

Wpływ długości fonacji na ilość informacji zawartej w sygnale głosu ludzkiego

Streszczenie. Skuteczność detekcji nieprawidłowości w głosie ludzkim wykorzystująca cyfrowe przetwarzanie sygnałów akustycznych w ostatnich latach wzrosła, lecz nadal nie ma uniwersalnego rozwiązania, które dostarczyłoby pewnej informacji na temat stanu badanego narządu głosu. Dotychczas nie istnieją żadne standardy rejestracji głosu uwzględniające określony czas fonacji mowy. W badaniach sprawdzono w jakim stopniu czas trwania fonacji wypowiedzianej samogłoski /a/ ma wpływ na ilość informacji zawartej w sygnale głosu z wykorzystaniem 28 parametrów opisujących sygnał akustyczny.

Abstract. The detection accuracy of irregularities in the human voice based on the digital processing of acoustic signals increased in recent years, but there is no universal solution that would provide a certain on the state of the examined voice organ. So far, there are no standards for voice registration including specified time of the phonation. In this study we examined the duration of the spoken vowel /a/ affects the amount of information contained in the voice signal using 28 parameters describing acoustic signal. (The effect of phonation time on the amount of information contained in the signal).

Słowa kluczowe: fonacja, analiza akustyczna, parametry sygnału
Keywords: phonation, acoustic analysis, voice parameters

Wstęp

Pomimo bardzo dobrze rozwiniętej technologii cyfrowej w analizie akustycznej nadal istnieje problem ekstrakcji istotnych informacji o stanie pacjenta. Głos jest dynamiczny i złożony, wypowiedzany jest w różnych intonacjach i w różnych emocjach. Nadal poszukiwane są rozwiązania jak pozyskać informacje dotyczące stanu narządu głosu pacjenta, które pomogą w skutecznym rozróżnieniu pacjenta chorego od zdrowego [1]. Analiza akustyczna może być przeprowadzona przy pomocy: metody klasycznej, wykorzystującej subiektywną ocenę głosu oraz metody opartej o nowoczesne technologie (analiza akustyczna, sonografia, obrazy rejestrujące trakt głosowy itd.). Jeśli diagnoza jest niejednoznaczna, metody te mogą być łączone [2].

Rejestracja sygnału głosu ludzkiego jest metodą nieinwazyjną umożliwiającą wykonanie ilościowej oceny głosu pozwalając przy tym na obiektywną ocenę dyskretnych, nie zawsze widocznych w metodzie odsłuchowej patologii głosu. W praktyce klinicznej, jakoś głosu pacjenta ocenia się za pomocą przedłużonych fonacji samogłosek i/lub analizy mowy. Samogłoski o przedłużonej fonacji są szczególnie przydatne, ponieważ nie wymagają one uwzględnienia artefaktów językowych, co powoduje że są uważane za wystarczające dla wielu aplikacji oceny głosu [3]. Podczas badania mowy pacjent jest proszony o utrzymanie fonacji danej samogłoski tak długo i stabilnie, jak to możliwe (utrzymując na tym samym poziomie intonację oraz amplitudę).

W literaturze można spotkać różne parametry akustyczne pomagające w analizie pomiaru niestabilności sygnału, hałasu i ogólnych problemów artykulacyjnych. Najczęściej używanymi parametrami są: częstotliwość podstawowa, współczynnik Jitter, Shimmer, współczynnik stosunku energii sygnału tonalnego do energii szumu oraz współczynniki Mel-cepstralne [4-10]. W ostatnich latach naukowcy skupiali swoje prace na automatycznych systemach rozpoznających nieprawidłowości w głosie ludzkim. Praca [11] prezentuje 95.74% skuteczność detekcji danego typu patologii na podstawie nagrań samogłoski /a/ o przedłużonej fonacji wypowiedzianej przez okres od 1 do 5 sekund przez 363 pacjentów. Artykuł [12] przedstawia różnice w sygnale akustycznym na podstawie 7 parametrów pomiędzy pacjentami zdrowymi i z polipami fałdów

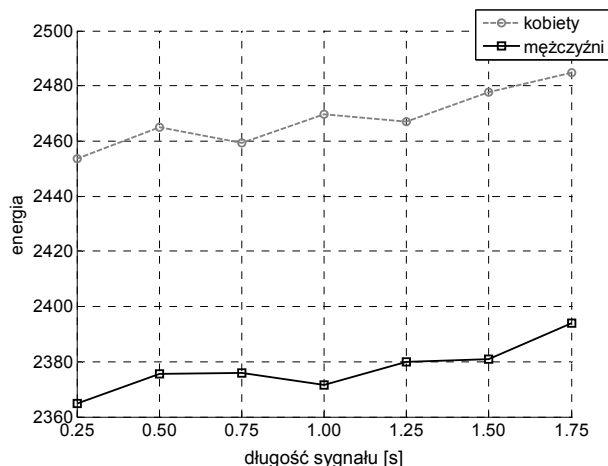
głosowych. Badania zostały przeprowadzone z wykorzystaniem nagrań samogłoski /a/ o przedłużonej fonacji trwającym przynajmniej 3 sekundy przez 46 kobiet. System do automatycznej oceny głosu przedstawiony przez [13] uzyskuje 65% skuteczność przy wykorzystaniu nagrań samogłoski /a/ o przedłużonej fonacji oraz zdania "es habil un solo dia" wypowiedzianych przez 648 pacjentów. Czas fonacji wynosił 3-4 sekundy. Klasyfikacja patologicznego głosu opublikowana w [14] przedstawia skuteczność na poziomie 97.33% wykorzystując nagrania 258 pacjentów wypowiadających przedłużoną fonację samogłoski /a/ trwającą 1 sekundę. W pracy [15], autorzy prezentują metodę detekcji polipów fałdów głosowych osiągając skuteczność 95-100%, gdzie w badaniach uwzględniono 710 nagrań trwających od 1 do 3 sekund samogłoski /a/ oraz przeczytania "Rainbow Passage". Ta sama samogłoska trwająca od 1 do 3 sekund została poddana badaniom [16] w celu poszukiwania patologii głosu osiągając skuteczność 90.7%. Prace [17-22] przedstawiają różne metody analizy sygnału akustycznego, których celem jest automatyczna detekcja patologii głosu oraz wspomaganie decyzji lekarza. Prace te nie uwzględniają czasu trwania nagrań. Żadna z tych prac nie przedstawia skuteczności detekcji patologii wyższej niż 96.1%.

Nadal nie ma uniwersalnego rozwiązania, które jednoznacznie wskazywałoby na stan traktu głosowego. W przeprowadzanych badaniach czas trwania fonacji jest zaniebywanym czynnikiem.

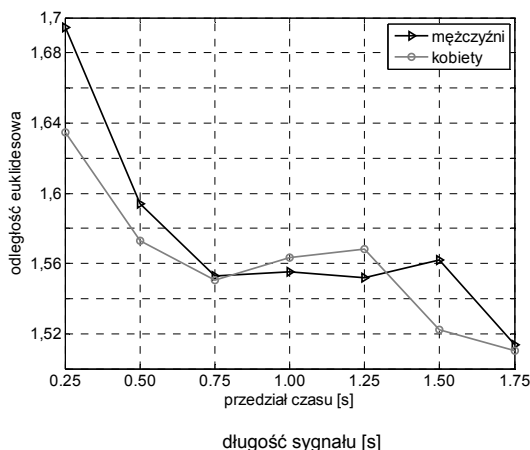
Niniejsza praca skupia uwagę na możliwościach jakie stwarza odpowiednie przeprowadzenie nagrań sygnału akustycznego mowy w obszarze diagnostyki akustycznej dla celów medycznych. Dotychczas nie istnieją żadne standardy rejestracji głosu uwzględniające określony czas fonacji mowy czy warunki rejestracji, co powoduje, że przy wykonywaniu oceny zmian w sygnale akustycznym konieczne jest indywidualne poszukiwanie cech diagnostycznych odnoszących się do rozważanego problemu. Głównym celem opisywanego eksperymentu było zbadanie w jakim stopniu czas trwania fonacji wypowiedzianej samogłoski ma wpływ na ilość informacji zawartej w sygnale głosu. Określenie optymalnej długości fonacji pozwoli przybliżyć się do ustalenia standardów badań akustycznych w kontekście detekcji patologii traktu głosowego.

Tabela 1. Średnie odchylenie standardowe w zadanych przedziałach czasowych

	0-0.25 [s]	0-0.50 [s]	0-0.75 [s]	0-1.00 [s]	0-1.25 [s]	0-1.50 [s]	0-1.75 [s]
kobiety	0.32	0.30	0.30	0.30	0.30	0.29	0.29
mężczyźni	0.31	0.30	0.30	0.30	0.30	0.29	0.29



Rys. 1. Średnia energia dla każdej z grup czasowych dla kobiet oraz mężczyzn dla samogłoski /a/ wypowiedzianej w intonacji normalnej



Rys. 2. Średnia odległość euklidesowa pomiędzy każdym z pomiarów oraz średnią dla każdej z grup czasowych dla kobiet oraz mężczyzn dla samogłoski /a/ wypowiedzianej w intonacji normalnej

Material

W badaniach wykorzystano nagrania z bazy Saarbruecken Voice Database (SVD) opublikowanej przez Instytut Fonetyki w Saarland w Niemczech [23]. Baza ta zawiera nagrania akustyczne od ponad 2000 pacjentów, w tym zdrowych oraz pacjentów z różnymi patologiami głosu. Baza SVD zawiera nagrania samogłosek /a/, /i/ oraz /u/ wypowiedzianych w intonacji wysokiej, niskiej, normalnej oraz modulowanej. Czas trwania nagrań samogłosek wynosi od 0.5 do 4 sekund. Nagrania zostały zarejestrowane z częstotliwością próbkowania 50 kHz i z rozdzielczością 16 bitów.

W niniejszych badaniach do analizy wykorzystano nagrania samogłoski /a/ wypowiedzianej w normalnej intonacji przez 30 zdrowych kobiet i 30 zdrowych mężczyzn. Czas trwania wybranych nagrań wynosił ponad 1.75 sekundy. Każde z nagrań zostało podzielone na odcinki czasu o długościach: 0.25 s, 0.5 s, 0.75 s, 1.00 s, 1.25 s, 1.5 s oraz 1.75 s.

Metody

W pierwszym etapie skupiono się na wstępnych przekształceniach sygnału głosu tak aby uzyskać zestaw parametrów. Podstawą do oceny stanu narządu głosu jest rejestracja nagrań oraz przetwarzanie sygnałów wydobywające z sygnału istotne cechy. Analiza akustyczna była przeprowadzona w oparciu o 28 parametrów: częstotliwość podstawową, współczynnik Jitter, Shimmer, energię sygnału, moment 0-, 1-, 2-, 3- rzędu, kurtozę, moc względna, amplituda 1-, 2-, 3- formantu, częstotliwość formantu 1-, 2- oraz 3- rzędu, wartość maksymalna i minimalna oraz 10 współczynników Mel-Cepstralnych (MFCC). Wyjaśnienie nieniejszych parametrów znajduje się w poprzedniej pracy autorów [24]. Na tym etapie, sygnał akustyczny jest reprezentowany przez zestaw wielowymiarowego wektora cech. Analiza ta została przeprowadzona osobno dla kobiet oraz dla mężczyzn w wyodrębnionych przedziałach czasu. Na rysunku 1 przedstawiono średnią energię sygnału dla kobiet i mężczyzn dla wybranych przedziałów czasowych.

W kolejnym etapie prac, każdy z parametrów został znormalizowany uwzględniając każdy z przedziałów czasowych. W niniejszym eksperymencie badano rozrzut danych w obrębie każdej z grup czasowych. Jako miarę rozrzutu obliczano odległość euklidesową w 28 wymiarowej przestrzeni pomiędzy wynikami każdego z nagrań i średniej w obrębie całej grupy sygnałów w danym przedziale czasowym. Takie podejście umożliwiło sprawdzenie koncentracji wektorów w przestrzeni parametrów. Jest to niezwykle istotne podczas późniejszej klasyfikacji. W związku z tym, iż dane pochodziły od pacjentów bez patologii głosu to ich wektory parametrów powinny leżeć blisko siebie w przestrzeni cech. Innymi słowy wariancja w ramach klasy powinna być jak najmniejsza.

Wyniki przedstawione na rys. 2 prezentują średnią odległość wektora od centroidu klasy. Sprawdzone również średnie odchylenie standardowe w obrębie każdej z grup (tabela 1).

Wyniki

Analizując wyniki przedstawione na rysunku 2 można zaobserwować, że odległość euklidesowa maleje wraz ze wzrostem długości sygnału dla kobiet oraz dla mężczyzn. Wskazuje to na fakt, że rozrzut w obrębie wybranych grup czasowych maleje wraz ze wzrostem czasu trwania sygnału akustycznego. Trend ten potwierdza również analiza średniego odchylenia standardowego (tabela 1), gdzie przy analizie dłuższego sygnału średnie odchylenie standardowe maleje, podczas gdy przy najkrótszym czasie trwania sygnału odchylenie jest najwyższe.

Wnioski

Poszukiwane są nadal nowe metody analizy sygnału akustycznego w celu wydobywania jak największej ilości informacji dotyczącej ewentualnych nieprawidłowości. Analizę cyfrowego przetwarzania sygnału poprzedza rejestracja głosu pacjenta. Obecnie nie ma standardów jak długo ma trwać rejestracja, nie ma określonego minimum, dla którego określona jest minimalna ilość informacji w sygnale niezbędna w celu postawienia diagnozy. Niewystarczająca ilość informacji w sygnale może skutkować zmniejszeniem skuteczności automatycznych metod diagnostycznych.

W przedstawianym eksperymencie analizie poddano łącznie 60 sygnałów akustycznych o czasie trwania min. 1.75s. Następnie każdy z sygnałów podzielono na 7 segmentów czasu, które odpowiednio trwały: 0.25 s, 0.5 s, 0.75 s, 1 s, 1.25 s, 1.5 s, 1.75 s. Następnie zbadano, czy długość sygnału poddanego analizie ma wpływ na rozrzut danych z wykorzystaniem stworzonego 28-wymiarowego wektora cech.

W wyniku analizy widoczne jest, że długość sygnału ma wpływ na ilość informacji zawartej w sygnale. W rezultacie, im dłuższy jest sygnał, tym mniejszy rozrzut wektorów parametrów reprezentujących sygnał.

Oszacowany na podstawie wybranych parametrów czas rejestracji głosu może pozwolić na dokładniejszą klasyfikację dowolnie badanego sygnału akustycznego mowy. Informacje zawarte w wyniku takich badań mogą służyć do celów diagnostycznych, profilaktycznych (prewencyjnych) bądź terapeutycznych.

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] Arroyave J.R.O., Bonilla J.F.V., Trejos E.D., Acoustic analysis and non-linear dynamics applied to voice pathology detection: A review, *Recent Patents on Signal Processing* 2, (2012), n.2, 133-137
- [2] Godino-Llorente J.I., S'aenz-Lech'on N., Osma-Ruiz V., Aguilera-Navarro S., G'omez-Vilda P., An integrated tool for the diagnosis of voice disorders, *Medical Engineering & physics* 28 (2006), n.3, 276-289
- [3] Tsanas A., Acoustic analysis toolkit for biomedical speech signal processing: concepts and algorithms, *Models and analysis of vocal emissions for biomedical applications*, 2 (2013), 37-40
- [4] Manfredi C., D'Aniello M., Brusciagioni P., Ismaelli A., A comparative analysis of fundamental frequency estimation methods with application to pathological voices, *Medical engineering & physics* 22 (2000), n.2, 135-147
- [5] Farrus M., Hernando J., Ejarque P., Jitter and shimmer measurements for speaker recognition, *INTER-SPEECH* (2007), 778-781
- [6] Lieberman P., Some acoustic measures of the fundamental periodicity of normal and pathologic larynges, *The Journal of the Acoustical Society of America* 35(1963), n.3, 344-353
- [7] Yumoto E., Gould W. J., Baer, T., Harmonics-to-noise ratio as an index of the degree of hoarseness, *The journal of the Acoustical Society of America* 71 (1982), n.6, 1544-1550
- [8] Rabiner L. R., Juang, B.H., *Fundamentals of speech recognition*, 14 (1993), PTR Prentice Hall Englewood Cliffs
- [9] Godino-Llorente J. I., Gomez-Vilda P., Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors, *Biomedical Engineering, IEEE Transactions on* 51(2004), n.2, 380-384
- [10] Saldanha J. C., Ananthakrishna T., Pinto, R., Vocal fold pathology assessment using mel-frequency cepstral coefficients and linear predictive cepstral coefficients features, *Journal of Medical Imaging and Health Informatics* 4(2014), n.2, 168-173
- [11] Jothilakshmi S., Automatic system to detect the type of voice pathology, *Applied Soft Computing* 21 (2014), 244-249
- [12] Petrovic-Lazic M., Babac S., Vukovic M., Kosanovic R., Ivankovic Z., Acoustic voice analysis of patients with vocal fold polyp, *Journal of Voice* 25 (2011), n.1, 94-97
- [13] Saenz-Lechon, N., Godino-Llorente J. I., Osma-Ruiz V., Blanco-Velasco M., Cruz-Roldan F., Automatic assessment of voice quality according to the GRBAS scale, *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE* (2006), 2478-2481.
- [14] Akbari A., Arjmandi M.K., 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 (2014), 209-223
- [15] Hariharan M., Polat K., Yaacob S., A new feature constituting approach to detection of vocal fold pathology, *International Journal of System Science*, 45 (2014), n.8, 1622-1634
- [16] 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* (2012), 99-109
- [17] Umapathy K., Krishnan S., Parsa V., Jamieson D. G., Discrimination of pathological voices using a time-frequency approach, *Biomedical Engineering, IEEE Transactions on* 52 (2005), n.3., 421-430
- [18] Godino-Llorente J. I., S'aenz-Lech'on N., Osma-Ruiz V., Aguilera-Navarro S., G'omez-Vilda, P., An integrated tool for the diagnosis of voice disorders, *Medical engineering & physics* 28 (2006), n.3, 276-289.
- [19] Drugman T., Dubuisson T., Dutoit T., Phase-based information for voice pathology detection, *Acoustics Speech and Signal Processing (ICASSP), International Conference on IEEE* 2011, 4612-4615
- [20] Orozco J.R., Vargas J.F., Alonso J.B., Ferrer M.A., Travieso C.M., Henriquez P., Voice pathology detection in continuous speech using nonlinear dynamics, *Information Science, Signal Processing and their Applications (ISSPA)* (2012), 1030-1033
- [21] Hadjitodorov S., Boyanov B., Teston B., Laryngeal pathology detection by means of class-specific neural maps, *Information Technology in Biomedicine, IEEE Transactions* 4.1 (2000), 68-73
- [22] Hadjitodorov S., Mitev P., A computer system for acoustic analysis of pathological voices and laryngeal diseases screening, *Medical engineering & physics* 24.6 (2002), 419-429.
- [23] Saarbruecken Voice Database, <http://www.stimmdatenbank.coli.uni-saarland.de/>
- [24] Panek D., Skalski A., Gajda J., Quantification of linear and non-linear acoustic analysis applied to voice pathology detection, *Information Technologies in Biomedicine* 4, (2014) 355-364

Autorzy: mgr inż. Daria Panek, AGH Akademia Górniczo-Hutnicza w Krakowie, Katedra Metrologii i Elektroniki, al. Mickiewicza 30, 30-059 Kraków, E-mail: dpanek@agh.edu.pl.
 dr inż. Andrzej Skalski, AGH Akademia Górniczo-Hutnicza w Krakowie, Katedra Metrologii i Elektroniki, al. Mickiewicza 30, 30-059 Kraków,
 prof. dr hab. inż. Janusz Gajda, AGH Akademia Górniczo-Hutnicza w Krakowie, Katedra Metrologii i Elektroniki, al. Mickiewicza 30, 30-059 Kraków,