Neuro-Spectral Audio Synthesis: Exploiting characteristics of the Discrete Fourier Transform in the real-time simulation of musical instruments using parallel Neural Networks

Carlos Tarjano and Valdecy Pereira - Universidade Federal Fluminense, Niterói, RJ, Brazil

Contact Author: Carlos Tarjano (tesserato@hotmail.com)

ICANN19
28th International Conference on Artificial Neural Networks

PROBLEM AND OBJECTIVES

Most methods involving the use of neural networks to sound synthesis fail to take full advantage of structural characteristics of sounds, missing opportunities to increase the efficiency of the final models. Drawing inspiration from established traditional methods, the present work takes advantage of patterns present in the frequency domain representation of harmonic sounds, enabling the use of neural networks in the efficient and realistic emulation of acoustic musical instruments.

THE MODEL

1 – Raw Samples

Raw samples are collected, representative of characteristics of a particular instrument (or hybrid of instruments) that should be present in the model; volume and other instrument dynamics are obvious candidates, but other variables can be used, such as microphone distance. Those samples are ordered according to fundamental frequency and numbered, considering, for convenience, the range of a standard grand piano. A polynomial estimate (I) of the relationship between numbers and fundamental frequencies is calculated.

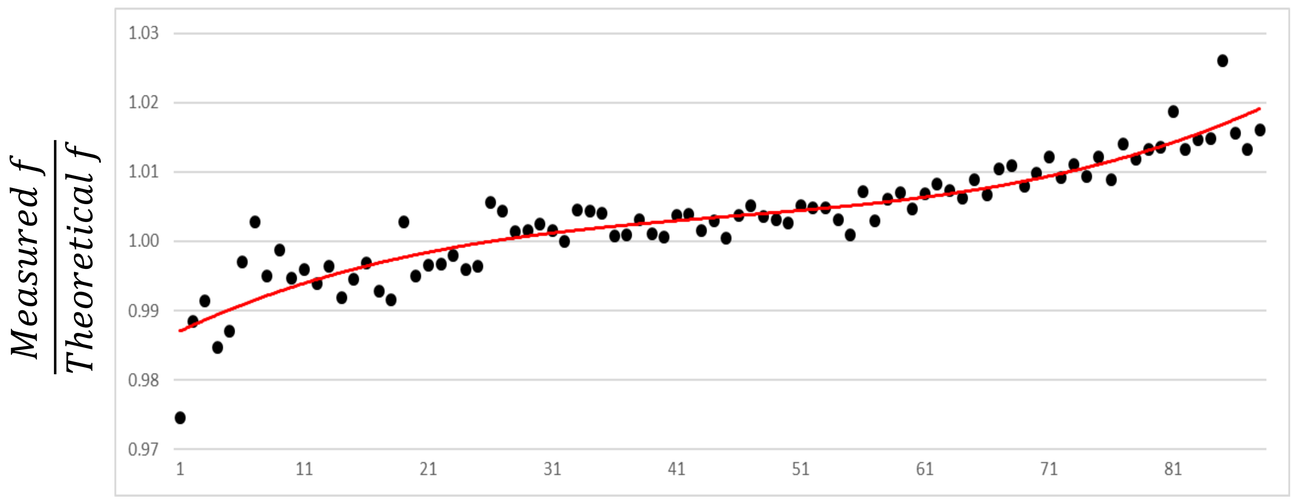

Fig. 1. Ratio between measured and theoretical fundamental frequencies as black dots and fitted polynomial as the red line. The effect of the tuning stretch can be observed

$$f_0[k,p] = 440.2^{\frac{k-49}{12}}(c_1k^3 + c_2k^2 + c_3k + c_4)(p+1)i[k,p] \quad \text{(I)}$$

3 – Training

The data is used to train a tree-like network: the common "trunk" allows for less redundancy in the process, while each parallel branch can be tailored to an optimal tradeoff between accuracy an efficiency.

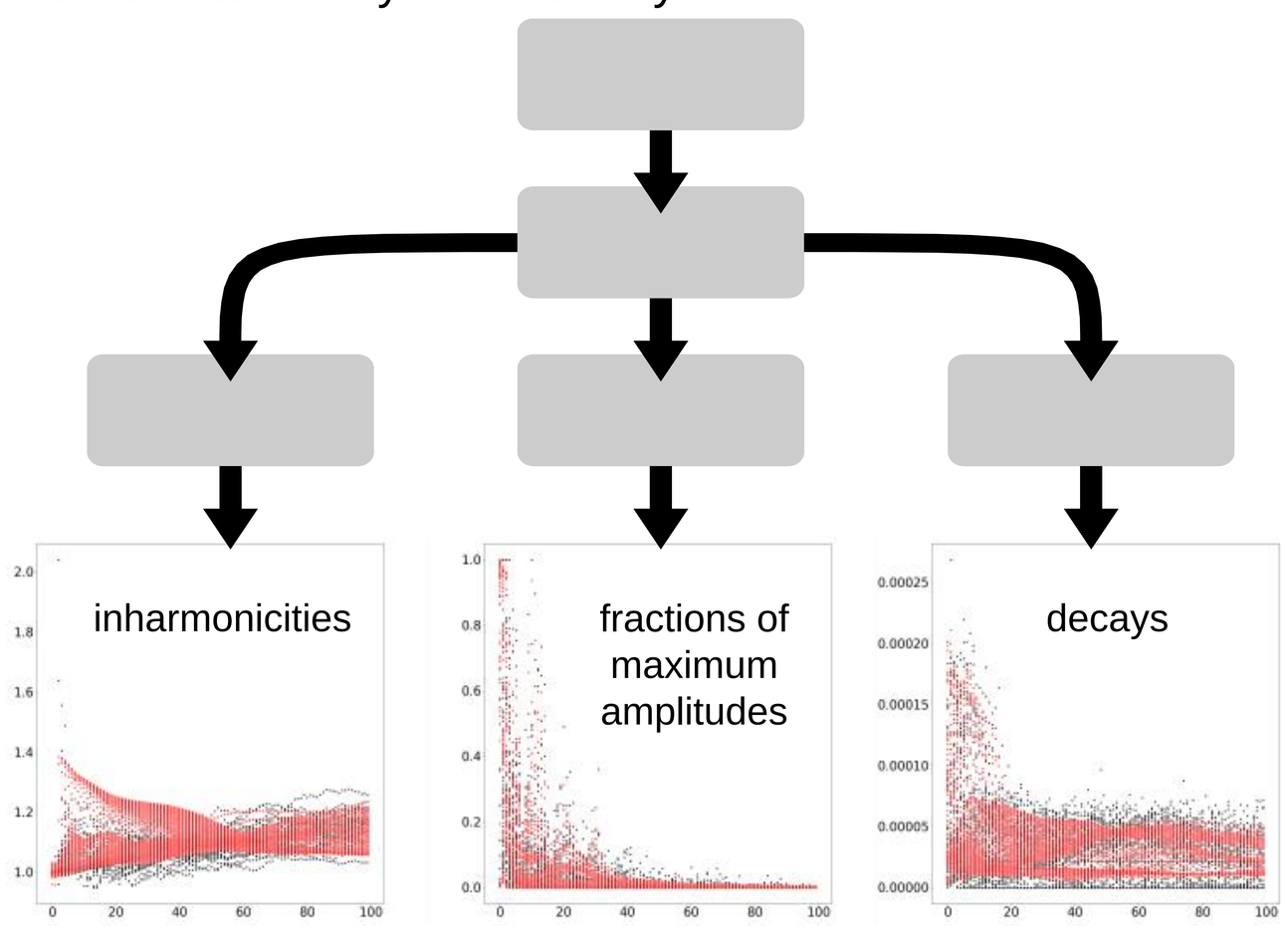

Fig. 2. Results of the model. Black dots are the original measured values and red dots are the values predicted by the network.

2 – Labelled Samples

For each of the now labelled samples, the behavior of each relevant partial is investigated in the frequency domain. Knowing that, in harmonic sounds, partials are close to integer multiples of the fundamental, the deviation from this ideal case, as well as amplitude proportion and exponential decay are calculated.

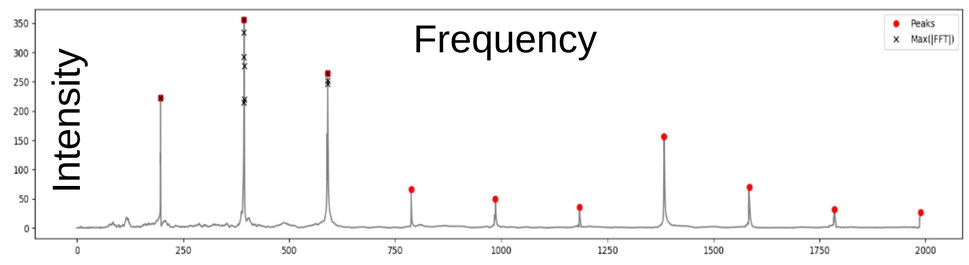

Fig. 2. Example of peak detection: peaks (red dot) x naïve maximum values (black “x”)

4 – Synthesis

Synthesis happens in real-time, in an additive way: the network receives as input a value related to the original samples and a number of additional values referring to specific dynamics and generates one sinusoid, with its respective decay, for each of the relevant partials. Any of the input values can represent interpolations or extrapolations: training the model with samples from a piano played in two different volumes will lead to a model capable of continuous control of volume dynamics, for example.

HEAR THE RESULTS
soundcloud.com/carlos-tarjano

Other contact Information:

+55 21 98232 4242

github.com/tesserato

ManuscriptID:330