

XVI Seminarium
ZASTOSOWANIE KOMPUTERÓW W NAUCE I TECHNICIE' 2006
Oddział Gdański PTETiS
Referat nr 12

ANALIZA SYGNAŁU MOWY W FONIATRII

Andrzej KACZMAREK

Politechnika Gdańska, ul. Gabriela Narutowicza 11/12
tel: (058) 3472444 fax: (058) 3471301 e-mail: akacz@sound.eti.pg.gda.pl

Praca zawiera opis algorytmów analizy sygnału mowy opracowanych do zastosowań medycznych. Mają one znaczenie wspomagające w diagnostyce schorzeń krtani, objawiających się jako np. chrypka, dwugłos oraz nieprawidłowa artykulacja tonu krtaniowego (głos drżący). Opisano algorytmy służące do ekstrakcji tonu krtaniowego (parametr F0 i jego przebieg zmienności w czasie) oraz do wygładzania widma (np. do estymacji parametrów formantowych lub do wyodrębnienia składowych harmonicznym tonu krtaniowego). Przykładowe wyniki porównano z wykresami obrazującymi zmienność czasowo-częstotliwościową (spektrogramami). Opisano także algorytmy tworzące histogram dla parametru F0 (intonogram) oraz obliczające wskaźniki jitter i shimmer. Przykładowe wyniki porównano z parametrami otrzymanymi na podstawie funkcji autokorelacji dla głosek dźwięcznych. Algorytmy te zostały zaimplementowane w ramach komputerowego analizatora mowy patologicznej. Na podstawie przedstawionych badań można oczekiwać stworzenia w przyszłości nowych obiektywnych miar obejmujących różnorodne patologie związane z narządami mowy (także w przypadku osób niedosłyszących), oraz w oparciu o te parametry, odpowiedniego systemu eksperckiego, służącego do wspomagania diagnostyki mowy.

1. WPROWADZENIE

Badania foniatryczne są skupione głównie na badaniu krtani, która w procesie artykulacji mowy odgrywa rolę generatora akustycznego. Wynikiem działania tego generatora jest tak zwany ton krtaniowy, dźwięk okresowy o częstotliwości podstawowej od kilkudziesięciu Hz dla basów, do kilkuset Hz dla sopranów. Cechą charakterystyczną tego dźwięku jest bogate widmo, zajmujące pasmo częstotliwości do kilku kHz. Zawiera ono wszystkie składowe harmoniczne - zarówno parzyste jak i nieparzyste. Obwiednia tego widma to krzywa malejąca monotonicznie, dająca się aproksymować przy pomocy linii prostej o nachyleniu od -6 dB/oktawę do -12 dB/oktawę. Ton krtaniowy bywa często modelowany jako przebieg piłokształtny, charakteryzujący się składem widmowym z wszystkimi składowymi harmonicznymi o malejących amplitudach w tempie 6 dB/oktawę. Cechy akustyczne tonu krtaniowego zależą od ciśnienia podgłośnia i sprawności interakcji fałdów głosowych.

W badaniach foniatrycznych nie ma bezpośredniego dostępu do tonu krtaniowego, korzysta się więc z wynikowego sygnału mowy, który jak wiadomo powstaje z udziałem całego traktu głosowego (język, usta, trakt nosowy). Konfiguracja traktu głosowego determinuje transmitancję przejścia w modelu fizycznym procesu artykulacji. Wnęki traktu powodują istnienie częstotliwości rezonansowych, objawiających się w widmie sygnału mowy jako uwypuklenie niektórych zakresów częstotliwości. Tego typu obszary nazywane są formantami i oznaczane są kolejno od małych częstotliwości zaczynając jako F1, F2, F3 itd. Formanty są opisywane przez ich parametry: częstotliwość i poziom. Częstotliwości formantowe to częstotliwości środkowe tych pasm, oznaczane jako F1, F2, F3 itd.

Parametry te bardzo dobrze identyfikują samogłoski i dlatego też mogą służyć do określenia metodyki analizy mowy w przypadkach patologicznych. W badaniach foniatrycznych formanty różnych samogłosek wykorzystywane są np. do diagnostyki chrypki (skala Yanagihary).

W przypadkach klinicznych spotyka się jedno lub dwustronne porażenie krtani, które zaburza w istotny sposób jej funkcjonowanie jako generatora głosu. Czasami nawet występuje sytuacja skrajna - afonia, czyli brak drgań, ale zazwyczaj spotyka się różne stopnie dysfonii, widoczne jako nieprawidłowości sygnału pobudzającego cały trakt głosowy.

2. ANALIZA AKUSTYCZNA MOWY PATOLOGICZNEJ

2.1. Podstawy

Zostało stwierdzone, że zmiany struktury anatomicznej i ograniczenie czynności krtani wpływają na zmianę parametrów fizycznych modelu źródła tonu krtaniowego i znajdują odzwierciedlenie w postaci odpowiednich zmian struktury akustycznej głosu. Stąd powstał wniosek, że na podstawie analizy parametrów akustycznych można określić anomalie struktury anatomicznej i przyczyny ograniczenia czynności ruchowych. W efekcie badanie akustyczne stało się integralną częścią oceny foniatrycznej, przy czym największe znaczenie mają metody czasowo-częstotliwościowe [1], [2].

Ciągle trwają poszukiwania najlepszej metodyki pomiarowej, wzorów i algorytmów komputerowych mogących dać odpowiedź na pytanie, czy w badanym głosie występują cechy patologiczne, czy też należy on do kategorii głosów prawidłowych [3].

Podstawową analizą stała się analiza przebiegu zmienności parametru F0, która stała się głównym narzędziem stosowanym przy ocenie stanu krtani a od dawna najwyższą ocenianą przydatność procentowego, średniego wskaźnika zaburzenia wysokości tonu postawowego, przydatna jest także analiza rozkładów statystycznych okresu tonu postawowego T0 i rozkładu procentowego jego zmian w czasie. Z parametrem F0 (lub T0) jest związany parametr jitter stosowany do szacowania stabilności tonu krtaniowego [4].

Histogramy F0 pozwalają na dokumentowanie poprawy stanu głosu po leczeniu i rehabilitacji, a także są dobrą metodą do śledzenia poprawy głosu u osób głuchych po przeprowadzonym protezowaniu słuchu.

2.2. Opracowane algorytmy

Prezentowane w pracy wyniki są uzyskane dla analizy widmowej opartej na krótkoterminowej transformacji Fouriera, przy segmentacji 1024 próbek (ok. 46,44 ms przy częstotliwości próbkowania 22,05 kHz) i wykorzystaniu okna Blackmana.

Podstawą dla wszystkich pozostałych analiz jest spektrogram, na podstawie którego obliczane jest widmo średnie, istnieje także możliwość wyboru widma chwilowego. Widma te po wyglądzeniu umożliwiają uzyskanie parametrów formantowych. Na podstawie po-



szczególnych widm chwilowych obliczane są cepstra mocy stanowiące w sumie cepstrogram, który jest podstawą dla estymacji częstotliwości tonu kraniowego. Analiza cepstralna jest realizowana w oparciu o transformację kosinusową:

$$\hat{X}_c(k) = \sum_{n=0}^{N-1} [\ln|X(n)|] \cdot \cos\left(\frac{(n-0.5) \cdot k \cdot \pi}{N}\right) \quad (1)$$

gdzie: $X(n)$ – dyskretne widmo mocy, n – numer prążka widma, N – numer maksymalnego prążka widma analizowanego pasma częstotliwości, k – numer współczynnika cepstralnego

Widmo wygładzone cepstralnie z kolei uzyskuje się ze wzoru:

$$X_c(n) = \sum_{k=0}^K \hat{X}_c(k) \cdot \cos\left(\frac{n \cdot k \cdot \pi}{N}\right) \quad (2)$$

gdzie: K – rząd wygładzania cepstralnego, inne oznaczenia j.w.

Stąd jest uzyskiwany wygładzony spektrogram, natomiast na podstawie przebiegu zmienności parametru F0 zostaje obliczony parametr jitter.

Jitter i shimmer można określić ilościowo poprzez iloraz zaburzenia częstotliwości dla jitter oraz iloraz zaburzenia energii dla shimmer wg. wzoru (perturbation quotient):

$$PQ = \frac{100\%}{N - K} \cdot \sum_{i=\frac{K-1}{2}}^{\frac{N-K-1}{2}} \left| \frac{u(i) - \frac{1}{K} \cdot \sum_{k=\frac{K-1}{2}}^{\frac{K-1}{2}} u(i+k)}{\frac{1}{K} \cdot \sum_{k=\frac{K-1}{2}}^{\frac{K-1}{2}} u(i+k)} \right| \quad (3)$$

gdzie: $u(i)$ - sekwencja długości pojedynczych okresów (lub częstotliwości) w przypadku jitter (PPQ – pitch perturbation quotient) albo sekwencja energii pojedynczych okresów w przypadku shimmer (EPQ – energy perturbation quotient)

Dodatkowo został opracowany algorytm obliczający funkcję autokorelacji, która może służyć jako miara okresowości i częstotliwości sygnału (wielkość i położenie głównego maksimum). Równolegle analizując przebieg czasowy uzyskuje się parametr shimmer. Funkcja autokorelacji to:

$$R(n) = \frac{\sum_{i=1}^k (X_i - \overline{X_{k,i}})(X_{i+n} - \overline{X_{k,i+n}})}{\sqrt{\sum_{i=1}^k (X_i - \overline{X_{k,i}})^2 \sum_{i=1}^k (X_{i+n} - \overline{X_{k,i+n}})^2}} \quad (4)$$

gdzie: k - długość ramki w próbkach, X_j – wartość j -tej próbki oraz $\overline{X_{k,i}} = \frac{1}{k} \sum_{j=i}^{k+i} X_j$ (średnia arytmetyczna).



W oparciu o ten algorytm uzyskuje się alternatywny do cepstralnego przebieg zmienności parametru F0 służący również do obliczenia parametru jitter.

2.3. Zastosowanie

Analiza akustyczna może służyć do badania schorzeń z następującymi objawami: chrypka (obecność składowych szumowych), dwugłos (skokowe zmiany parametru F0 o oktawę), nieprawidłowa artykulacja tonu krtaniowego (głos drżący – chaotyczne zmiany parametru F0).

Przyczynami chrypki są zmiany patologiczne związane ze strunami głosowymi (lub inne zmiany w ich otoczeniu - np. mały guz). Akustyczne cechy chrypki są zdeterminowane przez składowe szumowe w głównych formantach samogłosek oraz dla wyższych częstotliwości (ponad 3kHz) a także przez obecność składowych harmonicznnych.

Powyższe dane stanowią podstawę skali chrypki Yanagihary. Skala ta jest czterostopniowa i oparta jest na wynikach analizy widmowej samogłosek. Klasyfikacja chrypki wg. Yanagihary zawiera następujące typy:

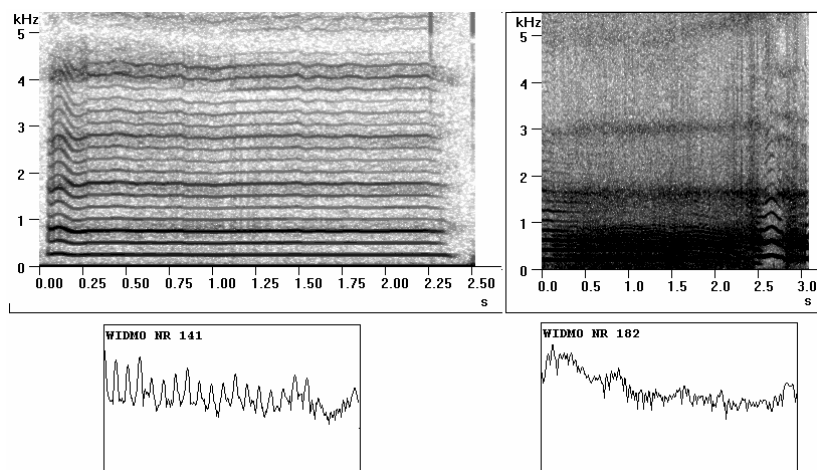
typ 1: składowe harmoniczne tonu krtaniowego są zmieszane z komponentami szumowymi w zakresie głównych formantów samogłosek,

typ 2: składowe szumowe w drugim formancie "e" oraz "i" dominują nad składowymi harmonicznymi, także pojawia się lekki dodatek składowych szumowych w zakresie ponad 3kHz,

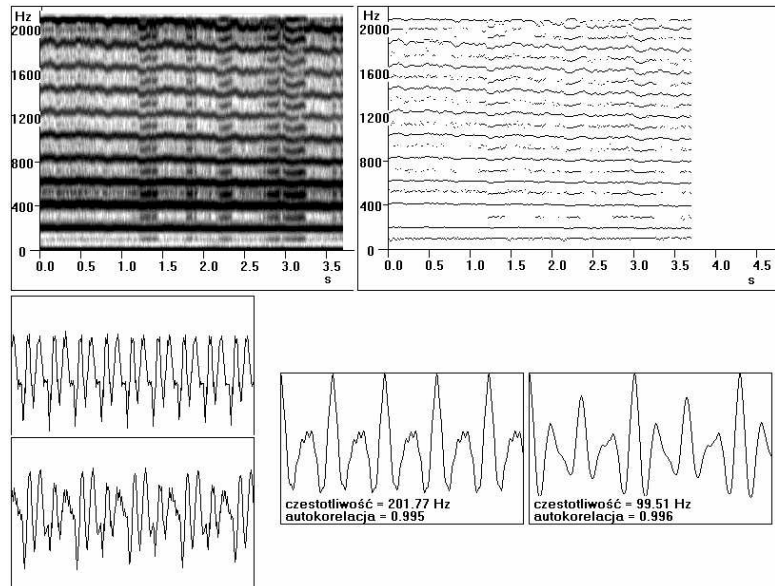
typ 3: drugi formant samogłosek "e" oraz "i" jest całkowicie przykryty przez składowe szumowe a dodatek składowych szumowych w zakresie ponad 3kHz jest widoczny w szerszym zakresie częstotliwości,

typ 4: drugi formant "a", "e", "i" jest przykryty przez składowe szumowe, nawet pierwsze formanty wszystkich samogłosek zwykle tracą swoje składowe periodyczne, na które nakładają się składowe szumowe. Jednocześnie widoczna jest większa intensywność komponentów szumowych w wysokich częstotliwościach.

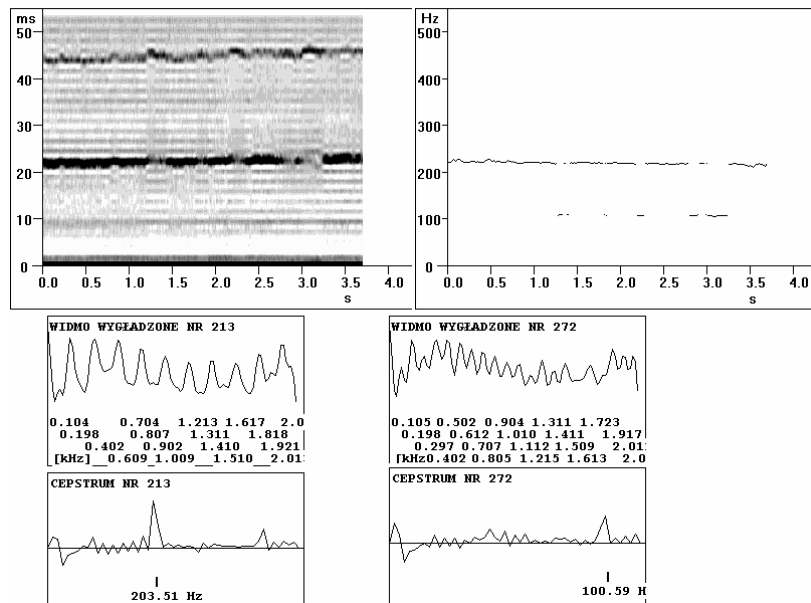
Przykładowe wyniki analiz dla tego typu schorzenia zostały pokazane na rys. 1.



Rys. 1. Porównanie wyników analizy sonograficznej dla wypowiedzi głoski "e": po lewej stronie głos zdrowy, po prawej głos obarczony chrypką, poniżej odpowiednio widma chwilowe. Widoczne są różnice w charakterach widm: widmo okresowe dla głosu zdrowego, widmo o charakterze szumowym dla głosu patologicznego.



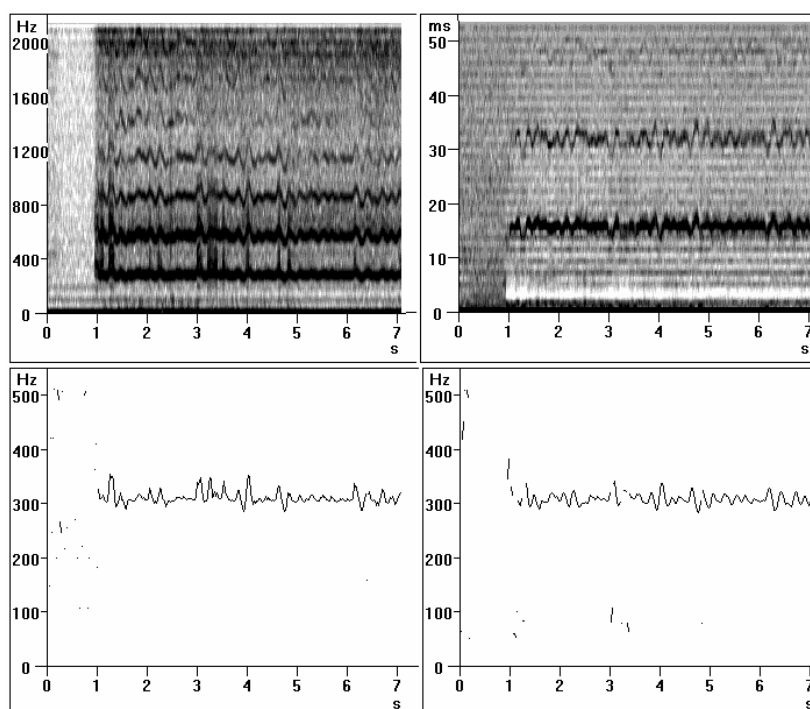
Rys. 2. Wyniki analizy sonograficznej i wykres ścieżek maksimów (powyżej) oraz wykresy przebiegów czasowych (poniżej po stronie lewej) oraz funkcji autokorelacji (po stronie prawej) dla przypadku dwugłosu (przekroje czasowe 2,5 s i 3,1 s)



Rys. 3. Wyniki analizy cepstralnej – cepstrogram (powyżej po lewej) i wykres zmienności parametru F0 (po prawej) oraz widma i cepstra chwilowe (odpowiednio poniżej) dla przypadku dwugłosu

Dwugłos i drżenie głosu to objawy nieprawidłowej artykulacji tonu krtaniowego. Dla dwugłosu przy wytwarzaniu tonu krtaniowego biorą udział nie tylko struny głosowe, ale także fałdy przedsionka. Objawia się to na wykresie zmienności parametru F0, który wykazuje przeskok o oktawę niżej. Przykładowe wyniki analiz dla tego typu schorzenia zostały pokazane na rys. 2 i 3.

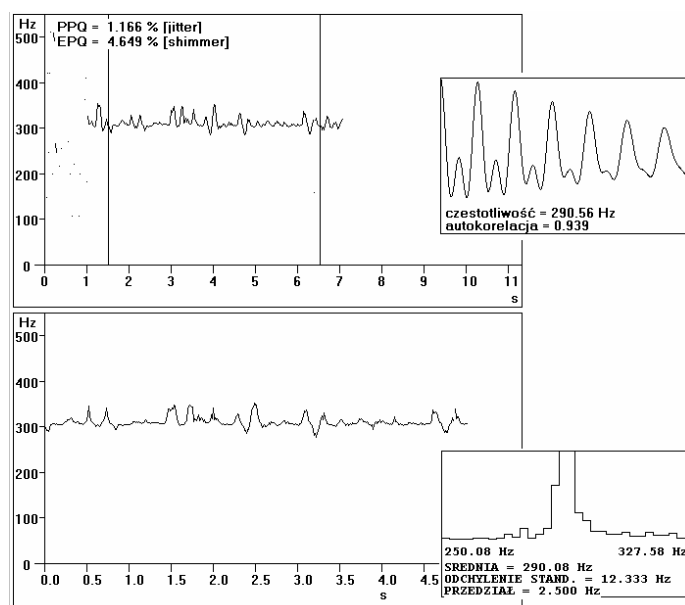
Przyczynami drżenia głosu są niedomagania mięśnia głosowego będące wynikiem nieprawidłowego napięcia, np. jako skutek zmęczenia. Obserwuje się w tym przypadku chaotyczne zmiany parametru F0 a także znajduje to odzwierciedlenie w histogramie F0. Przykładowe wyniki analiz dla tego typu schorzenia zostały pokazane na rys. 4 i 5.



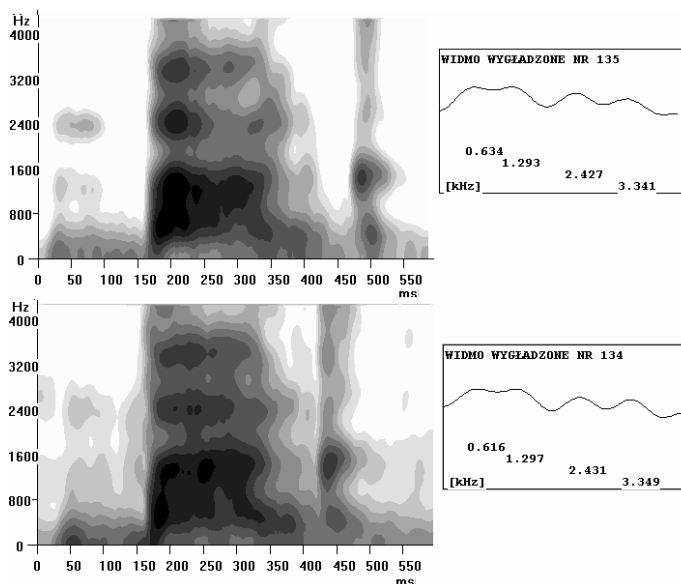
Rys. 4. Wyniki analizy sonograficznej i cepstralnej –spektrogram i cepstrogram (powyżej) oraz wykresy zmienności parametru F0 (poniżej: metoda cepstralna i autokorelacyjna) dla głosu drżącego (niestabilność F0), będącego efektem niedomagań mięśni głosowych.

Analizy akustyczne stosuje się także podczas rehabilitacji pacjentów po laryngektomii, pacjentów po zabiegach operacyjnych dotyczących języka, podniebienia miękkiego, żuchwy, policzków, warg oraz gardła i krtani oraz pacjentów z afazją. W toku rehabilitacji spotyka się różne przypadki począwszy od braku głosu i pseudoszeptu do głosu gardłowego i przelykowego, a także głosu zastępczego wykształconego poprzez chirurgiczne wytworzenie przetok głosowych i zastosowanie odpowiednich protez,

Analizy głosu są także wykonywane w celu obserwowania poprawy głosu u osób niedoświadczających korzystających z aparatu słuchowego. W tym przypadku zwraca się uwagę na parametry formantowe. Przykład takiej analizy pokazano na rys. 6.



Rys. 5. Wyniki analizy cepstralnej – wykres zmienności parametru F0 dla głosu drżącego. Podane zostały wskaźniki jitter (PPQ) i shimmer (EPQ) obliczone dla zaznaczonego odcinka czasowego. Po prawej stronie u góry wykres funkcji autokorelacji, poniżej histogram parametru F0.



Rys. 6. Wyniki analizy sonograficznej (wyraz „bank”)– spektrogramy widma wygładzone (po stronie lewej) oraz ich przekroje czasowe uwidaczniające częstotliwości formantowe. Powyżej bez aparatu słuchowego, poniżej z aparatem. Widoczne różnice można interpretować jako wpływ audytorijnego sprzężenia zwrotnego (m. in. efekt Lombarda)

4. WNIOSKI

Na podstawie uzyskanych wyników można zaplanować badania umożliwiające obiektywizację skali chrypki. Można także oczekiwać zdefiniowania skali afonii oraz drżenia głosu, co może przyczynić się do pomocy w diagnozie. Istnieje także potrzeba zdefiniowania parametrów sygnału mowy w przypadku dwugłosu. Dla wykonania powyższych zadań powinno się wykorzystać wszystkie możliwe parametry: jitter, shimmer, odchylenie standardowe dla rozkładu parametru F0 (lub T0), współczynniki lokalnych maksimów funkcji autokorelacji i innych. Istnieje także możliwość zastosowania systemów eksperckich do celów związanych z klasyfikacją głosów patologicznych.

PODZIĘKOWANIE

Wykonane prace badawcze zostały dofinansowane przez Ministerstwo Edukacji i Nauki w ramach projektu nr 3 T11E02829.

5. BIBLIOGRAFIA

1. Rogowski M., Ginzieńska E.: Czasowo-częstotliwościowa analiza wąskopasmowa wywołanych emisji otoakustycznych u noworodków, *Audiofonologia*, Tom VIII, 1996, str. 37-42, ISSN 1425-3089.
2. Grzanka A., Hatzopoulos S.: Przegląd metod czasowo-częstotliwościowych w odniesieniu do analiz sygnałów otoemisji akustycznych, *Audiofonologia*, Tom XIII, 1998, str. 41-53, ISSN 1425-3089.
3. Van Gogh C.D.L., Festen J.M., Verdonck-de Leeuw I.M., Parker A.J. et al: Acoustical analysis of tracheoesophageal voice, *Speech Communication*, Volume 47, Issue 1-2, September - May, 2005, pp. 160-168, ISSN 0167-6393.
4. Wuyts F.L., De Bodt M.S., Molenberghs G., Remacle M., Heylen L., Millet B., Van Lierde K., Raes Jan, Van de Heyning P.H.: The Dysphonia Severity Index: An Objective Measure of Vocal Quality Based on a Multiparameter Approach, *Journal of Speech, Language and Hearing Research*, No. 43, Jun 2000, pp. 796 - 809, ISSN 1092-4388.

SPEECH ANALYSIS IN PHONiatrics

The paper contains a description of methods of speech signal analysis for medical uses. They may help in larynx pathologies diagnosis such as eg. hoarseness, double voice, or tremulous voice. Algorithms for vocal tone extraction (F0 parameter and its time variation) and for spectral smoothing (formant parameters estimation or harmonic components) are presented. Some results are presented and compared with reference spectrograms analyzes. Algorithms for F0 parameter histogram, jitter and shimmer are presented. Results are compared with the autocorrelation parameters for vowels. All algorithms are implemented as a computer pathologic speech analyser. Basing on these investigations it is possible to create objective measures for various pathologies of speech organs (also in case of associated hearing disorder) designed as an expert system in speech diagnosis.

