

## **RECENZJA**

### **ROZPRAWY DOKTORSKIEJ mgr inż. Doroty Kamińskiej**

pod tytułem

#### **"ROZPOZNAWANIE EMOCJI NA PODSTAWIE MOWY NATURALNEJ"**

Niniejsza recenzja została sporządzona na zlecenie profesora dr hab. inż. Andrzeja Bartosiewicza, prodziekana Wydziału Elektrotechniki, Elektroniki, Informatyki i Automatyki Politechniki Łódzkiej. Przedstawiona do oceny praca doktorska mgr inż. Doroty Kamińskiej, napisana pod kierunkiem dr hab. inż., profesora Politechniki Łódzkiej Adama Pelikanta.

#### **1. Określenie aktualności tematu i oryginalności problemu badawczego podjętego w rozprawie**

Rozprawa doktorska mgr inż. Doroty Kamińskiej jest poświęcona opracowaniu systemu metod rozpoznawania i klasyfikacji stanów emocjonalnych mowy naturalnej. Opracowanie nowoczesnych metod i technik rozpoznawania mowy ludzkiej obecnie stanowi jeden z najważniejszych kierunków badań naukowych w zakresie projektowania interakcji człowiek-komputer oraz w robotyce, dotyczących komunikacji człowiek - robot. Zagadnienia rozpoznawania stanów emocjonalnych na podstawie analizy sygnału mowy są stosunkowo młode. Ale jest to intensywnie rozwijający się kierunek badań ze względu na rosnące zapotrzebowanie analiz baz danych zawierających nagrania głosowe. Więc, jest to problematyka nie tylko aktualna, ale również bardzo ważna ze względu na praktyczne zastosowania w różnych dziedzinach. Należy tu podkreślić również, że jest to bardzo trudna tematyka. W oparciu o publikacje naukowe można stwierdzić, że badania w zakresie rozpoznawania stanów emocjonalnych na podstawie sygnału mowy są dużym wyzwaniem dla naukowców.

W związku z powyższym, wybraną tematykę rozprawy uważam za w pełni uzasadnioną, bardzo interesującą i aktualną dla współczesnych prac w zakresie rozpoznawania mowy naturalnej i projektowania interaktywnych systemów informatycznych.

#### **2. Charakterystyka celów oraz tez badawczych rozprawy**

W rozprawie doktorskiej mgr inż. Doroty Kamińskiej, niestety, brak precyzyjnie sformułowanego celu pracy. Z wprowadzenia do problematyki wywnioskowałam, że celem

badani naukowych było stworzenie systemu automatycznego rozpoznawania stanów emocjonalnych na podstawie mowy naturalnej.

Przedstawione zostały tezy badawcze, jak doktorantka podkreśliła, „ściśle powiązane z zagadnieniami klasyfikacji mowy emocjonalnej” zawierają:

„**Teza 1:** Wykorzystanie hybrydowych współczynników percepcyjnych w procesie klasyfikacji dokonywanym przy użyciu komitetu klasyfikatorów pozwala na uzyskanie wysokiej skuteczności rozpoznawania emocji na podstawie mowy naturalnej.” Według mnie jest to wiarygodna teza, aczkolwiek, moim zdaniem, skuteczność rozpoznawania emocji w znacznym stopniu będzie zależała od opracowanego komitetu klasyfikatorów i wykorzystywanych w nim metod.

„**Teza 2:** Opierając się na teorii emocji Pluchnika można dokonać klasyfikacji emocji wtórnych.” – Przyznam się, że nawet po wielokrotnym przeczytaniu rozprawy nie bardzo wiem, w jaki sposób ta teza była badana. Poza tym, że w pracy, opierając się na kole R. Pluchnika stworzono modele emocji podstawowych oraz wtórnych. Czy ta teza dotyczy również wyodrębnienia profili emocjonalnych?

„**Teza 3:** Istnieje możliwość zwiększenia dokładności rozpoznawania emocji zawartych w głosie poprzez minimalizację cech osobniczych mówcy”. Ta teza, prawdopodobnie może być wiarygodna i może zostać zweryfikowana w badaniach.

### **3. Ogólna charakterystyka rozprawy. Ocena poprawności struktury rozprawy**

Opiniowana rozprawa doktorska obejmuje łącznie 107 stron. Na logiczny układ rozprawy składają się wstępna i trzy podstawowe części:

- część wstępna, obejmująca bardzo zwięzłe wprowadzenie do problemu oraz sformułowanie tez badawczych, ta część liczy ponad trzy strony;
- pierwsza część teoretyczna zawiera 3 rozdziały (2,3 i 4), w których zaprezentowano zagadnienia teoretyczne z prezentacją aktualnego stanu wiedzy, licząca 44 strony;
- druga część eksperymentalna, stanowiąca 2 rozdziały (5 i 6) opisujące proponowane rozwiązanie zadań (21 stron);
- ostatnia trzecia część zawiera 7 i 8 rozdziały, prezentujące weryfikacje wykorzystywanych algorytmów, badania (jest to 19 stron) i 9 rozdział - podsumowanie i wnioski (4 strony), oraz wykaz literatury, spis rysunków i spis tablic.

Taka struktura rozprawy jest logicznie poprawna, jak również kolejność rozdziałów również jest jak najbardziej uzasadniona.

Moim zdaniem, zostały nieznacznie naruszone proporcje między poszczególnymi częściami, a mianowicie, między zbyt obszerną częścią teoretyczną, która stanowi wprawdzie 48% całej rozprawy a częścią eksperymentalną (22%) i weryfikującą (około 20%).

W części teoretycznej w 2 rozdziale (13 stron) zostały bardzo szczegółowo opisane podstawy neurologiczne i psychologiczne rozpoznawania stanów emocjonalnych, w tym zarys historyczny badań nad emocjonalnością, taksonomia stanów emocjonalnych, ich źródeł i korelaty. Znalazły tam miejsce bardzo ważne do zrozumienia problematyki pojęcia, ale

moim zdaniem, bardzo szczegółowa prezentacja niektórych zagadnień takich, jak korelaty ludzkiego ciała, czy model narządu głosu są zbędne przy realizacji badań w zakresie rozpoznawania emocji na podstawie mowy w ujęciu informatycznym.

W 3 rozdziale, w którym przedstawiono "dotychczasowy stan wiedzy" w zakresie automatycznego rozpoznawania emocji na podstawie sygnału mowy, Autorka wykazała się dobrą orientacją w literaturze z zakresu podjętej tematyki. Bibliografia dość skromna, jak na rozprawę doktorską, zawiera 90 pozycji, w tym 5 publikacji Autorki rozprawy. W tym rozdziale na początku opisano przedmiot badań, czyli bazy nagrań mowy emocjonalnej, wśród których wyróżniono trzy kategorie: "mowa naturalna (emocje spontaniczne), emocje wywoływane oraz odegrane (symulowane)". Autorka dokonała przeglądu istniejących baz nagrań mowy emocjonalnej, nazwanej korpusem mowy emocjonalnej oraz badań naukowych w zakresie klasyfikacji emocji z wykorzystaniem tych baz mowy, realizowanych w różnych ośrodkach naukowych, zarówno w kraju, jak i za granicą. Aczkolwiek podrozdział 3.2 nosi tytuł: "Klasyfikacja mowy emocjonalnej", Autorka opisała w nim wyniki dokonanego przeglądu tych badań (jak wyżej), skupiając się na wykorzystywanych bazach nagrań i deskryptorach, czy współczynnikach wykorzystywanych do klasyfikacji emocji. Zabrakło mi w wielu przypadkach prezentacji wykorzystywanych przy tym metod klasyfikacji, czy algorytmów, m.in. to dotyczy analizy badań w [53] opisanej na str. 29, w pozycjach [54, 55, 56] - na stronie 30, jak również analiz poz. [58,59,60,61] opisanych na stronie 31. Nie jasne są dla mnie, w związku z powyższym, wnioski autorki zaprezentowane na końcu tego rozdziału, dotyczące wyboru metod klasyfikacji w procesie rozpoznawania stanów emocjonalnych.

Czwarty rozdział należy również jeszcze do części teoretycznej rozprawy i zawiera podstawowe zagadnienia teoretyczne dotyczące analizy sygnału mowy, w tym matematyczny opis sygnału mowy w dziedzinie czasu i częstotliwości, liniowe kodowanie predykcyjne oraz analizę cepstralną. Szczególnie ważny jest tu opis współczynników percepcyjnych, wykorzystywanych w systemach rozpoznawania mowy, mówcy, czy emocji. W podrozdziale 4.6.2. przedstawiony został algorytm generacji "innych, wykorzystywanych w niniejszej pracy współczynników percepcyjnych." Jest to niewątpliwie, bardzo istotny algorytm generacji współczynników, nie jasne jest dla mnie tylko, czy to jest autorski algorytm, opracowany w oparciu o istniejące, jak to autorka zauważa, "powszechnie znane" metody, czy jest to algorytm zaczerpnięty z innych badań. Niestety, brak odsyłacza do źródła umieszczonego tam na rysunku 4.8. algorytmu wyznaczania współczynników percepcyjnych.

W drugiej części pracy, którą można nazwać częścią eksperymentalną rozprawy, w 5 rozdziale autorka przedstawia autorską bazę próbek mowy emocjonalnej, opracowanej na potrzeby realizacji badań. Opisane zostało szczegółowo podejście i metody tworzenia korpusu mowy emocjonalnej, w tym metodykę doboru źródeł nagrań, gromadzenia próbek, sposobu etykietowania, tworzenia zbiorów treningowego i testowego. W tym rozdziale Autorka przedstawiła również bazę odegranej mowy emocjonalnej wykorzystaną w badaniach w celach porównawczych, udostępnianej przez Politechnikę Łódzką.

W 6 rozdziale tej że części eksperymentalnej doktorantka zaprezentowała proces tworzenia algorytmów rozpoznawania stanów emocjonalnych. Bazowym algorytmem jest

opisany w tym rozdziale PAKEmo (Podstawowy Algorytmie Klasyfikacji Emocji), który w procesie badań był przez Autorkę udoskonalany i modyfikowany w celu podniesienia jakości klasyfikacji. Ostatecznie opracowany przez Autorkę algorytm jest wielopoziomowym komitetem klasyfikatorów, „budowany poprzez dodawanie kolejnych poziomów podproblemów”, a mianowicie: „rozpoznawanie płci, budowa profili emocjonalnych, zależność emocji od długości wypowiedzi oraz badanie natężenia emocji”. Na poziomie zerowym algorytmu PAKEmo dokonywana jest oddzielna klasyfikacja na podstawie określonej grupy cech przy wykorzystaniu klasyfikatorów minimalnoodległościowych k-NN (k-najbliższych sąsiadów), po czym wykonywane są testy. Szczegółowo opisano działanie algorytmu bazowego.

Rozdział 7 zawiera opis realizowanych eksperymentów badawczych, a mianowicie testów wybranych deskryptorów mowy. Przedstawiono wyniki doświadczeń przeprowadzonych na puli cech powszechnie wykorzystywanej w rozpoznawaniu emocji, jak również wyniki doświadczeń realizowanych z wykorzystaniem rozszerzonej puli cech o współczynniki percepcyjne, które wykorzystywane są w rozpoznawaniu mowy, ale pomijane w rozpoznawaniu emocji. Testy przeprowadzone były na dwóch korpusach mowy naturalnej: na bazie emocji spontanicznych (opracowanych przez Autorkę) oraz bazie emocji odegranych. Przedstawione w tym rozdziale wyniki selekcji cech zostały dokonane z wykorzystaniem metody SFS w celu maksymalizacji skuteczności klasyfikatora. Dokonana selekcja cech pozwoliła Autorce znacząco zredukować wymiarowość zadań, zmniejszając ilość atrybutów nawet kilkakrotnie. Wyniki testów zostały starannie opracowane i zinterpretowane.

W rozdziale 8 zostały zaprezentowane przez Autorkę wyniki testów algorytmów systemu rozpoznawania emocji, przedstawione zostały wyniki eksperymentalnej weryfikacji algorytmu PAKEmo przy użyciu dwóch korpusów mowy emocjonalnej języka polskiego. Porównano wyniki rozpoznawania z użyciem algorytmu k-NN z autorskim podejściem opartym na zbiorze klasyfikatorów - algorytmem PEKAEmo. Opisany został w tym rozdziale przez Autorkę dobór sposobu głosowania w procesie rozpoznawania emocji. Ze względu na fakt, że każdy z podzbiorów cech ma różny wpływ na rozpoznawanie, zastosowano w algorytmie głosowanie ważone, co pozwoliło na wzmocnienie wpływu silniejszych cech na wynik rozpoznawania. W wyniku zastosowania głosowania ważonego, jak podkreśla Autorka rozprawy, polepszona została jakość rozpoznawania. W tym że rozdziale opisano rozszerzenie podstawowego algorytmu o moduł rozpoznawania płci. Klasyfikację płci mówcy przeprowadzono za pomocą również metody "k-NN z dziesięciokrotną walidacją krzyżową". W przypadku mowy naturalnej, czyli bazy emocji spontanicznych zauważono wzrost skuteczności rozpoznawania emocji. Poza tym przeanalizowano wyniki rozpoznawania intensywności emocji podstawowych.

#### **4. Ocena przeprowadzonych badań i merytoryczna ocena rozprawy**

Podstawą opracowanego przez doktorantkę systemu automatycznego rozpoznawania stanów emocjonalnych na podstawie mowy naturalnej jest wykorzystanie i udoskonalenie istniejących metod i podejść w tym zakresie, oraz opracowanie nowego podejścia,

wykorzystującego hybrydowe współczynniki percepcyjne w procesie klasyfikacji, dokonywanym przy użyciu autorskiego algorytmu, opartego na komitetach klasyfikatorów, zapewniającego lepsze wyniki rozpoznawania. Analiza wyników przeprowadzonych badań naukowych potwierdziły początkowe założenia i sformułowane przez Autorkę tezy.

Do najważniejszych osiągnięć Autorki rozprawy należy zaliczyć następujące:

1. W celu realizacji badań naukowych została stworzona autorska baza emocji spontanicznych, na którą składa się ponad siedemset próbek, podzielonych na siedem podzbiorów reprezentujących stany podstawowe. Opierając się na kole R. Pluchnika stworzono modele emocji podstawowych oraz wtórnych, będących wariacjami stanów podstawowych. Eksperymenty badawcze w recenzowanej pracy przeprowadzono na dwóch bazach emocji, porównując emocje odegrane z naturalnymi.
2. Opracowano autorski algorytm określenia intensywności danej emocji na podstawie stopnia jej podobieństwa do mowy neutralnej. Jest to istotny wkład w rozpoznawaniu stanów emocjonalnych, który daje możliwość rozróżnienia intensywności emocji, co jest szczególnie ważne w aplikacyjnych zastosowaniach.
3. Wykorzystanie przez Autorkę do ilościowego opisu problemu nie tylko grup deskryptorów mowy powszechnie używanych w tego typu badaniach, ale również hybrydowych współczynników percepcyjnych, które są używane w rozpoznawaniu mowy, ale pomijane w rozpoznawaniu emocji. Jak wykazały badania, zastosowanie hybrydowych współczynników percepcyjnych dla obu korpusów mowy pozwoliło osiągnąć najwyższe wyniki rozpoznawania emocji.
4. Autorski algorytm klasyfikacji oparty na komitecie klasyfikatorów, pozwalający uzyskać lepsze wyniki rozpoznawania.

Przedłożona do recenzji rozprawa została napisana w sposób przejrzysty i poprawny technicznie i metodologicznie. Autorka wykazała się obszerną wiedzą i umiejętnościami przy rozwiązywaniu trudnych problemów w zakresie rozpoznawania stanów emocjonalnych na podstawie mowy naturalnej.

Wszystkie uzyskane wyniki zostały poddane wnikliwej analizie porównawczej, w której Autorka stwierdziła zarówno poprawność swoich tez, jak też i zasadność wykorzystywanych metod i algorytmów.

Wyniki przedstawione w rozprawie zostały opublikowane w renomowanych czasopismach naukowych, takich jak czasopismo PAN "International Journal of Electronics and Telecommunications", "Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne", czy "Studia Informatica" oraz wygłoszone i zaaprobowane na międzynarodowych i krajowych konferencjach naukowych.

## **5. Ocena formalnej strony rozprawy**

Praca napisana jest na ogół poprawnym językiem, chociaż występują nieliczne błędy, tak zwane literówki (np., str. 47, str.63. w tytule podrozdziału 6.4, kilka błędów na stronie

69), czy błąd w numeracji wzorów matematycznych na str.64 zamiast "na podstawie równania 8.1" powinno być opisano w treści "6.6 i 6.7". Poza tym występują dość liczne błędy interpunkcyjne.

## 6. Uwagi, pytania i komentarze

### 6.1. Uwagi szczegółowe dotyczące treści pracy

1. Brak w dokonanym przeglądzie literatury (rozdział 3) w wielu opisywanych wynikach badań w innych ośrodkach wykorzystywanych w nich metod klasyfikacji, czy algorytmów, m.in. to dotyczy analizy badań w [53] opisanej na str. 29, w pozycjach [54, 55, 56] - na stronie 30, jak również analiz poz. [58,59,60,61] opisanych na stronie 31. Nie są jasne dla mnie, w związku z powyższym, wnioski Autorki zaprezentowane na końcu tego rozdziału, dotyczące wyboru metod klasyfikacji w procesie rozpoznawania stanów emocjonalnych. W ogóle, przegląd literatury w zakresie metod klasyfikacji jest według mnie zdecydowanie zbyt skromny. Brakuje wielu znaczących publikacji, w tym i monografii.
2. Merytorycznie błędne jest użycie terminologii i pojęć: na str.55 w podrozdziale 6.1. autorka stwierdza, że: "Mimo, że metody klasyfikacji to raczej standardowe narzędzia automatycznego rozpoznawania, obejmujące szereg podejść (systemy ekspertowe, drzewa decyzyjne, logika rozmyta, sieci neuronowe i inne)...". Więc, owszem, systemy ekspertowe czasami są wykorzystywane, m.in. do rozwiązania zadań klasyfikacji, natomiast logikę rozmytą zaliczać do tych "podejść" jest dużym błędem. Może to jest skrót myślowy, jeżeli wiemy, że z wykorzystaniem logiki rozmytej i teorii zbiorów rozmytych opracowywane metody grupowania i klasyfikacji, m.in., tak zwana, Metoda rozmytej klasyfikacji k-średniej - (Fuzzy c-means).
3. Brak uzasadnienia wybrania algorytmu k-NN (najbliższych sąsiadów) jako bazowego algorytmu, wykorzystywanego w Podstawowym Algorytmie Klasyfikacji Emocji - PAKEmo. Jak wiadomo algorytm k-NN posiada dość istotne wady, m.in, wadą algorytmu jest jednakowe (często nieuprawnione) traktowanie wszystkich wymiarów oraz duża, w porównaniu z wieloma innymi klasyfikatorami, wrażliwość na zbędne cechy.
4. W rozprawie brakuje odsyłaczy do źródeł przy wszystkich rysunkach, m.in. rys. 2.1, rys. 2.3, rys. 3.1., 4.1., 4.2 i t.d., zwłaszcza istotne jest to w celu wyróżnienia własnych opracowań, autorskich, np. w przypadku prezentowanego na rys. 4.8. str.45 algorytmu, czy rys. 6.2. na str. 59, gdzie przedstawiony jest Schemat podstawowego algorytmu klasyfikacji emocji.
6. W opiniowanej rozprawie nie zostały wyraźnie rozdzielone, czy podkreślone fragmenty, zawierające wyniki znane z literatury lub wprost z nich wynikające od tych, które można by uznać za istotny wkład Autorki w rozwój dyscypliny naukowej, w zakresie której jest realizowana praca.

## 6.2. Pytania do autorki

1. Czy Podstawowy Algorytm Klasyfikacji Emocji (PAKEmo) jest autorski? Z opisu na str. 58 nie zostało to wyraźnie przedstawiono.
2. Jak zakłada autorka, na poziomie zerowym algorytmu znajduje się m klasyfikatorów - gdzie każdy z nich "dokonuje oddzielnej klasyfikacji na podstawie określonej grupy cech" za pomocą algorytmu k-NN - czyli tej samej metody. Wobec czego, czym się różnią te klasyfikatory, wykorzystujące tylko jedną metodę?  
Tu należy nawiązać do poprzedniego stwierdzenia autorki na str.57: "Komitet zbudowany z identycznie działających modeli jest bezużyteczny, a ostateczny wynik nie może być lepszy od pojedynczego klasyfikatora".
3. Czy były podejmowane badania w zakresie oceny błędów klasyfikatorów, w oparciu o które można było dokonać wyboru odpowiednich klasyfikatorów?
4. Czy badane są możliwości zastosowania innych metryk w algorytmie k-NN, niż metryka euklidesowa? Czy były podejmowane badania w tym zakresie?
5. W oparciu o jakie badania jest uzasadnione zastosowanie metody najbliższej średniej w algorytmie rozpoznawania intensywności emocji podstawowych?
6. Czy opisany algorytm wyznaczania współczynników percepcyjnych jest zaczerpnięty z [66], czy jest on modyfikowany przez Autorkę i w jakim zakresie?

Pragnę zaznaczyć, że powyższe uwagi nie zmieniają mojej bardzo pozytywnej oceny pracy naukowej Doktorantki i nie pomniejszają zasadniczo wysokiej oceny całości wykonanych badań.

## 7. PODSUMOWANIE

W podsumowaniu należy podkreślić, że opiniowana rozprawa doktorska została napisana w sposób przejrzysty i poprawny technicznie i metodologicznie. Praca stanowi samodzielne rozwiązanie bardzo ciekawego, ale również niezwykle trudnego zadania naukowego. Autorka wykazała się bardzo dobrą znajomością tematyki podjętej w rozprawie, umiejętnościami i odpowiednim przygotowaniem do samodzielnej pracy naukowej oraz dowiodła prawdziwości postawionych w pracy tez.

Stwierdzam, że przedstawiona do oceny rozprawa doktorska mgr inż. Doroty Kamińskiej pt."Rozpoznawanie emocji na podstawie mowy naturalnej" spełnia wymagania stawiane rozprawom doktorskim w Ustawie o Stopniach Naukowych i Tytule Naukowym oraz stopniach i tytule w zakresie sztuki z dnia 14 marca 2003 roku (Dz. U. nr 65/2003, poz.595).

W związku z powyższym, jako recenzent pracy doktorskiej, zwracam się do Pana Dziekana oraz Rady Wydziału Elektrotechniki, Elektroniki, Informatyki i Automatyki Politechniki Łódzkiej z wnioskiem o dopuszczenie rozprawy do publicznej dyskusji i obrony.

  
Galina Setlak