July 23, 2019

Smart Speech Composition for Augmentative and Alternative Communication Using a Machine-Learned Model

Phil Weaver

Smart Speech Composition for Augmentative and Alternative Communication Using a Machine-Learned Model

Abstract:

This publication describes an augmentative and alternative communication (AAC) user equipment (UE) that enables a user to select autocompleted sentences displayed in a user interface (UI) of the AAC UE. The AAC UE scans ambient sounds to identify speech that is being spoken in the vicinity of the user. After the AAC UE scans for speech, the AAC UE converts the audible speech into digitized speech using a speech-recognition model. Also, the AAC UE identifies the audience in the conversation with the user by employing user input (*e.g.*, the user selects the audience), voice recognition, facial recognition, radar signature, biometric sensors (*e.g.*, a person may scan their thumb on the AAC UE before communicating with the user), media address control identification (MAC ID) (*e.g.*, the AAC UE can scan MAC IDs of smartphones used by the audience), radio-frequency identification (RFID) (*e.g.*, an employee's badge), or other sensors (*e.g.*, in car seats). The AAC UE feeds the digitized speech and the identity of the audience into a machine-learned (ML) model, which analyzes the speech and makes suggestions on sentences that the user may want to use. The UI of the AAC UE displays the suggested sentences and waits for user input. The user reads the suggested sentences and selects the sentences that are applicable to a conversation. In case the user does not like the suggested sentences, they can use a keyboard to compose a new sentence or to modify a suggested sentence. The digitized composed speech, aided by user input, is then converted to synthesized speech. In addition, the digitized composed speech becomes an input to the ML model. Reiteratively, the ML model is updated to make a better prediction in future conversations, thus speeding up the communication process.

Keywords:

Augmentative and alternative communication, AAC, aided AAC device, aided AAC user equipment, speech composition, smart reply, device for the impaired, machine learning, artificial intelligence, AI, neural network, model training, speaker identity, listener identity, voice recognition, speech recognition, synthesize.

Background:

Augmentative and alternative communication (AAC) is an umbrella term that encompasses communication methods used to supplement or replace speech or writing for those with impairments in the production or comprehension of spoken or written language. AAC aids a user who has a speech and language impairment due to a physical or mental condition, such as cerebral palsy, post-stroke conditions, intellectual impairment, some forms of autism, amyotrophic lateral sclerosis, Parkinson's disease, and other conditions that affect the user's ability to effectively communicate. The user may utilize AAC permanently, temporarily, some of the time during their day, or all the time during their day, depending on their condition and/or lifestyle. A user can utilize a user equipment (UE) device for AAC communication (an "AAC UE"). An AAC UE can be an unaided AAC UE that does not require external tools (*e.g.*, facial expressions, vocalizations, gestures, sign language) or an aided AAC UE that requires external tools (*e.g.*, user equipment (UE), communication board). This publication focuses on the AAC UE.

The AAC UE may be a tablet that displays a user interface (UI) with different options that the user may choose from to compose what they want to communicate. Depending on the user's cognitive abilities, the choices may be highly restricted (*e.g.*, two yes-or-no buttons) or general (*e.g.*, a keyboard). Figure 1 illustrates an AAC UE with a keyboard.

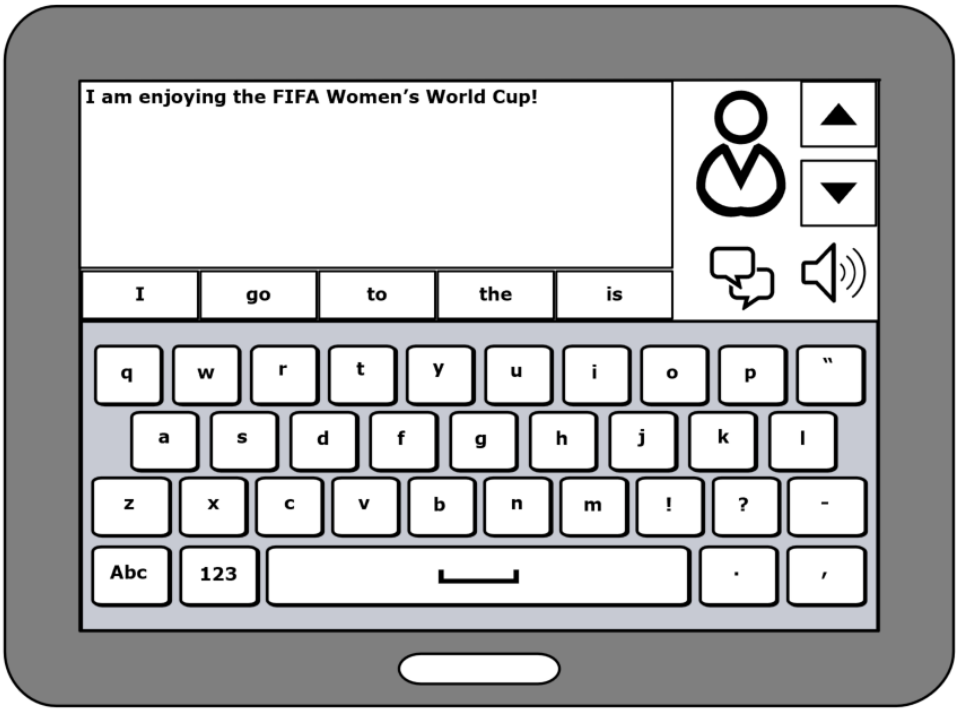

Figure 1

Assume Jane has a speech-impairment condition, and she uses an AAC UE to communicate with her family, friends, caregivers, and coworkers, as is illustrated in Figure 1. Jane has the cognitive ability to comprehend and compose general and complex sentences. Like people with no speech impairment, Jane has a variety of interests. Even though Jane's AAC UE incorporates a full keyboard that allows her to compose complex sentences, Jane struggles composing sentences in real-time. She often finds herself responding to a comment long after the people she is communicating with have commented on a topic. For example, assume Jane is watching the *Fédération Internationale de Football Association* (FIFA) Women's World Cup 2019 (France 2019) with her family. Her father comments on how exiting the championship is and, then, shifts the conversation to talk about a different topic. Three minutes after Jane's father commented on France 2019, Jane, using her AAC UE, exclaims, "I am enjoying the FIFA Women's World Cup!"

One can understand the added challenges that Jane faces to have a normal conversation in real-time even though she can comprehend and compose complex sentences.

To this end, some AAC UE original device manufacturers (ODMs) have shortened the time it takes the user to compose a sentence by incorporating pages with autocomplete sentences that the user can select using the UI of the AAC UE. These autocomplete sentences are pre-configured to allow the user to select words depending on the context of the conversation. For example, the user may use a food-related autocomplete page when they want to order food. As another example, the user may use a caregiver-related autocomplete page when they ask for help preparing to go to bed. Nevertheless, it is challenging for ODMs of the AAC UE to create pre-configured autocomplete pages for every scenario. In addition, it is challenging for the user to select the proper autocomplete page for every scenario.

Therefore, it is desirable to have a technological solution that allows users with speech impairment to take part in a close to real-time general-purpose conversation.

Description:

This publication describes an augmentative and alternative communication (AAC) user equipment (UE), such as a tablet, that uses a machine-learned (ML) model to auto-populate smart replies in a user interface (UI) of the AAC UE that are representative of a current conversation. Using the ML model to analyze the overall conversation, the UI of the AAC UE can auto-populate suggested long and complex sentences, which limit the need for a user to compose speech by using the keyboard of the AAC UE. The suggestions displayed in the UI of the AAC UE aid the user to take part in a close to real-time general-purpose conversation, as is illustrated in Figure 2.

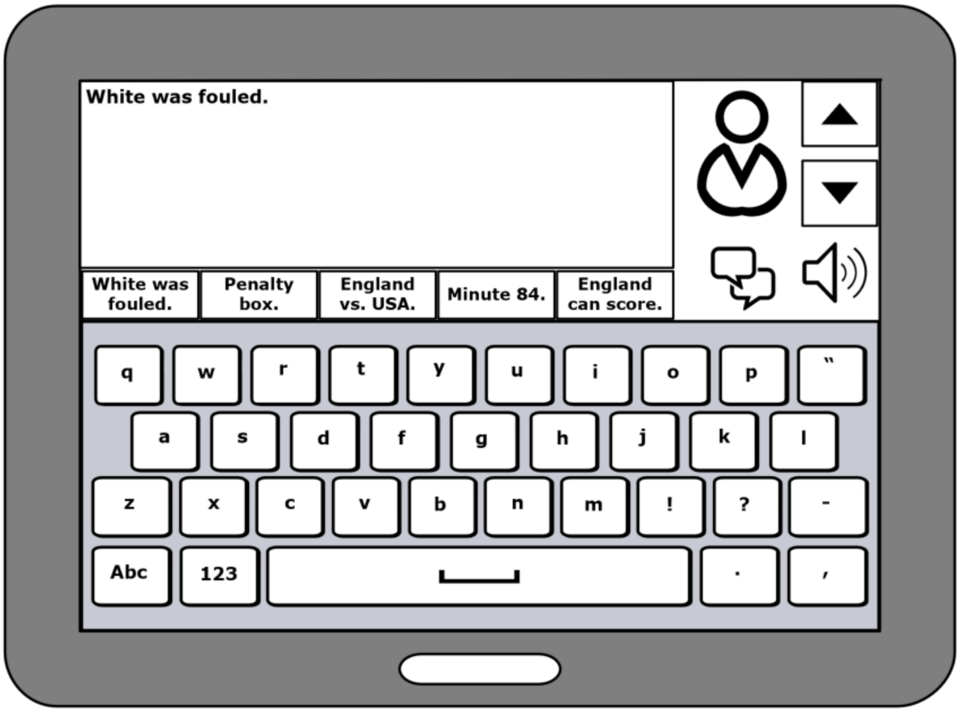

Figure 2

Assume Jane uses an AAC UE that can aid her in taking part in a close to real-time general-purpose conversation. Like the example illustrated in Figure 1, Jane is watching the France 2019 tournament with her family. Jane and her family are on "pins and needles" as they watch the match of England versus (vs.) the United States of America (USA). It is minute 84, and the score is "England 1, USA 2." England is proving to be a formidable opponent to the reigning champions (USA). With six minutes to go, at minute 84, England's forward, White (jersey number 18), is fouled in the penalty box. Jane's AAC UE analyzes the ambient sound (speech), such as her father's speech, her siblings' speeches, the television (TV) commentators, and other phrases representative of the situation. The TV commentators are discussing whether the USA defender fouled White, whether White took a dive, or whether White simply lost her balance. Jane's AAC UE displays different phrase selections that are representative of the ambient speech, such as "White was fouled," "Penalty box," "England vs. USA," "Minute 84," "England can score," as is illustrated in

Figure 2. Jane's family are arguing whether the referee made an incorrect penalty call. As the TV station replays the scenario from different angles, Jane clearly sees that the USA defender fouled White. So, in real-time, Jane adds to the conversation by selecting the phrase, "White was fouled." By using the AAC UE that displays suggested phrases that are representative of the ambient speech, Jane can truly be part of the conversation. Figure 3 illustrates how the AAC UE aids the user to produce close to real-time audible speech.

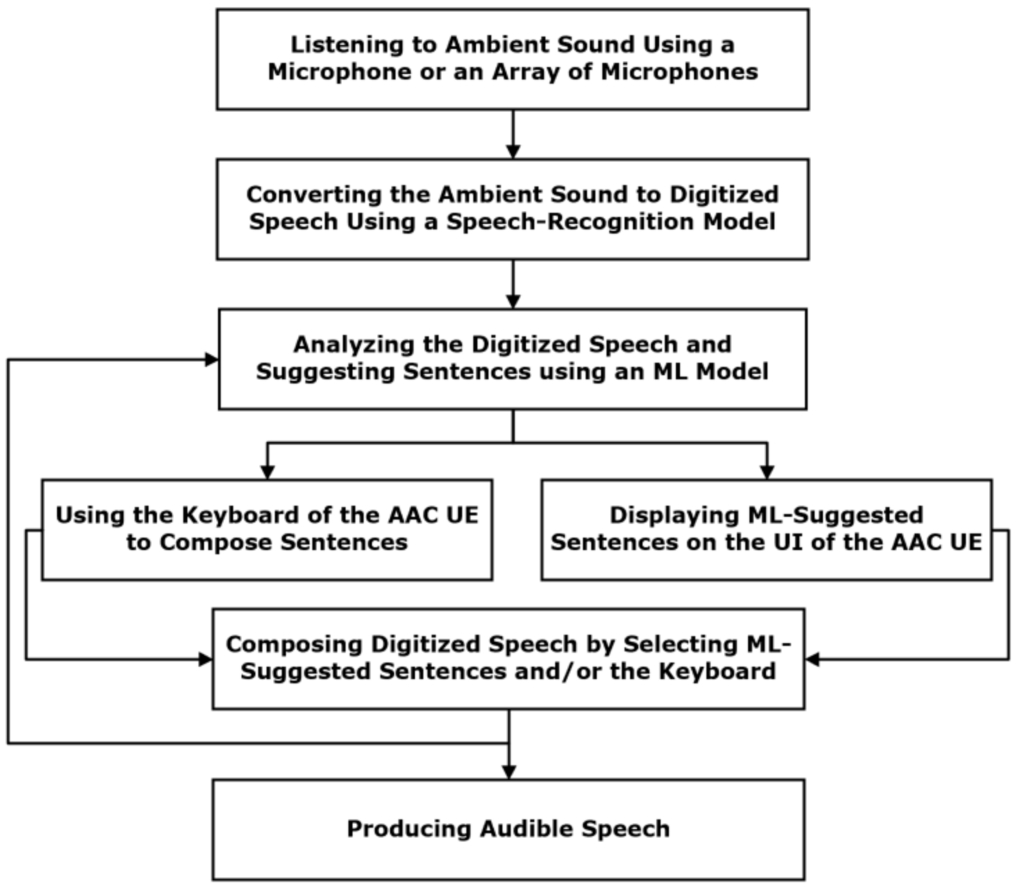

Figure 3

Using a microphone or an array of microphones, the AAC UE scans ambient sounds to identify speech that is being spoken in the vicinity of the user. After the AAC UE scans for speech, the AAC UE converts the audible speech to digitized speech using a speech-recognition model; refer to Figure 3. The digitized speech becomes an input to the ML model. The ML model may be a standard neural network-based model with corresponding layers required for processing input features, such as fixed side vectors, text embeddings, or variable length sequences. In addition, the ML model may be a support vector machine, a recurrent neural network (RNN), a convolutional neural network (CNN), a dense neural network (DNN), heuristics, or a combination thereof. The ACC UE uses the ML model to suggest sentences that the user may want to select, and the UI of the AAC UE displays the suggested sentences; refer to Figure 3. The user reads the suggested sentences and selects the sentences that are applicable to the current conversation. In case the user does not like the suggested sentences, they can use the keyboard to compose a new sentence or to modify a suggested sentence; refer to Figure 3. The digitized composed speech aided by user input is then converted to synthesized speech. In addition, the digitized composed speech becomes an input to the ML model. Reiteratively, the ML model is updated to make a better prediction in future conversations and, thus, speeding up the communication process; refer to Figure 3.

Given the large computational power that machine learning can use to train a model, the model training can be performed on a cloud, server, or other capable computing device or system. Periodic model updates are sent to each user's computing device, which allows the user's computing device to execute the ML model even if that device does not have the resources to update the model itself. Instead or in addition, some or all of the model training can be performed on the AAC UE.

Similar to people who communicate without the aid of an AAC UE, the AAC UE user chooses different sentences depending on the audience. For example, Jane may discuss movies with her friend Tom, talk about soccer with her father and her siblings, discuss personal feeling with her mother, discuss work-related topics with her supervisor and coworkers, discuss physical needs (*e.g.*, needing help to change her clothes) with her caregiver, discuss spiritual topics with her priest, and so forth. The AAC UE can "seed" the ML model with data from a variety of sources depending on the audience of the user. For example, assume Jane is speaking to her priest, the ML model can pre-load a Sunday's sermon, the Bible, or other spiritual literary works. After Mass, the priest may ask Jane, "Jane, what did you think of today's sermon?" Jane may reply using sentences related to the topic of the sermon. Therefore, identifying the audience and the context of the conversation, helps the ML model to make better suggestions, as is illustrated in Figure 4.

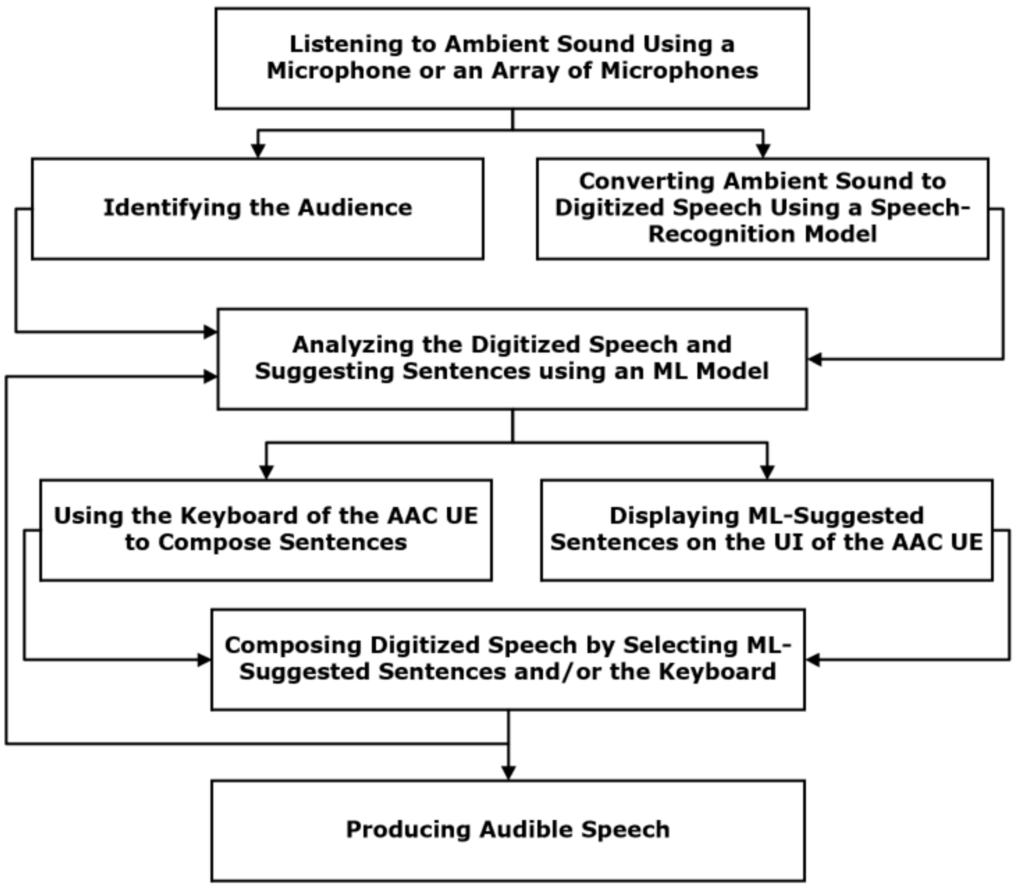

Figure 4

In addition to the features of the AAC UE illustrated in Figure 3, the AAC UE illustrated in Figure 4 tailors the input to the ML model depending on the audience of the user. To identify the audience of the AAC UE user, the AAC UE evaluates each person by employing user input (*e.g.*, the user selects the audience), voice recognition, facial recognition, radar signature, biometric sensors (*e.g.*, a person may scan their thumb on the AAC UE before communicating with the user), media address control identification (MAC ID) (*e.g.*, the AAC UE can scan MAC IDs of smartphones used by the audience), radio-frequency identification (RFID) (*e.g.*, an employee's badge), or other sensors (*e.g.*, in car seats). Once the AAC UE identifies the audience of the user, the ML model can narrow the topic of the conversation and help make better suggestions. As in the example illustrated in Figure 3, the AAC UE illustrated in Figure 4 reiteratively updates the ML model to make better suggestions in future conversations with the identified audience. Revisiting the example of Jane's interaction with her priest—over time, the ML model may learn that Jane's discussions with her priest are related specifically to the priest's sermon and are less related to the rest of the Mass. In that case, the ML model avoids pre-loading suggestions that are related to topics other than the sermon.

Further to the descriptions above, a user or may be provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable collection of user information (*e.g.*, information about a user's social network, social actions, social activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user and their audience, how that information is used, and what information is provided to the user.

In summary, an AAC UE that incorporates a microphone or an array of microphones, a speech-recognition model, the ability to identify the audience, a machine-learned model, the ability for a user to help train the model, and the ability to produce audible speech, can greatly improve the lives of people who live with a speech impairment.

References:

[1] Patent Publication: US20180018144A1. Leveraging environmental context for enhanced communication throughput. Filing Date: July 15, 2016.

[2] Patent Publication: WO2015112250A1. Visual-kinesthetic language construction. Filing Date: November 20, 2014.