Journal of Sound and Vibration 594 (2025) 118667

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

Active acoustic sensing for determining touch location on an elastic surface

Benjamin R. Thompson *, Tre DiPassio, Jenna Rutowski, Mark F. Bocko, Michael C. Heilemann

Departent of Electrical and Computer Engineering, University of Rochester, United States

ARTICLE INFO

Keywords:
Touch sensing
Touch interface
Active acoustic sensing
Vibration sensing
Vibroacoustics

ABSTRACT

In this paper, we describe the development and testing of an active acoustic sensing (AAS) touch interface. The interface detects subtle changes in the vibrational characteristics of an elastic panel when a user applies a small force to the surface at different locations. The system consists of a panel with an affixed force exciter and vibration sensor to monitor the panel vibrations. An automated system was used to manipulate a stylus to apply a light force to the panel at an array of known locations and the vibrational response of the panel was recorded by the affixed sensor. We employed a rectangular grid of 414 points with 10 mm spacing on a 2 mm thick acrylic panel of dimensions 18 cm by 23 cm. Features of the recordings were employed as training data for a deep neural network. The results demonstrate the viability of the AAS interface, and they show the relative performance of the system as a function of the selected features. The demonstration platform achieved a classification accuracy of 100% and a mean distance error of 0.20 mm for regression. The best-performing feature sets were those that contained sufficient spectral resolution to discriminate subtle changes in the center frequencies and amplitudes of the panel's modal resonances in response to small changes in touch location. The AAS touch interface is a practical and inexpensive means to provide accurate touch sensing for large surfaces, such as displays, televisions, and information kiosks.

1. Introduction

Over the last two decades, the touchscreen has become the dominant input modality for human–machine interfaces in devices such as tablets and smartphones [1]. Since its inclusion in the first iPhone by Apple in 2007 [2], projected capacitance (p-cap) has become the preferred approach to touch sensing for small-screen devices due to its high accuracy and resolution, capacity to function with minimal applied pressure, ruggedness, and ability to recognize multiple touches simultaneously [3–6]. Though widely used, the technology is limited in a number of ways as enumerated in [3]. For example, it is relatively expensive to implement compared with competing touch-sensing technologies such as resistive and some infrared approaches [1] and it can be difficult to construct a completely transparent p-cap screen for large devices. Additionally, devices relying on the p-cap sensing mechanism generally require that contact is made with a finger or a specialized stylus capable of modifying the fringing capacitance between electrodes, and the performance of p-cap systems tends to suffer in damp environments [7]. Finally, p-cap devices are incapable of detecting input pressure directly, so they are less suited to applications where input pressure sensitivity may be useful.

Acoustic-based touch sensors, which function by analyzing the change in the vibration response of a structure when contact is made by a user, have the potential to address some of the disadvantages of p-cap touch systems. In general, there are two approaches

* Corresponding author.
E-mail addresses: benjaminrthompson.mail@gmail.com (B.R. Thompson), mheilema@ur.rochester.edu (M.C. Heilemann).

https://doi.org/10.1016/j.jsv.2024.118667
Received 23 April 2024; Received in revised form 9 July 2024; Accepted 5 August 2024
Available online 6 August 2024

to acoustic touch sensing: active and passive. In passive acoustic sensing systems, waves are induced in the panel by the user when they touch it [8–11]. In active acoustic sensing (AAS) systems, waves are induced by a dynamic force exciter mounted to the panel and the user's touch changes the vibrational behavior of the surface [12–16]. In addition to the lower-frequency approaches described in these works, some active acoustic sensing systems make use of higher-frequency excitations in the form of Lamb waves and ultrasonic surface waves. Surface acoustic wave (SAW) approaches rely on the user obstructing the propagation of ultrasonic Rayleigh waves at the surface of a panel [17]. Lamb wave systems work either by extrapolating location from the time it takes for a wave to arrive at a sensor after being reflected off of a stylus or finger as in [18], or by monitoring changes to the spectral response of a plate to a lamb wave excitation that result from diffraction caused by a finger or stylus as in [19]. Time difference approaches, regardless of the frequencies at which they operate, require at least three sensors in order to triangulate a location. Implementation of spectral approaches can theoretically be accomplished with a single sensor; however, in practice, these systems often employ multiple sensors to increase reliability, for example, by rejecting a location determination if the data from different sensors gives conflicting estimations.

A drawback of the passive approach is that the signal-to-noise ratio of the system is strongly dependent upon how the user engages with the surface. For example, if a user places their finger lightly on the surface of a touch screen, that touch may be misinterpreted or go undetected. AAS systems are more robust in this regard but have the disadvantage that they require that a potentially audible, system-generated signal be used to excite the surface. Both acoustic approaches can be implemented using inexpensive sensing elements, which makes them a viable option for larger panels where a p-cap system would be cost-prohibitive. Additionally, acoustic approaches have the potential to outperform p-cap sensing in wet environments, and when the user is wearing a glove or touching the surface with an object not explicitly designed to serve as a p-cap stylus. Acoustic systems are also capable of measuring touch force, opening the door for pressure-sensitive touch surfaces that provide an additional dimension of input.

This work is intended to further the literature around lower-frequency active acoustic sensing by presenting an analysis of the physical principles on which these systems operate, utilizing the results of this analysis to inform the signal processing and machine learning techniques that are used in a prototype system, and then evaluating the performance of that prototype system. To minimize the hardware costs associated with integrating the system into commercial products, a single sensor was employed to measure the vibration response to a low-frequency excitation produced by a single actuator. A low-frequency excitation allows the prototype to operate at a low sample rate which has benefits in terms of data and computational overhead that may be significant if the system is deployed on low-cost, low-power embedded edge devices. Though a low-frequency excitation comes at the expense of an audible signal when compared with an ultrasonic approach, we believe that this effect can be harnessed as a form of feedback to inform the user that the system has recognized an interaction. In the next section, a physical model of a vibrating elastic surface is developed to provide context for the systems and methods presented for AAS.

2. Theoretical development

Cremer et al. [17] and Fahy et al. [18], give the equation of motion for a thin, damped, isotropic panel driven by an external pressure, $P(x, y, t)$,

$$D\nabla^4 u(x, y, t) + \rho h \frac{\partial^2 u(x, y, t)}{\partial t^2} + b \frac{\partial u(x, y, t)}{\partial t} = P(x, y, t), \tag{1}$$

where $u(x, y, t)$ is the displacement of the panel normal to the surface at position (x, y) at time t, ρ is the density of the panel, h is the panel thickness, b is the damping coefficient, and D is the panel's bending stiffness given by,

$$D = \frac{Eh^3}{12(1 - \nu^2)}, \tag{2}$$

where ν is the Poisson's ratio and E is the Young's modulus of the material.

The actuator used to excite an AAS system at location (x_0, y_0) may be approximated as a normal, harmonically driven point force, F, in the limit where the bending wavelength is large compared to the size of the actuator. Assuming the panel is rectangular, the solution to (1) may be presented in terms of a summation of bending modes (m, n) in the vertical and horizontal dimensions respectively as,

$$\widetilde{u}(x, y, \omega) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} F \widetilde{\alpha}_{mn} \boldsymbol{\Phi}_{mn}(x, y), \tag{3}$$

where $\widetilde{u}(x, y, \omega)$ is the panel's displacement as a function of position and frequency, $\boldsymbol{\Phi}_{mn}$ is the spatial response of the (m, n)th mode, and $\widetilde{\alpha}_{mn}$ is the complex amplitude of the (m, n)th mode given by,

$$\widetilde{\alpha}_{mn} = \frac{4\boldsymbol{\Phi}_{mn}(x_0, y_0)}{\rho h S(\omega_{mn}^2 - \omega^2 + j\omega_{mn}\omega/Q_{mn})}, \tag{4}$$

where S is the surface area of the panel, and ω_{mn} and Q_{mn} are the frequency and quality factor of the (m, n)th mode, respectively [19].

In the AAS system shown in Fig. 1a, an input signal $x(t)$ is imposed on a panel through a dynamic force exciter. The signal arriving at an affixed vibration sensor, $y(t)$, is filtered according to a transfer function $h_0(t)$, which may be adapted from (3) using the appropriate actuator and sensor locations. When a constant force is applied to the panel, by a stylus for example, additional constraints are imposed alongside the initial boundary conditions. The addition of these constraints alters the transfer function as illustrated in Fig. 1b. How the transfer function is altered is dependent on the location at which the force is applied. This dependence

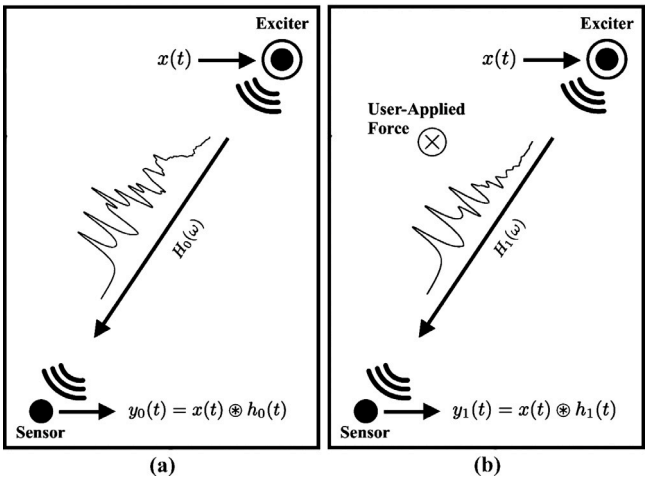

Fig. 1. Functional diagram of an AAS system (a) with and (b) without a user-applied force. The transfer function $H(\omega)$ from the signal played by the exciter $x(t)$ to the signal received by the sensor $y(t)$ is altered by the magnitude and location of the applied force.

can be leveraged by recording the signal with a force applied at many locations on the panel, and either using features from these recordings to train a machine learning model [16], or comparing incoming signals to this library of recorded signals directly [12] to determine where on the panel a new force is being applied.

An exact model describing the changes in the vibrational response of a panel contacted by a finger or stylus in different locations is outside of the scope of this paper. However, a general understanding of the way in which touch interactions at different locations alter the vibration response of a panel may allow the definition of feature sets that are more efficient and higher-performing, and allow designers to make informed decisions about materials, dimensions, and exciter and sensor locations. To that end, we present a brief empirical investigation into the effects of applying a constant, localized, normal force on the resonant properties of a panel.

3. Empirical analysis

A Polytec PSV 500 scanning laser vibrometer was used to measure the velocity response of a $180\,\text{mm} \times 230\,\text{mm}$ foam board panel while a $\sim 1\,\text{N}$ force was applied to various locations on the surface by the probe of a spring-loaded dial indicator. The panel was driven by an inertial exciter located at $(0.705L_x, 0.241L_y)$ to ensure coupling to the three lowest-frequency bending modes. The presence of the applied force produced three observable effects.

First, as the location at which the force is applied approaches the antinode of a mode, the magnitude of the resonance associated with that mode is attenuated. This is shown in Fig. 2 where it can be seen that as the force location approaches an antinode of the (1,1) and (1,2) modes illustrated in the insets at the left and center, the magnitudes of the resonances associated with those modes decrease. However, since all the force locations lie on a nodal line for the (2,1) mode illustrated by the inset on the right, the magnitude of the resonance associated with that mode is largely unaffected as the force location changes. The same phenomenon can be observed in Fig. 3, which shows that as the force location approaches the antinode of the mode shown in Fig. 3b–g, the magnitude is reduced, whereas the responses are almost identical when no force is applied in Fig. 3a and when the force is applied on a nodal line in Fig. 3h.

Second, the presence of an applied force can alter the spatial response in the bandwidth of a modal resonance. This is illustrated in Fig. 3, which shows the spatial response in the resonant bandwidth of the (1,2) mode being altered as the force location approaches one of its antinodes. As is the previous case, changes to the spatial response are most pronounced for a given mode when the force location lies in the antinodal region of that mode, and the effects are negligible when the force location is on a nodal line. This is evident in Fig. 3h where it can be seen that the spatial response is unchanged when the force is applied in the middle of the panel - a nodal point for the (1,2) mode.

It is possible that the first two effects are related. Since it is apparent from (3) that the spatial response of a panel is always given by a superposition of resonant modes, reducing the amplitude $\tilde{\alpha}_{mn}$ of a particular mode (especially at the resonant frequency of that mode) inherently increases the relative contributions of the other modes to the overall spatial response. This effect is demonstrated in [20,21], where vibrations localize on the surface of a plate to the area around the exciter in frequency regions of high modal overlap.

Finally, the measurements also reveal that the application of a localized, constant, normal force can cause a shift in the center frequencies of the modal resonances of a panel. As is the case with the magnitude suppression and mode shape distortion described above, the resonances that exhibit the largest shift are those associated with modes that have antinodes near the location at which the force is applied. This can be observed in Fig. 2, which shows that as the force location approaches an antinode of a mode, the center frequencies of the resonances associated with that mode increase. Again, because all the force locations lie on a nodal line

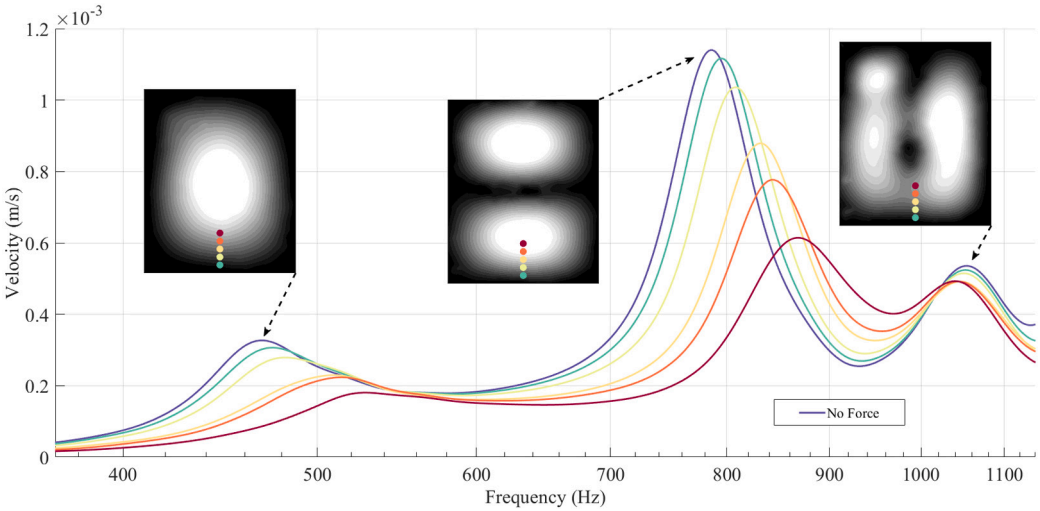

Fig. 2. Velocity magnitude response of a foam board panel with a ∼1 N normal force applied to different locations on the surface measured using a scanning laser vibrometer with 1/12th octave smoothing. The location at which the force was applied is indicated by the dots with the corresponding colors on the inset. The velocity magnitude response when no force is applied is shown by the violet trace, and the spatial response is depicted in each inset. The responses shown by the traces were measured at $(0.25L_x, 0.25L_y)$.

for the mode shown by the inset on the right, the center frequency of the resonance associated with that mode is largely unchanged between each force location. The increase in each resonant frequency is likely due to an increase in modal stiffness and/or a change in boundary conditions as a result of the applied force. A similar effect was shown in [22] to approximate the increase in modal resonant frequencies when the boundary conditions are changed from simply-supported to clamped.

From these observations, one can conclude that when designing a feature set to be used in an AAS system it is insufficient to incorporate the magnitude response only in bandwidths defined by the static modal resonances of the untouched panel [23,24]. Instead, the observed frequency range must be wide enough to accommodate the shifted resonances that will occur as a result of anticipated force interactions. Additionally, the physical resolution of the system will be limited by both the magnitude and frequency resolution of the feature set, as small changes in touch location will produce small changes to the response.

4. Experimental setup

A prototype AAS system was constructed using a 2 mm thick acrylic sheet affixed to a wooden frame with two-part epoxy. Its effective dimensions (excluding the frame) were $L_x = 180$ mm and $L_y = 230$ mm. A Billionsound 14H26 moving-coil actuator was placed with its center at (x, y) location $(153\,\text{mm}, 203\,\text{mm})$, and a PCB Piezotronics U352C66 accelerometer [25] was placed at position $(60\,\text{mm}, 76.7\,\text{mm})$. The sensor and actuator locations were chosen so that they lay in the antinodal region of a number of lower-order modes. This arrangement can be seen in Fig. 4.

To facilitate efficient data acquisition and repeatability for testing, a three-axis CNC machine with sub-millimeter resolution was used to provide the touch inputs to the panel. A spring-loaded Mitutoyo 575-123 electronic dial indicator [26] was mounted to the CNC machine to act as a stylus as shown in Fig. 5. At each location, the probe of the dial indicator was used to apply a constant, localized, normal force on the panel of ∼0.6 N.

The requirements for the excitation signal were that it be sufficiently wide-band to excite a large number of modes, and short enough in duration that the device could be deployed for real-time location sensing. With these criteria in mind, an excitation signal consisting of a full-scale sample followed by approximately 100 ms of silence, was chosen. The excitation of the panel created an acoustic impulse a few milliseconds in duration with a typical peak sound pressure level of less than 65 dB SPL measured at a distance of 20 cm from the panel. An attempt has not yet been made, however, to characterize the relationship between excitation level and the ability of the system to determine touch location, and it is possible that a significantly lower excitation level may yield acceptable performance. For training and testing data collection, a rectangular grid of 414 points with 10 mm spacing was defined on the surface of the panel as shown in Fig. 4. At each location, a force was applied using the stylus, and the panel's response to the excitation signal was recorded 1000 times for a total of 414,000 100 ms responses at a sampling frequency of 8 kHz resulting in a 4 kHz bandwidth. As the purpose of this exercise was to establish the performance of various feature sets under controlled conditions, the force magnitude, as well as the stylus shape and material, were kept constant throughout the training data, and the system was only tested using force locations on which the models were trained. Ultimately, to create a robust system that is effective

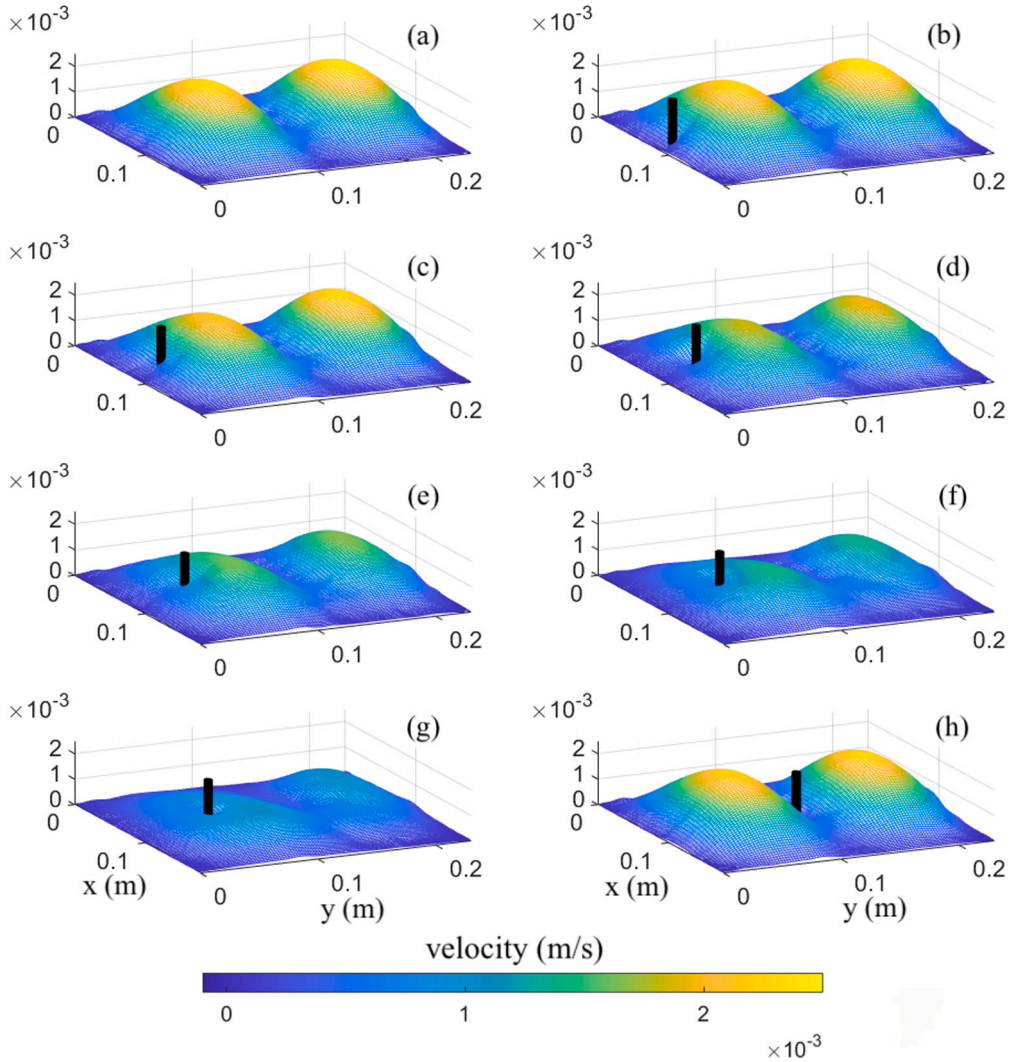

Fig. 3. The measured velocity magnitude response over the surface of a foam board panel from 630 Hz to 1 kHz, the bandwidth of the resonance associated with the (1,2) mode, (a) when no force is applied, and (b) - (h) when a ∼1 N normal force centered in the x dimension is applied at vertical locations $y = 10, 20, 30, 40, 50,$ and 60 mm, respectively. The black cylinders indicate the location at which the force was applied. Note that there is virtually no difference in the responses when no force is applied in (a) and when the force is applied at $(0.5L_x, 0.5L_y)$ in (h) because that location lies on a nodal line for this mode.

in real-world scenarios, the network will need to be trained on more diverse data, and testing will need to be done to ensure that the system can generalize to any input it is likely to encounter.

A design goal for the prototype was to create a system that was not reliant on any specific plate boundary conditions. For this reason, along with a desire to develop a system that might eventually be compatible with a human finger in addition to a number of different stylus materials and shapes, it is not practical to employ an analytic solution to the problem of determining where a force is being applied given the signal at a sensor. One approach that has been employed in this context is to develop a library of responses to different touch locations and then use an iterative method to determine which library response an unknown response most resembles, and therefore where the touch is likely located [12,27]. One drawback to this arrangement is that to collect many responses for the same touch location using different forces or different styluses, all the responses for a given location need to be weighted and aggregated. Using a machine learning approach, this aggregation and weighting is done as part of the training of the network, and it is done in such a way as to minimize the error in the validation set. A machine learning approach also has the potential advantage that a trained network can be deployed on low-power, dedicated hardware that is optimized for the task, with the upshot being extremely fast and efficient location predictions. Though in the system presented here the force and stylus were kept constant as it is meant only as a proof of concept, a machine learning approach provides a versatile platform on which the system could evolve.

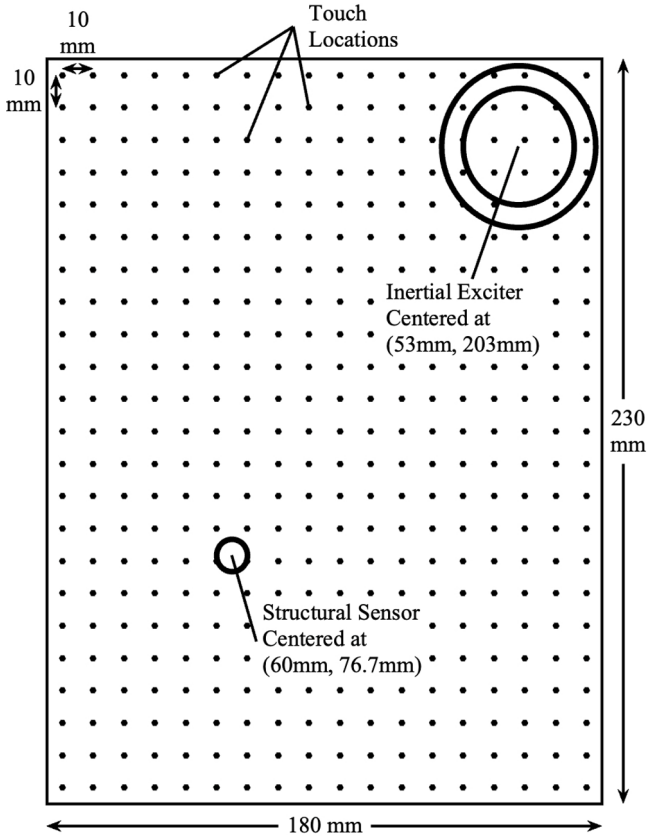

Fig. 4. Diagram of physical arrangement for prototype AAS system including locations of all touch points used for training and testing.

In addition to affecting the performance of a model, the choice of feature set has implications with regard to the feasibility of deploying an AAS system in real time on an embedded system, as the processing power and memory required to perform the extraction can vary substantially. Additionally, in much the same way that the measurements presented in Section 3 can inform a choice of feature set, the relative performance of different feature sets can inform our understanding of the dependence of a panel's vibration response on the location of an applied force. From the analysis in Section 3, it appears that any potential feature set must capture the bands of the spectrum that contain the modal resonances when no external force is applied and that these bands must be wide enough to capture the shifted resonances when an external force is applied. Furthermore, these portions of the spectrum must be captured with enough frequency resolution to differentiate small shifts in the resonant frequencies of the system and enough acceleration resolution to capture small changes to the magnitude of the resonant peaks. With these criteria in mind, ten feature sets were chosen, where the amount of spectral information carried by the features varied from none at all to the maximum amount possible given the sample rate, frame length, and bit depth. Additionally, some feature sets carrying the same spectral information, but in different forms, i.e. the time domain data and the one-sided Discrete Fourier Transform of that same data, were included to gauge the networks' dependence on how the information was presented to it. The intention is that the performance of these feature sets might be useful to future developers of AAS systems given the requirements and limitations of their particular use case. Descriptions of each feature set along with the equations for extracting them from a sampled time domain signal can be found in Appendix, and plots illustrating the features from several locations for some of the feature sets are shown in Fig. 6.

Each feature set was used to train four neural networks with distinct architectures. The networks employed included a feedforward neural network (FNN) for classification, and a convolutional neural network (CNN) for classification, as well as an FNN and CNN for regression. A block diagram detailing the specific architectures for each network is shown in Fig. 7. Note that the intention of this work is to explore whether it is possible to train a neural network to determine the location of a force on a surface using data from a structural sensor; it is not to determine the optimal configuration for such a network. The 1000 data points for each force location were split into 600 responses for training and 200 responses each for validation and testing. Each network/feature set combination was trained for 1000 epochs at 485 iterations per epoch, or until validation accuracy did not increase for 1000 consecutive iterations. The highest-performing version was retained for testing.

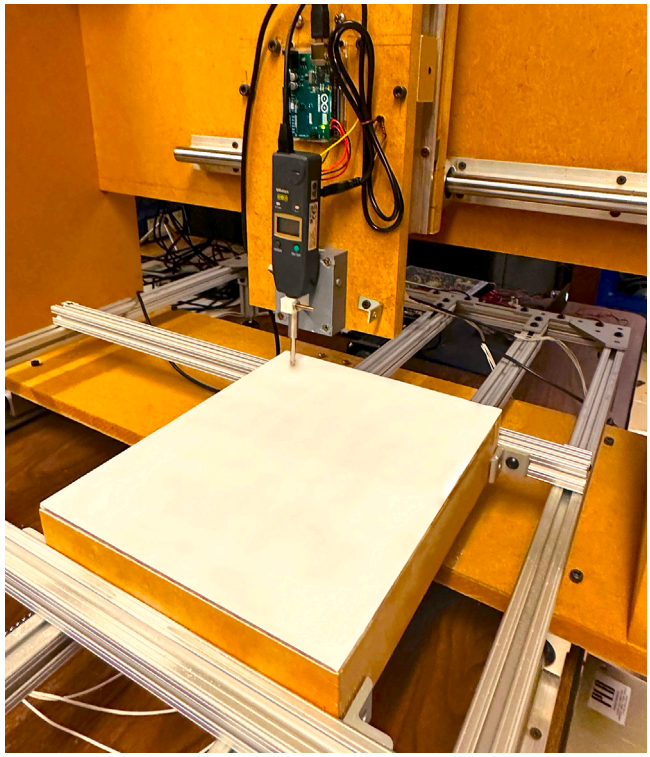

Fig. 5. Prototype AAS acrylic panel mounted on a CNC machine. A dial indicator acts as a stylus, applying a constant, localized, normal force to the surface.

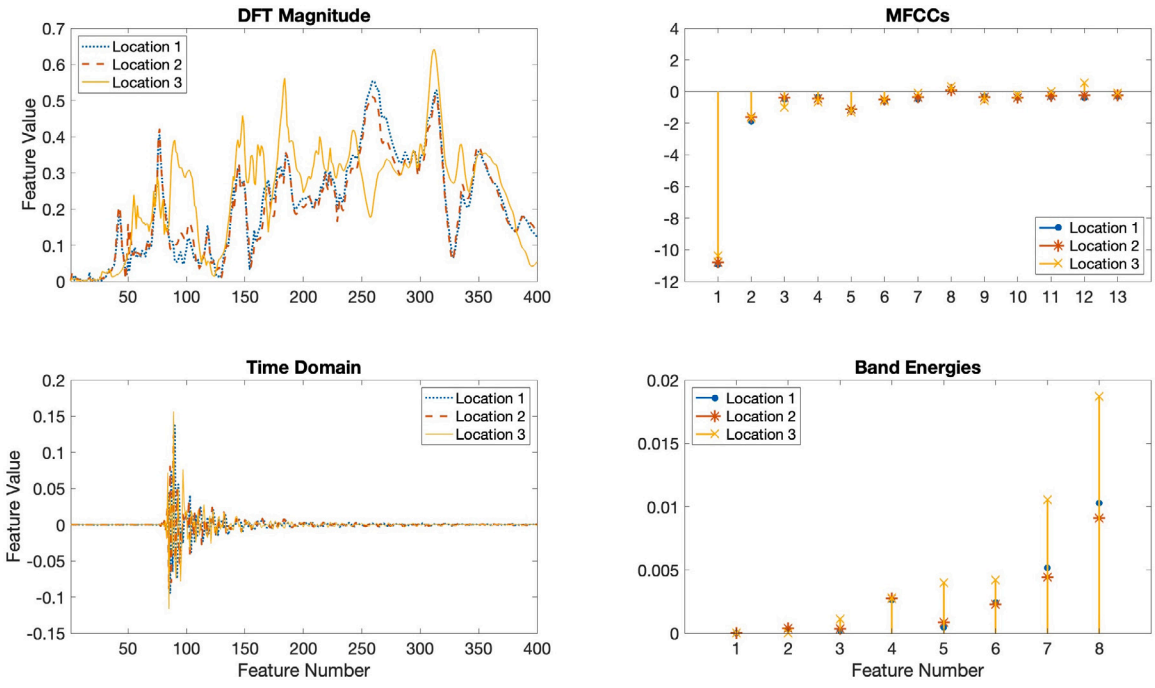

Fig. 6. Features for three force locations and four different feature sets. Locations 1 and 2 are physically close together on the plate, while location 3 is several centimeters away.

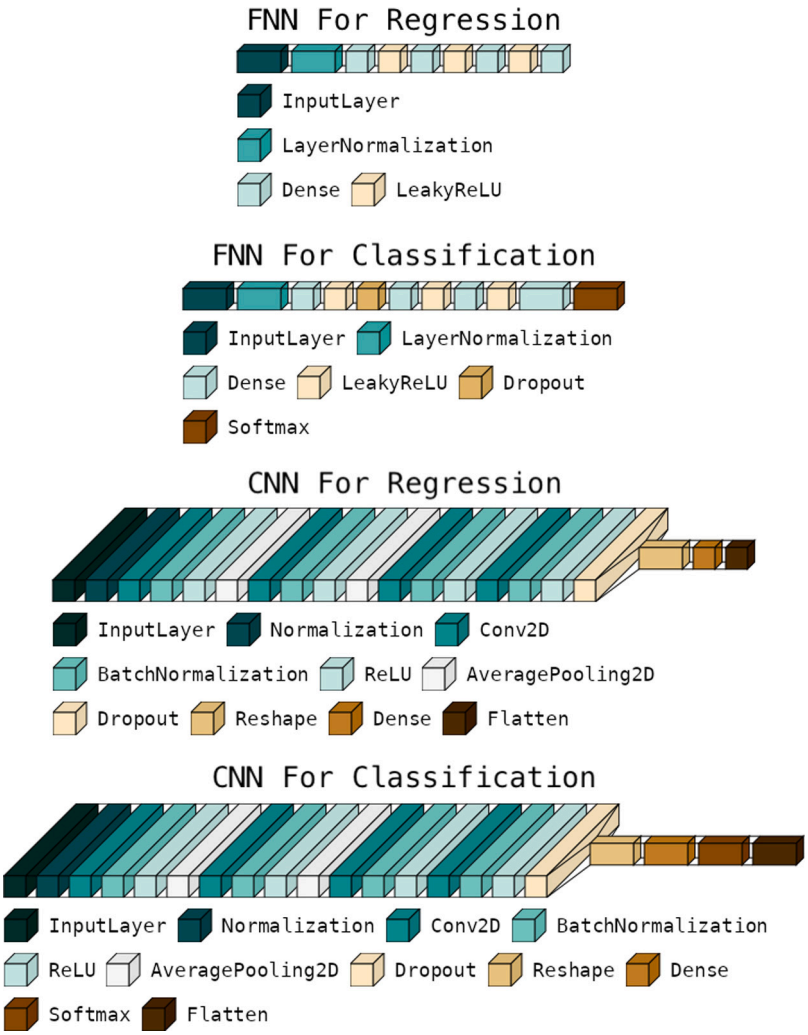

Fig. 7. Architecture for each neural network. The input size depends on the feature set but is always one-dimensional. These visualizations were produced using [28].

5. Results & discussion

The performances of each network using each feature set are shown in Table 1. The average accuracy of each classification network is determined by the number of correct predictions as a percentage of the total number of predictions. The accuracy of each regression network is determined by the average distance between the predicted force location and the actual force location, referred to as mean distance error (MDE) and given by,

$$\text{MDE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}, \tag{5}$$

where N is the number of observations, $\hat{x}$ and $\hat{y}$ are the horizontal and vertical coordinates of the predicted location, respectively, and x and y are the horizontal and vertical coordinates of the observed location, respectively. A more accurate network produces a smaller MDE.

Both classification network architectures performed at greater than 99% accuracy using all feature sets with the exception of Total Energies. Excluding the Total Energies feature, the worst-performing network and feature set combination was the FNN using the Band Energies feature set, which generated a test accuracy of 99.31%. The top four combinations, the CNN and FNN using Mel Spectrum and DFT magnitude, all displayed 100% test accuracy. In general, the CNN for classification outperforms the FNN

Table 1
Average accuracy and MDE across all force locations for each network architecture and feature set.

	Classification average accuracy		Regression MDE (mm)	
	FNN	CNN	FNN	CNN
Total Energies	0.24%	24.35%	78.72	67.30
DFT Points	99.33%	99.59%	3.10	9.42
Band Energies	99.31%	99.37%	3.79	12.51
MFCCs	99.69%	99.76%	2.05	5.58
Mel Spectrum	100.00%	100.00%	0.62	3.66
DFT Magnitude	100.00%	100.00%	0.22	1.39
DFT Phase	99.95%	99.96%	0.26	1.99
DFT Magnitude & Phase	99.94%	99.96%	0.21	1.59
Time domain	99.98%	99.99%	0.21	1.81
Complex DFT	99.97%	99.98%	0.20	1.41

for classification, though further testing would be required to verify that conclusion in general once each network architecture has been optimized to achieve maximum accuracy under the test conditions.

The results from the regression networks provide a more nuanced picture of the performance of each feature set, as the results for each individual prediction are not quantized to a binary state (correct or incorrect) as is the case for classification. In general, the more spectrally-complete feature sets outperform the less spectrally-complete ones. As with the classification networks, the Total Energies feature set, which contains no spectral information, performs the worst. The DFT Points, Band Energies, MFCCs, and Mel Spectrum feature sets, which contain limited spectral information perform better than the Total Energies feature set, but all have MDEs greater than 2 times the MDEs of the best-performing feature sets for each regression network architecture. The most spectrally-complete feature sets include DFT Magnitude, DFT Phase, DFT Magnitude & Phase, Time Domain, and Complex DFT, all of which display MDEs within a factor of 2 of the best-performing feature sets for each network architecture.

While the performance of these feature sets seems, in general, to be proportional to the amount of spectral information the features carry, more spectral information often means larger feature set sizes, so the models trained on more spectrally-complete feature sets were often also trained on larger amounts of data (in bytes) than the models trained using less spectrally-complete feature sets. This correlation between the amount of spectral information, feature set size, and performance does not always hold, however. The DFT Magnitude feature set, for example, outperforms the DFT Magnitude & Phase, Complex DFT, and Time Domain feature sets for three of four networks, even though it is roughly half the size. Additionally, the DFT Phase feature set underperforms the DFT Magnitude feature set for all four networks, though both sets contain the same number of features. The data suggests that the performance is determined to a larger degree by the amount of spectral information than by feature set size alone.

Within the top-performing feature sets, DFT Phase exhibits the highest MDE for both regression architectures, suggesting that more relevant information is carried by the magnitude portion of the spectrum. Additionally, the feature sets that contain phase information in some form (DFT Magnitude & Phase, Time Domain, and Complex DFT) did not systematically outperform the feature set that did not contain any phase information (DFT Magnitude) across both regression architectures. This suggests that unique information about force location may not be carried by the phase portion of the spectrum.

For regression, the highest-performing feature set/network architecture combination was the FNN using the Complex DFT feature set, which displayed a test MDE of $0.20\,\mathrm{mm}$. The FNN using the Time Domain feature set produced a test MDE of $0.21\,\mathrm{mm}$, which is an increase of less than 10% over the Complex DFT feature set. This result has potential implications on both the hardware constraints and the achievable latency performance of an AAS system, as the feature extraction portion of the process can be eliminated entirely, potentially reducing required processing power while decreasing the time required to return a prediction. Between the two regression architectures, the FNN consistently outperforms the CNN. However, as with the classification architectures where the reverse was true, further testing would be required to verify that conclusion in general once each network architecture has been optimized to achieve maximum accuracy under the test conditions.

The poor performance of the Total Energies feature set may be explained by the empirical analysis in Section 3. The spatial response of each bending mode and the frequency at which it resonates changes when a force is applied to the panel at different locations. Since the Total Energies feature set contains no direct information about the spectrum, and none can be inferred as it can with the Time Domain feature set, the network cannot distinguish between force locations that contain the same energy at the sensor location but have different spectra.

The fact that the DFT points feature set outperforms the band energies feature set for all four networks may seem counter-intuitive as the Band Energies feature set gathers information from a larger portion of the spectrum; the DFT feature set could be compared to a Band Energies feature set with very narrow, rectangular filters. One possible explanation for this result is that the Band Energies feature set condenses the behavior at a broad range of frequencies into a single feature, potentially masking behavior that is captured when a narrower band is used. Imagine, for example, a scenario in which the responses for two different force locations contain resonances at different frequencies. If both frequencies fall within a single band of the Band Energies filter bank (which they almost

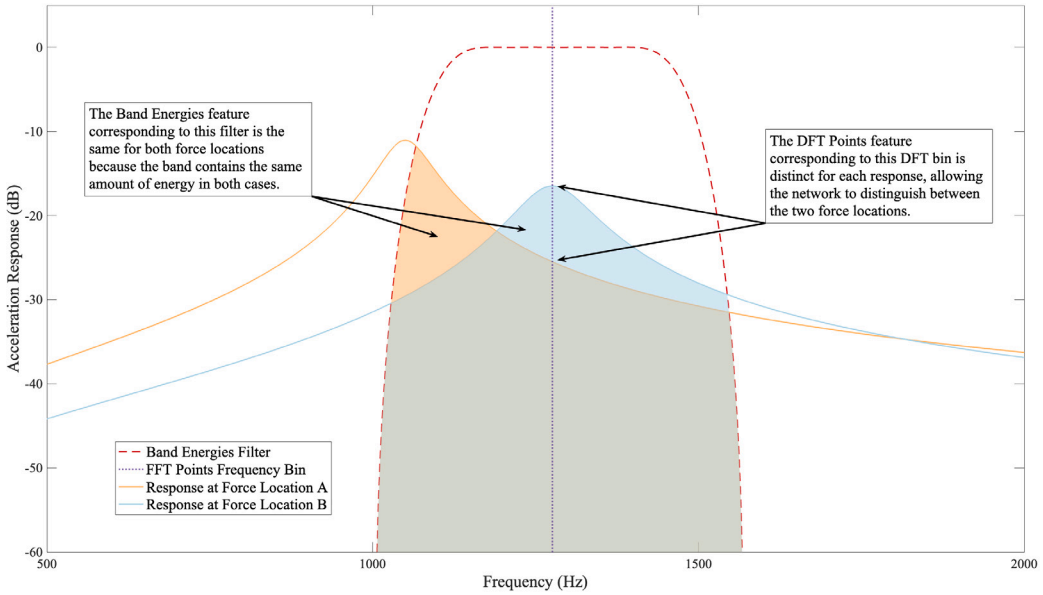

Fig. 8. A scenario in which a Band Energies feature would not reflect the difference in the responses produced by different force locations, whereas the DFT point feature would.

certainly would, given the typical amount of resonant frequency shift and the density of the filter bands) the band energies feature corresponding to that filter band might be the same for both force locations, whereas the DFT Points feature corresponding to a single frequency bin would be different for the two locations, allowing the network to differentiate between them. This scenario is depicted graphically in Fig. 8.

5.1. Error analysis

Since most commercial applications such as virtual buttons do not require sub-millimeter precision, it is important to analyze the accuracy of each feature set within a particular tolerance. For the FNN using the Complex DFT feature set, 99.77% of the test predictions were within 1 mm of the actual location, 99.36% were within 0.75 mm, 96.87% were within 0.5 mm, and 74.20% were within 0.25 mm. To put these results in context, consider the size of a home screen icon in Apple iOS, which is 180 × 180 pixels [29]. On an iPhone 14 with a resolution of 460 pixels per inch, and a native scale factor of 3x, this gives a square icon with a side length of approximately 9.94 mm [30]. The results for each threshold indicate how often the predicted force location must fall within the area of the icon, and thus result in a correctly interpreted input for force locations within a subset of the icon area as shown in Fig. 9. For example, when the threshold is 1 mm, the location of a force anywhere within a 7.94 mm square at the center of a home screen icon will be predicted to fall somewhere on the icon 99.77% of the time.

Additionally, certain touch locations on the panel are more likely to give incorrect predictions than others. The general trend is that the models tend to perform worse when the touch point is near the edges of the panel than they do when the touch point is closer to the center. This is shown in Fig. 10, which shows the MDE for each force location in the same arrangement as the physical locations on the panel. A darker band around the edge, indicating higher MDEs in this area, can be observed in the figure. This behavior is likely a result of the boundary conditions of the panel, which fix nodal lines at the panel edges for every vibrational mode. As discussed in Section 3, the changes in the vibrational behavior of a panel that result from the application of a constant, localized, normal force are most pronounced for a given mode when the force is applied in the antinodal region of that mode, and that vibrational behavior is relatively unaffected by a force applied at or near a nodal line of that mode. A force applied near the edge, then, will have a relatively small effect on the vibrational behavior of the panel, resulting in a response that is similar to the responses for other force locations near the edge. This results in a symmetry that makes it difficult for the model to distinguish between different force locations near the edge of the panel.

It is worth acknowledging that any lower-frequency AAS system requires that a sound be emitted as part of its core functionality. While the prototype system presented here uses a transient click, which is less acoustically obtrusive than a sweep or band-limited noise, this may represent a drawback to using this technology when compared with other touch technologies. One potential solution could be a two-stage sensing system, where the first stage serves to determine whether a touch is present anywhere on the surface. This could be done using a passive approach, which would not produce sound. Only when this stage is triggered will the AAS stage emit a transient excitation to identify the touch point. This signal will not only serve as the excitation required for measuring the vibration response, but it will also act as audible feedback to the user that the system has registered the interaction, which is desirable in many use cases.

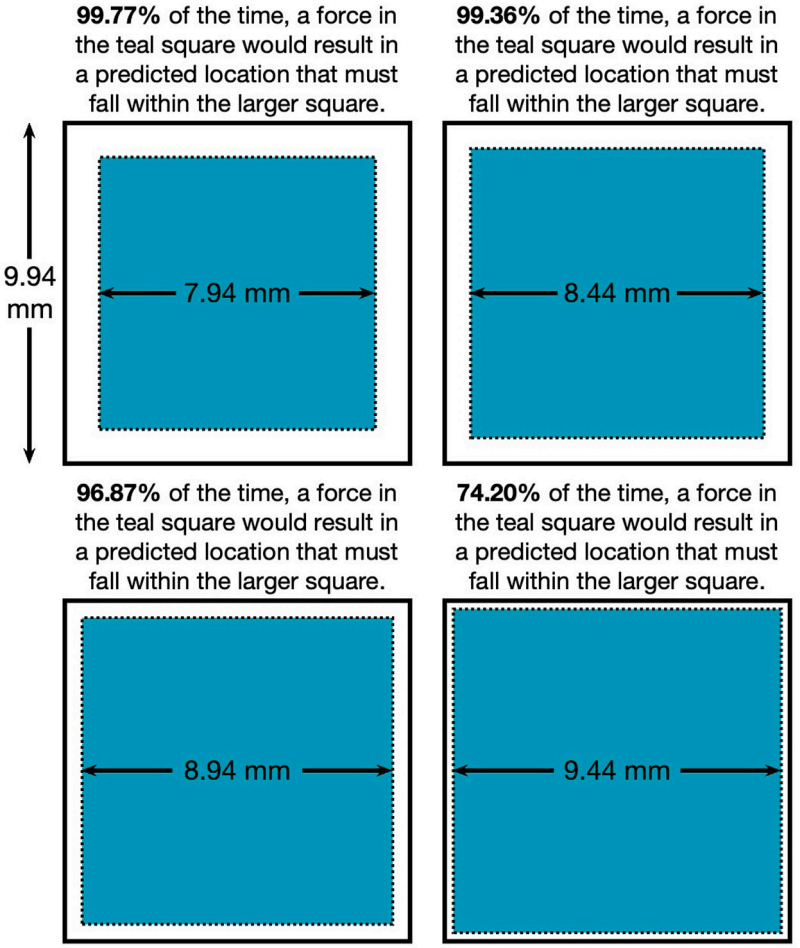

Fig. 9. Distance threshold results of a button rendered on a touchscreen for the FNN regression architecture using the Complex DFT feature set. Each large, 9.94 mm square represents an area the size of a home screen icon on an iPhone 14. An actual force location anywhere within the smaller, teal square would give a predicted force location within the icon area with probabilities indicated by the percentages.

6. Conclusion

A method and prototype design were described for an AAS system that utilizes machine learning to infer touch location on a surface from vibrations as monitored by a structural sensor. The performance of the prototype AAS system was also presented. Building on these results, continuing work to develop a robust AAS touch interface requires further investigations in several areas. First, the testing of the AAS system described here employed a spring-loaded steel stylus to apply force to an acrylic surface; further experiments will need to be carried out to investigate the effects of employing different stylus and surface materials. The human finger, for example, has a stiffness many orders of magnitude lower than steel or acrylic, and it is possible that the optimal surface material for use with a finger may have a smaller Young's modulus than acrylic. Second, these experiments were carried out with a constant touching force magnitude, and further research is required to explore the performance of this system as a function of the applied touch force. The ability to measure the force also provides an additional input degree of freedom which may be useful for some applications such as gaming and virtual musical instrument interfaces. Third, it is important that the vibration response of the system for all touch locations remains constant over time and under different environmental conditions. Further work needs to be done to establish the stability of the current system and investigate construction methods that ensure a consistent response over time. Additionally, further work will also be required to establish the maximum spatial resolution of AAS systems of the sort described in this paper and to quantify their ability to interpolate to touch input locations not included in the original training data. Finally, an assessment of the performance of the system as a function of the amplitude of the driven panel vibrations is needed as well as user testing to assess the active panel vibration and sound levels that users will deem to be acceptable.

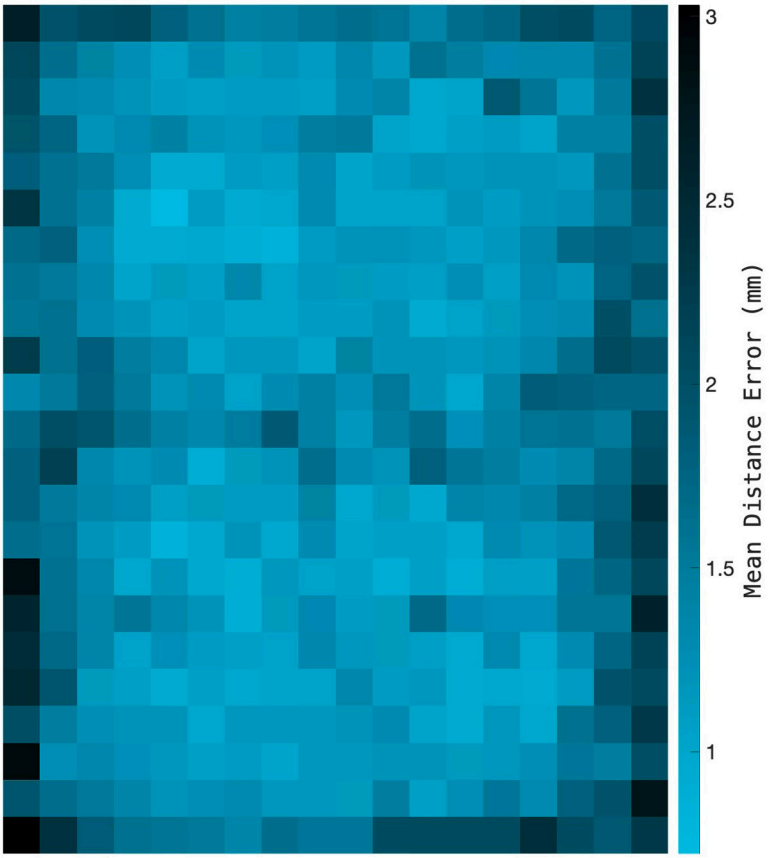

Fig. 10. Mean distance error for each force location for the regression CNN using the Complex DFT feature set. The force locations in the figure are arranged to mirror the physical distribution on the surface of the panel. Notice that the model tends to fail near the edges of the panel.

CRediT authorship contribution statement

Benjamin R. Thompson: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tre DiPassio:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Jenna Rutowski:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Mark F. Bocko:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Michael C. Heilemann:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by National Science Foundation, United States Award 2104758.

Appendix. Feature sets

See Table A.1.

Table A.1
Feature sets.

Feature set	Size[a] (Bytes)	Description	Equation (where β_m is the mth element of a feature array, $\boldsymbol{A}$, of length M)
Total energy	8	The time domain response, $x[n]$, is squared and summed.	$\beta_1 = \sum_n x[n]^2$
DFT points	96	The N-point discrete Fourier transform (DFT), $X[k]$, of the time domain signal, $x[n]$, is computed. The feature set is comprised of the magnitudes of a subset of M DFT bins, k_m, where $M \leq N$. These bins are chosen such that each bin contains the measured resonant frequency of one of the modes of the panel when no external force is applied. For the panel under test, 8 isolated modes were identified with resonant frequencies of 253 Hz, 429 Hz, 560 Hz, 755 Hz, 1043 Hz, 1276 Hz, 1470 Hz, and 1920 Hz.	$\beta_m = \lvert X[k_m] \rvert$
Band energies	96	The time domain signal, $x[n]$, is passed through an M-band filter bank, $h_m[n]$, and the energy in each band is summed. The center frequency of each filter is chosen to correspond to the measured resonant frequency of one of the modes of the panel when no external force is applied. For the panel under test, 8 isolated modes were identified. The resonant frequencies of these modes are the same frequencies used for DFT points, and listed above. The bandwidths for each filter were also chosen to match the bandwidths of the measured resonances of the panel when no external force was applied. For the panel under test, these bandwidths, listed here in the same order as the resonant frequencies, are 32 Hz, 54 Hz, 110 Hz, 162 Hz, 162 Hz, 274 Hz, 426 Hz, and 430 Hz.	$\beta_m = \sum_n (x[n] \circledast h_m[n])^2$
Mel-Frequency Cepstral Coefficients (MFCCs)	104	Following [31], the N-point DFT, $X[k]$, of the time domain signal, $x[n]$, is multiplied by an R-band mel filter bank, $H_r[k]$. The magnitudes of each filtered DFT signal are summed, and a discrete cosine transform (DCT) is performed on the logarithm L of those sums, to generate a feature set of M coefficients. A 40-band triangular filter bank was employed to generate 13 coefficients for this experiment. To extract these features, MATLAB's *mfcc* function was utilized and the details of the filter bank can be found in [32].	$L_r = \log_{10} \sum_k (\lvert X[k] H_r[k] \rvert)$ $\beta_m = \mathrm{DCT}(L)[m]$ for $r = 1:R$, $m = 1:M$
Mel Spectrum	320	The N-point DFT, $X[k]$, of the time domain signal, $x[n]$, is multiplied by an M-band mel filter bank, $H_m[k]$. The magnitudes of each filtered DFT signal are summed to give M features. The same 40-band triangular filter bank used for the MFCCs feature set was employed for this experiment.	$\beta_m = \sum_k (\lvert X[k] H_m[k] \rvert)$ for $m = 1:M$
DFT Magnitude	3208	The magnitude of the one-sided DFT of the length-N time domain response.	$\beta_m = \lvert X[k] \rvert$ for $m = k = 1 : \lfloor \frac{N}{2} \rfloor + 1$
DFT Phase	3208	The wrapped phase angle of the one-sided DFT of the length-N time domain response.	$\beta_m = \angle X[k]$ for $m = k = 1 : \lfloor \frac{N}{2} \rfloor + 1$
DFT Magnitude & Phase	6416	The magnitude and the wrapped phase angle of the one-sided DFT of the length-N time domain response. They are arranged in series; all magnitudes followed by all phases.	$\beta_m = \lvert X[k] \rvert$ $\beta_{m+\lfloor \frac{N}{2} \rfloor + 1} = \angle X[k]$ for $m = k = 1 : \lfloor \frac{N}{2} \rfloor + 1$
Time domain	6400	The unprocessed, length-N, time domain response.	$\beta_m = x[n]$
Complex DFT	6416	The complex-valued, one-sided DFT of the N-length time domain signal. When complex data is used to train the networks, it is split into its real and imaginary components which are treated as separate channels [33].	$\beta_m = X[k]$ for $m = k = 1 : \lfloor \frac{N}{2} \rfloor + 1$

[a] Feature set size is calculated for a single 100 ms time domain response at a sampling frequency of 8 kHz and assuming 64-bit (8-byte), double-precision floating point numbers.

References

[1] G. Walker, Touch technologies tutorial latest version, 2014, https://walkermobile.com/Touch_Technologies_Tutorial_Latest_Version.pdf. (Accessed 28 December 2022).

[2] G. Walker, A review of technologies for sensing contact location on the surface of a display, J. Soc. Inf. Disp. 20 (2012) http://dx.doi.org/10.1002/jsid.100.

[3] G. Walker, Fundamentals of projected-capacitive touch technology, 2014, http://walkermobile.com/SID_2014_Short_Course_S1.pdf. (Accessed 28 December 2022).

[4] G. Barrett, R. Omote, Projected-capacitive touch technology, Inf. Disp. 26 (3) (2010) 16–21, http://dx.doi.org/10.1002/j.2637-496X.2010.tb00229.x.

[5] H. Nam, K.-H. Seol, J. Lee, H. Cho, S.W. Jung, Review of capacitive touchscreen technologies: Overview, research trends, and machine learning approaches, Sensors 21 (14) (2021) 4776, http://dx.doi.org/10.3390/s21144776.

[6] H. Akhtar, Q. Kemao, R. Kakarala, A review of sensing technologies for small and large-scale touch panels, in: Fifth International Conference on Optical and Photonics Engineering, Vol. 10449, 2017, pp. 209–221, http://dx.doi.org/10.1117/12.2270631.
[7] K.-W. Lee, S.-C. Kim, S.-C. Lim, DeepTouch: Enabling touch interaction in underwater environments by learning touch-induced inertial motions, IEEE Sens. J. 22 (9) (2022) 8924–8932, http://dx.doi.org/10.1109/JSEN.2022.3163664.
[8] D.T. Pham, Z. Ji, M. Yang, Z. Wang, M. Al-Kutubi, A novel human-computer interface based on passive acoustic localisation, in: International Conference on Human-Computer Interaction, Springer, 2007, pp. 901–909, http://dx.doi.org/10.1007/978-3-540-73107-8_99.
[9] G. De Sanctis, D. Rovetta, A. Sarti, G. Scarparo, S. Tubaro, Localization of tactile interactions through TDOA analysis: Geometric vs. inversion-based method, in: 2006 14th European Signal Processing Conference, 2006, pp. 1–4.
[10] V.G. Reju, A.W.H. Khong, A.B. Sulaiman, Localization of taps on solid surfaces for human-computer touch interfaces, IEEE T. Multimed. 15 (6) (2013) 1365–1376, http://dx.doi.org/10.1109/TMM.2013.2264656.
[11] N.J. Harris, M. Colloms, Touch sensitive device employing bending wave vibration sensors that detect touch location and provide haptic feedback, 2015, US 2015/9035918B2.
[12] T. Collins, Active acoustic touch interface, Electron. Lett. 45 (20) (2009) 1055–1056, http://dx.doi.org/10.1049/el.2009.2054.
[13] T. Hiraki, M. Fukumoto, Y. Kawahara, Touchable wall: Easy-to-install touch-operated large-screen projection system, in: Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces, ISS '18, 2018, pp. 465–468, http://dx.doi.org/10.1145/3279778.3279916.
[14] M. Ono, B. Shizuki, J. Tanaka, Touch & activate: Adding interactivity to existing objects using active acoustic sensing, in: UIST 2013 - Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, 2013, pp. 31–40, http://dx.doi.org/10.1145/2501988.2501989.
[15] J. Liu, Y. Chen, M. Gruteser, Y. Wang, VibSense: Sensing touches on ubiquitous surfaces through vibration, in: 2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking, SECON, 2017, pp. 1–9, http://dx.doi.org/10.1109/SAHCN.2017.7964907.
[16] D.A. Anderson, Touch-sensing via excitation of a surface with a waveform, 2023, US 2023/11836319B2.
[17] L. Cremer, M. Heckl, Structure-Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies, Springer Science & Business Media, 2013, http://dx.doi.org/10.1007/b137728.
[18] F.J. Fahy, P. Gardonio, Sound and Structural Vibration: Radiation, Transmission and Response, Elsevier, 2007, http://dx.doi.org/10.1016/B978-0-12-373633-8.X5000-5.
[19] C.C. Fuller, S.J. Elliott, P.A. Nelson, Active Control of Vibration, Academic Press, 1996, http://dx.doi.org/10.1016/B978-0-12-269440-0.X5000-6.
[20] G. Rabbiolo, R. Bernhard, F. Milner, Definition of a high-frequency threshold for plates and acoustical spaces, J. Sound Vib. 277 (4–5) (2004) 647–667, http://dx.doi.org/10.1016/j.jsv.2003.09.015.
[21] D.A. Anderson, M.C. Heilemann, M.F. Bocko, Measures of vibrational localization on point-driven flat-panel loudspeakers, Proc. Meet. Acoust. 26 (1) (2016) 065003, http://dx.doi.org/10.1121/2.0000216.
[22] A. Mitchell, C. Hazell, A simple frequency formula for clamped rectangular plates, J. Sound Vib. 118 (2) (1987) 271–281, http://dx.doi.org/10.1016/0022-460X(87)90525-6.
[23] T. DiPassio, M.C. Heilemann, M.F. Bocko, Direction of arrival estimation of an acoustic wave using a single structural vibration sensor, J. Sound Vib. 553 (2023) 117671, http://dx.doi.org/10.1016/j.jsv.2023.117671.
[24] T. DiPassio, M.C. Heilemann, B. Thompson, M.F. Bocko, Estimating acoustic direction of arrival using a single structural sensor on a resonant surface, in: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2023, pp. 1–5, http://dx.doi.org/10.1109/ICASSP49357.2023.10095986.
[25] PCB Piezotronics Inc, Model 352c66 high sensitivity, miniature (2 gm), ceramic shear ICP®accel., 100 mv/g, 0.5 installation and operating manual, 2023, Accessed 1 January 2023.
[26] Mitutoyo, ABSOLUTE digimatic indicator ID-U, 2023, Accessed 30 May 2023.
[27] Y. Liu, J. Nikolovski, N. Mechbal, M. Hafez, M. Vergé, An acoustic multi-touch sensing method using amplitude disturbed ultrasonic wave diffraction patterns, Sensors Actuators A 162 (2) (2010) 394–399, http://dx.doi.org/10.1016/j.sna.2010.03.020.
[28] P. Gavrikov, Visualkeras, 2020, https://github.com/paulgavrikov/visualkeras.
[29] App icons, 2023, https://developer.apple.com/design/human-interface-guidelines/app-icons. (Accessed 25 July 2023).
[30] Iphone 14, 2023, https://www.apple.com/iphone-14/specs/. (Accessed 25 July 2023).
[31] P. Mermelstein, Distance measures for speech recognition, psychological and instrumental, Int. J. Pattern Recognit. 116 (1976) 374–388.
[32] Mfcc: Extract MFCC, log energy, delta, and delta-delta of audio signa, 2024, https://www.mathworks.com/help/audio/ref/mfcc.html. (Accessed 4-June-2024).
[33] Train network with complex-valued data, 2023, https://www.mathworks.com/help/deeplearning/ug/train-network-with-complex-valued-data.html. (Accessed 1 July 2023).