China Instrument
and Control Society
Instrumentation

OPEN ACCESS

INSTRUMENTATION, Vol. 11, No. 1, March 2024 https://doi.org/10.15878/j.instr.202300148

Article

Research on Detection of Food additives Based on Terahertz Spectroscopy and Analytic Hierarchy Process

Miaoyu Zhao, FangYan*, Wenwen Li and Yangshuo Liu

Inner Mongolia University of Science and Technology School of Automation and Electrical Engineering, Baotou, Inner Mongolia 014010, China

* Corresponding author email: 0472yanfang@163.com

Abstract: Terahertz time-domain spectroscopy is a kind of far-infrared spectroscopy technology, and its spectrum reflects the internal properties of substances with rich physical and chemical information, so the use of terahertz waves can be used to qualitatively identify food additives containing nitrogen elements. Analytic hierarchy process (AHP) was originally used to solve evaluation-type problems, and this paper introduces it into the field of terahertz spectral qualitative analysis, proposes a terahertz time-domain spectral qualitative identification method combined with analytic hierarchy process, and verifies the feasibility of the method by taking four common food additives (xylitol, L-alanine, sorbic acid, and benzoic acid) and two illegal additives (melamine, and Sudan Red No. I) as the objects of study. Firstly, the collected terahertz time-domain spectral data were pre-processed and transformed into a data set consisting of peaks, peak positions, peak numbers and overall trends; then, the data were divided into comparison and test sets, and a qualitative additive identification model incorporating analytic hierarchy process was constructed and parameter optimisation was performed. The results showed that the qualitative identification accuracies of additives based on single factors, i.e., overall trend, peak value, peak position, and peak number, were 80.23%, 70.93%, 67.44%, and 40.70%, respectively, whereas the identification accuracy of the analytic hierarchy process qualitative identification method based on multi-factors could be improved to 92.44%. In addition, the fuzzy characterisation of the absorption spectrum data was binarised in the data pre-processing stage and used as the base data for the overall trend, and the recognition accuracy was improved to 94.19% by combining the fuzzy characterisation method of such data with the hierarchical analysis qualitative recognition model. The results show that it is feasible to use terahertz technology to identify different varieties of additives, and this paper constructs a hierarchical analytical qualitative model with better effect, which provides a new means for food additives detection, and the method is simple in steps, with a small demand for samples, which is suitable for the rapid detection of small samples.

Keywords: terahertz time-domain spectroscopy; analytic hierarchy process; qualitative identification; food additives

Citation: Miaoyu Zhao, FangYan, Wenwen Li and Yangshuo Liu. "Research on Detection of Food Additives Based on Terahertz Spectroscopy and Analytic Hierarchy Process." *Instrumentation* 11, no. 1 (2023). https://doi.org/10.15878/j.instr.202300148

0 Introduction

Food additives are the "secret weapon" of today's food industry, used to improve the flavour of food, enhance nutrition or preserve freshness[1] . In recent years, problems related to the illegal use of food additives have been repeatedly reported in the press, and the issue of food quality and safety has become a hot topic in the society[2] . Terahertz time-domain spectroscopy, as a new non-destructive testing technology, has unique advantages in food testing[3] . Studies have shown that the low-frequency vibration frequencies of many biomolecules in nature happen to be located in the terahertz frequency band, which makes terahertz waves have unique spectral properties for the collective vibration of biomolecules and intermolecular interactions such as hydrogen-bonding oscillations[4-6]. Food additives are complex in composition, and most of the interactions between organic molecules are in the terahertz band, so their fingerprint spectroscopy can be utilised for the detection of additives.

In the early stage of applying terahertz technology to realise substance detection, it mostly relies on the identification of special absorption peaks in the spectrum, however, when faced with a wide variety of substances to be measured with similar structures and force majeure factors in the environment during the measurement process, it is no longer applicable to identify the substance species only by relying on the identification of absorption peaks. Later, with the continuous development and improvement of machine learning, the identification of substances based on terahertz spectral data also mostly adopts machine learning methods, for example, Hu et al. used particle swarm (PSO) algorithm to optimize the SVM model to achieve the rapid identification of coffee bean origin, and the recognition rate reaches 95%[7] ; Zhou Bo et al. used their terahertz time-domain spectra to establish the SVM for different tobacco recipe components, respectively. KNN and bagging tree classification models, the classification accuracy is more than 90%, for the application of terahertz technology in the non-destructive inspection of tobacco materials to provide a reference[8] ; but machine learning as a method of strong dependence on the data, usually need a large number of training sets and a longer period of time to train the model in order to get a better result, and for the different training data are required to find a specific optimal parameter[9] . In the actual detection, most of the data obtained are small samples, and there may be overfitting or underfitting problems in using them to train the model[10], which limits the application of terahertz time-domain spectroscopy in the field of material qualitative identification to a certain extent.

Finding suitable, accurate and rapid detection methods for different food ingredients is an important means to ensure the safety of food consumption[11] . The analytic hierarchy process introduced in this paper is a subjective evaluation method, which can be used for the characterisation based on spectral features by integrating the overall trend of the spectral data, peak position, peak value, number of peaks and other absorption spectral features, so as to convert the subjective evaluation into an objective identification based on the data, which not only reduces the reliance on the data pre-processing, but also effectively identifies the different substances with similar spectra.

1 Experimental Part

1.1 Materials and Laboratory Instruments

The experimental data collected in this paper were obtained from the Transmission THz-TDS Experimental System[12] of the Beijing Industrial Spectral Imaging Engineering and Technology Research Centre (BISIRC). The system uses a Mai Tai HP titanium-sapphire femtosecond laser from Spectra-Physics to output an ultrafast infrared laser[13] for generating and detecting terahertz waves with a central wavelength of 800 nm, an average infrared power of 2.95 W, a pulse width of less than 100 fs, and a repetition frequency of 79.3 MHz, with the output power varying within ±1%,the system principle and system objects are shown in Fig.1 and Fig.2. All the experimental samples were purchased from Aladdin Reagent Network, and their purity was ⩾99%, and the experiments were conducted to collect

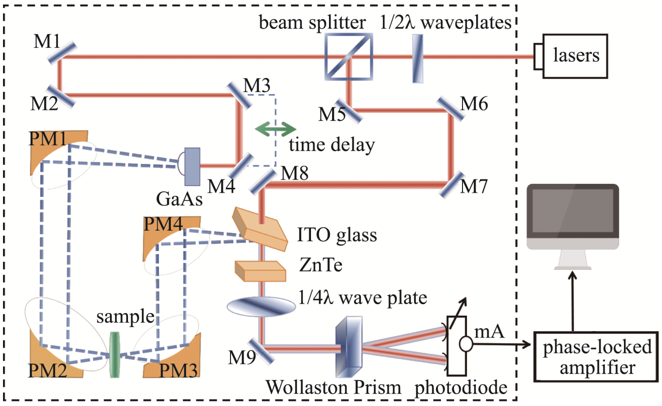


Fig.1 Schematic diagram of transmissive terahertz time-domain spectroscopy system

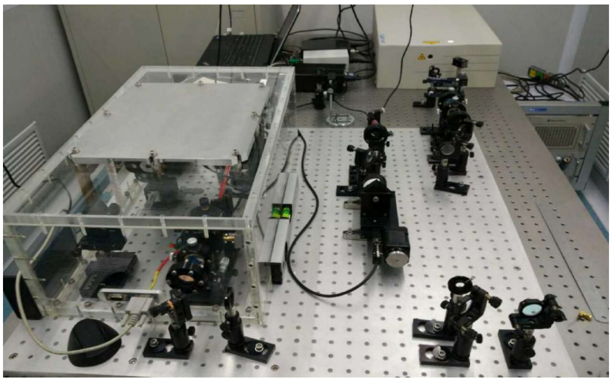
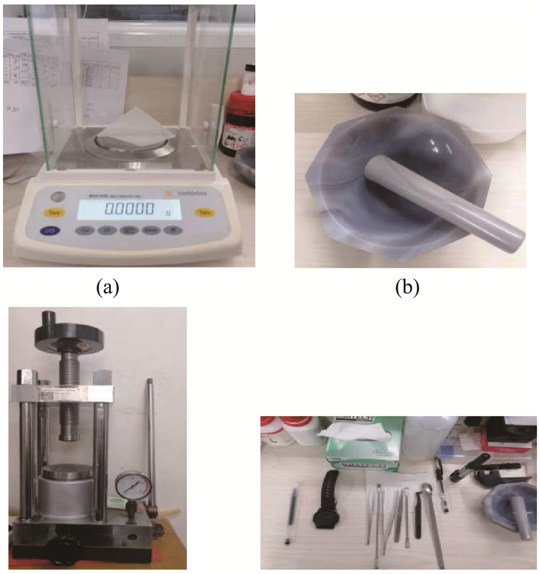


Fig.2 Transmission terahertz time-domain spectroscopy system diagram

the terahertz time-domain spectra of the monomers of six food additives, namely, L-alanine, benzoic acid, melamine, sorbic acid, Sudan 1, and xylitol, where the compensating sample was polyethylene (PE).

1.2 Sample Preparation and Data Acquisition

In this paper, sample preparation was carried out using the tablet pressing method and Fig.3 shows the apparatus required for the preparation of the tablets. Firstly, the sample and polyethylene were weighed and poured into an agate mortar according to the mass ratio set in the experiment, and then the well-mixed sample was poured into the tabletting mould, and finally pressed by a tabletting machine at a pressure of 5 MPa for 5 min, so that it was made into a round tablet with a thickness of about 1.5 mm and a diameter of 1.3cm. The samples were then checked for uniformity with light and bad slices were rejected, measured with vernier calipers, recorded with their thickness information and then put into a sample bag for storage. For the convenience of presentation, S1~S6 will be used to refer to these six additives, in which there are three concentrations of each additive sample, each concentration of three samples, a total of nine samples were prepared, a total of 54 samples were prepared, and the details of the sample information are shown in Table 1. During the data acquisition process, each sample was measured four times, and a total of 216 terahertz time-domain spectroscopic data were obtained.



(a) (b)

(c) (d)

Fig.3 Instruments required for sample preparation: (a)Electronic scales;(b) Agate mortar; (c) Tablet press ; (d) Gauges.

1.3 Optical Parameter Extraction

The absorption coefficients of each sample in the terahertz band were calculated according to the basic model of optical parameters proposed by T.D. Dorney and L. Duvillaret[14,15] $\alpha(\omega)$, the equations are shown in Eq. (1) Eq. (2).

$$n(\omega)=\frac{\varphi(\omega)c}{\omega d}+1 \tag{1}$$

$$\alpha(\omega)=\frac{2}{d}\ln\left\{\frac{4n(\omega)}{\rho(\omega)\left[n(\omega)+1\right]^{2}}\right\} \tag{2}$$

In Eq.(1), c is the vacuum speed of light; ω is the vibration frequency; d is the sample thickness; $\varphi(\omega)$ is the phase difference between the sample signal and the reference signal;

Table 1 Statistical table of additives samples

Category number	Name	Sample concentration/%	Sample quantity
S1	L-Alanine	10%	3
		25%	3
		35%	3
S2	Benzoic Acid	10%	3
		25%	3
		35%	3
S3	Melamine	10%	3
		25%	3
		35%	3
S4	2,4-Hexadienoic acid	10%	3
		25%	3
		35%	3
S5	Sudan-I	10%	3
		25%	3
		35%	3
S6	Xylitol	10%	3
		25%	3
		35%	3

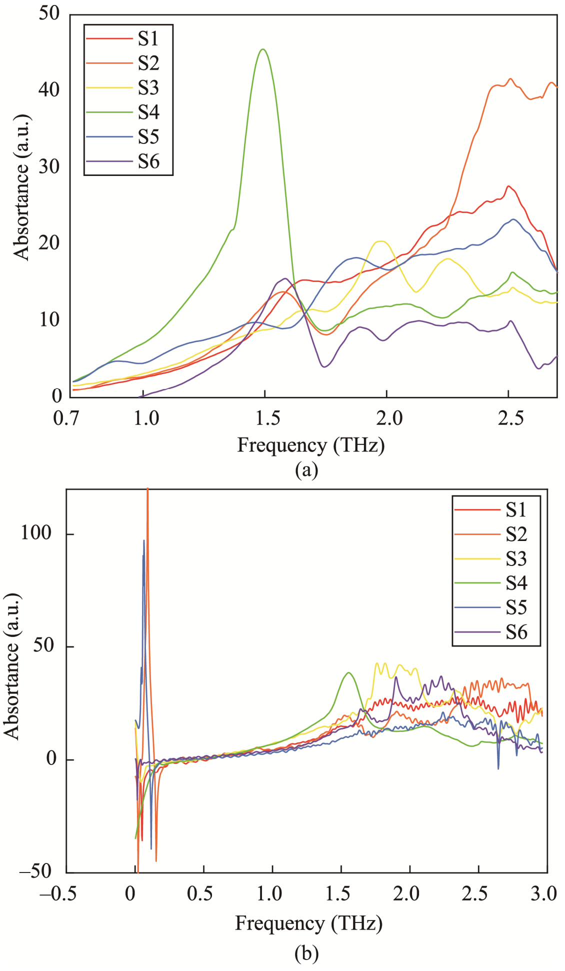
In Eq.(2), $\rho(\omega)$ is the amplitude ratio between the sample signal and the reference signal.

As an example, the frequency domain spectrum of a 35% sample of the six additive monomers after data preprocessing is shown in Fig.4 Among them. Fig.4(a) shows the original absorption coefficient spectrum after the extraction of optical parameters, the absorption peaks are more obvious, but there is more noise; Fig.4(b) shows the terahertz absorption coefficient spectrum after S-G smoothing and characteristic band interception, the interception band is 0.7 THz-2.7 THz, and the width of the moving window is 50, which retains the effective spectral information and filters out the noise.

2 Model Construction

2.1 Introduction to Analysis Hierarchical process

In the analytic hierarchy process (AHP), by decomposing the complex analytical object into the objective layer, criterion layer and solution layer, and judging the importance of the elements of the same layer, establishing a judgement matrix, and calculating the weight of each factor on the importance of the indicators of the higher layer, and obtaining the weight corresponding to the judgement matrix, which provides the basis for the optimal choice[16][17] . According to the analytic hierarchy process structure, the relative importance of each criterion is compared with that of the target level, starting from the criterion level. The relative importance scale is shown in Table 2.



(a)

(b)

Fig.4 Frequency domain spectrum after data preprocessing

Table 2 AHP relative importance scale

scale	define	clarification
1	equally important	Elements a and b are equally important for an attribute
3	slightly important	Element a is slightly more important than element b for an attribute
5	more important	Element a is more important than element b for an attribute
7	clearly important	Element a is significantly more important than element b for an attribute
9	special importance	Element a is particularly more important than element b for an attribute
2, 4, 6, 8.	midpoint	A compromise between the two neighbouring scales above
The reciprocal of the above scales	inverse comparison	For example, a compares to b as 3 and b compares to a as 1/3.

2.2 Modeling

The process of constructing a hierarchical analytical qualitative identification model that integrates multiple factors is shown in Fig. 5.

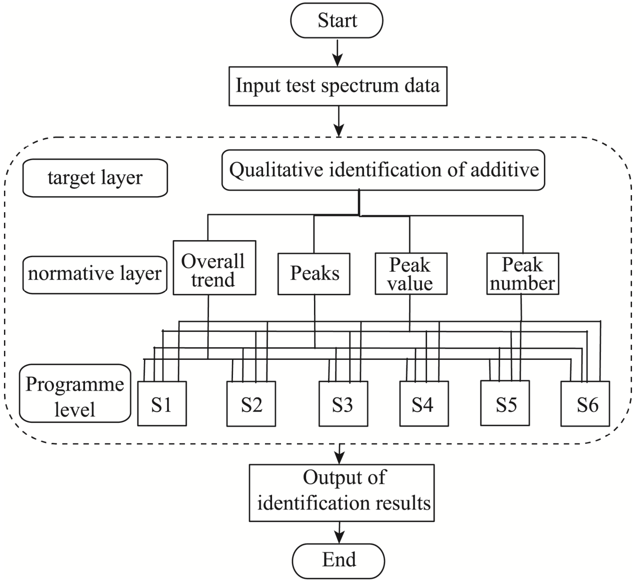


Fig.5 Flow chart of Qualitative Identification model with AHP based on multiple factors

2.3 Evaluation of Indicator Design

Evaluation index design: Definition 1: Let matrix $W=(w_{ij})_{(n+1)\times m}$, n is the number of criterion layers, m is the number of programme layers, if there is $0 \leqslant w_{ij} \leqslant 1$, then matrix W is the final weight matrix.

Definition 2: Set matrix $A = (a_{ij})_{n\times n}$, if there is $0 \leqslant A_{ij} \leqslant 1$, then matrix A for the criterion level to the target level of the judgement matrix, which a_{ij} expressed the significance of the indicator i relative to the target level of the importance of the indicator j relative to the importance of the target level.

Design of judgement matrix A from guideline layer to target layer: based on the

qualitative results of only a single factor to consider the overall trend, the peak, peak position, peak number of the initial importance of the decreasing order, the specific judgement matrix A, as shown in Table 3, the judgement matrix A to the calculation of the weight vector AA, as shown in Eq.(3). From the judgement matrix A, the initial weights of the four criteria in the target layer are vector AA=[0.6545, 0.2045, 0.0955, 0.0456], which is filled into the first column of the weight matrix W.

$$aa_i = \frac{1}{4}\sum_{j=1}^{4}\frac{a_{ij}}{\sum_{k=1}^{4} a_{kj}},\ i=1, 2, 3, 4;\ k=1, 2, 3, 4 \quad (3)$$

Table 3 Judgment matrix from criterion layer to target layer

norm	Overall trend	peak value	peaks	peak number
Overall trend	1	3	5	7
peak value	1/3	1	3	5
peaks	1/5	1/3	1	3
peak number	1/7	1/5	1/3	1

(2) Design of the weights of the programme layer with respect to the criterion layers: notate the matrix $B = (b_{ij})_{4\times6}$, as the matrix of weights of the programme layer with respect to the four criteria, and let R=[0.048, 0.092, 0.14, 0.19, 0.24, 0.29]$_{1\times6}$, with R being the normalised weights on the 2, 4, 6, 8, 10, 12 scales. The weights of the scenario layers in matrix B with respect to each criterion layer are specific orderings of the R vector, based on the similarity of the scenario layer to that criterion layer. For example, if the Euclidean distance of test data X about the overall trend on categories S1, S2, S3, S4, S5, S6 is 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, then it is determined that this data is most similar to S1 and least similar to S6, then the weights corresponding to the overall trend on the R=[0.29, 0.24, 0.19, 0.14, 0.092, 0.048], and so on filling all the elements of matrix B by rows into the final weight matrix W.

(3) Design of the final weight matrix W: define $AA = (aa_i)_{4\times1}$ as the weight vector of the criterion layer about the target layer derived from the judgement matrix A, assume the importance weight ratio of the overall trend of the spectrum, the peak position, the peaks, and the number of peaks in the qualitative identification. w is the combination of the vector AA and the weight matrix B of the scenario layer about the criterion layer, which is shown in Table 4. Finally, according to the weight matrix W, the comprehensive score of a single test data is derived, and the result is $C=AA^T\times B$, assume the first column of W is multiplied by the rest of the columns to sum up. C data is stored in the form of [c1,...c2], and the coordinates of the columns corresponding to the maximum value of the vector C is the final identification result, and so on to derive the identification result of all the test data.

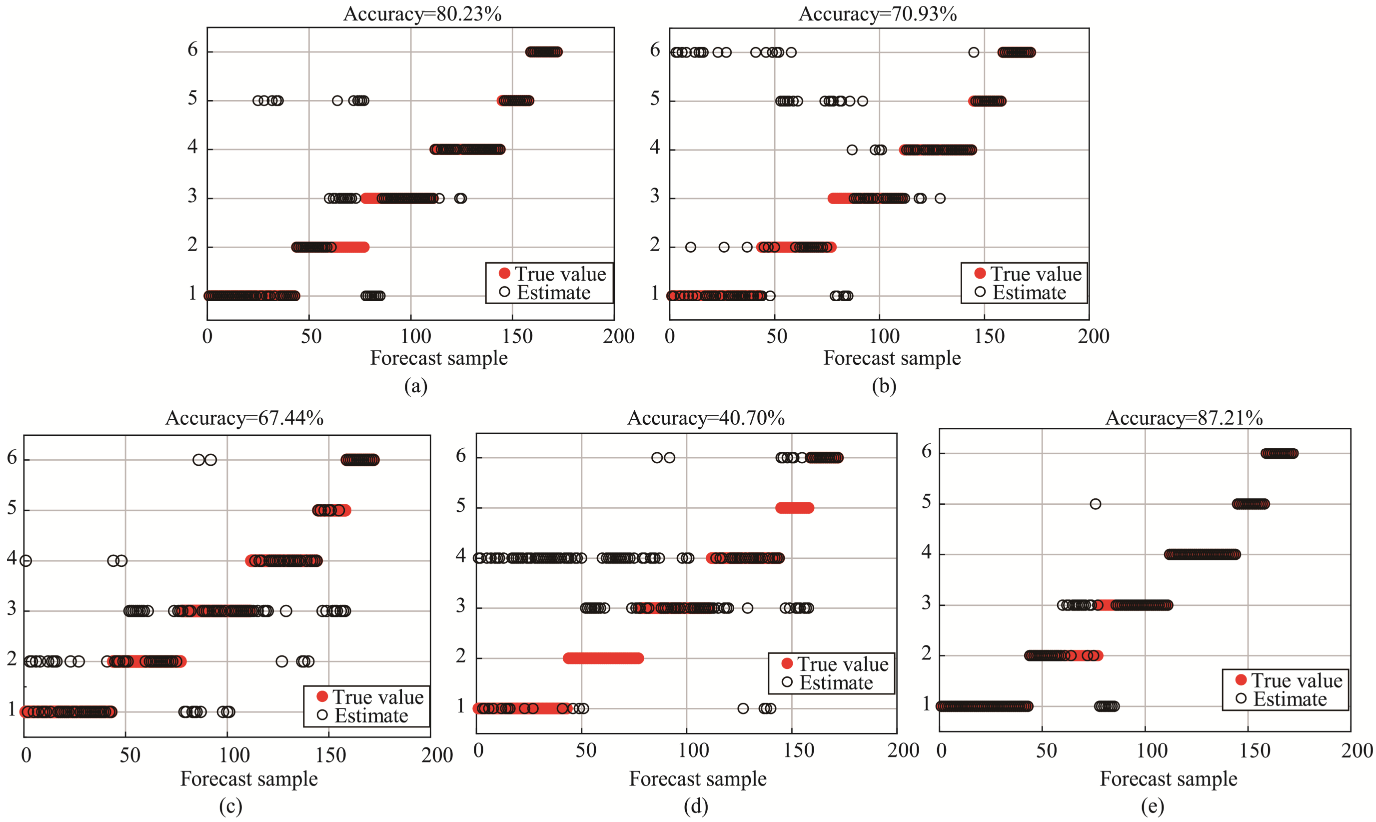
Table 4 Final weight matrix W

target	Indicator weights	S1	S2	S3	S4	S5	S6
Overall trend	0.6545	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}	b_{16}
peaks	0.2045	b_{21}	b_{22}	b_{23}	b_{24}	b_{25}	b_{26}
peak value	0.0955	b_{31}	b_{32}	b_{33}	b_{34}	b_{35}	b_{36}
peak number	0.0456	b_{41}	b_{42}	b_{43}	b_{44}	b_{45}	b_{46}

3 Analysis of Results and Optimisation

3.1 Analysis of Identification Results

The results of qualitative identification of additives based on different factors in terahertz spectroscopy are shown in Fig.6. As can be seen from Fig.6, the qualitative identification results based on a single factor are not satisfactory (Fig.6(a)-(d)), which can be attributed to the fact that the non-absolute consistency of the experimental environment and the thickness of the samples makes the similar data differ in a certain feature interval, and there are similar features between different types of samples; in



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

Fig.6 The recognition results of 6 types of additive samples based on different factors. (a) Overall trend; (b) Peak value; (c) Peak position; (d) Peak number; (e) Based on multiple factors

Fig.6(e), the identification accuracy is improved significantly compared with that of only considering a single factor, because the analytic hierarchy process algorithm based on multiple factors takes into account the influence of each factor on the identification according to the specific weights. is improved compared with that of the recognition result that only considers a single factor.

3.2 Optimisation of Recognition Results

3.2.1 Fuzzy Characterisation of Data

Comparing the identification results in Fig. 6, it can be seen that the overall trend of the spectral lines in the absorption spectrum has a higher degree of importance among the factors, so this paper proposes a method of characterising the overall trend of the spectrum in a fuzzy way, as follows:

Firstly, write the single raw spectrum as $H = (h_1, h_2, h_3, \ldots, h_j)$, j=1, 2, 3...,g, where g is the dimension of the raw spectral data; secondly, the difference of the neighbouring data in H is written as $E = (e)_{j1\times j} = (h_2 - h_1, h_3 - h_2, \ldots, h_j - h_{j-1})$,when $e_j \geqslant \lambda$ or $-\lambda < e_j < 0$, let $e_j = 1$, which represents an ascent, and let $e = -1$, which represents a descent, when $e_j < -\lambda$ or $0 < e_j < \lambda$, where λ represents the mean of the data difference, H is transformed into a dataset with only –1 and 1.

A comparison of some of the raw absorption spectral data of sample S5 with the data after fuzzy characterisation is shown in Fig.7. From Fig.7, it can be seen that S5 is a characteristic point at 1.46 THz, and the data on the right side of S5-3 at 1.46 THz in the original data show an upward trend, which does not match the trend of S5-1 and S5-2 on the right side of 1.46 THz, which may lead to the reduction of the recognition accuracy; while the fuzzy characterisation makes the three data at 1.46 THz reflect their characteristics, and in the area of the characteristics are not obvious There is also a certain feature performance, so the method can highlight the overall trend characteristics of the spectral data, and the integer arithmetic saves the arithmetic time of the procedure compared with the decimal arithmetic. In addition, in the subsequent layer of the programme, the similarity of the overall trend will no longer be based on the Euclidean distance to sort R, but use the repetition rate between the data to do the basis of R sorting.

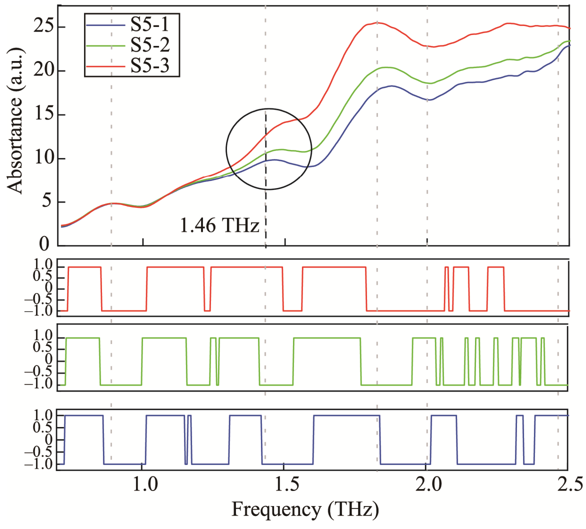


Fig.7 Comparison of original absorption spectra and fuzzy characterisation spectra of sample S5

3.2.2 Parameter Optimisation

The optimisation of the qualitative identification algorithm combined with analytic hierarchy process is mainly for the overall trend, peak value, peak position, peak number of the proportion of weighting indicators for parameter optimisation, starting from the initial parameters, the range of the search parameter is set to 0.2-0.7, with a step size of 0.05, and the enumeration method is used to optimise the parameter search for the analytic hierarchy process qualitative identification model after the overall trend is characterised with the original spectral data and fuzzy characterisation, and the results of the search for the optimisation are shown in Fig.8 .

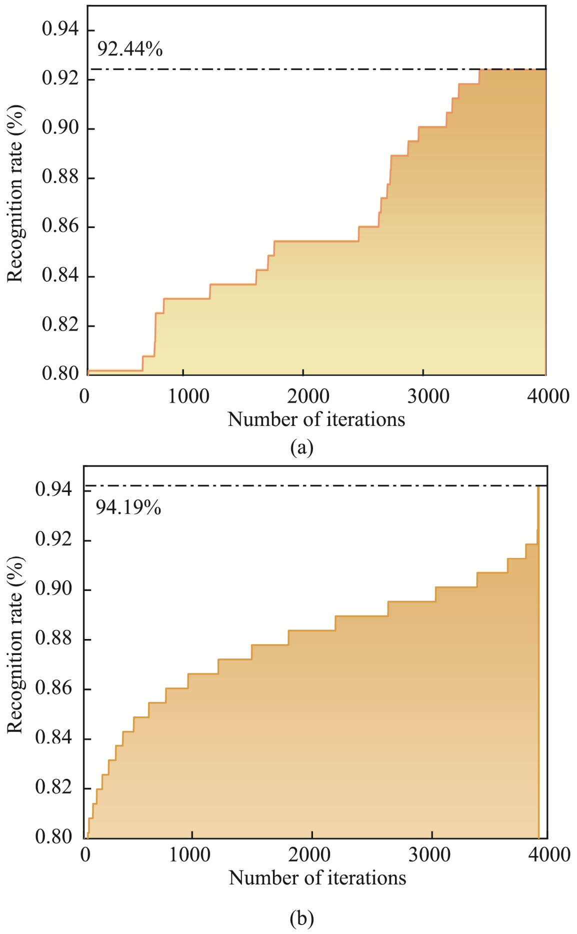


Fig.8 Optimisation results of AHP weight index
(a) YS+AHP; (b) MH+AHP

Table 5 shows the recognition accuracy and corresponding parameters of each of the fuzzy characterisation (denoted as MH) and raw data characterisation (denoted as YS) of the overall trend under the qualitative recognition model combined with analytic hierarchy process. As shown in Table 5, the optimal recognition rate is 92.44% for YS+AHP and 94.19% for

Table 5 Identification accuracy and corresponding parameters statistical table

Data characterisation methods	parameter weighting				Qualitative identification accuracy/%
	Overall trend	Peak value	Peaks	Peak number	
YS	1	0	0	0	80.23
YS	0.6	0.1	0.2	0.1	92.44
MH	1	0	0	0	93.60
MH	0.8	0.1	0.1	0.0	94.19

MH+AHP, and the recognition rate is 80.23% using the YS method and 93.60% using the MH method when only the overall trend is considered. For the YS method, the similarity is determined based on the Euclidean distance, and since each coordinate contributes equally to the Euclidean distance, when there is a large deviation in one coordinate of the two sets of values, it will cause the final Euclidean distance to be large and affect the discriminative results; for the MH method, it will be affected by the magnitude of the value of the threshold λ while highlighting the overall trend of the spectral data.

4 Conclusions

In this paper, four common food additives (xylitol, L-alanine, sorbic acid and benzoic acid) and two illegal additives (melamine and Sudan Red I) were studied as the objects of research, and the qualitative detection method of additives based on terahertz absorption spectroscopy was investigated. The proposed algorithm combined with the hierarchical analysis method, which takes into account the overall trend of the absorption spectra, peaks, peak positions, peak numbers and their corresponding weights, can be used for the qualitative detection of additives, and the recognition accuracy can reach up to 94.19%. The results show that the qualitative identification algorithm combined with hierarchical analysis effectively avoids the inconsistency of the overall trend caused by the phenomenon of "different spectra for the same substance" and can be effectively applied to the qualitative detection of food additives. This study is of some significance to the maintenance of food safety and food quality regulatory guidelines, and also provides new ideas for the identification of spectral data affected by multiple factors.

Author Contributions:

ZHAO Miaoyu, mainly completed the design and writing of algorithms for articles. YAN Fang, mainly gave the overall idea of the algorithm and guidance in the process of writing the article. LI Wenwen, mainly completed the collection and pre-processing of experimental data in the article. LIU Yangshuo, mainly accomplished the improvement of the algorithm and the drawing of the pictures in the article.

Funding Information:

This research was funded by Key Technology Tackling Programme of Inner Mongolia, grant number 2021GG0361

This research was funded by Basic Research Operating Costs of Colleges and Universities Directly Under the Inner Mongolia Autonomous Region Project.

Data Availability:

All data supporting the results of this study are included in the manuscript and are available upon request.

Conflict of Interest:

The authors declare no competing interests.

Dates:

Received 17 July 2023; Accepted 17 December 2023; Published online 31 March 2024

References

[1] Zheng W, Wang Y L, Liu T, et al. (2018). Discussion on Food Safety Problems Caused by Food Additives in China and Their Countermeasures. *China Condiment,* 43(04), pp. 195 - 197 +200.

[2] Peishen Z. (2015). Research Progress on Detection Methods of Food Additives in Sports Health Food. The Open Cybernetics & Systemics Journal 9(1).

[3] Yang Q L, Deng X J, Sun X D, et al.(2020). Application and Research Progress of Spectral Data Fusion Technology in Food Testing.*Science and Technology of Food Industry*, 41(18), PP. 324-329.

[4] Liu W, Liu Y, Huang J Q. (2018). Application of terahertz spectroscopy in biomolecule detection. *Frontiers in Laboratory Medicine*, 2 (4), PP. 127-133.

[5] Chen T, Li Z, Zhang H, et al. (2019). Terahertz spectroscopic investigation of D- and DL-tartaric acid. *Chemical Physics Letters*, 731.

[6] Zheng Z P, Li A D, Li C Y, et al.(2021).Terahertz Time-Domain Spectral Study of Paracetamol. *Spectroscopy and Spectral Analysis,* 41(12), PP. 3660-3664. 41(18), PP.324-329.

[7] Hu X H, Liu W, Liu C H, et al. (2017). Rapid identification of producing area of coffee bean based on terahertz spectroscopy

and support vector machine. *Transactions of the Chinese Society of Agricultural Engineering*, 33 (09), PP. 302-307.

[8] Zhou B, Zhu W K, Wang Z G, et al.(2022). Identification of tobacco materials based on terahertz time-domain spectroscopy. *Transactions of the Chinese Society of Agricultural Engineering*, 38(10), PP. 310-316.

[9] Cui X W, Shen T, Liu Y L, et al. (2021). Recognition of Small-Sample Terahertz Spectrum. *Laser & Optoelectronics Progress*, 58(01), PP. 321-328.

[10] Gan L, Sheng H F, Wang Y, et al.(2021) .Data augmentation method based on improved deep convolutional generative adversarial networks. *Journal of Computer Applications*, 41(05), PP. 1305.

[11] Zhan K, Chen J W, Xu Y, et al. (2023). Rapid detection method for freshness of crayfish during freezing storage: Based on near infrared spectroscopy. *Food Science*, http:// kns.cnki.net/kcms/detail/11.2206.ts.20230726.1622.052. html.

[12] Yan F, Liu T H, Zhang J L. (2022). Study on the formation mechanism of terahertz absorption peaks of sugar isomers. *Journal of Optics*, 42(05), PP. 228-234.

[13] Zhang T Y, Li B Y, Li X Y, et al.(2023). RefractiveIndex Measurement Using Continuous Wave Terahertz Frequency-DomainSpectroscop. *pectroscopy and Spectral Analysis*, 43 (02), PP. 495-502.

[14] Dorney T D, Baraniuk R G, Daniel M, et al.(2001). Material parameter estimation with terahertz time-domain spectroscopy. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 18(7), PP. 1562.

[15] Duvillare L, Garet F, Coutaz J L, etal.(1996). A reliable method for extraction of material parameters in terahertz time-domain spectroscopy. *in IEEE Journal of Selected Topics in Quantum Electronics,* 2(3), PP. 39.

[16] Zheng F Z, Ning F, Wang H L, et al.(2022). Qualitative analysis of reliability on servo stabilization platform of electro-optical system.*Journal of Applied Optics*, 43(05), PP. 853-858.

[17] Yanling M. (2023). Research on Safety Risk Assessment Method of Wind Power Enterprises Based on Hybrid Analytic Hierarchy Process. *Journal of Physics: Conference Series*, 2418(1).