

SoapBox Labs Fluency Assessment Platform for child speech

Amelia C. Kelly, Eleni Karamichali, Armin Saeb, Karel Veselý, Nicholas Parslow, Gloria Montoya Gomez, Agape Deng, Arnaud Letondor, Niall Mullally, Adrian Hempel, Robert O'Regan, Qiru Zhou

SoapBox Labs, Dublin, Ireland

amelia@soapboxlabs.com

Abstract

The SoapBox Labs Fluency API service allows the automatic assessment of a child's reading fluency. The system uses automatic speech recognition (ASR) to transcribe the child's speech as they read a passage. The ASR output is then compared to the text of the reading passage, and the fluency algorithm returns information about the accuracy of the child's reading attempt. In this show and tell paper we describe how the fluency cloud API is accessed and demonstrate how the fluency demo system processes an audio file, as shown in the accompanying video.

Index Terms: speech recognition, child speech, educational technology, oral reading fluency, child literacy

1. Introduction

In this demonstration we present the SoapBox Labs Fluency platform (SoapBox Fluency), a cloud API service that allows the assessment of a child's oral reading fluency. Fluency assessment is normally a manual process, where a child is asked to read a passage in the presence of a teacher, who makes a note of where the child deviates from the script, say by omitting a word in the passage, substituting one of the words for another, or inserting extra words. The child's reading ability is then expressed in terms of reading rate, and words read correctly. Reading rate and correct words can be combined into a single metric, Words Correct Per Minute (WCPM), calculated as follows:

mistakes = deletion_count + substitution_count

correct_count = word_count - mistakes

reading_time = audio_duration_in_minutes

wcpm = correct_count / reading_time

SoapBox Fluency offers teachers a voice-enabled alternative to manual assessment of fluency for K-5 children. The fluency system is underpinned by models trained using many hours of child speech from real-world situations. These models and data have been leveraged to create an accurate and robust ASR system tuned to the reading domain. The accuracy of ASR in education applications is extremely important as it determines the accuracy of any fluency solution and how well its output will correlate with human assessors. WCPM is calculated by comparing the ASR output to the section of text that was read within one minute, and counting the correct words, and insertions, and penalising for deletions and substitutions. The fluency endpoint returns an in-depth analysis of the audio file compared to the reference text. Data returned includes the respective counts of words read, insertions, substitutions, deletions (omissions), differences (compared to reference text), and correct words. In addition, a breakdown is also provided of each word/token identified in the transcription and reference text, along with an indication of its type (i.e. insertion, deletion,

etc). This allows the source text to then be marked up for review by a teacher for example. By enabling the integration of voice technology into existing formative and summative assessments, SoapBox Fluency allows educators and parents to receive both immediate and longitudinal feedback about a child's progress.

2. Directions for use

The fluency API service accepts an audio file and accompanying reference text (i.e. what the child was asked to read) and decodes the audio file, returning a detailed analysis of the speech contained in the audio file as well as a comparison of the text to the reference text.

SoapBox Fluency is officially released and is generally available as a RESTful Web Service. Once authenticated, fluency requests should be sent via HTTPS to: <https://api.soapboxlabs.com/v1/speech/fluency> The parameters required for a successful response are given in Table 1.

Table 1: Description of fields in CURL request

file	audio file to be analysed
reference_text	text file to compare with audio
user_token	unique id that represents the speaker

```
curl -H "X-App-Key:YOUR_API_KEY_HERE"
-F "file=@AudioFile.wav"
-F "reference_text=reftext.txt"
-F "user_token=abc123"
https://api.soapboxlabs.com/
v1/speech/fluency
```

SoapBox Fluency is a synchronous service and accepts uncompressed WAV audio files with a minimum sampling rate of 16kHz. The maximum duration currently accepted is 60 seconds. The reference_text file should be a plain text file. If the request is successful, SoapBox Fluency will return a JSON response containing the fluency analysis. It will look similar to the following:

```
{
  "audio_duration": 2.744687557220459,
  "user_id": "sandbox",
  "results": {
    "num_differences": 3,
    "substitution_count": 1,
    "insertion_count": 1,
    "correct_count": 2,
    "deletion_count": 1,
    "reference_text": "I like stripes really",
    "text_score": [{
      "transcription_index": 0,
```

```

    "reference_index": 0,
    "alignment_type": "CORRECT",
    "reference_word": "i",
    "transcription_word": "i"
  }, {
    "transcription_index": 1,
    "reference_index": 0,
    "alignment_type": "INSERTION",
    "reference_word": " ",
    "transcription_word": "don't"
  }, {
    "transcription_index": 2,
    "reference_index": 1,
    "alignment_type": "SUBSTITUTION",
    "reference_word": "like",
    "transcription_word": "hate"
  }, {
    "transcription_index": 3,
    "reference_index": 2,
    "alignment_type": "CORRECT",
    "reference_word": "stripes",
    "transcription_word": "stripes"
  }, {
    "transcription_index": 4,
    "reference_index": 3,
    "alignment_type": "DELETION",
    "reference_word": "really",
    "transcription_word": "<eps>"
  }
}, {
  "transcription": "i don't hate stripes",
  "word_count": 4
},
"language_code": "en-GB",
"result_id": "sandbox-1_1591291343755",
"time": "2020-06-04T17:22:24.828Z"
}

```

3. Demonstration of fluency assessment

For this demonstration we use some recordings of children reading passages from the Dynamic Indicators of Basic Early Literary Skills (DIBELS) program [1], aimed at children between 6-10 years of age. The children are provided with the DIBELS reference text and are asked to read it. The file is submitted to the API and processed. The results are returned and the data points provided in the JSON are then used to highlight the words in the reference text: a blue token indicates an insertion, a yellow highlighted word indicates a substitution, and a red highlighted word indicates an omission.

3.1. Fluency demo interface

In the first example shown in our demo video, the child reads the text almost perfectly. Our system successfully detects two substitution errors. The information returned in the JSON results object are displayed on screen by highlighting the substitutions and displaying them in a count at the bottom of the screen. Also shown are counts for correct words, omissions, substitutions insertions, total words, audio file duration, and WCPM. WCPM is not read directly from the JSON, but rather calculated by using the number of errors, the number of words read, and the duration of the audio file. This allows users to be flexible with their definition of WCPM, for example including a penalty for insertions.

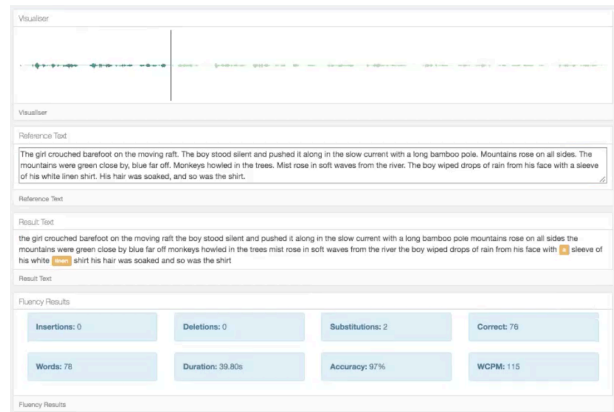


Figure 1: Example of the fluency assessment

3.2. Insertions, deletions, substitutions

In the next example we demonstrate how the system handles deletions, substitutions and insertions, by creating a situation where the audio differs from the text. In this instance, since we use the same audio file, we change the reference text that the child was supposed to read in order to make the comparison. The number of insertions, substitutions and deletions has increased, and this is reflected in the WCPM score.



Figure 2: Fluency with insertions, deletions and substitutions

4. Conclusions

Assessing a child's reading fluency is a manual and time consuming task for a teacher, making continuous assessment impossible given teacher work-load and class sizes. Using an ASR-based solution not only removes the teacher's workload, but allows the teacher to continuously and passively evaluate a child's reading progress. The data points returned from the SoapBox Fluency system provide all the information needed to drive a teacher dashboard or create a report of the child's reading fluency.

5. References

- [1] R. H. Good, R. A. Kaminski, S. Smith, and D. Laimon, *Dynamic indicators of basic early literacy skills: DIBELS*. Dynamic Measurement Group, 2002.