentale din punct de vedere a sintezei, și anume :

- evidențierea formelor de undă ale sunetelor fundamentale ale vorbirii (foneme), în domeniile timp și frecvență ;
- analiza modului de tranziție de la un fonem la altul, în cadrul vorbirii cursive ;
- evidențierea parametrilor vorbirii ce caracterizează un vorbitor dintre mai mulți ;
- evidențierea parametrilor care diferențiază un fonem pronunțat normal de unul accentuat ;
- determinarea coeficienților de predicție liniară pentru fonemele limbii române, etc.

În cadrul acestor cercetări s-au obținut o serie de rezultate experimentale, materializate prin grafice, reprezentând formele de variație în timp și frecvență, precum și valorile coeficienților de predicție liniară, specifice fonemelor limbii române.

Totodată, s-a reușit și obținerea unor așa zise "suprafețe spectrale", caracteristice anumitor foneme, sau cuvinte, prin reprezentarea pe același grafic a spectrelor determinate la intervale mici de timp și care oferă o informație mai bogată asupra dinamicii unora din parametrii fundamentali ce caracterizează vorbirea.

Rezultatele experimentale menționate sînt prezentate în anexă, împreună cu descrierea și listarea principalelor programe de analiză utilizate (achiziție și conversie semnale analogice, transformată Fourier, analiză prin predicție liniară, etc).

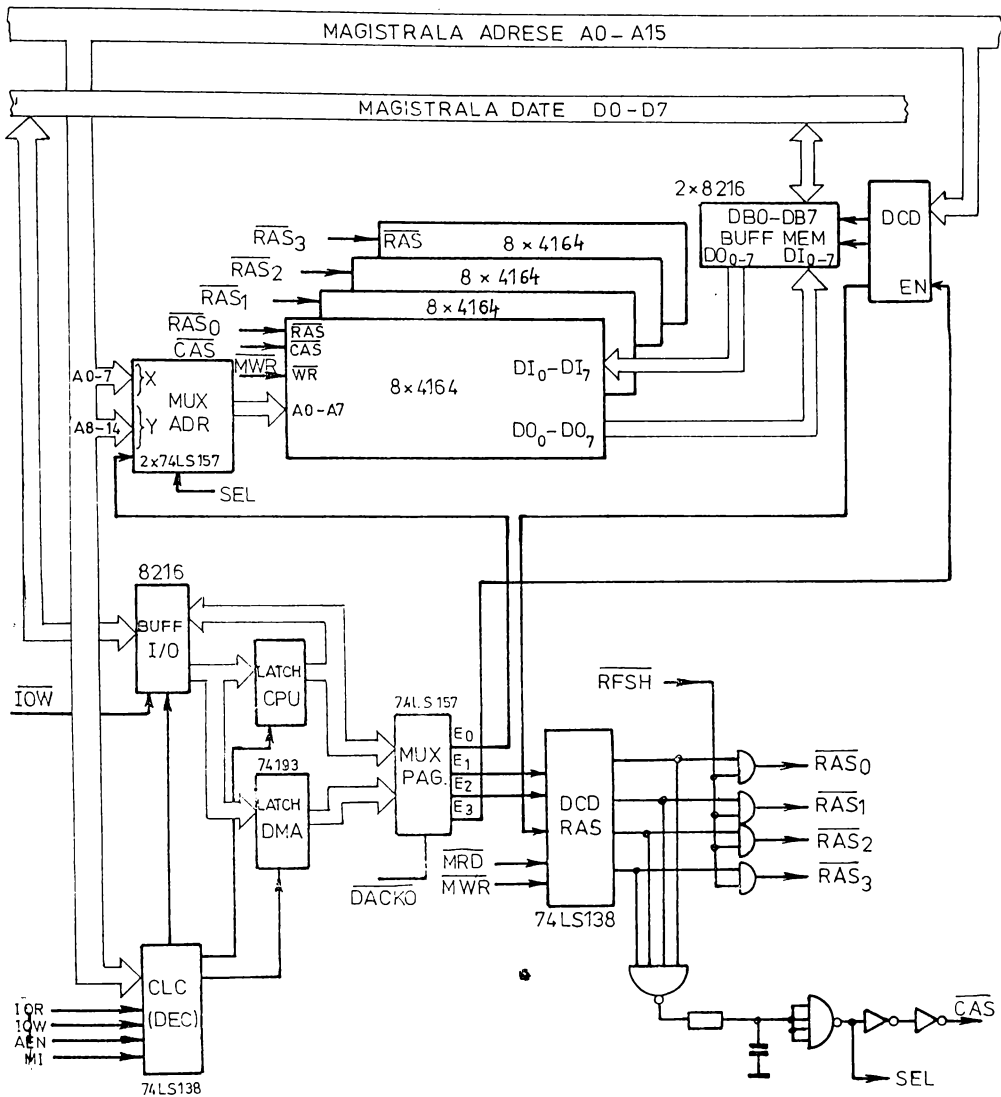


Figura 2.8. Schema bloc a extensiei de memorie „RAM-Disc”

3. SISTEME CU LOGICĂ PROGRAMATĂ PENTRU SINTEZA VORBIRII

3.1. Caracteristici generale ale sintetizoarelor de vorbire bazate pe logică programată

Pornind de la clasificarea generală a metodelor de sinteză a vorbirii în :

- metode cu codificarea (refacerea) sursei de semnal și
- metode cu codificarea (refacerea) formei de semnal, rezultă și principalul mod de clasificare al sintetizoarelor de vorbire bazate pe logică programată, care implementează aceste metode.

Sintetizoarele vor prezenta astfel o serie de caracteristici comune, impuse, în special, datorită utilizării logicii programate, precum și o serie de deosebiri, datorate metodelor diferite de sinteză pe care le implementează. Principala caracteristică comună a acestor sintetizoare o constituie arhitectura lor, care este, în principiu, similară celei de calculator, la care se adaugă și o interfață pentru conversia digital-analogică și amplificarea semnelului sintetizat, conform schemei bloc generale reprezentate în figura 3.1.

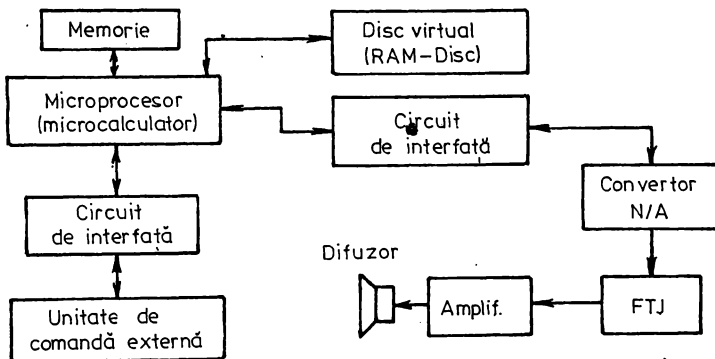


Fig.3.1. Schema bloc a unui sintetizor bazat pe utilizarea logicii programate

O altă caracteristică comună o reprezintă faptul că informația de bază necesară sintezei, provenită dintr-o etapă anterioară, de analiză, este rezidentă în memoria sistemului respectiv, fiind înregistrată în procesul de fabricare a sintetizorului, pentru sintetizoarele cu vocabular nelimitat, sau pentru cele cu vocabular limitat, dar fix, sau la fiecare inițializare a sistemului cu logică programată, în cazul sintetizoarelor cu vocabular limitat, dar variabil, ce necesită schimbarea informației de bază, odată cu cea a vocabularului.

Sintetizoarele cu logică programată care necesită reprogramarea informației de bază, funcție de vocabularul ce se dorește a fi sintetizat, vor trebui să includă și sistemul de analiză pentru obținerea acestei informații, adică să aibe o schemă bloc corespunzătoare celei din figura 1.19, sau să dispună măcar de un port de intrare, destinat conectării unui periferic de tip disc flexibil, sau unitate de casetă magnetică, prin intermediul căreia să se poată introduce în memoria sintetizorului informația de bază respectivă, obținută prin intermediul unui sistem de analiză separat.

Cea de-a treia caracteristică comună a sintetizoarelor cu logică programată o constituie faptul că informația de bază este înregistrată în memoria sistemului sub formă digitală, într-una din formele specifice de codificare utilizate în cadrul conversiei analogico-digitale a semnalului vorbit (MLC, delta, predicție liniară, etc). În funcție de dimensiunea vocabularului ce se dorește a putea fi sintetizat, depinde și cantitatea informației de bază, deci și volumul de memorie necesar stocării acestei informații.

În cazul în care acest volum este prea mare, se mai impune și aplicarea unor metode de compresie adecvată, care să reducă dimensiunea informației de bază.

Contextul în care trebuie evaluate aceste probleme îl constituie compromisul necesar a fi efectuat între calitatea vorbirii sintetizate și dimensiunea vocabularului dorit a se obține, cele două obiective impunând condiții antagoniste de rezolvare.

Diferența principală, ce deosebește sintetizoarele de vorbire bazate pe logică programată, este dictată de apartenența metodei de sintetizare pe care o implementează la una din cele două tipuri de metode menționate și la începutul acestui paragraf.

Astfel, metodele bazate pe codificarea sursei impun sistemului cu logică programată posibilitatea de a implementa struc-

turi de filtre digitale programabile, cu ajutorul cărora să se poată simula funcționarea tractului vocal, corespunzător pronunțării unui anumit fonem, în timp ce metodele bazate pe codificarea formei semnalului nu necesită decât extragerea din memorie a segmentelor necesare formării cuvântului dorit și alăturarea lor în ordinea corespunzătoare, urmată de conversia digital-analogică a secvenței respective.

Metodele de sinteză bazate pe codificarea sursei, fiind metode parametrice, vor necesita o informație de bază mult mai redusă, comparativ cu cele bazate pe codificarea formei, la același vocabular. În același timp însă și numărul de operații pe care-l vor necesita va fi cu mult mai mare, implementarea structurilor de filtre digitale, specifice acestor metode, reclamând operații aritmetice multiple, ce impun viteze de calcul ridicate sistemelor cu logică programată.

În consecință, nu cu orice sistem cu logică programată se va putea realiza un sintetizor de vorbire care să implementeze o metodă de sinteză din categoria celor cu codificarea sursei, ci numai cu cele bazate pe procesoare rapide (bit-slice, procesoare de semnal), care să poată asigura viteza necesară implementării în timp real a structurilor de filtre digitale comandabile /39/.

Microprocesoarele de uz general pot fi luate în considerare, în acest caz, numai dacă sînt însoțite de procesoarele aritmetice specifice familiei de dezvoltare a microprocesoarelor respective, în calitate de coprocesor, și dacă acest ansamblu, microprocesor-coprocesor aritmetic, satisface cerințele de viteză impuse.

În compensație, sintetizoarelor bazate pe codificarea formei semnalului li se cer condiții de viteză ușor de realizat, chiar și cu microprocesoare de uz general, dar, în acest caz, dimensiunea memoriei necesare stocării informației de bază devine prohibitivă, crescînd foarte repede cu dimensiunea vocabularului ce se dorește să fi sintetizat. Apare frecvent, de aceea, necesitatea introducerii unor metode de compresie a informației de bază, precum și utilizarea unor extensii de memorie, la aceste sintetizoare.

Deși volumul extensiei de memorie necesar poate depăși frecvent capacitatea direct adresabilă a microprocesoarelor de uz general, de opt biți, (64 KO), totuși acestea rămîn utilizabile și în aceste cazuri, prin folosirea metodei de paginare a memoriei, care permite unui microprocesor să utilizeze o memorie mai mare decît cea pe care o poate adresa direct, dacă aceasta este divizată

În pagini și dimensiunea unei pagini nu depășește valoarea maxim adresabilă. Microprocesorul respectiv va putea astfel avea acces la unități de memorie ce nu depășesc valoarea sa maximă de adresare directă, dar situate în pagini diferite, care se selectează adecvat, prin programare. Acest principiu poate fi combinat și cu ideea de a utiliza pentru exploatarea unei astfel de extensii softul specializat destinat exploatarea unei astfel de extensii de memorie externă, cum este discul flexibil, ajungându-se astfel la ideea realizării și implementării unei extensii de memorie de tip RAM DISC, ce a fost prezentată în cadrul capitolului anterior.

Din punct de vedere al aplicațiilor urmărite, se poate constata faptul că sintetizoarele bazate pe metodele de sinteză cu codificarea sursei se utilizează mai ales în cazurile în care se dorește realizarea unui vocabular nelimitat, dar cu concesii în ceea ce privește calitatea vorbirii sintetizate (nenaturală), în timp ce sintetizoarele bazate pe codificarea formei de semnal sînt utilizate în acele aplicații ce necesită un vocabular limitat, dar pretind o calitate deosebită pentru vorbirea sintetizată (naturală).

Trebuie menționat faptul că și prin metoda predicției liniare, încadrată în categoria metodelor de sinteză cu codificarea sursei, se poate obține o sinteză de calitate (apropiată de vorbirea naturală), prin creșterea numărului de coeficienți de predicție luați în considerare, care însă conduce și la creșterea informației de bază și, ca stare, la reducerea vocabularului. În acele situații însă, în care tehnologia de realizare a procesoarelor de mare viteză rămîne indisponibilă și nici procurarea lor nu este posibilă, singura cale de realizare a sintetizoarelor de vorbire bazate pe logică programată rămîne cea a implementării metodelor de sinteză bazate pe codificarea formei de semnal.

3.2. Sintetizoare de vorbire cu codificarea sursei de semnal, bazate pe logică programată

3.2.1. Probleme de bază ale sintetizoarelor de vorbire ce utilizează metode de sinteză cu codificarea sursei de semnal

Dintre metodele de sinteză cu codificarea sursei de semnal, pe sisteme cu logică programată au fost implementate mai ales sinteza fonemică, sinteza prin formantă și sinteza prin predicție

liniară. Aceeași tendință s-a manifestat și în domeniul realizării procesoarelor de semnal specializate pentru sinteza vorbirii prin metode din categoria celor cu codificarea sursei de semnal, /11/, /40/, /41/.

Schema bloc generală, valabilă pentru toate cele trei metode, este cea corespunzătoare sintetizoarelor parametrice de vorbire și este reprezentată în figura 3.2.

Sinteza unui anumit fonem, în acest caz, presupune comanda elementelor programabile ale sintetizorului cu anumiți parametri, specifici generării aceluși fonem, și care pot fi de două categorii:

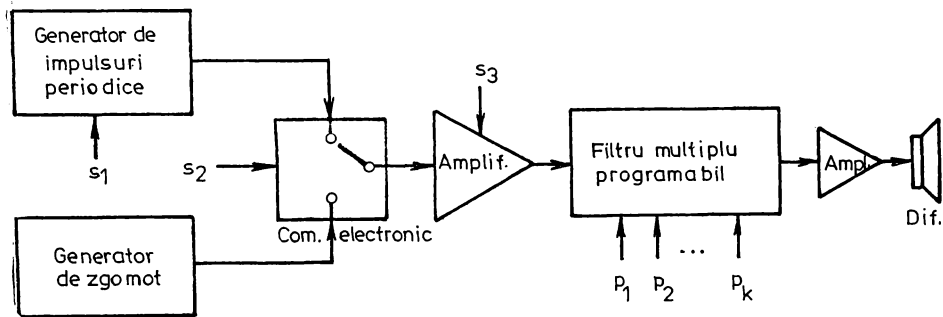


Fig.3.2. Schema bloc generală a unui sintetizor de vorbire, cu codificarea sursei de semnal

- parametri ce caracterizează sursa de excitație (s_1 , s_2 , s_3);
- parametri ce caracterizează filtrul multiplu programabil (p_1 , p_2, \dots, p_k).

Rezultă astfel că pentru a comanda sinteza unui anumit fonem, cu un sintetizor parametric, este necesar un vector (V_f) de forma :

$$V_f = V_f(s_1, s_2, s_3, p_1, p_2, \dots, p_k) \quad (3.1)$$

Parametrii sursei reprezintă :

- frecvența tonului fundamental (s_1);
 - tipul fonemului (conor-nesonor) (s_2);
 - nivelul semnalului furnizat de sursa de excitație (s_3),
- din care s_1 și s_3 reprezintă mărimi continue variabile, iar s_2 este un parametru binar.

Componentele vectorului V_f se determină în faza de analiză, iar mulțimea tuturor vectorilor similari formează informația de bază a sintetizorului, fiind înregistrată în cadrul acestuia și definind dimensiunea vocabularului pe care sintetizorul îl poate genera.

Generatorul de impulsuri periodice reprezintă sursa de excitație în cazul în care se sintetizează un fonem sonor. Perioada acestui semnal (s_1) va fi chiar perioada tonului fundamental al vorbitorului, a cărui vorbire a fost analizată.

Generatorul de zgomot furnizează un zgomot alb și reprezintă sursa de excitație în cazul în care se sintetizează un fonem nesonor.

Intrucât nivelul semnalului de excitație diferă, în cadrul sintezei, de la un fonem la altul, este necesar și parametrul s_3 , menit să comande amplificarea semnalului de excitație pînă la nivelul impus excitării filtrului în cazul respectiv.

Sursa de excitație simulează rolul plămînilor și al glotei, din cadrul procesului de generare a vorbirii umane, iar filtrul multiplu programabil simulează funcționarea tractului vocal și a sursei de radiație (gura, nările) /4/.

În cazul unei soluții analogice de realizare a sintetizorului, filtrul multiplu comandabil va fi alcătuit dintr-o structură de filtre în serie, sau în paralel, menită să aproximeze și să simuleze din punct de vedere electric tractul vocal. Se utilizează, în acest scop, structuri de filtre active, de ordin superior, iar pentru a facilita o programare cât mai flexibilă a acestora, soluțiile moderne apelează la tehnica capacităților comutate /1/, /42/ (figura 3.3).

În cazul soluțiilor digitale, filtrul multiplu programabil este realizat sub forma unui filtru digital, ce poate fi implementat pe sistemele cu logică programată.

Schema bloc generală a unui sintetizor parametric, în varianta digitală de realizare, este reprezentată în figura 3.4.

Se observă că, comparativ cu varianta analogică, în acest caz, amplificatorul sursei de excitație este înlocuit cu un circuit multiplicator, generatorul de zgomot cu un generator de impulsuri aleatoare, iar filtrul multiplu este realizat cu un filtru digital. Interfața dintre partea digitală și cea analogică a acestei variante de realizare este asigurată de un convertor numeric-analogic, iar informația de bază este înregistrată într-o memorie.

În cazul sintezei prin formanți, pentru generarea unui formant (F_i), caracterizat de frecvența centrală f_{Fi} și lățimea de bandă B_{Fi} , se utilizează frecvent structuri de filtre digitale recursive [1/].

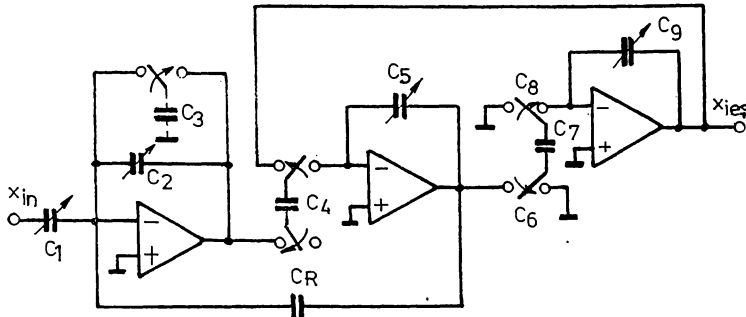


Fig.3.3. Schema de principiu a unei celule de filtru activ, de ordinul doi, cu capacități comutate

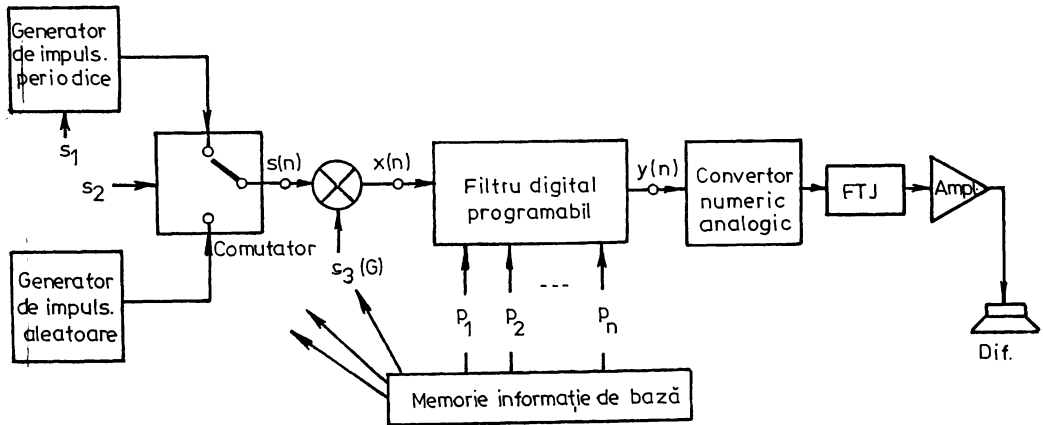
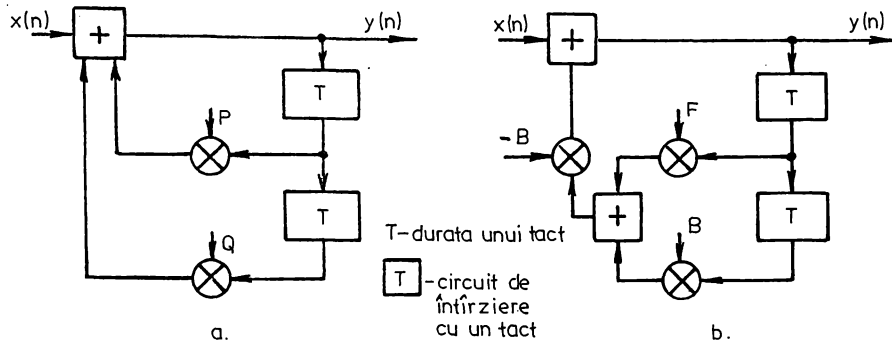


Fig.3.4. Schema bloc a unui sintetizor de vorbire, cu codificarea sursei de semnal, în variantă digitală de realizare

În figura 3.5 sînt reprezentate două variante de structuri de filtre digitale recursive folosite la sinteza formanților în cadrul sintetizozelor digitale, cu sinteză prin formanți.

Pentru comanda filtrelor reprezentate în figura 3.5 se folosesc mărimile :

$$P = 2 \cdot e^{-\pi \cdot B_{Fi}} \cdot \cos(2\pi \cdot F_{Fi} \cdot T) \quad (3.2)$$



Structură cu 2 multiplicatori

Structură cu 3 multiplicatori

Fig.3.5. Filtre formant digitale, de ordinul doi, utilizate în sintetizoarele digitale cu sinteză prin formanți

$$Q = -e^{-2\pi \cdot B \cdot f_i T} \quad (3.3)$$

$$F = 2 \cos(2\pi F_{f_i} T) \quad (3.4)$$

$$B = e^{-\pi \cdot B \cdot f_i T} \quad (3.5)$$

În cazul sintezei prin predicție liniară, semnalul rezultat la ieșirea filtrului digital este dat de relația :

$$y(n) = \sum_{k=1}^p s_k \cdot y(n-k) + G \cdot s(n), \quad (3.6)$$

în care $G = s$, reprezintă câștigul semnalului de excitație, aplicat prin intermediul unui circuit multiplicator.

Se arată /2/ că funcția de transfer a filtrului digital este de tip numai poli, în cazul fonemelor sonore și a unora din cele nesonore, dar include și zero-uri, în cazul fonemelor nazale și a celor fricative. Totuși și aceste cazuri pot fi caracterizate prin funcții de transfer numai poli, în cazul în care ordinul filtrului folosit este destul de mare /2/. În acest caz, funcția de transfer a filtrului digital este :

$$H(z) = \frac{Y(z)}{S(z)} = \frac{G}{1 - \sum_{k=1}^p a_k \cdot z^{-k}} \quad (3.7)$$

și un predictor liniar, caracterizat de coeficienții de predicție $\{a_k\}$, este dat de relație :

$$\tilde{y}(n) = \sum_{k=1}^p \alpha_k \cdot y(n-k) \quad (3.8)$$

S-a văzut că se pot calcula coeficienții de predicție $\{\alpha_k\}$ din condiția de minimizare a erorii de predicție :

$$e(n) = y(n) - \tilde{y}(n) = y(n) - \sum_{k=1}^p \alpha_k \cdot y(n-k) \quad (3.9)$$

La rândul ei, eroarea de predicție $e(n)$ poate fi considerată ca fiind ieșirea unui filtru având funcție de transfer

$$A(z) = 1 - \sum_{k=1}^p \alpha_k \cdot z^{-k} \quad (3.10)$$

Observând că între funcțiile de transfer ale celor două filtre digitale există relația :

$$H(z) = \frac{G}{A(z)}, \quad (3.11)$$

filtrul caracterizat prin funcția de transfer $A(z)$ poate fi considerat drept un filtru invers față de cel caracterizat prin funcția de transfer $H(z)$.

Dacă am lua condiția $\alpha_k = a_k$,
atunci relația (3.9) devine :

$$e(n) = G \cdot y(n), \quad (3.12)$$

ceea ce pune în evidență faptul important că eroarea de predicție este proporțională cu nivelul sursei de excitație și, ca atare, poate fi utilizată și ca semnal de excitație la sinteză.

Sintetizoarele de vorbire ce utilizează metoda predicției liniare se pot realiza în două variante /2/, /9/ ;

- sintetizoare care utilizează ca parametri coeficienții de predicție ;
- sintetizoare care utilizează ca parametri coeficienții de reflexie (PARCOR).

În primul caz, pentru a realiza filtrul digital se folosesc tot structuri de filtre recursive, iar în cel de-al doilea, se utilizează filtre de tip lattice (rețea în X).

În figurile 3.6 și 3.7 sînt exemplificate cele două tipuri de structuri de filtre.

În analiza vorbirii, efectuată în scopul determinării coeficienților de predicție liniară, trebuie să se țină cont de du-

ratele fonemelor pentru care se calculează acești coeficienți și care pot fi de la câteva zeci de milisecunde, pentru fonemele scurte, până la 200-300 ms pentru fonemele lungi și cele accentuate.

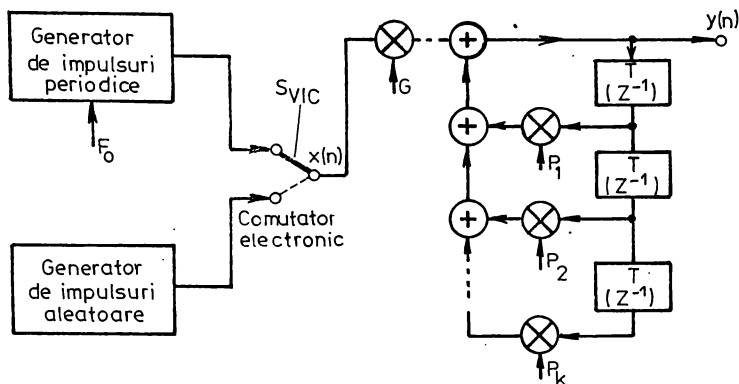


Fig.3.6. Tip de filtru recursiv utilizat în realizarea sintetizoarelor cu predicție liniară

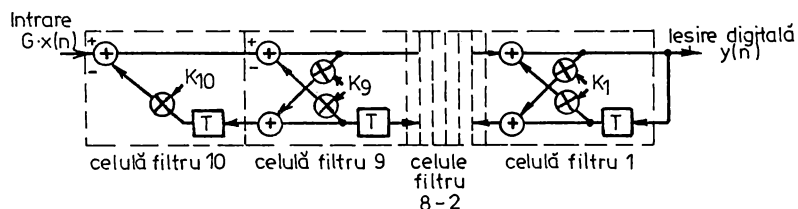


Fig.3.7. Filtru de tip lattice utilizat în realizarea sintetizoarelor cu predicție liniară

Cum analizorul auditiv uman nu sesizează sunetele vorbirii, decât dacă acestea depășesc aproximativ 5 ms /4/, rezultă că alegerca intervalelor de analiză trebuie să depășească această valoare. Pe de altă parte, lungimea de ordinul sutelor de milise-cunde a unor foneme accentuate se poate dovedi prea mare pentru o singură analiză prin predicție liniară, sinteza putând conduce, în acest caz, la erori mari de predicție și la o înrăutățire a vorbirii sintetizate. Este, de aceea, utilizată, în astfel de ca-zuri, caracterizarea fonemelor respective nu numai printr-un sin-gur vector, ci prin câțiva vectori consecutivi, iar segmentele

aceluiași fonem supus analizei nu se aleg disjuncte, ci astfel încât sfârșitul unui segment să fie inclus și în începutul segmentului următor.

Durata segmentelor supuse analizei se alege, de aceea, în general, între 5 ÷ 30 ms, durată pe care comportarea tractului vocal poate fi considerată staționară și caracteristică unui anumit fonem, durată rezultând și din valoarea perioadei tonului fundamental, situată în jurul valorii de 10 ms (pentru vorbitori masculini obișnuiți : $f_P \approx 120$ Hz), astfel încât, în intervalul supus analizei, să existe cel puțin una, sau două, perioade fundamentale.

Pentru fonemele ce depășesc acest interval se pot face mai multe analize întrepesute, conform reprezentării din figura 3.8 /43/.

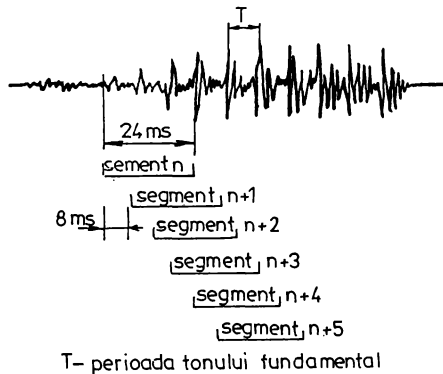


Fig.3.8. Exemplu de selecție a segmentelor de analiză prin predicție liniară, pentru foneme de durată mai mare

Sinteza propriuzisă, prin predicție liniară, cu ajutorul coeficienților de predicție, are la bază relația :

$$y(n) = \sum_{k=1}^p \alpha_k \cdot y(n-k) + g \cdot s(n) \quad (3.13)$$

obținută din combinarea relațiilor (3.6) și (3.8). Ca tip de eroare de predicție se estimează și se reprezintă grafic descoperirea de predicție normalizată (E_n), definită prin raportul dintre puterea erorii de predicție și puterea semnalului de intrare, supus analizei prin predicție liniară, conform relațiilor :

$$P(e_n) = \sum_{n=0}^{N-1} e^2(n) \quad (3.14)$$

$$P(y_n) = \sum_{n=0}^{N-1} y^2(n) \quad (3.15)$$

$$E_n = \frac{P(e_n)}{P(y_n)} \quad (3.16)$$

Pentru a obține o sinteză de calitate a vorbirii, prin metoda predicției liniare, sinteza trebuie să se facă prin intermediul a 10-12 coeficienți de predicție /9/,/13/.

Utilizarea filtrelor digitale recursive, pentru implementarea digitală a filtrului multiplu programabil și necesitatea de a diviza fonemele mai lungi în segmente, pentru analiza prin predicție liniară, impun o tehnică de interpolare între seturile consecutive de coeficienți de predicție, care caracterizează aceste segmente, pentru a sintetiza fonemul respectiv.

S-a constatat /1/, /2/ că în astfel de situații funcționarea filtrului digital nu poate fi garantată stabilă, putându-se amorsa oscilații care conduc la o sinteză eronată. De aceea, la sintetizările prin predicție liniară ce utilizează pentru sinteză chiar coeficienții de predicție propriuziși, trebuie luate măsuri de asigurare a stabilității filtrului digital, care complică structura circuitului integrat respectiv.

O măsură de îmbunătățire constă, de exemplu, în sinteza prin predicție liniară sincronă, în care seturile de coeficienți de predicție corespunzătorii aceluiași fonem se schimbă sincron cu perioada tonului fundamental. Sînt necesare însă pentru aceasta subansamble suplimentare în structura sintetizatorului. Acesta este motivul pentru care se utilizează din ce în ce mai mult, pentru sinteză, la acest tip de sintetizare, nu coeficienți de predicție, ci coeficienții de reflexie (PARCOR) /2/, /13/, a căror variație e limitată în intervalul (-1; 1) și care necesită o structură de filtru special, de tip lattice, pentru care s-a putut demonstra că se asigură o funcționare stabilă /1/,/13/, dar cu prețul dublării operațiilor de înmulțire și adunare, comparativ cazului în care s-ar fi utilizat coeficienții de predicție liniară.

Studiul erorii de predicție pune în evidență faptul că aceasta este mai mică pentru sunetele sonore și mai mare pentru cele nesonore./1/.

Mai mult, se constată chiar că pentru sinteza fonemelor ne-sonore, dublarea numărului de parametri de predicție luați în considerare (de exemplu : 8, în loc de 4) nu conduce la o îmbunătățire semnificativă a calității sintezei fonemului respectiv.

3.2.2. Procesoare de semnal pentru sinteza vorbirii prin metode de sinteză cu codificarea sursei

Considerații generale

Procesoarele de semnal pentru sinteza vorbirii prin metode de sinteză cu codificarea sursei implementează schema bloc generală a sintetizoarelor parametrice, reprezentată în figura 3.2. Analizând diferitele tipuri de sintetizoare integrate, existente la ora actuală în lume /1/, se poate constata faptul că marea lor majoritate utilizează metode de sinteză cu codificarea sursei de semnal (sinteza prin formanți și sinteza prin predicție liniară). Din punct de vedere a structurii lor, se mai poate constata că s-au adoptat două tehnici de realizare a schemei bloc și anume :

- o tehnică mixtă (digital-analogică), în care o parte din subansamble sînt digitale, iar altele analogice și
- o tehnică complet digitală.

Dintre subansamblele care se mai implementează sub formă analogică, se pot menționa : filtrul multiplu comandabil, sursele de semnal de excitație și circuitele lanțului de audiofrecvență (preamplificatoare, filtre trece jos, circuite corectoare de ton), etc.

Tehnica mixtă, analogico-digitală, în realizarea sintetizoarelor de vorbire, permite ca, pe baza utilizării tehnologiei capacităților comutate, să se integreze sub formă analogică și schemele tipic digitale (ca filtrele digitale), păstrînd însă modul de comandă digitală a acestora.

Se remarcă astfel /1/ faptul că toate cele trei componente de bază ale unui filtru digital : registrul, multiplicatorul și sumatorul, pot fi înlocuite, din punct de vedere analogic, cu circuitul de eșantionare și reținere (Sample & Hold), multiplicatorul analogic și respectiv sumatorul analogic, toate incluzînd și capacități comutate, pentru a permite o comandă digitală și a putea astfel înlocui efectiv subansamblele digitale respective.

Se poate exemplifica aplicarea acestei tehnologii în cadrul sintetizorului integrat SC-01 (Votrax), ce utilizează sinteza prin formanți și S 3610 (AMI), ce realizează o sinteză prin predicție liniară /1/.

În cadrul tehnicii mixte, se pot exemplifica procesoarele: SC-01 (Votrax), S3610 (AMI), /1/, SSI 263 (Silicon Systems Incorporated), /44/, etc, iar în cadrul celei digitale : SP0256 (General Instruments), HD51885 (Hitachi), MEA 8000 (Philips-Volvo)/1,60/, etc.

Inregistrarea informației de bază se face aproape în exclusivitate sub formă digitală, în unități de memorie.

Pentru sintetizoarele integrate cu un vocabular restrâns, dedicat unei anumite aplicații, memoria cu informația de bază este, de regulă, integrată în același circuit, dar în cazul sintetizoarelor cu un vocabular extins, volumul memoriei fiind mult mai mare, ea se atașează, de regulă, sintetizorului sub forma unui circuit integrat separat.

Primul caz se poate exemplifica prin sintetizoarele integrate SC-01 (Votrax) UAA 1003 (I.T.T.), iar cel de-al doilea prin procesoarele HD 61885 (Hitachi), TMS5220 (Texas Instruments), etc.

Varianta cu memorie externă este mai flexibilă, putând adapta sintetizorul, prin schimbarea memoriei, la diferite aplicații specifice, care impun un anumit vocabular. Ea mai permite utilizatorului să-și programeze singur memoria, funcție de anumite necesități, și, eventual, să adapteze sintetizorul pentru limbi diferite, dacă structura acestuia permite acest lucru și nu mai include și alte informații specifice unei anumite limbi, înregistrate într-o memorie internă.

Dintre sintetizoarele integrate, unele pot fi utilizate în mod independent, incluzând toate subsamblele necesare funcționării ca sintetizor (exemplu UAA1003 (ITT)), eventual cu excepția memoriei, iar altele sînt concepute pentru a fi utilizate împreună cu un microprocesor de uz general (exemplu : MEA 8000 (Philips-Volvo), putînd fi astfel considerate ca un circuit auxiliar din familia procesorului respectiv, cu rol de periferic, ce-i facilitează acestuia și posibilitatea de realizare a sintezei vorbirii.

În ultimul caz, procesorul integrat pentru sinteză poate fi, eventual, mai simplu, transferînd microprocesorului asociat efectuarea anumitor operații de comandă și control a sintezei, eventual chiar depozitarea informației de bază, într-o memorie

gestionată de microprocesor. Există, de asemenea, sintetizoare integrate, capabile să funcționeze atât independent, cât și în cuplaj cu un microprocesor.

3.3. Caracteristici generale ale sintetizoarelor de vorbire cu codificarea formei de semnal, bazate pe logică programată

Ca și la metodele de codificarea sursei, și în acest caz este nevoie de o etapă prealabilă de analiză, dar obiectul acestei analize are ca scop segmentarea formei de undă a vorbirii în domeniul timp, pentru a extrage segmentele necesare, a le codifica și a le înregistra într-o memorie, constituind astfel informația de bază necesară sintezei. Efectuarea acestor operații impune de asemenea utilizarea unui sistem complex de analiză a vorbirii, similar celui a cărui schemă bloc e reprezentată în figura 1.19. Cu ajutorul terminului grafic, forma în timp a semnalului de vorbire poate fi vizualizată și segmentată, iar segmentele respective se memorează, fiind codificate prin adresa de început și cea de sfârșit a zonei de memorie respective, sau prin adresa de început și numărul de octeți ocupat de segment.

Segmentele se vor alege astfel încât să nu se înregistreze de două ori segmente similare, dar provenite din cuvinte diferite.

Utilizându-se pentru sinteză segmente ale funcției amplitudine-timp a semnalului vorbit, sinteza va fi de o calitate mai bună, deoarece, spre deosebire de metodele de sinteză parametrică, în acest caz vorbirea sintetizată va păstra și informații referitoare la persoana care a pronunțat cuvintele supuse analizei și deci asemănarea cu vorbirea naturală va fi mai mare. Operația de segmentare a vorbirii, în scopul sintezei, prin segmente de forme de undă amplitudine-timp este denumită în limba engleză "concatenation" de unde s-a format și un cuvânt românesc, tot mai utilizat (concatenare), care desemnează aceeași operație și face ca acest gen de sinteză să fie cunoscut și sub denumirea de sinteză prin concatenare. Cum s-a mai precizat, segmentarea și, ca urmare, și sinteza, se pot face la unul din următoarele nivele :

- foneme (alofoni) ;
- difoneme ;
- morfeme ;
- cuvinte.

Pe măsură ce diviziunea este mai îngustă (foneme), volumul ocupat de informația de bază scade, dar scade și calitatea vorbirii, iar tehnica de interpolare a segmentelor în cursul sintezei devine mai dificilă. La polul opus, diviziunea maximă (cuvîntul) transformă informația de bază într-un dicționar, vorbirea rezultată are un caracter natural, interpolarea este simplă, dar volumul memoriei necesare este maxim.

Sinteza prin codificarea formei de undă se utilizează, de regulă, în cazurile în care se impune o calitate deosebită vorbirii generate (servicii automate de informare publică), la sistemele de calcul mari, la care dimensiunea memoriei necesare nu constituie o problemă (bănci de date, sisteme expert), și, mai ales, în cazurile în care tehnologia necesară realizării procesoarelor rapide, indispensabilă implementării metodelor de sinteză cu codificarea sursei, nu este disponibilă.

Una din problemele importante care se pune, de regulă, în cadrul sintezei cu codificarea formei de semnal, o constituie reducerea volumului informației de bază, problemă impusă mai ales în cazul în care sistemul de calcul ce implementează metoda de sinteză este un microsistem.

Posibilitatea de a realiza sintetizoare de acest tip, chiar și cu microprocesoare generale, de 8 biți, oferă o gamă largă de aplicații, în acest domeniu, multor utilizatori, dar impune limite prin dimensiunea memoriei necesare stocării informației de bază, impunând și niște soluții optime de realizare și exploatare a unor extensii de memorie.

Pentru a ilustra posibilitățile unor astfel de microsisteme este semnificativ de menționat că în memoria de 64 KO, pe care o poate adresa direct un microprocesor de 8 biți (8080; Z80 etc.), poate fi înregistrat un număr de segmente echivalent cu 8 secunde de vorbire, iar în cea pe care o poate adresa direct un microprocesor de 16 biți (1 MO) se pot înregistra un număr de segmente echivalente cu 120 secunde de vorbire.

Se poate remarca astfel faptul că microprocesoarele pe 16 biți (8086, 68000, etc) oferă deja posibilități avantajoase de realizare a sintetizoarelor de vorbire cu un vocabular mediu (sute de cuvinte) chiar și în cazul utilizării metodelor de sinteză cu codificarea formei semnalului, fără utilizarea nici unei metode de compresie. Progresul rapid al capacității circuitelor RAM dinamice întărește, de asemenea, această concluzie, indicînd

totodată și faptul că extensiile de memorie vor putea fi realizate la dimensiuni din ce în ce mai mici. Totuși, necesitatea introducerii unei compresii a informației de bază se impune încă în multe aplicații și chiar și pentru sistemele cu volum mare de memorie poate contribui la creșterea semnificativă a vocabularului ce poate fi sintetizat.

În realizarea acestei compresii, pot fi avute în vedere două direcții și anume :

- utilizarea unei metode de conversie analogico-digitală cu un debit mai redus (MICD, MICDA, delta, delta adaptivă, etc);
- exploatarea unor particularități a semnalului vorbit, puse în evidență de analiza vorbirii (repetări de forme de undă, valori ne semnificative, etc) ce permit eliminarea unor informații redundante.

Este de remarcat însă că aplicarea celei de-a doua soluții nu trebuie să conducă la o deteriorare semnificativă a calității vorbirii sintetizate, deoarece, în acest caz, s-ar intra și în domeniul sintezei parametrice, ce anulează avantajele de calitate pe care le oferă metodele de sinteză cu codificarea formei semnalului.

Se poate reține însă ideea de a combina compresia specifică metodelor de sinteză cu codificarea sursei cu cea specifică metodelor cu codificarea formei semnalului, pentru a reduce debitul binar al conversiei analogico-binare sub posibilitățile oferite numai de codificarea formei, dar cu o calitate mai bună ca în cazul codificării sursei /45/.

Se poate aminti, în acest context, metoda de compresie combinată Moser, numită astfel după inventatorul ei (Forest Moser) și care a fost utilizată și în cadrul unor sintetizoare integrate, ca : UAA 1003 și UAA 1103 produse de firma ITT /1/.

Din punct de vedere al integrării, se constată că și metodele de sinteză cu codificarea formei de semnal au fost implementate pe circuite integrate specializate, dar numărul lor este foarte redus, în comparație cu cele bazate pe codificarea sursei. Se poate exemplifica astfel cu sintetizorul integrat MN54104 (SPC), produs de firma National Semiconductor.

Există însă un număr destul de mare de sisteme standard de sinteză, implementate pe micro sisteme de calcul și destinate unor aplicații dedicate, în special realizării de servicii de informare automată /1/.

3.4. Sisteme complexe dedicate sintezei vorbirii

Odată cu extinderea sintezei vorbirii, au apărut diverse sisteme dedicate acestui scop, ce oferă o gamă foarte variată de aplicații, funcție de care depinde și complexitatea sistemului.

Există, astfel, de la sisteme simple, capabile să furnizeze doar câteva mesaje vorbite, referitoare la starea de funcționare a unei instalații, până la sisteme complexe și deosebit de complexe, menite să asiste învățarea unor limbi străine, să translateze texte scrise în vorbire, să permită persoanelor handicapate de vorbire, să vorbească și, mai ales, să ofere, automat, o gamă largă de informații din cele mai diferite domenii /1/,/47/

Ca și concepție, ele merg de la structura simplă a unor scheme ce nu conțin decât o memorie, citită prin intermediul unor registre comandate prin generatoare de tact și care furnizează doar un singur mesaj, până la sisteme de dezvoltare foarte complexe, ce pot furniza un vocabular nelimitat, sau specializat pentru anumite domenii, sau sisteme programabile, ce includ și partea de analiză, destinate a fi adaptate de utilizator aplicației dorite. Sînt frecvente, astfel, aplicațiile de genul sistemelor expert în domeniul tehnicii, al medicinei, sau în cel militar. Criteriile de clasificare ale acestor sisteme de sinteză pot urmări și ele diferite caracteristici, care au mai fost menționate, în legătură cu clasificarea sintetizoarelor.

Ar mai putea fi remarcată însă, în acest context, și tendința de a clasifica aceste sisteme, la nivel de utilizator, după criterii ce, de fapt, nu mai țin cont de tipul sintezei, ci de felul aplicației respective.

Se impune însă și o clasificare din punct de vedere al fabricantului de astfel de sisteme în :

- sisteme de sinteză bazate pe un sintetizor integrat și
- sisteme de sinteză bazate pe sisteme de calcul adaptate, precum și din punct de vedere al utilizatorului în
- sisteme de sinteză specifice unei anumite aplicații (programate de fabricant) și
- sisteme de sinteză programabile de către utilizator.

Se mai poate remarca și faptul că, în lipsa unei tehnologii care să permită realizarea unor procesoare specializate pentru sinteză, rămîne doar varianta sintetizoarelor bazate pe sisteme de calcul adaptate, realizate, de regulă, cu ajutorul microprocesoarelor de uz general.

Fabricanții unor astfel de sisteme aparțin, de regulă, următoarelor categorii :

- fabricanți de sintetizoare integrate, ce dezvoltă și sisteme de sinteză mai complexe, bazate pe integrarea acestora (exemplu : Texas Instruments ; National Semiconductor ; Votrax, etc),
- fabricanți de microprocesoare, ce dezvoltă și sisteme dedicate sintezei vorbirii (exemplu : Intel, NEC) ;
- fabricanți de sisteme de calcul complexe, ce includ și funcția de sinteză a vorbirii (exemplu IBM);
- fabricanți de echipamente industriale, sau de altă natură, prevăzute și cu sinteză de vorbire (exemplu : Fidelity Electronics, Ltd. cu echipamentul de șah vorbitor : Chess challenger) ;
- mici fabricanți specializați (exemplu : Ultratronik; Speech Design, etc).

3.5. Contribuții la realizarea sintetizoarelor de vorbire bazate pe logică programată

3.5.1. Precizări generale

Datorită condițiilor concrete de cercetare, în care elementele cu logică programată disponibile au fost microprocesoarele de uz general, pe 8 biți, cercetarea s-a axat pînă acum pe trei direcții principale și anume :

- identificarea și perfecționarea unor metode eficiente de compresie a vorbirii, în scopul reducerii volumului de memorie necesar stocării informației de bază ;
- realizarea unor sintetizoare de vorbire, cu codificarea formei de semnal, pe microsisteme dedicate, bazate pe microprocesoare de uz general, pe 8 biți ;
- identificarea și realizarea unor aplicații, pe bază de sinteză de vorbire, în telecomunicații.

3.5.2. Metodă de compresie a vorbirii bazată pe analiza semnalului vorbit /43/

3.5.2.1. Bazele experimentale ale metodei

Analizînd semnalul de vorbire corespunzător diferitelor foneme se pot constata următoarele caracteristici /3/, /4/, /9/ :

- forma de variație în timp a fonemelor sonore pune în evidență un fenomen de cvasiperiodicitate, evidențiind o perioadă, care este perioada tonului fundamental ;
- forma de variație în timp a fonemelor nesonore este asemănătoare unui semnal de zgomot ;
- amplitudinea fonemelor sonore este de câteva ori mai mare ca a celor nesonore ;
- spectrul fonemelor sonore pune în evidență maximuri de energie spectrală (formanți), ce sînt mai accentuați la vocale, decît la consoane ;
- spectrul fonemelor sonore este un spectru discret, avînd ca fundamentală frecvența tonului fundamental ;
- spectrul fonemelor nesonore este un spectru de zgomot, asemănător zgomotului alb ;
- accentuarea, în cadrul unui cuvînt, se manifestă prin accentuarea unei anumite vocale a cuvîntului ;
- fonemele accentuate durează mai mult ca cele similare, dar neaccentuate, și au o amplitudine mai mare ;
- între silabe și între cuvintele unei pronunțări cursive există pauze, care corespund, ca formă de undă, unui semnal de zgomot, de amplitudine redusă ;
- două foneme situate în cadrul aceluiași cuvînt, și consecutive, se influențează una pe alta, rezultînd o zonă comună, de tranziție ;
- consoanele explozive au o lungime mai scurtă ca celelalte consoane, iar forma lor de undă este destul de dificil de pus în evidență, fiind mascată de începutul, sau sfîrșitul vocalei, cu care se asociază ;
- prin pronunțarea cu un nivel mai puternic, la microfon, amplitudinea semnalului crește, dar forma sa în timp nu se schimbă și nici forma spectrului său ;
- spectrul semnificativ din punct de vedere a inteligibilității este mai redus la fonemele sonore, comparativ cu cele nesonore, primii trei formanți (cei mai semnificativi) fiind, de regulă, situați în banda 300 - 3000 Hz.

3.5.2.2, Prezentarea metodei de compresie /48/,/49/

Metoda de compresie propusă reprezintă o variantă adaptivă a metodei MIC, în care se condiționează atît valoarea frecvenței de eșantionare, cît și debitul informației rezultate în urma con-

versiei, de o analiză prealabilă a semnalului vorbit în domeniul timp și în domeniul frecvență. Pentru expunerea metodei se consideră conversia MIC cu banda de audiofrecvență de 300 - 4000 Hz, frecvența de eșantionare de 8 kHz și cuantizarea pe 8 biți. Metoda se bazează pe particularitățile acestui semnal, care, exploatare în mod eficient, pot conduce la o micșorare semnificativă a debitului binar rezultat în urma conversiei, în condițiile menținerii unei bune inteligibilități a vorbirii reconstituite.

Pe baza analizei vorbirii, corespunzătoare limbii române, și a observațiilor constatate și menționate în paragraful precedent, metoda propune :

- reducerea cu 1 bit a numărului de biți alocați cuantizării în cazul fonemelor nesonore (deci 7 biți/eșantion), avînd în vedere că amplitudinea acestora este de cel puțin două ori mai mică ca cea a fonemelor sonore ;
- limitarea spectrului fonemelor sonore la numai 3 kHz, avînd în vedere că pentru marea majoritate a vocalelor (excepție : "i") primii trei formanți, esențiali din punct de vedere a inteligibilității, sînt situați în gama 300 - 3000 Hz, pentru care se va utiliza o frecvență de eșantionare de 6 kHz, aceeași frecvență rămînînd însă la valoarea de 8 kHz pentru frecvențele nesonore ;
- transmiterea doar a unei singure perioade în cazul fonemelor sonore, însoțită de numărul care indică de cîte ori se repetă acea perioadă în fonemul respectiv ;
- compresia pauzelor existente între silabe și cuvinte, prin înregistrarea doar a unui fragment de pauză (zgomot de fond), ce se multiplică apoi de cîte ori este nevoie la sinteză, pentru a ocupa aceeași poziție ca în cadrul vorbirii supuse analizei.

Deoarece unitatea de conversie specifică sistemelor de conversie analogico-numerică asistate de microcalculatoare este octetul (8 biți), metoda propune, spre deosebire de conversia MIC standard, utilizarea a trei tipuri de octeți diferiți pentru cuantificare, după cum urmează :

- unul de sincronizare, de valoare -123, valoare interzisă pentru cuantizare, a cărui semnificație constă în a indica că începe, sau se termină o producție, sau urmează o schimbare de tip (de fonem (sonor, sau nesonor) ;
- unul pentru indicarea tipului de fonem, în care bitul cel mai semnificativ codifică tipul fonemului (0 pentru cele sonore

și 1 pentru cele nesonore), restul de 7 biți fiind rezervați numărului de repetări ai octeților care urmează până la următorul octet de sincronizare, cu specificarea că în cazul fonemelor nesonore, acest număr va fi 1 ;

- unul pentru a indica amplitudinea cantităților, similar ca în cazul metodei MIC.

Referitor la pauze, acestea vor fi codificate similar fonemelor sonore, considerându-se că o pauză reală poate fi refăcută prin multiplicarea de un număr de ori a unui segment al ei.

Reprezentarea caracteristicilor formelor de undă specifice fonemelor sonore și nesonore, în domeniul timp și în domeniul frecvență, a căror exploatare stă la baza metodei de compresie propuse, este efectuată în figurile 3.9 și 3.10, iar reprezentarea celor trei tipuri de octeți utilizați pentru cuantizare în figura 3.11.

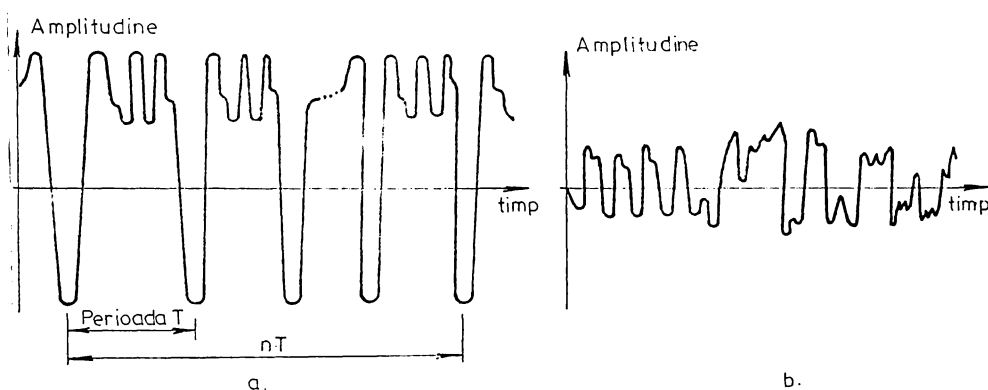


Fig.3.9. Forme de undă corespunzătoare tipurilor fundamentale de fonem. a) fonem sonor (vocală $T \approx 3$ ms, $f_0 \approx 125$ Hz, $n_{\max} \approx 20$; b) fonem nesonor (consonană)

3.5.2.3. Evaluarea eficacității metodei de compresie propuse

Deoarece reducerea unui bit pentru cuantizarea fonemelor nesonore este dificil de exploatat pe actualele micro sisteme de calcul, a căror bază de conversie este octetul, s-a renunțat la evaluarea avantajului introdus de această măsură, deși el este

destul de important, reprezentând, funcție de valoarea frecvenței de eșantionare, o economic de 6-8 kbiți/secundă. Ea poate fi însă luată în considerare în viitor, deoarece

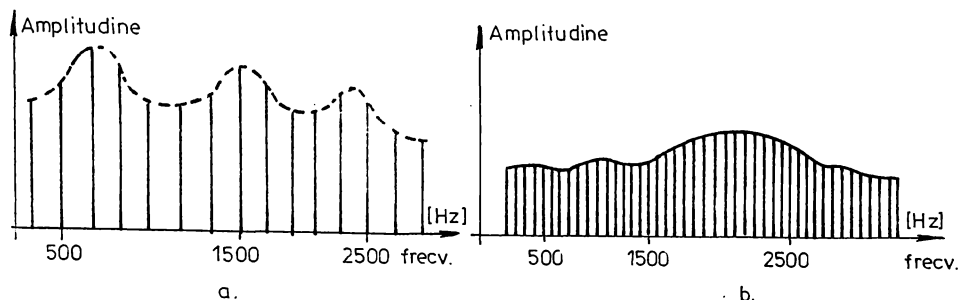


Fig.3.10. Diagrama spectrale corespunzătoare tipurilor fundamentale de foneme. a) fonem sonor (vocală)(spectrul discret); b) fonem nesonor (consoană)(spectru continuu)

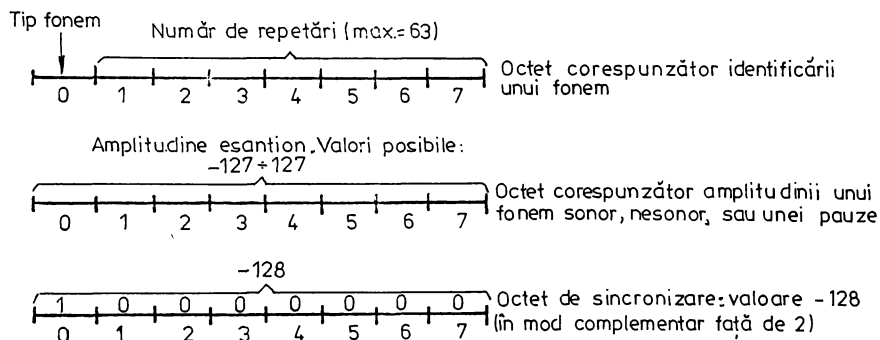


Fig.3.11. Tipuri de înregistrări propuse tipurilor fundamentale de foneme

La microprocesoarele noi (32 biți) există tendința de a se renunța la structura programării clasice, cu alocarea zonelor de transfer și a formatului instrucțiunilor pe octeți.

Pentru evaluarea eficacității în compresie a celorlalte măsuri, pe care le presupune metoda, s-au făcut următoarele ipoteze, care reflectă destul de bine condițiile reale :

- se presupune că într-un cuvânt numărul fonemelor sonore este egal cu cel al fonemelor nesonore ;

- se presupune că un cuvînt este alcătuit, în medie, din două silabe, adică este marcat de trei pauze, din care două marginale și una intersilabică, iar durata medie a unui cuvînt este 1 secundă ;

- se presupune lungimea unei pauze intersilabice de 10 ms, avînd în vedere că analizatorul auditiv uman necesită în jur de 5 ms pentru perceperea unui sunet /4/ ;

- se presupune lungimea unei pauze între cuvinte de 50 ms;

- se neglijează în evaluare numărul octeților necesari pentru indicarea tipului de fonem și pentru sincronizare, avînd în vedere că numărul lor total este de ordinul a cîtorva zeci, iar numărul celorlalți octeți este de ordinul zecilor de mii ;

- se presupune că, în cadrul fonemelor sonore, o perioadă se repetă de 6 ori ;

- se consideră ca unitate etalon, de comparație, debitul de 64 ko/secundă rezultat prin metoda MIC standard și cele 8 secunde de vorbire pe care le poate înregistra memoria maxim adresabilă a microprocesoarelor pe 8 biți (64 ko).

În ipotezele menționate, rezultă că în 64 ko se pot memora 8 cuvinte, care vor înloba 9 pauze între cuvinte și 8 pauze între silabe.

Din aceleași ipoteze, rezultă, pentru pauze, o durată totală de 530 ms la nivelul celor 8 s de vorbire și 66 ms la nivelul unei secunde.

Presupunînd că din totalitatea pauzelor se reține doar o pauză intersilabică (10 ms), din care, prin multiplicare, se obțin și celelalte, rezultă o reducere totală de 520 ms datorată eliminării pauzelor.

Analizînd forma de variație în timp a fonemelor sonore, se constată că numărul de repetări ale perioadei lor fundamentale depinde de tipul fonemului respectiv, cît și de faptul că este accentuat sau nu.

Aceeași analiză pune însă în evidență că o valoare destul de probabilă pentru numărul minim de repetări este cifra 6, iar o valoare minimă acoperitoare pentru durata fonemelor sonore este 60 ms. Alegînd, în cadrul evaluării efectuate, valorile minime corespunzătoare pentru numărul de repetări și perioada fundamentală și pentru durata vocalelor, se poate considera că se acoperă astfel și influența altor factori, ce au fost omiși, și care ar putea cauza efecte de sens contrar, adică de creștere a duratei lor avute în vedere la efectuarea calculelor. Transmiterea unei

singure perioade (≈ 10 ms), din cele 6 presupuse, înseamnă transmiterea doar a 10 ms, în loc de 60 ms, adică o economie de aproximativ 80%, pe care se poate conta în compresia fonemelor sonore.

Referită la durata maximă de 8 secunde de vorbire, ce se pot înregistra în memoria de 64 ko prin MIC standard, această economie se referă la jumătate din numărul fonemelor înregistrate, presupuse a fi de tip sonor, deci la 4 secunde de înregistrare, rezultând o economie de 3,2 secunde, care, privită prin prisma timpului total de 8 secunde, reprezintă o compresie de 50%.

Rezultatele evaluării sînt centralizate în tabelul reprezentat în figura 3.12, în comparație cu cele obținute prin conversia cu metoda MIC standard.

3.5.2.4. Concluzii cu privire la metoda de compresie propusă

Aplicarea metodei de compresie propusă poate conduce la o suplimentare a timpului de înregistrare în memorie standard de 64 ko, direct adresabilă, a unui microprocesor de uz general, pe 8 biți, cu o valoare în jur de 5 s, ceea ce reprezintă un grad de compresie de aproximativ 50%.

Chiar presupunînd ca evaluările efectuate nu au fost în totalitate corecte, și că anumite aspecte ce ar putea diminua avantajele acestei metode nu au fost sesizate, se poate considera că gradul de compresie obținut se va menține la o valoare importantă, ce depășește 50%. Aplicarea acestei metode nu este, în general, posibilă pentru conversia analog-numerică a semnelului vorbit, în sensul în care se realizează această conversie prin metodele clasice (MIC, delta), deoarece presupune analiza prealabilă a fiecărui segment de vorbire supus conversiei. Ea se justifică deocamdată doar prin prisma sintezei de vorbire, contribuind la reducerea informației de bază a sintetizatoarelor și, în consecință, la creșterea vocabularului ce poate fi sintetizat, dar și, eventual, la alte aplicații, ^{care,} în viitor, nu pot fi excluse.

Deși conversia, conform metodei, este complicată și durează un timp considerabil, aceste dezavantaje sînt compensate de creșterea performanțelor sintetizatorului. Setul de programe pentru aplicarea acestei metode de sinteză nu depășește, la înregistrare, cîteva sute de octeți.

Metoda propusă prezintă, în mod firesc, și o serie de dezavantaje. Afară de faptul că este mai complicată și impune o apa-

ratură complexă pentru realizare, calitatea vorbirii refăcute prin sinteză este inferioară celei în care s-ar fi folosit pentru conversie doar metoda MIC standard.

Principalul neajuns este acela că, prin multiplicarea unei singure perioade a fonemelor sonore, se pierde din calitatea vorbirii, deși ea păstrează o inteligibilitate și o naturalitate foarte bune. Dispar însă inflexiunile, apărând un aspect oarecum monoton, sesizat la audição.

În urma analizei au rezultat și câteva măsuri de îmbunătățire a acestei metode, ce mai pot fi luate în considerare, dar care nu au fost studiate și aplicate :

- filtrarea temporală a fonemelor sonore, după reenerarea lor, cu o funcție fereastră, aleasă astfel încît să realizeze o atenuare a nivelurilor de amplitudine la începutul și sfîrșitul fiecărui fonem vocalic, conform evoluției naturale a acestor foneme în cursul vorbirii ;

- eșantionarea vocalei "i" cu o frecvență de 8 kHz, ca și la fonemele nesonore, avînd în vedere că al treilea formant al acestei vocale poate depăși valoarea de 3 kHz ;

- stabilirea frecvenței de eșantionare nu doar la nivel de tip de fonem, ci pentru fiecare fonem în parte, funcție de spectrul semnificativ al acestuia.

Mai trebuie remarcat faptul că metoda de compresie a fost aplicată semnalului de vorbire, convertit în prealabil prin metoda MIC standard, și ar fi de așteptat că dacă conversia s-ar fi efectuat printr-o metodă de conversie adaptivă (MICD, MICDA, delta, delta adaptivă), s-ar fi obținut poate o compresie mai bună ca cea realizată.

Deși unele din caracteristicile utilizate în cadrul acestei metode au fost sesizate și aplicate și de alți autori (exemplu : metoda de codificare Locoer α 1/, s-a considerat totuși că, în ansamblul ei, metoda prezintă caracter de nouitate și a fost brevetată /49/.

3.5.3. Program de sinteză a vorbirii prin predicție liniară

În cadrul metodelor de sinteză a vorbirii cu codificarea sursei de semnal, s-a realizat un program de sinteză a vorbirii prin predicție liniară, ce realizează sinteza cu ajutorul parametrilor de predicție. Programul este scris în limbajul PASCAL și

Nr. crt.	Tip conversie/compresie	Debit eşantionare [mii eşant/sec]	Reducere faţă de MIC standard [mii eşant/sec]	Debit binar rezultat în urma conversiei analog-digiale [kbit/sec]	Reducere faţă de MIC standard [kbit/sec]	Dimensiune memorie ocupată de debitul binar rezultat [kocetei/sec]	Reducere memorie ocupată faţă de cazul MIC standard		Reducere memorie pentru 8 secunde [kocetei]	Durată posibilă de înregistrare în 64.koct. [sec]	Durată suplimentară de înregistrare faţă de MIC standard	
							[kocetei/sec]	[%]			[sec]	[%]
1.	MIC STANDARD (feşant = 8 kHz)	8	0	64	0	8	0	0	0	8	0	0
2.	Metodă de compresie propusă	4,3	3,7	34,64	29,36	4,3	3,7	45,2	29,6	14,03	6,88	85

Observații :

- metoda de compresie propusă se bazează în principal pe următoarele considerații :
- 1 sec.vorbire = 0,467 sec. vocale + 0,467 sec. consoane + 0,066 sec.pauze
- vocalele (fonemele sonore) se eşantionează cu frecvența de 6 kHz, se înregistrează doar eşantioanele corespunzătoare unei perioade care se vor repeta de un număr de ori, specificat în cadrul unui octet de identificare, ce precede eşantioanele fonemului respectiv. Pentru calculul estimativ s-a aproximat un coeficient de compresie de 0,2.
- consoanele (fonemele nesonore) se eşantionează cu frecvența de 8 kHz, înregistrându-se complet.
- pauzele se tratează ca și vocalele, înregistrându-se un număr de oceteți, care se vor repeta de un număr specificat de ori.
- conform metodei propuse numărul de eşantioane înregistrat pentru o secundă de vorbire, supusă analizei va fi $N = 0,467 \cdot 6 \cdot 0,2 + 0,467 \cdot 8 + 0,066 \cdot 6 \cdot 0,2 = 4,3$ mii eşantioane.

Figura 3.12. Tabel centralizator al performanțelor metode de compresie propuse, comparativ cu metoda MIC standard

utilizează rutinele de înmulțire din biblioteca limbajului respectiv, inclusă în sistemul de operare CP/M.

Dimensiunea programului este de ordinul sutelor de octeți, incluzând și rutine din programul de analiză, descris la capitolul precedent.

S-au calculat, cu acest program, coeficienții de predicție liniară corespunzători diferitelor tipuri de foneme ale limbii române, precum și erorile de predicție corespunzătoare. S-a putut constata că pentru a menține aceste erori la valori cât mai mici este necesar ca segmentul de analiză să corespundă, în cazul fonemelor sonore, duratei tonului fundamental.

Programul a fost rulat pe un microcalculator de tip Junior și calculul răspunsului filtrului digital, la 12 coeficienți de predicție, în cazul sintezei unui fonem sonor, a durat, funcție de fonem, în jur de 20-30 secunde, iar pe un calculator de tip ELLIX PC, realizat cu microprocesorul 8086, pe 16 biți, în jur de 5-10 secunde. Înlocuind rutina de calcul scrisă în limbajul PASCAL cu una scrisă în limbajul de asamblare a microprocesorului 8086, deci cea mai rapidă, a fost nevoie de un timp de 3-4 secunde pentru a calcula 12 coeficienți de predicție, ceea ce indică faptul că nici microprocesoarele uzuale pe 16 biți nu sînt apte pentru o sinteză în timp real.

Aceeași rulare pe un calculator de tip CORAL a durat cu mult sub o secundă, evidențiind astfel posibilitățile procesoarelor bit-slice de a contribui în cele mai bune condiții la realizarea sintetizoarelor de vorbire cu codificarea sursei.

Prezentarea programului de sinteză și a rezultatelor experimentale obținute figurează în anexă.

3.5.4. Sintetizor de vorbire bazat pe codificarea formei semnalului •

În cadrul metodelor de sinteză cu codificarea formei semnalului, s-a realizat un sintetizor de vorbire programabil, putînd efectua o sinteză la nivel de fonem, alofoni, difoni, morfeme sau cuvinte.

Sintetizorul este realizat cu ajutorul microprocesorului Z80 și a fost implementat în cadrul unui sistem complex de analiză și sinteză a vorbirii, a cărui schemă bloc este reprezentată în figura 2.7. Cu acest sistem s-a putut genera vocabularul nece-

sar unor aplicații dedicate și evalua calitatea vorbirii sintetizate, funcție de tipul sintezei.

Informația de bază a sintetizorului este programabilă și se obține pe principiul segmentării (concatenării) vorbirii analizate, la nivelul de segmentare dorit (foneme, alofoni, difoni, morfeme și suvinte).

Au fost efectuate diferite aplicații pentru stabilirea informației de bază necesare unor sintetizoare destinate pentru telecomunicații, roboți, supraveghere și control, informare publică, etc.

Programul de concatenare pentru obținerea informației de bază include multe rutine din cel de analiză, fiind prezentat, în anexă, în cadrul programelor realizate pentru analiza vorbirii, descrise în capitolul precedent.

3.5.5. Sisteme de sinteză a vorbirii dedicate unor aplicații industriale

Cu ajutorul sistemului de analiză și sinteză a vorbirii descris în paragraful precedent, s-a realizat informația de bază pentru două sintetizoare dedicate unor aplicații industriale și anume :

- sintetizor pentru roboți /50/ ;
- sintetizor pentru o instalație de supraveghere automată a funcționării unui reactor chimic /51/.

Prima aplicație realizează o sinteză prin codificarea formei semnalului, la nivel de morfeme (silabe), sau cuvinte, funcție de dimensiunea vocabularului dorit.

Informația de bază a fost obținută prin analiză și concatenare, rezultând o dimensiune de memorie de 16 ko, stocată pe o casetă magnetică și introdusă apoi în memoria RAM a calculatorului asociat. Aplicația a fost concepută pentru a cupla sintetizorul la magistralele microcalculatorului ce stă la baza realizării robotului, ca un periferic de ieșire a acestuia, conform schemei bloc din figura 3.13.

Softul necesar funcționării sintetizorului a fost scris în limbajul BASIC și nu depășește 1 ko memorie. Vocabularul ales pentru sintetizor este reprezentat din cifrele de la 0 la 15, corespunzător codului hexadecimal, pentru a permite exprimarea codificată a unor informații privind starea funcționării robotului,

sau pentru a putea realiza un dialog cu un operator uman, în perioada de învățare a robotului. Sistemul realizat a fost conectat la un calculator personal de tip ZX 81, ce a simulat rolul calculatorului atașat robotului.

Interfața cu magistrala de date a calculatorului a fost realizată prin intermediul unui circuit amplificator de magistrală și memorie tampon de tip 8212, urmat de un convertor numeric-analogic, realizat cu circuitul integrat specializat DAC 08, un filtru trece jos de tip K,C și un amplificator audio, realizat cu circuitul integrat specializat TBA 790 T /52/.

A doua aplicație a fost dezvoltată pentru echiparea unui sistem de supraveghere automată a funcționării unui reactor chimic.

De fapt, întregul sistem, realizat sub forma schemei bloc din figura 3.14, a fost implementat tot pe un calculator personal

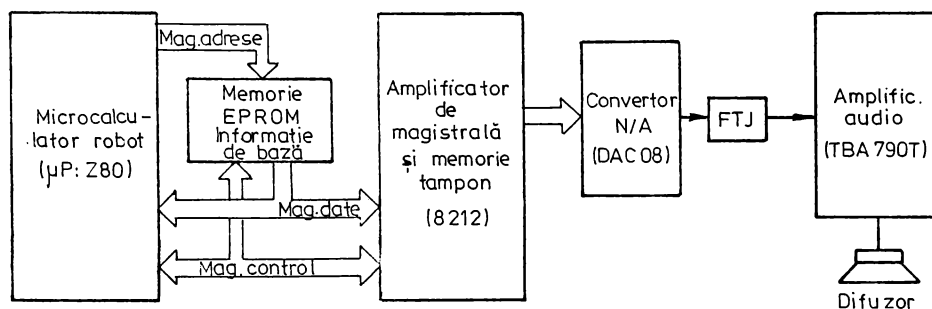


Fig.3.13. Schema bloc a sintetizorului de vorbire destinat echipării unui robot

de tip ZX 81 /53/, pentru a-i asigura portabilitatea, și a supraveghea funcționarea unui reactor chimic din dotarea unui laborator al Facultății de inginerie chimică, din cadrul Institutului politehnic Timișoara.

Sistemul poate urmări 8 canale analogice, prin intermediul unui multiplexor, convertor numeric-analogic și microcalculatorul ZX 81, bazat pe microprocesorul Z80. Softul aplicativ se introduce printr-o unitate periferică serie, de tip casetofon, și este

de ordinul a câteva sute de octeți, iar informația de bază pentru sinteză ocupă 16 ko.

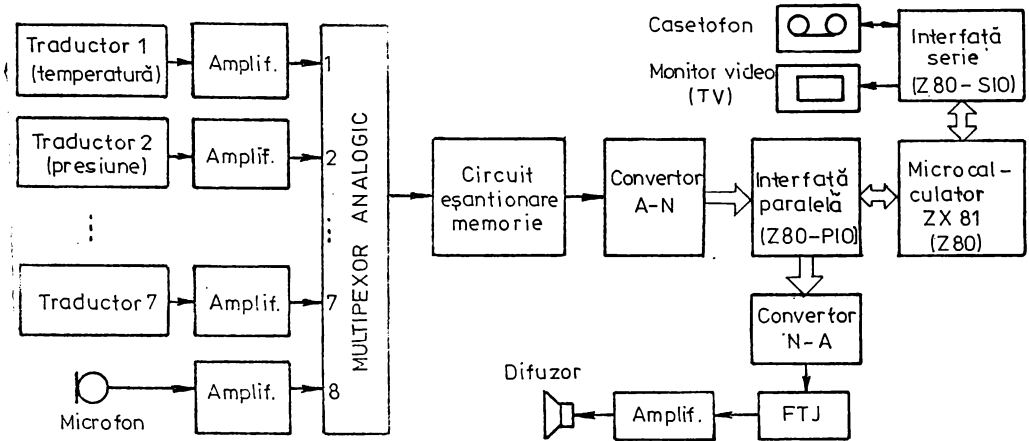


Fig.3.14. Schema bloc a sistemului de supraveghere automată a unui reactor chimic, prevăzut cu sinteză de vorbire

În aplicația practică, s-au urmărit doi parametri (temperatura și presiunea), sintetizorul fiind programat pentru a genera următoarele tipuri de mesaje :

- "crește (scade) temperatura (presiunea)" ;
- "atenție temperatură (presiune) maximă (minimă)" ;
- "alarmă temperatura (presiunea)".

Sinteza se bazează pe codificarea formei, la nivel de cuvinte, vocabularul incluzând cuvintele specificate anterior.

4. APLICATIILE ALE SINTEZEI VORBIRII ÎN TELECOMUNICATII

4.1. Domenii de utilizare a sintezei vorbirii în rețeaua telefonică

Utilizarea sintezei vorbirii în telecomunicații poate cuprinde o gamă destul de largă de aplicații, de la generarea unor scurte mesaje, privind stadiul efectuării selecției în cadrul unei transmisiuni telefonice, până la extinderea gamei de servicii speciale automate, bazate pe vorbire artificială și oferite abonaților telefonici. Restrângând aria de aplicare a vorbirii artificiale la rețeaua telefonică, se pot depista trei direcții principale de implementare a aplicațiilor de vorbire :

- transmiterea de informații abonaților telefonici cu privire la desfășurarea convorbirilor, de la efectuarea apelului și până la terminarea convorbirii ;
- transmiterea mesajelor privind starea funcționării unor echipamente telefonice, prin intermediul unor sisteme de supraveghere automată ;
- realizarea de servicii speciale de informare automată a abonaților telefonici, bazate pe vorbire artificială.

Pentru a exemplifica prima categorie, se poate menționa faptul că în multe rețele telefonice naționale și, în special, în cele ce au la bază sistemul de numerotare deschis /54/, se întâlnește frecvent situația în care cuplarea abonatului cu localitatea dorită este marcată de un mesaj de identificare, bazat pe vorbire artificială, transmitându-i abonatului chemător, pentru a-l informa că s-a conectat la localitatea dorită și poate efectua numărul local al abonatului chemat. Un sistem similar funcționează și în cadrul rețelei telefonice CTA din țara noastră.

O altă aplicație o constituie marcarea cu mesaje speciale a faptului că efectuarea apelului a condus la selecția unui număr ce nu corespunde unui abonat fizic, sau că abonatul căutat și-a schimbat numărul, sau că poate fi găsit temporar la numărul (serviciu de tip "urmează-mă" - "follow me").

În sfârșit, o altă aplicație poate fi transmiterea către abonatul chemător, la sfârșitul fiecărei convorbiri, a unui mesaj informând asupra duratei și costului convorbirii. Astfel de apli-

cații ar contribui semnificativ la reducerea numărului de apeluri eronate, precum și a numărului de reclamații privind achitarea taxelor telefonice.

Al doilea tip de aplicații ar include sisteme de supraveghere automată a echipamentelor unei centrale telefonice automate (roboți de supraveghere), de exemplu, sau sisteme de testare periodică a acestor echipamente. Tot aici ar putea fi incluse și sistemele de tip "flashing" pentru prezentarea identității centrelor telefonice automate, în cadrul unui apel de control efectuat de la distanță.

Implementarea acestor aplicații ar conduce la creșterea fiabilității echipamentelor respective, prin semnalarea mai rapidă a defecțiunilor, precum și la reducerea personalului de supraveghere, sau utilizarea acestui personal și la alte activități, având în vedere că supravegherea se efectuează prin intermediul auzului.

Aplicațiile cele mai importante rămân însă cele din categoria serviciilor speciale automate suplimentare, ce se pot introduce cu ajutorul vorbirii artificiale. Introducerea acestor servicii conduce atât la creșterea productivității muncii în sectorul telecomunicațiilor, cât și la creșterea gradului de confort al abonaților telefonici, putând constitui o alternativă de realizare parțială a serviciilor specifice rețelei de telecomunicații numerice cu integrarea serviciilor (RNIS) /55/, /56/, acolo unde această rețea încă nu a fost realizată.

Din punct de vedere al sintetizoarelor de vorbire, aceste aplicații necesită, de la sisteme simple, pentru generarea unor mesaje înregistrate digital cuvânt cu cuvânt (flashing, mesaje privind starea efectuării apelului, etc), până la sintetizoare complexe destinate transmiterii unor informații ce corespund unui anumit algoritm (robotul de oră exactă), sau sintetizoare programabile, destinate transmiterii unor informații ce trebuie mereu reactualizate (prognoza meteorologică, orarul instituțiilor publice, informații sportive, repertoriul spectacolelor, etc).

Mai trebuie remarcat că, din punct de vedere calitativ, pretenția pentru calitatea vorbirii generate în cadrul acestor servicii variază și ea, de la o inteligibilitate bună, pentru sistemele de informare privitor la starea selecției, sau a stării echipamentelor supravegheate, până la o calitate cât mai aproape de cea naturală, pentru sistemele de informare implementate ca servicii speciale automate taxabile.

4.2. Sisteme de sinteză a vorbirii destinate unor aplicații în domeniul rețelei telefonice

4.2.1. Considerații generale

Sintetizoarele de vorbire implementabile în cadrul rețelelor telefonice variază, ca și complexitate, de la niște simple generatoare de mesaje statice, realizabile și fără aportul logicii programate, cum ar fi sistemele de flashing, sau pentru transmiterea unor mesaje privind cauza nefinalizării unui apel, și pînă la sisteme complexe de sinteză a vorbirii, programabile, destinate transmiterii unor informații cu reactualizare periodică. Omitînd sistemele de supraveghere a funcționării echipamentelor, mai puțin uzitate deocamdată, se poate constata că pe celelalte două direcții de aplicare, sistemele cu vorbire artificială s-au impus și se extind. Sistemele programabile sînt, de regulă, sisteme cu logică programată pentru conversia textului în vorbire artificială, textul fiind introdus prin intermediul unei tastaturi, iar sistemele care le implementează pot utiliza microprocesoare de uz general, sau sintetizoare de vorbire realizate sub formă de circuit integrat specializat.

Procesoarele întregite specializate se utilizează în cazul în care vocabularul necesar a fi sintetizat este mare și asupra calității vorbirii artificiale rezultate nu se impun condiții deosebite.

Pentru aplicațiile în care se impune o calitate deosebită, ca și în cazul sintetizoarelor destinate transmiterii de informații în cadrul unor servicii speciale, se utilizează, de regulă, sintetizoare cu codificarea formei de undă, la nivel de morfeme, sau chiar cuvinte.

Mai este de remarcă și faptul că în gama de servicii speciale automate, prevăzute a fiura în dotarea centralelor telefonice electronice, există și așa numitul serviciu de "informații vorbite" (voice mail), realizat cu ajutorul sintetizoarelor de vorbire.

O caracteristică generală a sintetizoarelor de vorbire implementate pentru transmiterea unor mesaje, sau a unor informații, în cadrul rețelei telefonice este aceea că ele trebuie să permită cuplarea simultană a mai multor abonați care apelează la serviciul respectiv, căci numai așa utilizarea lor devine și eficientă.

4.2.2. Exemple de sintetizoare de vorbire cu logică programată dedicate unor aplicații în cadrul rețelei telefonice

4.2.2.1. Sintetizor destinat transmiterii informației de taxare a convorbirilor telefonice /57/

În cadrul aplicațiilor legate de transmiterea unor mesaje referitoare la modul de decodificare a unei convorbiri telefonice, se prezintă un exemplu de implementare a unui sintetizor pentru transmiterea informației de taxare, realizat de CNET (Franța). Sintetizorul utilizat este cu codificarea formei semnalului, iar informația de bază este înmagazinată în memorie la nivelul cuvintelor, conform schemei indicate în figura 4.1. Sînt înregistrate două fraze de introducere și cifrele cu ajutorul cărora, prin combinații, se pot forma restul cifrelor.

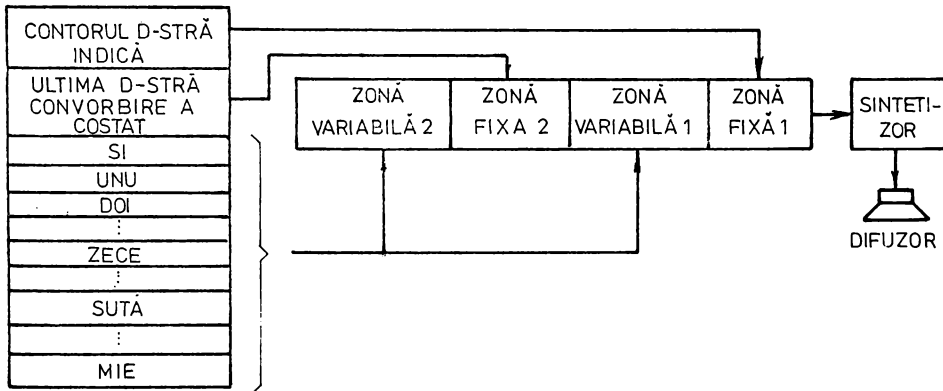


Fig.4.1. Modul de înregistrare a informației de bază și alcătuirea mesajului la un sintetizor destinat transmiterii informației de taxare

Sintetizorul poate fi un sintetizor integrat, conectabil la un microcalculator, sau un microcalculator special dedicat acestei funcții.

4.2.2.2. Sintetizor destinat informării abonaților telefonici cu privire la schimbarea numerelor de telefon /58/

Acest serviciu a fost realizat în cadrul rețelei telefonice maghiare. Se bazează pe un sintetizor cu logică programată,

cu codificarea formei de semnal, realizat cu un microcalculator ce include un microprocesor de tip Z80, pe 8 biti.

Sistemul a fost realizat astfel încât să poată fi cuplat printr-un port de intrare, de tip paralel, și o magistrală analogică, la mai multe linii telefonice. Schema bloc a sistemului este reprezentată în figura 4.2.

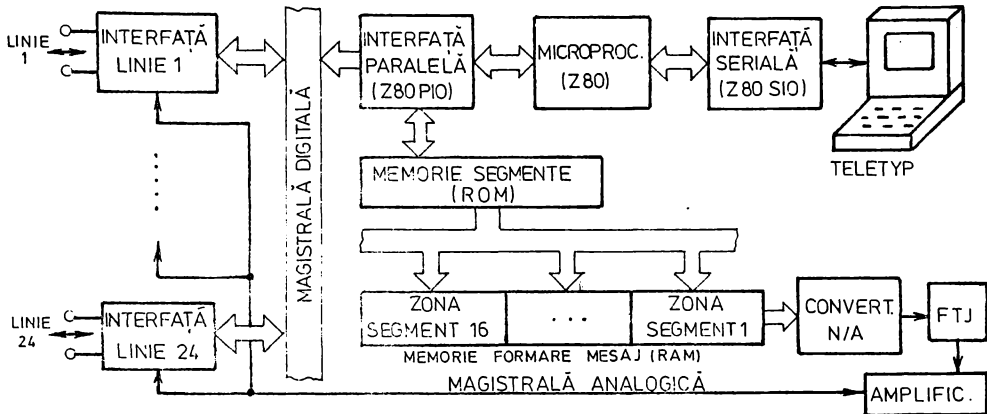


Fig.4.2. Schema bloc a sistemului de transmitere a informațiilor privitoare la schimbarea numărului unui abonat

Informația de bază o constituie morfemele, selectate la nivelul silabelor. Realizatorii menționează că într-o primă fază, la sistem au avut acces 24 de linii de abonat, că intenționează să extindă acest număr la 256 și că pînă la data apariției comunicării, acest sistem a decursiv în jur de 100.000 apeluri.

4.2.2.3. Sintetizor specializat pentru transmiterea orei exacte /11/,/59/.

Este realizat cu un circuit integrat specializat pentru transmiterea orei exacte, fabricat de firma International Telegraph & Telephone (UAA 1003). Circuitul folosește ca semnal de intrare semnalul de codificare a unui dispozitiv de afișare cu 7 segmente, necesitînd deci un cesc de precizie, cu afișaj pe dispozitive cu 7 segmente, de la care pleacă semnalul de comandă.

Schema bloc a acestui circuit este reprezentată în figura

4.3.

Informația de bază necesară transmiterii orei exacte, la nivel de oră și minute, este obținută prin codificarea formei semnalului, adică prin concatenarea cuvintelor respective, reținerea morfemelor necesare și aplicarea unei metode de compresie

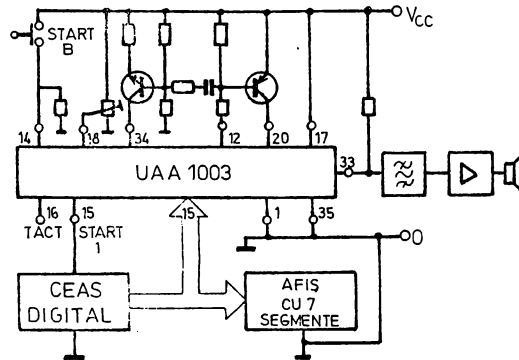


Fig.4.4. Schema electrică de conectare a circuitului integrat UAA 1003

4.3. Contribuții la aplicarea vorbirii artificiale în telecomunicații

4.3.1. Investigații de aplicare a vorbirii artificiale în rețeaua telefonică /61/, /62/

În cadrul unei activități de cercetare privind investigarea unor domenii de aplicare a sintezei vorbirii în rețeaua telefonică, precum și a proiectării și realizării unor sintetizatoare de vorbire dedicate acestui scop, s-a relevat o gamă destul de diversă de aplicații, materializate prin posibilitatea introducerii unor servicii suplimentare. Acestea pot fi clasificate după următoarele criterii :

- după modul de participare al abonatului :
 - servicii care nu necesită un dialog cu abonatul (servicii tip monolog, în care sistemul transmite și abonatul ascultă) ;
 - servicii care necesită un dialog cu abonatul (servicii în care se solicită mai întâi abonatului date cu privire la tipul serviciului, acesta le transmite, după care sistemul transmite și abonatul ascultă);
- după tipul mesajului transmis :
 - cu mesaje fixe ;
 - cu mesaje variabile ;

- după dimensiunea vocabularului :
 - cu număr foarte redus de cuvinte (< 10) ;
 - cu număr redus sau mediu de cuvinte (< 100) ;
 - cu număr extins de cuvinte (> 100) ;
- după tipul serviciului :
 - de supravechere și avertizare ;
 - de informare.

Este de remarcat faptul că serviciile care necesită un dialog, impun dotarea abonatului cu un dispozitiv capabil să transmită informația respectivă. Soluțiile ce pot fi aplicate în acest caz sînt :

- utilizarea în acest scop chiar a dispozitivului de transmitere a informației de selecție (disc, tastatură) ;
- utilizarea unui dispozitiv special (cuplor acustic, tastatură cu cod multifrecvență, sistem de recunoaștere a vorbirii, etc).

Prima categorie de soluții are avantajul că este mai simplă, dar impune modificări în centrală, la nivel de registru, pentru a prelungi legătura galvanică cu abonatul pînă la nivelul registrelor alocate pentru servicii speciale (în cazul transmiterii informației de selecție prin întreruperea tensiunii de alimentare).

Cea de-a doua este mai complexă, necesitînd un dispozitiv special de transmitere, dublat de un decodificator amplasat în centrală, dar nu impune modificări ale echipamentului centralei telefonice.

Realizarea unei game largi de servicii de informare, sub formă unor servicii speciale, taxabile sau nu, impune, în acest caz, ca o soluție generală, amplasarea unui sistem complex de analiză și sinteză a vorbirii în cadrul sectorului de servicii speciale din centralele telefonice, conform schemei bloc din figura 4.5.

O problemă deosebită pe care o ridică introducerea unor servicii complexe de informare o constituie necesitatea unei codificări adecvate pentru informația de bază, avînd în vedere că, în cadrul serviciilor de tip dialog, selecția informației va trebui să fie făcută direct de abonat.

În cadrul unui astfel de apel, sistemul cu vorbire artificială ce deservește serviciul special respectiv, va începe prin a transmite un mesaj de formă : "Serviciul special de informare în domeniul vă stă la dispoziție. Introduceți codul informației dorite".

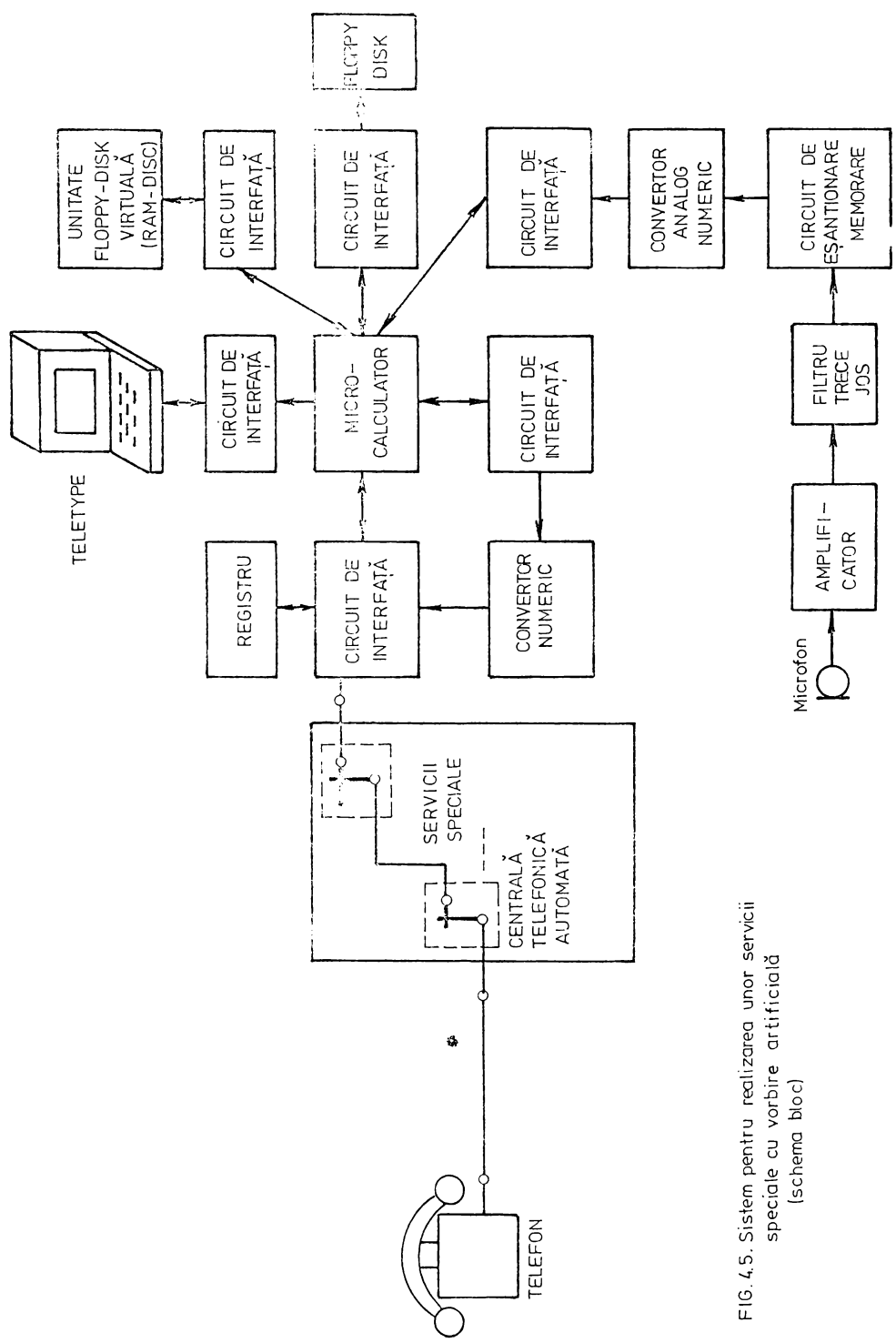


FIG. 4.5. Sistem pentru realizarea unor servicii speciale cu vorbire artificială (Ischema bloc)

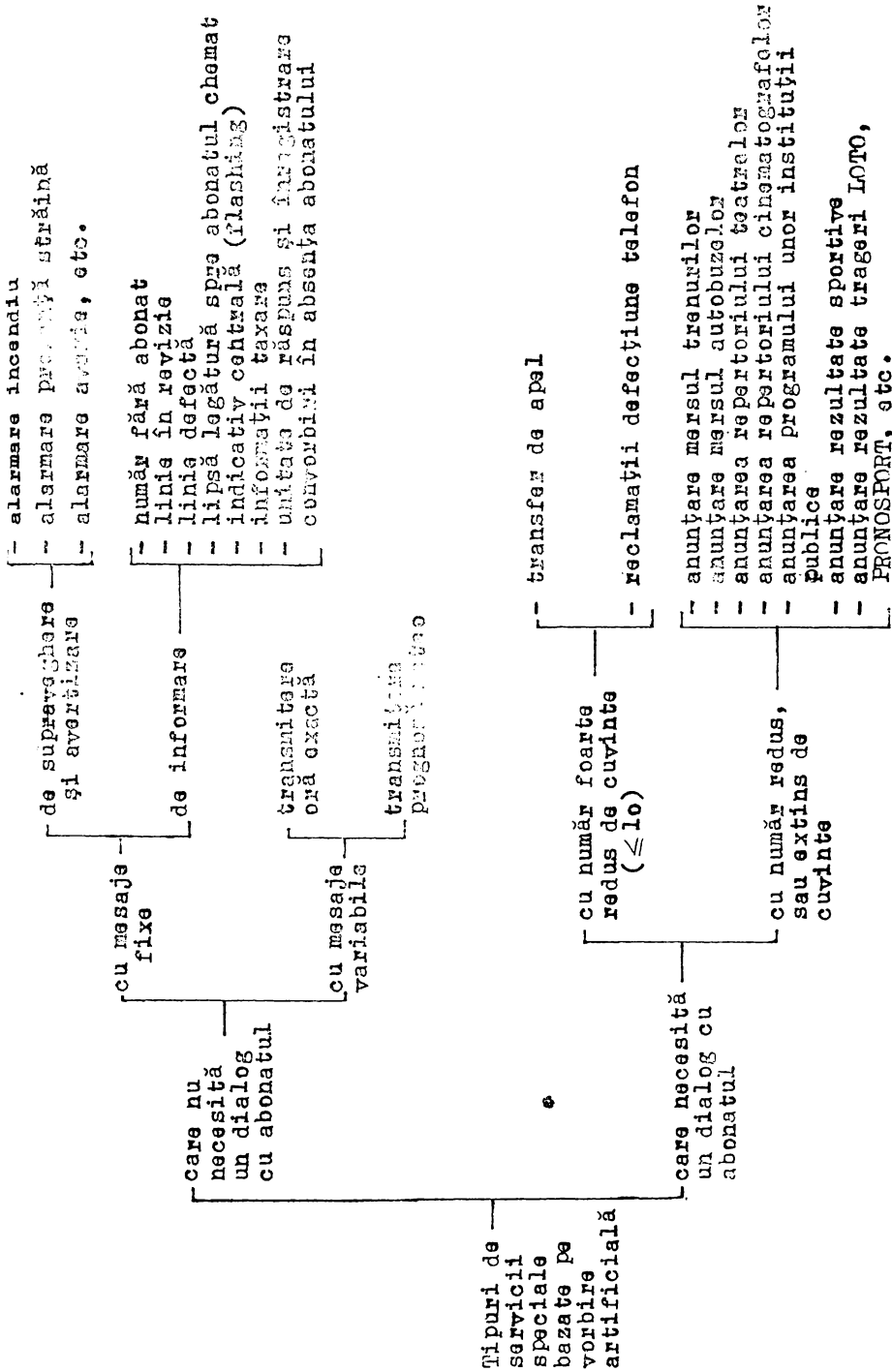


Fig.4.6. Tipuri de servicii speciale posibile a se realiza cu ajutorul vorbirii artificiale

Abonatul va dispune de o listă cu instrucțiuni de utilizare a serviciului respectiv, ce va conține și codurile alocate diverselor tipuri de informații. Dacă codul introdus de abonat nu corespunde uneia din valorile admise, sistemul va repeta comanda inițială, iar dacă apar incidente de dialog (abonatul nu răspunde într-un anumit timp, etc), sistemul se va elibera automat.

O altă problemă deosebită o ridică codificarea impusă serviciilor de anunțare a mersului trenurilor și autobuzelor, datorită marelui număr de localități de pe trasee, care ar conduce la un vocabular atât de mare încât n-ar putea fi implementat. O soluție "salvatoare", în acest caz, o constituie însă existența codificării poștale, care poate fi folosită. Abonatul va indica deci sistemului codul poștal al localității de destinație.

Un tabel centralizator al tipurilor de servicii speciale propuse și posibile a se realiza prin implementarea vorbirii artificiale în rețeaua telefonică este reprezentat în figura 4.6.

4.3.2. Echipeamente complexe, cu vorbire artificială, realizate pentru rețeaua telefonică

4.3.2.1. Sistem de supraveghere automată, prin telefon /62/

În realizarea acestui sistem s-a plecat de la ideea asocierii unui telefon, de concepție specială, la un sistem de supraveghere automată, astfel încât, la un semnal de comandă furnizat de sistemul de supraveghere, telefonul asociat să inițieze un apel și să transmită unui anumit abonat, programat în prealabil, un anumit mesaj, înregistrat într-o memorie.

Schema bloc a sistemului este prezentată în figura 4.7.

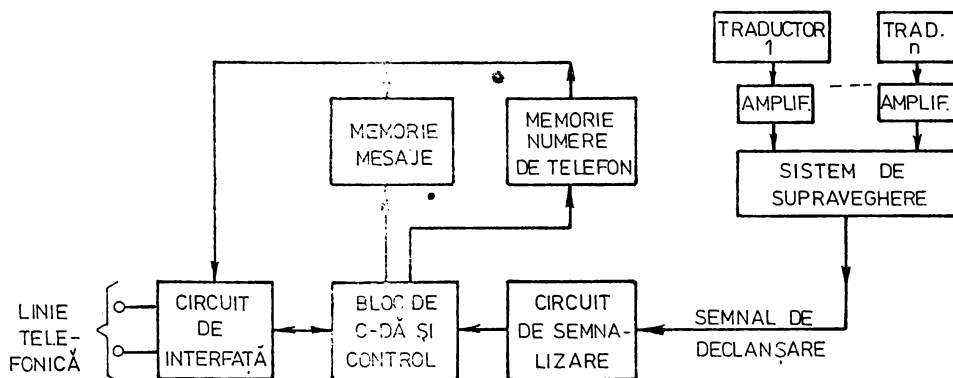


Fig.4.7. Schema bloc a sistemului de supraveghere prin telefon

Sistemul de supraveghere urmărește, prin intermediul tractoarelor sale, funcționarea anumitor procese. În momentul în care unul din procesele supravegheate iese din parametrii normali, sistemul furnizează circuitului de semnalizare un anumit cuvânt de cod, corespunzător evenimentului respectiv.

Acest circuit informează blocul de comandă și control, care declanșează apelul telefonic. Destinația apelului și tipul mesajului ce urmează a fi furnizat vor depinde de cuvântul de cod recepționat de la sistemul de supraveghere.

Prin intermediul circuitului de interfață, blocul de comandă va închide bucla de abonat, solicitând astfel centralei telefonice efectuarea unui apel. Apariția tonului de disc este sesizată cu un filtru de bandă îngustă și declanșează, printr-un releu, întreruperea tensiunii de alimentare, conform numărului ce urmează a fi apelat, extras din memoria de numere de telefon.

Se așteaptă apoi apariția tonului de revers apel și după dispariția acestuia se consideră apelul finalizat și se transmite, din memoria de mesaje, mesajul corespunzător cuvântului de cod ce a declanșat acționarea, după care sistemul se eliberează automat.

Evenimentele incidente de selecție, detectate prin reapariția tonului de disc sau, apoi, a tonului de revers apel, sau de dispariția acestuia, provoacă eliberarea sistemului și reluarea tentativei de apel.

Ca indicator de răspuns al abonatului chemat, se poate considera și inversarea tensiunii de alimentare, ce se produce în momentul în care acesta ridică telefonul, fiind utilizată pentru marcarea taxării.

Un astfel de sistem poate asigura supravegherea funcționării unor instalații, în unele cazuri în care nu se impune prezența permanentă a unui personal de supraveghere, sub forma unui serviciu special, oferit de administrația rețelei telefonice naționale.

Pot fi avute în vedere, astfel, sesizarea unor incendii, emanări de gaz, prezențe străine, întreruperea unor tensiuni de alimentare, etc., ce pot fi semnalate direct celor avizați, prin telefon.

În cadrul rețelei telefonice, o aplicație interesantă ar putea-o constitui supravegherea pe timp de noapte a centralelor telefonice automate.

Sistemul prezentat a fost realizat la nivel de model experimental, iar comanda sistemului de supraveghere a fost simulată prin aplicarea unei tensiuni de comandă circuitului de supraveghere.

Modelul a fost dezvoltat pentru un singur număr de telefon și transmite un singur cuvânt : "alarmă", Extinderea acestui model este însă posibilă prin implementarea logicii programate și a sintezei de vorbire, putându-se realiza un sistem complex, capabil să rezolve în mod eficient servicii de supraveghere automată din cele mai diverse și să intereseze o gamă largă de beneficiari.

4.3.2.2. Sisteme pentru prezentarea automată, prin telefon, a indicativului centrelor telefonice automate (Flashing)

4.3.2.2.1. Sistem de vorbire artificială destinat nunci funcției de "flashing" /53/

Sistemele de tip "flashing", cu care sînt dotate centralele telefonice automate din rețeaua telefonică națională, au devenit inadecvate prin creșterea masivă a numărului de centrale. La ora actuală, majoritatea centrelor telefonice automate dispun de sisteme de tip flashing care generează anumite sunete muzicale, sau sisteme bazate pe utilizarea unui casetofon, cu mesajul înregistrat pe casetă magnetică.

Primele au dezavantajul că sînt greu de memorat și recunoscute, netransmițînd o informație de identificare precisă, constituită dintr-un mesaj vorbit, iar celelalte au o fiabilitate foarte scăzută.

Se impune, deci, înlocuirea acestor sisteme cu altele noi și moderne, bazate pe vorbire artificială, fie sub forma unor mesaje înregistrate digital, fie sub forma unor mesaje sintetizate. În acest scop, a fost realizat un sistem de tip flashing care poate înlocui în mod avantajos actualele sisteme. Sistemul se poate conecta în centrală ca un abonat obișnuit și transmite un mesaj de identificare de forma :

"Aici centrala Pentaconta 2 Timișoara".

Mesajul este înregistrat digital, codificat MIC și comprimat prin eliminarea pauzelor, care se regenerează la transmisie prin intermediul unor întâzieri programate. Sistemul dispune de o memorie de tip EPROM, de capacitate maximă 16 KO, considerată suficientă pentru înregistrarea tuturor mesajelor de acest tip, specifice rețelei telefonice indigenă. Schema bloc a sistemului este reprezentată în figura 4.8, iar schema bloc a circuitului de generare a mesajului sonor în figura 4.9.

După recepționarea apelului, prin intermediul detectorului apel, sistemul transmite, de un număr de ori (programabil),

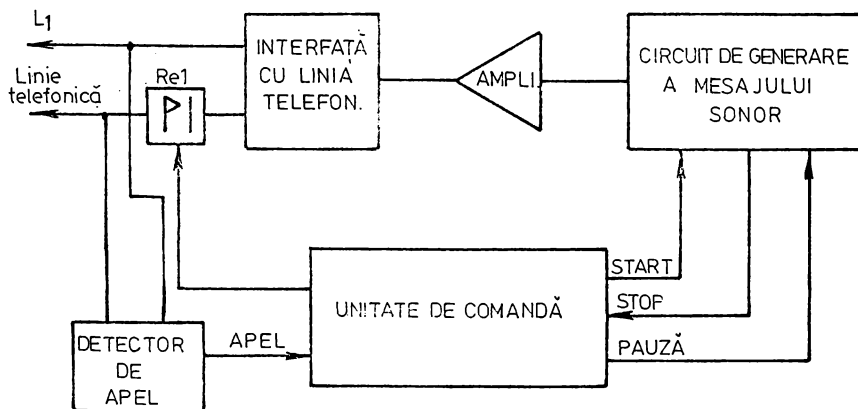
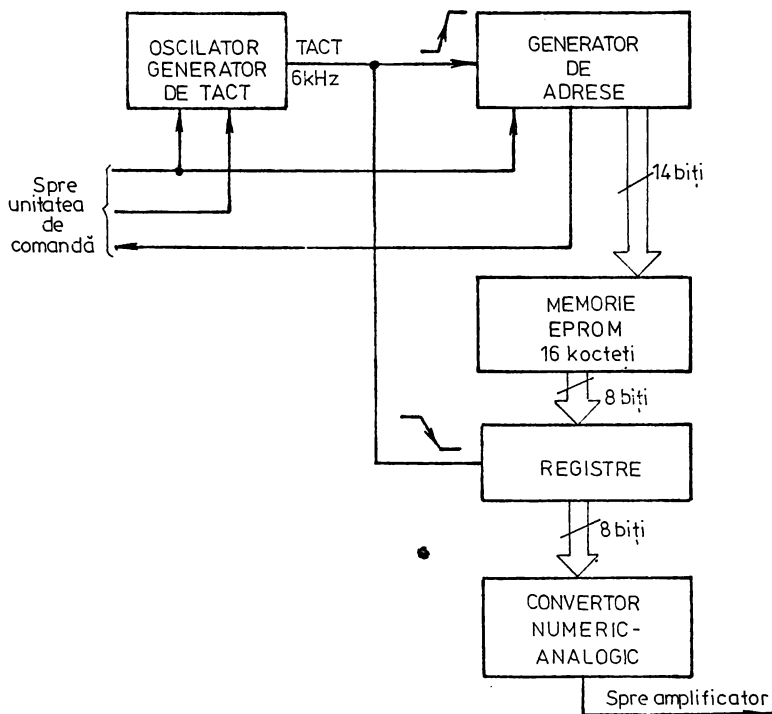


Fig.4.9. Schema bloc a sistemului tip flashing



... Schema bloc a sistemului tip flashing prezentare a 11 paginilor CIAI

indicativul centralei, după care se eliberează automat.
Sistemul descris a fost realizat la nivel de prototip.

4.3.2.2.2. Sistem complex de identificare și testare a centrelor telefonice automate

4.3.2.2.2.1. Probleme ale testării rețelei telefonice

Calitatea deservirii abonaților telefonici depinde, în mare măsură, de modul în care se realizează testarea preventivă a echipamentului de comutație din centralele automate, precum și a celui aferent sistemelor de transmisie și a canalelor de transmisie telefonică. Analizând modul în care se face, la ora actuală, testarea echipamentelor respective din cadrul rețelei telefonice naționale, se pot constata destule lipsuri, cu implicații directe asupra calității abonaților telefonici, bazate, în principal, pe următoarele cauze :

- utilizarea unor echipamente învechite din punct de vedere tehnologic și deseori chiar cu un grad de uzură destul de mare;
- lipsa unor sisteme eficiente pentru testare;
- utilizarea pe scară largă a unor metode de testare subiective, bazate pe conștientizarea personalului implicat în efectuarea acestor operații.

Dacă prima cauză este de ordin general și rezolvarea ei pozitivă reprezintă o problemă de termen lung, legată de schimbarea opticii de investiții în domeniul telecomunicațiilor, rezolvarea celorlalte două este mai ușoară și mai rapidă, putând conduce la ameliorarea situației existente în momentul de față. În acest sens, realizarea unor sisteme complexe și eficiente de testare ar contribui atât la depistarea rapidă a defectelor existente, cât și la prevenirea altora, pe cale de a se produce.

Problema este cu atât mai importantă, cu cât defectarea unui anumit echipament de bază poate antrena scoaterea din uz a unor echipamente dependente de acesta, cum este cazul defectării unui registru în centralele telefonice automate, sau a unui generator de purtătoare într-un sistem de transmisiuni.

Pe de altă parte, actualele metode de testare se bazează încă destul de mult pe încrederea acordată personalului verificator, care nu este întotdeauna justificată de cei în cauză. Afer frecvent situații în care se raportează efectuarea unui număr de testări nereale, precum și situații mai grave, în care operatorii de la un capăt și de la celălalt al unei linii testate se pun de acord asupra rezultatului testării, rezultat ce nu corespunde realității.

Introducerea unor echipamente de testare automată, sau chiar semiautomată, dar efectuate de la un centru dispecer, ar putea elimina aceste aspecte și ar contribui astfel la ameliorarea calității deservirii abonaților telefonici.

4.3.2.2.1.2. Sistem de tip "fleshing" prevăzut și cu posibilități de testare automată

Utilizarea unui sistem automat pentru identificarea centralelor telefonice, bazat pe vorbire artificială, conduce la ideea de a asocia această funcție cu cea de testare și de a realiza astfel un sistem complex de identificare și testare a unei game largi de echipamente din cadrul rețelei telefonice.

Implantat într-o centrală telefonică automată, ca un abonat obișnuit, un astfel de sistem poate fi apelat de la distanță, sau de la un centru dispecer de verificare, poate prelua, prelucra și transmite diferite semnale, oferind astfel informații cu privire atât la funcționarea echipamentelor din centrala în care este amplasat, cât și cu privire la echipamentele de transmisiuni și canalele telefonice prin care este legat cu apelantul.

Conceptia și realizarea unui astfel de sistem de testare s-a efectuat la cererea Direcției generale de poștă și telecomunicații București, pe baza unui contract de cercetare științifică. S-a prevăzut astfel realizarea unui sistem de identificare a centralelor telefonice automate, dotat și cu posibilitatea testării echipamentului de comutație, de transmisiuni și a canalelor de legătură. Pentru realizarea acestui sistem s-au stabilit următoarele cerințe inițiale :

- să fie conectat într-o centrală telefonică automată, ca un abonat obișnuit ;
- să răspundă la al doilea tren de impulsuri al semnalului de apel ;
- să prezinte un mesaj de identificare a centralei în care este amplasat, prin intermediul vorbirii artificiale, cu precizarea :
 - localității ;
 - tipului centralei ;
 - numărului centralei din localitatea respectivă ;
- durata maximă a mesajului : 3 secunde ;
- banda de frecvențe : 0,3 - 3,4 kHz ;
- transmiterea a 5 impulsuri, cu durata de la fiecare și cu factorul de umplere 1/2, pentru verificarea taxării ;
- transmiterea unui semnal de 300 Hz, cu nivel de +4 dBm, pentru un semnal recepționat, de 300 Hz, cu nivel ≥ -13 dBm ;
- întreruperea lucrării, telefonice dacă nivelul recepționat este < -13 dBm ;
- transmiterea unui semnal de test, de 300 Hz, cu nivel de +4 dBm, în trenuri de undă cu durata medie de 2-5 s ;
- realizarea unor întreruperi cu durata medie de 2-5 s (în secvențe) pentru verificarea stabilității ;
- impedanța spre linie :
 - 600 Ω în stare de răspuns ;
 - > 20 k Ω în stare de veghe ;
- alimentare 44-56 V/0,5 A c.c. prin convertor c.c./c.c. de tip ICRET ;
- capacitatea memoriei utilizate :
 - maxim 16 kC RAM ;
 - maxim 32 kC EPROM ;
- construcție și montaj într-un cadru standard tip Pentaconta.

Sistemul a fost conceput și realizat ca un sistem de achiziție și prelucrări de date, bazat pe microprocesor, fiind astfel un sistem programabil. La baza acestui sistem, a cărui schemă bloc este prezentată în figura 4.10, stă utilizarea microprocesorului Z30.

Intrarea în funcțiune a sistemului se produce în momentul în care detectorul de apel sesizează apariția celui de-al doilea tren din semnalul de apel, pentru a se evita declanșarea sistemului de eventuale semnale false, ce ar putea apărea pe linie.

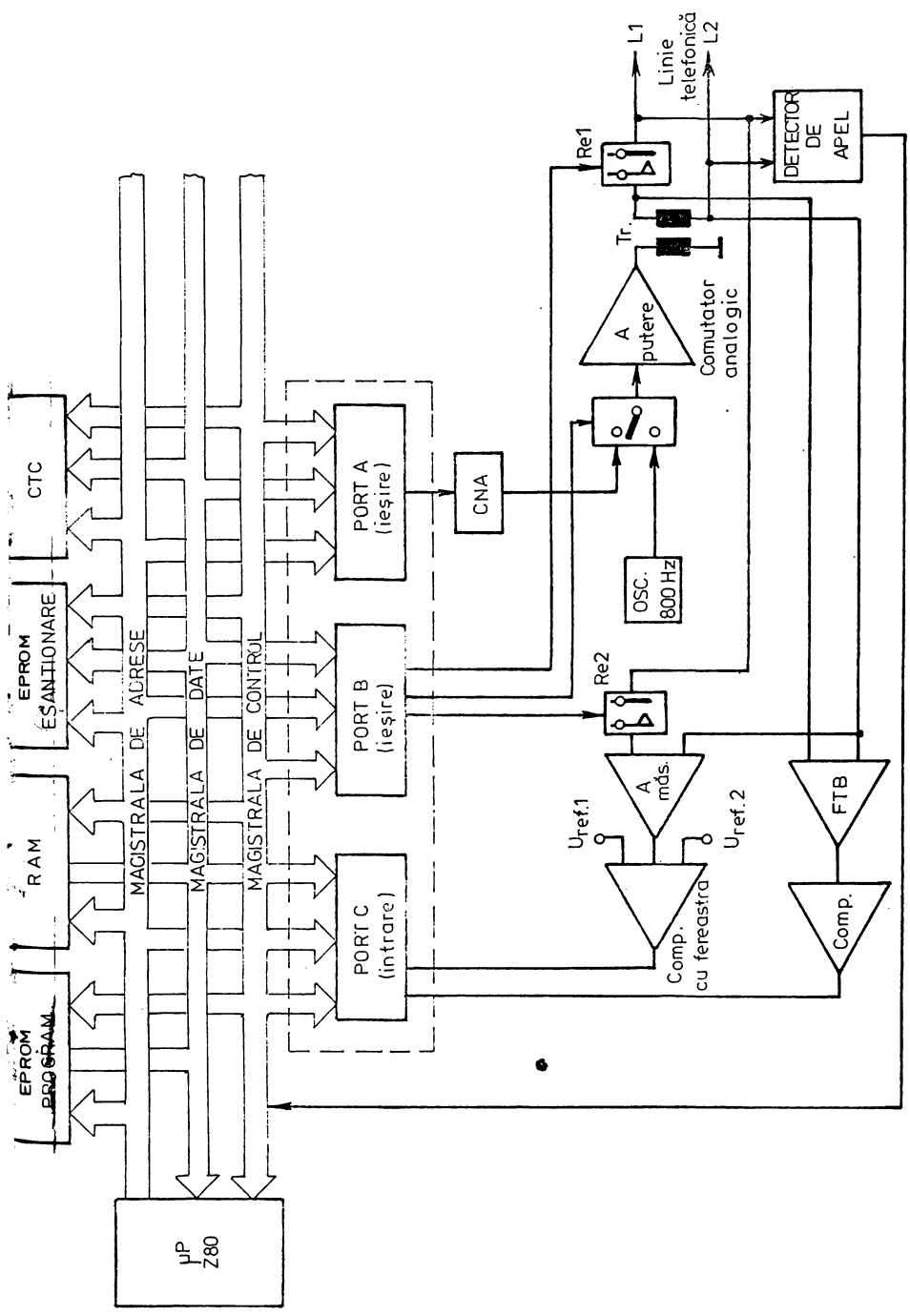


FIG.4.10.Schema bloc a sistemului cu microprocesor destinat prezentării denumirii centrelor telefonice automate, prevăzut cu posibilități de testare

Sesizarea apelului generează o întrerupere mascabilă către microprocesor, în lipsa căreia acesta rămâne într-o stare de așteptare (HALT), fiind programat pentru a răspunde la modul 1 de întrerupere, specific microprocesorului Z80, cu întreruperile validate /SA/.

Un circuit de tip Z80-CYC furnizează temporizările necesare, iar pentru interfatare este folosit un circuit de interfață paralel de tip 8255, ce dispune de trei porturi (A,B,C). Portul A este utilizat pentru conectarea unui convertor numeric-analogic, portul B pentru a transfera semnale de acționare a releelor, iar portul C pentru a vehicula semnale de la amplificatorul de măsură.

La sesizarea cererii de întrerupere, se apelează o rutină de tratare a întreruperii, care transmite, prin intermediul unui circuit de interfață de tip paralel (portul B al circuitului 8255), comanda de anclangare a unui releu (R_{el}), prin care se cuplează la linie un amplificator de putere, prin intermediul unui transformator diferențial, ce reprezintă interfața cu linia telefonică. Se revine apoi în programul principal, apelându-se la rutina de generare a mesajului de identificare, prin intermediul unui convertor numeric-analogic, cuplat la amplificatorul de putere printr-un comutator analogic. Informația de bază, pentru transmiterea mesajului de identificare, este înregistrată într-o memorie EPROM, fie direct, printr-o simplă conversie analogico-numerică, în cazul unor mesaje scurte, fie prin aplicarea unei metode de compresie, în cazul unor mesaje mai lungi.

Mesajul de identificare se repetă de 5 ori, cu o pauză de 1,5 s.

După terminarea transmiterii mesajului, se generează cele 5 impulsuri pentru verificarea taxării, după care sistemul intră într-o stare de așteptare a unui semnal de test, de 800 Hz, transmis de apelant.

Apelantul trebuie să transmită un astfel de semnal, cu nivelul +4 dBm, în termen de cel mult 30 s de la transmiterea de către sistem a impulsurilor de taxare. Dacă acest semnal nu este recepționat, sau este recepționat cu un nivel inferior valorii de -13 dBm, sistemul nu răspunde, semnalizând astfel apelantului că linia este defectă.

Dacă semnalul este recepționat cu o valoare mai mare de -13 dBm, sistemul transmite, la rîndul său, apelantului, un semnal

de 800 Hz, cu valoarea -13 dBm pentru a permite acestuia să testeze calea de recepție.

Pentru a putea măsura semnalul de 800 Hz recepționat, sistemul dispune de un amplificator de măsură, iar pentru a-l transmite, de un generator etalon de semnal. Măsurarea semnalului de 800 Hz recepționat se face cuplând, prin intermediul unui releu (R_{e2}), la linie, o sarcină de 600Ω , la bornele căreia se conectează apoi amplificatorul de măsură, urmat de un detector de valoare de vârf, a cărui semnal de ieșire este aplicat unui comparator cu fereastră, pentru a măsura dacă nivelul semnalului este cuprins în domeniul -13 dBm; $+4$ dBm.

Cuplarea la linie a generatorului de 800 Hz se face tot prin intermediul amplificatorului de putere și a transformatorului diferențial.

Senalele de test sunt transmise de 5 ori, în serie cu durată de 3 s și cu pauze de 1 s între serii. După terminarea transmiterii semnalului de test, se realizează 5 întreruperi ale buclei, cu durată de 3 s fiecare, în vederea verificării stabilității legăturii la întreruperi de scurtă durată, cu o pauză de 1 s între două întreruperi. Întreruperile se realizează prin intermediul releului R_{g1} .

După ultima întrerupere, se cuplează din nou convertorul N/A la intrarea amplificatorului de putere, și se mai transmite o sarcină dată mesajul de identificare, semn că testul s-a terminat, după care sistemul se elichează, prin decuplarea releului R_{g1} , intrând în starea de așteptare a unui nou apel.

Diagrama de timp a funcționării sistemului este prezentată în figura 4.11, iar în figura 4.12 organizarea funcționării.

Sistemul de identificare și testare descris urmează a intra în fabricație de serie la întreprinderea de construcții și reparații echipamente de telecomunicații București.

Prezentarea programelor ce alcătuiesc softul utilitar al sistemului este efectuată în anexă.

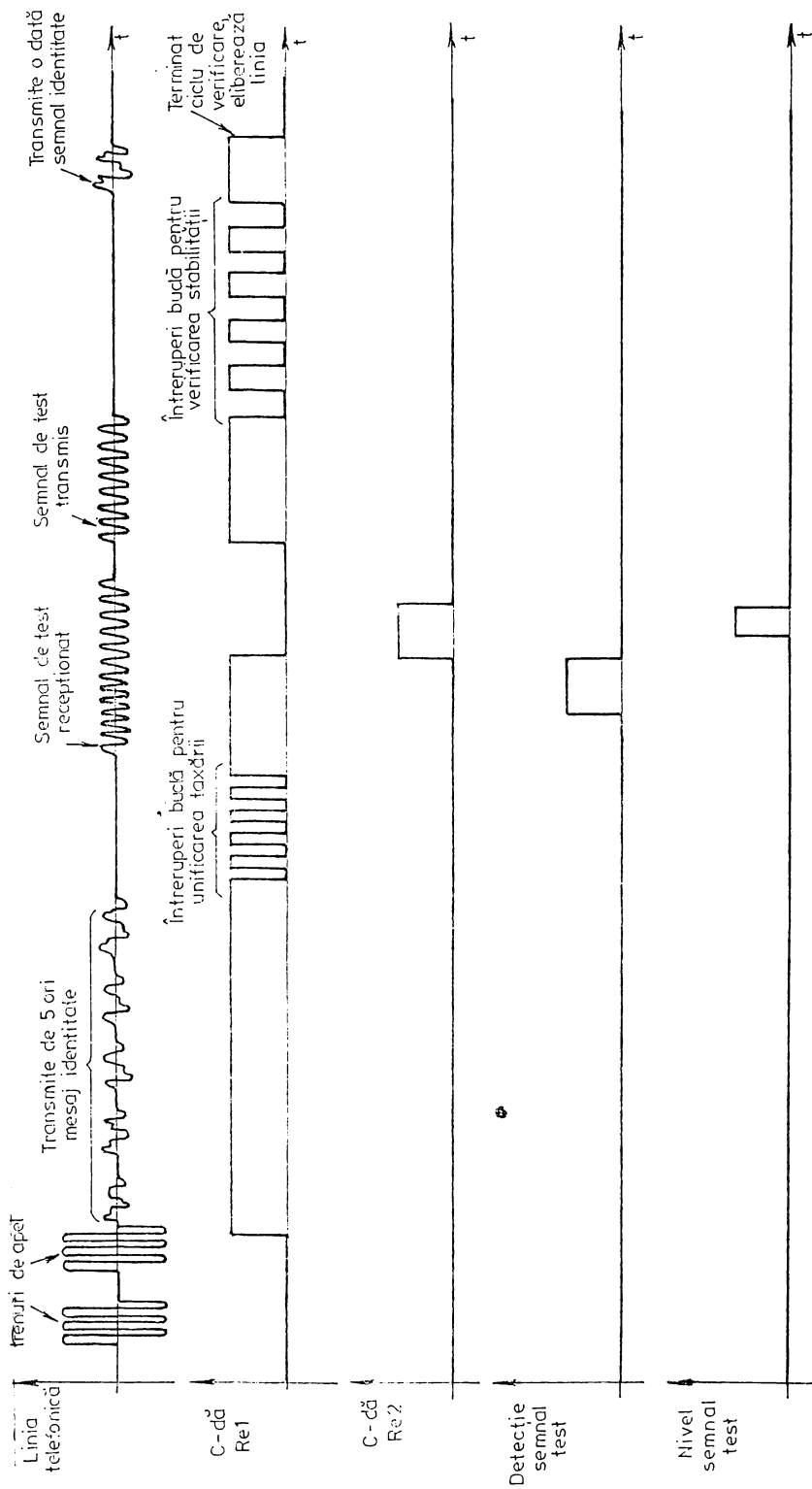


FIG. 4.11. Diagrama de timp a dispozitivului de prezentare a denumirii centralilor telefonice automate

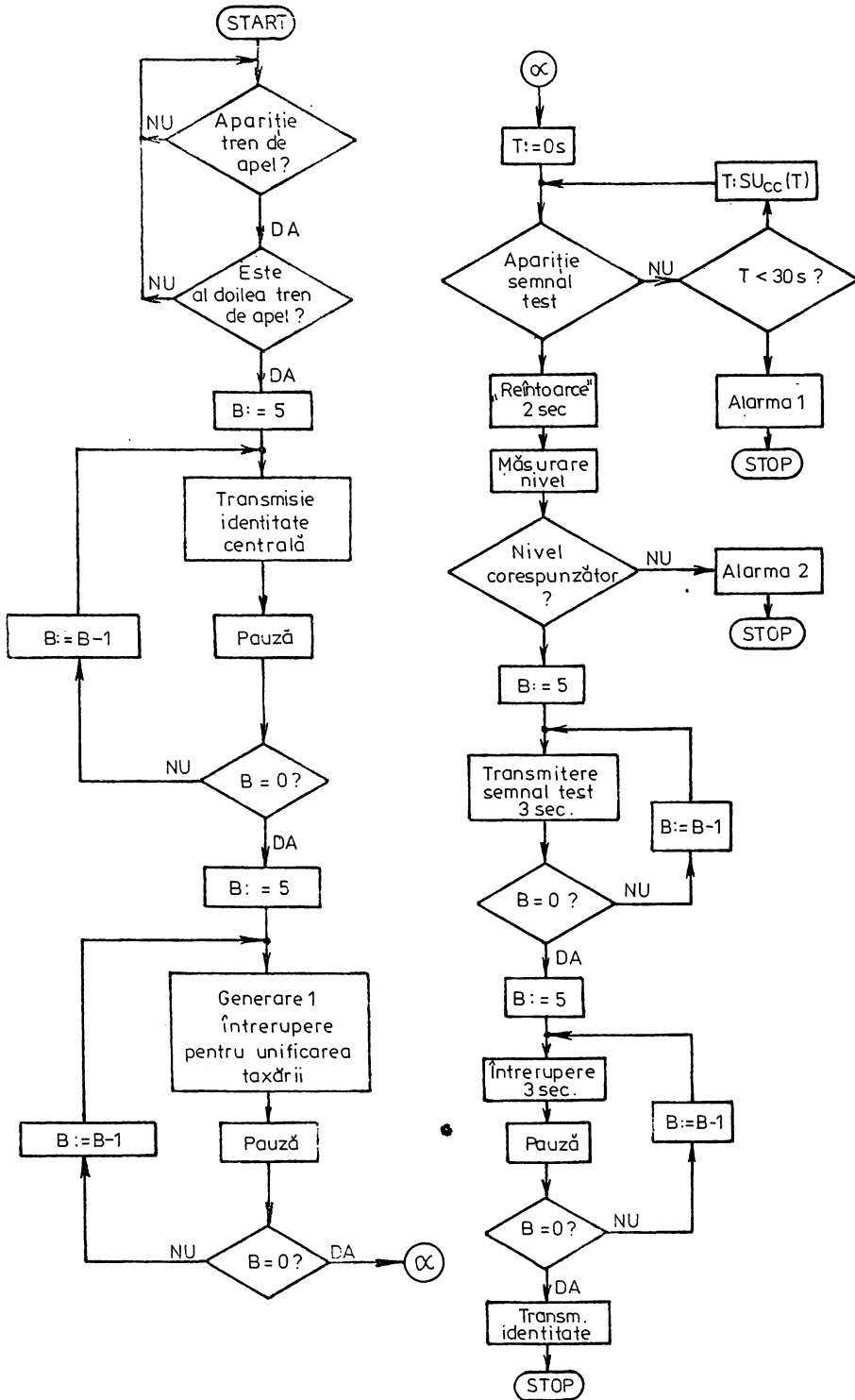


FIG.4.12.Diagrama de funcționare a dispozitivului de prezentare a denumirii centrelor telefonice automate

4.3.2.3. Sistem robot pentru transmiterea prin telefon a orei exacte /64/,/65/,/66/

4.3.2.3.1. Prezentare generală

Sistemul a fost proiectat și realizat pe baza unui contract de cercetare, avînd ca beneficiar Direcția generală de poștă și telecomunicații București, pentru a înlocui actuala instalație electromecanică, de fabricație Siemens, aflată în exploatare în cadrul rețelei noastre telefonice și amplasată în cadrul Direcției de telecomunicații a municipiului București, de unde semnalul de oră exactă este transmis, prin circuite interurbane rezervate, în toate reședințele de județ. Un prim model experimental a fost realizat încă în cursul anului 1985, pe baza asocierii unui calculator personal de tip ZX-81 /53/ și care a fost brevetat /64/.

Ulterior, pe baza acestui model, s-a realizat un sistem dedicat, cu logică programată și sinteză de vorbire, bazat pe microprocesorul Z80 /65/,/66/. Schema bloc generală a acestui robot este reprezentată în figura 4.13.

Unitatea de comandă și control este realizată de microprocesorul Z80, iar declanșarea momentelor de generare a mesajului de oră exactă se face din lo în lo secunde, sub controlul unui ceas cu afișaj digital, realizat prin intermediul circuitului temporizator programabil Z80 CTC.

Circuitul Z80 CTC solicită întreruperi microprocesorului Z80 la fiecare lo secunde, declanșînd astfel procedura de generare a mesajului.

Ceasul sistemului este sincronizat cu un cuarț exterior, fiind programabil, prin intermediul unei logici de programare, realizată cu comutatoare, la nivel de oră, minute și secunde, similar ca un ceas electronic digital, de mîină.

Rutina de programare a ceasului intră în funcțiune imediat după acționarea comutatorului de inițializare a sistemului (RESET). Sistemul transmite un mesaj de forma :

"La semnalul următor va fi ora HH, MM minute, SS secunde"
Momentul exact al orei anunțate este marcat de un semnal sonor.

Sinteza se realizează prin metoda codificării formei semnalului, pe trei nivele de înregistrare : cuvinte, morfeme și difoni. Astfel, textul fix, conținut în mesajul de oră exactă, este înregistrat la nivel de cuvinte, iar pentru sinteza cifrelor, ce indică orele, minutele și secunde, s-a procedat la o analiză

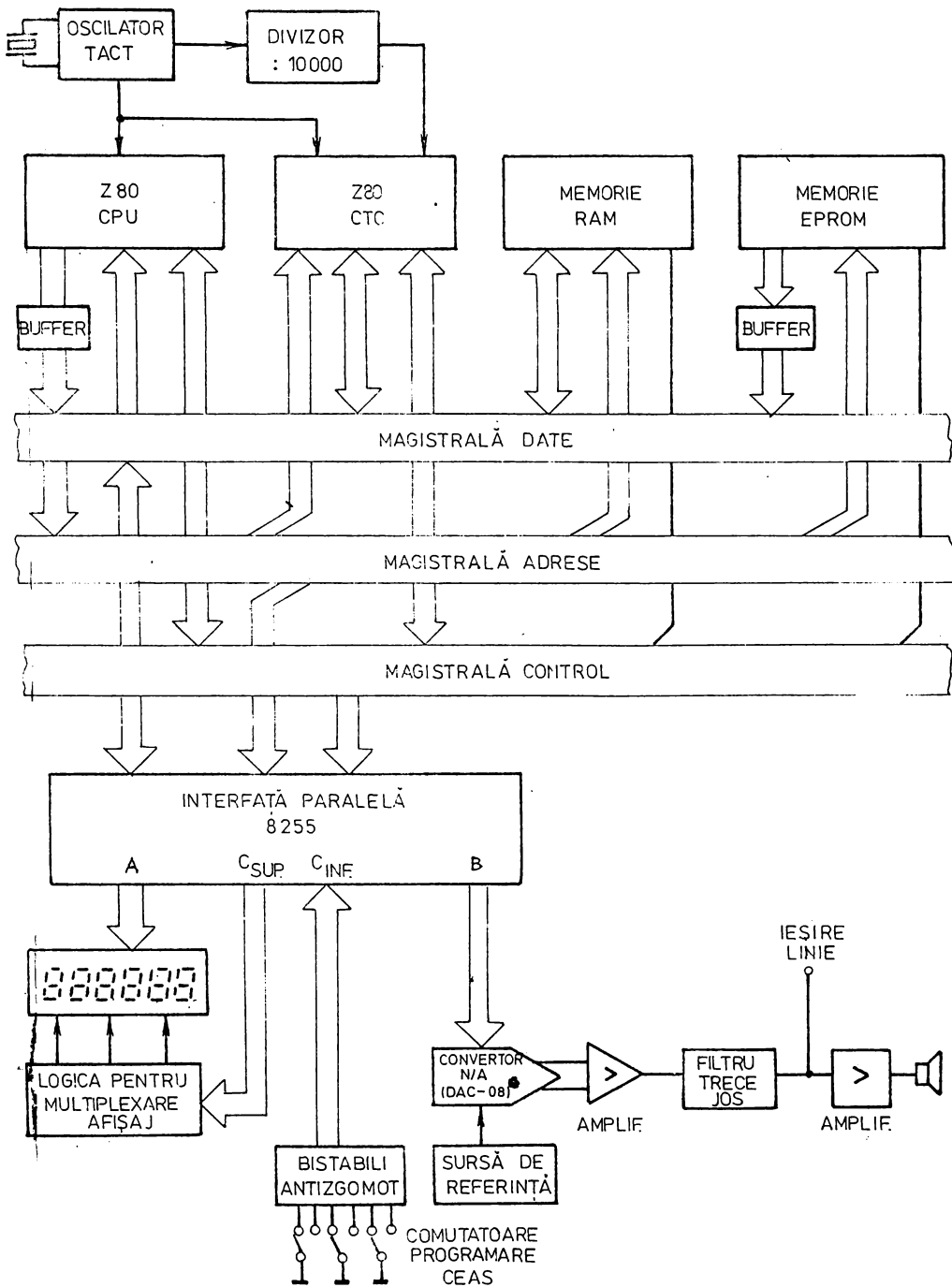


FIG.4.13.Schema bloc generală a robotului telefonic pentru transmiterea orei exacte

și selecție de morfeme și difoni.

Analiza s-a efectuat cu sistemul de analiză descris la capitolul 2.

Informația de bază a fost înregistrată într-o memorie EPROM de dimensiune 46 KO, realizată cu capsule de memorie de tip 2716.

La înregistrarea informației de bază s-a utilizat parțial și metoda de compresie expusă în paragraful 3.5.2. Aplicarea globală a acestei metode de compresie nu a fost posibilă, datorită condițiilor impuse de beneficiar, referitor la sintetizarea unei vorbiri cât mai naturale, asemănătoare unui anumit tip de voce (crainic feminin). Înregistrarea informației de bază a utilizat codificarea rezultată prin conversia analogico-digitală de tip MIC, cu cuantizare pe 8 biți, combinată cu eliminarea pauzelor, compresia logaritmică, frecvența de eșantionare variabilă, funcția de spectrul fonemelor, și înregistrarea unor foneme numai prin intermediul unei fracțiuni, repetabilă de un număr de ori.

Generarea mesajului se bazează pe existența unui tabel de corespondențe timp-adrese, pe care microprocesorul îl consultă la fiecare anunțare. Este mai întâi alcătuit mesajul, într-o zonă tampon a memoriei RAM, prin introducerea secvențială, în această zonă, a segmentelor ce formează mesajul următor. Fiecărui segment extras din memoria i se realizează și o decompresie, dacă a fost comprimat la înregistrare, acest fapt rezultând tot din tabelul de corespondențe. La alcătuirea cuvintelor din segmente, se mai aplică și o procedură de interpolare a nivelelor de început și sfârșit a segmentelor, dacă există diferențe semnificative între acestea.

Urmează apoi transferul mesajului, astfel format, unui port de ieșire, ce-l aplică unui convertor digital-analogic (DAC OS), care-l transformă în mesaj analogic și o interfață analogică, prevăzută cu un difuzor de control și cu o ieșire pentru linia telefonică. Comanda afișajului digital, programarea ceșului și conectarea convertorului numeric-analogic s-au efectuat prin intermediul unui circuit de interfață paralelă de tip 8255.

S-a ales acest circuit, nespecific familiei microprocesorului Z80, în locul unui circuit similar Z80-PIO, în ciuda unor dificultăți de adaptare, deoarece dispune de trei porturi, care au putut realiza conectarea celor trei subansamble. Cum circuitul

Z80-PIO dispune de numai două porturi, ar fi fost nevoie de două astfel de circuite pentru a satisface aceste cerințe.

În realizarea robotului de oră exactă, s-au avut în vedere și condițiile impuse acestui sistem de necesitatea de a funcționa continuu /66/, de a fi prevăzut cu un regim de rezervare și de a fi sincronizat central, la nivel de oră exactă, printr-un semnal de sincronizare aplicat din exterior.

Programele ce asigură funcționarea robotului au fost scrise în limbaj de asamblare Z80 și ocupă o dimensiune de memorie ce nu depășește 4 KO.

Descrierea lor, la nivel de algoritmi, precum și listingu-rile propriuzise, sînt incluse în anexă. Sub denumirea de DATOREX (Dispozitiv de anunțare prin telefon a orei exacte), sistemul prezentat a fost omologat în vara anului 1989 și a intrat în fabricație de serie la Întreprinderea de construcții și reparații echipamente de telecomunicații București. O copie a actului de omologare este inclus în anexă.

Mai trebuie menționat și faptul că, în proiectarea acestui echipament, s-a ținut cont și de posibilitatea de a se realiza o sinteză într-o altă limbă, fără modificări de hardware, ci numai la nivel software, iar în momentul de față este realizată deja o versiune software pentru sinteza în limba germană.

4.3.2.3.2. Descrierea funcțională a robotului telefonic pentru oră exactă

Funcționarea robotului se poate separa în două secțiuni relativ independente:

- funcționarea ceasului și a sistemului de afișare;
- generarea secvenței vorbite.

Sincronizarea ceasului poate fi făcută intern, de la oscilatorul cu cuarț, sau extern, cu un alt semnal de sincronizare.

Regimul normal de funcționare este asigurat de întreruperile generate de cele 4 secțiuni ale circuitului Z80-CTC, cu funcție de numărător-temporizator, ce primește ca semnal de tact semnalul de la oscilatorul cu cuarț, sau un semnal extern de sincronizare.

Circuitul Z80-CTC este programat să genereze impulsuri pentru a realiza patru tipuri de întreruperi, necesare funcționării, în următoarea ordine de prioritate:

- la interval de 1 secundă, ce constituie baza de timp pentru ceas;

- la aproximativ 30 milisecunde, utilizate pentru înprospătarea afișării multiplexate a orei;

- la 150 - 200 microsecunde, ce constituie semnalul de tact pentru redarea eșantioanelor ce alcătuiesc mesajul de oră exactă;

- la 10 secunde, ce constituie comanda pentru începerea generării mesajului de oră exactă.

Secvența numerică de eșantioane este transformată în semnal analogic de lanțul de circuite: interfața 8255, circuitul convertor numeric-analogic DAC 08, filtrul trece jos (format din 2 pînă la 4 celule de filtrare de ordinul 2, cu frecvența de tăiere de 3,6 kHz) și amplificatorul de putere realizat cu circuitul TBA 810, conform schemei bloc din fig.4.13.

Inițializarea sistemului presupune poziționarea ceasului pe o anumită oră de pornire și se realizează prin intermediul a trei comutatoare, a căror comenzi sînt preluate prin intermediul portului C al circuitului de interfață paralel 8255, după ce, în prealabil, au fost trecute printr-un set de circuite de eliminare a zgomotului electric, provocat de vibrația comutatoarelor, realizat cu circuite basculante bistabile de tip R-S. Afișarea orei se face cu 3 celule de afișaj, de cîte 2 digiți fiecare, multiplexate prin controlarea, comandată prin program, a tensiunii aplicate pe anozii fiecărei capsule, tot prin intermediul portului C al circuitului de interfață 8255.

Magistralele microsistemului sînt bufferate cu circuite de tip CDB 407, pentru magistrale de adrese și amplificatoare de magistrală, de tip 8216, pentru magistrala de date.

La conectarea alimentării, sau la acționarea butonului RESET, prin program, se anulează conținutul numărătorului de ore, minute și secunde, conectîndu-se o rutină de citire a stării comutatoarelor de programare a orei exacte.

Pornirea efectivă a ceasului, deci și a întreruperilor, are loc după programarea orei exacte și trecerea comutatorului de funcțiuni pe poziție de start.

Sincronizarea cu ora exactă se mai poate face și automat, prin intermediul unui impuls de sincronizare, interpretat ca o comandă de poziționare a ceasului la ora 0, 0 minute și 0 secunde, și acceptat de microsistem ca o întrerupere nemascabilă.

Detectarea apoi a unui număr de secunde multiplu de zece (0, 10, 20, 30, 40, 50) va avea, ca efect, generarea întreruperii

ce va declanșa procesul de formare a mesajului de anunțare a orai exacte și generarea acestuia sub formă de vorbire artificială, în mai multe etape. În prima etapă, se generează un semnal sonor, pentru marcarea exactă a orai anunțate anterior. În continuare, se calculează ora ce va fi actuală peste 10 secunde și pe baza acestui calcul se formează, în memoria RAM, un tabel cu adresele de început și lungimea secvențelor, extrase din memoria EPROM, ce vor alcătui segmentele cu care se construiește mesajul respectiv.

Fiecare segment este prevăzut și cu un cod care specifică dacă, la înregistrare, segmentul respectiv a fost supus unei compresii, deci dacă, la redare, trebuie efectuată, sau nu, operația inversă.

Apoi se generează întreruperile de 150-200 μsecunde (funcție de lungimea segmentelor) prin care segmentele a căror adrese figurează în tabel sînt extrase și trimise sistemului de conversie numeric-analogică, pentru generarea mesajului vorbit.

S-a preferat această metodă de generare unuia anterioare, ce forma, mai întîi, întregul mesaj în memoria RAM (nu doar un tabel de adrese) și care era apoi transferat blocului de conversie numeric-analogic, deoarece necesită mai puțină memorie RAM, deși presupune complicații de programare, impuse de introducerea unor intervale de așteptare variabile, datorate lungimii diferite a mesajelor.

4.3.2.3.3. Tehnici de testare implementate prin program la robotul telefonic de oră exactă

Avînd în vedere faptul că robotul de ceas reprezintă un echipament ce urmează a funcționa permanent, adică 24 de ore din 24, au fost realizate și puse la punct proceduri de testare, care să poată intra în funcțiune atît la cerere, cît și în mod automat, folosind perioadele în care anumite circuite de bază (memorie, porturi, etc.) nu sînt utilizate.

Cu sistemul dezactivat se pot testa memoria RAM, circuitul de interfață și funcționarea sistemului de întreruperi, adică funcționarea circuitului Z80-CTC.

Astfel, pentru funcționarea robotului, este strict necesară o zonă minimală de memorie RAM validă. În scopul de a testa existența unei astfel de zone, se testează primii 100 octeți ai memoriei RAM, conform algoritmului prezentat în fig.4.14. Dacă se identifică un octet defect, în sensul că nu poate fi înscris cu o

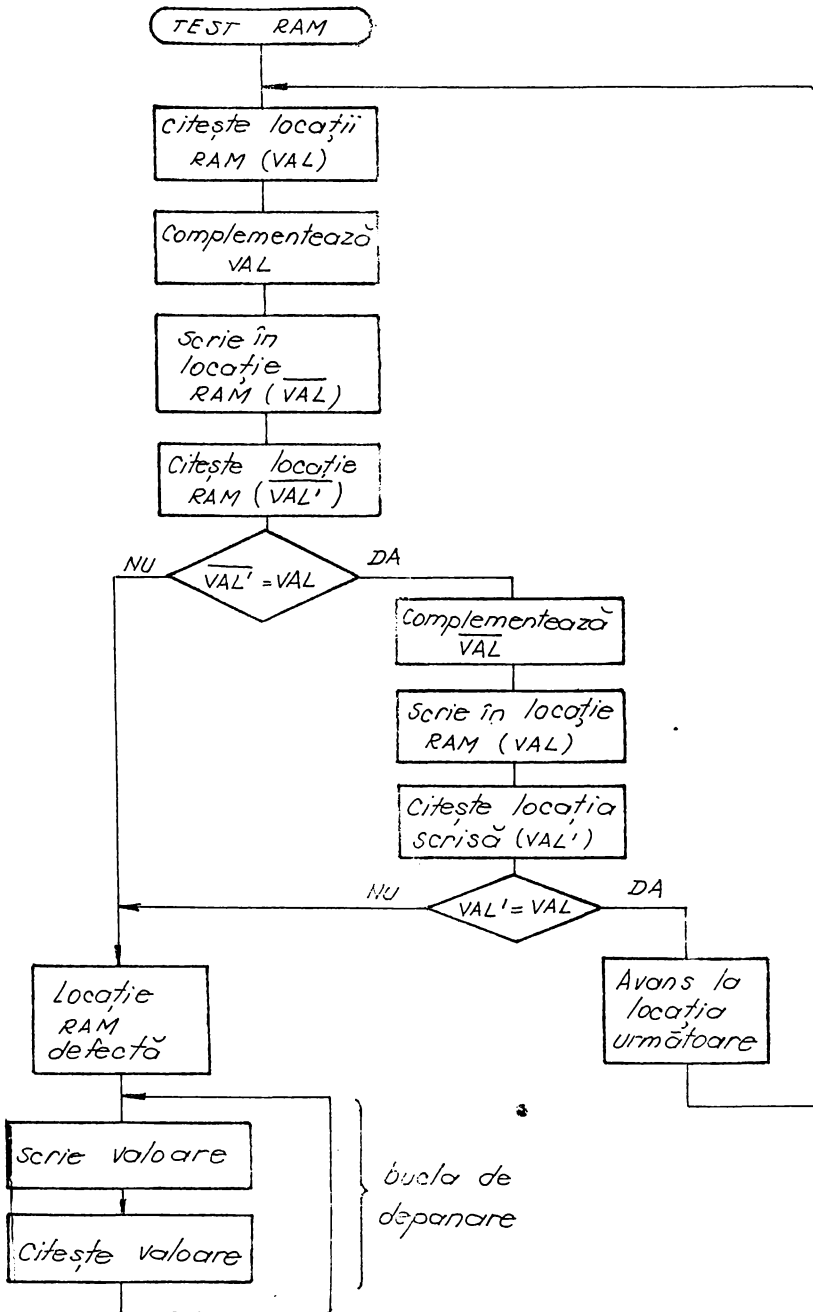


Figura 4.14. Algoritm de realizare a testului memoriei RAM.

anumită informație, și citit, se afișează pe cele trei celule de afișaj numărul de octeți valizi identificați în zona testată. Lungimea de loc octeți este limitată de numărul de digiți de care dispune sistemul de afișaj.

Pentru realizarea acestui test se utilizează numai microprocesorul, deoarece sistemul de întreruperi, bazat pe circuitul Z80-CTC, este întrerupt, urmînd a fi testat și el. Dacă se identifică loc octeți valizi, se consideră că memoria RAM este funcțională și se trece la testarea sistemului de întreruperi și a circuitului Z80-CTC.

Pentru a realiza acest test se plasează adresa de formare a memoriei stivă, specifică funcționării în regim de întreruperi, la începutul zonei de loc octeți a memoriei RAM, testată anterior.

Se inițializează apoi sistemul de întreruperi și se testează pe rînd cele 4 canale ale circuitului Z80-CTC.

Testarea se realizează în modul temporizator, prevăzut cu întreruperi, unul din modulele specifice de lucru ale acestui tip de circuit /64/.

Pentru comparație, se execută numărarea și cu microprocesorul. Etapele sînt următoarele:

- se poziționează adresele rutinelor de întrerupere pentru toate canalele pe o rutină de afișare a erorii;
- se lansează operația de numărare de către CTC și de către microprocesor;
- dacă intervine o întrerupere înainte de timpul programat, se activează rutina de afișare a erorii depistate, cauze fiind sau faptul că sistemul de întreruperi este invalid, sau că circuitul Z80-CTC numără incorect;
- rutina de eroare afișează la sistemul de afișaj numărul canalului defect.

Dacă și acest test este verificat și nu sînt erori, se inițializează sistemul de întreruperi, dar nu se trece încă la demararea funcționării întregului sistem, înainte de a se testa și validitatea restului de memorie RAM, în afara celor loc octeți testați inițial, cu același test.

Dacă nu este semnalată nici o eroare, se consideră că sistemul este valid, se mută adresa memoriei stivă la cea prevăzută în programul de funcționare, se fac inițializările curente și se demarează programul de bază pentru programarea orei de referință.

Pentru verificarea stării sistemului și în situația în care acesta este activ, deci funcționează, au fost prevăzute teste "on

lina". Se pot verifica în acest mod memoria RAM și memoria EPROM.

Testul memoriei RAM este similar celui realizat în cazul în care sistemul este dezactivat, afișarea efectuându-se alternativ și multiplexat cu cea a orei exacte.

Testul memoriei EPROM constă în efectuarea sumei modulo 256 pentru conținutul fiecărei capsule de memorie EPROM, de tip 2716, ce conține segmentele de vorbire și compararea ei cu valoarea martor, memorată în cadrul capsulei EPROM ce conține și softul de bază și care este netestabilă. Defectarea acestei capsule este însă ușor sesizabilă prin nefuncționarea completă a sistemului.

La depistarea unei neconcordanțe, se afișează numărul capsulei de memorie EPROM respective și se așteaptă o comandă, de la panou, de continuare, sau de abandonare a acestui test.

Efectuarea testării se desfășoară fără perturbarea funcționării robotului, fiind semnalizată prin stingerea afișajului pe durata acestei testări.

Algoritmul după care se face această testare este prezentat în fig.4.15.

Testele prezentate au fost astfel concepute încât să permită și o depanare rapidă a sistemului.

Astfel, testul memoriei RAM, în cazul semnalării unei erori, se oprește într-o buclă infinită, care citește și scrie succesiv, în octetul depistat ca defect, pentru a facilita depanarea cu osciloscopul, iar oprirea testului întreruperilor se face într-o stare de tip "halt", cu dezactivarea întreruperilor, pentru a se putea determina starea lanțului de întreruperi în care s-a produs eroarea.

4.3.2.3.4. Concepția tehnologică și concluzii cu privire la introducerea în fabricație de serie a robotului telefonic pentru oră exactă /67/

Întregul sistem a fost amplasat pe trei plăci distincte și anume:

- unitatea centrală, incluzând microprocesorul, memorie, generatorul de tact, temporizatorul și circuitul de interfață;

- unitatea de conversie numeric-analogică, incluzând convertorul numeric-analogic, filtrele și amplificatorul de audio-frecvență și interfața cu linia telefonică;

- unitatea de comandă și afișaj, incluzând partea de comandă a afișajului, precum și de programare a orei exacte.

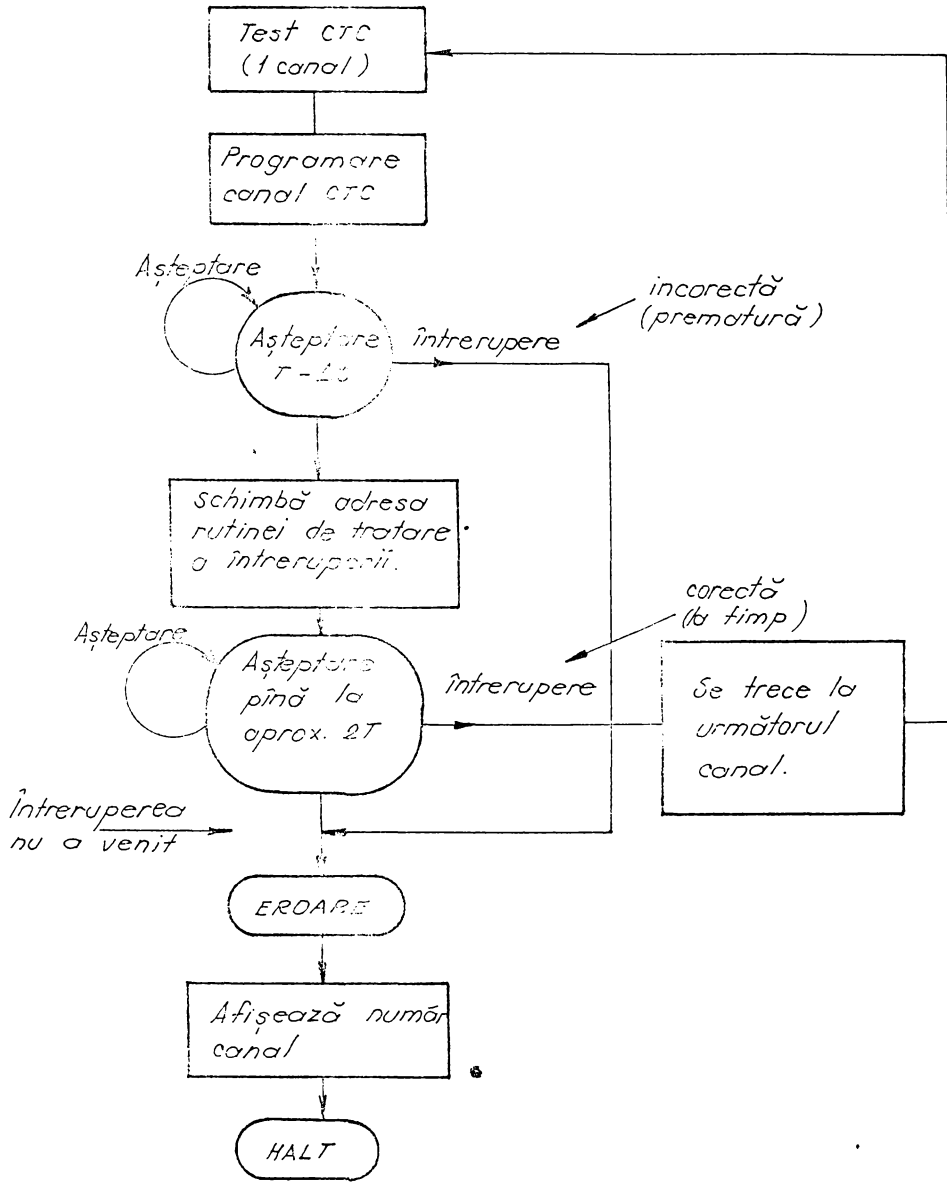


Figura 4.15. Algoritm de realizare a testului de verificare a funcționării circuitului Z80-CTC.

Această separare s-a făcut atât datorită tipurilor distincte de prelucrare pe care le efectuează unitățile respective, ce implică și teste specifice în procesul fabricației de serie, precum și modului diferit în care trebuie amplasate în structura aparatului, atât din motive constructive, cât și pentru evitarea introducerii de perturbații.

Robotul telefonic pentru prezentarea orei exacte, ca sistem cu vorbire artificială, reprezintă primul sistem de acest fel, cu un vocabular extins (70 cuvinte) realizat pentru limba română și destinat unei funcții de informare publică.

Introducerea acestui aparat în exploatare curentă poate aduce venituri de ordinul zecilor de milioane de lei anual Ministerului Poștelor și Telecomunicațiilor, contribuind totodată la sporirea confortului telefonic al abonaților și deschizând calea realizării unor sisteme similare, cu funcție informativă, și în alte domenii, ce pot conduce la o creștere semnificativă a productivității muncii în domeniul telecomunicațiilor, la noi în țară.

5. CONSIDERATII FINALE

În dezvoltarea acestei teze de doctorat, avînd ca temă: Sisteme cu vorbire artificială pentru telecomunicații, s-a pornit de la considerentul că, în cadrul tezei, trebuie să-și găsească locul o prezentare a metodelor de sinteză a vorbirii, care motivează tema, o prezentare a sistemelor cu logică programată pentru analiza vorbirii, care condiționează sinteza, o prezentare a sintetizatoarelor bazate pe logică programată, ce constituie obiectivul cercetării, precum și aplicații ale acestora în telecomunicații, ajungîndu-se astfel la cele patru capitole de bază ale tezei.

În capitolul I se prezintă o sistematizare a metodelor de sinteză a vorbirii, efectuîndu-se o clasificare a acestor metode în două categorii de bază, funcție de modul în care se realizează codificarea semnalului vorbit analizat și, în consecință, sinteza. Cele două categorii au ca obiect refacerea sursei generatoare a acestui semnal (echivalentă simulării tractului vocal), respectiv a formei finale a semnalului (echivalentă simulării undei acustice). Fiecare dintre acestea este apoi dezvoltată, prezentîndu-se, sub forma unor diviziuni, diferitele variante de sinteză care pot fi integrate în categoria respectivă, evaluîndu-se, la fiecare, avantajele, dezavantajele, precum și perspectivele de dezvoltare viitoare.

Astfel, din categoria metodelor de sinteză bazate pe codificarea sursei de semnal, se prezintă sinteza fonemică, sinteza prin refacerea înfășurătorii spectrale, sinteza prin formanți și sinteza prin predicție liniară.

O atenție aparte este acordată metodei de sinteză prin refacerea înfășurătorii spectrale, cît și sintezei bazată pe predicție liniară, datorită, în principal, ariei largi de răspîndire a acestor metode, cît și pentru că, mai ales ultima, în diferitele sale variante, a cunoscut o implementare masivă prin intermediul logicii programate, fie sub forma unor sisteme de procesare digitală dedicate, fie sub cea a unor circuite integrate pe scară largă, specializate în această funcție.

Au fost prezentate, în detaliu, două din variantele metodei de sinteză prin predicție liniară și anume cea în care determinarea coeficienților de predicție se realizează prin metoda corelației, precum și cea în care aceeași determinare are la bază metoda rețelelor, cu celule în X (lattice), cele două metode fiind și cele mai larg răspândite la ora actuală. La cea de-a doua categorie de metode de sinteză, referitoare la codificarea formei semnalului, s-au prezentat gradele de detaliere posibile ale analizei și, în consecință, ale sintezei respective, de la nivel de alofon, morfem, difonem și pînă la cel de cuvînt.

Prezentarea s-a efectuat, în acest caz, și prin prisma utilizării sistemelor cu logică programată, dedicate acestui scop, evidențiindu-se problemele specifice care apar și care se cer rezolvate.

În cadrul capitolului se revendică unele contribuții originale în sistematizarea și alegerea criteriilor de clasificare a metodelor de sinteză a vorbirii și implicit a sintetizoarelor de vorbire, care nu apar astfel prezentate, ca un tot unitar, în bibliografia de profil.

Tot în capitolul 1 se prezintă și cîteva detalii tehnice și performanțe ale unui sintetizor cu refacerea înfășurătorii spectrale, realizat de autor, cu care s-a putut efectua sinteza oricărui cuvînt al limbii române. Alegerea benzilor de trecere a filtrelor sintetizorului s-a făcut în urma unei analize a formelor limbii române, ținînd cont de frecvența lor statistică de apariție, astfel încît, filtrele să poată transmite, cu prioritate și cît mai fidel, formații celor mai utilizate foneme, fapt ce constituie un element de noutate printre cele cîteva încercări de sinteză, de acest tip, realizate pentru limba română.

În capitolul 2, dedicat sistemelor cu logică programată pentru analiza vorbirii, se prezintă obiectivele pe care sintezele impune analizei vorbirii, privind, în principal, determinarea informației de bază necesară sintezei și compresia acesteia, pentru a reduce volumul de memorie necesar înregistrării ei. Se prezintă apoi principalele metode numerice de prelucrare digitală a semnalului vorbit, cu accent pe transformata Fourier rapidă, în variantele ei cu decimare în timp și, respectiv, în frecvență, precum și ponderarea semnalului analizat prin funcții de tip fereastră, cu scopul de a se realiza un anumit tip de microprocesor, aplicabil la maximum facilitățile acestuia.

În scopul efectuării unei analize complexe a semnalului vorbit, pentru a putea determina caracteristicile fonemelor limbii române și a putea astfel stabili condițiile ce se cer rezolvate pentru efectuarea unei sinteze eficiente, s-a realizat o instalație complexă de analiză, bazată pe logică programată.

Această instalație constă dintr-un sistem complet de achiziție și prelucrare digitală a semnalului vorbit, cu posibilități de stocare a unui volum considerabil de semnal, care urmează a fi supus analizei, precum și cu posibilitatea vizualizării formelor de undă inițiale, sau rezultate din diversele faze ale prelucrării.

Descrierea părții hard a acestei instalații face, de asemenea, obiectul capitoului dai, în timp ce prezentarea, în extenso, a setului de programe de prelucrare elaborat este cuprinsă în anexă. Realizarea unei memorii rapide de dimensiuni mari (RAM-DISC), organizată sub formă unei unități de disc virtual, constituie o soluție inovativă, care permite sistemului o achiziție rapidă a semnalului, utilizând procedurile specializate pentru tratarea unităților de disc flexibil, de care dispune orice sistem de operare dezvoltat pentru microcalculatoare. Sistemul de analiză menționat reprezintă, la ora actuală, unul din cele mai performante sisteme de acest tip, realizate în țară, permițând achiziționarea și prelucrarea chiar și a altor tipuri de semnale, precum cele biologice, de vibrații etc.

Setul de programe include un sistem complet de rutine pentru achiziționarea, conversie analog-numerică, prelucrarea digitală, vizualizarea grafică, conversie numeric-analogică și restituirea semnalelor.

Modul de apelare al rutinelor este facilitat de prezentarea întregului set sub forma unui meniu și e unei tehnici de apelare conversațională.

Componenta soft, dedicată efectiv prelucrării digitale, conține rutine specializate pentru principalele tehnici de prelucrare numerică a semnalelor, cum sînt: transformata Fourier rapidă, directă și inversă, predicția liniară, funcția de corelație și autocorelație etc.

Programele respective au fost concepute și realizate pentru a fi cât mai performante, mai ales prîn reducerea vitezei de calcul și utilizarea, cu precădere, a limbajelor de asamblare, singurele în măsură să asigure viteza maximă de lucru pentru un anumit tip de microprocesor, explicatînd la maximum facilitățile acestuia.

S-au obținut astfel spectrele și formele de undă ale tuturor fonemelor limbii române, atât în pronunție izolată, cât și legată, iar ca un element inedit încă pentru limba română, s-a pus la punct și o rutină prin care s-a obținut imaginea acestor foneme într-un spațiu tridimensional, sub forma unor suprafețe spectrale, introduse în anexă.

În capitolul 3 se face o prezentare a sistemelor de sinteză a vorbirii, bazate pe implementarea logicii programate.

Prezentarea urmărește clasificarea de bază, descrisă în primul capitol, adică sinteza prin codificarea (refacerea) sursei și, respectiv, prin codificarea formei semnalului.

Sînt analizate problemele principale care se pun în proiectarea și realizarea unui sintetizor de vorbire, bazat pe logică programată, în fiecare din cele două cazuri, făcîndu-se referiri la circuite integrate specializate pentru sinteza vorbirii, sau la procesoare de semnal utilizabile în acest scop.

Ca și contribuție personală, în acest domeniu, este prezentată o metodă de compresie și condifiere a semnalului vorbit, bazată pe analiza vorbirii, care permite o suplimentare cu peste 30% a cantității de informație de bază ce poate fi înregistrată în memoria maxim adresabilă direct a unui microprocesor pe 8 biți.

Inedită este și realizarea unui sintetizor bazat pe refacerea formei semnalului, avînd la bază un calculator personal. Sintetizorul este programabil, fapt ce a permis dezvoltarea a două aplicații, una destinată roboților industriali și cealaltă destinată supravegherii funcționării unui reaktor chimic, menționate în cadrul aceluiași capitol.

Capitolul 4 este destinat prezentării unor aplicații ale sintezei vorbirii în domeniul telecomunicațiilor și, mai precis, în cel al rețelei telefonice. Sînt identificate mai întîi o serie de aplicații privind realizarea unor sisteme de supraveghere, sau de informare, concepute sub forma unor servicii speciale ce pot fi oferite abonaților telefonici, ca o cale importantă de creștere a productivității muncii în domeniul telecomunicațiilor, precum și a creșterii gradului de confort oferit abonaților.

Se procedează apoi la prezentarea modului de realizare a unor astfel de aplicații în alte țări, urmată de prezentarea contribuțiilor originale ale autorului în acest domeniu, privind concepția și realizarea unor sisteme complexe, bazate pe vorbire și aplicații privind analiza unor semnale analogice (exemplu: ritmul, respirația etc.):

artificială, care realizează automat serviciile de informare, supraveghere, verificare și testare în cadrul rețelei telefonice.

Sînt prezentate detaliat, în acest capitol, un robot telefonic pentru transmiterea automată a orei exacte, un sistem de tip flashing, precum și un sistem complex destinat identificării și testării centralelor telefonice automate.

Atît partea hard, cît și cea soft, a acestor echipamente includ soluții tehnice originale, precum realizarea sursei de excitație pentru sinteză, compresia informației de bază, modul de formare al mesajului de oră exactă, întreaga concepție de testare axată pe utilizarea instalației de tip flashing etc., certificate prin obținerea unor brevete de invenții.

Dintre echipamentele bazate pe vorbire artificială, prezentate și realizate de autor, robotul telefonic pentru ora exactă și sistemul de identificare și testare a centralelor telefonice automate au făcut obiectul unor contracte de cercetare finanțate de Ministerul Poștelor și Telecomunicațiilor și au fost preluate pentru producția de serie de către întreprinderi specializate.

S-a considerat, de asemenea, importantă și reliefaarea anumitor concluzii privitoare la acest domeniu, ce se desprind prin optica considerentelor legate de cazul specific al stadiului de dezvoltare a industriei autohtone, precum și a caracteristicilor limbii române.

Dintre acestea, se vor mai reaminti, la acest capitol următoarele:

- nu este posibilă realizarea unor procesoare specializate în sinteza vorbirii, prin metode cu codificarea sursei, fără tehnologia necesară realizării procesoarelor de semnal de uz general;

- implementarea unui procesor integrat specializat în sinteza vorbirii, pentru limba română, implică, de regulă, o colaborare tehnologică cu firma producătoare;

- nu este posibilă realizarea unui sintetizor pentru o anumită limbă, decît prin efectuarea prealabilă a unei analize a cuvintelor limbii respective, în scopul obținerii informației de bază;

- analiza vorbirii, în scopul sintezei, este o operație complexă și necesită un sistem de analiză complicat, bazat pe logică programată, sistem ce poate fi însă utilizat și în alte aplicații privind analiza unor semnale analogice (exemplu: vibrații, zgomote etc.);

- în condițiile tehnologice în care nu este posibilă, pentru realizarea sintetizoarelor de vorbire, decât implementarea unor microprocesoare de uz general, pe 8 biți, sinteza prin metode cu codificarea formei de semnal este singura posibilă;

- utilizarea vorbirii artificiale în rețeaua telefonică permite realizarea unor servicii noi, specifice, în general, numai rețelelor telefonice prevăzute cu integrarea serviciilor, de care dispun, la ora actuală, numai țările cu tehnologie foarte avansată;

- este posibilă și realizarea unei aparaturi complexe, prevăzute și cu vorbire artificială, pentru testarea obiectivă a echipamentelor de telecomunicații, cu implicații importante în ridicarea gradului de funcționalitate a rețelei telefonice;

- realizarea unor echipamente pentru telecomunicații, bazate pe vorbire artificială și programabile, permite adaptarea lor și pentru alte limbi, decât limba română, putând sta astfel la baza unor oferte pentru export din partea industriei de telecomunicații autohtone;

- implementarea sintetizoarelor de vorbire în telecomunicații și, în special, în rețeaua telefonică, reprezintă o sursă importantă de creștere a productivității muncii în acest sector, prin introducerea unor servicii automate taxabile, precum și a sporirii gradului de confort oferit abonaților telefonici.

BIBLIOGRAFIE

1. K.Sickart, Automatische Sprecheingabe und Sprachausgabe, Markt & Technik Verlag, Haar bei München, 1983.
2. L.Rabiner, R.Schafer, Digital processing of Speech Signals, Prentice Hall Inc. Englewood Cliffs, New Jersey, 1978.
3. J.L.Flanegan, Speech Analysis, Synthesis and Perception, Springer Verlag, 1972, vol.I și II.
4. A.Cancaea, Analiza și sinteza vorbirii, Editura militară, 1976.
5. L.Rabiner, Speech Synthesis by Rule : An Acoustic Domain Approach, Bell System Technical Journal, January, 1988.
6. L.Rabiner, A Model For Synthesizing Speech by Rule, IEEE Transactions on Audio and Electroacoustics. AU 17, March 1969.
7. L.Fortuna, L.Mărgulescu, Sintetizor analogic pentru sinteza vorbirii, în volumul: "Lucrările simpozionului: Contribuții la dezvoltarea aparaturii electronice medicale", ediția 3-a Timișoara, Institutul politehnic "Traian Vuia", 1982 "LSCDAEM - 82".
8. L.Fortuna, L.Mărgulescu, Analizor analogic pentru analiza și sinteza vorbirii, în volumul "LSCDAEM-02" (citată anterior).
9. L.Fortuna, Metode de sinteză a vorbirii, Referat pentru doctorat, Institutul Politehnic București, Facultatea Electronică și Telecomunicații.
10. D.W.Griffin, J.S.Lim, Multiband Excitation Vocoder, IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 36, Nr.8, August 1988.
11. B.Dancea, Schaltkreise zum Erzeugen und Erkennen von Sprachen. Funk und Technik, 36, Heft 4, 1981.
12. C.Radu, G.Săndulescu, Filtre numerice. Aplicații. Editura tehnică, 1979.
13. G.Stolojanu, V.Podaru, F.Cetină, Prelucarea numerică a semnalului vocal, Editura militară, 1984.
14. J.D.Markel, A.H.Gray, Linear Prediction of Speech, Springer Verlag, Berlin, Heidelberg, New York, 1976.
15. L.Toma, A.Stoian, Predicția liniară, Institutul Politehnic "Traian Vuia" Timișoara, 1987.
16. J.Makhoul, Spectral Analysis of Speech by Linear Prediction, IEEE Transaction on Audio and Electroacoustics, June 1973.

17. J.Makhoul, Stable and Efficient Lattice Methods for Linear Prediction, IEEE Transaction on Acoustics, Speech and Signal Processing, vol.ASSP 25, October 1977.
18. A.Mătescu, Semnale, circuite și sisteme. Mitura Didactică și Pedagogică, București, 1979.
19. J.Makhoul, Linear Prediction: A Tutorial Review, Proceedings of the IEEE, vol.63, Nr.4, April 1975.
20. F.Itakura, S.Saito, T.Ishikawa, H.Sawabe, T.Nishikawa, An Audio Response Unit Based on Partial Autocorrelation. IEEE Transaction on Communication, vol.COM, Nr.4, August, 1972.
21. H.Feșeanu, L.Fortuna, Căntărea fenomenelor limbii române utilizând metoda predicției liniare, în volumul "Lucrările celui de al 5-lea Colocviu de informatică INFO Iași 85", Iași, Universitatea A.I.Cuza, 1985.
22. L.Fortuna, P.Xenicos, Noi tipuri de servicii pentru supraveghere și informarea automată a abonaților telefonici, bazate pe utilizarea vorbirii artificiale, Simpozionul de "Creativitate în poștă și telecomunicații", Ploiești, oct.1988.
23. R.Randall, Frequency Analysis, Pröl & Kjaer, Denmark, 1987.
24. N.B.Pinto, D.G.Children, A.L.Lalwani, Formant Speech Synthesis: Improving Production Quality, IEEE Transactions on Acoustics, Speech and Signal Processing, vol.ASSP 37, Nr.12, Dec. 1989.
25. K.A.El-Isam, An Unrestricted Vocabulary Arabic Speech Synthesis System, IEEE Transactions on Acoustics, Speech and Signal Processing, vol.ASSP 37, Nr.12, Dec.1989.
27. L.S.Lee - Synthesis Rules in a Chinese Text to Speech System, IEEE Transactions on Acoustics, Speech and Signal Processing, vol ASSP 37, Nr.9, Sept.1989.
28. G.Olaszy, Elektronikus beszédelőállítás. A magyar beszéd akusztikája és formanszintézise. Műszaki Könyv-kiadó, Budapest, 1989.
29. Y.Medan, Pitch Synchronous Spectral Analysis Scheme for Voiced Speech, IEEE Transactions on Acoustics, Speech and Signal Processing, vol.ASSP 37, Nr.9, Sept.1989.
30. R.P.Ramachandran, Pitch Prediction Filters in Speech Coding, IEEE Transactions on ASSP, Vol.ASSP 37, Nr.4, April. 1989.
31. V.Croza, L.Fortuna, I.Căciu, Sistem de achiziție și prelucrare automată, în timp real, a vorbirii, specializat pentru recunoașterea fenomenelor de tip vocalic, în volumul "INFO IASI 85" (citât la nr.21).

32. L. Fortuna, H. Porțeanu, Sistem de achiziție și prelucrare digitală a semnalelor analogice de joasă frecvență, realizat cu un microcalculator de tip "personal", cu aplicații la prelucrarea semnalului vorbit, în volumul "INFO-IASI 85" (citată anterior).
33. L. Fortuna, K. Fazekas, A. Fülöp, MSOUC-Standard Board-Based Microcomputer, în volumul "Lucrările Simpozionului de microprocesare, microcalculatoare și aplicații în economie", Timișoara, Institutul Politehnic "Traian Vuia", 1985.
34. L. Fortuna, V. Groza, Cercetări și rezultate obținute în prelucrarea automată a semnalului vorbit, cu aplicații în inteligența artificială, în volumul "Lucrările celui de-al IV-lea Simpozion: Robotizarea în industrie", Timișoara, Institutul Politehnic "Traian Vuia", 1984.
35. L. Fortuna, H. Porțeanu, Microsistem de achiziție și prelucrare digitală a semnalelor biologice, în volumul "Lucrările Simpozionului de Contribuții la dezvoltarea aparatului electronic medical", ediția 5, Cluj, Institutul Politehnic, 1984.
36. L. Fortuna, M. Chilințan, C. Cosovan, F. Vancea, Sistem de achiziții de date pentru microcalculatoare personale, în volumul "Lucrările simpozionului de microprocesare, calculatoare și aplicații în economie", Institutul Politehnic "Traian Vuia", Timișoara, 1988.
37. L. Fortuna, C. Cosovan, M. Chilințan, F. Vancea, Interfață pentru achiziții și conversie analog-numerică a semnalelor analogice pentru microcalculatoare, în volumul "A patra conferință de electronică, telecomunicații, automatică și calculatoare CNFTAC-88", București, Institutul Politehnic, 1988.
38. F. Moraru, M. Atodiroaei, Programarea microcalculatoarelor în sistemul de operare CP/M, Editura Științifică și enciclopedică, 1989.
39. L. Fortuna, Sintetizare de vorbire cu logică programată. Referat pentru doctorat. Institutul Politehnic București, Facultatea de Electronică și Telecomunicații.
40. W. Li, C. Chan, C. Vernis, D. Messerschmidt, Erzeugen und Erkennen von Sprachsignalen mit Analogprozessoren. Elektronik, 12, 1981.
41. B. Le Bos, Speak I/O is making itself heard, Electronics, May 22, 1980.
42. A. Mateescu, A. Serbănescu, Circuite cu capacități comutate, Editura Militară, 1987.

43. K.Sickart, Hohe Sprachqualität geringer Speicherbedarf, Elektronik, Nr.4, 24.02.1984.
44. O.Soskutny, Phonem-Synthesizer - IC ist μ P kompatibel, Elektronik, 17, 24.08.1984.
45. L.Fortuna, Corelarea metodelor de conversie analog-numerică cu metodele de analiză parametrică a vorbirii în scopul reducerii debitului informațional binar - rezultat, cu aplicații în realizarea sistemelor cu inteligență artificială, în volumul "INFO-IASI 85" (citată la 21).
46. D.G.Childrens, M.Hahn, J.N.Larar, Silent and Voiced /Unvoiced/ Mixed Excitation Classification of Speech, IEEE Transactions on Acoustics, Speech and Signal Processing, vol.ASSP 37, Nr.11 Nov. 1989.
47. G.Kaplan, S.Lerner, Realism în Synthetic Speech, IEEE Spectrum, April 1985.
48. L.Fortuna, Metodă de compresie a semnalului vorbit bazată pe analiza vorbirii, în volumul "A patra Conferință de electronică, telecomunicații, automată și calculatoare - CNETAC 88", Institutul Politehnic București, 1988.
49. L.Fortuna, Metodă de compresie a semnalului vorbit, Brevet de invenție RSR, Nr.92653, 1986.
50. L.Fortuna, H.Porțeanu, Sintetizor de vorbire pentru roboți, în volumul "Al V-lea simpozion național de roboți industriali" Institutul politehnic București, 1985.
51. L.Fortuna, H.Porțeanu, Sistem automat de supraveghere a funcționării unui reactor chimic, prevăzut cu posibilitatea generării unor mesaje verbale privind starea funcționării, în volumul "Eficiență și rentabilitate în proiectarea, fabricarea și exploatarea mașinilor și aparatelor electrice", Întreprinderea Electromotor, Timișoara, 1985.
52. N.S.Jayant, P.Holl, Digital Coding of Waveforms, Englewood Cliffs, New Jersey, Prentice-Hall, 1984.
53. M.W.Marcellin, T.R.Fischer, J.D.Gibson, Predictive Trellis Coded Quantization of Speech, IEEE Transactions on Acoustics, Speech and Signal Processing, vol.38, Nr.1, Jan.1990.
54. T.Pădulescu, B.Gheorghiu, D.Stefănescu, Centrale telefonice automate Pentacross, Editura Tehnică, București, 1984.
55. A.Mateescu, Direcții și tendințe în dezvoltarea rețelei de telecomunicații, în volumul "Sesiunea anuală de comunicări științifice "Telecomunicații 87", Poiana Brașov, 21-22 Mai, 1987, Intr.Electromagnetica, CCSITTEC, București.

56. A.Mateescu, Telecomunicațiile și societatea, în volumul "Creativitatea științifică în slujba progresului social", Editura Politică, 1989.
57. J.Ganin, Les etudes de synthese de parole au CNST, L'echo des recherches, Juillet, 1976
58. G.Takacs, Synthetic speech aided subscriber services, Buda Vox Telecommunication Review, 1983, Nr.2.
59. R.Pickvance, Speech synthesis: the new frontiers, Electronic Engineering, July 1980.
60. J.P.Cater, Electronically Speaking: Computer Speech Generation, Howard W.Sams & Co. Inc. Indianapolis, 1983.
61. L.Fortuna, Posibilități de utilizare a vorbirii artificiale în scopul extinderii serviciilor speciale automate, oferite abonaților telefonici. Simpozionul anual de Telecomunicații, 26-28.05.1988, Poiana Brașov.
62. L.Fortuna, Noi tipuri de servicii pentru supravegherea și informarea automată a abonaților telefonici, bazate pe utilizarea vorbirii artificiale. Simpozionul de creativitate în poștă și telecomunicații: SICRSPT 88, Ploiești, 12-13,10.1988.
63. L.Fortuna, D.Brustur, M.Banfy, Sistem electronic de tip "flashing" pentru centrale telefonice automate, Simpozionul anual de Telecomunicații, "Telecomunicații 89", Poiana Brașov, 18-20 05.1989, Intr.electromagnetica, COSITC București.
64. M.Cornea Hașegan, D.Cornea Hașegan, Proiectarea sistemelor cu microprocessor. Z80, Editura Dacia, Cluj, 1988.
65. L.Fortuna, H.Forțeanu, K.Fazekas, Aparat electronic de generare și anunțare a orei exacte, prin vorbire sintetică. Brevet de invenție RSR, Nr.91073/1985.
66. L.Fortuna, H.Forțeanu, K.Fazekas, A.Fülöp, B.Fazekas, C.Denes, F.Vancea, Sintetizor de vorbire destinat anunțării orei exacte, în volumul "Conferința națională de electronică, automatizată, telecomunicații, calculatoare - ONSTAC 86", Institutul politehnic București, 1986.
67. L.Fortuna, F.Reisz, R.Sohlér, Aspecte tehnologice referitoare la realizarea sintetizoarelor de vorbire, în volumul "Lucrările simpozionului național de tehnologie electronică și fiabilitate, Băile Herculane, 1987.

C U P R I N S

1.	Metode de sinteză a vorbirii.	1
1.1.	Introducere. Clasificări.	1
1.2.	Metode de sinteză a vorbirii, cu codificarea sursei de semnal (Metoda parametrice de sin- tază)	3
1.2.1.	Introducere. Clasificări.	3
1.2.2.	Sinteza fonemică.	7
1.2.3.	Sinteza prin refacerea înfășurătorii spectrale. .	10
1.2.4.	Sinteza pe bază de formanți	13
1.2.5.	Sinteza prin predicție liniară	23
1.2.5.1.	Principiul predicției liniare	23
1.2.5.2.	Aplicarea predicției liniare la semnalul de vorbitură.	25
1.2.5.3.	Determinarea coeficienților de predicție prin metoda autocorelației	27
1.2.5.4.	Determinarea coeficienților de predicție prin metoda rețelilor cu celule în X (lattice)	29
1.2.5.5.	Considerații cu privire la sinteza vorbirii prin predicție liniară	32
1.3.	Metode de sinteză a vorbirii, cu codificarea forme de semnal	34
2.	Sisteme cu logică programată pentru analiza vorbirii.	39
2.1.	Obiective impuse de sinteză analizei vorbirii . .	39
2.2.	Metode numerice de prelucrare a semnalelor analogice eșantionate, utilizate frecvent în analiza vorbirii	41
2.2.1.	Transformata Fourier rapidă	42
2.2.1.1.	Clasificare a tipurilor de algoritmi TFR	42
2.2.1.2.	Algoritmi cu decimare pentru TFR	45
2.2.1.2.1.	Algoritmi TFR cu decimare în timp.	45
2.2.1.2.2.	Algoritmi TFR cu decimare în frecvență.	49
2.2.2.	<u>Tehnica ferestruirii în analiza semnalului</u> <u>vocalizării.</u>	51
2.2.3.	Spektrul de putere al TFD	54
2.3.	Contribuții la realizarea sistemelor cu logică programată pentru analiza vorbirii.	54

2.3.1.	Prezentarea generală a sistemului de analiză realizat.	54
2.3.2.	Extensia de memorie RAM-DISC	58
2.3.3.	Rezultate experimentale obținute	60
3.	Sisteme cu logică programată pentru sinteza vorbirii	63
3.1.	Caracteristici generale ale sintetizoarelor de vorbire bazate pe logică programată	63
3.2.	Sintetizoare de vorbire cu codificarea sursei de semnal, bazate pe logică programată	66
3.2.1.	Probleme de bază ale sintetizoarelor de vorbire ce utilizează metode de sinteză cu codificarea sursei de semnal.	66
3.2.2.	Procesoare de semnal pentru sinteza vorbirii prin metode cu codificarea sursei. Considerații generale.	75
3.3.	Caracteristici generale ale sintetizoarelor de vorbire cu codificarea formei de semnal, bazate pe logică programată	77
3.4.	Sisteme complexe dedicate sintezei vorbirii	80
3.5.	Contribuții la realizarea sintetizoarelor de vorbire bazate pe logică programată	81
3.5.1.	Precizări generale	81
3.5.2.	Metodă de compresie a vorbirii bazată pe analiza semnalului vorbit	81
3.5.2.1.	Bazele experimentale ale metodei	81
3.5.2.2.	Prezentarea metodei de compresie	82
3.5.2.3.	Evaluarea eficacității metodei de compresie propuse	84
3.5.2.4.	Concluzii cu privire la metoda de compresie propusă	87
3.5.3.	Program de sinteză a vorbirii prin predicție liniară	88
3.5.4.	Sintetizor de vorbire bazat pe codificarea formei semnalului	90
3.5.5.	Sisteme de sinteză a vorbirii dedicate unor aplicații industriale	91
4.	Aplicații ale sintezei vorbirii în telecomunicații	94
4.1.	Domenii de utilizare a sintezei vorbirii în rețeaua telefonică	94

4.2.	Sisteme de sinteză a vorbirii destinate unor aplicații în domeniul rețelei telefonice. . . .	96
4.2.1.	Considerații generale.	96
4.2.2.	Exemple de sintetizoare de vorbire, cu logică programată, dedicate unor aplicații în cadrul rețelei telefonice.	97
4.2.2.1.	Sintetizor destinat transmiterii informației de taxare a convorbirilor telefonice.	97
4.2.2.2.	Sintetizor destinat informării abonaților telefonici cu privire la schimbarea numerelor de telefon.	97
4.2.2.3.	Sintetizor specializat pentru transmiterea orei exacte	98
4.3.	Contribuții la aplicarea vorbirii artificiale în telecomunicații	100
4.3.1.	Investigarea unor domenii de aplicare a vorbirii artificiale în rețeaua telefonică. . .	100
4.3.2.	Echipamente complexe, cu vorbire artificială, realizate pentru rețeaua telefonică.	104
4.3.2.1.	Sistem de supraveghere automată, prin telefon .	104
4.3.2.2.	Sisteme pentru prezentarea automată, prin telefon a indicativului centralelor telefonice automate (flashing)	106
4.3.2.2.1.	Sistem cu vorbire artificială destinat numai funcției de flashing	106
4.3.2.2.2.	Sistem complex de identificare și testare a centralelor telefonice automate	108
4.3.2.2.2.1.	Probleme ale testării rețelei telefonice . . .	108
4.3.2.2.2.2.	Sistem de tip "flashing" prevăzut și cu posibilități de testare automată	109
4.3.2.3.	Sistem robot pentru transmiterea prin telefon a orei exacte.	116
4.3.2.3.1.	Prezentare generală	116
4.3.2.3.2.	Descrierea funcțională a robotului telefonic pentru ora exactă	119
4.3.2.3.3.	Tehnici de testare implementate prin program la robotul telefonic de oră exactă	121
4.3.2.3.4.	Conceptie tehnologică și concluzii cu privire la introducerea în fabricație de serie a robotului telefonic pentru oră exactă	124

5. Considerații finale 127
Bibliografie 133
Cuprins 138