ASHA American Speech-Language-Hearing Association

JSLHR

Research Note

Automated Analysis of Fluency Behaviors in Aphasia

Davida Fromm,[a] Steffi Chern,[b] Zihan Geng,[b] Mason Kim,[b] Joel Greenhouse,[b] and Brian MacWhinney[a]

[a] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA [b] Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA

ARTICLE INFO

Article History:
Received October 20, 2023
Revision received February 10, 2024
Accepted March 25, 2024

Editor-in-Chief: Julie A. Washington
Editor: William S. Evans

https://doi.org/10.1044/2024_JSLHR-23-00659

ABSTRACT

Purpose: This study explored the use of an automated language analysis tool, FLUCALC, for measuring fluency in aphasia. The purpose was to determine whether CLAN's FLUCALC command could produce efficient, objective outcome measures for salient aspects of fluency in aphasia.
Method: The FLUCALC command was used on CHAT transcripts of Cinderella stories from people with aphasia (PWA; *n* = 281) and controls (*n* = 257) in the AphasiaBank database.
Results: PWA produced significantly fewer total words, fewer words per minute, more pausing, more repetitions, more revisions, and more phonological fragments than controls, with only one exception: The Wernicke's group was similar to the control group in percentage of filled pauses. Individuals with Broca's aphasia had significantly longer inter-utterance pauses and fewer total words than all other aphasia groups. Both the Broca's and conduction aphasia groups had higher percentages of phrase repetitions than the NABW (NotAphasicByWAB) group. The conduction aphasia group also had a higher percentage of phrase revisions than the NABW and the anomic aphasia groups. Principal components analysis revealed two principal components that accounted for around 60% of the variance and related to quantity of output, rate of speech, and quality of output. The Gaussian mixture models showed that the participants clustered in three groups, which corresponded predominantly to the controls, the nonfluent aphasia group, and the remaining aphasia groups (all classically fluent aphasia types).
Conclusions: FLUCALC is an efficient way to measure objective fluency behaviors in language samples in aphasia. Automated analyses of objective fluency behaviors on large samples of adults with and without aphasia can produce measures that can be used by researchers and clinicians to better understand and track salient aspects of fluency in aphasia.
Supplemental Material: https://doi.org/10.23641/asha.25979863

In aphasia, the measurement of fluency is fundamental to assessment, diagnosis, and treatment (Gordon, 1998; Gordon & Clough, 2020). The easy, smooth flow of fluent speech can be disrupted in different ways for different reasons. For example, basic word-finding problems can manifest in frequent pauses, revisions, false starts, and incomplete utterances; agrammatism can manifest in telegraphic speech; coexisting apraxia of speech can manifest in effortful groping, paraphasias, and self-corrections. However, as Gordon (1998) explains, definition and measurement of fluency can be difficult. This study focuses on improving the objectivity and efficiency of fluency measurements in aphasia for clinical and research purposes.

This work adds a new tool to those being used to measure fluency in aphasia. Although many would agree that fluency is among the most salient features in diagnosis and treatment of aphasia, there is little agreement on a definition of fluency or its measurement. Gordon (1998) pointed out the "fuzziness of the fluency concept" (p. 674) and the need to improve its clarity and measurement for clinical and research purposes. Recently, D'Alesio and Roccaforte (2022) reviewed the literature on fluency in

Correspondence to Davida Fromm: fromm@andrew.cmu.edu. ***Disclosure:** Davida Fromm has received salary from the AphasiaBank grant (NIH-NIDCD R01-DC008524) since 2007. The other authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

aphasia and second language acquisition and concluded that no model has a satisfactory definition of fluency, pointing out its conflation of speech (e.g., pauses, rate) and language (e.g., impaired word finding) factors. Factors often associated with fluency are rate of speech, which includes pausing mean length of utterance (MLU), and grammatical complexity (Gordon & Clough, 2022). However, in the field of fluency disorders (stuttering), typical disfluencies include phrase repetitions, word revisions, phrase revisions, pause counts, phonological fragments, and filled pauses; stutter-like disfluencies include behaviors such as prolongations, broken words, blocks, part-word repetitions, and monosyllabic whole-word repetitions (Ratner & MacWhinney, 2018). Interestingly, revisions, repetitions, and sound fragments are not often mentioned as factors that influence fluency judgments in aphasia. The work presented here proposes the use of quantifiable, objective, and automated measures to improve the clarity and consistency in our teaching, clinical work, and research on fluency in aphasia.

Current diagnostic tests typically rely on subjective judgments. The Boston Diagnostic Aphasia Examination (BDAE; Goodglass et al., 2001) requires subjective ratings of fluency for six individual elements: melodic line, phrase length, articulatory agility, grammatical form, paraphasia, and word finding. The Western Aphasia Battery–Revised (WAB-R; Kertész, 2007) uses a single subjective fluency rating score that incorporates those multiple dimensions. Although these dimensions are indeed relevant to assessing fluency, numerous studies have reported issues with the reliability and validity of these subjective rating scores (e.g., Clough & Gordon, 2020; Gordon, 1998; John et al., 2017). Objective measures such as speech rate and various measures of phrase length are often used but typically alone rather than in combination to capture the multidimensional aspects of fluency (see Clough & Gordon, 2020). Recently, software programs have also been used to facilitate objective measurement of speech and language features relevant to fluency in aphasia. Gordon and Clough (2022) used the EVAL command in CLAN (https://talkbank.org/manuals/CLAN.pdf) as well as some acoustic analyses from Praat (https://www.fon.hum.uva.nl/praat/) to compare with clinicians' perceptual ratings of fluency. The current study aims to continue in this direction, using a CLAN command designed specifically to measure multiple aspects of fluency.

Ratner and MacWhinney (2018) reported on automated analysis of fluency behaviors using the FLUCALC command in CLAN, which is free to download (https://dali.talkbank.org/clan/) and allows for a variety of automated analyses of language samples (MacWhinney & Fromm, 2022). FLUCALC was initially developed to characterize patterns of disfluency in children. It provides preconfigured analyses of raw and proportioned counts of individual types of disfluencies. Six of those measures are relevant to aphasia discourse: filled pauses (e.g., *uh*, *um*), word and phrase revisions, word and phrase repetitions, and sound fragments. Two additional measures that are relevant to fluency behaviors in aphasia have recently been added to the FLUCALC program: intra-utterance pause time (total unfilled pause time within an utterance) and inter-utterance pause time (unfilled pause time between the end of one utterance and the beginning of the next by the same speaker).

The purpose of this study was to use FLUCALC on a large database of Cinderella storytelling transcripts from adults with and without aphasia who completed a standard discourse protocol as part of the AphasiaBank project (MacWhinney et al., 2011). More specifically, the research aimed to (a) improve the efficiency and objectivity of fluency measurement in aphasia, (b) determine how aphasia groups differ on automated fluency outcome measures, and (c) determine which automated fluency variables predict type of aphasia. The Cinderella task was chosen from the other protocol tasks for several reasons: It provides a longer and more complex language sample than the picture description and procedural discourse tasks (Stark, 2019), it is more tightly constrained than the free speech tasks, it is the second most frequently used sampling task in aphasia after the BDAE Cookie Theft picture (Bryant et al., 2016), and the storytelling processing demands (characters, sequence of events) plus linguistic and memory demands may produce more disfluent behaviors (Dede & Salis, 2020).

Method

Participants

All participants who produced a Cinderella story as part of the AphasiaBank standard discourse protocol (MacWhinney et al., 2011) were included in this study, yielding a total of 257 controls (151 females, 106 males, M_{age} = 55.1 years, range: 18–89) and 281 people with aphasia (PWA; 120 females, 161 males, M_{age} = 60.7 years, range: 25–90).[1] Participants were from multiple university clinics and aphasia centers in the United States as well as one each from Canada and the United Kingdom. Based on WAB-R Aphasia Quotient (AQ) scores, the PWA had

[1] Twenty participants who produced Cinderella stories were excluded for these reasons: participants (n = 10) who did not have aphasia types (no WAB-AQ administered) and therefore could not contribute to analyses for Aims 2 and 3 and participants with transcortical motor (n = 9) and transcortical sensory (n = 1) aphasia whose groups were too small to include in statistical analyses.

Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

these types of aphasia: 103 anomic, 72 Broca's, 57 conduction, 26 Wernicke's, and 31 who tested above the 93.7 AQ cutoff but were still aphasic (NotAphasicByWAB, NABW; Fromm et al., 2017).

Procedure

Administration of the AphasiaBank discourse protocol was videotaped and then transcribed by trained and experienced transcribers into CHAT files, which could then be analyzed using CLAN program commands (CHAT is the format for the editor in the CLAN program, https://talkbank.org/manuals/CHAT.pdf). Following the guidelines developed for use in the Quantitative Production Analysis, utterances were segmented on the basis of the following hierarchy of indices: syntax, intonation, pause, and semantics (Berndt et al., 2000; Saffran et al., 1989). Filled pauses (&-uh, &-um), word and phrase revisions ([//]), word and phrase repetitions ([/]), and word fragments (&+) were manually coded into the speaker line transcription. In the first example below, the participant (PAR) was about to say, *She's all excited about it*, but after starting the word *excited* (producing the phonological word fragment, &+e), she revised her utterance. The angle brackets surround the phrase that was revised. The second example shows a string of filled pauses and then one repetition of a single word, *the*.

> *PAR: <and &-um she's all> [//] &+e well they're all excited about it.
>
> *PAR: and &-um &-um &-um the [/] the king wants the prince to get married.

Two transcribers reviewed each transcription and reached forced-choice agreement on any discrepancies. Complete transcripts and videos are available at the AphasiaBank website (http://aphasia.talkbank.org/). Nontask-related utterances (e.g., *give me a second*, *what's the word?*) were excluded.

Intra-utterance and inter-utterance pause times were computed automatically from the word and utterance alignment information on the %wor tier in the CHAT transcript. The %wor tier has time stamps for each word. In the example below, the main speaker tier (*PAR) shows the participant's speech output with the utterance time stamp (in ms) at the end. Each time stamp marks the start time and the end time (for the utterance or word) with an underscore in the middle. Once a language sample is transcribed and linked to the media file, the %wor tier is created automatically using a command in the automated Batchalign system (Liu et al., 2023).

> ***PAR: and &-um &+sh they arrive in a glass &-um house**. •1534520_1572980•
> **%wor: and** •1534520_1534940•
> **&-um** •1534940_1535520•
> **& +sh they** •1548220_1548820•
> **arrive** •1548900_1549830• **in** •1550070_1550550• **a** •1550760_1551090• **glass** •1552470_1553320• **&-um** •1556190_1556760• **house** •1572660_1572980•.
>
> ***PAR: &-um and they go to the ball.**
> •1574490_1581940•
> **%wor: &-um** •1574490_1575040•
> **and** •1580180_1580440• **they** •1580440_1580750•
> **go** •1580780_1581000• **to** •1581000_1581170•
> **the** •1581170_1581370• **ball** •1581370_1581940•.

Intra-utterance pause time is computed as the total pause time between words for all utterances in the task divided by the speaker's total time; inter-utterance pause time is calculated as the total pause time between the end of an utterance (1572980 ms in the example) and the beginning of the next utterance (1574490 ms) divided by speaker's total time.

The following CLAN command was used to compute the outcome measures: flucalc +t*par +a +b *.cind.cex. In this command, *+t*par* selects the participant's speech output, *+a* gets the pause time values from the %wor tier in the transcript, *+b* selects word mode analysis (instead of syllable mode), and **.cind.cex* runs the command on all of the Cinderella transcripts extracted from the larger CHAT files using the GEM command. The 11 FLUCALC outcome measures relevant to this study[2] were total utterances, total words, words per minute, % word repetitions, % phrase repetitions, % word revisions, % phrase revisions, % fragments, % filled pauses, intra-utterance pause time, and inter-utterance pause time. The first three measures are typically included in studies of fluency as they reflect the amount and rate of output. The other measures capture behaviors that disrupt the production and perception of fluent speech. For the percent measures, the denominator is total words.

Statistical Method

Analysis of variance (ANOVA) tests (95% confidence level) were used to determine whether the means of the aphasia groups (all aphasia types including NABW)

[2] FLUCALC includes many measures relevant to stuttering (e.g., prolongations, broken words, part-word repetitions) that were not appropriate for this study.

Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

were significantly different for each measure. Tukey's honestly significant difference (HSD) test was used to identify specifically which pairs of aphasia groups had significantly different means in this multiple comparison setting. Principal components analysis (PCA) was used to reduce the dimensionality of the data. PCA reduces the dimensionality of a dataset by transforming a large set of variables into a smaller one that still contains most of the information in the large set. The first principal component (PC1) is formed as a linear combination of the original variables that explains the most variance, and the second principal component (PC2) explains the most variance in what is left once the effect of the first component is removed, and so on. Finally, based on the PCA results, Gaussian mixture models (GMMs) were used to better visualize the separation of the aphasia groups. GMM identifies clusters of subjects, assigns each patient to a cluster, and provides a probability estimate of each cluster that a patient belongs to. One challenge when using GMM is determining the optimal number of clusters to fit the data. To address this, we used the Bayesian information criterion (BIC), which is calculated by balancing the fit of the model with the complexity of the model. By performing GMM and choosing the optimal number of clusters, we were able to gain a deeper understanding of the structure of the data and identify underlying patterns that may not be apparent from the PCA results alone.

Results

The FLUCALC command produced results in spreadsheet format within a matter of seconds for all 538 transcripts. A log-transformation was performed on all outcome measures, because they were strongly right-skewed, yielding transformed distributions that were closer to being normally distributed. These log-transformed variables were used in the subsequent analyses and modeling. Results of the ANOVA indicated significant differences ($p < .001$) for all 11 outcome measures (see Table 1). Post hoc pairwise confidence intervals for group differences on the measures of output and speaking rate revealed that the control group had significantly more total words and more words per minute than every aphasia group, and they had more total utterances than the Broca's and anomic aphasia groups. Across the aphasia groups, the Broca's aphasia group had significantly fewer total words than every other group. Furthermore, their mean words per minute were significantly less than the NABW group. On the eight outcome measures reflecting fluency disruption behaviors, the control group had significantly lower mean values than every aphasia group with only one exception: They were not significantly different from the Wernicke's group on % filled pauses. That is, all groups with aphasia demonstrated significantly more disfluency behaviors than the control group, with that one exception. Among the aphasia groups, those with conduction aphasia had significantly higher percentages of phrase repetitions and phrase revisions than the NABW group. They also had a significantly higher percentage of phrase revisions than the group with anomic aphasia. Those with Broca's aphasia also had a significantly higher percentage of phrase repetitions than the NABW group. Finally, the group with Broca's aphasia had significantly larger inter-utterance pause time than all other aphasia groups. Tukey's HSD post hoc test results are illustrated graphically in Supplemental Material S1.

Table 2 shows the results of the PCA. Based on the scree plot, the first two principal components, PC1 and PC2, account for approximately 60% of the total variance in the data (34.76% and 25.29%, respectively). The values in each column represent the weight (i.e., the coefficient) that each variable contributes to PC1 and PC2, respectively. We highlighted the largest coefficients in absolute value using the gaps in the coefficient values to define these large values. The magnitude of the coefficient indicates the strength of the relationship between the variable and the principal component. We see, for example, that the first principal component is related to differences in total number of words produced, rate of speech (words per minute), and pausing. The second principal component is related to differences in the amount of output (total words and utterances) and the percentage of repetitions, revisions, and phonological fragments in that output.

Figure 1 is a scatterplot of each patient's PC2 versus PC1 score. The color and shape of the points correspond to the different aphasia types. The figure shows how each aphasia group is clustered based on the features in the PCA. A higher score on PC1 (the right side of the figure) indicates more words at a faster rate with less pausing (filled and unfilled). A higher score on PC2 (the top part of the figure) shows more words and utterances as well as more word and phrase repetitions, word and phrase revisions, and phonological fragments. The scatterplot shows relatively clear clusters for the control (teal/+) and Broca's aphasia (light brown/Δ) groups. Participants in the control group had positive scores on the first principal component (PC1), indicating that they differ from the aphasia groups in terms of greater output and fewer pauses. Participants in the Broca's aphasia group had negative scores on PC1, indicating that they differ from all groups in terms of less output and more pauses. The variability of this group on PC2 reflects the heterogeneity and variability in clinical profiles that has been well documented in Broca's aphasia (Alexander, 1988; Drai & Grodzinsky, 2006; Fridriksson et al., 2015; Fromm et al., 2022). With the exception of

Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 1. Analysis of variance results: group differences on FLUCALC outcome measures.

Measures	ANOVA	Control	NABW	Anomic	Broca
Total utterances	*F*(5, 538) = 25.12*	> Broca's > Anomic		< Control	< Control
Total words	*F*(5, 538) = 83.71*	> All aphasia groups			< All aphasia groups
Words/minute	*F*(5, 538) = 83.41*	> All aphasia groups	> Broca's		< NABW
% Word repetitions	*F*(5, 538) = 25.56*	< All aphasia groups			
% Phrase repetitions	*F*(5, 538) = 15.97*	< All aphasia groups	< Broca's < Conduction		> NABW
% Word revisions	*F*(5, 538) = 13.76*	< All aphasia groups			
% Phrase revisions	*F*(5, 538) = 16.23*	< All aphasia groups	< Conduction	< Conduction	
% Fragments	*F*(5, 538) = 33.20*	< All aphasia groups			
% Filled pauses	*F*(5, 538) = 33.71*	< All aphasia groups EXCEPT Wernicke's			
Intra-utterance pause time	*F*(5, 538) = 48.53*	< All aphasia groups			
Inter-utterance pause time	*F*(5, 538) = 62.13*	< All aphasia groups			> All aphasia groups

Note. ANOVA = analysis of variance; NABW = NotAphasicByWAB.
*p < .001.

the Broca's aphasia group, most of the other aphasia groups were clumped together in the lower middle part of the plot, although some participants in the NABW (blue/◊) and Wernicke's aphasia (pink/x) groups showed characteristics similar to the control group.

Based on their PC1 and PC2 scores, participants were assigned to the aphasia group that the GMM algorithm indicated that they most likely belonged to. The algorithm assumes that the data are generated by a mixture of Gaussian distributions, where each distribution corresponds to a different aphasia group in the data. In Figure 2, we let GMM fit six clusters corresponding to the six aphasia groups and assign the most prevalent aphasia type in that cluster to each one. We see three major clusterings: the control group (teal/+), the Broca's group (light brown/Δ), and the rest of the aphasia groups in the lower middle. In Figure 3, GMM was used to determine the optimal number of clusters to fit the data without knowing what the actual aphasia group labels were. The number was chosen based on BIC, which is a statistical measure used for model selection that balances the goodness-of-fit of a model with its complexity (James et al., 2013). This encourages the selection of models that are simpler. Based on the "mclust" R package used for the calculation, the model with the highest BIC value is considered the best model. As seen in Figure 3, the number of clusters chosen was three, which aligns with the PCA results (see Figure 1) and Figure 2, which also suggests that the data can be roughly clustered into three groups corresponding to the control group, the Broca's group, and the remaining aphasia groups. The latter cluster can be considered fluent types (according to the classic fluent–nonfluent dichotomy) of aphasia (anomic, conduction, Wernicke, and NABW), although that does not mean that the output is "normal" in flow.

Table 2. Results of principal components analyses.

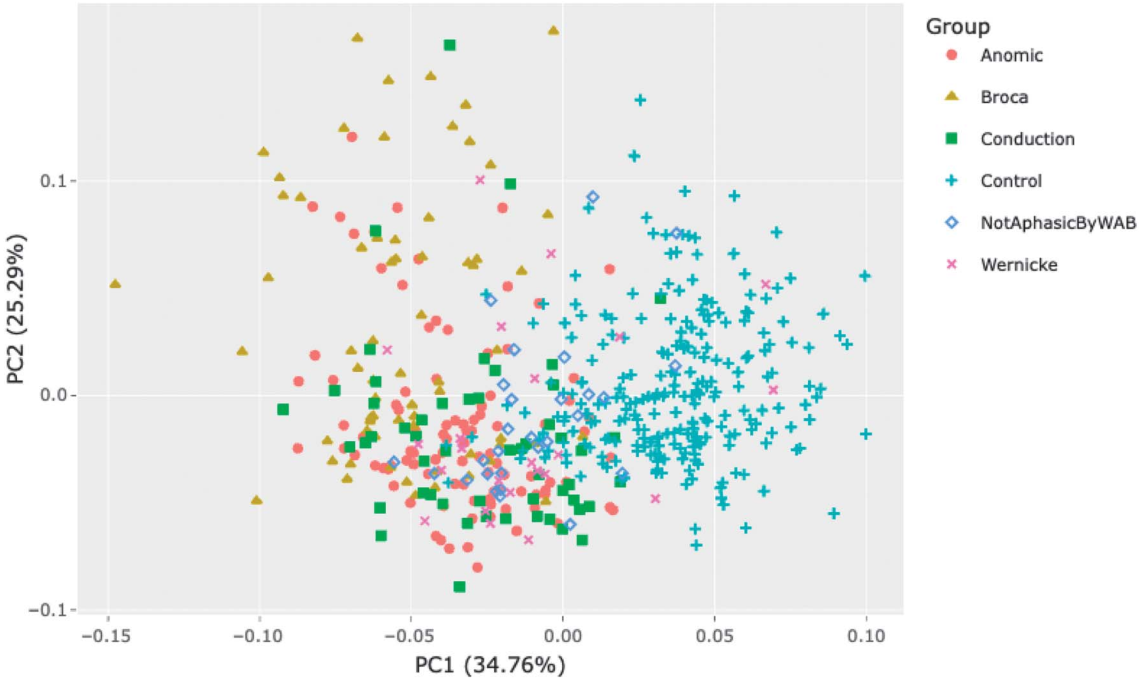
Measures	PC1 coefficients	PC2 coefficients
Log # utterances	0.27	**0.40**
Log # words	**0.36**	**0.37**
Log words per minute	**0.42**	0.09
Log % word repetition	–0.26	**0.36**
Log % phonological fragments	–0.28	**0.33**
Log % phrase repetitions	–0.16	**0.34**
Log % word revisions	–0.16	**0.38**
Log % phrase revisions	–0.07	**0.41**
Log % filled pauses	*–0.35*	0.07
Log % intra-utterance pause time (ms)	*–0.40*	–0.04
Log % inter-utterance pause time (ms)	*–0.37*	–0.16

Note. Numbers in bold are the largest positive coefficients and numbers in italics are the largest negative coefficients in absolute value, respectively, using the gaps in the coefficient values to define these large values. PC1 = first principal component; PC2 = second principal component.

A confusion matrix (see Table 3) was used to evaluate the performance of the GMM in grouping the PWA. The table compares the predicted cluster assignments from the GMMs with the actual aphasia types of the sample. The rows of the matrix correspond to the true aphasia types (according to the WAB-R AQ subtest scores), and the columns correspond to the predicted cluster

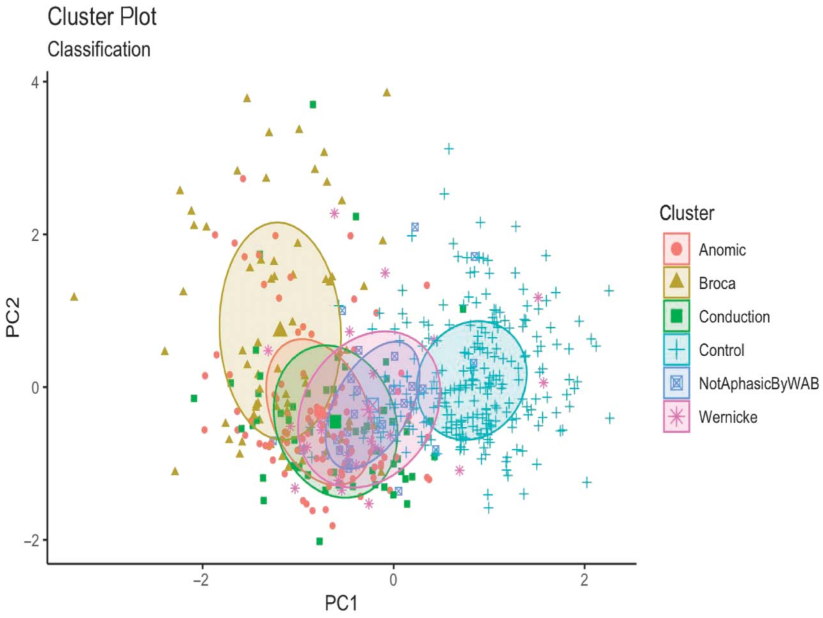
Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 1. Principal components analysis scatterplot by aphasia groups. PC1 = first principal component; PC2 = second principal component; WAB = Western Aphasia Battery.



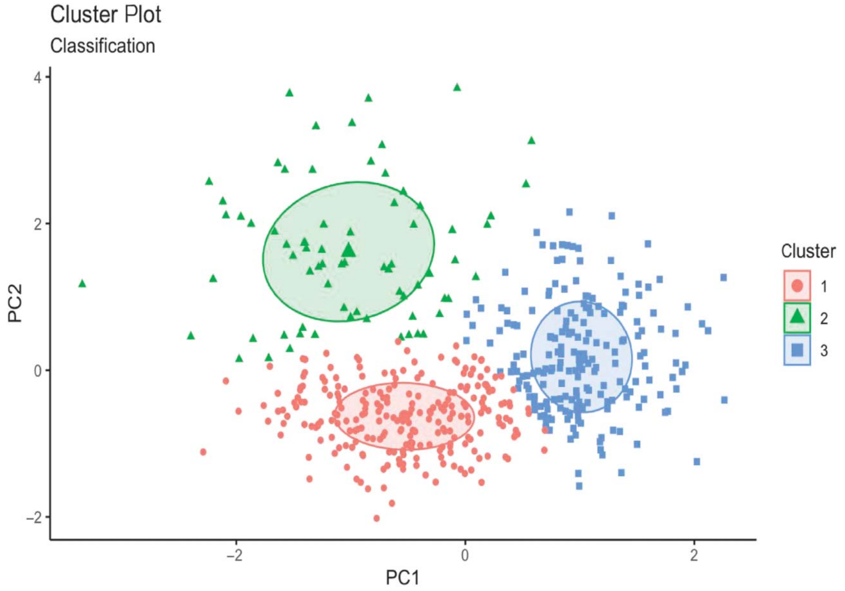
assignments from the GMM in Figure 2. The entries in the matrix represent the number of participants who are correctly or incorrectly classified. The absolute classification error was 22.81%, which was calculated by dividing the total number of misclassified subjects by the total number of subjects. It can be seen that the GMMs have a relatively high proportion of correctly classified patients as control or Broca's aphasia, which is likely due to the

Figure 2. Clustering with aphasia group labels. PC1 = first principal component; PC2 = second principal component; WAB = Western Aphasia Battery.



Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 3. Clustering without aphasia group labels. PC1 = first principal component; PC2 = second principal component.



distinct fluency speech characteristics of these groups compared to the other aphasia groups.

Discussion

This study demonstrates how a tool for automated measurement of fluency behaviors originally designed to characterize patterns of disfluency in children can be used to measure and understand fluency behaviors in aphasia. FLUCALC greatly increases the speed and efficiency of measuring objective fluency behaviors in language samples: amount of output, rate of speech, filled pauses, sound fragments, word and phrase revisions, word and phrase repetitions, and silent intra- and inter-utterance pauses. All aphasia groups produced fewer total words and fewer words per minute than the control group. All aphasia groups also produced more disfluency behaviors (filled and unfilled pauses, fragments, repetitions, and revisions) than the control group with only one exception. Interestingly, that exception was the percentage of filled pauses for the PWA with Wernicke's aphasia. Typically, speakers with Wernicke's aphasia are quite fluent, producing many words and long sentences, often with paraphasias or jargon but with normal rate and prosody and limited use of filled pauses (Buckingham & Kertesz, 1974; Damasio, 1992). The PWA with conduction aphasia showed significantly higher percentages of disfluent behavior than the NABW group for phrase repetitions and revisions. They also had a significantly higher percentage of phrase revisions than the anomic aphasia group. This result is consistent with the repetitive self-corrections (termed "conduites d'approches") that are characteristic of the language output of individuals with this type of aphasia (Bartha & Benke, 2003). Broca's aphasia, the one classically nonfluent aphasia group in the study, had significantly less output as evidenced by significantly fewer total words and longer inter-utterance pause time than all other aphasia groups. It is also worth mentioning that although participants in the NABW group tested above the normal cutoff on the WAB-R AQ subtests, they still produced connected speech that was significantly more disfluent than the controls. This is consistent with several other reports in the literature on the discourse of these individuals (Dalton & Richardson, 2015; Fromm et al., 2017; Gordon, 2020; Richardson et al., 2021).

PCA results revealed that the first principal component related to differences in speaking rate, total words, and pausing, whereas the second principal component related to differences in the amount of output and repair behaviors. The larger coefficients (both positive and negative weights) in PC1 were for total words, words per minute, intra- and inter-utterance pause time, and filled pauses. Those with higher PC1 coefficients produced more words at a faster rate with less pausing (filled or unfilled). The larger PC2 coefficients were for total words and utterances, word and phrase repetitions, word and phrase revisions, and phonological fragments. Thus, this output contained more overt fluency disruptors, perhaps indicating more self-monitoring and more attempts to repair errors (e.g., lexical, morphological, phonological) or more issues with word finding or other language production problems. This component suggests a relationship between greater

Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 3. Confusion matrix: number of correct and incorrect predictions.

Class	Predicted					
	Control	Anomic	Broca’s	Conduction	NABW	Wernicke’s
Control	232	6	0	1	3	2
Anomic	6	76	8	4	1	4
Broca’s	0	7	50	5	2	2
Conduction	3	22	4	20	1	3
NABW	12	7	0	1	7	1
Wernicke’s	2	6	1	3	0	11

Note. NABW = NotAphasicByWAB.

output and more behaviors reflecting some type of linguistic challenges (e.g., lexical retrieval, grammatical formulation) and/or repair.

The GMM suggested three major clusters based on the FLUCALC variables. The clusters aligned with the broad clinical categories of controls, nonfluent aphasia (Broca’s), and fluent aphasia (NABW, anomic, conduction, and Wernicke’s), illustrating the potential clinical relevance of these fluency outcome measures. That is, those with higher PC1 scores are mainly speakers from the control group. However, some participants in the NABW and Wernicke’s groups looked similar to participants in the control group, suggesting that the fluency disturbances of their connected speech symptoms were relatively mild. Those with higher PC2 scores are mainly speakers with Broca’s aphasia, likely the more mildly impaired ones with more output and the ability to self-monitor and make repairs. The fact that the conduction group did not yield its own distinct cluster suggests that individual measures (e.g., percent word and phrase revisions) may not be enough to differentiate them from the others. It may also mean that the impairment in conduction aphasia is a unique kind of aphasia with both fluent and disfluent features that overlap other aphasia types (Gordon & Clough, 2022). This suggests that we need to continue searching for the most sensitive combination of measures to capture the multidimensional behaviors of fluency.

The confusion matrix suggested that there are underlying patterns in speech fluency that are not fully accounted for by the aphasia types assigned based on WAB-R AQ subtest scores. Specifically, it revealed that some participants who were classified as having a particular aphasia type based on their AQ score actually exhibited speech characteristics that were more similar to a different aphasia group or the control group. In this case, the measured fluency characteristics (e.g., fillers, pauses, repetitions, revisions) of spontaneous speech overlapped among the various types of aphasia. This confirms known issues with the fluency scoring system in the Spontaneous Speech subtest of the WAB-R (Crary et al., 1992; Ferro & Kertesz, 1987; Fromm et al., 2022; John et al., 2017; Swindell et al., 1984; Trupe, 1984). It also confirms the difficulty inherent in defining, describing, and assessing fluency in aphasia and underscores the importance of paying attention to spontaneous speech behaviors that facilitate or interfere with successful communication. After examining four continuous fluency measures (WAB-R fluency scale, MLU, speech rate, and retracing) using linear regression, Gordon and Clough (2020) concluded that each represented different aspects of fluency. They emphasized the importance of considering the multidimensionality of fluency, which in the current study was constrained to fluency characteristics in spontaneous speech but did not consider other aspects of fluency such as utterance length, paraphasias, and syntax from the WAB-R fluency scoring scale. We cannot expect any particular test or classification system to capture all of the nuances and complexities of the condition. Further research is needed to investigate these potential patterns and to develop more accurate and comprehensive methods for classifying and characterizing aphasia.

This experimental application of FLUCALC for use in aphasia has demonstrated useful outcome measures that can efficiently and objectively measure salient features of fluency in aphasia, including overall amount and rate of output as well as frank fluency disruptors. Measures of total words, total utterances, words per minute, word and phrase revisions, inter- and intra-utterance pause time, word repetition, and filled pauses were among the important fluency variables identified in the PCA analysis. Clinicians are encouraged to use this tool to identify and measure the behaviors contributing to a patient’s disfluency and track changes over time. Such measurements could be relevant to a variety of treatment goals such as decreasing the use of filled pauses or improving self-monitoring and repair of errors. Even if aphasia treatment is not specifically focused on fluency but rather on naming, syntax, or increasing MLU, clinicians could measure the ways in which that treatment affects these various aspects of fluency. Improvements in automatic speech recognition allow for transcripts to be created with much less time and effort, making these automated analyses more convenient.

Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

The results also confirm much of what we already know to be challenging about this topic. That is, while all measures distinguished the control group from the groups with aphasia (with one exception), and phrase revisions and repetitions distinguished the group with conduction aphasia from the NABW and anomic groups, only two measures distinguished the group with Broca's aphasia from all the other aphasia groups. While this may have something to do with the multifaceted presentation of Broca's aphasia, this also reminds us of the multidimensional aspect of fluency and the fact that we may need to add different or additional measures (e.g., MLU, incomplete utterances) to the FLUCALC program specifically for aphasia.

Limitations of this study can help clarify directions for future research. For example, the storytelling task was selected for reasons including its processing demands. Results might be different for a simple picture description task, a free speech task, or a procedural discourse task. The findings here should also be considered a preliminary attempt to explore objective, automated measures that were developed for FLUCALC. Given that this program was originally developed for use in childhood stuttering, it would be important to establish reliability and validity for these FLUCALC measures as they relate to fluency in aphasia. It would also be useful to consider adding other measures to the program that may help capture more relevant aspects of fluency for this population. Continued development of FLUCALC will include norms and benchmarks for these measures to allow comparisons of results to those of a reference group in the AphasiaBank database. As with the EVAL program, a spreadsheet will display a participant's results compared with mean scores of the comparison group and indicate where the participant differs by one or more standard deviations. Finally, supervised learning methods can be used to build models that can accurately classify aphasia patients based on their speech characteristics. We encourage others to explore the utility of this tool in their work in this area and contribute to the development of psychometrically sound and clinically relevant fluency measures for this population.

Data Availability Statement

The data reported here are available to members of the AphasiaBank consortium https://aphasia.talkbank.org/. Established researchers and clinicians working with aphasia who are interested in joining the consortium should read the Ground Rules (https://talkbank.org/share/) and then send an email to macw@cmu.edu with contact information, affiliation, and a brief statement explaining reasons for joining.

Acknowledgments

This work was supported by the AphasiaBank grant, R01-DC008524, made to Brian MacWhinney from National Institute on Deafness and Other Communication Disorders for 2022–2027. The authors wish to thank the participants whose data were used in this study and the clinical researchers who collected and contributed data to the shared AphasiaBank database.

References

Alexander, M. P. (1988). Variability in the syndrome of Broca's aphasia in a rehabilitation hospital: Implications for research strategies. *Aphasiology, 2*(3–4), 219–223. https://doi.org/10.1080/02687038808248913

Bartha, L., & Benke, T. (2003). Acute conduction aphasia: An analysis of 20 cases. *Brain and Language, 85*(1), 93–108. https://doi.org/10.1016/S0093-934X(02)00502-3

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E., & Schwartz, M. (2000). *Quantitative Production Analysis: A training manual for the analysis of aphasic sentence production*. Psychology Press.

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics, 30*(7), 489–518. https://doi.org/10.3109/02699206.2016.1145740

Buckingham, H. W., Jr., & Kertesz, A. (1974). A linguistic analysis of fluent aphasia. *Brain and Language, 1*(1), 43–61. https://doi.org/10.1016/0093-934X(74)90025-X

Clough, S., & Gordon, J. K. (2020). Fluent or nonfluent? Part A. Underlying contributors to categorical classifications of fluency in aphasia. *Aphasiology, 34*(5), 515–539. https://doi.org/10.1080/02687038.2020.1727709

Crary, M. A., Wertz, R. T., & Deal, J. L. (1992). Classifying aphasias: Cluster analysis of Western Aphasia Battery and Boston Diagnostic Aphasia Examination results. *Aphasiology, 6*(1), 29–36. https://doi.org/10.1080/02687039208248575

D'Alesio, V., & Roccaforte, M. (2022). On fluency: Perspectives from aphasiology and second language acquisition research. *International Journal of Linguistics, 14*(2), 1–16. https://doi.org/10.5296/ijl.v14i2.19763

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology, 24*(4), S923–S938. https://doi.org/10.1044/2015_AJSLP-14-0161

Damasio, A. R. (1992). Aphasia. *The New England Journal of Medicine, 326*(8), 531–539. https://doi.org/10.1056/NEJM199202203260806

DeDe, G., & Salis, C. (2020). Temporal and episodic analyses of the story of Cinderella in latent aphasia. *American Journal of Speech-Language Pathology, 29*(1S), 449–462. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0210

Drai, D., & Grodzinsky, Y. (2006). A new empirical angle on the variability debate: Quantitative neurosyntactic analyses of a large data set from Broca's aphasia. *Brain and Language, 96*(2), 117–128. https://doi.org/10.1016/j.bandl.2004.10.016

Ferro, J. M., & Kertesz, A. (1987). Comparative classification of aphasic disorders. *Journal of Clinical and Experimental Neuropsychology, 9*(4), 365–375. https://doi.org/10.1080/01688638708405057

Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Fridriksson, J., Fillmore, P., Guo, D., & Rorden, C. (2015). Chronic Broca's aphasia is caused by damage to Broca's and Wernicke's areas. *Cerebral Cortex, 25*(12), 4689–4696. https://doi.org/10.1093/cercor/bhu152

Fromm, D., Forbes, M., Holland, A., Dalton, S. G., Richardson, J., & MacWhinney, B. (2017). Discourse characteristics in aphasia beyond the Western Aphasia Battery cutoff. *American Journal of Speech-Language Pathology, 26*(3), 762–768. https://doi.org/10.1044/2016_AJSLP-16-0071

Fromm, D., Greenhouse, J., Pudil, M., Shi, Y., & MacWhinney, B. (2022). Enhancing the classification of aphasia: A statistical analysis using connected speech. *Aphasiology, 36*(12), 1492–1519. https://doi.org/10.1080/02687038.2021.1975636

Goodglass, H., Kaplan, E., & Weintraub, S. (2001). *BDAE: The Boston Diagnostic Aphasia Examination*. Lippincott Williams & Wilkins.

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

Gordon, J. K. (2020). Factor analysis of spontaneous speech in aphasia. *Journal of Speech, Language, and Hearing Research, 63*(12), 4127–4147. https://doi.org/10.1044/2020_JSLHR-20-00340

Gordon, J., & Clough, S. (2020). How fluent? Part B. Underlying contributors to continuous measures of fluency in aphasia. *Aphasiology, 34*(5), 643–663. https://doi.org/10.1080/02687038.2020.1712586

Gordon, J. K., & Clough, S. (2022). How do clinicians judge fluency in aphasia? *Journal of Speech, Language, and Hearing Research, 65*(4), 1521–1542. https://doi.org/10.1044/2021_JSLHR-21-00484

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with applications in R. Springer. https://doi.org/10.1007/978-1-4614-7138-7

John, A. A., Javali, M., Mahale, R., Mehta, A., Acharya, P. T., & Srinivasa, R. (2017). Clinical impression and Western Aphasia Battery classification of aphasia in acute ischemic stroke: Is there a discrepancy? *Journal of Neurosciences in Rural Practice, 8*(1), 074–078. https://doi.org/10.4103/0976-3147.193531

Kertész, A. (2007). *Western Aphasia Battery*. PsychCorp.

Liu, H., MacWhinney, B., Fromm, D., & Lanzi, A. (2023). Automation of language sample analysis. *Journal of Speech, Language, and Hearing Research, 66*(7), 2421–2433. https://doi.org/10.1044/2023_JSLHR-22-00642

MacWhinney, B., & Fromm, D. (2022). Language sample analysis with TalkBank: An update and review. *Frontiers in Communication, 7*. https://doi.org/10.3389/fcomm.2022.865498

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

Ratner, N. B., & MacWhinney, B. (2018). Fluency Bank: A new resource for fluency research and practice. *Journal of Fluency Disorders, 56,* 69–80. https://doi.org/10.1016/j.jfludis.2018.03.002

Richardson, J. D., Dalton, S. G., Greenslade, K. J., Jacks, A., Haley, K. L., & Adams, J. (2021). Main concept, sequencing, and story grammar analyses of Cinderella narratives in a large sample of persons with aphasia. *Brain Sciences, 11*(1), Article 110. https://doi.org/10.3390/brainsci11010110

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37*(3), 440–479. https://doi.org/10.1016/0093-934X(89)90030-8

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American Journal of Speech-Language Pathology, 28*(3), 1067–1083. https://doi.org/10.1044/2019_AJSLP-18-0265

Swindell, C. S., Holland, A. L., & Fromm, D. (1984). Classification of aphasia: WAB type versus clinical impression. In *Clinical Aphasiology: Proceedings of the Conference 1984* (pp. 48–54). BRK Publishers.

Trupe, E. H. (1984). Reliability of rating spontaneous speech in the Western aphasia battery: Implications for classification. In *Clinical Aphasiology: Proceedings of the Conference 1984* (pp. 55–69). BRK Publishers.

Downloaded from: https://pubs.asha.org CASA Institution Identity on 06/19/2024, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions