

Soft Tissue Characterisation Using a Novel Robotic Medical Percussion Device with Acoustic Analysis and Neural Networks

Pilar Zhang Qiu¹, Yongxuan Tan¹, Oliver Thompson¹, Bennet Cobley and Thrishantha Nanayakkara

Abstract—Medical percussion is a common manual examination procedure used by physicians to determine the state of underlying tissues from their acoustic responses. Although it has been used for centuries, there is a limited quantitative understanding of its dynamics, leading to subjectivity and a lack of detailed standardisation. This paper presents a novel compliant two-degree-of-freedom robotic device inspired by the human percussion action, and validates its performance in two tissue characterisation experiments. In Experiment 1, spectro-temporal analysis using 1-D Continuous Wavelet Transform (CWT) proved the potential of the device to identify hard nodules, mimicking lipomas, embedded in silicone phantoms representing a patient's abdominal region. In Experiment 2, Gaussian Mixture Modelling (GMM) and Neural Network (NN) predictive models were implemented to classify composite phantom tissues of varying density and thickness. The proposed device and methods showed up to 97.5% accuracy in the classification of phantoms, proving the potential of robotic solutions to standardise and improve the accuracy of percussion diagnostic procedures.

I. INTRODUCTION

A. About Percussion

Manual percussion (Fig. 1.a) is a diagnostic procedure used in clinical examination whereby the chest, back and abdomen are tapped to determine the condition of underlying tissue through the acoustic response [1]. The perceived auditory feedback are known as percussion notes [2]. Subjectively listening to their amplitude and pitch (i.e. frequency) changes can help medical professionals in the identification of certain pathologies. For instance, cases of emphysema and pneumothorax cause hyperresonance during chest percussion; large areas creating dull notes in the abdomen suggest an enlarged organ or a mass [2].

Due to the subjective and skill-dependent nature of the assessment, misdiagnosis may occur even in skilled professionals and may lead to disagreements between clinicians [3]. Robotic devices can help quantify the mechanical percussion action and the analysis of the acoustic response, minimising misdiagnosis and potentially automating remote patient examinations.

* This work was supported by the EPSRC Robopatient Grant No EP/T00603X/1.

** For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising

¹ Authors contributed equally to this work.

All authors were part of Imperial College London, Dyson School of Design Engineering, London, SW7 1AL, U.K.

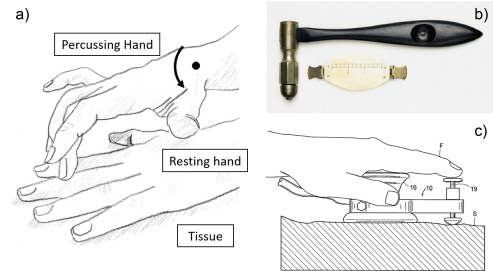


Fig. 1. Percussion methods and devices. (a) Manual Percussion [4]. (b) Percussor and Pleximeter [5]. (c) Stethoscope-attached device [6].

B. Studies on Medical Percussion

For complex soft tissue dynamics, percussion can be viewed as an impulse response analysis involving a process to associate percussion motor action with complex acoustic features. Whilst percussion is common in medical practice, there is limited knowledge on what auditory attributes determine a diagnosis. Quantitative understanding of this acoustic response of percussion will provide new insights into this procedure, allowing for more accurate diagnosis and potentially enabling the automation of percussion examinations.

Relevant models describing the transmission properties of chest percussion emerged in the mid 1900s. Rice's model theorised that the acoustic response during chest percussion was partially damped by the thoracic contents [7]. McKusick expanded on this theory through the use of sound spectrography [8]. Recently, Mansy [9] analysed the acoustic response of abdominal percussion by comparing the signal waveform for varying pneumoperitoneum conditions. The results showed significant differences in the frequency domain between moderate and severe conditions, suggesting the validity of auditory analysis for percussion diagnosis.

C. Automation and the Future of Percussion

Throughout centuries, manual devices have been used to augment percussion methods. In the 18th century, percussor and pleximeter tools were documented (Fig. 1.b) and a stethoscope-attached handheld percussion device was patented in 2004 (Fig. 1.c). However, variances in the posture or use of these tools may still lead to inconsistencies in the percussion actions. Thus, as of late, interest has been gathered around the automation and standardisation of percussion, both in physical action and diagnosis. [10], [11]

To develop an optimal percussion robot, it is crucial to identify what mechanical impact system generates the optimal acoustic response for tissue characterisation. In the

work of Krumpholz et. al [11], three percussion mechanisms were tested on a healthy human. The results showed that human participants were able to discern the nature of underlying regions more accurately when using a motorized device, versus a magnetized or a pneumatic approach. This shows that impact dynamics for tissue identification are best achieved using a motorized system with a rotary mechanism, instead of direct impact, linear mechanisms. This might be due to rotary mechanisms allowing a probe to bounce off to preserve natural restitution dynamics of the tissue.

Experimental results on percussion devices [12] also show that damping helps to increase the signal to noise ratio. Therefore, a certain level of damping at the pivot joint can reduce noise injected to the tissue during percussion.

Aside from the actuation systems, different sensors have been used to capture percussion responses. For instance, Peng et al. [13] modelled sound transmission by measuring the mechanical skin vibrations at the posterior chest. More compact applications have also been developed using microphones embedded in single actuator and measuring units [10], [14]. Ayodele et al. [10] used a one-degree-of-freedom push-pull solenoid plessor and analysed the acoustic response captured by a microphone. Their classifier reported at least 90% accuracy in classifying the percussion note into "Dull", "Resonant" and "Tympanic" responses, proving the practicality of using wavelet transform and neural networks for classification of acoustic percussion responses.

D. Key contributions in this paper

This paper presents a compliant, two degree-of-freedom robotic percussion system, featuring a motorized rotary system and a damping mechanism at its wrist joint. The device is designed to replicate a clinician's natural percussion movement while minimising the variability between consecutive percussion force profiles and simplifying the task of tissue characterisation. The use of Gaussian Mixture Models and Neural Networks allowed for a better abstraction of the audio features needed to understand tissue composition.

While similar projects, such as the one by Krumpholz [11], presented a unique remote-physician approach, the proposed solution explores the potential for fully-automated diagnosis methods, which could provide more regular home assessments to patients with low mobility (e.g. elderly) in saturated healthcare systems.

II. ROBOTIC DEVICE DEVELOPMENT

A. Mechanics of the robotic percussion device

We assume the hand and forearm act as rigid links and the force at the finger tips originates from the elbow joint and is damped by a compliant wrist joint. The forearm and hand movement of a percussion event can be summarised as follows:

- 1) Elbow joint initiates the rotation of the forearm and the hand with a compliant wrist joint.
- 2) When the hand is close to the target region, the elbow joint rotation stops, and the hand continues to rotate around the wrist joint.

- 3) The finger tip impacts the target surface and rebounds.

To mimic this action, a TowerPro MG996R servo motor acts as the driving unit initiating the elbow rotation. Extension springs are placed around the elbow joint for generating the initial torque to actuate the system, and at the wrist joint for compliance control (Fig. 2.A-B).

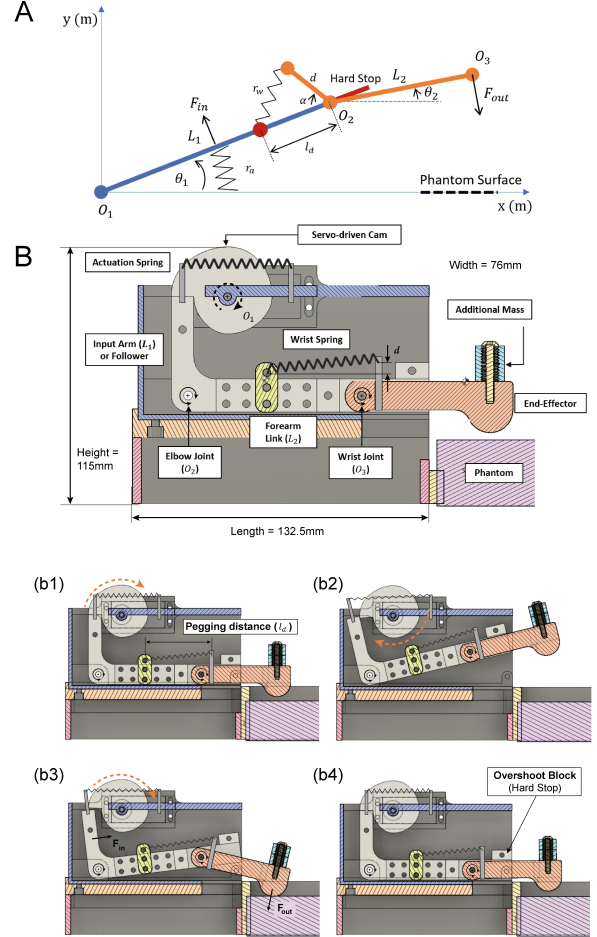


Fig. 2. (A) Percussion arm model. Blue represents the forearm and orange represents the hand. The actuation force F_{in} is stored in the extension spring with free length r_a . (B) Cross-section view of the CAD model. (b1) From the initial resting state, the servo-driven snail cam spins clockwise, controlling the actuation of the system and extending the springs. After reaching the peak state (b2), the input arm or follower "falls" and the wrist spring causes the percussion motion in which the end-effector hits the phantom once (b3), before the overshoot block stops any wrist joint oscillatory motions from re-hitting the phantom (b4).

The torque of the system at launch is:

$$\tau_S = \begin{bmatrix} \tau_e \\ \tau_w \end{bmatrix} = \begin{bmatrix} F_{in} L_{in} \\ F_w \frac{d \sin(a + \theta_1 - \theta_2)}{\tau_w} l_d \end{bmatrix} \quad (1)$$

where τ_e is the initial torque applied at the elbow joint, and τ_w is the resultant torque around the wrist joint.

In the initial torque (τ_e) formula, L_{in} is the input rod length, and F_{in} is the input force generated by extending the elbow or actuation spring, which is defined as:

$$F_{in} = k_a x_a \quad (2)$$

where k_a is the actuation spring constant and x_a is the displacement of the tension spring (i.e. drop distance of the cam).

From the resultant wrist torque (τ_w) formula, the spring pulling force at the wrist (F_w) is defined as:

$$F_w = k_w \delta r_w \quad (3)$$

where k_w is the wrist spring constant and δr_w is the displacement of the wrist tension spring from its free length, given a specific pegging distance l_d .

To produce compliant wrist action, a control rod of length d is attached at angle α to the hand rod. An extension spring with free length r_w connects the control rod to the forearm rod, setting a pegging distance l_d . This distance can be adjusted by pulling the peg away from the wrist joint, resulting in a new pegging distance $l_d + x$ and spring length $r_w + \delta r_w$.

Once the elbow joint extension spring is at the maximum length, the forearm movement is released and the momentum is transferred to the hand link and end-effector, which then makes impact with the tissue. An overshoot block mechanism also ensures that the device only impacts the target once by stopping any large wrist joint oscillations (Fig. 2.b4).

Further visual and verbal explanations of the inner workings of the percussion device are presented in Figure 2.b1-b4.

B. Estimation of Component Characteristics

The appropriate impact torque for the robotic device was estimated from Rao's study [15], where the wrist torque variation in thumb-to-finger percussion was found to be 8.3–12.6 Nm. A model of the percussion robot was developed from the prior mathematical formulation and manufactured using Fused Deposition Modeling (FDM) with PLA plastic.

With this information, the elbow and wrist spring rate were calculated to be 0.41 N/mm, which results in a wrist torque of 8.6 Nm, a value that aligns with the existing literature. These parameters were calculated based on a forearm link length (L_1) of 77.5mm, a hand link length (L_2) of 55.5mm and a pegging distance (l_d) of 42.5mm. The resulting pegging distance was adjusted to accommodate the springs used, with a free length of 34.5mm and maximum extended length of 59.5mm. An additional mass of 25g was added at the end-effector, as the PLA's low density was insufficient to output the required force range.

III. VALIDATION EXPERIMENTS

We conducted two experiments to quantify the potential of the device in medical percussion procedures:

- 1) Experiment 1 was conducted to test its ability to detect hard nodules embedded in silicone phantoms mimicking tissue anomalies through robotic percussion and acoustic analysis.
- 2) Experiment 2 was conducted to test its ability to identify the composition of underlying tissues in non-homogeneous silicone phantoms.

A. Experiment 1: Hard Nodule Depth Detection

A preliminary study was conducted to validate and determine the device's ability to detect the presence of an anomaly (e.g. hard nodule) embedded at different depths in silicone phantoms mimicking the human abdominal tissue.

1) **Method:** To represent an early-stage lipoma or cancerous lump, a hard acrylic nodule (diameter: 11 mm, Young's modulus: 3.2 GPa) was buried into a silicone phantom in a way that the phantom could be placed either way up, giving nodule depths of 2 mm and 15 mm (Fig. 3). The silicone phantom simulating abdominal tissue was made using a standard Ecoflex 00-10 mixture of 1:1 (Part A : B) [16].

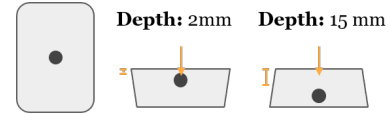


Fig. 3. Nodule depths of 2 mm ('close') and 15 mm ('far').

A TIMESETL piezo-electric contact microphone was attached to the surface of the silicone phantom. It was positioned at a distance of approximately 15 mm from the impact point, with the center of the impact point being on top of the hard nodule location (Fig. 4). The audio signal was transmitted with a Focusrite Scarlett 4i4 3rd Gen USB Audio Interface. To preserve the human hearing range of 20 Hz to 20 kHz and following Nyquist–Shannon sampling theorem, we recorded the signal at 44.1 kHz to avoid aliasing.

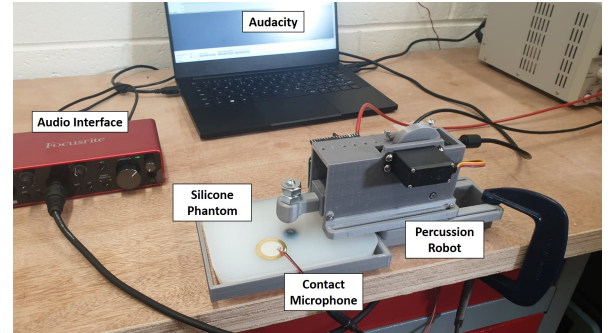


Fig. 4. Robotic percussion device experimental set-up

2) **Spectro-Temporal Analysis Using Wavelet Transform:** Signal pre-processing was performed to prepare and extract individual percussion instances from the raw audio recordings. Individual percussion events were extracted by locating the timestamp of the primary response or amplitude peak corresponding to each individual robot percussion. The time window for cropping the data was chosen to be 150 ms in range, with the individual events starting 20 ms before the maxima and ending 130 ms after the maxima.

In Experiment 1, a total of 1,688 percussion events were extracted from 20 minutes of testing.

The cropped and aligned percussion events were analysed through 1-D Continuous Wavelet Transform (CWT) coefficient scalograms. This was preferred over Fourier transform

mations as the time domain contained crucial information on time-localized spectral characteristics of the signal. We chose Symlet (Sym) mother wavelet family due to its wide use in audio analysis and signal damping [17].

The continuous wavelet transform $W(a, b)$ of the original signal $f(t)$ with respect to the Sym10 wavelet $\Psi(t)$ is:

$$W(a, b) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{|a|}} \Psi \left(\frac{t-b}{a} \right) dt \quad (4)$$

where a is the scale factor (determining resolution) and b is the time-localisation. The symbol $*$ denotes complex conjugation.

The CWT scalograms were split into different frequency bands and the average C coefficient values were computed for each band. This allowed to identify the frequency bands of interest when discerning the presence of close and far nodules.

3) Experimental results: Statistical analysis showed the strength of correlations between changes in the spectro-temporal domains and the two nodule depth groups. The results showed that frequency bands ranging from 175 – 400 Hz were the most representative (Table I).

TABLE I

T-TEST RESULTS FOR THE MOST RELEVANT FREQUENCY BANDS.

Frequency Band (Hz)	P - Value
0-175	0.02015 *
175-400	2.9391E-176 ***
400 - 700	0.3201
700 - 2000	1.38158E-04 ***
2000 - 15500	0.6909
15500 - 21000	3.76940E-08 ***

One asterisk (*) indicates a p-value smaller than 0.05 ($p < 0.05$). Three asterisks (***) indicate a p-value smaller than 0.001 ($p < 0.001$). The most statistically significant band (175-400 Hz) is highlighted in light gray.

For the purposes of visualisation and readability, the CWT scalograms in Figures 5 - 6 were plotted for a smaller frequency range containing the most significant frequency band (i.e. 175 – 400 Hz) instead of the full 0 – 22 kHz (human hearing) or 0 – 44.1 kHz (captured data) ranges.

In order to compare the acoustic response from close and far nodule instances, an average of the aligned signals from each group was generated.

Figure 6 shows a damping effect of the acoustic signal present in the close nodule recordings (depth = 2 mm), represented by faster-dying signal intensities following the percussion events, as opposed to the longer-lasting oscillations in both high and low frequency bands of the far nodules (depth = 15 mm).

A clearer visualization of the relevance of the frequencies from 175–400 Hz can be seen in Figure 7, where the largest difference in the magnitude between close and far nodule samples is shown to be between frequency band 9 (200–225 Hz) to band 16 (375–400 Hz). These frequencies might be of interest in the future when studying the presence of hard anomalies in abdominal-like tissues.

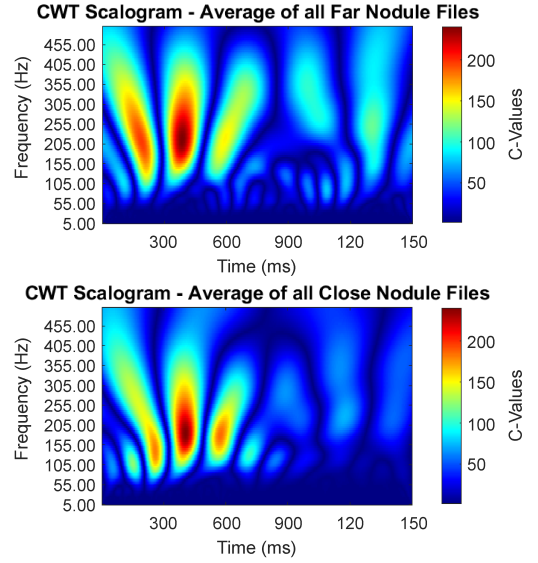


Fig. 5. Coefficient Scalograms for the average signal in close nodule (top) and far nodule (bottom) percussion instances. Graphs fitted to regions of interest (e.g. 5-500 Hz bands) rather than the full frequency scale.

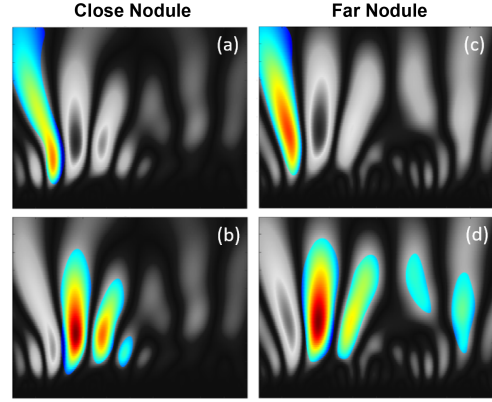


Fig. 6. (a) and (b) correspond to the close nodule (depth = 2 mm) scalograms, while (c) and (d) correspond to the far nodule data (depth = 15 mm). (a) and (c) show a common high-intensity area during early time frames. This represents the immediate time after device impact. The subsequent intensity areas highlighted in colour in (b) and (d) represent the acoustic response in the silicone phantom. These resonant waves diminish faster in the close nodule samples (b).

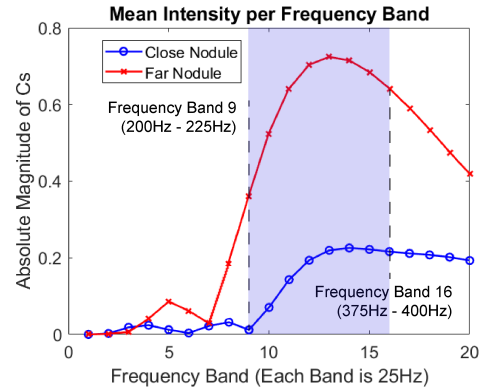


Fig. 7. Absolute, mean C-values for frequency bands in close and far nodule depth samples. Highlighted is the most statistically significant frequency bands region. In this graph, each frequency band has a range of 25Hz.

These preliminary results prove that the proposed system is able to discern different depths of embedded hard nodules with extremely high accuracy.

B. Experiment 2: Complex Tissue Characterisation and Classification

While Experiment 1 explored the potential of the system to identify the depth of hard nodules or anomalies, Experiment 2 aims to perform more complex tissue characterisation and classification using Neural Networks (NN).

1) **Method:** A total of 15 silicone abdomen phantoms were designed to showcase the potential of the system to characterise tissues with varying organ compositions (i.e. representing swollen, normal and hard organs) and distance from the surface (i.e. organ size).

Each phantom ($170 \times 115 \times 50$ mm) has two regions — an homogeneous "flesh" region occupying half the volume of the phantom, and a composite "flesh-organ" region occupying the other half, as shown in Figure 8.A.

In the "flesh-organ" region, 3 different silicone mixture ratios were used to recreate the underlying "organ" of the "flesh-organ" regions, represented by the colours red, blue and beige in Figure 8.A-B.

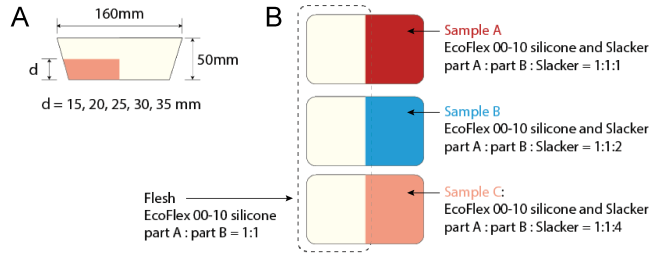


Fig. 8. Dimension and composition of the phantoms. (A) All phantom samples have the same overall dimension. They all feature a "flesh-organ" (left half) and a "flesh" region (right half). (B) Different silicone mixture ratios were used to mimic the three types of liver regions. The mixture utilised was a combination of Ecoflex 00-10 and Slacker (Smooth-on, USA), which displays similar physical properties to a human liver. The ratios chosen cover a wide range of Young's modulus values (A: 12.2 kPa, B: 9.8 kPa, C: 3.2 kPa) as previously studied by Liang et. al [18].

For each "organ" mixture, 5 different "organ" thicknesses (i.e. 15, 20, 25, 30, 35 mm) were generated, resulting in 15 phantoms with varying "flesh" depth and "organ" stiffness.

The "flesh" layer silicone mixture remained the same across all samples and was recreated using a standard Ecoflex 00-10 silicone 1:1 ratio (part A:B). Details on the silicone mixture ratios and Young's modulus values for the "organ" regions can be found in Figure 8.B.

The set-up system used in Experiment 1 was identical to that of Experiment 2. The TIMESETL piezo-electric contact microphone was attached to the surface of the new silicone phantoms. It was positioned at 40 mm from the short edge and 30 mm from the long edge of the phantom. The area of impact from the percussion device was positioned at 25 mm from the short edge and 30 mm from the long edge of the phantom.

For each phantom, two sessions of percussion were recorded, one at the homogeneous "flesh" side, and one at the composite "flesh-organ" side. The percussions on the "flesh" side were performed to understand if the system could capture the nearby presence of organs when impacting flesh alone.

The percussion robot was programmed to perform over 250 percussion actions for each phantom side. Similarly to experiment 1, audio signals were transmitted with a Focusrite Scarlett 4i4 3rd Gen USB Audio Interface, and recorded at 44.1 kHz.

In Experiment 2, a total of 7,500 percussion samples across the 15 different phantom combinations were recorded.

The resulting audio samples were pre-processed in a similar manner to experiment 1 — individual percussion events were extracted and the spectro-temporal features were analysed through CWT using the Symlet family. Key audio features were identified from the CWT through Gaussian Mixture Models and neural networks, creating a predictive model aimed at classifying the 15 different silicone phantoms.

2) **Exploration of Neural Networks for Prediction and Analysis:** In order to create an automatic classification method that can identify physical properties from the audio sample, it is necessary to understand which data features are the most useful in creating a predictive model. This also serves to validate that the observed differences from the scalogram plots are consistent and reliable enough for classification.

Through visual inspection of the CWT scalogram plots, it is clear that particular regions of the graph differ between samples, corresponding to particular time variant frequency bands. In order to extract these spatial features, Gaussian Mixture Modelling (GMM) was used to generate a number of optimal clusters to best represent the 3D data. This significantly reduces the dimensionality of the high resolution CWT plot. The method was applied to each of the 7,500 audio samples collected.

Firstly, the 3D scalogram plots were downsampled by a resize factor of 0.01, as shown in Figure 9.

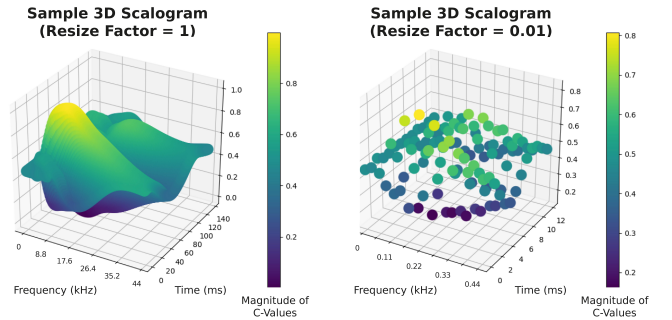


Fig. 9. Before (left) and after (right) downsampling the wavelet transform by a factor of 0.01. This was done to improve the runtime of the GMM computation. Experimentation with smaller scaling factors showed that it had negligible effect on the generated datapoints

Then, the GMM algorithm was applied to the average

scalogram from all samples. A total of 5 Gaussian Distributions (GD) were determined to be the optimal number using Bayesian Information Criteria (BIC). The Gaussian distributions were then calculated and their contours plotted in Figure 10.

This operation allowed to reduce the high resolution of the scalograms to mean regions or clusters in 3D space and a 3x3 covariance matrix, which was decomposed into 3 variances by taking the diagonal values.

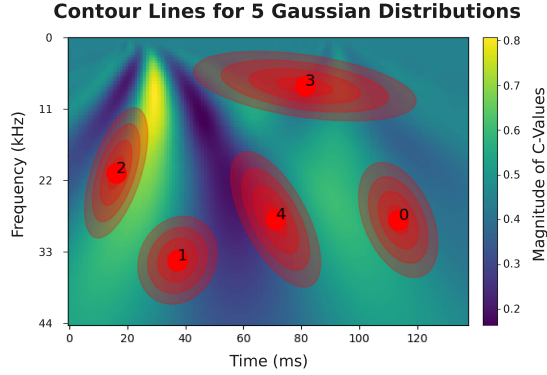


Fig. 10. An example of the 5 Gaussian distributions superimposed on the original scalogram data. The middle dots represent the means, and each of the ellipsoids represent a single standard deviation. This is a 'top down' view for clarity, the data is three-dimensional. Each distribution is numbered to allow consistent comparison of individual distributions across samples.

The dataset totals 30 features (Gaussian means and variances), 3 individual labels (sample type, sample letter and sample depth) and a composite label for 7,500 data points. Sample type describes whether the area percussed in the silicon sample is 'flesh' or 'liver-flesh', sample letter describes which mixture ratio of silicone was used to manufacture the sample, and sample depth describes the height of the liver part of the sample. The composite label combines the information from these three individual labels.

Neural networks were chosen to build predictive models of the dataset. Certain features were selectively omitted to better understand which spatial regions of the scalogram were important for an accurate classification. This experiment was repeated for each individual labels and for the composites of all labels. 30% of the randomized data was used for testing, with the remaining 70% used for training. Validation data was not used, as the objective of the experiment was not necessarily to fine-tune the most accurate model, but to explore which combination of labels and features could give the best results.

The first neural network considered the main label as a composite label of sample type, sample letter and sample depth, giving 30 classification categories.

The remaining neural networks were trained considering each of the 3 label types individually. The letter label (i.e. A, B, C) data set and the sample type (i.e. flesh –F, liver-flesh –L) data set consisted of 3 and 2 categories respectively and both were run with the same architecture as the previous label composite network, where categorical cross-entropy

was computed using a Softmax output activation function. As the depth label is a regression function ranging from 5 mm to 30 mm, an architecture with Mean Squared Error (MSE) loss function, linear activation function and mean squared error minimisation was chosen. The cross-entropy loss and categorical accuracy for the mentioned neural networks can be found in Figure 11 and Figure 12, respectively.

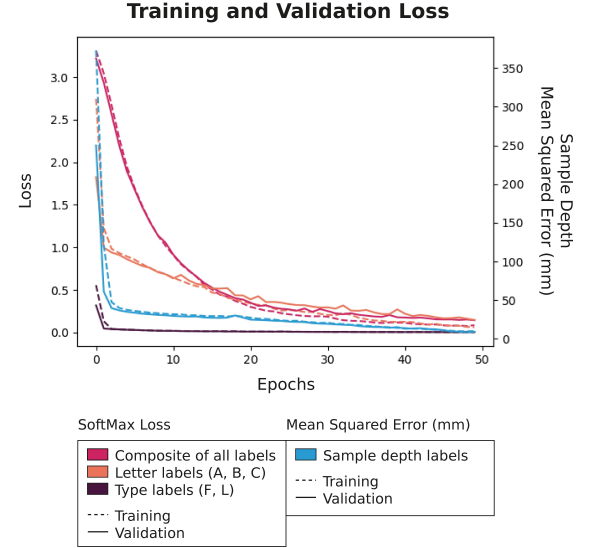


Fig. 11. Categorical cross-entropy loss against epochs for the NN on training and validation models. The composite of all labels represents the model totalling 30 categories, utilising the letter, type and depth labels (e.g. B25F - sample B, liver of 25 mm height, percussed on the flesh side).

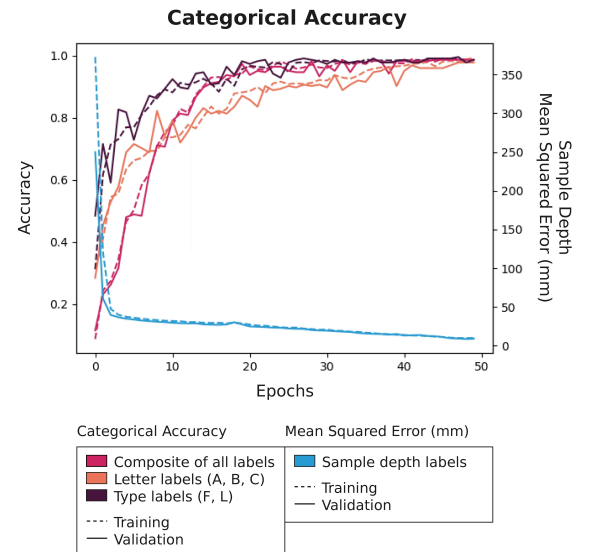


Fig. 12. Categorical accuracy and MSE against epochs for the NN on training and validation models. The composite of all labels represents the NN totalling 30 categories, utilising the letter, type and depth labels (e.g. B25F - sample B, liver of 25 mm in height, percussed on the flesh side).

The next stage of the experiment was to identify which time variant features were the most important ones for label classification. The features were grouped into 5 Gaussian distributions centered at optimal points in the 3D space of

the CWT scalogram. By analysing the change in performance of the neural networks on different Gaussian distributions, particular regions of space were effectively obscured from the neural networks' view, allowing for the identification of the Gaussian distributions of interest.

Each of the aforementioned networks were trained on data from either all the Gaussian distributions or only one of the 5 distributions, and the success of each network was evaluated by comparing their categorical accuracy or mean squared error. This resulted in a total of 24 networks trained in the study.

3) Experimental results: When considering all 5 Gaussian distribution areas, the classification achieved a 97.5% accuracy for the model trained on the composite of all labels (total: 30 categories, e.g. C15L) as shown in Table II. This proves that the neural network accurately fitted a function of the sample type, sample depth and sample letter, and that machine learning methods are suitable to accurately classify certain physical properties of a percussion sample.

The results of the neural networks trained on the individual labels (i.e. only letter, type or depth in Table II) allowed us to understand how classification performs in isolation, as the neural network has no perception of the scope of labels outside those on which it has been trained. The letter label data set achieved an accuracy of 96.3%. Likewise, the sample type neural network achieved 98.7% accuracy. The depth label presented a mean squared error of 9.07 mm. These experiments show that it is possible to classify labels accurately in isolation and that different physical properties can be predicted using neural networks.

TABLE II

CATEGORICAL ACCURACY OR MEAN SQUARED ERROR (MSE) OF NN.

	All GDs	GD 0	GD 1	GD 2	GD 3	GD 4
Composite Accuracy	0.975	0.502	0.977	0.891	0.397	0.946
Letter Accuracy	0.963	0.753	0.801	0.749	0.806	0.687
Type Accuracy	0.987	0.987	0.914	0.935	0.875	0.732
Depth MSE (mm)	9.07	9.097	3.968	6.926	13.01	6.775

Accuracy and MSE of the neural network models in the classification of sample labels when considering all Gaussian distributions (GD) or single distributions. These labels could be from composite nature (i.e. with letter, type and depth data), or contain only letter, type or depth data.

By analysing the performance of neural networks trained on different Gaussian distributions we can also understand which time variant frequencies are helpful for the classification of the different labels and for the composite labels. Understanding this can help reduce computational time in future studies by exploring only specific spectro-temporal regions. Figure 13 and Table II show the accuracy comparison between sibling networks.

The results show that higher frequencies occurring closer in time to the percussion event (GD 1 - see Fig. 10) are crucial at classifying the composite labels (97.7% accuracy). This Gaussian distribution is also particularly relevant when

predicting a sample's depth, providing a lower depth estimation error than when exploring all distributions (MSE: 3.97 and 9.07 mm, respectively). This is not the case for classifying the sample type, which showcases slightly better results when trained on frequencies higher in pitch but further in time from the moment of percussion (GD 0); however, the distribution effect of these frequencies in time does not appear to be as sizeable across networks as when predicting sample depth.

Classification Accuracy per Gaussian Distribution

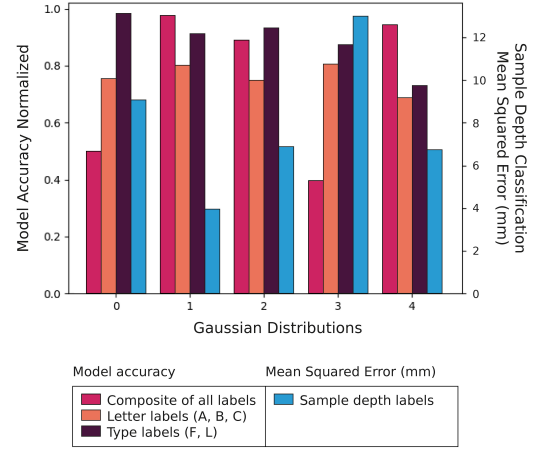


Fig. 13. Bar chart showing the accuracy of the neural networks trained on each of the 5 Gaussian distribution regions (see Fig. 10). Overall, distribution 1 showcases the best classification accuracy, specifically when predicting the depth label (MSE: 3.97 mm). This implies that, while the informative frequencies are distributed throughout the sample in time, the most immediate waves after the percussion event are overall the most significant for sample classification.

Moreover, from the categorical cross-entropy loss graph (Fig. 11), it is inferred that models trained on individual labels have a faster learning rate than the model trained on the 30 composite categories. Nevertheless, all models achieve a similar high accuracy after approximately 50 epochs (Fig. 12). These results are promising signs for the use of machine learning in robot-actuated percussion and tissue characterisation.

DISCUSSION

The proposed robotic percussion device and acoustic response analysis methods presented a 97.5% accuracy when classifying complex silicone phantoms in controlled experiments. The results also highlight, for the first time, that manual percussion must involve a complex sensorimotor integration skill. The embodied process involves fine tuning the stiffness of the wrist and fingers to deliver a controlled impulse input to the underlying soft tissue.

However, in reality, percussion examinations are far from a controlled environment and the size and structural composition of the silicone phantoms may not be representative of the complexity of the human abdominal region. Thus, some of the detected acoustic response characteristics may be due to the phantom's structure and not be present in a patient.

Furthermore, this experiment did not consider the impact of haptic feedback, and ignored the influence of the pleximeter (non-percussing) hand. It is possible that the pleximeter hand presents attenuating qualities, by absorbing and distributing the energy of the plexor. A preliminary comparison between the force profiles of a single medical student and the robotic device confirmed this suspicion. Their transient percussion force profiles are shown in Figure 14. A clear dominant peak followed by smaller ripples is present in the medical student data, whereas the robotic percussion leads to a less damped force profile. This might be due to the previously mentioned attenuation qualities of the pleximeter hand. In the case of the robotic device, the end-effector bounces off the percussed tissue without the presence of a pleximeter element to attenuate the tissue oscillations.

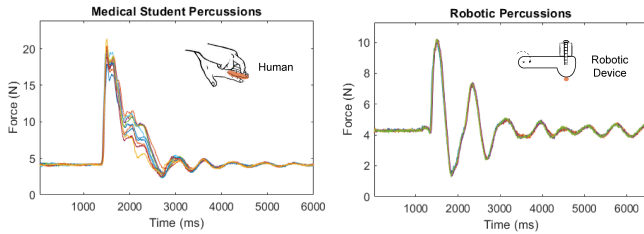


Fig. 14. Time-synchronised force readings of the human and robot performing percussion on the phantoms in Experiment 1, chosen to simulate abdominal tissue [16]. The force profile was recorded with a TedeA Huntleigh 1040 (20 Kg) load cell and NI USB-6341 interface.

The presented research proposes an initial exploration into the field of robot-assisted medical percussion diagnosis, and demonstrates the potential of robotic devices to provide accurate estimates of tissue properties (Table II) and less variability between percussion actions (Fig. 14). There is significant scope for further experimentation, expanding upon the methodology and considerations made in this paper.

CONCLUSION

This paper presents a robotic device that replicates manual medical percussion, with a compliant wrist joint. The robotic device showed comparable torque when performing percussion (8.6 Nm) to those reported in literature [15], and with better consistency between percussion actions.

Through two different experiments, we studied the ability of the system to predict the depth of a buried hard nodule or anomaly (Experiment 1) and to estimate complex tissue characterization (Experiment 2). The study found that accurate sample classification can be obtained through the acoustic feedback data alone, specifically when analysing spectro-temporal features extracted using 1-D Continuous Wavelet Transform. Identification of key Gaussian distributions allowed to reduce task dimensionality and minimised the model's computing time.

The proposed method, for the first time, provides design guidelines for how a robotic device and audio analysis might be used to automate percussion examination. Future studies will explore whether these methods are applicable in more representative abdomen models and human participants.

REFERENCES

- [1] J. B. P. Barth and H. Roger, *Traité pratique d'auscultation; suivi d'un précis de percussion*. Labé, 1860.
- [2] J. C. Yernault and A. Bohadana, "Chest percussion," *European Respiratory Journal*, vol. 8, no. 10, pp. 1756–1760, 1995.
- [3] E. J. Meidl and J. Ende, "Evaluation of liver size by physical examination," *Journal of general internal medicine*, vol. 8, no. 11, pp. 635–637, 1993.
- [4] A. Harris, "Listening-touch, affect and the crafting of medical bodies through percussion," *Body & society*, vol. 22, no. 1, pp. 31–61, 2016.
- [5] L. A. (1722-1809). Percusor and pleximeter, image ref:10289197.
- [6] J. S. Thierman, "Device for medical percussion," Sept. 14 2004, uS Patent 6,790,184.
- [7] D. A. Rice, "Transmission of lung sounds," in *Seminars in respiratory medicine*, vol. 6, no. 03. Copyright© 1985 by Thieme Medical Publishers, Inc., 1985, pp. 166–170.
- [8] V. A. McKusick, J. T. Jenkins, and G. N. Webb, "The acoustic basis of the chest examination; studies by means of sound spectrography," *American review of tuberculosis*, vol. 72, no. 1, pp. 12–34, 1955.
- [9] H. Mansy, T. Royston, and R. Sandler, "Use of abdominal percussion for pneumoperitoneum detection," *Medical and Biological Engineering and Computing*, vol. 40, no. 4, pp. 439–446, 2002.
- [10] K. Ayodele, O. Ogunlade, O. Olugbon, O. Akinwale, and L. Kehinde, "A medical percussion instrument using a wavelet-based method for archivable output and automatic classification," *Computers in Biology and Medicine*, vol. 127, p. 104100, 2020.
- [11] R. Krumpholz, J. Fuchtmann, M. Berlet, A. Hangleiter, D. Ostler, H. Feussner, and D. Wilhelm, "Telemedical percussion: objectifying a fundamental clinical examination technique for telemedicine," *International Journal of Computer Assisted Radiology and Surgery*, 2021.
- [12] J. G. Cherng, M. Eksioğlu, and K. Kızılaslan, "Vibration reduction of pneumatic percussive rivet tools: mechanical and ergonomic re-design approaches," *Applied ergonomics*, vol. 40, no. 2, pp. 256–266, 2009.
- [13] Y. Peng, Z. Dai, H. A. Mansy, R. H. Sandler, R. A. Balk, and T. J. Royston, "Sound transmission in the chest under surface excitation: an experimental and computational study with diagnostic applications," *Med Biol Eng Comput*, vol. 52, pp. 695–706, 2014.
- [14] A. Rao, J. Ruiz, C. Bao, and S. Roy, "Tabla: A proof-of-concept auscultatory percussion device for low-cost pneumonia detection," *Sensors*, vol. 18, no. 8, p. 2689, 2018.
- [15] M. S. Rao, "Thumb-to-finger method of percussion: a novel, modified technique that is less painful and more productive than the conventional method," *Singapore Med J*, vol. 1, p. 11, 2021.
- [16] M. Li, J. Konstantinova, G. Xu, B. He, V. Aminzadeh, J. Xie, H. Wurdemann, and K. Althoefer, "Evaluation of stiffness feedback for hard nodule identification on a phantom silicone model," *PloS one*, vol. 12, no. 3, 2017.
- [17] D. Yan *et al.*, "Detection of hmm synthesized speech by wavelet logarithmic spectrum," *Automatic Control and Computer Sciences*, vol. 53, no. 1, pp. 72–79, 2019.
- [18] L. He, N. Herzig, S. de Lusignan, and T. Nanayakkara, "Granular jamming based controllable organ design for abdominal palpation," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 2154–2157.

ACKNOWLEDGEMENTS

This work was supported by the EPSRC Robopatient Grant No EP/T00603X/1.

AUTHOR CONTRIBUTIONS STATEMENT

Y.T. and B.C. developed and analysed the initial robotic device. P.Z.Q., O.T. and Y.T. conducted the experiments and analysed the results, under the guidance of T.N. Finally, P.Z.Q., O.T., Y.T. and T.N. reviewed the manuscript.

ADDITIONAL INFORMATION

All authors declare no competing interests on this paper.