



Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI

# Singing Voice Detection with Deep Recurrent Neural Networks

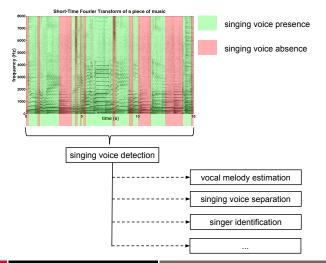
Simon Leglaive, Romain Hennequin and Roland Badeau

April 24, 2015

40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015, April 19-24 2015, Brisbane, Australia

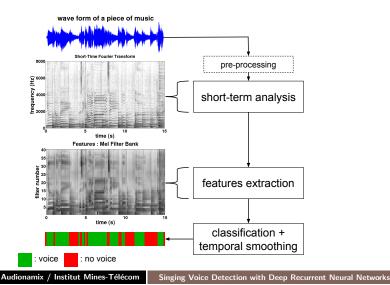
Recurrent Neural Networks and Long Short-Term Memory System Overview Results Conclusion

#### Introduction



Recurrent Neural Networks and Long Short-Term Memory System Overview Results Conclusion

#### Introduction



Recurrent Neural Networks and Long Short-Term Memory System Overview Results Conclusion



Our method :

 Bidirectional Long-Short Term Memory (BLSTM) Recurrent Neural Network (RNN)

 $\longrightarrow$  long past and future temporal context

With several hidden-layers

 $\longrightarrow$  extract simple useful information from low level features

Recurrent Neural Networks and Long Short-Term Memory System Overview Results Conclusion

# Outline

# Recurrent Neural Networks and Long Short-Term Memory

Artificial Neural Network Long Short-Term Memory Bidirectional Recurrent Neural Networks

#### System Overview

Double HPSS Global system Building the Network

#### Results

Dataset Network functioning Results

Artificial Neural Network Long Short-Term Memory Bidirectional Recurrent Neural Networks

# Outline

#### Recurrent Neural Networks and Long Short-Term Memory Artificial Neural Network Long Short-Term Memory Bidirectional Recurrent Neural Networks

#### System Overview

Double HPSS Global system Building the Network

#### Results

Dataset Network functioning Results

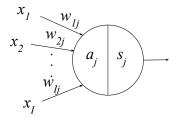
Artificial Neural Network

Long Short-Term Memory Bidirectional Recurrent Neural Networks

# **Formal Neuron**

• Activation : 
$$a = \sum_{i=1}^{l} w_i x_i$$

Output : s = f(a) with f the nonlinear activation function (e.g. step function, sigmoid, hyperbolic tangent, ...)

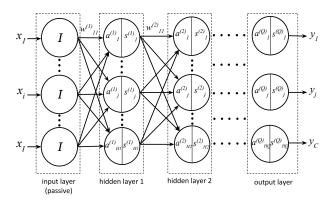


#### Artificial Neural Network

Long Short-Term Memory Bidirectional Recurrent Neural Networks

# **Multi-Layer Perceptron**

Conclusion

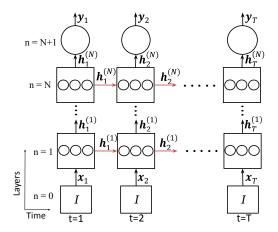


- Feedforward Artificial Neural Network
- Maps inputs to outputs by propagating data through the layers
- Training : Gradient descent using backpropagation

#### Artificial Neural Network

Long Short-Term Memory Bidirectional Recurrent Neural Networks

### **Recurrent Neural Network**



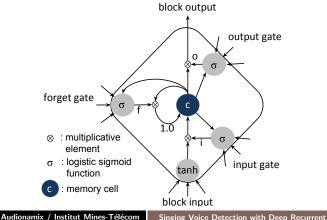
Recurrent Neural Network unfolded in time

Audionamix / Institut Mines-Télécom Singing Voice Detection with Deep Recurrent Neural Networks

Artificial Neural Network Long Short-Term Memory Bidirectional Recurrent Neural Networks

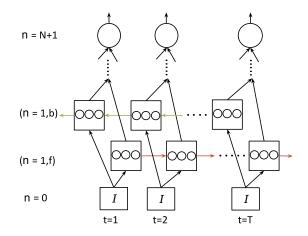
# Long Short-Term Memory

- Memory cell
- Input Gate, Output Gate, Forget Gate = Write, Read and Reset



Artificial Neural Network Long Short-Term Memory Bidirectional Recurrent Neural Networks

### **Bidirectional Recurrent Neural Network**



Bidirectional Recurrent Neural Network unfolded in time

Audionamix / Institut Mines-Télécom Singing Voice Detection with Deep Recurrent Neural Networks

Double HPSS Global system Building the Network

# Outline

Recurrent Neural Networks and Long Short-Term Memory Artificial Neural Network Long Short-Term Memory Bidirectional Recurrent Neural Networks

Conclusion

#### System Overview

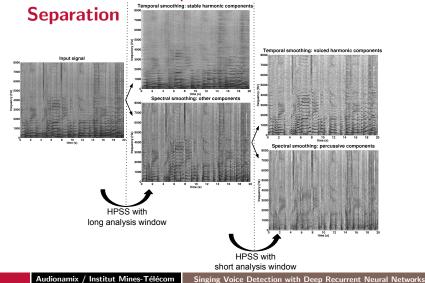
Double HPSS Global system Building the Network

#### Results

Dataset Network functioning Results

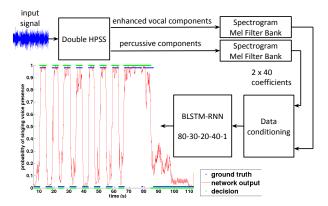
Double HPSS Global system Building the Network

# **Double Harmonic/Percussive Source**



Double HPSS Global system Building the Network

### **Global System**

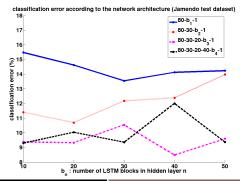


Conclusion

Double HPSS Global system Building the Network

# **Building the Network**

- ► No theoretical evidence → empirical approach, not much discussed in papers
- Incremental procedure : depth increased by progressively adding hidden layers



Dataset Network functioning Results

# Outline

# Recurrent Neural Networks and Long Short-Term Memory

Artificial Neural Network Long Short-Term Memory Bidirectional Recurrent Neural Networks

#### System Overview

Double HPSS Global system Building the Network

#### Results

Dataset Network functioning Results

Dataset Network functioning Results

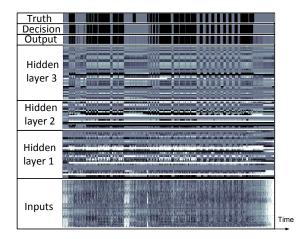
Jamendo : A Common Benchmark Dataset

- Publicly available dataset
- Singing voice activity annotations
- ▶ Training set : 61 files
- ► Validation and Test sets : 16 files each
- Common database  $\longrightarrow$  fair comparison of our approach

Recurrent Neural Networks and Long Short-Term Memory System Overview Dataset Network functioning Results

Results Conclusion

# **Internal Netwok Functioning**



Color scale between -1 (white) and 1 (black)

Recurrent Neural Networks and Long Short-Term Memory System Overview Dataset Network functioning Results

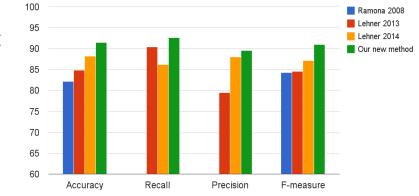
Results Conclusion

## **Consideration of a Temporal Context**

Truth		
Decision		
Output		
Hidden layer 3		
Hidden		
layer 2		
Hidden layer 1		
Inputs	nter of the second s	Time

Dataset Network functioning Results

**Results on Jamendo Dataset** 



# **Conclusion**

- New approach for singing voice detection
- ► We do not focus on defining a complex feature set → may be suboptimal
- We make use of neural networks to extract a simple representation, fitted to our task
- ► A past and future temporal context is considered by the classifier → no need for temporal smoothing
- The results we obtain encourage further work with BLSTM-RNN in MIR for sequence classification tasks, e.g. melody estimation

# Thank you