

More than noise: assessing the viscosity of food products based on sound emission

Dominik Schiller, Silvan Mertes, Marcel Achzet, Fabio Hellmann, Ruben Schlagowski, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Schiller, Dominik, Silvan Mertes, Marcel Achzet, Fabio Hellmann, Ruben Schlagowski, and Elisabeth André. 2024. "More than noise: assessing the viscosity of food products based on sound emission." In *Deep Learning Theory and Applications: 5th International Conference, DeLTA 2024, Dijon, France, July 10–11, 2024, proceedings, part I*, edited by Ana Fred,ALLEL Hadjali, Oleg Gusikhin, and Carlo Sansone, 210–18. Cham: Springer.
https://doi.org/10.1007/978-3-031-66694-0_13.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

Deutsches Urheberrecht

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



More Than Noise: Assessing the Viscosity of Food Products Based on Sound Emission

Dominik Schiller¹(✉)() , Silvan Mertes¹() , Marcel Achzet², Fabio Hellmann¹() ,
Ruben Schlagowski¹() , and Elisabeth André¹()

¹ Chair for Human-Centered Artificial Intelligence, Augsburg University, Augsburg, Germany
{dominik.schiller,silvan.mertes,fabio.hellmann,
ruben.schlagowski,elisabeth.andre}@uni-a.de

² Chair for Mechanical Engineering, Augsburg University, Augsburg, Germany
marcel.achzet@mrmm.uni-augsburg.de

Abstract. In the era of Industry 4.0, manufacturing is rapidly shifting towards automation, particularly in processes such as quality control, production lines, and logistics. However, the food industry poses distinctive challenges to automation due to the variability in raw materials and stringent hygiene standards. Sensory analysis is crucial in maintaining consistent quality and safety while manufacturing food products. This paper focuses on the automatic estimation of viscosity, a key parameter in many industry quality aspects of food products. An often overlooked aspect is the potential correlation between viscosity and sound emissions. While conventional methods for determining viscosity require expensive equipment, this research investigates the possibility of analyzing the acoustic emission when a liquid is sucked through a vacuum pump to determine the viscosity. By simulating industry-like food products with varying viscosity through different flour and water mixtures, we aim to investigate the feasibility of developing an automatic, deep-learning-based system for real-time viscosity estimation in manufacturing processes. Our results indicate that our proposed methodology can automatically determine the difference in viscosity, showing the feasibility of using sound emission analysis as a tool for viscosity estimation in manufacturing processes.

Keywords: Machine learning · Industry 4.0 · Viscosity estimation

1 Introduction

With the advent of *Industry 4.0*, manufacturing processes, be it quality control, production lines, or logistics, are undergoing a massive shift towards automation. The food industry presents unique challenges for automation, not only due to the significant variability in raw materials but also because of strict hygiene standards that must be carefully adhered to ensure consistent quality and safety for the end products. An integral component in this context is the sensory analysis of the products during the manufacturing process. This allows for replication and precise control of taste, texture, and visual appearance, ensuring consistent quality of the food products. A key characteristic for a plenitude of such products is *viscosity* - a parameter that can influence the

result's texture, mouthfeel, and look. However, analyzing a product's viscosity commonly uses highly specialized devices like viscometers, rheometers, or consistometers. These devices are costly to buy and time-intensive, making a continuous online analysis of the manufacturing process nearly impossible. However, a matter that is rarely considered is that viscosity correlates with specific characteristics related to fluid dynamics and can also affect the sound emission of a liquid substance during manufacturing. Imagine using a straw to draw water from the surface of your drinking glass. The sound produced differs from other substances, such as honey, which is more viscous. In this paper, we investigate if we can use this property to build an automatic, deep-learning-based system that can determine a liquid's viscosity by analyzing the sound that is produced when being sucked through a vacuum suction device. To this end, we simulate industrial-like food products with different viscosities by mixing flour and water in different predefined ratios.

2 Related Work

2.1 Viscosity Detection

Rheology is a crucial field of study in the food industry as it deals with substance flow, deformation, and consistency. Viscosity is a fundamental rheological parameter that characterizes a fluid's resistance to flow or deformation [4, 16]. One can conceptualize viscosity by considering a fluid element subjected to shear stress τ at its top while the bottom is held in place. Newton's law of viscosity now states a proportionality between shear stress and the induced velocity gradient $\frac{\partial v}{\partial y}$ resulting in

$$\tau = \nu \frac{\partial v}{\partial y}, \quad (1)$$

with the proportionality factor ν defined as the viscosity [12]. To measure the viscosity of the fluid and semi-fluid matter, there are a variety of devices based on different working principles and methods [4]. Typical devices used in the food industry are capillary viscometers, falling-ball viscometers, and rotational or oscillatory rheometers using the proportionality of the applied force and deformation of the sample to derive its rheological properties such as viscosity [14]. However, the diverse variety of these properties within food products requires carefully selecting an appropriate method for confident viscosity measurements. [4] provides an overview, describing the commonly used devices as mentioned and viscosity estimations based on models and correlations. Next to conventional methods for viscosity measurements, various techniques use sound waves, such as ultrasonic-type viscometers. Their non-destructive and non-invasive nature makes them well-suited for in-line measurements, and their low cost and energy consumption make them a popular choice for monitoring systems across manufacturing processes [5, 11].

2.2 Sound-Based Process Monitoring in Food Industry

Sound waves' acoustic properties strongly depend on the carrier medium's rheological properties. Consequently, the propagation of sound waves within a given sample,

whether induced through active or passive excitation, correlates closely with its viscous properties and enables the measurement of the sample's viscosity. Thus, ultrasound has been widely used for analysis and quality control of food products [3, 8, 11] by directly measuring the speed of sound, attenuation, or reflection coefficients. Other approaches correlate viscosity measurements to features extracted from the recorded ultrasonic data [1] or use long short-term memory neural networks combined with feature extraction to monitor fermentation, cleaning and mixing processes [5]. These methods rely on coupling ultrasound into the sample using external ultrasonic transducers. Goksu et al. [9] introduced an approach that involves analyzing sound emissions generated, i.e., by vibrations of the machinery and material during the process. To this end, they relied on wavelet packet analysis of complex sound emissions from fluid flow in household pipe systems to predict the flow speed, while [7] combines ultrasonic measurements and machine learning techniques to accurately determine the presence of fouling in food manufacturing equipment.

These examples illustrate the wide application of ultrasound and sound emission analysis in process monitoring through appropriate signal processing methods. Specifically, they show the potential of active ultrasound for viscosity detection. Nonetheless, its implementation is frequently impeded by challenges related to integrating requisite hardware into the process. Thus, parameters such as the resonance frequency and positioning of the transducer directly affect measurement quality [4, 8]. To avoid these issues, it is practical to use sound emission signals produced by the process to extract the substance's rheological properties under examination. To the authors' knowledge, there is no reliable method to predict the viscosity of matter by sound emission analysis in the audible range (2 Hz–20 kHz).

3 Approach

Deep neural networks are the de-facto-standard for sound classification tasks. Convolutional neural networks (CNNs) that operate on the frequency domain (e.g., spectrograms) to automatically learn a meaningful representation of the input data have been shown to yield superior results over more traditional approaches involving handcrafted features [2, 13, 15]. However, to train such models large amounts of annotated data are necessary. Given the lack of available research for the specific task at hand (see Sect. 2, and the associated lack of available training data, we recorded a new dataset.

3.1 Data Recording

The greatest challenge in creating such a dataset is accurately determining the ground truth (i.e., the viscosity). As described in Sect. 1, the viscosity determination of a fluid substance usually requires highly specialized and often costly equipment. We, therefore, decided to simulate food products of various viscosities by mixing flour with water in predefined ratios. The mixture with the highest viscosity consists of 60% flour, while the other end of the spectrum is represented by pure water without flour. A second aspect is the choice of the audio recording hardware and setup, which can ideally be replicated in

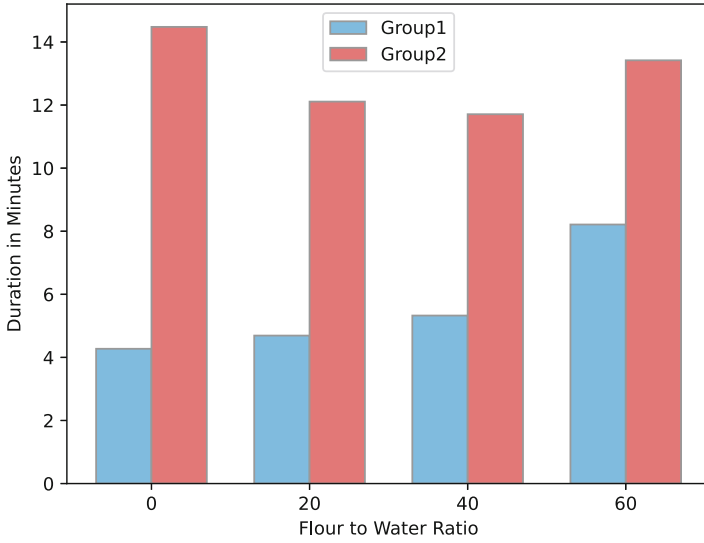


Fig. 1. Class-wise sample distribution of data recordings.

an industrial application. Depending on the actual employment scenario, the positioning options of the microphones might be very restricted. To consider this limitation, we record the vacuum sound with three microphones from various distances and directions. We use two overhead Rode NT-5 and one Sennheiser MKE 600 directed microphones. All devices have been specially selected for their relatively neutral frequency response curve, meaning no specific frequency ranges are emphasized or attenuated. The frequency patterns generated during the suction process differ depending on several execution parameters. For example, the speed of execution or the degree of immersion of the nozzle leads to audible changes in the recorded signals. We split the recording procedure into two groups to account for this variance in the data recording process. Each group consisted of three participants who carried out the suction process using a mixture of fixed and randomized microphone positions. The audio has been recorded using a sampling rate of 16 kHz. Overall, both groups recorded roughly 75 min of audio. Figure 1 shows the class-wise distribution of samples for both recording groups (Fig. 2).

3.2 Model and Training

To address the challenge of building a model that can predict the viscosity of a liquid mixture by the sound emission of the fluid when sucked through a suction device, we approach the task in two different ways. First, we model the problem as a four-class classification problem, where each viscosity ratio is represented as a categorical label expressing the respective flour-to-water ratio in percent (i.e., ‘0’, ‘20’, ‘40’, and ‘60’). For the second approach, we exploit the labels’ ordinality to model it as a regression problem, allowing us to assess the viscosity on a continuous scale.

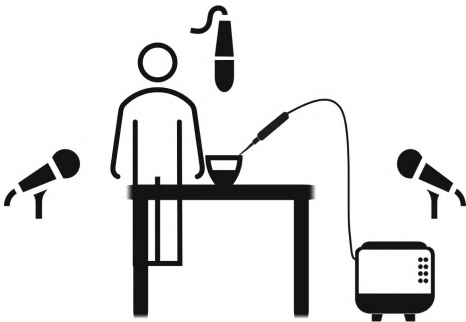


Fig. 2. Schematic overview of the data recording setup.

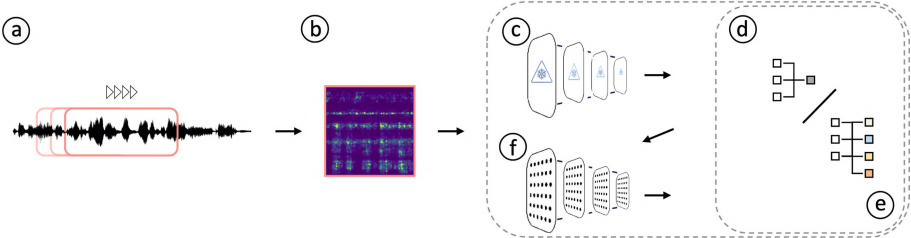


Fig. 3. Schematics of the processing pipeline. A sliding window processes the audio signal (a). From each sliding window, a mel spectrogram is computed (b) and then fed through a convolutional neural network with frozen weights (c). The training target is either formulated as a regression problem (d) or a classification problem (e). After the initial training, the CNN weights are unfrozen and the training process is repeated (f).

As mentioned before, our proposed approach utilizes on the application of CNNs, to automatically learn meaningful features from the input signal. To this end, we rely on the MobileNetV3 [10] architecture. The MobileNet architectures are a family of CNNs optimized for resource efficiency and execution speed instead of maximized predictive performance. We chose this architecture specifically to meet the requirements of our industrial use case, where the detection of an anomaly may require immediate action. Furthermore, the low demand for computing resources means that MobileNet can be easily installed on mobile hardware, resulting in a versatile and cost-effective deployment method. We employ the MobileNetV3 *large* variant for our pipeline, pre-trained on the imagenet dataset [6]. To process the input, we first chunk the audio recordings using a one-second sliding window with a 40-millisecond stride. We then extract a mel-spectrogram from each window and resize it to the input size required by MobileNetV3. The processing pipeline is depicted in Fig. 3. To train the network, we employ a standard two-step finetuning procedure. In the first step, we freeze all the model’s backbone layers and train only the classification head for 20 epochs. Subsequently, we continue training the whole model for another 100 epochs. Finally, we chose the best model for validation loss (regression task) and accuracy (classification task). To assess the predictive capabilities of our model, we employ a Two-Fold Validation procedure, using

one group for training and the other for testing. From now on, we will refer to folds using the terminology *fold G1* and *fold G2*, where the number indicates the respective group used for training. To avoid data leakage between sets and to prevent our model from unintentionally adapting to a specific group of experimenters, we further split the training data into a train and a validation partition. The validation partition is then used for hyperparameter tuning and model selection.

3.3 Results

To compare the models, we use the unweighted average accuracy metric. To this end, the predictions of the models trained on the regression task are binned into the categories of the classification task based on the numerical proximity between prediction and class value. We observe that both models trained on fold G1 and trained on fold G2 are performing equally well on their respective test sets with an unweighted average accuracy score of 0.81 . However, when looking at the test results and predicting unseen data from the other experiment groups, we see a sharp drop in performance. While the model trained on the data from group one still manages to achieve an accuracy score of 0.64 , the model trained on group 2 only manages to achieve a score of 0.58 .

Both models trained on the regression task are performing worse than their respective classification counterparts with 0.69 validation accuracy for fold G1 and a score 0.74 for fold G2. We can notice a similar drop between validation and test set in both cases. For fold G1, the test performance drops to 0.58 while the model trained on the data from group 2 achieves 0.49 .

Table 1. Accuracy scores for the classification and the regression model on the respective validation and test sets.

Fold	Accuracy		
	Setting	Val	Test
G1	CLS	0.81	0.64
	REG	0.69	0.58
G2	CLS	0.81	0.58
	REG	0.74	0.49

The results of all experiments are shown in Table 1.

The results of the respective confusion matrices are depicted in Fig. 4.

Figure 5 shows the boxplots for each regression model. For both plots, we observe that the median lines of all boxes are well-aligned between the models. We can also see that the median lines for the 20 and 40% ratios accurately align with the target values. At the same time, the median lines for the lower extreme of 0 are too high, and the median for the upper bound of 60 is a bit too low for both models. Overall, the median lines exhibit a staircase-like pattern, which shows that the models pick up the overall trend of viscosity. For the upper and lower quartiles of the boxes, we find that the



Fig. 4. Confusion matrices for all folds and tasks.

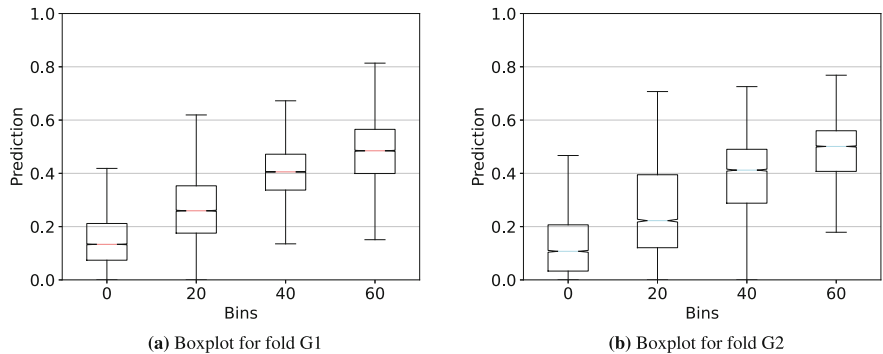


Fig. 5. Boxplot for the regression task.

neighboring categories are building the upper and lower bounds. This means that half of the predictions do not reach values higher than the next label or lower than the previous label. However, when we analyze the whiskers, we can see that the extreme values do not generally exceed the neighboring classes. This is in line with the results from the

classification models, where the confusion matrix shows that wrongly predicted classes are more likely to be the neighboring class of the actual label than more distant classes. While the predictions for both models exhibit a certain amount of noise, the overall trend is observable.

3.4 Discussion

The results of our experiments convincingly demonstrate the potential to derive viscosity characteristics from audio data emitted from a fluid substance sucked in by a suction device. However, it is crucial to address the implications associated with our setup. Since we used a controlled laboratory environment for data collection, things might play out differently in a real-world scenario. On the one hand, we largely exclude possible audio interference factors in our data through consistent audio settings. This would differ in a large-scale industry setting, where production lines commonly consist of various other machines and devices, resulting in a diverse soundscape. On the other hand, creating our data set was time-consuming due to the controlled laboratory conditions and the associated manual work, so the data recorded is comparatively small. In a natural setting, product throughput might be very high, opening up the possibility of also collecting large datasets for training—although annotating that training data might still involve traditional viscosity measurement techniques.

In addition, transferring our approach to industry requires further consideration of the specific use cases. An essential consideration lies in the distinction between Newtonian and non-Newtonian fluids. In contrast to non-Newtonian fluids, Newtonian fluids exhibit a linear correlation between the viscous stresses resulting from the flow and the local strain rate. While water is a classic example of a Newtonian fluid, mixtures of water and flour deviate from this behavior. It is crucial to acknowledge that factors such as the diameter of the suction device or its suction power can, depending on the liquid, interfere in unexpected ways with the viscosity of the material during suction - thereby exerting substantial effects on the emitted audio signals. When extending our approach to more complex food production scenarios, these interfering factors must be carefully considered.

4 Conclusion and Future Work

In this work, we investigated whether it is possible to determine a liquid's viscosity by analyzing the sound emissions during vacuum pump suction. To this end, we created an artificial dataset by mixing water with flour in predetermined ratios and then recording the sound signal during suction. We then trained two deep convolutional neural networks on the Mel spectrograms of the individual sound samples to estimate the viscosity from the recorded sound emissions of the suction process. For the first setting, we formulated the task at hand as a classification problem, with the training target of categorizing the given sound sample into a fixed viscosity class. For the second training, we also experimented with a regression model, aiming to predict the continuous viscosity value. Both approaches show promising results and demonstrate the feasibility of the task and the methodology employed. In the future, we would like to expand the data

set with further finer viscosity gradations. Also, given the continued advances in hardware computing capabilities, we believe further investigation of the online prediction capabilities of larger models is worthwhile.

References

1. Achzet, M., Schaller, B., Schlech, T., Linscheid, F.F., Sause, M.G.: Ultraschallbasierte untersuchung rheologischer eigenschaften von polymeren für die prozessüberwachung (2023)
2. Amiriparian, S., et al.: Snore sound classification using image-based deep spectrum features. In: Interspeech 2017, pp. 3512–3516. ISCA (2017)
3. Awad, T., Moharram, H., Shaltout, O., Asker, D., Youssef, M.: Applications of ultrasound in analysis, processing and quality control of food: a review. *Food Res. Int.* **48**(2), 410–427 (2012)
4. Bhattad, A.: Review on viscosity measurement: devices, methods and models. *J. Therm. Anal. Calorim.* **148**, 6527–6543 (2023)
5. Bowler, A., Pound, M., Watson, N.: Convolutional feature extraction for process monitoring using ultrasonic sensors. *Comput. Chem. Eng.* **155**, 107508 (2021)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
7. Escrig, J.E., Simeone, A., Woolley, E., Rangappa, S., Rady, A., Watson, N.: Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes. *Food Bioprod. Process.* **123**, 1–13 (2020)
8. Franco, E.E., Buiocchi, F.: Ultrasonic measurement of viscosity: signal processing methodologies. *Ultrasonics* **91**, 213–219 (2019)
9. Göksu, H.: Flow measurement by wavelet packet analysis of sound emissions. *Meas. Contr.* **51**(3–4), 104–112 (2018)
10. Howard, A., et al.: Searching for mobilenetv3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1314–1324 (2019)
11. Khairi, M.T.M., Ibrahim, S., Yunus, M.A.M., Faramarzi, M.: Contact and non-contact ultrasonic measurement in the food industry: a review. *Meas. Sci. Technol.* **27**(1), 012001 (2015)
12. Papanastasiou, T., Georgiou, G., Alexandrou, A.N.: *Viscous Fluid Flow*. CRC Press (2021)
13. Piczak, K.J.: Environmental sound classification with convolutional neural networks. In: 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6. IEEE (2015)
14. Tabilo-Munizaga, G., Barbosa-Cánovas, G.V.: Rheology for the food industry. *J. Food Eng.* **67**(1–2), 147–156 (2005)
15. Wagner, J., Schiller, D., Seiderer, A., André, E.: Deep learning in paralinguistic recognition tasks: are hand-crafted features still relevant? In: Interspeech 2018 (2018)
16. White, F.M.: *Fluid Mechanics*, 8th edn. The McGraw Hill Companies, New York, USA (2015)