

Jolanta Zielińska

Analiza sygnału mowy metodą elektrolotografii - zastosowania praktyczne

Analiza mowy jest procesem trudnym, dotyczy bowiem wysoce złożonego sygnału. Stąd jest ona przedmiotem zainteresowania specjalistów z wielu dziedzin naukowych, tak skrajnie odległych, jak językoznawstwo [4] i informatyka [8]. W aspekcie techniczno-fizycznym sygnałem mowy zajmuje się fonetyka akustyczna. Używając odpowiednio rozszerzonego aparatu akustyki analizuje ona sygnał mowy jako realizację dźwiękową kodu językowego. Prace z tego zakresu [5, 6, 2] przedstawiają klasyczne metody pomiarowe, których wyniki zawierają przede wszystkim opisy cech widmowych głosek polskich. Istotnym aspektem nowoczesnych badań fonetycznych jest analiza instrumentalna sygnału mowy, używająca takich urządzeń, jak: aparat rentgenowski, elektromyograf (EMG), spektrograf, minograf, Laryngograph. Jej celem jest wizualizacja sygnału mowy w sposób pozwalający na wychwycenie pewnych jego cech, a następnie przedstawienie w formie graficznej. Proces analizy wykonywany z użyciem komputera jest prostszy, szybszy i zapewnia wyższą jakość, aniżeli realizowany za pomocą innych urządzeń. Komputerowa rejestracja głosu umożliwia jego wizualizację i pozwala na obiektywność oceny i powtarzalność. Te zalety posiada metoda elektrolotografii. Umożliwia ona badanie pracy strun głosowych w sposób niezależny od wytwarzanego dźwięku na drodze pomiaru fizycznych przesunięć fałd głosowych, poprzez rejestrację drgań krtani. Jest to możliwe do uzyskania, gdyż większość dźwięków mowy ludzkiej wytwarzanych jest na drodze pobudzenia strun głosowych przez strugi powietrza pulsujące z częstotliwością, określoną przez drgania fałd głosowych mówiącego.

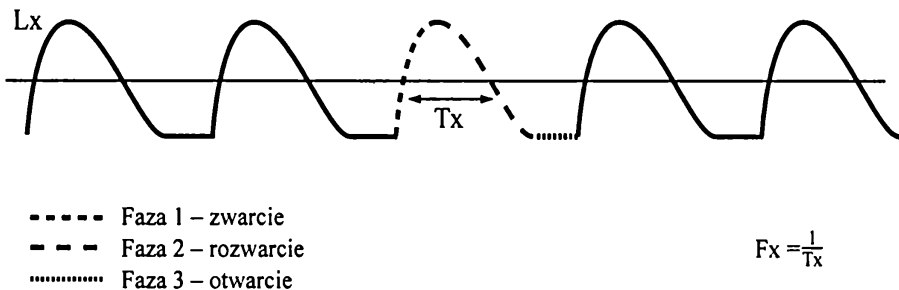
Metoda elektrogłotografii - ogólna charakterystyka

Metoda elektrogłotografii została oparta na pomiarze impedancji elektrycznej pomiędzy fałdami głosowymi podczas ich drgania w procesie tworzenia mowy dźwiękowej. Jest ona funkcją ich wzajemnego położenia i ulega zmianom podczas drgania krtani. W szczególności, jeżeli fałdy głosowe są zwarte – jest mniejsza niż w przypadku, gdy są rozwarne. Zjawisko to było wykorzystywane do monitorowania ruchów fałd głosowych w procesie normalnego mówienia już od dłuższego czasu. Jednakże negatywnymi cechami urządzeń wykorzystujących go bezpośrednio była konieczność indywidualnej regulacji dla każdego pacjenta oraz zniekształcenia pomiarów wynikające z ewentualnego przesunięcia elektrod służących do pomiaru impedancji na szyi osoby badanej. Technika stanowiąca udoskonaloną formę elektrogłotografii eliminującą wspomniane niedogodności nosi nazwę laryngograficznej i została praktycznie zastosowana w urządzeniu o nazwie Laryngograph. Eliminację błędów pomiarowych osiągnięto w nim poprzez:

- wprowadzenie elektrod z dodatkowymi otaczającymi przewodzącymi pierścieniami, zastosowanie ekranowanych połączeń oraz użycie trójbiegunowej sieci, której wyjście jest zależne od warunków elektrycznych na szyi mówiącego wyłącznie w bezpośrednim sąsiedztwie elektrod,
- zaprojektowanie obwodu przetwarzającego wyjście z elektrod w sposób samoczynnie kompensujący różne impedancje na szyi mówiącego tak, iż wyjście zależne jest tylko od szybkich zmian związanych z drganiami fałd głosowych.

Moc elektryczna rozpraszana na szyi osoby badanej wynosi około 20 mW (przy częstotliwości 1 MHz), a górną granicą częstotliwości jest około 5 kHz. W przypadku normalnego głosu męskiego stosunek końcowego sygnału wyjściowego do szumu wynosi około 40 dB. Dla małych dzieci i niemowląt względny szum jest znacznie większy, chociaż udane wyniki badań daje się uzyskać nawet dla nowo narodzonych dzieci.

Sygnał uzyskiwany za pomocą elektrogłotografii reprezentujący pracę fałd głosowych stanowi podstawę dla wyznaczenia, istotnej w badaniach fonetycznych, częstotliwości podstawowej F_x w sposób pokazany na rysunku 1.



Rys. 1. Przykładowy przebieg sygnału L_x , prezentacja sposobu wyznaczenia częstotliwości podstawowej F_x [9]

Analiza cyklu sygnału Lx pozwala wyróżnić trzy jego fazy. Pierwsza z nich to szybki wzrost sygnału, odpowiadający szybkiemu zamykaniu się fałd głosowych. Druga faza odzwierciedla nieco powolniejsze opadanie sygnału i związana jest z ich rozwieraniem. Faza trzecia o płaskim przebiegu odpowiada stanowi, w którym fałdy głosowe są otwarte. Częstotliwość Fx jest obliczana na drodze pomiaru długości każdego cyklu pracy fałd głosowych w połowie wartości amplitudy sygnału Lx (w praktyce dokonuje się pomiaru odstępu czasu pomiędzy przejściami sygnału przez ten poziom). Poszukiwana wartość Fx stanowi odwrotność długości cyklu sygnału Lx. Najczęściej przedstawia się zmiany sygnału Fx w formie logarytmicznej. Wartość tego sygnału jest ujemnym logarytmem z obserwowanej na bieżąco wartości cyklu sygnału Lx.

Praktyczne zastosowanie metody elektrolotografii – działanie przystawki komputerowej Laryngograph Processor PCLX

Przystawka Laryngograph Processor PCLX, której działanie zostało oparte na metodzie elektrolotografii jest podłączona do komputera PC przy użyciu karty interfejsu zainstalowanej bezpośrednio na magistrali komputera IBM PC, stanowiącej połączenie z Laryngographem oraz oprogramowaniem. Karta zawiera specjalne układy przetwarzania uzyskiwanych sygnałów tak, aby mogły one być wizualizowane na ekranie komputera i zapamiętywane na bieżąco. Oprogramowanie PCLX o nazwie Speech Studio pozwala na akwizycję sygnałów, ich wizualizację, analizę on-line oraz zapamiętywanie na dysku w celu późniejszej analizy i odtworzenia.

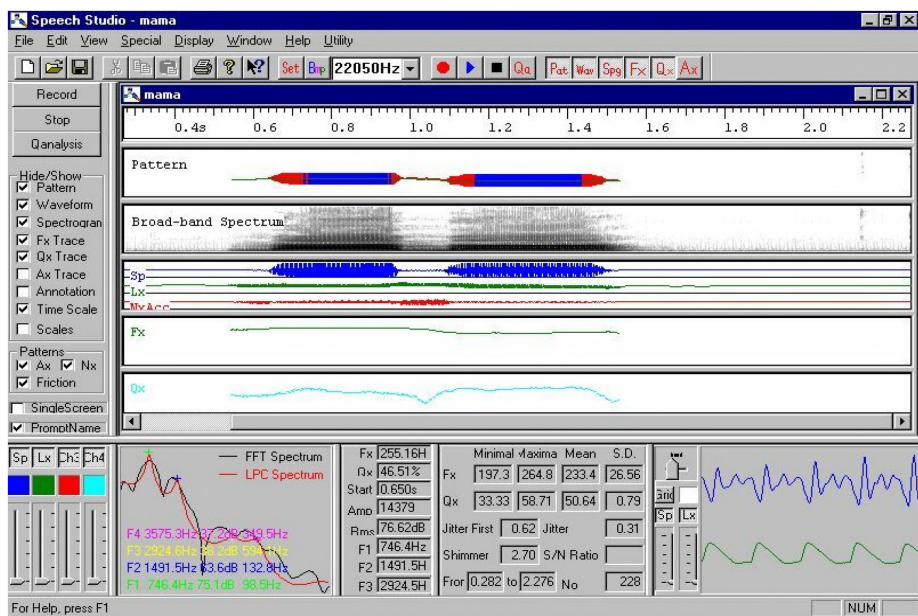
Istotnym elementem analizy sygnału mowy jest sposób pobierania i zapamiętywanych go przez komputer. Przystawka Laryngograph Processor PCLX posiada dwa detektory sygnału wejścia. Są to miedziane elektrody zakładane na szyję osoby badanej za pomocą regulowanej opaski i mikrofon, umożliwiający za pomocą kabłąka stabilne zamocowanie na głowie. Warunkiem poprawności wykonania badań jest sposób założenia elektrod na szyję osoby badanej. Zależy on od wieku diagnozowanej osoby: im młodsza, tym elektrody na szyję powinny zostać założone wyżej, przyjmując środkowe położenie pomiędzy czwartym a siódmym kręgiem szyjnym dopiero u dojrzałego pacjenta. Fakt ten wynika w bezpośredni sposób z fizjologii rozwoju krtani [7, 3]. Do 5. roku życia u wszystkich dzieci następuje szybki wzrost krtani, a potem jego zwolnienie. Kolejny okres wzrostu przypada na wiek pokwitania. Struktura krtani zmienia się, rozrasta się ona przede wszystkim u chłopców w płaszczyźnie poziomej, czemu towarzyszy wydłużenie i poszerzenie fałd głosowych (okres mutacji głosu). Obejmuje to wiek około 13.–15. roku życia, niekiedy przedłużając się do 19. roku życia. W okresie tym szyja wydłuża się a krtani przemieszcza ku dołowi. W tym wieku należy zwracać szczególną uwagę na fizjologicznie poprawne przypięcie elektrod Laryngographu. Drugi detektor wejściowy, czyli mikrofon, powinien się znajdować około 2,5 cm od ust osoby badanej.

Prototyp opisywanego urządzenia został opracowany po raz pierwszy na Wydziale Fonetyki University College w Londynie [1]. Obecnie jest ono produkowane w firmie Laryngograph Ltd. W ciągu kilku lat od wynalezienia urządzenie to znalazło zastosowanie w takich dziedzinach, jak:

- terapia mowy i logopedia, w celu doskonalenia procesu mowy,
- badania kliniczne nosa i gardła, do diagnozowania pracy fałd głosowych po zabiegach operacyjnych,
- doskonalenie głosu przez śpiewaków i spikerów,
- badania relacji pomiędzy mową a pracą traktu głosowego,
- nauczanie języków obcych.

Możliwości analizy sygnału mowy z użyciem przystawki komputerowej Laryngograph Processor PCLX

W procesie wytwarzania głosu można wyróżnić trzy zasadnicze etapy, które są w pełni monitorowane i analizowane przez oprogramowanie stanowiska komputerowego o nazwie Speech Studio. Należy do nich generacja drgania pobudzającego, jego modyfikacja w rezonatorze głosowym oraz wypromieniowanie [5, 6]. Sygnał wytwarzany w krtani podczas wypowiadania głosek dźwięcznych, przez struny głosowe pełniące rolę generatora pobudzającego, ma postać przebiegu okresowego, charakteryzowanego przez częstotliwość podstawową F_x , której sposób pozyskiwania został wcześniej opisany (rys. 1). W przewężeniach kanału głosowego sygnał jest modulowany w zależności od kształtu, jaki przyjmuje przestrzeń od krtani do ust podczas artykulacji. Ustawienie traktu głosowego, czyli konfiguracja artykulatorów jest charakteryzowana przez formanty, określane na podstawie średniego widma amplitudowo-częstotliwościowego danej wypowiedzi (rys. 3). Ruchy odpowiadające artykulacji danego fonemu zależą od fonemów z nim sąsiadujących. Podczas wymawiania danego dźwięku przygotowany jest stan traktu głosowego do artykulacji następnego fonemu. Zjawisko to nosi nazwę koartykulacji. Jego analiza jest bardzo istotna przy określaniu przyczyn nieprawidłowości artykulacyjnych osób mających problemy z motoryką oralną. Opisane stany przejściowe są monitorowane i określane na podstawie wartości częstotliwości i amplitudy formantów odczytywanych z chwilowych widm amplitudowo-częstotliwościowych. Ostatni etap procesu wytwarzania mowy to wytworzenie strumienia powietrza rozchodzącego się w postaci fali akustycznej. Jest ona charakteryzowana przez ciśnienie akustyczne, którego zmiany w czasie stanowią istotną informację o wytwarzanym sygnale mowy i podstawę jego dalszej analizy. Rysunek 2 prezentuje ekran główny zintegrowanego systemu badania sygnału mowy o nazwie Speech Studio, posiadającego wszystkie wymienione możliwości analizy sygnału mowy. Podczas pracy z programem dostępne są dwa paski narzędziowe (górny i boczny), pozwalające na wykonywanie szeregu operacji, od akwizycji sygnału, jego zapamiętania, odtwarzania, analizy, po wybór ilości i rodzaju widocznych przebiegów i skal.



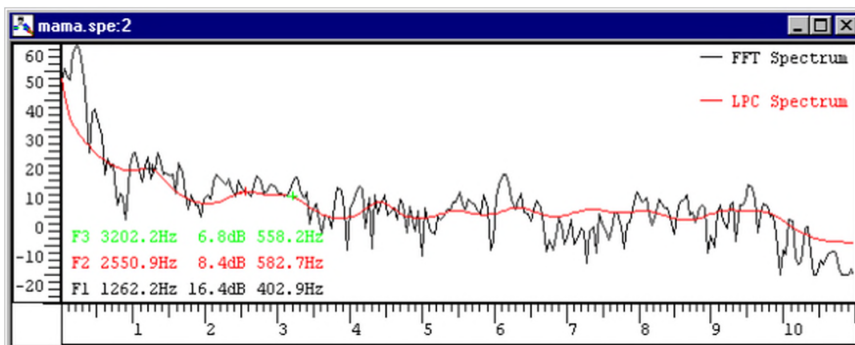
Rys. 2. Zintegrowany system badania sygnału mowy Speech Studio [9]

Zasadnicza część prezentowanego ekranu to siedem przebiegów charakteryzujących sygnał mowy. Dwa najistotniejsze diagnostycznie przebiegi, pokazujące zmiany amplitudy wypowiedzi w czasie oraz pracę fałd głosowych, widoczne są w powiększeniu w prawym dolnym oknie ekranu. Pozwalają one między innymi na natychmiastowe, wstępne ustalenie patologii głosu oraz stosowne do tego ustawienie parametrów aparatury badawczej. Wzmocnienie sygnałów można regulować za pomocą pokręteł przystawek lub suwaków umieszczonych w lewym dolnym rogu ekranu, poprzez odpowiednie ich przesunięcie. W dolnej części ekranu pomiędzy omówionymi oknami znajdują się jeszcze dwa okna. W jednym wyświetlane są średnie widma amplitudowo-częstotliwościowe, wyliczone przez program Speech Studio, z użyciem algorytmów FFT (szybkiej transformaty Fouriera) i LPC (algorytmu predykcji liniowej). Pozwalają one na określenie częstotliwości formantów w zaznaczonym przez prowadzącego badanie przedziale czasu. W drugim oknie znajdują się wartości charakteryzujące profil mowy osoby badanej. W ramach tego profilu zostaje określona wartość minimalna, maksymalna, średnia i odchylenie standardowe dla częstotliwości podstawowej Fx oraz wskaźnika zwarcia głosi Qx (WZG), a także parametry opisujące zaburzenia częstotliwości (ang. jitter) oraz amplitudy (ang. shimmer). Przebieg wskaźnika zwarcia głosi Qx zależy w sposób bezpośredni od przebiegu Lx i liczbowo podawany jest w procentach. Oznacza on procentowy udział czasu, w którym fałdy głosowe nie są otwarte, czyli czas ich zwarcia i rozzwarcia (suma czasu trwania fazy 1 i fazy 2 pokazanej na rys. 1) podczas artykułowania wypowiedzi.

Podczas tego typu badania sygnału mowy na ekranie komputera znajduje się jednocześnie wiele charakteryzujących go przebiegów. Można z nich określić zarówno parametry liczbowe cechujące badany głos, poprzez dokonanie stosownej analizy, jak i wyprowadzić wnioski jakościowe. Program umożliwia redukcję ilości przebiegów do wybranych przez prowadzącego badanie.

Uzyskiwany wynik pomiarowy zależy nie tylko od ruchów fałd głosowych, ale również od wielkości krtani oraz masy drgających mięśni. Sygnał pokazujący na ekranie komputera zmiany ciśnienia fali akustycznej w czasie oznaczony został na rysunkach symbolem Sp. Jest to oscylogram, z przebiegu którego można wyprowadzić wstępne informacje o patologii strukturalno-funkcjonalnej krtani. Im sygnał ten jest bardziej poszarpany i ma mniejszą regularność, tym większe zmiany fizjologiczne w obrębie diagnozowanego narządu. Z tego oscylogramu, poprzez analizę widocznych zmian amplitudy fali głosowej w czasie, można określić jej podstawowe własności akustyczne, w tym również dźwięczność i nosowość. W sygnale czasowym właściwości źródła i kształtującego dźwięk kanału głosowego są ze sobą powiązane. Występuje w nim równoprawność tworzących go składników przypadkowych i zeterminowanych, połączenie aspektu osobniczego i semantycznego [8]. Stąd przebiegi czasowe nie mogą być wykorzystywane do procesu automatycznego rozpoznawania mowy (ARM), natomiast są szeroko używane w badaniach nad akustycznymi cechami mowy, zwłaszcza uzyskiwana z nich obwiednia amplitudowa sygnału. Stanowi to szeroko przebadane zagadnienie opisane w wielu pracach z zakresu fonetyki [5, 6, 2]. Na rysunku 2 jako drugi przebieg prezentowany jest spektrograf dynamiczny zwany sonografem, będący szerokopasmowym widmem, służącym do akustycznej oceny normalnej oraz patologicznej fonacji, stanowiącej jeden z parametrów fonetycznych mowy w płaszczyźnie segmentalnej. Pozwala on na zróżnicowanie opozycji dźwięczność – bezdźwięczność oraz ocenę elementów prozodycznych wypowiedzi, głównie intonacji. Przebieg pokazuje zmiany częstotliwości w czasie oraz poziom, czyli amplitudę sygnału wyróżnioną przez stopień zaciemnienia. Taka reprezentacja sygnału mowy znalazła najszerokie zastosowanie w badaniach fonetycznych, w tym ARM, gdyż daje podstawy do wyznaczenia tak istotnej dla tych badań wartości częstotliwości podstawowej i formantów. Ich związek z oceną patologii czynności krtani, intonacją, pracą traktu głosowego, czy prawidłowością artykulacji jest dogłębnie przebadany i szeroko reprezentowany w literaturze przedmiotowej.

Kolejny przebieg widoczny na ekranie komputera (rys. 2) o nazwie Sp umożliwia dynamiczną wizualizację przebiegów czasowych wypowiedzi, określenie stopnia jej napięcia, ocenę głośności wypowiedzi. Natomiast odpowiednio wykonana analiza częstotliwości podstawowej F_x informuje o wszystkich elementach prawidłowej wypowiedzi. Są to: oddychanie dynamiczne, ekonomiczne gospodarowanie oddechem, praca fałd głosowych, wysokość głosu, elementy prozodyczne wypowiedzi, takie jak: tempo, rytm i akcent oraz dodatkowo ukazanie stopnia dźwięczności głosek (przy braku dźwięczności brak obrazu). W ocenie prawidłowej artykulacji, głównie kształtu kanału głosowego od krtani do ust, gdzie dźwięk jest modulowany, pomocne jest widmo amplitudowo-częstotliwościowe wypowiedzi, zaprezentowane na rysunku 3.



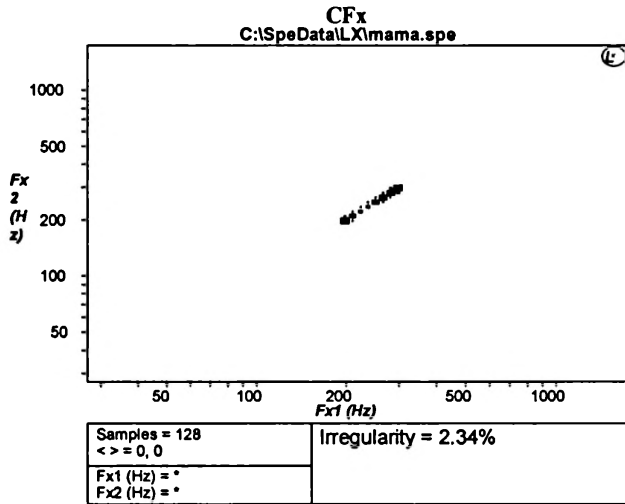
Rys. 3. Średnie widmo amplitudowo-częstotliwościowe wypowiedzi „mama”, program Speech Studio [9]

Prezentowane widmo sygnału mowy uzyskane jest poprzez odpowiednie procedury komputerowe, wykorzystujące algorytm FFT (szybkiej transformaty Fouriera) i LPC (algorytm predykcji liniowej). Pozwala ono na określenie formantów w zaznaczonym przez prowadzącego badanie przedziale czasu, widmo chwilowe w pewnym punkcie czasowym wypowiedzi. Dla każdego formantu podana zostaje wartość jego maksymalnej częstotliwości, amplituda oraz pasmo zajmowanych częstotliwości.

Komputerowa analiza materiału badawczego

Możliwości analizy uzyskanego materiału badawczego są bardzo różnorodne, od obliczeń o charakterze statystycznym, parametrycznym, do obróbki graficznej. W celu uzyskania dokładniejszego obrazu sygnału można wybrać dowolny jego fragment, a następnie prawie dowolnie powiększyć, wychwytyjąc nawet najmniejsze zmiany patologiczne. Oprogramowanie umożliwia także pomiar odstępu czasu pomiędzy poszczególnymi punktami sygnału. Dokonuje się go przez odpowiednie ustawienie na ekranie dwu dostępnych kursorów, pomiędzy którymi mierzony jest przedział czasu.

Analiza sygnału F_x polega na wyznaczeniu 23 histogramów, w tym dwuwymiarowego histogramu C_{F_x} pokazującego nieregularność pracy strun głosowych podczas wytwarzania sygnału mowy (rys. 4).



Rys. 4. Przykład histogramu CFx, pokazującego nieregularność pracy strun głosowych [9]

Komputerowa analiza sygnału mowy pozwala na uzyskanie profilu głosu osoby badanej w postaci parametrów akustycznych (rys. 5).

| Report – Vocal Profile Analysis | | | |
|---------------------------------|-----------|---------------|-----------|
| No. Samples | 8 | Duration | 0.714 s |
| Minimal Fx | 194.43 Hz | Maximal Fx | 211.95 Hz |
| Average Fx | 203.16 Hz | S.D. Fx | 5.08 Hz |
| Minimal Qx | 36.28 % | Maximal Qx | 47.27 % |
| Average Qx | 42.47 % | S.D. Qx | 2.19 % |
| Jitter First | 2.16 % | Jitter Second | 0.64 % |
| Shimmer+ | 24.34 % | Shimmer- | -15.69 % |

Rys. 5. Przykładowy profil głosu, wypowiedź „mama”, program Speech Studio [9]

Profil głosu zawiera średnią wartość częstotliwości podstawowej Fx, jej wartość minimalną i maksymalną, pozwalające na określenie wahań częstotliwości, odchylenie standardowe, współczynnik zaburzenia częstotliwości o nazwie Jitter, podaną w procentach średnią wartość współczynnika zwarcia głośni Qx, jego wartość minimalną i maksymalną oraz stosowne odchylenie standardowe. Dodatkowo, jako ostatnia informacja podawana jest wartość współczynnika zaburzenia amplitudy głosu, o nazwie Shimmer.

Podsumowanie

Zastosowanie metody elektrolotografii pozwala na dogłębną analizę sygnału mowy, zapamiętanego w pamięci komputera. Może to być przykładowo ogólna analiza widmowa lub na jej podstawie szczegółowa analiza wartości formantów. Można z przebiegu częstotliwości podstawowej F_x na ekranie komputera jedynie ocenić wysokość głosu, ale można też poprzez stosowną analizę statystyczną otrzymać pełny, akustyczny profil głosu badanego osobnika. Prowadzi to do wniosku ogólnego, że tak jak złożony jest sygnał mowy, tak złożone i szczegółowe są metody jego badania. Technika musi podążyć tym śladem. Prezentowana metoda elektrolotografii stanowi tego dowód. To, jak i w jakim zakresie zostanie ona wykorzystane zależy w dużej mierze od przyjętego celu badawczego i kompetencji osoby prowadzącej badania. Pozostaje niezaprzeczalnym faktem, że zaprezentowane możliwości, w procesie badania sygnału mowy są bardzo szerokie, a co najważniejsze stosunkowo proste w uzyskaniu i interpretacji wyników pomiarowych

Bibliografia

- [1] Abberton E., Howard D., Fourcin A., *Laryngograph assessment of normal voice: a tutorial*, *Clinical Linguistic and Fonetics* 3, 1996, 243–259
- [2] Basztura C., *Źródła, sygnały i obrazy akustyczne*, Warszawa 1988
- [3] Eberhardt G., *Zaburzenia głosu u dzieci w wieku rozwojowym*, [w:] *Zaburzenia głosu – badanie – diagnozowanie – metody usprawniania*, red. H. Mierzejewska, M. Przybysz-Piwkowska, Warszawa 1998
- [4] Grabias S., *Logopedyczna klasyfikacja zaburzeń mowy*, „*Audiofonologia*”, t. VI, 1994, s. 7–22
- [5] Jassem W., *Podstawy fonetyki akustycznej*, Warszawa 1973
- [6] Jassem W., *Mowa a nauka o łączności*, Warszawa 1974
- [7] Pruszewicz A., *Foniatria kliniczna*, Warszawa 1992
- [8] Tadeusiewicz R., *Sygnał mowy*, Warszawa 1988
- [9] Zielińska J., *Diagnoza i terapia sprawności ortofonicznej dzieci z uszkodzeniem słuchu wspomaganą techniką komputerową*, Kraków 2004

Practical applications of electroglottographic speech signal analysis

Abstract

This paper presents the essence of speech signal analysis performed with use of electroglottography. It is an instrumental analysis, which enables presentation of practical possibilities of using Laryngograph in the discussed problem area. Advantages of this device are clearly described. The main point is that visualization of speech signal with use of Laryngograph makes it possible to pick up and recognize its main acoustic and phonetic characteristics and to present results in the form of graphic charts. The analysis process performed with the use of an add-on computer device is much easier and faster than other methods and ensures higher quality. Computer voice registration enables not only visualization of its properties, but also objective valuation and its repeatability. This paper also presents issues connected with practical advantages of integrated voice diagnosis system Speech Studio.

Key words: speech signal, electroglottography, visualization, computer, laryngograph