

## ALGORYTMY PRZETWARZANIA WIDM RAMANA PODCZAS DETEKЦИИ SUBSTANCJI CHEMICZNYCH

Andrzej KWIATKOWSKI

Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Katedra Metrologii i Optoelektroniki,  
ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk Wrzeszcz.

tel.: +48 58 3476368, e-mail: andrzej.kwiatkowski@pg.gda.pl.

**Streszczenie:** W pracy zostaną przedstawione algorytmy wstępnego przetwarzania widm mające na celu usunięcie lub zmniejszenie intensywności zakłóceń obecnych w rejestrowanych widmach Ramana, ze szczególnym uwzględnieniem automatyzacji tego procesu (bez konieczności wyboru parametrów przetwarzania przez obsługę urządzenia).

Kolejną prezentowaną grupę algorytmów stanowią procedury parametryzacji widm, które opisują widma w postaci zbioru kilku parametrów, np. pozycji linii widmowych, jej względnej amplitudy oraz szerokości. Parametry te są niezbędną informacją wejściową dla niektórych algorytmów detekcji. W pracy zostaną również zaprezentowane najpopularniejsze algorytmy korelacyjne oraz algorytmy porównujące położenie i amplitudę linii widmowych.

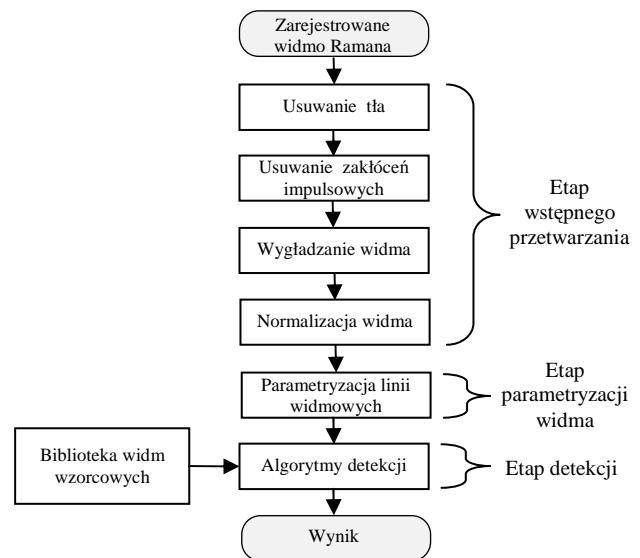
**Słowa kluczowe:** Spektroskopia Ramana, algorytmy detekcji, zakłócenia.

### 1. WSTĘP

Automatyczna detekcja substancji chemicznych w oparciu o pomiar ich widm Ramana polega na porównywaniu zarejestrowanego widma nieznanymi substancji chemicznych z widmami wzorcowymi zgromadzonymi w bazie danych. Algorytmy detekcji wskazują substancję z bazy danych, której widmo jest najbardziej podobne, według przyjętych kryteriów, do zarejestrowanego widma badanej substancji chemicznej. Niestety, poziom zakłóceń w rejestrowanych widmach Ramana często uniemożliwia skuteczną detekcję substancji chemicznych. Błędy te mają szczególne znaczenie w przypadku pomiarów polowych w warunkach terenowych, gdzie dysponuje się sprzętem przenośnym o ograniczonej dokładności wynikającej z uproszczonej budowy i ograniczeń związanych ze źródłem zasilania. Nie bez znaczenia pozostaje również fakt, że w pomiarach terenowych często jest niemożliwe zapewnienie optymalnych warunków pomiarowych takich jak np. zacielenie próbki na czas pomiaru. Wskutek tego zmierzone widma mają najczęściej znaczącą składową tła, są zaszumione oraz występują w nich zakłócenia o charakterze wąskich impulsów powodowane promieniowaniem kosmicznym. Dodatkowo w zależności od intensywności efektu Ramana zachodzącego w badanej próbce oraz od czasu ekspozycji matrycy CCD rejestrowane widma charakteryzują się różnymi poziomami sygnału.

Wspomniane problemy uniemożliwiają bezpośrednie porównywanie widm za pomocą automatycznych

algorytmów. Niezbędne jest zatem dokonanie wstępnego przetwarzania zarejestrowanych widm (Rys.1).



Rys. 1. Schemat blokowy kolejności przetwarzania widm Ramana podczas automatycznej detekcji substancji chemicznych

Następnie przetworzone widma mogą zostać poddane procedurze parametryzacji widma, która jest konieczna do wygenerowania danych wejściowych dla algorytmów detekcji opierających się na porównywaniu linii widmowych (ang. Peak Matching Algorithms). W przypadku algorytmów korelacyjnych porównujących wszystkie punkty zarejestrowanych widm (Full Spectrum Matching Methods) procedura parametryzacji widma jest zbędna.

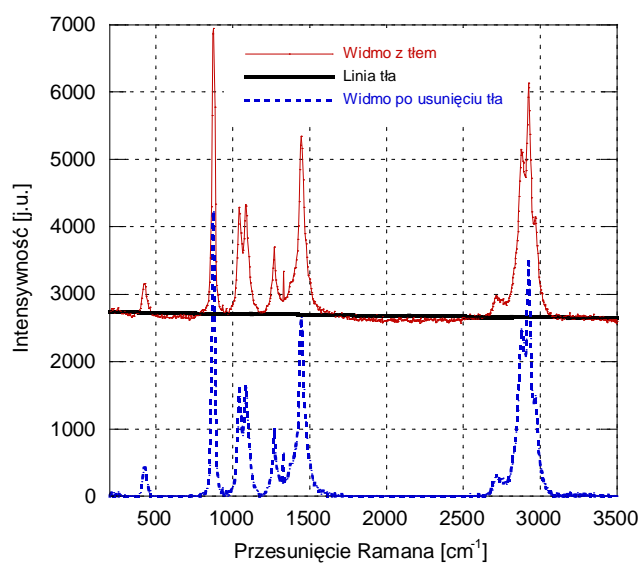
### 2. ALGORYTMY WSTĘPNEGO PRZETWARZANIA WIDM RAMANA

#### 2.1. Usunięcie tła widma

Przyjmuje się, że tło widma jest spowodowane fluorescencją, szumami przetworników CCD oraz obecnością zewnętrznego oświetlenia podczas pomiarów widm Ramana [1]. Korekta tła widma polega na aproksymacji linii tła znaną funkcją matematyczną, a następnie na odjęciu jej od zarejestrowanego widma.

Najprostszą metodą jest aproksymacją linii tła prostą lub łamaną przechodzącą przez punkty wskazane przez operatora [2]. Nieco lepsze rezultaty daje aproksymacja wielomianem. Niestety te algorytmy wymagają ingerencji operatora uniemożliwiając przez to automatyzację procesu detekcji badanej substancji chemicznej.

Jednym z algorytmów pozwalających na automatyczne wyznaczenie linii tła widma (ang. Baseline) jest algorytm GIFTS [2, 3, 4]. Algorytm opiera się na iteracyjnym aproksymowaniu widma linią prostą wyznaczaną metodą najmniejszych kwadratów i odrzucaniu punktów widma powyżej wyznaczonej linii prostej. Algorytm jest przerywany po osiągnięciu kryterium zbieżności lub zakończeniu zadanej liczby iteracji. Rysunek 2 obrazuje działanie opisanego algorytmu na przykładowym widmie Ramana.



Rys. 2. Efekt działania algorytmu GIFTS na przykładzie widma etanolu

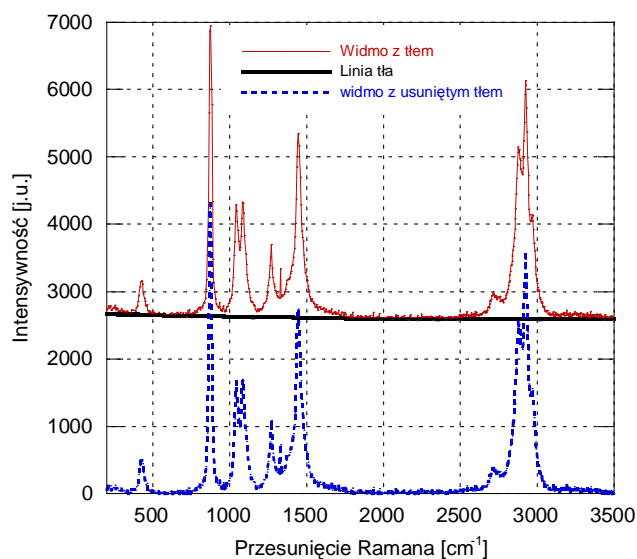
Lepsze efekty w usuwaniu tła widma daje iteracyjne dopasowanie wielomianem zadanego stopnia (Rys. 3.). W kolejnych iteracjach wszystkie punkty powyżej dopasowanej linii tła są odcinane i zastępowane punktami obliczonego wielomianu. Tak zmodyfikowane nowe tymczasowe widmo jest ponownie aproksymowane wielomianem, po czym procedura się powtarza. Algorytm przerywa się po zadanej liczbie iteracji. Należy podkreślić, że w niektórych przypadkach stosowanie wielomianów wyższych rzędów niż trzeciego może spowodować pojawienie się niekorzystnych zafalowań w linii tła.

## 2.2. Usuwanie zakłóceń impulsowych (pików)

Zakłócenia impulsowe pojawiające się w widmach Ramana mają charakter bardzo wąskich pasm (często szerokości jednego lub dwóch pikseli matrycy CCD). Ich źródłem jest najczęściej promieniowanie kosmiczne. Rysunek 2 i 3 ilustruje tego typu zakłócenie w widmie etanolu. Przy ok.  $1330\text{cm}^{-1}$  obserwowana jest linia widmowa, która jest zbyt wąska aby pochodziła od zjawiska Ramana.

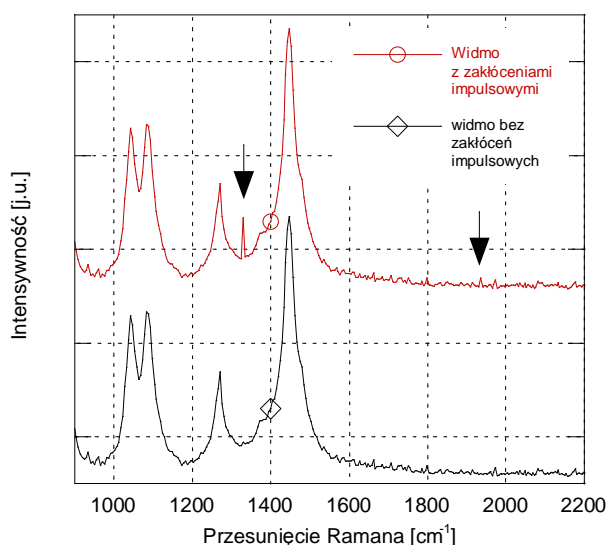
Algorytm usuwania takich zakłóceń opiera się na wykrywaniu szybko zmiennych linii widmowych i zastępowaniu ich aproksymowaną na podstawie kilku sąsiednich próbek wartością (Rys. 4.). Algorytm wpływa

tylko na zakłócone obszary widma, pozostawiając pozostałe bez zmian.



Rys. 3. Efekt działania algorytmu iteracyjnej aproksymacji widma wielomianem 3 stopnia na przykładzie widma etanolu

Wykrywanie pików (ang. Spikes) następuje poprzez badanie parametrów statystycznych (wartość średnia i odchylenie standardowe) w przyjętym oknie. Zakłada się, że pikiem jest punkt w którym lokalne odchylenie od wartości średniej jest większe niż 2,5-krotna wartość odchylenia standardowego w rozpatrywanym oknie (ramce) [5]. Wówczas wykryty punkt jest zamieniany wartością aproksymowaną (ważona aproksymacja wielomianem czwartego stopnia) na podstawie punktów sąsiednich. Wagi dla wszystkich sąsiednich punktów mają wartość 1, a w wykrytym punkcie waga równa jest 0. Procedura powtarzana jest dopóki wartość odchyżeń lokalnych w rozpatrywanym punkcie nie będzie mieścić się w zakresie  $1\sigma$  lub po osiągnięciu zadanej liczby iteracji.



Rys. 4. Wynik działania algorytmu usuwającego zakłócenia o charakterze impulsowym z rejestrowanych widm Ramana. Wykryte piki zaznaczono strzałkami.

### 2.3. Wygładzanie i normalizacja widma.

Wygładzanie ma na celu redukcję występującej w widmie składowej losowej – szumów. W spektroskopii Ramana najczęściej używanymi metodami filtracji sygnałów są: Transformata Fouriera (FFT oraz IFFT), wygładzanie dwumianowe oraz filtracje Savitzky'ego- Golay'a [3]. Obecnie upowszechniają się też techniki wykorzystujące transformację Falkową [5]. Zagadnienia wygładzania widm Ramana były przedmiotem publikacji [2, 6].

Normalizacja widma Ramana jest niezbędna do porównania widm algorytmami detekcyjnymi. Najczęściej widma normalizuje się do jedności. Po przeprowadzeniu procedury usunięcia tła widma oraz wygładzeniu wszystkie próbki widma są dzielone przez wartość maksymalnej próbki widma. W ten sposób otrzymuje się widmo znormalizowane w zakresie wartości próbek od zera do jedności niezależnie od intensywności obserwowanego efektu Ramana, mocy promieniowania pobudzającego i czasu ekspozycji matrycy światłoczułej (detektora). Normalizacja umożliwia bezpośrednią ocenę względnych amplitud wykrytych linii widmowych w porównywanych widmach.

### 2.4. Parametryzacja widma.

Parametryzacja widma jest niezbędna do pracy algorytmów określających podobieństwo widm na podstawie wybranych parametrów linii widmowych (Peak Matching Algorithms). Do tych parametrów należą najczęściej: względne amplitudy linii widmowych, ich położenie na osi przesunięcia Ramana oraz ich szerokość połówkowa (FWHH). Po wykonaniu parametryzacji linii widmowych (widma) otrzymuje się w zależności od ilości wykrytych linii widmowych od kilku do kilkudziesięciu liczb jednoznacznie opisujących widmo zamiast wektorów o długości nawet kilku tysięcy punktów. Metody określania wspomnianych parametrów linii widmowych były już przedmiotem publikacji [2 i 6].

## 3. Algorytmy detekcji

Detekcja substancji chemicznej następuje w oparciu o porównanie zarejestrowanego widma badanej substancji chemicznej ze wszystkimi widmami znajdującymi się w bibliotece widm wzorcowych. Algorytmy detekcyjne zwracają pewną wartość liczbowa, najczęściej w zakresie 0÷100 lub 0÷1000, która jest miarą stopnia podobieństwa porównywanych widm.

Widma wzorcowe najczęściej rejestruje się w warunkach laboratoryjnych za pomocą spektrometrów z wysoką rozdzielczością (nawet kilka tysięcy punktów dla spektrometrów fourierowskich) i na ogół widma te są poddane procedurze wstępnego przetwarzania oraz znormalizowane, aby umożliwić bezpośrednie porównanie widm.

W przypadku algorytmów korelacyjnych niezbędne jest takie dostosowanie biblioteki widm referencyjnych do stosowanego spektrometru aby możliwe było określenie wartości próbki widma w punktach pomiarowych (na osi przesunięcia Ramana) determinowanych rozdzielczością stosowanego spektrometru. W przypadku dyspersyjnych spektrometrów przenośnych należy dokonać transformacji widma z rozdzielczości kilku tysięcy punktów do kilkuset punktów, co jest zależne od rozdzielczości zastosowanej matrycy CCD. Najprostszą metodą takiej transformacji jest

liniowa aproksymacja wartości widma w wymaganym punkcie pomiarowym na podstawie otaczających go punktów widma wzorcowego.

Obecnie rozwijane są również inne rodzaje algorytmów, które są swoistymi klasyfikatorami. Przykładem takiego algorytmu jest metoda wektorów nośnych (Support Vector Machine) [7], w której punkty widm należących do zbioru uczącego (kilkaset widm) traktowane są jako punkty w wielowymiarowej przestrzeni. Zbiór uczący składa się z dwóch grup widm. Jedną grupę stanowią widma znanej substancji chemicznej w różnych realizacjach, drugą grupę stanowią widma innych różnych substancji chemicznych. W trakcie nauki algorytm dokonuje takich przekształceń tej przestrzeni aby możliwie z największym marginesem odległości rozdzielić dwie grupy widm. Na podstawie tych przekształceń jest budowany model, który zostanie wykorzystany w czasie etapu testowania. W czasie etapu testowania klasyfikator na podstawie modelu zwraca informację, czy wprowadzone widmo należy do zbioru znanej substancji chemicznej czy nie. Niewątpliwą zaletą metody jest jej wysoka zdolność detekcji (nawet do 99%). Niestety metoda wymaga zbudowania modelu dla każdej substancji chemicznej zatem w tym przypadku bibliotekę widm wzorcowych będzie stanowił biblioteka modeli. Zbudowanie takiej biblioteki jest bardzo czasochłonne i wymaga rejestracji co najmniej kilkaset widm dla każdej substancji w bibliotece.

### 3.1. Algorytmy korelacyjne

Nazwa tej grupy algorytmów wywodzi się z faktu, że formuły pozwalające na określenie współczynnika  $M$  określającego stopień podobieństwa porównywanych widm przypominają klasyczny współczynnik korelacji, który w istocie również jest jednym z algorytmów detekcyjnych stosowanym w spektrometrii Ramana. Szczegółowy opis algorytmów i ich zdolności detekcyjnych można odnaleźć w [6].

Większość z nich opiera się na formule (1), w której różną formę, zależnie od kryterium porównania, przyjmują współczynniki  $N$  i  $D$ .

$$M = \left(1 - \frac{N}{D}\right) \cdot 100 \quad (1)$$

Przykładowo kryterium porównania może stanowić wartość bezwzględna różnicy (algorytm ADV) lub kwadrat różnicy (algorytm LS), odpowiadających sobie punktów widma. Wówczas wyrażenie na współczynniki  $N$  i  $D$  przyjmą postać:

$$N_{ADV} = \sum_{i=1}^n |s_i - r_i|, \quad D_{ADV} = \sum_{i=1}^n |s_i|. \quad (2)$$

$$N_{LS} = \sum_{i=1}^n (s_i - r_i)^2, \quad D_{LS} = \sum_{i=1}^n s_i^2. \quad (3)$$

gdzie:  $s_i$  – kolejne próbki widma badanej substancji,  
 $r_i$  – kolejne próbki widma z biblioteki widm wzorcowych,  $i$  – indeks próbek,  $n$  – liczba próbek.

Istnieją również wersje podanych algorytmów, które porównują pierwszą pochodną kolejnych próbek widm (FDAV, FDLS). Algorytmy operujące na pierwszej

pochoďnej s bardziej odporne na blady pojawiajce sie podczas usuwania ta widma.

### 3.2. Algorytmy porownujce ksztalt i poozienie linii widmowych

Algorytmy korelacyjne mimo prostoty se czesto wystarczajco efektywne. Najwikszym ich wada jest fakt, e porownuj wszystkie punkty widma, rowniez fragmenty widma, w ktorych nie ma wyraznych linii widmowych. Zatem ich szybkosc dzialania bedzie silnie zaleza od ilosci probek w widmie. Dla biblioteki widm wzorcowych o wysokiej rozdzielczoci czas porownania moe byc stosunkowo dlugi.

Tej wady pozbawione se algorytmy porownujce tylko linie widmowe (Peak Matching Algorithms), w ktorych oceniane se poozienia i amplitudy linii widmowych. W pierwszym kroku nastepuje analizowanie pozycji odpowiadajcych sobie linii widmowych w porownywanych widmach [3]. W zaleznoci od roznicy w poozieniu porownywanych linii widmowych nastepuje przydzial punktow zgodnie z tablica 1.

Tablica 1. Kryteria przydzialu punktow za podobienstwo w poozieniu linii widmowych

| Wartoc bezwzgledna roznicy poozienia linii widmowych [px] lub [cm <sup>-1</sup> ] | Punkcja za podobienstwo w poozieniu linii widmowej [j.u.] |
|--|---|
| 0 lub 1  | 100   |
| 2  | 80  |
| 3  | 40  |
| 4  | 20  |
| 5  | 10  |
| 6  | 8   |
| 7  | 4   |
| 8  | 2   |
| 9  | 1   |
| 10 i wicej  | 0   |

W drugim kroku oceniana jest roznica amplitudy porownywanych wczeniej linii widmowych. Zakada sie, e porownywane widma se znormalizowane, zatem obliczona maksymalna roznica pomiedzy amplitudami linii widmowych nie moe przekraczac wartoci 1. Roznica w amplitudach mnozona jest przez 100. Nastepnie otrzymana wartoc jest odejmowana od wczeniej nadanych punktow za podobienstwo w poozieniu linii widmowych.

Otrzymane dla kadej linii widmowej punkty se sumowane a nastepnie normalizowane poprzez dzielenie otrzymanej sumy przez liczbe wykrytych pikow w badanej probce. W ten sposob otrzymujemy parametr w zakresie

0-100 oceniajcy podobienstwo poozienia pikow w calym widmie.

### 4. WNIOSKI KONCOWE

Przestawione w artykule zagadnienia opisuj proces automatycznego przetwarzania widm Ramana celem dokonania detekcji substancji chemicznych na podstawie porownania zarejestrowanego widma Ramana z widmami zgromadzonymi w bibliotece widm wzorcowych.

Algorytmy detekcji wskazuj substancj chemiczn (lub kilka) ktorych widmo jest najbardziej podobne wedug przyjetych kryteriow do widma badanej substancji chemicznej. Urzadzenia, ktorych dzialanie opiera sie na przedstawionych algorytmach, umozliwiaj szybki detekcj substancji chemicznej bez koniecznoci szerokiej wiedzy z zakresu chemii i fizyki.

### 5. PODZIEKOWANIA

Praca badawcza realizowana w ramach projektu systemowego „InnoDoktorant – stypendia dla doktorantow, V edycja”. Projekt jest wspofinansowany przez Uni Europejsk z Europejskiego Funduszu Spoecznego (Program Operacyjny Kapita Ludzki, Priorytet VIII, Dzialanie 8.2, Poddzialanie 8.2.2: „Regionalne Strategie Innowacji”).

### 6. BIBLIOGRAFIA

1. Laserna J.J.: Modern Techniques in Raman Spectroscopy, John Wiley & Sons, Chichester 1996.
2. Kwiatkowski A., Smulko J., Gnyba M., Wierzba P.: Algorytmy wykrywania substancji chemicznych na podstawie pomiarow widm Ramana, Pomiary Automatyka Kontrola 2010 nr 9, s.1081-1084.
3. Pomoc oprogramowania GRAMS Spectroscopy Software Suite 9.0.
4. Azarraga L.V., Hanna D.A.: GIFTS, Athens ERL GC/FT-IR Software and User's Guide.
5. Fochesatto J., Sloan J., „Signal processing of multicomponent raman spectra of particulate matter. International Journal of High Speed Electronics and Systems, Vol. 18, No.18, 2009
6. Kwiatkowski A., Gnyba M., Smulko J., „Algorithms of chemicals detection using Raman spectra”, Metrology and Measurement Systems, Vol. XVII, No. 4, 2010
7. Cristianini N., Shawe-Taylor J.: An Introduction to Support Vector Machines. Cambridge University Press, 2000.

## ALGORITHMS OF RAMAN SPECTRA PROCESSING DURING CHEMICALS DETECTION

Detection of chemicals on the basis of their Raman spectra measurement consists in comparison recorded spectra for the tested chemical substances with model spectra stored in the master database. Detection algorithms indicate the substance of the database whose spectrum is the most similar, according to established criteria, to the recorded spectrum of the tested chemical. In the paper, there will be presented spectra preprocessing algorithms to eliminate or reduce the intensity of the noise present in the recorded Raman spectra, with particular emphasis on automation of this process (no need to select the parameters processed by the operation of the device). Another presented group of algorithms are spectra parameterization procedures that describe the spectrum as a set of several parameters, such as the position of the spectral lines, the relative amplitude and width. These parameters are necessary input information for some detection algorithms. The paper will also be presented to the most popular correlation algorithms.

**Keywords:** Raman spectroscopy, detection algorithms, noise.