News Report Adaptation for Synthetic Voice Presentation

Beatrix Chik Wu, Master's in Information Science, Department of Information Science and Media Studies, University of Bergen

Supervisor: Bjørnar Tessem, University of Bergen
Co-Supervisors: Pablo A. Vidales Calderon & Saikiran Tharimena, Schibsted

Media Futures

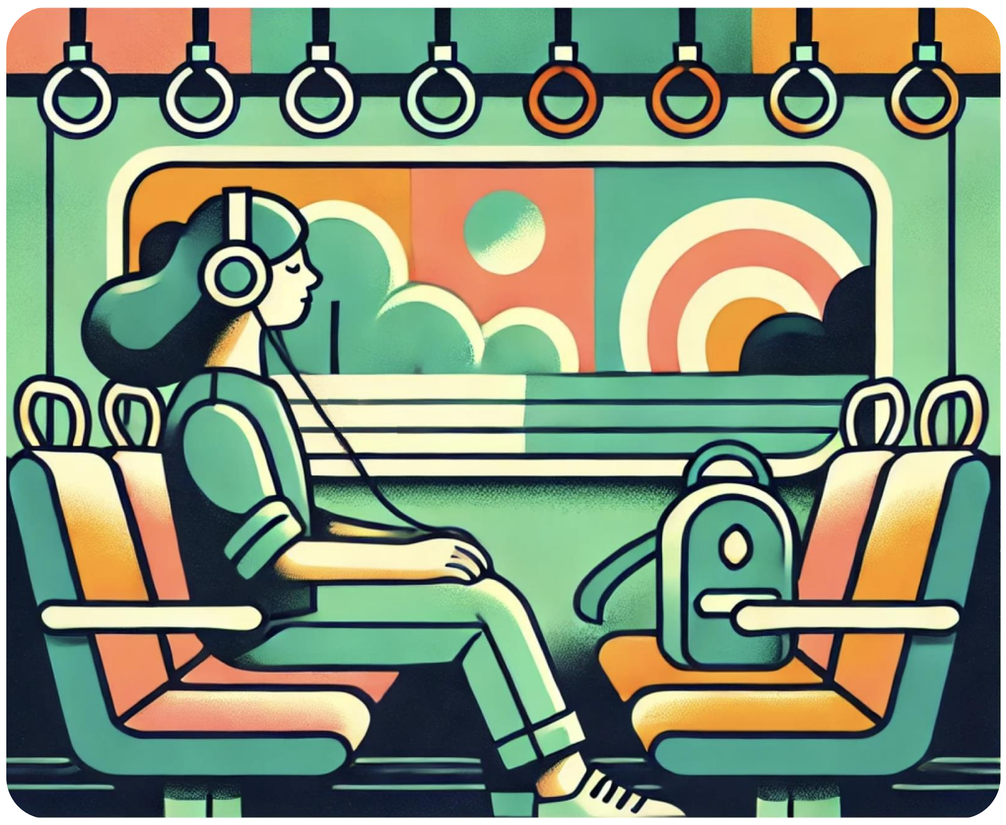

Image generated using ChatGPT and DALL-E

Introduction

The news media landscape is continuously evolving, with the consumption of auditory content, such as news podcasts, on the rise in recent years. According to Medietilsynet, daily podcast listening rose from 12% in 2019 to 17% in 2022, with news podcasts consistently ranking among the top spots, indicating a shift in how news are consumed in Norway.

This master's thesis, in collaboration with Schibsted, explores how written news articles can be adapted for presentation with synthetic voice. Schibsted's current «listen to the article» feature mainly reads out the written articles directly, which may not provide the optimal listening experience. To address this, this project aims to develop software that transforms written content into a spoken format, functioning like a script designed for text-to-speech systems. By specifically tailoring articles for oral presentation, the tool seeks to improve the overall auditory experience of consuming news through synthetic voice for a diverse audience.

Research Question

How to design and evaluate a system for adapting written news articles into oral formats with generative AI and rule-based approaches?

Methods

Data Collection and Preparation
Since no existing dataset directly pairs written articles with corresponding oral versions, the first step is to create a synthetic dataset using one-shot learning through some steps:

1. Identify an audio and a written version of the same news story
2. Transcribe the audio version
3. Adapt the written article into an oral format, use the audio transcript and a generative AI tool
4. Repeat steps 1-3 a limited number of times to create an initial foundational dataset
5. Expand the dataset with one-shot learning by providing the dataset to a generative AI tool to generate oral versions of new articles in a similar style

Model development & Evaluation
I will use a pre-trained generative AI model, fine-tuning using the synthetic dataset. The rule-based system will be integrated to address specific areas where manual control is necessary. The system will be evaluated through quantitative NLP metrics and qualitative assessment by users and experts.

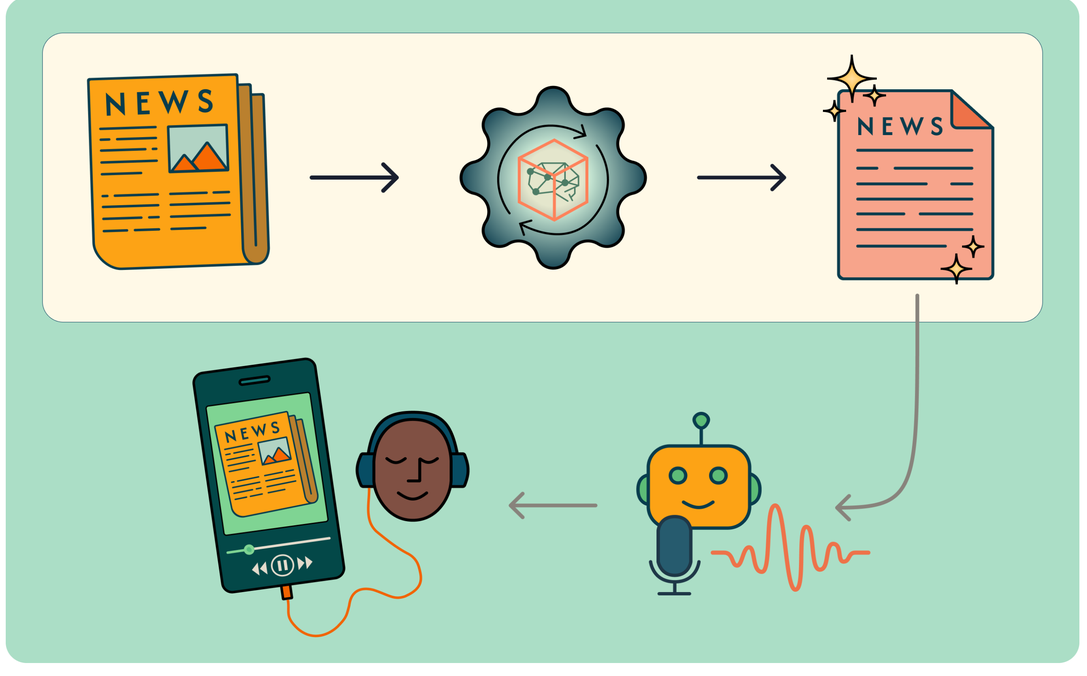

Illustration of the project workflow: My work focuses on the initial steps, where a news article is processed by the software developed, transforming it into an oral-friendly format. The adapted text can then be sent to a text-to-speech tool, generating an audio version that users can listen to.

Further Readings
Kaur, N., & Singh, P. (2022). Conventional and contemporary approaches used in text to speech synthesis: A review. *Artificial Intelligence Review, 56*(7), 5837–5880. https://doi.org/10.1007/s10462-022-10315-0

Hadi, M. U., Al Tashi, Q., Shah, A., Qureshi, R., Muneer, A., Irfan, M et al. (2024). Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects. *Institute of Electrical and Electronics Engineers (IEEE)*. http://dx.doi.org/10.36227/techrxiv.23589741.v6

PARTNERS

amedia · Schibsted · TV 2 · NRK · Universitas Osloensis · Bergens Tidende · nb.no · VIZRT · WOLFTECH · NORCE · UiS · Faktisk.

HOST

UNIVERSITY OF BERGEN

FUNDED BY

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

Forskningsrådet · Norwegian Centre for Research-based Innovation