

PARAMETRYCZNE KODOWANIE HARMONICZNYCH OBIEKTÓW DŹWIĘKOWYCH *

Streszczenie: Artykuł opisuje prototypowy system kodowania sygnałów fonii w oparciu o schemat obiektowy. W celu kompresji sygnał podlega dekompozycji na tzw. obiekty, które reprezentują odrębne źródła lub zdarzenia dźwiękowe. Obiekty harmoniczne (złożone z wielotonów) reprezentowane są przez dwuwymiarową strukturę czasowo-częstotliwościową (obwiednię harmoniczną) oraz jednowymiarowy tzw. prototyp wąskopasmowy. Oba elementy są następnie kodowane stratnie metodą transformatową. Opisywana technika zapewnia bardzo dobrą jakość dźwięku przy prędkości strumienia przypadającej na pojedynczy obiekt harmoniczny w zakresie 4-5kb/s.

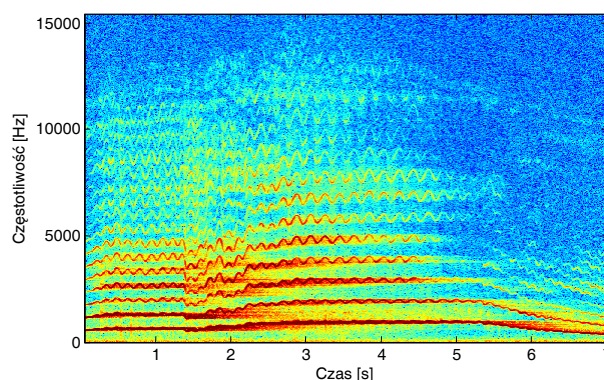
1. WSTĘP

Popularne schematy kompresji sygnału fonii zazwyczaj ignorują strukturę programu zawierającego dźwięki wielu instrumentów występujące równocześnie i w krótkim następstwie. Podstawową wadą klasycznego kodowania transformatowego [1,2] jest to, że jedynym sposobem redukcji prędkości bitowej jest ograniczenie pasma lub zwiększenie stopnia kwantyzacji, przy czym oba kroki powodują powstanie nienaturalnych zniekształceń (np. nieliniowych interferencji między składnikami widmowymi). Z kolei w przypadku kompresji parametrycznej [3,4,5] w strumieniu bitowym przesyłany jest zestaw parametrów modeli (np. sinusoidalnych), które aproksymują perceptualnie istotne składniki dźwięku. Redukcja strumienia polega zazwyczaj na rezygnacji z kodowania niektórych składników, co może również prowadzić do nienaturalnych artefaktów.

Kodowanie obiektowe jest nowym sposobem kompresji fonii, w którym dźwięki z indywidualnych źródeł, ich części lub grupy składowych wykazujących wspólne cechy są reprezentowane przez niezależne partie strumienia danych [6,7,8]. Takie kodowanie ma zastosowanie zarówno przy transmisji dźwięku wielokanałowego, szczególnie z bardzo dużą liczbą kanałów, np. towarzyszącego telewizji 3D (gdyż pozwala na dowolną manipulację geometrią sceny akustycznej i pozycją słuchacza poprzez modyfikację sposobu łączenia obiektów po stronie dekodera), jak i transmisji fonii cyfrowej z bardzo małą prędkością bitową (gdyż ułatwia ono świadomą selekcję kodowanych elementów, gdy część informacji trzeba odrzucić).

Potencjalną wadą kodowania obiektowego jest to, że łączny strumień generowany przez poszczególne obiekty może być większy niż strumień generowany przez tradycyjny kodek operujący na zmiksowanych

składnikach programu. Potrzebne są więc nowe metody kompresji minimalizujące tę nadmiarowość. Technika opisana w niniejszym artykule jest próbą odpowiedzi na tę potrzebę w odniesieniu do kodowania tzw. *obektów harmonicznych*, a więc dźwięków o strukturze wielotonów, posiadających znaczącą liczbę koherentnych składowych harmonicznych. Takie (często niestacjonarne) sygnały są zazwyczaj szerokopasmowe (rys. 1), a ich kompresja klasycznymi metodami jest mało efektywna i wprowadza słyszalne zniekształcenia [5,6]. Niniejszy artykuł nie dotyczy problematyki dekompozycji nagrań muzyki na obiekty oraz kodowania innych kategorii obiektów, które są przedmiotem wielu prac, np. [7,8,9].



Rys. 1. Spektrogram typowego obiektu harmonicznego (aria operowa- głos żeński solo)

2. OPIS TECHNIKI KODOWANIA

2.1. Podstawowa idea

Chwilowe widmo harmoniczne typowych dźwięków muzycznych jest stosunkowo gładkie (wolnozmiennie, w dziedzinie częstotliwości) i stabilne (wolnozmiennie w dziedzinie czasu), z wyjątkiem początkowego odcinka artykulacji dźwięku. Podstawową ideą zaproponowanego systemu kodowania obiektów jest maksymalne wykorzystanie tej cechy widma poprzez użycie dwuwymiarowej techniki kompresji, podobnej do kompresji obrazów. Jednak dźwięki naturalne nie składają się z czystych prążków harmonicznych. Często posiadają one zmienną częstotliwość podstawową i zawierają również składowe szumowe (rys. 1). Szerokości widm poszczególnych składowych tonalnych są znacząco zależne od drobnych modulacji amplitud i częstotliwości chwilowych. Cechy te znacząco wpływają na percepcję naturalności dźwięków.

* Praca finansowana ze środków publicznych na badania naukowe

Podstawowym problemem przy potraktowaniu obiektu harmonicznego jako sygnału dwuwymiarowego jest zatem wyodrębnienie informacji o zmiennych w czasie natężeniach i fazach poszczególnych prążków widmowych i uniezależnienie jej od zmiennej częstotliwości. Uzyskanie wysokiej wierności reprezentacji wymaga też zachowania mikro-modulacji składowych harmonicznym zawartych w sygnale.

W prezentowanym rozwiązaniu proponuje się odzielenie zgrubnej informacji o obwiedni widmowej sygnału od informacji opisującej wspomniane mikro-modulacje. Sygnał jest więc reprezentowany poprzez dwa zbiory zakodowanych danych. Technika ta jest rozwinięciem hybrydowej techniki kompresji opisywanej w [10] i opiera się na wykorzystaniu zalet dwóch klasycznych metod:

- modelowania sinusoidalnego, które pozwala na adaptację kodera do niestacjonarności sygnału poprzez mechanizmy śledzenia,
- dwuwymiarowego kodowania transformatowego, które pozwala na maksymalne wykorzystanie wolnozmiennego charakteru widma harmonicznego.

W celu identyfikacji poszczególnych składowych wielotonu, określenia ich zmiennych w czasie częstotliwości $f_k(t)=kf_0(t)$, amplitud $A_k(t)$ oraz fazy φ_k , proponowany koder wykorzystuje znany z literatury *harmoniczny* model sinusoidalny [11,12],

$$\hat{x}(t) = \sum_{k=1}^{K_{\max}} \tilde{A}_k(t) \cos\left(k2\pi \int_0^t \tilde{f}_0(\tau) d\tau + \varphi_k\right) \quad (1)$$

W tym i następnym wzorach symbole z tyldą oznaczają przebiegi zmiennych parametrów interpolowane na podstawie ich wartości chwilowych estymowanych w odstępach czasu jednej ramki (np. co około 12-20ms).

Warto nadmienić, że istnieje już technika parametrycznego kodowania dźwięku oparta na zastosowaniu modelu sinusoidalnego, którą można wykorzystać do kompresji indywidualnych obiektów harmonicznym [4,5]. Jednakże jakość dźwięku oferowana przez kodek standardu MPEG-4 SSC jest stosunkowo niska.

2.2. Reprezentacja obiektu

W proponowanym systemie parametry f_0 , A_k oraz φ_k otrzymane z modelu sinusoidalnego nie służą bezpośrednio do zakodowania sygnału, tak, jak we wspomnianej technice MPEG-4. Dla uzyskania maksymalnie wiernej reprezentacji potrzebna jest parametryzacja składowych dokładniejsza niż 1 wartość na ramkę. Jest to konieczne, by opisać mikro-modulacje amplitud i częstotliwości.

W tym celu przebiegi czasowe składowych harmonicznym są wyodrębniane z obiektu metodą analizy heterodynowej, gdzie częstotliwości poszczególnych trajektorii z modelu sinusoidalnego (kf_0) służą jedynie za częstotliwości nośne w demodulacji jednowstęgowej. Dzięki temu wyeliminowane są również błędy estymacji sygnału niestacjonarnego spowodowane analizą czasowo-częstotliwościową opartą na FFT w blokach próbek. Te błędy to jedno z istotnych źródeł niedokładności modelu sinusoidalnego.

Produkty demodulacji poszczególnych składowych harmonicznym stanowią dolnopasmową reprezentację tzw. obwiedni zespolonej tych składowych. Ponieważ

taka obwiednia jest wąskopasmowa, możliwa jest daleko idąca redukcja częstotliwości próbkowania. Próbki zdecymowanych produktów demodulacji zebrane w dwuwymiarowej tablicy opisują tzw. *obwiednię harmoniczną* dźwięku (rys. 5). Następnie tablica taka poddana zostaje dwuwymiarowemu kodowaniu transformatowemu.

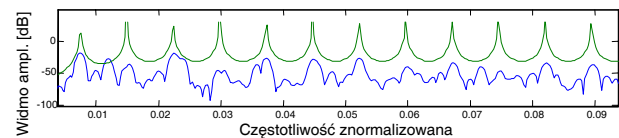
Drugim zbiorem danych proponowanej reprezentacji jest tzw. *sygnał prototypowy*, który niesie informacje o częstotliwości podstawowej oraz o resztkowej obwiedni czasowej sygnału (mikro-modulacji AM), zakodowane niezależnie poprzez modulację częstotliwości oraz modulację amplitudy. Sygnał prototypowy ma zatem postać podwójnie zmodulowanego (AM-FM) przebiegu sinusoidalnego,

$$x_p(t) = A_0(t) \cos\left(2\pi \int_0^t f_0(\tau) d\tau\right), \quad (2)$$

jest więc prostym sygnałem jednokomponentowym, którego efektywne zakodowanie jest stosunkowo łatwe. Również tutaj zastosowano kompresję transformatową.

2.3. Harmoniczny model sinusoidalny

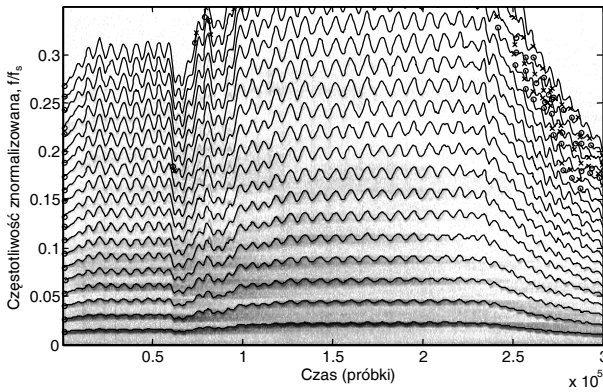
Wykorzystany w koderze blok modelowania sinusoidalnego służy do identyfikacji składowych harmonicznym dźwięku oraz estymacji ich częstotliwości, która może zmieniać się w czasie. Szczegółowy opis tej techniki znajduje się w pracy [12]. Algorytm działa krokowo, analizując oryginalny sygnał z wykorzystaniem FFT w kolejnych blokach próbek pobieranych z zakładką. Dla każdej ramki szacowany jest zbiór potencjalnych częstotliwości, które mogą być częstotliwością podstawową sygnału. Następnie w widmie amplitudowym wyszukiwane są maksima lokalne w okolicach wielokrotności przypuszczalnej częstotliwości podstawowej (rys. 2).



Rys. 2. Przykład analizy dla pojedynczej ramki sygnału: widmo amplitudowe oraz funkcja wiarygodności detekcji prążka harmonicznego (powyżej)

Dla każdego z maksimów estymowana jest związana z nim fizyczna częstotliwość prążka oraz amplituda. Zestawy danych zawierających częstotliwość podstawową oraz informacje o prążkach są porównywane z danymi otrzymanymi w poprzedniej i następnej ramce. Dane łączone są w sekwencje (trajektorie) przez zaawansowany algorytm śledzący, który znajduje najlepsze dopasowanie całych ciągów harmonicznym. Kryterium dopasowania jest minimalny błąd wielokanałowej predykcji adaptacyjnej danych parametrów. Niezależnej predykcji podlegają zbiory częstotliwości oraz amplitud, przy czym adaptacja współczynników predyktora jest oparta na kryterium najmniejszych kwadratów.

Ostatecznym wynikiem analizy jest sekwencja wartości częstotliwości podstawowej oraz dane o wszystkich wykrytych składowych harmonicznym tworzące tzw. harmoniczne trajektorie sinusoidalne (rys. 3).



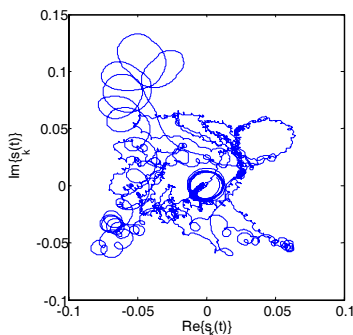
Rys. 3. Wynik harmonicznej analizy sinusoidalnej dla obiektu z rys. 1: trajektorie sinusoidalne (--) na tle spektrogramu sygnału

2.4. Analiza heterodynowa

Celem analizy heterodynowej jest dokładna ekstrakcja sygnałów obwiedni zespolonej komponentów sygnału związanych z poszczególnymi składowymi harmonicznymi. Procedura ekstrakcji wykorzystuje częstotliwości trajektorii sinusoidalnych z modelu sinusoidalnego. Ponieważ dane te są dostępne jedynie z rozdzielczością ramek (czyli raz na 12-20ms), są one interpolowane przy pomocy funkcji sklepanych (zastosowano tu wielomian interpolacyjny Hermite'a). Analiza heterodynowa (3) bazuje na demodulacji jednowstęgowej i wykorzystuje zespolone nośne harmoniczne.

$$s_k(t) = \left[x(t) \exp\left(-j2\pi \int_0^t \tilde{f}_k(\tau) d\tau\right) \right] * h_{LP}(t), \quad (3)$$

$\tilde{f}_k(t)$ oznacza tutaj zinterpolowaną częstotliwość k-tej składowej harmonicznej z modelu sinusoidalnego, $h_{LP}(t)$ jest odpowiedzią impulsową filtra dolnoprzepustowego, a $*$ jest symbolem operacji splotu liniowego. Zastosowany filtr dolnoprzepustowy o zerowej fazie (co jest krytyczne dla prawidłowego funkcjonowania systemu) złożony jest z dwóch filtrów IIR o charakterystyce Butterwortha 4-go rzędu, zaaplikowanych do fragmentu sygnału w dwóch kierunkach upływu czasu. Częstotliwość graniczna filtra jest kompromisem między tłumieniem niepożądanych produktów demodulacji a czasem narastania odpowiedzi filtra, który jest istotny z uwagi na reakcję na transjenty obecne w niektórych dźwiękach. W prototypowej implementacji zastosowano automatyczną regułę dobierającą częstotliwość graniczną filtra jako połowę częstotliwości podstawowej f_0 (a więc i odległości pomiędzy kolejnymi prążkami).



Rys. 4. Przykładowy przebieg obwiedni zespolonej jednej składowej harmonicznej (3) na płaszczyźnie Re-Im

Obwiednia zespolona $s_k(t)$ (rys. 4) odzwierciedla zmienną w czasie amplitudę danej składowej harmonicznej, zwykle zawiera ona też marginalną składową oscylacyjną (modulację kąta) wynikającą z niedokładności estymacji częstotliwości prążka w modelu sinusoidalnym. Z drugiej strony, analizowany sygnał $x(t)$ może posiadać więcej niż jedną składową sinusoidalną w okolicach częstotliwości harmonicznej f_k (np. dźwięk fortepianu złożony z wibracji nawet trzech bardzo blisko nastrojonych strun), co przekłada się na głęboką modulację amplitudową $s_k(t)$ będącą efektem zdudnienia.

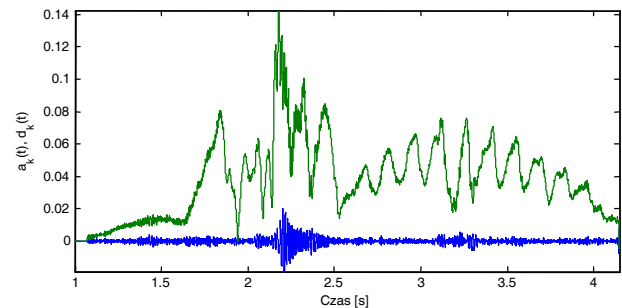
Dla efektywnej kompresji i z uwagi na znikomą słyszalność powolnych zmian fazy korzystne jest zredukowanie amplitudy zespolonej sygnału do wartości rzeczywistych. W tym celu dla każdego sygnału $s_k(t)$ wyznaczany jest wolnozmienny komponent rotacyjny,

$$\vartheta_k(t) = \exp\left(-j2\pi \int_0^t (\arg\{s_k(\tau)\} * h_{LP}(\tau)) d\tau\right) \quad (4)$$

W definicji (4) wygładzająca filtracja dolnoprzepustowa poprzedzona jest operacją rozwinięcia argumentu w celu zachowania ciągłości. Efektem rotacji

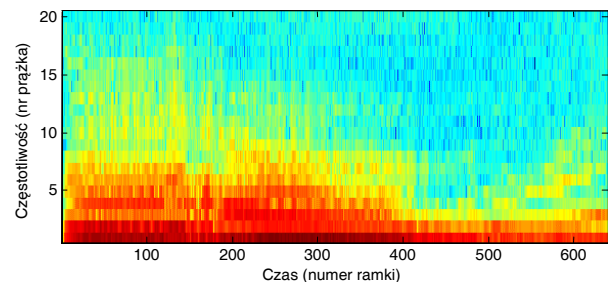
$$a_k(t) = \text{Re}\{s_k(t) \vartheta_k(t)\}. \quad (5)$$

jest skupienie zmienności amplitudy wzdłuż osi rzeczywistej, co pozwala odrzucić szczytkową część urojonej (rys. 5). Zaletą takiego rozwiązania w stosunku do nieliniowej operacji modułu jest uniknięcie zniekształceń reprezentacji sygnału dla sytuacji, gdy $a_k(t)$ przyjmuje wartości ujemne (wspomniana modulacja obwiedni).



Rys. 5. Przykładowy przebieg amplitudy chwilowej $a_k(t)$ otrzymanej z rotacji (5) obwiedni zespolonej z rys. 4, oraz przebieg resztkowej modulacji amplitudy $d_k(t)$ uzyskany z (6)

Ostateczną postacią obwiedni harmonicznej sygnału jest macierz złożona z próbek zdecydowanych wartości $a_k(t)$ (rys. 6). Dane te poddawane są kompresji stratnej.



Rys. 6. Reprezentacja przykładowej obwiedni harmonicznej dla sygnału z rys. 1, w postaci obrazu w skali [dB]

Stopień decymacji ma wpływ na dokładność, z jaką zmienność proporcji natężeń składowych harmonicz-

nych jest odwzorowana w tablicy i został dobrany eksperymentalnie na 1:1000. Oznacza to, że zmiany te (a więc zmiany barwy) są aktualizowane co około 25ms.

2.5. Sygnał prototypowy

W proponowanej technice kodowania sygnał prototypowy pełni funkcję nośnika równocześnie dwóch rodzajów informacji: o częstotliwości podstawowej obiektu harmonicznego oraz o mikro-modulacji amplitud wszystkich składowych harmonicznym. Ta druga informacja pozwala z dużą wiernością zakodować obiekty zawierające transjenty oraz składniki pseudo-szumowe (obecne w dźwiękach instrumentów dętych i w śpiewie). Ponieważ duże znaczenie dla percepcji mają *koherentne* modulacje składowych harmonicznym [13], przyjęto, że w amplitudzie sygnału prototypowego zakodowany zostanie szybkozmenny komponent wspólny dla wszystkich $a_k(t)$ stanowiący poprawkę do indywidualnych wolnozmiennych amplitud zakodowanych w obwiedni harmonicznym. Składnik ten wyznaczony jest za pomocą analizy składowych głównych (PCA). W tym celu obliczona jest różnica pomiędzy kolejnymi amplitudami $a_k(t)$ a zinterpolowaną obwiednią harmoniczną (która jest ich wolnozmiennym przybliżeniem),

$$d_k(t) = a_k(t) - \tilde{a}_k(t), \quad k = 1 \dots K, \quad (6)$$

oraz konstruowana jest macierz próbek $\mathbf{d} = [d_k(t)]_{K \times N}$. Przykładowy przebieg takiego sygnału pokazuje rys. 5.

Analiza PCA wymaga wyznaczenia wartości własnych λ_k macierzy autokowariancji $\mathbf{R}_{dd} = \mathbf{d}\mathbf{d}^T/N$ i skonstruowania macierzy \mathbf{W} , której wierszami są wektory własne związane z λ_k . Wektor własny \mathbf{W}_1 związany z największą wartością własną określa najmniej stratne przekształcenie redukujące wymiar macierzy \mathbf{d} . Rezultatem transformacji

$$\underline{a}_0 = \mathbf{W}_1 \mathbf{d} \quad (7)$$

jest wektor próbek $\underline{a}_0 = [a_0(t)]_{1 \times N}$, który stanowi komponent AM docelowego sygnału prototypowego. Z kolei częstotliwość chwilową (komponent FM) tego sygnału definiuje zinterpolowana częstotliwość podstawową f_0 uzyskana z modelu sinusoidalnego. Aby umożliwić bezproblemową rekonstrukcję dźwięku, a więc również odtworzenie parametrów sygnału prototypowego, konieczne jest złożenie obu modulacji pozwalające uniknąć zjawiska przemodulowania. Dlatego przebieg $a_0(t)$ uzupełniony jest o wartość stałą gwarantującą nieujemną amplitudę chwilową,

$$A_0(t) = a_0(t) + c \quad : \quad c > \max_{n: a_0(n) < 0} |a_0(n)| \quad (8)$$

Ostateczną postać sygnału prototypowego definiuje (2). Sygnał ten poddawany jest kompresji stratnej.

2.6. Kompresja danych

W celu efektywnego zakodowania danych obwiedni harmonicznym oraz prototypu wykorzystano klasyczne metody kompresji bazujące na przekształceniu kosinusowym. Dla obwiedni harmonicznym zastosowano dwuwymiarowy algorytm wzorowany na technice JPEG. Macierz próbek obwiedni poddana zostaje przekształceniu DCT-2D (w jednym bloku), współczynniki transformaty są uporządkowane zygzakowo i kwantowane, a następnie zakodowane kodem run-length. Ze względu na ty-

powo dużą gładkość obwiedni, efektywność takiego schematu jest bardzo wysoka, pozwalając zakodować typowy zbiór z równoważną prędkością bitową około 760-820b/s bez słyszalnego pogorszenia jakości dźwięku.

Dla sygnału prototypowego zastosowano klasyczne kodowanie wzorowane na technice AAC, a zatem bazujące na przekształceniu MDCT [1]. Współczynniki transformaty są kwantowane nierównomiernie i kodowane kodem Huffmana. Typowa uzyskiwana w tym schemacie równoważna prędkość bitowa mieści się w zakresie 2,8-4,2kb/s, w zależności od stacjonarności sygnału.

2.7. Rekonstrukcja sygnału

Poprawne odtworzenie sygnału fonicznego ze zdekodowanych danych wymaga odwrócenia zabiegów opisanych przez przekształcenia (3) oraz (6-8). W tym celu konieczne jest wyznaczenie parametrów chwilowych zdekodowanego sygnału prototypowego $\hat{x}_p(t)$. Wykorzystano tutaj klasyczną metodę Gabora bazującą na przekształceniu Hilberta,

$$\hat{\Phi}_p(t) = \arg\{\hat{x}_p(t) + j \mathcal{H}[\hat{x}_p(t)]\} \quad (9a)$$

$$\hat{A}_0(t) = |\hat{x}_p(t) + j \mathcal{H}[\hat{x}_p(t)]|. \quad (9b)$$

Po odjęciu stałej c , resztkowe przebiegi modulacji amplitudy indywidualnych składowych harmonicznym otrzymane są na drodze odwrotnej transformacji

$$\hat{\mathbf{d}} = \underline{G}_1 \hat{\underline{a}}_0, \quad (10)$$

gdzie \underline{G}_1 jest odpowiednią kolumną macierzy \mathbf{W}^{-1} . Przebiegi te uzupełniane są o zinterpolowane wartości ze zdekodowanej obwiedni harmonicznym,

$$\hat{a}_k(t) = \hat{d}_k(t) + \tilde{a}_k(t). \quad (11)$$

Ostateczną rekonstrukcję sygnału kończy proces modulacji

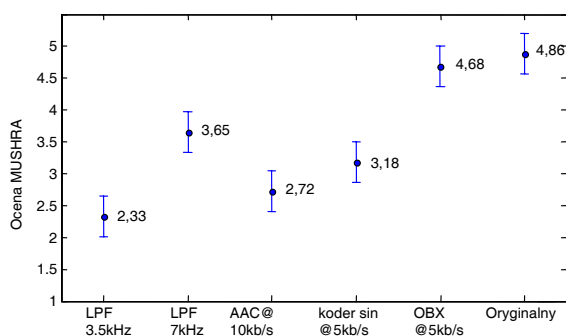
$$\hat{x}(t) = \sum_{k=1}^{K_{\max}} \hat{a}_k(t) \cos(k \hat{\Phi}_p(t)). \quad (12)$$

Odtworzony sygnał nie jest wierną repliką sygnału wejściowego. Sygnał ten nie zawiera składowych szumowych wypełniających przestrzeń w widmie pomiędzy składowymi harmonicznymi dźwięku (składowe te stanowią odrębny obiekt, który może być przesyłany przy pomocy dedykowanego koderu). Ponadto, wskutek odrzucenia wartości reprezentowanych przez $v_k(t)$, utraciona zostaje dokładna informacja o indywidualnych relacjach fazowych między harmonicznymi. Informacja ta jest nieistotna dla percepcji. Kolejnym elementem stratności jest redukcja indywidualnych modulacji amplitud wskutek transformacji (7). Również sama kompresja stratna danych obwiedni harmonicznym oraz sygnału prototypowego prowadzi do powstania błędów zależnych od prędkości bitowej. Należy podkreślić jednak, że wszystkie wymienione rodzaje stratności powodują powstawanie drobnych artefaktów w zakodowanym dźwięku, które są znacznie bardziej tolerowalne od klasycznych zniekształceń nieliniowych, utraty pasma, czy zniekształceń produkowanych przez kodery transformtowe.

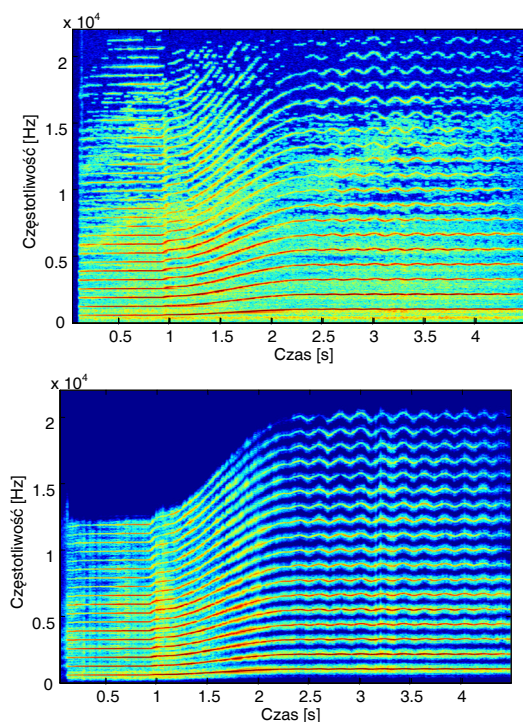
3. WYNIKI EKSPERYMENTALNE

Zaimplementowany system poddany został weryfikacji poprawności działania oraz efektywności. Prze-

proawdzono eksperymenty z różnymi sygnałami syntetycznymi (modulowane szeregi harmoniczne zakłócone szumem) uzyskując dokładność rekonstrukcji na poziomie od 38dB do 47dB (rekonstrukcja sygnału bez kompresji). Eksperymenty z udziałem dźwięków naturalnych poddano ocenie subiektywnej z zastosowaniem metodologii MUSHRA [14]. Na rysunku 7 pokazano typowe zestawienie wyniku uśrednionego testu odsłuchowego dla 7 słuchaczy i 6 fragmentów nagrań dźwięku zakodowanych jako jeden obiekt harmoniczny. Przykładowe spektrogramy pokazują rysunki 8 i 9.



Rys. 7. Uśredniony wynik testu odsłuchowego, porównanie jakości sygnału zakodowanego opisaną techniką (OBX) z jakością oferowaną przez koder sinusoidalny i koder MPEG-4 AAC-HE

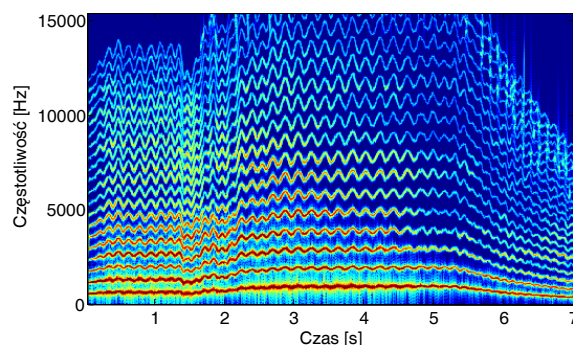


Rys. 8. Oryginalny i zrekonstruowany (bez rekonstrukcji szumu tła) obiekt harmoniczny (dźwięk skrzypiec) przy prędkości bitowej 4,85kb/s

4. PODSUMOWANIE

Na podstawie wyników przeprowadzonych testów można postawić tezę, że obiektowy schemat kompresji z zastosowaniem zaprezentowanej metody oferuje atrakcyjną alternatywę dla współczesnych technik kodowania fonii. Istotną fundamentalną zaletą przyjętej reprezentacji dwuskładnikowej jest separacja struktury makro i mikroskopowej sygnału, dzięki czemu ewentualne znie-

kształcenia mają charakter wysoko tolerowalnych zmian barwy, częstotliwości dźwięku i jego dynamiki. Pozwala to osiągnąć bardzo wysokie stopnie kompresji.



Rys. 9. Zrekonstruowany obiekt harmoniczny z rys. 1, przy prędkości bitowej 5,12kb/s

SPIS LITERATURY

- [1] T. Painter, A. Spanias, "Perceptual Coding of Digital Audio", *Proc. of the IEEE*, vol. 88, no. 4, Apr. 2000.
- [2] ISO/IEC JTC1/SC29/WG11 MPEG, "Int. Standard ISO/IEC 14496-3/Amd1, Coding of Audio-Visual Objects: Audio", 1999.
- [3] A.C. den Brinker, E. Schuijers, W. Oomen, "Parametric Coding for High-Quality Audio", *112th AES Convention*, Apr. 2002, Paper 5554.
- [4] ISO/IEC JTC1/SC29/WG11 MPEG, "Int. Standard 14496-3:2005/Dcor2, Sinusoidal Coding", 2005.
- [5] M. Bartkowiak, „Metody parametryczne we współczesnych technikach kodowania dźwięku szerokopasmowego”, KKRRiT, Poznań, 2-7 czerwca 2006.
- [6] T. Tolonen, "Object-based sound source modeling for musical signals," *109th AES Conv.*, Los Angeles, USA, 2000, paper 5174.
- [7] S.J. Welburn, M.D. Plumbley, E. Vincent, "Object-coding for resolution-free musical audio," *31st Int. AES Conf.: "New Directions In High Resolution Audio"*, London, UK, 2007.
- [8] E. Vincent, M.D. Plumbley, "Low bit-rate object coding of musical audio using Bayesian harmonic models", *IEEE Trans. on Audio, Sp., and Lang. Proc.*, vol. 15, nr 4, ss. 1273-1282, May 2007.
- [9] E. Vincent, N. Bertin, R. Badeau, "Adaptive Harmonic Spectral Decomposition for Multiple Pitch Estimation", *IEEE Trans. Audio, Sp., and Lang. Proc.*, vol.18, nr 3, March 2010, ss.528-537.
- [10] M. Bartkowiak, T. Żernicki, „Hybrydowy, parametryczno-transformatowy kodek dźwięku”, KKRRiT'08, Wrocław, 2008.
- [11] X. Serra, "A System for Sound Analysis/Transformation/Synthesis Based on Deterministic Plus Stochastic Decomposition", Raport STAN-M-58, CCRMA, Stanford University, 1989.
- [12] M. Bartkowiak, T. Żernicki, "Harmonic Sinusoidal+Noise Modeling of Audio Based on Multiple F0 Estimation", *125th AES Conv.*, 2-5 October 2008, San Francisco, USA, Paper 7510.
- [13] S. McAdams, *Spectral fusion, spectral parsing and the formation of auditory images*, rozprawa doktorska, Stanford University, 1984
- [14] ITU-R, "Method for the subjective assessment of intermediate quality level of coding systems", ITU-R, Tech. Rep. BS. 1534-1, 2003.