

Towards the Automated Evaluation of the Oral Components of the CXC Examinations: The ASR / AI Approach



James N. Carmichael

Presented at the ACTT Conference: 'Future of Online Learning Quality - Best Practices in Artificial Intelligence and Competency-Based Credentialing'

Marking CXC Oral Exams: 2nd Opinion Needed?

- Unlike some international exam boards – e.g. IELTS – the Caribbean Examinations Council (CXC) does not currently mandate any systematic audio recording of candidates' responses when attempting the reading and speaking tasks for the CSEC Oral Exams.
- In the event of any disputes / demands for re-evaluation of a candidate's oral performance, all that can be done in the absence of any audio recordings is to tot up the original "raw" scores again.
- So, what solutions can we offer in the case a candidate wants a re-marking of a CSEC oral examination attempt?

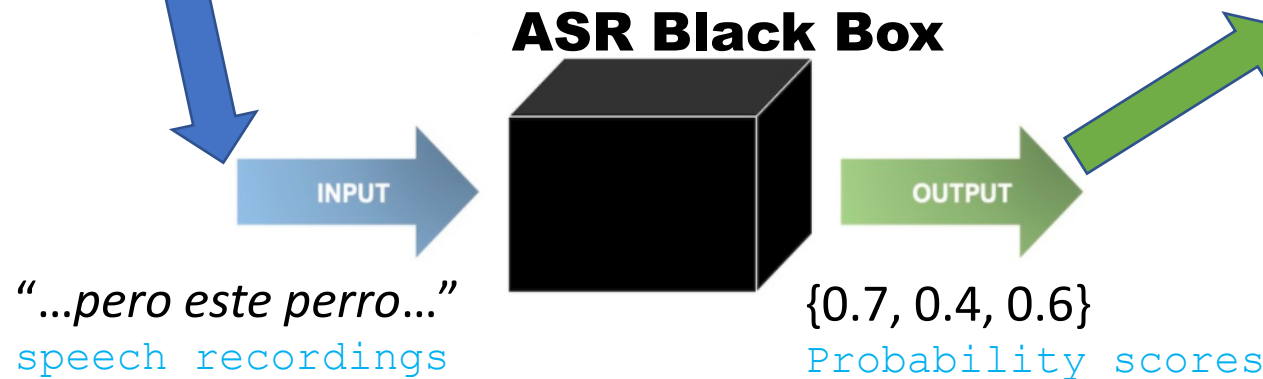
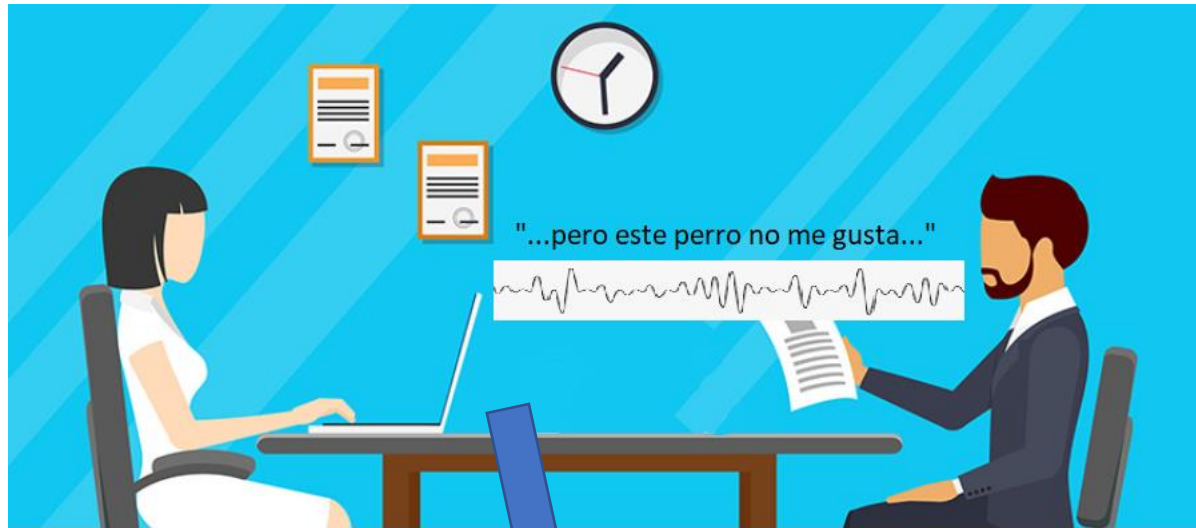
Proposed Solution: ASR-Assisted Automated Evaluation

- [Most Basic Solution](#): Make analog recordings of all attempts to be stored in some central repository .





- [More Advanced Solution](#): Make computer-based *digital* recordings of oral exam attempts to facilitate data transfer, but there's still the issue of overcoming the inherent inconsistencies of *psychometric* evaluation.
- [Our Proposed Solution \(for objective evaluation\)](#): Using *automatic speech recognition* (ASR) technology for automated evaluation of the digital audio recordings.

Typical ASR Automated Evaluation Scenario under Real-Life Exam Conditions



Assessment Report of Oral Exam Reading Passage

1. List of words mispronounced:
"perro" and "me" [Click on microphone icons to hear mispronounced words]  
2. Conversion score: 6 / 10

ASR Forced Alignment Evaluation of Speech: An Overview (1)

HMM-Based ASR speech recognition is a process which....



- i. ...builds “every-man” statistical models of isolated words or connected speech utterances by collecting a lot of speech samples and then training the ASR engine on the speech recordings in the *training corpus*.
- ii. ... tests the speech recognition accuracy of the ASR engine by having it process a collection of unseen speech samples in the *testing corpus*.
- iii. ... after demonstrating an acceptable level of speech recognition accuracy, the ASR algorithm is now fed either live or recorded incoming CSEC oral exam speech utterances for processing.

ASR Forced Alignment Evaluation of Speech: An Overview (2)

In a typical real-life CSEC oral exam, the ASR speech recognition engine would process the incoming stream of speech to decide...

- i. ... which incoming word or word sequence most closely matches some specific vocabulary item(s) in the ASR app's vocabulary list.
- ii. ... generate a list of probability scores which indicate how well the incoming speech sample aligns with each and every item in the ASR vocab list, for example .

iii. **Incoming speech sample:** 

- Probability that incoming speech sample represents  [perro] (probability score: 0.6)
- Probability that incoming speech sample represents  [pero] (probability score: 0.7)
- Based on the higher probability score of 0.6, the ASR deems the speech sample to be “*pero* [=but]”

ASR Forced Alignment Evaluation of Speech: An Overview (3)

- If it is known beforehand what the words in the speech sample should be (i.e. if the speaker is reading aloud from a prepared script), then we can configure the ASR engine to operate in *forced alignment* mode and use the resulting probability scores as a quantitative scoring tool to grade the candidates' oral performance in the CSEC reading exercises.
- NOTE: It would be necessary to conduct a standardization exercise to determine the correlation between the assessment of expert human assessors and automated ASR evaluation of candidates' attempt at the CSEC reading exercises.
- If there is a positive correlation between expert human and automated ASR assessment, then let's go for a pilot study!
- **DISCLAIMER**: The ASR evaluation should only be used in the event of a dispute concerning the evaluation of the primary (human) evaluator and / or if there is a discrepancy between the evaluations of the 1st and 2nd human markers.

Issues Concerning the Configuration of the ASR CSEC Evaluation Software

- Should the CSEC ASR Evaluation be trained exclusively on speech data from native speakers?
- How do we normalize for differences in microphone quality and ambient room noise which may significantly vary between oral examiners and the exam rooms where they will hold the oral examinations?
- When, where and how often will we hold training & standardization workshops to teach the unfamiliar technology to CSEC oral examiners?

Further Reading

- Carmichael, James. "Using articulatory visualisation techniques to improve pronunciation." *2013 International Conference on Current Trends in Information Technology (CTIT)*. IEEE, 2013.
- Benzeghiba, Mohamed, et al. "Automatic speech recognition and speech variability: A review." *Speech communication* 49.10-11 (2007): 763-786.
- Eskenazi, Maxine. "Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype." *Language learning & technology* 2.2 (1999): 62-76.

Thanks for Your Time!

