Dr hab. inż. Andrzej P. Dobrowolski
Wydział Elektroniki Wojskowej Akademii Technicznej
Dane kontaktowe: *Andrzej.Dobrowolski@wat.edu.pl*

12.02.2017 r.

Recenzja rozprawy doktorskiej

Radosława Weychana

zatytułowanej:

Speaker recognition based on transcoded speech for human-machine interfaces

Recenzja została sporządzona z inicjatywy Dziekana Wydziału Informatyki Politechniki Poznańskiej Pana dr. hab. inż. Andrzeja Jaszkiewicza, prof. nadzw., na podstawie Umowy nr 09/93/2016/54 z dnia 23.12.2016 r.

1. Problem badawczy i jego znaczenie

Problem automatycznego rozpoznawania mówcy jest bardzo ważny i aktualny zarówno z naukowego, jak i z praktycznego punktu widzenia, gdyż rozpoznawanie mówcy znajduje coraz szersze zastosowanie w wielu dziedzinach ludzkiej aktywności, np. w kontroli dostępu, kryminalistyce, głosowym sterowaniu urządzeniami technicznymi itp. Z tego powodu od wielu lat, w kraju i za granicą, w wielu ośrodkach naukowych prowadzone są prace badawcze nad udoskonalaniem istniejących metod rozpoznawania mówcy oraz poszukiwaniem metod nowych i doskonalszych. W tym kontekście wybór tematyki rozprawy uważam za właściwy i odpowiadający wymaganiom aktualności tematyki badawczej.

Recenzowana rozprawa doktorska powstała w zespole Pana Profesora Adama Dąbrowskiego, który od wielu lat zajmuje się m.in. projektowaniem coraz lepszych systemów rozpoznawania mówcy. Systemy te charakteryzują się dużą różnorodnością, głównie w zależności od obszarów zastosowań. Autor rozprawy podjął się poprawienia skuteczności rozpoznania mówcy w zastosowaniu do sterowania systemami człowiek-komputer. Takie zastosowanie wymusza optymalizację systemu rozpoznawania mówcy dla bardzo krótkich wypowiedzi, co rodzi szereg dodatkowych trudności.

Celem rozprawy doktorskiej mgr. inż. Radosława Weychana jest udowodnienie tezy, że *efektywność rozpoznawania mówcy bazująca na sygnałach niskiej jakości (szybkość próbkowania 8 kS/s i 10-cio bitowa rozdzielczość amplitudowa) może być poprawiona za pomocą opracowanych algorytmów detekcji aktywności głosowej oraz metody kodowania, a także za pomocą właściwego wyboru modelu z bazy mówców*. Po analizie treści rozprawy mogę stwierdzić, że cel ten został osiągnięty i teza rozprawy została udowodniona.

W celu udowodnienia tezy rozprawy Autor zdefiniował cele szczegółowe obejmujące badania wpływu skracania wypowiedzi oraz stosowania różnych technik detekcji aktywności głosowej i rodzajów kodowania na efektywność typowych systemów rozpoznawania mówcy, a także – w aspekcie praktycznym – badanie wpływu arytmetyki stało i zmiennoprzecinkowej w celu implementacji opracowanych metod w autonomicznym systemie z procesorem sygnałowym.

Rozprawa ma charakter eksperymentalny z wyraźnym rysem aplikacyjnym o dużym znaczeniu praktycznym, a podjęty w niej problem badawczy, aktualny zarówno z naukowego, jak i z technicznego punktu widzenia oceniam pozytywnie.

2. Wkład autora

Do najważniejszych, przedstawionych w rozprawie, osiągnięć Autora można zaliczyć:

- projekt i przeprowadzenie eksperymentu z wykorzystaniem różnych krótkich wypowiedzi w celu oceny przydatności modelowania głosu mówcy za pomocą kwantyzacji wektorowej i mieszanin gaussowskich;
- wybór optymalnej dla projektowanego systemu metody detekcji aktywności głosowej;
- ocenę wpływu zastosowanych kodeków mowy na skuteczność systemów rozpoznawania mówcy w różnych kombinacjach oraz wykorzystanie autorskiej metody detekcji kodeka w celu poprawy skuteczności rozpoznawania;
- implementację opracowanych metod rozpoznawania mówcy w systemach czasu rzeczywistego wykorzystującego stałoprzecinkowe procesory sygnałowe.

Z punktu widzenia zastosowań projektowanego systemu dość wątpliwa wydaje się sensowność przeprowadzenia testów z głosem kodowanym. Należy spodziewać się, że komendy głosowe będą w praktycznej aplikacji wydawane bezpośrednio, tzn. sygnał na wejście systemu rozpoznawania mówcy będzie pobierany bezpośrednio z toru mikrofonowego. Ostatecznie można wyobrazić sobie jeszcze wersje systemu sterowania głosem przez telefon GSM lub VoIP – na takie zastosowania właściwie wskazuje sam tytuł rozprawy, ale analiza wpływu kodeków audio (MP3, WMA,...) jest wg mnie nadmiarowa.

Z praktycznego punktu widzenia – jeśli celem jest interfejs człowiek-maszyna, w sensie interpretacji krótkich rozkazów w celu sterowania np. wózkiem inwalidzkim, czy telefonem komórkowym podczas jazdy samochodem – trzy sekundowy czas trwania wypowiedzi jest dość dyskusyjny. Z jednej strony jest to czas bardzo krótki, szczególnie w przypadku ludzi mówiących powoli, i wyniki rozpoznawania osiągnięte przez Autora można uznać za bardzo dobre, jednak z drugiej strony rozkazy typu: *odbierz*, *w lewo*, *w prawo*, *stój*, *start* itp. trwają jednak znacznie krócej, zatem wykorzystanie zaprojektowanego systemu w tej roli może być wątpliwe.

3. Poprawność

Postawiona w rozprawie doktorskiej teza, że *efektywność rozpoznawania mówcy ... może być poprawiona* została udowodniona (Autor odnotował nawet kilkunastoprocentową poprawę współczynnika błędu zrównoważonego) poprzez prawidłowo zaplanowane badania eksperymentalne opisane w rozdziałach trzecim i czwartym. Doświadczenia przeprowadzone z użyciem różnych baz głosów pozwoliły Autorowi na wybór optymalnych dla rozważanego zagadnienia sposobów modelowania oraz dobór i optymalizację algorytmów pomocniczych (detekcja aktywności głosowej, detekcja kodeka) poprawiających skuteczność zaprojektowanego i zaimplementowanego sprzętowo systemu rozpoznawania mówcy. Eksperymenty przedstawione w rozprawie są dobrze opisane, a kolejne wyniki prawidłowo interpretowane i wykorzystywane w kolejnych etapach.

Rozprawa doktorska mgr. inż. Radosława Weychana została napisana w sposób przejrzysty, z poprawną kolejnością omawiania zagadnień i wyników. W części teoretycznej nie znalazłem istotnych błędów analitycznych, chociaż występują pewne drobne nieścisłości merytoryczne, np. na stronie 14 Autor pisze, że K oznacza rozdzielczość transformaty, podczas gdy faktycznie jest to liczba próbek w dziedzinie częstotliwości (podobnie na stronie 17). W części eksperymentalnej również można natknąć się na przekłamania, np. na str. 66 Autor pisze, że w przypadku wcześniej nieprzetworzonego sygnału błąd będzie mniejszy niż w przypadku sygnału uprzednio zakodowanego, podczas gdy w istocie jest przeciwnie.

Recenzowana rozprawa została przygotowana w języku angielskim z wysoką starannością i dbałością o poziom edytorski, tym nie mniej Autor nie ustrzegł się błędów edycyjnych i językowych. Przykładowo:

- na str. 1 jest „smart phones", a powinno być „smartphones";
- na str. 3 jest „... system is use of the telephone network.", a powinno być „...is used...";
- na str. 6 jest „...effective speaker techniques", a powinno być „...effective speaker recognition techniques";
- na str. 11 jest *sampleng* zamiast *sampling*;
- na str. 22 jest „... na rys. 2.11...", a powinno być „...na rys. 2.12..."
- na str. 27 brakuje spacji we fragmencie: „...value.An...";
- ...
- na str. 164 jest „... 8 kSps and 16 kSps ...", ale na rys. 5.1. jest: 8 kSps i 22.050 kSps.

Uważam, że większości tych błędów można by uniknąć, gdyby praca nie była tak obszerna (ok. 200 stron) i obejmowała tylko kwestie istotne dla udowodnienia tezy rozprawy, a nie raportowała wszystkie prace autora w obszarze rozpoznawania mówców. Przytoczone usterki nie obniżają jednak zasadniczej wysokiej wartości merytorycznej rozprawy.

4. Wiedza kandydata

W rozdziałach pierwszym i drugim Autor dokonuje zasadniczego przeglądu literatury i prezentuje istniejący stan wiedzy w obszarze rozprawy wpisujący się w dyscypliny naukowe *Informatyka* oraz *Automatyka i Robotyka*. Poza zarysem ogólnym Autor prezentuje zagadnienia związane z mechanizmem powstawania ludzkiej mowy, analizą sygnałów, w szczególności analizą spektralną, cepstralną orasz mel-cepstralną oraz metodami klasyfikacji (*k*-średnich, kwantyzacji wektorowej, mieszanin gaussowskich). W mojej opinii opis jest profesjonalny i świadczy o biegłości Autora w tym obszarze oraz głębokim zrozumieniu prezentowanych zagadnień.

Krajowa i zagraniczna literatura naukowa dotycząca zagadnień przetwarzania sygnału mowy i w szczególności rozpoznawania mówcy jest bardzo rozległa, a ostatnio można zaobserwować rosnący trend liczby publikacji poświęconych tej tematyce. Autor dokonał trafnej selekcji źródeł ograniczając się do 143 reprezentatywnych pozycji, z których wiele to pozycje z ostatnich kilku lat, co świadczy o dobrej orientacji Autora w omawianej problematyce. Cytowania dotyczą literatury polskiej i zagranicznej oraz, co warto podkreślić, prac własnych ze znacznym udziałem Autora rozprawy, z których pierwsze zostały opublikowane w 2010 r. Autor formułuje poprawne wnioski z analizy źródeł wykazując zasadność realizacji celów wskazanych w rozprawie.

5. Inne uwagi

Za oryginalny dorobek Autora uważam całość osiągnięć związanych z problematyką dotyczącą rozpoznawania mówców w oparciu o krótkie wypowiedzi rejestrowane z niską jakością. Wprawdzie Autor zastosował klasyczne narzędzia badawcze i obliczeniowe, ale dokonał ich umiejętnego połączenia w kompletny i poddany weryfikacji system rozpoznawania mówców. W wyniku przeprowadzonych badań opracował oryginalny algorytm przetwarzania sygnałów, pozwalający uzyskać wysoką skuteczność rozpoznawania mówcy w utrudnionych warunkach. Wobec powyższego uważam, że Autor wniósł konkretny i samodzielny wkład do dyscyplin naukowych *Informatyka* oraz *Automatyka i Robotyka* w zakresie analizy sygnału mowy.

6. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez artykuł 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami) moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problem naukowego? (wybierz jedną opcję stawiając znak **X**)

[X]	[ ]	[ ]	[ ]	[ ]
Zdecydowanie TAK	*Raczej TAK*	*Trudno powiedzieć*	*Raczej NIE*	*Zdecydowanie NIE*

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka lub Automatyka i Robotyka?

[X]	[ ]	[ ]	[ ]	[ ]
Zdecydowanie TAK	*Raczej TAK*	*Trudno powiedzieć*	*Raczej NIE*	*Zdecydowanie NIE*

C. Czy kandydat ma umiejętność samodzielnego prowadzenia pracy naukowej?

[X]	[ ]	[ ]	[ ]	[ ]
Zdecydowanie TAK	*Raczej TAK*	*Trudno powiedzieć*	*Raczej NIE*	*Zdecydowanie NIE*

Na podstawie przeprowadzonej analizy stwierdzam, że rozprawa doktorska mgr. inż. Radosława Weychana pt. *Speaker recognition based on transcoded speech for human-machine interfaces* spełnia wymagania określone przez obowiązujące przepisy i dlatego wnioskuję o dopuszczenie jej do publicznej obrony.

Podpis