



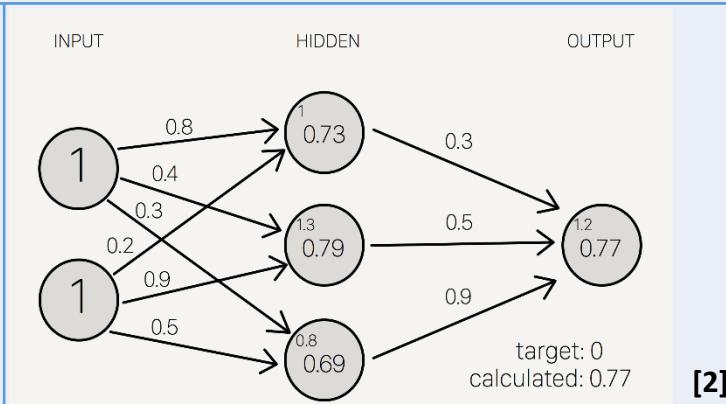
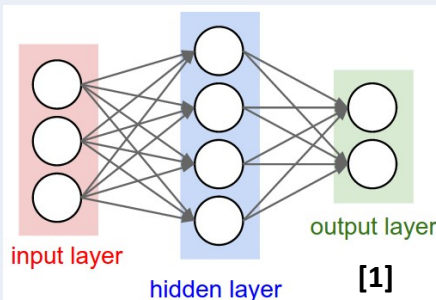
Realtime downbeat recognition using neural networks

Jack Hinton – p147676 – jhinton.147676@student.uwtsd.ac.uk

The aim of this project is to create a system that can keep time when music is played to it. This will then allow it to be combined with a drum machine or similar to allow musicians to practise with a drum track that keeps time with them, rather than the other way around. This presents unique challenges, as it is necessary for accurate realtime feedback. To this end, it has been decided that a neural network and accompanying feature extraction architecture is likely to be provided the needed rapidity of result. A neural network can be thought of as a 'net' of simple processes that feed their output into other processes in the net.

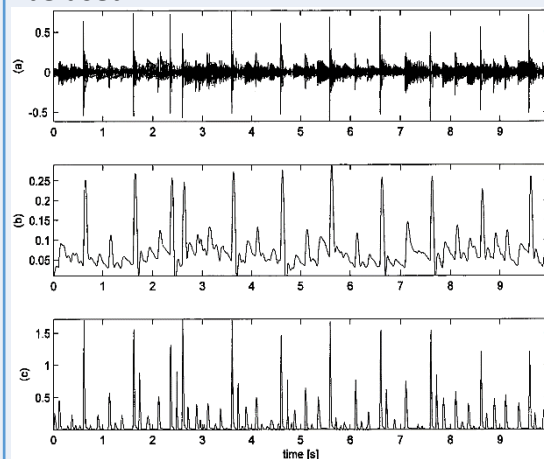
Single neurons typically have an imprecise binary or floating point outputs, but when many are networked together, the subtlety and precision of available output is greatly increased. Neural

To accomplish this, a system of weighting is used whereby neurons that give correct results are given progressively higher influence over the end result than others. In this manner, a neural network that is given many examples and the correct outcomes for these examples will eventually learn to give the correct output reliably.



In order to process a data stream, it is often useful to split it up into discrete units that may be individually addressed, rather than trying to continuously process the data on the fly. In music, the smallest quantum of time perceptible by the listener is known as a 'Tatum' [3]

To segment the audio stream into the smallest necessary components ("Tatums") in real time, many approaches may be used.



The most efficient method may be to use an approach that uses a Fourier transform to model the signal and thus segment it by comparing the resulting model against itself through time – when the model changes, a segment ends. [5]

After segmentation has been reliably achieved, the resulting segments must then be processed in order to extract the data pertaining to presence of a musical downbeat. There are many methods of doing this, the four most reliable of which will be used in conjunction. These methods are as follows: Analysis of Chroma change, detection of onsets, spectrum analysis, and melodic constant-Q transformation. [6]

The aforementioned data extraction methods will then feed their output into a neural network, which will have been trained prior using MIDI samples. If this project is successful, the neural network will then be able to keep time with music as it is played.

[1]"CS231n Convolutional Neural Networks for Visual Recognition", *Cs231n.github.io*, 2017. [Online]. Available: <http://cs231n.github.io/neural-networks-1/>. [Accessed: 04- Oct- 2017].
[2]"CS231n Convolutional Neural Networks for Visual Recognition", *Cs231n.github.io*, 2017. [Online]. Available: <http://cs231n.github.io/neural-networks-1/>. [Accessed: 04- Oct- 2017].
[3] Bilmes, J. (1993). Timing is of the essence. Massachusetts: Massachusetts Institute of Technology, p.21.

[4] J. Seppänen, A. Eronen and J. Hiipakka, "Method, apparatus and computer program product for providing rhythm information from an audio signal", US 7612275, 2009.
[5] J. Foote, "Automatic audio segmentation using a measure of audio novelty", 2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532), 2000.
[6] S. Durand, J. Bello, B. David and G. Richard, "Robust Downbeat Tracking Using an Ensemble of Convolutional Networks", IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 1, pp. 76-89, 2017.