

Autoreferat

Krzysztof Szklanny

Szklanny

Spis treści

1	Imię i nazwisko.....	3
2	Posiadane dyplomy i stopnie.....	3
3	Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych	3
4	Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.):	3
4.1	Tytuł osiągnięcia naukowego (cykl 6 publikacji).....	3
4.2	Wykaz publikacji wchodzących w skład osiągnięcia naukowego.....	3
5	Omówienie celu naukowego w/w prac i osiągniętych wyników	5
5.1	Wprowadzenie	5
5.2	Automatyczna klasyfikacja jakości głosu	5
5.3	Podsumowanie	13
5.4	Literatura	13
5.5	Publikacje niewchodzące w skład osiągnięcia naukowego znajdujące się w Web of Science.....	16
6	Omówienie pozostałych osiągnięć naukowo-badawczych	17
6.1	Badania niezwiązane z głównym celem badawczym.....	17
6.1.1	Realizacja systemu multimodalnej syntezy mowy.....	17
6.1.2	Przygotowanie nowej bazy akustycznej.....	18
6.1.3	Realizacja kursów m-learningowych przy użyciu metody Puzzle Learning oraz tworzenie aplikacji multimedialnych wspomagających naukę.....	19

1 Imię i nazwisko

Krzysztof Szklanny

2 Posiadane dyplomy i stopnie

Magistra: Wydział Informatyki, Polsko-Japońska Wyższa Szkoła Technik Komputerowych, „Przygotowanie bazy difonów języka polskiego dla realizacji syntezy mowy w systemie MBROLA”, 2002 r.

Doktora: Wydział Informatyki, Polsko-Japońska Wyższa Szkoła Technik Komputerowych, „Optymalizacja funkcji kosztu dla realizacji korpusowej syntezy mowy języka polskiego”, 2010 r.

3 Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

2002- 2010 — Polsko-Japońska Akademia Technik Komputerowych w Warszawie, Wydział Informatyki, Katedra Multimediów (asystent)

2002- obecnie — Profesjonalny fotograf, specjalista cyfrowej fotografii produktowej, komercyjnej, ultra szybkiej,

2005-2006 Instytut Badań Edukacyjnych w Warszawie, (wykładowca)

2005-2007 Akademia Humanistyczna w Pułtusku, (wykładowca)

2006-2007 Wszechnica Polska, Szkoła Wyższa w Warszawie, (wykładowca)

2006 - obecnie — Polsko-Japońska Akademia Technik Komputerowych w Warszawie, (Pełnomocnik Rektora ds. Praktyk Studenckich, dla pięciu Wydziałów Uczelni)

2010- obecnie — Polsko-Japońska Akademia Technik Komputerowych w Warszawie, Wydział Informatyki, Katedra Multimediów, (adiunkt)

4 Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.):

4.1 Tytuł osiągnięcia naukowego (cykl 7 publikacji)

„Analiza akustyczna i automatyczna klasyfikacja w szczegółowej ocenie jakości głosu”.

4.2 Wykaz publikacji wchodzących w skład osiągnięcia naukowego

Sumaryczny Impact Factor publikacji wchodzących w skład osiągnięcia naukowego wynosi **18,177** a sumaryczny pięcioletni Impact Factor równy jest **20,923**.

ORCID: 0000-0001-6540-1671

1. Szklanny K. [70%], Gubrynowicz R. [10%], Ratyńska J. [10%], Chojnacka-Wądołowska D. [10%] Electroglottographic and acoustic analysis of voice in children with vocal nodules, International Journal of Pediatric Otorhinolaryngology, 2019, DOI:10.1016/j.ijporl.2019.03.030 (IF=1,305, punkty MNiSW: 20)
Mój wkład w powstanie tej pracy polegał na opracowaniu koncepcji eksperymentu; uzyskaniu zgody Komisji Bioetycznej na realizację badań; przygotowaniu bazy nagrań pacjentów z guzkami głosowymi oraz osób zdrowych; analizie nagrań elektroglotograficznych i akustycznych; parametryzacji, klasyfikacji sygnału i oszacowaniu jakości głosu; opracowaniu wyników; realizacji

testów statystycznych; przygotowaniu, napisaniu i poprawie artykułu zgodnie z uwagami recenzentów. Mój udział szacuję na 70%.

2. **Szklanny K. [60%], Wrzeciono P. [40%]** Relation of RBH auditory-perceptual scale to acoustic and electroglottographic voice analysis in children with vocal nodules, IEEE ACCESS, 2019, DOI: 10.1109/ACCESS.2019.2907397 (**IF=3,557**, punkty MNiSW: 25)

Mój wkład w powstanie tej pracy polegał na opracowaniu koncepcji eksperymentu; przygotowaniu nagrań do bazy danych; ocenie perceptualnej nagrań; parametryzacji i klasyfikacji nagrań akustycznych i elektrogłotograficznych; przeprowadzeniu eksperymentu dotyczącego predykcji szczegółowej klasy skali RBH za pomocą parametrów akustycznych; realizacji testów statystycznych; opracowaniu wyników; napisaniu artykułu; jego poprawie zgodnie z uwagami recenzentów. Mój udział szacuję na 60%.

3. **Szklanny K. [53%], Wrzeciono P. [47%]** The application of a genetic algorithm in the noninvasive assessment of vocal nodules in children, IEEE ACCESS 2019, DOI: 2019.10.1109/ACCESS.2019.2908313 (**IF=3,557**, punkty MNiSW: 25)

Mój wkład w powstanie tej pracy polegał na opracowaniu koncepcji eksperymentu, realizacji nagrań; wyszukaniu skutecznych parametrów do oceny jakości głosu; przeprowadzeniu eksperymentów przy użyciu sieci DNN; napisaniu kilku rozdziałów artykułu; poprawie artykułu zgodnie z uwagami recenzentów. Mój udział szacuję na 53%.

4. **Szklanny K. [100%]** Acoustic Parameters in the Evaluation of Voice Quality of Choral Singers. Prototype of Mobile Application for Voice Quality Evaluation, Archives of Acoustics, 2019, DOI: 10.24425/aoa.2019.129257 (**IF=0,917**, punkty MNiSW: 15)

Mój wkład w powstanie tej pracy polegał na stworzeniu koncepcji eksperymentu; przygotowaniu bazy danych z nagraniami chórzystów oraz grupy kontrolnej; realizacji prototypu aplikacji webowej pozwalającej na monitorowanie zmian jakości głosu za pomocą smartfona; parametryzacji i klasyfikacji nagrań; ocenie jakości głosu i detekcji zmian w głosie u chórzystów i chórzystek; przeprowadzeniu eksperymentów; przygotowaniu testów statystycznych; opracowaniu wyników; przygotowaniu i poprawie artykułu zgodnie z uwagami recenzentów

5. **Szklanny K. [70%], Gubrynowicz R. [10%], Iwanicka-Pronicka K. [10%], Tylki-Szymańska A [10%]**. Analysis of voice quality in patients with late-onset Pompe disease. Orphanet J Rare Dis. 2016;11(1):99. DOI: 10.1186/s13023-016-0480-5, praca oryginalna (**IF = 3,478**, punkty MNiSW: 35)

Mój wkład w powstanie tej pracy polegał na stworzeniu koncepcji eksperymentu; przygotowaniu bazy danych z nagraniami pacjentów; parametryzacji i klasyfikacji sygnału głosu; ocenie jakości głosu; detekcji zmian w głosie; przeprowadzeniu eksperymentów; opracowaniu wyników; przygotowaniu i poprawie artykułu zgodnie z wymogami recenzentów. Mój udział procentowy szacuję na 70%.

6. **Szklanny K. [80%], Tylki-Szymańska A. [20%]** Follow-up analysis of voice quality in patients with late-onset Pompe disease, Orphanet J. Rare Dis., 2018, (**IF=3,607**, punkty MNiSW: 35), DOI: 10.1186/s13023-018-0932-1

Mój wkład w powstanie tej pracy polegał na stworzeniu koncepcji eksperymentu; przygotowaniu bazy danych z nagraniami pacjentów; oszacowaniu zmian jakości głosu za pomocą parametrów akustycznych; klasyfikacji jakości głosu; przeprowadzeniu eksperymentów; porównaniu wyników z nagraniami sprzed 3 lat; opracowaniu wyników; przygotowaniu, poprawie artykułu zgodnie z wymogami recenzentów. Mój udział procentowy szacuję na 80%.

7. **Szklanny K. [80%], Gubrynowicz, R. [10%], & Tylki-Szymańska, A. [10%]**, (2018). Voice alterations in patients with Morquio a syndrome. Journal of applied genetics, 59(1), 73-80 DOI: 10.1007/s13353-017-0421-6 (**IF = 1,756**, Punkty MNiSW: 20). Mój udział procentowy szacuję na 80%.

Mój wkład w powstanie tej pracy polegał na stworzeniu koncepcji eksperymentu; przygotowaniu bazy danych z nagraniami pacjentów oraz osób zdrowych, oszacowaniu zmian jakości głosu za pomocą parametrów akustycznych; klasyfikacji jakości głosu; przeprowadzeniu eksperymentów; opracowaniu wyników; przygotowaniu, poprawie artykułu zgodnie z wymogami recenzentów. Mój udział procentowy szacuję na 80%.

5 Omówienie celu naukowego w/w prac i osiągniętych wyników

5.1 Wprowadzenie

Współcześnie odnotowuje się coraz więcej przypadków chorób aparatu mowy. Szacuje się, że obejmują one 30% społeczeństwa (Roy 2004, Battle 2012, ASHA 2018). Liczne raporty na temat zdrowia również potwierdzają ten problem. Skala ta świadczy o konieczności poszukiwania nowych metod klasyfikacji sygnału akustycznego, które będą zgodne z metodami dotychczas stosowanymi w diagnostyce lekarskiej. Zgodnie z protokołem opracowanym przez Komitet Foniatrii Europejskiego Towarzystwa Laryngologicznego (ELS) analiza percepcyjna i analiza akustyczna głosu należą do podstawowych narzędzi badawczych stosowanych w diagnostyce zaburzeń głosu (Dejonckere 2001, 2004). Zastosowanie klasyfikatorów automatyzuje ten proces.

Celem pracy habilitanta było opracowanie takich rozwiązań automatycznej klasyfikacji sygnału, które będą prowadzić do uzyskania szczegółowej diagnozy głosu. Jest to temat ważny i nowatorski, jednak ciągle rzadko poruszany w literaturze z uwagi na trudność organizacji badań medycznych.

W celu przeprowadzenia eksperymentów, habilitant musiał przygotować odpowiednie bazy danych nagrań. Proces ich tworzenia trwał pięć lat. Nagrania te są unikalne, ponieważ obejmują również choroby rzadkie.

Habilitant opublikował prace z tej tematyki w prestiżowych czasopismach, co jest potwierdzeniem oryginalności zastosowanych metod. Mogą być one stosowane u pacjentów z guzkami głosowymi oraz chorobami rzadkimi takimi jak choroba MPS czy choroba Pompego. Habilitant potwierdził możliwość stosowania tych metod również u osób zdrowych, ale ze względów zawodowych mogących mieć tendencję do nadwyższenia głosu – chórzystów. Habilitant opracował również metodę pozwalającą na automatyczne uzyskanie oceny subiektywnej jakości głosu poprzez zastosowanie analizy akustycznej.

Powtarzalność wyników daje możliwość zastosowania opracowanych metod również w innych chorobach głosu. Wyniki eksperymentów opublikowane zostały w czasopismach o wysokim Impact Factorze.

Pierwszym krokiem zmierzającym do realizacji osiągnięcia naukowego było nawiązanie współpracy z zewnętrznym ośrodkiem naukowym — Instytutem "Pomnik – Centrum Zdrowia Dziecka" (IP-CZD). Na mocy porozumienia pomiędzy PJWSTK obecnie (PJATK), habilitant wystąpił przed Komisją Bioetyczną i uzyskał zgodę na prowadzenie badań (sygnatura dokumentu: 133/KBE/2014, załącznik 7).

5.2 Automatyczna klasyfikacja jakości głosu

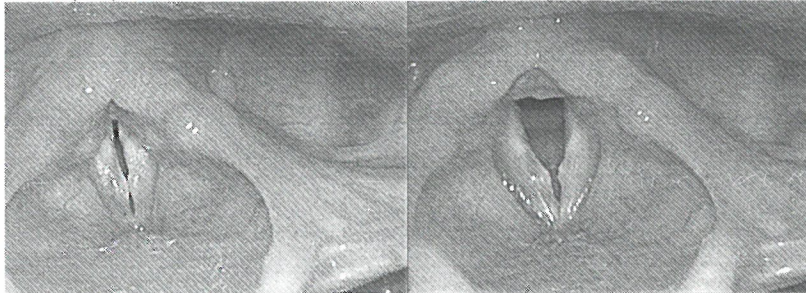
Pierwsza z przedstawianych prac (Szkłanny 2019a) podejmuje zagadnienie porównania analizy elektrolotograficznej z analizą akustyczną sygnału mowy u pacjentów z guzkami głosowymi. Guzki głosowe są jedną z najczęściej występujących chorób krtani wśród dzieci. Wielu autorów publikowało wyniki badań wskazujących jednoznacznie na skalę problemu. Częstotliwość występowania guzków głosowych jest wysoka, wynosi ok. 20-30% w populacji dzieci (Johns 2003, Kilić 2004, De Bodt 2007). Obecność guzków głosowych objawia się przewlekłą chrypką (Shah 2005, Shah 2007, Nuss 2012, Lee 2018, Mudd 2018, Guo 2018). Istnieją metody pozwalające na odróżnienie głosu pacjentów z guzkami głosowymi od głosu osób zdrowych, ale nadal brakuje metod porównujących jakość głosu za pomocą metod akustycznych i elektrolotograficznych, ze wskazaniem skuteczności tych metod oraz wnoszących precyzyjne informacje o jakości głosu u tych osób.

W celu realizacji tego eksperymentu habilitant przygotował bazę nagrań dzieci z guzkami głosowymi. Pacjent, u którego podejrzewano wystąpienie guzków głosowych, był kierowany do foniatrii przez lekarza pierwszego kontaktu. Każdy z pacjentów został poddany badaniu foniatrycznemu i wideolaryngoskopowemu. Badanie wideolaryngoskopowe pozwala na zapis pracy fałdów głosowych za pomocą kamery wideo. W opisie lekarskim krtani, realizowanym przez dwóch lekarzy foniatrów, wprowadzono anotację nagrania, zawierającą między innymi informację o stopniu braku zwarcia fałdów głosowych. (Jest to oryginalne rozwiązanie ponieważ pozwoli na szczegółową diagnozę w kolejnych eksperymentach habilitanta. Dotychczas takie rozwiązanie nie było opisane w literaturze). W przypadku potwierdzenia występowania guzków głosowych, pacjent kierowany był na tygodniowy turnus rehabilitacyjny w szpitalu. Podczas turnusu habilitant realizował nagrania pacjentów z guzkami głosowymi. Baza nagrań była realizowana w IP-CZD, przez okres dwóch lat. Nagrania były tworzone za pomocą elektrolotografu Glottal Enterprises EG-PCX2. Zapis elektrolotograficzny bazuje na pomiarze impedancji elektrycznej pomiędzy fałdami głosowymi. Metoda została opracowana przez Fabre (Fabre 1957), a znaczący wkład w jej rozwój mieli Frokjaer-Jensen i Fourcin (Frokjaer-Jensen 1968, Fourcin 1972). Pomiaru sygnału dokonuje się poprzez umieszczenie dwóch elektrod w okolicy fałdów głosowych. Jest to dokładna metoda pomiarowa, w tym przypadku referencyjna w odniesieniu do analizy akustycznej. Habilitant zarejestrował nagrania 80 pacjentów w tym 23 zdrowych, którzy stanowią część grupy kontrolnej.

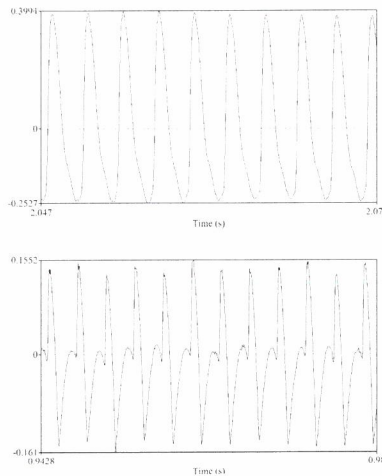
Uzupełnieniem grupy kontrolnej były nagrania osób zdrowych udostępnione w bazie Saarbrücken Voice Database (SVD) (Barry 1997). Warunkiem poprawności badań w diagnostyce medycznej jest zbadanie powtarzalności uzyskanych rezultatów, dlatego zastosowano dane pochodzące z innej bazy. Porównanie bazy SVD i stworzonej przez habilitanta zostanie przedstawione w kolejnej części.

Na rysunku 1 umieszczono przykładowy obraz fałdowych głosowych wraz z opisem lekarskim. Rysunek 2 zawiera zapis elektrolotograficzny u:

- a) osoby zdrowej (góra)
- b) pacjenta z guzkami (dół).



Rys. 1 Ocena foniatryczna pacjenta z guzkami głosowymi. Guzki położone w 1/3 części fałdów, wielkość 3, szerokie, symetryczne, brak zwania fałdów głosowych w stopniu 1. Po lewej fałdy głosowe podczas fonacji, po prawej podczas oddychania.



Rys. 2 Przykładowy zapis sygnału EGG u a) osoby zdrowej (góra), b) pacjenta z guzkami (dół).

Spośród wielu dostępnych parametrów akustycznych, do analizy używa się tych, które powinny wnosić kluczowe informacje o jakości głosu. Trudność polega na znalezieniu takich parametrów, które będą prawidłowo opisywały stan zaawansowania choroby z różnymi patologiami głosu np. guzki głosowe, cysta, polip. Przeprowadzono wiele eksperymentów, przy zastosowaniu następujących parametrów akustycznych: Peak Slope (PS) (Kane 2011), H1H2 (Titze 1992), Normalized Amplitude Quotient (NAQ) (Alku 2002), Parabolic Spectral Parameter (PSP) (Alku 1997), Quasi Open Quotient (QOQ) (Hacki 1996), Cepstral Peak Prominence (CPP) (Hillenbrand 1996), Harmonic Richness Factor (HRF) (Childers 1996), Rd (Fant 1995) w modelu toru głosowego Liljencrants-Fanta (LF) (RDS) (Fant 1985), Maxima Dispersion Quotient (MDQ) (Kane 2013), oraz parametru Closed Quotient (CQ H) zaproponowanego przez Howarda (Howard 1990, 1995) wyliczanego z sygnału EGG. Najbardziej istotnymi parametrami okazały się Peak Slope, NAQ oraz CQ H. Parametr Peak Slope charakteryzuje się 70% skutecznością w identyfikacji jakości głosu przy SNR = 10 dB, a parametr NAQ przy takim samym SNR pozwala na 65% skuteczność identyfikacji jakości głosu (Kane 2011).

Wyniki przeprowadzonych eksperymentów wskazały jednoznacznie na wyższą skuteczność analizy elektrolotograficznej w opisie sygnału u dzieci z guzkami głosowymi w odniesieniu do analizy akustycznej. U 54 z 57 pacjentów zaobserwowano w sygnale EGG odchylenie od kształtu cechującego głos modalny. Kształt ten jest charakterystyczny dla występowania guzków głosowych. Brak zwania fałdów głosowych zaobserwowano u 54 pacjentów (CQ H = 0,315 (dzieci z guzkami głosowymi) vs 0,411 (grupa kontrolna)).

Scun

Różnice w wymienionych wartościach parametrów dla grupy pacjentów oraz osób zdrowych są statystycznie istotne (Tabela 1).

W analizie akustycznej stwierdzono, u 36 pacjentów głos chuchający (Peak Slope = $-0,19$ vs. $-0,28$). Fonację wysiłkową wykryto u 14 pacjentów. Było to pośrednio wskazane przez średnią wartość NAQ wynoszącą $0,14$ w grupie pacjentów z guzkami w stosunku do NAQ = $0,12$ w grupie kontrolnej. Różnice między medianami a średnimi wartościami parametrów są statystycznie nieistotne, co świadczy o stabilności parametrów.

Tabela 1. Średnie wartości i odchylenie standardowe parametrów CQ H, Peak Slope, NAQ dla pacjentów z guzkami i osób zdrowych.

Parametr	Grupa 1	Grupa 2	p-value
CQ H	$0,315 \pm 0,061$	$0,411 \pm 0,066$	$<0,0001$
PS	$-0,193 \pm 0,095$	$-0,276 \pm 0,083$	$0,0001$
NAQ	$0,138 \pm 0,033$	$0,119 \pm 0,026$	$0,0017$

Otrzymane wyniki wskazały na wysoką zgodność analizy sygnału EGG z badaniem wideolaryngoskopowym. Analiza akustyczna okazała się mniej skuteczna od analizy EGG, ale wniosła istotne informacje na temat sposobu fonacji u tych pacjentów. Zastosowane parametry potwierdziły swoją skuteczność w kolejnych badaniach habilitanta. Wybór parametrów akustycznych i EGG pozwolił na osiągnięcie lepszych wyników niż prezentowanych w literaturze (Chandrashekarappa 2016). Otrzymane przez habilitanta rezultaty wyznaczyły nowy kierunek badań, dotyczących próby predykcji oceny perceptualnej głosu tych pacjentów za pomocą analizy akustycznej. Badania te stały się celem drugiej pracy (Szkłanny 2019b).

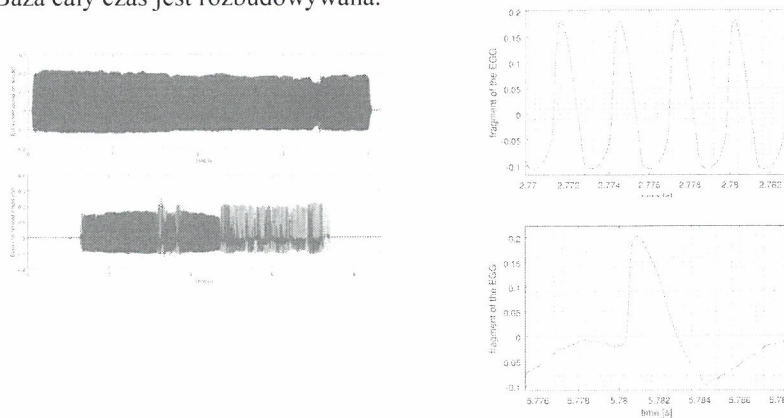
Habilitant stworzył bazę danych zawierającą nagrania pacjentów z guzkami głosowymi i opublikował ją udostępniając szerszemu gronu naukowców. Obecnie dostępne są następujące bazy zawierające nagrania głosu patologicznego: Saarbrücken Voice Database (SVD) oraz Massachusetts Eye and Ear Infirmary (MEEI) database (MEEI 1994, Barry 1997). Baza SVD jest jedyną bezpłatną i ogólnodostępną bazą danych zawierającą nagrania dźwiękowe i elektrogłotograficzne głosu osób zdrowych i chorych.

Baza Massachusetts Eye and Ear Infirmary (MEEI) jest płatną bazą zawierającą nagrania głosu patologicznego oraz zdrowego. Nagrania w bazie MEEI zostały zrealizowane w Kay's Computerized Speech Lab (MEEI 1994).

Baza stworzona przez habilitanta wyróżnia się następującymi cechami w porównaniu do dwóch wymienionych powyżej:

1. Zawiera nagrania głosu dzieci z guzkami głosowymi. W bazie SVD występuje tylko jedno takie nagranie.
2. Baza składa się z nagrań przedłużonej fonacji samogłoski /a:/ dziecka z guzkami głosowymi wraz z początkiem i końcem, co stanowi istotną informację w klasyfikacji sygnału akustycznego.
3. W bazie zarejestrowano kilka zdań, dzięki czemu zwiększyła się wiarygodność fonacji pacjenta.
4. Baza zawiera nagrania akustyczne i elektrogłotograficzne.
5. W bazie udostępniono opis lekarski krtani włącznie z informacją o niedomykalności fałdów głosowych. Jest to nowatorskie rozwiązanie, które powoli habilitantowi na przeprowadzenia eksperymentów w zakresie diagnozy szczegółowej opisanych w kolejnej pracy.
6. Baza posiada ocenę głosu w skali perceptualnej RBH (Nawka 1994).
7. Baza zawiera również zapis nagrań wideolaryngoskopowych, które będą udostępniane w przyszłości.
8. Baza została sparametryzowana przy użyciu 9 parametrów akustycznych i jednego parametru obliczonego z sygnału EGG.
9. W bazie udostępniono wizualizację nagrania dla samogłoski /a/. Fragment wizualizacji przedstawiono na rysunku 3. Wizualizacja pozwala na rozpoznanie nieprawidłowości sygnału po jej obejrzeniu. Składa się ona z 6 wykresów – oscylogramu i elektrogłotogramu dla całej fonacji, następnie 1% długości sygnału pobranego ze środka dla oscylogramu i EGG oraz dwóch wykresów przedstawiających kilka okresów krtaniowych EGG pobranych z 1/3 i 2/3 sygnału.
10. Dane dotyczące chorób pacjentów oraz umożliwiające jednoznaczną identyfikację osoby są szczególnie chronione. Są one przechowywane zgodnie z Rozporządzeniem Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r.

11. Baza została udostępniona pod adresem: <https://vpadb.wzim.sggw.pl/>.
12. Dostęp do bazy jest możliwy po wcześniejszej rejestracji.
13. Baza cały czas jest rozbudowywana.



Rysunek 3. Po lewej wizualizacja sygnału EGG u osoby zdrowej (góra) i pacjenta z guzkami (dół). Po prawej stronie kilka okresów kraniowych sygnału EGG u osoby zdrowej (góra) i pacjenta z guzkami (dół).

Baza danych została przygotowana w oparciu o model relacyjny zaimplementowany w MySQL. Wszystkie dane wrażliwe, podlegające ochronie prawnej, są przechowywane w formie zaszyfrowanej. Istotną funkcjonalnością bazy jest możliwość jej archiwizacji z wykorzystaniem SQL. Umożliwia to relatywnie mało pracochłonne przeniesienie bazy do innego systemu bazodanowego. Podstawowym interfejsem użytkownika jest strona internetowa, napisana jako aplikacja zgodna ze wzorcem MVC. Proces rejestracji odbywa się poprzez wypełnienie formularza i wysłanie go za pomocą strony bazy do administratora danych. Po zalogowaniu się do serwisu, użytkownik ma dostęp do danych według przyznanych uprawnień.

Istotnym aspektem pracy nad bazą była jej walidacja. Habilitant przeprowadził eksperyment sprawdzający czy jest możliwa predykcja stopni skali perceptualnej RBH za pomocą parametrów akustycznych. Istnieje kilka skal perceptualnych, między innymi skala GRBAS, CAPE-V i RBH (Hirano 1981, ASHA 2002, Kempster 2009, Nawka 1994). Skala RBH została opracowana na podstawie skali GRBAS i jest jej uproszczoną wersją. Skala percepcyjna RBH stosowana jest w niemieckich klinikach i jest rekomendowana przez Komitet ds. Foniatrii Europejskiego Towarzystwa Laryngologicznego (Dejonckere 2001, 2004). Litery RBH stanowią akronim następujących cech:

R - Rauigkeit – szorstkość głosu wynikająca z nieregularności drgań fałdów głosowych;

B - Behauchtheit – głos chuchający wynikający z wydobywania się powietrza podczas fonacji przez niezamkniętą głośnie;

H - Heiserkeit – stopień chrypki;

Skala RBH przypisuje ocenę 0, 1, 2 i 3 dla wszystkich parametrów, w zależności od stopnia zaburzeń głosu, gdzie "0" oznacza normalny głos, "1" – niewielki stopień zmian głosu, "2" – średnie zmiany, i "3" – duże zmiany.

Dwóch niezależnych ekspertów przeprowadziło ocenę jakości głosu dla pacjentów z guzkami głosowymi oraz dla osób zdrowych przy użyciu skali RBH. Eksperti poddani byli badaniu audiometrycznemu, które nie wykazało problemów ze słuchem. Wyniki przeprowadzonych analiz statystycznych, pokazały że różnice w anotacji pomiędzy ekspertami są statystycznie nieistotne (Chi square test, p value > 0.05). Przeprowadzone eksperymenty z użyciem parametrów akustycznych potwierdziły, że możliwe jest odróżnienie szczegółowej oceny perceptualnej przez zastosowanie odpowiednich parametrów akustycznych i elektroglograficznych:

- Klasy R0*, R0 i R1 można rozróżnić poprzez zastosowanie parametrów CQ H i NAQ dla eksperta 1
- Klasy R0*, R1 i R2 można rozróżnić poprzez zastosowanie parametrów CQ H i NAQ dla eksperta 2
- Klasy B0*, B0 i B1 można rozróżnić poprzez zastosowanie parametrów CQ H i NAQ dla eksperta 1 i 2
- Klasy H0*, H1 i H2 można rozróżnić poprzez zastosowanie parametrów CQ H, NAQ i PS dla eksperta 1 i 2

Wyniki analiz zostały potwierdzone testami statystycznymi ANOVA. Symbol * oznacza osoby zdrowe z grupy kontrolnej.

Signature

Opisane dwie prace udowodniły skuteczność zarówno wybranych metod jak i parametrów akustycznych i EGG w analizie sygnału głosu ze znaczącym stopniem patologii. Habilitant uzyskał pozytywny wynik walidacji bazy danych. Jednocześnie wyniki eksperymentu dotyczącego skali RBH wskazały nowy obszar badań w postaci wyznaczania diagnozy szczegółowej.

Spośród 10 parametrów akustycznych finalnie wybrano cztery – CQ H, Peak Slope, NAQ i RDS. Wyboru tych parametrów dokonano na podstawie wyników badań przedstawionych w pracach (Alku 1997, 2002, Kane 2011) oraz w pierwszym i drugim artykule. Habilitant utworzył klasyfikator ogólny, którego składowe uzyskał się za pomocą algorytmu genetycznego. Poprawność wybranych parametrów była weryfikowana przez klasyfikator po otrzymaniu postaci szczegółowej (proces ten opisany jest w dalszej części autoreferatu). Klasyfikator ogólny został przedstawiony we wzorze 1.

W wzorze (1) zmienna y oznacza wartość estymowaną, $o_1, \dots, o_7$ są operatorami ze zbioru $\{+, -, \cdot, /, f_{k_1}\}$, f_{k_1} oznacza funkcję z parametrem 1, f_{k_2} - funkcję z parametrem 2, f_{k_3} - funkcję z parametrem 3, f_{k_4} - funkcję z parametrem 4. Funkcje te pochodzą ze zbioru czterech funkcji: $\{ax+b; a\log_{|b|x}; ax^b; ab^x\}$. Indeks k będący dodatkowym identyfikatorem funkcji przyjmuje wartości 0 lub 1. Zmienna x oznacza wartość parametru. Operator „*” zarezerwowany jest dla spłotu dlatego używany jest symbol „.”.

Za pomocą funkcji ze zbioru $\{ax+b; \log_b x; ax^b; ab^x\}$ można konstruować wiele zależności. Funkcja liniowa $ax+b$, umożliwia opisanie zależności liniowych pomiędzy zmiennymi objaśniającymi a zmienną objaśnianą. Funkcja ax^b umożliwia tworzenie dużej liczby funkcji różnych rodzajów, np. wielomianów, funkcji wymiernych, itp. Funkcja logarytmiczna $\log_b x$ oraz potęgowa ab^x umożliwia realizację podstawowych zależności opisywanych prawem Webera-Fechnera (Fechner 1966). Tak zdefiniowany klasyfikator pozwala na znalezienie zależności wyrażonej w postaci wzoru.

Chromosom 0	f_{01}	f_{02}	f_{03}	f_{04}								
Chromosom 1	$a_{f_{01}}$	$h_{f_{01}}$	o_1	$a_{f_{02}}$	$b_{f_{02}}$	o_2	$a_{f_{03}}$	$b_{f_{03}}$	o_3	$a_{f_{04}}$	$b_{f_{04}}$	o_4
Chromosom 2	f_{11}	f_{12}	f_{13}	f_{14}								
Chromosom 3	$a_{f_{11}}$	$h_{f_{11}}$	o_5	$a_{f_{12}}$	$b_{f_{12}}$	o_6	$a_{f_{13}}$	$b_{f_{13}}$	o_7	$a_{f_{14}}$	$b_{f_{14}}$	

Do oceny jakości działania klasyfikatora służy tzw. funkcja oceny. Zdefiniowano ją jako średnią arytmetyczną modułu różnic pomiędzy diagnozą wyliczoną przez klasyfikator, a diagnozą postawioną przez

lekarzy (Michalewicz 1996, Muni 2006). Poszukiwanie klasyfikatora zostało zrealizowane na populacji o stałej liczbie osobników. Wybór osobników do krzyżowania został przeprowadzony za pomocą tzw. turnieju. W wyniku przeprowadzonych eksperymentów uzyskano dziesięć najlepszych klasyfikatorów. Najlepszą zgodność uzyskano dla klasyfikatora nr 1, dla którego wartość funkcji oceny wynosi około 0,85. Finalną postać klasyfikatora przedstawiono we wzorze 2.

$$y = \frac{1}{f_{13}(NAQ)} \cdot \left(f_{03}(NAQ)(f_{01}(CQH) - f_{02}(PS)) + f_{11}(CQH) + f_{12}(PS) \right) + 0.16$$

$$\begin{aligned} f_{01}(CQH) &= -1.5CQH - 3.65625 \\ f_{02}(PS) &= -4.875 \cdot 1.21875^{PS} \\ f_{03}(NAQ) &= -0.65625 \log_{(0.84375)}(NAQ) \\ f_{11}(CQH) &= 5.4375CQH - 5.4375 \\ f_{12}(PS) &= 3.46875PS + 3.28125 \\ f_{13}(NAQ) &= -1.5 \log_{(0.375)}(NAQ) \end{aligned}$$

$$D_1 = \frac{-1}{1.5 \log_{(0.375)}(NAQ)}$$

$$D_2 = -0.65625 \log_{(0.84375)}(NAQ)$$

$$y = D_1 \left((D_2) \cdot (1.5CQH - 3.65625 - 4.875 \cdot 1.21875^{PS}) \right) + 0.16 \quad (2)$$

Pierwsza ocena poprawności działania klasyfikatora została zrealizowana na podstawie porównania wyliczonych diagnoz w postaci zaokrąglonych wartości do najbliższej całkowitej z diagnozami lekarskimi.

Diagnozy lekarskie zostały zdefiniowane następująco: 1 – pełne zwarcie lub niewielki brak zwarcia, 2 – średni brak zwarcia, 3 – duży brak zwarcia fałdów głosowych, 4 – brak możliwości postawienia diagnozy w zakresie braku zwarcia fałdów głosowych.

Gdy wyliczona diagnoza po zaokrągleniu jest identyczna z diagnozą lekarską, uznajemy, że klasyfikator dokonał poprawnej oceny. W przypadku diagnozy o wartości 4, tylko diagnoza wystawiona przez klasyfikator jest prawidłowa, ponieważ pacjent posiada guzki, ale lekarz nie mógł jednoznacznie oszacować stopnia braku zwarcia fałdów na podstawie obrazu wideofibroskopowego.

Należy wspomnieć o pewnych wadach tego badania. Wideolaryngoskop musi być wprowadzony do gardła pacjenta. W przypadku wielu dzieci, przeprowadzenie tego rodzaju diagnozy jest utrudnione lub niemożliwe z uwagi na silny odruch gardłowy. W takiej sytuacji wykorzystuje wideofibroskop – wprowadzany przez nos – jednakże nie pozwala on na tak dokładną obserwację pracy fałdów, jak laryngoscopia sztywna. Kolejna wada związana jest ze zbyt niską częstotliwością próbkowania obrazu w stosunku do częstotliwości pracy fałdów głosowych. W podstawowej diagnostyce medycznej używa się kamer standardu PAL lub SECAM, co determinuje liczbę pełnych klatek na sekundę. Wartości standardowe wynoszą odpowiednio 50 lub 60 klatek na sekundę z przeplotem lub połowę mniej bez przeplotu obrazu. Średnie wartości częstotliwości podstawowej drgań fałdów głosowych u dziewczynek w wieku 6-10 lat wynoszą 281 Hz, a u chłopców 262 Hz (Sorenson 1989). W konsekwencji różnica częstotliwości próbkowania obrazu oraz drgań fałdów głosowych powoduje, że podczas diagnozowania guzków głosowych przez foniatrę za pomocą wideolaryngoskopu pojawia się aliasing w obrazie. Utrudnia to postawienie diagnozy przez lekarza i obserwację stopnia niedomykalności fałdów głosowych, współistniejącego przy guzkach głosowych. Istnieją dokładniejsze metody obrazowania takie jak stroboscopia czy wideokymografia. Stroboscopia sprawdza się dobrze w przypadku powtarzalności drgań fałdów głosowych a zatem wśród osób o zdrowym głosie. Wideokymografia jest najlepszym rozwiązaniem, ponieważ obraz jest zapisywany z częstotliwością ok. 8000 klatek na sekundę. Takie urządzenie jest bardzo drogie dlatego w wielu szpitalach diagnoza jest stawiana na podstawie obrazu wideolaryngoskopowego (Švec 2000).

Występowanie guzków głosowych wiąże się z niedomykalnością fałdów głosowych. Jeśli diagnoza wyliczona przez klasyfikator różniła się od diagnozy lekarskiej, weryfikowano ją z dokładnym opisem lekarskim pochodzącym z karty lekarskiej pacjenta. Jeśli opis lekarski wskazywał na niepewność odnośnie stopnia zwarcia fałdów głosowych oraz był zgodny z diagnozą wyliczoną przez klasyfikator, wtedy działanie klasyfikatora uznane zostało za poprawne. W przeciwnym przypadku klasyfikator mylił się. Wszystkie uzyskane klasyfikatory, charakteryzowały się redukcją parametru RDS, gdyż wyrażenia zawierające ten parametr albo były równe zero, albo posiadały niewielką wartość (z zakresu od 0,03 do 0,16).

W celu sprawdzenia poprawności działania algorytmu genetycznego, przeprowadzono test polegający na utworzeniu klasyfikatorów dla różnych grup zawierających cztery parametry (wymienione w opisie pierwszej pracy habilitanta). Grupy te nie zawierały parametrów (CQ H, PS, NAQ, RDS). Takich zestawów utworzono 15

(kombinacja bez powtórzeń). Otrzymane rezultaty okazały się znacząco gorsze. Klasyfikatory przypisywały jednakową diagnozę dla każdego pacjenta, albo klasyfikator diagnozował osoby chore jako zdrowe. Minimalna wartość funkcji oceny dla tak otrzymanych klasyfikatorów była zbyt wysoka, ponieważ wynosiła od 1,2 do 2,5.

W celu zweryfikowania opracowanej metody, habilitant porównał wyniki uzyskane przez klasyfikator z wynikami uzyskanymi z klasyfikatorem głębokich sieci neuronowych (DNN).

Do wytrenowania DNN habilitant użył toolkitu <https://deeppcognition.ai>. Ilość epok ustawiono na 600, jako funkcję straty wybrano Poisson, współczynnik nauki = 0,01, epsilon $1e-08$, wybrano optymalizator Adagrad.

W pierwszym eksperymencie habilitant wytrenował DNN na podstawie danych dla przypadków, dla których lekarz potrafił postawić diagnozę odnośnie zwarcia fałdów głosowych. W kolejnym eksperymencie habilitant wytrenował sieć z uwzględnieniem danych, kiedy lekarz nie potrafił postawić tej diagnozy (diagnoza oznaczona „4”). Diagnoza ta w przypadku pacjentów z guzkami głosowymi oznacza, że istnieje pewien stopień niedomykalności fałdów głosowych, ale nie zawsze jest możliwy do oszacowania. Powodem tego może być brak współpracy z pacjentem, czy też niewystarczająca ilość klatek na sekundę w obrazie wideofiberoskopowym.

Dokładność wytrenowanego modelu bez „4” okazała się wysoka 81,81 % ale niższa niż prezentowanego w tej pracy klasyfikatora 92 % . W kolejnym eksperymencie zawierająca diagnozę z „4” uzyskano wyższą wartość parametru Accuracy 86,67 %.

Dane zawierające diagnozę „4” są potrzebne do treningu, ponieważ odzwierciedlają rzeczywistą sytuację w praktyce lekarskiej. Bez tych danych nie udało się wytrenować klasyfikatora ewolucyjnego, a wartość parametru Accuracy w DNN była niższa o 4,86 punktów procentowych.

Po dokładnej analizie danych okazuje się, że klasyfikator głębokich sieci neuronowych nie zawsze potrafi podjąć poprawną diagnozę w przypadku „4”. To znaczy, podejmuje diagnozę taką jak lekarz, czyli nie wnosząc istotnej informacji, o zwarcie lub braku zwarcia fałdów głosowych. Dotyczy to 26,31 % pacjentów. Jest to jeden z powodów dla którego wartość parametru Accuracy podczas walidacji spadła do 50 % dla modelu wytrenowanego na danych bez „4” ale testowanych na wszystkich danych łącznie z „4”. Okazało się, że DNN ma znaczącą tendencję do klasyfikowania pacjenta chorego jako zdrowego. Spośród 26 pacjentów z diagnozą „4” 11 pacjentów zostało zakwalifikowanych jako zdrowych co stanowi 11,70 % całej populacji. W klasyfikatorze genetycznym nie było takich pacjentów.

DNN stanowią alternatywne rozwiązanie do zastosowanego algorytmu ewolucyjnego. Uzyskane wyniki z DNN wskazują również prawidłowy wybór cech i ich parametryzację w guzkach głosowych. Jednakże w porównaniu do klasyfikatora genetycznego habilitant uzyskał wyższą wartość rozpoznawania przy użyciu diagnozy „4”.

Temat automatycznej klasyfikacji jakości głosu na podstawie cech akustycznych podejmowany był w pracach (Muhammad 2016, Neto 2007, Akbari 2014, Fezari 2014, Souissi 2016, Heris 2009). Dotychczas powstała tylko jedna praca dotycząca pośrednio guzków głosowych i zastosowania klasyfikatora opartego na algorytmie genetycznym (Heris 2009). Żadna z wymienionych prac nie pozwala na postawienie szczegółowej diagnozy patologii głosu.

Zaprezentowany klasyfikator genetyczny pozwala na postawienie diagnozy dotyczącej występowania guzków głosowych wraz ze szczegółową diagnozą zawierającą informacje o stopniu braku zwarcia fałdów głosowych. Klasyfikator używa trzech parametrów: CQ H, PS oraz NAQ.

W czwartej pracy habilitant sprawdził czy opracowane metody analizy i klasyfikacji sygnału mowy mogą być przydatne do oszacowania jakości głosu osób zdrowych. Habilitant sprawdził czy jakość głosu chórzystów różni się od jakości głosu grupy kontrolnej przy użyciu analizy percepcyjnej i akustycznej. Następnie habilitant przygotował prototyp aplikacji webowej (WEBAA) pozwalającej na monitorowanie jego jakości za pomocą smartfona przy użyciu parametrów akustycznych pochodzących z repozytorium COVAREP (A Collaborative Voice Analysis Repository For Speech Technologies) (Degottex 2014). Aplikacja znajduje się pod adresem <https://webaa.thebb.pl>.

W kolejnym kroku habilitant zrealizował bazę nagrań chórzystów śpiewających w chórze w Polsko-Japońskiej Akademii Technik Komputerowych. Nagrania te zostały sparametryzowane i porównane z nagraniami osób niepracujących głosem na co dzień.

Ocena percepcyjna została wykonana przez dwóch ekspertów, przy użyciu skali RBH. Eksperti poddani byli badaniu audiometrycznemu, które nie wykazało problemów ze słuchem. Dla klas R, B i H różnice w anotacjach ekspertów okazały się statystycznie nieistotne. Analiza statystyczna nie wykazała różnic pomiędzy oceną perceptualną głosu chórzystów a grupą kontrolną.

Następnie habilitant sprawdził czy różnice pomiędzy częstotliwością podstawową u chórzystów i chórzystek a odpowiednio grupą kontrolną są statystycznie nieistotne. Potwierdzenie tego testu pozwoliło na porównanie wyników analizy akustycznej. Wskazały one, że możliwe jest odróżnienie głosu chórzystek od głosu osób z grupy kontrolnej za pomocą parametrów QOQ (p-value 0,0485) i HRF (p-value 0,0024).

W grupie chórzystów i odpowiednio grupie kontrolnej parametrami pozwalającymi na rozróżnienie głosu są PS (p value < 0,0001), NAQ (p value < 0,0001), H1H2 (p value < 0,0001), PSP (p value < 0,0001), QOQ (p value < 0,0001). Wśród wymienionych parametrów najbardziej skutecznymi okazały się PS i NAQ (Kane 2011, Alku 2002).

Przy użyciu skali percepcyjnej nie jest możliwe rozróżnienie głosu chociaż można wskazać na jego zmiany np. u trzech chórzystów (panowie) zaobserwowano szorstkość głosu (Roughness), u jednego z chórzystów chrypkę (Hoarseness), u jednej chórzystki szorstkość głosu (Roughness) i u dwóch chórzystek chrypkę (Hoarseness). Wyniki analizy percepcyjnej nie dają tak dokładnych informacji jak analiza akustyczna.

W grupie chórzystów (panowie) parametr Peak Slope pozwolił na obserwację przesunięcia jakości głosu w kierunku fonacji wysiłkowej w porównaniu z grupą kontrolną (PS = -0,39 vs -0,33). Wyniki testów statystycznych potwierdziły tę hipotezę (p-value < 0,0001). Może to świadczyć o niewłaściwie ustawionym głosie podczas śpiewu i w konsekwencji prowadzić do zmiany jego jakości. Chórzyści są grupą osób znacznie bardziej narażoną na choroby głosu niż osoby niepracujące nim na co dzień. Intensywne jego używanie powoduje przeciążenia mechanizmu głosowego. Nadmierne napięcie mięśni aparatu głosowego prowadzi do dyskomfortu lub bólu podczas mówienia. Obserwacja przesunięcia sposobu fonacji może pomóc chórzystom w korekcie nieprawidłowych nawyków podczas śpiewania. Opracowany prototyp aplikacji pozwala na dokonanie pomiaru jakości głosu za pomocą smartfona i może być przydatny na co dzień dla osób pracujących głosem.

Wyniki drugiej i trzeciej pracy habilitanta są zgodne z rezultatami uzyskanymi w artykule piątym, szóstym i siódmym. Prace te dotyczą rzadkiej choroby genetycznej Pompego oraz syndromu Morquio (MPS IVA). Prace są oryginalne i zostały opublikowane w wysoko ocenianych czasopismach (Szkłanny 2016, Szkłanny 2018a, Szkłanny 2018b). Realizacja nagrań do wymienionych artykułów trwała trzy lata.

Późna choroba Pompego (LOPD) jest rzadką chorobą genetyczną, chorobą metaboliczną, powodowaną przez niedobór enzymu alfa-glukozydazy. Efektem tego jest gromadzenie się glikogenu w organizmie między innymi w sercu i w płucach, co ma wpływ na pracę aparatu głosowego. W piątej pracy habilitant zbadał jakość głosu u osób dotkniętych tą chorobą a w szóstej porównał jak zmieniła się jakość głosu po trzech latach od pierwszego badania (praca piąta).

Wyniki zaprezentowane w piątym artykule jednoznacznie wskazały, że u pacjentów występują zmiany w głosie. Dotyczą one między innymi niedomykalności fałdów głosowych (EGG CQ H) oraz występowanie fonacji wysiłkowej (NAQ). Wyniki analiz są zgodne z badaniem wideolaryngoskopowym. Analiza sygnału EGG (CQ H) wskazała, że niedomykalność jest widoczna również w krótszych fragmentach fonacji, odpowiednio dwu- i jednosekundowych. Wideolaryngoskopia nie dostarczyła tak precyzyjnej informacji. Ta obserwacja była inspiracją do badań habilitanta nad diagnozą szczegółową (trzecia praca).

W artykule szóstym (Szkłanny 2018a), habilitant udowodnił, że po trzech latach od pierwszych nagrań, pomimo leczenia enzymatycznego, w chorobie Pompego, następuje progresja choroby w zakresie pracy aparatu głosowego. Badania zostały przeprowadzone na tej samej grupie pacjentów. Odnotowano statystycznie znaczące, zwiększenie stopnia niedomykalności fałdów głosowych (CQ H p-value = 0.0399) oraz przesunięcie fonacji w kierunku głosu wysiłkowego (Peak Slope p-value = 0.0417). U pacjentów, którzy rozpoczęli leczenie enzymatyczne w okresie przed objawowym odnotowano stabilizację jakości głosu. Parametry te okazały się czułe na zmiany jakości głosu (już po trzech latach od pierwszego nagrania).

Wyniki eksperymentów są zgodne z oceną percepcyjną jakości głosu u tych pacjentów, nie mniej jednak ocena percepcyjna nie wniosła tak szczegółowych informacji jak analiza akustyczna i EGG.

W siódmej pracy habilitant dokonał analizy jakości głosu w rzadkiej chorobie genetycznej MPS IV A. W chorobie tej stwierdzono zmiany w budowie toru głosowego oraz funkcjonowaniu krtani (Walker 2003, Nakarat 2014, Harmatz 2015). Habilitant za pomocą analizy sygnału akustycznego, elektrogłotograficznego wskazał na występowanie znaczących zmian w głosie u pacjentów z chorobą MPS IV w porównaniu do grupy kontrolnej. U pacjentów tych zaobserwowano głos wysiłkowy, którego konsekwencja prowadzi do tworzenia guzków głosowych, niedomykalności fałdów głosowych, dysfonii i chrypki.

Zastosowane metody analizy sygnału akustycznego i EGG pozwoliły na oszacowanie jakości głosu. Bezinwazyjna diagnoza u tych osób jest szczególnie ważna, ponieważ przeprowadzenie wideolaryngoskopii jest bardzo utrudnione lub niemożliwe z uwagi na budowę ciała (Yeung 2008, Szkłanny 2018b). Opublikowana praca jest oryginalna. Jest to pierwszy taki artykuł wnoszący precyzyjne informacje o jakości głosu u tych pacjentów.

5.3 Podsumowanie

W ramach osiągnięcia naukowego przedstawione zostały metody wnoszące oryginalny wkład w rozwój informatyki w zakresie analizy i metod klasyfikacji sygnału akustycznych:

1. Ważnym aspektem jest podjęta przez habilitanta diagnoza szczegółowa jakości głosu za pomocą narzędzi informatycznych. To nowatorskie rozwiązanie obejmuje oszacowanie stopnia braku zwarcia fałdów głosowych za pomocą klasyfikatora genetycznego.
2. Zastosowana parametryzacja sygnału akustycznego pozwoliła również opracować metodę, za pomocą której możliwa jest predykcja stopni skali RBH za pomocą parametrów akustycznych.
3. Analiza sygnału akustycznego pozwala na odróżnienie jakości głosu chórzystów i chórzystek od grupy osoby zdrowych. Zaprezentowana aplikacja webowa może pomóc chórzystom w korekcie nieprawidłowych nawyków podczas śpiewania.
4. Bezinwazyjność zaprezentowanych metod pozwala na ich stosowanie u osób u których utrudnione lub niemożliwe jest badanie wideolaryngoskopowe.
5. Parametry analizy akustycznej (Peak Slope, NAQ) i EGG (CQ H), użyte do analizy jakości głosu u osób z guzkami, oraz w chorobie Pompego i MPS IV A znalazły swoje zastosowanie w klasyfikatorze genetycznym.
6. Impact Factor z czterech najlepszych prac habilitanta wynosi 14,199, Impact Factor 5 letni 16,757, co świadczy o tym że habilitant koncentrował się na publikowaniu w prestiżowych czasopismach.
7. Opracowane metody automatycznej klasyfikacji mogą być zastosowane w innych chorobach głosu do określenia zmian i śledzenia ich w czasie.

5.4 Literatura

- (Akbari 2014) Akbari, A., & Arjmandi, M. K. (2014). An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features. *Biomedical Signal Processing and Control*, 10, 209-223.
- (Alku 1997) Alku, P., Strik, H., & Vilkman, E. (1997). Parabolic spectral parameter—a new method for quantification of the glottal flow. *Speech Communication*, 22(1), 67-79.
- (Alku 2002) Alku, P., Bäckström, T., & Vilkman, E. (2002). Normalized amplitude quotient for parametrization of the glottal flow. *the Journal of the Acoustical Society of America*, 112(2), 701-710.
- (ASHA 2002) Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) ASHA, 2002.
- (ASHA 2018) American Speech-Language-Hearing Association (<https://www.asha.org/slp/clinical/voice-disorders/>) 2018
- (Battle 2012) Battle, D. E. (2012). *Communication disorders in multicultural and international populations*. Elsevier Health Sciences.
- (Barry 1997) W.J. Barry, M. Putzer, Saarbrücken Voice Database, Institute of Phonetics, Univ. of Saarland, Accessed 2017 <http://www.stimmdatenbank.coli.uni-saarland.de/>
- (Chandrashekarappa 2016) Chandrashekarappa S., Sandeep S., Jain S., Guruswamy M., & Shetty S. T., (2016). Correlation of electroglottograph with vocal cord pathology. *International Journal of Otorhinolaryngology and Head and Neck Surgery*, 3(1), 46-48.
- (Childers 1991) Childers, D. G., & Lee, C. K. (1991). Vocal quality factors: Analysis, synthesis, and perception. *the Journal of the Acoustical Society of America*, 90(5), 2394-2410.
- (De Bodt 2007) De Bodt, M. S., et al. "Evolution of vocal fold nodules from childhood to adolescence." *Journal Voice* 21.2 (2007): 151-156
- (Degottex 2014) Degottex, G., Kane, J., Drugman, T., Raitio, T., & Scherer, S. (2014, May). COVAREP—A collaborative voice analysis repository for speech technologies. In *2014 IEEE international conference on acoustics, speech and signal processing (icassp)* (pp. 960-964). IEEE.
- (Dejonckere 2001) Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., Van De Heyning P., Remacle M., Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques: Guideline elaborated by the Committee

- on Phoniatrics of the European Laryngological Society (ELS). European Archives of Oto-Rhino-Laryngology. <https://doi.org/10.1007/s004050000299>
- (Dejonckere 2004) Dejonckere, P. H., Crevier-Buchman, L., Marie, J. P., Moerman, M., Remacle, M., & Woisard, V. (2003). Implementation of the European Laryngological Society (ELS) basic protocol for assessing voice treatment effect. *Revue de laryngologie-otologie-rhinologie*, 124(5), 279-283.
- (Fabre 1957) Fabre P., "Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence; premiers résultats,," *Bull. Acad. Natl. Med.*, vol. 141, no. 3-4, pp. 66-69, 1957.
- (Fant 1985) Fant, G., Liljencrants, J., & Lin, Q. G. (1985). A four-parameter model of glottal flow. *STL-QPSR*, 4(1985), 1-13.
- (Fant 1995) Fant, G. (1995). The LF-model revisited. *Transformations and frequency domain analysis. Speech Trans. Lab. Q. Rep.*, Royal Inst. of Tech. Stockholm, 2(3), 40.
- (Fechner 1966) Fechner, G. T., Howes, D. H., & Boring, E. G. (1966). *Elements of psychophysics* (Vol. 1). New York: Holt, Rinehart and Winston.
- (Fezari 2014) Fezari, M., Amara, F., & El-Emary, I. M. (2014, February). Acoustic analysis for detection of voice disorders using adaptive features and classifiers. In *Proc. 2014th Int. Conf. on Circuits, Systems and Control*, Switzerland, Feb. 22-24 2014.
- (Fourcin 1972) Fourcin, A. J., & Abberton, E. (1972). First applications of a new laryngograph. *The Volta Review*.
- (Frokjaer-Jensen 1968) Frokjaer-Jensen, B., & Thorvaldsen, P. (1968). Construction of a Fabre glottograph. *ARIPUC*, 3, 1.
- (Gramuglia 2014) Gramuglia, A. C. J., Tavares, E. L., Rodrigues, S. A., & Martins, R. H. (2014). Perceptual and acoustic parameters of vocal nodules in children. *International journal of pediatric otorhinolaryngology*, 78(2), 312-316.
- (Guo 2017) Guo, W. J., Yu, W. Y., & Wang, B. Q. (2017). Research progress of intervention strategies on voice disorders in children with vocal nodules. *Lin chuang er bi yan hou tou jing wai ke za zhi*= *Journal of clinical otorhinolaryngology, head, and neck surgery*, 31(13), 1043.
- (Hacki 1989) Hacki, T. (1989). Klassifizierung von glottiscysfunktionen mit hilfe der elektrogglottographie [classification of glottal dysfunctions on the basis of electroglottography]. *Folia phoniatrica*, 41(1), 43-48.
- (Harmatz 2015) Harmatz, P. R., Mengel, K. E., Giugliani, R., Valayannopoulos, V., Lin, S. P., Parini, R., ... & Martins, A. M. (2015). Longitudinal analysis of endurance and respiratory function from a natural history study of Morquio A syndrome. *Molecular genetics and metabolism*, 114(2), 186-194.
- (Hillenbrand 1996) Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39(2), 311-321.
- (Heris 2009) Heris, H. K., Aghazadeh, B. S., & Nikkhah-Bahrami, M. (2009). Optimal feature selection for the assessment of vocal fold disorders. *Computers in Biology and Medicine*, 39(10), 860-868.
- (Hirano 1981) Hirano, M. (1981). Clinical examination of voice. *Disorders of human communication*, 5, 1-99.
- (Howard 1990) Howard, D. M., Lindsey, G. A., & Allen, B. (1990). Toward the quantification of vocal efficiency. *Journal of Voice*, 4(3), 205-212.
- (Howard 1995) Howard, D. M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice*, 9(2), 163-172.
- (Johns 2003) Johns, M. M. (2003). Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts. *Current opinion in otolaryngology & head and neck surgery*, 11(6), 456-461.
- (Kane 2011) Kane, J., & Gobl, C. (2011). Identifying regions of non-modal phonation using features of the wavelet transform. In *Twelfth Annual Conference of the International Speech Communication Association*.
- (Kane 2013) Kane, J., & Gobl, C. (2013). Wavelet maxima dispersion for breathy to tense voice discrimination. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(6), 1170-1179.
- (Kempster 2009) Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*.
- (MEEI 1994) *Disordered Voice Database (CD-ROM)*, Version 1.03, Massachusetts Eye and Ear Infirmary, Kay Elemetrics Corporation, Boston, MA, Voice and Speech Lab., October 1994

- (Muhammad 2016) Muhammad, G., Altuwajiri, G., Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., Malki K. H., & Al-nasheri, A. (2016). Automatic voice pathology detection and classification using vocal tract area irregularity. *Biocybernetics and Biomedical Engineering*, 36(2), 309-317.
- (Nakarat 2014) Nakarat, T., Läßig, A. K., Lampe, C., & Keilmann, A. (2014). Alterations in speech and voice in patients with mucopolysaccharidoses. *Logopedics Phoniatrics Vocology*, 39(1), 30-37.
- (Neto 2007) Neto, B. G. A., Fechine, J. M., Costa, S. C., & Muppa, M. (2007, October). Feature estimation for vocal fold edema detection using short-term cepstral analysis. In 2007 IEEE 7th International Symposium on BioInformatics and BioEngineering (pp. 1158-1162). IEEE.
- (Kiliç 2004) Kiliç, M. A., Okur, E., Yildirim, I., & Güzelsoy, S. (2004). The prevalence of vocal fold nodules in school age children. *International Journal of Pediatric Otorhinolaryngology*, 68(4), 409-412.
- (Lee 2018) Lee, J. M., Roy, N., & Dietrich, M. (2018). Personality, Psychological Factors, and Behavioral Tendencies in Children With Vocal Nodules: A Systematic Review. *Journal of Voice*.
- (Michalewicz 1996) Zbigniew, M. (1996). Genetic algorithms+ data structures= evolution programs. 2 nd, extended edition, Springer-Verlag Berlin Heidelberg, (1994).
- (Mudd 2018) Mudd, P., & Noelke, C. (2018). Vocal fold nodules in children. *Current opinion in otolaryngology & head and neck surgery*, 26(6), 426-430.
- (Muni 2006) Muni, D. P., Pal, N. R., & Das, J. (2006). Genetic programming for simultaneous feature selection and classifier design.
- (Nawka 1994) Nawka T, Anders LC, Wendler J. Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. *Sprache-Stimme-Geho* r. 1994;18:130-3.
- (Nuss 2012) Nuss, R. C., Ward, J., Recko, T., Huang, L., & Woodnorth, G. H. (2012). Validation of a pediatric vocal fold nodule rating scale based on digital video images. *Annals of Otology, Rhinology & Laryngology*, 121(1), 1-6.
- (Ravindran 2013) Ravindran, S., Siew-Chin, N., & Muthusamy, H. (2013). Optimal selection of long time acoustic features using GA for the assessment of vocal fold disorders. In *Applied Mechanics and Materials* (Vol. 239, pp. 65-70). Trans Tech Publications.
- (Roy 2004) Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D., & Smith, E. M. (2004). Prevalence of voice disorders in teachers and the general population. *Journal of Speech, Language, and Hearing Research*.
- (Shah 2005) Shah, R. K., Woodnorth, G. H., Glynn, A., & Nuss, R. C. (2005). Pediatric vocal nodules: correlation with perceptual voice analysis. *International journal of pediatric otorhinolaryngology*, 69(7), 903-909.
- (Shah 2007) Shah, R. K., Feldman, H. A., & Nuss, R. C. (2007). A grading scale for pediatric vocal fold nodules. *Otolaryngology—Head and Neck Surgery*, 136(2), 193-197
- (Sorenson 1989) Sorenson, D. N. (1989). A fundamental frequency investigation of children ages 6–10 years old. *Journal of Communication Disorders*, 22(2), 115-123.
- (Souissi 2016) Souissi, N., & Cherif, A. (2016). Artificial Neural Networks and Support Vector Machine for Voice Disorders Identification. *International Journal of Advanced Computer Science and Applications*, 7(5), 339-344.
- (Szkłanny 2016) Szkłanny, K., Gubrynowicz, R., Iwanicka-Pronicka, K., & Tylki-Szymańska, A. (2016). Analysis of voice quality in patients with late-onset Pompe disease. *Orphanet journal of rare diseases*, 11(1), 99.
- (Szkłanny 2018a) Szkłanny, K., & Tylki-Szymańska, A. (2018). Follow-up analysis of voice quality in patients with late-onset Pompe disease. *Orphanet journal of rare diseases*, 13(1), 189.
- (Szkłanny 2018b) Szkłanny, K., Gubrynowicz, R., & Tylki-Szymańska, A. (2018). Voice alterations in patients with Morquio a syndrome. *Journal of applied genetics*, 59(1), 73-80.
- (Szkłanny 2019a) Szkłanny K., Gubrynowicz R., Ratyńska J., and Chojnacka-Wądołowska D., “Electroglottographic and acoustic analysis of voice in children with vocal nodules,” *Int. J. Pediatric Otorhinolaryngol.*, 2019.
- (Szkłanny 2019b) Szkłanny K, Wrzeciono P, Relation of RBH auditory-perceptual scale to acoustic and electroglottographic voice analysis in children with vocal nodules, doi: 10.1109/ACCESS.2019.2907397, pp. 1-12, Online ISSN: 2169-3536 IEEE ACCESS 2019.
- (Szkłanny 2019c) Szkłanny K, Wrzeciono P, The application of a genetic algorithm in the noninvasive assessment of vocal nodules in children, doi: 10.1109/ACCESS.2019.2908313, IEEE ACCESS 2019.
- (Švec 2000) Švec, J. G., Šram, F., & Schutte, H. K. (2000). Development and application of videokymography for high-speed examination of vocal-fold vibration. *Proceedings of LP*, 3-10.

- (Titze 1992) Titze, I. R., & Sundberg, J. (1992). Vocal intensity in speakers and singers. the Journal of the Acoustical Society of America, 91(5), 2936-2946.
- (Tezcaner 2009) C. Z. Tezcaner, S. K. Ozgursoy, I. Sati, and G. Dursun, "Changes after voice therapy in objective and subjective voice measurements of pediatric patients with vocal nodules," Eur. Arch. Oto-Rhino-Laryngology, 2009
- (Valadez 2012) Valadez, V., Ysunza, A., Ocharan-Hernandez, E., Garrido-Bustamante, N., Sanchez-Valerio, A., & Pamplona, M. C. (2012). Voice parameters and videonasolaryngoscopy in children with vocal nodules: a longitudinal study, before and after voice therapy. International journal of pediatric otorhinolaryngology, 76(9), 1361-1365.
- (Walker 2003) Walker, P. P., Rose, E., & Williams, J. G. (2003). Upper airways abnormalities and tracheal problems in Morquio's disease. Thorax, 58(5), 458-459.
- (Yeung 2008) Yeung, A. H., Cowan, M. J., Horn, B., & Rosbe, K. W. (2009). Airway management in children with mucopolysaccharidoses. Archives of Otolaryngology-Head & Neck Surgery, 135(1), 73-79.

5.5 Publikacje niewchodzące w skład osiągnięcia naukowego znajdujące się w Web of Science

Gonera J., Szklanny K., Wichrowski M., Wieczorkowska A. (2018) Mobile Application with Image Recognition for Persons with Aphasia. In: Ceci M., Japkowicz N., Liu J., Papadopoulos G., Raś Z. (eds) Foundations of Intelligent Systems. ISMIS 2018. Lecture Notes in Computer Science, vol 11177. Springer, Cham.

Wkład habilitanta: Opracowanie koncepcji aplikacji dla osób dotkniętych afazją, napisanie kilku rozdziałów artykułu. Mój udział szacuję na 20%.

Kosiński, J., Szklanny, K., Wieczorkowska, A., & Wichrowski, M. (2018, September). An Analysis of Game-Related Emotions Using EMOTIV EPOC. In 2018 Federated Conference on Computer Science and Information Systems (FedCSIS) (pp. 913-917). IEEE.

Wkład habilitanta: Opracowanie koncepcji badania emocji w grach komputerowych za pomocą EMOTIV EPOC, napisanie artykułu. Mój udział szacuję na 25%.

Szklanny, K., & Koszuta, S. (2017, September). Implementation and verification of speech database for unit selection speech synthesis. In Computer Science and Information Systems (FedCSIS), 2017 Federated Conference on (pp. 1262-1267). IEEE.

Wkład habilitanta: Opracowanie metody poprawy jakości korpusu dla języka polskiego. Przygotowanie korpusu do nagrań, oraz metody wzbogacającej korpus o rzadko występujące fonemy. Realizacja testu weryfikującego bazę. Napisanie artykułu i poprawie artykułu zgodnie z wytycznymi recenzentów.

Mój udział szacuję na 55%.

Szklanny, K., Homoncik, Ł., Wichrowski, M., & Wieczorkowska, A. (2017, September). Creating an interactive and storytelling educational physics app for mobile devices. In Computer Science and Information Systems (FedCSIS), 2017 Federated Conference on (pp. 1269-1273). IEEE.

Wkład habilitanta: Opracowanie koncepcji badań nad interaktywną grą do nauki fizyki. Realizacja testów UX dla zrealizowanej aplikacji. Napisanie kilku rozdziałów w artykule. Mój udział szacuję na 40%.

Wichrowski, M., Korżinek, D., & Szklanny, K. (2015, June). Google glass development in practice: UX design sprint workshops. In Proceedings of the Multimedia, Interaction, Design and Innovation (p. 11). ACM.

Wkład habilitanta: Realizacja testów użyteczności dla opracowanych aplikacji na Google Glass i tablet. Analiza i anotacja nagranych testów. Napisanie części artykułu. Mój udział szacuję na 20%.

Szklanny K. (2014) Multimodal Speech Synthesis for Polish Language. In: Gruca D., Czachórski T., Kozielski S. (eds) Man-Machine Interactions 3. Advances in Intelligent Systems and Computing, vol 242. Springer, Cham

Wkład habilitanta: Implementacja systemu korpusowej syntezy mowy polskiej. Oszacowanie funkcji kosztu dla języka polskiego. Realizacja awatara. Napisanie i korekta artykułu. Mój udział wynosi 100%.

Szklanny K., Wichrowski M. (2014) Prototyping M-Learning Course on the Basis of Puzzle Learning Methodology. In: Zaphiris P., Ioannou A. (eds) Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration. LCT 2014. Lecture Notes in Computer Science, vol 8524. Springer, Cham

Wkład habilitanta: Przygotowanie raportu dotyczącego technologii mobilnych w Polsce. Opracowanie kursu Puzzle Learning. Napisanie kilku rozdziałów artykułu. Mój udział szacuję na 60%.

Szklanny K., Stańczak A., Wojtków P., Cosentino S., Wieczorkowska A. (2014) SyncBox - Synchronizer and Interface for High-Speed Macro Photography. In: Kurosu M. (eds) Human-Computer Interaction. Theories, Methods, and Tools. HCI 2014. Lecture Notes in Computer Science, vol 8510. Springer, Cham

Wkład habilitanta: Opracowanie metody zastosowania Arduino w fotografii ultraszybkiej. Napisanie artykułu. Mój udział szacuję na 30%.

6 Omówienie pozostałych osiągnięć naukowo-badawczych

6.1 Badania niezwiązane z głównym celem badawczym

W ramach badań naukowych prowadzonych po uzyskaniu stopnia doktora (2010-2019) niewyszczególnionych w ramach osiągnięcia w rozdziale 4, habilitant uzyskał znaczące wyniki prowadzone w ramach trzech zagadnień:

1. Realizacji systemu multimodalnej syntezy mowy (Szklanny 2014a)
2. Przygotowania baz akustycznych przeznaczonych dla systemów syntezy i rozpoznawania mowy (Szklanny 2014a)
3. Realizacja kursów m-learningowych przy użyciu metody Puzzle Learning oraz tworzenie aplikacji multimedialnych wspomagających naukę (Szklanny 2014b, Szklanny 2017a)

6.1.1 Realizacja systemu multimodalnej syntezy mowy

W pracy doktorskiej habilitant zrealizował pełny, funkcjonalny system korpusowej syntezy mowy dla języka polskiego, pozwalający na wygenerowanie naturalnie brzmiącej mowy. Był to pierwszy niekomercyjny system syntezy mowy polskiej. Najistotniejszym elementem w takim systemie jest funkcja kosztu, ponieważ odpowiada między innymi za jakość syntezy. Składa się dwóch części – kosztu doboru jednostki oraz kosztu konkatenacji. Koszt doboru jednostki wyszukuje te segmenty, które będą najbardziej spełniały cechy lingwistyczne zdania docelowego. Koszt konkatenacji bazuje na parametrach numerycznych takich jak: energia sygnału, częstotliwość podstawowa (F0). Jej celem, jest minimalizowanie kosztu nieciągłości spektralnej na granicach łączonych ze sobą fragmentów. Konstrukcja funkcji kosztu jest zadaniem trudnym, ponieważ znalezienie wag parametrów oraz zależności między nimi ma umożliwić wyszukanie takich jednostek, których łączenie ze sobą, pozwoli na uzyskanie naturalnej mowy. Istotnym problemem badawczym była między innymi optymalizacja tej funkcji. Autor zastosował metodę heurystyczną w postaci algorytmu ewolucyjnego. Do oszacowania parametrów wybrano strategię elitarną, pozwalającą wyszukiwać najlepszego potomka wśród dzieci (μ) jak i rodziców (λ) ($\mu+\lambda$). Znak $/+/-$ w strategii oznacza, że kandydat będzie wybierany zarówno z rodziców jak i z potomków. Tak utworzona populacja jest łączona z populacją bazową (Michalewicz 2004). Przeprowadzona optymalizacja wskazała, że najważniejszymi elementami w funkcji kosztu dla projektowanego głosu są:

- koszt pozycji w sylabie,
- koszt nieciągłości melodii,
- koszt prawego kontekstu,
- koszt nieciągłości energii,
- koszt akcentu,
- koszt złego doboru F0 dla kosztu doboru jednostki

Wyniki testów perceptualnych (MOS - Mean Opinion Score) potwierdziły skuteczność zoptymalizowanej funkcji kosztu. MOS jest subiektywną metodą stosowaną do testowania jakości dźwięku (mowy) np. w telefonii czy też w systemach syntezy mowy (ITU 1996, Viswanathan 2005). Mierzone były dwie cechy sygnału mowy - zrozumiałość oraz naturalność. Ocena mieści się w skali od 1 do 5. Z porównania trzech funkcji kosztu wynika, że proces estymacji parametrów przyniósł oczekiwany efekt. Ekspertcy ocenili domyślną funkcję kosztu na 2,185, najgorsza funkcja kosztu uzyskała podczas estymacji parametrów 1,97, dla najlepszej funkcji kosztu wartość średniej wzrosła do 2,7111. Porównując system do komercyjnego systemu korpusowej syntezy mowy IVONA stworzonego przez firmę IVOSOFTWARE (obecnie Amazon) jest to wynik bardzo dobry. System IVONA dla języka angielskiego otrzymał ocenę 3,6 (dla języka polskiego nie istniało bezpośrednie porównanie). Używając

głosu profesjonalnego lektora można otrzymać jakość zbliżoną do profesjonalnych systemów syntezy mowy. Podsumowując habilitant opracował narzędzie heurystyczne oraz metodę pozwalającą na optymalizację parametrów funkcji kosztu przy zastosowaniu dowolnej bazy akustycznej. Metoda ta z dużym prawdopodobieństwem może być również wdrożona do estymacji dowolnej funkcji kosztu dla innych języków.

Kontynuacja prac nad syntezą mowy po doktoracie, doprowadziła do implementacji pełnego multimodalnego systemu syntezy mowy polskiej. System ten posiada dwa zintegrowane modele awatarów (Rysunek 5). Zaimplementowano w nich reakcje rozglądania się po ekranie, ruchy mimiczne, podążanie wzrokiem i kursorem oraz możliwością ustawienia wyrazu twarzy. Poprawiono również jakość korpusowej syntezy mowy, eliminując błędy na złączeniach segmentów akustycznych. Ten pierwszy niekomercyjny system korpusowej syntezy mowy w Polsce do dziś jest używany i stosowany, między innymi w norweskim projekcie Nomad "Łamigłówki dla Nomada – metoda nauczania przez całe życie na miarę XXI wieku". Syntezator był używany na wydarzeniach takich jak dni otwarte uczelni PJATK, Festiwal Przemiany organizowane przez Wolskie Centrum Kultury. System używany jest również przez osoby niewidome. System dostępny jest pod adresem: <http://syntezamowy.pjwstk.edu.pl/korpus.html>.

Wyniki prac nad syntezą opublikowano w artykułach (Szklanny 2014a, Marasek 2009). Bazę akustyczną, czyli nagrany korpus zawierający 2150 zdań, autor opublikował na stronie międzynarodowego projektu European Language Resources Association (ELRA). Warty podkreślenia jest fakt, że taka publikacja jest możliwa dopiero po spełnieniu rygorystycznych kryteriów jej jakości. Autor opublikował tam również bazę akustyczną służącą do realizacji konkatenacyjnej syntezy mowy. Bazy te dostępne są pod następującymi linkami:

http://catalog.elra.info/product_info.php?cPath=37_39&products_id=1164

http://catalog.elra.info/product_info.php?cPath=37_39&products_id=1168



Rysunek 5. Trójwymiarowy model awatara zaimplementowany w systemie multimodalnej syntezy mowy polskiej.

6.1.2 Przygotowanie nowej bazy akustycznej

Uzyskane wyniki poprzednich eksperymentów badań były na tyle dobre, że habilitant przygotował nową wersję bazy akustycznej, wykorzystując udowodnione tezy z pracy doktorskiej, m.in. wybór mówcy i jakość bazy akustycznej mają znaczący wpływ na jakość syntetycznej mowy. Lektor do nowej bazy został wybrany na podstawie nagranych próbek głosu zebranych od 30 kandydatów. Każdy z kandydatów miał profesjonalne doświadczenie w pracy głosem. Celem było znalezienie lektora o energicznym i stabilnym głosie. Oceny głosu dokonało ośmiu ekspertów z dziedziny przetwarzania dźwięku oraz mowy. Wybrany został głos damski. Habilitant zapewnił profesjonalną jakość realizowanym nagraniom (Szklanny 2017b). Jakość przygotowanej bazy zweryfikowano w prototypowym syntezatorze. Program pozwala na wstępne przetestowanie jakości korpusu. Umożliwia syntezę korpusową z wykorzystaniem trzech różnych algorytmów: „Random“, „Forward“ oraz „Viterbi“ (algorytm Viterbiego) (Viterbi 1967). Algorytmy te odpowiadają za sposób wyboru sekwencji segmentów z bazy akustycznej. Głównym kryterium brany pod uwagę przy wyborze segmentów jest ich bezpośrednie sąsiedztwo w bazie akustycznej, co zmniejsza prawdopodobieństwo wystąpienia artefaktów w postaci nieciągłości energii i generuje naturalniej brzmiącą mowę. Dodatkowo uwzględnia się podobieństwo F_0 przy granicach konkatenowanych jednostek. Wyniki testu MOS wskazały, że głos lektora otrzymał ocenę 4,3, w skali od 1 do 5. Ocena głosu lektora wyraża opinię respondentów na temat możliwości przyszłego syntezatora. Jest to maksymalna ocena jaką mógłby uzyskać najlepszy syntezator przy użyciu tego głosu. Resyntezę zdań czyli syntezowanie zdań nagranych w bazie, eksperci ocenili na poziomie 3,41 co jest rezultatem dobrym. Syntezę oceniono na 2,07. Ocena ta dotyczy bazy posegmentowanej metodami automatycznymi, bez

scen

zastosowania korekty ręcznej. Mimo tego baza otrzymała w teście MOS podobne wyniki, jak baza w której korekta ręczna została zastosowana (2.07 vs 2.18) oraz ocenę wyższą niż baza BOSS¹ dla bogatych fonetycznie zdań (2,07 vs 1.70). Zatem poprzez uzupełnienie i poszerzenie bazy, uzyskano wyniki poprawiające jakość bazy akustycznej. W praktyce oznacza to, że podczas implementacji tej bazy w systemie syntezy mowy, jakość generowanej mowy będzie znacząco lepsza, co na tym etapie zostało potwierdzone przez ekspertów. Wyniki prac zostały opublikowane w artykule (Szklanny 2017b).

6.1.3 Realizacja kursów m-learningowych przy użyciu metody Puzzle Learning oraz tworzenie aplikacji multimedialnych wspomagających naukę

W pracy (Szklanny 2014b) opublikowano metodologię tworzenia kursów m-learningowych podczas realizacji międzynarodowego projektu "Łamigłówki dla Nomada – metoda nauczania przez całe życie na miarę XXI wieku". Jego celem było poprawienie sposobu nauczania, poprzez wypełnienie luk w kompetencjach, dla osób poszukujących pracy. Wykorzystano innowacyjną metodę nauczania opartą na łamigłówkach (Puzzle Learning). Metoda ta polega na uczeniu się poprzez rozwiązywanie różnego rodzaju problemów za pomocą zagadek (Michalewicz 2008). W artykule (Szklanny 2014b) przedstawiono sposób realizacji pierwszego prototypowego kursu. Wynikiem pracy habilitanta są kursy z zakresu multimediiów, dostępne na platformie m-learningowej <https://kursy.nomad.pja.edu.pl/index.do>.

Tematyka tworzenia kursów m-learningowych była kontynuowana w pracy (Szklanny 2017a). Jej wynikiem jest aplikacja pozwalająca na wsparcie procesu nauki fizyki. Jest przeznaczona dla gimnazjalistów i licealistów. Większość istniejących aplikacji na urządzenia mobilne koncentruje się na prezentacji wiedzy pozbawionej animacji, interakcji i multimediiów. Zrealizowana aplikacja ma charakter edutainmentu i została przedstawiona jako gra przygodowa. Przeprowadzone wyniki testów użyteczności gry Apollo wskazały, że metodologia zastosowana do stworzenia tej aplikacji została właściwie zaprojektowana. Opinie większości testerów potwierdzają, że aplikacja może wspierać naukę fizyki. Aplikacja została oceniona jako łatwa w użyciu, nawigacja w niej jest prosta, elementy interfejsu zostały przedstawione czytelnie. Jakość dźwięku została oceniona wysoko.

Literatura

- (ITU 1996) Telecom, I. T. U. (1996). Standardization Sector of ITU, "Video coding for low bitrate communication," Draft ITU-T Recommendation H, 263.
- (Michalewicz 2004) Michalewicz Z. (2004) Algorytmy genetyczne + struktury danych = programy ewolucyjne wyd. Wydawnictwa Naukowo-Techniczne ISBN: 83-204-2881-5
- (Michalewicz 2008) Michalewicz, Z., & Michalewicz, M. (2008). Puzzle-based learning. Hybrid Publishers.
- (Marasek 2009) Marasek, K., Brocki, Ł., Korżinek, D., Szklanny, K., & Gubrynowicz, R. (2009). User-centered design for a voice portal. In *Aspects of Natural Language Processing* (pp. 273-293). Springer, Berlin, Heidelberg
- (Szklanny 2014a) Szklanny K. (2014) Multimodal Speech Synthesis for Polish Language. In: Gruca D., Czachórski T., Kozielski S. (eds) Man-Machine Interactions 3. Advances in Intelligent Systems and Computing, vol 242. Springer, Cham
- (Szklanny 2014b) Szklanny K., Wichrowski M. (2014) Prototyping M-Learning Course on the Basis of Puzzle Learning Methodology. In: Zaphiris P., Ioannou A. (eds) Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration. LCT 2014. Lecture Notes in Computer Science, vol 8524. Springer, Cham
- (Szklanny 2017a) Szklanny, K., Homoncik, Ł., Wichrowski, M., & Wieczorkowska, A. (2017, September). Creating an interactive and storytelling educational physics app for mobile devices. In Computer Science and Information Systems (FedCSIS), 2017 Federated Conference on (pp. 1269-1273). IEEE.

¹ Syntezator mowy BOSS jest pierwszym systemem syntezy mowy korpusowej stworzonym w Polsce w ramach współpracy pomiędzy Uniwersytetem Adama Mickiewicza oraz IKP (Institut für Kommunikationsforschung und Phonetik) w Bonn.



- Szklanny (2017b) Szklanny, K., & Koszuta, S. (2017, September). Implementation and verification of speech database for unit selection speech synthesis. In Computer Science and Information Systems (FedCSIS), 2017 Federated Conference on (pp. 1262-1267). IEEE.
- (Viswanathan 2005) Viswanathan M., Viswanathan M. (2005) Measuring speech quality for text-to-speech systems: development and assessment of a modified mean opinion score (MOS) scale, Computer Speech and Language 19, 55–83.
- (Viterbi 1967) Viterbi A. J. (1967) Error bounds for convolutional codes and an asymptotically optimal decoding algorithm, IEEE Transactions on Information Processing, 13:260-269.

