



Published in final edited form as:

Data (Basel). 2022 November ; 7(11): . doi:10.3390/data7110148.

An open dataset of connected speech in aphasia with consensus ratings of auditory-perceptual features

Zoe Ezzes¹, Sarah M. Schneck¹, Marianne Casilio¹, Davida Fromm², Antje Mefford¹, Michael R. de Riesthal¹, Stephen M. Wilson¹

¹Department of Hearing and Speech Sciences, Vanderbilt University Medical Center;

²Department of Psychology, Carnegie Mellon University

Abstract

Purpose—Auditory-perceptual rating of connected speech in aphasia (APROCSA) involves trained listeners rating a large number of perceptual features of speech samples, and has shown promise as an approach for quantifying expressive speech and language function in individuals with aphasia. The aim of this study was to obtain consensus ratings for a diverse set of speech samples, which can then be used as training materials for learning the APROCSA system.

Method—Connected speech samples were recorded from six individuals with chronic post-stroke aphasia. A segment containing the first five minutes of participant speech was excerpted from each sample, and 27 features were rated on a five-point scale by five researchers. The researchers then discussed each feature in turn to obtain consensus ratings.

Results—Six connected speech samples are made freely available for research, education, and clinical uses. Consensus ratings are reported for each of the 27 features, for each speech sample. Discrepancies between raters were resolved through discussion, yielding consensus ratings that can be expected to be more accurate than mean ratings.

Conclusions—The dataset will provide a useful resource for scientists, students, and clinicians to learn how to evaluate aphasic speech samples with an auditory-perceptual approach.

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, such as lexical access, phonological encoding, syntactic encoding, and speech motor programming (Vermeulen et al., 1989; Prins & Bastiaanse, 2004; Casilio et al., 2019). 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 (MacWhinney et al., 2011; Yagata et al., 2017; Casilio et al., 2019; Stark et al., 2021).

Consent for publication

All participants provided written informed consent to make audiovisual recordings of their speech samples freely available online in connection with this publication.

Recently, Casilio and colleagues (2019) described a novel method for auditory-perceptual rating of connected speech in aphasia (APROCSA). Inspired by the auditory-perceptual approach to motor speech assessment (Darley et al., 1969), they defined 27 features that commonly occur in aphasic connected speech, such as Anomia, Abandoned utterances, Empty speech, Semantic paraphasias, and so on, and they specified a five-point scale on which each feature is to be scored (not present, mild, moderate, marked, severe). They showed that most features could be rated with good-to-excellent reliability by researchers or by 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.

The goal of the present study was to acquire a set of sharable connected speech samples from a diverse group of individuals with aphasia, so that they can be used as training materials for researchers, students, and clinicians interested in learning the APROCSA system. We report consensus ratings of each APROCSA feature, for each speech sample.

Method

Participants

Six individuals with chronic post-stroke aphasia were recruited at Vanderbilt University Medical Center, and provided written informed consent to take part in the study. The study was approved by the institutional review board at Vanderbilt University Medical Center. Demographic, neurological, and behavioral data are provided in Table 1.

We recruited only individuals we had worked with previously and whom we anticipated would be comfortable with allowing their speech samples to be shared freely. Three patients were originally recruited at the bedside in the first few days after stroke, and took part in a longitudinal study of the neural correlates of language processing for one year, before later consenting separately to participate in the present study. The other three patients 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, before later consenting separately to participate in the present study. One additional patient consented to provide a speech sample, but not to freely share it, so they were not included in the study.

Connected speech samples

Three connected speech samples were recorded in quiet testing rooms, and three were recorded in participants' homes. We used the AphasiaBank discourse elicitation protocol (MacWhinney et al., 2011). This protocol includes free speech samples about participants' personal experiences with their strokes and an important life event, three picture descriptions, a narrative retell (Cinderella story), and a procedural discourse. Participants were also administered the Quick Aphasia Battery (Wilson et al., 2018) to quantify the nature and severity of their aphasia (Table 1). Each session was recorded with a Canon VIXIA HF S20 camcorder and a Marantz PMD661MKII digital audio recorder. The

audiovisual recordings were edited to remove personally identifying information as far as possible, except that first names were retained.

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. SMS, ASM, and MDR were licensed and certified speech-language pathologists. ZE 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 hours of clinical experience in aphasia. SMW was an experienced aphasia researcher.

Rating procedure

The six speech samples were individually analyzed in six separate meetings, each attended by all five raters. Excerpts containing approximately five minutes of participant speech (plus some examiner speech) were clipped from the beginning of each speech sample. The five raters listened together to the excerpt, twice in succession. During and immediately after listening to the excerpt, each rater independently rated each of the 27 APROCSA features and flagged any noteworthy utterances. Each feature was then discussed in sequence, in the order listed on the APROCSA rating form. For each feature without perfect agreement, we discussed our scores until we reached consensus, listening back to certain informative parts of the speech sample as necessary. This process took approximately 75 minutes per sample.

After the consensus rating procedure was complete, each sample was transcribed, but the transcriptions were not used in the consensus rating process.

Results

The six speech samples are made available at <https://langneurosci.org/aprocsa-dataset> and on AphasiaBank at <https://doi.org/10.21415/KT40-EA41>. Access to these materials is unrestricted, however permission is granted only for research, education, and clinical uses.

The consensus ratings for the 27 features in the six participants are provided in Table 2. All but two features (Neologisms and Jargon) were observed in at least one participant, and of these, all were variable across participants except for Paragrammatism, which was judged to be mild in all six participants.

We compared the consensus ratings to simple averages of ratings across the raters, such as were used in Casilio et al. (2019). As expected, for each patient, 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 patient 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 make average ratings less accurate, but were able to be resolved through discussion. Second, for 12 of the 27 features, there was at least one patient 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.

Discussion

This dataset should prove to be a useful resource for scientists, students, and clinicians who are interested in learning how to evaluate aphasic speech samples with an auditory-perceptual approach.

The main limitation of the dataset is that only six participants were included. The participants who were included were diverse in terms of the nature of their aphasia, but a comprehensive training protocol for rating of connected speech in aphasia may ultimately require more examples from more individuals. 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. This entails that additional speech samples will be required to obtain examples of how these features should be rated.

Finally, per the APROCSA protocol, 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). Although there is evidence to suggest that five minutes is a sufficient minimum for observing relevant behaviors of connected speech in aphasia (Boles & Bombard, 1998; Casilio et al., 2019), connected speech features have been observed to differ across elicitation methods (Armstrong, 2000; Fergadiotis & Wright, 2011), which we also observed when reviewing the speech samples in their entirety.

Acknowledgments

We thank the six individuals with aphasia who participated in this study, and generously agreed to freely share audiovisual recordings of their connected speech samples.

Funding

This research was supported in part by the National Institute on Deafness and Other Communication Disorders (R01 DC013270).

Availability of data and materials

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>.

References

- Armstrong E (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, 14(9), 875–892.
- Boles L, & Bombard T (1998). Conversational discourse analysis: Appropriate and useful sample sizes. *Aphasiology*, 12(7–8), 547–560.
- Casilio M, Rising K, Beeson PM, Bunton K, & Wilson SM (2019). Auditory-perceptual rating of connected speech in aphasia. *American Journal of Speech-Language Pathology*, 28(2), 550–568. [PubMed: 31136232]

- Darley FL, Aronson AE, & Brown JR (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 12(2), 246–269. [PubMed: 5808852]
- Fergadiotis G, & Wright HH (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, 25(11), 1414–1430. [PubMed: 23125474]
- MacWhinney B, Fromm D, Forbes M, & Holland A (2011). AphasiaBank: methods for studying discourse. *Aphasiology*, 25(11), 1286–1307. [PubMed: 22923879]
- Prins R, & Bastiaanse R (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, 18(12), 1075–1091.
- Stark BC, Dutta M, Murray LL, Bryant L, Fromm D, MacWhinney B, Ramage AE, Roberts A, den Ouden DB, Brock K, McKinney-Bock K, Paek EJ, Harmon TG, Yoon SO, 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. [PubMed: 32585117]
- Vermeulen J, Bastiaanse R, & Van Wagensingen B (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language*, 36(2), 252–274. [PubMed: 2465808]
- Wilson SM, Eriksson DK, Schneck SM, & Lucanie JM (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PloS One*, 13(2), e0192773. [PubMed: 29425241]
- Yagata SA, Yen M, McCarron A, Bautista A, Lamair-Orosco G, & Wilson SM (2017). Rapid recovery from aphasia after infarction of Wernicke’s area. *Aphasiology*, 31(8), 951–980. [PubMed: 29051682]

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;

* = Acute lesion extent.

Table 2

Consensus ratings of APROCSA features for the 6 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
Sample duration (total; min:sec)	39:07	56:50	36:15	58:03	46:22	74:26
Sample duration (analyzed; min:sec)	6:56	6:02	5:54	8:48	7:20	7:23

0 = Not present; 1 = Mild; 2 = Moderate; 3 = Marked; 4 = Severe. See Casilio et al. (2019) for detailed definitions of connected speech features and scores.