# Wordification — Dialectical Speech Synthesis Using Machine Learning Systems (Project Proposal)

Shashank Comandur
Clay Crews

September 5, 2023

## 1 Problem Statement

Wordification is a research project at the University of South Carolina (mentored by Dr. Stanley Dubinsky and Dr. Homayoun Valafar) that both team members are currently developers on. Its goal is to gamify the teaching of the English language to young children, utilizing the linguistic properties of words to instruct in more effective ways. Computerized spelling instruction has proven to be far more effective for young children than rote memorization of words, and this is especially true for speakers of regional dialects, such as African American English (AAE), or Southern White English (SWE).

## 2 Context

The games, in their current form, involve the use of spoken audio of English words. The goal of the project is to address how machine learning systems can be used to create the game audio of spoken words in different dialects, such as AAE or SWE. There are already tools that exist to achieve this goal: see resemble.ai, or speechify.com. However, these tools are proprietary and generally cost a fair share to generate audio at the scale we wish — the database of words utilized by the app contains ∼850 entries, and we wish to generate audio of each word, as well as 2-4 sample sentences for each, thus necessitating the use of a custom model. We plan to read up on how the algorithms behind the aforementioned tools work to see what we can learn, and have also spent time examining the existing published literature surrounding machine learning systems for audio generation.

## 3 Data

In order to generate audio in different dialects (SWE and AAE) using a machine learning model, we need to first create training data consisting of audio of people speaking in those dialects. We aim to sit down with speakers of these dialects for about an hour, use high-fidelity recording equipment to obtain samples of their speech, and of course, compensate them for their efforts. This data will be used in training the model we develop.

## 4 Method

As previously mentioned, we plan to analyze the existing services and recently published speech synthesis literature to develop our approach to this problem. We are currently unsure of whether to develop one, central model for all audio generation, or whether it would be more effective to create a separate model for each dialect. These questions, among others, will be answered as we begin the development process.

## 5 Results

Given the nature of the task, a qualitative evaluation of the output data would be preferred. We will perform a comparative analysis of the model output against the test data and evaluate its effectiveness for use in the games. Because the goal is precise instruction on the dialectical level, we must manually comb through the generated audio and make sure it matches with our expectations for what speech in those dialects should sound like. Quantitative analysis, though not ideal, could be carried out through analysis of the output waveforms, by counting for artifacts, blemishes, and mistakes in the audio.