

**dr inż. Rafał Korycki**

*ekspert z zakresu inżynierii dźwięku i obrazu Biura Badań Kryminalistycznych ABW*

*r.korycki@abw.gov.pl*

## **Badanie autentyczności cyfrowych nagrań fonicznych utrwalonych w plikach MP3**

---

### **Streszczenie**

W pracy nakreślono problem wykrywania nieciągłości w nagraniach dźwiękowych poddanych stratnej kompresji i zaprezentowano nowe metody, które mogą być wykorzystane do badania autentyczności cyfrowych zapisów fonicznych. Prezentowane rozwiązania bazują na statystycznej analizie danych obliczonych na podstawie wartości współczynników MDCT. Wyznaczone wektory składające się z 228 cech zostały użyte jako sekwencje treningowe dwóch algorytmów uczenia maszynowego z nadzorem: liniowej analizy dyskryminacyjnej (LDA) oraz maszyny wektorów nośnych (SVM). Detekcja wielokrotnej kompresji została wykorzystana zarówno do wykrywania modyfikacji zapisu, jak również do ujawniania śladów montażu w cyfrowych nagraniach fonicznych. Skuteczność algorytmów służących do wykrywania nieciągłości została przetestowana na specjalnie przygotowanej w tym celu bazie nagrań muzycznych składającej się z blisko miliona plików MP3. Wyniki badań zostały omówione w kontekście wpływu parametrów kompresji na możliwość detekcji ingerencji w zapis cyfrowych nagrań fonicznych.

**Słowa kluczowe** badanie autentyczności nagrań cyfrowych, wykrywanie montażu, badanie dowodów cyfrowych, podwójna i wielokrotna kompresja MP3, MDCT, metody uczenia maszynowego z nadzorem, maszyna wektorów nośnych (SVM), liniowa analiza dyskryminacyjna (LDA)

---

### **Wstęp i cel pracy**

Badania autentyczności nagrań fonicznych, które polegają na badaniu oryginalności zapisów oraz na wykrywaniu śladów ingerencji w ich ciągłość, należą do istotnych zadań współczesnej kryminalistyki. Podstawowy zakres badań związany z analizą autentyczności nagrań fonicznych został zdefiniowany orzeczeniem Sądu Najwyższego w sprawie o sygn. akt III K 49/61 z 10 marca 1961. W sentencji zapisano wówczas [1]:

„...dowód z taśmy magnetofonowej stanowiącej dowód rzeczowy wymaga przeprowadzenia z kolei dowodu na okoliczność identyczności zarówno utrwalonych głosów, jak i samej taśmy, a także braku w niej zmian”.

Zgodnie z przesłaniem cytowanego orzeczenia zadaniem biegłych wykonujących opinie na zlecenie organów wymiaru sprawiedliwości i organów ścigania jest ustalenie, czy analizowane nagranie jest oryginalne (swoiste, identyczne), jak również ujawnienie ewentualnych śladów ingerencji w jego ciągłość. Można tu również przytoczyć definicję nagrania autentycznego według standardu AES27–1996 opra-

cowanego przez międzynarodową organizację AES (Audio Engineering Society), która brzmi następująco (przekład własny autora) [2]:

„Nagranie wykonane jednocześnie wraz ze zdarzeniami akustycznymi, które powinny zostać zarejestrowane w sposób w pełni i całkowicie spójny z technikami rejestracji, o których poinformowała strona, która je wytworzyła; nagranie wolne od niewyjaśnionych artefaktów, przeróbek, wstawień, wykasowań lub edycji”.

Powyższa definicja również wskazuje na konieczność badania oryginalności zapisów oraz wykrywania śladów wykonanego montażu, precyzując przy tym potrzebę analizy zgodności zastosowanych technik rejestracji z zastosowanymi podczas wytwarzania nagrania. Ponadto zwrócono tu uwagę na integralność zapisu z całokształtem zdarzeń akustycznych towarzyszących nagraniu.

Metody badania autentyczności zapisów fonicznych ulegały zmianie na przestrzeni lat. Większość z nich była tworzona z myślą o analizie nagrań zapisanych w technice analogowej na taśmach magnetycznych. Po upowszechnieniu się zapisu cyfrowego części dotychczas wykorzystywanych metod, np.

obserwacja przebiegu czasowego i spektrogramu nagrania, używano także do oceny integralności nagrań poddanych cyfryzacji [3, 4]. Ingerencje w ciągłość zapisu cyfrowego można wykryć, stosując metodę polegającą na analizie wahań częstotliwości prądu sieci energetycznej (*ENF Criterion*) [3, 5, 6, 7]. Metoda badawcza została zatwierdzona i jest powszechnie stosowana przez ekspertów należących do Europejskiej Sieci Instytutów Nauk Sądowych (European Network of Forensic Science Institutes – ENFSI) [8]. Może być jednak stosowana tylko wówczas, gdy sygnał o częstotliwości sieci elektroenergetycznej zaindukuje się w urządzeniu rejestrującym i zostanie ujawniony w obrębie nagrania. Wówczas należy wyznaczyć chwilowe wartości częstotliwości tegoż sygnału, a następnie porównać ich zmiany z wartościami częstotliwości prądu sieci elektroenergetycznej pochodzącymi z bazy danych (np. własnej lub uzyskanej od operatora sieci elektroenergetycznej).

Sygnał o częstotliwości prądu sieci elektroenergetycznej jest z punktu widzenia producentów urządzeń rejestrujących sygnałem zakłócającym, w związku z tym stosowane są różne zabiegi techniczne mające na celu utrudnienie jego utrwalania w nagraniach, np. ekranowanie elementów toru analogowego lub stosowanie górnoprzepustowych filtrów cyfrowych. Ponadto zdarza się, że zaindukowany sygnał związany z przepływem prądu w sieci elektroenergetycznej charakteryzuje się małym odstępem od szumu lub innych zakłóceń, co utrudnia lub niekiedy uniemożliwia skuteczne wykorzystanie metody *ENF Criterion*. W związku z powyższym oraz mając na uwadze rosnącą liczbę zleceń badania autentyczności nagrań cyfrowych ze strony organów wymiaru sprawiedliwości i organów ścigania, konieczne było opracowanie skutecznych metod wykrywania montażu w cyfrowych nagraniach fonicznych.

W przypadku analizy zapisów wizyjnych dostępnych jest dużo więcej rozwiązań służących do badania ich autentyczności niż ma to miejsce w przypadku nagrań fonicznych. Zaproponowano wiele metod polegających m.in. na detekcji procesu podwójnego kwantowania w nagraniach wideo skompresowanych koderem MPEG-2 [9], analizie blokowych artefaktów na podstawie różnic w błędzie kwantowania pomiędzy sąsiednimi blokami [10] czy też badaniu zmian rozdzielczości świadczących o wkomponowaniu w badany obraz innego obrazu [11, 12]. Opracowano również metody służące zarówno do analizy zapisów wizyjnych poddanych stratnej kompresji, jak i niepoddanych kompresji [13]. Dzięki temu możliwe jest obecnie wykrywanie naruszenia integralności pojedynczych zdjęć, a także całych sekwencji filmowych.

Wspomnianych metod nie można jednak wykorzystać do badania nagrań fonicznych, m.in. ze względu na fundamentalne różnice w budowie koderów obrazu i dźwięku. Pomimo to można znaleźć pewne podobieństwa między nimi. Zastosowanie stratnej kompresji cyfrowych danych fonicznych powoduje, że sąsiednie próbki zapisu fonicznego stają się w większym stopniu zależne od siebie (skorelowane). Także podział na ramki i dyskretyzacja w dziedzinie częstotliwości pozostawiają pewne „ślady” w nagraniu, które można zidentyfikować. W wielu publikacjach autorzy zwracali już uwagę na możliwość pozyskiwania informacji o zastosowanym koderze na podstawie skompresowanego nagrania [14, 15]. Opisywano sposoby określania przepływności bitowej (*bit rate*), z jaką zakodowane zostało nagranie [16, 17] oraz metody wykrywania podwójnej kompresji [18, 19]. Prezentowano również rozwiązania umożliwiające detekcję nieciągłości w nagraniach poddanych kompresji i zdekodowanych, wykorzystujące właściwości zmodyfikowanej transformacji kosinusowej [20].

Celem przedmiotowej pracy jest przedstawienie wyników badań, które polegały na opracowaniu nowych metod badania autentyczności cyfrowych nagrań fonicznych. Autor skoncentrował się na analizie nagrań poddanych stratnej kompresji z użyciem algorytmu MPEG-1 Layer 3 [21]. Wybór ten został podyktowany bogatą literaturą przedmiotu, dostępnością dokumentacji oraz powszechnym zastosowaniem algorytmu MP3. Warto nadmienić, że zaprezentowane w niniejszej pracy metody można stosować również w odniesieniu do innych algorytmów kompresji, w których użyto zespołów filtrów DCT (*Discrete Cosine Transform*) o perfekcyjnej lub quasi-perfekcyjnej rekonstrukcji, np. AAC (*Advanced Audio Coding*) lub Ogg Vorbis [20, 22]. Bazując na rozwiązaniach opisanych w wymienionych publikacjach, autor zaproponował własne algorytmy służące zarówno do predykcji pierwotnej przepływności bitowej, detekcji podwójnej i potrójnej kompresji, jak również do wykrywania montażu w nagraniach poddanych wielokrotnym procesom kodowania i dekodowania. Aby sprawdzić skuteczność opracowanych metod, stworzono własną bazę danych nagrań fonicznych składającą się z blisko miliona plików MP3. W opinii autora, takie podejście sprzyja obiektywizacji prowadzonych badań oraz umożliwia łatwe porównywanie metod badawczych. Ponadto wykorzystano wciąż popularne algorytmy uczenia maszynowego, które mogą być przydatne w rozpoznawaniu i klasyfikacji rodzajów modyfikacji, jakim poddano nagrania (np. ponowna kompresja z użyciem innych parametrów, usunięcie fragmentu zapisu itp.). Badania pokazały, że dzięki odpowiednio przygotowanym sekwencjom

uczającym jest możliwe stworzenie algorytmu umożliwiającego badanie autentyczności nagrań poddanych kompresji MP3.

Na potrzeby opisu przeprowadzonych badań autor postanowił wprowadzić następujące określenia: „nagrania źródłowe”, odnoszące się do fragmentów utworów muzycznych pochodzących z oryginalnych płyt CD oraz „nagrania oryginalne” w odniesieniu do tychże fragmentów poddanych pojedynczej kompresji.

Badania były prowadzone w ramach projektu pt.: „Projektowanie badań empirycznych i analizy materiałów dotyczących specyfiki metod kryminalistyki w pracy służb specjalnych służb porządku publicznego” finansowanego przez Narodowe Centrum Badań i Rozwoju w formie grantu nr 0023/R/ID3/2012/02.

### Wybrane zagadnienia związane z kompresją MP3

Standard MPEG-1 powstał jako efekt pracy grupy badawczej MPEG (Moving Pictures Expert Group), która została utworzona w wyniku połączenia komisji technicznych dwóch organizacji międzynarodowych: ISO (International Organization for Standardization) oraz IEC (International Electrotechnical Commission). Jej zadaniem było opracowanie wytycznych dla algorytmów kompresji obrazów ruchomych wraz z dźwiękiem, wymagających prędkości transmisji mniejszej niż 1,5 Mbps (megabita na sekundę) [22]. W standardzie MPEG-1 sygnał foniczny może być kodowany z wykorzystaniem trzech różnych algorytmów, tzn. na trzech różnych poziomach/warstwach (*layers*), nazywanych MP1, MP2 oraz MP3. Norma dopuszcza kompresję sygnału monofonicznego lub stereofonicznego, próbkowanego z częstotliwością 32 kHz, 44,1 kHz oraz 48 kHz, zapewniając przepływność bitową od 32 do 448 kbps (w przypadku MP3 do 320 kilobitów na sekundę) [21]. Biorąc pod uwagę warstwę trzecią standardu MPEG-1, oznacza to dla nagrań stereofonicznych próbkowanych z częstotliwością 44,1 kHz współczynnik kompresji w przybliżeniu od 1 : 44 do 1 : 4.

Sygnał akustyczny na wejściu koderza MP3 jest grupowany w ramki o długości 1152 próbek i podawany do filtra polifazowego (pierwszy zespół filtrów), który dokonuje podziału tego sygnału w dziedzinie częstotliwości na 32 podpasma o identycznej szerokości (wyrażonej w hercach). Każdy z 32 sygnałów podpasmowych jest następnie rozkładany na 6 lub 18 podkanałów (drugi zespół filtrów), tzn. próbki sygnału uzyskane dla każdego z podpasm są mnożone przez jedno z czterech predefiniowanych okien

czasowych, po czym wyznaczana jest dla nich zmodyfikowana dyskretna transformacja kosinusowa (MDCT – *Modified Discrete Cosine Transform*) [23]. W efekcie dla każdej ramki sygnału wejściowego powstaje 576 (okna *długie* oraz dwa rodzaje okien przejściowych: *start* i *stop*) lub trzy realizacje po 192 (okna *krótkie*) prążki widma. Decyzję o tym, który z podziałów zastosować, podejmuje się na podstawie wartości entropii psychoakustycznej widma fragmentu sygnału, wyznaczonej w bloku modelu psychoakustycznego [22].

Równolegle sygnał wejściowy jest grupowany w ramki o długości 1024 próbki (pozostałe próbki będące dopełnieniem do wartości 1152 są odrzucane), na podstawie których obliczane są wartości prążków widma szybkiej transformacji Fouriera FFT (*Fast Fourier Transform*) w celu wyznaczenia modelu psychoakustycznego. Zadaniem modelu jest dostarczenie informacji pomocnej w przełączaniu długości okna czasowego w drugim zespole filtrów oraz współczynników wykorzystywanych do kwantowania 576 próbek podpasmowych (po drugim zespole filtrów) zgrupowanych w predefiniowanych podzakresach częstotliwości odzwierciedlających występowanie pasm krytycznych (*scalefactor bands*) [22]. Model psychoakustyczny wykorzystuje efekt maskowania dźwięku opierający się na właściwości słuchu ludzkiego polegającej w uproszczeniu na różnej percepcji danych sygnałów fonicznych w obecności innych dźwięków. Dla każdego pasma krytycznego o danej szerokości, w którym występuje sygnał o zadanej częstotliwości, można wyznaczyć krzywą maskowania, poniżej której inne dźwięki (występujące w tym paśmie oraz poza nim) nie będą słyszalne [22]. Odległość między poziomem tonu maskującego a krzywą maskowania jest określana jako SMR (*Signal to Mask Ratio*). Jeśli sygnałem maskowanym będzie szum kwantowania, to w tym paśmie krytycznym nie będzie on słyszalny, dopóki wartość SNR (*Signal to Noise Ratio*) będzie większa niż SMR [22]. Następnie wartości 576 próbek widma są kwantowane z dokładnością zależną od estymowanych parametrów związanych z krzywą maskowania, aby możliwie najlepiej „ukryć” szum kwantowania. Ostatnim etapem procesu kodowania stratnego jest tworzenie strumienia danych i zapis do pliku.

Niezwykle istotny z punktu widzenia wykrywania nieciągłości w nagraniach poddanych stratnej kompresji MP3 (jak również w odniesieniu do innych algorytmów kompresji, w których użyto zespołów filtrów DCT o perfekcyjnej lub quasi-perfekcyjnej rekonstrukcji) jest proces zerowania części z 576 próbek widma (uzyskiwanych po drugim zespole filtrów), realizowany w koderze podczas wykonywania operacji kwantowania. Ponadto, dzięki jednej z własności



transformacji MDCT istnieje możliwość obserwacji minimów charakterystyki widmowej zdekodowanego nagrania, występujących w miejscach zerowanych prążków widma. Poniżej przedstawiono zależność pozwalającą na wyznaczenie transformacji kosinowej na podstawie  $N$  próbek sygnału  $x[n]$  pogrupowanych w ramki z 50% nakładaniem [24]:

$$X_{(p)}[k] = \frac{2}{N} \sum_{n=0}^{N-1} x_{(p)}[n] \cdot h[n] \cdot \cos\left(\frac{\pi}{N} \cdot \left(n + \frac{N+1}{2}\right) \cdot (k+0,5)\right) \quad (1),$$

gdzie  $0 \leq k \leq N-1$ ,  $p$  jest numerem ramki, natomiast  $h[n]$  jest funkcją okna czasowego. Po zastosowaniu odwrotnej transformacji MDCT uzyskuje się  $2N$  próbek, które także nakładają się na siebie [24]:

$$\hat{x}_{(p)}[n] = \frac{2}{N} \sum_{k=0}^{N-1} X_{(p)}[k] \cdot \cos\left(\frac{\pi}{N} \cdot \left(n + \frac{N+1}{2}\right) \cdot (k+0,5)\right) \quad (2),$$

gdzie  $0 \leq n \leq 2N-1$ . W celu wyeliminowania zniekształceń (aliasingu) konieczne jest zastosowanie procedury OLA (*overlap-and-add*), w której odwrotna transformacja MDCT wyznaczana jest dla poprzedniej i kolejnej ramki [25]. Następnie każdy z nakładających się segmentów jest mnożony przez odpowiadające mu okno, po czym segmenty w dziedzinie czasu są sumowane [26]:

$$x_{(p)}[n] = \begin{cases} \hat{x}_{(p-1)}[n+N] \cdot h[N-n-1] + \hat{x}_{(p)}[n] \cdot h[n], & 0 \leq n \leq N-1 \\ \hat{x}_{(p)}[n] \cdot h[2N-n-1] + \hat{x}_{(p+1)}[n-N] \cdot h[n-N], & N \leq n \leq 2N-1 \end{cases} \quad (3).$$

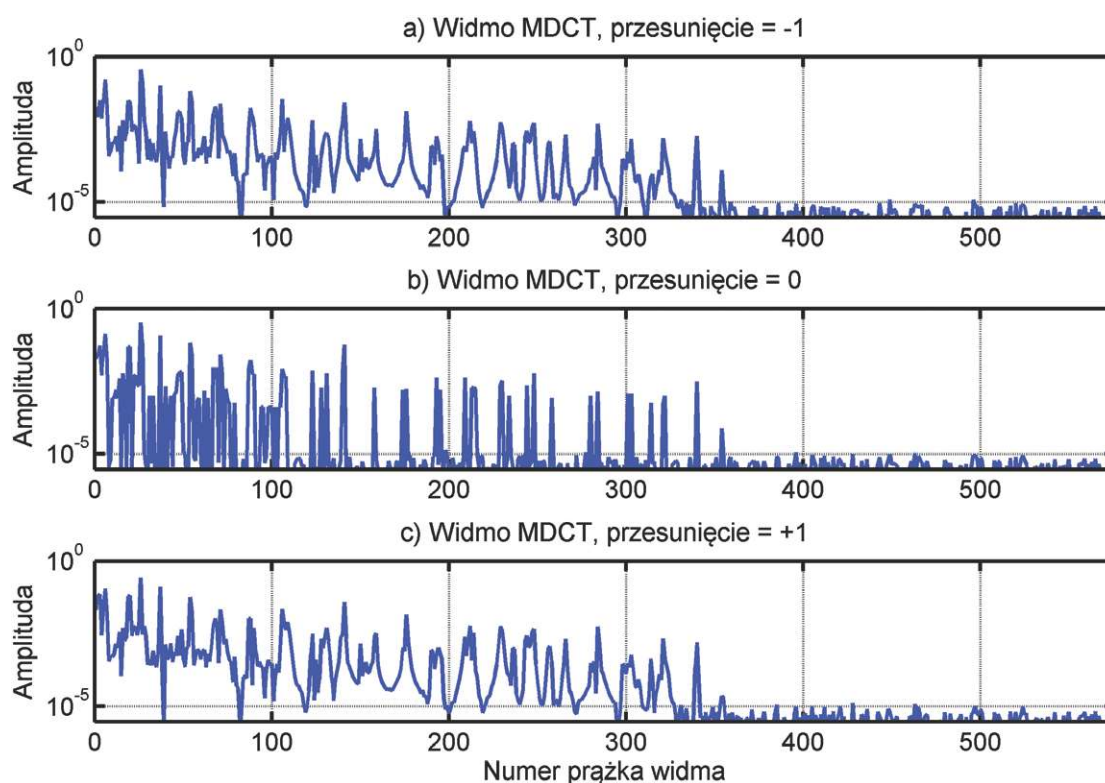
Jeśli sygnał wykazuje lokalną symetrię [26]:

$$\begin{cases} \tilde{x}_p[n] = \tilde{x}_p[N-n-1], & 0 \leq n \leq N-1 \\ \tilde{x}_p[n] = -\tilde{x}_p[3N-n-1], & N \leq n \leq 2N-1 \end{cases} \quad (4),$$

gdzie:

$$\tilde{x}_p[n] = x_p[n] \cdot h[n], \quad 0 \leq n \leq 2N-1 \quad (5),$$

wówczas wartości prążków widma po wykonaniu transformacji MDCT są zerowane [26]. Dzięki tej własności jest możliwe obserwowanie minimów charakterystyki widmowej nagrania poddanego kompresji i zdekodowanego tylko przy takiej samej pozycji okna analizy, jaka była użyta podczas procesu kodowania. Na rycinie 1 przedstawiono charakterystyki widmowe wyznaczone dla jednej wybranej ramki nagrania z wykorzystaniem okna czasowego umieszczonego w takiej samej pozycji jak podczas procesu kodowania oraz przesuniętego o jedną próbkę odpowiednio „w lewo” (przesunięcie = -1) oraz „w prawo” (przesunięcie = +1). Jak można zauważyć, brak synchronizacji na skutek przesunięcia okna nawet o jedną próbkę nie pozwala na obserwację zjawiska kwantowania poziomów próbek widma.



**Ryc. 1.** Widmo MDCT zdekodowanego fragmentu nagrania wyznaczone z wykorzystaniem okna czasowego przesuniętego: a) jedną próbkę w lewo (przesunięcie = -1), b) bez przesunięcia, c) jedną próbkę w prawo (przesunięcie = +1) w stosunku do układu ramek użytego w koderze. Amplituda widma została wyznaczona w skali logarytmicznej.

## Ekstrakcja wektora cech

Na podstawie analizy standardu ISO [21] oraz wstępnych badań doświadczalnych został skonstruowany wektor cech umożliwiający zarówno badanie oryginalności, jak i wykrywanie montażu w nagraniach fonicznych poddanych stratnej kompresji MP3. Do jego budowy posłużyły wartości zmodyfikowanej dyskretniej transformacji kosinusowej wyznaczone na podstawie danych binarnych zakodowanych w pliku MP3 z wykorzystaniem algorytmu kompresji Huffmana.

W procesie kwantowania algorytm kompresji dokonuje zamiany wartości prążków widma MDCT na wartości dyskretne. Oznacza to, że każdej dowolnej zmiennej z zadanego przedziału zostają przypisane wartości pochodzące ze skończonego (przeliczalnego) zbioru poziomów. Proces ten ma charakter nieliniowy, co oznacza, że więcej poziomów kwantowania jest przypisywanych wartościom małym niż dużym. W związku z tym można przyjąć założenie, że małe wartości współczynników MDCT będą wykazywały większą wrażliwość na podwójną kompresję [19]. W publikacji [17] autorzy zbadali możliwość rozróżniania oryginalnych (poddanych pojedynczej kompresji) oraz skompresowanych ponownie plików MP3 na podstawie liczby współczynników MDCT o najmniejszych wartościach. Zaobserwowano, że nagrania oryginalne charakteryzują się większą liczbą współczynników przyjmujących poziomy  $\pm 1$  niż nagrania poddane podwójnej kompresji. Kolejnym spostrzeżeniem jest możliwość analizy tych współczynników w powiązaniu z zakresami częstotliwości, w jakich występują. Jest to związane z procesem filtracji, jaki jest wykonywany przed wyznaczeniem transformacji MDCT (podział na 32 podpasma w pierwszym zespole filtrów) oraz z ponownym podziałem na podzakresy częstotliwości odzwierciedlające występowanie pasm krytycznych (*scalefactor bands*) w trakcie procesu kwantowania i kodowania. Można bowiem założyć, że wartości próbek widma znajdujące się w tym samym podzakresie są ze sobą w większym stopniu skorelowane niż występujące poza danym pasmem.

Podczas analizy każdego z plików MP3 uzyskuje się macierz  $\mathbf{M}$  skwantowanych współczynników  $m_{p,i}$  zmodyfikowanej dyskretniej transformacji kosinusowej, gdzie  $0 \leq p \leq N$  oznacza numer ramki,  $0 \leq i \leq 576$  natomiast oznacza indeks prążka widma MDCT [19]:

$$\mathbf{M} = \begin{bmatrix} m_{0,0} & \cdots & m_{0,575} \\ \vdots & \ddots & \vdots \\ m_{N-1,0} & \cdots & m_{N-1,575} \end{bmatrix} \quad (6).$$

Mając na uwadze powyższe obserwacje, autor zaproponował wyznaczenie wektora 228 cech słu-

żącego do opisu własności nagrań poddanych kompresji. Na podstawie wartości elementów macierzy  $\mathbf{M}$  obliczono wektor niezerowych wartości unikalnych ( $U_i$ ), wektor zawierający liczbę wystąpień tych wartości ( $H_i$ ) oraz wektor będący miarą odległości między sąsiednimi niezerowymi wartościami unikalnymi ( $D_i$ ). Pierwszy element wektora cech ( $F1$ ) stanowił wartość średnią liczby zerowych współczynników MDCT przypadających na jedną ramkę [19]:

$$F1 = \frac{1}{N} \sum_{i=0}^{575} \sum_{p=0}^{N-1} \delta(m_{p,i} = 0) \quad (7),$$

gdzie  $N$  stanowi liczbę ramek w badanym pliku MP3, natomiast funkcja  $\delta(m_{p,i} = 0)$  przyjmuje wartości równe jedności dla  $m_{p,i} = 0$  i zero w przeciwnym przypadku. Kolejne elementy wektora cech odzwierciedlały liczbę niezerowych prążków widma ( $F2$ ), liczbę unikalnych wartości MDCT ( $F3$ ), średnią odległość między sąsiednimi niezerowymi wartościami unikalnymi ( $F4$ ) oraz zmienność i ciągłość współczynników MDCT ( $F5$ – $F7$ ) [19].

$$F2 = \frac{1}{N} \sum_{i=0}^{575} \sum_{p=0}^{N-1} |\delta(m_{p,i} \neq 0) \times m_{p,i}| \quad (8),$$

$$F3 = \dim(U_i) \quad (9),$$

$$F4 = \frac{1}{N-1} \sum_{i=0}^{N-2} D_i \quad (10),$$

$$F5 = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-2} (D_i - F4)^2} \quad (11),$$

$$F6 = \frac{1}{N-1} \sum_{i=0}^{N-2} [D_i \times (h_i + h_{i+1})] \quad (12),$$

$$F7 = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-2} \{ [D_i \times (h_i + h_{i+1})] - F6 \}^2} \quad (13).$$

Autor zaproponował również wyznaczenie trzech cech opisujących średnią liczbę wystąpień prążków skwantowanego widma MDCT przyjmujących wartości  $\pm 1$ ,  $\pm 2$  oraz  $\pm 3$ :

$$F8 = \frac{1}{N} \sum_{i=0}^{575} \sum_{p=0}^{N-1} \delta(|m_{p,i}| = 1) \quad (14),$$

$$F9 = \frac{1}{N} \sum_{i=0}^{575} \sum_{p=0}^{N-1} \delta(|m_{p,i}| = 2) \quad (15),$$

$$F10 = \frac{1}{N} \sum_{i=0}^{575} \sum_{p=0}^{N-1} \delta(|m_{p,i}| = 3) \quad (16).$$

Ponadto, ze względu na podział widma MDCT na 22 podzakresy dla okien długich (*scalefactor bands*) odzwierciedlające występowanie pasm krytycznych, dla każdego z nich wyznaczono dodatkowo cechy  $F3$ – $F7$  [19]. Podczas wstępnych badań doświad-

czalnych zauważono również wzrost skuteczności algorytmów klasyfikacji w przypadku, gdy w każdym z wymienionych podzakresów obliczono również cechy  $F1$  oraz  $F8$ . Pozostałe 64 zmienne wektora cech zostały wyznaczone na podstawie metody zaproponowanej w artykule [18]. Przedstawiony algorytm został nieznacznie zmodyfikowany przez autora w związku z koniecznością zastąpienia progu detekcji 0,00001 wartością 1, albowiem współczynniki MDCT zdekodowane na podstawie danych Huffmana mają wartości całkowite. Tak przygotowany wektor 228 cech wyznaczany dla każdego badanego nagrania (lub fragmentu zapisu) mógł następnie posłużyć jako dane wejściowe dla algorytmów klasyfikacji.

## Algorytmy uczenia maszynowego

Maszynowe uczenie z nadzorem polega na wnioskowaniu o nieznanej funkcji na podstawie sekwencji danych treningowych. Dane treningowe składają się z zestawu przykładów, które zawierają wektor danych oraz pożądany stan wyjściowy (zwany sygnałem kontrolnym). Algorytm uczący analizuje sekwencje treningowe i wnioskuje o nieznanej funkcji (modelu), którą określa się mianem klasyfikatora (w przypadku danych dyskretnych) lub funkcji regresji (dla danych ciągłych). Właściwie dobrany klasyfikator powinien poprawnie przewidywać stan wyjściowy dla każdego poprawnego zestawu danych [27].

Wielowarstwowe sieci perceptronów (MLP – *MultiLayer Perceptron*) są obecnie najpopularniejszą odmianą sztucznych sieci neuronowych (*artificial neural networks*). Model perceptronu stanowi pojedynczy neuron z siecią liniowych funkcji wag  $w$  oraz z funkcją progowej aktywacji [29]. Dane wejściowe stanowią wektor cech  $x = (x_1, x_2, \dots, x_n)$  w  $n$ -wymiarowej przestrzeni cech. Funkcja sieci stanowi ważoną sumę wejść [29]:

$$u(x) = w_0 + \sum_{i=1}^n w_i x_i \quad (17),$$

natomiast funkcję wyjścia otrzymuje się na podstawie funkcji aktywacji [29]:

$$y(x) = \begin{cases} 1 & u(x) \geq 0, \\ 0 & u(x) < 0 \end{cases} \quad (18).$$

Proces uczenia odbywa się na podstawie sekwencji treningowej w sposób następujący [29]:

$$w(k+1) = w(k) + \eta(d(k) - y(k))x(k) \quad (19),$$

gdzie  $\eta$  określa prędkość adaptacji,  $d(i) \in \{0,1\}$  jest pożądanym stanem na wyjściu klasyfikatora dla poszczególnych sekwencji treningowej  $\{(x(i), d(i))\}$ . Indeks  $k$  używany jest tu w celu wskazania, że wektory

uczące przekazywane są do perceptronu sekwencyjnie, jednakże w porządku losowym. Jeśli stan na wyjściu perceptronu  $y(k)$  jest zgodny ze stanem pożądanym  $d(k)$ , to wówczas wektor wag nie ulega zmianie. W przeciwnym przypadku wektor  $w$  jest aktualizowany.

Klasyczne sztuczne sieci neuronowe z algorytmami uczącymi wykorzystującymi minimalizację nieliniowej funkcji błędu wykazują pewne wady [30]. Funkcja błędu może przyjmować wiele lokalnych minimów, w których algorytm uczący może się zatrzymać, nie osiągając optymalnego rozwiązania. Maszyna wektorów nośnych (SVM – *Support Vector Machine*), która pojawiła się w ostatnich latach, prezentuje nowe podejście do konstruowania i trenowania sieci neuronowych, wolne od wad dotychczas stosowanych algorytmów [31]. SVM wykorzystuje algorytm uczący, maksymalizujący margines separujący pomiędzy dwiema klasami definiowanymi przez zestaw  $p$  par zawierających wektor danych wejściowych  $x$  oraz klasę  $d$ , do której przynależy [32]:

$$D = \{(x_i, d_i)\}_{i=1,2,\dots,p}, \quad x \in X, \quad d_i \in \{-1,1\} \quad (20).$$

Optymalna hiperpłaszczyzna jest zdefiniowana jako liniowa funkcja decyzyjna z maksymalnym marginesem separacji między wektorami obu klas. Konstruując taką hiperpłaszczyznę, należy użyć tylko niektórych wybranych wektorów treningowych tzw. wektorów nośnych lub wektorów wspierających (*support vectors*), które definiują margines separacji. W przypadku wektorów treningowych, które są liniowo separowalne, hiperpłaszczyzna przyjmuje postać [32]:

$$g(x) = w^T x + b = 0 \quad (21),$$

gdzie  $w$  są wektorami normalnymi do hiperpłaszczyzny,  $b$  stanowi stałą przesunięcia, natomiast  $T$  oznacza operację transpozycji. Wyznaczenie maksymalnie szerokiego marginesu sprowadza się do rozwiązania głównego problemu optymalizacyjnego [3]:

$$\min_w \frac{1}{2} \|w\|^2 \quad (22)$$

z następującym ograniczeniem:

$$d_i(w^T x_i + b) \geq 1, \quad i = 1, \dots, K \quad (23).$$

Powyższy problem optymalizacji można przedstawić za pomocą jednego wyrażenia, które minimalizuje się ze względu na  $w$  i  $b$  oraz maksymalizuje ze względu na mnożniki Lagrange'a  $(\alpha_1, \alpha_2, \dots, \alpha_K)$  [3]:

$$\min_{w,b} \max_{\alpha} \left\{ \frac{1}{2} \|w\|^T - \sum_{i=1}^K \alpha_i (d_i (w^T x_i + b) - 1) \right\} \quad (24).$$

Zależność między  $\alpha_i$  a położeniem wektora  $x_i$  względem marginesu opisują warunki Karusha-



-Kuhna-Tuckera (KKT). Wynika z nich, że współczynniki  $\alpha_i$  przyjmują wartości różne od zera tylko w tych równaniach, które zawierają wektory nośne [3, 4]. Wobec tego funkcja klasyfikacji jest zależna jedynie od niewielkiej liczby wektorów nośnych  $S$  w porównaniu z ogólną liczbą wektorów treningowych.

Inną popularną metodą wykorzystywaną do odnajdywania liniowej kombinacji cech, które najlepiej charakteryzują lub separują dwie lub więcej klas obiektów lub zdarzeń, jest analiza dyskryminacyjna (*Discriminant Analysis*) [33]. Polega ona na rzutowaniu wielowymiarowych wektorów danych wejściowych na przestrzeń o mniejszej liczbie wymiarów przez jednoczesną maksymalizację separacji danych pochodzących od różnych klas oraz minimalizację dyspersji danych należących do tej samej klasy. W ten sposób uzyskuje się maksymalną zdolność klasyfikacji wektorów danych przy jednoczesnej redukcji ich wymiarowości.

Liniowa analiza dyskryminacyjna (LDA) jest realizowana przez zebranie informacji o klasach, do których mają być przypisywane nowe obserwacje za pomocą wskaźników położenia i rozproszenia podprób próby uczącej. Wartości średnie dla klas można zdefiniować jako [34, 35]:

$$\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki} \quad (25),$$

gdzie  $k$  oznacza liczbę populacji, natomiast  $n_k$  podpróbę obserwacji z danej klasy. Przy założeniu, że  $x_k$  jest średnią grupową i wariancja wektora danych jest taka sama we wszystkich  $g$  populacjach, wspólna macierz kowariancji może zostać wyznaczona na podstawie macierzy kowariancji wewnątrzgrupowej [34, 35]:

$$W = \frac{1}{n-k} \sum_{k=1}^g (n_k - 1) S_k = \frac{1}{n-k} \sum_{k=1}^g \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)(x_{ki} - \bar{x}_k)^T \quad (26),$$

gdzie  $S_k$  jest macierzą kowariancji próby w  $k$  tej populacji, natomiast  $n = n_1 + n_2$ . Celem metody LDA jest znalezienie takiego kierunku  $a$ , który maksymalizuje odległość między zrzutowanymi średnimi prób przy uwzględnieniu wariancji rzutu [34, 35]:

$$\arg \max_a \left( \frac{(a^T \bar{x}_2 - a^T \bar{x}_1)^2}{a^T W a} \right) \Rightarrow a = W^{-1} (\bar{x}_2 - \bar{x}_1) \quad (27)$$

przy założeniu występowania dwóch klas. Klasyfikacja jest wykonywana poprzez rzutowanie obserwacji  $x$  na kierunek  $a$  i przypisanie określonej klasy w zależności od tego, czy rzut jest bliższy rzutowi środka próby pierwszej, czy drugiej. W przypadku większej liczby klas poszukuje się takiego kierunku  $a$ , który mógłby maksymalizować następujące wyrażenie [34, 35]:

$$\arg \max_a \left( \frac{a^T B a}{a^T W a} \right) \quad (28),$$

w którym  $B$  jest macierzą kowariancji międzygrupowej, natomiast  $W$  jest macierzą kowariancji wewnątrzgrupowej [34, 35]:

$$B = \frac{1}{g-1} \sum_{k=1}^g n_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})^T \quad (29),$$

$$W = \frac{1}{n-g} \sum_{k=1}^g (n_k - 1) S_k \quad (30).$$

## Baza nagrań poddanych kompresji

Do oceny skuteczności wykrywania ingerencji w zapis nagrania przez zaproponowane algorytmy detekcji wykorzystano specjalnie w tym celu zaprojektowaną i stworzoną bazę danych. Przedmiotem badań miały być materiały utrwalone w warunkach najbliższych rzeczywistym za pomocą rejestratorów o nieznanym parametrach. Takie nagrania charakteryzują się obecnością szumu i zakłóceń, jak również trudnym do przewidzenia kształtem charakterystyki widmowej. Odstąpiono wobec tego od próby dokładnego odwzorowania tych „idealnych” warunków rzeczywistych i jako materiał odniesienia wykorzystano różnego rodzaju nagrania muzyczne. Takie rozwiązanie daje możliwość zapewnienia powtarzalności badań, a podobne rozwiązania są praktykowane w środowisku naukowym, w szczególności w odniesieniu do problemu wykrywania modyfikacji w cyfrowych nagraniach fonicznych [4, 17, 18, 19]. Ponadto, należy zwrócić uwagę na brak dostępnych baz nagrań służących do badania autentyczności zapisów fonicznych. Istniejące bazy zakłóconych nagrań mowy zawierają pliki utrwalone z częstotliwościami próbkowania mniejszymi lub równymi 16 kHz, co nie odpowiada typowym wartościom opisanym w standardzie ISO dla algorytmu MP3 [21]. Warto również podkreślić, że niektóre zapisy muzyczne w doskonały sposób oddają wiele różnych artefaktów, które pojawiają się w rzeczywistych nagraniach dowodowych, np. trzaski, uderzenia (perkusia), szum pasmowy (instrumenty klawiszowe, efekty, gitary elektryczne), zakłócenia sinusoidalne (instrumenty smyczkowe).

W związku z powyższym do badań wykorzystano 2940 fragmentów utworów muzycznych (o długości 10 sekund każdy) pochodzących z 36 albumów z prywatnej kolekcji autora. Wśród plików dodanych do bazy znalazły się m.in. zapisy symfonii Ludwiga van Beethovena wykonane przez London Symphony Orchestra, trzy symfonie Wolfganga Amadeusza Mozarta pod batutą Rafaela Kubelika, kompilacja

**Tabela 1**  
**Wykaz koderów MP3 wykorzystanych w badaniach**

| Ozna-<br>czenie | Nazwa kodera/wydawca lub<br>twórca       | Wersja<br>(rok)     |
|-----------------|------------------------------------------|---------------------|
| Lame 396        | LAME/Mike Cheng, Mark Taylor<br>et al.   | 3.96.1<br>(2004)    |
| Lame 398        | LAME/Mike Cheng, Mark Taylor<br>et al.   | 3.98<br>(2008)      |
| Lame 399        | LAME/Mike Cheng, Mark Taylor<br>et al.   | 3.99.5<br>(2012)    |
| Adobe 2         | Adobe Audition/Adobe Systems<br>Inc.     | 2.0 (2005)          |
| Adobe<br>CS6    | Adobe Audition CS6/Adobe<br>Systems Inc. | 5.0 (2012)          |
| iTunes          | iTunes/Apple Inc.                        | 10.3.1.55<br>(2011) |
| 8Hz             | 8Hz MP3 Encoder/8Hz Production           | 02b<br>(1998)       |
| BladeEnc        | BladeEnc/Tord Jansson                    | 0.94.2<br>(2001)    |
| GoGo            | Gogo-no-coda/Herumi and Pen              | 3.13<br>(2004)      |
| Helix           | Helix/RealNetworks Inc.                  | 5.1 (2005)          |
| Mp3enc          | Mp3Enc (Demo)/Fraunhofer IIS-A           | 3.1 (1998)          |
| Plugger         | Plugger+ Pro/Alberto Demichelis          | 0.4 (1998)          |
| SCMPX           | SCMPX/Shinji Chiba                       | 1.51<br>(1999)      |
| Shine           | Shine/Gabriel Bouvigne                   | 0.1.4<br>(2001)     |
| SoloH           | Mpeg Encoder/SoloH                       | 0.07a<br>(1998)     |
| Real            | Real Player/RealNetworks Inc.            | 15.0.6.14<br>(2012) |
| Xing            | Xing MP3 Encoder/Xing Tech.<br>Corp.     | 1.5 (1999)          |
| Zoom            | Zoom Handy Recorder/Zoom                 | H4n (b.d.)          |
| Tascam          | Linear PCM Recorder/Tascam               | DR-40<br>(b.d.)     |
| Olympus         | Digital Voice Recorder/Olympus           | WS-812<br>(b.d.)    |
| Sony            | ICD Recorder/Sony                        | UX513F<br>(b.d.)    |

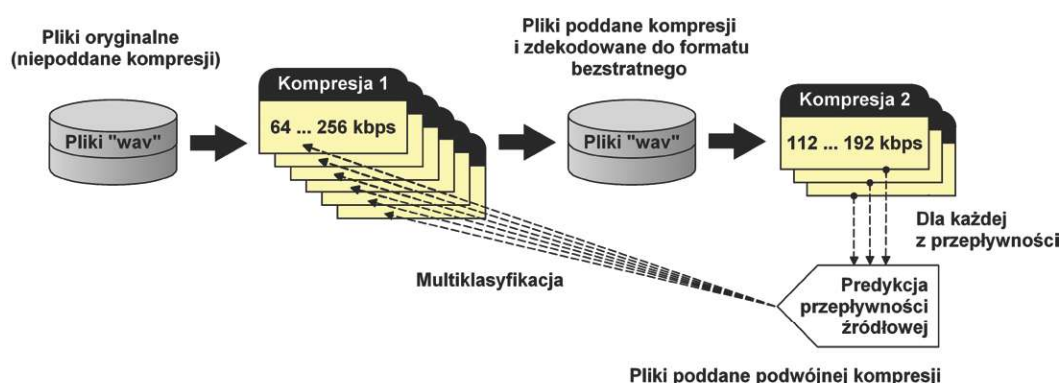
*Adagios* sygnowana przez dyrygenta Herberta von Karajana, odnowione cyfrowo wersje dzieł Raya Charlesa, po dwa albumy Norah Jones i Iron Maiden, koncertowa płyta „Pulse” Pink Floyd oraz album Beyonce pt. „4”. Wszystkie pliki źródłowe (2940) utwalone w liniowym i bezstratnym formacie LPCM (*Linear Pulse Code Modulation*) i zapisane w formacie „wav” zostały poddane kompresji za pomocą 21 różnych

realizacji koderów MP3 (zob. tabela 1), w każdym z przypadków przy zastosowaniu sześciu dostępnych przepływności: 64, 96, 112, 128, 192 oraz 256 kbps (kilobitów na sekundę). Cztery z wymienionych koderów były realizacją sprzętową w postaci przenośnych urządzeń rejestrujących (Zoom, Tascam, Olympus, Sony), zatem w ich przypadku utwory muzyczne odtwarzano za pomocą głośników aktywnych Adam P22 i nagrywano z wykorzystaniem wbudowanych mikrofonów.

Nagrania poddane kompresji za pomocą trzech wybranych implementacji algorytmu MP3: Lame 399, iTunes oraz Adobe CS6 były następnie dekodowane do formatu bezstratnego (LPCM z rozszerzeniem „wav”) z wykorzystaniem odpowiadających tym koderom dekodów zawartych w tych samych aplikacjach (w przypadku kodera Lame 399 występującego również samodzielnie użyto aplikacji Lame Front-End 1.8 zawierającej zarówno koder Lame 399, jak i dekod). Następnie kopie zdekodowanych fragmentów nagrań zostały przesunięte z wykorzystaniem bufora cyklicznego (funkcja *circshift* w środowisku obliczeniowym MATLAB [36]) o 10 ms. Wprowadzone przesunięcie miało na celu odwzorowanie procesu montażu polegającego na wycięciu lub wstawieniu fragmentu wypowiedzi, który może zostać wykonany w dziedzinie czasu po uprzednim zdekodowaniu zapisu do formatu bezstratnego. W procesie kompresji dane wejściowe są porządkowane w ramki, a następnie w takiej formie przetwarzane. Jeśli zostanie wykonany montaż, to istniejący porządek ramek zostanie naruszony. Aby przywrócić pierwotny układ, należałoby wprowadzić stosowne przesunięcie, identyczne dla każdej następnej ramki, począwszy od miejsca wykonania edycji. Zatem przez wprowadzenie przesunięcia, np. o 10 ms, można modelować zjawisko montażu w stosunku do tych fragmentów nagrania, które następują po miejscu, w którym wykonano montaż.

Dzięki powyższym zabiegom uzyskano materiał do badań nad plikami, w których wykonano montaż w postaci usunięcia lub wstawienia fragmentu zapisu. Tak przygotowane nagrania w formacie bezstratnym były następnie poddawane ponownej kompresji z wykorzystaniem tych samych trzech koderów (Lame 399, iTunes oraz Adobe CS6) z przepływnościami: 112, 128 oraz 192 kbps. Utworzona baza nagrań składała się łącznie z blisko miliona plików MP3. Badania wykonane w ramach projektu „Projektowanie badań empirycznych i analizy materiałów dotyczących specyfiki metod kryminalistyki w pracy służb specjalnych służb porządku publicznego” obejmowały różne zagadnienia, m.in. identyfikację rodzaju i wersji kodera oraz urządzeń rejestrujących na podstawie analizy plików MP3 oraz wpływ





**Ryc. 2.** Schemat tworzenia nagrań oraz prowadzenia porównań w badaniach przepływności źródłowej plików MP3 poddanych podwójnej kompresji.

modeli treningowych na dokładność klasyfikacji. Niniejsze opracowanie, ze względu na konieczność ograniczenia jego objętości, zawiera wyniki badań wykonanych na podstawie plików poddanych kompresji jedynie za pomocą kodera Lame 399. Warto dodać, że kodery z rodziny Lame są najbardziej rozpowszechnionymi implementacjami algorytmu MP3 i korzysta z nich większość darmowych programów, które mogą służyć do edycji cyfrowych nagrań fonicznych.

W badaniach wykorzystano środowisko obliczeniowe MATLAB, w szczególności narzędzia z pakietów *Signal Processing Toolbox*, *Parallel Computing Toolbox*, *Statistics Toolbox* oraz *Bioinformatics Toolbox* [36]. Do klasyfikacji wykorzystano liniową analizę dyskryminacyjną z funkcją odwracania macierzy za pomocą metody Moora–Penrose’a [33, 36] oraz maszynę wektorów nośnych z liniową funkcją jądra. W przypadku każdej z wykonywanych klasyfikacji wektory cech reprezentujące jedną grupę (np. nagrania poddane pojedynczej i podwójnej kompresji lub nagrania niezmontowane i poddane edycji) były dzielone na dwie równe sekwencje: treningową oraz testową. Sekwencja treningowa wykorzystywana była do tworzenia modeli przez algorytmy LDA oraz SVM, natomiast skuteczność klasyfikacji, a w związku z tym również dokładność wykrywania ingerencji w zapis nagrania, była sprawdzana na sekwencji testowej.

### Ujawnianie przepływności źródłowej

Za miarę jakości nagrania utrwalonego w pliku MP3 oprócz m.in. częstotliwości próbkowania przyjmuje się wartość jego przepływności bitowej (*bit rate*) wyrażonej w kilobitach na sekundę (*kbps*). Na tej podstawie można np. podjąć decyzję o zakupie cyfrowej wersji albumu w sklepie internetowym. Ponieważ wartość przepływności jest zapisana w pliku i wskazuje jedynie

na parametry ostatniej wykonanej kompresji, można się spodziewać, że niektóre nagrania mogą być celowo transkodowane (czyli dekodowane do formatu bezstratnego i kodowane ponownie z innymi parametrami, np. inną wartością przepływności) z niższej przepływności do wyższej, aby wykazać ich lepszą jakość. Takie działania stanowią oszustwo i znacznie obniżają zaufanie do handlu w sieci Internet. Nagrania mogą być również transkodowane w celu ukrycia

**Tabela 2**  
**Macierz klasyfikacji nagrań MP3 poddanych podwójnej kompresji z przepływnością wynikową 112 kbps**

| Przeptywność przewidywana | Przeptywność zastosowana podczas pierwszej kompresji [kbps] |     |       |       |       |       |
|---------------------------|-------------------------------------------------------------|-----|-------|-------|-------|-------|
|                           | 64                                                          | 96  | 112   | 128   | 192   | 256   |
| 64                        | 100                                                         | –   | –     | –     | –     | –     |
| 96                        | –                                                           | 100 | –     | –     | –     | –     |
| 112                       | –                                                           | –   | 96,26 | 2,38  | 0,34  | 1,02  |
| 128                       | –                                                           | –   | 1,36  | 92,52 | 1,36  | 2,04  |
| 192                       | –                                                           | –   | 0,68  | 1,70  | 71,43 | 15,31 |
| 256                       | –                                                           | –   | 1,70  | 3,40  | 26,87 | 81,63 |

**Tabela 3**  
**Macierz klasyfikacji nagrań MP3 poddanych podwójnej kompresji z przepływnością wynikową 128 kbps**

| Przeptywność przewidywana | Przeptywność zastosowana podczas pierwszej kompresji [kbps] |       |     |       |       |       |
|---------------------------|-------------------------------------------------------------|-------|-----|-------|-------|-------|
|                           | 64                                                          | 96    | 112 | 128   | 192   | 256   |
| 64                        | 99,66                                                       | –     | –   | –     | –     | –     |
| 96                        | 0,34                                                        | 99,32 | –   | –     | –     | –     |
| 112                       | –                                                           | 0,68  | 100 | –     | –     | –     |
| 128                       | –                                                           | –     | –   | 94,90 | 1,36  | 2,38  |
| 192                       | –                                                           | –     | –   | 1,70  | 85,71 | 10,20 |
| 256                       | –                                                           | –     | –   | 3,40  | 12,93 | 87,41 |

**Tabela 4**  
**Macierz klasyfikacji nagrań MP3 poddanych podwójnej kompresji z przepływnością wynikową 192 kbps**

| Przepływność przewidywana | Przepływność zastosowana podczas pierwszej kompresji [kbps] |       |       |     |       |       |
|---------------------------|-------------------------------------------------------------|-------|-------|-----|-------|-------|
|                           | 64                                                          | 96    | 112   | 128 | 192   | 256   |
| 64                        | 100                                                         | –     | –     | –   | –     | –     |
| 96                        | –                                                           | 99,66 | –     | –   | –     | –     |
| 112                       | –                                                           | 0,34  | 99,66 | –   | –     | –     |
| 128                       | –                                                           | –     | 0,34  | 100 | –     | –     |
| 192                       | –                                                           | –     | –     | –   | 81,97 | 17,35 |
| 256                       | –                                                           | –     | –     | –   | 18,03 | 82,65 |

śladów ingerencji wykonanej w ich treść. Wówczas można spodziewać się np. celowego obniżenia jakości nagrania przez poddanie go ponownej kompresji z zastosowaniem niższej przepływności niż pierwotna.

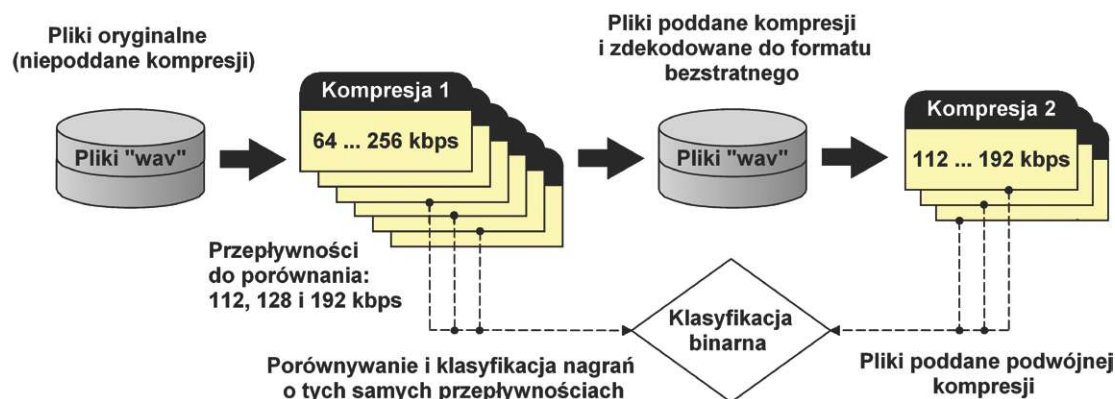
Aby sprawdzić możliwość wykrywania ingerencji w zapis nagrania polegających na transkodowaniu plików MP3, do badań wykorzystano nagrania z utworzonej przez autora bazy danych wykonane według schematu przedstawionego na rycinie 2. Pliki źródłowe będące fragmentami utworów poddano kompresji z wykorzystaniem sześciu przepływności, zdekodowano je do formatu bezstratnego, a następnie poddano ponownej kompresji, stosując tym razem trzy wybrane przepływności: 112, 128 oraz 192 kbps. Z utworzonych plików MP3 wyodrębniono wektory cech (każdy składający się z 228 elementów), po jednym dla każdego z plików. Do klasyfikacji wykorzystano liniową analizę dyskryminacyjną. W przypadku każdej z wykonanych klasyfikacji wektory cech były dzielone na dwie równe sekwencje: treningową oraz testową. Wykonywana w tym przypadku multiklasyfikacja nagrań poddanych podwójnej kompresji polegała na predykcji przepływności, jaka była użyta podczas

pierwszej kompresji. W tabelach 2–4 zamieszczono wyniki predykcji przepływności pierwotnej (podawane w procentach) w zależności od przepływności wtórnej, odpowiednio 112 kbps (tabela 2), 128 kbps (tabela 3) oraz 192 kbps (tabela 4). Przykładowo, wszystkie pliki transkodowane z przepływności 64 kbps oraz 96 kbps do przepływności 112 kbps zostały sklasyfikowane prawidłowo, natomiast liczba prawidłowych konwersji z 256 kbps do 112 kbps wyniosła 81,63% (zob. tabela 2).

Analiza powyższych wyników pokazuje, że predykcja przepływności źródłowej w nagraniach poddanych podwójnej kompresji jest dokładniejsza w przypadku przewidywania mniejszych wartości przepływności, niż używane w trakcie drugiego kodowania. Jednakże w odniesieniu do nagrań transkodowanych z przepływności większej na mniejszą, wciąż możliwe jest wykonanie prawidłowej klasyfikacji konwertowanych plików.

### Wykrywanie podwójnej i potrójnej kompresji

Predykcja przepływności źródłowej plików MP3 poddanych podwójnej kompresji, jakkolwiek interesująca, jest prowadzona przy założeniu, że nagranie po jego utrwaleniu było jeszcze raz kodowane. Konieczne jest zatem umożliwienie odróżniania nagrań oryginalnych (poddanych pojedynczej kompresji) od tych, które zostały poddane kompresji ponownie. Ponadto osoba dokonująca edycji nagrania zwykle kieruje się potrzebą modyfikacji treści wypowiedzi, zaniedbując przy tym charakterystyczne ustawienia plików fonicznych, w tym m.in. dane dotyczące kompresji. Można się zatem spodziewać, że po wykonaniu edycji, nagranie zostanie utwalone z domyślnym zestawem opcji lub z takim samym wyborem parametrów, jakie charakteryzowały zapis pierwotny.



**Ryc. 3.** Schemat tworzenia nagrań oraz prowadzenia porównań w badaniach polegających na wykrywaniu podwójnej kompresji w plikach MP3.

Mając na uwadze powyższe rozważania, postanowiono sprawdzić możliwość wykrywania zapisów poddanych podwójnej kompresji. Wykorzystano w tym celu nagrania z utworzonej przez autora bazy danych, wykonane według schematu przedstawionego na rycinie 3. Fragmenty utworów poddano kompresji z użyciem sześciu przepływności, zdekodowano je do formatu bezstratnego, a następnie ponownie skompresowano, stosując te same trzy wybrane wartości przepływności (112, 128 oraz 192 kbps). Z utworzonych plików MP3 wyodrębniono wektory cech, które poddano klasyfikacji z wykorzystaniem liniowej analizy dyskryminacyjnej oraz maszyny wektorów nośnych z liniową funkcją jądra.

W przypadku każdej z wykonanych klasyfikacji wektory cech były dzielone na dwie równe sekwencje: treningową oraz testową. Realizowana w tym przypadku klasyfikacja binarna polegała na rozróżnianiu nagrań poddanych pojedynczej i podwójnej kompresji, przy czym porównywano ze sobą wektory wyznaczone na podstawie plików o tych samych parametrach, np. pliki poddane kompresji z przepływnością 96 kbps i przekonwertowane do 128 kbps były porównywane z nagraniami poddanymi pojedynczej kompresji z przepływnością 128 kbps. W tabelach 5 i 6 zamieszczono wyniki detekcji podwójnej kompresji (podawane w procentach) w zależności od przepływności wtórnej (odpowiednio: 112 kbps, 128 kbps oraz 192 kbps). Dla przykładu, w trakcie klasyfikacji z wykorzystaniem algorytmu SVM nagrań poddanych konwersji z przepływności 192 kbps do 128 kbps oraz ich oryginalnych odpowiedników (poddanych pojedynczej kompresji z przepływnością 128 kbps) 91,16% plików zostało prawidłowo przypisanych do właściwej im klasy (zob. tabela 5).

Osoba edytująca nagranie może jednak zechcieć wykonać więcej niż jedną dodatkową kompresję, np. w celu ukrycia śladów wykonanej modyfikacji treści zapisu. W związku z tym konieczne okazało się podjęcie badań nad możliwością wykrywania tak

Tabela 5

Klasyfikacja binarna nagrań poddanych pojedynczej i podwójnej kompresji z wykorzystaniem algorytmu SVM oraz różnych wartości pierwotnej przepływności

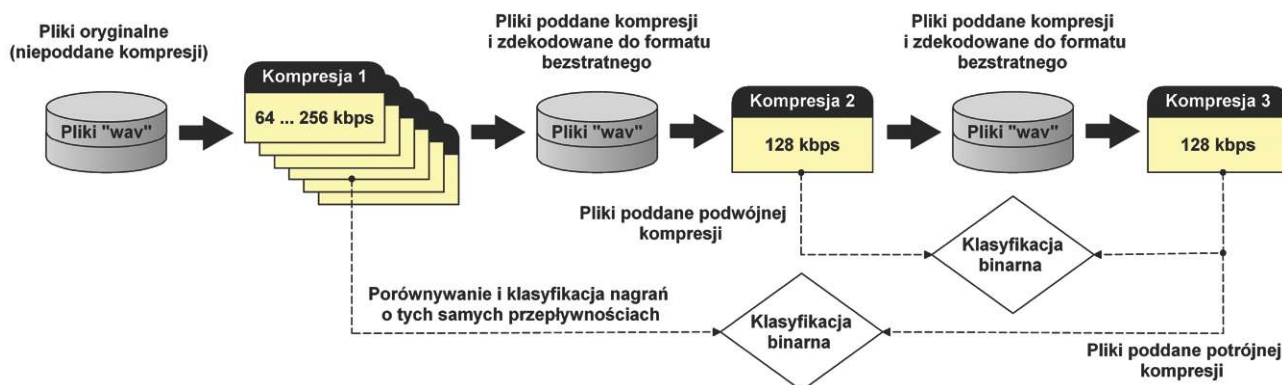
| Przepływność podczas pierwszej kompresji | Przepływność podczas ponownej kompresji [kbps] |       |       |
|------------------------------------------|------------------------------------------------|-------|-------|
|                                          | 112                                            | 128   | 192   |
| 64                                       | 100                                            | 100   | 100   |
| 96                                       | 99,83                                          | 100   | 100   |
| 112                                      | 98,64                                          | 100   | 100   |
| 128                                      | 96,26                                          | 96,09 | 100   |
| 192                                      | 89,63                                          | 91,16 | 89,80 |
| 256                                      | 75,85                                          | 82,82 | 83,67 |

Tabela 6

Klasyfikacja binarna nagrań poddanych pojedynczej i podwójnej kompresji z wykorzystaniem algorytmu LDA oraz różnych wartości pierwotnej przepływności

| Przepływność podczas pierwszej kompresji | Przepływność podczas ponownej kompresji [kbps] |       |       |
|------------------------------------------|------------------------------------------------|-------|-------|
|                                          | 112                                            | 128   | 192   |
| 64                                       | 100                                            | 99,83 | 100   |
| 96                                       | 98,83                                          | 99,83 | 100   |
| 112                                      | 98,81                                          | 100   | 100   |
| 128                                      | 97,62                                          | 98,30 | 100   |
| 192                                      | 91,67                                          | 94,39 | 87,93 |
| 256                                      | 72,79                                          | 79,25 | 86,56 |

zmodyfikowanych nagrań oraz określenie, ile razy przedmiotowy materiał był przetwarzany. Podobnie jak poprzednio wykorzystano w tym celu nagrania z bazy danych wykonane według schematu przedstawionego na rycinie 4. Fragmenty utworów poddano kompresji z wykorzystaniem sześciu przepływności, zdekodowano je do formatu bezstratnego, a następnie poddano dwa razy kompresji z przepływnością



Ryc. 4. Schemat tworzenia nagrań oraz prowadzenia porównań w badaniach polegających na wykrywaniu potrójnej kompresji w plikach MP3.



Tabela 7

**Klasyfikacja binarna nagrań poddanych pojedynczej, podwójnej i potrójnej kompresji dla różnych wartości pierwotnej przepływności. Pliki poddano drugiej i trzeciej kompresji z przepływnością 128 kbps**

| Przepływność podczas pierwszej kompresji | Porównywane pliki z różnych etapów kompresji |                           |                           |                           |
|------------------------------------------|----------------------------------------------|---------------------------|---------------------------|---------------------------|
|                                          | SVM                                          |                           | LDA                       |                           |
|                                          | Kompresja 1 – Kompresja 3                    | Kompresja 2 – Kompresja 3 | Kompresja 1 – Kompresja 3 | Kompresja 2 – Kompresja 3 |
| 64                                       | 100                                          | 83,33                     | 99,83                     | 79,59                     |
| 96                                       | 100                                          | 89,12                     | 100                       | 84,18                     |
| 112                                      | 99,83                                        | 90,48                     | 100                       | 88,78                     |
| 128                                      | 95,92                                        | 77,38                     | 99,15                     | 75,51                     |
| 192                                      | 99,49                                        | 98,13                     | 99,32                     | 98,47                     |
| 256                                      | 96,94                                        | 97,45                     | 98,64                     | 98,30                     |

128 kbps. Z utworzonych plików MP3 wyodrębniono wektory cech, po jednym dla każdego z plików. Do klasyfikacji wykorzystano liniową analizę dyskryminacyjną oraz maszynę wektorów nośnych.

W przypadku każdej z wykonanych klasyfikacji wektory cech były dzielone na dwie równe sekwencje: treningową oraz testową. Realizowana w tym przypadku klasyfikacja binarna polegała na rozróżnianiu nagrań poddanych pojedynczej i potrójnej kompresji oraz podwójnej i potrójnej kompresji, przy czym porównywano ze sobą wektory cech wyznaczone na podstawie plików o tych samych parametrach. W tabeli 7 zamieszczono wyniki wykonanych porównań (podawane w procentach) w zależności od liczby modyfikacji, jakim zapisy były poddawane. Przykładowo, w trakcie klasyfikacji za pomocą algorytmu LDA nagrań poddanych pojedynczej kompresji (128 kbps) oraz plików poddanych konwersji z przepływności 192 kbps do 128 kbps, a następnie ponownie skompresowanych (z przepływnością 128 kbps), 99,32% nagrań zostało prawidłowo przypisanych do właściwej im klasy (zob. tabela 7).

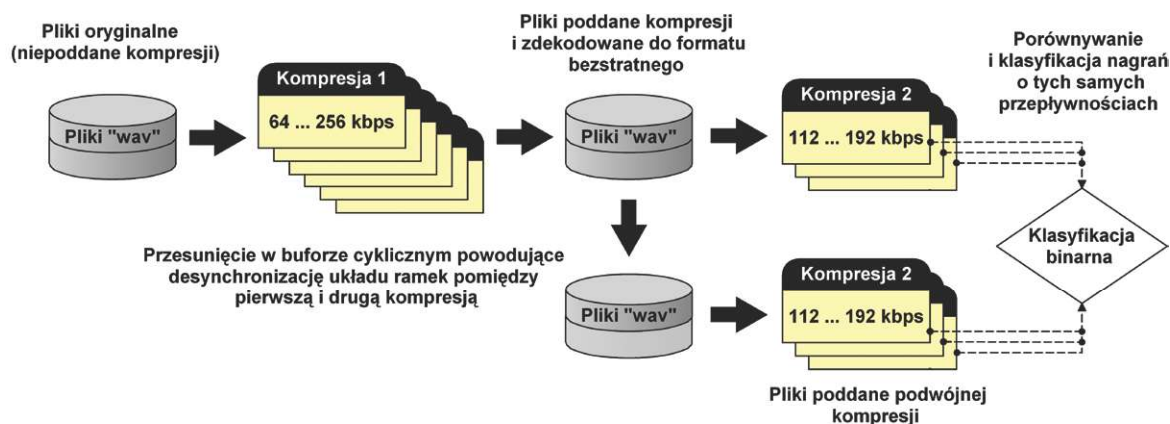
Na podstawie analizy uzyskanych wyników można zauważyć, że istnieje możliwość detekcji nagrań, które zostały poddane podwójnej oraz potrójnej kompresji. Co ciekawe, wprowadzenie dodatkowego przetwarzania powoduje, że trzykrotnie kodowane zapisy są łatwiejsze do odróżnienia od nagrań oryginalnych (poddanych pojedynczej kompresji), niż ma to miejsce w przypadku plików kodowanych dwa razy. Gorsze wyniki uzyskano w przypadku porównania plików poddanych podwójnej i potrójnej kompresji. Jednak mimo to wykonane doświadczenia pokazują, że istnieje możliwość odtworzenia rodzaju oraz kolejności modyfikacji, jakim był poddawany badany materiał. Istotne jest tylko stworzenie właściwych modeli dla algorytmów klasyfikacji.

### Wykrywanie nieciągłości w nagraniach poddanych kompresji

Edycja nagrania utrwalonego w pliku MP3 polegająca na wycięciu lub wstawieniu fragmentu wypowiedzi może zostać wykonana w dziedzinie czasu, po uprzednim zdekodowaniu zapisu do formatu bezstratnego. Jak wspomniano, w procesie kompresji dane wejściowe są porządkowane w ramki, a następnie w takiej formie przetwarzane. Jeśli zostanie wykonany montaż, to istniejący porządek ramek zostanie naruszony. Aby przywrócić pierwotny układ, należałoby wprowadzić stosowne przesunięcie, identyczne dla każdej następnej ramki, począwszy od miejsca wykonania edycji. Zagadnienie pomiaru wspomnianego przesunięcia i wykorzystania go do wykrywania ingerencji w zapis nagrania poddanego kompresji i zdekodowanego do formatu bezstratnego było już przedmiotem rozważań autora i nie będzie tu szczegółowo omawiane [3, 20].

Wykonując badania na potrzeby niniejszego opracowania, dostrzeżono możliwość wykorzystania wymienionego zjawiska do badania integralności zapisów poddanych podwójnej i – w ogólności – wielokrotnej kompresji. Naruszenie siatki ramek powstałej w procesie stratnego kodowania nagrania przed ponowną kompresją zapisu spowoduje bowiem zmianę kształtu widma wyznaczanego przez algorytm MP3, podobnie jak w przypadku zilustrowanym na rycinie 1. W związku z powyższym wszelkie operacje wykonywane przez koder, w szczególności estymacja progu maskowania oraz kwantowanie zależne od estymowanej charakterystyki maskowania, będą determinowane rozkładem prążków transformacji MDCT.

W nawiązaniu do powyższych rozważań postanowiono sprawdzić możliwość wykrywania edycji w nagraniach poddanych podwójnej kompresji, wykorzystując w tym celu pliki ze stworzonej bazy danych wykonane według schematu przedstawio-



**Ryc. 5.** Schemat tworzenia nagrań oraz prowadzenia porównań w badaniach polegających na wykrywaniu nieciągłości w plikach MP3 poddanych podwójnej kompresji.

nego na rycinie 5. Fragmenty utworów poddano kompresji z wykorzystaniem sześciu przepływności, zdekodowano je do formatu bezstratnego, a następnie ich kopie przesunięto w buforze cyklicznym o 10 milisekund. Tak przygotowane pliki (bez przesunięcia i przesunięte) poddano ponownej kompresji, stosując również trzy wybrane przepływności (112, 128 oraz 192 kbps). Do klasyfikacji wektorów cech wyodrębnionych z plików MP3 wykorzystano, podobnie jak w poprzednich przypadkach, liniową analizę dyskryminacyjną oraz maszynę wektorów nośnych.

W przypadku każdej z wykonanych klasyfikacji wektory cech były dzielone na dwie równe sekwencje – treningową oraz testową. Klasyfikacja binarna polegała na rozróżnianiu nagrań poddanych edycji (w których nastąpiło przesunięcie ramek o 10 ms) i niezmontowanych (bez przesunięcia ramek), przy czym porównywano ze sobą wektory cech wyznaczone na podstawie plików o tych samych parametrach. W tabelach 8 i 9 zamieszczono wyniki detekcji nieciągłości (podawane w procentach) w zależności od przepływności

wtórnej (odpowiednio: 112 kbps, 128 kbps oraz 192 kbps). Przykładowo, w trakcie klasyfikacji nagrań za pomocą algorytmu SVM (pierwsza kompresja wykonana z przepływnością 128 kbps) przesuniętych w buforze cyklicznym oraz niepoddanych edycji, które następnie skompresowano ponownie z przepływnością 112 kbps, 96,43% nagrań zostało prawidłowo przypisanych do właściwej im klasy (zob. tabela 8).

Podobnie jak w przypadku detekcji potrójnej kompresji sprawdzono możliwość wykrywania nieciągłości również w tych nagraniach, które po wykonaniu w nich edycji zostały skompresowane jeszcze dwa razy. Badania przeprowadzono zgodnie ze schematem zamieszczonym na rycinie 5, wykonując jednak dodatkowe dekodowanie i trzecią kompresję (z przepływnością 128 kbps). Z utworzonych plików MP3 wyodrębniono wektory cech, a do ich klasyfikacji wykorzystano również liniową analizę dyskryminacyjną oraz maszynę wektorów nośnych. Klasyfikacja binarna polegała na rozróżnianiu nagrań poddanych edycji i niezmontowanych. W tabeli 10 zamieszczono

**Tabela 8**

**Klasyfikacja binarna z wykorzystaniem algorytmu SVM oraz różnych wartości pierwotnej przepływności polegająca na wykrywaniu montażu w nagraniach poddanych podwójnej kompresji**

| Przepływność podczas pierwszej kompresji | Przepływność podczas ponownej kompresji [kbps] |       |       |
|------------------------------------------|------------------------------------------------|-------|-------|
|                                          | 112                                            | 128   | 192   |
| 64                                       | 100                                            | 100   | 100   |
| 96                                       | 99,66                                          | 100   | 100   |
| 112                                      | 98,13                                          | 100   | 100   |
| 128                                      | 96,43                                          | 96,26 | 100   |
| 192                                      | 89,46                                          | 90,31 | 92,69 |
| 256                                      | 71,26                                          | 78,74 | 88,33 |

**Tabela 9**

**Klasyfikacja binarna z wykorzystaniem algorytmu LDA oraz różnych wartości pierwotnej przepływności polegająca na wykrywaniu montażu w nagraniach poddanych podwójnej kompresji**

| Przepływność podczas pierwszej kompresji | Przepływność podczas ponownej kompresji [kbps] |       |       |
|------------------------------------------|------------------------------------------------|-------|-------|
|                                          | 112                                            | 128   | 192   |
| 64                                       | 100                                            | 99,83 | 100   |
| 96                                       | 100                                            | 100   | 99,83 |
| 112                                      | 99,32                                          | 100   | 99,66 |
| 128                                      | 97,79                                          | 98,81 | 100   |
| 192                                      | 90,82                                          | 94,22 | 92,18 |
| 256                                      | 69,73                                          | 79,76 | 83,67 |

Tabela 10

**Klasyfikacja binarna polegająca na wykrywaniu montażu w nagraniach poddanych potrójnej kompresji**

| Przepływność podczas pierwszej kompresji | Przepływność podczas ponownej kompresji [kbps] |              |
|------------------------------------------|------------------------------------------------|--------------|
|                                          | SVM                                            | LDA          |
| 64                                       | 100                                            | 99,83        |
| 96                                       | 100                                            | 100          |
| 112                                      | 99,66                                          | 100          |
| 128                                      | <b>84,01</b>                                   | <b>84,69</b> |
| 192                                      | 90,48                                          | 89,63        |
| 256                                      | 64,63                                          | 64,29        |

wyniki detekcji nieciągłości (podawane w procentach) w zależności od użytego algorytmu klasyfikacji. Przykładowo, w trakcie klasyfikacji nagrań za pomocą algorytmu SVM (pierwsza kompresja wykonana z przepływnością 112 kbps) przesuniętych w buforze cyklicznym oraz niepoddanych edycji, które następnie skompresowano ponownie dwa razy z przepływnością 128 kbps, 99,66% nagrań zostało prawidłowo przypisanych do właściwej im klasy (zob. tabela 10).

Analizując uzyskane wyniki, można zauważyć, że dzięki wykorzystaniu zjawiska polegającego na zmianie rozkładu prążków widma MDCT w przypadku naruszenia siatki ramek powstałej w procesie kodowania stratnego, istnieje możliwość skutecznego wykrywania nieciągłości w nagraniach skompresowanych. Ponadto wprowadzenie dodatkowego przetwarzania powoduje, że identyfikacja montażu jest nadal możliwa, w szczególności dla małych wartości przepływności bitowej użytej podczas pierwszego kodowania.

## Podsumowanie

Wyniki badań wykonanych w ramach projektu badawczego pokazują, że istnieje możliwość wykrywania różnego rodzaju ingerencji w zapis nagrania poddanego stratnej kompresji MP3. Bazując na dostępnych rozwiązaniach, autor zaproponował własne algorytmy służące zarówno do predykcji oryginalnej przepływności bitowej, detekcji podwójnej i potrójnej kompresji, jak i do wykrywania montażu w nagraniach poddanych wielokrotnym procesom kodowania i dekodowania. W celu sprawdzenia skuteczności stworzonych metod opracowano własną bazę danych nagrań fonicznych składającą się z blisko miliona plików MP3. Wyniki wykonanych

badan pokazują, że dzięki odpowiednio przygotowanym sekwencjom uczącym jest możliwe stworzenie skutecznego algorytmu umożliwiającego wykrywanie nieciągłości w nagraniach poddanych kompresji MP3. Ponadto, nawiązując do definicji nagrania autentycznego według standardu AES27–1996 [2], można zauważyć, że w ramach wykonanych badań dokonano nie tylko analizy oryginalności zapisów, lecz także podjęto skuteczną próbę ujawnienia śladów wykonanego montażu.

Omówione w niniejszej pracy metody badania autentyczności nagrań poddanych stratnej kompresji można wykorzystać zarówno do badania plików MP3, jak również innych algorytmów kompresji, w których użyto zespołów filtrów DCT o perfekcyjnej lub quasi-perfekcyjnej rekonstrukcji, np. AAC lub Ogg Vorbis. Warto również podkreślić, że w związku z wykorzystaniem do stworzenia wektora cech przede wszystkim małych wartości próbek widma MDCT skuteczność zaproponowanych metod nie może być zależna od kształtu obwiedni charakterystyki widmowej badanych zapisów fonicznych, co niewątpliwie ma miejsce w przypadku m.in. metod rozpoznawania mówców na podstawie głosu oraz innych rozwiązań polegających na analizie i klasyfikacji treści nagrań (w tym nagrań muzycznych). W związku z powyższym przedstawione metody badania autentyczności nagrań poddanych kompresji MP3 mogą być wykorzystane również w odniesieniu do rzeczywistych nagrań dowodowych.

Ze względu na coraz większe zainteresowanie współczesnej kryminalistyki badaniami autentyczności cyfrowych nagrań fonicznych konieczny jest ciągły rozwój oraz automatyzacja nowych metod badawczych. Dostępne rozwiązania, jak np. analiza zmian częstotliwości prądu sieci elektroenergetycznej, badanie zmian tła akustycznego za pomocą metod czasowo-częstotliwościowych, czy wyznaczanie parametrów nagrań poddanych kompresji stratnej mogą i – w miarę możliwości – powinny być wykorzystywane łącznie. Należy bowiem mieć na uwadze, że brak wykrycia miejsc montażu nie wyklucza faktu nieuprawnionej modyfikacji nagrania. Może świadczyć m.in. o zastosowaniu niezbadanej jeszcze techniki edycji zapisu lub braku dostępnych metod analitycznych.

Badania prowadzone były w ramach projektu pt.: „Projektowanie badań empirycznych i analizy materiałów dotyczących specyfiki metod kryminalistyki w pracy służb specjalnych służb porządku publicznego” finansowanego przez Narodowe Centrum Badań i Rozwoju w formie grantu nr 0023/R/ID3/2012/02.



## Źródła rycin i tabel

Ryciny 1–5: autor

Tabele 1–10: opracowanie własne; wyniki opublikowane w [37]

## Bibliografia

1. Wyrok SN z dnia 10 marca 1961 r., sygn. III K 49/61, OSNKW 1962, nr 1, poz. 8.
2. Audio Engineering Society: AES27–1996 (r2007) „Recommended Practice for Forensic Purposes – Managing Recorded Audio Materials Intended for Examination”, 1996 (2007).
3. R. Korycki: Zastosowanie metod czasowo-częstotliwościowej analizy sygnałów w wykrywaniu śladów ingerencji w ciągłość cyfrowych zapisów fonicznych, rozprawa doktorska, Instytut Radioelektroniki, Politechnika Warszawska, 2012.
4. R. Korycki: Time and Spectral Analysis Methods with Machine Learning for the Authentication of Digital Audio Recordings, „Forensic Science International” 2013, vol. 230, nr 1–3, s. 117–126.
5. I. Biernacka, R. Korycki, J. Rzeszotarski: Analiza wahań częstotliwości prądu sieciowego w badaniach autentyczności nagrań cyfrowych, „Problemy Kryminalistyki” 2007, nr 258, s. 36–40.
6. C. Grigoros: Digital Audio Recording Analysis: The Electric Network Frequency (ENF) Criterion, „The International Journal of Speech Language and the Law” 2005, vol. 12, nr 2, s. 63–76.
7. M. Michałek: Automatyzacja badania autentyczności cyfrowych nagrań dźwiękowych z wykorzystaniem kryterium ENF”, „Problemy Kryminalistyki” 2010, nr 267.
8. C. Grigoros, A.J. Cooper, M. Michałek: Best Practice Guidelines for ENF Analysis in Forensic Authentication of Digital Evidence, European Network of Forensic Science Institutes, Forensic Speech and Audio Analysis Working Group, 2009.
9. H. Farid, W. Wang: Exposing Digital Forgeries in Video by Detecting Double Quantization, in Proc. 11th ACM workshop on Multimedia and Security – Intl. Multimedia Conf., Princeton NJ, Sep. 7–8, 2009.
10. Y. Chen, C. Hsu: Image tampering detection by blocking periodicity analysis in JPEG compressed images, in Proc. 10th Workshop on Multimedia Signal Processing, Cairns, Queensland, Oct. 8–10, 2008, s. 803–808.
11. H. Farid: Exposing Digital Forgeries from JPEG Ghosts, „IEEE Transactions on Information Forensics and Security” 2009, nr 4 (1), s. 154–160.
12. H. Farid, A.C. Popescu: Exposing Digital Forgeries by Detecting Traces of Resampling, „IEEE Transactions on Signal Processing” 2005, nr 53 (2), s. 758–767.
13. H. Farid, E. Kee: Exposing digital forgeries from 3-d lightning environments, „IEEE International Workshop on Information Forensics and Security” 2010.
14. R. Geiger, J. Herre, S. Moehrs: Analysing decompressed audio with the Inverse Decoder – towards an operative algorithm, in Proc. 112th AES Convention, Munich, May 10–13, 2002.
15. J. Herre, M. Schug: Analysis of Decompressed Audio – The Inverse Decoder, in Proc. 109th AES Convention, Los Angeles, USA, Sept. 22–25, 2000.
16. B. D'Alessandro, Y. Q. Shi: MP3 Bit Rate Quality Detection through Frequency Spectrum Analysis, in Proc. 11th ACM workshop on Multimedia and Security – Intl. Multimedia Conf., Princeton NJ, Sep. 7–8, 2009.
17. J. Huang, Y.Q. Shi, R. Yang: Defeating Fake-Quality MP3, [w:] Proc. 11th ACM workshop on Multimedia and Security – Intl. Multimedia Conf., Princeton NJ, Sep. 7–8, 2009.
18. Q. Liu, M. Qiao, A. H. Sung: Detection of double MP3 compression, „Cognitive Computing” 2010, vol. 2, nr 4, s. 291–296.
19. Q. Liu, M. Qiao, A. H. Sung: Revealing real quality of double compressed MP3 audio, [w:] Proc. ACM Workshop on Multimedia and Security – Intl. Multimedia Conf., Firenze, Italy 2010.
20. R. Korycki: Detection of tampering in lossy compressed digital audio recordings, [w:] Proc. of the IEEE and AES Joint Conference – New Trends in Audio and Video & Signal Processing: Algorithms, Architectures, Arrangements, and Applications, Łódź, Poland, Sep. 27–29, 2012, s. 97–102.
21. MPEG, Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s, part 3: Audio, International Standard IS 11172–3, ISO/IEC JTC1/SC29 WG11, 1992.
22. T. Zieliński: Cyfrowe przetwarzanie sygnałów – od teorii do zastosowań, Wydawnictwa Komunikacji i Łączności sp. z o.o., Warszawa 2005, 2007.
23. M. Vaananen, M. Vilermo, Y. Wang, L. Yaroslavsky: Some Peculiar Properties of the MDCT, [w:] Proc. of 16th IFIP World Computer Congress, Beijing, China, 2000, s. 61–64.
24. V. Britanak: A survey of efficient MDCT implementations in MP3 audio coding standard: Retrospective and state-of-the-art, „Signal Processing” 2011, nr 91, s. 624–672.
25. A.B. Bradley, A.W. Johnson, J.P. Princen: Subband/transform coding using filter bank designs based on time domain aliasing cancellation, „IEEE Proc. Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)” 1987, nr 12, s. 2161–2164.
26. J. Huang, Z. Qu, and R. Yang: Detecting digital audio forgeries by checking frame offsets”, [w:] Proc. of the 10th ACM workshop on Multimedia and security, International Multimedia Conference, Oxford 2008, s. 21–26.
27. N.J. Nilsson: Introduction to machine learning, Artificial Intelligence Laboratory, Department of Computer Science, Stanford University 2010.
28. J.C. Principe: Artificial Neural Networks, The Electrical Engineering Handbook, CRC Press LLC, 2000.
29. Y.H. Hu, J.N. Hwa: Handbook of Neural Network Signal Processing, CRC Press 2002.
30. A. Cichocki, R. Unbehauen: Neural Networks for Optimization and Signal Processing, Wiley, New York 1993.

31. P. Janik, T. Lobos: Automated Classification of Power Quality Disturbances using SVM and RBF Networks, „IEEE Transaction on Power Delivery” 2006, vol. 21, nr 3, s. 1663–1669.
32. P. Walendowski: Zastosowanie sieci neuronowych typu SVM do rozpoznawania mowy, rozprawa doktorska, Politechnika Wrocławska, 2008.
33. S. Ji, J. Ye: Discriminant Analysis for Dimensionality Reduction: An Overview of Recent Developments, [w:] Biometrics: Theory, Methods, and Applications, N.V. Boulgouris, K. N. Plataniotis, E. Micheli-Tzanakou, [red.], John Wiley & Sons, Inc., Hoboken, NJ, USA, 2009, ch. 1, s. 1–17.
34. J. Ćwik, J. Koronacki: Statystyczne systemy uczące się, Wydawnictwo Exit, Warszawa 2008, rozdział 3.
35. J. Ćwik, J. Mielniczuk: Statystyczne systemy uczące się – ćwiczenia w oparciu o pakiet R, Oficyna Wydawnicza PW, Warszawa 2009.
36. Mathworks R2013a Documentation, Statistics Toolbox and Bioinformatics Toolbox, Available: <http://www.mathworks.com/help/documentation-center.html>
37. R. Korycki: Authenticity examination of compressed audio recordings using detection of multiple compression and encoders' identification, „Forensic Science International” 2014, vol. 238, nr 1–3, s. 33–46.