

## Research Article

# An Active Feature Selection Strategy for DWT in Artificial Taste

Tao Liu , Yanbing Chen, Dongqi Li, and Mengya Wu

College of Communication Engineering, Chongqing University, No. 174 Shazheng Street, Shapingba District, Chongqing 400044, China

Correspondence should be addressed to Tao Liu; [cquliutao@cqu.edu.cn](mailto:cquliutao@cqu.edu.cn)

Received 30 November 2017; Revised 28 February 2018; Accepted 13 March 2018; Published 10 April 2018

Academic Editor: Maria Luz Rodríguez-Méndez

Copyright © 2018 Tao Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A discrete wavelet transform (DWT) extracts meaningful information in a time-frequency domain and is a favorable feature extraction approach from pulse-like responses in large pulse voltammetry (LAPV) electronic tongues (e-tongue). A regular DWT generates lots of coefficients to describe signal details and approximations at different scales. Thus, coefficient selection is necessary to reduce the feature size. However, the common DWT-based feature selection follows a passive mode: manipulation through human experience or exhaustive trials. It is subjective, time consuming, and barely works in nonlaboratory conditions. In this paper, we present an active feature selection strategy consisting of a dispersion ratio computation and optimal searching search. To evaluate the performance of the proposed method, we prepared several beverage samples and performed experiments with a LAPV e-tongue. Meanwhile, the features of raw response, peak-inflection point, referenced DWT method, and our proposed method were presented to indicate the effects of the refined features of the proposed method. Furthermore, we utilized several classifiers such as the  $k$ -nearest neighbor ( $k$ -NN), support vector machine (SVM), and random forest (RF) to evaluate the improvement of recognition by the refined features. Compared with other regular feature extraction methods, the proposed method can automatically explore high-quality features with an acceptable feature size. Moreover, the highest average accuracy was achieved by the proposed method for each classifier. It is an alternative feature extraction approach for a LAPV e-tongue without any manipulation in real applications.

## 1. Introduction

An artificial taste system named electronic tongue (e-tongue) has become a potential approach for liquid-phase evaluations [1, 2]. A sensor array and the proper pattern recognition algorithm are the two main parts of an e-tongue. The sensor array imitates a human's taste cells to sense substances, while the pattern recognition algorithm functions as the human brain to handle judgments. Compared with traditional chemical devices, e-tongues have evident advantages including lower cost, lower latency, and simpler operations. Wide applications, such as honey identification [3], rice discrimination [4], and beverage classification [5, 6], have been a concern in recent years. In beverage classification, scholars mainly focus their attention on the substances with specific aromatic flavors such as tea and liquor [7, 8] since the e-tongue identifications are more objective and reproducible than human judgments [9].

The pattern recognition part composed of classification and feature extraction is an important section in an e-tongue system. Various classifiers attract concerns in e-tongue studies. The principal component analysis (PCA) [10, 11], support vector machine (SVM) [12, 13], autoregressive-modeling technique [14], linear discriminant analysis [6, 15], and back-propagation neural network [13, 16] have appeared to discriminate different analytes. On the other hand, feature extraction absorbs meaningful features from sensor responses for classifiers. There are regular feature extraction methods for large pulse voltammetry (LAPV) [17–19], a common e-tongue type with pulse-like responses from sensors. Features of e-tongue have two categories: geometric and time-frequency approaches. For the geometric method, the PCA can obtain valuable features according to the correlation of the sensor responses [20]. Furthermore, some scholars used peak and inflection values of raw response curves as features to express signal characteristics

[10, 21]. Sometimes, even the original sensor response was used directly as geometric features for classifiers [22]. Meanwhile, time-frequency analysis is another popular means for e-tongue feature extraction. The discrete wavelet transform (DWT) is a promising technology due to its capacity for multiscale analysis [23, 24]. Yin et al. introduced the power ratio to DWT for feature refining and optimization [25]. Considering that the e-tongue system utilized in experiments follows the large pulse voltammetry (LAPV) mode and generates a series of multifrequency pulse-like responses, we pay more attention to feature extraction by the DWT approach. Normally, DWT would present a great deal of coefficients (features) along with the increase of the decomposition level, thus a feature selection process is needed to remove the redundant and meaningless DWT coefficients. However, a conventional DWT feature selection depends on human manipulation. The favorable features are achieved according to either human experience or exhaustive trials. Subjective decision and massive computation would hinder e-tongue systems from handling real industry applications. Accordingly, we need to explore an objective and simple strategy to guide DWT feature selection under unmanned conditions.

In this study, we proposed a DWT feature extraction method with an active feature selection strategy (AFS-DWT) to obtain suitable features from DWT coefficients. Both dispersion ratio computation and optimal searching stage are introduced to AFS-DWT. Meanwhile, we performed beverage experiments with a LAPV e-tongue to classify red wine, white spirit, beer, black tea, oolong tea, maofeng tea, and pu'er tea samples. Recognition models such as the  $k$ -nearest neighbor ( $k$ -NN) [26], support vector machine (SVM) [27], and random forest (RF) [28] were adopted in evaluations to explore the performance of feature extraction methodologies. The experimental results show that AFS-DWT is qualified in unmanned working conditions and shows similar performance to the best one from the reference methods. This paper is composed of five sections: Section 1 introduces the literature and background of this study. Section 2 describes the hardware of our LAPV e-tongue. The recognition and feature extraction methods used in this study are presented in Section 3. Section 4 shows the experimental results and analysis. Finally, some conclusions are summarized in the last section.

## 2. e-Tongue System

As shown in Figure 1, we designed a LAPV e-tongue composed of an electrolytic cell, an electrode sensor array, a LAPV transformation board, a control unit, and a higher-level computer. Eight metal electrodes fixed on the top of the electrolytic cell are selected to form a three-electrode system. The details of the electrode sensor array are described in Table 1. The LAPV transformation board has two functions: one is generating excitation signals to the auxiliary and reference electrode sensors synchronously and the other one is transforming LAPV responses of the working electrode sensors from microcurrents to proper voltages. The control unit receives the transformed LAPV responses and handles

response sampling, data storage, data transferring, and excitation signal control via an onboard 32 bit microcontroller. The sampling circuit consists of a 6-channel analog-to-digital converter with a 16 bit resolution to corresponding working electrode sensors. We save the digitalized data in a TF card for further processing on a higher-level computer. Particularly, the original excitation signal is generated from a digital-to-analog converter integrated on the microcontroller; as a result, the frequency, amplitude, and time interval of excitation signals are programmable.

When the LAPV e-tongue is working, we use a ribbon cable to transfer transformed LAPV responses and excitation signals between the LAPV transformation board and control unit. Shielded cables and bayonet-nut connectors link the electrode sensor array and LAPV transformation board to reduce signal interference. Algorithm implementation and evaluation are ordinarily performed on a higher-level computer (desktop or laptop). We also designed external memories on the control unit for the microcontroller to execute lightweight algorithms if necessary.

## 3. Materials and Methods

**3.1. Principal Component Analysis.** PCA is widely used in dimensionality reduction and data visualization [29]. It performs orthogonal transformation by converting a group of correlated signals into uncorrelated linear components.

In practice, eigenvalue decomposition is utilized to compute principal components (PCs) and their loadings. Here, we assume that the PCs denote the uncorrelated linear components while the loadings mean the amplitudes of PCs. Thus, it is possible to choose a small amount of PCs to generally express high-dimensional data in low-dimensional space. The PCs with the largest two or three loadings are customarily selected to visualize original data in 2D or 3D space, respectively.

**3.2. Peak-Inflection Point Method.** Considering the geometric shapes of signals, the LAPV e-tongue responses regularly consist of a series of peaks and troughs. Therefore, it is reasonable to take some representative points from electrode responses as characteristics. A typical one is picking up four points in each pulse cycle: one peak point, one valley point, and two inflection points as response features [19], and we name this feature extraction approach the peak-inflection point method (PIPM). This geometric access for feature extraction is very convenient, and the selected points intuitively contain distinct characteristics of e-tongue responses.

**3.3. Discrete Wavelet Transform.** DWT handles time-frequency analysis for discrete signals. Its use is considerable in multiscale analysis when working on digital data. This method divides original signals to approximation and the detailed parts through low-frequency and high-frequency filters, respectively. The resolution of approximation and detailed parts depend on the DWT decomposition level while the lengths of approximation and detailed parts are halved according to those of the parent

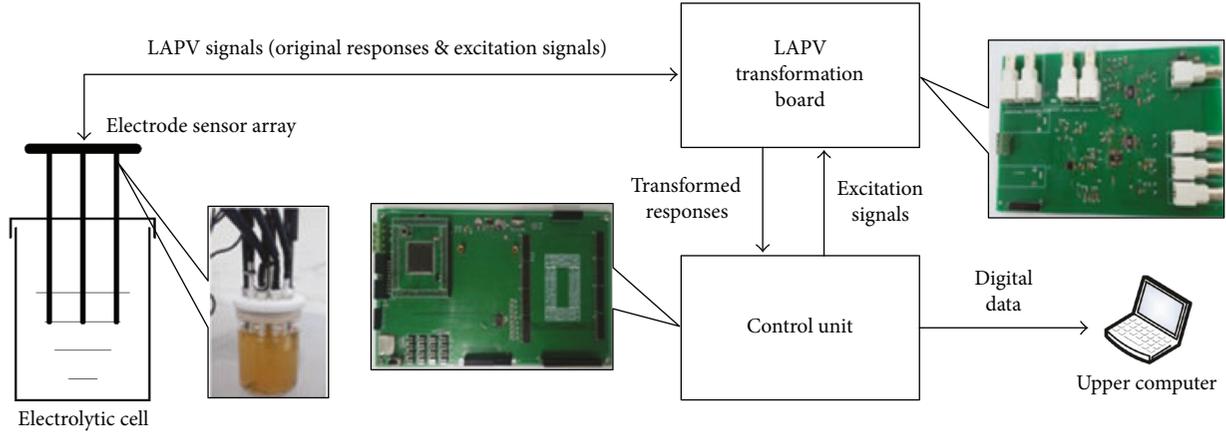


FIGURE 1: Structure of a LAPV e-tongue.

TABLE 1: Details of an electrode sensor array.

Electrode sensor types	Materials
Auxiliary electrode sensor	Pillar platinum
Reference electrode sensor	Ag/AgCl Gold Platinum
Working electrode sensor	Palladium Titanium Tungsten Silver

coefficients. As shown in Figure 2, the DWT is implemented using Mallat's pyramidal algorithm [30]. CA1 and CD1 denote the approximation and detailed DWT coefficients of the first decomposition level. In the second decomposition level, CAA2 and CAD2 are the approximation and detailed DWT coefficients from CA1 while CDA2 and CDD2 are the ones from CD1.

Considering that the pulse-like responses of the LAPV mode have rich frequency components, the DWT method fits such a case due to the ability of varied time-scale analyses.

**3.4. Relative Power Ratio of Discrete Wavelet Transform.** A relative power ratio (RPR) can be calculated from each DWT coefficient according to a previous study [25]; we call this feature extraction strategy RPR-DWT. Assume that  $C_{ij}$  represents the  $j$ th DWT coefficient in the  $i$ th decomposition level. We can compute the power of  $C_{ij}$  as follows:

$$\begin{aligned} e_{ij} &= \|C_{ij}\|^2 \\ r_{ij} &= \frac{e_{ij}}{\sum_j e_{ij}}, \end{aligned} \quad (1)$$

where  $e_{ij}$  and  $r_{ij}$  denote the power and RPR of  $C_{ij}$  respectively. In feature extraction, the maximum RPR of each decomposition level is calculated as a candidate feature and then the best one is selected from among the candidates as

features. We should note that RPR-DWT is a refined method for DWT coefficients, the coefficient selection process still relies on a passive mode: human designation based on either experience or exhaustive calculation.

**3.5. Active Feature Selection Strategy for Discrete Wavelet Transform.** As far as we are concerned, when DWT is executed as a feature extraction method, lots of wavelet coefficients are generated as the decomposition level is increased. Therefore, choosing the necessary features from DWT coefficients becomes crucial for the effectiveness and conciseness of the following classifications. The current strategy to solve this issue depends on manual selection. However, this manner is detrimental for the practicability of DWT. In this part, we introduce an active feature selection (ASF) strategy for DWT coefficients. There are two phases contained in the ASF process.

**3.5.1. Phase One: Dispersion Ratio Computation for DWT Coefficients.** In order to reduce and provide a uniform feature dimensionality, we use RPR to preprocess each DWT coefficient obtained from digital responses. For the  $k$ th category, we separate the training samples into two sections:  $k$  and non- $k$  categories. Suppose that  $M$  and  $N$  are the sample sizes of the  $k$  and non- $k$  categories, respectively. We calculate inner-class distance  $S_w^{ij}$  and inter-class distance  $S_b^{ij}$  of  $C_{ij}$  as follows:

$$\begin{aligned} S_w^{ij} &= \frac{1}{M(M-1)} \sum_{p \in k} \sum_{q \in k} \|r_{ij}^p - r_{ij}^q\|_2, \quad p \neq q, \\ S_b^{ij} &= \frac{1}{M \cdot N} \sum_{p \in k} \sum_{q \notin k} \|r_{ij}^p - r_{ij}^q\|_2, \end{aligned} \quad (2)$$

where  $r_{ij}^p$  and  $r_{ij}^q$  denote the RPR of the  $p$ th and  $q$ th samples in the corresponding categories, respectively. Thus, the dispersion ratio  $S_{ij}$  of  $C_{ij}$  is defined as follows:

$$S_{ij} = \frac{S_w^{ij}}{S_b^{ij}}, \quad (3)$$

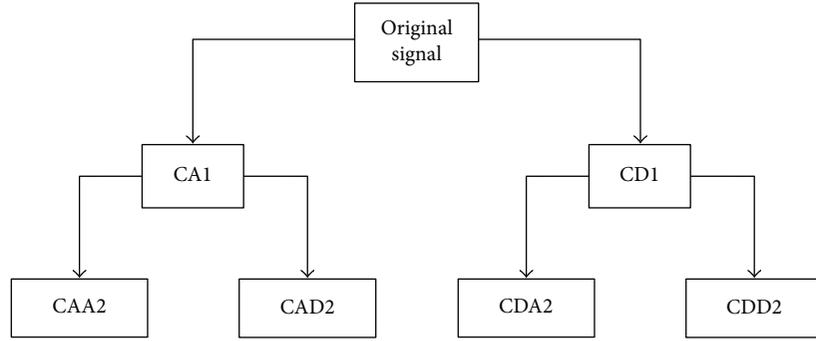


FIGURE 2: Structure and relationship of DWT coefficients.

where the smaller dispersion ratio  $S_{ij}$  means it is easier to separate the sample space with the DWT coefficient  $C_{ij}$  and vice versa.

**3.5.2. Phase Two: Optimal Feature Searching.** In this stage, we aim to find the optimal DWT coefficients for classification with an automatic search method summarized in Algorithm 1 which ensures that the outcomes of the whole searching process are the most useful features for the following classification.

Through the above phases, the original DWT features are refined and selected according to their RPR values. Concise features are finally achieved in an active manner with no human designation.

## 4. Results and Discussion

**4.1. Sample Preparation.** We chose seven kinds of drink such as red wine, white spirit, beer, oolong tea, black tea, maofeng tea, and pu'er tea as test objects. For each kind of tea, we measured 2 g of dried tea leaves with an electronic microbalance, then we soaked the leaves with 200 ml of boiling water for 5 minutes. After that, we filtered out the liquid as the original solution. For liquor objects, such as red wine, white spirit, and beer, we regarded the liquors themselves as the original solutions. In the experiments, we formulated three different concentrations using the original solution and distilled water for each drink. Low, medium, and high concentrations were made up by adjusting the ratio of the original solution at 14%, 25%, and 100%, respectively. The experiments of one concentration were performed three times to increase the reliability of the sampled data. Thus, we collected a total of 63 (7 kinds  $\times$  3 concentrations  $\times$  3 times) samples in this study.

**4.2. Electronic Tongue Settings.** As aforementioned, the electrode sensor array of the designed e-tongue consists of six working electrode sensors, one reference sensor, and one auxiliary electrode sensor. We adopted the multifrequency LAPV (MLAPV) method [19] to generate a multifrequency excitation signal on the reference electrode sensor while the auxiliary and working electrode sensors form six electron loops via test solutions during the working process. The excitation signal of a MLAPV includes several frequency segments in one time cycle to stimulate different transient

pulse-like responses. Consequently, the fingerprints of substances can be achieved in the form of a series of pulses. As shown in Figure 3, we set three segments at frequencies of around 0.2 Hz, 1 Hz, and 2 Hz. To avoid interference between adjacent segments, we added bank areas between the adjacent segments. In each segment, five pulses are arranged at the voltages of 3.3 V, 3.1 V, 2.9 V, 2.7 V, and 2.5 V according to a reference voltage of 2.3 V (the DC voltage on the reference electrode sensor). In other words, the actual pulse amplitudes are 1 V, 0.8 V, 0.6 V, 0.4 V, and 0.2 V. Thus, a total of 15 pulses are generated on the reference electrode in one signal time cycle. It should be noted that the duration time of each pulse remains constant due to keeping the same excitation time for all the working electrode sensors.

Figure 3(b) shows a typical reaction of one working electrode sensor in the experiment. We set the sampling rate at 200 Hz for the sensor array to reduce the distortion according to the highest frequency of the excitation signal. Thus, we sampled a total of 12,300 points for 6 working electrode sensors in one time cycle of 61.5 s and the response size of one sensor was sorted into  $12,300/6 = 2050$  in one test.

**4.3. Feature Visualization and Distribution.** In this part, we intend to demonstrate the extracted features by different methods including no feature extraction (NFE), PIPM, RPR-DWT, and the proposed ASF-DWT. Here we adopted NFE to indicate the data distributions before the feature extraction process. For two DWT-based feature extraction methods, we chose Daubechies 1 (Harr) wavelet as the wavelet basis due to its simplicity and wide application in DWT analysis. Considering high dimensionalities of raw responses and extracted features, we implement PCA transformation for the results of feature extraction methods mentioned above to visualize their distributions [31, 32]. We used red markers to denote the beverages belonging to each liquor type: beer (block), spirit (asterisk), and red wine (pentagram), while the sample distributions of black tea, maofeng tea, pu'er tea, and oolong tea are shown in black color by a triangle, asterisk, block, and circle, respectively. Figures 4–7 show the feature distributions of NFE, PIPM, RPR-DWT, and ASF-DWT, respectively, in PCA spaces.

As shown in Figure 8, we arranged all the responses of the 6 working sensors one by one to form a raw sample for PCA transformation directly. According to 12,300 points, we organized each raw sample as a 12,300-dimensional vector which

Input: all dispersion ratios of DWT coefficients  
Output: selected RPRs of DWT coefficients  
Procedure:  
(1) Initialize maximum decomposition level of DWT and set  $i = 1$   
(2) Initialize optimal RPRs and set  $O = \{r_{11}, r_{12}\}$   
(3) While  $i \leq$  maximum decomposition level  
do  
For each RPR in  $i$ th decomposition level, compare the RPR value with its parent DWT coefficients.  
If current RPR value  $<$  parent's RPR value  
Save current RPR to  $O$ ;  
Delete parent's RPR from  $O$ ;  
If no new RPR added in  $i$ th decomposition level  
Stop and go to Step 4;  
 $i = i + 1$ ;  
(4) Output the elements of  $O$  as the selected features

ALGORITHM 1: ASF for DWT coefficients.

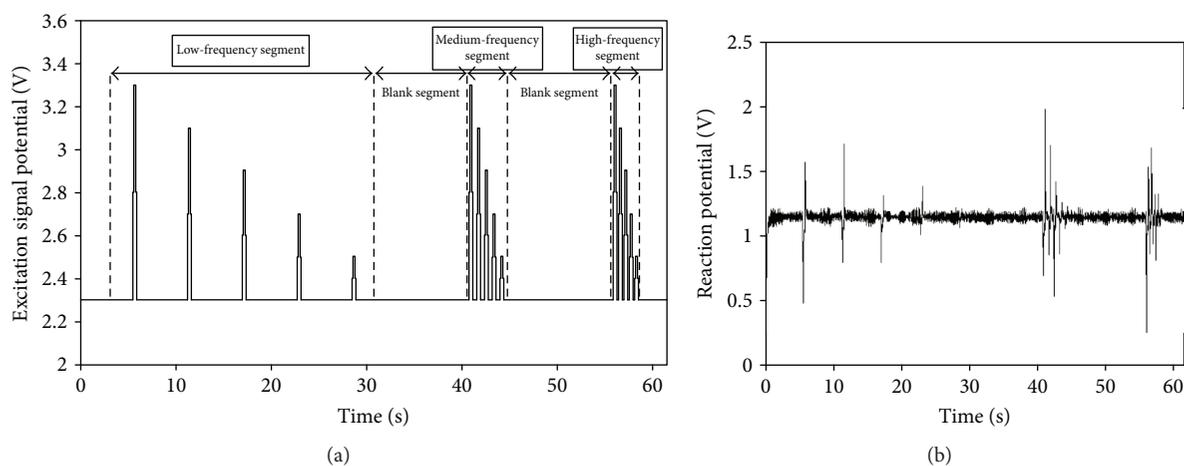


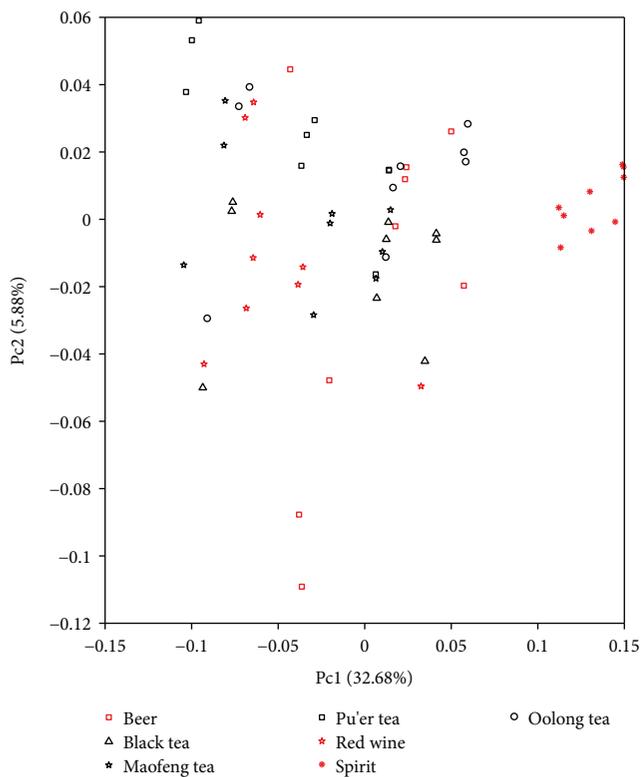
FIGURE 3: (a) Excitation signal of MLAPV. (b) Typical response of one working electrode sensor.

entered PCA in NFE mode. Figures 4(a) and 4(b) show the results of NFE: spirit can be lightly separated from other samples while other beverage distributions have more or less overlaps between each other. It is observed that the contribution rates of PC1–PC3 are 32.68%, 5.88%, and 4.58%, respectively, in Figure 4. The total contribution rates of the first two or three PCs dramatically seem low. We infer that this phenomenon may be caused by a large variance of the same kind of raw sample vectors which have 12,300 dimensions. In other words, the quality of raw samples is unsatisfactory for further recognition and feature extraction is needed to achieve meaningful information hidden in raw sample vectors.

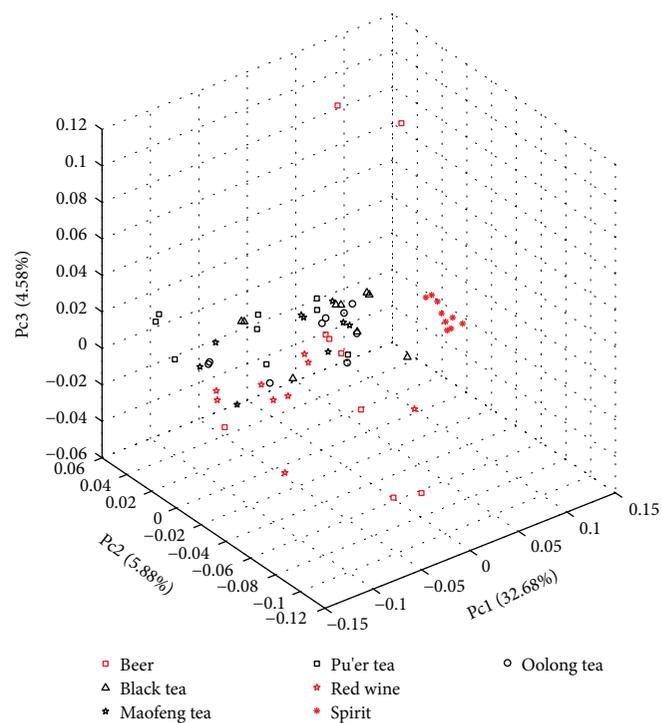
As for PIPM, we took four points from each pulse. Considering 15 pulses in one time cycle,  $15 \times 4 = 60$  points could be extracted from each sensor response, and a total of  $60 \times 6 = 360$  points for the six sensors had been collected; thus, the feature vector of PIPM has 360 dimensions. From Figures 5(a) and 5(b), we found larger contribution rates for the first three PCs (66.79%, 7.69%, and 4.78%) of PIPM compared with the ones of NFE. It means PIPM features have

improved the data quality by reducing the deviation of homogeneous samples. On the other hand, the distributions of different classes are still overlapped; even the red wine samples are no longer separated due to a distribution similar to oolong tea samples.

In terms of RPR-DWT, considering that the 1st coefficient of each decomposition level empirically has the greatest power among the coefficients belonging to the same decomposition level, we specified the selection scope in the 1st coefficients from the 1st to the 6th decomposition levels. For each working electrode sensor, we extracted a RPR value from its response. Thus, the feature size of RPR-DWT is 6 according to six working electrode sensors. From Figures 6(a)–6(f), we could find that the contribution rates (more than 85%) of the first two PCs with RPR-DWT are higher than the ones with the former feature extraction methods. We believe the power calculation process in RPR has decreased the interference caused by tiny disturbances in signals and condensed the useful messages from raw sample vectors. As Figure 6(a) demonstrated, both spirit and red wine samples are clearly located in separate areas with RPR features of the 1st DWT coefficients

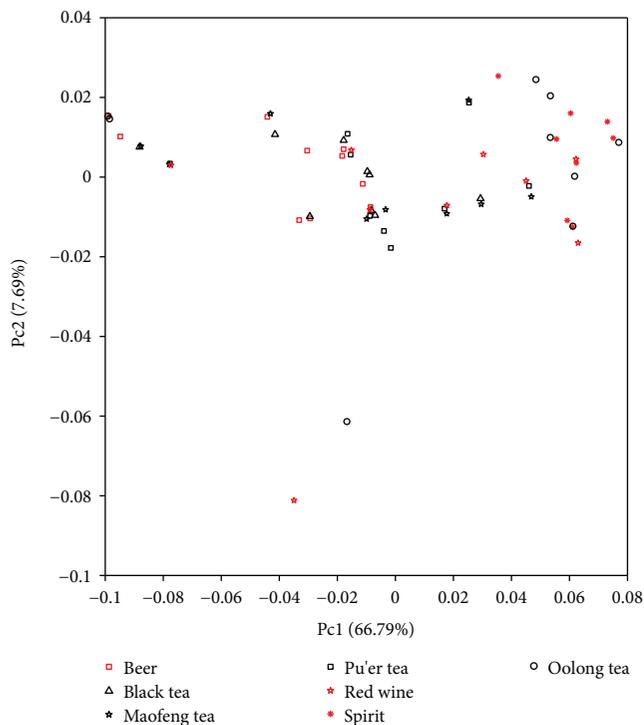


(a)

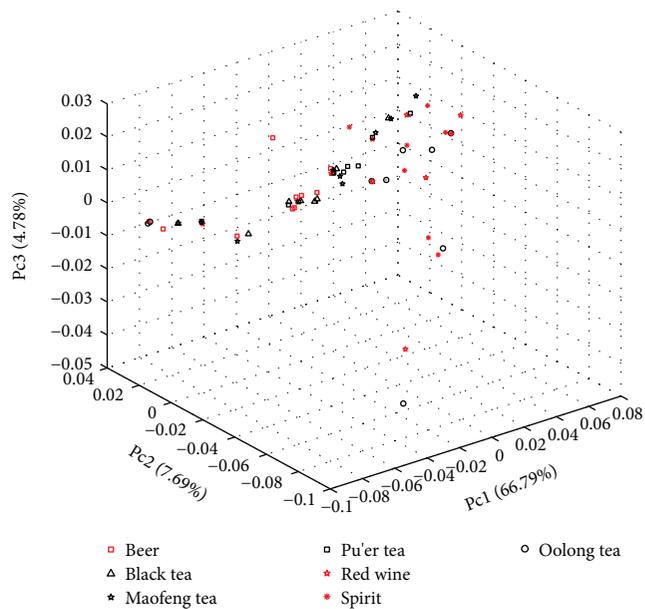


(b)

FIGURE 4: (a) 2D PCA plot of samples with NFE. (b) 3D PCA plot of samples with NFE.



(a)



(b)

FIGURE 5: (a) 2D PCA plot of samples with PIPM. (b) 3D PCA plot of samples with PIPM.

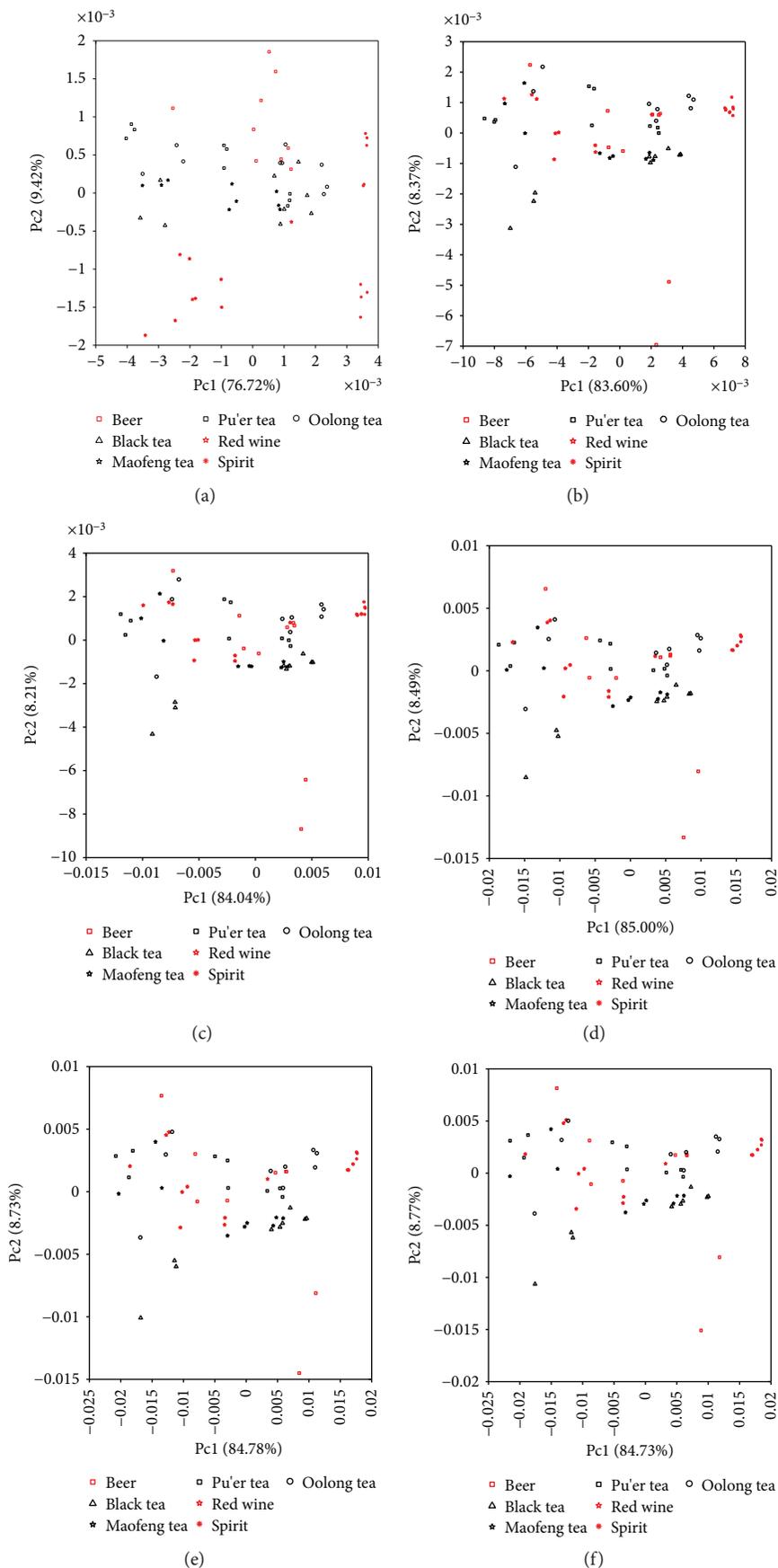


FIGURE 6: (a)–(f) 2D PCA plots of samples with 1st RPR-DWT coefficients in the 1st–6th decomposition layers.

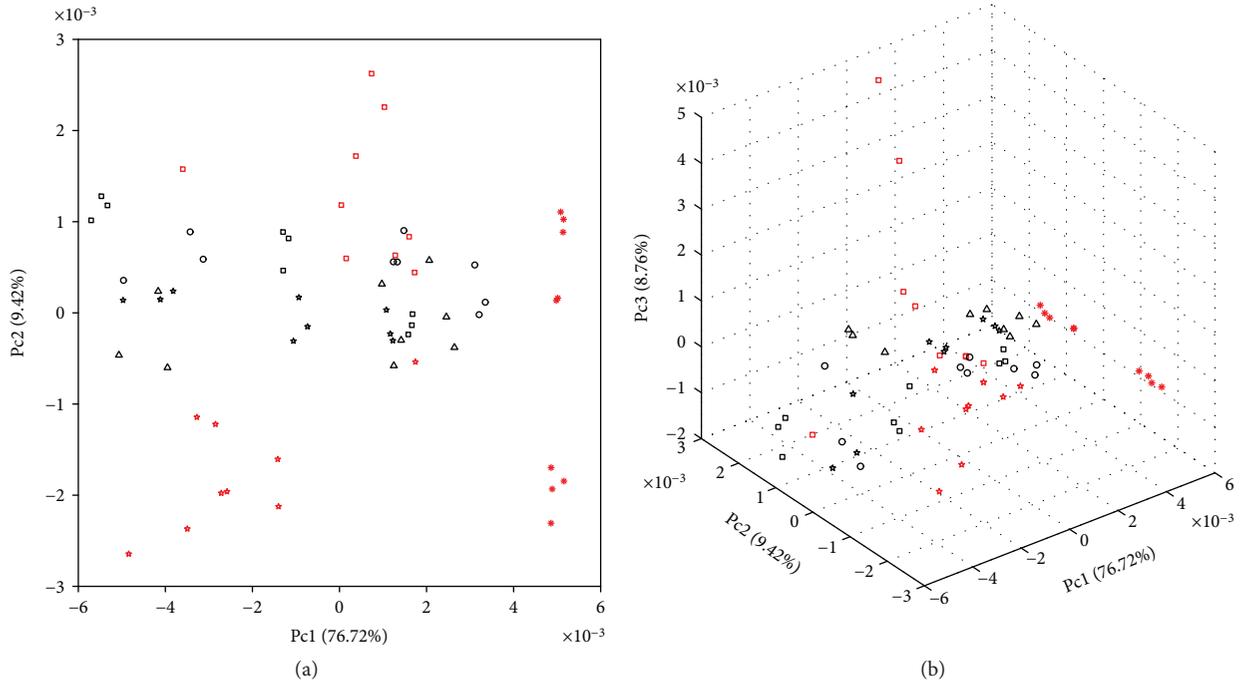


FIGURE 7: (a) 2D PCA plot of samples with ASF-DWT. (b) 3D PCA plot of samples with ASF-DWT.

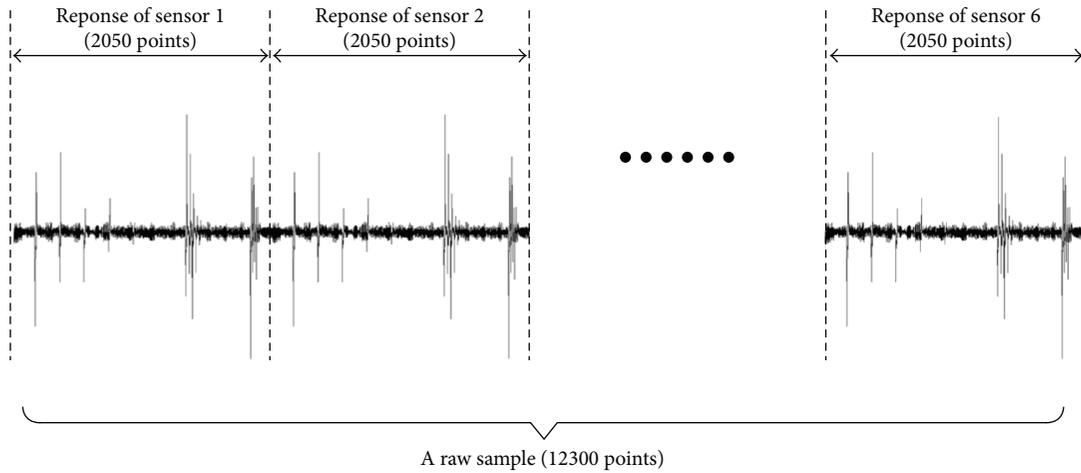


FIGURE 8: Structure of one raw sample.

in the 1st decomposition layer. Meanwhile, only spirit samples can be divided from any other sample distributions in Figures 6(b)–6(f). It seems that the 1st RPR value in decomposition layer 1 is more effective for recognition. In a word, RPR-DWT is a capable feature extraction method for LAPV e-tongues to enhance the quality of extracted features with a little feature size of 6. However, suitable DWT coefficients for RPR features are mainly selected by human experience or a lot of trials. It lacks objectivity and has high costs.

Figures 7(a) and 7(b) demonstrate 2D and 3D PCA plots of ASF-DWT features, respectively. According to Algorithm 1 in Section 3.5, feature exploration starts from the 1st decomposition layer. For the beverage data set, we found that the dispersion ratios of RPRs in the 2nd decomposition layer are not smaller than the ones of their parents in the 1st

decomposition layer. Thus, the RPR and dispersion ratio computation can be stopped in the 2nd decomposition layer. At the same time, the 1st and 2nd RPRs in the 1st decomposition layer were chosen as ASF-DWT features for one sensor. A total of 12 feature values were extracted for one test. Considering that the selected features are based on the coefficients in the 1st decomposition layer, it is reasonable that the sample distributions on ASF-DWT features (shown in Figure 7(a)) are similar to the ones of RPR-DWT with the 1st RPR in the 1st decomposition level (shown in Figure 6(a)). Although the feature size of ASF-DWT is a little larger than the one of RPR-DWT, it can work automatically in unmanned situations with limited calculation.

Generally speaking, PIPM, RPR-DWT, and ASF-DWT are effective at improving the poor quality of raw sensor

TABLE 2: Average recognition rates with leave-one-out validation.

Classifier	Feature extraction methods			
	NFE	PIPM	RPR-DWT	ASF-DWT
RF	71.43% $\pm$ 0.0612	17.46% $\pm$ 0.0397	74.60% $\pm$ 0.0142	<b>76.19%</b> $\pm$ 0.0306
SVM1	46.03% $\pm$ 0.0193	30.16% $\pm$ 0.0329	<b>58.73%</b> $\pm$ 0.0125	<b>58.73%</b> $\pm$ 0.0329
SVM2	14.28% $\pm$ 0.0000	14.28% $\pm$ 0.0000	<b>77.18%</b> $\pm$ 0.0108	76.19% $\pm$ 0.0102
KNN	66.67% $\pm$ 0.0153	22.22% $\pm$ 0.0516	82.54% $\pm$ 0.0091	<b>84.13%</b> $\pm$ 0.0125

TABLE 3: *t*-Test results for the average recognition rates of RPR-DWT and ASF-DWT.

<i>t</i> -Test results	ASF-DWT			
	RF	SVM1	SVM2	KNN
RF	0			
RPR-DWT		0		
			0	
				0

response. In terms of deviation, the effects of feature refinement is gradually raised in the order of NFE, PIPM, and DWT-based methods (RPR-DWT and ASF-DWT). Furthermore, regarding feature size, DWT-based methods provide more compact expressions than NFE and PIPM. It has lower computational complexity for following recognition. Especially, ASF-DWT realizes an active and automatic feature selection mechanism in spite of a little bit of dimensionality increase compared with RPR-DWT. With respect to sample distribution, we cannot pick one where all the classes are separated distinctly. Therefore, we need to utilize proper classifiers to numerically evaluate the effects of the feature extraction methods mentioned above on classification.

**4.4. Evaluation Setups.** In the evaluation part, we adopted the random forest (RF), support vector machine (SVM), and *k*-nearest neighbor (*k*-NN) as classifiers to recognize the features from AFS-DWT. Besides, other feature extraction methods including NFE, PIPM, and RPR-DWT are selected as references.

RF is a powerful ensemble classification method proposed by Breiman in 2001 based on decision trees, and the bagging strategy [28], excellent robustness, and recognition ability of RF [13] make scholars tend to improve and apply this classifier. In the following evaluation, the number of decision trees in a random forest was set to 200, and the final category of the sample is determined by the voting results of the 200 decision trees.

The basic idea of SVM is to divide samples in accordance with the structure and experiential risk minimization. We chose two kinds of kernel functions for the SVM learner: linear kernel and radial basis function (RBF) kernel. The linear kernel divides samples in the original space, while the RBF kernel maps samples into a nonlinear high-dimensional space. We denote SVM with the linear and RBF kernel as SVM1 and SVM2, respectively, in the following sections. In the SVM model, a

penalty coefficient *C* is introduced to adjust the tolerance for error classification. A larger *C* value represents a smaller tolerance for classification errors. We scanned the *C* value from 0.1 to 1 with steps 0.1 and found that *C* = 0.5 is a suitable choice comprehensively. Additionally, we set  $\sigma = 1$  for the RBF kernel and use a one-versus-one strategy to execute multiclass identification.

*k*-NN is a typical classifier based on sample density in a local area. It determines the class of an unknown sample according to *k* labeled samples with the nearest *k* distances. In the following discussion, we set *k* = 1 and took Euclidean distance as the measure metric of *k*-NN.

Considering a total of 63 selected samples, we use a leave-one-out strategy [33] to perform validation. We adopted this strategy to index the effects of feature extraction methods theoretically. With regard to the label balance of training samples, equal sizes of samples were followed for each class in training and validation.

**4.5. Evaluation Results.** Each feature extraction method should be combined with one classifier mentioned in the last part to compute recognition rates accordingly. Considering that the size of a validation set of a certain beverage is 9, we obtain 9 percentage values for each combination of feature extraction and classifier consequently.

According to the validation results shown in Table 2, the highest average recognition rate can be achieved with ASF-DWT in most cases compared with other feature extraction methods. Regarding SVM1, two DWT-based methods achieved the same average recognition rates and clearly exceeded NFE and PIPM while RPR-DWT obtained the best rate with SVM2. As to the rates with RF and *k*-NN, the DWT-based methods have performed much better than others. Somewhat differently, the rate of ASF-DWT is a little bit higher than the one of RPR-DWT. It may have actively benefited from the searching features of the ASF strategy with favorable dispersion ratios. Considering that the effects of the two DWT-based approaches are very close, we performed a *t*-test for the recognition rates of the same classifier and the results were collected in Table 3. We set the significance level of the *t*-test to 0.05. Result “1” denotes that a significant difference existed between two cases in distribution while “0” means the opposite. In other words, “1” denotes that the comparison result is not equal and depends on the average recognition rates in Table 2 while “0” shows that the recognition performance of the two methodologies is equal. The results of the *t*-test are all 0 which implies the recognition performance of both RPR-DWT and ASF-DWT are statistically equal.

Concerning the details of DWT coefficient selection, we use  $x:y$  to denote the  $y$ th DWT coefficient in the  $x$ th decomposition level. We selected 3:1 RPR values by traversing the first RPR of the 1st–10th decomposition layers for RPR-DWT while the 2:1, 6:1, and 8:1 RPR values for ASF-DWT were automatically explored. Thus, the feature sizes are 6 and 18 for RPR-DWT and ASF-DWT, respectively. It is worthwhile, we assume, to exchange automatic operation with a small dimensional addition.

In view of the recognition learner, all the classifiers performed ineffectively on either raw sample vectors or PIPM features except RF on NFE. It is apparent that excellent results can hardly be achieved on the data with a large dispersion of the same class. For two DWT-based feature extraction methods, we discovered that the recognition rates become gradually higher in the order of SVM1, RF, SVM2, and  $k$ -NN. It is reasonable that SVM1 provides poor results due to its linear kernel which divides samples linearly while RF and SVM2 obtain much higher recognition rates by implementing nonlinear classification. We notice that  $k$ -NN reaches the highest rates of 82.54% and 84.13% with RPR-DWT and ASF-DWT features, respectively, among four feature extraction modes. However, in traditional opinions, the  $k$ -NN classifier suffers interference from the local sample distribution while SVM can achieve optimal results globally. We believe this contradiction can be explained by the sample distribution shown in Figures 6 and 7. Take Figure 7 as an example, black tea (black triangle), maofeng tea (black pentagram), and pu'er tea (black square) samples are overlapping in the plot. It is difficult to distinguish them entirely according to certain separating hyperplanes, even in nonlinear spaces. Under this restriction, recognition that depends on its neighbors ( $k$ -NN) seems more effective and feasible.

## 5. Conclusions

In this study, we propose ASF-DWT to automatically extract features from raw responses of a LAPV e-tongue to deal with the inconvenience and huge computation required by manual judgments. Dispersion ratio calculation and optimal feature search are combined together to obtain favorable features. Furthermore, we used a LAPV e-tongue to collect beverage samples and identify categories. The experimental results show that ASF-DWT is a very helpful feature extraction tool for LAPV e-tongue responses and it outperforms conventional feature extraction methods with acceptable feature size.

Future works should concentrate on the optimal rules of ASF-DWT for various applications and its compatibility with various classification methods. At the same time, the bandwidth of an excitation signal can be further increased to achieve more useful information from transient responses.

## Disclosure

The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in writing the manuscript; and in the decision to publish the results.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC, 61403049) in 2015 at P. R. China and by the National International Science and Technology Cooperation Project (2014DFA31560) in 2014 at P. R. China.

## Supplementary Materials

The dataset used in this paper is included in the supplementary materials. We saved the original data (hexadecimal numbers) of 63 tests separately in txt files. Thus, each file corresponds one test and information about the subject, concentration, and ordinal are provided by file name “A-B-C” where “A” shows the beverage name, “B” denotes the concentration level, and “C” refers to the order of experiments. (*Supplementary Materials*)

## References

- [1] Y. Tahara and K. Toko, “Electronic tongues—a review,” *IEEE Sensors Journal*, vol. 13, no. 8, pp. 3001–3011, 2013.
- [2] D. Ha, Q. Sun, K. Su et al., “Recent achievements in electronic tongue and bioelectronic tongue as taste sensors,” *Sensors and Actuators B Chemical*, vol. 207, no. 2015, pp. 1136–1146, 2015.
- [3] Z. Wei and J. Wang, “Tracing floral and geographical origins of honeys by potentiometric and voltammetric electronic tongue,” *Computers and Electronics in Agriculture*, vol. 108, pp. 112–122, 2014.
- [4] L. Wang, Q. Niu, Y. Hui, and H. Jin, “Discrimination of rice with different pretreatment methods by using a voltammetric electronic tongue,” *Sensors*, vol. 15, no. 7, pp. 17767–17785, 2015.
- [5] P. Ciosek, Z. Brzózka, and W. Wróblewski, “Classification of beverages using a reduced sensor array,” *Sensors and Actuators B: Chemical*, vol. 103, no. 1–2, pp. 76–83, 2004.
- [6] R. B. Domínguez, L. Morenobarón, R. Muñoz, and J. Gutiérrez, “Voltammetric electronic tongue and support vector machines for identification of selected features in Mexican coffee,” *Sensors*, vol. 14, no. 9, pp. 17770–17785, 2014.
- [7] M. Gutiérrez, A. Llobera, A. Ipatov et al., “Application of an E-tongue to the analysis of monovarietal and blends of white wines,” *Sensors*, vol. 11, no. 5, pp. 4840–4857, 2011.
- [8] M. Palit, B. Tudu, N. Bhattacharyya et al., “Comparison of multivariate preprocessing techniques as applied to electronic tongue based pattern classification for black tea,” *Analytica Chimica Acta*, vol. 675, no. 1, pp. 8–15, 2010.
- [9] M. Palit, B. Tudu, P. K. Dutta et al., “Classification of black tea taste and correlation with tea taster’s mark using voltammetric electronic tongue,” *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 8, pp. 2230–2239, 2010.
- [10] A. Ghosh, A. K. Bag, P. Sharma et al., “Monitoring the fermentation process and detection of optimum fermentation time of black tea using an electronic tongue,” *IEEE Sensors Journal*, vol. 15, no. 11, pp. 6255–6262, 2015.

- [11] Z. Wei, W. Zhang, Y. Wang, and J. Wang, "Monitoring the fermentation, post-ripeness and storage processes of set yogurt using voltammetric electronic tongue," *Journal of Food Engineering*, vol. 203, pp. 41–52, 2017.
- [12] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "A novel technique of black tea quality prediction using electronic tongue signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 10, pp. 2472–2479, 2014.
- [13] M. Liu, M. Wang, J. Wang, and D. Li, "Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar," *Sensors and Actuators B: Chemical*, vol. 177, no. 2, pp. 970–980, 2013.
- [14] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "Tea quality prediction by autoregressive modeling of electronic tongue signals," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4470–4477, 2016.
- [15] X. Cetó, A. González-Calabuig, J. Capdevila, A. Puig-Pujol, and M. del Valle, "Instrumental measurement of wine sensory descriptors using a voltammetric electronic tongue," *Sensors and Actuators B: Chemical*, vol. 207, pp. 1053–1059, 2015.
- [16] K. Tiwari, B. Tudu, R. Bandyopadhyay, and A. Chatterjee, "Identification of monofloral honey using voltammetric electronic tongue," *Journal of Food Engineering*, vol. 117, no. 2, pp. 205–210, 2013.
- [17] F. Winquist, P. Wide, and I. Lundström, "An electronic tongue based on voltammetry," *Analytica Chimica Acta*, vol. 357, no. 1-2, pp. 21–31, 1997.
- [18] P. Ivarsson, S. Holmin, N. Höjer, C. Krantz-Rülcker, and F. Winquist, "Discrimination of tea by means of a voltammetric electronic tongue and different applied waveforms," *Sensors and Actuators B: Chemical*, vol. 76, no. 1–3, pp. 449–454, 2001.
- [19] S. Y. Tian, S. P. Deng, and Z. X. Chen, "Multifrequency large amplitude pulse voltammetry: a novel electrochemical method for electronic tongue," *Sensors and Actuators B: Chemical*, vol. 123, no. 2, pp. 1049–1056, 2007.
- [20] L. Lu, X. Hu, S. Tian, S. Deng, and Z. Zhu, "Visualized attribute analysis approach for characterization and quantification of rice taste flavor using electronic tongue," *Analytica Chimica Acta*, vol. 919, pp. 11–19, 2016, 2016..
- [21] Z. Wei and J. Wang, "The evaluation of sugar content and firmness of non-climacteric pears based on voltammetric electronic tongue," *Journal of Food Engineering*, vol. 117, no. 1, pp. 158–164, 2013.
- [22] B. Adhikari, M. Mahato, T. Sinha, A. Halder, and N. Bhattacharya, "Development of novel polymeric sensors for taste sensing: electronic tongue," in *SENSORS2013 IEEE*, vol. 1-4, Baltimore, MD, USA, 2013, IEEE.
- [23] X. Cetó, J. M. Gutiérrez, M. Gutiérrez et al., "Determination of total polyphenol index in wines employing a voltammetric electronic tongue," *Analytica Chimica Acta*, vol. 732, no. 1, pp. 172–179, 2012.
- [24] L. Moreno-Barón, R. Cartas, A. Merkoçi et al., "Application of the wavelet transform coupled with artificial neural networks for quantification purposes in a voltammetric electronic tongue," *Sensors and Actuators B: Chemical*, vol. 113, no. 1, pp. 487–499, 2006.
- [25] Y. Yin, H. Yu, and H. Zhang, "A feature extraction method based on wavelet packet analysis for discrimination of Chinese vinegars using a gas sensors array," *Sensors and Actuators B: Chemical*, vol. 134, no. 2, pp. 1005–1009, 2008.
- [26] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [27] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [28] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [29] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1–3, pp. 37–52, 1987.
- [30] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [31] X. Cetó, J. M. Gutiérrez, L. Moreno-Barón, S. Alegret, and M. del Valle, "Voltammetric electronic tongue in the analysis of Cava wines," *Electroanalysis*, vol. 23, no. 1, pp. 72–78, 2011.
- [32] J. M. Gutiérrez, L. Moreno-Baró, M. I. Pividori, S. Alegret, and M. del Valle, "A voltammetric electronic tongue made of modified epoxy-graphite electrodes for the qualitative analysis of wine," *Microchimica Acta*, vol. 169, no. 3–4, pp. 261–268, 2010.
- [33] M. Kearns and D. Ron, "Algorithmic stability and sanity-check bounds for leave-one-out cross-validation," *Neural Computation*, vol. 11, no. 6, pp. 1427–1453, 1999.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

