Pertanika J. Sci. & Technol. 31 (6): 2703 - 2722 (2023)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Recognition of Fruit Types from Striking and Flicking Sounds

Rong Phoophuangpairoj

Department of Computer Engineering, College of Engineering, Rangsit University, Pathumthani 12000, Thailand

ABSTRACT

This paper proposes a method to recognize fruits whose quality, including their ripeness, grades, brix values, and flesh characteristics, cannot be determined visually from their skin but from striking and flicking sounds. Four fruit types consisting of durians, watermelons, guavas, and pineapples were studied in this research. In recognition of fruit types, preprocessing removes the non-striking/non-flicking parts from the striking and flicking sounds. Then the sequences of frequency domain acoustic features containing 13 Mel Frequency Cepstral Coefficients (MFCCs) and their 13 first- and 13 second-order derivatives were extracted from striking and flicking sounds. The sequences were used to create the Hidden Markov Models (HMMs). The HMM acoustic models, dictionary, and grammar were incorporated to recognize striking and flicking sounds. When testing the striking and flicking sounds obtained from the fruits used to create the training set but were collected at different times, the recognition accuracy using 1 through 5 strikes/flicks was 98.48%, 98.91%, 99.13%, 98.91%, and 99.57%, respectively. For an unknown test set, of which the sounds obtained from the fruits that were not used to create the training set, the recognition accuracy using 1 through 5 strikes/flicks were 95.23%, 96.82%, 96.82%, 97.05%, and 96.59%, respectively. The results also revealed that the proposed method could accurately distinguish the striking sounds of durians from the flicking sounds of watermelons, guavas, and pineapples.

Keywords: Flicking sounds, fruit grading, fruit recognition, Hidden Markov Models, striking

ARTICLE INFO

Article history:
Received: 15 September 2022
Accepted: 06 March 2023
Published: 08 September 2023

DOI: https://doi.org/10.47836/pjst.31.6.04

E-mail address:
rong.p@rsu.ac.th

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Fruits are vital for health which supply necessary nutrition supplements to life. When buying fruits, customers anticipate getting their desired fruit quality, including ripeness, sweetness, and characteristics of the flesh inside. However, when cutting or

peeling them, many customers were unsatisfied with the quality of the fruits they bought. For certain kinds of fruits, for example, durians, watermelons, guavas, and pineapples (Figure 1), the characteristics of fruit pulp and physiochemical properties were generally unknown by observing from their outer skins.

(a) (b) (c) (d)

Figure 1. Four types of fruits in the recognition: (a) durians; (b) watermelons; (c) guavas; and (d) pineapples

Experienced fruit merchants usually determine the fruit quality by striking or flicking them and listening to the generated sounds. When flicking, the index or middle finger is released from the thumb against an object (Figure 2). Flicking can assess the quality of fruits, e.g., watermelon, guava, and pineapple. However, flickering a durian- the thorny king of fruit- is not practical as it can injure the finger. Striking or tapping the durian with a tapping stick and listening to the sounds to determine the ripeness of durians are shown in Figure 3.

Figure 2. Flicking a guava

Figure 3. Tapping a durian

Figure 4 shows the characteristics of the striking/flicking sounds of durians, watermelons, guavas, and pineapples. They are quite similar and difficult to be visually distinguished. They consist of non-striking/non-flicking parts and striking/flicking parts. Each striking/flicking sound normally begins with a non-striking/non-flicking part, followed by a striking/flicking part, and ends with a non-striking/non-flicking part. The flicking/

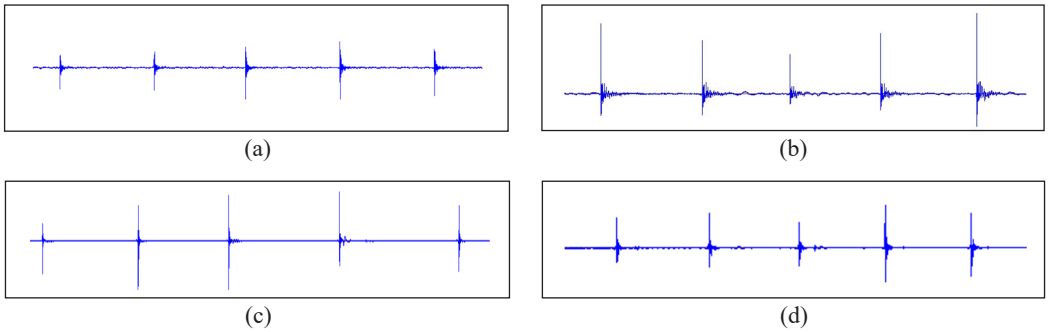

Figure 4. Striking and flicking sounds of fruits five times: (a) durian; (b) watermelon; (c) guava; and (d) pineapple

striking parts have higher amplitude than the non-flicking/non-striking parts. The striking/flicking parts, usually much shorter than the non-striking/non-flicking parts, contain more information about the types of fruits and fruit characteristics. In addition, the duration and the amplitude of flicking/striking signals derived from the same or different fruits fluctuate, resulting in difficulties in recognition. The pace of flicking/striking the fruits can affect the results, especially if the duration of non-flicking/non-striking parts is long. In order to overcome these differences, the preprocessing method of reducing the flickering parts based on the amplitude of the signals was proposed (Phoophuangpairoj, 2014a). The duration of striking/flicking parts depends on when the finger or the tapping stick hits the fruits. The hardness or impact of flicking and striking affects the amplitude of signals.

Therefore, it is not practical to determine the quality of fruits from the amplitude but from the frequency-based features extracted from signals and models that can efficiently capture acoustic phenomena. It can be seen from how some merchants recognize short flicking and striking sounds, which have some frequency differences to predict the internal fruit flesh. Based on the results, using HMMs with frequency-based features could efficiently handle the different impacts of watermelon flicking and durian tapping (Phoophuangpairoj, 2014a; Phoophuangpairoj, 2014b). For guavas, repeatedly flicking the same area can affect the recognition results. The flicking should be applied to the different areas of the guavas to classify the freshness of the guavas.

There was research applying speech recognition technologies to recognize the quality of watermelons and guavas using flicking sounds (Phoophuangpairoj, 2014a; Phoophuangpairoj, 2013). For the recognition of firm flesh and flesh with cracks, the average watermelon quality recognition rates of 95.0%, 97.0%, 98.0%, 98.0%, and 98.0% were achieved by using 1 through 5 flicks, respectively. For guavas stored in a normal refrigerator, the average correct freshness recognition rates of 92.0%, 88.0%, and 94.0% were obtained from fresh, 3-day-kept, and 6-day-kept guavas, respectively. The striking sounds were also used to recognize ripe and unripe durians using a dictionary and grammar (Phoophuangpairoj, 2014b) and an N-gram language model (Phoophuangpairoj, 2014c).

The recognition system using MFCC-based features and HMMs efficiently recognized the quality of watermelons and durians. When using grammar, the durian ripeness recognition rates of 91.0%, 92.0%, 90.0%, 92.0%, and 92.0% were achieved using 1 through 5 strikes, respectively. Using an N-gram language model, the durian ripeness recognition rates of 88.0%, 91.0%, 91.0%, 92.0%, 92.0%, and 90.0% were achieved using 1 through 5 strikes, respectively. The flicking sounds were also studied to classify pineapples and their physicochemical properties. Even though predicting pineapple grades using flicking sounds cannot be done efficiently, the results showed that pineapples classified as grade 1 and grade 3 differed significantly in terms of total soluble solid (TSS), pH value, and water content (Phoophuangpairoj & Srikun, 2014).

When recognizing some fruit types whose internal characteristics cannot be visually determined from their skin, flicking and striking sounds can also be applied. Studying the feasibility of recognizing the types of fruits from the flicking and striking sounds without using image processing is beneficial. The image processing requires another different source of data, while the proposed method merely utilizes a source of data, which is more efficient. Hence, this work proposed a novel method to recognize the fruit types without using image processing but did not stress the quality of fruits because this issue had already been researched.

LITERATURE SURVEY

Automation in food processing plays a crucial role in increasing the productivity, quality, and profitable growth of countries. Fruit grading is a process for producers which affects fruit quality evaluation and export markets (Raja et al., 2018). Automatic fruit classification is an interesting issue in the retailing and fruit-growing industry because it can help farmers and supermarkets identify the status of fruits from stock or containers (Shahi et al., 2022). Computer vision and machine learning methods have been applied for fruit detection, ripeness, and categorization in the past decade (Fan et al., 2020; Hossain et al., 2018). The problem of classifying fruits and vegetables in computer vision remains a challenge because some fruits look alike and have similar colors, shapes, and textures. CNNs (Convolutional Neural Networks) and transfer learning have obtained impressive results in image classification (Albarrak et al., 2022). Based on the previous work, CNN recognized 26 categories of fruits and vegetable images (Zeng, 2017) and orange grades (Asriny et al., 2020). A system has to extract image features and use them as a source to recognize the fruits to recognize the fruit quality from a video. Meanwhile, the system has to use striking or flicking sounds as the other source to recognize their qualities. Such a system requires two different sources, and this may not be as practical, resulting in creating a more complex heterogeneous system when compared to the system using only one source of flicking or striking sounds.

The success of MFCC acoustic features combined with their cost-effective and robust computation turned them into a standard choice in speech recognition applications. In speech recognition systems such as an Arabic speech recognition system, 39 MFCC-based acoustic features were extracted by partitioning the speech signals into frames (Elharati et al., 2020). HMM is a model used to represent the acoustic phenomenon and acoustic changes according to time. HMMs provide a highly reliable method of recognizing spoken signals (Chavan & Sable, 2013; Naithani et al., 2018; Najkar et al., 2010). HMMs were also applied to recognize inhaling and exhaling signals (Phoophuangpairoj, 2020) and sleep spindles (Stevner et al., 2019). For HMM, Gaussian Mixture Models (GMMs), which are the components within each HMM state, were primarily utilized to compute the probabilistic distribution of each phone or phoneme (or any speech signal atom), and the fusion of GMMs-HMMs has led to many successful automatic speech recognition (ASR) applications (Kiranyaz et al., 2021). Phonemes or syllables were combined into words and sentences using a dictionary and language model. The main reasons for this success are this model's analytic ability in the speech phenomenon and its accuracy in practical speech recognition systems (Najkar et al., 2010).

Viterbi is an algorithm that searches HMM states to find the most probable phone, phoneme, word, and sentence from the acoustic models of phones or phonemes connected based on a dictionary and grammar. Viterbi algorithm was applied to recognize or search the possible phones in a speech recognition system (Hatala & Puturuhu, 2019).

MATERIALS AND METHODS

Data

Data were collected from four different types of fruits: 100 durians, 100 watermelons, 150 guavas, and 110 pineapples. The striking/flicking sounds were recorded at 11,025 Hz. The data were collected from fruits of different quality, grades, and ripeness. Nonetheless, the work studied the differentiation of the fruit types. Striking sounds were derived from 100 durians struck five times each to train the HMM acoustic models. The flicking sounds were obtained from 100 watermelons and 150 guavas, all flicked five times each, while the pineapple flicking sounds were obtained from 110 pineapples, ten times each. For testing, untrained and unknown sets were used. The untrained set was collected from the same set of fruits used in the training but at different times, whereas the unknown set was collected from different fruits not included in the training. The untrained set consisted of 1 through 5 striking/flicking sounds, each collected from 100 durians, 100 watermelons, 150 guavas, and 110 pineapples. The unknown set consisted of 1 through 5 striking/flicking sounds, each collected from 100 durians, 100 watermelons, 150 guavas, and 90 pineapples.

Preprocessing

As a rule, the striking/flicking parts contain information about fruit quality and types. Preprocessing was performed to reduce the non-striking/non-flicking parts. The digitized signals contain positive, negative, and zero values. As a result, it is easier to set a removing threshold by computing their absolute values. Additionally, a clipping or cut-off threshold was applied to handle the high difference in the signal amplitude. The threshold (*Th*) to reduce non-flicking/striking parts was computed from all frames using Equation 1:

$$framesize = \frac{frame\_duration}{1000} \times sampling\_rate \quad (1)$$

where: *framesize* is the number of points or values in each frame; *frame_duration* is the frame duration or size in milliseconds; *sampling_rate* is the recording sampling rate (11,025 Hz).

The frame duration (*frame_duration*) was set to 2 milliseconds. The number of samples (*num_smp_file*) was obtained from each wav file. Then the *nFrame,* which is the number of frames in a striking/flicking file, was computed using Equation 2.

$$nFrame = \frac{num\_smp\_file}{framesize} \quad (2)$$

Clip() was a function used to clip the signals. If any value of $|s_i|$ was higher than a clipping threshold (Th_{clip}) (e.g., 10,000), the value was set to the threshold (Equations 3-5).

$$Frame_{abs_n} = \sum_{i=(n-1)\times framesize\ +1}^{n\times framesize} clip(|s_i|), \quad 1 \le n \le nFrame \quad (3)$$

$$AvgFrame_{abs} = \frac{\sum_{i=1}^{\mathrm{nFrame}} Frame_{abs_i}}{nFrame} \quad (4)$$

$$Th = C * AvgFrame_{abs} \quad (5)$$

where: $Frame_{abs_n}$ is the sum of clipped absolute values computed from the nth frame; $AVGFrame_{abs}$ is the average of $Frame_{abs_n}$ computed from all frames; *C* is a constant (e.g., 3); *It* is the threshold to reduce non-striking/non-flicking parts.

Then the following method was applied to remove non-striking/non-flicking parts (frames) from the signals.

```
for(n=1;n<=nFrame;n++)
    KeepFrame[n] = 0
    if((vFrame_abs_n >= Th) OR (n equals to 1, 2, nFrame-1 or nFrame))
        KeepFrame[n] = 1;  // The n^th frame is flagged to be kept.
```

The method scanned all frames. If it was the first or last two frames or the sum of clipped absolute values computed from the nth frame ($Frame_{abs_n}$) was greater than or equal to the threshold to reduce non-striking/non-flicking parts, the KeepFrame[n] was set to 1. When the KeepFrame[n] was equal to 1, the nth frame was kept. When the KeepFrame[n] was equal to 0, the nth frame was removed. Removing all non-striking/non-flicking parts without losing some precious short striking/flicking parts is difficult. According to the algorithm below, some parts before and after each striking/flicking were kept to ensure that the precious data in striking/flicking parts were not removed and to obtain longer signals for recognizing HMMs, which contained a higher number of states. The algorithm kept the frames before and after each striking/flicking part, as described below.

```
for(n=2;n<nFrame;n=n+1)
    if(KeepFrame[n] equals to 1 AND KeepFrame[n-1] equals to 0)
        KeepFrame[n-1] = 1;

for(n=nFrame-1;n>1;n=n-1)
    if(KeepFrame[n] equals to 1 AND KeepFrame[n+1] equals to 0)
        KeepFrame[n+1] = 1;
```

The data in the nth frame of which KeepFrame[n] equals 1 were written in the preprocessed file. Thereafter, the Hidden Markov Model Toolkit (HTK) (http://htk.eng.cam.ac.uk/) was used to extract acoustic features from the preprocessed signals, train HMM acoustic models and detect the types of fruits. As for the evaluation, HResult, a tool in HTK, was used to find the fruit type recognition accuracy.

Extracting Acoustic Features

The time-domain or the visual characteristics of the durian striking, watermelon, guava, and pineapple flicking were similar. Therefore, the frequency-domain features of the signals were computed and used instead. Acoustic features consisting of 13 MFCCs and their 13 first- and 13 second-order derivatives were extracted from each particular time or window. The feature extraction used a window size of 4 milliseconds and a window shift rate of 1 millisecond. Left-to-right HMMs were used to model striking/flicking parts of each fruit type and a shared model of non-striking/non-flicking parts.

Creating Acoustic Models

The sounds were transcribed without providing the position of each part. For each fruit type, the non-flicking and non-striking parts which remained after the preprocessing were represented using sil (silence). Each striking/flicking part was represented based on the types of fruits, drstrike for each striking signal of a durian, wmflick, gvflick, and paflick for

each flicking signal of watermelon, guava, and pineapple, respectively. The transcriptions of each 5-striking/5-flicking signal used in the training are shown in Table 1.

The transcriptions without the positions of striking/flicking parts were used in the training. HTK tried to find acoustic features of striking/flicking and non-striking/non-flicking parts to create the acoustic models according to the transcriptions. In training, the number of HMM states varied from 4 to 7, and the number of Gaussian mixtures in each state was from 2 to 6.

Table 1
Transcriptions of each 5-striking/5-flicking signal used in the training

Types of fruits	Transcriptions for each 5-striking/5-flicking sound
durian	sil drstrike sil drstrike sil drstrike sil drstrike sil drstrike sil
watermelon	sil wmflick sil wmflick sil wmflick sil wmflick sil wmflick sil
guava	sil gvflick sil gvflick sil gvflick sil gvflick sil gvflick sil
pineapple	sil paflick sil paflick sil paflick sil paflick sil paflick sil

Creating a Dictionary to Recognize Fruit Types

Words representing 1 through 5 strikes/flicks were defined according to the characteristics of the striking and flicking sounds. For example, the one-flick word consists of a non-flicking/non-striking model followed by each fruit type's striking or flicking model and the sil model. In the Thai language, words can be pronounced in different ways. The words in the dictionary consisting of durian, watermelon, guava, and pineapple were created from different numbers of phones. The words in the dictionary to represent the sounds of 1 through 5 strikes/flicks are shown below.

durian [durian] sil drstrike sil
watermelon [watermelon] sil wmflick sil
guava [guava] sil gvflick sil
pineapple [pineapple] sil paflick sil
durian [durian] sil drstrike sil drstrike sil
watermelon [watermelon] sil wmflick sil wmflick sil
guava [guava] sil gvflick sil gvflick sil
pineapple [pineapple] sil paflick sil paflick sil
durian [durian] sil drstrike sil drstrike sil drstrike sil
watermelon [watermelon] sil wmflick sil wmflick sil wmflick sil
guava [guava] sil gvflick sil gvflick sil gvflick sil
pineapple [pineapple] sil paflick sil paflick sil paflick sil
durian [durian] sil drstrike sil drstrike sil drstrike sil drstrike sil
watermelon [watermelon] sil wmflick sil wmflick sil wmflick sil wmflick sil
guava [guava] sil gvflick sil gvflick sil gvflick sil gvflick sil

pineapple [pineapple] sil paflick sil paflick sil paflick sil paflick sil
durian [durian] sil drstrike sil drstrike sil drstrike sil drstrike sil drstrike sil
watermelon [watermelon] sil wmflick sil wmflick sil wmflick sil wmflick sil wmflick sil
guava [guava] sil gvflick sil gvflick sil gvflick sil gvflick sil gvflick sil
pineapple [pineapple] sil paflick sil paflick sil paflick sil paflick sil paflick sil

Creating Grammar to Recognize Fruit Types

Grammar was applied to constrain the recognition results. The following grammar was created to recognize the fruit types.

$fruittype = durian | watermelon | guava | pineapple;
($fruittype)

The | means "or" while the () means no repetition. The first line of recognition grammar states that $fruittype can be durian, watermelon, guava, or pineapple, while the second line states that only the $fruittype without repetition can be derived.

Recognizing Types of Fruits

The acoustic models, dictionary, and grammar were integrated to recognize the extracted acoustic features. The dictionary and grammar make the recognition more flexible and easy to handle the arbitrary numbers of strikes and flicks. The Viterbi algorithm decided which hypotheses or word-connected paths comprising phones were most likely to be the correct textual interpretation of the signals.

RESULTS

To investigate the results, the number of strikes/flicks, the number of HMM states, and the number of Gaussian mixtures per state were varied. Table 2 shows the accuracy obtained from the untrained set. The results reported that of N testing times in which each time used F strikes or flicks, H times were recognized correctly. There were S substitution errors. When using 1 strike/flick, the accuracy of 98.48% was obtained by using 6 states and 5 Gaussian mixtures per state (6S5M). For 2 through 5 strikes/flicks, 98.91%, 99.13%, 98.91%, and 99.57% accuracy rates were obtained using 6 states and 4 Gaussian mixtures per state (6S4M), respectively. However, when using one strike or flick with HMMs containing 7 states, there were six striking/flicking sounds that the system was not provided the recognition results because the number of the feature vectors extracted from very short signals was not adequate to be recognized using the HMMs.

Table 3 shows the fruit type recognition accuracy obtained from the unknown set.

Table 2
Accuracy of fruit type recognition obtained from the untrained set based on the number of strikes/flicks, states, and Gaussian mixtures

Number of strike(s)/flick(s) (T)	Number of states	Number of Gaussian mixtures	Accuracy (%)	H	S	N
1	4	2	93.48	430	30	460
	4	3	93.70	431	29	460
	4	4	93.04	428	32	460
	4	5	93.70	431	29	460
	4	6	96.74	445	15	460
	5	2	95.65	440	20	460
	5	3	95.65	440	20	460
	5	4	95.00	437	23	460
	5	5	95.87	441	19	460
	5	6	96.09	442	18	460
	6	2	95.87	441	19	460
	6	3	97.39	448	12	460
	6	4	98.04	451	9	460
	6	5	98.48	453	7	460
	6	6	98.26	452	8	460
	7	2	94.93	431	23	460
	7	3	96.04	436	18	454
	7	4	96.70	439	15	454
	7	5	97.14	441	13	454
	7	6	97.58	443	11	454
2	4	2	94.13	433	27	460
	4	3	95.43	439	21	460
	4	4	95.22	438	22	460
	4	5	95.43	439	21	460
	4	6	96.52	444	16	460
	5	2	95.87	441	19	460
	5	3	96.74	445	15	460
	5	4	95.87	441	19	460
	5	5	95.43	439	21	460
	5	6	97.17	447	13	460
	6	2	98.04	451	9	460
	6	3	98.70	454	6	460
	6	4	98.91	455	5	460
	6	5	98.70	454	6	460
	6	6	98.26	452	8	460
	7	2	96.30	443	17	460
	7	3	98.04	451	9	460
	7	4	98.04	451	9	460

Table 2 *(continue)*

Number of strike(s)/flick(s) (T)	Number of states	Number of Gaussian mixtures	Accuracy (%)	H	S	N
	7	5	97.61	449	11	460
	7	6	98.04	451	9	460
3	4	2	93.48	430	30	460
	4	3	94.35	434	26	460
	4	4	95.22	438	22	460
	4	5	95.22	438	22	460
	4	6	96.96	446	14	460
	5	2	96.74	445	15	460
	5	3	96.74	445	15	460
	5	4	96.09	442	18	460
	5	5	95.87	441	19	460
	5	6	97.39	448	12	460
	6	2	97.83	450	10	460
	6	3	98.91	455	5	460
	6	4	99.13	456	4	460
	6	5	98.70	454	6	460
	6	6	98.70	454	6	460
	7	2	97.17	447	13	460
	7	3	99.13	456	4	460
	7	4	98.91	455	5	460
	7	5	98.70	454	6	460
	7	6	99.13	456	4	460
4	4	2	94.78	436	24	460
	4	3	94.57	435	25	460
	4	4	94.78	436	24	460
	4	5	95.22	438	22	460
	4	6	96.74	445	15	460
	5	2	96.52	444	16	460
	5	3	97.17	447	13	460
	5	4	95.43	439	21	460
	5	5	96.30	443	17	460
	5	6	97.83	450	10	460
	6	2	97.83	450	10	460
	6	3	98.70	454	6	460
	6	4	98.91	455	5	460
	6	5	98.48	453	7	460
	6	6	98.70	454	6	460
	7	2	97.61	449	11	460
	7	3	98.26	452	8	460

Table 2 *(continue)*

Number of strike(s)/flick(s) (T)	Number of states	Number of Gaussian mixtures	Accuracy (%)	H	S	N
	7	4	98.48	453	7	460
	7	5	98.48	453	7	460
	7	6	98.70	454	6	460
5	4	2	94.35	434	26	460
	4	3	95.00	437	23	460
	4	4	95.00	437	23	460
	4	5	95.43	439	21	460
	4	6	96.52	444	16	460
	5	2	97.17	447	13	460
	5	3	97.39	448	12	460
	5	4	96.30	443	17	460
	5	5	96.52	444	16	460
	5	6	98.26	452	8	460
	6	2	98.04	451	9	460
	6	3	98.91	455	5	460
	6	4	99.57	458	2	460
	6	5	98.70	454	6	460
	6	6	98.70	454	6	460
	7	2	97.39	448	12	460
	7	3	98.91	455	5	460
	7	4	99.13	456	4	460
	7	5	98.70	454	6	460
	7	6	99.13	456	4	460

Table 3
Accuracy of fruit type recognition obtained from the unknown set based on the number of strikes/flicks, states, and Gaussian mixtures

Number of strike(s)/flick(s) (T)	Number of states	Number of Gaussian mixtures	Accuracy (%)	H	S	N
1	4	2	92.50	407	33	440
	4	3	92.27	406	34	440
	4	4	92.73	408	32	440
	4	5	92.73	408	32	440
	4	6	93.41	411	29	440
	5	2	92.73	408	32	440
	5	3	92.73	408	32	440
	5	4	94.09	414	26	440
	5	5	95.23	419	21	440
	5	6	93.86	413	27	440

Table 3 *(continue)*

Number of strike(s)/flick(s) (T)	Number of states	Number of Gaussian mixtures	Accuracy (%)	H	S	N
	6	2	94.55	416	24	440
	6	3	94.09	414	26	440
	6	4	94.55	416	24	440
	6	5	95.23	419	21	440
	6	6	95.00	418	22	440
	7	2	93.74	404	27	431
	7	3	94.66	408	23	431
	7	4	94.20	406	25	431
	7	5	93.27	402	29	431
	7	6	93.50	403	28	431
2	4	2	92.27	406	34	431
	4	3	94.55	416	24	440
	4	4	94.55	416	24	440
	4	5	94.32	415	25	440
	4	6	94.77	417	23	440
	5	2	94.32	415	25	440
	5	3	95.23	419	21	440
	5	4	94.55	416	24	440
	5	5	95.68	421	19	440
	5	6	95.45	420	20	440
	6	2	95.23	419	21	440
	6	3	96.36	424	16	440
	6	4	95.91	422	18	440
	6	5	96.82	426	14	440
	6	6	96.36	424	16	440
	7	2	94.32	415	25	440
	7	3	94.77	417	23	440
	7	4	95.45	420	20	440
	7	5	95.68	19	19	440
	7	6	95.91	422	18	440
3	4	2	92.05	405	35	440
	4	3	93.64	412	28	440
	4	4	93.64	412	28	440
	4	5	94.09	414	26	440
	4	6	94.77	417	23	440
	5	2	94.09	414	26	440
	5	3	95.23	419	21	440
	5	4	95.23	419	21	440
	5	5	94.77	417	23	440

Table 3 *(continue)*

Number of strike(s)/flick(s) (T)	Number of states	Number of Gaussian mixtures	Accuracy (%)	H	S	N
	5	6	95.68	421	19	440
	6	2	95.23	419	21	440
	6	3	97.05	427	13	440
	6	4	96.82	426	14	440
	6	5	96.82	426	14	440
	6	6	96.59	425	15	440
	7	2	94.55	416	24	440
	7	3	95.00	418	22	440
	7	4	95.23	419	21	440
	7	5	96.14	423	17	440
	7	6	96.59	425	15	440
4	4	2	92.73	408	32	440
	4	3	93.18	410	30	440
	4	4	94.09	414	26	440
	4	5	94.77	417	23	440
	4	6	94.55	416	24	440
	5	2	94.32	415	25	440
	5	3	95.00	418	22	440
	5	4	95.23	419	21	440
	5	5	95.45	420	20	440
	5	6	94.55	416	24	440
	6	2	95.00	418	22	440
	6	3	96.14	423	17	440
	6	4	95.91	422	18	440
	6	5	97.05	427	13	440
	6	6	96.36	424	16	440
	7	2	95.00	418	22	440
	7	3	95.00	418	22	440
	7	4	96.14	423	17	440
	7	5	95.68	421	19	440
	7	6	96.14	423	17	440
5	4	2	92.27	406	34	440
	4	3	93.41	411	29	440
	4	4	93.64	412	28	440
	4	5	95.00	418	22	440
	4	6	95.00	418	22	440
	5	2	94.09	414	26	440
	5	3	94.77	417	23	440
	5	4	95.23	419	21	440

Table 3 *(continue)*

Number of strike(s)/flick(s) (T)	Number of states	Number of Gaussian mixtures	Accuracy (%)	H	S	N
	5	5	95.00	418	22	440
	5	6	95.00	418	22	440
	6	2	94.55	416	24	440
	6	3	96.14	423	17	440
	6	4	96.14	423	17	440
	6	5	96.14	423	17	440
	6	6	95.91	422	18	440
	7	2	95.45	420	20	440
	7	3	95.91	422	18	440
	7	4	96.59	425	15	440
	7	5	96.59	425	15	440
	7	6	96.36	424	16	440

When using 1 through 4 strikes/flicks, the highest accuracy of 95.23%, 96.82%, 96.82%, and 97.05% were respectively obtained by using 6 states and 5 Gaussian mixtures per state (6S5M). For 5 strikes/flicks, the highest accuracy of 96.59% was yielded when using 7 states and 5 Gaussian mixtures per state (7S5M). However, when using one strike or flick with HMMs containing 7 states, there were nine striking/flicking sounds that the system was not provided the recognition results because the number of the feature vectors extracted from very short signals was not adequate to be recognized using the HMMs. Next, the results were further investigated, and the confusion matrix derived from recognizing the untrained set was shown in Table 4.

For the untrained set, the highest accuracy was derived when using 5 strikes/flicks. The results showed high watermelon, guava, and durian recognition rates. The errors occurred when recognizing pineapple flicking sounds. The errors occurred when recognizing pineapple flicking sounds. There were 1.89% pineapples incorrectly recognized as watermelons.

Table 5 shows the confusion matrix of fruit-type recognition obtained from the unknown set.

The results revealed that although the striking and flicking sounds look similar, the proposed method could correctly distinguish durian striking sounds from those of watermelons, guavas, and pineapples. Although there were some errors when recognizing the flicking sounds of the different types of fruits, the overall accuracy was higher than 90%. When using 4 flicks, the highest recognition accuracy rates of 91%, 99.33%, 97.27%, and 100% were achieved for watermelons, guavas, pineapples, and durians, respectively.

Table 4
Confusion matrix (untrained set)

Number of strike(s)/flick(s)	Actual type of fruit	Recognized as			
		Durian	Watermelon	Guava	Pineapple
1 strike/flick (6S5M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	98% (98)	0% (0)	2% (2)
	Guava	0% (0)	0.67% (1)	99.33% (149)	0% (0)
	Pineapple	0% (0)	3.64% (4)	0% (0)	96.36% (106)
2 strikes/flicks (6S4M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	98% (98)	0% (0)	2% (2)
	Guava	0% (0)	0% (0)	98.67% (148)	1.33% (2)
	Pineapple	0% (0)	0.91% (1)	0% (0)	99.09% (109)
3 strikes/flicks (6S4M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	99% (99)	0% (0)	1% (1)
	Guava	0% (0)	0% (0)	98.67% (148)	1.33% (2)
	Pineapple	0% (0)	0.91% (1)	0% (0)	99.09% (109)
4 strikes/flicks (6S4M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	99% (99)	0% (0)	1% (1)
	Guava	0% (0)	0% (0)	100% (150)	0% (0)
	Pineapple	0% (0)	3.64% (4)	0% (0)	96.36% (106)
5 strikes/flicks (6S4M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	100% (100)	0% (0)	0% (0)
	Guava	0% (0)	0% (0)	100% (150)	0% (0)
	Pineapple	0% (0)	1.89% (2)	0% (0)	98.18% (108)

Table 5
Confusion matrix (unknown set)

Number of strike(s)/flick(s)	Actual type of fruit	Recognized as			
		Durian	Watermelon	Guava	Pineapple
1 strike/flick (6S5M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	93% (93)	0% (0)	7% (7)
	Guava	0% (0)	0% (0)	98.67% (148)	1.33% (2)
	Pineapple	0% (0)	7.27% (8)	3.64% (4)	89.10% (98)
2 strikes/flicks (6S5M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	94% (94)	0% (0)	96% (6)
	Guava	0% (0)	0% (0)	98.67% (148)	1.33% (2)
	Pineapple	0% (0)	3.64% (4)	1.82% (2)	94.55% (104)
3 strikes/flicks (6S5M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	93% (93)	0% (0)	7% (7)
	Guava	0% (0)	0% (0)	98.67% (148)	1.33% (2)
	Pineapple	0% (0)	2% (3)	1.33% (2)	96.67% (145)
4 strikes/flicks (6S5M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	91% (91)	0% (0)	99% (9)
	Guava	0% (0)	0% (0)	99.33% (149)	0.67% (1)
	Pineapple	0% (0)	0.91 (1)	1.82% (2)	97.27% (107)
5 strikes/flicks (6S5M)	Durian	100% (100)	0% (0)	0% (0)	0% (0)
	Watermelon	0% (0)	90% (90)	0% (0)	10% (10)
	Guava	0% (0)	0% (0)	99.33% (149)	0.67% (1)
	Pineapple	0% (0)	1.82% (2)	1.82% (2)	96.36% (106)

DISCUSSION

Even though the types of fruits can be manually provided for recognizing the fruit quality, it will be better if we can simultaneously recognize the types of fruits and their quality from the striking and flicking signals. The duration of the flicking and striking sounds of the fruits is very short, and the sounds resemble. Therefore, the method to recognize the different types of fruits should be studied. HMMs can model striking and flicking signals for recognizing the quality of fruits and fruit types. Nowadays, the prices of fruits are much higher when compared with those in the past decade, which makes non-destructive fruit grading more important. According to the derived results and previous studies (Phoophuangpairoj, 2014a, 2014b), it revealed some possibility of creating an application that can recognize the quality of durians and watermelons from the striking ad flicking sounds without recognizing the types of fruits from images or manually giving the type of fruits in advance.

When using HMMs, if the number of feature vectors extracted from the signals is insufficient for HMM states, the HMM decoder will not give the result. Therefore, it is suggested that the large number of HMM states is inappropriate for recognizing short signals such as 1 striking or flicking sound.

In the future, CNN, which has been used in several computer vision applications and will be more widely used in processing sequential data, including natural language processing and speech recognition (Kiranyaz et al., 2021), long short-term memory networks (LSTM), and deep learning techniques should be explored along with frequency domain features such as MFCCs for the recognition of the fruits from striking and flicking sounds.

CONCLUSION

This paper proposes using preprocessing, acoustic models, a dictionary, and grammar to recognize the fruit types from flicking/striking sounds. The dictionary and grammar provide flexibility to design the recognition system and can be used to recognize arbitrary duration of flicking/striking sounds. The parameters to extract acoustic features, including the window size, have to be adjusted to fit the problem. The preprocessing acoustic models, dictionaries, and grammar have to be designed based on the characteristic of the striking and flicking sounds. The results when using the different number of flicking and striking sounds, number of states, and number of Gaussian mixtures were compared. The method could correctly differentiate durian striking sounds from watermelon, guava, and pineapple flicking sounds. Averagely, more than 95% of recognition accuracy was obtained from recognizing striking and flicking sounds. The findings shed light on the feasibility of recognizing the durian ripeness of fruits and watermelon flesh from the flicking and striking sounds without image processing.

ACKNOWLEDGEMENT

The author thanks everyone who had helped and supported completing the study and publication. The author also thanks the reviewers for their dedication, timing, and fruitful comments in improving and increasing the quality of the manuscript. Sincerely, the author expresses gratitued to Rangsit University, Thailand, for supporting the publication fee.

REFERENCES

Albarrak, K., Gulzar, Y., Hamid, Y., Mehmood, A., & Soomro, A. B. (2022). A deep learning-based model for date fruit classification. *Sustainability*, *14*(10), Article 6339. https://doi.org/10.3390/ su14106339

Asriny, D. M., Rani, S., & Hidayatullah, A. F. (2020). Orange fruit images classification using convolutional neural networks. *IOP Conference Series: Materials Science and Engineering, 803*(1), Article 012020. https://doi.org/10.1088/1757-899X/803/1/012020

Chavan, R. S., & Sable, G. S. (2013). An overview of speech recognition using HMM. *International Journal of Computer Science and Mobile Computing*, *2*(6), 233-238.

Elharati, H. A., Alshaari, M., & Këpuska, V. Z. (2020) Arabic speech recognition system based on MFCC and HMMs. *Journal of Computer and Communications*, *8*(3), 28-34. https://doi.org/10.4236/jcc.2020.83003

Fan, S., Li, J., Zhang, Y., Tian, X., Wang, Q., He, X., Zhang, C., & Huang, W. (2020). On line detection of defective apples using computer vision system combined with deep learning methods. *Journal of Food Engineering*. *286,* Article 110102. https://doi.org/10.1016/j.jfoodeng.2020.110102

Hatala, Z., & Puturuhu, F. (2021). Viterbi algorithm and its application to Indonesian speech recognition. *Journal of Physics: Conference Series*, *1752*(1), Article 012085. https://doi.org/10.1088/1742-6596/1752/1/012085

Hossain, M. S., Al-Hammadi, M., & Muhammad, G. (2018). Automatic fruit classification using deep learning for industrial applications. *IEEE Transactions on Industrial Informatics*, *15*(2), 1027-1034. https://doi.org/10.1109/TII.2018.2875149

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, *151*, Article 107398. https://doi.org/10.1016/j.ymssp.2020.107398

Naithani, K., Thakkar, V. M., & Semwal, A. (2018, August 22-24). *English language speech recognition using MFCC and HMM*. [Paper presentation]. International Conference on Research in Intelligent and Computing in Engineering (RICE), Salvador, El Salvador. https://doi.org/10.1109/RICE.2018.8509046

Najkar, N., Razzazi, F., & Sameti, H. (2010). A novel approach to HMM-based speech recognition systems using particle swarm optimization. *Mathematical and Computer Modelling, 52*(11-12), 1910-1920. https://doi.org/10.1016/j.mcm.2010.03.041

Phoophuangpairoj, R. (2013). Determining guava freshness by flicking signal recognition using HMM acoustic models. *International Journal of Computer Theory and Engineering*, *5*(6), 877-884. https://doi.org/10.7763/IJCTE.2013.V5.815

Phoophuangpairoj, R. (2014a). Automated classification of watermelon quality using non-flicking reduction and HMM sequences derived from flicking sound characteristics. *Journal of Information Science and Engineering*, *30*(4), 1015-1033.

Phoophuangpairoj, R. (2014b). Computerized unripe and ripe durian striking sound recognition using syllable-based HMMs. *Applied Mechanics and Materials*, *446-447*, 927-935. https://doi.org/10.4028/www.scientific.net/amm.446-447.927

Phoophuangpairoj, R. (2014c). Durian ripeness striking sound recognition using N-gram models with N-best lists and majority voting. In S. Boonkrong, H. Unger & P. Meesad (Eds), *Recent Advances in Information and Communication Technology: Proceedings of the 10th International Conference on Computing and Information Technology (IC2IT2014)* (pp. 167-176). Springer

Phoophuangpairoj, R., & Srikun, N. (2014). Computerized recognition of pineapple grades using physicochemical properties and flicking sounds. *International Journal of Agricultural and Biological Engineering*, *7*(3), 93-101.

Phoophuangpairoj, R. (2020, October 21-22). *Recognizing breathing sounds using HMMs and grammar.* [Paper presentation]. Proceedings of the 5th International Conference on Information Technology (InCIT2020), ChonBuri, Thailand. https://doi.org/10.1109/InCIT50588.2020.9310966

Raja S. L., Ambika, N, Divya, V., & Kowsalya, T, (2018). Fruit classification system using computer vision: A review. *International Journal of Trend in Research and Development*, *5*(1), 22-26.

Shahi, T. B., Sitaula, C., Neupane, A., & Guo, W. (2022). Fruit classification using attention-based MobileNetV2 for industrial applications. *Plos One 17*(2), Article e0264586. https://doi.org/10.1371/journal. pone.0264586

Stevner, A. B. A., Vidaurre, D., Cabral, J., Rapuano, K., Nielsen, S. F. V., Tagliazucchi, E., Laufs, H., Vuust, P., Deco, G., Woolrich, M. W., Someren, E. V. & Kringelbach (2019). Discovery of key whole-brain transitions and dynamics during human wakefulness and non-REM sleep. *Nature Communications, 10*(1), Article 1035. https://doi.org/10.1038/s41467-019-08934-3

Zeng, G. (2017, October 3-5). *Fruit and vegetables classification system using image saliency and convolutional neural network*. [Paper presentation]. *IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC)*, Chongqing, China. https://doi.org/10.1109/ITOEC.2017.8122370.