Modul universal pentru sinteza mesajelor cu voce multilingve

Sergiu TINCOVAN, Iurie SOROCEANU, Vitalie SECRIERU, Eugeniu MUNTEANU, Valerian DOROGAN
Technical University of Moldova
s_tincovan@mail.ru , iuis@mail.md , primcast@mail.com, eugeniumunteanu@gmail.com , dorogan_lme@yahoo.com

Abstract — **Most voice messaging devices and systems are provided for a single language, and changing messages in to another language represents a technical difficulties. Elaborated method supports two and more languages at the same time, as will as gives possibility of completing and correcting the dictionary with reasonable expenses. The method of synthesizing voice messages is based on the original procedure for using standard methods.**

Index Terms — **Microcontroller, delta-modulation, spectral analysis, phonemes, frequency formats, vocal sound, consonant sound.**

I. INTRODUCERE

În tehnică tot mai frecvent necesitatea de a transforma mesajele de text sau codul mesajului de eroare în expresii cu voce. Ca exemplu pot servi sistemele de diagnosticare, robot-secretat al serviciilor de dispecerat, sistemul de bord în transportului, telefonie, bioacustica, medicină, sistem de educaţie, etc.

Majoritatea sistemelor de tipul TTS (Text To Speech, sau text în voce) sunt prevăzute pentru menţinerea unei singure limbi, ce nu întotdeauna poate asigura schimbarea operativă a limbii mesajelor în sistem. Schimbarea limbii în dispozitivele (sistemele) se sinteză a mesajelor cu voce necesită schimbarea chip-ului specializat sau reprogramarea circuitelor de memorie cu altă bază de date a altei limbi. Altă problemă asemenea sistem este imposibilitatea de menţinere în paralel a 2 şi mai multe limbi, unde e posibilă complectarea dicţionarului pe parcurs.

În lucrarea dată sunt analizate modalităţile de realizare a modulelor de sinteză a mesajelor cu voce de uz universal, ce permite soluţionat majoritatea problemelor menţionate.

II. Formularea problemei

Pentru mărirea eficienţei proceselor este necesar de soluţionat următoarele sarcini:

1) De analizat metodele de sinteză a mesajelor la nivel de algoritmi şi mijloace tehnice.

2) De obţinut o variantă optimală a structurii şi algoritmului de funcţionare pentru modului de sinteză, ce permite de modificat operativ limba şi dicţionarul ei.

3) De obţinut structura şi algoritmul de funcţionare optimal a modulului de sinteză a sunetelor şi mesajelor cu voce, unde dicţionarul (vocabularul) de sunete şi foneme este limitat după volum.

Pe parcursul soluţionării este necesar de ţinut cont de repartizarea funcţiilor executate între partea HARD şi SOFT a modulului de sinteză mesajelor cu voce, specificul de preparare a bancului de foneme şi sunete cu ajutorul mijloacelor SOFT specializate al PC aparte.

III. Căile de soluţionare şi implementare

Variantele posibile de realizare a sintezei sunetelor speciale şi vocii pot utiliza următoarele tehnologii:

1) Metoda directă de codare-restabilire a semnalelor sunetelor şi vocii;

2) Modelarea digitală a tractului vocii;

3) Sinteza analogică a formantelor de frecvenţă;

4) Sinteza digitală a fonemelor.

Pentru metoda directă codare-restabilire a semnalelor sunetelor şi vocii codarea semnalelor vocii se efectuează cu ajutorul PCM (Pulse Code Modulation) sau delta-modulare [1, 2].

Informaţia despre semnal vine în formă succesiunii de eşanţionări cu frecvenţa de discretizare f_{discr} , Conform teoremei Kotelnicov este necesar de respectat $2 \cdot f_{max} \le f_{discr}$, unde f_{max} este frecvenţa maximă a spectrului semnalului vocii. Această relaţie este valabilă pentru filtrare ideală a funcţiei restabilite, pentru condiţiile reale este acceptată condiţia $f_{discr}=45 \cdot f_{max}$.

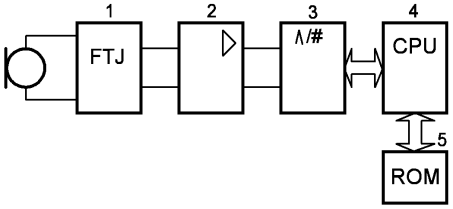

Fig. 1. Convertirea semnalului cu ajutorul PCM.

Pentru intervalul de frecvenţe, unde valoarea maximă nu depăşeşte 8-10 KHz în tractul preliminar de prelucrare a semnalului se include filtru trece jos cu scopul de a elimina frecvenţele mai înalte $f_{discr}/2$ pentru intrarea CAD. Pentru imprimarea şi redarea vocii este necesar să fie asigurată compatibilitatea CAD şi DAC după binaritate şi frecvenţa discretizării. Rata de transfer a datelor constituie circa 96 Kbit/s pentru o secundă de sonorizare, care permite de simplificat partea HARD a echipamentului şi asigură calitate suficientă a sintezei sunetului.

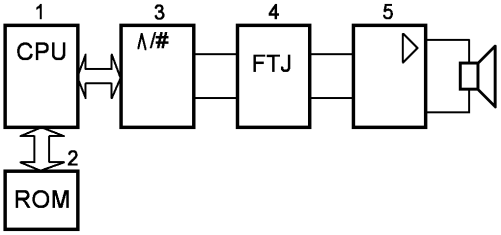

Fig. 2. Schema de structură a sintezării vocii cu ajutorul PCM.

În calitate de alternativă a metodei directe de codare-restabilire a semnalelor sunetelor şi vocii poate servi delta-modularea, care permite de redus rata de transfer, care se bazează pe variaţia relativă a amplitudinii şi nu pe valorile

absolute. Semnalul vocii ca şi pentru PCM este prelucrat preliminar, care apoi este supus delta-modulării.

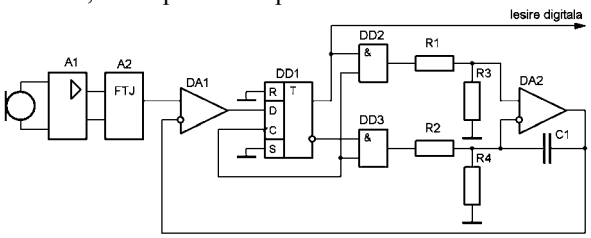

Fig. 3. Schema funcţională a delta-modulatorului.

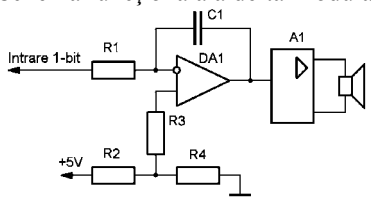

Fig. 4. Schema funcţională a delta-demodulatorului.

Sinteza analogică a formantelor de frecvenţă se bazează pe cunoaşterea detaliată a fonemelor şi descompunerii fonetice a mesajelor, la temelia cărora stau două noţiuni: lingvistică – foneme, şi acustică – formantă.

Sub noţiunea de formantă se subînţeleg frecvenţele de rezonanţă (polurile funcţiei de transfer) a sistemului acustic vocal. Parametrii formantelor (frecvenţa, banda de trecere şi amplitudinea) sunt determinate de parametrii tractului vocal. Cel mai important parametru, frecvenţa este strâns legat de configuraţia geometrică a tractului vocal, unde în procesul vorbirii odată cu schimbarea parametrilor configuraţiei se modifică şi valorile frecvenţelor formantelor (fig. 5.).

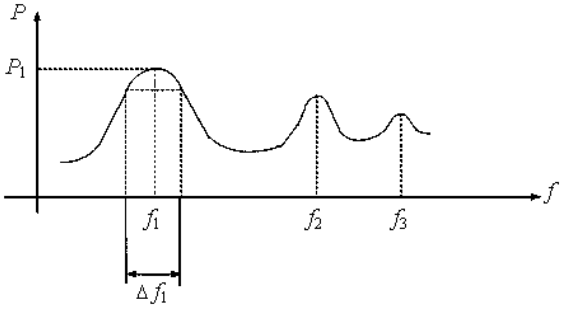

Fig. 5. Exemplu al spectrului vocii.

Pentru sinteza vocii sunt suficiente 2-4 frecvenţe a formantelor, unde prima formantă are frecvenţa 200 Hz (prima formantă a vocii bărbăteşti) până la 2000 Hz (a treia formantă a vocii femenine). În procesul vorbirii frecvenţele formantelor şi obertoanelor sunt prezente simultan şi se deplasează pe axa spectrului, ce corespund particularităţilor cuvântului pronuţat. Deaceia mesajele vocale sau sunetele lumii animale se percep nu ca o singură frecvenţă, ci o mulţime de armonici, care se formează la filtrarea impulsurilor, formate la ieşirea tractului vocal (fig. 6).

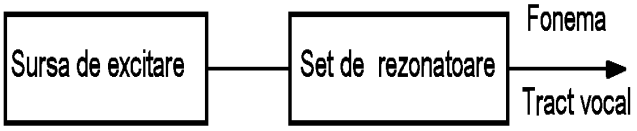

Fig. 6. Modelul tractului vocal.

În proces de vorbire tractul vocal funcţionează în calitate de un set de rezonatoare ce filtrează spectrul semnalului de excitare, ca rezultat apare un tablou pe caracteristica spectrală, în care se conţin un şir de maxime (ele şi reprezintă rezonanţele formantelor). Ca exemplu este format un tabel pentru vocalele „o”, „a” şi „i”.

Tabelul 1.

Fonema	Frecvenţele formantelor		
	F_1	F_2	F_3
o	275	850	2400
i	250	2300	3000
a	575	900	2450

Prin generarea simultană a frecvenţelor F_1, F_2, şi F_3 conform tabelului 1 putem obţine sunetele vocale, schema de structură a sintezatorului este dată mai jos (fig. 7.)

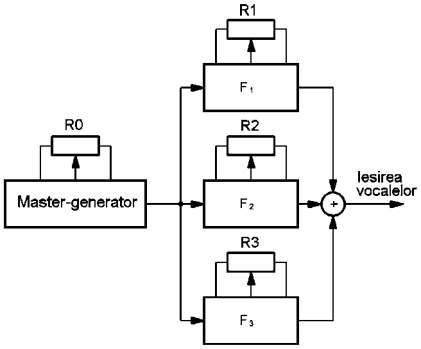

Fig.7. Schema de structură a sintezei sunetelor vocale.

Schema sintezatorului de sunete vocale conţine master-generator al tonului principal şi filtre trece bandă reglabile, ce pot fi acordare pe frecvenţele formantelor (cu rezistenţele R_1-R_3) şi sumator, ce mixează semnalele de la ieşirea filtrelor. Corespunzător spectograma la ieşire va conţine trei frecvente a formantelor identice acelor sunete vocale ce sunt pronuţate pe viu.

Este mult mai dificil de sintezat sunetele consonante, ele se formează când în cavitatea bucală se formează obstacole pentru aerul expirat, aşa tipuri de consonante sunt:

- explozive (sunetele „p”, „t”, „k”)
- fricative (sunetele „s”, „f”, „h”)
- nazale (sunetele „n”, „m”)
- affricata (sunete „ci”, „t-ş-ci”, „ţ”, „ts”)

Schema de structură unui asemenea modul permite sinteza a sunetelor vocale şi consonante, din care se poate de format sunete speciale şi mesaje (fig. 8), unde sunetele sunt sintezate în ordinea prescrisă de vocabular (dicţionar) [3].

Sinteza digitală a fonemelor se bazează pe generarea fonemelor şi compilarea ulterioară din ele a sunetelor speciale, cuvinte, propoziţii şi fraze. Realizarea ei combină metode compacte de prelucrare digitale şi flexibilitatea de gestionare cu parametrii principali a vocii, ce este specific modelelor de formante (fig. 9). Procesul codării vocabularului necesar este substituit cu compilarea mesajelor arbitrare dintr-un set de elemente a vocii, pregătit preliminar [4].

Sintezatorul de foneme conţine 3 nivele de prelucrare, unde la primul nivel se efectuează translarea simbolurilor orfografice în codurile fonemelor, la nivelul doi se calculă setul parametrilor acustici, care serveşte pentru gestionarea nivelului trei – formarea semnalului vocii.

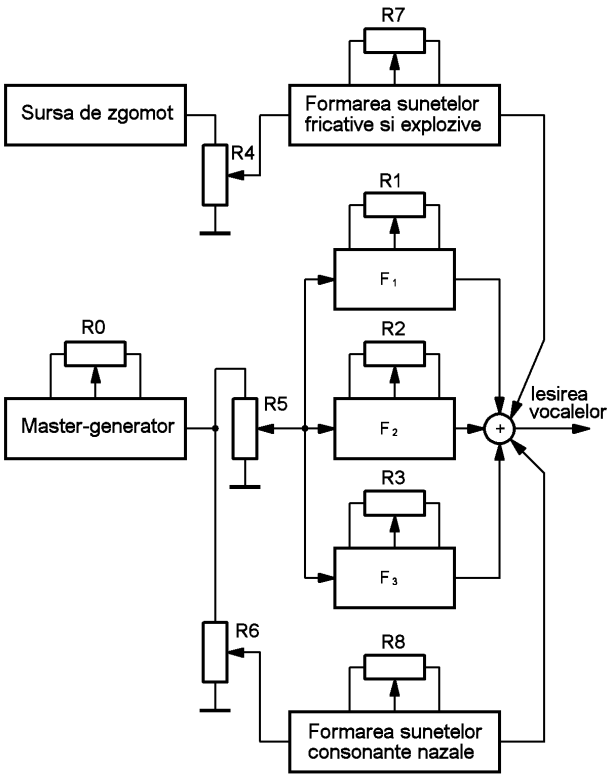

Fig. 8. Schema de structură a sintezatorului sunetelor vocale şi consoane.

Pentru nivelul doi de prelucrare se calculă setul parametrilor de gestionare, care reprezintă frecvenţele formantelor (F_1 ,F_2 ,F_3), banda de frecvenţe (ΔF_1, ΔF_2, ΔF_3), frecvenţa tonului principal şi amplitudinea de vocalizare. Setul parametrilor regenerează cu interval de 6,4 ms, ce asigură de urmărit cele mai rapide variaţii între foneme, rata de transfer de la nivelul doi la nivelul trei constituie circa 45 Kbit/s. Sintezarea propriu zisă se efectuează la nivelul trei, unde semnalele de excitare (armonic şi de zgomot) se filtrează cu ajutorul filtrelor de rezonanţă a formantelor a tractului vocii.

Frecvenţa de eşanţionare constitue 10 kHz, ce permite de redat componentele de frecvenţă până la 5 kHz. Ca exemplu poate servi procesorul DSP de tipul TMS32010 a firmei „Texas Instruments" (SUA)

Codarea vocii cu ajutorul coeficienţilor liniari de predicţie (CPL) se bazează pe teoria analizei statistice a seriilor de timp. Seria de timp reprezintă o succesivitate de observaţii (referinţe) aranjate în timp. Esenţa acestei metode este următoare. Admitem succesivitatea seriei a semnalului vocal x_1, x_2, … x_t. Pentru această mulţime se calculă "valoare medie". Pentru un interval prestabilit (de exemplu 1020 ms) proprietăţile statistice a semnalului vocal sunt neschimbate. Acest interval se codează cu un set de coeficienţii a_s , care minimizează eroarea medie patratică a predicţiei, cu alte cuvinte reduce la minim eroarea între seria iniţială şi "netezită". Calcularea coeficienţilor constituie o procedură dificilă (se calculă ecuaţiile diferenţiale cu metoda patratelor minimale). Principiul analizei CPL şi codării sunt date pe fig. 10.

Ca rezultat analiza CPL a vocii codate PCM, alcătuită din eşanţionări cu frecvenţa 1020kHz, se convertează în serie de vectori a parametrilor cu frecvenţa 50100Hz, unde compresia descrierii vocii se comprimă de 50-100 ori cu pierdere neesenţială a calităţii.

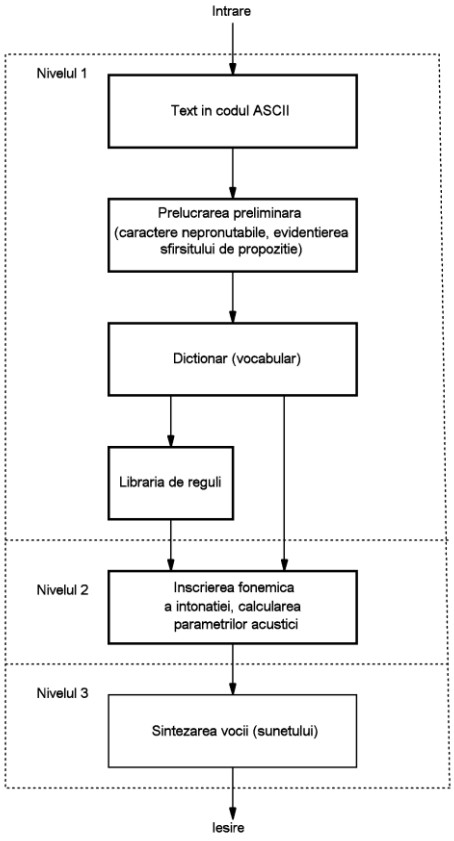

Fig. 9. Structura sintezatorului digital de foneme.

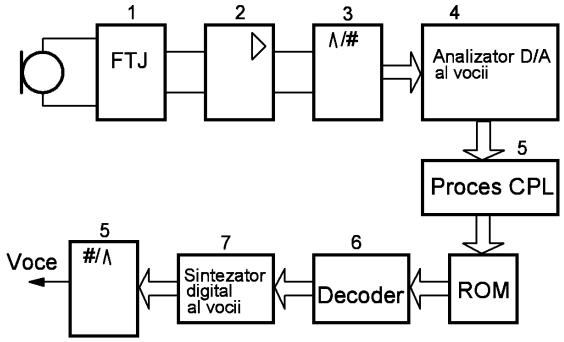

Fig. 10. Schema de structură a codări/decodări CPL.

Metoda dată de sinteză a vocii combină în sine performanţele PCM şi sintezei cu formante. În CPL sinteză a vocii se utilizează filtre liniare recursive de imitare a tractului vocal (fig. 11).

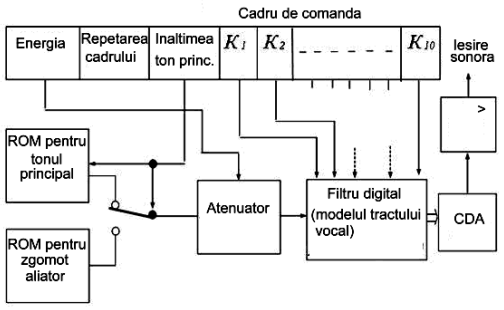

Fig. 11. Structura sintezatorului CPL.

Ultimele 10 elemente a cadrului CPL corespund coeficienţilor filtrului digital pentru sinteza vocii. Pentru implementare sunt utilizate procesoare specializate sau procesoare programabile, deoarece viteza de lucru a procesoarelelor tradiţionale este insuficientă pentru asemenea scop.

Metoda dată este destul de complicată în realizare, deoarece ea cere echipament de implementare foarte rapid. Însă această metodă se consideră de perspectivă, deoarece el se bazează pe procedee perfecte de formare a vocii. Posibilitatea de gestionat cu parametrii modelului permite de acordat sunetele şi cuvintele după nivelul energiei, tempou şi tonalitate. Acest lucru permite de sintezat mesaje mai complicate pe baza setului de elemente cu ajutorul regulilor.

Pentru implementarea modulului de sinteză din variantele menţionate pentru cerinţele de menţinere dicţionarului nelimitat cele mai reuşite metode este analiza formantelor (poate sinteza orice sunet a limbilor existente) şi metoda fonemică (poate sinteza mesaje în mai multe limbi). Alt argument în folosul sintezei fonemice este cea mai joasă rată transfer pentru canalele de legătură (tabelul 2.), unde se păstrează calitatea sunetului sintezat.

Tabelul 2.

Metoda de codare a vocii	Cheltuieli informaţionale, bit/s
PCM	$(40\text{-}100)\cdot 10^3$
Delta-modulare	$(20\text{-}50)\cdot 10^3$
Delta-modulare adaptată	$(10\text{-}25)\cdot 10^3$
Analiza formantelor	$(2\text{-}4)\cdot 10^3$
Metodă fonemică	50-100

IV. CONCLUZII

Pentru implementarea modulului de sinteză a mesajelor cu voce este necesar de utilizat algoritmul din fig. 9, unde la nivelul 3 de recurs la redarea samplerelor fonemelor, ce sunt extrase dintr-un banc foneme, unde tipul şi succesiunea este definită de către program la nivelul 2 (fig. 9.).

1) Pentru sinteza sunetelor şi mesajelor este suficient să fie la dispoziţie samplere a circa 300-600 silabe în diferite combinaţii, pentru obţinerea parametrilor mai performanţi este necesar de mărit numărul samplerelor a variantelor de silabe.

2) Calitatea sunetului va fi determinată de binaritarea şi frecvenţa de eşanţionare a semnalului vocii-donor, pentru telefonie, sisteme de bord, servicii de dispecerat, etc. este suficient de utilizat binaritatea discretizării 8 biţi şi frecvenţa de eşanţionare 22500 Hz.

3) Pentru convertirea codului în semnal acustic e suficient de recurs la metoda PCM sau PWM, alte modalităţi sunt mai costisitoare în resurse HARD

Ca un neajuns a variantei propuse este necesitatea de a obţine semnal a vocii-donor, unde dictorul citeşte un text special, ce conţine maximul combinaţiilor de silabe, apoi imprimarea este prelucrată minuţios. Rezultatul prelucrării imprimării va fi acel banc de foneme, ce se obţine prin decuparea silabelor din fonograma textului citit.

Pentru înlăturarea acestor neajunsuri este necesar de inclus elemente de intelect artificial ce va constitui direcţia de cercetări ulterioare.

REFERINŢE

1. Г. Нуссбаумер. Быстрое преобразование Фурье и алгоритмы вычисления сверток.— М.: Радио и связь, 1985.

2. А. В. Фролов, Г. В. Фролов. Мультимедиа для Windows. Библиотека системного программиста, т. 15 М: Диалог-МИФИ, 1994

3. Л. Захаров. Проблемы создания аллофонной базы автоматического синтеза речи (http://art.bdk.com.ru/govor/rasp.htm).

4. Black A.W., Taylor P. Automatically clustering similar units for unit selection in speech synthesis // In Proceedings of Eurospeech 97. Rhodes, Greece, 1997. Vol.2, pp. 601-604.