25.11.2009, Poznań

dr hab. inż. prof.UAM Grażyna Demenko

Zaklad Fonetyki
Instytut Językoznawstwa
Uniwersytet im. A. Mickiewicza
Międzychodzka 5, 60-371 Poznań

Laboratorium Technologii Języka i Mowy
Poznańskie Centrum Komputerowo-Sieciowe
Zwierzyniecka 20, 61-704 Poznań

Recenzja rozprawy doktorskiej mgr inż. Krzysztofa Szklannego

Optymalizacja funkcji kosztu w korpusowej syntezie mowy polskiej

1. Ocena wyboru tematu, tezy i zakresu pracy

Jedną z ważniejszych bardzo dynamicznie rozwijających się dziedzin łączności słownej w układzie człowiek komputer jest synteza mowy. Nie ma wątpliwości, iż język będzie stanowił podstawowy element multimedialnej komunikacji i w niedługiej przyszłości wraz ze wzrostem wszechstronności systemów technicznych nastąpi coraz ściślejsza integracja przetwarzania języka mówionego z innymi gałęziami technologii informacyjnej. Jednakże, pomimo kilkudziesięciu lat badań w wielu ośrodkach naukowych na całym świecie, relacje między sygnałem akustycznym i strukturą języka nie zostały w pełni ustalone, a implementacje syntezy mowy są ciągle ograniczone. Złożoność problematyki zarówno na etapie wytwarzania, percepcji, jak i analizy akustycznej mowy wynika – niezależnie od tego, czy układem rozpoznającym jest mózg człowieka, czy komputer – z jej specyficznych własności. Powstałe w procesie artykulacji zespoły dźwięków są nośnikami różnorodnych informacji językowych, paralingwistycznych oraz pozajęzykowych. Określenie źródeł zmienności sygnału mowy i opisanie ich funkcjonowania jest zadaniem tak skomplikowanym, iż istnieje pogląd sceptyczny, według którego sformułowanie odpowiednich algorytmów przetwarzania mowy wyłącznie na bazie teorii fonetyczno-lingwistycznych jest możliwe.

W związku z tym, zauważa się w ostatnich latach rozwiązania oparte głównie na metodach korpusowych, które umożliwiają uczenie systemów na bazie statystyczno-matematycznych algorytmów bez konieczności bezpośredniego uwzględniania złożonych związków między językowymi i akustycznymi cechami sygnału.

Należy więc stwierdzić, że ukierunkowanie badań na syntezę korpusową, które przyjął Doktorant jest słuszne, zgodne z obowiązującymi tendencjami na świecie i dające możliwość dalszego rozwoju prac.

Temat rozprawy *Optymalizacja funkcji kosztu w korpusowej syntezie mowy polskiej* jest niezwykle ważny, zarówno z uwagi na teoretyczne aspekty przetwarzania sygnału mowy, jak i praktyczne implementacje. Mgr inż. Krzysztof Szklanny jako cel badań przyjął zoptymalizowanie funkcji kosztu w korpusowej syntezie mowy dla języka polskiego. Proces syntezy obejmował etap przygotowania korpusu, realizacje nagrań, segmentacje bazy językowej, zawierał realizacje nowych modułów, jak i dostosowanie już istniejących do wymogów syntezy mowy polskiej w systemie Festival.

Dla optymalizacji funkcji kosztu mgr Szklanny zaproponował nowatorską metodę wykorzystującą algorytm ewolucyjny. Efekt badań potwierdził percepcyjnym testem jakości syntetycznej mowy typu MOS.

W pracy zostały postawione trzy tezy:

– funkcje kosztu można optymalizować za pomocą metod heurystycznych. Jedną z metod optymalizacji jest metoda oparta na algorytmie ewolucyjnym,

– optymalizacja funkcji kosztu ma istotny wpływ na poprawienie jakości syntezy korpusowej,

– wybór odpowiedniego mówcy oraz jakość bazy akustycznej ma bardzo duży wpływ na finalną jakość generowanej mowy.

W pierwszej kolejności silną stroną opiniowanej pracy jest - według mojej oceny - wybór jej tematu, a także dobór metodologii, jaką Autor zdecydował się zastosować.

Przechodząc od mojej zdecydowanej aprobaty dla tematu opiniowanej pracy, do stosowanej metodologii, pragnę stwierdzić, że Doktorant łączy w swojej pracy bardzo dobrze stosowane podejście teoretyczne z niezwykle rzetelnymi badaniami eksperymentalnymi, w których z kolei znakomicie łączy metody subiektywne (odsłuchowe) i metody obiektywne (wykorzystujące aparaturę akustyczną dla śledzenia zmian w czasowym przebiegu, w widmie i w parametrycznym opisie sygnału).

W pracy zwrócono szczególną uwagę na aspekty przetwarzania sygnału mowy, praktyczne zastosowanie modeli HMM w systemach segmentacji mowy. Należy jednakże z uznaniem podkreślić, iż autor z powodzeniem przeprowadza również analizy fonetyczne i w pełni uświadamia sobie trudności związane z analizą specyficznych informacji lingwistycznych.

Tezy pracy związane z funkcją kosztu, jej optymalizacją oraz strukturą bazy zostały przez Doktoranta w zupełności udowodnione. Metoda optymalizacji funkcji kosztu zaproponowana przez mgr inż. Szklannego potencjalnie otwiera nowe możliwości, które mogłyby być wykorzystywane w systemach syntezy mowy.

Oceniając ogólny zakres rozprawy należy stwierdzić, iż jest on w pełni kwalifikujący niniejszą pracę jako dysertację doktorską.

2. Ocena osiągnięć i uwagi dyskusyjne

Przed przystąpieniem do omówienia poszczególnych rozdziałów należy podkreślić, że poszczególne etapy badań zostały na ogół 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 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 7 rozdziałów, spis literatury oraz załącznik w postaci płyty CD, na której umieszczono tekst rozprawy, przykłady akustyczne, a co najważniejsze wyniki testów.

Dwa pierwsze rozdziały (rozdziały 1 – 2) stanowią wprowadzenie do pracy, następne (rozdziały 3 - 4) zawierają opis zastosowanej metodologii, końcowe rozdziały (rozdziały 5 – 7) prezentują badania własne wyniki oraz wnioski i podsumowanie rozprawy.

Rozdział pierwszy stanowiący syntetyczne wprowadzenie do zagadnień analizy i syntezy świadczy o bardzo dobrym zorientowaniu Autora rozprawy w dziedzinie przedmiotu. Szkoda jednak, że Autor nie zamieścił w tej części pracy również krótkiego dodania wyjaśnień poświęconych terminologii fonetycznej, a szczególnie pojęciom lingwistycznym stosowanym w całej pracy np. głosce, fonemowi, sylabie oraz wyrazowi. Problemy terminologiczne występują w większości prac z zakresu technologii mowy, trudno bowiem tę interdyscyplinarną dziedzinę opisać wspólnym dla wszystkich dyscyplin językiem. W rezultacie nie zawsze udaje się uniknąć niejednoznaczności, jako że na przeszkodzie stoi również wiele kontrowersji reprezentowanych przez uznane autorytety w dziedzinie przedmiotu. Przy okazji omawiania zagadnień fonetycznych, zwłaszcza struktury prozodycznej wypowiedzi, transkrypcji fonetycznej warto byłoby w celu ilustracji złożoności problematyki fonetyki języka polskiego, przytoczyć różne poglądy prezentowane przez polskich językoznawców (np. Marię Steffen Batogową, Leokadię Dukiewicz, Wiktora Jassema).

Ponieważ zwykle przedstawiciele nauk technicznych wykazują awersję (częściowo zrozumiałą np. powodu wymienionego powyżej) do studiowania teoretycznych zagadnień lingwistycznych, wskazane byłoby, aby Doktorant zapoznał się z pracami ujmującymi w sposób syntetyczny i sformalizowany problemy lingwistyczne w technologii mowy, przedstawianymi przez inżynierów i dla inżynierów. Sugeruję więc, aby Doktorant przygotowując się do obrony zechciał przestudiować np. podrozdział 2.5 (zwłaszcza pojęcie fonemu), rozdział 4 (metody opisu sygnału mowy) oraz rozdział 5 (sygnał mowy w automatyce) w podręczniku *Sygnał mowy* napisanym przez prof. Ryszarda Tadeusiewicza. Przydatną lekturą byłby również podręcznik: Hess W. (1983) *Pitch determination of speech signals - algorithms and devices*, Springer Verlag, Berlin oraz do analiz prozodycznych opracowanie Marii Steffen-Batogowej (1996) *Struktura przebiegu melodii polskiego języka ogólnego*, wyd. UAM. Poznań lub Leokadii Dukiewicz (1978) *Intonacja wypowiedzi polskich*, Prace Instytutu Języka Polskiego, wyd. PAN, Wrocław.

Powyższe uwagi nie odnoszą się do zasadniczych treści rozprawy, mają charakter raczej koniecznego komentarza i nie obniżają ogólnej wartości dokonań Doktoranta.

W rozdziale drugim zaprezentowano historie syntetyzatorów mowy. Opisano podstawowe rodzaje syntezy, a także dokonano analizy działania systemu TTS oraz jego poszczególnych modułów. Dla porządku jednak chcę tu również odnotować, jak się wydaje na podstawie mojego oglądu tej tematyki, iż pierwsze udane opracowania na temat syntezy mowy polskiej powstawały w grupie Prof. Leonarda Bolca na Uniwersytecie Warszawskim,

inne bardzo ciekawe prace (ale raczej badawcze, niż użytkowe) tworzono także na Politechnice Wrocławskiej, zaś bardzo ciekawe użytkowe systemy dla niewidomych budowano również we wczesnych latach 80 tych w Instytucie Biocybernetyki i Inżynierii Biomedycznej PAN. Jako jeden z pierwszych syntezatorow regułowych należy uznać Synfor powstały w latach 70 tych w IPPT (Kacprowski J., Mikiel W. (1968) *Realizacja procesu syntezy mowy za pomocą syntezatora Synfor II*, Prace IPPT, 25/1968, Warszawa). Syntezator formantowy „Kubuś" opracowany w IPPT w Poznaniu stanowił przez wiele lat pierwszą wersję komercyjnego syntezatora mowy polskiej.

Uwaga ta również nie ma oczywiście znaczenia z punktu widzenia oceny rozprawy, ale myślę, że dbając o sprawiedliwą ocenę naukowego wkładu w rozwój syntezy, powinno się wspomnieć o osiągnięciach polskich zespołów w ostatnich kilkudziesięciu latach.

W łącznej ocenie rozdziałów wstępnych stwierdzam, że są one przykładem samodzielnej, wszechstronnie przeprowadzonej syntezy różnorodnej problematyki począwszy od analizy parametryzacji akustycznej sygnału mowy aż do optymalizacji metody syntezy w zakresie doboru funkcji kosztu.

Rozdział trzeci stanowi wprowadzenie do jednej z najważniejszej funkcji w korpusowym syntezatorze mowy - funkcji kosztu. Rozdział ten zdecydowanie powinien zostać poszerzony. W podrozdziale 3.3 Autor wymienia kilkanaście parametrów definiujących koszt konkatenacji systemie Festival, nie podaje jednakże żadnej ich interpretacji. W związku z tym, czytelnik nie znający instrukcji Festivala musi się domyślać co może oznaczać np. koszt Fo, koszt niewłaściwego doboru melodii, akcentu, POS, itp.
Przed ewentualną publikacją rozprawy należałoby temu fragmentowi poświęcić zdecydowanie więcej uwagi, a przede wszystkim podać interpretacje optymalizowanych parametrów funkcji kosztu.
Znając strukturę języka angielskiego i polskiego można byłoby tu włączyć krótką dyskusję odnośnie potrzeby optymalizacji funkcji kosztu pod kątem specyficznych cech językowych.

W rozdziale czwartym przedstawiono szereg zadań związanych z tworzeniem korpusu, rejestracją nagrań oraz ich segmentacją. Przedstawiono również automatyczną metodę korekty posegmentowanych nagrań. W systemach korpusowych istnieje kilka sposobów optymalizacji funkcji kosztu. Pierwszy z nich polega na intuicyjnym dobieraniu parametrów oraz przeprowadzaniu kontrolnych testów percepcyjnych, które maja umożliwić wyznaczenie najlepszych pod względem percepcyjnym współczynników wag. Drugim sposobem jest metoda automatyczna polegająca na trenowaniu poszczególnych wag kosztu doboru jednostki.

Fragment odnoszący się do balansowania korpusu stanowi bardzo wartościową część rozprawy. Przydatne byłyby tu jednak szersze wyjaśnienia odnośnie założeń balansowania korpusu w zakresie struktury segmentalnej (podane na str.82).

Autor przyjął następujące założenia:

– minimalna długosc fonetyczna zdania to 30 znaków

– maksymalna długosc fonetyczna zdania to 80 znaków

–korpus powinien zawierać około 2500 zdań

–w korpusie każdy fonem powinien wystąpić, co najmniej 40 razy

– każdy difon powinien wystąpić, co najmniej 4 razy

– każdy trifon powinien wystąpić, co najmniej 3 razy, to wymaganie jest dostepne tylko dla najczęściej występujących trifonów

Ciekawa uwaga pojawia się również w odniesieniu do balansowania prozodycznego korpusu (str.82): *W praktyce okazało sie, iż 76 zdań pytających oraz 13 zdań wykrzyknikowych to zbyt mało, aby stworzyc odpowiedni model intonacyjny wypowiedzi jezyka polskiego. Z tych powodów w finalnej wersji korpusu zrezygnowano z tych promptów, czyli korpus został zmniejszony o 89 zdań.*

Warto byłoby ten fragment poszerzyć w dyskusji o analizę specyfikacji dla zbalansowania struktury prozodycznej korpusu.

Bardzo ważny fragment rozprawy stanowi podrozdział 4.3 poświęcony segmentacji sygnału mowy. Porównanie poprawności segmentacji różnych zestawów modeli Markova przeprowadzono w programie Praat. Przeanalizowano segmentacje wygenerowaną przez różne modele fonemów i wybrano najlepsze z nich.

Problem segmentacji automatycznej sygnału mowy jest ważny, często zasadniczy dla różnych aplikacji, warto byłoby więc poszerzyć ten fragment rozprawy o znacznie bardziej systematyczną analizę błędów (tab. 4.7). Rys. 4.11-4.14 wymagają szerszego omówienia i bardziej czytelnej prezentacji niepoprawnych segmentacji. Być może warto rozważyć odrębną publikacje tego fragmentu rozprawy.

W rozdziale piątym opisano strukturę i sposób działania algorytmu ewolucyjnego. Przedstawiono strategie optymalizacji funkcji kosztu oraz sposób przeprowadzenia badań optymalizacyjnych. Rozdział ten stanowi niewątpliwie nowatorski własny wkład pracy Autora i charakteryzuje się wysokim stopniem oryginalności.

W rozdziale szóstym dokonano analizy wyników badań. Wyniki testu wskazały, iż strategie ewolucyjne są skuteczne w procesie optymalizacyjnym i wygenerowane parametry dla funkcji kosztu potwierdziły to w badaniach testowych.

Rozdział siódmy zawiera opis testu percepcyjnego MOS, którego wyniki potwierdziły efektywność wykonanych badan optymalizacyjnych, dzięki którym uzyskano lepszą jakość syntetycznej mowy polskiej. Przedmiotem badań testowych było porównanie 3 różnych funkcji kosztu oraz ocena jakości sygnału syntetycznego.

Podsumowanie dysertacji zawiera syntetyczne ujęcie wyników i propozycje dalszych badań. Do ich kontynuacji chciałabym pana mgr inż. Krzysztofa Szklannego bardzo zachęcić. Propozycja optymalizacji modeli wydaje się być obiecująca do zastosowań w syntezie mowy. W celu oceny ich przydatności praktycznej należałoby przeprowadzić bardziej szczegółową analizę fonetyczno-akustyczną systematycznie uwzględniającą uwarunkowania prozodyczne. Rozwiązanie tego zagadnienia zdecydowanie przekracza ramy dysertacji doktorskiej. Po przestudiowaniu niniejszej rozprawy, jestem jednakże przekonana, iż pan mgr Szklanny zainteresuje się poszerzeniem swoich badań również w tym zakresie.

Synteza mowy zarówno na poziomie akustyczno-fonetycznym, jak i lingwistycznym wymaga akustycznych analiz sygnału mowy i odpowiedniej ich interpretacji. Analizujemy dźwięki języka naturalnego, dźwięki mowy, która ma określona strukturę nie tylko akustyczną, ale i lingwistyczną. Nie wolno ani na moment zapominać, iż analizujemy jedną z płaszczyzn języka, a nie przypadkowe ciągi dźwięków.

Wszystkie systemy rozpoznawania mowy muszą więc zawierać moduł akustyczny oraz lingwistyczny obejmujący w różnym stopniu poziomy: fonologiczny, fonetyczny, morfologiczny, leksykalny i składniowy.

Zastrzeżenia, które ujęłam w niniejszej recenzji związane są głównie z interpretacją fonetyczną i lingwistyczną wyników, nie stanowią przedmiotu pracy i dlatego nie mają żadnego wpływu na moją wysoką ocenę dokonań autora zarówno na poziomie teoretycznym, jak i praktycznym. Uważam opiniowaną rozprawę jako bardzo wartościową i stanowiąca ważne osiągnięcie w zakresie korpusowej syntezy mowy

3. Uwagi szczegółowe

3.1. Akapit umieszczony na str. 90 (punkt 4.1) powinien zapewne zostać włączony do punktu 4.2., w którym autor omawia realizację bazy akustycznej: „ *Baza akustyczna została zweryfikowana pod katem prozodycznym. Wykonano stylizacje Insint (Hirst*

1999), bedaca systemem anotacji wzorców prozodycznych. (Rozdział 1.5.3) Dodatkowo przeprowadzono prozodyczna anotacje dla przerw miedzy frazami w zdaniach. Przeprowadzone badania wskazuja ze stworzony korpus jest prozodycznie bogaty.

Warto tu byłoby umieścić również kilka przykładów anotacji prozodycznej przeanalizowanego korpusu.

3.2. Opis na stronie 16 odnoszący się do omówienia podstawowych problemów organizacji dźwiękowej wypowiedzi języka naturalnego jest bardzo niejasny. Należałoby bezwzględnie wyjaśnić jak Doktorant rozumie podstawowe pojęcia fonetyczne takie jak: fonem, alofon, sylaba, akcent, iloczas, rytm, intonacja, struktura prozodyczna.

3.3. Na stronie 18 podano rozważania odnośnie iloczasu: „ *Czas trwania głoski (iloczas) związany jest również ze sposobem artykulacji. Nieco krócej trwaja głoski ustne a spółgłoski nosowe sa najkrótszymi głoskami. Iloczas trwania głoski jest zawiazany z czasem jej artykulacji i jest użyteczny przy określaniu zmiany iloczasu głoski odpowiednio do otaczającej jej dzwieków mowy.*

Konieczna byłaby tu kompleksowa analiza czynników wpływających na iloczas i rytm oraz odpowiedź na pytanie jak mierzymy iloczas, jak normalizujemy iloczas, jak definiujemy i mierzymy rytm wypowiedzi.

3.4. Na stronie 20 autor stwierdza: „ *W zależności od tego, który z tych czynników przeważa, akcent jest określany, jako:*

dynamiczny, – gdy czynnikiem dominującym w płaszczyźnie akustycznej są chwilowe zmiany intensywności

rytmiczny – gdy o wrażeniu akcentu decydują zmiany iloczasów sylab,

melodyczny – gdy akcentowanie sylaby jest realizowane poprzez chwilową zmianę wysokości głosu.

Na podstawie więc jakich wyznaczników akustycznych oraz w jaki sposób można określić praktyczną możliwość automatycznego wyznaczania akcentu.

3.5. Podrozdział 1.4 warto byłoby uzupełnić o dwa podstawowe warianty wymowy: warszawską oraz poznańsko-krakowską.

Reguły transkrypcji wydają się też nie w pełni kompletne (np. ę, ą przed trącymi; np. jak transkrybujemy takie wyrazy jak: *ciąża, więź*).

3.6. Bardzo ciekawe stwierdzenie pojawiło się na str.31: *„Sylaba jest fonetyczno-fonologiczna jednostka słowa jak i jednym z bardziej spornych zagadnien w fonetyce. Definicje sylaby podano w 1.3.6. Należy dodac, iż segmentacja sylab jest względnie łatwa".*

Interesujące jest, więc zagadnienie praktycznego podziału sygnału mowy na sylaby akustyczne. Według mojej wiedzy, jak dotychczas dla języków europejskich nie udało się rozwiązać tego problemu w sposób zadawalający.

3.7. Na str.124 Autor stwierdza: *„Kryterium wyboru najlepiej brzmiącego zdania, – czyli osobnika w przeszukiwanej populacji, było znalezienie syntetycznego zdania z najmniejsza ilością bledów łączeniowych, prozodycznych, intonacyjnych. Najwiekszy wpływ miało prawidłowe łączenie sąsiadujących ze sobą elementów, niż dobre odtworzenie cech prozodycznych.*

Potrzebne byłoby tu wyjaśnienie, czy eksperci kierowali się liczbą błędów, czy ogólną, jakością syntezy.

3.8 Na str.110 Autor stwierdza: *„ Jeśli mówca mówi zbyt szybko, wówczas dokładność segmentacji maleje, ponieważ przewidywana z tekstu sekwencja wystąpienia określonych głosek jest znacznie mniej prawdopodobna, niż ta, która została wypowiedziana. W wyniku tego następuje desynchronizacja sygnału z tekstem ortograficznym oraz jego transkrypcja fonetyczna. W pierwszym prototypowym głosie nie zwrócono uwagi na wpływ tempa mowy na dokładność segmentacji i etykietyzacji, dlatego jakość mowy syntetycznej była początkowo niezadowalająca.*

Jakie tempo wypowiedzi uważa Doktorant za właściwe dla syntezy korpusowej? Czy wolne tempo wypowiedzi zapewni poprawną segmentacje?

3.9. Czy rozważano jakąkolwiek metodę wygładzania zsyntezowanego sygnału (np. w zakresie widma oraz/lub fluktuacji Fo).

3.10 Nie jest jasny sposób wykorzystania informacji prozodycznych w module lingwistycznym. W jaki sposób wykorzystano informacje o POS w modelowaniu intonacyjnym wypowiedzi.

4. Uwagi techniczne

4.1. Układ pracy

Ogólnie dysertacja stanowi spójną całość. Jednakże układ rozdziałów wymaga weryfikacji. Przykładowo, podrozdział 1.3.1 *Klasyfikacja dźwięków mowy* zawiera tylko jedno zdanie. Warto byłoby również pomyśleć o włączeniu do rozprawy zagadnienia percepcji mowy, choćby w zakresie podstawowym. Bardzo byłoby to przydatne do interpretacji wyników. Ocena jakości syntezy wymaga zarówno fonetyczno-akustycznych analiz, jak i percepcyjnych ocen. Nie zawsze tak jest, iż duże zniekształcenia akustyczne powodują znaczne zaburzenia percepcji. Często również relatywnie małe nieciągłości akustyczne (z punktu widzenia pomiaru fizycznego) wywołują drastyczne pogorszenie percepcji sygnału.

4.2. Rysunki przedstawiające sygnał mowy należy opracować w sposób przejrzysty. Zupełnie pozbawiona jest sensu prezentacja przebiegów formantowych w ciszy (np. rys.1.6., rys.1.7.). Praat nie posiada dobrego modułu wyznaczania formantów, a więc do wyników tej analizy należy podchodzić z dużą ostrożnością.

4.3. Prezentacja niektórych tabel np. 4.2, 4.3 wymaga opracowania graficznego (właściwego formatu liczb, opisu tabeli, wizualizacji).

4.4. Podpisy pod rysunkami są niekiedy nieadekwatne do treści rysunku (np. rys.4.5, str.86). Na str.54 podano analizę nieodpowiadających sobie wypowiedzi.

4.5. W pracy pojawiły się różnego typu błędy gramatyczne; powtarzającym się błędem jest brak kropki na końcu zdania. Zdania niekiedy rozpoczynają się z małej litery lub zaczynają się od nawiasu.

Np. na str 4 czytamy:

Na brzegach fałdów głosowych znajduja sie wiazadła głosowe. (Stevens 1998) W tyle krtani wiazadła głosowe sa przymocowane do wyrostków głosowych, które mogą sie od siebie oddalać lub przybliżać.

Zaś na str.13: Poniżej *opisano krótko klasyfikacje samogłosek na podstawie czworoboku samogłoskowego, opracowanego przez angielskiego fonetyka Daniela. Jonesa.(Jones 1918) (Rysunek 1.9)*

4.6 Błędy stylistyczne, które można i trzeba byłoby usunąć przed ewentualną publikacją rozprawy, ponieważ w niektórych przypadkach prowadzą one do rozmaitych wątpliwości np.

Str.18. *Iloczas trwania głoski jest zawiązany z czasem jej artykulacji i jest użyteczny przy określaniu zmiany iloczasu głoski odpowiednio do otaczającej jej dźwięków mowy;*

str. 67. *„Z przeprowadzonych badan wynika, że najlepsza korelacja pomiędzy kosztem akustycznym oraz percepcją ekspertów lingwistycznych nie przekracza 0, 66 co jest niezadowalającym wynikiem z naukowego punktu widzenia.*

Str.*32: „ W rozdziale przedstawione zostały ogólne zagadnienia związane z fonetyka akustyczna obrazującą sposób opisu dźwięków człowieka mowy w płaszczyźnie artykulacyjnej. Przedstawiono budowę narządu człowieka oraz klasyfikacje dźwięków przez niego artykułowanych.*

4.7. Autor nie zawsze powołuje się na rysunki, brak jest opisów do niektórych rysunków. W spisie bibliografii podaje pozycje, na które również się również nie powołuje, np.:

Koza J.R., Rice J. P. (1991) *Genetic generation of both the weights and architecture for a network Neural Networks,* 1991 IEEE international conference pp: 397-044.

Fant G. (1970) *Acoustic Theory of Speech Production*, The Hagues, Mounton.

Występują również niewłaściwe powołania, np. str.35:

Artykulacyjna synteza mowy polega na modelowaniu rzeczywistego narządu artykulacyjnego. Model ten wymaga dynamicznego modelu toru głosowego, który pozwala na symulacje ruchu artykulatorów podczas procesu generowania mowy. (Wagner 2008). Praca pani Wagner zdecydowanie więcej wnosi informacji o strukturze prozodycznej mowy, niż o syntezie artykulacyjnej.

Podsumowanie

W posumowaniu mojej oceny dysertacji doktorskiej zatytułowanej: *Optymalizacja funkcji kosztu w korpusowej syntezie mowy polskiej,* pana mgr inż. Krzysztofa Szklannego stwierdzam, co następuje. Autor samodzielnie rozwiązał istotny problem naukowy związany z optymalizacją funkcji kosztu za pomocą algorytmu ewolucyjnego. Zastosowana strategia ewolucyjna oraz przeprowadzone badania wskazują, iż funkcje kosztu można optymalizować za pomocą metod heurystycznych, a proces optymalizacji funkcji kosztu ma wpływ na jakość syntezy korpusowej. Jako ważny należy uznać praktyczny wynik w postaci w pełni funkcjonującego systemu korpusowej syntezy mowy w środowisku Festival. Opracowana została również innowacyjna technika poprawiająca jakość segmentacji automatycznej sygnału mowy.

W konkluzji stwierdzam, że opiniowana rozprawa z naddatkiem spełnia wymogi stawiane przez ustawę o stopniach i tytułach naukowych dla prac doktorskich, a jej autor zasługuje na przyznanie stopnia naukowego doktora nauk technicznych.

Stawiam więc wniosek o dopuszczenie tej rozprawy do publicznej obrony.