Learning to See Inside Opaque Liquid Containers using Speckle Vibrometry: Supplementary materials

Matan Kichler Shai Bagon Mark Sheinin
Weizmann Institute of Science, Israel
matankic@gmail.com, shai.bagon@weizmann.ac.il, mark.sheinin@weizmann.ac.il

A. Volume effect on sensed vibration's SNR

This experiment shows the relation between the sound volume at the container's surface and the resulting SNR of the recovered vibration. In other words, we aim to assess the minimum threshold amplitude of the excitation at which our system can no longer reliably measure the vibrations. In general, this threshold depends on the container's size and material. To get a ballpark understanding of this relationship, we used a Coke can positioned 50 cm from a single speaker as a reference. Specifically, we played a single tone at various volumes, recovered the container's vibration using our system, and measured the sound level at the surface of the Coke can (using a calibrated microphone).

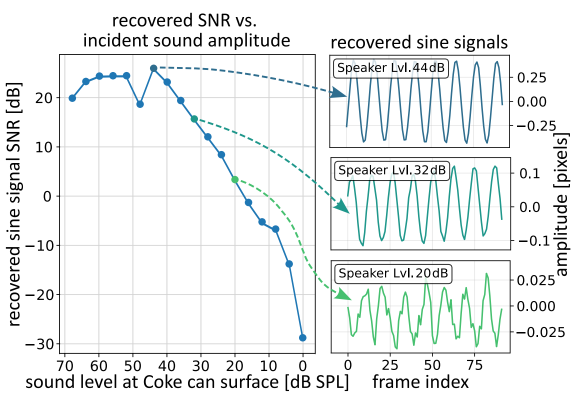


Figure 1. Vibration SNR vs. excitation sound amplitude.

Fig. 1 shows the recovered vibration's SNR as a function of the sound level measured close to the can's surface (in dB SPL units). Let $\mathrm{dB_{rec.}}$ and $\mathrm{dB_{sound}}$ denote the recovered vibration units and the speaker surface level, respectively. The SNR plot remains flat at approximately $25\,\mathrm{dB_{rec.}}$ until the sound level reaches about $40\,\mathrm{dB_{sound}}$ (typical home noise), beyond which the SNR declines, hitting $0\,\mathrm{dB_{SNR}}$ at $17\,\mathrm{dB_{sound}}$ of sound. Remarkably, even at $20\,\mathrm{dB_{sound}}$ (rustling leaves), the signal is still detectable (around $3\,\mathrm{dB_{rec.}}$). The recovered signal SNR was computed using

$$\mathrm{SNR_{dB}} = 20\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right),$$

where P_{signal} is the DFT magnitude of the test tone while P_{noise} is the root sum square of the other frequency components (excluding the DC component).

B. Industrial container experiment

In addition to various everyday drink containers, we evaluated our method on a large industrial water tank weighing about 100 kg when full. We excited the container vibrations using slightly more powerful speakers (see Fig. 5(c) of the main manuscript), which we manually moved to six different locations. To increase the amount of data samples for the single industrial container, we projected the full 6×6 laser grid onto the container's surface and used three columns of points as separate measurements [1]. As with the smaller containers, we recorded vibrations at six discrete fill levels using a chirp, a song segment, and ambient sound. However, no intermediate levels were acquired for this container. Tab. 1 shows the inference result for tests *(a)* and *(d)* on the industrial container. These results come from a model trained for 7500 epochs on the full training set, including all other containers in training but omitting speaker position three in the water-tank data (reserved for the within-distribution test) and excluding ambient-sound recordings.

	Level Pred.		Container
Test name	Acc. ↑	MAE ↓	Acc. ↑
(a) within distribution	1.0000	0.0078 ±0.0108	1.0000
(d) ambient sound	1.0000	0.0226 ±0.0631	0.9444

Table 1. Water-tank experiment.

[1]Each column here had three laser points, in correspondence to the other containers in our dataset

C. Visualizing the inference results

Here we provide a comprehensive breakdown of our evaluation results across all dataset subsets. First, we visualize the inference process on the different test sets in Figs. 2 and 3. In Fig. 2, the rows of each test visualize the predicted logit values for the given test, while the ground truth fill levels are indicated on the y-axis. For example, the second row of the third column in Fig. 2 shows an erroneous classification of 80% level for a true level of 20%. Fig. 3 shows the inference of the liquid level using Eq. (5) of the main paper, which predicts a continuous fill level instead of a discrete one. The predicted continuous liquid level is marked using a red dot, while the true level is marked on the y-axis, as before. Fig. 3(a) shows several successful predictions, while the second row in Fig. 3(b) shows a failure case.

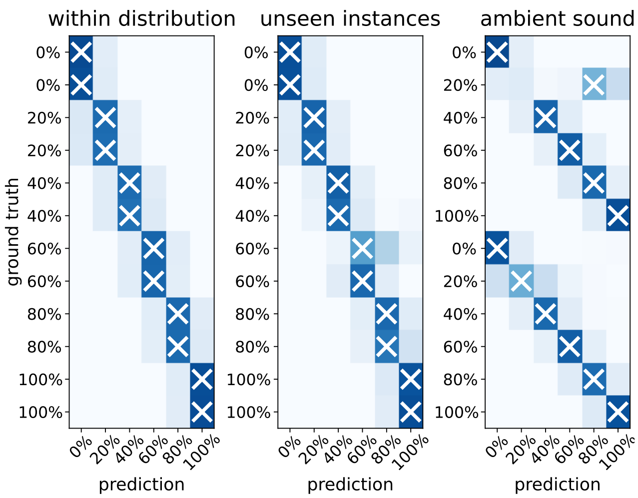
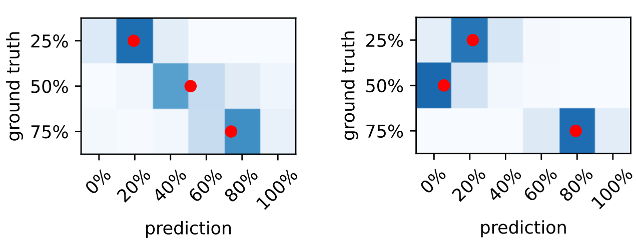


Figure 2. Liquid fill-level prediction for various containers across different test sets: "(a) within-distribution", "(b) unseen instances", and "(d) ambient sound". Each row displays the logits corresponding to a given input. The X sign indicates the argmax of that row.



(a) success case

(b) failure case

Figure 3. Liquid fill-level prediction for test "(c) unseen liquid levels". The MAE for this test set is measured with respect to the expected value of the logits, indicated by a red dot in each row. **(a)** The model approximately manages to interpolate the unseen liquid levels (50 % and 75 %). **(b)** The model fails to predict the 50 % level in the second row (on a different container).

D. Detailed liquid-level inference results

Here we provide tables detailing the results on the inference task across our various test sets. The tables report the mean absolute error (MAE) and accuracy for each container type along with the number of samples in each set. Specifically, Tab. 8 details our training set (3,347 samples); Tab. 2 covers the within-distribution subset (694 samples); Tab. 3 reports the unseen instances subset (549 samples); Tab. 4 presents the unseen liquid levels (486 samples); Tab. 5 shows the ambient sound subset (449 samples); Tab. 6 details unseen liquid levels with ambient sound (54 samples); and Tab. 7 lists the unseen instances with ambient sound (61 samples), for a total of 5,640 samples.

In addition, Tab. 9 separately presents additional results of our ablation study, evaluating the performance of the VibrationTransformer when trained using only the bottom or top measurement points.

Finally, we report the performance of VibrationTransformer trained on variations of the input signal V_i: either by supplementing the magnitude with additional phase information (Tab. 10), or by directly processing the raw complex coefficients of V_i (Tab 11).

Container	#Samp.	Level Pred. Acc.↑	Level Pred. MAE↓	Container Acc.↑
Shampoo Plastic	12	1.0000	0.0077 ±0.0106	1.0000
Coconut Water Carton	12	1.0000	0.0078 ±0.0107	1.0000
Almond Milk Carton	6	1.0000	0.0076 ±0.0106	1.0000
Tomato Juice Carton	6	1.0000	0.0077 ±0.0107	1.0000
Ananas Juice Carton	6	1.0000	0.0077 ±0.0106	1.0000
Rice Milk Carton	12	1.0000	0.0078 ±0.0107	1.0000
Energy Drink	12	1.0000	0.0079 ±0.0109	1.0000
Short Beer Can	36	1.0000	0.0078 ±0.0107	1.0000
Coke Can	72	1.0000	0.0078 ±0.0107	1.0000
Tall Beer Can	12	1.0000	0.0078 ±0.0106	1.0000
Pineapple Nectar Tin	24	1.0000	0.0078 ±0.0107	1.0000
Green Tea Plastic	24	1.0000	0.0078 ±0.0107	1.0000
Oil Tin Can	6	1.0000	0.0077 ±0.0107	1.0000
Silver Vacuum flask	12	0.9167	0.0234 ±0.0449	1.0000
Black Vacuum flask	6	1.0000	0.0078 ±0.0109	1.0000
Champagne Glass	12	0.7500	0.0586 ±0.0717	1.0000
Orange Vacuum flask	36	1.0000	0.0078 ±0.0106	1.0000
Pitcher Vacuum flask	6	1.0000	0.0079 ±0.0108	1.0000
Conditioner	120	1.0000	0.0078 ±0.0107	1.0000
Oatly	60	1.0000	0.0078 ±0.0107	1.0000
Contigo	72	1.0000	0.0078 ±0.0107	1.0000
Matte Vacuum Flask	130	0.9385	0.0233 ±0.0526	1.0000
Water Tank	36	1.0000	0.0078 ±0.0106	1.0000
Total:	730	0.9836	0.0116 ±0.0279	1.0000

Table 2. Detailed results for the "(a) within-distribution" test set. The second column shows the number of samples in the set for each container.

Container	#Samp.	Level Pred. Acc.↑	Level Pred. MAE↓	Container Acc.↑
Coke Can	108	1.0000	0.0077 ±0.0106	1.0000
Conditioner	108	1.0000	0.0078 ±0.0108	1.0000
Oatly	108	0.6574	0.1287 ±0.2103	0.7870
Contigo	108	0.8426	0.0623 ±0.1575	0.9907
Matte Vacuum Flask	117	0.4786	0.2392 ±0.2565	0.9829
Total:	549	0.7905	0.0916 ±0.1886	0.9526

Table 3. Detailed results for the "(b) unseen instances" test set. The second column shows the number of samples in the set for each container.

Container	#Samp.	Level Pred.		Container
		Acc.↑	MAE↓	Acc.↑
Shampoo Plastic	27	N/A	0.1176 ±0.1054	0.8148
Coconut Water Carton	27	N/A	0.1813 ±0.1027	0.8148
Almond Milk Carton	27	N/A	0.0773 ±0.0338	1.0000
Tomato Juice Carton	27	N/A	0.0927 ±0.0745	0.8889
Ananas Juice Carton	27	N/A	0.0810 ±0.0351	0.8519
Rice Milk Carton	27	N/A	0.0770 ±0.0478	0.9259
Energy Drink	27	N/A	0.1585 ±0.1596	0.3704
Short Beer Can	27	N/A	0.1163 ±0.0594	0.6296
Coke Can	27	N/A	0.0775 ±0.0460	0.9259
Tall Beer Can	27	N/A	0.1569 ±0.1402	0.5926
Pineapple Nectar Tin	27	N/A	0.1560 ±0.1099	0.7778
Green Tea Plastic	27	N/A	0.1054 ±0.0578	0.6667
Oil Tin Can	27	N/A	0.1117 ±0.0428	0.3704
Silver Vacuum flask	27	N/A	0.0709 ±0.0249	1.0000
Black Vacuum flask	27	N/A	0.1621 ±0.2024	1.0000
Champagne Glass	27	N/A	0.1316 ±0.0767	1.0000
Orange Vacuum flask	27	N/A	0.1575 ±0.2157	1.0000
Pitcher Vacuum flask	27	N/A	0.0683 ±0.0314	1.0000
Total:	486	N/A	0.1167 ±0.1104	0.8128

Table 4. Detailed results for the "(c) unseen liquid levels" test set. The second column shows the number of samples in the set for each container.

Container	#Samp.	Level Pred.		Container
		Acc.↑	MAE↓	Acc.↑
Shampoo Plastic	12	1.0000	0.0077 ±0.0104	1.0000
Coconut Water Carton	12	1.0000	0.0140 ±0.0207	1.0000
Almond Milk Carton	6	1.0000	0.0191 ±0.0249	1.0000
Tomato Juice Carton	6	1.0000	0.0081 ±0.0104	1.0000
Ananas Juice Carton	6	0.8333	0.0411 ±0.0717	0.6667
Rice Milk Carton	6	0.6667	0.0740 ±0.0880	1.0000
Energy Drink	6	1.0000	0.0078 ±0.0107	1.0000
Short Beer Can	18	1.0000	0.0091 ±0.0111	1.0000
Coke Can	72	1.0000	0.0078 ±0.0107	1.0000
Tall Beer Can	6	1.0000	0.0078 ±0.0107	0.8333
Pineapple Nectar Tin	12	1.0000	0.0079 ±0.0107	0.9167
Green Tea Plastic	12	1.0000	0.0128 ±0.0178	1.0000
Oil Tin Can	6	1.0000	0.0075 ±0.0105	1.0000
Silver Vacuum flask	12	0.7500	0.0709 ±0.1419	1.0000
Black Vacuum flask	6	1.0000	0.0078 ±0.0108	1.0000
Champagne Glass	6	0.3333	0.2011 ±0.2635	0.8333
Orange Vacuum flask	18	1.0000	0.0077 ±0.0107	1.0000
Pitcher Vacuum flask	6	1.0000	0.0078 ±0.0106	1.0000
Conditioner	60	0.9167	0.0208 ±0.0476	0.9667
Oatly	60	0.9500	0.0282 ±0.0858	0.9333
Contigo	36	1.0000	0.0077 ±0.0106	1.0000
Matte Vacuum Flask	65	0.6769	0.1483 ±0.2468	0.9846
Water Tank	18	1.0000	0.0132 ±0.0229	0.8889
Total:	467	0.9165	0.0377 ±0.1183	0.9700

Table 5. Detailed results for the "(d) ambient sound" test set. The second column shows the number of samples in the set for each container.

Container	#Samp.	Level Pred. Acc.↑	MAE↓	Container Acc.↑
Shampoo Plastic	3	N/A	0.1003 ±0.0410	0.6667
Coconut Water Carton	3	N/A	0.1922 ±0.1414	0.6667
Almond Milk Carton	3	N/A	0.2302 ±0.2271	0.0000
Tomato Juice Carton	3	N/A	0.1978 ±0.1087	0.6667
Ananas Juice Carton	3	N/A	0.0970 ±0.0445	0.6667
Rice Milk Carton	3	N/A	0.0822 ±0.0479	1.0000
Energy Drink	3	N/A	0.2534 ±0.2142	0.0000
Short Beer Can	3	N/A	0.1149 ±0.0893	0.3333
Coke Can	3	N/A	0.0612 ±0.0290	1.0000
Tall Beer Can	3	N/A	0.1564 ±0.0739	0.6667
Pineapple Nectar Tin	3	N/A	0.2923 ±0.1893	0.3333
Green Tea Plastic	3	N/A	0.1712 ±0.1095	0.6667
Oil Tin Can	3	N/A	0.1310 ±0.0600	0.6667
Silver Vacuum flask	3	N/A	0.0673 ±0.0241	1.0000
Black Vacuum flask	3	N/A	0.1051 ±0.0469	1.0000
Champagne Glass	3	N/A	0.2679 ±0.2879	0.6667
Orange Vacuum flask	3	N/A	0.1664 ±0.1558	1.0000
Pitcher Vacuum flask	3	N/A	0.0953 ±0.0369	1.0000
Total:	54	N/A	0.1546 ±0.1503	0.6667

Table 6. Detailed results for the "(e) unseen liquid levels + ambient sound" test set. The second column shows the number of samples in the set for each container.

Container	#Samp.	Level Pred. Acc.↑	MAE↓	Container Acc.↑
Coke Can	12	1.0000	0.0077 ±0.0105	1.0000
Conditioner	12	0.8333	0.0701 ±0.0923	0.7500
Oatly	12	0.3333	0.1964 ±0.2003	0.0833
Contigo	12	0.4167	0.1415 ±0.1319	1.0000
Matte Vacuum Flask	13	0.3846	0.3604 ±0.3276	1.0000
Total:	61	0.5902	0.1586 ±0.2275	0.7705

Table 7. Detailed results for the "(f) unseen instances + ambient sound" test set. The second column shows the number of samples in the set for each container.

Container	#Samp.	Level Pred. Acc.↑	Level Pred. MAE↓	Container Acc.↑
Shampoo Plastic	96	1.0000	0.0077 ±0.0106	1.0000
Coconut Water Carton	96	1.0000	0.0078 ±0.0107	1.0000
Almond Milk Carton	48	1.0000	0.0076 ±0.0106	1.0000
Tomato Juice Carton	48	1.0000	0.0078 ±0.0107	1.0000
Ananas Juice Carton	48	1.0000	0.0077 ±0.0106	1.0000
Rice Milk Carton	42	1.0000	0.0078 ±0.0107	1.0000
Energy Drink	42	1.0000	0.0078 ±0.0108	1.0000
Short Beer Can	126	1.0000	0.0078 ±0.0107	1.0000
Coke Can	576	1.0000	0.0078 ±0.0107	1.0000
Tall Beer Can	42	1.0000	0.0079 ±0.0107	1.0000
Pineapple Nectar Tin	84	1.0000	0.0078 ±0.0107	1.0000
Green Tea Plastic	84	1.0000	0.0078 ±0.0107	1.0000
Oil Tin Can	48	1.0000	0.0077 ±0.0108	1.0000
Silver Vacuum flask	96	1.0000	0.0077 ±0.0106	1.0000
Black Vacuum flask	48	1.0000	0.0078 ±0.0109	1.0000
Champagne Glass	42	1.0000	0.0077 ±0.0108	1.0000
Orange Vacuum flask	126	1.0000	0.0078 ±0.0107	1.0000
Pitcher Vacuum flask	48	1.0000	0.0079 ±0.0108	1.0000
Conditioner	420	1.0000	0.0078 ±0.0107	1.0000
Oatly	480	1.0000	0.0078 ±0.0107	1.0000
Contigo	252	1.0000	0.0078 ±0.0107	1.0000
Matte Vacuum Flask	455	1.0000	0.0090 ±0.0110	1.0000
Water Tank	180	1.0000	0.0078 ±0.0106	1.0000
Total:	3527	1.0000	0.0079 ±0.0107	1.0000

Table 8. Evaluating the model on the training set. Our VibrationTransformer is expressive enough to fit the training data perfectly. The second column shows the number of samples in the set for each container.

Test Name	Full Model	Single Point Bottom	Single Point Center	Single Point Top
(a) within distribution	0.02 ±0.05	0.02 ±0.05	0.03 ±0.07	0.03 ±0.07
(b) unseen instances	0.09 ±0.15	0.12 ±0.18	0.11 ±0.19	0.14 ±0.21
(f) unseen instances + ambient sound	0.16 ±0.21	0.14 ±0.16	0.18 ±0.21	0.23 ±0.25

Table 9. Ablation study on liquid level prediction using MAE (chance ≈ 0.30). The full model is compared against variants using only a single measurement (bottom, center, or top). Results are reported for tests (a) within distribution, (b) unseen instances, and (f) unseen instances + ambient sound, demonstrating increased difficulty with each added challenge factor.

Test name	Level Pred. Acc. ↑	Level Pred. MAE ↓	Container Acc. ↑
within distribution	0.9251	0.0315 ±0.0802	0.9957
unseen instances	0.7377	0.0909 ±0.1546	0.9781
unseen liq. levels	N/A	0.1373 ±0.1314	0.7757
ambient sound	0.6949	0.1046 ±0.1631	0.8241
unseen liq. levels + ambient sound	N/A	0.1554 ±0.1137	0.4630
unseen instances + ambient sound	0.4426	0.2024 ±0.2033	0.8033

Table 10. Ablating for adding phase information. Instead of using only the *magnitude* of V_i, we trained a version of our network on a concatenation of magnitude and phase of V_i, resulting in patches of dimension 400 linearly projected to 512 tokens. Results show that adding phase information did not significantly contribute to the performance of the model.

Test name	Level Pred. Acc. ↑	Level Pred. MAE ↓	Container Acc. ↑
within distribution	0.7911	0.0899 ±0.1705	0.9438
unseen instances	0.7814	0.0989 ±0.1926	0.9107
unseen liq. levels	N/A	0.1930 ±0.1771	0.6214
ambient sound	0.4009	0.2371 ±0.2320	0.4922
unseen liq. levels + ambient sound	N/A	0.1970 ±0.1384	0.2222
unseen instances + ambient sound	0.3934	0.2258 ±0.2225	0.4918

Table 11. Ablating for working on the raw complex Fourier coefficients. We tested whether our PointTransformer can directly reason with the raw complex V_i. We trained a version of our network with a linear projection of the complex 200-dimension patches to real 512-dimensional tokens. Results show that the PointTransformer struggles to extract meaningful information from the raw complex signal.