Poznań, 23.05.2016 r.

Prof. dr hab. inż. Adam Dąbrowski
Politechnika Poznańska
Wydział Informatyki
Katedra Sterowania i Inżynierii Systemów
Pracownia Układów Elektronicznych
i Przetwarzania Sygnałów

OCENA

rozprawy doktorskiej pt.: „*Wykorzystanie szerokopasmowej matrycy wielomikrofonowej w rozpoznawaniu mówcy*"
Pana **mgra inż. Rafała Samborskiego**

1 Ocena doboru tematu i zakresu przeprowadzonych badań

Oceniana rozprawa doktorska powstała w zespole Pana Profesora Mariusza Ziółki w Katedrze Elektroniki Wydziału Informatyki, Elektroniki i Telekomunikacji Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie, czyli pod opieką wybitnego — i to nie tylko w skali kraju, ale także w skali międzynarodowej — specjalisty w dziedzinie analizy i przetwarzania sygnału mowy. Rozprawa ta doskonale wpisuje się w nurt dotychczasowych prac z tej dziedziny, opracowanych w tym zespole, porządkując i w wielu aspektach rozszerzając uzyskane do tej pory wyniki, dotyczące rozpoznawania mówcy.

Tematem rozprawy jest zbadanie możliwości poprawy skuteczności rozpoznawania mówcy spośród kilku mówców podczas ich rozmowy w stacjonarnych warunkach akustycznych (np. podczas zebrania w sali konferencyjnej) przy dodatkowym zastosowaniu systemu wielomikrofonowego do oceny kierunków a nawet obszarów nadchodzenia poszczególnych sygnałów dźwiękowych. Doktorant zakłada, że mówcy nie przemieszczają się w trakcie rozmowy a w pomieszczeniu nie zmieniają się warunki akustyczne, w tym nie zmienia się liczba mówców.

Mimo iż zarówno rozpoznawanie mówcy jak i zastosowania systemów wielomikrofonowych w akustyce technicznej należą do stosunkowo dobrze rozpoznanych zagadnień naukowych, to tematykę rozprawy oraz zakres przeprowadzonych badań oceniam wysoko. Szczegółowe zbadanie wpływu uwzględniania kierunków nadchodzących dźwięków na skuteczność rozpoznawania mówcy jest bowiem w pełni aktualnym problemem naukowym a ponadto zagadnieniem o dużym znaczeniu aplikacyjnym.

2 Ocena celu badań i sformułowanej tezy naukowej

W punkcie 1.2 pt. „Motywacja pracy", który jest częścią wstępu (pierwszego rozdziału rozprawy) Doktorant uzasadnił istotność podjętych badań, mających na celu poprawę jakości komunikacji głosowej za pomocą systemów wielomikrofonowych i na stronie 15 sformułował następującą tezę naukową:

„Wykorzystanie kilku strumieni cech w znaczący sposób polepsza skuteczność działania systemu diaryzacji nagrań. Poprzez dynamiczny dobór proporcji pomiędzy informacją pochodzącą z klasycznego systemu identyfikacji mówcy opartego o cechy częstotliwościowe (MFCC) a informacją związaną z lokalizacją mówcy (TDOA) następuje znacząca poprawa wyników algorytmu w stosunku do istniejących rozwiązań."

Zarówno cele badań (które można wywieźć z tekstu, choć nie zostały jawnie sformułowane), jak i podaną tezę naukową, a także objęty rozprawą zakres badań i eksperymentów oceniam bardzo pozytywnie. Uważam, że Pan mgr inż. Rafał Samborski podjął się opracowania ciekawego i ważnego zadania polegającego na zautomatyzowanej diaryzacji wielomikrofonowych nagrań rozmów prowadzonych przez kilka osób w stacjonarnych warunkach akustycznych i przedstawił nowoczesne oraz przyszłościowe rozwiązanie techniczne tego problemu.

3 Ocena tekstu rozprawy i sposobu przedstawienia wyników

Recenzowana rozprawa jest zwięzła (ma 94 strony tekstu) i oprócz umieszczonych na samym początku podziękowań oraz spisu treści zawiera: spis rysunków, spis tablic, wykaz skrótów i oznaczeń, 8 rozdziałów stanowiących główną część tekstu, w tym wstęp (rozdział 1) oraz podsumowanie (rozdział 8), a także bibliografię.

We wstępie, o czym już pisałem, Doktorant uzasadnił istotność podjętych badań, określił ich przedmiot oraz sformułował tezę naukową. W rozdziale drugim opisał technologię matryc wielomikrofonowych. Rozdział 3-ci został poświęcony lokalizacji źródeł akustycznych, rozdział 4-ty — filtracji adaptacyjnej, a rozdział 5-ty — kształtowaniu wiązki.

Główną część rozprawy, w której Pan mgr inż. Rafał Samborski przedstawił oryginalne wyniki swoich badań i eksperymentów, stanowią rozdziały: 6-ty, pt.: „Matryce wielomikrofonowe w rozpoznawaniu mówcy" i 7-my, pt. „Wyniki eksperymentów".

Kompozycja rozprawy jest właściwa. Doktorant z dużą skrupulatnością opisał zarówno dotychczasowy stan badań jak i własne osiągnięcia. Język rozprawy jest poprawny, choć niektóre rozważania matematyczne nie zostały przeprowadzone w pełni prawidłowo i z zachowaniem pełnego formalizmu, wymaganego zazwyczaj w pracach badawczych dotyczących nauk technicznych.

Reasumując, tekst rozprawy (poza drobnymi i stosunkowo nielicznymi błędami oraz niedociągnięciami, o których piszę w uwagach szczegółowych i które nie zmniejszają ogólnego pozytywnego wrażenia) a także sposób przedstawienia wyników zasługują na ocenę pozytywną.

4 Ocena uzyskanych wyników

Zamieszczone w rozprawie wyniki prac, w których upatruję istotny wkład jej Autora, to:

- propozycja określania położenia mówcy za pomocą systemu cztero-mikrofonowego przy użyciu wektora zdefiniowanego wzorem (6.13), złożonego z trzech opóźnień względem par kolejnych mikrofonów: pierwszego z drugim, drugiego z trzecim i trzeciego z czwartym

- pomysł fuzji klasycznych cech MFCC's (mel frequency cepstral coefficients) mówców z parametrami ich położeń do tworzenia modeli typu GMM-UMB (Gaussian mixture model – universal background model)

- opracowanie opisanego w p. 6.6 rozprawy hybrydowego algorytmu klasyfikacji mówcy w oparciu o ważony według zależności (6.15) logarytmiczny wskaźnik wiarygodności (log-likelihood ratio)

- zbadanie optymalnych wartości współczynnika wagowego, sterującego przebiegiem tego algorytmu w zależności od stosunku sygnału do szumu dla wybranych typów szumu (szumu białego i szumu typu „speech babble" (rysunek 7.4).

Oceniając te wyniki uważam, że są one w pełni oryginalne i bardzo ważne dla dalszego rozwoju technologii rozpoznawania mówców i diaryzacji nagrań. Ich Autor wykazał się pomysłowością i znajomością rozpatrywanej problematyki, a także pracowitością, systematycznością i konsekwencją w pracy.

5 Wybrane uwagi szczegółowe

Poniżej zebrałem niektóre uwagi szczegółowe, które nasunęły mi się podczas czytania tekstu rozprawy:

- str. 9 i 10: zamieszczony w rozprawie wykaz skrótów i oznaczeń jest niepełny i zawiera błędy, np. σ to nie jest wariancja, jak podaje Doktorant, a odchylenie standardowe, przez E oznaczono zarówno charakterystykę sensora dźwięku jak również błąd średniokwadratowy, G to nie transformacja Fouriera funkcji R, jak podaje Autor, a ewentualnie transformata Fouriera, h to nie współczynniki filtru a ewentualnie współczynnik filtru, m to albo sensor dźwięku albo sygnał zebrany przez sensor, a nie jedno i drugie, podobnie s to albo źródło sygnału albo sygnał emitowany przez to źródło, a nie jedno i drugie

- str. 15 (wiersz 15^{g}): Doktorant w całym tekście pracy posługuje się pojęciem „mikstury gaussowskie" — moim zdaniem określenie „mikstura" o pejoratywnym zabarwieniu chemiczno-farmaceutycznym powinno być zastąpione neutralnym i powszechnie w tym kontekście stosowanym określeniem „mieszanina"

- str. 19 (wiersz 5_{d}): kąty θ i ϕ w zależności (2.1) nie zostały zdefiniowane, wskazane jest zamieszczenie odpowiedniego rysunku ilustrującego definicję tych kątów

- str. 20 (wiersz 8_{d}): wielkość x_a we wzorze (2.3) i dalszych nie została zdefiniowana, wskazane jest zamieszczenie odpowiedniego rysunku ilustrującego

- str. 21 (wiersz 8_{d}): emocjonalne określenie „niezwykle" nie powinno być stosowane w tekstach naukowych

- str. 22 (wiersz 8_{d} i dalsze): wielkość $e_n(f)$ określona jako charakterystyka częstotliwościowa nie występuje w wykazie oznaczeń

- str. 22 (wiersz 4_{d} i dalsze): wielkość k_x nie została zdefiniowana i nie występuje w wykazie oznaczeń

- str. 23 (wiersz 7_{d}): Autor stosuje określenie „chwile czasowe", czy istnieją „chwile nie czasowe"?

- str. 23 (wiersz 5_{d}): zamiast „Niquista" powinno być „Nyquista"

- str. 24: wszystkie charakterystyki pokazane na rysunku 2.7 mają przy pewnych kątach dziwne „wżery" o wartościach ok. 3 dB, świadczące o niezidentyfikowanych błędach pomiarowych — to dziwne zjawisko powinno być wyjaśnione w tekście rozprawy

- str. 25 (wiersz 9_{d}): zamiast „ilość wymiarów" powinno być „liczbę wymiarów"

- str. 29 i 30: w przeprowadzonych rozważaniach założono model w dziedzinie czasu ciągłego (wzory (3.3)) a kolejne zależności (3.4) i (3.5) dotyczą już sygnałów czasu dyskretnego, ponadto jeśli przyjmuje się definicję (3.4), to skąd w wyrażeniu (3.5) pojawiły się argumenty $\tau + \tau_0$ zamiast $\tau - \tau_0$?

- str. 30 (wiersz 5^{g}): zależność (3.4), wbrew temu co podaje Doktorant, nie jest splotem, splot sygnałów wymaga bowiem odwrócenia jednego z nich w osi czasu

- str. 34 (wiersz 9_{d}): rysunek 3.3 nie przedstawia „architektury filtru adaptacyjnego" a „schemat blokowy z filtrem adaptacyjnym"

- str. 35 (rys. 3.4): uzyskana przez Doktoranta zbieżność filtru adaptacyjnego dopiero po ok. 1 sekundzie przy zaledwie 100 współczynnikach (rys. 3.5), to wynik, który moim zdaniem należy ocenić jako niezadowalający

- str. 35 (wiersz 1 i 2_{d} i dalsze): Doktorant stosuje niepoprawne określenie „częstotliwość próbkowania" zamiast poprawnego „szybkość próbkowania" i mierzy ją w Hz zamiast w próbkach na sekundę (S/s)[1]

- str. 39 (rys. 4.1): na tym rysunku należy wyraźnie zaznaczyć sygnał wyjściowy filtru

- str. 51: puste miejsce pod rys. 5.1 powinno być zagospodarowane np. powinno być przeznaczone na tekst

- str. 78 i 79: na rys. 7.4 brakuje podania znaczenia wartości w dB przypisanych do poszczególnych krzywych (można domyślać się, że są to wartości SNR), ponadto wykresy

[1] Jestem przeciwnikiem stosowania tej nazwy i podawania jej wartości w hercach (Hz). Jest to niepoprawne i prowadzi do nieporozumień. Należy bowiem rozróżniać pojęcia: „częstotliwość" (jako właściwość sygnału) i „szybkość" jako właściwość procesu wykonywanego nad sygnałem (w omawianym przypadku procesu próbkowania) i powstałego strumienia danych. Dlatego należy wyłącznie stosować określenie „szybkość próbkowania" i należy ją mierzyć w próbkach na sekundę (S/s), podobnie jak szybkość binarnego strumienia danych mierzy się w bitach na sekundę (b/s), a nie w hercach. Dzięki temu można jednoznacznie rozróżniać powszechnie stosowane pojęcia: „częstotliwość Nyquista" i „szybkość Nyquista". Przez częstotliwość Nyquista rozumie się maksymalną częstotliwość widma sygnału dolnopasmowego, który można próbkować z określoną szybkością, a przez szybkość Nyquista — minimalną szybkość, z którą można próbkować sygnał dolnopasmowy o określonej maksymalnej częstotliwości widma. Skutkiem stosowania niepoprawnego terminu „częstotliwość próbkowania" jest niejednoznaczność pojęcia „częstotliwość Nyquista". Na usprawiedliwienie Doktoranta powinienem jednak dodać, że posługiwanie się „częstotliwością próbkowania" i mierzenie jej w hercach jest bardzo powszechną „manierą", choć na szczęście zanikającą, nie tylko w pracach polskojęzycznych. Rolą osób zajmujących się profesjonalnie przetwarzaniem sygnałów, do których należy przecież Doktorant, powinno być jednak jej zwalczanie, a nie rozpowszechnianie.

optymalnych wartości parametru α przedstawione na rys. 7.5 nie obejmują pełnego zakresu SNR przeanalizowanego na rys. 7.4 — uwzględnienie SNR od $-10\,\mathrm{dB}$ zamiast od $5\,\mathrm{dB}$, istotne ze względów poznawczych, jeśli nawet bez większego znaczenia technicznego, drastycznie zmieniłoby przebieg krzywych na rys. 7.5 i wymagałoby uzupełnienia zależności (7.3), w której błędnie określono zakres SNR $\in (-\infty, 13)\,\mathrm{dB}$

- str. 84 (tablica 7.2): szkoda, że Doktorant nie przeprowadził testów z bazą NIST, co utrudnia rzetelne porównania opracowanego algorytmu z wynikami uzyskanymi w innych ośrodkach naukowych
- str. 85 (wiersz 10^{g}): zamiast „ilość" powinno być „liczba"
- str. 86 (wiersz 1-4_{d}): tekst pracy zamyka uwaga wskazująca na istotną wadę aktualnego algorytmu, polegającą na konieczności „ręcznego" wskazywania chwil, w których poszczególne osoby kończą wypowiedzi — ta wada powinna być w przyszłości wyeliminowana np., jak pisze Autor, za pomocą metod uczenia nienadzorowanego.

6 Konkluzja

Podsumowując powyższą charakterystykę recenzowanej rozprawy doktorskiej stwierdzam, że Pan mgr inż. Rafał Samborski zamieścił w niej bardzo wartościowe, oryginalne wyniki prac przeprowadzonych przez Niego pod kierunkiem Promotora rozprawy Pana Profesora Mariusza Ziółki. Oceniam, że Doktorant osiągnął zakładane cele badawcze i wykazał prawdziwość sformułowanej w rozprawie tezy naukowej. Wymienione przeze mnie w poprzednim punkcie szczegółowe uwagi krytyczne mają znaczenie drugorzędne i nie zmniejszają wartości osiągniętych i przedstawionych w rozprawie wyników.

Na zakończenie stwierdzam, że przedłożona praca spełnia wymagania stawiane przez stosowne przepisy rozprawom doktorskim. Uważam zatem, że Doktorant powinien być dopuszczony do dalszych etapów procedury doktorskiej i do publicznej obrony.