

22.09.2017, Poznań

**dr hab. inż. prof. UAM Grażyna Demenko**

Instytut Językoznawstwa  
Uniwersytet im. A. Mickiewicza  
60-371 Poznań

Recenzja rozprawy doktorskiej mgr Michała Lenarczyka

*Akustyczne i fonetyczne metody przemiany głosu*

**Promotor: prof. Ryszard Tadeusiewicz**

#### **I. Ocena wyboru tematu, tezy i zakresu pracy**

Przedmiotem opiniowanej rozprawy doktorskiej są metody konwersji sygnału mowy prowadzące do transpozycji cech osobniczych jednego głosu na głos innego mówcy. Zmienność osobnicza w głosach mówców jest zasadniczym problemem w rozwoju aplikacji syntezy mowy, jak i rozpoznawaniu mowy i mówcy. W technologii mowy zauważa się w ostatnich latach coraz częściej rozwiązania oparte głównie na algorytmach umożliwiającym uczenie systemów bez konieczności bezpośredniego uwzględniania złożonych źródeł zmienności wpływających na lingwistyczne, fonetyczne i akustyczne cechy sygnału. Tego typu podejście nie umożliwia interpretacji czynników wpływających na zróżnicowania osobnicze i nie prowadzi do wyjaśnienia procesu tworzenia określonych struktur akustyczno-fonetycznych wypowiedzi.

Pozytywnie więc należy ocenić przedstawioną w rozprawie próbę modelowania struktur mowy na podstawie analizy procesu wytwarzania sygnału.

Warto podkreślić, że mgr Lenarczyk zajmuje się problemem, któremu jak dotąd w dziedzinie przedmiotu poświęcano stosunkowo mało uwagi, trudnym i ważnym, z punktu zaś widzenia rozwoju dalszych aplikacji technologii mowy zasadniczym. Obrany kierunek badań jest słuszny, zgodny z obowiązującymi tendencjami obowiązującymi w tym zakresie na świecie i dający możliwość właściwej kontynuacji pracy, co oczywiście wpłynęło na moją pozytywną ocenę opiniowanej dysertacji.

Mgr Michał Lenarczyk jako cel badań przyjął opracowanie takich metod konwersji głosu, które umożliwią mu odpowiedź na postawione pytanie: *Czy głos można uznać za cechę niepodrabialną, biometryczną - czy można zmieniać tożsamość mówcy?*

Przedmiotem pracy są więc metody przemiany głosu polegające na transformacji cech osobniczych w mowie jednego mówcy na cechy głosu innego mówcy. Mgr Lenarczyk zastosował różne modele obwiedni widmowej: liniowy model predykcyjny, model homomorficzny i mel-cepstralny. Do realizacji tego ambitnego zadania wykorzystał model generowania mowy w układzie źródło-filtr opisywany przy pomocy niewielkiej liczby parametrów i na tej podstawie poprawnie sformalizował algorytm kodowania mowy. W celu odwzorowania przestrzeni cech mówcy na przestrzeń innego mówcy wykorzystał sieć neuronową oraz metodę wektorów nośnych (SVM). Dla transformacji wysokości tonu zastosował metodę parametryczną oraz nieparametryczną.

Jako bardzo ważne osiągnięcie Autora należy uznać nie tylko rozwiązanie teoretyczne analizowanego zagadnienia, ale również praktyczne w postaci wstępnie przetestowanego programu konwersji wybranych par głosów (głos męski, kobiecy i dziecięcy). Wyniki konwersji głosy poddano wstępnym ocenom percepcyjnym. Zaproponowana metoda kodowania mowy przez mgr Lenarczyka potencjalnie może zostać wykorzystana do doskonalenia metodologii rozpoznawania mówcy oraz syntezy mowy.

Cześć teoretyczną oraz implementacyjną pracy można przyjąć bez zastrzeżeń (ewentualnie przydatna byłaby analiza źródeł zmienności sygnału mowy), jednakże część eksperymentalna jest dość fragmentaryczna, zwłaszcza brak jest oceny błędów w odniesieniu do fonetyczno-akustycznej struktury przekonwertowanych wypowiedzi, brak jest również wskazania korelacji między jakością przekonwertowanych sygnałów a odległością fonetyczno-akustyczną między poszczególnymi parami głosów. Doświadczenie percepcyjne przeprowadzono raczej tylko w celu uzyskania wstępnych ocen efektywności zastosowanej metody (co zresztą stanowiło wskazany cel pracy) niż dla systematycznej oceny jakości przekonwertowanych wypowiedzi.

W swojej rozprawie mgr Lenarczyk zwrócił uwagę na te aspekty modelowania zmienności sygnału mowy, które odnosiły się bezpośrednio do postawionej tezy: *możliwe jest uzyskanie efektu przemiany głosu przez zastosowanie wyselekcjonowanej w pracy parametryzacji sygnału mowy i opracowanych metod uczenia maszynowego. Przy określonych założeniach możliwa jest realizacja tego zadania w czasie rzeczywistym.*

Analiza treści rozprawy oraz ocena załączonych próbek dźwiękowych pozwala stwierdzić, że teza została przez mgr Lenarczyka w zupełności udowodniona.

Oceniając ogólny zakres należy uznać, iż jest on w pełni kwalifikujący rczowaną pracę, jako dysertację doktorską.

## II. Ocena osiągnięć i uwagi dyskusyjne

Przed przystąpieniem do omówienia poszczególnych rozdziałów należy podkreślić, że poszczególne etapy badań zostały starannie zrealizowane i połączone w spójną całość. Praca zawiera znaczny materiał informacyjny, który jest wartościową obudową treści oryginalnej stanowiącej własne osiągnięcia doktoranta.

Rozprawa została dobrze napisana pod względem kompozycji całości, zrozumiałym stylem, zawiera jednak pewne uchybienia redakcyjne, które w przypadku publikacji fragmentów rozprawy należałoby usunąć. Została również dość starannie opracowana graficznie, umieszczone wzory, rysunki i wykresy dokumentują właściwie poszczególne etapy badań. W kilku jednakże przypadkach staranność redakcyjna zawodzi i pojawiają się nieoczekiwane potknięcia, jak niejasne opisy, omyłki (do których ustosunkuję się w uwagach szczegółowych).

Rozprawa obejmuje łącznie 6 rozdziałów, spis literatury oraz suplement zawierający 4 załączniki: załącznik A opisujący proces artykulacji, B – wyjaśniający symbole fonetyczne, C – charakteryzujący problematykę rezonansów w torze głosowym oraz cenny załącznik D ilustrujący implementację opracowanych algorytmów.

W celu poszerzenia opisu eksperymentu w uzupełniającym suplemencie dodano jeszcze dwa załączniki E oraz F ilustrujące odpowiednio opis syntezy mowy z użyciem metody *true envelope* estymacji obwiedni widma oraz wyniki dodatkowych testów odsłuchowych, które miały na celu porównanie parametrycznej i nieparametrycznej metody kodowania.

Dodatkowo umieszczone na płycie CD wykresy zmian struktur widmowych oraz przykłady akustycznej konwersji wybranych głosów pozwalają na wstępną ocenę zastosowanej metodologii.

Wprowadzenie pracy (Wstęp) zawiera omówienie tematu, celu pracy, metodyki i układu rozprawy. Dwa następne rozdziały (rozdziały 2 – 3) stanowią podstawy modelowania i przetwarzania mowy, rozdział 4 zawiera opis metodologii transformacji cech osobniczych oraz uczenie, końcowe rozdziały (rozdziały 5 – 6) prezentują ocenę jakości i skuteczności zastosowanych własnych metod oraz podsumowanie rozprawy.

We *Wstępie* (6 stron) mgr Lenarczyk przejrzyście określił założenia oraz tezę pracy.

Sformułował przy tym trzy szczegółowe pytania:

1. *Jak odpowiednio reprezentować sygnał mowy w postaci pozwalającej z jednej strony na możliwie wierną jego rekonstrukcję, a z drugiej nadającej się do transformacji?*
2. *Jaką metodą transformować parametry sygnału i wyznaczać funkcje transformacji?*
3. *Jak wykorzystać dostępny materiał dźwiękowy- korpus nagrań różnych głosów?*

W zakresie metodyki mgr Lenarczyk wyróżnił 3 zasadnicze etapy:

1. *Analizę wstępną sygnału – parametryzację dogodną do transformacji*
2. *Właściwą transformację*
3. *Syntezę, w której następuje rekonstrukcja sygnału o zmienionych cechach osobniczych.*

Obszerne wprowadzenie do zagadnienia konwersji głosu, pozwala stwierdzić, iż Autor jest w pełni świadomy skali trudności oraz złożoności podjętych zadań badawczych.

W rozdziale drugim **Sygnał mowy i głos** (25 stron) mgr Lenarczyk przedstawia podstawy analizy akustycznej oraz problem modelowania sygnału mowy. Podane w podrozdziale 2.1 wiadomości wstępne są jednak zbyt syntetyczne (jedna strona). Szkoda, że Autor nie zamieścił w tej części pracy również krótkiego omówienia segmentalnych oraz suprasegmentalnych cech mowy (np. dla języka polskiego), które byłoby przydatnym wprowadzeniem czytelnika w zagadnienie praktycznego przetwarzania sygnału, w szczególności w zakresie modelowania osobniczych zróżnicowań głosu. Pozwoliłoby to zapewne na uniknięcie nieporozumień, np. odnośnie interpretacji stwierdzenia na str.5 (pierwszy akapit): *Przemiana głosu, która stanowi temat rozprawy, jest techniką manipulacji brzmieniem sygnału mowy w celu sztucznej zmiany oryginalnego głosu mówcy na inny, wybrany głos, przy zachowaniu nienaruszonej struktury fonetycznej i prozodycznej wypowiedzi.* Stwierdzenie to może prowadzić do przypuszczenia, że cechy fonetyczne głosek przy konwersji głosu nie ulegają modyfikacji, co stanowi zbyt duże uproszczenie (przykładowo allofony głosek płynnych oraz trących niosą dużo zróżnicowań osobniczych). Dla potrzeby parametryzacji mowy wyłącznie na poziomie akustycznym przetwarzanie informacji jest niespecyficzne. Tak można opisywać w zasadzie dowolny sygnał akustyczny zmienny w czasie i tak właśnie potraktowany jest sygnał mowy w niniejszej pracy. Autor skoncentrował się wyłącznie na zagadnieniach opisu akustycznego, mimo to pewne elementy analizy fonetycznej są konieczne, ponieważ po odpowiednim przetworzeniu i parametryzacji sygnału akustycznego następuje bezpośrednia operacja przyporządkowania struktur językowych oraz cech osobniczych.

W podrozdziale 2.2 zatytułowanym: *Akustyczne cechy głosu* pojawia się następująca uwaga (str.25): *„statystyczne cechy 2 pierwszych formantów wystarczą do identyfikacji mówcy”.* Stwierdzenia tego nie można zaakceptować, wydaje się że Doktorant nie dość

wystarczająco dokładnie przestudiował pracę, na którą się zresztą sam powołuje (poz. 51: Jassem, W., M. Krzyśko and P. Stolarski. 1984. *Częstotliwości formantowe samogłosek polskich, jako cechy fonematyczne i osobnicze w świetle statystycznej analizy dyskryminacyjnej*. Prace IPPT PAN27/1984). W pracy tej autorzy na str. 17 stwierdzają: „Częstotliwości F1 i F2 silniej różnicują fonemy niż głosy”. Oczywiście przy niewielkiej liczbie mówców istnieje potencjalna szansa różnicowania głosów na podstawie formantów F1 i F2. Wyniki tych badań są autorce niniejszej recenzji dobrze znane, ponieważ doprowadziły one do ostatecznego ustalenia przez prof. Jassema i współpracowników cech widmowych głosek polskich. Pierwszy i drugi formant służą głównie **rozdzielaniu** głosek. Oczywiście istnieją w tym zakresie również różnicowania osobnicze, jednakże nie są one zasadniczym powodem zmienności. Absolutnie konieczna jest staranniejsza analiza w tym przedmiocie wyników badań nowszych (nie sprzed 40 lat!). Nie tylko dla języka polskiego można stwierdzić, iż dla ekstrakcji cech indywidualnych głosu ważniejsze są wyższe pasma częstotliwości (np. 3 i 4 formant). Jednakże wyznaczanie zarówno częstotliwości jak i szerokości pasm formantowych obarczone jest bardzo dużym błędem i w związku z tym obecnie nie wykorzystuje się tych cech w praktyce rozpoznawania mówców.

W kolejnych podrozdziałach: podrozdziale 2.3 Autor rozprawy przedstawia zagadnienie modelowania mowy, w podrozdziale 2.3.1 omawia syntezę formantową, w 2.3.2 liniowy model predykcyjny, w 2.3.3. metody widmowe oraz metody falkowe, zaś w podrozdziale 2.3.5 przedstawia metody synchroniczne z tonem podstawowym. Warto również zaktualizować podrozdział 2.3, np. synteza konkatenacyjna nie jest techniką najnowszą (obecnie szeroko stosuje się w syntezie modele Markowa HMM). W podrozdziale 2.4 *transformacja cech w przemieszanie głosu* należałoby wskazać jakie PSOLA ma ograniczenia w odniesieniu do zakresu modyfikacji zmian tonu podstawowego, ponieważ powodują one określone ograniczenia praktyczne. W podrozdziale 2.5 poświęconym metodom oceny konwersji głosu przydatne byłoby szersze omówienie testów percepcyjnych (częściowo już uzupełnione w załączniku F) oraz włączenie oceny fonetyczno-akustycznej przekonwertowanych struktur mowy. Poszerzenie rozdziału 2 powinno również obejmować podstawową analizę źródeł zmienności mowy przyczyniających się do określonych różnicowań sygnału akustycznego.

Można założyć, iż w charakterystyce i rozpoznawaniu mówcy przydatny będzie kompleksowy zbiór różnego typu parametrów: (1) uwarunkowanych fizjologicznie (poprzez anatomię i fizjologię narządu głosu) długo- i krótkoterminowych akustycznych, (2) uwarunkowanych fonetycznie (na podstawie czynników językowych i pozajęzykowych), (3) cech gramatycznych (określonych regułami gramatyki dla określonego języka i mówcy).

Oczekuję, że Doktorant przygotowując się do obrony przestudiuje jeden z wielu podręczników fonetyki, aktualnych publikacji z dziedziny przedmiotu (przykładowe opracowania podałam w końcowym fragmencie recenzji) i będzie mógł ustosunkować się (zgodnie z aktualnym stanem wiedzy) do następujących zagadnień:

- Fonetyczno-akustyczna struktura głosek polskich.
- Jakie są źródła zmienności sygnału mowy, jakie są źródła zmienności międzyosobniczych.
- Czy istnieje interakcja między częstotliwością podstawową i cechami widmowymi w sygnale mowy?
- Metody normalizacji zmian wysokości tonu.

Zaniepokoiła mnie dość skromna literatura poświęcona rozwojowi problematyki rozpoznawania mówcy w Polsce, nie znalazłem w wykazie literatury podanym w rozprawie mgr Lenarczyka ani jednego odwołania do publikacji uznanych nie tylko w Polsce specjalistów w dziedzinie przedmiotu np. Czesława Basztury, Wojciecha Majewskiego.

Podsumowując całościowo ten fragment pracy należy stwierdzić, że obszerne wprowadzenie do zagadnień konwersji głosu oraz przegląd literaturowy świadczy o bardzo dobrym zorientowaniu Autora rozprawy w dziedzinie przedmiotu. Praktycznie bezbłędnie opracowane są fragmenty rozprawy związane z cyfrowym przetwarzaniem sygnału mowy, co zapewne stanowiło wynik starannego przestudiowania prac promotora prof. Ryszarda Tadeusiewicza, nieco gorzej opracowane są fragmenty bezpośrednio związane z fonetyką akustyczną oraz różnicowaniami osobniczymi głosów.

Moje uwagi krytyczne nie odnoszą się bezpośrednio do zasadniczych treści rozprawy, mają charakter raczej koniecznego komentarza i nie obniżają ogólnej wartości dokonań doktoranta. W łącznej ocenie rozdziałów wstępnych stwierdzam, że są one przykładem samodzielnej, starannie przeprowadzonej syntezy różnorodnej problematyki począwszy od procesu wytwarzania mowy, parametryzacji akustycznej aż do modelowania i kodowania sygnału.

*Metody kodowania mowy* omówione w rozdziale trzecim (38 stron) charakteryzują się zróżnicowanym stopniem oryginalności. Przedstawione sposoby przekodowania głosu oraz opracowany własny mechanizm kodowania oparty na modelu artykulacji w układzie źródło-filtr w sposób przejrzysty ilustrują przyjętą w rozprawie koncepcję. Na podstawie analizy różnych modelei obwiedni widmowej (liniowy model predykcyjny, homomorficzny i mel-cepstralny) mgr Lenarczyk zaproponował dwa oryginalne podejścia do przemiany wysokości głosu: parametryczną metodę syntezy i nieparametryczną, polegającą na wykorzystaniu

wokodera fazowego. Zamieszczone przykłady są jednak bardzo nieczytelne, np. na rys.3.1 ilustrującym wstępną obróbkę sygnału zasadnicze szczegóły nie są dobrze widoczne. Trudno czytelna jest również ilustracja na rys. 3.11, która ma obrazować brak odwzorowania dolnego i górnego krańca częstotliwości.

W podrozdziale 3.3.1. mgr Lenarczyk przedstawia wyznaczenie okresu metodą korelacji własnej zgodnie z przyjętym zakresem zmian wysokości tonu od 100Hz do 640Hz. Głosy męskie często jednak mają znacznie niższą częstotliwość (70 Hz, a nawet 60 Hz) a więc dla obliczania funkcji korelacji własnej długość okna musiałaby być rzędu około 45 ms (3\*15 ms). Nasuwają się więc wątpliwości: Czy ta sama długość okna stosowana była do wszystkich głosów. Czy Autor rozprawy przeprowadził analizę błędów wyznaczenia wysokości tonu na podstawie swojej metody.

W podrozdziale 3.4 nie jest jasne stwierdzenie (str.75): *modyfikowane są tylko cechy osobnicze, a do tych należy barwa (obwiednia częstotliwościowa) oraz wysokość (częstotliwość podstawowa). Obwiednia czasowa sygnału oraz dźwięczność są cechami związanymi przede wszystkim z prozodią i fonetyką. Chociaż mogą one w pewnym stopniu zależeć od głosu mówcy i jego sposobu artykulacji, ich modyfikacja wiązałaby się z nieuniknioną degradacją zrozumiałości i naturalności głosu.*

Trudno potwierdzić, że cechy osobnicze nie są związane np. z prozodią (ściślej z fonetyką segmentalną i z fonetyką suprasegmentalną).

W rozdziale czwartym *Uczenie i transformacja cech osobniczych* (29 stron) mgr Lenarczyk skupił się na zagadnieniu transformacji cech i wyznaczeniu parametrów transformacji. Dla konwersji głosów wykorzystał wybrane wypowiedzi (istotnie różniącymi się segmentalnie) z bazy Corpora. Dla transformacji cech, uczenia zastosował sieć neuronową i wektory nośne SVM. Warto może byłoby rozważyć konwersje głosów w zakresie określonych zakresów wysokości tonu dla następujących par: 1) głos męski-głos męski, 2) głos żeński-głos żeński, głos dziecięcy-głos dziecięcy. Można przypuszczać, że wyniki konwersji byłyby znacznie lepsze niż dla przeanalizowanych w rozprawie par głosów.

Warto byłoby też przemyśleć ponownie metody normalizacji, nie jest np. jasne dlaczego nie została wykorzystana DTW, która w tym przypadku mogłaby się okazać efektywna.

Rys. 4.3 na str. 93 przedstawia histogramy cech LSP dla analizowanych mówców, na str. 95 pojawia się interpretacja fonetyczna tych histogramów, jednak trochę zbyt ogólna; warto byłoby się przyjrzeć dokładniej odnoszonym skupieniom punktów przyporządkowanych do określonych głosek.

W rozdziale piątym *Ocena jakości i skuteczności* mgr Lenarczyk przedstawia metody

odsluchowe przetransformowanych sygnałów (ABX i MOS) celem oceny skuteczności przemiany i jakości wyników konwersji. Zagadnienie oceny jakości syntezy znacznie poszerzono w załączniku F, dodano omówienie metodologii (załącznik F1) oraz wyników doświadczeń (załącznik F2). Uzupełnienia te w pewnym stopniu wyjaśniły przeprowadzane w tym zakresie eksperymenty. Jednakże dokonane oceny jakości, z uwagi na bardzo skromny materiał eksperymentalny (bardzo ograniczony materiał lingwistyczny, kilka par głosów) należy traktować jako wstępne i stanowiące podstawę do dalszych badań.

**Podsumowanie** dysertacji w rozdziale 6 (4 strony) zawiera syntetyczne ujęcie wyników i propozycje dalszych badań. Do ich kontynuacji chciałabym pana mgr Lenarczyka zachęcić. Przyjęte metody konwersji głosów wydają się obiecujące do zastosowań np. w syntezie mowy. Mgr Lenarczyk przeprowadza krótką podsumowującą dyskusję swojej rozprawy. Zauważa, że wyniki mogłyby być lepsze jeśli zmiany wysokości tonu byłyby w obrębie podobnej skali głosu. Zastosowane podejście parametryczne Doktorant oparł na oryginalnej własnej metodzie estymacji współczynnika dźwięczności chwilowej jako wielkości ciągłej. Źródłem znaczącej straty jakości jest transformacja obwiedni. Zastosowane modele obwiedni częstotliwościowej według Autora mają pewną wariancję związaną z następującymi czynnikami: a) procesem okienkowania, b) sprzężeniem estymaty obwiedni i wzbudzenia oraz c) zastosowaniem dla dopasowania próbek korpusu nieoptymalnej metody. Analizę błędów umożliwiłby zapewne obszerniejszy systematycznie dobierany materiał eksperymentalny.

Zastrzeżenia, które ujęłam w niniejszej recenzji związane są głównie z interpretacją fonetyczną wyników, nie stanowią przedmiotu pracy i dlatego nie mają znaczącego wpływu na moją ocenę dokonań autora odnośnie konwersji głosów.

### III . Uwagi szczegółowe

1. Badania eksperymentalne przeprowadzono tylko dla 3 par głosów. Czy zaistniały przeszkody aby poszerzyć materiał badawczy uwzględniający konwersje par głosów o podobnym zakresie zmian wysokości tonu.
2. W streszczeniu pojawiło się bardzo nieprecyzyjne stwierdzenie: *W pracy skupiono się na cechach segmentalnych (krótkoczasowych głosu, które nadają się do transformacji w czasie rzeczywistym. Należą do nich wysokość tonu i barwa głosu.* Zmiany wysokości tonu bezpośrednio określają strukturę suprasegmentalną sygnału.
3. Na str. 11 pojawiło się stwierdzenie: *Informacja niesiona przez sygnał mowy obejmuje cechy osobnicze oraz językową i pozajęzykową treść komunikatu.*



Czy cechy osobnicze można uznać jako wyłącznie pozajęzykowe?

4. Czy dla sygnału mowy polskiej częstotliwość próbkowania 16KHz jest wystarczająca?
5. Rysunki przedstawiające sygnał mowy należałoby opracować w sposób bardziej przejrzysty. Nie jest to zadanie trudne. Wystarczy wykorzystać jakikolwiek program (np. Praat, Wavesurfer; obecnie dostępnych jest co najmniej kilka takich darmowych i bardzo dobrych programów). Zaprezentowane w pracy rysunki przedstawiające sygnał mowy, w zasadzie bardzo niewiele ilustrujące, zdecydowanie należałoby je powiększyć, umieścić np. jeden pod drugim.

### 5. Literatura uzupełniająca

- Basztura, C. (1991). Experiments of automatic speaker recognition in open sets. *Speech Communication*, 10(2), 117-127.
- Majewski, W., & Basztura, C. (2014). Parametrical patterns of voices in forensic applications. *Archives of Acoustics*, 23(4), 463-480.
- Majewski, W., & Basztura, C. (1996). Integrated approach to speaker recognition in forensic applications. *Forensic linguistics*, 3, 50-64.
- Drygajło, A. (2007). Forensic automatic speaker recognition [Exploratory DSP]. *IEEE Signal Processing Magazine*, 24(2), 132-135
- Jassem, W. (1973). *Podstawy fonetyki akustycznej*. Państwowe Wydawnictwo Naukowe.
- Steffen-Batogowa, M. (1975). *Automatyzacja transkrypcji fonematycznej tekstów polskich*. Państwowe Wydawnictwo Naukowe.
- Stefan Grocholewski (2001) „Statystyczne podstawy systemu ARM dla języka polskiego”, Wydawnictwo Politechniki Poznańskiej w serii Rozprawy (nr 362)
- Demenko, G. (2015). *Korpusowe badania języka mówionego*. Akademicka Oficyna Wydawnicza EXIT.
- Dukiewicz, L. (1995). Fonetyka. *Henryk Wróbel (ed.)*, 9-103.
- Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). MIT press.

### III. Podsumowanie

W posumowaniu swojej oceny dysertacji doktorskiej pana mgr Michała Lenarczyka *Akustyczne i fonetyczne metody przemiany głosu* stwierdzam co następuje. Pomimo zastrzeżeń dotyczących głównie aspektów fonetycznych i lingwistycznych, a więc nie stanowiących istoty pracy, uważam opiniowaną rozprawę jako wartościową i stanowiącą ważne osiągnięcie w zakresie metodologii konwersji głosu. Autor samodzielnie rozwiązał istotny problem naukowy związany z przekształcaniem struktury widmowej sygnału mowy. Jako ważny należy uznać również praktyczny wynik w postaci wstępnie przetestowanego programu dokumentującego poprawność przyjętych koncepcji.

Ze względu na systematyczność opracowanej przez autora metody naukowej ważnej dla dalszego ukierunkowania badań oraz przyszłej użyteczności praktycznej, uważam opiniowaną pracę jako spełniającą wymagania stawiane dysertacjom doktorskim i stawiam wniosek o dopuszczenie jej do publicznej obrony.

