This is the post print version of the article, which has been published in Journal of speech,language and hearing research. 2018, 61 (5), 1086-1103. http://dx.doi.org/10.1044/2018_JSLHR-S-17-0326 .

Auditory-Perceptual Assessment of Fluency in Typical and Neurologically Disordered Speech

Nelly Penttilä,[a] Anna-Maija Korpijaakko-Huuhka,[a] and Raymond D. Kent[b]

[a] Faculty of Social Sciences, University of Tampere, Tampere, Finland

[b] Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin

Correspondence to Nelly Penttilä

nelly.penttila@uta.fi

University of Tampere,

Kalevantie 4

Main Building/Room D227

33100 Tampere Finland

Abstract

Purpose: The aim of this study is to investigate how speech fluency in typical and atypical speech is perceptually assessed by speech-language pathologists (SLPs). Our research questions were: (a) how do SLPs rate fluency in speakers with and without neurological communication disorders; (b) do they differentiate the speaker groups; and (c) what features do they hear impairing speech fluency?

Method: Ten SLPs specialized in neurological communication disorders volunteered as expert judges to rate 90 narrative speech samples on a Visual Analogue Scale (VAS). The samples – randomly mixed – were from 70 neurologically healthy speakers (the control group) and 20 speakers with traumatic brain injury (TBI), 10 of whom had neurogenic stuttering (designated as clinical groups A and B).

Results: The fluency rates were higher for typical speakers than for speakers with TBI; however, the agreement among the judges was higher for atypical fluency. Auditory-perceptual assessment of fluency was significantly impaired by the features of *stuttering* and *something else* but not by *speech rate*. Stuttering was also perceived in speakers not diagnosed as stutterers. A borderline between typical and atypical fluency was found.

Conclusions: Speech fluency is a multifaceted phenomenon, and based on this study, we suggest a more general approach to fluency and its deviations that will take into account, in addition to the motor and linguistic aspects of fluency, the metalinguistic component of expression as well. The results of this study indicate a need for further studies on the precise nature of borderline fluency and its different disfluencies.

Keywords: Fluency, Fluency Disorders, Speech- and Language Pathology, Traumatic Brain Injury, Stuttering

Auditory-Perceptual Assessment of Fluency in Typical and Neurologically Disordered Speech

Spontaneous speech is disfluent by its very nature due to the linguistic and cognitive processing demands that affect speech-motor performance (Brown, Ingham, Ingham, Laird & Fox, 2005; Kent, 2000; Lehtonen, 1976). Fox Tree (1995) estimated that in typical speech approximately 6% of words are disfluent. Disfluent speech also tends to increase with advanced age or when the planning demands of the speaking task become greater (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001). In natural speech, such as ordinary conversation, speakers may hesitate, choose wrong words, interrupt themselves and repeat words and syllables – and still, their speech is considered fluent. Similar disfluencies are found in the speech of people who are suffering from various communication disorders after traumatic brain injury (TBI), for example, apraxia of speech (AOS), dysarthria, aphasia, and in some rare cases, neurogenic stuttering (McDonald, Togher, & Code, 1999; Van Borsel, 1997). When evaluating speech fluency in actual clinical practice, speech-language pathologists (SLPs) can approach the assessment of speech disorders using three methods. Specifically, speech is assessed perceptually by the SLP, subjectively using self-assessment by the client, and “objectively” using published tests and other measurements. This current study focuses on the auditory-perceptual assessment of speech fluency by SLPs in both the typical and disordered speech of adults. Topics concerning auditory judgments, the varying terminology concerning fluency, and tasks used in the assessments are discussed in the following sections.

Auditory-Perceptual Judgments in Speech-Language Pathology

In speech-language pathology, both in clinical and research practice, auditory-perceptual judgments (APJ) are commonly used for identifying and classifying the different aspects of disordered speech, for example, voice and fluency disorders, and also for assessing the outcome of interventions provided for clients with these disorders (Chan & Yiu, 2002; Cordes & Ingham, 1995; Gerratt, Till, Rosenbek, Wertz & Boysen, 1991; Kent, 1996). Often APJ are considered the gold standard by which other approaches are validated. SLPs are trained to identify and differentiate communication disorders by observing behaviors and using instruments, such as those permitting acoustic analysis. These perceptual observations are combined with instrumental data to gain a comprehensive view of the client's communication disorder (McCauley, 1989). One might think that SLPs' ears are calibrated to be extremely sensitive to disturbances in speech, and their experience and training are expected to improve this sensitivity (Kreiman, Gerratt, and Precoda, 1990). However, this increased sensitivity may complicate and even mislead their actual perceptual evaluation. In a study by Dagenais, Watts, Turnage, and Kennedy (1999), SLPs ranked speech intelligibility and acceptability more strictly – compared to naïve listeners – when assessing dysarthric speech. In addition, knowing the diagnosis of the speaker may lead to overpathologizing the assessed features of speech and voice as was shown in a study by Eadie, Stroka, Wright, and Merati (2011). In it, listeners judged speech samples with and without dysphonia as being more severely disordered when they were aware of the speakers' diagnoses. Thus, focusing on speech disorders may result in hypersensitivity for disfluencies, which in turn causes the listener to pathologize normal variation and the different aspects of speech fluency in typical speakers. Therefore, the border between typical and atypical may become even fuzzier as attention is focused on the speech pattern.

As proposed by Kreiman, Gerratt, Kempster, Erman, and Berke (1993), listeners make perceptual judgments based on stored mental representations that then serve as internal standards. Expert background, continuous training, and diagnostic information on the speakers are examples of these so-called internal standards that can influence auditory perception and evaluation (Chan & Yiu, 2002; Kreiman *et al.*, 1993). To modify and secure the internal standards and increase the reliability of auditory perceptual evaluation, external standards, like audio speech samples, are needed. The quantity of training samples does vary across studies, but common for all the training materials used is that these samples are considered to act as anchors for shaping the judges' internal standards and seek to represent the varying degrees of the measured object – or at least – what is considered to be typical.

The Problem of Agreement When Assessing Fluency

The agreement levels in APJ studies of speech fluency and disfluencies have been problematic (Brainbridge, Stavros, Ebrahimian, Wang, & Ingham, 2015; Kent, 1996; Young, 1984). Different types of tasks used in APJ studies seem to produce different results (Kempster, Gerratt, Verdolini Abbott, Barkmeier Kraemer, & Gillman, 2009) and it appears to be difficult to obtain high agreement levels independent of the task type. First, in *discrimination tasks*, the agreement levels in stuttering research have been very low (e.g., Finn, Ingham, Ambrose, and Yairi, 1997; Hoffman, Wilson, Copley, Hewart, and Lim. 2014). Secondly, agreement levels have also been unreliable for *identification tasks*, as in the study by Cordes and Ingham (1995) wherein 10 experienced judges disagreed markedly when identifying stuttering. Further, Brundage, Bothe, Lengeling, and Evans (2006) compared agreement levels in three judgment groups in a task on identifying stuttering. The experienced judges were 10 researchers who had participated in Cordes and Ingham's (1995) study, 31 clinicians (SLPs) and 41 students of the communication sciences and disorders. Both students and clinicians under-identified stuttering and neither the clinicians' intrajudge agreement nor their interjudge agreement differed from the students' agreement levels. The identification of AOS seems to be equally difficult. In a study by Haley, Jacks, Riesthal, Abou-Khalil, and Roth (2012), three SLPs assessed individuals with aphasia for AOS. The results showed that the inter-observer agreement for AOS diagnosis was very low and the reliability of perceptual scaling related to various speech dimensions was limited.

Methodological refinements have been suggested as a way to improve the agreement between judges. In addition to the above-mentioned training and designated anchors to calibrate the rating scale (Chan & Yiu, 2002), the concepts and dimensions used in the assessment should be clarified for the judges and agreed on (Kent, 1996). For example, in a study by Bothe (2008), the interjudge agreement for highly experienced judges for the identification task of stuttering varied from 39.0% to 89.1%, depending on the definitions being used. In addition, the type of scale should be suitable for the task chosen. Chan and Yiu (2002) used a Visual Analog Scale (VAS) in their study of perceptual voice evaluation. Cannito, Burch, Watts, Rappold, Hood, and Sherrard (1997) also found VAS to be a good tool when scaling disfluency in their study of disfluencies in spasmodic dysphonia. Finn and colleagues (1997) used a 9-point scale to assess speech naturalness, while Susca and Healey (2001, p. 64) decided to use a Likert scale (1=strongly disagree to 7=strongly agree) to estimate how listeners agreed with the statement, "*this person is a fluent speaker*". Of all the scales, VAS has been found to be a reliable and user-friendly tool to evaluate impressions, and because the measured data are in millimeters, the data are suitable for parametric statistical analysis. In addition, VAS can accommodate variations in ratings rather than simply forcing judges to pick values from an interval scale (Cannito *et al.*, 1997; Chan & Yiu 2002; Kempster *et al.*, 2009).

Variations in Terminology

As mentioned earlier, the low agreement levels in APJ studies have been problematic. Besides overpathologizing and producing possible bias toward what is atypical, the listeners' internal standards and methodological choices, like task effects, may result in these low reliability rates. In addition, one possible and maybe the most elementary difficulty when reaching satisfactory agreement levels in APJ studies is the wide variety in the terminology regarding fluency. As the concept of fluency is utilized in different branches of science concerned with human speech, its meaning needs to be considered from each of those perspectives in order to understand the various aspects of naturally fluent – or should we say – naturally disfluent speech.

Disfluencies of speech seem to have different meanings depending on the focus of the research. Some researchers use the term *dis*fluencies only when referring to "normal disturbances" of speech and the term *dys*fluencies for an abnormal flow of speech (Quesal, 1988). On the other hand, Nicolosi, Harryman, and Kresheck (1987) define disfluency as an alternative spelling for dysfluency. In spite of this terminological ambiguity, the most common disruptions of fluent speech flow in typical speech are agreed upon as being interjections, hesitations, revisions, interruptions, and word and phrase repetitions (Ambrose & Yairi, 1999; Campbell & Hill, 1994). The dysfluencies in atypical speech are most commonly classified according to their relationship to stuttering, frequency, location, or by the distinction between what is typical and pathological (see Table 1).

Disturbances in the Motor Coordination of Speech. In the context of motor speech disorders, speech fluency is defined as the smooth coordination of different motoric processes related to respiration, phonation, and articulation (Duffy, 2005; Guenther, Gosh, Nieto-Castanon, & Tourville, 2006). The classic fluency disorder stuttering is, therefore, defined as the opposite: Non-fluent speech where the smooth coordination is disturbed by repetitions, prolongations, and blocks (Ambrose & Yairi, 1999; Guitar, 2006). The motoric coordination is also disturbed in AOS where difficulties in programming movements are thought to trigger false starts and revisions (Deger & Ziegler, 2002; van Lieshout, Bose, Square, & Steele, 2007). Acquired stuttering is a generic term used to describe stuttering that is not developmental, but instead manifests in a formerly fluent speaker after a neurological or psychological episode (Van Borsel, 2014). Neurogenic stuttering can have its origin in a myriad of lesion sites in the brain, but the clinical observations show similarities with developmental stuttering, and often there do not seem to be distinguishing features between acquired and developmental stuttering (Krishnan & Tiwari, 2013; Van Borsel, 2014; Van Borsel & Taillieu, 2001). Also, patients with neurogenic stuttering tend to show additional signs associated with aphasia and dysarthria (Lundgren, Helm-Estabrooks & Klein, 2010; Tani & Sakai, 2011). Therefore, it is questionable whether neurogenic stuttering is a distinct disorder or just a manifestation of other fluency disorders or a general breakdown in fluency related to factors that impede the processes of spoken communication. In this present study, we distinguished neurogenic stuttering after TBI from palilalia, cluttering, and AOS with an excluding process and following general characteristics of neurogenic stuttering (Lundgren, Helm-Estabrooks, & Klein, 2010; Tani and Sakai, 2011; see flowchart in Appendix A).

Disfluencies of Linguistic Origin. In aphasiology, the term fluency refers to *language-fluency*, and aphasias are categorized as fluent and non-fluent types according to the classic Wernicke-Lichtheim-Geschwind paradigm (Basso, 2003; Goodglass & Kaplan, 1983; Park *et al.,* 2011). Language-fluency is defined as the ability to produce semantically and grammatically appropriate and uninterrupted words and phrases that are easily articulated. Accordingly, the classic non-fluent aphasia, Broca's aphasia, is characterized by the effortful production of short and agrammatic utterances. In fluent aphasia, as in Wernicke's aphasia, language-fluency is compromised by word retrieval difficulties, i.e., paraphasias and neologisms (Basso, 2003; Harmon, Jacks, Haley, & Faldowski, 2015). Word-finding difficulties in general may result in such disfluencies as pauses and hesitations during word-search. Language formulation difficulties, like phonological approximations and impaired sentence processing, also have an impact on speech rhythm and may, therefore, sound like stuttering disfluencies with repetitions, false starts, and revisions.

When differentiating between fluent and non-fluent aphasias, many features have been examined, such as the number of words produced (Gordon, 2006; Kreindler, Mihailescu, & Fradis, 1980), total speaking time (Feyereisen, Verbeke-Dewitte, & Seron, 1986), and perceptible struggling (Benson, 1967). The influence of these factors on listener judgment of speaker fluency was examined by Park *et al*. (2011). Speech samples were collected from adult patients with different etiologies, such as stroke, dementia, and TBI. Additional, 20 healthy adults were recruited to represent typical language fluency. The results showed that higher speech rate and higher productivity in addition to lack of audible struggle discriminated fluent speakers from non-fluent ones. Fillers, pauses, and repeated syllables were not significant when discriminating between these speaker groups. Similarly, Gordon (2006) found, that a higher speech rate separated fluent aphasic speakers from non-fluent ones.

Fluency as a Component of Speaking Proficiency. In second-language-learning studies, the term fluency is used to describe native-like and non-native language performance (Cucchiarini, Strik,

& Boves, 2000). In this context, fluency in the broader sense means grammatically error-free and lexically rich output with native-like pronunciation (Lennon, 1990). From a narrower perspective, fluency is seen as a component of speaking proficiency. According to Segalowitz (2010), second-language fluency has three facets: cognitive fluency (language formulation), utterance fluency (articulation) and perceived fluency (listeners' inference). The respective disfluencies are classified as lexical and grammatical errors, mispronunciations, and problems with prosody, such as having a strong accent.

In a classic Finnish study, Jaakko Lehtonen (1976; p. 66) concluded that "*fluency is not the sort of physically objective and unambiguous property of speech that it could be measured by means of some simple phonetic parameters such as speech rate or quantity of pauses.*" Instead, he indicated that fluency is a complicated concept, and its description should include linguistic, psychological, and sociolinguistic factors. This can be interpreted to mean that in addition to the smooth motor continuity of speech flow and the linguistic aspects of lexical and grammatical productivity, way of expression should be taken into account, including the stylistic, functional, and textual aspects of speech.

Purpose of the Present Study

The foregoing review indicates that disfluency is a multi-faceted concept that is conceived somewhat differently in different fields of study. There is much debate about the assessment of disfluency in developmental stuttering, but the challenges are even greater in the assessment of disfluency in individuals with acquired neurological disease or damage because disfluencies occur not as an isolated condition, but rather potentially co-occur with aphasia, apraxia of speech, and/or dysarthria. Another challenge is the discrimination of these disorders from each other perceptually. In judgments by SLPs, the border between what is typical and what is not might also be biased in that assessments of typical speakers tend to be based on a stricter criterion that can often lead to overpathologizing. For these reasons, it is important to determine how well SLPs agree in their assessments of disfluency in those individuals with acquired neurogenic communication disorders. We also need to increase our knowledge and awareness of the criteria that SLPs do apply when perceptually assessing fluency, to develop sustainable methodologies for APJ studies, and most importantly, develop more reliable clinical diagnostics and effective intervention methods for communication disorders after TBI. The specific aim of this study is to examine how speech fluency is auditory-perceptually assessed by experienced clinical evaluators, SLPs, in people with and without TBI. Our research questions were the following: 1) how do SLPs rate fluency in speakers with and without neurological communication problems, 2) do SLPs differentiate the speaker groups on the basis of fluency ratings, and 3) what features do SLPs hear that impair speech fluency?

Method

Participants

The participants of this study included a group of SLPs as listeners (a jury of speech pathologists) and three groups of speakers: A control group representing typical Finnish speakers, and two clinical groups representing speakers with TBI with and without neurogenic stuttering. Study participation was voluntary, and all participants provided informed written consent before their inclusion. The ethics related to consent forms, flyers, and the experiment procedures, were approved by the Faculty of Social Sciences at the University of Tampere (Finland).

Control Group. Participants in the control group (CG) were recruited based on convenience sampling to represent typical Finnish speakers. Primary inclusion criteria were Finnish language as the mother tongue, an age of 18 years or more, and a self-reported absence of developmental or acquired communication disorders, learning disability, hearing loss or neurodegenerative diseases. Seventy-six persons volunteered for the study, and they were thoroughly interviewed by a SLP (1st author) to ascertain that they met the inclusion criteria. Based on that interview, six persons were excluded, two due to rhotacism and four because of one of the following, namely, treated developmental stuttering, bilingual background, functional voice disorder, or being underage. Thus, 70 participants with no communication disorder were finally selected for this study. All were considered to be typically fluent speakers in the interview situation. Of the 70 participants in the CG, 36 were female and 34 male. Their mean age was 41.1 years (SD=14, range 18–89 years). Most of the participants were right-handed; 13% were left-handed. Educational backgrounds varied; however, most of the participants (64.2%) had completed a higher-level education: Secondary education was completed by 33% of the participants, and only elementary school by 3%. It was considered desirable to include a relatively large group of typical speakers in order to establish normative patterns in speech fluency for the selected research task.

Clinical Group A. For the clinical group A (CGA), the participants were recruited from the Finnish Brain Injury Association and by SLPs who were working in neurological departments. The major inclusion criteria were TBI established by neuroimaging and the presence of a communication disorder. Other criteria were Finnish as the mother tongue, age over 18 years, no communication disorders before the head trauma and no background of bilingualism or hearing loss. Twenty-two subjects with TBI volunteered as candidates for the CGA. Twelve were excluded, three due to brain tumor, five due to a stroke, three because of missing results from the neuroimaging, and one because the etiology of his injury was poisoning. Ten participants with TBI comprised CGA in this study. Nine of them were male and one was female. The participants' mean age was 46.3 years (SD=11.6, range 34–64 years). All were right-handed. Typical educational background for this group was a secondary education (6/10); three participants had completed a Bachelor's degree and one completed elementary school only. The participants in CGA showed various communication disorders due to their injury, and all participants suffered from multifaceted cognitive deficits, including changes in attention, problem-solving, executive functioning and/or memory (see Table 2). Aphasia and speech motor disorders were assessed using the Finnish version of the Western Aphasia Battery (WAB; Pietilä, Lehtihalmes, Klippi & Lempinen, 2005) or the Boston Diagnostic Aphasia Examination (Laine, Koivuselkä-Sallinen, Hänninen & Niemi, 1997) by those SLPs who recruited the candidates for this study. Neurological and neuropsychological data were collected from medical files with the participants' permission. All participants had one of their significant others (spouse, friend, aide, adult child) with them during the test and interview session to confirm the validity of the information gathered.

Clinical Group B. Clinical group B (CGB) was also recruited from the Finnish Brain Injury Association and by SLPs' working in neurological departments. The two major inclusion criteria were TBI established by neuroimaging and onset of stuttering subsequent to head injury with or without aphasia and/or dysarthria. Other criteria for inclusion were Finnish as the mother tongue, age over 18 years, no communication disorders before the head trauma (especially no developmental stuttering), no AOS, cluttering, echolalia or palilalia after the head trauma, and no hearing impairment before or after the injury. Eighteen subjects with TBI and neurogenic stuttering (NS) volunteered as candidates for the CGB. They had been originally assessed for neurogenic stuttering, aphasia, and other related motor speech disorders by the SLPs who informed the candidates of this study. Another SLP (1st author) further confirmed neurogenic stuttering using the excluding process described in Appendix A. Consequently, eight candidates were excluded because their stuttering-like disfluencies were related to either word repetitions due to word-finding difficulty (n = 4), AOS (n = 2), palilalia (n = 1) or cluttering (n = 1), instead of neurogenic stuttering. Thus, 10 participants with NS after TBI comprised the CGB. Of these, seven were male and three female. The participants' mean age was 40.9 years (SD=15.9, range 19-61 years). All of the participants in CGB were right-handed. Their educational backgrounds varied from a Bachelor's degree (4/10) and secondary education (3/10) to elementary school only (3/10). All participants also had self-identified the onset of their stuttering. All participants also had one of their significant others (spouse, friend, aide, adult child) with them in the test and interview session to confirm the validity of the information gathered. Data from the neurological and speech-language pathology assessments for CGB are summarized in Table 3.

Jury of Speech-Language Pathologists. For the auditory-perceptual evaluation of speech fluency of the three speaker groups, 17 SLPs' working in neurology departments were invited to form an expert jury. Criteria for inclusion were 1) a minimum of 2 years' experience as a SLP; 2) current position in a clinic where most patients are adults; and 3) clinical specialization in neurogenic speech and language disorders in adults. Ten SLPs volunteered for this study and provided informed written consent of their participation. Most importantly, professional secrecy with relation to the audio samples for the experiment and other jury members was demanded. All jury members were women, with a mean age of 41.6 years of age (SD=12.1, range 27–60 years) (see Table 4). Their mean professional work experience as a SLP was approximately 14 years.

Speech Samples

The speech samples were audio recorded from a narrative story based on a comic strip with a Zoom H2 –device (Zoom Corpoartion, Tokyo, Japan) wherein the microphone was positioned approximately 20 cm from the participant's mouth. In the narrative task, the speakers were asked to generate a story based on a 9-frame comic strip depicting a little man (Ferd'nand by Henning Dahl Mikkelsen) in his garden, first sowing seeds and later chasing birds with a scarecrow (hence the scarecrow story). The scarecrow story task has been commonly used in speech and language studies in Finland (see Korpijaakko-Huuhka & Lind, 2012). The speakers delivered stories of varying length (with a range of 21–352 seconds). To make the texts more comparable, we decided to edit the narratives on a semantic basis. Thus, all talk related to the first and the two last pictures was cut, and the new samples covered pictures 2–7. This decision was based on the presumption that at the beginning of storytelling, a speaker's speech rate may be unnaturally slow due to simultaneous planning of the storyline and expressions, and especially because brain injured speakers may also show some difficulties getting started (Biddle, McCabe & Bliss, 1996). On the other hand, in this task toward the end of the story, speakers tend to increase their speech rate, as the need to process the storyline diminishes (Korpijaakko-Huuhka & Aulanko, 1994). The edited speech samples had durations of between 12 and 196 seconds. In the Laboratory of Music Studies at the University of Tampere, the audio samples were normalized for peak intensity, and their loudness levels were equalized by defining the RMS (root mean square) level and modifying the loudness to -10dB RMS using MAGIX® Sound Forge Pro 10.0. The samples were then randomly mixed and burned on a CD for the listening experiment.

Perceptual Evaluation of Fluency

First, the SLPs completed a background information form. Then they were instructed to rate speech fluency on a 120 mm long VAS by drawing a vertical line on the spot that best represented their rating. Anchors were located on the scale to represent the most severely disordered speech fluency (0 mm at the left end), moderately disordered fluency (40 mm), and average fluency (80 mm). The right end (120 mm) represented a high, if not exceptional, degree of speech fluency. The line lengths in the millimeters measured (from 0 to 120) on the scale were transformed into scores for the statistical analysis. Each 10 mm corresponded to a whole score (i.e., 10 mm equaled 1 point), and, thus, the total variation of the scores ranged from 0 to 12. If the judges rated speech fluency as less than average (under 80 mm) they were asked to report if *speech rate*, *pauses*, *stuttering*, *language difficulties* or *something else* alone or in different combinations had affected their decision. Language difficulties were defined as word-finding difficulties, interrupted words and phrases, paraphasias, neologisms, and agrammatism. The rationale for including the feature entitled *something else* was to give the judges the opportunity to react more than to the investigator-assigned categories, since open-ended ratings have been used because "*they have the advantage of allowing respondents to answer in their own frames of reference, implicitly reducing the priming influences of researcher-suggested alternatives*" (Mossholder, Settoon, Harris, & Armenakis, 1995, p. 338).

Before the experiment, the SLPs listened to four educational audio-samples that corresponded to the anchors mentioned above. The length of the total audio for the perceptional evaluation task was 77 minutes. To maintain the jury's attention, the listening experiment was completed in two parts.

First, the jury evaluated 45 samples, and then they took a break. During the break, they were cautioned not to discuss the evaluation task or any of the audio samples they had heard. After the break, the remaining 45 samples were evaluated. After the evaluation session, the task and the speech samples were discussed in an informal debriefing session.

Data Analyses and Statistics

The data management and analyses were performed using SPSS 24.0 (2016). Using the One-way ANOVA, the assessment of homogeneity between the judges' ratings was measured. For a deeper analysis, the agreement between SLPs was tested using a post hoc-test (Bonferroni) and Fleiss' Kappa for each speaker group. To study how the SLPs differentiated the speaker groups, the speech data were divided into groups in two ways. First, the groups were formed by using centroid linkage and creating a dendrogram, and secondly, speakers were grouped based on the number of identified features that affected the fluency ratings. With an independent samples *t*-test, the equality of samples in CGA and CGB was assessed. In order to assess the differentiation between the control group and clinical groups, the Kruskal-Wallis test was also used as well as pairwise comparisons. Agreement of the features observed by the jury members was analyzed using One-way ANOVA. Correlations between given features and fluency ratings in each speaker group were first examined with the Pearson's Correlation test (sig. 2-tailed), where correlation was significant at the 0.05 level. After that, the Pearson product moment correlation coefficient was used for a regression analysis to determine the relationship between the given features and fluency ratings in the aggregate data.

Results

Results of the study are reported in the following order: In the first section, the fluency rates in different speaker groups are observed; in the second section, features that influenced fluency ratings are reported; and in the third section, the agreement levels between the judges are presented.

Speech Fluency in the Three Groups

The fluency ratings for the aggregate data (all three speaker groups) ranged from very severely disordered speech fluency to a high, if not exceptional, degree of speech fluency. The mean fluency ratings also differed between the groups (see Table 5), and the ratings in all groups were distributed normally by the Kolmogrov-Smirnov Test (CG p= .149, CGA p= .303, CGB p= .958). The control group clearly differed in fluency from the two clinical groups for the Kruskal-Wallis Test (p=.001, significance level .05); however, the two clinical groups did not differ from each other (p= .871). Therefore, using the independent samples *t*-test, the two clinical groups were compared, and they did not differ from each other (*t*-test: p=.409).

The centroid linkage analysis revealed six natural clusters which we labeled by their degree of fluency disorder. These clusters differed significantly (p=.001) from each other, both according to the Kruskal-Wallis Test and Tukey's Honest Significant Difference Test. In Table 6, the clusters are presented in their order of severity. The first cluster – representing the highest degree of fluency – included four participants from the CG (Mean = 9.6). The second cluster represents average fluency (Mean = 8.2) and included 61 participants from the CG. The third cluster (Mean = 6.5) contains speakers from all three groups with a fluency below average. Of the 13 participants, five belonged to the CG, six to the CGA and two to the CGB. The fourth cluster is a group with speech fluency rated as moderately disordered (Mean = 3.4). This group contained seven speakers: two from CGA and five from CGB. The fifth cluster (Mean = 1.5) represents severely disordered fluency in two participants from CGB. The sixth cluster represented very severely disordered fluency (Mean = 0.4). It contained only three participants: two from the CGA and one from the CGB.

As an overall comparison, the mean fluency rating for typical speech was 8.2, for disordered speech it was 4.3 (speakers from the CGA with AOS, aphasia, dysarthria or cluttering) and for speech with neurogenic stuttering it was 3.4 (speakers from the CGB). The CG differed statistically from the clinical groups, but the clinical groups did not differ from each other. The data were divided into six natural clusters according to the severity of the fluency disorder. The most interesting finding was the third cluster "*below average*" which included participants from all three groups.

Features' Affecting Fluency Ratings

In the aggregate data (*N*=90) *speech rate* was the most commonly reported feature that impaired speech fluency, *something else* was the second common feature identified, and *pauses* was the third feature. There were 10 diagnosed stutterers in the participants of this study, but the SLPs considered *stuttering* as impairing speech fluency in 14 cases.

Of the participants in the CG, 55.7% were not observed to have any features in their speech that impaired the rating of fluency (see Table 7). When one feature was reported (in 24.3%) it was most commonly *speech rate*, and secondly, it was *something else*. Two features that were impairing speech fluency were observed in 11.4% of the participants in CG, and the most common combination was *speech rate* and *pauses*, secondly, it was *pauses* and *something else*. If three disturbing features were observed (5.7%), the most common combination was *speech rate, pauses,* and *something else*. Four features' impairing fluency were present in only 2.9% of the participants in CG, one combination including also *stuttering*. Six participants were judged to have language difficulties, and their mean fluency rating was 7.4 (6.9–7.9).

Participants in CGA had at least two disturbing features, but 60% showed 4–5 features impairing speech fluency (see Table 7), most commonly *pauses* and *language difficulties*. According to the centroid linkage analysis, these participants were divided into three different clusters: Very severely disordered fluency (*n*=2), moderately disordered fluency (*n*=2), and fluency below average (*n*=6) (Table 8). The SLPs perceived stuttering in three participants with AOS and aphasia. Language difficulties were observed to impair fluency in only those six participants who had aphasia. A total of five participants in the CGA had dysarthria, three of them solely. For these three, the most typically mentioned features' impairing fluency were *pauses+something else*, *speech rate* and *something else*, respectively.

In CGB, nine participants out of ten showed at least four features that were disturbing fluency (see Table 7). Stuttering was perceived to impair fluency in all 10 participants (see Table 9). According to the centroid linkage analysis, the participants in CGB were divided into four different clusters: Very severely disordered fluency (*n*=1), severely disordered fluency (*n*=2), moderately disordered fluency (*n*=5), and fluency below average (*n*=2). Four participants had been diagnosed only as neurogenic stutterers. Their mean fluency rating as a subgroup was 5.1 (3.3–6.8), and they all had moderate TBI. *Language difficulties* were not perceived solely for those participants with aphasia, contrary to what was the case in CGA. *Pauses* were perceived more often as impairing fluency in CGB than in CGA.

The different features seemed to correlate with fluency rates differently in various speaker groups (see Table 10). In the CG, a significant negative correlation (r= -.589, p= .044) between *pauses* and fluency ratings was found as well as a promising correlation with *something else* and fluency ratings (r= .506, p= .054). In CGA, no statistically significant correlations were found between the features and the fluency ratings (Table 10). Still, a suggestive nonsignificant correlation was noticed between fluency ratings and *language difficulties* (r= -.758, p= -.081). For the CGB, a negative and significant correlation between *pauses* and fluency was found (r= -.745, p= .013). Surprisingly, *stuttering* was not significantly correlated with fluency ratings (r= -.661, p= .079) in CGB (see Table 10). A regression-analysis (see Table 11) between fluency ratings and the given features for the aggregate data (*N*=90) revealed strong associations between fluency rating and *stuttering* (p= .001) and also fluency rating and *something else* (p= .014).

In summary, 55.7% of the participants in the CG were perceived not to have any feature that was impairing fluency. As the number of perceived features increased, the fluency rate decreased in all speaker groups. For the aggregate data (*N*=90), *speech rate* was the most commonly reported feature that was impairing speech fluency. The two most common features for the CGA were *language difficulties* and *pauses*. *Language difficulties* were considered to impair fluency only in participants with aphasia. Supposedly, for the CGB, the most common feature was *stuttering*. There were 10 diagnosed stutterers in the aggregate data, but the judges considered *stuttering* as impairing fluency in 14 cases. A significant correlation with fluency ratings and *pauses* was found in both CG and CGB. An encouraging correlation was also found for *something else* in CG. Regression analysis of the aggregate data revealed an association between fluency and *stuttering* and the feature *something else*.

Agreement of Perceptual Assessments

The SLPs' agreement on judging fluency and the features' impairing their ratings varied in the different groups. The total range of fluency ratings in CG was 3.6–12.0, in CGA, it was 0.0–9.3, and in CGB, it was 0.0–8.0. The homogeneity between SLPs ratings for speakers in different groups was examined using Levene's Test of Homogeneity of Variances, One-way ANOVA, Post Hoc Test (Bonferroni) and Fleiss' Kappa. These analyses showed that variances in the fluency ratings of the CG participants were not equal [F (9, 690) = 6.170, p= .001]. Multiple comparisons for the dependent variables in the CG (Bonferroni) also showed a significant level of disagreement between the SLPs. Lastly, the Fleiss' Kappa showed parallel results with low inter-rater agreement (IRR=66.1%). The low homogeneity between the raters manifested, for example, with two SLPs utilized a wider rating scale for speech fluency (SDs 1.13 and 1.23) compared to the rest of the SLPs (SD range=0.32–0.97).

For the two clinical groups, the variances in fluency ratings were equal (CGA [F (9, 90) = .655, p= .747; CGB [F (9, 90) = .763, p= .650), while the inter-rater agreement in fluency ratings between the SLPs was good: 90.4% and 89.2%, respectively. Also, the multiple comparison tests (Bonferroni) showed high agreement (p= 1.000 in all compared cases) in the ratings for both clinical groups.

To explore the agreement among the SLPs in assessing the features' impairing fluency, an *ad hoc* classification was needed based on the number of SLPs that were reporting the same observation in each participant's speech (*N*=90) (see Table 12). As shown in Table 12, the agreement level was highest in over half of the cases for *stuttering* (57.2%), and in about one- third of the cases for *pauses* and *language difficulties*. *Speech rate* and the feature *something else* were commonly observed, but seldom agreed on. The SLPs also seemed to weight different features slightly differently: Five SLPs most commonly mentioned *pauses* as impairing their rating, two recognized most typically *language difficulties* while *stuttering*, *speech rate*, and "*something else*" were regularly mentioned by one SLP, respectively. Because half of the judges most commonly rated pauses to be the distracting feature in fluency, it was also a statistically significant feature correlating with fluency ratings in both the control group and clinical group B, as mentioned above.

In conclusion, the SLPs highly agreed when rating fluency in CGA and CGB but the homogeneity between the raters was low for CG. The best agreement for the given features was gained in *stuttering*, where the consensus was more than 70% in over half of the cases (8/14; mean=63%). For the feature *pauses*, the mean agreement level was 45%, for *language difficulties* 44%, for *speech rate* 24% and for *something else* 22%.

Discussion

This study investigated the auditory-perceptual ratings of SLPs to determine fluency in speech samples acquired from TBI patients and neurologically intact control speakers. The findings indicate that the three speaker groups differed in their fluency ratings and the features contributing to the SLPs' evaluation. *Speech rate* was a commonly reported feature contributing to fluency, but it did not separate fluent and non-fluent speakers. The feature *pauses* correlated negatively with the fluency ratings in both CG and CGB, and – interestingly –the feature *something else* impaired fluency in the aggregate data according to the regression analysis. *Stuttering*, however, was the feature with the highest agreement among judges, and its impact on fluency was also the largest statistically. *Stuttering* was also perceived in speakers with no diagnosis of stuttering. The control group was assessed statistically to be more fluent than the two clinical groups, but the two clinical groups did not differ from each other. Because disorders like AOS and neurogenic stuttering can be difficult to discriminate perceptually, it is important to increase our awareness of the criteria that SLPs apply when assessing fluency in order to develop sustainable methodologies for APJ studies.

The difference between the control group and the two clinical groups was obvious according to the centroid linkage analysis. Most speakers from the clinical groups were clustered into the lower (more severe) end of the fluency continuum, while most participants from the control group represented the higher (less severe) end. Most importantly, however, there were participants from all three speaker groups who were clustered into a borderline group where the fluency ratings were less than the average, but higher than in the clearly non-fluent groups. This finding encourages us to develop more reliable clinical diagnostics and effective intervention methods for communication disorders after TBI.

Methodological Considerations

Fluency of speech has been an interest in many fields, including linguistics, phonetics, speech-language pathology, and speech technologies such as automatic speech recognition. Still we lack the data to establish typical fluency of speech and effectively delineate it from atypical fluency, as many of the disfluencies in typical speech are similar to non-fluencies in disordered speech. As speech fluency can be impaired due to different etiologies, assessing fluency in clients with mixed pathologies is especially demanding. In this study, samples from those participants with and without TBI were perceptually evaluated by experienced SLPs who were instructed to consider fluency as a multidimensional phenomenon, not just a speech motor problem.

The need for a multidimensional perspective on fluency rose from the detection of a conceptual bias in the vast literature where fluency is most typically connected to stuttering (e.g., Campbell & Hill, 1994; Guitar, 2006; Lundgren, Helm-Estabrooks, & Klein, 2010; Yaruss, 1997). As fluency may be impaired in various communication disorders (e.g., Bailey, Blomgren, DeLong, Berggren, & Wambaugh, 2017; Harmon *et al*., 2015; Kent, 2000; St. Louis, Myers, Faragasso, Townsend, & Gallaher, 2004; Park *et al.*, 2011), we adopted a more holistic view of speech fluency and examined it in speakers who were suffering from mixed communication disorders, in addition to typical speakers. Because the data to be analyzed came from a relatively large sample of 90 speakers, we wanted to keep the experiment feasible, and therefore, we chose a modest number of five features for the SLPs to consider. From the mainstream concepts related to fluency, we chose *stuttering*, *speech rate*, and *pauses* to represent those variables that may influence the perception of fluency. We also included in the definition *language difficulties*, that is, word-finding difficulties, interrupted words and phrases, paraphasias, neologisms, and agrammatism as linguistic-cognitive processing problems that might impair speech-motor performance (e.g., Gordon, 1998; McNeil & Copland, 2011). The fifth dimension of fluency in this study was labeled *something else*. This collective and deliberately nonspecific concept was included to determine whether the SLP's judgments were influenced by something other than the specified features (Mossholder, Settoon, Harris, & Armenakis, 1995). In the informal debriefing after the evaluation experiment, the judges' voluntarily commented that voice quality was one of the categories that was missed in the listing of explicit features. Based on clinical experience and informal debriefing after the perceptual evaluation experiment, the feature "*something else*" was meant to cover, for example, voice quality, prosody of speech, and stylistic aspects of speech production, such as the speakers' narrative style, and quality of vocabulary being used. Based on the findings of Cucchiarini, Strik, and Boves (2000) and Derwing, Rossiter, Munro, and Thomson (2004), no definitions for the concept *something else* were given to the judges. Both Cucchiarini *et al*. and Derwing *et al*. observed that without these specific instructions, the judges used their own definitions when judging fluency, while specifically instructed judges did merely what they were asked to do. Because *something else* significantly impaired fluency ratings in the aggregate data, further research is needed to explore *what else* may actually impair the ratings of fluency in addition to the features already defined more explicitly. One direction of study could, therefore, include acoustic analyses of, for example, voice quality using Praat-software (Boersma & Weenink, 2005), prosody and intonation patterns using ProsodyPro-software (Xu, 2005–

2010), or narrative styles and vocabulary using qualitative methods to find out if differences from typical are found in speech samples impaired by the feature *something else*. Another direction would be to ask the SLPs for insights on what the *something else* impairing fluency includes in addition to the given features.

The task also needs to be discussed. In this study, speech fluency was evaluated on the basis of narrative discourse samples. The story generation task was expected to guide the speakers to maintain the same topic, contrary to spontaneous speech task where speakers, especially those with TBI, may have difficulties in maintaining the storyline (Biddle, McCabe, & Bliss, 1996). Also, spontaneous speech tasks can possibly evoke personal details, which then distract listeners' attention and skew their perception about the speech itself and its fluency. Sentence repetition, however, was not considered relevant for fluency rating, as we wanted to make evaluations from natural-like speech, and therefore, the narrative discourse represented semi-spontaneous speech and enabled these comparisons.

A 120 mm Visual Analog Scale (VAS) was chosen as the instrument of evaluation. The VAS allowed us to analyze fluency in millimeters and utilize parametric testing (Cannito *et al.*, 1997). Unlike interval scales, we found VAS to be more genuine in accommodating different variations in ratings (Kempster *et al*., 2009). Typically, researchers use a 100 mm long VAS, but in this study we decided to extend it by 20 mm to give the judges the option to rate the "high or exceptional degree of fluency". We used anchors set at 0 mm, 40 mm, 80 mm and 120 mm to represent severely non-fluent, moderately non-fluent, typical and very fluent speech, respectively. Audio-samples corresponding to these anchors were played for the judges to calibrate their ratings. However, with anchors located as such, caution must be applied, as the anchors at the both ends of the scale did not provide extra room, and the "end effect", reported by Kempster *et al.* (2009), may, therefore, have influenced the results.

Typical fluency

Speech-language pathologists often make judgments about the speech fluency of clients with various pathologies (Park *et al.*, 2011) but they are rarely asked to assess the fluency of typical speakers' utterances. The 70 healthy adults in this study represented such typical speakers. Their ages were between 18 and 89 years, and their educational backgrounds varied from elementary school to holding a Master's degree. The participants were geographically from different areas of Finland, so the varying dialects created natural diversity in the data. The mean fluency rating for the control group (CG) was 8.2 (range of the means 6.7–9.9 and range of given values 3.6–-12.0), and the data was normally distributed. Of the 70 participants, 61 belonged to a cluster labeled "average" based on a centroid linkage analysis. The fluency rating for this group was also 8.2, but the range was narrower (7.5–-9.2) compared to the total CG data. The nine other speakers from the CG belonged either to the cluster labeled "exceptionally high fluency" (*n*=4) or the borderline cluster with fluency ratings "under the average" (*n*=5).

No agreement on the fluency ratings for the CG was gained (IRR=66.1%). The One-way ANOVA, Bonferroni Post Hoc Test and Fleiss' Kappa all together indicated that the perceptual assessment of typical fluency is a difficult task for even expert judges. The range of assigned ratings for typical speakers was wide, and individual SLPs seemed to utilize different scales in their ratings. For example, the highest intrajudge variation ranged from severely disordered fluency (3.6) to a high degree of fluency (10.7), in samples of typical speech, while the smallest range fell between 7.2 and 9.2. The low fluency ratings given by some of the SLPs may have reflected their strict internal standards for typical and normal speech, even more strict compared to the standards for disordered speech as was also shown in a study by Dagenais and colleagues (1999).

As to the different features that impaired the fluency ratings, this study suggests that *speech rate* is not a significant factor in the auditory-perceptual assessment of typical fluency although it is considered of primary importance when differentiating fluent and non-fluent speakers (Goodglass &

Kaplan, 1983; Kerschensteiner, Poeck, & Brunner, 1972). Instead, it was *pauses* that significantly correlated with fluency in typical speakers. Of course, pauses are strongly related to speech rate and articulation rate, but they also can result from features of typical language processing, such as word search and sentence formulation. Therefore, it would be interesting to investigate the quality of pauses in those speech samples that the SLPs rated to be less fluent due to pauses.

Deviant Fluency Following TBI

To examine how fluency is perceived in patients with multiple communication disorders, we recruited participants with TBI to participate in this study. Some limitations of this study concerning disordered fluency, however, need to be kept in mind. First, the findings of this study are based on small sample sizes with only 10 participants in each clinical group. Secondly, the severity of TBI varied from mild to very severe, and various communication disorders in different combinations impaired the SLPs' perceptual evaluation of fluency. Therefore, further research is needed to investigate the perceptual assessment with larger sample sizes and identify distinct disorders, especially between speakers with only neurogenic stuttering and speakers with only AOS – as far as it is possible in natural clinical groups of speakers suffering from TBI or stroke.

Fluency in Speakers with Aphasia, AOS, and Dysarthria Following TBI. The mean fluency rating for participants in the CGA was 4.3 (range of mean values 0.2–7.2; range of given values 0.0–9.3). The SLPs' ratings were very consistent according to the One-way ANOVA, Bonferroni Post Hoc and Fleiss' kappa tests (IRR=90.4%). The most common features that impaired the ratings were *pauses* and *language difficulties* (see Table 8) but the correlations were not statistically significant (see Table 10). According to the centroid linkage analysis, the participants in CGA represented different degrees of non-fluency: Very severe (*n*=2) and moderate non-fluency (*n*=2), and below average fluency (*n*=6). The most severely non-fluent speakers suffered from both AOS and aphasia, while the moderately non-fluent speakers had aphasia with AOS or dysarthria – and many SLP's heard the three speakers with AOS also stuttering.

The relation of AOS and stuttering has been for long discussed in the literature (Bailey *et al*., 2017; Johns & Darley, 1970). Using the term "stuttering" in relation to AOS is, however, controversial. For example, Bailey *et al.* (2017) reported stuttering-like disfluencies (repetitions, prolongations, and blocks; discussed in detail below) in samples from speakers with AOS, but Van Borsel (2014) has advised not to call disfluencies in people with AOS actual stuttering. Nevertheless, the disfluencies in AOS are mainly considered to have a motor origin. Because AOS is commonly co-morbid with aphasia, as in this current study, certain researchers (e.g. Bailey *et al*., 2017; Jacks and Haley. 2015) have discussed the possibility of an overlap of certain disfluencies, such as revisions, that may originate from both motor and linguistic deviations. In a recent study by Bailey *et et al.* (2017), stuttering was reported in 20 speakers with AOS and aphasia. The frequency of stuttering-like disfluencies in a word repetition task ranged from 2.6% to 5.4% and in a connected speech task from 4.4% to 5.2% (percentage of words with stuttering-like disfluencies). Syllable repetitions occurred more frequently than prolongations and blocks did. Thus, if stuttering-like behaviors can be measured by apraxic and aphasic speech, disfluencies in AOS with aphasia can also be perceived as stuttering, as was obvious in this study. Therefore, this study has raised important questions in a reference to other studies in terms of the quality and quantity of acquired stuttering and the relationship between stuttering and AOS.

The role of dysarthria in fluency assessment is also of interest. Three participants in the CGA had only dysarthria and were rated as being below average in fluency according to centroid linkage analysis. The SLP's reported that *pauses, speech rate,* and *something else* impaired the ratings of these speakers. This result is understandable as pauses contribute to speech rate and because dysarthria typically is manifested as slow speech rate in addition to weakness, rigidity, or dyscoordination of speech movements (Kent, 2000). The feature *something else* raises the question of the role of speech proficiency and expression style. In the dysarthria research, the terms commonly used include *speech intelligibility* (see Weismer & Martin, 1992) and *naturalness of speech* (see

Bellaire, Yorkston, & Beukelman, 1986; Yorkston, Hammen, Beukelman, & Traynor, 1990). Thus, in the context of dysarthria, one might speculate that the feature *something else* may relate to speech that sounds unnatural or unintelligible.

Interestingly, the SLPs perceived *language difficulties* only in those participants with aphasia in the CGA. This finding might indicate that aphasic features – word-finding difficulties and sentence production problems – are easily identifiable in semi-spontaneous narrative samples. The negative correlation between fluency and *language difficulties* was not significant in this group, but it was suggestive. However, the specific definitions of language difficulties given before the listening experiment were more comprehensive than the definitions given for other features, possibly impairing the SLPs' ratings. This limitation needs to be kept in mind when comparing the role of different features in any fluency assessment. Further research should be undertaken to investigate the disfluencies in speech samples quantitatively and compare these findings with the SLPs' perceptions of fluency and features that impaired their ratings.

Six participants in the CGA belonged in the cluster "below average fluency", that is, the so-called borderline group between non-fluent and fluent speakers. This finding will be examined in the latter part of this discussion.

Fluency in Neurogenic Stuttering. The clinical group B (CGB), which represented speakers with neurogenic stuttering, did not differ statistically in fluency from the CGA, but it had a lower mean (3.4). In addition, according to the centroid linkage analysis, more participants in CGB were clustered into the severely disordered (n=3) and moderately disordered (n=5) groups than were in the CGA, and only two participants were evaluated as "below average" in fluency. It thus seems that neurogenic stuttering has a specific impact on fluency, and the SLPs also considered stuttering as the most common feature impairing their fluency ratings of this group of participants. However, there were only four participants diagnosed solely with neurogenic stuttering. The other six represented speakers with other neurogenic communication disorders in addition to neurogenic stuttering, in this case aphasia or dysarthria, as has also been reported in early studies by Canter (1971), and Mazzucchi, Moretti, Carpeggiani Parma, and Paini (1981), respectively. AOS can also co-occur with neurogenic stuttering (Rosenbek, 1980), but in the CGB, no participant was diagnosed as having AOS. Instead, three participants with AOS in the CGA were perceived as stutterers, as noted earlier.

The relationship of stuttering and aphasia also seemed unclear (Van Borsel, 2014). Three participants in this study had both aphasia and neurogenic stuttering. In two cases, it was *language difficulties* that impaired most the SLPs' evaluation of fluency in addition to *stuttering*, but in one case the major reasons for the low rating were *stuttering* and *pauses*. This latter participant had conduction aphasia with frequent phonemic paraphasias and attempts to self-correct the output, which seems to have been perceived as stuttering. Here again, AOS may have played a role given that AOS has been connected to conduction aphasia (afferent motor aphasia; Luria, 1973). In another case, transcortical motor aphasia (dynamic aphasia; Luria, 1973) heavily impaired verbal planning in the narrative task (Costello & Warrington, 1989) with wrong starts and pauses that most probably sounded like stuttering. These mixed pathologies, like a combination of aphasia and neurogenic stuttering, may be the interpretative factor that produced the low correlation between perceived feature *stuttering* and fluency rating in CGB.

All of these combinations of disorders co-occurring with stuttering raise the question of whether neurogenic stuttering is a distinct disorder or an epiphenomenon of other communication disorders (Lundgren *et al.*, 2010; Lundie, Erasmus, Zsilavecz, & van der Linde, 2014). Neurogenic stuttering is a broad term, which in the simplest terms refers to stuttering after a head trauma (Van Borsel, 2014). It seems difficult to distinguish stuttering from AOS and certain aphasic features, and it seems also hard to detect it from verbal output as was shown in an experiment by Van Borsel and Taillieu (2001) where expert evaluators were asked to decide if speakers were developmental stutterers or neurogenic stutterers. Almost one- fourth of the judges placed patients in wrong diagnostic groups. In our study, neurogenic stuttering was originally diagnosed in 10 participants after TBI, while stuttering was perceived in 14 participants, one of which represented the control group without any observed communication disorder. Thus, disfluencies of various origins may well be perceived as stuttering.

The Borderline Between Typical and Atypical Fluency

Perhaps the most interesting finding in this study was the detection of a borderline group situated between average fluency and moderate non-fluency speaker groups. The centroid linkage analysis clustered 13 participants into the below-average group with the representatives of all three speaker groups: Five from the control group of typical speakers (CG), six from the clinical group A (CGA), and two from the clinical group B (CGB). Of the six participants from the CGA, four had dysarthria and two had aphasia after a mild or moderate brain injury while the two speakers from the CGB had only neurogenic stuttering as a consequence of a moderately severe brain injury.

This borderline group statistically differed from the other clusters. The SLPs ratings were impaired by various factors, including *speech rate* and *pauses*, individually or in combination, as well as *stuttering* and *language difficulties*. However, the most frequently mentioned factor was *something else*. Thus, the SLPs perceived various types of disfluencies that were obviously mild, but distinct from typical fluency. The judges likely selected the comprehensive category *something else* when the given features were not sufficient to describe a perceived departure from typical fluency. Further research using open-ended questions or the Delphi technique would help reveal the judges' conception of this feature.

These findings indicate that fluency is not an on-off phenomenon, but rather it is a continuum. The borderline group having both dysfluent and fluent speakers contributes to our knowledge of the normal variations in speech. One direction to take to grasp at these “grey areas” would be to investigate both the quantitative features of speech disfluencies and acoustical parameters to find the crucial markers that separate these speakers from speakers with fluency rated as average.

Implications for Future Research

Returning to the introduction, fluency can be triangulated using three methods: The clinician's assessment, the client's self-assessment, and instrumental measures, each of which has advantages and disadvantages as a tool for fluency evaluation. Taken together, these three approaches have the potential to yield valid, reliable, and sensitive assessments of fluency. Based on the findings of this study, the possible directions for future research are discussed as follows.

The data in the present study show that levels of agreement among the SLPs varied for both VAS ratings and the features identified as contributing to the fluency judgment. For the latter, agreement decreased in the following order: Stuttering, pauses and language difficulties, speech rate, and something else. Future research is needed to gain an understanding of the sources of such variance when identifying the features that are critical for perceptual evaluation, as such information could be helpful in developing training materials and anchors to increase the reliability of the assessment.

Another direction for future research is to identify instrumental measures that complement or extend perceptual ratings. For example, instrumental measures of pause, including some measures that are automated, have been indicated in studies of stuttering (Love & Jeffress, 1971), second-language learning (Cucchiarini *et al.*, 2000), dysarthria (Green, Beukelman, & Ball, 2004; Rosen *et al.*, 2010), primary progressive aphasia (Rohrer, Rossor, & Warren, 2010), and childhood apraxia of speech (Shriberg, Green, Campbell, McSweeny, & Scheer, 2003). Instrumental measures of speaking rate have also been developed and can be performed automatically or semi-automatically (Cucchiarini *et al.*, 2000; De Jong & Wempe, 2009). These studies indicate the potential for automatic instrumental measures of pause and speaking rate that can complement the perceptual judgments by clinicians or the self-assessments by clients.

On the theoretical level, we propose a novel approach that may be worth exploring. By definition, a fluency disorder is a condition wherein the flow of natural speech is interrupted. Because fluency is compromised in aphasia, AOS, dysarthria or neurogenic stuttering – all of which can be present simultaneously – and as the origin of various disfluencies is not always clear (motor or linguistic), we suggest that disfluencies should be perceived as interruptions in speech flow. Therefore, instead of inspecting stuttering-like disfluencies (repetitions, blocks, and prolongations) or other-disfluencies (revisions, hesitations, incomplete words or phrases) we should concentrate on the quality of these interruptions – which are secondarily manifested as audible or inaudible breaks in the speech flow (see Figure 1). The basic idea of this proposition is that fluency of speech is disturbed either because the speech flow slows down, becomes fragmentary or because the speech flow breaks entirely. *Prolongation segments* include prolongations (e.g., *hooorse*) and hesitations (e.g., *mmm*), that are audible interruptions and tend to slow the speech flow. *Dysrhythmic segments* include repetitions (e.g., *do-do-dog*) and revisions (e.g., *ble-bleu-blue*), that are perceived both as audible interruptions affecting the speech rhythm, and thereby fragmenting the flow of speech. Fluency can be severed in two major ways. First, s*ilent segments* include blocks (e.g., *op'(1.2)portunity*), incomplete utterances (*I we..did something*) and inaudible hesitations, which cause silent periods and cut the speech flow. Second, w*ord-finding segments*, such as word-finding phrases (e.g., *oh well you know what they call for things like apple and things)*, paraphasias (e.g., *tevilision* for television) and neologisms (e.g., *pinwad* for light), also sever the speech flow with audible expression, but the information does not necessarily change. This suggested classification does not categorize interruptions into typical or atypical nor does it label them under a specific communication disorder. Rather, the classification is based on observations of behavior that can be applied broadly to the speech samples produced by individuals with varying degrees of communication ability.

In clinical practice, SLPs should be aware of the internal and external standards that affect fluency assessment and the evaluation of treatment outcomes. One way to modify the external standards would be to consider fluency a continuum and investigate interruptions of speech flow more functionally (see Figure 1). In addition, it would be worth studying therapy outcomes if the aim of speech-language therapy interventions for people with neurogenic communication disorders was not set on the average level of fluency, but slightly below it as what is considered typical or functional speech may also sound somewhat disfluent. Lastly, in clinical practice, the client's self-evaluation of the experienced communication handicap should also be taken into account.

Conclusions

The results of this study show that the 10 SLPs perceptually differentiated most of the typical speakers from most of those with disordered speech. Most interestingly, however, this study identified a small group of speakers who were evaluated as being less fluent than the average, but more fluent than what was considered to be moderately deviant. This borderline group included typical speakers in addition to speakers with communication disorders. In the aggregate data, the most significant features that impaired fluency ratings were *stuttering* and *something else*, although *pauses* as well as *language difficulties* – in different combinations – were also found to impair the ratings. The results of this study indicate the need for further studies on the nature of borderline fluency and different disfluencies, with continuing discussion of the relationship between perceived fluency and traditional quantitative measures. Fluency is a multifaceted phenomenon, and we suggest taking a holistic approach to fluency and its deviations, an approach that can take into account both the motoric and linguistic aspects of fluency, but also any variations in speech styles – the affective and metalinguistic components of fluency.

Acknowledgments

This research was supported by a grant from the Emil Aaltonen Foundation (25.4.2016 / 20 025e). We thank all the participants who volunteered their time for recording the speech samples and for acting as evaluators in the study. We would also like to gratefully acknowledge the statistician, Juho Luoma, for his advice on the statistical designs of this study.

References

Ambrose, N., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language and Hearing Research, 42,* 895-909.

Bailey, D., Blombgren, M., DeLong, C., Berggren, K., & Wambaugh, J. (2017). Quantification and Systematic Characterization of Stuttering-Like Disfluencies in Acquired Apraxia of Speech. *American Journal of Speech-Language Pathology, 26,* 641-648.

Basso, A. (2003). *Aphasia and its therapy.* Oxford: Oxford University Press.

Bellaire, K., Yorkston, K., & Beukelman, D. (1986). Modification of breath patterning to increase naturalness of a mildly dysarthric speaker. *Journal of Communication Disorders, 19,* 271-280.

Benson, D. (1967). Fluency in aphasia: Correlation with radioactive scan localization. *Cortex, 3,* 373-394.

Biddle, K., McCabe, A., & Bliss, L. (1996). Narrative skills following traumatic brain injury in children and adults. *Journal of Communication Disorders, 29(6),* 447-468.

Boersma, P. & Weenik, D. (2005). *Praat: Doing phonetics by computer.* Available from: http://www.fon.hum.uva.nl/praat/.

Bortfeld, H., Leon, S., Bloom, J., Schober, M., & Brennan, S. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech, 44(2),* 123-147.

Bothe, A. (2008). Identification of children's stuttered and nonstuttered speech by highly experienced judges: Binary judgements and comparisons with disfluency-types definitions. *Journal of Speech, Language and Hearing Research, 51,* 867-878.

Brainbridge, L., Stavros, C., Ebrahimian, M., Wang, Y., & Ingham, R. J. (2015). The efficacy of stuttering measurement training: evaluating two training programs. *Journal of Speech, Language and Hearing Research, 58,* 278-286.

Brown, S., Ingham, R.J., Ingham, J.C., Laird, A., & Fox, P. (2005). Stuttered and Fluent Speech Production: An ALE Meta-Analysis of Functional Neuroimaging Studies. *Human Brain Mapping, 25,* 105-117.

Brundage, S., Bothe, A., Lengeling, A., & Evans, J. (2006). Comparing judgments of stuttering made by students, clinicians and highly experienced judges. *Journal of Fluency Disorders, 31,* 271-283.

Campbell, J., & Hill, D. (1994). *Systematic disfluency analysis.* Evanston, IL: Northwestern University.

Canter, G. (1971). Observations on neurogenic stuttering: A contribution to differential diagnosis. *British Journal of Communication Disorders, 6,* 139-143.

Cannito, M., Burch, A., Watts, C., Rappold, P., Hood, S., & Sherrard, K. (1997). Disfluency in Spasmodic Dysphonia: A Multivariate Analysis. *Journal of Speech, Language and Hearing Research, 40,* 627-641.

Chan, K., & Yiu, E. (2002). The effect of anchors and training on the reliability of perceptual voice evaluation. *Journal of Speech, Language and Hearing Research, 45,* 111-126.

Conture, E. G. (1990). Childhood stuttering: What is it and who does it? *ASHA Reports, 18*, 2-14.

Cordes, A. K., & Ingham, R. J. (1995). Judgements of stuttered and nonstuttered intervals by recognized authorities in stuttering research. *Journal of Speech and Hearing Research, 38,* 33-41.

Costello. A. de L., & Warrington, E. (1989). Dynamic aphasia. The selective impairment of verbal planning. *Cortex, 25*, 103-114.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of Acoustical Society of America, 107 (2),* 989-999.

Dagenais, P., Watts, C., Turnage, L., & Kennedy, S. (1999). Intelligibility and acceptability of moderately dysarthric speech by three different types of listeners. *Journal of Medical Speech-Language Pathology, 7,* 91-96.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41(2),* 385-390.

Deger, K., & Ziegler, W. (2002). Speech motor programming in apraxia of speech. *Journal of Phonetics, 30(3),* 321-335.

Derwing, T., Rossiter, M., Munro, M., & Thomson, R. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54,* 655-679.

Duffy, J. R. (2005). *Motor speech disorders.* St. Louis, Missouri: Elsevier-Mosby.

Eadie, T., Stroka, A., Wright, D. R., & Merati, A. (2011). Does knowledge of medical diagnosis bias auditory-perceptual judgments of dysphonia? *Journal of Voice, 25: 4,* 420-429

Feyereisen, P., Verbeke-Dewitte, C., & Seron, X. (1986). On fluency measures in aphasic speech. *Journal of Clinical and Experimental Neuropsychology, 8(4),* 393-404.

Finn, P. Ingham, R. Ambrose, N., & Yairi, E. (1997). Children recovered from stuttering without formal treatment: Perceptual assessment of speech normalcy. *Journal of Speech, Language & Hearing Research, 40(4),*867-876.

Fox Tree, J. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34,* 709-738.

Gerratt, B. R., Till, J. A., Rosenbek, J. C., Wertz, R. T., & Boysen, A. E. (1991). Use and perceived value of perceptual and instrumental measures in dysarthria management. In C. A. Moore, K. M. Yorkston, & D. R. Beukelman (Ed.), *Dysarthria and Apraxia of Speech:Perspectives on Management* (pp. 77-93). Baltimore: Paul H. Brookes.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders.* Philadelphia, PA: Lea & Febiger.

Gordon, J. (1998). The fluency dimension in aphasia. *Aphasiology, 12(7),* 673-688.

Gordon, J. (2006). A quantitative production analysis of picture description. *Aphasiology, 20(2),* 188-204.

Green, J. R., Beukelman, D. R., & Ball, L. J. (2004). Algorithm estimation of pauses in extended speech samples of dysarthric and typical speech. *Journal of Medical Speech-Language Pathology, 12(4),* 149-154.

Gregory, H. (1993). A clinician's perspective: Comment of identification of stuttering, prevention and early intervention. *Journal of Fluency Disorders, 18,* 389-402.

Guenther, F.H., Gosh, G. H., Nieto-Castanon, A., & Tourville, J. A. (2006). A neural model of speech production. In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes and techniques.* UK: Psychology Press.

Guitar, B. (2006). *Stuttering. An integrated approach to its nature and treatment.* Baltimore: Lippincott Williams & Wilkins.

Haley, K.L., Jacks, A., de Riesthal, M., Abou-Khalil, R., & Roth, H.L. (2012). Toward a quantitative basis for assessment and diagnosis of apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55,* 1502-1517

Harmon, T.G., Jacks, A., Haley, K.L., & Faldowski, R.A. (2015). Listener perceptions of simulated fluent speech in non-fluent aphasia. *Aphasiology, 30(8),* 922-942.

Hoffman, L., Wilson, L., Copley, A., Hewat, S., & Lim, V. (2014). The reliability of a severity rating scale to measure stuttering in an unfamiliar language. *International Journal of Speech and Language Pathology*, 16(3), 317-326.

Jacks, A., & Haley, K. (2015). Auditory masking effects on speech fluency in apraxia of speech and aphasia: Comparison to altered auditory feedback. *Journal of Speech, Language, and Hearing Research, 58,* 1670-1686.

Johns, D., & Darley, F. (1970). Phonemic variability in apraxia of speech. *Journal of Speech and Hearing Research, 13,* 556-583.

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. R. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol. *American Journal of Speech-Language Pathology, 18,* 124-132.

Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5(3),* 7-23.

Kent, R. D. (2000). Research on speech motor control and its disorders: a review and prospective. *Journal of Communication Disorders, 33,* 391-428.

Kerschensteiner, M., Poeck, K., & Brunner, E. (1972). The fluency-non fluency dimension in the classification of aphasic speech. *Cortex, 8(2),* 233-247.

Korpijaakko-Huuhka, A-M., & Aulanko, R. (1994). Auditory and acoustic analysis of prosody in the clinical evaluation of narrative speech. In R. Aulanko & A-M. Korpijaakko-Huuhka (Ed.) *Proceedings of the third Congress of the International Clinical Phonetics and Linguistic Association* (p. 91-98) Helsinki: Publications of the Department of Phonetics. University of Helsinki.

Korpijaakko-Huuhka, A-M., & Lind, M. (2012). The impact of aphasia on textual coherence: Evidence from two typologically different languages. *Journal of Interactional Research in Communication Disorders, 3.1,* 47-70.

Kreiman, J., Gerratt, B., and Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research, 33,* 103-115.

Kreiman, J., Gerratt, B.R, Kempster, G. B, Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research, 36,* 21-40.

Kreindler, A., Mihailescu, L., & Fradis, A. (1980). Speech fluency in aphasics. *Brain and language,* *9(2)*, 199-205.

Krishnan, G., & Tiwari, S. (2013). Differential diagnosis in developmental and acquired neurogenic stuttering: Do fluency-enhancing conditions dissociate the two? *Journal of Neurolinguistics, 26*(2), 252-257.

Laine, M., Koivuselkä-Sallinen, P., Hänninen, R., & Niemi, J. (1997). *Bostonin nimentätesti.* Helsinki: Psykologien kustannus.

Lehtonen, J. (1976). On the problems of measuring fluency. *AFinLAn vuosikirja*, 53-68.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 3,* 387-417.

Love, L., & Jeffress, L. (1971). Identification of brief pauses in the fluent speech of stutterers and nonstutterers. *Journal of Speech and Hearing Research, 14(2)*, 229-240.

Lundgren, K., Helm-Estabrooks, N., & Klein, R. (2010). Stuttering Following Acquired Brain Damage: A Review of the Literature. *Journal of Neurolinguistics, 23(5),* 447-454.

Lundie, M., Erasmus, Z., Zsilavecz, U., & van der Linde, J. (2014). Compilation of a preliminary checklist for the differential diagnosis of neurogenic stuttering. *South African Journal of Communication Disorders, 61(1),* 1-10.

Luria, A. R. (1973). *The working brain.* New York: Basic Books.

Mazzucchi, A., Moretti, G., Carpeggiani, P., Parma, M., & Paini, P. (1981). Clinical observations on acquired stuttering. *British Journal of Disorders of Communication, 16(1),* 19-30.

McCauley, R. (1989). Measurement as a dangerous activity. *Journal of Speech-Language Pathology and Audiology, 13,* 29-32.

McDonald, S., Togher, L., & Code, C. (1999). Communication problems following traumatic brain injury. In S. McDonald, L. Togher & C. Code (Ed) *Communication problems following traumatic brain injury* (p. 1-18). Hove: Psychology Press.

McNeil, M. & Copland, D. (2011). Aphasia theory, models and classification.

In L. L. LaPointe (Ed.),*Handbook of aphasia and brain-based cognitive-language disorders*. New York: Thieme.

Meyers, S. (1986). Qualitative and quantitative differences and patterns of variability in disfluencies emitted by preschool stutterers and nonstutterers during dyadic conversations. *Journal of Fluency Disorders, 11*, 293-306.

Meyers, F., Bakker, K., St. Louis, K., & Raphael, L. (2012). Disfluencies in cluttered speech. *Journal of Fluency Disorders, 37*, 9-19.

Mossholder, K. W., Settoon, R. P., Harris, S. G., & Armenakis, A. A. (1995). Measuring emotion in open-ended survey responses: An application of textual data analysis. *Journal of Management*, 21(2), 335-355.

Nicolosi, L., Harryman, E., & Kresheck, J. (1987). *Terminology of communication disorders: Speech-language-hearing* (2nd ed.). Baltimore: Williams & Wilkins.

Park, H., Rogalski, Y., Rodriguez, A.D., Zlatar, Z., Benjamin, M., Harnish, S., Bennett, J., Rosenbek, J.C., Crosson, B., & Reilly, J. (2011). Perceptual cues used by listeners to discriminate fluent from non-fluent narrative discourse. *Aphasiology, 25*, 998-1015.

Pietilä, M-L., Lehtihalmes, M., Klippi, A., & Lempinen, M. (2005). *Western Aphasia Battery. Käsikirja*. Helsinki: Psykologien kustannus.

Rohrer, J., Rossor, M., & Warren, J. (2010). Syndromes of non-fluent primary progressive aphasia. A clinical and neurolinguistic analysis. *Neurology*, *75*(7), 603-610.

Rosenbek, J. (1980). Apraxia of speech- relationship to stuttering. *Journal of Fluency Disorders,5,* 233-235.

Quesal, R. W. (1988). Inexact use of “disfluency” and “dysfluency” in stuttering research. *Journal of Speech and Hearing Disorders, 53,* 349-350.

Rosen, K., Murdoch, B., Folker, J., Vogel, A., Cahill, L., Delatycki, M., & Corben, L. (2010).

Automatic method of pause measurement for normal and dysarthric speech. *Clinical Linguistics & Phonetics,24* (2), 141-154.

Segalowitz, N. (2010). *Cognitive bases of second language fluency.* New York: Routledge.

Shriberg, L. D., Green, J. R., Campbell, T. F., Mcsweeny, J. L., & Scheer, A. R. (2003). A diagnostic marker for childhood apraxia of speech: The coefficient of variation ratio. *Clinical Linguistics & Phonetics,17* (7), 575-595.

St. Louis, K., Myers, F., Faragasso, K., Townsend, P., & Gallaher, A. (2004). Perceptual aspects of cluttered speech. *Journal of Fluency Disorderds, 29,* 213-235.

Susca, M., & Healey, E. (2001). Perceptions of simulated stuttering and fluency. *Journal of Speech, Language and Hearing Research, 44(1),* 61-72.

Tani, T., & Sakai, Y. (2011). Analysis of five cases with neurogenic stuttering following brain injury in the basal ganglia. *Journal of Fluency Disorders, 36,* 1-16.

Van Borsel, J. (1997). Neurogenic stuttering: A review. *Journal of Clinical Speech and Language Studies, 7,* 16-33.

Van Borsel, J. (2014). Acquired stuttering: A note on terminology. *Journal of Neurolinguistics, 27,* 41-49.

Van Borsel, J., & Taillieu, C. (2001). Neurogenic stuttering versus developmental stuttering. An observer judgement study. *Journal of Communication Disorders, 34,* 385-395.

Van Lieshout, P., Bose, A., Square, P., & Steele, C. (2007). Speech motor control in fluent and dysfluent speech production of an individual with apraxia of speech and Broca's aphasia. *Clinical Linguistics & Phonetics, 21(3),* 159-188.

Weismer, G., & Martin, R. (1992). Acoustic and perceptual approaches to the study of intelligibility In R. D. Kent (Ed. p. 68-118), *Intelligibility in speech disorders: Theory, measurement and management.* Amsterdam, The Netherlands: John Benjamins.

Xu, Y. (2005-2010). *ProsodyPro.praat.* Available from: http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/

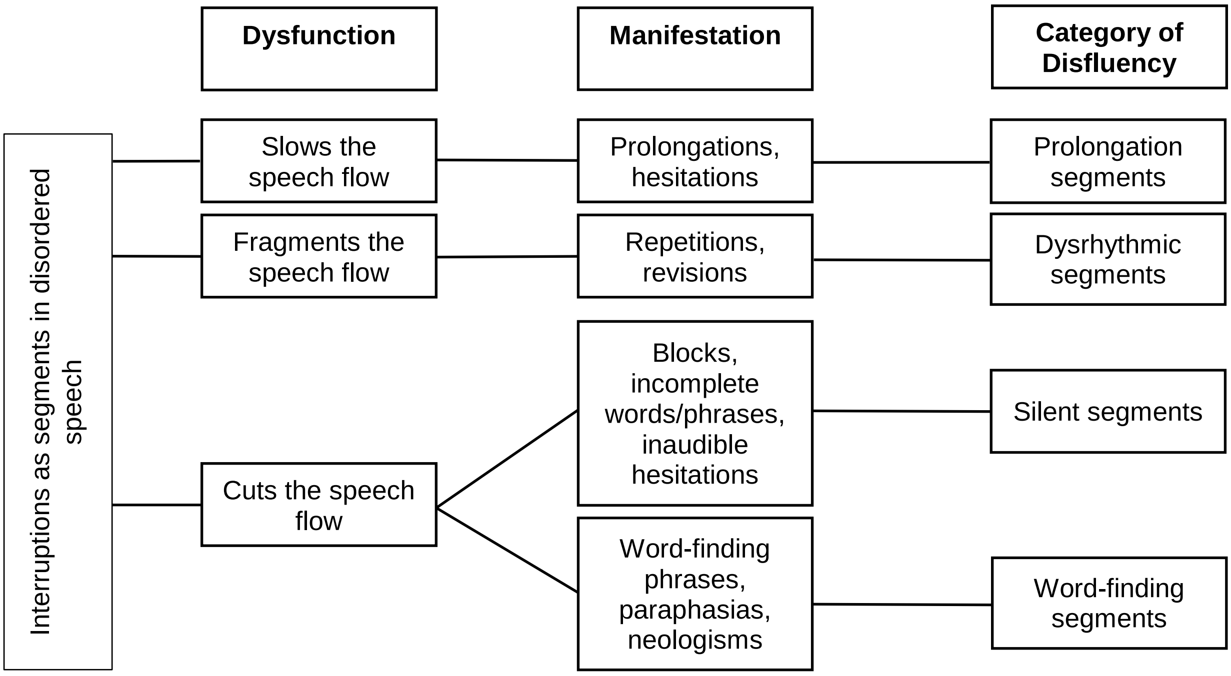
Yaruss, J. S. (1997). Clinical Measurement of Stuttering Behaviors. *Contemporary Issues in Communication Science and Disorders, 24,* 33-44.

Yorkston, K., Hammen, V., Beukelman, D., & Traynor, C. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders, 55,* 550-561.

Young, M. A. (1984). Identification of stuttering and stutterers. In R. F. Curlee & W. H. Perkins (Ed)., *Nature and treatment of stuttering: New directions* (pp. 13-30). San Diego: College-Hill Press.

Figures

Figure 1. Primary interruptions observed in disordered speech



Tables

Table 1.
The variety in the terminology related to disfluencies in stuttering research (modified after Yaruss, 1997, p.37)

Reference	Typical disfluencies in speech	Atypical disfluencies in speech	Classification is based on
Ambrose & Yairi (1999)	**Other disfluencies:** interjections, phrase repetitions, revisions	**Stuttering-like disfluencies:** part-word and single syllable word repetitions, dysrhythmic phonation	**Relation to stuttering**
Conture (1990)	**Between-word disfluencies**: interjections, phrase-, and polysyllabic whole word repetitions, revisions	**Within-word disfluencies:** sound-, syllable- and monosyllabic whole word repetitions, audible and inaudible prolongations	**Location**
Gregory (1993)	**More-typical disfluencies:** interjections, phrase-, monosyllabic word-, and part word syllable repetitions, revisions, hesitations	**Less-typical disfluencies:** sound-, monosyllabic- or part word syllabic repetitions, prolongations, blocks	**Incidence**
Meyers (1986)	**Normal-type disfluencies:** interjections, whole-word and phrase repetitions, revisions, incomplete phrases	**Stutter-type disfluencies**: part-word repetitions, prolongations, broken words, tense pauses	**Distinction between typical and pathological**

Table 2

Characteristics of participants in clinical group A

ID	Age	Sex	Injury	Severity of TBI	Aphasia	Dysarthria	Cluttering	AOS	Cognitive deficits			
									A	**P**	**E**	**Me**
A1	39	M	Fall	moderate		x						x
A2	41	M	MVA	mild			x		x			x
A3	34	M	PA	moderate	anomic	x				x		x
A4	50	M	MVA	very severe	global			x			x	x
A5	36	M	Fall	moderate		x					x	x
A6	35	M	Fall	moderate	anomic				x			x
A7	43	M	PA	mild		x						x
A8	60	F	MVA	severe	wernicke's	x				x	x	x
A9	61	M	MVA	very severe	broca's			x		x		x
A10	64	M	Fall	very severe	conduction			x		x		x

Note. Columns include participant identification (ID), age in years (at the moment of testing), sex (M = male; F = female), the type of injury (MVA = motor vehicle accident, PA = physical abuse), severity of the injury, type of the communication disorder (AOS= apraxia of speech) and the type of cognitive deficit (A = changes in attention, P = changes in problem-solving, E = changes in executive functioning, Me = changes in memory).

Table 3

Characteristics of participants in clinical group B with neurogenic stuttering

ID	Age	Sex	Injury	Severity of TBI	Aphasia	Dysarthria	Cognitive deficits			
							A	**P**	**E**	**Me**
B1	59	M	Fall	severe	dynamic		x			x
B2	22	M	MVA	moderate		x		x		x
B3	26	M	MVA	very severe		x			x	x
B4	61	F	MVA	moderate					x	x
B5	30	M	MVA	mild	anomic			x		x
B6	56	M	MVA	moderate			x		x	x
B7	19	M	PA	mild		x				x
B8	42	F	Fall	moderate						x
B9	59	F	PA	very severe	conduction			x		x
B10	42	M	MVA	moderate				x		x

Note. Columns include participant identification (ID), age in years (at the moment of testing), sex (M = male; F = female), the type of injury (MVA = motor vehicle accident, PA = physical abuse), severity of the injury, type of communication disorder and the type of cognitive deficit (A = changes in attention, P = changes in problem solving, E = changes in executive functioning, Me = changes in memory).

Table 4
Characteristic of the SLPs in the expert jury

ID	Working years (years and months)	Experience in the evaluation/rehabilitation in following disorders								
		ds	**ns**	**cl**	**ec**	**pa**	**aos**	**dy**	**aph**	**vd**
slp1	34 y	x	x	x	x	x	x	x	x	x
slp2	12 y	x			x	x	x	x	x	x
slp3	14y 8m		x				x	x	x	x
slp4	5y 9m		x		x	x	x	x	x	x
slp5	2y 10m	x		x			x	x	x	x
slp6	33y 1m	x		x	x		x	x	x	x
slp7	17y 11m	x	x				x	x	x	x
slp8	4y 8m			x	x	x	x	x	x	x
slp9	5y 2m	x	x		x	x	x	x	x	x
slp10	10y 6m	x			x		x	x	x	x

Note. Columns include participant identification (ID), working years (years and months), and information of SLPs work experience with communication disorders (ds = developmental stuttering, ns = neurogenic stuttering, cl = cluttering, ec = echolalia, pa = palilalia, aos = apraxia of speech, dy = dysarthria, aph = aphasia, vd = voice disorders).

Table 5
Fluency ratings for the three speaker groups

	Aggregate data	Control group	Clinical group A	Clinical group B
N	90	70	10	10
Mean	7.2	8.2	4.3	3.4
Median	7.9	8.1	5.7	3.3
Range	0.2-9.9	6.7-9.9	0.2-7.2	0.3-6.8
Std. Deviation	2.2	0.6	2.5	2.0

Table 6
Mean fluency ratings for dendrogram clusters

	High fluency	Average fluency	Below average fluency	Moderate impairment	Severe impairment	Very severe impairment
N	4	61	13	7	2	3
CG	4	61	5	0	0	0
CGA	0	0	6	2	0	2
CGB	0	0	2	5	2	1
Mean	9.6	8.2	6.5	3.4	1.5	0.4
Std. Deviation	0.2	0.4	0.6	0.6	0.3	0.2
Range	9.4-9.9	7.5-9.2	5.6-7.3	2.7-4.3	1.3-1.8	0.2-0.6

Table 7
Percentage of participants exhibiting features of fluency impairment

Number of features	Control group (*n*=70)		Clinical group A (*n*=10)		Clinical group B (*n*=10)	
	%	mean fluency rating	%	mean fluency rating	%	mean fluency rating
0	55.7	8.4				
1	24.3	8.2				
2	11.4	7.7	20	6.6		
3	5.7	7.5	20	5.8	10	4.1
4	2.9	7.2	30	5.1	40	3.8
5			30	1.2	50	3.1
Total	100	8.2	100	4.3	100	3.4

Table 8.

CGA and the clusters from the dendrogram

ID	Cluster	Fluency rating Mean (SD)	Communication disorders	Severity of TBI	S	SR	P	LD	SE
A7	BTA	7.2 (0.637)	DY	Mi		1			6
A6	BTA	6.1 (1.083)	A	M		2	7	4	2
A1	BTA	5.9 (1.134)	DY	M		1	6		6
A2	BTA	5.9 (1.113)	Cl	Mi		6			3
A3	BTA	5.9 (1.620)	A, DY	M		4	4	2	3
A5	BTA	5.6 (1.188)	DY	M		10	1		1
A8	M	2.9 (1.495)	A, DY	S		3	9	9	1
A10	M	2.7 (1.649)	A, AOS	VS	7	2	3	10	2
A9	VS	0.6 (0.670)	A, AOS	VS	3	7	10	10	2
A4	VS	0.2 (0.201)	A, AOS	VS	5	3	9	7	4
total					15	39	49	42	30

Note. Columns include participant identification (ID), cluster from the dendrogram (BTA = below average, M = moderate, VS = very severe), mean fluency rating, type of diagnosed communication disorder (DY = dysarthria, Cl = cluttering, A = aphasia, AOS = apraxia of speech), severity of the injury (Mi = mild, M = moderate, S = severe, VS = very severe), and the number of how many SLPs out of ten rated specific feature for the speaker (S = stuttering, SR = speech rate, P = pauses, LD = language difficulties, SE = something else).

Table 9.

CGB and the clusters from the dendrogram

ID	Cluster	Fluency rating Mean (SD)	Communication disorders	Severity of TBI	S	SR	P	LD	SE
B6	BTA	6.8 (0.901)	NS	M	2	1	1		2
B10	BTA	6.1 (0.921)	NS	M	8	1	5	2	1
B5	M	4.3 (0.579)	NS, A	Mi	1	1	4	5	1
B8	M	4.1 (1.297)	NS	M	10	5	7		
B7	M	3.4 (2.152)	NS, DY	Mi	4	1	1	2	5
B4	M	3.3 (1.377)	NS	M	9	6	6		2
B1	M	3.2 (1.263)	NS, A	S	8	4	7	9	
B2	S	1.8 (1.470)	NS, DY	M	10	1	8		3
B3	S	1.3 (1.351)	NS, DY	VS	9	6	10	2	5
B9	VS	0.3 (0.441)	NS, A	VS	10	4	9	6	1
total					71	30	58	26	20

Note. Columns include participant identification (ID), cluster from the dendrogram (BTA = below average, M = moderate, S = severe, VS = very severe), mean fluency rating, type of diagnosed communication disorder (NS = neurogenic stuttering, A = aphasia, DY = dysarthria), severity of the injury (Mi = mild, M = moderate, S = severe, VS = very severe), and the number of how many SLPs out of ten rated specific feature for the speaker (S = stuttering, SR = speech rate, P = pauses, LD = language difficulties, SE = something else).

Table 10

The significances (2-tailed) of the correlations between fluency ratings and features in each group

Group	Stuttering	Speech rate	Pauses	Language difficulties	Something else
CG	.120	.307	.044*	.317	.054
CGA	.429	.699	.115	.081	.365
CGB	.079	.157	.013*	.659	.446

Note. Columns include groups (CG = control group, CGA = clinical group A, CGB = clinical group B), * = significant at level $p < .05$. *Note 2.* Correlation between stuttering and fluency rating in control group is based on one participant only.

Table 11

Coefficients of regression analysis of fluency ratings and features of fluency impairment

Features	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
(Constant)	8.551	.162		52.772	.000
Stuttering	-3.170	.424	-.535	-7.483	.001
Speech rate	-.477	.278	-.110	-1.715	NS
Pauses	-.551	.370	-.121	-1.489	NS
Language difficulties	-.708	.388	-.132	-1.823	NS
Something else	-.784	.311	-.176	-2.516	.014

Note. Non-significant p-values are coded with abbreviation “NS”, significant at level p < .05.

Table 12

Level of SLP agreement by fluency feature

Agreement level	Stuttering	Speech rate	Pauses	Language difficulties	Something else
	***N* (%)**	***N* (%)**	***N* (%)**	***N* (%)**	***N* (%)**
No agreement (0-1 SLPs)	1 (7.1)	22 (56.4%)	8 (26.7%)	3 (16.7%)	14 (42.4%)
Low agreement (2 SLPs)	2 (14.3)	5 (12.8%)	3 (10.0%)	6 (33.3%)	10 (30.3%)
Moderate agreement (3-6 SLPs)	3 (21.4)	10 (25.6%)	9 (30.0%)	4 (22.2%)	9 (27.3%)
High agreement (7-10 SLPs)	8 (57.2)	2 (5.2%)	10 (33.3%)	5 (27.8%)	-
Total number of cases	14 (100)	39 (100%)	30 (100%)	18 (100%)	33 (100%)

Appendix 1. The process of distinguishing neurogenic stuttering after TBI from cluttering, palilalia, echolalia and AOS. (see Duffy, 2005: 358-359, 362-363, 417, 427; Lundgren, Helm-Estabrooks & Klein, 2010; Myers, Bakker, St. Louis & Raphael, 2012; Tani & Sakai, 2011)