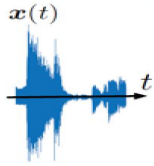


**Un cadre  
perceptual-neuronal-physique pour l'  
écoute artificielle**

Han Han  
LS2N, CNRS, École Centrale Nantes

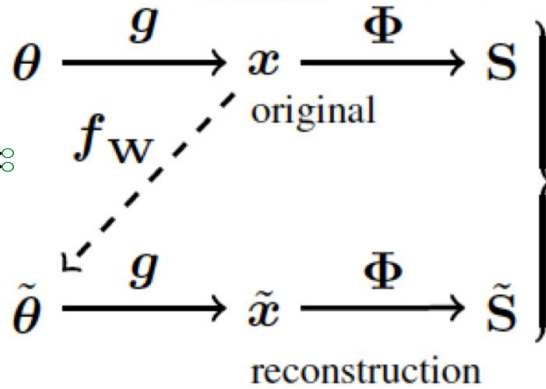
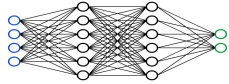
# Problématique: Appariement Sonore (Sound Matching)



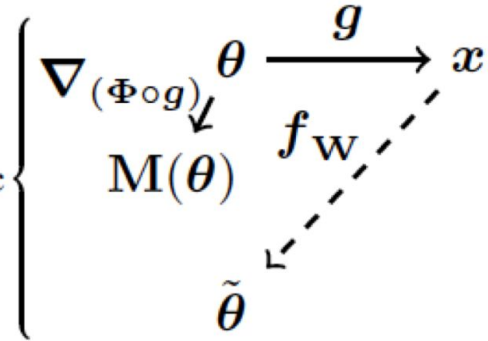
Physical domain

Audio domain

Perceptual domain



Riemannian metric



## Idée Clé

$$L_{\theta}^{DDSP}(W) = \frac{1}{2} \|\Phi(\tilde{x}) - \Phi(x)\|_2^2$$

Taylor expand  $\Phi(\tilde{x})$

$$\Phi(\tilde{x}) = \Phi \circ g(\tilde{\theta}) = \Phi \circ g(\theta) + \nabla_{(\Phi \circ g)}(\theta) \cdot (\tilde{\theta} - \theta) + O(\|\tilde{\theta} - \theta\|_2^2)$$

$$L_{\theta}^{DDSP} = \frac{1}{2} \langle \tilde{\theta} - \theta \mid \underbrace{\nabla_{(\Phi \circ g)}^T(\theta) \nabla_{(\Phi \circ g)}(\theta)}_{M(\theta)} \tilde{\theta} - \theta \rangle + O(\|\tilde{\theta} - \theta\|_2^3)$$

$L_{\theta}^{PNP}$

P-loss:  $M(\theta) = I$

Weighted P-loss:  $M(\theta) = D$

Spectral loss:  $M(\theta) = \nabla^T \Phi \circ g(\tilde{\theta}) \nabla \Phi \circ g(\tilde{\theta})$

## Application sur les sons percussifs

$\Phi$  : Joint Time-frequency Scattering coefficient **~20k coefficients**

$\mathcal{g}$  : Functional Transformation method PDE solver **3s audio**

$\theta$  : Physical parameters of a rectangular drum model **5 parameters**

$f_W$  : EfficientNet B0 **4M weights**

$$L_{\theta}^{PNP} = \frac{1}{2} \langle \tilde{\theta} - \theta | M(\theta) + \lambda I | \tilde{\theta} - \theta \rangle$$

