GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Speech Processing, PG_00058857		
Field of study	Informatics		
Date of commencement of studies	February 2024	Academic year of realisation of subject	2023/2024
Education level	second-cycle studies	Subject group	Optional subject group Subject group related to scientific research in the field of study
Mode of study	Full-time studies	Mode of delivery	at the university
Year of study	1	Language of instruction	Polish
Semester of study	1	ECTS credits	2.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Intelligent Interactive Systems -> Faculty of Electronics, Telecommunications and Informatics		
Name and surname of lecturer (lecturers)	Subject supervisor	dr hab. inż. Jan Daciuk	
	Teachers	dr hab. inż. Jan Daciuk	

Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	0.0	0.0	15.0	30
	E-learning hours included: 0.0						

Learning activity and number of study hours	Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
	Number of study hours	30	2.0	18.0	50

Subject objectives	The course concerns issues in biometrics and in speech processing. In biometrics, the aim is to familiarize students with methods and devices used in that domain. In speech processing, the aim is to get students aquainted with properties of speech, and to familiarize them with methods and particular solutions used in speech recognition and in speech synthesis.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K7_U01] can apply mathematical knowledge to formulate and solve complex and non-typical problems related to the field of study by:n-appropriate selection of source information and its critical analysis, synthesis, creative interpretation and presentation,n-application of appropriate methods and toolsn	The student can use the acquired mathematical knowledge in solving biometrics and speech processing problems.	[SU2] Assessment of ability to analyse information
[K7_W04] Knows and understands, to an advanced extent, the principles, methods and techniques of programming and the principles of computer software development or programming devices or controllers using microprocessors or programmable elements or systems specific to the field of study, and organisation of systems using computers or such devices	The student knows and understands in advanced degree the foundations, methods, and programming techniques as well as rules for developing software for biometrics and for speech processing.	[SW2] Assessment of knowledge contained in presentation [SW1] Assessment of factual knowledge
[K7_U06] can analyse the operation of components, circuits and systems related to the field of study; measure their parameters; examine technical specifications; interpret obtained results and draw conclusions	The student can measure characteristics and analyze biometrics and speech processing systems.	[SU3] Assessment of ability to use knowledge gained from the subject
[K7_W03] Knows and understands, to an increased extent, the construction and operating principles of components and systems related to the field of study, including theories, methods and complex relationships between them and selected specific issues - appropriate for the curriculum.	The student knows the theory and methods in biometrics and speech processing as well as the foundations of functioning of the systems used in those domains.	[SW3] Assessment of knowledge contained in written work and projects [SW2] Assessment of knowledge contained in presentation
[K7_U42] can solve engineering and research problems including design, assessment and maintenance of information systems and applications, using experimental methods and management techniques	The student can solve engineering and research problems in design, evaluation and maintenance of biometrics and speech processing systems.	[SU3] Assessment of ability to use knowledge gained from the subject

Subject contents

Biometrics (lecture):

1. Foundations
2. Measures
3. Biometric features as well as methods and systems that use them (partially at a seminar)

Speech processing (lecture):

1. Speech features (relation with ortography, phonetics nad phonology)
2. Speech recognition
3. Speech synthesis
4. Standards for speech markup
5. Examples of speech processing systems (seminar)

Prerequisites and co-requisites

Basic knowledge of machine learning, and neural networks ijn particular.

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
lecture (exam)	50.0%	70.0%
seminar (evaluation of a presentation and its preparation)	50.0%	30.0%

Recommended reading	Basic literature	1. Marek Wiśniewski, Zarys fonetyki i fonologii współczesnego języka polskiego, Wydawnictwo Uniwersytetu Mikołaja Kopernika, wydanie IV, Toruń 2001. 2. Daniel Jurafsky, James H. Martin, Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, second edition, Prentice-Hall, 2008. 3. Bartosz Ziółko, Mariusz Ziółko, Przetwarzanie mowy, Wydawnictwa AGH, Kraków 2011. 4. Ruud M. Bolle, Jonathan H. Connell, Sharath Pankanti, Nalini K. Ratha, Andrew W. Senior, Biometria, Wydawnictwa Naukowo-Techniczne, Warszawa, 2008. 5. Krzysztof Ślot, Wybrane zagadnienia biometrii, Wydawnictwa Komunikacji i Łączności, Warszawa, 2008. 6. Krzysztof Ślot, Rozpoznawanie biometryczne. Nowe metody ilościowej reprezentacji obiektów, Wydawnictwa Komunikacji i Łączności, Warszawa, 2010. 7. Zygmunt Ciota, Metody przetwarzania sygnałów akustycznych w komputerowej analizie mowy, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2010. 8. Mariusz Kubanek, Wybrane metody i systemy biometryczne bazujące na ukrytych modelach Markowa, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2013. 9. Deep Learning in Biometrics, Mayank Vatsa, Richa Singh, Angshul Majumdar (eds.), CRC Press, 2018. 10. Yaniv Taigman, Ming Yang, MarcAurelio Ranzato, Lior Wolf, Deepface: Closing the Gap to Human-Level Performance in Face Verification, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1701-1708, 2014.

	Supplementary literature	1. Danuta Ostaszewska, Jolanta Tambor, Fonetyka i fonologia współczesnego języka polskiego, PWN, Warszawa, 2002. 2. Alicja Nagórko, Podręczna gramatyka języka polskiego, PWN, Warszawa, 2010. 3. Zygmunt Ciota, Metody przetwarzania sygnałów akustycznych w komputerowej analizie mowy, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2010. 4. Ryszard Tadeusiewicz, Sygnał mowy, Wydawnictwa Komunikacji i Łączności, Warszawa 1988, książka dostępna pod adresem: http://winntbg.bg.agh.edu.pl/skrypty/0004/ 5. Tomasz Zieliński, Cyfrowe przetwarzanie sygnałów. Od teorii do zastosowań, Wydawnictwa Komunikacji i Łączności, 2009. 6. Grażyna Demenko, Korpusowe badania języka mówionego, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2015. 7. Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, Spoken Language Processing: A Guide to Theory, Algorithm and System Development, Prentice Hall, 2001. 8. Paul Taylor, Text-to-speech synthesis, Cambridge University Press, 2009. 9. Stefan Breuer, Multifunktionale und Multilinguale Unit-Selection-Sprachsynthese. Designprinzipien für Architektur und Sprachbausteine, rozprawa doktorska, Rheinischen Friedrich-Wilhelms-Universität Bonn, 2009. Dostępna pod adresem: http://hss.ulb.uni-bonn.de/2009/1650/1650.pdf 10. Peter Birkholz, 3D-Artikulatorische Sprachsynthese, rozprawa doktorska, Universität Rostock, 2005. Dostępna pod adresem: http://www.vocaltractlab.de/publications/birkholz-2005-dissertation.pdf 11. Takashi Masuko, HMM-Based Speech Synthesis and Its Applications, rozprawa doktorska, Tokyo Institute of Technology, 2002. Dostępna pod adresem: http://www.kbys.ip.titech.ac.jp/masuko/masuko-doctor.pdf 12. Sercan Ö. Arik et al., Deep Voice: Real Time Neural Text-To-Speech, ICML 2017. Dostępny pod adresem: https://arxiv.org/abs/1702.07825 13. Sercan Ö. Arik et al., Deep Voice: Multi-Speaker Real Time Neural Text-To-Speech, NIPS 2017. Dostępny pod adresem: https://arxiv.org/abs/1705.08947 14. Wei Ping et al., Deep Voice 3: Scaling Text-To-Speech With Convolutional Sequence Learning, ICLR 2018. Dostępny pod adresem: https://arxiv.org/abs/1710.07654 15. Sercan Ö. Arik et al., Neural Voice Cloning with a Few Samples, NIPS 2018. Dostępny pod adresem: https://arxiv.org/abs/1802.06006 16. Aäron van den Oord et al., WaveNet: A Generative Model for Raw Audio, arXiv preprint, Dostępny pod adresem: https://arxiv.org/abs/1609.03499 17. Jose Sotelo et al., Char2wav: End-to-End Speech Synthesis, 2017. Dostępny pod adresem: https://openreview.net/pdf?id=B1VWyySKx 18. Alex Graves, Abdel-rahman Mohamed, Geoffrey Hinton, Speech Recognition with Recurrent Neural Networks. Dostępny pod adresem: http://www.cs.toronto.edu/~fritz/absps/RNN13.pdf 19. Dario Amodei et al., Deep Speech 2: End-to-End Speech Recognition in English and Mandarin, ICML 2016. Dostępny pod adresem: http://proceedings.mlr.press/v48/amodei16.pdf 20. Alex Graves, Navdeep Jaitly, Towards End-To-End Speech Recognition with Recurrent Neural Networks, ICML14, pp. 1764-1772. 2014. Dostępny pod adresem: http://proceedings.mlr.press/v32/graves14.pdf 21. Haşim Sak, Andrew Senior, Kanishka Rao, Françoise Beaufays, Fast and Accurate Neural Network Acoustic Models for Speech Recognition, Interspeech 2015. Dostępny pod adresem: https://arxiv.org/abs/1507.06947
	eResources addresses	Adresy na platformie eNauczanie: Biometria i przetwarzanie mowy 2024 - Moodle ID: 36652 https://enauczanie.pg.edu.pl/moodle/course/view.php?id=36652
Example issues/ example questions/ tasks being completed		
Work placement	Not applicable	