



A powerful and modern AAC composition tool for impaired speakers

Aanchan Mohan^{1,2,4}, Monideep Chakraborti¹, Katelyn Eng^{1,3}, Nailia Kushaeva¹, Mirjana Prpa²,
Jordan Lewis², Tianyi Zhang^{1,4}, Vince Geisler, Carol Geisler

¹Happy Prime Inc; ²Northeastern University; ³Mercury Speech and Language; ⁴University of
Victoria, Canada

aanchan@happyprime.io, mo.chakraborti@northeastern.edu

Abstract

Augmentative and alternative communication (AAC) software assists impaired speakers to communicate. Enabling context awareness, authenticity and ease of use goes a long way in empowering communication. Our work presents an easy to use AAC tool that allows for message composition with text or emoji input by typing as well as speaking with contextually relevant word-level suggestions. Any transcription errors are corrected, and contextually relevant phrases with the appropriately chosen emotional tone are suggested by large-language models. The user is then able to use modern text-to-speech synthesis to be able to synthesize the composed message in the user's own voice. Our system additionally maintains features from AAC software such as word-cards, pre-composed and frequently used messages.

Index Terms: augmentative and alternative communication, atypical speech recognition, large language models, speech synthesis, human computer interaction

1. Introduction

To serve populations with motor speech disorders and language disorders our augmentative and alternative communication (AAC) application allows for message composition with text or emoji input by typing as well as speaking while offering contextually relevant word-level suggestions. Any transcription errors are corrected, and contextually relevant phrases with the appropriately chosen emotional tone are suggested by large-language models (LLMs) to compose a complete message. The user is then able to use modern text-to-speech synthesis to be able to synthesize the composed message in the user's own voice. Our application uses a combination of automatic speech recognition (ASR), text-to-speech (TTS) synthesis, and Large Language Models (LLMs).

Past conversational context with a particular communication partner is useful in interpreting and interpolating current (possibly incomplete) conversational phrases. Current AAC solutions do not allow for nuances in personal expression [1]. The emergence of LLMs have ushered in a range of possibilities especially for AAC applications. Valencia et al. [2] conducted a study for AAC users to use LLMs for various tasks. The study concludes that while AI generated phrases do save time and cognitive and physical effort, the generated phrases do need to reflect their own communication style and preferences. Recent work by Cai et al. [3], has used communication context in dialogue to allow for LLMs to perform abbreviation expansion of short phrases typed AAC users. Additionally error-correction solutions allow applications to work with transcription errors from users [4]. Last but not the least users feel extremely empowered when the message that is composed sounds like them.

This is possible as recent models like VALL-E [5] have zero-shot capabilities for TTS to generate voice with only 3 seconds of enrollment audio data.

There is certainly a gap in the software available on the market for AAC users that allows for: (1) multi-modal text and speech input (using ASR), (2) contextual composition and correction (using LLMs), (3) audio playback in the user's own voice (using TTS and Voice Conversion (VC)), and (4) include functionalities of traditional AAC software. Additionally, while recent literature has looked at ASR, TTS/VC and LLM technologies individually in the context of AAC, there is very little research that has explored their combined use for message composition. The rest of this paper gives an overview of our application in Section 2 and then provides a discussion and conclusion in Section 3.

2. Application Overview

This section gives an overview of our application. Figure 1 shows a logical block diagram of the user journey. A user with impaired speech can either choose to speak or type and easily switch between the two input modalities. They user is able to add emotions, and specify whom they are composing their message. ASR technology is able to transcribe their speech. It is likely the transcription errors at this point. Along with past conversational context, a large language model is able to provide relevant suggestions. The final message is available both as a completed text phrase as well as a clear audio message in the users' own voice. User voice samples are collected during the enrollment phase.

The application is currently under development. Figure 2 shows a screenshot of the interface to our mobile application. This application has been collectively co-created by community members diagnosed with Autism, a speech and language pathologist and a technical team which includes a user experience professional and a human computer interaction expert. The screenshot shows the user intending to compose a message in a happy tone to their father, with as few clicks as possible. The LLM, knowing past conversational context is able to give relevant contextual suggestions. The mic button in the figure, labelled as 'Record', allows them to go into the so-called 'speak mode' if they wanted to in order to activate ASR transcription. Once the user is done composing their message, they are able, to play back the message in their own voice using TTS using the 'Play' button. Our application also allows the user to save this message for future use. It is very common for AAC users to compose their messages ahead of time to either prepare for an appointment, or perhaps an interaction that they might anticipate. In case the 'Type' mode is cumbersome, our application also allows for the use of 'Cards' so that words appear on cards

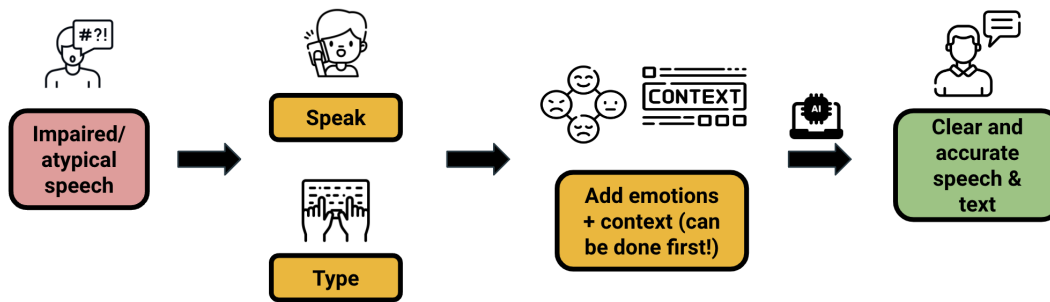


Figure 1: Overview of our AAC application

that make it for easier input.

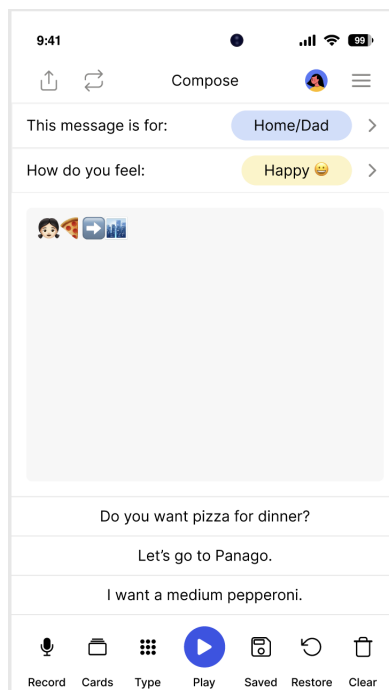


Figure 2: Prototype screenshot : A user in type mode, communicating with dad, wanting to go for pizza with LLM suggestions below. The mic button allows for audio input and transcription as well. Personalized voice output is available with the play button.

The usage of this application has garnered a lot of interest with parents of children with Autism and Down Syndrome, speech and language pathologists, older adults and school teachers locally in the province of British Columbia in Canada. Existing AAC software is unable to meet the needs of the population we have spoken with. Also, with recent accessibility legislation in BC, all workplaces need to be compliant to provide adequate software and tools for differently abled individuals within the next 4 years.

The mobile application has its front-end is written in ReactNative, which at the moment is dependent on an application programming interface (API) backend. Our ASR functional-

ity is open sourced and is available as BoltX¹. BoltX extends the WhisperX² project with additional windowing for efficient ASR. Our roadmap includes a similar custom LLM backend integration, along with TTS. Over time our goal is to be able to move all of our backend technology to work on-device as the technology evolves. While backend integration with our custom ASR is pending our demo shows a current implementation of our design using 'react-native-voice'. Ethics is currently underway, and work is currently pending evaluation from parents, caregivers and speech and language pathologists.

3. Conclusion

A powerful new tool for communication for AAC was presented in this show-and-tell paper as a mobile application. The tool really aims to serve its target users by allowing them to express themselves in a way that empowers them to sound authentic while using as few clicks to compose their message as possible.

4. References

- [1] S. K. Kane, M. R. Morris, A. Paradiso, and J. Campbell, ““at times avuncular and cantankerous, with the reflexes of a mongoose” understanding self-expression through augmentative and alternative communication devices,” in *Proceedings of the 2017 acm conference on computer supported cooperative work and social computing*, 2017, pp. 1166–1179.
- [2] S. Valencia, R. Cave, K. Kallarackal, K. Seaver, M. Terry, and S. K. Kane, ““The less I type, the better”: How AI language models can enhance or impede communication for AAC users,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–14.
- [3] S. Cai, S. Venugopalan, K. Tomanek, A. Narayanan, M. Morris, and M. Brenner, “Context-aware abbreviation expansion using large language models,” in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle, United States: Association for Computational Linguistics, Jul. 2022, pp. 1261–1275.
- [4] R. Ma, M. Qian, P. Manakul, M. Gales, and K. Knill, “Can generative large language models perform asr error correction?” *arXiv preprint arXiv:2307.04172*, 2023.
- [5] C. Wang, S. Chen, Y. Wu, Z. Zhang, L. Zhou, S. Liu, Z. Chen, Y. Liu, H. Wang, J. Li *et al.*, “Neural codec language models are zero-shot text to speech synthesizers,” *arXiv preprint arXiv:2301.02111*, 2023.

¹<https://github.com/SlangLab-NU/whisperX-webtrc>

²<https://github.com/m-bain/whisperX>