

prof. dr hab. inż. Andrzej Dobrucki
Politechnika Wrocławska
Wydział Elektroniki
Katedra Akustyki i Multimediów
Wybrzeże Wyspiańskiego 27
50-370 Wrocław
Tel. 71 320 30 68
e-mail: andrzej.dobrucki@pwr.edu.pl

RECENZJA

rozprawy doktorskiej mgr. inż. Piotra Kardasza
pt. „ Wykorzystanie metod nierównomiernego próbkowania i inteligentnych algorytmów
przetwarzania sygnałów do rekonstrukcji archiwalnych nagrań dźwiękowych”

Podstawą wykonania recenzji jest uchwała Rady Wydziału Elektrycznego Politechniki Białostockiej nr 37/2016 z dnia 25.05.2016 r. Promotorem rozprawy jest dr hab. inż. Mirosław Świercz, prof. nadzw. PB. Rozprawa podzielona jest na 6 rozdziałów, z których pierwszy stanowi wstęp, zaś w ostatni podsumowanie wraz z wytyczeniem kierunków dalszych badań. Rozprawa zawiera streszczenie w języku angielskim oraz wykaz literatury liczący 97 pozycji. W spisie literatury zawartych jest 10 jednoautorskich pozycji Doktoranta dotyczących tematu rozprawy. Wszystkie pozycje są artykułami opublikowanymi w dwóch czasopismach: *Pomiary, Automatyka i Robotyka* oraz *Poznań University of Technology Academic Journals. Electrical Engineering*. Publikacje te są w języku polskim i powstały w latach 2013 – 2016. Rozprawa liczy 166 stron, zawiera 65 rysunków i 16 tabel.

We wstępie Autor podaje wprowadzenie do tematyki rozprawy, przedstawia pokrótce zawartość kolejnych rozdziałów oraz formułuje tezę rozprawy. Teza ta brzmi:

Możliwe jest opracowanie metod, wykorzystujących nierównomierne próbkowanie sygnałów i inteligentne algorytmy ich przetwarzania, które pozwolą na usprawnienie i poprawę wyników procesu rekonstrukcji archiwalnych nagrań dźwięku.

Stwierdzam, że tematyka rozprawy związana z rekonstrukcją uszkodzonych archiwalnych nagrań dźwięku jest ważna z naukowego punktu widzenia. Mimo, że istnieją komercyjne programy komputerowe pozwalające na usunięcie z nagrań powstałych skażeń, jednakże metody rekonstrukcji są stale rozwijane i są przedmiotem wielu współczesnych publikacji. Zastosowanie do procesu rekonstrukcji metod nierównomiernego próbkowania jest oryginalnym wkładem Autora rozprawy. Tzw. algorytmy inteligentne są stosowane do procesu rekonstrukcji, jednakże Autor rozprawy opracował i wdrożył wiele oryginalnych pomysłów. W związku z tym stwierdzam, że teza rozprawy jest sformułowana poprawnie. Pragnę w tym miejscu zwrócić uwagę na błędy terminologiczne, które pojawiają się w tym rozdziale, ale występują również w rozdziałach kolejnych. Otóż Autor pisze (str. 7, w. 3 i 4): „Wykonane metodami analogowymi nagrania charakteryzują się także szumami i zakłóceniami.”. W dalszym ciągu zaś na tej samej stronie:

„Zakłócenia występujące w archiwalnych nagraniach dźwięku można, stosując różne kryteria, podzielić na:

- Liniowe, zarówno amplitudowe jak i fazowe,
- Nieliniowe analogowe (..) i cyfrowe (...),
- Quasistacjonarne (np. przydźwięk sieciowy, pilot stereo, częstotliwość odchylenia poziomego),
- Zniekształcenia wynikające ze zmian prędkości przesuwu nośnika (tzw. wow)

- Impulsowe (...),
- Szумы,
- Itd.”

Ponieważ występuje tu wiele nieporozumień, muszę zrobić mały wykład.

Sygnal akustyczny i foniczny w procesie transmisji przez kanał elektroakustyczny podlega rozmaitym skażeniom. Skażenia te możemy podzielić na:

- Zniekształcenia
- Zakłócenia.

Zniekształcenia definiuje się jako deformacje sygnału, spowodowane niedoskonałością toru elektroakustycznego. Zniekształcenia dzielą się na:

- Liniowe, do których zalicza się zniekształcenia amplitudowe oraz fazowe (opóźnieniowe). Zniekształcenia amplitudowe wynikają z transmisji sygnałów o różnych częstotliwościach z różnym wzmocnieniem lub tłumieniem. Są charakteryzowane za pomocą częstotliwościowej charakterystyki amplitudowej (amplituda vs częstotliwość). Zniekształcenia fazowe wynikają z transmisji sygnałów o różnych częstotliwościach z różnymi opóźnieniami. Charakterystyczną cechą zniekształceń liniowych jest to, że w wyniku tych zniekształceń w sygnale nie pojawiają się składowe o innych częstotliwościach, niż były w sygnale oryginalnym.
- Nieliniowe, spowodowane nieliniowymi właściwościami toru elektroakustycznego. W wyniku tych zniekształceń w sygnale pojawiają się składowe o innych częstotliwościach, niż były w sygnale oryginalnym. Dlatego zniekształcenia te mierzy się poprzez określenie zawartości tych składowych (harmonicznych, intermodulacyjnych, różnicowych itd.) . Tak naprawdę, nieliniowy tor elektroakustyczny charakteryzowany jest przez wielowymiarowe jądra szeregowoltery lub Wienera albo ich wielowymiarowe transformaty Fouriera.

Zakłócenia definiuje się jako sygnały obce, dostające się do transmitowanego przez tor elektroakustyczny sygnału. Do zakłóceń należą szумы, przydźwięk i inne zakłócenia quasistacjonarne wymienione przez Autora, zakłócenia impulsowe, przesłuchy międzykanałowe itd.

Z tych powodów nie ma sensu mówić „szумы i zakłócenia”, ponieważ szумы są zakłóceniami. Nie można natomiast wymieniać zniekształceń liniowych, nieliniowych i *wow* (specyficzne zniekształcenia nieliniowe) jako zakłóceń. Do sprawy tej terminologii nie będę wracał w dalszej części recenzji, aczkolwiek te nieporozumienia są powielane.

Rozdział 2 dotyczy sygnału dźwiękowego, metod jego zapisu oraz sposobów analizy. Na początku rozdziału Autor przedstawia podstawowe informacje o dźwięku. Autor rozprawy nie jest akustykiem, i zapewne stąd w tej części występuje wiele informacji bałamutnych. I tak, moc przypadająca na jednostkę powierzchni nazywa się natężeniem dźwięku. Jest to wielkość wektorowa, skierowana prostopadle do czoła fali. Dla jednowymiarowej fali płaskiej z reguły pomija się wektorowy charakter natężenia dźwięku. Nie można tej wielkości mylić z mocą akustyczną, ani też gęstością mocy, jak to czyni Autor na str. 11. Jednostka ciśnienia akustycznego po polsku nazywa się po prostu paskal, a nie pascal. Ciśnienie nie jest amplitudą drgań dźwięku, można tutaj uznać za wielkość charakteryzującą drgania akustyczne prędkość cząstek drgających, zwaną również prędkością akustyczną. Charakteryzując poziom dźwięku, nie używa się pojęcia amplitudy ciśnienia (w przypadku przebiegów niesinusoidalnych często trudno ją określić), a wartości skutecznej. Poziom określa się jako 20 logarytmów dziesiętnych stosunku wartości skutecznej ciśnienia akustycznego do wartości poziomu odniesienia 20 μPa . Poziom natężenia definiuje się jako 10 logarytmów dziesiętnych stosunku natężenia dźwięku do wartości poziomu odniesienia 10^{-12} W/m^2 . Dla tak zdefiniowanych poziomów odniesienia, poziom ciśnienia i poziom natężenia dźwięku w powietrzu są równe. Dla szumu

(szmeru), zamiast posługiwać się terminem „Intensywność” (str. 11 w. 9 od dołu), Autor powinien użyć precyzyjnego określenia „widmowa gęstość mocy”. W kolejnym podrozdziale Autor przedstawia pokrótce budowę ucha oraz podstawowe wiadomości z psychologii słyszenia. Omawiając funkcję ucha zewnętrznego, warto może dodać, że budowa małżowiny usznej wpływa na słyszenie kierunkowe, charakteryzowane przez tzw. funkcję transmitancji odniesioną do głowy (HRTF – Head Related Transfer Function). Ponadto, rezonans akustyczny kanału słuchowego powoduje wzrost czułości słuchu dla częstotliwości 2 – 3 kHz. Omawiając z kolei ucho środkowe, warto byłoby wspomnieć, że kosteczki słuchowe pełnią rolę transformatora akustycznego. Nie powinno się też pisać, że ucho wewnętrzne składa się ze ślimaka oraz błony podstawnej (str. 12, w. 4 od dołu), ponieważ błona podstawna jest częścią ślimaka. Ucho wewnętrzne składa się ze ślimaka i kanałów półkolistych, które pełnią funkcję organu równowagi. Na str. 14 zamieszczony jest rys. 2.2, przedstawiający krzywe jednakowego poziomu głośności (a nie krzywe jednakowej głośności, jak pisze Autor!) wg Fletchera i Munsona. Liczby, położone na krzywych są właśnie poziomami głośności, wyrażonymi w fonach, natomiast głośność jest wyrażana w sonach. Krzywe wg Fletchera i Munsona nie są obecnie stosowane. Przez długi czas stosowano krzywe wg Robinsona i Dadsona, na podstawie artykułu opublikowanego w roku 1956. Obecnie obowiązująca norma międzynarodowa **ISO 226: 2003 Acoustics -- Normal equal-loudness-level contours** definiuje jeszcze inne krzywe. Na str. 15, w. 11 od góry, Autor wprowadza jeszcze inne nieprecyzyjne pojęcie częstotliwości próbkowania, wyrażonej w hercach [Hz]. Jest to stosowane dość powszechnie w literaturze popularnej, ale w pracy naukowej powinien być stosowany termin „szybkość próbkowania”, która wyrażana jest w próbkach na sekundę. Częstotliwość jest cechą sygnału, natomiast szybkość próbkowania jest cechą procesu. Ma to również odzwierciedlenie w literaturze anglojęzycznej, w której stosuje się termin „bitrate” lub „sampling rate” zamiast „frequency”. Ta nieścisłość również przewija się przez całą pracę. W dalszej części rozdziału 2 Autor przedstawia metodę mechanicznego i magnetycznego zapisu dźwięku, jako najczęściej występujących metod rejestracji dźwięku w nagraniach archiwalnych, a także będących obiektem badań Autora rozprawy. Omówiona jest historia obu metod zapisu, oraz podstawowe ich właściwości, w tym typowe zniekształcenia i zakłócenia występujące w tych nagraniach. Omawiając historię zapisu mechanicznego warto może by wspomnieć o Emilu Berlinerze (1851–1929), wynalazcy płyty gramofonowej, oraz o duńskim inżynierze Valdemarze Poulsenie (1869 - 1942), wynalazcy zapisu magnetycznego na drucie stalowym. W ostatnim podrozdziale 2.4 przedstawione są metody analizy sygnałów akustycznych. Myślę, że bardziej właściwe byłoby użycie terminu „sygnał foniczny”, ponieważ sygnał akustyczny jest przenoszony w postaci fal dźwiękowych, zaś sygnał foniczny powstaje w wyniku przeniesienia transmisji sygnału w dziedzinę elektryczną. Sygnał foniczny to jest właśnie sygnał elektryczny, odpowiadający sygnałowi akustycznemu. Omówione są metody czasowe, częstotliwościowe oraz czasowo-częstotliwościowe analizy sygnałów. Analiz dokonuje się zwykle przenosząc sygnał z postaci analogowej do cyfrowej. Omawiając podstawową dla analizy metodę ciągłej i dyskretnej transformaty Fouriera, Autor popełnia we wzorach na transformatę odwrotną błąd (wzory 2.2 i 2.4) polegający na pozostawieniu znaku minus (-) w wykładniku eksponenty. Dla transformaty odwrotnej powinien występować tam znak plus (+). Autor w dalszym ciągu rozprawy używa transformaty Fouriera bardzo długich sekwencji próbek ($2^{18} - 2^{23}$). Chcę się w tym miejscu zapytać, jak długo trwały obliczenia tych transformat? Ze względu na konieczną rozdzielczość częstotliwościową w stosowanych metodach rekonstrukcji uznaję postępowanie Autora za zasadne, pytanie wynika z ciekawości.

W tym miejscu pragnę skomentować pewną informację dotyczącą tabeli 2.2, a konkretnie okien czasowych nazwanych tam oknem Hanna i Hanninga. Nigdzie, poza książką prof. T. Zielińskiego nie

spotkałem rozróżnienia okien Hanna i Hanninga, różniących się jedną próbką. Zwykle stosuje się okno wg wzoru podanego w wierszu trzecim (okno Hanna). Okno zawiera N próbek. Jednakże w MATLABIE, stosuje się jako okno Hanna zgodne ze wzorem w wierszu 4 tabeli. Okno to zawiera $N+1$ próbek. Jak sądzę, dla długich sekwencji różnica jest znikoma. Trzeba w tym miejscu dodać, że nie istniał badacz o nazwisku Hanning. Nazwa tego okna czasowego jest związane z nazwiskiem austriackiego meteorologa Juliusa von Hanna (1831 – 1921). Okno to nazwali na jego cześć Blackman i Tukey w pracy z roku 1959. Użyli oni terminu „hanning” na zastosowanie funkcji okna do danego ciągu. Po polsku można to tłumaczyć jako „hannowanie”. Termin się przyjął, ale nieporozumienie zostało powielone, nawet w literaturze anglojęzycznej i okno to zaczęto nazywać „Hanning” (z dużej litery), co sugerowało, że nazwa pochodzi od nazwiska Hanning. Nieporozumienie pogłębiał fakt, że istniał uczony amerykański Richard Hamming (1915 – 1998), od którego nazwiska nazwano inny typ okna – okno Hamminga właśnie. Ten przydługi komentarz nie jest zarzutem w stosunku do Autora, chociaż w dalszym ciągu pracy używał on okna Hanninga, jak wynika np. ze wzoru (4.31), chodzi o okno wg 4 wiersza tabeli 2.2.

W dalszym ciągu podrozdziału Autor prezentuje pokrótce analizę za pomocą falek, przedstawia też ideę próbkowania nierównomiernego i oszczędnego. Opisuje też tzw. algorytmy inteligentne (genetyczne i ewolucyjne) oraz algorytmy równoległe służące do przyspieszenia obliczeń. Ostatnia część tego rozdziału jest zatytułowana „Klasyfikacja nagrań archiwalnych”. Tytuł ten jest niezbyt szczęśliwy, ponieważ właściwie nie dotyczy klasyfikacji, a jedynie przedstawia kilka istniejących programów komercyjnych służących rekonstrukcji starych nagrań. Lista ta jest mocno niekompletna, brak jest m.in. bardzo dobrych programów firmy Algorithmix (<http://www.algorithmix.com/>), np. Sound Laundry. Na końcu tej części znajduje się passus wskazujący na potrzebę stworzenia narzędzia do klasyfikacji nagrań na potrzeby automatycznej rekonstrukcji. Właściwa klasyfikacja nagrań jest przedmiotem rozważań w rozdziale 4.3. Cały rozdział 2 został opracowany na podstawie literatury, chociaż niektóre przykłady ilustrujące rozważania zostały opracowane samodzielnie przez Autora.

Rozdział 3 jest zatytułowany „Charakterystyka zakłóceń sygnałów akustycznych”. Moje uwagi dotyczące rozróżnienia zakłóceń i zniekształceń, a także sygnału akustycznego i fonicznego pozostają w mocy. Rozdział 3, aczkolwiek opisuje znane zakłócenia (i zniekształcenia) pojawiające się w nagraniach dźwiękowych, jest ilustrowany przykładami opracowanymi przez Autora rozprawy. W kolejnych podrozdziałach przedstawione są quasistacjonarne zakłócenia harmoniczne (przydźwięk, sygnał pilota stereo, odchylenie poziome w odbiornikach TV), szумы, zakłócenia impulsowe i zaniki sygnału, oraz zniekształcenia liniowe i nieliniowe, a także wynikające ze zmian prędkości przesuwu nośnika (*wow*), które stanowią formę zniekształceń nieliniowych. Opisane są również mniej znane efekty, takie, jak „loudness war” oraz „musical noise”. Przykłady są dobrane na ogół trafnie i dobrze ilustrują wpływ rozmaitych typów zakłóceń (i zniekształceń) na charakterystyki sygnału. Przykład przedstawiony na rys. 3.1., mający ilustrować przydźwięk, dotyczy sytuacji, która nigdy nie miała miejsca w praktyce zasilania urządzeń elektronicznych, również archiwalnych. Prostowanie jedno- lub dwupołówkowe bez odfiltrowania tętnień i stabilizacji stosuje się co najwyżej do ładowania akumulatorów. W podrozdziale 3.5 dotyczącym zniekształceń liniowych Autor omawia typy stosowanych wkładek gramofonowych: „moving magnet” i „moving coil”, nazywając je wkładkami elektromagnetycznymi. Nazwa ta odnosi się jedynie do wkładek typu „moving magnet” (lub części jej „moving armature”). Wkładki typu „moving coil” pracują na zupełnie innej zasadzie – mianowicie indukcji elektromagnetycznej. Wkładki pracujące na tej zasadzie nazywają się magnetoelektryczne (nie elektromagnetyczne!), albo elektrodynamiczne, patrz moja książka „Przetworniki elektroakustyczne”, WNT Warszawa 2007. Gwoli ścisłości trzeba wspomnieć, że stosuje się również

wkładki piezoelektryczne (krystaliczne), w których powstała siła elektromotoryczna jest proporcjonalna do wychylenia igły, a nie do prędkości, i w tym przypadku nie jest wymagana korekcja krzywą RIAA. W moich młodych latach miałem NRD-owski gramofon Ziphona, który miał właśnie wkładkę piezoelektryczną. W ostatnim podrozdziale 3.7, Autor opisuje inne rodzaje zniekształceń i zakłóceń, w tym głównie *mp3*. Rzeczywiście, stosowanie kompresji stratnej jest procesem silnie nieliniowym i generującym dodatkowy szum, jednakże artefakty te są zamaskowane. Wydaje mi się, że nagania *mp3*, stanowią jeszcze w dalszym ciągu pewną nowość i nie wymagają rekonstrukcji, a jeśli już, to przy użyciu zupełnie innych metod. Nie jest to przedmiotem dalszych rozważań Autora rozprawy. „Peak clipping” jest klasycznym przykładem zniekształceń nieliniowych.

Rozdział 4 jest najważniejszym i najdłuższym rozdziałem rozprawy. Został on opracowany w całości samodzielnie przez Autora. Przedstawiono w nim algorytmy używane do identyfikacji i usuwania rozmaitych artefaktów pojawiających się w starych nagraniach dźwiękowych. W rozdziale 4.1 przedstawiony jest w postaci opisu i schematu blokowego ogólny przebieg procesu rekonstrukcji, począwszy od przetworzenia postaci analogowej nagrania na postać cyfrową z wymaganą szybkością próbkowania, a skończywszy na zapisie zrekonstruowanego nagrania w postaci cyfrowej z wybranymi przez użytkownika parametrami zapisu. Jeśli sygnał wejściowy jest już przetworzony do postaci cyfrowej, to zachodzi zwykle potrzeba jego przepróbkowania do szybkości 192000 próbek/s, używanej w programie rekonstrukcyjnym. Autor opisuje algorytm takiego przepróbkowania z zapisu o szybkości 44100 próbek/s (standard CD) . Algorytm wykorzystuje przepróbkowanie do szybkości pośredniej równej 2822400 próbek/s, poprzez wstawienie między kolejne próbki 63 zer, i poddanie uzyskanego sygnału filtracji dolnoprzepustowej za pomocą filtru Czebyszewa 6 rzędu. Filtr ten zapewnia odpowiednie tłumienie w paśmie zaporowym. Następnie wybiera się dwie próbki na miejscach najbardziej zbliżonych do miejsca, w którym występuje próbka o docelowej szybkości próbkowania i dokonuje się interpolacji liniowej. Szybkość pośrednia jest dziesięciokrotnie mniejsza niż najmniejsza wspólna wielokrotność 44100 i 192000. Filtr Czebyszewa ma największe nachylenie zboczy w pasmie zaporowym i w praktyce jest najczęściej stosowany jako filtr antyaliasowy. W podobny sposób można dokonać przepróbkowania sygnałów o szybkościach będących stosowanymi wielokrotnościami lub podwielokrotnościami szybkości 44100 próbek/s. Przepróbkowanie nagrań o szybkości 48000 próbek/s (standard DAT) oraz jej wielokrotności i podwielokrotności nie wymaga stosowania szybkości pośrednich. Testy wykazały poprawność metody i mały poziom wprowadzonych zakłóceń. Kolejnym etapem działania algorytmu jest klasyfikacja nagrania na podstawie występujących w nim zakłóceń i zniekształceń. Sygnał zalicza się do jednej z 7 klas o różnym stopniu zakłóceń lub zniekształceń oraz pierwotnym źródle nagrania (płyta gramofonowa lub taśma magnetyczna). Na podstawie dokonanej klasyfikacji program proponuje operatorowi różne korekcje. Operator może z części z nich zrezygnować, lub też poszerzyć program rekonstrukcji. Może też zdać się na program, który automatycznie będzie wykonywał odpowiednie działania rekonstrukcyjne. Klasyfikację dokonuje się z użyciem sieci neuronowych.

W podrozdziale 4.4. Autor opisuje algorytmy identyfikacji składowych harmonicznym (sinusoidalnym) występujących w sygnale. Ta identyfikacja jest wykorzystywana w kolejnych algorytmach rekonstrukcyjnych, zarówno takich, w których zidentyfikowane pasożytnicze składowe sinusoidalne usuwa się z sygnału, jak i takich, w których używa się tych składowych w procesie syntezy zrekonstruowanego sygnału. Autor rozprawy wdrożył dwa algorytmy, algorytm ewolucyjny, stosowany w optymalizacji, oraz algorytm z zastosowaniem tzw. nieortogonalnej bazy. Algorytm ewolucyjny jest dobrze znany, należy on do grupy algorytmów inteligentnych. Badania Autora pokazały jego dobre działanie, ale jest on też stosunkowo powolny. Działanie algorytmu z

nieortogonalną bazą wykorzystuje fakt, że całka z iloczynu dwóch funkcji trygonometrycznych (sinus i cosinus) o różnych argumentach jest funkcją okresową, natomiast całka z kwadratu funkcji sinus lub cosinus ma składową liniowo rosnącą. Po dostatecznie długim czasie składowa ta dominuje, a składowa okresowa może być pominięta. Identyfikacja składowych sinusoidalnych wymaga w tym przypadku stosowania dużej liczby funkcji o gęsto rozłożonych częstotliwościach. Algorytm ten działa szybciej niż algorytm ewolucyjny, ale też jest mniej dokładny. Do testowania wybrano składowe niezaszumione oraz zaszumione szumem białym o poziomie SNR=20 dB. Jak należało się spodziewać, szybciej i dokładniej algorytm wykrywał składowe w sygnale niezaszumionym. Nie bardzo wiem, dlaczego w tabeli 4.8 występują sygnały zaszumione z poziomem SNR=0 dB, nieopisane w tekście. Oba algorytmy, ewolucyjny i wykorzystujący nieortogonalną bazę do identyfikacji składowych sinusoidalnych potrzebują stosunkowo niewielkiej liczby próbek, przy czym algorytm ewolucyjny potrzebuje ich mniej. Konieczna liczba próbek wzrasta wraz ze wzrostem poziomu szumu. W końcowej części tego podrozdziału Autor przetestował funkcjonowanie algorytmu identyfikacji składowych sinusoidalnych opartego o nieortogonalną bazę w warunkach obliczeń równoległych. Określił stopień przyspieszenia obliczeń w stosunku do działania jednowątkowego dla różnych komputerów. Różnice między zastosowanymi komputerami okazały się znaczne.

W kolejnych podrozdziałach opisane są algorytmy poszczególnych działań rekonstrukcyjnych: identyfikacja i usuwanie niepożądanych składowych sinusoidalnych, zniekształceń *wow*, szumów i zakłóceń impulsowych.

Algorytm polegający na wymnażaniu sygnału przez sygnał sinusoidalny o danej częstotliwości oraz całkowaniu iloczynu stał się podstawą algorytmu wykrywającego niepożądane składowe sinusoidalne w nagraniu archiwalnym oraz usuwającego te składowe. Jest to przedmiotem rozważań w podrozdziale 4.5. Niepożądanymi składowymi są zarówno składowe niskoczęstotliwościowe (przydźwięk i jego harmoniczne) jak i składowe wysokoczęstotliwościowe (częstotliwość odchylenia poziomego w nagraniach z odbiornika TV, pilot stereo). Badania testowe potwierdziły poprawność działania algorytmu, chociaż nie wiem, dlaczego nie wykorzystano sygnałów testowych 3, 5 i 6 z tab. 4.11. Jeśli nie były wykorzystane, to po co w ogóle je umieszczano? Nie bardzo rozumiem też rozważań w końcowej części podrozdziału 4.5 dotyczących niepotrzebnego usuwania harmonicznego przydźwięku z sygnału. Czy chodzi tu o składowe użyteczne sygnału, których częstotliwości przypadkowo pokrywają się z harmonicznymi przydźwięku? Pod koniec podrozdziału Autor pisze o badaniach odsłuchowych. Te badania pojawiają się w dalszej części rozprawy, przy opisie testowania kolejnych elementów algorytmu rekonstrukcyjnego. Brak jest w rozprawie jakichkolwiek informacji na temat tych badań: metodyki badań, liczebności i składu ekipy odsłuchowej, stosowanych metod statystycznych weryfikacji wyników. Uważam to za niedociągnięcie rozprawy.

Kolejny podrozdział (4.6) przedstawia metodę wykrycia i usunięcia z nagrania zniekształceń *wow*, związanych z nierównomiernością prędkości nośnika. Metoda ta opiera się o algorytm detekcji składowych harmonicznymi. Po wykryciu zakłócenia, objawiającego się zmianą częstotliwości obserwowanej harmonicznej, dokonuje się korekcji poprzez zmianę szybkości próbkowania (próbkowanie nierównomierne). Rozdział napisany jest bardzo niejasno. Z jego treści wynika, że częstotliwością, której zmienność się bada, jest 50 Hz, czyli przydźwięk. Częstotliwość ta została dodana do nagrania, a następnie symulowano jej zmiany. W oryginalnym nagraniu sytuacja nie jest tak komfortowa. Po pierwsze, przydźwięk jest zakłóceniem, które może wystąpić w nagraniu lub nie. Sztuczne dodawanie przydźwięku do archiwalnego nagrania nie ma sensu. Gdy przydźwięku nie ma, detekcję zniekształcenia *wow* trzeba oprzeć o inną częstotliwość, która występuje w sygnale, niekoniecznie pasożytniczą. W takim razie problemem staje się naturalna zmiana częstotliwości. Gdy

przydźwięk występuje w nagraniu, to ewentualne usunięcie zniekształceń wow należy przeprowadzić przed jego usunięciem. Informacja o korekcji zniekształceń wow poprzez zmianę szybkości próbkowania występuje w schemacie na rys. 4.19, nie powołanym w tekście rozprawy, natomiast nie ma jej w tekście zasadniczym.

Podrozdział 4.7 dotyczy redukcji szumów występujących w nagraniu. W pierwszej części tego podrozdziału Autor opisuje tzw. klasyczne metody redukcji szumów oraz ich wady, takie jak możliwość wystąpienia „wzdychania” szumu na skutek niedoskonałości algorytmu, lub wystąpienia tzw. „musical noise”. W tym krótkim podrozdziale brak jest jakichkolwiek odniesień do ogromnej przecież literatury przedmiotu. Autor nie wspomina również o takich klasycznych metodach redukcji, jak odejmowanie widmowe czy zastosowanie filtrów Wienera lub Kalmana. W drugiej części podrozdziału 4.7 opisany jest autorski algorytm redukcji szumów. Polega on po pierwsze na określeniu poziomu tła szumów, np. z nienagranej części nośnika, a następnie, przy zastosowaniu uprzednio wybranych algorytmów, wyodrębnienie z sygnału składowych harmonicznym przekraczających ustalony próg. Z pozostałej, stochastycznej części sygnału usuwa się w poszczególnych pasmach (w danym przypadku oktawowych) szum o poziomie tła używając klasycznych metod. Pozostała część zawiera składowe harmoniczne o niskich poziomach oraz użyteczne składowe stochastyczne. Tę część sumuje się następnie z wyodrębnionymi uprzednio składowymi harmonicznymi. Ten sposób postępowania zabezpiecza proces rekonstrukcji przed wystąpieniem wyżej wspomnianych artefaktów, takich jak *musical noise* czy „wzdychanie” szumowe. Algorytm przeszedł pozytywnie testy. Jak pisze Autor, uzyskano redukcję poziomu szumu o kilkadziesiąt decybeli. Co to znaczy? Kilkadziesiąt, to jest zarówno pięćdziesiąt, jak i dziewięćdziesiąt. Były też przeprowadzone badania słuchowe, ale znów nie podano o nich żadnych informacji.

Redukcja zakłóceń impulsowych jest przedmiotem rozważań w podrozdziale 4.8. Zakłócenia impulsowe są charakterystyczne dla starych nagrań na płytach. Ich usuwanie jest poważnym wyzwaniem dla rekonstruktora. Poważnym problemem jest tu łatwa możliwość pomyłki takiego zakłócenia z chwilowym wzrostem poziomu nagrania. Jest to przedstawione na rys. 3.7. W podrozdziale przedstawiono autorski algorytm polegający na wyodrębnieniu z nagrania wolnozmiennych składowych harmonicznym metodą opisaną w podrozdziale 4.7. Szybkozmienne zakłócenia impulsowe pozostają w części stochastycznej sygnału. Następnie porównuje się poziomy sąsiadujących próbek i jeśli różnica poziomów przekracza określony próg, próbki o dużych poziomach traktuje się jak zakłócenia impulsowe i się je usuwa. Powstałą lukę wypełnia się próbkami estymowanymi lub w prostszej wersji algorytmu - szumem kolorowym. Jednym z sygnałów testowych jest prostokąt, na który nałożono gasnący przebieg sinusoidalny. Ten przebieg jest impulsową składową szybkozmienną i został on skutecznie usunięty za pomocą opisanego algorytmu. Moja wątpliwość dotyczy przebiegu prostokątnego. Narastające lub opadające zbocze też jest przecież składową szybkozmienną, a jednak nie została ona usunięta z sygnału. Może to wynikać z tego, że algorytm usuwa te próbki, których amplituda przekracza znacznie wartości dwóch próbek sąsiadujących z nią z obu stron, a w przypadku prostokąta takie przekroczenie występuje jednostronnie. Ale z drugiej strony, algorytm w przypadku szerokiego (tzn. o szerokości przekraczającej trzy próbki) zakłócenia impulsowego może spowodować nie tyle usunięcie zakłócenia, ile spłaszczenie jego wierzchołków. Znowu brak jest opisu przeprowadzonych badań odsłuchowych.

Ostatni podrozdział 4.9 rozdziału czwartego dotyczy poprawy brzmienia nagrań. W starym nagraniu, ze względu na zastosowaną aparaturę, często jest obcięte pasmo, powodując pogorszenie wrażenia estetycznego. Dotyczy to zwłaszcza zubożenia wysokich tonów. Zagadnienie polega na

dodaniu do rekonstruowanego nagrania brakujących składowych wysokoczęstotliwościowych. Jeśli traktuje się proces rekonstrukcji archiwalnego nagrania jako usunięcie z niego artefaktów wynikających z długotrwałej eksploatacji (szumy, trzaski) i powstałych w procesie nagrania (składowe tonalne, *wow*), to nie wiem, czy dodawanie do niego dodatkowych składowych można nazwać rekonstrukcją. W wyniku takich działań powstaje może nagranie lepszej jakości niż oryginalne, ale różniące się od niego – czyli zupełnie nowe nagranie. Autor opisuje tu dwa algorytmy poprawy brzmienia. Pierwszy z nich polega na identyfikacji składowych tonalnych występujących w nagraniu oryginalnym, w paśmie przeniesionym przez nagranie, a następnie dodanie do nagrania estymowanych harmonicznym tych składowych, które występują poza pasmem przeniesionym przez nagranie. Amplitudy estymowanych składowych są obliczane wg ustalonej reguły. Widmo nagrania przedstawiono na rys. 4.20. W podpisie pod rysunkiem powołano się też na rys. 4.14, który jednak dotyczy czegoś innego. Być może chodzi o rys. 4.22, w którym mały poziom widma powyżej 4 kHz uzyskano na skutek procesu odszumiania, a na rys. 4.28 poziom w zakresie dużych częstotliwości został zwiększony poprzez dodanie składowych harmonicznym. Nieprawidłowe powołanie się na rys. 4.14 występuje również w tab. 4.13, która ma również nieprawidłowy numer, ponieważ tabela o tym numerze wystąpiła poprzednio na str. 122. Drugi algorytm jest prostszy. Polega on na poddaniu nagrania przekształceniu nieliniowemu. Pojawiające się w wyniku tego przekształcenia składowe harmoniczne i sumacyjne, leżące w paśmie nieprzenoszonym przez oryginalne nagranie, są do niego dodawane, wzbogacając brzmienie w zakresie dużych częstotliwości. Warto może dodać, że podobne przekształcenie, z użyciem składowych różnicowych używane jest do wzbogacenia basów w odtwarzanych nagraniach. Algorytm nieliniowy ma jeszcze mniej wspólnego z rekonstrukcją, ponieważ składowe są dodawane w sposób całkowicie sztuczny. Widmo nagrania z użyciem przekształcenia nieliniowego jest przedstawione na rys. 4.30. Rysunek ten jest chyba jednak pomyłony, ponieważ poziom w zakresie dużych częstotliwości po przekształceniu jest niższy niż przed przekształceniem, ponadto pojawia się w nim artefakt w postaci pojedynczej składowej sinusoidalnej o częstotliwości ok. 16 kHz. Znowu Autor pisze o pozytywnych wynikach testów odsłuchowych, zaś nic o samych testach.

W rozdziale 5 Autor opisuje praktyczną implementację opracowanych algorytmów. Rozdział składa się z dwóch części. W pierwszej znajdują się rozważania dotyczące możliwości wdrożenia algorytmów na różnych platformach hardware'owych: komputerach klasy PC, kilku wielordzeniowych mikrokontrolerach i systemach mikroprocesorowych oraz układach FPGA. Autor opracował i przetestował system wykorzystujący układ FPGA i mikrokontroler Propeller firmy Parallax Semiconductor (USA) . W drugiej części rozdziału 5 opisany jest opracowany program do rekonstrukcji nagrań archiwalnych w oparciu o przedstawione w rozdziale 4 algorytmy. Program w obecnej wersji wymaga wprowadzenia nagrań przewidzianych do rekonstrukcji już przekonwertowanych na postać cyfrową (plik *.wav). Program dokonuje przepróbkowania do szybkości 192 kpróbk/s, i klasyfikacji nagrania. Na jego podstawie proponuje wybór określonych działań rekonstrukcyjnych. Użytkownik może przeprowadzić je ręcznie wybierając kolejne opcje, albo też zdać się na program, który dokonuje automatycznej rekonstrukcji. W przypadku plików o jakości zakwalifikowanej jako dobra program nie podejmuje żadnego działania. Po zakończeniu program zapisuje zrekonstruowany plik z żadaną przez użytkownika rozdzielczością i szybkością próbkowania. W końcowej części rozdziału 5 znajdują się wytyczne do dalszych prac nad opracowanym programem.

Rozdział 6 stanowi podsumowanie rozprawy. Autor stwierdza, że cel pracy został osiągnięty, a teza – udowodniona. Zgadzam się z Autorem.

Autor rozprawy doktorskiej wykonał dużą pracę. Opracował i wdrożył szereg oryginalnych algorytmów. Ich badanie pokazało poprawne funkcjonowanie. Rozprawa stanowi istotny przyczynek do badań nad rekonstrukcją nagrań archiwalnych. Oprócz znaczenia teoretycznego, wynikiem rozprawy jest oprogramowanie, mające duże walory użytkowe. Jeśli chodzi o formalną stronę pracy, to jest ona zredagowana dosyć niestarannie. W tekście recenzji wskazałem na główne niedostatki. Użyta jest niewłaściwa, można powiedzieć „amatorska” terminologia. Niektóre fragmenty, na przykład te dotyczące podstaw akustyki i słyszenia, są bardzo spłycone. Autor nągminnie nie powołuje się na rysunki zamieszczone w tekście. Wskazałem to w odniesieniu do rys. 4.19, ale brak takich odniesień występuje znacznie częściej. Brak jest jakichkolwiek informacji dotyczących oceny słuchowej. Wady te obniżają niewątpliwie jakość rozprawy, ale jej nie dyskwalifikują. Osiągnięte pozytywne rezultaty przeważają nad wadami. Dlatego też z czystym sumieniem mogę zaproponować ostateczną konkluzję.

Stwierdzam, że rozprawa doktorska mgr. inż. Piotra Kardasza pt. „ Wykorzystanie metod nierównomiernego próbkowania i inteligentnych algorytmów przetwarzania sygnałów do rekonstrukcji archiwalnych nagrań dźwiękowych” spełnia wymagania Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki (Dz. U. 2003 Nr 65 poz. 595 z późniejszymi zmianami) i wnioskuję o jej dopuszczenie do publicznej obrony.

Wrocław, 7.08.2016