

Ocena parametrów analizy akustycznej w detekcji patologii mowy

Streszczenie. Diagnostyka stanu dróg głosowych wymaga stworzenia wektora, który składa się z różnych parametrów akustycznych, co może pomóc w szybkiej oraz automatycznej detekcji patologii głosu. W niniejszym artykule przedstawiono wektor cech złożony z 31 parametrów. Parametry mowy zostały wyodrębnione w dziedzinie czasu, częstotliwości oraz cepstralnej. Wybór parametrów niezbędnych w ocenie patologii głosu został potwierdzony w analizie głównych składowych, jądrowej analizie głównych składowych (kernel PCA) oraz liniowej analizie dyskryminacyjnej (LDA).

Abstract. The diagnosis of the current state of the vocal tract requires the creation of a feature vector that consists of various acoustic parameters, which can help in rapid and automatic detection of voice pathologies. Vector consisting of 31 parameters was done in this project. Speech parameters were extracted in the time, frequency and cepstral domain. Essential parameters were selected and analysed using principal component analysis, kernel principal component analysis and linear discriminant analysis. (**Parameters evaluation of acoustic analysis in speech pathology detection**)

Słowa kluczowe: analiza akustyczna, patologiczny sygnał mowy, analiza głównych składowych, liniowa analiza dyskryminacyjna.

Keywords: acoustic analysis, pathological speech signal, principal component analysis, linear discriminant analysis.

doi:10.12915/pe.2014.05.29

Wstęp

Obecny rozwój technologii, w tym cyfrowej rejestracji oraz metod przetwarzania głosu, spowodował postęp w automatycznej detekcji większości patologii i chorób aparatu głosowego człowieka. Automatyczna analiza akustyczna jest coraz bardziej ceniona jako nieinwazyjne i obiektywne narzędzie diagnostyczne. Do przeprowadzenia analizy akustycznej można wykorzystać: metodę klasyczną, która oparta jest na subiektywnej ocenie głosu oraz metodę nowoczesną, wykorzystującą analizę sygnałów akustycznych, spektrografię, sonografię czy obrazy przedstawiające trakt głosowy. Obecnie, nadal poszukiwane są metody pomiarów, modele, algorytmy i nowe podejścia do klasyfikacji, które pomogą w skutecznym rozróżnieniu osoby chorej od zdrowej [1,2]. Cyfrowe przetwarzanie sygnału jest procesem, który umożliwia pomiar coraz większej liczby mierzalnych cech ludzkiego głosu, a także pozwala na dokładną, obiektywną ocenę dyskretnych, niekoniecznie widocznych w metodzie odsłuchowej, patologii głosu oraz mowy.

Rozpoznanie stanu traktu głosowego wymaga stworzenia modelu matematycznego składającego się z różnych parametrów opisujących sygnał akustyczny. W artykule przedstawiono analizę akustyczną przeprowadzoną dla 31 różnych parametrów wyodrębnionych z zapisów sygnałów ludzkiej mowy. Analizę przeprowadzono w dziedzinie czasu, częstotliwości oraz cepstralnej. Przy wykorzystaniu metod wielowymiarowej analizy statystycznej, takich jak liniowa Analiza Głównych Składowych PCA (ang. Principal Component Analysis), jądrowa Analiza Głównych Składowych kPCA (ang. Kernel Principal Component Analysis) oraz liniowa analiza dyskryminacyjna LDA (ang. Linear Discriminant Analysis), przeprowadzono redukcję danych oraz zbadano możliwości postawienia diagnozy mowy patologicznej u kobiet oraz u mężczyzn.

Materiał

W pracy wykorzystano nagrania pochodzące z *Saarbruecken Voice Database* (SVD), udostępnione przez Instytut Fonetyki w Saarland w Niemczech [2,3]. SVD jest kolekcją nagrań samogłosek /a/, /i/, /u/ wypowiedzianych w wysokiej, niskiej, normalnej oraz modulowanej tonacji. Nagrania zostały zarejestrowane od ponad 2000 osób.

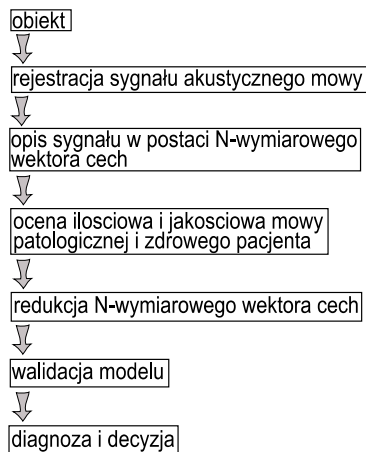
Czas trwania nagrań samogłosek wykorzystanych w analizie wynosił od 1 do 4 s. Wszystkie nagrania zostały zarejestrowane z częstotliwością próbkowania 50 kHz i z rozdzielczością 16 bitów. Baza SVD składa się z nagrań dotyczących 71 różnych chorób.

Nagrania, które były niekompletne lub posiadały uszkodzone części zostały wyłączone z zestawu danych. W niniejszej pracy do analizy wykorzystano nagrania samogłoski /a/ dla normalnej intonacji 850 kobiet, z których 425 były zdrowe, a 425 ze stwierdzoną różną patologią głosu oraz 510 mężczyzn, z których 255 było zdrowych, a 255 należało do grupy patologicznej. Tylko dla kobiet analiza samogłoski /a/ została rozszerzona o intonację wysoką oraz niską. Ze względu na istotne różnice w głosie pomiędzy kobietami i mężczyznami analizę akustyczną przeprowadzono osobno dla każdej z grup.

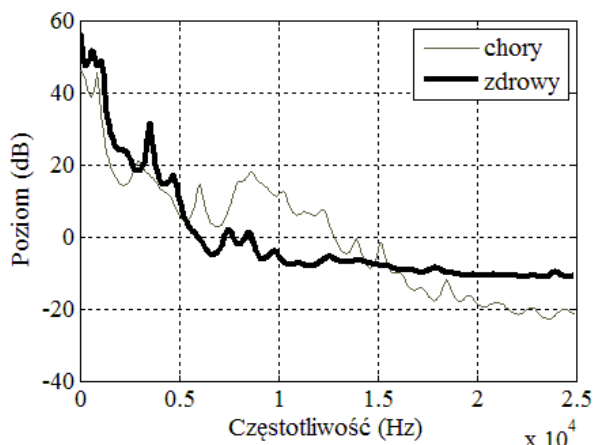
Analiza sygnału akustycznego

Celem niniejszej pracy była ocena przydatności różnych sposobów analizy sygnału mowy w wykrywaniu zmian patologicznych głosu. W pierwszym etapie skupiono się na wstępnych przekształceniach sygnału mowy w celu uzyskania zestawu parametrów, których wartości stanowią podstawę diagnozy choroby pacjenta. Sama rejestracja i jej wstępne przetwarzanie nie czyni analizy w pełni przydatnej w procesie identyfikacji, oceny zmian deformacji i patologii głosu. Dlatego konieczne staje się opracowanie oraz opisanie nagrań fonetycznych przy użyciu zestawu parametrów, które następnie uporządkowane w odpowiedniej konstrukcji – wektora cech – będą używane do opracowania modelu deformacji mowy. Takie modele mogą być podstawą oceny zmian patologicznych.

Analiza sygnału akustycznego została przeprowadzona w celu wyznaczenia 31 parametrów: częstotliwość podstawowa, współczynnik Jitter i Shimmer, energia, moc średnia, wartość skuteczna, amplituda 1- , 2- oraz 3-ciego rzędu formantu, częstotliwość formantu 1-, 2- oraz 3-ciego rzędu, kurtoza, moment 0-, 1-, 2- i 3-ciego rzędu, moc względna, wartość średnia, wartość maksymalna i minimalna sygnału oraz 10 współczynników Mel-cepstralnych (MFCC) [16-18]. Na rysunku 1 przedstawiono schemat ideowy proponowanej metody analizy, natomiast na rysunku 2 przykładową prezentację graficzną formant, wykorzystanych w niniejszej analizie.



Rys. 1 Schemat blokowy przeprowadzonej analizy akustycznej



Rys. 2 Graficzna prezentacja formant samogłoski /a/, wypowiedzianej przez mężczyznę zdrowego oraz chorego

Wybór cech

Ponieważ wektor składał się z wielu parametrów, istotne było zastosowanie metody, która pozwoliłaby wyodrębnić z nich te, które zawierały najbardziej istotne informacje.

Aby wykonać redukcję liczby parametrów, należało uporządkować ich cechy według zdolności dyskryminacyjnej, tak aby uzyskać stabilne oraz spójne wyniki, a w rezultacie otrzymać lepsze odzwierciedlenie w ogólnej wydajności systemu [3,4]. Do wyboru cech wykorzystano metodę Analizy Głównych Składowych (PCA). PCA jest procedurą matematyczną powszechnie stosowaną w rozpoznawaniu wzorców i przetwarzaniu sygnałów. Umożliwia przekształcenie zestawu zmiennych obserwacji, które mogą być skorelowane, w zestaw zmiennych nieskorelowanych, tak zwanych składowych głównych. PCA jest stosowana jako narzędzie w analizie badawczej oraz modelowaniu, kompresji i wizualizacji danych wielowymiarowych. Celem analizy PCA w tym badaniu była redukcja liczby podstawowych parametrów otrzymanych z analizy sygnału akustycznego, w wyniku czego uzyskano składowe główne opisujące przeważającą część zmienności danych. W związku z tym Analiza Głównych Składowych stosowana jest jako etap wstępnej obróbki danych przed dalszą analizą oraz do redukcji parametrów w celu przyspieszenia i uproszczenia dalszych obliczeń weryfikujących jakość algorytmu [4-8].

W wyniku przeprowadzonej analizy PCA otrzymano 31 składowych głównych, uporządkowanych odpowiednio od 1-ej składowej głównej, zawierającej największą część wariancji, do 31-ej składowej głównej, zawierającej najmniejszą część wariancji. Oznacza to, że najważniejsze

dane ekstrahowano z minimalnym zakłóceniem dla oryginalnych danych. W rezultacie wybrano 17 z 31 parametrów, które pokrywały ponad 90% zmienności początkowych parametrów.

Jądrowe PCA

Analiza głównych składowych jest metodą liniową, która służy do redukcji wymiarów poprzez liniową transformację zbioru zmiennych zależnych w zbiór mniej licznych zmiennych niezależnych. W przypadku, kiedy układ analizowanych danych nie może zostać poddany redukcji w liniowy sposób, nie odzwierciedla oczekiwanej redukcji danych do ich rzutów na wyliczone kierunki, wówczas tradycyjna PCA może być niewystarczająca [6]. W takim wypadku, redukcja wymiaru oryginalnych danych wymaga zrzutowania na krzywą, a nie na prostą. W tym celu wykorzystuje się inną metodę, tzw. jądrowe PCA (kPCA) [9-11,14]. Do obliczeń kPCA w niniejszej pracy wykorzystano funkcję jądra Gaussa. Stosując nieliniowe jądro, uzyskano granice separujące klasy nieliniowo w przestrzeni danych, ale liniowo w przekształconej za pomocą jądra przestrzeni o wyższym wymiarze. Późniejszą walidację za pomocą kroswalidacji (tzw. walidacja krzyżowa) oraz klasyfikatora k -najbliższych sąsiadów (ang. *k-nearest neighbours*) przeprowadzono dla 31 parametrów.

Liniowa analiza dyskryminacyjna

Metoda używana do znalezienia liniowej kombinacji cech, które najlepiej rozróżniają klasy obiektów lub zdarzeń, nazywana jest liniową analizą dyskryminacyjną (LDA). Metoda ta określana jest mianem metody klasyfikacji pod nadzorem [12,13]. Głównym celem zastosowania analizy dyskryminacyjnej jest przewidywanie klasyfikacji przypadków. Metoda ta rozdziela klasy (parametry) poprzez rzutowanie obserwacji na jeden kierunek kanoniczny na podstawie liniowej funkcji dyskryminacyjnej lub liniowych funkcji klasyfikacyjnych [4]. Zadaniem tej metody jest określenie przynależności do klas (tutaj: klasa zdrowy oraz chory) badanych obiektów opisanych tymi samymi parametrami. Podstawą LDA jest znalezienie predyktora przynależności dla każdej obserwacji (pacjenta) do klasy osób zdrowych oraz osób chorych. Analizę LDA przeprowadzono dwuetapowo. Najpierw przeprowadzono etap uczenia, podczas którego na podstawie zbioru uczącego znaleziono reguły klasyfikacyjne. Następnie, w drugim etapie, na podstawie znalezionych charakterystyk klas dokonano klasyfikacji zasadniczego zbioru obiektów, których przynależność była nieznana. W finalnym punkcie, za pomocą kroswalidacji sprawdzono poprawność identyfikacji patologii głosu.

Walidacja

Podczas walidacji wektora parametrów przeprowadzono ocenę jakości operatora klasyfikacji za pomocą kroswalidacji. Pod pojęciem operatora klasyfikacji rozumie się wyniki analizy dla poszczególnych pacjentów. W niniejszej pracy wykorzystano k -krotną walidację, a zbiór danych został podzielony na k podzbiorów, gdzie $k=10$. Taki jeden podzbiór zawierający 10% danych stanowił wówczas zbiór testowy, a pozostałe 90% danych były zbiorem uczącym. Analiza została wykonana odpowiednio k razy z losową przynależnością do grup. W celu klasyfikacji danych do danej grupy wykorzystano algorytm centroidów (k -średnich, ang. *k-means*).

Kroswalidacja została przeprowadzona osobno dla nagrań samogłoski /a/ wypowiedzianej w intonacji normalnej przez kobiety oraz mężczyzn, zdrowych oraz chorych oraz dla kobiet, które utrzymały tonację wyższą oraz niższą niż naturalna.

Wyniki krosvalidacji przedstawiono przy wykorzystaniu macierzy pomyłek będącej standardowym narzędziem do analizy operatora klasyfikacji, w tym przypadku parametrów.

Tabela 1. Przykład macierzy pomyłek, stosowanej w analizie

		wyniki klasyfikacji		
		zdrowy	chory	
diagnoza	zdrowy	TP (True Positive)	FP (False Positive)	Precyzja zdrowy: $P_z = \frac{TP}{TP + FP}$
	patologiczny	FN (False Negative)	TN (True Negative)	
	Czułość: $C = \frac{TP}{TP + FN}$		Swoistość: $S = \frac{TN}{FP + TN}$	

Precyzja oznaczała stosunek liczby przypadków faktycznie poprawnych, które zostały zaklasyfikowane przez system

Tabela 2. Wyniki analizy dla 31 parametrów, PCA, kPCA oraz LDA dla kobiet wypowiadających samogłoskę /a/ w normalnej intonacji

samogłoska /a/	31 pierwotnych parametrów		PCA		kPCA		LDA	
	zdrowa	chora	zdrowa	chora	zdrowa	chora	zdrowa	chora
normalna intonacja								
dokładność [%]	81,41		74,75		82,94		71,52	
precyzja [%]	74,64	88,08	80,43	69,08	73,56	94,48	63,24	81,67
czułość [%]	86,06	-	72,15	-	94,26	-	80,87	-
swoistość [%]	-	77,89	-	78,04	-	74,38	-	64,40

Tabela 3. Wyniki analizy dla wysokiej intonacji

samogłoska /a/	31 pierwotnych parametrów		PCA		kPCA		LDA	
	zdrowa	chora	zdrowa	chora	zdrowa	chora	zdrowa	chora
wysoka intonacja								
dokładność [%]	81,53		77,51		87,88		68,94	
precyzja [%]	75,93	86,57	83,74	71,28	80,09	96,12	59,88	82,93
czułość [%]	83,60	-	73,40	-	95,62	-	93,84,42	-
swoistość [%]	-	79,95	-	79,9	-	82,02	-	57,23

Tabela 4. Wyniki analizy dla niskiej intonacji

samogłoska /a/	31 pierwotnych parametrów		PCA		kPCA		LDA	
	zdrowa	chora	zdrowa	chora	zdrowa	chora	zdrowa	chora
niska intonacja								
dokładność [%]	85,06		73,26		85,17		77,29	
precyzja [%]	80,40	89,05	78,97	67,54	76,43	95,2	72,11	81,85
czułość [%]	86,33	-	72,12	-	94,80	-	77,77	-
swoistość [%]	-	84,09	-	78,61	-	77,89	-	76,92

W celu uzyskania oceny jakościowej poszczególnych analiz obliczono parametry takie jak: dokładność, precyzję, czułość oraz swoistość (tab. 1). Dokładność klasyfikatora (ACC) określono jako stosunek parametrów ze zbioru testowego, które zostały właściwie przypisane do odpowiednich klas, do wszystkich parametrów ze zbioru testowego:

$$(1) \quad ACC = \frac{TP + TN}{TP + FN + TN + FP}$$

jako poprawne oraz w rzeczywistości były poprawne do wszystkich zaklasyfikowanych przez system jako poprawne. Czułość dotyczyła zdolności testu do jednoznacznej identyfikacji dodatnich wyników, natomiast swoistość dotyczyła identyfikacji wyników negatywnych.

Wyniki

Aby przyspieszyć oraz uprościć obliczenia 31 parametrów, w badaniach wykorzystano metodę Analizy Głównych Składowych. W rezultacie wybrano 17 parametrów oraz poddano je ostatecznej analizie. Ponadto, wykorzystano metodę nieliniowego jądrowego PCA oraz metodę liniowej analizy dyskryminacyjnej.

W wyniku analizy wykrywalność patologii u mężczyzn potwierdzona parametrami takimi jak: dokładność, precyzja, czułość oraz swoistość we wszystkich rodzajach analizy, oprócz metody LDA, wynosiła 100%. W metodzie LDA dokładność wynosiła 96,4%, precyzja dla mężczyzn zdrowych 92,96%, a dla patologicznych 100%. Czułość oraz swoistość wyniosła również 100%. Oznacza to, że metoda analizy głównych składowych oraz jądrowe PCA pozwalają skutecznie zidentyfikować patologie w męskim głosie. U kobiet, które wymawiały samogłoskę /a/ w normalnej wysokiej oraz niskiej intonacji (tab. 2,3,4) najwyższą dokładność oraz precyzję otrzymano dla metody kPCA.

Podczas analizy parametru czułości najwięcej poprawnych wyników również otrzymano dla metody kPCA, natomiast dla swoistości metoda PCA okazała się być najbardziej użytecznym narzędziem, gdy kobiety wymawiały samogłoskę przy normalnej i niskiej intonacji. Dla wysokiej intonacji jądrowe PCA charakteryzował wynik o około 2% większy niż dla metody analizy głównych składowych.

Wnioski

Dostępna w internecie oraz darmowa baza z nagraniami akustycznymi została wykorzystana w niniejszej pracy w celu identyfikacji patologii głosu. W oparciu

o przeprowadzone badania pokazano, że obliczone parametry oraz ich redukcja umożliwiły rozróżnienie cech akustycznych dla mowy patologicznej oraz "zdrowej". W niniejszym artykule omówiono metody umożliwiające walidację utworzonego wektora cech, który miał potwierdzić status zdrowia lub wykryć patologię głosu u pacjenta. U mężczyzn detekcja stanu zdrowego lub zaburzeń w sygnale mowy była bezbłędna dla metod PCA oraz kPCA. Liniowa analiza dyskryminacyjna wykazała dokładność, która wynosiła 96%. Analiza przeprowadzona u kobiet, które wypowiadały samogłoskę /a/ w intonacji normalnej, niskiej oraz wysokiej, wykazała, że dokładność, precyzja oraz czułość obliczona na podstawie metody kPCA ma najwyższe wartości. Parametr swoistość wykazywał nieco wyższe wyniki dla rezultatów otrzymanych metodą standardowej Analizy Głównych Składowych niż kPCA.

W wyniku przeprowadzenia metody LDA, otrzymano najmniej poprawnych wyników.

Podsumowując, najlepsze wyniki, co miało przełożenie na najlepszą dokładność w identyfikacji patologii mowy, otrzymano dla analizy kPCA, podczas gdy PCA pozwoliło na zredukowanie liczby parametrów, co skróciło czas obliczeń. Podczas analizy akustycznej głosu kobiet, porównując różne intonacje, największy procent poprawnych wyników otrzymano dla wysokiej intonacji. Powodem tego może być to, że dla tej samej samogłoski wypowiedzianej w różnych intonacjach związek pomiędzy drugą harmoniczną i pierwszą formantą może się zmieniać, powodując, że amplituda harmonicznym jest sztucznie wzmocniona [15].

Praca była finansowana przez Ministerstwo Nauki i Szkolnictwa Wyższego w ramach programu Diamentowy Grant, nr decyzji: 0136/DIA/2013/42 (AGH 68.68.120.364).

LITERATURA

- [1] Gogh C.D.L., Festen J.M., Verdonck-de Leeuw I.M., Parker A.J., *Acoustical analysis of tracheoesophageal voice*, *Speech Communication*, Vol. 47, Issue 1-2 (2005), 160-168
- [2] Martinez D., Lleida E., Ortega A., Miguel A., Villalba J., *Voice Pathology Detection on the Saarbruecken Voice Database with Calibration and Fusion of Scores Using MultiFocal Toolkit*, *Advances in Speech and Language Technologies for Iberian Languages Communications in Computer and Information Science*, vol. 328 (2012), 99-109
- [3] Orozco-Arroyave J.R., Murillo-Rendon S., Alvarez-Meza A.M., Arias-Londono J.D., Delgado-Trejos E., Vargas-Bonilla J.F., Castellanos-Domingues C.G., *Automatic Selection of Acoustic*

- and Non-linear Dynamic Features in Voice Signals for Hypernasality Detection*, *Interspeech* (2011), 529-532
- [4] Bishop Ch.M., *Pattern Recognition and Machine Learning*, Springer Science (2006), Singapur, 559-599
- [5] Vetter R., Virag N., Renevey P., Vesin J.M., *Single channel speech enhancement using principal component analysis and MDL subspace selection*, *Eurospeech* (1999), ISCA
- [6] Yang J., Yang J.-Y., *Why can LDA be performed in PCA transformed space?*, *Pattern Recogn.* 36(2) (2003), 563-566
- [7] Liu C., *Gabor-based kernel PCA with fractional power polynomial models for face recognition*, *IEEE Trans, Pattern Anal. Mach. Intell.*, 26(5) (2004), 572-581
- [8] Gramacki J., Gramacki A., *Wybrane metody redukcji wymiarowości danych oraz ich wizualizacje*, XIV Konferencja PLOUG (2008), Szczyrk
- [9] Scholkopf B., Smola A., Muller K.R., *Kernel principal component analysis*, *Artificial Neural Networks - iCANN'97*, Berlin (1997), 583-588
- [10] Scholkopf B., Smola A., Muller K.R., *Nonlinear component analysis as a kernel eigenvalue problem*, *Neural Comput.*, 10(5) (1998), 1299-1319
- [11] Wu J., Wang J., Liu L., *Feature extraction via KPCA for classification of gait patterns*, *ScienceDirect, Human Movement Science* 26(2007), 393-411
- [12] Balakrishnama S., Ganapathiraju A., *Linear Discriminant Analysis - a brief tutorial*, Institute for Signal and Information Processing, Department of Electrical and Computer Engineering, Mississippi State University (2014)
- [13] Wu Y., *Principal Component Analysis and Linear Discriminant Analysis*, Electrical Engineering and Computer Science, Northwestern University, Evanston, wykład (2014)
- [14] Young X., Zhang D., Song F., Yang J.-Y., Jing Z., Li M., *A method for speeding up feature extraction based on KPCA*, *Neurocomputing* 70 (2007), 1056-1061
- [15] Epstein M.A., Payri B.G., *The effects of vowel quality and pitch on spectral and glottal flow measurements of the voice source*, Lecture, University of California, Los Angeles, http://www.linguistics.ucla.edu/people/grads/melissa/effects_of_vowel_quality.PDF (2014)
- [16] Maciel C.D., Pereira J.S., *Identifying healthy and pathologically affected voice signals*, *IEEE Signal Processing Magazine* (2010), 120-123
- [17] Osowski S., *Sieci neuronowe w ujęciu algorytmicznym*, WNT, (1996)
- [18] Engel Z.W., Klaczynski M., Wszolek W., *A Vibroacoustic Model of Selected Human Larynx Diseases*, *International Journal of Occupational Safety and Ergonomics (JOSE)*, 13(4) (2007), 367-379

Autorzy: mgr inż. Daria Panek, AGH Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, Katedra Metrologii i Elektroniki, al. Mickiewicza 30, 30-060 Kraków, E-mail: dpanek@agh.edu.pl.