Tailored Automatic Speech Recognition for a Patient with Global Aphasia and Dysarthria: Single-Case Proof of Concept

Davide Mulfari, Serena Campana, Loredana Voci, Carmelo Mario Vicario, Rocco Salvatore Calabrò, Francesco Corallo, Salvatore Mulfari, Francesco Tomaiuolo

Submitted to: JMIR Rehabilitation and Assistive Technologies
on: August 07, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript ... 4
Supplementary Files ... 16
Multimedia Appendixes ... 17
Multimedia Appendix 1 ... 17

Preprint
JMIR Publications

Tailored Automatic Speech Recognition for a Patient with Global Aphasia and Dysarthria: Single?Case Proof of Concept

Davide Mulfari[1]; Serena Campana[2]; Loredana Voci[2]; Carmelo Mario Vicario[3]; Rocco Salvatore Calabrò[4]; Francesco Corallo[4]; Salvatore Mulfari[1]; Francesco Tomaiuolo[5]

[1] Assistive Technology for Special Needs Office, University of Messina Messina IT
[2] Neurorehabilitation Unit, Auxilium Vitae Volterra Volterra IT
[3] 5. Department of Cognitive Sciences, Psychology, Education and Cultural Studies, University of Messina Messina IT
[4] IRCCS Centro Neurolesi Bonino-Pulejo Messina IT
[5] Department of Clinical and Experimental Medicine, University of Messina Messina IT

Abstract

Background: Severe dysarthria and global aphasia drastically reduce speech intelligibility, confining communication to familiar partners. Commercial automatic speech recognition (ASR) engines typically fail with such atypical speech.

Objective: To determine whether a speaker?dependent Voice?Input Voice?Output Communication Aid (VIVOCA) embedded in the CapisciAMe app can decode the speech of a person with severe dysarthria and aphasia more accurately than human listeners.

Methods: We conducted a single?case proof?of?concept study. A 34?year?old woman, 15?years post?stroke, recorded 1,120 utterances of 13 command words across five prompting modalities. A compact convolutional neural network (cnn?trad?fpool3) was trained on these samples and evaluated on an independent set of 936 utterances. Intelligibility was benchmarked against 12 rehabilitation professionals familiar with the patient. The primary outcome was word?level accuracy.

Results: The tailored ASR achieved 72.7?% accuracy, outperforming human listeners (mean?=?56.2?%, SD?=?13.0).

Conclusions: Customized ASR systems represent a transformative assistive technology for individuals with severe dysarthria and aphasia, restoring aspects of communicative independence that may be unattainable through unaided human interaction. The CapisciAMe VIVOCA exemplifies how user-tailored technology can enhance autonomy, social participation, and overall quality of life, and the trajectory from early voice-input/voice-output systems to large-scale personalization efforts supports continued development in this direction.

(JMIR Preprints 07/08/2025:81999)
DOI: https://doi.org/10.2196/preprints.81999

Preprint JMIR Publications

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

Preprint
JMIR Publications

Tailored Automatic Speech Recognition for a Patient with Global Aphasia and Dysarthria: Single-Case Proof of Concept

Davide Mulfari[1], Davide Cardile[2,3*], Campana Serena[4], Loredana Voci[4], Carmelo Mario Vicario[5], Rocco Salvatore Calabrò[2], Francesco Corallo[2], Salvatore Mulfari[1], Francesco Tomaiuolo[6]

1. Assistive Technology for Special Needs Office, University of Messina, Piazza Pugliatti, 1, 98122 Messina, Italy.
2. IRCCS Centro Neurolesi Bonino-Pulejo, S.S. 113 Via Palermo, C. da Casazza; 98124- Messina, Italy.
3. Department of Health Sciences, 'Magna Graecia' University of Catanzaro, viale Europa, 88110, Catanzaro, Italy.
4. Neurorehabilitation Unit, Auxilium Vitae Volterra, Via Borgo San Lazzero 5, 56048 Volterra, Italy;
5. Department of Cognitive Sciences, Psychology, Education and Cultural Studies, University of Messina, Piazza Pugliatti, 1, 98125 Messina, Italy.
6. Department of Clinical and Experimental Medicine, University of Messina, Piazza Pugliatti, 1, 98122 Messina, Italy.

davide.mulfari@unime.it; davide.cardile@irccsme.it; s.campana@riabilitazione-volterra.it, l.voci@riabilitazione-volterra.it, cvicario@unime.it, roccos.calabro@irccsme.it, francesco.corallo@irccsme.it, salvatore.mulfari@gmail.com, ftomaiuolo@unime.it

*** Corresponding author:** davide.cardile@irccsme.it

Keywords: dysarthria; aphasia; automatic speech recognition; speaker-dependent ASR; VIVOCA; assistive technology; rehabilitation

Abstract

Background: Severe dysarthria and global aphasia drastically reduce speech intelligibility, confining communication to familiar partners. Commercial automatic speech recognition (ASR) engines typically fail with such atypical speech.

Objective: To determine whether a speaker-dependent Voice-Input Voice-Output Communication Aid (VIVOCA) embedded in the CapisciAMe app can decode the speech of a person with severe dysarthria and aphasia more accurately than human listeners.

Methods: We conducted a single-case proof-of-concept study. A 34-year-old woman, 15 years post-stroke, recorded 1,120 utterances of 13 command words across five prompting modalities. A compact convolutional neural network (cnn-trad-fpool3) was trained on these samples and evaluated on an independent set of 936 utterances. Intelligibility was benchmarked against 12 rehabilitation professionals familiar with the patient. The primary outcome was word-level accuracy.

Results: The tailored ASR achieved 72.7 % accuracy, outperforming human listeners (mean = 56.2 %, SD = 13.0). Conclusions: A personalized ASR system can exceed the intelligibility of human listeners for profoundly disordered speech, supporting its use as an assistive communication technology.

Keywords: dysarthria; aphasia; automatic speech recognition; speaker-dependent ASR; VIVOCA; assistive technology; rehabilitation

1. Introduction

Dysarthria is a motor speech disorder arising from impaired neuromuscular control [1]. Although various types exist [2], these speech impairments can result from damage at various locations of the neuroaxis, including: the motor cortex, basal ganglia, brainstem nuclei, cerebellum, or descending corticobulbar pathways, often affecting cranial nerves V, VII, IX, X, and XII which are critical for respiration, articulatory precision, prosody and speech timing [3,4]. Beyond motor impairment, individuals with dysarthria frequently face significant psychosocial challenges, including social withdrawal, depression, and reduced quality of life, especially when intelligibility is severely compromised [5,6].

These effects are even more pronounced when dysarthria co-occurs with global aphasia after large left-hemispheric lesions a condition marked by profound deficits in expressive and receptive language.

Following large left-hemispheric strokes, especially involving the perisylvian cortex and deep white matter, verbal communication becomes drastically limited, and the patient may rely entirely on familiar partners for interaction [7,8].

Recent advances in machine learning have rekindled interest in automatic speech recognition (ASR) as a tool for augmentative and alternative communication (AAC) [9].

In clinical populations with atypical speech, such as post-stroke dysarthria, generic ASR systems often fail due to mismatched acoustic models and high speech variability [10,11]. However, speaker-dependent systems (models trained exclusively on a single individual's voice) offer a promising solution by adapting to the unique phonetic and articulatory patterns of the user [12]. These systems have demonstrated improved word-level accuracy even in the presence of severe motor speech impairments [13].

The CapisciAMe app was developed in Italy to explore personalized ASR vocabularies for people with motor speech disorders [14]. We report a single-case evaluation of a CapisciAMe VIVOCA configured for an adult patient with global aphasia and severe dysarthria. System performance [15] was compared with that of rehabilitation clinicians who routinely communicate with her. This case adds to emerging evidence supporting customized ASR systems in augmenting communication for motor speech disorders.

2. Method

2.1 Case Description

The patient was a 34 year-old right-handed adult (education: 12 years) who sustained a massive left-hemispheric hemorrhage when the patient was 19 years old. Surgical excision of an arteriovenous malformation and occlusion of the left middle cerebral artery were performed. Follow-up MRI at 18 months revealed an extensive lesion encompassing the opercular, temporo-fronto-parietal and putaminal regions (figure 1). The patient presented with right hemiparesis, global aphasia and dysarthria. Despite minimal verbal output, non-verbal reasoning was preserved (Raven's = 43/60). Language abilities were evaluated using the Esame Neuropsicologico per l'Afasia. (E.N.P.A.; Capasso & Miceli, 2001), at two time-points at 14 and 28 months form injury. This longitudinal clinical profile underscores a marked recovery in cognitive and communicative intentionality, despite persistent severe verbal limitations. The preserved reasoning skills and enhanced social interaction provide a strong foundation for introducing assistance. A more detailed clinical and linguistic profile is available in Appendix 1.

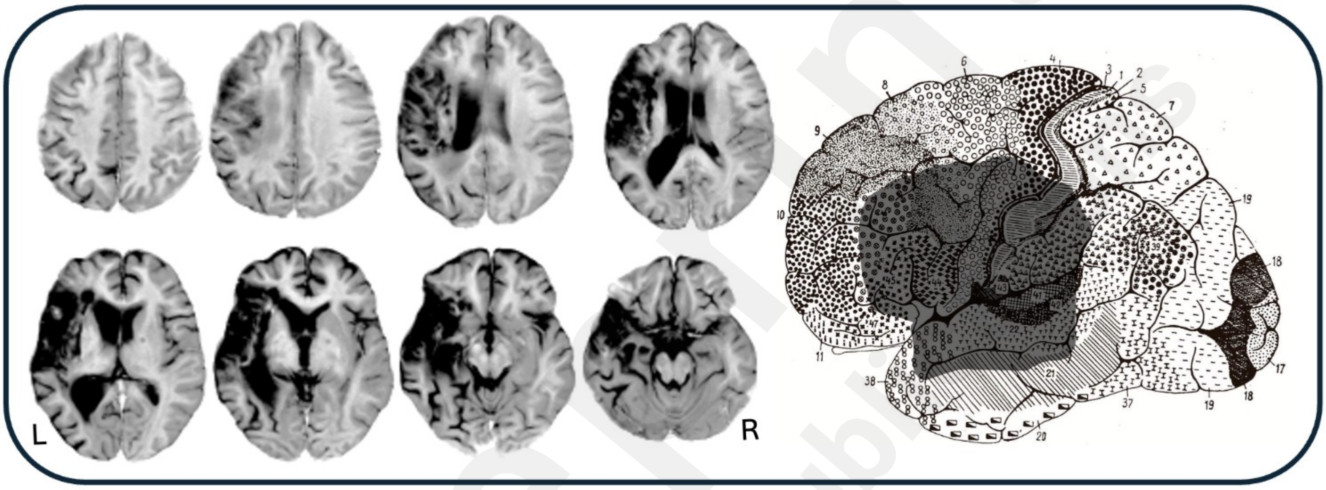

Figure 1: Axial brain MRI scans highlight an extensive neuroanatomical disruption affecting the left opercular region of the patient. The figure shows MRI scans performed 18 months from the cerebrovascular accident. The first image series from left to right shows progressive slices of the brain at different levels from top to bottom. The localization and extent of the lesion was depicted on the left hemisphere with reference to Broadman's map areas. The most posterior part of the superior and middle temporal gyri was partially spared as well as the angular gyrus and part of the supramarginal gyrus.

2.2 Vocabulary and Prompting Modalities

A 13 Italian word set (destra: right, su: up, okay: okay, stop: stop, dietro: back, sinistra: left, esci: exit, avanti: forward, giù: down, chiudi: close, volume: volume, entra: enter, apri: open) previously validated in a home-automation project was adopted [14] keeping the vocabulary small enough to train a reliable speaker-dependent model.

Patient speech production was elicited using the following five prompting modalities:

Modality	Test	Task
M1	Text-to-caption and Text-to-speech	Word was both displayed on the screen and played aloud using text-to-speech synthesis
M2	Text-to-speech only	The word delivered exclusively through synthesized audio without any visual cue
M3	Pictures only	An image representing the word was shown to prompt a naming response
M4	Text caption only	The word appeared on the screen without accompanying audio
M5	Antonym-Based Prompting	A semantically opposite word was presented, either visually or orally, to prompt production of the target word through contrast.

Table 2: Summary of the five prompting modalities used to train the patient's personalized CapisciAMe ASR system.

2.3 Data Collection and Model Training

Over one month the patient completed 33 recording sessions (about three times a day, three days a week) in a quiet therapy room using a headset microphone (44.1 kHz, 16-bit). For every item of the Italian word set, a single target word was elicited with one of the five prompting modalities reported in Table 1. The resulting 1,120 utterances formed the training set of 18 recording sessions. A separate 936-item set and strictly excluded from training (M1 = 94, M2 = 208, M3 = 97, M4 = 165, M5 = 372) served for testing. The cnn-trad-fpool3 network, a baseline architecture for keyword spotting presented at Google [16], was trained with Adam (learning rate = 1×10^{-4}, batch = 64) for up to 40 epochs with early stopping.

Model training employed the Adam optimizer with a learning rate of 1e-4, a batch size of 64, and was run for 40 epochs with early stopping triggered after 3 consecutive non-improving validation steps.

The primary evaluation metric was word-level classification accuracy, calculated as the proportion of correctly identified words among the total utterances produced by the patient.

2.4. Human Speech Intelligibility Benchmark

To compare ASR performance with human perception, twelve Italian rehabilitation professionals (6 female, 6 male; mean age = 38.92; DS 9.35), all members of the clinical rehabilitation unit and familiar with the speech profile of the participant, with the exception of the speech-language therapists directly involved in her treatment, were recruited to assess intelligibility. All participants provided informed consent. Except for the patient's current speech-language therapists, all clinicians reported familiarity with the patient's communicative style but had not been directly involved in speech training. Each clinician was presented with 26 utterances (two per keyword), randomly selected from the recorded dataset. Recordings were played without any visual or contextual cues, and listeners were instructed to verbally report the word they believed the patient intended to say. Responses were scored for open-set accuracy.

3. Results

Across all modalities the personalized ASR correctly identified 72.7 % of the patient's spoken words, compared to a mean of 56.2 % (SD = 13.0) for human listeners (Figure 2).

To assess whether the tailored ASR outperformed the listener-control group, we compared aggregated word-level accuracies with a two-proportion test. The ASR correctly recognized 680 of 936 words (72.65%, 95% CI 69.71–75.41). The listener-control group correctly recognized 175 of 312 words (56.09%, 95% CI 50.54–61.49). This yields an absolute advantage of 16.56 percentage points for the ASR (95% CI 10.36–22.76). A two-sided z-test confirmed that the difference is highly significant ($z = 5.454$, $p = 5.0 \times 10^{-8}$). The effect size was Cohen's $h = 0.348$, indicating a medium practical effect.

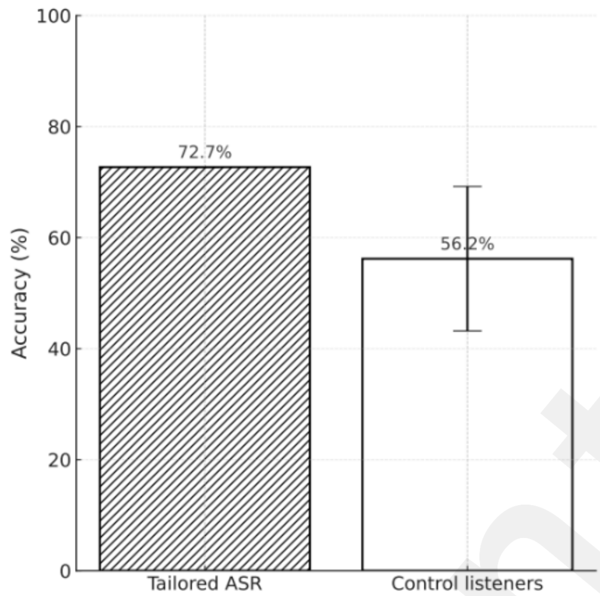

Figure 2. Mean accuracy for Tailored ASR output (left) and control listener performance (right). The dashed error bar represents the standard deviation (13%), while the bar heights reflect the group means (Tailored ASR = 72.7%, Listeners = 56.2%).

4. Discussion

This single-case study shows that a personalized VIVOCA integrated into the CapisciAMe app decodes profoundly disordered speech more accurately than familiar human listeners. Across an independent test set, the tailored recognizer consistently outperformed the clinician control group. For this speaker with severe dysarthria and global aphasia, a speaker-dependent ASR therefore functions as an amplifier of intelligibility, supporting more reliable communication in everyday contexts [17,18].

Beyond the motor speech impairment, this case is marked by global aphasia, which severely constrains linguistic formulation and comprehension. Within this context, the observed benefit of a personalized VIVOCA reflects not only better decoding of highly atypical articulatory patterns but also a pragmatic workaround for the lexical–semantic bottlenecks typical of global aphasia. By centering communication on a compact, functionally salient vocabulary, the system lowers linguistic load and supports reliable expression of high-priority intents, an approach consistent with frameworks in augmentative and alternative communication that prioritize communicative competence and participation over restoration of normal language. Evidence from picture-based, computer-supported programs for global aphasia shows that several individuals convey more information with aided communication than with natural communication alone, underscoring the practical value of such supports [19].

This study employs a single-case design, centered on a person with language and motor speech impairments. Although the personalized ASR system yielded promising outcomes, the results are not immediately generalizable to broader populations with more variable profiles. The limited training vocabulary also constrains the communicative breadth of the system. Additionally, the intelligibility benchmark involved only rehabilitation professionals, which may introduce bias due to their ongoing interactions with the individual.

These considerations align with broader observations that generic recognizers struggle with dysarthric speech, while personalization tends to close the gap [20].

Future directions should include expanding the system's lexicon to support more flexible and spontaneous communication. Comparative evaluations against strong, general-purpose recognizers and against unfamiliar listeners would help establish more robust performance benchmarks. In parallel, advances in large-scale speech and language modeling, such as weakly supervised models trained on hundreds of thousands of hours, offer practical routes to improved robustness when combined with speaker-specific adaptation; personalization with limited data has already shown large error-rate reductions for atypical speech [21].
Finally, as proposed by Tröger et al. [22], future evaluations should pair recognition accuracy with person-centered outcomes such as daily use, communicative effectiveness, caregiver burden, and satisfaction and include validated digital indicators of intelligibility. One practical option like an automated speech-intelligibility score, developed and validated across several motor-speech disorders and languages; it provides a scalable, objective outcome that complements clinician ratings and supports sensitive tracking over time.
In conclusion, customized ASR systems represent a transformative assistive technology for individuals with severe dysarthria and aphasia, restoring aspects of communicative independence that may be unattainable through unaided human interaction. The CapisciAMe VIVOCA exemplifies how user-tailored technology can enhance autonomy, social participation, and overall quality of life, and the trajectory from early voice-input/voice-output systems to large-scale personalization efforts supports continued development in this direction [23].

5. Funding & Ethical Statements:

This research was supported by PRIN 2022 (CUP J53D23008320006). Per local policy, the study was classified as low-risk and only required notification to the ethics committee, which was duly completed. Written informed consent was obtained from the participant, that cannot be identified from the information presented.

6. Conflicts of Interest

The authors declare no conflicts of interest.

5. References

1. Palmer, R., & Enderby, P. (2007). Methods of speech therapy treatment for stable dysarthria: A review. *Advances in Speech Language Pathology*, *9*(2), 140–153. https://doi.org/10.1080/14417040600970606

2. Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of speech and hearing research*, *12*(2), 246–269. https://doi.org/10.1044/jshr.1202.246.

3. Enderby P. (2013). Disorders of communication: dysarthria. Handbook of clinical neurology, 110, 273–281. https://doi.org/10.1016/B978-0-444-52901-5.00022-8

4. Jayaraman DK, Das JM. Dysarthria. [Updated 2023 Jun 5]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK592453/

5. Dickson, S., Barbour, R. S., Brady, M., Clark, A. M., & Paton, G. (2008). Patients' experiences of disruptions associated with post-stroke dysarthria. International Journal of Language & Communication Disorders, 43(2), 135–153.

6. Whitehill, Tara & Ma, Estella & Tse, Francis. (2010). Environmental Barriers to Communication for Individuals with Dysarthria. Journal of Medical Speech-Language Pathology. 18. 141-144.

7. Summaka, M., Hannoun, S., Harati, H., Daoud, R., Zein, H., Estephan, E., Naim, I., & Nasser, Z. (2022). Neuroanatomical regions associated with non-progressive dysarthria post-stroke: a systematic review. *BMC neurology*, *22*(1), 353. https://doi.org/10.1186/s12883-022-02877-x.

8. Kim, G., Min, D., Lee, E. O., & Kang, E. K. (2016). Impact of Co-occurring Dysarthria and Aphasia on Functional Recovery in Post-stroke Patients. *Annals of rehabilitation medicine*, *40*(6), 1010–1017. https://doi.org/10.5535/arm.2016.40.6.1010

9. Chen, S. K., Saeli, C., & Hu, G. (2024). A proof-of-concept study for automatic speech recognition to transcribe AAC speakers' speech from high-technology AAC systems. Assistive technology : the official journal of RESNA, 36(4), 319–326. https://doi.org/10.1080/10400435.2023.2260860.

10. Jaddoh, A., Loizides, F., & Rana, O. (2023). Interaction between people with dysarthria and speech recognition systems: A review. *Assistive technology : the official journal of RESNA*, *35*(4), 330–338. https://doi.org/10.1080/10400435.2022.2061085.

11. Shor, J., Emanuel, D., Lang, O., Tuval, O., Brenner, M., Cattiau, J., Vieira, F., McNally, M., Charbonneau, T., Nollstadt, M., Hassidim, A., & Matias, Y. (2019). *Personalizing ASR for dysarthric and accented speech with limited data* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1907.13511

12. He, Y., Seng, K. P., & Ang, L. M. (2025). Collaborative AI Dysarthric Speech Recognition System

With Data Augmentation Using Generative Adversarial Neural Network. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, *33*, 2097–2111. https://doi.org/10.1109/TNSRE.2025.3570383

13. Mulfari, D., Meoni, G., Marini, M., & Fanucci, L. (2021). Machine learning assistive application for users with speech disorders. *Applied Soft Computing*, *103*, 107147. https://doi.org/10.1016/j.asoc.2021.107147

14. Mulfari, D., La Placa, D., Rovito, C., Celesti, A., & Villari, M. (2022). Deep learning applications in telerehabilitation speech therapy scenarios. Computers in biology and medicine, 148, 105864. https://doi.org/10.1016/j.compbiomed.2022.105864

15. Capasso, R., & Miceli, G. (2001). *Esame neuropsicologico per l'afasia (E.N.P.A.)*. Springer Milano.

16. Sainath, T.N., Parada, C. (2015) Convolutional neural networks for small-footprint keyword spotting. Proc. Interspeech 2015, 1478-1482, doi: 10.21437/Interspeech.2015-352.

17. Hawley, M. S., Cunningham, S. P., Green, P. D., Enderby, P., Palmer, R., Sehgal, S., & O'Neill, P. (2013). A voice-input voice-output communication aid for people with severe speech impairment. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, *21*(1), 23–31. https://doi.org/10.1109/TNSRE.2012.2209678.

18. Green, J. R., MacDonald, R. L., Jiang, P.-P., Cattiau, J., Heywood, R., Cave, R., Seaver, K., Ladewig, M. A., Tobin, J., Brenner, M. P., Nelson, P. C., & Tomanek, K. (2021). Automatic speech recognition of disordered speech: Personalized models outperforming human listeners on short phrases. In *Proceedings of INTERSPEECH 2021*. Graz, Austria

19. Nicholas, M., Sinotte, M. P., & Helm-Estabrooks, N. (2011). C-Speak Aphasia alternative communication program for people with severe aphasia: importance of executive functioning and semantic knowledge. *Neuropsychological rehabilitation*, *21*(3), 322–366. https://doi.org/10.1080/09602011.2011.559051.

20. Tobin, J., Nelson, P., MacDonald, B., Heywood, R., Cave, R., Seaver, K., Desjardins, A., Jiang, P.-P., & Green, J. R. (2024). Automatic speech recognition of conversational speech in individuals with disordered speech. *Journal of Speech, Language, and Hearing Research*, *67*(11), 4176–4185.

21. Tomanek, K., Beaufays, F., Cattiau, J., Chandorkar, A., & Sim, K. C. (2021). On-device personalization of automatic speech recognition models for disordered speech. *arXiv preprint*. https://doi.org/10.48550/arXiv.2106.10259.

22. Tröger, J., Dörr, F., Schwed, L., Linz, N., König, A., Thies, T., Barbe, M. T., Orozco-Arroyave, J. R., & Rusz, J. (2024). An automatic measure for speech intelligibility in dysarthrias-validation across

Preprint JMIR Publications

multiple languages and neurological disorders. *Frontiers in digital health*, *6*, 1440986. https://doi.org/10.3389/fdgth.2024.1440986.

23. MacDonald, R.L., Jiang, P.-P., Cattiau, J., Heywood, R., Cave, R., Seaver, K., Ladewig, M.A., Tobin, J., Brenner, M.P., Nelson, P.C., Green, J.R., Tomanek, K. (2021) Disordered Speech Data Collection: Lessons Learned at 1 Million Utterances from Project Euphonia. Proc. Interspeech 2021, 4833-4837, doi: 10.21437/Interspeech.2021-697.

Preprint JMIR Publications

Supplementary Files

Preprint
JMIR Publications

Multimedia Appendixes

Untitled.
URL: http://asset.jmir.pub/assets/8a1c5a9671d622d22c1e8aeb77669feb.docx

Preprint
JMIR Publications