Data Descriptor

An Open Dataset of Connected Speech in Aphasia with Consensus Ratings of Auditory-Perceptual Features

Zoe Ezzes [1] , **Sarah M. Schneck** [1], **Marianne Casilio** [1], **Davida Fromm** [2] , **Antje S. Mefferd** [1],
Michael de Riesthal [1] **and Stephen M. Wilson** [1,*]

[1]   Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[2]   Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
*   Correspondence: stephen.m.wilson@vanderbilt.edu

Abstract: Auditory-perceptual rating of connected speech in aphasia (APROCSA) is a system in which trained listeners rate a variety of perceptual features of connected speech samples, representing the disruptions and abnormalities that commonly occur in aphasia. APROCSA has shown promise as an approach for quantifying expressive speech and language function in individuals with aphasia. The aim of this study was to acquire and share a set of audiovisual recordings of connected speech samples from a diverse group of individuals with aphasia, along with consensus ratings of APROCSA features, for future use as training materials to teach others how to use the APROCSA system. Connected speech samples were obtained from six individuals with chronic post-stroke aphasia. The first five minutes of participant speech were excerpted from each sample, and five researchers independently evaluated each sample using APROCSA, rating its 27 features on a five-point scale. The researchers then discussed each feature in turn to obtain consensus ratings. The dataset will provide a useful, freely accessible resource for researchers, clinicians, and students to learn how to evaluate aphasic speech with an auditory-perceptual approach.

Dataset: The dataset can be freely accessed on the Language Neuroscience Laboratory website at: https://langneurosci.org/aprocsa-dataset or through AphasiaBank at: https://doi.org/10.21415/KT40-EA41.

Dataset License: The dataset may be freely used, but only for research, clinical, and educational purposes. Appropriate credit must be given. The dataset may not be used for commercial purposes, nor distributed further.

Keywords: aphasia; connected speech; auditory-perceptual rating

1. Summary

Connected speech is a valuable source of information in aphasia assessment, because it is easy to acquire, yet can reveal underlying impairments in a number of speech/language domains, including lexical access, phonological encoding, syntactic encoding, and speech motor programming [1–3]. Moreover, connected speech is potentially more ecologically valid than the speech and language tasks that are typically performed in aphasia batteries. However, the quantification of speech and language function based on connected speech samples can be time-consuming, and requires considerable expertise and training [3–6].

Recently, Casilio and colleagues [3] described a novel method for auditory-perceptual rating of connected speech in aphasia (APROCSA). Inspired by the auditory-perceptual approach to motor speech assessment [7], they defined 27 features that commonly occur in connected speech in aphasia (e.g., Anomia, Abandoned utterances, Empty speech, Semantic paraphasias), and they specified a five-point scale on which each feature is to be scored: Not present, Mild, Moderate, Marked, or Severe. They then developed a procedure for using

APROCSA, whereby five-minute audiovisual recordings of participant speech are reviewed twice by raters with pre-existing expertise in aphasia. Using data from AphasiaBank [4], they demonstrated that most features could be rated with good-to-excellent interrater reliability by both researchers and student clinicians, and that most features demonstrated excellent concurrent validity with respect to quantitative connected speech measures derived from transcripts. A factor analysis accounted for 79% of the observed variance, with factor loadings supporting four underlying constructs, which were labeled Paraphasia, Logopenia, Agrammatism, and Motor speech.

One potential future direction identified by Casilio and colleagues [3] was the development of materials to support the use of APROCSA in research settings, and to work toward implementation in clinical practice. To support these goals, the aim of the present study was to acquire a set of freely shareable audiovisual recordings of connected speech samples from a diverse group of individuals with aphasia, along with consensus ratings of APROCSA features. This dataset should prove to be a useful resource for scientists, clinicians, and students who are interested in learning how to evaluate aphasic speech samples with an auditory-perceptual approach.

2. Data Description

This dataset contains audiovisual recordings from six individuals with aphasia completing a standardized protocol for connected speech elicitation, as well as an aphasia battery. In addition to the recordings, consensus APROCSA feature ratings and coded transcriptions are available for the connected speech samples, as well as subscores on the aphasia battery, and relevant demographic and clinical variables. The dataset is available on our lab website at: https://langneurosci.org/aprocsa-dataset and through Aphasia-Bank [4] at: https://doi.org/10.21415/KT40-EA41. Access to these materials is unrestricted, however permission is granted only for research, clinical, and educational uses.

2.1. Connected Speech Samples

Connected speech samples were elicited from six individuals with aphasia using the AphasiaBank protocol [4] as described in detail below. Audiovisual recordings of these speech samples are available in mp4 format.

2.2. Demographic, Neurological, and Behavioral Data

Demographic, neurological, and behavioral data for each participant are provided in Table 1. Each individual's aphasia profile was characterized using the Quick Aphasia Battery [8], an efficient, reliable, and multidimensional speech and language evaluation. The administration of this battery is included in each audiovisual recording.

2.3. Consensus Ratings of APROCSA Features

Per the APROCSA protocol, we excerpted segments for analysis comprising approximately five minutes of patient speech. Consensus ratings for all 27 APROCSA features for each participant's connected speech excerpt are presented in Table 2.

2.4. Transcriptions

Complete transcriptions of each speech sample are provided in CHAT format [4], at the same URLs as the audiovisual recordings. Note that the speech samples were transcribed after the APROCSA ratings were completed, so the transcripts played no role in the rating process.

Table 1. Demographic, neurological, and behavioral characteristics of the participants.

	1738	1944	1713	1554	1833	1731
Age	72	71	63	46	67	48
Sex	M	F	F	F	M	M
Handedness	R	R	R	R	A	R
Education (years)	14	16	14	15	14	18
Race	W	B	W	W	W	W
Time post onset (months)	120	151	23	35	18	52
Stroke etiology	I	I	I	I	H	I
Lesion extent (cm^3)	147.2	51.1	29.2	17.8	9.7	218.6
Quick Aphasia Battery						
Word comprehension	9.38	10.00	10.00	10.00	10.00	8.54
Sentence comprehension	9.38	8.13	9.58	9.58	7.71	2.71
Word finding	7.00	5.50	9.00	8.00	7.00	1.50
Grammatical construction	7.75	7.13	7.50	5.13	5.75	0.75
Speech motor programming	5.00	7.50	7.50	7.50	7.50	5.00
Repetition	7.50	8.75	9.17	7.08	7.92	4.58
Reading	7.50	9.17	9.17	8.75	7.92	0.83
Overall	7.72	7.69	8.84	7.96	7.52	3.74

M = Male; F = Female; R = Right; A = Ambidextrous; W = White; B = Black; I = Ischemic; H = Hemorrhagic; Guidelines for interpretation of QAB scores: 0.00–4.99 = Severe; 5.00–7.49 = Moderate; 7.50–8.89 = Mild; 8.90–10.00 = No aphasia.

Table 2. Consensus ratings of APROCSA features for the six participants.

	1738	1944	1713	1554	1833	1731
Anomia	1	3	2	2	2	3
Abandoned utterances	0	2	1	1	2	1
Empty speech	0	2	0	1	1	1
Semantic paraphasias	0	0	1	1	1	2
Phonemic paraphasias	0	0	1	0	0	1
Neologisms	0	0	0	0	0	0
Jargon	0	0	0	0	0	0
Perseverations	0	0	0	0	0	1
Stereotypies and automatisms	0	0	0	0	0	2
Short and simplified utterances	0	1	0	2	1	4
Omission of bound morphemes	0	1	1	1	0	3
Omission of function words	0	0	1	2	2	4
Paragrammatism	1	1	1	1	1	1
Pauses between utterances	1	2	0	2	1	1
Pauses within utterances	2	3	2	2	2	2
Halting and effortful	2	1	1	1	1	2
Reduced speech rate	2	3	1	2	2	2
Retracing	1	3	1	1	2	1
False starts	1	2	1	1	2	1
Conduite d'approche	1	0	1	0	1	0
Target unclear	1	1	0	0	0	1
Meaning unclear	1	1	0	1	1	3
Off-topic	0	0	0	0	0	1
Expressive aphasia	1	2	1	2	2	3
Apraxia of speech	2	1	1	1	1	2
Dysarthria	1	0	0	0	0	0
Overall communication impairment	2	2	1	2	2	3
Total video duration (min:s)	39:07	56:50	36:15	58:03	46:22	74:26
Analyzed sample duration (min:s)	6:56	6:02	5:54	8:48	7:20	7:23

0 = Not present; 1 = Mild; 2 = Moderate; 3 = Marked; 4 = Severe.

3. Methods

3.1. Participants

Six individuals with chronic post-stroke aphasia were recruited at Vanderbilt University Medical Center. The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Vanderbilt University Medical Center (IRB #160847, approved 7 July 2016, amended 15 May 2019). All participants provided written informed consent to take part in the study, and to freely share audiovisual recordings of their connected speech samples.

We recruited only individuals we had worked with previously and, specifically, those who we anticipated would be comfortable with allowing their speech samples to be shared freely. Three participants were originally recruited at the bedside in the first few days after their stroke for a one-year longitudinal study of the neural correlates of language processing, later consenting separately to participate in the present study. The other three participants were originally recruited through the Aphasia Group of Middle Tennessee for a study of the neural correlates of language processing in chronic post-stroke aphasia, also later consenting separately to participate in the present study. One additional participant consented to provide a speech sample, but not to freely share it, so they were not included in the study.

Demographic, neurological, and behavioral data are provided in Table 1. Five participants were monolingual native speakers of English, while the sixth (Participant 1731) spoke Spanish as a first language, moved to the United States at age 7 and learned English, and now reported not remembering much Spanish. Five of the participants had mild to moderate aphasia overall per clinical impression, while the sixth (Participant 1731) had severe aphasia. Word comprehension was largely spared in all participants except for 1731, in whom it was mildly impaired. Sentence comprehension was mildly impaired in all participants except for 1731, in whom it was severely impaired. All patients had expressive deficits that varied considerably across speech and language domains, as can be seen in Tables 1 and 2. All participants had received at least some speech-language therapy in the year(s) following their strokes, however the nature and amount of therapy varied and was not probed in detail.

3.2. Connected Speech Samples

Three participants' connected speech samples were recorded in quiet testing rooms, and three were recorded in their homes due to distance from the testing site. The samples were elicited using the AphasiaBank protocol [4], which includes free speech samples about participants' personal experiences with their strokes and an important life event, three picture descriptions, a narrative storytelling (Cinderella), and a procedural discourse:

1. A free speech sample about participants' personal experiences with their strokes

    Prompt 1: How do you think your speech is these days?
    Prompt 2: Do you remember when you had your stroke?
    Prompt 3: Tell me about your recovery. What kinds of things have you done to get better since your stroke?

2. A free speech sample about an important life event

    Prompt: Thinking back, can you tell me a story about something important that happened to you in your life? It could be happy or sad or from any time—from when you were a kid or more recently.

3. Three picture descriptions (Broken Window; Refused Umbrella; Cat Rescue)
4. A narrative storytelling (Cinderella)
5. A procedural discourse

    Prompt: Tell me how you would make a peanut butter and jelly sandwich.

The connected speech samples were elicited by Z.E., who was a second-year master's student in speech-language pathology who had completed graduate coursework in aphasia

and motor speech disorders and had more than 50 h of clinical experience in aphasia. All data collection was supervised by S.M.S., a licensed speech-language pathologist, who was present for each session.

Each session was recorded with a Canon VIXIA HF S20 camcorder and a Marantz PMD661MKII digital audio recorder. Videos were reviewed and edited using Kdenlive to remove personally identifiable audio content, apart from participants' first names. Video recordings were made because visual information can yield important information in speech/language assessment, such as facial expressions, gestures, and multimodal communication strategies, as well as groping behaviors associated with apraxia of speech.

3.3. Raters

Five of the authors of this article served as the raters, all of whom had substantial experience in assessment of connected speech in aphasia. S.M.S., A.S.M., and M.d.R. were licensed speech-language pathologists. Z.E. was a second-year master's student, as described above. S.M.W. was an experienced aphasia researcher.

3.4. Rating Procedure

Each of the five raters was oriented to the APROCSA system by reading the original article and reviewing the list of features and their definitions [3].

We excerpted segments for analysis comprising approximately the first five minutes of patient speech. For four of the participants, these segments included only content from the discussion of their experience with their strokes, while for Participants 1738 and 1554, parts of the discussions of important life events were included. This is similar to the content included in the samples analyzed in the original APROCSA study [3].

The six participants' speech samples were individually analyzed across six separate meetings, each attended by all raters. The raters listened to the excerpt together, twice in succession. During and immediately after listening to the excerpt, each rater independently rated each of the 27 APROCSA features and wrote down any noteworthy utterances that highlighted particular auditory-perceptual features. Each feature was then discussed in sequence, in the order listed on the APROCSA rating form. Each rater stated their numerical severity rating (i.e., 0–4), then for each feature without perfect agreement, the raters discussed their scores until reaching consensus, listening back to informative parts of the speech sample as necessary. This process took approximately 75 min per sample.

3.5. Analysis of Consensus Ratings

We compared the consensus ratings to simple averages of ratings across the raters, such as were used in the original APROCSA study [3]. As expected, for each participant, the consensus ratings were highly correlated with mean ratings (range $r = 0.87$–0.96). However, the consensus ratings are preferable to the average ratings for two reasons.

First, for 8 of the 27 features, there was at least one expert rating for at least one participant that deviated from the consensus rating by 2 or more points; there were a total of 13 such deviant ratings. These ratings, which significantly differed from the ultimate consensus, would have made average ratings less accurate, but were able to be resolved for consensus scores through discussion.

Second, for 12 of the 27 features, there was at least one participant who was rated as 0 by consensus but non-zero by at least one rater, implying that mean ratings would indicate that a feature was present in the sample, while our consensus determination was that the feature was not present in the sample.

3.6. Limitations

Our dataset has several noteworthy limitations. First, it includes only six individuals with aphasia. Every individual with aphasia has a unique connected speech profile, so any small sample of patients will inevitably not provide exposure to all of the phenomena that will be encountered when assessing aphasic speech.

Second, although the participants who were included were quite diverse in terms of the nature of their aphasias, not all APROCSA features were well demonstrated in the set of speech samples. In particular, two features were not observed at all in the six participants we studied—Neologisms and Jargon—and one feature (Paragrammatism) was considered to be present to the same extent in all six participants. Moreover, many features showed only a limited range of scores in the samples. This entails that a comprehensive training protocol for rating of connected speech in aphasia will require these samples to be supplemented with other aphasic speech samples. In future work, we hope to elicit additional freely sharable speech samples from individuals with aphasia to extend the present dataset.

Finally, per the APROCSA protocol, our ratings were based on the first five minutes of connected speech of the AphasiaBank protocol and, as such, contained only one of the four elicitation methods, free speech. Our previous study demonstrated that five minutes is a sufficient minimum for observing relevant behaviors of connected speech in aphasia [3], and other studies of different kinds of discourse and perceptual features have similarly found that five-minute samples generally suffice [9,10]. However, connected speech features have been observed to differ across elicitation methods [11,12], which we also observed when reviewing the speech samples in their entirety. Future research could further investigate the quantitative and qualitative differences between speech samples obtained by different means of eliciting connected speech.

Author Contributions: Conceptualization, Z.E., S.M.S., M.C. and S.M.W.; investigation: Z.E. and S.M.S.; formal analysis: Z.E., S.M.S., A.S.M., M.d.R. and S.M.W.; data curation, M.C. and D.F.; writing—original draft preparation, Z.E. and S.M.W.; writing—review and editing, S.M.S., M.C., D.F., A.S.M. and M.d.R.; supervision, S.M.W.; funding acquisition, S.M.W. and D.F. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Vanderbilt University Medical Center (IRB #160847, approved 7 July 2016, amended 15 May 2019).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: The dataset can be freely accessed on the Language Neuroscience Laboratory website at https://langneurosci.org/aprocsa-dataset or through AphasiaBank at https://doi.org/10.21415/KT40-EA41.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Vermeulen, J.; Bastiaanse, R.; Van Wageningen, B. Spontaneous speech in aphasia: A correlational study. *Brain Lang.* **1989**, *36*, 252–274. [CrossRef]
2. Prins, R.; Bastiaanse, R. Analysing the spontaneous speech of aphasic speakers. *Aphasiology* **2004**, *18*, 1075–1091. [CrossRef]
3. Casilio, M.; Rising, K.; Beeson, P.M.; Bunton, K.; Wilson, S.M. Auditory-perceptual rating of connected speech in aphasia. *Am. J. Speech-Lang. Pathol.* **2019**, *28*, 550–568. [CrossRef] [PubMed]
4. MacWhinney, B.; Fromm, D.; Forbes, M.; Holland, A. AphasiaBank: Methods for studying discourse. *Aphasiology* **2011**, *25*, 1286–1307. [CrossRef] [PubMed]
5. Yagata, S.A.; Yen, M.; McCarron, A.; Bautista, A.; Lamair-Orosco, G.; Wilson, S.M. Rapid recovery from aphasia after infarction of Wernicke's area. *Aphasiology* **2017**, *31*, 951–980. [CrossRef] [PubMed]

6. 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.; et al. Standardizing assessment of spoken discourse in aphasia: A working group with deliverables. *Am. J. Speech-Lang. Pathol.* **2021**, *30*, 491–502. [CrossRef] [PubMed]

7. Darley, F.L.; Aronson, A.E.; Brown, J.R. Differential diagnostic patterns of dysarthria. *J. Speech Hear. Res.* **1969**, *12*, 246–269. [CrossRef] [PubMed]

8. Wilson, S.M.; Eriksson, D.K.; Schneck, S.M.; Lucanie, J.M. A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLoS ONE* **2018**, *13*, e0192773. [CrossRef] [PubMed]

9. Boles, L.; Bombard, T. Conversational discourse analysis: Appropriate and useful sample sizes. *Aphasiology* **1998**, *12*, 547–560. [CrossRef]

10. Correll, A.; van Steenbrugge, W.; Scholten, I. Judging conversation: How much is enough? *Aphasiology* **2010**, *24*, 612–622. [CrossRef]

11. Armstrong, E. Aphasic discourse analysis: The story so far. *Aphasiology* **2000**, *14*, 875–892. [CrossRef]

12. Fergadiotis, G.; Wright, H.H. Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology* **2011**, *25*, 1414–1430. [CrossRef] [PubMed]