Magdalena IGRAS, Wiesław WSZOŁEK
AGH AKADEMIA GÓRNICZO-HUTNICZA IM.STANISŁAWA STASZICA W KRAKOWIE,
Al. Mickiewicza 30, 30-059 KRAKÓW

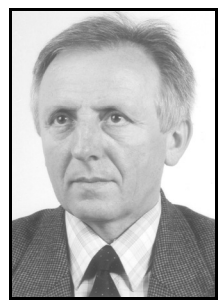
Pomiary parametrów akustycznych mowy emocjonalnej – krok ku modelowaniu wokalnej ekspresji emocji

Mgr inż. Magdalena IGRAS

Absolwentka Inżynierii Biomedycznej na specjalności Informatyka i Elektronika Medyczna w Akademii Górniczo-Hutniczej. Od 2011 r. doktorantka Akademii Górniczo-Hutniczej na specjalności Biocybernetyka i Inżynieria Biomedyczna, w Zespole Przetwarzania Sygnałów. Zajmuje się badaniem zawartości afektywno-kognitywnej sygnału mowy.

e-mail: migras@agh.edu.pl

Dr inż. Wiesław WSZOŁEK



Absolwent Wydziału Elektrotechniki, Automatyki, Informatyki i Elektroniki Akademii Górniczo-Hutniczej. Adiunkt w Katedrze Mechaniki i Wibroakustyki wydziału Inżynierii Mechanicznej i Robotyki AGH. Działalność naukowa w zakresie modelowania procesów i analizy sygnałów wibroakustycznych. Stosowanie tych metod diagnostyce technicznej i medycznej. Analiza przydatności metod sztucznej inteligencji w klasyfikacji i rozpoznawaniu mowy zdeformowanej, a także innych zdarzeń akustycznych.

e-mail: wieslaw.wszolek@agh.edu.pl

Streszczenie

Niniejsza praca podejmuje próbę pomiaru cech sygnału mowy skorelownych z jego zawartością emocjonalną (na przykładzie emocji podstawowych). Zaprezentowano korpus mowy zaprojektowany tak, by umożliwić różnicową analizę niezależną od mówcy i treści oraz przeprowadzono testy mające na celu ocenę jego przydatności do automatyzacji wykrywania emocji w mowie. Zaproponowano robocze profile wokalne emocji. Artykuł prezentuje również propozycje aplikacji medycznych opartych na pomiarach emocji w głosie.

Słowa kluczowe: rozpoznawanie emocji, wokalne korelaty emocji, przetwarzanie sygnału mowy.

Measurements of emotional speech acoustic parameters – a step towards vocal emotion expression modelling

Abstract

The paper presents an approach to creating new measures of emotional content of speech signals. The results of this project constitute the basis for further research in this field. For analysis of differences of the basic emotional states independently of a speaker and semantic content, a corpus of acted emotional speech was designed and recorded. The alternative methods for emotional speech signal acquisition are presented and discussed (Section 2). Preliminary tests were performed to evaluate the corpus applicability to automatic emotion recognition. On the stage of recording labeling, human perceptual tests were applied (using recordings with and without semantic content). The results are presented in the form of the confusion table (Tabs. 1 and 2). The further signal processing: parametrisation and feature extraction techniques (Section 3) allowed extracting a set of features characteristic for each emotion, and led to developing preliminary vocal emotion profiles (sets of acoustic features characteristic for each of basic emotions) – an example is presented in Tab. 3. Using selected feature vectors, the methods for automatic classification (k nearest neighbours and self organizing neural network) were tested. Section 4 contains the conclusions: analysis of variables associated with vocal expression of emotions and challenges in further development. The paper also discusses use of the results of this kind of research for medical applications (Section 5).

Keywords: emotion recognition, vocal correlates of emotions, speech signal processing.

1. Wprowadzenie

Emocje pełnią kluczową rolę w interakcjach międzyludzkich. Jako że głos to jeden z naturalnych, spontanicznych środków ekspresji emocji, sygnał mowy może posłużyć do skutecznej detekcji i identyfikacji stanów emocjonalnych mówcy.

Problematyka badania emocji w głosie jest zagadnieniem interdyscyplinarnym, które angażuje nauki humanistyczne, medyczne i techniczne. Wraz z rozwojem technologii mowy (systemy automatycznego rozpoznawania mowy i automatycznego rozpoznawania mówcy) podejmowane są próby opracowania systemów automatycznie rozpoznających emocje w głosie mówcy. Pierwszym krokiem ku temu jest opracowanie technicznego opisu cech sygnału akustycznego znamiennych dla ekspresji poszczególnych emocji, a kolejnym – konieczność uniezależnienia metody pomiaru takich cech od mówcy i treści wypowiedzi. Pośród licznych opracowań dla innych języków [1, 6], badania takie w polskiej nauce wciąż należą do rzadkości [3]. Większość badań stanów afektywnych w głosie skupia się na emocjach uważanych za podstawowe: smutek, radość, złość, strach, zdziwienie [8, 9].

Dotychczasowe badania dotyczące cech sygnału skorelowanych ze stanami emocjonalnymi mówcy wskazują na następujące grupy cech: związane z energią sygnału i parametrami opisującymi zmiany energii (minimum, maksimum, zakres, średnia, wariancja i in.), częstotliwości podstawowej (tonu krtaniowego – F0) i parametrów przebiegu F0 i jej pochodnej (minimum, maksimum, zakres, średnia, wariancja i in.), współczynniki MFCC, cechy związane z czasem (ilość i długość pauz, tempo wypowiedzi). Cechy związane z intonacją, akcentem i iloczasem wypowiedzi określane są mianem cech prozodycznych. [1, 20, 22]

Jako klasyfikatory najczęściej stosowane są metody minimalnoodległościowe, ukryte modele Markowa i klasyfikator Bayesa, sztuczne sieci neuronowe (zarówno wielowarstwowe perceptrony, jak i samoorganizujące się sieci neuronowe) oraz drzewa decyzyjne. Średnia skuteczność rozpoznawania emocji dla takich metod sięga nawet 80%, przy czym wyniki znacznie różnią się dla poszczególnych emocji [1, 10-13, 18, 22]. Dla porównania, skuteczność rozpoznawania emocji w głosie przez człowieka wynosi ok. 50-70% [8, 19].

2. Materiał akustyczny

Uzyskanie nagrań mowy emocjonalnej stanowiących dobry materiał badawczy okazuje się nie lada wyzwaniem. Najczęściej stowanymi metodami są: pozyskiwanie nagrań emocji odgrywanych z udziałem aktorów, wywoływanie emocji spontanicznych lub pozyskiwanie baz np. z infolinii (call center) lub nagrań radiowych i telewizyjnych. [1, 17-18] Kryterium autentyczności emocji w prowadzonych nagraniach mowy jest trudne do spełnienia, gdyż indukowanie prawdziwych, spontanicznych emocji o zbyt dużym natężeniu jest niepożądane z etycznego punktu widzenia. Dodatkową sztuczność narzuca stosowanie metody badawczej i towarzyszące jej okoliczności. Jak wykazały badania pilotażowe, działanie bodźcami afektogennymi (głównie audiowizualnymi) wywołuje co najwyżej zmianę nastroju osoby nagrywanej, a więc uzyskany sygnał mowy byłby niedostatecznie silnie nacechowany emocjonalnie. Z kolei nagrania pochodzące z mediów, zawierające emocje spontaniczne i dostatecznie silne, charakteryzowało zróżnicowanie jakości akustycznej oraz brak materiału porównawczego poszczególnych emocji dla danej osoby.

Aby uzyskać nagrania sygnału mowy emocjonalnej przy zachowaniu poprawności etycznej, odpowiedniej jakości sygnału, a zarazem najbardziej korzystnej struktury nagrań (gdzie jedyną zmienną będzie stan emocjonalny), zaprojektowano korpus Emotive. Do nagrań użyto mikrofonu pojemnościowego AKG C5 Vocal oraz rejestratora Zoom H4N, uzyskując nagrania w formacie .wav 16 bit o częstotliwości próbkowania 44 100 Hz i poziomie SNR > 30 dB. Dokonano rejestracji akustycznego sygnału mowy emocjonalnej kilkunastu mówców – aktorów, studentów aktorstwa, osób o przygotowaniu teatralnym. Dla każdej osoby zarejestrowano te same zdania typu dialogowego w każdym z podstawowych stanów emocjonalnych (radość, smutek, strach, złość, zdziwienie, stan neutralny). Uzyskano dzięki temu szereg usystematyzowanych nagrań w postaci próbek (*.wav) czasowego sygnału mowy, etykietowanych intencją mówcy.

Nagrania poddano selekcji w oparciu o wyniki testów percepcyjnych, w których grupa słuchaczy oceniła każde z nich pod kątem zawartości emocjonalnej nagrania. W dalszym przetwarzaniu użyto tych spośród nagrań, które zostały jednoznacznie ocenione przez przeważającą część grupy statystycznej i traktowano je jako etykietowane zarówno intencją mówcy, jak i oceną słuchaczy.

Tabela 1 prezentuje procentowo klasyfikację oryginalnych nagrań dokonaną przez słuchaczy (w kolumnach – nagrania etykietowane intencją mówcy, w wierszach – ocena słuchaczy).

Tab. 1. Tablica błędu dla testów percepcyjnych nagrań z treścią. Oznaczenia: ne – stan neutralny, ra – radość, sm – smutek, st – strach, zd – zdziwenie, zl – złość, nie – nie rozpoznano
Tab. 1. Confusion table for perceptual tests for recordings with semantic content. Abbreviations: ne – neutral, ra – joy, sm – sadness, st – fear, zd – surprise, zl – anger, nie – not recognized

	ne	*ra*	*sm*	*st*	*zd*	*zl*
ne	**59**	7	10	1	2	0
ra	3	**42**	0	0	7	1
sm	7	0	**78**	13	0	3
st	1	0	3	**38**	4	3
zd	7	4	0	10	**75**	0
zl	1	2	0	3	4	**76**
nie	21	46	9	35	9	17

Analogiczne testy przeprowadzono dla wybranych nagrań, pozbawionych komputerowo treści (z pozostawioną jedynie linią intonacyjną zdania). Wyniki prezentuje tabela 2.

Tab. 2. Tablica błędu dla testów percepcyjnych nagrań pozbawionych treści (wartości procentowe). Oznaczenia: ne – stan neutralny, ra – radość, sm – smutek, st – strach, zd – zdziwenie, zl – złość, nie – nie rozpoznano
Tab. 2. Confusion table for perceptual tests of recording without semantic content. Abbreviations: ne – neutral, ra – joy, sm – sadness, st – fear, zd – surprise, zl – anger, nie – not recognized

	ne	*ra*	*sm*	*st*	*zd*	*zl*
ne	**41**	25	33	0	0	25
ra	12	**38**	0	0	0	12
sm	24	0	**34**	0	0	0
st	6	13	0	**78**	0	12
zd	0	0	0	0	**89**	0
zl	6	0	22	0	11	**38**
nie	11	24	11	22	0	13

Otrzymane wyniki są również podstawą do oceny jakości korpusu mowy emocjonalnej pod kątem wiarygodności odgrywanych emocji. Mogą też stanowić punkt wyjścia do interpretacji ludzkich zdolności percepcyjnych poszczególnych emocji w mowie. Sugerują one, które z emocji przejawiają się w mniejszym stopniu w sygnale mowy (a wyraźniejsza jest ich ekspresja i percepcja np. w mimice twarzy). Najwyższy poziom trafności oceny przez słuchaczy rzeczywistych wypowiedzi wystąpił dla smutku, złości i zdziwienia, a najniższy dla radości i strachu). Wyniki pokazują też, które emocje łatwiej (smutek, zdziwienie, złość), a które trudniej imitować (radość, strach).

Natomiast wyniki testów percepcyjnych dla nagrań pozbawionych treści ilustrują, że sama melodia wypowiedzi może być wystarczająca dla określenia zawartości emocjonalnej komunikatu. Szczególną rolę odgrywa w rozpoznawaniu zdziwienia ze względu na charakterystyczne uniesienie tonu w końcowej części wypowiedzi.

Łącznie, wykonany korpus mowy emocjonalnej aktorów został zweryfikowany oceną percepcyjną słuchaczy z wynikiem 61 % skuteczności rozpoznawania.

3. Przetwarzanie sygnałów

3.1. Parametryzacja i ekstrakcja cech

Wyselekcjonowane nagrania poddano preemfazie przy pomocy filtru górnoprzepustowego. Następnie dokonano ekstrakcji przebiegów częstotliwości podstawowej tonu krtaniowego F0 – podstawowej harmonicznej sygnału odzwierciedlającej częstotliwość drgania fałdów głosowych, używając metody autokorelacji, ze względu na jej dokładność i odporność na szum w sygnale (przykładowe przebiegi wygenerowane przy pomocy programu Praat zaprentowano na rys. 1-3) [5, 23]. Analiza różnicowa tak otrzymanych wykresów zmienności F0 pozwoliła zaobserwować zmienność intonacji w zdaniu dla poszczególnych emocji oraz wskazać ogólne prawidłowości zmian cech sygnału skorelowanych z emocjami.

Dalsza parametryzacja sygnału obejmowała obliczenie dla próbek *x(n)* sygnału mowy szeregu cech, m.in.:

- średnia częstotliwość podstawowa:

$$\bar{F}_0 = \frac{1}{N}\sum_{i=1}^{N} F_i \quad ,$$

- energia sygnału:

$$E = \sum_{n=n1}^{n2} x^2(n),$$

- największa i najmniejsza F0:

$$F_{0max}=\max\,(\{F_1, F_2, \ldots F_N\}),$$
$$F_{0min}=\min\,(\{F_1, F_2, \ldots F_N\}),$$

- zakres F0:

$$F_{0za} = F_{0max} - F_{0min},$$

- odchylenie standardowe F0:

$$\mathrm{F_{0od}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - \bar{F}_0)^2} \quad ,$$

- jitter:

$$J = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N-1}(F_i - F_{i-1})^2}}{\frac{1}{N}\sum_{i=1}^{N=1} F_i} \cdot 100\% \,,$$

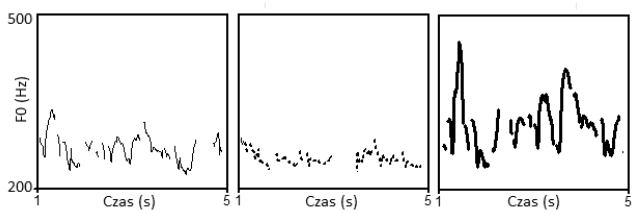
- shimmer:

$$S = \frac{\sqrt{\frac{1}{2N-1}\sum_{i=1}^{2N-1}(A_i - A_{i-1})^2}}{\frac{1}{N}\sum_{i=1}^{N=1} A_i} \cdot 100\%,$$

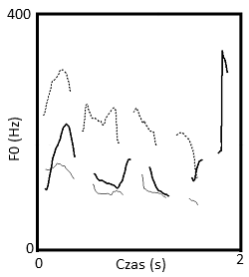
gdzie: F_i – i-ta częstotliwość podstawowa,
A_i – i-ta amplituda częstotliwości podstawowej.

a ponadto ilość i długość pauz oraz tempo mówienia.

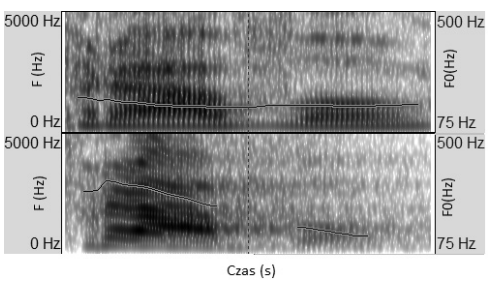
Następnie wszystkie wartości znormalizowano dla każdej emocji do stanu neutralnego dla każdego mówcy jako referencji, otrzymując znormalizowane wektory cech opisujących każde nagranie. [2, 4-5, 15].



Rys. 1. Przykład przebiegów zmienności intonacji (F0) fragmentu wypowiedzi tego samego mówcy o tej samej treści w stanach: neutralny (linia ciągła), smutek (linia przerywana), radość –(linia ciągła pogrubiona)
Fig. 1. Example intonation contours (F0) of the same utterance of the same speaker in states: neutral (solid line), sadness (dotted line) and joy (solid bold line)



Rys. 2. Porównanie przebiegów zmienności intonacji (F0) fragmentu wypowiedzi tego samego mówcy o tej samej treści (‘On jest najlepszy na świecie’) w stanach: neutralny (linia ciągła), złość (linia przerywana), zdziwienie (linia ciągła pogrubiona)
Fig. 2. Comparison of example intonation contours (F0) of the same utterance (‘He is the best in the world’) of the same speaker in states: neutral (solid line), anger (dotted line) and surprise (solid bold line)

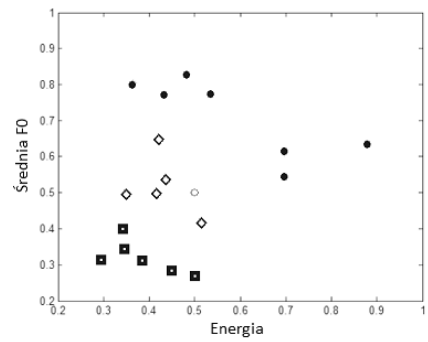


Rys. 3. Spektrogram dla pojedynczego słowa (‘prawo’) wypowiedzianego przez tego samego mówcę (kobietę) w stanach: neutralny (na górze) i złość (na dole). Na wykresach naniesiono przebieg F0 (linią ciągłą)
Fig. 3. Spectrogram for single word (‘right’) uttered by the same speaker (female) in neutral state (top figure) and in anger (bottom figure). Solid line presents F0 contour

3.2. Klasyfikacja

Z nagrań trafnie rozpoznanych przez słuchaczy wyselekcjonowano grupy nagrań jednolitych pod kątem treści i mówcy. Podzielono je na zestaw treningowy (90%) i testowy (10%). Przeprowadzono testy klasyfikatorów: kNN (k-najbliższych sąsiadów, ang. k nearest neighbours) oraz samoorganizującej się sieci neuronowej SOM (ang. self organizing map), używając programu Matlab [3, 14, 22, 23]. Do klasyfikacji wykorzystano wektory cech opisujących każde nagranie. Przykładowy wynik prezentacji poszczególnych stanów emocji, na płaszczyźnie średniej F0 i energii sygnału, zaprezentowano na rys. 4. Przykład obrazuje, jak reprezentacje nagrań tej samej emocji grupują się w obszary możliwe do rozgraniczenia. Testy powtarzano dla różnych zestawów emocji, parametrów i nagrań.

Jako wynik klasyfikacji otrzymano średnią skuteczność rozpoznawania metodą kNN zawierała się dla poszczególnych prób w granicach 50-70%, a przy pomocy sieci neuronowej – 60-80%.



Rys. 4. Przykład klasyfikacji przy pomocy kNN. Wartości znormalizowane do stanu neutralnego (punkt 0.5, 0.5) na płaszczyźnie: oś x – intensywność sygnału, oś y – średnia F0. Oznaczenia: kółko – radość, deltoid – złość, kwadrat – smutek
Fig. 4. Example classification using kNN. Values normalised to neutral state as reference (point 0.5, 0.5), x axis – intensity, y axis – mean F0. Circle – joy, diamond – anger, square – sadness

Wstępne testy klasyfikatorów dowodzą, że na podstawie pozyskanych nagrań i wyekstrahowanych cech możliwa jest automatyczna klasyfikacja zawartości emocjonalnej w nagraniach mowy.

3.3. Wokalne profile emocjonalne

Jako wynik analizy zbiorczej cech różnicujących emocje, sporządzono robocze wokalne profile wybranych emocji podstawowych (fragment zawiera tab. 1).

Tab. 3. Fragment zestawienia względnych zmian cech akustycznych wybranych emocji (∨ - spadek; ∧- wzrost) wraz ze średnią różnicą względem stanu neutralnego (procentowo)
Tab. 3. A part of summary of relative changes of selected emotion acoustic parameters (∨ - decrease; ∧- increase) with average change in reference to neutral state (%)

Parametr	*Smutek*	*Neutralny*	*Radość*
F0 średnia	$\bar{F}_0$ ∨ (-15%)	$\bar{F}_0$	$\bar{F}_0$ ∧ (+23%)
F0 – odchylenie standardowe	F_{0od}∨ (-32%)	F_{0od}	F_{0od}∧ (+40%)
Zakres F0	F_{0za}∨ (-43%)	F_{0za}	F_{0za}∧ (+46%)
Jitter	J -∨ (brak zmian lub niewielki spadek)	J	J ∧ (+5%)
Shimmer	S -∨(brak zmian lub niewielki spadek)	S	S ∧(+3%)
Energia	E∨ (-5%)	E	E∧ (+18%)
Długość pauz	P∧ (+12%)	P	P∨ (-4%)
Tempo mówienia	T∨ (-13%)	T	T∧ (+6%)

Ogólne zależności otrzymanych parametrów (spadek/wzrost wartości względem stanu neutralnego), są zgodne z danymi literaturowymi [1, 6, 8-13, 16-21]. Z kolei dla procentowych zmian wartości tych parametrów nie można określić poprawności ze względu na brak określenia przedziałów zmienności.

4. Podsumowanie i wnioski

Badania nad automatyczną detekcją emocji w sygnale mowy opierają się na założeniu, że istnieją uniwersalne wzorce wokalnego komunikowania poszczególnych emocji [1]. W praktyce bardzo dużym utrudnieniem jest subiektywność i różnice indywidualne zarówno w ekspresji jak i percepcji emocji w mowie, uwarunkowane różnicami osobniczymi. Na etapie etykietowania nagrań ocena pewnej grupy statystycznej nie zapewnia obiektywności – poziom odbioru i oceny emocji zależy od indywidualnej wrażliwości, empatii i inteligencji emocjonalnej. Dodatkowo jednoznaczna klasyfikacja była często niemożliwa, ponieważ najczęściej występują interkorelacje i współwystępowanie złożonych stanów emocjonalnych [8].

Dywersyfikacja indywidualnego sposobu ekspresji i percepcji emocji powoduje nieodzowną potrzebę kalibrowania systemu automatycznej detekcji emocji pod kątem charakterystyki profilu danego użytkownika afektywnego interfejsu głosowego [1, 3].

Wykonany korpus nagrań mowy emocjonalnej aktorów został poddany ocenie percepcyjnej, z wynikiem 61% oraz wstępnej analizie akustycznej. Otrzymane do tej pory rezultaty w zakresie cech charakterystycznych dla emocji podstawowych są zgodne z danymi literaturowymi [1, 6, 8-13, 16-21].

Jednocześnie należy pamiętać, że przetwarzanie nagrań z emocjami sztucznymi jest tylko modelem wokalnej ekspresji emocji, i dla niektórych cech może odbiegać od wyników prowadzonym na nagraniach zawierających emocje autentyczne [1].

Dalsze badania nad sporządzonym korpusem nagrań będą obejmowały ekstrakcję większej liczby parametrów oraz rozszerzenie wokalnych profilów emocji o kolejne cechy, jak również opis w dziedzinie częstotliwości. Zostaną również opracowane metody klasyfikacji, które posłużą do automatyzacji procesu rozpoznawania emocji w mowie. Na etapie weryfikacji takich algorytmów wskazane będzie testowanie ich na nagraniach zawierających emocje autentyczne.

5. Zastosowania

Badania nad rozpoznawaniem emocji w mowie mają znaczenie zarówno poznawcze, jak i wdrożeniowe. Obok całego spektrum zastosowań technologicznych (m.in. jako moduł systemów automatycznego rozpoznawania mowy i mówcy), aplikacje zbudowane w oparciu o algorytmy identyfikujące stan emocjonalny pacjenta w oparciu o parametry akustyczne jego mowy mają duży potencjał w dziedzinie diagnostyki i terapii medycznej.

W psychologii i psychiatrii pomiary parametrów głosu pacjentów mogą służyć jako metoda diagnostyczna (np. depresji, choroby afektywnej dwubiegunowej, ADHD). [3] Periodyczne badanie zabarwienia emocjonalnego głosu może służyć jako jedno z narzędzi wspomagających monitoring przebiegu leczenia zaburzeń psychologicznych i neurologicznych (np. zwiększenie tempa wypowiedzi i skrócenie długości pauz może świadczyć o sukcesie terapeutycznym i wychodzeniu pacjenta z choroby). Kolejnym potencjalnym zastosowaniem jest wczesne diagnozowanie stresu stanowiącego podłoże wielu chorób cywilizacyjnych [1].

Praca naukowa finansowana ze środków na naukę jako projekt badawczy POIG INSIGMA 01.01.02-00-062/09.

6. Literatura

[1] Red. Izdebski K.: Emotions in the human voice, Vol. I-III, Plural Publishing, San Diego 2008.
[2] Tadeusiewicz R.: Sygnał mowy, WKiŁ, Warszawa 1987.
[3] Ciota Z.: Metody przetwarzania sygnałów akustycznych w komputerowej analizie mowy, Wyd. EXIT Warszawa 2010.
[4] Ziółko M., Ziółko B.: Przetwarzanie mowy; Wyd. AGH, Kraków 2011.
[5] Boersma Paul: Praat, A system for doing phonetics by computer. Glot International 5:9/10, 341-345.
[6] Waaramaa-Maki-Kulmala T.: Emotions in voice - acoustic and perceptual analysis of voice quality in the vocal expression of emotions, University of Tampere 2009.
[7] Demenko G.: Analiza cech suprasegmentalnych języka polskiego na potrzeby technologii mowy, Wyd. Naukowe UAM, Poznań 1999.
[8] Lewis M., Haviland-Jones J.M.: Psychologia emocji, Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2005.
[9] Ekman P., Davidson R.: Natura emocji – podstawowe zagadnienia, Gdańskie Wydawnictwo Psychologiczne, Gdańsk 1999.
[10] Sidorova J.: Speech Emotion Recognition, DEA report, doctoral program Ci`encia Cognitiva i Llenguatge, Universitat Pompeu Fabra, 2007.
[11] Schuller B., Rigoll G., and Lang M.: Hidden markov model-based speech emotion recognition, Institute for Human-Computer Communication, 2003.
[12] Petrushin V. A.: Emotion recognition in speech signal: experimental study, development, and application, Center for Strategic Technology Research (CSTaR), 2000.
[13] Amir N.: Classifying emotions in speech: a comparison of methods. Holon Academic Institute of technology, EUROSPEECH 2001, Escandinavia.
[14] Demuth H., Beale M: Neural Network Toolbox for use with Matlab, mathworks.com
[15] Kłaczyński M.: Zjawiska wibroakustyczne w kanale głosowym człowieka – rozprawa doktorska, AGH Kraków 2007.
[16] Alter K., Rank E., Kotz S.A., Pfeifer E., Besson M., Friederici A.D., Matiasek J.: On the relations of semantic and acoustic properties of emotions. In Proceedings of the 14th International Conference of Phonetic Sciences (ICPhS-99), San Francisco, California, 1999.
[17] Campbell N.: Databases of Emotional Speech. In Cowie R. Douglas-Cowie E. & Schröder M. (Eds.) Proceedings of the ICSA Workshop on Speech and Emotion. Belfast, 2000.
[18] Petrushin V.A.: Emotion in speech: Recognition and Application to Call Centers". Artificial Neu. Net. In Engr. (ANNIE'99), Nov. 1999.
[19] Pittam J., Scherer K. R.: Vocal expression and communication of emotion. In M. Lewis & J. M. Haviland (Eds.), Hand-book of emotions. New York: Guilford Press. 1993.
[20] Zetterholm E.: Prosody and voice quality in the expression of emotions. Lund University. In SST Proceedings of the 7th Australian International conference on Speech Science And Technology. Sydney, 1998.
[21] Tato R., Santos R., Kompe R., Pardo J.: Emotion Recognition in Speech Signal,http://wwwgth.die.upm.es/partners/sony/ main.html
[22] Tadeusiewicz R., Flasiński M.: Rozpoznawanie obrazów, PWN, Warszawa 1991.
[23] Wszołek W.: Metody kognitywnej kategoryzacji w zastosowaniu do analizy i klasyfikacji wybranych przypadków mowy patologicznej, Wyd. AGH, Kraków 2011.

otrzymano / received: 16.01.2012
przyjęto do druku / accepted: 02.03.2012 *artykuł recenzowany / revised paper*