Magdalena Igras[1], **Bartosz ZIÓŁKO**[2]
AGH AKADEMIA GÓRNICZO-HUTNICZA W KRAKOWIE, KATEDRA ELEKTRONIKI
{[1]migras@agh.edu.pl, [2]bziolko@ agh.edu.pl }

Ekstrakcja i kwantyfikacja cech sygnału oddechowego z sygnału mowy[1]

Streszczenie

Artykuł zawiera charakterystykę oddechów w sygnale audio oraz opracowanie statystyczne ich cech akustycznych oraz częstości występowania. Zaproponowano algorytm automatycznej detekcji oddechów. Proponowana metoda w pierwszej fazie wykorzystuje cechy prozodyczne (czas trwania i energię sygnału), w drugiej weryfikuje je poprzez sprawdzenie kryterium wartości formantów pierwszego (*F1*) i drugiego (*F2*) oraz ich odchyleń standardowych. Testy wykazały metody skuteczność na poziomie: pewność 80-90%, precyzja 45%. Dzięki prostocie algorytmu metoda wykazuje małą złożoność obliczeniową i dobrą wydajność czasową działania. Zaprezentowano potencjalne zastosowania metody w medycynie.

Abstract

An algorithm for automatic detection of breath events in speech signal is suggested in this paper. The issues of breath events occurrences in recordings are discussed as well as their statistical parameters. In the beginning of the detection procedure, preliminary hypotheses of breath are indicated in the analyzed speech signal using temporal features (duration and energy). Then, values of formants F1 and F2 (with their standard deviations) are calculated to establish final recognition. The method achieves recall 80-90% and precision 45% and needs to be improved in further works. Thank to its simplicity, the algorithm performance time is very good. Potential applications of the method in medicine are presented.

Słowa kluczowe: detekcja oddechu, sygnał mowy.

Keywords: breath detection, speech signal

Title: Modelling and detection of breath in acoustic signal

1. Wprowadzenie

Średnia fizjologiczna częstość oddechu wynosi 12-20/min u osób dorosłych. Stanami patologicznymi są zwiększenie tempa oddychania powyżej 35 oddechów/min lub zmniejszenia poniżej 8 oddechów na minutę [1, 2]. Do pomiaru tempa oddechu stosowane są m.in. metoda akcelerometryczna, pulsoksymetryczna, lub wizyjna. Pomiary cyklu oddychania podczas mowy były badane m.in. przy wykorzystaniu sygnału pletyzmograficznego klatki piersiowej [3]. Niniejsza praca skupia się automatyzacji pomiaru wystąpień zjawiska oddychania (wdechów) w sygnale mowy, w celu dokonania analizy częstości oddychania. Proponowane badania pozwolą w przyszłości identyfikować inne czynniki fizjologiczne towarzyszące oddychaniu, np. głębokość oddechu czy występowanie szmerów oddechowych.

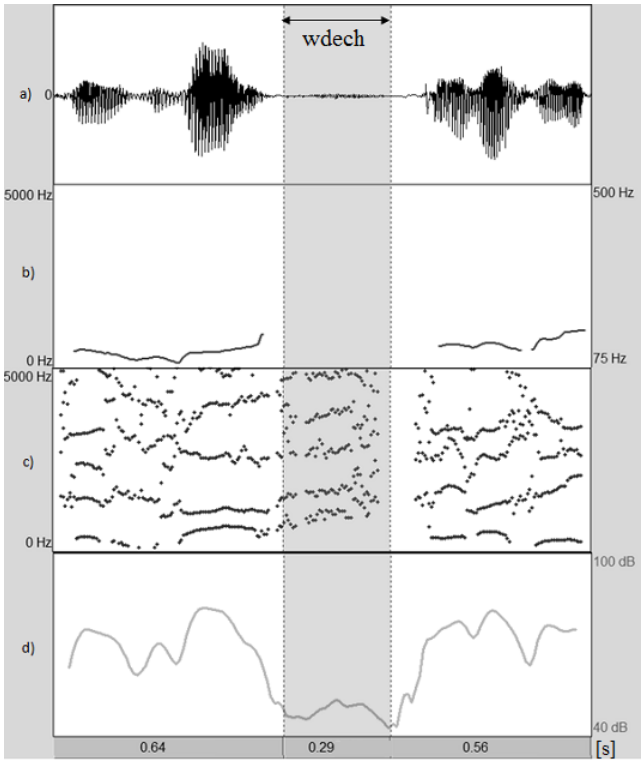
2. Zbiór danych

Do celów pracy wyselekcjonowano zbiór nagrań sygnału mowy zawierających słyszalne oddechy, łącznie 30 min nagrań 12 mówców (format .wav 16 bit PCM, 44100 Hz). Nagrania częściowo zawierają mowę spontaniczną, a częściowo czytaną. W każdym nagraniu zjawiska słyszalnych oddechów zostały poddane anotacji czasowej za pomocą narzędzia *Anotator* [4], w standardzie HTK 3.0 do formatu .mlf (*Master Label File*). Nagrania podzielono na zbiór treningowy (15 min) oraz zbiór testowy (15 min).

3. Analiza akustyczna i algorytm detekcji

Zjawisko oddechu rejestrowane w nagraniach akustycznych sygnału mowy jest słyszalne przy wysokiej jakości nagrania. Regularność oddychania podczas produkcji mowy jest zdeterminowana synchronizacją procesu oddychania z intencjonalną emisją głosu [5]. Częstotliwość wdechów dla nagrań ze zbioru treningowego wyniosła średnio 11/min. Dla zbioru treningowego zmierzono czasy trwania odnotowanych oddechów (średnia: 366 ms). Iloczas wdechów przyjmuje wartości z zakresu 120 - 820 ms (dla porównania - średnie czasy trwania polskich fonemów wynoszą od 70 do 170 ms [6]). Przykładowy przebieg sygnału zawierającego oddech pokazano na rys. 1 a). Rysunki 1 b)-d) ilustrują przebiegi zmienności następujących cech sygnału *x(n):* częstotliwości podstawowej (krtaniowej) *F0* (rys. 1b), formantów - maksimów obwiedni widma sygnału (rys. 1c) oraz energii (rys 1 d). W miejsca wystąpienia wdechu obserwuje się brak częstotliwości krtaniowej, natomiast przebieg formantów charakteryzuje się stosunkowo dużymi odchyleniami standardowymi. Średnie wartości formantów *F1* i *F2* oraz ich odchyleń standardowych, zmierzone dla zbioru treningowego, wynoszą: *F1* - 1004 Hz (odchylenie standardowe 339 Hz), *F2* - 2028 Hz (odchylenie standardowe 318 Hz).

Dotychczas najczęściej stosowanymi metodami detekcji oddechu w sygnale akustycznym były MFCC oraz LPC [7, 8]. Większość z tych rozwiązań wykorzystuje zarówno cechy czasowe, jak i spektralne sygnału mowy.



Rys. 1. Charakterystyka przykładu sygnału mowy zawierającego oddech: a) przebieg sygnału, b) przebieg zmian częstotliwości podstawowej F0, c) przebieg formantów d) przebieg energii sygnału

Proponowany algorytm jest dwustopniowy. W fazie wstępnej każde nagranie zawierające sygnał mowy zostaje poddane normalizacji amplitudy względem średniej energii sygnału. Następnie sygnał jest analizowany w obrębie ramek o długości 20 ms z zakładką 10 ms (wartości dobrane eksperymentalnie) i wskazane zostają regiony sygnału spełniające kryterium czasu i energii (utrzymywanie się lokalnej energii sygnału na poziomie 0.05 - 0.4 maksymalnej amplitudy sygnału przez czas dłuższy niż 150 ms. W drugiej fazie detekcji dla wskazanych fragmentów sygnału wyznaczane są wartości *F0*, *F1* oraz *F2* (wraz z odchyleniami standardowymi formantów *F1* i *F2*). Kryteria klasyfikacji jako oddech to brak *F0*, *F1* w granicach od 490 do 1600 Hz, odchylenie standardowe *F1* w granicach 200 - 600 Hz, *F2* w granicach od 1500 do 2500 Hz, odchylenie standardowe *F2* w granicach 200 - 600 Hz. Metodę zaimplementowano z wykorzystaniem środowiska Matlab oraz programu Praat. [9]

4. Wyniki

Działanie opisanego algorytmu przetestowano dla zbioru testowego nagrań. Zliczono poprawne rozpoznania momentów wystąpienia oddechów (*tp*, ang. *true positives*), fałszywe rozpoznania (*fp*, ang. *false positives*) w momentach, w których nie stwierdzono oddechu oraz braki rozpoznań w momentach, w których oddech wystąpił (*fn*, ang. *false negatives*). Na podstawie ich wartości obliczono współczynniki: pewność (ang. *recall*) $r=tp/(tp+fn)$, precyzja (ang. *precision*) $p=tp/(tp+fp)$ oraz F-miara (ang. *F-measure*) $f=2pr/(p+r)$. Otrzymane wyniki dla całego zestawu testowego to: pewność 83,4 %, precyzja 44,7%. Metoda umożliwia uzyskanie wysokiego stopnia prawidłowych rozpoznań, jednak na niezadowalającym poziomie kształtuje się wskaźnik precyzji. Z jednej strony udoskonalenia wymaga sama metoda (zarówno w fazie parametryzacji, jak i klasyfikacji). Z drugiej strony, weryfikacji musi zostać poddany sposób anotowania nagrań, który w niniejszym teście opierał się na subiektywnej ocenie. Na etapie projektowania metody detekcji oddechu z sygnału mowy, weryfikację i zwiększenie obiektywności anotacji można uzyskać poprzez zastosowanie niezależnej metody pomiaru oddechu podczas wykonywania nagrania.

5. Zastosowania

Algorytm detekcji oddechów w zapisie audio wykazuje potencjał dla zastosowań medycznych jako nieinwazyjna, tania i prosta w obsłudze metoda pomiaru częstotliwości oddechu dla pomiaru i monitorowania procesów fizjologicznych związanych z czynnością respiracyjną. Częstotliwość oddechu jest również skorelowana ze stopniem intensywności reakcji emocjonalnej. Dla technologii mowy detekcja oddechu ma znaczenie m.in. dla poprawienia efektywności działania systemów automatycznego rozpoznawania mowy (m.in. wspomaganie automatycznej detekcji interpunkcji w transkryptach języka mówionego).

6. Literatura

[1] Konturek S.: Fizjologia człowieka. Podręcznik dla studentów medycyny, *Elsevier Urban & Partner*, 2007

[2] Ratan. V: Handbook of Human Physiology, *Jaypee* 1993

[3] Li C., Parham D.F., Ding Y.: Cycle detection in speech breathing signals, *Proc. BSEC 2011*, Knoxville, Tennessee

[4] Ziółko B., Miga B., Jadczyk T.: Semisupervised production of speech corpora using existing recordings, *International 24. Seminar on Speech Production (ISSP'11)* , Montreal, 2011

[5] Pawłowski Z.: Emisja głosu - struktura, funkcja, diagnostyka, pedagogizacja, *Wydawnictwo Salezjańskie*, Warszawa, 2008.

[6] Ziółko B., Ziółko M.: Time durations of phonemes in Polish language for speech and speaker recognition, *Human language technology : challenges for computer science and linguistics : 4th Language and Technology Conference, LTC 2009* : Poznań, Nov. 6–8, 2009.

[7] Ruinskiy D., Lavner Y.: An effective algorithm for automatic detection and exact demarcation of breath sounds in speech and song signals, *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 838–850, 2007.

[8] Nakano T., Ogata J., Goto M., Hiraga Y.: Analysis and automatic detection of breath sounds in unaccompanied singing voice, *Proceedings of the 10th International Conference on Music Perception and Cognition*, pp. 387–390, 2008.

[9] Boersma, P.: Praat, a system for doing phonetics by computer. *Glot International 5:9/10*, 2001, 341-345.

[1] Wykonano w ramach projektu nr: 18.18.230.009 (NCN)