Iwona Christop

Adam Mickiewicz University
Wieniawskiego 1
61-712 Poznań
iwochr2@st.amu.edu.pl

1 Research interests

1.1 Deep learning in audio classification

My main area of research interest is the use of **deep learning** methods in the classification of audio recordings, with a particular focus on **emotion recognition in speech** and **deep fake detection**.

Speech emotion recognition is an interdisciplinary problem combining psychology, physics, and computer science. By incorporating deep learning techniques, we can enhance spoken dialogue systems and improve their ability to understand users' emotional states. This solution has many applications, including adaptable interfaces, speech analysis for research purposes, and healthcare applications, where it could provide insight into patients' well-being.

As part of my Master's thesis, I am using pre-trained neural models such as Whisper by OpenAI (Radford et al. (2022)) and Wav2Vec2 by Facebook AI (Baevski et al. (2020)) to develop a robust system that can classify emotions in speech. As there is a visible lack of resources in this area, I am also creating a dataset of emotionally charged speech in Polish. By doing so, I hope to contribute to the advancement of research in the future.

As mentioned, my research interests also include the detection of deep fake audio. **Voice conversion techniques** are developing rapidly, and nowadays, one can easily preserve linguistic and semantic information of an utterance while manipulating the speaker's identity, prosody, and emotions. While these techniques have enormous potential, they also raise concerns. As such recordings are difficult to distinguish from real ones, it becomes harder to protect people against the spread of fake news, identity theft, and reputational damage (Kawa et al. (2022)).

I believe that developing models that use **deep learning techniques** to classify audio recordings is particularly important today. It is incredibly easy to come across fake news, which can greatly affect the society.

1.2 Sign languages

My second area of interest is **sign language**, which is often forgotten in the context of **dialogue systems**.

The world familiar to most people is unsuitable for deaf people, mainly due to communication barriers. For this reason, taking measures to guarantee better access to goods and services is crucial to enable deaf people to participate in social and public life.

A significant problem, which most people are not aware of, is that there is a large group of deaf people who cannot read or write in the native languages such as English or Polish.

In many institutions, deaf people cannot use an interpreter, making contact with a doctor, teacher, or police officer often impossible. The personnel of institutions is not familiar with the needs and problems of the Deaf, so they often demand written communication or expect lip reading.

While researchers around the world pay attention to this group of languages, we must remember that each nationality not only has its own **sign language**, but also dialects. Thus, it is important that **Polish Sign Language** also receive attention.

In addition, **sign languages** are considered **low-resource languages** – currently there is not enough annotated data to do more extensive research.

The development of a **real-time translation** system would provide a significant change in accessibility for the Deaf. I think it's worth paying attention to **low-resource languages** and making **dialogue systems** available to everyone.

1.3 Multimodality in dialogue systems

During my master's studies, I had the opportunity to participate in a research and development project that resulted in creating AMUseBot, a **task-oriented dialogue system** designed to assist the user in completing multi-step tasks.

The main goal of the project was to create a system that will provide engaging experience and keep the user focused throughout the conversation. In order to meet these objectives, we introduced two novel approaches – **dynamic multimodal communication** and **graph-based task management**.

The system's architecture follows standard baseline and it includes several modules responsible for specific **natural language processing** tasks – **automatic speech recognition**, **natural language understanding**, **dialogue management**, **natural language generation**, and **text-to-speech**.

As mentioned, the primary novelty is **graph-based**

Proceedings of the 19th Annual Meeting of the Young Reseachers' Roundtable on Spoken Dialogue Systems, pages 35–36
September 11–12, 2023.

task management. It effectively organizes the flow of dialogue and provides the user with visual cues during the conversation. A graph consists of a conversation history, with edges containing the user's statements and vertices containing the system's responses. With this representation, the visualization significantly improves user experience.

In addition, to ensure an engaging conversation, we gave the system different personalities. In the basic version, the user can choose from three options – default short commands, a kind chef who pays attention to details, and Gordon Ramsay.

Tests with users showed that the **multimodal approach** significantly increased their engagement and made the conversation more realistic.

The system was awarded an honorable mention in the research and development project competition at the AI Tech Summer School. The process of developing the system was described in the article "AMUseBot: Towards making the most out of a task-oriented dialogue system", published in the monograph "Progress in Polish Artificial Intelligence Research 4" (Christop et al. (2023)).

2 Spoken dialogue system (SDS) research

Currently, the biggest issue with **dialogue systems** is hallucination – they generate grammatically correct texts that are contentually incorrect. Research should therefore focus on creating chatbots that are able to back up their statements with relevant sources. To achieve this, young researchers should focus first and foremost on developing good quality data and extracting valuable information. This is the basis for obtaining substantively correct and satisfactory results.

The accuracy of information is also important for users. They should be able to get reliable answers with prompts that do not require specialized knowledge. In addition, **multimodality**, such as 3D models or human-like robots, should be used to provide a more realistic experience.

Nowadays, academic research is more focused on low-resource issues. These are matters that require attention but do not generate income. Companies, on the other hand, prefer to focus on more profitable ventures. Finding balance between both approaches is crucial to the development of **dialogue systems**.

It is worth pursuing research on **dialogue systems**, as they will be used even more extensively in the future – especially in the form of virtual assistants, helpline assistance, or emergency calls.

3 Suggested topics for discussion

- Using anthropomorphism to improve human-machine interaction – is speech emotion recognition too much?
- The gap between spoken and sign language: data acquisition methods and technological solutions.
- Leveraging multimodality in times of narrowing attention span.

References

Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, and Michael Auli. 2020. Wav2vec 2.0: A framework for self-supervised learning of speech representations. Curran Associates Inc., Red Hook, NY, USA, NIPS'20.

Iwona Christop, Kacper Dudzic, and Mikołaj Krzymiński. 2023. Amusebot: Towards making the most out of a task-oriented dialogue system. *Progress in Polish Artificial Intelligence Research 4* .

Piotr Kawa, Marcin Plata, and Piotr Syga. 2022. Specrnet: Towards faster and more accessible audio deepfake detection.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust speech recognition via large-scale weak supervision .

Biographical sketch

Iwona Christop is a MSc student in the field of Artificial Intelligence. Her Master thesis topic is "Speech Emotion Recognition using pre-trained neural models". She is a holder of a BSc in Computer Science, and completed courses in Acoustics and Polish Sign Language.

She is a co-author of an article published in the monograph "Progress in Polish Artificial Intelligence Research 4" and has had the opportunity to present her work at conferences such as Polish Conference on Artificial Intelligence 2023 and the US-Poland Science and Technology Symposium 2023 in Silicon Valley. In addition, she has been a participant in such events as Data Science Summit 2022, the 17th Conference of the European Chapter of the Association for Computational Linguistics, and the 24th Annual Conference of The European Association for Machine Translation.