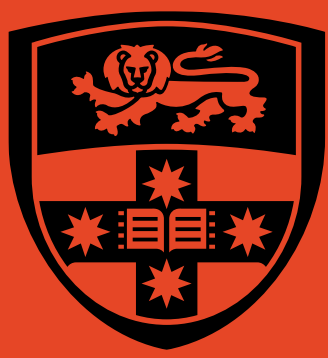




Investigating Speakers' Performative Vocal Behaviour in Manual Versus Automated Voice Data Collection Environments

Rebecca Li, Brian Stasak, Cate Madill

Voice Research Laboratory, Sydney School of Health Sciences, Faculty of Medicine and Health, NSW Australia

Introduction

The human voice is a complex biosignal carrying information about a person's health, mood, and demographic attributes. Gathering high-quality voice data can help advance speech processing technologies in many contexts, such as practical use in automated healthcare tech^[1], migration to automated voice applications^[2], voice biometrics^[3], and speech research utility (e.g., forensics, paralinguistics).

Growing demand for cost-efficient voice collection methods^[4] necessitates secure, open-source automatic collection tools.

Aim

This study test automatic voice data collection reliably gathers high-quality, natural-behaviour voice recordings comparable to, or better than, human-assisted traditional manual method.

Materials & Methods

Dataset

Table 1: Comparing Voice Acquisition Techniques (CVAT) dataset demographics.

Gender	F	M	Total
# of Participants	28	13	41
Avg. Age (yrs)	23.0 ±1.1	25.4 ±2.6	—
Age Range (yrs)	16–48	18–53	—

In a soundproof studio, participants recorded 16 voice tasks (CAPE-V Phrases, counting, Happy Birthday, maximum phonation time {/a/_1,_2,_3, /s/, /z/}, pitch range (low-to-high, high-to-low), popular song, 3 sound dynamics, The Rainbow Passage, and Vowel Onset Phrases) in 2 different conditions:

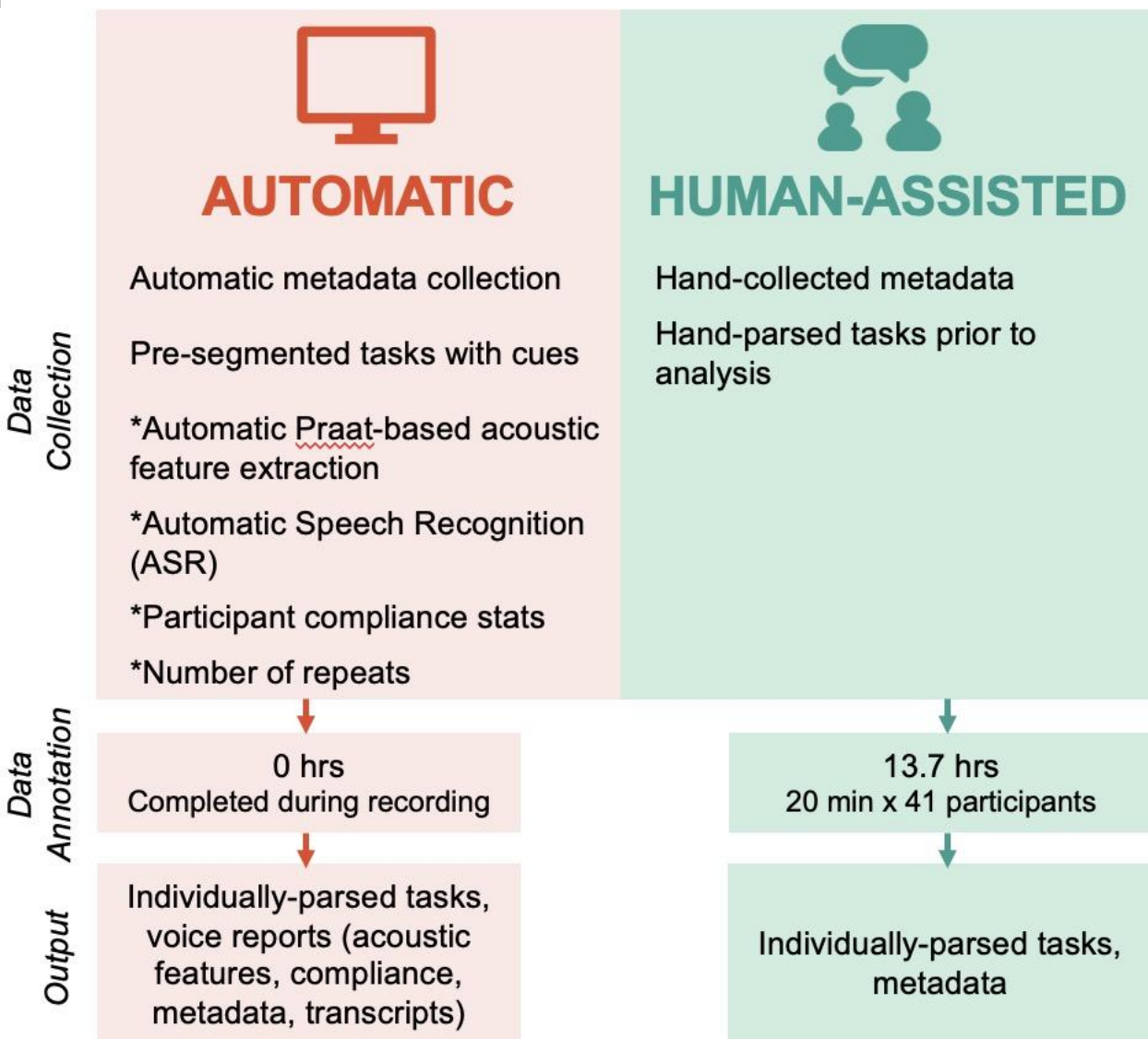


Figure 1: Automatic and human-assisted CVAT database collection processes. In total the 41 participants generated 1387 individual voice recordings (7 hours, 10 minutes of speech data). * indicates additional voice-related information collected.

Results

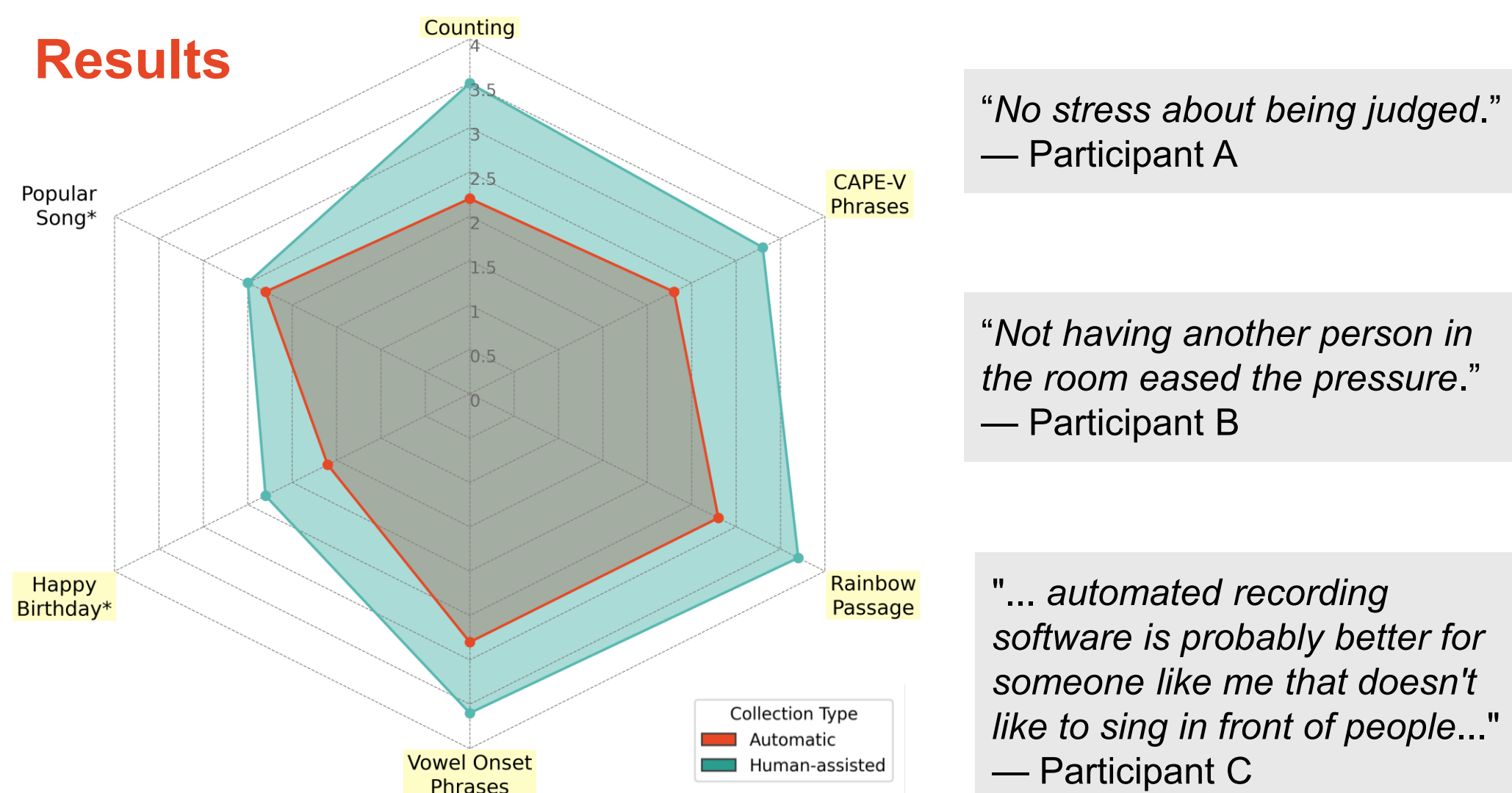


Figure 2: Average speech rate (syllables per second) for **automatic** and **human-assisted** CVAT database, for read tasks and performative tasks (*). Yellow highlights indicate statistical significance (t -tests ($p \leq 0.01$), Hedges' ($g \geq 0.50$)).

- **Automatic** collection produced a slower speech rate – often indicative of less nervousness/stress^[5].
- 51% of participants reported less nervousness during **automatic** recordings, especially in singing tasks.

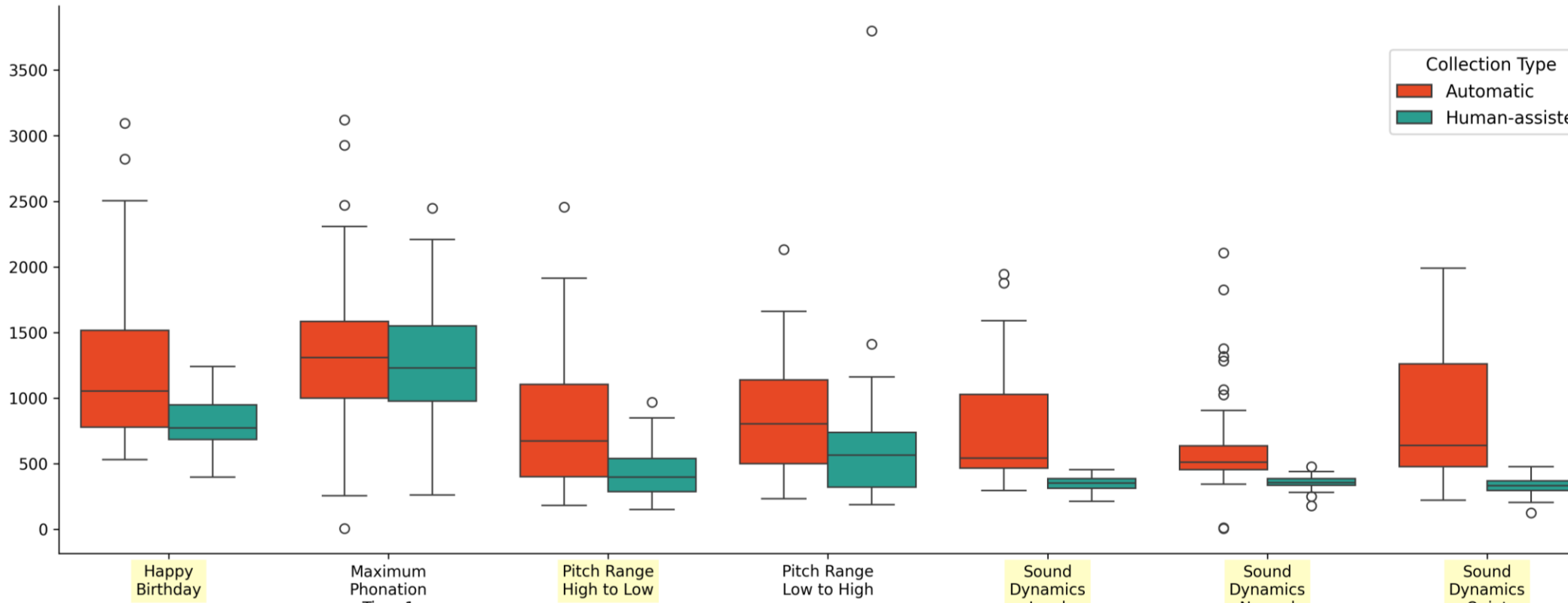


Figure 3: Number of voiced frames (amount of speech data) for **automatic** and **human-assisted** CVAT database. These results show performative tasks. Yellow highlights indicate statistical significance (t -tests ($p \leq 0.01$), Hedges' ($g \geq 0.50$)).

- **Automatic** collection often produced longer voice recordings (i.e., more effort, larger data sample).
- During **automatic** collection, participants chose to re-attempt voice tasks more often than **human-assisted** collection.

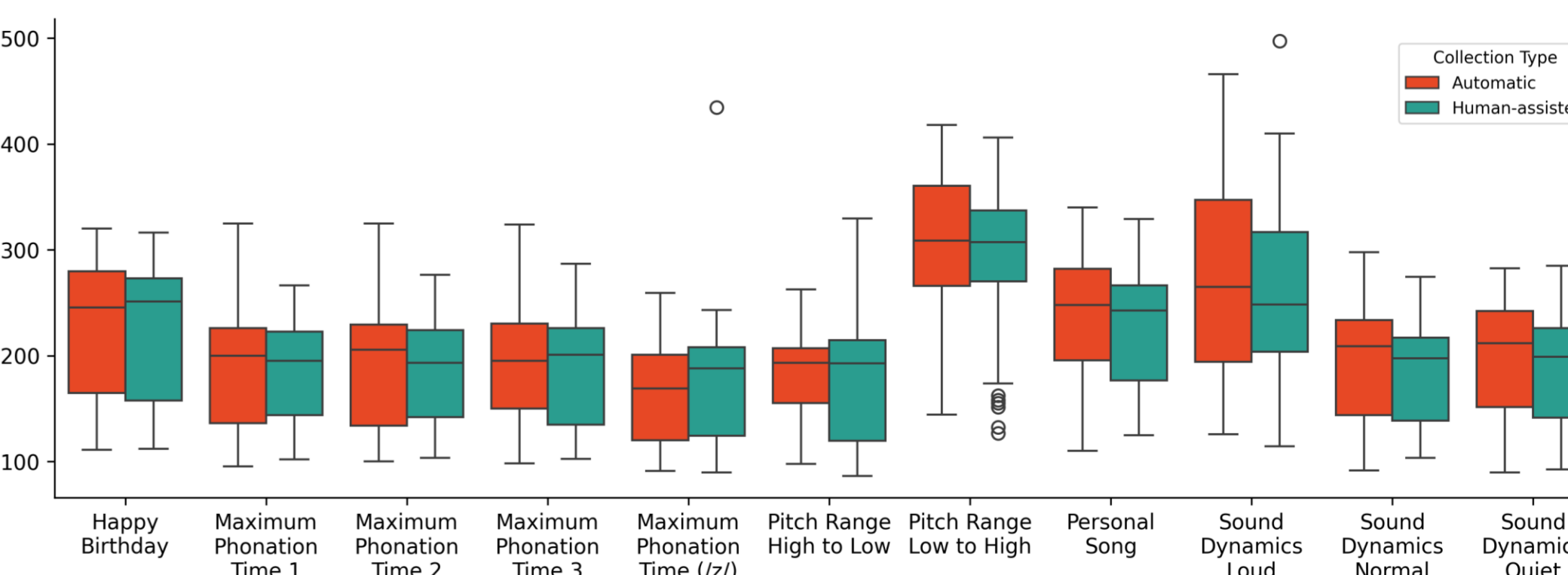


Figure 4: Average pitch (Hz) for **automatic** and **human-assisted** CVAT database. These results show performative tasks.

- No statistically significant average pitch differences between **automatic** and **human-assisted** recorded performative tasks. The same trend was observed across read tasks.
- Statistically significant differences for pitch standard deviation were observed for read tasks (e.g., CAPE-V, Rainbow Passage, and Vowel Onset Phrases) and maximum phonation time /a/_3.

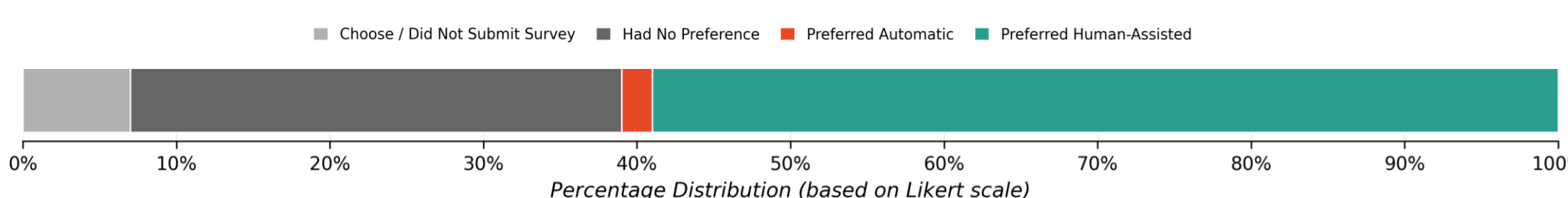


Figure 4: 5-point Likert ratings for **automatic** and **human-assisted** collection conditions. 1 – Very Negative, 2 – Negative, 3 – Neutral, 4 – Positive, 5 – Very Positive. Overall, ratings for both conditions were neutral-to-positive, with higher ratings for human-assisted collection (4.45) than automatic collection (3.63).

Conclusions

- ❖ Compiled a new open-source human–computer interaction voice dataset to explore automated and human-assisted audio-voice recording behaviour.
- ❖ Investigated the viability of using a newly available automated voice data collection tool that logs systematic participant metadata and different voice task recordings with cued instructions.
- ❖ The majority of surveyed CVAT participants still preferred human-assisted collection over automatic. Despite a higher average Likert rating (4.45 vs. 3.63), the automatic method was more cost/time efficient and generated high-quality recordings with similar task compliance to the human-assisted method.
- ❖ CVAT participants also indicated feeling less nervous completing the automatic condition.

Future Research

Additional validation of this automated data collection is needed, especially in clinical and/or remote settings. The CVAT will soon be publicly released (2025/HE001398).

In late 2026, a new study will explore practical use of a voice collection tool in a clinical healthcare patient setting. Python code for CalliOpenNLP can be downloaded for free from the QR code below.



Literature Cited

- [1] A. Kalia, M. Boyer, G. Fagherazzi, Jean-Christophe B  lisle-Pipon, and Y. Bensoussan, “Master protocols in vocal biomarker development to reduce variability and advance clinical precision: a narrative review,” *Frontiers in Digital Health*, vol. 7, pp. 1619183–1619183, June 2025, doi: 10.3389/fdgh.2025.1619183.
- [2] G. Fagherazzi and Y. Bensoussan, “The Imperative of Voice Data Collection in Clinical Trials,” *Digital Biomarkers*, vol. 8, no. 1, pp. 207–209, Nov. 2024, doi: 10.1159/000541456.
- [3] M.-F. Malo, S. Bouhouita-Guermech, H. Gallois, V. Ravitsky, and J.-C. B  lisle-Pipon, “Using voice and speech data in healthcare: a scoping review of the ethical, legal and social implications,” *Frontiers in Digital Health*, vol. 8, Feb. 2026, doi: 10.3389/fdgh.2026.1750111.
- [4] A. P. Vogel and P. Maruff, “Comparison of voice acquisition methodologies in speech research,” *Behavior Research Methods*, vol. 40, no. 4, pp. 982–987, Nov. 2008, doi: 10.3758/bm.40.4.982.
- [5] L. Schewski, M. M. Doss, G. Beldi, and S. Keller, “Measuring negative emotions and stress through acoustic correlates in speech: A systematic review,” *PLOS One*, vol. 20, no. 7, Jul. 2025, doi:10.1371/journal.pone.0328833.

Contact Information

Rebecca Li (BAppSci., Hons.)

E: reli2670@uni.sydney.edu.au

