**Nazwa przedmiotu:**

Analiza sygnału audio i rozpoznawanie mowy

**Koordynator przedmiotu:**

Rajmund Kożuszek

**Status przedmiotu:**

Fakultatywny dowolnego wyboru

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Informatyka

**Grupa przedmiotów:**

Przedmioty techniczne - zaawansowane

**Kod przedmiotu:**

EASAR

**Semestr nominalny:**

2 / rok ak. 2021/2022

**Liczba punktów ECTS:**

4

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

1. liczba godzin kontaktowych – 47 godz., w tym
• obecność na wykładach i dwóch sprawdzianach 18 godz.,
• obecność na ćwiczeniach 12 godz.,
• obecność na ćwiczeniach laboratoryjnych 15 godz.,
• konsultacje 2 godz.,
2. praca własna studenta – 60 godz., w tym
• przygotowanie do ćwiczeń 20 godz.,
• przygotowanie do laboratoriów 20 godz.,
• przygotowanie do dwóch sprawdzianów 20 godz.
Łączny nakład pracy studenta wynosi 107 godz., co odpowiada 4 pkt. ECTS.

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

1.76 pkt. ECTS, co odpowiada 47 godz. kontaktowym

**Język prowadzenia zajęć:**

polski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

2.50 pkt. ECTS, co odpowiada 32 godz. ćwiczeń i przygotowaniu do ćwiczeń plus 35 godz. przygotowań i realizacji laboratorium

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład:  | 15h |
| Ćwiczenia:  | 15h |
| Laboratorium:  | 15h |
| Projekt:  | 0h |
| Lekcje komputerowe:  | 0h |

**Wymagania wstępne:**

Od osób uczęszczających na przedmiot wymagana jest wiedza z zakresu matematyki i rachunku prawdopodobieństwa, znajomość podstaw programowania i struktur danych oraz podstaw przetwarzania sygnałów (w zakresie odpowiedniego przedmiotu na studiach I stopnia). Wymagana jest umiejętność programowania w języku wysokiego poziomu (np. C++, C#, Java, Python lub Matlab).

**Limit liczby studentów:**

30

**Cel przedmiotu:**

Celem jest zapoznanie z metodami komputerowej analizy sygnałów audio, w tym metodami rozpoznawania ludzkiej mowy, przeznaczonymi do tworzenia systemów automatycznego rozpoznawania komend/zdań i identyfikacji mówcy, a także nabycie umiejętności projektowania takich systemów.

**Treści kształcenia:**

Wykład (18h) :
W1. Wprowadzenie – zastosowania, przegląd problematyki. Reprezentacja cyfrowego sygnału audio. (2h)
W2. Przetwarzanie wstępne sygnału audio (1h)
W3. Metody separacji i rozplatania sygnałów mieszanin (1h)
W4. Lokalizacja źródeł dźwięku i ekstrakcja sygnału zainteresowania (1h)
W5. Transformaty sygnału dźwięku i klasyfikacja dźwięków (2h)
W6. Struktura systemu ASR – automatycznego rozpoznawania mowy (1h)
W7. Sprawdzian (1h)
W8. Model fonetyczny języka (1h)
W9. Detekcja i parametryzacja sygnału mowy (1h)
W10. Modele DTW i HMM w rozpoznawaniu słów i fraz zdaniowych (2h)
W11. N-gramy i rozpoznawanie zdań (1h)
W12. Rozpoznawanie mówcy – zagadnienia, struktura podstawowego rozwiązania (1h)
W13. Zaawansowane metody modelowania i rozpoznawania mówcy (1h)
W14. Sieci neuronowe w rozpoznawaniu mowy (1h)
W15. Sprawdzian końcowy (1h)

Ćwiczenia (6 x 2h) :
C1. Przetwarzanie, separacja, ekstrakcja i lokalizacja źródeł
C2. Transformacje i klasyfikacja dźwięku
C3. Detekcja i przetwarzanie sygnału mowy
C4. Analiza akustyczno-fonetyczna mowy
C5. Rozpoznawanie fraz i zdań
C6. Modelowanie i rozpoznawanie mówcy

Laboratorium (1h + 7 x 2h) :
L1. Wprowadzenie (1h)
L2. Przetwarzanie, separacja i ekstrakcja źródeł
L3. Lokalizacja źródeł
L4. Transformacje i klasyfikacja dźwięku
L5. Detekcja i przetwarzanie sygnału mowy
L6. Analiza akustyczno-fonetyczna mowy
L7. Rozpoznawanie fraz i zdań
L8. Modelowanie i rozpoznawanie mówcy

**Metody oceny:**

Wykład prowadzony jest w tradycyjnej formie z wykorzystaniem materiałów w postaci elektronicznej. Przewiduje się 18 godzin części wykładowej realizowanej podczas 8 spotkań 2 godzinnych i 2 sprawdzianów po 1-godzinie.
Ćwiczenia mają charakter rozwiązywania zadań o charakterze algorytmiczno-obliczeniowym przez Prowadzącego wspólnie ze studentami. Zadania stanowią praktyczną ilustrację zagadnień omawianych wcześniej na wykładzie. Przewiduje się 12 godzin ćwiczeń realizowanych podczas 6 spotkań 2 godzinnych. Oceniana jest aktywność studentów podczas zajęć, w tym rozwiązywanie podanych wcześniej zadań.
Laboratorium polega na projektowaniu programistycznych rozwiązań wybranych zagadnień z wykorzystaniem narzędzi i bibliotek programów zainstalowanych na komputerach w laboratorium. Przewiduje się pracę w grupach 2-osobowych. Zajęcia w laboratorium składają się z 7 zajęć po 2 godziny każda, poświęcone odrębnym zagadnieniom przedstawianym podczas wykładu i ćwiczeń, oraz na jednych zajęciach 1 godzinnych wprowadzających. Oceniany jest stopień realizacji zadanych ćwiczeń laboratoryjnych.

**Egzamin:**

nie

**Literatura:**

- Literatura podstawowa
1. W. Kasprzak: Rozpoznawanie obrazów i sygnału mowy. Oficyna Wydawnicza PW, Politechnika Warszawska, 2009.
2. L. R. Rabiner and R. W. Schafer: Introduction to Digital Speech Processing. Foundations and Trends in Signal Processing, vol. 1, no. 1-2 (2007), pp. 1-194, NOW - the essence of knowledge, Boston - Delft. (Rozdziały: 1-6, 9)
3. J. Benesty, M. M. Sondhi, Y. Huang (eds): Handbook of Speech Processing. Springer, Berlin Heidelberg, 2008. (wybrane artykuły):
- R.W. Schafer: Homomorphic Systems and Cepstrum Analysis of Speech. In [3], pp. 161-180.
- S. Young: HMMs and related speech recognition technologies. In [3], 539-555.
- A. E. Rosenberg, F. Bimbot, S. Parthasarathy: Overview of Speaker Recognition. In [3], pp. 725-741
4. W. Kasprzak: Adaptive computation methods in digital image sequence analysis. (Rozdziały 5-9). Prace Naukowe - Elektronika, Nr. 127 (2000), Oficyna Wydawnicza PW, Warszawa.
- Literatura uzupełniająca
5. I. Goodfellow, Y. Bengio, A. Courville: Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org (wybrane rozdziały)
6. G.E. Hinton et al.: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Process. Mag., 29(6), 82–97, 2014
7. S. Makino, T.-W. Lee, H. Sawada: Blind Speech Separation. Berlin: Springer, 2007.
8. A. Cichocki, J. Karhunen, W. Kasprzak, R. Vigario: Neural Networks for Blind Separation with Unknown Number of Sources, Neurocomputing, Elsevier, NL, vol. 24 (1999), 55-93.
9. W. Kasprzak, N. Ding, N. Hamada: Relaxing the WDO assumption in blind extraction of speakers from speech mixtures. Journal of Telecommunications and Information Technology, Warsaw, Poland, vol. 2010, No. 4, 50-58.
10. Man-Wai Mak and Jen-Tzung Chien: Machine Learning for Speaker Recognition, INTERSPEECH 2016 Tutorial.
11. K. Fujimoto, N. Hamada, W. Kasprzak: Estimation and Tracking of Fundamental, 2-nd and 3-d Harmonic Frequencies for Spectrogram Normalization in Speech Recognition. Bulletin of the Polish Academy of Sciences. Technical Sciences, vol. 60(1) 2012, 71-8. DOI: 10.2478/v10175-012-0011-z
12. W. Kasprzak, P. Przybysz: Stochastic modelling of sentence semantics in speech recognition. Computer Recognition Systems 4, Advances in Intelligent and Soft Computing, vol. 95, pp. 737-746, Springer-Verlag, Berlin Heidelberg, 2011. (60%) (WoS 1 cyt.), (GS), link.springer.com, DOI: 10.1007/978-3-642-20320-6\_75
13. C. Zieliński et al.: Variable structure robot control systems: The RAPP approach. Robotics and Autonomous Systems, 94 (2017) 226–244, North-Holland.
14. W.Kasprzak et al.: Agentowa struktura wielomodalnego interfejsu do Narodowej Platformy Cyberbezpieczeństwa. Pomiary Automatyka Robotyka, tom 23, nr 3/2019, 41-54 (część I), tom 23, nr 4/2019 (część II) 5-18.

- Oprogramowanie
1. Carnegie Mellon University: Biblioteka Sphinx4. 2004.
 http://cmusphinx.sourceforge.net/doc/sphinx4/
2. Projekt KALDI: http://kaldi.sourceforge.net/index.html
3. Projekt CLARIN: http://mowa.clarin-pl.eu/
4. Projekt ALIZE: biblioteka „open source”. http://alize.univ-avignon.fr
Kod programu: https://github.com/ALIZE-Speaker-Recognition
Publikacje: http://mistral.univ-avignon.fr/publications.html
5. Platforma biometryczna Mistral: http://mistral.univ-avignon.fr/

**Witryna www przedmiotu:**

https://usosweb.usos.pw.edu.pl/kontroler.php?\_action=katalog2/przedmioty/pokazPrzedmiot&prz\_kod=103A-INISY-MSP-EASAR

**Uwagi:**

(-)

## Charakterystyki przedmiotowe