Measuring prosody in child speech using SoapBox Fluency API

Mauro Nicolao, Brenda McGuirk, Declan Moore, Niall Mullally,
Lora Lynn O'Mahony, Emma O'Neill, Amelia C. Kelly

SoapBox Labs, Dublin

`hello@soapboxlabs.com`

Abstract

SoapBox Fluency API uses automatic speech recognition (ASR) to return data points that enable educators to assess a child's oral reading fluency (ORF). Prosody is a key component of ORF assessment together with accuracy and speed, but manual assessment of prosody is largely subjective. A quantitative prosody score would benefit fair assessment. In this show and tell paper and accompanying video, we describe the new prosody feature of the API and demonstrate how the outputs can be used to assess expressiveness in ORF.

Index Terms: automatic speech recognition, child speech, edtech, oral reading fluency, child literacy, prosody

1. Introduction

SoapBox Fluency is a speech solution designed specifically for children's speech and can be used by educators to assess children's ORF in both formal and informal settings. Given a text passage and an audio file containing a child reading that passage, the system will return an analysis of the audio file in reference to the text passage. SoapBox Fluency returns several data points, such as duration; phoneme-level confidence; and insertion, deletion and substitution information. These are designed to be condensed into high-level scores and metrics that are used in ORF assessment, such as reading accuracy and speed (words correct per minute, WCPM).

Prosody is linked to comprehension and can be measured by examining the reader's expressiveness, specifically information about their intonation, emphasis and phrasing. Educators face significant challenges in capturing prosody data, for example:

- It's difficult to simultaneously track accuracy and prosody during observed assessments.
- Applying an evidence-based approach to measuring prosody is a challenge because it's highly intuitive by nature.
- Prosody skills develop at a different pace for kids than other fluency skills, requiring independent progress reporting.

SoapBox Fluency addresses these challenges, and supports creators of education technology by automating prosody scoring and generating at-a-glance prosody data that's actionable in the classroom.

1.1. Prosody features in SoapBox Fluency

SoapBox Fluency can return several prosody-related data points, as demonstrated in the example shown in Figure 1.

`pitch`: The fundamental frequency (f_0) values for that word.

`time_since_previous`: Reports the number of seconds


```
    "confidence": 99.51,
    "duration": 0.75,
    "end": 2.79,
    "phone_breakdown": [{...}],
    "phonetic_transcription": "l ay k",
    "pitch": {
        "values": [289, 299, 306, 306, 299, 288, 274, 266, 259, 253, 243,
        244, 259, 253, 246, 242, 237, 232, 230, 230, 228, 226, 225, 223,
        221, 221, 221, 222, 222, 222, 220, 217, 214, 213, 213, 213,
        205, 190, 190, 188, 196, 204, 211, 215, 219, 221, 222, 223, 223,
        223, 223, 223, 224, 225, 226, 228, 229, 230, 231, 232, 233, 234,
        236, 237, 238, 239, 240, 242, 243, 244, 245, 246, 248, 249]
    },
    "start": 2.04,
    "time_since_previous": 0.24,
    "word": "like"
}
```

Figure 1: *Prosody data points in Fluency JSON output.*

from the end of the previous word to the start of the current word.

`punctuation`: Any punctuation marks detected in the reference text

These data points can be either used individually, or combined in order to derive a quantitative comprehensive analysis of a read passage, depending on a user's specific needs.

We faced the following challenges in quantifying and surfacing this prosodic information:

- The lack of a unanimous prosody definition outside expert circles. Educators often follow an evidence-based approach to measuring prosody.
- The subjectivity of prosody and the resulting difficulty of finding expert-validated prosody annotations.
- The multi-dimensionality of prosody – the interpretation of prosody data is just as important as collecting the data points themselves.
- Translating the complexity of some prosody-related measurements (e.g., pitch signal) into machine-readable speech data points.

2. Use cases for prosody

In this section we describe some common use cases of the prosody feature.

2.1. Expressiveness

Expressiveness is a measure of how much a student's pitch modulates while reading. High expressiveness can indicate the ability to control pitch modulation, engagement in the reading content, and understanding and conveying meaning through emphasis on the appropriate words. Low expressiveness, on the other hand, can indicate a student is struggling or has yet to fully grasp the content of the passage.

SoapBox Fluency can be used to highlight the contour of a student's pitch in an audio file by returning data points on
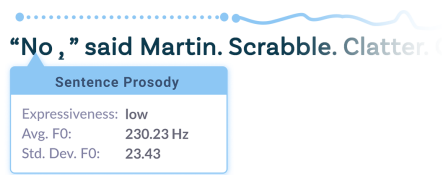
Figure 2: *Example of student's expressiveness assessment.*

pitch measurements at fixed intervals and timestamps, to indicate whether pitch changed, at which point in a sentence, and at what rate. This is shown in Figure 2. All the visuals in the Figures report elements of the reading assessment exercises in the live demo user interface used to showcase SoapBox Fluency API capabilities.

2.2. Phrasing

Another important aspect of prosody is ensuring that a student groups relevant words together by pausing at appropriate points in a passage, typically demarcated by punctuation (e.g., periods, commas, question marks, and exclamation marks).

SoapBox Fluency can return accurate time alignment of speech and text, which allows to calculate how long, in seconds, the child paused between a word with a punctuation mark and the next word. Pause thresholds can be customized for a given task to identify whether a student paused appropriately at particular punctuation marks. Data for pauses and hesitations across a sentence or passage can be matched to prompt text or modeled oral responses.

When pauses are matched to punctuation and logical breaks, this can indicate understanding and fluency. When pauses are too frequent or not matched to punctuation, it may be that the student is struggling to decode or is reading with less fluency and prosody than expected at their grade level.
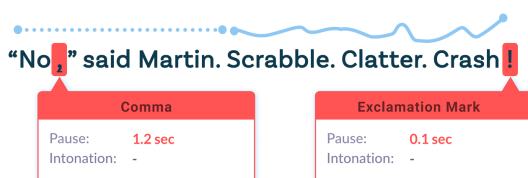

Figure 3: *Example of punctuation pause assessments.*

2.3. Punctuation and expressiveness

Ensuring that a student varied their intonation at appropriate points in a passage is another important prosody measure. For instance, when reading a sentence that ends in a question mark the intonation may be expected to rise. Similarly, for a sentence ending in a period, the intonation would be expected to fall.

SoapBox Fluency can measure student's intonation at punctuation boundaries. Once a punctuation boundary is identified, the pitch values can be used to check if the student raised or lowered their intonation as expected, and the slope of a pitch value array can be derived.

Once the slope is calculated, a threshold can be used to determine if the incline or decline was sharp enough to qualify

as a rise or fall in intonation. Thresholds (on the slope trend) can be adapted according to the different use cases and reader proficiency.

Figure 4: *Example of intonation assessment on question marks.*

2.4. Mapping prosody to a rubric

Prosody rubrics, such as the widely used NAEP[1] rubric, provide quick, clear, and actionable feedback for educators. Figure 5 shows a hierarchical sample scheme to support users in translating prosody data into:

- scores for individual prosody measurements (e.g., expressiveness or phrasing).
- a rubric score (an aggregate of multiple prosody measures).

This provides flexibility and choice in how users can generate and surface prosody scores.
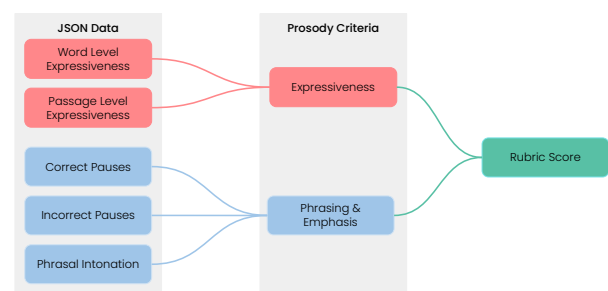

Figure 5: *Example hierarchy to calculate prosody rubric score*

3. Conclusion

In this Show and Tell paper and accompanying video, we present a live demo of the features of SoapBox Fluency API that enables users to measure prosody aspects of a child's ORF. The system returns granular pitch and phrasing data, which can be combined to create a comprehensive analysis of a read passage, or used individually. Education companies and research groups can use this data to:

- Automatically and objectively measure the quality of a student's prosody.
- Gauge whether a student's prosody performance is at expected level.
- Map a student's score to a prosody rubric as well as surface the most granular aspects of the prosody production down to word and phoneme level.
- Surface prosody data directly to teacher dashboards, eliminating the need for manual assessment.

---

[1] https://nces.ed.gov/nationsreportcard/studies/orf/scoring.aspx