# neural audio: music information retrieval using deep neural networks

Grey Patterson (Linfield, CCT), Andrew Pfalz, and Jesse Allison (LSU School of Music/CCT)

## introduction

Scan the QR code for the more verbose digital version of this poster.

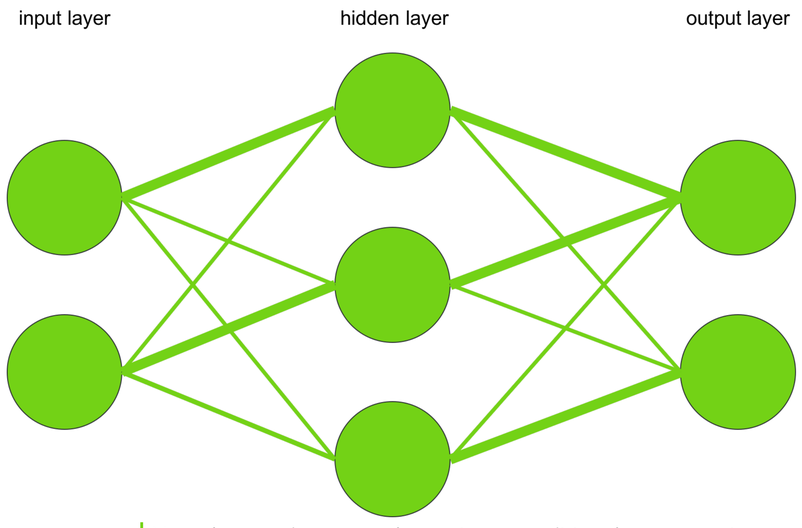
Neural networks are a form of machine learning that consist of nodes connected by weighted edges. Input goes in as floating-point numbers, and is propagated through the network from node to node along those edges.

Music is rich in information - from things like what key and time signature are being played in up to the sociocultural context in which the lyrics were written.

The field of music information retrieval exists to provide musicologists with automated tools. Complex tasks, though, remain out of reach. There is no algorithm to identify the genre of a piece, and no machine can identify which instrument is playing as easily as a human can.

## neural networks

**structure** | Neural networks are based on the structure of the human mind: nodes serve as neurons, and weighted lines between them alter the input values as they flow through the network.




**training** | Neural networks are an alternative to traditional programming. Instead of writing an algorithm yourself, you provide training data - inputs, and the output expected: `([0, 821, 1643,...],"art")`

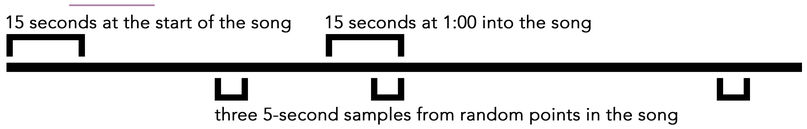
The software uses stochastic gradient descent to minimize the cross-entropy of the correct answers versus the calculated answers.

**output** | The output can be either a categorization ('2') or a softmax ('[0.2, 0.2, 0.6]'). We chose to use softmax throughout.

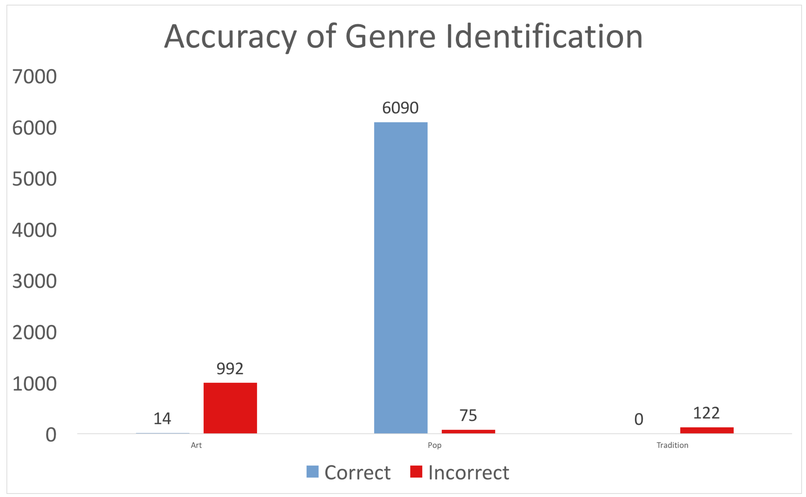
## genre identification

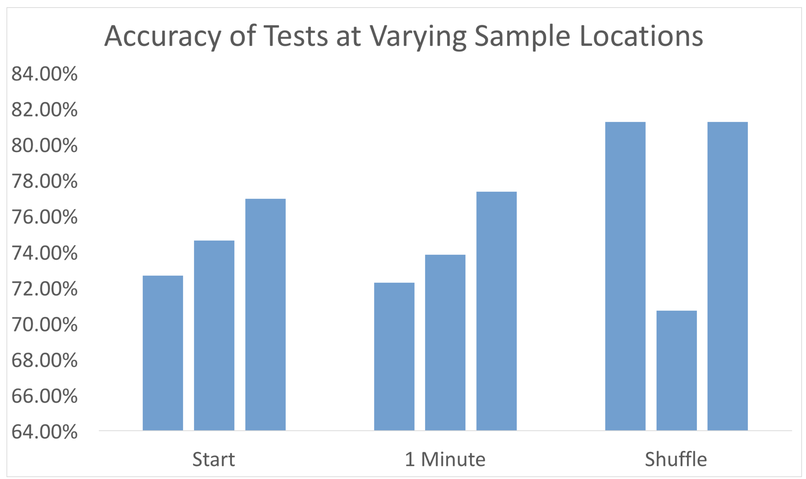
**methods** | Based on "Automatic Musical Pattern Feature Extraction Using Convolutional Neural Network" (Li, T., Chan, A., Chun, A..), we wanted to split music into at most four categories. We opted to use Philip Tagg's axiomatic triangle, giving us three categories: art, popular, or traditional.

Three windows were used:




**results** | Overfitting proves to be a significant problem, but adjusting for it shows that randomizing samples within the song increases accuracy.



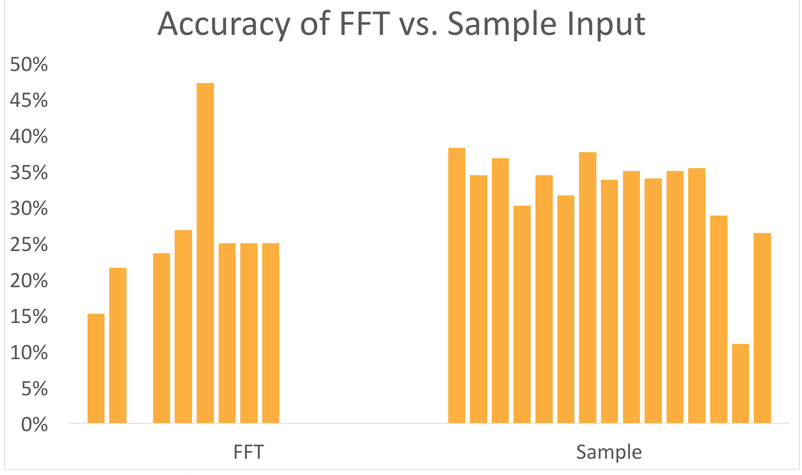

**conclusion** | Identifying which of the three categories a song falls into is a difficult task at times. Christmas music is a great example of how difficult this can be: Pentatonix' cover of *Carol of the Bells* is a three-year-old version of a song that was penned in 1914, which was itself based on an even older Ukranian folk song. Is this popular or traditional music?

## instrument identification

**methods** | We broke song stems and mixes from MedleyDB (R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam and J. P. Bello) into four categories: vocal solo, guitar solo, drum solo, and non-solo 'other.'

By breaking the songs into samples consisting of 1-3 seconds, with .5-1.5 seconds of overlap, we were able to process entire songs into an array of categorization results. With that, we generated an array of start/stop times and their categories.


**results** | We found that using raw samples, rather than FFT data, yielded higher overall accuracy (ignoring the outlier).




**conclusion** | Instrument identification remains an interesting problem, and one that we feel should be given further consideration. Increasing the size of the windows that the network was given yielded slight increases in accuracy, but we were unable to explore the extent to which this continues. Overall, we recommend further research with the aim of identifying which factors have the largest effect on accuracy.

## references


- Chollet, F. *Keras*, 2015, https://github.com/fchollet/keras/
- Li, T., Chan, A., Chun, A.. *"Automatic Musical Pattern Feature Extraction Using Convolutional Neural Network"*
- R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam and J. P. Bello, *"MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research"*, in 15th International Society for Music Information Retrieval Conference, Taipei, Taiwan, Oct. 2014.


- Work supported by the National Science Foundation, award OCI-1263236. Additional support from the Center for Computation & Technology at LSU.