An Evaluation of Expedited Transcription Methods for School-Age Children's Narrative Language: Automated Speech Recognition & Real-Time Transcription

Carly Fox, M.S.
Utah State University

Introduction

Language sample analysis (LSA) is a critical component in conducting a comprehensive assessment of developmental language disorders (DLD)

- **Many clinicians report** that LSA is too time-consuming
- **To reduce the time-cost** many clinicians practice real-time transcription (RTT)
- **There is limited evidence** for the efficacy of RTT
- **Automated Speech Recognition** (ASR) may serve as an alternative means of expedited transcription

The aims of the current study were to **1)** evaluate the accuracy of RTT from both clinicians and trained transcribers (TT), **2)** compare the accuracy of RTT and ASR produced transcripts, and **3)** evaluate the reliability of LSA indices produced from each transcription type

Table 1- Word Error Rate by Method

Transcription Method	Mean	Median	Min	Max
ASR (n = 42)	.30 (.11)	.30	.08	.51
S-RT (n = 42)	.42 (.19)	.40	.11	.83
T-RT (n = 41)	.43 (.19)	.45	.10	.74

Note. ASR = automated speech recognition, S-RT = real-time transcription, clinician, T-RT = real-time transcription, trained transcriber. The higher the WER, the lower the transcription accuracy.

Methods

A total of 14 participants (clinicians = 7, TTs = 7) took part in this study. Each were asked to transcribe 6 narrative language recordings in real-time. Recordings were elicited from school-age children (7-11 years) with DLD. The same 42 recordings were also transcribed with Google Cloud Speech ASR.

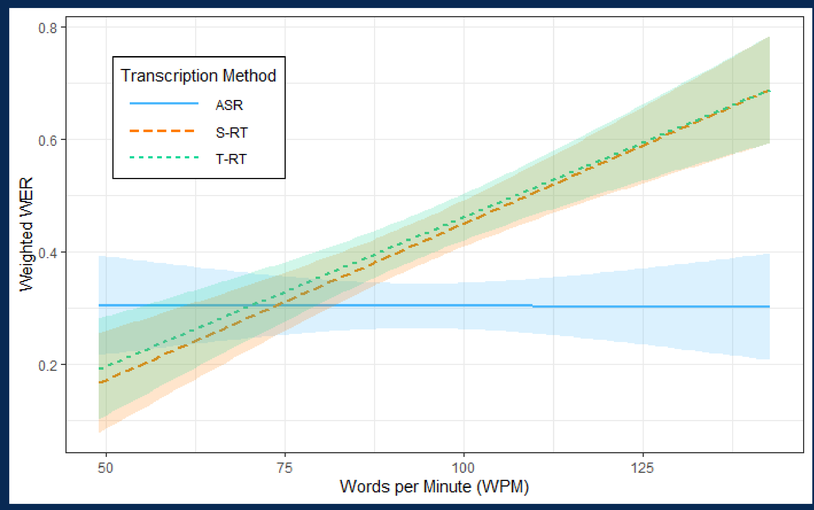
1. **Screen recordings** for quality and length
2. **Produce** ground-truth reference transcripts
3. **Evaluate** transcription accuracy of each method + accuracy of LSA indices

Table 2- Reliability on LSA Indices

LSA Index	ASR	S-RT	T-RT
Number of Utterances	.99	.77	.82
Mean Length of Utt.	.94	.80	.74
Lexical Diversity	.98	.72	.78
Number of Words	.98	.66	.76
Type-Token Ratio	.87	.74	.71

Note. Interrater reliability of LSA indices produced using each transcription method with the reference transcripts was determined via Pearson moment product correlation coefficients. ASR = automated speech recognition, S-RT = real-time transcription, clinician, T-RT = real-time transcription, trained transcriber.

Figure 1 – Cross-Level Interaction



Note. ASR = automated speech recognition, S-RT = real-time clinician, T-RT = real-time trained transcriber. The higher the word error rate (WER), the lower the transcription accuracy.

Results

Multi-level analyses indicated significant differences in transcription accuracy between methods (ASR, RTT) moderated by speech rate. ASR had the lowest WER (*M* = .30, *SD* = .11) and was the only method not significantly impacted by speech rate.

Correlation analyses revealed LSA indices were most reliable on transcripts produced with ASR

Conclusion

ASR outperformed RTT for both transcription accuracy and reliability of LSA indices. ASR may serve as a better option for clinicians looking to reduce the time associated with LSA for their school-age clients. Additional research is needed to determine whether these findings generalize to other populations of interest (e.g., at-risk children, older age-range, etc.).

UtahStateUniversity

Carly Fox
Utah State University
Special Education & Rehabilitation
Carly.fox@usu.edu