

Politechnika Warszawska
Wydział Elektroniki i Technik Informacyjnych

Warszawa, 24 kwietnia 2018 r.

D z i e k a n a t

Uprzejmie informuję, że na Wydziale Elektroniki i Technik Informacyjnych Politechniki Warszawskiej odbędzie się w dniu 5 czerwca 2018 r. publiczna obrona rozprawy doktorskiej

mgr inż. Rafała Romaniuka

temat: „Kompresja mowy z wykorzystaniem aproksymacji rzadkiej”

promotor – dr hab. inż. Przemysław Dymarski, prof. Politechniki Warszawskiej

recenzenci:

prof. dr hab. inż. Ryszard Stasiński z Politechniki Poznańskiej

dr hab. inż. Jacek Misiurewicz, prof. Politechniki Warszawskiej

Obrona odbędzie się w dniu 5 czerwca 2018 r. w sali 116 na Wydziale Elektroniki i Technik Informacyjnych – Gmach im. Janusza Groszkowskiego, Warszawa, ul. Nowowiejska 15/19; początek godz. 9.30.

Po adresem: www.elka.pw.edu.pl/Wydzial/Rada-Wydzialu/Harmonogram-obron-doktorskich-streszczenia-i-recenzje zapewniony jest na stronie Wydziału dostęp do tekstów streszczenia rozprawy i recenzji, jak również do tekstu rozprawy umieszczonej w Bazie Wiedzy Politechniki Warszawskiej.

Dziekan



prof. dr hab. inż. Krzysztof Zaremba

Mgr inż. Rafał Romaniuk

Promotor- dr hab. inż. Przemysław Dymarski, prof. PW

**Tytuł rozprawy doktorskiej: "Kompresja mowy z wykorzystaniem
aproksymacji rzadkiej**

Streszczenie

Niniejsza rozprawa doktorska została poświęcona badaniom koderów predykcyjnych m.in. CELP (ang. *Code-excited Linear Prediction*) i ACELP (ang. *Algebraic Code-excited Linear Prediction*). W szczególności skupiono się na omówieniu i zbadaniu możliwości wykorzystania algorytmów tzw. rzadkiej aproksymacji (ang. *sparse approximation*). W części wprowadzającej do zagadnień rozprawy przedstawiono podstawowe informacje związane z przetwarzaniem sygnału mowy. Szczegółowo opisano działanie wokodera i kodeka CELP.

W części praktycznej przedstawiono algorytmy z rodziny MP (ang. *Matching Pursuit*), oraz zaproponowano nowe ich warianty, cechujące się dobrą jakością modelowania sygnału mowy przy akceptowalnej złożoności obliczeniowej. Przeprowadzono symulacje w celu porównania możliwości modelowania sygnału mowy z wykorzystaniem algorytmów rzadkiej aproksymacji.

W dalszej części pracy porównano algorytmy konstruowania sygnału pobudzającego filtr predykcyjnej syntezy w koderach MP-MLQ (ang. *Multipulse Maximum Likelihood Quantization*) i ACELP. Zaproponowano kilka nowych algorytmów, opartych na równoległym obliczaniu kilku suboptymalnych sygnałów pobudzających i na wymianie pojedynczych impulsów pobudzenia lub podzbiorów tych impulsów. Algorytmy przetestowano na symulatorze kodera o strukturze zbliżonej do standardowego kodera G.723.1, jednak umożliwiającym regulację przepływności binarnej w celu uzyskania lepszej jakości mowy.

W ostatniej części pracy skupiono się na algorytmach generowania pobudzenia impulsowego dla koderów mowy typu CELP i Low Delay CELP o zmiennej przepływności binarnej. Porównano algorytmy wykorzystujące jedno wspólne wzmocnienie dla wszystkich impulsów (MLQ) i osobne wzmocnienia poszczególnych impulsów (MPE – ang. *Multipulse*

Excitation). Ze względu na fakt, że kodery MLQ zapewniają lepszą jakość mowy w zakresie niższych przepływności, a kodery MPE – w zakresie wyższych przepływności, zaproponowano hybrydowe algorytmy generowania pobudzenia. Dzięki zastosowaniu zmodyfikowanych algorytmów dekodowania sfery obliczono optymalne pobudzenia impulsowe dla kodera Low Delay CELP o zmiennej przepływności binarnej. Obecnie nie jest możliwa implementacja algorytmów optymalnych w czasie rzeczywistym, jednak pozwoliło to na zbadanie, jak daleko proponowane w niniejszej pracy rozwiązania realizowalne w czasie rzeczywistym odbiegają od rozwiązań optymalnych.

słowa kluczowe: aproksymacja rzadka, CELP, ACELP, dekodowanie sfery a

**KWESTIONARIUSZ-RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY
WYDZIAŁU ELEKTRONIKI I TECHNIK INFORMACYJNYCH
POLITECHNIKI WARSZAWSKIEJ**

Tytuł rozprawy: „Kompresja mowy z wykorzystaniem aproksymacji rzadkiej”

Autor rozprawy: mgr inż. Rafał Romaniuk

1. Jakie zagadnienie naukowe jest rozpatrzone w pracy /teza rozprawy/ i czy zostało ono dostatecznie jasno sformułowane przez Autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?

Rozprawa doktorska mgr inż. Rafała Romaniuka dotyczy jednego z najczęściej wykorzystywanego w praktyce zastosowania cyfrowego przetwarzania sygnałów, a mianowicie kodowania mowy. Jest to zbiór metod o długiej historii, cyfrowa transmisja mowy była historycznie pierwszą powszechnie stosowaną techniką tej dziedziny wiedzy. Oznacza to, iż najbardziej burzliwy okres rozwoju algorytmów kodowania mowy, a następnie ich standaryzacji, miał miejsce relatywnie dawno temu. Paradoksalnie autor rozprawy znalazł w tym fakcie sposobność dla siebie - dawne i następnie niewiele modyfikowane algorytmy zostały opracowane pod kątem wydajności obliczeniowej ówczesnych systemów cyfrowych. Jednocześnie w międzyczasie rozwinięte zostały techniki aproksymacji rzadkiej, które, jak dowiódł jego promotor, potencjalnie dają możliwość podniesienia jakości, lub podwyższenia stopnia kompresji kodowanej mowy, co prawda kosztem zwiększenia złożoności obliczeniowej kodera. Taki był punkt wyjścia badań opisanych w rozprawie doktorskiej mgr inż. Rafała Romaniuka.

Teza rozprawy jest następująca:

Zaproponowane algorytmy cyklicznego i równoległego przeszukiwania słownika pobudzeń w koderze CELP i ACELP zapewniają zmniejszenie błędu kodowania mowy przy akceptowalnej złożoności obliczeniowej.

Jest ona jasna, precyzyjna i dokładnie odpowiada temu co zrobiono w ramach badań. Rozprawa ma charakter konstrukcyjno-doświadczalny: zaproponowano i zaimplementowano nowe algorytmy koderów mowy, a następnie porównano ich efektywność z efektywnością dotychczasowych rozwiązań dla powszechnie stosowanych do tego celu próbek mowy.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł / w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle/ świadczący o dostatecznej wiedzy autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Bibliografia rozprawy liczy 96 tytułów, w tym 6 prac własnych, 4 są wspólne z promotorem doktoranta. Liczba prac z dziedziny rozprawy jest ogromna, tak więc załączoną literaturę można określić jako minimalistyczną. Niemniej wybór cytowań jest rozsądny, są to prace istotne z punktu widzenia badań. Można dyskutować odnośnie wyboru niektórych pozycji,

niemniej nie ma to większego znaczenia dla oceny doktoratu. Wnioski doktoranta z przeglądu źródeł sprowadzają się do opisu różnych wersji kodera mowy CELP, oraz zastosowanych w rozprawie metod aproksymacji rzadkiej, wskazano podzespoły kodera, których wydajność te metody mogą poprawić. Jak więc widać, wnioski są konkretne i jasno sformułowane.

3. Czy Autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Cel pracy został określony następująco:

Opracowanie algorytmów „rzadkiej” aproksymacji sygnału mowy, umożliwiających zwiększenie dokładności (zmniejszenie błędu aproksymacji) przy akceptowalnej złożoności obliczeniowej.

Cel ten został zrealizowany, jednak z pewną uwagą, co to znaczy akceptowalna złożoność obliczeniowa. Rzeczywiście zrealizowano algorytmy aproksymacji mowy oparte o algorytmy aproksymacji rzadkiej, które mają istotnie mniejszy błąd niż metody tradycyjne. Niemniej ocena czy większa złożoność kodeka jest do zaakceptowania zostanie ocenione dopiero przez twórców praktycznych realizacji. Autor podaje sensowną argumentację wskazując, iż tak powinno się stać: oszacowanie liczby operacji zmiennoprzecinkowych na sekundę dla tych algorytmów (strona 72), oraz dla algorytmów rozmieszczania impulsów (strona 88). Wielkości te są 2 do 4 rzędów mniejsze od teoretycznej wydajności procesorów telefonów, w przypadku niektórych algorytmów rozmieszczania impulsów dysproporcja jest jeszcze większa. Jednakże teoretyczna wydajność procesora jest niemożliwa do uzyskania w praktyce, w dodatku procesory smartfonów są poważnie obciążone coraz to liczniejszymi zadaniami związanymi z różnymi aplikacjami, w tym bardzo kosztownym przetwarzaniem sygnałów wideo dla różnych celów. Oczywiście przynajmniej niektórzy konstruktorzy przywiązują szczególną wagę do jakości przesyłanej mowy, niemniej pojawia się tu inny aspekt sprawy - na ile proponowane udoskonalenia poprawią jej subiektywną ocenę. W pracy jakość kodowanej mowy oceniano na podstawie ogólnie przyjętych wskaźników obiektywnych (SNR, PESQ i POLQA), a nie w oparciu o ocenę odsłuchujących mowę ludzi. Osobne ważne zagadnienie nie poruszane w pracy to zgodność zmodyfikowanych koderów mowy z przyjętymi standardami. Niemniej abstrahując od praktycznego wykorzystania wyników podsumowując ten punkt stwierdzam, że autor rozwiązał postawione zagadnienie, użył do tego skutecznej metody, a przyjęte założenia były uzasadnione.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Jeśli chodzi o oryginalny wkład doktoranta do nauki, to w pierwszym rzędzie trzeba stwierdzić, że jest on prekursorem zastosowania algorytmów aproksymacji rzadkiej do udoskonalania metod kompresji mowy. Podejście to zaowocowało następującymi wynikami:

- Opracowano nowe algorytmy obliczania sygnału pobudzającego filtr predykcyjny w koderze CELP: algorytm M-best wykorzystujący procedurę RMGS i algorytm cyklicznej optymalizacji - rozdział 4.2.
- Opracowano nowe algorytmy dla koderów MP-MLQ i ACELP, są to algorytmy typu M-best wykorzystujące algorytm minimalizacji kąta - rozdział 5.2.
- Opracowano wielostopniowy algorytm skalowania wzmocnień w koderze CELP o zmiennej przepływności – rozdział 6.3.
- Zaproponowano hybrydowy model pobudzenia dla koderów CELP w celu poprawienia ich skalowalności - rozdział 6.4.

- Zaproponowano hybrydowy model pobudzenia w koderze LD-CELP - rozdział 7.2.
- Opracowano zmodyfikowany algorytm dekodowania sfery do obliczania optymalnego pobudzenia w koderze LD-CELP - rozdział 7.3.

Algorytmy te zaimplementowano w ramach opracowanego przez autora zmodyfikowanego kodera CELP bazującego na standardzie ITU-T G.723.1, a następnie przetestowano. Osobnym wynikiem jest analiza szerokiej gamy algorytmów aproksymacji rzadkiej pod kątem ich zastosowania w kompresji mowy, rozdziały 4.3 i 5.3. Zaproponowane rozwiązania są nowatorskie w skali światowej i mają szansę być wykorzystane w praktyce.

5. Czy Autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników /zwięzłość, jasność, poprawność redakcyjna rozprawy/?

Ogólna redakcja rozprawy jest poprawna, choć trochę niestaranna. Prezentacja wyników jest zrobiona jasno i przekonująco, natomiast trudno czyta się wszelkiego rodzaju opisy. Autor ma niechęć do wzorów i do tłumaczenia co kryje się pod fachowymi nazwami. Ma do tego prawo, ale czytania to nie ułatwia. W paru miejscach odwołuje się do pojęć, rysunków i wzorów, które znajdują się wiele stron dalej (przykładowo: pojęcie analizy przez syntezę, rysunki 3.2 i 3.3 anonsowane są już na stronie 29, odwołanie do wzoru (5.4) pojawia się trzy strony wcześniej). Określiłbym to jako źle rozumianą dbałość o zwięzłość rozprawy i to przy powtórzeniach w innych miejscach, jak na przykład w opisach kodera CELP w rozdziałach 1 i 3. Skutek jest taki, że czytając rozdział 3 trzeba wracać do rysunków z rozdziału 1. Uważam zresztą, że rozdział 1 jest niepotrzebnie szczegółowy, zwykle w rozprawach jest to tylko rys historyczny. W rozdziałach 5.2.1 i 5.2.2 doktorant przekroczył granicę dobrych zasad prezentacji przedstawiając kilka autorskich algorytmów w sposób opisowy. W efekcie w dalszych rozdziałach niejasności nawarstwiają się - mamy zdawkowe uwagi odnośnie modyfikacji algorytmów, które opisano w tych dwóch rozdziałach nieformalnie. Warianty algorytmu dekodowania sfery też zostały przedstawione w sposób opisowy. Praca ma 129 stron, a w ramach badań autor zrobił naprawdę dużo, tak więc nic by się nie stało gdyby była ona o kilka-kilkanaście stron dłuższa.

Redakcja tekstu i język rozprawy są na ogół poprawne, choć są też i uchybienia. Z błędów redakcyjnych: rysunek 4.11 jest rozdzielony na dwie strony, przy czym tak samo wygląda on w zbiorze PDF pracy, w tekście jest trochę literówek (np. "of" zamiast "od" na stronie 48), problemów ze spójnikami i przecinkami, źle sformatowanych liter w objaśnieniach do wzorów (zakres "i" na stronie 87 w formacie indeksu dolnego, b_g i α w formacie indeksów na stronach 95 i 96). Katastrofalny jest akapit na górze strony 111: "wyniów" zamiast wyników, a potem niezbyt zrozumiałe następne zdanie. Doktorant nieprawidłowo formułuje stwierdzenia opisujące liczbę bitów użytych do kodowania: na stronie 35 pisze o współczynnikach zapisanych na 41 bitach, co jeszcze można przyjąć, ale mówienie o "kodowaniu w 24" bitach na stronie 93 jest błędem językowym. Powtarza się on dalej w tekście na stronach 93, 95, 100, 104, 105 i 106. Na stronie 57 pojawiają się operacje długie, z wytłumaczeniem, że chodzi o mnożenie i akumulację. To także nie jest ścisłe określenie. W polskim tekście pisząc o funkcji kosinus lepiej używać pisowni z literą "k".

6. Jakie są słabe strony rozprawy i jej główne wady?

Jeśli chodzi o słabe strony rozprawy, to nie są one specjalnie istotne dla jej oceny, jak choćby przedstawione powyżej uwagi do jej strony redakcyjnej. W tym miejscu wróciłbym do sprawy ewentualnego wykorzystania wyników pracy w praktyce. Po pierwsze, brakuje mi tu

oszacowania złożoności obliczeniowej współcześnie stosowanych koderów mowy. Oszacowanie to dałoby punkt odniesienia i lepiej scharakteryzowałoby kwestię akceptowalnej złożoności poruszony w celu pracy. O przemilczeniu sprawy zgodności proponowanych zmian z obowiązującymi standardami już wspominałem, ten czynnik decyduje o tym jak szybko proponowane udoskonalenie mogłyby być wprowadzone do kodeków mowy. Po drugie, bardzo by się przydały choćby wstępne i wybiórcze badania jakości sygnału mowy z udziałem ludzi. Sam autor w pewnym momencie wspomina, że wyniki testu PESQ nie za bardzo zgadzały się z wrażeniami z nieformalnego odsłuchu próbek mowy, tak więc zdecydował się na nowszy test POLQA, też przecież nie idealny.

Na koniec została sprawa zaskakujących błędów przy omawianiu metod znajdowania współczynników predyktora liniowego sygnału. Otóż przejście od wzorów (3.8) i (3.9) do (3.10) nie da się uzasadnić inaczej jak tylko zakładając, że funkcje $c(k,i)$ i $R(m)$ opisują funkcję autokorelacji ergodycznego procesu stacjonarnego. Odwrotnie jak twierdzi autor oznacza to, że posiadamy nieskończenie wiele próbek sygnału do wyznaczenia funkcji $c(k,i)$, tak jak sugerują poprzedzające wzory. Ze wzoru (3.11) wynika jednak, że $R(m)$ jest estymatorem (jednowymiarowej) funkcji autokorelacji, a więc wzór (3.10) nie ma prawa być poprawny niezależnie od tego czy $c(k,i)$ jest estymatorem, czy dokładną postacią funkcji autokorelacji. Prawda wygląda nieco inaczej - w metodzie kowariancyjnej do wyliczenia współczynników predyktora używamy funkcji $c(k,i)$ takiej, jaka nam doświadczalnie wyjdzie. Natomiast w metodzie korelacyjnej popełniamy małe oszustwo - zamiast $c(k,i)$ używamy macierzy Toeplitza o współczynnikach równych próbkom estymatora funkcji autokorelacji $R(m)$. Dół strony 30: "...współczynniki predykcji na początku i końcu ramki obliczone są z pewnym błędem". Przecież współczynniki obliczane są raz dla całej ramki! Zgaduję, że chodzi o to, iż błąd predykcji sygnału w środku ramki jest najmniejszy. Na koniec jeszcze jeden szczegół - podręcznikowe metody rozwiązywania układu równań (3.18) mają złożoność $O(p^3)$, a więc nie ma większego sensu szukanie tutaj innej, jak to autor określił, efektywnej techniki, jeżeli jej złożoność także wynosi $O(p^3)$. Na szczęście dla doktoranta zagadnienie znajdowania współczynników predykcji nie leży w zakresie merytorycznej oceny rozprawy.

7. Jaka jest przydatność rozprawy dla nauk technicznych?

Rozprawa ma potencjalnie duże znaczenie dla nauk technicznych z dwóch powodów. Po pierwsze, kodowanie mowy będzie zawsze istotną funkcjonalnością urządzeń służących komunikacji osobistej. Jeśli więc poprawa jakości sygnału mowy wprowadzana przez metody opisane w rozprawie zostanie uznana za istotną, to nastąpią dalsze badania zakończone ich implementacją w telefonach. Po drugie, praca jest interesująca dla badaczy działających w obszarze aproksymacji rzadkiej: daje ona porównanie wyników zastosowania algorytmów tej dziedziny przy rozwiązywaniu praktycznego problemu o określonej specyfice. Rozprawa zawiera analizę szerokiego spektrum metod aproksymacji rzadkiej.

8. Do której kategorii Recenzent zalicza rozprawę:

Rozprawę zaliczam do kategorii

d) spełniająca z wyraźnym nadmiarem wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy

W związku z powyższym wnoszę o dopuszczenie rozprawy mgr inż. Rafała Romaniuka do publicznej obrony przed Radą Wydziału Elektroniki i Technik Informacyjnych Politechniki Warszawskiej.

R. Romaniuk

Warszawa, dn. 5 grudnia 2017 r.

dr hab. inż. Jacek Misiurewicz, prof. PW
Instytut Systemów Elektronicznych
Wydział Elektroniki
Politechnika Warszawska

KWESTIONARIUSZ- RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY
WYDZIAŁU ELEKTRONIKI I TECHNIK INFORMACYJNYCH
POLITECHNIKI WARSZAWSKIEJ

Tytuł rozprawy: Kompresja mowy z wykorzystaniem aproksymacji rzadkiej

Autor rozprawy: mgr inż. Rafał Romaniuk

1. *Jakie zagadnienie naukowe jest rozpatrzone w pracy /teza rozprawy/ i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?*

Zgodnie z deklaracją Autora, celem pracy było opracowanie algorytmów rzadkiej aproksymacji sygnału mowy, umożliwiających zwiększenie dokładności (zmniejszenie błędu aproksymacji) przy akceptowalnej złożoności obliczeniowej.

W ramach pracy Autor zaproponował całą gamę nowych lub zmodyfikowanych algorytmów aproksymacji pobudzenia w koderach mowy. Zaproponowane algorytmy i modyfikacje oparte są na zaawansowanych podstawach matematycznych, a w ramach pracy zostały zaimplementowane w modelowych koderach i przebadane. W związku z tym charakter pracy należy określić jako teoretyczno-doświadczalny.

Autor postawił tezę, iż zaproponowane algorytmy cyklicznego i równoległego przeszukiwania słownika pobudzeń w koderze CELP i ACELP zapewniają zmniejszenie błędu kodowania mowy przy akceptowalnej złożoności obliczeniowej. Tezę tę uzasadnił wynikami badań przeprowadzonych z wykorzystaniem rzeczywistych nagrań ludzkiej mowy.

2. *Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł / w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle /świadczący o dostatecznej wiedzy autora. Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?*

Przedstawiona przez doktoranta analiza źródeł obejmuje kilka obszarów wiedzy. Zasadniczym obszarem analizy jest tematyka kodowania mowy. Przywoływane w tym obszarze źródła stosowane są głównie do opisu znanych i stosowanych powszechnie koderów, co stanowi bardzo potrzebny wstęp do zagadnień rozpatrywanych w dalszej części pracy.

Drugim analizowanym obszarem wiedzy są problemy aproksymacji rzadkiej, rozpatrywane przez Autora pod kątem ich przydatności do rozpatrywanych zagadnień. Dlatego opisując analizowane problemy Autor posługuje się przykładami związanymi z kodowaniem mowy, co jest w tej pracy jak najbardziej właściwe, gdyż ułatwia inżynierowi praktyczne zrozumienie pojęć wprowadzonych przez matematyków w sposób abstrakcyjny.

3. *Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?*

W mojej ocenie, postawiony w pracy problem opracowania nowych algorytmów aproksymacji pobudzenia w koderach mowy z wykorzystaniem metod aproksymacji rzadkiej został w pełni rozwiązany. Autor zaproponował bardzo wiele nowych algorytmów i przebadał ich praktyczną przydatność. Algorytmy oparte na równoległym obliczaniu kilku suboptymalnych rozwiązań (M-best), na cyklicznej optymalizacji wstępnego rozwiązania, a także algorytmy hybrydowe gwarantują dobrą jakość mowy przy akceptowalnej złożoności obliczeniowej, także w warunkach zmiennej przepływności binarnej.

W celu przebadania zaproponowanych algorytmów Autor zmodyfikował ogólnodostępny symulator kodera mowy tak, aby móc do niego włączać własne rozwiązania. Otrzymał w ten sposób uniwersalne narzędzie, które posłużyło mu do badań przeprowadzanych w stosunku do kolejnych proponowanych rozwiązań. Uważam taką metodę za w pełni właściwą.

Również uzasadnione jest użycie przez Autora metod oceny jakości kompresji poprzez zastosowanie kryteriów MOS LQQ i PESQ, a także dokonanie modyfikacji koderów mających na celu zwiększenie obiektywności w/w kryteriów.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Badania porównawcze metod rzadkiej aproksymacji były wielokrotnie podejmowane w literaturze, jednak prowadzono je w kontekstach innych niż opisany w rozprawie kontekst algorytmów kodowania mowy, choć we wprowadzeniu do pracy Autor wykazuje, że takie zastosowanie jest jak najbardziej naturalne. Dlatego uważam, że przedstawiona rozprawa słusznie wypełnia tę lukę w aktualnym stanie wiedzy.

Można wskazać szereg elementów pracy stanowiących oryginalny dorobek Autora. Polegają one na wykorzystaniu metod rzadkiej aproksymacji do opracowania nowych algorytmów obliczania pobudzenia w różnych koderach mowy.

Autor zaproponował i zbadał wiele algorytmów obliczania pobudzenia w koderach CELP o pobudzeniu wieloimpulsowym i stochastycznym. Na tej podstawie zaproponował algorytm typu M-best, obliczający równolegle kilka suboptymalnych sygnałów pobudzających. Inną propozycją jest algorytm cyklicznej optymalizacji sygnału pobudzającego filtr syntezy. Oba algorytmy cechuje dobra jakość mowy (oszacowana na podstawie wartości SNR w ujęciu segmentowym) i umiarkowana złożoność obliczeniowa (od kilkudziesięciu do kilkuset Mflops).

Po przebadaniu opisanych w literaturze i zaimplementowanych w standardowych koderach mowy algorytmów obliczania pobudzenia w modelach MP-MLQ i ACELP Autor wykazał, że bardzo prosty algorytm „zachłanny”, a mianowicie algorytm minimalizacji kąta (zaproponowany dla kodera CELP już w 1993r.) niewiele ustępuje o wiele bardziej złożonym algorytmom, np. globalnej wymiany impulsów. Na tej podstawie Autor opracował algorytm typu M-best, wykorzystujący wspomniany algorytm minimalizacji kąta który, według Autora, może zostać zarekomendowany jako właściwe rozwiązanie dla konstruktorów koderów mowy wykorzystujących modele MP-MLQ i ACELP.

W zastosowaniu do koderów o pobudzeniu impulsowym i zmiennej przepływności binarnej Autor zaproponował nowe modele pobudzenia filtru syntezy predykcyjnej w koderach CELP i LD-CELP, opierające się na wielostopniowym skalowaniu wzmocnień impulsów i na tworzeniu modeli hybrydowych, łączących dwa różne modele pobudzenia. Wykorzystując koder modelowy w konfiguracji o zmiennej przepływności Autor przetestował zaproponowane modele pobudzenia.

Ponadto Autor zaproponował zastosowanie zmodyfikowanych algorytmów dekodowania sfery do obliczania optymalnego pobudzenia w koderze LD-CELP.

Potwierdzeniem wartości i oryginalności wyników uzyskanych przez doktoranta są trzy artykuły opublikowane w materiałach konferencji EUSIPCO (European Signal Processing Conference) i kolejne trzy – w czasopiśmie Przegląd Telekomunikacyjny- Wiadomości Telekomunikacyjne. Jeden z tych artykułów jest samodzielny, pozostałe są wspólne z promotorem, przy czym w jednym doktorant jest głównym autorem.

5. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników /zwięzłość, jasność, poprawność redakcyjna rozprawy/?

Analiza układu rozprawy oraz jej zawartości wskazuje na: (a) poprawne określenie obszaru badań, (b) solidną podstawę metodyczną prowadzonych badań, uzyskaną dzięki dobremu wprowadzeniu aparatu pojęciowego związanego z metodami kodowania (kompresji) mowy oraz (c) dużą swobodę autora w projektowaniu szerokiej gamy nowych algorytmów obliczania pobudzenia, a następnie w eksperymentalnym badaniu ich właściwości.

Autor bardzo dobrze porusza się w obszarze kodowania mowy i wielu istniejących wariantów koderów z rodziny CELP. Przekonywająco uzasadnia swoje propozycje zastosowania nowych algorytmów obliczania pobudzenia i stosuje poprawne metody oceny ich jakości.

Niestety, poprawność redakcyjna rozprawy przedstawia sporo do życzenia. Już przy wstępnym przeglądaniu pracy rzuca się w oczy jej słaba czytelność i pewien brak staranności redakcyjnej. Cele i teza pracy zdefiniowane są dopiero na 28 stronie rozprawy, nie można ich też znaleźć w streszczeniu. Znacznego wysiłku wymaga też zorientowanie się, gdzie w pracy Autor opisuje aktualny stan wiedzy, a gdzie wprowadza własne, nowatorskie rozwiązania. Niektóre fragmenty tekstu są prawie bez zmian przetłumaczone na język polski z własnych artykułów Autora, a rysunki – bezpośrednio przeniesione, podczas gdy kontekst rozprawy doktorskiej jednak zasadniczo różni się od kontekstu artykułu konferencyjnego.

Sama strona graficzna pracy też budzi duże zastrzeżenia. Autor przyjął nietypową konwencję ujmowania wzorów matematycznych w ramki, co bardzo pogarsza czytelność całości. Rysunki i wykresy wykonane są niespójnie – często nawet w tym samym rozdziale stosowane są zupełnie różne style rysowania zarówno schematów blokowych jak i wykresów wyników eksperymentalnych (np. rys. 4.10 i 4.11).

6. Jakie są słabe strony rozprawy i jej główne wady?

Niewątpliwie, pomimo wymienionych powyżej zalet pracy, można w niej wskazać także słabsze miejsca i pewne elementy warte wyjaśnienia.

Otóż widoczny jest ogrom włożonej przez Autora pracy badawczej, natomiast nie widać w rozprawie podążania prostą ścieżką do wytkniętego celu. Autor bada dziesiątki różnych algorytmów, zazwyczaj już znanych z literatury, i do prawie każdego proponuje większe lub mniejsze modyfikacje. Na podstawie badań porównawczych wyciąga wnioski o przewadze jednych algorytmów nad drugimi, ale trudno się w tekście doszukać końcowej rekomendacji. Takie rekomendacje przedstawione w czytelny sposób można znaleźć dopiero w podsumowaniu pracy.

Nieco brakuje też w rozprawie odniesienia określonej przez Autora złożoności obliczeniowej do aktualnych możliwości procesorów sygnałowych stosowanych w różnych klasach urządzeń kompresujących mowę. Takie odniesienie ułatwiłoby ocenę, które z proponowanych rozwiązań są już możliwe do zastosowania, a które mają szansę być wdrożone w bliskiej przyszłości.

Zauważalną wadą rozprawy jest wspomniana już słaba czytelność i brak staranności redakcyjnej, co utrudnia jej odbiór przez czytelnika.

Drobne uchybienia można znaleźć w istotnym dla pracy rozdziale dotyczącym wprowadzenia do zagadnień aproksymacji rzadkiej. Pokazując przykład słownika, w którym niemożliwe jest otrzymanie zerowej koherencji Autor nie wyjaśnia, czemu do wygenerowania słownika używa generatora liczb pseudolosowych, zupełnie pomijając w ten sposób ważny fragment teorii matematycznej dotyczącej aproksymacji rzadkiej. Jednocześnie nieściśle określa generowane liczby jako „liczby losowe” (str. 52).

Wyjaśnienia wymaga też, czemu Autor unika stosowania pojęcia iloczynu skalarnego lub współczynnika korelacji, posługując się zamiast tego pojęciem kąta między wektorami, co, choć nie jest nieściśle, nadaje pracy charakter popularny (str. 51 i n.).

W innym miejscu (str. 65) zamiast „iloczyn skalarny” użyte jest określenie „skalar” - być może jest to po prostu przeoczenie.

W całej pracy Autor na przemian zapisuje słowo „rzadki” w cudzysłowach lub bez, i nie widać przyczyny takiego różnicowania.

Należy zauważyć, że wskazane usterki edycyjne oraz formalne nie wpływają w istotny sposób na moją merytoryczną ocenę pracy jako całości – Autor wykonał znaczną ilość pracy badawczej uzyskując ciekawe i przydatne praktyczne wyniki, które opisał w sposób co najmniej akceptowalny.

7. Jaka jest przydatność rozprawy dla nauk technicznych?

Wyniki uzyskane w pracy mają duże praktyczne znaczenie. Można się spodziewać, że proponowane metody wyznaczania pobudzenia zostaną zaimplementowane w nowych opracowaniach koderów mowy, umożliwiając poprawę jakości kodowania.

Ponadto rozprawa pokazuje nowy, nie opisany dotąd w literaturze światowej obszar zastosowań dla zyskującej coraz większą popularność techniki modelowania rzadkiego.

8. Do której z następujących kategorii Recenzent zalicza rozprawę?

Tematyka rozprawy jest aktualna i ważna. Autor rozwiązał postawiony przez siebie problem naukowy, używając do tego celu nowoczesnych metod. Rozprawa ukazuje szeroką wiedzę Autora w zakresie metod kodowania mowy i umiejętność wprowadzania w tym zakresie nowych rozwiązań.

Uważam, że autor zrealizował cel rozprawy oraz wykazał się umiejętnościami i odpowiednim przygotowaniem do samodzielnej pracy naukowej w dyscyplinie Telekomunikacja. Na tej podstawie stwierdzam, że przedstawiona do oceny rozprawa doktorska mgr. inż. Rafała Romaniuka pt. „Kompresja mowy z wykorzystaniem aproksymacji rzadkiej” spełnia wymagania stawiane rozprawom doktorskim w Ustawie o stopniach naukowych i tytule naukowym z dnia 14 marca 2003 roku (Dz. U nr 65/2003, poz. 595 z późn. zm.) i wnoszę o dopuszczenie jej Autora do publicznej obrony.

