



**UNIVERSITATEA NAȚIONALĂ
DE ȘTIINȚĂ ȘI TEHNOLOGIE
POLITEHNICA BUCUREȘTI**



**Școala Doctorală de Electronică, Telecomunicații și
Tehnologia Informației**

**REZUMAT
TEZĂ DE
DOCTORAT**

Bogdan MOROȘANU

**CONTRIBUȚII ÎN DOMENIUL PRODUCȚIEI MUZICALE
FOLOSIND REȚELE NEURONALE**

**CONTRIBUTIONS IN THE FIELD OF MUSIC PRODUCTION
USING NEURAL NETWORKS**

COMISIA DE DOCTORAT

Prof. Univ. Politehnica din București	Președinte
Prof. dr. ing. Constantin PALEOLOGU Univ. Politehnica din București	Conducător de doctorat
Prof. dr. ing. Cristian NEGRESCU Univ. Politehnica din București	Referent
Prof. dr. ing. Radu-Mihnea UDREA Univ. Politehnica din București	Referent
Dr. ing. Daiana-Amelia SARU Univ. Politehnica din București	Referent

BUCUREȘTI 2025

Abstract

Această lucrare, intitulată "Contribuții în Domeniul Producției Muzicale Folosind Rețele Neuronale", propune soluții inovatoare la provocările contemporane ale producției audio, într-un context caracterizat de o creștere exponențială a conținutului muzical și de resurse tehnice limitate, în special în România. Cercetarea abordează problema discrepanței semnificative dintre numărul crescut de artiști și producții muzicale și infrastructura tehnică profesională disponibilă, explorând potențialul rețelelor neuronale de a optimiza procesele critice din lanțul de producție audio. Lucrarea dezvoltă o fundamentare teoretică solidă a principiilor percepției auditive și a tehnicilor de producție muzicală, înainte de a propune patru contribuții originale majore: un sistem de recunoaștere automată a instrumentelor muzicale în înregistrări complexe; un algoritm de control personalizat al nivelului pentru înregistrări multi-canal, bazat pe optimizarea intensității percepțuale; un compresor multi-bandă cu învățare automată a parametrilor, capabil să adapteze dinamic setările în funcție de materialul sonor; și o tehnologie de expansiune spațială de la stereo la multi-canal, utilizând un model hibrid de separare a surselor și extracție ambientală. Fiecare dintre aceste soluții a fost implementată, testată și evaluată atât prin metrice obiective, cât și prin evaluări subiective realizate de profesioniști din industria muzicală. Rezultatele demonstrează că integrarea rețelelor neuronale în fluxurile de lucru audio poate îmbunătăți semnificativ eficiența și calitatea producțiilor muzicale, oferind inginerilor de sunet și producătorilor instrumente avansate, accesibile și intuitive pentru a face față provocărilor pieței muzicale actuale. Contribuțiile aduse prin această cercetare au potențialul de a democratiza accesul la calitate sonoră profesională și de a stimula competitivitatea producțiilor muzicale românești pe piața internațională, în contextul unei industrii în care volumul de conținut și așteptările calitative sunt în continuă creștere.

Cuprins

1	Introducere	1
1.1	Prezentarea domeniului tezei de doctorat	1
1.2	Scopul tezei de doctorat	2
1.3	Conținutul tezei de doctorat	2
2	Fundamentele percepției și analizei sonore în producția audio	3
2.1	Ascultarea critică	3
2.2	Urechea și psihoacustica în contextul producției audio	4
2.2.1	Anatomia și funcționarea sistemului auditiv	4
2.2.2	Principii fundamentale de psihoacustică	4
2.2.3	Funcția de transfer a capului (HRTF)	5
2.3	Camera de control	6
3	Producția muzicală	7
3.1	Procesul de mixaj	7
3.1.1	Stabilirea proporțiilor sonore fundamentale	8
3.1.2	Panoramarea	8
3.1.3	Procesarea spectrală	8
3.1.4	Procesarea dinamică	9
3.2	Procesul de masterizare	9
3.2.1	Echilibrul tonal	10
3.2.2	Saturație	10
3.2.3	Optimizarea tăriei sonore	10
3.3	Concluzii	11
4	Rețele neuronale în producția audio	13
4.1	Clasificarea instrumentelor muzicale	13
4.2	Controlul personalizat al nivelului pentru înregistrări multi-canal	15
4.3	Compresorul multi-bandă cu învățare automată a parametrilor	16
4.4	Expandare spațială de la stereo la multi-canal	18

Cuprins

5	Concluzii	21
5.1	Rezultate obținute	21
5.2	Contribuții originale	23
5.3	Lista lucrărilor originale	24
5.4	Perspectivă de dezvoltare ulterioară	25
	Bibliografie	29

Capitolul 1

Introducere

Domeniul producției muzicale traversează o perioadă de transformări fără precedent, caracterizată prin democratizarea instrumentelor de producție și o revoluție digitală în distribuția conținutului. În 2022, peste 34,1 milioane de piese noi au fost încărcate pe platformele de streaming [50], iar la nivel mondial aproximativ 120.000 de piese noi sunt adăugate zilnic [7], estimându-se un total anual de peste 43 de milioane de piese.

Piața globală include peste 3 milioane de artiști doar pe Spotify [41], iar conform UNESCO, aproape 4 milioane de locuri de muncă sunt direct legate de industria muzicală [3]. În România, datele indică peste 7.000 de lansări și 5.000 de artiști unici în 2022 [39], cu specificul că piața locală operează cu resurse tehnice limitate și un deficit de specialiști în ingineria de sunet avansată.

1.1 Prezentarea domeniului tezei de doctorat

Teza se poziționează la intersecția a trei domenii: ingineria audio, inteligența artificială și producția muzicală contemporană. Ingineria audio a migrat de la paradigma analogică la ecosisteme predominant digitale, în timp ce producția muzicală operează într-un mediu extrem de competitiv, cu volume masive de conținut.

În România sunt aproximativ 10.000 de muzicieni activi, din care 88% sunt interpreți emergenți sau de nișă [38], în contrast cu doar 600-1.200 de profesioniști în domeniul ingineriei de sunet [25]. Această discrepanță generează o presiune semnificativă asupra infrastructurii tehnice existente, cu aproximativ 5.000-6.000 de noi compoziții înregistrate anual [5].

Contextul digital al consumului de muzică din România, o țară cu peste 91% penetrare a internetului [51] și clasificată printre primele 5 cele mai dezvoltate piețe digitale din regiune, creează un mediu propice pentru implementarea tehnologiilor avansate bazate pe inteligență artificială.

1.2 Scopul tezei de doctorat

Scopul principal este dezvoltarea, implementarea și evaluarea unor soluții bazate pe rețele neuronale pentru optimizarea proceselor critice din producția muzicală modernă. Din perspectivă științifică, teza își propune să avanseze cunoașterea în domeniul aplicării inteligenței artificiale în procesarea semnalelor audio, contribuind cu modele originale adaptate provocărilor specifice ale producției muzicale.

Din perspectivă practică, cercetarea urmărește să ofere comunității profesionale instrumente tehnologice avansate și accesibile, într-un context global în care peste 3 milioane de artiști concurează pentru atenția ascultătorilor [41].

1.3 Conținutul tezei de doctorat

Teza este structurată în cinci capitole principale:

- **Capitolul 2 - Fundamentele Percepției și Analizei Sonore în Producția Audio** stabilește baza teoretică privind percepția sonoră, psiho-acustica și mediile de monitorizare.
- **Capitolul 3 - Producția Muzicală** detaliază procesele fundamentale ale lanțului de producție audio, de la captură la masterizare.
- **Capitolul 4 - Rețele Neuronale în Producția Audio** prezintă contribuțiile originale, incluzând un sistem de control personalizat al nivelului pentru înregistrări multi-canal, un compresor multi-bandă cu învățare automată și o tehnologie de expansiune spațială de la stereo la multi-canal.
- **Capitolul 5 - Concluzii** sintetizează rezultatele obținute și perspectivele de dezvoltare ulterioară.

În ansamblu, lucrarea propune o abordare integrată a producției muzicale contemporane, în care tehnologiile bazate pe rețele neuronale sunt utilizate strategic pentru a depăși limitările tradiționale și pentru a democratiza accesul la calitate sonoră profesională, cu aplicabilitate particulară în contextul industriei muzicale românești.

Capitolul 2

Fundamentele percepției și analizei sonore în producția audio

2.1 Ascultarea critică

Ascultarea critică reprezintă fundamentul întregului proces de producție audio, fiind o abilitate esențială care distinge inginerii de sunet profesioniști de ascultătorii obișnuiți. Transcendând simpla plăcere a audiției muzicale, aceasta implică o analiză metodică a componentelor sonore pentru a obține rezultatul artistic și tehnic dorit.

În esența sa, ascultarea critică se bazează pe dezvoltarea unei conștientizări extraordinare a detaliilor sonore. Izhaki *et al.* [28] numește această capacitate "*the golden ear*" - abilitatea de a discrimina între cele mai subtile detalii dintr-o producție, ce ar trece neobservate pentru ascultătorii neantrenați.

Ascultarea critică operează simultan pe două planuri complementare:

- **Planul obiectiv** - concentrat pe aspectele tehnice și măsurabile ale sunetului (frecvență, amplitudine, fază, dinamică) utilizând analizatoare de spectru, măsurători de nivel și indicatori de fază;
- **Planul subiectiv** - evaluează impactul emoțional și muzicalitatea materialului sonor.

Studiile arată că profesioniștii experimentați pot identifica frecvențele cu o marjă de eroare mai mică de un sfert de octavă [33], iar în domeniul procesării dinamice, pot percepe modificări ale raportului de compresie de până la 0.5:1 și variații ale timpilor de atac și eliberare de ordinul milisecundelor [18].

2.2 Urechea și psihoacustica în contextul producției audio

2.2.1 Anatomia și funcționarea sistemului auditiv

Sistemul auditiv uman este structurat în trei secțiuni principale:

Urechea Externă Pavilionul auricular direcționează undele sonore realizând o amplificare naturală de 5-10 dB în intervalul 2-7 kHz [2]. Una dintre cele mai remarcabile funcții ale urechii externe este capacitatea sa de a modifica spectrul sunetului incident în funcție de direcția sa de propagare. Acest fenomen, cunoscut în literatura de specialitate sub numele de "*Head Related Transfer Function*" - Funcție de Transfer Relațională cu Capul (HRTF), este fundamental pentru capacitatea noastră de a localiza surse sonore în spațiul tridimensional [35].

Urechea Medie Funcționează ca un transformator acustic-mecanic, amplificând vibrațiile sonore prin principiul pârghiei și diferența dintre ariile membranei timpanice și ferestrei ovale, producând o amplificare de aproximativ 26 dB. Urechea medie prezintă și un mecanism de protecție sofisticat - reflexul stapedian. Acest reflex involuntar este declanșat de sunete puternice (peste aproximativ 85 dB SPL) și implică contracția a doi mușchi minuscule: mușchiul stapedian și mușchiul tensor al timpanului.

Urechea Internă Cohleea convertește vibrațiile mecanice în semnale nervoase. Membrana bazilară variază în rigiditate și masă, permițând analiza spectrală conform relației lui Greenwood [20]. Celulele ciliate externe acționează ca amplificatoare moleculare, îmbunătățind sensibilitatea și selectivitatea sonică.

Sistemul nervos auditiv, începând de la ieșirea din cohlee și până la cortexul auditiv primar, reprezintă o rețea complexă de procesare a informației sonore organizată ierarhic [46]. Această cale neurală nu este un simplu releu de transmitere a informației, ci un sistem sofisticat de analiză și extragere a caracteristicilor relevante ale stimulului sonor.

2.2.2 Principii fundamentale de psihoacustică

Psihoacustica studiază relația dintre proprietățile fizice ale sunetului și percepția acestora de către sistemul auditiv uman, esențială pentru producția audio de înaltă calitate.

Percepția frecvenței La frecvențe joase (sub 500 Hz), sistemul auditiv prezintă o diferențiere spectrală remarcabil de fină, capacitate care se degradează progresiv pe măsură ce frecvența crește, urmând o scală aproximativ logaritmică [23]. Această caracteristică explică percepția constantă a intervalelor muzicale atunci când raportul frecvențelor rămâne neschimbat.

Percepția intensității Legea Weber-Fechner descrie cum sistemul auditiv comprimă gama largă de intensități fizice într-un interval perceptual mai restrâns, permițând funcționarea eficientă atât în medii silențioase, cât și în medii zgomotoase [23].

Mascarea spectrală și temporală Prezența unui sunet afectează capacitatea de a percepe alte sunete, cu o asimetrie caracteristică: sunetele de frecvență joasă pot masca efectiv cele de frecvență înaltă ("*upward spread of masking*"). Mascarea temporală se extinde până la 20 ms înainte (pre-mascare) și 200 ms după (post-mascare) sunetul mascator.

2.2.3 Funcția de transfer a capului (HRTF)

HRTF reprezintă o funcție complexă de frecvență și direcție care caracterizează modificările acustice introduse de anatomia ascultătorului asupra sunetelor:

$$\text{HRTF}(f, \theta, \varphi) = \frac{P(f, \theta, \varphi)}{P_0(f)} \quad (2.1)$$

unde P reprezintă presiunea sonoră la timpan pentru un sunet cu frecvența f ce provine din direcția definită de azimutul θ și elevația φ , iar P_0 este presiunea sonoră în câmp liber.

Cercetările recente [43] demonstrează că măsurarea HRTF-urilor este posibilă și în camere reverberante, utilizând tehnici avansate de procesare a semnalelor. Răspunsul acustic total măsurat poate fi exprimat ca:

$$h_{total,L/R}(t) = h_{amp,L/R} * h_{spk,L/R} * h_{room,L/R} * h_{L/R} * h_{mic,L/R} \quad (2.2)$$

unde componentele sunt:

- $h_{amp,L/R}(t)$ - răspunsul la impuls al amplificatorului audio
- $h_{spk,L/R}(t)$ - răspunsul la impuls al difuzorului
- $h_{room,L/R}(t)$ - răspunsul la impuls al camerei
- $h_{L/R}(t)$ - HRIR dorit pentru urechea stângă/dreaptă
- $h_{mic,L/R}(t)$ - răspunsul la impuls al microfoanelor

În producția audio, HRTF-urile permit crearea unei imagini sonore tridimensionale folosind doar două canale de redare, fiind esențiale pentru mixajul spațial și producția de conținut pentru realitatea virtuală și augmentată.

2.3 Camera de control

Camera de control, situată la intersecția dintre artă și tehnologie, reprezintă spațiul esențial unde se evaluează și procesează materialul sonor. Standardele internaționale ITU-R BS.1116 și EBU Tech 3276 stabilesc parametri riguroși pentru aceste spații.

Geometria camerei Raporturile dimensionale sunt guvernate de relația:

$$1.1 \frac{w}{h} \leq \frac{l}{h} \leq 4.5 \frac{w}{h} - 4 \quad (2.3)$$

unde l , w și h reprezintă lungimea, lățimea și înălțimea camerei, asigurând o distribuție optimă a modurilor proprii.

Timpul de reverberație Calibrat în funcție de volumul camerei conform ecuației:

$$T_m = 0.25 \left(\frac{V}{V_0} \right)^{1/3} \quad (2.4)$$

unde V reprezintă volumul real al camerei, iar V_0 este volumul de referință de 100 m³.

Distribuția modurilor proprii Poate fi analizată utilizând:

$$F = \frac{c}{2} \sqrt{\frac{x^2}{l^2} + \frac{y^2}{w^2} + \frac{z^2}{h^2}} \quad (2.5)$$

unde c este viteza sunetului, iar x , y , z sunt numere întregi ce definesc tipul modului.

Tratamentul acustic Conceptul LEDE (Live-End Dead-End) implică absorbție predominantă în partea frontală și difuzie controlată în zona posterioară. Pentru controlul frecvențelor joase se utilizează capcane de bas și rezonatoare Helmholtz calibrate conform:

$$f = \frac{100R}{\sqrt{V(l + 1.6R)}} \quad (2.6)$$

unde f este frecvența de rezonanță dorită, R este raza deschiderii, V este volumul cavității, iar l reprezintă lungimea gâtului rezonatorului.

Metodologia de cercetare adoptată îmbină fundamentarea teoretică cu validarea practică, utilizând ca studiu de caz procesul de dezvoltare a laboratorului specializat din cadrul Facultății de Electronică, Telecomunicații și Tehnologie Informației, Universitatea Națională de Știință și Tehnologie POLITEHNICA București.

Experiența demonstrează că, deși provocator, crearea unui spațiu conform cu standardele internaționale este realizabilă prin aplicarea sistematică a principiilor acustice, combinată cu măsurători precise și ajustări progresive [43].

Capitolul 3

Producția muzicală

Producția muzicală reprezintă procesul complex prin care ideile muzicale sunt transformate în înregistrări finite, gata pentru distribuție și consum [24]. Este punctul de convergență între artă și tehnologie, unde viziunea creativă întâlnește expertiza tehnică. În era digitală, democratizarea tehnologiei a făcut posibilă crearea de muzică de înaltă calitate în studiouri personale, crescând totodată complexitatea procesului [53].

Procesul constă în două etape majore interconectate: mixajul și masterizarea. Mixajul reprezintă îmbinarea elementelor înregistrate într-un tot unitar și coerent [28], manipulând nivelurile relative, poziția în câmpul stereo, benzile de frecvențe și dinamica. Masterizarea, etapa finală, reprezintă procesul de finisare și optimizare pentru distribuție [30], asigurând că materialul sună consistent pe orice sistem de redare.

3.1 Procesul de mixaj

Mixajul reprezintă punctul crucial unde arta întâlnește tehnica, transformând surse audio multiple într-un produs final coerent și expresiv [28]. Procesul se construiește pe fundamentul etapei de captare, întrucât calitatea materialului înregistrat definește limitele și posibilitățile mixajului.

Matematic, pentru un sistem stereo cu n surse, mixajul poate fi modelat ca o funcție de transfer:

$$M_{L,R}(t) = \sum_{i=1}^n [S_i(t) \cdot A_i(t) \cdot P_i(t) \cdot F_i(\omega, t) \cdot D_i(t)] \quad (3.1)$$

unde $M_{L,R}(t)$ reprezintă semnalele mixate pentru canalele stânga și dreapta, $S_i(t)$ este semnalul audio sursă, $A_i(t)$ este amplitudinea, $P_i(t)$ este funcția de panoramare, $F_i(\omega, t)$ reprezintă procesarea spectrală, și $D_i(t)$ procesarea dinamică aplicată.

3.1.1 Stabilirea proporțiilor sonore fundamentale

Stabilirea proporțiilor sonore fundamentale reprezintă unul dintre primii și cei mai critici pași în construirea unui mixaj de succes [48]. Aceasta începe cu elementele care definesc caracterul și energia piesei, în muzica modernă fiind adesea secțiunea ritmică (tobe și bas) și vocea principală.

Inginerul trebuie să identifice centrul de greutate al piesei - frecvența sau zona de frecvențe pe care se sprijină întreaga construcție sonoră [19]. Pentru muzica cu tempo rapid, acest centru tinde să fie poziționat mai sus în spectrul de frecvențe, în timp ce muzica mai lentă poate beneficia de un centru de greutate mai jos.

Echilibrul tonal optim nu poate fi definit în termeni absoluți, ci doar în raport cu contextul muzical specific și intenția artistică [42], fiind influențat de genul muzical, epoca de producție și așteptările audienței.

3.1.2 Panoramarea

Panoramarea configurează distribuția energiei sonore între canalele stânga și dreapta, având un rol important în crearea iluziei de spațiu și clarificarea arhitecturii sonore. Acest proces este guvernat de mecanisme psihoacustice: diferențele interaurale de intensitate (IID) și diferențele interaurale de timp (ITD) [6].

În practica actuală, panoramarea îndeplinește două funcții esențiale:

- Separarea spectrală și reducerea mascării dintre instrumente, în special între cele care ocupă regiuni spectrale similare
- Crearea unei ierarhii perceptuale, sunetele plasate central beneficiind de o atenție crescută, motiv pentru care elementele-cheie (vocea principală, toba mare) sunt poziționate central

3.1.3 Procesarea spectrală

Procesarea spectrală constituie fundamentul modelării sonore în producția audio contemporană. Egalizarea transcende simpla corecție tehnică, devenind un instrument artistic esențial pentru definirea caracterului unui mixaj.

Matematic, funcția de transfer a unui egalizator parametric poate fi reprezentată prin:

$$H(f) = A_0 \cdot \prod_{i=1}^n \frac{1 + \frac{Q_i}{g_i} \cdot \left(\frac{f}{f_i}\right) + \left(\frac{f}{f_i}\right)^2}{1 + \frac{Q_i}{g_i \cdot A_i} \cdot \left(\frac{f}{f_i}\right) + \left(\frac{f}{f_i}\right)^2} \quad (3.2)$$

unde A_0 reprezintă amplificarea generală, f_i este frecvența centrală a benzii i , Q_i este factorul de calitate care determină lățimea benzii, iar g_i reprezintă amplificarea sau atenuarea aplicată benzii respective.

Dincolo de formularea matematică, egalizarea implică înțelegerea comportamentului spectral al instrumentelor muzicale, fiecare ocupând un spațiu spectral specific ce poate fi modelat strategic [28]. În acest sens, egalizarea devine un limbaj expresiv prin care inginerul de sunet poate:

- Defini ierarhia și rolurile elementelor într-un mixaj, evidențiind anumite instrumente în detrimentul altora
- Crea senzația de spațialitate și adâncime prin manipularea conținutului spectral
- Modela caracterul timbral al fiecărui element pentru a se integra în viziunea estetică generală
- Sublinia tensiunea și relaxarea muzicală prin accentuarea sau atenuarea strategică a anumitor regiuni spectrale

3.1.4 Procesarea dinamică

Procesarea dinamică influențează direct percepția intensității, impactului și coeziunii materialului sonor. Funcția de transfer a unui compresor poate fi reprezentată prin:

$$y(t) = \begin{cases} x(t), & \text{pentru } |x(t)| \leq T \\ T + \frac{|x(t)| - T}{R}, & \text{pentru } |x(t)| > T \end{cases} \quad (3.3)$$

unde $x(t)$ reprezintă semnalul de intrare, $y(t)$ semnalul de ieșire, T pragul ("threshold") și R raportul de compresie.

Efectul procesării dinamice nu este o simplă modificare a nivelului semnalului, acesta influențează direct percepția tridimensională și ierarhizarea elementelor într-un mixaj. Impactul procesării dinamice este strâns legat de mecanismele psihoacustice fundamentale [56], incluzând:

- Percepția intensității sunetului, care nu este liniară și depinde de conținutul spectral
- Mascare temporală, prin care sunetele puternice pot masca sunetele mai slabe care le precedă sau urmează
- Integrarea temporală a energiei acustice în sistemul auditiv
- Adaptarea la niveluri constante și sensibilitatea la modificări

3.2 Procesul de masterizare

Masterizarea reprezintă ultimul pas creativ din lanțul producției muzicale, transformând mixajul în produsul finit destinat consumatorului [30]. În era contemporană, optimizarea

pentru platformele de streaming a revoluționat abordarea prin introducerea normalizării intensității sonore conform standardelor EBU R128 și ITU BS.1770 [47].

3.2.1 Echilibrul tonal

Echilibrul tonal conferă producției muzicale claritate, profunzime și coerență spectrală [30], referindu-se la distribuția optimă a energiei sonore pe întreg spectrul de frecvențe.

Din punct de vedere tehnic, procesul începe cu o analiză atentă a mixajului utilizând instrumente precum analizoarele de spectru a căror rezultate sunt comparate cu materiale de referință din același gen muzical. Aceasta permite identificarea eventualelor anomalii spectrale, cum ar fi acumulările excesive de energie în anumite benzi de frecvență sau carențele din altele [10]. De exemplu, un mixaj cu prea multă energie în gama 200-300 Hz va suna înfundat sau "turbure", în timp ce lipsa energiei în zona 3-5 kHz poate face ca detaliile vocale să fie insuficient prezente.

Un aspect important de menționat este că echilibrul tonal în masterizare presupune interacțiunea complexă cu celelalte procese, în special cu compresia dinamică [14]. Compresoarele multi-bandă, de exemplu, influențează semnificativ echilibrul tonal prin modificarea raportului dintre diversele game de frecvență în funcție de conținutul dinamic al materialului. Această interconectare necesită o abordare holistică, în care ajustările tonale și cele dinamice se completează reciproc.

3.2.2 Saturație

Saturația contribuie la modelarea caracterului sonic și îmbogățirea paletelor armonice [30], fiind un proces de distorsiune controlată care emulează comportamentul circuitelor analogice. Tipurile principale includ:

- **Saturația de bandă magnetică** - generează armonici preponderent pare și comprimă ușor tranziții
- **Saturația de tub cu vid** - adaugă armonici predominant pare într-un mod progresiv
- **Saturația de transformator** - introduce modificări subtile ale formei de undă
- **Saturația digitală waveshaping** - utilizează algoritmi matematici pentru distorsiune controlată

3.2.3 Optimizarea tăriei sonore

Optimizarea tăriei sonore urmărește maximizarea impactului perceptual fără a compromite calitatea artistică și dinamica muzicală [30]. Standardul ITU-R BS.1770 [26] a oferit fundamentul matematic pentru măsurarea tăriei sonore percepute, fiind adoptat și extins de EBU R128 [16].

Compresia multi-stadiu Oferă control nuanțat asupra dinamicii fără sacrificarea caracterului natural [14], bazându-se pe principiul că multiple etape moderate generează un rezultat mai muzical decât o compresie agresivă.

Limitarea Limitatoarele au dublu rol: prevenirea distorsiunii digitale și maximizarea nivelului perceptiv [30]. În era normalizării tăriei percepute, strategia de maximizare a evoluat spre menținerea unui nivel de limitare moderat, care păstrează dinamica și impactul [54].

3.3 Concluzii

Producția muzicală reprezintă o simbioză complexă între artă și tehnologie. Mixajul și masterizarea se bazează pe principii tehnice riguroase, dar transcend dimensiunea pur matematică, încorporând elemente artistice indispensabile.

În era contemporană, dominată de platformele de streaming cu algoritmi de normalizare a intensității sonore, abordarea tradițională a masterizării a suferit transformări fundamentale. Accentul s-a deplasat de la simpla maximizare a nivelului sonor către menținerea dinamicii naturale și a caracteristicilor timbrale.

Valoarea artistică a unei producții muzicale nu derivă din simpla sa intensitate percepută, ci din capacitatea sa de a transmite emoție și expresivitate printr-o calitate sonoră optimă.

Capitolul 4

Rețele neuronale în producția audio

Revoluția inteligenței artificiale a transformat fundamental producția audio în ultimii ani, extinzându-se la aproape toate aspectele procesului: de la sinteză și procesare până la mixaj, masterizare și restaurare [45]. În inima acestei transformări se află rețelele neuronale, care oferă capacități fără precedent de a modela, genera și manipula date audio complexe [15].

Cu toate acestea, majoritatea abordărilor propuse urmăresc să dezvolte soluții universale, aplicabile unei game largi de probleme audio [9]. Această orientare generalistă, deși atractivă din perspectiva aplicabilității și comercializării, conduce adesea la rezultate sub-optime când este confruntată cu provocările specifice ale diferitelor subdomenii ale producției audio [22].

În acest context, lucrarea își propune o reevaluare a paradigmei curente, pledând pentru dezvoltarea unor arhitecturi neuronale specializate, optimizate pentru provocări specifice din producția audio. Această schimbare de direcție nu reprezintă un pas înapoi în universalitatea soluțiilor bazate pe inteligență artificială, ci mai degrabă o maturizare necesară a domeniului, similar cu evoluția altor tehnologii care au progresat de la soluții generaliste la instrumente specializate de înaltă performanță [55].

4.1 Clasificarea instrumentelor muzicale

Recunoașterea și clasificarea instrumentelor reprezintă o componentă fundamentală în procesarea audio modernă, cu aplicații multiple în mixajul automat, organizarea bibliotecilor de sunet și analiza înregistrărilor muzicale [21]. În această secțiune, prezentăm un sistem de clasificare ierarhic care utilizează atât caracteristici spectrale, cât și informații textuale pentru identificarea precisă a instrumentelor muzicale în înregistrări multi-canal [36].

Implementarea unui sistem robust de recunoaștere a instrumentelor permite automatizarea etapelor laborioase, facilitând organizarea pistelor în grupuri funcționale, aplicarea șabloanelor adecvate și stabilirea nivelurilor inițiale optimizate pentru fiecare tip de in-

strument. Această automatizare nu doar accelerează fluxul de lucru, ci și standardizează practica de mixaj, reducând inconsistențele între proiecte și minimizând erorile umane.

O limitare semnificativă a sistemelor actuale constă în faptul că majoritatea modelelor sunt antrenate pe înregistrări izolate de înaltă calitate, care nu reflectă caracteristicile semnalelor întâlnite în sesiunile de înregistrare reale. Acestea conțin frecvent interferențe între canale, poziționări variate ale microfoanelor și acustică de cameră care modifică semnificativ amprente spectrale, complicând procesul de clasificare [30].

Arhitectura propusă este structurată pe două niveluri de clasificare: inițial, un clasificator primar separă materialul audio în patru categorii fundamentale (tobe, bas, voce și altele), urmând ca ulterior clasificatoare specializate secundare să ofere identificări mai detaliate pentru categoriile cu diversitate ridicată, precum tobele și categoria „altele”.

Pentru îmbunătățirea performanței de clasificare, în special în cazul înregistrărilor cu interferențe sau caracteristici spectrale ambigue, s-a dezvoltat o abordare multi-modală care îmbină analiza spectrală cu analiza metadatelor textuale (denumirile pistelor audio). Această metodă se bazează pe observația că, în practica inginerilor de sunet, denumirile pistelor conțin frecvent indicii valoroase despre tipul instrumentului înregistrat.

Procesul de inferență al sistemului selectează cele mai reprezentative trei ferestre cu energia RMS maximă din fiecare pistă audio și procesează paralel atât conținutul audio, cât și denumirea pistei. Clasificatorul primar categorizează materialul audio în cele patru clase principale, iar pentru categoriile Tobe sau Altele, un clasificator secundar oferă identificare detaliată. Modelul multi-modal combină spectrogramele audio cu textul denumirii pistei. Clasa finală este determinată prin compararea scorurilor de încredere din predicțiile valide, cu preferință pentru rezultatul cu încrederea cea mai ridicată.

Performanța sistemului de clasificare, detaliată în Tabelul 4.1, a instrumentelor a fost evaluată utilizând multiple seturi de date, incluzând MUSDB18HQ [11], colecția Cambridge Multitrack [12] și seturi de date personalizate pentru categorii specializate de instrumente. Modelul de clasificare primară a atins o acuratețe de 83,89% pe setul de validare MUSDB18HQ, demonstrând capacitatea sa de a distinge eficient între cele patru categorii principale de instrumente. Clasificatorul specializat pentru tobe a înregistrat o performanță excelentă cu o acuratețe de 95,67%, în timp ce clasificatorul pentru categoria „altele” a atins 89,10% acuratețe.

Tabel 4.1 Acuratețea de validare pentru fiecare model de clasificare.

Model	Set de date	Tip intrare	Clase	Acuratețe [%]
ResNet-18	MUSDB18HQ [11]	STFT	4	83,89
ResNet-18	Tobe personalizat	STFT	5	95,67
ResNet-18	Altele personalizat	STFT	3	89,10
Multi-modal	Cambridge [12]	Text + STFT	27	82,19

Aplicat în proiecte reale multi-canal, sistemul a demonstrat o capacitate de adaptare variabilă în funcție de complexitatea scenariului de producție muzicală. Pentru proiecte de dimensiuni reduse (10-22 piste), rezultatele au fost remarcabile, cu o acuratețe de clasificare între 90% și 100%. Însă în cazul proiectelor de amploare medie și mare (34-60 piste), sistemul a întâmpinat dificultăți semnificative, acuratețea scăzând la valori între 47,72% și 70,58%.

În ciuda limitărilor identificate, sistemul de recunoaștere a instrumentelor demonstrează potențialul aplicațiilor de învățare automată în producția audio profesională. Prin combinarea analizei spectrale cu procesarea informațiilor textuale într-un cadru ierarhic de decizie, obținem un sistem robust de clasificare care poate funcționa eficient în diverse contexte de producție muzicală, reducând semnificativ timpul necesar pentru organizarea și etichetarea manuală a sesiunilor audio complexe.

4.2 Controlul personalizat al nivelului pentru înregistrări multi-canal

Domeniul mixajului se află într-o continuă transformare datorită introducerii soluțiilor automate, cu scopul de a imita sau a asista procesul complex și artistic care, în mod tradițional, depinde de expertiza unui profesionist. Această expertiză este esențială pentru ajustarea dinamicii, a caracteristicilor spațiale, timbrului și tonalității înregistrărilor multi-canal. Modificarea acestor componente implică o serie de etape procedurale care utilizează procese liniare și neliniare pentru a obține rezultatul auditiv dorit [28].

Pentru acest proiect, a fost construit un set de date compus din 20 de piese muzicale, combinând piese din baza de date *Cambridge Music Technology* [48] și înregistrări ale unor trupe românești locale, oferite de inginerii participanți la experiment. Piese selectate acoperă o varietate largă de genuri și stiluri, oferind o bază solidă pentru analiza particularităților procesului de mixaj audio.

S-a optat pentru utilizarea algoritmilor genetici [29, 32] pentru a gestiona eficient această problemă complexă cu soluții multiple. Metoda utilizează un algoritm genetic pentru a rafina iterativ coeficienții de mixaj, urmărind obținerea unui mixaj optim. Algoritmul genetic personalizat asigură că doar cele mai bune soluții supraviețuiesc și își transmit mai departe caracteristicile.

Pentru a evalua fezabilitatea metodei propuse, în completarea algoritmului genetic, am implementat și o rețea neuronală care rezolvă problema controlului automat al intensității perceptuale într-un mod adaptiv. Am ales o arhitectură bazată pe rețele neuronale convoluționale (CNN) pentru analiza semnalele audio deoarece acestea excelează în identificarea tiparelor din date secvențiale, precum sunetul.

Ambele metode propuse (algoritmul genetic și rețeaua neuronală) au fost aplicate pentru fiecare dintre cei 10 ingineri de sunet, cu scopul de a surprinde preferințele

fiecăruia dintre aceștia. Astfel, au rezultat 10 soluții ale algoritmului genetic personalizat și 10 modele de rețea neuronală antrenate.

În urma analizei rezultatelor testului subiectiv, detaliat în Tabelul 4.2, pe cele 15 piese utilizate pentru antrenare, se observă că algoritmul genetic a obținut scoruri egale sau superioare mixajului realizat de ingineri în jumătate dintre cazuri, mai exact pentru 5 din cei 10 ingineri. Rețeaua neuronală CNN a avut de asemenea o performanță bună, egalând sau depășind MIX-ul în 4 din cele 10 cazuri. În situațiile în care GA și CNN nu au depășit scorurile MIX-ului, diferențele au fost minore.

Tabel 4.2 Scorurile medii pentru setul de date de antrenare format din 15 melodii. Cele mai bune rezultate sunt evidențiate cu albastru.

Versiune	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Medie
RAW	3,5	3,9	3,5	3,1	3,5	2,9	3,7	1,2	3,5	1,5	3,0
MIX	4,3	4,1	4,0	3,7	3,3	3,3	4,3	2,2	4,0	4,0	3,7
GA	4,6	3,9	4,2	3,6	3,1	3,4	3,9	2,9	4,2	4,5	3,9
CNN)	4,5	4,2	4,5	3,7	3,1	3,7	4,1	3,3	4,5	3,7	3,9

Rezultatele obținute în urma implementării și evaluării algoritmului pentru controlul personalizat al nivelului în înregistrările multi-canal, rezultate prezentate și în lucrarea publicată anterior în cadrul revistei *Algorithms* [37], demonstrează potențialul considerabil al metodelor automate în domeniul mixajului audio.

4.3 Compresorul multi-bandă cu învățare automată a parametrilor

Compresia multi-bandă reprezintă una dintre cele mai versatile și complexe tehnici de procesare audio utilizate în producția muzicală modernă. Spre deosebire de compresoarele convenționale care procesează întregul spectru de frecvențe în mod uniform, compresoarele multi-bandă împart semnalul audio în benzi de frecvență separate, permițând astfel o procesare dinamică individualizată pentru fiecare regiune spectrală [17].

Configurarea optimă a parametrilor unui compresor multi-bandă reprezintă însă o provocare considerabilă chiar și pentru inginerii experimentați. Interacțiunile complexe între parametrii multiplelor benzi, alături de comportamentul dinamic al materialului sonor, creează un spațiu decizional vast și neliniar. Acest context a stimulat cercetări intensive în direcția dezvoltării sistemelor automate de asistență sau configurare a procesării dinamice.

Implementarea software a compresorului multi-bandă cu învățare automată a fost realizată utilizând JUCE (*Jules' Utility Class Extensions*), o alegere strategică fundamentată pe multiple avantaje tehnice și practice. JUCE reprezintă în prezent standardul industrial

pentru dezvoltarea aplicațiilor audio profesionale, oferind un ecosistem complet și matur pentru implementarea procesoarelor de efecte audio [52].

Implementarea componentei de procesare multi-bandă a reprezentat una dintre provocările centrale ale dezvoltării compresorului, necesitând soluții tehnice complexe pentru a asigura o separare spectrală precisă fără introducerea de artefacte sesizabile. Împărțirea spectrului audio în patru benzi distincte (frecvențe joase, medii-joase, medii și înalte) a fost realizată utilizând filtre Linkwitz-Riley de ordinul al patrulea (LR4) [31], o alegere determinată de caracteristicile lor superioare în aplicații de trecere între benzi audio.

Arhitectura sistemului de învățare automată pentru parametrii compresorului multi-bandă se bazează pe o soluție hibridă ce îmbină rețele neuronale convoluționale (CNN) de tip siamez cu regresie prin arbori decizionali (*Random Forest*). Această structură a fost concepută pentru a extrage caracteristici relevante din semnalele audio și a modela relațiile complexe dintre parametrii compresorului și modificările spectrale introduse de acesta.

Antrenarea modelului pentru predicția parametrilor compresorului multi-bandă a implicat o abordare complexă bazată pe rețele siameze și învățare prin transfer de stil. Procesul a utilizat înregistrări audio ale 3 ingineri de sunet, dispunând atât de versiunile neprocesate cât și pe cele procesate pentru cinci cântece.

Rezultatele obținute, detaliate în Tabelul 4.3, ne prezintă erorile de predicție și dezvăluie aspecte semnificative privind comportamentul modelului. Cu o eroare de predicție medie totală de 10,2% pentru ansamblul parametrilor, sistemul demonstrează o capacitate promițătoare de a aproxima setările optime ale compresorului multi-bandă. Această precizie globală, coroborată cu performanța excelentă pentru parametri critici precum timpul de atac, validează abordarea bazată pe rețele neuronale siameze pentru automatizarea parțială a procesului de compresie dinamică în producția muzicală.

Tabel 4.3 Eroare medie de a predicției per parametru și eroare totală

Parametru	Eroare Medie
Atac	3,17%
Revenire	7,44%
Câștig	3,76%
Rație	20,36%
Prag	16,34%
Frecvență	10,26%
Media erorilor	10,2%

Evaluarea perceptuală a rezultatelor, realizată prin teste de ascultare critică cu trei ingineri de sunet (doi profesioniști și unul amator), oferă perspective valoroase asupra calității sonore a procesării realizate de modelul implementat. Materialul audio procesat prin rețeaua neuronală (NN) a obținut un scor mediu de 3,7 pe o scală de la 1 la

5, poziționându-se semnificativ peste materialul neprocesat (RAW) și apropiindu-se considerabil de procesarea profesională realizată manual (MIX), care a obținut 4,1.

4.4 Expandare spațială de la stereo la multi-canal

Dezvoltarea sistemelor de sunet multi-canal a evoluat semnificativ de-a lungul timpului, începând cu sistemele mono simple și progresând către configurațiile stereo. Ulterior, această evoluție a continuat prin adăugarea mai multor canale și implementarea unor tehnologii avansate de procesare audio, precum formatele 5.1 (standardizat prin ITU 775) [49] sau 7.1, care sunt frecvent utilizate în cinematografe și sisteme personale *home theater*.

În ciuda apariției acestor tehnologii moderne, formatul stereo, dezvoltat de Blumlein în anii 1930 [1], continuă să fie cel mai răspândit sistem audio, constând în două difuzoare (stâng și drept) concepute pentru a simula modul în care auzul uman percepe sunetele. Procesul de expandare multi-canal reprezintă transformarea unui conținut audio de la n la m canale, unde $n < m$.

Diverse metode de expandare multi-canal au fost dezvoltate anterior, bazate pe separarea conținutului în componente de tip primar și ambiental [4, 44]. Implementarea software a modelului hibrid de separare a surselor și extracție ambientală combină unele dintre cele mai eficiente strategii de extragere a mai multor canale dintr-un mixaj stereo. S-a utilizat un algoritm de separare a surselor care poate extrage 4 surse muzicale separate: tobe, bas, voce și alte sunete.

Acest modul se bazează pe o rețea neuronală (DNN) dezvoltată de Défossez *et al.* [13], un model ce primește ca intrare o mostră audio stereo de 44 secunde la 44,1 kHz și extrage 4 surse audio (tobe, bas, altele, voci), folosind o abordare hibridă bazată pe separarea surselor în domeniul spectral și a formei de undă. Acest algoritm a fost clasat pe primul loc la *Music Demixing Challenge* (MDX) în 2021, organizată de Sony [34].

În plus față de cele patru surse obținute prin modulul de separare a surselor, a fost utilizat și un algoritm de extragere primar-ambiental, însă au fost folosite doar sunetele ambientale din semnalul stereo de intrare, deoarece acestea au o contribuție semnificativă în crearea efectului imersiv. Pentru aceasta, s-a utilizat algoritmul *Adaptive Panning* (ADP) [8, 27], bazat pe filtrul adaptiv LMS, care estimează sunetele primare și ambientale din canalele stereo.

Semnalul audio rezultat în urma procesului de expandare multi-canal trebuie să combine toate sursele într-un mod armonios, pentru a crea o experiență de sunet multi-canal captivantă. Sursele audio separate trebuie distribuite atent în sistemul 5.1 pentru a asigura o reprezentare fidelă a semnalului original și păstrarea relațiilor complexe dintre sursele sonore.

Rezultatele experimentale, publicate în cadrul conferinței *SpeD 2023* [40], demonstrează în mod convingător superioritatea metodei propuse. Comparativ cu abordările

clasice, metoda propusă reflectă avantajele utilizării combinației dintre tehnicile de separare a surselor și extracție ambientală pentru generarea unui sunet imersiv natural.

Algoritmul de expandare spațială SSPAE a demonstrat superioritatea sa față de metodele tradiționale, oferind îmbunătățiri semnificative în redarea imersivă a materialelor stereo prin formate multi-canal. Integrarea canalului LFE, combinată cu tehnicile avansate de separare a surselor, a condus la rezultate superioare în testele obiective bazate pe metrica SDR.

Capitolul 5

Concluzii

În această teză au fost prezentate contribuții în domeniul producției muzicale folosind rețele neuronale, cu accent pe dezvoltarea unor soluții automatizate și adaptive pentru îmbunătățirea proceselor de mixaj audio și a experienței ascultătorului. Investigațiile și soluțiile propuse acoperă aspecte esențiale ale producției audio moderne, de la percepția și analiza sonoră, la recunoașterea instrumentelor, controlul nivelurilor audio, procesarea dinamică și expandarea spațială.

Cercetarea a fost orientată spre crearea unor soluții specializate pentru provocări specifice din producția audio, cu accent pe tehnici de învățare automată capabile să emuleze deciziile experților umani, de asemenea să îmbunătățească fluxurile de lucru existente. Acest demers științific a reunit cunoștințe din domenii variate precum psihoacustica, ingineria audio, procesarea semnalelor și inteligența artificială.

O atenție deosebită a fost acordată contextului ascultării critice, înțelegerii mecanismelor de percepție auditivă și mediului fizic de audiere, considerate fundații esențiale pentru dezvoltarea unor algoritmi eficienți. Pe parcursul cercetării, metodele propuse au fost evaluate atât prin metrici obiective, cât și prin teste subiective, asigurând astfel validarea soluțiilor în contexte reale de utilizare.

5.1 Rezultate obținute

În capitolul doi a fost realizată o analiză detaliată a fundamentelor percepției și analizei sonore, esențiale pentru dezvoltarea algoritmilor din domeniul producției muzicale:

- Au fost investigate și documentate principiile ascultării critice, evidențiind atât aspectele obiective, cât și cele subiective implicate în evaluarea calității audio.
- A fost analizată anatomia și funcționarea sistemului auditiv uman, cu accent pe mecanismele de percepție spațială și procesare neurală a sunetului.
- Au fost prezentate principiile psihoacustice fundamentale, inclusiv percepția frecvenței, intensității și fenomenele de mascare spectrale și temporale.

- A fost elaborat un studiu detaliat privind funcția de transfer relațională cu capul (HRTF), cu implicațiile sale în reproducerea spațială a sunetului.
- A fost realizat un studiu de caz privind proiectarea și implementarea unei camere de control pentru ascultare critică, incluzând aspecte de izolare acustică, tratament acustic și conformitate cu standardele internaționale.

În capitolul trei a fost realizată o analiză cuprinzătoare a proceselor de mixaj și masterizare, integrând atât perspectiva tehnică, cât și cea artistică:

- Au fost sistematizate principiile fundamentale care guvernează procesul de mixaj, cu accent pe stabilirea proporțiilor sonore, panoramarea strategică și manipularea perspectivei spectrale.
- A fost dezvoltat și documentat un cadru conceptual pentru procesarea spectrală în mixaj, incluzând tehnici de egalizare aditivă și substractivă, precum și metode de construire a unei biblioteci intuitive de asocieri spectrale.
- Au fost analizate tehnicile avansate de procesare dinamică, inclusiv compresie serială, paralelă și *side-chain*, evaluând impactul acestora asupra percepției spațiale și a coeziunii sonore.
- A fost prezentat un model matematic pentru optimizarea tăriei sonore în contextul standardelor moderne de normalizare (EBU R128, ITU BS.1770), oferind o perspectivă integratoare asupra evoluției standardelor în ecosistemul platformelor de streaming.
- Au fost investigate tehnicile de saturație armonică în masterizare, analizând rolul distorsiunilor controlate (bandă magnetică, tuburi, transformatoare) în modelarea caracterului sonic și în optimizarea densității spectrale.

În al patrulea capitol au fost prezentate contribuții originale în aplicarea rețelelor neuronale pentru automatizarea și îmbunătățirea unor procese esențiale din producția audio:

- A fost conceput și implementat un sistem ierarhic de recunoaștere automată a instrumentelor muzicale pentru optimizarea organizării sesiunilor de lucru în stațiile audio digitale (DAW). Sistemul combină multiple modele de clasificare pentru identificarea precisă a 27 de tipuri distincte de instrumente, atingând o acuratețe de până la 100% pentru proiectele de dimensiuni reduse și o acuratețe medie de 68.33% pentru proiectele complexe înregistrate în multiple studiouri cu convenții variate de denumire a pistelor.

- A fost dezvoltat un sistem de control personalizat al nivelului pentru înregistrări multi-canal, folosind atât algoritmi genetici (GA), cât și rețele neuronale (CNN), obținând performanțe competitive cu mixajele realizate de inginerii profesioniști. Sistemul a demonstrat capacitatea de a reproduce preferințele sonore ale inginerilor de sunet, cu o eroare medie sub 5% pentru parametrul de atac.
- A fost implementat un compresor multi-bandă cu învățare automată a parametrilor, utilizând o arhitectură de rețea neuronală siameză, capabilă să determine optim setările de dinamică pentru diferite tipuri de instrumente muzicale. Testele percepțuale au arătat o diferență medie de doar 0.4 puncte între procesarea manuală și cea automată pe o scală de 5 puncte.
- A fost dezvoltat un algoritm hibrid pentru expandarea spațială a semnalelor audio stereo la format multi-canal 5.1, combinând tehnici de separare a surselor și extracție ambientală. Metoda propusă (SSPAE) a demonstrat performanțe superioare metodelor existente, cu îmbunătățiri de până la 6 dB în metrica SDR (Signal to Distortion Ratio).

5.2 Contribuții originale

Principalele contribuții originale ale acestei teze se încadrează în patru direcții majore:

1. Proiectarea și implementarea unei camere de control pentru ascultare critică [1, 2]

- A fost dezvoltată o metodologie de proiectare acustică bazată pe standarde internaționale și cercetare experimentală.
- Au fost implementate și validate tehnici moderne de măsurare și tratament acustic pentru obținerea unui mediu de audiție neutru și precis.
- A fost elaborată o metodă de măsurare a HRTF-urilor în camere reverberante, eliminând necesitatea utilizării camerelor anecoide costisitoare.

2. Clasificarea și recunoașterea automată a instrumentelor muzicale în înregistrări multi-canal [3]

- A fost dezvoltat un sistem ierarhic de clasificare a instrumentelor muzicale, bazat pe modele CNN specializate pentru diferite categorii de instrumente.
- A fost implementată o abordare multimodală care integrează atât analiza spectrală cât și procesarea metadatelor textuale din denumirile pistelor.
- A fost conceput un algoritm decizional bazat pe praguri de încredere adaptive pentru maximizarea acurateții clasificării în scenarii de producție reale.

- A fost evaluată performanța sistemului pe înregistrări complexe realizate în multiple studiouri, demonstrând robustetea soluției în fața inconsistențelor de denumire și a interferențelor acustice între piste.

3. Dezvoltarea unor algoritmi de control automat al nivelului audio în producția multi-canal [4, 5]

- A fost proiectat și implementat un algoritm genetic pentru optimizarea intensității perceptuale, capabil să aproximeze preferințele subiective ale inginerilor de sunet.
- A fost dezvoltată o rețea neuronală (CNN) adaptivă pentru controlul nivelurilor, oferind flexibilitate sporită în funcție de materialul audio procesat.
- Au fost validate experimental ambele metode, demonstrând performanțe comparabile sau superioare mixajelor realizate manual de profesioniști.

4. Sisteme adaptive pentru procesarea dinamică și expandarea spațială audio [6, 7]

- A fost implementat un compresor multi-bandă cu învățare automată a parametrilor, utilizând o arhitectură de rețea neuronală siameză pentru transferul stilistic între ingineri de sunet.
- A fost dezvoltat un sistem hibrid de expandare spațială stereo-la-5.1, integrând tehnici de separare a surselor bazate pe rețele neuronale cu algoritmi de extracție ambientală.
- Au fost propuse și validate metode de evaluare a calității acustice pentru sistemele de expandare spațială, demonstrând superioritatea soluției propuse față de metodele existente.

5.3 Lista lucrărilor originale

[1] **Moroșanu, Bogdan**, Victor Popa, Cristian Negrescu, and Ionuț Ficîu. "*Control room design for subjective audio critical listening.*" In *Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies XI*, vol. 12493, pp. 414-421. SPIE, 2023.

[2] Popa, Victor, **Bogdan Moroșanu**, and Cristian Negrescu. "*Head related transfer function measurement in reverberant rooms.*" In *Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies XI*, vol. 12493, pp. 617-622. SPIE, 2023.

[3] **Moroșanu, Bogdan**, Marian Negru, Georgian Nicolae, Horia Ioniță, and Constantin Paleologu. "A Machine Learning-Assisted Automation System for Optimizing Session Preparation Time in Digital Audio Workstations." Information 16, under review (2025). Journal Rank: CiteScore - **Q2** (Information Systems).

[4] **Moroșanu, Bogdan**, Marian Negru, and Constantin Paleologu. "Automated Personalized Loudness Control for Multi-Track Recordings." Algorithms 17, no. 6 (2024): 228. (Journal Rank: JCR - **Q2** (Computer Science, Theory and Methods) / CiteScore - **Q1** (Numerical Analysis)), WOS:001254517300001.

[5] **Moroșanu, Bogdan**, Marian Negru, Ana Neacșu, Cristian Negrescu, and Constantin Paleologu. "Personalized Multi-Track Leveling Algorithm." In 2023 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), pp. 24-29. IEEE, 2023.

[6] Negru, Marian, **Bogdan Moroșanu**, Ana Neacșu, Dragoș Drăghicescu, and Cristian Negrescu. "Automatic Audio Upmixing Based on Source Separation and Ambient Extraction Algorithms." In 2023 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), pp. 12-17. IEEE, 2023.

[7] Ioniță, Horia Sebastian, Victor Popa, and **Bogdan Moroșanu**. "Multiband Dynamics Compressor with Automatic Parameter Learning." In 2024 15th International Conference on Communications (COMM), pp. 1-4. IEEE, 2024.

5.4 Perspective de dezvoltare ulterioară

Cercetările prezentate în această teză deschid multiple direcții de dezvoltare ulterioară, dintre care se menționează:

1. Extinderea algoritmilor de inteligență artificială pentru producția muzicală

- Dezvoltarea unor sisteme complete de mixaj automat care să integreze recunoașterea instrumentelor, controlul nivelului, procesarea dinamică și poziționarea spațială într-un cadru unitar.
- Conceperea unor modele de învățare profundă capabile să capteze și să reproducă stilurile specifice unor producători sau genuri muzicale.
- Implementarea unor algoritmi de optimizare pentru setările parametrilor diverselor procesoare bazate pe analiza contextuală a materialului sonor.
- Perfecționarea modelelor multimodale de recunoaștere a instrumentelor prin integrarea informațiilor temporale și spectrale la nivel de proiect complet, nu doar la nivel de pistă individuală.

2. Sisteme adaptive pentru experiențe audio imersive

- Elaborarea unor algoritmi de expandare spațială adaptați pentru formate avansate precum Dolby Atmos sau sistemele audio *object-based*.
- Integrarea tehnicilor de personalizare a HRTF-urilor pentru îmbunătățirea reproducerii spațiale a sunetului pe căști.
- Dezvoltarea unor sisteme de producție muzicală interactive, care să se adapteze în timp real la preferințele utilizatorilor.
- Extinderea sistemului de recunoaștere a instrumentelor pentru a identifica automat pozițiile spațiale optime în mixaje surround sau *object-based*.

3. Evaluare perceptuală avansată

- Elaborarea unor protocoale standardizate pentru evaluarea subiectivă a soluțiilor automatizate de producție muzicală.
- Dezvoltarea unor metrici obiective îmbunătățite pentru cuantificarea calității procesării dinamice și expandării spațiale.
- Integrarea feedback-ului utilizatorilor în ciclul de învățare al modelelor, pentru rafinarea continuă a performanțelor.
- Dezvoltarea unor metode de evaluare a impactului clasificării automate a instrumentelor asupra fluxului de lucru și eficienței procesului de mixaj.

4. Construirea unei baze de date cuprinzătoare pentru cercetarea audio în context românesc

- Colectarea sistematică de materiale audio prin implicarea studenților în proiecte practice de înregistrare, mixaj și masterizare, valorificând astfel potențialul educațional și de cercetare al laboratoarelor universitare.
- Stabilirea unor parteneriate strategice cu studiouri de înregistrare, case de producție și artiști din industria muzicală românească, pentru obținerea accesului la înregistrări profesionale multi-pistă din diverse genuri muzicale.
- Documentarea și standardizarea metadatelor asociate înregistrărilor, incluzând informații despre tehnicile de captare, echipamentele utilizate și parametrii de procesare, pentru a facilita cercetări comparative și dezvoltarea algoritmilor specializați pe caracteristicile specifice producțiilor românești.
- Integrarea acestei baze de date într-o platformă colaborativă deschisă comunității academice, care să permită evaluarea și validarea noilor algoritmi de procesare audio, contribuind astfel la evoluția ecosistemului de cercetare audio din România.

- Elaborarea unui standard de nomenclatură pentru instrumentele muzicale în context românesc, care să faciliteze recunoașterea automată a instrumentelor în înregistrările locale și să contribuie la optimizarea fluxurilor de lucru în studiourile autohtone.

Aceste direcții de cercetare vor contribui la avansarea domeniului producției muzicale asistate de inteligență artificială, facilitând democratizarea expertizei de producție audio de înaltă calitate și deschizând noi posibilități creative pentru artiști și producători.

Bibliografie

- [1] Alexander, R. (2013). *The inventor of stereo: The life and works of Alan Dower Blumlein*. CRC Press.
- [2] Alvord, L. S. and Farmer, B. L. (1997). Anatomy and orientation of the human external ear. *Journal of the American Academy of Audiology*, 8(6).
- [3] Askwonder (2017). Innovators choose Wonder. <https://askwonder.com/research/total-addressable-market-music-makers-instrumentalists-rappers-producers-ebtqmsw4k>. [Accessed 13-04-2025].
- [4] Bai, M. R. and Shih, G.-Y. (2007). Upmixing and downmixing two-channel stereo audio for consumer electronics. *IEEE Transactions on Consumer Electronics*, 53(3):1011–1019.
- [5] Biem (2025). BIEM. <https://www.biem.org/members/directorySociety.do?method=detail&societyId=52&rMethod=membersDirectoryList&by=society&domain=&alpha=>. [Accessed 13-04-2025].
- [6] Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge.
- [7] Buckle, B. (2023). 120,000 new tracks released on streaming services every day, report finds — mixmag.net. <https://mixmag.net/read/120-000-new-tracks-are-released-on-streaming-services-every-day-report-finds-tech>. [Accessed 13-04-2025].
- [8] Chun, C. J., Kim, Y. G., Yang, J. Y., and Kim, H. K. (2009). Real-time conversion of stereo audio to 5.1 channel audio for providing realistic sounds. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2(4):85–94.
- [9] Colonel, J., Javed, S., and Valero-Fernandez, L. (2021). A general approach to transfer learning for audio classification tasks. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.
- [10] Cousins, M. and Hepworth-Sawyer, R. (2013). *Practical mastering: A guide to mastering in the modern studio*. Routledge.
- [11] De Man, B., Mora-Mcginity, M., Fazekas, G., and Reiss, J. D. (2014a). The open multitrack testbed. In *Audio Engineering Society Convention 137*. Audio Engineering Society.
- [12] De Man, B., Mora-Mcginity, M., Fazekas, G., and Reiss, J. D. (2014b). The open multitrack testbed: A new approach to the study of multitrack mixing. In *Audio Engineering Society Convention 137*, Los Angeles, USA. Audio Engineering Society.
- [13] Défossez, A. (2021). Hybrid spectrogram and waveform source separation. *arXiv preprint arXiv:2111.03600*.

- [14] Deruty, E. and Tardieu, D. (2011). About dynamic processing in mainstream music. *Journal of the Audio Engineering Society*, 59(5):300–311.
- [15] Engel, J., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D., and Simonyan, K. (2017). Neural audio synthesis of musical notes with wavenet auto-encoders. In *International Conference on Machine Learning*, pages 1068–1077. PMLR.
- [16] European Broadcasting Union (2014). Loudness normalisation and permitted maximum level of audio signals. Recommendation R128, EBU, Geneva, Switzerland. Retrieved from <https://tech.ebu.ch/publications/r128>.
- [17] Giannoulis, D., Massberg, M., and Reiss, J. D. (2012). Digital dynamic range compressor design—a tutorial and analysis. *Journal of the Audio Engineering Society*, 60(6):399–408.
- [18] Giannoulis, D., Massberg, M., and Reiss, J. D. (2013). Parameter automation in a dynamic range compressor. *Journal of the Audio Engineering Society*, 61(10):716–726.
- [19] Gibson, D. (2019). *The Art of Mixing: A Visual Guide to Recording, Engineering, and Production*. Routledge, New York, 3 edition.
- [20] Greenwood, D. D. (1991). Critical bandwidth and consonance in relation to cochlear frequency-position coordinates. *Hearing research*, 54(2):164–208.
- [21] Han, Y. and Lee, K. (2016). Acoustic scene classification using convolutional neural network and multiple-width frequency-delta data augmentation. *IEEE Signal Processing Letters*, 23(12):1649–1653.
- [22] Hawley, S. H. and Scheirer, W. J. (2020). Speech enhancement using deep learning and its application to the detection, classification and segmentation of sound events. *Frontiers in Computer Science*, 2:5.
- [23] Howard, D. and Angus, J. (2013). *Acoustics and psychoacoustics*. Routledge.
- [24] Huber, D. M., Caballero, E., and Runstein, R. (2023). *Modern Recording Techniques: A Practical Guide to Modern Music Production*. CRC Press.
- [25] IBISWorld, I. (2024). Sound Recording & Music Publishing in Romania - Market Research Report (2014-2029). <https://www.ibisworld.com/romania/industry/sound-recording-music-publishing/200261/#IndustryStatisticsAndTrends>. [Accessed 13-04-2025].
- [26] International Telecommunication Union (2015). Algorithms to measure audio programme loudness and true-peak audio level. Recommendation BS.1770-4, International Electrotechnical Commission, Geneva, Switzerland. Retrieved from <https://www.itu.int/rec/R-REC-BS.1770>.
- [27] Irwan, R. and Aarts, R. M. (2002). Two-to-five channel sound processing. *Journal of the Audio Engineering Society*, 50(11):914–926.
- [28] Izhaki, R. (2017). *Mixing Audio: Concepts, Practices, and Tools*. Focal Press, London, 3 edition.
- [29] Katoch, S., Chauhan, S. S., and Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80(5):8091–8126.

- [30] Katz, B. and Katz, R. A. (2003). *Mastering audio: the art and the science*. Butterworth-Heinemann.
- [31] Linkwitz, S. H. (1978). Passive crossover networks for noncoincident drivers. *Journal of the Audio Engineering Society*, 26(3):149–150.
- [32] Mirjalili, S. (2019). Genetic algorithm: Theory, literature review, and application in image reconstruction. *Nature-Inspired Optimizers*, pages 69–85.
- [33] Miskiewicz, A. (1992). Timbre solfege: A course in technical listening for sound engineers. *Journal of the Audio Engineering Society*, 40(7/8):621–625.
- [34] Mitsufuji, Y., Fabbro, G., Uhlich, S., Stöter, F.-R., Défossez, A., Kim, M., Choi, W., Yu, C.-Y., and Cheuk, K.-W. (2022). Music demixing challenge 2021. *Frontiers in Signal Processing*, 1:18.
- [35] Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (1995). Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, 43(5):300–321.
- [36] Moroşanu, B., Negru, M., Nicolae, G., Ioniţă, H., and Paleologu, C. (2025). Automated personalized loudness control for multi-track recordings. *Information*, 16(5).
- [37] Moroşanu, B., Negru, M., and Paleologu, C. (2024). Automated personalized loudness control for multi-track recordings. *Algorithms*, 17(6):228.
- [38] Motoc, G. (2024). Sizing up the romanian music industry. <https://georgemotoc.com/2024/01/01/sizing-up-the-romanian-music-industry/>. [Accessed 13-04-2025].
- [39] MusicBrainz (2023). MusicBrainz - the open music encyclopedia. <https://musicbrainz.org/>. [Accessed 13-04-2025].
- [40] Negru, M., Moroşanu, B., Neacşu, A., Drăghicescu, D., and Negrescu, C. (2023). Automatic audio upmixing based on source separation and ambient extraction algorithms. In *2023 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pages 12–17. IEEE.
- [41] Niall (2023). How Many Official Musicians Are There in the World? <https://bigtimemusicians.com/how-many-official-musicians-are-there-in-the-world/>. [Accessed 13-04-2025].
- [42] Owsinski, B. (2006). *The mixing engineer's handbook*. Boston: Thomson Course Technology.
- [43] Popa, V., Moroşanu, B., and Negrescu, C. (2023). Head related transfer function measurement in reverberant rooms. In *Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies XI*, volume 12493, pages 617–622. SPIE.
- [44] Pulkki, V. (2006). Directional audio coding in spatial sound reproduction and stereo upmixing. In *Audio Engineering Society Conference: 28th International Conference: The Future of Audio Technology—Surround and Beyond*. Audio Engineering Society.
- [45] Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S.-Y., and Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 13(2):206–219.

- [46] Robles, L. and Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiological reviews*, 81(3):1305–1352.
- [47] Rumsey, F. (2013). Mastering for today’s media. *Journal of the Audio Engineering Society*, 61(1/2):79–83.
- [48] Senior, M. (2018). Mixing secrets for the small studio. *Taylor & Francis*.
- [49] Series, B. (2010). Multichannel stereophonic sound system with and without accompanying picture. *International Telecommunication Union Radiocommunication Assembly*.
- [50] Stassen, M. (2023). There are now 120,000 new tracks hitting music streaming services each day. <https://www.musicbusinessworldwide.com/there-are-now-120000-new-tracks-hitting-music-streaming-services-each-day/>. [Accessed 13-04-2025].
- [51] Stassen, M. (2024). ‘Romania is an emerging music market that’s growing at a fast pace. It’s catching up with more developed markets.’ - Music Business Worldwide — [musicbusinessworldwide.com](http://disq.us/t/4omwokx). <http://disq.us/t/4omwokx>. [Accessed 13-04-2025].
- [52] Storer, J. (2004). Home - JUCE. <https://juce.com/>. [Accessed 01-04-2025].
- [53] Théberge, P. (2012). The end of the world as we know it: The changing role of the studio in the age of the internet. *The art of record production: An introductory reader for a new academic field*, pages 77–90.
- [54] Vickers, E. (2010). The loudness war: Background, speculation, and recommendations. In *Audio Engineering Society Convention 129*. Audio Engineering Society.
- [55] Wang, S., Germain, F. G., Simionato, R., and Bello, J. P. (2023). Neural parametric equalizer matching using differentiable biquads. *Applied Sciences*, 13(2):800.
- [56] Zwicker, E. and Fastl, H. (2013). *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media.