Measuring Inner Speech Objectively and Subjectively in Aphasia

Julianne M. Alexander[1]*, Peter Langland-Hassan[2], and Brielle C. Stark[1]

[1] Department of Speech, Language and Hearing Sciences and Program in Neuroscience, Indiana University, Bloomington, Indiana, USA

[2] Department of Philosophy, University of Cincinnati, Cincinnati, Ohio, USA

Abstract

Background

Many people with aphasia and people without brain injury talk to themselves in their heads, i.e., have "inner speech." Inner speech may be more preserved compared with spoken speech for some people with aphasia and may serve a variety of functions (e.g., emotion regulation), which motivates us to provide a high-fidelity characterization of it. Researchers have used multiple methods to measure this internal phenomenon in the past, which we combine here for the first time in a single study.

Aims

We compare performance between individuals with and without aphasia on inner speech tasks that assess inner speech "in-the-moment" to general subjective impressions of inner speech to tease apart the relationship of aphasia severity to inner speech.

Methods and Procedures

Twenty people with mild-moderate aphasia and twenty neurotypical controls completed several inner speech tasks, including objective silent rhyme judgements (picture, written, and auditory), subjective reports of inner speech during naming, and subjective rating scales about inner speech experience more generally.

Outcomes and Results

In-the-moment inner speech during silent rhyming tasks was associated with aphasia severity only for picture and written rhyming but not auditory rhyming. In-the-moment inner speech reports during silent naming were not associated with aphasia severity, nor were the subjective ratings about general inner speech experience. Individuals with and without aphasia demonstrated a variety of subjective general inner speech experiences, demonstrating heterogeneity of this phenomenon more broadly.

Conclusions

Methods of measuring inner speech complement each other and speak to different facets of the inner speech phenomenon, and clinicians and researchers must carefully choose the method(s) that will provide the information about inner speech that they desire.

Keywords

Aphasia, inner speech, methods, severity

Introduction

Inner speech is the familiar experience of a "little voice in the head" (Langland-Hassan, 2021; Perrone-Bertolotti et al., 2014). Research in neurotypical individuals and individuals with aphasia suggests that inner speech retains lexical and phonological properties (Oppenheim & Dell, 2008), and is associated with brain areas closely associated with language production (Fama et al., 2017; Geva & Fernyhough, 2019) and with cognitive processes such as phonological working memory (Fama, Henderson, et al., 2019; Stark et al., 2017). Inner speech is not experienced by all, however; some estimates suggest that people experience inner speech with a wide range (0-75%) of frequency (Heavey et al., 2019; Hurlburt et al., 2021). Compellingly, we and others have shown that inner speech, compared to overt speech, is relatively preserved in a majority of individuals with aphasia (Fama et al., 2017; Geva, Bennett, et al., 2011; Stark et al., 2017), which has led to the hypothesis that inner speech may have a consequential role in language recovery in aphasia (Fama & Turkeltaub, 2020; Stark et al., 2017).

Inner speech is complex. One method of conceptualizing inner speech's multifaceted nature is the ConDialInt model, which describes three dimensions of inner speech: condensation, dialogality, and intentionality (Grandchamp et al., 2019). On the condensation dimension, inner speech ranges from condensed (without all acoustic, phonological, and syntactic information) to expanded (including articulatory and auditory properties). Dialogality dimension refers to whether the inner speech captures just one voice/perspective, like a monologue or inner soliloquy, or whether it is dialogical (e.g., an inner verbal conversation in one's own voice or with the addition of other voices). For the intentionality dimension, inner speech can be purposeful, as when intentionally rehearsing information or manipulating it silently, or it can be unintentional/spontaneous, as in daydreaming or mind wandering (Grandchamp et al., 2019). Therefore, different methods for measuring inner speech likely also target different facets of inner speech.

Within the aphasia literature, there are two definitions of inner speech, differing based on whether inner speech involves active manipulation of the word form in the head (as in Geva, Bennett, et al., 2011; Stark et al., 2017), or access to the word form without the need for manipulation (as in, e.g., Fama, Snider, et al., 2019). Notably, both definitions evaluate what we call 'in-the-moment' inner speech, meaning that they are aiming to tap into inner speech generated during isolated moments of time.

It is not surprising, given the differing definitions of inner speech from the aphasia literature, that researchers have elected to measure inner speech using different, though complementary, methods: objective and subjective tasks. Inner speech involving manipulation has been most often measured objectively using silent rhyming decisions, where individuals are shown either two pictures, two written words, or listen to two aurally presented words and then make a decision via button press about whether or not they rhyme. Therefore, within objective tasks measuring inner speech, cognitive processes differ. Auditory judgement of rhymes relies on phonological access and the phonological loop (sometimes termed phonological working memory; Baddeley & Hitch, 2019). Written judgement of rhymes relies on orthographic-to-phonologic conversion, especially in instances where two words rhyme but have dissimilar orthography (e.g., bear, chair), in addition to the phonological loop. Rhyme judgement using pictures requires access to lexical-semantic and phonological information via the visual route, and also the phonological loop. A benefit to objective measurement is that these methods assure that the participants are completing the tasks asked of them (e.g., by responding via button press to a rhyme judgement) and because the experimenters largely control the situation (e.g., by providing stimuli that range in orthographic match and by monitoring the subject to ensure that they are not moving their lips or tongue) (Geva, Jones, et al., 2011). Objective methods like rhyme judgment require the extra step of manipulation beyond access to the word, making it more difficult to separate inner speech access from cognitive-linguistic abilities such as phonological working memory. Yet, manipulation is not exactly equivalent to phonological working memory. While Geva and colleagues (2011) identified that rhyme judgement requires working memory, they did not find that homophone judgement relied on phonological working memory to the same extent. Their definition of inner speech, however, requires manipulation.

Subjective assessment of inner speech in aphasia has likewise relied on measuring in-the-moment experience of inner speech, most commonly during confrontation naming. For example, when presented with a picture and asked to name it silently in one's head, a participant would say 'yes' to having subjective inner speech if they could access the word and it sounded right (Fama, Henderson, et al., 2019). Silent picture naming has been related to phonological access skills, but not necessarily phonological working memory given that no manipulation is necessary (in contrast to the objective tasks, described above) (Fama, Snider, et al., 2019). As with all subjective tasks, whether in the moment or generalized, there is a lack of experimental control inherent to measuring internal, phenomenological experience (Lupyan et al., 2023).

The other way in which inner speech has been measured subjectively is by evaluating the function and uses of inner speech more generally. This is the method most commonly used to measure inner speech in neurotypical adults by employing questionnaires/surveys (Alderson-Day et al., 2018; Heavey et al., 2019; Racy et al., 2020), thought listing (Racy et al., 2020), interviews and other dynamic sampling methods (e.g., experience sampling methodology, Hurlburt et al., 2021). The goal of these methodologies is to evaluate how often inner speech occurs across some timeframe (e.g., the last week), for what reasons, and in what contexts. Research in non-brain-damaged adults suggests inner speech frequency ranges from 0-75% of sampled instances (Hurlburt et al., 2016); is used for a variety of functions, including problem-solving (Wallace et al., 2017), self-awareness (Morin, 2011b, 2011a) and emotion regulation (Morin & Michaud, 2007), and varies by context (e.g., may be more present during cognitively difficult scenarios) (Morin et al., 2018; Racy et al., 2020). A benefit to this way of measuring inner speech is these methods have high face validity for how inner speech is perceived in everyday life.

The choice to measure inner speech objectively and subjectively is experimenter-specific and based on the goals of the study, for example, whether the goal is to evaluate inner speech in the moment (using rhyme [objective] or inner naming [subjective]) or more generally (e.g., via survey). The tasks tell us slightly different information about inner speech because they theoretically invoke different cognitive processes. However, there is also, theoretically, considerable overlap in the cognitive processes employed by some objective and subjective tasks. For example, objective picture rhyme judgment and subjective inner naming are both thought to require phonological access via visual routes and have both been shown to implicate the dorsal stream of language, thereby reflecting some shared neural components (e.g., Fama, Snider, et al., 2019; Geva, Jones, et al., 2011).

To our knowledge, our study is the first to employ both objective and subjective means of measuring inner speech in aphasia. In the current study, we collect data in adults with and without aphasia for the three in-the-moment objective rhyming tasks described above (picture, written word, auditory word; Section 1.1 of Methods) as well as a subjective in-the-moment task (silent naming; Section 1.2 of Methods) and subjective Likert-style questions about inner speech experience more generally (Section 2 of Methods). Few studies have contrasted performance between neurotypical adults and individuals with aphasia across objective and subjective means of measuring inner speech, and thus we will examine this in Research Question 1. To flesh out the differences and similarities between tasks all meant to evaluate some facet of inner speech, we will examine the relationship between performance on objective tasks, on tasks evaluating inner speech in-the-moment, and between subjective tasks (Research Question 2). We will compare objective and subjective performance whenever possible.

An interesting phenomenon that has been observed in at least two studies using objective rhyming tasks to measure inner speech is that inner speech is relatively preserved compared to overt speech, even in individuals with more severe aphasia (Geva, Bennett, et al., 2011; Stark et al., 2017). Note, though, that the individuals with more severe aphasia, whilst having relatively preserved inner speech compared with their overt speech, still demonstrated worse performance on both inner and overt speech tasks compared to individuals with milder aphasia. There is a theoretical basis for explaining the impact of aphasia severity on inner speech performance, and it is related to the cognitive processes used during these objective inner speech tasks. Two of the objective tasks – picture and written rhyme decisions – are thought to engage the "inner voice" whilst auditory rhyme decisions are thought to engage the "inner ear" (Hubbard, 2010; Smith et al., 1995). The "inner ear" is understood to be a more passive process, involving just auditory information, with the "inner voice" being a more active process related to subvocalization, thus including pre-articulatory processes and, often, rehearsal. Another group makes a similar distinction between "inner speaking" and "inner hearing" (Hurlburt et al., 2013), emphasizing the difference between whether a person is truly saying something in their head or if they are simply hearing it, as if played back on a recording. There are likely different underlying processes involved in inner voice/speaking versus inner ear/hearing (Hubbard, 2010). The "inner voice" requires lexical access and retrieval and other cognitive processes, like phonological working memory, compared to the "inner ear," which does not. Because decision making based on inner voice tasks likely requires access to more cognitive-linguistic processes than decision making based on inner ear tasks, inner voice tasks might be more affected in people with more severe aphasia. Tree and Playfoot (2019), in an individual with conduction aphasia and impaired phonological short-term memory, and Langland-Hassan et al. (2015) in a large group of individuals with aphasia, both showed that inner voice tasks are typically more difficult for individuals with aphasia, though they did not explicitly look at aphasia severity as further moderating performance on inner voice tasks. Studies evaluating inner speech using subjective methods like silent naming have likewise demonstrated that individuals with varying aphasia severities demonstrate varying inner speech abilities (Fama, 2019; Fama et al., 2017; Fama, Snider, et al., 2019), though the impact of aphasia severity on subjective inner speech has had limited evaluation. Therefore, we empirically evaluate the relationship of aphasia severity with inner speech performance across objective and subjective tasks in Research Question 3.

Research Questions

Below is a summary of our three research questions, with hypotheses generated from the literature discussed above:

Research Question 1: *How do neurotypical controls differ from persons with aphasia regarding inner speech*? This is largely a confirmatory research question, as prior evidence has demonstrated that individuals with aphasia, overall, do worse than their neurotypical peers on objective measures of inner speech, like rhyming tasks. Therefore, we expect to confirm these results for objective inner speech tasks. However, comparatively less literature has evaluated differences between individuals with and without aphasia using subjective measurements of inner speech. Our hypothesis regarding group differences during subjective inner speech tasks is therefore exploratory, but we also anticipate finding that the aphasia group demonstrates different performance (likely, more variability) than the neurotypical group. Given that some of our subjective measurements don't have a "better" or "worse," as the objective measurements do, we cannot make a directional hypothesis in the same way. Therefore, this exploratory hypothesis about subjective inner speech is derived from the fact that inner speech is typically variable in neurotypical adults (e.g., Hurlburt et al., 2021; Roebuck & Lupyan, 2020) and inner speech in aphasia may be more variable (Fama, Snider, et al., 2019; Fama & Turkeltaub, 2020; Geva, Bennett, et al., 2011; Langland-Hassan et al., 2015; Stark et al., 2017).

Research Question 2: *To what extent does performance on each inner speech task relate to performance on other inner speech tasks within similar genres (i.e., objective, subjective)? Relatedly, to what extent do subjective measures and objective measures relate to one another?* Inner speech tasks tap into different cognitive processes but share underlying processes. We therefore expected to find positive correlations between inner speech tasks within the same genre (objective, subjective) and which share underlying cognitive processes (i.e., picture rhyme decision [objective] and silent naming of pictures [subjective]).

Research Question 3: *Is performance on inner speech tasks mediated by aphasia severity*? We expected individuals with aphasia to do worse, overall, on inner voice tasks (measured objectively using picture and written word rhyme judgments) compared to inner ear tasks (measured objectively using auditory word rhyme judgments). We also expected there to be an impact of severity, where individuals with most severe aphasia also demonstrate the lowest accuracy on inner voice tasks. For hypotheses regarding the impact of aphasia severity on subjective measurements of inner speech, there is sparse evidence, so these are largely exploratory. We hypothesized that individuals with more severe aphasia would show more impairments with silently naming objects (described in Section 1.2 of Methods, "Naming SIS"), and that aphasia severity would impact general interpretation of inner speech in daily life (described in Section 2 of Methods).

Method

Participants

Twenty-three participants with chronic post-stroke aphasia were recruited by author PLH for a related study (details can be found in Langland-Hassan et al. (2021)), which was approved by the University of Cincinnati's ethical review board (IRB #2012-4185). The original study collected the Western Aphasia Battery-Revised (WAB-R) (Kertesz, 2007) to establish aphasia severity and type. To ensure adequate comprehension of experimental tasks, the original study included participants only if their auditory-verbal comprehension score on the WAB-R was ≥4 (out of 10). To provide a general overview of cognitive functioning, the original study also had individuals with aphasia complete the non-linguistic subtests of the Cognitive Linguistic Quick Task (CLQT) (Helm, 2003), and the participants were subsequently classified as within normal limits, mildly impaired, moderately impaired, or severely impaired. If a participant obtained ratings of "within normal limits" and/or "mild" impairment on the non-linguistic subtests of the CLQT, this was taken as evidence of auditory and visual perception adequate to complete the experimental tasks and participants were included in the study.

N = 20 individuals with aphasia were included in the following analyses, excluding three of the original 23 individuals with aphasia due to Attention and Visuospatial Skill scores on the CLQT indicating more than mild impairment. They were aged 35-72 years (M=56.4, SD=9.09), and all had chronic aphasia (1.5-31 years with aphasia, with a mean of 8.85±8.09). Four types of aphasia were represented: anomic (n=7), Broca's (n=6), conduction (n=6), and transcortical motor (n=1). Only one participant was classified as having severe aphasia according to the WAB-R Aphasia Quotient (AQ) (an AQ score of <50), while all others were classified as mild-moderate (M=73.1±13.9). The average level of education was approximately equivalent to a bachelor's degree (M=15.9±1.79 years).

A sample of prospectively matched non-brain damaged individuals served as a control group. This group was at least 18 years of age, had no history of language impairments, had no history of major psychotic or neurologic episodes or intractable substance abuse, had at least a high school education, and were native speakers of American English. Twenty adults were included (Table 1).

Between the aphasia and control groups, there were no significant differences in age, gender/sex, or education. Table 1 includes demographic information and statistical comparisons.

[TABLE 1 HERE]

Materials and Procedure

An inner speech battery, described below, was presented on a 21" touchscreen using a JavaScript application stored on a secure external web server. There are two major sections to the battery. The first section evaluates "in-the-moment inner speech," which is further divided into subjective and objective methods, whilst the second section evaluates how participants generally, subjectively interpret their inner speech.

Section 1: In-the-moment inner speech

Section 1 was further subdivided into Objective inner speech tasks (S1.1) and Subjective inner speech tasks (S1.2).

Section 1.1: Objective tasks. The objective portion of the inner speech battery (see Figure 1 for a visual representation) included the following three subtests:

1. Picture Rhyme Judgement, in which the participants saw two pictures of objects and were asked if the words for them rhyme (as in Langland-Hassan et al., 2015). This task requires access to semantic, lexical and phonological information via visual routes, as well as phonological working memory. It is commonly referred to as an "inner voice" task;
2. Written Rhyme Judgement, in which the participants saw two written words, some of which varied in orthography (e.g., Bear / Chair), and were asked if they rhyme. This task requires access to orthographic and phonological information as well as phonological working memory. It is commonly referred to as an "inner voice" task;
3. Auditory Rhyme Judgement, in which the participants heard two words and were asked if they rhyme. Participants could hear the words an unlimited number of times. This task requires access to phonological representations and phonological working memory. It is commonly referred to as an "inner ear" task.

The materials for the picture rhyming task were adapted from the materials of Geva and Warburton (2019). However, the sets used were slightly different from the published version of Geva and Warburton (2019)'s materials (those materials were normed on British English-speaking adults) and so cannot be considered strictly equivalent.

The words used in each of the three inner rhyming tasks did not differ significantly in terms of their frequency, familiarity, or imageability, based on t-tests ($p > 0.3$ for all comparisons) of ratings from the online MRC Psycholinguistic Database (Coltheart, 1981). For all tasks, participants were instructed to give a yes/no response via button press on the touchscreen and to complete the task in silence and without moving their mouth or tongue. A research assistant monitored performance. There were 38 trials involving picture rhyming, 40 trials for written word rhyming, and 40 trials for auditory rhyme judgements with no time limits for each trial and no instructions about speed of answering.

[FIGURE 1 HERE]

Section 1.2: Subjective task: Naming Successful Inner Speech (SIS). Participants also completed subjective reports of inner speech during a silent naming task in which they were shown one picture at a time and were asked to indicate whether they were able to say the word for the object silently in their head and have it sound right ("naming SIS"). Again, they responded with a yes/no button press to the prompt. The subjective report of single word inner speech requires only access to phonological information but not manipulation of phonological information.

In order to compare inner speech to overt speech ability, participants completed overt (out loud) naming of the same items. They were shown one picture at a time and asked to name it out loud. The experimenter recorded their response in English orthography.

Section 2: General, subjective interpretation of inner speech

This section included four subjective rating questions, which were presented on the screen and read aloud by the experimenter. In contrast to the in-the-moment objective and subjective tasks above, these ratings were about participant's experience in general, not just during the task. The subjective rating questions used a categorical visual rating scale (continuous arrow with numbers and descriptions) from 1-5 with verbal descriptions and are as follows:

1. <u>Idea without Word (IwW)</u>, "How often do you see something and know what it is but can't say the word for it out loud?" with 1 being almost never and 5 being almost

always;

2. Post-Stroke, given to aphasia group only, "Do you think you talk to yourself in your head less now, more now, or about the same as before your stroke?" with 1 being much less now and 5 being much more now;
3. Successful Inner Speech (SIS), "A few minutes ago we talked about how we sometimes see something and know what it is but can't say the word for it out loud. When that happens to you, how often can you say the right word for it in your head, and have it sound right?" with 1 being almost never and 5 being almost always.
4. Inner vs. Overt Speech, "In general, how do you think your abilities to speak in your head compare to your abilities to speak out loud?" with 1 being much better in head and 5 being much better out loud.

Two of the subjective rating questions (IwW and SIS) were adapted from Fama's work (e.g., Fama et al., 2019, Fama & Turkeltaub, 2020). As noted above, the subjective responses were on a 5-point scale, which changed based on the question, and were about general abilities rather than subjective reports about specific single words. Therefore, they served as a complement to the subjective single word inner speech data also collected. The questions were read aloud and presented visually. Participants selected their answer on a touchscreen. individuals with aphasia answered all four subjective questions, while the control group answered all but the post-stroke question.

To summarize, we distinguish the inner speech tasks in two ways: whether the method is objective or subjective, and whether the measurement is happening "in-the-moment" of inner speech or is a general interpretation of the overall inner speech experience.

Experimental Procedure

Participants with aphasia attended two sessions: a screening session in which they completed neuropsychological assessments (e.g., WAB-R), and a second session in which they completed the inner speech tasks and a categorization task relevant to a larger abstract thought study (Langland-Hassan et al., 2021). Sessions were scheduled one to three weeks apart, depending on participant availability. Neurotypical controls attended only one session and completed the inner speech tasks and categorization task in that session.

Regarding the inner speech task ordering, the subjective ratings described in Section 2 were always presented to participants first, and the order of the subjective questions was not randomized (they were presented in the order described above). Naming SIS (Section 1.2) came after the general subjective ratings from Section 2. Next, participants completed the rhyme judgement tasks (Section 1.1). The questions from Section 1.1 were presented in pseudo-randomized order by participant, alternating among six different orderings of the tasks. Overt naming (also described in Section 1.2) was always the last task.

Steps were taken to ensure that individuals with aphasia understood the questions. In all cases, the question prompts were both displayed visually and slowly read aloud by the experimenter. Participants had unlimited time to consider and answer the prompts. The important concept of "knowing what something is without knowing the word for it," was explained in a training trial that preceded the experimental trials. That training trial showed a picture of a watch and read, in part, "For example, a person might look at the object pictured below and know that it is something used for keeping track of the time, and that you wear on your wrist, but not be able to remember that it is called a **watch**. This would be a case of knowing what something is without remembering the word for it." Additionally, we want to highlight that the mean Auditory Verbal Comprehension score from the WAB-R was 8.9 out of 10, indicating comprehension suitable for understanding the questions.

Data Analysis

All data analysis was completed in R version 4.1.1 with RStudio Desktop 2022.07.1+554. The data analysis was designed to examine the research questions, below.

Research Question 1: Neurotypical Controls compared to Participants with Aphasia

To ensure that the objective inner speech tasks were valid (i.e., that controls performed near ceiling / with high accuracy on all tasks) and to replicate findings from Geva and Warburton (2019), we computed descriptive statistics for the objective inner speech tasks. We also statistically compared the neurotypical controls to the individuals with aphasia on the inner speech tasks that both groups completed (i.e., excluding the subjective post-stroke question because only the stroke group completed this question). Note that two members of the control group were missing data for the general, subjective rating about SIS, and the missing values were ignored in analysis. For continuous metrics, we used either independent t-tests or Wilcoxon rank sum tests dependent on data normality, and for categorical metrics (the subjective questions), we used Wilcoxon rank sum tests.

Research Question 2: Inner Speech Task Relationship with one another

We estimated the magnitude of the relationship between performance on each of the Section 1.1 (objective "in-the-moment" inner speech) tasks using Spearman correlations. Additionally, we investigated the relationship between tasks from Section 1.2 (subjective inner speech "in-the-moment" and overt naming) using the same methods. We used Spearman's correlation to evaluate the relationship between select Section 1.1 and Section 1.2 tasks, specifically the successful inner speech in-the-moment (Section 1.2) and the objective picture rhyming task (Section 1.1). We chose these two tasks because they both require inner speech achieved through picture-based visual semantic access, and thus should partially share cognitive mechanisms. Finally, we used Spearman correlations to explore the relationship between Section 2 tasks (general, subjective interpretation of inner speech using Likert scales).

For Research Question 2, we only evaluated the aphasia data, given that the data from neurotypical controls approached ceiling for most Section 1.1 and 1.2 measures.

Research Question 3: Relationship between Aphasia Severity and Inner Speech

We first computed Spearman correlations between severity, as measured by the WAB-R AQ, and each of the inner speech in-the-moment tasks (objective, subjective). In addition to computing correlations with this continuous metric of aphasia severity, we separated the aphasia group into two severity groups based on *a priori* severity groups identified by the WAB-R: a "mild" group (AQ scores of >= 76, but note that one person included in this group scored a 94, which could be considered "latent", i.e., non-clinical aphasia) and "moderate" (AQ scores of 51-75, but note that we included one person who scored 47.8 in this group). There were n=9 in the mild group and n=11 in the moderate group. Mild and moderate aphasia are commonly used distinctions from the WAB-R, relevant to clinical practice. We then analysed the extent to which inner speech scores differed by severity group using the Wilcoxon rank sum test with continuity correction, given that the data did not satisfy normality assumptions (Meek et al., 2007). To analyse aphasia severity impact on Section 2 tasks (subjective, general interpretation of inner speech), we computed the Wilcoxon rank sum test with continuity correction for each question to compare aphasia severity groups.

We had originally planned to also evaluate the impact of aphasia type (fluent vs. non-fluent) on inner speech, but we ultimately decided that due to the scoring criteria on the WAB-R (in which severity/AQ has a large bearing on type of aphasia), evaluating by type would not add sufficient additional information to our study and would be conflated with aphasia severity.

Results

Research Question 1: Descriptive statistics and group comparisons

Section 1 Tasks: In-the-moment inner speech

Controls' data on Section 1 assessments confirmed that the tasks were valid (i.e., performance was high/ at ceiling for all tasks) (see Table 2). To assess the differences between individuals with aphasia and controls on all Section 1 in-the-moment inner speech tasks, Wilcoxon rank sum tests were conducted. As expected, the control group significantly outperformed the individuals with aphasia group on all Section 1 tasks except for Naming SIS (Section 1.2), as both groups were near ceiling.

[TABLE 2 HERE]

Section 2 Tasks: General, subjective interpretation of inner speech

We conducted Wilcoxon rank sum tests to assess the differences between individuals with aphasia and controls on the general, subjective questions. For the subjective questions that asked more generally about inner speech, we saw similar answers for the controls and individuals with aphasia (IwW: $W = 236.5$, p-value = 0.31, SIS: $W = 165$, p-value = 0.89), excepting that individuals with aphasia more commonly reported that their inner speech was better or much better than their speech out loud, while the controls sometimes reported better speech out loud than in their heads ($W = 74$, p-value = 0.0003). Figure 2 represents the subjective answers from the individuals with aphasia group and the control group, visually demonstrating that individuals with aphasia and control groups only differ on their answers of whether they speak better in their head vs. out loud. Additionally, the individuals with aphasia and control groups' answers vary considerably within groups.

[FIGURE 2 HERE]

Research Question 2: Inner speech task relationship with one another

Figure 3 shows a comprehensive representation of Spearman correlations between general subjective inner speech questions (Section 2) and inner speech in-the-moment tasks (Section 1) as well as the demographic variables of age, years with aphasia, education (in years), and aphasia severity (WAB-R AQ).

[FIGURE 3 HERE]

Section 1 Tasks: In-the-moment inner speech

As hypothesized, the objective rhyme tasks from Section 1.1 were significantly correlated with one another (picture and written, r_s=0.82, p<.001; picture and auditory, r_s=0.64, p=0.004; written and auditory, r_s=0.71, p=0.0004). Overt naming was also related to the two objective rhyming judgements which required the "inner voice," picture (r_s=0.61, p=0.004) and written (r_s=0.59, p=0.007), but not the "inner ear" task, auditory (r_s=0.29, p=0.22).

We only had one in-the-moment, subjective inner speech task (Section 1.2), Naming SIS. Therefore, we could not truly compare this to the more general subjective tasks from Section 2. However, we did evaluate the extent to which silent naming (Naming SIS) was related to overt naming in the aphasia group and found a not-significant relationship (r_s = -0.17, p = 0.47, Figure 4a). This finding shows that overt naming abilities were not related to reports of silent naming in individuals with aphasia, with few reporting impoverished silent naming.

We evaluated the relationship between an objective in-the-moment method (picture rhyme judgment) and a subjective in-the-moment method (Naming SIS) that we hypothesized involved substantially overlapping cognitive processes, namely picture-based visual access to lexical and phonological information. Using Spearman's correlations, we identified a non-significant relationship between the picture rhyme task and Naming SIS ($r_s = 0.14$, $p = 0.55$, Figure 4b).

Section 2 Tasks: General, subjective interpretation of inner speech

We found only a single significant correlation between subjective, general tasks, which was that IwW was significantly related to Post-Stroke ($r_s=0.51$, $p=0.02$). This suggested that people with aphasia who tended to have the idea without word (IwW) also tended to experience more inner speech post-stroke.

[FIGURE 4 HERE]

Research Question 3: Evaluating the impact of aphasia severity on inner speech

Section 1 Tasks: in-the-moment inner speech

We first compared performance on the objective rhyme judgement tasks using Wilcoxon rank sum tests, because we hypothesized that individuals with aphasia would do worse on inner voice compared with inner ear tasks. As expected, individuals with aphasia performed better on the auditory task than the written and picture tasks (Table 2 for means), and there was no difference in performance between the written and picture tasks (auditory vs picture: $W = 279.5$, p-value = 0.03; auditory vs written: $W = 286$, p-value = 0.02; picture vs written: $W = 202.5$, p-value = 0.96). It was also of interest how individuals with aphasia did when we parsed apart the residuals from each inner speech task, i.e., evaluated performance on picture rhyming when considering variance from written and auditory rhyming performances, and so on. For example, given that the picture rhyming task is thought to involve different cognitive processes, with some overlapping with the cognitive processes required in the written and auditory rhyming tasks, considering variance from the other objective tasks enabled us to evaluate performance uniquely per task. When accounting for variance from each of the other objective tasks, a one-way ANOVA showed that individuals with aphasia did not perform significantly differently on any task (three paired t-tests, all $p>0.99$). This suggests that, when evaluating each objective task as comprised of unique cognitive components, individuals with aphasia performed similarly on auditory, written, and picture inner speech tasks. The similar performance when accounting for variance suggested that the poor performance on picture and written rhyme tasks by the individuals with aphasia group as a whole (no regression of other tasks) may be due to the specific cognitive processes engaged by these tasks.

Next, we explicitly assessed the role of aphasia severity. To assess how the continuous variable of aphasia severity was related to the objective inner speech tasks, we conducted Spearman's correlations. People with milder aphasia (based on the WAB-R AQ) tended to perform better on the objective inner speech tasks, especially picture ($r_s= .31$, $p = .01$) and written ($r_s = .23$, $p = .03$) tasks, but aphasia severity was not related to performance on the auditory task ($r_s = .17$, $p = .06$). As WAB-R AQ increases (indicating milder aphasia severity), performance on the picture and written tasks also increases, but more of the participants at all aphasia severity levels are at or near ceiling for the auditory task (Figure 5).

[FIGURE 5 HERE]

When we grouped the individuals with aphasia into mild and moderate aphasia severity groups based on their WAB-R AQ, we found similar results using the Wilcoxon rank sum tests, with the mild group performing more accurately than the moderate group on the picture ($W = 89$, $p = 0.003$) and written ($W = 77$, $p = 0.04$) inner speech rhyme judgement tasks, but not on the auditory rhyme judgement task ($W = 71$, $p = 0.11$). As seen in Figure 6, the severity distinction appears related to inner speech abilities, as the mild group outperformed the moderate group on the tasks which required inner “voice” (picture, written), but not inner “ear” (auditory).

When evaluating subjective in-the-moment inner speech (Naming SIS), we did not find a significant association with aphasia severity, as measured by AQ ($r_s = 0.02$, $p = 0.60$) or using severity groups ($W = 47.5$, $p = 0.90$).

[FIGURE 6 HERE]

Section 2 Tasks: General, subjective interpretation of inner speech

Aphasia severity groups answered similarly on all the general subjective questions about inner speech (Section 2) based on the Wilcoxon rank sum test with continuity correction. Individuals from the different severity groups reported similar frequency of successful inner speech ($W = 48.5$, $p = 0.97$) and idea without word ($W = 37.5$, $p = 0.36$). Both participant groups reported using inner speech more often now or at the same frequency as they did before their stroke, with no significant differences between the groups ($W = 38$, $p = 0.36$). Both groups reported equal or better inner speech than overt speech, with no significant differences between the groups ($W = 33.5$, $p = 0.20$). As seen in Figure 7, subjective ratings were similar between the two severity groups, with intra- and inter-group variability.

[FIGURE 7 HERE]

Cases examining individual differences in inner speech performance

Looking at individual participants illuminates the differences in abilities and reports of inner speech. Hence, we wanted to take time in the Results to discuss specific examples. For example, in participants with moderate aphasia, there are varying strengths and weaknesses on tasks requiring lexical-semantic access with pictures (Table 3). Three of these participants (1003, 1008, and 1011) performed similarly on picture rhyme judgement but differed on the other tasks, while the fourth participant (1016) performed better on picture rhyme judgement, allowing for another contrast. First consider participant 1008, who made more phonemic naming errors (i.e., similar speech sounds, e.g., “bowl” for “bow”) rather than semantic errors (i.e., similar meaning, e.g., “bowl” for “cup”). Other errors were either unrelated to the target or no response. Participant 1008 endorsed having 100% naming SIS, but overt naming score for the same items was 46%. They also scored poorly on the objective picture task (50%). On the rating scales for general subjective inner speech questions (Section 2), participant 1008 marked a “5” for how often they experience SIS, suggesting agreement between Naming SIS and general SIS experiences. Participant 1011 scored 50% on picture rhyme judgement, 13.5% on overt naming (with more semantic than phonemic errors), but 100% on reports of successful inner speech for the same pictures and rated the frequency of successful inner speech as 5 (almost always). In contrast, participant 1003, who also has moderate aphasia, scored 55% on picture rhyme judgement but 92% on overt naming (with no phonemic errors and some semantic) and 95% on reports of successful inner speech and rated the frequency of successful inner speech as lower (3, sometimes). It is possible that participants have difficulty with different cognitive processes required for picture rhyme judgement (semantic and phonological access, phonological awareness, working memory) even though they performed similarly on that task (~50%), and aphasia severity does not completely explain performance on these tasks. Those differences appear in overt naming scores and their perceptions of inner speech.

Another contrast is from participant 1016, who performed well on the picture rhyme judgement task (82%) but poorly on overt naming (54%, with more semantic than phonemic errors), while rating the frequency of SIS in cases of anomia out loud as 2 (rarely). We postulate that participant 1016’s primary impairments are in articulatory and motor components, but if they cannot find a word, it is because they usually do not have the semantic or phonetic access.

Participant 1008 may have a different issue because they make more phonemic errors, with difficulty in phonetic access which makes both the picture rhyming and overt naming challenging.

[TABLE 3 HERE]

Discussion

Inner speech is difficult to measure by nature, as it occurs entirely in one's head. Nevertheless, it has been measured using both subjective and objective paradigms, in-the-moment and more generally, with individuals demonstrating its validity by showing that perception of inner speech relates to expected brain areas (Fama et al., 2017). Understanding how inner speech relates to severity when it is measured differently has downstream implications for more comprehensive study of inner speech in aphasia. It is important to evaluate the relationship of aphasia severity and inner speech, as aphasia severity is related to quality of life and participation (Williamson et al., 2011), and inner speech is likely related to quality-of-life factors, such as self-awareness (Morin, 2009, 2011; Morin & Michaud, 2007) and psychological distress (Heavey & Hurlburt, 2008). By evaluating how aphasia severity is related to inner speech, we can gain insight into the mechanisms behind changes in self-awareness and psychological health with aphasia. With the many methods of measuring inner speech that have been utilized in prior research, it is important to reveal what information we can gain from each method and how the methods complement each other.

How do neurotypical controls differ from persons with aphasia on inner speech tasks?

Regarding objective tasks, we expected and found that individuals with aphasia did worse on objective measures of inner speech than individuals without aphasia. This makes sense, given that inner speech measured using silent rhyme judgement has been closely tied to phonological awareness and manipulation and most individuals with aphasia have impairments with these language functions (Stark et al., 2017; Fama et al., 2019). Regarding group differences for subjective inner speech tasks, our exploratory hypothesis was that the aphasia group would demonstrate different performance (likely, more variability) than the neurotypical group. For the subjective, general inner speech questions, we found that both the aphasia and the control group had high within-group variability with no significant group differences. This suggests that people have a range of experiences of inner speech, regardless of history of brain injury, which has been shown before in neurotypical adults (Alderson-Day et al., 2018; Racy et al., 2020). We likewise did not identify a significant difference on the in-the-moment subjective measure of inner speech, Naming SIS, between the two groups. In the case of Naming SIS, both groups reported accuracy near ceiling (see Table 2 for means). Therefore, our hypothesis regarding wider variability for the aphasia group on subjective questions did not come to fruition. A null finding related to group differences during subjective tasks may be because the subjective tasks did not employ the phonological working memory component that the objective rhyming tasks did, resulting in a higher performance from the aphasia group.

Inner speech measured across tasks

We first expected to find positive correlations between inner speech tasks within the same genre (objective, subjective) and modality (in-the-moment) for the aphasia group. We found this to be the case for the objective rhyming tasks, which showed positive correlations with one another. We also found that, within the more general subjective questions, IwW (idea without word) was significantly, positively related to using their inner speech post-stroke. It may have been the case that, more generally, people are thinking more about concepts related to speaking (i.e., IwW) after the stroke.

We also expected to find positive correlations between tasks that shared underlying cognitive processes despite being collected using different methods (i.e., picture rhyme decision [objective] and silent naming of pictures [subjective]). Surprisingly, we did not find a significant relationship between picture rhyme judgment and silent picture naming, which suggests that, even whilst at face value these tasks appear to substantially overlap in cognitive demands, this may not be the case in practice. Indeed, we found that individuals with aphasia had a much higher rate of Naming SIS (all > 80%, refer to Table 2) than they did accuracy on picture rhyming (mean of 64%, refer to Table 2), suggesting that this null relationship may be due to the difficulty of the picture rhyming task. This difficulty could be the result of the added cognitive process of phonological working memory, the fact that they were asked to come up with two names (e.g., Bear vs. Chair) versus one name (Naming SIS), or inflated reporting of accuracy in Naming SIS.

We did not identify a significant relationship between Naming SIS and overt naming performance in the aphasia group. At face value, this seems strange, but this finding jives with prior literature (e.g., Geva et al., 2011). To be related in a positive manner, this would entail that everyone with very good naming also has very good inner speech (not rare, and we do find this) but also that everyone with poor naming has poor inner speech (rare). Our sample did not report poor inner speech during naming, as you'll see in Figure 4A. That is, the median for Naming SIS was 97% (range ~84-100%) whilst overt naming had a median of 86.5% (range ~13.5 – 100%). The markedly preserved inner speech alongside relatively impoverished overt naming has been shown in a variety of papers evaluating inner speech in aphasia (e.g., Stark et al., 2017; Geva et al., 2011), where the majority of participants had more preserved inner speech compared to overt speech. Thus, we suspect that the non-significant relationship between Naming SIS and overt naming is driven by a lack of people reporting poor inner speech.

Is performance on inner speech tasks mediated by aphasia severity?

We expected individuals with aphasia to do worse, overall, on inner voice tasks (measured objectively using picture and written word rhyme judgments) compared to inner ear tasks (measured objectively using auditory word rhyme judgments). Confirming our hypothesis, we found that individuals with aphasia performed better on the auditory rhyme task (inner ear) than the written or picture rhyme tasks (inner voice tasks). We also expected there to be an impact of severity, where more severe individuals would show the worst performance on inner voice tasks, in particular. Individuals with milder aphasia performed better on the objective inner speech tasks involving the inner voice (picture, written). We did not identify a significant relationship between severity and performance on the auditory rhyme task, likely because the performance was near ceiling for most participants. These findings confirm prior research (Fama, Henderson, et al., 2019; Langland-Hassan et al., 2015; Tree & Playfoot, 2019) and, coupled with complementary findings from other labs, lead us to conclude that objective methods that require several cognitive-linguistic steps for accessing inner speech (e.g., picture rhyme, written rhyme) are more liable to be testing the language production-like component of inner voice and more likely to be related to aphasia severity.

Our exploratory hypothesis was that individuals with more severe aphasia would show more impairments with silently naming objects and that aphasia severity would impact general interpretation of inner speech in daily life. In contrast to what we identified for the inner voice tasks, above, we did not find that the subjective task of Naming SIS was related to aphasia severity for our sample, even though it is also an in-the-moment inner speech task. One likely explanation is that this is due to the difference in cognitive resources required for access alone

(Naming SIS task) vs. access and manipulation (inner voice objective inner speech tasks). That is, aphasia severity may become an important factor when internal manipulation is required. Note that overt naming was related to aphasia severity (as shown in Figure 3), even though overt naming does not require manipulation. This correlation makes sense given that the aphasia quotient from the WAB-R is partially derived from performance on word-finding and object naming. Further, overt naming requires post-phonological processes (articulation, motor) that inner speech does not, and these may be particularly related to the way that the WAB-R quantifies aphasia severity. Alternatively, Naming SIS reports may have been inflated if participants were inaccurate with reporting inner speech.

We likewise explored the extent to which aphasia severity mediated the more general subjective inner speech tasks. We found that aphasia severity groups did not demonstrate significantly different responses for any of the subjective questions about how they generally experience inner speech. There was high within-group variability for most subjective questions, and we did not find that ratings were related to severity. We identified some trends suggesting that severity of aphasia may play a role in general subjective perception of inner speech if evaluated in a larger sample: for example, participants with moderate aphasia reported using inner speech more often now than they did before their stroke, while participants with milder aphasia tended to report having the same frequency of inner speech use now as they did prior to their stroke. This could be because people with moderate aphasia are intentionally using inner speech more often due to their more impaired overt speech. Both groups reported equal or better inner speech than overt speech. This finding strongly encourages further investigation into the nature and uses of inner speech across a diverse sampling of individuals with aphasia.

Figure 7 is an example of heterogeneity on the general subjective perceptions about inner speech in the mild and moderate aphasia groups. When asked about having an Idea Without Word ("IwW"), there was wide variation within the moderate aphasia group (all scale ratings used), whereas the mild group used only three of five of the scale ratings. This suggests greater homogeneity of experience in the mild group as it relates to Idea Without Word. For Successful Inner Speech ("SIS") frequency, both the mild and moderate groups showed high intra-group heterogeneity, using all scale options (moderate) or four out of five scale options (mild). Both the mild and moderate groups trended in feeling that their inner speech was better than their overt speech, demonstrating intra-group homogeneity. For the frequency of inner speech pre-stroke, the mild group tended to be most homogeneous, with nearly all members selecting option three ("about the same as before my stroke") whereas the moderate group tended to show higher intra-group heterogeneity but with many participants choosing a higher scale option ("experience inner speech more now"). While these subjective scales did not significantly relate to aphasia severity or demonstrate significant differences between participant groups (control, individuals with aphasia), they serve to clearly illustrate individual differences in the experience of inner speech. Some of these individual differences were likely present pre-stroke, because as we saw, the control group also had heterogeneity in their answers (Figure 2). Individual differences in inner speech have also been shown in individuals with aphasia in past literature (Fama, Henderson, et al., 2019; Geva, Bennett, et al., 2011; Langland-Hassan et al., 2015; Stark et al., 2017), and often in individuals without aphasia (e.g., Hurlburt et al., 2013).

Contextualizing the differences between inner speech tasks

We want to conclude this Discussion by mapping back how these tasks relate to a multidimensional framework used to understand inner speech, ConDialInt (Grandchamp et al., 2019). If the relationships between aphasia severity and the ways of measuring inner speech are dissimilar, it is possible that they are measuring different facets or experiences of inner speech, as described in the ConDialInt model (Grandchamp et al., 2019). Picture and written rhyme judgements likely use expanded, monologue, and intentional inner speech, based on the ConDialInt model. These rhyme judgements require phonetic information, are not conversational, and are purposeful, and the auditory task may share these qualities, although it may be less "intentional." While the general subjective rating questions were designed to test perception of inner speech, it is possible that participants were thinking beyond inner speech used for isolated tasks (e.g., rhyming, naming/word-finding) and more about their experience of an inner narrative, which may be more condensed and less intentional but likely still monologue. Inner speech throughout daily life, measured by some of our subjective general questions, relates to work by Morin and colleagues on the relationship between inner narrative and extra-linguistic processes such as self-regulation and problem-solving (Morin, 2011a; Morin & Michaud, 2007), which are likely less related to aphasia severity. The differences in the condensation and intentionality dimensions may contribute to the results which indicate that there are not strong relationships between the subjective, general methods of measuring inner speech and how they relate to aphasia severity. For example, condensed inner speech would not require total phonetic access, so participants with difficulty accessing the phonemes would not necessarily report worse inner speech more generally.

Conclusion and Clinical Implications

Aphasia severity was related to in-the-moment objective inner speech tasks that involve the inner voice, including picture and written rhyme judgments, but not to other in-the-moment tasks, including subjective inner speech during naming and objective auditory rhyme judgment. Aphasia severity was also not related to subjective interpretation of the general frequency/use of inner speech. We therefore suggest that the difference may have been due to the cognitive-linguistic processes involved and alternative interpretations of the questions on the subjective rating scales.

Individuals with aphasia may retain inner speech despite impoverished lexical or phonological abilities (Fama et al., 2017; Fama, 2019; Fama, Snider, et al., 2019; Fama & Turkeltaub, 2020; Geva, Bennett, et al., 2011; Stark et al., 2017). It has long been understood that inner speech plays a role in aphasia, e.g., Brumfitt (1993) states that one must understand that the 'impact on the aphasic person's experience may be the lost ability to talk to oneself'. Moreover, in brain damaged patients who eventually recover from their trauma, self-awareness often returns in parallel with inner speech (Ojemann et al., 1996). The importance of inner speech in aphasia recovery is emphasized by Dr. Jill Bolte Taylor in her book, *My Stroke of Insight*, in which she describes her experience recovering from a left hemispheric stroke (Bolte Taylor, 2006). Dr. Taylor recalls a total lack of inner speech: "The dramatic silence that had taken residency inside my head" (pp. 75–76) and emphasizes that this silence greatly impacted her self-awareness and sense of self (p. 67).

Various methods of measuring inner speech provide meaningful information about the experience of inner speech in both neurotypical controls and people with aphasia. These methods are likely capturing unique facets of inner speech. Because there is the possibility of relatively strong inner speech abilities in comparison to overt speech in individuals with aphasia (Geva, Bennett, et al., 2011; Stark et al., 2017; Fama et al., 2019; Fama & Turkeltaub, 2020), characterizing inner speech in aphasia has clinical implications. It is important, however, to consider the specific facets of inner speech that are relevant to the clinical situation. For example, if a clinician is interested in the underlying naming abilities of their client, they may be more likely to use in-the-moment methods of measuring inner speech to assess the accuracy of access to the word (Fama, Snider, et al., 2019). The clinician could then gain valuable information about where their client was having difficulty in the process of speech production (i.e., difficulty with access alone vs. difficulty with access and manipulation of language information). If a clinician wanted to know more about the frequency of inner speech in daily life, they may use subjective rating scales and adapt them to their desired questions. Finally, if the clinician were interested in the contents and functions of their client's inner speech, they could opt for open-ended subjective questions or adapt pre-existing surveys (Racy et al., 2020) for individuals with aphasia (our lab is presently working on this). As always, clinicians and researchers must carefully select their methods of measuring a phenomenon based on the desired outcome.

Limitations

Not all aphasia severity levels were represented in the aphasia group, so relationships with severity may be stronger or appear differently if people with more severe aphasia as well as more mild aphasia (e.g., latent aphasia) participate in future studies. At a case study level, we explored the relationship between phonemic and semantic errors in overt naming and inner speech abilities, but this may benefit from a more detailed analysis with a larger set of naming data.

While we did match the stimuli in the objective tasks for frequency, familiarity, and imageability, we did not calculate Levenshtein's distance, which may have informed us further on task difficulty.

Some of the subjective questions about general inner speech may not have been answered in the same way by the two groups, as some members of the control group answered that they sometimes or often cannot say a word out loud but have successful inner speech (SIS), meaning that they had access to the word and its phonemes in their head. This seems unlikely for people who do not have a speech or language disorder, except possibly if someone were extremely exhausted or recently had oral surgery. Because the experience of anomia is likely uncommon for the control group, they may have assumed that the question was addressing a different experience, such as having partial phonetic access (tip-of-the-tongue phenomenon).

When using subjective reports of an internal phenomenon, there is always the possibility that the reports are inaccurate. We recognize that self-report may be harder to interpret in our population with aphasia, which may affect comprehension and self-monitoring/error detection, as discussed in Fama and Turkeltaub (2020). Readers can be encouraged that with similar methods, subjective reports of inner speech were meaningful in individuals with aphasia (Fama, Snider, et al., 2019). We acknowledge that some of the subjective questions may also have had difficult wording, making them linguistically difficult for individuals with aphasia to interpret.

Future Directions

Future work should focus not only on inner speech in the moment for isolated tasks (e.g., rhyming, naming), but as a part of the daily experience of people with aphasia. The focus may also shift to include more clinical applications of relatively strong inner speech compared to overt speech, such as how successful inner speech may predict therapy outcomes. We are presently evaluating both of these directions in our lab as part of a National Institute on Disability, Independent Living and Rehabilitation Research (NIDILRR)-funded grant to Stark.

Data Availability Statement

The data that support the findings of this study are openly available in The Open Science Framework at https://osf.io/8sjny/, reference number DOI 10.17605/OSF.IO/8SJNY.

Acknowledgments

We thank Aimee Dietz, Frank R. Faries, and Maxwell Gatyas for their support in facilitating and conducting some of the experimental sessions reported here.

References

Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The varieties of inner speech questionnaire – Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition*, *65*, 48–58. https://doi.org/10.1016/j.concog.2018.07.001

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update. *Cortex*, *112*, 91–106. https://doi.org/10.1016/j.cortex.2018.05.015

Bolte Taylor, J. (2006). *My stroke of insight: A brain scientist's personal journey*.

Brumfitt, S. (1993). Losing your sense of self: What aphasia can do. *Aphasiology*, *7*(6), 569–575. https://doi.org/10.1080/02687039308248631

Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505. https://doi.org/10.1080/14640748108400805

Fama, M. E. (2019). Self-reported inner speech relates to phonological retrieval ability in people with aphasia. *Consciousness and Cognition*, 12.

Fama, M. E., Hayward, W., Snider, S. F., Friedman, R. B., & Turkeltaub, P. E. (2017). Subjective experience of inner speech in aphasia: Preliminary behavioral relationships and neural correlates. *Brain and Language*, *164*, 32–42. https://doi.org/10.1016/j.bandl.2016.09.009

Fama, M. E., Henderson, M. P., Snider, S. F., Hayward, W., Friedman, R. B., & Turkeltaub, P. E. (2019). Self-reported inner speech relates to phonological retrieval ability in people with aphasia. *Consciousness and Cognition*, *71*, 18–29. https://doi.org/10.1016/j.concog.2019.03.005

Fama, M. E., Snider, S. F., Henderson, M. P., Hayward, W., Friedman, R. B., & Turkeltaub, P. E. (2019). The Subjective Experience of Inner Speech in Aphasia Is a Meaningful Reflection of Lexical Retrieval. *Journal of Speech, Language, and Hearing Research*, *62*(1), 106–122. https://doi.org/10.1044/2018_JSLHR-L-18-0222

Fama, M. E., & Turkeltaub, P. E. (2020). Inner Speech in Aphasia: Current Evidence, Clinical Implications, and Future Directions. *American Journal of Speech-Language Pathology*, *29*(1), 560–573. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0212

Geva, S., Bennett, S., Warburton, E. a., & Patterson, K. (2011). Discrepancy between inner and overt speech: Implications for post-stroke aphasia and normal language processing. *Aphasiology*, *25*(3), 323–343. https://doi.org/10.1080/02687038.2010.511236

Geva, S., & Fernyhough, C. (2019). A penny for your thoughts: Children's inner speech and its neuro-development. *Frontiers in Psychology*, *10*(JULY), 1708. https://doi.org/10.3389/fpsyg.2019.01708

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain*, *134*(10), 3071–3082. https://doi.org/10.1093/brain/awr232

Geva, S., & Warburton, E. A. (2019). A Test Battery for Inner Speech Functions. *Archives of Clinical Neuropsychology*, *34*(1), 97–113. https://doi.org/10.1093/arclin/acy018

Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., Lachaux, J.-P., Dohen, M., Perrier, P., Garnier, M., Baciu, M., & Lœvenbruck, H. (2019). The ConDialInt Model: Condensation, Dialogality, and Intentionality Dimensions of Inner

Speech Within a Hierarchical Predictive Control Framework. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02019

Heavey, C. L., Moynihan, S. A., Brouwers, V. P., Lapping-Carr, L., Krumm, A. E., Kelsey, J. M., Turner, D. K., & Hurlburt, R. T. (2019). Measuring the Frequency of Inner-Experience Characteristics by Self-Report: The Nevada Inner Experience Questionnaire. *Frontiers in Psychology*, *9*(JAN), 2615. https://doi.org/10.3389/fpsyg.2018.02615

Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin*, *136*(2), 302. https://doi.org/10.1037/a0018436

Hurlburt, R. T., Alderson-Day, B., Kühn, S., & Fernyhough, C. (2016). Exploring the Ecological Validity of Thinking on Demand: Neural Correlates of Elicited vs. Spontaneously Occurring Inner Speech. *PLOS ONE*, *11*(2), e0147932. https://doi.org/10.1371/journal.pone.0147932

Hurlburt, R. T., Heavey, C. L., & Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Consciousness and Cognition*, *22*(4), 1477–1494. https://doi.org/10.1016/j.concog.2013.10.003

Hurlburt, R. T., Heavey, C. L., Lapping-Carr, L., Krumm, A. E., Moynihan, S. A., Kaneshiro, C., Brouwers, V. P., Turner, D. K., & Kelsey, J. M. (2021). Measuring the Frequency of Inner-Experience Characteristics. *Perspectives on Psychological Science*, 1745691621990379. https://doi.org/10.1177/1745691621990379

Langland-Hassan, P. (2021). Inner speech. *Wiley Interdisciplinary Reviews. Cognitive Science*, *12*(2), e1544. https://doi.org/10.1002/wcs.1544

Langland-Hassan, P., Faries, F. R., Gatyas, M., Dietz, A., & Richardson, M. J. (2021). Assessing abstract thought and its relation to language with a new nonverbal paradigm: Evidence from aphasia. *Cognition*, *211*. https://doi.org/10.1016/j.cognition.2021.104622

Langland-Hassan, P., Faries, F. R., Richardson, M. J., & Dietz, A. (2015). Inner speech deficits in people with aphasia. *Frontiers in Psychology*, *6*(528), 1–10. https://doi.org/10.3389/fpsyg.2015.00528

Lupyan, G., Uchiyama, R., Thompson, B., & Casasanto, D. (2023). Hidden Differences in Phenomenal Experience. *Cognitive Science*, *47*(1), e13239. https://doi.org/10.1111/cogs.13239

Meek, G., Ozgur, C., & Dunning, K. (2007). Comparison of the t vs. Wilcoxon Signed-Rank Test for Likert Scale Data and Small Samples. *Journal of Modern Applied Statistical Methods*, *6*(1). https://doi.org/10.22237/jmasm/1177992540

Morin, A. (2011a). Self-Awareness Part 1: Definition, Measures, Effects, Functions, and Antecedents. *Social and Personality Psychology Compass*, *5*(10), 807–823. https://doi.org/10.1111/j.1751-9004.2011.00387.x

Morin, A. (2011b). Self-Awareness Part 2: Neuroanatomy and Importance of Inner Speech. *Social and Personality Psychology Compass*, *5*(12), 1004–1017. https://doi.org/10.1111/j.1751-9004.2011.00410.x

Morin, A., Duhnych, C., & Racy, F. (2018). Self-reported inner speech use in university students. *Applied Cognitive Psychology*, *32*(3), 376–382. https://doi.org/10.1002/acp.3404

Morin, A., & Michaud, J. (2007). Self-awareness and the left inferior frontal gyrus: Inner speech use during self-related processing. *Brain Research Bulletin*, *74*(6), 387–396. https://doi.org/10.1016/j.brainresbull.2007.06.013

Ojemann, J. G., John W., M., & Daniel L., S. (1996). Preserved Function in Brain Invaded by

Tumor. *Neurosurgery*, *39*(2), 253–259. https://doi.org/10.1097/00006123-199608000-00003

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528–537.

Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, *261*, 220–239. https://doi.org/10.1016/j.bbr.2013.12.034

Racy, F., Morin, A., & Duhnych, C. (2020). Using a Thought Listing Procedure to Construct the General Inner Speech Questionnaire: An Ecological Approach. *Journal of Constructivist Psychology*, *33*(4), 385–405. https://doi.org/10.1080/10720537.2019.1633572

Roebuck, H., & Lupyan, G. (2020). The Internal Representations Questionnaire: Measuring modes of thinking. *Behavior Research Methods*, *52*(5), 2053–2070. https://doi.org/10.3758/s13428-020-01354-y

Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, *33*(11), 1433–1454. https://doi.org/10.1016/0028-3932(95)00074-D

Stark, B. C., Geva, S., & Warburton, E. A. (2017). Inner speech's relationship with overt speech in poststroke aphasia. *Journal of Speech, Language, and Hearing Research*, *60*(9). https://doi.org/10.1044/2017_JSLHR-S-16-0270

Tree, J. J., & Playfoot, D. (2019). How to get by with half a loop – An investigation of visual and auditory codes in a case of impaired phonological short-term memory (pSTM). *Cortex*, *112*, 23–36. https://doi.org/10.1016/j.cortex.2018.08.008

Wallace, G. L., Peng, C. S., & Williams, D. (2017). Interfering With Inner Speech Selectively Disrupts Problem Solving and Is Linked With Real-World Executive Functioning. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3456–3460. https://doi.org/10.1044/2017_JSLHR-S-16-0376

Appendix: Tables and Figures

Table 1

Table 1: Demographic information of all included participants

	Aphasia (N=20)	Control (N=20)	Statistics
Age			
Mean (SD)	56.4 (9.09)	55.2 (9.66)	Independent t-test $t(38) = -0.42$, $p = 0.68$
Median [Min, Max]	55.5 [35.0, 72.0]	55.0 [41.0, 70.0]	
Sex			
F	11 (55.0%)	13 (65.0%)	Chi-square test $\chi^2 = 0.38$, $p = 0.54$
M	9 (45.0%)	7 (35.0%)	
Years Education			
Mean (SD)	15.9 (1.79)	16.5 (1.82)	Wilcoxon Rank Sum Test $W = 239.5$, $p = 0.27$
Median [Min, Max]	16.0 [12.0, 18.0]	18.0 [14.0, 18.0]	
Years with Aphasia			
Mean (SD)	8.85 (8.09)	NA	NA
Median [Min, Max]	5.00 [1.50, 31.0]	NA	
Aphasia Type			
Anomic	7 (35.0%)	NA	NA
Broca's	6 (30.0%)	NA	
Conduction	6 (30.0%)	NA	
Transcortical Motor	1 (5.0%)	NA	
WAB-R Aphasia Quotient			
Mean (SD)	73.1 (13.9)	NA	NA
Median [Min, Max]	70.9 [47.8, 94.0]	NA	

WAB-R = Western Aphasia Battery – Revised.

Table 2

Table 2: Descriptive Statistics for individuals with aphasia and Controls on Objective and Subjective Inner Speech Tasks, with Section 1 as proportions and Section 2 as ratings from 1-5.

Section	Task	Aphasia (N=20)	Control (N=20)	Statistics
Section 1.1	**Inner Picture Rhyming**			
	Mean (SD)	0.643 (0.166)	0.946 (0.0646)	W = 18, p < .001
	Median [Min, Max]	0.592 [0.368, 0.947]	0.970 [0.790, 1.00]	
	Inner Written Rhyming			
	Mean (SD)	0.634 (0.169)	0.948 (0.0409)	W = 22, p < .001
	Median [Min, Max]	0.625 [0.350, 0.950]	0.950 [0.880, 1.00]	
	Inner Auditory Rhyming			
	Mean (SD)	0.755 (0.148)	0.961 (0.0341)	W = 28.5, p < .001
	Median [Min, Max]	0.750 [0.450, 1.00]	0.950 [0.880, 1.00]	
Section 1.2	**Naming SIS**			
	Mean (SD)	0.972 (0.0397)	0.993 (0.0133)	W = 142.5, p = .08
	Median [Min, Max]	0.973 [0.838, 1.00]	1.00 [0.970, 1.00]	
	Overt Naming			
	Mean (SD)	0.786 (0.217)	0.992 (0.0170)	W = 28, p < .001
	Median [Min, Max]	0.865 [0.135, 1.00]	1.00 [0.950, 1.00]	
Section 2	**Idea without Word**			
	Mean (SD)	2.70 (1.08)	2.35 (0.933)	W = 236.5, p = .30
	Median [Min, Max]	3.00 [1.00, 5.00]	2.00 [1.00, 4.00]	
	Successful Inner Speech			
	Mean (SD)	3.55 (1.32)	3.39 (1.61)	W = 184, p = .91
	Median [Min, Max]	4.00 [1.00, 5.00]	4.00 [1.00, 5.00]	
	Inner vs. Outer			
	Mean (SD)	1.90 (0.718)	2.95 (0.826)	W = 74, p < .001
	Median [Min, Max]	2.00 [1.00, 3.00]	3.00 [2.00, 5.00]	
	Post-Stroke			
	Mean (SD)	3.25 (1.02)	NA	NA
	Median [Min, Max]	3.00 [1.00, 5.00]	NA	

Table 3

Table 3: Scores on tasks and ratings requiring lexical-semantic access, highlighting the individual differences in participants with moderate aphasia.

Participant	**Aphasia Severity (AQ)**	**Overt Naming Errors (%)**	**Overt Naming Score (%)**	**Naming SIS (%)**	**Rating- SIS (Scale 1-5)**	**Objective- Picture Rhyming (%)**
1003	Moderate (67.6)	Phonemic: 0 Semantic: 50 Other: 50	92	95	3 - sometimes	55
1008	Moderate (47.8)	Phonemic: 28 Semantic: 0 Other: 72	46	100	5 - almost always	50
1011	Moderate (60.2)	Phonemic: 3 Semantic: 22 Other: 75	14	100	5 - almost always	50
1016	Moderate (52.4)	Phonemic: 0 Semantic: 24 Other: 76	54	100	2 - rarely	82

Figure 1

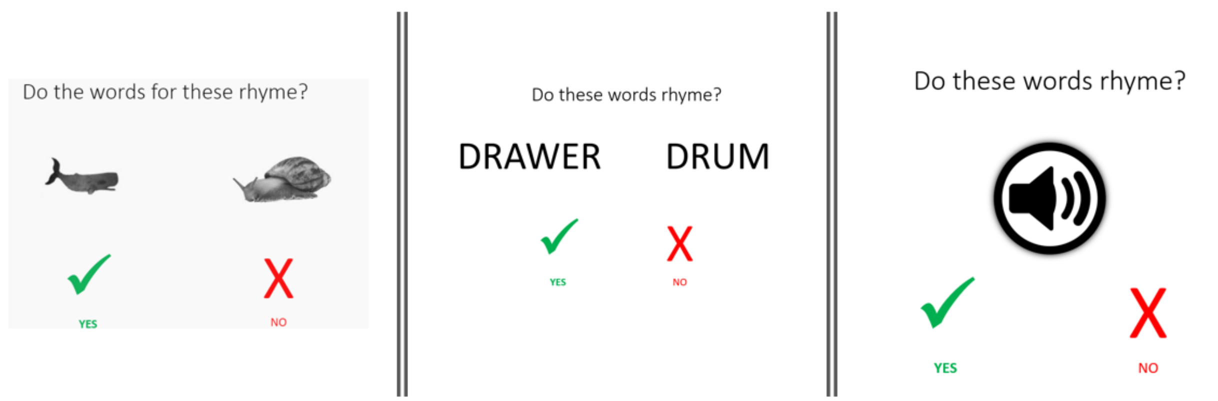

Figure 1: Examples of displays for picture rhyme judgement, written rhyme judgement, and auditory rhyme judgement. Participants responded by touching the green check mark or the red x.

Figure 2

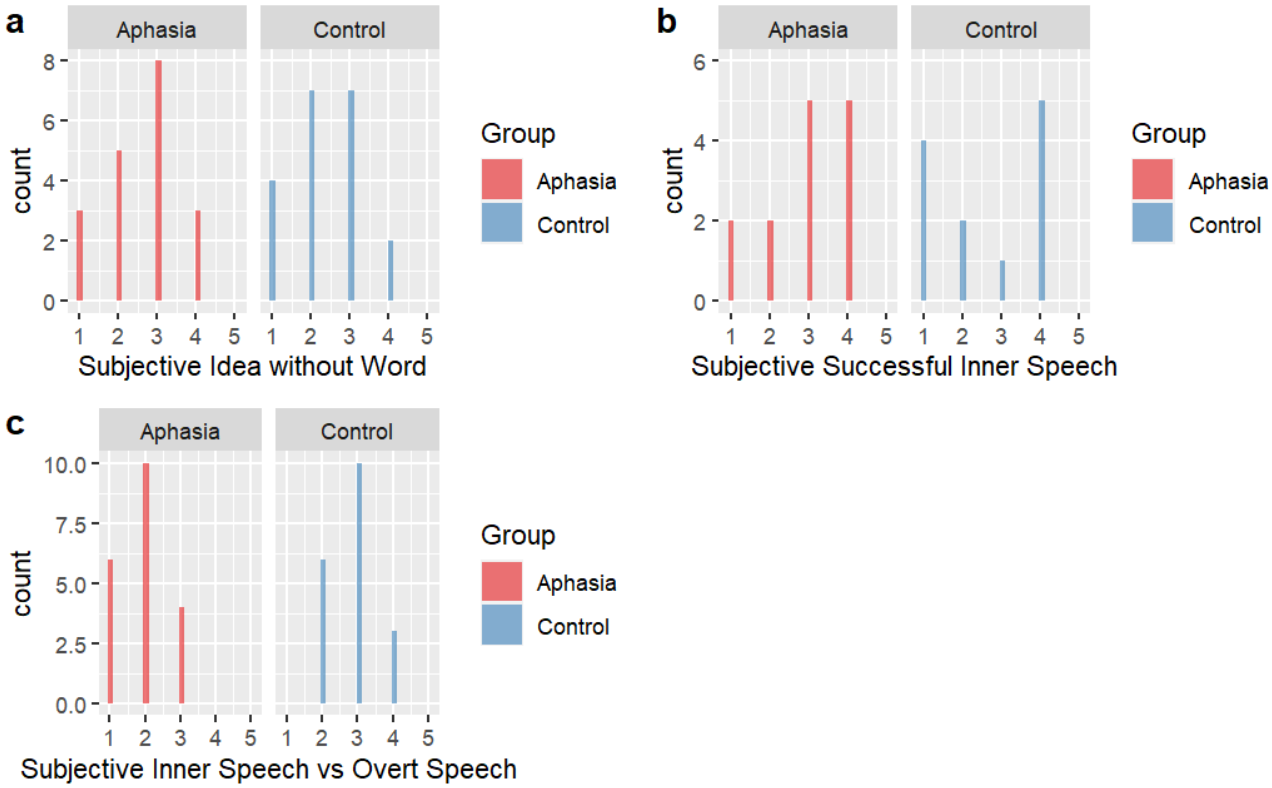

Figure 2: Subjective Ratings with persons with aphasia (individuals with aphasia) and neurotypical controls comparison.
Idea without Word (IwW), "How often do you see something and know what it is but can't say the word for it out loud?" with 1 being almost never and 5 being almost always; Successful Inner Speech (SIS), "A few minutes ago we talked about how we sometimes see something and know what it is but can't say the word for it out loud. When that happens to you, how often can you say the right word for it in your head, and have it sound right?" with 1 being almost never and 5 being almost always; Inner vs. Overt Speech, "In general, how do you think your abilities to speak in your head compare to your abilities to speak out loud?" with 1 being much better in head and 5 being much better out loud

Figure 3

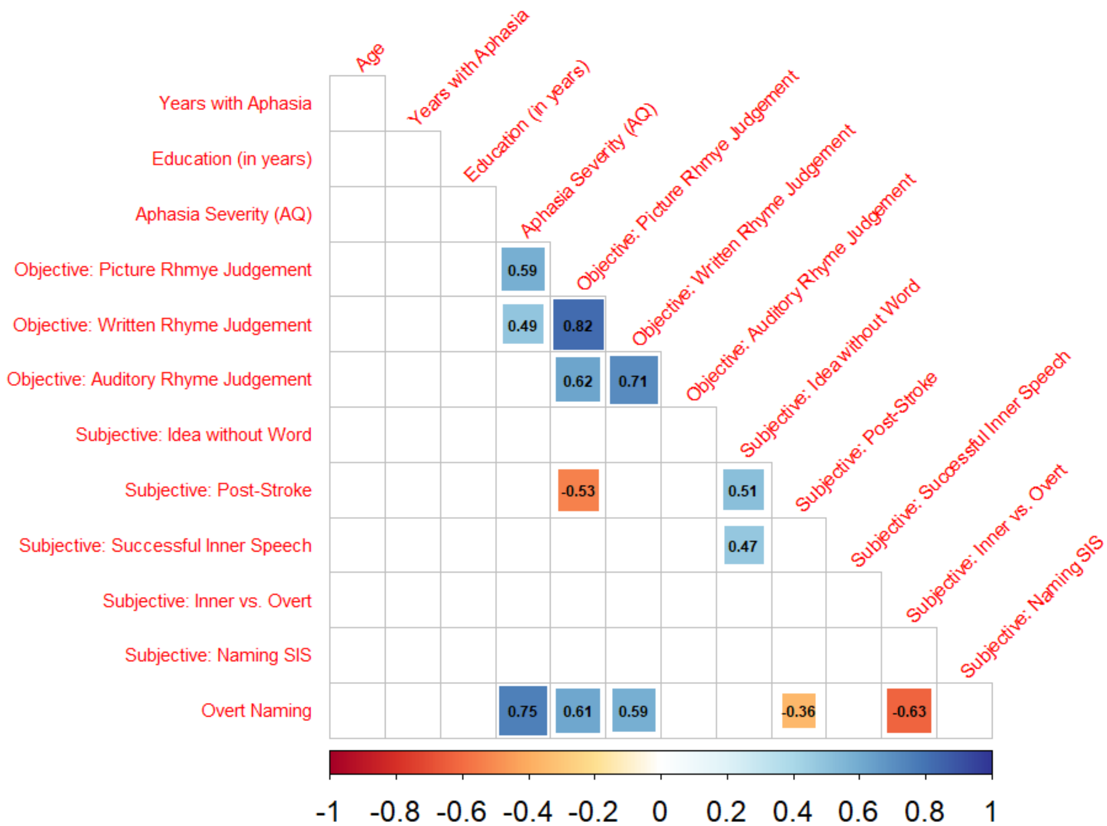

Figure 3: Spearman's correlations between objective and subjective measures, as well as demographic variables. All correlations shown are significant with a 95% confidence interval. Larger squares and darker hues indicate stronger correlations, also shown by the numbers in each square.

Figure 4

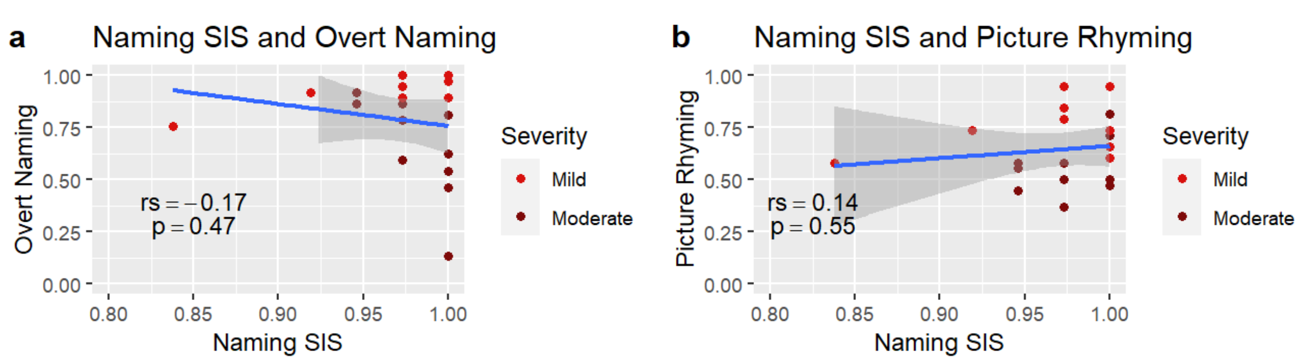

Figure 4: Correlations between reports of successful inner speech (SIS) in-the-moment of silent

naming ("Naming SIS") with a) overt (out loud) naming performance and b) performance on the objective picture rhyme judgement task for individuals with aphasia. Spearman's correlation coefficient (rs) and significance values are provided. Additionally, dots represent each participant, with light red dots representing those with mild aphasia severity, and dark red dots representing those with moderate aphasia severity. The trend line (in blue) includes all participants with aphasia.

Figure 5

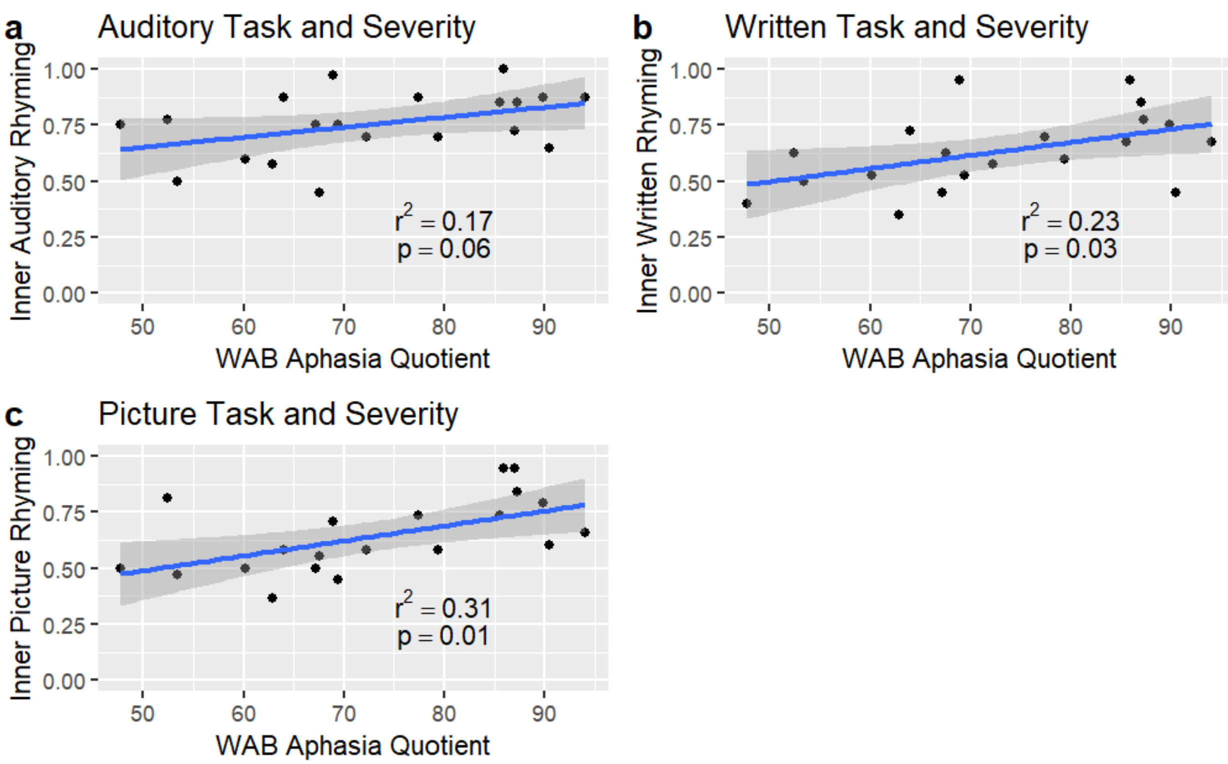

Figure 5: Objective Inner Rhyming Judgement Tasks - a) Auditory: participants hear two words; b) Written: participants see two words; c) Picture: participants see drawings of two objects. x-axis is WAB-R AQ, where a higher score is milder aphasia severity.

Figure 6

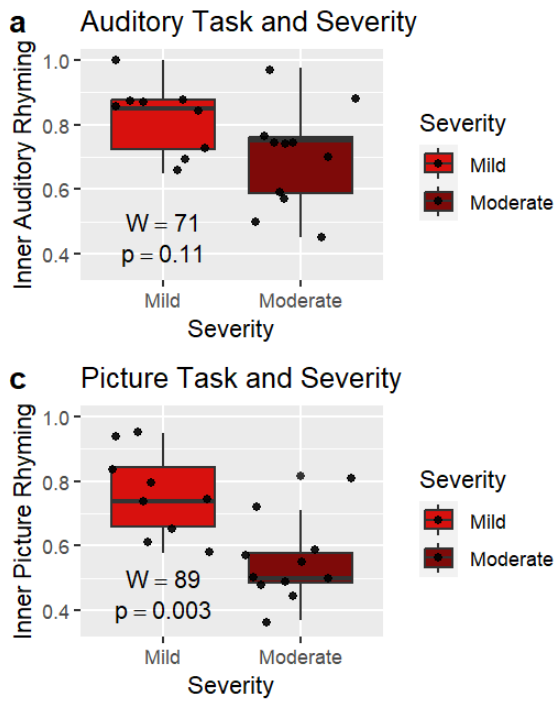

Figure 6: Objective Inner Rhyme Judgement and Categorical Severity - a) Auditory; b) Written; c) Picture. The groups were split into mild severity and moderate severity based on the cut-offs from the WAB-R. Jittered dots represent individual performance.

Figure 7

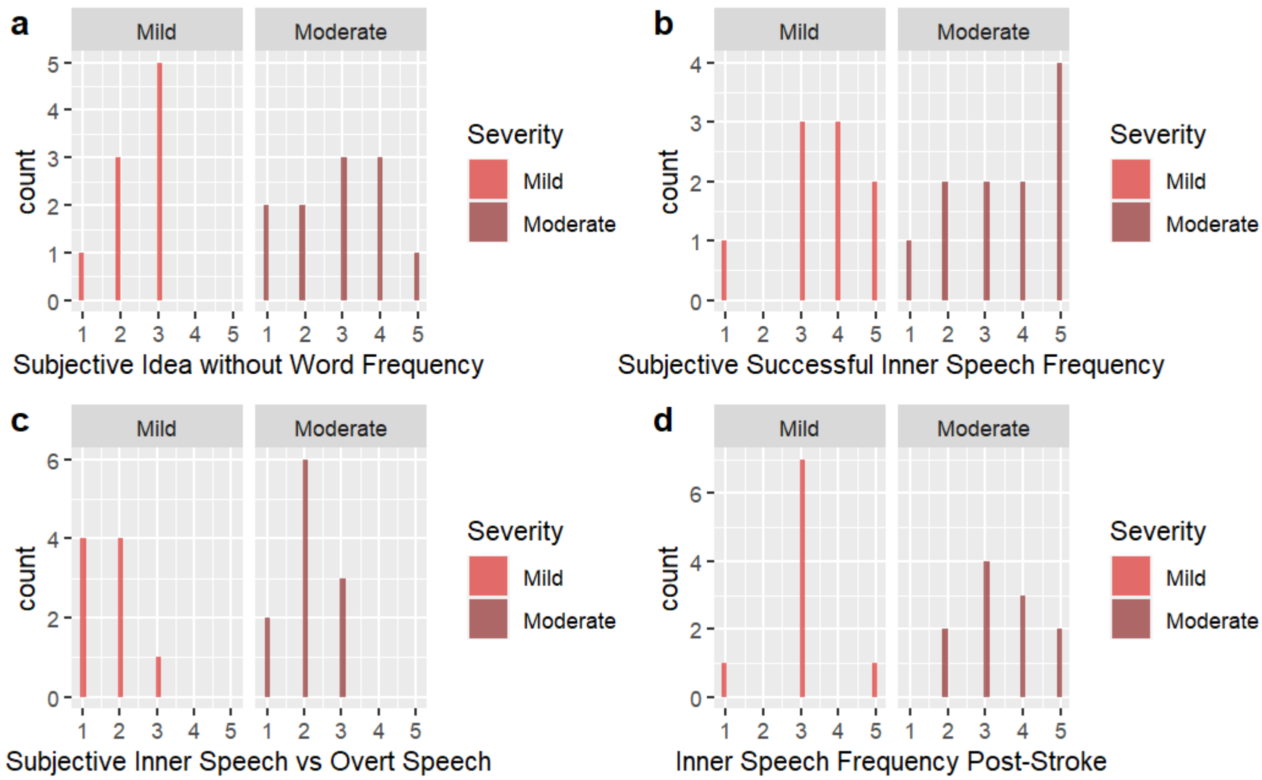

Figure 7: Subjective Ratings, comparisons between mild and moderate aphasia. Idea without Word (IwW), "How often do you see something and know what it is but can't say the word for it out loud?" with 1 being almost never and 5 being almost always; Post-Stroke, "Do you think you talk to yourself in your head less now, more now, or about the same as before your stroke?" with 1 being much less now and 5 being much more now; Successful Inner Speech (SIS), "A few minutes ago we talked about how we sometimes see something and know what it is but can't say the word for it out loud. When that happens to you, how often can you say the right word for it in your head, and have it sound right?" with 1 being almost never and 5 being almost always; Inner vs. Overt Speech, "In general, how do you think your abilities to speak in your head compare to your abilities to speak out loud?" with 1 being much better in head and 5 being much better out loud