Computational Quantification of Oral Fluency in English

Abdur Rahman[1], Subaraksha R T[2]

Research Scholar, Vellore Institute of Technology, Vellore - 632014

Vellore Institute of Technology, Vellore - 632014

Abstract

The research paper aims to propose a base for the measurement of oral fluency through computational analysis in English Language. The analysis of speech sounds could be dissimilated into multiple variables to convert the speech sounds into decimals. The proposed idea could provide the future researchers and scholars a way to analyse and evaluate Oral Fluency through multiple variables such as words per minute, pauses, fillers, reputations and syllables per minute and segments. The tool could give feedback on specific areas of speech sounds, therefore supporting the researchers with a detailed output. The goal of the proposal is to provide a valid and reliable tool for the measurement of fluency.

Keywords: Oral fluency, Measurement, Computational, English as a Second Language, Speaking Fluency

Introduction

Oral Fluency in English is a broad research area which is being researched by many researchers all over the world. According to BBC, fluency of the people is measured through 'speech rate' and 'utterance length'(Sandoval, n.d.). Fluency, on the other hand, does not only incorporate speech rate and utterance length, it deals with multiple variables of speech including grammar, pronunciation, words per minute (WPM), pauses, prop words, repetitions,and hesitations. It is a tedious task to measure the fluency level of learners manually, especially in research with large sample size. Research conducted previously with manual measurement have made a mention of the difficulty in measuring fluency without computations (Podesva & Sharma, 2014). Even if it is done manually, it is prone to accurate results of the analysed data. Availability of other softwares such as Cool Edit 2.0, Pratt, Audacity, and Adobe Audition focus on analysing the audio rather than analysing fluency. There is no particular application or software available for evaluating and analysing the fluency of a person in English.

English being a global language has four main language skills (LSRW). Among which speaking skill is given more importance because of its need in the employment market. Considering its significance, multiple researches focus on the skill of speaking (Aziz &Kashinathan, 2021). Fluency is defined through multiple philosophies And so, identification of the appropriate philosophy to adhere is quite a challenge for the formulation of assessment tools. A software that could import, analyse, process the audio with the given set of procedures and give the user an end result will immensely benefit the contemporary and future researchers who work in the field of applied linguistics.

Researchers use various methods to evaluate fluency. Some researchers take in the number of words along with pauses and fillers to carry out the analysis and some look out for mispronounced words, syllables and length. And so there is no standard method which is universally accepted. Formation of a software (tool) without a standard method in existence might only benefit a few researchers. Researchers who disagree with that philosophy of analysing might find the software partially futile. A plausible solution to the above mentioned limitation could let all the researchers, who investigates on Fluency, analyse their samples with this software that provides multiple sets of variables and allows the users to create a custom set of variables. The end result of the analysis will be customised according to the customised preset of the researcher.

Given set of Variables
I. Number of Words uttered per Minute (WPM)
II. Transcription
III. Pronunciation
IV. Filler words
V. Proper nouns
VI. Repetitions
VII.   Pauses
VIII.   Syllables
IX. Segments
X. Grammar

XI. Transcription

(Tag: Window 2), (automatic, manual)

The imported audio will be opened in a window which will have a Oscilloscope and a text tool bar. The words uttered in the audio will be automatically displayed in the text tool bar. This will be done through speech recognition. The text appeared through speech recognition will have options paving way for

manual edit to modify the uttered words in case of an error in speech recognition. When the subject's level, accent, and pronunciation of English differs, there are possibilities for misinterpretation of words. However, this can be nullified through manual edit option. The researcher should be able to manually edit the text that appears in the text toolbar after automatic transcription. (Auto-corrections will not be made while transcription)

Tools required

(a) Oscilloscope - A device that is used to visualise audio in the form of waves.

(b) Text tool bar - The text tool bar, in this context, will contain text and this tool bar should be editable.

(c) Speech recognition - The ability of the computer to comprehend and capture the audio waves produced through human speech.

Options on the window

(a) Automatic transcription - Once clicked, will transcribe the audio format into text

(b) Manual edit - once clicked, the text will become editable

(c) Proceed - Once clicked, Makes it to the next window

II. Words per Minute (WPM)

(Window 3), (automatic)

The number of words and characters in (window 2) will be counted and shown in the 'Total number of words' and 'Total number of characters'. An 'edit' option will be made available to customise the numbers. But the 'View' option in this window will let the researcher hide or unhide the 'Total number of characters' toolbar. The total number of words per minute will be calculated through the number of words uttered per minute. If the audio exceeds one minute, the respective formulas will be used.

$$n = Number\ of\ Words$$
$$V = Variables\ in\ Time\ (Number\ of\ minutes,\ 1\ minute = 60\ seconds).$$
$$WPM = Words\ Per\ Minute$$
$$WPM = \frac{n}{v}$$

III. Pronunciation

(Window 4), (automatic, manual)

Mispronouncing words can lay as a barrier to effectively communicate and therefore, this could cause miscommunication or misinterpretation. A person cannot be considered fluent if s(he) happens to mispronounce all the words. So, there should be a count to measure the total number of words mispronounced to evaluate and analyse oral fluency of a person even more accurately. Total number of mispronounced words will be calculated by finding out the number of words that mismatches with the auto transcribed text (unedited text in window 1). Once the mismatch is found the user can listen to the right pronunciation and the actual pronunciation of the subject. This will further help the user to evaluate

it. If the user still finds the mismatched word to be correct. He/she will have an option to ignore the mispronounced word.

The mismatched words will be found by the analysing two different datasets of texts. The first data will have manually edited words uttered (data 1). And the second data will have the words recognised through speech recognition (data 2).

For example,

(*A single unit from Data 1 will be closed when it finds an equivalent in Data 2. If the word 'tree' is present twice in data 1 and only once in data 2, it identifies the mismatch of one unit)

Common words = mw

Mismatched words = mw

Dataset of (Data 1) - { He, is, a, good, boy, and, he, is, very, smart }

Dataset of (Data 2) - { **She**, is, a, **food**, boy, and, he, is, **berry**, smart }

Total number of units (TNU) in Data 1 - (10)

cw = (7)

mw = (3)

(*Therefore;* TNU = (cw + mw)

Total number of all mispronounced words = (TNU - mw); (10-3) = 7

Tools required

(a) Oscilloscope - A device that is used to visualise audio in the form of waves.

(b) Text tool bar - The text tool bar, in this context, will contain text and this tool bar should be editable.

(c) Speech recognition - The ability of the computer to comprehend and capture the audio waves produced through human speech.

Options on the window

(a) Manual markings - This will let the user to manually mark the words mispronounced.

(b) Automatic identification - This will let the user instantly identify the words mismatched.

IV. Proper-nouns

(Window 5) (manual)

A noun that names a particular place, thing or a person is called a proper noun. Most of the Proper nouns can be pronounced in any way. For example the proper noun 'Ramesh' can be pronounced as 'Ra-meah-sh', 'Ramesh', 'R-uh-m-ai-sh', and 'Rama-M-SH'. Most of the proper nouns have different pronunciation with no particular correct way of pronouncing it. And so mispronunciation of Proper nouns should be neglected from the total numbers of words mispronounced.

n = Total number of all mispronounced words

V = *Total number of proper nouns mispronounced*

$$d = Total\ number\ of\ mispronounced\ words$$
$$d = (n-v)$$

V. Filler-words

(Window 6) (automatic, manual)

A filler word is often used by the speakers to avoid pauses and hesitations while speaking. It is also habitual for some speakers. A filler word has little or mostly no semantic content in a sentence. For example: (umm, Uh, you know, okay, like, right). The software will automatically identify words that are repeated more than twice. It will then be manually approved as a filler word by the user in order for the software to calculate the number of filler words used in the entire text.

Sample paragraph: { Oral communication is an important skill because it promotes leadership and confidence. **You know**, When children are confident **in** their abilities they will succeed academically and socially. **You know**, It is important for students to generate their own ideas using outside references **in** the community. **You know**, Oftentimes some students have trouble making friends. If they can learn to be more confident **in** their speaking abilities they will be more likely to approach their peers. **You know**, On the other hand, when students feel comfortable expressing their ideas **in** front of people they will most likely contribute more to class discussions.}

Fillers identified = { You know, in }
Approve as Filler = { You know }
Reject as Filler = { in }
Fillers Per Minute (FPM)
Total number of Effective words used = (WPM - FPM)

VI. Repetitions

(Window 7), (automatic, manual)

Words that are repeated continuously adding no meaning to the context can be considered as Repetition. This will be identified by spotting out words that are repeated continuously. For example: { He **is isis** a good boy }: Here, the word 'is' is repeated consequently right next to each other twice. This will be spotted and identified as Repetition. The repeated words will be removed from the effective Number of Words Uttered in the analysis. However, not all the words that are repeated will be neglected. Whenever repeated words are found, the first word will be still counted effective. But when the word is used more than once, it will be removed. *(1<=Repetition).*

However, there is a possibility for the continuous words to give meaning as well. If the continuous usage of the same word gives meaningful context to the sentence, it will not be considered as repetition. This can be manually ignored from repetition. For example { No sentence starts with **because becausebecause** is a conjunction }: In this sentence, though the word 'because' is used consequently next to each other twice, the sentence is still meaningful and it is used contextually. In this case, these
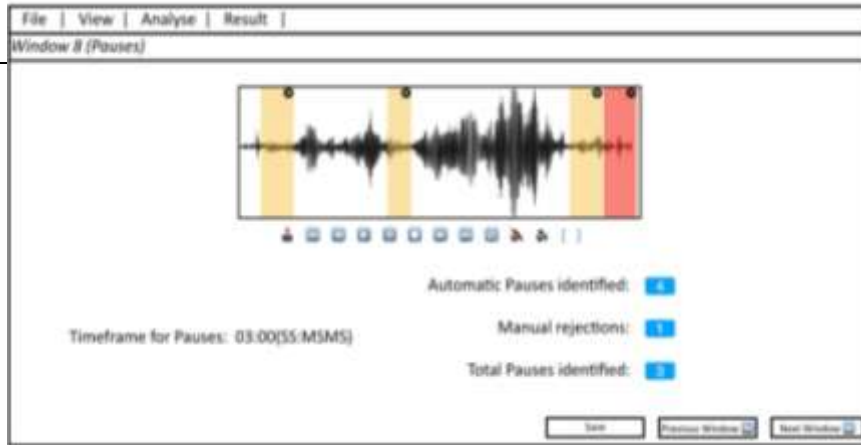
repetitions will be manually ignored. If the same paragraph has another 'because' with no context it will be taken as a repetition too. And will be coded as 'because^{2}'.

$$\text{Repetitions identified} = \{ \text{is, because} \}$$
$$\text{Number of times 'is' identified} = \{ 2 \}$$
$$\text{Number of times 'because' identified} = \{2\}$$
$$\text{Number of times 'because}^{2}\text{' identified} = \{2\}$$
$$\text{'because' can be ignored manually as it has context}$$
$$\text{'is', 'because}^{2}\text{' is manually approved as Repetition}$$
$$\text{Total number of words repeated} = 4$$
$$\text{(TRW) - Total number of repeated words}$$
$$\text{Total number of Effective words used} = \text{(WPM - TRW)}$$

VII. Pauses

(Window 8), (automatic, manual)

Pauses are the gaps that speaker makes in between words or sentences. Pauses are important but too much of a pause makes the speech influent. A pause between utterances could help the speaker communicate effectively, however, if a person speaks with no pause between utterances, it becomes hard for the receptor to understand it. For example, 'Iamveryveryhappytoday' is a group of words that has no gap in between. But the gap while pronouncing two words or two sentences should not be long enough to be noticed. The pauses is required to differentiate but not to delay speech. Pauses can make and break a conversation. Long pauses can break a conversation and very short pauses may make the speech difficult to comprehend. To some speakers more than a quarter or half a second is considered a pause and to some speakers four seconds is considered a pause. Considering variance in opinion in measuring fluency, the software will be customisable between (0:01 - 7:00) seconds to consider a stop as a pause. The processing can be manually done or through the help of Oscilloscope and Artificial Intelligence of identifying silence/pause through low pitch in voice for the specified filtered time.

$$\text{A pause} = \{ \text{(time specified)} <= \},$$
$$\text{sample [time specified 3 seconds]}$$
$$\text{A pause} = \{ 3.00, 3.01, 3.02, < \}$$
$$\text{Total number of Pauses (TNP)}$$

VIII.   Segments

(Window 9), (Manual)

Segments are sentences or clauses that are produced in an utterance or in a single breadth. It can either have a complete meaning or might be incomplete. These will be marked manually by the user to identify the segments of the speech. These segments are coded to identify the grammatical structure of the language. Every segment will be coded through numbers. For example {There is a cow[1] it has four legs and two horns[2] it is very very small[3]}. The above paragraph will be individually coded with a period (.) in the end and a capital letter in the start.

{There is a cow it has four legs and two hornsit is very very small}

↓

{There is a cow[1] it has four legs and two horns[2] it is very very small[3]}.

↓

{There is a cow.}, {It has four legs and two horns.}, {It is very very small.}

IX. Grammatical check

(Window 10), (Automatic, Manual)

Grammatical check without punctuations and proper identification of sentences could not be done through computer automation alone. To counter this, in the previous segment, the entire speech got classified as sentences with capital letter in the start and a period in the end. This will form sentences through segments from the entire speech. Each sentence will be analysed separately through grammar check software and the number of grammatically wrong sentences will be identified. The wrong sentences will appear before screen with two options available.

Option 1: To ignore the grammatical error that gives enough significance. There is a possibility for the computer to mark a properly communicated sentence as a statement of error. To avoid this, the user will be given the leverage to avoid the grammar check manually.

Option 2: To approve the segment indicating the grammatical error in the sentence.

X.   Syllables

The number of syllables in the speech will be identified through an existing python package. This is done to identify the speech rate of the given speech. Most of the temporal variables of fluency are measured through Syllables Per Minute (SPM). And thus, this is not a crucial aspect to look into while measuring fluency.

Window 1:

Window 1 has two main aspects for the users to analyse and evaluate the end result of Oral Fluency. Two toolbars will be available. One is the time frame of the audio and the other one is the inclusion or exclusion of variables.

Option 1: (Time Frame: {1:00; 2:00; 3:00; 4:00; 5:00})

Option 2 : Will have a set of variables, among which the user will enable the needed and disable the unnecessary elements of his study.

Given set of variables are as follows in order;

- ☐ Transcription *(default)*
- ☐ Words Per Minute *(default)*
- ☐ Pronunciation *(optional)*
- ☐ Proper nouns *(default when pronunciation is enabled)*
- ☐ Filler words *(optional)*
- ☐ Repetitions *(optional)*
- ☐ Pauses *(optional)*
- ☐ Segments *(optional), (default to grammar check)*
- ☐ Grammar check *(optional)*
- ☐ Syllables

Conclusion

The proposed assessment tool for the measurement of oral fluency is a reliable model that was specifically developed based on the requirement of existing research. The idea behind the model could instigate the development of multiple alternative models and tools in upcoming research. With the given set of variables (Transcription, Words Per Minute, Pronunciation, Proper nouns, Filler words, Reputations, Pauses, Segments, Grammar check and syllabus), a number specific temporal variables of both fluency and accuracy could be derived. The future researchers could work on the acoustical analysis of speech sounds to have a in-depth analysis of speech samples. Despite availability of softwares that could measure speech sounds such as PRAAT and Cooledit 2.0, there is a need for tools that are more accessible and user-friendly. Though the said softwares give in-depth analysis, these are extremely complicated softwares for language teachers. Thus, software models that are user-friendly and efficient enough to derive variables from speech samples could help track the progress or level of second language learners.

Reference

1. Sandoval, E. (n.d.). *How do we measure language fluency?* BBC Future. https://www.bbc.com/future/article/20190903-linguistic-fluency-proficiency-second-language-learning
2. Aziz, A. A., &Kashinathan, S. (2021). ESL Learners' Challenges in Speaking English in Malaysian Classroom. *Development*, *10*(2), 983-991.
3. Podesva, R. J., & Sharma, D. (Eds.). (2014). *Research methods in linguistics*. Cambridge University Press.