

Sincronizarea secvențelor video cu fluxuri de tip text

Dârdală Marian

Academia de Studii Economice
București, Piața Romană, nr 6
dardala@ase.ro

Reveiu Adriana

Academia de Studii Economice
București, Piața Romană, nr 6
reveiua@ase.ro

Furtună Titus Felix

Academia de Studii Economice
București, Piața Romană, nr 6
titus@ase.ro

REZUMAT

Construirea de aplicații multimedia presupune și controlul precum și gestiunea fluxurilor media. Cea mai complexă dată implicată în aplicațiile multimedia este data de tip video. Pentru a utiliza o astfel de resursă multimedia există soluții orientate spre utilizatorul final ca în cazul componentelor software specializate din cadrul produselor software de creație multimedia, în acest caz accentul se pune pe scenariul de prezentare, pe compoziție, pe sincronizarea elementelor implicate în prezentare. În ceea ce privește dezvoltarea aplicațiilor de tip *low level* adică atunci când dezvoltatorul dorește să-și construiască propriul său modul de derulare a secvențelor video, el trebuie să implementeze și partea de sincronizare a fluxurilor și eventual sincronizarea celor native, existente în cadrul resursei multimedia cu cele externe resursei multimedia. Prin acest articol ne propunem să prezentăm modalitățile de reprezentare a fluxurilor externe de tip text și sincronizarea lor cu fluxurile audio și de imagini ale unei date video.

Cuvinte cheie

Tip Media, Flux, Audio, Video, Text

Clasificare ACM

H5.2. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCERE

În cadrul articolului sunt prezentate câteva dintre cele mai cunoscute și folosite formate de reprezentare a fluxurilor de tip text precum și modul în care este definită informația de sincronizare. Astfel, există formate de reprezentare care permit descrierea adnotărilor numai într-o singură limbă cât și formate capabile să redea textele în mai multe limbi de circulație, urmând ca utilizatorul să poată selecta limba în care textul să fie redat. În acest fel se răspunde și unei cerințe a dezvoltării sistemelor informatice, mai precis este vorba de accesul universal la informație. Alegerea limbii se poate face în mod implicit pe baza informațiilor de tip *Cultural*, setate prin aplicația *Regional and Language Options*, existente pe calculatorul utilizatorului.

Validarea aspectelor teoretice s-a realizat prin dezvoltarea unei aplicații de tip *Video Player* prin intermediul căreia s-a urmărit sincronizarea fluxurilor native ale unei date video cu un flux extern de tip text. Tehnologia software multimedia folosită pentru implementare a fost *Video for Windows* accesibilă programatorilor prin intermediul unui set de funcții API ce lucrează cu dispozitivele specifice.

ACCESUL DIRECT LA SECVENȚELE VIDEO

Player-ele video pot accesa în mod direct o subsecvență video sau chiar un cadru individual din fluxul de imagini. Accesul direct la diferite cadre ale unui film presupune indexarea invariabilă în timp, a imaginilor fixe ce compun

mișcarea. Acest lucru se realizează cu ajutorul sistemelor de reperaj cunoscute sub denumirea de *Time Code* și de *Frame Code*[1].

Time Code-ul este o secvență de coduri numerice generate la un anumit interval de timp de către sistem. Sistemul de indexare *Time Code* este o metodă care folosește timpul pentru codificarea semnalului video. Metoda a fost perfectată de Society of Motion Picture and Television Engineers (SMPTE) ca un standard de sincronizare universal, utilizat în editarea video, audio și a filmului. Conform acestei codificări, fiecărui cadru video individual i se asociază câte un număr unic, ce permite controlul exact al benzii și indexarea secvențelor. Codificarea se poate realiza în diferite formate: time code *liniar* (LTC) și *vertical* (VTC). Prin aceste numere se pot adresa cadrele individuale ale unei secvențe prin codul *time* și apoi se pot accesa direct în timpul procesului de editare. Numărul unic al fiecărei imagini se codifică într-un bloc de 80 de biți, ce reprezintă în fapt, momentul captării sau înregistrării acesteia. Astfel, *Time Code* indică prin codajul BCD (Binary Coded Decimal) ora, minutul, secunda și numărul imaginii din fluxul video. Deoarece semnalul *Time Code* este înregistrat pe o pistă audio paralelă, codul s-a mai numit și *Liniar Time Code* sau *LTC*. Acest mod de a ține evidența cadrelor, poate pierde cu ușurință sincronismul imagine - numărul unic *Time Code*, datorită paralelismului semnalului cu banda video, la o simplă defazare a semnalului. Datorită acestui lucru, un alt format de codificare *Time Code* este cel *vertical* care apare ca o pereche de linii verticale negre și albe incluse în semnalul video. Aceste linii sunt de obicei inserate în intervalul neutilizat dintre două cadre consecutive ale unui flux video și exprimă același lucru ca și LTC.

Un alt sistem de reperaj și indexare a cadrelor video este *Frame Code*. Din punctul de vedere al codificării, el este mai simplu și se bazează pe numerotarea crescătoare a imaginilor fixe din fluxul video, începând cu primul cadru până la 300000.

În cadrul interfeței API sunt puse la dispoziția utilizatorilor funcții care permit conversia între cele două sisteme de reperaj[5]. Astfel, funcția:

```
STDAPI_ (LONG) AVISampleToTime
(PAVISTREAM p, LONG f);
```

convertește poziția cadrului f din fluxul p dată prin numărul său (Frame Code) în *Time Code*-ul corespunzător, în sensul că funcția returnează timpul exprimat în milisecunde. Conversia, în sens invers, se realizează prin funcția:

```
STDAPI_ (LONG) AVISampleToTime
(PAVISTREAM p, LONG t);
```

care primește poziția cadrului, din fluxul p , exprimată în milisecunde (t) și returnează poziția acestuia prin *Frame Code*-ul său.

Având disponibile aceste funcții de conversie putem face ca orice poziționare în flux să o exprimăm fie prin *Time Code* fie prin *Frame Code*.

FLUXURI DE TIP TEXT

Speculând tehnicile de acces direct la resursele multimedia de tip video acestea au putut fi adnotate prin secvențe de tip text furnizând informațiile de sincronizare în diferite forme (Time Code sau Frame Code). Adnotările secvențelor de tip video reprezintă o formă alternativă de transmitere a informațiilor către utilizator[2]. Spre exemplu, dacă explicațiile nu pot fi asimilate auditiv, într-o secvență video, fie datorită unei dizabilități a utilizatorului, fie datorită faptului că utilizatorul nu înțelege limba în care este prezentată informația, textul, în acest caz reprezintă o formă alternativă de redare a informației. Prin intermediul secvențelor de tip text se pot substitui secvențe video, se pot insera comentarii în prezentare, utilizând motoare de sinteză vocală se pot reda auditiv textele, în diferite limbi de circulație etc.

Dintre cele mai cunoscute formate de reprezentare a fluxurilor de tip text putem aminti:

- formatul *MPSub*, rezidă în fișiere ce au extensia *sub* și au descrierea bazată pe text;
- formatul *SubRip*, rezidă în fișiere ce au extensia *srt* și au descrierea bazată pe text;
- formatul *SAMI*, rezidă în fișiere ce au extensia *smi* și au descrierea bazată pe formatul HTML.

Formatele *sub* sau *srt* conțin, în esență, limitele intervalelor și textele care trebuie afișate sincron cu celelalte fluxuri din secvența video. Intervalul se poate exprima prin unități de timp sau prin numărul cadrelor.

De exemplu, un fișier care conține informațiile de sincronizare prin unități de timp are forma:

```
1
00:00:23,130 --> 00:00:27,090
Care animale locuiesc
în cele mai elegante case din Paris?
2
00:00:27,170 --> 00:00:31,000
Care animale dorm în catifea?
3
00:00:31,070 --> 00:00:35,030
Care animale abia de au vreo pată?
```

Se poate observa că paragrafele sunt numerotate, limitele intervalului de timp se precizează în forma:

HH:MM:SS,MIL, unde HH reprezintă ora, MM minutul, SS secunda și MIL milisecunde.

Un exemplu de fișier ce conține informații de sincronizare exprimate prin numărul cadrelor este:

```
{98}{110} Ce are de face figuri?
{214}{232} Ma enerveaza ca nu merge!!
{242}{271} La naiba cu e!!
```

Informația de sincronizare s-a redat în forma $\{ci\}\{cf\}$, unde prin *ci* se înțelege cadrul inițial iar prin *cf* cadrul final {de la care} {la care} textul va fi vizualizat.

În ceea ce privește formatul *SAMI*, un astfel de document conține următoarele elemente[4]:

- definirea caracteristicilor celor două stiluri folosite pentru scrierea textului;
- definirea variabilelor de lucru asociate limbilor folosite pentru redarea prezentării;
- definirea secțiunilor <SYNC> folosite pentru sincronizarea textului cu secvența video. În această secțiune se specifică momentul de timp, raportat la începutul secvenței video, în care va fi afișată fiecare secvență de text (subtitrare sau descriere) și textul propriu-zis în cele două limbi indicate.

Un exemplu de document *SAMI* este:

```
<SAMI>
<HEAD>
<STYLE TYPE = "text/css"><!--
<!--P definește stilurile corespunzătoare celor două opțiuni ale
utilizatorului: scris mic și scris mare-->
P {font-family:sans-serif; color:white;}
#Small {Name:SmallPrint; font-size:10pt; color:yellow;}
#Big {Name:BigPrint1; font-size:14pt; color:magenta;}
<!--ENUSCC și ROROCC definesc variabilele folosite pentru a
indica limbile folosite pentru transcrierea și afișarea textului-->
.ENUSCC {Name:English; lang: en-US; SAMIType:CC;}
.ROROCC {Name:Romanian; lang: ro-RO; SAMIType:CC;}-->
</STYLE>
</HEAD>
<BODY>
<!--Primul set de texte va fi afișat la 100 ms-->
<SYNC Start = 100 ID=Source>
<--Textul în engleză-->
<P Class = ENUSCC ID = Source>Presenter
<P Class = ENUSCC> the text in english
<--text în română-->
<P Class = ROROCC ID = Source>Prezentator
<P Class = ROROCC> textul 1 în română
<!--Al 2-lea set de texte va fi afișat la 1200 ms-->
<SYNC Start = 1200 ID=Source>
<--Textul în engleză-->
<P Class = ENUSCC ID = Source>Producer
<P Class = ENUSCC> the text in english
<--Text în română-->
<P Class = ROROCC ID = Source>Producător
<P Class = ROROCC>textul 2 în română
.....
</BODY>
</SAMI>
```

Pentru sincronizarea fluxurilor native ale unei resurse multimedia cu fluxuri de tip text este necesară modelarea unitară a tipului media. Se prezintă în paragraful următor un model orientat obiect pentru controlul și gestiunea fluxurilor multimedia existente într-o resursă de tip video digital.

MODEL ORIENTAT OBIECT PENTRU GESTIUNEA FLUXURILOR UNEI SECVENȚE VIDEO

Conceptul de *tip media* este folosit pentru a referi unitar elementele care formează acest tip și anume: imagine, animație, sunet și video. Acest tip, în diferite pachete software este descris obiectual sub diferite nume, ca de exemplu obiectul *clip* în Multimedia Toolbook. Prin intermediul conceptului de *clip* în Multimedia Toolbook se gestionează unitar resursele multimedia chiar dacă ele sunt de tipuri diferite.

Pornindu-se de la componentele tipului media se pot face clasificări utile pentru a surprinde acele elemente care sunt utile în vederea modelării lui obiectuale[6].

Din punct de vedere al temporalității, elementele ce-l formează se pot clasifica în:

- *statice*, care nu au componentă temporală (imaginea);
- *dinamice*, care depind de factorul timp (animația, sunetul, video-ul).

Din punct de vedere al componente vizuale, există:

- elementele care necesită spațiu ecran pentru afișare (fereastră) - imagine, animație și video;
- elementele care nu necesită spațiu pentru afișare (sunet);

Din aceste clasificări deducem faptul că trebuie să lucrăm cu două dispozitive și anume unul pentru vizualizarea imaginilor (ecranul prin intermediul unei ferestre) și altul pentru audiția sunetului (placa de sunet).

Tipului media, în integralitatea lui, i se pot aplica operații de:

- deschidere (*open*) care implică preluarea informațiilor atât dintr-un fișier existent cât și dintr-un periferic specializat (ex. scanner pentru imagini fixe, microfon pentru sunet și webcam pentru video); pentru animație deschiderea se face numai pe baza furnizării fluxului de imagini;
- închidere (*close*) a fișierului sau a dispozitivului.

Pe de altă parte, datelor temporale existente în fișiere corespunzătoare li se asociază operații de tip:

- *play* – pentru derularea secvenței;
- *stop* – pentru oprirea secvenței cu posibilitatea reluării ei de la început;
- *pause* – pentru oprirea secvenței cu posibilitatea reluării ei din punctul în care ea a fost oprită.

Elementului static de tip imagine i se poate asocia operația de tip *play* cu sensul de vizualizare a imaginii dar operații de genul *stop* și *pause* nu au sens în acest caz.

În figura 1 se prezintă principalele clase cu interacțiunile dintre ele în vederea modelării tipului media[1].

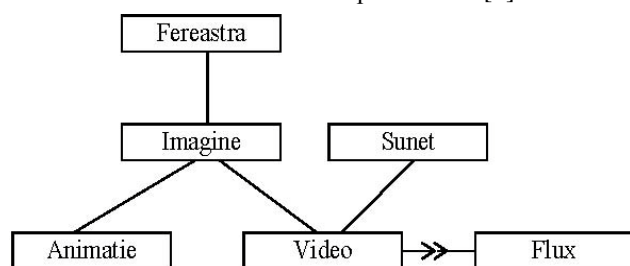


Figura 1 Modelarea obiectuală a tipului media

Se observă în figura 1 două tipuri de relații între clase:

- cele simbolizate prin — descriu relații de moștenire;
- cele simbolizate prin —>>— descriu relații de tip colecții.

Clasa *Fereastră* este utilizată pentru a afișarea imaginii pe suprafața client de lucru în diferite forme (cu posibilitatea scalării imaginii în raport de dimensiunea ferestrei). Se observă că ea este clasa de bază pentru orice resursă multimedia care necesită a fi vizualizată pe ecran.

Clasa *Animație* derivă din clasa *Fereastră* și gestionează fluxul de imagini care se vor afișa în fereastră.

Clasa *Sunet* este utilizată pentru a gestiona fluxul audio și transmiterea lui spre audiere.

Între clasa *Video*, pe de o parte și clasele *Fereastră* și *Sunet*, pe de altă parte, există o relație de moștenire multiplă în sensul că, clasa *Video* moștenește caracteristici atât de la clasa *Fereastră* cât și de la clasa *Sunet*. Logica acestei proiectări este explicată prin faptul că resursa de tip video este alcătuită, de cele mai multe ori, din flux de imagini și flux audio. Semnalul video în format AVI este accesibil la nivel de flux, de aceea el a fost modelat printr-o clasă separată numită *Flux*. În aceste condiții obiecte de tip *Flux* accesează fluxurile din cadrul unei secvențe AVI sau externe ei, iar prin intermediul clasei *Video* le transmite către obiecte specializate, în vederea derulării lor (*Fereastră* și *Sunet*). O sarcină deosebit de importantă a clasei *Video* constă în sincronizarea fluxurilor.

SINCRONIZAREA FLUXURILOR

Problema sincronizării fluxurilor în procesul de vizualizare se poate rezolva în mod diferit în funcție de numărul și natura fluxurilor existente într-o resursă de tip video.

În cazul în care resursa multimedia de tip video conține doar fluxul de imagini nu se poate vorbi de sincronizare între fluxurile native ale ei, însă palyer-ul trebuie să asigure controlul vitezei de derulare a cadrelor în procesul de vizualizare. Informații cu privire la rata de derulare a imaginilor exprimată în frame-uri pe secundă se găsesc în header-ul fluxului de imagini.

În cazul în care resursa multimedia de tip video integrează, pe lângă fluxul de imagini și fluxul audio, sincronizarea se face pe baza fluxului audio, adică se trimite fluxul audio plăcii de sunet pentru a fi redat și în mod repetat se interoghează cadrul audio emis de dispozitiv iar pe baza lui se identifică cadrul din fluxul de imagini care se afișează în fereastra de vizualizare. Corespondența dintre numărul cadrelor audio (eșantioane) ce corespund unui cadru individual din fluxul de imagini (*ncapv*) se determină astfel:

$$ncapv = AVIStreamLength(pavia) / AVIStreamLength(pavi);$$

unde, $AVIStreamLength(pavia)$ determină numărul cadrelor fluxului audio, identificat prin variabila *pavia*, iar $AVIStreamLength(pavi)$ determină numărul cadrelor fluxului de imagini, identificat prin variabila *pavi*.

Procesul de sincronizare a fluxurilor native a unei date video implică identificarea cadrului de imagine ce trebuie afișat la un moment dat în procesul de derulare a secvenței. Pe baza acestei informații se poate afișa textul din fluxul de tip text:

```
sir = get_text(ncimg);
```

unde, *sir* este sirul returnat prin apelul funcției *get_text* care primește ca parametru numărul cadrului video *ncimg* ce este afișat în fereastră.

În cazul în care sirul returnat este *vid*, atunci în fereastră nu se afișează nici un text, iar când șirul nu este *vid* în fereastră, se va desena textul în partea de jos a imaginii vizualizate.

În figura 2 se ilustrează sincronizarea fluxurilor native a unei date de tip video cu un flux extern de tip text.



Figura 2 Afișarea sincronă a fluxului de tip text

CONCLUZII

Sincronizarea mediilor, în general, și a fluxurilor video și text, în special, are aplicabilitate într-o multitudine de domenii precum: globalizarea accesului la date

multimedia – “oriunde” și “oricând”, instruirea asistată, indexare video, biblioteci multimedia digitale.

Principala direcție de dezvoltare a cercetărilor este utilizarea secvențelor video drept componente hypermedia.

REFERINȚE

1. Dârdală, M., Smeureanu, I., Reveiu, A., *Tehnologii multimedia*, Editura ASE, București, 2008;
2. Demetriades G., *Streaming Media. Building and Implementing a Complete Streaming System*, Wiley Publishing, Inc., 2003;
3. Pagani M., *Encyclopedia of multimedia technology and networking*, Idea Group Inc., 2005;
4. Reveiu A., Surcel T., *Distribuirea de conținut multimedia accesibil*, Volumul de lucrări al Conferinței Naționale de Tehnologii Informatice Inclusive, ProInclusiv 2008, ASE București, 2008;
5. Rimmer, S., *Multimedia Programing for Windows*, McGraw-Hill, 1994;
6. Stein J., *Digital Signal Processing. A Computer Science Perspective*, John Wiley & Sons, Inc., 2000.