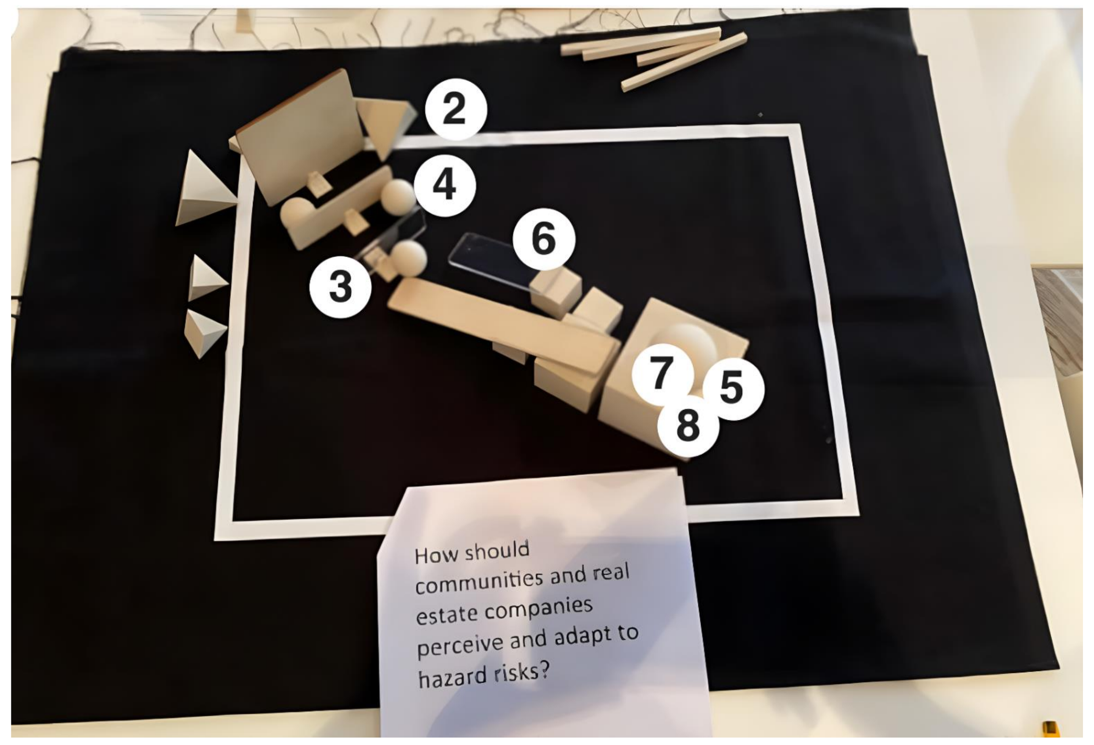
Meeting audio data summarization and visualization using ASR and NLP tools within the context of captured meeting data of the Shape Language

Author: Ella Milinovic; (e.milinovic@student.tudelft.nl) Supervisor & Responsible professor: Stephanie Tan, Edgar Salas Girones

Background

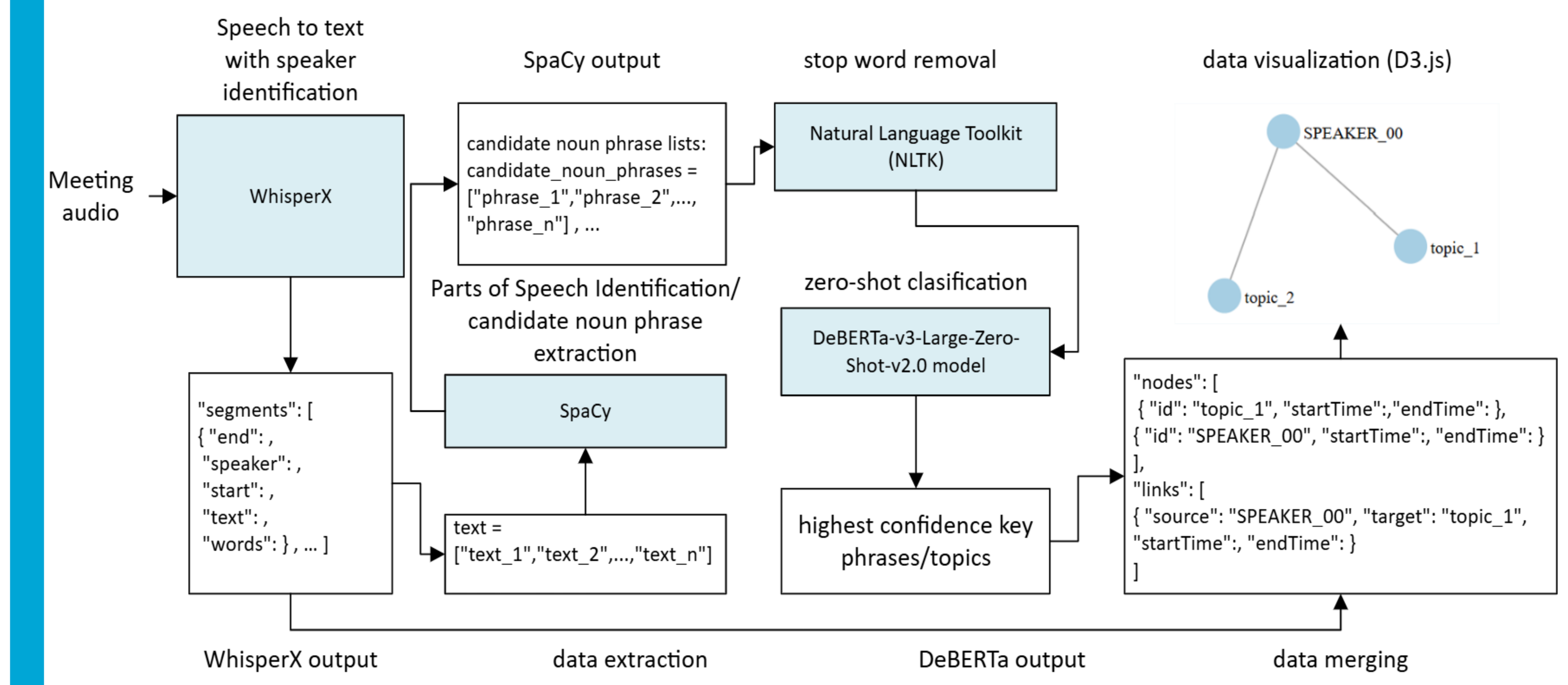


- Meetings are a key component in discussions, planning, and negotiations
- Problem: People are forgetful and often leave meetings with a different or incomplete understanding of the topic/s discussed
- Inspiration: Previous research that uses architectural shapes (Shape Language) to improve understanding of discussions [1].
- Existing tools and research:
- Tools: WhisperX [2], SpaCy [3], NLTK [4], DeBERTa [5], D3.js [6]
- Research: sliding window for meeting minute summaries [7], ChatGPT with prompts for dialogue summaries [8], unsupervised approaches for keyword extraction [9], graph-based meeting summary UI concept [10]
- Issue: None of the above can provide a key topic-focused visual meeting summary
- Proposed solution: Combining ASR models with NLP tools to create a visual key topic summary of meetings

Research question

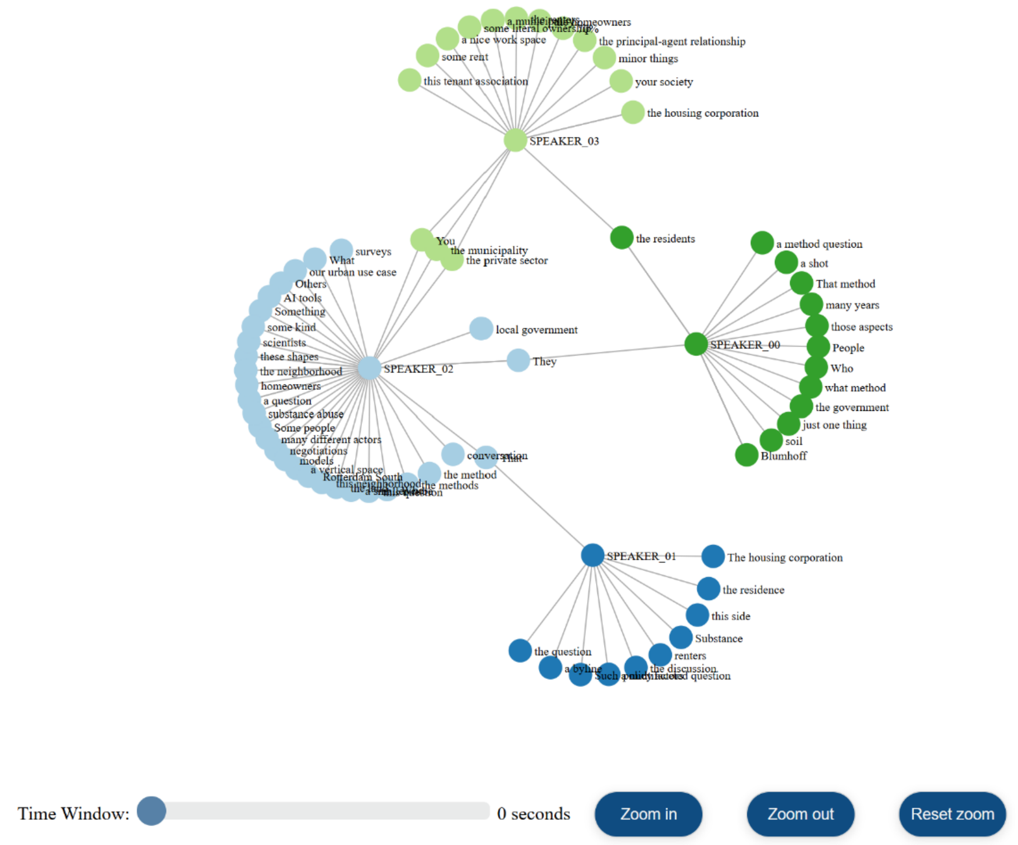
How can automatic speech recognition and natural language processing tools extract key meeting topics to create a visual summary of meeting audio data?

Process



1. Perform speech-to-text conversion with speaker diarization (recognition) using the WhisperX model on meeting audio
2. Create lists of candidate key phrases using spaCy.
3. Remove stop words using NLTK
4. Extract a list of key discussion points with the DeBERTa-v3-Large-Zero-Shot-v2.0 model
5. Combine data into a format accepted by D3.js
6. Create a graph-based web UI that shows which meeting participants are discussing which key topics in a certain time frame

Results

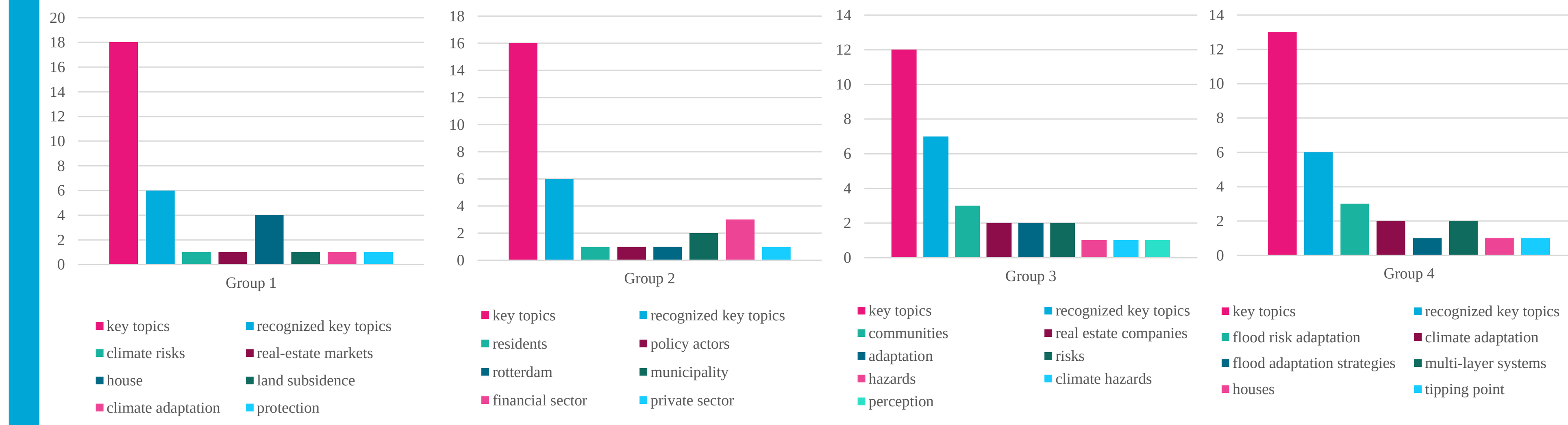


- 4 meetings with 4 participants
- Meeting topics: Climate change adaptation strategies in The Netherlands
- Meetings lasted between 12-15 minutes
- Meetings had between 12-18 key topics
- 78 – 93 discussed topics identified
- Average of 6 discussion points per minute
- The system identified 33%-58% of key topics
- Majority of the key topics were identified multiple times

Group	Total number of extracted topics in summary	Number of key topics	Correctly identified key topics (%)	Total number of occurrences of key topics
Group 1	93	18	33.33%	9
Group 2	78	16	37.50%	10
Group 3	87	12	58.33%	12
Group 4	92	13	46.15%	11

Table 1: Key topics extraction results

Groups:	Key topics:
Group 1	**climate risks**, **real-estate markets**, **house**, flood risk, urban environment, pole rot, **land subsidence**, **climate adaptation**, elevation, sea waves, dikes, **protection**, floating house, coastline, retreat strategy, inhabitants, migration, protected area
Group 2	**residents**, **policy actors**, subsidence, Bloemhof, **Rotterdam**, fragile situation, dynamic situation, environment, infrastructure, social housing renters, private homeowners, **municipality**, **financial sector**, **private sector**, real-estate developers, insurers
Group 3	**communities**, **real estate companies**, **adaptation**, **risks**, **hazards**, **climate hazards**, barriers, **perception**, pathways, society, integration
Group 4	**flood risk adaptation**, **climate adaptation**, **flood adaptation strategies**, maladaptation, **multilayer systems**, **houses**, subsidence, river, bottleneck, **tipping point**, decisions, black box, decentralized decision making



Conclusions

- Proposed system can identify 33%-58% key meeting topics confidently
- Results show that the system can be used to improve user understanding and retention of key meeting data
- Future research: optimizing system performance, extensive user studies into user behaviour and comprehension during meetings, optimal web UI design for summarizing meeting data

References

[1] A. Kamp. "Speaking: Part I- Speaking architecture / Part II- Speaking architecture / Part III- Speaking through form , 2018.
[2] MaxBain, Jaesung Huh, Tengda Han, and Andrew Zisserman. WhisperX: Time-Accurate Speech Transcription of Long-Form Audio, 3 2023.
[3] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spaCy: Industrial-strength Natural Language Processing in Python. 2020.
[4] Steven Bird, Ewan Klein, and Edward Loper. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.",2009.
[5] Moritz Laurer, Wouter van Atteveldt, Andreu Casas, and Kasper Welbers. Building Efficient Universal Classifiers with Natural Language Inference, December 2023. arXiv:2312.17543 [cs].
[6] Mike Bostock. D3.js- data-driven documents, 2012.
[7] Jia Jin Koay, Alexander Roustai, Xiaojin Dai, and Fei Liu. A Sliding-Window approach to automatic creation of meeting minutes, 4 2021.
[8] Yongxin Zhou, Fabien Ringeval, and Franc ̧ois Portet. Can GPT models Follow Human Summarization Guidelines? Evaluating ChatGPT and GPT-4 for Dialogue Summarization, 10 2023.
[9] Feifan Liu, Deana Pennell, Fei Liu, and Yang Liu. Unsupervised approaches for automatic keyword extraction using meeting transcripts. In Mari Ostendorf, Michael Collins, Shri Narayanan, Douglas W. Oard, and Lucy Vanderwende, editors, Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 620–628, Boulder, Colorado, June 2009. Association for Computational Linguistics.
[10] Yurdaer Doganata and Mercan Topkara. Visualizing meetings as a graph for more accessible meeting artifacts. In CHI '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, page 1939–1944, NewYork, NY, USA, 2011. Association for Computing Machinery.