Determining Guava Freshness by Flicking Signal Recognition Using HMM Acoustic Models

Rong Phoophuangpairoj

Abstract—Being able to determine the freshness or quality of fruit automatically is significant because people in the world consume fruit. Countless fruit buyers can be disappointed when they purchase stale, old or sub-standard produce. Studying and developing a computerized method that helps to determine the freshness of fruit without cutting, destroying or tasting is interesting because it could be of benefit to people worldwide. A method using non-flicking reduction preprocessing and acoustic models of different freshness levels is proposed to recognize fresh and not fresh guava flicking signals. In the recognition process, first, the non-flicking parts of the signals are reduced. Then, spectral features of the signals are extracted. Finally, 1) acoustic models are created using Hidden Markov Models (HMM), 2) acoustic sequences of fresh and not fresh guavas are defined and 3) defined possible freshness recognition results are applied to determine guava freshness. The proposed method resulted in average correct freshness recognition rates of 92.00%, 88.00% and 94.00% from fresh, 3 and 6-day-kept guava unknown test sets, respectively. Average correct freshness recognition rates of 90.00%, 90.67%, 92.00%, 92.00% and 92.00% were obtained when using one through five flicks, respectively. An average recognition time of less than 50 milliseconds was taken when using any number of flicks from one to five. The results indicate that the proposed method using three to five flicks is time-efficient and accurate enough to be used to determine the quality of guavas.

Index Terms—Guava, guava freshness, flicking signals, acoustic models, different freshness levels, freshness recognition, HMM.

I. INTRODUCTION

Food, including agricultural produce, is essential for everyday life. Selecting agricultural produce from supermarket shelves or produce stands is routine for shoppers around the globe. If it were possible to ensure that produce was fresh, less fruit would be discarded. For several kinds of fruit, it is difficult for buyers to determine the freshness or ripeness of the fruit from the external appearance. Sounds generated by flicking may be a useful indicator of the conditions inside some agricultural produce. It is hoped that in the future, with the help of tablets or smart phones, buyers can accurately choose fresh and good quality fruit. Furthermore, the fruit industry will have an automated system that can recognize large quantities of fruit such as guavas not only by size but also by freshness quality.

Signal processing methods have been studied and applied in various fields. For example, there has been research that applied signal processing to animal sounds [1]-[3]. In speech recognition, digitized human speech signals are converted to words or text using: signal processing, a pronunciation dictionary and domain recognition grammar or language models. HMM is one of the most efficient techniques used in speech recognition to create acoustic speech models. HMM of phoneme and syllable units are created and applied to speech recognition systems [4], [5]. Mel Frequency Cepstral Coefficients (MFCCs) are high-performance acoustic features that have been widely used to recognize speech signals [6]-[8]. Additionally, MFCC-based acoustic features and pitch contours have been applied to speech recognition in tonal languages such as Thai and Cantonese [9]-[11]. In emotion recognition, acoustic features including fundamental frequencies, spectral features, energy features and their augmentations have been studied to recognize emotional states of the human voice [12]. In gender classification, fundamental frequencies and MFCCs have been widely used [13]-[16]. The duration of human speech segments has also been studied for gender classification [17].

A computerized method that uses flicking sounds to recognize the freshness of guavas and the effect the number of flicks has on the guava freshness recognition rate is interesting because it has not yet been studied. In this research, flicking sounds are investigated to determine the freshness of guavas. A guava freshness recognition method using acoustic models of different freshness levels is proposed to achieve a high freshness recognition rate within an acceptable amount of time.

II. BACKGROUND INFORMATION RELATING TO FLICKING AND GUAVA FLICKING SIGNALS

To create an understanding of guava freshness recognition, background information relating to flicking and guava flicking signals is provided. Flicking is moving the index or middle finger off the thumb against an object, as illustrated in Fig. 1. Flicking may be a practical method that can be used to determine the freshness of guavas. Therefore, flicking sounds are required to be collected to assess their suitability for determining guava quality.

The flicking signals consist of two parts, namely, a non-flicking and a flicking part. The non-flicking part is longer and contains no or a small amount of spectral frequency information. Contrarily, the much shorter flicking part contains much more valuable spectral information that may be used to differentiate between fresh guavas and not fresh ones. Fig. 2 shows one-flick signal and the duration of the flicking part and non-flicking part resulting from guava flicking.
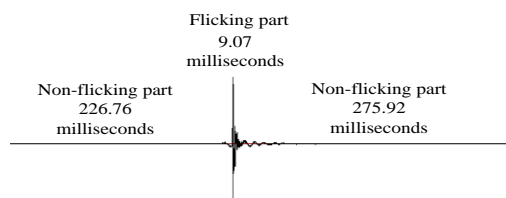
Fig. 1. Flicking a guava.

Fig. 2. One-flick signal resulting from guava flicking.

As shown in the figure, the duration of the guava flicking part is only 9.07 milliseconds and at several points, the duration is shorter than 7 milliseconds. It is difficult to capture the spectral information because the flicking part is of such short duration. Therefore, using more than one flick to determine guava freshness may result in a higher freshness recognition rate. Signal processing methods are required to be studied to develop a computerized freshness recognition method that can be used to capture freshness information and efficiently determine the freshness of guavas from short duration flicking signals.

III. PROPOSED METHOD

The proposed method is composed of 3 stages: 1) preprocessing signals using non-flicking reduction, 2) extracting acoustic features from flicking signals, and 3) recognizing fresh and not fresh flicking signals, as shown in Fig. 3.
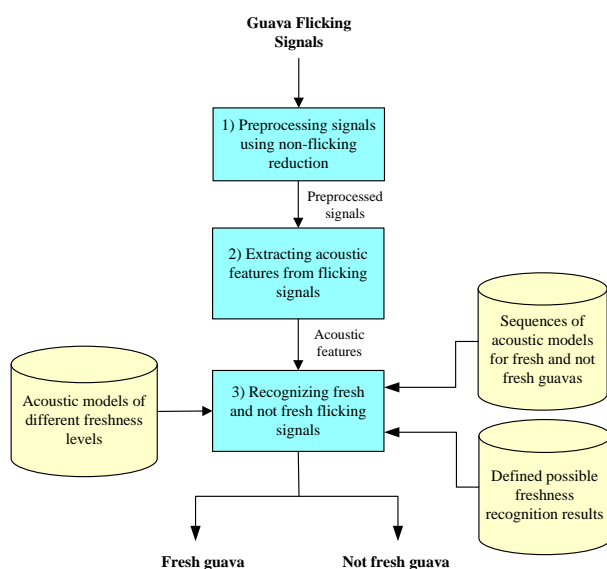

Fig. 3. Proposed method.

Before the recognition, acoustic models of different freshness levels and data for freshness recognition consisting

of the sequences of acoustic models for fresh and not fresh guavas and the defined possible freshness recognition results are prepared. At the first stage of the process, long non-flicking parts are reduced. At the second stage, acoustic features, which are MFCCs and their delta and accelerator coefficients, are extracted from the guava flicking signals. At the final stage, the freshness of the guava is determined using the created acoustic models and freshness recognition data.

A. Acoustic Models of Different Freshness Levels

To determine the freshness of guavas, acoustic models of different freshness levels are created using 1st and 2nd level not fresh guava flicking signals, as illustrated in Fig. 4.
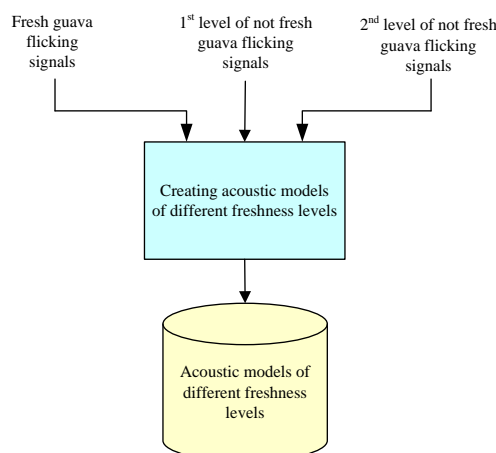

Fig. 4. Creating acoustic models of three different freshness levels.

Instead of using a single not fresh acoustic model, 1st and 2nd level not fresh acoustic models are used in the proposed method to reduce acoustic model variation and improve the freshness recognition rate. In this work, flicking signals recorded from fresh guavas and guavas that were kept on ice for three and six days, represent fresh guava flicking signals and 1st and 2nd level not fresh guava flicking signals, respectively. The acoustic models are created using HMM. To create the acoustic models, flicking signals and their transcription without the matched positions between signals and acoustic model labels are used. Since the duration of the guava flicking part is quite short, whole flicking parts are used to create acoustic models of different freshness levels. For example, five flicking sounds derived from fresh guavas prepared for the acoustic model creation are transcribed as "sil FRESH sil FRESH sil FRESH sil FRESH sil FRESH sil". Five-flicking sounds obtained from sub-standard 3-day-kept guavas are transcribed as "sil NOTFRESH3 sil NOTFRESH3 sil NOTFRESH3 sil NOTFRESH3 sil NOTFRESH3 sil" and five-flicking sounds obtained from 6-day-kept guavas are transcribed as "sil NOTFRESH6 sil NOTFRESH6 sil NOTFRESH6 sil NOTFRESH6 sil NOTFRESH6 sil". The sil (silent part) represents each non-flicking part, while FRESH, NOTFRESH3 and NOTFRESH6 represent each flicking part of three different freshness levels. To create the acoustic models the non-flicking parts of the flicking signals are reduced during the preprocessing. Then the acoustic features are extracted from the preprocessed signals. Finally, the obtained acoustic features with the transcription are used to train three different freshness levels of acoustic models and a silent model (non-flicking part model). After the acoustic models are obtained, they are further applied to define

sequences of acoustic models for fresh and not fresh guavas.

B. Sequences of Acoustic Models for Fresh and Not Fresh Guavas

Before recognizing guava freshness, different sequences of acoustic models representing fresh and not fresh guava flicking signals are defined based on acoustic models of different freshness levels. The flicking sound characteristics

and the allowed number of flicks are considered when creating the acoustic sequences for fresh and not fresh guavas. Typically, to test internal characteristics, it is not necessary to flick fruit more than five times. Five sequences for fresh guavas and ten sequences for not fresh guavas are defined to handle variation in the number of flicks from one to five. The defined sequences are shown below.

FRESH	sil FRESH sil
NOTFRESH	sil NOTFRESH3 sil
NOTFRESH	sil NOTFRESH6 sil
FRESH	sil FRESH sil FRESH sil
NOTFRESH	sil NOTFRESH3 sil NOTFRESH3 sil
NOTFRESH	sil NOTFRESH6 sil NOTFRESH6 sil
FRESH	sil FRESH sil FRESH sil FRESH sil
NOTFRESH	sil NOTFRESH3 sil NOTFRESH3 sil NOTFRESH3 sil
NOTFRESH	sil NOTFRESH6 sil NOTFRESH6 sil NOTFRESH6 sil
FRESH	sil FRESH sil FRESH sil FRESH sil FRESH sil
NOTFRESH	sil NOTFRESH3 sil NOTFRESH3 sil NOTFRESH3 sil NOTFRESH3 sil
NOTFRESH	sil NOTFRESH6 sil NOTFRESH6 sil NOTFRESH6 sil NOTFRESH6 sil
FRESH	sil FRESH sil FRESH sil FRESH sil FRESH sil FRESH sil
NOTFRESH	sil NOTFRESH3 sil NOTFRESH3 sil NOTFRESH3 sil NOTFRESH3 sil NOTFRESH3 sil
NOTFRESH	sil NOTFRESH6 sil NOTFRESH6 sil NOTFRESH6 sil NOTFRESH6 sil NOTFRESH6 sil

Fresh sequences of one through five flicks are defined using sil and FRESH acoustic models while the not fresh sequences are defined using sil, NOTFRESH3 and NOTFRESH6 acoustic models. As the system can manage thousands of allowed sequences for fresh and not fresh guavas, additional fresh and not fresh flicking sequences may be added to support more flicks.

C. Defined Possible Freshness Recognition Results

To determine whether a guava is fresh or not, in the syntax below, only "FRESH" and "NOTFRESH" are defined as the only allowed possible freshness recognition results.

```
$FreshLevel = FRESH|NOTFRESH;
($FreshLevel)
```

After the defined possible freshness recognition results are prepared, preprocessing is applied to reduce the non-flicking parts first.

D. Preprocessing Signals Using Non-flicking Reduction
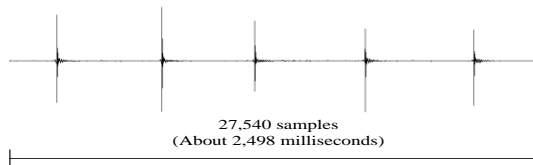

27,540 samples
(About 2,498 milliseconds)

Fig. 5. Guava flicking signals (before preprocessing).

Currently, it is possible to model both the non-flicking and flicking parts of the signals and use them to model whole signals resulting from flicking by applying the same signal processing techniques that are used in continuous speech recognition. However, the difference in duration between non-flicking and flicking parts means that it is difficult to automatically create accurate HMM acoustic models. Hence, a preprocessing method consisting of 5 steps is proposed to reduce the non-flicking parts of the guava flicking sounds. In

the first step, the number of samples and sample values are read from a digitized guava flicking sound file, as shown in Fig. 5.

In the second step, the number of samples in each frame based on the defined frame size is computed using the equation below.

$$NS = \frac{SF}{1000} \times FS \qquad (1)$$

NS: Number of samples in each frame
SF: Sampling frequency (11,025 Hz)
FS: Frame size used for preprocessing (10 milliseconds)

After that, the number of frames in the flicking sound file is computed using the equation below.

$$NF = \frac{NA}{NS} \qquad (2)$$

NF: Number of frames in a flicking sound file
NA: Number of all samples in a flicking sound file
(Obtained from reading the header of the digitized flicking sound file in the first step)

In the third step, the sum of the amplitudes of clipped samples (*SA*) found in each frame is computed. Clipping signals helps to reduce the amplitude variation of the signals, which makes it simpler to set a threshold that differentiates between non-flicking and flicking parts. To compute the *SA* in each frame, the amplitude or absolute value of each sample is calculated first, using the following equation.

$$A(k) = \left| S(k) \right| \qquad (3)$$

A(k) : the amplitude of the kth sample value in the digitized flicking signals
S(k) : the kth sample value in the digitized flicking signals

The *SA* of the ith frame is computed using the equation below.

$$SA_i = \sum_{k=ixNS}^{(i+1)xNS-1} Clip(A(k)) \qquad 0 \le i \le NF-1 \qquad (4)$$

$Clip(A(k))$: the value of the clipped amplitude of the kth sample (obtained using a clipping threshold (σ_{Clip}))

Then, the mean of the frame (*MFA*) amplitude is found for the whole signal using the equation below.

$$MFA = \frac{\sum_{i=0}^{NF-1} SA_i}{NF} \qquad (5)$$

The algorithm for the third step is shown below.

```
MFA = 0;

for i=0 to NF-1 step by 1
{
k = i * NS;
SAi = 0.0;

for j=0 to NS-1 step by 1
{
if(Frame[i].data[j] ≤ σClip)
Frame[i].data[j] = A[k+j];
else
Frame[i].data[j] = σClip;

SAi = SAi + Frame[i].data[j];
}

MFA = MFA+SAi;
}
 MFA = MFA/NF;
```

This algorithm is used to calculate SA_i of the 0th through $(NF-1)^{th}$ frames. Next, the MFA is computed from the obtained SA_i. The variable Frame[i].data[j] is used to represent the amplitude data of the jth sample value in the ith frame. A σ_{Clip} (of 10,000) is used to clip the amplitude or absolute values of the samples in the signals that are higher than the σ_{Clip}.

In the fourth step, the preprocessing gathers the information required for the reduction of the non-flicking signals. The algorithm is shown below.

```
for i=0 to NF-1 step by 1
{
if (SAi ≥ σFTh x MFA)
F[i] = 1;
else
F[i] = 0;
}
```

For each frame, σ_{FTh}, which can be equal to 3, is used together with the computed *MFA* to discriminate between flicking and non-flicking frames. The ith frame that has a SA_i higher than or equal to (σ_{FTh} x *MFA*) is designated as a flicking frame ($F[i]$ is set to 1). Otherwise, it is designated as a non-flicking frame ($F[i]$ is set to 0).

In the final step, the flicking frames are kept and the non-flicking frames that are not adjacent to the flicking frames are removed. Then, the reduced non-flicking signals are derived as shown in Fig. 6.

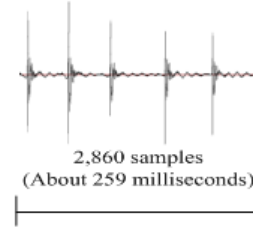2,860 samples
(About 259 milliseconds)

Fig. 6. Guava flicking signals after preprocessing

The duration of the signals after preprocessing is only about 259 milliseconds, which is much shorter than the 2,498 milliseconds taken for the original signals (shown in Fig. 5). In training, after the preprocessing, the signals are further used to extract acoustic features and train acoustic models of different freshness levels while in the recognition, they are used to extract acoustic features and determine the guava freshness.

E. Extracting Acoustic Features from Flicking Signals

Unlike the human voice, which consists of sounds produced by the vibration of vocal cords, fundamental frequencies or pitch contours cannot be accurately computed from flicking sounds. Therefore, MFCCs and their derivatives are used as acoustic features. As the duration of guava flicking signals is quite short, a 5-ms frame size with a 1-ms frame shift interval is used in the feature extraction. Firstly, a pre-emphasis coefficient of 0.97 and the Hamming window are applied. Then, the Fast Fourier Transform (FFT) is used to compute the frequency spectra of the flicking signals. Next, the log amplitudes of the spectra are mapped onto the Mel scale using a filter bank with 26 channels. Later, the discrete cosine transform (DCT) is applied to obtain 12 MFCCs, and then the energy is calculated. Finally, the first and second derivatives of the MFCCs and the energy are computed. 39-dimension acoustic features, consisting of 12 MFCCs with energy and their 1st and 2nd order derivatives are obtained then used for fresh and not fresh flicking signal recognition.

F. Recognizing Fresh and Not Fresh Flicking Signals

To recognize fresh and not fresh guavas, methods, HMM acoustic models of different freshness levels are connected according to the sequences of the acoustic models for fresh and not fresh guavas and defined possible freshness recognition results to create possible recognition paths. The path that has the highest probability is determined and its corresponding freshness recognition result is used as the final result.

IV. EXPERIMENTAL RESULTS

Experiments were conducted to evaluate the proposed method. In the experiments, guava flicking sounds were collected from 100 guavas and recorded using the 16-bit PCM format at 11,025 Hz. First, five-flick sounds were collected from 50 guavas for training. Then, after 3 days, they were flicked again to obtain five more flick sounds from each guava. Finally, after six days, five more flick sounds were collected

from each of the guavas. For testing, there were two sets: untrained and unknown. The untrained set was recorded from the 50 guavas that were used in training, but the sounds were collected by flicking at different times. The unknown set was recorded from the remaining 50 guavas that were not included in the training set. For both untrained and unknown sets, each guava was flicked from one to five times, respectively. The preprocessing algorithm was developed using Microsoft Visual C++. The Hidden Markov Toolkit (HTK) [18] was used to extract the acoustic features, train the HMM acoustic models of the three different freshness levels and determine the freshness of the guavas. In the experiments, HMM acoustic models, comprising of three emitting states with two Gaussian mixtures per state, were used for the freshness recognition. The experimental results are reported in 3 parts: 1) the duration of flicking signals before and after the

preprocessing, 2) freshness recognition rates, and 3) freshness recognition time.

A. Duration of Flicking Signals before and after the Preprocessing

The average duration of flicking signals before and after the preprocessing is shown in Table I. Before the preprocessing, the average duration of one through to five flicks was 523.71, 886.11, 1208.29, 1474.51 and 1786.73 milliseconds, respectively. After the preprocessing, the average duration of one through five flicks decreased to 36.28, 69.28, 104.03, 138.52 and 173.17 milliseconds, respectively. The results show that the proposed preprocessing reduces the non-flicking parts and makes the duration of non-flicking parts similar to that of the flicking parts, which results in more accurate acoustic models.

TABLE I: AVERAGE DURATION OF FLICKING SIGNALS BEFORE AND AFTER THE PREPROCESSING

Number of Flicks	Average Duration (Milliseconds)	
	Before Preprocessing	After Preprocessing
1	523.71	36.28
2	886.11	69.28
3	1208.29	104.03
4	1474.51	138.52
5	1786.73	173.17

B. Freshness Recognition Rates

The effect the number of flicks from the proposed method had on the freshness recognition rate was investigated and compared to the method using only fresh and not fresh acoustic models. The fresh model was created using flick signals collected from fresh guavas, whereas the not fresh

model was created by combining flicking signals collected from 3- and 6-day-kept not fresh guavas. The sequences of fresh and not fresh guavas were defined based on the fresh and not fresh models. The freshness recognition rates of the untrained set are shown in Table II.

TABLE II: COMPARISON OF FRESHNESS RECOGNITION RATES (UNTRAINED SET)

Number of Flicks	Correct Freshness Recognition Rate (%)							
	Using Fresh and Not Fresh Acoustic Models				Using Fresh, 3- and 6-day-kept Not Fresh Acoustic Models			
	Fresh Guava	Not Fresh Guava		Average	Fresh Guava	Not Fresh Guava		Average
		3-day-kept	6-day-kept			3-day-kept	6-day-kept	
1	84.00%	78.00%	96.00%	86.00%	84.00%	90.00%	92.00%	88.67%
2	88.00%	78.00%	100.00%	88.67%	88.00%	90.00%	94.00%	90.67%
3	88.00%	80.00%	98.00%	88.67%	90.00%	90.00%	98.00%	92.67%
4	88.00%	80.00%	98.00%	88.67%	90.00%	90.00%	98.00%	92.67%
5	90.00%	80.00%	98.00%	89.33%	90.00%	90.00%	98.00%	92.67%
Average	87.60%	79.20%	98.00%	88.27%	88.40%	90.00%	96.00%	91.47%

When using the fresh and not fresh acoustic models, average correct freshness recognition rates of 86.00%, 88.67%, 88.67%, 88.67% and 89.33% were obtained by flicking the guavas one through five times, respectively.

When using the fresh, 3- and 6-day-kept not fresh acoustic models, higher average correct freshness recognition rates of 88.67%, 90.67%, 92.67%, 92.67% and 92.67% were achieved for one through five flicks, respectively. The

findings show that using only one or two flicks to determine guava freshness may not be efficient because it results in lower freshness recognition rates. The proposed method using three to five flicks resulted in a higher average correct freshness recognition rate of 92.67%. For all numbers of flicks, the method using fresh and not fresh acoustic models resulted in an average correct freshness recognition rate of 88.27%, whereas the proposed method using fresh 3- and 6-day-kept acoustic models achieved a significantly higher average freshness recognition rate of 91.47%. Average correct freshness recognition rates from the fresh, 3- and 6-day-kept guavas were 88.40%, 90.00% and 96.00%, respectively. When using the fresh and not fresh acoustic models, the not fresh model was created from flicking signals obtained from both 3- and 6-day-kept not fresh guavas. The

created not fresh model had more acoustic variation than the separately created 3- and 6-day-kept not fresh models. The acoustic variation made it difficult to obtain accurate acoustic models and resulted in lower freshness recognition rates. The results show that the proposed method, using two separate 3- and 6-day-kept not fresh models, could achieve much higher correct average recognition rates than using only one not fresh acoustic model to recognize the freshness of sub-standard 3-day-kept guavas (90.00% vs. 79.20%). The results indicate that the proposed method, which used acoustic models of different freshness levels, was better than the method that used only fresh and not fresh acoustic models. Next, the proposed method was evaluated using the unknown set, as shown in Table III.

TABLE III: COMPARISON OF FRESHNESS RECOGNITION RATES (UNKNOWN SET)

Number of Flicks	Correct Freshness Recognition Rate (%)							
	Using Fresh and Not Fresh Acoustic Models				Using Fresh, 3- and 6-day-kept Not Fresh Acoustic Models			
	Fresh Guava	Not Fresh Guava		Average	Fresh Guava	Not Fresh Guava		Average
		3-day-kept	6-day-kept			3-day-kept	6-day-kept	
1	92.00%	64.00%	92.00%	82.67%	92.00%	84.00%	94.00%	90.00%
2	90.00%	66.00%	96.00%	84.00%	90.00%	88.00%	94.00%	90.67%
3	92.00%	68.00%	94.00%	84.67%	92.00%	90.00%	94.00%	92.00%
4	92.00%	68.00%	94.00%	84.67%	92.00%	90.00%	94.00%	92.00%
5	94.00%	66.00%	94.00%	84.67%	94.00%	88.00%	94.00%	92.00%
Average	92.00%	66.40%	94.00%	84.13%	92.00%	88.00%	94.00%	91.33%

When using the fresh and not fresh acoustic models, average correct freshness recognition rates of 82.67%, 84.00%, 84.67%, 84.67% and 84.67% were obtained by flicking guavas one through five times, respectively. When using the fresh, 3- and 6-day-kept not fresh acoustic models, higher average correct freshness recognition rates of 90.00%, 90.67%, 92.00%, 92.00% and 92.00% were achieved for one through five flicks, respectively. This is in parallel with the findings from the untrained set, as the results from the unknown set also show that using only one or two flicks to determine guava freshness results in low correct recognition rates. The proposed method using three to five flicks resulted in a higher average correct freshness recognition rate of 92.00%. For all numbers of flicks, the method using fresh and not fresh acoustic models resulted in an average correct freshness recognition rate of 84.13%, whereas the proposed method using the fresh, 3- and 6-day-kept not fresh acoustic models achieved a significantly higher average freshness recognition rate of 91.33%. Average correct freshness recognition rates from the fresh, 3- and 6-day-kept guavas were 92.00%, 88.00% and 94.00%, respectively. The proposed method achieved significantly higher average correct recognition rates when determining the freshness of sub-standard 3-day kept guavas (88.00% vs. 66.40%). For both the untrained and unknown sets, the proposed method

was better than the method that used the fresh and not fresh acoustic models. Additionally, the results indicate that three to five flicks yields higher freshness recognition rates than only one or two flicks.

A support Vector Machine (SVM) was also used as a comparison to evaluate the efficiency of the proposed method. The 39-dimension acoustic features, comprising of 12 MFCCs with energy as well as their 1st- and 2nd order derivatives, were used. In both training and testing, the acoustic features were extracted from a single 20-ms part of the flicking signal that had the highest sum of the amplitude of sample values. The SVM was trained and the freshness of the guavas was recognized using the LIBSVM [19]. Using the radial basis SVM, which is one of the most efficient and widely used SVMs, the recognition rates of 69.33%, 70.00%, 69.33%, 69.33% and 68.33% were obtained from the untrained set when using one to five flicks, respectively. For the unknown set, recognition rates of 68.00%, 68.67%, 67.33%, 67.33% and 68.00% were obtained when using one to five flicks, respectively. The correct recognition rates of around 70.00% were obtained using the SVM with the acoustic features extracted from the flicking part having the highest amplitude. The baseline method using the SVM gave lower recognition rates than the proposed method, which achieved the significantly higher average correct recognition

rates of 91.33%.

C. Freshness Recognition Time

The time recognition process is divided into three parts: 1) preprocessing guava flicking signals, 2) extracting acoustic features and 3) determining guava freshness. The recognition time of the proposed method was measured and averaged from both the untrained and unknown sets. The results are shown in Table IV.

TABLE IV: AVERAGE FRESHNESS RECOGNITION TIME

Number of Flicks	Average Time Taken to Preprocess Guava Flicking Signals (Milliseconds)	Average Time Taken to Extract Acoustic Features (Milliseconds)	Average Time Taken to Determine Guava Freshness (Milliseconds)	Average Total Time (Milliseconds)
1	15.30	3.77	3.08	22.15
2	16.16	5.67	6.13	27.96
3	16.99	7.49	9.61	34.09
4	17.41	9.51	12.80	39.72
5	18.21	10.81	16.28	45.30

The average time taken to preprocess flicking signals was 15.30, 16.16, 16.99, 17.41 and 18.21 milliseconds and the average time taken to extract acoustic features was 3.77, 5.67, 7.49, 9.51 and 10.81 milliseconds for one through five flicks, respectively. For any number of flicks from one through five, the average time spent on determining guava freshness was less than or equal to 16.28 milliseconds. The average total time was 22.15, 27.96, 34.09, 39.72 and 45.30 milliseconds for one through five flicks, respectively. The results indicate that the proposed guava freshness recognition method is time-efficient.

V. CONCLUSIONS

Freshness recognition by flicking sounds is a practical method that can be used to determine the quality of guavas because it requires no cutting, slicing or tasting. It is difficult to use guava flicking parts to discern guava freshness because they are audible for only a short period of time. Therefore, a method that used the preprocessing and acoustic models of different freshness levels was proposed to recognize flicking sounds. When two different not fresh acoustic models were used in place of a single not fresh acoustic model, acoustic variation was reduced and higher overall freshness recognition rates were achieved. The proposed method was more accurate than the method that used only fresh and not fresh acoustic models. For the unknown test set, an average correct guava freshness recognition rate of 92.00% was obtained when the number of flicks was three, four and five. For 6-day-kept guavas, a higher average correct guava freshness recognition rate of 94.00% was achieved. The results show that using only one or two flicks to determine guava freshness yields lower recognition rates than using three to five flicks. An average total time of less than or equal to 45.30 milliseconds was taken to recognize the guava freshness. The relatively high guava freshness recognition rates and the relatively short amount of time needed to quantify the freshness of the guava demonstrate that the proposed computerized method is both viable and accurate enough to be used to determine guava quality reliably.

REFERENCES

[1] C. Y. Yeo, S. A. R. Al-Haddad, and C. K. Ng, "Animal voice recognition for identification (ID) detection system," in *Proc. the IEEE 7th International Colloquium on Signal Processing and Its Applications*, 2011, pp. 198-201.

[2] D. Mitrovic, M. Zeppelzauer, and C. Breiteneder, "Discrimination and retrieval of animal sounds," in *Proc. the 12th International Multi-Media Modelling Conf.*, 2006, pp. 339-343.

[3] G. Guo and Z. Li, "Content-based classification and retrieval by support vector machines," *IEEE Trans. on Neural Networks*, vol. 14, pp. 209-215, 2003.

[4] S. Tangwongsan and R. Phoophuangpairoj, "Boosting Thai syllable speech recognition using acoustic models combination," in *Proc. the International Conf. on Computer and Electrical Engineering*, 2008, pp. 568-572.

[5] S. Tangruamsub, P. Punyabukkana, and A. Suchato, "Thai speech keyword spotting using heterogeneous acoustic modeling," in *Proc. the IEEE International Conf. on Research, Innovation and Vision for the Future*, 2007, pp. 253-260.

[6] A. Deemagarn and A. Kawtrakul, "Thai connected digit speech recognition using hidden Markov models," in *Proc. the 9th International Conf. on Speech and Computer*, 2004.

[7] L. Fuhai, M. Jinwen, and D. Huang, "MFCC and SVM based recognition of Chinese vowels," *Lecture Notes in Computer Science*, vol. 3802, pp. 812-819, 2005.

[8] R. Phoophuangpairoj, "Using multiple HMM recognizers and the maximum method to improve voice-controlled robots," in *Proc. the International Conf. on Intelligent Signal Processing and Communication Systems*, 2011.

[9] S. Tangwongsan, P. Po-Aramsri, and R. Phoophuangpairoj, "Highly efficient and effective techniques for Thai syllable speech recognition," *Lecture Notes in Computer Sciences*, vol. 3321, pp. 259-270, 2004.

[10] N. Thubthong and B. Kijsirikul, "Tone recognition of continuous Thai speech under tonal assimilation and declination effects using half-tone model," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 6, pp. 815-825, 2001.

[11] T. Lee, W. Lau, Y. W. Wong, and P. C. Ching, "Using tone information in Cantonese continuous speech recognition," *ACM Trans. on Asian Language Information Proc. (TALIP)*, vol. 1, no. 1, pp. 83-102, 2002.

[12] D. Ververidis and C. Kotropoulos, "Automatic speech classification to five emotional states based on gender information," in *Proc. the European Signal Processing Conf.*, 2004, vol. 1, pp. 341-344.

[13] R. Phoophuangpairoj, S. Phongsuphap, and S. Tangwongsan, "Gender identification from Thai speech signal using a neural network," *Lecture Notes in Computer Science*, vol. 5863, pp. 676-684, 2009.

[14] H. Ting, Y. Yingchun and W. Zhaohui, "Combining MFCC and pitch to enhance the performance of the gender recognition," in *Proc. the 8th International Conf. on Signal Processing*, 2006.

[15] S. M. R. Azghadi, M. R Bonyadi, and H. Sliahhosseini, "Gender classification based on feedforward backpropagation neural Network," *IFIP International Federation for Information Proc.*, vol. 247, pp. 299-304, 2007.

[16] M. H. James and J. C. Michael, "The role of F0 and formant frequencies in distinguishing the voices of men and women," *Attention, Perception, & Psychophysics*, vol. 71, no. 5, pp. 1150-1166, 2009.

[17] M. Sigmund, "Gender distinction using short segments of speech signal," *International Journal of Computer Science and Network Security*, vol. 8, no. 10, pp. 159-162, 2008.

[18] The hidden Markov model toolkit (HTK). [Online]. Available: http://htk.eng.cam.ac.uk/

[19] C. C. Chang and C. J. Lin. (2011). LIBSVM: a library for support vector machines. *ACM Trans. on Intelligent Systems and Technology* [Online]. 2(3). Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

Rong Phoophuangpairoj graduated from Chulalongkorn University with a B. Eng, he also has an M.Sc and a Ph.D. from Mahidol University, Bangkok, Thailand. Currently, he is employed by the Rangsit University Computer Engineering Department where he lectures for the Electrical and Computer Engineering Master Degree Program. He has published several research papers focused on: speech recognition, gender classification and signal processing. His research interests include speech processing, multimodal interaction with users, signal processing in language learning and other applications.