Revista Română de Interacţiune Om-Calculator 5 (1) 2012, 19-36

Detectarea automată a genurilor muzicale

Adrian Simion[1], Ştefan Trăuşan-Matu[1,2]

[1]Universitatea Politehnica din Bucureşti
Splaiul Independenţei nr. 313, sector 6, 060042, Bucureşti,
E-mail: simion.adrian@gmail.com

[2]Institutul de Crcetări în Inteligența Artificială
Calea 13 septembrie nr. 13, Bucureşti

Rezumat. Această lucrare descrie şi aplică diferite metode ce folosesc un calculator pentru a determina apartenența la un anumit gen muzical a unui fişier audio. Algoritmii au fost testați pe colecțiile de date MagnaTune şi MARSYAS, dar instrumentele software implementate pot fi folosite pe o gamă variată de surse. Instrumentele vor face parte dintr-un sistem software mai larg numit ADAMS (Advanced Dynamic Analysis of Music Software) dezvoltat de autori. Acest sistem este bazat pe biblioteca open-source MARSYAS şi conține un modul similar cu WEKA pentru sarcini de „data mining" şi „machine learning".

Cuvinte cheie: segmentare automată, clasificare audio, analiză muzicală bazată pe conținut, detectarea acordurilor, regiuni vocale şi instrumentale, MIR.

1. Introducere

Cantitatea de informație digitală a cunoscut o creştere exponențială, acest fapt fiind susținut de progresul tehnologiei dar, în cazul muzicii, şi de interesul utilizatorilor ce au la dispoziție din ce în ce mai multe surse de acces la acest tip de informație.

Trendul actual este explicat, în primul rând, de o componentă importantă a muzicii: "caracteristica existențială". Muzica din România poate fi apreciată în SUA, iar în Japonia putem găsi adepți ai muzicii clasice indiene. Această formă de exprimare poate fi înțeleasă şi apreciată fără etape intermediare, cum ar fi traducerea în cazul textelor. Putem afirma că „muzica este o formă de exprimare ce poate fi împărtăşită de oameni ce aparțin unor culturi diferite, deoarece aceasta depăşeşte granițele limbilor naționale sau a fundalului cultural." (Orio 2006)

În cele din urmă, muzica este o artă cultă şi populară în acelaşi timp, ba chiar, câteodată,este imposibil de a separa cele două aspecte, un exemplu fiind jazz-ul şi muzica tradiţională.

Disponibilitatea imediată şi cererea pentru conţinutul muzical a indus noi cerinţe pentru administrarea, publicitatea şi distribuirea acestuia. Din această cauză a apărut nevoia de o analiză mai sofisticată şi directă a conţinutului decât cea făcută prin intermediul meta-datelor catalogate de oameni.

Noile tehnici au permis abordări ce fuseseră folosite până în acel moment în analiza muzicală teoretică. Una dintre problemele esenţiale a fost prezentată de Frank Howes: „Există un fond vast de material muzical disponibil pentru studii comparative.” (Howes 1948) Ar fi foarte interesant de a descoperi şi de a stabili o corelare între muzică şi fenomenul social. Cu puterea de calcul actuală şi cu progresul făcut în această direcţie putem răspunde la întrebări de genul: Care este fondul etnic al unei anumite piese muzicale, din ce culturi face parte?

În lumina acestor posibilităţi şi dezvoltări ale tehnologiei aveam nevoie de o nouă disciplină care să acopere şi să răspundă diferitelor probleme. MIR (Music Information Retrieval) este o ştiinţă interdisciplinară ce îşi extrage informaţiile din muzică. Originea MIR se află în domenii ca: muzicologie, psihologie cognitivă, ştiinţa calculatoarelor şi lingvistică.

Un domeniu activ de cercetare este compus din noi metode şi instrumente pentru regăsirea modelelor sau pentru compararea conţinutului muzical. În acest sens a fost formată ISMIR (International Society for Music Information Retrieval) care este strâns legată de MIREX (Music Information Retrieval Evaluation eXchange). Sarcinile de evaluare includ detectarea automată a genurilor, detectarea acordurilor, segmentarea, extragerea liniilor melodice, interogarea prin fredonare, doar ca să amintim câteva. Această lucrare se va concentra mai mult pe detectarea automată a genurilor şi pe segmentarea automată.

În cazul de faţă, segmentarea automată presupune împărţirea unei surse audio în regiuni omogene. Rezultatele obţinute în urma unei detecţii cat mai eficiente ar putea face clasificarea informaţiei audio mult mai facilă, sau chiar o pot automatiza complet.

2. Studii anterioare şi înrudite legate de segmentarea automată muzicală

Subiectul clasificării limbajului/muzicii a fost studiat de către mulţi cercetători. Chiar dacă aplicabilitatea rezultatelor poate fi foarte variată, multe din aceste studii folosesc colecţii similare de caracteristici, cum ar fi „energia de scurtă durată", ZCR – Zero-Crossing-Rate (Kulkarni 2007), coeficienţi cepstrali, desfăşurarea spectrală, centroidul şi intensitatea sonoră, alături de unele caracteristici unice, cum ar fi „dinamismul". Un cepstrum, în general, se obţine prin aplicarea transformatei Fourier inverse pe logaritmul spectrului unui semnal. Totuşi, combinaţiile de caracteristici ce au fost folosite pot diferi foarte mult, la fel ca şi mărimea colecţiei.

În mod uzual câteva caracteristici de termen lung, cum ar fi media sau variaţia, şi nu caracteristicile propriu-zise sunt folosite pentru discriminare.

Diferenţele majore dintre diferitele studii constau în algoritmul pentru clasificarea exactă, chiar dacă în mod normal se folosesc unii clasificatori consacraţi ca bază de pornire: KNN (K-Nearest Neighbor), reţele neuronale sau gausieni multidimensionali.

Pentru studii, de obicei, sunt folosite diferite colecţii de date pentru antrenarea şi testarea algoritmului. Merită menţionat că pentru aceste studii, în special pentru cele timpurii, colecţiile de date erau destul de mici. Următorul tabel descrie câteva din studiile precedente.

*Tabel 1.*Câteva studii anterioare

Autor	Aplicabilitate	Caracteristici	Metoda de clasificare
Saunders, 1996	Monitorizare radio automatizată în timp real	STE (short-time energy), parametrii statistici ZCR	Clasificator multispaţial gausian
Scheirer şi Slaney, 1997	Separarea discursului/muzicii pentru recunoaşterea automată a limbajului	13 caracteristici temporale, spectrale şi cepstrale (e.g., modulare la 4Hz, % din cadre cu energie joasa, derularea spectrală, spectroid central, fluxul spectral, ZCR, "ritmicitate")	GMM (Gaussian Mixture Model), (KNN), arbori K-D, multidimensional, estimator gausian MAP
Foote, 1997	Regăsirea documentelor audio pe baza asemănărilor acustice	12 MFCC, STE (short time energy)	Regăsirea modelelor cu ajutorul histogramelor, cuantificator de vectori bazat pe arbori antrenat pentru a maximiza informaţia reciproca
Liu et al., 1997	Analiza semnalelor audio pentru clasificarea scenelor din programele	Coeficient pentru linişte, câmp dinamic pentru volum, frecvenţă 4Hz, media şi diferenţa de înălţime,	O reţea neuronală ce foloseşte o structură de tip OCON (One-class-in-one

	televizate	raporturi pentru zgomot, centroidul frecvenţelor, lăţimea de bandă, energia în 4 sub-benzi	network)
Zhang şi Kuo, 1999	Segmentarea/regăsirea audio pentru clasificarea scenelor video, indexarea înregistrărilor audio, parcurgerea bazelor de date	Caracteristici bazate pe energia de timp scurt, media ZCR, frecvenţa fundamentală pe durate scurte de timp	O procedură bazată pe reguli euristice pentru prima etapă, HMM pentru a doua etapă
Williams şi Ellis, 1999	Segmentarea discursului vs. segmentarea semnalelor non verbale în sarcinile de recunoaştere automată a limbajului	Entropia medie per-cadru şi probabilitatea medie „dinamică", coeficient pt. energia de fundal, HMM	Raportul de probabilitate Gaussiană
El-Maleh al., 2000	Programarea automată şi regăsirea vizuală/audio bazată pe conţinut	LSF, LSF diferenţial, filtru trece sus bazat pe măsurători ZCR	Clasificatori KNN şi QCG (Quadratic Classifier Gaussian)
Bugatti et al., 2002	„Cuprins" pentru un document multimedia	Caracteristici bazate pe ZCR, flux spectral, energie pe termen scurt, coeficienţi cepstrali, centroidul spectral, dimensiunea frecvenţei silabice, raportul frecvenţei înalte a spectrului puterii	Clasificator multispaţial gausian, reţele neuronale MLP
Lu, Zhang, şi Jiang, 2002	Analiza conţinutului audio în segmentarea video	Perechi liniare spectrale, periodicitatea benzilor, raport zgomot-cadru (NFR – noise-frame ratio), HZCRR (High zero-crossing rate ratio)	Clasificare în 3 paşi: 1. KNN şi perechi spectrale liniare – cuantificări vectoriale (LSP-VQ) pentru discriminarea discurs / non-discurs. 2. Reguli euristice pentru clasificarea non-discursului în muzică/zgomot de fond/linişte. 3. Segmentarea limbajului
Ajmera et al., 2003	Transcrierea automată a ştirilor radiofonice	Măsurarea entropiei medii şi „dinamismului" estimat la ieşirea unui perceptron multistrat (MLP – multilayer perceptron) antrenat pentru a emite probabilităţi posterioare. Date de intrare MLP: primii 13 coeficienţi cepstrali ai ordinului al 12-a din filtrul de predicţie perceptual liniar	HMM format din 2 stări cu limitări de durată (fără prag, supervizat, nesupervizat).
Burred şi Lerch, 2004	Clasificarea audio (limbaj/ muzică /	Măsurători statistice a cadrelor pe timp scurt: ZCR, desfăşurare/flux	

	zgomot de fond), clasificarea muzicală în genuri	centroid spectral, primii 5 MFCC, netezimea/centroidul spectrului audio, intensitatea ritmului, regularitatea ritmică, energia RMS, dimensiunea temporală, intensitatea sonoră	Clasificator KNN, clasificator GMM format din 3 componente
Barbedo şi Lopes, 2006	Segmentarea automată pentru aplicaţii în timp real	Caracteristici bazate pe ZCR, derularea spectrală, intensitate şi frecvenţe fundamentale	KNN, SOM, reţele neuronale MLP, combinaţii liniare
Mu˜noz- Exp´ osito et al., 2006	Sistem inteligent pentru codarea audio automata	Centroid spectral LPC răsucit	GMM format din 3 componente cu sau fără sistem bazat pe reguli fuzzy
Alexandre et al, 2006	Clasificarea discursului / muzicii pentru clasificarea genurilor muzicale	Desfăşurarea/centroidul spectral, ZCR, energie pe termen scurt, LSTER (low short time energy ratio), MFCC, VTW (voice to white)	Discriminant liniar Fisher, KNN

3. Semnale digitale audio

În momentul în care muzica este înregistrată, presiunea continuă a undei sonore este măsurată cu ajutorul unui microfon. Aceste măsurători sunt luate la intervale regulate de timp şi fiecare măsurătoare este cuantificată.

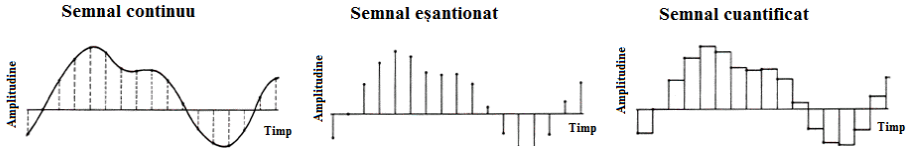

Figura 1. Reprezentare digitală a sunetului (în timp)

a. Muzica este un semnal b. care este eşantionat c. şi cuantificat continuu;

Sunetele pot fi reprezentate ca o sumă de sinusoide. Un semnal format din N eşantioane poate fi scris astfel:

$$x = \sum_{k=0}^{N/2} a_k^{(r)} \cos(2\pi(\frac{k}{N})) + a_k^{(i)} \sin(2\pi(\frac{k}{N})). \quad (1)$$

Semnalul poate fi reprezentat în domeniul *frecvență* folosind coeficienții $\{(a_1^{(y)}, a_1^{(i)}),...,(a_{N/2}^{(y)}, a_{N/2}^{i})\}$.

Amplitudinea şi faza frecvenței k sunt date de:

$$X_M[k] = \sqrt{(a_k^{(r)})^2 + (a_k^{(i)})^2} \qquad (2)$$

$$X_p[k] = \arctan(\frac{a_k^{(i)}}{a_k^{(r)}}) \quad (3)$$

Studii perceptuale bazate pe auzul uman demonstrează că informația dată de fază este relativ neimportantă atunci când aceasta este comparată cu informația preluată din amplitudine, astfel componenta fazică este de obicei ignorată în momentul extragerii caracteristicilor. [19]

Spectroidul central este o altă caracteristică spectrală ce este utilă în procesul de extragere şi de analiză. În tabelul 1 se pot observa diferitele utilizări. Spectroidul central este centrul de gravitate al spectrului şi este dat de:

$$C = \frac{\sum_{k=1}^{N/2} X_M[k] * k}{\sum_{k=1}^{N/2} X_M[k]} \qquad (4)$$

Spectroidul central poate fi înțeles ca fiind o unitate de măsură a luminozității deoarece cântecele sunt considerate mai luminoase atunci când au componente ce conțin frecvențe mai înalte.

3.1 Transformări în domeniile timp-frecvență

În MIR şi în analiza sunetelor în general este foarte uzual să facem o transformare între domeniile timp şi frecvență. Pentru acest lucru aparatul matematic ne dă transformata Fourier discretă, transformata Fourier rapidă, transformata Fourier, transformata cosinus discretă, transformata undină discretă şi transformata gammaton.

Analiza muzicală nu se ocupă cu transformarea în mulțimea complexă, deoarece muzica este întotdeauna o serie de valori reale în timp şi are doar frecvențe pozitive.

Fiind dat un semnal x cu N eşantioane, funcțiile de bază pentru transformata Fourier discretă va fi N/2 unde sinus şi N/2 unde cosinus ce corespund coeficienților anteriori.

Operatorul de proiecție este corelarea, o unitate a similarității între două serii temporale. Coeficienții pot fi aflați folosind:

$$a_k^{(r)} = \frac{2}{N}\sum_{i=0}^{N-1} x[i]\cos(2\pi\frac{k}{N}i) \quad (5)$$

$$a_k^{(i)} = -\frac{2}{N}\sum_{i=0}^{N-1} x[i]\sin(2\pi\frac{k}{N}i) \quad (6)$$

Transformata Fourier discretă este calculată într-un mod eficient cu ajutorul transformatei Fourier rapide. O limitare a celor două reprezentări, cea temporală şi cea spectrală, este dată de faptul ca niciuna dintre reprezentări nu prezintă simultan informația obținută din analiza frecvenței şi din analiza timpului. O reprezentare timp-frecvență este găsită folosind transformata Fourier de timp-scurt: Mai întâi, semnalul audio este divizat în secvențe (suprapuse) de segmente. Fiecare segment este multiplicat de o *funcție fereastră*.

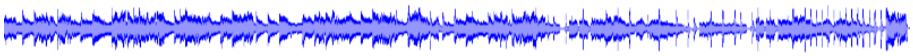

Figura 2. Magnatune apa_ya-apa_ya-14-maani-59-88.wav (domeniu temporal)

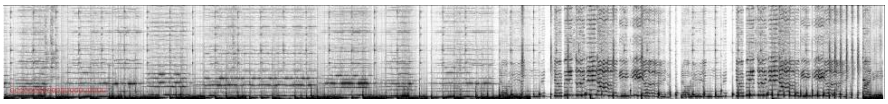

Figura 3. Magnatune apa_ya-apa_ya-14-maani-59-88.wav (spectrogramă)

Figurile 2 şi 3 au fost obținute cu ajutorul unei versiuni modificate a executabilului MARSAYS sound2png cu ajutorul următoarelor comenzi:

- ./sound2png -m waveform ../audio/magnatune/0/apa_ya-apa_ya-14-maani-59-88.wav ../saveres/magnatunewav.png -ff Adventure.ttf
- ./sound2png -m spectogram ../audio/magnatune/0/apa_ya-apa_ya-14-maani-59-88.wav ../saveres/magnatunespec.png -ff Adventure.ttf

O altă transformare utilă este cea bazată pe transformata undină.

3.2 MFCC (Mel-Frequency Cepstral Coefficients)

Cel mai uzual set de caracteristici folosit în recunoaşterea limbajului şi în sistemele de adnotare muzicală sunt MFCC. Aceştia sunt de fapt trăsături de scurtă durată ce caracterizează amplitudinea spectrală a semnalului audio. Pentru fiecare interval scurt de timp (25 ms), este găsit vectorul de caracteristici folosind un algoritm explicat în Alg1. Primul pas este de a obține amplitudinea fiecărei frecvențe în domeniul frecvență folosind transformata Fourier discretă. După acest pas se logaritmează amplitudinea deoarece intensitatea perceptuală s-a dovedit a fi aproximativ logaritmică. În pasul următor componentele ce formează frecvența sunt unite în 40 de cadre ce au fost determinate conform scării Mel. Scara Mel este maparea dintre frecvența reală şi modelul perceput de frecvență ca fiind aproximativ logaritmic. Deoarece o secvență temporală a acestor vectori 40-dimensionali Mel-frecvență ar fi fost foarte redundantă, am putea reduce aceste dimensiuni folosind PCA. În locul acestei abordări, comunitatea a adoptat transformata cosinus discretă, ce aproximează PCA dar nu are nevoie de date de antrenament, pentru a reduce numărul de dimensiuni la un număr de 13 MFCC. (Turnbull 2005)

```
1: Calcularea spectrului folosind transformata Fourier discretă
2: Logaritmarea spectrului
3: Aplicarea scării Mel netezirea
4: Decorelare folosind transformata cosinus discretă
```

*Algoritmul 1.*Calcularea vectorului de caracteristici MFCC

4. Descrierea problemei

O caracteristică obişnuită ce-i ajută pe producătorii de discuri să răspundă cerințelor celor din publicul țintă, pe muzicologi să studieze influențele muzicale şi pe entuziaşti să-şi organizeze colecțiile muzicale este identificarea genurilor.

Conceptul de gen este în mod natural subiectiv deoarece influențele, ierarhia şi intersecția unui cântec cu un anumit gen nu este un lucru stabilit de comun acord. Acest punct de vedere este susținut de o comparație a trei furnizori de servicii muzicale online ce a arătat că există mari diferențe în

numărul genurilor, cuvintele ce le descriu şi structura ierarhiei acestora.(Pachet 2000)

Chiar dacă există multe contradicții cauzate de natura subiectivă, conceptul de gen a stârnit interesul comunității MIR. Diferite lucrări legate de acest subiect reflectă presupunerile autorilor în legătură cu genurile. Legile drepturilor de autor au împiedicat cercetătorii să stabilească o bază de date comună de cântece, făcând foarte dificilă compararea rezultatelor.

5. Descrierea experimentului

Colecțiile de date folosite pentru antrenament şi testare au fost MAGNATUNE (2012) şi două colecții ce au fost construite în etapele timpurii ale MARSYAS.

Autorii au creat un sistem software plecând de la biblioteca open-source MARSYAS pentru a facilita acest experiment şi pentru alte sarcini de analiză audio ulterioare. Deoarece sistemul ADAMS (Advanced Dynamic Analysis of Music Software) este construit într-un design modular, diferitele sarcini (descrise mai jos) pot fi automatizate, iar sunetul poate „curge" prin aceste module până când este făcută întreaga analiză.

Structura directorului ADAMS poate fi observata în figura următoare:

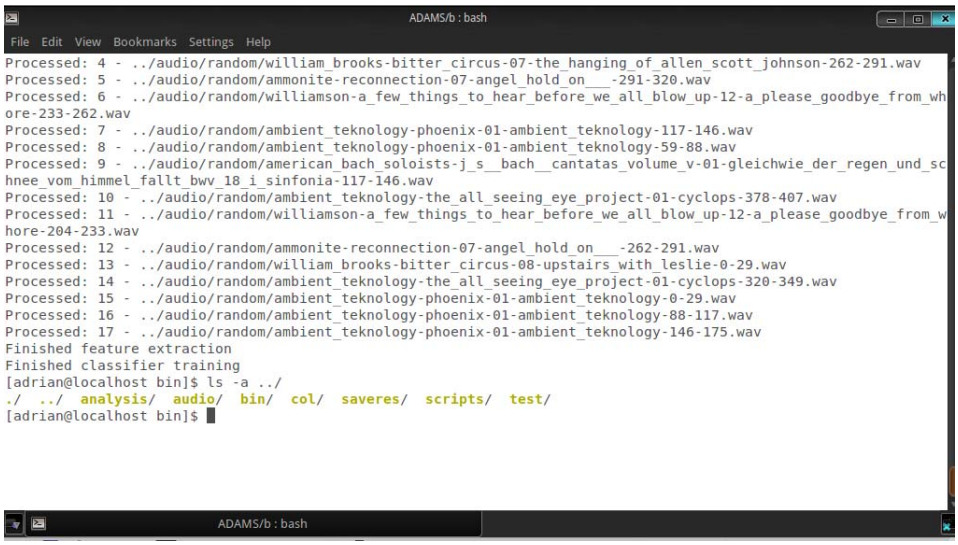

Figura 4. Structura principală a sistemului ADAMS

Sarcinile de „învățare automată” (machine learning) au fost realizate prin intermediul WEKA (2012), încărcând fişierele compatibile arff create cu ajutorul MARSYAS.

Sistemul de operare ales pentru aceste experimente a fost Mandriva Linux 2011, versiunea compilatorului fiind „gcc (GCC) 4.6.1 20110627 (Mandriva)”.

Au fost folosiți următorii extractori:

- BEAT: Caracteristici histogramice legate de ritm (Beat histogram features)

- LPCC: Coeficienți cepstrali LPC derivați (LPC derived Cepstral coefficients)

- LSP: Perechi spectrale liniare (Linear Spectral Pairs)

- MFCC: Mel-Frequency Cepstral Coefficients

- SCF: Spectral Crest Factor (MPEG-7)

- SFM: Unitate de măsură spectrală a netezimii (Spectral Flatness Measure MPEG-7)

- SFMSCF: caracteristici SCF şi SFM

- STFT: Centroid, Desfăşurare spectrală, Flux, ZC – Valori pozitive (Zero-Crossings)

- STFTMFCC: Centroid, Flux-Desfăşurare spectrală, ZC, MFCC

La fiecare experiment pentru extractorii specificați sunt prezentate de asemenea matricele de confuzie pentru a avea o idee despre clasificarea reală şi cea prezisă realizată cu ajutorul sistemului de clasificare.

5.1 Experimentul 1: Clasificarea folosind „Caracteristici timbrale”

Acest experiment foloseşte următorii extractori: TZC (Time Zero-Crossings), Centroidul Spectral, Flux-Desfăşurare spectrală şi MFCC. Extragem aceste caracteristici cu opțiunea –timbral şi creăm de asemenea un fişier ce va fi încărcat cu ajutorul mediului WEKA pentru analiză cu ajutorul comenzii:

```
./adamsfeature -sv -timbral ../col/all.mf -w
    ../analysis/alltimbral.arff
```

Pentru celelalte etape s-au folosit comenzi similare. În urma experimentelor s-au ales următorii clasificatori: BN (Bayes Network), NB (Naive Bayes), DT (Decision Table), FC (Filtered Classifier) şi NNGE.

Rezultatele sunt evidenţiate în următorul tabel:

*Tabel 2.*Caracteristici Timbrale – Rezultatele clasificatorilor

Clasificator	Timp Cons model	Clasif. corect	Clasif. incorect	Er. Med. Abs.	Er. Med. Pătr.	Er. Rel. Abs.	Er. Rel. Pătr.
Bayes Network	1.78	62.5%	37.5%	0.0753	0.2648	41.82%	88.28%
Naïve Bayes	0.04	55%	45%	0.0902	0.2925	50.09%	97.51%
Decision Table	15.49	51.6%	48.4%	0.1467	0.2599	81.53%	86.64%
Filtered Classifier	4.55	87.8%	12.2%	0.0348	0.1318	19.31%	43.94%
NNGE	10.69	100%	0%	0	0	0	0

Tabelul 2 a fost construit încărcând fişierul alltimbral.arff în WEKA şi antrenând clasificatorii existenţi.

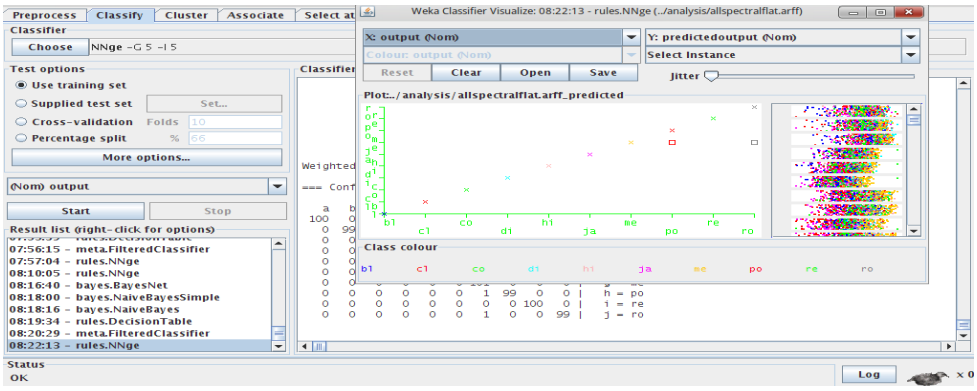

Figura 5. Grafic de predicţie a erorilor în mediul WEKA

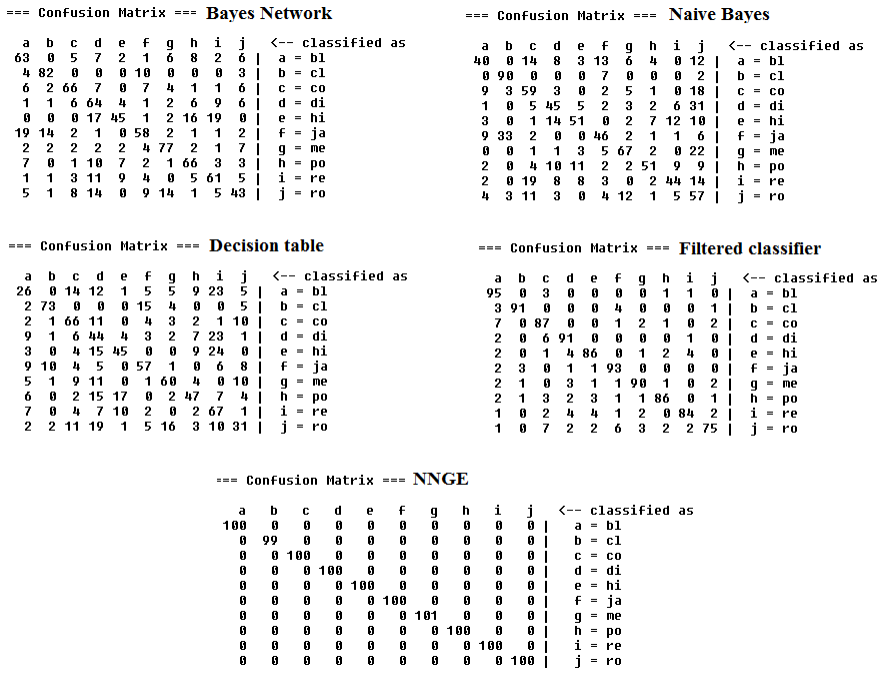

Figura 6. Matricele de confuzie pentru clasificarea cu ajutorul caracteristicilor timbrale

5.2 Experimentul 2: Clasificare folosind „Caracteristici spectrale”

Acest experiment foloseşte următorii extractori: Centroidul Spectral şi Flux-Desfăşurare spectrală.

Folosind aceiaşi clasificatori rezultatele sunt:

*Tabel 3.*Caracteristici Spectrale – Rezultatele clasificatorilor

Clasificator	Timp Cons model	Clasif. corect	Clasif. incorect	Er. Med. Abs.	Er. Med. Pătr.	Er. Rel. Abs.	Er. Rel. Pătr.
Bayes Network	1.78	46.5%	53.5%	0.1192	0.2742	66.21%	91.41%
Naïve Bayes	0.23	42.5%	57.5%	0.1205	0.2924	66.92%	97.47%
Decision Table	0.72	46.1%	53.9%	0.1491	0.2655	82.82%	88.49%
Filtered Classifier	0.41	63.6%	36.4%	0.099	0.2225	54.98%	74.15%
NNGE	20.2	100%	0%	0	0	0	0

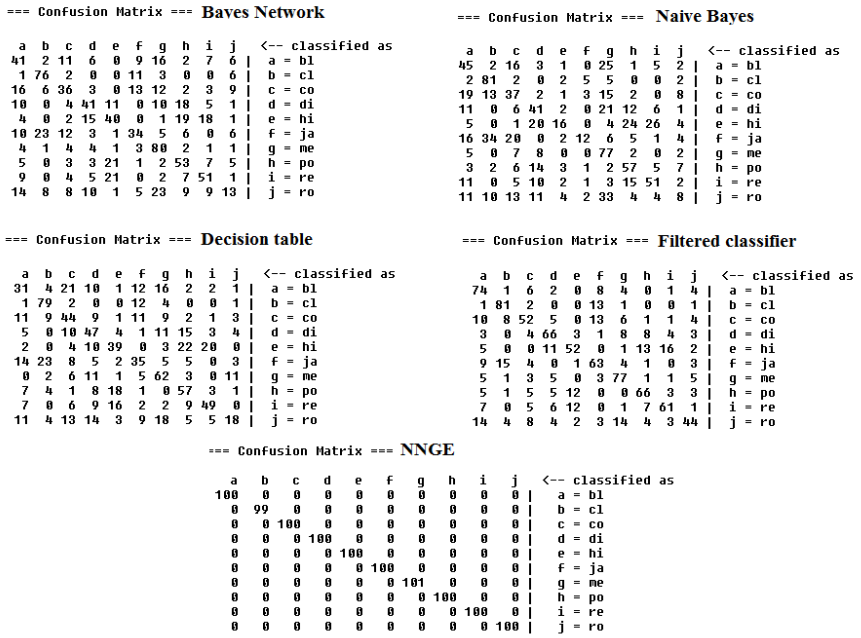

Figura 7. Matricele de confuzie pentru clasificare folosind caracteristici spectrale

5.3 Experimentul 3: Clasificare folosind „MFCC”

Acest experiment foloseşte extractorii bazați pe MFCC (Mel-Frequency Cepstral Coefficients).

*Tabel 4.*Caracteristici MFCC – Rezultatele clasificării

Clasificator	Timp Cons model	Clasif. corect	Clasif. incorect	Er. Med. Abs.	Er. Med. Pătr.	Er. Rel. Abs.	Er. Rel. Pătr.
Bayes Network	1.23	63.3%	36.7%	0.0764	0.2475	42.42%	82.50%
Naïve Bayes	0.22	58.5%	41.5%	0.0847	0.2694	47.07%	89.80%
Decision Table	6.4	49.1%	50.9%	0.1481	0.2638	82.27%	87.94%
Filtered Classifier	0.81	87.1%	12.9%	0.0363	0.1348	20.18%	44.92%
NNGE	3.74	99.8%	0.2%	0.0004	0.02	0.22%	6.66%

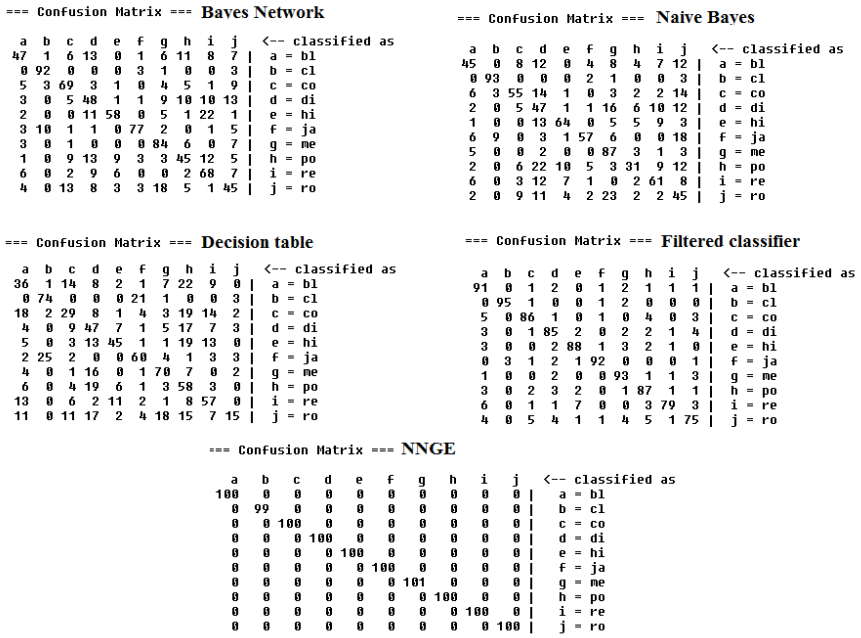

Figura 8. Matricele de confuzie pentru clasificarea bazată pe caracteristici MFCC

5.4 Experimentul 4: Clasificare folosind „ZC”

Acest experiment foloseşte extractorii bazaţi pe ZC.

*Tabel 5.*Caracteristici ZC – Rezultatele clasificării

Clasificator	Timp Cons model	Clasif. corect	Clasif. incorect	Er. Med. Abs.	Er. Med. Pătr.	Er. Rel. Abs.	Er. Rel. Pătr.
Bayes Network	0.09	34.7%	65.3%	0.1437	0.2789	79.83%	82.50%
Naïve Bayes	0.01	34.5%	65.5%	0.1441	0.2869	80.06%	89.80%
Decision Table	0.22	42.4%	57.6%	0.1511	0.2691	83.95%	87.94%
Filtered Classifier	0.15	44%	56%	0.1403	0.2649	77.94%	44.92%
NNGE	0.52	99.8%	0.2%	0.0004	0.02	0.22%	6.66%

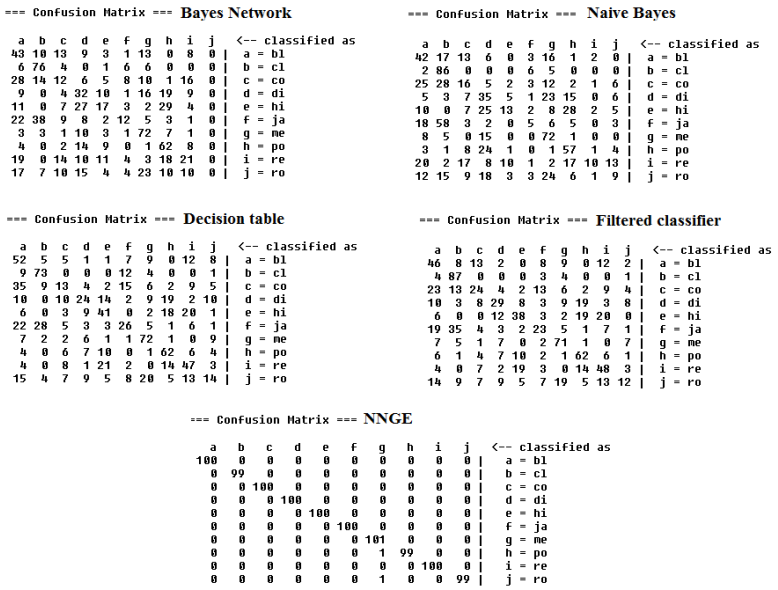

Figura 9. Matricele de confuzie pentru clasificarea bazată pe caracteristici ZC

5.5 Experimentul 5: Clasificare folosind „SFM”

*Tabel 6.*Caracteristici SFM – Rezultatele clasificării

Clasificator	Timp Cons model	Clasif. corect	Clasif. incorect	Er. Med. Abs.	Er. Med. Pătr.	Er. Rel. Abs.	Er. Rel. Pătr.
Bayes Network	1.78	58.4%	41.6%	0.0838	0.2738	46.53%	91.28%
Naïve Bayes	0.15	53.2%	46.8%	0.0935	0.294	51.96%	97.99%
Decision Table	12.35	50.4%	49.6%	0.1472	0.2621	81.78%	87.37%
Filtered Classifier	2.1	83.8%	16.2%	0.045	0.15	25.01%	50.12%
NNGE	9.24	99.8%	0.2%	0.0004	0.02	0.22%	6.66%

Acest experiment foloseşte extractorii bazaţi pe SFM (Spectral Flatness Measure).

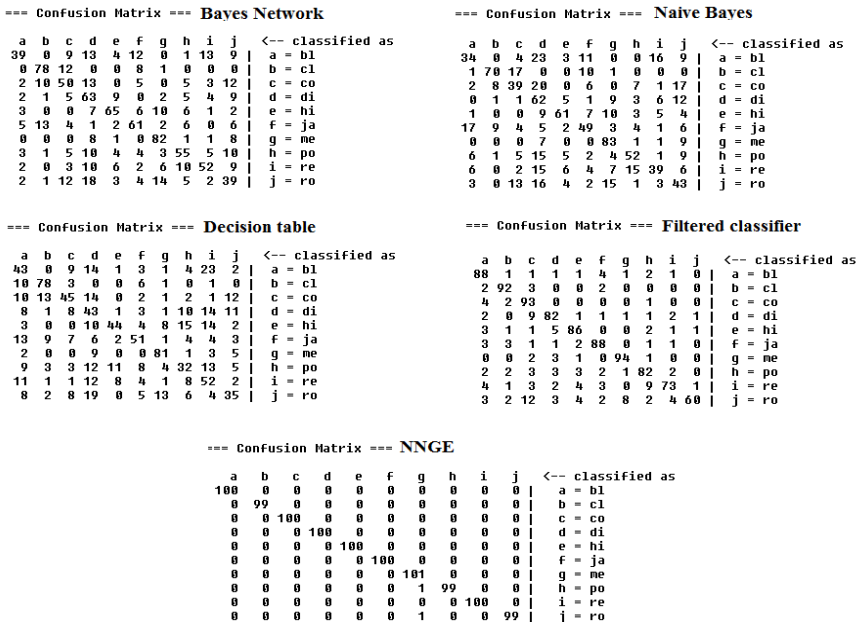

Figura 10. Matricele de confuzie pentru clasificarea bazată pe caracteristici SFM

6. Concluzii

Au fost făcute cinci experimente pentru determinarea genului muzical al fişierelor audio. Caracteristicile extrase au variat de la un experiment la altul pentru a determina care se pretează mai bine pentru colecţia de date folosită. Cei cinci clasificatori au dat rezultate diferite în funcţie de caracteristicile extrase, iar acestea au fost evaluate cu unelte software de „machine-learning" bine-cunoscute cum ar fi WEKA şi de asemenea cu un sistem de analiză dezvoltat pe baza bibliotecii MARSYAS.

Rezultatele arată că se pot obţine concluzii satisfăcătoare chiar şi prin abordări simpliste cum ar fi clasificarea NB (Naive Bayes), dar rezultate cu o relevanţă mai mare au fost obţinute folosind tehnici mai avansate. Faptul ca NN (cel mai apropiat vecin – Nearest Neighbor) a produs rezultate foarte bune nu înseamnă ca va avea acelaşi comportament pe o altă colecţie de date.

Îmbunătățiri ale metodelor prezentate pot fi obținute testându-le pe o colecție extinsă de date şi determinând influențele intrinsece ale fiecărui gen asupra altuia.

Concluziile acestor influențe pot avea un înțeles mai mare din punct de vedere social, cum ar fi analiza blues-ului şi modul în care au apărut derivatele sale. În unele cazuri putem găsi rezultate improbabile cum ar fi că genul muzical death-metal îşi are rădăcinile în muzica jazz.

Referințe

Ajmera J., McCowan I., şi Bourlard H., *Speech/music segmentation using entropy and dynamism features in a HMM classification framework*, Speech Communication, vol. 40, nr. 3, pp. 351-363, 2003

Alexandre E., Rosa M., L. Caudra, şi Gil-Pita R., Application of Fisher linear discriminant analysis to speech/music classification, Proceedings of the 120th Audio Engineering Society Convention (AES '06), Paris, France, Mai 2006, nr 6678

Barbedo J. G. A. şi Lopes A, *A robust and computationally efficient speech/music discriminator,* Journal of the Audio Engineering Society, vol. 54, nr. 7-8, pp. 571–588, 2006

Bugatti A., Flammini A., şi Migliorati P., *Audio classification in speech and music: a comparison between a statistical and a neural approach*, EURASIP Journal on Applied Signal Processing, vol. 2002, nr. 4,pp. 372–378, 2002.

Burred J. J. şi Lerch A., *Hierarchical automatic audio signal classification*, Journal of the Audio Engineering Society, vol. 52, no. 7-8, pp. 724–739, 2004

CONFUSION MATRIX

http://www2.cs.uregina.ca/~hamilton/courses/831/notes/confusion_matrix/confusion_matrix.html (Vizitat pe 2012/01/23)

El-Maleh K., Klein M., Petrucci G., şi Kabal P., *Speech/music discrimination for multimedia applications*, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00), vol. 6, pp. 2445–2448, Istanbul, Turcia, Iunie 2000.

Foote J. T., *A similarity measure for automatic audio classification*, in Proceedings of the AAAI Spring Symposium on Intelligent Integration and Use of Text, Image, Video, and Audio Corpora, Stanford, Calif, USA, Martie 1997.

Howes, F. *Man Mind and Music. Marin* Secker & Warbug LTD., 1948

ISMIR. http://www.ismir.net/ (Vizitat pe 2012/01/23)

Kulkarni K., Iyer D. si Sridharan S.R., *Audio Segmentation*, Stanford University, Stanford, 2007

Liu Z., Huang J., Wang Y., şi Chen I. T., *Audio feature extraction and analysis for scene classification*, Proceedings of the 1st IEEE Workshop on Multimedia Signal Processing

(MMSP ’97), pp. 343–348, Princeton, NJ, USA, Iunie 1997.

[19] Logan B., *Mel-Frequency Cepstral Coefficients for music modeling*, ISMIR ’00: International Symposium on Music Information Retrieval, 2000

Lu L., Zhang H.-J., şi Jiang H., *Content analysis for audio classification and segmentation*, IEEE Transactions on Speech and Audio Processing, vol. 10, no. 7, pp. 504–516, 2002.

Mangatune. http://tagatune.org/Magnatagatune.html (Vizitat pe 2012/01/23)

MARSYAS. http://marsyas.info/ (Vizitat pe 2012/01/23)

MIREX http://www.music-ir.org/mirex/wiki/MIREX_HOME

(Vizitat pe 2012/01/23)

Munoz-Exposito J. E, Galan S. G., Reyes N. R., P. V. Candeas şi F. R. Pena, *A fuzzy rules-based speech/music discrimination approach for intelligent audio coding over the Internet*, Proceedings of the 120th Audio Engineering Society Convention (AES ’06), Paris, Franța, Mai 2006, nr. 6676

Orio N., *Music Retrieval: A Tutorial and Review*, Department of Information Engineering, University of Padova, Via Gradenigo, 6/b, Padova 35131, Italy, 2006

Pachet F. şi Cazaly D., *A taxonomy of musical genres*, RIAO ’00: Content-Based Multimedia Information Access, 2000.

Saunders J., *Real-time discrimination of broadcast speech/music*, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’96), vol. 2, pp. 993–996, Atlanta, Ga, USA, Mai 1996.

Scheirer E. şi Slaney M., *Construction and evaluation of a robust multifeature speech/music discriminator*, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’97), vol. 2, pp. 1331–1334, Munich, Germany, Aprilie 1997.

Turnbull D., *Automatic music annotation*, Department of Computer Science, UC San Diego, 2005

WEKA. http://www.cs.waikato.ac.nz/ml/weka/ (Vizitat pe 2012/01/23)

Williams G. şi Ellis D. P. W., *Speech/music discrimination based on posterior probability features*, Proceedings of the 6th European Conference on Speech Communication and Technology (EUROSPEECH ’99), pp. 687–690, Budapest, Ungaria, Septembrie 1999.

Zhang T. şi Kuo C.-C. J., *Hierarchical classification of audio data for archiving and retrieving*, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’99), vol. 6, pp. 3001–3004, Phoenix, Ariz, USA, Martie 1999.