Prototype App for Automatic Speech Intelligibility Estimation

Learning Outcomes

- **DESCRIBE** the difference between reference-free and reference-based approaches to automated speech intelligibility estimation
- **IDENTIFY** clinical scenarios where automated intelligibility estimation tools could support speech assessment and progress monitoring
- **EVALUATE** the potential benefits and limitations of prototype automated intelligibility tools for clinical use

Keywords

speech intelligibility, automated assessment, dysarthria, clinical tool, speech technology

Session Description (max 120 words)

Speech intelligibility is one of the most meaningful measures of communicative function, yet routine assessment remains time-consuming and resource-intensive. This interactive poster presents a prototype application that automates intelligibility estimation using three approaches: reference-audio methods, which compare a speaker's production against a typical speaker recording; reference-text methods, which score speech against a known transcription; and reference-free methods, which estimate intelligibility directly from the speech signal. Attendees can try the app, provide speech samples, and observe how each approach works. The session welcomes clinicians, speech scientists, engineers, and other professionals, and aims to gather feedback on usability, relevance, and remaining challenges. No prior technical background is required.

Abstract (max 1000 words)

Background

Assessing speech intelligibility in individuals with speech disorders, caused by neurological conditions (such as Parkinson's), structural changes following head and neck cancer surgery, or other pathologies, is central to tracking disease progression, monitoring rehabilitation, and evaluating the effectiveness of interventions (Miller, 2013).

Despite its clinical importance, routine intelligibility assessment remains underutilized in practice. Surveys of speech-language pathologists (SLPs) working with adults with dysarthria have found that many lack access to formal assessment tools, with cost and time constraints cited as primary barriers (Gurevich & Scamihorn, 2017). Most clinicians rely on informal perceptual estimates rather than systematic measurement, which introduces subjectivity and limits the ability to track change over time. While speech can be characterized along multiple dimensions, including articulatory precision (Tu et al., 2016), voice quality, and overall impairment severity (Halpern et al., 2023), these measures correlate very strongly with intelligibility ($r \geq 0.9$) in individuals with dysarthria and head-and-neck cancer (Halpern et al., 2025; Tu et al., 2016), likely because patients rarely exhibit isolated speech deficits. Intelligibility therefore serves as a practical unifying target for clinical assessment.

Automated Intelligibility Estimation

Automated approaches to intelligibility estimation are not new. Methods utilizing automatic speech recognition to quantify speech intelligibility in individuals with speech pathologies have already been reported in the 90s (Sy & Horowitz, 1993), and have been shown to work for several kinds of speech disorders for more than three decades including dysarthria (Choi et al., 2026), head and neck cancer (Windrich et al., 2008), cleft lip and palate (Schuster et al., 2006b), and laryngectomy (Schuster et al., 2006a). More recently, advances in artificial intelligence have substantially expanded possibilities. Large-scale self-supervised speech models, trained on vast amounts of multilingual audio, can extract rich representations of speech without requiring language-specific resources or labelled pathological data. This has enabled approaches that move beyond ASR-based transcription matching, estimating intelligibility directly from the speech signal and opening the door to evaluation across languages and clinical populations with minimal adaptation (Halpern et al., 2026; Yeo et al., 2026). While these methods have shown very promising results, they are not yet ready for clinical use. As recently outlined (Liss & Berisha, 2024), key challenges remain around generalizability and interpretability of the output, and addressing these requires direct input from the clinicians and researchers who would use these tools in practice.

Prototype Application

This gap between what is technically possible and what is practically available in the clinic motivates the present work. The purpose of this presentation is to introduce a prototype application for automated speech intelligibility estimation and to gather direct feedback about its potential utility for research and clinic.

The app implements three complementary estimation approaches. Reference-audio methods compare the patient's speech against a recording of a typical speaker producing the same utterance, enabling fine-grained quantification of acoustic deviation from a target. This yields high sensitivity to subtle changes but requires matched audio reference material, limiting applicability to structured elicitation tasks where stimulus control is possible. Reference-text methods use a known target transcription to score the patient's speech via automatic speech recognition, comparing the recognized output against the intended text. This is less demanding to set up than audio references (requiring only the stimulus text rather than a matched recording) but depends on ASR accuracy, and its resolution is capped at the phoneme level, meaning it cannot capture finer acoustic detail such as within-phoneme distortions. Reference-free methods estimate intelligibility directly from the patient's speech signal without any comparison material, making them applicable to spontaneous or conversational speech and removing the burden of collecting reference material entirely, at the cost of somewhat reduced precision relative to reference-based approaches.

By offering all three modes within a single interface, the application allows the user to select the approach best suited to their context. The prototype is designed to be simple and accessible, requiring no technical expertise to operate: the user records or uploads a speech sample, selects the estimation mode, and receives an intelligibility estimate along with supporting information. The interface abstracts away all technical complexity, presenting results in clinically interpretable terms.

Poster Session

During the poster session, attendees can try the prototype app, compare the output of reference-based and reference-free estimation modes, and provide structured feedback on usability, interpretability of results, and perceived clinical relevance. The presenter is particularly interested in understanding attendee perspectives on what constitutes a clinically meaningful intelligibility score, how automated estimates might complement rather than replace perceptual judgment, and crucially, what practitioners themselves identify as the current barriers to adopting tools of this kind in routine practice. This feedback will directly inform continued development of the tool, including decisions about output format, score interpretation, and integration with existing clinical workflows.

Conclusion

This presentation aligns with the 2026 ASHA Convention theme, "Momentum: Where Vision Meets Action," by bridging speech technology research and clinical practice. While automated intelligibility estimation has been explored in the research literature for several years, few tools have been developed with direct input from clinicians and researchers or made accessible to practitioners outside research settings. This presentation represents a step toward translating these advances into practical clinical tools, with the explicit goal of involving clinicians and speech researchers as collaborators in the design process. The presentation does not claim that the prototype is ready for clinical deployment. Rather, it aims to demonstrate what is achievable with current technology and to initiate a dialogue between technology developers and the clinicians/researchers who will ultimately determine whether such tools are useful in practice.

References

Choi, J., Moya-Galé, G., Hwang, K., Hirschberg, J., & Levy, E. S. (2026). Automatic speech recognition for intelligibility assessment in children with dysarthria. *Journal of Speech, Language, and Hearing Research*, 1–17.

Gurevich, N., & Scamihorn, S. L. (2017). Speech-language pathologists' use of intelligibility measures in adults with dysarthria. *American Journal of Speech-Language Pathology*, *26*(3), 873–892.

Halpern, B. M., Feng, S., van Son, R., van den Brekel, M., & Scharenborg, O. (2023). Automatic evaluation of spontaneous oral cancer speech using ratings from naive listeners. *Speech Communication*, *149*, 84–97.

Halpern, B. M., Tienkamp, T., Abur, D., & Toda, T. (2026). PathBench: Speech intelligibility benchmark for automatic pathological speech assessment. *arXiv preprint arXiv:2603.08097*.

Halpern, B. M., Tienkamp, T., Rebernik, T., van Son, R. J., Wieling, M., Abur, D., & Toda, T. (2025). Relationship between objective and subjective perceptual measures of speech in individuals with head and neck cancer. *Interspeech 2025*, 3733–3737. https://doi.org/10.21437/Interspeech.2025-1127

Liss, J., & Berisha, V. (2024). Operationalizing clinical speech analytics: Moving from features to measures for real-world clinical impact. *Journal of Speech, Language, and Hearing Research*, *67*(11), 4226–4232.

Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders*, *48*(6), 601–612.

Schuster, M., Haderlein, T., Nöth, E., Lohscheller, J., Eysholdt, U., & Rosanowski, F. (2006a). Intelligibility of laryngectomees' substitute speech: Automatic speech recognition and subjective rating. *European Archives of Oto-Rhino-Laryngology*, *263*(2), 188–193. https://doi.org/10.1007/s00405-005-0974-6

Schuster, M., Maier, A., Haderlein, T., Nkenke, E., Wohlleben, U., Rosanowski, F., Eysholdt, U., & Nöth, E. (2006b). Evaluation of speech intelligibility for children with cleft lip and palate by means of automatic speech recognition. *International Journal of Pediatric Otorhinolaryngology*, *70*(10), 1741–1747. https://doi.org/10.1016/j.ijporl.2006.05.016

Sy, B. K., & Horowitz, D. M. (1993). A statistical causal model for the assessment of dysarthric speech and the utility of computer-based speech recognition. *IEEE Transactions on Biomedical Engineering*, *40*(12), 1282–1298.

Tu, M., Wisler, A., Berisha, V., & Liss, J. M. (2016). The Relationship Between Perceptual Disturbances in Dysarthric Speech and Automatic Speech Recognition Performance. *The Journal of the Acoustical Society of America*, *140*(5), EL416–EL422.

Windrich, M., Maier, A., Kohler, R., Nöth, E., Nkenke, E., Eysholdt, U., & Schuster, M. (2008). Automatic quantification of speech intelligibility of adults with oral squamous cell carcinoma. *Folia Phoniatrica et Logopaedica*, *60*(3), 151–156.

Yeo, E., Liss, J. M., Berisha, V., & Mortensen, D. R. (2026). Multilingual dysarthric speech assessment using universal phone recognition and language-specific phonemic contrast modeling. *arXiv preprint arXiv:2601.21205*.