future internet

MDPI

Article

Snack Texture Estimation System Using a Simple Equipment and Neural Network Model †

Shigeru Kato *, Naoki Wada, Ryuji Ito, Takaya Shiozaki, Yudai Nishiyama and Tomomichi Kagawa

Niihama College, National Institute of Technology, Niihama City, Ehime Prefecture 792-8580, Japan; wada@ele.niihama-nct.ac.jp (N.W.); ri.ei.nnct17@gmail.com (R.I.); sozktky.4096@gmail.com (T.S.); ny.15.nnct@gmail.com (Y.N.); kagawa@ele.niihama-nct.ac.jp (T.K.)

* Correspondence: skatou@ele.niihama-nct.ac.jp; Tel.: +81-897-37-7862

† This paper is a revised and expanded version of a paper entitled "Texture Estimation System of Snacks Using Neural Network Considering Sound and Load" presented at The 13th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2018), Taichung, Taiwan, 27–29 October 2018.

Received: 28 December 2018; Accepted: 4 March 2019; Published: 8 March 2019

check for updates

Abstract: Texture evaluation is manually performed in general, and such analytical tasks can get cumbersome. In this regard, a neural network model is employed in this study. This paper describes a system that can estimate the food texture of snacks. The system comprises a simple equipment unit and an artificial neural network model. The equipment simultaneously examines the load and sound when a snack is pressed. The neural network model analyzes the load change and sound signals and then outputs a numerical value within the range (0,1) to express the level of textures such as "crunchiness" and "crispness". Experimental results validate the model's capacity to output moderate texture values of the snacks. In addition, we applied the convolutional neural network (CNN) model to classify snacks and the capability of the CNN model for texture estimation is discussed.

Keywords: food texture; neural network; human sensibility; artificial intelligence; CNN

1. Introduction

Food texture is a typical sensory experience influencing an individual's preference or distaste of a food product. The relationship between sensory perception and texture, described by the adjectives such as "crispy", "crunchy", and "crackly", has been analyzed previously [1]. According to Hayakawa [2], there are 445 terms related to texture in the Japanese language. For example, "crunchy" is equivalently represented by several onomatopoeic words, including "Kali-Kali" or "Boli-Boli". Japanese people might be comparatively sensitive to the texture of foods.

In Japan, snack texture is typically examined by human inspectors and quality evaluators of manufacturing firms. In the process, sensory evaluation is based on the person's sensibility. In cases of discrepancy, an artificial Intelligence (AI) might be supportive in finalizing the decision of an evaluation task. For example, given differing evaluations of the human inspectors, the AI can propose objective numerical values in the form of "crispness = 0.8" or "crunchiness = 0.5". With such criteria, an agreement would be met among the inspectors, thus alleviating their burden in the course. AI-based systems, such as a neural network model, are potentially capable of learning from a large amount of data, while simultaneously inferring moderate estimation; more available data provide a more precise estimation. Therefore, such intelligent systems could be effective aids in the management and inspection of food texture quality.

Research on automatic food-texture-estimation methods has been conducted previously. Sakurai et al. [3–5] proposed a method for texture diagnosis, which analyzes the sound of a sharp metal probe stabbing the food for texture estimation. Liu and Tan [6] applied a neural network model to evaluate the "crispness" score of snacks by magnifying the crushing sound with a microphone; they crushed the snacks using a pair of pliers by hand and inputted the produced sound signal features to the neural network model. Moreover, Srisawas and Jindal [7] used a neural network to estimate the grade of snacks by estimating "crispness", the snack was manually crushed with a pair of pincers, and then the resulting sound was inputted to the neural network.

While several food-texture-estimation studies [3–7] inferred texture considering only sound, humans tend to evaluate texture by also considering the load on their teeth. The load signal is as equally important as sound in texture estimation. We therefore developed novel equipment capable of examining both signals simultaneously. We applied a neural network model for numerical texture level inference for vegetables such as cucumbers and radishes, in terms of "munch-ness" and "crunchiness" [8].

Similarly, Okada and Nakamoto [9] developed a human-tooth-imitating-sensor that could sense the vibration and load on the tooth, complemented by a recurrent neural network model, which inferred the numerical classification value of the snacks or the sweets into "biscuits", "gummy candy", or "corn snack". Conversely, our study does not aim to categorize snacks but to quantify the texture level of their "crunchiness" or "crispness" within the numerical range (0,1), and develop simple and durable equipment and an intelligent model for the texture level estimation.

The load sensor used in our previous equipment [8] was very narrow and fragile, as shown in Figure 1a, and could accommodate only small food specimens. The current paper developed equipment built from the ground up [10]. Figure 1b illustrates the system, in which the load sensor is attached to the top of the probe for examining differently sized food specimens. Rheometers are commonly used in food manufacturing companies for measuring the force response of the food [11]; nevertheless, rheometers are not built with a sound sensor, unlike our equipment. Therefore, we could regard our equipment as a next-generation rheometer capable of observing load change and sound signals simultaneously. The equipment is very simple, durable, user-friendly, and inexpensive. The novelty of the current study is emphasized on the fusion of our original simple equipment and the artificial neural network model. In particular, the proposed model considers both the sound and load.

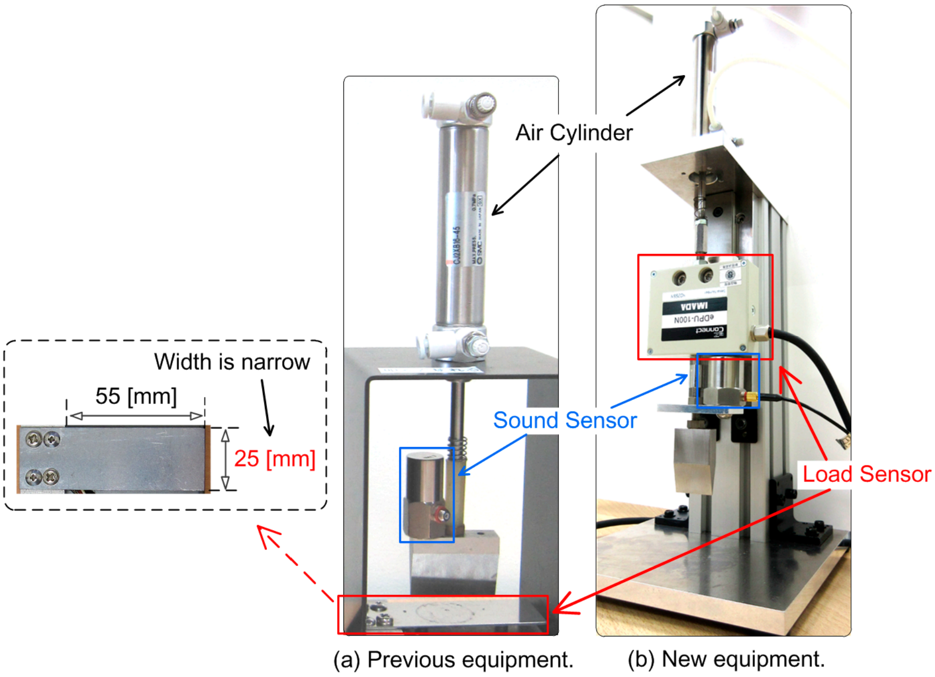


Figure 1. Equipment for food texture evaluation: (**a**) Previous equipment. The plate of the load sensor is small; (**b**) proposed equipment. The load sensor is above the probe, thereby accommodating differently sized food specimens.

We first constructed a simple system using a neural network model that can quantify texture into numerical levels, e.g., “crispness = 0.8 or crunchiness = 0.5”. If such quantification is realized, food texture evaluation can be elevated to a merchandise level, and further development can be accomplished. We conducted an experiment to validate the efficiency of our proposed simple system.

As this study focuses on sound analysis, we investigated recent studies on sound signal processing using AI techniques, and then discovered several studies on audio feature analysis using a spectrogram. A spectrogram image obtained by short-time Fourier transform contains rich information regarding sound characteristics [12–14]; for this, a convolution neural network (CNN) [15] is employed to classify the sound from the inputted spectrogram. Justin and Juan [16] addressed the classification of an environmental sound using CNN, into which a spectrogram-like image (mel-spectrogram) is inputted. CNN is useful for texture analysis of biometrics such as finger prints, palm texture [17], and the iris [18] to identify persons.

Shervin et al. [19] suggested the application of CNN in distinguishing commercial scenes from main video contents, considering video slide images and spectrogram images. Likewise, Shawn et al. [20] classify the musical performance scenes from video stream by CNN, which deals with performance scene images and spectrograms of audio. Such an interesting cross-modal method is possible in our study by combining the sound spectrogram and load change curve images. Afterward, these images will be processed by CNN for texture estimation. This study also addressed applying CNN for classifying snacks. The CNN analyzes an image that comprises the spectrogram of sound and visualized load intensity with color gradation. The capability of CNN for texture estimation is discussed.

2. System Structure

Figure 2 shows the system construction. In Figure 2a, the equipment comprises an air cylinder that moves the metal probe up and down; under the flat and round metal probe is the food sample, i.e., potato chips. The air cylinder moves the probe up and down once it gains air pressure. On the other hand, the load sensor is a load cell fixed between the probe and the air cylinder rod, while the sound sensor is fixed on the metal probe. The system structure is shown in Figure 2b. Signals from the sound and load sensors are amplified and transmitted to the computer via a data acquisition device. As noise is not filtered, the experiment should be performed in a quiet environment. The computer calculates input parameters of the neural network model: W1–W5 and F1–F5 express the characteristics of the load change and the sound features, respectively. The model then outputs the texture level range of (0,1) for “crunchiness” and “crispness”. Here, “crunchy” texture refers to a feeling with a certain load accompanied by a loud sound, while “crispy” texture is the feeling with small load accompanied by high-frequency sound.

Food viscosity or elasticity are measured by a rheometer [11], which is generally used to measure only the force response of the food through an electrical motor that moves the probe. In our proposed equipment, we focus on measuring sound and relate it to “crunchiness” or “crispness” of the food specimen. For this reason, we employ the air cylinder instead of the electrical motor, which causes mechanical noise.

As shown in Figure 3, if such an intelligent system is connected to the internet, then a user holding the equipment can obtain texture information from the server quickly due to the neural network model trained by a big amount of data.

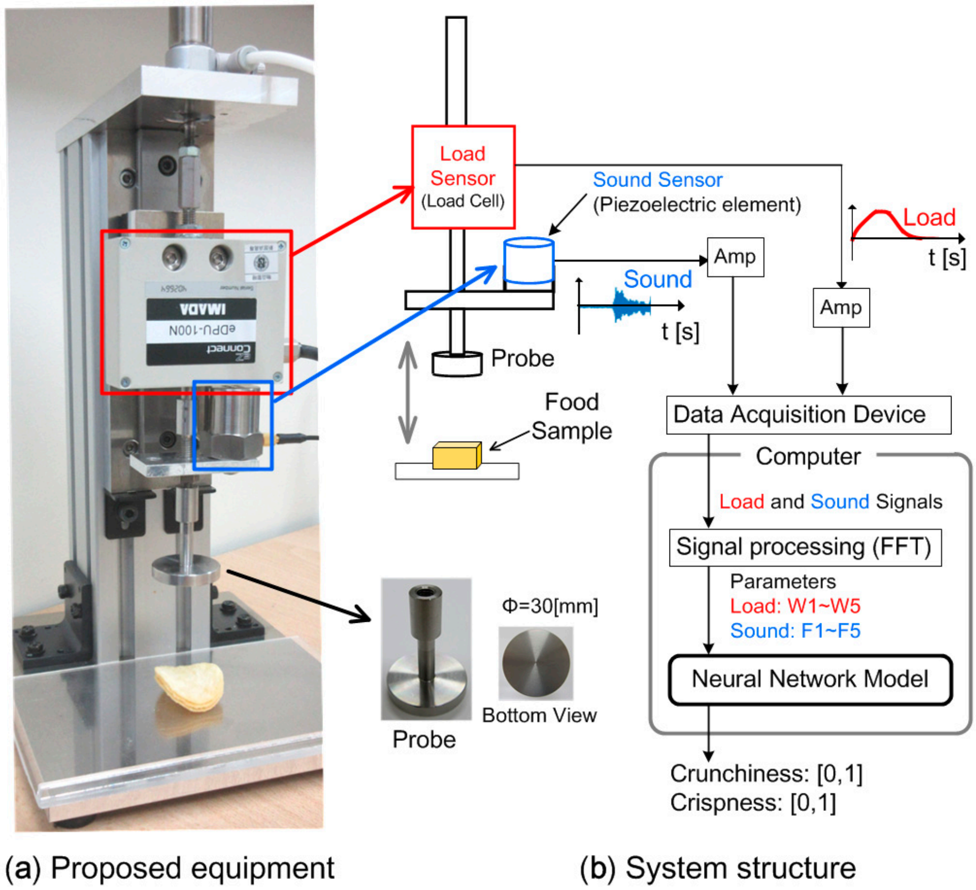


Figure 2. System construction: (**a**) Image of the proposed equipment and; (**b**) system structure for estimation of "crunchiness" and "crispness" textures.

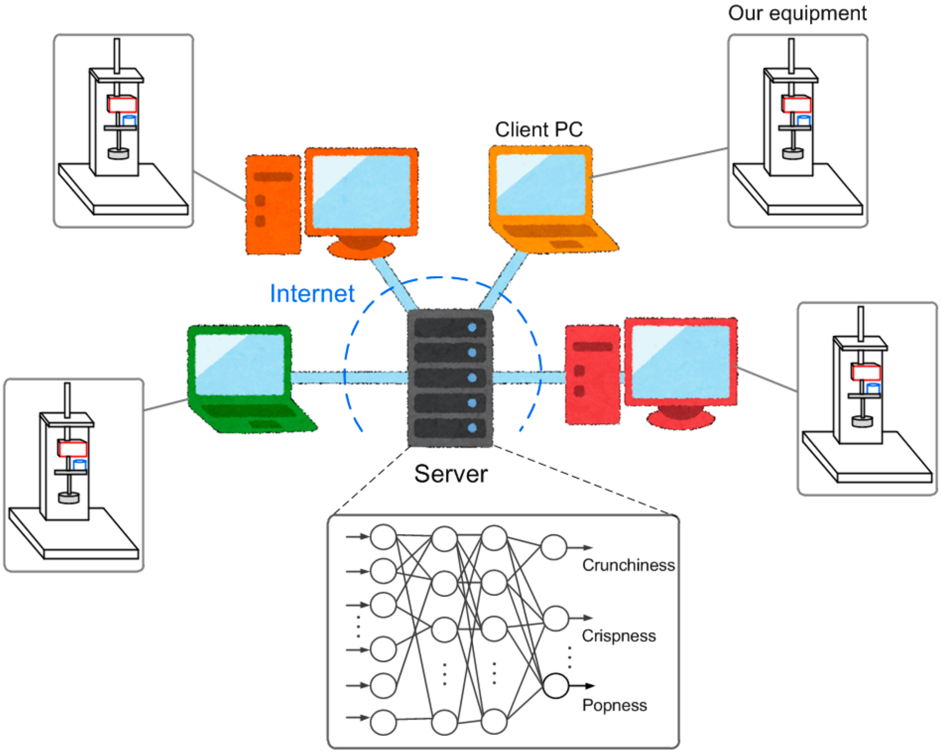


Figure 3. Schematic of the proposed system connected to the internet.

3. Experiment

Rice crackers, potato chips, cookies, biscuits, and corn snacks purchased from a local supermarket were the food specimens evaluated by the system (Figure 4).

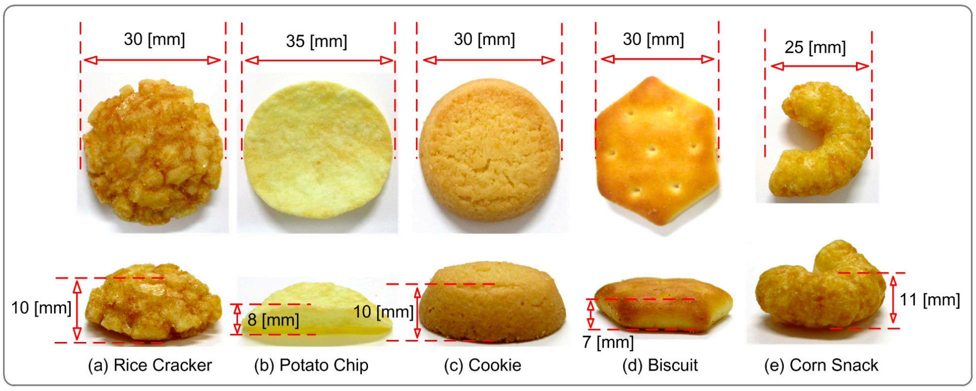


Figure 4. Food specimens: (**a**) A popular rice cracker in Japan; (**b**) smaller sized potato chip compared with ordinary merchandise; (**c**) a butter cookie; (**d**) small sized biscuit and; (**e**) a corn snack.

Table 1 shows the number of samples and texture information of each food specimen after the experiment. Specifically, 200 rice crackers, 200 potato chips, 200 cookies, 200 biscuits, and 200 corn snacks were sampled. Table 2 enumerates the conditions for the load and sound evaluation of the food samples.

Table 1. Number and texture information of the snack samples.

	Rice Crackers	Potato Chips	Cookies	Biscuits	Corn Snacks
Number of Samples	200	200	200	200	200
Sample Number	No.1–200	No.201–400	No.401–600	No.601–800	No.801–1000
Crunchiness	0.9	0.2	0.7	0.5	0.4
Crispness	0.8	0.9	0.7	0.7	0.8

Table 2. Texture evaluation conditions for the samples.

Parameter	Value/Condition
Cylinder air pressure	0.4 [MPa]
Temperature	19~23 [°C]
Humidity	26~32 [%]
Weather	Fine
Sampling rate	25 [k Samples/s]
Probe speed	12 [mm/s]

Figure 5a shows the image of three stacked potato chips being crushed by the equipment; this was carried out 200 times, leading to 200 data for the potato chips samples.

The experiment of each sample was conducted 200 times, for a total of 200 data. The samples are numbered accordingly, as illustrated in Table 1. Figure 6 displays graphs of five specimens of the experiment. The graph at the top illustrates the curve of the load (red line) and the sound (blue line), as shown in Figure 6a, which indicates that as the probe touched the sample, the load increased and a loud sound occurred. The middle graph in Figure 6a shows automatically extracted signals for 2.0 s; the extraction method is explained in the following paragraph. By focusing on the 2.0 s period while the snack is being pressed, we do not have to consider the noise influence except at the 2.0 s period. The bottom graph in Figure 6a shows the FFT (Fast Fourier Transform) results of the extracted 2.0 s sound data. Likewise, the results for the potato chips, cookie, biscuit, and corn snack samples are displayed in Figure 6b–e, respectively.

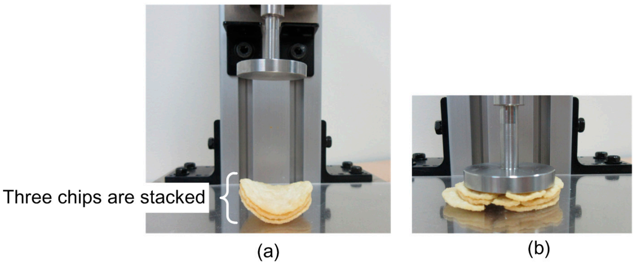
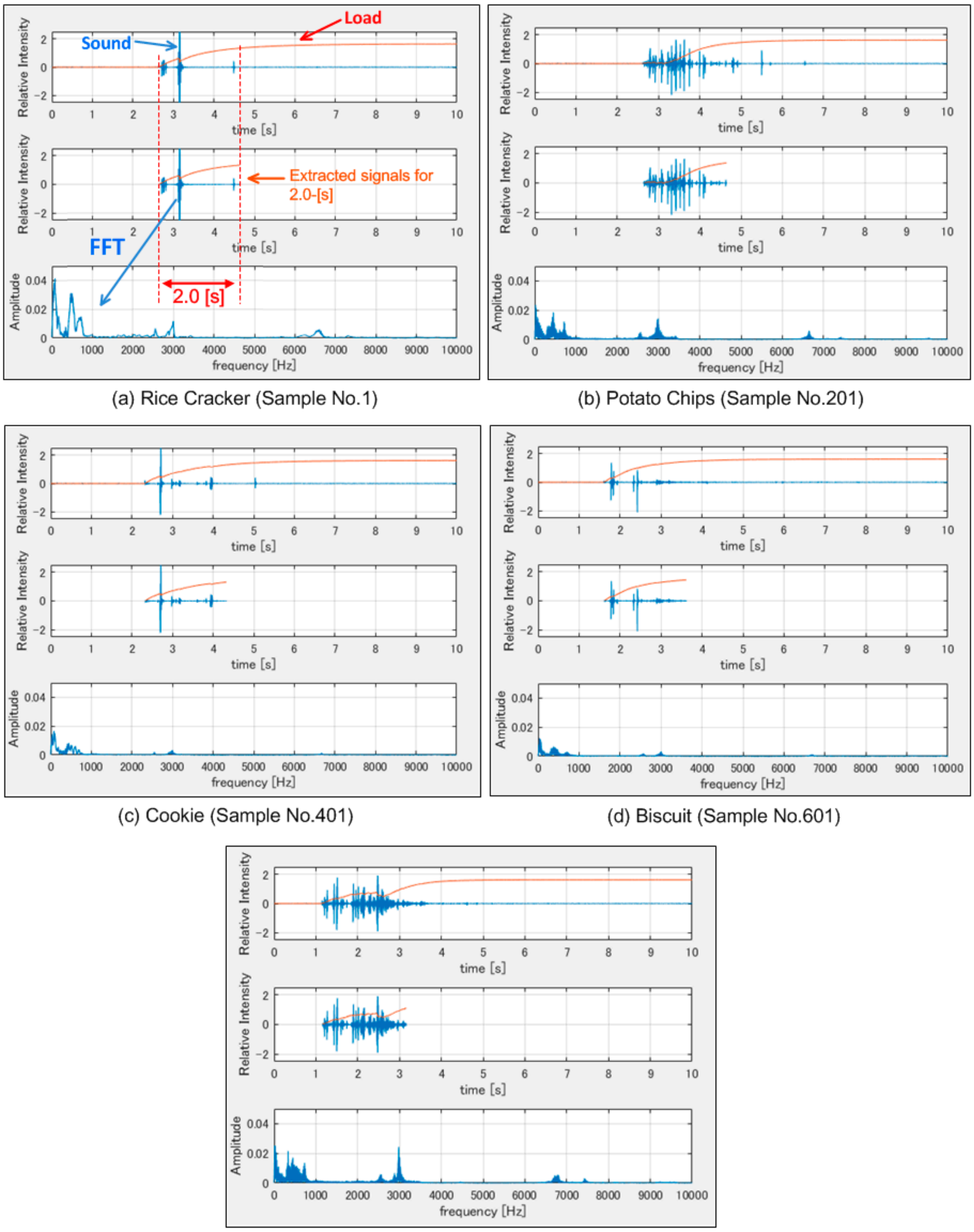


Figure 5. Images of stacked potato chips under equipment evaluation: (**a**) Three stacked potato chips prior to evaluation and; (**b**) potato chips being crushed.



(a) Rice Cracker (Sample No.1)

(b) Potato Chips (Sample No.201)

(c) Cookie (Sample No.401)

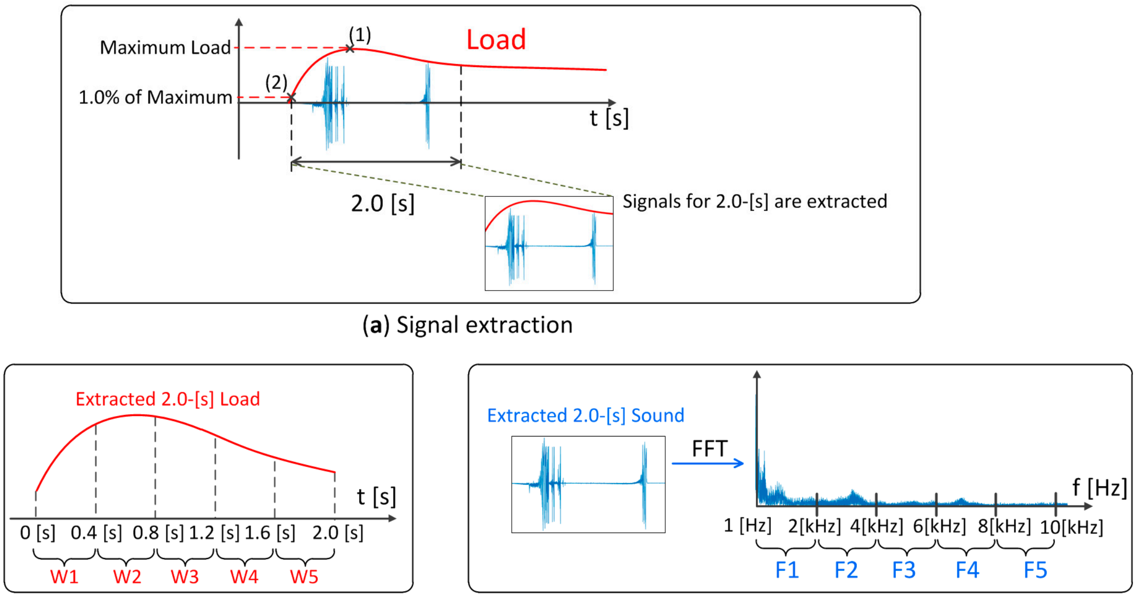
(d) Biscuit (Sample No.601)

(e) Corn Snack (Sample No.801)

Figure 6. Graphs of texture examination of all sorts of samples: (**a**) Rice cracker; (**b**) potato chips; (**c**) cookie; (**d**) biscuit and; (**e**) corn snack.

The signal data for 2.0 s were extracted from the entire 10 s data as shown in Figure 7a.

(i) The maximum load point (1) was determined.
(ii) Point, at 1.0 % of maximum load, was identified.
(iii) Finally, signals for 2.0 s from point (2) were extracted.



(**a**) Signal extraction

(**b**) Parameters in Load

(**c**) Parameters in Sound

Figure 7. Signal Processing: (**a**) 2.0 s signal extraction; (**b**) parameters W1–W5 and; (**c**) parameters F1–F5.

The load curve of the extracted 2.0 s signals was divided into five sections, as shown in Figure 7b. Subsequently, parameters W1–W5 in the load were calculated as follows:

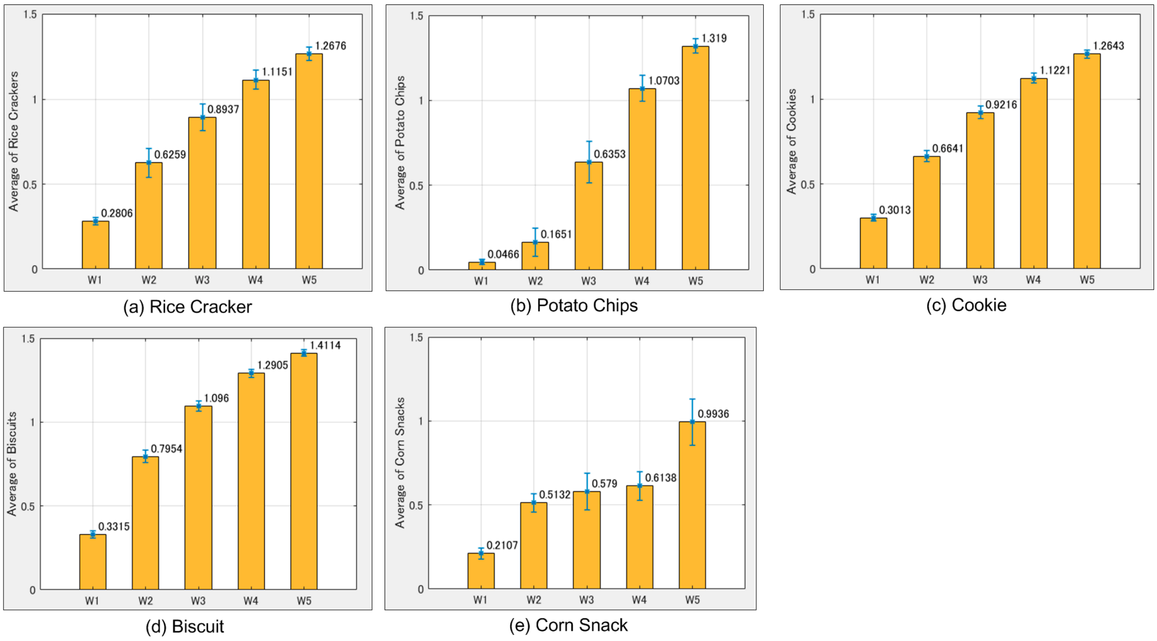
- W1 is the average of the load between 0.0 s and 0.4 s.
- W2 is the average between 0.4 s and 0.8 s.
- W3 is the average between 0.8 s and 1.2 s.
- W4 is the average between 1.2 s and 1.6 s.
- W5 is the average between 1.6 s and 2.0 s.

The extracted 2.0 s sound signal data were converted by FFT, resulting in a frequency range of 1–10 kHz, which was also divided into five sections, as shown in Figure 7c. F1–F5 in the sound were calculated as follows:

- F1 is the integration of FFT result between 1 Hz and 2 kHz.
- F2 is the integration of results between 2 kHz and 4 kHz.
- F3 is the integration of results between 4 kHz and 6 kHz.
- F4 is the integration of results between 6 kHz and 8 kHz.
- F5 is the integration of results between 8 kHz and 10 kHz.

Figure 8 shows the average values and standard deviations (STDs) of the parameters W1–W5 for (a) rice cracker, (b) potato chips, (c) cookie, (d) biscuit, and (e) corn snack, respectively. As observed, averages of W were different depending on each specimen.

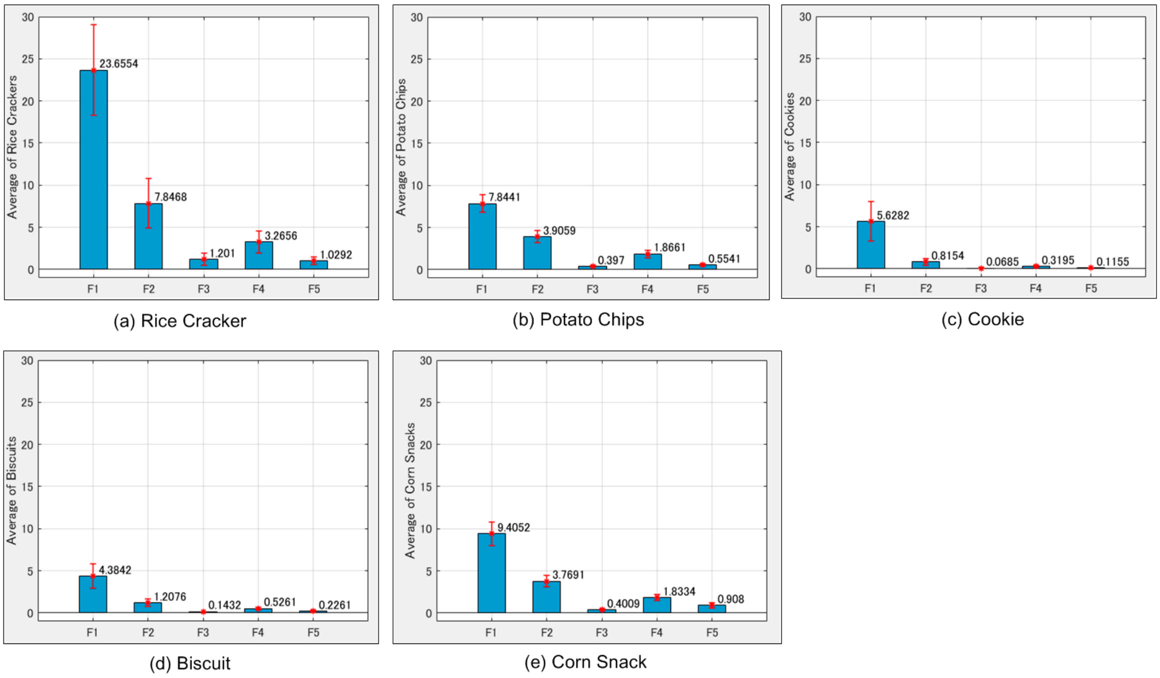
Figure 9 shows the average values and STDs of F1–F5 for the (a) the rice cracker, (b) potato chips, (c) cookie, (d) biscuit, and (e) corn snack, respectively. Likewise, F1–F5 differed in the specimens.



(a) Rice Cracker (b) Potato Chips (c) Cookie

(d) Biscuit (e) Corn Snack

Figure 8. The average and standard deviations (STDs) of W1–W5 for (**a**) the rice cracker; (**b**) potato chips; (**c**) cookie; (**d**) biscuit and; (**e**) corn snack.



(a) Rice Cracker (b) Potato Chips (c) Cookie

(d) Biscuit (e) Corn Snack

Figure 9. Averages and STDs of F1–F5 for (**a**) the rice cracker; (**b**) potato chips; (**c**) cookie; (**d**) biscuit and; (**e**) corn snack.

The STDs showed that the parameters characterized different types of specimens, with respect to the form, size, and density. For texture estimation, we employed the neural network model. In a conventional texture analysis, multiple regression is employed as enormous sample data are analyzed [21]; determining characteristics related to a target texture is a very complicated task. The neural network model simplifies this task and works well with the parameters W1–W5, and F1–F5 is obtained by a very simple calculation.

4. Neural Network Model

The neural network model for texture degree estimation is shown in Figure 10. The input layer comprises 10 nodes for W1–W5 and F1–F5, and a bias node. Hidden layers 1 and 2 comprise 10 nodes and a bias node, respectively. The output layer comprises two nodes expressing the degree range (0,1) of "crunchiness" and "crispness".

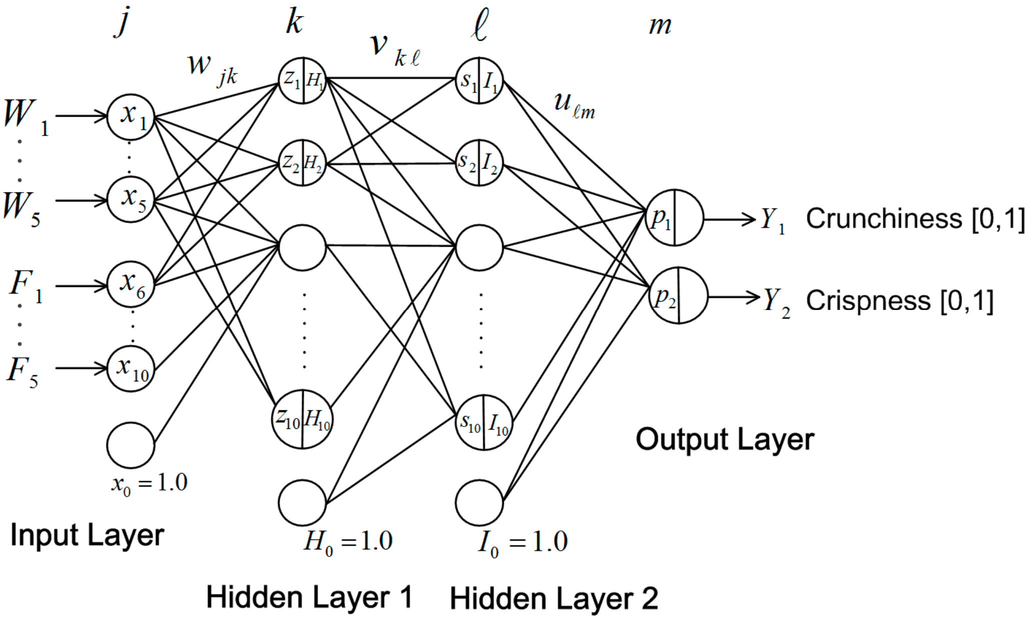


Figure 10. A neural network model for estimating food texture.

The transfer function of hidden layers 1, 2, and the output layer is expressed in Equations (1)–(3), respectively, where x_j is j-th input node value, H_k is the output value of k-th node of hidden layer 1, I_ℓ is the output value of the ℓ-th node of hidden layer 2, and Y_1 and Y_2 are the model outputs:

$$H_k = \frac{1}{1+\exp(-z_k)},\ z_k = \sum_{j=0}^{10} x_j w_{jk},\ x_0 = 1.0 \tag{1}$$

$$I_\ell = \frac{1}{1+\exp(-s_\ell)},\ s_\ell = \sum_{k=0}^{10} H_k v_{k\ell},\ H_0 = 1.0 \tag{2}$$

$$Y_m = \frac{1}{1+\exp(-p_m)},\ p_m = \sum_{\ell=0}^{10} I_\ell u_{\ell m},\ I_0 = 1.0 \tag{3}$$

where w, v and u are the connection weights between the input layer and hidden layer 1, between hidden layers 1 and 2, and between hidden layer 2 and the output layer, respectively. The connection weights were adjusted to minimize the difference (i.e., error) between the expected value and the actual neural network output Y_m via the back-propagation algorithm [22].

For a limited amount of data, such as in our case, the neural network model is verified via cross-validation [23]. We implemented the leave-one-out cross-validation (Figure 11) normally adopted for obtaining a reliable estimation [24].

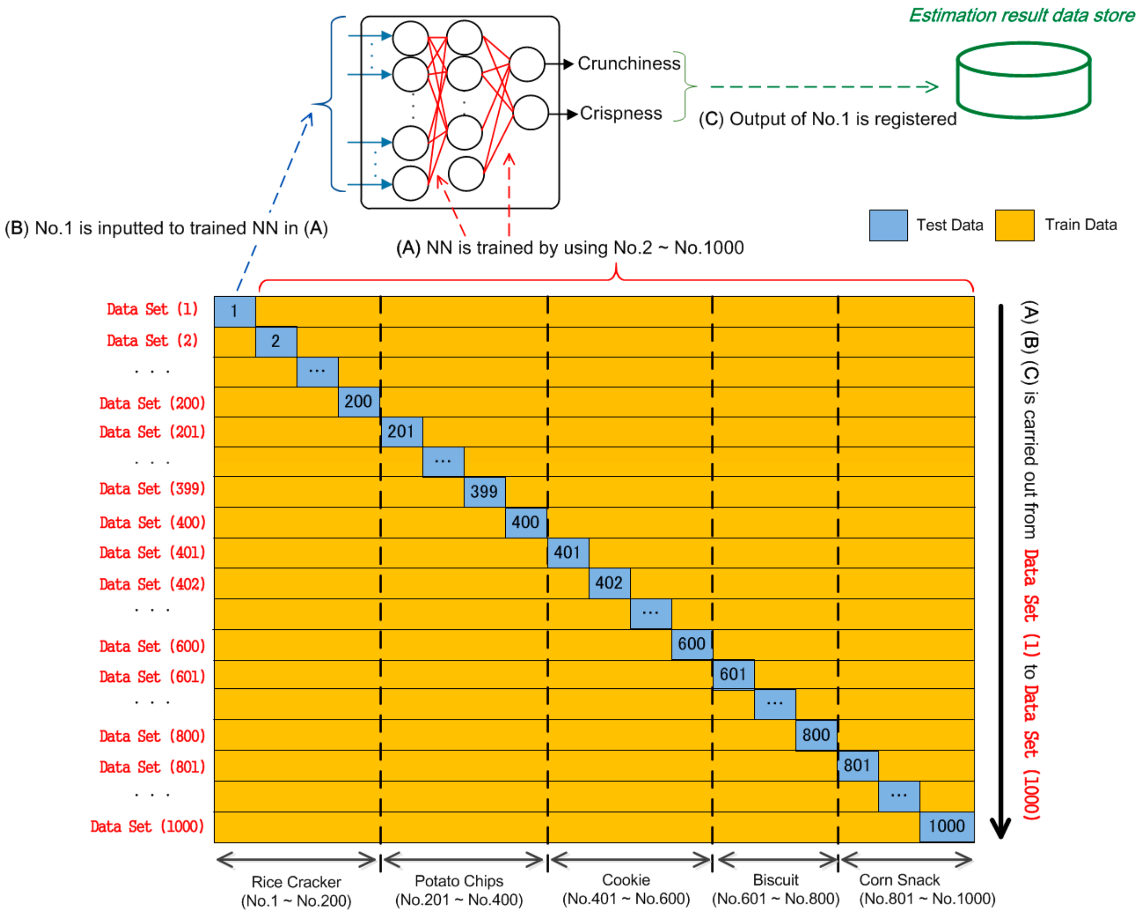

Estimation result data store
Crunchiness
Crispness
(C) Output of No.1 is registered
(B) No.1 is inputted to trained NN in (A)
(A) NN is trained by using No.2 ~ No.1000
Test Data
Train Data
Data Set (1)
Data Set (2)
Data Set (200)
Data Set (201)
Data Set (399)
Data Set (400)
Data Set (401)
Data Set (402)
Data Set (600)
Data Set (601)
Data Set (800)
Data Set (801)
Data Set (1000)
1
2
200
201
399
400
401
402
600
601
800
801
1000
(A) (B) (C) is carried out from Data Set (1) to Data Set (1000)
Rice Cracker (No.1 ~ No.200)
Potato Chips (No.201 ~ No.400)
Cookie (No.401 ~ No.600)
Biscuit (No.601 ~ No.800)
Corn Snack (No.801 ~ No.1000)

Figure 11. Schematic of the leave-one-out cross-validation.

For Data Set (1), initially, (W1–W5, F1–F5) of sample Nos. 2–1000 were used for training the connection weights of the neural network in step (A). Subsequently, (W1–W5, F1–F5) of No. 1 were inputted to the trained network in step (B). Finally, crunchiness and crispness outputs of the network were registered in the estimation result data store in step (C). Similarly, (A)–(C) were repeated from Data Sets (2)–(1000). This *Leave-One-Out Cross-Validation (LOOCV)* was thus considered as 1000-fold cross-validation. The above procedure could be explained as follows:

Step 0: $i \leftarrow 1$

Step 1: Select i-th sample data out of all 1000 samples

Step 2: Prepare the next 999 train input vectors except the i-th data

$$X_{train}^{(n)} = \begin{bmatrix} W_1^{(n)} \\ \vdots \\ W_5^{(n)} \\ F_1^{(n)} \\ \vdots \\ F_5^{(n)} \end{bmatrix} \quad \text{for } n = 1,\ 2, \cdots,\ 999.$$

Step 3: Prepare the next 999 correct output vectors

$$Y_{train}^{(n)} = \begin{bmatrix} Y_1^{(n)} \\ Y_2^{(n)} \end{bmatrix} \quad \text{for } n = 1,\ 2, \cdots,\ 999.$$

where, $Y^{(n)}_{train} = \begin{bmatrix} 0.9 \\ 0.8 \end{bmatrix}$ is assigned for the rice crackers, $Y^{(n)}_{train} = \begin{bmatrix} 0.2 \\ 0.9 \end{bmatrix}$ is assigned for the potato chips, $Y^{(n)}_{train} = \begin{bmatrix} 0.7 \\ 0.7 \end{bmatrix}$ is assigned for the cookies, $Y^{(n)}_{train} = \begin{bmatrix} 0.5 \\ 0.7 \end{bmatrix}$ is assigned for the biscuits, and $Y^{(n)}_{train} = \begin{bmatrix} 0.4 \\ 0.8 \end{bmatrix}$ is assigned for the corn snacks. These values were in accordance with Table 1.

Step 4: Initiate the connection weights w, v, and u, which are random values. Train the neural network model by the back-propagation algorithm by adjusting w, v, and u so that $Y^{(n)}_{train}$ is outputted when the corresponding $X^{(n)}_{train}$ is inputted. In particular, the iteration to train the network is 100 epochs. It is necessary to observe the decreasing error as the epoch proceeds.

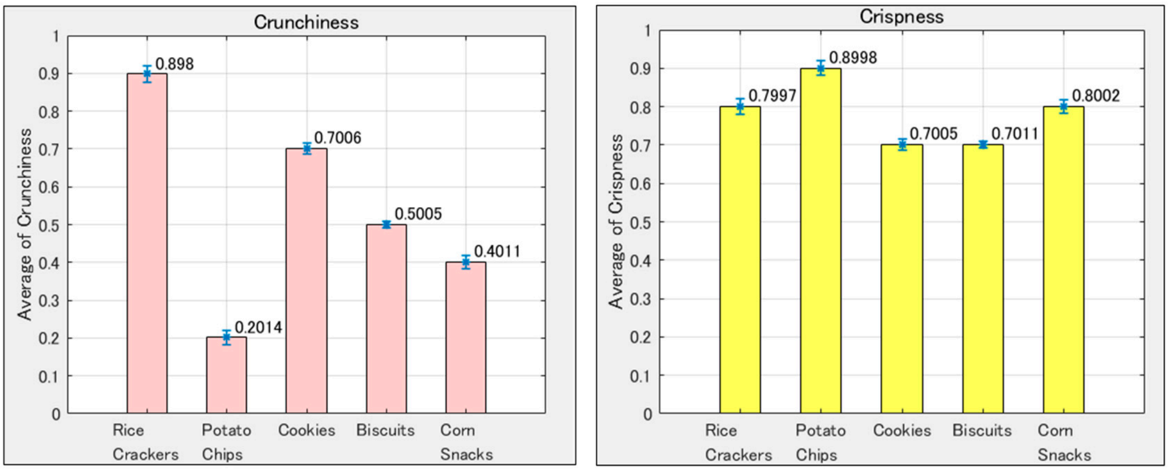
Step 5: Input W1–W5 and F1–F5 of the i-th sample data into the neural network model trained in previous step i.e., **Step 4**. (*Note that the i-th sample data is not used to train the neural network in **Step 4.**) The output texture value set (i.e., estimated texture result of the i-th sample) is registered in the estimation result data store.

When $i = 1000$, the routine is completed, otherwise $i \leftarrow i + 1$ and return to **Step 1**. Table 3 shows the averages of texture values in the estimation result data store. Although the sample data not used for training were inputted to the neural network model, the model outputs generally expected values.

Table 3. Averages of texture values in the estimation result data store.

	Rice Crackers	Potato Chips	Cookies	Biscuits	Corn Snacks
Crunchiness	0.8980	0.2014	0.7006	0.5005	0.4011
(expected)	(0.9)	(0.2)	(0.7)	(0.5)	(0.4)
Crispness	0.7997	0.8998	0.7005	0.7011	0.8002
(expected)	(0.8)	(0.9)	(0.7)	(0.7)	(0.8)

Based on the result of this implementation, the neural network model estimated the expected textures almost correctly, even though there was a dispersion in parameters W1–W5 and F1–F5. Figure 12 displays the averages and STDs in Table 3.



(**a**) Mean of estimated crunchiness (**b**) Mean of estimated crispness

Figure 12. Texture evaluation results of the samples: (**a**) The mean of estimated crunchiness degree and (**b**) crispness degree.

To validate the general flexibility of the neural network model, we performed another 10-fold cross-validation for Data Sets (1)–(10), as shown in Figure 13. For Data Set (1), Nos. 101–1000 were used to train the neural network, while Nos. 1–100 were used as test data and thus, inputted to the trained neural network afterward. The outputted results were registered in the estimation result data store. Similarly, for Data Set (2), Nos. 101–200 were used to test the neural network, previously trained by Nos. 1–100 and Nos. 201–1000. This process was iterated from Data Sets (1)–(10) in the same way explained above from Step 0 to Step 5. For each data set, the test data output were preserved in the estimation result data store. The estimation result data store for the validation is illustrated in Figure 14.

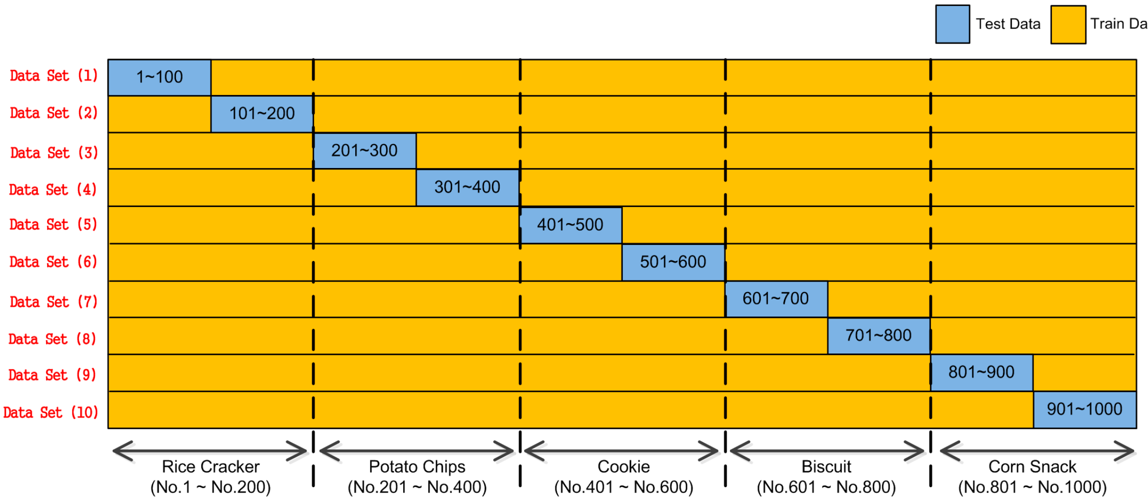
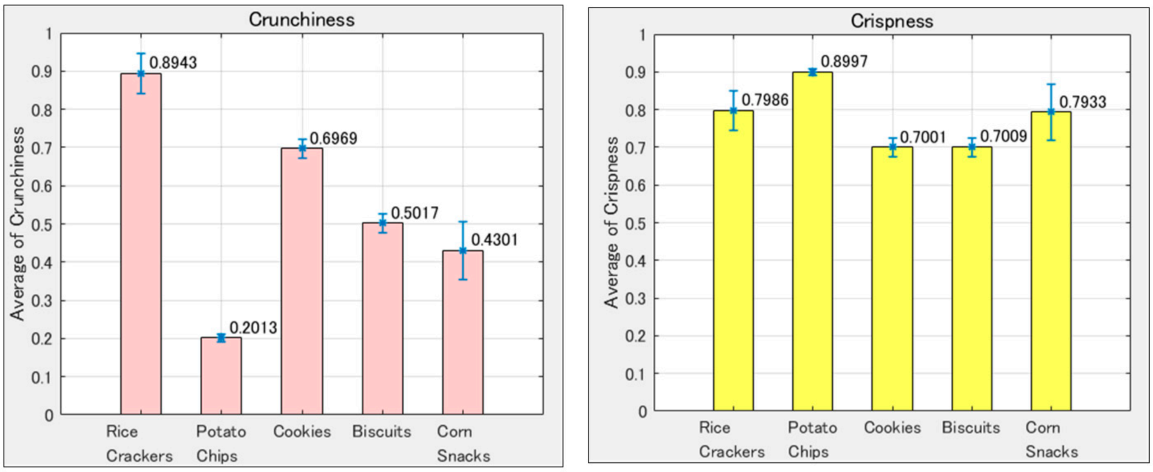


Figure 13. Data Sets (1)–(10) employed for the neural network model validation.



(**a**) Mean of estimated crunchiness

(**b**) Mean of estimated crispness

Figure 14. Texture evaluations of the samples using data sets of Figure 13: (**a**) The mean of estimated crunchiness degree and (**b**) crispness degree.

Since the training data were decreased compared with the Leave-One-Out Cross-Validation (LOOCV) in Figure 11 (i.e., LOOCV), the estimated result dispersed as STDs are shown in Figure 14. However, estimation values generally gathered around the expected values, even though W1–W5 and F1–F5 showed dispersion even for the same food specimen, as previously shown in Figures 8 and 9. In summary, the combination of our equipment and NN (Neural Network) worked well to quantify the texture levels.

5. CNN

The spectrogram has rich sound features [12–14]. To enhance the system capabilities, we considered introducing CNN [15], which deals with images. As a first step, we tried to classify snacks using CNN.

Since we have only 1000 data points, we adopted AlexNet [25], which is a pre-trained CNN. We employed the Neural Network Toolbox in MATLAB developed by MathWorks. We conducted Transfer Learning using AlexNet [26]. Figure 15 shows the CNN classifying the snacks. Note that the ReLU operation, cross-channel normalization, and max-pooling layers are omitted in Figure 15. CNN comprises the input layer for the 227 × 227 RGB image and soft-max output layer. The soft-max layer outputs classified values of rice cracker, potato chips, cookies, biscuit, and corn snacks within a range of (0,1). The item with the highest classified value is judged as an inputted snack. For instance, the value of the rice cracker is highest in Figure 15; thereby, the inputted image is classified as a rice cracker.

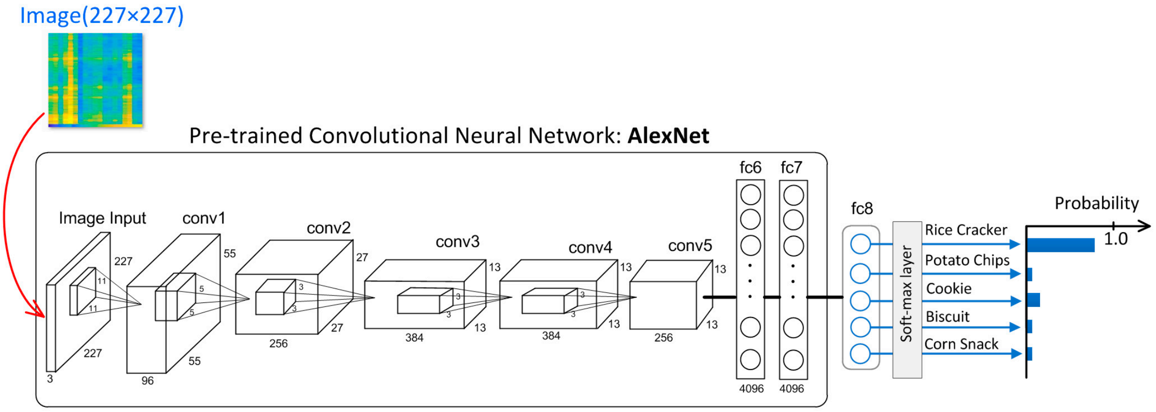
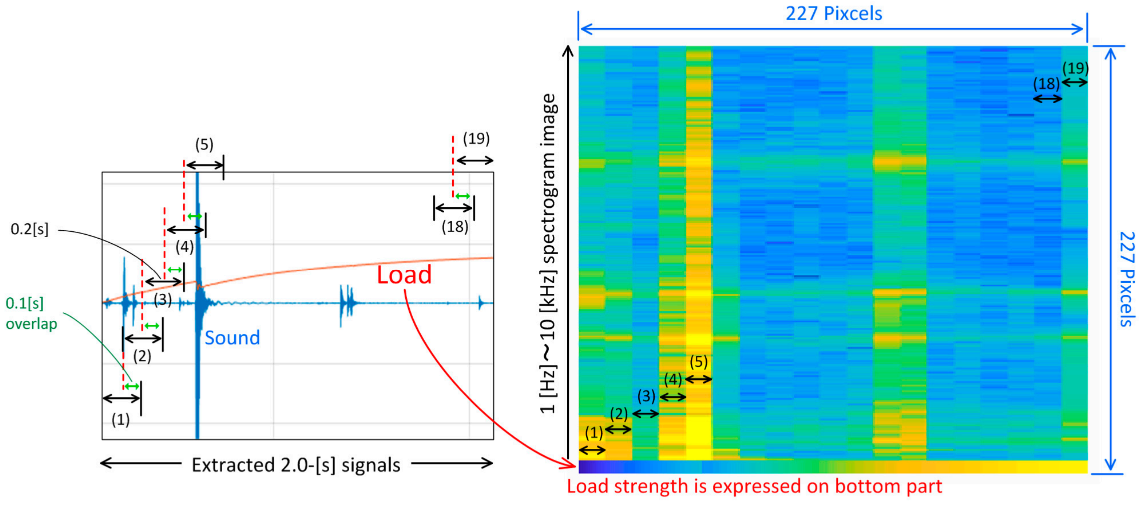


Figure 15. Convolutional neural network (CNN) for classifying snacks.

The convolution layers (conv1–conv5) and fully connected layers (fc6 and 7) are original parts which are pre-trained with a massive number of image data. We only attached fc8 and the soft-max layer with original AlexNet.

The input image (227 × 227 RGB) comprises the spectrogram and visualized load intensity (Figure 16b), which are obtained from extracted 2.0 s signals (Figure 16a). FFT is performed in each section from (1)–(19), as shown in Figure 16a, where the period of each section is 0.2 s, and then there are 0.1 s overlaps between the adjacent sections. FFT results of (1)–(19) are visualized, as shown in Figure 16b. In addition, this image also includes information in the load intensity, which is visualized in the bottom part of the image.



(**a**) FFT sections to make spectrogram.

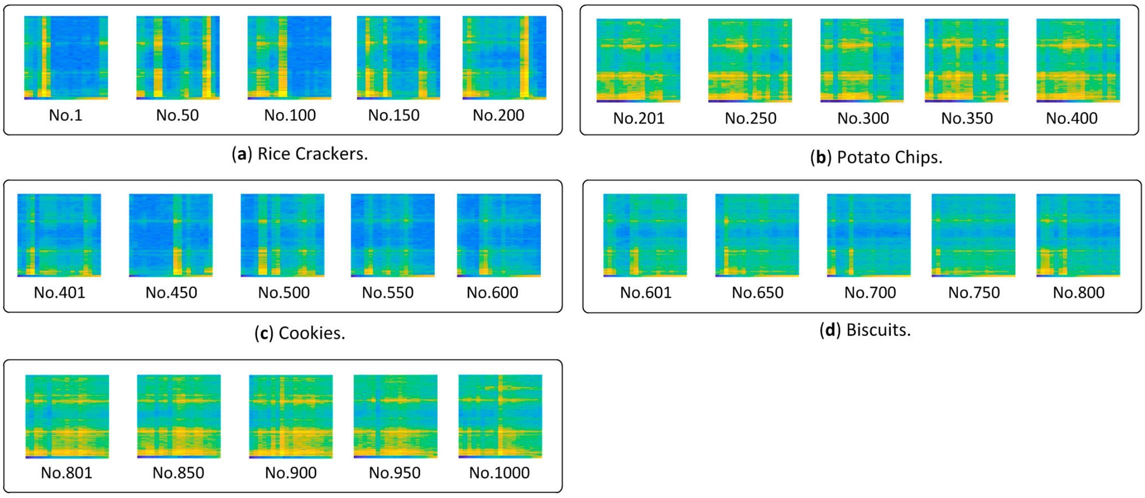
(**b**) Spectrogram (upper part), and load change (bottom part.)

Figure 16. Calculation of input image to CNN: (**a**) FFT (Fast Fourier Transform) period and (**b**) spectrogram and load change image.

The images of snacks are generated automatically and numbered as shown in Table 4, and then some of the generated images are shown in Figure 17.

Table 4. Snack image information.

	Rice Cracker	Potato Chips	Cookie	Biscuit	Corn Snack
Number of Images	200	200	200	200	200
Number	No. 1–200	No. 201–400	No. 401–600	No. 601–800	No. 801–1000



(**a**) Rice Crackers.

(**b**) Potato Chips.

(**c**) Cookies.

(**d**) Biscuits.

(**e**) Corn Snacks.

Figure 17. Images inputted to CNN: (**a**) Rice crackers; (**b**) potato chips; (**c**) cookies; (**d**) biscuits and; (**e**) corn snacks.

Table 5 enumerates the conditions for validation of CNN. The 10-fold cross-validation in Figure 13 is performed. The result is shown in Table 6. CNN performed very well. The mean of the accuracies in Data Sets (1)–(10) is 98.30%. The CNN is useful to modify or revise the outputted texture values from NN model described in previous Section 4. It is found that incorporating CNN must improve our system.

Table 5. Transfer learning settings.

Parameter	Value/Condition
Solver	sgdm
Learning Rate	0.0001
Max Epochs	20
Mini Batch Size	100
Total Iterations	180
Number of Train Data/Test Data	900/100

Table 6. The result of the 10-fold cross-validation, as shown in Figure 13.

Data Set	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Accuracy [%]	95	100	98	98	95	99	99	99	100	100
Mini-batch Loss at Epoch 1	1.8634	2.1723	2.1096	2.3062	1.9875	2.2507	2.3755	2.1629	2.1397	1.7647
Mini-batch Loss at Epoch 20	0.0034	0.0048	0.0035	0.0080	0.0123	0.0210	0.0042	0.0044	0.0071	0.0110

6. Conclusions

This paper proposed a food-texture-estimation system which is comprised of improved equipment and a simple neural network model. The system can process load changes and sound signals simultaneously for estimating textures, i.e., crunchiness and crispness. The classical neural network model was applied in the experiments to estimate the expected texture values of food specimens, i.e., rice crackers, potato chips, cookies, biscuits, and corn snacks. The model works well. Moreover, CNN is applied to classify the spectrogram image, including rich sound features, to expand our model capabilities. The CNN model performed very well. In the future, we will address the texture evaluation using CNN. Conventionally, food texture evaluation is manually performed by humans and is cumbersome. The neural network model simplifies this task. In this work, a moderate estimation was performed.

Author Contributions: The author's name and contributions are as follows: S.K.: conceptualization, methodology, software, validation, formal analysis, investigation, writing—original draft preparation, writing—review and editing, and visualization; N.W.: conceptualization, methodology, validation, formal analysis, investigation, supervision, and project administration; R.I.: conceptualization, methodology, software, validation, formal analysis, and investigation; T.S.: methodology, software, validation, formal analysis, and investigation; Y.N.: methodology, software, validation, formal analysis, and investigation; T.K.; conceptualization, methodology, validation, formal analysis, and investigation.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank reviewers for giving us fruitful comments, Enago (www.enago.jp) for the English language review, and Maeda in MathWorks for giving us technical advice.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Duizer, L. A Review of Acoustic Research for Studying the Sensory Perception of Crisp, Crunchy, and Crackly Textures. *Trends Food Sci. Technol.* **2001**, *12*, 17–24. [CrossRef]
2. Hayakawa, F.; Kazami, Y.; Nishinari, K.; Ioku, K.; Akuzawa, S.; Yamano, Y.; Baba, Y.; Kohyama, K. Classification of Japanese Texture Terms. *J. Texture Stud.* **2013**, *44*, 140–159. [CrossRef]
3. Sakurai, N.; Iwatani, S.; Terasaki, S.; Yamamoto, R. Texture Evaluation of Cucumber by a New Acoustic Vibration Method. *J. Jpn. Soc. Hortic. Sci.* **2005**, *74*, 31–35. [CrossRef]
4. Sakurai, N.; Iwatani, S.; Terasaki, S.; Yamamoto, R. Evaluation of 'Fuyu' Persimmon Texture by a New Parameter, Sharpness index. *J. Jpn. Soc. Hortic. Sci.* **2005**, *74*, 150–158. [CrossRef]
5. Taniwaki, M.; Hanada, T.; Sakurai, N. Development of Method for Quantifying Food Texture Using Blanched Bunching Onions. *J. Jpn. Soc. Hortic. Sci.* **2006**, *75*, 410–414. [CrossRef]
6. Liu, X.; Tan, J. Acoustic Wave Analysis for Food Crispness Evaluation. *J. Texture Stud.* **1999**, *30*, 397–408. [CrossRef]
7. Srisawas, W.; Jindal, V.K. Acoustic Testing of Snack Food Crispness Using Neural Networks. *J. Texture Stud.* **2003**, *34*, 401–420. [CrossRef]
8. Kato, S.; Wada, N.; Murakami, N.; Ito, R.; Kondo, R.; Goto, Y. The Estimation System of Food Texture Considering Sound and Load Using Neural Networks. In Proceedings of the 2017 International Conference on Biometrics and Kansei Engineering, Kyoto, Japan, 15–17 September 2017; pp. 104–109.
9. Okada, S.; Nakamoto, H.; Kobayashi, F.; Kojima, F. A Study on Classification of Food Texture with Recurrent Neural Network. In Proceedings of the Intelligent Robotics and Applications (ICIRA 2016), Lecture Notes in Computer Science, Tokyo, Japan, 22–24 August 2016; Volume 9834, pp. 247–256.
10. Kato, S.; Wada, N.; Ito, R.; Shiozaki, T.; Nishiyama, Y.; Kagawa, T. Texture Estimation System of Snacks Using Neural Network Considering Sound and Load. In Proceedings of the 13th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2018), Lecture Notes on Data Engineering and Communications Technologies, Taichung, Taiwan, 27–29 October 2018; Volume 24, pp. 48–61.
11. Tabilo-Munizaga, G.; Barbosa-Canovas, G.V. Rheology for the food industry. *J. Food Eng.* **2005**, *67*, 147–156. [CrossRef]

12. He, D.-C.; Wang, L. Texture Unit, Texture Spectrum, And Texture Analysis. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 509–512.
13. Sharan, R.V.; Moir, T.J. Robust audio surveillance using spectrogram image texture feature. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 1956–1960.
14. Nanni, L.; Costa, Y.M.G.; Lucio, D.R.; Silla, C.N., Jr.; Brahnam, S. Combining visual and acoustic features for audio classification tasks. *Pattern Recognit. Lett.* **2017**, *88*, 49–56. [CrossRef]
15. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
16. Salamon, J.; Bello, J.P. Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Process. Lett.* **2017**, *24*, 279–283. [CrossRef]
17. Minaee, S.; Abdolrashidi, A. Multispectral palmprint recognition using textural features. In Proceedings of the 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 13 December 2014.
18. Minaee, S.; Abdolrashidi, A.; Wang, Y. Iris recognition using scattering transform and textural features. In Proceedings of the 2015 IEEE Signal Processing and Signal Processing Education Workshop (SP/SPE), Salt Lake City, UT, USA, 9–12 August 2015.
19. Minaee, S.; Bouazizi, I.; Kolan, P.; Najafzadeh, H. Ad-Net: Audio-Visual Convolutional Neural Network for Advertisement Detection In Videos. *arXiv* **2018**, arXiv:1806.08612.
20. Hershey, S.; Chaudhuri, S.; Ellis, D.P.W.; Gemmeke, J.F.; Jansen, A.; Moore, R.C.; Plakal, M.; Platt, D.; Saurous, R.A.; Seybold, B.; Slaney, M.; Weiss, R.J.; Wilson, K. CNN architectures for large-scale audio classification. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017.
21. Sesmat, A.; Meullenet, J.-F. Prediction of Rice Sensory Texture Attributes from a Single Compression Test, Multivariate Regression, and a Stepwise Model Optimization Method. *J. Food Sci.* **2001**, *66*, 124–131. [CrossRef]
22. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Representations by Back-propagating Errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
23. Priddy, K.L.; Keller, P.E. *Artificial Neural Networks—An Introduction*; Dealing with Limited Amounts of Data; SPIE Press: Bellingham, WA, USA, 2005; Chapter 11, pp. 2839–2846.
24. Wong, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross-validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [CrossRef]
25. Alex, K.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105.
26. MathWorks, Transfer Learning Using AlexNet. Available online: https://www.mathworks.com/help/deeplearning/examples/transfer-learning-using-alexnet.html?lang=en (accessed on 28 February 2019).

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).