DOI: 10.61782/fa.2023.1053

MACHINE LEARNING OF FINITE-DIFFERENCE TIME DOMAIN (FDTD) PHYSICAL MODELLING SOUND SIMULATIONS OF DRUMHEAD PASTE PATTERN DISTRIBUTIONS

Chrisoula Alexandraki[1*] **Michael Starakis**[1] **Rolf Bader**[2] **Panagiotis Zervas**[3]
[1] Department of Music Technology and Acoustics, Hellenic Mediterranean University, Greece
[2] Institute for Systematic Musicology, University of Hamburg, Germany
[3] Department of Electrical and Computer Engineering, University of Peloponnese, Greece

ABSTRACT

Drummers attach different kinds of material on their drumheads to either increase damping or to tune them and adjust the relationships of sound partials. The former is a common practice for drummers, while the latter may be found in percussion instruments of various ethnic traditions, such as the Myanmar *pat wain* drum circle or the Indian tabla. A Finite-Difference Time Domain (FDTD) physical model of a drumhead was used to compute more than 2000 sounds simulating membrane vibration, which was adjusted by adding varied amounts of paste, distributed in different surface patterns. These sounds were analysed using Self-Organizing Maps (SOMs) as well as a Convolutional Neural Network (CNN). The SOMs were used to cluster the partial relationships of the generated sounds. It is demonstrated that different paste patterns correspond to different clusters. Furthermore, the CNN was trained to identify the damping approach, yielding an accuracy of 94% for paste pattern classification and a mean error of +/-11% for the estimation of membrane mass increase. These tools can be used to identify damping patterns used in historical drum recordings or as suggestions to percussionists for deriving a desired sound texture.

Keywords: *drumhead damping, membrane tuning, machine learning, physical modelling*

[*] ***Corresponding author***: *chrisoula@hmu.gr*

1. INTRODUCTION

Circular membranes, produce unpitched sounds as their vibrational modes are not harmonically related. It is well known [1] that the eigenfrequencies of an ideal circular membrane are provided by Eqn. (1).

$$f_{mn} = \frac{1}{2\pi R}\sqrt{\frac{T}{\mu}}J_{mn} \tag{1}$$

fmn denotes the frequency of the mode corresponding to *m* nodal diameters and *n* radial nodes, R denotes the radius of the membrane, *T* the tension applied, μ its mass distribution and $J_{mn}(kr)$ denotes the n-th root of an m-order Bessel function. Therefore, the fundamental frequency is determined by the radius, the mass density and the Tension applied on the membrane, while the relationship of overtones has a constant non-harmonic relationship. Realistic drumheads exhibit an overtone structure that has many deviations to that of the ideal circular membrane. This is commonly attributed to the vibration of the shell, the air cavity within the shell, as well as the fact that mass and tension are not uniformly distributed throughout the surface. Different sounds may be produced by varying the radius R, the distribution of tension $T(x, y)$, the distribution of mass density $\mu(x, y)$, as well as by the hitting point that excites membrane vibration. Changing these attributes does not only affect the frequency of the fundamental, as shown on equation (1), but also the entire overtone spectrum in terms of frequency and amplitude relationships, as well as the decay rate of different partials.
Percussionists employ different strategies to adjust the sound texture of their drums. As shown on Fig. 1, there are several types of drum dampeners available in the market, such as damping pads covering the entire area of the membrane, drum clips, adhesive gels, gaffer tapes and

muffle rings to name a few. There are no standard instructions on how drums are damped, so the choice of which dampers to use and how to position them on the instrument is entirely up to the performer. The general idea is that a dampener will reduce the energy and duration of the modes that have antinodes at the membrane location of the dampener [2]. So circular dampeners aim at damping angular modes, small rectangular pads aim at damping diametric modes, while damping pads covering the entire membrane area provide volume reduction, without considerable changes of the sound texture.

Figure 1. Commercial drum dampeners including damping pads, drum clips, muffle rings and adhesive gels.

As an alternative to membrane damping, modifying the distribution of mass on the drumhead is often used for tuning an instrument, as is the case of the Myanmar pat wain instrument. *Pat wain* is a drum circle consisting of 20 or 21 pitched drums, which are tuned by applying a paste called pa sa. *Pa sa* is a mixture of rice and ashes. When cooked, the rice varieties of Southeast and East Asia become particularly sticky, the so-called sticky rice, which means that they have a high amount of viscoelasticity. To tune the pat wain drums, paste is applied at the center of the instrument. This decreases the frequency of the fundamental mode, as it applies additional mass without any adjustment to tension. Then, the tuner redistributes the amount of paste to control the overtone spectrum by checking timbre and comparing it with the other drums of the instrument [3].

The rest of this paper is structured as follows. The next section presents our efforts for assembling a sound dataset corresponding to the vibrations produced by a drumhead of predefined geometry, which is tuned by varying the amount of paste applied and the paste distribution pattern. Then, section 3 presents a methodology for investigating the effect of damping patterns on the spectral envelop of the corresponding sounds. A SOM is used to cluster sounds according to the relationships of spectral overtones. Section 4 presents a CNN which was trained with the dataset, to estimate the amount of added mass due to paste and to recognize the distribution pattern from a given sound signal. Finally, section 5 presents some concluding remarks of our work, as well as our ideas for further investigation.

2. SOUND DATASET

To calculate the sound of a membrane when damping or tuning material is applied, a FDTD algorithm was implemented as the numerical solution of the wave equation describing the vibration of the membrane. FDTD models have been previously used for complete geometries of a guitar, a violin, and several other instruments [4][5]. This section provides an overview of the FDTD model that was implemented to compute sound signals of a circular membrane, as well as the modification of the model to generate sounds when tuning paste is applied on the surface of the drumhead.

2.1 FDTD Model

The FDTD model used in this study provides a numerical solution of the wave equation of the circular membrane, provided by Eqn. (2).

$$\frac{T}{\mu(x,y)}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)=\frac{\partial^2 u}{\partial t^2}+D\,\frac{\partial u}{\partial t} \qquad (2)$$

The function $u(x, y, t)$ represents the instant displacement of a point (x, y) at time t, which is perpendicular to the surface of the membrane. T is the tension of the membrane, $\mu(x, y) = m(x, y)/A$ denotes the area density, i.e. mass over membrane surface A, and D is a damping constant. To derive a numerical solution for Eqn. (2), the membrane is modelled as a finite number of rectangular cell grids for which $\Delta x = \Delta y = h$, as shown in Fig. 2. The fundamental time step is defined as $\Delta t = k$. The membrane assumes boundary conditions $u(x_b, y_b, t) = 0$ for all points x_b, y_b around the circumference of the membrane and initial conditions $u(x, y, 0) = 0$ for all points except from the striking point of the membrane, for which $u(x_p, y_p, 0) = 1$.

The Newton–Störmer–Verlet, also known as leapfrog algorithm [6] is used to derive the displacement of every grid cell at every time step as described in [3]. Finally, a radiated sound signal is computed by assuming a microphone at distance *d* above the center of the membrane. This computation integrates the displacements at microphone position with a time delay and an attenuation determined by the virtual microphone position above the drumhead. Thereby the attenuation *d/r(x,y)* with *r(x,y)* the distances between the respective points on the membrane and the microphone, and a delay *r(x,y)/c* is used, with *c* denoting the speed of sound in the air.

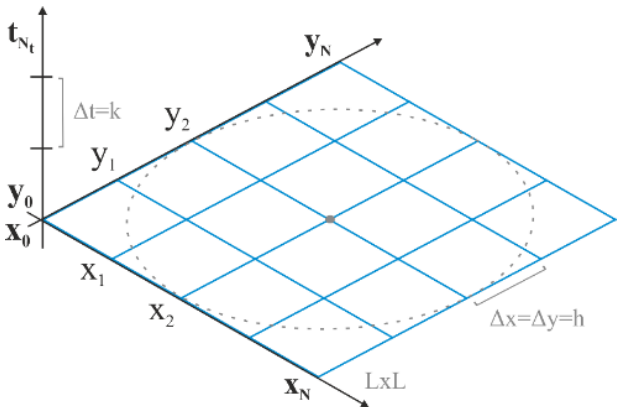

Figure 2. The cell grid of the circular membrane.

2.2 Membrane geometry and added paste

To generate a dataset of sound signals, the parameters concerning the geometry and the material of the membrane were kept constant. Paste was described in terms of the amount of paste mass applied on the membrane, as well as in terms of its distribution on the membrane surface, which followed seven patterns, inspired by common drum tuning and/or damping strategies. These paste patterns are illustrated in Fig. 3 and are indexed as: 1-diameter, 2-radius, 3-cross, 4-disc, 5-ring and 6-point. An extra pattern, i.e., 0-no_paste was included to account for the 'bare' membrane, for which different sounds were generated by varying the membrane thickness.

The 'bare' membrane was modeled having a thickness value of 3 mm, a radius of 0.25 m and a volume density of 300 kgr/m^3. These values describe a membrane having a total mass of 177 gr. The tension of the membrane was kept constant at T=800 Nt, the point of excitation was fixed on the center of the membrane, the virtual microphone position was set at distance d=0.7 m above the center, the damping factor was set at a value of D=0.9999 and the sound of speed was set as c=343 m/sec.

The produced dataset comprised 2331 sound samples, corresponding to approximately 333 sounds per pattern. For each pattern, different sounds were produced by adjusting the geometry of the area having paste, as well as the percentage of mass increase due to paste for the grid points for which paste was applied. Specifically, for radial patters (i.e., diameter, radius, cross) different sounds were computed by varying the line width and paste percentage. For the disc pattern, the radius of the disc and the paste percentage were adjusted, while for the ring pattern the outer radius, the width of the ring and the percentage of paste were varied. Finally, for the point pattern, representing an adhesive pad on the membrane, four grid points were chosen having different distances from the center of the membrane. For each of these the radial width and the paste percentage were varied to provide 332 combinations for the point pattern.

To generate the audio dataset, the FDTD model was implemented using the CUDA architecture on an Intel i7-4790k 4GHz/32GB RAM computer system using the NVIDIA GeForce GTX 970 4GB GPU on Windows 7. A uniform, square grid of 104x104=10816 nodal points was used to discretize the wave equation. The computation time was estimated around 3-4 seconds for the generation of 1sec of monophonic audio at a sampling rate of 96 kHz. The audio signals were resampled to the sampling rate 22050 Hz and stored in wav format. Ground truth annotations accounted for the label of the paste pattern as well as the amount of added mass in kg, as computed for the cells where paste was applied.

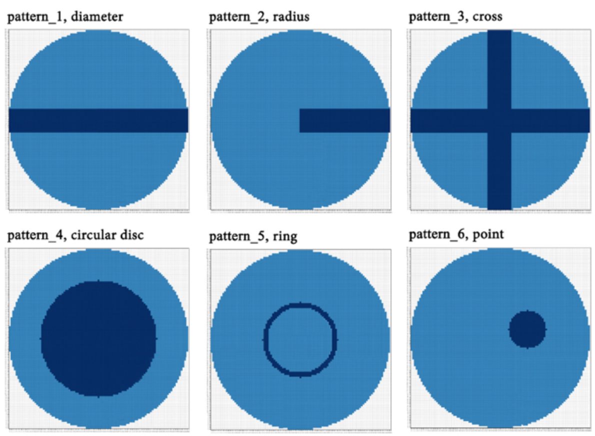

Figure 3. The paste distribution patterns simulated by the FDTD model.

3. DATA INVESTIGATION

Different paste patterns alter the sound spectrum with respect to the frequency relationships of the modal frequencies. As an example, Fig. 4 presents the FFT spectra for no added paste (pattern-0) and for paste distributed according to the cross pattern (pattern-3). By peak-picking the FFT spectra, it was found that for pattern-0 the fundamental frequency is 40Hz and the frequency ratios of the overtones to the fundamental are [1, 2.31, 3.62, 4.93, 6.24, 7.53, 8.84, 10.13]. For pattern-3 the estimated fundamental was at 30 Hz, and the frequency ratios were [1, 2.26, 4.06, 4.50, 5.53, 5.77, 6.67, 7.07].

The fundamental of 40 Hz for pattern-0 agrees with Eqn. (1). Moreover, the second and third partial have a frequency relationship to the fundamental that corresponds to mode four (0,2) and mode nine (0, 3). Note that due to the arbitrary choice of using the center of the membrane as the hitting point of the vibration, modes two, three and five up to eight have not been excited because they present nodal points on the center of the membrane. The fundamental of 30Hz for pattern-3 suggests a total mass increase by a factor of 2.25. This can be estimated by considering Eqn. (1) and the fundamental of the plain (no-paste) membrane. Moreover, the spectral envelopes of the two signals are significantly different, both in terms of frequency as well as in terms of amplitude relationships.

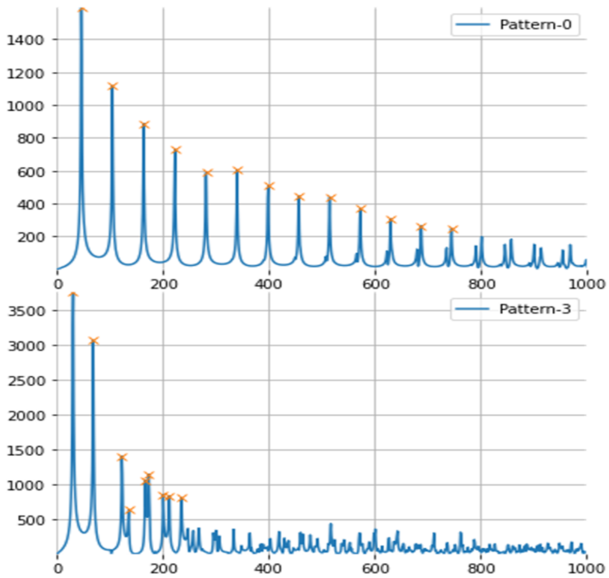

Figure 4. FFT spectra of two sounds corresponding to the bare membrane (no paste) and pattern 3 (cross) pattern.

3.1 SOM clusters

To investigate whether the relationships of spectral overtones and their amplitudes can be clustered in agreement with paste patterns, a Kohonen or self-organizing map (SOM) [7] was calculated from (a) the frequency ratios and (b) the amplitude ratios of partials for all sounds in the dataset.

A SOM reduces the multidimensional input data space into a two-dimensional map of size $N \times N$. Each neuron $i,j = 1,2,3,\ldots N$ on the map is a vector $S_{i,j} = \{s^{i,j}_1, s^{i,j}_2, s^{i,j}_3, \ldots s^{i,j}_M\}$, where M is also the length of the feature vectors $X_k = \{x^k_1, x^k_2, x^k_3, \ldots x^k_M\}$ used for training the map and $k=1,2,3,\ldots R$ with the R the number of training vectors. Initially, the $S_{i,j}$ are chosen randomly with normalized random vectors. Both neurons $S_{i,j}$ as well as the feature vector values X_k are normalized. The map is trained with T iterations, where each iteration is training of all feature vectors of the map. Each feature vector X_k is trained to the map by first calculating the distances $d_{i,j} = < S_{i,j} , X_k >$ between each neuron and the respective feature vector k, where $<,>$ denotes the scalar product. The neuron i,j with minimum distance $min(d_{i,j})$ is detected and this neuron, as well as its neighboring neurons, is changed to $S^{t+1}_{i,j} = S^t_{i,j} + \varepsilon\, h(S^t_{i,j} - X_k)$, where t is the training iteration number. Here, ε is the learning rate and h is a function determining the learning strength between the best-match neuron and its neighbors. In this study, a Mexican hat function was used.

The map is expected to unfold, i.e., converge, at iteration time t into regions of similar neurons, i.e., into clusters. These clusters often have boundaries to neighboring clusters, where neighboring map neurons change considerably. To estimate clusters, a u-matrix is calculated from distances between neurons, where the maximum distance is one and similarity is zero. Each neuron is assigned a u-matrix number according to its distance to neighboring neurons, again using the scalar product and the Mexican hat function. In the plots shown in this paper, the background color shows the u-matrix, in which dark colors represent similarity and bright colors strong dissimilarities. As can be seen in Fig. 5 and Fig. 6 below, larger regions of blue represent clusters and yellow or light green ridges show boundaries between the clusters. The trained map is then used for detecting the best matches of each input vector again calculating the distances $d_{i,j}$ as a scalar product and choosing the neuron with minimum distance as best match.

The Apollon and Computational Phonogram Archiving (COMSAR) frameworks [8-12] were used to train and analyze the SOMs. For each simulated drum sound, the first 16 partials were detected by pick-peaking the FFT spectra and the ratio of partial frequencies to the frequency of the fundamental were calculated and used as the feature vector for training. To account for the amplitude of different partials, the absolute amplitudes of the first 16 partials were used as training vectors. For each feature vector set, a SOM grid of 55 x 55 neurons was chosen and trained for 1000 training epochs. After training, all stimuli were best fitted into the trained SOM.

Fig. 5 shows the trained SOM and fitted sounds according to the frequency ratios of partials. Note that due to the overlap of fitted sounds on single neurons, some of the 2331 sounds are not represented as separate dots. The background is the u-matrix, displaying similarity between neighboring trained neuron weights. Dark blue regions represent similar regions, and light-yellow displays strong dissimilarity. The patterns are marked with colors; black is no paste, yellow, red, and magenta represent line patterns, i.e., diameter, radius and cross respectively, while cyan, green, and blue display circular patterns disc, ring and point. At first sight, the patterns cluster quite well, pointing to distinct partial relations within patterns. In the background neuron-similarity plot, a strong yellow ridge can be seen, distinguishing between the circular patterns ring (green) and disc (cyan) as well as the no-paste pattern (black), from the line patterns diameter(yellow), cross (magenta), radius (red) and the point pattern (black). The point pattern is on the side of line patterns mixed with the radius patterns, most likely because they are both asymmetric with respect to the membrane mid-point over all diameters. Besides these two groups, several subclusters appear, e.g., the cross (magenta) and the diameter pattern (yellow) subclusters at the left and the distribution of no-paste (black) pattern over several fields within disc (cyan) and ring (green) pattern clusters. Interestingly, the ring pattern is split into two fields, possibly related to the percentage of paste coverage of the membrane surface.

It is important to note the distribution of the no paste pattern. This is surprising at first, as the no-paste case is expected to have the identical frequency relationships. As the different sounds of pattern-0 correspond to increased thickness compared to the reference membrane, it appears that the decrease of f0, due to the increase of the total mass, introduces computational artifacts in peak-picking the higher order partials. As depicted by the top spectrum of Fig. 4, a second peak appears above the ninth overtone, which corresponds to a separate vibrational mode of a very close frequency. The peak-picking algorithm has a threshold related to the distance between successive peaks, therefore when identifying peaks, it appears that only one of these ‘double’ modes is selected. The reduced fundamental reduces the distances between successive modes therefore choosing either the first or the second peak as the next partial. Although these artifacts may be small, the SOM differentiates some of these sounds. Nevertheless, a closer look, reveals that there are fewer black dots compared to any other color, approximately 50-60 out of 335 no-paste sounds, suggesting that there is significant overlap.

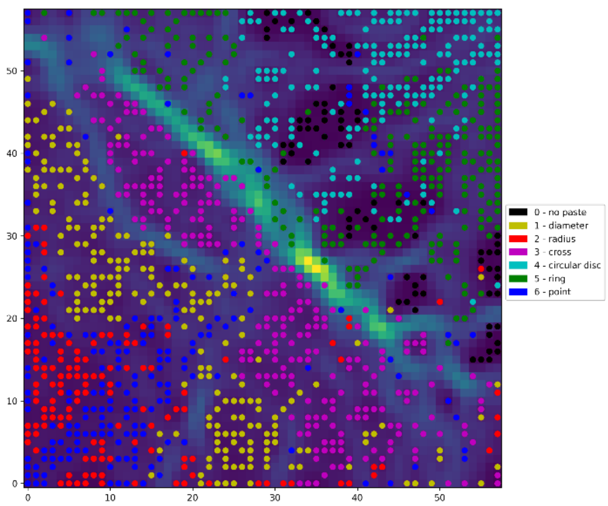

Figure 5. Trained SOM and fitted sounds according to the frequency ratios of their partials. Dark background colors reveal similarities, and light colors reveal strong neuron similarities.

Fig. 6 demonstrates the SOM fitted by training on the absolute amplitude values of the first sixteen partials. In this case, there is no clear ridge in the neuron similarity matrix of the background plot. However, the basic relationships are similar: line patterns diameter (yellow) and radius (red) are again close, asymmetric patterns point (blue) and radius (red) are combined, although to a lesser extent with a new blue cluster at the upper right corner. The circular disc pattern (cyan) and the no-paste pattern (black) are again combined. In this Figure, omitting paste in pattern-0 yields more instances of black dots compared to Figure 5, as the different thickness values results in different amplitude values but similar frequency relationships.

Yet a third SOM was calculated, combining the partial relationships with the amplitude values. This third SOM yields only slight differences compared to Fig. 5 and Fig. 6 and therefore has been omitted.

Overall, the SOM results show that using the patterns suggested in this paper leads to distinct drum sounds found in the distinct clusters. Asymmetric patterns are distinct from line patterns, again distinct from circular symmetric patterns.

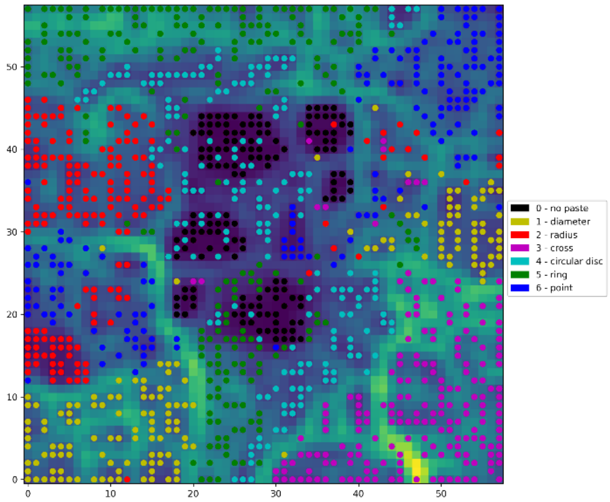

Figure 6. Trained SOM and fitted sounds for the amplitude values of the first sixteen partials. The clear separation between line and circular patterns is no longer present, although many clusters are similar to the clustering of frequency ratios.

The disc (cyan) pattern is close to the case of no-paste (black), therefore suggested for slight or imperceptible change of partial relationships. So, drummers only seeking to damp their membranes and not change the basic partial relations find their best choice in the circular disc pattern.

4. CNN INFERENCE

A deep neural network was implemented to identify the damping strategy for deriving a desired sound texture. Dataset sounds were provided as input to the network, which was trained to recognize the paste pattern as well as to estimate mass increase of the membrane owing to paste, thus accounting for a classification and a regression task respectively.

A multi-output network [13] was implemented to drive the training process towards making a combined inference for paste pattern and mass increase. The fact that each SOM cluster spans a considerable area, as well as the fact that clusters are not isolated from each other, reveal that pattern and mass increase have a combined effect on the resulting sound texture, which was the reason for opting for a multi-output network instead of separately training a classification and a regression task.

4.1 Network architecture

The architecture of the final multi-output CNN is shown on Fig. 7. It comprises multiple layers, including convolutional, pooling, dense, and flatten layers. The input layer accepts 1 sec of an audio signal at the sampling rate of 22050 Hz.

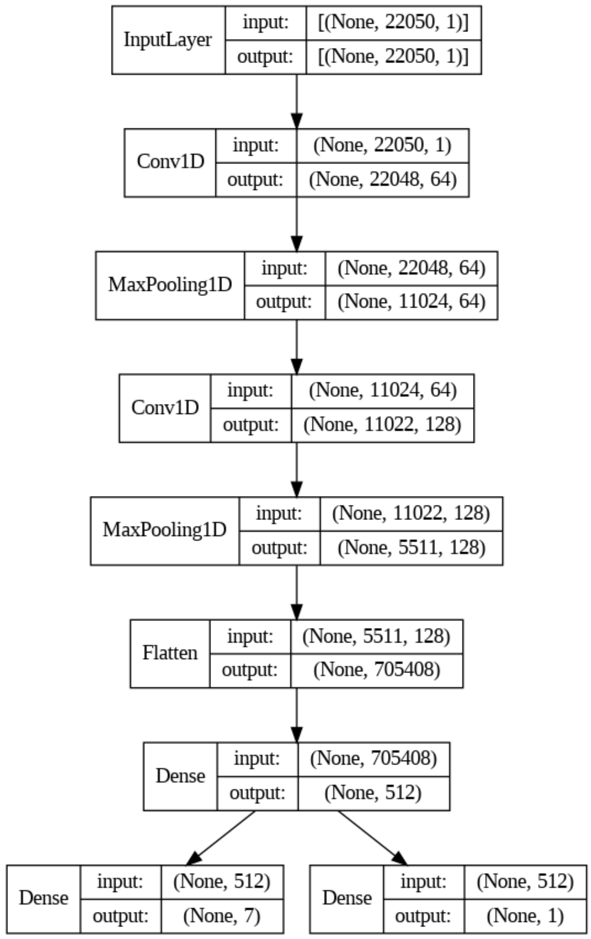

Figure 7. The architecture of the CNN used for inferring paste pattern and mass increase.

The output of the input layer is driven to a 1D Convolutional layer of 64 filters and a kernel size of 3, and a Max Pooling layer with a pool size of 2, followed by a second convolutional layer of 128 filters and a pooling layer having a pool size of 2. The output of the second pooling

layer is flattened and fetched to a dense layer of 512 units, which uses ReLU as the activation function. The network then splits in two separate outputs, one for classification and one for regression. The classification output is a Dense layer of 7 units and uses Softmax as the activation function and outputs the class corresponding to the paste pattern. The regression output is a Dense layer with one regression unit and a linear activation function, which outputs the predicted amount of paste.
The model used the Adaptive Moment Estimation (ADAM) algorithm for optimization, and the cost functions were based on Categorical Cross-Entropy (CCE) loss for multiclass classification, Mean Squared Error (MSE) for regression.

4.2 Network training

To infer drumhead tuning the dataset was split into train and test sets in an analogy of 67% (1561 samples) to 33% (770 samples) respectively. The CNN model was implemented in Python using TensorFlow and Keras on the Google Colaboratory environment, which made use of a Tesla T4 GPU.
Training used a batch size of 20 samples and was stopped after 220 training iterations (epochs). To assess the performance of the model and prevent overfitting, 5-fold cross-validation was employed.

4.3 Results

The classification accuracy reached a performance of 94.03% for the test set, as detailed by the confusion matrix of Fig. 8.

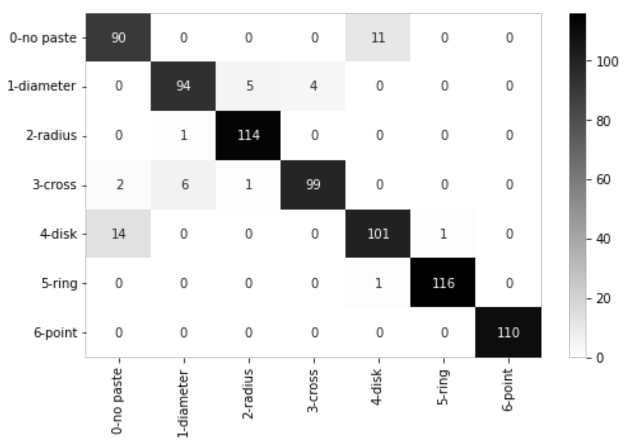

Figure 8: The confusion matrix of paste pattern classification in the test set.

The performance of the regression task was evaluated using the values of Mean Absolute Error (MAE) and Mean Squared Error (MSE). Table 1 provides the values of these measures per paste pattern. It appears that the added mass was estimated with an improved precision for the patterns Diameter, Disk and Point, while the weighted average of 0.01889 for MAE demonstrates an overall error of +/- 10.67% compared to the total mass of the membrane (177gr).

The high performance of the classification task may suggest that the overtone spectrum, i.e., frequency and amplitude relationships is dominated by the way mass is distributed on the surface, rather than the amount of mass used, which predominantly affects the energy and the frequency of the fundamental mode of vibration. This is in agreement with relevant studies on musical acoustics ([2] and [3]), as well as with the claims of professional percussionists. To confirm this assumption, further computational experiments are being performed. These involve the generation of more sounds, as well as the comparison of the performance of different neural network architectures to account for the variability of information encoded by deep learning architectures [14].

Table 1. The performance of the regression task per paste pattern in the test set.

Pattern	#	MAE	MSE
0-No Paste	101	0.02627	0.00162
1-Diameter	103	0.00754	0.00010
2-Radius	115	0.00406	0.00003
3-Cross	108	0.01792	0.00062
4-Disk	116	0.04699	0.00601
5-Ring	117	0.02363	0.00124
6-Point	110	0.00481	0.00004
Sum or W. A.	**770**	**0.01889**	**0.01417**

5. CONCLUSIONS

This work attempts to provide insight into drumhead tuning and damping techniques often used by percussionists by repeatedly altering tension or mass distribution on their drumheads till they achieve a satisfactory timbre. The methodology includes the computation of a sound dataset of more than 2000 sounds of a physical model of a membrane having pre-defined geometric characteristics and generating different sounds by altering mass distribution according to seven patterns. The variation of mass distribution has been inspired by the Myanmar *pat wain* instrument, which is tuned by adding paste on the surface of the drumhead.

Self-Organizing Maps were used as a visualization tool to reveal two dimensional similarities of the relationships of the first sixteen spectral overtones. It is shown that each pattern corresponds to a different cluster of frequency relationships, while different patterns lead to highly dissimilar patterns. As revealed by the SOMs, the disk pattern leads to similar vibrations of the membrane when no damping is applied, suggesting that it does not introduce significant changes in the spectral envelop. In contrast, asymmetric patterns are distinct from line patterns, again distinct from circular symmetric patterns.

As a further step, a deep neural network was trained to infer the damping approach when provided with an input sound. The network produces two outputs suggesting a paste pattern and an amount of extra mass to apply on the drumhead to derive the given sound. The paste pattern inference task has a very high accuracy of 94% again confirming that a specific paste pattern results in highly correlated sounds.

A limitation of this study concerns the fact that the sounds used for SOM clustering and CNN inference have been derived by purely computational methods. Future investigations will focus on applying this methodology on realistic sounds that will include a mixture of recorded drumhead sounds and computed sounds derived by physical models and data augmentation techniques. Moreover, a perceptual evaluation of the similarity of a desired sound texture to the sound texture provided by the physical model suggested by the CNN, while further enhance the validity of the target application, namely suggesting damping techniques to percussionists.

6. REFERENCES

[1] M. Schroeder, T. D. Rossing, F. Dunn, W. M. Hartmann, D. M. Campbell, and N. H. Fletcher: *Springer Handbook of Acoustics* (1st. ed.). Springer Publishing Company Inc, 2007.

[2] R. Worland and W. Miyahira, “Physics of musical drumhead damping using externally applied products,” in *Proc. of Meetings on Acoustics* 35, 035004, 2018.

[3] R. Bader, “Finite-Difference model of Mode shape changes of the Myanmar Pat WAIN drum circle using tuning paste,” in *Proc. of Meetings on Acoustics*, vol. 29, 2016.

[4] S. Bilbao: *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics,* John Wiley and Sons, 2009.

[5] R. Bader: *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology*, Springer-Verlag, Berlin, Heidelberg, Current Research in Systematic Musicology, 2013.

[6] E. Hairer, C. Lubich and G. Wanner: “Geometric numerical integration illustrated by the Störmer–Verlet method,” in *Acta Numerica*, vol. 12, pp. 399–450, 2003.

[7] Kohonen, T.: *Self-organizing maps.* 3rd edition, Berlin: Springer, 2001.

[8] R. Bader (ed.): *Computational Phonogram Archiving*. Springer Series Current Research in Systematic Musicology Vol. 5, 2019.

[9] M. Blaß, and R. Bader, R: *Content Based Music Retrieval and Visualization System for Ethnomusicological Music Archives.* In: R. Bader (ed.): Computational Phonogram Archiving, Springer Series ‘Current Research in Systematic Musicology’, Vol. 5, 145-174, 2019.

[10] N. Plath and R. Bader: “Piano Timbre Development Analysis using Machine Learning” arXiv:submit/4061895 [q-bio.NC] 6 Dec 2021

[11] R. Bader, M. Blaß, M. and J. Franke, J.: “Computational timbre and tonal system similarity analysis of the music of Northern Myanmar-based Kachin compared to Xinjiang-based Uyghur ethnic groups” Preprint: arXiv:submit/3650053 [q-bio.NC] 15 Mar 2021.

[12] M. Blaß, J. Fischer, and N. Plath, N: “Computational Phonogram Archiving: a generic framework for knowledge discovery in music archives,” *Physics Today* 73, 12, 2020.

[13] D. Xu, Y. Shi, I. Tsang, Y. Ong, C. Gong and X. Shen: “Survey on Multi-Output Learning,” *IEEE Transactions on Neural Networks and Learning Systems*. 31, pp. 2409–2429, 2020.

[14] L. ten Bosch and L. Boves, “Information encoding by Deep Neural Networks: What can we learn?,” in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH.*, pp. 1457–1461, 2018.