



Imię i nazwisko autora rozprawy: mgr inż. Marcin Ciołek
Dyscyplina naukowa: Automatyka i Robotyka

ROZPRAWA DOKTORSKA

Tytuł rozprawy w języku polskim: Usuwanie zakłóceń impulsowych z archiwalnych nagrań fonicznych

Tytuł rozprawy w języku angielskim: Elimination of impulsive disturbances from archive audio recordings

Promotor <i>podpis</i>	Drugi promotor <i>podpis</i>
prof. dr hab. inż. Maciej Niedźwiecki	
Promotor pomocniczy <i>podpis</i>	Kopromotor <i>podpis</i>

Dziękuję mojemu Promotorowi
prof. dr. hab. inż. Maciejowi Niedźwieckiemu
za cenne rady, wsparcie, życzliwość
oraz niezwykłą cierpliwość i wyrozumiałość.

Dziękuję wszystkim pracownikom
Katedry Systemów Automatyki
za życzliwość, pomoc
i wspaniałą atmosferę pracy.

Dziękuję całej mojej Rodzinie i Przyjaciołom
za to, że zawsze mnie wspierali
gdy ich potrzebowałem.

Badania realizowane w ramach tej pracy były częściowo finansowane z projektu NCN nr UMO-2013/09/B/ST7/01582 pt. „Usuwanie zakłóceń impulsowych z sygnałów fonicznych”.

Streszczenie

W niniejszej pracy doktorskiej opisano nowe metody usuwania zakłóceń impulsowych z archiwalnych nagrań dźwiękowych odtworzonych z płyt gramofonowych. Przedstawione w rozprawie rozwiązania, takie jak analiza dwukierunkowa, dwuzakresowe modelowanie autoregresyjne oraz modelowanie wektorowe umożliwiają przetwarzanie zarówno sygnałów muzycznych jak sygnałów mowy przy zachowaniu wysokiej jakości dźwięku po obróbce. Testy odsłuchowe przeprowadzone w grupie 20 słuchaczy potwierdziły, że opracowany algorytm poprawia subiektywnie odczuwalną jakość przetworzonego sygnału. Opracowana metoda pozwala na całkowite lub prawie całkowite wyeliminowanie zakłóceń impulsowych w sposób nie powodujący nowych, niepożądanych efektów dźwiękowych. Otrzymywana jakość zrekonstruowanych nagrań fonicznych jest porównywalna, a w niektórych przypadkach nawet znacząco lepsza od tej uzyskiwanej przez komercyjny system CEDAR, uznawany powszechnie za najlepszy produkt komercyjny służący do rekonstrukcji archiwalnych nagrań dźwiękowych.

Przedstawiona w pracy metoda wraz z wynikami eksperymentów potwierdzają jednoznacznie tezę rozprawy dowodząc, że możliwe jest skuteczne usuwanie zakłóceń impulsowych, a w rezultacie uzyskiwanie wysokiej jakości zrekonstruowanego sygnału fonicznego.

W rozprawie zaprezentowano również podejście oparte na wykorzystaniu wiedzy *a priori* o typowych kształtach i rozmiarach zakłóceń impulsowych do dokładniejszego wykrywania położenia zakłóceń w sygnale. Otrzymane wyniki testów symulacyjnych oraz testów odsłuchowych potwierdzają, że podejście to jest użyteczne w przypadku analizy sygnałów dźwiękowych „w czasie rzeczywistym”.

Spis treści

Spis treści	vii
Notacja i skróty	1
1 Wstęp	3
Cel i teza rozprawy	4
Innowacyjny wkład autora	5
Przegląd treści rozprawy	6
2 Wprowadzenie	9
2.1 Metody odczytu dźwięku: stykowe oraz bezstykowe	9
2.2 Przyczyny degradacji archiwalnych nagrań fonicznych	11
2.3 Przegląd znanych metod eliminacji zakłóceń impulsowych	14
2.4 Podsumowanie	19
3 Wybrane metody eliminacji zakłóceń impulsowych	21
3.1 Model sygnału fonicznego	21
3.2 Rekonstrukcja sygnału	22
3.2.1 Interpolator w wersji wymagającej odwracania macierzy	23
3.2.2 Interpolator w wersji rekurencyjnej	24
3.2.3 Algorytm filtru Kalmana o zmiennym rzędzie	27
3.3 Identyfikacja sygnału fonicznego	28
3.3.1 Ważone estymatory najmniejszych kwadratów	29
3.3.2 Estymatory gradientowe	33
3.3.3 Porównanie estymatorów WLS i LMS	34
3.3.4 Estymator wariancji białego szumu	35
3.3.5 Monitorowanie stabilności modelu AR	36
3.3.6 Eliminacja składowej stałej	38
3.4 Metody oparte na analizie błędów predykcji	38
3.4.1 Detekcja w pętli otwartej	40
3.4.2 Detekcja z decyzyjnym sprzężeniem zwrotnym	41
3.5 Metody oparte na analizie błędów resztowych	44
3.6 Porównanie algorytmów detekcji	47
3.7 Podsumowanie	49



4	Analiza dwukierunkowa	51
4.1	Analiza sygnału w czasie odwrotnym	51
4.2	Jedno- i dwukierunkowa rekonstrukcja sygnału	53
4.3	Dwukierunkowa detekcja zakłóceń impulsowych	54
4.3.1	Technika poszerzania alarmów detekcyjnych	55
4.3.2	Globalne reguły decyzyjne	56
4.3.3	Lokalne reguły decyzyjne	56
4.3.4	Wybór zoptymalizowanego zestawu lokalnych reguł decyzyjnych	60
4.4	Podsumowanie	60
5	Model dwuzakresowy	61
5.1	Autoregresyjny model dwuzakresowy	62
5.2	Model SAR w postaci zdekomponowanej	63
5.3	Rekonstrukcja oparta na modelu SAR	64
5.4	Detekcja oparta na modelu SAR	66
5.5	Identyfikacja modelu SAR	67
5.6	Porównanie modeli AR i SAR	70
5.7	Schemat detekcji zakłóceń impulsowych	72
5.8	Podsumowanie	73
6	Model wektorowy	75
6.1	Identyfikacja sygnału stereofonicznego	75
6.2	Detekcja zakłóceń impulsowych i interpolacja sygnału	82
6.2.1	Opis problemu w przestrzeni stanów	82
6.2.2	Predykcja sygnału i detekcja zakłóceń impulsowych	83
6.2.3	Algorytm filtracji Kalmana o zmiennym rzędzie oparty na modelu VAR	86
6.3	Kwestie implementacyjne	88
6.3.1	Eliminacja składowych stałych	88
6.3.2	Monitorowanie stabilności modelu VAR	89
6.4	Rozszerzenie na analizę dwukierunkową	90
6.5	Rozszerzenie na model dwuzakresowy	91
6.5.1	Ogólny model SVAR	91
6.5.2	Zdekomponowany model SVAR	91
6.5.3	Rekonstrukcja sygnału oparta na modelu SVAR	93
6.5.4	Detekcja oparta na modelu SVAR	94
6.5.5	Identyfikacja modelu SVAR	94
6.5.6	Porównanie modeli VAR i SVAR	96
6.5.7	Wektorowy schemat detekcji zakłóceń impulsowych	96
6.6	Podsumowanie	98
7	Technika dopasowywania wzorców	99
7.1	Wzorce zakłóceń impulsowych	100
7.1.1	Przykładowe zakłócenia impulsowe	100
7.1.2	Analiza podobieństwa kształtu	101
7.1.3	Tworzenie wzorców zakłóceń impulsowych	102
7.2	Lokalizacja zakłóceń o typowych kształtach	104
7.2.1	Wybielanie	104
7.2.2	Wykrywanie wzorców wtórnych	107

7.3	Technika kompensacji	108
7.4	Podsumowanie	111
8	Wyniki eksperymentów	113
8.1	Metodologia	113
8.1.1	Nagrania testowe	114
8.1.2	Obiektywne miary jakości	115
8.1.3	Testy odsłuchowe	118
8.2	Testy symulacyjne	120
8.2.1	Analiza dwukierunkowa – wyniki badań	122
8.2.2	Model dwuzakresowy – wyniki badań	142
8.2.3	Model wektorowy – wyniki badań	153
8.2.4	Technika dopasowywania wzorców – wyniki badań	174
9	Zakończenie	183
	Dodatek A Wybrane operacje na macierzach	185
A.1	Lemat o odwracaniu macierzy	185
	Dodatek B Wyprowadzenie rekurencyjnych algorytmów estymacji	187
B.1	Wyprowadzenie zależności (3.58)-(3.59)	187
B.2	Wyprowadzenie zależności (3.61)	188
	Dodatek C Tabele z wynikami	191
	Dodatek D Zawartość dołączonej płyty CD	213
	Bibliografia	215

Notacja i skróty

Notacja	Opis
t	unormowany (bezwymiarowy) czas dyskretny
$y(t)$	spróbkowany sygnał monofoniczny
$s(t)$	niezakłócony, czysty sygnał monofoniczny
$\delta(t)$	ciąg zakłóceń impulsowych w sygnale $y(t)$
$n(t)$	biały szum tworzący o zerowej wartości oczekiwanej i wariancji ρ
$d(t)$	funkcja lokalizacji zakłóceń impulsowych w sygnale $y(t)$
a_1, \dots, a_r	współczynniki autoregresji
$\mathbf{y}(t) = [y_1(t), y_2(t)]^T$	spróbkowany sygnał stereofoniczny
$\mathbf{s}(t) = [s_1(t), s_2(t)]^T$	niezakłócony, czysty sygnał stereofoniczny
$\boldsymbol{\delta}(t) = [\delta_1(t), \delta_2(t)]^T$	ciąg zakłóceń impulsowych, odpowiednio w lewym i w prawym kanale sygnału $\mathbf{y}(t)$
$\mathbf{n}(t) = [n_1(t), n_2(t)]^T$	dwuwymiarowy biały szum tworzący o zerowej wartości oczekiwanej i macierzy kowariancji $\boldsymbol{\rho}$
$\mathbf{d}(t) = [d_1(t), d_2(t)]^T$	funkcja lokalizacji zakłóceń impulsowych w sygnale $\mathbf{y}(t)$
$\mathbf{A}_1, \dots, \mathbf{A}_r$	macierze współczynników autoregresji o wymiarach 2×2
$\Phi(t)$	blokowa macierz regresji
z^{-1}	operator przesunięcia wstecz
$\lambda, s(t), y(t)$	wartości bądź sygnały skalarne
$\boldsymbol{\varphi}(t), \boldsymbol{\theta}, \mathbf{x}(t)$	wektory
$\mathbf{A}, \mathbf{P}(t)$	macierze
$\mathcal{Y}, Y_\delta, Y_s$	zbiory danych
$\hat{\boldsymbol{\theta}}, \hat{d}(t)$	estymaty
\mathbf{I}_q	macierz jednostkowa o wymiarze $q \times q$
$\mathbf{0}_q$	wektor zerowy o wymiarze $q \times 1$
\mathbf{O}_q	macierz zerowa o wymiarze $q \times q$
$A \cup B$	suma zbiorów
$A \cap B$	iloczyn zbiorów
$A \subset B$	zbiór A jest podzbiorem zbioru B
\emptyset	pusty zbiór
$p(\cdot)$	funkcja gęstości rozkładu prawdopodobieństwa
$ \cdot $	wartość bezwzględna
$\ \cdot\ $	norma euklidesowa
$E(\cdot)$	wartość oczekiwana
$\overline{(\cdot)}$	lokalne uśrednianie po czasie
$\text{var}(\cdot)$	wariancja



Notacja	Opis
$\text{cov}(\cdot)$	macierz kowariancji
$\text{det}(\cdot)$	wyznacznik macierzy
$\text{tr}(\cdot)$	śląd macierzy
$\nabla(\cdot)$	gradient funkcji
\equiv	operator równości tożsamościowej
\cong	operator przybliżonej równości tożsamościowej

Skróty	Opis
AR	model autoregresyjny (z ang. <i>autoregressive</i>)
VAR	wektorowy model autoregresyjny (z ang. <i>vector autoregressive</i>)
ARX	model autoregresyjny z zewnętrznym wejściem (z ang. <i>autoregressive with exogenous input</i>)
VARX	wektorowy model autoregresyjny z zewnętrznym wejściem (z ang. <i>vector autoregressive with exogenous input</i>)
SAR	autoregresyjny model dwuzakresowy, należący do klasy tzw. modeli rzadkich (z ang. <i>sparse autoregressive</i>)
SVAR	wektorowy autoregresyjny model dwuzakresowy (z ang. <i>sparse vector autoregressive</i>)
WLS	metoda ważonych najmniejszych kwadratów (z ang. <i>weighted least squares</i>)
EWLS	metoda ważonych najmniejszych kwadratów w wersji z oknem wykładniczym (z ang. <i>exponentially weighted least squares</i>)
SWLS	metoda ważonych najmniejszych kwadratów w wersji z oknem prostokątnym (z ang. <i>sliding window least squares</i>)
LMS	metoda najmniejszych średnich kwadratów (z ang. <i>least mean squares</i>)
NLMS	unormowany algorytm najmniejszych średnich kwadratów (z ang. <i>normalized least mean squares</i>)
CML	warunkowy algorytm największej wiarygodności (z ang. <i>conditional maximum likelihood</i>)
LD	algorytm Levinsona-Durbina wykorzystywany do rozwiązywania układu równań Yule'a-Walkera
WWR	algorytm Whittle'a-Wigginsa-Robinsona, wielowymiarowe rozszerzenie algorytmu Levinsona-Durbina
KF	algorytm filtracji Kalmana (z ang. <i>Kalman filter</i>)
PEAQ	algorytm obiektywnej oceny jakości nagrania fonicznego (z ang. <i>Perceptual Evaluation of Audio Quality tool</i>)

Rozdział 1

Wstęp

Archiwa Polskiego Radia¹ i Polskiej Biblioteki Narodowej² zgromadziły dotychczas łącznie ponad milion archiwalnych dokumentów dźwiękowych i audiowizualnych z lat 1889 – 2007, utrwalających między innymi: historyczne przemówienia, wywiady, audycje radiowe, koncerty i studyjne nagrania muzyczne. W zbiorze tym możemy odnaleźć bardzo cenne nagrania zapisane na historycznych nośnikach dźwięku takich jak wałki fonograficzne, rolki pianolowe, płyty patefonowe i gramofonowe czy też taśmy magnetyczne. Dla porównania, Archiwum Brytyjskiej Biblioteki Dźwięku³, które posiada jedną z największych kolekcji dokumentów dźwiękowych na świecie, zawiera już ponad 3 miliony nagrań, w tym ponad milion płyt i ponad 200 tys. taśm magnetofonowych. Cyfrowa archiwizacja dokumentów dźwiękowych jest szczególnie ważna w przypadku nagrań unikatowych, cennych, zniszczonych lub wrażliwych na negatywne skutki ich udostępniania. W Polsce cyfrową archiwizacją takich dokumentów zajmuje się Narodowe Archiwum Cyfrowe⁴. Kolejnym, oczywistym etapem będzie cyfrowa rekonstrukcja dokumentów dźwiękowych. Dlatego badania nad technologią rekonstrukcji archiwalnych nagrań fonicznych są ważne z praktycznego punktu widzenia i wciąż aktualne.

Archiwalne nagrania foniczne zawierają szum szerokopasmowy oraz zakłócenia impulsowe. Obecność tych zakłóceń w nagraniu zmniejsza zrozumiałość przekazywanych treści, a w przypadku nagrań muzycznych dodatkowo utrudnia lub wręcz uniemożliwia ich właściwy odbiór estetyczny. Dlatego usuwanie obu typów zakłóceń w procesie rekonstrukcji archiwalnych nagrań fonicznych jest ważnym elementem ochrony naszego dziedzictwa kulturowego.

Niniejsza praca doktorska koncentruje się na opisie nowych podejść do automatycznej eliminacji zakłóceń impulsowych z archiwalnych nagrań dźwiękowych odtworzonych z płyt gramofonowych. Choć w literaturze zaproponowano szereg metod pozwalających na wykrywanie i usuwanie tego typu zakłóceń, niewiele z nich może być stosowanych z powodzeniem do przetwarzania zarówno sygnałów muzycznych jak sygnałów mowy. Głównymi wyzwaniem dla algorytmu eliminacji zakłóceń impulsowych pozostają wciąż: niestacjonarność przetwarzanego sygnału oraz jego „dwuzakresowa” struktura korelacyjna. Chęć sprostania tym wyzwaniom często wiąże się z generowaniem dużej liczby fałszywych wykryć przez detektor zakłóceń impulsowych. Konsekwencją tego, w przypadku niezadowalającej rekonstrukcji sygnału w miejscach fałszywych wykryć, jest powstawanie nowych efektów dźwiękowych, które obniżają jakość rekonstruowanego nagrania. Oprócz tego, do niepożądanych wyników przetwarzania sygnału można zaliczyć

¹Oficjalna strona: www2.polskieradio.pl/archiwum/

²Oficjalna strona: www.bn.org.pl

³Oficjalna strona: <http://sounds.bl.uk/>

⁴Oficjalna strona: www.nac.gov.pl



pomijanie mniejszych zakłóceń impulsowych oraz niedokładne wykrywanie tych większych, co również obniża jakość rekonstruowanego nagrania. Całkowite lub prawie całkowite wyeliminowanie zakłóceń impulsowych z sygnałów fonicznych bez wprowadzania nowych, niepożądanych efektów dźwiękowych nie jest zadaniem trywialnym.

Cel i teza rozprawy

Celem niniejszej rozprawy jest opracowanie skutecznej, uniwersalnej metody eliminacji zakłóceń impulsowych z archiwalnych nagrań dźwiękowych odtworzonych z płyt gramofonowych. Metoda taka powinna umożliwiać przetwarzanie zarówno sygnałów muzycznych jak sygnałów mowy przy zachowaniu wysokiej jakości dźwięku po obróbce. W przypadku nagrań zarejestrowanych w wersji stereofonicznej, metoda ta powinna być oparta na łącznej analizie lewego i prawego kanału stereofonicznego w celu uzyskania możliwie najlepszych rezultatów. Opracowany algorytm powinien poprawiać subiektywnie odczuwalną jakość przetworzonego sygnału poprzez całkowite lub prawie całkowite wyeliminowanie zakłóceń impulsowych w sposób nie powodujący nowych, niepożądanych efektów dźwiękowych.

Otrzymane wyniki badań pozwoliły na postawienie następującej tezy rozprawy:

Zastosowanie techniki dwuzakresowego modelowania autoregresyjnego w połączeniu z techniką dwukierunkowego przetwarzania sygnału fonicznego pozwala na skuteczne eliminowanie zakłóceń impulsowych, a w rezultacie na uzyskanie wysokiej jakości zrekonstruowanego sygnału.

Aby uzasadnić przyjętą tezę określono następujące cele szczegółowe:

- Opracowanie podejścia pozwalającego na łączenie wyników detekcji otrzymywanych na drodze dwukierunkowej analizy sygnału: od pierwszej do ostatniej próbki oraz od próbki ostatniej do pierwszej (tj. w odwróconym porządku czasowym).
- Opracowanie podejścia opartego na dwuzakresowym autoregresyjnym modelowaniu sygnału.
- Opracowanie podejścia opartego na łącznej analizie obu kanałów nagrania stereofonicznego przy użyciu techniki wektorowego modelowania autoregresyjnego.
- Opracowanie podejścia wykorzystującego wiedzę *a priori* o typowych kształtach i rozmiarach zakłóceń impulsowych.
- Symulacyjna i eksperymentalna weryfikacja opracowanych podejść.



Innowacyjny wkład autora

Do najważniejszych elementów będących innowacyjnym wkładem autora należą:

- Opracowanie metody eliminacji zakłóceń impulsowych z archiwalnych nagrań fonicznych opartej na dwukierunkowym przetwarzaniu sygnału fonicznego (patrz rozdział 4). W ramach tego podejścia zaproponowano eksperymentalną procedurę wyboru reguł najlepiej „dopasowanych” do danego schematu detekcji (patrz punkt 8.2.1). Uzyskane wyniki opublikowano w czasopiśmie z listy JCR [1] oraz w dwóch referatach konferencyjnych ([2], [3]).
- Opracowanie metody eliminacji zakłóceń impulsowych opartej na dwuzakresowym modelowaniu sygnału fonicznego (patrz rozdział 5). W ramach tego podejścia zaproponowano również schemat kooperacyjnej detekcji zakłóceń impulsowych zmniejszający liczbę fałszywych alarmów (patrz podrozdział 5.7). Uzyskane wyniki opublikowano w czasopiśmie z listy JCR [1] oraz w trzech referatach konferencyjnych ([2], [4], [5]).
- Opracowanie podejścia opartego na wektorowym modelowaniu autoregresyjnym (patrz rozdział 6). Rozszerzenie techniki dwukierunkowej analizy sygnału (patrz podrozdział 6.4) oraz dwuzakresowego modelowania autoregresyjnego (patrz podrozdział 6.5) na przypadek wektorowy. Otrzymane wyniki opublikowano w czasopiśmie z listy JCR [6] oraz w dwóch referatach konferencyjnych ([7], [8]).
- Opracowanie podejścia opartego na dopasowywaniu wzorców zakłóceń impulsowych (patrz rozdział 7). Opracowanie procedury tworzenia bazy wzorców zakłóceń (patrz podrozdział 7.1). Opracowanie podejścia do detekcji zakłóceń impulsowych opartego na połączeniu techniki filtracji dopasowanej z techniką detekcji wykorzystującą adaptacyjne modelowanie autoregresyjne (patrz podrozdział 7.2). Porównanie dwóch technik rekonstrukcji sygnału: klasycznej, opartej na interpolacji sygnału oraz nowej, opartej na kompensacji zakłóceń impulsowych poprzez odjęcie od zakłóconego sygnału odpowiednio przeskalowanych wzorców zakłóceń (patrz podrozdział 7.3). Otrzymane wyniki opublikowano w czasopiśmie z listy JCR [9] oraz w jednym referacie konferencyjnym ([10]).

Przegląd treści rozprawy

Rozprawa została podzielona na dziewięć rozdziałów.

W drugim rozdziale omówiono różne sposoby zapisu i odczytu dźwięku oraz główne przyczyny powstawania zakłóceń impulsowych. Rozdział kończy się przeglądem znanych metod eliminacji zakłóceń impulsowych. Krytyczna analiza tych metod uzasadnia potrzebę opracowania nowych, skuteczniejszych podejść, które będą cechować się dokładniejszą detekcją zakłóceń i lepszą rekonstrukcją sygnału.

W trzecim rozdziale szerzej opisano wybrane podejścia do detekcji zakłóceń impulsowych, tj. metody oparte na analizie błędów predykcji oraz metody oparte na analizie błędów resztowych. Rozdział kończy się wskazaniem ograniczeń omówionych metod detekcji.

Główną część pracy stanowią cztery kolejne rozdziały. W rozdziale czwartym przedstawiono nowe podejście do detekcji zakłóceń impulsowych oparte na dwukierunkowej analizie sygnału. Tradycyjne metody wykorzystujące detektor oparty na analizie błędów predykcji działają zadowalająco gdy charakterystyki sygnału zmieniają się wolno w czasie, ale mogą zawieść w obecności ich nagłych zmian. Jedną z kluczowych obserwacji jest to, że zdarzenia niedające się przewidzieć na podstawie lewostronnej analizy sygnału, tj. w kierunku od jego pierwszej do ostatniej próbki, są na ogół przewidywalne gdy analiza prowadzona jest w czasie odwrotnym. W proponowanym podejściu zastosowano kontekstowe reguły łączenia wyników detekcji otrzymywanych na drodze dwukierunkowego przetwarzania sygnału. Aby wybrać zestaw reguł, który najlepiej pasuje do danego schematu detekcji, każda kontekstowa reguła sprawdzona została w sposób doświadczalny.

W rozdziale piątym omówiono problem eliminacji zakłóceń impulsowych z archiwalnych sygnałów mowy i sygnałów muzycznych z udziałem śpiewu i/lub instrumentów dętych takich jak trąbka, saksofon czy klarnet. Naturalnym podejściem do opisu struktury korelacyjnej tych sygnałów jest użycie autoregresyjnego modelu dwuzakresowego. Stosowane do tej pory metody identyfikacji modelu dwuzakresowego nie gwarantują jego stabilności, co może rodzić szereg problemów na etapie detekcji zakłóceń i interpolacji sygnału. Dlatego przedstawiono nowe podejście, w którym sygnał reprezentowany jest przez model dwuzakresowy identyfikowany jako kaskadowe połączenie filtra formantowego i filtra wysokości dźwięku. Technika ta pozwala na zapewnienie stabilności tworzonego modelu.

Rozdział szósty poświęcono analizie sygnałów stereofonicznych. W rozdziale tym omówiono nowe podejście do detekcji uszkodzonych próbek sygnału. Podejście to oparte jest na łącznej analizie lewego i prawego kanału stereofonicznego przy użyciu techniki wektorowego modelowania autoregresyjnego. Ponieważ zakłócenia impulsowe mogą być asymetrycznie ułożone w lewym i prawym kanale, nieuszkodzony materiał w jednym kanale może być pomocny w wykrywaniu i interpolowaniu uszkodzonych próbek w kanale drugim. Rozdział ten zawiera także rozszerzenie techniki modelowania dwuzakresowego oraz techniki analizy dwukierunkowej na przypadek wektorowy.

W rozdziale siódmym przedstawiono nowe podejście do problemu lokalizacji zakłóceń impulsowych w sygnałach fonicznych, oparte na połączeniu techniki filtracji dopasowanej z techniką detekcji wykorzystującą adaptacyjne modelowanie autoregresyjne. Podejście to wykorzystuje w procesie detekcji wiedzę *a priori* o typowych kształtach i rozmiarach zakłóceń impulsowych. Okazuje się, że duża liczba zakłóceń impulsowych daje się podzielić ze względu na podobieństwo kształtu na stosunkowo niewielką liczbę grup. W wyniku analizy takiej grupy podobnych zakłóceń impulsowych można zaproponować reprezentujący je wzorzec zakłócenia. W ramach nowego podejścia zaproponowano procedurę tworzenia bazy wzorców zakłóceń impulsowych.



Proponowane podejście wykorzystuje klasyczną metodę filtracji dopasowanej do lokalizacji zakłóceń impulsowych podobnych do wzorców zakłóceń. W celu wykrycia zakłócenia impulsowego wyznacza się splot odpowiednio zmodyfikowanych wzorców z sekwencją błędów jednokrokowej predykcji otrzymaną za pomocą predyktora opartego na modelu autoregresyjnym, a następnie „proguje” się otrzymane wyniki. Ważną cechą proponowanego podejścia jest jego zdolność adaptacji do źródła zakłóceń impulsowych, wzorce zakłóceń mogą być bowiem tworzone dla konkretnego nagrania otrzymanego przy użyciu konkretnego sprzętu fonicznego.

Rozdział ósmy zawiera opis i wyniki eksperymentów z użyciem opracowanych technik eliminacji zakłóceń impulsowych. Omówiono w nim przyjętą w pracy metodologię przeprowadzania eksperymentów oraz metodologię oceny otrzymywanych wyników. Jakość zrekonstruowanych nagrań oceniono przy użyciu miar obiektywnych oraz w oparciu o wyniki testów odsłuchowych.

Rozprawę kończy dyskusja i podsumowanie uzyskanych wyników, wraz ze wskazaniem możliwości rozszerzenia zaproponowanych metod.

Rozprawa zawiera cztery dodatki. W dodatku A umieszczono wybrane operacje na macierzach. Dodatek B zawiera wyprowadzenia rekurencyjnych algorytmów estymacji. W dodatku C zamieszczono tabele z wynikami. W dodatku D opisano zawartość dołączonej płyty CD.

Rozdział 2

Wprowadzenie

W rozdziale tym omówiono różne sposoby zapisu i odczytu dźwięku oraz główne przyczyny powstawania zakłóceń impulsowych, związane z degradacją powierzchni nośnika. Przedstawiono klasyfikację zakłóceń impulsowych ze względu na ich różną amplitudę, długość oraz częstotliwość występowania w sygnale. Przedyskutowano również inne czynniki, które mogą silnie wpływać na kształt oraz czas trwania zakłóceń impulsowych. Rozdział kończy się przeglądem znanych metod eliminacji zakłóceń impulsowych, zarówno analogowych jak cyfrowych.

2.1 Metody odczytu dźwięku: stykowe oraz bezstykowe

Dźwięk zapisany na archiwalnych nośnikach analogowych, takich jak wałki fonograficzne, płyty patefonowe oraz płyty gramofonowe, może być odczytany przy użyciu tradycyjnych metod stykowych (mechanicznych) bądź nowszych metod bezstykowych (optycznych).

Metody stykowego odczytu dźwięku są bezpośrednio związane ze sposobem jego zapisu. W przypadku wymienionych wyżej nośników, dźwięk odczytywany jest poprzez zamianę drgań końcówki (igły lub – w przypadku płyt patefonowych – kulki), umieszczonej w rowku obracającej się płyty/wałka, na impulsy elektryczne. Końcówka wprawiana jest w ruch drgający przez rzeźbę rowka. Drgania końcówki umożliwiają odtwarzanie sygnału akustycznego zapisanego na nośniku. W przypadku płyt gramofonowych zapisu dokonuje się na bocznych ścianach rowka (tzw. technika zapisu w bocznego), zaś w przypadku wałków fonograficznych oraz płyt patefonowych zapis realizowany był poprzez zmianę głębokości rowka (tzw. technika zapisu wgłębnego). Jakość odczytanego nagrania zależy od jakości wykonania sprzętu odgrywającego, warunków jego pracy, jakości i doboru odpowiedniej końcówki (np. w zależności od typu płyty gramofonowej dobierana jest igła o odpowiednich parametrach [11]) oraz, przede wszystkim, od stopnia degradacji nośnika. Na stan nośnika wpływają wszelkiego rodzaju zabrudzenia, pęknięcia materiału, rysy, obtarcia, drobne ubytki materiału oraz zużycie. W skrajnych przypadkach zły stan nośnika nie pozwala nawet na odtworzenie dźwięku.

Idea bezstykowego odczytu dźwięku zaproponowana została już w latach 80. ubiegłego stulecia [12]. Podejście to polegało na wykorzystaniu wiązki lasera, zamiast tradycyjnej igły, do odczytu dźwięku zapisanego na zboczach rowka płyty gramofonowej. Komercyjne gramofony firmy ELP wyposażone w laser dostępne są w sprzedaży już od 1997 roku [13], [14]. Urządzenie to odtwarza większość typów płyt gramofonowych, tj. odtwarza płyty koloru czarnego (istnieją jeszcze płyty kolorowe i/lub z nadrukami), okrągłe (są jeszcze nagrania zarejestrowane na folii, którą



owijano zwykle prostokątne pocztówki) o średnicy od 7 do 12 cali (są jeszcze płyty o średnicy 5, 6 oraz 16 cali). Jakość odtwarzanych nagrań jest znacznie lepsza niż w przypadku użycia tradycyjnego gramofonu. Niestety, technologia odczytu jest bardzo czuła na zabrudzenia powierzchni płyty. Wiązka lasera reaguje na każdy pyłek kurzu i każdą grudkę brudu, które znajdują się w jej zasięgu, generując zakłócenia impulsowe, w przeciwieństwie do tradycyjnej igły, która jest mniej wrażliwa na tego rodzaju zabrudzenia. W sytuacji gdy wiązka lasera napotyka na swojej drodze ubytek materiału spowodowany rysą, pęknięciem czy innym uszkodzeniem, powstaje „nieciągłość” w odczytywanym sygnale, która może być traktowana jako zakłócenie impulsowe. Dlatego odczytane nagrania wymagają dalszej rekonstrukcji przy użyciu algorytmów przetwarzania sygnału w celu eliminacji zakłóceń impulsowych. Pewną zaletą odczytu optycznego jest to, że powstałe zakłócenia impulsowe obejmują znacznie mniej próbek (nie występują drgania układu odczytującego) niż analogiczne zakłócenia impulsowe powstałe przy odczycie stykowym.

Nowsze metody bezstykowego odczytu dźwięku opierają się na analizie zeskanowanego obrazu rzeźby rowka. Ciekawym rozwiązaniem jest zastosowanie optycznego skanera I.R.E.N.E (z ang. *Image, Reconstruct, Erase Noise, Etc.*) do rekonstrukcji nagrań zarejestrowanych na płytach gramofonowych z wykorzystaniem technik przetwarzania obrazu. Podejście to opisano w pracy [15]. Zaproponowany tam algorytm odtwarza sygnał na podstawie analizy położenia dna rowka. Jakość odtwarzanego sygnału zależy od zdolności algorytmu do śledzenia dna rowka oraz od liczby lokalnych zakłóceń (których źródłem mogą być np. pyłki kurzu osiadłe na dnie rowka) zniekształcających pozyskiwany obraz. Podejście to wprawdzie redukuje liczbę zakłóceń impulsowych w odtwarzanym nagraniu, lecz dzieje się to kosztem utraty jakości nagrania w porównaniu z metodą odczytu stykowego. Inne podejścia do analizy zeskanowanego obrazu powierzchni płyty, uwzględniające boczny profil rowka, opisane zostały w pracach [16], [17], [18], [19]. W podejściach tych algorytm analizuje ściany rowka, które są znacznie bogatszym źródłem informacji o charakterystyce dźwięku, niż dno rowka. Dalsze prace nad rozwojem technik bezstykowego odczytu doprowadziły do opracowania metody rekonstrukcji nagrania w oparciu o analizę zeskanowanego profilu 3D powierzchni nośnika (wałki fonograficzne, płyty patefonowe oraz płyty gramofonowe) [20], [21]. Metody te pozwalają na wykorzystanie pełnej informacji zarejestrowanej w wyżłobionych rowkach zarówno techniką zapisu w bocznego jak i w głębokiego. Jakość odtwarzanego nagrania zależy od jakości uzyskanego obrazu/profilu 3D powierzchni nośnika, sposobu odczytu sygnału fonicznego na podstawie analizy obrazu oraz zastosowanych metod rekonstrukcji. Zaletą tego podejścia jest to, że eliminacja zakłóceń impulsowych może być przeprowadzona zarówno na etapie przetwarzania obrazu jak i na etapie przetwarzania sygnału fonicznego.

Metody stykowego i bezstykowego odczytu dźwięku są wrażliwe na kurz i brud zalegający na dnie/ścianach bocznych rowka. Jeżeli jest taka możliwość, nośniki dźwięku powinny zostać dokładnie oczyszczone przed rozpoczęciem odczytywania/odtworzenia nagrania. W przypadku gramofonu, kilkukrotne odtworzenie nagrania przy użyciu odpowiednio dobranych igieł może pomóc w oczyszczeniu rowka z zalegającego w nim kurzu [11]. W sytuacjach, gdy nośnik jest już poważnie uszkodzony nie zaleca się nawet czyszczenia jego powierzchni, a w skrajnych przypadkach nawet jego odtwarzania przy użyciu metod stykowych. W tych sytuacjach metody bezstykowe mogą być dalej stosowane, co potwierdza ich uniwersalność i użyteczność.

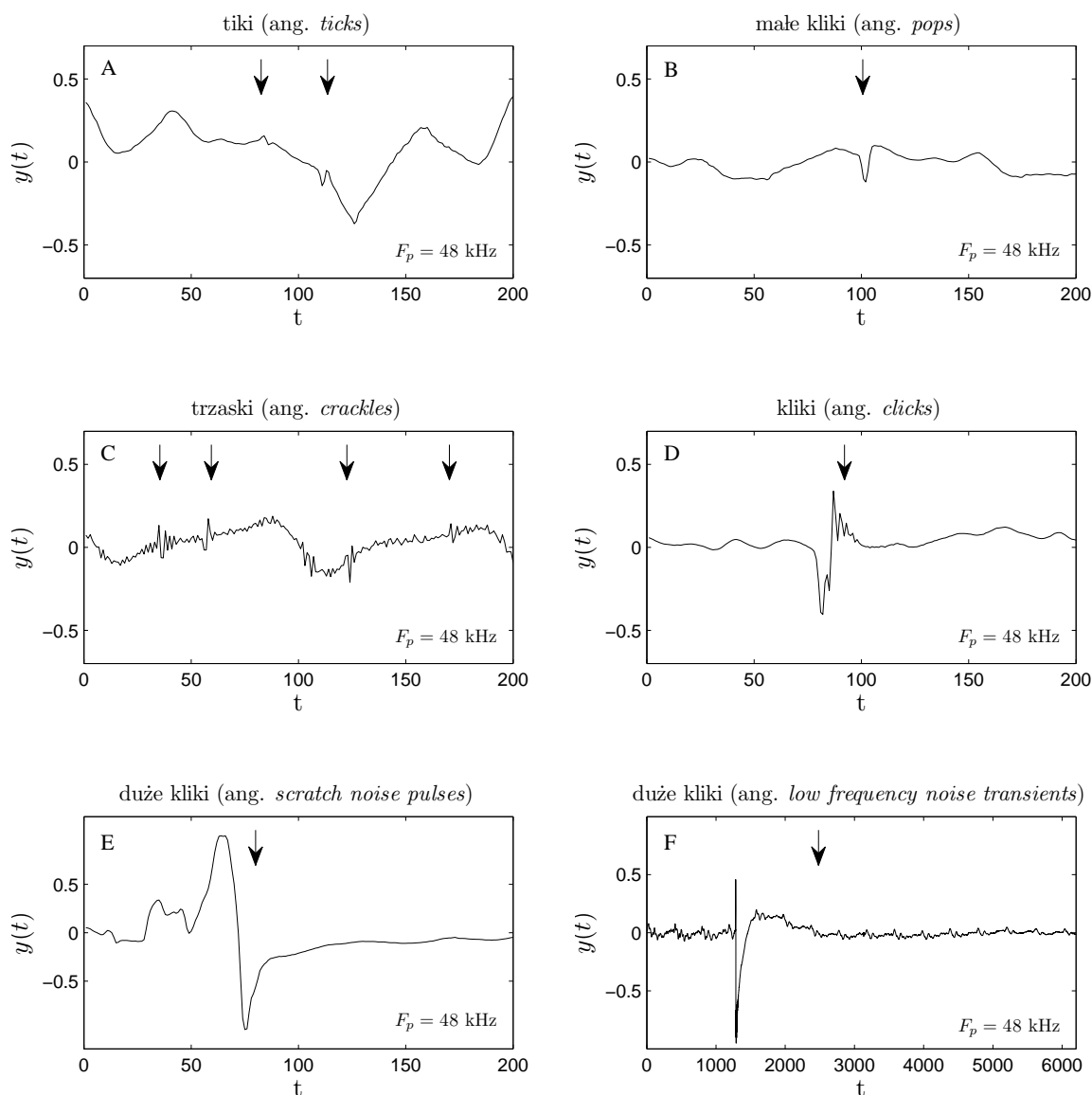
2.2 Przyczyny degradacji archiwalnych nagrań fonicznych

Zakłócenia impulsowe pojawiające się w odtwarzanych archiwalnych nagraniach fonicznych związane są najczęściej z uszkodzeniem powierzchni nośnika dźwięku. Powierzchnia nośnika ulega degradacji w wyniku procesu starzenia, negatywnego wpływu warunków atmosferycznych (temperatura, wilgotność), niewłaściwego sposobu obchodzenia się z nośnikiem, naturalnego zużycia wynikającego z wielokrotnego odtwarzania, błędów na etapie tłoczenia płyty (pęcherzyki gazu) czy nawet w wyniku występowania pewnego rodzaju grzybów, żywiących się składnikiem mineralnym, z którego wykonana jest płyta. Przykładami degradacji powierzchni płyty gramofonowej mogą być zakurzenie, zabrudzenia, zarysowania, pęknięcia czy też drobne ukruszenia materiału. Wymienione uszkodzenia wpływają bezpośrednio na częstotliwość występowania zakłóceń impulsowych w nagraniu, ich amplitudę oraz długość. Stąd, w zależności od typu zakłócenia impulsowego, używa się różnych określeń, takich jak tiki, trzaski czy kliki. Terminy te, niezbyt dobrze „zadomowione” w krajowej literaturze naukowej dotyczącej przedmiotu, mają swoje pierwowzory w literaturze angielskiej [11], [22],[23], [24], [25]:

- ang. *ticks* (tiki) – bardzo krótkie zakłócenia impulsowe o względnie małej amplitudzie. Mogą pojawiać się w niewielkich odstępach. Przyczyną ich powstawania są zabrudzenia (kurz, pył) lub mikrouszkodzenia powierzchni płyty gramofonowej.
- ang. *pops* (małe kliki) – krótkie zakłócenia impulsowe pojawiające się w większych odstępach czasowych, zazwyczaj o unimodalnym kształcie. Przyczyną ich powstawania są na przykład drobne rysy na powierzchni płyty gramofonowej.
- ang. *crackles* (trzaski) – krótkie zakłócenia impulsowe pojawiające się w niewielkich odstępach czasowych, o względnie małej amplitudzie. Ten rodzaj zakłóceń jest typowy dla bardzo starych (produkowanych do lat 30. XX wieku) płyt gramofonowych. Zakłócenie te charakteryzowane są często jako skwierczenie/trzeszczenie.
- ang. *clicks* (kliki) – zakłócenia impulsowe pojawiające się w większych odstępach czasowych, o złożonym przebiegu i dużej amplitudzie. Przyczyną ich powstawania są uszkodzenia (np. rysy) widoczne na powierzchni płyty gramofonowej.
- ang. *low frequency noise transients/scratch noise pulses* (duże kliki) – długie zakłócenia impulsowe, których źródłem są głębokie rysy lub pęknięcia widoczne na powierzchni płyty gramofonowej. Napotkane uszkodzenia wprawiają głowicę wraz z igłą w ruch drgający, będący – z systemowego punktu widzenia – jej odpowiedzią impulsową. W typowym przebiegu zakłócenia impulsowego wyróżnić możemy dwie fazy: w pierwszej fazie drgania układu odczytującego są najsilniejsze i skutkują powstaniem zasadniczej części zakłócenia impulsowego o dużej amplitudzie i nieliniowej charakterystyce, po czym następuje faza relaksacji drgań skutkująca powstaniem zanikającego przebiegu oscylacyjnego o niskiej częstotliwości i liniowej charakterystyce. Zakłócenia te charakteryzowane są często jako „głuche tąpnięcia”.

Wymienione typy zakłóceń impulsowych zilustrowano na rysunku 2.1.

Oprócz wspomnianych wyżej przyczyn powstawania zakłóceń impulsowych, należy również wziąć pod uwagę to, że kształt oraz czas trwania zakłóceń impulsowych mogą silnie zależeć od takich czynników jak [11]:



Rysunek 2.1: Przykłady różnych typów zakłóceń impulsowych znalezionych w archiwalnych nagraniach gramofonowych. Od góry, na kolejnych wykresach zostały wskazane zakłócenia impulsowe: tiki (A), małe kliki (B), trzaski (C), kliki (D) oraz duże kliki (E,F). W dolnej części każdego wykresu podana została częstotliwość próbkowania sygnału F_p .

- *typ nośnika*

Zakłócenia impulsowe nazywane trzaskami są typowe dla nośników takich jak płyty gramofonowe wykonane z szelaku oraz cylindry akustyczne. W przypadku płyt gramofonowych, wykonanych z tworzywa sztucznego – polichlorku winylu (tzw. płyta winylowa) – typowymi zakłóceniami impulsowymi są tiki oraz kliki. Materiał ten charakteryzuje się mniejszą ziarnistością oraz większą odpornością na czynniki zewnętrzne, niestety jest też bardziej podatny na zarysowania.

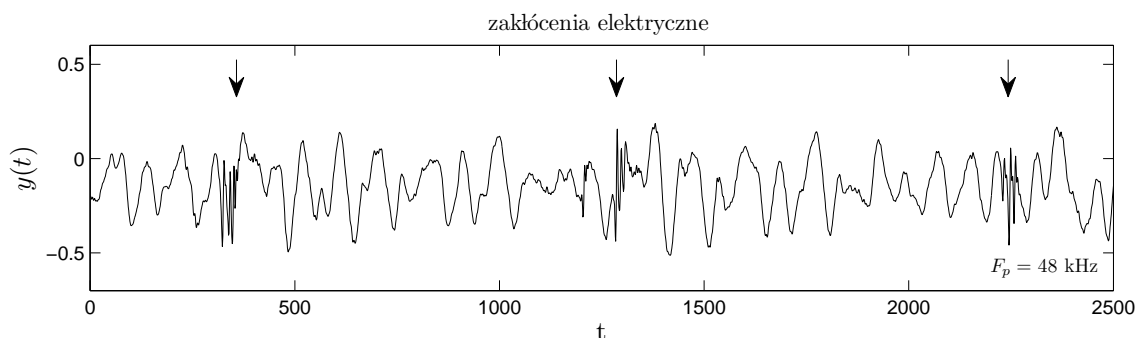


- *użyty sprzęt odgrywający*
Przykładowo, w przypadku użycia gramofonu ważne są tryb pracy przedwzmacniacza, prędkość obrotu talerza i typ igły gramofonowej.
- *sposób zamiany na postać cyfrową*
Wybór odpowiedniej częstotliwości próbkowania oraz rozdzielczości może mieć wpływ na przebieg rekonstrukcji sygnału fonicznego. Częstotliwości próbkowania 11,025 kHz oraz 22,05 kHz mogą zostać użyte w przypadku materiału fonicznego, w którym zawarty jest wyłącznie sygnał mowy, zaś w przypadku sygnału muzycznego zalecane są wyższe częstotliwości próbkowania 44,1 kHz oraz 48 kHz, a nawet 96 kHz. Należy jednak pamiętać, że wraz ze wzrostem częstotliwości próbkowania zakłócenia impulsowe obserwowane w sygnale fonicznym tracą swój impulsowy charakter, co może powodować problemy z ich wykrywaniem.

Podczas zgrywania materiału fonicznego, oprócz wymienionych, typowych zakłóceń impulsowych, mogą wystąpić różne lokalne zniekształcenia sygnału, o których warto wspomnieć, mimo iż nie są one przedmiotem rozprawy:

- *przeskok igły gramofonowej na sąsiedni rowek*
W przypadku przeskoku na wewnętrzny rowek mamy do czynienia z brakiem części materiału fonicznego. Rekonstrukcja sygnału może polegać na skopiowaniu i wklejenia fragmentu podobnego w miejsce, w którym wystąpiła nieciągłość sygnału fonicznego. W przypadku przeskoku na zewnętrzny rowek pojawi się zdublowany materiał foniczny. Rekonstrukcja sygnału może polegać na „wycięciu” powtarzającego się fragmentu, a następnie scaleniu ze sobą dwóch części sygnału fonicznego.
- *nierównomierność odczytu*
Ze względu na ziarnistą strukturę materiału, z którego wykonana jest płyta gramfonowa, na ograniczenia mechaniczne/elektryczne powodujące niewielkie zmiany prędkości obrotu talerza oraz ze względu na lokalne uszkodzenia rowka, prędkość igły przemieszczającej się wzdłuż rowka nie jest stała. Może się zdarzyć, że igła gramofonowa przeskoczy do przodu pomijając część materiału fonicznego zapisanego na płycie. W zarejestrowanym sygnale zaobserwujemy wówczas nieciągłość związaną z brakującym blokiem próbek. Rekonstrukcja sygnału może polegać na „wstawieniu” w miejsce brakujących próbek fragmentu podobnego lub zastąpieniu bloku brakujących próbek interpolacją opartą na modelu sygnału.
- *zakłócenia elektryczne przy nagrywaniu i/lub transmisji*
W przypadku magnetofonu, zakłócenia elektryczne mogą wynikać z silnego oddziaływania sieci elektrycznej przy nagrywaniu materiału fonicznego z radia. Powstałe lokalne zakłócenia impulsowe będą charakteryzowały się wysokoczęstotliwościowym oscylacyjnym przebiegiem – patrz rysunek 2.2. Dodatkowo, częstotliwość występowania zakłóceń impulsowych może odpowiadać częstotliwości sygnału w sieci energetycznej, tj. 50 Hz w Europie oraz 60 Hz w USA. Zakłócenia te postrzegane są w nagraniu jako brzęczenie. Zakłócenia elektryczne mogą powstawać też przy transmisji sygnałów, a ich przyczyną mogą być na przykład wyładowania atmosferyczne występujące podczas burzy.





Rysunek 2.2: Przykład zakłóceń elektrycznych powstałych w wyniku oddziaływania sieci elektrycznej przy nagrywaniu utworu z radia. Zakłócenia te postrzegane są w nagraniu jako brzęczenie. W dolnej części wykresu podana została częstotliwość próbkowania sygnału F_p .

2.3 Przegląd znanych metod eliminacji zakłóceń impulsowych

Techniki eliminacji zakłóceń impulsowych można podzielić na analogowe oraz cyfrowe.

Techniki analogowe stosowano odkąd wprowadzono do użytku taśmy magnetyczne. Usuwanie zakłóceń impulsowych polegało na ręcznym wycięciu fragmentu taśmy i ponownym jej sklejeniu. W przypadku płyt gramofonowych praktykowało się rekonstrukcję ścian rowków uszkodzonych przez głębokie rysy. Rekonstrukcja ta polegała na ręcznym uzupełnieniu brakującego materiału i nadaniu mu odpowiedniego kształtu. Technika ta była czasochłonna dlatego stosowana była tylko w przypadku unikatowych płyt. Znane są również automatyczne metody redukcji zakłóceń impulsowych polegające na zastosowaniu dodatkowego układu elektronicznego na wyjściu gramofonu [26], [27], [28], [29], [30]. Układ ten wykrywał zakłócenia impulsowe, a następnie tłumił ich skutki. Testowano różne techniki tłumienia w miejscu wykrycia zakłócenia impulsowego, polegające na: chwilowym wytworzeniu sygnału na poziomie bezpośrednio poprzedzającym zakłócenia, zastosowaniu liniowej interpolacji sygnału, wprowadzeniu dodatkowej linii z opóźnionym sygnałem i przełączaniu pomiędzy dwiema liniami, przełączaniu między lewym, a prawym kanałem nagrania odtwarzanego przez gramofon z wkładką stereofoniczną, zastosowaniu liniowej lub nieliniowej filtracji dolnoprzepustowej czy też na lokalnym wyciszaniu zarówno sygnału jak i zakłócenia. Jednak żadna z tych technik nie pozwalała na skuteczne usunięcie zakłóceń impulsowych bez wprowadzania nowych zniekształceń pogarszających jakość sygnału [25]. Podobny schemat postępowania – detekcja zakłócenia impulsowego, a następnie rekonstrukcja sygnału – wykorzystany został w przypadku wielu technik cyfrowych.

Techniki cyfrowe zyskiwały coraz większą popularność wraz z rozwojem komputerów. Pozwalały na swobodę w przetwarzaniu sygnałów fonicznych oraz dawały znacznie większe możliwości w ich analizie [31]. W literaturze przedmiotu opisano wiele podejść do eliminacji zakłóceń impulsowych, które można podzielić na dwie główne kategorie: metody działające w dziedzinie częstotliwości oraz metody działające w dziedzinie czasu [22], [23].

Do pierwszej kategorii należą algorytmy oparte na adaptacyjnej filtracji Wienera [32], [33], odejmowaniu widmowym [34], [35], [36] oraz bardziej zaawansowane algorytmy CASA wykorzystujące wiedzę o systemie słuchowym człowieka [37], [38], [39], [40] (z ang. *Computational Auditory Scene Analysis*). Efektem działania metod częstotliwościowych jest podniesienie jakości sygnału fonicznego poprzez zwiększenie stosunku sygnału do szumu i/lub poprawienie

rozumiałości. Algorytmy te mogą być stosowane do jednoczesnej filtracji szumu szerokopasmowego oraz eliminacji małych zakłóceń impulsowych. Jednak w przypadku dużych zakłóceń impulsowych algorytmy te radzą sobie już znacznie gorzej, powodując jedynie częściowe stłumienie skutków zakłócenia impulsowego. Dalsza eliminacja „pozostałości” po zakłóceniu impulsowym może okazać się bardzo trudna lub wręcz niemożliwa. W tym przypadku właściwa kolejność etapów rekonstrukcji dźwięku posiada istotne znaczenie dla jakości uzyskanych wyników [25]. W pierwszej kolejności należy usunąć większe zakłócenia impulsowe, a następnie drobne zakłócenia impulsowe i szum szerokopasmowy. Algorytmy te mogą również wprowadzać zniekształcenia i słyszalne artefakty, które są nie do zaakceptowania z punktu widzenia zasad poprawnej rekonstrukcji archiwalnych nagrań fonicznych (instrumentalnych, wokalnych). Niepożądanym efektem dźwiękowym w przetworzonym nagraniu może być szum muzyczny (typowy dla metody odejmowania widm) lub zniekształcenia wynikające ze stłumienia składowych wysokoczęstotliwościowych (typowe dla filtracji Winera).

Do metod eliminacji zakłóceń impulsowych działających w dziedzinie czasu należą:

- klasyczne techniki filtracji (liniowe oraz nieliniowe). Wadą klasycznych metod liniowej filtracji dolnoprzepustowej jest to, że wprowadzają one zmiany w sygnale usuwając składowe wysokoczęstotliwościowe, nawet wtedy gdy zakłócenie impulsowe nie występuje. Lepszym rozwiązaniem może okazać się zastosowanie techniki nieliniowej filtracji takiej jak filtracja medianowa. W podejściu tym próbka centralna w analizowanej sekwencji danych zastępowana jest przez medianę tej sekwencji [41]. Podejście to dobrze sprawdza się w przypadku eliminacji bardzo krótkich zakłóceń impulsowych, takich jak tiki oraz małe kliki, przy jednoczesnym zachowaniu składowych wysokoczęstotliwościowych. Jednak w przypadku dłuższych zakłóceń impulsowych o złożonym przebiegu, filtracja medianowa może okazać się niewystarczająco skuteczna. W pracy [42] opisano podejście oparte na adaptacyjnej filtracji medianowej. W podejściu tym szerokość okna filtru medianowego ustalana była adaptacyjnie, a najlepsze rezultaty uzyskiwano wtedy, gdy zakłócenia impulsowe eliminowane były przy użyciu najmniejszego możliwego okna.
- techniki oparte na autoregresyjnym modelowaniu sygnałów fonicznych, takie jak metody wnioskowania bayesowskiego oparte na modelu sygnału [23], [43], [44], [45], [46], metody oparte na analizie błędów resztowych [47], [48], [49], [50], [51], [52], [53], [54] oraz metody oparte na analizie błędów predykcji [55], [56], [57], [58], [59], [60], [61]. Ważną cechą algorytmów opartych na modelu autoregresyjnym jest ich zdolność do równoczesnego wykrywania zakłóceń impulsowych oraz interpolowania uszkodzonych próbek sygnału na podstawie wiedzy *a priori* o sygnale i charakterze zakłóceń [62], [63]. Metody te działają zadowalająco, gdy charakterystyki sygnału zmieniają się wolno w czasie, często jednak zawodzą w przypadku ich nagłych zmian myląc naturalne dźwięki z zakłóceniami impulsowymi. Kolejnym wyzwaniem dla tych metod jest przetwarzanie sygnałów fonicznych, dla których założenie o ich gaussowskim charakterze nie jest na ogół spełnione.
 - a) **Metody wnioskowania bayesowskiego** cechują się tym, że najbardziej prawdopodobna lokalizacja zakłóceń impulsowych w sygnale ustalana jest na drodze maksymalizacji prawdopodobieństwa *a posteriori*, bez konieczności wybierania progów detekcyjnych. W tych metodach stosuje się iteracyjną procedurę detekcji zakłóceń i rekonstrukcji sygnału przez co metody te są również obliczeniowo bardziej wymagające. W pracy [43] opisano podejście, w którym detekcja zakłóceń impulsowych dokonywana jest dwuetapowo. Podejście to zakłada przetwarzanie blokowe sygnału. Najpierw wstępnie ocenia się położenie zakłóceń impulsowych przy użyciu prostej

techniki opartej na autoregresyjnym modelowaniu sygnału i progowaniu ciągu błędów resztowych otrzymanego w wyniku filtracji odwrotnej. Dopiero na drugim etapie przetwarzania, przy użyciu teorii wnioskowania bayesowskiego, określa się dokładnie położenie zakłóceń wybierając ich najbardziej prawdopodobną lokalizację na podstawie całej dostępnej wiedzy o sygnale. Opisana metoda sprawdza się dobrze w przypadku większych zakłóceń impulsowych, niestety jej skuteczność spada w przypadku małych zakłóceń. W pracy [44] rozwinięto powyższe podejście na przypadek sekwencyjnej analizy sygnału, co pozwoliło na przetwarzania sygnału w czasie rzeczywistym. Z kolei, w pracy [45] opisano podejście, w którym problem łącznej estymacji parametrów modelu i rekonstrukcji sygnału rozwiązany jest przy użyciu metody próbkowania Gibbsa. Proponowany tam schemat detekcji pozwala na wykrywanie małych zakłóceń impulsowych przy jednoczesnej zwiększonej odporności na fałszywe wykrycia. Metoda jest obliczeniowo wymagająca, ale pozwala na uzyskiwanie bardzo dobrych wyników. W pracy [46] zaproponowano podejście, które uwzględnia wiedzę *a priori* o zakłóceniach impulsowych w procesie ich detekcji i eliminacji. Autorzy pracy zaproponowali bayesowską procedurę dopasowywania wzorców opartą na sekwencyjnie dostrajanych modelach: autoregresyjnym modelu czystego sygnału oraz sparametryzowanym modelu zakłócenia impulsowego. Problem łącznej detekcji i estymacji uszkodzonych próbek rozwiązywany jest tu przy użyciu metody próbkowania Gibbsa – za pomocą wariantu algorytmu Metropolisa-Hastingsa szukany jest łączny rozkład *a posteriori* dla czystego sygnału, parametrów modelu autoregresyjnego i parametrów modelu zakłócenia.

- b) **Metody oparte na analizie błędów resztowych** realizowane są w formie przetwarzania blokowego. Zaletą tych metod jest prosty sposób wykrywania obecności większych zakłóceń impulsowych w sygnale. Polega on na progowaniu ciągu błędów resztowych, otrzymanego w wyniku odwrotnej filtracji sygnału przy użyciu współczynników modelu autoregresyjnego. Do estymacji parametrów modelu używane są wszystkie dane z analizowanej ramki sygnału, zarówno te niezakłócone jak i te zakłócone. Taki sposób estymacji parametrów może negatywnie wpływać na zdolności predykcyjne i interpolacyjne otrzymywanego modelu, dlatego zaleca się ponowną estymację parametrów modelu po wykryciu obecności zakłóceń impulsowych. Metody oparte na analizie błędów resztowych często oznaczają jako uszkodzone znacznie dłuższe fragmenty sygnału, ponieważ zakłócenia impulsowe zostają „rozmyte” w wyniku filtracji odwrotnej, co utrudnia wykrycie ich rzeczywistej długości w sygnale. To z kolei prowadzi do rekonstrukcji nadmiernie długich ciągów próbek i może powodować niepożądane efekty dźwiękowe obniżające jakość rekonstruowanego sygnału. W pracy [48] opracowano metodę eliminacji zakłóceń impulsowych z sygnałów mowy opartą na dwuzakresowym modelowaniu sygnału. Stosowany model opisywał zarówno krótkozakresową strukturę korelacyjną sygnału jak i jego strukturę długozakresową. Zdolności predykcyjne i interpolacyjne otrzymywanego modelu silnie zależały od tego, czy zidentyfikowany model był stabilny oraz czy uzyskana ocena okresu podstawowego sygnału używana w części długozakresowej modelu była dokładna. Innym, ciekawym podejściem stosowanym do przetwarzania sygnałów muzycznych, opisanym w pracy [50], było zastosowanie dwóch progów detekcyjnych w odniesieniu do analizowanego ciągu błędów resztowych. Oba progi detekcyjne zależały od stałych współczynników, dobieranych eksperymentalnie oraz od średniego odchylenia standardowego błędów resztowych, szacowanego w sposób odporny na

obecność zakłóceń impulsowych. Pierwszy, wyższy próg detekcyjny służył do wykrycia obecności zakłócenia impulsowego, natomiast drugi, niższy próg pozwalał na ustalenie początku i końca zakłócenia. Metoda ta cechuje się dużą odpornością na fałszywe wykrycia podczas przetwarzania sygnału. W kolejnych pracach [51], [52], [53], [54] progi detekcyjne ustalane były na podstawie statystycznej analizy ciągu błędów resztowych, tak aby w efekcie zmniejszyć odsetek fałszywych wykryć i poprawić zdolności detektora do wykrywania mniejszych zakłóceń impulsowych.

- c) **Metody oparte na analizie błędów predykcji** znane są ze swoich dobrych właściwości śledzących wynikających z tego, że model sygnału aktualizowany jest na bieżąco. Cechuje je duża wrażliwość na nagłe zmiany w sygnale, co pozwala na dokładniejsze wykrywanie zakłóceń impulsowych, nawet tych bardzo małych, ale jeszcze słyszalnych. Metody oparte na analizie błędów predykcji wstrzymują estymację parametrów modelu w momencie wykrycia zakłócenia, a następnie wykorzystują technikę wielokrokowej predykcji sygnału do oceny długości zakłócenia impulsowego. Po oznaczeniu zakłócenia uszkodzony blok próbek jest interpolowany w oparciu o dostępny model sygnału, pochodzący z chwili poprzedzającej moment wykrycia zakłócenia impulsowego. Zaletą tych metod jest to, że do estymacji parametrów modelu wykorzystywane są tylko przeszłe, już przetworzone próbki sygnału. Wpływa to pozytywnie na zdolności predykcyjne i interpolacyjne używanego modelu. W pracy [56] opisano podejście umożliwiające jednoczesną eliminację szumu szerokopasmowego i zakłóceń impulsowych. Pokazano, że zadania równoległej filtracji/śledzenia parametrów modelu autoregresyjnego oraz eliminacji zakłóceń impulsowych mogą być rozwiązane przy użyciu teorii rozszerzonego filtru Kalmana (EKF, z ang. *Extended Kalman Filter*). Następnie, podejście to rozszerzono w pracy [55] na przypadek dwuwymiarowej analizy autoregresyjnej. W tym celu wykorzystano monofoniczne płyty gramofonowe odczytane przy użyciu stereofonicznego przetwornika, tj. przy użyciu odpowiedniej wkładki gramofonowej. Otrzymano w ten sposób w lewym i prawym kanale dwa, nieco różniące się między sobą, sygnały. Oba te sygnały poddano łącznej analizie w celu zwiększenia skuteczności detekcji zakłóceń i interpolacji sygnału. Podejście oparte na analizie błędów predykcji zostało wykorzystane również w przypadku analizy sygnałów mowy – patrz praca [58]. W miejscach występowania silnych pobudzeń krtaniowych blokowano detektor zakłóceń oparty na analizie błędów jednokrokowej predykcji aby uniknąć problemów związanych z fałszywymi wykryciami. Wadą tego podejścia jest jednak to, że zakłócenia pojawiające się w miejscach blokowania detektora nie są wykrywane. Inne rozszerzenie podejścia opartego na EKF, zostało opisane w pracy [59]. Pokazano tam, że łączenie wyników rekonstrukcji uzyskanych przez algorytm EKF działający „w przód” oraz w czasie odwrotnym może przynieść dalszą poprawę uzyskiwanych wyników. Przetwarzanie sygnału fonicznego w czasie odwrotnym daje zazwyczaj lepsze rezultaty w eliminacji zakłóceń impulsowych niż w przypadku analizy prowadzonej wyłącznie „w przód” [23].
- **metody nieliniowe**, takie jak metody wykorzystujące do predykcji i interpolacji sygnału sieci neuronowe [64], [65], [66], mogą mieć lepsze zdolności predykcyjne oraz interpolacyjne od tych obserwowanych dla algorytmów opartych na liniowym modelowaniu sygnału. Zaletą metod nieliniowych jest możliwość odróżnienia zakłóceń impulsowych od gwałtownych zmian w charakterystykach sygnału, pod warunkiem wcześniejszego wytrenowania sieci neuronowej. Przykładowo, gwałtowne zmiany w charakterystykach sygnału mogą być spowodowane udziałem perkusji w nagraniu dźwiękowym. W pracy [65] zaproponowano

procedurę nieliniowego łączenia wyników rekonstrukcji sygnału otrzymanych przez analizę sygnału „w przód” oraz „w tył”. Procedura ta może poprawiać jakość zrekonstruowanego sygnału fonicznego oraz może być stosowana w połączeniu z innymi algorytmami eliminacji zakłóceń impulsowych.

- **inne metody**, których przykładem jest to opisane w pracy [67]. W podejściu tym detekcja zakłóceń impulsowych i rekonstrukcja sygnału odbywa się na drodze przetwarzania sygnałów pochodzących z kilku kopii płyt gramofonowych wytłoczonych przy użyciu tej samej matrycy. Sygnały są najpierw synchronizowane, a następnie różniczkowane, aby uwypuklić obecność zakłóceń impulsowych. W przypadku wykrycia zakłócenia impulsowego w jednym sygnale, do jego rekonstrukcji wykorzystuje się informację pochodzącą z pozostałych, niezakłóconych sygnałów. Zaletą tej metody jest to, że umożliwia zadawalającą rekonstrukcję długich ciągów brakujących (zakłóconych) próbek sygnału, co nie jest możliwe w przypadku użycia tradycyjnych metod rekonstrukcji opartych na modelowaniu.

Efektom rozwoju metod eliminacji zakłóceń impulsowych jest wiele profesjonalnych programów do rekonstrukcji nagrań fonicznych. Do takich programów należą: CEDAR opracowany na Uniwersytecie Cambridge¹ [23], DartPro opracowany w Katedrze Systemów Automatyki Politechniki Gdańskiej dla firmy DARTECH, Inc. [68], internetowy serwis rekonstrukcji nagrań on-line *YouArchive.net*² opracowany w Katedrze Systemów Multimedialnych Politechniki Gdańskiej oraz wiele innych. Skuteczność algorytmów i programów komercyjnych stosowanych do eliminacji zakłóceń impulsowych może być oceniana przy użyciu obiektywnych miar, ale tylko w przypadku gdy znane jest niezakłócone nagranie wzorcowe. W wielu przypadkach takie wzorcowe nagranie nie istnieje, co powoduje, że jakość zrekonstruowanego nagrania może być oceniona jedynie za pomocą testów odsłuchowych. Podczas oceniania jakości nagrania należy zadać sobie następujące pytania [25], [69]:

- czy algorytm wpływa na poprawę subiektywnie odczuwanej jakości sygnału?
- czy zakłócenia impulsowe zostały całkowicie wyeliminowane, czy też może jedynie ich liczba została zredukowana?
- czy pojawiły się nowe, niepożądane efekty dźwiękowe będące wynikiem cyfrowej obróbki sygnału?
- czy osiągnięto akceptowalny kompromis uwzględniający wszystkie przedstawione powyżej kryteria oceny?

Automatyczne metody eliminacji zakłóceń impulsowych powinny być stosowane z rozwagą. Lepiej jest pozostawić nieliczne, drobne zakłócenia impulsowe w przetworzonym nagraniu zachowując jednocześnie jego wyrazistość, niż próbować całkowicie wyeliminować zakłócenia powodując przy tym liczne, niepotrzebne zmiany sygnału obniżające jego jakość.

¹CEDAR Audio Ltd. uważany jest za wiodącego producenta w dziedzinie urządzeń i systemów komputerowych do restauracji dźwięku, a opracowane przez niego algorytmy włączone zostały w postaci „wtyczek” do takich systemów jak: SADiE, Pyramix, Soundscape oraz AMS Neve. Oficjalna strona: www.cedaraudio.com

²Oficjalna strona: www.youarchive.net

2.4 Podsumowanie

W rozdziale tym dokonano przeglądu metod eliminacji zakłóceń impulsowych z sygnałów fonicznych. Podstawowymi ograniczeniami znanych metod są: niestacjonarność analizowanego sygnału oraz jego dwuzakresowa struktura. Ograniczenia te powodują problemy z dokładnym wykrywaniem zakłóceń impulsowych oraz z odróżnianiem naturalnych dźwięków sygnału od zakłóceń impulsowych. Przegląd literatury wskazuje na obiecujące kierunki rozwoju metod opartych na analizie błędów predykcji obejmujące dwukierunkowe przetwarzanie sygnału, dwuzakresowe modelowanie sygnału, wektorowe modelowanie sygnału stereofonicznego oraz wykorzystanie wiedzy *a priori* o zakłóceniach impulsowych do ich dokładniejszego lokalizowania w sygnale. W dalszej części pracy opisane zostaną nowe podejścia do eliminacji zakłóceń impulsowych uwzględniające powyższe kierunki rozwoju dla metody opartej na analizie błędów predykcji.

Rozdział 3

Wybrane metody eliminacji zakłóceń impulsowych

Wprowadzenie

W niniejszym rozdziale przedstawimy szerzej opis metod eliminacji zakłóceń impulsowych wykorzystujących autoregresyjne modelowanie sygnału fonicznego i opartą na tym modelu liniową, adaptacyjną predykcję. W takich podejściach na bieżąco prowadzona jest identyfikacja modelu sygnału, a jej rezultaty używane są do predykcji nowych próbek w oparciu o próbki wcześniejsze. Jeżeli wartość bezwzględna błędu predykcji jest zbyt duża, próbka jest klasyfikowana jako zakłócona i podlega interpolacji.

3.1 Model sygnału fonicznego

Założmy, że spróbkowany sygnał foniczny $y(t)$ ma postać

$$y(t) = s(t) + \delta(t), \quad (3.1)$$

gdzie $t = \dots, -1, 0, 1, \dots$ oznacza unormowany (bezwymiarowy) czas dyskretny, $s(t)$ oznacza niezakłócony (czysty) sygnał foniczny, a $\delta(t)$ jest sekwencją zakłóceń impulsowych.

Założmy, że czysty sygnał foniczny $s(t)$ w warunkach stacjonarności może być opisany za pomocą następującego autoregresyjnego (AR) modelu rzędu r

$$s(t) = \sum_{i=1}^r a_i s(t-i) + n(t), \quad (3.2)$$
$$\text{var}[n(t)] = \rho,$$

gdzie a_1, \dots, a_r są tak zwanymi współczynnikami autoregresji, a $n(t)$ oznacza biały szum tworzący o zerowej wartości oczekiwanej i wariancji ρ . Stosowanie techniki modelowania autoregresyjnego wymaga przyjęcia następującego założenia co do charakteru szumu:

Z1. $\{n(t)\}$ jest ciągiem niezależnych zmiennych losowych o zerowej wartości oczekiwanej i wariancji ρ .



W przypadku braku zakłóceń impulsowych w sygnale $y(t)$, model (3.2) możemy zapisać w następującej skróconej postaci wektorowej

$$y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta} + n(t), \quad (3.3)$$

gdzie $\boldsymbol{\theta} = [a_1, \dots, a_r]^T$ oznacza wektor parametrów, zaś $\boldsymbol{\varphi}(t) = [y(t-1), \dots, y(t-r)]^T$ jest wektorem regresji złożonym z r ostatnich próbek sygnału.

Założymy również, że nie jest dostępny żaden statystyczny model zakłócenia impulsowego (określający częstotliwość występowania, długość bądź kształt zakłócenia). Zamiast tego przyjmujemy następujący model statystyczny

$$\delta(t) \sim \mathcal{N}(0, \gamma(t)), \quad \gamma(t) = \begin{cases} \infty, & \text{jeżeli } d(t) = 1, \\ 0, & \text{jeżeli } d(t) = 0, \end{cases} \quad (3.4)$$

gdzie $d(t)$ oznacza funkcję określającą położenie zakłóceń impulsowych w sygnale $y(t)$

$$d(t) = \begin{cases} 1, & \text{jeżeli } \delta(t) \neq 0, \\ 0, & \text{jeżeli } \delta(t) = 0. \end{cases} \quad (3.5)$$

Zauważmy, że przyjęcie $\gamma(t) = \infty$ oznacza że pomiar $y(t)$ nie zawiera żadnych użytecznych informacji o sygnale $s(t)$. W takiej sytuacji zakłócona próbka sygnału $y(t)$ jest traktowana tak jak próbka brakująca.

Problem usuwania zakłóceń impulsowych może być podzielony na dwa podproblemy:

1. Lokalizacja zakłóceń impulsowych

$$\hat{d}(t) = \begin{cases} 1, & \text{jeżeli próbka jest sklasyfikowana} \\ & \text{jako zakłócona,} \\ 0, & \text{w pozostałych przypadkach.} \end{cases} \quad (3.6)$$

2. Interpolacja próbek sklasyfikowanych jako zakłócone $Y_\delta = \{y(t) : \hat{d}(t) = 1\}$ na podstawie próbek sąsiadujących bezpośrednio z nimi, sklasyfikowanych jako niezakłócone $Y_s = \{y(t) : \hat{d}(t) = 0\}$.

3.2 Rekonstrukcja sygnału

Zdefiniujmy problem rekonstrukcji sygnału jako problem interpolacji n odizolowanych bloków nieodwracalnie uszkodzonych próbek stacjonarnego procesu AR opisanego równaniem (3.2). Załóżmy, że znamy dokładne położenie sekwencji zakłóconych próbek, tj. $\hat{d}(t) \equiv d(t)$. Zatem sygnał $\hat{d}(t)$ może być traktowany jako sekwencja n bloków, oznaczona dalej symbolem $D(i)$:

$$\hat{d}(t) = \begin{cases} 1, & \text{jeżeli } t \in \cup_{i=1}^n D(i), \\ 0, & \text{w pozostałych przypadkach,} \end{cases}$$

gdzie

$$D(i) = [\underline{t}(i), \bar{t}(i)].$$

Wielkości $\underline{t}(i)/\bar{t}(i)$ spełniające warunek

$$\underline{t}(i) \leq \bar{t}(i)$$

oznaczają odpowiednio początek i koniec i -tego bloku. Załóżmy, że sąsiadujące bloki spełniają warunek

$$\underline{t}(i+1) - \bar{t}(i) \geq r+1, \quad (3.7)$$

tj. oddzielone są od siebie przez co najmniej r niezakłóconych próbek sygnału. Jest to minimalna odległość między blokami pozwalająca na zdekomponowanie problemu interpolacji n bloków brakujących próbek na n lokalnych zadań interpolacji.

Uwaga 3.1. Można pokazać, że aby otrzymać na podstawie modelu AR (3.2) optymalną, w sensie minimum błędu średniokwadratowego, interpolację odizolowanego bloku danych, wystarczy wziąć pod uwagę r próbek sygnału bezpośrednio poprzedzających ten blok i r próbek sygnału bezpośrednio występujących po nim [62], [70]. Próbki sygnału oddalone od tego bloku o $r+1$, $r+2$, itd., nie wpływają na otrzymaną interpolację. Dlatego każdy taki segment może być rozpatrywany niezależnie od pozostałych.

Biorąc pod uwagę ten fakt, rozważmy lokalny problem interpolacji bloku m nieodwracalnie uszkodzonych próbek w oparciu o model AR (3.2). Załóżmy, że rozważany blok zaczyna się w chwili $t_1 = t_0 + 1$, zaś kończy w chwili $t_2 = t_0 + m$. Niech $\mathcal{T}_0 = [t_1, t_2]$. Sytuację tę możemy opisać w następujący sposób

$$\begin{aligned} \hat{d}(t_1) &= \dots = \hat{d}(t_2) = 1, \\ \hat{d}(t_1 - r) &= \dots = \hat{d}(t_1 - 1) = 0, \\ \hat{d}(t_2 + 1) &= \dots = \hat{d}(t_2 + r) = 0. \end{aligned} \quad (3.8)$$

Interpolacja brakującej sekwencji próbek może zostać otrzymana za pomocą interpolatora w wersji „wsadowej”, wymagającej odwracania macierzy, lub w wersji rekurencyjnej, nie wymagającej jej odwracania. Obie interpolacje są identyczne.

3.2.1 Interpolator w wersji wymagającej odwracania macierzy

Optymalna, w sensie średniokwadratowym, estymata brakującego bloku danych $\{s(t), t \in \mathcal{T}_0\}$ ma postać [70]

$$\begin{aligned} \{\hat{s}(t_1), \dots, \hat{s}(t_2)\} &= \arg \min_{s(t_1), \dots, s(t_2)} \sum_{t=t_1}^{t_2+r} \left[s(t) - \sum_{i=1}^r a_i s(t-i) \right]^2 = \\ &= \arg \min_{\boldsymbol{\psi}_m} \boldsymbol{\psi}^T \mathbf{B} \mathbf{B}^T \boldsymbol{\psi}, \end{aligned} \quad (3.9)$$

gdzie $\boldsymbol{\psi} = [s(t_1 - r), \dots, s(t_2 + r)]^T$ oznacza wektor wszystkich próbek branych pod uwagę w (3.9), $\boldsymbol{\psi}_m = [s(t_1), \dots, s(t_2)]^T$ oznacza wektor brakujących próbek, zaś \mathbf{B} oznacza macierz o wymiarach $(r+m) \times (2r+m)$ zbudowaną ze współczynników autoregresji

$$\mathbf{B} = \begin{bmatrix} a_r & a_{r-1} & \dots & -1 & 0 & 0 & \dots & 0 \\ 0 & a_r & \dots & a_1 & -1 & 0 & \dots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & \dots & 0 & a_r & \dots & a_1 & -1 \end{bmatrix}.$$

Niechaj $\mathcal{K} = \{1, \dots, r, r + m + 1, \dots, 2r + m\}$ oznacza zbiór indeksów charakteryzujących położenie $2r$ znanych próbek w analizowanym fragmencie sygnału fonicznego $\{s(t_1 - r), \dots, s(t_2 + r)\}$ o długości $2r + m$. Podobnie, niech $\mathcal{U} = \{r + 1, \dots, r + m\}$ oznacza zbiór indeksów charakteryzujących położenie m nieznanymi próbek. Symbolem $\mathbf{B}_m = \mathbf{B}_{|\mathcal{K}>}$ będziemy oznaczali macierz o wymiarach $(r + m) \times m$ otrzymaną poprzez usunięcie z macierzy \mathbf{B} kolumn o indeksach ze zbioru \mathcal{K} . Podobnie, symbolem $\mathbf{B}_o = \mathbf{B}_{|\mathcal{U}>}$ będziemy oznaczali macierz o wymiarach $(r + m) \times 2r$ otrzymaną z \mathbf{B} po usunięciu kolumn o indeksach ze zbioru \mathcal{U} .

Zgodnie z [71], optymalna estymata (3.9), która może być interpretowana jako rzut prostopadły wektora nieznanymi próbek $\boldsymbol{\psi}_m$ na wektor znanych próbek $\boldsymbol{\psi}_o$

$$\boldsymbol{\psi}_o = [s(t_1 - r), \dots, s(t_1 - 1), s(t_2 + 1), \dots, s(t_2 + r)]^T,$$

jest dana następującym równaniem

$$\hat{\boldsymbol{\psi}}_m = - \left(\mathbf{B}_m^T \mathbf{B}_m \right)^{-1} \mathbf{B}_m^T \mathbf{B}_o \boldsymbol{\psi}_o, \quad (3.10)$$

gdzie $\hat{\boldsymbol{\psi}}_m = [\hat{s}(t_1), \dots, \hat{s}(t_2)]^T$. Macierz $\mathbf{B}_m^T \mathbf{B}_m$ jest nieosobliwa, dodatnio określona oraz pasmowa [71]

$$\begin{aligned} \det[\mathbf{B}_m^T \mathbf{B}_m] &\neq 0, \\ \mathbf{B}_m^T \mathbf{B}_m &> 0, \end{aligned} \quad (3.11)$$

co pozwala na jej odwrócenie przy użyciu algorytmu opartego na dekompozycji Choleskiego.

3.2.2 Interpolator w wersji rekurencyjnej

Podejście do rekonstrukcji sygnału wykorzystujące algorytm filtracji Kalmana [70] będzie dla nas punktem początkowym dla wyprowadzenia bardziej realistycznego algorytmu łączącego w sobie procedurę detekcji zakłóconych próbek sygnału z procedurą interpolacji sygnału.

W celu zaprojektowania algorytmu filtracji Kalmana wprowadzimy opis w przestrzeni stanów równoważny opisowi typu wejście-wyjście (3.1)-(3.2). Niechaj $q = 2r + m$ oznacza rząd modelu w przestrzeni stanów. Zdefiniujemy wektor stanu $\mathbf{x}_q(t) = [s(t), \dots, s(t - q + 1)]^T$ o wymiarach $q \times 1$ złożony z q ostatnich próbek sygnału. Symbolem $\mathbf{0}_q$ będziemy oznaczać wektor zerowy o wymiarach $q \times 1$, zaś symbolami \mathbf{O}_q i \mathbf{I}_q – odpowiednio, macierz zerową oraz macierz jednostkową o wymiarach $q \times q$.

W celu rekonstrukcji wydzielonego bloku próbek za pomocą filtru Kalmana, niezbędne jest użycie nadokreślonego (o rzędzie $q > r$) modelu w przestrzeni stanów (zwróćmy uwagę, że do opisu zależności (3.1)-(3.2) wystarczające jest użycie $q = r$). Model taki możemy zapisać w następującej postaci

$$\begin{aligned} \mathbf{x}_q(t + 1) &= \mathbf{A}_q \mathbf{x}_q(t) + \mathbf{c}_q n(t + 1), \\ y(t) &= \mathbf{c}_q^T \mathbf{x}_q(t) + \delta(t), \end{aligned} \quad (3.12)$$

gdzie

$$\mathbf{A}_q = \begin{bmatrix} a_1 & a_2 & \dots & a_r & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & & 0 & 0 & \dots & 0 & 0 \\ \vdots & & & & & \ddots & & \vdots \\ 0 & 0 & & 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

oznacza macierz przejścia o wymiarach $q \times q$, a $\mathbf{c}_q = [1, 0, \dots, 0]^T$ oznacza wektor wyjścia o wymiarach $q \times 1$.

Stosowanie algorytmu filtracji Kalmana wymaga przyjęcia następujących założeń co do charakteru szumu:

Z2. $\{n(t)\}$ jest ciągiem niezależnych zmiennych gaussowskich o zerowej wartości oczekiwanej i znanej wariancji ρ

$$n(t) \sim \mathcal{N}(0, \rho).$$

Z3. Ciąg zakłóceń $\delta(t_0 + 1), \dots, \delta(t_0 + m)$ może być potraktowany jako sekwencja wzajemnie nieskorelowanych zmiennych gaussowskich, niezależnych od $\{n(t)\}$, o znanej wariancji

$$\gamma(t) = \text{var}[\delta(t)], \quad t_0 + 1 \leq t \leq t_0 + m.$$

Na podstawie (3.12) oraz dostępnej wiedzy *a priori* o sygnale, rekurencyjny algorytm filtracji/predykcji Kalmana możemy zapisać w następującej postaci

$$\begin{aligned} \hat{\mathbf{x}}_q(t|t-1) &= \mathbf{A}_q \hat{\mathbf{x}}_q(t-1|t-1), \\ \mathbf{P}_q(t|t-1) &= \mathbf{A}_q \mathbf{P}_q(t-1|t-1) \mathbf{A}_q^T + \rho \mathbf{c}_q \mathbf{c}_q^T, \\ e(t) &= y(t) - \mathbf{c}_q^T \hat{\mathbf{x}}_q(t|t-1), \\ g(t) &= \mathbf{c}_q^T \mathbf{P}_q(t|t-1) \mathbf{c}_q + \gamma(t), \\ \mathbf{l}_q(t) &= \mathbf{P}_q(t|t-1) \mathbf{c}_q / g(t), \\ \hat{\mathbf{x}}_q(t|t) &= \hat{\mathbf{x}}_q(t|t-1) + \mathbf{l}_q(t) e(t), \\ \mathbf{P}_q(t|t) &= \mathbf{P}_q(t|t-1) - g(t) \mathbf{l}_q(t) \mathbf{l}_q^T(t). \end{aligned} \quad (3.13)$$

Dopiero w chwili $t_0 + 1$, kiedy $\delta(t_0 + 1) \neq 0$ i przy założeniu (Z3), algorytm rozpoczyna rekonstrukcję zakłóconych próbek z warunkami początkowymi

$$\begin{aligned} \hat{\mathbf{x}}_q(t_0|t_0) &= [y(t_0), \dots, y(t_0 - q + 1)]^T, \\ \mathbf{P}_q(t_0|t_0) &= \mathbf{O}_q. \end{aligned} \quad (3.14)$$

Zauważmy, że gdy $\gamma(t) \rightarrow \infty$, $t \in [t_0 + 1, t_0 + m]$, próbka sygnału $y(t)$ traktowana jest jako zakłócenie o nieskończonej wariancji i w następstwie tego, informacja zawarta w tym pomiarze jest ignorowana. Próbka ta będzie podlegała interpolacji na podstawie sąsiadujących próbek sygnału, sklasyfikowanych jako wolne od zakłóceń. Natomiast gdy $\gamma(t) = 0$, $t \in [t_0 + m + 1, t_0 + m + r]$, próbka sygnału $y(t)$ traktowana jest jako wolna od zakłóceń i jest akceptowana. Na podstawie tych rozważań możemy podać równania aktualizacji pomiaru w przypadku $\hat{d}(t) = 1$ dla $t \in [t_0 + 1, t_0 + m]$

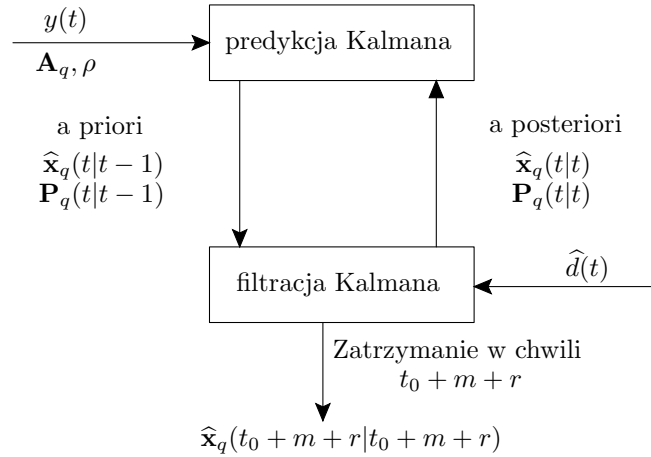
$$\begin{aligned} \hat{\mathbf{x}}_q(t|t) &= \hat{\mathbf{x}}_q(t|t-1), \\ \mathbf{P}_q(t|t) &= \mathbf{P}_q(t|t-1) \end{aligned} \quad (3.15)$$

Inicjalizacja:

$$\hat{\mathbf{x}}_q(t_0|t_0) = [y(t_0), \dots, y(t_0 - q + 1)]^T$$

$$\mathbf{P}_q(t_0|t_0) = \mathbf{O}_q$$

$$t \in [t_0 + 1, t_0 + m + r]$$



Rysunek 3.1: Schemat ideowy procedury interpolacji realizowanej za pomocą algorytmu filtracji/predykcji Kalmana. Oznaczenia: $t_0 + 1$ oznacza chwilę rozpoczęcia pracy algorytmu rekonstrukcji, m oznacza długość rekonstruowanego bloku danych, r oznacza rząd modelu autoregresyjnego, $q = 2r + m$ oznacza rząd modelu w przestrzeni stanu, \mathbf{O}_q oznacza macierz zerową o wymiarach $q \times q$, \mathbf{A}_q oznacza macierz przejścia o wymiarach $q \times q$, ρ oznacza wariancję białego szumu tworzącego $n(t)$, $\hat{\mathbf{x}}_q(t|t-1)/\hat{\mathbf{x}}_q(t|t)$ oznacza ocenę wektora stanu a priori/a posteriori, $\mathbf{P}_q(t|t-1)/\mathbf{P}_q(t|t)$ oznacza macierz kowariancji a priori/a posteriori, a $\hat{d}(t)$ jest funkcją lokalizacji zakłócenia wskazującą, które pomiary są wykorzystywane/ignorowane na etapie filtracji Kalmana. Algorytm interpolacji zostaje zatrzymany w chwili $t_0 + m + r$, kiedy r kolejnych pomiarów $y(t)$ z rzędu zostanie wykorzystanych na etapie filtracji Kalmana.

oraz w przypadku $\hat{d}(t) = 0$ dla $t \in [t_0 + m + 1, t_0 + m + r]$

$$\begin{aligned} \hat{\mathbf{x}}_q(t|t) &= \hat{\mathbf{x}}_q(t|t-1) + \mathbf{l}_q(t)e(t), \\ \mathbf{P}_q(t|t) &= \mathbf{P}_q(t|t-1) - g(t)\mathbf{l}_q(t)\mathbf{l}_q^T(t). \end{aligned} \quad (3.16)$$

Algorytm kończy swoją pracę w chwili $t_0 + m + r$ po wykorzystaniu r niezakłóconych próbek sygnału występujących bezpośrednio po bloku uszkodzonych próbek. W chwili zatrzymania algorytmu ocena filtracyjna wektora stanu ma postać

$$\begin{aligned} \hat{\mathbf{x}}_q(t_0 + m + r | t_0 + m + r) &= [y(t_0 + m + r), \dots, y(t_0 + m + 1), \\ &\quad \hat{s}(t_0 + m), \dots, \hat{s}(t_0 + 1), y(t_0), \dots, y(t_0 + 1 - r)]^T, \end{aligned} \quad (3.17)$$

gdzie sekwencja $\{\hat{s}(t_0 + 1), \dots, \hat{s}(t_0 + m)\}$ stanowi blok interpolowanych próbek. Schemat ideowy procedury interpolacji realizowanej za pomocą algorytmu filtracji/predykcji Kalmana został przedstawiony na rysunku 3.1.

Zwróćmy uwagę na to, że w rozważanym przypadku otrzymane interpolacje nie zależą od pomiarów zebranych w chwilach $t_0 + m + r + 1$, $t_0 + m + r + 2$, itd., zatem nie ma powodu, aby kontynuować proces filtracji Kalmana po osiągnięciu chwili $t_0 + m + r$ – patrz uwaga 3.1.

3.2.3 Algorytm filtru Kalmana o zmiennym rzędzie

Wykorzystując wiedzę o specjalnej strukturze macierzy/wektorów \mathbf{A}_q , \mathbf{c}_q i $\mathbf{P}_q(t_0|t_0)$ można pokazać, że rząd filtru Kalmana może być stopniowo zwiększany bez żadnego wpływu na wyniki estymacji. Stopniowe zwiększanie rzędu rozpoczyna się od $q = r + 1$ w chwili $t_0 + 1$ i kończy się na $q = 2r + m$ w chwili $t_0 + m + r$. Filtr Kalmana o zmiennym rzędzie oferuje znaczącą oszczędności obliczeniowe w porównaniu z jego wersją o stałym rzędzie ($q = q_{\max} = 2r + m_{\max}$).

Niech $\boldsymbol{\theta}_r = [a_1, \dots, a_r]^T$, natomiast przez $\boldsymbol{\theta}_q = [\boldsymbol{\theta}_r^T, \mathbf{0}_{q-r}^T]^T$, $q > r$, będziemy oznaczali powiększony wektor parametrów. Ponadto, niech $q(t) = r + t - t_0$ oznacza zmienny w czasie rząd modelu. Algorytm filtracji Kalmana o zmiennym rzędzie rozpoczyna pracę w chwili $t = t_0 + 1$ z warunkami początkowymi

$$\begin{aligned}\hat{\mathbf{x}}_r(t_0|t_0) &= [y(t_0), \dots, y(t_0 - r + 1)]^T, \\ \mathbf{P}_r(t_0|t_0) &= \mathbf{O}_r.\end{aligned}\quad (3.18)$$

W pierwszym kroku, po aktualizacji czasu ($t \geq t_0 + 1$), otrzymujemy wielkości o powiększonym rzędzie: ocenę wektora stanu *a priori* $\hat{\mathbf{x}}_{q(t)}(t|t-1)$ oraz macierz kowariancji *a priori* $\mathbf{P}_{q(t)}(t|t-1)$:

$$\begin{aligned}\hat{y}(t|t-1) &= \boldsymbol{\theta}_{q(t)-1}^T \hat{\mathbf{x}}_{q(t)-1}(t-1|t-1), \\ e(t) &= y(t) - \hat{y}(t|t-1), \\ \hat{\mathbf{x}}_{q(t)}(t|t-1) &= \begin{bmatrix} \hat{y}(t|t-1) \\ \hat{\mathbf{x}}_{q(t)-1}(t-1|t-1) \end{bmatrix}, \\ \mathbf{h}_{q(t)-1}(t) &= \mathbf{P}_{q(t)-1}(t-1|t-1)\boldsymbol{\theta}_{q(t)-1}, \\ \rho_e(t) &= \boldsymbol{\theta}_{q(t)-1}^T \mathbf{h}_{q(t)-1}(t) + \rho, \\ \mathbf{P}_{q(t)}(t|t-1) &= \begin{bmatrix} \rho_e(t) & \mathbf{h}_{q(t)-1}^T(t) \\ \mathbf{h}_{q(t)-1}(t) & \mathbf{P}_{q(t)-1}(t-1|t-1) \end{bmatrix}.\end{aligned}\quad (3.19)$$

Zwróćmy uwagę na specjalną budowę macierzy kowariancji *a priori* o powiększonym rzędzie, która powstaje w wyniku „obsunięcia” macierzy kowariancji *a posteriori* $\mathbf{P}_{q(t)-1}(t-1|t-1)$ oraz dodania pierwszej kolumny i pierwszego wiersza [lewy górny element tej macierzy odpowiada wariancji błędu predykcji $\rho_e(t)$].

W kolejnym kroku w przypadku ignorowania pomiaru, $\hat{d}(t) = 1$, $t \in [t_0 + 1, t_0 + m]$, aktualizacja polega na przepisaniu informacji *a priori*

$$\begin{aligned}\hat{\mathbf{x}}_{q(t)}(t|t) &= \hat{\mathbf{x}}_{q(t)}(t|t-1), \\ \mathbf{P}_{q(t)}(t|t) &= \mathbf{P}_{q(t)}(t|t-1).\end{aligned}\quad (3.20)$$

Natomiast w przypadku akceptacji pomiaru, $\hat{d}(t) = 0$, $t \in [t_0 + m + 1, t_0 + m + r]$, aktualizacja przebiega następująco

$$\begin{aligned}\mathbf{l}_{q(t)}(t) &= \frac{1}{\rho_e(t)} \mathbf{P}_{q(t)}^{(1)}(t|t-1), \\ \hat{\mathbf{x}}_{q(t)}(t|t) &= \hat{\mathbf{x}}_{q(t)}(t|t-1) + \mathbf{l}_{q(t)}(t)e(t), \\ \mathbf{P}_{q(t)}(t|t) &= \mathbf{P}_{q(t)}(t|t-1) - \rho_e(t)\mathbf{l}_{q(t)}(t)\mathbf{l}_{q(t)}^T(t),\end{aligned}\quad (3.21)$$

gdzie $\mathbf{P}_{q(t)}^{(1)}(t|t-1)$ oznacza pierwszą kolumnę macierzy kowariancji *a priori*.

Algorytm przerywa pracę w chwili $t = t_0 + m + r$ po zaakceptowaniu r kolejnych pomiarów. W chwili zatrzymania algorytmu ocena filtracyjna wektora stanu ma postać

$$\hat{\mathbf{x}}_{q(t)}(t|t) = [y(t), \dots, y(t_0 + m + 1), \hat{s}(t_0 + m), \dots, \hat{s}(t_0 + 1), y(t_0), \dots, y(t_0 + 1 - r)]^T \quad (3.22)$$

i jest identyczna z (3.17). Adaptacyjna wersja algorytmu, przedstawionego powyżej, może zostać otrzymana przez zastąpienie wielkości $\boldsymbol{\theta}_r$ i ρ ich oszacowaniami.

Zabezpieczenia numeryczne

Gdy $\hat{d}(t) = 0$, zastosowanie reguły (3.21) sprawia, że filtr Kalmana stawiany jest w trudnej sytuacji numerycznej. Krok aktualizacji pomiaru powinien ustawić górny element wektora $\mathbf{I}_{q(t)}(t)$ na wartość równą jeden. To z kolei skutkuje ustawieniem pierwszego elementu wektora $\hat{\mathbf{x}}_{q(t)}(t|t)$ na $y(t)$ i wyzerowaniem lewej górnej wartości w macierzy $\mathbf{P}_{q(t)}(t|t)$. Ponieważ „teoretyczna” macierz kowariancji *a posteriori* (tj. macierz obliczana z nieskończoną precyzją) jest w tym przypadku osobliwa, jej wersja „numeryczna” – z powodu ograniczonej dokładności obliczeniowej – może w łatwy sposób stracić własność nieujemnej określoności, powodując w następstwie błędne zachowanie filtru Kalmana. Problem złego uwarunkowania wskazany powyżej może zostać rozwiązany przy użyciu filtracji pierwiastkowej [72], ale biorąc pod uwagę specyficzne przeznaczenie algorytmu, prostszym rozwiązaniem wydaje się być zastosowanie techniki zaokrąglania poprawiającej odporność numeryczną algorytmu – po każdym cyklu obliczeń (prowadzonych w MATLABie) wyniki zaokrąglane zostają do 12-tego miejsca po przecinku. Bezpośrednią konsekwencją zaokrąglania jest to, że wszystkie elementy o wartościach równych „prawie zero” oraz „prawie jeden” w obliczanych macierzach/wektorach zastąpione zostają odpowiednio zerami i jedynkami. Pozwala to uniknąć problemów numerycznych zachowując strukturę algorytmu filtracji Kalmana o zmiennym rzędzie.

3.3 Identyfikacja sygnału fonicznego

W niniejszym podrozdziale omówimy „tradycyjne” metody identyfikacji modelu AR, które mogą być stosowane w adaptacyjnych algorytmach eliminacji zakłóceń impulsowych.

Do tej pory, w celu uproszczenia opisu procedury interpolacji, zakładaliśmy, że parametry modelu AR są stałe i znane. Aby otrzymać adaptacyjną wersję procedury, należy zastąpić wielkości $\boldsymbol{\theta}$ i ρ ich oszacowaniami $\hat{\boldsymbol{\theta}}(t)$ i $\hat{\rho}(t)$, otrzymanymi w wyniku bieżącej identyfikacji modelu sygnału. Identyfikacja może być prowadzona przez algorytmy estymacji o skończonej pamięci: takie jak dobrze znane ważone estymatory najmniejszych kwadratów (z ang. *Weighted Least Squares*, WLS), estymatory gradientowe takie jak algorytm najmniejszych średnich kwadratów (z ang. *Least Mean Squares*, LMS) oraz jego modyfikacje, tj. unormowany algorytm najmniejszych średnich kwadratów (z ang. *Normalized Least Mean Squares*, NLMS) [72], [73].

Aby uwzględnić niestacjonarność sygnału fonicznego, stosować będziemy następujący odpowiednik (3.3), zakładający zmienność parametrów modelu w czasie

$$\begin{aligned} y(t) &= \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}(t) + n(t), \\ \text{var}[n(t)] &= \rho(t), \end{aligned} \quad (3.23)$$

gdzie $\boldsymbol{\theta}(t) = [a_1(t), \dots, a_r(t)]^T$ oznacza wektor wolno zmieniających się w czasie współczynników autoregresji, zaś $\rho(t)$ jest wariancją białego szumu tworzącego, podlegającą również wolnym zmianom w czasie. W warunkach niestacjonarności sygnału rząd autoregresji może być dobrany jednorazowo na stałe lub aktualizowany adaptacyjnie na podstawie kryterium informacyjnego Akaikiego [74].

3.3.1 Ważone estymatory najmniejszych kwadratów

W tym punkcie omówione zostaną ważne estymatory najmniejszych kwadratów w wersji z oknem prostokątnym (z ang. *Sliding Window Least Squares*, SWLS) oraz z oknem wykładniczym (z ang. *Exponentially Weighted Least Squares*, EWLS).

Estymator WLS minimalizuje ważoną sumę kwadratów błędów modelowania

$$\hat{\boldsymbol{\theta}}(t) = \arg \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}), \quad (3.24)$$

gdzie $J(\boldsymbol{\theta})$ oznacza funkcję strat

$$J(\boldsymbol{\theta}) = \sum_{i=0}^{t-1} w(i) [y(t-i) - \boldsymbol{\varphi}^T(t-i)\boldsymbol{\theta}]^2, \quad (3.25)$$

w której $\{w(i)\}$ oznacza ciąg ważący o nieujemnych współczynnikach

$$\begin{aligned} w(0) = 1 \geq w(1) \geq \dots \geq 0, \\ \sum_{i=0}^{\infty} w(i) < \infty. \end{aligned} \quad (3.26)$$

Minimalizacja odbywa się na drodze przyrównania do zera gradientu funkcji strat $J(\boldsymbol{\theta})$

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(t)} &= 2 \left[\sum_{i=0}^{t-1} w(i) \boldsymbol{\varphi}(t-i) \boldsymbol{\varphi}^T(t-i) \right] \hat{\boldsymbol{\theta}}(t) - 2 \left[\sum_{i=0}^{t-1} w(i) y(t-i) \boldsymbol{\varphi}(t-i) \right], \\ \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(t)} &= \mathbf{0}_r. \end{aligned} \quad (3.27)$$

Z zależności (3.27) otrzymujemy ogólną postać ważonego estymatora najmniejszych kwadratów

$$\hat{\boldsymbol{\theta}}(t) = \left[\sum_{i=0}^{t-1} w(i) \boldsymbol{\varphi}(t-i) \boldsymbol{\varphi}^T(t-i) \right]^{-1} \left[\sum_{i=0}^{t-1} w(i) y(t-i) \boldsymbol{\varphi}(t-i) \right] = \mathbf{R}^{-1}(t) \mathbf{p}(t) \quad (3.28)$$

przy założeniu, że macierz regresji jest odwracalna

$$\det[\mathbf{R}(t)] = \det \left[\sum_{i=0}^{t-1} w(i) \boldsymbol{\varphi}(t-i) \boldsymbol{\varphi}^T(t-i) \right] \neq 0, \quad (3.29)$$

co jest warunkiem identyfikowalności modelu AR.

Pamięć estymatora WLS

Pamięć estymatora WLS jest istotnym parametrem wpływającym na jego zdolności śledzące. Estymatory o krótkiej pamięci są „szybkie” (małe systematyczne błędy estymacji w trakcie śledzenia), ale „niedokładne” (duża wariancja śledzenia), podczas gdy estymatory o długiej pamięci

są „wolne”, ale „dokładne”. Najlepsze wyniki otrzymujemy, gdy pamięć estymacji algorytmu śledzącego jest „dopasowana” do stopnia niestacjonarności identyfikowanego procesu [72]. Pamięć estymatora można regulować poprzez zabieg zapominania lub ważenia odległych danych identyfikowanego procesu. W rezultacie estymator jest nieczuły (lub mniej czuły) na dane zebrane w odległej przeszłości. W praktyce najczęściej stosowanymi ciągami ważącymi (tzw. oknami) są: okno prostokątne

$$w(i) = \begin{cases} 1, & \text{jeżeli } 0 \leq i < N, \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad (3.30)$$

oraz okno wykładnicze

$$w(i) = \lambda^i, \quad 0 < \lambda < 1. \quad (3.31)$$

W pierwszym przypadku tylko N ostatnich danych (z jednakowymi wagami równymi 1) jest branych pod uwagę przy estymacji parametrów. Podejście to jest często nazywane podejściem z przesuającym się oknem danych.

W drugim przypadku wagi maleją wykładniczo, stopniowo zmniejszając wpływ starszych danych na najświeższe oceny parametrów. Podejście to nazywane jest często zapominaniem wykładniczym. Stała λ oznacza tak zwaną stałą zapominania.

Estymator SWLS

W przypadku zastosowania w zależności (3.28) okna prostokątnego (3.30) otrzymujemy estymator SWLS postaci

$$\hat{\boldsymbol{\theta}}(t, N) = \left[\sum_{i=0}^{N-1} \boldsymbol{\varphi}(t-i) \boldsymbol{\varphi}^T(t-i) \right]^{-1} \left[\sum_{i=0}^{N-1} y(t-i) \boldsymbol{\varphi}(t-i) \right] = \mathbf{R}^{-1}(t, N) \mathbf{p}(t, N). \quad (3.32)$$

Rekurencyjna procedura dla tego estymatora ma postać algorytmu realizowanego w dwóch krokach. W pierwszym kroku przedział analizy jest poszerzany o jedną próbkę w celu uwzględnienia nowej obserwacji $y(t)$

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t, N+1) &= \mathbf{R}^{-1}(t, N+1) \mathbf{p}(t, N+1), \\ \mathbf{R}(t, N+1) &= \mathbf{R}(t-1, N) + \boldsymbol{\varphi}(t) \boldsymbol{\varphi}^T(t), \\ \mathbf{p}(t, N+1) &= \mathbf{p}(t-1, N) + \boldsymbol{\varphi}(t) y(t). \end{aligned} \quad (3.33)$$

W drugim kroku przedział analizy jest zmniejszony o jedną próbkę w celu usunięcia zależności oceny parametrów $\hat{\boldsymbol{\theta}}(t, N+1)$ od obserwacji $y(t-N)$ zebranej $N+1$ chwil wcześniej

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t, N) &= \mathbf{R}^{-1}(t, N) \mathbf{p}(t, N), \\ \mathbf{R}(t, N) &= \mathbf{R}(t, N+1) - \boldsymbol{\varphi}(t-N) \boldsymbol{\varphi}^T(t-N), \\ \mathbf{p}(t, N) &= \mathbf{p}(t, N+1) - \boldsymbol{\varphi}(t-N) y(t-N). \end{aligned} \quad (3.34)$$

Aby otrzymać rekurencyjną wersję tego algorytmu bez konieczności odwracania macierzy regresji, należy zastosować lemat 1 o odwracaniu macierzy, a dokładniej jego specjalny przypadek zwany zależnością Shermana-Morrisona (A.2) – rekurencyjny algorytm SWLS został opisany między innymi w pracy [73].

Estymator EWLS

W przypadku zastosowania okna wykładniczego (3.31) otrzymujemy estymator EWLS postaci

$$\hat{\boldsymbol{\theta}}(t) = \left[\sum_{i=0}^{t-1} \lambda^i \boldsymbol{\varphi}(t-i) \boldsymbol{\varphi}^T(t-i) \right]^{-1} \left[\sum_{i=0}^{t-1} \lambda^i y(t-i) \boldsymbol{\varphi}(t-i) \right] = \mathbf{R}^{-1}(t) \mathbf{p}(t). \quad (3.35)$$

Jest to estymator „wsadowy”, ponieważ bieżąca ocena wektora parametrów $\hat{\boldsymbol{\theta}}(t)$ jest wyznaczana niezależnie od poprzedniej oceny wektora $\hat{\boldsymbol{\theta}}(t-1)$. Rozwiązanie to jest mało efektywne i wymaga odwracania macierzy regresji w każdym kroku.

Rekurencyjna wersja tego estymatora może być wyprowadzona w następujących krokach. Zauważmy, że zachodzą rekursje

$$\begin{aligned} \mathbf{R}(t) &= \lambda \mathbf{R}(t-1) + \boldsymbol{\varphi}(t) \boldsymbol{\varphi}^T(t), \\ \mathbf{p}(t) &= \lambda \mathbf{p}(t-1) + \boldsymbol{\varphi}(t) y(t), \end{aligned} \quad (3.36)$$

co pozwala na wyrażenie zależności (3.35) w postaci rekurencyjnej

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \mathbf{R}^{-1}(t) \mathbf{p}(t) = \mathbf{R}^{-1} \left[\lambda \mathbf{R}(t-1) \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{\varphi}(t) y(t) \right] = \\ &= \mathbf{R}^{-1}(t) \left[\mathbf{R}(t) \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{\varphi}(t) (y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t-1)) \right] = \\ &= \hat{\boldsymbol{\theta}}(t-1) + \mathbf{Q}(t) \boldsymbol{\varphi}(t) \varepsilon(t|t-1), \end{aligned} \quad (3.37)$$

gdzie

$$\mathbf{Q}(t) = \mathbf{R}^{-1}(t) = [\lambda \mathbf{Q}^{-1}(t-1) + \boldsymbol{\varphi}(t) \boldsymbol{\varphi}^T(t)]^{-1}. \quad (3.38)$$

Po zastosowaniu lematu 1 o odwracaniu macierzy [podstawiając w zależności (A.2) $\mathbf{A} = \lambda \mathbf{Q}^{-1}(t)$, $\mathbf{x} = \boldsymbol{\varphi}(t)$] otrzymujemy dobrze znaną postać algorytmu EWLS

$$\begin{aligned} \varepsilon(t|t-1) &= y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t-1), \\ \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \mathbf{k}(t) \varepsilon(t|t-1), \\ \mathbf{k}(t) &= \frac{\mathbf{Q}(t-1) \boldsymbol{\varphi}(t)}{\lambda + \boldsymbol{\varphi}^T(t) \mathbf{Q}(t-1) \boldsymbol{\varphi}(t)}, \\ \mathbf{Q}(t) &= \frac{1}{\lambda} [\mathbf{I}_r - \mathbf{k}(t) \boldsymbol{\varphi}^T(t)] \mathbf{Q}(t-1), \end{aligned} \quad (3.39)$$

gdzie $\mathbf{k}(t)$ oznacza wektor wzmocnień, a $\mathbf{Q}(t)$ jest macierzą kowariancji błędów estymacji. Porównując równania (3.37) i (3.39) możemy zauważyć, że

$$\mathbf{k}(t) = \mathbf{Q}(t) \boldsymbol{\varphi}(t). \quad (3.40)$$

Aby algorytm rozpoczął poprawnie pracę od chwili $t = 1$, należy wybrać odpowiednie wartości początkowe $\hat{\boldsymbol{\theta}}(0)$ i $\mathbf{Q}(0)$ oraz przyjąć zerowe wartości $y(t)$ odpowiadające chwilom czasu $t < 1$. Gdy nie dysponujemy żadną wiedzą wstępną, przyjmujemy

$$\hat{\boldsymbol{\theta}}(0) = \mathbf{0}_r, \quad \mathbf{Q}(0) = \rho_0 \mathbf{I}_r, \quad (3.41)$$

gdzie $\rho_0 > 0$ jest odpowiednio „dużą” liczbą. Jeśli macierz $\mathbf{Q}(0)$ jest „duża” [co oznacza małe zaufanie do oceny początkowej $\hat{\boldsymbol{\theta}}(0)$], to kolejne oceny wektora parametrów $\hat{\boldsymbol{\theta}}(t)$ stają się bliskie ocenie wsadowej (3.35) już w chwili $t = r + 1$. Wpływ wartości początkowych na ocenę wektora parametrów $\hat{\boldsymbol{\theta}}(t)$ jest analizowany w pracy [75].

Porównanie estymatorów SWLS i EWLS

Dla estymatorów WLS, przy założeniu stacjonarności sygnału, możemy wyznaczyć dwie charakterystyczne wielkości: efektywną szerokość okna $l_e(t)$ (efektywną liczbę obserwacji) oraz równoważną szerokość okna $l_r(t)$ (równoważną liczbę obserwacji). Wielkości te umożliwiają porównanie właściwości statycznych oraz dynamicznych estymatorów WLS o różnym kształcie okien ważących.

Efektywna szerokość okna opisana jest wzorem

$$l_e(t) = \sum_{i=0}^{t-1} w(i) \quad (3.42)$$

i może być interpretowana jako „powierzchnia okna”. Wielkość ta charakteryzuje ilość energii wyodrębnionej z sygnału w wyniku zastosowania ciągu ważącego $\{w(i)\}$. Dla estymatorów SWLS oraz EWLS współczynnik ten przyjmuje następującą postać

$$l_e^{\text{SWLS}}(t) = \min(t, N), \quad (3.43)$$

$$l_e^{\text{EWLS}}(t) = \frac{1 - \lambda^t}{1 - \lambda} \xrightarrow{t \rightarrow \infty} \frac{1}{1 - \lambda}. \quad (3.44)$$

Estymatory WLS, charakteryzujące się tą samą wartością współczynnika $l_e(t)$, wykazują jednakową efektywną długość pamięci. Z kolei równoważna szerokość okna $l_r(t)$ opisana jest wzorem

$$l_r(t) = \frac{\left[\sum_{i=0}^{t-1} w(i) \right]^2}{\sum_{i=0}^{t-1} w^2(i)} \quad (3.45)$$

i charakteryzuje ilość informacji o wektorze parametrów θ , która jest wydobywana z sygnału w wyniku zastosowania metody ważonych najmniejszych kwadratów. Dla estymatorów SWLS oraz EWLS współczynnik ten przyjmuje następującą postać

$$l_r^{\text{SWLS}}(t) = \min(t, N), \quad (3.46)$$

$$l_r^{\text{EWLS}}(t) = \frac{(1 - \lambda^t)(1 + \lambda)}{(1 + \lambda^t)(1 - \lambda)} \xrightarrow{t \rightarrow \infty} \frac{1 + \lambda}{1 - \lambda}. \quad (3.47)$$

Estymatory WLS charakteryzujące się tą samą wartością współczynnika $l_r(t)$, są równoważne z punktu widzenia dokładności estymacji, bez względu na kształt okna ważącego.

Z uwagi na zależność (3.26) otrzymujemy

$$l_r(t) \geq l_e(t) \quad (3.48)$$

przy czym równość zachodzi tylko dla okna prostokątnego. Oznacza to, że ilość informacji wydobywanej z sygnału poprzez zastosowanie okna ważącego innego niż prostokątne jest zawsze większa niż odpowiednia ilość energii. Rozbieżność pomiędzy efektywną, a równoważną liczbą obserwacji może być istotna. Przykładowo, dla okna wykładniczego (dla wartości λ bliskiej 1) otrzymujemy

$$\frac{1}{l_c(t)} = \frac{l_r(t)}{l_e(t)} = \frac{(1 + \lambda)}{(1 + \lambda^t)} \xrightarrow{t \rightarrow \infty} 1 + \lambda \cong 2, \quad (3.49)$$



gdzie $l_c(t)$ oznacza stopień koncentracji okna w pobliżu $i = 0$ (im mniejsza wartość tego współczynnika tym większy stopień koncentracji okna).

Porównanie właściwości statycznych oraz dynamicznych estymatorów WLS o różnym kształcie okien ma sens tylko wtedy, gdy charakteryzują się one tą samą równoważną szerokością okna $l_r(t)$. Przy spełnieniu tego warunku oraz w warunkach stacjonarności, estymatory WLS wykazują podobne właściwości statyczne – odpowiadające im macierze kowariancji błędów estymacji są w przybliżeniu takie same

$$\text{cov}[\hat{\boldsymbol{\theta}}_{\text{EWLS}}(t)] \approx \text{cov}[\hat{\boldsymbol{\theta}}_{\text{SWLS}}(t)]. \quad (3.50)$$

Analizując właściwości dynamiczne estymatorów WLS można wykazać, że estymatory EWLS mają nieco lepsze własności śledzące niż estymatory SWLS, co wynika z większego stopnia koncentracji okna wykładniczego w pobliżu $i = 0$. Skutkiem tego jest mniejsze opóźnienie estymacji i w konsekwencji mniejsze systematyczne błędy estymacji parametrów w trakcie śledzenia. Głębszą oraz obszerniejszą analizę właściwości estymatorów WLS można znaleźć w pracy [73].

3.3.2 Estymatory gradientowe

Metoda najmniejszych średnich kwadratów jest rekurencyjnym algorytmem gradientu stochastycznego stosowanym do śledzenia wolno zmieniającego się w czasie wektora parametrów $\boldsymbol{\theta}(t)$ niestacjonarnego procesu AR opisanego równaniem (3.23). Klasyczny estymator LMS zdefiniowany jest w następujący sposób

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \nu \boldsymbol{\varphi}(t) \varepsilon(t|t-1), \\ \varepsilon(t|t-1) &= y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t-1), \end{aligned} \quad (3.51)$$

gdzie ν , $\nu > 0$, oznacza krok adaptacji.

Analityczne badanie właściwości estymatora LMS wymaga przyjęcia silnych założeń co do sekwencji $\{\boldsymbol{\varphi}(t)\}$, $\{\boldsymbol{\theta}(t)\}$, $\{n(t)\}$ oraz zależności między nimi. W pracy [73] wykazano, że estymator LMS jest wykładniczo stabilny w warunkach wolnej adaptacji (małe wartości ν). Analiza właściwości estymatora LMS ujawnia też, że na szybkość zbieżności tego algorytm wpływa struktura kowariancyjna wektora regresji $\boldsymbol{\varphi}(t)$.

Niepożądaną z praktycznego punktu widzenia właściwością estymatora LMS jest zmienność długości jego pamięci, która zależy od struktury kowariancyjnej wektora regresji $\boldsymbol{\varphi}(t)$ oraz od mocy analizowanego sygnału [73].

Estymator NLMS

Ponieważ skalowanie sygnału nie wpływa na zawartość informacyjną analizowanych danych, dla ustalonego kroku adaptacji ν oczekivalibyśmy identycznych wyników estymacji bez względu na to jak dane zostały „wyskalowane” – niestety estymator LMS nie posiada tej właściwości. Wady tej pozbawiony jest unormowany algorytm najmniejszych średnich kwadratów. Algorytm NLMS, który jest niewrażliwy na skalowanie sygnału może zostać otrzymany przez zastąpienie kroku adaptacji ν poprzez unormowany krok adaptacji

$$\frac{\bar{\nu}}{\|\boldsymbol{\varphi}(t)\|^2}. \quad (3.52)$$

W rezultacie otrzymujemy estymator postaci

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\bar{\nu}}{\|\boldsymbol{\varphi}(t)\|^2} \boldsymbol{\varphi}(t) \varepsilon(t|t-1). \quad (3.53)$$

Aby uniknąć problemów numerycznych pojawiających się, gdy norma kwadratowa $\|\boldsymbol{\varphi}(t)\|^2$ przyjmie zbyt małe (bliskie lub równe zero) wartości, mianownik w (3.53) jest zazwyczaj zastępowany przez

$$\epsilon_0 + \|\boldsymbol{\varphi}(t)\|^2, \quad (3.54)$$

gdzie ϵ_0 jest stałą liczbą o małej, dodatniej wartości, pełniącą rolę zabezpieczenia przed zbyt dużym, niekontrolowanym wzrostem unormowanego kroku adaptacji.

Kolejny wariant estymatora LMS, również nieczuły na skalowanie sygnału, realizuje normalizację kroku adaptacji poprzez wykorzystanie techniki zapominania wykładniczego dając

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \frac{\bar{\nu}}{r(t)} \boldsymbol{\varphi}(t) \varepsilon(t|t-1), \\ r(t) &= \lambda r(t-1) + \|\boldsymbol{\varphi}(t)\|^2. \end{aligned} \quad (3.55)$$

Nazwa tego algorytmu – z ang. *Trace Least Mean Square*, w skrócie TLMS – jest związana z faktem, że współczynnik normalizujący $r(t)$ jest identyczny ze śladem macierzy $\mathbf{R}(t)$ w algorytmie EWLS

$$r(t) = \sum_{i=0}^{t-1} \lambda^i \|\boldsymbol{\varphi}(t-i)\|^2 = \text{tr} \left\{ \sum_{i=0}^{t-1} \lambda^i \boldsymbol{\varphi}(t-i) \boldsymbol{\varphi}^T(t-i) \right\} = \text{tr} \{ \mathbf{R}(t) \}. \quad (3.56)$$

3.3.3 Porównanie estymatorów WLS i LMS

Próbując ocenić zakres stosowalności estymatorów WLS i LMS z punktu widzenia praktycznego, należy wziąć pod uwagę między innymi takie aspekty algorytmu jak: szybkość zbieżności, zdolności predykcyjne, szybkość reakcji na nagłe zmiany, złożoność obliczeniowa, bezpieczeństwo numeryczne oraz wygoda stosowania.

Mimo iż estymatory WLS i LMS różnią się szybkością zbieżności w początkowej fazie działania na korzyść estymatora WLS, w stanie ustalonym wykazują podobne zdolności predykcyjne oraz cechują się podobną szybkością reakcji na gwałtowne zmiany estymowanych parametrów [73]. Można wskazać zastosowania wymagające szybkiej początkowej zbieżności estymatora (np. eliminacja echa w zastosowaniach telekomunikacyjnych), oraz takie, w których ten aspekt jest pomijalny na rzecz złożoności obliczeniowej estymatora (np. aktywne tłumienie hałasu). W przypadku eliminacji zakłóceń impulsowych brane są pod uwagę takie aspekty jak: zdolności predykcyjne, bezpieczeństwo numeryczne oraz wygoda stosowania. Jak już wspomniano, pamięć estymatora LMS zależy nie tylko od nastaw parametrów projektowych, ale również od struktury kowariancyjnej wektora regresji $\boldsymbol{\varphi}(t)$ oraz od mocy analizowanego sygnału co oznacza, że może się zmieniać w trakcie działania algorytmu w sposób niekontrolowany przez użytkownika. Problem ten nie występuje w przypadku estymatora EWLS. Efektywna długość jego pamięci jest stała i zależy wyłącznie od przyjętej wartości stałej zapominania λ . Ani szybkość zbieżności, ani pamięć estymatora nie zależą od struktury kowariancyjnej wektora regresji $\boldsymbol{\varphi}(t)$. Warto też zwrócić uwagę na to, że wyniki estymacji nie zależą od „wyskalowania” sygnału.



3.3.4 Estymator wariancji białego szumu

Pokażemy teraz, że ocena wariancji białego szumu tworzącego $\hat{\rho}(t)$ może być liczona rekurencyjnie w dwóch wariantach: sprzężonym oraz niesprzężonym z algorytmem EWLS.

Wariant sprzężony z EWLS

W pierwszym podejściu ocena wariancji szumu tworzącego może być otrzymywana na drodze uśredniania sekwencji ważonych kwadratów błędów resztowych w następujący sposób

$$\hat{\rho}(t) = \frac{1}{l_e(t)} \sum_{i=1}^t \lambda^{t-i} \left[y(i) - \boldsymbol{\varphi}^T(i) \hat{\boldsymbol{\theta}}(t) \right]^2. \quad (3.57)$$

Zwróćmy uwagę na to, że pamięć tego estymatora jest identyczna z pamięcią estymatora EWLS. Estymator $\hat{\rho}(t)$ w postaci (3.57) wymaga, aby za każdym razem obliczać nową sekwencję błędów resztowych względem aktualnej oceny wektora parametrów $\hat{\boldsymbol{\theta}}(t)$, co w praktyce jest obliczeniowo nieefektywne. Rekurencyjny algorytm estymacji ma następującą postać (patrz wyprowadzenie zależności w dodatku B)

$$\hat{\rho}(t) = \lambda \hat{\rho}(t-1) + (1-\lambda) \varepsilon^2(t|t-1) \varsigma(t), \quad (3.58)$$

gdzie

$$\varsigma(t) = \left[1 - \boldsymbol{\varphi}^T(t) \mathbf{Q}(t) \boldsymbol{\varphi}(t) \right] = \left[1 - \mathbf{k}^T(t) \boldsymbol{\varphi}(t) \right] = \frac{\lambda}{\lambda + \boldsymbol{\varphi}^T(t) \mathbf{Q}(t-1) \boldsymbol{\varphi}(t)} \quad (3.59)$$

jest współczynnikiem łatwym do obliczenia w oparciu o wyniki estymacji otrzymane przy użyciu algorytmu EWLS.

Wariant niesprzężony z EWLS

W drugim podejściu wykładniczo ważona estymata $\hat{\rho}(t)$ może być otrzymana na drodze uśredniania ostatnio zaobserwowanych kwadratów błędów jednokrokowej predykcji według następującej formuły

$$\hat{\rho}(t) = \frac{1}{l_e(t)} \sum_{i=1}^t \lambda_0^{t-i} \left[y(i) - \boldsymbol{\varphi}^T(i) \hat{\boldsymbol{\theta}}(i-1) \right]^2 = \frac{1}{l_e(t)} \sum_{i=1}^t \lambda_0^{t-i} \varepsilon^2(i|i-1), \quad (3.60)$$

gdzie λ_0 , $0 < \lambda_0 < 1$, jest kolejną stałą zapomnienia określającą pamięć estymacji algorytmu uśredniania. Pamięć tego estymatora jest niezależna od pamięci estymatora EWLS. Można pokazać, że estymator w wersji rekurencyjnej ma w stanie ustalonym postać (patrz wyprowadzenie zależności w dodatku B)

$$\hat{\rho}(t) = \lambda_0 \hat{\rho}(t-1) + (1-\lambda_0) \varepsilon^2(t|t-1). \quad (3.61)$$

Uwaga 3.2. Estymatory (3.58) i (3.61) dają w stanie ustalonym zbliżone oszacowania pod warunkiem, że $\lambda_0 = \lambda$ oraz $\varsigma(t)$ przyjmuje wartości bliskie 1. Gdy algorytm EWLS jest używany do identyfikacji modelu AR (3.23), obie wielkości $\boldsymbol{\theta}(t)$ i $\rho(t)$ są śledzone z tą samą szybkością/dokładnością, odpowiednio przez estymatory $\hat{\boldsymbol{\theta}}(t)$ (3.37) i $\hat{\rho}(t)$ (3.58) związane z tą samą stałą zapomnienia λ . Ponieważ eksperymenty na rzeczywistych nagraniach fonicznych pokazują, że wariancja szumu tworzącego $\rho(t)$ podlega często szybszym zmianom

niż wektor parametrów $\theta(t)$, w celu detekcji zakłóceń impulsowych korzystniej jest stosować estymator $\hat{\rho}(t)$ o krótszej pamięci co możliwe jest w przypadku użycia algorytmu (3.61) ze stałą zapominania λ_0 , $\lambda_0 < \lambda$. Taka modyfikacja umożliwi detektorowi zakłóceń impulsowych szybsze reagowanie na nagłe zmiany wartości $\rho(t)$.

Uwaga 3.3. Obie wielkości $\theta(t)$ i $\rho(t)$ są śledzone dopóki pomiary sygnału są klasyfikowane jako niezakłócone. Dopiero gdy alarm detekcyjny zostanie podniesiony, tj. gdy $\hat{d}(t_0 + 1) = 1$, najświeższe oszacowania $\hat{\theta}(t_0)$ i $\hat{\rho}(t_0)$ zostają czasowo „zamrożone”. Estymacja obu wielkości wznowiana jest gdy rekonstrukcja zakwestionowanego fragmentu zostanie zakończona (z wykorzystaniem interpolowanych próbek w miejsce tych uszkodzonych).

3.3.5 Monitorowanie stabilności modelu AR

W większości aplikacji fonicznych, do których również zalicza się problem adaptacyjnej detekcji/rekonstrukcji, stabilność modelu sygnału musi być zagwarantowana, aby oparta na nim analiza była dobrze uwarunkowana.

Twierdzenie 3.1. Model AR rzędu r o współczynnikach autoregresji a_i , $i = 1, \dots, r$, jest stabilny asymptotycznie wtedy i tylko wtedy gdy wszystkie zera z_i , $i = 1, \dots, 2r$, jego wielomianu charakterystycznego $\mathbb{A}(z^{-1})$

$$\mathbb{A}(z^{-1}) = 1 - \sum_{i=1}^r a_i z^{-i} \quad (3.62)$$

leżą wewnątrz okręgu jednostkowego na płaszczyźnie zespolonej:

$$|z_i| < 1, \quad i = 1, \dots, 2r. \quad (3.63)$$

Test stabilności wielomianu $\mathbb{A}(z^{-1})$ może polegać na sprawdzeniu wartości bezwzględnych jego „współczynników odbicia” $a_{k,k}$, $k = 1, \dots, r$, wyliczonych przy użyciu tzw. „wstecznego” algorytmu Levinsona-Durbina [75]. Jeżeli

$$|a_{k,k}| < 1 \quad \text{dla każdego } k = 1, \dots, r, \quad (3.64)$$

warunek stabilności (3.63) jest spełniony. Współczynniki odbicia mogą zostać otrzymane na podstawie następującej rekurencyjnej procedury

$$a_{k,i} = (a_{k+1,i} - a_{k+1,k+1} a_{k+1,k+1-i}) / (1 - a_{k+1,k+1}^2), \quad (3.65)$$

$$i = 1, \dots, k, \quad k = r - 1, \dots, 1,$$

z warunkami początkowymi: $a_{r,i} = -a_i$, $i = 1, \dots, r$. Innym podejściem do testowania stabilności wielomianu $\mathbb{A}(z^{-1})$ może być użycie jednego z wariantów procedury Schura-Cohna, patrz praca [76] oraz zawarte w niej odniesienia. Oba testy stabilności są równoważne – [72], [76].

Niestety, gdy prawdziwe parametry modelu zostają zastąpione przez ich estymaty EWLS $\hat{a}_i(t)$ nie ma gwarancji, że otrzymany model jest stabilny. Ponieważ niestabilne modele mogą



proceedzić do błędów na etapie detekcji i interpolacji (na przykład mogą być źródłem artefaktów o oscylacyjnym charakterze, pojawiających się gdy interpolacja sygnału oparta jest na niestabilnym modelu), stabilność modelu jest ważnym praktycznym wymogiem w przypadku analizy sygnałów fonicznych. Dlatego, za każdym razem gdy podniesiony zostaje alarm detekcyjny, zalecane jest sprawdzenie stabilności modelu. Test stabilności wielomianu $\hat{A}(z^{-1}, t)$,

$$\hat{A}(z^{-1}, t) = 1 - \sum_{i=1}^r \hat{a}_i(t) z^{-i}, \quad (3.66)$$

przeprowadzany jest zgodnie z (3.64)-(3.65), z warunkami początkowymi: $a_{r,i} = -\hat{a}_i(t)$, $i = 1, \dots, r$. Jeżeli warunek stabilności nie jest spełniony, parametry modelu estymowane są ponownie przy użyciu algorytmu Levinsona-Durbina (LD) gwarantującego stabilność modelu – szczegółowy opis tego algorytmu i dyskusję jego właściwości znaleźć można w monografii [75].

W podejściu tym metoda najmniejszych kwadratów stosowana jest do sekwencji danych $\{y(t-N+1), \dots, y(t-1), y(t)\}$. Po obustronnym rozszerzeniu tej sekwencji przez r próbek o zerowych wartościach (tzw. autokorelacyjna technika estymacji), optymalna w sensie średniokwadratowym estymata wektora parametrów może zostać wyrażona w postaci

$$\hat{\boldsymbol{\theta}}(t) = \hat{\mathbf{R}}^{-1}(t) \hat{\mathbf{p}}(t), \quad (3.67)$$

gdzie

$$\hat{\mathbf{R}}(t) = \begin{bmatrix} \hat{r}_0(t) & \cdots & \hat{r}_{r-1}(t) \\ \vdots & \ddots & \vdots \\ \hat{r}_{r-1}(t) & \cdots & \hat{r}_0(t) \end{bmatrix}, \quad \hat{\mathbf{p}}(t) = \begin{bmatrix} \hat{r}_1(t) \\ \vdots \\ \hat{r}_r(t) \end{bmatrix}$$

i gdzie $\hat{r}_k(t)$, $k = 0, \dots, r$, oznacza lokalną ocenę współczynnika autokorelacji

$$r_k = E[y(t)y(t-k)]$$

otrzymywaną w następujący sposób

$$p_k(t, N) = \sum_{i=0}^{N-k-1} y(t-i)y(t-i-k), \quad (3.68)$$

$$\hat{r}_k(t) = \frac{1}{N} p_k(t, N).$$

Zauważmy, że w przypadku stosowania tej procedury w każdej chwili czasu $t > N$ dla przesuwającej się sekwencji danych (tzw. podejście SWLS), współczynniki $p_k(t, N)$ mogą być zaktualizowane przy użyciu następującej rekurencyjnej procedury, analogicznej do (3.33)-(3.34),

$$p_k(t, N+1) = p_k(t-1, N) + y(t)y(t-k),$$

$$p_k(t, N) = p_k(t, N+1) - y(t-N)y(t-N+k). \quad (3.69)$$

Zauważmy także, że podczas gdy wielkości $p_0(t), \dots, p_r(t)$ powinny być na bieżąco aktualizowane, to już oceny $\hat{\boldsymbol{\theta}}(t)$ mogą być wyznaczone „na żądanie”.

Aby uwzględnić długość pamięci algorytmu EWLS, wielkości N nadaje się wartość równą równoważnej szerokości $l_r(\infty)$ stosowanego okna wykładniczego. Ważną właściwością algorytmu LD jest to, że gwarantuje on stabilność otrzymywanego modelu AR pod warunkiem, że macierz regresji $\hat{\mathbf{R}}(t)$ jest dodatnio określona [co jest zapewnione gdy używane są obciążone estymaty współczynników autokorelacji (3.68)].



Warto w tym miejscu podkreślić, że algorytm LD jest traktowany jako „ratunkowa” procedura estymacji wielkości $\theta(t)$, używana *tylko* wtedy gdy model EWLS jest niestabilny w momencie zgłoszenia alarmu detekcyjnego (co nie zdarza się często). Gdy algorytm LD używany jest przez cały czas, tj. zamiast algorytmu EWLS, wyniki detekcji/interpolacji ulegają pogorszeniu z powodu gorszych zdolności predykcyjnych otrzymywanego modelu AR. Wydaje się, że jest to „cena” jaką ponosi się za gwarancję stabilności modelu. Z drugiej strony, jeżeli etap sprawdzania/wymuszania stabilności zostanie pominięty, otrzymane od czasu do czasu modele niestabilne mogą powodować duże błędy w rekonstrukcji sygnału.

3.3.6 Eliminacja składowej stałej

Stosując model AR (3.2), zakładamy, że modelowany sygnał ma zerową wartość oczekiwaną: $E[s(t)] = 0$. Ponieważ dla typowego nagrania fonicznego założenie to nie jest spełnione, problem niezerowej składowej stałej (czyli tzw. offsetu) powinien zostać w pewien sposób rozwiązany.

Bezpośrednie rozwiązanie tego problemu polega na uwzględnieniu offsetu w modelu AR, tj. na użyciu następującego opisu w miejsce (3.2)

$$s(t) = \sum_{i=1}^r a_i s(t-i) + \bar{s} + n(t), \quad (3.70)$$

gdzie \bar{s} oznacza nieznaną stałą. Ponieważ po uwzględnieniu dodatkowego elementu w wektorach $\theta = [a_1, \dots, a_r, \bar{s}]^T$ i $\varphi(t) = [y(t-1), \dots, y(t-r), 1]^T$ skrócony opis sygnału (3.3) pozostaje niezmienny, identyfikacji rozszerzonego modelu (3.70) można dokonać w dokładnie ten sam sposób jak to opisano wcześniej. Z tego samego powodu algorytm filtru Kalmana o zmiennym rzędzie nie wymaga żadnej modyfikacji.

Pośrednie rozwiązanie problemu offsetu polega na usunięciu niezerowej składowej stałej przed zastosowaniem procedury detekcji/interpolacji. Taki zabieg centrowania sygnału może zostać w łatwy sposób zrealizowany na drodze przepuszczenia go przez filtr górnoprzepustowy $H(z^{-1})$ postaci

$$H(z^{-1}) = \frac{k_c(1 - z^{-1})}{1 - k_c z^{-1}}, \quad (3.71)$$

gdzie k_c , $0 < k_c < 1$, oznacza stałą kontrolującą szerokość pasma, która powinna być bliska wartości 1. Przewaga rozwiązania pośredniego nad bezpośrednim, polega na większej elastyczności wynikającej z faktu, że stałe λ i k_c mogą być dobrane niezależnie od siebie (w przypadku użycia rozszerzonego modelu (3.70), szybkość śledzenia, o której decyduje stała zapominania λ jest taka sama dla wszystkich współczynników modelu).

3.4 Metody oparte na analizie błędów predykcji

Procedura detekcji zakłóceń impulsowych związana jest z analizą kolejnych błędów predykcji. Zgodnie z przyjętym modelem (3.2) optymalną (minimalnowariancyjną) jednokrokową predykcję sygnału $s(t)$ opisuje zależność

$$\hat{s}(t|t-1) = \sum_{i=1}^r a_i s(t-i). \quad (3.72)$$

Zauważmy, że w przypadku gdy ostatnich r pomiarów stanowią próbki niezakłócone, co odpowiada sytuacji gdy $d(t-r) = \dots = d(t-1) = 0$, równanie (3.72) możemy zastąpić przez

$$\hat{y}(t|t-1) = \sum_{i=1}^r a_i y(t-i), \quad (3.73)$$

gdzie błąd jednokrokowej predykcji definiowany jest następująco

$$\varepsilon(t|t-1) = y(t) - \hat{y}(t|t-1). \quad (3.74)$$

Ponieważ charakterystyki sygnału zmieniają się w czasie, równanie (3.73) zastąpione zostaje przez

$$\hat{y}(t|t-1) = \sum_{i=1}^r \hat{a}_i(t) y(t-i), \quad (3.75)$$

gdzie oszacowania parametrów modelu otrzymane są w wyniku bieżącej identyfikacji modelu przez algorytm EWLS (3.39).

Adaptacyjna procedura detekcji jest rozpoczynana za każdym razem, gdy zostanie podniesiony alarm detekcyjny: $\hat{d}(t_0+1) = 1$, co oznacza, że próbka w chwili t_0+1 jest traktowana jako zakłócona. Dzieje się tak wtedy, gdy wartość bezwzględna błędu jednokrokowej predykcji przekroczy μ razy jego średnie odchylenie standardowe $\sigma_\varepsilon(t_0+1|t_0)$

$$\hat{d}(t_0+1) = \begin{cases} 1, & \text{jeżeli } |\varepsilon(t_0+1|t_0)| > \mu\sigma_\varepsilon(t_0+1|t_0), \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad (3.76)$$

gdzie μ oznacza mnożnik dobierany eksperymentalnie (zazwyczaj najlepsze wyniki otrzymuje się dla $\mu \in [3, 5]$), a $\sigma_\varepsilon(t_0+1|t_0) = \sqrt{\hat{\rho}(t_0)}$ – wielkość $\hat{\rho}(t)$ otrzymana jest w wyniku bieżącej estymacji przez algorytm (3.61).

Uwaga 3.4. Wartość $\mu = 3$ odpowiada dobrze znanej regule „3-sigma” używanej do detekcji próbek „odstających” (z ang. *outliers*) w sygnałach gaussowskich. Ponieważ sygnały foniczne są z reguły niegaussowskie, bardzo często lepsze wyniki otrzymuje się dla $\mu > 3$.

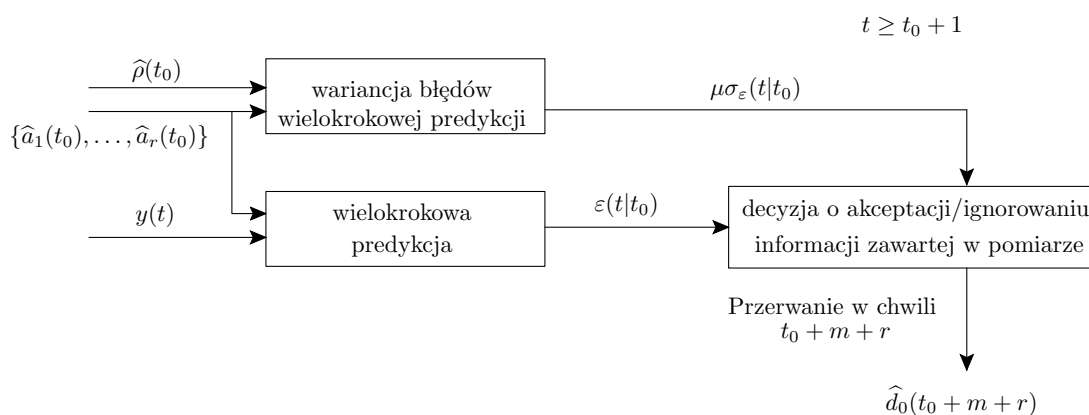
Test ten zostaje rozszerzony na kolejne błędy predykcji $\varepsilon(t)$ dla $t > t_0+1$

$$\hat{d}_0(t) = \begin{cases} 1, & \text{jeżeli } |\varepsilon(t)| > \mu\sigma_\varepsilon(t), \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad (3.77)$$

gdzie $\hat{d}_0(t)$ przechowuje informacje o tymczasowo podjętych decyzjach dotyczących akceptacji/odrzućcia kolejnych pomiarów. Wielkości $\varepsilon(t)$, $\sigma_\varepsilon(t)$ mogą zostać obliczone przy użyciu schematu detekcji w pętli otwartej lub przy użyciu schematu detekcji z decyzyjnym sprzężeniem zwrotnym. Wspomniane schematy detekcji zostaną opisane w kolejnych dwóch punktach tego rozdziału.

Alarm detekcyjny jest przerywany w chwili t_0+m+1 jeżeli r kolejnych błędów predykcji przyjmie odpowiednio małe wartości

$$|\varepsilon(t_0+m+i)| \leq \mu\sigma_\varepsilon(t_0+m+i), \quad i = 1, \dots, r, \quad (3.78)$$



Rysunek 3.2: Schemat ideowy adaptacyjnej procedury detekcji w pętli otwartej. Oznaczenia: $t_0 + 1$ oznacza chwilę rozpoczęcia pracy algorytmu detekcji, m oznacza długość alarmu detekcyjnego, r oznacza rząd modelu autoregresyjnego, $\{\hat{a}_1(t_0), \dots, \hat{a}_r(t_0)\}$ oznaczają oceny współczynników autoregresji, $\hat{\rho}(t_0)$ oznacza lokalną ocenę wariancji białego szumu tworzącego, $\varepsilon(t|t_0)$ oznacza błąd wielokrokowej predykcji, μ jest parametrem projektowym (mnożnikiem), $\sigma_\varepsilon(t|t_0)$ oznacza średnie odchylenie standardowe błędu wielokrokowej predykcji, a $\hat{d}_0(t)$ przechowuje informacje o tymczasowo podjętych decyzjach. Algorytm detekcji zostaje przerwany w chwili $t_0 + m + r$, gdy r kolejnych pomiarów $y(t)$ z rzędu zostało zaakceptowanych, tzn. gdy r razy z rzędu spełniony został warunek $|\varepsilon(t|t_0)| \leq \mu\sigma_\varepsilon(t|t_0)$.

lub gdy liczba zakwestionowanych próbek osiągnie swoją maksymalną dopuszczalną (ustaloną z góry) wartość m_{\max} . Ostatecznie, alarm detekcyjny przyjmuje postać nieprzerwanego bloku „jedynek”: $\hat{d}(t) = 1$ w całym przedziale $t \in [t_0 + 1, t_0 + m]$.

Gdy procedura detekcji zostanie zakończona, sekwencja uszkodzonych próbek sygnału $\{s(t_0 + 1), \dots, s(t_0 + m)\}$ zostaje zastąpiona sekwencją interpolacji $\{\hat{s}(t_0 + 1), \dots, \hat{s}(t_0 + m)\}$ otrzymaną na podstawie dostępnego modelu sygnału – patrz podrozdział (3.2).

3.4.1 Detekcja w pętli otwartej

W tym podejściu procedura detekcji jest kontynuowana w oparciu o wyniki wielokrokowej predykcji, $\varepsilon(t) = \varepsilon(t|t_0)$, $\sigma_\varepsilon(t) = \sigma_\varepsilon(t|t_0)$ dla $t > t_0 + 1$, tzn. wartości bezwzględne błędów k -krokowej predykcji

$$\varepsilon(t_0 + k|t_0) = y(t_0 + k) - \hat{y}(t_0 + k|t_0), \quad k = 2, 3, \dots, \quad (3.79)$$

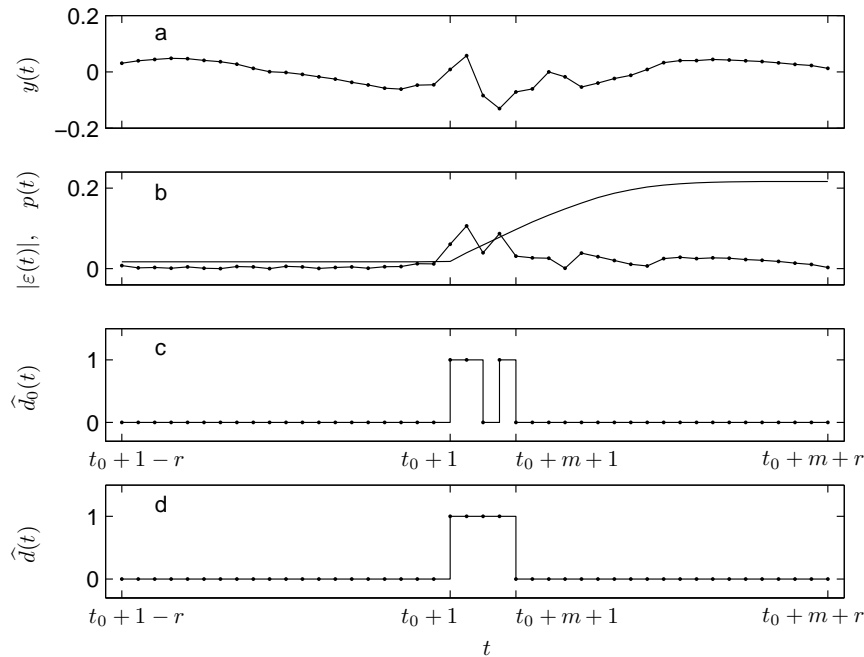
są porównywane z odpowiadającymi im progami według następującego schematu

$$\hat{d}_0(t_0 + k) = \begin{cases} 1, & \text{jeżeli } |\varepsilon(t_0 + k|t_0)| > \mu\sigma_\varepsilon(t_0 + k|t_0), \\ 0, & \text{w pozostałych przypadkach.} \end{cases} \quad (3.80)$$

Wielkość $\hat{y}(t_0 + k|t_0)$ może zostać otrzymana w wyniku złożenia k operacji jednokrokowej predykcji

$$\hat{y}(t_0 + j|t_0) = \sum_{i=1}^r \hat{a}_i(t_0) \hat{y}(t_0 + j - i|t_0), \quad j = 1, \dots, k, \quad (3.81)$$

gdzie $\hat{y}(t_0 + j|t_0) = y(t_0 + j|t_0)$ dla $j \leq 0$. Natomiast $\sigma_\varepsilon(t_0 + k|t_0) = \sqrt{\rho_\varepsilon(t_0 + k|t_0)}$ oznacza średnie odchylenie standardowe błędu wielokrokowej predykcji, gdzie wielkość $\rho_\varepsilon(t_0 + k|t_0)$ można



Rysunek 3.3: Typowe zachowanie algorytmu detekcji w pętli otwartej w obecności rzeczywistego zakłócenia impulsowego. Kolejne wykresy przedstawiają: sygnał foniczny (a), wartości bezwzględne błędów predykcji (b – linia ciągła z naniesionymi kropkami), próg detekcyjny (b – linia ciągła), tymczasowe decyzje podjęte na etapie detekcji (c), funkcję lokalizacji zakłóceń (d). Praca algorytmu detekcji rozpoczyna się w chwili $t_0 + 1$ i zostaje przerwana w chwili $t_0 + m + r$. Alarm detekcyjny kończy się w chwili $t_0 + m + 1$, gdy r kolejnych błędów predykcji przyjmuje wystarczająco małe wartości.

wyznaczyć rekurencyjnie za pomocą następującego algorytmu zaproponowanego przez Stoicę [77]

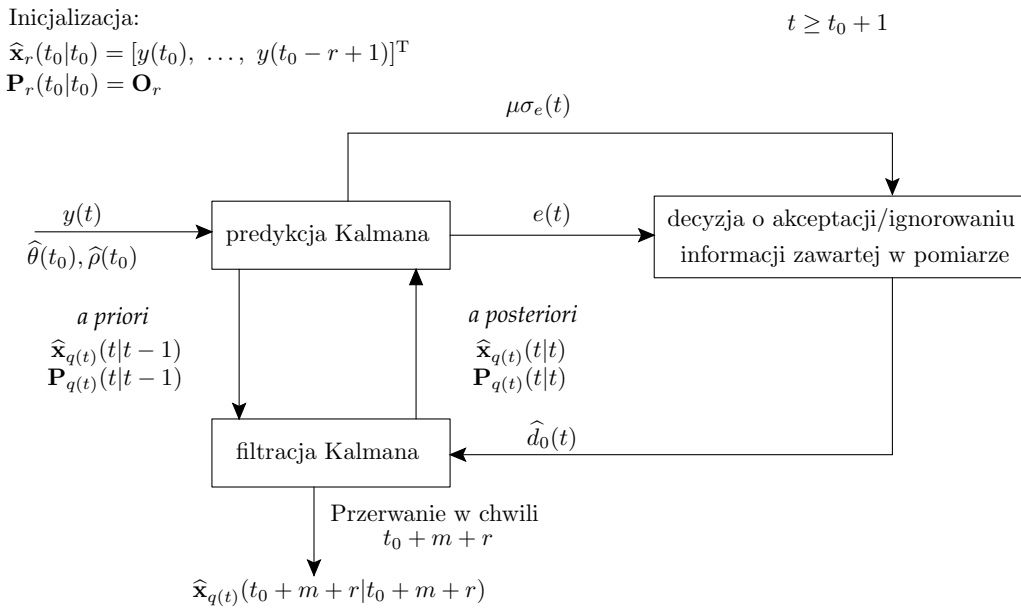
$$\begin{aligned}
 \rho_\varepsilon(t_0 + k|t_0) &= \rho_\varepsilon(t_0 + k - 1|t_0) + \hat{\rho}(t_0)f_{k-1}^2, \\
 f_{k-1} &= g_{k-1}^0, \\
 g_k^i &= g_{k-1}^{i+1} + \hat{a}_{i+1}(t_0)f_{k-1}, \\
 i &= 0, \dots, r - 1, \\
 k &= 2, \dots, m + r,
 \end{aligned} \tag{3.82}$$

z warunkami początkowymi: $\rho_\varepsilon(t_0 + 1|t_0) = \hat{\rho}(t_0)$, $f_0 = 1$ i $g_1^i = \hat{a}_{i+1}(t_0)$, $i = 0, \dots, r - 1$, $g_{k-1}^r = 0$, $k = 2, \dots, m + r$.

Schemat ideowy procedury detekcji w pętli otwartej został przedstawiony na rysunku 3.2. Z kolei, na rysunku 3.3 została zilustrowana typowa praca opisanego powyżej algorytmu detekcji.

3.4.2 Detekcja z decyzyjnym sprzężeniem zwrotnym

Prosty schemat detekcji w pętli otwartej może zostać zastąpiony bardziej wyszukany podejściem opartym na predykcji sygnału z decyzyjnym sprzężeniem zwrotnym. W podejściu tym błędy predykcji $\varepsilon(t) = e(t)$ oraz odpowiadające im średnie odchylenia standardowe



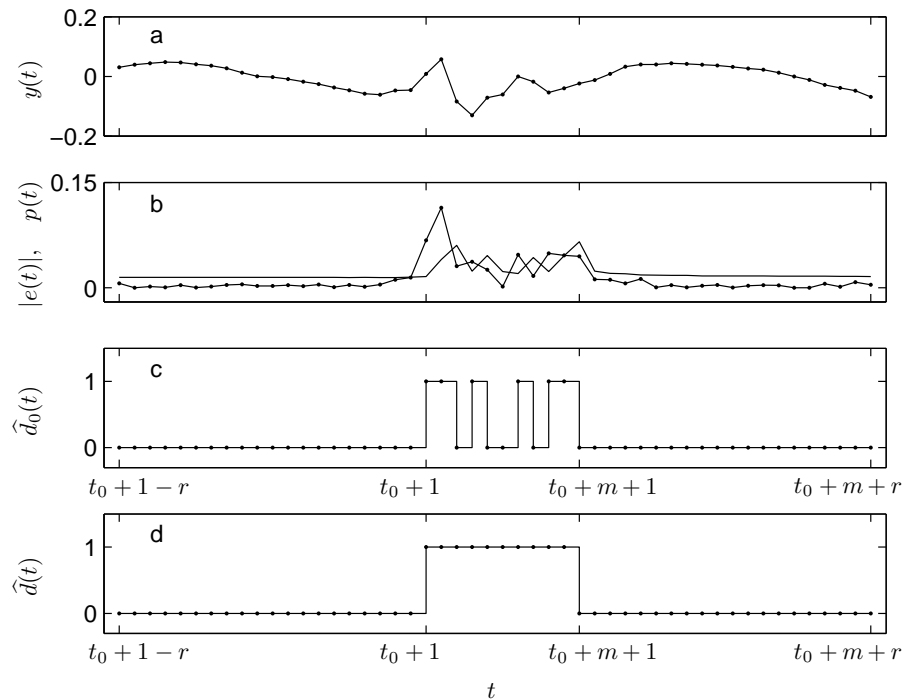
Rysunek 3.4: Schemat ideowy procedury detekcji z decyzyjnym sprzężeniem zwrotnym zrealizowany za pomocą algorytmu filtracji Kalmana o zmiennym rzędzie. Oznaczenia: $t_0 + 1$ oznacza chwilę rozpoczęcia pracy algorytmu detekcji, m oznacza długość alarmu detekcyjnego, r oznacza rząd modelu autoregresyjnego, $q(t)$ oznacza zmienny w czasie rząd filtru Kalmana, \mathbf{O}_r oznacza macierz zerową o wymiarach $r \times r$, $\hat{\boldsymbol{\theta}}(t_0)$ oznacza ocenę wektora parametrów, $\hat{\rho}(t_0)$ oznacza lokalną ocenę wariancji białego szumu tworzącego, $\hat{\mathbf{x}}_{q(t)}(t|t-1)/\hat{\mathbf{x}}_{q(t)}(t|t)$ oznacza ocenę wektora stanu a priori/a posteriori, $\mathbf{P}_{q(t)}(t|t-1)/\mathbf{P}_{q(t)}(t|t)$ oznacza macierz kowariancji a priori/a posteriori, $e(t)$ oznacza błąd predykcji, μ jest parametrem projektowym (mnożnikiem), $\sigma_e(t)$ oznacza średnie odchylenie standardowe błędu predykcji, a $\hat{d}_0(t)$ przechowuje informacje o tymczasowo podjętych decyzjach dotyczących akceptacji/ignorowania kolejnych pomiarów w algorytmie filtracji Kalmana. Algorytm detekcji zostaje przerwany w chwili $t_0 + m + r$, gdy r kolejnych pomiarów $y(t)$ z rzędu zostało zaakceptowanych, tzn. gdy r razy z rzędu spełniony został warunek $|e(t)| \leq \mu\sigma_e(t)$.

$\sigma_\varepsilon(t) = \sigma_e(t) = \sqrt{\rho_e(t)}$ są na bieżąco obliczane za pomocą algorytmu filtracji Kalmana o zmiennym rzędzie, zainicjowanego w chwili t_0 – patrz opis w punkcie 3.2.3. W każdej chwili pracy algorytmu stosowana jest następująca reguła decyzyjna

$$\hat{d}_0(t) = \begin{cases} 1, & \text{jeżeli } |e(t)| > \mu\sigma_e(t), \\ 0, & \text{w pozostałych przypadkach.} \end{cases} \quad (3.83)$$

W przypadku ignorowania pomiaru, $\hat{d}_0(t) = 1$, aktualizacja wielkości $\hat{\mathbf{x}}_{q(t)}(t|t)$ i $\mathbf{P}_{q(t)}(t|t)$ odbywa się zgodnie z (3.20). Z kolei w przypadku akceptacji pomiaru, $\hat{d}_0(t) = 0$, aktualizacja przeprowadzana jest według (3.21).

Schemat ideowy procedury detekcji z decyzyjnym sprzężeniem zwrotnym został przedstawiony na rysunku 3.4. Rysunek 3.5 ilustruje działanie opisanego powyżej algorytmu detekcji.



Rysunek 3.5: Typowe zachowanie algorytmu detekcji z decyzyjnym sprzężeniem zwrotnym w obecności rzeczywistego zakłócenia. Kolejne wykresy przedstawiają: sygnał foniczny (a), wartości bezwzględne błędów predykcji (b – linia ciągła z naniesionymi punktami), próg detekcyjny (b – linia ciągła), tymczasowe decyzje podjęte na etapie detekcji (c), funkcję lokalizacji zakłóceń (d). Praca algorytmu detekcji rozpoczyna się w chwili t_0+1 i zostaje przerwana w chwili t_0+m+r . Alarm detekcyjny kończy się w chwili t_0+m+1 , gdy r kolejnych błędów predykcji przyjmuje wystarczająco małe wartości.

Związki między podejściami

Na początek zauważmy, że zachowanie obu algorytmów detekcji jest identyczne

$$\begin{aligned} e(t_0+k) &= \varepsilon(t_0+k|t_0), \quad 1 \leq k \leq m, \\ \sigma_e(t_0+k) &= \sigma_\varepsilon(t_0+k|t_0), \end{aligned} \quad (3.84)$$

jedynie wtedy, gdy

$$\hat{d}_0(t_0+1) = \dots = \hat{d}_0(t_0+m) = 1, \quad (3.85)$$

tj. gdy wszystkie próbki z przedziału $[t_0+1, t_0+m]$ traktowane są na etapie detekcji jako zakłócone. Gdy warunek (3.85) nie jest spełniony, zachowanie algorytmu detekcji z decyzyjnym sprzężeniem zwrotnym różni się od zachowania algorytmu detekcji w pętli otwartej. Wynika to z faktu, że każda decyzja o akceptacji pomiaru $y(t)$ z wnętrza przedziału $t \in [t_0+2, t_0+m-1]$

$$|e(t)| \leq \mu \sigma_e(t) \Rightarrow \hat{d}_0(t) = 0 \quad (3.86)$$

ma istotny wpływ zarówno na kolejne predykcje sygnału jak i na obniżenie progu detekcji dla kolejnych błędów predykcji. Dzieje się tak dlatego, że przewidywana wartość sygnału otrzymywana z zależności $\mathbf{c}_{q(t)}^T \hat{\mathbf{x}}_{q(t)}(t|t-1)$, zależy nie tylko od próbek zgromadzonych do chwili t_0 , ale

także od próbek, które zostały zaakceptowane po chwili $t_0 + 1$. Stąd przyjęta nazwa – detekcja z decyzyjnym sprzężeniem zwrotnym.

Natomiast w przypadku algorytmu detekcji w pętli otwartej, wraz ze wzrostem horyzontu predykcji, obserwujemy wzrost wariancji błędu wielokrokowej predykcji, która zmierza do estymowanej wariancji sygnału. Gdy model jest stabilny, wartości wielokrokowej predykcji zmierzają z czasem do zera. Oznacza to, że detektor zakłóceń staje się coraz bardziej tolerancyjny wobec powstających rozbieżności między obserwowanymi, a przewidywanymi wartościami sygnału.

Zwróćmy uwagę, że w obu omówionych podejściach nie możemy polegać na tymczasowych decyzjach o akceptacji próbek $[\hat{d}_0(t) = 0]$ z przedziału $t \in [t_0 + 2, t_0 + m - 1]$, ponieważ w większości przypadków są to uszkodzone próbki, które zostały tylko „przypadkowo zaakceptowane” – patrz rysunki 3.3 i 3.5.

Uwaga 3.5. Istotne jest, aby alarm detekcyjny $\hat{d}(t)$ nie uwzględniał tymczasowych decyzji o akceptacji próbki $[\hat{d}_0(t) = 0]$ z przedziału $t \in [t_0 + 2, t_0 + m - 1]$. Dlatego ostateczna decyzja ma postać

$$\begin{aligned} \hat{d}(t_0 + 1 - r) &= \dots = \hat{d}(t_0) = 0, \\ \hat{d}(t_0 + 1) &= \dots = \hat{d}(t_0 + m) = 1, \\ \hat{d}(t_0 + m + 1) &= \dots = \hat{d}(t_0 + m + r) = 0 \end{aligned} \quad (3.87)$$

identyczną z (3.8), która gwarantuje, że alarmy detekcyjne tworzą zawsze nieprzerwany blok „jedynek” poprzedzonych i zakończonych przez co najmniej r „zer”.

Zwróćmy uwagę jeszcze na to, że w przypadku detekcji z decyzyjnym sprzężeniem zwrotnym procedura ta jest sprzężona z procedurą interpolacji sygnału (co wynika z porównania schematów na rysunkach 3.1 i 3.4), ponieważ w chwili przzerwania pracy algorytmu dysponujemy już oceną filtracyjną wektora stanu $\hat{\mathbf{x}}_{q(t)}(t_0 + m + r | t_0 + m + r)$. Będzie ona identyczna z oceną (3.17) pod warunkiem, że na etapie detekcji nie wystąpiły „przypadkowe akceptacje” próbek zlokalizowanych w środku przedziału $[t_0 + 2, t_0 + m - 1]$. W pozostałych przypadkach zalecane jest ponowne uruchomienie procedury interpolacji z uwzględnieniem ostatecznej postaci alarmu detekcyjnego (3.87).

3.5 Metody oparte na analizie błędów resztowych

W odróżnieniu od sekwencyjnej analizy błędów predykcji metody oparte na analizie błędów resztowych wykorzystują przetwarzanie blokowe. Sygnał $y(t)$ dzielony jest na równe bloki (zazwyczaj częściowo nachodzące na siebie), z których każdy jest analizowany oddzielnie.

Rozważmy blok składający się z N próbek $Y(N) = \{y(1), \dots, y(N)\}$, dla którego sekwencja błędów resztowych $\eta(t)$ obliczana jest w następujący sposób

$$\eta(t) = y(t) - \sum_{i=1}^r \hat{a}_i(N) y(t - i), \quad t = 1, \dots, N, \quad (3.88)$$

gdzie $\{\hat{a}_1(N), \dots, \hat{a}_r(N)\}$ oznaczają oceny współczynników autoregresji w chwili N , otrzymywane na drodze identyfikacji modelu sygnału (3.2) po uwzględnieniu całej historii pomiarów w danym bloku $Y(N)$. Warunki początkowe $\{y(1 - r), \dots, y(0)\}$ pochodzą z poprzedniego

bloku danych; wyjątek stanowi pierwszy blok, dla którego warunki początkowe przyjmują wartości zerowe.

Obecność zakłóceń impulsowych wykrywana jest na drodze progowania otrzymanego ciągu błędów resztowych $\eta(t)$. Poniżej omówimy jedną z ciekawszych procedur detekcji wykorzystującą dwa progi, opisaną w pracy [50].

Detekcja wykorzystująca dwa progi

Omawiana procedura jest dwuetapowa. Celem pierwszego etapu jest znalezienie „dużych” wartości błędów resztowych $\eta(t)$, tzn. wartości znacząco odstających od pozostałych (w większości przypadków duże błędy mogą być przypisywane obecności zakłóceń impulsowych) przy użyciu następującej reguły decyzyjnej

$$\hat{d}_0(t) = \begin{cases} 1, & \text{jeżeli } |\eta(t)| > \mu_0 \hat{\sigma}(N), \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad t = 1, \dots, N, \quad (3.89)$$

gdzie μ_0 oznacza parametr projektowy (mnożnik) ustawiany celowo na względnie dużą wartość $\mu_0 = 5$, a $\hat{\sigma}(N)$ jest lokalną oceną średniego odchylenia standardowego wyznaczoną w sposób odporny na obecność zakłóceń impulsowych

$$\begin{aligned} \hat{\sigma}(N) &= \sqrt{\hat{\rho}(N)}, \\ \hat{\rho}(N) &= 1,4 \operatorname{med}\{\eta^2(t), t = 1, \dots, N\}. \end{aligned} \quad (3.90)$$

Funkcja $\operatorname{med}(\cdot)$ oznacza medianę – środkowy element uporządkowanego ciągu liczb.

Celem drugiego etapu jest dokładniejsza lokalizacja początku i końca każdego alarmu detekcyjnego. Załóżmy, że sygnał $\hat{d}_0(t)$ może zostać zapisany jako sekwencja n odseparowanych od siebie bloków detekcyjnych, tzn.

$$\hat{d}_0(t) = \begin{cases} 1, & \text{dla } t \in \bigcup_{i=1}^n D_0(i), \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad (3.91)$$

gdzie

$$D_0(i) = [\underline{t}_0(i), \bar{t}_0(i)], \quad \underline{t}_0(i) \leq \bar{t}_0(i), \quad i = 1, \dots, n. \quad (3.92)$$

Każdy wstępnie zgłoszony alarm detekcyjny $D_0(i)$ jest sprawdzany pod kątem potencjalnego rozszerzenia. Rozszerzony (wydłużony z obu stron) alarm detekcyjny ma postać

$$D^*(i) = [\underline{t}^*(i), \bar{t}^*(i)], \quad \underline{t}^*(i) \leq \underline{t}_0(i), \quad \bar{t}^*(i) \geq \bar{t}_0(i), \quad i = 1, \dots, n, \quad (3.93)$$

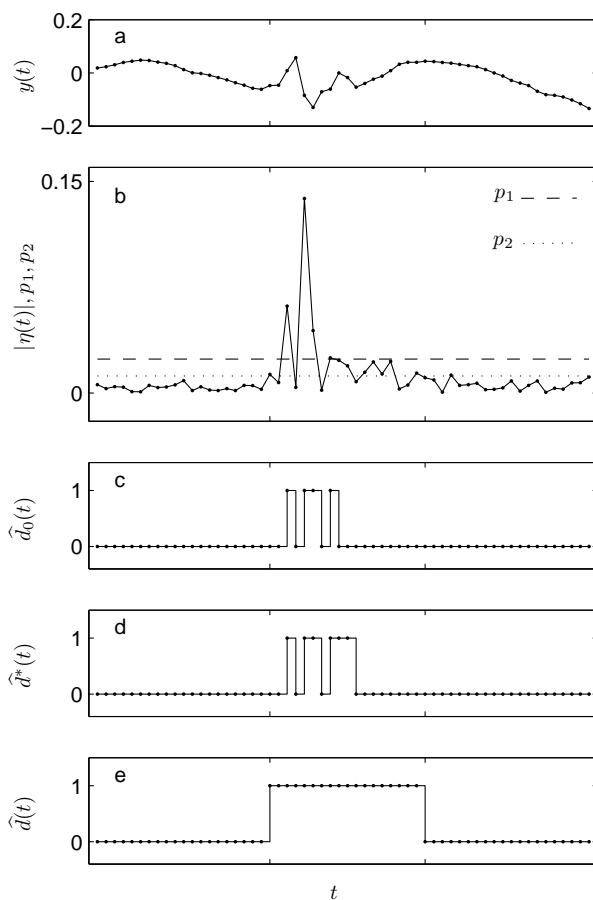
gdzie $\underline{t}^*(i)/\bar{t}^*(i)$ oznaczają minimalną/maksymalną wartość taką, że

$$|\eta(t)| > \mu \hat{\sigma}(N) \quad \forall t \in D^*(i), \quad (3.94)$$

gdzie $\mu = \mu_0/2$ jest kolejnym mnożnikiem, tym razem przyjmującym względnie małą wartość. Zastosowanie drugiego progu detekcyjnego pozwala na uzyskanie nieco lepszej lokalizacji początku i końca każdego alarmu

$$\hat{d}^*(t) = \begin{cases} 1, & \text{dla } t \in \bigcup_{i=1}^n D^*(i), \\ 0, & \text{w pozostałych przypadkach.} \end{cases} \quad (3.95)$$





Rysunek 3.6: Typowe zachowanie algorytmu detekcji wykorzystującego dwa progi w obecności rzeczywistego zakłócenia impulsowego. Kolejne wykresy przedstawiają: sygnał foniczny (a), wartości bezwzględne błędów resztowych $|\eta(t)|$ (b – linia ciągła z naniesionymi punktami), pierwszy próg detekcyjny p_1 (b – linia przerywana), drugi próg detekcyjny p_2 (b – ciąg punktów), kolejne etapy detekcji zakłóceń $\hat{d}_0(t)$, $\hat{d}^*(t)$, $\hat{d}(t)$ odpowiednio wykresy (c,d,e). Progi detekcyjne $p_1 = \mu_0 \hat{\sigma}(N)$ i $p_2 = \frac{1}{2} p_1$ zależą od mnożnika μ_0 oraz od oceny średniego odchylenia standardowego $\hat{\sigma}(N)$.

Ostatni stosowany zabieg poszerza alarmy detekcyjne $D^*(i)$ do momentu, aż sekwencja błędów resztowych o długości l_0 , np. $l_0 = 3$, z obu stron końcowego alarmu detekcyjnego nie będzie przekraczała drugiego progu. Końcowy alarm detekcyjny ma postać

$$D(i) = [\underline{t}(i), \bar{t}(i)], \quad \underline{t}(i) \leq \underline{t}^*(i), \quad \bar{t}(i) \geq \bar{t}^*(i), \quad i = 1, \dots, n, \quad (3.96)$$

gdzie $\underline{t}(i)/\bar{t}(i)$ oznaczają minimalną/maksymalną wartość taką, że

$$\begin{aligned} |\eta(\underline{t}(i) - j)| &\leq \mu \hat{\sigma}(N), \quad j = 1, \dots, l_0, \\ |\eta(\bar{t}(i) + j)| &\leq \mu \hat{\sigma}(N). \end{aligned} \quad (3.97)$$

W wyniku zastosowania zabiegów poszerzania otrzymujemy

$$\hat{d}(t) = \begin{cases} 1, & \text{dla } t \in \bigcup_{i=1}^n D(i), \\ 0, & \text{w pozostałych przypadkach.} \end{cases} \quad (3.98)$$

Rysunek 3.6 ilustruje procedurę detekcji zakłócenia wykorzystującą dwa progi.

Po zakończeniu detekcji, zakwestionowane próbki zastępowane są ich interpolacjami uzyskanymi na podstawie dostępnego modelu sygnału – patrz podrozdział 3.2.

3.6 Porównanie algorytmów detekcji

W podrozdziale tym porównamy zachowanie trzech algorytmów detekcji: algorytmu pracującego w pętli otwartej, algorytmu pracującego z decyzyjnym sprzężeniem zwrotnym oraz algorytmu wykorzystującego dwa progi.

Na rysunku 3.7 przedstawiono porównanie wyników pracy wymienionych algorytmów detekcji. Wspólnym problemem dla wszystkich trzech algorytmów jest brak wystarczającej precyzji w ustalaniu położenia zakłócenia.

Algorytm detekcji pracujący w pętli otwartej wykazuje tendencję do podnoszenia zbyt krótkich alarmów detekcyjnych, tj. alarm detekcyjny często kończy się zanim przebieg zakłócenia impulsowego zostanie wykryty w całości. Wyniki ulegają poprawie, gdy użyjemy algorytmu detekcji pracującego z decyzyjnym sprzężeniem zwrotnym. Alarmy podniesione przez ten algorytm są zazwyczaj dłuższe niż te otrzymane w wyniku zastosowania algorytmu detekcji działającego w pętli otwartej.

Algorytm detekcji wykorzystujący dwa progi wykazuje tendencję do podnoszenia znacznie dłuższych alarmów detekcyjnych niż te otrzymane za pomocą algorytmów wykorzystujących predykcję sygnału. Efekt ten daje się łatwo wytłumaczyć. Załóżmy, że zakłócenie zaczyna się w chwili $t_0 + 1$ i kończy w chwili $t_0 + m$. Zauważmy, że nawet gdy próbka $y(t_0 + m + 1)$ jest niezakłócona, to odpowiadająca jej wartość błędu resztowego $\eta(t_0 + m + 1)$ zazwyczaj pozostaje nadal duża. Dzieje się tak, ponieważ do jej obliczenia wykorzystuje się r ostatnich próbek sygnału. W przypadku gdy przyjęty rząd modelu AR jest duży (dla częstotliwości próbkowania 44,1 kHz i 48 kHz zaleca się wybór $r \geq 10$), długość generowanych alarmów detekcyjnych jest zazwyczaj o wiele większa niż rzeczywista długość wykrytego zakłócenia.

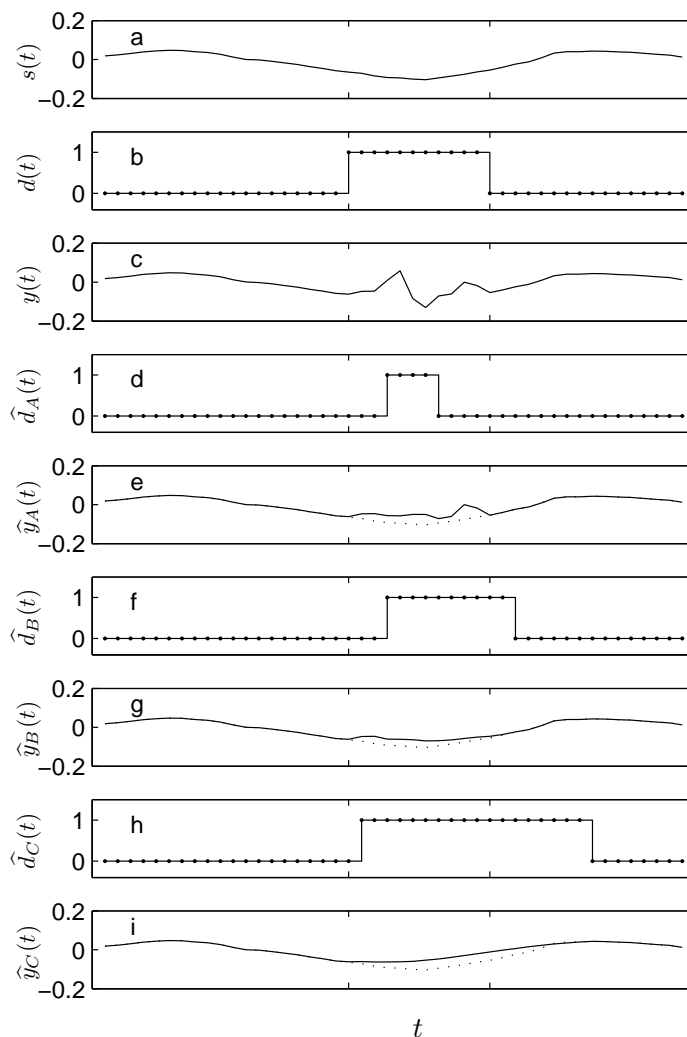
Wpływ mnożnika μ

Parametr μ ma bezpośredni wpływ na czułość detektora. Jeżeli jego wartość jest zbyt duża, detektor będzie miał tendencję do „przepuszczania” niewielkich zakłóceń impulsowych oraz do zbyt późnego wykrywania obecności większych zakłóceń impulsowych. Z kolei gdy jego wartość jest zbyt mała, detektor staje się „nadwrażliwy” i generuje wiele fałszywych alarmów detekcyjnych, choć zarazem szybko reaguje na obecność zakłóceń impulsowych.

Ponieważ zbyt duża liczba fałszywych alarmów (szczególnie tych długich) może powodować słyszalne zniekształcenia rekonstruowanego sygnału, wynikające z ograniczonej dokładności interpolacji, wybór mnożnika μ jest ważnym elementem procesu decyzyjnego dokonywanym na drodze kompromisu.

W przypadku algorytmów detekcji wykorzystujących sekwencyjną predykcję sygnału, najczęściej wybiera się wartość mnożnika z przedziału $\mu \in [3, 5, 4, 5]$, tak aby alarm detekcyjny podnoszony był w pobliżu początku zakłócenia impulsowego. Z kolei w przypadku metod opartych na analizie błędów resztowych wybiera się wartość mnożnika $\mu > 4, 5$, ponieważ pierwszym





Rysunek 3.7: Porównanie wyników pracy trzech algorytmów detekcji w obecności rzeczywistego zakłócenia impulsowego. Kolejne wykresy przedstawiają: czysty sygnał foniczny (a), położenie zakłócenia impulsowego (b), zakłócony sygnał foniczny (c), alarm detekcyjny zgłoszony przez algorytm detekcji pracujący w pętli otwartej (d) i odpowiadającą mu rekonstrukcję sygnału (e), alarm detekcyjny zgłoszony przez algorytm detekcji pracujący z decyzyjnym sprzężeniem zwrotnym (f) i odpowiadającą mu rekonstrukcję sygnału (g), alarm detekcyjny zgłoszony przez algorytm detekcji wykorzystujący dwa progi (h) i odpowiadającą mu rekonstrukcję sygnału (i). Na wykresach (e,g,i) ciąg punktów oznacza trajektorię czystego sygnału.

krokiem wieloetapowej analizy jest zlokalizowanie głównej aktywności zakłócenia impulsowego. Dopiero w kolejnych etapach próbuje się ustalić początek i koniec zakłócenia.

Ograniczenia algorytmów detekcji

Omówione metody detekcji posiadają kilka istotnych ograniczeń:

- uwzględniają jedynie wyniki pochodzące z jednokierunkowej analizy sygnału fonicznego,
- radzą sobie słabo w przypadku analizy sygnałów mowy,
- nie uwzględniają pełnej informacji zawartej w nagraniach stereofonicznych gdyż analiza każdego kanału fonicznego prowadzona jest niezależnie, tak jak w przypadku nagrania monofonicznego,
- nie wykorzystują żadnej wiedzy *a priori* o zakłóceniach impulsowych np. o ich typowych kształtach.

Wymienione ograniczenia w znaczący sposób wpływają na efektywność eliminacji zakłóceń impulsowych z archiwalnych nagrań fonicznych.

3.7 Podsumowanie

Na początku rozdziału założyliśmy, że problem usuwania zakłóceń impulsowych może być podzielony na dwa podproblemy: lokalizację zakłóceń impulsowych oraz rekonstrukcję sklasyfikowanych jako uszkodzone próbek sygnału. Dotychczas opracowane metody eliminacji zakłóceń impulsowych popełniają jeszcze zbyt wiele błędów na etapie detekcji i na etapie rekonstrukcji, dając w rezultacie niezadowalające wyniki (co zostanie obiektywnie pokazane w dalszej części niniejszej pracy). Do błędów detekcji zaliczamy:

- niewykryte zakłócenia,
- częściowo wykryte zakłócenia,
- fałszywe alarmy detekcyjne.

Natomiast błędy rekonstrukcji wynikają z:

- interpolacji bloku brakujących próbek na podstawie sąsiadujących z nim, uszkodzonych próbek sygnału (sklasyfikowanych jako „dobre”). Tu winę ponoszą błędy popełnione na etapie detekcji (np. częściowo wykryte zakłócenia). W wyniku takiej interpolacji otrzymamy „hybrydę” będącą skutkiem połączenia dobrych i uszkodzonych próbek na podstawie dostępnego modelu sygnału – patrz rysunek 3.7, wykres (e).
- interpolacji bloku brakujących próbek, związanego z silnym pobudzeniem tonalnym o okresowym charakterze (typowe dla sygnałów mowy), w oparciu o krótkozakresowy model sygnału.

Wspomniane problemy, zarówno na etapie detekcji zakłóceń jak rekonstrukcji sygnału, uzasadniają potrzebę dalszych prac nad rozwojem (udoskonalaniem) metod eliminacji zakłóceń impulsowych z archiwalnych nagrań fonicznych, wykorzystujących technikę modelowania autoregresyjnego. W kolejnych rozdziałach niniejszej pracy zaproponowane zostaną nowe podejścia, które znacząco redukują liczbę popełnianych błędów zarówno na etapie detekcji jak i rekonstrukcji, podnosząc tym samym jakość rekonstruowanych nagrań fonicznych.



Rozdział 4

Analiza dwukierunkowa

Wprowadzenie

W przypadku analizy sygnałów fonicznych w czasie rzeczywistym wielkim wyzwaniem dla detektora, opartego na analizie błędów predykcji, jest odróżnienie naturalnych, nowych dźwięków sygnału od zakłóceń impulsowych. Większość znanych procedur detekcji i interpolacji opartych jest na hipotezie lokalnej stacjonarności analizowanych danych. Metody te działają zadowalająco, gdy charakterystyki sygnału zmieniają się wolno w czasie, ale mogą zawieść w obecności ich nagłych zmian.

Jedną z kluczowych obserwacji jest to, że zdarzenia niedające się przewidzieć na podstawie lewostronnej analizy sygnału, tj. w kierunku od jego pierwszej do ostatniej próbki, są na ogół przewidywalne w czasie odwrotnym. Zastosowanie analizy dwukierunkowej pozwala na lepsze rozróżnienie między naturalnymi dźwiękami, a zakłóceniami impulsowymi. Zgodnie z posiadaną wiedzą idea dwukierunkowego przetwarzania sygnału fonicznego oparta na analizie błędów predykcji wykorzystana była tylko raz, w pracy Canazzy *et al.* [59]. Zaproponowana tam metoda oparta jest na łączeniu wyników rekonstrukcji, otrzymanych niezależnie dla analizy lewo- i prawostronnej.

Proponowane w niniejszej pracy podejście jest inne. Zakłada ono opracowanie lokalnych, zależnych od danej sytuacji, reguł łączenia wyników detekcji otrzymywanych na drodze dwukierunkowej analizy sygnału fonicznego. Umożliwia to dokładniejszą lokalizację zakłóceń impulsowych oraz wpływa na ogólną poprawę statystyk (lepsze dopasowanie alarmów detekcyjnych, mniejsza liczba przeoczonych zakłóceń impulsowych), a w ostatecznym rezultacie podnosi jakość zrekonstruowanych sygnałów fonicznych.

4.1 Analiza sygnału w czasie odwrotnym

Dwukierunkowa analiza zakłada, że od samego początku dysponujemy pełną historią analizowanego sygnału fonicznego $\mathcal{Y} = \{y(1), \dots, y(N)\}$, tj. że sygnał przetwarzany jest w tzw. trybie *offline*. Podejście to zakłada również niezależne przetwarzanie sygnału przy użyciu tego samego algorytmu w kierunku od „najmłodszej” próbki $y(1)$ do próbki „najstarszej” $y(N)$, co będzie dalej nazywane analizą lewostronną, oraz w kierunku od próbki „najstarszej” $y(N)$ do próbki „najmłodszej” $y(1)$, co będzie dalej nazywane analizą prawostronną lub analizą w czasie odwrotnym.



Wprowadźmy następujące oznaczenia dla sygnałów obserwowanych w czasie odwrotnym

$$\begin{aligned}\tilde{y}(t) &= y(N - t + 1), \\ \tilde{s}(t) &= s(N - t + 1), \\ \tilde{\delta}(t) &= \delta(N - t + 1), \quad t = 1, \dots, N.\end{aligned}\tag{4.1}$$

Stąd, otrzymujemy analogiczny do (3.1) opis typu wejście-wyjście

$$\tilde{y}(t) = \tilde{s}(t) + \tilde{\delta}(t).\tag{4.2}$$

Zauważmy dalej, że dowolny stacjonarny sygnał AR opisany modelem (3.2) ma także następującą reprezentację w czasie odwrotnym

$$\tilde{s}(t) = \sum_{i=1}^r a_i \tilde{s}(t - i) + \tilde{n}(t),\tag{4.3}$$

gdzie $\{\tilde{n}(t)\}$ oznacza biały szum tworzący, który jest różny od $\{n(t)\}$, ale ma tę samą wariancję

$$\text{var}[\tilde{n}(t)] = \text{var}[n(t)] = \rho.\tag{4.4}$$

Dowód jest bezpośredni – ponieważ funkcja autokorelacji procesu stacjonarnego jest symetryczna, otrzymujemy

$$\begin{aligned}r_{\tilde{s}}(\tau) &= \text{E}[\tilde{s}(t)\tilde{s}(t - \tau)] = \text{E}[s(N - t + 1)s(N - t + \tau + 1)] = \\ &= r_s(-\tau) = r_s(\tau), \quad \forall t,\end{aligned}\tag{4.5}$$

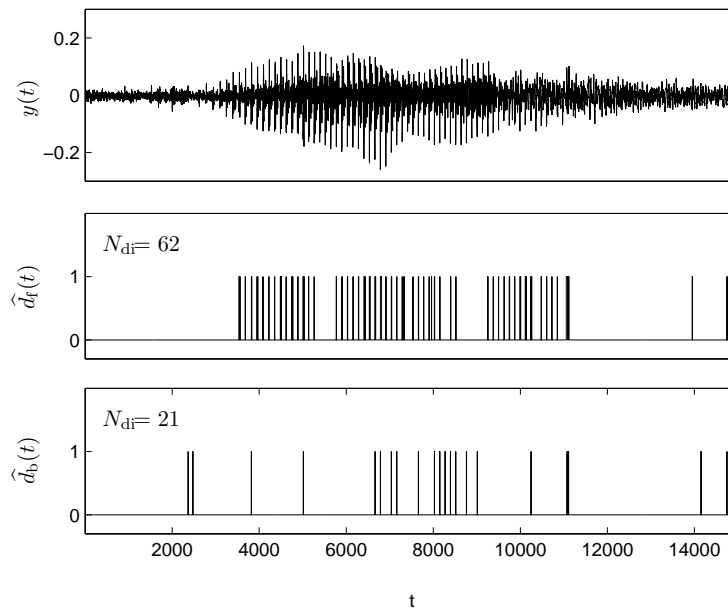
gdzie τ oznacza przesunięcie w czasie. Wynika z tego, że sygnały $s(t)$ i $\tilde{s}(t)$ mają tę samą funkcję autokorelacji, a więc opisane są tymi samymi równaniami Yule'a-Walkera. Zauważmy, że model (4.3) może zostać równoważnie zapisany w postaci

$$s(t) = \sum_{i=1}^r a_i s(t + i) + \tilde{n}(t),\tag{4.6}$$

która wiąże bieżącą wartość sygnału z jego „przyszłymi” wartościami.

Niech wynikiem analizy lewostronnej będą: lewostronny sygnał detekcyjny $\hat{d}_f(t)$, lewostronne oszacowania wektora parametrów $\hat{\theta}_f(t)$ oraz lewostronna estymata wariancji innowacji $\hat{\rho}_f(t)$. Aby zagwarantować zgodność z wynikami analizy lewostronnej, wszystkie analogiczne sygnały otrzymane w wyniku analizy prawostronnej zostają ponownie odwrócone i będą dalej nazywane: prawostronnym sygnałem detekcyjnym $\hat{d}_b(t)$, prawostronnym oszacowaniem wektora parametrów $\hat{\theta}_b(t)$ oraz prawostronną estymatą wariancji innowacji $\hat{\rho}_b(t)$.

Testy odsłuchowe potwierdzają, że wyniki prawostronnej analizy (zarówno wyniki detekcji jak i wyniki interpolacji) są lepsze niż te otrzymane przez analizę lewostronną. Najbardziej prawdopodobnym wytłumaczeniem tego faktu jest to, że naturalne dźwięki mają asymetryczne cechy, mianowicie ich czasy narastania są zazwyczaj o wiele krótsze niż czasy opadania. Wynika z tego, że w przypadku analizy sygnałów o zmiennych w czasie charakterystykach, predyktor oparty na analizie prawostronnej ma łatwiejsze zadanie niż jego odpowiednik oparty na analizie lewostronnej. Na rysunku 4.1 przedstawione zostały typowe wyniki detekcji otrzymane dla czystego sygnału fonicznego o zmiennych w czasie charakterystykach. Zwróćmy uwagę na liczbę fałszywych alarmów podniesionych przez detektor oparty na modelu AR, która wynosi odpowiednio $N_{\text{di}} = 62$ dla analizy lewostronnej oraz $N_{\text{di}} = 21$ dla analizy prawostronnej. Ponadto, fałszywe



Rysunek 4.1: Typowe wyniki detekcji otrzymane dla czystego sygnału fonicznego o zmiennych w czasie charakterystykach, próbkowanego z częstotliwością 22,050 kHz. Wykresy przedstawiają (od góry): czysty sygnał foniczny, wyniki adaptacyjnej detekcji oparte na analizie lewostronnej oraz wyniki adaptacyjnej detekcji oparte na analizie prawostronnej. Na wykresach pokazujących sygnały detekcyjne zamieszczona została informacja o liczbie fałszywych alarmów (N_{di}).

alarmy podnoszone przez detektor oparty na analizie prawostronnej są z reguły o wiele krótsze niż w przypadku analizy lewostronnej – dla rozważanego przykładu średnia długość alarmu wynosi odpowiednio 7 próbek dla analizy lewostronnej oraz 3 próbki dla analizy prawostronnej.

Dokładniejsza analiza pokazuje, że najlepszą efektywność możemy otrzymać, jeżeli wyniki detekcji/interpolacji pochodzące z analizy lewostronnej/prawostronnej zostają odpowiednio połączone. Odpowiednie reguły łączenia zostaną zaprezentowane w dalszej części niniejszego rozdziału.

4.2 Jedno- i dwukierunkowa rekonstrukcja sygnału

Kontynuujemy rozważania nad rekonstrukcją sygnału opisane w punkcie 3.2.1.

Ponieważ $s(t) = y(t)$ dla $t \in [t_1 - r, t_1 - 1] \cup [t_2 + 1, t_2 + r]$, równanie interpolacji (3.10) zapisać możemy w symbolicznej postaci

$$\{\hat{s}(t), t \in \mathcal{T}_0\} = h[\varphi_f(t_1), \varphi_b(t_2), \theta], \quad (4.7)$$

gdzie $\varphi_f(t_1) = [y(t_1 - 1), \dots, y(t_1 - r)]^T$ i $\varphi_b(t_2) = [y(t_2 + 1), \dots, y(t_2 + r)]^T$ oznaczają odpowiednio lewostronne i prawostronne wektory regresji. Zauważmy, że pierwsza część wektora ψ_o w równaniu (3.10) pokrywa się z wektorem $\varphi_f(t_1)$, a jego druga część jest złożona z elementów wektora $\varphi_b(t_2)$.

Gdy parametry modelu są nieznanne, mogą zostać zastąpione ich oszacowaniami – patrz podrozdział 3.3. Rozważmy trzy podejścia do adaptacyjnej procedury interpolacji:

a) *Rekonstrukcja lewostronna*

Interpolacja oparta jest na analizie lewostronnej, tj. wektor θ zastąpiony zostaje lewostronną estymatą $\hat{\theta}_f(t_1 - 1)$:

$$\{\hat{s}_f(t), t \in \mathcal{T}_0\} = h \left[\varphi_f(t_1), \varphi_b(t_2), \hat{\theta}_f(t_1 - 1) \right]. \quad (4.8)$$

b) *Rekonstrukcja prawostronna*

Interpolacja wygląda podobnie jak w poprzednim przypadku, z tą różnicą, że bierzemy pod uwagę prawostronne oceny parametrów $\hat{\theta}_b(t_2 + 1)$:

$$\{\hat{s}_b(t), t \in \mathcal{T}_0\} = h \left[\varphi_f(t_1), \varphi_b(t_2), \hat{\theta}_b(t_2 + 1) \right]. \quad (4.9)$$

c) *Rekonstrukcja dwukierunkowa*

Śledząc tok rozumowania przedstawiony w pracy [59], interpolacja dwukierunkowa może być otrzymana jako kombinacja wypukła rekonstrukcji lewostronnej i prawostronnej:

$$\hat{s}_{fb}(t) = w_f \hat{s}_f(t) + w_b \hat{s}_b(t), \quad t \in \mathcal{T}_0, \quad (4.10)$$

gdzie

$$w_f = \frac{\hat{\rho}_b(t_2 + 1)}{\hat{\rho}_f(t_1 - 1) + \hat{\rho}_b(t_2 + 1)},$$

$$w_b = \frac{\hat{\rho}_f(t_1 - 1)}{\hat{\rho}_f(t_1 - 1) + \hat{\rho}_b(t_2 + 1)}$$

są wagami, które zależą od lokalnej zdolności predyktora opartego na modelu AR działającego w „przód” oraz „w tył”. Zauważmy, że $w_f + w_b = 1$.

4.3 Dwukierunkowa detekcja zakłóceń impulsowych

W podrozdziale tym przedstawione zostaną reguły „łączenia” sygnałów detekcyjnych $\hat{d}_f(t)$ i $\hat{d}_b(t)$. Oba sygnały mogą być traktowane jako sekwencje alarmów detekcyjnych, oznaczonych dalej symbolami $D_f(i)$ i $D_b(i)$:

$$\hat{d}_f(t) = \begin{cases} 1, & \text{jeżeli } t \in \cup_{i=1}^{n_f} D_f(i), \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad (4.11)$$

$$\hat{d}_b(t) = \begin{cases} 1, & \text{jeżeli } t \in \cup_{i=1}^{n_b} D_b(i), \\ 0, & \text{w pozostałych przypadkach,} \end{cases}$$

gdzie

$$D_f(i) = [t_f(i), \bar{t}_f(i)], \quad D_b(i) = [t_b(i), \bar{t}_b(i)].$$

Wielkości $t_f(i)/t_b(i)$ oraz $\bar{t}_f(i)/\bar{t}_b(i)$, spełniające warunki

$$t_f(i) \leq \bar{t}_f(i), \quad t_b(i) \leq \bar{t}_b(i),$$

oznaczają odpowiednio początek i koniec i -tego lewostronnego/prawostronnego alarmu detekcyjnego.

Ze względu na użyty schemat detekcji (patrz uwaga 3.5), sąsiadujące alarmy detekcyjne spełniają warunek separowalności

$$\underline{t}_f(i+1) - \overline{t}_f(i) > r, \quad \underline{t}_b(i+1) - \overline{t}_b(i) > r, \quad (4.12)$$

tj. oddzielone są od siebie przez co najmniej r próbek sygnału sklasyfikowanych jako niezależne. Jest to minimalna odległość między alarmami detekcyjnymi pozwalająca na zdekomponowanie problemu interpolacji n_f/n_b bloków brakujących próbek na n_f/n_b lokalnych zadań interpolacji, analizowanych w poprzednim podrozdziale. Zauważmy, że analogiczna separacja nie jest zagwarantowana, gdy analizujemy łącznie lewo- i prawostronne alarmy detekcyjne, co oznacza, że mogą one niekiedy tworzyć złożone konfiguracje. Dlatego ustalenie dwukierunkowego sygnału detekcyjnego $\hat{d}_{fb}(t)$, opartego na wynikach analizy lewo- i prawostronnej, nie jest zadaniem prostym.

Gdy opracowany zostanie sygnał detekcyjny $\hat{d}_{fb}(t)$, przeprowadzona zostanie ponownie lewostronna/prawostronna interpolacja sygnału z wykorzystaniem nowych oszacowań położenia zakłóceń impulsowych. Aby uzyskać jeszcze lepszą jakość zrekonstruowanego sygnału, wyniki lewo- i prawostronnej interpolacji połączone zostaną zgodnie z regułą (4.10).

4.3.1 Technika poszerzania alarmów detekcyjnych

Typowa geometria lokalnych uszkodzeń powierzchni płyty gramofonowej (np. uszkodzeń rowka) powoduje powstawanie zakłóceń impulsowych, które zazwyczaj zaczynają się i kończą w dość łagodny sposób. Efekt ten jest tym widoczniejszy im wyższa jest częstotliwość próbkowania. Ponieważ metody oparte na analizie błędów predykcji rzadko podnoszą alarmy detekcyjne na samym początku zakłóceń o łagodnych brzegach, skutkuje to małymi, ale słyszalnymi zniekształceniami rekonstruowanego sygnału. Efekt ten może być złagodzony przez obniżenie wartości mnożnika μ , wpływając tym samym na zwiększenie wrażliwości detektora na „nieprzewidywalne” zmiany zachodzące w sygnale – niestety rozwiązanie to powoduje znaczący wzrost liczby fałszywych alarmów.

Proponowane jest praktyczne rozwiązanie tego problemu polegające na przesunięciu przedniego zbocza każdego alarmu detekcyjnego o małą, stałą liczbę próbek Δ_1 na etapie wstępnego przetwarzania – zanim jeszcze zostanie ustalony sygnał detekcyjny $\hat{d}_{fb}(t)$. Zazwyczaj $\Delta_1 \in [2, 4]$ jest rozsądnym wyborem dla nagrań próbkowanych z częstotliwością: 22,05 kHz, 44,1 kHz lub 48 kHz.

Zwróćmy uwagę, że po odwróceniu sygnału detekcyjnego pochodzącego z analizy w czasie odwrotnym, przednie zbocze alarmu detekcyjnego staje się jego tylnym zboczem. Zmodyfikowane lewostronne/prawostronne alarmy detekcyjne przyjmują postać

$$D_f^*(i) = [\underline{t}_f^*(i), \overline{t}_f(i)], \quad D_b^*(i) = [\underline{t}_b(i), \overline{t}_b^*(i)], \quad (4.13)$$

gdzie¹

$$\underline{t}_f^*(i) = \underline{t}_f(i) - \Delta_1, \quad \overline{t}_b^*(i) = \overline{t}_b(i) + \Delta_1. \quad (4.14)$$

¹Ważne jest, aby przy poszerzaniu alarmów detekcyjnych nie naruszyć warunku separowalności. Dlatego modyfikacja nie jest wprowadzona jeżeli odległość pomiędzy sąsiednimi alarmami detekcyjnymi jest mniejsza niż $r + \Delta_1$, tj. gdy $\underline{t}_f(i) - \overline{t}_f(i-1) - \Delta_1 < r$ (dla alarmów lewostronnych) i $\underline{t}_b(i+1) - \overline{t}_b(i) - \Delta_1 < r$ (dla alarmów prawostronnych). W takim przypadku stosowane jest krótsze rozszerzenie, mianowicie takie, które nie narusza warunku separowalności.

Zmodyfikowane sygnały detekcyjne odnoszące się do analizy lewo- i prawostronnej oznaczane będą dalej odpowiednio przez $\hat{d}_f^*(t)$ i $\hat{d}_b^*(t)$.

4.3.2 Globalne reguły decyzyjne

Najprostszym podejściem do łączenia wyników detekcji lewo- i prawostronnej jest podejście oparte na zastosowaniu globalnych reguł decyzyjnych nawiązujących do reguł wykonywania operacji na wartościach logicznych, takich jak iloczyn logiczny (\cap)

$$\hat{d}_{fb}(t) = \begin{cases} 1, & \text{jeżeli } \hat{d}_f^*(t) = 1 \cap \hat{d}_b^*(t) = 1, \\ 0, & \text{w pozostałych przypadkach} \end{cases} \quad (4.15)$$

lub suma logiczna (\cup)

$$\hat{d}_{fb}(t) = \begin{cases} 1, & \text{jeżeli } \hat{d}_f^*(t) = 1 \cup \hat{d}_b^*(t) = 1, \\ 0, & \text{w pozostałych przypadkach.} \end{cases} \quad (4.16)$$

W pierwszym przypadku dwukierunkowy alarm detekcyjny $\hat{d}_{fb}(t)$ podniesiony zostanie tylko wtedy gdy ta sama próbka sygnału zakwestionowana zostanie przez oba detektory, z kolei w drugim przypadku – gdy zakwestionowana zostanie przez co najmniej jeden z detektorów. Wstępne testy pokazały, że żadna z tych reguł nie działa zadowalająco w praktyce. Reguła iloczynu logicznego jest zbyt ostrożna – ma tendencję do przeoczenia wielu małych zakłóceń impulsowych i tworzenia niedopasowanych (zbyt krótkich) alarmów detekcyjnych. Reguła sumy logicznej jest z kolei zbyt tolerancyjna – daje wiele przeszacowanych (zbyt długich) alarmów detekcyjnych, które po interpolacji skutkują słyszalną degradacją sygnału.

Dokładniejsza analiza pokazuje, że najlepszą efektywność możemy otrzymać, jeżeli wyniki lewostronnej i prawostronnej detekcji zostają połączone przy użyciu pewnej liczby lokalnych reguł decyzyjnych dopasowanych do poszczególnych sytuacji.

4.3.3 Lokalne reguły decyzyjne

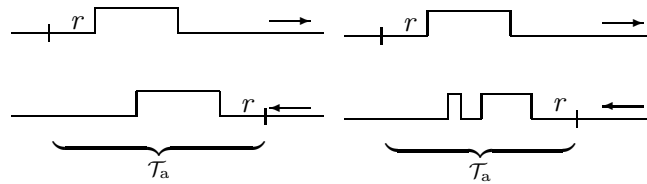
Aby uniknąć problemów typowych dla globalnych reguł decyzyjnych, różne konfiguracje lewostronnych i prawostronnych alarmów detekcyjnych podzielone zostaną na klasy i podklasy. Każda klasa analizowana będzie osobno w celu ustalenia najlepszego sposobu łączenia alarmów detekcyjnych.

Przestrzegając zasadę separowalności alarmów detekcyjnych, alarmy rozpatrywane będą w kolejnych oknach analizy $\mathcal{T}_a(k) = [\underline{t}_a(k), \overline{t}_a(k)]$, $k = 1, \dots, n_{fb}$ zdefiniowanych jako przedziały *minimalnej długości*, które zaczynają się i kończą r decyzjami pozytywnymi (bez alarmu)

$$\begin{aligned} \hat{d}_f^*(t) = \hat{d}_b^*(t) = 0 & \quad \text{dla } t \in [\underline{t}_a(k), \underline{t}_a(k) + r - 1], \\ \hat{d}_f^*(t) = \hat{d}_b^*(t) = 0 & \quad \text{dla } t \in [\overline{t}_a(k) - r + 1, \overline{t}_a(k)] \end{aligned} \quad (4.17)$$

i zawierają co najmniej jeden alarm detekcyjny lewostronny bądź prawostronny:

$$\begin{aligned} \hat{d}_f^*(\underline{t}_a(k) + r) = 1 \quad \text{i/lub} \quad \hat{d}_b^*(\underline{t}_a(k) + r) = 1, \\ \hat{d}_f^*(\overline{t}_a(k) - r) = 1 \quad \text{i/lub} \quad \hat{d}_b^*(\overline{t}_a(k) - r) = 1. \end{aligned} \quad (4.18)$$



Rysunek 4.2: Przykład elementarnej (lewy rysunek) i złożonej (prawy rysunek) konfiguracji alarmów. Wykresy przedstawiają wyniki lewo (\rightarrow) i prawostronnej (\leftarrow) detekcji. Klamry pokazują okna analizy \mathcal{T}_a .

Sytuacje gdy okno analizy zawiera co najwyżej jeden lewostronny alarm detekcyjny oraz co najwyżej jeden prawostronny alarm detekcyjny nazywać będziemy konfiguracjami elementarnymi; pozostałe określane będą mianem konfiguracji złożonych – patrz rysunek 4.2. Zauważmy, że sąsiednie okna analizy mogą częściowo nachodzić na siebie (mogą mieć do r próbek wspólnych na swoim początku i/lub końcu).

Elementarne konfiguracje alarmów mogą zostać podzielone na kilka klas oraz podklas.

Konfiguracje tworzące klasę A

Do klasy A zalicza się konfiguracje tworzone przez pary „nachodzących na siebie” alarmów detekcyjnych – jednego lewostronnego $D_f^*(i)$ oraz jednego prawostronnego $D_b^*(j)$

$$D_f^*(i) \cap D_b^*(j) \neq \emptyset. \quad (4.19)$$

Klasę tę można podzielić na 5 różnych, wzajemnie wykluczających się, podklas przedstawionych na rysunku 4.3:

A_1 : alarm lewostronny pokrywa się z alarmem prawostronnym

$$D_f^*(i) = D_b^*(j), \quad (4.20)$$

A_2 : alarm prawostronny jest podzbiorem alarmu lewostronnego

$$D_b^*(i) \subset D_f^*(j) \text{ i } D_f^*(i) \neq D_b^*(j), \quad (4.21)$$

A_3 : alarm lewostronny jest podzbiorem alarmu prawostronnego

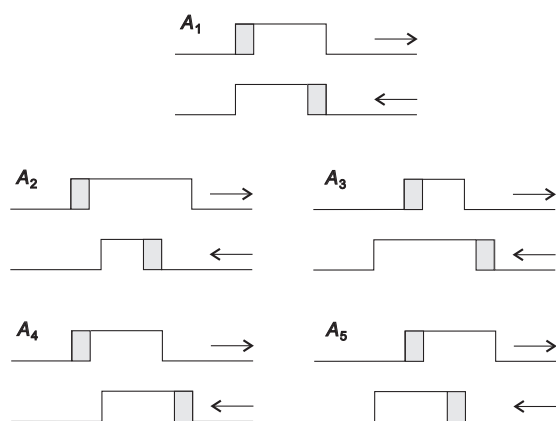
$$D_f^*(i) \subset D_b^*(j) \text{ i } D_f^*(i) \neq D_b^*(j), \quad (4.22)$$

A_4 : alarm lewostronny zaczyna/kończy się zanim alarm prawostronny zacznie/skończy się

$$\underline{t}_f^*(i) < \underline{t}_b(j), \quad \bar{t}_f(i) < \bar{t}_b^*(j), \quad (4.23)$$

A_5 : alarm prawostronny zaczyna/kończy się zanim alarm lewostronny zacznie/skończy się

$$\underline{t}_b(j) < \underline{t}_f^*(i), \quad \bar{t}_b^*(j) < \bar{t}_f(i). \quad (4.24)$$



Rysunek 4.3: Konfiguracje alarmów należące do klasy A . Wykresy przedstawiają wyniki lewostronnej (\rightarrow) i prawostronnej (\leftarrow) detekcji. Zaciemniony obszar oznacza rozszerzenie alarmu dodane na etapie wstępnego przetwarzania.

W każdej z wymienionych wyżej podklas sprawdzone zostały trzy reguły łączenia lewo- i prawostronnych alarmów detekcyjnych – reguła sumy logicznej (\cup):

$$D_{fb}(k) = D_f^*(i) \cup D_b^*(j), \quad (4.25)$$

reguła iloczynu logicznego (\cap):

$$D_{fb}(k) = D_f^*(i) \cap D_b^*(j) \quad (4.26)$$

oraz reguła „przednie zbocze-przednie zbocze” (FF):

$$D_{fb}(k) = [\underline{t}_f^*(i), \overline{t}_b^*(j)]. \quad (4.27)$$

W ostatnim przypadku wynikowy alarm detekcyjny zaczyna się w miejscu przedniego zbocza alarmu lewostronnego oraz kończy się w miejscu tylnego zbocza alarmu prawostronnego (które w czasie odwróconym było jego przednim zboczem). Reguła FF wynika z pewnych obserwacji praktycznych – wiadomo, że moment zgłoszenia alarmu detekcyjnego ustalany jest zazwyczaj dokładniej niż moment jego zakończenia.

Konfiguracje tworzące klasę B

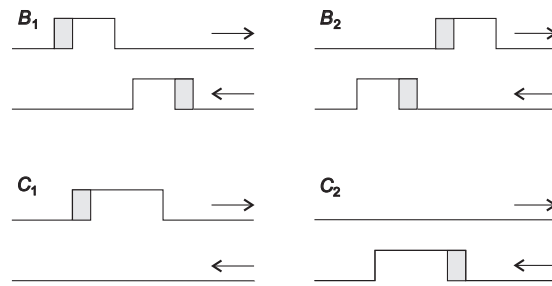
Konfiguracje alarmów należące do klasy B złożone są z par nienachodzących na siebie alarmów detekcyjnych

$$D_f^*(i) \cap D_b^*(j) = \emptyset, \quad (4.28)$$

w przypadkach gdy dzieląca je odległość jest mniejsza niż r próbek. Klasa ta podzielona została na dwie podklasy – patrz rysunek 4.4 – związane z kolejnością pojawienia się alarmów detekcyjnych:

B_1 : alarm lewostronny poprzedza alarm prawostronny

$$\overline{t}_f(i) < \underline{t}_b(j), \quad (4.29)$$



Rysunek 4.4: Wzorce konfiguracji alarmów należące do klas B oraz C. Wykresy przedstawiają wyniki lewo (→) i prawostronnej (←) detekcji. Zaciemniony obszar oznacza rozszerzenie dodane na etapie wstępnego przetwarzania.

B_2 : alarm prawostronny poprzedza alarm lewostronny

$$\overline{t_b^*}(j) < \underline{t_f^*}(i). \quad (4.30)$$

Dla konfiguracji alarmów należących do klasy B sprawdzone zostały dwie reguły: reguła sumy logicznej „z wypełnieniem” (\sqcup)

$$D_{fb}(k) = [\min\{\underline{t_f^*}(i), \underline{t_b}(j)\}, \max\{\overline{t_f}(i), \overline{t_b^*}(j)\}] \quad (4.31)$$

oraz reguła iloczynu logicznego (skutkująca brakiem alarmu)

$$D_{fb}(k) = \emptyset. \quad (4.32)$$

Konfiguracje tworzące klasę C

Konfiguracje alarmów należące do klasy C składają się z pojedynczego alarmu detekcyjnego: albo lewostronnego (C_1) albo prawostronnego (C_2) – patrz rysunek 4.4. W tym przypadku rozważono trzy reguły „łączenia” alarmów: regułę sumy logicznej (niezmieniony alarm)

$$D_{fb}(k) = D_f^*(i), \quad D_{fb}(k) = D_b^*(i), \quad (4.33)$$

regułę iloczynu logicznego (brak alarmu)

$$D_{fb}(k) = \emptyset \quad (4.34)$$

oraz regułę typu „przednie zbocze” (F)

$$D_{fb}(k) = [\underline{t_f^*}(i), \overline{t_f^*}(i)], \quad D_{fb}(k) = [\underline{t_b^*}(j), \overline{t_b^*}(j)], \quad (4.35)$$

gdzie²

$$\overline{t_f^*}(i) = \underline{t_f}(i) + \Delta_1, \quad \underline{t_b^*}(j) = \overline{t_b}(j) - \Delta_1. \quad (4.36)$$

²Jeżeli pozycjonowanie przedniego/tylnego zbocza alarmu zaburza warunek separowalności, wtedy powinno być zastosowane mniejsze przesunięcie.

Dokładniejsza analiza konfiguracji alarmów z klasy C pokazała, że w większości przypadków zakłócenia impulsowe (jeżeli w ogóle są obecne) występują w bliskim otoczeniu przedniego zbocza alarmu detekcyjnego. Zgodnie z regułą F, nowe tylne/przednie zbocze alarmów lewostronnych/prawostronnych zaliczających się do klasy C zostało ustalone w miejscach odległych o Δ_1 próbek, odpowiednio od oryginalnych położenia przedniego/tylnego zbocza tych alarmów. Oznacza to, że pierwotne położenie przedniego/tylnego zbocza alarmu otoczone zostaje przez Δ_1 próbek poprzedzających (dodanych na etapie wstępnego przetwarzania) i Δ_1 występujących bezpośrednio po. Wynika z tego, że dopóki warunek separowalności alarmów nie wymusza ograniczenia w poszerzaniu, długość nowych alarmów detekcyjnych zawsze równa będzie $2\Delta_1 + 1$.

Konfiguracje tworzące klasę D

Klasa D utworzona jest ze wszystkich pozostałych, bardziej złożonych konfiguracji alarmów, tj. tych które łączą więcej niż dwa alarmy detekcyjne lewostronne/prawostronne – przykład takiego wzorca pokazany jest na rysunku 4.2. Dla przypadku złożonych wzorców sprawdzone zostały trzy reguły łączenia alarmów detekcyjnych: reguła sumy logicznej „z wypełnieniem”, reguła iloczynu logicznego oraz reguła „przednie zbocze-przednie zbocze”.

4.3.4 Wybór zoptymalizowanego zestawu lokalnych reguł decyzyjnych

Zestaw lokalnych reguł decyzyjnych, dopasowany do danego schematu detekcji zakłóceń impulsowych, może zostać otrzymany przy użyciu testów odsłuchowych i/lub obiektywnych miar oceny jakości dźwięku, których wyniki korelują się z ocenami subiektywnymi. Testy odsłuchowe są bardzo czasochłonne oraz wymagają zaangażowania ekspertów z dziedziny rekonstrukcji dźwięku, ale pozwalają na osiągnięcie dobrych rezultatów. W przypadku użycia obiektywnych miar należy zachować pewną ostrożność przy próbie interpretacji otrzymywanych wyników.

W rozdziale 8 omówiono dokładnie sposób przygotowania nagrań testowych pozwalający na wybranie najlepszej reguły dla każdej konfiguracji alarmów. Przedstawiono również zestaw reguł wytypowany przez trzech ekspertów i potwierdzony wynikami obiektywnej miary jakości dźwięku. Podejścia wykorzystujące proponowany zestaw reguł uzyskały znacznie lepsze wyniki w testach symulacyjnych, niż te same schematy detekcji, ale oparte na jednej globalnej regule decyzyjnej: iloczynu bądź sumy. Proponowane rozwiązanie dało również znacząco lepsze wyniki niż podejście zaproponowane w pracy Canazzy *et al.* [59].

4.4 Podsumowanie

Proponowana tu technika łączenia wyników detekcji oparta jest na zastosowaniu lokalnych reguł łączenia lewostronnych alarmów detekcyjnych z ich prawostronnymi odpowiednikami, jeszcze przed zastosowaniem interpolacji sygnału. Otrzymane w ten sposób lokalizacje zakłóceń impulsowych są bardziej dokładne i jednocześnie bardziej wiarygodne, wpływając tym samym na poprawę wyników w porównaniu do jednokierunkowej (lewostronnej lub prawostronnej) analizy sygnału.



Rozdział 5

Model dwuzakresowy

Wprowadzenie

Chociaż opisane do tej pory procedury detekcji/interpolacji oparte na modelu AR dają bardzo dobre wyniki w przypadku sygnałów muzycznych, to dość często zawodzą w przypadku archiwalnych sygnałów mowy (takich jak np. historyczne przemówienia), szczególnie we fragmentach mowy dźwięcznej. Ponieważ fragmenty te formowane są na drodze pobudzenia traktu głosowego (reprezentowanego przez model AR) okresowym ciągiem impulsów powietrza [78], detektory oparte na analizie błędów predykcji często mylą silne tony krtaniowe z zakłóceniami impulsowymi generując dużą liczbę fałszywych alarmów. Ponieważ model sygnału uwzględnia tylko krótkozakresową korelację sygnału, otrzymywane interpolacje oparte na tym modelu często powodują słyszalne zniekształcenia dźwięku. Interesujący jest też fakt, że podobny efekt może zostać zaobserwowany dla sygnałów muzycznych z udziałem śpiewu i/lub instrumentów dętych takich jak trąbka, saksofon czy klarnet [79].

Jeden ze sposobów radzenia sobie z tym problemem opisany został w pracy [58]. Zaproponowana tam metoda oparta jest na blokowaniu detektora zakłóceń impulsowych w miejscach, gdzie spodziewana jest aktywność tonu krtaniowego. Oczywistą wadą tego podejścia jest jednak to, że zakłócenia pojawiające się w miejscach blokowania detektora nie są wykrywane.

Naturalnym podejściem do problemu opisanego powyżej jest użycie autoregresyjnych modeli dwuzakresowych, należących do klasy tzw. modeli rzadkich (z *ang.* *sparse autoregressive model*, SAR), [48]. Niestety, stosowane do tej pory metody identyfikacji jednorodnych modeli SAR nie gwarantują ich stabilności, co może rodzić szereg problemów na etapie detekcji zakłóceń i interpolacji sygnału.

W rozdziale tym omówimy nowe podejście, którego wspomniane wyżej problemy nie dotyczą. W podejściu tym sygnał mowy reprezentowany jest przez model SAR identyfikowany jako kaskadowe połączenie filtra formantowego i filtra wysokości dźwięku. Technika ta jest szeroko stosowana w systemach liniowego predykcyjnego kodowania sygnałów mowy, ponieważ gwarantuje stabilność modelu [80], [81]. Przeprowadzone testy potwierdzają bardzo dobre zdolności predykcyjne/interpolacyjne modelu SAR, również w miejscach występowania silnych tonów krtaniowych. Aby zwiększyć odporność schematu detekcji zakłóceń impulsowych i jednocześnie zmniejszyć liczbę fałszywych alarmów, decyzja o podniesieniu alarmu wypracowywana jest wspólnie przez trzy działające równolegle detektory zakłóceń impulsowych.



5.1 Autoregresyjny model dwuzakresowy

Naturalnym podejściem do problemu opisanego we wprowadzeniu jest użycie autoregresyjnych modeli dwuzakresowych [48]. Dwuzakresowy model SAR ma postać

$$s(t) = \sum_{i=1}^r a_i s(t-i) + \sum_{i=\tau_0+1}^{\tau_0+v} a_i s(t-i) + n(t), \quad (5.1)$$

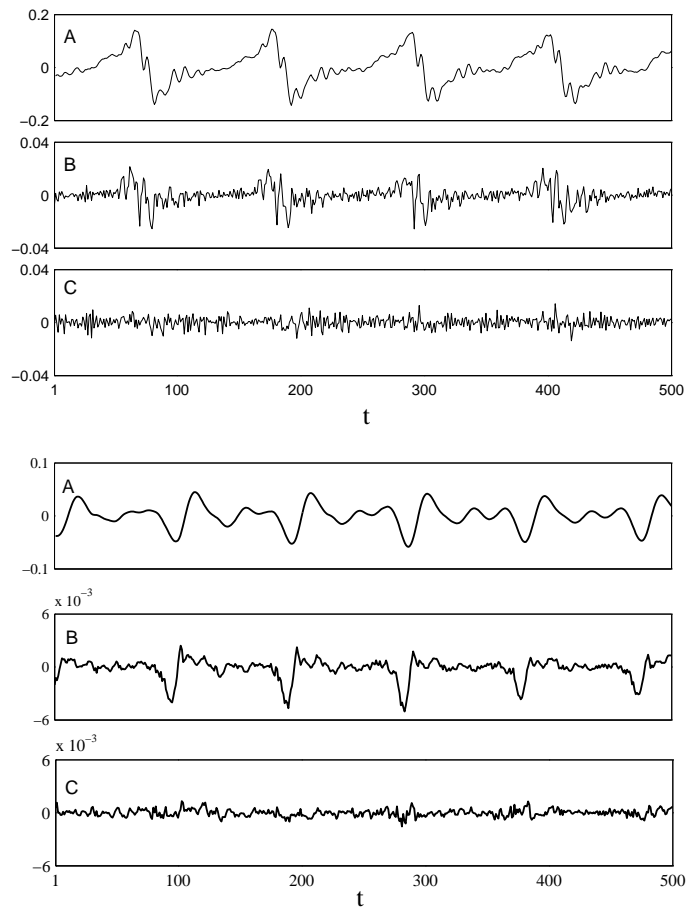
gdzie wielkości τ_0 ($\tau_0 \gg r$) i v wybierane są w taki sposób, aby $\tau_0+1 \leq T \leq \tau_0+v$, gdzie T oznacza okres podstawowy sygnału, tj. w przypadku sygnałów mowy jest to okres tonu krtaniowego (jeżeli jest w ogóle obecny). Choć formalnie rząd modelu równy $p = \tau_0+v$ jest wysoki, model taki może być nazywany rzadkim, ponieważ zawiera tylko $r+v \ll p$ niezerowych współczynników. Rozkład niezerowych współczynników w modelu pozwala na uwzględnienie zarówno korelacji krótkozakresowej [odpowiada za nią pierwszy składnik po prawej stronie równania (5.1)] jak i korelacji długozakresowej [odpowiada za nią drugi składnik po prawej stronie równania (5.1)] analizowanego sygnału fonicznego.

Aby lepiej zrozumieć zalety modelowania dwuzakresowego, rozważmy sygnał opisany modelem (3.2) w przypadku gdy $\{n(t)\}$ nie jest białym szumem lecz ciągiem okresowych impulsów o ustalonym kształcie. Oznaczmy symbolem T okres takiego zewnętrznego pobudzenia. Ponieważ w stanie ustalonym $s(t)$ jest także sygnałem okresowym, o okresie równym T , sygnał ten może zostać opisany następującym modelem rzadkim

$$s(t) = a_T s(t-T), \quad a_T = 1. \quad (5.2)$$

Zauważmy, że taki model charakteryzuje się zerowymi błędami predykcji we wszystkich chwilach czasu t , włącznie z momentami odpowiadającymi okresowej aktywności sygnału pobudzającego. Wyjaśnia to dlaczego modele SAR posiadają dobre zdolności predykcyjne w obecności mieszanego pobudzenia (stochastyczne + okresowe) – oczywiście pod warunkiem, że okres T jest znany. Rysunek 5.1 przedstawia błędy generowane przez adaptacyjne predyktory, oparte na modelu AR i modelu SAR, zastosowane do dwóch fragmentów utworów fonicznych zawierających odpowiednio: mowę dźwięczną i muzykę instrumentalną (trąbka). W obu przypadkach wyniki otrzymane dla modelu AR rzędu 10 ukazują obecność okresowego pobudzenia, które może być łatwo pomyłone z zakłóceniem impulsowym. Gdy ten sam model rozszerzony zostanie o tylko jeden składnik długozakresowy ($v = 1$) i gdy wartość T jest zgodna z okresem podstawowym sygnału, predykcje sygnału stają się bardziej dokładne, a błędy predykcji nie podlegają już okresowym gwałtownym zmianom związanym z pobudzeniem.

Głównym problemem związanym z użyciem modelu (5.1) jest to, że nie istnieją algorytmy identyfikacji, które gwarantowałyby jego stabilność. Problem ze stabilnością modelu pojawia się, gdy prawdziwe parametry modelu a_i zastąpione zostają ich estymatami $\hat{a}_i(t)$, co pociąga za sobą konieczność monitorowania stabilności i jej wymuszania w wymaganych sytuacjach – w przypadku modeli AR sposób rozwiązania tego problemu opisany został w punkcie 3.3.5. Niestety, zaproponowane tam rozwiązanie nie może zostać bezpośrednio zastosowane dla modelu SAR. Ponieważ rząd modelu SAR p jest duży (dla typowych częstotliwości próbkowania zazwyczaj przekracza on wartość 200), testy stabilności, które mogłyby ewentualnie wyeliminować niestabilne modele są obliczeniowo bardzo kosztowne i dlatego nie są stosowane w praktyce.



Rysunek 5.1: Wyniki otrzymane dla dwóch sygnałów fonicznych: mowa dźwięczna (trzy górne wykresy) i muzyka instrumentalna (trąbka; trzy dolne wykresy). Kolejne wykresy pokazują: oryginalny sygnał foniczny (A), błędy predykcji generowane przez identyfikowany na bieżąco model AR (B) i błędy predykcji generowane przez identyfikowany na bieżąco model SAR (C). Liczba współczynników autoregresji w części krótkozakresowej dla obu modeli jest taka sama i równa wartości $r = 10$.

5.2 Model SAR w postaci zdekomponowanej

Wspomniany powyżej problem stabilności może być łatwo rozwiązany jeżeli model SAR zostanie zdekomponowany do następującej kaskadowej postaci, często stosowanej w układach predykcyjnego kodowania sygnałów mowy, takich jak kodery CELP (z ang. *Code Excited Linear Prediction*) [80], [81]

$$s(t) = \sum_{i=1}^r b_i s(t-i) + x(t), \quad (5.3)$$

$$x(t) = cx(t-T) + n(t). \quad (5.4)$$

W kontekście kodowania sygnałów mowy równanie (5.3) opisuje tzw. filtr formantowy charakteryzowany przez współczynniki b_1, \dots, b_r , a równanie (5.4) opisuje filtr wysokości dźwięku,

charakteryzowany przez współczynnik c ¹. Filtr formantowy i filtr wysokości dźwięku tworzą kaskadę. Opis ten jest asymptotycznie równoważny opisowi (5.1) po podstawieniu

$$\mathbb{A}(z^{-1}) = \mathbb{B}(z^{-1})\mathbb{C}(z^{-1}), \quad (5.5)$$

gdzie

$$\mathbb{B}(z^{-1}) = 1 - \sum_{i=1}^r b_i z^{-i}, \quad \mathbb{C}(z^{-1}) = 1 - cz^{-T}.$$

Model w postaci zdekomponowanej (5.3)-(5.4) może zostać łatwo przekształcony do jednorodnej postaci rzadkiej (5.1) w oparciu o zależność

$$\begin{cases} a_i = b_i, & i = 1, \dots, r, \\ a_i = 0, & i = r+1, \dots, T-1, \\ a_T = c, \\ a_{T+i} = -cb_i, & i = 1, \dots, r, \end{cases} \quad (5.6)$$

z której wynika, że $v = r + 1$ oraz $p = T + r$.

Stabilność modelu w postaci zdekomponowanej jest zapewniona jeżeli oba te filtry są stabilne. Warunek ten łatwo jest spełnić w przypadku użycia odpowiednich narzędzi estymacji i prostych mechanizmów wymuszających stabilność. W przypadku modelu formantowego (5.3) stosowana będzie technika opisana w punkcie 3.3.5, z kolei w przypadku modelu wysokości dźwięku (5.4) wystarczający warunek stabilności ma następującą postać

$$|c| < 1, \quad (5.7)$$

co można łatwo sprawdzić przy użyciu procedury Schura-Cohna testowania stabilności. Już po pierwszej iteracji algorytmu stopień wielomianu charakterystycznego $\mathbb{C}(z^{-1})$ redukuje się do wielomianu stopnia zerowego, co kończy procedurę. Zgodnie z [80], warunek stabilności modelu długozakresowego może zostać rozszerzony na większą liczbę współczynników.

Jeżeli warunek (5.7) jest spełniony, nie jest wymagana żadna interwencja. W przeciwnym razie stabilność modelu może być w łatwy sposób wymuszona poprzez zastąpienie współczynnika c jego przeskalowaną wersją

$$(1 - \epsilon) \frac{c}{|c|}, \quad (5.8)$$

gdzie $\epsilon > 0$ oznacza mały margines stabilności, np. $\epsilon = 0,01$.

5.3 Rekonstrukcja oparta na modelu SAR

Rozważmy problem rekonstrukcji odizolowanego bloku m nieodwracalnie uszkodzonych próbek – patrz podrozdział 3.2 – w przypadku gdy interpolacja oparta jest na modelu SAR, w następującej postaci

$$s(t) = \sum_{i=1}^r a_i s(t-i) + \sum_{i=T}^{T+r} a_i s(t-i) + n(t), \quad (5.9)$$

¹Filtr wysokości dźwięku może być charakteryzowany przez więcej niż tylko jeden współczynnik – patrz [80]. Wówczas dodatkowe współczynniki przypisane do próbek sąsiadujących z $x(t-T)$ zabezpieczają model przed ewentualną niedokładną oceną okresu podstawowego T . Jednak nadmierna liczba współczynników powoduje pogorszenie zdolności predykcyjnych modelu w porównaniu do modelu z jednym współczynnikiem.

który powstał w wyniku przekształcenia modelu w postaci zdekomponowanej (5.3)-(5.4) do jednorodnej postaci rzadkiej (5.1) zgodnie z zależnością (5.6). Zauważmy, że model (5.9) możemy zapisać w postaci, która przypomina klasyczny model AR o wysokim rzędzie p

$$s(t) = \sum_{i=1}^p a_i s(t-i) + n(t), \quad (5.10)$$

ale posiada jedynie $2r + 1 \ll p$ niezerowych współczynników autoregresji.

Uwaga 5.1. W celu interpolacji brakujących próbek sygnału będziemy wykorzystywali $p = T + r$ próbek poprzedzających rekonstruowany blok oraz r próbek występujących bezpośrednio po tym bloku – inaczej niż to opisano w pracy [48], gdzie zachowano symetrię. Wprowadzona asymetria wynika bezpośrednio z faktu, że „przyszłe” próbki sygnału nie zostały jeszcze przetworzone i mogą być zakłócone. Ponadto, wartość przesunięcia T wyznaczona na podstawie analizy poprzednich próbek sygnału może już nie być aktualna dla „przyszłych” zmian sygnału, co mogłoby prowadzić do zniekształceń w rekonstruowanym bloku próbek.

W przypadku rekonstrukcji sygnału przy użyciu interpolatora (3.10)

$$\hat{\psi}_m = -(\mathbf{B}_m^T \mathbf{B}_m)^{-1} \mathbf{B}_m^T \mathbf{B}_o \psi_o,$$

należy uwzględnić wspomnianą asymetrię w następujący sposób: macierz \mathbf{B} będzie miała wymiary $(r+m) \times (p+r+m)$, zbiór indeksów $\mathcal{K} = \{1, \dots, p, p+m+1, \dots, p+r+m\}$, zbiór indeksów $\mathcal{U} = \{p+1, \dots, p+m\}$ oraz $\psi_o = [s(t_0+1-p), \dots, s(t_0), s(t_0+m+1), \dots, s(t_0+m+r)]^T$. Analogiczną asymetrię wprowadza się w przypadku rekonstrukcji sygnału przy użyciu algorytmu filtracji Kalmana (3.13), zaprojektowanego w oparciu o model stanowy (3.12) równoważny opisowi (3.1) i (5.10), gdzie $q = p + r + m$ oraz

$$\mathbf{A}_q = \begin{bmatrix} a_1 & a_2 & \dots & a_r & 0 & \dots & a_T & \dots & a_p & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & & 0 & 0 & & 0 & & 0 & 0 & & 0 & 0 \\ \vdots & & & & & & & & & & \ddots & & \vdots \\ 0 & 0 & & 0 & 0 & \dots & & \dots & 0 & 0 & & 1 & 0 \end{bmatrix}. \quad (5.11)$$

Algorytm kończy swoją pracę w chwili $t_0 + m + r$ po wykorzystaniu *tylko* r niezakłóconych próbek sygnału występujących bezpośrednio po bloku uszkodzonych próbek. Ponieważ oba podejścia stają się już niepraktyczne ze względu na rozmiary występujących macierzy, wprowadzone zostaną pewne uproszczenia zmniejszające złożoność obliczeniową algorytmu rekonstrukcji.

Założmy, że $q_{\max} < T_{\min} + r$, gdzie T_{\min} oznacza minimalną dozwoloną wartość okresu podstawowego T . Przy tym założeniu model SAR (5.9) może być traktowany jako autoregresyjny model z zewnętrznym wejściem (z ang. *autoregressive with exogenous input*, ARX)

$$s(t) = \sum_{i=1}^r a_i s(t-i) + u(t) + n(t), \quad (5.12)$$

gdzie wielkość

$$u(t) = \sum_{i=T}^{T+r} a_i s(t-i) \quad (5.13)$$

traktowana jest jako zewnętrzny (mierzalny) sygnał wejściowy, a nie jako wewnętrzna zmienna stanu. Zauważmy, że nawet gdy blok próbek podlegających rekonstrukcji osiągnie swoją maksymalną dopuszczalną (ustaloną z góry) długość m_{\max} , zakwestionowane próbki sygnału z przedziału $[t_0 + 1, t_0 + m_{\max}]$ nie zostaną wykorzystane przy obliczaniu wielkości $u(t)$. Zwróćmy uwagę, że sygnał $u(t)$ może zostać zastąpiony przez $\hat{x}(t) = cx(t - T)$.

Równoważna „pseudo-reprezentacja” w przestrzeni stanów może zostać zapisana w następującej postaci

$$\begin{aligned} \mathbf{x}_q(t+1) &= \mathbf{A}_q \mathbf{x}_q(t) + \mathbf{c}_q u(t+1) + \mathbf{c}_q n(t+1), \\ y(t) &= \mathbf{c}_q^T \mathbf{x}_q(t) + \delta(t), \end{aligned} \quad (5.14)$$

gdzie $q = 2r + m$. Projektując filtr Kalmana w oparciu o taki „pseudo-model” w przestrzeni stanów, uzyskujemy następujący algorytm o zredukowanej złożoności obliczeniowej

$$\begin{aligned} \hat{\mathbf{x}}_q(t|t-1) &= \mathbf{A}_q \hat{\mathbf{x}}_q(t-1|t-1) + \mathbf{c}_q u(t), \\ \mathbf{P}_q(t|t-1) &= \mathbf{A}_q \mathbf{P}_q(t-1|t-1) \mathbf{A}_q^T + \rho \mathbf{c}_q \mathbf{c}_q^T, \\ e(t) &= y(t) - \mathbf{c}_q^T \hat{\mathbf{x}}_q(t|t-1), \\ g(t) &= \mathbf{c}_q^T \mathbf{P}_q(t|t-1) \mathbf{c}_q + \gamma(t), \\ \mathbf{l}_q(t) &= \mathbf{P}_q(t|t-1) \mathbf{c}_q / g(t), \\ \hat{\mathbf{x}}_q(t|t) &= \hat{\mathbf{x}}_q(t|t-1) + \mathbf{l}_q(t) e(t), \\ \mathbf{P}_q(t|t) &= \mathbf{P}_q(t|t-1) - g(t) \mathbf{l}_q(t) \mathbf{l}_q^T(t), \end{aligned} \quad (5.15)$$

który zachowuje zdolności interpolacyjne filtra Kalmana (3.13) zaprojektowanego w oparciu o poprawny model stanowy (3.12) – wyniki interpolacji nie ulegają zmianie.

5.4 Detekcja oparta na modelu SAR

Na początek zauważmy, że równanie jednokrokowej predykcji opartej na zdekomponowanym modelu SAR (5.3)-(5.4) przyjmuje postać

$$\begin{aligned} \hat{y}(t|t-1) &= \sum_{i=1}^r b_i y(t-i) + \hat{x}(t|t-1), \\ \hat{x}(t|t-1) &= cx(t-T), \end{aligned} \quad (5.16)$$

gdzie sygnał resztowy $x(t)$ otrzymywany jest w wyniku filtracji odwrotnej

$$x(t) = y(t) - \sum_{i=1}^r b_i y(t-i). \quad (5.17)$$

W algorytmach detekcji, opisanych w punktach 3.4.1 i 3.4.2, modyfikacji ulegają jedynie równania odpowiadające za (wielokrokową) predykcję sygnału, odpowiednio równanie (3.81) i pierwsze równanie w bloku (3.19). Wszystkie pozostałe kroki wykonywane są w ten sam sposób jak dla modelu AR, włącznie z warunkiem końca alarmu detekcyjnego (3.78).

W przypadku schematu detekcji w pętli otwartej, równanie wielokrokowej predykcji (3.81) zastąpione zostaje przez następujący rekurencyjny algorytm

$$\begin{aligned} \hat{y}(t+k|t) &= \sum_{i=1}^r b_i \hat{y}(t+k-i|t) + \hat{x}(t+k|t), \\ \hat{x}(t+k|t) &= cx(t+k-T|t), \quad k = 1, 2, \dots, \end{aligned} \quad (5.18)$$

gdzie dla $j \geq 0$ $\hat{y}(t-j|t) = y(t-j)$. Oceny wariancji błędów wielokrokowej predykcji mogą być otrzymane przy użyciu algorytmu (3.82), po przepisaniu modelu z postaci zdekomponowanej do postaci (5.1).

Podobnie w przypadku schematu detekcji z decyzyjnym sprzężeniem zwrotnym, równanie predykcji w filtrze Kalmana o zmiennym rzędzie – pierwsze równanie w bloku (3.19) – zostaje rozszerzone o dodatkowy składnik

$$\hat{y}(t|t-1) = \boldsymbol{\theta}_{q(t)-1}^T \hat{\mathbf{x}}_{q(t)-1}(t-1|t-1) + \hat{x}(t|t-1)\mathbf{c}_{q(t)-1}. \quad (5.19)$$

Po uwzględnieniu powyższej modyfikacji, stopniowe zwiększanie rzędu w algorytmie filtracji Kalmana o zmiennym rzędzie rozpoczyna się od $r+1$ w chwili t_0 , tak jak w klasycznym przypadku.

5.5 Identyfikacja modelu SAR

Identyfikacja modelu SAR w postaci zdekomponowanej przebiega w dwóch krokach. W pierwszym kroku identyfikowane są parametry modelu krótkozakresowego (5.3) i wykorzystane zostają do obliczenia sygnału resztowego $x(t)$ (5.17). Do estymacji współczynników modelu formantowego użyty zostaje algorytm EWLS (3.39). W kolejnym kroku estymowane są parametry modelu długozakresowego (5.4).

Identyfikacja modelu długozakresowego

Zakładając, że okres podstawowy T jest znany, spełniona jest następująca korelacyjna zależność dla modelu długozakresowego

$$\mathbf{r}_T^x = c\mathbf{r}_0^x, \quad (5.20)$$

gdzie $\mathbf{r}_k^x = E[x(t)x(t-k)]$. Lokalna ocena współczynnika c może być otrzymana na drodze rozwiązania równania

$$\hat{\mathbf{r}}_T^x(t) = \hat{c}(t)\hat{\mathbf{r}}_0^x(t), \quad (5.21)$$

gdzie

$$\hat{\mathbf{r}}_k^x(t) = \frac{1}{N} \sum_{i=0}^{N-1} x(t-i)x(t-i-k), \quad k = 0, T,$$

są lokalnymi nieobciążonymi ocenami współczynników $\mathbf{r}_0^x(t)$ i $\mathbf{r}_T^x(t)$. Niestety w większości aplikacji fonicznych okres podstawowy jest nieznan i zmienny w czasie. Ponieważ model długozakresowy wykorzystuje tylko jeden składnik, wybór „właściwej” wartości T jest kluczowy – nawet małe odchylenie oszacowanej wartości T od jej prawdziwej wartości może skutkować znaczącym spadkiem zdolności predykcyjnych modelu.

Na przestrzeni lat opracowano bardzo wiele algorytmów śledzących okres podstawowy sygnału (z ang. *Pitch Determination*, PD) – patrz prace i zawarte w nich odniesienia [82], [83], [84]. W ogólności, algorytmy te mogą być podzielone na trzy kategorie: algorytmy działające w dziedzinie czasu, algorytmy działające w dziedzinie częstotliwości oraz algorytmy łączące analizę w dziedzinie czasu z analizą w dziedzinie częstotliwości. Wybór odpowiedniego algorytmu estymacji okresu T oraz obiektywnego kryterium oceny uzyskiwanych wyników nie jest prostym zadaniem i zależy od przeznaczenia danej aplikacji, na co zwrócił już uwagę Rabiner w swojej pracy [85]. W naszym przypadku tylko dokładne oszacowania T skutkować będą polepszeniem

zdolności predykcyjnych modelu wysokości dźwięku (5.4), a w konsekwencji też modelu SAR. Dlatego dobrym kryterium obiektywnej oceny algorytmu PD jest minimalizacja sumy kwadratów błędów jednokrokowej predykcji. Zwróćmy uwagę, że nawet gdy algorytm PD wybierze wielokrotność okresu podstawowego, nie wpływa to negatywnie na zdolności predykcyjne modelu, wręcz przeciwnie – w niektórych sytuacjach może okazać się korzystniejsze. Zwróćmy uwagę na jeszcze jeden istotny fakt – „przyszłe” (nieprzetworzone) próbki sygnału mogą być uszkodzone przez zakłócenia impulsowe, dlatego algorytm PD, który ocenia okres T na podstawie przeszłych próbek sygnału i pewnej liczby próbek „przyszłych” będzie działać zadowalająco dopóki w jego zasięgu nie pojawi się zakłócenie impulsowe. Z kolei algorytm oparty tylko na przeszłych (przetworzonych) próbkach sygnału może czasami dawać niedokładne oszacowania wielkości T . Pewnym kompromisem może być równoległa estymacja przyczynowa i nieprzyczynowa okresu T , a w momencie zgłoszenia alarmu, dokonanie wyboru pomiędzy nimi na podstawie kryterium oceniającego dopasowanie modelu do sygnału. Rozważane były różne algorytmy PD działające w dziedzinie czasu i oparte na: funkcji autokorelacji, ważonej funkcji autokorelacji [86], funkcji wartości bezwzględnych różnic [87] oraz funkcji wartości kwadratów różnic [88]. Na podstawie wspomnianego kryterium wybrano podejście oparte na minimalizacji sumy kwadratów różnic pomiędzy dwiema przesuniętymi względem siebie sekwencjami próbek – patrz praca [88].

A) Estymacja przyczynowa okresu T

W celu estymacji okresu T wykorzystany zostanie sygnał $x(t)$ otrzymywany w wyniku filtracji odwrotnej. Technika szukania najlepszego oszacowania okresu T oparta jest na dopasowywaniu sekwencji N ostatnio zaobserwowanych próbek sygnału $\mathcal{F}_N(t) = \{x(t), \dots, x(t - N + 1)\}$ do jego przeszłych próbek w ustalonym przedziale szukania. Niech

$$\begin{aligned} J_t(T, N) &= \sum_{i=0}^{N-1} |x(t-i) - x(t-T-i)|^2 = \\ &= p_k^x(t, N) - 2p_T^x(t, N) + p_0^x(t-T, N), \end{aligned} \quad (5.22)$$

gdzie wielkość $p_k^x(t, N)$, $k = 0, T$, po unormowaniu, jest identyczna z $\hat{r}_k^x(t)$. Zauważmy, że gdy funkcja $J_t(T, N)$ liczona jest przy użyciu współczynników $p_k^x(t, N)$, jednocześnie uzyskujemy potrzebne wielkości do obliczenia $\hat{c}(t)$. Dodatkowo, współczynniki te mogą być liczone rekurencyjnie przy użyciu następującej procedury

$$\begin{aligned} p_k^x(t, N+1) &= p_k^x(t-1, N) + x(t)x(t-k), \\ p_k^x(t, N) &= p_k^x(t, N+1) - x(t-N)x(t-N+k), \end{aligned} \quad (5.23)$$

co zmniejsza koszt obliczeniowy proponowanego podejścia.

Oznaczmy symbolami T_{\min}/T_{\max} najmniejszą/największą wartość okresu podstawowego dla danej częstotliwości próbkowania. Proponowana procedura estymacji jest dwukrokowa. W pierwszym kroku w wyniku przeszukania całego zakresu dla porównywanej sekwencji próbek o długości M_0 otrzymywana jest przybliżona ocena okresu T , oznaczona symbolem $\hat{T}_0(t)$

$$\hat{T}_0(t) = \arg \min_{T \in \Delta_0} J_t(T, M_0), \quad (5.24)$$

gdzie $\Delta_0 = [T_{\min}, T_{\max}]$. Następnie, w wyniku lokalnego sprawdzenia wokół wartości $\hat{T}_0(t)$, przeprowadzonego dla krótszej sekwencji próbek o długości $M_* \ll M_0$, ustalona zostaje dokładniejsza ocena okresu, oznaczona symbolem $\hat{T}(t)$

$$\hat{T}(t) = \arg \min_{T \in \Delta(t)} J_t(T, M_*), \quad (5.25)$$



gdzie $\Delta(t) = [\max\{T_{\min}, \hat{T}_0(t) - k_0\}, \min\{T_{\max}, \hat{T}_0(t) + k_0\}]$, a k_0 jest małą liczbą całkowitą.

Łącząc (5.24) i (5.25) otrzymujemy następującą ocenę współczynnika modelu długozakresowego

$$\hat{c}(t) = \hat{r}_{\hat{T}(t)}^x(t) [\hat{r}_0^x(t)]^{-1}. \quad (5.26)$$

Po podstawieniu oszacowanych parametrów otrzymujemy następującą postać modelu długozakresowego

$$x(t) = \hat{c}(t)x[t - \hat{T}(t)] + n(t). \quad (5.27)$$

B) Nieprzyczynowa estymacja okresu T

Jak już wspomniano, zdolności predykcyjne modelu SAR silnie zależą od dokładnego ustalenia okresu podstawowego T . Za każdym razem gdy okres podstawowy podlega nagłym zmianom, ocena przyczynowa $\hat{T}(t)$, oparta wyłącznie na przeszłych wartościach sygnału, może nie być wystarczająco dokładna dla potrzeb śledzenia $T(t)$. Znacznie lepsze oceny otrzymywane są gdy porównywana sekwencja próbek obejmuje zarówno przeszłe próbki jak i pewną liczbę „przyszłych” (jeszcze nieprzetworzonych) próbek – chyba że „przyszłe” próbki są poważnie uszkodzone.

Odpowiednia nieprzyczynowa ocena $\hat{T}^*(t)$ może zostać otrzymana przy użyciu dwukrokowej procedury analogicznej do (5.24)-(5.25)

$$\hat{T}_0^*(t) = \arg \min_{T \in \Delta_0} J_t^*(T, M_0, m_0), \quad (5.28)$$

$$\hat{T}^*(t) = \arg \min_{T \in \Delta^*(t)} J_t^*(T, M_*, m_*), \quad (5.29)$$

gdzie

$$J_t^*(T, N, l) = \sum_{i=0}^{N-1} |x(t+l-i|t) - x(t+l-T-i|t)|^2,$$

$$\Delta^*(t) = [\max\{T_{\min}, \hat{T}_0^*(t) - k_0\}, \min\{T_{\max}, \hat{T}_0^*(t) + k_0\}]$$

i

$$x(t+k|t) = \begin{cases} y(t+k) - \sum_{i=1}^r \hat{b}_i(t)y(t+k-i), & \text{dla } k > 0, \\ x(t+k), & \text{dla } k \leq 0. \end{cases}$$

Liczby całkowite m_0 i m_* , $m_* \ll m_0$ oznaczają liczbę „przyszłych” próbek $x(t)$ uwzględnianych gdy obliczane są odpowiednio $\hat{T}_0^*(t)$ i $\hat{T}^*(t)$. Zauważmy, że dla $k > 0$ wielkość $x(t+k|t)$ różni się od $x(t+k) = y(t+k) - \sum_{i=1}^r \hat{b}_i(t+k-1)y(t+k-i)$, ponieważ zależy od „zamrożonych” parametrów modelu w chwili t . Nieprzyczynowa ocena $\hat{c}^*(t)$ współczynnika c może zostać otrzymana w sposób analogiczny do (5.26)

$$\hat{c}^*(t) = \hat{r}_{\hat{T}^*(t)}^x(t|t+m_*) [\hat{r}_0^x(t|t+m_*)]^{-1}, \quad (5.30)$$

gdzie

$$\hat{r}_k^x(t|t+m_*) = \frac{1}{N} \sum_{i=0}^{N-1} x(t+m_*-i|t)x(t+m_*-i-k|t).$$

Po podstawieniu otrzymujemy model postaci

$$x(t) = \hat{c}^*(t)x[t - \hat{T}^*(t)] + n(t). \quad (5.31)$$

Model (5.31) pozwala na znaczącą redukcję liczby fałszywych alarmów detekcyjnych występujących gdy – z powodu mniejszej dokładności w śledzeniu zmian okresu podstawowego – model (5.27) generuje zbyt duże błędy predykcji. Zdarza się to zazwyczaj gdy okres podstawowy podlega szybkim zmianom – w przeciwnym razie oceny $\hat{T}(t)$ i $\hat{T}^*(t)$ często są identyczne.

5.6 Porównanie modeli AR i SAR

W podrozdziale tym porównano zdolności predykcyjne i interpolacyjne modeli AR i SAR na krótkim fragmencie sygnału mowy, spróbkowanym z częstotliwością 22,050 kHz. Model formantowy o rzędzie $r = 10$ identyfikowany był na bieżąco przy użyciu algorytmu EWLS wyposażonego w stałą zapomnienia $\lambda = 0,998$. Do identyfikacji modelu wysokości dźwięku użyte zostały następujące wartości parametrów: $M_0 = 160$, $M_* = 40$, $m_0 = 20$, $m_* = 5$, $k_0 = 3$, $T_{\min} = 110$, $T_{\max} = 1000$.

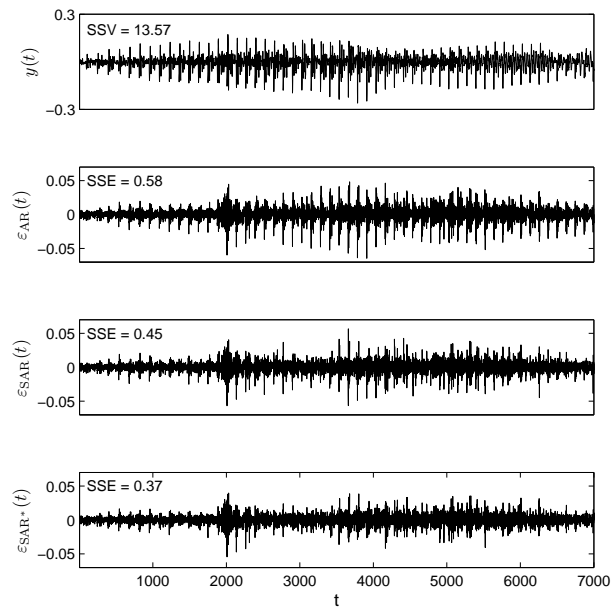
Oznaczmy symbolami $\varepsilon_{\text{AR}}(t+1|t)$, $\varepsilon_{\text{SAR}}(t+1|t)$ i $\varepsilon_{\text{SAR}^*}(t+1|t)$ ciągi błędów jednokrokowej predykcji generowane odpowiednio przez modele: AR (5.17), przyczynowy jednorodny model SAR (5.17), (5.27) i nieprzyczynowy jednorodny model SAR* (5.17), (5.31) (zauważmy, że w ostatnim przypadku „predykcja” nie jest już odpowiednim terminem):

$$\begin{aligned} \varepsilon_{\text{AR}}(t+1|t) &= y(t+1) - \sum_{i=1}^r \hat{b}_i(t)y(t-i+1), \\ \varepsilon_{\text{SAR}}(t+1|t) &= x(t+1) - \hat{c}(t)x[t+1 - \hat{T}(t)], \\ \varepsilon_{\text{SAR}^*}(t+1|t) &= x(t+1) - \hat{c}^*(t)x[t+1 - \hat{T}^*(t)], \end{aligned}$$

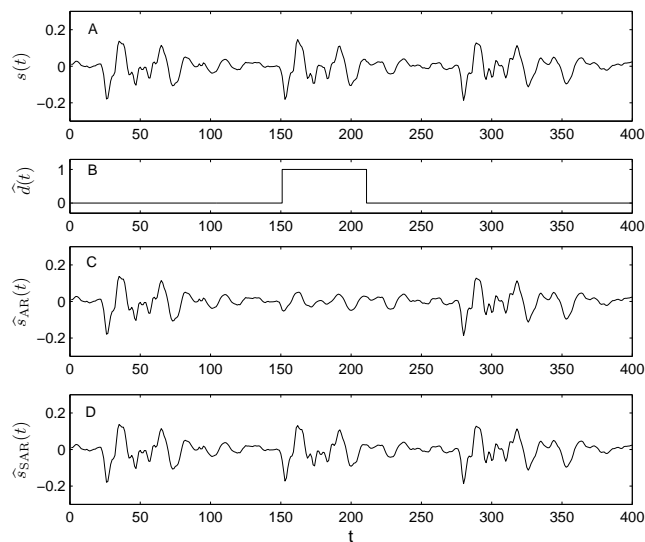
gdzie $x(t+1) = \varepsilon_{\text{AR}}(t+1|t)$. Rysunek 5.2 pokazuje porównanie błędów jednokrokowej predykcji otrzymanych dla krótkiego fragmentu sygnału mowy, w trzech rozważanych powyżej przypadkach. Zauważmy, że zdolności predykcyjne modelu SAR zostały zwiększone w porównaniu z modelem AR, z kolei model SAR* posiada lepsze zdolności predykcyjne niż model SAR.

Rysunek 5.3 ilustruje zdolności interpolacyjne modeli AR i SAR w miejscu występowania silnego pobudzenia krtaniowego. Blok próbek podlegający rekonstrukcji ma długość 60 próbek. Zauważmy, że zdolności interpolacyjne modelu SAR (wykres D) zostały znacząco zwiększone w porównaniu z modelem AR (wykres C), a uzyskana interpolacja przypomina „kopię” sygnału z przeszłości. Testy odsłuchowe nie wykazały słyszalnych różnic pomiędzy sygnałem oryginalnym, a jego rekonstrukcją opartą na modelu SAR. Z kolei rekonstrukcja sygnału oparta na modelu AR powoduje zniekształcenie dźwięku przypominające „głuche tąpnięcie”.

Zauważmy na koniec, że przydatność modelu długozakresowego zależy przede wszystkim od obecności w sygnale $x(t)$ silnej długozakresowej struktury korelacyjnej, takiej jak ta pokazana na rysunku 5.2. Jeżeli jej nie ma, zdolności predykcyjne i interpolacyjne modelu SAR wynikającego z kaskadowego połączenia modeli (5.3) i (5.4) niewiele różnią się od tych obserwowanych dla modelu formantowego (5.3).



Rysunek 5.2: Porównanie błędów jednokrokowej predykcji otrzymanych dla krótkiego fragmentu sygnału mowy. Od góry: sygnał mowy, ciąg błędów jednokrokowej predykcji generowany przez model AR ($\varepsilon_{AR}(t)$), ciąg błędów jednokrokowej predykcji generowany przez model SAR ($\varepsilon_{SAR}(t)$) oraz ciąg błędów jednokrokowej predykcji generowany przez model SAR* ($\varepsilon_{SAR^*}(t)$). SSV oznacza sumę kwadratów wartości sygnału, a SSE – sumę kwadratów błędów jednokrokowej predykcji. Zauważmy różnicę w skali pomiędzy wykresem sygnału, a wykresami błędów.



Rysunek 5.3: Sygnał mowy (A), lokalizacja bloku 60 próbek podlegającego interpolacji (B), interpolacja oparta na modelu AR (C) i interpolacja oparta na modelu SAR (D).

5.7 Schemat detekcji zakłóceń impulsowych

Aby zwiększyć odporność schematu detekcji zakłóceń impulsowych i jednocześnie zmniejszyć liczbę fałszywych alarmów, decyzja o podniesieniu alarmu detekcyjnego wypracowywana jest wspólnie przez trzy działające równolegle detektory zakłóceń impulsowych. Dwa z nich oparte są na modelach SAR o tych samych współczynnikach modelu formantowego $\hat{b}_1(t), \dots, \hat{b}_r(t)$ lecz różnych parametrach modelu wysokości dźwięku odpowiednio $\{\hat{T}(t), \hat{c}(t)\}$ i $\{\hat{T}^*(t), \hat{c}^*(t)\}$. Trzeci detektor oparty jest na klasycznym modelu AR wykorzystującym tylko współczynniki modelu formantowego.

Po zastosowaniu reguły detekcji zakłócenia impulsowego (3.76) do ciągu błędów predykcji generowanych przez każdy z konkurujących modeli,

$$\hat{d}_{AR}(t_0 + 1) = \begin{cases} 1, & \text{jeżeli } |\varepsilon_{AR}(t_0 + 1|t_0)| > \mu\hat{\sigma}_{\varepsilon_{AR}}(t_0 + 1|t_0), \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad (5.32)$$

$$\hat{d}_{SAR}(t_0 + 1) = \begin{cases} 1, & \text{jeżeli } |\varepsilon_{SAR}(t_0 + 1|t_0)| > \mu\hat{\sigma}_{\varepsilon_{SAR}}(t_0 + 1|t_0), \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad (5.33)$$

$$\hat{d}_{SAR^*}(t_0 + 1) = \begin{cases} 1, & \text{jeżeli } |\varepsilon_{SAR^*}(t_0 + 1|t_0)| > \mu\hat{\sigma}_{\varepsilon_{SAR^*}}(t_0 + 1|t_0), \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad (5.34)$$

otrzymujemy trzy tymczasowe wyniki detekcji $\hat{d}_{AR}(t_0+1)$, $\hat{d}_{SAR}(t_0+1)$ i $\hat{d}_{SAR^*}(t_0+1)$. Zauważmy, że dla każdego ciągu błędów jednokrokowej predykcji wyznaczana jest jego lokalna ocena wariancji przy użyciu rekursji (3.61). Alarm detekcyjny podnoszony jest dopiero gdy wszystkie trzy błędy predykcji mogą być traktowane jako nadmiarowe, mianowicie jeżeli

$$\hat{d}_{AR}(t_0 + 1) \neq 0, \quad \hat{d}_{SAR}(t_0 + 1) \neq 0 \quad \text{i} \quad \hat{d}_{SAR^*}(t_0 + 1) \neq 0. \quad (5.35)$$

Jeżeli alarm detekcyjny nie został podniesiony, ponieważ tylko jeden z trzech lub dwa z trzech błędów predykcji sklasyfikowane zostały jako nadmiarowe, przykładowo gdy

$$\hat{d}_{AR}(t_0 + 1) = 1, \quad \hat{d}_{SAR}(t_0 + 1) = 0 \quad \text{i} \quad \hat{d}_{SAR^*}(t_0 + 1) = 0, \quad (5.36)$$

wtedy błędy predykcji, traktowane jako nadmiarowe, nie są używane przy aktualizacji odpowiadającej im lokalnej oceny wariancji, tj. $\hat{\rho}_{AR}(t) = \hat{\rho}_{AR}(t-1)$. Zapobiega to sytuacji, w której w wyniku odwołania tymczasowej decyzji detektora, staje się on w konsekwencji zbyt tolerancyjny na przyszłe zmiany w sygnale.

Gdy alarm detekcyjny jest podniesiony w chwili $t_0 + 1$ (po sekwencji co najmniej r decyzji o braku alarmu), zadania detekcji, a następnie interpolacji sygnału mogą zostać wykonane przez odpowiednio zaprojektowany filtr Kalmana o zmiennym rzędzie, uruchamiany w chwili $t_0 + 1$. Alarm detekcyjny jest przerywany w chwili $t_0 + m$ jeżeli r kolejnych błędów predykcji przyjmuje odpowiednio małe wartości: $\hat{d}_0(t_0 + m + i) = 0$ dla $i = 1, \dots, r$. Algorytm filtru Kalmana oparty jest na modelu (AR, SAR lub SAR*), który minimalizuje sumę kwadratów błędów jednokrokowej predykcji obserwowanych w niezbyt dalekiej przeszłości $\sum_{i=1}^r |\varepsilon(t_0 + 1 - i)|^2$.

Uwaga 5.2. Pierwotnie stosowano nieco inny schemat detekcji zakłóceń impulsowych opisany w pracy [5]. Ponieważ na podstawie łącznej analizy sygnałów $y(t)$ i $\hat{y}(t)$ otrzymywano w tym podejściu aż cztery różne oceny okresu podstawowego T , końcowa decyzja o położeniu zakłócenia impulsowego wypracowywana była wspólnie przez pięć działających równolegle detektorów zakłóceń impulsowych, z których cztery były oparte na modelach SAR, zaś jeden

– na modelu AR. Alarm detekcyjny rozpoczynał się, gdy wszystkich pięć detektorów wskazywało obecność zakłócenia impulsowego, zaś kończył gdy co najmniej jeden z detektorów zaakceptował r kolejnych próbek sygnału. Końcowa interpolacja oparta była na modelu, który był odpowiedzialny za przerwanie alarmu detekcyjnego. Zastosowanie dwukrokowej procedury estymacji wielkości T na podstawie analizy sygnału $x(t)$ pozwoliło na zredukowanie liczby detektorów przy jednoczesnej poprawie otrzymywanych ocen okresu podstawowego. Ponadto, uproszczono sposób wyboru lokalnie „najlepszego” modelu, na którym oparta jest procedura detekcji zakłócenia impulsowego oraz procedura rekonstrukcji sygnału.

5.8 Podsumowanie

W rozdziale tym opisano nowe podejście do eliminacji zakłóceń impulsowych z sygnałów fonicznych wykorzystujące autoregresyjne modelowanie rzadkie. Zaproponowany sposób identyfikacji modelu SAR umożliwił łatwe monitorowanie jego stabilności i wymuszanie jej w wymaganych sytuacjach. Do detekcji, a następnie interpolacji sygnału wykorzystano odpowiednio zaprojektowany filtr Kalmana o zmiennym rzędzie, uruchamiany w chwili $t_0 + 1$, oparty na „pseudo-reprezentacji” modelu stanowego. Zaproponowany tu schemat detekcji zakłóceń impulsowych pozwala na znaczącą redukcję liczby fałszywych alarmów detekcyjnych.

Rozdział 6

Model wektorowy

Wprowadzenie

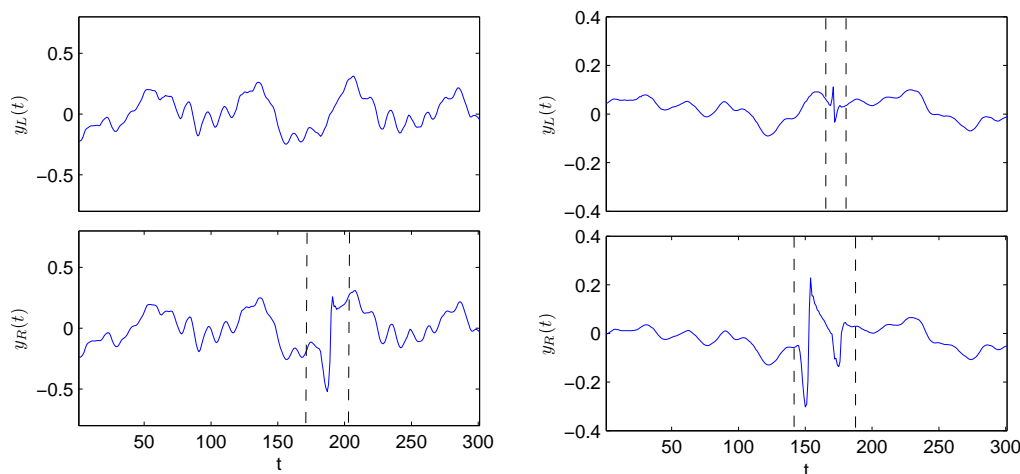
Chociaż dwie składowe stereofonicznego sygnału, odpowiadające lewemu i prawemu kanałowi, mogą być przetwarzane oddzielnie, nie jest to z pewnością najlepsze podejście do rekonstrukcji nagrań stereofonicznych. W niniejszym rozdziale omówimy nowe podejście do problemu eliminacji zakłóceń impulsowych w sygnałach stereofonicznych. Można pokazać, że zarówno detekcji jak i rekonstrukcji (interpolacji) uszkodzonych próbek sygnału można dokonać lepiej w przypadku gdy oba kanały analizowane są łącznie przy użyciu techniki wektorowego modelowania autoregresyjnego.

W przypadku rekonstrukcji sygnału dokładniejsze modelowanie daje dwie potencjalne korzyści. Po pierwsze, ponieważ większość istniejących procedur wykrywania zakłóceń impulsowych opartych jest na adaptacyjnym progowaniu sygnału błędów predykcji, przetwarzanie wektorowe sprawia, że detektory stają się bardziej wrażliwe na zmiany charakterystyk sygnału. Po drugie, dokładniejsze modele gwarantują zazwyczaj lepszą interpolację nieodwracalnie uszkodzonych próbek sygnału, zakwestionowanych przez detektor zakłóceń impulsowych.

Gdy materiał foniczny podlegający rekonstrukcji pochodzi z gramofonowego nagrania stereofonicznego, można spodziewać się dodatkowych korzyści z zastosowania podejścia wektorowego. W przypadku takich nagrań lokalne uszkodzenia rowka lub różnego rodzaju niedoskonałości powierzchni płyty, powodujące zakłócenia impulsowe, są zazwyczaj asymetryczne, tj. są bardziej widoczne na lewym lub prawym brzegu rowka, albo umiejscowione są tylko po jego jednej stronie. Ponieważ typowy system nagrywania/odtwarzania stereofonicznego wykorzystuje oba zbocza tego samego rowka (tzw. system 45/45), tj. zachowuje wspomnianą asymetrię (patrz rysunek 6.1), rekonstrukcja może być przeprowadzona lepiej w przypadku gdy oba kanały są analizowane łącznie – nieuszkodzony materiał w jednym kanale może być bowiem pomocny w wykrywaniu i interpolowaniu uszkodzonych próbek w drugim kanale.

6.1 Identyfikacja sygnału stereofonicznego

Zarejestrowany sygnał stereofoniczny oznaczajmy symbolem $\mathbf{y}(t) = [y_1(t), y_2(t)]^T$, gdzie $y_1(t)$ i $y_2(t)$ oznaczają odpowiednio lewy i prawy kanał. Dla uproszczenia przyjmujemy, że



Rysunek 6.1: Typowe konfiguracje zakłóceń impulsowych spotykane w archiwalnych nagraniach stereofonicznych: sytuacja gdy tylko jeden z dwóch kanałów stereofonicznych został uszkodzony przez zakłócenie impulsowe (lewy rysunek) oraz sytuacja gdy oba kanały stereofoniczne zostały uszkodzone przez zakłócenia impulsowe (prawy rysunek). Przerywane linie wskazują początek/koniec każdego zakłócenia impulsowego.

sygnał $\mathbf{y}(t)$ można zapisać w postaci

$$\mathbf{y}(t) = \mathbf{s}(t) + \boldsymbol{\delta}(t), \quad (6.1)$$

gdzie $\mathbf{s}(t) = [s_1(t), s_2(t)]^T$ oznacza sygnał czysty, a $\boldsymbol{\delta}(t) = [\delta_1(t), \delta_2(t)]^T$ jest ciągiem pojawiających się sporadycznie zakłóceń impulsowych.

Czysty sygnał stereofoniczny będzie modelowany jako wektorowy dwuwymiarowy proces autoregresyjny (z ang. *vector autoregressive*, VAR) rzędu r [89], [90], [91]

$$\mathbf{s}(t) = \sum_{i=1}^r \mathbf{A}_i \mathbf{s}(t-i) + \mathbf{n}(t), \quad (6.2)$$

gdzie

$$\mathbf{A}_i = \begin{bmatrix} a_{11,i} & a_{12,i} \\ a_{21,i} & a_{22,i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_{1i}^T \\ \boldsymbol{\alpha}_{2i}^T \end{bmatrix}, \quad i = 1, \dots, r,$$

są macierzami współczynników autoregresji o wymiarach 2×2 , a $\{\mathbf{n}(t)\}$,

$$\mathbf{n}(t) = [n_1(t), n_2(t)]^T,$$

oznacza dwuwymiarowy szum biały o zerowej wartości oczekiwanej i macierzy kowariancji

$$\text{cov}[\mathbf{n}(t)] = \begin{bmatrix} \rho_1^2 & \rho_{12} \\ \rho_{12} & \rho_2^2 \end{bmatrix} = \boldsymbol{\rho}.$$

Oznaczmy przez $\boldsymbol{\theta}_j = [\boldsymbol{\alpha}_{j1}^T, \dots, \boldsymbol{\alpha}_{jr}^T]^T$ wektor współczynników charakteryzujący j -ty kanał, a przez $\boldsymbol{\varphi}(t) = [\mathbf{y}^T(t-1), \dots, \mathbf{y}^T(t-r)]^T$ – odpowiadający mu wektor regresji (ten sam dla obu kanałów). Oznaczmy w tym rozdziale symbolem $\mathbf{0}_r$ wektor zerowy o wymiarach $2r \times 1$, zaś

symbolami \mathbf{O}_r i \mathbf{I}_r odpowiednio – macierz zerową i macierz jednostkową o wymiarach $2r \times 2r$. Następnie, niech

$$\Phi(t) = \begin{bmatrix} \varphi(t) & \mathbf{0}_r \\ \mathbf{0}_r & \varphi(t) \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}. \quad (6.3)$$

Na podstawie wprowadzonej wyżej skróconej notacji model (6.2), w przypadku nieobecności zakłóceń impulsowych, może zostać przepisany w postaci

$$\mathbf{y}(t) = \Phi^T(t)\theta + \mathbf{n}(t). \quad (6.4)$$

Aby uwzględnić niestacjonarność sygnałów fonicznych stosowany będzie następujący zależny od czasu wariant modelu (6.4)

$$\mathbf{y}(t) = \Phi^T(t)\theta(t) + \mathbf{n}(t), \quad \text{cov}[\mathbf{n}(t)] = \rho(t), \quad (6.5)$$

gdzie $\rho(t)$ jest macierzą kowariancji białego szumu, wolno zmieniającą się w czasie.

Estymacja wektora parametrów $\theta(t)$ prowadzona będzie przy użyciu wykładniczo ważonej metody najmniejszych kwadratów (patrz podrozdział 3.3)

$$\hat{\theta}(t) = \arg \min_{\theta} \sum_{k=1}^t \lambda^{t-k} \|\mathbf{y}(k) - \Phi^T(k)\theta\|^2. \quad (6.6)$$

Rozwiązując zadanie minimalizacji (6.6) otrzymamy estymator znanej postaci

$$\hat{\theta}(t) = \left[\sum_{k=1}^t \lambda^{t-k} \Phi(k)\Phi^T(k) \right]^{-1} \left[\sum_{k=1}^t \lambda^{t-k} \Phi(k)\mathbf{y}(k) \right]. \quad (6.7)$$

Ponieważ macierz $\Phi(k)$ ma strukturę blokowo-diagonalną, estymator (6.7) może zostać przepisany w postaci zdekomponowanej

$$\hat{\theta}_j(t) = \left[\sum_{k=1}^t \lambda^{t-k} \varphi(k)\varphi^T(k) \right]^{-1} \left[\sum_{k=1}^t \lambda^{t-k} \varphi(k)y_j(k) \right], \quad j = 1, 2. \quad (6.8)$$

Podobna technika może zostać użyta do śledzenia macierzy kowariancji $\rho(t)$. Lokalna ocena $\rho(t)$ może zostać otrzymana analogicznie jak dla przypadku skalarnego (3.57)

$$\hat{\rho}(t) = \frac{\mathbf{W}(t)}{l_e(t)}, \quad (6.9)$$

gdzie $\mathbf{W}(t)$ oznacza wykładniczo ważoną sumę kwadratów błędów resztowych

$$\mathbf{W}(t) = \sum_{k=1}^t \lambda^{t-k} [\mathbf{y}(k) - \Phi^T(k)\hat{\theta}(t)] [\mathbf{y}(k) - \Phi^T(k)\hat{\theta}(t)]^T. \quad (6.10)$$

Rekurencyjne algorytmy estymacji

Wielkości $\hat{\boldsymbol{\theta}}(t)$ i $\hat{\boldsymbol{\rho}}(t)$ mogą być estymowane rekurencyjnie. Do śledzenia $\hat{\boldsymbol{\theta}}_j(t)$ wykorzystywany jest rekurencyjny algorytm EWLS

$$\begin{aligned}\varepsilon_j(t|t-1) &= y_j(t) - \boldsymbol{\varphi}^T(t)\hat{\boldsymbol{\theta}}_j(t-1), \\ \hat{\boldsymbol{\theta}}_j(t) &= \hat{\boldsymbol{\theta}}_j(t-1) + \mathbf{k}(t)\varepsilon_j(t|t-1), \\ j &= 1, 2, \\ \mathbf{k}(t) &= \frac{\mathbf{Q}(t-1)\boldsymbol{\varphi}(t)}{\lambda + \boldsymbol{\varphi}^T(t)\mathbf{Q}(t-1)\boldsymbol{\varphi}(t)}, \\ \mathbf{Q}(t) &= \frac{1}{\lambda} \left[\mathbf{I}_r - \mathbf{k}(t)\boldsymbol{\varphi}^T(t) \right] \mathbf{Q}(t-1).\end{aligned}\tag{6.11}$$

Zauważmy, że wektor wzmocnienia $\mathbf{k}(t)$ nie zależy od j , tj. jest taki sam dla obu kanałów. Wielkość $\hat{\boldsymbol{\rho}}(t)$ może być obliczana według następującej rekurencyjnej zależności, analogicznej do (3.58)

$$\hat{\boldsymbol{\rho}}(t) = \lambda\hat{\boldsymbol{\rho}}(t-1) + (1-\lambda)\boldsymbol{\varepsilon}(t|t-1)\boldsymbol{\varepsilon}^T(t|t-1)\boldsymbol{\varsigma}(t)\tag{6.12}$$

lub analogicznej do (3.61)

$$\hat{\boldsymbol{\rho}}(t) = \lambda_0\hat{\boldsymbol{\rho}}(t-1) + (1-\lambda_0)\boldsymbol{\varepsilon}(t|t-1)\boldsymbol{\varepsilon}^T(t|t-1),\tag{6.13}$$

gdzie $\boldsymbol{\varepsilon}(t|t-1) = [\varepsilon_1(t|t-1), \varepsilon_2(t|t-1)]^T$.

Związek z metodą największej wiarygodności

Załóżmy, że identyfikowany proces jest stacjonarny, tj. że podlega opisowi (6.4). Przy gausowskich założeniach statystycznie efektywne estymatory $\boldsymbol{\theta}$ i $\boldsymbol{\rho}$, przy dostępnej historii pomiarów $\mathcal{Y}(t) = \{\mathbf{y}(1), \dots, \mathbf{y}(t)\}$, mogą być otrzymane w postaci

$$\{\boldsymbol{\theta}^*(t), \boldsymbol{\rho}^*(t)\} = \arg \max_{\{\boldsymbol{\theta}, \boldsymbol{\rho}\}} p(\mathcal{Y}(t)|\mathcal{Y}_0, \boldsymbol{\theta}, \boldsymbol{\rho}),$$

gdzie $\mathcal{Y}_0 = \{\mathbf{y}(1-r), \dots, \mathbf{y}(0)\}$ oznacza zbiór warunków początkowych, zaś

$$\begin{aligned}p(\mathcal{Y}(t)|\mathcal{Y}_0, \boldsymbol{\theta}, \boldsymbol{\rho}) &= \prod_{k=1}^t p(\mathbf{y}(k)|\mathbf{y}(k-1), \dots, \mathbf{y}(1), \mathcal{Y}_0, \boldsymbol{\theta}, \boldsymbol{\rho}) = \\ &= (2\pi)^{-t} \{\det[\boldsymbol{\rho}]\}^{-t/2} \times \exp \left\{ -\frac{1}{2} \sum_{k=1}^t \|\mathbf{y}(k) - \boldsymbol{\Phi}^T(k)\boldsymbol{\theta}\|_{\boldsymbol{\rho}^{-1}}^2 \right\}\end{aligned}$$

oznacza tzw. warunkową funkcję wiarygodności. Warunkowe oceny największej wiarygodności (z ang. *conditional maximum likelihood*, CML) mogą być wyznaczone w sposób iteracyjny przy użyciu następujących zależności [89]

$$\begin{aligned}\boldsymbol{\theta}_{i+1}^*(t) &= \left\{ \sum_{k=1}^t \boldsymbol{\Phi}(k) [\boldsymbol{\rho}_i^*(t)]^{-1} \boldsymbol{\Phi}^T(k) \right\}^{-1} \left\{ \sum_{k=1}^t \boldsymbol{\Phi}(k) [\boldsymbol{\rho}_i^*(t)]^{-1} \mathbf{y}(k) \right\}, \\ \boldsymbol{\rho}_{i+1}^*(t) &= \frac{1}{t} \sum_{k=1}^t \left[\mathbf{y}(k) - \boldsymbol{\Phi}^T(k)\boldsymbol{\theta}_{i+1}^*(t) \right] \left[\mathbf{y}(k) - \boldsymbol{\Phi}^T(k)\boldsymbol{\theta}_{i+1}^*(t) \right]^T, \\ \boldsymbol{\theta}^*(t) &= \lim_{i \rightarrow \infty} \boldsymbol{\theta}_i^*(t), \quad \boldsymbol{\rho}^*(t) = \lim_{i \rightarrow \infty} \boldsymbol{\rho}_i^*(t).\end{aligned}$$

Kashyap and Rao [89] udowodnili, że gdy wszystkie równania wykorzystują ten sam wektor regresji (co ma miejsce w rozważanym przypadku), estymatory CML $\boldsymbol{\theta}^*(t)$ i $\boldsymbol{\rho}^*(t)$ pokrywają się z estymatorami LS (najmniejszych kwadratów) $\hat{\boldsymbol{\theta}}(t)$ i $\hat{\boldsymbol{\rho}}(t)$ otrzymanych z (6.6) po podstawieniu $\lambda = 1$ – patrz Twierdzenie 6a.1 w pracy [89]. Jest to interesujący wynik, ponieważ zgodnie z (6.8), estymator $\hat{\boldsymbol{\theta}}_j(t)$ może być otrzymany w wyniku analizy wyłącznie j -tego równania w (6.4), bez żadnego odniesienia do drugiego równania.

Gdy estymowane współczynniki zmieniają się wolno w czasie, mogą być śledzone przy użyciu wariantu estymatora CML z ograniczoną pamięcią, otrzymanego na drodze maksymalizacji następującej wykładniczo ważonej funkcji wiarygodności

$$\begin{aligned} \prod_{k=1}^t [p(\mathbf{y}(k)|\mathbf{y}(k-1), \dots, \mathbf{y}(1), \mathcal{Y}_0, \boldsymbol{\theta}, \boldsymbol{\rho})] \lambda^{t-k} &= \\ &= (2\pi)^{-l(t)} \{\det[\boldsymbol{\rho}]\}^{-l(t)/2} \times \exp \left\{ -\frac{1}{2} \sum_{k=1}^t \lambda^{t-k} \|\mathbf{y}(k) - \boldsymbol{\Phi}^T(k)\boldsymbol{\theta}\|_{\boldsymbol{\rho}^{-1}}^2 \right\}. \end{aligned}$$

Ponieważ dowód wspomnianej równości, podany w pracy [89], może być w łatwy sposób rozszerzony na wykładniczo ważony warunkowy estymator największej wiarygodności (z ang. *exponentially weighted conditional maximum likelihood*, EWCML), estymatory (6.8) i (6.9) mogą być traktowane jako estymatory EWCML.

Miarodajne porównanie modeli VAR i AR

Aby rzetelnie porównać wyniki detekcji/rekonstrukcji oparte na modelowaniu wektorowym z tymi otrzymanymi przy zastosowaniu modelowania skalarne, należy upewnić się, że w warunkach stacjonarnych i w stanie ustalonym algorytmy identyfikacji oparte na analizie wektorowej/skalarnej mają te same zdolności śledzące – w przeciwnym przypadku porównanie nie byłoby miarodajne [73].

Jako miarę zdolności śledzących algorytmu zastosować można wariancję nadmiarowych błędów predykcji. Rozważmy stacjonarny sygnał VAR podlegający opisowi (6.2). Zauważmy, że błędy jednokrokowej predykcji otrzymywane w trakcie śledzenia sygnału za pomocą algorytmu EWLS (6.11) mogą zostać zapisane w postaci

$$\varepsilon_j(t+1|t) = \eta_j(t+1|t) + n_j(t+1), \quad j = 1, 2,$$

gdzie

$$\eta_j(t+1|t) = \boldsymbol{\varphi}^T(t+1)[\boldsymbol{\theta}_j - \hat{\boldsymbol{\theta}}_j(t)]$$

oznacza tzw. nadmiarowy błąd predykcji, tj. składową błędu predykcji, której źródłem są wyłącznie błędy estymacji parametrów.

Gdy algorytm śledzący parametry ma „wystarczającą długą” pamięć estymacji, wariancja nadmiarowej części błędu predykcji może być obliczona w przybliżeniu przy użyciu techniki uśredniania – ponieważ zmiany w błędach estymacji parametrów $\hat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j$ są w rozważanym przypadku o wiele wolniejsze niż zmiany składowych wektora regresji $\boldsymbol{\varphi}(t)$, otrzymujemy $\overline{\eta_j^2(t+1|t)} \cong [\hat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j]^T \overline{\boldsymbol{\varphi}(t+1)\boldsymbol{\varphi}^T(t+1)} [\hat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j]$, gdzie $\overline{(\cdot)}$ oznacza lokalne uśrednianie po czasie. Prowadzi to do następującego przybliżenia

$$\begin{aligned} E[\eta_j^2(t+1|t)] &\cong E \left\{ [\hat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j]^T \boldsymbol{\Phi}_0 [\hat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j] \right\} = \\ &= \text{tr} \left[\text{cov}[\hat{\boldsymbol{\theta}}_j(t)] \boldsymbol{\Phi}_0 \right], \end{aligned}$$

gdzie $\Phi_0 = E[\varphi(t)\varphi^T(t)]$. Ponadto, ponieważ zachodzi zależność [73]

$$\lim_{t \rightarrow \infty} \text{cov}[\hat{\theta}_j(t)] \cong \frac{1 - \lambda}{1 + \lambda} \rho_j^2 \Phi_0^{-1},$$

wartość w stanie ustalonym dla $E[\eta_j^2(t+1|t)]$ może być wyrażona w postaci

$$\lim_{t \rightarrow \infty} E[\eta_j^2(t+1|t)] \cong \frac{2r(1 - \lambda)}{1 + \lambda} \rho_j^2. \quad (6.14)$$

W celu otrzymania analogicznego równania dla przypadku gdy oba kanały foniczne modelowane są oddzielnie, załóżmy, że $s_1(t)$ i $s_2(t)$ są stacjonarnymi sygnałami AR opisanymi w następujący sposób

$$\begin{aligned} s_j(t) &= \sum_{i=1}^r b_{ji} s_j(t-i) + n_j(t) = \\ &= \psi_j^T(t) \beta_j + n_j(t), \quad j = 1, 2, \end{aligned} \quad (6.15)$$

gdzie $\psi_j(t) = [s_j(t-1), \dots, s_j(t-r)]^T$ jest wektorem regresji, a $\beta_j = [b_{j1}, \dots, b_{jr}]^T$ oznacza wektor współczynników AR. Zauważmy, że model VAR (6.2) redukuje się do (6.15) jeżeli wszystkie elementy macierzy $\mathbf{A}_i, i = 1, \dots, r$, znajdujące się poza główną przekątną są równe zero. Załóżmy, że wektor parametrów β_j jest estymowany/śledzony przy użyciu skalarnej wersji algorytmu EWLS

$$\hat{\beta}_j(t) = \arg \min_{\beta} \sum_{k=1}^t \lambda_j^{t-k} [y_j(k) - \psi_j^T(t) \beta]^2, \quad (6.16)$$

gdzie $\lambda_j, 0 < \lambda_j < 1$, oznacza stałą zapominania użytą do celu identyfikacji j -tego kanału. Stosując technikę uśredniania można pokazać, że

$$\begin{aligned} E[\eta_j^2(t+1|t)] &\cong \text{tr} [\text{cov}[\hat{\beta}_j(t)] \Psi_j], \\ \lim_{t \rightarrow \infty} \text{cov}[\hat{\beta}_j(t)] &\cong \frac{1 - \lambda_j}{1 + \lambda_j} \rho_j^2 \Psi_j^{-1}, \end{aligned}$$

gdzie $\Psi_j = E[\psi_j(t)\psi_j^T(t)]$.

Prowadzi to do następującego wyrażenia

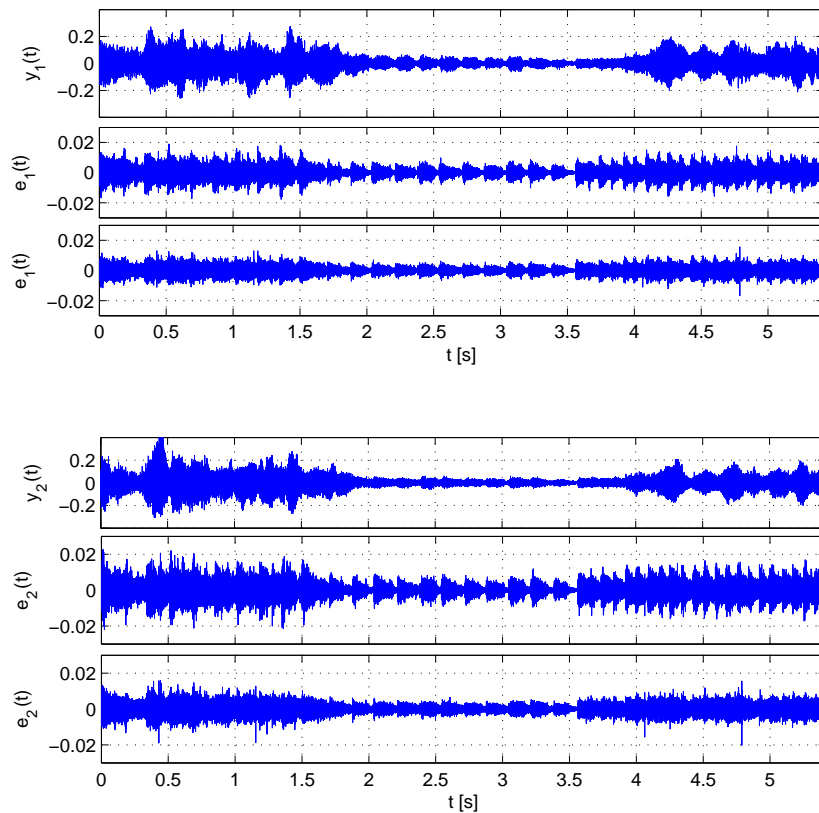
$$\lim_{t \rightarrow \infty} E[\eta_j^2(t+1|t)] \cong \frac{r(1 - \lambda_j)}{1 + \lambda_j} \rho_j^2, \quad (6.17)$$

które powinno być porównane z wyrażeniem (6.14). Zgodnie z przyjętym założeniem, o tym że wariancja nadmiarowych błędów predykcji powinna być w obu przypadkach taka sama, otrzymujemy następujący warunek „rzetelnego porównania”

$$\frac{2(1 - \lambda)}{1 + \lambda} = \frac{1 - \lambda_j}{1 + \lambda_j}. \quad (6.18)$$

Ponieważ stałe zapominania λ i λ_j są zazwyczaj bliskie jedności, tj. $1 + \lambda \cong 1 + \lambda_j \cong 2$, warunek (6.18) jest w przybliżeniu równoważny warunkowi

$$l_e^{\text{VAR}}(\infty) \cong 2l_e^{\text{AR}}(\infty), \quad (6.19)$$



Rysunek 6.2: Porównanie ciągów błędów jednokrokowej predykcji otrzymanych – dla typowego sygnału stereofonicznego (górnny wykres w obu grupach) – w przypadku skalarnego modelowania sygnałów (środkowy wykres w obu grupach) oraz w przypadku wektorowego modelowania sygnałów (dolny wykres w obu grupach). Górne grupa wykresów odpowiada lewemu kanałowi stereofonicznego, zaś dolna grupa – kanałowi prawemu. Warto odnotować różnicę skali pomiędzy górnym wykresem, a wykresami dolnymi (w każdej z grup).

gdzie $l_e^{\text{VAR}}(\infty)$ i $l_e^{\text{AR}}(\infty)$ oznaczają wartości otrzymywane w stanie ustalonym, reprezentujące efektywne długości pamięci estymatorów odpowiednio VAR i AR. Zauważmy, że ponieważ w przypadku wektorowym liczba estymowanych współczynników przypadających na jeden kanał równa jest $2r$, tj. jest dwa razy większa niż w analogicznym przypadku skalarnym, gdy spełniony jest warunek (6.19) średnia efektywna liczba próbek użyta do estymacji przypadająca na jeden współczynnik modelu jest w obu przypadkach taka sama.

Rysunek 6.2 porównuje ciągi błędów jednokrokowej predykcji otrzymane – dla typowego sygnału stereofonicznego – z użyciem skalarnego i wektorowego modelowania sygnału. W obu przypadkach parametry modelu były estymowane za pomocą algorytmu EWLS wyposażonego w różne stałe zapomnienia, dobrane w taki sposób aby zrównoważyć długości pamięci estymacji dla porównywanych podejść. Zauważmy, że łączna analiza lewego i prawego kanału pozwala na dokładniejsze modelowanie sygnału fonicznego – w rozważanym przypadku współczynnik redukcji wariancji błędów predykcji wyniósł 1,7 dla lewego kanału i 2,1 dla prawego kanału.

6.2 Detekcja zakłóceń impulsowych i interpolacja sygnału

Oceny EWLS otrzymane zostały przy założeniu, że $\boldsymbol{\delta}(t) \equiv \mathbf{0}$, tj. że w zarejestrowanym sygnale nie ma zakłóceń impulsowych. Aby estymator mógł działać w obecności zakłóceń impulsowych można zastosować analogiczną modyfikację jak w omówionym przypadku skalarnym. Oznaczmy przez $\mathbf{d}(t) = [d_1(t), d_2(t)]^T$ funkcję lokalizacji zakłóceń

$$d_j(t) = \begin{cases} 0, & \text{jeżeli } \delta_j(t) = 0, \\ 1, & \text{jeżeli } \delta_j(t) \neq 0, \end{cases} \quad j = 1, 2,$$

zaś przez $\hat{\mathbf{d}}(t) = [\hat{d}_1(t), \hat{d}_2(t)]^T$ – wyjście detektora zakłóceń impulsowych (który opisany zostanie w dalszej części rozdziału)

$$\hat{d}_j(t) = \begin{cases} 1, & \text{jeżeli zakłócenie zostało wykryte,} \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad j = 1, 2.$$

Aby uchronić oceny parametrów przed negatywnym wpływem zakłóceń impulsowych, estymacja parametrów modelu wstrzymywana jest za każdym razem gdy podniesiony zostanie alarm detekcyjny, tj. gdy $\hat{\mathbf{d}}(t) \neq \mathbf{0}$. Estymacja wznowiana jest gdy rekonstrukcja zakwestionowanego fragmentu zostanie zakończona (z wykorzystaniem interpolowanych próbek w miejsce tych uszkodzonych).

6.2.1 Opis problemu w przestrzeni stanów

Podobnie jak dla przypadku skalarnego (patrz punkt 3.2.2) zaczniemy od rozwiązania problemu rekonstrukcji odizolowanego bloku m nieodwracalnie uszkodzonych próbek stacjonarnego procesu AR opisanego równaniem (6.2). Załóżmy, że blok uszkodzonych próbek zaczyna się w chwili $t_0 + 1$, a kończy w chwili $t_0 + m$ (tj. $\mathbf{d}(t_0 + 1) = \dots = \mathbf{d}(t_0 + m) = \mathbf{1}$, gdzie $\mathbf{1} = [1, 1]^T$). Dodatkowo przyjmijmy, że pozostałe próbki sygnału są nieuszkodzone (tj. $\mathbf{d}(t) = \mathbf{0}$ dla $t \leq t_0$ i $t > t_0 + m$). Załóżmy, że lokalizacja sekwencji zakłóceń impulsowych jest dokładnie znana [tj. $\hat{\mathbf{d}}(t) \equiv \mathbf{d}(t)$]. Ponadto załóżmy, że zakłócenia impulsowe $\boldsymbol{\delta}(t_0 + 1), \dots, \boldsymbol{\delta}(t_0 + m)$ mogą być modelowane jako ciąg wewnętrznie nieskorelowanych zmiennych gaussowskich, niezależnych od $\{\mathbf{n}(t)\}$, ze znanymi macierzami kowariancji

$$\boldsymbol{\Delta}(t) = \text{cov}[\boldsymbol{\delta}(t)], \quad t_0 + 1 \leq t \leq t_0 + m.$$

Aby zaprojektować filtr Kalmana musimy wprowadzić opis w przestrzeni stanów równoważny opisowi wejściowo-wyjściowemu (6.1)-(6.2). Niech $q = 2r + m$. Zdefiniujmy wektor stanu $\mathbf{x}_q(t) = [\mathbf{s}^T(t), \dots, \mathbf{s}^T(t - q + 1)]^T$ o wymiarach $2q \times 1$ złożony łącznie z ostatnich q próbek sygnału $s_1(t)$ i $s_2(t)$. Model stanowy równoważny modelowi (6.1)-(6.2) może być zapisany w postaci

$$\begin{aligned} \mathbf{x}_q(t+1) &= \mathbf{A}_q \mathbf{x}_q(t) + \mathbf{C}_q \mathbf{n}(t+1), \\ \mathbf{y}(t) &= \mathbf{C}_q^T \mathbf{x}_q(t) + \boldsymbol{\delta}(t), \end{aligned} \quad (6.20)$$

gdzie

$$\mathbf{A}_q = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_r & \mathbf{O} & \dots & \mathbf{O} & \mathbf{O} \\ \mathbf{I} & \mathbf{O} & \dots & \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} & & \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} & \mathbf{O} \\ \vdots & & & & & \ddots & & \vdots \\ \mathbf{O} & \mathbf{O} & & \mathbf{O} & \mathbf{O} & \dots & \mathbf{I} & \mathbf{O} \end{bmatrix}$$

jest blokową macierzą przejścia o wymiarach $2q \times 2q$, a $\mathbf{C}_q = [\mathbf{I}, \mathbf{O}, \dots, \mathbf{O}]^T$ oznacza blokową macierz wyjścia o wymiarach $2q \times 2$, zaś $\mathbf{O} = \mathbf{O}_1$ i $\mathbf{I} = \mathbf{I}_1$ oznaczają odpowiednio macierz zerową i macierz jednostkową o wymiarach 2×2 .

Na podstawie (6.20) i dostępnej wiedzy wstępnej o sygnale filtr Kalmana może zostać zapisany w formie rekursywnej

$$\begin{aligned} \hat{\mathbf{x}}_q(t|t-1) &= \mathbf{A}_q \hat{\mathbf{x}}_q(t-1|t-1), \\ \mathbf{P}_q(t|t-1) &= \mathbf{A}_q \mathbf{P}_q(t-1|t-1) \mathbf{A}_q^T + \mathbf{C}_q \boldsymbol{\rho} \mathbf{C}_q^T, \\ \mathbf{e}(t) &= \mathbf{y}(t) - \mathbf{C}_q^T \hat{\mathbf{x}}_q(t|t-1), \\ \mathbf{G}(t) &= \mathbf{C}_q^T \mathbf{P}_q(t|t-1) \mathbf{C}_q + \boldsymbol{\Delta}(t), \\ \mathbf{L}_q(t) &= \mathbf{P}_q(t|t-1) \mathbf{C}_q \mathbf{G}^{-1}(t), \\ \hat{\mathbf{x}}_q(t|t) &= \hat{\mathbf{x}}_q(t|t-1) + \mathbf{L}_q(t) \mathbf{e}(t), \\ \mathbf{P}_q(t|t) &= \mathbf{P}_q(t|t-1) - \mathbf{L}_q(t) \mathbf{G}(t) \mathbf{L}_q^T(t). \end{aligned} \quad (6.21)$$

Ponieważ założyliśmy, że $\boldsymbol{\delta}(t) = \mathbf{0}$ dla $t \leq t_0$, algorytm powinien rozpocząć swoją pracę w chwili $t_0 + 1$ z warunkami początkowymi $\hat{\mathbf{x}}_q(t_0|t_0) = [\mathbf{y}^T(t_0), \dots, \mathbf{y}^T(t_0 - q + 1)]^T$ oraz $\mathbf{P}_q(t_0|t_0) = \mathbf{O}_q$, i zakończyć ją w chwili $t_0 + m + r$, po zaakceptowaniu r nieuszkodzonych próbek sygnału zlokalizowanych bezpośrednio za zakwestionowanym fragmentem. Filtracyjna ocena wektora stanu w chwili zakończenia filtracji ma postać $\hat{\mathbf{x}}_q(t_0 + m + r|t_0 + m + r) = [\mathbf{y}^T(t_0 + m + r), \dots, \mathbf{y}^T(t_0 + m + 1), \hat{\mathbf{s}}^T(t_0 + m), \dots, \hat{\mathbf{s}}^T(t_0 + 1), \mathbf{y}^T(t_0), \dots, \mathbf{y}^T(t_0 - r + 1)]^T$, gdzie $\hat{\mathbf{s}}^T(t_0 + 1), \dots, \hat{\mathbf{s}}^T(t_0 + m)$ jest blokiem interpolowanych próbek.

Na rysunku 6.3 pokazano wyniki rekonstrukcji sygnału opartych na modelach AR i VAR. W rozważanej sytuacji oba kanały nagrania stereofonicznego są silnie skorelowane, a uszkodzeniu uległ fragment sygnału tylko w lewym kanale. W pierwszym przypadku otrzymana rekonstrukcja sygnału oparta na modelu AR jest niezadowolająca – nie przypomina kształtem fragmentu z drugiego, nieuszkodzonego kanału. Gdy do rekonstrukcji brakującego bloku próbek w pierwszym kanale wykorzystano również informacje pochodzące z drugiego kanału, otrzymana rekonstrukcja przypomina już w dużym stopniu wzorcowy fragment z kanału drugiego.

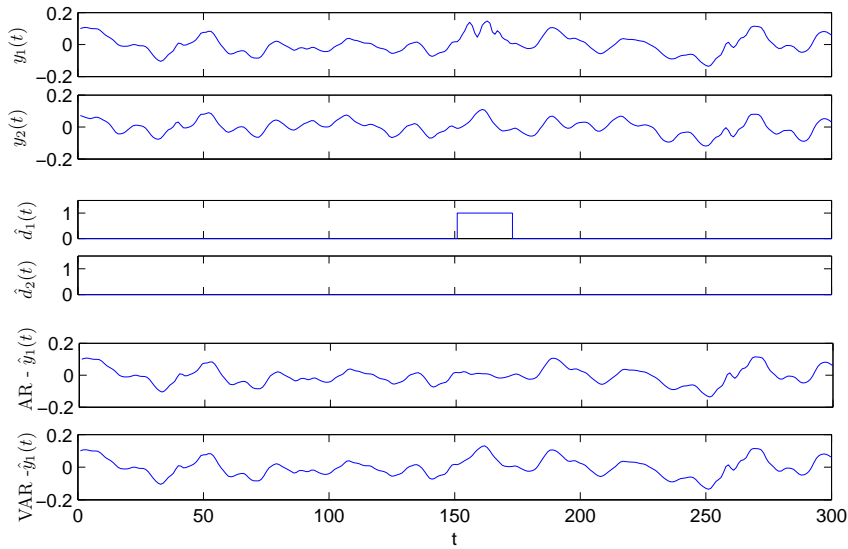
6.2.2 Predykcja sygnału i detekcja zakłóceń impulsowych

Technika detekcji opisana w punkcie 3.4.2 może być rozszerzona na przypadek wielu zmiennych. Warunek zgłoszenia alarmu (3.76) ma następujący wielowymiarowy odpowiednik

$$\boldsymbol{\varepsilon}^T(t_0 + 1|t_0) \boldsymbol{\Sigma}_\varepsilon^{-1}(t_0 + 1|t_0) \boldsymbol{\varepsilon}(t_0 + 1|t_0) > \mu^2, \quad (6.22)$$

gdzie $\boldsymbol{\Sigma}_\varepsilon(t_0 + 1|t_0) = \boldsymbol{\rho}$ oznacza macierz kowariancji błędów jednokrokowej predykcji. Warunek zakończenia detekcji (3.78) może zostać przeformułowany w analogiczny sposób

$$\begin{aligned} \boldsymbol{\varepsilon}^T(t|t_0) \boldsymbol{\Sigma}_\varepsilon^{-1}(t|t_0) \boldsymbol{\varepsilon}(t|t_0) &\leq \mu^2, \\ t &= t_0 + m + 1, \dots, t_0 + m + r, \end{aligned} \quad (6.23)$$



Rysunek 6.3: Porównanie wyników rekonstrukcji sygnału opartych na modelach AR i VAR. Dwa górne wykresy pokazują fragment archiwalnego nagrania stereofonicznego z uszkodzoną lewą ścieżką dźwiękową $y_1(t)$. Kolejne dwa wykresy pokazują decyzje detektora zakłóceń impulsowych $\hat{d}_1(t)/\hat{d}_2(t)$. Dwa dolne wykresy pokazują wyniki rekonstrukcji lewego sygnału $y_1(t)$ otrzymane na podstawie modelu skalarnego [AR- $\hat{y}_1(t)$] oraz na podstawie modelu wektorowego [VAR- $\hat{y}_1(t)$].

gdzie $\varepsilon(t|t_0)$ oznacza $(t-t_0)$ -krokowy błąd predykcji sygnału, zaś $\Sigma_\varepsilon(t|t_0)$ oznacza odpowiadającą mu macierz kowariancji błędu predykcji. Obie wielkości mogą zostać w łatwy sposób obliczone przy użyciu filtru Kalmana (6.21). Aby zrealizować to zadanie, należy przyjąć $q = 2r + m$ i wymusić

$$\Delta(t) = \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}, \quad \gamma \rightarrow \infty, \quad \text{dla } t > t_0. \quad (6.24)$$

Warunek (6.24) oznacza, że próbki $\mathbf{y}(t_0 + 1)$, $\mathbf{y}(t_0 + 2)$, ... powinny być traktowane jako uszkodzone przez zakłócenie o nieskończonej wariancji co jest równoznaczne z ich całkowitą eliminacją z procesu estymacji. Ponieważ w takich przypadkach $\mathbf{G}^{-1}(t) = \mathbf{O}$, otrzymujemy

$$\begin{aligned} \hat{\mathbf{x}}_q(t|t) &= \hat{\mathbf{x}}_q(t|t-1), \\ \mathbf{P}_q(t|t) &= \mathbf{P}_q(t|t-1), \end{aligned} \quad (6.25)$$

dla wszystkich $t > t_0$. W takich warunkach filtr Kalmana staje się wielokrotnym predyktorem generującym $\varepsilon(t|t_0) = \mathbf{e}(t)$ i

$$\begin{aligned} \Sigma_\varepsilon(t|t_0) &= \text{cov}[\mathbf{e}(t)] = \mathbf{C}_q^T \mathbf{P}_q(t|t-1) \mathbf{C}_q = \\ &= \begin{bmatrix} \sigma_1^2(t) & \sigma_{12}(t) \\ \sigma_{12}(t) & \sigma_2^2(t) \end{bmatrix} = \Sigma(t). \end{aligned} \quad (6.26)$$

Niestety powyższe podejście nie pozwala na rozróżnienie pomiędzy poszczególnymi kanałami nagrania stereofonicznego (oba kanały analizowane są łącznie) i dlatego nie jest odpowiednie

dla naszych celów. Zostanie ono zastąpione następującą regułą decyzyjną

$$\hat{d}_{0,j}(t) = \begin{cases} 0, & \text{jeżeli } |e_j(t)| \leq \mu\sigma_j(t), \\ 1, & \text{jeżeli } |e_j(t)| > \mu\sigma_j(t), \end{cases} \quad j = 1, 2, \quad (6.27)$$

gdzie $\hat{d}_{0,j}(t)$, $j = 1, 2$, oznaczają tymczasowe decyzje podjęte przez detektor, oraz bardziej selektywnie dobieraną macierz kowariancji błędu predykcji

$$\Delta(t) = \begin{cases} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, & \text{jeżeli } \hat{d}_{0,1}(t) = 0 \wedge \hat{d}_{0,2}(t) = 0, \\ \begin{bmatrix} 0 & 0 \\ 0 & \gamma \end{bmatrix}, & \text{jeżeli } \hat{d}_{0,1}(t) = 0 \wedge \hat{d}_{0,2}(t) = 1, \\ \begin{bmatrix} \gamma & 0 \\ 0 & 0 \end{bmatrix}, & \text{jeżeli } \hat{d}_{0,1}(t) = 1 \wedge \hat{d}_{0,2}(t) = 0, \\ \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}, & \text{jeżeli } \hat{d}_{0,1}(t) = 1 \wedge \hat{d}_{0,2}(t) = 1, \end{cases} \quad (6.28)$$

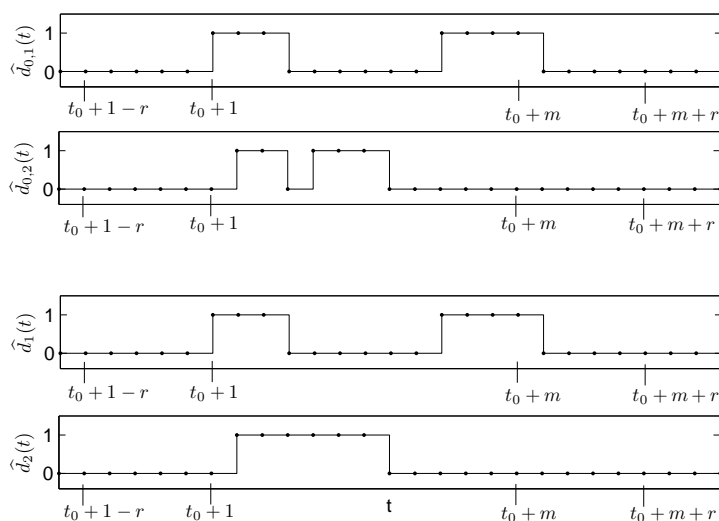
$\gamma \rightarrow \infty$.

Łatwo sprawdzić, że w przypadku przyjęcia (6.28) odpowiednie wartości macierzy $\mathbf{G}^{-1}(t)$ mają postać

$$\mathbf{G}^{-1}(t) = \begin{cases} \Sigma^{-1}(t), & \text{jeżeli } \hat{d}_{0,1}(t) = 0 \wedge \hat{d}_{0,2}(t) = 0, \\ \begin{bmatrix} \frac{1}{\sigma_1^2(t)} & 0 \\ 0 & 0 \end{bmatrix}, & \text{jeżeli } \hat{d}_{0,1}(t) = 0 \wedge \hat{d}_{0,2}(t) = 1, \\ \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\sigma_2^2(t)} \end{bmatrix}, & \text{jeżeli } \hat{d}_{0,1}(t) = 1 \wedge \hat{d}_{0,2}(t) = 0, \\ \mathbf{O}, & \text{jeżeli } \hat{d}_{0,1}(t) = 1 \wedge \hat{d}_{0,2}(t) = 1. \end{cases} \quad (6.29)$$

Pozwala to na: akceptację obu wielkości $\mathbf{y}(t) = [y_1(t), y_2(t)]^T$ jeżeli próbki pochodzące z obu kanałów traktowane są jako niezakłócone [$\hat{d}_{0,1}(t) = \hat{d}_{0,2}(t) = 0$], odrzucenie $y_2(t)$ jeżeli tylko drugi kanał jest uszkodzony [$\hat{d}_{0,1}(t) = 0 \wedge \hat{d}_{0,2}(t) = 1$], odrzucenie $y_1(t)$ jeżeli tylko pierwszy kanał jest uszkodzony [$\hat{d}_{0,1}(t) = 1 \wedge \hat{d}_{0,2}(t) = 0$] lub odrzucenie obu składników $\mathbf{y}(t)$ jeżeli oba kanały są uszkodzone [$\hat{d}_{0,1}(t) = \hat{d}_{0,2}(t) = 1$]. Zwróćmy uwagę na to, że aż w trzech rozważanych sytuacjach predykcja sygnału wykorzystuje informacje zawarte w sygnale $y_1(t)$ i/lub $y_2(t)$, co sprawia że jest ona dokładniejsza niż w przypadku skalarnym.

Przeanalizujemy pojedynczy alarm detekcyjny, rozpoczynający się gdy przynajmniej jeden z dwóch obliczonych błędów predykcji dla sygnału stereofonicznego przyjmuje zbyt dużą wartość, i kończy się gdy r kolejnych błędów predykcji przyjmuje w obu kanałach jednocześnie wystarczająco małe wartości. Po spełnieniu warunku końca alarmu, tymczasowe decyzje podjęte przez detektory, $\hat{d}_{0,1}(t)$ i $\hat{d}_{0,2}(t)$, podlegają sprawdzeniu niezależnie dla lewego i prawego kanału. W przypadku gdy odległość pomiędzy dwoma sąsiadującymi blokami zakwestionowanych próbek jest mniejsza niż r próbek, sąsiadujące bloki należy scalić – patrz rysunek 6.4. Końcowe decyzje detektora oznaczane będą dalej przez $\hat{d}_1(t)$ i $\hat{d}_2(t)$, odpowiednio dla lewego i prawego kanału. Jeżeli końcowe decyzje detektora różnią się od jego tymczasowych decyzji, filtr Kalmana musi być ponownie uruchomiony, aby uzyskać nową interpolację odtwarzanych próbek sygnału. Alternatywą może być użycie do tego celu nierekurencyjnej procedury opisanej w pracy [71].



Rysunek 6.4: Przykładowe sygnały detekcyjne generowane przez detektor oparty na modelu VAR. Dwa górne wykresy ilustrują tymczasowe decyzje podjęte przez detektor, odpowiednio dla lewego i prawego kanału. Dwa dolne wykresy ilustrują końcowe decyzje detektora, odpowiednio dla lewego i prawego kanału.

Zgodnie z tą procedurą, rekonstrukcja próbek zakwestionowanych przez detektor, oparta na modelu VAR, może być przeprowadzona niezależnie – bez utraty jakiegokolwiek informacji – dla każdego lokalnego okna analizy obejmującego r próbek „przed” i r próbek „po” zakwestionowanym fragmencie sygnału $\mathbf{y}(t)$. Zauważmy jeszcze na koniec, że wszystkie próbki sygnału $\mathbf{y}(t)$ z przedziału $t \in [t_0 + 1 - r, t_0 + m + r]$, które nie zostały zakwestionowane na etapie detekcji, biorą udział w rekonstrukcji oznaczonych bloków próbek wewnątrz tego przedziału.

6.2.3 Algorytm filtracji Kalmana o zmiennym rzędzie oparty na modelu VAR

Analogicznie jak miało to miejsce dla przypadku skalarne (patrz punkt 3.2.3), filtr Kalmana o stałym rzędzie zastąpiony zostanie jego wariantem o zmiennym rzędzie. Modyfikacja ta jest możliwa ze względu na szczególną strukturę macierzy \mathbf{A}_q , \mathbf{C}_q i $\mathbf{P}_q(t_0|t_0)$ w (6.21). Z uwagi na tę strukturę rząd filtru Kalmana (6.21) może być – bez wpływu na wyniki estymacji – stopniowo rozszerzany, w miarę trwania detekcji poczynając od rzędu $r + 1$.

Oznaczmy symbolem $\Theta_r = [\theta_1 | \theta_2]$ macierz złożoną z współczynników autoregresji (o wymiarach $2r \times 2$), natomiast symbolem

$$\Theta_q = \begin{bmatrix} \theta_1 & \theta_2 \\ \mathbf{0}_{q-r} & \mathbf{0}_{q-r} \end{bmatrix}, \quad q > r$$

– analogiczną macierz rozszerzoną poprzez dodanie elementów zerowych. Adaptacyjny algorytm o zmiennym rzędzie, który łączy (6.21) z (6.27)-(6.29), rozpoczyna pracę w chwili $t_0 + 1$ z wa-

runkami początkowymi

$$\begin{aligned}\hat{\mathbf{x}}_r(t_0|t_0) &= [\mathbf{y}^T(t_0), \dots, \mathbf{y}^T(t_0 - r + 1)]^T, \\ \mathbf{P}_r(t_0|t_0) &= \mathbf{O}_r.\end{aligned}\quad (6.30)$$

W pierwszym kroku, po aktualizacji czasu ($t \geq t_0 + 1$), wyznaczamy wielkości o powiększonym rzędzie: ocenę wektora stanu *a priori* $\hat{\mathbf{x}}_{q(t)}(t|t-1)$ oraz macierz kowariancji *a priori* $\mathbf{P}_{q(t)}(t|t-1)$:

$$\begin{aligned}\hat{\mathbf{y}}(t|t-1) &= \Theta_{q(t)-1}^T \hat{\mathbf{x}}_{q(t)-1}(t-1|t-1), \\ \mathbf{e}(t) &= \mathbf{y}(t) - \hat{\mathbf{y}}(t|t-1) = [e_1(t), e_2(t)]^T, \\ \hat{\mathbf{x}}_{q(t)}(t|t-1) &= \begin{bmatrix} \hat{\mathbf{y}}(t|t-1), \\ \hat{\mathbf{x}}_{q(t)-1}(t-1|t-1) \end{bmatrix}, \\ \mathbf{H}_{q(t)-1}(t) &= \mathbf{P}_{q(t)-1}(t-1|t-1) \Theta_{q(t)-1}, \\ \Sigma(t) &= \Theta_{q(t)-1}^T \mathbf{H}_{q(t)-1}(t) + \boldsymbol{\rho} = \\ &= \begin{bmatrix} \sigma_1^2(t) & \sigma_{12}(t) \\ \sigma_{12}(t) & \sigma_2^2(t) \end{bmatrix}, \\ \mathbf{P}_{q(t)}(t|t-1) &= \begin{bmatrix} \Sigma(t) & \mathbf{H}_{q(t)-1}^T(t) \\ \mathbf{H}_{q(t)-1}(t) & \mathbf{P}_{q(t)-1}(t-1|t-1) \end{bmatrix}.\end{aligned}\quad (6.31)$$

Zwróćmy uwagę na specjalną budowę macierzy kowariancji *a priori*, która powstaje w wyniku „obsunięcia” macierzy kowariancji *a posteriori* $\mathbf{P}_{q(t)-1}(t-1|t-1)$ oraz dodania pierwszych dwóch kolumn i pierwszych dwóch wierszy [lewy górny blok o wymiarach 2×2 tej macierzy odpowiada macierzy kowariancji błędu predykcji $\Sigma(t)$].

W kolejnym kroku podejmowane są niezależnie tymczasowe decyzje o akceptacji lub ignorowaniu informacji zawartych w pomiarach $y_1(t)/y_2(t)$

$$\hat{d}_{0,j}(t) = \begin{cases} 0, & \text{jeżeli } |e_j(t)| \leq \mu\sigma_j(t), \\ 1, & \text{jeżeli } |e_j(t)| > \mu\sigma_j(t), \end{cases} \quad j = 1, 2. \quad (6.32)$$

Decyzje te będą miały istotny wpływ na oceny wektora stanu *a posteriori* $\hat{\mathbf{x}}_{q(t)}(t|t)$ i macierzy kowariancji *a posteriori* $\mathbf{P}_{q(t)}(t|t)$. W przypadku akceptacji obu pomiarów $\hat{d}_{0,1}(t) = \hat{d}_{0,2}(t) = 0$ lub osiągnięcia maksymalnej długości alarmu detekcyjnego ($t \geq t_0 + m_{\max}$), aktualizacja przebiega następująco

$$\begin{aligned}\mathbf{L}_{q(t)}(t) &= \mathbf{P}_{q(t)}^{(1,2)}(t|t-1) \Sigma^{-1}(t), \\ \hat{\mathbf{x}}_{q(t)}(t|t) &= \hat{\mathbf{x}}_{q(t)}(t|t-1) + \mathbf{L}_{q(t)}(t) \mathbf{e}(t), \\ \mathbf{P}_{q(t)}(t|t) &= \mathbf{P}_{q(t)}(t|t-1) - \mathbf{L}_{q(t)}(t) \Sigma(t) \mathbf{L}_{q(t)}^T(t),\end{aligned}\quad (6.33)$$

gdzie $\mathbf{P}_{q(t)}^{(1,2)}(t|t-1)$ oznacza macierz złożoną z pierwszych dwóch kolumn macierzy kowariancji *a priori*. Natomiast w przypadku zakwestionowania tylko jednego z pomiarów możliwe są dwie sytuacje. Jeśli $\hat{d}_{0,1}(t) = 0$ i $\hat{d}_{0,2}(t) = 1$, tj. zakłócenie zostaje wykryte tylko w prawym kanale, wektor wzmocnień $\mathbf{L}_{q(t)}(t)$ obliczany jest z wykorzystaniem pierwszej kolumny macierzy

a priori $\mathbf{P}_{q(t)}^{(1)}(t|t-1)$

$$\begin{aligned} \mathbf{l}_{q(t)}(t) &= \frac{1}{\sigma_1^2(t)} \mathbf{P}_{q(t)}^{(1)}(t|t-1), \\ \hat{\mathbf{x}}_{q(t)}(t|t) &= \hat{\mathbf{x}}_{q(t)}(t|t-1) + \mathbf{l}_{q(t)}(t)e_1(t), \\ \mathbf{P}_{q(t)}(t|t) &= \mathbf{P}_{q(t)}(t|t-1) - \sigma_1^2(t)\mathbf{l}_{q(t)}(t)\mathbf{l}_{q(t)}^T(t). \end{aligned} \quad (6.34)$$

Podobnie, jeśli $\hat{d}_{0,1}(t) = 1$ i $\hat{d}_{0,2}(t) = 0$, tj. zakłócenie zostaje wykryte wyłącznie w lewym kanale, do aktualizacji wektora wzmocnień $\mathbf{l}_{q(t)}(t)$ wykorzystywana jest druga kolumna macierzy *a priori* $\mathbf{P}_{q(t)}^{(2)}(t|t-1)$

$$\begin{aligned} \mathbf{l}_{q(t)}(t) &= \frac{1}{\sigma_2^2(t)} \mathbf{P}_{q(t)}^{(2)}(t|t-1), \\ \hat{\mathbf{x}}_{q(t)}(t|t) &= \hat{\mathbf{x}}_{q(t)}(t|t-1) + \mathbf{l}_{q(t)}(t)e_2(t), \\ \mathbf{P}_{q(t)}(t|t) &= \mathbf{P}_{q(t)}(t|t-1) - \sigma_2^2(t)\mathbf{l}_{q(t)}(t)\mathbf{l}_{q(t)}^T(t). \end{aligned} \quad (6.35)$$

Wreszcie, w przypadku zakwestionowania obu pomiarów $\hat{d}_{0,1}(t) = \hat{d}_{0,2}(t) = 1$, aktualizacja polega na przepisaniu informacji *a priori*

$$\begin{aligned} \hat{\mathbf{x}}_{q(t)}(t|t) &= \hat{\mathbf{x}}_{q(t)}(t|t-1), \\ \mathbf{P}_{q(t)}(t|t) &= \mathbf{P}_{q(t)}(t|t-1). \end{aligned} \quad (6.36)$$

Algorytm przerywa pracę w chwili $t_0 + m + r$ po spełnieniu warunku końca alarmu detekcyjnego (6.23). Algorytm ten podlega tym samym zabezpieczeniom numerycznym, o których była mowa dla przypadku skarlanego – patrz punkt 3.2.3.

Adaptacyjne wersje procedur detekcji i interpolacji opisanych powyżej mogą być otrzymane przez zastąpienie prawdziwych parametrów modelu $\boldsymbol{\theta}$ i $\boldsymbol{\rho}$, które wcześniej traktowane były jako znane, ich najbardziej aktualnymi oszacowaniami $\hat{\boldsymbol{\theta}}(t)$ i $\hat{\boldsymbol{\rho}}(t)$.

Z tych samych powodów co w przypadku skalarnym, do celów detekcji używany będzie niesprężony z algorytmem EWLS estymator $\hat{\boldsymbol{\rho}}(t)$ (6.13) – patrz uwaga 3.2. Mimo iż aplikacje ze zmodyfikowanym estymatorem kowariancji (6.13) dają lepsze wyniki na etapie detekcji, to już na etapie interpolacji sygnału lepsze wyniki osiągane są gdy stosowany jest oryginalny estymator EWLS (6.12) – w przypadku skalarnym analogiczna wielkość nie wpływała na wyniki interpolacji. Dlatego, aby osiągnąć najlepszy efekt, algorytm filtru Kalmana uruchamiany jest dwukrotnie: pierwszy przebieg ma na celu wykrycie zakłócenia impulsowego przy użyciu estymaty (6.13), zaś drugi – aby zrekonstruować nieodwracalnie uszkodzone próbki z użyciem estymaty (6.12).

6.3 Kwestie implementacyjne

Implementacja opisanego powyżej algorytmu wymaga rozwiązania dwóch kwestii: eliminacji składowych stałych oraz monitorowania i wymuszania stabilności modelu VAR w wymaganych sytuacjach.

6.3.1 Eliminacja składowych stałych

Podobnie jak miało to miejsce dla skalarnego modelu AR, stosując model VAR (6.2), zakładamy, że modelowany sygnał ma zerową wartość oczekiwaną: $E[\mathbf{s}(t)] = \mathbf{0}$. Niestety, założenie

to nie jest spełnione dla typowego nagrania stereofonicznego. Problem niezerowych składowych stałych obecnych w lewym/prawym kanale może zostać rozwiązany podobnie jak dla przypadku skalarne.

Bezpośrednie rozwiązanie tego problemu polega na uwzględnieniu offsetów w modelu VAR, tj. na użyciu następującego opisu w miejsce (6.2)

$$\mathbf{s}(t) = \sum_{i=1}^r \mathbf{A}_i \mathbf{s}(t-i) + \bar{\mathbf{s}} + \mathbf{n}(t), \quad (6.37)$$

gdzie $\bar{\mathbf{s}} = [\bar{s}_1, \bar{s}_2]^T$, a \bar{s}_1 i \bar{s}_2 oznaczają nieznane stałe. Ponieważ po uwzględnieniu dodatkowych elementów w wektorach $\boldsymbol{\theta}_j = [\boldsymbol{\alpha}_{j1}^T, \dots, \boldsymbol{\alpha}_{jr}^T, \bar{s}_j]^T$ i $\boldsymbol{\varphi}(t) = [\mathbf{y}^T(t-1), \dots, \mathbf{y}^T(t-r), 1]^T$ skrócony opis sygnału (6.4) pozostaje niezmienny, identyfikacji rozszerzonego modelu VAR (6.37) może dokonać w dokładnie ten sam sposób jak to opisano wcześniej. Z tego samego powodu algorytm filtru Kalmana o zmiennym rzędzie nie wymaga żadnej modyfikacji.

Pośrednie rozwiązanie problemu offsetu, które było zalecane dla przypadku skalarne i jest również zalecane dla przypadku wektorowego, polega na usunięciu niezerowej składowej stałej przed zastosowaniem procedury detekcji/interpolacji wykorzystując filtr górnoprzepustowy postaci (3.71). Takie rozwiązanie zapewni większą elastyczność w doborze parametrów λ i k_c .

6.3.2 Monitorowanie stabilności modelu VAR

Dla skalarne modelu AR zaproponowane zostało podejście polegające na monitorowaniu jego stabilności i w przypadku stwierdzenia jej braku, ponownej estymacji parametrów modelu przy użyciu algorytmu Levinsona-Durbina – patrz punkt 3.3.5. Analogiczna procedura stosowana będzie dla modelu VAR. Model VAR (6.2) jest stabilny asymptotycznie wtedy i tylko wtedy gdy wszystkie zera $z_i, i = 1, \dots, 2r$, wielomianu charakterystycznego $\mathcal{A}(z^{-1})$,

$$\mathcal{A}(z^{-1}) = \det \left[\mathbf{I} - \sum_{i=1}^r \mathbf{A}_i z^{-i} \right],$$

leżą wewnątrz okręgu jednostkowego na płaszczyźnie zespolonej: $|z_i| < 1, i = 1, \dots, 2r$. Za każdym razem gdy podniesiony zostanie alarm detekcyjny, sprawdzona jest stabilność modelu VAR. Jeżeli warunek stabilności nie jest spełniony, współczynniki modelu estymowane są ponownie przy użyciu algorytmu Whittlea-Wigginsa-Robinsona (WWR) [92], [93] gwarantującego stabilność modelu (wielowymiarowe rozszerzenie algorytmu Levinsona-Durbina) – szczegółowy opis tego algorytmu i dyskusję jego właściwości znaleźć można w Uzupełnieniu C8.6 w monografii [75]. Algorytm WWR estymuje $\hat{\mathbf{A}}_1(t), \dots, \hat{\mathbf{A}}_r(t)$ i $\hat{\boldsymbol{\rho}}(t)$ rozwiązując poniższy układ równań Yule'a-Walkera

$$[\mathbf{I}, -\hat{\mathbf{A}}_1(t), \dots, -\hat{\mathbf{A}}_r(t)] \hat{\boldsymbol{\mathcal{R}}}(t) = [\hat{\boldsymbol{\rho}}(t), \mathbf{O}, \dots, \mathbf{O}], \quad (6.38)$$

gdzie

$$\hat{\boldsymbol{\mathcal{R}}}(t) = \begin{bmatrix} \hat{\mathbf{R}}_0(t) & \hat{\mathbf{R}}_1(t) & \dots & \hat{\mathbf{R}}_r(t) \\ \hat{\mathbf{R}}_1^T(t) & \hat{\mathbf{R}}_0(t) & & \vdots \\ \vdots & & & \hat{\mathbf{R}}_1(t) \\ \hat{\mathbf{R}}_r^T(t) & \dots & & \hat{\mathbf{R}}_0(t) \end{bmatrix},$$

natomiast $\widehat{\mathbf{R}}_k(t), k = 0, \dots, r$, oznacza lokalną ocenę macierzy autokowariancji

$$\mathbf{R}_k = E[\mathbf{y}(t)\mathbf{y}^T(t-k)] \quad (6.39)$$

otrzymywaną w następujący sposób

$$\widehat{\mathbf{R}}_k(t) = \frac{1}{N} \sum_{i=0}^{N-k-1} \mathbf{y}(t-i)\mathbf{y}^T(t-i-k). \quad (6.40)$$

Ważną właściwością algorytmu WWR jest to, że gwarantuje on stabilność otrzymywanego modelu VAR pod warunkiem, że macierz $\widehat{\mathbf{R}}(t)$ jest dodatnio określona [co jest zapewnione gdy używane są obciążone estymaty (6.40)].

Z tych samych powodów co w przypadku skalarnym, algorytm WWR jest traktowany jako „ratunkowa” procedura estymacji, używana *tylko* wtedy gdy model EWLS jest niestabilny w momencie zgłoszenia alarmu detekcyjnego.

6.4 Rozszerzenie na analizę dwukierunkową

Omówione w tym rozdziale podejście wektorowej jednostronnej analizy sygnału fonicznego może zostać w łatwy sposób rozszerzone na analizę wektorową dwukierunkową – patrz rozdział 4. Łączenie lewostronnych i prawostronnych sygnałów detekcyjnych przebiega niezależnie dla lewego/prawego kanału.

Na początku, lewostronne i prawostronne alarmy detekcyjne zostają rozszerzone o małą, stałą liczbę próbek Δ_1 , tj. dla lewostronnego alarmu przesuwane jest jego przednie zbocze, zaś dla prawostronnego alarmu – jego tylne zbocze, zgodnie z (4.14). Zmodyfikowane lewostronne i prawostronne sygnały detekcyjne będą oznaczane dalej odpowiednio przez $\widehat{d}_j^f(t)$ i $\widehat{d}_j^b(t)$. Następnie taka konfiguracja rozszerzonych alarmów łączona jest przy użyciu odpowiedniej reguły decyzyjnej.

Dla przykładu, rozważmy konfigurację alarmów w kanale j należąca do klasy A (alarmy nachodzą na siebie częściowo)

$$\begin{aligned} \widehat{d}_j^f(t) &= 1 \quad \text{dla } t \in [\underline{t}_j^f, \overline{t}_j^f], \\ \widehat{d}_j^b(t) &= 1 \quad \text{dla } t \in [\underline{t}_j^b, \overline{t}_j^b], \\ [\underline{t}_j^f, \overline{t}_j^f] \cap [\underline{t}_j^b, \overline{t}_j^b] &\neq \emptyset, \end{aligned}$$

w której najlepsze wyniki uzyskiwane są dla reguły „przednie zbocze-przednie zbocze” – patrz punkt 8.2.1 w części pracy poświęconej omówieniu wyników eksperymentów. Zgodnie z tą regułą, nowy dwustronny alarm rozpoczyna się w chwili \underline{t}_j^f nawiązującej do przedniego zbocza lewostronnego alarmu, zaś przerywany jest w chwili \overline{t}_j^b nawiązującej do tylnego zbocza prawostronnego alarmu (które, pierwotnie w czasie odwrotnym było jego przednim zboczem)

$$\widehat{d}_j^{fb}(t) = 1 \quad \text{dla } t \in [\underline{t}_j^f, \overline{t}_j^b].$$

Założmy, że otrzymany alarm detekcyjny uwzględniający oba kanały zaczyna się w chwili $t_0 + 1$, zaś kończy się w chwili $t_0 + m$. W obu kanałach alarm ten poprzedza i występuje bezpośrednio

po nim sekwencja co najmniej r niezakłóconych próbek:

$$\begin{aligned}\widehat{d}_1^{fb}(t_0 + 1) = 1 \quad \text{lub} \quad \widehat{d}_2^{fb}(t_0 + 1) = 1, \\ \widehat{d}_1^{fb}(t_0 + m) = 1 \quad \text{lub} \quad \widehat{d}_2^{fb}(t_0 + m) = 1, \\ \widehat{d}_1^{fb}(t) = \widehat{d}_2^{fb}(t) = 0, \\ t \in [t_0 - r + 1, t_0] \cup [t_0 + m + 1, t_0 + m + r].\end{aligned}$$

Ponieważ uzyskane alarmy różnią się od ich lewostronnych i prawostronnych alarmów składowych, sygnał podlega ponownej rekonstrukcji na podstawie nowych ocen lokalizacji zakłóceń impulsowych. W tym przypadku algorytm filtru Kalmana uruchamiany jest ponownie, ale z pominięciem procedury detekcji, tj. z wykorzystaniem opracowanych sygnałów detekcyjnych $\widehat{d}_1^{fb}(t)$ i $\widehat{d}_2^{fb}(t)$, odpowiednio dla lewego i prawego kanału.

6.5 Rozszerzenie na model dwuzakresowy

W podrozdziale tym omówimy możliwość rozszerzenia techniki wektorowego modelowania autoregresyjnego na wektorowe modele dwuzakresowe (z ang. *sparse vector autoregressive*, SVAR).

6.5.1 Ogólny model SVAR

Założmy, że gdy sygnały $s_1(t)$ i $s_2(t)$ są skorelowane ze sobą, okres podstawowy T (jeżeli sygnał $\mathbf{s}(t)$ ma znamiona okresowości) jest identyczny dla lewego i dla prawego kanału, tj. wielkość T jest skalarem. Biorąc to pod uwagę, model SVAR może zostać zapisany w ogólnej postaci

$$\mathbf{s}(t) = \sum_{i=1}^r \mathbf{A}_i \mathbf{s}(t-i) + \sum_{i=\tau_0+1}^{\tau_0+v} \mathbf{A}_i \mathbf{s}(t-i) + \mathbf{n}(t), \quad (6.41)$$

gdzie wielkości $\tau_0 \gg r$ i v wybierane są w taki sam sposób jak dla przypadku skalarowego, tj. tak aby $\tau_0 + 1 \leq T \leq \tau_0 + v$.

Podobnie jak miało to miejsce dla przypadku skalarowego, niestabilny model SVAR może być przyczyną wielu problemów na etapie predykcji i/lub na etapie rekonstrukcji sygnału. W przypadku niestabilnego modelu VAR, dla którego oceny parametrów otrzymane zostały przy użyciu algorytmu EWLS (6.11), stosowano „ratunkowy” algorytm identyfikacji WWR gwarantujący stabilność otrzymywanego modelu – patrz punkt 6.3.2 – niestety, nie ma takiego algorytmu identyfikacji dla uogólnionego modelu SVAR. Z tego powodu w dalszych rozważaniach model w ogólnej postaci (6.41) zastąpiony zostanie jego zdekomponowanym odpowiednikiem, dla którego można już w łatwy sposób monitorować i wymuszać stabilność modelu.

6.5.2 Zdekomponowany model SVAR

Podobnie jak w przypadku skalarowego modelu SAR, wykorzystamy tu model SVAR będący kaskadowym połączeniem dwóch prostszych modeli opisujących odpowiednio korelację krótkozakresową sygnału oraz jego korelację długozakresową. W najprostszym przypadku opis taki

może zostać przedstawiony w następującej postaci

$$\mathbf{s}(t) = \sum_{i=1}^r \mathbf{B}_i \mathbf{s}(t-i) + \mathbf{x}(t), \quad (6.42)$$

$$\mathbf{x}(t) = \mathbf{C} \mathbf{x}(t-T) + \mathbf{n}(t), \quad (6.43)$$

gdzie $\mathbf{B}_1, \dots, \mathbf{B}_r$ są macierzami współczynników, o wymiarach 2×2 , dwuwymiarowego filtru formantowego (6.42), \mathbf{C} jest macierzą współczynników, o wymiarach 2×2 , dwuwymiarowego filtru wysokości dźwięku (6.43) oraz $\mathbf{x}(t) = [x_1(t), x_2(t)]^T$ jest dwuwymiarowym sygnałem resztowym. Opis ten jest asymptotycznie równoważny opisowi (6.41) po podstawieniu

$$\mathcal{A}(z^{-1}) = \mathcal{B}(z^{-1})\mathcal{C}(z^{-1}), \quad (6.44)$$

gdzie

$$\mathcal{B}(z^{-1}) = \mathbf{I} - \sum_{i=1}^r \mathbf{B}_i z^{-i}, \quad \mathcal{C}(z^{-1}) = \mathbf{I} - \mathbf{C} z^{-T}.$$

Zgodnie z (6.44) zachodzą następujące zależności

$$\begin{aligned} \mathbf{A}_i &= \mathbf{B}_i, \quad i = 1, \dots, r, & \mathbf{A}_i &= \mathbf{O}, \quad r < i < T, \\ \mathbf{A}_T &= \mathbf{C}, & \mathbf{A}_{T+i} &= -\mathbf{C}\mathbf{B}_i, \quad i = 1, \dots, r. \end{aligned} \quad (6.45)$$

Zdekomponowany model (6.42)-(6.43) jest stabilny pod warunkiem, że jego oba „podmodele” składowe są stabilne. Stabilność modelu krótkozakresowego, opisanego równaniem (6.42), może być w łatwy sposób monitorowana i w razie potrzeby zapewniona przez odpowiedni dobór algorytmu identyfikacji – patrz punkt 6.3.2. Również stabilność modelu długozakresowego, opisanego równaniem (6.43), jest łatwa do sprawdzenia i, jeżeli zachodzi taka potrzeba, „wymuszenia”. Aby zbadać stabilność modelu (6.43) można użyć techniki diagonalizacji.

Oznaczmy przez \mathbf{V} nieosobliwą macierz, która przekształca macierz \mathbf{C} w macierz diagonalną

$$\mathbf{V}\mathbf{C}\mathbf{V}^{-1} = \mathbf{\Lambda},$$

gdzie $\mathbf{\Lambda} = \text{diag}\{\lambda_1^\circ, \lambda_2^\circ\}$, zaś λ_1° i λ_2° oznaczają wartości własne macierzy \mathbf{C} . Niech $\check{\mathbf{x}}(t) = \mathbf{V}\mathbf{x}(t) = [\check{x}_1(t), \check{x}_2(t)]^T$ i $\check{\mathbf{n}}(t) = \mathbf{V}\mathbf{n}(t) = [\check{n}_1(t), \check{n}_2(t)]^T$. Zauważmy, że model opisany równaniem (6.43) może zostać przepisany w równoważnej postaci zdekomponowanej

$$\check{x}_j(t) = \lambda_j^\circ \check{x}_j(t-T) + \check{n}_j(t), \quad j = 1, 2$$

i dlatego warunek stabilności dla rozważanego modelu jest prostym rozszerzeniem jego skalar-nego przypadku (patrz [80]):

$$|\lambda_j^\circ| < 1, \quad j = 1, 2.$$

Jeżeli warunek

$$|\lambda^\circ|_{\max} = \max\{|\lambda_1^\circ[\mathbf{C}]|, |\lambda_2^\circ[\mathbf{C}]|\} < 1$$

jest spełniony, nie jest wymagana żadna interwencja. W przeciwnym razie stabilność modelu może być w łatwy sposób wymuszona poprzez zastąpienie macierzy \mathbf{C} jej przeskalowaną wersją

$$[1 - \epsilon/|\lambda^\circ|_{\max}]\mathbf{C}. \quad (6.46)$$

6.5.3 Rekonstrukcja sygnału oparta na modelu SVAR

Rozważmy problem rekonstrukcji odizolowanego bloku m nieodwracalnie uszkodzonych próbek stacjonarnego procesu AR w przypadku gdy opisany jest modelem SVAR w następującej postaci

$$\mathbf{s}(t) = \sum_{i=1}^r \mathbf{A}_i \mathbf{s}(t-i) + \sum_{i=T}^{T+r} \mathbf{A}_i \mathbf{s}(t-i) + \mathbf{n}(t). \quad (6.47)$$

Model ten może zostać otrzymany w wyniku przekształcenia modelu w postaci zdekomponowanej (6.42)-(6.43) do jednorodnej postaci rzadkiej (6.41) zgodnie z zależnością (6.45). Przyjmijmy te same założenie odnośnie lokalizacji bloku uszkodzonych próbek oraz początkowej wiedzy o sygnałach $\mathbf{y}(t)$, $\mathbf{n}(t)$ i $\boldsymbol{\delta}(t)$, jak w przypadku rekonstrukcji opartej na modelu VAR – patrz punkt 6.2.1.

Podobnie jak w przypadku skalarnego modelu SAR, przy założeniu że $q_{\max} + r < T_{\min}$, model SVAR (6.47) może być traktowany jako wektorowy dwuwymiarowy proces autoregresyjny z zewnętrznym wejściem (z ang. *vector autoregressive with exogenous input*, VARX)

$$\mathbf{s}(t) = \sum_{i=1}^r \mathbf{A}_i \mathbf{s}(t-i) + \mathbf{u}(t) + \mathbf{n}(t), \quad (6.48)$$

gdzie wielkość

$$\mathbf{u}(t) = \sum_{i=T}^{T+r} \mathbf{A}_i \mathbf{s}(t-i) \quad (6.49)$$

traktowana jest jako zewnętrzny (mierzalny) dwuwymiarowy sygnał wejściowy, a nie jako wewnętrzna zmienna stanu. Równoważna modelowi (6.1) i (6.48) „pseudo-reprezentacja” w przestrzeni stanów może zostać zapisana w następującej postaci

$$\begin{aligned} \mathbf{x}_q(t+1) &= \mathbf{A}_q \mathbf{x}_q(t) + \mathbf{C}_q \mathbf{u}(t+1) + \mathbf{C}_q \mathbf{n}(t+1), \\ \mathbf{y}(t) &= \mathbf{C}_q^T \mathbf{x}_q(t) + \boldsymbol{\delta}(t), \end{aligned} \quad (6.50)$$

gdzie wielkości $\mathbf{x}_q(t)$, \mathbf{A}_q , \mathbf{C}_q oraz q są identyczne z tymi zdefiniowanymi dla przypadku modelu VAR. Zmodyfikowany filtr Kalmana oparty na modelu VARX przyjmuje następującą postać

$$\begin{aligned} \hat{\mathbf{x}}_q(t|t-1) &= \mathbf{A}_q \hat{\mathbf{x}}_q(t-1|t-1) + \mathbf{C}_q \mathbf{u}(t), \\ \mathbf{P}_q(t|t-1) &= \mathbf{A}_q \mathbf{P}_q(t-1|t-1) \mathbf{A}_q^T + \mathbf{C}_q \boldsymbol{\rho} \mathbf{C}_q^T, \\ \mathbf{e}(t) &= \mathbf{y}(t) - \mathbf{C}_q^T \hat{\mathbf{x}}_q(t|t-1), \\ \mathbf{G}(t) &= \mathbf{C}_q^T \mathbf{P}_q(t|t-1) \mathbf{C}_q + \boldsymbol{\Delta}(t), \\ \mathbf{L}_q(t) &= \mathbf{P}_q(t|t-1) \mathbf{C}_q \mathbf{G}^{-1}(t), \\ \hat{\mathbf{x}}_q(t|t) &= \hat{\mathbf{x}}_q(t|t-1) + \mathbf{L}_q(t) \mathbf{e}(t), \\ \mathbf{P}_q(t|t) &= \mathbf{P}_q(t|t-1) - \mathbf{L}_q(t) \mathbf{G}(t) \mathbf{L}_q^T(t). \end{aligned} \quad (6.51)$$

Podobnie jak w przypadku skalarnym, algorytm filtracji Kalmana kończy swoją pracę w chwili $t_0 + m + r$, po zaakceptowaniu r nieuszkodzonych próbek w obu kanałach sygnału stereofonicznego zlokalizowanych bezpośrednio za zakwestionowanym fragmentem – patrz uwaga 5.1.

6.5.4 Detekcja oparta na modelu SVAR

Jednokrokowa predykcja sygnału oparta na zdekomponowanym modelu SVAR (6.42)-(6.43) może zostać opisana w następującej postaci

$$\begin{aligned}\varepsilon(t|t-1) &= \mathbf{y}(t) - \hat{\mathbf{y}}(t|t-1), \\ \hat{\mathbf{y}}(t|t-1) &= \sum_{i=1}^r \mathbf{B}_i \mathbf{y}(t-i) + \hat{\mathbf{x}}(t|t-1), \\ \hat{\mathbf{x}}(t|t-1) &= \mathbf{C} \mathbf{x}(t-T),\end{aligned}\tag{6.52}$$

gdzie dwuwymiarowy sygnał resztowy $\mathbf{x}(t)$ otrzymywany jest w wyniku filtracji odwrotnej

$$\mathbf{x}(t) = \mathbf{y}(t) - \sum_{i=1}^r \mathbf{B}_i \mathbf{y}(t-i).\tag{6.53}$$

Alarm detekcyjny podniesiony zostaje gdy błąd jednokrokowej predykcji dla lewego i/lub prawego kanału może być traktowany jako nadmiarowy, tj. po spełnieniu warunku (6.22). Gdy alarm detekcyjny został podniesiony, dalsze kroki związane z predykcją sygnału i detekcją zakłócenia impulsowego mogą zostać zrealizowane według schematu opisanego w punkcie 6.2.2, jeżeli wielkości $\mathbf{e}(t)$ i $\boldsymbol{\Sigma}(t)$ otrzymywana będzie przy użyciu zmodyfikowanego filtru Kalmana (6.51) opartego na modelu VARX (6.48). Dodatkowo, jeżeli dokonamy podstawienia $\mathbf{u}(t) = \hat{\mathbf{x}}(t|t-1)$, równanie prognozy stanu w zmodyfikowanym filtrze Kalmana upraszcza się do postaci

$$\hat{\mathbf{x}}_q(t|t-1) = \mathbf{A}_q \hat{\mathbf{x}}_q(t-1|t-1) + \mathbf{C}_q \hat{\mathbf{x}}(t|t-1).\tag{6.54}$$

Zauważmy, że detekcja zakłócenia impulsowego oparta na modelu VARX (6.48) może być przeprowadzona przy użyciu algorytmu filtracji Kalmana o zmiennym rzędzie, opisanego w punkcie 6.2.3, po uwzględnieniu dodatkowego składnika $\hat{\mathbf{x}}(t|t-1)$ w równaniu predykcji sygnału, otrzymując jego nową, następującą postać

$$\hat{\mathbf{y}}(t|t-1) = \boldsymbol{\Theta}_{q(t)-1}^T \hat{\mathbf{x}}_{q(t)-1}(t-1|t-1) + \hat{\mathbf{x}}(t|t-1).\tag{6.55}$$

Algorytm detekcji kończy swoją pracę po spełnieniu warunku (6.23).

6.5.5 Identyfikacja modelu SVAR

Identyfikacja zdekomponowanego modelu (6.42)-(6.43) przeprowadzona będzie w dwóch krokach. W pierwszym kroku, zakładając, że (6.42) jest procesem VAR rzędu r , oceny $\hat{\mathbf{B}}_1, \dots, \hat{\mathbf{B}}_r$ (otrzymane przy użyciu algorytmu identyfikacji opisanego w podrozdziale 6.1) użyte zostaną do wygenerowania sygnału resztowego $\mathbf{x}(t)$

$$\mathbf{x}(t) = \mathbf{y}(t) - \sum_{i=1}^r \hat{\mathbf{B}}_i(t) \mathbf{y}(t-i).\tag{6.56}$$

W drugim kroku estymowane będą parametry modelu długozakresowego (6.43).

Identyfikacja modelu długozakresowego

Identyfikacja modelu długozakresowego (6.43) przebiega analogicznie jak w przypadku skalarnym – patrz podrozdział 5.5. Zakładając, że okres podstawowy T jest znany, spełniona jest następująca kowariancyjna zależność

$$\mathbf{R}_T^x = \mathbf{C}\mathbf{R}_0^x, \quad (6.57)$$

gdzie $\mathbf{R}_k^x = E[\mathbf{x}(t)\mathbf{x}^T(t-k)]$. Lokalna ocena macierzy \mathbf{C} może być otrzymana na drodze rozwiązania równania

$$\widehat{\mathbf{R}}_T^x(t) = \widehat{\mathbf{C}}(t)\widehat{\mathbf{R}}_0^x(t), \quad (6.58)$$

gdzie

$$\widehat{\mathbf{R}}_k^x(t) = \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{x}(t-i)\mathbf{x}^T(t-i-k), \quad k = 0, T,$$

są lokalnymi nieobciążonymi ocenami macierzy \mathbf{R}_0^x i \mathbf{R}_T^x . Ponieważ okres podstawowy jest nieznan i zmienny w czasie, w celu jego dokładnej oceny będą używane dwa estymatory: estymator przyczynowy oraz estymator nieprzyczynowy.

A) Przyczynowa estymacja okresu T

W celu estymacji okresu T wykorzystany zostanie sygnał $\mathbf{x}(t)$ otrzymywany w wyniku filtracji odwrotnej (6.56). Technika szukania najlepszego oszacowania okresu T oparta jest na minimalizacji funkcji $J_t(T, N)$

$$J_t(T, N) = \sum_{i=0}^{N-1} \|\mathbf{x}(t-i) - \mathbf{x}(t-T-i)\|^2. \quad (6.59)$$

Postępując zgodnie z dwukrokową procedurą estymacji (5.24) i (5.25) ustalona zostaje ocena okresu $\widehat{T}(t)$. Następnie, dla tej oceny okresu otrzymujemy ocenę macierzy współczynników modelu długozakresowego

$$\widehat{\mathbf{C}}(t) = \widehat{\mathbf{R}}_{\widehat{T}(t)}^x(t) \left[\widehat{\mathbf{R}}_0^x(t) \right]^{-1}. \quad (6.60)$$

Po podstawieniu oszacowanych parametrów otrzymujemy następującą postać modelu długozakresowego

$$\mathbf{x}(t) = \widehat{\mathbf{C}}(t)\mathbf{x}[t - \widehat{T}(t)] + \mathbf{n}(t). \quad (6.61)$$

Eksperymentalnie sprawdzono, że gdyby funkcję $J_t(T, N)$ (6.59) zastąpić jej wersją uwzględniającą tylko jeden kanał, lewy lub prawy

$$J_t(T, N) = \sum_{i=0}^{N-1} |x_j(t-i) - x_j(t-T-i)|^2, \quad j = 1, 2, \quad (6.62)$$

zdolności predykcyjne otrzymanego modelu długozakresowego (6.61) ulegają pogorszeniu w porównaniu do modelu, dla którego ocena okresu wyznaczona została przy łącznej analizie lewego i prawego kanału (6.59).

B) Nieprzyczynowa estymacja okresu T

Odpowiednia nieprzyczynowa ocena $\hat{T}^*(t)$ może zostać otrzymana przy użyciu procedury (5.24)-(5.25) minimalizującej funkcję $J_t^*(T, N, l)$

$$J_t^*(T, N, l) = \sum_{i=0}^{N-1} \|\mathbf{x}(t+l-i|t) - \mathbf{x}(t+l-T-i|t)\|^2, \quad (6.63)$$

gdzie

$$\mathbf{x}(t+k|t) = \begin{cases} \mathbf{y}(t+k) - \sum_{i=1}^r \hat{\mathbf{B}}_i(t)\mathbf{y}(t+k-i), & \text{dla } k > 0, \\ \mathbf{x}(t+k), & \text{dla } k \leq 0. \end{cases}$$

Nieprzyczynowa ocena $\hat{\mathbf{C}}^*(t)$ macierzy \mathbf{C} może zostać otrzymana w sposób analogiczny do (6.60)

$$\hat{\mathbf{C}}^*(t) = \hat{\mathbf{R}}_{\hat{T}^*(t)}^{\mathbf{x}}(t|t+m_*) \left[\hat{\mathbf{R}}_0^{\mathbf{x}}(t|t+m_*) \right]^{-1}, \quad (6.64)$$

gdzie

$$\hat{\mathbf{R}}_k^{\mathbf{x}}(t|t+m_*) = \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{x}(t+m_*-i|t)\mathbf{x}^T(t+m_*-i-k|t).$$

Po podstawieniu otrzymujemy model postaci

$$\mathbf{x}(t) = \hat{\mathbf{C}}^*(t)\mathbf{x}[t - \hat{T}^*(t)] + \mathbf{n}(t). \quad (6.65)$$

6.5.6 Porównanie modeli VAR i SVAR

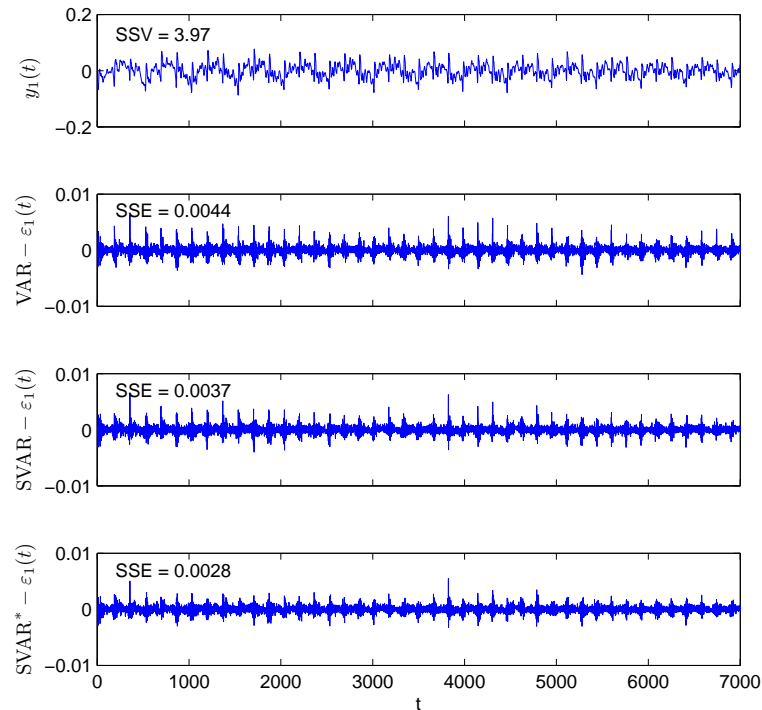
Analogicznie jak w przypadku skalarnym, oznaczymy symbolami $\varepsilon_{\text{VAR}}(t+1|t)$, $\varepsilon_{\text{SVAR}}(t+1|t)$ i $\varepsilon_{\text{SVAR}^*}(t+1|t)$ ciągi błędów jednokrokowej predykcji generowane odpowiednio przez modele: VAR (6.56), przyczynowy model SVAR (6.56), (6.61) i nieprzyczynowy model SVAR* (6.56), (6.65):

$$\begin{aligned} \varepsilon_{\text{VAR}}(t+1|t) &= \mathbf{y}(t+1) - \sum_{i=1}^r \hat{\mathbf{B}}_i(t)\mathbf{y}(t-i+1), \\ \varepsilon_{\text{SVAR}}(t+1|t) &= \mathbf{x}(t+1) - \hat{\mathbf{C}}(t)\mathbf{x}[t+1 - \hat{T}(t)], \\ \varepsilon_{\text{SVAR}^*}(t+1|t) &= \mathbf{x}(t+1) - \hat{\mathbf{C}}^*(t)\mathbf{x}[t+1 - \hat{T}^*(t)], \end{aligned}$$

gdzie $\mathbf{x}(t+1) = \varepsilon_{\text{VAR}}(t+1|t)$. Rysunek 6.5 pokazuje porównanie błędów jednokrokowej predykcji otrzymanych dla krótkiego fragmentu nagrania stereofonicznego, w trzech rozważanych powyżej przypadkach. Zauważmy, że zdolności predycyjne modelu SVAR zostały zwiększone w porównaniu z modelem VAR, z kolei model SVAR* posiada lepsze zdolności predycyjne niż model SVAR.

6.5.7 Wektorowy schemat detekcji zakłóceń impulsowych

Wektorowy schemat detekcji jest prostym rozszerzeniem schematu detekcji dla przypadku skalarnego. Analogicznie, uwzględnia on trzy równoległe działające detektory zakłóceń impulsowych, dwa z nich oparte na modelu SVAR oraz jeden na model VAR. Po zastosowaniu reguły



Rysunek 6.5: Lewy kanał sygnału stereofonicznego (górny wykres) i błędy jednokrokowej predykcji (pozostałe wykresy) otrzymane na podstawie wektorowego modelowania autoregresyjnego (VAR) i wektorowych modeli dwuzakresowych wykorzystujących przyczynową (SVAR) i nieprzyczynową (SVAR*) ocenę okresu podstawowego. SSV oznacza sumę kwadratów wartości sygnału, a SSE – sumę kwadratów błędów predykcji. Zauważmy różnicę w skali pomiędzy wykresem sygnału a wykresami błędów.

decyzyjnej (6.22) do ciągu błędów predykcji generowanych przez każdy z konkurujących modeli, otrzymujemy trzy tymczasowe wyniki detekcji $\hat{\mathbf{d}}_{\text{VAR}}(t)$, $\hat{\mathbf{d}}_{\text{SVAR}}(t)$ i $\hat{\mathbf{d}}_{\text{SVAR}^*}(t)$. Alarm detekcyjny podnoszony jest w chwili $t_0 + 1$ dopiero gdy wszystkie trzy błędy predykcji mogą być traktowane jako nadmiarowe, mianowicie jeżeli

$$\hat{\mathbf{d}}_{\text{VAR}}(t_0 + 1) \neq \mathbf{0}, \quad \hat{\mathbf{d}}_{\text{SVAR}}(t_0 + 1) \neq \mathbf{0} \quad \text{i} \quad \hat{\mathbf{d}}_{\text{SVAR}^*}(t_0 + 1) \neq \mathbf{0}. \quad (6.66)$$

Gdy alarm detekcyjny jest podniesiony w chwili $t_0 + 1$, test błędów predykcji oparty jest na regule decyzyjnej (6.27) i może być przeprowadzony dla wielokrokowej predykcji. Zadanie te może zostać wykonane przez odpowiednio zaprojektowany filtr Kalmana o zmiennym rzędzie, uruchamiany w chwili $t_0 + 1$. Alarm detekcyjny jest przerywany w chwili $t_0 + m$ jeżeli r kolejnych błędów predykcji przyjmuje dla obu kanałów odpowiednio małe wartości: $\hat{\mathbf{d}}(t_0 + m + i) = \mathbf{0}$ dla $i = 1, \dots, r$. Algorytm filtracji Kalmana oparty jest na modelu (VAR, SVAR lub SVAR*), który minimalizuje sumę kwadratów błędów jednokrokowej predykcji obserwowanych w niezbyt długiej przeszłości $\sum_{i=1}^r \|\boldsymbol{\varepsilon}(t_0 + 1 - i)\|^2$. W przypadku modeli SVAR/SVAR* filtr Kalmana wykorzystuje pseudo reprezentację modelu VARX w przestrzeni stanów co pozwala na zachowanie struktury algorytmu filtru Kalmana o zmiennym rzędzie (6.2.3) poprzez zastąpienie równania

predykcji (6.31) jego następującym odpowiednikiem w wersji przyczynowej SVAR

$$\hat{\mathbf{y}}(t|t-1) = \hat{\Theta}_{q(t)-1}^T(t_0) \hat{\mathbf{x}}_{q(t)-1}(t-1|t-1) + \hat{\mathbf{C}}(t-1) \mathbf{x}[t - \hat{T}(t-1)] \quad (6.67)$$

oraz wersji nieprzyczynowej SVAR*

$$\hat{\mathbf{y}}(t|t-1) = \hat{\Theta}_{q(t)-1}^T(t_0) \hat{\mathbf{x}}_{q(t)-1}(t-1|t-1) + \hat{\mathbf{C}}^*(t-1) \mathbf{x}[t - \hat{T}^*(t-1)]. \quad (6.68)$$

6.6 Podsumowanie

W rozdziale tym omówiono nowe podejście oparte na wektorowym modelowaniu autoregresyjnym. Podejście to wykorzystuje łączną analizę lewego i prawego kanału nagrania stereofonicznego i w efekcie otrzymywany model sygnału cechuje się lepszymi zdolnościami predykcyjnymi i interpolacyjnymi niż model sygnału otrzymywany w wyniku analizy tylko lewego lub tylko prawego kanału. W proponowanym podejściu zakłócenia impulsowe eliminowane są przez specjalnie zaprojektowany algorytm filtracji Kalmana o zmiennym rzędzie, który pozwala na rozróżnienie pomiędzy poszczególnymi kanałami nagrania stereofonicznego. Na koniec dokonano rozszerzenia tego podejścia na analizę dwukierunkową oraz na dwuzakresowe wektorowe modelowanie autoregresyjne.

Rozdział 7

Technika dopasowywania wzorców

Wprowadzenie

W tym rozdziale omówiono nowe podejście do problemu eliminacji zakłóceń impulsowych z sygnałów fonicznych, oparte na połączeniu techniki filtracji dopasowanej z klasyczną metodą detekcji wykorzystującą modelowanie autoregresyjne. Podejście to wykorzystuje w procesie detekcji wiedzę *a priori* o typowych rozmiarach i kształtach zakłóceń impulsowych.

Technika filtracji dopasowanej proponowana była już we wcześniejszych pracach poświęconych eliminacji zakłóceń impulsowych [48], [49]. Autorzy wspomnianych prac analizowali jaki wpływ ma wyidealizowane zakłócenie impulsowe (modelowane jako impuls Kroneckera) na wyjście filtru odwrotnego. Sugerowali, że w celu wykrycia zakłócenia w sygnale można wyznaczyć splot ciągu błędów resztowych, otrzymanych w wyniku filtracji odwrotnej, z ciągiem współczynników autoregresji, a następnie dokonać progowania. Podejście to nie korzysta jednak z żadnej wiedzy o typowych kształtach zakłóceń impulsowych. Opisana procedura może być więc użyta tylko do wykrywania krótkich, unimodalnych impulsów zakłócających. Z kolei technika opisana w pracy [49] jest bardziej podobna do proponowanego podejścia, ale używana może być tylko do wykrywania bardzo długich zakłóceń o dużej energii, takich jak zakłócenia których źródłem są głębokie rysy na powierzchni płyty winylowej.

Zgodnie z posiadaną wiedzą, oprócz metody opisanej w pracy [49], jedynym rozwiązaniem zaproponowanym do tej pory, które uwzględnia wiedzę *a priori* o zakłóceniach impulsowych w procesie ich detekcji i eliminacji, jest podejście opisane w pracy Ávila i Biscainho [46]. Autorzy pracy zaproponowali bayesowską procedurę dopasowywania wzorców opartą na sekwencyjnie dostrajanych modelach: model AR czystego sygnału (z dostrajanymi współczynnikami autoregresji oraz z dostrajaną wariancją białego szumu tworzącego) oraz sparametryzowany model zakłócenia impulsowego (wykładniczo zanikające zakłócenie z dostrajanymi parametrami kształtu). Problem łącznej detekcji i estymacji uszkodzonych próbek rozwiązywany jest tu przy użyciu metody próbkowania Gibbsa – za pomocą wariantu algorytmu Metropolisa-Hastingsa szukany jest łączny rozkład *a posteriori* dla czystego sygnału, współczynników autoregresji i parametrów modelu zakłócenia. Proponowana, iteracyjna procedura jest obliczeniowo bardzo złożona.

Nowe podejście opiera się na obserwacji, że znaczna część zakłóceń impulsowych w nagraniach fonicznych wyróżnia się typowymi, powtarzającymi się kształtami pasującymi do niewielkiej liczby „wzorców”. W ramach tego podejścia zaproponowano procedurę pozwalającą na tworzenie wzorców na podstawie analizy reprezentatywnego zbioru zakłóceń impulsowych. Zbiór ten w wielu przypadkach może zostać otrzymany w wyniku „wycięcia” przykładowych zakłóceń



impulsowych z fragmentów ciszy poprzedzających dany utwór i/lub występujących tuż po nim. W wyniku zastosowania wspomnianej wyżej procedury otrzymujemy niewielką liczbę wzorców zakłóceń w porównaniu do znacznie liczniejszego zbioru przykładowych zakłóceń impulsowych. Proponowane podejście wykorzystuje klasyczną metodę filtracji dopasowanej do lokalizacji zakłóceń impulsowych podobnych do wzorców zakłóceń. W celu wykrycia zakłócenia impulsowego wyznacza się splot odpowiednio zmodyfikowanych wzorców z sekwencją błędów jednokrokowej predykcji otrzymaną za pomocą predyktora opartego na modelu AR, a następnie „proguje” się otrzymane wyniki. W wyniku zastosowania takiego podejścia otrzymane lokalizacje zakłóceń impulsowych są dokładniejsze niż w przypadku zastosowania algorytmu detekcji w pętli otwartej lub algorytmu detekcji z decyzyjnym sprzężeniem zwrotnym. Ważną cechą proponowanego podejścia jest jego zdolność adaptacji do źródła zakłóceń impulsowych, wzorce zakłóceń mogą być bowiem tworzone dla konkretnego nagrania otrzymanego przy użyciu konkretnego sprzętu fonicznego.

7.1 Wzorce zakłóceń impulsowych

W archiwalnych nagraniach fonicznych spotkać możemy zakłócenia impulsowe o nietypowych kształtach, w większości przypadków jednak są to zakłócenia o kształtach powtarzalnych, należących do względnie małej liczby klas reprezentowanych przez wzorce.

W niniejszym podrozdziale omówimy procedurę tworzenia wzorców na podstawie zbioru przykładowych zakłóceń impulsowych. Procedura ta stanowi oryginalną część proponowanego podejścia.

7.1.1 Przykładowe zakłócenia impulsowe

Jeżeli chcemy stworzyć wzorce reprezentujące zakłócenia o typowych kształtach należy wziąć pod uwagę to, że typowy kształt oraz czas trwania zakłóceń impulsowych mogą silnie zależeć wielu czynników – patrz podrozdział 2.2. Proponowana procedura pozwala na adaptację do źródła zakłóceń – wzorce zakłóceń mogą być tworzone dla konkretnej grupy nagrań [np. tych pochodzących z płyt długogrających (LP, z ang. *Long play*)] otrzymanych przy użyciu konkretnego sprzętu fonograficznego. Aby otrzymać wiarygodne wzorce zakłóceń impulsowych, typowych dla danego źródła, należy zebrać co najmniej kilkaset przykładowych zakłóceń impulsowych. Choć dla pojedynczego nagrania spełnienie tego wymogu może okazać się trudne lub nawet niemożliwe, to już w przypadku kolekcji wielu nagrań, zgromadzonych na płytach LP lub taśmach, zadanie zebrania wystarczającej liczby przykładowych zakłóceń impulsowych staje się względnie proste.

Przykładowe zakłócenia impulsowe mogą zostać „wycięte” z fragmentów ciszy poprzedzających nagrane utwory i/lub występujących bezpośrednio po nich. Ponieważ zakłócenia impulsowe są bardziej widoczne na tle ciszy, niż na tle utworu, ich precyzyjne wykrycie jest stosunkowo łatwe – można tego dokonać stosując klasyczne metody detekcji oparte na adaptacyjnym progowaniu sygnału predykcji zgodnie z regułą „3-sigma”. W celu eliminacji ewentualnych błędów detekcji sekwencja próbek otrzymana w wyniku odjęcia od sygnału oryginalnego jego rekonstrukcji, jest dodatkowo poddawana ocenie wzrokowej. W ten sposób otrzymujemy zbiór przykładowych zakłóceń impulsowych.



7.1.2 Analiza podobieństwa kształtu

Celem tego etapu jest ocena stopnia podobieństwa kształtu dla każdej pary zakłóceń impulsowych pochodzących z wcześniej otrzymanego zbioru przykładowych zakłóceń.

Do analizy podobieństwa kształtu używać będziemy wielkości znanej w statystyce jako współczynnik korelacji Pearsona – unormowany współczynnik oceniający zależność liniową między dwiema zmiennymi losowymi X i Y . Współczynnik Pearsona, dla realizacji dwóch zmiennych losowych X i Y zapisanych jako ciąg pomiarów $x(k)$ i $y(k)$, $k = 1, \dots, N$, może zostać obliczony w następujący sposób

$$\hat{r}_{XY} = \frac{\sum_{k=1}^N [x(k) - \bar{x}][y(k) - \bar{y}]}{\sqrt{\sum_{k=1}^N [x(k) - \bar{x}]^2 \sum_{k=1}^N [y(k) - \bar{y}]^2}} = \sum_{k=1}^N \tilde{x}(k)\tilde{y}(k), \quad (7.1)$$

gdzie

$$\bar{x} = \frac{1}{N} \sum_{k=1}^N x(k), \quad \bar{y} = \frac{1}{N} \sum_{k=1}^N y(k) \quad (7.2)$$

oznaczają odpowiednio oceny średnich wartości X i Y , a

$$\tilde{x}(k) = \frac{x(k) - \bar{x}}{\sqrt{\sum_{k=1}^N [x(k) - \bar{x}]^2}}, \quad \tilde{y}(k) = \frac{y(k) - \bar{y}}{\sqrt{\sum_{k=1}^N [y(k) - \bar{y}]^2}} \quad (7.3)$$

oznaczają unormowane pomiary. Współczynnik Pearsona przyjmuje wartości z przedziału od -1 do 1 i jest niewrażliwy na skalowanie pomiarów, co czyni go dobrym narzędziem do oceny stopnia podobieństwa kształtu.

Niech $\mathcal{P} = \{\mathcal{P}_1, \dots, \mathcal{P}_{n_t}\}$ oznacza zbiór składający się z n_t „wyciętych” zakłóceń impulsowych, gdzie

$$\mathcal{P}_i = \{p_i(1), \dots, p_i(k_i)\} \quad (7.4)$$

jest sekwencją próbek o długości k_i tworzącą i -te zakłócenie. Symbolem $\tilde{\mathcal{P}}_i = \{\tilde{p}_i(1), \dots, \tilde{p}_i(k_i)\}$ oznaczmy sekwencję unormowanych próbek przebiegu zakłócenia impulsowego, otrzymaną w sposób analogiczny do (7.3). Aby znaleźć najlepsze dopasowanie położenia \mathcal{P}_j względem \mathcal{P}_i , obliczać będziemy współczynnik korelacji między $\tilde{\mathcal{P}}_i$ a sekwencją $\tilde{\mathcal{P}}_j$ przesuniętą o τ próbek, gdzie τ jest liczbą całkowitą. Jeżeli założymy, że próbki poprzedzające $\tilde{p}_i(1)$ i występujące po $\tilde{p}_i(k_i)$ mają zerowe wartości, to ocenę stopnia podobieństwa kształtu między \mathcal{P}_i a \mathcal{P}_j dla przesunięcia τ możemy wyrazić w postaci

$$r_{ij}(\tau) = \sum_{\substack{k=1 \\ 1 \leq k+\tau \leq k_j}}^{k_i} \tilde{p}_i(k)\tilde{p}_j(k+\tau) = \sum_{k=\max(1, 1-\tau)}^{\min(k_i, k_j-\tau)} \tilde{p}_i(k)\tilde{p}_j(k+\tau), \quad (7.5)$$

gdzie $\tau \in \mathcal{T}_{ij} = [1 - k_i, k_j - 1]$. Zauważmy, że przedział sumowania w (7.5) uwzględnia fakt, że porównywane sekwencje mogą mieć różną długość. Cały zbiór współczynników korelacji $r_{ij}(\tau)$, $\tau \in \mathcal{T}_{ij}$, może być efektywnie obliczony jeśli splot wyznaczany jest w dziedzinie częstotliwości za pomocą algorytmu FFT. W wyniku maksymalizacji $r_{ij}(\tau)$ otrzymujemy optymalne przesunięcie τ_{ij}

$$\tau_{ij} = \arg \max_{\tau \in \mathcal{T}_{ij}} r_{ij}(\tau), \quad (7.6)$$

które gwarantuje najlepsze dopasowanie porównywanych sekwencji. Zatem, aby ocenić stopień podobieństwa kształtu między \mathcal{P}_i a \mathcal{P}_j , używać będziemy następującego współczynnika korelacji

$$r_{ij}^* = \max_{\tau \in \mathcal{T}_{ij}} r_{ij}(\tau) = r_{ij}(\tau_{ij}). \quad (7.7)$$

7.1.3 Tworzenie wzorców zakłóceń impulsowych

Na tym etapie pokażemy w jaki sposób zakłócenia impulsowe ocenione jako podobne mogą zostać pogrupowane, unormowane, wyrównane, a na koniec uśrednione, tworząc wzorce zakłóceń.

W celu zgrupowania wzajemnie podobnych zakłóceń impulsowych wprowadzimy nieskierowany graf podobieństwa G , zbudowany na podstawie zbioru współczynników korelacji. Graf ten pokazuje wewnętrzną strukturę podobieństwa analizowanego zbioru zakłóceń impulsowych, gdzie kolejne przebiegi zakłóceń $\mathcal{P}_1, \dots, \mathcal{P}_{n_t}$ przypisane są do kolejnych wierzchołków grafu. Jeżeli stopień podobieństwa między \mathcal{P}_i a \mathcal{P}_j , ($i \neq j$) jest wystarczająco wysoki ($r_{ij}^* \geq \kappa$, gdzie κ jest progiem bliskim wartości 1, np. $\kappa = 0,95$), wtedy wierzchołki związane z \mathcal{P}_i i \mathcal{P}_j połączone zostają krawędzią. Macierz sąsiedztwa dla grafu G ma więc postać

$$\mathbf{L} = [l_{ij}]_{n_t \times n_t}, \quad l_{ij} = \begin{cases} 1, & \text{jeżeli } r_{ij}^* \geq \kappa \text{ i } i \neq j, \\ 0, & \text{w pozostałych przypadkach.} \end{cases} \quad (7.8)$$

Wzorce zakłóceń impulsowych mogą zostać otrzymane na drodze uśredniania przebiegów zakłóceń należących do tych samych maksymalnych klik¹ znalezionych w grafie G . Proponowana procedura szukania maksymalnych, niezależnych klik jest rekurencyjna i może być przedstawiona w następujący sposób:

Inicjalizacja: $\mathbf{L}_1 \leftarrow \mathbf{L}$, $G_1 \leftarrow G$, $i \leftarrow 1$.

Krok 1: Znajdź maksymalną klikę Q_i w grafie G_i zdefiniowanym przez macierz sąsiedztwa \mathbf{L}_i . Jeżeli znaleziono więcej niż jedną maksymalną klikę (z tą samą liczbą wierzchołków), wybierz tę, dla której suma wartości współczynników korelacji r_{ij}^* przyjmuje największą wartość (sumowanie odbywa się po wszystkich krawędziach Q_i). Alternatywą może być użycie bardziej skomplikowanego algorytmu rozwiązującego zadanie szukania ważonych maksymalnych klik. Jeżeli rozmiar klik Q_i jest wystarczająco duży, np. jeżeli liczba zgrupowanych zakłóceń impulsowych $n_i \geq 10$, przejdź do **Kroku 2** – w pozostałych przypadkach następuje **Przerwanie procedury**.

Krok 2: Usuń z G_i wszystkie wierzchołki i krawędzie należące do klik Q_i . W wyniku tego otrzymamy nowy graf G_{i+1} z odpowiadającą mu macierzą sąsiedztwa \mathbf{L}_{i+1} (\mathbf{L}_{i+1} może być otrzymana przez wyzerowanie wierszy i kolumn w macierzy \mathbf{L}_i związanych z kliką Q_i). Zwiększ $i \leftarrow i + 1$ i wróć do **Kroku 1**.

Gdy już wszystkie klik Q_i zostaną znalezione, ustala się elementy centralne dla każdej z nich. Niech $S_i = \{\tilde{\mathcal{P}}_j, j \in J_i\}$ oznacza zbiór unormowanych przebiegów zakłóceń impulsowych przypisanych do klik Q_i (J_i jest zbiorem wskazującym, które wierzchołki grafu G należą do Q_i). Elementem centralnym zbioru S_i , oznaczonym przez $\tilde{\mathcal{P}}_{j_i}$, jest element o największej sumie

¹Klika reprezentuje nieskierowany, skończony (pod)graf, w którym każda para wierzchołków połączona jest krawędzią. Maksymalna klika to taka, która posiada największą liczbę wierzchołków.

wag (współczynników korelacji) przypisanych wychodzącym krawędziom

$$j_i = \arg \max_{j \in J_i} \sum_{\substack{l \in J_i \\ l \neq j}} r_{jl}^* \quad (7.9)$$

Taki element może być interpretowany jako „najbardziej podobny” do pozostałych elementów ze zbioru S_i .

Następnie wszystkie przebiegi zgrupowane w S_i zostają uzupełnione z obu stron próbkami o zerowych wartościach. Zabieg ten umożliwia dalsze wyrównanie przebiegów względem przebiegu centralnego $\tilde{\mathcal{P}}_{j_i}$. Ostatecznie po uśrednieniu przebiegów otrzymujemy nowy przebieg zakłócenia impulsowego. Zauważmy, że wszystkie optymalne przesunięcia $\tau_{j_i l}, l \in J_i$, potrzebne do wyrównania względem siebie każdej pary przebiegów, obliczone już zostały na etapie analizy podobieństwa kształtu (7.6).

Ponieważ uśrednianie wykazuje tendencję do tworzenia długich „ogonów” (małe, niezerowe wartości poprzedzające i występujące po zasadniczej części zakłócenia impulsowego), i ponieważ te „ogony” mają pomijalny wpływ na dalszą analizę podobieństwa kształtu, wzorce zakłóceń impulsowych otrzymywane są poprzez obustronne skrócenie uśrednionych przebiegów. Zalecane jest usunięcie początkowych i końcowych próbek, których bezwzględna wartość znajdzie się poniżej pięcioprocentowego progu wyznaczonego w odniesieniu do maksymalnej bezwzględnej wartości przebiegu.

Ostatecznie otrzymujemy zestaw L_w wzorców zakłóceń impulsowych

$$\mathcal{C}_i = \{c_i(1), \dots, c_i(m_i)\}, \quad 1 \leq i \leq L_w, \quad (7.10)$$

gdzie m_i oznacza długość i -tego wzorca.

Uwaga 7.1. Gdy długość otrzymanego wzorca jest zbyt mała, problem dopasowywania kształtu staje się źle uwarunkowany. Dla przykładu, rozważmy wzorcowe zakłócenie o długości 1 (zakłócenie w postaci delty Kroneckera). Za każdym razem gdy taki wzorzec porównywany jest z dowolnym fragmentem sygnału o długości 1 (pojedyncza próbka), zachodzi pełna zgodność podobieństwa kształtu (współczynnik korelacji przyjmuje wartość 1), co świadczy o utracie jego zdolności dyskryminacyjnych. Z tego powodu ustalono minimalną długość wzorców zakłóceń równą 4 – pozostałe, krótsze wzorce usunięte zostają ze zbioru wzorców. Nie powoduje to żadnych komplikacji gdyż krótkie zakłócenia są łatwo wykrywane przy użyciu tradycyjnych detektorów zakłóceń, takich jak te omówione w podrozdziale 3.4.

Uwaga 7.2. Wszystkie maksymalne kliki w macierzy sąsiedztwa \mathbf{L} mogą zostać znalezione w obliczeniowo oszczędny sposób przy użyciu algorytmu opisanego w pracy [94] lub za pomocą dobrze znanego algorytmu Brona-Kerbosha [95] używanego do szukania maksymalnej kliki w grafie (kod napisany w MATLABie *maximalcliques.m* dostępny jest w repozytorium Mathwork mathworks.com/matlabcentral). W przypadku gdy poszukiwania zawężone zostaną do tylko jednej (dowolnej) maksymalnej kliki, zadanie to może zostać znacznie szybciej rozwiązane przez algorytmy opisane w pracy [96] (kod algorytmu w języku C++ można znaleźć na stronie pierwszego autora sicmm.org/~konc/maxclique).

Uwaga 7.3. Metoda filtracji dopasowanej może działać bez tworzenia wzorców zakłóceń impulsowych. W takim przypadku, za każdym razem gdy wykryte zostanie zakłócenie impulsowe, należałoby korelować sekwencję błędów predykcji z każdym przebiegiem w bazie zakłóceń impulsowych. Ponieważ baza taka może składać się z setek lub nawet tysięcy przykładowych przebiegów zakłóceń, takie „siłowe” podejście jest obliczeniowo bardzo wymagające. Dodatkowo, ponieważ wiele przebiegów zakłóceń impulsowych wpasowuje się z wysokim współczynnikiem korelacji tylko do kilku grup, użycie wzorców zakłóceń wydaje się być naturalnym i eleganckim rozwiązaniem.

Wspomnieliśmy wcześniej, że zarówno długość jak i kształt przebiegu zakłócenia impulsowego mogą silnie zależeć od źródła materiału fonicznego. Na rysunku 7.1(a) przedstawiamy pierwszych 14 wzorców zakłóceń otrzymanych w wyniku przetwarzania zbioru 500 przykładowych zakłóceń impulsowych „wyciętych” z archiwalnego nagrania gramofonowego. Z kolei na rysunku 7.1(b) przedstawiamy analogiczny zbiór wzorców zakłóceń otrzymany dla archiwalnego nagrania pochodzącego z taśmy magnetycznej, zakłóconego oddziaływaniem sieci elektrycznej (wyładowania elektryczne). W obu przypadkach częstotliwość próbkowania wynosiła 48 kHz. Zauważmy, że w przypadku zakłóceń gramofonowych spotykane typowe kształty mają najczęściej charakter unimodalny lub bimodalny. Z kolei w przypadku zakłóceń elektrycznych kształty te mają charakter oscylacyjny. Zaletą proponowanego podejścia jest jego zdolność do wykorzystania w procesie detekcji specyficznej wiedzy związanej ze źródłem zakłóceń impulsowych.

7.2 Lokalizacja zakłóceń o typowych kształtach

Proponowana procedura lokalizacji w sygnale fonicznym zakłóceń impulsowych o typowych kształtach oparta jest na technice znanej w telekomunikacji jako filtracja dopasowana. Klasyczna filtracja dopasowana używana jest do detekcji znanych symboli, przesyłanych przez zakłócony kanał. Zakłada się, że sygnał użyteczny zmieszany jest z addytywnym białym szumem pomiarowym [72]. Detekcja odbywa się na drodze korelowania znanych wzorców symboli z odebranym sygnałem, a następnie progowania otrzymanych w ten sposób wyników. Gdy szum pomiarowy nie jest biały, technika filtracji dopasowanej może być nadal używana, pod warunkiem że analizowany sygnał zostanie „wybielony” przed zastosowaniem filtracji dopasowanej. W dalszej części podrzdziału pokażemy w jaki sposób można użyć tego podejścia do lokalizacji zakłóceń o typowych kształtach.

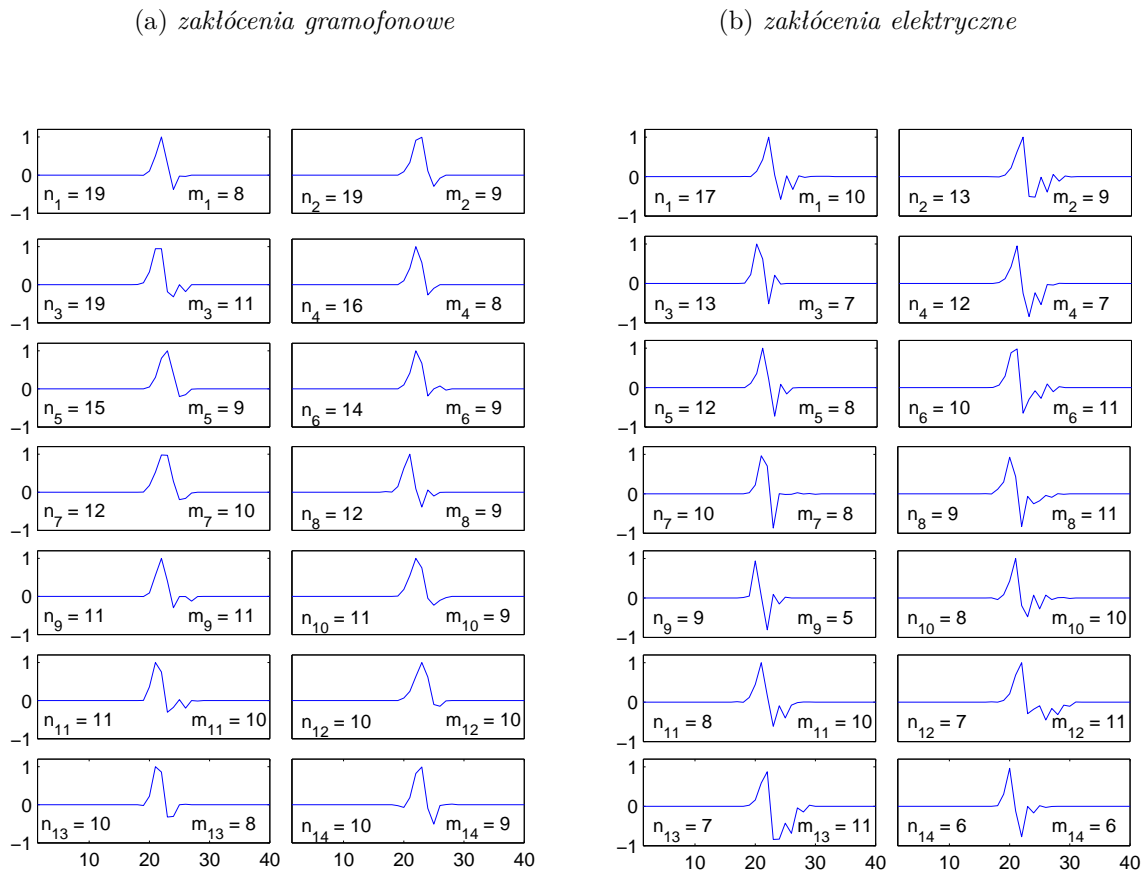
7.2.1 Wybielanie

Gdy czysty sygnał $s(t)$ reprezentowany jest przez model AR (3.2), wybielenie realizowane jest przez przepuszczenie sygnału $y(t)$ przez filtr odwrotny postaci

$$A(z^{-1}) = 1 - \sum_{i=1}^r a_i z^{-i}. \quad (7.11)$$

Zgodnie z zależnościami (3.1) i (3.2) zachodzi równość

$$\begin{aligned} \varepsilon(t|t-1) &= A(z^{-1})y(t) = n(t) + \delta^f(t), \\ \delta^f(t) &= A(z^{-1})\delta(t). \end{aligned} \quad (7.12)$$



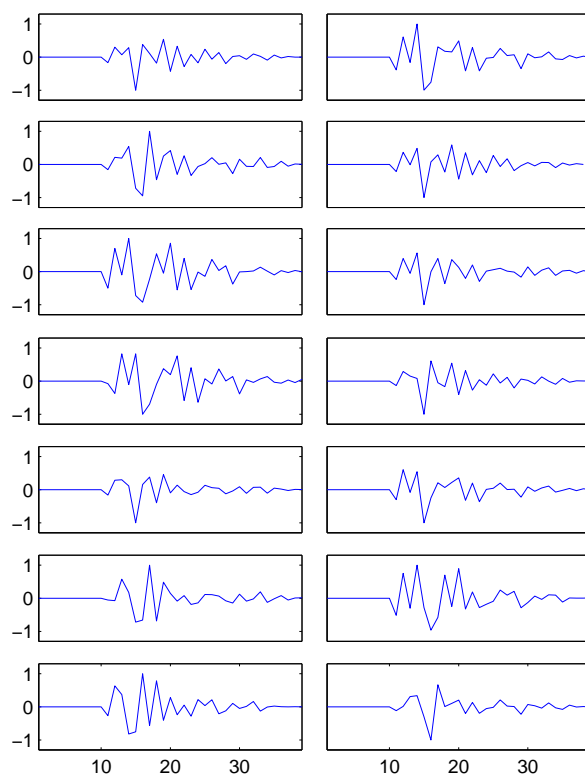
Rysunek 7.1: Po lewej zbiór 14 wzorców zakłóceń impulsowych otrzymanych dla archiwalnego nagrania gramofonowego. Po prawej zbiór 14 wzorców zakłóceń impulsowych otrzymanych dla nagrania z taśmy magnetycznej zakłóconego oddziaływaniem sieci elektrycznej. Informacje o rozmiarze kolejnych klik (n_i) i o długości wzorców zakłóceń (m_i) reprezentujących daną klikę, umieszczone zostały pod każdym z wykresów. Aby zachować oryginalny wygląd zakłóceń impulsowych, wszystkie przebiegi zostały unormowane amplitudowo z zachowaniem niezerowej wartości średniej.

W przypadku gdy $\delta(t) \neq 0$, $\delta^f(t)$ jest odpowiednio zmodyfikowanym zakłóceniem impulsowym poddanym działaniu filtracji odwrotnej. Biorąc to pod uwagę, problem detekcji w sygnale $y(t)$ zakłóceń impulsowych o typowych kształtach można sprowadzić do problemu detekcji odpowiednio zmodyfikowanych wzorców (wzorców poddanych filtracji odwrotnej) w sygnale błędów predykcji $\varepsilon(t|t-1)$. Ponieważ wielkość $n(t)$ pojawiająca się w (7.12) oznacza biały szum, drugi sposób sformułowania problemu jest zgodny z klasycznym problemem filtracji dopasowanej.

Zanim zastosowana zostanie procedura filtracji dopasowanej należy odpowiednio zmodyfikować zestaw wzorców poddając każdy z nich filtracji odwrotnej. Oznaczmy symbolem

$$\tilde{\mathcal{C}}_i = \{\tilde{c}_i(1), \dots, \tilde{c}_i(m_i)\}, \quad 1 \leq i \leq L_w, \quad (7.13)$$

zbiór unormowanych wzorców pierwotnych, który może być otrzymany w wyniku zastosowania (7.3). Niech M , $M = \max_{1 \leq i \leq L_w} m_i$, oznacza długość najdłuższego wzorca pierwotnego. Przyjmijmy, że każdy wzorec pierwotny rozszerzony zostaje z obu stron poprzez dodanie próbek



Rysunek 7.2: Przykładowy zestaw 14 wzorców wtórnych otrzymanych w wyniku filtracji odwrotnej wzorców pierwotnych.

o zerowych wartościach

$$\bar{c}_i(k) = \begin{cases} 0, & k \leq 0, \\ \tilde{c}_i(k), & 1 \leq k \leq m_i, \\ 0, & k > m_i, \end{cases} \quad (7.14)$$

a następnie poddany filtracji odwrotnej

$$c_i^f(k) = \bar{c}_i(k) - \sum_{j=1}^r a_j \bar{c}_i(k-j), \quad k = 1, \dots, k_i, \quad (7.15)$$

gdzie $k_i = m_i + r$. Zbiór wzorców wtórnych

$$\tilde{C}_i^f = \{\tilde{c}_i^f(1), \dots, \tilde{c}_i^f(k_i)\}, \quad 1 \leq i \leq L_w, \quad (7.16)$$

może być otrzymany poprzez unormowanie sekwencji $\{c_i^f(1), \dots, c_i^f(k_i)\}$ wygenerowanej za pomocą (7.15). Długość najdłuższego wzorca wtórnego oznaczana będzie symbolem k_e , $k_e = \max_{1 \leq i \leq L_w} k_i = M + r$.

Na rysunku 7.2 przedstawiono zestaw wzorców wtórnych otrzymanych w wyniku zastosowania filtracji odwrotnej do wzorców pierwotnych zaprezentowanych wcześniej na rysunku 7.1(a).

7.2.2 Wykrywanie wzorców wtórnych

Procedura wykrywania wzorców wtórnych rozpoczyna się za każdym razem, gdy podniesiony zostaje alarm detekcyjny, tj. w chwili $t + 1$, takiej że: $\hat{d}(t + 1) = 1$. W celu wykrycia zakłócenia impulsowego o typowym kształcie sprawdzeniu podlega sekwencja uwzględniająca k_b ostatnich (gdzie k_b oznacza małą liczbą całkowitą) oraz k_e „przyszłych” błędów jednokrokowej predykcji

$$\{\varepsilon(t - k_b + 1|t - k_b), \dots, \varepsilon(t + k_e|t + k_e - 1)\}. \quad (7.17)$$

W sekwencji tej szuka się odpowiednio przesuniętych w czasie wzorców wtórnych \tilde{C}_i^f badając współczynniki korelacji wzajemnej

$$g_i(\tau^f) = \sum_{k=1}^{k_i} \tilde{c}_i^f(k) \tilde{\varepsilon}(t_0 + \tau^f + k|t_0 + \tau^f + k - 1), \quad (7.18)$$

gdzie $t_0 = t - k_b$, $\tau^f \in [0, k_b]$ jest liczbą całkowitą, oznaczającą przesunięcie w czasie wzorca wtórnego, zaś

$$\{\tilde{\varepsilon}(t_0 + \tau^f + 1|t_0 + \tau^f), \dots, \tilde{\varepsilon}(t_0 + \tau^f + k_i|t_0 + \tau^f + k_i - 1)\} \quad (7.19)$$

jest unormowanym ciągiem błędów predykcji [zauważmy, że normalizacja musi być dokonana niezależnie dla każdego przesunięcia τ^f]. Wprowadzenie przesunięcia w czasie τ^f jest niezbędne, aby skorygować ewentualną niedokładność związaną ze zbyt późnym rozpoczęciem alarmu detekcyjnego, tj. aby ustalić moment, w którym powinien rozpocząć się proces dopasowywania. Gdy test oceny podobieństwa kształtu odbywa się tylko dla $k_b = \tau^f = 0$, wyniki ulegają pogorszeniu.

Oznaczmy symbolem τ_i^f optymalne przesunięcie w czasie wzorca wtórnego \tilde{C}_i^f

$$\tau_i^f = \arg \max_{\tau^f \in [0, k_b]} |g_i(\tau^f)|. \quad (7.20)$$

Zauważmy, że w celu zapewnienia niewrażliwości wyników na biegunowość wykrytych zakłóceń impulsowych, sprawdzaniu podlegają bezwzględne wartości współczynników korelacji. Najlepiej pasującym wzorcem wtórnym $\tilde{C}_{i_0}^f$ jest ten, który maksymalizuje zoptymalizowaną ocenę podobieństwa

$$i_0 = \arg \max_{1 \leq i \leq L_w} |g_i(\tau_i^f)|. \quad (7.21)$$

Po znalezieniu najlepiej pasującego wzorca należy jeszcze sprawdzić czy stopień korelacji jest wystarczający. Jest tak w przypadku gdy

$$|g_{i_0}(\tau_{i_0}^f)| \geq \kappa_0, \quad (7.22)$$

gdzie κ_0 , $\kappa_0 < \kappa$, jest przyjętym progiem podobieństwa. Aby uwzględnić wpływ szumu pomiarowego na błędy predykcji, współczynniki κ_0 nadaje się celowo mniejszą wartość niż współczynnikowi κ . Gdy warunek (7.22) jest spełniony, wykryte zakłócenie traktowane jest jako zakłócenie o typowym kształcie, pasujące do wzorca C_{i_0} . W takiej sytuacji alarm detekcyjny przyjmuje postać

$$\begin{aligned} \hat{d}(t) &= 1 \quad \text{dla } t \in D = [t_1, t_2], \\ t_1 &= t_0 + \tau_{i_0}^f + 1, \quad t_2 = t_0 + \tau_{i_0}^f + m_{i_0}. \end{aligned} \quad (7.23)$$

Opisany schemat detekcji uzupełnia (lecz nie zastępuje) klasyczne podejście oparte na modelowaniu autoregresyjnym. Za każdym razem kiedy zakłócenie impulsowe ma typowy kształt, pasujący do jednego ze wzorców, jego lokalizacja może być zazwyczaj dokładniej określona za pomocą filtracji dopasowanej niż za pomocą podejścia klasycznego. W przypadku braku dopasowania wzorca, wykrywanie zakłócenia odbywa się w oparciu o opisany wcześniej schemat wykorzystujący detekcję w pętli otwartej lub detekcję z decyzyjnym sprzężeniem zwrotnym – patrz podrozdział 3.4.

Alarmy detekcyjne ustalone w wyniku dopasowywania wzorców mogą wymagać poszerzenia. Mimo iż wprowadzono w tej metodzie przesunięcie w czasie τ^f , aby skorygować ewentualną niedokładność związaną ze zbyt późnym rozpoczęciem alarmu detekcyjnego, należy pamiętać, że na etapie tworzenia wzorców zakłóceń uśrednione przebiegi zakłóceń zostały obustronnie skrócone, dlatego korzystne jest poszerzenie uzyskanych alarmów detekcyjnych przed interpolacją. W tym przypadku poszerzenie o małą, stałą liczbę próbek Δ_2 obejmuje zarówno przednie jak i tylne zbocze alarmu detekcyjnego: $t_1 \leftarrow (t_1 - \Delta_2)$, $t_2 \leftarrow (t_2 + \Delta_2)$. Dla nagrań próbkowanych z częstotliwością 44,1 kHz i 48 kHz zalecaną wartością (sprawdzoną eksperymentalnie) jest $\Delta_2 = 1$.

7.3 Technika kompensacji

W klasycznym podejściu, stosowanym do eliminacji zakłóceń impulsowych o nietypowych kształtach, sekwencja uszkodzonych próbek sygnału $\{y(t_1), \dots, y(t_2)\}$ interpolowana jest na podstawie r próbek poprzedzających i r występujących bezpośrednio za rekonstruowanym fragmentem.

Gdy wykryte zakłócenie impulsowe pasuje do jednego ze wzorców, można rozważyć dwa sposoby rekonstrukcji uszkodzonego fragmentu sygnału: na drodze interpolacji (tak, jak w klasycznym podejściu) lub kompensacji. W ramach drugiego podejścia czysty sygnał jest odtwarzany poprzez odjęcie od zakłóconego fragmentu sygnału odpowiednio zmodyfikowanego (poprzez dopasowanie skali i wartości średniej) wzorca zakłócenia.

Założmy, że zakłócenie impulsowe poddane filtracji odwrotnej pokrywa się z odpowiednio skorygowanym i -tym wzorcem wtórnym, mianowicie

$$\delta^f(t_0 + \tau_i^f + k) = \alpha_i \tilde{c}_i^f(k) + \beta_i, \quad k = 1, \dots, k_i, \quad (7.24)$$

gdzie α_i i β_i oznaczają współczynniki korekcji skali i wartości średniej. Niech

$$\mathbf{w}_i^f = \begin{bmatrix} w_i^f(1) \\ \vdots \\ w_i^f(k_i) \end{bmatrix}, \quad \mathbf{e}_i = \begin{bmatrix} e_i(1) \\ \vdots \\ e_i(k_i) \end{bmatrix}, \quad \boldsymbol{\eta}_i = \begin{bmatrix} \eta_i(1) \\ \vdots \\ \eta_i(k_i) \end{bmatrix}, \quad \mathbf{c}_i = \begin{bmatrix} \tilde{c}_i^f(1) \\ \vdots \\ \tilde{c}_i^f(k_i) \end{bmatrix}, \quad \mathbf{h}_i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

gdzie

$$\begin{aligned} w_i^f(k) &= \delta^f(t_0 + \tau_i^f + k), \\ e_i(k) &= \varepsilon(t_0 + \tau_i^f + k | t_0 + \tau_i^f + k - 1), \\ \eta_i(k) &= n(t_0 + \tau_i^f + k), \\ k &= 1, \dots, k_i. \end{aligned} \quad (7.25)$$

Używając wyżej wprowadzonej notacji wektorowej, zależność (7.24) można przedstawić w postaci

$$\mathbf{w}_i^f = \alpha_i \mathbf{c}_i + \beta_i \mathbf{h}_i. \quad (7.26)$$

Zauważmy, że w zgodzie z zależnością (7.12)

$$\mathbf{e}_i = \boldsymbol{\eta}_i + \mathbf{w}_i^f. \quad (7.27)$$

Na podstawie (7.26) i (7.27), optymalna w sensie średniokwadratowym ocena współczynników α_i i β_i może być otrzymana na drodze minimalizacji kwadratowej funkcji kosztu

$$J_i(\alpha_i, \beta_i) = \| \mathbf{e}_i - \alpha_i \mathbf{c}_i - \beta_i \mathbf{h}_i \|^2. \quad (7.28)$$

Niech

$$\{\hat{\alpha}_i, \hat{\beta}_i\} = \arg \min_{\alpha_i, \beta_i} J_i(\alpha_i, \beta_i). \quad (7.29)$$

Można łatwo pokazać, że

$$\hat{\alpha}_i = \frac{f_1 f_5 - f_3 f_4}{f_1 f_2 - f_3^2}, \quad \hat{\beta}_i = \frac{f_2 f_4 - f_3 f_5}{f_1 f_2 - f_3^2}, \quad (7.30)$$

gdzie współczynniki f_1, \dots, f_5 otrzymywane są z zależności

$$\begin{aligned} f_1 &= \| \mathbf{h}_i \|^2 = k_i, \\ f_2 &= \| \mathbf{c}_i \|^2 = \sum_{k=1}^{k_i} [\tilde{c}_i^f(k)]^2, \\ f_3 &= \mathbf{h}_i^T \mathbf{c}_i = \sum_{k=1}^{k_i} \tilde{c}_i^f(k), \\ f_4 &= \mathbf{h}_i^T \mathbf{e}_i = \sum_{k=1}^{k_i} e_i(k), \\ f_5 &= \mathbf{c}_i^T \mathbf{e}_i = \sum_{k=1}^{k_i} \tilde{c}_i^f(k) e_i(k). \end{aligned} \quad (7.31)$$

Ponieważ wzorce wtórne są unormowane, otrzymujemy $f_2 = 1$ i $f_3 = 0$, co prowadzi do zależności

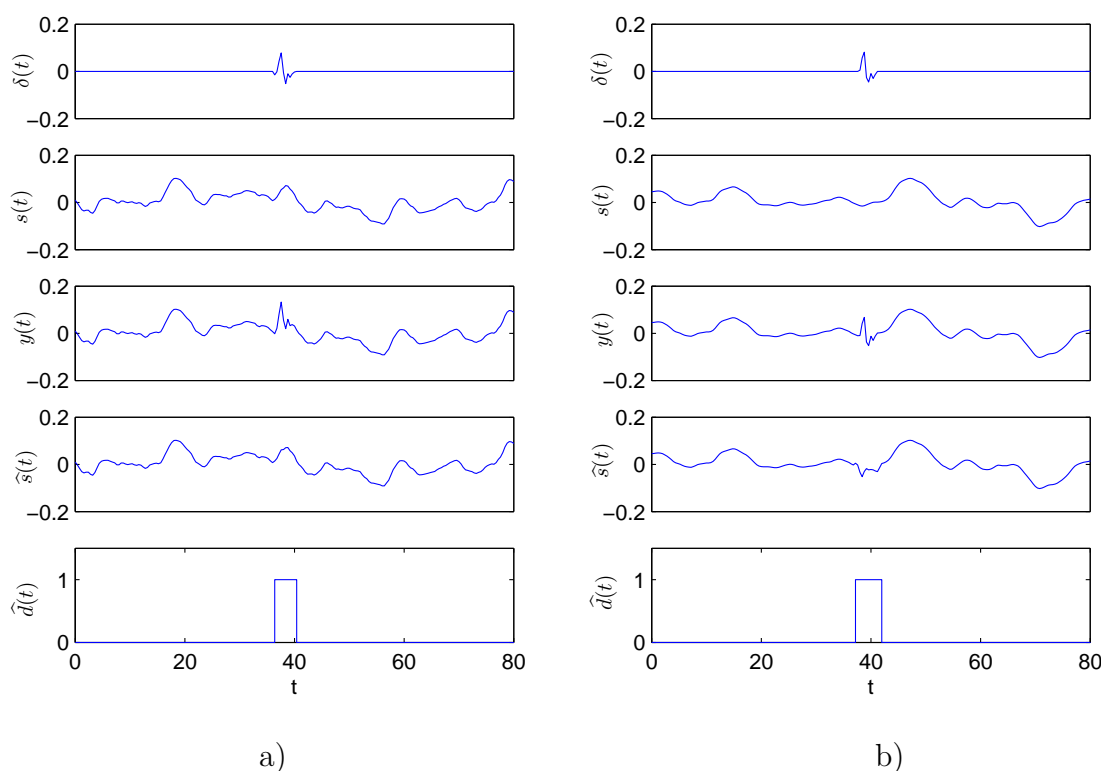
$$\hat{\alpha}_i = f_5, \quad \hat{\beta}_i = \frac{f_4}{k_i}. \quad (7.32)$$

Używając powyższych ocen otrzymujemy następujące oceny wtórnych zakłóceń impulsowych

$$\hat{w}_i^f(k) = \hat{\delta}_i^f(t_0 + \tau_i^f + k) = \hat{\alpha}_i \tilde{c}_i^f(k) + \hat{\beta}_i, \quad k = 1, \dots, k_i. \quad (7.33)$$

Ostatecznie, aby otrzymać oceny pierwotnych wzorców zakłóceń

$$\hat{w}_i(k) = \hat{\delta}_i(t_0 + \tau_i^f + k), \quad k = 1, \dots, m_i, \quad (7.34)$$



Rysunek 7.3: Przykład udanej (a) i nieudanej (b) eliminacji zakłócenia impulsowego za pomocą techniki kompensacji. W każdej grupie kolejne wykresy pokazują (od góry do dołu): przebieg zakłócenia impulsowego, czysty sygnał foniczny, zakłócony sygnał foniczny, zrekonstruowany sygnał foniczny oraz oszacowane położenie zakłócenia impulsowego.

należy cofnąć zmiany wprowadzone przez filtr odwrotny. Rekurencyjna formuła pozwalająca to uczynić, ma postać

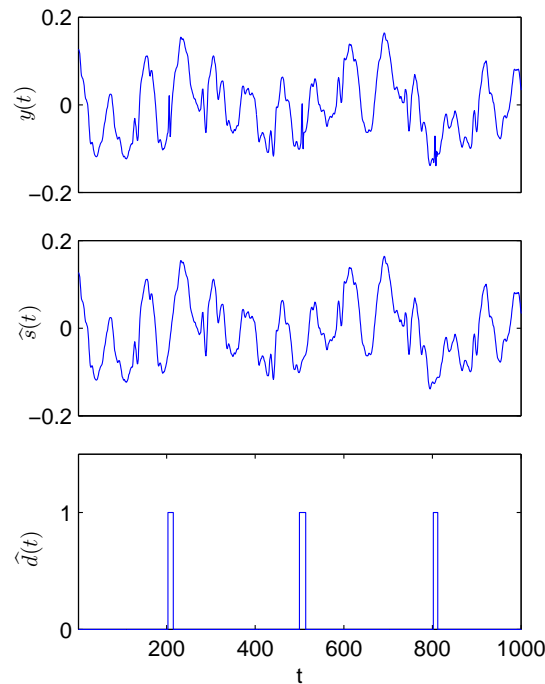
$$\hat{w}_i(k) = \hat{w}_i^f(k) + \sum_{j=1}^{m_i} \hat{w}_i(k-j), \quad k = 1, \dots, m_i, \quad (7.35)$$

gdzie $\hat{w}_i(k) = 0$ dla $k \leq 0$.

Opisana procedura powinna być zastosowana do najlepiej pasującego wzorca wtórnego $\tilde{C}_{i_0}^f$. Równanie kompensacji przyjmuje postać

$$\hat{s}(t) = y(t) - \hat{\delta}_{i_0}(t), \quad t \in D. \quad (7.36)$$

Główną wadą kompensacji jest to, że często pozostawia ona w miejscach rekonstrukcji niewielkie lecz słyszalne artefakty, co prowadzi do ogólnego pogorszenia jakości dźwięku zrekonstruowanego materiału fonicznego – patrz rysunek 7.3. Artefakty te pojawiają się, ponieważ kształt zakłócenia impulsowego przypomina, ale też zazwyczaj nieznacznie różni się od kształtu najlepiej dopasowanego wzorca. Dlatego zalecane jest przeprowadzenie rekonstrukcji uszkodzonego fragmentu sygnału na drodze interpolacji. Na rysunku 7.4 przedstawiono typowe wyniki detekcji/interpolacji uzyskane dla proponowanego podejście.



Rysunek 7.4: Typowe wyniki detekcji/interpolacji uzyskane dla proponowanego podejścia. Zakłócenia impulsowe dodane zostały do czystego sygnału w chwilach czasu 200, 500 i 800.

7.4 Podsumowanie

Proponowana w tym rozdziale technika lokalizacji zakłóceń impulsowych oparta jest na obserwacji, że większość zakłóceń impulsowych degradujących archiwalne nagrania posiada wysoce powtarzający się kształt, który pasuje do względnie małej liczby typowych kształtów zakłóceń, tzw. wzorców zakłóceń impulsowych. Wzorce te mogą być tworzone na podstawie zbioru przykładowych przebiegów zakłóceń wyciętych z fragmentów ciszy poprzedzających utwory foniczne. Aby wykryć typowy kształt zakłócenia w analizowanym sygnale, wtórne wzorce korelowane są z sekwencją błędów jednokrokowej predykcji generowaną przez predyktor oparty na modelu AR. Na podstawie wyników symulacyjnych – zamieszczonych w rozdziale 8 – pokazano, że gdy tradycyjne metody detekcji połączone zostaną z tą selektywną techniką lokalizacji wyniki detekcji i rekonstrukcji ulegają poprawie.

Rozdział 8

Wyniki eksperymentów

Na początku rozdziału omówiono przyjętą w pracy metodologię przeprowadzania eksperymentów oraz metodologię oceny otrzymywanych wyników. Następnie, przedstawiono wyniki eksperymentów z wykorzystaniem algorytmów eliminacji zakłóceń impulsowych opisanych w poprzednich rozdziałach oraz z wykorzystaniem algorytmu detekcji opartego na dwóch progach, opisanego w pracy [50].

8.1 Metodologia

Aby wiarygodnie ocenić skuteczność opracowanych algorytmów przeprowadzono liczne testy symulacyjne z wykorzystaniem sztucznie spreparowanych nagrań oraz nagrań archiwalnych odczytanych z płyt gramofonowych.

W pierwszym przypadku nagrania testowe uzyskano poprzez dodanie do czystych sygnałów fonicznych sekwencji rzeczywistych zakłóceń impulsowych, „wyciętych” z fragmentów ciszy archiwalnego nagrania. W ten sposób w przeprowadzanych eksperymentach dostępne były: czyste nagranie wzorcowe, nagranie zakłócone oraz pełna wiedza *a priori* o zakłóceniach impulsowych, tj. ich dokładna lokalizacja w sygnale, kształt, amplituda oraz częstotliwość występowania. Umożliwiło to przeprowadzenie symulacji, których wyniki mogły zostać ocenione przy użyciu obiektywnych, pośrednich i bezpośrednich, miar jakości dźwięku. Pośrednie miary dostarczyły informacji w postaci statystyk opisujących skuteczność algorytmów detekcji, natomiast bezpośrednia miara dostarczała informacji o jakości zrekonstruowanego sygnału.

W drugim przypadku archiwalne nagranie pełniło rolę nagrania wzorcowego, stąd jedyną możliwością oceny uzyskanych wyników było przeprowadzenie testów odsłuchowych. W teście poproszono 20 słuchaczy o ocenę jakości eliminacji zakłóceń impulsowych z archiwalnych nagrań fonicznych.

8.1.1 Nagrania testowe

W przeprowadzonych testach symulacyjnych posłużono się zarówno nagraniami sztucznie spreparowanymi jak archiwalnymi nagraniami gramofonowymi.

Sztucznie spreparowane nagrania

Sztuczne dodawanie zakłóceń impulsowych do czystych nagrań fonicznych daje możliwość obiektywnej oceny skuteczności algorytmów detekcji i rekonstrukcji. Zakłócenia te mogą zostać zamodelowane bądź „wycięte” z archiwalnego nagrania.

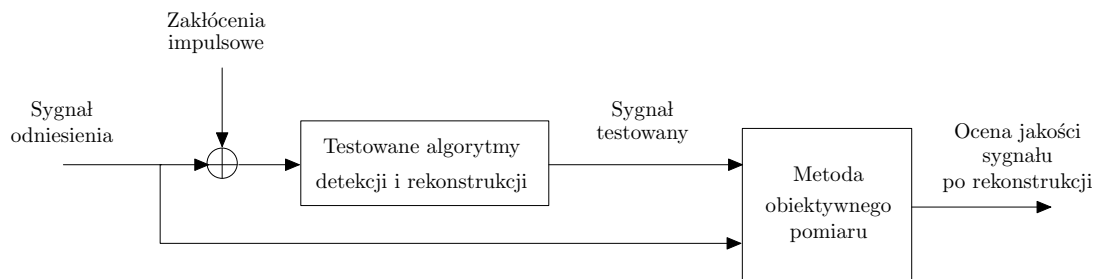
W pierwszym przypadku zakłócenia mogą być reprezentowane przez pojedyncze impulsy o dużej amplitudzie i o losowym miejscu wystąpienia. Mimo iż tak zakłócone nagranie brzmi podobnie do archiwalnego nagrania, rzeczywiste zakłócenia impulsowe są o wiele bardziej wymagające dla algorytmów detekcji i rekonstrukcji, niż pojedyncze impulsy.

W drugim przypadku zakłócenia impulsowe zostają „wycięte” z archiwalnych nagrań fonicznych. Procedura „wycinania” zakłóceń impulsowych może być zastosowana zarówno do fragmentów ciszy poprzedzających utwory muzyczne i/lub występujących bezpośrednio po nich. Na początku fragment takiego nagrania przetworzony zostaje przy użyciu algorytmu do eliminacji zakłóceń impulsowych lub przy użyciu gotowego komercyjnego pakietu do rekonstrukcji nagrań fonicznych. Następnie wyznacza się różnicę pomiędzy oryginalnym sygnałem a jego przetworzoną wersją, aby znaleźć lokalizację i kształt zakłóceń impulsowych. Na koniec, przeprowadza się sprawdzenie polegające na bezpośredniej wizualnej ocenie wspomnianych wyżej sygnałów, aby wyeliminować oczywiste błędy wynikające z błędów detekcji lub niskiej jakości interpolacji – odpowiednie przebiegi zakłóceń są wtedy usuwane. W ten sposób uzyskuje się nagranie zawierające wyizolowane zakłócenia impulsowe. Zebrane zakłócenia mogą być dalej podzielone ze względu na ich kształt, długość bądź wielkość.

Aby przeprowadzić testy symulacyjne należy do czystego sygnału sztucznie wprowadzić (dodać) sekwencję zakłóceń impulsowych. Zakłócenia te mogą być dodawane do sygnału w losowych odstępach bądź w niewielkich, ale jednakowych odstępach. W drugim przypadku popełniane w seriach błędy detekcji/rekonstrukcji stają się wyraźniej słyszalne niż gdyby zastosować nierównomierne (losowe) rozłożenie zakłóceń w sygnale. Równomierny sposób dodawania zakłóceń sprawia, że spreparowane nagrania brzmią nieco mniej naturalnie, ale za to są bardziej wymagające – uzyskanie wysokich ocen jakości dźwięku wymaga prawie bezbłędnego przetworzenia takiego nagrania.

Zanim przygotowana sekwencja zakłóceń impulsowych zostanie dodana, wszystkie sygnały foniczne powinny zostać odpowiednio przeskalowane w taki sposób, aby stosunek sygnału do „szumu” w zakłóconych fragmentach był identyczny dla wszystkich sygnałów. Choć wszystkie spreparowane nagrania powstały przez dodanie do czystych nagrań tej samej sekwencji przykładowych zakłóceń impulsowych, niektóre zakłócenia mogą być łatwiejsze bądź trudniejsze do wykrycia i zlokalizowania, ponieważ analizowane materiały foniczne stanowią różne „tło” muzyczne dla zakłóceń impulsowych.

Przeprowadzono również próby z dodawaniem szumu szerokopasmowego, aby testowe sygnały lepiej przypominały te rzeczywiste. W tym przypadku przyjmowano, że czysty sygnał zakłócony szumem szerokopasmowym pełni rolę sygnału wzorcowego. Niestety, zabieg ten nie przyniósł oczekiwanych korzyści. Otrzymywane rekonstrukcje sygnału nawet, przy pełnej wiedzy o położeniu zakłóceń impulsowych, oceniane były nisko przez stosowany algorytm do obiektywnej oceny jakości dźwięku. Powód tego był oczywisty – otrzymywane interpolacje oparte na autoregresyjnym modelu sygnału nie uwzględniały składowej losowej, przez co algorytm oceniający



Rysunek 8.1: *Koncepcja obiektywnego pomiaru jakości dźwięku.*

jakość dźwięku traktował je jako „nieudane”. Dlatego zabieg ten pominięto w przygotowywaniu spreparowanych nagrań testowych.

Archiwalne nagrania gramofonowe

Mimo iż badania z użyciem spreparowanych nagrań testowych mają pewne niepodważalne zalety – dają możliwość uzyskania wyników przy pełnej wiedzy o położeniu zakłóceń impulsowych, a także otrzymania ocen jakościowych i ilościowych przy użyciu obiektywnych miar jakości – końcowe testy powinny być przeprowadzane zawsze przy użyciu rzeczywistych nagrań archiwalnych. Nagrania te są zazwyczaj bardziej wymagające, ponieważ oprócz zakłóceń impulsowych zawierają szum szerokopasmowy, taki jak szum powierzchniowy płyt gramofonowych.

Posiadany zbiór liczył około 80 archiwalnych nagrań (całe utwory lub ich początkowe fragmenty) pochodzących z monofonicznych oraz z stereofonicznych płyt gramofonowych. W zbiorze tym znalazły się również silnie zakłócone nagrania pochodzące ze starych płyt wykonanych z szelaku, dla których typowymi zakłóceniami były trzaski oraz kliki. Materiał foniczny obejmował sygnały mowy (audycje radiowe) oraz sygnały muzyczne (arie operowe, utwory muzyki klasycznej, pieśni chóralne, utwory bluesowe oraz jazzowe z udziałem śpiewu i instrumentów dętych). Sygnały te charakteryzowały się dużą intensywnością i różnorodnością zakłóceń impulsowych od małych i cichych po duże i głośne. W tej różnorodności można doszukać się wszystkich typów zakłóceń impulsowych takich jak tiki, trzaski oraz kliki. W nieformalnych testach odsłuchowych korzystano ze wszystkich tych nagrań, aby ocenić skuteczność opracowanych algorytmów. Niestety, ze względu na ograniczenia formalnych testów odsłuchowych, ocenie formalnej poddano tylko kilkanaście wybranych archiwalnych nagrań.

8.1.2 Obiektywne miary jakości

Koncepcja obiektywnego pomiaru jakości została zilustrowana na rysunku 8.1. Do czystego sygnału dodane zostają zakłócenia impulsowe. Tak spreparowany sygnał podlega następnie adaptacyjnej detekcji i rekonstrukcji, której wynikiem jest sygnał testowany. W celu otrzymania oceny jakości sygnału testowanego porównuje się go z sygnałem odniesienia przy użyciu obiektywnych metod pomiaru jakościowego oraz ilościowego. Miary jakościowe dostarczają informacji w bezpośredni sposób o jakości zrekonstruowanego sygnału, zaś miary ilościowe – informacji w postaci statystyk opisujących skuteczność algorytmów detekcji.

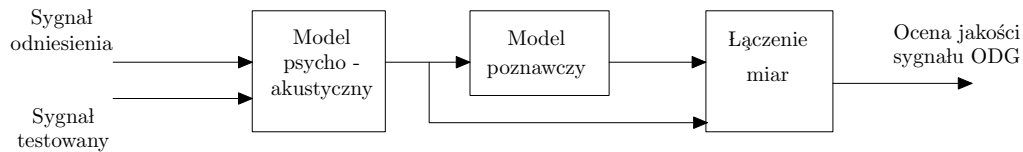
Miary jakościowe

Tradycyjne metody obiektywnego pomiaru jakości takie jak stosunek sygnału do szumu, współczynnik zawartości harmonicznych (z ang. *Total Harmonic Distortion*) czy błąd średniokwadratowy (z ang. *Mean Square Error*) nie korelują się dobrze z ocenami subiektywnymi i dlatego nie są użyteczne w tym zastosowaniu. Bardziej zaawansowane metody obiektywnego pomiaru zostały zaprojektowane z przeznaczeniem do zastosowań w aplikacjach telekomunikacyjnych. Dobrze znany algorytm PESQ (z ang. *Perceptual Evaluation of Speech Quality*) może być używany do oceny jakości wąskopasmowych sygnałów mowy zdegradowanych przez błędy systemów kodowania, błędy transmisji, utratę pakietów, pojawiające się opóźnienia czy szum otoczenia [97]. Niestety, algorytm PESQ nie może być używany do oceny jakości algorytmów do redukcji szumów [98]. W przypadku głośnych, krótkich zakłóceń impulsowych wyniki algorytmu nie korelują się dobrze z ocenami subiektywnymi. Nowsze badania nad problemem obiektywnej oceny jakości dźwięku w przypadku obecności w sygnale niepożądanych efektów dźwiękowych znaleźć można w pracy [99], niestety zakłócenia impulsowe nie występują wśród rozważanych typów degradacji sygnału. Dalsze prace nad rozwojem algorytmu PESQ doprowadziły do powstania nowego, aktualnego standardu POLQA (z ang. *Perceptual Objective Listening Quality Assessment*) opisanego w rekomendacji [100], lecz również nie jest on odpowiedni do oceny jakości algorytmów do redukcji szumów. Inny, dobrze znany algorytm PEAQ (z ang. *the Perceptual Evaluation of Audio Quality tool*) został zaprojektowany do obiektywnej oceny jakości przetworzonych szerokopasmowych sygnałów muzycznych (próbekowanych z częstotliwością 48 kHz), np. do oceny jakości sygnałów fonicznych po kompresji [101]. Algorytm PEAQ wykorzystuje liczne psychoakustyczne miary, które po połączeniu w odpowiedni sposób dają obiektywny pomiar stopnia zniekształcenia (z ang. *Objective Difference Grade*, ODG) sygnału testowanego w porównaniu z sygnałem wzorcowym. Wynik PEAQ przyjmuje ujemne wartości z przedziału od -4 do 0 , a jego interpretacja została przedstawiona w tabeli 8.1. Metoda ta okazała się użyteczna w rozważnym zastosowaniu, ponieważ generowane wyniki korelują się dobrze z ocenami subiektywnymi. W rozprawie wykorzystano tę metodę do bezpośredniej oceny skuteczności algorytmów detekcji i rekonstrukcji w przypadku przetwarzania sygnałów muzycznych. W przypadku przetwarzania sygnałów mowy, zrekonstruowane sygnały oceniane były wyłącznie w testach odsłuchowych.

Zniekształcenia	Ocena
niesłyszalne	0
słyszalna ale nie przeszkadzające	-1
lekko przeszkadzające	-2
przeszkadzające	-3
bardzo przeszkadzające	-4

Tablica 8.1: Skala obiektywnej oceny degradacji sygnałów fonicznych.

Metoda PEAQ oparta jest na modelu działania ludzkiego narządu słuchu oraz na modelu poznawczym reprezentującym sposób, w jaki mózg ocenia jakość słyszanego dźwięku. Ogólny model obiektywnej metody pomiaru PEAQ przedstawia rysunek 8.2. Zadaniem modelu psychoakustycznego jest przetworzenie sygnału w taki sposób, w jaki dźwięk przetwarzany jest przez ludzkie ucho. Najpierw wyznacza się reprezentację sygnału w dziedzinie częstotliwości przy użyciu dyskretnej transformaty Fouriera. Następnie, widmo sygnału zostaje wstępnie przetworzone tak aby uwzględnić charakterystykę częstotliwościową ucha oraz charakterystykę częstotli-



Rysunek 8.2: Model metody obiektywnego pomiaru jakości dźwięku (PEAQ).

ściową szumu spowodowanego przepływem krwi. W kolejnych krokach uwzględnia się zjawisko maskowania, progowania oraz zjawisko różnego postrzegania wysokości dźwięku na skali częstotliwości. Na koniec wyselekcjonowane zostają cechy widmowe ułatwiające dalsze badanie różnic występujących pomiędzy dwoma sygnałami. Z kolei zadaniem modelu poznawczego jest przetworzenie informacji, które dostarcza model psychoakustyczny. W wyniku przetwarzania otrzymuje się zestaw miar charakteryzujących słyszalne różnice w porównywanych sygnałach, np. zmiany w postrzeganej głośności. Na koniec parametry te zostają przetworzone przy użyciu wytrenowanej sieci neuronowej w celu wyznaczenia obiektywnej oceny jakości dźwięku. W celu wytrenowania takiej sieci przeprowadzono subiektywne testy odsłuchowe w bardzo licznej grupie słuchaczy, a następnie na drodze nadzorowanego uczenia przy użyciu algorytmu wstecznej propagacji błędu otrzymano optymalny zestaw wag, dający wyniki dobrze skorelowane z ocenami subiektywnymi.

Dostępne są dwie wersje implementacji algorytmu PEAQ: wersja podstawowa oraz wersja rozszerzona. W wersji podstawowej model ucha ludzkiego oparty został na transformacji FFT, natomiast w wersji rozszerzonej połączono ze sobą dwa różne modele oparte odpowiednio na transformacji FFT oraz na banku filtrów. Wersja rozszerzona wykorzystuje dokładniejszy opis modelu ucha ludzkiego, co umożliwiłoby uzyskiwanie nieco dokładniejszych wyników. Dzieje się to jednak kosztem zwiększenia złożoności obliczeniowej metody. W zależności od zastosowania stosuje się odpowiednią wersję metody, przy czym tylko wersja podstawowa umożliwia pomiar jakości sygnału w czasie rzeczywistym. W pracy wykorzystano implementację metody PEAQ w wersji podstawowej, która została rzetelnie przedyskutowana i opisana w dokumentacji [102]. Algorytm ten pozwala na analizę sygnałów fonicznych o częstotliwości próbkowania 48 kHz, nagranych w systemie monofonicznym oraz w systemie stereofonicznym.

Po przeprowadzeniu licznych testów odsłuchowych, mających na celu sprawdzenie użyteczności metody PEAQ w kontekście obiektywnej oceny jakości sygnałów otrzymywanych w wyniku eliminacji zakłóceń impulsowych, poczyniono pewne spostrzeżenia. W przypadku oceny jakości dźwięku w rekonstruowanym sygnale na podstawie otrzymanego wyniku testu PEAQ, należy zachować pewną ostrożność w jego interpretacji. Podczas gdy w aplikacjach telekomunikacyjnych zniekształcenia sygnału są rozciągnięte w czasie, w rozważanym przypadku obejmują one tylko odizolowane fragmenty materiału fonicznego. Oznacza to, że aby degradacja sygnału mogła być interpretowana jako np. „słyszalna ale nieprzeszkadzająca”, ocena PEAQ musi być znacznie wyższa (wartości bliższe zera) niż w przypadku zastosowań telekomunikacyjnych. Zauważono, że gdy oceny PEAQ uzyskane dla sygnału testowanego przekraczają próg $-0,1$, zniekształcenia sygnału mogą być traktowane jako niesłyszalne. Podobnie, zauważono, że pomiędzy dwoma sygnałami testowanymi będącymi wynikiem zastosowania różnych algorytmów detekcji i/lub rekonstrukcji słyszalna różnica jakości dźwięku występuje dopiero, gdy odpowiadające tym sygnałom oceny PEAQ różnią się od siebie o co najmniej $0,1$. Ustalona rozdzielczość metody pomiaru pokrywa się więc z tą podaną w rekomendacji

$$|\text{PEAQ}_1 - \text{PEAQ}_2| \geq 0,1. \quad (8.1)$$

Miary ilościowe

W pracy zaproponowano dwie pośrednie miary skuteczności algorytmów detekcji: miarę podobieństwa wykryć oraz miarę pokrycia energetycznego zakłóceń.

Pierwsza miara pomaga ustalić w jakim stopniu oszacowane lokalizacje zakłóceń impulsowych pokrywają się z ich rzeczywistym położeniem w sygnale. Rozważmy dwa alarmy detekcyjne: oszacowanie lokalizacji zakłócenia $\hat{D}_t = [\hat{t}_B, \hat{t}_E]$ oraz jego dokładną lokalizację $D_t = [t_B, t_E]$. W przypadku gdy oba alarmy pokrywają się częściowo $\hat{D}_t \cap D_t \neq \emptyset$, stopień ich podobieństwa może być zmierzony przy użyciu następującego wskaźnika

$$s_t = \frac{\min(t_E, \hat{t}_E) - \max(t_B, \hat{t}_B) + 1}{\max(t_E, \hat{t}_E) - \min(t_B, \hat{t}_B) + 1} \quad [\%]. \quad (8.2)$$

Wskaźnik s_t osiąga maksymalną wartość, równą 100%, gdy wykrycie jest bezbłędne, tj. $\hat{t}_B = t_B$ i $\hat{t}_E = t_E$. We wszystkich innych przypadkach współczynnik ten przyjmuje wartości mniejsze niż 100%. Im jego wartość jest mniejsza, tym większa rozbieżność w lokalizacji i w rozmiarze porównywanych binarnych sekwencji detekcyjnych. W ogólnym przypadku stopień podobieństwa jest wyznaczany na drodze uśredniania wyników otrzymanych dla wszystkich alarmów detekcyjnych.

Kolejną miarą jest wskaźnik pokrycia energetycznego zakłóceń, wyrażony w następujący sposób

$$c_t = \frac{\sum_{t \in T_{\hat{d}}} \delta^2(t)}{\sum_{t \in T_d} \delta^2(t)} \quad [\%], \quad (8.3)$$

gdzie $T_{\hat{d}} = \{t : \hat{d}(t) = 1 \wedge d(t) = 1\}$, $T_d = \{t : d(t) = 1\}$. Jest to miara względna zależna od całkowitej energii dodanych do sygnału zakłóceń impulsowych. Wskaźnik ten osiąga najwyższy wynik procentowy w sytuacji, gdy wszystkie zakłócone próbki sygnału wykryte zostały przez detektor zakłóceń impulsowych.

Przy próbie interpretacji wskaźników należy analizować je łącznie, ponieważ wskaźnik pokrycia energetycznego może być łatwo zwiększony przez zwykłe poszerzenie alarmów detekcyjnych. Jednoczesne zwiększenie obu wskaźników może być jednak osiągnięte tylko dzięki dokładniejszej lokalizacji zakłóceń impulsowych.

8.1.3 Testy odsłuchowe

Specyfika testu odsłuchowego przeprowadzanego w oparciu o archiwalne nagrania dźwiękowe polega na tym, że czyste nagranie wzorcowe nie jest dostępne. W tym przypadku rolę nagrania wzorcowego pełni nagranie, w którym występują nie tylko zakłócenia impulsowe lecz również szum szerokopasmowy. O ile łatwo jest ustalić fakt usunięcia zakłóceń impulsowych, o tyle dużo trudniej jest ustalić czy zauważone niepożądane efekty dźwiękowe w przetworzonym nagraniu są wynikiem obróbki cyfrowej czy też może były już obecne w nagraniu oryginalnym. Podjęcie takich decyzji wymaga od słuchacza dużego skupienia i jest czasochłonne. Dodatkowym czynnikiem utrudniającym proces podejmowania decyzji jest obecność szumu szerokopasmowego, który może wywierać wpływ na subiektywną ocenę słuchacza. Podczas oceniania skuteczności algorytmu służącego do usuwania zakłóceń impulsowych słuchacz musi zadać sobie następujące pytania:

- czy algorytm wpływa na poprawę subiektywnie odczuwanej jakości sygnału?

- czy zakłócenia impulsowe zostały całkowicie wyeliminowane, czy też może jedynie ich liczba została zredukowana?
- czy pojawiły się nowe, niepożądane efekty dźwiękowe będące wynikiem cyfrowej obróbki sygnału?
- czy osiągnięto akceptowalny kompromis uwzględniający wszystkie przedstawione powyżej kryteria oceny?

Odpowiedź na te pytania i podjęcie subiektywnej decyzji o skuteczności algorytmu nie jest zadaniem prostym. Jeszcze większy problem pojawia się gdy należy jednocześnie ocenić oraz porównać ze sobą dwa lub więcej algorytmów. W literaturze przedmiotu nie opracowano do tej pory rekomendacji, która opisywałaby zalecany sposób przygotowywania i wykonywania testów odsłuchowych w przypadku eliminacji zakłóceń impulsowych. Dlatego scenariusz testów wzorowany był na wytycznych z dziedziny telekomunikacji, opisanych szczegółowo w rekomendacjach ITU-R BS.1534-2 [103] i ITU-R BS.1116 [104].

Podczas każdego testu, każda z 20 osób poproszona została o ocenienie porównywanych nagrań (przetworzonych nagrań oraz nagrania oryginalnego) w skali od 0 do 100. Przedział ocen podzielony został na 5 równych podprzedziałów ([0,20],[21,40], itd.) z opisem podanym w tabeli 8.2. Przed rozpoczęciem właściwej części badania słuchaczy zapoznano z próbkami ocenianego materiału fonicznego. Słuchacze poproszeni zostali, aby w trakcie oceniania skuteczności usuwania zakłóceń impulsowych ignorowali szerokopasmowy szum powierzchniowy, obecny we wszystkich nagraniach. Kolejność nagrań testowych była za każdym razem wybierana w sposób losowy. Wszystkie przesłuchania odbyły się przy użyciu tego samego zestawu audio wyposażonego w wysokiej klasy słuchawki Beyerdynamic DT 770 PRO i dedykowany do nich wzmacniacz HEADPOD 4TM firmy APHEX. Aby ustalić końcową ocenę słuchacze mogli wielokrotnie odsłuchiwać całe nagrania lub ich wybrane fragmenty.

Rozważane były również inne perceptualne testy. Test MOS (z ang. *Mean Opinion Score*) często stosowany do oceny efektów usuwania szumu szerokopasmowego [59], został celowo pominięty ponieważ daje niespójne wyniki w przypadku oceniania efektów eliminacji zakłóceń impulsowych. Z kolei test MUSHRA (z ang. *Multi Stimulus test with Hidden Reference and Anchor*) często stosowany do oceny efektów kompresji [103], nie może być stosowany w sytuacji gdy nie jest znane czyste nagranie wzorcowe lub gdy porównywane nagrania są zbliżonej jakości. Obie sytuacje mają miejsce w naszym przypadku.

Jakość	Przedział ocen
doskonała	81 – 100
dobra	61 – 80
średnia	41 – 60
słaba	21 – 40
zła	0 – 20

Tablica 8.2: Skala subiektywnej oceny jakości usuwania zakłóceń impulsowych.

8.2 Testy symulacyjne

W tym podrozdziale przedstawiono wyniki eksperymentów z użyciem opisanych w pracy podejść do eliminacji zakłóceń impulsowych, którymi są: podejście wykorzystujące analizę dwukierunkową, podejście oparte na dwuzakresowym modelowaniu autoregresyjnym, podejście oparte na wektorowym modelowaniu autoregresyjnym oraz podejście wykorzystujące technikę dopasowywania wzorców. Otrzymane wyniki oceniono przy użyciu miar obiektywnych oraz w testach odsłuchowych. Algorytm, który okazał się najlepszy w końcowym porównaniu, został użyty do przetworzenia rzeczywistych archiwalnych nagrań gramofonowych, a otrzymane rekonstrukcje porównano z analogicznymi wynikami uzyskanymi przy użyciu komercyjnego systemu CEDAR przeznaczonego do redukcji zakłóceń impulsowych.

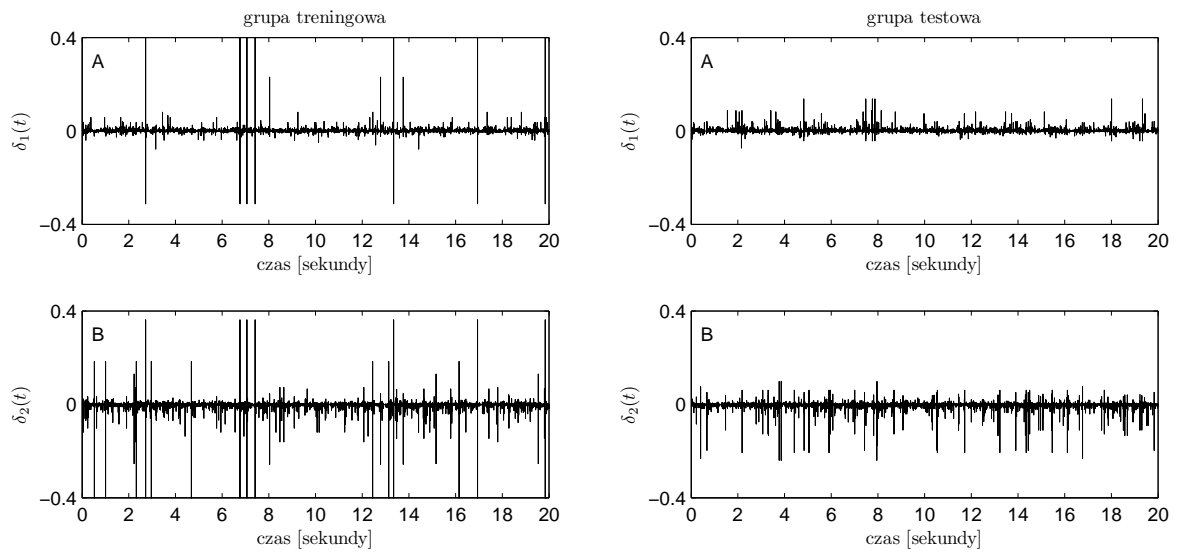
Aby rzetelnie porównać skuteczność działania pierwszych trzech podejść niewykorzystujących wiedzy *a priori* o typowych kształtach zakłóceń impulsowych, przygotowano jedną wspólną bazę 50 czystych nagrań stereofonicznych oraz dwa różne szablony zakłóceń impulsowych, użyte odpowiednio w fazie treningowej oraz w fazie testowej. Do otrzymania wszystkich wyników użyto tych samych parametrów symulacji. Wprawdzie prezentowane wyniki różnią się liczbowo od tych opublikowanych w referatach konferencyjnych oraz artykułach naukowych, ale wnioski płynące z analizy otrzymanych wyników są identyczne z tymi opublikowanymi. W przypadku podejścia wykorzystującego technikę dopasowywania wzorców zamieszczono wyniki zgodne z tymi opublikowanymi w artykule naukowym. Należy zwrócić uwagę, że w tym przypadku użyta baza nagrań testowych, baza zakłóceń impulsowych oraz parametry symulacji były inne niż dla pozostałych ocenianych podejść.

Spreparowane nagrania

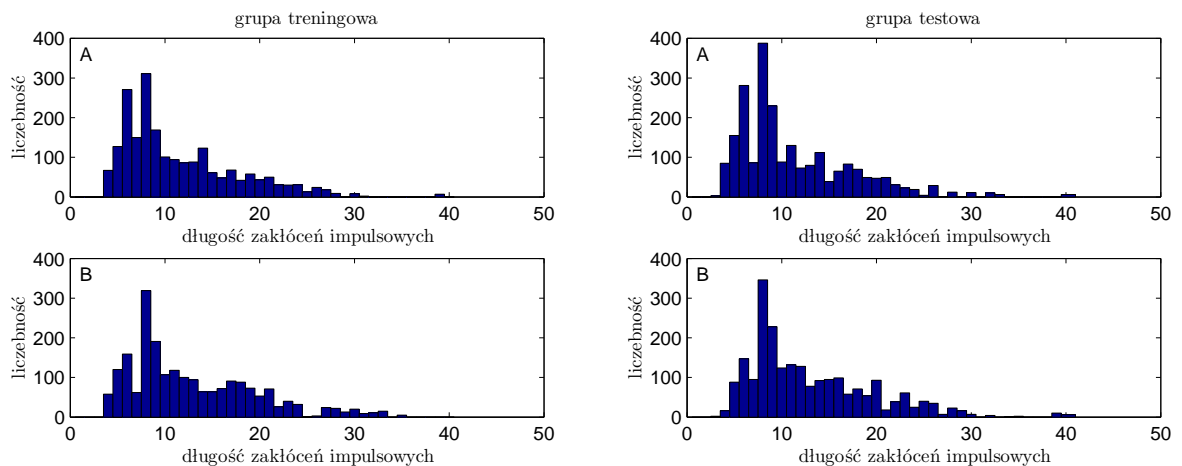
Baza nagrań testowych składała się z 50 czystych nagrań stereofonicznych: 10 z muzyką jazzową (wokalną/instrumentalną), 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną (z dominującą obecnością takich instrumentów jak kontrabas, klarnet czy skrzypce) oraz 10 z muzyką klasyczną, próbkowanych z częstotliwością 48 kHz z 16-bitową rozdzielczością. Wszystkie nagrania obejmowały 22 sekundy materiału fonicznego. Zanim dodano zakłócenia impulsowe, wszystkie sygnały foniczne zostały przeskalowane tak aby ich średnia moc była jednakowa. Następnie nagrania te spreparowane zostały poprzez dodanie rzeczywistych zakłóceń impulsowych „wyciętych” uprzednio z cichych partii stereofonicznego nagrania gramofonowego zawierającego etiudy Fryderyka Chopina op. 25 nr 8, 9 i 10 (przy częstotliwości próbkowania 48 kHz). Zbiór przykładowych zakłóceń impulsowych obejmował 1420 par przebiegów, znalezionych odpowiednio w lewym i prawym kanale nagrania stereofonicznego. Zbiór ten został podzielony w sposób losowy na dwa mniejsze podzbiory \mathcal{P}_A i \mathcal{P}_B zawierające po $n_t = 710$ par przykładowych zakłóceń impulsowych, które użyte zostały odpowiednio w fazie treningowej i fazie testowej.

Aby przygotować bazę nagrań użytych na etapie uczenia, do przeskalowanych sygnałów dodano 20 sekundowy szablon zakłóceń (taki sam dla wszystkich nagrań testowych) składający się z 2867 par zakłóceń impulsowych, losowo oddalonych od siebie i losowo wybranych z bazy zakłóceń \mathcal{P}_A , gdzie: 742 zakłóceń impulsowych uszkadzało tylko lewy kanał, 737 zakłóceń impulsowych uszkadzało tylko prawy kanał i 1388 zakłóceń impulsowych uszkadzało oba kanały jednocześnie. Prawdopodobieństwo wystąpienia pary zakłóceń impulsowych wynosiło $1/300$, przy czym zachowana została minimalna odległość między kolejnymi zakłóceniami – wynosząca r niezakłóconych





Rysunek 8.3: Szablony zakłóceń impulsowych wykorzystane do utworzenia grupy nagrań treningowych (szablony A i B od lewej) oraz grupy nagrań testowych (szablony A i B od prawej). Szablon A dodawany jest do lewego kanału sygnału stereofonicznego, a szablon B – do prawego kanału.



Rysunek 8.4: Histogramy przedstawiają liczebność zakłóceń impulsowych o długości z przedziału od 1 do 50 próbek. Histogramy A i B (od lewej strony) odpowiadają szablonowi zakłóceń A i B z grupy treningowej, natomiast histogramy A i B (od prawej strony) odpowiadają szablonowi zakłóceń A i B z grupy testowej. Całkowita liczba zakłóceń impulsowych w grupie treningowej wynosi 4255, zaś w grupie testowej wynosi 4506.

próbek. Całkowita liczba uszkodzonych próbek pochodzących z lewego i prawego kanału wyniosła 50731, co stanowi 2,64% wszystkich próbek w zdegradowanych 20 sekundowych fragmentach. Zdegradowane fragmenty zaczynały się po czasie 1 sekundy od początku każdego nagrania. W analogiczny sposób przygotowano bazę nagrań użytych w fazie testowej. W tym przypadku szablon zakłóceń składał się z 3006 par zakłóceń impulsowych, losowo oddalonych od siebie i losowo wybranych z bazy zakłóceń \mathcal{P}_B , gdzie: 766 zakłóceń impulsowych uszkadzało tylko lewy kanał, 740 zakłóceń impulsowych uszkadzało tylko prawy kanał i 1500 zakłóceń impulsowych uszka-

działo oba kanały jednocześnie. Całkowita liczba uszkodzonych próbek w obu kanałach wyniosła 53416, co stanowi 2,78% wszystkich próbek w zdegradowanych 20 sekundowych fragmentach.

Otrzymane szablony zakłóceń impulsowych pokazano na rysunku 8.3. Może się wydawać, że grupa treningowa jest trudniejsza niż grupa testowa ze względu na występujące zakłócenia impulsowe o dużej amplitudzie – ale sytuacja ta nie ma wpływu na uzyskiwane wyniki, bo zakłócenia o dużej amplitudzie są łatwiej i dokładniej wykrywane w sygnale, niż te o małej amplitudzie. Rysunek 8.4 pokazuje rozkład liczebności zakłóceń impulsowych w zależności od ich długości w obu używanych szablonach. Minimalna długość zakłócenia impulsowego wynosiła 3 próbki a maksymalna 41.

8.2.1 Analiza dwukierunkowa – wyniki badań

Aby ocenić skuteczność opracowanej techniki wykorzystującej dwukierunkowe przetwarzania sygnału fonicznego, przeprowadzono pięć eksperymentów symulacyjnych. Pierwszy eksperyment skupiał się na wyborze jednego z trzech podejść do rekonstrukcji sygnału, którymi są: rekonstrukcja lewostronna, rekonstrukcja prawostronna oraz rekonstrukcja dwukierunkowa. Drugi eksperyment pozwolił na wybranie zoptymalizowanego zestawu lokalnych reguł łączenia alarmów detekcyjnych na podstawie miar obiektywnych i subiektywnych. Trzeci eksperyment sprawdzał uniwersalność wybranych reguł na przykładzie dwóch schematów detekcji: podejścia opartego na detektorze pracującym w pętli otwartej oraz podejścia opartego na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym. W czwartym eksperymencie sprawdzono wpływ zabiegu poszerzania alarmów detekcyjnych w analizie dwukierunkowej na jakość uzyskiwanych wyników. W ostatnim eksperymencie porównano proponowane podejście z podejściem wykorzystującym analizę dwukierunkową opartym na łączeniu wyników rekonstrukcji lewostronnej i prawostronnej, opisanym w pracy Canazzy *et al.* [59].

Parametry symulacji

Dwa przebiegi nagrania stereofonicznego reprezentujące odpowiednio lewy i prawy kanał dźwiękowy, zostały rozdzielone i osobno przetworzone. Porównywane adaptacyjne algorytmy detekcji i interpolacji zakładały, że analizowany sygnał foniczny opisany jest za pomocą modelu AR rzędu $r = 10$. Identyfikacja sygnału prowadzona była na bieżąco przy użyciu algorytmu EWLS wyposażonego w stałą zapomnienia $\lambda = 0,998$ (3.39). W sytuacji zgłoszenia alarmu detekcyjnego sprawdzano stabilność aktualnego modelu. W przypadku stwierdzenia niestabilności dokonywano ponownej identyfikacji modelu przy użyciu algorytmu gwarantującego stabilność – algorytmu Levinsona-Durbina. Oceny współczynników autokorelacji estymowane były dla okna prostokątnego o szerokości $N = 1000$ – patrz punkt 3.3.5. Współczynnik szerokości pasma filtru górnoprzepustowego ustawiony był na $c = 0,995$, a stała zapomnienia użyta do aktualizacji wariancji wynosiła $\lambda_0 = 0,993$. Maksymalna długość alarmu detekcyjnego ustawiona była na $m_{\max} = 50$, a parametr związany z rozszerzaniem alarmów detekcyjnych – na $\Delta_1 = 2$. Mnożnik średniego odchylenia standardowego ustawiony został na $\mu = 4,5$.

Eksperyment 1

Ponieważ detekcja zakłóceń impulsowych poprzedza interpolację nieodwracalnie uszkodzonych próbek sygnału, jakość zrekonstruowanych sygnałów fonicznych zależy od osiągniętej na obu etapach skuteczności. Aby sprawdzić w jakim stopniu adaptacyjny algorytm interpolacji



Tablica 8.3: Porównanie skuteczności działania 3 podejść do adaptacyjnej interpolacji sygnału na podstawie otrzymanych wyników PEAQ: podejście wykorzystujące rekonstrukcję lewostronną (GT_f), podejście wykorzystujące rekonstrukcję prawostronną (GT_b) oraz podejście wykorzystujące rekonstrukcję dwukierunkową (GT_{fb}). Rekonstrukcja sygnału przeprowadzona została przy pełnej wiedzy o położeniu zakłóceń impulsowych. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W tabeli zamieszczono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. $\bar{S}r_{.10}$ i $\bar{S}r_{.50}$ oznaczają uśredniony wynik dla 10 wyszczególnionych nagrań oraz dla wszystkich 50 nagrań.

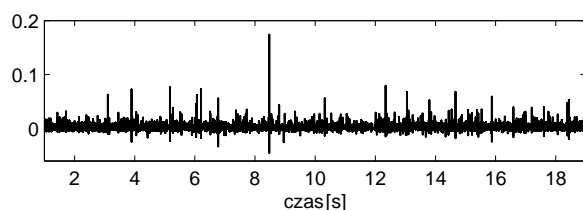
Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

podejście	1	2	3	4	5	6	7	8	9	10	$\bar{S}r_{.10}$	$\bar{S}r_{.50}$
GT_f	-0,38	-0,27	-0,32	-0,13	-0,65	-0,32	-0,76	-0,47	-0,43	-0,21	-0,39	-0,33
GT_b	-0,38	-0,26	-0,32	-0,13	-0,68	-0,34	-0,76	-0,48	-0,44	-0,22	-0,40	-0,34
GT_{fb}	-0,37	-0,26	-0,31	-0,13	-0,63	-0,31	-0,73	-0,46	-0,43	-0,21	-0,38	-0,32

wpływa na końcowy wynik i wybrać najlepsze podejście do rekonstrukcji sygnału, zbadano wyniki interpolacji otrzymane dla przypadku pełnej wiedzy o położeniu zakłóceń impulsowych, tj. $\hat{d}(t) \equiv d(t)$.

W tym eksperymencie porównano skuteczność działania trzech podejść do adaptacyjnej interpolacji sygnału opartych na klasycznym modelu AR, opisanych w podrozdziale 4.2. Pierwsze podejście wykorzystuje rekonstrukcję lewostronną (4.8) i będzie dalej oznaczane symbolem GT_f . Drugie podejście oparte jest na rekonstrukcji prawostronnej (4.9) i będzie dalej oznaczane symbolem GT_b . Trzecie podejście wykorzystuje rekonstrukcję dwukierunkową (4.10), tj. łączy wyniki rekonstrukcji lewostronnej z wynikami rekonstrukcji prawostronnej. Podejście to będzie oznaczane symbolem GT_{fb} . Wszystkie wyniki PEAQ otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Wyniki dla wszystkich nagrań zamieszczono w tabeli C.2 – patrz dodatek C. W tabeli 8.3 zamieszczono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. Porównując jakościowo otrzymane wyniki PEAQ można zauważyć, że najlepszym rozwiązaniem jest podejście GT_{fb} . Prawie we wszystkich przypadkach uzyskało ono najlepsze wyniki PEAQ. Jednak, w żadnym przypadku poprawa jakości nie okazała się znacząca. Przetworzone nagrania oceniono również w testach odsłuchowych. Chociaż różnice w wynikach PEAQ pomiędzy porównywanymi podejściami są nieznaczne, zauważano jednak lokalną poprawę jakości sygnału w miejscach, gdzie długość zrekonstruowanego fragmentu przekraczała 20 próbek. Ponadto, we wszystkich przypadkach skutki zastosowania interpolacji dwukierunkowej opartej na modelu AR były ledwo słyszalne bądź w ogóle niesłyszalne. Najlepszy efekt uzyskano dla grupy nagrań reprezentujących muzykę chóralną oraz muzykę klasyczną. W przypadku grupy nagrań reprezentujących muzykę jazzową, muzykę operową oraz muzykę instrumentalną, słyszalne skutki interpolacji wynikały z użycia prostego, krótkozakresowego modelu AR. Oznacza to, że znacząca większość słyszalnych artefaktów, występujących w przypadku gdy interpolacja jest połączona z adaptacyjną detekcją zakłóceń impulsowych, powodowana jest przez błędy detekcji.

Ponieważ uśrednione wyniki w grupie wszystkich 50 nagrań testowych okazały się najlepsze dla algorytmu GT_{fb} , ten właśnie algorytm stosowany był w kolejnych eksperymentach.



Rysunek 8.5: Zakłócenia impulsowe „wycięte” z archiwalnego nagrania gramofonowego użyte na etapie wyboru lokalnych reguł łączenia alarmów detekcyjnych.

Eksperyment 2

W tym eksperymencie, w charakterze sygnałów testowych do wyboru odpowiednich reguł decyzyjnych, użyto czystych sygnałów (innych niż te z grupy treningowej) spreparowanych przez dodanie zakłóceń impulsowych, „wyciętych” z archiwalnego nagrania gramofonowego. Wykorzystano 10 czystych sygnałów z muzyką klasyczną (Bach, Mozart, Vivaldi, Smetana). Każde nagranie, spróbkowane z częstotliwością 22,05 kHz, zawierało od 23 do 29 sekund materiału fonicznego. Z kolei zakłócenia impulsowe zostały „wycięte”, zgodnie z procedurą opisaną w punkcie 8.1.1, z archiwalnego nagrania gramofonowego¹ – pieśń F. Schuberta (lied) „An die Musik” (opus 88, No. 4). To silnie uszkodzone nagranie pozwoliło na wyizolowanie dużej liczby różnych zakłóceń impulsowych począwszy od małych, krótkich impulsów do impulsów bardzo dużych i długich, których źródłem były głębokie rysy na powierzchni płyty. W ten sposób utworzono 19 sekundowe nagranie zawierające 2674 wyizolowanych zakłóceń impulsowych obejmujących 13428 próbek – patrz rysunek 8.5. Następnie dodano je do przeskalowanych sygnałów. Zdegradowane fragmenty zaczynały się po czasie 1,5 sekundy od początku każdego nagrania.

Spreparowane sygnały foniczne zostały najpierw przetworzone lewostronnie i prawostronnie przez algorytm oparty na detektorze pracującym w pętli otwartej. Następnie, uzyskane sygnały detekcyjne poddano łącznej analizie w celu zlokalizowania wszystkich segmentów zawierających poszczególne konfiguracje alarmów detekcyjnych – patrz punkt 4.3.3. Na koniec, przygotowano nagrania do testu odsłuchowego w specjalny sposób, tak aby uwzględniały tylko jedną, konkretną konfigurację alarmów. Wszystkie segmenty, uwzględniające pozostałe konfiguracje alarmów, zastąpione zostały ich odpowiednikami pochodzącymi z czystych (niezakłóconych) nagrań fonicznych. Dla przykładu, aby porównać 3 potencjale reguły decyzyjne (\cup , \cap , FF), związane z klasą A_2 , przygotowano zostały trzy warianty tego samego nagrania testowego. W każdym z trzech nagrań okna analizy ograniczone zostały tylko do konfiguracji alarmów detekcyjnych związanych z klasą A_2 .

Zestaw lokalnych reguł łączenia alarmów detekcyjnych wybrany został na podstawie wyników testu odsłuchowego z udziałem trzech ekspertów z dziedziny rekonstrukcji dźwięku. Test odsłuchowy polegał na uporządkowaniu nagrań według ich subiektywnie postrzeganej jakości – każdy z trzech ekspertów proszony był o wskazanie najlepszego nagrania w każdej z analizowanych grup nagrań. W przypadku gdy jakość dwóch lub większej liczby nagrań w grupie była porównywalna, więcej niż jedno nagranie mogło być wskazane jako najlepsze. Dla każdego z 10 nagrań należało ocenić 24 warianty jego przetworzenia (zebrane w 10 grupach). Tabela 8.4 pokazuje wyniki testu odsłuchowego. Zauważmy, że z powodu możliwości wielokrotnego wyboru, liczba subiektywnych wyborów zazwyczaj nie sumuje się do 10 (liczba nagrań w analizowanej

¹Nagranie to przetworzone zostało przy użyciu komercyjnego pakietu do rekonstrukcji nagrań fonicznych (CEDAR).

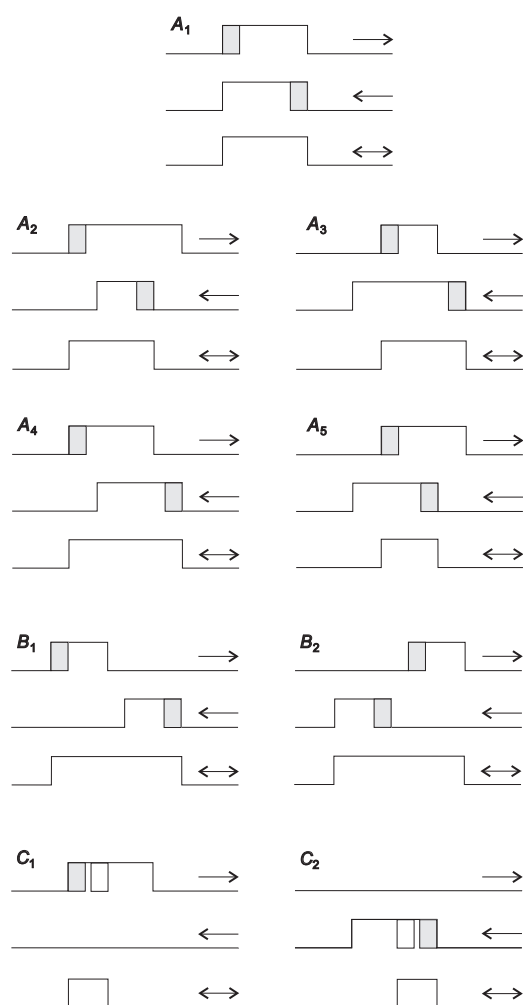
Tablica 8.4: Wybór lokalnych reguł łączenia alarmów detekcyjnych na podstawie obiektywnych i subiektywnych miar. $W[\%]$ oznacza częstość występowania danej konfiguracji alarmów. O oznacza wyniki otrzymane na podstawie obiektywnego pomiaru PEAQ. $E1$, $E2$ i $E3$ oznaczają wyniki pochodzące od trzech ekspertów. Liczby podane w kolumnach O , $E1$, $E2$, $E3$ określają ile razy oceniana reguła otrzymała najlepszą ocenę w analizowanej grupie nagrań (eksperti mogli wybrać więcej niż jedną regułę w każdej kategorii, to samo tyczy się miary PEAQ gdy wyniki były identyczne).

Konfiguracja alarmów	W [%]	Reguła	O	E1	E2	E3	Wybór PEAQ	Wybór Ekspertów
A_1	3,4	$\cup/\cap/FF$	10	10	10	10	×	×
A_2	8,4	\cup	3	8	8	7		
		\cap	0	0	0	3		
A_3	7,5	FF	7	9	8	8	×	×
		\cup	4	8	7	8		
A_4	40,2	\cap	0	0	0	1		
		FF	6	10	9	10	×	×
A_5	3,0	\cup/FF	10	10	9	10	×	×
		\cap	0	0	1	0		
B_1	3,8	\cup	7	7	8	6	×	
		\cap/FF	3	8	7	10		×
B_2	0,1	\sqcup	10	10	10	10	×	×
		\cap	0	0	0	0		
C_1	13,1	\sqcup	10	10	10	9	×	×
		\cap	7	7	7	9		
C_2	12,5	\cup	1	3	1	5		
		\cap	0	0	0	0		
D	8,1	F	9	9	9	9	×	×
		\cup	0	3	2	5		
D	8,1	\cap	0	0	0	0		
		FF	10	10	8	8	×	×
D	8,1	\sqcup	1	6	9	7		
		\cap	0	0	0	0		
D	8,1	FF	9	10	7	9	×	×

Tablica 8.5: Zestaw lokalnych reguł łączenia alarmów detekcyjnych wytypowany przez ekspertów, potwierdzone przez wyniki obiektywnej miary PEAQ. Wybrane reguły: „przednie zbocze-przednie zbocze” (FF), „suma z wypełnieniem” (\sqcup) oraz „przednie zbocze” (F).

Klasa konfiguracji	Reguła łączenia
A	FF
B	\sqcup
C	F
D	FF

grupie). Dla przykładu, wyniki pochodzące od pierwszego eksperta ($E1$) dla reguł \cup , \cap i FF w rozważanej klasie konfiguracji alarmów A_2 , równe są odpowiednio 8, 0 oraz 9. Oznacza to, że co najmniej 7 nagrań z 10, w których porównano regułę \cup i FF , uznanych zostało przez eksperta za porównywalne. Wybór dokonany przez ekspertów przedstawiony został w ostatniej kolumnie



Rysunek 8.6: Lokalne reguły łączenia alarmów detekcyjnych wytypowane przez ekspertów, potwierdzone przez wyniki otrzymane przy użyciu obiektywnej miary PEAQ. Wykresy pokazują wyniki analizy lewostronnej (→), analizy prawostronnej (←) oraz analizy dwukierunkowej (↔) dla elementarnych konfiguracji alarmów oznaczonych symbolami: A₁, A₂, A₃, A₄, A₅, B₁, B₂, C₁ oraz C₂. Zacięte obszary symbolizują poszerzenie alarmów detekcyjnych na etapie wstępnego przetwarzania.

tabeli 8.4, a także – w bardziej syntetycznej formie – w tabeli 8.5 i w graficznej postaci na rysunku 8.6.

W badaniu wykorzystano również obiektywną miarę PEAQ do oceny jakości przetworzonych nagrań². Wyniki pochodzące z bezpośredniego porównania ocen PEAQ, w ramach każdej kategorii, przedstawione zostały w kolumnie O. Zamieszczone liczby pokazują ile razy oceniana reguła uzyskała najlepszą ocenę PEAQ w analizowanej grupie nagrań. W ramach tego badania starano się również odpowiedzieć na pytanie czy wybór lokalnych reguł decyzyjnych oparty na opiniach ekspertów może zostać zastąpiony automatycznym procesem wyboru opartym na

²Ponieważ początkowe badania skupiały się na nagraniach fonicznych próbkowanych z częstotliwością 22,050 kHz, a dopiero na późniejszym etapie prac zaczęto stosować algorytm PEAQ wymagający sygnałów o częstotliwości próbkowania 48 kHz, wszystkie zrekonstruowane już nagrania przepróbkowane zostały z 22,050 kHz na 48 kHz w celu wykonania obiektywnego pomiaru jakości dźwięku.

Tablica 8.6: Wskaźnik częstości ($W[\%]$) występowania różnych konfiguracji alarmów detekcyjnych otrzymanych dla dwóch algorytmów detekcji: algorytmu detekcji pracującego w pętli otwartej (algorytm A) oraz algorytmu detekcji pracującego w pętli z decyzyjnym sprzężeniem zwrotnym (algorytm B). W tabeli umieszczono uśrednione wyniki dla wszystkich 50 spreparowanych sygnałów fonicznych z bazy treningowej.

		$W [\%]$									
Konfiguracja alarmów		A_1	A_2	A_3	A_4	A_5	B_1	B_2	C_1	C_2	D
algorytm A		1,73	20,99	19,84	7	30,86	0,13	0,54	9,78	6,73	2,4
algorytm B		0,23	14,5	14,23	11,97	23,17	1,92	0,37	11,79	8,33	13,49

porównaniu wyników PEAQ. Wybór dokonany przez ekspertów pokrywa się większości przypadków z wyborem dokonany na podstawie wyników PEAQ. Tylko w jednym przypadku, dla konfiguracji alarmów A_5 , dokonane wybory różniły się – wybór ekspertów to reguła FF, zaś wybór na podstawie wyników PEAQ to reguła \cup . Należy jednak zwrócić uwagę na to, że wybór ekspertów nie był jednogłośny, tj. dwóch ekspertów wskazało na regułę FF, a jeden na regułę \cup . Ponadto, eksperci E_1 oraz E_2 uzyskali wyniki porównania \cup vs. FF odpowiednio 7 vs. 8 oraz 8 vs. 7. Oznacza to, że jakość przetworzonych nagrań dla obu reguł była porównywalna. Można zatem stwierdzić, że zastosowanie algorytmu PEAQ pozwala na przeprowadzenie bardziej systematycznego poszukiwania najlepszych reguł łączenia alarmów detekcyjnych oraz pozwala na uniknięcie czasochłonnych testów odsłuchowych.

Eksperyment 3

W tym eksperymencie sprawdzono uniwersalność wybranych reguł decyzyjnych w dwóch grupach podejść, którymi są: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Aby zweryfikować wybrany zestaw reguł użyto innej bazy spreparowanych sygnałów fonicznych niż ta, która posłużyło do ich wyboru. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z bazy treningowej.

Na początek sprawdzono częstość występowania różnych konfiguracji alarmów detekcyjnych w przypadku użycia detektora pracującego w pętli otwartej (algorytm A) oraz detektora pracującego w pętli z decyzyjnym sprzężeniem zwrotnym (algorytm B). W tabeli 8.6 pokazany został procentowy wskaźnik częstości ($W[\%]$) występowania różnych konfiguracji alarmów detekcyjnych, uśredniony dla wszystkich 50 nagrań. Zauważmy, że nawet gdy liczby różnią się od tych wyświetlonych w drugiej kolumnie tabeli 8.4, ogólna tendencja pozostaje bez zmian – najczęściej obserwowane konfiguracje to te sklasyfikowane jako: A_2 , A_3 , A_4 , A_5 , C_1 , C_2 i D (97,6 % oraz 97,48 % wszystkich przypadków, odpowiednio dla algorytmu A oraz dla algorytmu B).

Następnie porównano skuteczność działania 5 podejść do eliminacji zakłóceń impulsowych opartych na modelu AR na podstawie otrzymanych wyników PEAQ. Porównywanymi podejściami były: podejście wykorzystujące analizę lewostronną (L), podejście wykorzystujące analizę prawostronną (P), podejście wykorzystujące analizę dwukierunkową z globalną regułą sumy (S), podejście wykorzystujące analizę dwukierunkową z globalną regułą iloczynu (I) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem reguł lokalnych (RL). Wyniki PEAQ uzyskane dla wszystkich nagrań zamieszczono w tabeli C.3 – patrz dodatek C. W tabeli 8.7 zamieszczono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. W kolumnie oznaczonej symbolem GT_{fb} zamieszczono wyniki PEAQ otrzymane



Tablica 8.7: Porównanie skuteczności działania 5 podejść do eliminacji zakłóceń impulsowych opartych na modelu AR na podstawie otrzymanych wyników PEAQ. Porównywanymi podejściami były: podejście wykorzystujące analizę lewostronną (L), podejście wykorzystujące analizę prawostronną (P), podejście wykorzystujące analizę dwukierunkową z globalną regułą sumy (S), podejście wykorzystujące analizę dwukierunkową z globalną regułą iloczynu (I) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem reguł lokalnych (RL). Porównanie przeprowadzono w dwóch grupach: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów z bazy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W tabeli zamieszczono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. Śr.₁₀ i Śr.₅₀ oznaczają uśredniony wynik dla 10 wyszczególnionych nagrań oraz dla wszystkich 50 nagrań. Symbolem GT_{fb} oznaczono wyniki PEAQ otrzymane przez adaptacyjny algorytm rekonstrukcji sygnału w wariancie dwukierunkowym, przy pełnej wiedzy o położeniu zakłóceń impulsowych. REF oznacza początkową ocenę PEAQ dla spreparowanego nagrania. W obu grupach najlepsze wyniki dla każdego nagrania zostały pokazane wytłuszczoną czcionką.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

nagranie			grupa A					grupa B				
Lp.	GT _{fb}	REF	L	P	S	I	RL	L	P	S	I	RL
1	-0,37	-3,61	-3,63	-3,63	-1,57	-3,45	-0,80	-0,96	-0,83	-0,82	-0,68	-0,51
2	-0,26	-2,80	-1,72	-1,72	-1,73	-0,96	-1,12	-1,05	-0,91	-1,75	-0,63	-0,93
3	-0,31	-3,91	-3,84	-3,81	-2,09	-3,85	-1,07	-2,51	-1,51	-0,85	-1,69	-0,54
4	-0,13	-3,91	-3,83	-3,81	-2,02	-3,84	-0,62	-0,84	-0,80	-0,56	-0,78	-0,27
5	-0,63	-3,80	-3,63	-3,62	-1,81	-3,20	-0,84	-1,18	-1,06	-0,98	-0,61	-0,60
6	-0,31	-3,82	-3,75	-3,71	-1,66	-3,50	-0,74	-0,98	-0,97	-0,78	-0,45	-0,44
7	-0,73	-3,68	-3,61	-3,61	-1,46	-3,02	-0,89	-1,03	-1,07	-0,91	-0,72	-0,71
8	-0,46	-3,38	-2,91	-3,14	-0,72	-2,01	-0,45	-0,54	-0,57	-0,46	-0,39	-0,36
9	-0,43	-3,82	-3,75	-3,66	-1,92	-3,75	-0,85	-1,06	-1,12	-0,85	-0,75	-0,42
10	-0,21	-3,66	-3,24	-3,38	-0,58	-2,86	-0,33	-0,42	-0,40	-0,27	-0,25	-0,21
Śr. ₁₀	-0,38	-3,64	-3,39	-3,41	-1,56	-3,04	-0,77	-1,06	-0,93	-0,82	-0,69	-0,50
Śr. ₅₀	-0,32	-3,56	-3,26	-3,27	-1,57	-2,95	-0,83	-0,97	-0,89	-0,84	-0,68	-0,50

przez adaptacyjny algorytm rekonstrukcji sygnału w wariancie dwukierunkowym, przy pełnej wiedzy o położeniu zakłóceń impulsowych. Natomiast w sąsiedniej kolumnie, oznaczonej symbolem REF, pokazano początkową ocenę PEAQ dla spreparowanego nagrania. Analizując wyniki z tabeli C.3 można zauważyć, że najwięcej wyników PEAQ wyróżnionych jako najlepsze dla ocenianych nagrań zarówno w grupie A jak w grupie B, odpowiada proponowanemu podejściu RL. W grupie A proponowane podejście RL okazało się najlepsze 49 razy, a w grupie B – 40 razy (w tym, w dwóch przypadkach wynik był taki sam jak dla innego podejścia). Otrzymane wyniki PEAQ dla podejścia RL są w wielu przypadkach zbliżone do wyników z kolumny GT_{fb}, a nawet lepsze bądź równe – patrz grupa nagrań z muzyką instrumentalną. W grupie A zaobserwowano 2 przypadki, w których otrzymany wynik okazał się lepszy niż ten z kolumny GT_{fb}, a w grupie B – 8 przypadków, w których wynik okazał się lepszy i 1 przypadek, w którym wynik był taki sam. Kolejnym ważnym spostrzeżeniem jest to, że prawie we wszystkich przypadkach wyniki PEAQ z grupy B są lepsze niż odpowiadające im wyniki PEAQ z grupy A. Poniżej przedstawiono szczegółowe porównanie jakościowej oraz ilościowej przeprowadzone na podstawie wyników PEAQ.

I. Porównanie jakościowe

Jakościowe porównanie wyników z tabeli C.3 zaprezentowano w tabeli 8.8 pokazującej liczbę przypadków, w których dane podejście otrzymało lepszą bądź równą ocenę PEAQ niż jego konkurent. Liczbę „wyraźnych zwycięstw” danego podejścia, tj. liczbę przypadków gdy jego ocena PEAQ była co najmniej o 0,1 lepsza od oceny podejścia konkurencyjnego, podano w nawiasach klamrowych. Na podstawie wyników bezpośredniego porównania pomiędzy poszczególnymi podejściami można wyciągnąć trzy najważniejsze wnioski.

Po pierwsze, porównanie podejść opartych na analizie jednokierunkowej (L vs. P) pokazuje, że podejście P uzyskało w większej liczbie przypadków lepszy wynik PEAQ niż podejście L, zarówno w grupie A (20 vs. 34) jak grupie B (18 vs. 33). Z kolei porównując liczbę „wyraźnych zwycięstw”, w grupie A jest ona na korzyść podejścia L (12 vs. 5), natomiast w grupie B – na korzyść podejścia P (3 vs. 17).

Po drugie, proponowane podejście wykorzystujące analizę dwukierunkową RL uzyskało w większości przypadków lepsze wyniki niż podejścia wykorzystujące analizę jednokierunkową – patrz L vs. RL oraz P vs. RL w tabeli 8.8. W grupie A, proponowane podejście RL uzyskało we wszystkich przypadkach znacząco lepsze oceny PEAQ niż porównywane podejścia konkurencyjne. W grupie B, w przypadku porównania L vs. RL, podejście RL było również bezkonkurencyjne uzyskując we wszystkich przypadkach znacząco lepsze wyniki. W drugim porównaniu P vs. RL, przewaga podejścia RL była również zdecydowana – 47 lepszych ocen PEAQ, w tym 41 „wyraźnych zwycięstw”.

Po trzecie, podejście oparte na analizie dwukierunkowej z proponowanym zestawem lokalnych reguł decyzyjnych uzyskało w znaczącej większości przypadków lepsze wyniki niż podejścia oparte na analizie dwukierunkowej z globalną regułą sumy lub z globalną regułą iloczynu – patrz S vs. RL oraz I vs. RL w tabeli 8.8. W grupie A, podejście RL uzyskało w 49 przypadkach, w obu porównaniach, wyraźnie lepsze wyniki. W grupie B, podejście RL również uzyskało w prawie we wszystkich przypadkach lepsze wyniki, w tym 38 znacząco lepszych wyników dla porównania S vs. RL oraz 22 znacząco lepszych wyników dla porównania I vs. RL. Wyniki wszystkich porównań zostały także zilustrowane na rysunku 8.7 – patrz pierwsza para wykresów.

II. Porównanie ilościowe

Ilościowe porównanie wyników z tabeli C.3 zaprezentowano na rysunkach 8.7, 8.8 oraz 8.9. W tym punkcie sprawdzono w jakim stopniu wystąpiła poprawa ilościowa w wynikach PEAQ po zastosowaniu analizy dwukierunkowej z proponowanym zestawem lokalnych reguł decyzyjnych. Sprawdzono również czy występuje poprawa ilościowa w wynikach PEAQ po zastosowaniu analizy prawostronnej w porównaniu do podejścia opartego na analizie lewostronnej. W tym celu porównano parami testowane podejścia (X vs. Y) w ramach grupy A i grupy B. Różnice w odpowiadających im wynikach PEAQ: $PEAQ_Y - PEAQ_X$, zaznaczono punktami na wykresach – patrz dolna para wykresów na rysunku 8.7. Analizując zamieszczone wykresy można wyciągnąć kilka wniosków.

Po pierwsze, proponowane podejście RL wykazuje znaczącą poprawę, zarówno jakościową jak ilościową, we wszystkich porównaniach: L vs. RL, P vs. RL, S vs. RL oraz I vs. RL – patrz rysunek 8.7.

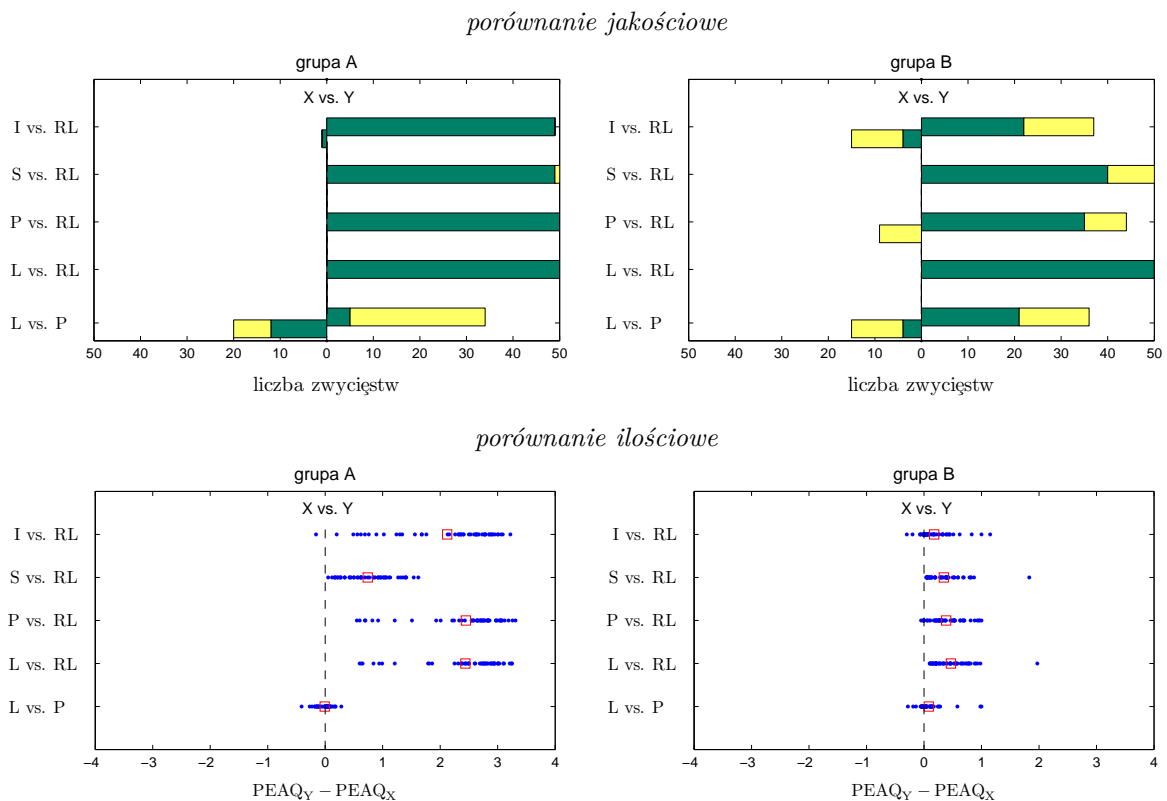
Tablica 8.8: Bezpośrednie porównanie skuteczności działania 5 podejść na podstawie uzyskanych wyników PEAQ: podejście wykorzystujące analizę lewostronną (L), podejście wykorzystujące analizę prawostronną (P), podejście wykorzystujące analizę dwukierunkową z globalną regułą sumy (S), podejście wykorzystujące analizę dwukierunkową z globalną regułą iloczynu (I) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem reguł lokalnych (RL). Porównanie przeprowadzono w dwóch grupach podejść, którymi są: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Porównywane parami podejścia podane zostały w drugiej kolumnie. W kolejnych kolumnach umieszczone zostały wyniki, mówiące ile razy dane podejście otrzymało lepszą ocenę PEAQ w bezpośrednim porównaniu. Wyniki umieszczone w nawiasach klamrowych informują o liczbie zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z bazy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną.

porównywane podejścia		muzyka jazzowa	muzyka chóralna	muzyka operowa
grupa A	L vs. P	5 {2} vs. 7 {3}	6 {2} vs. 5 {0}	0 {0} vs. 10 {0}
	L vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
	P vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
	S vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
	I vs. RL	1 {1} vs. 9 {9}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
grupa B	L vs. P	1 {1} vs. 9 {8}	4 {1} vs. 6 {2}	4 {0} vs. 6 {2}
	L vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
	P vs. RL	4 {0} vs. 7 {3}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
	S vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {6}	0 {0} vs. 10 {9}
	I vs. RL	5 {2} vs. 5 {4}	1 {0} vs. 10 {7}	1 {0} vs. 9 {4}

porównywane podejścia		muzyka instrumentalna	muzyka klasyczna	łączy wynik
grupa A	L vs. P	6 {5} vs. 5 {2}	3 {3} vs. 7 {0}	20 {12} vs. 34 {5}
	L vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 50 {50}
	P vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 50 {50}
	S vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {9}	0 {0} vs. 50 {49}
	I vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	1 {1} vs. 49 {49}
grupa B	L vs. P	4 {0} vs. 6 {3}	5 {1} vs. 6 {2}	18 {3} vs. 33 {17}
	L vs. RL	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 50 {50}
	P vs. RL	0 {0} vs. 10 {9}	1 {0} vs. 10 {9}	5 {0} vs. 47 {41}
	S vs. RL	0 {0} vs. 10 {7}	0 {0} vs. 10 {6}	0 {0} vs. 50 {38}
	I vs. RL	0 {0} vs. 10 {3}	3 {0} vs. 8 {4}	10 {2} vs. 42 {22}

Po drugie, w przypadku porównań L vs. RL oraz P vs. RL, obserwowane różnice w wynikach PEAQ są znacznie większe w grupie A niż w grupie B. Ponieważ detektor pracujący w pętli otwartej (grupa A) ma tendencje do podnoszenia zbyt krótkich alarmów detekcyjnych, co powoduje duże błędy na etapie rekonstrukcji, i podejścia oparte na analizie jednokierunkowej, L i P, uzyskują bardzo niskie oceny PEAQ. Sytuacja ulega znaczącej poprawie, gdy zastosujemy proponowane podejście oparte na analizie dwukierunkowej RL. W obu porównaniach L vs. RL oraz P vs. RL obserwowane różnice w wynikach PEAQ mieszczą się w przedziale od 0,55 do 3,31 ze średnią wartością 2,44 dla pierwszego porównania oraz ze średnią wartością 2,45 dla drugiego porównania. Oznacza to średni wzrost jakości aż o ponad dwa stopnie na czterostopniowej skali PEAQ. W grupie B, w obu porównaniach L vs. RL oraz P vs. RL, obserwowane





Rysunek 8.7: Jakościowe i ilościowe porównanie skuteczności działania testowanych podejść opartych na modelu AR wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście wykorzystujące analizę lewostronną (L), podejście wykorzystujące analizę prawostronną (P), podejście wykorzystujące analizę dwukierunkową z globalną regułą sumy (S), podejście wykorzystujące analizę dwukierunkową z globalną regułą iloczynu (I) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem reguł lokalnych (RL). Porównanie przeprowadzono w dwóch grupach podejść, którymi są: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Wszystkie wyniki uzyskano dla 50 nagrań testowych: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W przypadku porównania jakościowego żółte słupki pokazują ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu X vs. Y. Zielone słupki pokazują liczbę zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść. W przypadku porównania ilościowego dla każdej porównywanej pary X vs. Y zaznaczono punktami wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$. Wyniki dodatnie są na korzyść podejścia Y, a wyniki ujemne na korzyść podejścia X. Wartość średnia z uzyskanych 50 wyników została zaznaczona kwadratem.

różnice w wynikach PEAQ głównie koncentrują się w przedziale od $-0,05$ do $1,00$. Dla tych porównań średnie wartości wyniosły odpowiednio $0,47$ i $0,39$.

Po trzecie, dobór odpowiednich lokalnych reguł decyzyjnych w analizie dwukierunkowej znacząco poprawia wyniki. Proponowany zestaw lokalnych reguł decyzyjnych uzyskał w większości przypadków znacznie lepsze wyniki niż zastosowanie jednej globalnej reguły decyzyjnej w postaci sumy lub w postaci iloczynu – patrz porównania S vs. RL oraz I vs. RL. W grupie A uzyskane wyniki mieszczą się w przedziale od $0,05$ do $1,62$ dla pierwszego porównania oraz w przedziale od $-0,16$ do $3,22$ dla drugiego porównania. Dla tych porównań średnie wartości wyniosły od-

powiednio 0,74 i 2,12. W przypadku grupy B analogiczne wyniki mieszczą się w przedziale od 0,04 do 1,83 oraz w przedziale od -0,3 do 1,15. Wartości średnie dla tych porównań wyniosły 0,34 i 0,18. Należy zwrócić uwagę na to, że w grupie A zastosowanie globalnej reguły sumy daje lepsze rezultaty niż zastosowanie globalnej reguły iloczynu. W przypadku grupy B jest na odwrót – to globalna reguła iloczynu daje lepsze wyniki.

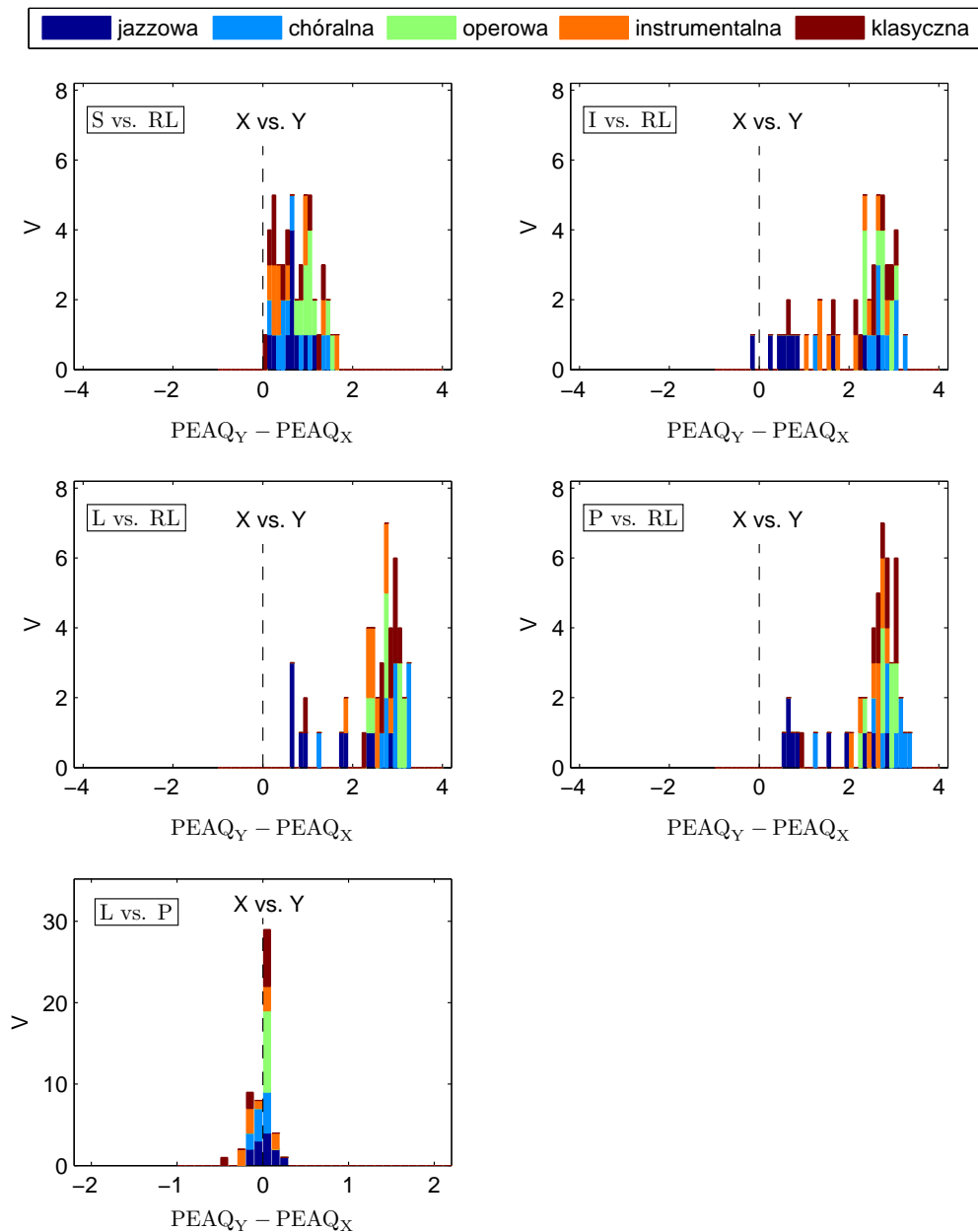
Na koniec, wyniki porównania podejść L vs. P mieszczą się w przedziale od -0,41 do 0,28 dla grupy A oraz od -0,28 do 1,00 dla grupy B. Średnie wartości wyniosły odpowiednio -0,01 oraz 0,08. Dla tej pary podejść, zarówno wyniki ilościowe jak wyniki jakościowe wskazują na to, że podejście oparte na analizie prawostronnej daje lepsze wyniki niż jego konkurent tylko w przypadku gdy porównywane podejścia wykorzystują detektor pracujący w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B).

Rysunki 8.8 oraz 8.9 przedstawiają szczegółowe porównanie ilościowe dla testowanych podejść, odpowiednio dla grupy A i dla grupy B. Dla każdej porównywanej pary podejść sporządzono histogram. Histogram ten pokazuje liczbę (V) uzyskanych różnic w wynikach PEAQ w przedziałach o równej szerokości wynoszącej 0,1. Przedziały dla wartości ujemnych są prawostronnie domknięte, zaś dla wartości dodatnich – lewostronnie domknięte. Poszczególne kolory na histogramie reprezentują wyniki otrzymane dla określonego rodzaju muzyki: muzyka jazzowa, muzyka chóralna, muzyka operowa, muzyka instrumentalna oraz muzyka klasyczna. W grupie A, w przypadku porównania proponowanego podejścia RL z podejściami wykorzystującymi analizę jednokierunkową – L vs. RL oraz P vs. RL (patrz rysunek 8.8) – największe różnice w wynikach PEAQ obserwowane są dla grupy nagrań z muzyką chóralną, muzyką operową, muzyką instrumentalną oraz muzyką klasyczną. Natomiast w przypadku porównania z pozostałymi dwoma podejściami wykorzystującymi analizę dwukierunkową – S vs. RL oraz I vs. RL – największe różnice w wynikach PEAQ obserwowane są dla grupy nagrań z muzyką chóralną oraz muzyką operową. Podobnie dla grupy B, w przypadku porównania proponowanego podejścia RL z podejściami wykorzystującymi analizę jednokierunkową – L vs. RL oraz P vs. RL (patrz rysunek 8.9) – największe różnice w wynikach PEAQ obserwowane są dla grupy nagrań z muzyką chóralną oraz muzyką operową. Natomiast w przypadku porównania S vs. RL, największe różnice w wynikach PEAQ obserwowane są dla grupy nagrań z muzyką jazzową oraz muzyką operową, zaś w przypadku porównania I vs. RL – dla muzyki chóralnej.

W tabeli 8.9 pokazano wyniki, uzyskane dla dwóch miar ilościowych, reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki, wyniki średnie dla 10 wyszczególnionych nagrań oraz wyniki średnie dla wszystkich 50 nagrań z bazy treningowej. Wyniki te pozwalają na dokonanie ciekawych spostrzeżeń dotyczących porównywanych schematów detekcji, przedstawionych w grupie A i w grupie B. Na początek zauważmy, że podejścia z grupy B dają dla obu wskaźników lepsze wyniki niż ich odpowiedniki z grupy A. Dowodzi to, że detektor pracujący w pętli z decyzyjnym sprzężeniem zwrotnym lepiej dopasowuje alarmy detekcyjne niż detektor pracujący w pętli otwartej. Następnie zauważmy, że w obu grupach zastosowanie analizy dwukierunkowej (S, I, RL) znacząco poprawia wskaźnik podobieństwa w porównaniu do analizy jednokierunkowej (L, R). Poprawie ulega również druga statystyka, ale tylko w przypadku podejść S oraz RL. W przypadku podejścia I występuje znaczące pogorszenie drugiej miary w grupie A oraz niewielkie pogorszenie tego samego wskaźnika w grupie B. Wreszcie, podejście RL uzyskało zdecydowanie najlepsze wyniki dla wskaźnika podobieństwa oraz drugie w kolejności wyniki dla wskaźnika pokrycia energetycznego. W przypadku drugiej miary różnice pomiędzy wynikami dla podejścia S oraz podejścia RL są bardzo małe. Dowodzi to, że proponowany schemat wykrywania zakłóceń impulsowych gwarantuje najlepsze dopasowanie alarmów detekcyjnych spośród testowanych schematów detekcji, tj. lepsze oznaczenie początku i końca zakłócenia impulsowego.

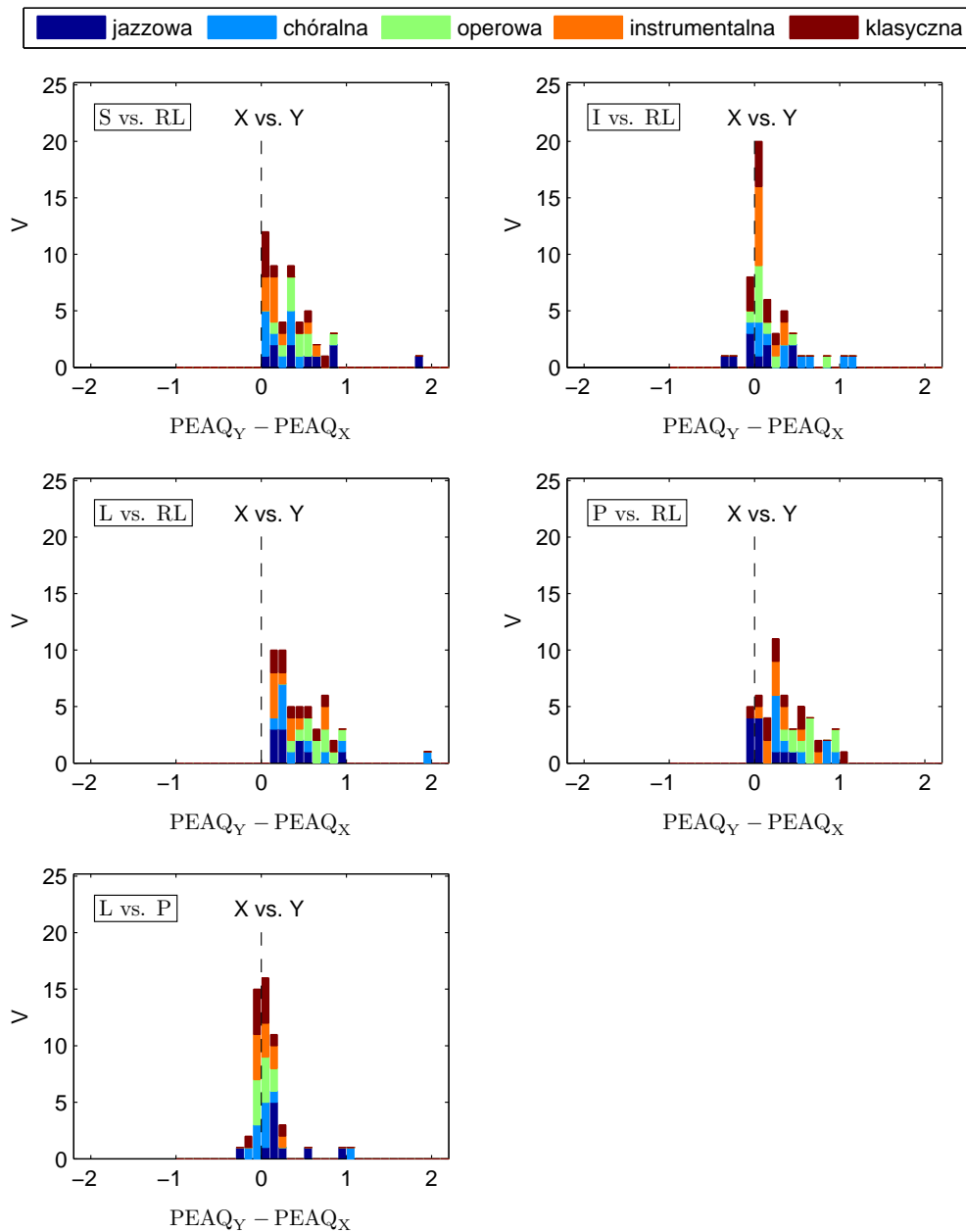


grupa A



Rysunek 8.8: Ilościowe porównanie skuteczności działania testowanych podejść opartych na modelu AR wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście wykorzystujące analizę lewostronną (L), podejście wykorzystujące analizę prawostronną (P), podejście wykorzystujące analizę dwukierunkową z globalną regułą sumy (S), podejście wykorzystujące analizę dwukierunkową z globalną regułą iloczynu (I) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem reguł lokalnych (RL). Wszystkie testowane podejścia wykorzystują detekcję działającą w pętli otwartej (grupa A). Dla każdej porównywanej pary X vs. Y otrzymano wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$ w grupie 50 nagrań testowych: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Wyniki dodatnie są na korzyść podejścia Y , a wyniki ujemne na korzyść podejścia X . Histogram pokazuje liczbę (V) uzyskanych wyników w przedziałach o równej szerokości wynoszącej 0,1. Przedziały dla wartości ujemnych są prawostronnie domknięte, zaś dla wartości dodatnich – lewostronnie domknięte. Poszczególne kolory na histogramie reprezentują wyniki otrzymane dla określonego rodzaju muzyki.

grupa B



Rysunek 8.9: Ilościowe porównanie skuteczności działania testowanych podejść opartych na modelu AR wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście wykorzystujące analizę lewostronną (L), podejście wykorzystujące analizę prawostronną (P), podejście wykorzystujące analizę dwukierunkową z globalną regułą sumy (S), podejście wykorzystujące analizę dwukierunkową z globalną regułą iloczynu (I) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem reguł lokalnych (RL). Wszystkie testowane podejścia wykorzystują detekcję działającą w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Dla każdej porównywanej pary X vs. Y otrzymano wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$ w grupie 50 nagrań testowych: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Wyniki dodatnie są na korzyść podejścia Y, a wyniki ujemne na korzyść podejścia X. Histogram pokazuje liczbę (V) uzyskanych wyników w przedziałach o równej szerokości wynoszącej 0,1. Przedziały dla wartości ujemnych są prawostronnie domknięte, zaś dla wartości dodatnich – lewostronnie domknięte. Poszczególne kolory na histogramie reprezentują wyniki otrzymane dla określonego rodzaju muzyki.

Tablica 8.9: Porównanie skuteczności działania 5 podejść na podstawie dwóch miar pośrednich: wskaźnika podobieństwa oraz wskaźnika pokrycia energetycznego. Porównywane podejścia: podejście wykorzystujące analizę lewostronną (L), podejście wykorzystujące analizę prawostronną (P), podejście wykorzystujące analizę dwukierunkową z globalną regułą sumy (S), podejście wykorzystujące analizę dwukierunkową z globalną regułą iloczynu (I) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem reguł lokalnych (RL). Porównanie przeprowadzono w dwóch grupach podejść, którymi są: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z bazy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W tabeli zamieszczono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. $\bar{S}r_{.10}$ i $\bar{S}r_{.50}$ oznaczają uśredniony wynik dla 10 wyszczególnionych nagrań oraz dla wszystkich 50 nagrań. W obu grupach najlepsze wyniki dla każdego nagrania zostały pokazane wytłuszczoną czcionką.

wskaźnik podobieństwa [%]

Lp.	grupa A					grupa B				
	L	P	S	I	RL	L	P	S	I	RL
1	36,44	37,11	50,40	50,25	72,43	57,36	57,48	70,40	72,36	76,78
2	26,24	27,29	49,72	31,86	58,55	47,46	45,16	63,45	56,49	61,63
3	41,72	43,18	44,75	61,20	80,21	62,05	62,40	62,95	81,18	83,39
4	41,62	43,02	45,42	62,48	81,14	64,25	64,43	66,34	82,06	84,56
5	41,35	42,25	53,90	54,08	74,25	58,50	58,87	67,52	75,01	78,19
6	41,40	42,39	52,61	55,21	76,71	60,98	61,40	67,84	78,13	81,02
7	36,83	37,42	57,02	45,68	69,68	57,16	56,93	72,36	71,49	75,47
8	35,57	35,86	57,67	42,10	66,97	54,33	54,07	72,04	66,49	72,00
9	42,94	44,32	56,24	53,50	78,73	63,01	63,28	68,32	79,65	82,30
10	39,50	40,18	55,87	51,11	73,32	58,07	57,75	71,05	72,81	77,45
$\bar{S}r_{.10}$	38,36	39,30	52,36	50,75	73,20	58,32	58,18	68,23	73,57	77,28
$\bar{S}r_{.50}$	38,76	39,43	51,35	51,86	73,37	57,45	57,31	67,14	72,67	76,99

wskaźnik pokrycia energetycznego [%]

Lp.	grupa A					grupa B				
	L	P	S	I	RL	L	P	S	I	RL
1	98,01	98,83	99,77	96,24	99,73	99,77	99,76	99,88	99,35	99,86
2	97,81	98,10	99,34	97,12	99,32	99,05	98,98	99,49	98,94	99,39
3	98,04	99,40	99,97	97,56	99,95	99,86	99,94	99,98	99,92	99,97
4	98,51	99,32	99,98	96,27	99,95	99,95	99,96	99,98	99,94	99,97
5	97,04	98,40	99,91	95,74	99,86	99,85	99,82	99,92	99,87	99,90
6	98,51	99,04	99,93	97,71	99,90	99,88	99,87	99,94	99,90	99,93
7	97,25	98,25	99,79	96,08	99,74	99,80	99,78	99,90	99,82	99,88
8	97,10	98,03	99,81	95,71	99,75	99,75	99,74	99,89	99,79	99,87
9	92,27	99,08	99,96	91,56	99,76	99,92	99,93	99,98	99,92	99,96
10	96,94	97,95	99,84	94,89	99,81	99,74	99,74	99,88	99,79	99,85
$\bar{S}r_{.10}$	97,15	98,64	99,83	95,89	99,78	99,76	99,75	99,89	99,72	99,86
$\bar{S}r_{.50}$	97,43	98,57	99,81	96,07	99,64	99,49	99,68	99,85	99,29	99,81

Tablica 8.10: Porównanie skuteczności działania 5 podejść na podstawie uzyskanych wyników PEAQ. Testowanymi podejściami wykorzystującymi analizę dwukierunkową z proponowanym zestawem lokalnych reguł decyzyjnych były: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Porównywane podejścia, różniące się między sobą wartością parametru Δ_1 , oznaczono symbolami RL_i , gdzie $i = \Delta_1 \in \{0, 1, 2, 3, 4\}$. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z bazy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W tabeli zamieszczono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. $\bar{S}r_{.10}$ i $\bar{S}r_{.50}$ oznaczają uśredniony wynik dla 10 wyszczególnionych nagrań oraz dla wszystkich 50 nagrań. W obu grupach najlepsze wyniki PEAQ dla każdego nagrania zostały pokazane wytłuszczoną czcionką.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

Lp.	grupa A					grupa B				
	RL_0	RL_1	RL_2	RL_3	RL_4	RL_0	RL_1	RL_2	RL_3	RL_4
1	-1,38	-0,81	-0,80	-0,87	-0,96	-0,91	-0,48	-0,51	-0,62	-0,76
2	-1,07	-1,01	-1,12	-1,18	-1,44	-0,87	-0,86	-0,93	-0,96	-1,12
3	-2,41	-1,37	-1,07	-1,03	-1,06	-1,54	-0,61	-0,54	-0,58	-0,68
4	-1,85	-0,92	-0,62	-0,63	-0,68	-1,06	-0,36	-0,27	-0,31	-0,41
5	-1,72	-1,02	-0,84	-0,82	-0,89	-1,03	-0,56	-0,60	-0,60	-0,75
6	-1,63	-0,81	-0,74	-0,70	-0,78	-1,00	-0,46	-0,44	-0,50	-0,62
7	-1,41	-0,87	-0,89	-0,89	-1,01	-1,04	-0,66	-0,71	-0,78	-0,91
8	-0,77	-0,47	-0,45	-0,45	-0,51	-0,57	-0,34	-0,36	-0,39	-0,46
9	-2,01	-1,04	-0,85	-0,85	-1,04	-1,14	-0,45	-0,42	-0,52	-0,73
10	-0,88	-0,62	-0,33	-0,31	-0,33	-0,48	-0,26	-0,21	-0,22	-0,26
$\bar{S}r_{.10}$	-1,51	-0,89	-0,77	-0,77	-0,87	-0,96	-0,50	-0,50	-0,55	-0,67
$\bar{S}r_{.50}$	-1,51	-0,96	-0,83	-0,84	-0,90	-0,95	-0,52	-0,50	-0,56	-0,67

Eksperyment 4

W analizie dwukierunkowej, zanim lewostronne alarmy detekcyjne zostaną połączone z alarmami prawostronnymi zgodnie z proponowanym zestawem lokalnych reguł decyzyjnych, alarmy detekcyjne zostają wstępnie poszerzone o stałą liczbę próbek Δ_1 – patrz punkt 4.3.1. Zabieg ten jest bardzo istotny oraz znacząco wpływa na jakość przetworzonego nagrania. Aby to pokazać, porównano skuteczność działania 5 podejść. Testowanymi podejściami wykorzystującymi analizę dwukierunkową z proponowanym zestawem lokalnych reguł decyzyjnych były: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejście oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Porównywane podejścia, różniące się wartością parametru Δ_1 , oznaczono symbolami RL_i , gdzie $i = \Delta_1 \in \{0, 1, 2, 3, 4\}$. Sprawdzone eksperymentalnie i polecane w pracy wartością dla tego parametru jest $\Delta_1 = 2$.

Otrzymane wyniki PEAQ dla treningowej bazy nagrań zamieszczono w tabeli C.4 – patrz dodatek C. W tabeli 8.10 zamieszczono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. Analizując wyniki można zauważyć, że w grupie A podejścia RL_2 i RL_3 uzyskały najwięcej najlepszych wyników, równo po 24. W grupie B, podejście RL_2 zwyciężyło uzyskując w 34 przypadkach najlepszy wynik. Należy zwrócić uwagę na dużą różnicę w wynikach PEAQ pomiędzy podejściami RL_0 vs. RL_1 .

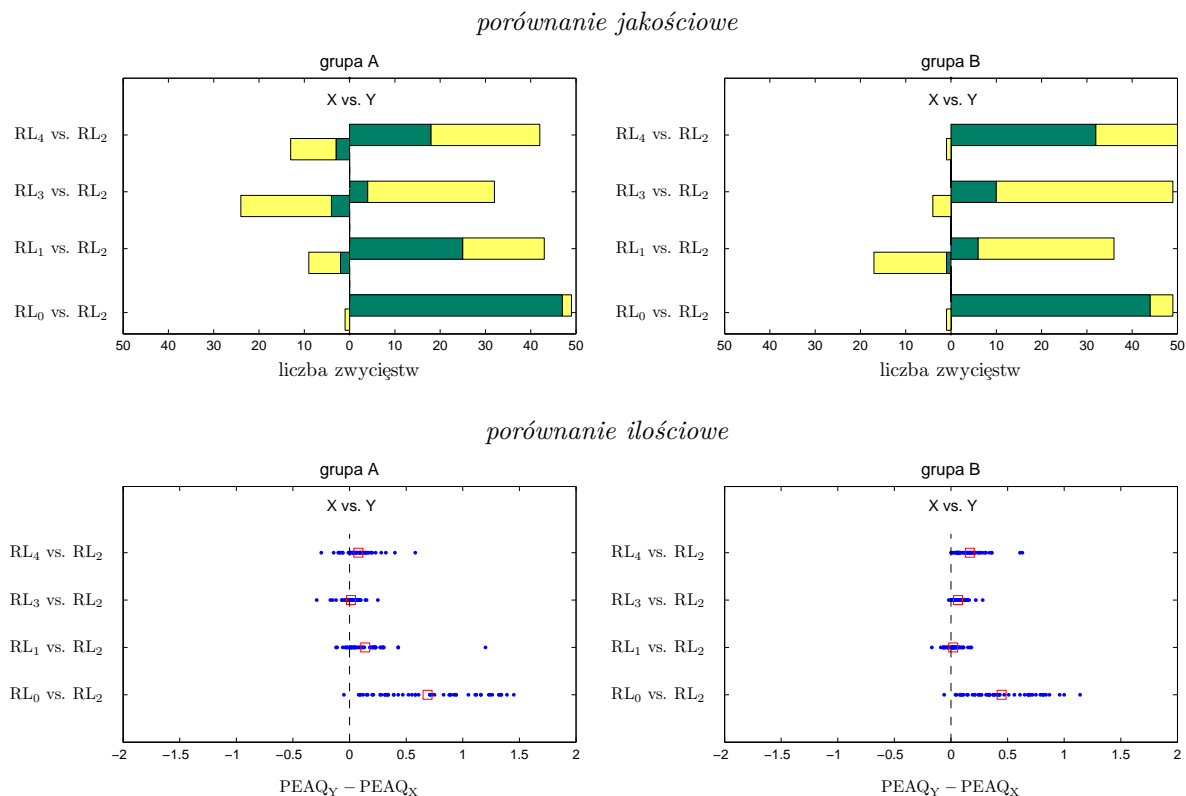
Tablica 8.11: Bezpośrednie porównanie skuteczności działania 5 podejść na podstawie uzyskanych wyników PEAQ. Testowanymi podejściami wykorzystującymi analizę dwukierunkową z proponowanym zestawem lokalnych reguł decyzyjnych były: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Porównywane podejścia, różniące się między sobą wartością parametru Δ_1 , oznaczono symbolami RL_i , gdzie $i = \Delta_1 \in \{0, 1, 2, 3, 4\}$. Porównywane parami podejścia podane zostały w drugiej kolumnie. W kolejnych kolumnach umieszczone zostały wyniki, mówiące ile razy dane podejście otrzymało lepszą ocenę PEAQ w bezpośrednim porównaniu. Wyniki umieszczone w nawiasach klamrowych informują o liczbie zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z bazy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną.

porównywane podejścia		muzyka jazzowa	muzyka chóralna	muzyka operowa
grupa A	RL_0 vs. RL_2	1 {0} vs. 9 {8}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
	RL_1 vs. RL_2	6 {1} vs. 6 {0}	1 {1} vs. 9 {6}	0 {0} vs. 10 {8}
	RL_3 vs. RL_2	1 {0} vs. 9 {2}	6 {1} vs. 4 {1}	5 {1} vs. 6 {0}
	RL_4 vs. RL_2	1 {0} vs. 10 {6}	3 {1} vs. 8 {3}	2 {1} vs. 8 {3}
grupa B	RL_0 vs. RL_2	1 {0} vs. 9 {6}	0 {0} vs. 10 {9}	0 {0} vs. 10 {10}
	RL_1 vs. RL_2	8 {0} vs. 4 {0}	1 {1} vs. 9 {2}	2 {0} vs. 8 {1}
	RL_3 vs. RL_2	1 {0} vs. 10 {3}	1 {0} vs. 10 {2}	1 {0} vs. 10 {2}
	RL_4 vs. RL_2	0 {0} vs. 10 {6}	0 {0} vs. 10 {5}	0 {0} vs. 10 {10}

porównywane podejścia		muzyka instrumentalna	muzyka klasyczna	łączy wynik
grupa A	RL_0 vs. RL_2	0 {0} vs. 10 {10}	0 {0} vs. 10 {9}	1 {0} vs. 49 {47}
	RL_1 vs. RL_2	2 {0} vs. 8 {4}	0 {0} vs. 10 {7}	9 {2} vs. 43 {25}
	RL_3 vs. RL_2	4 {0} vs. 9 {0}	8 {2} vs. 4 {1}	24 {4} vs. 32 {4}
	RL_4 vs. RL_2	1 {0} vs. 9 {3}	6 {1} vs. 7 {3}	13 {3} vs. 42 {18}
grupa B	RL_0 vs. RL_2	0 {0} vs. 10 {10}	0 {0} vs. 10 {9}	1 {0} vs. 49 {44}
	RL_1 vs. RL_2	5 {0} vs. 6 {1}	1 {0} vs. 9 {2}	17 {1} vs. 36 {6}
	RL_3 vs. RL_2	1 {0} vs. 9 {1}	0 {0} vs. 10 {2}	4 {0} vs. 49 {10}
	RL_4 vs. RL_2	0 {0} vs. 10 {6}	1 {0} vs. 10 {5}	1 {0} vs. 50 {32}

Jakościowe porównanie wyników z tabeli C.4 przedstawiono w tabeli 8.11. Na podstawie wyników bezpośredniego porównania pomiędzy poszczególnymi podejściami można wyciągnąć trzy wnioski. Po pierwsze, podejście RL_2 uzyskało w obu grupach w większości przypadków najlepsze wyniki – patrz porównania RL_0 vs. RL_2 , RL_1 vs. RL_2 , RL_3 vs. RL_2 oraz RL_4 vs. RL_2 . Po drugie, największa przewaga podejścia RL_2 jest widoczna dla porównania RL_0 vs. RL_2 – w prawie wszystkich przypadkach podejście to uzyskało znacząco lepsze wyniki. Ostatnim, ważnym wnioskiem jest to, że dalsze zwiększanie wartości parametru Δ_1 , np. $\Delta_1 = 4$, powoduje pogorszenie uzyskiwanych wyników.

Jakościowe i ilościowe porównanie wyników z tabeli C.4 zaprezentowano także na rysunku 8.10. Pierwsza para wykresów przedstawia porównanie jakościowe, a druga para wykresów – porównanie ilościowe. W przypadku porównania ilościowego RL_0 vs. RL_2 różnice w wynikach PEAQ mieściły się w przedziałach od $-0,05$ do $1,45$ oraz od $-0,06$ do $1,14$. Wartości średnie dla tych porównań wyniosły odpowiednio $0,69$ oraz $0,45$. Wyniki te potwierdzają, że zabieg poszerzania alarmów detekcyjnych znacząco wpływa na poprawę uzyskiwanych wyników. W przypadku kolejnych porównań – patrz porównania RL_1 vs. RL_2 , RL_3 vs. RL_2 oraz RL_4 vs. RL_2 – różnice w wynikach były już mniejsze, ale nadal w wielu przypadkach znaczące.



Rysunek 8.10: Jakościowe i ilościowe porównanie skuteczności działania 5 podejść na podstawie wyników PEAQ. Testowanymi podejściami wykorzystującymi analizę dwukierunkową z proponowanym zestawem lokalnych reguł decyzyjnych były: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Porównywane podejścia, różniące się między sobą wartością parametru Δ_1 , oznaczono symbolami RL_i , gdzie $i = \Delta_1 \in \{0, 1, 2, 3, 4\}$. Wszystkie wyniki otrzymano dla grupy 50 nagrań testowych: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W przypadku porównania jakościowego żółte słupki pokazują ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu X vs. Y. Zielone słupki pokazują liczbę zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść. W przypadku porównania ilościowego dla każdej porównywanej pary X vs. Y zaznaczono punktami wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$. Wyniki dodatnie są na korzyść podejścia Y, a wyniki ujemne na korzyść podejścia X. Wartość średnia z uzyskanych 50 wyników została zaznaczona kwadratem.

W tabeli 8.12 pokazano wyniki, uzyskane dla dwóch miar ilościowych, reprezentujące po 2 nagrania z każdego rodzaju muzyki, wyniki średnie dla 10 wyszczególnionych nagrań oraz wyniki średnie dla wszystkich 50 spreparowanych sygnałów fonicznych z bazy treningowej. Wyniki te pozwalają na dokonanie dwóch spostrzeżeń dotyczących zabiegu poszerzania alarmów detekcyjnych. Po pierwsze, w obu grupach poszerzanie alarmów detekcyjnych o stałą liczbę próbek z przedziału od $\Delta_1 = 1$ do $\Delta_1 = 4$ poprawia obie statystyki w porównaniu do sytuacji, gdy nie poszerzono alarmów detekcyjnych $\Delta_1 = 0$. Po drugie, zbyt małe poszerzenie ($\Delta_1 = 1$) lub zbyt duże poszerzenie ($\Delta_1 = 4$) alarmów detekcyjnych nie prowadzi do uzyskania najlepszych wyników w pierwszej statystyce. W grupie A najlepsze wyniki w pierwszej statystyce uzyskują podejścia RL_2 oraz RL_3 , zaś w grupie B – podejście RL_2 . Natomiast druga statystyka, zgodnie z oczekiwaniami, rośnie w miarę większego poszerzania alarmów detekcyjnych. Wyniki potwierdzają, że dokonany wybór $\Delta_1 = 2$ jest wyborem optymalnym dla obu grup.

Tablica 8.12: Porównanie skuteczności działania 5 podejść na podstawie miar pośrednich: wskaźnika podobieństwa oraz wskaźnika pokrycia energetycznego. Testowanymi podejściami wykorzystującymi analizę dwukierunkową z proponowanym zestawem lokalnych reguł decyzyjnych były: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Porównywane podejścia, różniące się między sobą wartością parametru Δ_1 , oznaczono symbolami RL_i , gdzie $i = \Delta_1 \in \{0, 1, 2, 3, 4\}$. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z bazy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W tabeli wyszczególniono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. $\bar{S}r_{.10}$ i $\bar{S}r_{.50}$ oznaczają uśredniony wynik dla 10 wyszczególnionych nagrań oraz dla wszystkich 50 nagrań.

wskaźnik podobieństwa [%]

Lp.	grupa A					grupa B				
	RL_0	RL_1	RL_2	RL_3	RL_4	RL_0	RL_1	RL_2	RL_3	RL_4
1	46,19	60,28	72,43	73,93	69,55	53,93	65,92	76,78	76,62	70,71
2	32,16	45,23	58,55	64,97	65,02	37,46	48,77	61,63	66,96	66,18
3	58,95	73,89	80,21	74,39	66,85	67,04	80,43	83,39	75,63	67,05
4	59,02	74,70	81,14	75,37	67,56	66,65	81,44	84,56	76,73	67,80
5	49,04	63,38	74,25	74,39	69,32	57,50	69,92	78,19	76,27	69,71
6	51,92	66,81	76,71	75,22	69,25	60,49	74,06	81,02	77,33	69,89
7	42,82	55,76	69,68	74,51	71,36	53,12	62,88	75,47	78,04	72,75
8	40,91	52,95	66,97	72,59	70,88	50,40	59,20	72,00	75,68	72,35
9	55,24	69,89	78,73	75,83	69,06	64,04	76,80	82,30	77,06	69,11
10	46,09	60,67	73,32	75,06	70,48	54,28	66,89	77,45	77,19	71,19
$\bar{S}r_{.10}$	48,23	62,36	73,20	73,63	68,93	56,49	68,63	77,28	75,75	69,67
$\bar{S}r_{.50}$	48,50	62,70	73,37	73,35	68,67	56,01	68,43	76,99	69,37	69,37

wskaźnik pokrycia energetycznego [%]

Lp.	grupa A					grupa B				
	RL_0	RL_1	RL_2	RL_3	RL_4	RL_0	RL_1	RL_2	RL_3	RL_4
1	99,55	99,68	99,73	99,76	99,78	99,72	99,81	99,86	99,88	99,88
2	98,81	99,15	99,32	99,42	99,47	98,97	99,24	99,39	99,48	99,54
3	99,88	99,93	99,95	99,96	99,96	99,85	99,95	99,97	99,97	99,98
4	99,89	99,93	99,95	99,96	99,97	99,94	99,97	99,97	99,98	99,98
5	99,71	99,81	99,86	99,88	99,90	99,80	99,87	99,90	99,92	99,93
6	99,79	99,87	99,90	99,92	99,93	99,86	99,91	99,93	99,94	99,94
7	99,52	99,67	99,74	99,77	99,79	99,76	99,84	99,88	99,90	99,91
8	99,49	99,68	99,75	99,78	99,81	99,72	99,82	99,87	99,89	99,90
9	99,66	99,73	99,76	99,78	99,78	99,91	99,95	99,96	99,96	99,97
10	99,59	99,75	99,81	99,84	99,85	99,71	99,81	99,85	99,87	99,88
$\bar{S}r_{.10}$	99,59	99,72	99,78	99,81	99,82	99,72	99,82	99,86	99,88	99,89
$\bar{S}r_{.50}$	99,41	99,56	99,64	99,67	99,69	99,65	99,76	99,81	99,86	99,86

Tablica 8.13: Porównanie skuteczności działania 2 podejść na podstawie wyników PEAQ: podejście oparte na łączeniu rekonstrukcji lewostronnej i prawostronnej (M) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem lokalnych reguł (RL). Porównanie przeprowadzono w dwóch grupach: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z bazy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W tabeli zamieszczono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. $\bar{S}r_{.10}$ i $\bar{S}r_{.50}$ oznaczają uśredniony wynik dla 10 wyszczególnionych nagrań oraz dla wszystkich 50 nagrań. W obu grupach najlepsze wyniki dla każdego nagrania zostały pokazane wytłuszczoną czcionką.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

Nagranie		grupa A		grupa B	
Lp.	REF	M	RL	M	RL
1	-3,61	-3,45	-0,80	-0,60	-0,51
2	-2,80	-1,38	-1,12	-0,90	-0,93
3	-3,91	-3,82	-1,07	-0,77	-0,54
4	-3,91	-3,83	-0,62	-0,42	-0,27
5	-3,80	-3,57	-0,84	-0,80	-0,60
6	-3,82	-3,71	-0,74	-0,58	-0,44
7	-3,68	-3,55	-0,89	-0,76	-0,71
8	-3,38	-2,31	-0,45	-0,39	-0,36
9	-3,82	-3,75	-0,85	-0,58	-0,42
10	-3,66	-2,95	-0,33	-0,25	-0,21
$\bar{S}r_{.10}$	-3,64	-3,23	-0,77	-0,61	-0,50
$\bar{S}r_{.50}$	-3,56	-3,09	-0,83	-0,61	-0,50

Eksperyment 5

W ostatnim eksperymencie porównano proponowane podejście RL z podejściem wykorzystującym analizę dwukierunkową opartym na łączeniu wyników rekonstrukcji lewostronnej i prawostronnej (M), opisanym w pracy Canazzy *et al.* [59]. Należy podkreślić, że podejście to nie łączy w żaden sposób alarmów detekcyjnych otrzymanych przez analizę „w przód” i analizę „w tył”. Wyniki PEAQ otrzymane dla wszystkich nagrań pokazano w tabeli C.5, zaś w tabeli 8.13 – wyniki dla 10 wybranych nagrań, po dwa nagrania z każdego rodzaju muzyki. Porównanie przeprowadzono w dwóch grupach podejść, którymi są: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Analizując wyniki można zauważyć, że podejście RL zdecydowanie wygrało porównanie w obu grupach. W grupie A podejście RL uzyskało bardzo dużą przewagę dla każdego nagrania, zaś w grupie B różnice są mniejsze, ale systematyczne i słyszalne. W tej grupie podejście RL uzyskało w 49 przypadkach lepszy bądź równy wynik PEAQ, w tym 24 znacząco lepsze wyniki oraz 2 równe wyniki.

Podsumowanie

Wyniki przeprowadzonych eksperymentów pokazały, że proponowany tu zestaw reguł łączenia alarmów detekcyjnych, wybrany dla konkretnej realizacji zakłóceń impulsowych, jest uniwersalny, tj. algorytm oparty na wskazanych regułach działa zadowalająco w przypadku obróbki nagrań fonicznych spreparowanych inną realizacją zakłóceń impulsowych. Należy podkreślić, że najważniejszym elementem opisanej procedury jest sposób wyboru zestawu reguł i ich weryfikacji. Stosując tę procedurę do innego zbioru danych, można w łatwy sposób uzyskać nowe reguły, lepiej dostosowane do konkretnego problemu np. konkretnego zbioru nagrań fonicznych i/lub zakłóceń impulsowych.

8.2.2 Model dwuzakresowy – wyniki badań

Przetestowane w poprzednim punkcie podejścia do detekcji/interpolacji oparte na modelu AR dały bardzo dobre wyniki w przypadku dużej części sygnałów fonicznych z treningowej bazy nagrań. Niestety zawiodły w przypadku niektórych sygnałów fonicznych z udziałem śpiewu i/lub instrumentów dętych takich jak trąbka, saksofon czy klarnet. Choć zastosowanie analizy dwukierunkowej poprawiło znacząco wyniki, w niektórych przypadkach uzyskana jakość przetworzonego nagrania nadal pozostawała niezadowolająca.

W tym punkcie sprawdzono skuteczność dwuzakresowego modelowania sygnałów fonicznych zarówno w analizie jednokierunkowej jak w analizie dwukierunkowej. W pierwszym eksperymencie sprawdzono skuteczność działania adaptacyjnego algorytmu interpolacji opartego na modelu SAR. W drugim eksperymencie porównano skuteczność działania 4 podejść w przypadku gdy interpolacja połączona została z adaptacyjną detekcją zakłóceń impulsowych. Aby zwiększyć odporność schematu detekcji zakłóceń impulsowych i jednocześnie zmniejszyć liczbę fałszywych alarmów, końcowa decyzja o położeniu zakłócenia impulsowego wypracowywana jest w proponowanym podejściu wspólnie przez trzy działające równoległe detektory zakłóceń impulsowych. Dwa z nich oparte są na modelach SAR o tych samych parametrach modelu krótkozakresowego lecz różnych parametrach modelu długozakresowego. Trzeci detektor oparty jest na modelu krótkozakresowym. Ponieważ narzędzie PEAQ daje wiarygodne wyniki tylko dla sygnałów muzycznych, sygnały mowy oceniono w testach odsłuchowych.

Parametry symulacji

W tym podejściu sygnał foniczny reprezentowany jest przez model SAR identyfikowany jako kaskadowe połączenie modelu krótkozakresowego i modelu długozakresowego. Na model krótkozakresowy składało się 10 współczynników, zaś na model długozakresowy – tylko 1 współczynnik. Identyfikacja modelu SAR przebiegała w dwóch krokach. W pierwszym kroku identyfikowane były parametry modelu krótkozakresowego przy użyciu algorytmu EWLS wyposażonego w stałą zapominania $\lambda = 0,998$ (3.39). Model ten wykorzystany był następnie do obliczenia sygnału resztowego $x(t)$. W kolejnym kroku estymowane były parametry modelu długozakresowego: $T(t)$ oraz $c(t)$. W tym kroku użyto następujących wartości parametrów: $M_0 = 160$, $M_* = 40$, $m_0 = 20$, $m_* = 5$, $T_{\min} = 110$, $T_{\max} = 1000$, $k_0 = 3$ oraz $L = 40$. W sytuacji zgłoszenia alarmu detekcyjnego sprawdzano stabilność aktualnego modelu SAR. W przypadku stwierdzenia niestabilności dokonywano ponownej identyfikacji modelu krótkozakresowego przy użyciu algorytmu Levinsona-Durbina. Oceny współczynników autokorelacji estymowane były dla okna prostokątnego o szerokości $N = 1000$ – patrz punkt 3.3.5. W części długozakresowej przyjęty zapas stabilności wynosił $\epsilon = 0,01$. Współczynnik szerokości pasma filtru górnoprzepustowego ustawiony był na $c = 0,995$, a stała zapominania użyta do aktualizacji wariancji wynosiła $\lambda_0 = 0,993$. Maksymalna długość alarmu detekcyjnego ustawiona była na $m_{\max} = 50$, a parametr związany z rozszerzaniem alarmów detekcyjnych – na $\Delta_1 = 2$. W symulacjach użyto trzech typowych wartości mnożnika średniego odchylenia standardowego $\mu \in \{3, 5, 4, 4, 5\}$.

Eksperyment 1

W poprzednim punkcie stwierdzono na podstawie wyników pierwszego eksperymentu, że w wielu przypadkach słyszalne skutki interpolacji nieodwracalnie uszkodzonych próbek sygnału, wynikają z użycia prostego, krótkozakresowego modelu AR. W tej części rozszerzono poprzedni eksperyment na przypadek modelowania dwuzakresowego. Ponownie, porównano 3 podejścia



Tablica 8.14: Porównanie skuteczności działania 3 podejść do adaptacyjnej interpolacji sygnału na podstawie otrzymanych wyników PEAQ: podejście wykorzystujące rekonstrukcję lewostronną (GT_f), podejście wykorzystujące rekonstrukcję prawostronną (GT_b) oraz podejście wykorzystujące rekonstrukcję dwukierunkową (GT_{fb}). Porównanie przeprowadzono dla 4 wariantów: interpolacja oparta na modelu AR (A), interpolacja oparta na modelu SAR wykorzystującym przyczynową ocenę okresu podstawowego (B), interpolacja oparta na modelu SAR wykorzystującym nieprzyczynową ocenę okresu podstawowego (C) oraz interpolacja oparta na lokalnie wybieranym modelu A, B lub C (D). Rekonstrukcja sygnału przeprowadzona została przy pełnej wiedzy o położeniu zakłóceń impulsowych. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W tabeli zamieszczono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. $\bar{S}r_{.10}$ i $\bar{S}r_{.50}$ oznaczają uśredniony wynik dla 10 wyszczególnionych nagrań oraz dla wszystkich 50 nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

wariant A

podejście	1	2	3	4	5	6	7	8	9	10	$\bar{S}r_{.10}$	$\bar{S}r_{.50}$
GT_f	-0,38	-0,27	-0,32	-0,13	-0,65	-0,32	-0,76	-0,47	-0,43	-0,21	-0,39	-0,33
GT_b	-0,38	-0,26	-0,32	-0,13	-0,68	-0,34	-0,76	-0,48	-0,44	-0,22	-0,40	-0,34
GT_{fb}	-0,37	-0,26	-0,31	-0,13	-0,63	-0,31	-0,73	-0,46	-0,43	-0,21	-0,38	-0,32

wariant B

podejście	1	2	3	4	5	6	7	8	9	10	$\bar{S}r_{.10}$	$\bar{S}r_{.50}$
GT_f	-0,34	-0,24	-0,33	-0,15	-0,59	-0,25	-0,62	-0,41	-0,38	-0,21	-0,35	-0,29
GT_b	-0,33	-0,22	-0,32	-0,16	-0,56	-0,26	-0,61	-0,43	-0,38	-0,22	-0,35	-0,29
GT_{fb}	-0,29	-0,20	-0,28	-0,14	-0,51	-0,21	-0,56	-0,38	-0,34	-0,19	-0,31	-0,25

wariant C

podejście	1	2	3	4	5	6	7	8	9	10	$\bar{S}r_{.10}$	$\bar{S}r_{.50}$
GT_f	-0,36	-0,25	-0,32	-0,13	-0,62	-0,32	-0,72	-0,44	-0,41	-0,21	-0,38	-0,32
GT_b	-0,37	-0,23	-0,32	-0,13	-0,61	-0,33	-0,73	-0,46	-0,43	-0,21	-0,38	-0,32
GT_{fb}	-0,35	-0,23	-0,30	-0,13	-0,57	-0,30	-0,70	-0,43	-0,40	-0,20	-0,36	-0,30

wariant D

podejście	1	2	3	4	5	6	7	8	9	10	$\bar{S}r_{.10}$	$\bar{S}r_{.50}$
GT_f	-0,34	-0,24	-0,32	-0,14	-0,60	-0,26	-0,63	-0,42	-0,38	-0,21	-0,36	-0,29
GT_b	-0,33	-0,21	-0,32	-0,15	-0,57	-0,28	-0,65	-0,44	-0,42	-0,21	-0,36	-0,29
GT_{fb}	-0,31	-0,20	-0,28	-0,14	-0,53	-0,23	-0,60	-0,40	-0,36	-0,19	-0,32	-0,26

do adaptacyjnej interpolacji sygnału na podstawie otrzymanych wyników PEAQ: podejście wykorzystujące rekonstrukcję lewostronną (GT_f), podejście wykorzystujące rekonstrukcję prawostronną (GT_b) oraz podejście wykorzystujące rekonstrukcję dwukierunkową (GT_{fb}). Porównanie przeprowadzono dla czterech wariantów, w których interpolacja była oparta: na modelu AR (wariant A), na modelu SAR wykorzystującym przyczynowe oszacowanie okresu podstawowego (wariant B), na modelu SAR wykorzystującym nieprzyczynowe oszacowanie okresu podstawowego (wariant C) oraz na lokalnie wybieranym modelu (wariant A, wariant B lub wariant C), który minimalizuje sumę kwadratów błędów jednokrokowej predykcji obserwowanych w niezbyt dalekiej przeszłości. Wyniki otrzymane dla wszystkich nagrań zamieszczono odpowiednio w ta-

belach: tabeli C.2 (wariant A), tabeli C.6 (wariant B), tabeli C.7 (wariant C) oraz tabeli C.8 (wariant D) – patrz dodatek C. Wyniki reprezentujące kolejno po dwa nagrania z każdego rodzaju muzyki zamieszczono w tabeli 8.14.

Analizując uzyskane wyniki można wyciągnąć następujące wnioski. Po pierwsze, w każdym wariancie adaptacyjnej interpolacji podejście GT_{fb} uzyskało w prawie wszystkich przypadkach najlepsze wyniki PEAQ. W żadnym przypadku nie zaobserwowano jednak znaczącej różnicy w wynikach PEAQ – patrz, w ramach każdego wariantu, GT_f vs. GT_{fb} oraz GT_b vs. GT_{fb} . Po drugie, podejścia wykorzystujące dwuzakresowe modelowanie sygnału fonicznego uzyskały w prawie wszystkich przypadkach lepsze wyniki niż podejście wykorzystujące modelowanie krótkozakresowe – patrz wariant B vs. wariant A, wariant C vs. wariant A oraz wariant D vs. wariant A. W 17 przypadkach uzyskano wyniki znacząco lepsze na korzyść modelowania dwuzakresowego. Jakość przetworzonych nagrań oceniono w testach odsłuchowych. W wielu przypadkach słyszalne wcześniej skutki interpolacji (wariant A), tym razem już nie występowały (wariant B i wariant D). Po trzecie, najlepsze średnie wyniki, zarówno w grupie 10 wyszczególnionych nagrań jak w grupie wszystkich 50 nagrań, uzyskano w wariancie B. Na drugim miejscu znalazł się proponowany wariant D z niewielką różnicą w wynikach. Na trzecim miejscu znalazł się wariant C, a na ostatnim – wariant adaptacyjnej interpolacji opartej na klasycznym modelu AR. Uzyskana kolejność jest zgodna z oczekiwaniami.

Eksperyment 2

W tym eksperymencie połączono interpolację z adaptacyjną detekcją zakłóceń impulsowych i sprawdzono odporność różnych schematów detekcji na podnoszenie fałszywych alarmów dla trzech typowych ustawień parametru μ , gdzie $\mu \in \{3, 5, 4, 4, 5\}$. Porównano skuteczność działania 4 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie wyników PEAQ: podejście oparte na modelu krótkozakresowym (A/A^*), podejście oparte na modelu SAR wykorzystującym przyczynową ocenę okresu podstawowego (B/B^*), podejście oparte na modelu SAR wykorzystującym nieprzyczynową ocenę okresu podstawowego (C/C^*) oraz proponowane podejście oparte na modelu, który najlepiej w danym momencie opisuje sygnał foniczny (D/D^*) – model wybierany z trzech powyższych na podstawie kryterium minimalizacji sumy kwadratów błędów jednokrokowej predykcji obserwowanych w niezbyt dalekiej przeszłości. We wszystkich 4 podejściach użyto detektora pracującego w pętli z decyzyjnym sprzężeniem zwrotnym. Wszystkie podejścia wykorzystują ten sam model krótkozakresowy. Otrzymane wyniki dla wszystkich nagrań zamieszczono w tabeli C.9 – patrz dodatek C, zaś wyniki reprezentujące kolejno po dwa nagrania z każdego rodzaju muzyki w tabeli 8.15.

Na podstawie zebranych wyników można poczynić następujące obserwacje. Po pierwsze podejścia oparte na modelach SAR wykorzystujące analizę dwukierunkową uzyskały we wszystkich przypadkach lepsze wyniki (prawie wszystkie wyniki były znacząco lepsze) niż ich odpowiedniki wykorzystujące analizę jednokierunkową – patrz B vs. B^* , C vs. C^* oraz D vs. D^* . Po drugie, podejścia oparte na modelu dwuzakresowym uzyskały w większości przypadków lepsze wyniki PEAQ niż podejście oparte na ich modelu krótkozakresowym, zarówno w analizie jednokierunkowej jak i w analizie dwukierunkowej. Po trzecie, ponieważ obniżanie wartości μ wpływa bezpośrednio na obniżanie progu detekcyjnego (większa wrażliwość detektora), obserwujemy w przypadku niektórych nagrań polepszenia wyników, a w przypadku innych – pogorszenia. Na koniec, proponowane podejście D/D^* uzyskało łącznie najlepszy wynik w grupie wszystkich testowanych wartości μ , tj. najczęściej uzyskiwało najlepszy średni wynik PEAQ w grupie wszystkich 50 nagrań – patrz tabela 8.15.



Tablica 8.15: Porównanie skuteczności działania 4 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR (A/A^*), podejście oparte na modelu SAR z przyczynowym oszacowaniem okresu podstawowego (B/B^*), podejście oparte na modelu SAR z nieprzyczynowym oszacowaniem okresu podstawowego (C/C^*) oraz proponowane podejście oparte na współpracy powyższych trzech podejść (D/D^*). W tabeli zamieszczono wyniki porównania dla trzech typowych wartości współczynnika średniego odchylenia standardowego $\mu \in \{3, 5, 4, 4, 5\}$. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. W tabeli wyszczególniono jedynie wyniki reprezentujące kolejno po 2 nagrania z każdego rodzaju muzyki. $\bar{S}r_{.10}$ i $\bar{S}r_{.50}$ oznaczają uśredniony wynik dla 10 wyszczególnionych nagrań oraz dla wszystkich 50 nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

analiza jednokierunkowa

Lp.	$\mu = 3, 5$				$\mu = 4$				$\mu = 4, 5$			
	A	B	C	D	A	B	C	D	A	B	C	D
1	-1,94	-1,54	-1,83	-1,73	-1,18	-1,48	-1,07	-1,00	-0,96	-0,93	-0,87	-0,90
2	-2,37	-1,92	-1,73	-1,02	-1,60	-1,27	-1,01	-0,74	-1,05	-0,89	-0,73	-0,62
3	-1,35	-1,34	-2,44	-1,45	-1,26	-1,30	-1,24	-1,66	-2,51	-2,53	-2,45	-2,55
4	-0,81	-0,82	-0,80	-0,80	-0,78	-0,81	-0,77	-0,84	-0,84	-0,89	-0,80	-0,91
5	-1,36	-1,21	-1,35	-1,32	-1,27	-1,13	-1,32	-1,18	-1,18	-1,03	-1,15	-1,18
6	-1,75	-1,68	-1,83	-1,00	-1,00	-1,64	-1,06	-1,68	-0,98	-0,97	-0,99	-1,04
7	-1,11	-0,95	-1,08	-1,02	-0,97	-0,93	-1,01	-0,99	-1,03	-0,94	-1,01	-0,98
8	-0,60	-0,55	-0,61	-0,56	-0,56	-0,53	-0,56	-0,54	-0,54	-0,53	-0,55	-0,54
9	-1,11	-1,03	-1,14	-1,08	-1,02	-0,97	-1,01	-1,01	-1,06	-1,04	-1,03	-1,08
10	-0,40	-0,39	-0,40	-0,39	-0,37	-0,39	-0,37	-0,39	-0,42	-0,41	-0,43	-0,43
$\bar{S}r_{.10}$	-1,28	-1,14	-1,32	-1,04	-1,00	-1,05	-0,94	-1,00	-1,06	-1,02	-1,00	-1,02
$\bar{S}r_{.50}$	-1,20	-1,11	-1,22	-1,05	-0,99	-0,96	-0,98	-0,93	-0,97	-0,93	-0,92	-0,93

analiza dwukierunkowa

Lp.	$\mu = 3, 5$				$\mu = 4$				$\mu = 4, 5$			
	A*	B*	C*	D*	A*	B*	C*	D*	A*	B*	C*	D*
1	-1,29	-1,04	-1,34	-1,12	-0,73	-0,94	-0,75	-0,65	-0,51	-0,56	-0,58	-0,53
2	-2,39	-1,38	-1,36	-0,76	-1,43	-0,76	-0,75	-0,53	-0,93	-0,56	-0,53	-0,43
3	-0,71	-0,67	-1,07	-0,62	-0,53	-0,51	-0,69	-0,51	-0,54	-0,79	-1,05	-0,80
4	-0,42	-0,41	-0,42	-0,36	-0,31	-0,30	-0,32	-0,30	-0,27	-0,26	-0,27	-0,26
5	-0,85	-0,66	-0,81	-0,61	-0,68	-0,53	-0,67	-0,55	-0,60	-0,46	-0,62	-0,52
6	-0,71	-0,55	-0,70	-0,49	-0,48	-0,37	-0,57	-0,38	-0,44	-0,34	-0,41	-0,37
7	-0,91	-0,69	-0,88	-0,66	-0,76	-0,58	-0,74	-0,61	-0,71	-0,56	-0,70	-0,58
8	-0,49	-0,44	-0,49	-0,40	-0,39	-0,34	-0,39	-0,34	-0,36	-0,31	-0,36	-0,32
9	-0,65	-0,53	-0,64	-0,46	-0,48	-0,39	-0,48	-0,39	-0,42	-0,35	-0,42	-0,37
10	-0,35	-0,34	-0,32	-0,28	-0,24	-0,23	-0,22	-0,21	-0,21	-0,21	-0,21	-0,21
$\bar{S}r_{.10}$	-0,88	-0,67	-0,80	-0,57	-0,60	-0,50	-0,56	-0,45	-0,50	-0,44	-0,52	-0,44
$\bar{S}r_{.50}$	-0,86	-0,70	-0,82	-0,59	-0,60	-0,50	-0,57	-0,47	-0,50	-0,43	-0,49	-0,43

W tabeli 8.16 pokazano uśrednione wyniki dla 50 nagrań, zaokrąglone do najbliższej wartości całkowitej, ilustrujące wpływ parametru μ na trzy statystyki detekcji: wskaźnik poprawnie wykrytych zakłóceń impulsowych ($\overline{N_{dc}}$), wskaźnik fałszywie wykrytych ($\overline{N_{di}}$) oraz wskaźnik niewykrytych zakłóceń impulsowych ($\overline{N_n}$). Całkowita liczba dodanych zakłóceń wynosiła 4255. We wszystkich przypadkach wartość wskaźnika $\overline{N_{dc}}$ jest największa (im wyższa tym lepiej) dla podejścia C/C*. Dodatkowo, zastosowanie analizy dwukierunkowej zwiększa wartość tego wskaźnika w porównaniu do analizy jednokierunkowej – patrz C vs. C*. Z kolei, podejście D/D* cechuje się największą odpornością na podnoszenie fałszywych alarmów – liczba fałszywych alarmów jest średnio o co najmniej 36% mniejsza niż w przypadku pozostałych podejść. Wartości wskaźnika $\overline{N_{di}}$ dla tego podejścia są najmniejsze (im mniejsze tym lepiej). Należy zwrócić również uwagę na szybki wzrost liczby fałszywych alarmów dla coraz niższych wartości μ . Zwiększona odporność na podnoszenie fałszywych alarmów w przypadku podejścia D/D* ma też swoją cenę – większa liczba niewykrytych zakłóceń impulsowych, niż w przypadku pozostałych podejść ($\overline{N_n}$). Liczba niewykrytych zakłóceń impulsowych może być w tym przypadku zmniejszana w kontrolowany sposób poprzez niewielką zmianę wartości μ . Wskaźniki te dostarczają ogólnej wiedzy na temat skuteczności działania detektora, lecz nie pozwalają na ocenę jakości zrekonstruowanego nagrania. W przypadku krótkich fałszywych alarmów detekcyjnych algorytm interpolacji jest w stanie zrekonstruować wskazany fragment bez znaczącej utraty jakości, wyzwaniem dla algorytmu interpolacji są tylko długie alarmy detekcyjne.

Zmiana wartości parametru projektowego z $\mu = 4,5$ na niższą, $\mu = 4$ lub $\mu = 3,5$, może prowadzić zarówno do pogorszenia jak do polepszenia uzyskiwanych wyników PEAQ. W tabeli 8.17 umieszczone zostały wyniki, mówiące ile razy podejście X dla wybranej wartości μ otrzymało najlepszą ocenę PEAQ w bezpośrednim porównaniu: X($\mu = 3,5$) vs. X($\mu = 4$) vs. X($\mu = 4,5$). W przypadku analizy jednokierunkowej najwięcej najlepszych wyników uzyskały testowane podejścia dla $\mu = 4$ oraz $\mu = 4,5$. W przypadku analizy dwukierunkowej wszystkie testowane podejścia uzyskały zdecydowanie najwięcej najlepszych wyników dla $\mu = 4,5$.

Szczegółowe jakościowe porównanie wyników z tabeli C.9 przedstawiono w tabeli 8.18. Porównywane parami podejścia podane zostały w drugiej kolumnie. W kolejnych kolumnach umieszczone zostały wyniki, mówiące o tym ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu. Wyniki umieszczone w nawiasach klamrowych informują o liczbie zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść. Porównania przeprowadzono dla podejść wykorzystujących analizę jednokierunkową/dwukierunkową z różnymi ustawieniami parametru μ . Na podstawie analizy uzyskanych wyników wyciągnięto następujące wnioski. Po pierwsze, wyniki porównań stają się korzystniejsze dla proponowanego podejścia D/D* dla coraz mniejszych wartości μ – patrz porównania A/A* vs. D/D*, B/B* vs. D/D* oraz C/C* vs. D/D*. Po drugie, proponowane podejście uzyskuje lepsze wyniki w porównaniach gdy wykorzystuje ono analizę dwukierunkową – patrz A* vs. D*, B* vs. D* oraz C* vs. D*. Należy zwrócić uwagę, że tylko w jednym przypadku łączny wynik porównania jest na korzyść innego podejścia – patrz B* vs. D* dla $\mu = 4,5$ – jednak różnica w wynikach jest niewielka, 33 vs. 28. W pozostałych przypadkach podejście D* uzyskało najlepsze łączne wyniki. Dodatkowo, podejście D* uzyskało najwięcej znacząco lepszych wyników we wszystkich porównaniach. Po trzecie, proponowane podejście uzyskało w większości przypadków lepsze wyniki niż podejście oparte na modelu krótkozakresowym – patrz porównania A/A* vs. D/D* w tabeli 8.18. Na koniec, podejście B/B* uzyskało zdecydowanie lepsze wyniki PEAQ niż podejście C/C* dla prawie wszystkich sprawdzanych ustawień μ . Jedynie dla porównania B vs. C z $\mu = 4,5$, nieco lepszy łączny wynik uzyskało podejście C.

Tablica 8.16: Porównanie skuteczności działania 4 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie trzech wskaźników detekcji: podejście oparte na modelu AR (A/A^*), podejście oparte na modelu SAR z przyczynowym oszacowaniem okresu podstawowego (B/B^*), podejście oparte na modelu SAR z nieprzyczynowym oszacowaniem okresu podstawowego (C/C^*) oraz proponowane podejście oparte na współpracy powyższych trzech podejść (D/D^*). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. \overline{N}_{dc} , \overline{N}_{di} oraz \overline{N}_n oznaczają kolejno: średnią liczbę prawdziwie dodatnich alarmów detekcyjnych, średnią liczbę fałszywie dodatnich alarmów detekcyjnych oraz średnią liczbę fałszywie ujemnych alarmów detekcyjnych. W tabeli zamieszczono uśrednione wyniki w grupie wszystkich 50 nagrań, zaokrąglone do najbliższej liczby całkowitej. W tabeli uwzględniono porównania dla trzech typowych wartości parametru $\mu \in \{3, 5, 4, 4, 5\}$. Całkowita liczba zakłóceń impulsowych w każdym spreparowanym nagraniu wynosiła 4255.

analiza jednokierunkowa

	\overline{N}_{dc}				\overline{N}_{di}				\overline{N}_n			
	A	B	C	D	A	B	C	D	A	B	C	D
$\mu = 3, 5$	3843	3838	3848	3793	3177	3156	3052	1772	412	417	407	462
$\mu = 4$	3732	3727	3744	3681	1200	1183	1116	670	523	528	511	574
$\mu = 4, 5$	3622	3616	3638	3571	555	552	496	318	634	639	617	684

analiza dwukierunkowa

	\overline{N}_{dc}				\overline{N}_{di}				\overline{N}_n			
	A*	B*	C*	D*	A*	B*	C*	D*	A*	B*	C*	D*
$\mu = 3, 5$	3940	3939	3945	3898	4937	5007	4842	2759	315	316	310	357
$\mu = 4$	3839	3839	3851	3797	1734	1725	1631	942	416	416	404	459
$\mu = 4, 5$	3741	3739	3757	3698	743	727	653	389	514	516	498	557

Tablica 8.17: Bezpośrednie porównanie skuteczności działania testowanego podejścia dla trzech typowych wartości parametru $\mu \in \{3, 5, 4, 4, 5\}$ na podstawie wyników PEAQ. Zamieszczone w tabeli wyniki pokazują ile razy testowane podejście uzyskało najlepszy wynik PEAQ dla wybranej wartości μ . Testowane podejścia wykorzystujące analizę jednokierunkową/dwukierunkową: podejście oparte na modelu AR (A/A^*), podejście oparte na modelu SAR z przyczynowym oszacowaniem okresu podstawowego (B/B^*), podejście oparte na modelu SAR z nieprzyczynowym oszacowaniem okresu podstawowego (C/C^*) oraz proponowane podejście oparte na współpracy powyższych trzech podejść (D/D^*). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań.

	A	B	C	D	A*	B*	C*	D*
$\mu = 3, 5$	2	7	0	15	0	0	0	1
$\mu = 4$	23	26	22	25	3	5	3	8
$\mu = 4, 5$	26	23	30	20	47	48	47	45

Tablica 8.18: Bezpośrednie porównanie skuteczności działania 4 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR (A/A*), podejście oparte na modelu SAR z przyczynowym oszacowaniem okresu podstawowego (B/B*), podejście oparte na modelu SAR z nieprzyczynowym oszacowaniem okresu podstawowego (C/C*) oraz proponowane podejście oparte na współpracy powyższych trzech podejść (D/D*). W tabeli zamieszczono wyniki porównania dla trzech typowych wartości współczynnika średniego odchylenia standardowego $\mu \in \{3, 5, 4, 4, 5\}$. Porównywane parami podejścia podane zostały w drugiej kolumnie. W kolejnych kolumnach umieszczone zostały wyniki, mówiące o tym ile razy dane podejście otrzymało lepszą ocenę PEAQ w bezpośrednim porównaniu. Wyniki umieszczone w nawiasach klamrowych informują o liczbie zdecydowanych ocen na korzyść jednego z porównywanych podejść. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną.

analiza jednokierunkowa

porównywane podejścia		muzyka jazzowa	muzyka chóralna	muzyka operowa
$\mu = 3,5$	B vs. C	5 {2} vs. 5 {4}	5 {3} vs. 5 {0}	10 {9} vs. 0 {0}
	A vs. D	0 {0} vs. 10 {8}	7 {2} vs. 4 {1}	0 {0} vs. 10 {5}
	B vs. D	1 {1} vs. 9 {8}	4 {2} vs. 7 {1}	8 {3} vs. 2 {1}
	C vs. D	0 {0} vs. 10 {7}	6 {0} vs. 5 {2}	0 {0} vs. 10 {7}
$\mu = 4$	B vs. C	3 {1} vs. 7 {5}	0 {0} vs. 10 {2}	9 {4} vs. 1 {1}
	A vs. D	2 {0} vs. 8 {7}	9 {1} vs. 1 {1}	2 {1} vs. 9 {1}
	B vs. D	0 {0} vs. 10 {7}	8 {1} vs. 5 {2}	10 {2} vs. 0 {0}
	C vs. D	2 {0} vs. 8 {5}	9 {2} vs. 1 {0}	3 {1} vs. 8 {2}
$\mu = 4,5$	B vs. C	1 {1} vs. 9 {5}	0 {0} vs. 10 {2}	8 {3} vs. 2 {0}
	A vs. D	2 {0} vs. 9 {4}	9 {1} vs. 1 {0}	4 {0} vs. 7 {0}
	B vs. D	1 {0} vs. 10 {7}	9 {0} vs. 3 {1}	10 {4} vs. 0 {0}
	C vs. D	5 {0} vs. 6 {3}	9 {3} vs. 1 {0}	9 {0} vs. 2 {0}

porównywane podejścia		muzyka instrumentalna	muzyka klasyczna	łączy wynik
$\mu = 3,5$	B vs. C	10 {4} vs. 0 {0}	9 {3} vs. 3 {0}	39 {21} vs. 13 {4}
	A vs. D	0 {0} vs. 10 {4}	3 {0} vs. 8 {2}	10 {2} vs. 42 {20}
	B vs. D	6 {1} vs. 5 {2}	8 {2} vs. 3 {0}	27 {9} vs. 26 {12}
	C vs. D	0 {0} vs. 10 {4}	2 {0} vs. 8 {1}	8 {0} vs. 43 {21}
$\mu = 4$	B vs. C	10 {2} vs. 2 {0}	7 {3} vs. 3 {0}	29 {10} vs. 23 {8}
	A vs. D	5 {0} vs. 8 {2}	5 {0} vs. 7 {2}	23 {2} vs. 33 {13}
	B vs. D	10 {2} vs. 2 {0}	7 {0} vs. 4 {0}	35 {5} vs. 21 {9}
	C vs. D	4 {0} vs. 9 {1}	6 {0} vs. 6 {3}	24 {3} vs. 32 {11}
$\mu = 4,5$	B vs. C	8 {2} vs. 2 {0}	4 {1} vs. 6 {0}	21 {7} vs. 29 {7}
	A vs. D	5 {0} vs. 7 {1}	7 {0} vs. 5 {1}	27 {1} vs. 29 {6}
	B vs. D	10 {1} vs. 2 {0}	9 {0} vs. 3 {0}	39 {5} vs. 18 {8}
	C vs. D	3 {0} vs. 7 {0}	9 {0} vs. 3 {1}	35 {3} vs. 19 {4}

analiza dwukierunkowa

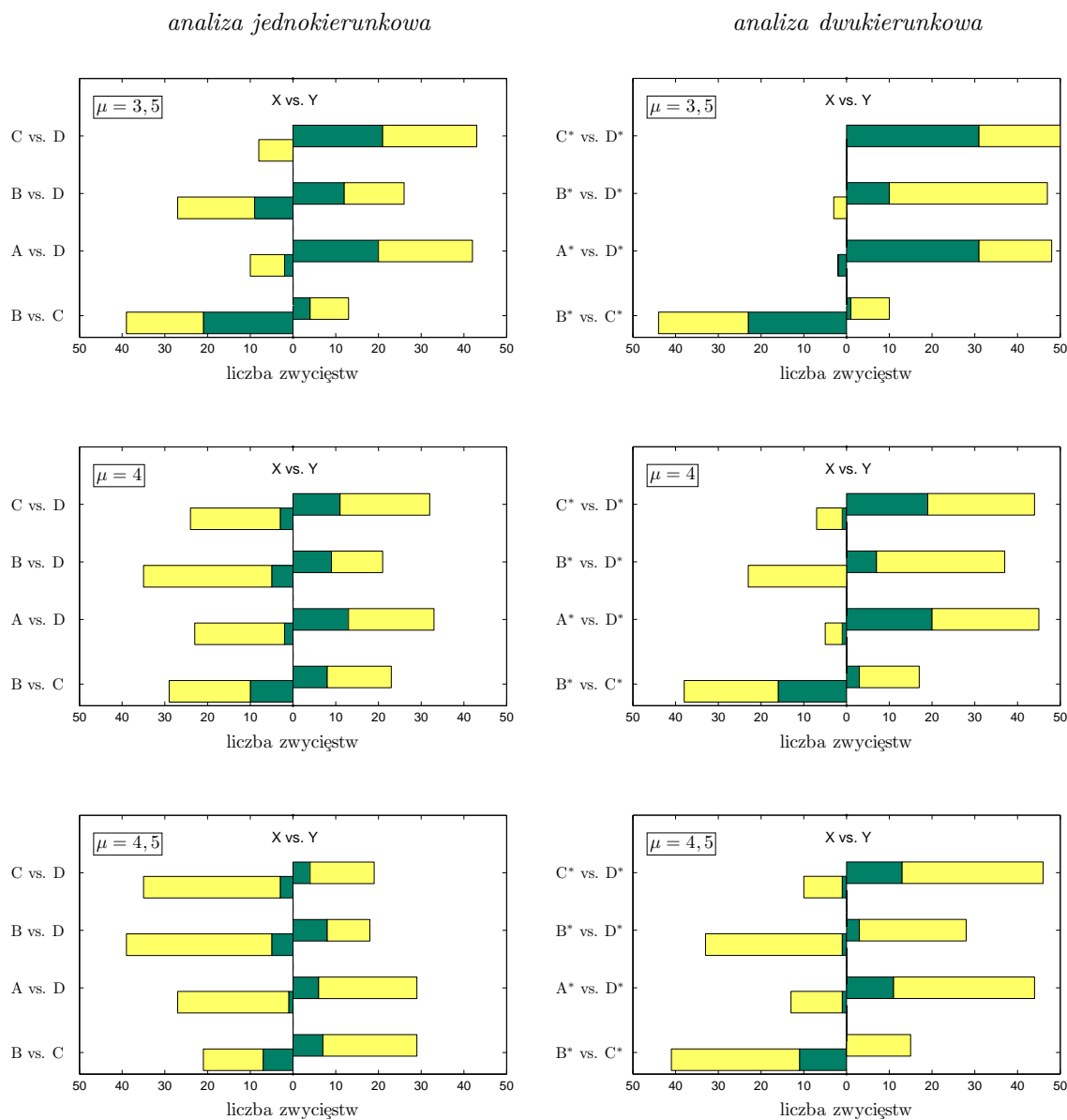
porównywane podejścia		muzyka jazzowa	muzyka chóralna	muzyka operowa
$\mu = 3,5$	B* vs. C*	7 { 2 } vs. 4 {1}	9 { 2 } vs. 4 {0}	10 { 10 } vs. 0 {0}
	A* vs. D*	0 {0} vs. 10 { 8 }	0 {0} vs. 10 { 2 }	0 {0} vs. 10 { 10 }
	B* vs. D*	1 {0} vs. 9 { 6 }	1 {0} vs. 9 { 1 }	1 {0} vs. 9 {0}
	C* vs. D*	0 {0} vs. 10 { 7 }	0 {0} vs. 10 { 3 }	0 {0} vs. 10 { 10 }
$\mu = 4$	B* vs. C*	5 {1} vs. 6 { 2 }	6 {1} vs. 6 {0}	10 { 9 } vs. 0 {0}
	A* vs. D*	2 {1} vs. 8 { 6 }	0 {0} vs. 10 {1}	0 {0} vs. 10 { 8 }
	B* vs. D*	2 {0} vs. 8 { 4 }	3 {0} vs. 9 {1}	8 {0} vs. 3 {0}
	C* vs. D*	3 {1} vs. 7 { 5 }	2 {0} vs. 9 { 2 }	0 {0} vs. 10 { 6 }
$\mu = 4,5$	B* vs. C*	6 {1} vs. 5 {0}	6 {1} vs. 7 {0}	10 { 4 } vs. 0 {0}
	A* vs. D*	4 {0} vs. 8 { 5 }	5 {1} vs. 8 {1}	0 {0} vs. 10 { 2 }
	B* vs. D*	2 {1} vs. 9 { 2 }	6 {0} vs. 8 {1}	10 {0} vs. 0 {0}
	C* vs. D*	2 {1} vs. 9 { 3 }	5 {0} vs. 7 { 2 }	0 {0} vs. 10 { 4 }

porównywane podejścia		muzyka instrumentalna	muzyka klasyczna	łączy wynik
$\mu = 3,5$	B* vs. C*	10 { 6 } vs. 0 {0}	8 { 3 } vs. 2 {0}	44 { 23 } vs. 10 {1}
	A* vs. D*	0 {0} vs. 10 { 7 }	2 {2} vs. 8 { 4 }	2 {2} vs. 48 { 31 }
	B* vs. D*	0 {0} vs. 10 { 2 }	0 {0} vs. 10 {1}	3 {0} vs. 47 { 10 }
	C* vs. D*	0 {0} vs. 10 { 6 }	0 {0} vs. 10 { 5 }	0 {0} vs. 50 { 31 }
$\mu = 4$	B* vs. C*	8 { 2 } vs. 2 {1}	9 { 3 } vs. 3 {0}	38 { 16 } vs. 17 {3}
	A* vs. D*	1 {0} vs. 9 { 3 }	2 {0} vs. 8 { 2 }	5 {1} vs. 45 { 20 }
	B* vs. D*	7 {0} vs. 8 { 1 }	3 {0} vs. 9 {1}	23 {0} vs. 37 { 7 }
	C* vs. D*	1 {0} vs. 9 { 3 }	1 {0} vs. 9 { 3 }	7 {1} vs. 44 { 19 }
$\mu = 4,5$	B* vs. C*	9 { 3 } vs. 1 {0}	10 { 2 } vs. 2 {0}	41 { 11 } vs. 15 {0}
	A* vs. D*	1 {0} vs. 10 { 2 }	3 {0} vs. 8 {1}	13 {1} vs. 44 { 11 }
	B* vs. D*	8 {0} vs. 4 {0}	7 {0} vs. 7 {0}	33 {1} vs. 28 { 3 }
	C* vs. D*	1 {0} vs. 10 { 3 }	2 {0} vs. 10 {1}	10 {1} vs. 46 { 13 }

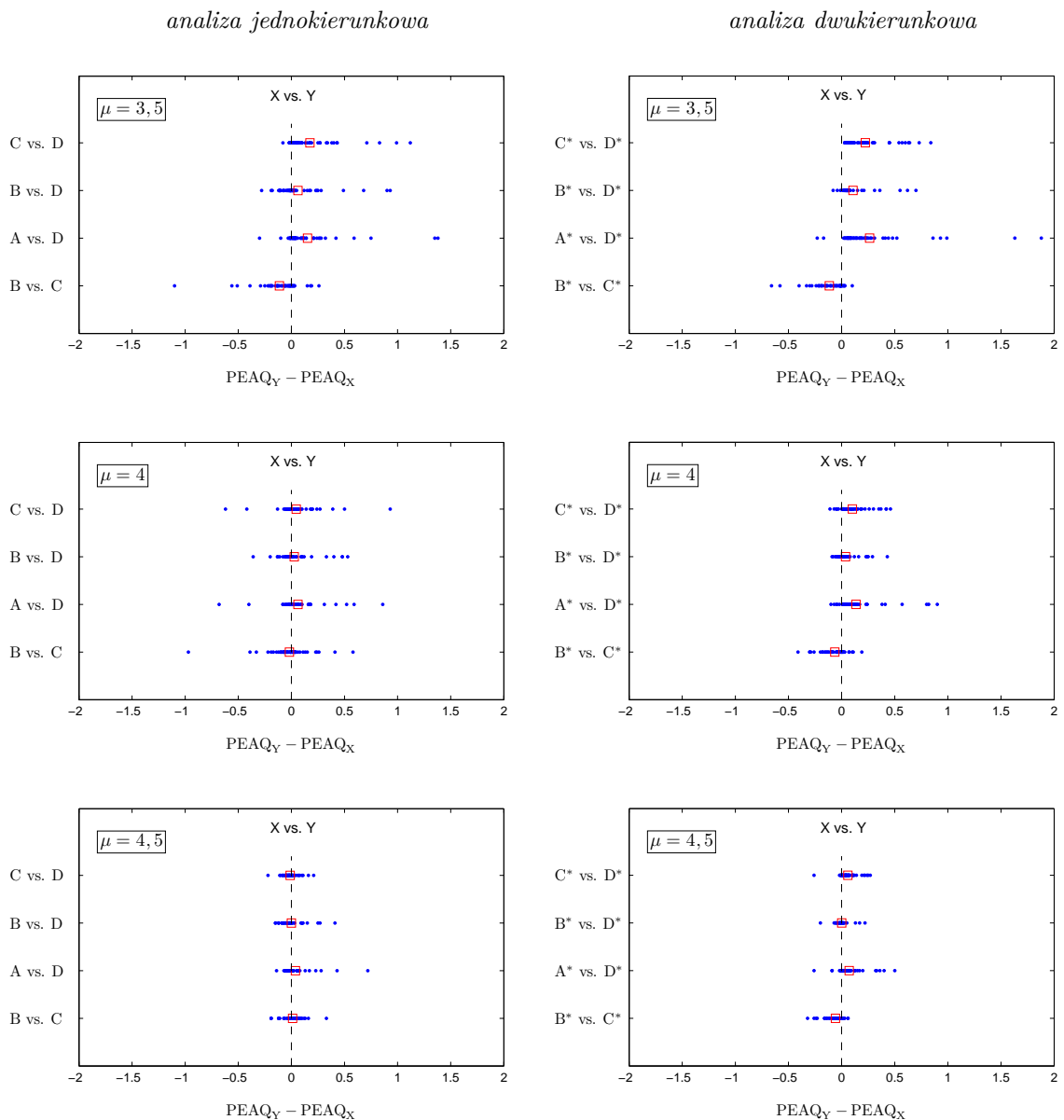
Rysunek 8.11 przedstawia łączne wyniki porównań uzyskane dla analizy jednokierunkowej oraz dla analizy dwukierunkowej.

Na rysunku 8.12 przedstawiono porównanie ilościowe wyników zamieszczonych w tabeli C.9. Analizując zamieszczone wykresy parami, dla tych samych wartości μ , można zauważyć, że po zastosowaniu analizy dwukierunkowej przewaga ilościowa w wynikach PEAQ staje się wyraźniejsza dla proponowanego podejścia – patrz A/A* vs. D/D*, B/B* vs. D/D* oraz C/C* vs. D/D*. Dla tych samych porównań różnice w wynikach PEAQ stają się coraz większe na korzyść podejścia D/D* w przypadku gdy wartość parametru μ maleje. Zwróćmy uwagę na porównanie A/A* vs. D/D*. W wielu przypadkach różnice w wynikach PEAQ są bardzo znaczące na korzyść podejścia D/D*, dla kilku nagrań przekraczają nawet wartość 0,5.





Rysunek 8.11: Jakościowe porównanie skuteczności działania 4 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR (A/A*), podejście oparte na modelu SAR z przyczynowym oszacowaniem okresu podstawowego (B/B*), podejście oparte na modelu SAR z nieprzyczynowym oszacowaniem okresu podstawowego (C/C*) oraz proponowane podejście oparte na współpracy powyższych trzech podejść (D/D*). Porównania przeprowadzono dla trzech typowych wartości współczynnika średniego odchylenia standardowego $\mu \in \{3, 5, 4, 4, 5\}$. Wszystkie wyniki otrzymano dla grupy 50 nagrań testowych: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Żółte słupki pokazują ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu X vs. Y. Zielone słupki pokazują liczbę zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść.



Rysunek 8.12: Ilościowe porównanie skuteczności działania 4 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR (A/A*), podejście oparte na modelu SAR z przyczynowym oszacowaniem okresu podstawowego (B/B*), podejście oparte na modelu SAR z nieprzyczynowym oszacowaniem okresu podstawowego (C/C*) oraz proponowane podejście oparte na współpracy powyższych trzech podejść (D/D*). Porównania przeprowadzono dla trzech typowych wartości współczynnika średniego odchylenia standardowego $\mu \in \{3, 5, 4, 4, 5\}$. Wszystkie wyniki otrzymano dla grupy 50 nagrań testowych: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Dla każdej porównywanej pary X vs. Y zaznaczono punktami wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$. Wyniki dodatnie są na korzyść podejścia Y, a wyniki ujemne na korzyść podejścia X. Wartość średnia z uzyskanych 50 wyników została zaznaczona kwadratem.

Podsumowanie

Przedstawione wyniki porównań jakościowych i ilościowych potwierdzają fakt, że proponowane podejścia oparte na modelowaniu dwuzakresowym – w przypadku obecności okresu podstawowego w sygnale – dają lepsze wyniki niż podejście oparte na modelowaniu krótkozakresowym. Przewaga proponowanego podejścia staje się jeszcze bardziej wyraźniejsza, gdy stosowane jest ono w połączeniu z analizą dwukierunkową. Proponowany schemat detekcji wykorzystujący równoległe działanie trzech detektorów cechuje zwiększona odporność na podnoszenie fałszywych alarmów. W efekcie, w przetwarzanym sygnale dokonywane są interwencje głównie tam gdzie rzeczywiście pojawiają się zakłócenia impulsowe.

8.2.3 Model wektorowy – wyniki badań

Ocenione w poprzednich eksperymentach symulacyjnych podejścia oparte na skalarnym modelu AR/SAR wykorzystujące analizę jednokierunkową/dwukierunkową przetwarzają niezależnie lewy i prawy kanał nagrania stereofonicznego. W tej części pokażemy, że w przypadku zastosowania łącznej analizy lewego i prawego kanału sygnału stereofonicznego można uzyskać jeszcze lepsze wyniki zarówno w przypadku adaptacyjnej interpolacji sygnału jak wtedy gdy interpolacja połączona zostaje z adaptacyjną detekcją zakłóceń impulsowych.

Porównywane podejścia

W pierwszych dwóch eksperymentach porównano skuteczność podejść opartych na skalarnym (AR, SAR) i wektorowym (VAR, SVAR) modelowaniu sygnałów fonicznych. W pierwszym przypadku, dwa przebiegi nagrania stereofonicznego reprezentujące odpowiednio lewy i prawy kanał dźwiękowy, zostały rozdzielone i osobno przetworzone. W drugim przypadku oba przebiegi nagrania stereofonicznego były analizowane łącznie. W ramach badania porównano skuteczność 6 podejść do eliminacji zakłóceń impulsowych wykorzystujących analizę jednokierunkową (lewostronną)/dwukierunkową: podejście oparte na skalarnym modelu AR wykorzystujące detekcję z dwoma progami (A/A^*), podejście oparte na skalarnym modelu AR wykorzystujące detekcję w pętli otwartej (B/B^*), podejście oparte na skalarnym modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C^*), podejście oparte na skalarnym modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D^*), podejście oparte na wektorowym modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E^*) oraz podejście oparte na wektorowym modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F^*).

W ramach kolejnych dwóch eksperymentów przeprowadzono formalne testy odsłuchowe. W pierwszym teście wykorzystano czyste sygnały mowy spreparowane poprzez dodanie rzeczywistych zakłóceń impulsowych. W drugim teście wykorzystano już rzeczywiste, archiwalne nagrania gramofonowe.

Parametry symulacji

We wszystkich podejściach rząd modelu krótkozakresowego AR/VAR był taki sam i wynosił $r = 10$. Aby rzetelnie porównać wyniki detekcji/rekonstrukcji oparte na wektorowym modelowaniu sygnału z tymi uzyskanymi dla przypadku skalarnego, należy upewnić się, że w warunkach stacjonarności odpowiadające im wektorowe/skalarne algorytmy identyfikacji zapewniają tę samą dokładność estymacji [73]. Ponieważ modele wektorowe wymagają estymacji dwa razy większej liczby współczynników niż ich skalarne odpowiedniki, aby zagwarantować rzetelne porównanie, ich efektywne szerokości okien estymacji były dwa razy większe. Identyfikacja sygnału przeprowadzona została przy użyciu algorytmu EWLS [przypadek skalarny (3.39) oraz przypadek wektorowy (6.11)] wyposażonego w stałe zapomnienia: $\lambda = 0,999$ – w przypadku przetwarzania wektorowego i [w zgodzie z (6.18)] $\lambda_1 = \lambda_2 = 0,998$ – w przypadku przetwarzania skalarnego. W sytuacji zgłoszenia alarmu detekcyjnego sprawdzano stabilność aktualnego modelu. W przypadku stwierdzenia niestabilności dokonywano ponownej identyfikacji modelu przy użyciu algorytmu gwarantującego stabilność: w przypadku skalarnym stosowano algorytm Levinsona-Durbina dla okna prostokątnego o szerokości $N = 1000$, natomiast w przypadku wektorowym – algorytm Whittle'a-Wigginsa-Robinsona dla okna prostokątnego o szerokości $N = 2000$.

Tablica 8.19: Porównanie skuteczności działania 4 podejść do adaptacyjnej interpolacji sygnału na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR, podejście oparte na modelu SAR, podejście oparte na modelu VAR oraz podejście oparte na modelu SVAR. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań oraz dla 50 spreparowanych sygnałów fonicznych z testowej bazy nagrań. Rekonstrukcja nagrań została przeprowadzona przy pełnej wiedzy o położeniu zakłóceń impulsowych. Śr.₁₀ i Śr.₅₀ oznaczają uśrednione wyniki w grupie 10 wyszczególnionych nagrań oraz w grupie wszystkich 50 nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

model	1	2	3	4	5	6	7	8	9	10	Śr. ₁₀	Śr. ₅₀
AR	-0,38	-0,27	-0,32	-0,13	-0,65	-0,32	-0,76	-0,47	-0,43	-0,21	-0,39	-0,33
SAR	-0,34	-0,24	-0,32	-0,14	-0,60	-0,26	-0,63	-0,42	-0,38	-0,21	-0,36	-0,29
VAR	-0,18	-0,31	-0,26	-0,11	-0,53	-0,23	-0,47	-0,28	-0,36	-0,20	-0,29	-0,22
SVAR	-0,20	-0,31	-0,26	-0,14	-0,53	-0,21	-0,43	-0,26	-0,34	-0,21	-0,29	-0,21

testowa baza nagrań

model	1	2	3	4	5	6	7	8	9	10	Śr. ₁₀	Śr. ₅₀
AR	-0,38	-0,28	-0,41	-0,18	-0,63	-0,32	-0,83	-0,49	-0,40	-0,27	-0,42	-0,35
SAR	-0,35	-0,25	-0,39	-0,19	-0,58	-0,27	-0,69	-0,47	-0,36	-0,25	-0,38	-0,31
VAR	-0,17	-0,19	-0,31	-0,16	-0,42	-0,20	-0,51	-0,33	-0,35	-0,23	-0,29	-0,23
SVAR	-0,15	-0,17	-0,31	-0,17	-0,38	-0,18	-0,44	-0,30	-0,35	-0,21	-0,27	-0,21

W przypadku podejścia A/A* przyjęte zostały domyślne wartości parametrów zalecane w pracy [50]. W pozostałych podejściach mnożnik średniego odchylenia standardowego ustalony został na $\mu = 4,5$, a maksymalna długość alarmów detekcyjnych ustawiona została na $m_{\max} = 50$. Współczynnik szerokości pasma filtru górnoprzepustowego ustalony był na $c = 0,995$, a stała zapominania użyta do aktualizacji wariancji w przypadku skalarnym oraz macierzy kowariancji w przypadku wektorowym wynosiła $\lambda_0 = 0,993$. W przypadku podejść D/D* i F/F* użyto następujących wartości parametrów do estymacji okresu podstawowego: $M_0 = 160$, $M_* = 40$, $m_0 = 20$, $m_* = 5$, $T_{\min} = 110$, $T_{\max} = 1000$, $k_0 = 3$. Przy identyfikacji części długozakresowej modelu szerokość okna wynosiła $L = 40$ dla przypadku modelowania skalarnego oraz $L = 80$ dla przypadku modelowania wektorowego. Przyjęty zapas stabilności wynosił $\epsilon = 0,01$.

We wszystkich podejściach wykorzystujących analizę dwukierunkową stosowano proponowany w pracy zestaw lokalnych reguł łączenia alarmów detekcyjnych, a parametr wpływający na rozszerzenia alarmów detekcyjnych ustawiono na $\Delta_1 = 2$.

Eksperyment 1

Pierwsze badanie porównuje skuteczność działania 4 podejść do adaptacyjnej interpolacji sygnału opartych na skalarnych modelach (AR, SAR) oraz modelach wektorowych (VAR, SVAR). Ponieważ stosowany w poprzednich eksperymentach schemat łączenia rekonstrukcji lewostronnej i rekonstrukcji prawostronnej nie może być użyty w przypadku podejść opartych na modelach wektorowych, porównanie przeprowadzono na podstawie wyników otrzymanych dla rekonstrukcji lewostronnej. Wszystkie sygnały rekonstruowane były przy pełnej wiedzy o położeniu zakłóceń

impulsowych: $\hat{\mathbf{d}}(t) \equiv \mathbf{d}(t)$. Wyniki rekonstrukcji oceniono przy użyciu obiektywnej miary PEAQ. W tabeli C.10 zamieszczono wszystkie wyniki PEAQ dla treningowej bazy nagrań, zaś w tabeli C.11 – wszystkie wyniki PEAQ dla testowej bazy nagrań. W tabeli 8.19 zamieszczono wyniki dla 10 nagrań z bazy treningowej oraz dla 10 nagrań z bazy testowej, uśrednione wyniki w grupie 10 wyszczególnionych nagrań oraz w grupie wszystkich 50 nagrań. Otrzymane wyniki dla obu baz nagrań pozwalają na wyciągnięcie tych samych wniosków, mimo iż użyte sekwencje zakłóceń impulsowych w grupie testowej i w grupie treningowej różniły się.

Wyniki zebrane w tabeli C.10 i w tabeli C.11 wyraźnie pokazują przewagę wektorowego modelowania (VAR vs. AR, SVAR vs. SAR) i przewagę modelowania dwuzakresowego (SAR vs. AR, SVAR vs. VAR). Najlepsze wyniki otrzymujemy gdy obie techniki zostają połączone, tj. gdy użyte jest podejście SVAR. Najmniejsze różnice w wynikach PEAQ pomiędzy modelowaniem wektorowym a skalarnym zauważyć można dla grupy nagrań reprezentującej muzykę chóralną oraz muzykę klasyczną. Natomiast największa przewaga ujawnia się dla grup nagrań reprezentujących muzykę jazzową, operową oraz instrumentalną. W przypadku grupy nagrań z muzyką instrumentalną możemy wskazać przykład nagrania pierwszego z dominującą obecnością skrzypiec, dla którego podejścia oparte na modelach AR, SAR, VAR oraz SVAR otrzymały następujące wyniki PEAQ: $-0,83$, $-0,69$, $-0,51$ oraz $-0,44$. Dla tego nagrania przewaga podejścia opartego na modelu SVAR w porównaniu do podejścia opartego na klasycznym modelu AR jest prawie dwukrotna. Różnica ta jest wyraźnie zauważalna w testach odsłuchowych.

Eksperyment 2

Drugie badanie porównuje skuteczność działania 6 podejść wykorzystujących analizę jednokierunkową/dwukierunkową, gdy interpolacja sygnału połączona zostaje z adaptacyjną detekcją zakłóceń impulsowych. W tabelach C.12 i C.13 zamieszczono wszystkie oceny PEAQ uzyskane przez 6 podejść, odpowiednio dla grupy nagrań treningowych oraz dla grupy nagrań testowych. W tabeli 8.20 zamieszczono jedynie po 10 nagrań z obu baz. Dla każdego nagrania najlepszy wynik PEAQ został wyróżniony wytłuszczoną czcionką. Najwięcej najlepszych ocen uzyskało podejście oparte na wektorowym modelu dwuzakresowym wykorzystujące analizę dwukierunkową (F^*), odpowiednio 33 i 35 dla grupy treningowej i dla grupy testowej. Wektorowe podejście E^* znalazło się na drugim miejscu otrzymując łącznie 29 najlepszych ocen, a na trzecim miejscu z łącznym wynikiem 21 – podejście oparte na skalarnym modelu dwuzakresowym (D^*). W każdej analizowanej grupie nagrań fonicznych z muzyką jazzową, muzyką chóralną, muzyką operową, muzyką instrumentalną oraz muzyką klasyczną, podejście F^* uzyskało najlepszy średni wynik PEAQ, a największa przewaga nad pozostałymi podejściami widoczna jest w przypadku grupy nagrań z muzyką jazzową. Otrzymane wyniki porównano jakościowo oraz ilościowo.

Porównanie jakościowe

Jakościowe porównanie wyników z tabel C.12 i C.13 zaprezentowano w tabelach 8.21 i 8.22. Na podstawie wyników bezpośredniego porównania pomiędzy poszczególnymi podejściami można wyciągnąć cztery najważniejsze wnioski.

Po pierwsze, prawie we wszystkich przypadkach podejścia oparte na wektorowej analizie sygnału uzyskały znacząco lepsze wyniki niż metody oparte na detekcji z dwoma progami oraz na detekcji w pętli otwartej – patrz A/A^* vs. E/E^* , B/B^* vs. E/E^* oraz A/A^* vs. F/F^* , B/B^* vs. F/F^* .

Po drugie, prawie wszystkie testowane podejścia oparte na analizie błędów predykcji jednokrokowej uzyskały lepsze wyniki gdy zastosowano analizę dwukierunkową – patrz B vs. B^* ,

Tablica 8.20: Porównanie skuteczności działania 6 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A/A*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B/B*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E*) oraz podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F*). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań oraz dla 50 spreparowanych sygnałów fonicznych z testowej bazy nagrań. REF oznacza początkową ocenę PEAQ dla zakłóconego nagrania. Śr.₁₀ i Śr.₅₀ oznaczają uśrednione wyniki w grupie 10 wyszczególnionych nagrań oraz w grupie wszystkich 50 nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,61	-2,32	-3,63	-0,96	-0,90	-0,82	-0,86	-1,73	-0,80	-0,51	-0,53	-0,38	-0,36
2	-2,80	-1,23	-1,72	-1,05	-0,62	-0,68	-0,56	-1,24	-1,12	-0,93	-0,43	-0,57	-0,40
3	-3,91	-3,67	-3,84	-2,51	-2,55	-3,09	-3,14	-3,46	-1,07	-0,54	-0,80	-0,51	-0,48
4	-3,91	-3,36	-3,83	-0,84	-0,91	-0,77	-0,85	-2,82	-0,62	-0,27	-0,26	-0,25	-0,27
5	-3,80	-2,11	-3,63	-1,18	-1,18	-1,09	-1,06	-1,57	-0,84	-0,60	-0,52	-0,52	-0,53
6	-3,82	-2,36	-3,75	-0,98	-1,04	-0,97	-0,88	-1,80	-0,74	-0,44	-0,37	-0,38	-0,34
7	-3,68	-1,27	-3,61	-1,03	-0,98	-0,78	-0,77	-1,22	-0,89	-0,71	-0,58	-0,48	-0,44
8	-3,38	-0,74	-2,91	-0,54	-0,54	-0,45	-0,45	-0,79	-0,45	-0,36	-0,32	-0,25	-0,24
9	-3,82	-2,91	-3,75	-1,06	-1,08	-0,99	-1,03	-2,30	-0,85	-0,42	-0,37	-0,42	-0,39
10	-3,66	-1,40	-3,24	-0,42	-0,43	-0,39	-0,41	-0,98	-0,33	-0,21	-0,21	-0,19	-0,18
Śr. ₁₀	-3,64	-2,14	-3,39	-1,06	-1,02	-1,00	-1,00	-1,79	-0,77	-0,50	-0,44	-0,39	-0,36
Śr. ₅₀	-3,56	-2,12	-3,26	-0,97	-0,93	-0,85	-0,84	-1,75	-0,83	-0,50	-0,43	-0,41	-0,38

testowa baza nagrań

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,71	-1,94	-3,55	-1,25	-1,21	-0,98	-1,00	-1,69	-1,12	-0,62	-0,67	-0,61	-0,37
2	-1,89	-1,23	-1,60	-1,09	-0,64	-0,67	-0,55	-1,28	-0,86	-0,87	-0,45	-0,57	-0,41
3	-3,91	-3,67	-3,84	-1,48	-1,53	-1,62	-1,39	-3,55	-2,29	-0,53	-0,51	-0,47	-0,47
4	-3,91	-3,38	-3,84	-0,94	-1,01	-0,89	-0,98	-2,87	-0,72	-0,29	-0,30	-0,28	-0,28
5	-3,85	-2,11	-3,63	-1,12	-1,07	-0,91	-0,89	-1,63	-0,88	-0,56	-0,50	-0,41	-0,41
6	-3,87	-2,46	-3,71	-1,12	-1,10	-1,08	-1,13	-2,02	-0,86	-0,44	-0,50	-0,32	-0,31
7	-3,82	-1,26	-3,59	-1,04	-1,00	-0,82	-0,81	-1,58	-0,97	-0,68	-0,59	-0,50	-0,46
8	-3,60	-0,70	-3,17	-0,61	-0,60	-0,46	-0,46	-0,81	-0,54	-0,38	-0,34	-0,26	-0,25
9	-3,89	-2,92	-3,79	-1,18	-1,17	-1,09	-1,14	-2,65	-1,17	-0,47	-0,43	-0,45	-0,45
10	-3,80	-1,42	-3,29	-0,44	-0,45	-0,42	-0,45	-1,02	-0,35	-0,21	-0,19	-0,22	-0,19
Śr. ₁₀	-3,63	-2,11	-3,40	-1,02	-0,98	-0,89	-0,88	-1,91	-0,98	-0,51	-0,45	-0,41	-0,36
Śr. ₅₀	-3,68	-2,11	-3,25	-0,98	-0,94	-0,86	-0,85	-1,87	-0,94	-0,53	-0,44	-0,45	-0,38

Tablica 8.21: Bezpośrednie porównanie skuteczności działania 6 podejść wykorzystujących analizę jednokierunkową/dwukierunkową: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A/A*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B/B*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E*), podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F*). Porównywane parami podejścia podane zostały w pierwszej kolumnie. W pozostałych kolumnach umieszczone zostały wyniki, mówiące ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu. Wyniki umieszczone w nawiasach klamrowych informują o liczbie zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść. Wyniki otrzymano dla 50 spreparowanych sygnałów fonicznych z grupy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną.

porównywane podejścia	muzyka jazzowa	muzyka chóralna	muzyka operowa
A vs. F	1 {0} vs. 9 { 9 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B vs. F	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
C vs. F	0 {0} vs. 10 { 6 }	2 {1} vs. 8 { 2 }	0 {0} vs. 10 { 8 }
D vs. F	2 { 1 } vs. 8 { 1 }	2 {1} vs. 8 { 2 }	0 {0} vs. 10 { 8 }
E vs. F	4 {0} vs. 8 { 4 }	10 { 1 } vs. 1 { 0 }	7 {0} vs. 3 { 0 }
A* vs. F*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B* vs. F*	0 {0} vs. 10 { 9 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
C* vs. F*	2 {0} vs. 10 { 7 }	1 {0} vs. 10 { 1 }	0 {0} vs. 10 { 9 }
D* vs. F*	4 {0} vs. 9 { 2 }	2 {0} vs. 8 { 1 }	1 {0} vs. 9 { 1 }
E* vs. F*	3 {0} vs. 8 { 4 }	8 {0} vs. 4 { 1 }	2 {0} vs. 9 { 0 }
A vs. A*	2 {1} vs. 8 { 7 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B vs. B*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
C vs. C*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
D vs. D*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
E vs. E*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
F vs. F*	0 {0} vs. 10 { 9 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
porównywane podejścia	muzyka instrumentalna	muzyka klasyczna	łączy wynik
A vs. F	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	1 {0} vs. 49 { 49 }
B vs. F	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
C vs. F	0 {0} vs. 10 { 6 }	2 {0} vs. 8 { 3 }	4 {1} vs. 46 { 25 }
D vs. F	0 {0} vs. 10 { 6 }	2 {0} vs. 8 { 3 }	6 {2} vs. 44 { 20 }
E vs. F	7 {0} vs. 4 { 0 }	9 {0} vs. 2 { 1 }	37 {1} vs. 18 { 5 }
A* vs. F*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
B* vs. F*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 9 }	0 {0} vs. 50 { 48 }
C* vs. F*	0 {0} vs. 10 { 7 }	2 {0} vs. 8 { 1 }	5 {0} vs. 48 { 25 }
D* vs. F*	0 {0} vs. 10 { 3 }	6 {0} vs. 5 { 1 }	13 {0} vs. 41 { 8 }
E* vs. F*	2 {0} vs. 9 { 0 }	5 {0} vs. 7 { 1 }	20 {0} vs. 37 { 6 }
A vs. A*	3 {0} vs. 7 { 5 }	0 {0} vs. 10 { 9 }	5 {1} vs. 45 { 41 }
B vs. B*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
C vs. C*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
D vs. D*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
E vs. E*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 9 }	0 {0} vs. 50 { 49 }

kontynuacja tabeli

porównywane podejścia	muzyka jazzowa	muzyka chóralna	muzyka operowa
A vs. D	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
B vs. D	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
C vs. D	2 {0} vs. 9 {4}	9 {1} vs. 1 {0}	4 {0} vs. 7 {0}
A vs. E	1 {1} vs. 9 {9}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
B vs. E	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
C vs. E	0 {0} vs. 10 {4}	1 {1} vs. 9 {2}	0 {0} vs. 10 {8}
A* vs. D*	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
B* vs. D*	0 {0} vs. 10 {8}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
C* vs. D*	4 {0} vs. 8 {5}	5 {1} vs. 8 {1}	0 {0} vs. 10 {2}
A* vs. E*	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
B* vs. E*	0 {0} vs. 10 {7}	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}
C* vs. E*	1 {0} vs. 9 {5}	0 {0} vs. 10 {1}	0 {0} vs. 10 {5}

porównywane podejścia	muzyka instrumentalna	muzyka klasyczna	łączy wynik
A vs. D	0 {0} vs. 10 {9}	0 {0} vs. 10 {10}	0 {0} vs. 50 {49}
B vs. D	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 50 {50}
C vs. D	5 {0} vs. 7 {1}	7 {0} vs. 5 {1}	27 {1} vs. 29 {6}
A vs. E	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	1 {1} vs. 49 {49}
B vs. E	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 50 {50}
C vs. E	0 {0} vs. 10 {6}	1 {0} vs. 10 {3}	2 {1} vs. 49 {23}
A* vs. D*	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 50 {50}
B* vs. D*	0 {0} vs. 10 {9}	0 {0} vs. 10 {9}	0 {0} vs. 50 {46}
C* vs. D*	1 {0} vs. 10 {2}	3 {0} vs. 8 {1}	13 {1} vs. 44 {11}
A* vs. E*	0 {0} vs. 10 {10}	0 {0} vs. 10 {10}	0 {0} vs. 50 {50}
B* vs. E*	0 {0} vs. 10 {10}	0 {0} vs. 10 {9}	0 {0} vs. 50 {46}
C* vs. E*	1 {0} vs. 10 {6}	4 {0} vs. 8 {2}	6 {0} vs. 47 {19}

C vs. C*, D vs. D*, E vs. E* oraz F vs. F* w tabelach 8.21 i 8.22 oraz na rysunku 8.13. Wyjątek stanowi porównanie E vs. E* w grupie testowej, które dla jednego nagrania dało taki sam wynik PEAQ w analizie jednokierunkowej jak dwukierunkowej oraz w jednym przypadku dało wynik gorszy po zastosowaniu analizy dwukierunkowej. Wyraźną poprawę skuteczności stwierdzono niemal we wszystkich przypadkach zarówno dla nagrań treningowych jak testowych. W przypadku podejścia opartego na analizie błędów resztowych A vs. A* poprawa skuteczności wystąpiła łącznie w 85 na 100 nagrań testowych, w tym w 78 była to poprawa znacząca.

Po trzecie, podejścia oparte na wektorowej analizie sygnału wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym uzyskały w znaczącej większości przypadków lepsze wyniki niż ich skalarne odpowiedniki – patrz C vs. E i C* vs. E* oraz D vs. F i D* vs. F* w tabelach 8.21 i 8.22 oraz na rysunku 8.13. W przypadku porównania C vs. E i C* vs. E* podejście wektorowe dało lepsze wyniki aż 91 razy dla analizy jednokierunkowej i 83 razy dla analizy dwukierunkowej. Ponadto podejście wektorowe dało łącznie 4 razy ten sam wynik co podejście skalarne. Znacząca przewaga podejścia wektorowego obserwowana była łącznie w 44 przypadkach dla analizy jednokierunkowej i w 39 przypadkach dla analizy dwukierunkowej. W przypadku porównania modelowania dwuzakresowego D vs. F i D* vs. F* wektorowe podejście okazało się lepsze aż 85 razy dla analizy jednokierunkowej i 74 razy dla analizy dwukierunkowej. Ponadto wynik nie

Tablica 8.22: Bezpośrednie porównanie skuteczności działania 6 podejść wykorzystujących analizę jednokierunkową/dwukierunkową: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A/A*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B/B*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E*), podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F*). Porównywane parami podejścia podane zostały w pierwszej kolumnie. W pozostałych kolumnach umieszczone zostały wyniki, mówiące ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu. Wyniki umieszczone w nawiasach klamrowych informują o liczbie zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść. Wyniki otrzymano dla 50 spreparowanych sygnałów fonicznych z grupy testowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną.

porównywane podejścia	muzyka jazzowa	muzyka chóralna	muzyka operowa
A vs. F	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B vs. F	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
C vs. F	2 {0} vs. 9 { 5 }	1 {0} vs. 9 { 2 }	2 {0} vs. 8 { 6 }
D vs. F	3 {1} vs. 8 { 3 }	1 {0} vs. 9 { 4 }	2 {0} vs. 8 { 5 }
E vs. F	4 {0} vs. 7 { 4 }	8 {0} vs. 2 { 1 }	6 {0} vs. 4 { 1 }
A* vs. F*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B* vs. F*	0 {0} vs. 10 { 9 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
C* vs. F*	0 {0} vs. 10 { 7 }	0 {0} vs. 10 { 1 }	0 {0} vs. 10 { 9 }
D* vs. F*	5 {0} vs. 6 { 1 }	1 {0} vs. 9 { 0 }	0 {0} vs. 10 { 4 }
E* vs. F*	2 {0} vs. 8 { 6 }	6 {0} vs. 7 { 2 }	1 {0} vs. 10 { 1 }
A vs. A*	3 {1} vs. 7 { 6 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B vs. B*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
C vs. C*	0 {0} vs. 10 { 9 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
D vs. D*	0 {0} vs. 10 { 8 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
E vs. E*	1 {0} vs. 9 { 9 }	1 {0} vs. 10 { 9 }	0 {0} vs. 10 { 10 }
F vs. F*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
porównywane podejścia	muzyka instrumentalna	muzyka klasyczna	łączy wynik
A vs. F	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 9 }	0 {0} vs. 50 { 49 }
B vs. F	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
C vs. F	0 {0} vs. 10 { 8 }	5 {0} vs. 7 { 1 }	10 {0} vs. 43 { 22 }
D vs. F	0 {0} vs. 10 { 8 }	3 {0} vs. 8 { 2 }	9 {1} vs. 43 { 22 }
E vs. F	6 {0} vs. 6 { 0 }	8 {0} vs. 2 { 1 }	32 {0} vs. 21 { 7 }
A* vs. F*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
B* vs. F*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 9 }	0 {0} vs. 50 { 48 }
C* vs. F*	0 {0} vs. 10 { 7 }	2 {0} vs. 9 { 3 }	2 {0} vs. 49 { 27 }
D* vs. F*	1 {0} vs. 10 { 5 }	6 {0} vs. 6 { 1 }	13 {0} vs. 41 { 11 }
E* vs. F*	2 {0} vs. 8 { 1 }	3 {0} vs. 10 { 2 }	14 {0} vs. 43 { 12 }
A vs. A*	7 { 4 } vs. 3 { 2 }	0 {0} vs. 10 { 9 }	10 {5} vs. 40 { 37 }
B vs. B*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
C vs. C*	0 {0} vs. 10 { 9 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 48 }
D vs. D*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 48 }
E vs. E*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 9 }	2 {0} vs. 49 { 47 }
F vs. F*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }

kontynuacja tabeli

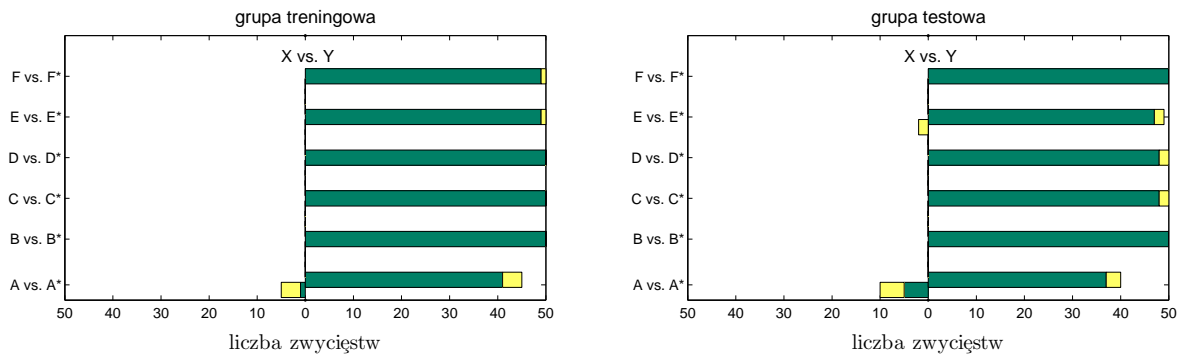
porównywane podejścia	muzyka jazzowa	muzyka chóralna	muzyka operowa
A vs. D	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B vs. D	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
C vs. D	3 {0} vs. 8 { 4 }	9 {0} vs. 1 {0}	1 {0} vs. 9 { 1 }
A vs. E	1 {0} vs. 10 { 9 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B vs. E	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
C vs. E	3 {0} vs. 9 { 4 }	2 {1} vs. 9 { 3 }	1 {0} vs. 9 { 7 }
A* vs. D*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B* vs. D*	0 {0} vs. 10 { 8 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
C* vs. D*	3 {0} vs. 8 { 6 }	5 {0} vs. 6 { 1 }	1 {0} vs. 9 { 3 }
A* vs. E*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }
B* vs. E*	1 {1} vs. 9 { 6 }	0 {0} vs. 10 { 9 }	0 {0} vs. 10 { 10 }
C* vs. E*	2 {1} vs. 8 { 3 }	3 {1} vs. 9 { 2 }	0 {0} vs. 10 { 7 }

porównywane podejścia	muzyka instrumentalna	muzyka klasyczna	łączy wynik
A vs. D	0 {0} vs. 10 { 9 }	0 {0} vs. 10 { 9 }	0 {0} vs. 50 { 48 }
B vs. D	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
C vs. D	4 {0} vs. 9 { 1 }	8 {0} vs. 4 {0}	25 {0} vs. 31 { 6 }
A vs. E	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 9 }	1 {0} vs. 50 { 48 }
B vs. E	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
C vs. E	0 {0} vs. 10 { 8 }	1 {0} vs. 9 { 2 }	7 {1} vs. 46 { 24 }
A* vs. D*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
B* vs. D*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 9 }	0 {0} vs. 50 { 47 }
C* vs. D*	0 {0} vs. 10 { 2 }	2 {1} vs. 9 { 1 }	11 {1} vs. 42 { 13 }
A* vs. E*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 10 }	0 {0} vs. 50 { 50 }
B* vs. E*	0 {0} vs. 10 { 10 }	0 {0} vs. 10 { 9 }	1 {1} vs. 49 { 44 }
C* vs. E*	1 {0} vs. 9 { 7 }	5 {1} vs. 6 { 1 }	11 {3} vs. 42 { 20 }

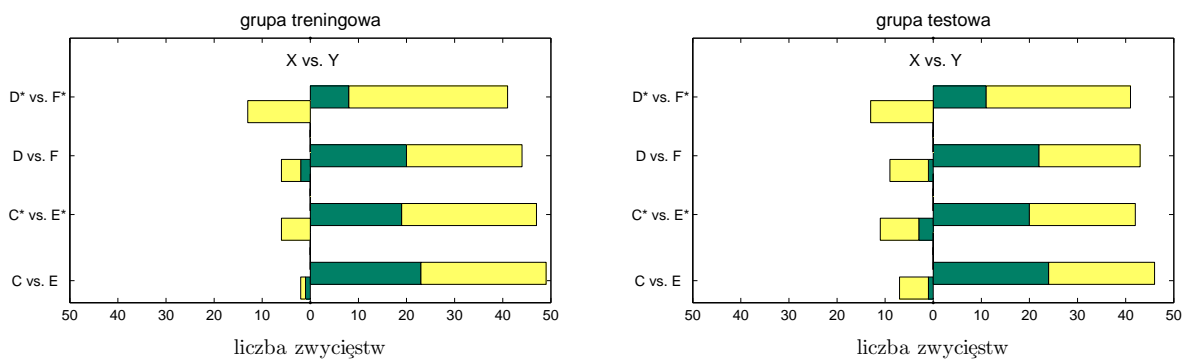
uległ zmianie łącznie dla 10 przypadków. Zaobserwowana, znacząca przewaga podejścia wektorowego wystąpiła w 42 przypadkach dla analizy jednokierunkowej i w 19 przypadkach dla analizy dwukierunkowej. Oba podejścia wektorowe wypadły najlepiej w grupie nagrań z muzyką operową i muzyką instrumentalną (zazwyczaj w tych, w których lokalna korelacja pomiędzy lewym, a prawym kanałem była silna).

Wreszcie, zastosowanie podejścia opartego na dwuzakresowym modelowaniu sygnału fonicznego daje lepsze rezultaty niż w przypadku podejścia opartego na modelowaniu krótkozakresowym – patrz C vs. D i C* vs. D* oraz E vs. F i E* vs. F* w tabelach 8.21 i 8.22 oraz na rysunku 8.13. Porównując podejścia C vs. D i C* vs. D* widać nieznaczącą przewagę w ogólnej liczbie lepszych wyników na korzyść podejścia dwuzakresowego dla analizy jednokierunkowej. Z kolei w przypadku analizy dwukierunkowej przewaga modelowania dwuzakresowego staje się już zdecydowanie wyraźniejsza, podejście to uzyskało bowiem 76 razy lepszy wynik i 10 razy ten sam wynik co podejście krótkozakresowe. Porównując podejścia E vs. F i E* vs. F* można zaobserwować, że dla analizy jednokierunkowej model krótkozakresowy uzyskał wprawdzie znacznie więcej lepszych wyników, jednak tylko w jednym przypadku poprawa była znacząca w porównaniu z 12 „wyraźnymi zwycięstwami” uzyskanymi przez podejście dwuzakresowe. W przypadku analizy dwukierunkowej sytuacja staje się jednoznaczna – model dwuzakresowy okazał się 67 razy lepszy i 14 razy uzyskał ten sam wynik co model krótkozakresowy. W tej analizie model

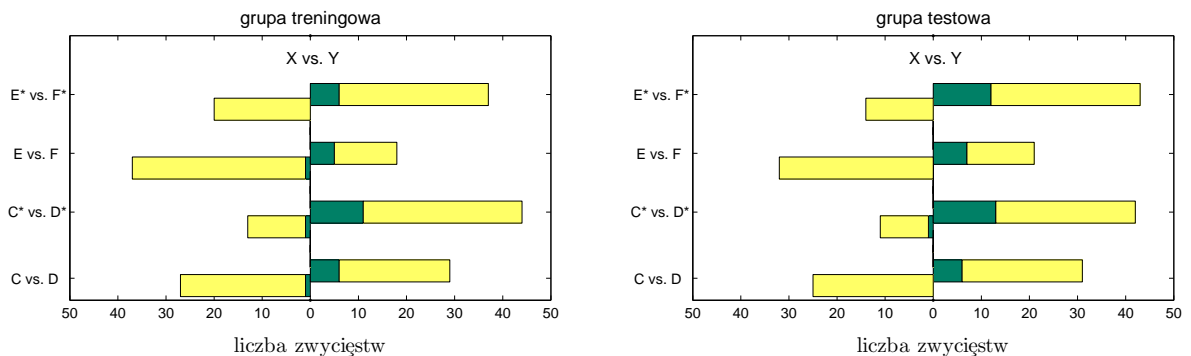
analiza jednokierunkowa vs. analiza dwukierunkowa



modelowanie skalarne vs. modelowanie wektorowe

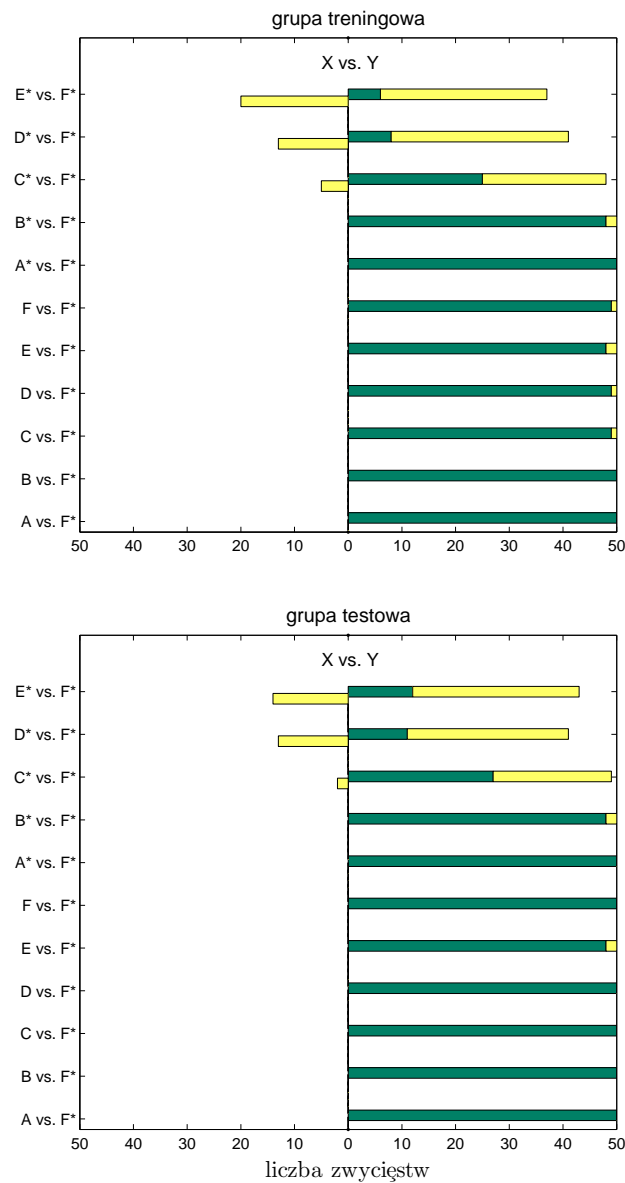


modelowanie krótkozakresowe vs. modelowanie dwuzakresowe



Rysunek 8.13: Jakościowe porównanie skuteczności działania testowanych podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A/A*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B/B*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E*), podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F*). Żółte słupki pokazują ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu X vs. Y. Zielone słupki pokazują liczbę zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść.

testowane podejścia vs. proponowane podejście (F^*)



Rysunek 8.14: Jakościowe porównanie skuteczności działania proponowanego podejścia w stosunku do pozostałych testowanych podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A/A^*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B/B^*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C^*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D^*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E^*), podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F^*). Żółte słupki pokazują ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu X vs. Y. Zielone słupki pokazują liczbę zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść.

dwuzakresowy osiągnął łącznie 18 razy wynik znacząco lepszy. Podejścia oparte na modelowaniu dwuzakresowym, zarówno skalarne jak wektorowe, wypadły zdecydowanie najlepiej w przypadku nagrań z muzyką jazzową, operową oraz instrumentalną.

Na rysunku 8.14 porównano proponowane w pracy podejście F^* z pozostałymi testowanymi podejściami, osobno dla grupy treningowej oraz dla grupy testowej. W obu grupach proponowane podejście osiągnęło najlepsze wyniki. W przypadku porównań A vs. F^* , B vs. F^* , C vs. F^* , D vs. F^* , E vs. F^* , F vs. F^* , A^* vs. F^* oraz B^* vs. F^* , proponowane podejście okazało się bezkonkurencyjne osiągając prawie we wszystkich przypadkach znacząco lepsze wyniki. W przypadku porównań C^* vs. F^* , D^* vs. F^* oraz E^* vs. F^* liczba znacząco lepszych wyników stopniowo maleje. Należy zauważyć, że pozostałe podejścia ani razu nie uzyskały znacząco lepszego wyniku.

Porównanie ilościowe

Ilościowe porównanie wyników z tabel C.12 i C.13 zaprezentowano na rysunku 8.15. Sprawdzono w jakim stopniu wystąpiła poprawa ilościowa w wynikach PEAQ po zastosowaniu analizy dwukierunkowej, modelowania wektorowego oraz modelowania dwuzakresowego. Ponieważ otrzymane wyniki dla obu badanych grup są podobne i prowadzą do tych samych wniosków, podawane poniżej wartości/przedziały wartości będą uwzględniały wszystkie 100 wyników z obu grup. Analizując zamieszczone wykresy można wyciągnąć trzy najważniejsze wnioski.

Po pierwsze, zastosowanie analizy dwukierunkowej przyczynia się do największej poprawy, zarówno jakościowej jak ilościowej – patrz pierwsza para wykresów na rysunku 8.15. W przypadku porównania podejść B vs. B^* , C vs. C^* , D vs. D^* , E vs. E^* oraz F vs. F^* , stwierdzono prawie dla wszystkich podejść wykorzystujących analizę dwukierunkową poprawę ilościową w wynikach PEAQ. W przypadku porównania B vs. B^* wszystkich 100 otrzymanych wyników było znacząco lepszych dla podejścia B^* . Obserwowane różnice w wynikach PEAQ mieszczą się w przedziale od 0,6 do 3,25. Największe skupienie wyników, bo aż 76 pomiarów, znalazło się w przedziale od 2 do 3,25. Wartość średnia dla tego porównania wyniosła 2,37. W przypadku podejść C vs. C^* , D vs. D^* , E vs. E^* oraz F vs. F^* , największe skupienie wyników wystąpiło w przedziale od $-0,07$ do 1,22. Dla każdego podejścia zaobserwowano po jednym pomiarze znacząco odstającym od reszty. Wartości średnie dla tych porównań wyniosły kolejno: 0,46, 0,50, 0,42 oraz 0,47. Powyżej progu 0,5 znalazło się 40 wyników dla porównania C vs. C^* , 42 wyniki dla porównania D vs. D^* , 36 wyników dla porównania E vs. E^* oraz 40 wyników dla porównania F vs. F^* . W przypadku podejścia A vs. A^* obserwowane różnice w wynikach PEAQ mieszczą się w przedziale od $-0,64$ do 0,81. Największa koncentracja pomiarów – aż 95 – wystąpiła w przedziale od $-0,2$ do 0,8. Dla tego porównania średnia wartość była najmniejsza i wyniosła 0,31.

Po drugie, podejścia oparte na wektorowej analizie sygnału wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym uzyskały w większości przypadków znacznie lepsze wyniki PEAQ niż ich skalarne odpowiedniki – patrz C vs. E i C^* vs. E^* oraz D vs. F i D^* vs. F^* . Porównania zilustrowano na rysunku 8.15, patrz druga para wykresów. W przypadku porównań C vs. E i C^* vs. E^* różnice w wynikach PEAQ mieściły się w przedziale od $-0,58$ do 0,48. Największa koncentracja wyników zauważalna jest w przedziale od $-0,1$ do 0,48 – odpowiednio 98 wyników dla porównania C vs. E oraz 97 wyników dla porównania C^* vs. E^* . Średnia poprawa w wynikach PEAQ dla podejść wektorowych wyniosła w tych przypadkach odpowiednio 0,12 i 0,08. Powyżej progu 0,2 znalazło się 25 wyników dla porównania C vs. E oraz 9 wyników dla porównania C^* vs. E^* . W przypadku porównania podejść D vs. F i D^* vs. F^* wszystkie wyniki mieściły się w przedziale od $-0,59$ do 0,43. Najwięcej wyników znalazło się w przedziale od $-0,2$ do 0,43, odpowiednio 98 dla porównania D vs. F oraz wszystkie 100 dla porównania D^* vs. F^* .

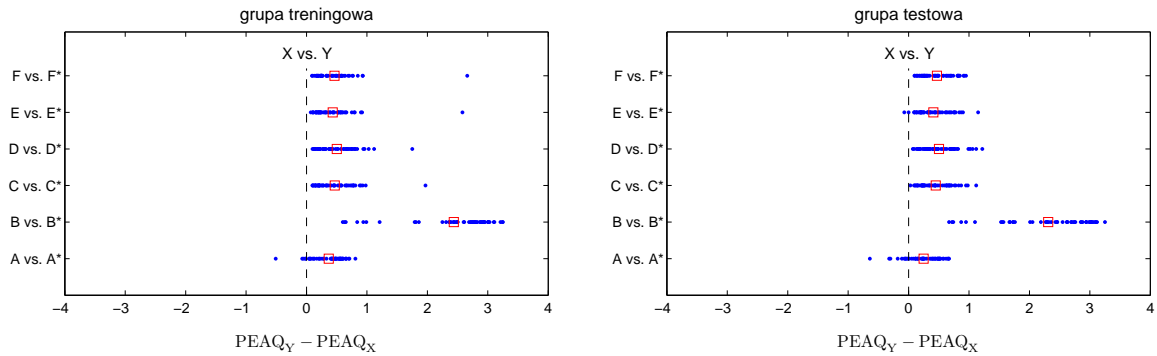
W tych przypadkach wartości średnie wyniosły 0,09 i 0,05. Powyżej progu 0,2 znalazło się 20 wyników dla porównania D vs. F oraz 5 wyników dla porównania D* vs. F*. Na podstawie otrzymanych wyników można stwierdzić, że największa względna poprawa ma miejsce w przypadku podejść opartych na wektorowym modelu wykorzystujących analizę jednokierunkową. Gdy podejścia oparte na wektorowym modelu wykorzystują analizę dwukierunkową, wyraźna poprawa w wynikach PEAQ występuje już w mniejszej liczbie przypadków.

W końcu, zastosowanie podejścia opartego na dwuzakresowym modelowaniu sygnału fonicznego daje lepsze rezultaty niż w przypadku podejścia opartego na modelowaniu krótkozakresowym – patrz C vs. D i C* vs. D* oraz E vs. F i E* vs. F* dolna para wykresów na rysunku 8.13. W przypadku obu podejść C vs. D i C* vs. D*, wyniki mieściły się w przedziale $-0,26$ do $0,72$. Największe skupienie wyników odnotowano w przedziale $-0,1$ do $0,5$, odpowiednio dla 98 wyników w porównaniu C vs. D oraz dla 97 wyników w porównaniu C* vs. D*. Wartości średnie dla tych porównań wyniosły odpowiednio 0,04 oraz 0,08. Powyżej progu 0,2 znalazło się 8 wyników dla porównania C vs. D oraz 15 wyników dla porównania C* vs. D*. W przypadku podejść E vs. F i E* vs. F*, uzyskane wyniki mieściły się w przedziale $-0,1$ do $0,43$. W pierwszym podejściu prawie wszystkie wyniki (96 wyników) koncentrowały się w przedziale $-0,1$ do $0,2$. W drugim przypadku wszystkie wyniki mieściły się w przedziale $-0,1$ do $0,32$ bez odstających od reszty pomiarów. Uzyskane wartości średnie wyniosły odpowiednio 0,01 oraz 0,05. Powyżej progu 0,2 znalazły się 4 wyniki dla porównania E vs. F oraz 12 wyników dla porównania E* vs. F*. Na podstawie otrzymanych wyników można stwierdzić, że podejście oparte na dwuzakresowym modelowaniu sygnału fonicznego daje względnie lepsze wyniki, gdy stosuje się je w analizie dwukierunkowej niż w sytuacji gdy stosuje się je w analizie jednokierunkowej.

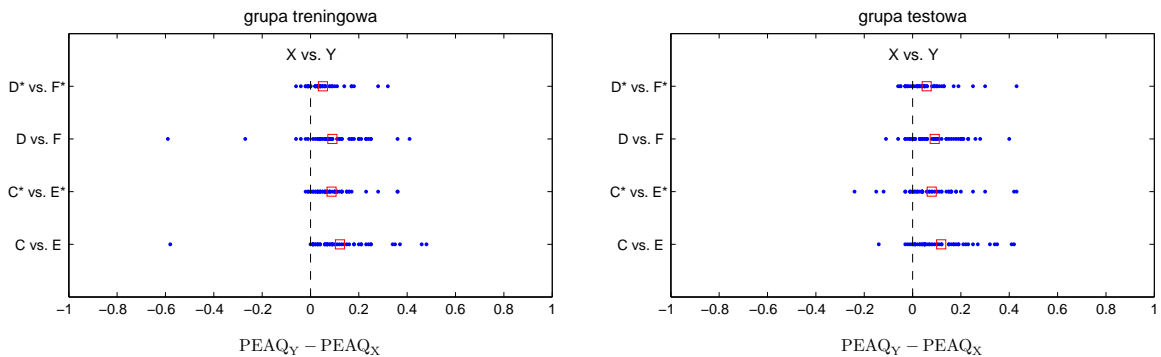
Na rysunku 8.16 porównano proponowane w pracy podejście F* z pozostałymi testowanymi podejściami, osobno dla grupy treningowej oraz dla grupy testowej. W obu grupach proponowane podejście osiągnęło najlepsze wyniki ilościowe. W przypadku porównań A vs. F*, B vs. F*, C vs. F*, D vs. F*, E vs. F*, F vs. F*, A* vs. F* oraz B* vs. F*, proponowane podejście okazało się bezkonkurencyjne osiągając we wszystkich przypadkach lepsze wyniki. Wartości średnie w różnicach wyników PEAQ wyniosły odpowiednio: 1,74, 2,87, 0,59, 0,56, 0,47, 0,47, 1,43, 0,51. W przypadku porównań C* vs. F*, D* vs. F* oraz E* vs. F* przewaga ilościowa w wynikach PEAQ zmalała. Wartości średnie w różnicach wyników PEAQ wyniosły odpowiednio: 0,13, 0,05, 0,05.

Rysunki 8.17 i 8.18 przedstawiają szczegółowe porównanie ilościowe proponowanego w pracy podejścia F* z pozostałymi testowanymi podejściami dla grupy testowej. W przypadku porównania proponowanego podejścia z podejściami wykorzystującymi analizę jednokierunkową – A vs. F*, B vs. F*, C vs. F*, D vs. F*, E vs. F*, F vs. F* (patrz rysunek 8.17) – największe różnice w wynikach PEAQ obserwowane są dla grupy nagrań z muzyką chóralną oraz z muzyką operową. W przypadku porównania proponowanego podejścia z podejściami wykorzystującymi analizę dwukierunkową – A* vs. F* oraz B* vs. F*, C* vs. F*, D* vs. F* oraz E* vs. F* (patrz rysunek 8.18) – największe różnice w wynikach PEAQ obserwowane są dla grupy nagrań z muzyką jazzową, z muzyką operową oraz z muzyką instrumentalną.

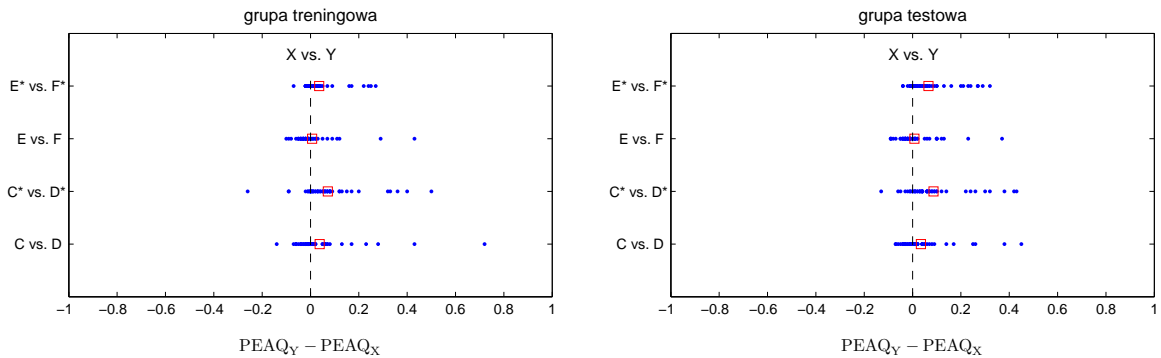
analiza jednokierunkowa vs. analiza dwukierunkowa



modelowanie skalarne vs. modelowanie wektorowe

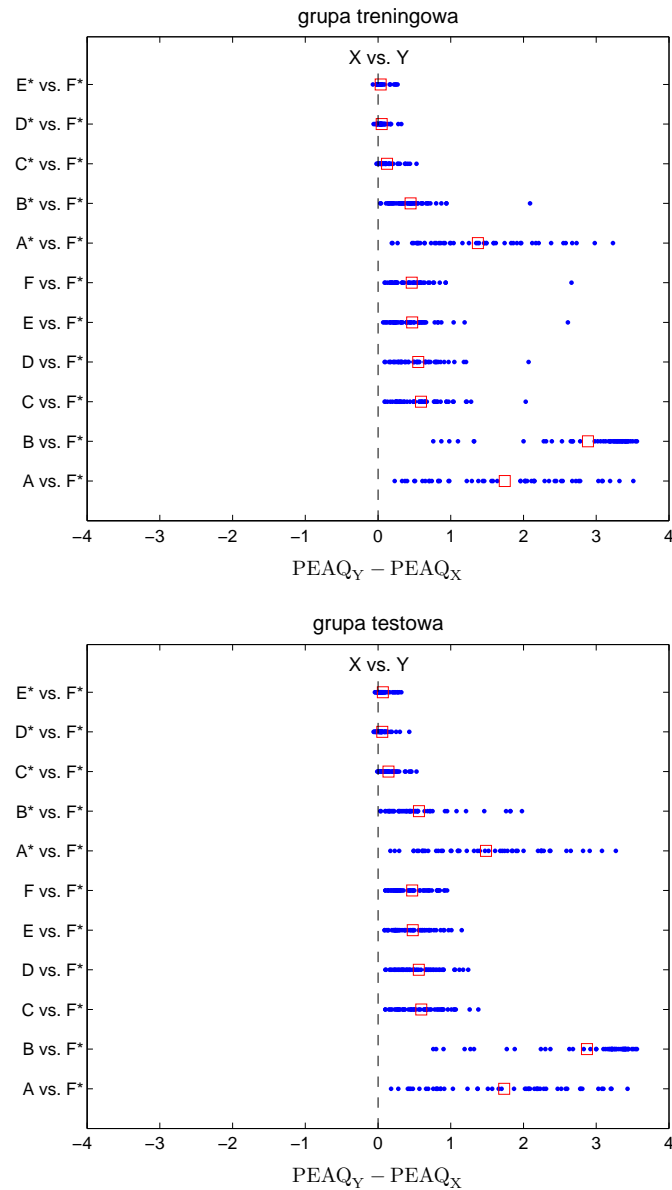


modelowanie krótkozakresowe vs. modelowanie dwuzakresowe

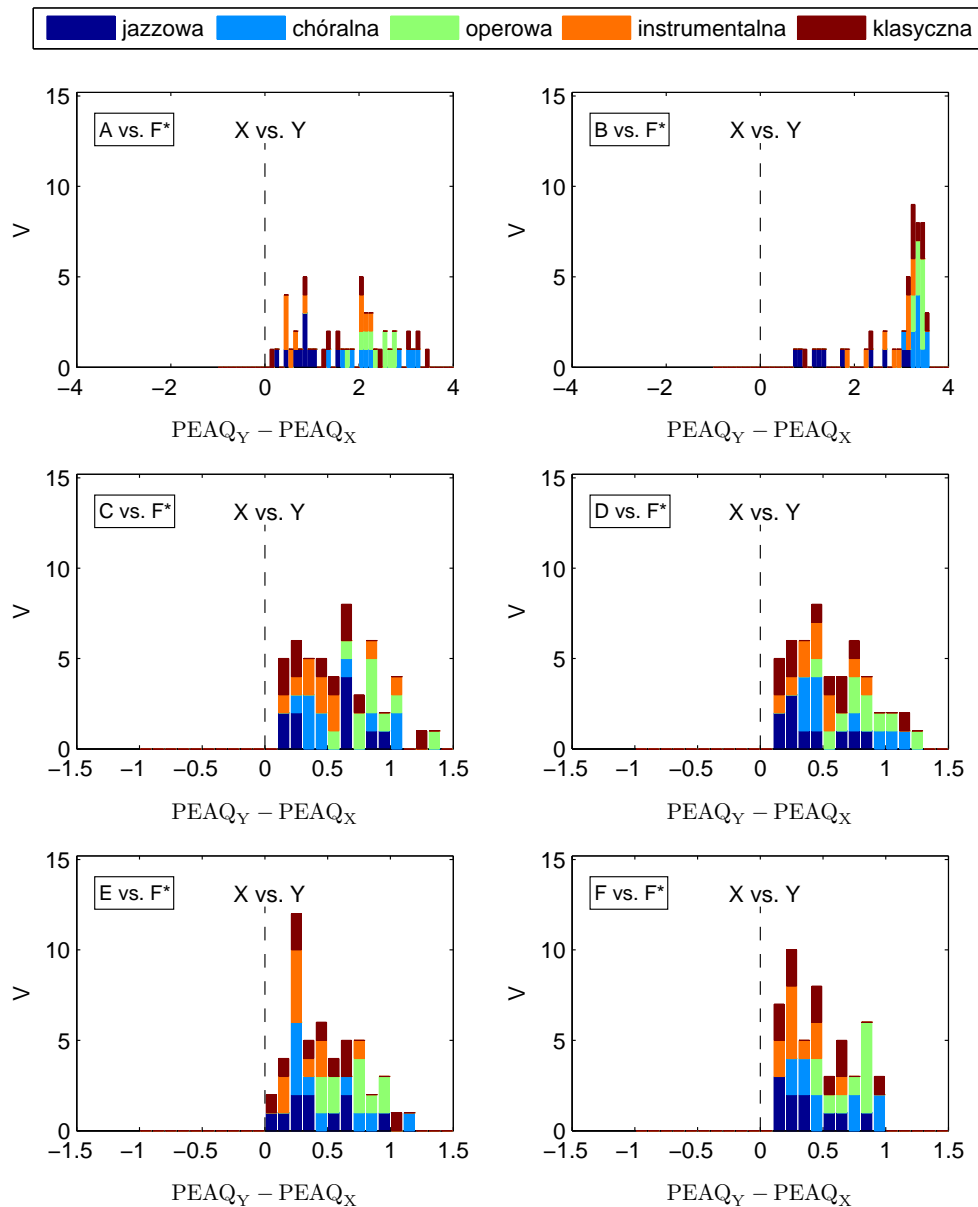


Rysunek 8.15: Ilościowe porównanie skuteczności działania testowanych podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie wyników PEAQ: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A/A*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B/B*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E*), podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F*). Dla każdej porównywanej pary X vs. Y zaznaczono punktami wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$. Wyniki dodatnie są na korzyść podejścia Y, a wyniki ujemne na korzyść podejścia X. Wartość średnia z uzyskanych 50 wyników została zaznaczona kwadratem.

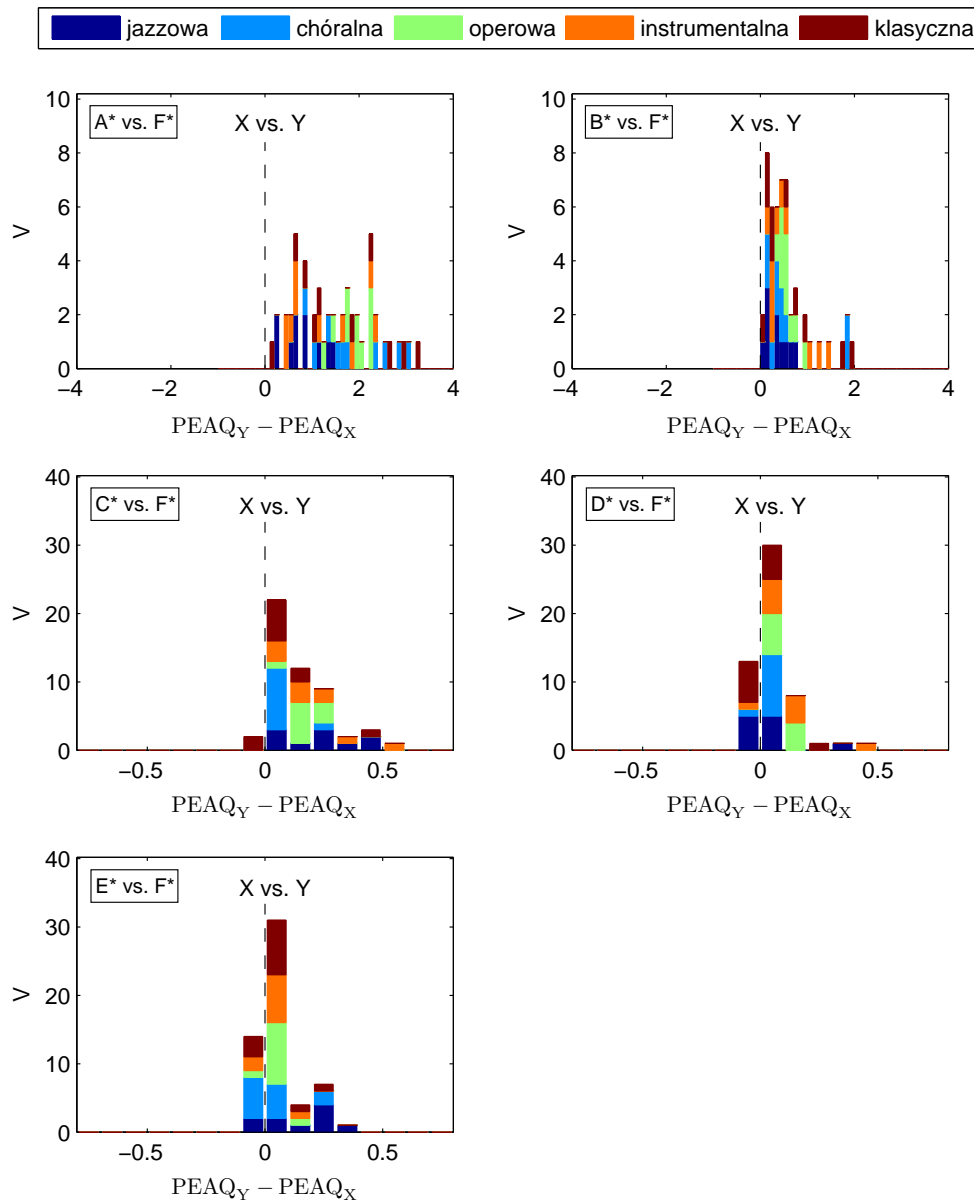
testowane podejścia vs. proponowane podejście (F^*)



Rysunek 8.16: Ilościowe porównanie skuteczności działania proponowanego podejścia w stosunku do pozostałych testowanych podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A/A^*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B/B^*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C^*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D^*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E^*), podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F^*). Dla każdej porównywanej pary X vs. Y zaznaczono punktami wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$. Wyniki dodatnie są na korzyść podejścia Y , a wyniki ujemne na korzyść podejścia X . Wartość średnia z uzyskanych 50 wyników została zaznaczona kwadratem.



Rysunek 8.17: Ilościowe porównanie skuteczności działania proponowanego podejścia (F^*) w stosunku do pozostałych testowanych podejść wykorzystujących analizę jednokierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E), podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F). Dla każdej porównywanej pary X vs. Y otrzymano wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$ w grupie 50 nagrań testowych: 10 z muzyką jazzową, 10 z muzyką choralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Wyniki dodatnie są na korzyść podejścia Y, a wyniki ujemne na korzyść podejścia X. Histogram pokazuje liczebność (V) uzyskanych wyników w przedziałach o równej szerokości wynoszącej 0,1. Przedziały dla wartości ujemnych są prawostronnie domknięte, zaś dla wartości dodatnich – lewostronnie domknięte. W przypadku, gdy wynik porównania wyniósł zero, oba sąsiadujące przedziały zwiększyły swoją liczebność. Poszczególne kolory na histogramie reprezentują wyniki pochodzące z danej grupy muzycznej.



Rysunek 8.18: Ilościowe porównanie skuteczności działania proponowanego podejścia (F^*) w stosunku do pozostałych testowanych podejść wykorzystujących analizę dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A^*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B^*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C^*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D^*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E^*), podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F^*). Dla każdej porównywanej pary X vs. Y otrzymano wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$ w grupie 50 nagrań testowych: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Wyniki dodatnie są na korzyść podejścia Y , a wyniki ujemne na korzyść podejścia X . Histogram pokazuje liczebność (V) uzyskanych wyników w przedziałach o równej szerokości wynoszącej 0,1. Przedziały dla wartości ujemnych są prawostronnie domknięte, zaś dla wartości dodatnich – lewostronnie domknięte. W przypadku, gdy wynik porównania wyniósł zero, oba sąsiadujące przedziały zwiększyły swoją liczebność. Poszczególne kolory na histogramie reprezentują wyniki pochodzące z danej grupy muzycznej.

Eksperyment 3

W poprzednim eksperymencie do oceny jakości eliminacji zakłóceń impulsowych wykorzystano spreparowane sygnały muzyczne. Wyniki przetworzenia spreparowanych nagrań mogły być ocenione przy użyciu obiektywnej miary PEAQ. Ponieważ w przypadku sygnałów mowy nie znaleziono obiektywnej miary, której wyniki korelowałyby się dobrze z subiektywnymi ocenami, przeprowadzono testy odsłuchowe.

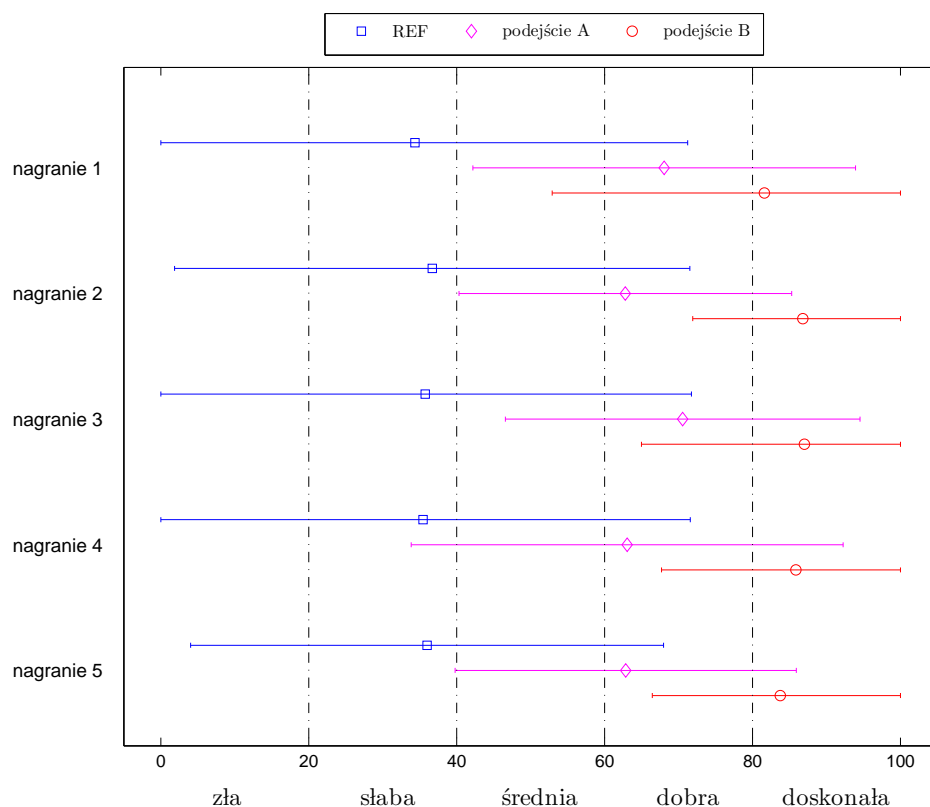
W teście odsłuchowym wykorzystano 5 czystych fragmentów słuchowiska radiowego pt. „Kochanek dr Luli” w reżyserii Jana Warenyci. Każdy fragment nagrania posiadał około 22 sekundy materiału fonicznego nagranych w systemie stereofonicznym i spróbkowanego z częstotliwością 48 kHz. Przeskalowane sygnały testowe spreparowano poprzez dodanie do lewego i prawego kanału szablonów zakłóceń utworzonych dla grupy treningowej – patrz rysunek 8.3. Następnie spreparowane nagrania przetworzono przy użyciu podejścia opartego na klasycznym modelu AR (podejście C) oraz przy użyciu proponowanego podejścia opartego na wektorowym modelu SVAR i wykorzystującego przetwarzanie dwukierunkowe (podejście F*).

Najpierw poproszono słuchaczy o ocenę jakości spreparowanych nagrań według przyjętej skali ocen rozciągającej się od 0 (zła) do 100 (doskonała) – patrz tabel 8.2. Następnie, korzystając z tej samej skali, słuchacze ocenili dwie propozycje przetworzenia spreparowanych nagrań pod kątem jakości eliminacji zakłóceń impulsowych. Wyniki porównania zamieszczono w tabeli 8.23. Pierwszy algorytm, oznaczony w tabeli literą A, uzyskał średnią ocenę jakości mieszczącą się w podprzedziale od 61 do 80 (jakość dobra). Z kolei proponowany algorytm, oznaczony w tabeli literą B, uzyskał we wszystkich przypadkach najlepszą średnią ocenę jakości mieszczącą się w najwyższym podprzedziale przyjętej skali, tj. od 81 do 100 (jakość doskonała). Słuchaczy w zdecydowanej większości preferowali nagrania przetworzone przez algorytm B. Podane w nawisach kwadratowych liczby pokazują ile razy oceniony algorytm otrzymał równą, bądź lepszą ocenę niż jego rywal (ponieważ niektóre wyniki były równe, liczby nie sumują się do 20).

Tablica 8.23: Porównanie wyników otrzymanych przy użyciu algorytmu opartego na klasycznym modelu AR (A) z tymi uzyskanymi przy użyciu polecanego w pracy algorytmu opartego na modelu SVAR (B). Wszystkie testy przeprowadzone zostały na fragmentach czystych nagrań słuchowiska radiowego spreparowanych poprzez dodanie rzeczywistych zakłóceń impulsowych. Test przeprowadzono w grupie 20 słuchaczy. Podane w pierwszym wierszu (REF) liczby odnoszą się do średniej oceny jakości spreparowanego nagrania według przyjętej skali ocen rozciągającej się od 0 (zła) do 100 (doskonała). Z kolei, podane w wierszach A i B liczby odnoszą się do średniej oceny jakości eliminacji zakłóceń impulsowych według tej samej skali. Podane w nawisach kwadratowych liczby pokazują ile razy oceniony algorytm otrzymał równą, bądź lepszą ocenę niż jego rywal (ponieważ niektóre wyniki były równe, liczby nie sumują się do 20). Najlepsze wyniki zostały pogrubione.

Skala ocen jakości: 0-20 = zła, 21-40 = słaba, 41-60 = średnia, 61-80 = dobra, 81-100 = doskonała.

Nagranie	słuchowisko radiowe				
	1	2	3	4	5
REF	34,4	36,7	35,8	35,5	36,0
A	68,1 [3]	62,8 [0]	70,6 [4]	63,1 [1]	62,9 [1]
B	81,6 [18]	86,8 [20]	87,0 [18]	85,9 [19]	83,8 [19]



Rysunek 8.19: Porównanie wyników otrzymanych przy użyciu algorytmu opartego na klasycznym modelu AR (A) z tymi uzyskanymi przy użyciu polecanego w pracy algorytmu finalnego opartego na modelu SVAR (B). Wszystkie testy przeprowadzone zostały na fragmentach czystych nagrań słuchowiska radiowego spreparowanych poprzez dodanie rzeczywistych zakłóceń impulsowych (REF). Test odsłuchowy przeprowadzono w grupie 20 słuchaczy. Zaznaczone przedziały reprezentują 95% przedziały ufności, gdy rozkład uzyskanych ocen przybliżony został rozkładem t Studenta. Symbolem kwadratu oznaczono wartość średnią oceny jakości spreparowanego nagrania (REF) według przyjętej skali ocen rozciągającej się od 0 (zła) do 100 (doskonała). Z kolei, symbolami romb i kółko oznaczono wartości średnie ocen jakości eliminacji zakłóceń impulsowych według tej samej skali, odpowiednio dla algorytmu A i B.

Na rysunku 8.19 pokazano 95% przedziały ufności w przypadku gdy rozkład uzyskanych ocen przybliżony jest rozkładem t Studenta. Średnie oceny jakości odpowiednio dla nagrania spreparowanego (REF) oraz dwóch wariantów jego przetworzenia oznaczono odpowiednio za pomocą kwadratu, rombu i kółka. Najszersze przedziały ufności uzyskano dla ocen jakości nagrania spreparowanego, zaś najwęższe – dla ocen jakości nagrań przetworzonych przez algorytm B. W żadnym z przypadków nie zaobserwowano, by odsetek ocen znajdujących się poza zaznaczonymi przedziałami przekroczył próg 5%.

Eksperyment 4

Ostatnie badanie miało na celu porównanie wyników otrzymanych przy użyciu proponowanego algorytmu, F* dla nagrań stereofonicznych oraz D* dla nagrań monofonicznych, z analogicznymi wynikami uzyskanymi przy użyciu komercyjnego systemu przeznaczanego do redukcji zakłóceń impulsowych. Do porównania wybrano system CEDAR – komercyjny pakiet do rekonstrukcji nagrań znany z bardzo wysokiej skuteczności w eliminowaniu zakłóceń impulsowych³.

Baza testowa składała się z 10 fragmentów archiwalnych nagrań gramofonowych (część z nich została odtworzona z płyty gramofonowej przy użyciu wkładki monofonicznej [3, 4, 6, 8, 9, 10], a część – przy użyciu wkładki stereofonicznej [1, 2, 5, 7]). Fragmenty nagrań zawierały od 13 do 27 sekund materiału fonicznego spróbkowanego z częstotliwością 44,1 kHz: 5 z nich o wysokim stopniu uszkodzenia (1, 2, 8, 9, 10) i 5 o średnim stopniu uszkodzenia (3, 4, 5, 6, 7). Nagrania te obejmowały różne style muzyki: klasyczna (6, 8), operowa (3, 4), popularna (1, 2, 5, 7) i bluesowa (9, 10).

Tablica 8.24: Porównanie wyników otrzymanych przy użyciu komercyjnego algorytmu the Auto Declick CEDAR (A) z tymi uzyskanymi przy użyciu polecanego w pracy algorytmu finalnego (B). Wszystkie testy przeprowadzone zostały na fragmentach rzeczywistych nagrań gramofonowych w grupie 20 słuchaczy. Podane w pierwszym wierszu liczby odnoszą się do średniej oceny jakości oryginalnego nagrania według przyjętej skali ocen rozciągającej się od 0 (zła) do 100 (doskonała). Z kolei, podane w wierszach A i B liczby odnoszą się do średniej oceny jakości eliminacji zakłóceń impulsowych według tej samej skali. Podane w nawisach kwadratowych liczby pokazują ile razy oceniony algorytm otrzymał równą, bądź lepszą ocenę niż jego rywal (ponieważ niektóre wyniki były równe, liczby nie sumują się do 20). Najlepsze wyniki zostały pogrubione.

Skala ocen jakości: 0-20 = zła, 21-40 = słaba, 41-60 = średnia, 61-80 = dobra, 81-100 = doskonała.

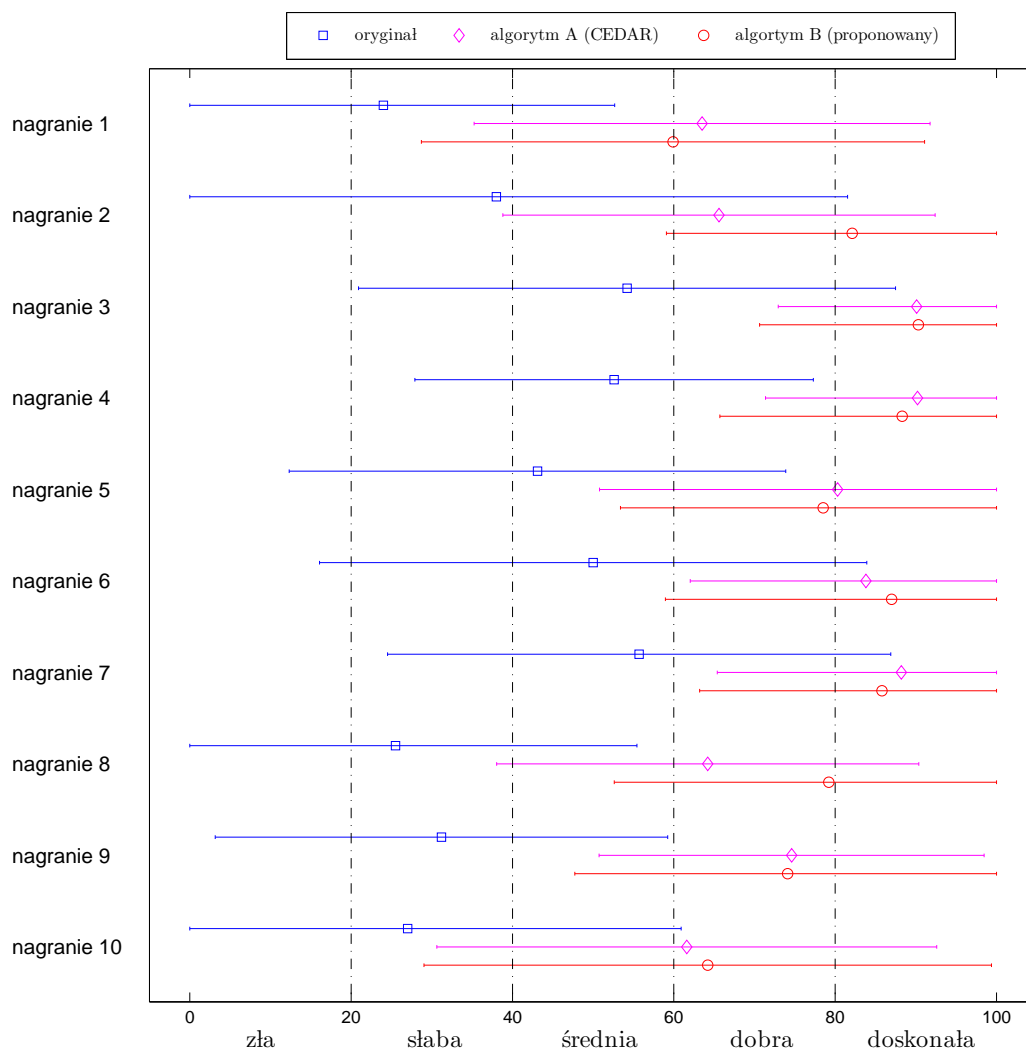
Nagranie archiwalne	1	2	3	4	5
oryginał	24,0	38,0	54,2	52,6	43,1
A	63,5 [15]	65,6 [1]	90,1 [15]	90,2 [16]	80,3 [17]
B	59,9 [6]	82,1 [19]	90,3 [14]	88,3 [13]	78,5 [10]

Nagranie archiwalne	6	7	8	9	10
oryginał	50,0	55,7	25,5	31,2	27,0
A	83,8 [9]	88,2 [18]	64,2 [2]	74,6 [13]	61,6 [12]
B	87,0 [17]	85,8 [13]	79,2 [18]	74,1 [13]	64,2 [14]

Wyniki porównania, pokazane w tabeli 8.24, są różne dla różnych nagrań, a przez to częściowo niejednoznaczne. W przypadku porównania średnich ocen, proponowany algorytm uzyskał lepszy wynik w pięciu nagraniach (2, 3, 6, 8, 10). W pozostałych pięciu nagraniach lepszy okazał się CEDAR. W przypadku bezpośredniego porównania ocen, w czterech przypadkach (2, 6, 8, 10)

³system CEDAR został opracowany na Uniwersytecie w Cambridge dla Brytyjskiej Narodowej Biblioteki Dźwięku. W pracy [59] przeprowadzono porównanie 8 komercyjnych produktów, spośród których CEDAR okazał się być najlepszym systemem do rekonstrukcji archiwalnych nagrań fonicznych. Narzędzie „Auto Declick” udostępnione przez CEDAR jest w pełni automatyczną procedurą, która nie wymaga od użytkownika wyboru żadnych nastaw. Żadne szczegóły dotyczące sposobu działania tego algorytmu nie zostały ujawnione.





Rysunek 8.20: Porównanie wyników otrzymanych przy użyciu komercyjnego algorytmu the Auto Declick CEDAR (A) z tymi uzyskanymi przy użyciu polecanego w pracy algorytmu finalnego (B). Wszystkie testy przeprowadzone zostały na fragmentach archiwalnych nagrań gramofonowych w grupie 20 słuchaczy. Zaznaczone przedziały reprezentują 95% przedziały ufności, gdy rozkład uzyskanych ocen przybliżony został rozkładem t Studenta. Symbolem kwadratu oznaczono wartość średnią oceny jakości oryginalnego nagrania według przyjętej skali ocen rozciągającej się od 0 (zła) do 100 (doskonała). Z kolei, symbolami romb i kółko oznaczono wartości średnie ocen jakości eliminacji zakłóceń impulsowych według tej samej skali, odpowiednio dla algorytmu A i B.

proponowany algorytm otrzymał więcej lepszych ocen niż CEDAR, w jednym przypadku (9) wynik porównania był identyczny, a w pozostałych pięciu przypadkach wynik porównania wypadł na korzyść CEDARA. W przypadku dwóch nagrań (2,8) proponowany algorytm uzyskał zdecydowaną przewagę. W przypadku pozostałych nagrań słuchacze podkreślili, że różnice były subtelne, ale słyszalne – raz na korzyść proponowanego podejścia raz na korzyść CEDARA.

Należy zwrócić uwagę na względnie dużą liczbę neutralnych decyzji (prawie 28%), tj. słuchacze przyznali aż 55 razy te same oceny obu algorytmom.

Na rysunku 8.20 pokazano 95% przedziały ufności wyznaczone dla przypadku gdy rozkład zebranych ocen przybliżono rozkładem t Studenta. Dla obu algorytmów uzyskane przedziały ufności są zbliżone. Wynika z nich, że słuchacze w większości przypadków oceniali jakość eliminacji zakłóceń jako dobrą lub doskonałą. W przypadku nagrań, których początkowa jakość oceniona została jako słaba, obserwujemy szersze przedziały ufności, zaś w przypadku nagrań o średniej jakości – uzyskane przedziały są węższe.

Podsumowanie

Przedstawione wyniki porównań jakościowych i ilościowych pokazują, że w przypadku przetwarzania sygnałów monofonicznych/stereofonicznych podejście oparte na modelu SAR/SVAR i wykorzystujące analizę dwukierunkową cechuje najlepsza skuteczność działania spośród porównywanych podejść. Proponowane podejścia sprawdzają się zarówno w przypadku analizy sygnałów muzycznych jak i sygnałów mowy. Testy odsłuchowe z wykorzystaniem rzeczywistych, archiwalnych nagrań pokazały, że proponowane podejścia skutecznie usuwają zakłócenia impulsowe, a uzyskiwana jakość zrekonstruowanych nagrań jest wysoka i porównywalna z tą uzyskiwaną przy użyciu komercyjnego systemu CEDAR.

8.2.4 Technika dopasowywania wzorców – wyniki badań

W tej części pracy przedstawiono wyniki symulacyjne, uzyskane zarówno dla sztucznie spreparowanych jak i rzeczywistych nagrań fonicznych, z wykorzystaniem algorytmów eliminacji zakłóceń impulsowych opisanych w rozdziale 3 oraz algorytmów wynikających z ich rozszerzeń wykorzystujących technikę dopasowywania wzorców zakłóceń, opisaną w rozdziale 7. W pierwszym eksperymencie zbadano skuteczność działania adaptacyjnego algorytmu interpolacji przy założeniu, że położenie zakłóceń impulsowych jest znane. W drugim eksperymencie oceniono skuteczność działania algorytmów na spreparowanych sygnałach fonicznych w przypadku gdy interpolacja połączona została z adaptacyjnymi detektorami zakłóceń opartymi na modelu AR. W ostatnim eksperymencie oceniono skuteczność działania proponowanej techniki na rzeczywistych nagraniach gramofonowych.

Spreparowane nagrania

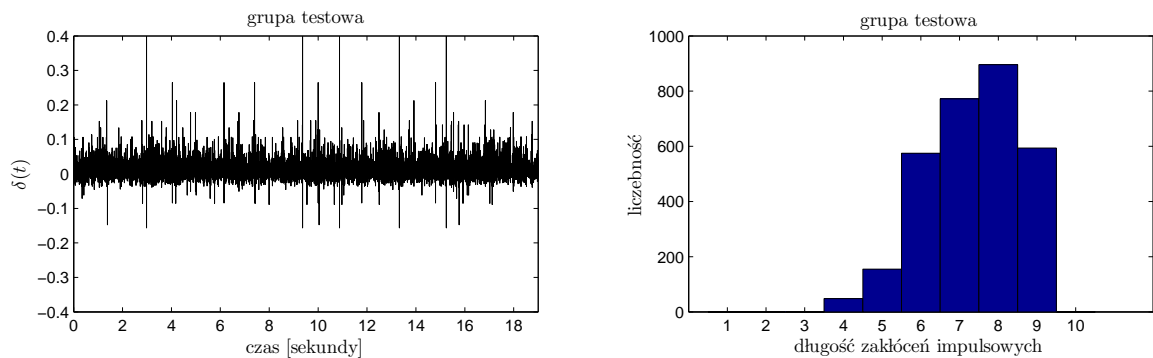
Zbiór przykładowych zakłóceń impulsowych, wyciętych z fragmentu ciszy pewnego archiwalnego nagrania gramofonowego (przy częstotliwości próbkowania 48 kHz), obejmował 1000 przebiegów. Zbiór ten został podzielony w sposób losowy na dwa mniejsze podzbiory \mathcal{P}_C i \mathcal{P}_D zawierające po $n_t = 500$ przykładowych zakłóceń impulsowych, które użyte zostały odpowiednio w fazie treningowej i fazie testowej.

Na podstawie zbioru treningowego uzyskano 14 wzorców zakłóceń impulsowych ($\kappa = 0,95$), pokazanych na rysunku 7.1(a). Czas potrzebny na realizację tego zadania na standardowym komputerze PC wyniósł kilka sekund. Na każdym wykresie tego rysunku podana została informacja o długości wzorca zakłóceń (m_i) oraz o rozmiarze kliku (n_i), którą reprezentuje ten wzorec. Zauważmy, że $\sum_{i=1}^{14} n_i = 189$, co oznacza, że aż 38% przykładowych przebiegów ze zbioru \mathcal{P}_C posłużyło do formowania wzorców zakłóceń.

Baza testowa składała się z 60 czystych sygnałów monofonicznych. Materiał muzyczny wyselekcjonowany został w taki sposób, aby reprezentował sygnały foniczne o różnych charakterystykach czasowych i częstotliwościowych. W tym celu wybranych zostało 29 fragmentów muzyki jazzowej (15 wokalnych, 14 instrumentalnych), 23 fragmenty muzyki klasycznej, 3 fragmenty muzyki chóralnej oraz 5 fragmentów muzyki operowej. Nagrania te trwały od 22 do 30 sekund i zostały spróbkowane z częstotliwością 48 kHz. Zanim dodano zakłócenia impulsowe, wszystkie sygnały foniczne zostały przeskalowane tak aby ich średnia moc była jednakowa. Przebiegi zakłóceń impulsowych zostały wybrane w sposób losowy ze zbioru \mathcal{P}_D , a następnie dodane do sygnałów w równych odstępach co 300 próbek. Liczba zakłóceń impulsowych dodanych do każdego nagrania testowego równa była $N_0 = 3040$. W ten sposób degradacji uległy fragmenty sygnałów testowych trwające po 19 sekund. Zdegradowane fragmenty zaczynały się po czasie 1,5 sekundy od początku każdego nagrania. Dodane zakłócenia obejmowały w sumie 22335 próbek, co stanowi 2,45% długości zdegradowanego fragmentu sygnału. Na rysunku 8.21 pokazano użyty szablon zakłóceń impulsowych oraz histogram wylosowanych zakłóceń impulsowych.

Porównywane podejścia

Przeprowadzony test sprawdzał skuteczność działania 4 podejść do eliminacji zakłóceń impulsowych: podejścia opartego na detekcji w pętli otwartej (A), podejścia wynikającego z połączenia podejścia A z proponowaną techniką dopasowywania wzorców zakłóceń (A^*), podejścia opartego na detekcji w pętli z decyzyjnym sprzężeniem zwrotnym (B) oraz podejścia wynikającego z połączenia podejścia B z proponowaną techniką dopasowywania wzorców zakłóceń (B^*).



Rysunek 8.21: Po lewej, szablony zakłóceń impulsowych wykorzystane do utworzenia spreparowanych nagrań testowych. Po prawej, histogram przedstawiający rozkład długości wylosowanych zakłóceń impulsowych z przedziału od 1 do 10 próbek.

Parametry symulacji

Wszystkie porównywane algorytmy detekcji i rekonstrukcji zakładały, że analizowany sygnał foniczny opisany jest za pomocą modelu AR rzędu $r = 20$. Identyfikacja sygnału prowadzona była na bieżąco przy użyciu algorytmu EWLS (3.39) wyposażonego w stałą zapomnienia $\lambda = 0,995$. Estymator wariancji białego szumu tworzącego (3.61) wyposażony był w stałą zapomnienia $\lambda_0 = 0,991$. Nastawy parametrów projektowych przyjmowały następujące wartości: mnożnik średniego odchylenia standardowego $\mu = 4,5$, próg weryfikacji podobieństwa $\gamma_0 = 0,8$, parametr określający liczbę wcześniejszych błędów predykcji uwzględnianych przy dopasowywaniu wzorców zakłóceń $k_b = 4$, maksymalna długość alarmu detekcyjnego $m_{\max} = 100$ oraz parametry związane z rozszerzaniem alarmów detekcyjnych $\Delta_1 = 4$ i $\Delta_2 = 1$.

Eksperyment 1

Zanim dokonano porównania skuteczności adaptacyjnych algorytmów detekcji, najpierw sprawdzono skuteczność adaptacyjnego algorytmu interpolacji wspieranego przez informacje o dokładnym położeniu wstawionych zakłóceń impulsowych [$\hat{d}(t) \equiv d(t)$]. Otrzymane w ten sposób wyniki rekonstrukcji wszystkich sygnałów testowych oceniono przy użyciu obiektywnej miary PEAQ – patrz kolumna GT w tabeli 8.25, a następnie w testach odsłuchowych. Celem tego eksperymentu było sprawdzenie w jakim stopniu interpolacja sygnału wpływa na pogorszenie jakości dźwięku w zrekonstruowanych sygnałach testowych. Ponieważ w żadnym z przypadków testy odsłuchowe nie wykazały słyszalnych różnic między oryginalnym materiałem fonicznym, a jego rekonstrukcją (towarzyszące temu wyniki PEAQ są bliskie wartości zerowej), potwierdzono w ten sposób fakt, że wszystkie słyszalne zniekształcenia (o ile występują) obserwowane po połączeniu procedury adaptacyjnej interpolacji z procedurą adaptacyjnej detekcji, są wyłącznie konsekwencją błędów detekcji, takich jak niewykryte zakłócenia, niedokładnie wykryte zakłócenia lub fałszywe alarmy detekcyjne.

Eksperyment 2

W tym eksperymencie porównano skuteczność działania 4 podejść do eliminacji zakłóceń impulsowych na podstawie uzyskanych wyników PEAQ. Wszystkie wyniki zamieszczone zostały

Tablica 8.25: Porównanie skuteczności działania 4 podejść na podstawie wyników PEAQ: podejście oparte na detekcji w pętli otwartej (A), podejście oparte na detekcji w pętli otwartej połączone z techniką dopasowywania wzorców (A*), podejście oparte na detekcji w pętli z decyzyjnym sprzężeniem zwrotnym (B) oraz podejście oparte na detekcji w pętli z decyzyjnym sprzężeniem zwrotnym połączone z techniką dopasowywania wzorców (B*). Dodatkowo, tabela pokazuje wyniki otrzymane dla spreparowanych nagrań fonicznych (REF) oraz wyniki odniesienia (GT) otrzymane gdy sygnał jest rekonstruowany przy pełnej wiedzy o lokalizacji zakłóceń impulsowych. Śr.₁₀ i Śr.₆₀ oznaczają uśrednione wyniki w grupie 10 wyszczególnionych nagrań i w grupie wszystkich 60 nagrań.

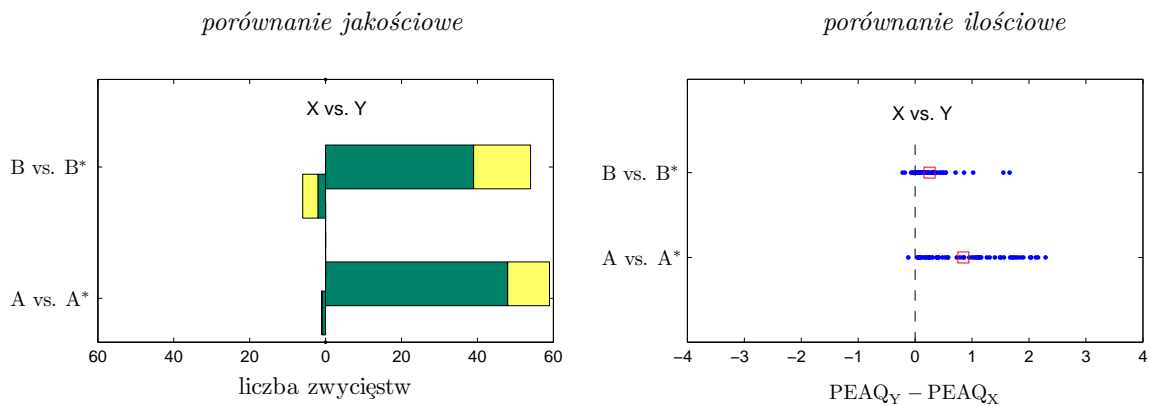
Interpretacja wyników PEAQ: 0 = niesłyszalne (znieszczenia sygnału), -1 = zauważalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

nagranie	GT	REF	A	A*	B	B*
1	-0,03	-3,68	-3,22	-1,10	-0,20	-0,16
2	-0,06	-3,69	-3,34	-1,05	-0,16	-0,14
3	-0,05	-3,70	-3,70	-2,97	-0,63	-0,47
4	-0,01	-3,59	-3,43	-1,39	-0,32	-0,28
5	-0,09	-3,77	-3,69	-2,60	-0,36	-0,20
6	-0,06	-3,84	-3,65	-3,48	-1,48	-1,05
7	-0,02	-3,82	-3,45	-2,14	-0,80	-0,49
8	-0,03	-3,55	-3,14	-2,76	-1,78	-0,92
9	-0,02	-3,13	-3,11	-0,95	-0,46	-0,37
10	-0,01	-3,01	-2,86	-1,30	-0,92	-0,68
Śr. ₁₀	-0,04	-3,58	-3,36	-1,97	-0,71	-0,47
Śr. ₆₀	-0,05	-3,75	-3,53	-2,69	-0,87	-0,62

w tabeli C.1 w dodatku C. W tabeli 8.25 wyszczególniono jedynie wyniki dla 10 pierwszych nagrań, podano również wyniki średnie dla tych 10 nagrań oraz dla wszystkich 60 nagrań. W tabeli podano również wyniki PEAQ dla nieprzetworzonych spreparowanych nagrań, kolumna REF, oraz wyniki PEAQ dla nagrań zrekonstruowanych przy pełnej wiedzy o położeniu zakłóceń impulsowych, kolumna GT. Porównywane algorytmy uszeregowano według rosnących wyników PEAQ. Najgorsze wyniki PEAQ otrzymano dla algorytmu A, a najlepsze dla algorytmu B*. Zauważmy, że proponowana technika dopasowywania wzorców zakłóceń poprawia skuteczność metod, z którymi jest łączona. W przypadku algorytmu A poprawa jest duża, z kolei w przypadku algorytmu B poprawa jest znacznie mniejsza, ale systematyczna i słyszalna.

Wyniki PEAQ z tabeli C.1 porównano jakościowo oraz ilościowo. Porównanie to zilustrowano na rysunku 8.22. W przypadku porównania jakościowego żółte słupki pokazują ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu X vs. Y . Z kolei zielone słupki pokazują liczbę zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść. Podejścia A* oraz B* wypadły najlepiej w swoich porównaniach uzyskując w większości przypadków znacząco lepsze wyniki PEAQ. W przypadku porównania ilościowego dla każdej porównywanej pary X vs. Y zaznaczono punktami wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$. Wyniki dodatnie są na korzyść podejścia Y , a wyniki ujemne na korzyść podejścia X . Wartość średnia z uzyskanych 60 wyników została zaznaczona kwadratem. Należy zwrócić uwagę, że w przypadku porównania A vs. A* różnice w wynikach PEAQ skupiły się w przedziale od -0,12 do aż 2,29. Wartość średnia dla tego porównania wyniosła 0,85. W przypadku porównania B vs. B* wyniki mieściły się w przedziale od -0,22 do 1,66, gdzie wartość średnia wyniosła 0,25. Wyniki te potwierdzają, że technika dopasowywania wzorców





Rysunek 8.22: Jakościowe i ilościowe porównanie 4 algorytmów detekcji: podejście oparte na detekcji w pętli otwartej (A), podejście oparte na detekcji w pętli otwartej połączone z techniką dopasowywania wzorców (A*), podejście oparte na detekcji w pętli z decyzyjnym sprzężeniem zwrotnym (B) oraz podejście oparte na detekcji w pętli z decyzyjnym sprzężeniem zwrotnym połączone z techniką dopasowywania wzorców (B*). Wszystkie wyniki uzyskano dla 60 spreparowanych nagrań testowych. W przypadku porównania jakościowego żółte słupki pokazują ile razy dane podejście otrzymało lepszą bądź równą ocenę PEAQ w bezpośrednim porównaniu X vs. Y. Zielone słupki pokazują liczbę zdecydowanie lepszych ocen na korzyść jednego z porównywanych podejść. W przypadku porównania ilościowego dla każdej porównywanej pary X vs. Y zaznaczono punktami wyniki będące różnicą ich ocen PEAQ: $PEAQ_Y - PEAQ_X$. Wyniki dodatnie są na korzyść podejścia Y, a wyniki ujemne na korzyść podejścia X. Wartość średnia z uzyskanych 60 wyników została zaznaczona kwadratem.

pozwała na dokładniejsze oznaczanie zakłóceń impulsowych niż tradycyjne algorytmy detekcji, co bezpośrednio wpływa na jakość przetworzonych nagrań.

Skuteczność detekcji oceniono także przy użyciu dwóch miar pośrednich: wskaźnika dopasowania oraz wskaźnika pokrycia energetycznego. Wyniki zamieszczone w tabeli 8.26 pozwalają na dokonanie ciekawych spostrzeżeń dotyczących porównywanych schematów detekcji. Na początek zauważmy, że w obu przypadkach rozszerzenie algorytmów A i B o proponowaną metodę dopasowywania wzorców zakłóceń poprawia obie statystyki. Dowodzi to, że proponowany schemat wykrywania zakłóceń impulsowych gwarantuje lepsze dopasowanie alarmów detekcyjnych, tj. lepsze oznaczenie początku i końca zakłócenia impulsowego. Zauważmy również, że największa poprawa obserwowana dla obu wskaźników ma miejsce dla podejścia A*. Klasyczne podejście A ma tendencje do podnoszenia krótkich alarmów detekcyjnych, gdzie początek zakłócenia jest wykrywany dokładniej niż jego koniec. W przypadku podejścia B* zachodzi mniejsza poprawa, ale jest ona również znacząca. Na koniec należy podkreślić, że jednoczesne zwiększenie obu wskaźników może być osiągnięte tylko dzięki dokładniejszej lokalizacji zakłóceń impulsowych.

Tabela 8.27 podsumowuje statystyki dla polecanego schematu detekcji B*. W tabeli podano całkowitą liczbę alarmów detekcyjnych (N_d) oraz liczbę alarmów, w przypadku których stwierdzono podobieństwo do jednego ze wzorców (N_m). Dodatkowo, pokazuje ona liczbę poprawnych wykryć (N_{dc} , N_{mc}), liczbę fałszywych wykryć (N_{di} , N_{mi}) oraz dwie miary wskaźnika sukcesu (N_{mc}/N_0 , N_{mc}/N_{dc}). Zgodnie z tabelą 8.27, dla wyróżnionych 10 nagrań procentowy wynik poprawnie wykrytych zakłóceń impulsowych o typowych kształtach N_{mc}/N_0 mieści się w przedziale od 81% do 93%. Ponieważ procedura dopasowywania wzorców zakłóceń uruchamiana jest tylko wtedy gdy alarm detekcyjny zostanie podniesiony, statystyka ta uwzględnia „straty” wpro-

Tablica 8.26: Porównanie skuteczności działania 4 algorytmów detekcji na podstawie dwóch pośrednich miar (wskaźnika podobieństwa oraz wskaźnika pokrycia energetycznego): podejście oparte na detekcji w pętli otwartej (A), podejście oparte na detekcji w pętli otwartej połączone z techniką dopasowywania wzorców (A*), podejście oparte na detekcji w pętli z decyzyjnym sprzężeniem zwrotnym (B) oraz podejście oparte na detekcji w pętli z decyzyjnym sprzężeniem zwrotnym połączone z techniką dopasowywania wzorców (B*). $\bar{S}_{r.10}$ i $\bar{S}_{r.60}$ oznaczają uśrednione wyniki w grupie 10 wyszczególnionych nagrań i w grupie wszystkich 60 nagrań.

wskaźnik podobieństwa [%]

nagranie	A	A*	B	B*
1	31,86	52,55	51,46	56,59
2	32,08	53,76	54,33	58,34
3	38,96	53,12	57,81	56,57
4	33,59	51,17	50,98	54,54
5	47,41	54,80	65,13	58,67
6	29,79	42,62	36,26	47,50
7	30,21	41,95	38,45	46,21
8	23,27	31,27	29,82	37,23
9	27,85	43,68	44,72	49,84
10	22,46	29,60	28,16	31,91
$\bar{S}_{r.10}$	31,75	45,45	45,71	49,74
$\bar{S}_{r.60}$	34,67	48,61	43,06	52,60

wskaźnik pokrycia energetycznego [%]

nagranie	A	A*	B	B*
1	88,07	97,58	93,55	98,82
2	88,84	98,19	95,07	99,52
3	92,14	98,56	96,43	99,48
4	86,09	97,15	91,81	98,34
5	95,22	99,34	98,53	99,93
6	95,08	98,82	99,91	99,84
7	93,76	97,81	98,61	99,26
8	92,32	96,43	97,69	98,88
9	87,78	97,20	98,73	98,88
10	86,45	92,98	96,86	96,96
$\bar{S}_{r.10}$	90,58	97,40	96,72	98,99
$\bar{S}_{r.60}$	94,49	98,06	98,57	99,05

dzane przez klasyczny detektor zakłóceń impulsowych. Bardziej miarodajny wskaźnik miary sukcesu, zdefiniowany jako N_{mc}/N_{dc} , mieści się w przedziale od 89% do 97%.

Tabela 8.28 przedstawia porównanie wyników, uśrednionych w grupie 60 nagrań, dla polecanej metody B* z uwzględnieniem rosnącej liczby wzorców zakłóceń L_w . Wyniki podsumowują, przy użyciu miary bezpośredniej (PEAQ) oraz miar pośrednich (wskaźniki statystyczne), skuteczność metody w zależności od liczby uwzględnianych wzorców. Zgodnie z oczekiwaniem,

Tablica 8.27: Statystyki otrzymane dla proponowanego schematu detekcji B*: całkowita liczba zgłoszonych alarmów detekcyjnych (N_d) oraz liczba alarmów, w przypadku których stwierdzono podobieństwo do jednego ze wzorców (N_m). Całkowita liczba dodanych zakłóceń impulsowych była równa $N_0 = 3040$. Dodatkowo, tabela pokazuje liczbę poprawnych wykryć (N_{dc} , N_{mc}), liczbę fałszywych wykryć (N_{di} , N_{mi}) oraz dwie miary wskaźnika sukcesu (N_{mc}/N_0 , N_{mc}/N_{dc}). Śr.₁₀ i Śr.₆₀ oznaczają uśrednione wyniki otrzymane odpowiednio dla prezentowanych 10 nagrań oraz dla wszystkich 60 nagrań. Uśrednione statystyki (oprócz dwóch ostatnich) zaokrąglone zostały do najbliższej liczby całkowitej.

nagranie	N_d	N_{dc}	N_{di}	N_m	N_{mc}	N_{mi}	N_{mc}/N_0	N_{mc}/N_{dc}
1	2877	2839	38	2575	2551	24	0,84	0,90
2	3011	2965	46	2816	2779	37	0,91	0,94
3	3037	2933	104	2702	2625	77	0,86	0,89
4	2779	2754	25	2465	2451	14	0,81	0,89
5	3068	3027	41	2776	2749	27	0,90	0,91
6	3732	3030	702	3291	2736	555	0,90	0,90
7	3632	2942	690	3260	2733	527	0,90	0,93
8	4361	2887	1474	3832	2654	1178	0,87	0,92
9	3349	2923	426	3174	2831	343	0,93	0,97
10	4553	2727	1826	3972	2513	1459	0,83	0,92
Śr. ₁₀	3440	2903	537	3086	2662	424	0,88	0,92
Śr. ₆₀	3212	2921	290	2820	2634	186	0,87	0,90

Tablica 8.28: Porównanie wyników dla polecanej metody B* z uwzględnieniem rosnącej liczby wykorzystanych wzorców zakłóceń L_w . Podane statystyki detekcji oznaczają: liczba alarmów detekcyjnych, w przypadku których stwierdzono podobieństwo do jednego ze wzorców (N_m), liczba poprawnych wykryć (N_{mc}) oraz liczba fałszywych wykryć (N_{mi}), $N_m = N_{mc} + N_{mi}$. Całkowita liczba dodanych zakłóceń impulsowych wyniosła 3040. Wyniki uśredniono w grupie 60 nagrań i zaokrąglono do najbliższej liczby całkowitej.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = zauważalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

L_w	N_m	N_{mc}	N_{mi}	PEAQ
0	0	0	0	-0,87
1	915	914	1	-0,78
2	1818	1672	146	-0,69
3	2060	1916	144	-0,69
4	2143	2000	143	-0,66
5	2354	2210	145	-0,67
6	2427	2283	144	-0,65
7	2455	2308	147	-0,66
8	2481	2335	146	-0,65
9	2521	2376	145	-0,65
10	2556	2398	157	-0,66
11	2720	2561	158	-0,63
12	2735	2576	158	-0,62
13	2799	2624	174	-0,61
14	2820	2634	186	-0,62

Tablica 8.29: Porównanie wyników PEAQ dla dwóch wariantów postępowania: metoda dopasowywania wzorców z następującą po niej kompensacją zakłóceń (B^-) oraz metoda dopasowywania wzorców z następującą po niej interpolacją sygnału (B^*). $\bar{S}r_{.10}$ i $\bar{S}r_{.60}$ oznaczają uśrednione wyniki otrzymane odpowiednio dla prezentowanych 10 nagrań oraz dla wszystkich 60 nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = zauważalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

nagranie	B^-	B^*
1	-3,47	-0,16
2	-3,56	-0,14
3	-3,68	-0,47
4	-3,51	-0,28
5	-3,70	-0,20
6	-3,83	-1,05
7	-3,78	-0,49
8	-3,65	-0,92
9	-2,98	-0,37
10	-3,00	-0,68
$\bar{S}r_{.10}$	-3,52	-0,47
$\bar{S}r_{.60}$	-3,70	-0,62

stopień poprawy mierzony wynikami PEAQ maleje wraz z rosnącą liczbą branych pod uwagę wzorców. W rozważanym przypadku okazuje się, że dla $L_w = 6$ otrzymujemy już w pełni zadowalające wyniki. Ważne jest też to, że wraz z dalszym zwiększaniem liczby wzorców, wyniki PEAQ (prawie) systematycznie rosną (są bliższe wartości zerowej), co oznacza, że nadmierna liczba użytych wzorców nie wprowadza kary w postaci obniżenia skuteczności proponowanej metody. W rozważanym przypadku, gdy L_w zostaje zwiększone z 6 do 14, można osiągnąć jeszcze małą poprawę.

Jak już wspomniano w podrozdziale 7.3, gdy wykryte zakłócenie impulsowe ma typowy kształt pasujący do jednego ze wzorców, uszkodzony fragment sygnału może zostać zrekonstruowany na dwa sposoby – przy użyciu interpolacji sygnału lub w wyniku zastosowania techniki kompensacji polegającej na odjęciu od zakłóconego sygnału odpowiednio skorygowanego wzorca zakłócenia impulsowego. Tabela 8.29 przedstawia porównanie wyników PEAQ dla dwóch wspomnianych wyżej wariantów postępowania. Zgodnie z tymi wynikami, interpolacja daje znacznie lepsze wyniki niż kompensacja – spostrzeżenie to potwierdzone zostało również w testach odsłuchowych.

Eksperyment 3

Ostatni test przeprowadzony został na 8 rzeczywistych nagraniach gramofonowych, spróbkowanych z częstotliwością 48 kHz i zawierających od 24 do 31 sekund materiału fonicznego. Nagrania testowe pochodzą z dwóch różnych źródeł: albumu zawierającego największe arie operowe w wykonaniu Maria del Monaco (cztery nagrania, 58 wzorców), oraz albumu będącego kompilacją bluesowych utworów Mississippi Delta Blues, w wykonaniu 4 różnych artystów (4 nagrania, 24 wzorce). W ramach testu porównano ze sobą dwa podejścia: podejście oparte na detekcji z de-



Tablica 8.30: Wyniki otrzymane dla dwóch porównywanych algorytmów eliminujących zakłócenia impulsowe, opartych na detekcji z decyzyjnym sprzężeniem zwrotnym (B) oraz detekcji z decyzyjnym sprzężeniem zwrotnym połączonej z techniką dopasowywaniem wzorców (B*). Wszystkie testy przeprowadzone zostały na fragmentach rzeczywisty nagrań gramofonowych w grupie 20 słuchaczy. Podane w każdym wierszu liczby odnoszą się do średniej oceny jakości eliminacji zakłóceń impulsowych według przyjętej skali ocen rozciągającej się od 0 (zła) do 100 (doskonała). W przypadku nagrania oryginalnego podano średnią ocenę jakości nagrania. Podane w nawisach kwadratowych liczby pokazują ile razy ocenione nagranie otrzymało równą, bądź lepszą ocenę niż jego rywal (ponieważ niektóre wyniki były równe, liczby nie sumują się do 20). Najlepsze wyniki zostały pogrubione.

nagranie archiwalne	blues			
	1	2	3	4
oryginał	29,3 [0]	32,4 [0]	26,0 [0]	26,7 [0]
B	65,6 [10]	66,0 [10]	60,1 [10]	66,2 [4]
B*	65,7 [13]	68,6 [15]	61,3 [12]	73,0 [17]

nagranie archiwalne	aria			
	1	2	3	4
oryginał	41,5 [0]	40,2 [0]	37,3 [0]	37,0 [0]
B	78,2 [9]	75,2 [4]	76,5 [10]	70,0 [7]
B*	78,3 [12]	79,9 [17]	80,2 [14]	77,4 [17]

Tablica 8.31: Statystyki wykrywania zakłóceń impulsowych otrzymane dla algorytmu B* zastosowanego do archiwalnego nagrania fonicznego: liczba alarmów detekcyjnych (N_d), liczba alarmów, w przypadku których stwierdzono podobieństwo do jednego ze wzorców (N_m) oraz miara wskaźnika sukcesu (N_m/N_d).

nagranie archiwalne	N_d	N_m	N_m/N_d
blues 1	1151	831	0,72
blues 2	966	751	0,78
blues 3	2931	1948	0,66
blues 4	1002	453	0,45
aria 1	888	811	0,91
aria 2	842	721	0,86
aria 3	1737	1493	0,86
aria 4	2680	2344	0,87

cyzyjnym sprzężeniem zwrotnym (B) oraz podejście oparte na detekcji z decyzyjnym sprzężeniem zwrotnym połączone z metodą dopasowywania wzorców (B*). Wnioski z testów odsłuchowych, podsumowanych w tabeli 8.30, są podobne do tych wyciągniętych dla spreparowanych nagrań: większość słuchaczy preferowała wyniki otrzymane przy użyciu proponowanego algorytmu B*.

W tabeli 8.31 przedstawiono statystyki dotyczące skuteczności detekcji zakłóceń impulsowych otrzymane dla algorytmu B*. Zwróćmy uwagę na wysokie wartości wskaźnika sukcesu N_m/N_d . Oznacza to, że w prawie wszystkich przypadkach większość alarmów detekcyjnych obsłużona została przez technikę dopasowywania wzorców. Potwierdza to fakt, że zakłócenia impulsowe występujące w archiwalnych nagraniach gramofonowych mają podobny kształt.

Podsumowanie

Wyniki przeprowadzonych eksperymentów potwierdzają, że skuteczność działania tradycyjnych detektorów zakłóceń impulsowych, takich jak detektor pracujący w pętli otwartej czy detektor pracujący w pętli z decyzyjnym sprzężeniem zwrotnym, jest zwiększana gdy działanie detektorów połączone zostanie z techniką dopasowywania wzorców. W efekcie, zakłócenia impulsowe o typowym kształcie są dokładniej wykrywane, a jakość zrekonstruowanych nagrań staje się lepsza. Uzyskiwana poprawa jakości jest systematyczna i słyszalna, co potwierdzają wyniki testów odsłuchowych. Dalsze badania nad tą techniką pokazały, że gdy testowane detektory wykorzystują analizę dwukierunkową technika dopasowywania wzorców nie przynosi już dodatkowych korzyści, dlatego polecana jest tylko do przetwarzania sygnałów fonicznych w trybie „on-line”.

Rozdział 9

Zakończenie

W niniejszej rozprawie zaprezentowano skuteczną, uniwersalną metodę eliminacji zakłóceń impulsowych z archiwalnych nagrań dźwiękowych odtworzonych z płyt gramofonowych.

Przedstawione w rozprawie rozwiązania, takie jak analiza dwukierunkowa (rozdział 4), dwuzakresowe modelowanie autoregresyjne realizowane w postaci kaskady dwóch filtrów (rozdział 5) oraz modelowanie wektorowe (rozdział 6) umożliwiają przetwarzanie zarówno sygnałów muzycznych jak sygnałów mowy przy zachowaniu wysokiej jakości dźwięku po obróbce. W przypadku nagrań zarejestrowanych w wersji stereofonicznej łączna analiza lewego i prawego kanału stereofonicznego pokazała, że można uzyskać znacząco lepsze rezultaty niż te uzyskiwane gdy lewy i prawy kanał analizowane są oddzielnie. Testy odsłuchowe przeprowadzone w grupie 20 słuchaczy potwierdziły, że opracowany algorytm poprawia subiektywnie odczuwalną jakość przetworzonego sygnału. Opracowana metoda pozwala na całkowite lub prawie całkowite wyeliminowanie słyszalnych zakłóceń impulsowych w sposób nie powodujący nowych, niepożądanych efektów dźwiękowych. Otrzymywana jakość zrekonstruowanych nagrań fonicznych jest porównywalna, a w niektórych przypadkach nawet znacząco lepsza od tej uzyskiwanej przez komercyjny system CEDAR uznawany powszechnie za najlepszy produkt komercyjny służący do rekonstrukcji archiwalnych nagrań dźwiękowych.

Przedstawiona w pracy metoda wraz z wynikami eksperymentów (rozdział 8) potwierdzają jednoznacznie tezę rozprawy dowodząc, że możliwe jest skuteczne eliminowanie zakłóceń impulsowych, a w rezultacie uzyskiwanie wysokiej jakości zrekonstruowanego sygnału fonicznego.

W rozprawie zaprezentowano również podejście oparte na wykorzystaniu wiedzy *a priori* o typowych kształtach i rozmiarach zakłóceń impulsowych do dokładniejszego wykrywania położenia zakłóceń w sygnale (rozdział 7). Otrzymane wyniki testów symulacyjnych oraz testów odsłuchowych potwierdzają, że podejście to jest użyteczne w przypadku analizy sygnałów dźwiękowych „w czasie rzeczywistym”.

Możliwe kierunki rozwoju

W rozdziale 4 zaproponowano dwukierunkową analizę sygnału fonicznego z wykorzystaniem lokalnych reguł decyzyjnych. Metoda ta uwzględnia wstępne poszerzanie alarmów detekcyjnych o stałą, niewielką liczbę próbek. W rozdziale 8 pokazano, że zabieg ten ma istotne znaczenie na jakość uzyskiwanych rezultatów. W przyszłości można byłoby zaproponować adaptacyjną procedurę poszerzania alarmów detekcyjnych na podstawie analizy jednokrokowych błędów predykcji.

W rozdziale 5 zaproponowano schemat detekcji zakłóceń impulsowych oparty na współpracy trzech detektorów, który cechuje się zwiększoną odpornością na podnoszenie fałszywych alar-



mów. W przyszłości można byłoby ideę współpracy detektorów opartych na analizie błędów jednokrokowej predykcji rozszerzyć o dodatkowe detektory oparte na analizie błędów interpolacji. Wskazany kierunek badań miałby na celu zredukowanie liczby fałszywych alarmów do minimum przy jednoczesnym zachowaniu wysokiej wrażliwości schematu detekcji na obecność zakłóceń impulsowych.

W rozdziale 6 zaprezentowano możliwość łącznej analizy dwuwymiarowych sygnałów fonicznych. W przyszłości rozważania na temat usuwania zakłóceń impulsowych z sygnałów stereofonicznych można byłoby rozszerzyć na przypadek procesów wielowymiarowych. Tego typu zastosowanie może być użyteczne np. przy analizie sygnałów biomedycznych.

W rozdziale 7 zaprezentowano procedurę tworzenia bazy wzorców. W przyszłości można byłoby zaproponować adaptacyjną procedurę tworzenia bazy wzorców w oparciu o przebiegi zakłóceń impulsowych wykrywanych na bieżąco.

Dodatek A

Wybrane operacje na macierzach

A.1 Lemat o odwracaniu macierzy

Powyższy lemat pozwala uniknąć odwracania macierzy regresji w każdym kroku pracy algorytmu estymacji opartego na metodzie najmniejszych kwadratów. Umożliwia to istotne zmniejszenie jego złożoności obliczeniowej [75].

Lemat 1. Jeżeli wszystkie odwrotności macierzy w zależności (A.1) istnieją, to

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[C^{-1} + DA^{-1}B]^{-1}DA^{-1}. \quad (\text{A.1})$$

Dowód. Aby dowieść poprawności wzoru (A.1), dokonamy mnożenia odwrotności lewej strony równania przez prawą

$$\begin{aligned} & [A + BCD][A^{-1} - A^{-1}B[C^{-1} + DA^{-1}B]^{-1}DA^{-1}] = \\ & = I + BCDA^{-1} - [B + BCDA^{-1}B][C^{-1} + DA^{-1}B]^{-1}DA^{-1} = \\ & = I + BCDA^{-1} - BC[C^{-1} + DA^{-1}B][C^{-1} + DA^{-1}B]^{-1}DA^{-1} = I \end{aligned}$$

◇

Szczególnym przypadkiem tego lematu jest tzw. zależność Shermana-Morrisona [105]. Jeżeli wszystkie odwrotności macierzy w zależności (A.2) istnieją oraz spełniony jest warunek $1 \pm \mathbf{x}^T \mathbf{A}^{-1} \mathbf{x} \neq 0$, to

$$[A \pm \mathbf{x}\mathbf{x}^T]^{-1} = A^{-1} \mp \frac{A^{-1}\mathbf{x}\mathbf{x}^T A^{-1}}{1 \pm \mathbf{x}^T A^{-1} \mathbf{x}}, \quad (\text{A.2})$$

gdzie A jest macierzą o wymiarach $n \times n$, a \mathbf{x} jest wektorem o wymiarach $n \times 1$.



Dodatek B

Wyprowadzenie rekurencyjnych algorytmów estymacji

B.1 Wyprowadzenie zależności (3.58)-(3.59)

Ocena wariancji białego szumu może być otrzymana na drodze uśredniania sekwencji ważonych kwadratów błędów resztowych. Niech $r(t)$ oznacza sumę ważoną kwadratów błędów resztowych

$$r(t) = \sum_{i=1}^t \lambda^{t-i} [y(i) - \varphi^T(i) \hat{\boldsymbol{\theta}}(t)]^2. \quad (\text{B.1})$$

Zatem, zgodnie z (3.57)

$$\hat{\rho}(t) = \frac{r(t)}{l_e(t)}. \quad (\text{B.2})$$

Zauważmy, że

$$r(t) = \sum_{i=1}^t \lambda^{t-i} y^2(i) - 2\hat{\boldsymbol{\theta}}^T(t) \sum_{i=1}^t \lambda^{t-i} y(i) \boldsymbol{\varphi}(i) + \hat{\boldsymbol{\theta}}^T(t) \left[\sum_{i=1}^t \lambda^{t-i} \boldsymbol{\varphi}(i) \boldsymbol{\varphi}^T(i) \right] \hat{\boldsymbol{\theta}}(t). \quad (\text{B.3})$$

Uwzględniając zależność (3.36) oraz wprowadzając oznaczenie $\mathbf{R}(t) = \mathbf{Q}^{-1}(t)$, wzór (B.3) możemy przedstawić w postaci

$$r(t) = \sum_{i=1}^t \lambda^{t-i} y^2(i) - 2\hat{\boldsymbol{\theta}}^T(t) \mathbf{s}(t) + \hat{\boldsymbol{\theta}}^T(t) \mathbf{Q}^{-1}(t) \hat{\boldsymbol{\theta}}(t). \quad (\text{B.4})$$

Ponieważ $\hat{\boldsymbol{\theta}}(t) = \mathbf{Q}(t) \mathbf{s}(t)$, to wzór (B.4) upraszcza się do postaci

$$r(t) = \sum_{i=1}^t \lambda^{t-i} y^2(i) - \hat{\boldsymbol{\theta}}^T(t) \mathbf{Q}^{-1}(t) \hat{\boldsymbol{\theta}}(t). \quad (\text{B.5})$$

Stąd rekurencja

$$r(t) - \lambda r(t-1) = y^2(t) + \lambda \hat{\boldsymbol{\theta}}^T(t-1) \mathbf{Q}^{-1}(t-1) \hat{\boldsymbol{\theta}}(t-1) - \hat{\boldsymbol{\theta}}^T(t) \mathbf{Q}^{-1}(t) \hat{\boldsymbol{\theta}}(t), \quad (\text{B.6})$$



gdzie warunek początkowy $r(0) = 0$. Uwzględniając zależności

$$\begin{aligned}\mathbf{Q}^{-1}(t) &= \lambda \mathbf{Q}^{-1}(t-1) + \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t), \\ \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \mathbf{Q}(t)\boldsymbol{\varphi}(t)\varepsilon(t|t-1)\end{aligned}\quad (\text{B.7})$$

możemy pokazać, że

$$\begin{aligned}\hat{\boldsymbol{\theta}}^T(t)\mathbf{Q}^{-1}(t)\hat{\boldsymbol{\theta}}(t) &= \lambda \hat{\boldsymbol{\theta}}^T(t-1)\mathbf{Q}^{-1}(t-1)\hat{\boldsymbol{\theta}}(t-1) + \\ &+ \hat{\boldsymbol{\theta}}^T(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t)\hat{\boldsymbol{\theta}}(t-1) + \\ &+ \boldsymbol{\varphi}^T(t)\mathbf{Q}(t)\boldsymbol{\varphi}(t)\varepsilon^2(t|t-1) + \\ &+ 2\boldsymbol{\varphi}^T(t)\hat{\boldsymbol{\theta}}(t-1)\varepsilon(t|t-1).\end{aligned}\quad (\text{B.8})$$

Ostatecznie, po połączeniu (B.6) z (B.8) oraz po uwzględnieniu zależności

$$\begin{aligned}y(t) &= \boldsymbol{\varphi}^T(t)\hat{\boldsymbol{\theta}}(t-1) + \varepsilon(t|t-1), \\ \mathbf{k}(t) &= \mathbf{Q}(t)\boldsymbol{\varphi}(t) = \frac{\mathbf{Q}(t-1)\boldsymbol{\varphi}(t)}{\lambda + \boldsymbol{\varphi}^T(t)\mathbf{Q}(t-1)\boldsymbol{\varphi}(t)},\end{aligned}\quad (\text{B.9})$$

otrzymujemy

$$\begin{aligned}r(t) &= \lambda r(t-1) + [1 - \boldsymbol{\varphi}^T(t)\mathbf{Q}(t)\boldsymbol{\varphi}(t)]\varepsilon^2(t|t-1) = \\ &= \lambda r(t-1) + [1 - \mathbf{k}^T(t)\boldsymbol{\varphi}(t)]\varepsilon^2(t|t-1) = \\ &= \lambda r(t-1) + \frac{\lambda \varepsilon^2(t|t-1)}{\lambda + \boldsymbol{\varphi}^T(t)\mathbf{Q}(t-1)\boldsymbol{\varphi}(t)}.\end{aligned}\quad (\text{B.10})$$

W stanie ustalonym, po połączeniu (B.2) z (B.10) oraz po uwzględnieniu zależności (3.44), otrzymujemy estymator wariancji białego szumu tworzącego $n(t)$ postaci

$$\hat{\rho}(t) = \lambda \hat{\rho}(t-1) + (1 - \lambda)\varepsilon^2(t|t-1)\varsigma(t), \quad (\text{B.11})$$

gdzie

$$\varsigma(t) = [1 - \boldsymbol{\varphi}^T(t)\mathbf{Q}(t)\boldsymbol{\varphi}(t)] = [1 - \mathbf{k}^T(t)\boldsymbol{\varphi}(t)] = \frac{\lambda}{\lambda + \boldsymbol{\varphi}^T(t)\mathbf{Q}(t-1)\boldsymbol{\varphi}(t)}. \quad (\text{B.12})$$

B.2 Wyprowadzenie zależności (3.61)

Ocena wariancji białego szumu może być również otrzymana na drodze uśredniania sekwencji ważonych kwadratów błędów jednokrokowej predykcji. Niech $r(t)$ oznacza sumę ważoną kwadratów błędów jednokrokowej predykcji

$$r(t) = \sum_{i=1}^t \lambda^{t-i} \varepsilon^2(i|i-1). \quad (\text{B.13})$$

Zatem, zgodnie z (3.60)

$$\hat{\rho}(t) = \frac{r(t)}{l_e(t)}. \quad (\text{B.14})$$

Zauważmy, że

$$r(t) - \lambda r(t-1) = \varepsilon^2(t|t-1). \quad (\text{B.15})$$

Stąd rekurencja

$$r(t) = \lambda r(t-1) + \varepsilon^2(t|t-1), \quad (\text{B.16})$$

gdzie warunek początkowy $r(0) = 0$.

W stanie ustalonym, po połączeniu (B.14) z (B.16) oraz po uwzględnieniu zależności (3.44), otrzymujemy estymator wariancji białego szumu tworzącego $n(t)$ postaci

$$\hat{\rho}(t) = \lambda \hat{\rho}(t-1) + (1-\lambda) \varepsilon^2(t|t-1). \quad (\text{B.17})$$

Dodatek C

Tabele z wynikami

W tym dodatku zamieszczono tabele zawierające oceny PEAQ uzyskane dla wszystkich nagrań testowych. Wyniki dla metody wykorzystującej technikę dopasowywania wzorców zamieszczono w tabeli C.1. Wyniki odnoszące się do analizy dwukierunkowej zamieszczono w tabelach C.2, C.3, C.4 oraz C.5. Wyniki związane z modelowaniem dwuzakresowym zamieszczono w tabelach C.6, C.7, C.8 oraz C.9. Wyniki odnoszące się do wektorowego modelowania sygnałów fonicznych zamieszczono w tabelach C.10, C.11, C.12 oraz C.13.

Tablica C.1: Porównanie skuteczności działania 5 podejść na podstawie wyników PEAQ: podejście oparte na detektorze pracującym w pętli otwartej (A), podejście A rozszerzone o wykorzystanie techniki dopasowywania wzorców (A*), podejście oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (B), podejście B rozszerzone o wykorzystanie techniki dopasowywania wzorców (B*) oraz podejście B*, w którym interpolacja zastąpiona została kompensacją zakłóceń (B⁻). Dodatkowo, tabela pokazuje wyniki otrzymane dla spreparowanych nagrań fonicznych (REF) oraz wyniki odniesienia (GT) otrzymane gdy sygnał jest rekonstruowany przy pełnej wiedzy o położeniu zakłóceń impulsowych. Śr.₆₀ oznacza uśrednione wyniki w grupie wszystkich 60 nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (znieszczenia sygnału), -1 = zauważalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

nagranie	GT	REF	A	A*	B	B*	B ⁻
1	-0,03	-3,68	-3,22	-1,10	-0,20	-0,16	-3,47
2	-0,06	-3,69	-3,34	-1,05	-0,16	-0,14	-3,56
3	-0,05	-3,70	-3,70	-2,97	-0,63	-0,47	-3,68
4	-0,01	-3,59	-3,43	-1,39	-0,32	-0,28	-3,51
5	-0,09	-3,77	-3,69	-2,60	-0,36	-0,20	-3,70
6	-0,06	-3,84	-3,65	-3,48	-1,48	-1,05	-3,83
7	-0,02	-3,82	-3,45	-2,14	-0,80	-0,49	-3,78
8	-0,03	-3,55	-3,14	-2,76	-1,78	-0,92	-3,65
9	-0,02	-3,13	-3,11	-0,95	-0,46	-0,37	-2,98
10	-0,01	-3,01	-2,86	-1,30	-0,92	-0,68	-3,00
11	-0,07	-3,77	-3,57	-2,17	-0,97	-0,62	-3,71
12	-0,08	-3,59	-3,00	-1,28	-0,40	-0,35	-3,47
13	-0,03	-3,88	-3,65	-2,58	-0,82	-0,66	-3,79
14	-0,06	-3,82	-3,42	-2,27	-0,91	-0,76	-3,71
15	-0,01	-3,73	-3,41	-1,51	-0,70	-0,59	-3,67
16	-0,02	-3,68	-3,38	-1,53	-0,65	-0,53	-3,50
17	-0,06	-3,57	-3,19	-1,17	-0,37	-0,29	-3,40
18	-0,04	-3,34	-3,19	-2,81	-0,28	-0,19	-3,57
19	-0,05	-3,81	-3,57	-1,81	-0,62	-0,67	-3,63
20	-0,02	-3,78	-2,98	-3,10	-0,98	-0,76	-3,80
21	-0,01	-3,64	-3,05	-1,54	-0,49	-0,45	-3,53
22	-0,03	-3,82	-3,49	-2,91	-0,35	-0,32	-3,73
23	-0,08	-3,83	-3,51	-2,47	-0,84	-0,61	-3,72
24	-0,01	-3,80	-3,44	-2,33	-0,30	-0,29	-3,72
25	-0,05	-3,89	-3,67	-2,72	-1,11	-0,78	-3,82
26	-0,04	-3,64	-3,29	-1,49	-0,92	-0,58	-3,43
27	-0,02	-3,73	-3,26	-2,14	-1,22	-0,90	-3,55
28	-0,09	-3,91	-3,74	-3,65	-1,77	-1,33	-3,88
29	-0,26	-3,37	-3,74	-3,53	-1,49	-1,27	-3,16
30	-0,04	-3,66	-3,40	-2,54	-0,24	-0,21	-3,51

kontynuacja tabeli

nagranie	GT	REF	A	A*	B	B*	B ⁻
31	-0,11	-3,91	-3,88	-3,60	-2,52	-0,86	-3,90
32	-0,01	-3,82	-3,44	-2,28	-1,77	-1,48	-3,70
33	-0,17	-3,79	-3,63	-2,37	-1,54	-1,16	-3,70
34	-0,02	-3,87	-3,64	-2,85	-1,38	-1,09	-3,77
35	0,00	-3,80	-3,26	-2,85	-0,55	-0,44	-3,76
36	-0,07	-3,78	-3,32	-1,65	-0,15	-0,08	-3,70
37	-0,02	-3,83	-3,21	-2,97	-0,40	-0,30	-3,80
38	0,00	-3,85	-3,51	-3,26	-0,69	-0,47	-3,83
39	0,00	-3,80	-3,37	-2,80	-0,35	-0,21	-3,78
40	-0,07	-3,81	-3,36	-2,36	-0,21	-0,17	-3,77
41	-0,04	-3,91	-3,91	-3,87	-0,72	-0,76	-3,91
42	-0,04	-3,86	-3,64	-3,56	-1,06	-0,74	-3,82
43	-0,06	-3,91	-3,86	-3,77	-0,29	-0,51	-3,89
44	-0,07	-3,74	-3,56	-1,85	-0,30	-0,22	-3,68
45	-0,01	-3,91	-3,87	-3,81	-0,37	-0,44	-3,88
46	-0,08	-3,78	-3,66	-2,82	-0,86	-0,69	-3,78
47	-0,03	-3,81	-3,64	-3,11	-0,22	-0,20	-3,77
48	-0,12	-3,86	-3,82	-3,65	-2,34	-1,81	-3,82
49	-0,06	-3,74	-3,46	-1,97	-0,15	-0,16	-3,71
50	-0,11	-3,75	-3,77	-3,46	-1,10	-0,92	-3,64
51	-0,10	-3,77	-3,86	-3,73	-2,40	-1,38	-3,76
52	-0,01	-3,81	-3,83	-3,35	-1,06	-0,35	-3,79
53	-0,08	-3,84	-3,85	-3,78	-2,87	-1,32	-3,83
54	-0,02	-3,87	-3,84	-3,62	-0,95	-0,41	-3,84
55	-0,07	-3,87	-3,81	-3,40	-0,87	-0,39	-3,85
56	0,00	-3,91	-3,91	-3,86	-0,63	-0,27	-3,91
57	-0,02	-3,91	-3,91	-3,84	-1,20	-0,71	-3,91
58	-0,05	-3,91	-3,91	-3,82	-0,77	-0,66	-3,91
59	-0,10	-3,91	-3,91	-3,86	-1,31	-1,49	-3,91
60	-0,01	-3,91	-3,91	-3,86	-0,73	-0,40	-3,91
Śr. ₆₀	-0,05	-3,75	-3,53	-2,69	-0,87	-0,62	-3,69

Tablica C.2: Porównanie skuteczności działania 3 podejść, opartych na modelu AR, do adaptacyjnej interpolacji sygnału na podstawie otrzymanych wyników PEAQ: podejście wykorzystujące rekonstrukcję lewostronną (GT_f), podejście wykorzystujące rekonstrukcję prawostronną (GT_b) oraz podejście wykorzystujące rekonstrukcję dwukierunkową (GT_{fb}). Rekonstrukcja sygnału przeprowadzona została przy pełnej wiedzy o położeniu zakłóceń impulsowych. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Śr. oznacza uśredniony wynik dla 10 wyszczególnionych nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

muzyka jazzowa

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,38	-0,31	-0,71	-0,22	-0,24	-0,27	-0,46	-0,30	-0,34	-0,39	-0,36
GT_b	-0,38	-0,33	-0,72	-0,22	-0,25	-0,26	-0,46	-0,31	-0,33	-0,41	-0,37
GT_{fb}	-0,37	-0,30	-0,70	-0,21	-0,23	-0,26	-0,45	-0,29	-0,33	-0,39	-0,35

muzyka chóralna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,32	-0,12	-0,21	-0,33	-0,26	-0,13	-0,19	-0,19	-0,40	-0,19	-0,23
GT_b	-0,32	-0,14	-0,21	-0,34	-0,25	-0,13	-0,18	-0,21	-0,39	-0,18	-0,23
GT_{fb}	-0,31	-0,12	-0,21	-0,32	-0,25	-0,13	-0,18	-0,19	-0,38	-0,18	-0,23

muzyka operowa

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,65	-0,26	-0,46	-0,34	-0,37	-0,32	-0,58	-0,33	-0,39	-0,44	-0,41
GT_b	-0,68	-0,25	-0,49	-0,37	-0,36	-0,34	-0,63	-0,35	-0,39	-0,42	-0,43
GT_{fb}	-0,63	-0,24	-0,46	-0,34	-0,35	-0,31	-0,58	-0,33	-0,37	-0,42	-0,40

muzyka instrumentalna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,76	-0,39	-0,32	-0,65	-0,23	-0,47	-0,33	-0,40	-0,25	-0,19	-0,40
GT_b	-0,76	-0,39	-0,32	-0,66	-0,23	-0,48	-0,33	-0,39	-0,22	-0,18	-0,40
GT_{fb}	-0,73	-0,38	-0,31	-0,65	-0,22	-0,46	-0,32	-0,39	-0,23	-0,18	-0,39

muzyka klasyczna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,43	-0,26	-0,12	-0,12	-0,35	-0,21	-0,11	-0,52	-0,19	-0,29	-0,26
GT_b	-0,44	-0,25	-0,11	-0,12	-0,35	-0,22	-0,11	-0,55	-0,19	-0,30	-0,26
GT_{fb}	-0,43	-0,24	-0,11	-0,12	-0,34	-0,21	-0,10	-0,52	-0,18	-0,28	-0,25

Tablica C.3: Porównanie skuteczności działania 5 podejść do eliminacji zakłóceń impulsowych opartych na modelu AR na podstawie otrzymanych wyników PEAQ: podejście wykorzystujące analizę lewostronną (L), podejście wykorzystujące analizę prawostronną (P), podejście wykorzystujące analizę dwukierunkową z globalną regułą sumy (S), podejście wykorzystujące analizę dwukierunkową z globalną regułą iloczynu (I) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem reguł lokalnych (RL). Porównanie przeprowadzono w dwóch grupach podejść, którymi są: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów z bazy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Symbolem GT_{fb} oznaczono wyniki PEAQ otrzymane przez adaptacyjny algorytm rekonstrukcji sygnału w wariancie dwukierunkowym, przy pełnej wiedzy o położeniu zakłóceń impulsowych. REF oznacza początkową ocenę PEAQ dla spreparowanego nagrania. Śr. oznacza uśredniony wynik dla 10 wyszczególnionych nagrań. W obu grupach najlepsze wyniki dla każdego nagrania zostały pokazane wytłuszczoną czcionką.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

muzyka jazzowa

nagranie			grupa A					grupa B				
Lp.	GT_{fb}	REF	L	P	S	I	RL	L	P	S	I	RL
1	-0,37	-3,61	-3,63	-3,63	-1,57	-3,45	-0,80	-0,96	-0,83	-0,82	-0,68	-0,51
2	-0,30	-3,48	-1,71	-1,63	-1,64	-1,28	-1,08	-0,94	-0,66	-1,53	-0,60	-0,66
3	-0,70	-2,25	-3,45	-3,42	-1,72	-3,41	-1,09	-1,42	-1,25	-1,11	-0,95	-0,97
4	-0,21	-3,22	-1,18	-1,04	-0,46	-0,96	-0,34	-0,41	-0,37	-0,36	-0,34	-0,31
5	-0,23	-3,54	-2,81	-2,53	-1,71	-1,78	-1,02	-0,73	-0,59	-2,36	-0,95	-0,53
6	-0,26	-2,80	-1,72	-1,72	-1,73	-0,96	-1,12	-1,05	-0,91	-1,75	-0,63	-0,93
7	-0,45	-3,46	-3,33	-3,48	-2,01	-1,79	-0,90	-1,00	-1,28	-1,48	-0,59	-0,79
8	-0,29	-3,11	-1,22	-1,26	-0,78	-1,13	-0,57	-0,71	-0,61	-0,68	-0,54	-0,56
9	-0,33	-3,35	-1,87	-1,70	-1,50	-1,37	-0,88	-1,76	-0,78	-1,22	-1,27	-0,83
10	-0,39	-3,77	-3,55	-3,67	-2,68	-3,42	-1,74	-2,18	-1,60	-2,20	-1,78	-1,61
Śr.	-0,35	-3,26	-2,45	-2,41	-1,58	-1,95	-0,95	-1,12	-0,89	-1,35	-0,83	-0,77

muzyka chóralna

nagranie			grupa A					grupa B				
Lp.	GT_{fb}	REF	L	P	S	I	RL	L	P	S	I	RL
1	-0,31	-3,91	-3,84	-3,81	-2,09	-3,85	-1,07	-2,51	-1,51	-0,85	-1,69	-0,54
2	-0,12	-3,78	-3,58	-3,65	-0,84	-3,40	-0,34	-0,48	-0,47	-0,27	-0,59	-0,20
3	-0,21	-3,76	-3,51	-3,62	-1,22	-3,20	-0,79	-0,44	-0,47	-0,29	-0,40	-0,22
4	-0,32	-3,90	-3,81	-3,81	-3,05	-3,84	-2,60	-1,55	-1,45	-0,97	-1,19	-0,57
5	-0,25	-3,86	-3,73	-3,67	-1,96	-3,62	-1,12	-1,15	-1,29	-1,40	-1,01	-1,01
6	-0,13	-3,91	-3,83	-3,81	-2,02	-3,84	-0,62	-0,84	-0,80	-0,56	-0,78	-0,27
7	-0,18	-3,86	-3,69	-3,70	-1,07	-3,53	-0,44	-0,58	-0,57	-0,36	-0,64	-0,25
8	-0,19	-3,71	-3,37	-3,55	-0,77	-3,10	-0,43	-0,47	-0,48	-0,27	-0,29	-0,22
9	-0,38	-3,89	-3,80	-3,73	-2,18	-3,73	-0,87	-1,27	-1,31	-0,87	-1,50	-0,50
10	-0,18	-3,78	-3,48	-3,57	-0,71	-3,15	-0,53	-0,50	-0,49	-0,29	-0,31	-0,24
Śr.	-0,23	-3,84	-3,66	-3,69	-1,59	-3,52	-0,88	-0,98	-0,88	-0,61	-0,84	-0,40

muzyka operowa

nagranie			grupa A					grupa B				
Lp.	GT _{fb}	REF	L	P	S	I	RL	L	P	S	I	RL
1	-0,63	-3,80	-3,63	-3,62	-1,81	-3,20	-0,84	-1,18	-1,06	-0,98	-0,61	-0,60
2	-0,24	-3,83	-3,66	-3,62	-1,39	-3,28	-0,65	-0,78	-0,76	-0,58	-0,84	-0,39
3	-0,46	-3,84	-3,71	-3,65	-2,28	-3,64	-1,27	-1,17	-1,19	-1,01	-0,55	-0,51
4	-0,34	-3,83	-3,75	-3,69	-1,69	-3,64	-0,64	-1,13	-1,10	-0,84	-1,26	-0,43
5	-0,35	-3,87	-3,76	-3,73	-1,80	-3,68	-0,76	-1,10	-1,06	-0,82	-0,50	-0,43
6	-0,31	-3,82	-3,75	-3,71	-1,66	-3,50	-0,74	-0,98	-0,97	-0,78	-0,45	-0,44
7	-0,58	-3,88	-3,79	-3,72	-3,02	-3,80	-1,48	-1,82	-1,87	-1,73	-0,85	-0,92
8	-0,33	-3,73	-3,68	-3,66	-1,42	-3,53	-0,58	-0,85	-0,88	-0,67	-0,41	-0,39
9	-0,37	-3,87	-3,77	-3,72	-2,43	-3,75	-1,02	-1,37	-1,42	-1,08	-0,74	-0,49
10	-0,42	-3,86	-3,77	-3,73	-2,11	-3,76	-0,98	-1,23	-1,11	-0,88	-0,58	-0,48
Śr.	-0,40	-3,83	-3,73	-3,68	-1,96	-3,58	-0,90	-1,16	-1,14	-0,94	-0,68	-0,51

muzyka instrumentalna

nagranie			grupa A					grupa B				
Lp.	GT _{fb}	REF	L	P	S	I	RL	L	P	S	I	RL
1	-0,73	-3,68	-3,61	-3,61	-1,46	-3,02	-0,89	-1,03	-1,07	-0,91	-0,72	-0,71
2	-0,38	-3,49	-2,91	-3,18	-0,69	-2,18	-0,42	-0,50	-0,51	-0,41	-0,35	-0,32
3	-0,31	-3,49	-2,71	-2,89	-0,68	-1,64	-0,34	-0,43	-0,39	-0,35	-0,31	-0,28
4	-0,65	-3,51	-3,11	-3,25	-0,95	-1,95	-0,61	-0,69	-0,70	-0,62	-0,53	-0,49
5	-0,22	-3,55	-2,15	-2,30	-0,46	-1,31	-0,29	-0,32	-0,30	-0,25	-0,30	-0,21
6	-0,46	-3,38	-2,91	-3,14	-0,72	-2,01	-0,45	-0,54	-0,57	-0,46	-0,39	-0,36
7	-0,32	-2,91	-3,66	-3,55	-2,26	-3,54	-0,91	-1,34	-1,08	-1,09	-0,89	-0,56
8	-0,39	-2,70	-3,71	-3,53	-2,94	-3,67	-1,32	-1,32	-1,29	-1,27	-0,63	-0,59
9	-0,23	-2,12	-3,64	-3,57	-2,06	-3,54	-1,14	-0,96	-0,82	-0,79	-0,81	-0,60
10	-0,18	-2,91	-3,75	-3,72	-1,85	-3,73	-0,88	-1,01	-0,86	-0,67	-0,85	-0,54
Śr.	-0,39	-3,17	-3,22	-3,27	-1,41	-2,66	-0,73	-0,81	-0,76	-0,68	-0,58	-0,47

muzyka klasyczna

nagranie			grupa A					grupa B				
Lp.	GT _{fb}	REF	L	P	S	I	RL	L	P	S	I	RL
1	-0,43	-3,82	-3,75	-3,66	-1,92	-3,75	-0,85	-1,06	-1,12	-0,85	-0,75	-0,42
2	-0,24	-3,71	-3,42	-3,38	-0,89	-2,89	-0,73	-0,47	-0,49	-0,34	-0,30	-0,25
3	-0,11	-3,91	-3,83	-3,79	-2,22	-3,83	-0,95	-0,89	-0,89	-0,76	-0,35	-0,37
4	-0,12	-3,90	-3,78	-3,75	-1,62	-3,76	-0,75	-1,07	-0,83	-0,81	-0,34	-0,29
5	-0,34	-3,52	-3,59	-3,50	-1,28	-3,57	-0,78	-0,83	-0,80	-0,61	-0,55	-0,42
6	-0,21	-3,66	-3,24	-3,38	-0,58	-2,86	-0,33	-0,42	-0,40	-0,27	-0,25	-0,21
7	-0,10	-3,55	-2,54	-2,95	-0,51	-1,97	-0,29	-0,28	-0,30	-0,17	-0,39	-0,13
8	-0,52	-3,90	-3,80	-3,78	-2,59	-3,79	-1,20	-1,62	-1,81	-1,60	-0,81	-0,81
9	-0,18	-2,98	-1,28	-1,26	-0,39	-1,03	-0,34	-0,39	-0,29	-0,34	-0,23	-0,29
10	-0,28	-3,81	-3,42	-3,54	-0,90	-2,74	-0,48	-0,69	-0,63	-0,64	-0,50	-0,35
Śr.	-0,25	-3,68	-3,27	-3,30	-1,29	-3,02	-0,67	-0,77	-0,76	-0,64	-0,45	-0,35

Tablica C.4: Porównanie skuteczności działania 5 podejść wykorzystujących analizę dwukierunkową na podstawie wyników PEAQ. Porównywane podejścia, różniące się między sobą wartością parametru Δ_1 , oznaczono symbolami RL_i , gdzie $i = \Delta_1 \in \{0, 1, 2, 3, 4\}$. Porównanie przeprowadzono w dwóch grupach podejść, którymi są: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z bazy treningowej: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Śr. oznacza uśredniony wynik dla 10 wyszczególnionych nagrań. W obu grupach najlepsze wyniki PEAQ dla każdego nagrania zostały pokazane wytłuszczoną czcionką.

Interpretacja wyników PEAQ: 0 = niesłyszalne (znieskształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

muzyka jazzowa

Lp.	grupa A					grupa B				
	RL_0	RL_1	RL_2	RL_3	RL_4	RL_0	RL_1	RL_2	RL_3	RL_4
1	-1,38	-0,81	-0,80	-0,87	-0,96	-0,91	-0,48	-0,51	-0,62	-0,76
2	-1,16	-1,02	-1,08	-1,16	-1,13	-0,77	-0,66	-0,66	-0,74	-0,73
3	-1,48	-1,09	-1,09	-1,15	-1,26	-1,31	-0,94	-0,97	-1,06	-1,14
4	-0,44	-0,37	-0,34	-0,33	-0,34	-0,39	-0,33	-0,31	-0,31	-0,32
5	-1,18	-1,06	-1,02	-1,17	-1,10	-0,58	-0,54	-0,53	-0,55	-0,58
6	-1,07	-1,01	-1,12	-1,18	-1,44	-0,87	-0,86	-0,93	-0,96	-1,12
7	-1,18	-0,86	-0,90	-1,15	-1,48	-1,04	-0,73	-0,79	-1,01	-1,40
8	-0,69	-0,58	-0,57	-0,60	-0,63	-0,65	-0,56	-0,56	-0,59	-0,61
9	-1,03	-0,85	-0,88	-0,95	-1,02	-0,93	-0,80	-0,83	-0,90	-0,96
10	-2,21	-1,74	-1,74	-1,83	-1,97	-1,95	-1,55	-1,61	-1,74	-1,91
Śr.	-1,18	-0,94	-0,95	-1,04	-1,13	-0,94	-0,75	-0,77	-0,85	-0,95

muzyka chóralna

Lp.	grupa A					grupa B				
	RL_0	RL_1	RL_2	RL_3	RL_4	RL_0	RL_1	RL_2	RL_3	RL_4
1	-2,41	-1,37	-1,07	-1,03	-1,06	-1,54	-0,61	-0,54	-0,58	-0,68
2	-0,95	-0,46	-0,34	-0,33	-0,37	-0,58	-0,25	-0,20	-0,21	-0,27
3	-1,18	-0,84	-0,79	-0,78	-0,79	-0,55	-0,25	-0,22	-0,26	-0,28
4	-3,43	-3,03	-2,60	-2,69	-2,70	-1,71	-0,75	-0,57	-0,60	-0,73
5	-2,06	-1,00	-1,12	-1,22	-1,40	-1,05	-0,84	-1,01	-1,15	-1,35
6	-1,85	-0,92	-0,62	-0,63	-0,68	-1,06	-0,36	-0,27	-0,31	-0,41
7	-1,15	-0,55	-0,44	-0,43	-0,48	-0,72	-0,29	-0,25	-0,27	-0,33
8	-1,16	-0,48	-0,43	-0,42	-0,46	-0,60	-0,25	-0,22	-0,22	-0,28
9	-2,21	-1,14	-0,87	-0,93	-1,00	-1,37	-0,60	-0,50	-0,62	-0,74
10	-0,87	-0,58	-0,53	-0,38	-0,39	-0,61	-0,26	-0,24	-0,25	-0,30
Śr.	-1,73	-1,04	-0,88	-0,88	-0,93	-0,98	-0,45	-0,40	-0,45	-0,54



muzyka operowa

	grupa A					grupa B				
Lp.	RL ₀	RL ₁	RL ₂	RL ₃	RL ₄	RL ₀	RL ₁	RL ₂	RL ₃	RL ₄
1	-1,72	-1,02	-0,84	-0,82	-0,89	-1,03	-0,56	-0,60	-0,60	-0,75
2	-1,40	-0,95	-0,65	-0,66	-0,70	-0,81	-0,40	-0,39	-0,43	-0,50
3	-1,98	-1,35	-1,27	-1,10	-1,18	-1,21	-0,62	-0,51	-0,67	-0,87
4	-1,80	-0,85	-0,64	-0,67	-0,79	-1,18	-0,48	-0,43	-0,49	-0,64
5	-1,88	-0,99	-0,76	-0,76	-0,85	-1,11	-0,46	-0,43	-0,51	-0,65
6	-1,63	-0,81	-0,74	-0,70	-0,78	-1,00	-0,46	-0,44	-0,50	-0,62
7	-2,87	-1,71	-1,48	-1,55	-1,68	-1,74	-0,83	-0,92	-1,06	-1,28
8	-1,47	-0,80	-0,58	-0,60	-0,67	-0,90	-0,46	-0,39	-0,46	-0,56
9	-2,28	-1,23	-1,02	-1,01	-1,14	-1,45	-0,56	-0,49	-0,58	-0,75
10	-2,22	-1,26	-0,98	-1,01	-0,88	-1,29	-0,56	-0,48	-0,55	-0,67
Śr.	-1,93	-1,10	-0,90	-0,89	-0,96	-1,17	-0,54	-0,51	-0,59	-0,73

muzyka instrumentalna

	grupa A					grupa B				
Lp.	RL ₀	RL ₁	RL ₂	RL ₃	RL ₄	RL ₀	RL ₁	RL ₂	RL ₃	RL ₄
1	-1,41	-0,87	-0,89	-0,89	-1,01	-1,04	-0,66	-0,71	-0,78	-0,91
2	-0,74	-0,44	-0,42	-0,48	-0,52	-0,53	-0,32	-0,32	-0,37	-0,45
3	-0,61	-0,36	-0,34	-0,37	-0,38	-0,43	-0,26	-0,28	-0,29	-0,34
4	-0,92	-0,59	-0,61	-0,63	-0,70	-0,68	-0,43	-0,49	-0,55	-0,65
5	-0,50	-0,33	-0,29	-0,29	-0,30	-0,35	-0,22	-0,21	-0,23	-0,25
6	-0,77	-0,47	-0,45	-0,45	-0,51	-0,57	-0,34	-0,36	-0,39	-0,46
7	-1,85	-1,13	-0,91	-0,92	-1,00	-1,17	-0,60	-0,56	-0,61	-0,77
8	-2,37	-1,55	-1,32	-1,36	-1,44	-1,29	-0,64	-0,59	-0,74	-0,87
9	-1,72	-1,27	-1,14	-1,15	-1,17	-1,01	-0,62	-0,60	-0,62	-0,68
10	-2,33	-2,08	-0,88	-0,82	-0,82	-1,22	-0,71	-0,54	-0,52	-0,56
Śr.	-1,32	-0,91	-0,73	-0,74	-0,78	-0,83	-0,48	-0,47	-0,51	-0,59

muzyka klasyczna

	grupa A					grupa B				
Lp.	RL ₀	RL ₁	RL ₂	RL ₃	RL ₄	RL ₀	RL ₁	RL ₂	RL ₃	RL ₄
1	-2,01	-1,04	-0,85	-0,85	-1,04	-1,14	-0,45	-0,42	-0,52	-0,73
2	-0,93	-0,78	-0,73	-0,72	-0,74	-0,56	-0,26	-0,25	-0,27	-0,33
3	-2,27	-1,38	-0,95	-0,66	-0,70	-1,21	-0,52	-0,37	-0,39	-0,45
4	-1,67	-0,95	-0,75	-0,68	-0,69	-0,94	-0,40	-0,29	-0,35	-0,41
5	-2,10	-0,96	-0,78	-0,66	-0,70	-0,85	-0,45	-0,42	-0,48	-0,59
6	-0,88	-0,62	-0,33	-0,31	-0,33	-0,48	-0,26	-0,21	-0,22	-0,26
7	-0,63	-0,39	-0,29	-0,28	-0,29	-0,37	-0,15	-0,13	-0,16	-0,17
8	-2,31	-1,30	-1,20	-1,34	-1,60	-1,62	-0,74	-0,81	-1,09	-1,44
9	-0,43	-0,35	-0,34	-0,34	-0,34	-0,37	-0,30	-0,29	-0,32	-0,29
10	-0,91	-0,57	-0,48	-0,56	-0,62	-0,61	-0,36	-0,35	-0,41	-0,52
Śr.	-1,41	-0,83	-0,67	-0,64	-0,70	-0,82	-0,39	-0,35	-0,42	-0,52

Tablica C.5: Porównanie skuteczności działania 2 podejść na podstawie wyników PEAQ: podejście oparte na łączeniu rekonstrukcji lewostronnej i prawostronnej (M) oraz podejście wykorzystujące analizę dwukierunkową z proponowanym zestawem reguł lokalnych (RL). Porównanie przeprowadzono w dwóch grupach podejść, którymi są: podejścia oparte na detektorze pracującym w pętli otwartej (grupa A) oraz podejścia oparte na detektorze pracującym w pętli z decyzyjnym sprzężeniem zwrotnym (grupa B). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z bazy treningowej. Śr.₅₀ oznacza uśredniony wynik dla wszystkich 50 nagrań. W obu grupach najlepsze wyniki dla każdego nagrania zostały pokazane wytłuszczoną czcionką.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

nagranie		grupa A		grupa B		nagranie		grupa A		grupa B	
Lp.	REF	M	RL	M	RL	Lp.	REF	M	RL	M	RL
1	-3,61	-3,45	-0,80	-0,60	-0,51	26	-3,82	-3,71	-0,74	-0,58	-0,44
2	-3,48	-1,35	-1,08	-0,88	-0,66	27	-3,88	-3,78	-1,48	-1,10	-0,92
3	-2,25	-3,14	-1,09	-1,12	-0,97	28	-3,73	-3,63	-0,58	-0,53	-0,39
4	-3,22	-0,81	-0,34	-0,36	-0,31	29	-3,87	-3,75	-1,02	-0,74	-0,49
5	-3,54	-2,70	-1,02	-0,68	-0,53	30	-3,86	-3,74	-0,98	-0,68	-0,48
6	-2,80	-1,38	-1,12	-0,90	-0,93	31	-3,68	-3,55	-0,89	-0,76	-0,71
7	-3,46	-3,27	-0,90	-0,97	-0,79	32	-3,49	-2,47	-0,42	-0,34	-0,32
8	-3,11	-0,96	-0,57	-0,61	-0,56	33	-3,49	-2,13	-0,34	-0,28	-0,28
9	-3,35	-1,24	-0,88	-0,85	-0,83	34	-3,51	-2,74	-0,61	-0,51	-0,49
10	-3,77	-3,30	-1,74	-1,70	-1,61	35	-3,55	-1,41	-0,29	-0,22	-0,21
11	-3,91	-3,82	-1,07	-0,77	-0,54	36	-3,38	-2,31	-0,45	-0,39	-0,36
12	-3,78	-3,47	-0,33	-0,26	-0,20	37	-2,91	-3,68	-0,91	-0,68	-0,56
13	-3,76	-3,35	-0,79	-0,26	-0,22	38	-2,70	-3,71	-1,32	-0,78	-0,59
14	-3,90	-3,82	-2,60	-0,83	-0,57	39	-2,12	-3,61	-1,14	-0,63	-0,60
15	-3,86	-3,71	-1,12	-1,01	-1,01	40	-2,91	-3,66	-0,88	-0,62	-0,54
16	-3,91	-3,83	-0,62	-0,42	-0,27	41	-3,82	-3,75	-0,85	-0,58	-0,42
17	-3,86	-3,62	-0,43	-0,31	-0,25	42	-3,71	-3,31	-0,73	-0,27	-0,25
18	-3,71	-2,92	-0,43	-0,25	-0,22	43	-3,91	-3,81	-0,95	-0,47	-0,37
19	-3,89	-3,76	-0,87	-0,72	-0,50	44	-3,90	-3,78	-0,75	-0,46	-0,29
20	-3,78	-3,12	-0,53	-0,26	-0,24	45	-3,52	-3,55	-0,78	-0,53	-0,42
21	-3,80	-3,57	-0,84	-0,80	-0,60	46	-3,66	-2,95	-0,33	-0,25	-0,21
22	-3,83	-3,61	-0,65	-0,45	-0,39	47	-3,55	-2,24	-0,29	-0,17	-0,13
23	-3,84	-3,69	-1,27	-0,73	-0,51	48	-3,90	-3,78	-1,19	-1,00	-0,81
24	-3,83	-3,70	-0,64	-0,64	-0,43	49	-2,98	-0,92	-0,34	-0,35	-0,29
25	-3,87	-3,74	-0,76	-0,59	-0,43	50	-3,81	-3,33	-0,48	-0,42	-0,35
Śr. ₅₀		-3,56		-3,09	-0,82			-0,61		-0,50	

Tablica C.6: Porównanie skuteczności działania 3 podejść, opartych na modelu SAR wykorzystującym przyczynową ocenę okresu podstawowego, do adaptacyjnej interpolacji sygnału na podstawie wyników PEAQ: podejście wykorzystujące rekonstrukcję lewostronną (GT_f), podejście wykorzystujące rekonstrukcję prawostronną (GT_b) oraz podejście wykorzystujące rekonstrukcję dwukierunkową (GT_{fb}). Rekonstrukcja sygnału przeprowadzona została przy pełnej wiedzy o położeniu zakłóceń impulsowych. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Śr. oznacza uśredniony wynik dla 10 wyszczególnionych nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

muzyka jazzowa

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,34	-0,29	-0,53	-0,21	-0,26	-0,24	-0,28	-0,28	-0,26	-0,34	-0,30
GT_b	-0,33	-0,37	-0,54	-0,21	-0,26	-0,22	-0,24	-0,28	-0,26	-0,33	-0,30
GT_{fb}	-0,29	-0,26	-0,48	-0,19	-0,23	-0,20	-0,21	-0,25	-0,23	-0,30	-0,26

muzyka chóralna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,33	-0,14	-0,22	-0,34	-0,26	-0,15	-0,20	-0,21	-0,39	-0,18	-0,24
GT_b	-0,32	-0,14	-0,22	-0,35	-0,23	-0,16	-0,19	-0,21	-0,41	-0,18	-0,24
GT_{fb}	-0,28	-0,12	-0,19	-0,32	-0,21	-0,14	-0,17	-0,19	-0,35	-0,16	-0,21

muzyka operowa

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,59	-0,22	-0,34	-0,29	-0,27	-0,25	-0,36	-0,28	-0,29	-0,33	-0,32
GT_b	-0,56	-0,22	-0,35	-0,29	-0,27	-0,26	-0,42	-0,27	-0,29	-0,30	-0,32
GT_{fb}	-0,51	-0,19	-0,28	-0,23	-0,22	-0,21	-0,33	-0,23	-0,23	-0,28	-0,27

muzyka instrumentalna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,62	-0,32	-0,31	-0,58	-0,23	-0,41	-0,23	-0,27	-0,20	-0,19	-0,33
GT_b	-0,61	-0,33	-0,29	-0,57	-0,23	-0,43	-0,24	-0,25	-0,21	-0,17	-0,33
GT_{fb}	-0,56	-0,29	-0,27	-0,53	-0,21	-0,38	-0,20	-0,24	-0,18	-0,16	-0,30

muzyka klasyczna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,38	-0,25	-0,12	-0,13	-0,32	-0,21	-0,11	-0,29	-0,20	-0,25	-0,23
GT_b	-0,38	-0,24	-0,14	-0,11	-0,32	-0,22	-0,12	-0,28	-0,19	-0,28	-0,23
GT_{fb}	-0,34	-0,20	-0,11	-0,10	-0,29	-0,19	-0,10	-0,22	-0,18	-0,24	-0,20

Tablica C.7: Porównanie skuteczności działania 3 podejść, opartych na modelu SAR wykorzystującym nieprzyczynową ocenę okresu podstawowego, do adaptacyjnej interpolacji sygnału na podstawie wyników PEAQ: podejście wykorzystujące rekonstrukcję lewostronną (GT_f), podejście wykorzystujące rekonstrukcję prawostronną (GT_b) oraz podejście wykorzystujące rekonstrukcję dwukierunkową (GT_{fb}). Rekonstrukcja sygnału przeprowadzona została przy pełnej wiedzy o położeniu zakłóceń impulsowych. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Śr. oznacza uśredniony wynik dla 10 wyszczególnionych nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

muzyka jazzowa

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,36	-0,30	-0,58	-0,21	-0,23	-0,25	-0,43	-0,28	-0,29	-0,41	-0,33
GT_b	-0,37	-0,35	-0,60	-0,21	-0,24	-0,23	-0,42	-0,28	-0,29	-0,42	-0,34
GT_{fb}	-0,35	-0,27	-0,56	-0,20	-0,21	-0,23	-0,40	-0,26	-0,27	-0,36	-0,31

muzyka chóralna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,32	-0,13	-0,20	-0,32	-0,25	-0,13	-0,19	-0,20	-0,40	-0,18	-0,23
GT_b	-0,32	-0,14	-0,21	-0,34	-0,24	-0,13	-0,17	-0,20	-0,39	-0,17	-0,23
GT_{fb}	-0,30	-0,12	-0,19	-0,32	-0,23	-0,13	-0,16	-0,19	-0,38	-0,17	-0,22

muzyka operowa

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,62	-0,25	-0,45	-0,33	-0,35	-0,32	-0,55	-0,32	-0,38	-0,43	-0,40
GT_b	-0,61	-0,25	-0,49	-0,35	-0,35	-0,33	-0,62	-0,32	-0,37	-0,36	-0,40
GT_{fb}	-0,57	-0,23	-0,45	-0,31	-0,32	-0,30	-0,55	-0,30	-0,35	-0,37	-0,38

muzyka instrumentalna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,72	-0,36	-0,31	-0,62	-0,22	-0,44	-0,32	-0,40	-0,25	-0,19	-0,38
GT_b	-0,73	-0,36	-0,31	-0,61	-0,22	-0,46	-0,32	-0,40	-0,23	-0,17	-0,38
GT_{fb}	-0,70	-0,34	-0,29	-0,59	-0,20	-0,43	-0,31	-0,38	-0,23	-0,18	-0,37

muzyka klasyczna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,41	-0,25	-0,12	-0,11	-0,35	-0,21	-0,11	-0,48	-0,18	-0,29	-0,25
GT_b	-0,43	-0,24	-0,13	-0,12	-0,34	-0,21	-0,11	-0,51	-0,19	-0,30	-0,26
GT_{fb}	-0,40	-0,23	-0,11	-0,12	-0,33	-0,20	-0,10	-0,47	-0,17	-0,27	-0,24

Tablica C.8: Porównanie skuteczności działania 3 podejść, opartych na lokalnie wybranym modelu (AR, SAR wykorzystujący przyczynową ocenę okresu podstawowego lub SAR wykorzystujący nieprzyczynową ocenę okresu podstawowego), do adaptacyjnej interpolacji sygnału na podstawie wyników PEAQ: podejście wykorzystujące rekonstrukcję lewostronną (GT_f), podejście wykorzystujące rekonstrukcję prawostronną (GT_b) oraz podejście wykorzystujące rekonstrukcję dwukierunkową (GT_{fb}). Rekonstrukcja sygnału przeprowadzona została przy pełnej wiedzy o położeniu zakłóceń impulsowych. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Śr. oznacza uśredniony wynik dla 10 wyszczególnionych nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

muzyka jazzowa

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,34	-0,29	-0,55	-0,20	-0,24	-0,24	-0,32	-0,29	-0,27	-0,34	-0,31
GT_b	-0,33	-0,30	-0,55	-0,20	-0,26	-0,21	-0,25	-0,27	-0,25	-0,31	-0,29
GT_{fb}	-0,31	-0,26	-0,50	-0,19	-0,23	-0,20	-0,23	-0,25	-0,23	-0,29	-0,27

muzyka chóralna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,32	-0,13	-0,21	-0,33	-0,25	-0,14	-0,19	-0,20	-0,39	-0,17	-0,24
GT_b	-0,32	-0,15	-0,22	-0,34	-0,22	-0,15	-0,18	-0,21	-0,41	-0,19	-0,24
GT_{fb}	-0,28	-0,12	-0,19	-0,32	-0,21	-0,14	-0,16	-0,19	-0,36	-0,16	-0,21

muzyka operowa

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,60	-0,22	-0,35	-0,29	-0,29	-0,26	-0,46	-0,27	-0,31	-0,36	-0,34
GT_b	-0,57	-0,22	-0,36	-0,30	-0,30	-0,28	-0,46	-0,27	-0,31	-0,30	-0,34
GT_{fb}	-0,53	-0,20	-0,31	-0,25	-0,25	-0,23	-0,41	-0,23	-0,25	-0,30	-0,30

muzyka instrumentalna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,63	-0,33	-0,30	-0,60	-0,23	-0,42	-0,26	-0,30	-0,20	-0,18	-0,35
GT_b	-0,65	-0,33	-0,30	-0,58	-0,23	-0,44	-0,27	-0,28	-0,20	-0,18	-0,34
GT_{fb}	-0,60	-0,30	-0,28	-0,55	-0,21	-0,40	-0,23	-0,26	-0,19	-0,17	-0,32

muzyka klasyczna

podejście	1	2	3	4	5	6	7	8	9	10	Śr.
GT_f	-0,38	-0,26	-0,11	-0,12	-0,33	-0,21	-0,11	-0,33	-0,20	-0,27	-0,23
GT_b	-0,42	-0,25	-0,15	-0,12	-0,34	-0,21	-0,12	-0,32	-0,19	-0,29	-0,24
GT_{fb}	-0,36	-0,21	-0,11	-0,11	-0,30	-0,19	-0,10	-0,26	-0,17	-0,26	-0,21

Tablica C.9: Porównanie skuteczności działania 4 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR (A/A*), podejście oparte na modelu SAR z przyczynowym oszacowaniem okresu podstawowego (B/B*), podejście oparte na modelu SAR z nieprzyczynowym oszacowaniem okresu podstawowego (C/C*) oraz proponowane podejście oparte na lokalnie wybieranym modelu (D/D*). We wszystkich 4 podejściach użyto detektora pracującego w pętli z decyzyjnym sprzężeniem zwrotnym. W tabeli zamieszczono wyniki porównania 4 podejść dla trzech typowych wartości współczynnika średniego odchylenia standardowego $\mu \in \{3, 5, 4, 4, 5\}$. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Śr. oznacza uśredniony wynik dla wyszczególnionych 10 nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (znieszczenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

muzyka jazzowa (analiza jednokierunkowa)

Lp.	$\mu = 3, 5$				$\mu = 4$				$\mu = 4, 5$			
	A	B	C	D	A	B	C	D	A	B	C	D
1	-1,94	-1,54	-1,83	-1,73	-1,18	-1,48	-1,07	-1,00	-0,96	-0,93	-0,87	-0,90
2	-1,21	-1,35	-1,17	-1,10	-1,03	-1,16	-1,05	-1,11	-0,94	-1,02	-0,90	-0,92
3	-1,76	-1,72	-1,74	-1,55	-1,55	-1,46	-1,47	-1,37	-1,42	-1,48	-1,36	-1,37
4	-0,55	-0,55	-0,56	-0,52	-0,45	-0,44	-0,45	-0,42	-0,41	-0,41	-0,40	-0,41
5	-0,87	-1,11	-0,85	-0,83	-0,78	-0,93	-0,78	-0,83	-0,73	-0,84	-0,74	-0,74
6	-2,37	-1,92	-1,73	-1,02	-1,60	-1,27	-1,01	-0,74	-1,05	-0,89	-0,73	-0,62
7	-2,62	-2,17	-2,36	-1,24	-1,38	-1,19	-1,36	-0,86	-1,00	-0,87	-0,98	-0,77
8	-1,05	-0,99	-0,97	-0,84	-0,90	-0,85	-0,77	-0,73	-0,71	-0,72	-0,64	-0,63
9	-2,67	-2,57	-2,42	-2,08	-2,20	-2,09	-1,85	-1,61	-1,76	-1,45	-1,12	-1,04
10	-3,34	-3,31	-3,34	-3,07	-2,78	-2,76	-2,75	-2,36	-2,18	-2,15	-2,06	-1,90
Śr.	-1,84	-1,72	-1,70	-1,40	-1,39	-1,36	-1,26	-1,10	-1,12	-1,08	-0,98	-0,93

muzyka jazzowa (analiza dwukierunkowa)

Lp.	$\mu = 3, 5$				$\mu = 4$				$\mu = 4, 5$			
	A*	B*	C*	D*	A*	B*	C*	D*	A*	B*	C*	D*
1	-1,29	-1,04	-1,34	-1,12	-0,73	-0,94	-0,75	-0,65	-0,51	-0,56	-0,58	-0,53
2	-0,91	-0,83	-0,84	-0,77	-0,72	-0,70	-0,68	-0,79	-0,66	-0,63	-0,62	-0,61
3	-1,28	-1,08	-1,08	-0,89	-1,02	-0,85	-0,90	-0,79	-0,97	-0,82	-0,83	-0,80
4	-0,42	-0,40	-0,41	-0,38	-0,35	-0,34	-0,34	-0,32	-0,31	-0,31	-0,31	-0,31
5	-0,81	-0,80	-0,82	-0,77	-0,62	-0,67	-0,68	-0,72	-0,53	-0,63	-0,66	-0,62
6	-2,39	-1,38	-1,36	-0,76	-1,43	-0,76	-0,75	-0,53	-0,93	-0,56	-0,53	-0,43
7	-2,52	-1,34	-1,48	-0,64	-1,25	-0,69	-0,87	-0,45	-0,79	-0,44	-0,60	-0,39
8	-0,87	-0,75	-0,72	-0,60	-0,67	-0,58	-0,55	-0,50	-0,56	-0,49	-0,48	-0,44
9	-1,89	-1,45	-1,35	-0,90	-1,24	-0,90	-0,79	-0,86	-0,83	-0,63	-0,57	-0,83
10	-3,18	-2,87	-2,95	-2,32	-2,37	-1,98	-2,01	-1,55	-1,61	-1,50	-1,52	-1,28
Śr.	-1,56	-1,20	-1,24	-0,91	-1,04	-0,84	-0,83	-0,72	-0,77	-0,66	-0,67	-0,63

muzyka chóralna (analiza jednokierunkowa)

Lp.	$\mu = 3,5$				$\mu = 4$				$\mu = 4,5$			
	A	B	C	D	A	B	C	D	A	B	C	D
1	-1,35	-1,34	-2,44	-1,45	-1,26	-1,30	-1,24	-1,66	-2,51	-2,53	-2,45	-2,55
2	-0,45	-0,46	-0,43	-0,45	-0,44	-0,45	-0,42	-0,45	-0,48	-0,49	-0,46	-0,50
3	-0,46	-0,48	-0,45	-0,47	-0,43	-0,44	-0,43	-0,45	-0,44	-0,46	-0,42	-0,46
4	-1,33	-1,35	-1,55	-1,63	-1,38	-1,46	-1,33	-1,46	-1,55	-1,60	-1,47	-1,69
5	-1,97	-1,78	-1,98	-1,55	-1,42	-1,34	-1,33	-1,24	-1,15	-1,23	-1,10	-1,08
6	-0,81	-0,82	-0,80	-0,80	-0,78	-0,81	-0,77	-0,84	-0,84	-0,89	-0,80	-0,91
7	-0,54	-0,57	-0,54	-0,56	-0,53	-0,55	-0,52	-0,57	-0,58	-0,63	-0,56	-0,64
8	-0,44	-0,45	-0,44	-0,45	-0,43	-0,45	-0,42	-0,45	-0,47	-0,50	-0,45	-0,50
9	-1,24	-1,19	-1,28	-1,22	-1,21	-1,41	-1,18	-1,22	-1,27	-1,27	-1,23	-1,31
10	-0,42	-0,45	-0,46	-0,43	-0,44	-0,44	-0,40	-0,45	-0,50	-0,51	-0,47	-0,52
Śr.	-0,90	-0,89	-1,04	-0,90	-0,83	-0,87	-0,80	-0,88	-0,98	-1,01	-0,94	-1,01

muzyka chóralna (analiza dwukierunkowa)

Lp.	$\mu = 3,5$				$\mu = 4$				$\mu = 4,5$			
	A*	B*	C*	D*	A*	B*	C*	D*	A*	B*	C*	D*
1	-0,71	-0,67	-1,07	-0,62	-0,53	-0,51	-0,69	-0,51	-0,54	-0,79	-1,05	-0,80
2	-0,31	-0,30	-0,30	-0,25	-0,22	-0,22	-0,21	-0,21	-0,20	-0,20	-0,20	-0,20
3	-0,36	-0,35	-0,35	-0,31	-0,24	-0,26	-0,24	-0,23	-0,22	-0,23	-0,22	-0,23
4	-0,70	-0,67	-0,71	-0,68	-0,56	-0,54	-0,52	-0,55	-0,57	-0,55	-0,55	-0,57
5	-1,63	-1,51	-1,72	-1,15	-1,25	-1,09	-1,02	-0,84	-1,01	-0,82	-0,76	-0,65
6	-0,42	-0,41	-0,42	-0,36	-0,31	-0,30	-0,32	-0,30	-0,27	-0,26	-0,27	-0,26
7	-0,38	-0,36	-0,36	-0,30	-0,27	-0,26	-0,26	-0,25	-0,25	-0,25	-0,24	-0,24
8	-0,36	-0,34	-0,36	-0,29	-0,23	-0,24	-0,24	-0,22	-0,22	-0,23	-0,21	-0,22
9	-0,74	-0,68	-0,75	-0,60	-0,55	-0,51	-0,55	-0,49	-0,50	-0,47	-0,49	-0,47
10	-0,35	-0,34	-0,33	-0,29	-0,27	-0,24	-0,25	-0,22	-0,24	-0,24	-0,24	-0,23
Śr.	-0,60	-0,56	-0,64	-0,48	-0,44	-0,42	-0,43	-0,38	-0,40	-0,40	-0,42	-0,39

muzyka operowa (analiza jednokierunkowa)

Lp.	$\mu = 3,5$				$\mu = 4$				$\mu = 4,5$			
	A	B	C	D	A	B	C	D	A	B	C	D
1	-1,36	-1,21	-1,35	-1,32	-1,27	-1,13	-1,32	-1,18	-1,18	-1,03	-1,15	-1,18
2	-0,84	-0,76	-0,89	-0,72	-0,76	-0,68	-0,74	-0,74	-0,78	-0,80	-0,74	-0,82
3	-1,26	-1,13	-1,35	-1,17	-1,19	-1,13	-1,20	-1,19	-1,17	-1,11	-1,18	-1,23
4	-1,23	-1,10	-1,23	-1,13	-1,10	-0,96	-1,06	-1,04	-1,13	-1,06	-1,08	-1,11
5	-1,24	-1,13	-1,25	-1,20	-1,06	-0,97	-1,06	-0,99	-1,10	-0,96	-1,02	-1,04
6	-1,75	-1,68	-1,83	-1,00	-1,00	-1,64	-1,06	-1,68	-0,98	-0,97	-0,99	-1,04
7	-2,06	-1,70	-2,09	-1,82	-1,82	-1,52	-1,91	-1,72	-1,81	-1,60	-1,79	-1,75
8	-0,84	-0,77	-0,86	-0,78	-0,77	-0,73	-0,78	-0,76	-0,85	-0,80	-0,79	-0,84
9	-1,33	-1,09	-1,34	-1,19	-1,32	-1,12	-1,24	-1,25	-1,37	-1,19	-1,31	-1,31
10	-1,19	-1,08	-1,27	-1,10	-1,11	-1,00	-1,07	-1,02	-1,23	-1,10	-1,14	-1,17
Śr.	-1,31	-1,16	-1,35	-1,14	-1,14	-1,09	-1,14	-1,16	-1,16	-1,06	-1,12	-1,15

muzyka operowa (analiza dwukierunkowa)

Lp.	$\mu = 3,5$				$\mu = 4$				$\mu = 4,5$			
	A*	B*	C*	D*	A*	B*	C*	D*	A*	B*	C*	D*
1	-0,85	-0,66	-0,81	-0,61	-0,68	-0,53	-0,67	-0,55	-0,60	-0,46	-0,62	-0,52
2	-0,55	-0,46	-0,56	-0,40	-0,44	-0,30	-0,40	-0,36	-0,39	-0,30	-0,37	-0,31
3	-0,81	-0,62	-0,83	-0,58	-0,58	-0,45	-0,60	-0,45	-0,51	-0,38	-0,52	-0,42
4	-0,69	-0,57	-0,75	-0,51	-0,51	-0,40	-0,50	-0,39	-0,43	-0,35	-0,41	-0,37
5	-0,67	-0,49	-0,67	-0,44	-0,48	-0,34	-0,46	-0,37	-0,43	-0,32	-0,41	-0,35
6	-0,71	-0,55	-0,70	-0,49	-0,48	-0,37	-0,57	-0,38	-0,44	-0,34	-0,41	-0,37
7	-1,39	-0,83	-1,41	-0,87	-0,99	-0,71	-1,12	-0,75	-0,92	-0,67	-0,91	-0,72
8	-0,60	-0,49	-0,62	-0,41	-0,45	-0,32	-0,42	-0,33	-0,39	-0,31	-0,38	-0,32
9	-0,73	-0,52	-0,74	-0,45	-0,53	-0,39	-0,52	-0,38	-0,49	-0,36	-0,48	-0,37
10	-0,68	-0,52	-0,66	-0,47	-0,53	-0,43	-0,51	-0,44	-0,48	-0,40	-0,48	-0,42
Śr.	-0,77	-0,57	-0,77	-0,52	-0,57	-0,42	-0,58	-0,44	-0,51	-0,39	-0,50	-0,42

muzyka instrumentalna (analiza jednokierunkowa)

Lp.	$\mu = 3,5$				$\mu = 4$				$\mu = 4,5$			
	A	B	C	D	A	B	C	D	A	B	C	D
1	-1,11	-0,95	-1,08	-1,02	-0,97	-0,93	-1,01	-0,99	-1,03	-0,94	-1,01	-0,98
2	-0,52	-0,48	-0,53	-0,49	-0,49	-0,46	-0,50	-0,49	-0,50	-0,48	-0,49	-0,51
3	-0,47	-0,46	-0,49	-0,44	-0,41	-0,41	-0,41	-0,41	-0,43	-0,41	-0,42	-0,41
4	-0,78	-0,73	-0,76	-0,73	-0,72	-0,67	-0,73	-0,68	-0,69	-0,69	-0,71	-0,70
5	-0,42	-0,40	-0,42	-0,39	-0,34	-0,34	-0,34	-0,34	-0,32	-0,32	-0,31	-0,32
6	-0,60	-0,55	-0,61	-0,56	-0,56	-0,53	-0,56	-0,54	-0,54	-0,53	-0,55	-0,54
7	-1,36	-1,14	-1,65	-1,22	-1,48	-1,21	-1,31	-1,32	-1,34	-1,22	-1,34	-1,28
8	-1,58	-1,19	-1,70	-1,30	-1,36	-1,06	-1,39	-1,19	-1,32	-1,07	-1,26	-1,19
9	-1,63	-1,43	-1,64	-1,31	-1,08	-1,02	-1,09	-1,03	-0,96	-0,93	-0,96	-0,95
10	-1,24	-1,30	-1,37	-1,12	-1,04	-1,04	-1,05	-1,05	-1,01	-1,00	-0,96	-1,02
Śr.	-0,97	-0,86	-1,03	-0,86	-0,85	-0,77	-0,84	-0,80	-0,81	-0,76	-0,80	-0,79

muzyka instrumentalna (analiza dwukierunkowa)

Lp.	$\mu = 3,5$				$\mu = 4$				$\mu = 4,5$			
	A*	B*	C*	D*	A*	B*	C*	D*	A*	B*	C*	D*
1	-0,91	-0,69	-0,88	-0,66	-0,76	-0,58	-0,74	-0,61	-0,71	-0,56	-0,70	-0,58
2	-0,45	-0,37	-0,42	-0,34	-0,35	-0,30	-0,34	-0,30	-0,32	-0,28	-0,31	-0,29
3	-0,38	-0,35	-0,36	-0,30	-0,31	-0,27	-0,30	-0,27	-0,28	-0,25	-0,28	-0,26
4	-0,65	-0,54	-0,64	-0,52	-0,55	-0,46	-0,55	-0,47	-0,49	-0,43	-0,51	-0,43
5	-0,29	-0,28	-0,29	-0,26	-0,23	-0,21	-0,23	-0,21	-0,21	-0,19	-0,20	-0,19
6	-0,49	-0,44	-0,49	-0,40	-0,39	-0,34	-0,39	-0,34	-0,36	-0,31	-0,36	-0,32
7	-1,00	-0,72	-1,00	-0,69	-0,68	-0,74	-0,67	-0,72	-0,56	-0,46	-0,62	-0,48
8	-1,21	-0,84	-1,50	-0,77	-0,76	-0,52	-0,82	-0,52	-0,59	-0,43	-0,66	-0,44
9	-1,42	-1,22	-1,46	-1,01	-0,78	-0,71	-0,79	-0,64	-0,60	-0,54	-0,59	-0,52
10	-0,89	-0,83	-1,02	-0,72	-0,62	-0,74	-0,64	-0,58	-0,54	-0,55	-0,54	-0,54
Śr.	-0,77	-0,63	-0,81	-0,57	-0,54	-0,49	-0,55	-0,47	-0,47	-0,40	-0,48	-0,41

muzyka klasyczna (analiza jednokierunkowa)

Lp.	$\mu = 3,5$				$\mu = 4$				$\mu = 4,5$			
	A	B	C	D	A	B	C	D	A	B	C	D
1	-1,11	-1,03	-1,14	-1,08	-1,02	-0,97	-1,01	-1,01	-1,06	-1,04	-1,03	-1,08
2	-0,47	-0,47	-0,47	-0,48	-0,45	-0,44	-0,45	-0,45	-0,47	-0,48	-0,46	-0,50
3	-1,84	-1,81	-1,87	-1,82	-0,83	-0,82	-1,79	-0,86	-0,89	-0,94	-0,88	-0,94
4	-1,36	-1,25	-1,43	-1,36	-1,16	-1,04	-1,18	-1,00	-1,07	-1,03	-1,04	-1,06
5	-0,79	-0,76	-0,82	-0,77	-0,78	-0,76	-0,77	-0,78	-0,83	-0,81	-0,79	-0,83
6	-0,40	-0,39	-0,40	-0,39	-0,37	-0,39	-0,37	-0,39	-0,42	-0,41	-0,43	-0,43
7	-0,30	-0,31	-0,31	-0,33	-0,28	-0,30	-0,27	-0,29	-0,28	-0,30	-0,28	-0,30
8	-2,20	-1,76	-2,32	-1,94	-1,71	-1,37	-1,59	-1,40	-1,62	-1,36	-1,55	-1,45
9	-0,47	-0,49	-0,48	-0,46	-0,41	-0,42	-0,39	-0,40	-0,39	-0,41	-0,37	-0,39
10	-0,79	-0,73	-0,77	-0,68	-0,68	-0,61	-0,64	-0,62	-0,69	-0,65	-0,68	-0,68
Śr.	-0,97	-0,90	-1,00	-0,93	-0,77	-0,71	-0,85	-0,72	-0,77	-0,74	-0,75	-0,77

muzyka klasyczna (analiza dwukierunkowa)

Lp.	$\mu = 3,5$				$\mu = 4$				$\mu = 4,5$			
	A*	B*	C*	D*	A*	B*	C*	D*	A*	B*	C*	D*
1	-0,65	-0,53	-0,64	-0,46	-0,48	-0,39	-0,48	-0,39	-0,42	-0,35	-0,42	-0,37
2	-0,40	-0,35	-0,38	-0,31	-0,28	-0,24	-0,28	-0,23	-0,25	-0,22	-0,24	-0,21
3	-0,45	-0,69	-0,77	-0,68	-0,33	-0,35	-0,61	-0,35	-0,37	-0,34	-0,36	-0,33
4	-0,69	-0,92	-1,08	-0,86	-0,40	-0,39	-0,68	-0,33	-0,29	-0,29	-0,35	-0,28
5	-0,57	-0,50	-0,56	-0,46	-0,47	-0,44	-0,47	-0,52	-0,42	-0,44	-0,55	-0,51
6	-0,35	-0,34	-0,32	-0,28	-0,24	-0,23	-0,22	-0,21	-0,21	-0,21	-0,21	-0,21
7	-0,26	-0,24	-0,25	-0,21	-0,18	-0,17	-0,17	-0,15	-0,13	-0,14	-0,14	-0,14
8	-1,64	-1,02	-1,35	-0,71	-1,06	-0,61	-0,91	-0,49	-0,81	-0,44	-0,76	-0,49
9	-0,42	-0,45	-0,42	-0,39	-0,34	-0,34	-0,34	-0,33	-0,29	-0,26	-0,29	-0,26
10	-0,52	-0,44	-0,51	-0,40	-0,41	-0,34	-0,39	-0,30	-0,35	-0,28	-0,35	-0,28
Śr.	-0,60	-0,55	-0,63	-0,48	-0,42	-0,35	-0,45	-0,33	-0,35	-0,30	-0,37	-0,31

Tablica C.10: Porównanie skuteczności działania 4 podejść wykorzystujących adaptacyjną interpolację na podstawie otrzymanych wyników PEAQ: adaptacyjna interpolacja oparta na modelu AR, adaptacyjna interpolacja oparta na modelu SAR, adaptacyjna interpolacja oparta na modelu VAR oraz adaptacyjna interpolacja oparta na modelu SVAR. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Rekonstrukcja nagrań została przeprowadzona przy pełnej wiedzy o położeniu zakłóceń impulsowych. Śr. oznacza uśrednione wyniki w grupie 10 wyszczególnionych nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

muzyka jazzowa

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,38	-0,31	-0,71	-0,22	-0,24	-0,27	-0,46	-0,30	-0,34	-0,39	-0,36
SAR	-0,34	-0,29	-0,55	-0,20	-0,24	-0,24	-0,32	-0,29	-0,27	-0,34	-0,31
VAR	-0,18	-0,23	-0,37	-0,14	-0,17	-0,31	-0,17	-0,20	-0,21	-0,20	-0,22
SVAR	-0,20	-0,25	-0,29	-0,14	-0,19	-0,31	-0,15	-0,21	-0,19	-0,19	-0,21

muzyka chóralna

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,32	-0,12	-0,21	-0,33	-0,26	-0,13	-0,19	-0,19	-0,40	-0,19	-0,23
SAR	-0,32	-0,13	-0,21	-0,33	-0,25	-0,14	-0,19	-0,20	-0,39	-0,17	-0,24
VAR	-0,26	-0,06	-0,14	-0,24	-0,23	-0,11	-0,13	-0,12	-0,29	-0,11	-0,17
SVAR	-0,26	-0,07	-0,15	-0,25	-0,21	-0,14	-0,13	-0,13	-0,30	-0,11	-0,17

muzyka operowa

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,65	-0,26	-0,46	-0,34	-0,37	-0,32	-0,58	-0,33	-0,39	-0,44	-0,41
SAR	-0,60	-0,22	-0,35	-0,29	-0,29	-0,26	-0,46	-0,27	-0,31	-0,36	-0,34
VAR	-0,53	-0,20	-0,27	-0,24	-0,21	-0,23	-0,28	-0,23	-0,26	-0,26	-0,27
SVAR	-0,53	-0,19	-0,26	-0,23	-0,19	-0,21	-0,27	-0,22	-0,24	-0,25	-0,26

muzyka instrumentalna

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,76	-0,39	-0,32	-0,65	-0,23	-0,47	-0,33	-0,40	-0,25	-0,19	-0,40
SAR	-0,63	-0,33	-0,30	-0,60	-0,23	-0,42	-0,26	-0,30	-0,20	-0,18	-0,35
VAR	-0,47	-0,27	-0,15	-0,44	-0,13	-0,28	-0,19	-0,24	0,00	-0,11	-0,23
SVAR	-0,43	-0,24	-0,17	-0,41	-0,15	-0,26	-0,17	-0,20	0,01	-0,12	-0,21

muzyka klasyczna

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,43	-0,26	-0,12	-0,12	-0,35	-0,21	-0,11	-0,52	-0,19	-0,29	-0,26
SAR	-0,38	-0,26	-0,11	-0,12	-0,33	-0,21	-0,11	-0,33	-0,20	-0,27	-0,23
VAR	-0,36	-0,22	-0,08	-0,10	-0,27	-0,20	-0,10	-0,40	-0,15	-0,21	-0,21
SVAR	-0,34	-0,22	-0,08	-0,09	-0,28	-0,21	-0,11	-0,28	-0,17	-0,22	-0,20

Tablica C.11: Porównanie skuteczności działania 4 podejść wykorzystujących adaptacyjną interpolację na podstawie otrzymanych wyników PEAQ: adaptacyjna interpolacja oparta na modelu AR, adaptacyjna interpolacja oparta na modelu SAR, adaptacyjna interpolacja oparta na modelu VAR oraz adaptacyjna interpolacja oparta na modelu SVAR. Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z testowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. Rekonstrukcja nagrań została przeprowadzona przy pełnej wiedzy o położeniu zakłóceń impulsowych. Śr. oznacza uśrednione wyniki w grupie 10 wyszczególnionych nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

testowa baza nagrań

muzyka jazzowa

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,38	-0,32	-0,71	-0,23	-0,29	-0,28	-0,47	-0,33	-0,34	-0,43	-0,38
SAR	-0,35	-0,31	-0,53	-0,21	-0,30	-0,25	-0,30	-0,30	-0,27	-0,36	-0,32
VAR	-0,17	-0,22	-0,36	-0,16	-0,25	-0,19	-0,22	-0,21	-0,19	-0,27	-0,22
SVAR	-0,15	-0,22	-0,30	-0,15	-0,25	-0,17	-0,17	-0,20	-0,17	-0,27	-0,21

muzyka chóralna

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,41	-0,15	-0,20	-0,33	-0,31	-0,18	-0,20	-0,18	-0,42	-0,14	-0,25
SAR	-0,39	-0,15	-0,22	-0,35	-0,30	-0,19	-0,19	-0,20	-0,41	-0,14	-0,25
VAR	-0,31	-0,08	-0,14	-0,24	-0,29	-0,16	-0,14	-0,12	-0,29	-0,08	-0,18
SVAR	-0,31	-0,08	-0,14	-0,25	-0,29	-0,17	-0,13	-0,13	-0,28	-0,06	-0,18

muzyka operowa

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,63	-0,29	-0,48	-0,37	-0,37	-0,32	-0,52	-0,52	-0,46	-0,38	-0,43
SAR	-0,58	-0,26	-0,40	-0,31	-0,31	-0,27	-0,40	-0,42	-0,35	-0,30	-0,36
VAR	-0,42	-0,23	-0,31	-0,26	-0,23	-0,20	-0,31	-0,31	-0,30	-0,26	-0,28
SVAR	-0,38	-0,22	-0,28	-0,24	-0,20	-0,18	-0,28	-0,28	-0,25	-0,25	-0,26

muzyka instrumentalna

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,83	-0,40	-0,32	-0,62	-0,21	-0,49	-0,41	-0,41	-0,24	-0,20	-0,41
SAR	-0,69	-0,35	-0,29	-0,56	-0,21	-0,47	-0,36	-0,29	-0,23	-0,19	-0,36
VAR	-0,51	-0,24	-0,19	-0,39	-0,12	-0,33	-0,25	-0,26	0,01	0,03	-0,23
SVAR	-0,44	-0,21	-0,18	-0,36	-0,12	-0,30	-0,22	-0,19	0,03	0,06	-0,19

muzyka klasyczna

model	1	2	3	4	5	6	7	8	9	10	Śr.
AR	-0,40	-0,30	-0,14	-0,21	-0,37	-0,27	-0,14	-0,45	-0,17	-0,33	-0,28
SAR	-0,36	-0,27	-0,15	-0,20	-0,34	-0,25	-0,15	-0,32	-0,17	-0,30	-0,25
VAR	-0,35	-0,26	-0,11	-0,14	-0,32	-0,23	-0,14	-0,34	-0,14	-0,26	-0,23
SVAR	-0,35	-0,26	-0,12	-0,13	-0,29	-0,21	-0,14	-0,24	-0,15	-0,24	-0,21

Tablica C.12: Porównanie skuteczności działania 6 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A/A*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B/B*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E*) oraz podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F*). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z treningowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. REF oznacza początkową ocenę PEAQ dla zakłóconego nagrania. Śr. oznacza uśrednione wyniki w grupie 10 wyszczególnionych nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

treningowa baza nagrań

muzyka jazzowa

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,61	-2,32	-3,63	-0,96	-0,90	-0,82	-0,86	-1,73	-0,80	-0,51	-0,53	-0,38	-0,36
2	-3,48	-1,31	-1,71	-0,94	-0,92	-0,87	-0,85	-1,25	-1,08	-0,66	-0,61	-0,59	-0,61
3	-2,25	-2,09	-3,45	-1,42	-1,37	-1,19	-1,28	-1,66	-1,09	-0,97	-0,80	-0,80	-0,80
4	-3,22	-0,64	-1,18	-0,41	-0,41	-0,40	-0,40	-0,51	-0,34	-0,31	-0,31	-0,30	-0,31
5	-3,54	-1,51	-2,81	-0,73	-0,74	-0,70	-0,70	-1,09	-1,02	-0,53	-0,62	-0,54	-0,53
6	-2,80	-1,23	-1,72	-1,05	-0,62	-0,68	-0,56	-1,24	-1,12	-0,93	-0,43	-0,57	-0,40
7	-3,46	-1,09	-3,33	-1,00	-0,77	-0,66	-0,54	-1,60	-0,90	-0,79	-0,39	-0,51	-0,35
8	-3,11	-0,83	-1,22	-0,71	-0,63	-0,70	-0,65	-0,73	-0,57	-0,56	-0,44	-0,53	-0,46
9	-3,35	-1,26	-1,87	-1,76	-1,04	-1,74	-1,31	-1,09	-0,88	-0,83	-0,83	-0,82	-0,55
10	-3,77	-2,21	-3,55	-2,18	-1,90	-2,11	-1,82	-1,97	-1,74	-1,61	-1,28	-1,46	-1,24
Śr.	-3,26	-1,45	-2,45	-1,12	-0,93	-0,99	-0,90	-1,29	-0,95	-0,77	-0,63	-0,65	-0,56

muzyka chóralna

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,91	-3,67	-3,84	-2,51	-2,55	-3,09	-3,14	-3,46	-1,07	-0,54	-0,80	-0,51	-0,48
2	-3,78	-2,22	-3,58	-0,48	-0,50	-0,40	-0,43	-1,52	-0,34	-0,20	-0,20	-0,16	-0,17
3	-3,76	-1,82	-3,51	-0,44	-0,46	-0,40	-0,43	-1,34	-0,79	-0,22	-0,23	-0,17	-0,18
4	-3,90	-3,59	-3,81	-1,55	-1,69	-1,34	-1,44	-3,18	-2,60	-0,57	-0,57	-0,44	-0,51
5	-3,86	-2,73	-3,73	-1,15	-1,08	-1,14	-1,14	-2,45	-1,12	-1,01	-0,65	-0,95	-0,71
6	-3,91	-3,36	-3,83	-0,84	-0,91	-0,77	-0,85	-2,82	-0,62	-0,27	-0,26	-0,25	-0,27
7	-3,86	-2,37	-3,69	-0,58	-0,64	-0,54	-0,56	-1,71	-0,44	-0,25	-0,24	-0,22	-0,22
8	-3,71	-1,62	-3,37	-0,47	-0,50	-0,41	-0,43	-1,10	-0,43	-0,22	-0,22	-0,18	-0,18
9	-3,89	-3,23	-3,80	-1,27	-1,31	-1,09	-1,14	-2,83	-0,87	-0,50	-0,47	-0,43	-0,45
10	-3,78	-1,76	-3,48	-0,50	-0,52	-0,44	-0,45	-1,19	-0,53	-0,24	-0,23	-0,19	-0,20
Śr.	-3,84	-2,64	-3,66	-0,98	-1,01	-0,96	-1,00	-2,16	-0,88	-0,40	-0,39	-0,35	-0,34

muzyka operowa

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,80	-2,11	-3,63	-1,18	-1,18	-1,09	-1,06	-1,57	-0,84	-0,60	-0,52	-0,52	-0,53
2	-3,83	-2,43	-3,66	-0,78	-0,82	-0,68	-0,71	-1,78	-0,65	-0,39	-0,31	-0,29	-0,29
3	-3,84	-2,55	-3,71	-1,17	-1,23	-0,97	-0,98	-1,99	-1,27	-0,51	-0,42	-0,44	-0,40
4	-3,83	-2,86	-3,75	-1,13	-1,11	-0,98	-1,03	-2,15	-0,64	-0,43	-0,37	-0,35	-0,32
5	-3,87	-2,77	-3,76	-1,10	-1,04	-0,89	-0,91	-2,29	-0,76	-0,43	-0,35	-0,33	-0,32
6	-3,82	-2,36	-3,75	-0,98	-1,04	-0,97	-0,88	-1,80	-0,74	-0,44	-0,37	-0,38	-0,34
7	-3,88	-3,21	-3,79	-1,82	-1,75	-1,36	-1,39	-2,71	-1,48	-0,92	-0,72	-0,56	-0,54
8	-3,73	-2,24	-3,68	-0,85	-0,84	-0,73	-0,77	-1,67	-0,58	-0,39	-0,32	-0,30	-0,28
9	-3,87	-3,10	-3,77	-1,37	-1,31	-1,12	-1,11	-2,55	-1,02	-0,49	-0,37	-0,37	-0,34
10	-3,86	-3,09	-3,77	-1,23	-1,17	-0,98	-0,99	-2,49	-0,98	-0,48	-0,42	-0,38	-0,37
Śr.	-3,83	-2,67	-3,73	-1,16	-1,15	-0,98	-0,98	-2,10	-0,90	-0,51	-0,42	-0,39	-0,37

muzyka instrumentalna

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,68	-1,27	-3,61	-1,03	-0,98	-0,78	-0,77	-1,22	-0,89	-0,71	-0,58	-0,48	-0,44
2	-3,49	-0,88	-2,91	-0,50	-0,51	-0,41	-0,42	-0,81	-0,42	-0,32	-0,29	-0,24	-0,23
3	-3,49	-0,69	-2,71	-0,43	-0,41	-0,34	-0,35	-0,71	-0,34	-0,28	-0,26	-0,17	-0,18
4	-3,51	-0,73	-3,11	-0,69	-0,70	-0,56	-0,57	-0,80	-0,61	-0,49	-0,43	-0,36	-0,33
5	-3,55	-0,76	-2,15	-0,32	-0,32	-0,26	-0,28	-0,64	-0,29	-0,21	-0,19	-0,15	-0,15
6	-3,38	-0,74	-2,91	-0,54	-0,54	-0,45	-0,45	-0,79	-0,45	-0,36	-0,32	-0,25	-0,24
7	-2,91	-2,72	-3,66	-1,34	-1,28	-0,86	-0,87	-2,23	-0,91	-0,56	-0,48	-0,41	-0,37
8	-2,70	-2,67	-3,71	-1,32	-1,19	-0,97	-0,95	-2,34	-1,32	-0,59	-0,44	-0,43	-0,38
9	-2,12	-2,56	-3,64	-0,96	-0,95	-0,85	-0,78	-1,91	-1,14	-0,60	-0,52	-0,54	-0,45
10	-2,91	-2,95	-3,75	-1,01	-1,02	-0,76	-0,79	-2,14	-0,88	-0,54	-0,54	-0,54	-0,52
Śr.	-3,17	-1,60	-3,22	-0,81	-0,79	-0,62	-0,62	-1,36	-0,73	-0,47	-0,41	-0,36	-0,33

muzyka klasyczna

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,82	-2,91	-3,75	-1,06	-1,08	-0,99	-1,03	-2,30	-0,85	-0,42	-0,37	-0,42	-0,39
2	-3,71	-1,60	-3,42	-0,47	-0,50	-0,44	-0,46	-1,12	-0,73	-0,25	-0,21	-0,22	-0,22
3	-3,91	-3,80	-3,83	-0,89	-0,94	-0,73	-0,74	-3,52	-0,95	-0,37	-0,33	-0,27	-0,29
4	-3,90	-3,60	-3,78	-1,07	-1,06	-0,83	-0,83	-3,01	-0,75	-0,29	-0,28	-0,31	-0,28
5	-3,52	-2,37	-3,59	-0,83	-0,83	-0,74	-0,76	-1,93	-0,78	-0,42	-0,51	-0,38	-0,34
6	-3,66	-1,40	-3,24	-0,42	-0,43	-0,39	-0,41	-0,98	-0,33	-0,21	-0,21	-0,19	-0,18
7	-3,55	-1,00	-2,54	-0,28	-0,30	-0,28	-0,34	-0,79	-0,29	-0,13	-0,14	-0,13	-0,15
8	-3,90	-3,43	-3,80	-1,62	-1,45	-1,44	-1,33	-2,98	-1,20	-0,81	-0,49	-0,65	-0,40
9	-2,98	-0,53	-1,28	-0,39	-0,39	-0,37	-0,40	-0,49	-0,34	-0,29	-0,26	-0,30	-0,30
10	-3,81	-1,75	-3,42	-0,69	-0,68	-0,61	-0,65	-1,29	-0,48	-0,35	-0,28	-0,28	-0,29
Śr.	-3,68	-2,24	-3,27	-0,77	-0,77	-0,68	-0,69	-1,84	-0,67	-0,35	-0,31	-0,32	-0,28

Tablica C.13: Porównanie skuteczności działania 6 podejść wykorzystujących analizę jednokierunkową/dwukierunkową na podstawie otrzymanych wyników PEAQ: podejście oparte na modelu AR wykorzystujące detekcję z dwoma progami (A/A*), podejście oparte na modelu AR wykorzystujące detekcję w pętli otwartej (B/B*), podejście oparte na modelu AR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (C/C*), podejście oparte na modelu SAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (D/D*), podejście oparte na modelu VAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (E/E*), podejście oparte na modelu SVAR wykorzystujące detekcję z decyzyjnym sprzężeniem zwrotnym (F/F*). Wszystkie wyniki otrzymane zostały dla 50 spreparowanych sygnałów fonicznych z testowej bazy nagrań: 10 z muzyką jazzową, 10 z muzyką chóralną, 10 z muzyką operową, 10 z muzyką instrumentalną oraz 10 z muzyką klasyczną. REF oznacza początkową ocenę PEAQ dla zakłóconego nagrania. Śr. oznacza uśrednione wyniki w grupie 10 wyszczególnionych nagrań.

Interpretacja wyników PEAQ: 0 = niesłyszalne (zniekształcenia sygnału), -1 = słyszalne, ale nieprzeszkadzające, -2 = lekko przeszkadzające, -3 = przeszkadzające, -4 = bardzo przeszkadzające.

testowa baza nagrań

muzyka jazzowa

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,71	-1,94	-3,55	-1,25	-1,21	-0,98	-1,00	-1,69	-1,12	-0,62	-0,67	-0,61	-0,37
2	-3,57	-1,29	-1,87	-0,87	-0,91	-0,90	-0,90	-1,14	-0,77	-0,62	-0,55	-0,58	-0,60
3	-3,29	-1,76	-3,41	-1,39	-1,42	-1,29	-1,31	-1,83	-1,12	-0,95	-0,73	-0,79	-0,73
4	-3,30	-0,60	-1,08	-0,42	-0,42	-0,41	-0,42	-0,55	-0,36	-0,33	-0,33	-0,31	-0,32
5	-3,75	-1,42	-2,33	-0,73	-0,72	-0,70	-0,69	-1,19	-0,66	-0,62	-0,65	-0,77	-0,56
6	-1,89	-1,23	-1,60	-1,09	-0,64	-0,67	-0,55	-1,28	-0,86	-0,87	-0,45	-0,57	-0,41
7	-3,34	-1,12	-3,36	-1,04	-0,79	-0,69	-0,59	-1,76	-1,00	-0,81	-0,39	-0,56	-0,36
8	-3,32	-0,86	-1,25	-0,73	-0,67	-0,73	-0,66	-0,74	-0,58	-0,56	-0,44	-0,52	-0,45
9	-3,36	-1,30	-1,81	-1,16	-0,78	-1,16	-0,79	-1,14	-0,86	-0,78	-0,46	-0,81	-0,49
10	-3,84	-2,14	-3,53	-2,19	-1,93	-2,14	-2,04	-2,03	-1,77	-1,60	-1,17	-1,52	-1,23
Śr.	-3,34	-1,37	-2,38	-1,09	-0,95	-0,97	-0,89	-1,34	-0,91	-0,78	-0,58	-0,70	-0,55

muzyka chóralna

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,91	-3,67	-3,84	-1,48	-1,53	-1,62	-1,39	-3,55	-2,29	-0,53	-0,51	-0,47	-0,47
2	-3,80	-2,35	-3,61	-0,54	-0,57	-0,44	-0,48	-1,69	-0,52	-0,20	-0,22	-0,44	-0,17
3	-3,86	-2,07	-3,56	-0,50	-0,53	-0,44	-0,47	-1,57	-0,53	-0,22	-0,22	-0,20	-0,20
4	-3,91	-3,47	-3,83	-1,49	-1,56	-1,30	-1,39	-3,26	-2,26	-0,51	-0,50	-0,40	-0,44
5	-3,89	-2,71	-3,69	-1,16	-1,12	-1,16	-1,15	-2,47	-1,23	-0,96	-0,66	-0,96	-0,69
6	-3,91	-3,38	-3,84	-0,94	-1,01	-0,89	-0,98	-2,87	-0,72	-0,29	-0,30	-0,28	-0,28
7	-3,89	-2,45	-3,71	-0,62	-0,66	-0,54	-0,61	-1,78	-0,46	-0,21	-0,22	-0,18	-0,17
8	-3,83	-1,58	-3,42	-0,50	-0,56	-0,45	-0,46	-1,10	-0,36	-0,24	-0,23	-0,20	-0,21
9	-3,91	-3,27	-3,77	-1,33	-1,36	-1,16	-1,20	-2,82	-0,91	-0,55	-0,54	-0,45	-0,46
10	-3,85	-1,82	-3,41	-0,50	-0,54	-0,45	-0,49	-1,17	-0,31	-0,19	-0,20	-0,19	-0,16
Śr.	-3,88	-2,68	-3,67	-0,91	-0,94	-0,85	-0,86	-2,23	-0,96	-0,39	-0,36	-0,38	-0,33

muzyka operowa

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,85	-2,11	-3,63	-1,12	-1,07	-0,91	-0,89	-1,63	-0,88	-0,56	-0,50	-0,41	-0,41
2	-3,88	-2,51	-3,67	-0,78	-0,77	-0,73	-0,71	-2,02	-0,68	-0,33	-0,30	-0,29	-0,28
3	-3,88	-2,67	-3,74	-1,26	-1,21	-0,94	-0,98	-2,28	-1,04	-0,59	-0,53	-0,41	-0,36
4	-3,88	-2,92	-3,75	-1,12	-1,10	-1,14	-1,16	-2,34	-0,73	-0,45	-0,38	-0,38	-0,34
5	-3,89	-2,85	-3,77	-1,18	-1,14	-1,03	-1,04	-2,23	-0,83	-0,48	-0,38	-0,36	-0,33
6	-3,87	-2,46	-3,71	-1,12	-1,10	-1,08	-1,13	-2,02	-0,86	-0,44	-0,50	-0,32	-0,31
7	-3,90	-3,20	-3,81	-1,98	-1,84	-1,57	-1,44	-2,85	-1,52	-0,86	-0,62	-0,70	-0,60
8	-3,87	-2,35	-3,68	-0,98	-0,90	-0,79	-0,81	-1,78	-0,74	-0,50	-0,41	-0,36	-0,31
9	-3,90	-3,17	-3,78	-1,45	-1,43	-1,28	-1,23	-2,66	-1,09	-0,58	-0,44	-0,44	-0,38
10	-3,89	-3,13	-3,78	-1,24	-1,25	-1,08	-1,17	-2,60	-0,90	-0,48	-0,45	-0,39	-0,35
Śr.	-3,88	-2,74	-3,73	-1,22	-1,18	-1,06	-1,06	-2,24	-0,93	-0,53	-0,45	-0,40	-0,37

muzyka instrumentalna

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,82	-1,26	-3,59	-1,04	-1,00	-0,82	-0,81	-1,58	-0,97	-0,68	-0,59	-0,50	-0,46
2	-3,67	-0,88	-3,05	-0,56	-0,55	-0,44	-0,45	-0,91	-0,50	-0,34	-0,27	-0,24	-0,22
3	-3,51	-0,66	-2,42	-0,41	-0,42	-0,37	-0,36	-0,67	-0,37	-0,27	-0,23	-0,19	-0,18
4	-3,60	-0,74	-2,94	-0,73	-0,71	-0,55	-0,56	-0,92	-0,62	-0,48	-0,42	-0,33	-0,31
5	-3,54	-0,71	-2,02	-0,33	-0,33	-0,26	-0,29	-0,63	-0,34	-0,19	-0,18	-0,13	-0,14
6	-3,60	-0,70	-3,17	-0,61	-0,60	-0,46	-0,46	-0,81	-0,54	-0,38	-0,34	-0,26	-0,25
7	-3,69	-2,62	-3,70	-1,26	-1,19	-0,92	-0,91	-2,67	-1,51	-0,64	-0,56	-0,48	-0,43
8	-3,70	-2,60	-3,72	-1,58	-1,41	-1,24	-1,18	-2,90	-1,99	-0,91	-0,65	-0,49	-0,53
9	-3,08	-2,53	-3,66	-1,00	-1,00	-0,88	-0,88	-2,29	-1,65	-0,97	-0,87	-0,54	-0,44
10	-3,67	-2,73	-3,69	-1,02	-1,02	-0,79	-0,81	-2,21	-0,96	-0,59	-0,53	-0,62	-0,53
Śr.	-3,59	-1,54	-3,20	-0,85	-0,82	-0,67	-0,67	-1,56	-0,94	-0,55	-0,47	-0,38	-0,35

muzyka klasyczna

nagranie		analiza jednokierunkowa						analiza dwukierunkowa					
Lp.	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
1	-3,89	-2,92	-3,79	-1,18	-1,17	-1,09	-1,14	-2,65	-1,17	-0,47	-0,43	-0,45	-0,45
2	-3,83	-1,63	-3,51	-0,53	-0,54	-0,50	-0,53	-1,26	-0,48	-0,27	-0,23	-0,28	-0,26
3	-3,91	-3,75	-3,83	-0,96	-0,99	-0,91	-0,89	-3,59	-2,30	-0,33	-0,30	-0,45	-0,32
4	-3,91	-3,53	-3,81	-0,95	-0,97	-0,96	-0,99	-3,23	-2,08	-0,44	-0,57	-0,40	-0,32
5	-3,82	-2,44	-3,65	-0,89	-0,89	-0,78	-0,86	-2,21	-0,89	-0,45	-0,41	-0,38	-0,37
6	-3,80	-1,42	-3,29	-0,44	-0,45	-0,42	-0,45	-1,02	-0,35	-0,21	-0,19	-0,22	-0,19
7	-3,73	-1,04	-2,55	-0,33	-0,35	-0,32	-0,36	-0,79	-0,28	-0,18	-0,18	-0,18	-0,18
8	-3,91	-3,49	-3,82	-1,67	-1,58	-1,42	-1,32	-3,06	-1,36	-0,84	-0,46	-0,64	-0,41
9	-2,50	-0,49	-1,21	-0,41	-0,42	-0,40	-0,41	-0,48	-0,34	-0,30	-0,26	-0,31	-0,31
10	-3,86	-1,77	-3,48	-0,71	-0,71	-0,64	-0,67	-1,36	-0,48	-0,37	-0,29	-0,33	-0,26
Śr.	-3,72	-2,25	-3,29	-0,81	-0,81	-0,74	-0,76	-1,96	-0,97	-0,39	-0,33	-0,36	-0,31

Dodatek D

Zawartość dołączonej płyty CD

Do pracy dołączono płytę CD, na której znajdują się: rozprawa doktorska w wersji elektronicznej, pliki dźwiękowe zapisane w formacie WAV oraz poniższe informacje o nagraniach dźwiękowych wykorzystanych w trzech testach odsłuchowych:

- *Test odsłuchowy 1*

W tym teście wykorzystano 5 czystych fragmentów słuchowiska radiowego pt. „Kochanek dr Luli” w reżyserii Jana Warenyci. Każdy fragment nagrania posiadał około 22 sekundy materiału fonicznego nagranych w systemie stereofonicznym i spróbkowanego z częstotliwością 48 kHz. Wszystkie sygnały testowe spreparowano poprzez dodanie do lewego i prawego kanału tych samych szablonów zakłóceń. W ramach testu odsłuchowego porównano wyniki otrzymane przy użyciu algorytmu opartego na klasycznym modelu AR (algorytm A) z tymi uzyskanymi przy użyciu polecanego w pracy algorytmu opartego na modelu SVAR (algorytm B). Wyniki testu zamieszczono w tabeli 8.23. W katalogu *Test odsłuchowy 1* znajdują się następujące pliki dźwiękowe: oryginalne wersje nagrań oznaczone jako *Nagranie oryginalne X.wav*, spreparowane wersje nagrań oznaczone jako *Nagranie spreparowane X.wav* oraz dwie wersje przetworzenia spreparowanych nagrań, odpowiednio przez algorytm A oraz przez algorytm B oznaczone jako *Rekonstrukcja X - algorytm A.wav* oraz *Rekonstrukcja X - algorytm B.wav*, gdzie *X* oznacza numer porządkowy nagrania.

- *Test odsłuchowy 2*

W tym teście wykorzystano 10 fragmentów rzeczywistych nagrań dźwiękowych odtworzonych z płyt gramofonowych (część z nich została odtworzona z płyty gramofonowej przy użyciu wkładki monofonicznej [3, 4, 6, 8, 9, 10], a część – przy użyciu wkładki stereofonicznej [1, 2, 5, 7]). Fragmenty nagrań zawierają od 13 do 27 sekund materiału fonicznego spróbkowanego z częstotliwością 44,1 kHz. Nagrania te obejmują różne style muzyki: klasyczna (6, 8), operowa (3, 4), popularna (1, 2, 5, 7) i bluesowa (9, 10). W ramach testu odsłuchowego porównano wyniki otrzymane przy użyciu komercyjnego algorytmu the Auto Declick CEDAR (algorytm A) z tymi uzyskanymi przy użyciu polecanego w pracy algorytmu opartego na wektorowym modelu SVAR w przypadku nagrań stereofonicznych lub na skalarnym modelu SAR w przypadku nagrań monofonicznych (algorytm B). Wyniki testu zamieszczono w tabeli 8.24. W katalogu *Test odsłuchowy 2* znajdują się następujące pliki dźwiękowe: oryginalne wersje nagrań oznaczone jako *Nagranie oryginalne X.wav*, oraz dwie wersje przetworzenia archiwalnych nagrań, odpowiednio przez algorytm A oraz przez algorytm B oznaczone jako *Rekonstrukcja X - algorytm A.wav* oraz *Rekonstrukcja X - algorytm B.wav*, gdzie *X* oznacza numer porządkowy nagrania.



- *Test odsłuchowy 3*

W tym teście wykorzystano 8 fragmentów rzeczywistych nagrań dźwiękowych odtworzonych z płyt gramofonowych, spróbkowanych z częstotliwością 48 kHz i zawierających od 24 do 31 sekund materiału fonicznego. Nagrania testowe pochodzą z dwóch różnych źródeł: albumu będącego kompilacją bluesowych utworów Mississippi Delta Blues, w wykonaniu 4 różnych artystów (4 pierwsze nagrania), oraz albumu zawierającego największe arie operowe w wykonaniu Maria del Monaco (cztery ostatnie nagrania). W ramach testu porównano ze sobą dwa podejścia: podejście oparte na detekcji z decyzyjnym sprzężeniem zwrotnym (algorytm A) oraz podejście oparte na detekcji z decyzyjnym sprzężeniem zwrotnym połączone z metodą dopasowywania wzorców (algorytm B). Wyniki testu zamieszczono w tabeli 8.31. W katalogu *Test odsłuchowy 3* znajdują się następujące pliki dźwiękowe: oryginalne wersje nagrań oznaczone jako *Nagranie oryginalne X.wav*, oraz dwie wersje przetworzenia archiwalnych nagrań, odpowiednio przez algorytm A oraz przez algorytm B oznaczone jako *Rekonstrukcja X - algorytm A.wav* oraz *Rekonstrukcja X - algorytm B.wav*, gdzie *X* oznacza numer porządkowy nagrania.

Bibliografia

- [1] M. Niedźwiecki, M. Ciołek, Elimination of impulsive disturbances from archive audio signals using bidirectional processing, *IEEE Trans. Audio, Speech, Language Process.* 21 (5) (2013) 1046–1059.
- [2] M. Niedźwiecki, M. Ciołek, Renovation of archive audio recordings using sparse autoregressive modeling and bidirectional processing, in: *38th International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, 2013, 5949–5953.
- [3] M. Niedźwiecki, M. Ciołek, Elimination of impulsive disturbances from archive audio files – comparison of three noise pulse detection schemes, in: *Forum Acusticum*, Kraków, Polska, 2014, stron 9.
- [4] M. Ciołek, Sparse autoregressive modeling, in: *II Konferencja Studentów i Doktorantów ICT Young*, Gdańsk, Polska, 2012, 499–504.
- [5] M. Niedźwiecki, M. Ciołek, Elimination of clicks from archive speech signals using sparse autoregressive modeling, in: *20th European Signal Processing Conference*, Bucharest, Romania, 2012, 2615–2619.
- [6] M. Niedźwiecki, M. Ciołek, K. Cisowski, Elimination of impulsive disturbances from stereo audio recordings using vector autoregressive modeling and variable-order Kalman filtering, *IEEE Trans. Audio, Speech, Language Process.* 23 (6) (2015) 970–981.
- [7] M. Niedźwiecki, M. Ciołek, Sparse vector autoregressive modeling of audio signals and its application to the elimination of impulsive disturbances, in: *17th IFAC Symposium on System Identification*, Beijing, China, 2015, 1202–1207.
- [8] M. Niedźwiecki, M. Ciołek, Elimination of impulsive disturbances from stereo audio recordings, in: *22th European Signal Processing Conference*, Lisbon, Portugal, 2014, 66–70.
- [9] M. Niedźwiecki, M. Ciołek, Localization of impulsive disturbances in audio signals using template matching, *Digital Signal Processing* 46 (11) (2015) 253–265.
- [10] M. Niedźwiecki, M. Ciołek, Localization of impulsive disturbances in archive audio signals using predictive matched filtering, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, 2014, 2888–2892.
- [11] Diamond Cut Productions, *Diamond Cut Audio Restoration Tools – Users Manual*, 1996.
- [12] W. Heine, A laser scanning phonograph record player, in: *Presented at the 57th Convention of the AES*, 1977, Preprint 1257.



- [13] R. Orban, G. Ogonowski, *Maintaining Audio Quality in the Broadcast Facility – 2008 Edition*, Orban, 2008.
- [14] J. Hayes, Lasers get groovy, *Engineering and Technology* 6 (11) (2011) 58–59.
- [15] V. Fadeyev, C. Haber, Reconstruction of mechanically recorded sound by image processing, *J. of Audio Eng. Soc.* 51 (12) (2003) 1172–1185.
- [16] S. Cavaglieri, O. Johnsen, F. Bapst, Optical retrieval and storage of analog sound recordings, in: *Proc. The AES 20th International Conference*, Budapest, Hungary, 2001, Preprint 1948.
- [17] S. Stotzer, O. Johnsen, F. Bapst, C. Milan, C. Sudan, S. Cavaglieri, P. Pellizzari, Visu-
audio: an optical technique to save the sound of phonographic records, *IASA Journal* 21 (July) (2003) 38–47.
- [18] S. Stotzer, O. Johnsen, F. Bapst, C. Sudan, R. Ingol, Phonographic sound extraction using image and signal processing, in: *Proc. ICASSP*, Montreal, Quebec, Canada, 2004, 289–292.
- [19] J. C. L. Laborelli, Dust detection by colour analysis in an optical method of phonographic discs digitisation, in: *Proc. IEEE International Conference on Image Processing*, 2007, 25–28.
- [20] V. Fadeyev, C. Haber, C. Maul, J. McBride, M. Golden, Reconstruction of recorded sound from an edison cylinder using three-dimensional non-contact optical surface metrology, *J. of Audio Eng. Soc.* 53 (6) (2005) 485–508.
- [21] B. Tian, J. Barron, Using computer vision technology to play gramophone records, *J. Audio Eng. Soc.* 59 (7–8) (2011) 514–538.
- [22] S. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction*, Wiley, 1996.
- [23] J. Godsill, P. Rayner, *Digital Audio Restoration – A Statistical Model Based Approach*, Springer-Verlag, 1998.
- [24] P. Wilson, Record contamination: Causes and cure, *J. Audio Eng. Soc.* 13 (2) (1965) 166–176.
- [25] J. Biliński, Odnawianie starych nagrań. Czy można zrobić to (jeszcze) lepiej?, *Estrada i Studio*, (7) 2004, stron 4.
- [26] J. G. R. Kinzie, D. W. Gravereaux, Automatic detection of impulse noise, *J. Audio Eng. Soc.* 21 (3) (1973) 331–336.
- [27] P. G. Craven, M. A. Gerzon, The elimination of scratch noise from 78 rpm records, in: *Presented at the 50th Convention of the AES*, 1975, Preprint L–37.
- [28] M. J. Carrey, I. Buckner, A system for reducing impulsive noise on gramophone reproduction equipment, *The Radio Electronic Engineer* 50 (7) (1976) 331–336.
- [29] M. R. Sacks, J. M. Bullingham, Audio-disc scratch filter, *Electronics Letters* 12 (25) (1976) 656–657.



- [30] R. S. Burwen, Suppression of low level impulsive noise, in: Presented at the 61st Convention of the AES, New York, USA, 1978, Preprint 1388.
- [31] A. Czyżewski, Dźwięk cyfrowy: wybrane zagadnienia teoretyczne, technologia, zastosowania, Exit, 1998.
- [32] J. Lim, A. Oppenheim, All-pole modeling of degraded speech, *IEEE Trans. Acoust., Speech, Signal Process.* 26 (3) (1978) 197–210.
- [33] D. Ephraim, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, *IEEE Trans. Acoust., Speech, Signal Process.* 32 (6) (1984) 1109–1121.
- [34] S. Böll, Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. Acoust., Speech, Signal Process.* 27 (2) (1979) 113–120.
- [35] R. McAulay, M. Malpass, Speech enhancement using a soft decision noise suppression filter, *IEEE Trans. Acoust., Speech, Signal Process.* 28 (2) (1980) 137–145.
- [36] P. Vary, Noise suppression by spectral amplitude estimation – mechanism and theoretical limits, *Signal Process.* 8 (4) (1985) 387–400.
- [37] D. Rosenthal, H. Okuno, *Computational Auditory Scene Analysis*, Mahwah, NJ, USA: Lawrence Erlbaum, 1998.
- [38] D. Wang, G. Brown, Separation of speech from interfering sounds based on oscillatory correlations, *IEEE Trans. Neural Networks* 10 (3) (1999) 684–697.
- [39] G. Hu, D. Wang, Monaural speech segregation based on pitch tracking and amplitude modulation, *IEEE Trans. Neural Networks* 15 (5), (2004) 1135–1150.
- [40] D. Wang, G. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms and applications*, Wiley, 2006.
- [41] I. Pitas, A. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*, Springer-Verlag, New York, 1990.
- [42] T. Kasparys, J. Lane, Adaptive scratch noise filtering, *IEEE Transactions on Consumer Electronics* 39 (4) (1993) 917–922.
- [43] S. J. Godsill, P. J. W. Rayner, A Bayesian approach to the detection and correction of error bursts in audio signals, in: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1992, 261–264.
- [44] S. Godsill, P. Rayner, A bayesian approach to the restoration of degraded audio signals, *IEEE Trans. Speech, Audio Process.* 3 (4) (1995) 267–278.
- [45] S. Godsill, P. Rayner, Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler, *IEEE Trans. Speech, Audio Process.* 6 (4) (1998) 352–372.
- [46] F. Ávila, L. Biscainho, Bayesian restoration of audio signals degraded by impulsive noise modeled as individual pulses, *IEEE Trans. Audio, Speech Language Process.* 20 (9) (2012) 2470–2481.

- [47] S. V. Vaseghi, P. J. W. Rayner, A new application of adaptive filters for restoration of archived gramophone recordings, in: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 1988, 2548–2551.
- [48] S. Vaseghi, P. Rayner, Detection and suppression of impulsive noise in speech communication systems, Communications, Speech and Vision, IEE Proceedings I. 137 (1) (1990) 38–46.
- [49] S. Vaseghi, R. Frayling-Cork, Restoration of old gramophone recordings, J. Audio Eng. Soc. 40 (10) (1992) 791–801.
- [50] P. Esquef, L. Biscainho, P. Diniz, F. Freeland, A double-threshold-based approach to impulsive noise detection in audio signals, in: Proc. European Signal Process. Conf., Tampere, Finland, 2000, 2041–2044.
- [51] K. Cisowski, Zastosowanie testów statystycznych do detekcji zakłóceń impulsowych w sygnałach fonicznych, in: Proc. IX Poznańskie Warsztaty Telekomunikacyjne, 2004, 131–136.
- [52] K. Cisowski, Detekcja zakłóceń impulsowych w sygnałach fonicznych z wykorzystaniem wygładzonych lokalnych histogramów sygnału. In: Proc. X Poznańskie Warsztaty Telekomunikacyjne, 2005, 109–114.
- [53] K. Cisowski, Dwuetapowa metoda detekcji zakłóceń impulsowych w sygnałach fonicznych, in: Proc. XI Poznańskie Warsztaty Telekomunikacyjne, 2006, 59–62.
- [54] K. Cisowski, Parametric Impulsive Noise Detector for Corrupted Audio Signals Based on Hidden Markov Model, in: Proceedings of the 2008 1st International Conference on Information Technology, 2008, 1–4.
- [55] M. Niedźwiecki, Elimination of clicks and background noise from archive gramophone recordings using the two track mono approach, in: Proc. European Signal Processing Conference, Trieste, Italy, 1996, 1749–1752.
- [56] M. Niedźwiecki, K. Cisowski, Adaptive scheme for elimination of broadband noise and impulsive disturbances from AR and ARMA signals, IEEE Trans. Signal Process. 44 (3) (1996) 528–537.
- [57] M. Niedźwiecki, Identification of time-varying processes in the presence of measurement noise and outliers, in: Proc. 11th IFAC Symposium on System Identification, Fukuoka, Japan, 1997, 1765–1770.
- [58] M. Niedźwiecki, Elimination of clicks from archive speech signals, in: 6th IEEE International Workshop on Intelligent Signal Processing and Communication Systems, Melbourne, Australia, 1998, 1116–1120.
- [59] S. Canazza, G. De Poli, G. Mian, Restoration of audio documents by means of extended Kalman filter, IEEE Trans. Audio, Speech Language Process. 18 (6) (2010) 1107–1115.
- [60] M. Niedźwiecki, K. Cisowski, Adaptive scheme for elimination of broadband noise and impulsive disturbances from audio signals, in: Proc. 14th Colloque GRETSI, Juan-les-Pins, France, 1993, 519–522.

- [61] K. Cisowski, *Adaptacyjna filtracja i rekonstrukcja sygnałów fonicznych*, Rozprawa Doktorska, Politechnika Gdańska, 2001.
- [62] A. Janssen, R. Veldhuis, L. Vries, Adaptive interpolation of discrete-time signals that can be modelled as autoregressive processes, *Acoustics, Speech and Signal Processing, IEEE Transactions on* 34 (2) (1986) 317–330.
- [63] R. Veldhuis, *Restoration of Lost Samples in Digital Signals*, Prentice-Hall, 1990.
- [64] A. Czyżewski, Artificial intelligence-based processing of old audio recordings, in: *Proc. 97th Audio Engineering Convention*, San Francisco, 1994, 1–16.
- [65] A. Czyżewski, Some methods for detection and interpolation of impulsive distortions in old audio recordings, in: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1995, 139–142.
- [66] A. Czyżewski, Learning Algorithms for Audio Signal Enhancement, Part I: Neural Network Implementation for the Removal of Impulsive Distortions, *J. Audio Eng. Soc.* 45 (10) (1997) 815–831.
- [67] P. Sprechmann, A. Bronstein, J.-M. Morel, G. Sapiro, Audio restoration from multiple copies, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, 878–882.
- [68] J. Biliński, Dart XP Pro odnawianie nagrań dla zaawansowanych, *Estrada i Studio*, (8) 2003, stron 5.
- [69] S. Canazza, G. Coraddu, G. D. Poli, G. A. Mian, Objective and subjective comparison of audio restoration systems, in: *Proc. Int. Cultural Heritage Informatics Meeting, ICHIM'01*, 2001, 273–281.
- [70] F. Lewis, *Optimal Estimation*, Wiley, 1986.
- [71] M. Niedźwiecki, Statistical reconstruction of multivariate time series, *IEEE Trans. Signal Process.* 41 (1) (1993) 451–457.
- [72] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1979.
- [73] M. Niedźwiecki, *Identification of Time-varying Processes*, Wiley, 2001.
- [74] M. Niedźwiecki, On the localized estimators and generalized Akaike's criteria, *IEEE Trans. Automat. Contr.* 29 (11) (1981) 970–983.
- [75] T. Söderström, P. Stoica, *System Identification*, Prentice-Hall, 1988.
- [76] P. Stoica, R. Moses, On the unit circle problem: the Schur–Cohn procedure revisited, *Signal Processing* 26 (1) (1992) 95–118.
- [77] P. Stoica, Multistep prediction of autoregressive signals, *Electron. Lett.* 29 (6) (1993) 554–555.
- [78] J. Deller, J. Proakis, J. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan, 1993.



- [79] J. Wolfe, M. Garnier, J. Smith, Vocal tract resonances in speech, singing, and playing musical instruments, *HFSP J.* 3 (2009) 6–23.
- [80] P. Ramachandran, P. Kabal, Stability and performance analysis of pitch filters in speech coders, *IEEE Trans. Acoust., Speech, Signal Process.* 35 (7) (1987) 937–946.
- [81] P. Ramachandran, P. Kabal, Pitch prediction filters in speech coding, *IEEE Trans. Acoust., Speech, Signal Process.* 37 (4) (1989) 467–478.
- [82] W. Hess, *Pitch Determination of Speech Signals*, Springer, 1993.
- [83] M. Wu, D. Wang, G. J. Brown, A multipitch tracking algorithm for noisy speech, *IEEE Trans. Speech Audio Process.* 11 (3) (2003) 229–241.
- [84] J. Xu, J. Principe, A pitch detector based on a generalized correlation function, *IEEE Trans. Audio, Speech Language Process.* 16 (8) (2008) 1420–1432.
- [85] L. Rabiner, M. Cheng, A. Rosenberg, C. McGonegal, A comparative performance study of several pitch detection algorithms, *IEEE Trans. Acoust., Speech, Signal Process.* 24 (5) (1976) 399–418.
- [86] T. Shimamura, H. Kobayashi, Weighted autocorrelation for pitch extraction of noisy speech, *IEEE Trans. Speech Audio Process.* 9 (7) (2001) 727–730.
- [87] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, H. J. Manley, Average magnitude difference function pitch extractor, *IEEE Trans. Acoust., Speech, Signal Process.* 22 (5) (1974) 353–362.
- [88] A. de Cheveigné, H. Kawahara, YIN, a fundamental frequency estimator for speech and music, *J. Acoust. Soc. Amer.* 111 (4) (2002) 1917–1930.
- [89] R. Kashyap, R. Rao, *Dynamic Stochastic Models from Empirical Data*, Academic Press, 1976.
- [90] H. Lüthepohl, *Introduction to Multiple Time Series Analysis*, Springer-Verlag, 1991.
- [91] J. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.
- [92] P. Whittle, The analysis of multiple stationary time series, *J. Roy. Statist. Soc. B* 15 (1) (1953) 125–139.
- [93] R. Wiggins, E. Robinson, Recursive solution to the multichannel filtering problem, *J. Geophys. Res.* 70 (8) (1965) 1885–1891.
- [94] F. Cazals, C. Karande, A note on the problem of reporting maximal cliques, *Theor. Comput. Sci.* 407 (1–3) (2008) 564–568.
- [95] C. Bron, J. Kerbosch, Algorithm 457: Finding all cliques of an undirected graph, *Comm. ACM* 16 (9) (1973) 575–577.
- [96] J. Konc, D. Janezic, An improved branch and bound algorithm for the maximum clique problem, *MATCH Commun. Math. Comput. Chem.* 58 (2007) 569–590.



- [97] ITU-T Recommendation P.862, Perceptual evaluation of speech quality (PESQ): Objective method for end-to-end speech quality assessment of narrow band telephone networks and speech codecs, 2001.
- [98] J. Beerends, A. Hekstra, A. Rix, M. Hollier, Perceptual evaluation of speech quality (PESQ), the new ITU standard for end-to-end speech quality assessment, part II – Psychoacoustic Model, *J. Audio Eng. Soc.* 50 (10) (2002) 765–778.
- [99] P. Esquef, L. Biscainho, L. Nunes, B. Lee, A. Said, T. Kalker, R. Schafer, Quality assessment of audio: increasing applicability scope of objective methods via prior identification of impairment type, in: *Proc. IEEE Int. Workshop Multimedia Signal Process.*, 2009, 1–6.
- [100] ITU-T Recommendation P.863 : Perceptual Objective Listening Quality Assessment, 2014.
- [101] ITU-R Recommendation BS.1387-1: Method for Objective Measurements of Perceived Audio Quality, 2001.
- [102] P. Kabal, An examination and interpretation of ITU-R recommendation BS.1387: Perceptual Evaluation of Audio Quality, Department of Electrical & Computer Engineering, McGill University, Canada, 2003.
- [103] ITU-R Recommendation BS.1534-2: Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems, 2014.
- [104] ITU-R Recommendation BS.1116-3: Methods for the Subjective Assessment of Small Impairments in Audio Systems, 2015.
- [105] D. Tylavsky, G. Sohie, Generalization of the matrix inversion lemma, *Proceedings of the IEEE* 74 (7) (1986) 1050–1052.