ASHA American Speech-Language-Hearing Association

JSLHR

Research Note

Feasibility of Auditory-Perceptual Rating of Four Dimensions of Connected Speech in Aphasia

Cheryl Chong,[a] Marianne Casilio,[b] and Stephen M. Wilson[a]

[a] School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia [b] Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN

ARTICLE INFO

Article History:
Received May 15, 2025
Revision received July 24, 2025
Accepted August 13, 2025

Editor-in-Chief: Julie A. Washington
Editor: Amy D. Rodriguez

https://doi.org/10.1044/2025_JSLHR-25-00369

ABSTRACT

Purpose: The purpose of this preliminary research study was to evaluate the feasibility of directly rating four dimensions of connected speech in aphasia—Paraphasia, Logopenia, Agrammatism, and Motor Speech—using an auditory-perceptual approach.
Method: Seventeen student clinicians rated the four dimensions in eight speech samples of individuals with aphasia drawn from AphasiaBank. There was minimal initial training, and the only feedback provided consisted of expert scores, which participants were asked to compare to their own ratings.
Results: Of the 68 total dimensions rated (4 dimensions × 17 students), about half were rated with good or excellent accuracy. However, ratings did not improve over the course of the session, suggesting that the feedback provided was insufficient.
Conclusion: Directly rating four key dimensions with an auditory-perceptual approach shows promise as a highly efficient way to analyze connected speech in aphasia, but more comprehensive training and individualized feedback will be necessary to develop this approach for clinical applications.
Supplemental Material: https://doi.org/10.23641/asha.30501776

Connected speech is valuable for aphasia assessment and diagnosis because it provides insight into everyday language skills, is straightforward to elicit, and is sensitive to impairments in multiple speech and language domains (Prins & Bastiaanse, 2004; Vermeulen et al., 1989). However, although connected speech samples are relatively easy to obtain (e.g., through conversation, picture description, narrative recount, etc.), their subsequent analysis remains a bottleneck for widespread clinical applications. Quantitative approaches are time consuming and require extensive training and linguistic expertise (Stark et al., 2021; Yagata et al., 2017). On the other hand, qualitative rating scales are often minimally supported by psychometric data (Stark & Dalton, 2024), may have poor interrater reliability outside of research contexts (Gordon, 1998; Trupe, 1984), and frequently conflate features in an ad hoc manner; for example, the fluency scale on the Western Aphasia Battery–Revised (WAB-R; Kertesz, 2006) incorporates grammatical, motor speech, and paraphasic phenomena.

In response to this challenge, Casilio et al. (2019) proposed a scheme for auditory-perceptual rating of connected speech in aphasia (APROCSA), analogous to the auditory-perceptual approaches used for the assessment of motor speech disorders (Duffy, 2012; Strand et al., 2014) and voice disorders (Sapienza & Hoffman, 2020). In APROCSA, 27 common features of connected speech in aphasia (e.g., abandoned utterances, false starts, phonemic paraphasias) are each rated on a 5-point scale. In an initial feasibility study based on 24 diverse patients from AphasiaBank, most features were rated with good-to-excellent interrater reliability by both experienced listeners and student clinicians and demonstrated strong concurrent validity with respect to quantitative measures (Casilio et al., 2019). Analysis with APROCSA takes about 10–15 min, making it feasible for clinical applications.

Intriguingly, a factor analysis showed that four underlying dimensions—labeled Paraphasia (misselection of words and sounds), Logopenia (paucity of speech), Agrammatism (morphosyntactic omissions and simplifications), and Motor Speech (impaired speech motor processing)—explained 79% of the variance in connected speech features in this patient cohort (Casilio et al., 2019). This suggests

Correspondence to Stephen M. Wilson: smwilson@uq.edu.au. ***Disclosure:*** *Stephen M. Wilson is a mentoring editor for JSLHR, Language. The authors have declared that no other competing financial or nonfinancial interests existed at the time of publication.*

that these four dimensions reflect coherent underlying deficits that account for many of the observable surface features of connected speech in aphasia, and the four dimensions have subsequently been shown to have distinct neural correlates (Casilio et al., 2025). We propose that characterizing a patient's connected speech along these four dimensions, instead of the traditional fluent/nonfluent binary, provides a more nuanced understanding of their specific strengths and weaknesses, which can better inform management and intervention.

Although scores on the four dimensions can be derived after rating the 27 APROCSA features, an alternative approach would be to rate the four dimensions directly using an auditory-perceptual approach. Such "gestalt" ratings of the four dimensions could yield an even more efficient scheme for assessing connected speech in aphasia and could be valuable in time-constrained settings such as acute care (Shrubsole et al., 2019).

In the present study, we aimed to test the feasibility of this general approach by asking student clinicians to rate the four dimensions in connected speech samples from eight individuals with diverse aphasia profiles. Our study addressed two objectives. First, how accurately could student clinicians rate the four dimensions with only minimal instruction beforehand? Second, would they improve over the course of rating the eight samples if they were provided with ongoing feedback in the form of expert dimension scores?

Method

Participants

Seventeen fourth-year undergraduate speech pathology students at The University of Queensland took part in the study (aged 22–44 years; $M = 23.2 \pm 5.3$ years; 14 women, three men). One additional participant also took part but was excluded due to not complying with task instructions. The study was approved by The University of Queensland's Human Research Ethics Committee (2023/HE000950), and all participants provided written informed consent and were compensated for their time.

The participants were recruited by word of mouth among the fourth-year cohort, of which the first author was a member. All were in their final semester of their programs and had recently passed clinical placement courses in which they were required to perform at entry level on the majority of competencies on Speech Pathology Australia's Competency Assessment in Speech Pathology. All had completed courses in aphasia and motor speech disorders. Their clinical experience in aphasia ranged from 0 to 50 hr ($M = 24$ hr). A few months after taking part in the study, all participants graduated and were eligible to become certified practicing speech pathologists in Australia. The participants thus served as a representative cohort of near–entry-level clinicians, possessing relevant knowledge and understanding of aphasia acquired through the completion of required courses in adult neurogenic communication disorders and, in many cases, clinical placements.

Connected Speech Samples From Individuals With Aphasia

Eight videotaped connected speech samples of individuals with chronic poststroke aphasia (aged 53–76 years; four men, four women) were selected from the AphasiaBank database (see Table 1; MacWhinney et al., 2011). All were right-handed monolingual English speakers and were recorded at universities or outpatient clinics across the United States. The connected speech samples were

Table 1. Patient characteristics.

Patient	Age (years)	Sex	Race	Education (years)	Time postonset (months)	WAB-R AQ (/100)	Aphasia type (WAB-R)	Apraxia of speech	Clip
kurland07a	70	F	W	16	13	83.0	Anomic	N	0:21–6:26
ACWT09a	56	F	W	13	94	80.1	Conduction	Y	0:00–5:25
ACWT02a	53	F	W	14	39	74.6	Transcortical motor	Y	0:02–6:07
elman14a	76	F	AA	17	55	65.7	Wernicke	N	0:00–5:27
kurland18a	74	M	AA	16	9	44.0	Wernicke	N	0:25–6:03
TCU08a	57	M	AA	14	95	63.9	Broca	Y	0:00–6:21
TAP09a	71	M	W	16	36	20.5	Global	Y	0:00–6:24
scale09a	66	M	W	12	240	20.3	Global	Y	0:00–6:10

Note. WAB-R = Western Aphasia Battery–Revised; AQ = aphasia quotient; Clip = portion of the audiovisual file excerpted for this study; F = female; W = White; N = no; Y = yes; AA = African American; M = male.

clipped excerpts containing approximately the first 5 min of patient speech, during which each individual with aphasia talked about their speaking abilities, stroke, and recovery and sometimes recounted a memorable life event.

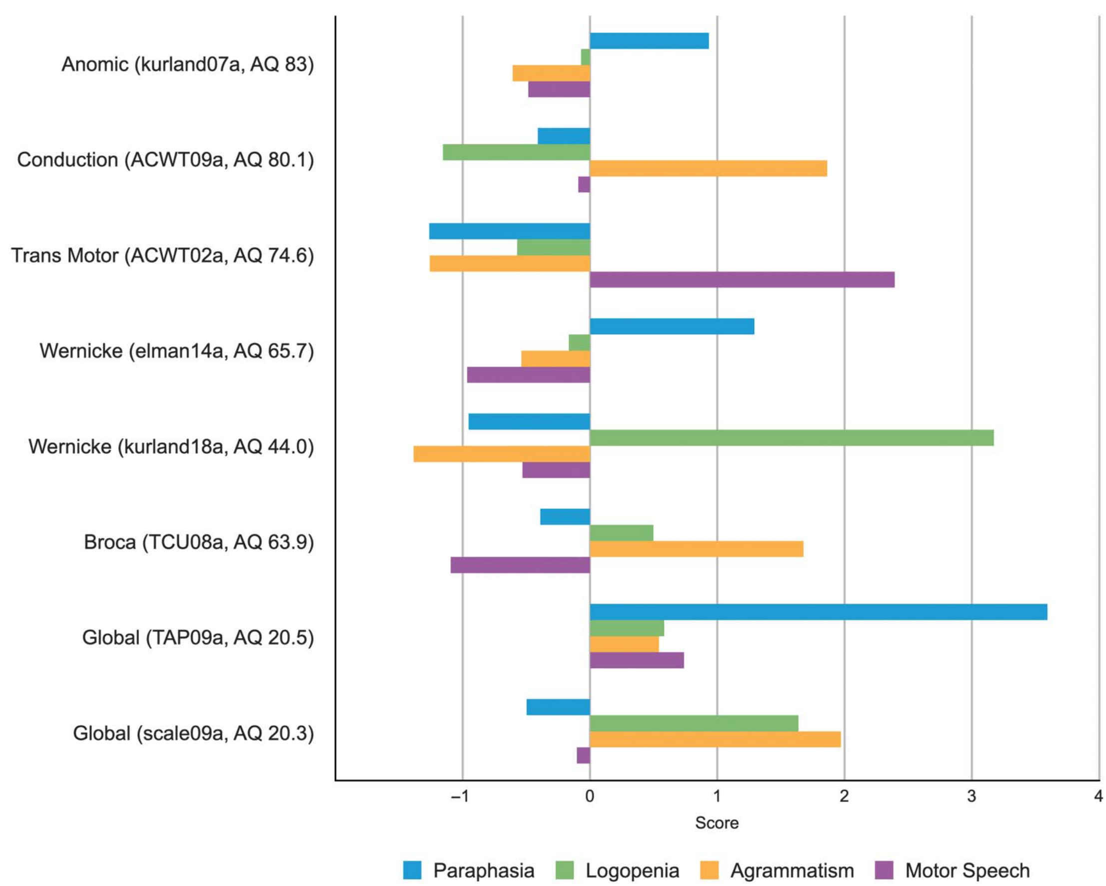
The eight individuals selected were a subset of the 24 included in Casilio et al. (2019). For each individual, scores for the four dimensions—Paraphasia, Logopenia, Agrammatism, and Motor Speech—were derived from the expert ratings of the 27 APROCSA features and the factor analysis reported by Casilio et al. We will refer to these scores as "expert scores," even though they were derived from feature ratings rather than being rated directly. The eight individuals were chosen such that they formed a diverse subset in terms of dimension profiles (distinct degrees of impairment across the four dimensions), aphasia types, and aphasia severities (WAB-R aphasia quotients ranged from 20.3 to 80; see Table 1 and Figure 1). Note that patients of the same aphasia classification (per the WAB-R) can present with dramatically different dimension profiles; for example, one participant with Wernicke's aphasia was predominantly paraphasic, whereas another was predominantly logopenic. This variability appears to reflect genuine differences between individuals in the characteristics of their connected speech, as supported by the strong concurrent validity of APROCSA features with respect to quantitative measures (Casilio et al., 2019).

Rating Procedure

Each participant completed the experiment within a 2-hr period in a quiet room. First, they read a brief two-page manual, which described the conceptual basis of four interacting dimensions, the prototypical features of each dimension, and the rating procedure (see the Appendix). This document emphasized that dimensions may co-occur and encouraged participants to approach the task as they would an auditory-perceptual rating of motor speech or voice disorders.

The four dimensions were each explained in just one paragraph, given our goal of exploring the feasibility of

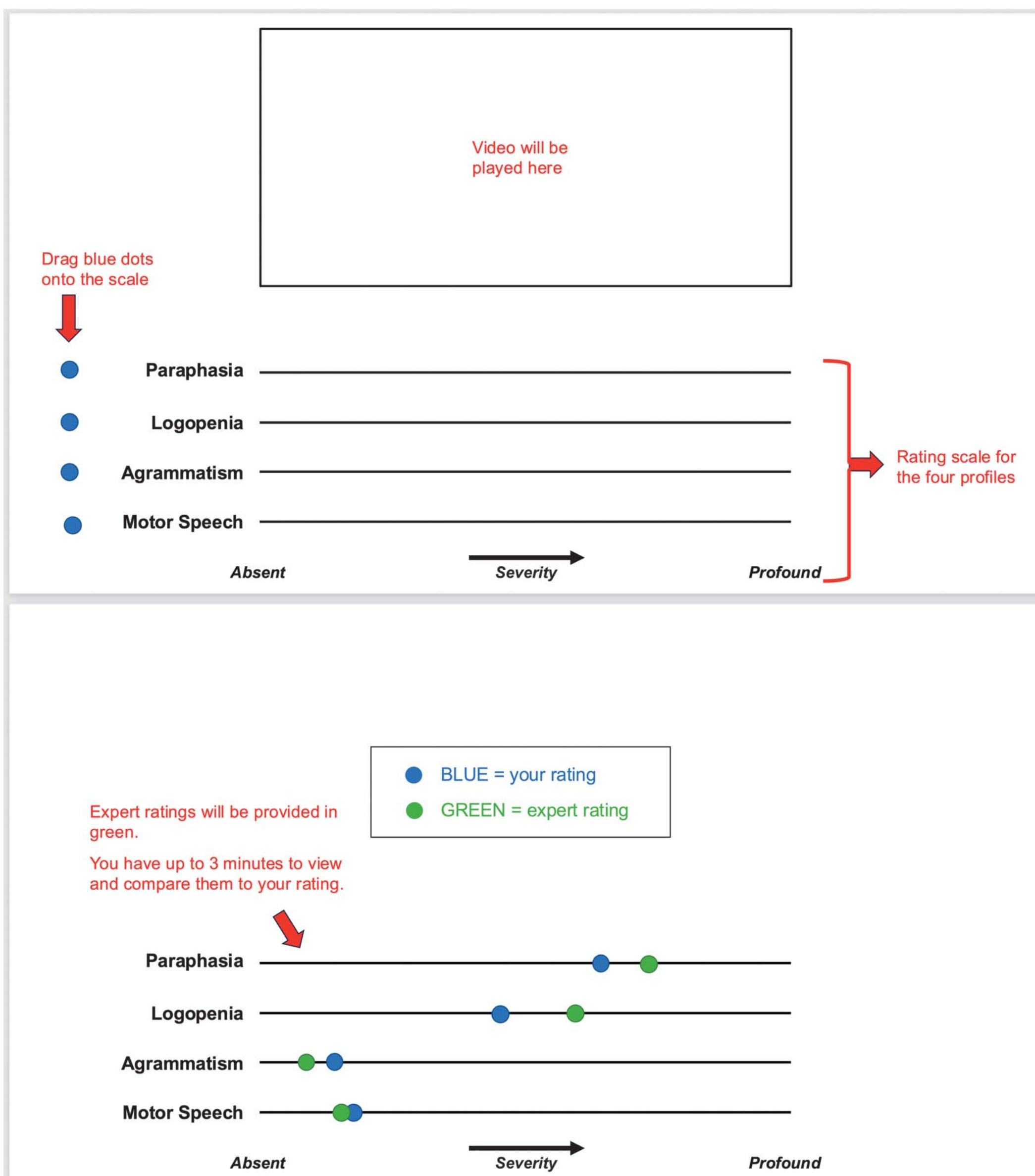
Figure 1. Patient characteristics. For each of the eight patients, the expert scores on each of the four dimensions are shown. Patients are ordered by aphasia type per the Western Aphasia Battery–Revised, with less severe types first, and then by descending aphasia quotient (AQ) within type. Trans Motor = transcortical motor.



rating the dimensions based on foundational knowledge, with minimal additional training. Key defining features within each dimension were highlighted and described individually and in relation to one another. Paraphasia features included semantic and phonemic paraphasia, neologisms, jargon, empty speech, paragrammatism, and retracing. Logopenia features included anomia, pauses within and between utterances, abandoned utterances, and circumlocution. Agrammatism features included short and simplified utterances and omission of bound morphemes and function words. Motor speech encompassed apraxia of speech and dysarthria, but note that none of the eight patients evidenced significant dysarthria.

The eight audiovisual connected speech samples were played in a different random order for each participant on a laptop computer (Apple MacBook Pro [15-in., 2018]). Participants listened to each of the samples once without pausing using Sennheiser HD 280 headphones and then rated the four dimensions by dragging blue dots to four visual analog scales that were labeled "Absent" on one end and "Profound" on the other end (see Figure 2,

Figure 2. Rating procedure. The top panel shows the instructions provided to participants at the start of the experiment. The bottom panel shows how participants were provided with feedback in the form of expert scores (green) that they could compare to their own ratings (blue).



top panel). They were allowed to make their dimension ratings while listening or during a 2-min period after listening. Then, they were shown the expert scores for the speech sample, which were represented with green dots (see Figure 2, bottom panel). They were given up to 3 min to compare the expert scores with their own ratings and were encouraged to reflect on any discrepancies and to seek to rate subsequent samples more closely in line with the expert scores. This was self-reflection only and did not involve any discussion with the experimenter.

Data Analysis

To determine how accurately the student clinicians could rate the four dimensions, we calculated the intraclass correlation coefficient (ICC; McGraw & Wong, 1996) for each participant's ratings of each dimension, in comparison to the expert scores. We calculated ICCs of type (A,1) to quantify absolute agreement, using separate one-way models for each student's ratings of each dimension, with speech sample as a random effect. The visual analog scales were interpreted as extending linearly from −2.2 at the "Absent" end to +3.8 at the "Profound" end in terms of the dimension scores reported by Casilio et al. (2019), since these were the minimum and maximum scores on any dimension among the eight individuals with aphasia in the sample. ICCs were interpreted as poor (ICC < .40), fair (.40 ≤ ICC < .60), good (.60 ≤ ICC < .75), or excellent (ICC ≥ .75), following Cicchetti (1994). ICCs were calculated in MATLAB R2024A using the icc function from the MATLAB File Exchange (Salarian, 2016).

To compare students' accuracy in rating the four dimensions, we first calculated the absolute value of the discrepancy between each student clinician rating and the corresponding expert score. Then, we constructed a linear mixed-effects model of rating discrepancy (544 observations) as a function of dimension (fixed effect with four levels, modeled with dummy variables) and order (fixed effect, treating sequential order as linear and mean centered), with random effects of participant and speech sample, and random slopes for order by participant. Model fitting was performed with the lmer function in R. We assessed the effect of dimension with a likelihood ratio test comparing the full model to a reduced model.

To determine whether participant ratings improved over the course of rating the speech samples with feedback, we evaluated the fixed effect of order (i.e., the potential learning effect) based on the same linear mixed-effects model. We again used a likelihood ratio test comparing the full model to a different reduced model.

Results

Accuracy

The ICCs quantifying similarity between the student clinician ratings and the expert scores are shown in Figure 3. Of the 68 total dimensions rated (4 dimensions × 17 students), about half were rated with good or excellent accuracy. For Paraphasia, the correspondence was excellent (ICC ≥ .75) for three students, good (.60 ≤ ICC < .75) for four students, fair (.40 ≤ ICC < .60) for six students, and poor (ICC < .40) for four students, and the mean ICC across students was .52, 95% CI [.37, .67]. For Logopenia, the correspondence was excellent for one student, good for three students, fair for eight students, and poor for five students, and the mean ICC was .45, 95% CI [.34, .56]. For Agrammatism, the correspondence was excellent for eight students, good for one student, fair for four students, and poor for four students, and the mean ICC was .58, 95% CI [.45, .71]. For Motor Speech, the correspondence was excellent for seven students, good for three students, fair for five students, and poor for two students, and the mean ICC was .63, 95% CI [.51, .74].

The mean absolute discrepancy between the student clinician ratings and the expert scores was 20.5% of the length of the scale, 95% CI [18.8%, 22.2%]. The linear mixed-effects model showed a significant effect of dimension, $\chi^2(3) = 15.26$, $p = .0016$, such that the Motor Speech dimension was rated most accurately, followed by Agrammatism, Paraphasia, and Logopenia. Full model results are reported in Supplemental Material S1.

Learning

The effect of order was not significant, $\beta = -.26 \pm .34$, $\chi^2(1) = 0.55$, $p = .46$, indicating that there was no learning effect (see Figure 4 and Supplemental Material S1). Although there was a slight numerical decrease in absolute discrepancies, this could not be distinguished from chance, so there was no evidence that the feedback provided resulted in more accurate ratings as the participants progressed through the eight speech samples.

Discussion

We found that near–entry-level speech pathology students were able to directly rate the four dimensions using the auditory-perceptual approach to a reasonable extent, with some dimensions rated better than others. However, the students' rating accuracy did not improve over the course of the session, meaning that the minimal training and implicit feedback provided did not have

Figure 3. Accuracy of rating the four dimensions. Box plots of intraclass correlation coefficients (ICCs) show the distributions of rating accuracy across raters. Note that the individual ICC observations were estimated with wide confidence intervals since they were based on just eight speech samples. The central lines indicate the medians, the boxes indicate the interquartile ranges, and the whiskers extend to the most extreme nonoutlier data points.

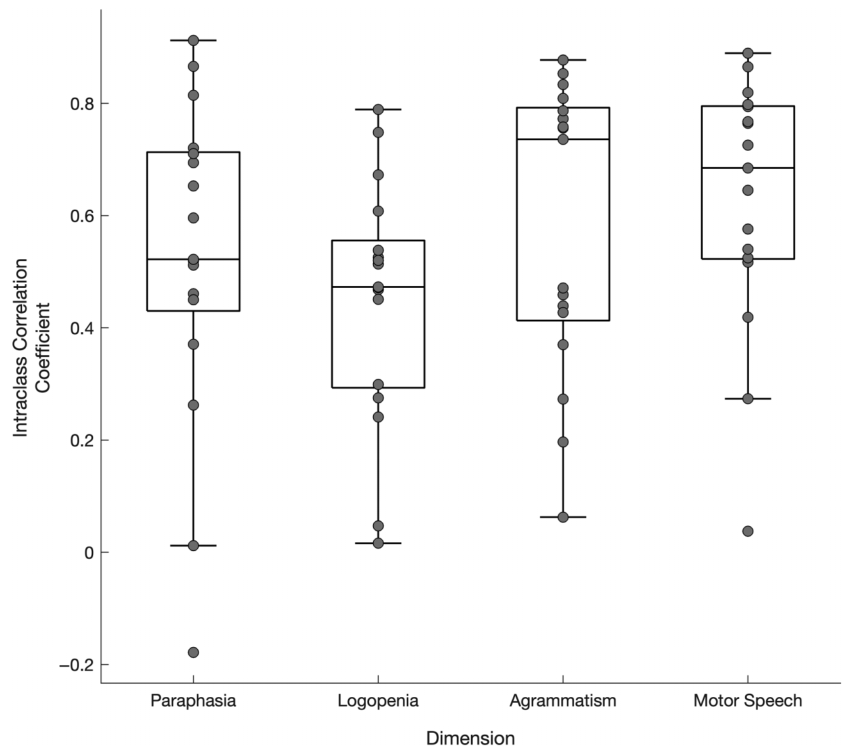
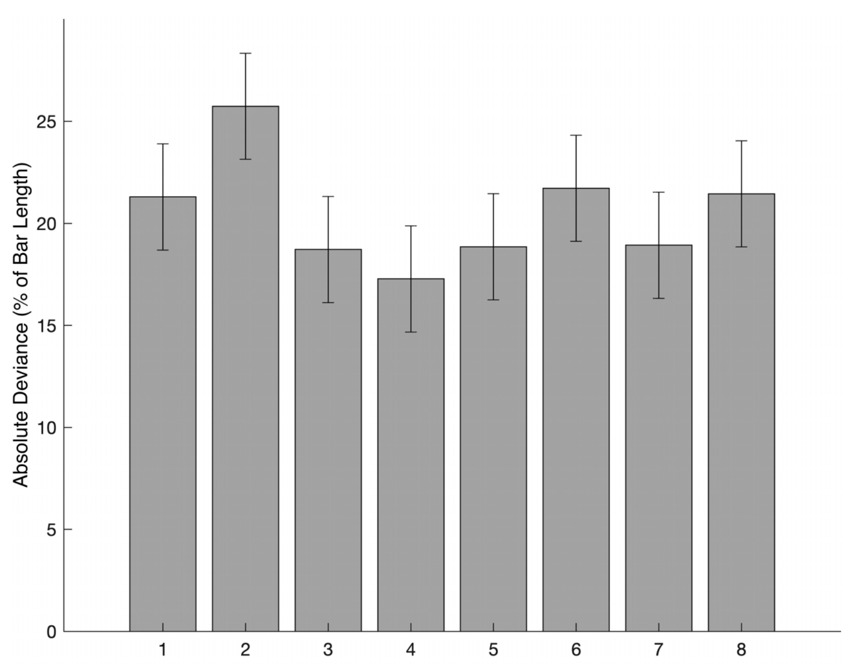


Figure 4. Accuracy of rating as a function of order. Bars show the mean absolute deviation between student clinician ratings and export scores, averaged across the four dimensions. Error bars show standard error of the mean.



significant impacts on their ratings. Although our approach demonstrates potential, it is not yet ready for clinical application.

It is apparent that significantly more training than what was provided in the present study will be required. It is noteworthy that the student clinician participants did best on Motor Speech, where the auditory-perceptual approach is well established and is taught as part of the clinical training program (Duffy, 2012; Strand et al., 2014), and Agrammatism, which is one of the simpler and more salient features of aphasia. In contrast, Paraphasia and Logopenia are concepts that would be less familiar to student clinicians.

To build a clinically useful assessment based on rating these four dimensions directly, it will be necessary to design and validate a training program. Critical components may include more detailed descriptions of the dimensions, examples of speech samples and expert dimension ratings with overt explanations of rationales for the ratings, practice with rating the dimensions, and individualized discussions of discrepant ratings so that the reasons for the discrepancies can be identified and corrected (Wong et al., 2021). Of note, similar principles have been implemented in a validated training program for rating the APROCSA features that give rise to the dimensions described herein (Casilio et al., in press). In addition to developing a training program, it will be necessary to characterize the psychometric properties of the resulting approach and to empirically investigate implementation potential (Stark & Dalton, 2024).

Directly rating the four dimensions could substantially widen the clinical utility of the auditory-perceptual approach. Accurately rating all 27 features of APROCSA requires deep knowledge of aphasia and comes with a higher cognitive load on the part of the rater, who must tease apart individual features, quantify them, and then arrive at an interpretation of how those feature ratings map onto the four dimensions. By contrast, rating the dimensions directly may require less expertise and fewer mental operations to arrive at a clinical profile. Consequently, a dimension rating system could be more feasible for fast-paced clinical settings such as acute care (Shrubsole et al., 2019) and open up the system to other health care providers (e.g., speech-language pathology assistants) who often fill critical gaps in providing speech-language services in rural, marginalized, or underresourced communities (Meeks, 2023).

Limitations

Our study had several limitations that should be acknowledged. First, we had a small number of student clinician participants and a small number of speech samples from individuals with aphasia. Second, the students varied in many important ways, including their knowledge and clinical experience of aphasia, interest in aphasia, and academic abilities. We did not quantify this variability, so we were unable to determine whether any of these features might be associated with performance. Third, the uninterrupted 2-hr sessions may have induced listener fatigue. The demanding nature of engaging in back-to-back auditory-perceptual ratings could have led to diminished focus and motivation of the participants (Walden & Khayumov, 2022), potentially impacting their ratings and/or countering any learning effects. Fourth, the individuals with aphasia were specifically chosen to present with diverse dimension profiles (distinct degrees of impairment across the four dimensions). More variability in the rating targets will tend to increase ICCs (Hallgren, 2012), so observed ICCs might be lower if participants were asked to rate an unselected sample of patients whose speech samples may be less diverse. Fifth and finally, our expert scores did not come from experts rating the dimensions directly but were derived from a factor analysis of expert ratings of the individual APROCSA features. To develop a training protocol and to more precisely assess participant accuracy, it would be better to obtain expert consensus ratings of the dimensions directly.

Conclusions

Our findings indicate that it could be feasible to clinicians to directly rate the four dimensions—Paraphasia, Logopenia, Agrammatism, and Motor Speech—which we have argued reflect coherent underlying deficits that account for many of the observable surface features of connected speech in aphasia (Casilio et al., 2019), have distinct neural correlates (Casilio et al., 2025), and may be useful in informing management and intervention. This could in principle be a highly time-efficient assessment of connected speech. However, to further develop this approach, it will be necessary to develop a training program, likely involving individualized feedback.

Data Availability Statement

The complete deidentified data set is available from the corresponding author on request.

Acknowledgments

The authors gratefully acknowledge support from the School of Health and Rehabilitation Sciences, The

University of Queensland, and the National Institute on Deafness and Other Communication Disorders through Grant R01 DC013270, awarded to Stephen M. Wilson.

References

Casilio, M., Dutta, M., Bryan, K., Crouch, K., Dewall, Z., Feld, I., Kile, J., McMahan, K., Samora, J., Shibata, K., Steel, S., Kasdan, A. V., Walljasper, L., de Riesthal, M., & Wilson, S. M. (in press). A clinical training program for auditory-perceptual rating of connected speech in aphasia. *Journal of Speech, Language, and Hearing Research*.

Casilio, M., Kasdan, A. V., Bryan, K., Shibata, K., Schneck, S. M., Levy, D. F., Entrup, J. L., Onuscheck, C., de Riesthal, M., & Wilson, S. M. (2025). Four dimensions of naturalistic language production in aphasia after stroke. *Brain, 148*(1), 291–312. https://doi.org/10.1093/brain/awae195

Casilio, M., Rising, K., Beeson, P. M., Bunton, K., & Wilson, S. M. (2019). Auditory-perceptual rating of connected speech in aphasia. *American Journal of Speech-Language Pathology, 28*(2), 550–568. https://doi.org/10.1044/2018_AJSLP-18-0192

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Duffy, J. R. (2012). *Motor speech disorders: Substrates, differential diagnosis, and management*. Mosby.

Gordon, J. K. (1998). The fluency dimension in aphasia. *Aphasiology, 12*(7–8), 673–688. https://doi.org/10.1080/02687039808249565

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Kertesz, A. (2006). *Western Aphasia Battery–Revised (WAB-R)*. Pearson.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

Meeks, J. C. (2023). Native American speech-language pathology assistants: Expanding culturally responsive services in reservation schools. *Language, Speech, and Hearing Services in Schools, 54*(2), 395–408. https://doi.org/10.1044/2022_LSHSS-22-00081

Prins, R., & Bastiaanse, R. (2004). Review. *Aphasiology, 18*(12), 1075–1091. https://doi.org/10.1080/02687030444000534

Salarian, A. (2016). *Intraclass correlation coefficient* (ICC). MATLAB Central File Exchange.

Sapienza, C., & Hoffman, B. (2020). *Voice disorders*. Plural.

Shrubsole, K., Worrall, L., & Power, E. (2019). Closing the evidence–practice gaps in aphasia management: Are we there yet? Where has a decade of implementation research taken us? A review and guide for clinicians. *Aphasiology, 33*(8), 970–995. https://doi.org/10.1080/02687038.2018.1510112

Stark, B. C., & Dalton, S. G. (2024). A scoping review of transcription-less practices for analysis of aphasic discourse and implications for future research. *International Journal of Language & Communication Disorders, 59*(5), 1734–1762. https://doi.org/10.1111/1460-6984.13028

Stark, B. C., Dutta, M., Murray, L. L., Bryant, L., Fromm, D., MacWhinney, B., Ramage, A. E., Roberts, A., den Ouden, D. B., Brock, K., McKinney-Bock, K., Paek, E. J., Harmon, T. G., Yoon, S. O., Themistocleous, C., Yoo, H., Aveni, K., Gutierrez, S., & Sharma, S. (2021). Standardizing assessment of spoken discourse in aphasia: A working group with deliverables. *American Journal of Speech-Language Pathology, 30*(1S), 491–502. https://doi.org/10.1044/2020_AJSLP-19-00093

Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. (2014). The Apraxia of Speech Rating Scale: A tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders, 51,* 43–50. https://doi.org/10.1016/j.jcomdis.2014.06.008

Trupe E. H. (1984). Reliability of rating spontaneous speech in the Western Aphasia Battery: Implications for classification. In Brookshire R. (Ed.), *Clinical aphasiology: Proceedings of the conference* (pp. 55–69). BRK.

Vermeulen, J., Bastiaanse, R., & Van Wageningen, B. (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language, 36*(2), 252–274. https://doi.org/10.1016/0093-934X(89)90064-3

Walden, P. R., & Khayumov, J. (2022). The use of auditory-perceptual training as a research method: A summary review. *Journal of Voice, 36*(3), 322–334. https://doi.org/10.1016/j.jvoice.2020.06.032

Wong, D. W.-M., Chan, R. W., & Wu, C.-H. (2021). Effect of training with anchors on auditory-perceptual evaluation of dysphonia in speech-language pathology students. *Journal of Speech, Language, and Hearing Research, 64*(4), 1136–1156. https://doi.org/10.1044/2020_JSLHR-20-00214

Yagata, S. A., Yen, M., McCarron, A., Bautista, A., Lamair-Orosco, G., & Wilson, S. M. (2017). Rapid recovery from aphasia after infarction of Wernicke's area. *Aphasiology, 31*(8), 951–980. https://doi.org/10.1080/02687038.2016.1225276

Appendix (p. 1 of 2)

Participant Manual

Feasibility of Auditory-Perceptual Rating of Four Dimensions Proposed to Characterize Connected Speech in Aphasia

Welcome and thank you for your interest in our study. As a fourth-year speech pathology student, you will be familiar with a range of aphasic speech characteristics, such as anomia, jargon, circumlocution, and so on. In this experiment, you will be rating the connected speech of individuals with aphasia across just four "dimensions," which we call **Paraphasia**, **Logopenia**, **Agrammatism**, and **Motor Speech**. You can think of these dimensions as umbrella terms for groups of aphasic speech characteristics that tend to go together. The four dimensions can co-occur and are not isolated from one another.

This means that the speech of each individual with aphasia may reflect different mixtures of one, two, three, or even all four of the dimensions, each of which can be present to different extents. Your task is to rate all four dimensions, rather than classify an individual into one of the dimensions. The four dimensions are described in further detail in the next section.

You will rate the speech of eight individuals with poststroke aphasia using an auditory-perceptual approach. This means that you will simply listen to their speech samples and rate their speech based on what you hear, just like you would do for a patient with a motor speech or a voice disorder.

The Four Dimensions

As you read through this section, you are encouraged to highlight the key points of each dimension and/or take notes to help yourself understand what to look out for when listening to the speech samples. Each dimension has several key defining features, which will be underlined and described further. You will be able to refer back to this manual when rating the samples.

Paraphasia—Misselection of Words and Sounds

Paraphasia involves the production of unintended substitutions of words and/or sounds. At the word level, semantic paraphasia is when a target word is substituted by a real word that may be related or unrelated to the target, while at the sound level, phonemic paraphasia is when one or more phonemes are substituted, inserted, deleted, or transposed. At the extreme, patients produce neologisms, where the semantic and/or phonemic disruptions are so severe that the target word may be unrecognizable and the speech of individuals who are most impaired on this dimension will take the form of jargon. Several other features commonly contribute to the Paraphasia dimension. Individuals may produce empty speech, whereby there are more function words (e.g., determiners, pronouns, etc.) than content words. This means that their speech conveys little or no meaning due to lack of specificity. Paragrammatism involves the inappropriate juxtaposition of words and phrases and/or misuse of function words and morphemes, such that grammatical elements are present but are jumbled up. The Paraphasia dimension is also characterized by retracing and revisions of sounds, words, and phrases. Overall, the main idea is that the words and sounds in speech are mixed up due to these defining features, thus making it difficult for a listener to understand the intended meaning.

Logopenia—Paucity of Speech

Logopenia is characterized by difficulty finding/retrieving words when speaking, also known as anomia. As a result, pauses within and between utterances are a defining feature as the individual searches for the right word. These pauses may be unfilled (silent) or filled (e.g., "um," "uh") and typically occur before content words. Individuals who are more severely impacted in this dimension will have longer and more frequent pauses when speaking, and word finding is less likely to be successful. Utterances can be abandoned, or individuals may engage in circumlocution or comment on their inability to retrieve words. These word-finding difficulties may also contribute to reduced speech rate, and speech can be perceived as effortful due to strenuous searching for the right words.

Agrammatism—Morphosyntactic Omissions

The Agrammatism dimension is characterized by short and simplified utterances that contain mostly content words. This is due to the omission of bound morphemes (i.e., affixes) and the omission of function words (e.g., determiners, auxiliary verbs). At the extreme, individuals may produce utterances that consist solely of single content words. The main idea is that productions are often short phrases/sentences or single words, are ungrammatical, and contain relatively more content words than function words. This means that a listener may still understand the basic meaning of the individual's responses, as their use of content words is preserved, although their productions are reduced in length and complexity and contain morphosyntactic errors.

Motor Speech—Impaired Speech Motor Processing

The Motor Speech dimension encompasses both apraxia of speech (AOS) and dysarthria. Salient features of AOS include phoneme distortions making the target unclear, reduced overall speech rate, syllable segregations, elongated phonemes, and atypical prosody. Dysarthria is characterized by speech that is slurred, choppy, or mumbled. This impacts speech intelligibility and/or speech naturalness, as one or more of the five speech subsystems (respiration, phonation, articulation, resonance, and prosody) are impaired. In stroke survivors, articulation is almost always the primary subsystem impaired. AOS and dysarthria often co-occur after stroke. Individuals with motor speech impairments often have a halting and effortful quality to their speech and reduced overall speech rate, and their utterances may be difficult to understand.

Appendix (p. 2 of 2)

Participant Manual

Rating Procedure

You will watch and listen to eight speech samples presented in random order, each containing approximately 5 min of patient speech. Then, you will use sliders to make ratings for each of the four dimensions. After viewing each video, you will have 2 min to finalize your ratings before being provided with expert ratings for the video. Your goal should be to rate each speech sample as closely as possible to the expert rating. You will then have 3 min to view the expert ratings and compare them with yours. During this time, you should reflect on how your ratings are similar and/or different to the expert ratings, your reasoning for your ratings, and how you may adjust the way you rate subsequent samples to be closer to the expert ratings.

Important points to consider:

- Each video can only be viewed once, so watch and listen to them carefully.
- You will not be able to discuss with the researcher about the speech sample or ratings.
- You are allowed to take notes while the video is being played if you find this helpful.

If you have any questions, please ask the researcher prior to beginning the rating activity. The researcher will now provide a demonstration of the rating procedure.