

Rafał Korycki

Analiza autentyczności cyfrowych nagrań fonicznych

Wstęp i cel pracy

Badania autentyczności nagrań fonicznych, które polegają na wykrywaniu śladów ingerencji w ciągłość zapisu, należą do ważnych zadań współczesnej kryminalistyki. W przypadku nagrań analogowych istnieje możliwość analizy przebiegu czasowego i spektrogramu sygnału fonicznego, a także zapisu pola magnetycznego utrwalonego na samej taśmie. Pojawienie się nagrań cyfrowych¹ znacznie skomplikowało, a w wielu przypadkach uniemożliwiło ujawnianie śladów wykonanego montażu. Przełomem w rozwiązywaniu tego problemu okazało się zastosowanie analizy wahań częstotliwości prądu sieci elektroenergetycznej. Wykorzystano w tym celu zarówno zaobserwowaną dużą korelację zmian częstotliwości prądu (ENF – *Electric Network Frequency*) w różnych obszarach podłączonych do tej samej sieci, jak również fakt rejestrowania sygnału sieci energetycznej przez niektóre urządzenia nagrywające. Problemem pozostaje wyodrębnienie sygnału sieci energetycznej z nagrania dowodowego, pomiar chwilowych wartości częstotliwości oraz porównanie sekwencji tych wartości z wartościami pochodzącymi z bazy danych. Pomiaru można dokonywać zarówno korzystając z metod opartych na wyznaczaniu transformacji Fouriera, jak i innych rozwiązań, takich jak modelowanie parametryczne, metody podprzestrzeni czy wyznaczanie częstotliwości na podstawie pomiaru czasu w miejscach, w których sygnał przechodzi przez zero.

Celem pracy jest prezentacja metod wykorzystywanych do wykrywania ingerencji w ciągłość zapisów. Oprócz nowych technik pomiaru częstotliwości, których zastosowanie jest propozycją autora mającą na celu usprawnienie badań z użyciem sygnału sieci energetycznej, zaprezentowano również możliwość graficznego wykrywania śladów montażu za pomocą analiz czasowo-częstotliwościowych. Przedstawiono przykłady wykorzystania graficznej analizy widmowej na podstawie pochodnych transformacji Wignera-Ville'a oraz inne przekształcenia. Opisano także metodę przemieszczenia (*reassignment method*) pozwalającą na zwiększenie czytelności wykresów będących wynikiem zastosowanych transformacji. Dzięki wykorzystaniu powyższych rozwiązań możliwe jest obserwowanie trudno dostrzegalnych zmian tła akustycznego, które mogą świadczyć o wykonanym montażu. Trzecią przedstawioną metodą jest analiza przesunięcia ramek w pli-

kach dźwiękowych skompresowanych z wykorzystaniem stratnych koderów psychoakustycznych. Wpływ parametrów algorytmu kompresji oraz kształtu użytego okna opisano na przykładzie kodera MP3. Ostatnie prezentowane zagadnienie to techniki analizy polispektralnej. Omówiono możliwość wykorzystania powyższych narzędzi w detekcji śladów ingerencji w ciągłość zapisów oraz ich efektywność w analizowaniu sygnału mowy.

Zasady stosowania metody analizy zmian częstotliwości prądu sieci energetycznej

Prąd przemienny płynący w instalacji elektrycznej oraz w urządzeniach podłączonych do niej generuje zmienne pole elektromagnetyczne. Częstotliwość zmian pola wynosi 50 Hz lub 60 Hz i zależy od systemu energetycznego obejmującego dany obszar. Polski system energetyczny jest częścią zgrupowania systemów przesyłowych Europy oraz północnej Afryki, które pracują ze sobą synchronicznie, to znaczy z tą samą częstotliwością prądu w sieci wynoszącą 50 Hz. Operatorzy tych systemów zrzeszeni są w Unii ds. Koordynacji Przesyłu Energii Elektrycznej (UCTE – *Union for the Coordination of Transmission of Electricity*) [1]. W Polsce operatorem systemu przesyłowego jest PSE-Operator S.A.

Podobnie jak przemienny prąd płynący w przewodniku generuje zmienne pole elektromagnetyczne, tak zmienne pole elektromagnetyczne powoduje indukowanie przemiennego prądu w przewodniku znajdującym się w zasięgu tego pola. Ze względu na sinusoidalny charakter zmian amplitudy prądu sieci energetycznej oraz częstotliwość (50 Hz) pokrywającą się z pasmem akustycznym (20–20 000 Hz), istnieje możliwość indukcji prądu o tej częstotliwości w elementach elektronicznych bloków wzmacniaczy sygnałów akustycznych znajdujących się w urządzeniach rejestrujących.

Odchylenie częstotliwości prądu sieci energetycznej od wartości zadanej powodowane jest różnicą między obciążeniem systemu energetycznego a wytwarzaniem energii elektrycznej w elektrowniach. Energia w systemie nie może być magazynowana, wobec tego zapotrzebowanie musi być pokryte w czasie rzeczywistym energią wyprodukowaną w elektrowniach [1]. Każde niezbilansowanie wytwarzania z zapotrzebowaniem

powoduje odchylenie częstotliwości od wartości zadanej. Częstotliwość rośnie, jeśli w systemie jest nadmiar mocy wyprodukowanej, lub maleje, jeśli zapotrzebowanie przewyższa produkcję [2]. Przeprowadzone dotychczas badania [3], [4] pokazują, że powyższe odchylenia mają charakter przypadkowy, a ich wartości wykazują silną korelację w obrębie całego obszaru synchronicznego. W związku z tym wartości częstotliwości prądu sieci energetycznej zmierzone w różnych lokalizacjach na terenie tego samego obszaru synchronicznego (np. w Brukseli, Warszawie czy Bukareszcie) będą niemal identyczne.

Zastosowanie metody sprowadza się do ekstrakcji z nagrania dowodowego sygnału indukowanego z sieci energetycznej oraz precyzyjnego pomiaru wartości częstotliwości zbliżonej do 50 Hz. Uzyskane wartości porównuje się dla każdej chwili czasowej np. $t = 1$ sek. z bazą danych – własną lub pochodzącą od operatora systemu energetycznego. Pełna synchronizacja, czyli zgodność sekwencji zmian wartości częstotliwości z bazy danych z wartościami uzyskanymi z materiału dowodowego, pozwala na wsparcie hipotezy potwierdzającej autentyczność nagrania. Natomiast możliwość synchronizacji jedynie fragmentów daje podstawy do zakwestionowania integralności zapisu i podejrzenia wykonania montażu.

Zewnętrzne pole elektromagnetyczne traktowane jest przez producentów rejestratorów dźwięku jako potencjalne źródło zakłóceń. W związku z tym urządzenia konstruowane są w taki sposób, aby wyeliminować, lub przynajmniej ograniczyć do minimum, możliwość nagrywania zakłóceń. Ponadto w wielu przypadkach stosowane są filtry cyfrowe odcinające sygnały o częstotliwościach poniżej kilkuset Hz, co uniemożliwia rejestrację sygnału pochodzącego z sieci energetycznej. W przypadku każdego badania konieczne jest zatem indywidualne sprawdzenie parametrów rejestratora i zastosowanego w nim algorytmu zapisu oraz kompresji sygnału mowy w celu oszacowania ich wpływu na możliwość rejestracji sygnału sieci energetycznej.

Możliwość rejestracji sygnału sieci energetycznej w nagraniu zależy między innymi od następujących czynników:

- rodzaju zasilania urządzenia rejestrującego (w większości przypadków zasilanie sieciowe zwiększa prawdopodobieństwo rejestracji sygnału sieci energetycznej);
- rodzaju algorytmu kompresji zastosowanego w urządzeniu podczas zapisu nagrania do pliku;
- filtrów górnoprzepustowych oraz układów redukcji szumu i zakłóceń;
- rodzaju użytego mikrofonu (a także długości i rodzaju połączenia przewodowego mikrofonu z układami rejestratora) oraz struktury bloków wzmacnia-

czy analogowych występujących przed układami przetworników analogowo-cyfrowych;

- lokalizacji rejestratora, w szczególności odległości od urządzeń zasilanych z sieci energetycznej lub odległości od sieci strukturalnej lub linii wysokiego napięcia.

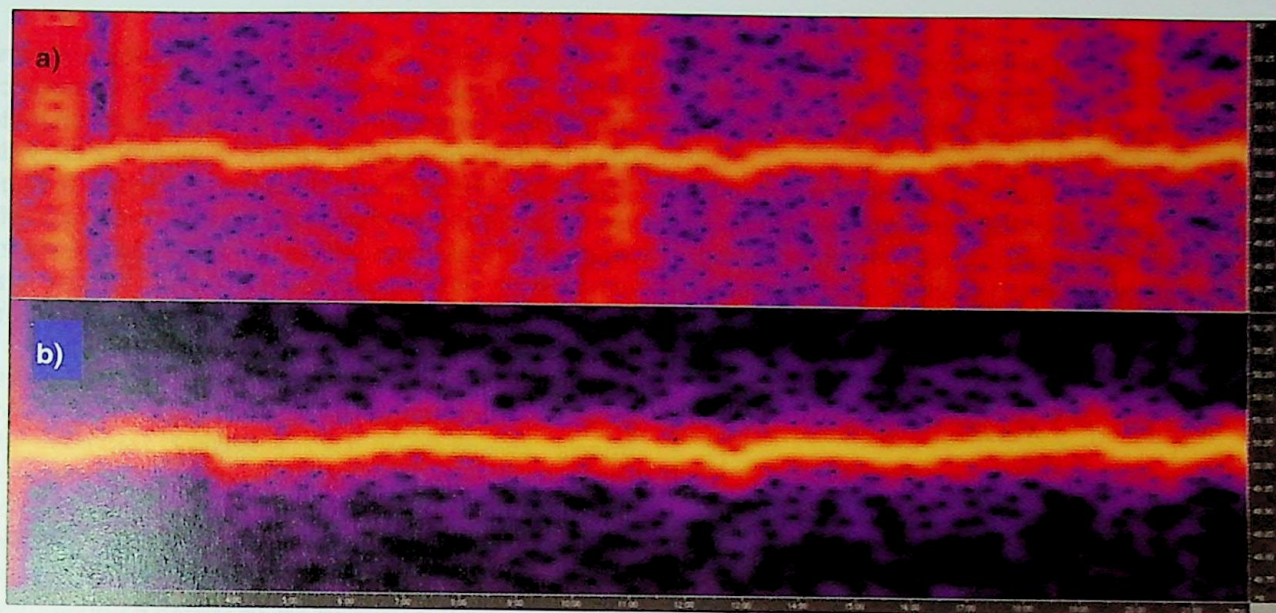
Metody nie powinno się stosować odnośnie do nagrań zapisanych w sposób analogowy na taśmie magnetycznej. Wiąże się to z fluktuacjami prędkości przesuwu taśmy, powodującymi przesunięcia częstotliwości mierzonego sygnału o wartości niekiedy znacznie przekraczające zakres wahań częstotliwości prądu sieci energetycznej. Typowa wartość nierównomierności przesuwu taśmy dla profesjonalnych magnetofonów kasetowych (np. Technics RS-B965, Pioneer CT-W806DR) wynosi ok. 0,03–0,09%, co przy założeniu liniowego wzrostu częstotliwości w stosunku do zwiększenia prędkości odtwarzania daje bezwzględny błąd pomiaru na poziomie ok. 15–45 mHz. Należy nadmienić, że typowe wartości odchylenia częstotliwości prądu sieci energetycznej od częstotliwości znamionowej rejestrowane są na poziomie ± 100 mHz, natomiast przenośne rejestratory charakteryzują się współczynnikiem nierównomierności przesuwu taśmy znacznie przekraczającym wartości dla profesjonalnych odtwarzaczy kasetowych. Mogą one dodatkowo ulec zwiększeniu na skutek rozładowywania się baterii zasilających urządzenie rejestrujące.

Pomiar wartości częstotliwości prądu sieci energetycznej

Porównanie spektrogramów to najszybsza metoda pozwalająca na sprawdzenie występowania sygnału sieci energetycznej w nagraniu dowodowym. Wykonuje się ją, wykorzystując dowolne oprogramowanie pozwalające na wyznaczenie spektrogramu, czyli krótko-czasowej transformacji Fouriera (STFT – *Short-time Fourier Transform*). Po wczytaniu pliku z materiałem dowodowym wykonywane są następujące operacje:

- konwersja częstotliwości próbkowania nagrania dowodowego;
- wyznaczenie spektrogramu na podstawie STFT z długim oknem analizy ($N \geq 4096$ próbek);
- analiza sygnałów sinusoidalnych w zakresie $f_{\text{sieci energetycznej}}$ i porównanie z odpowiednio skalowanym wykresem zmian częstotliwości prądu sieci energetycznej pochodzącym z bazy danych, np. bazy operatora systemu energetycznego lub z materiałem porównawczym (ryc. 1 i 2).

Zgodność wzorców zmian wartości częstotliwości na analizowanych wykresach (ryc. 1) pozwala przypuszczać, że nagranie jest autentyczne i nie występują w nim ślady ingerencji w jego ciągłość. Rozbieżności



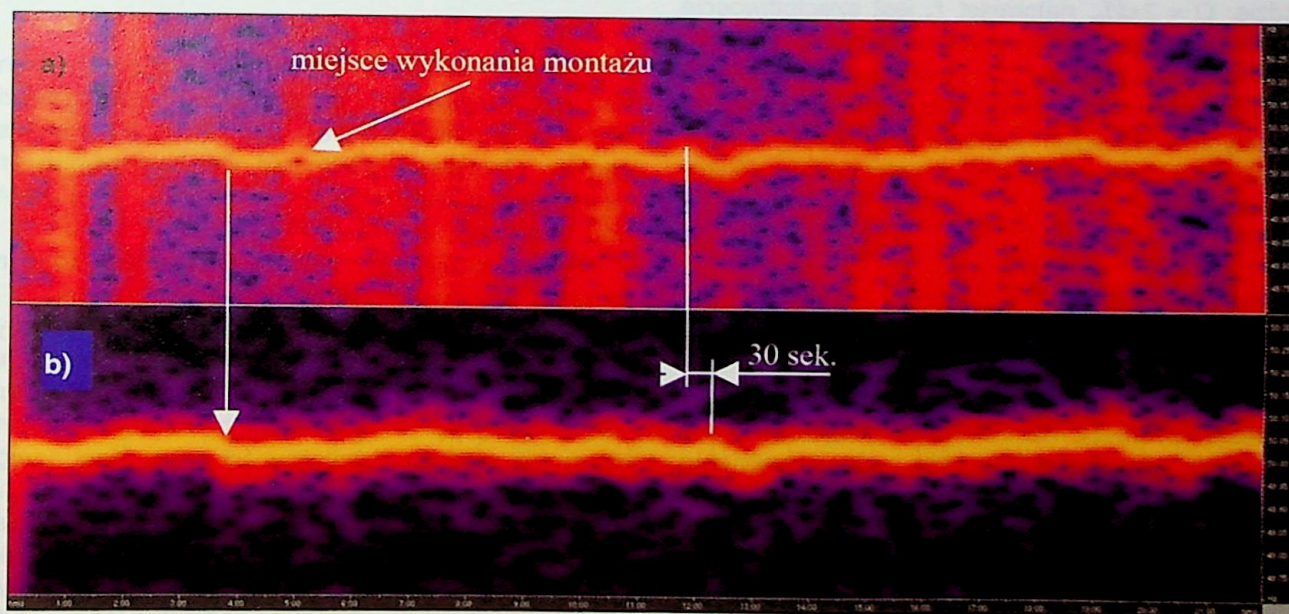
Ryc. 1. Porównanie spektrogramów:

- a) sygnału sieci energetycznej wydrebnionego z materiału dowodowego;
 b) nagrania referencyjnego zarejestrowanego bezpośrednio z sieci energetycznej.
 Parametry analizy: okno Blackmanna o długości 8192 próbek, 50% nakładanie ramek.

Fig. 1. Spectrograms comparison

- a) electric network frequency signal extracted from evidential recording;
 b) reference signal recorded directly from electric network.

Analysis parameters: Blackmann window, 8192 samples long, 50% overlapping.



Ryc. 2. Porównanie spektrogramów:

- a) sygnału sieci energetycznej wydrebnionego z materiału dowodowego poddanego montażowi;
 b) nagrania referencyjnego zarejestrowanego bezpośrednio z sieci energetycznej.
 Parametry analizy: okno Blackmanna o długości 8192 próbek, 50% nakładanie ramek.

Z materiału dowodowego we wskazanym miejscu wycięto 30 sekund nagrania. Widoczny jest brak synchronizacji zmian częstotliwości sygnału sieci energetycznej od miejsca, w którym wykonano montaż.

Fig. 2. Spectrograms comparison

- a) electric network frequency signal extracted from tampered evidential recording;
 b) reference signal recorded directly from electric network.

Analysis parameters: Blackmann window, 8192 samples long, 50% overlapping.

A 30 second time frame was removed from the recording as indicated. Lack of synchronization is visible beginning with the area where tampering occurred.

będące wynikiem miejscowej desynchronizacji przebiegów wartości zmian częstotliwości sygnału sieci energetycznej stanowią o braku ciągłości nagrania dowodowego (ryc. 2).

Transformacja Fouriera jest podstawowym narzędziem analizy częstotliwościowej sygnałów. Widmo Fouriera dla sygnałów ciągłych (np. sygnałów analogowych) zawierające informację o zawartości częstotliwościowej sygnału $x(t)$ opisywane jest zależnością:

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft} dt \quad (1)$$

Innymi słowy, wartość $X(f)$ niesie informację o tym, ile znajduje się w sygnale zespolonej² składowej harmonicznej $e^{j2\pi ft} = \cos(2\pi ft) + j\sin(2\pi ft)$ o konkretnej częstotliwości f . Natomiast widmo Fouriera dla sygnałów dyskretnych, odpowiednie do analizy sygnałów cyfrowych, przedstawia wzór:

$$X(e^{j\Omega}) = \sum_{n=-\infty}^{N-1} x(n) e^{-j\Omega n} \quad (2)$$

gdzie: $\Omega = 2\pi f/f_p$, natomiast f_p jest częstotliwością próbkowania. Baza fourierowska ma swoje rozwinięcie: $e^{j\Omega n} = \cos(\Omega n) + j\sin(\Omega n)$ [5]. Ze względu na konieczność odseparowania sygnału sieci energetycznej od pozostałych składowych nagrania dowodowego, zalecane jest wykonanie konwersji częstotliwości próbkowania do wartości:

$$f_{ds} = 2 \cdot f_{sieci\ energetycznej} + 20\% \quad (3)$$

a następnie przeprowadzenie filtracji pasmowej w zakresie np. 48–52 Hz. Konieczne jest także „nadpróbkowanie” transformacji Fouriera, czyli uzupełnienie zerami analizowanego sygnału. Jest to spowodowane tym, że dla $N = 240$ próbek rozdzielczość transformacji wynosić będzie tylko $\Delta f = (60:240 \text{ Hz}) = 250 \text{ mHz}$, co jest znacznie poniżej wymagań dla opisywanej metody. Liczba próbek transformacji Fouriera powinna wynosić teoretycznie przynajmniej:

$$N_{int} = \frac{f_{ds}}{2 \cdot \Delta f} = \frac{120 \text{ Hz}}{2 \cdot 1 \text{ mHz}} = 60000 \quad (4)$$

w przypadku konieczności zapewnienia rozdzielczości na poziomie 1 mHz.

Na rycinie 3 przedstawione zostało porównanie zmian wartości częstotliwości prądu sieci energetycznej, które wyznaczone zostały z wykorzystaniem metody FFT z uzupełnieniem zerami, z danymi z bazy danych PSE-Operator S.A. Nagranie dowodowe o długości 23 minut zostało zarejestrowane za pomocą zasilanego bateryjnie magnetofonu DAT z podłączonym zewnętrznym mikrofonem miniaturowym. Średni błąd bezwzględny wyznaczony na podstawie porównania danych wyniósł $\Delta_{FFT} = 1,2 \text{ mHz}$.

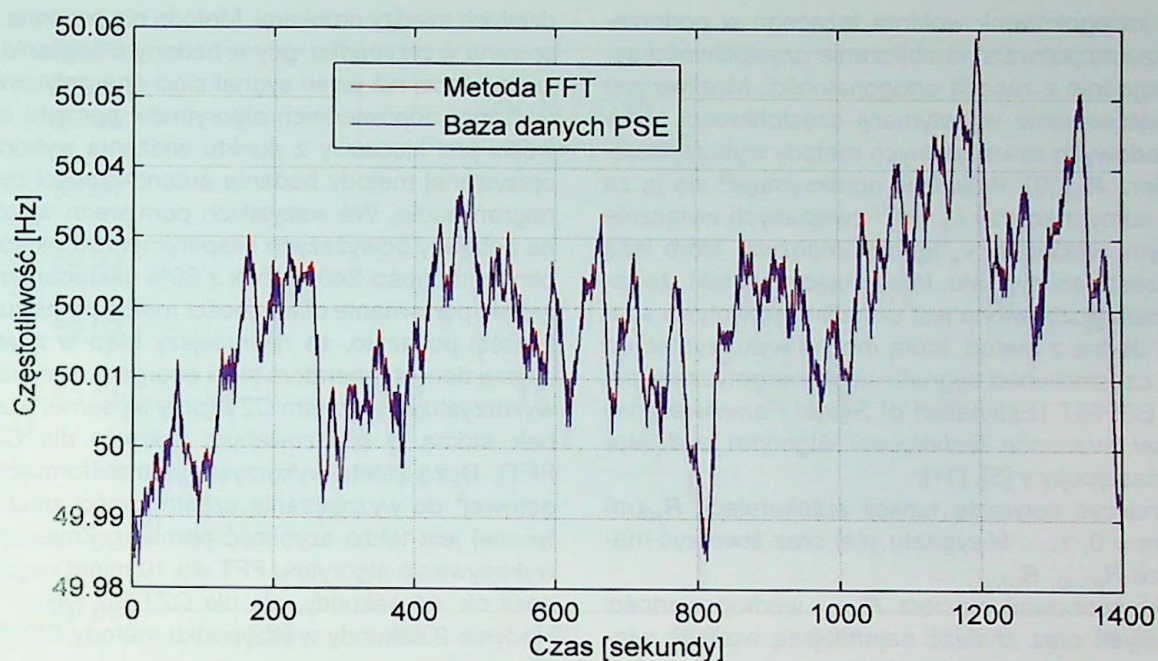
W celu wyznaczenia częstotliwości sygnału sieci energetycznej można zastosować dyskretną transformację „świergotową” (CZT – *Chirp-Z Transform*) [5]. Wykorzystuje się ją do obliczania widma Fouriera w zadanym paśmie częstotliwości z dowolną rozdzielczością zdefiniowaną przez użytkownika [6]. Innymi słowy, stanowi ona „lupę” w dziedzinie częstotliwości [7], [8]. Widmo sygnału wyznacza się za pomocą transformacji Fouriera sygnału $x(n)$ o długości N , dla $M + 1$ częstotliwości unormowanych: $f_k = f_0 + k \cdot \Delta f$, gdzie $k = 0, \dots, M$, wówczas:

$$X(k) = W^{k^2} \sum_{n=0}^{N-1} [x(n) A^n W^{n^2}] W^{-(k-n)^2} \quad (5)$$

Powyższa zależność prezentuje spłot dwóch sygnałów: $y_1(n) = x(n) A^n W^{n^2}$ oraz $y_2(n) = W^{-n^2}$, czemu w dziedzinie częstotliwości odpowiada iloczyn widm tych sygnałów. Opisywana zależność jest realizowana przy następujących założeniach: $A \equiv e^{-j2\pi f_0}$, $W \equiv e^{-j2\pi \Delta f/2}$ oraz $2kn = n^2 + k^2 - (k-n)^2$ [5].

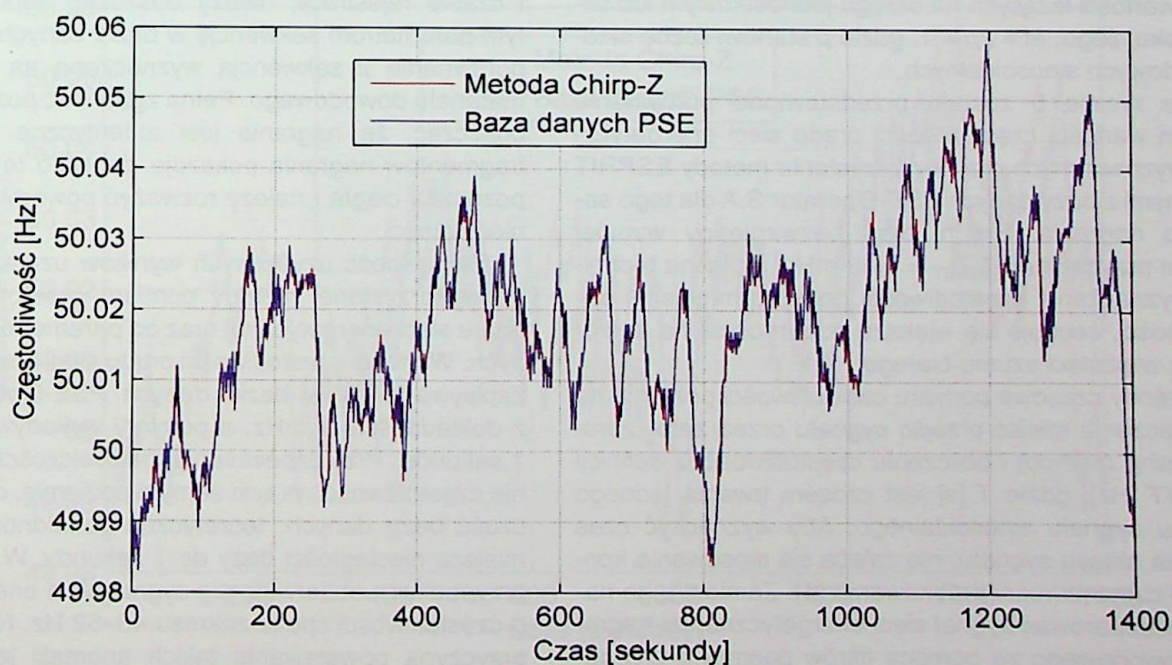
Wykorzystanie transformacji CZT teoretycznie nie wymaga stosowania filtracji pasmowej. Ponadto ze względu na wykonywanie obliczeń jedynie w zadanym paśmie algorytm wykorzystujący transformację jest szybszy niż stosujący „nadpróbkowaną” transformację Fouriera. Na rycinie 4 przedstawione zostało porównanie zmian wartości częstotliwości prądu sieci energetycznej wyznaczonych z wykorzystaniem metody CZT z danymi z bazy danych PSE-Operator S.A dla tego samego nagrania dowodowego o długości 23 minut. Średni błąd bezwzględny wyznaczony na podstawie porównania danych wyniósł $\Delta_{CZT} = 0,81 \text{ mHz}$.

Widmo można także wyznaczyć za pomocą estymaty³ funkcji autokorelacji $R_{xx}(m)$ analizowanego sygnału. Na podstawie wartości próbek estymaty tworzy się macierz autokorelacji R_{xx} , następnie obliczane są wartości własne λ_k i wektory własne v_k , a macierz R_{xx} przedstawiana jest w postaci sumy macierzy elementarnych $v_k v_k^T$ z wagami λ_k . Wektory własne są ortogonalne⁴ i tworzą dwie dopełniające się przestrzenie: sygnału i szumu [5]. Wyznaczenie na podstawie R_{xx} badanego



Ryc. 3. Porównanie wartości częstotliwości prądu sieci energetycznej wyznaczonych za pomocą transformacji FFT (z uzupełnieniem zerami) z wartościami z bazy danych PSE-Operator S.A.

Fig. 3. Electric network frequency values computed using FFT transform and compared with Polish electric system operator (PSE-Operator S.A.) database



Ryc. 4. Porównanie wartości częstotliwości prądu sieci energetycznej wyznaczonych za pomocą transformacji Chirp-Z z wartościami z bazy danych PSE-Operator S.A.

Fig. 4. Electric network frequency values computed using Chirp-Z transform and compared with Polish electric system operator (PSE-Operator S.A.) database

sygnału jakiegokolwiek wektora leżącego w podprzestrzeni szumu pozwala na obliczenie częstotliwości sygnałów zgodnie z zasadą ortogonalności. Możliwe jest także zastosowanie w estymacji częstotliwości sumy kilku składowych sinusoidalnych metody wykorzystującej macierz R_{xx} [9]. Wówczas aproksymuje⁵ się ją za pomocą sumy macierzy $\lambda_k v_k v_k^T$ związanych wyłącznie z głównymi wektorami v_k tej dekompozycji, które leżą w podprzestrzeni sygnału. Interesujący jest fakt, że taka estymata pozbawiona jest szkodliwego wpływu szumu [10]. Jedną z metod, którą można wykorzystać do pomiaru częstotliwości sygnału sieci energetycznej jest metoda ESPRIT (*Estimation of Signal Parameters via Rotational Invariance Techniques*). Algorytm jej działania jest następujący [5], [11]:

- wyznaczyć estymatę funkcji autokorelacji $R_{y0}(m)$ dla $m = 0, 1, \dots, M$ sygnału $y(n)$ oraz stworzyć macierze $R_{y0,y0}$, $R_{y0,y1}$;
- zdekomponować macierz $R_{y0,y0}$ według wartości własnych oraz znaleźć najmniejszą wartość własną, którą przyjmuje się za estymatę wariancji szumu σ_s^2 ;
- utworzyć macierze: $C_{y0,y0} = R_{y0,y0} - \sigma_s^2 I$ oraz $C_{y0,y1} = R_{y0,y1} - \sigma_s^2 J$, gdzie $I = \text{eye}(M)$, J – przesunięta macierz I z zerami w pierwszym wierszu;
- wyznaczyć uogólnione wartości własne dla macierzy $C_{y0,y0}$ oraz $C_{y0,y1}$: $D = \text{eig}(C_{y0,y0}, C_{y0,y1})$;
- znaleźć częstotliwości składowe f_k na podstawie p wartości leżących na okręgu jednostkowym lub blisko niego; $M = 2p + 1$, gdzie p stanowi liczbę składowych sinusoidalnych.

Na rycinie 5 zostało przedstawione porównanie zmian wartości częstotliwości prądu sieci energetycznej wyznaczonych przy wykorzystaniu metody ESPRIT z danymi z bazy danych PSE-Operator S.A dla tego samego nagrania. Średni błąd bezwzględny wyniósł w tym przypadku $\Delta_{\text{ESPRIT}} = 1,32$ mHz. Opisana technika wyznaczania częstotliwości pomimo mniejszej dokładności, cechuje się większą odpornością na zakłócenia w postaci szumu białego.

Metody czasowe pomiaru częstotliwości polegają na wyznaczeniu miejsc przejść sygnału przez zero (*zero-crossing method*) i obliczeniu częstotliwości z definicji $f = 1/T$ [Hz], gdzie T [s] jest czasem trwania jednego okresu sygnału sinusoidalnego. Aby wyznaczyć czas trwania okresu sygnału, nie zaleca się stosowania konwersji częstotliwości próbkowania [4]. Zamiast tego należy odseparować sygnał sieci energetycznej od nagrania dowodowego za pomocą filtrów pasmowo przepustowych wysokiego rzędu⁶. Można także stosować funkcje interpolacji⁷, aby zwiększyć rozdzielczość wynikającą ze skwantowanych poziomów próbkowania sygnału. Możliwe jest także stosowanie krzywych aproksymacji i wyznaczanie miejsc przejść przez zero w prze-

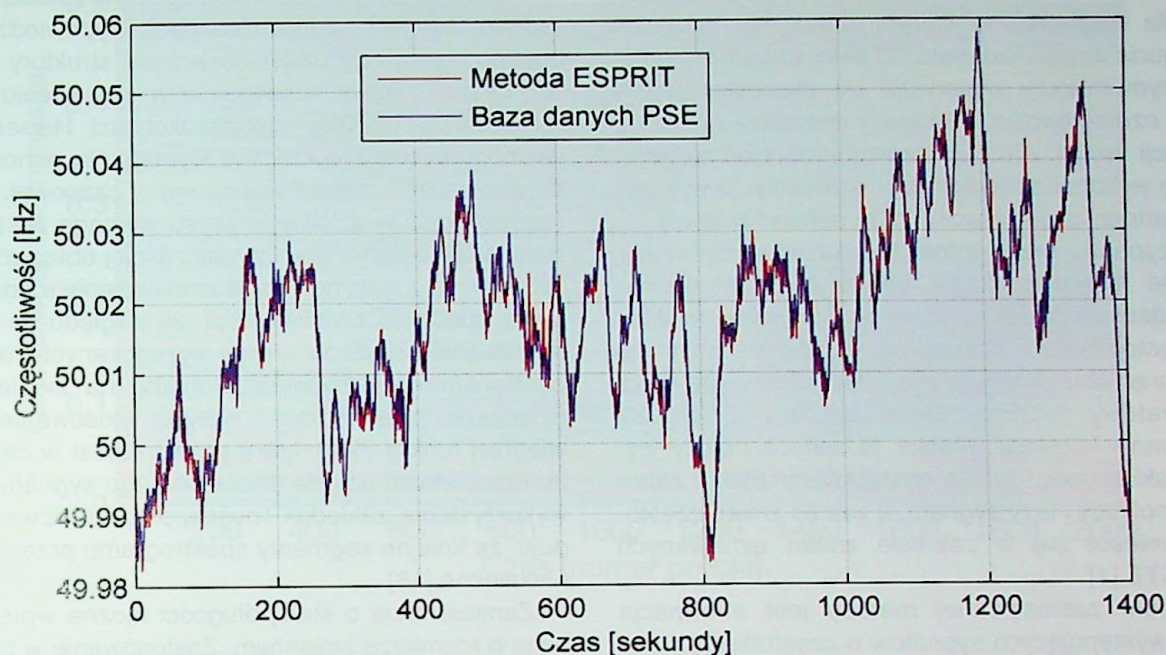
działach między próbkami. Metoda nie powinna być stosowana w przypadku, gdy w badanym nagraniu znajduje się więcej niż jeden sygnał sieci energetycznej.

Dobór odpowiednich algorytmów pomiaru częstotliwości jest kluczowy z punktu widzenia wykorzystania opisywanej metody badania autentyczności cyfrowych nagrań audio. We wszystkich pomiarach wykonanych na potrzeby powyższych eksperymentów zastosowano okno o długości 240 próbek z 50% nakładaniem. Praktyczne porównanie dokładności metod pomiaru częstotliwości pokazało, że najmniejszy błąd w zestawieniu z bazą danych operatora sieci energetycznej uzyskano, wykorzystując algorytm CZT (przy tej samej liczbie próbek widma w analizowanym zakresie dla CZT oraz FFT). Dużą zaletą wykorzystania transformacji „świergotowej” do wyznaczania częstotliwości sieci energetycznej jest także szybkość pomiaru. Zmierzony czas wykonywania algorytmu FFT dla 10 minut nagrania wynosił ok. 44 sekundy, ale dla CZT już tylko 7 sekund i jedynie 2 sekundy w przypadku metody ESPRIT.

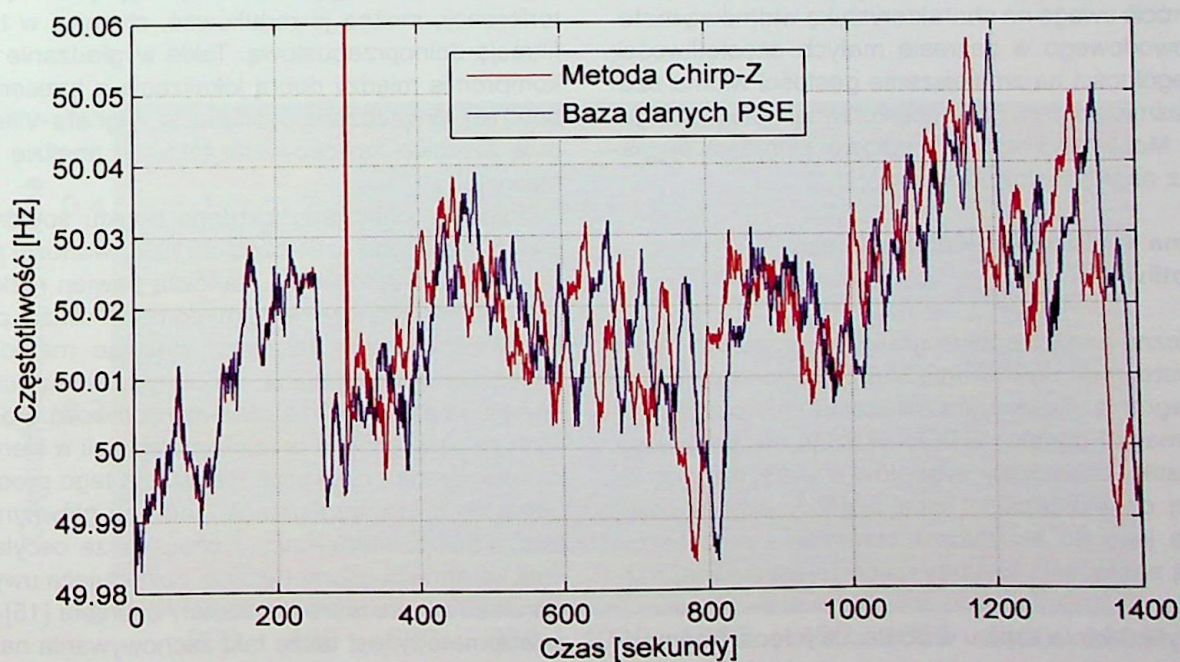
Interpretacja wyników badań

Po wykonaniu pomiaru częstotliwości prądu sieci energetycznej z materiału dowodowego należy porównać sekwencję zmian wartości częstotliwości z bazą danych – własną lub pochodzącą od operatora systemu energetycznego. Dysponując informacją o dacie i czasie rejestracji, należy odszukać odpowiadającą tym parametrom sekwencję w bazie danych i dokonać porównania z sekwencją wyznaczoną na podstawie materiału dowodowego. Pełna zgodność pozwala przypuszczać, że nagranie jest autentyczne. Zgodność fragmentów nagrania pokazuje, że tylko te fragmenty pozostają ciągłe i należy rozważyć powody powstania nieciągłości.

Niepewność uzyskanych wyników uzależniona jest od wykorzystanej metody pomiaru częstotliwości sygnału sieci energetycznej oraz od parametrów bazy danych. Wartości częstotliwości prądu sieci energetycznej zapisywane są w bazie danych PSE-Operator S.A. z dokładnością 1 mHz, a pomiary wykonywane są co 1 sekundę. Przy zapewnieniu rozdzielczości w dziedzinie częstotliwości na tym samym poziomie, co rozdzielczość bazy danych, teoretyczna dokładność detekcji miejsca nieciągłości dąży do 1 sekundy. W niektórych przypadkach obserwuje się sygnał sieci energetycznej o częstotliwości spoza zakresu 48–52 Hz. Najczęstszą przyczyną powstawania takich anomalii jest błędnie funkcjonujący układ zegarowy w urządzeniu rejestrującym, który ma wpływ na niewłaściwą wartość częstotliwości próbkowania. Stosowne poprawki można wówczas wprowadzić na etapie pomiaru i porównania wartości częstotliwości prądu sieci energetycznej.



Ryc. 5. Porównanie wartości częstotliwości prądu sieci energetycznej wyznaczonych za pomocą metody ESPRIT z wartościami z bazy danych PSE-Operator S.A.
 Fig. 5. Electric network frequency values computed using ESPRIT method and compared with Polish electric system operator (PSE-Operator S.A.) database



Ryc. 6. Porównanie wartości częstotliwości prądu sieci energetycznej wyznaczonych za pomocą transformacji Chirp-Z z wartościami z bazy danych PSE-Operator S.A. Z materiału dowodowego w 300 sekundzie (5 minucie) zapisu dokonano wycięcia 30 sekund nagrania. Widoczny jest brak synchronizacji zmian częstotliwości sygnału sieci energetycznej od miejsca, w którym wykonano montaż.

Fig. 6. Electric network frequency values computed using Chirp-Z transform and compared with Polish electric system operator (PSE-Operator S.A.) database. A 30 second time frame was removed in the 5th minute of the recording. Lack of synchronization is visible beginning with the area where tampering occurred.

Na rycinie 6 zilustrowano zmiany wartości częstotliwości dla nagrania, w którym dokonano montażu. W 5 minucie zapisu usunięto 30 sekund rozmowy. We wskazanym miejscu obserwuje się skokową zmianę wartości częstotliwości. Widoczny jest także brak synchronizacji zmian wartości częstotliwości od miejsca, w którym wykonano montaż (tzn. przesunięcie wykresu zmian wartości częstotliwości o 30 sekund w lewo).

W przypadku braku informacji o czasie rejestracji, konieczne jest wyszukanie sekwencji częstotliwości w bazie danych. Należy wybrać dowolny wektor N danych częstotliwości i dokonać przeszukania bazy danych, biorąc pod uwagę jedno z kryteriów: błąd średniokwadratowy (MSE – *Mean Squared Error*) lub współczynnik korelacji. Wektor N danych należy wybrać w taki sposób, by nie występowały żadne zaburzenia amplitudy i fazy sygnału, a zakres zmian częstotliwości mieścił się w zakresie zmian uznawanych przez UCTE [1].

Wynikiem zastosowanej metody jest informacja o liczbie występujących sygnałów o częstotliwości sieci energetycznej, dacie i czasie rejestracji materiału dowodowego oraz o liczbie fragmentów nagrania, które mogą zostać uznane za ciągłe. W przypadku stwierdzenia więcej niż jednego ciągłego fragmentu nagrania, należy stwierdzić ingerencję w ciągłość zapisu – jeśli nie występują inne okoliczności usprawiedliwiające powstanie tego zjawiska. Podczas analizy nagrania należy zwrócić uwagę na charakterystykę widmową materiału dowodowego w zakresie małych częstotliwości; w szczególności na zmniejszenie gęstości widma szumu w paśmie zbliżonym do częstotliwości sieci energetycznej. Może to świadczyć o celowej eliminacji sygnału sieci z nagrania dowodowego.

Graficzna analiza w dziedzinie czasu i częstotliwości

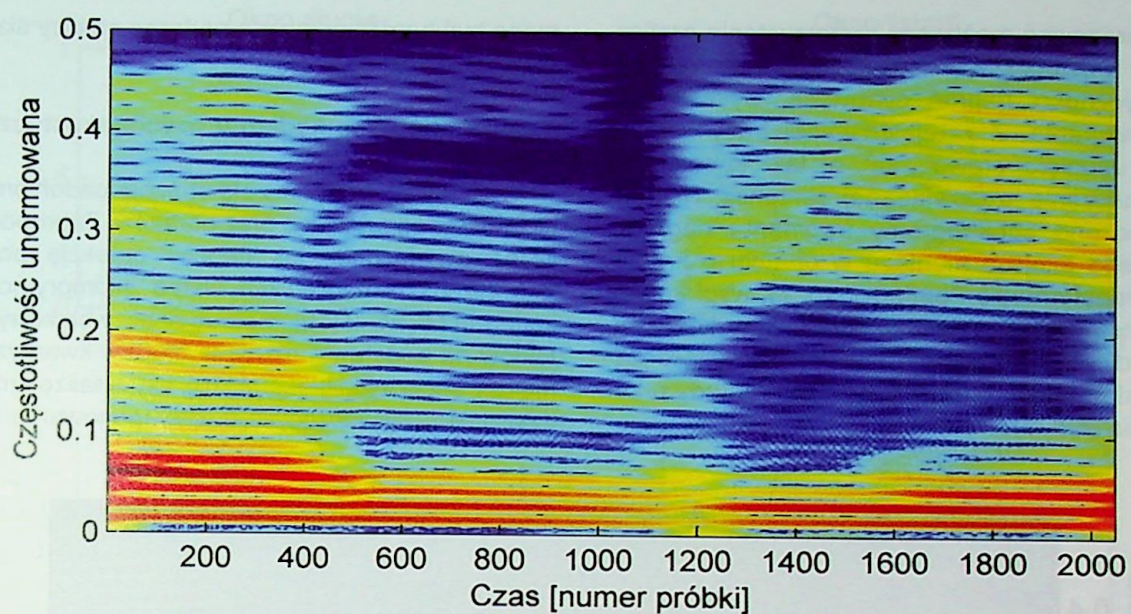
Graficzna analiza spektrograficzna może być także wykorzystana do wykrywania śladów ingerencji w ciągłość nagrania. Tradycyjna realizacja krótkoczasowej transformacji Fouriera (STFT) okazuje się przydatna w przypadku obserwacji sygnałów o stałej lub wolnozmiennej częstotliwości. Takie sygnały często pojawiają się jako tło akustyczne nagrania i najczęściej stanowią zakłócenia towarzyszące rejestrowanej rozmowie.

Zmodyfikowanie zapisu w postaci wycięcia fragmentu lub wstawienia innego także narusza ciągłość sygnałów zakłócających, co można zaobserwować w postaci zmiany fazy sygnału [12] i charakterystycznego rozmycia (por. ryc. 2). Skokowej zmianie amplitudy sygnału w dziedzinie czasu towarzyszy bowiem szeroki zakres widmowy.

Podobnie jak w przypadku sygnałów zakłócających, montaż zarejestrowanych ciągłych wypowiedzi może zostać zauważony dzięki obserwacji struktury widmowej sygnału mowy, szczególnie w odniesieniu do samogłosek. Z zasady nieoznaczoności Heisenberga-Gabora wynikają sprzeczne wymagania odnośnie do rozdzielczości częstotliwościowej i czasowej. Dobra rozdzielczość w dziedzinie czasu wymaga zastosowania krótkiego okna, aby jak najczęściej obliczać widmo sygnału. To z kolei powoduje zmniejszenie rozdzielczości w dziedzinie częstotliwości, ze względu na zmniejszenie liczby prążków widma wyznaczanych za pomocą transformacji Fouriera. Ponadto na zmniejszenie czytelności spektrogramu wpływa stosowanie odpowiedniej funkcji okna, która używana jest w celu minimalizacji efektu ucięcia analizowanego sygnału. Także wykorzystanie zakładek (*overlapping windows*) powoduje, że kolejne segmenty spektrogramu przestają być niezależne [13].

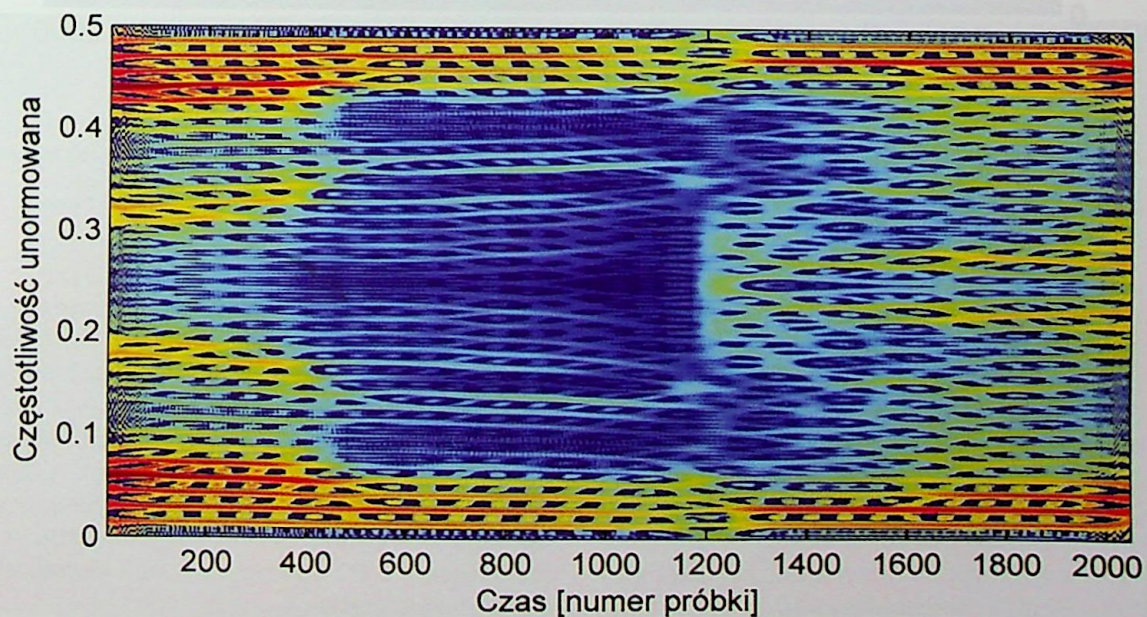
Zamiast okna o stałej długości można wprowadzić okno o rozmiarze zmiennym. Zastosowanie w tym celu odwróconej w czasie wersji badanego sygnału wykorzystane zostało w transformacji Wignera-Ville'a [14], [15]. Zaletą przekształcenia jest bardzo dokładne odwzorowanie zmian częstotliwości w czasie, jednakże kwadratowa natura dystrybucji powoduje powstawanie dodatkowych składowych zlokalizowanych pomiędzy sygnałami. Ze względu na oscylacyjną naturę tych interferencji, można je redukować, stosując w tym celu filtrację dolnoprzepustową. Takie wygładzanie stanowi kompromis między dobrą lokalizacją a brakiem zakłóceń [16]. Wygładzone dystrybucje Wignera-Ville'a znajdują szerokie zastosowanie także w analizie sygnału mowy [17].

Podczas obliczania każdego punktu spektrogramu o współrzędnych czas-częstotliwość, wartości potrzebne do jego wyznaczenia stanowią pewien rozkład wokół geometrycznego środka domeny. Dużą poprawę czytelności można uzyskać, stosując metodę przemieszczenia polegającą na korygowaniu lokalizacji energii na płaszczyźnie czas-częstotliwość. Podstawowym założeniem jest tu relokacja energii w kierunku lokalnego środka ciężkości. Realizacja tego procesu odbywa się przez wygładzenie, którego głównym celem jest redukcja interferencji o charakterze oscylacyjnym oraz kompresja geometryczna, powodująca uwydatnienie składowych, które nie zostały usunięte [15]. Istotną zaletą metody jest także fakt zachowywania na drodze przekształcenia informacji o fazie sygnału. Metoda przemieszczenia wykorzystywana była przeważnie do poprawy czytelności krótkoczasowej transformacji Fouriera [17], [18], jednakże można ją zastosować także do innych przekształceń, np. transformacji Wignera-Ville'a.



Ryc. 7. Wykres przedstawiający analizowany fragment wypowiedzi wyznaczony dzięki zastosowaniu metody krótkoczasowej transformacji Fouriera (spektrogram)

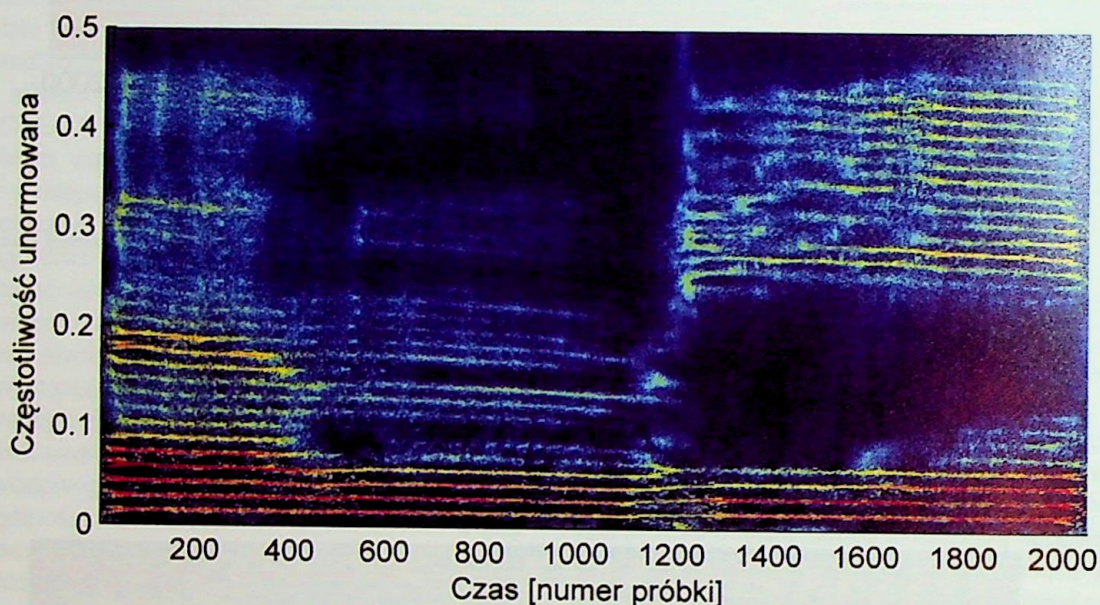
Fig. 7. Time-frequency plot of analyzed recorded utterance computed using short-time Fourier transform (spectrogram)



Ryc. 8. Wykres przedstawiający analizowany fragment wypowiedzi wyznaczony dzięki zastosowaniu metody przemieszczenia do wygładzonej pseudoprezentacji Wignera-Ville'a

Fig. 8. Time-frequency plot of analyzed recorded utterance computed using reassignment method applied to the smoothed-pseudo Wigner-Ville representation

Aby zobrazować możliwość wykorzystania graficznej analizy w dziedzinie czasu i częstotliwości do oceny autentyczności nagrań cyfrowych, wykasowano fragment nagrania. Z zarejestrowanej frazy „jestem niewinny” usunięto słowo „nie” w taki sposób, by zmienić znaczenie wypowiedzianej sentencji. Następnie wyznaczono spektrogram fragmentu nagrania obejmującego miejsce montażu w oparciu o krótkoczasową transformację Fouriera (ryc. 7), wygładzoną pseudotransformację Wignera-Ville’a z wykorzystaniem metody przemieszczenia (ryc. 8) oraz krótkoczasową transformację Fouriera z użyciem metody przemieszczenia (ryc. 9).



Ryc. 9. Wykres przedstawiający analizowany fragment wypowiedzi wyznaczony dzięki zastosowaniu metody przemieszczenia do krótkoczasowej transformacji Fouriera (spektrogram)

Fig. 9. Time-frequency plot of analyzed recorded utterance computed using reassignment method applied to the short-time Fourier transform (spectrogram)

Miejsce ingerencji w ciągłość zapisu jest widoczne na wszystkich wymienionych rycinach dla czasu równego 1200. próbkę. W przypadku STFT pojawia się jednak problem związany z możliwością pomylenia miejsca montażu ze śladami pochodzącymi od sygnałów zakłócających. Zastosowanie metody przemieszczenia znacznie poprawia czytelność spektrogramu i pozwala na łatwiejszą lokalizację miejsc, w których sygnał mowy został zdeformowany. W przypadku reprezentacji Wignera-Ville’a – nawet w wersji wygładzonej i zmodyfikowanej – ujawnia się szkodliwy wpływ interferencji na czytelność wyznaczonego wykresu. Mimo to można dostrzec miejsce ingerencji w ciągłość nagrania.

Należy nadmienić, że tego typu obserwacje, jakkolwiek nie przesądzają o braku autentyczności nagrania,

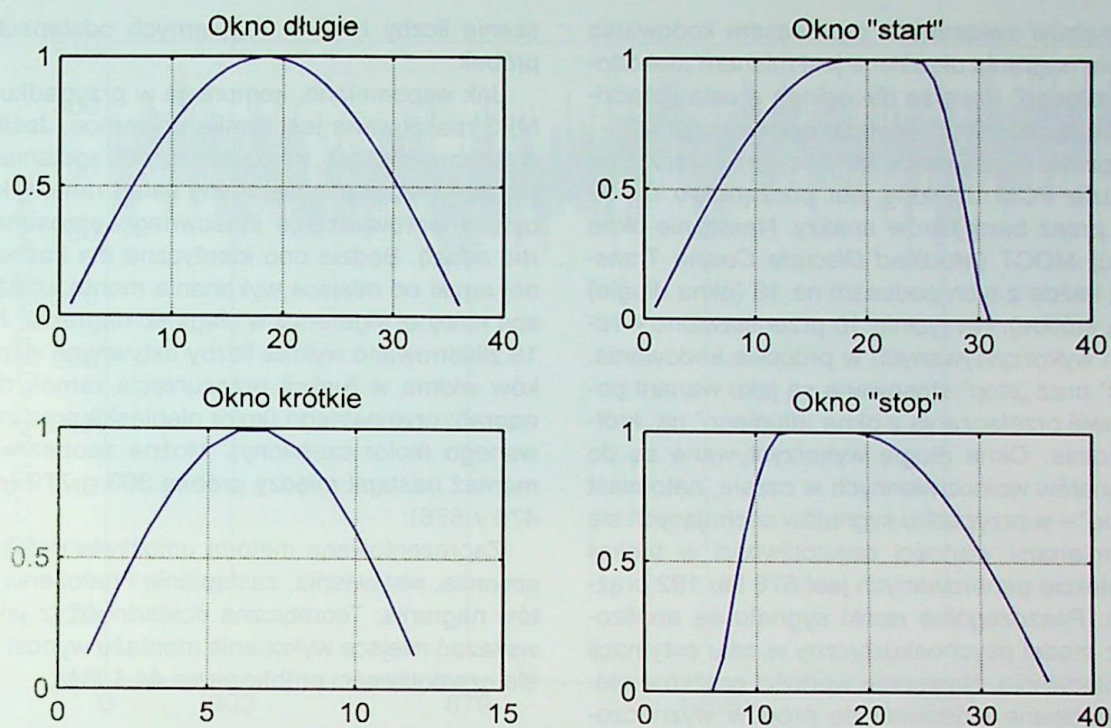
mogą być bardzo pomocne podczas analizy śladów ingerencji w ciągłość zapisu.

Analiza stratnych koderów psychoakustycznych

Wiele uwagi poświęca się obecnie badaniom autentyczności nagrań wideo oraz obrazów. Zaproponowano liczne rozwiązania, jak choćby detekcję podwójnej kwantyzacji w nagraniach wideo skompresowanych koderem MPEG-2 [19], metodę analizy blokowych artefaktów na podstawie różnic w błędzie kwantyzacji pomiędzy sąsiednimi blokami [20], czy analizę zmian rozdzielczości, świadczących o wkomponowaniu w bada-

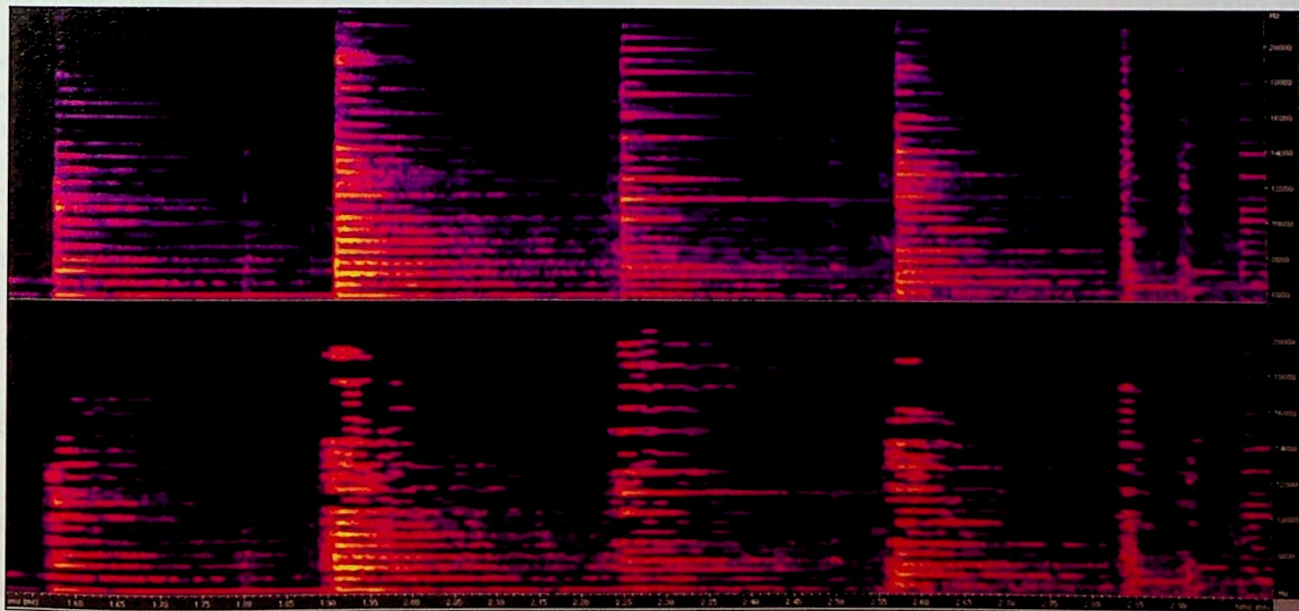
ny obraz innego obrazu [21]. Niestety, nie ma możliwości bezpośredniej adaptacji powyższych metod do kryminalistycznej analizy plików MP3. Spowodowane jest to znacznymi różnicami pomiędzy tymi standardami kompresji. Przykładowo koder MP3 grupuje próbki sygnału w ramki z 50% nakładaniem, podczas gdy w kompresji JPEG nakładanie nie występuje. Na skutek tego wykrywanie blokowych artefaktów w przypadku kompresji MP3 z zastosowaniem metod JPEG staje się problematyczne.

Nie jest możliwe bezpośrednie wykorzystanie metod zaadaptowanych do analizy obrazów skompresowanych koderem JPEG, jednakże można zaadaptować samą ideę badania śladów, jakie pozostawiły algorytmy kompresji stratnej. Takie podejście polegające na usta-



Ryc. 10. Wykres przedstawiający różne kształty okien stosowanych w koderze MPEG Layer 3 (MP3)

Fig. 10. Plot with four different analysis windows used in MPEG Layer 3 (MP3) encoder



Ryc. 11. Wykres przedstawiający spektrogramy fragmentu nagrania oryginalnego (wykres górny) oraz skompresowanego (wykres dolny) z wykorzystaniem kodera MP3. Zakres analizy: 6–22,05 kHz, rozmiar okna: 512 próbek.

Fig. 11. Spectrograms of recordings: non-encoded (upper diagram) and encoded using MP3 encoder (lower diagram). Analysis in range: 6–22,05 kHz, window size: 512 samples.

laniu parametrów związanych z procesem kodowania na podstawie nagrania określane jest mianem „dekodowania odwrotnego” (*inverse decoding*), a cała dziedzina – „inżynierią odwrotną” (*inverse engineering*) [22].

W przypadku kodowania MP3 sygnał wejściowy w standardzie PCM dzielony jest początkowo na 32 podpasma przez bank filtrów analizy. Następnie okno transformacji MDCT (*Modified Discrete Cosine Transform*) dzieli każde z tych podpasm na 18 (okna długie) lub 6 (okna krótkie). Na rycinie 10 przedstawiono 4 rodzaje okien wykorzystywanych w procesie kodowania. Okna „start” oraz „stop” stosowane są jako wariant pośredni w chwili przełączania z okna „długiego” na „krótkie” i odwrotnie. Okna długie wykorzystywane są do analizy sygnałów wolnozmiennych w czasie, natomiast okna „krótkie” – w przypadku sygnałów cechujących się szybkimi zmianami wartości częstotliwości w funkcji czasu. W efekcie generowanych jest 576 lub 192 prążków widma. Poszczególne ramki sygnału są analizowane przez model psychoakustyczny w celu estymacji progów maskowania. Następnie wartości prążków widma są kwantowane stosownie do progów wyznaczonych na podstawie modelu. Dekodowanie następuje przez odwrotną kwantyzację, a współczynniki są ponownie przenoszone do podpasm przez zastosowanie odwrotnej transformacji MDCT. Na rycinie 11 zilustrowano proces zerowania niektórych składników widma przez porównanie spektrogramu nagrania nieskompresowanego oraz skompresowanego z wykorzystaniem koderów MP3.

W procesie kodowania i kwantyzacji wiele prążków widma jest zerowanych, ze względu na ich maskowanie. Różnicę między widmem skwantowanym a nieskwantowanym można zaobserwować, stosując skalę logarymiczną. Cechą charakterystyczną jest to, że widmo skwantowane można zaobserwować jedynie, stosując okno analizy z takim samym przesunięciem, jakie wykorzystane zostało w procesie kodowania [23]. Brak synchronizacji ramek na skutek przesunięcia nawet o jedną próbkę nie pozwala na obserwację zjawiska. Istotne jest także zastosowanie tego samego algorytmu dekompozycji oraz takich samych kształtów okien analizy. To utrudnia badanie, ponieważ konieczne jest wykonanie analiz dla wszystkich 4 okien w przypadku koderów.

W celu sprawdzenia możliwości praktycznego wykorzystania opisanego zjawiska dokonano dekodowania wstecznego nagrania skompresowanego za pomocą koderów MPEG Layer 3 z przepływnością 128 kbps oraz z częstotliwością próbkowania 44100 Hz. Analizę przeprowadzono, przesuując okno „długie” co jedną próbkę i zliczając liczbę aktywnych współczynników widma (NAC – *Number of Active Coefficients*). Jak można zaobserwować na rycinie 12, następuje skokowe zmniejszenie

liczby NAC w regularnych odstępach co 576 próbek.

Jak wspomniano, kompresja w przypadku koderów MP3 realizowana jest ramka po ramce. Jeśli zostanie wykonany montaż, to rozkład ramek zostanie naruszony. Aby przywrócić pierwotny układ ramek, konieczne byłoby wprowadzenie stosownego przesunięcia (*frame offset*). Będzie ono identyczne dla każdej następnej ramki od miejsca wykonania montażu, aż do miejsca kolejnej ingerencji w ciągłość nagrania. Na rycinie 13 zilustrowano wykres liczby aktywnych współczynników widma w funkcji przesunięcia ramek dla dwóch nagrań: oryginalnego (kolor niebieski) oraz zmodyfikowanego (kolor czerwony). Można zaobserwować, że montaż nastąpił między próbką 303 a 779 (ponieważ: $476 \neq 576$).

Zaprezentowana metoda umożliwia detekcję wykaszowania, wstawienia, zastąpienia i nałożenia fragmentów nagrania. Teoretyczna dokładność, z jaką można wskazać miejsce wykonania montażu wynosi ok. 13 ms dla częstotliwości próbkowania 44,1 kHz.

Analiza statystyczna wyższego rzędu

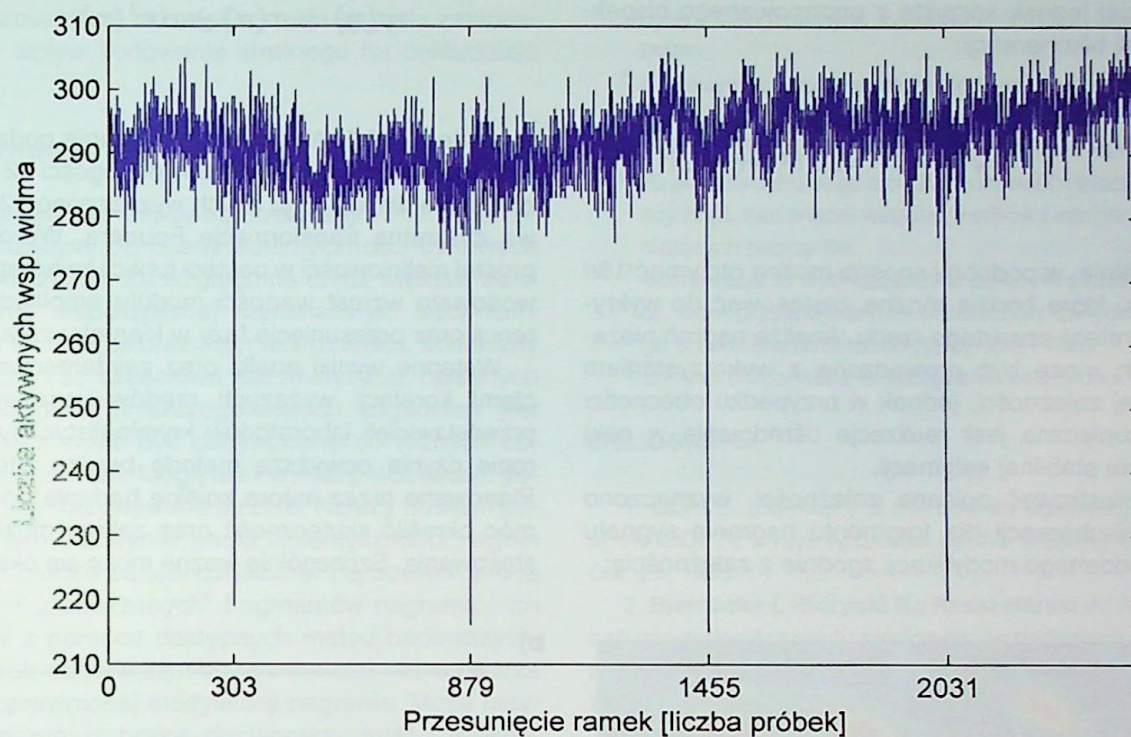
Kolejną metodą, która może służyć do oceny autentyczności cyfrowych nagrań audio jest znana od wielu lat analiza widmowa wyższego rzędu. To technika, która dotychczas wykorzystywana była w pracach badawczych przede wszystkim do analizy obrazów [24], [25]. Zakłada się, że naturalne sygnały nie posiadają lub posiadają w dziedzinie częstotliwości bardzo słabe korelacje wyższego rzędu. Z kolei wprowadzenie do takiego sygnału nieliniowości, co często następuje na skutek wykonywania montażu poprzez sklejanie lub wycinanie fragmentów nagrania, powoduje powstawanie pewnych nienaturalnych korelacji. Wykorzystanie opisanego zjawiska może mieć istotne znaczenie dla badań integralności cyfrowych nagrań dźwiękowych.

Widmo mocy sygnału, które często jest wykorzystywane do wykrywania korelacji drugiego rzędu, można zapisać jako:

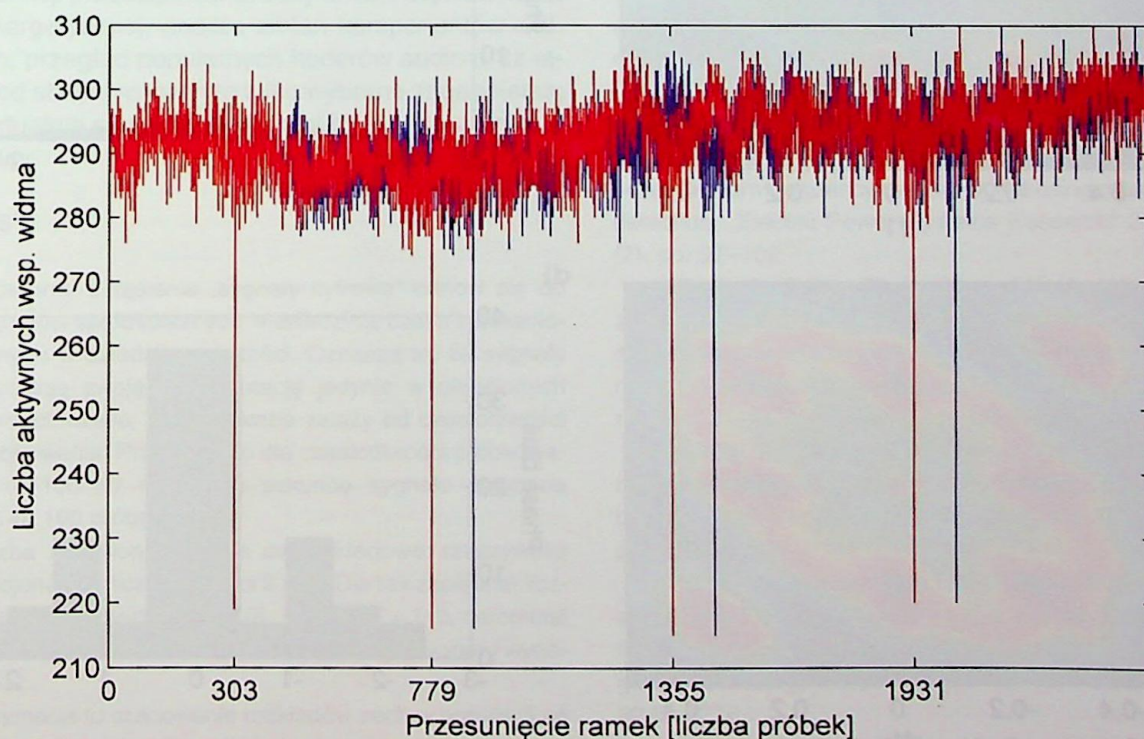
$$P(\omega) = X(\omega)X^*(\omega) \quad (6)$$

Aby mieć możliwość detekcji korelacji wyższego rzędu, konieczne jest zastosowanie bardziej złożonych przekształceń. Do obserwacji korelacji trzeciego rzędu używane jest bispektrum [26]:

$$B_{bisp}(\omega_1, \omega_2) = X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2) \quad (7)$$



Ryc. 12. Zmiana liczby aktywnych współczynników widma w funkcji przesunięcia (synchronizacji) ramki obserwacji w koderze MP3
 Fig. 12. Number of active spectral coefficients in a function of frame offset, according to MP3 encoder



Ryc. 13. Zmiana liczby aktywnych współczynników widma w funkcji przesunięcia (synchronizacji) ramki obserwacji dla fragmentu nagrania bez modyfikacji (wykres koloru niebieskiego) oraz fragmentu zmodyfikowanego nagrania (wykres koloru czerwonego)
 Fig. 13. Number of active spectral coefficients in a function of frame offset, according to MP3 encoder (red plot – tapered recording, blue plot – recording without modification)

Częściej jednak korzysta z unormowanego bispektrum, czyli bikoherencji:

$$B_{bicoh}(\omega_1, \omega_2) = \frac{X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2)}{\sqrt{|X(\omega_1)X(\omega_2)|^2|X(\omega_1+\omega_2)|^2}} \quad (8)$$

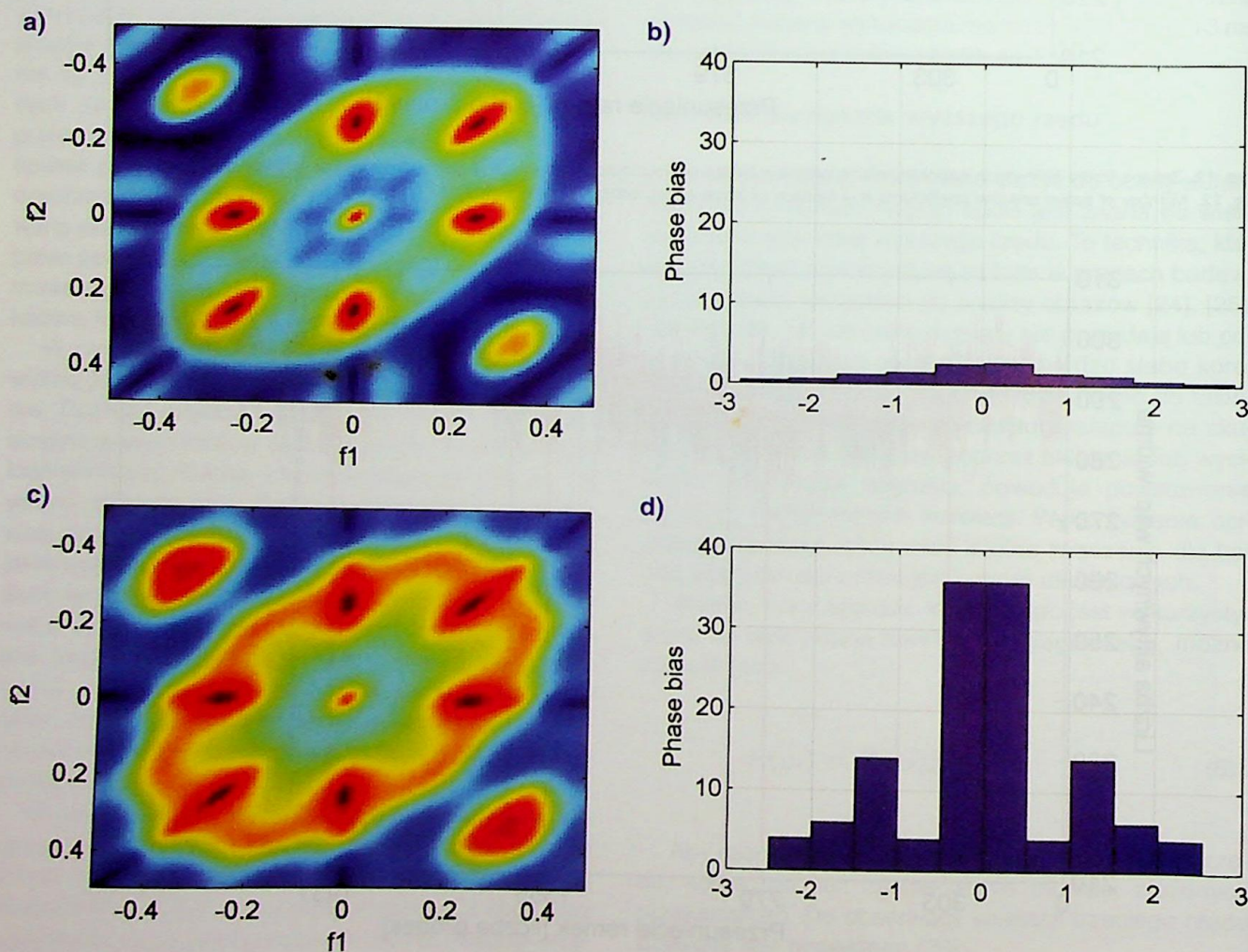
Naturalnie, w podobny sposób można otrzymać 18ri spektrum, które będzie można zastosować do wykrywania korelacji czwartego rzędu. Analiza nagrań niezakłóconych może być prowadzona z wykorzystaniem powyższej zależności, jednak w przypadku obecności szumu konieczna jest realizacja uśredniania w celu otrzymania stabilnej estymacji.

Aby zilustrować opisane zależności, wyznaczono funkcję bikoherencji dla fragmentu nagrania sygnału mowy poddanego modyfikacji zgodnie z zależnością:

$$y(n) = x(n) + ax^2(n) \quad (9)$$

gdzie $a = 0,5$. Analizowane nagranie podzielono na krótkie, nakładające się odcinki o długości 32 próbek, a następnie dla każdego z nich wyznaczono 128-punktową dyskretną transformację Fouriera. Wprowadzenie prostej nieliniowości w postaci funkcji kwadratowej spowodowało wzrost wartości modułu amplitudy bikoherencji oraz przesunięcie fazy w kierunku zera.

Wstępne wyniki analiz oraz zainteresowanie funkcjami korelacji wyższych rzędów wyrażane przez przedstawicieli laboratoriów kryminalistycznych w Europie czynią powyższą metodę bardzo interesującą. Planowane przez autora kolejne badania powinny pomóc określić skuteczność oraz zakres możliwości jej stosowania. Szczególnie ważne może się okazać usta-



Ryc. 14. Ilustracja przedstawiająca analizy nagrań: oryginalnego (wykresy górne) oraz zmodyfikowanego (wykresy dolne). Dla obydwu nagrań wyznaczono moduł amplitudy (a, c) oraz histogram przesunięcia fazy (b, d). Nieliniowość została wprowadzona zgodnie z funkcją opisaną przez równanie (9).
Fig. 14. Diagram showing an original (upper plots) and modified (lower plots) recording. Both types of recordings were subjected to magnitude (a, c) and phase histogram analysis (b, d). Non-linearity was introduced into the recording by function described by equation (9).

lenie podatności na różnego rodzaju sygnały zakłócające oraz wpływ kodowania stratnego na dokładność metody.

Podsumowanie

Przedstawione metody są obecnie głównymi narzędziami służącymi do oceny autentyczności cyfrowych nagrań fonicznych. Ze względu na coraz większe zainteresowanie współczesnej kryminalistyki badaniami z zakresu fonoskopii oraz inżynierii dźwięku, konieczny jest rozwój oraz częściowa automatyzacja opisanych metod. Każde z wyszczególnionych zagadnień jest specyficzne i może być dedykowane do różnych obszarów zastosowań. Mogą też i w miarę możliwości powinny być wykorzystywane łącznie. Należy bowiem pamiętać o tym, że wyniki analiz autentyczności nagrań cyfrowych rzadko są jednoznaczne. Są bowiem próbą znalezienia „podejrzanych” fragmentów nagrania i ich weryfikacji z pomocą dostępnych metod badawczych. Z kolei brak obserwacji miejsc montażu nie wyklucza faktu nieuprawnionej modyfikacji nagrania. Może jedynie świadczyć o braku dostępnych metod analizy. W związku z powyższym każdy z wymienionych obszarów będzie przedmiotem dalszych badań. Rozwój metod pomiaru częstotliwości odpornych na zakłócenia, dobór funkcji przeszukiwania bazy zmian częstotliwości sieci energetycznej, analiza zmian komponentów widmowych, przegląd popularnych koderów audio oraz testy metod statystycznych to tylko wybrane zagadnienia, które aktualnie stanowią przedmiot badań autora niniejszej pracy.

PRZYPISY

- ¹ Popularne określenie „sygnały cyfrowe” odnosi się do sygnałów próbkowanych w dziedzinie czasu i skwantowanych w dziedzinie wartości. Oznacza to, że sygnały posiadają swoją reprezentację jedynie w określonych chwilach czasu, których liczba zależy od częstotliwości próbkowania. Przykładowo dla częstotliwości próbkowania 44100 Hz na każdą sekundę sygnału przypada $n = 44\ 100$ próbek.
- ² Liczba zespolona zawiera dwie składowe: rzeczywistą i urojoną, np. liczba postaci $2 + j2$. Dla tak zapisanej liczby jej moduł wynosi: $|2 + j2| = \sqrt{2^2 + 2^2} = 2\sqrt{2}$, natomiast argument: $\pi/4 = 45^\circ$. Wartość składowej urojonej j wynosi $\sqrt{-1}$.
- ³ Estymacja to szacowanie rozkładów cech w populacji na podstawie cech jednostek wchodzących w skład losowo dobranej próby (źródło: Encyklopedia PWN).
- ⁴ Ortogonalność to wspólna nazwa dla prostokątności i jej wszystkich uogólnień; mówi się m.in. o ortogonalnych krzywych, wektorach, macierzach, funkcjach ortogonal-

nych i ortogonalnych szeregach (źródło: Encyklopedia PWN).

- ⁵ Aproksymacja to zastąpienie jednych wielkości matematycznych przez inne, przybliżone (źródło: Encyklopedia PWN).
- ⁶ Rząd filtru cyfrowego określa złożoność układu; im większy rząd, tym więcej współczynników i elementów opóźniających tworzy filtr.
- ⁷ Interpolacja to wyznaczenie w pewnym przedziale funkcji, która przyjmuje znane wartości dla danych liczb z tego przedziału (źródło: Encyklopedia PWN).
- ⁸ Symbol „*” oznacza tu sprzężenie zespolone.

BIBLIOGRAFIA

1. UCPTE, „Summary of the Current Operating Principles of the UCPTE”, Text Approved by the Steering Committee, Oct. 28, 1998.
2. **Biernacka I., Korycki R., Rzeszotarski J.**: Analiza wahań częstotliwości prądu sieciowego w badaniach autentyczności nagrań cyfrowych, „Problemy Kryminalistyki” 2007, nr 258, s. 36–40.
3. **Grigoras C.**: Digital Audio Recording Analysis: The Electric Network Frequency (ENF) Criterion, „The International Journal of Speech Language and the Law” 2005, vol. 12 (2), pp. 63–76.
4. **Grigoras C.**: Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis, „Forensic Science International” 2007, vol. 167, pp. 136–145.
5. **Zieliński P.**: Cyfrowe przetwarzanie sygnałów, Warszawa, WKŁ 2005, s. 239–241.
6. **Salcic Z., Mikhael R.**: A new method for instantaneous power system frequency measurement using reference points detection, „Electric Power Systems Research” 2000, vol. 55 (2), pp. 97–102.
7. **Zang C., Dai L., Zheng H. and He J.**: Using frequency zoom technology to realize high precision and adaptive frequency measurement for power system, in Power System Technology, International Conference – PowerCon 2004, Singapore, Nov. 21–24 2004, vol. 1, pp. 155–159.
8. **Aiello M., Cataliotti A.**: Chirp-Z Transform-Based Synchronizer for Power System Measurements, „IEEE Transactions on Instrumentation and Measurement” 2005, vol. 54 (3), pp. 1025–1032.
9. **Stoica P., Soderstrom T.**: Statistical analysis of MUSIC and ESPRIT estimates of sinusoidal frequencies, in Proc. The IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, Apr. 14–17, 1991, vol. 5, pp. 3273–3276.
10. **Tjader A., Gu I., Bollen M., Ronnberg S.**: Performance evaluation for frequency estimation of transients using the ESPRIT: Measured noise versus white noise, in Proc. 13th International Conference of Harmonics and Quality of Power, Wollongong, Sept. 28, 2008, pp. 1–8.

11. **Leonowicz Z., Lobos T.:** Parametric Spectral Estimation for Power Quality Assessment, in Proc. The International Conference on „Computer as a Tool”, Warsaw, Sept. 9–12, 2007, pp. 1641–1647.
12. **Nicolalde D., Apolinario J.:** Evaluating digital audio authenticity with spectral distances and ENF phase change, in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Apr. 19–24, 2009, pp. 1417–1420.
13. **Korycki R.:** Zastosowanie metod czasowo-częstotliwościowych w badaniach autentyczności cyfrowych nagrań fonicznych, 13th International Symposium on Sound Engineering and Tonmeistering, ISSET 2009, Warsaw, Oct. 16–18, 2009, pp. 73–80.
14. **Flandrin P., Auger F., Chassande-Mottin E.:** Time-Frequency Reassignment: From Principles to Algorithms, Tempe, Arizona: CRC Press, 2003.
15. **Papandreou-Suppappola A.:** Application in Time-Frequency Signal Processing, Arizona, CRC Press, 2003.
16. **Korycki R.:** Methods of Time-Frequency Analysis in Authentication of Digital Audio Recordings, „International Journal of Electronics and Telecommunications” 2010, 56 (3), pp. 257–261.
17. **Devedeux D., Duchene J. and Marque C.:** Use of synthetic uterine signals for an optimum choice of time/frequency representation, in Proc. 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Engineering Advances: New Opportunities for Biomedical Engineers, Baltimore, Nov. 3–6, 1994, vol. 2, pp. 1246–1247.
18. **Auger F. and Flandrin P.:** Improving the readability of time-frequency and time-scale representations by the reassignment method, „IEEE Transactions of Signal Processing” 1995, vol. 43 (5), pp. 1068–1089.
19. **Farid H. and Wang W.:** Exposing Digital Forgeries in Video by Detecting Double Quantization, in Proc. ACM Multimedia and Security Workshop, Princeton 2009.
20. **Chen Y. and Hsu C.:** Image tampering detection by blocking periodicity analysis in JPEG compressed images, in Proc. IEEE 10th Workshop on Multimedia Signal Processing, Cairns, Queensland, Oct. 8–10, 2008, pp. 803–808.
21. **Farid H. and Popescu A.C.:** Exposing Digital Forgeries by Detecting Traces of Re-sampling, „IEEE Transactions on Signal Processing” 2005, vol. 53 (2), pp. 758–767.
22. **Geiger R., Herre J. and Moehrs S.:** Analysing decompressed audio with the Inverse Decoder – towards an operative algorithm, in Proc. 112th AES Convention, Munich, May 10–13, 2002.
23. **Yang R., Qu Z., Huang J.:** Detecting digital audio forgeries by checking frame offsets. International Multimedia Conference, Proceedings of the 10th ACM workshop on Multimedia and security table of contents, Oxford, United Kingdom, 21–26, 2008.
24. **Ng T., Chang S., and Sun Q.:** Blind detection of photomontage using higher order statistics, In Proceedings of the 2004 International Symposium on Circuits and Systems, 2004. ISCAS'04, 2004.
25. **Farid H., Lyu S.:** Higher-Order Wavelet Statistics and their Application to Digital Forensics, IEEE Workshop on Statistical Analysis in Computer Vision (in conjunction with CVPR), Madison, Wisconsin 2003.
26. **Farid H.:** Detecting digital forgeries using bispectral analysis, Massachusetts Institute of Technology, Cambridge, MA, 1999.

Streszczenie

Pojawienie się cyfrowych nagrań audio sprawiło, że wykonywanie badań autentyczności stało się dużo trudniejsze, a w wielu wypadkach wręcz niemożliwe. Powszechnie dostępne rozwiązania oraz darmowe oprogramowanie służące do edycji dźwięku pozwalają fałszerzom na wycinanie i wstawianie poszczególnych słów w sposób dowolny i bez wprowadzania słyszalnych niekształceń. Obecnie najczęściej wykorzystywanym rozwiązaniem używanym do wykrywania montażu jest analiza wahań częstotliwości prądu sieci energetycznej.

Celem niniejszej pracy jest przedstawienie problemu detekcji śladów ingerencji w ciągłość zapisu oraz omówienie podstawowych metod wykorzystywanych do badania autentyczności cyfrowych nagrań fonicznych. Przedstawiono metody pomiaru częstotliwości, które następnie porównano z tradycyjnie używaną transformacją Fouriera, wykorzystywaną podczas wyodrębniania i analizy sygnału sieci energetycznej. Zaprezentowano możliwość graficznego wykrywania śladów montażu za pomocą analiz czasowo-częstotliwościowych wykonanych z użyciem wybranych transformacji czasowo-częstotliwościowych, których czytelność może być poprawiona dzięki zastosowaniu metody przemieszczenia. Rozwiązanie to umożliwia analizę minimalnych zmian tła akustycznego, które to zmiany mogą być oznaką modyfikacji nagrania. Trzecią przedstawioną metodą jest analiza przesunięcia ramek w plikach dźwiękowych skompresowanych z wykorzystaniem stratnych koderów psychoakustycznych. Wpływ parametrów algorytmu kompresji oraz kształtu użytego okna zostanie zaprezentowany na przykładzie kodera MP3. Ostatnim prezentowanym zagadnieniem są techniki analizy polispektralnej. W skrócie przedstawiono możliwość wykorzystania powyższych narzędzi w detekcji śladów ingerencji w ciągłość zapisów oraz możliwość ich zastosowania do analizy sygnału mowy.

Słowa kluczowe: badanie autentyczności, częstotliwość sieci energetycznej, detekcja montażu, analiza czasowo-częstotliwościowa, metoda przemieszczenia, dekodek odwrotny, MP3, analiza bispektralna, HOSA.

Summary

Since digital audio recordings appeared, audio authentication has become more difficult and in most cases impossible. Currently available technologies and free editing software allow the forger to cut or paste any single word without

audible artifacts. Nowadays, the most frequently used solution in tampering detection is the Electric Network Frequency (ENF) method.

This paper describes the problem of tampering detection and discusses the main methods used for authenticity analysis of digital audio recordings. For the first topic, frequency measurement algorithms are described and compared with a simple Fourier transform generally used in forensic ENF extraction. Time-frequency analysis plots computed by selected transforms are presented and improved with reassignment method in purpose of visual inspection of modified recordings. Using these solutions, one can analyze minimal changes of

background sounds, which can indicate tampering. The third approach is based on checking frame offsets in compressed audio files by using perceptual audio coding. The influence of compression algorithm parameters and window shapes are presented using MP3 encoder as an example. Additionally, the techniques from polyspectral analysis are described shortly. The usage of these tools in detecting forgeries is shown and the effectiveness in analyzing human speech is discussed.

Keywords: authenticity analysis, electric network frequency, tampering detection, time-frequency analysis, reassignment method, inverse decoder, MP3, bispectral analysis, HOSA.