Sistem de Sinteza Vizual-Spectrala

O Abordare Hibrida (DSP + Deep Learning) pentru Reconstructia Fazei

Izabela Jilavu Theodor Ancuta

Universitatea din Bucuresti

5 februarie 2026

Cuprins

Motivatia: De la Imagine la Sunet

Sonificarea: Transmiterea informației vizuale prin semnale acustice non-verbale.

Domenii de Aplicabilitate:

1. **Substituție Senzorială (Nevăzători):**
   - Maparea obstacolelor în sunet.
   - Sus $\rightarrow$ Frecvențe înalte / Jos $\rightarrow$ Joase.
2. **Design Sonor:**
   - Desenarea timbrului direct pe spectrogramă.
3. **Steganografie Acustică:**
   - Ascunderea imaginilor în spectrul audio.

Provocarea

Transformarea inversă (Imagine $\rightarrow$ Audio) este mult mai dificilă decât Vizualizarea (Audio $\rightarrow$ Imagine) din cauza pierderii informației temporale.

Problema Matematica: Phase Retrieval

Semnalul audio în domeniul timp-frecvență (STFT):

$$S(t,f) = \underbrace{|S(t,f)|}_{\text{Magnitudine (Imagine)}} \cdot \underbrace{e^{j\phi(t,f)}}_{\text{Faza (NECUNOSCUTA)}}$$

Problema Inversă (Ill-posed)

O imagine digitală oferă doar **Magnitudinea** (M).

- Fără fază (ϕ), nu putem aplica transformata inversă (ISTFT).
- Dacă presupunem $\phi = 0$, sunetul rezultat este metalic și neclar.

Obiectiv: Estimarea $\hat{\phi}$ pentru o reconstrucție fidelă.

State-of-the-art: De la GL la abordari hibride

- **Griffin-Lim (GL):** standard clasic, dar convergenta lenta si minime locale.
- **Fast Griffin-Lim (FGL):** momentum ($\alpha = 0.99$) $\rightarrow$ 30 iteratii (**3.1x** mai rapid).
- **Retele neurale:** rapide, dar uneori inflexibile (conditionare pe reprezentari specifice).
- **Hybrid (AI + DSP):** PhaseUNet + rafinare Griffin-Lim cu putine iteratii.

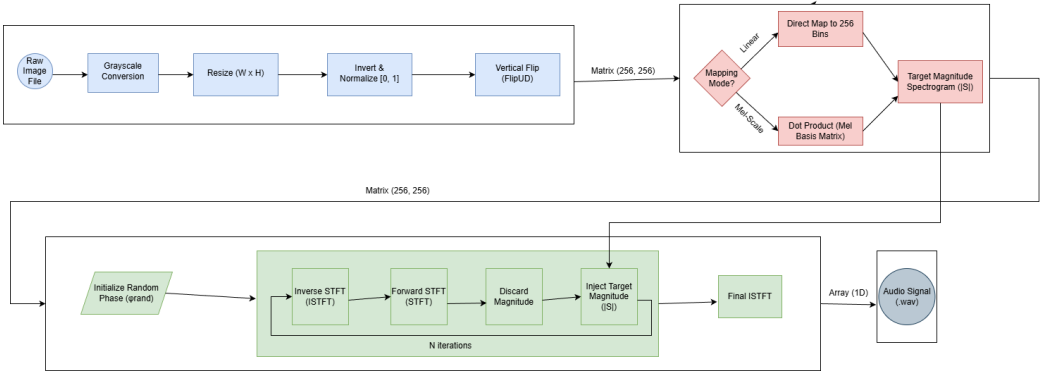
Ideea cheie

Warm-start reduce costul iterativ pastrand consistenta spectrala.

Pipeline-ul de Procesare DSP

Implementare *from scratch* a lanțului de procesare:

1. **Preprocesare:** Imagine 256×256.
2. **Analiza STFT:** $N_{FFT} = 510$ (pentru mapare exactă pe 256 pixeli).
3. **Recuperare Fază:** Algoritmul Griffin-Lim (GL).
4. **Sinteză ISTFT:** Metoda Overlap-Add.



Algoritmul Griffin-Lim: Pași (Phase Retrieval)

Input: Magnitudinea spectrului $M(t,f) = |S(t,f)|$ (din imagine) și o fază inițială $\phi^{(0)}$.

Iterația $k \to k+1$

1. **Recombinare:** $\hat{S}^{(k)}(t,f) = M(t,f)\, e^{j\phi^{(k)}(t,f)}$.
2. **Timp:** $\hat{x}^{(k)} = \mathrm{ISTFT}(\hat{S}^{(k)})$.
3. **Înapoi în TF:** $\tilde{S}^{(k)} = \mathrm{STFT}(\hat{x}^{(k)})$.
4. **Proiecție pe magnitudine:** $\hat{S}^{(k+1)} = M\, e^{j\angle \tilde{S}^{(k)}}$.
5. **Update fază:** $\phi^{(k+1)} = \angle \tilde{S}^{(k)}$.

Observație

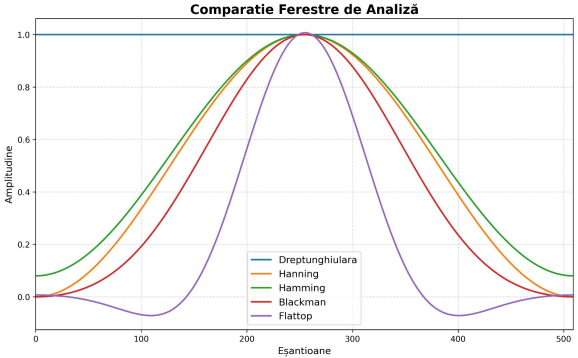
Algoritmul alternează între două constrângeri: *magnitudine fixă* în TF și *consistență STFT* în timp.

Analiza Ferestrelor si Spectral Leakage

Alegerea ferestrei $w[n]$ este un compromis între rezoluție și "curățenie" spectrală.

- **Rectangular:** Rezoluție bună, scurgeri mari (-13dB).
- **Hanning:** Echilibru optim, reconstrucție perfectă.
- **Blackman:** Atenuare excelentă (-58dB), dar lob lat.

Decizie: S-a utilizat fereastra **Hanning**.



Optimizare: Fast Griffin-Lim (FGL)

Griffin-Lim Clasic: Convergență lentă, necesită multe iterații.
Fast Griffin-Lim (Implementat): Utilizează *momentum* pentru accelerare ($\alpha = 0.99$).

Tabela: Benchmark Comparativ (CPU i7-10700K)

Metoda	Timp (s)	SNR (dB)	Eroare (SC)
GL-100	0.208	0.73	0.149
FGL-30	**0.066**	**1.37**	**0.123**

Rezultat

FGL-30 este de **3.1x mai rapid** și oferă o calitate superioară (SNR mai mare) față de GL standard cu 100 iterații.

"The Pivot": Schimbarea de Paradigma

1. **Abordarea Inițială (Eșec)**
   - Tip: Image Restoration (Deblurring).
   - Problemă: Imaginile naturale dense spectral produc zgomot. CNN-ul nu găsește o fază coerentă.

2. **Abordarea Finală (Succes)**
   - Tip: **Sinteză Spectrală (Sparse)**.
   - Preprocesare: **Canny Edge Detection**.
   - Logică: Fourier / Far-field. O linie $\rightarrow$ Un ton pur.

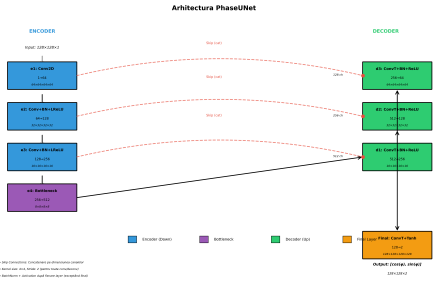
Arhitectura PhaseUNet

Retea de tip **U-Net** cu Skip Connections.

- **Input:** Imagine 128×128 (Edges).
- **Encoder:** 4 blocuri convolutionale (downsampling la 8×8).
- **Decoder:** Refacerea dimensiunii spațiale.

Inovație Output (Cartezian): Pentru a evita discontinuitatea fazei $[-\pi, \pi]$, rețeaua prezice:

$$\text{Output} = [\cos\phi, \sin\phi]$$



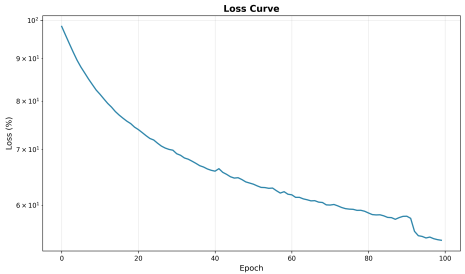
Antrenare si Date Sintetice

Dataset: Generat procedural "on-the-fly" (10.000 exemple).

1. **Tonuri Pure (Sinusoide):** Linii orizontale.
2. **Chirps (Glissando):** Linii diagonale (tranzitii de frecventa).
3. **Mixaje Polifonice:** Interferențe complexe.

Configurare:

- Optimizer: Adam
- Loss: MSE pe $(\cos, \sin)$
- Epoci: 100



Abordarea Hibrida: AI + DSP

Deși AI-ul este rapid, poate introduce mici erori. Soluția optimă este combinarea celor două lumi.

Strategia "Warm Start"

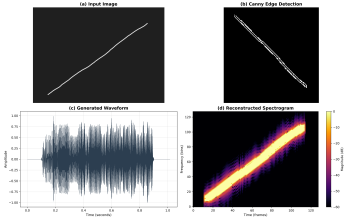
1. **Initializare:** PhaseUNet estimează structura globală a fazei (15ms).
2. **Rafinare:** Algoritmul Griffin-Lim rulează doar **30 de iterații**.

Rezultate:

- Evitarea minimelor locale (fără sunet metalic).
- Reducerea timpului total cu **68%** (0.8s vs 2.5s).

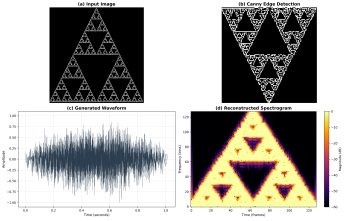
Studii de Caz: Reconstructie Vizuala

Diagonala



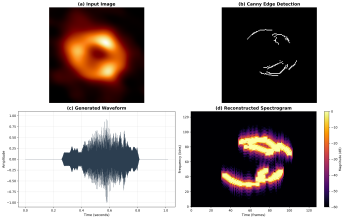
Glissando continuu

Fractal Sierpinski



Structură auto-similară

Gaură Neagră



Gradient radial

Sistemul respectă Limita Gabor (Incertitudinea Timp-Frecvență).

Concluzii

1. **Implementare DSP:** Validarea ferestrelor și algoritmului STFT.
2. **Performanță:** FGL-30 depășește GL-100 (**3.1x** mai rapid, SNR mai bun).
3. **Deep Learning:** PhaseUNet (output $[\cos\phi, \sin\phi]$) stabilizeaza predictia fazei.
4. **Hybrid:** warm-start + 30 iteratii GL ofera un compromis optim viteza/calitate.

Directii Viitoare

- Extindere la **Stereofonie** (Mapare axa X $\rightarrow$ Panoramare).
- Implementare **Real-Time** (TensorRT / Mobile).

Vă mulțumim!

Întrebări?

Demo Live: `https://academo.org/demos/spectrum-analyzer/`
Cod sursă disponibil la: `https://github.com/izabelamaria24/visual-spectral-synthesis`