

Deep learning for musical structure analysis and generation

DYCI2 Meeting - Ircam 20/09/2017

Superviseur :

Frederic Bimbot Emmanuel Vincent

Problématique

Peut-on capturer, générer la structure d'un morceau de musique avec des méthodes d'apprentissage profond ?

Spécificité ?

Contrainte ?

Méthode ?

Modèle ?

Plan

1. Corpus
2. Représentation des données
3. Architecture de modèle d'apprentissage
4. Résultat
5. Perspective

Corpus

Données brutes

Lakh MIDI dataset :

- 176,581 fichier MIDI
- Tous genres confondus
- Multipiste
- Polyphonique
- Toute structure



Données expérimentales

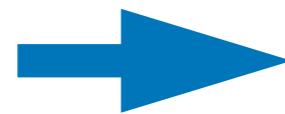
Sous-ensemble :

- 786 fichier MIDI
- Tous genres confondus
- Mélodie
- Monophonique
- Structure 8 mesures

Représentation

Transformation des données

Multipiste



Extraction

Melodie



Matrice



Moins de dimension
Moins de complexité

TIME

```
[0001000000000]  
[0100000000000]  
[0000001000000]  
[1000000000000]  
.  
.  
.  
[1000000000000]  
[0000100000000]  
[0000010000000]
```

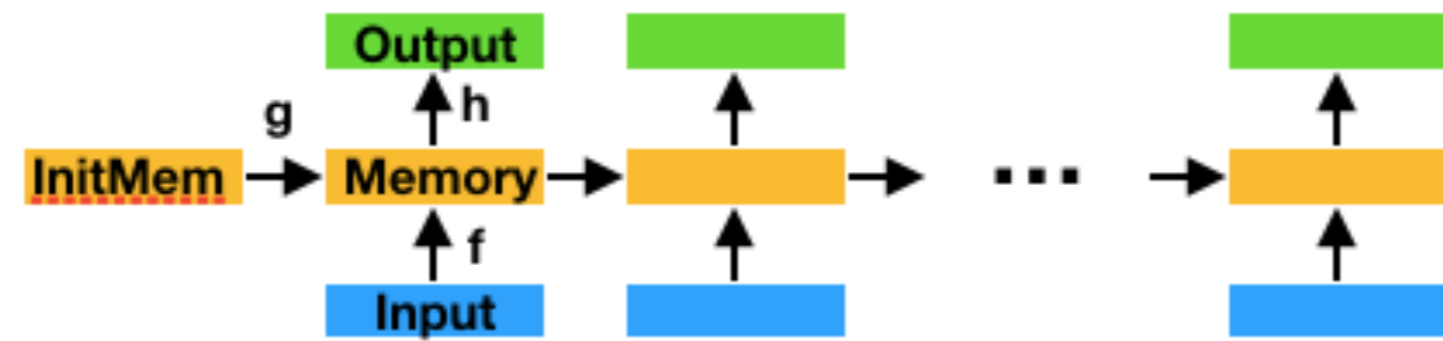
NOTE

Représentation vectorielle des notes

Architecture de base

Réseau récurrent

On veut générer de la musique à partir d'une graine

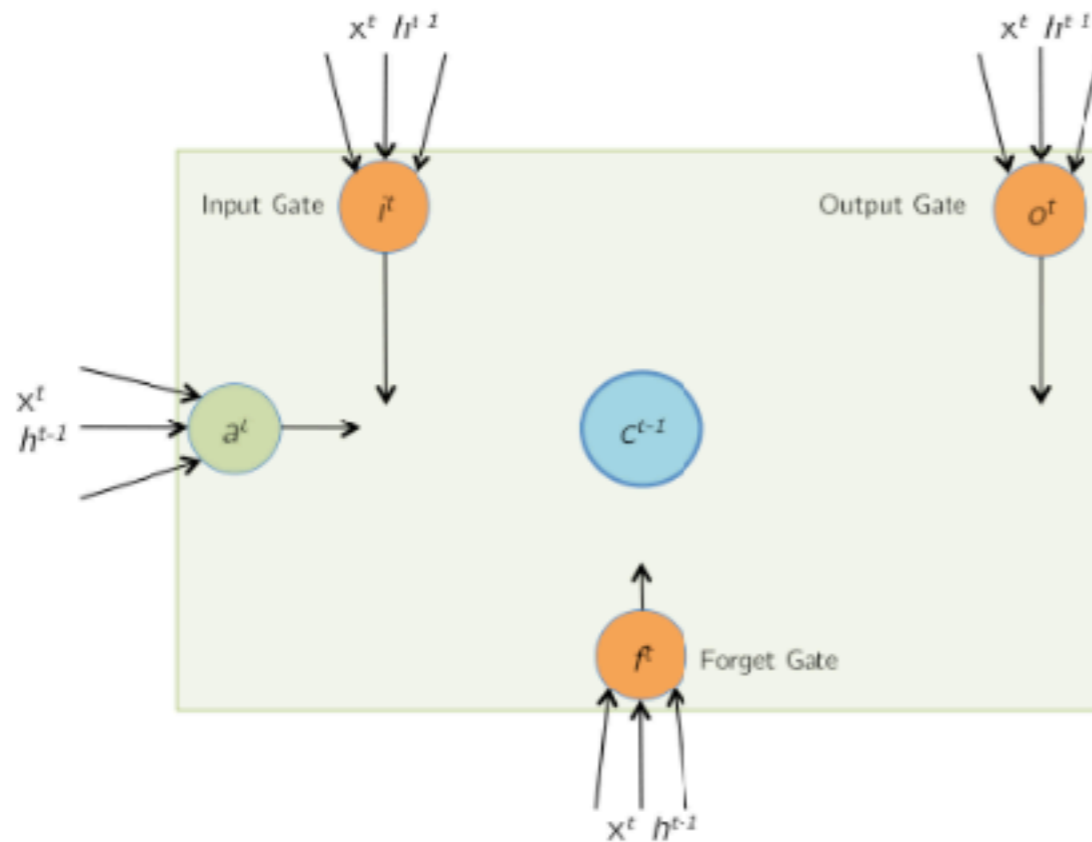


Hypothèse : la mémoire capture des éléments d'ordre structurel

Architecture LSTM

Long Short Term Memory

4 portes



$$a^t = \tanh(W_c x^t + U_c h^{t-1}) = \tanh(a^{\wedge})$$

$$i^t = \sigma(W_i x^t + U_i h^{t-1}) = \sigma(i^{\wedge})$$

$$f^t = \sigma(W_f x^t + U_f h^{t-1}) = \sigma(f^{\wedge})$$

$$o^t = \sigma(W_o x^t + U_o h^{t-1}) = \sigma(o^{\wedge})$$

Ignoring the non-linearities,

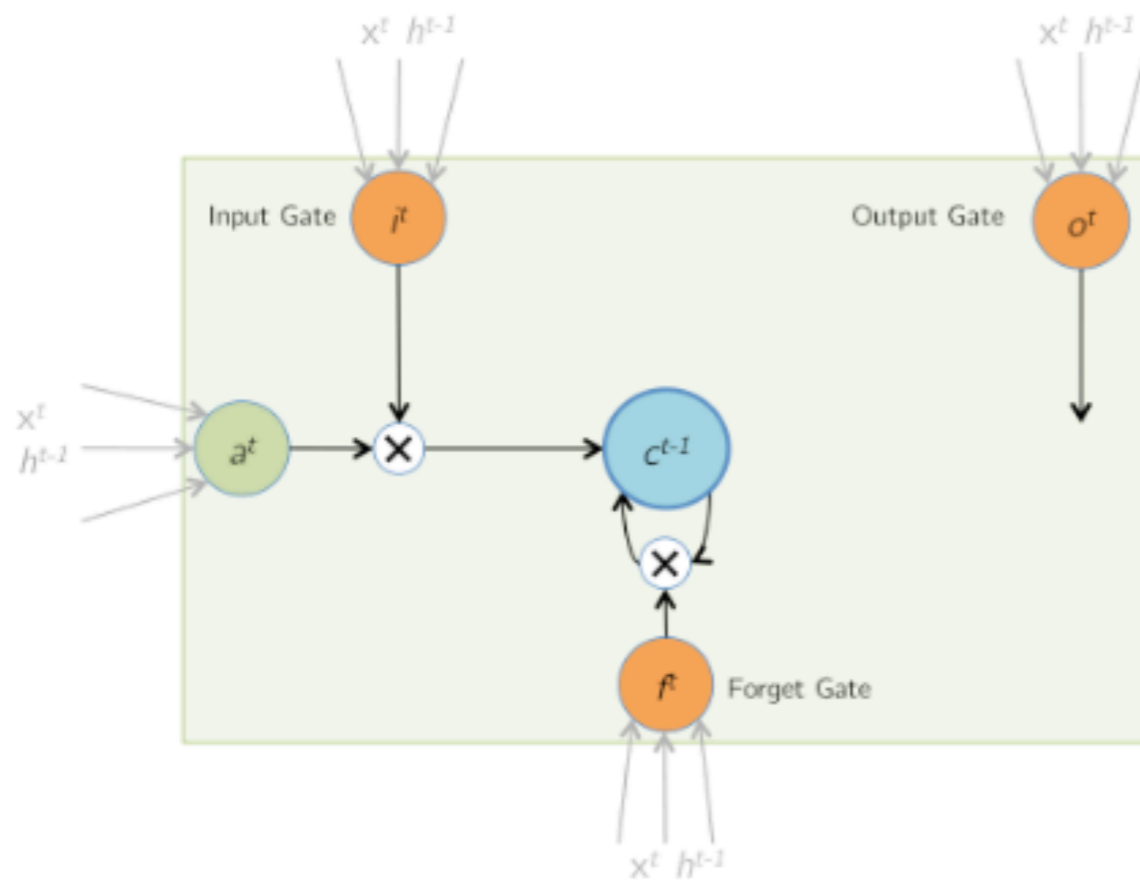
$$z^t = \begin{bmatrix} a^{\wedge} \\ i^{\wedge} \\ f^{\wedge} \\ o^{\wedge} \end{bmatrix} = \begin{bmatrix} W^c & U^c \\ W^i & U^i \\ W^f & U^f \\ W^o & U^o \end{bmatrix} \times \begin{bmatrix} x^t \\ h^{t-1} \end{bmatrix} = W \times I^t$$

C_t -> vecteur représentant la mémoire (état)

Architecture LSTM

Long Short Term Memory

Mise à jour de la mémoire courante

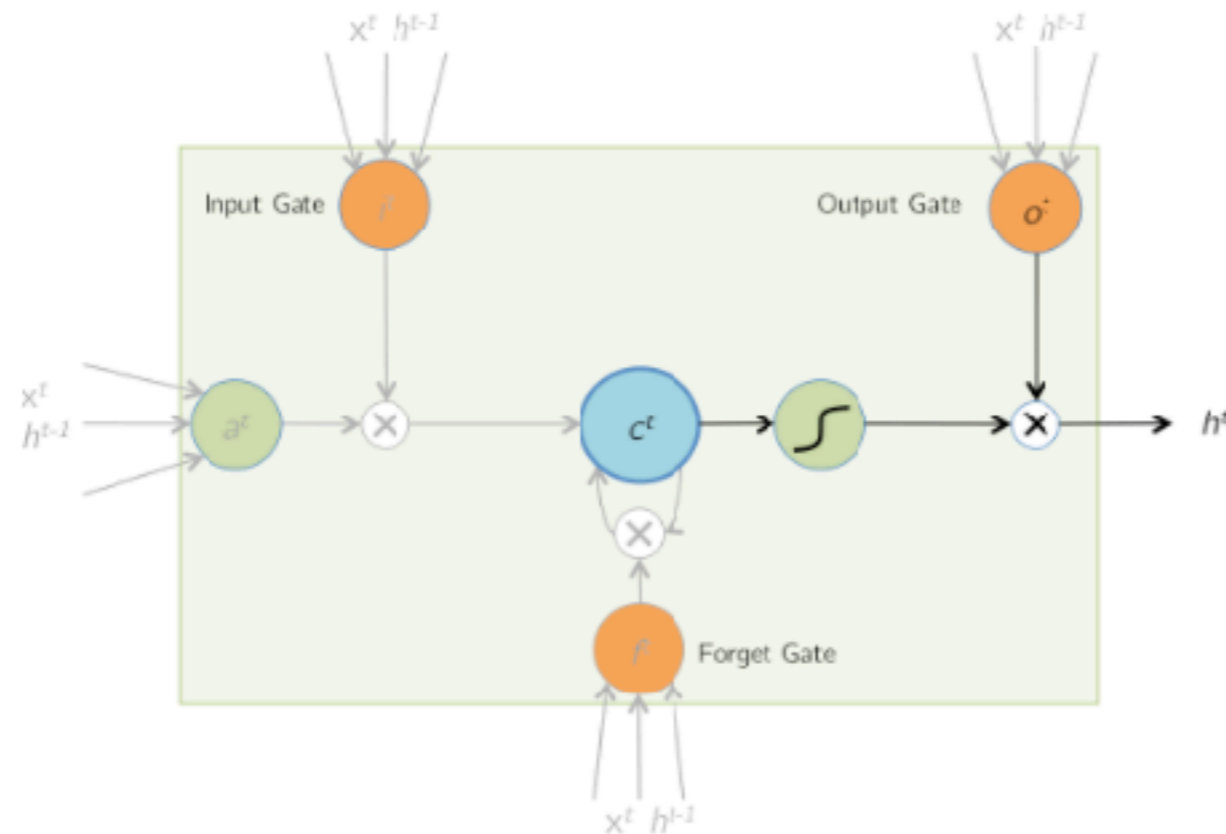


$$c^t = i^t \odot a^t + f^t \odot c^{t-1}$$

Architecture LSTM

Long Short Term Memory

La sortie finale est filtrée par la porte de sortie



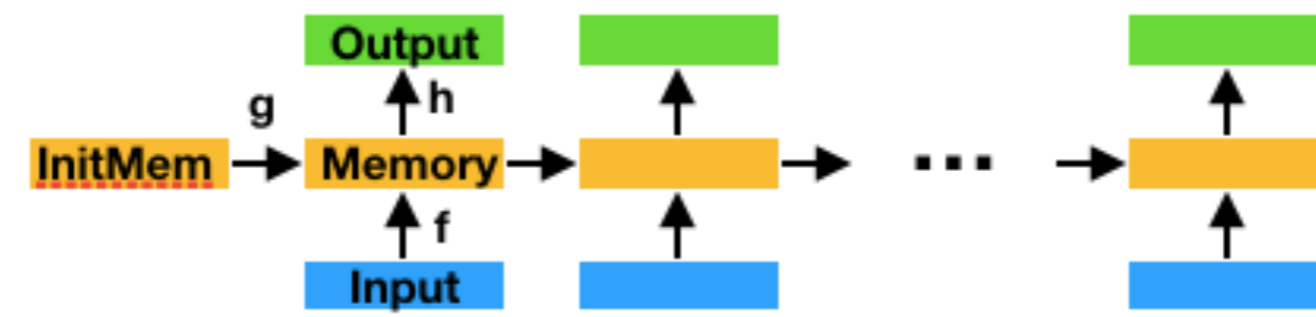
$$h^t = o^t \odot \tanh(c^t)$$

Architecture LSTM

Long Short Term Memory

Démo

Entraîné sur les données brutes (~1000)

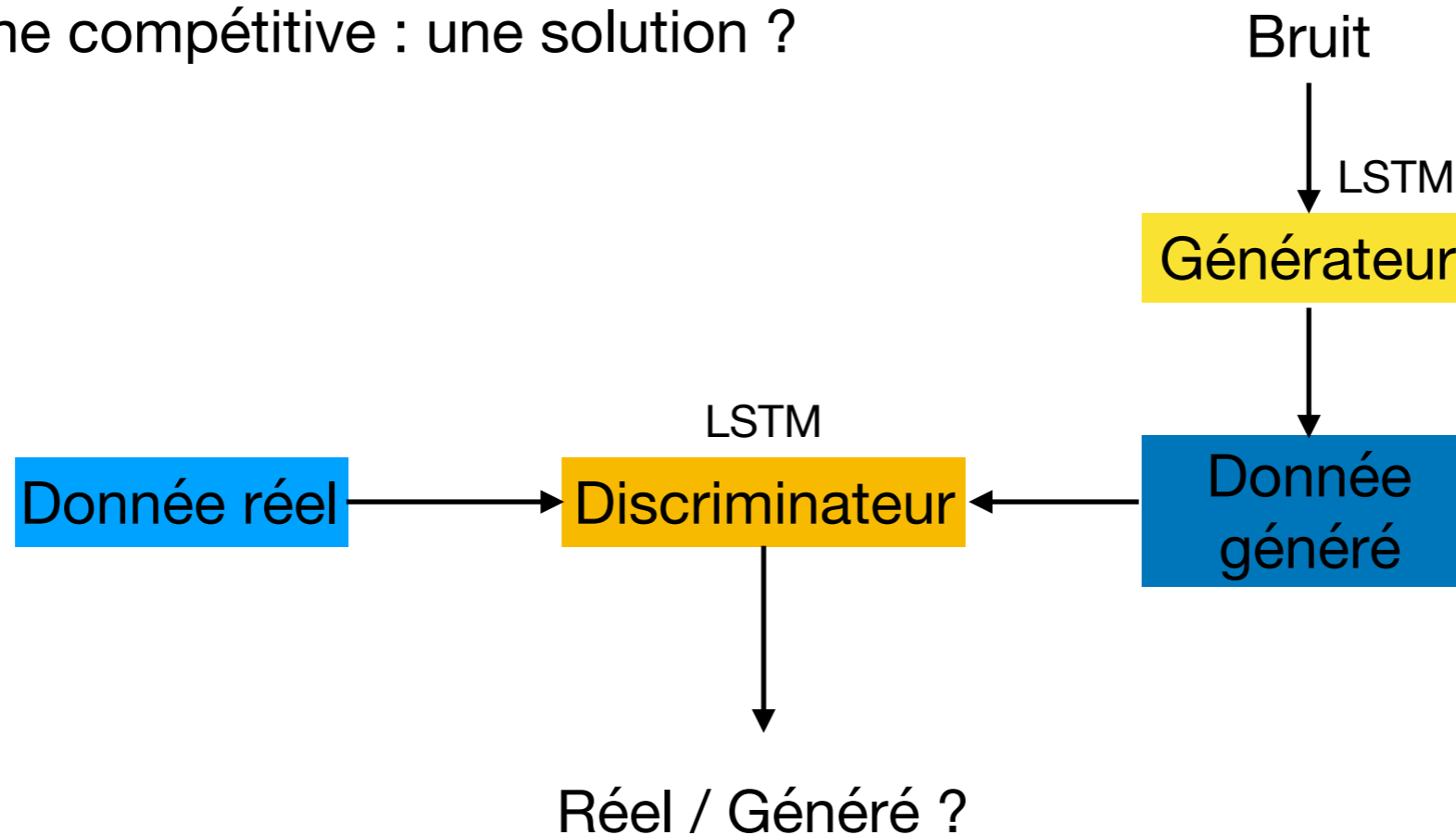


Testé par l'écoute (problématique)

Architecture GAN

Generative Adversarial Network

Approche compétitive : une solution ?



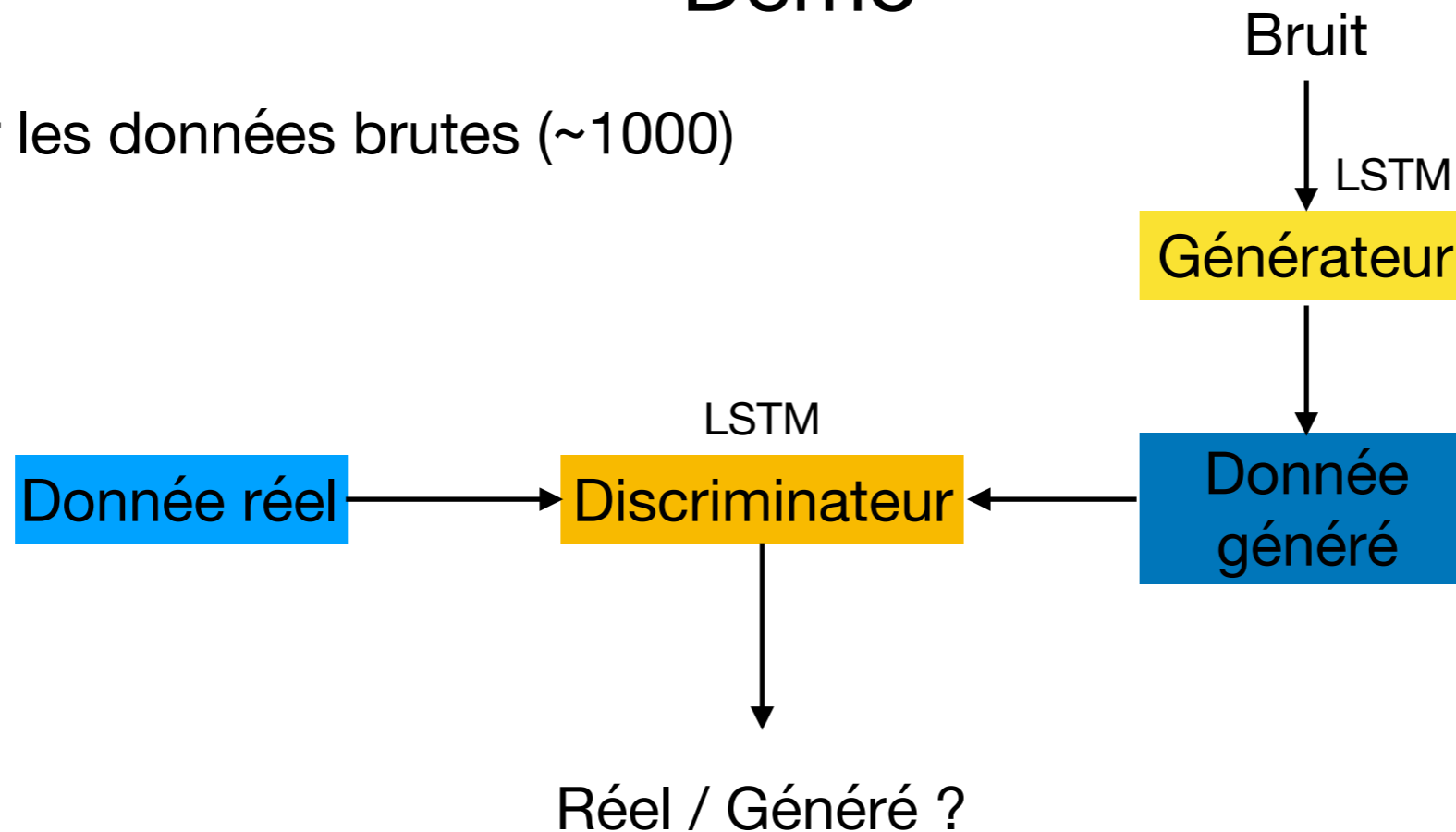
Le discriminateur est appris en même temps que le générateur

Architecture GAN

Generative Adversarial Network

Démo

Entrainé sur les données brutes (~1000)



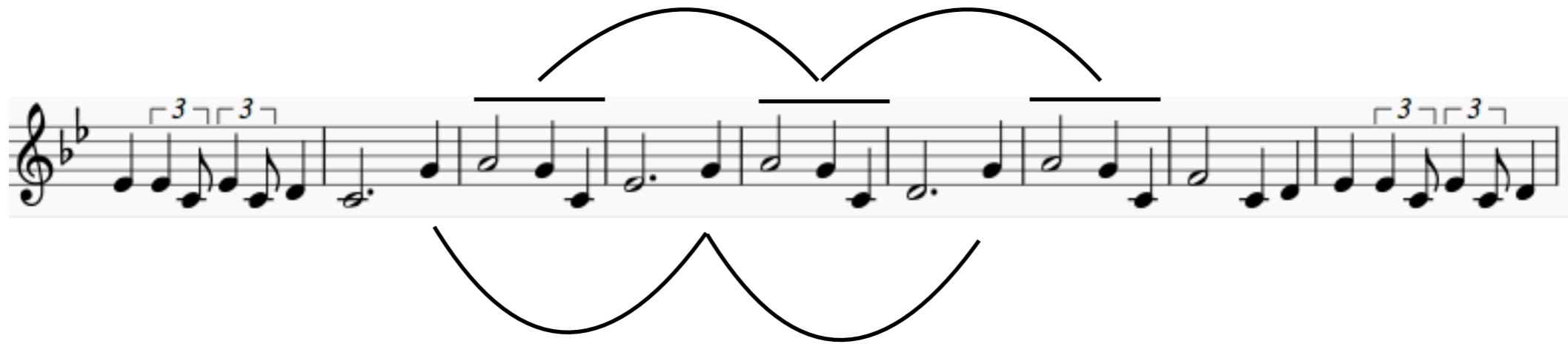
Ne parvient pas à capturer la structure

Comparer discriminateur LSTM et GAN

Invariance

Enseignement tiré des expériences :

La contrainte d'invariance par translation dans le temps



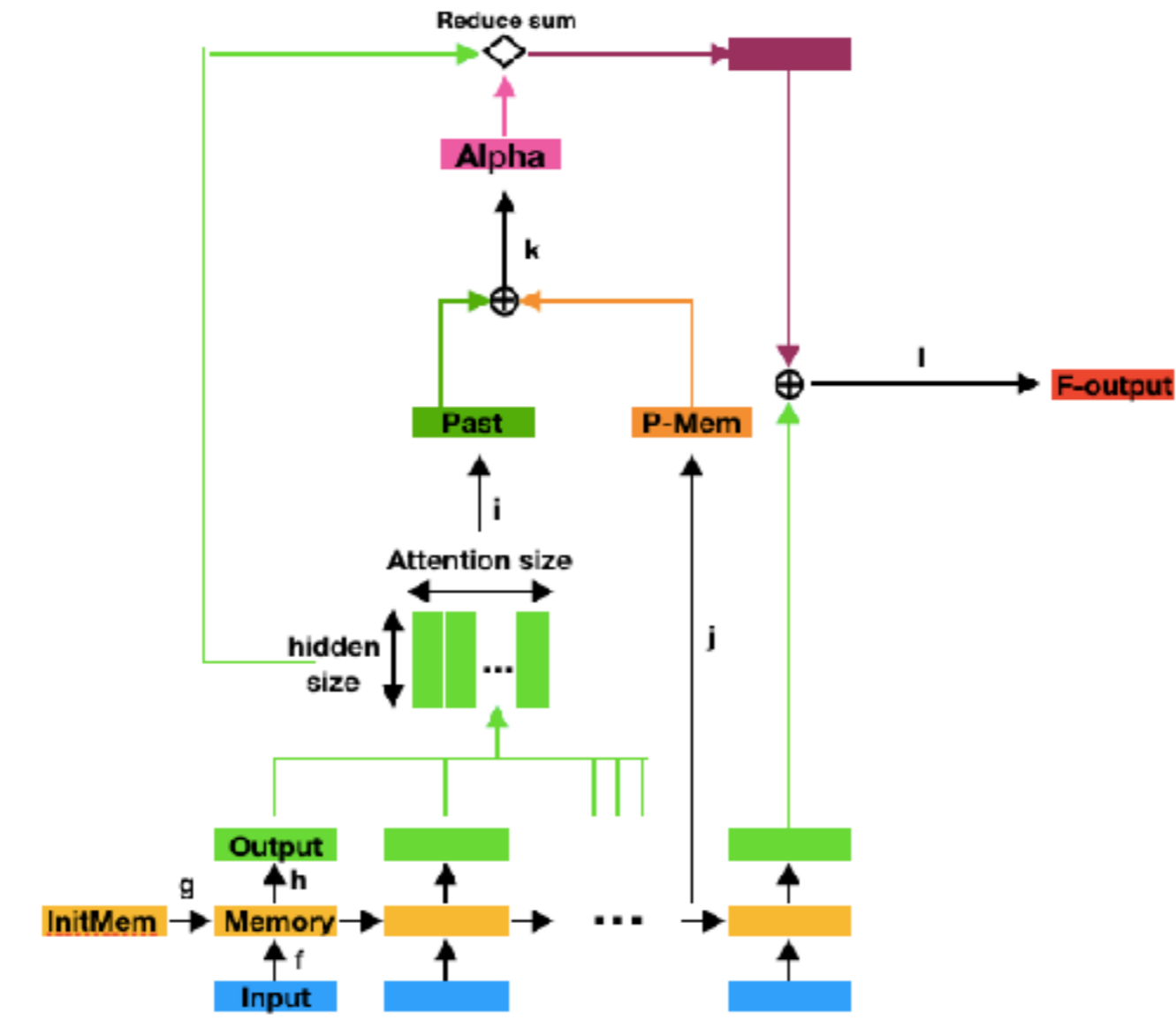
Model de sélection d'historique

-Un réseaux de neurones par emplacement de note

-Modèle d'attention

Modèle d'Attention

Apprendre à sélectionner l'historique



Plusieurs possibilités de comparaison, mémoire/historique

Perspective

- Inférence et Analyse de différents objets de comparaison pour le modèle d'attention
- Un réseau par placement temporel pour l'analyse de la structure
- Représentation des données, embedding