NEW ACOUSTICAL PATTERN RECOGNITION APPROACH TO IDENTIFY DIFFERENT STAGES OF A COOKING PROCESS. THE BOILING WATER CASE.

M. Tabacchi (1), C. Asensio (2), I. Pavón (2) and M. Recuero (2)

(1) Hibbs & Associates Pty Ltd, Unit 48, 378 Parramatta Rd, Homebush, NSW 2140, Australia
(2) Universidad Politécnica de Madrid (I2A2), ETSI TGC, Campus Sur – UPM, Ctra. Valencia Km. 7, 28031 Madrid, Spain

ABSTRACT

Although pattern recognition technique has been largely used in many fields, it seems that very few studies have applied this technique to cooking processes. In this preliminary research, a new methodology has been developed and tested on a simple case of water boiling. Besides defining and analysing the efficacy and the performance of a statistical pattern recognition approach when applied to different signals (sound and vibration), an optimisation module has been proposed to boost the classification rates by adding syntactical analysis that enables to consider the inertia of the process. In the specific case of boiling water, almost 100% successful recognition has been reached. These results prove the validity of this methodology, opening new research lines for new scenarios such as different cooking process, acoustically polluted environments, sensors optimisation, etc.

INTRODUCTION

Pattern recognition currently comprises a vast body of methods supporting the development of numerous applications and researches in many different areas such as civil and environmental engineering, industrial process control, communication science, etc. In the field of acoustics, pattern recognition has been applied to speech recognition (Fazel & Chakrabartty 2011; Garner 2011), environmental noise sources detection and classification (Gaunard et al. 1998; Cowling & Sitte 2003). However, very few studies have applied acoustic pattern recognition to detect different stages of a cooking process (Gutierrez et al.; Doney 1994). Although these studies try to exploit acoustic signals to recognise different stages of a cooking process, none of them seems to stabilise the results taking into account the intrinsic inertia of a generic cooking process.

For this reason, besides defining and analysing the efficacy of a pattern recognition methodology, and its performance when applied to different signals (sound or vibration), this research proposes a new optimisation module that improves the classification rates by adding a syntactical analysis of the phenomenon to classify. In order to analyse the potentials of this classification system a very simple case of water boiling was first studied. The optimisation module in this case took into account the inertia of the process.

METHODOLOGY

Figure 1 outlines the process followed to fulfil the aims of this research.

Figure 1. Outline of the whole process.

The water-boiling phenomenon was recorded by a microphone near the pot and by an accelerometer mounted on the stove. The recordings of the signals were manually labelled into the 4 stages of boiling water described below. Then, the significant signal features were extracted to train and test a statistical classification system. To improve the results, the classifier's results were optimised taking into account the characteristics of the full boiling process.

Audio and vibration recordings

The water-boiling phenomenon was recorded by a cardioid microphone (AKG – SE300B, frequency range 20-20000 Hz, sampling rate 44100 Hz) located 50 cm from the cooking vessel with an angle of 45º with respect to the vertical axis and by accelerometer (model 352C33, frequency range 0.5-10Khz, sampling rate 44100 Hz) placed right in the centre of the cooking induction stove.

As this project is the very first approach for future research and in order to reduce the influence of the side factors, a simplified and a very specific case study were considered:

- 1.5 litres of distilled water;
- power of the induction stove set on boost (maximum);
- water at room temperature;
- enamelled cookware of 18 cm diameter.

The measurements were undertaken filling the uncovered cook- ware with 1.5 l of distilled water at room temperature and heating up at boost power. The duration of each measurement was approximately 7–8 min to measure the whole boiling process without letting all the water evaporate. Finally, 28 experiments were undertaken measuring audio and vibration at the same time (i.e. 28 recordings of two channels).

Labelling

All the boiling recordings were split from the beginning into two groups, one to be used for training (19 boiling recordings) and another one used for testing (9 boiling recordings). 100ms samples were extracted from the recordings and then labelled into one of these 4 classes of boiling water boiling (i.e. heating, nucleate boiling, transition boiling and film

boiling). A total of 70650 and 33000 samples were available for training and testing of the system respectively.

Feature selection and classificiation

Many studies and articles show that the formation of bubbles in a heated liquid like water often causes cavitations and acoustic effects (Nesis 2008). In particular, each boiling water stage presents a unique acoustic pattern that can be used to discriminate one stage of the phenomenon from the others (Lawrence 2008). Once we know that there are some acoustic differences between the sound events, or, in our case, between the 4 stages of water boiling, this information needs to be extracted and analysed using feature vectors. Mel-frequency cepstrum coefficients (MFCC) are generally employed for speech recognition as they are based on human auditory perception (Lee et al. 2003; Sahidullah & Saha 2012; Kraaijveld 1996). However, since previous studies (Crowling & Sitte 2003; Asensio, Ruiz & Recuero) have demonstrated that they can be successfully used even for non-speech sound recognition (e.g. environmental aircraft noise), we decided to exploit them for our case study. MFCC were derived from each input samples previously labelled using an extended bandwidth from 0 Hz to 15kHz and considering 20 coefficients instead of the more usual 13 coefficients. This extraction technique works quite well in this case. In fact, for each coefficient the probability density function of each class can be easily distinguished from the others.

To compare the performance of the different input signals, features were extracted from:
1. audio (20 features);
2. vibration (20 features).

These two separate sets of features were used as input for the classification.

Several nonlinear supervised training algorithms such as Parzen, ANN or SVC were applied to the training samples. The training and testing processes were carried out using PRtool for Matlab. The Parzen classifier was finally deemed to be the more suitable for the case study.

Optimisation of the results

For each labelled sample in the testing dataset, the trained classifier estimates its probability of belonging to each class, yielding for each sample 4 outcomes varying from 0 to 1. The high fluctuation of these results leads the detection system to be a little bit unstable and unreliable especially because it does not fit the slow variations of the boiling water phenomenon well.

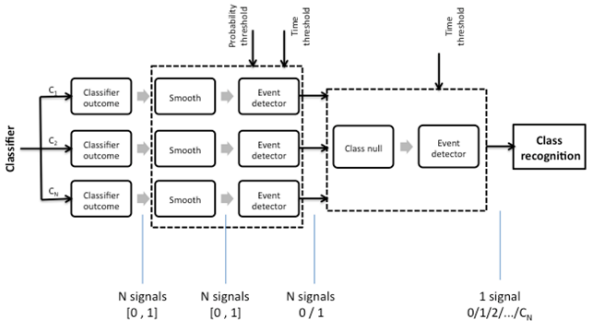

Figure 2. Outline of the optimisation system.

In the scenario, an optimisation of recognition system is needed, bearing in mind the inertia of the phenomenon we are trying to represent. Figure 2 shows the basic outline of the optimisation system used.

Boiling is a pseudo-stationary process with long-term changes. It is not worth classifying the labelled samples separately, because the time correlation of the samples is very strong. So the time sequence of each of them needs to be considered, hence the posterior probabilities for each sample to belong to each one of the 4 classes will be analysed as a soft output of the classifier with strong time dependency.

For this reason, to avoid sudden and unrealistic changes in the classes and probabilities (Figure 3, dotted blue line), each of the four soft outputs of the classifier was smoothed with a moving average (Figure 3, red line).

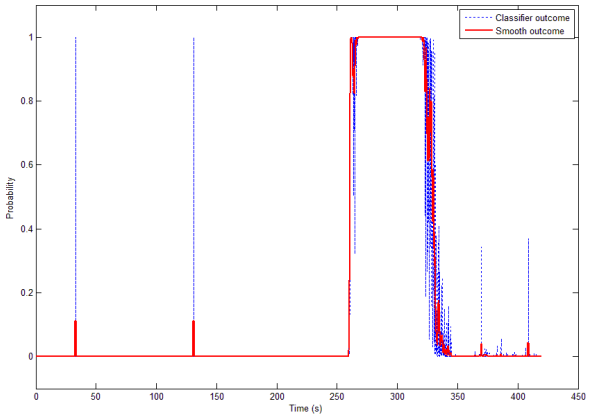

Figure 3. Smoothing of the classifier estimations.

Afterwards, a threshold detector was applied in order extract events from the smooth outcome. This module detects a class-event if its probability exceeds a certain threshold (in this case 0.5) during more than a defined time interval (in this case 1 second). The output of this block is a Boolean signal that takes the value 1 when an event is detected and 0 if it is not (Figure 4).

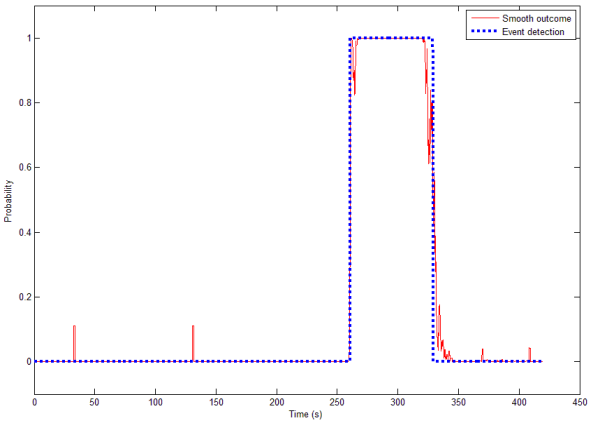

Figure 4. Class-event detector.

After applying the described constraints, the events detector marks the events of each class for steady time intervals, when the boiling class is certain. But due to the strict constraints, it can be observed that sometimes for a specific time interval no class-event is detected, and the class to be assigned is uncertain. Therefore a null class was created to state that in these particular instants the classification system is unable to decide about the specific boiling stage with sufficient accuracy. Due to the inertia of the water boiling process, it is not possible to get quick and unsteady stage changes. Taking ad-

vantage of this, a null-class event will be triggered if the "uncertainty interval" lasts for more than one second (i.e. more than 10 samples of 100ms). If the uncertainty of the system is sporadic (less than a second), the system will keep the class previously detected. Only when a null-class event is triggered will the optimiser output show "class 0" as result.

This optimisation process improves the understanding and the interpretation of the phenomenon of boiling water adjusting the results to its inertia. The outcomes of this system are therefore more stable and reliable, boosting the opportunity to implement this recognition system for future developments.

Testing

Different types of tests were undertaken to find out the best signal (audio and vibration) and especially to determine the improvements of the optimisation of the recognition process.

In the first test, we only wanted to check the efficacy of the trained classifier itself. Therefore, we simply applied the trained Parzen classifier to the testing dataset of labelled samples.

In the second test, we wanted to check the improvements achieved with the optimisation, so all the classification processes explained above were applied to the same labelled and sequential samples. It is worth remembering that only the parts of the signal far from the transitions from one boiling stage to another were considered. In this way we avoid any possible errors during the evaluation of the system deriving from incorrectly labelled samples.

Therefore, a third test was needed to see how the recognition system developed actually behaves in the stage transitions. For this purpose, the whole input signal of each recording was analysed comparing, even in this case, the results yielded with and without the optimisation.

RESULTS AND DISCUSSION

Audio

Table 1 shows the results of the test for audio input signal where the columns represent the classified classes and the rows the real classes. It can be seen that even without the optimisation the percentage of successful recognition reaches 99% especially in the first (heating) and fourth class (film boiling). However, with the optimisation system all the classes achieve a better recognition ratio of up to 100% for the first and fourth class. The misclassification (<2%) of the second (nucleate boiling) and third (transition boiling) class is mainly due to the boundary problems described previously.

Table 1. Results of the test (%) using audio.

Real class	Classified class 1	2	3	4	Null
Before optimisation					
1	*99.25*	0.68	0.01	0.06	-
2	0.00	*98.12*	1.86	0.02	-
3	0.00	1.90	*98.10*	0.00	-
4	0.00	0.35	0.32	*99.32*	-
After classification					
1	*100.00*	0.00	0.00	0.00	0.00
2	0.00	*98.90*	0.86	0.00	0.24
3	0.00	1.70	*98.30*	0.00	0.00
4	0.00	0.00	0.00	*100.00*	0.00

Figure 5 presents the results of the test run on the entire input signal of each recording. The first row represents the root mean square of the audio signal. The second, third, fourth and fifth rows represent the probabilities of the signal belonging respectively to first (heating), second (nucleate boiling), third (transition boiling) and fourth (film boiling) class, while the sixth row stands for the final decision of the classification system. The figure below highlights the outcomes due to the trained classifier only (without optimisation). We can observe that the output of the system is quite unstable (for instance time 260-270, or 300-310), showing sporadic misclassification samples not only in the boundary (which might be more acceptable) but also in the middle of a stage. Moreover, the system mixes up not only one class with the next or previous ones but also with further ones (e.g. confusing the fourth class with the first). This outcome may lead the system to be somehow unreliable.

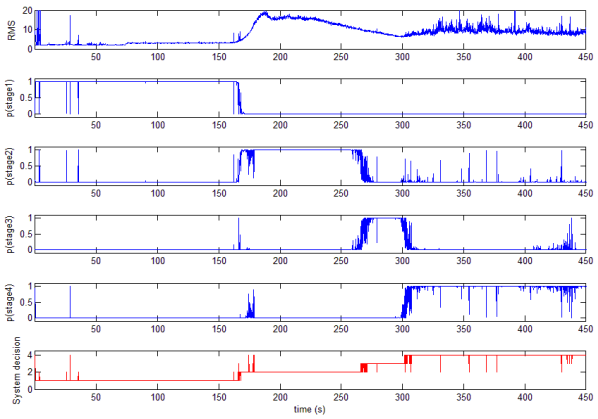

Figure 5. Outcome of the recognition without optimisation (audio input).

In this scenario, the need for an optimisation can be easily understood. In fact, Figure 6 presents the classification after applying the optimisation. As can be seen in the figure, the outcome is perfect: there is a clear transition from a boiling stage to the next and there is not a single class misclassified. Although the limit among stages is fuzzy, the system takes feasible and reliable decisions, removing instantaneous or sporadic changes in the output.

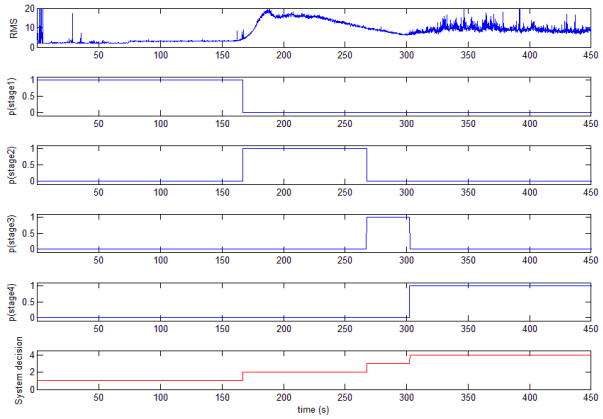

Figure 6. Outcome of the recognition with optimisation (audio input).

Vibration

Table 2 shows the results of the test for the vibration signal where the columns represent the classified classes and the rows the real classes. It can be seen that even without the optimisation the percentage of successful recognition reaches

at least 99% in the first class (heating). However, with the optimisation system all the classes increase the recognition ratio up to 100% for the fourth class. The misclassification (<3%) of the second (nucleate boiling) and third (transition boiling) class is mainly due to the boundary problems described above.

Table 2. Results of the test (%) using vibration.

Real class	Classified class				Null
	1	2	3	4	
Before optimisation					
1	*99.77*	0.22	0.00	0.01	-
2	0.00	*96.77*	3.23	0.00	-
3	0.00	0.00	*94.60*	5.40	-
4	0.00	0.11	1.67	*98.22*	-
After classification					
1	*99.79*	0.10	0.00	0.00	0.11
2	0.00	*97.44*	2.32	0.00	0.24
3	0.00	0.00	*97.80*	2.20	0.00
4	0.00	0.00	0.00	*100.00*	0.00

Figure 7 presents the results of the test run on the entire input signal of each recording. The first row represents the root mean square of the vibration signal. The second, third, fourth and fifth rows represent the probabilities of the signal belonging respectively to first (heating), second (nucleate boiling), third (transition boiling) and fourth (film boiling) class, while the sixth row stands for the final decision of the classification system. This figure highlights the outcomes due to the trained classifier only. We can observe that the output of the system is quite unstable detecting the wrong classes not only in the boundary (which might be more acceptable) but also in the middle of the class.

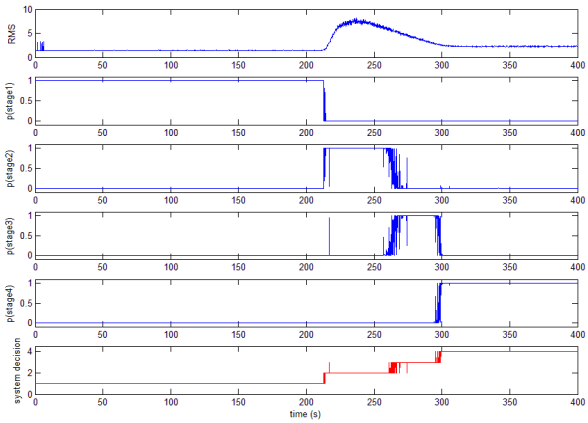

Figure 7. Outcome of the recognition without optimisation (vibration input).

After applying the optimisation, the result is outstanding. In fact all the transitions from one class to another are clearly defined and, above all, no misclassified class is detected in the middle of the class (Figure 8).

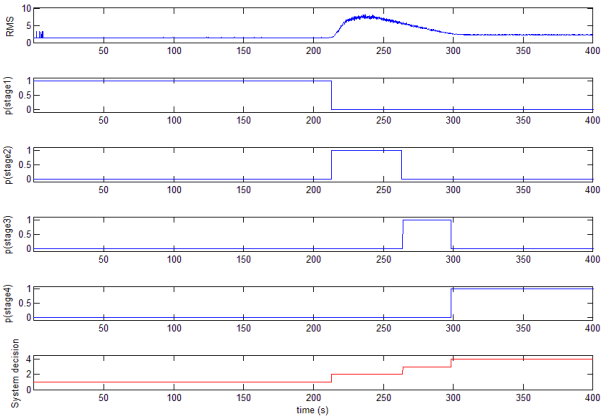

Figure 8. Outcome of the recognition with optimisation (vibration input).

The reason why the vibration cannot give as many good results as the audio can be found in the sample labelling. In fact, since the 2 channel samples (audio and vibration) used for training and testing the system were labelled considering only the audio time signal, it is possible that some vibration parts of the samples had not been properly labelled. This may explain the reduced recognition rates that can be noted especially in the second and third stage.

CONCLUSION

After analysing the results, we can state that the developed pattern recognition system can be a useful tool to detect different stages of a cooking process such as the boiling water with very high recognition rates. In general for both audio and vibration, we reach almost 100% of successful recognition for the first (heating) and fourth (film boiling) class. In particular, the best recognition rates can be achieved by using audio signal as input. In this case, the misclassification of the second (nucleate boiling) and third (transition boiling) class is mainly due to a boundary problem. As the transition from one class to another is not defined crisply, some samples might have been wrongly labelled and may be mistaken for the next or previous boiling stage. For this reason, more research is needed to find a more precise way to define boundaries (e.g. using temperature sensors, pressure sensors, etc.).

Nonetheless, the optimisation has been demonstrated to be the key to improve the recognition rates. For all the different inputs and in all the classes, the optimisation increases up to 2% the recognition rate.

With this very simple case of study, we have demonstrated that our optimisation module can avoid all the sudden and unrealistic changes of stages and yielding a more reliable and stable outcomes. This is the very first step of our research that aims to analyse and exploit the potentials of using this acoustic patter recognition system for a real detection of different stages of a generic cooking process.

REFERENCES

Asensio, C, Ruiz, M, & Recuero, M 2010, 'Real-time aircraft noise likeness detector", *Applied Acoustics*, vol. 71, no. 6, pp. 539-545.

Cowling, M & Sitte, R 2003, 'Comparison of techniques for environmental sound recognition', *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2895–2907.

Doney, GD 1994, 'Acoustic Boiling Detection', PhD thesis, Massachusetts Institute of Technology.

Fazel, A & Chakrabartty, S 2011, 'An Overview of Statistical Pattern Recognition Techniques for speaker verification', *IEEE Circuits and system magazine*.

Garner, PN 2011, 'Cepstral normalisation and the signal to noise ratio spectrum in automatic speech recognition', *Speech Communication*, vol. 53, no. 8, pp. 991–1001.

Gaunard, P, Mubikangiey, GC, Couvreur, C & Fontaine, V 1998, 'Automatic classification of environmental noise events by hidden Markov models', *Applied Acoustics*, vol. 54, no. 3, pp. 187–206.

Gutierrez, F, Paris, L, Gil, L, Gadea, P, Peman, R & Tabuenca, S, *Dispositivo de aparato de cocción*, Spain Patent P2011317652011.

Kraaijveld, MA 1996, 'A parzen classifier with an improved robustness against deviations between training and test data', *Pattern Recognition Letters*, vol. 17, no. 7, pp. 679-689.

Lawrence, R 2008, *Fundamental of speech recognition*, Pearson Education.

Lee, C, Hyun, D, Choi, E, Go, J, & Lee, C 2003. 'Optimizing feature extraction for speech recognition', *IEEE Trans Speech Audio Process*, vol. 11, no. 1, pp. 80-87.

Nesis, YI 1990, 'Acoustic noise of a boiling liquid', *Heat Transfer – Sov Res*, vol. 22, no. 6, pp. 789-795.

Sahidullah, M, & Saha, G 2012, 'Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition', *Speech Communications*, vol. 54, no. 4, pp. 543-565.