

prof. dr hab. inż. Marek Domański

Instytut Telekomunikacji Multimedialnej

Politechniki Poznańskiej

ul. Polanka 3, 60 965 Poznań

Poznań, 3 lipca 2026r.

RECENZJA

rozprawy doktorskiej mgr. inż. Stanisława Gmyrka

pt. „Propozycje algorytmów parametryzacji, odpornych na zmienność tonu krtaniowego, dla automatycznego rozpoznawania mowy”

Rozprawa doktorska pana Stanisława Gmyrka jest zwięźle napisana w języku polskim, liczy 98 stron łącznie z bogatym spisem literatury (114 pozycji). Należy pozytywnie ocenić zamieszczenie wykazu skrótowców użytych w rozprawie. Tekst rozprawy został uzupełniony trzema dodatkami opisującymi wykorzystane algorytmy oraz własne wyniki eksperymentalne dotyczące rozpoznawania fonemów.

W rozdziale pierwszym zwięźle przedstawiono znaczenie podjętego tematu, tezę pracy i układ pracy. Praca dotyczy efektywnego rozpoznawania mowy, a w szczególności uniezależnienia procesów rozpoznawania fonemów i całych wyrazów od częstotliwości tonu krtaniowego. Celem pracy jest sformułowanie metod uniezależniających automatyczne wyznaczanie parametrów mowy od częstotliwości tonu krtaniowego i jej zmienności. Należy stwierdzić, że tak określona tematyka rozprawy jest bardzo aktualna i ważna ze względu na praktyczne zastosowania w coraz częściej używanych systemach automatycznego rozpoznawania mowy.

Rozdział 2 przedstawia bardzo dobrze uporządkowany i treściwy przegląd wiedzy i literatury na temat parametryzacji sygnałów w systemach automatycznego rozpoznawania mowy. Ten rozdział dowodzi szerokiej wiedzy autora i jego znacznych umiejętności posługiwania się zaawansowanymi metodami przetwarzania sygnałów.

Aczkolwiek oceniam, że rozdział jest dobrze napisany, to dostrzegłem w nim drobne niedociągnięcia.

Przegląd literatury obejmuje sporą liczbę prac, a autor oczywiście nie może wymienić wszystkich prac dotyczących omawianej tematyki. Jednak wydaje się, że na przykład warto było

RDNAEETK/07137/2026

odnieć się przynajmniej do dwóch dodatkowych prac dosyć blisko związanych z tematyką rozprawy:

1. R. D. Zilca, B. Kingsbury, J. Navratil and G. N. Ramaswamy, "Pseudo Pitch Synchronous Analysis of Speech With Applications to Speaker Recognition," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 467-478, March 2006, doi: 10.1109/TSA.2005.857809.
2. M. Cutajar, E. Gatt, I. Grech, O. Casha and J. Micallef, "A study on pitch variation on the use of DWT with SVM for speaker independent phoneme recognition," *2012 5th International Symposium on Communications, Control and Signal Processing*, Rome, Italy, 2012, pp. 1-5, doi: 10.1109/ISCCSP.2012.6217773.

Ponadto w rozdziale drugim zauważyłem kilka usterek redakcyjnych:

1. We wzorze (2.14) brakuje nawiasów.
2. Po wzorze (2.15) brakuje objaśnienia wielkości e ,
3. W szeregu geometrycznym występuje iloraz a nie wykładnik, jak czytamy na stronie 29.
4. Termin „kształt filtru” (strona 29 i inne) nie jest poprawny. Pod tym terminem autor zazwyczaj rozumie kształt charakterystyki amplitudowej filtru.
5. Na str. 32 występuje nieobjaśniony skrótowiec „DNN-HNN”.
6. Zamiast pisać „sygnał audio”, „encoder” i „sieć konwolucyjna” (str. 33) chyba lepiej byłoby pisać „sygnał fonii”, „koder” i „sieć splotowa”.

Wymienione usterki mają małą wagę i nie obniżają ogólnie dobrej oceny tego rozdziału rozprawy.

Krótki rozdział trzeci opisuje problem badawczy i dwa autorskie rozwiązania w zakresie odpornej parametryzacji: metoda zmiennej długości ramki analizy sygnału oraz metoda wykorzystująca filtrację odwrotną. Te dwie autorskie propozycje są głównym osiągnięciem pracy. Jest to osiągnięcie oryginalne i dojrzałe oraz mające istotne znaczenie praktyczne. Wyniki badań obu metod są ujęte w rozdziale czwartym oraz w dodatku C.

Badania eksperymentalne zostały przeprowadzone starannie i ze znajomością techniki prowadzenia badań w zakresie rozpoznawania mowy. Pominięcie w badaniach fonemów szczelinowych i płozyjnych było uzasadnione celem badań. Oczywiście badania eksperymentalne są żmudne i czasochłonne, jednak chyba trzeba uznać, że mankamentem przeprowadzonych eksperymentów było ich ograniczenie tylko do głosów męskich.

Autor przedstawił analizy wskazujące, że w dużej liczbie przypadków zaproponowane metody ułatwiają prawidłowe automatyczne rozpoznawanie fonemów. Opisał też sytuacje, w których zaproponowane metody nie poprawiają lub wręcz pogarszają warunki rozpoznawania fonemów. Szkoda, że w rozprawie nie udało się sformułować konkretnych ilościowych kryteriów umożliwiających automatyczne podejmowanie decyzji o zastosowaniu metody zaproponowanej przez autora.

W tabeli 4.2 na podstawie literatury podano prawdopodobieństwa występowania poszczególnych fonemów. Tymi danymi można było zważyć zamieszczone w dodatku C wyniki eksperymentalne dotyczące wyznaczania fonemów, co pozwoliłoby bardziej realistycznie ocenić praktyczną przydatność zaproponowanych rozwiązań.

Dalszym ciekawym zagadnieniem raczej wykraczającym poza zakres rozprawy byłoby uwzględnienie w zaproponowanych rozwiązaniach wykorzystania informacji fazowych.

W uwagach szczegółowych do rozdziału 4 można wymienić nieobjaśniony skrótowiec UBM (str. 68).

Układ rozdziałów rozprawy oraz poszczególnych treści wewnątrz rozdziałów jest prawidłowy, logiczny i czytelny. Praca jest napisana w sposób bardzo jasny oraz łatwy do czytania i zrozumienia.

Wyniki uzyskane przez autora zostały opublikowane w sześciu pracach:

- w dwóch artykułach w czasopiśmie wydawanym przez PAN *International Journal of Electronics and Telecommunications*
- w artykule w czasopiśmie wydawanym przez PAN *Archive of Acoustics*,
- w artykule w czasopiśmie *Vibrations in Physical Systems*,
- w pracy zamieszczonej w materiałach konferencji *2023 Signal Processing Symposium (SPSympo)* ujętych w bazie IEEE Xplore,
- w rozdziale wydanej przez Politechnikę Śląską monografii *Acoustics, electroacoustics and electrical engineering*.

Z wyjątkiem ostatnio wymienionej samodzielnej pozycji wszystkie pozostałe prace zostały opublikowane wspólnie z promotorem i/lub promotorem pomocniczym. Podsumowując te informacje stwierdzam, że dorobek publikacyjny doktoranta należy uznać za znaczny.

W konkluzji recenzji jednoznacznie stwierdzam, że rozprawa pana mgr. inż. Stanisława M. Gmyrka przedstawia oryginalne i kompleksowe rozwiązanie ważnego problemu naukowego. Przedstawiona rozprawa stanowi opis istotnych osiągnięć naukowych mieszczących się w dziedzinie nauk inżynieryjno-technicznych,

a w szczególności w dyscyplinie „Automatyka, elektronika, elektrotechnika i technologie kosmiczne”. Rozprawa niewątpliwie spełnia warunki stawiane rozprawom doktorskim przez odpowiednie przepisy. W związku z tym proszę o dopuszczenie jej autora do publicznej obrony rozprawy doktorskiej w dyscyplinie „Automatyka, elektronika, elektrotechnika i technologie kosmiczne”.

Ponadto biorąc pod uwagę oryginalne i ważne wyniki rozprawy stanowiące istotny wkład do rozwoju metod automatycznego rozpoznawania mowy proponuję wyróżnienie rozprawy. Według mojej najlepszej wiedzy autor spełnił także obowiązujący na Politechnice Wrocławskiej w dyscyplinie „Automatyka, elektronika, elektrotechnika i technologie kosmiczne” warunek opublikowania ściśle związanych z tematyką rozprawy trzech artykułów w czasopismach z listy JCR i odpowiednio punktowanych według wykazu ministerialnego.

