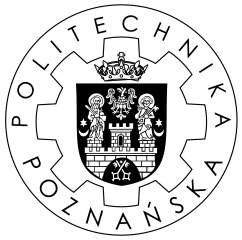


Politechnika Poznańska
Wydział Elektroniki i Telekomunikacji
Katedra Telekomunikacji Multimedialnej i Mikroelektroniki

Autoreferat rozprawy doktorskiej

Kompresja cyfrowych sygnałów fonicznych z łącznym wykorzystaniem rozszerzania widma i modelowania

Tomasz Żernicki

Promotor: prof. dr hab. inż. Marek Domański

Poznań, 2010

Spis treści autoreferatu*

* Układ rozdziałów nie odpowiada układowi rozprawy.

Rozdział 1

Wstęp

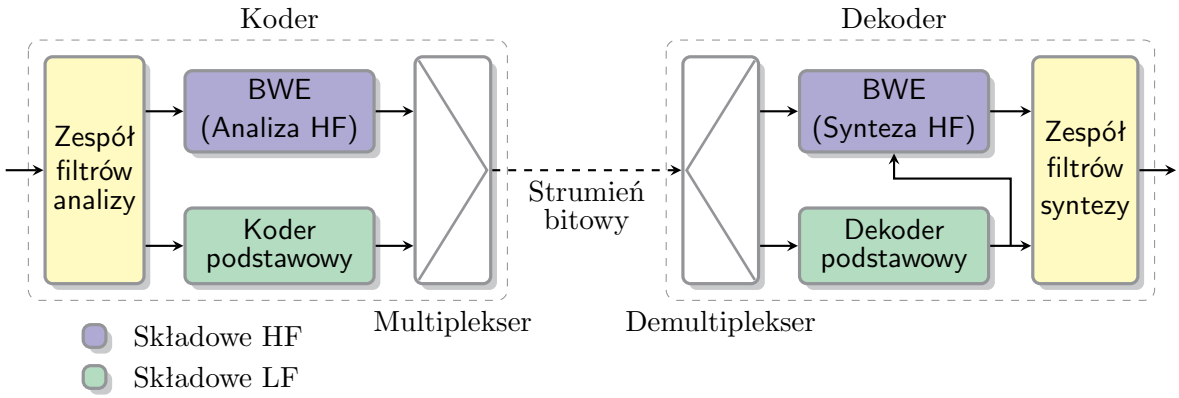
1.1. Zakres rozprawy

W pracy poruszono zagadnienia kompresji dźwięku szerokopasmowego przesyłanego z małą prędkością bitową. W szczególności praca dotyczy ulepszeń zaawansowanych technik reprezentacji i kompresji cyfrowych sygnałów dźwiękowych dla zastosowań w systemach teleinformatycznych nowych generacji, w tym w systemach bezprzewodowych. Przez zaawansowane techniki kompresji cyfrowych sygnałów dźwiękowych, rozumie się najnowsze techniki rozwinięte w XXI wieku.

Po wdrożeniu kodeka standardu MPEG-1 Layer 3 (popularne MP3) prowadzono intensywne badania naukowe mające na celu znalezienie bardziej efektywnej techniki kompresji szerokopasmowych sygnałów fonicznych. Wspomniane badania zaowocowały powstaniem kilku nowych generacji kodeków fonicznych. Każda kolejna generacja charakteryzuje się wyższą efektywnością kompresji uzyskiwaną kosztem wzrostu złożoności obliczeniowej kodeka. Jednym z kolejnych etapów rozwoju był opracowany w końcu lat 90 XX wieku kodek MPEG-4 AAC (ang. *Advanced Audio Coding*). Kodek ten jest udoskonaloną wersją kodera transformatowego, w którym ujęto dodatkowe narzędzia usprawniające efektywność kompresji, np. kształtowanie obwiedni czasowej (ang. *Temporal Noise Shaping*) lub percepcyjne zastąpienie szumu (ang. *Perceptual Noise Substitution*).

W rozprawie odniesieniem jest technika kompresji sygnałów fonicznych ujęta w standardzie kompresji dźwięku szerokopasmowego ISO/IEC 14496-3 (określanego również jako MPEG-4 część 3) uwzględniająca kompresję z rozszerzaniem widma Spectral Band Replication (SBR). SBR w języku polskim można nazwać techniką powielania pasm widmowych.

W schemacie kompresji z wykorzystaniem technik rozszerzania widma BWE (ang. *bandwidth extension*) (rysunek 1.1) koduje się sygnał o zawężonym paśmie oraz przesyła dodatkowe parametry opisujące pozostałą część sygnału (rysunek 1.2). Na tej podstawie możliwe jest zrekonstruowanie sygnału pełnopasmowego w dekoderze. W ten sposób **techniki BWE pozwalają na znaczną redukcję ilości danych w stosunku do metod**



Rysunek 1.1: Schemat kompresji z rozszerzaniem widma. W standardzie MPEG-4 AAC HE składowe małych częstotliwości (LF) kodowane są za pomocą techniki AAC (koder podstawowy), natomiast składowe dużych częstotliwości (HF) - z wykorzystaniem techniki SBR.

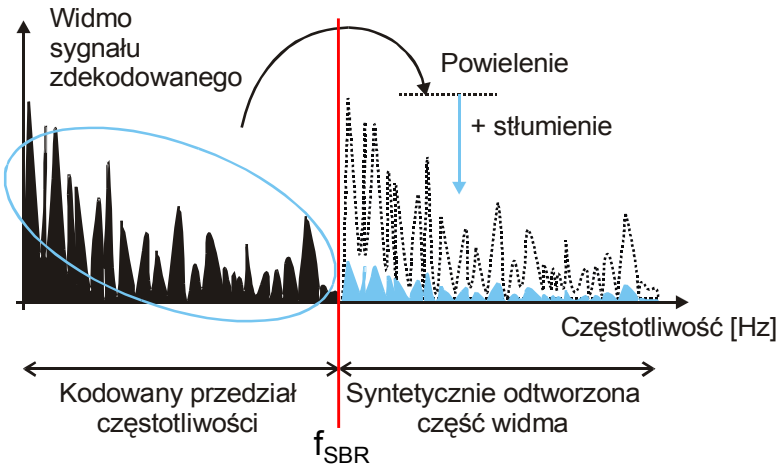
kompresji transformatowej. Zastosowanie technik BWE umożliwia konstrukcję kodeków, które przy tej samej jakości dźwięku pozwalają na dwukrotne zmniejszenie prędkości bitowej w stosunku do technik kompresji transformatowej (rysunek 1.3).

Przykładem techniki wykorzystującej schemat kompresji z rozszerzaniem widma jest MPEG-4 AAC High Efficiency (MPEG-4 AAC HE), która powstała z połączenia algorytmu SBR z techniką kodowania transformatowego Advanced Audio Coding (AAC). Wspomniana technika znalazła zastosowania w cyfrowym radiu, np. Digital Radio Mondiale (DRM+), HD Radio oraz XM Satellite Radio. Należy podkreślić, iż algorytm MPEG-4 AAC HE jest odzwierciedleniem aktualnego stanu techniki w zakresie kodowania sygnałów fonicznych. Technologia ta nadal jest przedmiotem wielu prac badawczych koncentrujących się na zapewnieniu wysokiej postrzeganej jakości zdekodowanego dźwięku szerokopasmowego i mowy, dla małych prędkości bitowych. **Również niniejsza rozprawa dotyczy schematu kompresji sygnałów fonicznych z rozszerzaniem widma**. Przy czym w pracy skoncentrowano się na usprawnieniu schematu kodowania przedstawionego na rysunku 1.3c.

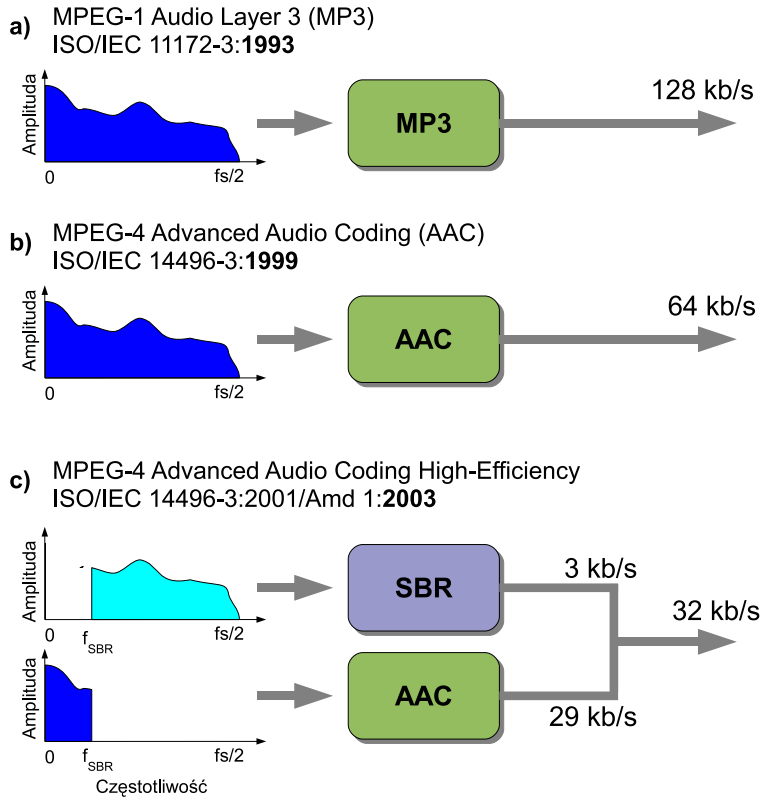
Schemat kompresji z rozszerzaniem widma wykorzystywany jest w zakresie kodowania szerokopasmowych sygnałów fonicznych jak i sygnałów mowy. W rozprawie skupiono się na technikach BWE dotyczących szerokopasmowych sygnałów fonicznych.

Terminem małe częstotliwości LF (ang. *low frequency*) w rozprawie nazwano składowe sygnału w zakresie od 0 do f_{SBR}, gdzie f_{SBR} jest granicą podziału pomiędzy zakresem działania *kodera podstawowego* (ang. *core encoder*) oraz kodera BWE. Składowe LF kodowane są z wykorzystaniem kodera podstawowego. Terminem duże częstotliwości HF (ang. *high frequency*) nazywać będziemy składowe sygnału powyżej f_{SBR}.

Opublikowanie standardu kompresji dźwięku szerokopasmowego MPEG-4 AAC HE (zawierającego SBR) w roku 2003, spowodowało duże zainteresowanie technikami BWE, w tym SBR. Powstało wiele artykułów opisujących metody rozszerzania widma. Jednak przeważnie są to techniki oferujące niską jakość dźwięku dla prędkości bitowych poniżej



Rysunek 1.2: Idea rekonstrukcji składowych dużych częstotliwości w technice SBR, jako przykład techniki rozszerzania widma.



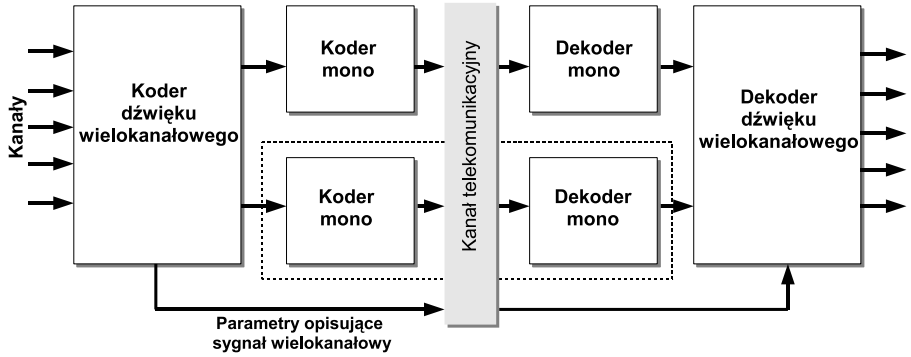
Rysunek 1.3: Porównanie wielkości generowanego strumienia dla tej samej jakości dekodowanego sygnału dla technik: a) MPEG-1 Layer 3, b) MPEG-4 AAC oraz c) MPEG-4 AAC HE.

30 kb/s. Technika SBR nie zawsze jest w stanie poprawnie odtworzyć brakujące składowe HF, nawet przy zwiększeniu puli bitów. W widmie składowych dużych częstotliwości mogą występować składniki, których nie da się odtworzyć poprzez mechanizm przesunięcia fragmentów widma w dziedzinie częstotliwości (rysunek 1.2) (np. źródła dźwięku o częstotliwości podstawowej powyżej częstotliwości podziału f_{SBR}). Dodatkowo istotnym problemem jest zachowanie harmonicznej struktury składników tonalnych, których

częstotliwości powinny być wielokrotnościami częstotliwości tonów podstawowych. W wielu technikach pojawia się również problem związany z poprawną reprezentacją szybkich zmian częstotliwości (efekty glissando i vibrato) oraz ich odtworzeniem.

Brakuje technik rozszerzania widma, które oferują wysoką (porównywalną z oryginałem) jakość zdekodowanego dźwięku szerokopasmowego dla prędkości poniżej 30 kb/s. Dlatego opracowanie efektywnej techniki lub wskazanie dalszych kierunków badań w tym zakresie będzie miało duży wkład w rozwój nauki. Szczególnie istotne jest poszukiwanie nowych, parametrycznych reprezentacji składowych HF, które umożliwią efektywną kompresję sygnałów fonicznych. **Rozprawa również podejmuje tematykę rozszerzania widma dźwięku, kodowania parametrycznego oraz modelowania sinusoidalno-szumowego w zastosowaniu do kompresji dźwięku szerokopasmowego. Praca w szczególności zajmuje się problemem poprawy reprezentacji struktury tonalnej składowych HF**.

Praca dotyczy sygnałów monofonicznych, ale proponowane narzędzia można również zastosować w połączeniu z kodowaniem wielokanałowym. W procesie kodowania dźwięku wielokanałowego sygnały pochodzące z wielu źródeł przekształcane są do postaci jedno lub dwu kanałowej i poddawane kompresji w koderze sygnału mono. Parametryczne dane niezbedne do rekonstrukcji pozostałych kanałów przesyłane sa jako strumień dodatkowych parametrów o niewielkiej prędkości bitowej (rysunek 1.4). W dalszym ciągu istotne jest poszukiwanie efektywnych technik kompresji sygnałów monofonicznych, czego dotyczy również niniejsza rozprawa.



Rysunek 1.4: Typowa struktura kodeka wielokanałowego. Przykładowo, dla MPEG-D Surround całkowita prędkość bitowa to 48 kb/s z czego 8 kb/s to dodatkowe parametry opisujące sygnał wielokanałowy. Linią przerywana zaznaczono obszar tematyczny pracy - kodowanie sygnału monofonicznego.

1.2. Cele i teza pracy

Celem pracy jest **stworzenie efektywnej metody kompresji wykorzystującej techniki rozszerzania widma oraz modelowania sinusoidalno-szumowego, które łącznie pozwolą uzyskać dobrą subiektywną jakość zrekonstruowanego sygnału przy małych prędkościach bitowych**.

Dotychczas opisane w literaturze prace związane z kompresją dźwięku szerokopasmowego z rozszerzaniem widma dotyczą głównie techniki SBR oraz jej modyfikacji. Technika SBR jest powszechnie uznanym standardem, jak również dostępna jest jej referencyjna implementacja. Z tych powodów została wybrana do badań przeprowadzonych przez autora rozprawy.

Dla techniki SBR otwartym problemem jest poprawna rekonstrukcja tonalnych składowych dużych częstotliwości, a zwłaszcza odtworzenie szybkich zmian w dziedzinie czasu i częstotliwości. W rozprawie pokazano, jak rozwiązać powyższy problem poprzez zastosowanie modelowania sinusoidalno-szumowego.

Teza rozprawy jest następująca:

> *Można stworzyć efektywną metodę kompresji cyfrowych sygnałów fonicznych, która wykorzystuje w tym samym paśmie częstotliwości rozszerzanie widma oraz modelowanie sinusoidalno-szumowe. Dla prędkości bitowych mniejszych od 30 kb/s, zastosowanie nowej metody prowadzi do uzyskania sygnału fonicznego o jakości lepszej niż przy zastosowaniu techniki MPEG-4 AAC HE. W szczególności poprawie ulegnie odtwarzanie składowych tonalnych dużych częstotliwości.*

1.3. Przegląd pracy

Rozprawa składa się z sześciu rozdziałów, które łącznie wchodzą w skład dwóch podstawowych części pracy. Pierwsza część dotyczy studiów literaturowych na temat zaawansowanych technik kompresji sygnałów fonicznych, ze szczególnym uwzględnieniem metod rozszerzania widma oraz modelowania sinusoidalno-szumowego. W tej części pracy zawarte są również wyniki badań dotyczące efektywności kompresji techniki MPEG-4 AAC HE, uzyskanych przez autora rozprawy. Druga część pracy zawiera opis nowych, autorskich algorytmów kompresji sygnałów fonicznych oraz wyniki przeprowadzonych eksperymentów.

Rozdział 2 :

- Przegląd wybranych zagadnień z dziedziny modelowania sinusoidalno-szumowego.
- Klasyfikacja technik rozszerzania widma oraz szczegółowy opis metody SBR wchodzącej w skład standardu MPEG-4 AAC HE.
- Określenie głównych problemów pojawiających się przy implementacji technik BWE, tj. poprawnej rekonstrukcji składowych tonalnych.

Rozdział 3 :

- Opis oryginalnej, zaproponowanej przez autora, metody rozszerzania widma bazującej na modelowaniu sinusoidalno-szumowym, która może stanowić uzupełnienie aktualnych technik BWE, zwiększając ich efektywność kompresji.
- Opis sposobu zaadoptowania nowej techniki do aktualnego standardu kodowania MPEG-4 AAC HE.

- Spektrogramy przykładowych zdekodowanych sygnałów.

Rozdział 4 :

- Opis zaproponowanego przez autora rozprawy nowatorskiego podejścia do rekonstrukcji składowych HF, w którym składowe tonalne reprezentowane są przez sygnały wąskopasmowe.
- Opis procesu kodowania oraz rekonstrukcji składowych HF na podstawie sygnału prototypowego uzyskiwanego w dekoderze.

Rozdział 5 :

- Opis i analiza wyników otrzymanych podczas testów odsłuchowych.
- Porównanie technik zaproponowanych w rozprawie ze standardową techniką MPEG-4 AAC HE.

Rozdział 6 :

- Najważniejsze wnioski.
- Oryginalne osiągnięcia.
- Wskazanie kierunku dalszych badań w tej dziedzinie.

Rozdział 2

Zasadnicze osiągnięcia rozprawy

2.1. Metodologia badań

Technika SBR została dokładnie zbadana pod kątem poprawnej rekonstrukcji składowych harmonicznych dużych częstotliwości. Wyniki własnych badań przedstawionych w rozprawie, potwierdzają znane z literatury trudności dotyczące poprawnej rekonstrukcji składowych harmonicznych HF. Przeprowadzono również dodatkowe eksperymenty, których celem było pokazanie natury powstawania zniekształceń. Analiza wyników pozwoliła zaproponować dwie nowe techniki rozszerzania widma, w których:

- zastosowano łącznie, w tym samym paśmie częstotliwości modelowanie sinusoidalne oraz technikę SBR. Proponowana technika stanowi uzupełnienie techniki MPEG-4 AAC HE.

- zastosowano skalowanie częstotliwości składowych tonalnych z pasma LF w celu odtworzenia brakujących wyższych składowych harmonicznych. Składowe tonalne reprezentowane są jako sygnały wąskopasmowe. Dodatkowo rekonstrukcję składowych HF uzupełniono modelowaniem szumu. Jest to alternatywne rozwiązaniem dla techniki SBR.

Ponieważ nie ma metody analitycznej pozwalającej określić efektywność proponowanych rozwiązań, w rozprawie zostały one poddane badaniom eksperymentalnym. W celu ich przeprowadzenia przygotowano modele kodeków związane z nowymi technikami. Jednym z elementów pracy była implementacja modelu sinusoidalno-szumowego, a następnie zbadanie możliwości zastosowania modelowania sinusoidalnego wraz z techniką rozszerzania widma.

Celem badań eksperymentalnych było określenie efektywności kompresji kodeków, w których wykorzystano proponowane w rozprawie techniki BWE.

Podczas prac wykorzystano najnowsze implementacje techniki MPEG-4 AAC HE oraz parametrycznej reprezentacji sygnałów fonicznych:

- Implementacja kodeka MPEG-4 AAC HE opracowana przez konsorcjum 3rd Generation Partnership Project (3GPP)*.
- Implementacja kodeka MPEG-4 SSC (ang. *Sinusoidal Coding*) opracowana przez firmę Philips oraz wybrana przez MPEG na wzorcową implementację techniki kodowania sinusoidalno-szumowego.
- Autorska implementacja kodera sinusoidalno-szumowego w środowisku Matlab.

Spośród dostępnych implementacji kodeka SBR, autor wybrał implementację 3GPP, którego kod źródłowy jest publicznie dostępny. Dzięki temu możliwa była analiza algorytmów kodowania oraz ich modyfikacja w celach eksperymentalnych. Kod źródłowy algorytmu SBR jest na bieżąco aktualizowany przez 3GPP i Coding Technologies. Implementacja kodera MPEG-4 AAC HE opublikowana przez konsorcjum 3GPP uznawana jest jako jedna z najlepszych w tej klasie, co potwierdzają testy odsłuchowe oraz opinie twórców standardu MPEG-4 AAC HE.

Autor rozprawy zaimplementował narzędzie modelowania sinusoidalnego oraz kodek sinusoidalny, którego efektywność została porównana z kodekiem MPEG-4 SSC. Badania eksperymentalne wykazały, że powyższa autorska implementacja umożliwia dekodowanie sygnału z jakością wyższą niż kodek SSC, dla tych samych prędkości bitowych.

Wymienione wyżej implementacje posłużyły do zbudowania modeli dwóch kodeków implementujących oryginalne metody kompresji zaproponowane w rozprawie. Autor zrealizował dwa pełne kodeki, czyli dla obu metod powstały zarówno koder jak i dekoder. Dzięki temu możliwe było wszechstronne zbadanie właściwości zaproponowanych technik kompresji.

Krytycznym elementem przeprowadzonych badań jest wybór odpowiednich sygnałów fonicznych. W tym celu wykorzystano sekwencje testowe pochodzące z płyty EBU SQAM (ang. *Sound Quality Assessment Material*), bazy danych Uniwersytetu Iowa oraz sekwencje pochodzące z zestawu testowego zaproponowanego przez komitet standaryzacyjny MPEG. Dodatkowo wykorzystano nagrania, które uwypuklają problemy zniekształceń wprowadzanych przez technik rozszerzania widma. Nagrania pochodzące ze wspomnianych źródeł zostały zarejestrowane w profesjonalnych studiach nagraniowych. Wszystkie nagrania są sygnałami monofonicznymi próbkowanymi z częstotliwością 44100 Hz. W rozprawie nie zostały poruszane zagadnienia związane z kodowaniem dźwięku wielokanałowego.

W trakcie badań skonfrontowano efektywność znanych z literatury technik rekonstrukcji składowych HF z zaproponowanymi technikami. Zbadano również możliwość rozszerzenia aktualnego standardu kodowania dźwieku szerokopasmowego MPEG-4 AAC HE o techniki zaproponowane w pracy doktorskiej. W tym celu przeprowadzono szereg testów

*3GPP jest wspólnym projektem kilku organizacji standaryzacyjnych mający na celu rozwój systemów telefonii komórkowej trzeciej generacji 3G. Organizacje uczestniczące w projekcie to: ARIB, CWTS, ETSI, T1, TTA, TTC.

subiektywnych według zaleceń ITU-R/EBU BS.1534, w których zbadano różne warianty proponowanych rozwiązań oraz ich wpływ na jakość zdekodowanego sygnału.

Wykorzystana podczas badań miara subiektywna ITU-R/EBU BS.1534 MUSHRA (ang. *MUltiple Stimuli with Hidden Reference and Anchor*) stosowana jest dla średnich i dużych zniekształceń sygnału. Metodologia MUSHRA definiuje szereg warunków związanych z wykonaniem testów subiektywnych, takich, jak akustyka pomieszczenia, w którym wykonywane są testy odsłuchowe, parametry techniczne sprzętu nagłaśniającego, dobór materiałów testowych oraz sam przebieg eksperymentu. Każdy słuchacz musi przejść wstępną fazę treningową podczas której zostaje zaznajomiony z procesem testów odsłuchowych.

Podczas przygotowania rozprawy doktorskiej przeprowadzono wiele sesji odsłuchowych, mających na celu sprawdzenie różnych parametrów badanych systemów. Dlatego **łączny czas wszystkich sesji odsłuchowych wynosił kilkaset godzin**.

W celu weryfikacji subiektywnych badań eksperymentalnych, przeprowadzono dodatkowe badania, posługując się obiektywną miarą zniekształceń - logarytmiczna odległość widmowa (ang. *log-spectral distance*) (LSD).

2.2. Oryginalne techniki rozszerzania widma

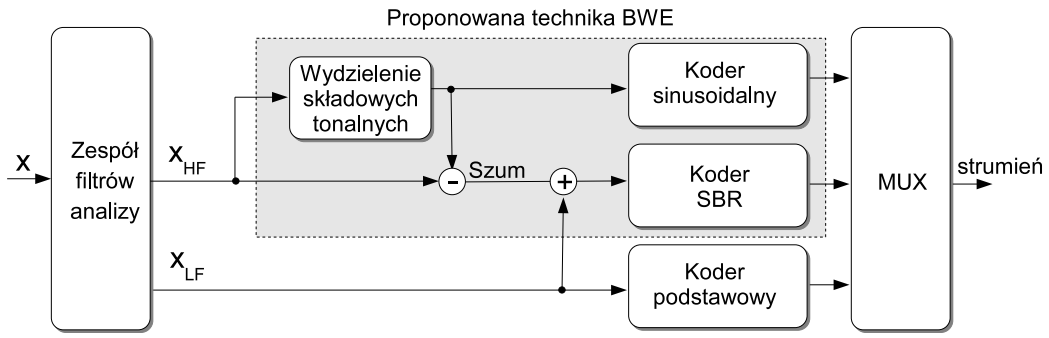
Autor rozprawy opracował dwie nowe techniki rozszerzania widma, które w połączeniu z istniejącymi standardami kodowania dźwięku szerokopasmowego umożliwiają wyższą efektywność kompresji przy tej samej docelowej prędkości bitowej.

2.2.1. Metoda rozszerzania widma wykorzystująca modelowanie sinusoidalne

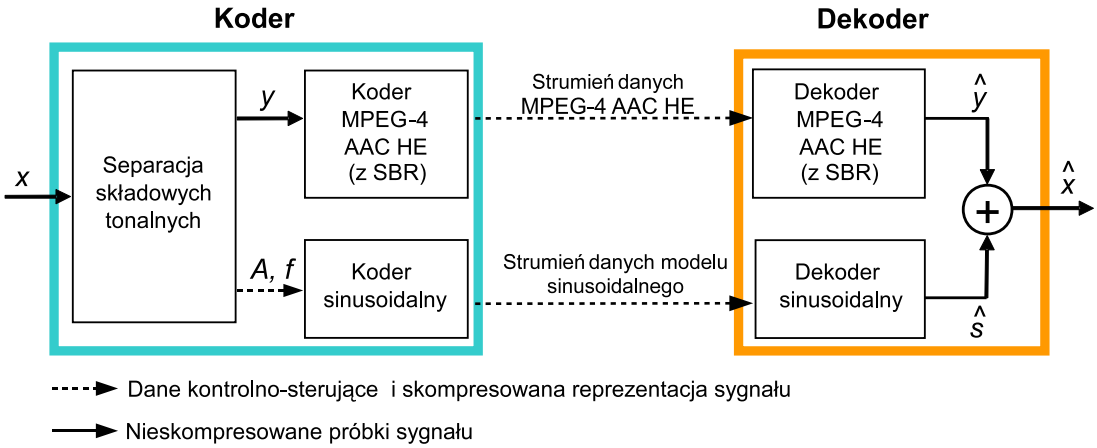
Poprzez łączne zastosowanie techniki SBR oraz modelowania sinusoidalnego uzyskano technikę efektywnej kompresji, umożliwiającą dokładniejszą reprezentację składowych tonalnych HF w stosunku do MPEG-4 AAC HE.

Podstawową ideą proponowanego rozwiązania jest zastosowanie odrębnych narzędzi do rekonstrukcji składowych tonalnych oraz szumu. Na tej podstawie, schemat kompresji z rozszerzaniem widma (rysunek 1.1) został uzupełniony o dodatkowy blok kodera sinusoidalnego, zgodnie z rysunkiem 2.1. Tego typu podejście jest uniwersalne i możliwe w połączeniu z dowolną techniką BWE. Dla potrzeb badań eksperymentalnych, proponowana technika została zaimplementowana jako opcjonalne rozszerzenie standardu MPEG-4 AAC HE.

Celem proponowanej metody jest zachowanie poprawnych zależności pomiędzy harmonicznymi w sygnale oraz zachowanie szybkozmienne składowe sinusoidalne dużych częstotliwości. Granicę pomiędzy pasmem małych i dużych częstotliwości wyznacza częstotliwość podziału f_{SBR}, określająca równocześnie zakres częstotliwości, powyżej których wykorzystywana jest technika SBR. Sposób wykorzystania zaproponowanej techniki wraz kodekiem MPEG-4 AAC HE pokazuje rysunek 2.2.



Rysunek 2.1: Ogólny schemat proponowanej techniki.



Rysunek 2.2: Rozszerzenie techniki MPEG-4 AAC HE o dodatkowy blok modelowania sinusoidalnego.

Proponowana technika wykorzystuje proces wyboru oraz separacji składowych tonalnych. Przez separację, w tym wypadku, rozumie się wyodrębnienie oraz usunięcie składowych tonalnych HF z sygnału oryginalnego. W efekcie, składowe tonalne podlegają kompresji z zastosowaniem wyspecjalizowanego kodeka sinusoidalnego. Natomiast kodek MPEG-4 AAC HE koduje sygnał y, z którego usunięto składowe tonalne HF.

W proponowanym rozwiązaniu koder wydziela z sygnału wejściowego x składowe tonalne HF. Są to składowe, dla których technika SBR często wprowadza zniekształcenia. Składowe LF sygnału y kodowane są w koderze podstawowym (AAC), natomiast mające charakter szumu składowe HF kodowane są za pomocą techniki replikacji widma SBR w standardowy sposób. Składowe HF sygnału y w dalszym ciągu mogą zawierać składowe tonalne.

Odseparowane składowe tonalne przesyłane są do kodera sinusoidalnego, który koduje grupy parametrów reprezentujących chwilowe częstotliwości i amplitudy sinusoid. Formowany jest strumień binarny zawierający zakodowane parametry modelu sinusoidalnego. Ten strumień binarny stanowi poboczną informację przesyłaną do dekodera wraz ze standardowym strumieniem binarnym MPEG-4 AAC HE. Ten dodatkowy strumień binarny może być traktowany jako opcjonalne rozszerzenie, pozwalające na uzyskanie wyższej jakości zdekodowanego sygnału w stosunku do techniki MPEG-4 AAC HE.

Należy jednak zauważyć, że jeśli dekoder nie będzie przystosowany do odbioru parametrów modelu sinusoidalnego, wówczas może nastąpić obniżenie jakości zdekodowanego sygnału. Stanie się tak, ponieważ zdekodowany sygnał będzie pozbawiony składowych tonalnych HF. Natomiast sygnały o charakterze szumu będą odtwarzane prawidłowo.

W rozprawie zaprezentowano ogólną koncepcję kodeka, który odtwarza harmoniczne dużych częstotliwości wierniej niż kodek MPEG-4 AAC HE. Autor zaprezentował również konstrukcję poszczególnych elementów kodeka implementującą powyżej opisaną ideę. Wybór elementów konstrukcji przedstawionej w pracy wynika z doświadczenia autora rozprawy zdobytego podczas konstrukcji i implementacji kodeków sygnałów fonicznych oraz studiów literaturowych.

Istotnym elementem proponowanej techniki jest proces wydzielenia tonalnych składowych dużych częstotliwości z sygnału wejściowego x. W wyniku separacji otrzymujemy sygnał y oraz zbiór parametrów opisujących składowe sinusoidalne, czyli amplitudy i częstotliwości. W proponowanej technice ciężar poprawnej rekonstrukcji tonalnych składowych HF został przeniesiony na stronę modelowania i kodowania sinusoidalnego. Autor rozprawy zaproponował również algorytm, którego zadaniem jest wybór składowych tonalnych podlegających separacji, a następnie ich usuwanie.

W rozdziale 3 rozprawy doktorskiej zaproponowano podstawowe elementy funkcjonalne techniki BWE oraz przykład implementacji bloków kodeka. Dodatkowo zaproponowano schemat wyboru składowych tonalnych, który zapewnia skuteczną separację składowych tonalnych i szumowych. Przedstawiono również sposób ograniczenia złożoności obliczeniowej algorytmu poprzez zastosowanie detektora tonalności i możliwe pominięcie przez koder sinusoidalny wskazanych fragmentów sygnału.

Zastosowanie modelowania sinusoidalnego w celu analizy i reprezentacji parametrów sygnału powoduje generowanie dodatkowych danych opisujących trajektorie sinusoidalne. Z tego powodu autor rozprawy zaproponował mechanizm sterownia przydziałem bitów oraz strategię kodowania parametrów modelu sinusoidalnego. W szczególności zaproponowano dwa warianty pracy kodera:

- **I wariant – bez pętli sterowania przepływnością** - w tym wariancie sterowania pracą kodera zakładamy, że percepcyjnie istotniejszą informacją, z punktu widzenie rekonstrukcji sygnału, jest poprawna generacja jak największej liczby składowych tonalnych, kosztem obniżenia częstotliwości podziału f_{SBR}. W konsekwencji może dojść do zmniejszenia wartości f_{SBR}, przez co poszerza się zakres pasma częstotliwości na którym operuje technika BWE.

- **II wariant – z pętlą sterowania przepływnością** - obliczany jest zasób bitów, który można przeznaczyć na zakodowanie parametrów modelu sinusoidalnego B_{Sin}. Jednocześnie, nie może nastąpić przełączenie granicy pracy kodera SBR - częstotliwość podziału f_{SBR}. Może to prowadzić do odrzucenia części trajektorii, a w konsekwencji niewłaściwą reprezentację składowych tonalnych HF. Z drugiej strony, koder podstawowy AAC operuje na maksymalnym, dopuszczalnym przez koder MPEG-4 AAC HE, zakresie częstotliwości pasma LF.

Należy podkreślić, że autor zaproponował również metodę selekcji percepcyjnie najistotniejszych trajektorii sinusoidalnych, która uwzględnia czas trwania trajektorii, jej całkowitą energię oraz częstotliwości poszczególnych składowych sinusoidalnych.

Przeprowadzono liczne badania symulacyjne oraz testy subiektywne, mające na celu ustalenie odpowiednich poziomów kwantyzacji parametrów modelu sinusoidalnego. **W efekcie ustalono schemat kompresji z rozszerzaniem widma, przy pomocy którego możliwa jest kompresja sygnału dla prędkości bitowych rzędu 16 – 24 kb/s, przy jednoczesnej wysokiej jakości zdekodowanego sygnału.** Prędkość bitowa strumienia danych opisującego składowe tonalne jest rzędu 3 – 6 kb/s.

W pracy przedstawiono nową, autorską metodę rozszerzania widma wykorzystującą modelowanie sinusoidalno-szumowe, która umożliwia poprawną rekonstrukcję składowych tonalnych dużych częstotliwości. **Nowa technika może być wykorzystana w połączeniu z istniejącymi już technikami BWE, pozwalając na zmniejszenie zniekształceń powodowanych przez replikację widma stosowaną w obecnych standardach kompresji sygnałów fonicznych (np. Dolby EAC-3, MPEG-4 AAC HE).** W pracy zastosowano połączenie kodeka standardu MPEG-4 AAC HE oraz modelowania sinusoidalnego.

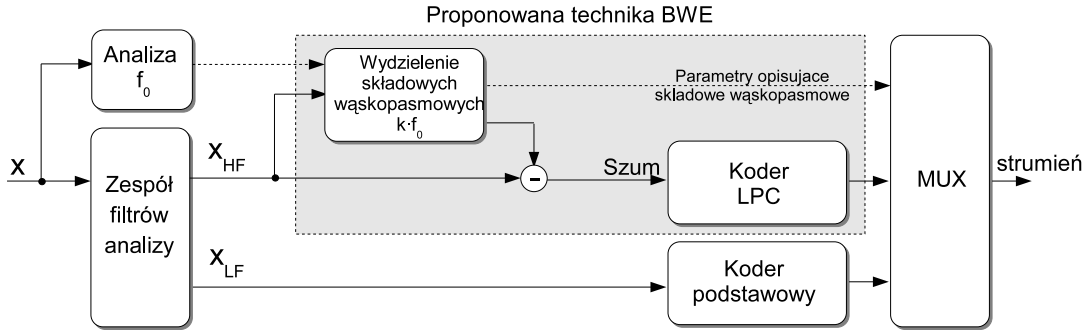
Do najważniejszych zalet proponowanej techniki rozszerzania widma należy:

- poprawna rekonstrukcja struktury harmonicznej oraz zmian na płaszczyźnie czasowo-częstotliwościowej składowych tonalnych dużych częstotliwości,
- uzyskanie wyższej jakości zdekodowanego sygnału w stosunku do kodeka MPEG-4 AAC HE.

2.2.2. Metoda rozszerzania widma wykorzystująca skalowanie częstotliwości

W pracy zaproponowano dodatkową metodę rozszerzania widma, która jest alternatywą dla replikacji widma (SBR) stosowanej w MPEG-4 AAC HE. Główna idea metody dotyczy rekonstrukcji struktury tonalnej składowych dużych częstotliwości z wykorzystaniem informacji o częstotliwościach podstawowych wyznaczonych w koderze (rysunek 2.3). Sygnał jest dekomponowany na grupy sygnałów wąskopasmowych, uzyskanych w procesie demodulacji przy wykorzystaniu chwilowych częstotliwości składowych tonalnych. Składowe dużych częstotliwości (HF) rekonstruowane są poprzez modulację sygnału wąskopasmowego z uwzględnieniem odpowiednio przeskalowanej częstotliwości chwilowej tego sygnału. Podejście to oferuje poprawną syntezę szybkozmiennych składowych tonalnych sygnału, jak również poprawną rekonstrukcję szeregów harmonicznych. Dla zachowania właściwej obwiedni widma decydującej o brzmieniu muzyki zastosowano również korekcję energii przeskalowanych składowych tonalnych. Podczas badań, do kodowania małych częstotliwości zastosowano koder podstawowy AAC (MPEG-4 AAC HE).

Rzeczywiste sygnały (w tym muzyka) zawierają składowe pseudo-sinusoidalne o szybko zmiennych wartościach amplitudy, częstotliwości i fazy, które reprezentują cechy wielu



Rysunek 2.3: Ogólny schemat proponowanej techniki. Linią ciągłą zaznaczono nieskompresowane próbki sygnału, linią przerywaną - dane kontrolno-sterujące.

dźwięków takie jak dźwięczność dysząca, skrzypiąca, szorstka (pojęcia fonologiczne), odróżniające je od produktu modelowania sinusoidalnego ocenianego przez wielu słuchaczy jako brzmiące syntetycznie. Tego typu cechę sygnału w dalszej części pracy nazywać będę mikromodulacjami AM i FM. Mikro-modulacje mają charakter losowy, co powoduje, że widmo amplitudowe takiego sygnału pewną szerokość i może być nawet interpretowane jako wąskopasmowy szum. Model parametryczny opisujący tego typu zjawiska powinien być w stanie reprezentować centralną częstotliwość oraz szerokość widma składowej tonalnej.

W dalszej części rozdziału przez składową tonalną rozumieć będziemy wąskopasmowy sygnał będący sinusoidą, której chwilowe parametry amplitudy i częstotliwości zawierają mikro modulacje AM i FM.

Dla sygnału wąskopasmowego reprezentującego składową tonalną można zdefiniować chwilową częstotliwość $f(t)$, fazę $\varphi(t)$ i amplitudę $A(t)$. Jeśli dla sygnału wąskopasmowego:

$$x(t) = A(t)\cos(\varphi(t)), \tag{2.1}$$

zapiszemy odpowiadający mu sygnał analityczny:

$$z(t) = x(t) + j\mathcal{H}\{x(t)\}, \tag{2.2}$$

gdzie $\mathcal{H}\{x(t)\}$ jest transformatą Hilberta sygnału $x(t)$. Wówczas:

$$z(t) = A(t)e^{j\varphi(t)}. \tag{2.3}$$

Amplitudę chwilową sygnału $x(t)$ możemy zapisać jako:

$$A(t) = |z(t)|. \tag{2.4}$$

Natomiast częstotliwość chwilowa sygnału $x(t)$ jest pochodną argumentu sygnału analitycznego:

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\arg\{z(t)\}. \tag{2.5}$$

Źródła dźwięku wielu naturalnych instrumentów rozpatrywać możemy jako rezonatory, które generują wąskopasmową energię, gdy są pobudzone przez losowy proces, np. podmuch powietrza, tarcie struny. Wygenerowany w ten sposób sygnał można interpretować jako sumę szeregów harmonicznych o wartościach chwilowych częstotliwości i amplitud,

które zmieniają się losowo w niewielkim zakresie. W tego typu sygnałach, pojedyncze składowe identyfikowane są w dziedzinie krótkookresowego widma jako wąskopasmowe skupienie energii.

W pracy rozpatruje się szeregi harmoniczne $s_h(t)$ będące sumami składowych pseudo-sinusoidalnych (czyli składowych sinusoidalnych o zmiennej w czasie amplitudzie i częstotliwości):

$$s_h(t) = \sum_{k=1}^{K} A_k(t) \cos\left(2\pi k \int_0^t f_0(\tau)d\tau + \varphi_k(0)\right), \tag{2.6}$$

gdzie k jest indeksem harmonicznej, K jest liczbą harmonicznych, $A_k(t)$ oznacza amplitudę chwilową, $f_0(\tau)$ jest częstotliwością chwilową, $\varphi_k(0)$ fazą początkową składowej pseudo-sinusoidalnej.

Główna idea techniki prezentowanej w bieżącym rozdziale polega na odtworzeniu wąskopasmowych sygnałów reprezentujących składowe tonalne dużych częstotliwości. Można to zrobić przekształcając składowe małych częstotliwości poprzez skalowanie ich parametrów chwilowych ($A(t)$ i $f(t)$). Jeśli dla danego szeregu harmonicznego (2.6) częstotliwość chwilowa k-tej składowej harmonicznej wynosi:

$$f_k(t) = k \cdot f_0(t), \qquad k = 1, ..., K, \tag{2.7}$$

gdzie $f_0(t)$ jest chwilową częstotliwości podstawową szeregu harmonicznego, wówczas sygnał wąskopasmowy reprezentujący k-tą harmoniczną opisany jest wzorem:

$$x_k(t) = Re\{A_k(t)\, e^{j2\pi k f_0(t)}\}. \tag{2.8}$$

Dodatkowo wąskopasmowy sygnał reprezentujący składową zmieniającą się wraz z częstotliwością podstawową opisujemy:

$$x_1(t) = Re\{A_1(t)\, e^{j2\pi f_1(t)}\}. \tag{2.9}$$

Analiza sygnału w dekoderze

W proponowanej technice koder estymuje wartości częstotliwości chwilowych $f(t)$ oraz amplitudy $A(t)$ danego szeregu harmonicznego. Zakodowane parametry przesyłane są do dekodera, w którym odbywa się proces analizy sygnału na podstawie informacji z kodera. Przykład procesu analizy sygnału skalowania częstotliwości (wzór (2.9)) pokazany jest na rysunku 2.4. Analiza sygnału polega na wyodrębnieniu sygnału $x_0(t)$ i na jego podstawie rekonstrukcji składowych dużych częstotliwości. Pierwszym krokiem algorytmu jest demodulacja sygnału wejściowego $x(t)$ zgodnie z przebiegiem częstotliwości podstawowej $f_0(t)$ (rysunek 2.4b). Następnie, sygnał poddany jest filtracji dolnoprzepustowej, w celu otrzymania reprezentacji sygnału $x_0(t)$ w paśmie podstawowym. Otrzymany sygnał $x_0(t)$ poddany jest modulacji zgodnie z częstotliwością $f_0(t)$ przez co uzyskujemy wąskopasmową reprezentację składowej tonalnej. Następnie częstotliwość sygnału $x_0(t)$ jest przeskalowywana kolejnymi, całkowitymi wielokrotnościami częstotliwości f_0 zgodnie ze wzorem (2.9). W wyniku modulacji otrzymujemy sygnał wąskopasmowy $x_{HT}(t)$ będący sumą składowych tonalnych dużych częstotliwości (rysunek 2.4c). Powyższa procedura powtarzana jest dla każdej trajektorii f_0. Sygnał przedstawiony na rysunku 2.4d stanowi

sumę wąskopasmowych składowych tonalnych $x_k(t)$. Wszystkie składowe $x_k(t)$ mają tą samą energię, aby więc poprawnie odtworzyć sygnał $x_{HT}(t)$ należy przeskalować obwiednię amplitudową widma otrzymanego sygnału.

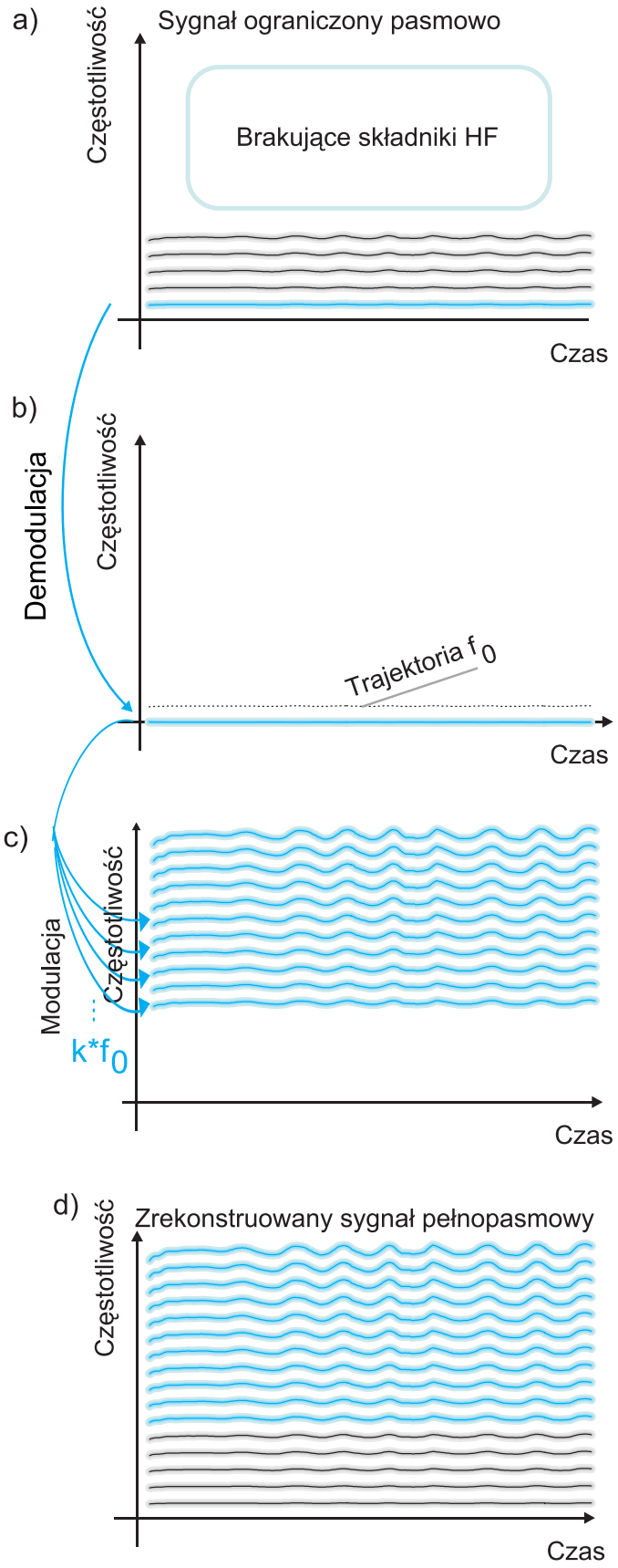
Współczynniki skalujące wyznaczane są w kolejnych, następujących po sobie oknach analizy. W tym celu zastosowano metodę predykcji liniowej LPC (ang. *Linear Predictive Coding*), w której dla każdego krótkookresowego widma (STFT) wyznaczane są współczynniki predyktora LPC. Następnie charakterystyka amplitudowa predyktora $|G_{LPC_m}(f)|$ jest próbkowana w miejscach $p = kf_0$, gdzie p jest częstotliwością składowej tonalnej, k jest indeksem składowej. Otrzymane współczynniki skalujące poszczególnych składowych tonalnych p w danym oknie m reprezentowane są przez funkcję skalującą widmo $|G_{LPC_m}(p)|$. W pracy zastosowano filtr LPC 10 rzędu. Współczynniki LPC reprezentowane przez częstotliwości linii widmowych - LSF (ang. *Line Spectral Frequencies*). Współczynniki filtru oraz wzmocnienie podlegają kompresji i są przesyłane do dekodera.

Dodatkowym elementem w proponowanej technice jest wyznaczenie składowej szumowej dużych częstotliwości $x_{HN}(t)$. Sygnał $x_{HN}(t)$ uzyskiwany jest poprzez usunięcie składowych tonalnych $x_{HT}t$ z sygnału wejściowego stosując algorytm tłumienia widma (opisany szczegółowo w rozprawie). Otrzymany sygnał $x_{HN}(t)$ traktowane jest jak szum i poddany kodowaniu LPC.

Po stronie dekodera składowe dużych częstotliwości rekonstruowane są poprzez skalowanie częstotliwości chwilowych sygnału $x_0(t)$ uzyskanego poprzez demodulację z sygnału małych częstotliwości $\hat{x}_L(t)$. Proces skalowania częstotliwości realizowany jest w podobny sposób jak w koderze.

Strumień bitowy generowany przez proponowany koder uwzględnia dane trajektorii częstotliwości podstawowych $f_0(t)$, współczynniki skalujące obwiednię widma oraz współczynniki predyktora LP opisujące składową szumową dużych częstotliwości.

Zaproponowana metoda rozszerzania widma wykorzystująca skalowanie częstotliwości prezentuje nowe, oryginalne podejście do realizacji techniki BWE oferującej wysoką jakość dekodowanego sygnału przy jednoczesnej małej prędkości bitowej, rzędu 16 – 24 kb/s, natomiast 1 – 3 kb/s dla składowych HF. Główną zaletą proponowanej techniki jest poprawna rekonstrukcja energii tonalnych składowych dużych częstotliwości oraz prawidłowa rekonstrukcja zmiennych w czasie i częstotliwości składowych sygnału. Kodek wykorzystujący proponowaną technikę BWE pozwala na osiągnięcie wyższej jakości zdekodowanego sygnału w porównaniu do obecnie istniejących metod, czyli MPEG-4 AAC HE, zwłaszcza dla sygnałów posiadających silne składowe tonalne w zakresie dużych częstotliwości.



Rysunek 2.4: Proces analizy i syntezy sygnału w proponowanej technice BWE: a) spektrogram sygnału ograniczonego pasmowo rekonstruowanego w dekoderze, b) wąskopasmowa składowa tonalna, c) zrekonstruowane składowe tonalne HF, d) zrekonstruowany sygnał stanowiący sumę składowych HF oraz sygnału zdekodowanego w dekoderze podstawowym.

2.3. Badanie zaproponowanych technik rozszerzania widma z wykorzystaniem subiektywnej miary jakości

Celem badań eksperymentalnych było określenie efektywności kompresji kodeków, w których wykorzystano zaproponowane w rozprawie techniki BWE dla prędkości bitowych w zakresie 16 – 24 kb/s. Jako punkt odniesienia posłużył kodek MPEG-4 AAC HE. Podczas badań wykorzystano metodologię subiektywnych testów odsłuchowych dla małych i średnich zniekształceń według standardu ITU-R BS.1534-1 (MUSHRA). Opis metody przeprowadzenia testów odsłuchowych oraz analiza statystyczna wyników znajdują się w pełnym tekście rozprawy.

Wykorzystane sygnały testowe zostały podzielone na dwie kategorie:

- zestaw 1 - fragmenty nagrań utworów zagranych na *wielu instrumentach* – muzyka.
- zestaw 2 - fragmenty nagrań utworów zagranych na *pojedynczym instrumencie*,

Sygnały foniczne użyte w badaniach są sygnałami monofonicznymi o częstotliwości próbkowania 44,1 kHz. W trakcie badań ocenie podlegała jakość dźwięku sygnałów zdekodowanych przez testowane kodeki, dla prędkości bitowych 16, 20 i 24 kb/s.

Zastosowano następujące oznaczenia technik BWE:

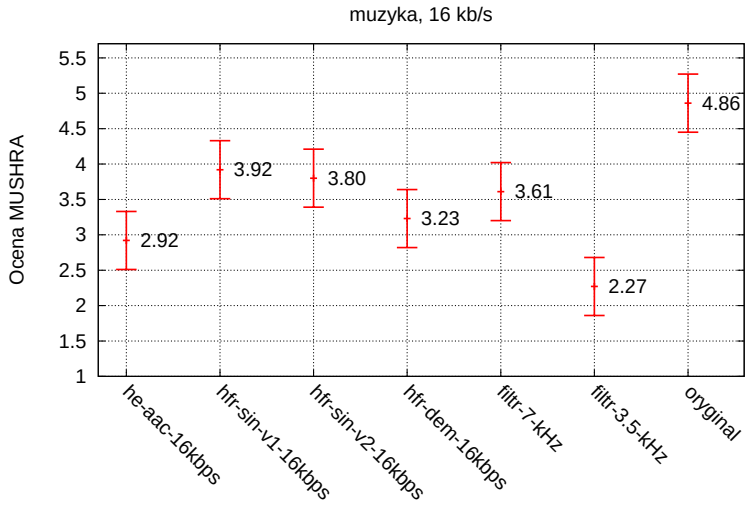
- hfr-sin – łączna metoda wykorzystująca MPEG-4 AAC HE oraz modelowanie sinusoidalne (opisana w punkcie 2.2.1). Technika została przebadana pod kątem zastosowania pętli sterowania przepływnością. Zaproponowano dwa warianty:
  - hfr-sin-v1 – wariant bez pętli sterowania przepływnością,
  - hfr-sin-v2 – wariant z pętlą sterowania przepływnością,
- hfr-dem – technika rozszerzania widma poprzez skalowanie częstotliwości (opisana w punkcie 2.2.2),
- he-aac – MPEG-4 AAC HE zawierający technikę SBR.

Dodatkowo podczas badań wykorzystano nieskompresowane sygnały odniesienia (kotwice) o częstotliwości próbkowania 44100 Hz, rozdzielczości 16 bitów na próbkę, w tym sygnał oryginalny o jakości nagrania na płycie CD (oryginał), sygnał oryginalny ograniczony pasmowo do 7 kHz (filtr-7-kHz) oraz sygnał oryginalny ograniczony pasmowo do 3,5 kHz (filtr-3,5-kHz).

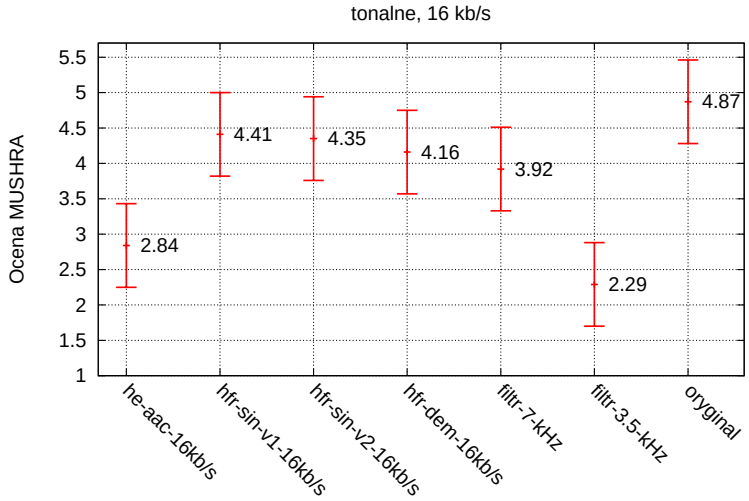
W badaniach wzięło udział 17 osób w wieku od 20 do 38 lat. Uczestnicy testów mieli już doświadczenie związane z testami odsłuchowymi. Osoby te zostały również zaznajomione z metodologią badań, sygnałami testowymi oraz rodzajem zniekształceń wprowadzanych przez techniki kompresji sygnałów fonicznych.

Testy zostały przeprowadzone w specjalnie do tego celu przygotowanym pomieszczeniu odsłuchowym w Katedrze Telekomunikacji Multimedialnej i Mikroelektroniki, Politechniki Poznańskiej. Pomieszczenie to spełnia wymagania akustyczne stawiane przez normę ITU-R BS.1534-1 (MUSHRA).

Na rysunkach 2.5 oraz 2.6 przedstawiono przykładowe zestawienie wyników dla sygnałów muzycznych (zestaw 1) oraz sygnałów zawierających silne składowe tonalne (zestaw 2) zdekodowanych za pomocą proponowanych technik BWE.



Rysunek 2.5: Porównanie technik rozszerzania widma dla sygnałów muzycznych (zestaw 1). Wyniki testów odsłuchowych MUSHRA dla prędkości transmisji 16 kb/s z 95% przedziałem ufności, przy założeniu rozkładu normalnego.



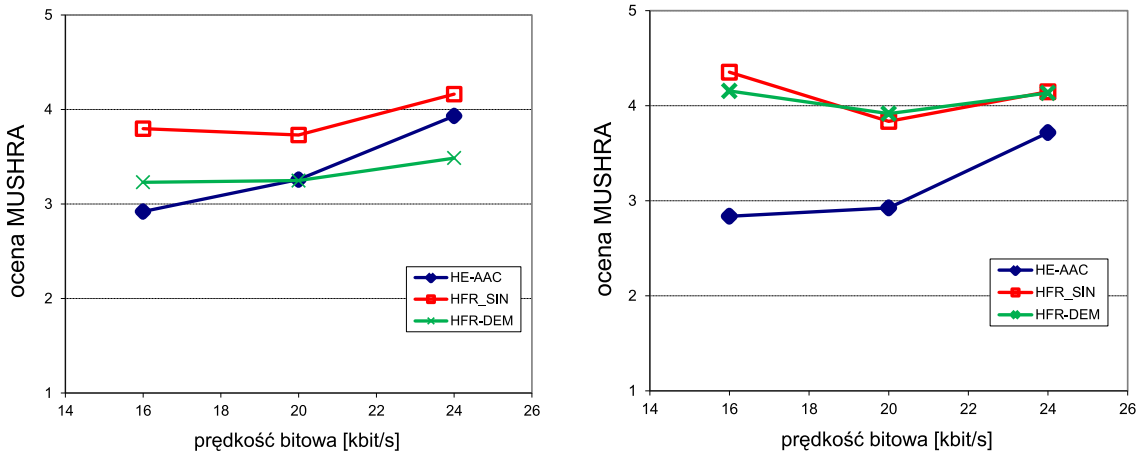
Rysunek 2.6: Porównanie technik rozszerzania widma dla sygnałów zawierających silne składowe tonalne (zestaw 2). Wyniki testów odsłuchowych MUSHRA dla prędkości transmisji 16 kb/s z 95% przedziałem ufności, przy założeniu rozkładu normalnego.

Zaprezentowana w punkcie 2.2.1 technika rozszerzania widma wykorzystująca modelowanie sinusoidalne, zbadana została w dwóch wariantach z i bez pętli sterowania przepływnością. Na podstawie analizy wyników można stwierdzić, że wariant z pętlą sterowania przepływnością („hfr-sin-v2") pozwala na wyższą efektywność kompresji w stosunku

do wariantu bez pętli sterowania. Dlatego w zagregowanych wynikach (rysunki 2.7a, 2.7b i 2.8) symbolem „hfr-sin" oznaczany będzie kodek z pętlą sterowania przepływnością.

Wyniki **dla zestawu 1 sygnałów muzycznych** wykorzystanych w eksperymentach pokazują, że **technika „hfr-sin" otrzymała znacznie wyższe oceny niż porównywana technika MPEG-4 AAC HE dla prędkości bitowych 16 - 24 kb/s** (rysunek 2.7a). Dla tego samego zestawu nagrań, technika „hfr-dem" otrzymała wyniki nieznacznie wyższe od MPEG-4 AAC HE dla 16 kb/s i zbliżone dla pozostałych prędkości bitowych.

Dla zestawu 2 zawierającego **sygnały o silnych składowych tonalnych** (rysunek 2.7b) **obie proponowane techniki są wyraźnie lepsze od MPEG-4 AAC HE**, zwłaszcza dla prędkości bitowej 16 kb/s.



(a) zestaw 1

(b) zestaw 2

Rysunek 2.7: Zależność subiektywnej oceny od prędkości bitowej dla: a) zestaw 1 - wszystkich badanych sygnałów, b) zestaw 2 - nagrań zawierających silne składowe tonalne. „hfr-sin" oznacza technikę opisaną w rozdziale 2.2.1 z pętlą sterowania przepływnością.

Wyniki badań opisane w rozdziale pokazują, że **kodeki wykorzystujące proponowane techniki BWE uzyskują szczególnie dobre rezultaty dla sygnałów fonicznych zawierających nagrania instrumentów o silnych składowych tonalnych oraz składowych o zmiennej częstotliwości** (zestaw 2). Dla tego typu sygnałów subiektywna ocena proponowanych rozwiązań mieści się w przedziale pomiędzy „dobrą" a „doskonałą" jakością na skali MUSHRA (około 4,5 / 5 punktów). Porównywana technika MPEG-4 AAC HE otrzymuje dla tych sygnałów, ocenę poniżej jakości „zadawalającej" na skali MUSHRA (poniżej 3 punktów).

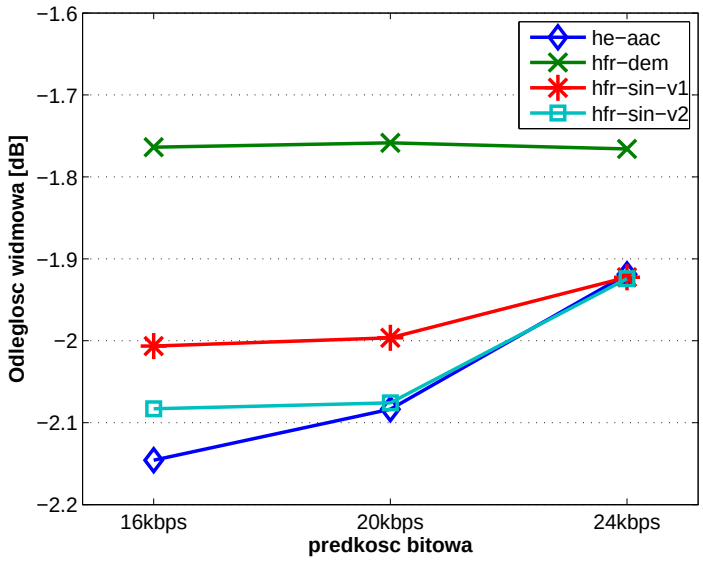
Zaproponowane w rozprawie rozszerzenie (rozdział 2.2.1) zwiększa efektywność kompresji standardowej techniki MPEG-4 AAC HE dla badanych sygnałów muzycznych o około $0,2 - 0,7$ w skali MUSHRA, w zależności do prędkości bitowej. **Szczególnie wyraźny wzrost efektywności można zaobserwować dla sygnałów zawierających silne składowe tonalne - nawet** $1,5$ **w skali MUSHRA**.

Kodek wykorzystujący drugą z zaproponowanych technik BWE - „hfr-dem" (rozdział 2.2.2) otrzymał szczególnie wysokie oceny subiektywne dla drugiego zestawu nagrań. Oznacza to, że **zaproponowana technika może stanowić dobrą alternatywę dla kodeka SBR w zastosowaniu dla kodowania sygnałów o silnych składowych tonalnych**.

Istotną zaletą proponowanych technik jest uzyskanie „dobrej" jakości dźwięku bez względu na wybraną prędkość bitową w badanym zakresie prędkości bitowych 16-24 kb/s, dla szerokiej klasy sygnałów muzycznych (rysunek 2.7).

2.4. Badanie zaproponowanych technik rozszerzania widma z wykorzystaniem obiektywnej miary jakości

W celu weryfikacji subiektywnych badań eksperymentalnych, przeprowadzono dodatkowe badania porównujące techniki rozszerzania widma, posługując się obiektywną miarą zniekształceń. W tym przypadku zastosowano miarę logarytmicznej odległości widmowej (ang. *log-spectral distance*) (LSD). LSD jest dobrze opisaną w literaturze miarą, która bazuje na informacji, że sposób postrzegania głośności przez człowieka ma charakter logarytmiczny. Jednak ignoruje ważne właściwości narządu słuchu, na przykład maskowanie w dziedzinie częstotliwości i czasu. Z tego powodu, wyniki uzyskane za pomocą miary LSD mają charakter poglądowy i służą weryfikacji oceny subiektywnej. Zagregowane wyniki uzyskane za pomocą miary odległości widmowej przedstawione na rysunku 2.8 potwierdzają wysoką efektywność proponowanych technik BWE. Technika „hfr-dem" uzyskuje wyraźnie wyższą ocenę obiektywną w stosunku do MPEG-4 AAC HE. W przypadku technik „hfr-sin-v1" oraz „hfr-sin-v2" różnica pomiędzy porównywanymi technikami nie jest tak znacząca, jednak w dalszym ciągu wyższa od MPEG-4 AAC HE.



Rysunek 2.8: Zależność obiektywnej oceny (odległości widmowej) od prędkości bitowej dla wszystkich badanych sygnałów.

Rozdział 3

Podsumowanie i dyskusja

3.1. Wnioski

Uzyskane wyniki badań potwierdzają tezę pracy. W rozprawie zostały zaproponowane techniki kompresji sygnału fonicznego, które łącznie wykorzystują rozszerzanie widma oraz modelowanie sinusoidalno-szumowe. Opisane w rozprawie nowe techniki zostały zaimplementowane jako rozszerzenia standardowej techniki MPEG-4 AAC HE. Nowe techniki kompresji umożliwiają rekonstrukcję sygnału fonicznego o jakości lepszej niż przy zastosowaniu MPEG-4 AAC HE dla prędkości bitowych rzędu 16 – 24 kb/s. Równocześnie, zaproponowane rozwiązania poprawnie rekonstruuje składowe tonalne sygnału.

Poniżej przedstawione zostały najważniejsze wnioski rozprawy.

- Zastosowanie modelowania sinusoidalnego pozwala na poprawną reprezentację składowych tonalnych sygnału dużych częstotliwości oraz ich efektywną kompresję.

- Zastosowanie modelowania sinusoidalnego w połączeniu z techniką - MPEG-4 AAC HE, pozwala na poprawę jakości zdekodowanego sygnału dla tej samej prędkości bitowej.

- Efektywność kompresji dla techniki wykorzystującej łączny mechanizm rozszerzania widma i modelowania (rozdział 2.2.1) w dużym stopniu zależy od poprawnej klasyfikacji składowych tonalnych. Na proces klasyfikacji istotny wpływ ma obecność szumu w sygnale oraz charakter poszczególnych składowych tonalnych. Szybkozmienne składowe w krótkookresowym widmie mogą zostać zakwalifikowane jako szum. Parametry trajektorii sinusoidalnych kodowane są różnicowo, jeśli więc zostaną zakłócone przez szum wówczas efektywność kompresji zmniejszy się.

- W rozprawie pokazano, że istnieje minimalna wartość $\overline{N}_{Sin,max}$ średniej liczby sinusoid na okno analizowanego sygnału, umożliwiająca poprawną rekonstrukcję sygnału, czyli uzyskania wyniku „dobry” na skali MUSHRA. Dla zaproponowanej techniki granica ta wynosi 20 sinusoid na okno analizy.

- W pracy przedstawiono wyniki badań, które wskazują, że istnieje górna granica liczby kodowanych trajektorii, której przekroczenie nie poprawia jakości zrekonstruowanego sygnału fonicznego. W tej sytuacji dochodzi do redukcji jakości dekodowanego sygnału. Spowodowane jest to błędną klasyfikacją składowych tonalno-szumowych. Podczas badań ustalono, że dla wykorzystanych sygnałów testowych redukcja jakości następuje po przekroczeniu średniej liczby 40 – 50 sinusoid na okno analizy.

- Na podstawie wyników badań przedstawionych w pracy ustalono, że parametry modelu sinusoidalnego powinny być kwantowane równomiernie w skali logarytmicznej z krokiem 40 centów dla częstotliwości oraz 3 dB dla amplitud. W zaproponowanych rozwiązaniach zrezygnowano z kodowania parametru fazy dla dużych częstotliwości. Zastosowana strategia kodowania pozwoliła na efektywną kompresję oraz poprawną rekonstrukcję sygnałów.

- Istotnym parametrem wpływającym na jakość zrekonstruowanego sygnału jest ciągłość parametrów modelu sinusoidalnego, zarówno amplitud jak i częstotliwości. Niekompletne lub krótkotrwałe trajektorie postrzegane są jako składowe szumowe lub powodują nieprzyjemne artefakty związane z gwałtowną zmiana głośności na początku i końcu trajektorii.

- Sterowanie pracą kodera sinusoidalnego musi odbywać się poprzez zastosowanie odpowiedniego warunku psychoakustycznego. W innym wypadku, koder może odrzucić trajektorie percepcyjnie istotne.

- W zaproponowanych rozwiązaniach zastosowano kodowanie Huffmana z dynamicznie tworzoną tablicą przesyłaną co pewien interwał czasu. Na podstawie wyników badań przedstawionych w pracy stwierdzono, że dla tablicy Huffmana przesyłanej co około 0,5 s możliwa jest rekonstrukcja sygnału o wysokiej jakości. Jednocześnie jest to odstęp czasu umożliwiający właściwą synchronizację ze strumieniem obrazów ruchomych.

- Na podstawie badań wstępnych ustalono, że składowe tonalne warto rozpatrywać jako sygnały wąskopasmowe o szerokości pasma zależnej od lokalnych modulacji AM i FM. Ponieważ reprezentacja oparta na takim podejściu oferuje wierniejszą, naturalniej brzmiącą rekonstrukcję sygnału, w porównaniu do reprezentacji typowej dla modelu sinusoidalnego, która brzmi syntetycznie z powodu zbyt uproszczonego przedstawienia zmienności amplitud i częstotliwości jako funkcji wolnozmiennej.

- Badania wstępne przeprowadzone z wykorzystaniem referencyjnej implementacji kodera MPEG-4 AAC HE wykazały istotne znaczenie modulacji amplitudy i częstotliwości poszczególnych składowych tonalnych dla subiektywnej oceny jakości sygnału. Ucho ludzkie wrażliwe jest na zniekształcenia modulacji częstotliwości wprowadzane przez koder standardu MPEG-4 AAC HE w zakresie dużych częstotliwości. Wnioski te zostały również potwierdzone podczas badań przedstawionych w rozprawie.

3.2. Oryginalne osiągnięcia

Głównym osiągnięciem autora jest opracowanie dwóch efektywnych technik kompresji wykorzystujących rozszerzanie widma i jednocześnie mogących być uzupełnieniem aktualnie stosowanych standardów w tej dziedzinie (MPEG-4 AAC HE).

Technika opisana w rozdziale 2.2.1 umożliwia dokładniejszą rekonstrukcję składowych tonalnych w stosunku do standardowej techniki SBR. Zrealizowano to, poprzez łączne zastosowanie w tym samym paśmie częstotliwości modelowania sinusoidalnego oraz techniki SBR. Proponowana technika stanowi uzupełnienie techniki MPEG-4 AAC HE.

Druga technika przedstawiona w rozdziale 2.2.2 wykorzystuje skalowanie częstotliwościowe składowych tonalnych z dolnej części pasma w celu odtworzenia brakujących wyższych składowych harmonicznych. Wąskopasmowy sygnał reprezentuje składową tonalną jako składową sinusoidalną przekształconą mikro-modulacjami AM i FM. Technika ta jest alternatywnym rozwiązaniem dla techniki SBR.

Obie techniki rozszerzania widma są autorskimi rozwiązaniami i stanowią oryginalny wkład w dziedzinie kompresji sygnałów wykorzystujących rozszerzanie widma.

Nowe techniki zostały przebadane eksperymentalnie zgodnie z obowiązującymi standardami testów subiektywnych. Parametry proponowanych kodeków ustalono na podstawie wielu badań eksperymentalnych. Uzyskane wyniki pozwoliły na implementację kodeków, a następnie poddanie ich badaniom eksperymentalnym mającym na celu porównanie proponowanych rozwiązań z techniką MPEG-4 AAC HE.

Wspomniane wyżej prace eksperymentalne wykonanie około 10 serii testów odsłuchowych, w których łączny czas trwania prezentowanych nagrań wynosił ponad 150 godzin. Na ten czas składa się końcowe porównanie efektywności proponowanych rozwiązań, jak również czas poświęcony na dobór wstępnych parametrów kodeków.

Należy podkreślić fakt, iż testy odbywały się w specjalnie do tego celu przygotowanym pomieszczeniu odsłuchowym. W tym miejscu w danym czasie mogła przebywać jedna osoba, co znacząco komplikowało i wydłużało czas badań. Realny czas trwania eksperymentów szacowany jest na 3 - 4 miesiące codziennych badań.

Przeprowadzenie i przygotowanie tak dużej liczby eksperymentów było zadaniem trudnym ze względów logistycznych i dużej liczby osób uczestniczących w eksperymentach. Na trud powstania pracy składa się również czas poświęcony na przetworzenie i analizę wyników.

Wyniki badań przedstawione w rozdziale 2.3 pokazują, że proponowane techniki oferują wyższą efektywność kompresji w stosunku do standardowej techniki MPEG-4 AAC HE. Szczególnie dobre rezultaty uzyskały dla sygnałów fonicznych zawierających dźwięki o silnych składowych tonalnych i/lub zmiennej częstotliwości. Dla tego typu sygnałów subiektywna ocena proponowanych rozwiązań mieści się w przedziale pomiędzy „dobrą" a „doskonałą" jakością na skali MUSHRA (około 4,5 / 5 punktów). Porównywana technika MPEG-4 AAC HE otrzymuje, dla tych sygnałów, ocenę

poniżej jakości „zadawalającej” na skali MUSHRA (poniżej 3 punktów). Należy podkreślić, że dla prędkości bitowych poniżej 24 kb/s i dla pozostałych klas sygnałów uwzględniających muzykę, proponowane techniki uzyskały wyższą efektywność kompresji w stosunku do MPEG-4 AAC HE. Dla prędkości bitowych równych 24kb/s porównywane techniki oferują podobną efektywność kompresji.

Poniżej przedstawiono **pozostałe oryginalne osiągnięcia autora pracy**.

- Zaproponowanie metody klasyfikacji składowych tonalnych bazującej na łącznej analizie parametrów widma chwilowego oraz trajektorii sinusoidalnych.
- Zaproponowanie klasyfikatora tonalności, którego celem jest przełączanie pracy kodera pomiędzy aktualnym trybem, a proponowanymi technikami.
- Przedstawienie strategii ograniczenia liczby kodowanych trajektorii do percepcyjnie najistotniejszych.
- Przedstawienie algorytmu sterownia pracą kodera sinusoidalnego, bazując na percepcyjnym kryterium ograniczania liczby kodowanych trajektorii sinusoidalnych.
- Pokazanie aktualnego stanu wiedzy w dziedzinie kompresji z rozszerzaniem widma.
- Implementacja aplikacji służącej do prowadzenia testów odsłuchowych subiektywnej oceny jakości sygnałów fonicznych według metodologii MUSHRA.

Istotnym osiągnięciem przedstawionym w pracy jest analiza zniekształceń wprowadzanych przez technikę MPEG-4 AAC HE. Na podstawie badań wskazano mechanizmy powodujące powstawanie zniekształceń oraz zaproponowano możliwe rozwiązania.

Autor przedstawił w pracy również analizę właściwości modelowania sinusoidalno-szumowego, które uwzględniały problemy klasyfikacji składowych, estymację parametrów oraz wskazał źródła zniekształceń kompresji dla modelownania sinusoidalnego.

Autor rozprawy na 93-cim spotkaniu grupy MPEG (Genewa 2010) zgłosił propozycję rozszerzenia auktualnie powstającego standardu kodowania sygnałów fonicznych MPEG-D Unified Speech and Audio Coding zgodnie z opisem w rozdziale 2.2.1. Wyniki pracy zostały bardzo wysoko ocenione przez grupę ekspertów MPEG. Aktualnie zgłoszona łączna technika modelowania sinusoidalnego i MPEG-D USAC została przyjęta do etapu weryfikacji, po którego zakończeniu zapadnie decyzja o przyjęciu proponowanego rozszerzenia jako części nowego standardu MPEG-D USAC. W proces weryfikacji proponowanego rozszerzenia zaangażowane są następujące instytucje: Panasonic, Yonsei University, Philips, Fraunhoffer IIS, Telcordia Poland, Politechnika Poznańska.

Rozdział 4

Dorobek naukowy autora

4.1. Publikacje naukowe autora

Konferencje międzynarodowe

1. T. Żernicki and M. Bartkowiak, "Audio bandwidth extension by frequency scaling of sinusoidal partials," in *125th Convention of the Audio Engineering Society*, ser. AES Preprint 7622, San Francisco, USA, 2-5 Oct. 2008.

2. M. Bartkowiak and T. Żernicki, "Harmonic sinusoidal + noise modeling of audio based on multiple f0 estimation," in *125th Convention of the Audio Engineering Society*, ser. AES Preprint 7510, San Francisco, USA, 2-5 Oct. 2008.

3. M. Bartkowiak and T. Żernicki, "A simple adaptive matrixing scheme for efficient coding of stereo sound," in *Proceedings of IX European Signal Processing Conference, EUSIPCO'05*, Antalya, Turkey, Sep. 2005.

4. P. Garstecki, A. Łuczak, T. Żernicki, "A bit-serial architecture for H.264/AVC interframe decoding," in *Proceedings of IX European Signal Processing Conference, EUSIPCO'06*, Florence, Italy, Sep. 4-8, 2006.

5. T. Żernicki and M. Domański, "Improved coding of tonal components in MPEG-4 AAC with SBR," in *Proceedings of IX European Signal Processing Conference, EUSIPCO'08*, Lausanne, Switzerland, Aug. 25-29 2008.

6. M. Bartkowiak and T. Żernicki, "The impact of finite word length on IMDCT computation in MPEG-2 AAC decoding," in *International Workshop on Systems, Signals and Image Processing, IWSSIP'04*, Poznań, 2004, pp. 343–346.

7. M. Bartkowiak and T. Żernicki, "Software audio codec for interactive television," in *International Workshop on Systems, Signals and Image Processing, IWSSIP'04*, Poznań, 2004, pp. 469–472.

Dokumenty standaryzacyjne

8. T. Żernicki, M. Bartkowiak, M. Domański, "Improved coding of tonal components in audio techniques utilizing the SBR tool", ISO/IEC JTC1/SC29/WG11 MPEG 2010 / M17914, Geneva, Switzerland, July 2010.

9. T. Żernicki, M. Bartkowiak, M. Domański, "Telcordia and PUT listening test results for CE on improved tonal component coding in eSBR (USAC)", ISO/IEC JTC1/SC29/WG11 MPEG 2010 / M18532, Guangzhou, China, October 2010.

Czasopisma krajowe

10. T. Żernicki i M. Bartkowiak, "Zastosowanie modelowania sinusoidalnego do regeneracji wysokoczęstotliwościowych składowych tonalnych kodera MPEG-4 HE-AAC", *Przegląd telekomunikacyjny 4/2008*, Kwiecień 2008, s. 511–514 (*Krajowa Konferencja Radiokomunikacji, Radiofonii i Telewizji, KKRRiT*, Wrocław, Polska, 9-11 Kwiecień 2008, s. 511–514).

11. M. Bartkowiak i T. Żernicki, "Hybrydowy parametryczno-transformatowy kodek dźwięku", *Przegląd telekomunikacyjny 4/2008*, Kwiecień 2008, s. 507–510 (*Krajowa Konferencja Radiokomunikacji, Radiofonii i Telewizji, KKRRiT*, Wrocław, Polska, 9-11 Kwiecień 2008, s. 507–510).

Konferencje krajowe

12. M. Bartkowiak i T. Żernicki, "Hybrydowy psychoakustyczny kodek dźwięku dla transmisji strumieniowej", *XI Krajowe Sympozjum Nauk Radiowych URSI*, Poznań, Polska, 7–8 Kwiecień 2005, s. 395–400.

13. M. Bartkowiak i T. Żernicki, "Improved partial tracking technique for sinusoidal modeling of speech and audio", *Poznańskie Warsztaty Telekomunikacyjne - PWT'07*, Poznań, Polska, 2007. [Online]: http://www.multimedia.edu.pl/publications/

4.2. Uzyskane przez autora i nie ujęte w rozprawie oryginalne wyniki naukowe

1. **Autorska implementacja i optymalizacja dekodera fonicznego MPEG-4 AAC HE**. Projekt dekodera sygnałów fonicznych MPEG–4 Advanced Audio Coding High Efficiency dla procesora w architekturze VLIW - praca zespołowa pod kierunkiem prof. dr. hab. inż. M. Domańskiego.

   Wynik: dekoder **zakupiony i wdrożony** w Advanced Digital Broadcast Polska, 2004.

2. **Autorska implementacja i optymalizacja kodeka fonicznego zintegrowanego w systemie telewizji interaktywnej iTVP**. Projekt kodeka sygnałów fonicznych realizowany dla TVP S.A. — praca zespołowa pod kierunkiem prof. dr. hab. inż. M. Domańskiego.

Wynik: efekt pracy **zakupiła** Telewizja Polska S.A., 2003 - 2005.

3. **Autorska implementacja i optymalizacja kodeka wizyjnego zintegrowanego na platformie FPGA**. Projekt: „Wirtualne komponenty elektroniczne dla scalonych koderów i dekoderów wizyjnych niskich przepływności" - praca zespołowa pod kierunkiem prof. dr. hab. inż. M. Domańskiego.

   Projekt celowy KBN Nr 6 T11 2004C/06330, finansowany przez Ministerstwo Nauki i Szkolnictwa Wyższego.

   Wynik: efekt pracy **zakupiony i wdrożony** przez Evatronix S.A, 2005.

4. **Autorska implementacja i optymalizacja kodeka fonicznego wykorzystującego parametryczny opis sygnału**. Projekt: „Nowe techniki parametrycznego kodowania dźwięku szerokopasmowego dla bardzo małych prędkości transmisji" - praca zespołowa pod kierunkiem dr. inż. M. Bartkowiaka.

   Projekt badawczy własny, KBN Nr 3T11D01730, finansowany przez Ministerstwo Nauki i Szkolnictwa Wyższego.

5. **Autorska implementacja i optymalizacja kodeka fonicznego wykorzystującego techniki rozszerzania widma**. Projekt: „Kompresja cyfrowych sygnałów fonicznych z łącznym wykorzystaniem poszerzania widma i modelowania" - praca pod kierunkiem prof. dr. hab. inż. M. Domańskiego.

   Projekt badawczy promotorski, Nr N N516 228435, finansowany przez Ministerstwo Nauki i Szkolnictwa Wyższego.

6. **Prace standaryzacyjne grupy ISO/IEC JTC1/SC29/WG11**

   Propozycja techniki kodowania składowych tonalnych dużych częstotliwości dla nowo powstającego standardu kompresji sygnałów fonicznych MPEG-D Unified Speech and Audio Coding. Dokumenty MPEG: M17914, M18532

   Praca w trakcie realizacji.

4.3. Nagrody

III nagroda w Konkursie Młodych Naukowców za referat:
„Zastosowanie modelowania sinusoidalnego do regeneracji wysokoczęstotliwościowych składowych kodera MPEG-4 HE-AAC",
przedstawiony na Krajowej Konferencji Radiokomunikacji, Radiofonii i Telewizji, KKRRIT-2008, przyznaną przez Fundację Wspierania Rozwoju Radiokomunikacji i Technik Multimedialnych.