STUTTERING AS THE RESULT OF A MISALLOCATION OF ATTENTION DURING SPEECH. A THEORY.

by Torsten Hesse, June 2018
www.stuttering-theory.eu

Contents

1. Introduction

1.1. Questions a theory of stuttering should answer

A lot of empirical data about stuttering, about people who stutter (PWS), and about their brains was collected in the last decades without having led to a convincing, received causal theory of the disorder up to now. It is, however, impossible to logically (in a strict sense) derive a theory from empirical data. Data can, at best, falsify a given theory. In theory building, empirical data can serve as suggestions, and they are bottom-up constraints confining the number of possible assumptions.

Theories always result from thinking and modeling, So is also the theory presented in this paper, and naturally it can be false. Theories about real phenomena cannot be proved in a strict sense (in contrast to mathematical theories which are about abstract phenomena). However, many features of stuttering are well known, thus we can formulate many questions of how these features come about. A theory of stuttering has to answer these questions in a coherent and consistent manner. If a theory is able to do this, then it should be considered a candidate for the true theory. The questions are the following ones:

- How do the typical symptoms repetition, prolongation, and silent block come about? When something goes wrong in the brain of PWS, why does it manifest in different kinds of symptoms, and why just in these kinds? What is the 'mechanism' behind that?
- Why does stuttering mostly and earliest onset when children start forming sentences, whereas previously, during the babbling phase and the period of one- or two-word utterances, the same children did not show any symptoms?
- Why do most of the affected children spontaneously recover from stuttering after some months or years, whereas a minority persists in the disorder, and why is the prediction of recovery much better for girls than for boys?
- Why is stuttering so variable, so strongly influenced by situations, by the environment, or by the number and the kind of listeners? If something goes wrong in the brain of PWS – why does it affect speech in some situations, but not in others?
- What relationship does exist between stuttering and linguistic structure? Why is a sort of phonemes (plosives, vowels...) more likely stuttered by one individual, another sort by another individual? Why does stuttering occur mostly at the onset of words and always

at the initial part of syllables? Why are stressed syllables more likely stuttered than unstressed ones, long words, and words with high information load more likely than other words?

- What about the structural and functional abnormalities found in the brain of PWS? What of them is causal for the onset of stuttering, what is causal for persistence? What is only a concomitant or a consequence of the disorder?
- Why is stuttering strongly influenceable by alterations of auditory feedback? Why can altered auditory feedback or auditory masking dramatically reduce or even completely inhibit stuttering immediately as long as the alteration continues, and remains unfamiliar?
- Why is stuttering, at least in some cases, successfully treatable by an alteration of behavior, e.g., by applying a special speech technique? The question is not trivial: If something goes wrong in the brain and causes the disorder – how can a change of behavior repair this?

All these questions should be answered by a theory of stuttering. Packman and Attanasio (2004, p. 55) claim that also the role of genes in stuttering should be explained, However, the assumption that genetic mutations directly cause stuttering is hardly consistent with the variability of the disorder. Therefore, not stuttering itself, but rather the predisposition for stuttering may be genetically caused.

Subject of the present theory is developmental stuttering, but not neurogenic stuttering caused by stroke, brain injury, or medication. I do not use the term 'psychogenic stuttering', as there may be cases of late-onset stuttering in which it is difficult to distinguish whether they were caused more by an inherent predisposition or more by psychological factors, e.g., a trauma. A causal theory of developmental stuttering should include cases of late-onset stuttering, because the symptoms as well as the abnormalities in the brain were found to be similar (Chang et al., 2010).

1.2. Basic assumptions about the control of fluent speech

1.2.1. Feedforward- and feedback-based control

Speaking is to produce a sensorimotor sequence: a sequence of movements as well as a sequence of perceptions resulting from the movements. The sequence of movements consists of segments of different complexity, as there are the movements by which a phoneme, a syllable, a word, or a phrase is produced. Less complex units are strung together to more complex units. The connections between units come about by practice: We hear a new word, repeat it and speak it a few times, the word becomes familiar to

us and part of our active vocabulary. There is no reason to assume that the motor sequence needed for speaking a word is learned in a way basically different from that in which other motor sequences are learned, e.g., the sequence needed for making a bow knot: It is learned by imitation and repetition, until the connections in the brain have been established.

How are sensorimotor sequences controlled so that they are executed fluently? This issue, of course, is important for a theory of stuttering. In the period of learning, sensory feedback is the basis of control: The perception of the current step being complete triggers the next step to start, i.e., each next step is a reaction to the perception of the end of the preceding step. However, execution cannot become fluent in this way, because every reaction implies a reaction time. A further argument against a purely feedback-based control is the fact that a well automatized motor sequence can be executed even if sensory feedback is interrupted. Hence, Lashley (1951) concluded that automatized motor sequences are feedforward-controlled by plans or programs, and that might also be the case in speaking.

Some experiments, however, showed that sensory feedback still plays a certain role in the control of speech: Lee (1951) and Stromsta (1959) demonstrated that a healthy person's speech flow can be disrupted by manipulation of auditory feedback, and Kalveram and Jäncke (1989) showed that the duration of long-stressed syllables is controlled depending on the auditory feedback of the syllable start ("audio-phonatory coupling"). I do not assume that these mechanisms play a role in stuttering, but they indicate that sensory and especially auditory feedback is involved in the control of fluent speech.

A further function of auditory feedback in speech control is the detection of errors and the immediate response to them. It is important to understand that this is part of the control of fluent speech: to ensure that the next speech unit (word, phrase, or clause) is executed only if the preceding unit is correct and complete. If an error is detected in the auditory feedback, the control system interrupts the sequence in order to enable a repair, according to Levelt's (1995) Main Interruption Rule. This is better than the complete utterance becomes unintelligible or misunderstood and must be repeated entirely.

Thus, even if speaking becomes more and more feedforward controlled in the course of learning and automatization, some elements of feedback control survive, probably because they allow to combine both, high automatization and high variability. Quick and automatic performance would be nothing particular if we always spoke one and the same sentence just as, in tying a bow, always the same comes out: a bow knot. In speaking, by contrast, we produce always new sentences. Words are arranged automatically and in high speed in an always new and meaningful order – that's an amazing ability. Elements of feedback control operate 'in the background'

and intervene only if necessary: when a syllable is prolonged in calling or in order to emphasize a word (audio-phonatory coupling), or when error repair requires an interruption of speech flow. The latter mechanism is suspected to be responsible for stuttering in the present theory.

1.2.2. Speech motor programs

A basic assumption of the present theory is that articulation is controlled by motor programs. The idea is not new. Van Riper (1983, p. 393) wrote that "words are learned motor sequences", and Segawa, Tourville, Beal, and Guenther (2015) define speech motor programs as "stored neural representations that encode the sequence of movements required to produce the utterance". These programs are acquired by practice: When a word from a foreign language is learned, or in special tasks like nonword repetition, the motor program first must be pieced together of subprograms for phonemes or familiar syllables. When, by contrast, familiar words or phrases are produced in everyday talking, then we can assume that the respective motor programs became established and are retrieved from long-term memory.

Empirical evidence for speech motor programs comes from Sternberg et al. (1988, 1980), who examined the performance of rapid movement sequences in speaking. Engelkamp and Rummer (1999) propose a model of the mental lexicon which is divided into an input- and an output system: Each word is represented twice therein: as an acoustic sound sequence allowing word recognition in listening, and as a motor program that controls articulation. The authors justify their model by results of aphasia research: cases in which either speech production, especially word access, is severely impaired with intact speech comprehension, or in turn, speech is no longer understood, but word production and speech fluency are not or less impaired (Broca's and Wernicke's aphasia, resp.).

How long can motor programs be when composed of several units (subprograms)? Sternberg et al. (1988) found that, in rapid reading of word lists, the maximum number of units was depending on unit size, e.g., 10 monosyllabic or 6 trisyllabic words in one program, which roughly coincides with a long clause or a sentence. As we however know, humans are able to memorize much longer speech sequences like poems, lyrics, or the lines of an actor's part, These might be speech motor programs as well – the difference is that such overlong programs require to commit them in long-term memory, which was probably not the case with the programs in the experiment conducted by Sternberg et al. (1988).

In the present theory, it is assumed that an utterance is controlled by only one program if only one decision is required, that is, after the program has started and is reeling off, no further decision on wording, grammar, or syntax is necessary. As mentioned, speech motor programs can be rather

long, e.g., in the case of a well memorized poem, but the usual length in spontaneous speech is that of a word or a familiar phrase.

Pauses for respiration as well can be part of a speech motor sequence. Similar to the movement sequence producing a familiar word, also the movement sequence of breathing in – as an automatized behavior – might be controlled by a motor program. Like motor programs for the production of words, that for breathing in must be included in the process of sequence forming in order to produce fluent speech: The end of the program for inspiration must be linked to the start of the program for the next word.

The basic idea of speech motor programs is that the production of familiar words and phrases in spontaneous speech is not a matter of planning or computation, but of learned and automatized motor routines. The concept of speech motor programs is central in the present theory, in which a stuttering event will be defined as the inhibition of a speech motor program, which will enable us to explain some astonishing features of the symptomatology of stuttering.

1.2.3. Self-monitoring and error repair

Civier, Tasko, and Guenther (2010) distinguish between two kinds of error correction: (1) online adjustment during ongoing speech, and (2) offline repair after an interruption of speech flow. We adopt the distinction, however, we assume that not the size but the sort of an error is crucial for whether it is corrected online or offline: If purely physical features like pitch, volume, or speech rate differ from a target level, then they are adjusted online. By contrast, linguistically relevant errors like mispronounced or mistaken words require an interruption of speech flow for a revision.

Subtle deficits in online-adjustment were found in PWS, as compared to normal fluent speakers, in several studies (e.g., Cai et al., 2012, 2014a; Loucks, Chon, and Han, 2012; Tourville, Cai, and Guenther, 2013). This may be due to a deficit in central auditory processing which we consider a factor in the predisposition for stuttering (see Section 3). However, a slightly delayed or reduced online-adjustment can hardly cause stuttering: The adaptation to auditory feedback manipulated in pitch is indeed reduced in adults, but not in children who stutter (Daliri et al., 2018). In the present theory, the second mechanism, i.e., interruption of speech flow for error repair, is suspected to cause stuttering. It is therefore necessary to consider this mechanism in more detail. A question of special interest is: What kind of self-monitoring allows the detection of phonological errors?

Phonological errors are detected by comparing the perception of a word or phrase just spoken with an expectation of its correct sound sequence. From there, the question arises: How are these expectations generated? Levelt (1995) believes that a ‘phonetic plan’ provides the basis for

that. A variety of this approach is the efference copy theory: An efference copy is defined as the projection of a motor plan onto that sensory system, by which the movement controlled by the motor plan is perceived (Holst, 1954; see, e.g., Beal et al., 2010; Brown et al., 2005 for this approach in the context of stuttering).

There are, however, two strong arguments against the hypothesis that efference copies are the basis of detecting errors in speech. First, the copy of the motor plan necessarily contains the movement sequence of the erroneous word or phrase, otherwise the error could not occur – and the efference copy of a wrong plan cannot be the basis of a correct prediction. Second, humans detect errors immediately, nearly as quickly as they detect their own errors, when listening to the speech of someone else, without having any copy of that speaker's plan. EEG studies showed that syntax errors elicit responses in a listener's brain after ca. 120ms, semantic errors after ca. 400ms (Friederici, 1999). Phonological errors can be equaled with syntax errors, as they are structure violations as well. Obviously, a listener, on the basis of auditory input and his/her knowledge of language, is able to generate an expectation of what a speaker is going to say and what it should sound like. So we can summarize: efference copies are neither useful nor necessary in the detection of speech:errors.

Obviously, auditory information is necessary and sufficient for error detection in speech, which might be true for one's own speech as well. Correct expectations are generated on the basis of auditory feedback and the speaker's (implicit) knowledge of language. The initial words of a sentence constrain the options to continue and allow to predict the subsequent words that are syntactically and semantically possible. A familiar word is often recognized after hearing its initial sounds, especially if the word is embedded in the semantic context of a sentence. So the word's correct phoneme sequence can be predicted from its initial portion. Astheimer and Sanders (2009, 2012) found that both, adults and preschool-aged children, when listening to speech, temporally modulated selective attention to preferentially process the initial portions of words. Already Halle and Stevens (1959) had developed a model of how words can be recognized and their phoneme sequence can be predicted based on a few initial phonemes. This 'analysis-by-synthesis' model was updated by Poeppel and Monaban (2011).

In the present theory, it is therefore presupposed that: (1) On the basis of the auditory feedback of the initial portion of a speech unit (word, phrase) just produced and the speaker's knowledge of language, the speech unit is recognized, and a correct prediction of its sound sequence is generated. (2) This prediction (expectation) is then compared with the sound sequence actually perceived. Every phonological mismatch between expectation and perception is an error and results in an interruption of speech flow in order to enable a correction. That is, both components needed for

the self-monitoring of speech – correct expectations and the perception of the output – depend on auditory feedback. A difference between the two components is important for the present theory: Only the initial portion of a word in usually needed for the generation of a correct expectation, but the complete sound sequence of the word is needed for checking whether the word was produced correctly and completely.

One may argue that the somatosensory feedback of articulation as well plays a role in the self-monitoring of speech. That is correct; especially the perception of self-produced consonants and the distinctness of articulation may be supported by that. However, somatosensory information is obviously not necessary for error detection as such, as we are able to detect all relevant errors in another person's speech without somatosensory information. Speaking is learned by imitation, thus the principle of verbal self-monitoring is: my speech is correct as long as it sounds like that of the rest of native speakers. Therefore, auditory information is crucial.

1.2.4. External and internal feedback

Auditory feedback, so I wrote above, is involved in the control and self-monitoring of speech. However, speaking as well as error detection are possible also when auditory feedback is interrupted, e.g., with auditory masking (whereby one's own speech signal is overlaid by loud noise) or hearing loss. Astonishingly, stuttering usually disappears in this condition (e.g., Maraist and Hutton, 1957; Van Riper, 1982; Webster and Lubker, 1968). These facts were alleged as arguments against theories claiming a disruption of auditory feedback in PWS (see, e.g., Postma and Kolk, 1992). It is therefore necessary to treat this issue in detail.

In his Perceptual Loop Theory, Levelt (1995) has described two feedback loops serving for the self-monitoring of speech: an external loop (hearing one's own speech externally), and an internal loop which is working when one's own speech cannot be heard externally. The latter is the case during silent verbal thinking (inner speech), but also with auditory masking (see above). Only feedback via the external loop is usually referred to as auditory feedback, but the internal feedback of speech is perceived in the same sensory modality: it is internally heard. Inner speech (verbal thinking) includes the imagery of an inner voice (McGuire et al., 1996), and it is processed in nearly the same auditory cortical areas as a speech signal perceived outside is processed (Brumberg et al., 2016; Kell et al., 2017a; Martin et al., 2014; Palmer et al., 2001; Shergill et al., 2002; Tian and Poeppel, 2012; Tian, Zarate, and Poeppel, 2016). Feedback via the internal loop should therefore regarded as auditory feedback as well,

That means, a speaker always gets auditory feedback, as long as his or her speech comprehension is not impaired (as it is in a Wernicke's aphasia). Thus, error detection, audio-phonatory coupling, and incremental

sentence planning work normally also when the external feedback loop is interrupted. When, however, the external feedback loop is working (because one's own speech is heard), then the internal loop is off – otherwise we would perceive our own speech twice, as the two kinds of feedback need different time (Levelt, 1995). That, however, means: If the central processing of external feedback is disrupted or insufficient (which is assumed in the present theory), then internal feedback cannot compensate for that. The internal feedback loop is off as soon as one's own voice is externally heard, regardless of whether the external feedback is properly processed or not on higher levels of processing

1.2.5. A simple model of normal speech production

In the present theory, stuttering is assumed to be a disorder of normal speech production, because (i) even individuals with severe stuttering show phases of fluent speech and are able to speak fluently under certain conditions, e.g., with altered auditory feedback (see below), and (ii) some normal fluent speakers stutter sometimes in awkward situations. Therefore, the present theory needs to be embedded in a model of normal speech production, not least because it should explain what goes wrong in the process of speech production such that stuttering comes about. On the other hand, the emergence of stuttering must be possible in such a model, otherwise it is neither realistic nor useful. The model underlying the present theory is sketched in Figure 1.

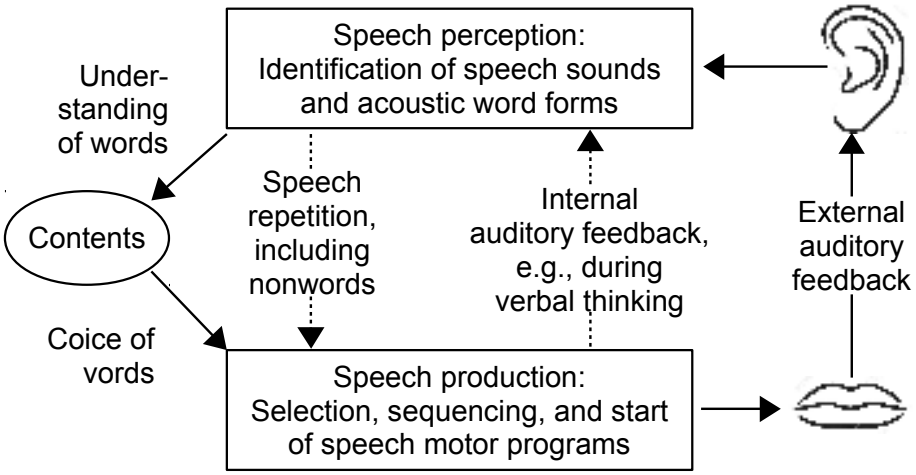

Figure1: Model of normal speech production. Not all functions work at the same time. Solid arrows symbolize the process in normal spontaneous speaking.

The model is derived from Levelt's depiction of his Perceptual Loop Theory (Levelt, 1995, p. 470), however, with an important alteration: Levelt

assumes that formulation and articulatory control are separate from each other in two different modules (Formulator and Articulator) with the Formulator controlling the Articulator, but without the Articulator influencing the Formulator. In the model sketched above, by contrast, formulation and articulatory control take place in the same speech production system. Consequently, external and internal feedback loop have the same origin, and the internal loop cannot serve for a 'pre-articulatory monitoring'.

Pre-articulatory monitoring, however, takes place only in non-spontaneous speech: a phrase or clause is first silently thought, i.e., internally spoken and monitored, and subsequently spoken out aloud. We behave so only in very special situations, and such thinking prior to speaking works in the same way as verbal thinking/ inner speech in general does. That is, there is no specific mechanism of 'pre-articulatory monitoring'. The idea behind the model sketched in Fig. 1 is that – in spontaneous speech – sentences are not first formulated by the mind and subsequently articulated by muscles, but that sentences are formulated by speaking them out, such that formulation and articulation can interact with each other, as was also proposed by Dell and Reich (1981).

There ere some arguments supporting the position that formulation and articulation are not separate: First, the position is consistent with our above assumptions that motor programs are the basis of speech control (Section 1.2.2). This is true not only for overt speech, but also for inner speech/ verbal thinking, as both are controlled by the same sort of motor programs (Tian and Poeppel, 2010). with the only difference that, in inner speech, the execution of these programs is partly inhibited, at least in terms of phonation (tongue and lips are often moved in silent thinking or reading). Our model is further consistent with the above assumption that error detection does not work on the basis of a pre-articulatory plan (Section 1.2.3); there is no such plan in spontaneous (!) speech.

A fact suggesting an interaction between articulation and formulation is the frequent occurrence of a sort of speech errors: malapropism. We often confuse similarly sounding words i.e., words that are similar in articulation. A further suggestion is the 'tip of the tongue' phenomenon: You cannot remember someone's name at the moment, but you remember its initial sound. These things can hardly be explained in a model, in which formulation is separate from articulation. In our model, by contrast, word selection is nothing else than the selection of a motor program, and similar motor programs, particularly such with similar beginning, are confused more likely than dissimilar ones.

Finally, as claimed above, stuttering must be possible in a realistic model of speech production. Considering Levelt's (1995) model, the question arises: Is stuttering caused by a dysfunction of the Formulator or of the Articulator? The Formulator can hardly be responsible, as PWS usually

have no difficulty selecting words and forming correct sentences. The Articulator as well can hardly be responsible: PWS mostly are fluent when not or less required to formulate by themselves, e.g., in singing, shadowing, and chorus reading (see below). Additionally, stuttering does not yet occur in young children as long as they produce only single words. This suggests that articulation as such is not the problem in stuttering. A third possibility is: The signal transfer from Formulator to Articulator is impaired. That would, however, mean: A person stuttering does not know what he or she is going to say, which is just the contrary of what PWS experience. They exactly know the word they want to say and the sound it begins with, but they do not get it out. Thus the most plausible account would be: Stuttering is anyway caused by an adverse interaction of Formulator and Articulator – but just an interaction is not provided in Levelt's model.

2. The mechanism of stuttering

2.1. Inhibition of speech motor programs by invalid error signals

This section is about the mechanism by which stuttering events are immediately caused. A core assumption of the present theory is that a stuttering event is caused by the inhibition of a speech motor program because of an invalid error signal elicited in the monitoring system: The system 'believes' that a new speech unit (a word, a phrase, or a clause) is going to start, but the preceding unit has been erroneous or incomplete. That is, stuttering is not regarded a breakdown of control in the present theory, even though the affected person subjectively may experience a loss of control. Instead, stuttering is regarded a regular response of the control system to an error signal, as postulated in the Main Interruption Rule (Levelt, 1995, p. 478; Levelt, 1983). Figure 2 shows the system's response to a real speech error and to an invalid error signal.

The core symptoms of stuttering come about in the following way: When a speech motor program is inhibited because of an invalid error signal, then the speaker is not aware of an error, as he/she did not misdo. Thus the speaker automatically and spontaneously tries to continue, and only by this, the externally observable symptoms of stuttering come about. If the motor program is inhibited immediately at its onset, either a silent block occurs (with onset on a vowel, on /h/, or on a stop consonant). or a prolongation of the first sound (with onset on a prolongable consonant). In the latter case, the program gets caught on the initial sound.

Since the response to an error signal needs a reaction time, it often happens that a motor program is not inhibited immediately at its onset, such that the beginning of the speech unit affected can still be produced. Then,

after the motor program is inhibited at a point, it restarts, driven by the speaker's will to continue, and runs until it is inhibited again at the same point. This recurs again and again as long as the inhibition is effective and as long as the speaker tries to continue. In all cases, stuttering behavior consists of two components: the inhibition of a speech motor program and the person's spontaneous attempt to continue.

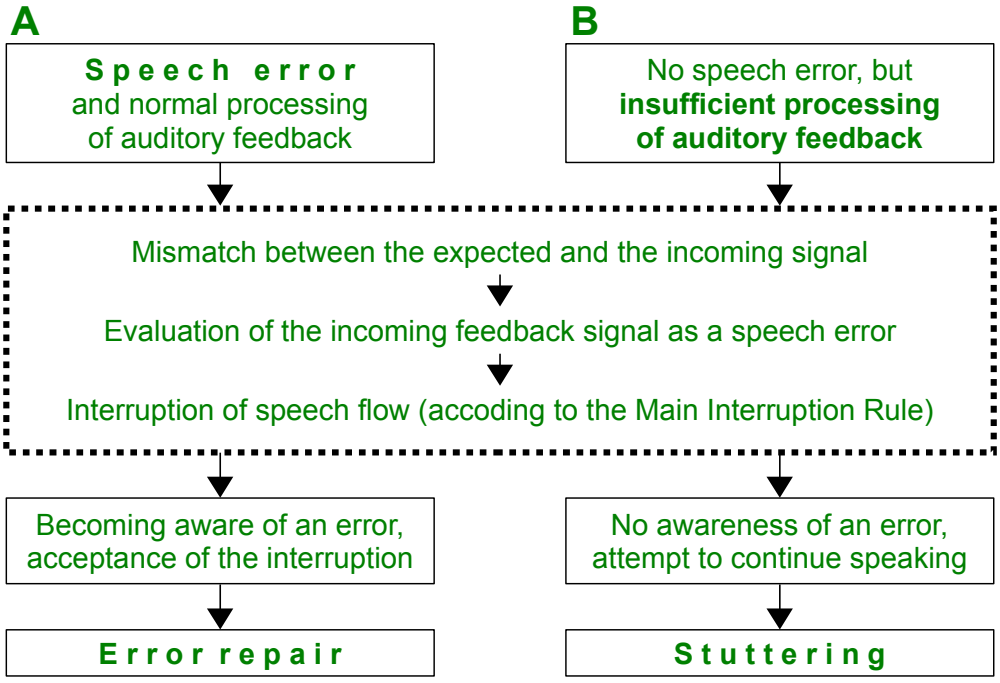

Figure 2: Detection of a speech error (A) and stuttering (B) are based on the same mechanism: the monitoring system (in the dashed frame) behaves equally in both cases.

Why is always the start of the next program inhibited, such that stuttering mostly occurs at word onset and always at the beginning of a syllable? Why does the error signal not stop the running program just before its end, but only the start of the next program? Probably because a mismatch between expectation and perception is not necessarily an error – maybe, for example, the speaker only hesitates and then completes an incomplete word. It makes therefore sense for the system to wait with an interruption of speech flow until the speaker is going to start the next program despite the mismatch. Then this program is inhibited during or just after its start, depending on the system's reaction time.

2.1.1. Explanation for some features of stuttering

The above description of a stuttering event provides an explanation for the fact that stuttering usually occurs at word initial position. For example,

Natke et al. (2004) found that, in preschool children 97.8% of all stuttering events occurred on the first syllable of words. Sheeban (1974) found 96% and Taylor (1966) 97% of stuttering events on the first syllable of words in adults who stutter. This pattern is explained by our assumption that a word is stuttered because of an error signal concerning the preceding word or speech unit, with effect on the beginning of the next unit, as described above. If stuttering occurs at the onset of an utterance, the preceding speech unit is the phase of breathing in prior to speech (see Section 2.4).

Sometimes, not the first syllable of a word is stuttered, but another one. The cause might be that the speech motor program was fragmented (dechunked), which happens when speaking is controlled less automatically, but more by the will. Bloodstein (1975) reported that fragmentation of words into smaller units was observed in the speech of most young children when they attempted to produce utterances that they found motorically or linguistically demanding. As a consequence, stuttering can occur also in word-medial position: If every syllable starts separately, then each of these starts can be inhibited if an error signal occurs in the preceding syllable.

Repetitions of monosyllabic words as well are regarded as symptoms of stuttering (Yairi, 1996). Such symptoms are possible because, as mentioned above, also a familiar phrase consisting of several words can be controlled by only one motor program. Many common phrases in English start with a monosyllabic function word: an article, a pronoun, or a preposition. If the motor program of such a phrase is inhibited – because of an error signal in the preceding speech unit and after a reaction time – then it can happen that the first word of the program has already been spoken when the inhibition onsets. The program then restarts at its beginning, and the word is repeated.

A peculiarity of stuttering is the fact that some PWS are fluent when reciting a memorized poem or when speaking the memorized lines of an actor's part on stage. As mentioned above, very long motor programs can be assembled and entrenched in the brain by means of memorizing by repetition. Once such a program has successfully started, no invalid error signals can occur in it, because they are always related to the preceding program. The effect is contrary to that of the above-mentioned fragmentation: If the motor program of a word is fragmented into programs of syllables, then stuttering can occur even in word-medial position, on the second or third syllable – if , by contrast, many programs of words are assembled to a long 'super program', then stuttering will hardly occur within it. Memorized lyrics as well are such super programs, which might be one reason why stuttering does not occur in singing. Likewise, the 'adaptation effect' – stuttering is reduced if a text is repeatedly read (Johnson and Knott, 1937; Van Riper and Hull, 1955) – can be explained in a similar way: The more familiar the text, the fewer words are fragmented, and the more words are

bundled (chunked together) to phrases, each produced by only one motor program . The fewer programs need to start, the fewer stuttering can occur.

The present theory clearly differentiates stuttering from normal disfluencies occurring when someone does not exactly know what to say or how to say it After the present theory, stuttering can only occur if a speech motor program is just starting or has just started, i.e., the speaker is about or has begun to articulate a specific sound sequence, and the motor program is inhibited at this moment. This approach is in agreement with the experience of every person stuttering: He or she exactly knows what to say, but does not get it out. The clear distinction between stuttering and normal speech disfluencies is supported by Jiang et al. (2012), who found stuttering-typical disfluencies and normal disfluencies to be associated with different activation patterns in the brain. This suggests that they are caused differently.

2.2. The cause of invalid error signals: insufficient processing of sensory feedback

Invalid error signals, so I assume, are elicited in the monitoring system, when the sensory feedback of speech is insufficiently processed by the brain. As described in Section 1.2.3, automatic self-monitoring of speech depends on the comparison between a correct expectation and the actual perception, i.e., the sensory feedback. If now the sensory feedback is poorly processed, e.g. not completely kept in working memory, or not completely transmitted to the monitoring system, then a mismatch between expectation and perception occurs: The internal monitor ‘believes’ that the sequence (e.g., a word or phrase) was not completely produced, thus the system automatically inhibits the motor program trying to start.

The assumption that invalid error signals are elicited in the monitoring system because of a mismatch between expectation and perception is in agreement with the theory proposed by Hickok, Houde, and Rong (2011) and by Tian and Poeppel (2012). There is, however, an important difference: Those researchers suspect incorrect predictions to cause the mismatch – the present theory assumes that not prediction, but perception or, more precisely, the processing of sensory feedback is impaired in developmental stuttering. This is also not a quite new idea: Maraist and Hutton (1957, p. 385) already assumed that “the stutterer misevaluates his own speech output at some point in the control system and finds error where, in reality, no error exists”. However, Maraist and Hutton believed that stuttering is caused by the attempt to repair those nonexistent errors (a similar theory was proposed by Vasic and Wijnen, 2005). By contrast, I think that the interruption of speech flow because of an invalid error signal is an automatic response, according to the Main Interruption Rule. The speaker (or the brain) does not

attempt to repair an error, but the speaker attempts to continue talking despite the inhibition of the motor program, and this causes overt stuttering.

Two modalities of sensory feedback are thought to be affected by deficits in processing: auditory feedback and the somatosensory feedback of breathing (perception of tension and relaxation of the muscles involved in breathing). The assumption that stuttering is caused by insufficient processing of auditory feedback is supported by the results of several brain imaging studies: Differences were found in the activation of brain areas in which auditory verbal input is processed, particularly in the left posterior temporal cortex:

- During speech, activation was reduced in PWS as compared with normal fluent controls (e.g., Braun et al., 1997; Chang et al., 2009; Fox et al., 1996; Ingham et al., 2003; see also the meta-analyses by Brown et al., 2005 and by Budde, Baron, and Fox, 2014).
- Likewise, activation was reduced in PWS (persistently) as compared with those who had naturally recovered from stuttering in childhood (Ingham et al., 2003).
- Within the group of PWS, auditory activation was reduced during stuttered speech as compared with activation during fluent speech evoked by chorus reading, paced speech, or in other fluency-enhancing conditions (Braun et al., 1997; Fox et al., 1996; Stager, Jeffries, and Braun, 2003; Toyomura, Fujii, and Kuriki, 2011; see also the meta-analysis by Budde, Baron, and Fox, 2014).
- Within the group of PWS, activation in auditory areas was reduced in individuals with severe stuttering, as a group, compared to those with mild stuttering (Fox et al., 2000; Ingham et al., 2004; Neumann et al., 2003),
- Auditory activation was reduced in PWS prior to a fluency-shaping therapy as compared with the same group after therapy (De Nil et al., 2003; Ingham et al., 2003; Neumann et al, 2003).

In studies of functional connectivity, reduced auditory-motor coupling was found in adults who stutter (Kell et al., 2017b; Watkins, 2011) and in boys who stutter, 3-9 years of age (Chang and Zhu, 2013). Successful treatment of stuttering increased the functional connectivity between auditory and motor areas, and in individuals who had recovered from stuttering in adulthood, auditory-motor coupling was normalized (Kell et al., 2017b). Cai et al. (2012; 2014) found weak responses to auditory feedback perturbation during articulation in PWS, compared to normal fluent controls. All these results strongly support the assumption that poor involvement of auditory feedback in speech control is a causal factor in stuttering.

2.1.2. The role of the cerebellum in error response

Until now, there is no empirical evidence of the invalid error signals which are assumed to cause stuttering in the present theory. There is, however, a part of the brain that plays a important role in motor sequence learning and in the detection and repair of motor errors on the basis of sensory feedback: the cerebellum (see overview by Seidler et al., 2014). Zheng et al. (2013) identified a brain network that appears to encode an 'error signal' in reaction to distorted auditory feedback during articulation. The network includes right angular gyrus, right supplementary motor area (SMA), and bilateral cerebellum.

In the light of these general findings, cerebellar activity in stuttering is of interest: Several brain imaging studies have shown that cerebellar overactivation is one of the neural signatures of stuttering (see meta-analyses by Brown et al., 2005 and by Budde, Baron, and Fox, 2014). Moreover, cerebellar activity in PWS was found to be positively correlated with stuttering severity (Fox et al., 2000; Ingham et al., 2004). Yang et al. (2016) found several abnormalities in resting-state functional connectivity between cerebellum and frontal cortex as well as within the cerebellum, some of them correlated with stuttering severity. Kell et al. (2017b) report that, in individuals who spontaneously recovered from stuttering, activity in the superior cerebellum appeared to be uncoupled from the rest of the speech production network.

Wymbs et al. (2013) investigated individual differences in the brain activation patterns in four PWS during stuttered and fluent word production. They found many brain regions overactivated during stuttered speech, but across-subject agreement for overactivated regions was minimal. The only region which was overactivated in all the four participants during stuttered speech was the left hemisphere of the cerebellum. A further hint at the crucial role of the cerebellum in stuttering is provided by cases in which lifelong stuttering disappeared because of impaired cerebellar function (Bakheit, 2011; Miller, 1995; a further case was in 2012 in Germany, but the study has not been published). Tani and Sakai (2010) report a case of stuttering (including typical secondary symptoms) after right cerebellar infarction. Perhaps, the impairment of cerebellar function here resulted in the generation of invalid error signals by the cerebellum itself, without a deficit in feedback processing. The findings are at least well consistent with the assumption that stuttering is triggered by invalid error signals because of poor processing of sensory feedback.

2.3. The cause of poor feedback processing: a misallocation of attention

Insufficient processing of sensory feedback in PWS is assumed to be caused

by a misallocation of attention, i.e., of perceptual and processing capacity (which might be overall limited) during speech, with too little attention being directed to auditory feedback and/or to the somatosensory feedback of breathing. Result is the causal chain depicted in Figure 3.

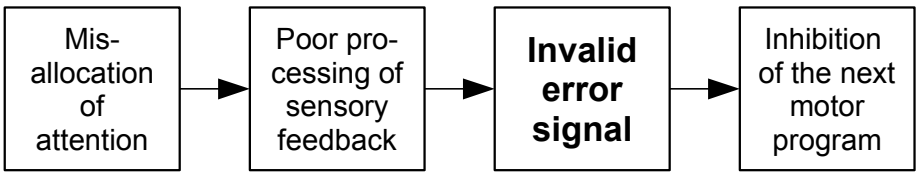

Figure 3: Causal chain leading to a stuttering event.

Has the allocation of attention during speech an impact on the processing of auditory feedback? It is evident that attention modulates perceptual speech processing. Cherry (1953) carried out a behavioral experiment in which two different spoken messages were presented simultaneously via headphones, one to the left, the other one to the right ear. Participants were asked to pay attention to the message in only one designated ear and to repeat it aloud. Later, when asked for the ignored message which was presented to the other ear, participants were not only unable to render its content, they even appeared not to have noticed a change in language (from English to German), or a change to speech played backward. Only relatively gross physical changes in the stimuli were noticed without attention, e.g., a switch from a male to a female voice.

Brain imaging studies confirmed a modulation of perceptive speech processing by selective attention. Certain auditory areas were found to be greater activated when participants were actively listening to verbal stimuli, and activation was lower when these stimuli were perceived only passively (Jäncke, Mirzazade, and Shah, 1999; Hugdahl et al., 2003; Sabri et al., 2008). Moreover, some cortical areas probably involved in speech comprehension showed activation only when participants were listening to the verbal stimuli, but not when attention was directed to something else (Sabri et al., 2008)). This suggests that the processing of verbal information is very limited without sufficient attention to the auditory channel.

We can assume that another person's speech and the auditory feedback of one's own speech are processed by the same speech comprehension system (Levelt, 1995) localized in the posterior temporal cortex (McGuire, Silbersweig, and Frith, 1996; Price et al., 1996; Zheng, Munhall, and Johnsrude, 2010), and we can assume that they are processed in a similar way, at least regarding the recognition of phonemes and words. Therefore, attention might affect the processing of the auditory feedback of one's own speech in a similar degree as the processing of speech produced by someone else.

This assumption is supported by Scheerer, Tumber, and Jones (2016),

who investigated how attention modulates the utilization of auditory feedback for speech control. They tested the response to a feedback manipulation in pitch, i.e., in a simple physical property the perception of which was found to be less attention-depending in the Cherry (1953) experiment mentioned above. Nevertheless, Scheerer and her colleagues found attention to have an impact on the degree in which auditory feedback was used for the control of speech. Hu et al. (2015) as well as Tumber, Scheerer, and Jones (2014) had obtained similar results. Processing and utilization of higher-order, i.e. phonological, lexical, and semantic feedback information might even more be depending on sufficient attention to the auditory channel.

If, however, (1) the processing of the auditory feedback of speech is attention-depending, and if (2) a speaker's attention is turned away from the auditory channel, then, auditory feedback may not be sufficiently processed (maybe some phonemes are not recognized, not kept in working memory, or not transmitted to the monitoring system). This would result in a mismatch between expectation and perception in the monitoring system and in an invalid error signal when the next speech motor program tries to start. Importantly, the detection of a mismatch between perception and expectation is independent of attention: The mismatch-related brain potential was found also when a person's attention was distracted from the unexpected stimulus (Näätänen et al., 2007; the brain response to an error is as a special case of the response to an unexpected stimulus). Therefore, the interaction between the attention-dependent processing of the auditory feedback of speech and the attention-independent mismatch response can lead to invalid error signals when too little attention is directed to the auditory channel.

2.3.1. The causes of insufficient attention to auditory feedback

Now the question arises: Why should PWS turn attention away from the auditory channel during speech? Three factors seem to play a role here: (1) a deficit in central auditory processing with impact on the control of auditory attention, (2) a general deficit in the control of attention with tendency towards hyperactivity or impulsivity, and (3) the attempt to actively avoid stuttering symptoms. The three factors can act in combination and can be present in various degrees in individuals. The two first-mentioned factors may be caused genetically and form the basis of a predisposition for persistent stuttering (see also Section 3.2)

There is growing evidence for deficient auditory processing of both, non-linguistic and linguistic acoustic stimuli in PWS (e.g., Andrade et al, 2008; Chang et al., 2009, Corbera et al. 2005; Halag-Milo et al., 2016; Hampton and Weber-Fox, 2009; Kikuchi et al., 2011, 2017; Neef et al., 2012; Prestes et al., 2017; Salmelin et al, 1998; Tahaei et al., 2014). In some studies, a relationship was found between auditory processing and stuttering frequency or -severity (Beal et al., 2010; 2011; Jansson-Verkasalo et al.,

2014; Kikuchi et al., 2016; Liotti et al., 2010). Moreover, teenagers who persisted in stuttering (as a group) were found to differ from those who recovered in terms of auditory processing; the persistent group was less able to hear a target sound prior to masking noise (Howell, Davis, and Williams, 2006).

It seems however unlikely that deficient auditory processing immediately causes invalid error signals and thereby stuttering – the variability of stuttering and the influence of situation and environment on the speech disorder can hardly be explained as the effect of a constant impairment. Deficient auditory processing on a low level, e.g., a poor auditory gating which was found in PWS (Kikuchi et al. 2011; Saltuklaroglu et al., 2017) might rather impair the control of auditory attention; for example, affected individuals may develop a habit of turning their attention away from the auditory channel as long as they are not actively listening. in order to prevent acoustic overstimulation. However, Tahaei et al. (2014) point to the opposite possibility that auditory processing even in the brainstem, i.e,, on a low level of processing can be modulated by attention via a top-down pathway from the cortex. In every case, a deficit in central auditory processing seems to be a factor in stuttering, and an interaction seems to exist between deficient auditory processing and the control of auditory attention.

The second factor leading to a misallocation of attention during speech in PWS seems to be an overall deficit in attention regulation, a tendency towards hyperactivity, impulsivity, and emotional reactivity – personality traits more prevalent in PWS than in the normal fluent population (see, e.g., Alm, 2014; Anderson and Wagovich, 2010; Eggers, De Nil, and Van den Bergh, 2012; Karras et al., 2006 for an overview). Particularly, behavioral tests showed that PWS, on average, were more prone to premature responses and false alarms than normal fluent controls (Bosshardt, 1993; Eggers, De Nil, and Van den Bergh, 2013; Jansson-Verkasalo et al., 2012; Markett et al., 2016).

A third factor contributing to a misallocation of attention during speech in not a few PWS is their attempt to actively avoid stuttering, either by excessive speech planning (substitution of words on which trouble is feared, reformulation of clauses in order to defer such words) or by volitional control of articulation and too much attention to its sensory (kinesthetic, tactile) feedback. Some results of brain research my be related to suchlike behaviors; Overactivation of right frontal speech areas (Neumann et al., 2003; Preibisch et al., 2003) suggests increased effort for speech planning. Greater and premature activation of motor and premotor areas especially in the right hemisphere (e.g., Salmelin et al., 2000; meta-analyses by Brown et al., 2005, and by Budde, Baron, and Fox, 2014) suggests voluntary motor control. Kell et al. (2017b) found that persistent stuttering was associated with reduced auditory-motor coupling in the brain, but increased integration

of somatosensory feedback. This hyper-connectivity was reduced after a successful stuttering therapy and was not found in a group of subjects who had spontaneously recovered from stuttering.

2.3.2. Modulation of attention during word recognition

As posed above in Section 1.2.3, auditory feedback is the basis of the recognition of the words actually produced. For quick word recognition after the analysis-by-synthesis model (Halle and Stevens, 1959; Poeppel and Monaban, 2011), attention is focused on the beginning or the initial part of a word (Astheimer and Sanders, 2009). This mechanism is essential for the self-monitoring of speech as it generates the expectations of the correct sound sequences. The mechanism cannot be substantially impaired in PWS, otherwise they would have difficulty detecting their speech errors (which is not the case). We must therefore exclude deficits in auditory attention and in the processing of auditory feedback at the beginning of words and of longer speech units. The normal modulation of auditory attention from the beginning to the end of speech units may be increased in PWS by a tendency towards shifting attention prematurely to the planning of the next clause, or to a word that attracts attention because of its length, informational or emotional content, or because of its initial sound on which stuttering is anticipated. Decreasing auditory attention at the ends of words or longer units results in a higher risk of insufficient processing of auditory feedback and invalid error signals.

The fact that PWS have no difficulty detecting their own speech errors has been taken as an evidence against all theories which claim that distortions of auditory feedback cause stuttering (e.g., Postma and Kolk, 1992). It is however assumed in the proposed theory that auditory feedback is poorly processed only in the back portion of words, especially of unstressed, 'unimportant' words . Speech errors occurring at such positions would then not detected. However, speech errors ('slips of the tongue') affecting only these positions are rare, at least with talking in one's native language and with normal grammatical ability. Already Meringer and Mayer (1895) found that errors were more likely on positions of more 'weight', and that "the sounds with most weight are the onset of the root syllable, the word onset and the stressed vowel or vowels" (cf. Nooteboom & Queene, 2015, p. 33). In other words, common speech errors mostly occur at the onset of words or word stems (initial consonants) and on stressed syllables (see also Dell, Juliano, and Govindjee, 1993; Fromkin, 1971). Thus, speech errors are unlikely to occur at positions at which auditory feedback is assumed to be poorly processed after the present theory, and the theory does not predict significant differences between PWS and normal fluent speakers in the detection and repair of slips of the tongue.

The present theory assumes that stuttering is caused by invalid error

signals, which has some similarity with the Covert Repair Hypothesis proposed by Postma and Kolk (1993); however, they have believed that the error signals responsible for stuttering were valid: Real errors in speech planning were detected in pre-articulatory monitoring and covertly repaired, and those covert repairs (which are unconscious to the speaker) cause the stutter. By contrast, the present theory does not assume any error of speech planning in stuttering, but invalid error signals due to a dysfunctional self-monitoring.

2.3.3. The variability of attention allocation explains the variability of stuttering

In the proposed theory, the allocation of the speaker's attention is the interface between the pathomechanism underlying stuttering, on one hand, and all the psychological, situational, and environmental factors which can influence the frequency, distribution, and severity of stuttering, on the other hand. The present theory can therefore be referred to as 'attention allocation theory of stuttering'. In this section, some aspects of the variability of stuttering are explained in the framework of the theory. Generally, the cause of stuttering – misallocation of attention during speech – is not a stable state, but is permanently changing, becoming greater or smaller depending on several variables.

Stuttering often occurs on 'salient' words: on long words, on stressed (emphasized) words, on words with high information load, on words rarely used by the speaker, and on words on which trouble is expected (see Bloodstein and Bernstein Ratner, 2008, for an overview). Probably, the speaker's attention prematurely shifts to the planning of such words, which reduces the attention (processing capacity) for auditory feedback especially of virtually unimportant, inconspicuous, unstressed words or word ends prior to salient words. In this way, invalid error signals because of poor feedback processing occur predominantly in speech units prior to salient words, with the effect that the motor program of the salient word is inhibited.

Likewise, the influence of situation and environment on the frequency of stuttering can be explained by their impact on the allocation of the speaker's attention. For example, talking to a superior or to a large audience versus talking to oneself, to a pet, or to a toddler impose different requirements on speech planning (correct wording, grammar) and are associated with different levels of stress (fear of saying something wrong, fear of stuttering). Thus, different speech situations require different amounts of attention for speech planning. Maxfield et al. (2016) found heightened demands on speech planning (word selection) to be associated with reduced capacity for auditory perception in PWS, but not in not in normal fluent controls.

It is not very likely that stress immediately causes stuttering, but it

influences attention allocation by drawing attention to the stressor. For example, Kawahara, Sato, and Takenaka (2011) showed that stress reduced attentional ressources for nontarget processing. The processing of the sensory feedback of speech is nontarget processing, at least when the speaker focuses on formulation, on the avoidance of stuttering, or on the impression he or she makes on the listeners. In sum, speech planning demands as well as arousal or stress might influence the allocation of the speaker's attention to the detriment of feedback processing and thereby increase stuttering.

2.3.4. The role of the basal ganglia

Some researchers have suspected the basal ganglia to play a crucial role in the causation of stuttering (e.g., Alm, 2004). The present theory, however, assumes that invalid error signals and/or the resulting inhibition of speech motor programs originate from the cerebellum (see Section 2.2.1). However, a relationship seems to exist between high dopamine reception in the basal ganglia and stuttering. For example, Metzger et al. (2018) found a correlation between substantia nigra activation and stuttering severity in a non-speech (!) motor task (dopamine is released from the substantia nigra to the striatum), and medication blocking dopamine reception (e.g., with haloperidol) was found to reduce stuttering in many cases (see, e.g., Alm 2004).

Children in general have a peak in the number of dopamine receptors type D2 in the striatum at the time when stuttering typically onsets (Alm, 2004). At the same time, it is the age at which children start forming sentences. The latter might be the reason for a higher risk of stuttering to onset at this age in general, because of higher demands of speech planning and feedback processing. An especially high dopamine reception in the basal ganglia may heighten a child's individual risk of stuttering, because of a relationship between basal ganglia activity and attention control:

Basal ganglia together with SMA, thalamus, and motor cortex form the 'medial premotor system' (Goldberg, 1985) which controls voluntary movements – whereas the 'lateral premotor system' connecting cerebellum and lateral premotor cortex with the motor cortex controls movements that are elicited not internally by the person's will, but externally by perception, e.g., by sensory feedback. Voluntary movements are usually associated with selective (goal-directed, top-down) attention, therefore, a dominance of the medial system leads to an attentional imbalance to the detriment of sensory feedback which rather depends on bottom-up attention. This (after the present theory) results in a higher risk of stuttering.

Besides the predisposition for stuttering, the activity of the basal ganglia and of the overall medial premotor system seems to influence the severity of overt stuttering behavior. When a speech motor program is inhibited because of an invalid error signal, the speaker automatically tries

to overcome the inhibition and to continue talking. This behavior is driven by the medial premotor system, as it is depending on the speaker's will (the 'drive'). In stuttering modification therapy, the patient learns to suppress the 'speech drive' when feeling an internal barrier and, in this way, to reduce or to avoid overt stuttering symptoms. Dopamine blockers like haloperidol seem to operate in a similar way, calming the medial premotor system. Alm (2004) wrote: "The drug seems to exert its main effect on the severity of stuttering behavior and not so much on the frequency of stuttering",

In sum, basal ganglia seem to play a double role in stuttering: (1) their activity seems to influence the risk of stuttering, as the medial premotor system has an impact on the allocation of attention, and (2) their activity seems to influence overt stuttering behaviors and their severity without being the underlying trigger.

2.4. Stuttering and the sensory feedback of breathing

If stuttering was caused by problems in the processing of auditory feedback only, then stuttering events at the onset, i.e., on the first syllable of an utterance would be impossible, but such symptoms are common at least in preschoolers who stutter (Buhr and Zebrowski, 2009; Richels et al., 2010). Stuttering at the onset of an utterance can be explained by taking the role of breathing into account. Respiration during speech differs from basal respiration: In basal respiration, inspiratory and expiratory phases are of similar duration; by contrast, during speech, inspiration is short and intensive, but expiration is slow and produces the constant air flow needed for speaking (Conrad and Schönle, 1979).

When speaking starts, respiration must switch from the basal mode to the speech mode. Further, inspiratory pauses must be included in a longer speech sequence such that (1) inspiration starts not before the preceding phrase or clause is complete, and that (2) the subsequent word starts not before inspiration is finished. To ensure the latter and to prevent premature phonation during inspiration, an internal automatic and widely unconscious monitoring mechanism might exist that inhibits a speech motor program to start as long as a preceding inspiratory movement has not been finished, similar as with the inhibition of a speech motor program after an error signal in the preceding word or phrase. That is, in the production of a speech sequence, a respiratory pause is treated like a speech unit.

The automatic monitoring of respiration might depend on the somatosensory feedback of the movement of thoracic and abdominal muscles (diaphragmatic movement is imperceptible). As with the processing of the auditory feedback of speech, also the processing of the feedback of breathing may require some degree of attention, i.e., a minimum of perceptual-

and processing capacity. If now the speaker's attention is very strained by other things, e.g., by speech planning at sentence onset, then it may happen that the feedback of breathing is insufficiently processed. This, again, may result in invalid error signals: The internal monitor fails to detect the end of inspiration (despite inspiration has been finished) and inhibits the subsequent speech motor program when it is starting. In this way, stuttering can come about at the onset of an utterance as well as at the onset of sentences or clauses after a respiratory pause.

One may argue that humans are quite able to speak during inspiration without an automatic mechanism preventing this. That is right, but humans are also able to deliberately mispronounce words or to deliberately make grammar or syntax mistakes without Levelt's (1995) Main Interruption Rule taking effect. Our internal automatic monitoring system is effective only as long as our behavior is controlled automatically; the person's conscious will can override this mechanism. However, our ability to speak fluently and correctly in terms of phonology, grammar, and syntax, and all this spontaneously, without thinking of rules, even without explicit knowledge of rules – this amazing ability is depending on automatic control and automatic self-monitoring.

Up to now, there is no evidence of a deficit in the processing of the somatosensory feedback of breathing in PWS. Chang et al. (2015) found reduced fractional anisotropy (FA) in the left supramarginal gyrus in children who stutter, as compared to normal fluent controls, and a negative correlation between FA and stuttering severity. Supramarginal gyrus is involved in the integration of the somatosensory feedback of articulation, breathing, and phonation (Simonyan and Horwitz, 2011). Hence, reduced FA possibly indicating a structural deficit in the white matter in this brain area could be related to poor involvement of the sensory feedback of breathing in speech control, specially because PWS are often focused on phonation and articulation, thus deficient processing of these sorts of feedback is rather unlikely).

Apart from that, a relationship between stuttering and respiration is suggested by the fact that some therapy programs include breathing exercises and -techniques in the treatment of stuttering, e.g., the Regulated Breathing Therapy (Azrin and Nunn, 1974; Conelea, Rice, and Woods, 2006). The effect that a change in breathing has on stuttering may result not so much from the breathing technique itself, as there might be many people who do not breathe in an ideal manner, but do not stutter. The effect may rather result from the necessity to pay attention to breathing during speech when applying such a technique. In Regulated Breathing, clients learn to start speaking following a small exhale (Conelea, Rice, and Woods, 2006), which is also an element in the Passive Airflow Technique (Falkowski, Guilford, and Sandler, 1982; Schwartz, 1974). After the present theory, this

method ensures that the switch from inspiration to expiration is not only happened but also perceived, such that invalid error signals are prevented.

2.5. Fluency-enhancing conditions operate by improving the processing of auditory feedback

In this section, the effect of some well-known conditions in which stuttering usually disappears or significantly decreases is explained by means of the attention allocation theory. Fluency-enhancing conditions can be divided into two groups: Group 1 comprises conditions in which no external auditory feedback is available, as in inner speech (verbal thinking), in silent mouthing, with auditory masking, or with hearing loss. Group 2 comprises conditions in which external auditory feedback is available and,(a) either the person is required to monitor whether his/her verbal output meets a target, e.g., in chorus reading, paced speech (metronome), and speech shadowing, or (b) the auditory feedback sounds odd or unfamiliar , e.g., in whispering, speaking with blocked ears, speaking in a disguised voice or in an unfamiliar dialect, or with altered auditory feedback (AAF).

Temporarily enhanced fluency in such conditions has often been interpreted as the result of either the elimination of auditory feedback or a distraction of the speaker's attention from it. The conclusion then has usually been that auditory feedback, or too much attention to it, is harmful for speech fluency in PWS (e.g., Max et al., 2004; Van Riper, 1973). In brain imaging studies, however, fluent speech of PWS induced, e.g., by chorus reading, paced speech, er singing was found to be associated with increased activation in auditory brain areas, that is, auditory under-activation typical of stuttered speech disappeared (see meta-analysis by Budde, Baron, and Fox, 2014). This suggests that auditory input was processed more intensively during induced fluency at least in the investigated conditions, which does not support the distraction hypothesis.

The present theory provides a coherent explanation for all fluency-enhancing conditions mentioned above: They prevent invalid error signals in the monitoring system resulting from poor processing of auditory feedback. In the conditions of Group 1, auditory feedback is not at all eliminated, but shifts to the internal auditory feedback loop (see Section 1.2.4). Internal auditory feedback of speech seems to be necessarily associated with sufficient attentional capacity (the thread of thoughts breaks off when you are distracted from thinking), thus invalid error signals due to insufficient feedback processing do not occur. In the conditions of Group 2, the speaker's attention is not distracted from, but drawn to his or her verbal output: Either (a) the speaker must continuously monitor whether his/her verbal output meets a target, e.g., in terms of pace (metronome), speech rate

(chorus reading), or wording (shadowing), which requires to focus not only on the target, but also on auditory feedback, or (b) attention is drawn to the auditory feedback because one's own speech sounds odd (DAF, FAF, speaking in a disguised voice or in an unfamiliar dialect) or is difficult to hear (whispering, incomplete auditory masking, blocked ears).

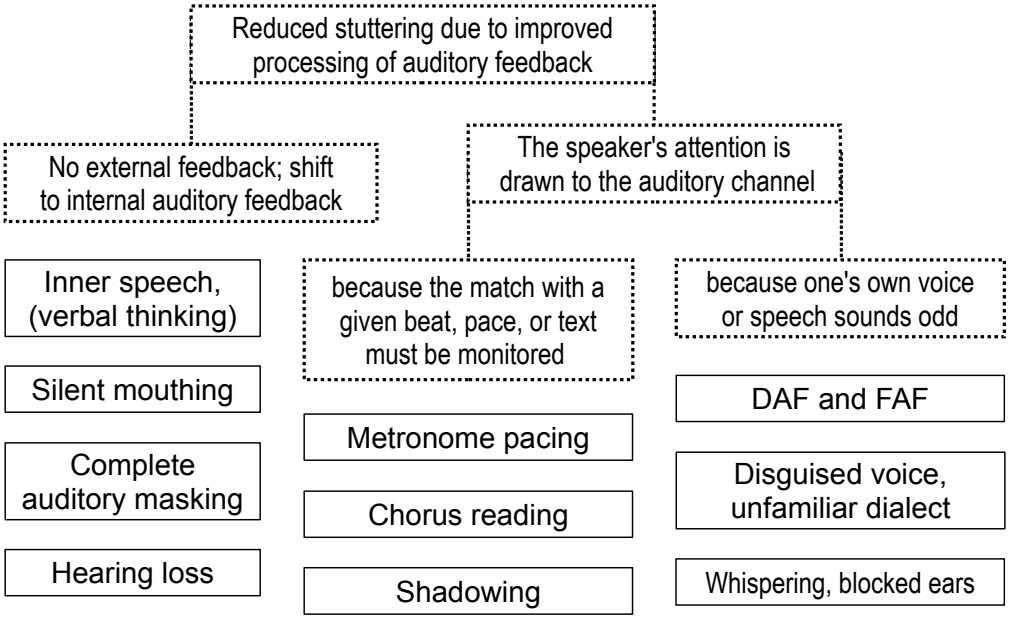

Figure 4: Fluency-inducing conditions

Particularly, speaking paced by a metronome and chorus reading have often been interpreted in the sense that PWS get external cues for syllable starts, which helps them producing a proper speech rhythm. However, PWS are well able to generate their own rhythm, e.g., in singing or in speaking with rhythmic hand movements. Further, paced speech and chorus reading cannot work in the way that the speaker each time waits for an external cue and responds to it: Because of the reaction time, the speaker will never be in sync with the metronome or the co-speakers, but always too late. Instead, the speaker must capture the given pace or speech rate such that he/she can predict and anticipate it. Then the speaker must adjust his/her own rhythm or rate and continuously monitor whether it is still in sync with that given by the metronome or the co-readers.

This view is supported by the result of an experiment conducted by Howell and El-Yaniv (1987): PWS were reading a short story (1) normally, (2) while listening to clicks of a metronome and adjusting their speech rhythm to it, and (3) while listening to clicks that sounded at the beginning of every syllable, triggered by the intensity of the speaker's voice. The third condition in which the participants spoke in their natural rate and rhythm

reduced stuttering nearly as effectively as the second condition: The mean number of disfluencies in the story was 20.25 in the normal condition, 0.6 with metronome, and 2.5 with click at syllable onset (standard deviations: 9.15, 1.35, and 2.01, resp). These results show that it is not mainly the given rhythm which reduces stuttering in the metronome condition, but rather the fact that attention is drawn to the auditory channel.

In speech shadowing, the third condition in Group 2a, the speaker does not get an external pace or rhythm for speaking in unison. Instead, he or she repeats what another person (leader speaker) says with a short time lag such that both are speaking at the same time, but not synchronously (e.g., Cherry and Sayers, 1956). In shadowing, one is required to listen to the leader's speech and, at the same time, to monitor whether one's own speech exactly follows. In all three conditions of Group 2a, the speaker is required to listen not only to the target but also to his/her own speech, and in this way, the processing of auditory feedback is improved, and invalid error signals in the monitoring system are prevented.

Regarding the conditions of Group 2b, Foundas et al. (2013) as well as Unger, Glück, and Cholewa (2012) examined the effect of electronic devices which reduce stuttering by altered auditory feedback (AAF, i.e., delayed and/or frequency-shifted auditory feedback, DAF/FAF). In both studies, speech fluency was found to be significantly improved by the devices even in a control condition without feedback alteration: It might have been somewhat unfamiliar to hear one's own voice not in the natural way, but through the device, and this may have drawn the participants' attention to the auditory channel. DAF and FAF may only increase this effect by making the feedback even more unfamiliar

Auditory feedback naturally can be altered also by speaking in an altered manner, e.g., in a disguised voice or in an unfamiliar dialect. In doing so, one will automatically listen to auditory feedback to hear whether the goal, the intended alteration of voice or dialect is reached. A response of spontaneous attentive listening also occurs when speech is difficult to hear. This may account for the fluency-enhancing effect of whispering, speaking with blocked ears (i.e., with external auditory feedback via bone conduction) and with incomplete masking. Generally, "any changes in stutterer's accustomed way of hearing themselves speak is likely to alleviate their speech difficulty." (Bloodstein and Bernstein Ratner, 2008, p. 301). The cause, after the present theory, is that every unfamiliar auditory feedback draws the speaker's attention to the auditory channel.

Two fluency-inducing conditions related to auditory feedback are not included in Figure 4: Singing and prolonged speech. In singing, probably several factors work at the same time: (1) Like the conditions of Group 2, also singing makes the person listening to his/her own voice in order to meet a target, the melody. (2) Memorized lyrics are stored in memory as

only one speech motor program (see Section 1.2.2). (3) Beat and rhythm may play a role, but, as in the metronome condition, not a very important one. (4) Syllables are often prolonged and evenly stressed in singing, As mentioned in Section 1.2.1, the duration of long-stressed syllables in normal speech is controlled by audio-phonatory coupling, i.e., the start of the next syllable depends on the auditory feedback of the start of the preceding long syllable (Kalveram and Jäncke, 1989). This mechanism might be effective in singing and even more in prolonged speech: When all syllables are long spoken, then all syllables (except the first one) are started on the basis of auditory feedback, which causes stuttering to disappear.

3. The development of stuttering

The preceding chapter was about the pathomechanism of stuttering; let us now look at the development of the disorder and at the factors involved – factors that trigger the mechanism, and those that contribute to its maintenance or overcoming, respectively. Recent investigation of young stuttering children near the onset of the disorder (Chow and Chang, 2017) showed some differences in brain structure between children who eventually recovered from stuttering and those who persisted, as groups, even at this early stage of development. This suggests that no gradual chronification of stuttering takes place; recovery or persistence seem to be rather predetermined, with external factors having little influence. In this chapter, it is therefore distinguished between the development of transient stuttering, on one hand, and of persistent stuttering, on the other hand, Psychogenic stuttering is dealt with in the section on transient stuttering, because the physical predisposition for stuttering is rather low in these cases, and appropriate psychotherapy often leads to complete recovery.

3.1. Transient developmental stuttering

Idiopathic stuttering is today regarded as a neurodevelopmental disorder, however, brain development especially in young children interacts with learning, i.e., with the development of behavioral routines and habits. An aspect of behavior often overlooked is the allocation of attention and, with that, of perceptual and processing capacity. For every more complex behavioral routine, and also for speaking, an adequate allocation of attention must be learned and automatized. In the present theory, it is assumed that just this rather unconscious aspect of behavioral control is the crucial factor in the development of stuttering. Figure 5 shows the development from onset to recovery (persistent stuttering is not included in this figure).

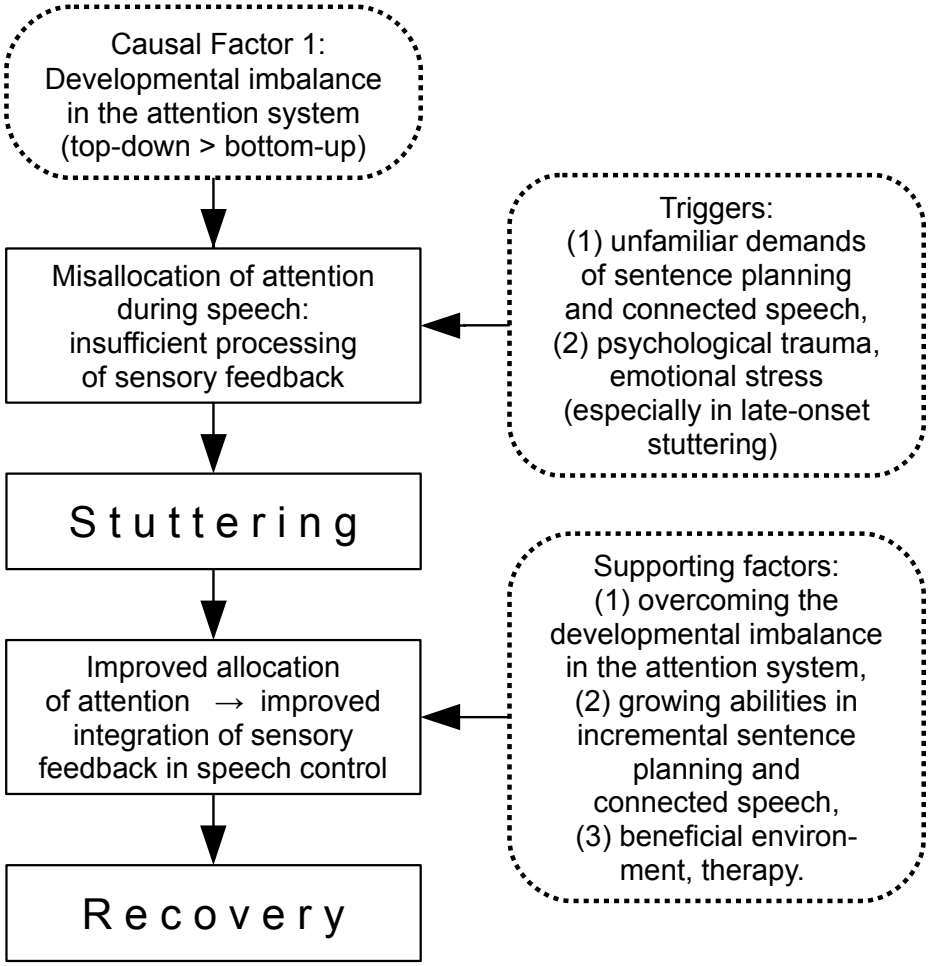

Figure 5: The development of transient stuttering. Causal and influencing factors are dashed-framed.

The cause of childhood stuttering, in most cases, seems to be an imbalance in the development of the attention system; a preponderance of attention to to action, i.e., to goals (top-down), to the detriment of the attention to passive perception, including the attention to sensory feedback (bottom-up). This becomes problematic in speech development at the change from one-word utterances to connected speech and sentence production, when auditory feedback and the sensory feedback of breathing must be more involved in speech control. The brain needs to know which constituents of a sentence have already been produced in order to correctly complete the sentence, and speech errors in a sentence must be detected immediately (which depends on auditory feedback; see Section 1.2.3). Breathing times must be included in the linguistic sequence at proper positions. Children naturally are not aware of changing behavior when they start forming sentences, and they cannot know that they need to change their attentional behavior.

There are some empirical findings suggesting an imbalance in the

attention system: first, a high number of dopamine D2 receptors at age 2.5 to 3 (Alm 2004; Rothmond et al. 2012). Alm (2004) has assumed that a high dopamine reception in the basal ganglia immediately causes stuttering, but the high dopamine level may rather manifest an imbalance in behavioral and attentional control which may cause a higher risk of stuttering at this age (see Section 2.3.4 about the role of the basal ganglia in stuttering).

A second,finding is a lower fractional anisotropy in the left arcuate fasciculus in children who stutter. as compared with their non-stuttering peers (Chow and Chang, 2017). This finding was, in part, more significant in children who eventually recovered from stuttering than in those who persisted (see Fig. 1, Clusters 1 and 2 in the study). Fractional anisotropy (FA) means the anisotropy of the diffusion of water molecules; it is a measure allowing to investigate the course of nerve fiber bundles in the brain and their structural features, e.g., the degree of myelination or 'maturation'. Arcuate fasciculus is a bundle of nerve fibers connecting sensory cortical areas with premotor and motor areas. Lower FA can be interpreted as a delay in fiber maturation which, in turn, can result from less frequent activation of the fibers when sensory information is less involved in motor planning and -control (a correlation between the activation of nerve fibers and FA was shown in several studies, e.g., Bengtsson et al., 2005; Keller and Just, 2009; Scholz et al., 2009).

A further finding suggesting an imbalance in the attention system is an atypical functional connectivity between neuronal networks in children who stutter as compared with normal developed children (Chang et al ., 2018). Functional connectivity between brain areas means that these areas tend to be activated or deactivated synchronously without the need of a structural connection. Particularly the hyperconnectivity between ventral attention network (VAN, mainly responsible for bottom-up attention) and default mode network (DMN; see Fig. 5A in the aforementioned study) suggests an imbalance in the control of attention. DMN is a set of brain regions that are deactivated during goal-directed tasks, thus hyperconnectivity between DMN and VAN means that also VAN is tendentially deactivated during goal-directed tasks. Since VAN is responsible for the automatic processing of sensory feedback, deactivation of VAN during goal-directed tasks means that sensory feedback tendentially is poorly processed during goal-directed tasks. Interestingly, Chang et al. (2018) even found a reduced functional connectivity in the visual network in stuttering children, suggesting a general deficit in the involvement of sensory input in the control of behavior.

The most important trigger for the onset of childhood stuttering, after the present theory, is the child's change to connected speech and sentence production. This idea is not new: Bloodstein (2006, p. 185) pointed to the facts that "early stuttering seldom occurs on one-word utterances; the

earliest age at which stuttering is reported is 18 months, with the beginning of grammatical development; the age at which most onset of stuttering is reported, 2-5 years, coincides with the period during which children acquire syntax; considerable spontaneous recovery takes place at the time most children have mastered syntax; incipient stuttering is influenced by the length and grammatical complexity of utterances..."

Late-onset stuttering and the so called psychogenic stuttering were included in the scheme for transient stuttering (Fig. 5) because the underlying mechanism might be the same as in developmental stuttering, but individuals affected seem to have no strong physical predisposition for stuttering, especially not for persistent stuttering. In a person whose attention system is susceptible, strong negative emotions, distress, fear, or the aftermath of a trauma may result in a misallocation of attention also during speech and by that in stuttering. Complete recovery is often reached in such cases by a supporting environment and/or by therapy, including psychotherapy (see Table 1 in Chang et al.,2010).

Spontaneous recovery from stuttering is probably caused by a kind of unconscious learning effect: Children eventually learn to adapt the allocation of their attention to the new demands of connected speech and sentence production. This learning effect manifests in brain structure, see, e.g., the upward developmental trajectories of FA in the recovered group in Chow and Chang (2017) in the Clusters 3, 5, and 6. This progress in the structural integrity of nerve fibers may result from learning, as it was shown that even a few weeks of practice, e.g., in reading or in juggling, can result in enhanced FA values in the white matter of involved brain regions (Keller and Just, 2009; Scholz et al., 2009).

The assumption that most stuttering children eventually learn to adapt their attentional allocation to the demands of connected speech and, by that, overcome the disorder does not mean that all things function completely in the same way as in children who have never stuttered. The findings obtained by Chang et al. (2008, 2018) showed some differences in brain structure and -function between children who had recovered from stuttering and their normal developed peers. However,, Chang et al. (2018) found the functional connectivity especially between,default mode network, attention networks, and executive control networks to be normalized in recovered children.

3.2. Persistent developmental stuttering

Given our assumptions about transient stuttering are true, then the question arises: Why does a minority of the children affected by stuttering not learn to adapt their attention to the demands of connected speech? A twofold

cause seems to be responsible for that: (1) an imbalance in the attention system, similar as in transient stuttering, and (2) a deficit in central auditory processing, probably not present in transient stuttering. Perhaps both factors are related to each other or interact with each other. The development of persistent stuttering is sketched in Figure 6:

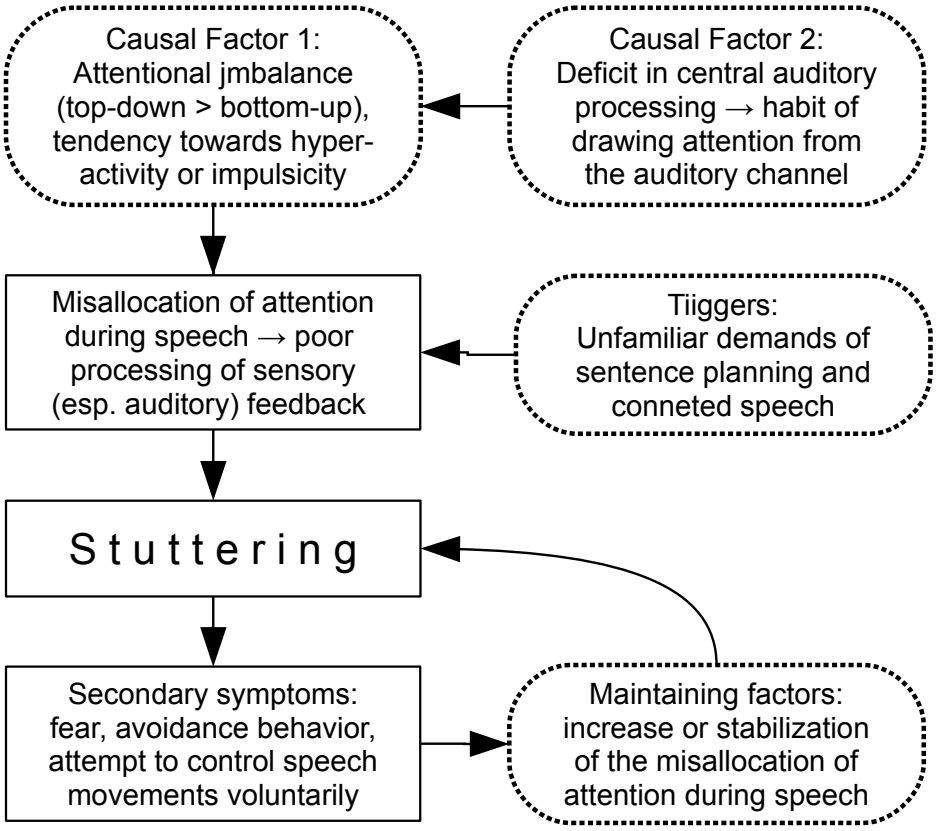

Figure 6: The development of persistent stuttering. Causal and influencing factors are dashed-framed.

3.2.1. Imbalance in the attention system

A finding by Chow and Chang (2017) which distinguished the group of stuttering children who eventually persisted from the group that eventually recovered is an initially (i.e., in the youngest children) higher FA value and its abnormal, stagnating or descending developmental trajectory in the thalamic radiation (see Fig. 2, Clusters 8, 9, and 10 in the study). These findings can hardly be explained as a consequence of stuttering because the differences are greatest in the youngest children. Since the thalamus plays a central role in attention regulation, the findings may be related to an abnormal development of the attention system.

Likewise, an anomalous functional connectivity between the attention networks and the default mode network (DMN) particularly in children

who eventually persisted in stuttering (Fig. 3 in Chang et al., 2018) indicates problems in the attention system in these children, however, the problems seem to partly differ from those in children with transient stuttering. Normal development proceeds in the way that other brain networks segregate from the DMN with age. Thus, the hyperconnectivity between VAN and DMN in children with transient stuttering (mentioned in the last section) can be interpreted as a delay in development which is partly overcome, partly compensated in the course of recovery. In children who persist in stuttering, by contrast, the connectivity between VAN and DMN and also between dorsal attention network and DMN is mainly reduced, suggesting a premature segregation of the networks in these children. This may have negative consequences for the further development of the attention system or for its integration in the control of behavior.

In addition, there are many behavioral studies in which deficits in attention regulation became apparent in PWS (see Alm, 2014; Anderson and Wagovich, 2010; Eggers, De Nil, and Van den Bergh, 2012; Ntourou, Anderson, and Wagovich, 2018 for overviews). Findings indicate that PWS, on average, are less efficient in attention regulation, less able to divide attention under dual-task conditions, less able to quickly shift attention and to suppress a planned motor action. For example, Eggers, De Nil, and Van den Bergh (2012) concluded that the orienting network important for the allocation of attention appears to be less efficient in children who stutter, and Kaganovich, Hampton Wray, and Weber-Fox (2010) concluded from an examination of auditory processing in preschool-aged children who stutter, that stuttering may be associated with less efficient attention allocation.

Further, PWS are more prone to exhibit attention disorders: The prevalence of attention-deficit hyperactivity disorder (ADHD) in school-age children in general is 3–6 %, depending on diagnostic criteria (Donaher, Healey, and Soffer, 2013). Data on the prevalence of ADHD in school-aged children who stutter range from 4 % to 26 % (Healey and Reid, 2003; Alm, 2014); Conture (2001) has suggested that 10–20 % of children who stutter might exhibit ADHD. Donaher and Richels (2012) found 58 % of children and adolescents who stuttered to show symptoms that could warrant a referral to a specialist for a possible ADHD diagnosis. Interestingly, both, persistent developmental stuttering and ADHD are more often diagnosed in males than in females (for ADHD, see Ramtekkar et al., 2010), which can be taken as a suggestion that the Factor 1 – especially by its components hyperactivity and impulsivity – is responsible for the greater prevalence of persistent stuttering in males

A central assumption of the attention allocation theory of stuttering is an imbalance between two components of behavioral control: between (1) internally initiated goal-directed action, controlled by the will, and associated with selective (top-down) attention, and (2) externally initiated re-action

based on sensory input including sensory feedback, and more depending on non-selective (bottom-up) attention. A finding consistent with that central assumption is a heightened FA in the frontal aslant tract linking SMA to Broca's area (Kronfeld-Duenias et al., 2016). The FA values of some subregions of the left frontal aslant tract were negatively correlated with speech fluency. Since the SMA is responsible for the voluntary control of behavior (after the model proposed by Goldberg, 1985; see also Section 2.3.4), we can interpret this finding as the manifestation of an overly strong influence of voluntary control on speech production in PWS. A further finding points to the same direction: Xuan et al. (2012) found increased activation in the dorsolateral prefrontal cortex in adult males who stutter. They conclude: "Increased activity or activation in these prefrontal areas appears to reflect increased attention to action or increased attempts to deal with stuttering."

3.2.2. Deficit in central auditory processing

Let us now look at the second, additional causal factor which seems to play a role in persistent stuttering: There are many empirical findings of deficits in the processing of linguistic acoustic stimuli in PWS (e.g., Beal et al., 2010, 2011; Chang et al., 2009; Maxfield et al., 2010, 2012; Neef et al., 2012; Tahaei et al., 2014), including the processing of auditory feedback (Cai et al., 2012, 2014a; Loucks, Chon, and Han, 2012; Natke, Grosser, and Kalveram, 2001; Salmelin et al., 1998), but also in the processing of non-linguistic acoustic stimuli (e.g., Arcuri, Schiefer, and Azevedo, 2017; Chang et al., 2009; Hampton and Weber-Fox, 2009; Howell et al., 2000; Howell, Davis, and Williams, 2006; Kikuchi et al., 2011, 2017: Saltuklaroglu et al., 2017).

Unfortunately, there are few data about differences between those who persisted in stuttering and those who recovered in terms of auditory processing. The only study I know is that by Howell, Davis, and Williams (2006). They compared children of both groups in a backward-masking task the performance of which is assumed to reflect the operation of central auditory processing. They found an appropriately 10 decibel higher mean backward-masking threshold in the persistent group, i.e., the probe tone needed to be about 10dB louder on average for these children to hear it prior to the masking noise. The group difference was statistically significant, but there was a high variability in the persistent group, thus the authors conclude that an auditory deficit may be sufficient, but not necessary, for the disorder to persist.

In a preceding study, Howell et al, (2000) had found a higher backward-masking threshold on average in stuttering children as compared to normal fluent controls, with that threshold being positively correlated with stuttering frequency in the stuttering group; that is, the poorer central auditory processing, the greater stuttering frequency. A correlation between

deficits in central auditory processing and stuttering frequency or severity was found also in other studies (Beal et al., 2010, 2011; Jansson-Verkasalo et al., 2014; Kikuchi et al., 2017; Liotti et al., 2010). Given that stuttering severity in children is (1) a predictor of eventual recovery or persistence (as concluded by Howell and Davis, 2011) and that (2) stuttering severity is related to central auditory processing, then we can assume that also auditory processing in young stuttering children can predict persistence or recovery. This hypothesis is consistent with the results of Howell, Davis, and Williams (2006) reported above and can be tested in a longitudinal study.

There are some further findings suggesting that a deficit in central auditory processing is the factor which distinguishes transient and persistent developmental stuttering: Usler and Weber-Fox (2015) as well as Mohan and Weber (2015) found differences between children with persistent stuttering and those who recovered in the processing of acoustically presented linguistic stimuli. Chow and Chang (2017) found a structural deficit (lower FA) in the splenium (the posterior part of the corpus callosum) in children who eventually persisted in stuttering, but not in those who eventually recovered (Fig. 1, Cluster 4 in the study). The affected fibers probably connect bilateral temporal regions (Kuvazeva, 2013), and lower FA may be related to less effective labor division between the hemispheres in auditory processing (callosal fibers often have inhibitory function, such that an activation in one hemisphere inhibits the homologous area in the other hemisphere).

The difference in FA in the mentioned Cluster 4 in Chow and Chang (2017) is great already with the youngest children, and there is no much overlap between the persistent group, on one hand, and the recovered and the control group, on the other hand. Therefore, the structural deficit in the splenium can hardly be a consequence of stuttering (while we can assume this when structural abnormalities grow with the duration of stuttering). Interestingly, Chow, Liu, Bernstein Ratner, and Braun found a strong relation between reduced FA in the splenium and stuttering severity in adults who stutter (unpublished study, presented at the 2014 ASHA Convention).

The theory that transient and persistent developmental stuttering are in core the same disorder, and that persistence is caused by one or more additional factors has already been proposed by Ambrose, Cox, and Yairi (1997) in a study of the genetic basis of persistence and recovery. They write: “ It was found that recovery or persistence is indeed transmitted, and further, that recovery does not appear to be a genetically milder form of stuttering, nor do the two types of stuttering appear to be genetically independent disorders. Data are most consistent with the hypothesis that persistent and recovered stuttering possess a common genetic etiology, and that persistence is, in part, due to additional genetic factors.”

4. Summary, consequences for treatment

The present attention allocation theory of stuttering makes the following assumptions:

1. Core symptoms of stuttering , i.e., repetition, prolongation, and silent block are caused by the inhibition of a speech motor program and the speaker's spontaneous attempt to continue despite the inhibition.
2. The inhibition of a speech motor program is caused by an invalid error signal: the monitoring system believes that a speech error has happened and interrupts the speech flow in order to enable a correction. This response is mainly controlled by the cerebellum.
3. The invalid error signals are triggered by insufficient processing of the sensory (mainly auditory) feedback of speech: Parts of the feedback are not transmitted or not stored in working memory.
4. Insufficient processing of sensory feedback is caused by a misallocation of attention, i.e., of perceptual and processing capacity during speech: Attention is too much focused on action (speech planning or motor control) to the detriment of the processing of auditory feedback and/or the sensory feedback of breathing.
5. The misallocation of attention is caused by a transient or persistent imbalance in the attention system (in the latter case often associated with hyperactivity or impulsivity) and, in persistent stuttering, by a deficit in central auditory processing.

More generally said: The speech production system as such is intact and able to work well in PWS, which becomes apparent by the fact that all of them speak fluently at times. Stuttering occurs when the system does not, or not in sufficient quality, get a special sort of sensory feedback. Fluency-enhancing conditions like chorus reading, altered auditory feedback, or – only a virtual paradox – auditory masking ensure that the system gets the feedback information needed (see Section 2.5). – If the proposed theory is true, then stuttering should be reduced most effectively by a change in the allocation of the speaker's attentional ressour-ces in order to improve the processing of auditory feedback and/or of the sensory feedback of breathing. That means:

- less attention to speech planning, e.g., by short sentences or short units of meaning with pauses between them; no substitutions;
- less attention to fear and avoidance of stuttering ,e.g., by desensitization and the acceptance of one's own stuttering;
- more attention to auditory feedback: listening to one's own voice and words, especially to the ends of words and to unstressed words;

- more attention to the sensory feedback of breathing: awareness of the coupling between exhalation and phonation (for those who often stutter at utterance onset or after breathing in).

Only a few practical possibilities can be mentioned here: Fluency-inducing means like chorus reading, shadowing, or altered auditory feedback can be used in order to practice listening during speech (see Section 2.5). In such a training, transcranial direct current stimulation (noninvasive) can support sustainable changes in the brain (Chesters, Möttönen, and Watkins, 2018). Another way to improve auditory attention in everyday talking is to try for speaking in a powerful and sonorous voice and for varied and expressive prosody, as all this needs to be monitored via auditory feedback. Besides, it should be investigated whether acoustically controlled auditory training aiming for an improvement of central auditory processing can help to improve speech fluency, as proposed by Prestes et al. (2016).

5. References

Alm, P. A. (2014). Stuttering in relation to anxiety, temperament, and personality: Review and analysis with focus on causality. *Journal of Fluency Disorders, 40*, 5–21.

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. *Journal of Communication Disorders, 37,* 325–369.

Ambrose, N. G., Cox, N. J., & Yairi, E. (1997). The genetic basis of persistence and recovery in stuttering. *Journal of Speech, Language, and Hearing Research, 40,* 567–580.

Anderson, J. D. & Wagovich, S. A. (2010). Relationships among linguistic processing speed, phonological working memory, and attention in children who stutter. *Journal of Fluency Disorders, 35,* 216–234.

Andrade, A. N. de, Gil, D., Schiefer, A. M., & Pereira, L. D.(2008). Behavioral auditory processing evaluation in individuals with stuttering. *Pró-Fono Revista de Atualização Científica, 20,* 43–49.

Arcuri, C. F., Schiefer, A. M., & Azevedo, M. F. (2017). Research about suppression effect and auditory processing in individuals who stutter. *Codas, 29,* e20160230.

Astheimer, L. B. & Sanders, L. C. (2012). Temporally selective attention supports speech processing in 3- to 5-year-old children. *Developmental Cognitive Neuroscience, 2,* 120–128.

Astheimer, L. B. & Sanders, L. C. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology, 80,* 23–34.

Azrin, N. H., & Nunn, R. G. (1974). A rapid method of eliminating stuttering by a regulated breathing approach. *Behaviour Research and Therapy, 12,* 279–286.

Bakheit, A. M. O. (2011). Remission of life-long stammering after posterior circulation stroke. *Neurocase, 17,* 41–45.

Beal, D. S., Quraan, M. A., Cheyne, D. O., Taylor, M. J., Gracco, V. L., & De Nil, L. F. (2011). Speech-induced suppression of evoked auditory fields in children who stutter. *NeuroImage, 54,* 2994–3003.

Beal, D. S., Cheyne, D. O., Gracco, V. L., Quraan, M. A., Taylor, M. J., & De Nil, L. F. (2010). Auditory evoked fields to vocalization during passive listening and active generation in adults who stutter. *NeuroImage, 52,* 1645–1653.

Bengtsson, S.L., Nagy, Z., Skare. S. Forsman, L., Forssberg, H., Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience, 8,* 1148–1150.

Bloodstein, O. (2006). Some empirical observations about early stuttering: a possible link to language development.
*Journal of Communication Disorders, 39*185–191.

Bloodstein, O. (1975). Stuttering as tension and fragmentation. In Eisenson (Ed.). *Stuttering: A second symposium* (pp. 1–95). New York: Harper & Row.

Bosshardt, H. G. (1999). Effects of concurrent mental calculation on stuttering, inhalation and speech timing. *Journal of Fluency Disorders, 24,* 43–72.

Braun, A. R., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J. M., Carson, R. F., & Ludlow, C. L. (1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering. *Brain, 120,* 761–784.

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE Meta-Analysis of functional neuroimaging studies. *Human Brain Mapping, 25,* 105–117.

Brumberg, J. S., Krusienski, D. J., Chakrabarti, S., Gunduz, A., Brunner, P., Ritaccio, A. L., & Schalk, G. (2016). Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. *PLoS ONE, 11,* e0166872.

Budde, K. S., Barron, D. S., & Fox, P. T. (2014). Stuttering, induced fluency, and natural fluency: A hierarchical series of activation likelihood estimation meta-analyses. *Brain & Language, 139,* 99–107.

Buhr, A. P. & Zebrowski, P. M. (2009). Sentence position and syntactic complexity of stuttering in early childhood: A longitudinal study. *Jounal of Fluency Disorders, 34,* 155–172.

Cai, S., Beal, D. S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2014). Impaired timing adjustments in response to time-varying auditory perturbation during connected speech production in persons who stutter. *Brain & Language, 129,* 24–29.

Cai, S., Beal, D. S., Ghosh, S.S., Tiede, M. K., Guenther, F. H., et al. (2012) Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *PLoS ONE, 7,* e41830.

Chang, S.-E., Angstadt M., Chow H. M., Etchell, A. C., Garnett, E. O., Choo, A. L., Kessler, D., Welsh, R. C., & Sripada, C. (2018). Anomalous network architecture of the resting brain in children who stutter. *Journal of Fluency Disorders, 55,* 46–57.

Chang, S.-E., Zhu, D.C., Choo, A. L., & Angstadt, M. (2015). White matter neuroanatomical differences in young children who stutter. *Brain, 138,* 694–711.

Chang, S.-E. & Zhu, D. C. (2013). Neural network connectivity differences in children who stutter. *Brain, 136,* 3709–3726.

Chang S.-E., Synnestvedt, A., Ostuni, J., & Ludlow, C. L. (2010). Similarities in speech and white matter characteristics in idiopathic developmental stuttering and adult-onset stuttering. *Journal of Neurolinguistics, 23,* 455–469.

Chang, S.-E., Kenney, M. K., Loucks, T. M., & Ludlow, C. L. (2009). Brain activation abnormalities during speech and non-speech in stuttering speakers. *NeuroImage, 15,* 201–212.

Chang, S.-E., Erickson, K. I., Ambrose, N. G., Hasegawa-Johnson, M. A., & Ludlow, C. (2008). Brain anatomy differences in childhood stuttering. *NeuroImage, 39,* 1333–1344.

Cherry, E. C., (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 15,* 975–979.

Cherry, E. C. & Sayers, B. M. (1956). Experiments upon the total inhibition of stammering by external control, and some clinical results. *Journal of Psychosomatic Research, 1,* 233–246.

Chesters, J., Möttönen, R., & Watkins, K. E. (2018). Transcranial direct current stimulation over left inferior frontal cortex improves speech fluency in adults who stutter. *Brain, 141,* 1161–1171.

Chow, H. M. & Chang, S.-E. (2017). White matter developmental trajectories associated with persistence and recovery of childhood stuttering during childhood. *Human Brain Mapping, 38,* 3354–3359.

Conelea, C. A., Rice, K. A., & Woods, D. W. (2006). Regulated breathing as a treatment for stuttering: a review of the empirical evidence. *Journal of Speech Language Pathology and Applied Behavior Analysis, 1.* 94–102.

Conrad, B. & Schönle, P. (1979). Speech and respiration. *Archiv für Psychiatrie und Nervenkrankheiten, 226,* 251–268.
Conture, E. G. (2001). *Stuttering: its Nature, Diagnosis, and Treatmant.* Needham Heights, MA: Allyn & Bacon.
Corbera, S., Corral, M.-J., Escera, C., & Idiazábal, M. A.(2005). Abnormal speech sound representation in persistent developmental stuttering. *Neurobiology, 65,* 1246–1252.
Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Sciences, 21,* e12521.
Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science, 17.* 149–195.
Dell, Gary S., & Reich, Peter A. (1981): *Stages in sentence* production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behaviour, 20.* 611–629.
De Nil, L. F., Kroll, R. M., Lafaille, S. J., & Houle, S. (2003). A positron emission tomography study of short- and long-term treatment effects on functional brain activation in adults who stutter. *Journal of Fluency Disorders, 28,* 357–380.
Donaher, J., Healey, E. C., & Soffer, S. (2013). *ADHD and Stuttering. Information Brochure*. Memphis, TN: The Stuttering Foundation of America.
Donaher, J. & Richels, C. (2012). Traits of attention deficit/hyperactivity disorder in school-age children who stutter.. *Journal of Fluency Disorders, 37*, 242–252.
Eggers, K., De Nil, L. F., & Van den Bergh, B. R. (2013). Inhibitory control in childhood stuttering. *Journal of Fluency Disorders, 38,* 311–313.
Eggers, K., De Nil, L. F., & Van den Bergh, B. R. (2012). The efficiency of attentional networks in children who stutter. *Journal of Speech, Language, & Hearing Research, 55,* 946–959.
Engelkamp, J. & Rummer, R. (1999). The architecture of the mental lexicon. In Friederici, A. D. (Ed.), *Language comprehension: A biological perspective* (pp. 133–174). Heidelberg: Springer.
Falkowski, G. L., Guilford, A. M., & Sandler, J. (1982). Effectiveness of a modified version of airflow therapy: case studies. *Journal of Speech and Hearing Disorders, 47*, 460–464.
Foundas, A. L., Mock, J. R., Corey, D. M., Golob, E. J., & Conture, E. G. (2013) The SpeechEasy device in stuttering and nonstuttering adults: Fluency effects while speaking and reading. *Brain & Language, 126,* 141–150.

Friederici, A. D. (1999). The neurobiology of language comprehension. In A. D. Friederici (Ed.), *Language comprehension: a biological perspective.* (2nd ed., pp. 265–292). Berlin: Springer.

Fromkin, V. A. (1971). The non-anomalo us nature of anomalous utterances. *Language 47,* 27–52.

Fox, P. T., Ingham, R. J., Ingham, J. C., Zamarrina, F., Xiong, J. H., & Lancaster, L. J. (2000). Brain correlates of stuttering and syllable production: a PET performance-correlation analysis. *Brain, 123,* 1985–2004.

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C. et al. (1996). A PET study of the neural systems of stuttering. *Nature, 382,* 158–162.

Goldberg, G. (1985). Supplementary motor area structure and function: Review and hypotheses. *The Behavioral and Brain Sciences, 8,* 567–616.

Halag-Milo, T., Stoppelman, N,, Kronfeld-Duenias, V., Civier, O., Amir, O., Ezrati-Vinacour, R., & Ben-Shachar, M. (2016). Beyond production: Brain responses during speech perception in adults who stutter. *NeuroImage: Clinical, 11,* 328–338.

Halle, M., & Stevens, K. N. (1959). Analysis by synthesis. In W. Wathen-Dunn & L. E. Woods (Eds.), *Proceedings of the seminar on speech compression and processing.* USAF Cambridge Research Center. 2: Paper D7.

Hampton, A. & Weber-Fox, C. (2009). Non-linguistic auditory processing in stuttering: Evidence from behavior and event-related brain potentials. *Journal of Fluency Disorders, 33,* 253–273.

Healey, E. C. & Reid, R. (2003). ADHD and stuttering: a tutorial. *Journal of Fluency Disorders, 28,* 79–93.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 60,* 407–422.

Holst, E. v. (1954). Relation between the central nervous system and the peripheral organs. *Britisch Journal of Animal Behavior, 2,* 89–94.

Howell, P. & Davis, S. (2011). Predicting persistence of and recovery from stuttering by the teenage years based on information gathered at age 8 years. *Journal of Developmental and Behavioral Pediatrics, 32*, 196–205.

Howell, P. Davis, S., & Williams, S. M. (2006). Auditory abilities of speakers who persisted, or recovered, from stuttering. *Journal of Fluency Disorders, 31,* 257–270.

Howell, P. & Williams, S. M. (2004). Development of auditory sensitivity in children who stutter and fluent children. *Ear and Hearing, 25,* 265–274.

Howell, P., Rosen, S., Hannigan, G., & Rustin, L. (2000). Auditory backward-masking performance by children who stutter and its relation to dysfluency rate. *Perceptual & Motor Skills, 90,* 355–363.
Howell, P., & El-Yaniv (1987). The effects of presenting a click in syllable-initial position on the speech of stutterers. Comparison with a metronome click. *Journal of Fluency Disorders, 12,* 249–256.
Hu, H., Liu, Y., Guo, Z., Li, W., Liu, P., Chen, S., & Liu, H. (2015). Attention modulates cortical processing of pitch feedback errors in voice control. *Science Reports, 5,* 7812.
Hugdahl, K., Thomsen, T., Ersland, L., Rimol, L. M., & Niemi, J. (2003). The effects of attention on speech perception: an fMRI study. *Brain and Language., 85,* 37–48.
Ingham, R. J., Fox, P. T., Ingham, J. C., Xiong, J., Zamarrina, F., Hardies, L. J., & Lancaster, J. L. (2004). Brain corellates of stuttering and syllable production: gender comparison and replication. *Journal of Speech, Language, and Hearing Research., 47,* 321–341.
Ingham, R. J., Ingham, J. C., Finn, P., & Fox, P. T. (2003). Towards a functional neural systems model of developmental stuttering. *Journal of Fluency Disorders, 28,* 297–318.
Jäncke, L., Mirzazade, S., & Shah, N. J. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neuroscience Letters, 266,* 125–128.
Jansson-Verkasalo, E., Eggers, K., Järvenpää, A., Van den Bergh, B., De Nil, L., & Kujala, T. (2014). Atypical central auditory speech-sound discrimination in children who stutter as indexed by the mismatch negativity. *Journal of Fluency Disorders, 41*, 1–11.
Jansson-Verkasalo, E., Eggers, K., Aro, K., De Nil, L. F., & Van den Bergh, B. R. (2012). Auditory attention shifting in children who stutter. *European Symposium on Fluency Disorders,* Poster retrieved from http://www.ecsf.eu/
Jiang, J., Lu, C., Peng, D., Zhu, C., & Howell, P. (2012). Classification of types of stuttering symptoms based on brain activity. *PLoS ONE, 7,* e39747.
Johnson, W. & Knott, J. R. (1937). Studies in the psychology of stuttering: I. The distribution of stuttering in successive readings of the same material. *Journal of Speech Disorders, 2,* 17–19.
Kaganovich, N., Hampton Wray, A., & Weber-Fox, C. (2010). Non-linguistic auditory processing and working memory update in pre-school children who stutter: an electrophysiological study. *Developmental Neuropsychology, 35,* 712–736.

Kalveram, K. T. & Jäncke, L. (1989). Vowel duration and voice onset time for stressed and nonstressed syllables in stotterers under delayed auditory feedback condition. *Folia Phoniatrica et Logopaedica, 41,* 30–42.

Karrass, J., Walden, T. A., Conture, E. G., Graham, C. G., Arnold, H. S., Hartfield, K. N., & Schwenk, K. A. (2006). Relation of emotional reactivity and regulation to childhood stuttering. *Journal of Communication Disorders, 39,* 402–423.

Kawahara, J., Sato, H., & Takenaka, I. (2011). Does stress enhance or impair selective attention? The effects of stress and perceptual load on distractor interference. Journal of Vision, 11, article 99.

Kell, C. A., Darquea, M., Behrens, M., Cordani, L., Keller, C., & Fuchs, S. (2017a). Phonetic detail and lateralization of reading-related inner speech and of auditory and somatosensory feedback processing during overt reading. *Human Brain Mapping, 38,* 493–508.

Kell, C. A., Neumann, K., Behrens, M., von Gudenberg, A. W., Giraud, A. L. (2017b). Speaking-related changes in cortical functional connectivity associated with assisted and spontaneous recovery from developmental stuttering. *Journal of Fluency Disorders, 55,* 135–144.

Keller, T. A. & Just, M. A. (2009). Altering cortical connectivity: remediation-induced changes in the white matter of poor readers. *Neuron, 5*, 624–631.

Kikuchi, Y., Okamoto, T., Ogata, K., Hagiwara, K., Umezaki, T., Kenjo, M., et al. (2017). Abnormal auditory synchronization in stuttering: A magnetoencephalographic study. *Hearing Research, 344,* 82–89.

Kikuchi, Y., Ogata, K., Umesaki, T., Yoshiura, T., Kenjo, M., Hirano, Y., et al. (2011). Spatiotemporal signatures of an abnormal auditory system in stuttering. *NeuroImage, 55,* 891–89.

Kronfeld-Duenias, V., Amir, O., Ezrati-Vinacour, R., Civier, O., & Ben-Shachar, M. (2016). The frontal aslant tract underlies speech fluency in persistent developmental stuttering. *Brain Structure and Function, 221,* 365–381.

Kuvazeva, M. G. (2013). Splenium of corpus callosum: patterns of interhemispheric interaction in children and adults. *Neural Plasticity,* article ID 639430.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.). *Cerebral mechanisms in behavior* (pp. 112–146). New York: Wiley.

Lee, B. (1951). Artificial Stutter. *Journal of Speech and Hearing Disorders, 16,* 53–55.

Levelt, W. J. M. (1995). *Speaking. From Intention to Articulation* (4th ed.). Cambridge: MIT Press, (Chapter 12)

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition,14,* 41–104.

Liotti, M., Ingham, J. C., Takai, O., Paskos, D. K., Perez, R., & Ingham, R. J. (2010). Spatiotemporal dynamics of speech sound perception in chronic developmental stuttering. *Brain and Language, 115,* 141–147.

Loucks, T.,, Chon, H. & Han, W. (2012). Audiovocal integration in adults who stutter. *International Journal of Language and Communication Disorders, 47,* 451–456.

Lu, C., Peng, D., Chen, C., Ning, N., Ding, G., Li, K., Yang, Y., & Lin, C. (2010). Altered effective connectivity and anomalous anatomy in the basal ganglia-thalamocortical circuit of stuttering speakers. *Cortex, 46,* 49–67.

Lu, C., Ning, N., Peng, D., Ding, G., Li, K., Yang, Y.,& Lin, C. (2009). The role of large-scale neural interactions for developmental stuttering. *Neuroscience, 161,* 1008–1026.

Maraist, J. A. & Hutton, C. (1957). Effects of auditory masking upon the speech of stutterers. *Journal of Speech and Hearing Disorders, 22,* 385–389.

Markett, S., Bleek, B., Reuter, M., Prüss, H., Richardt, K., Müller, T., et al. (2016). Impaired motor inhibition in adults who stutter – evidence from speech-free stop-signal reaction time tasks. *Neuropsychologia, 91,* 444–450.

Martin, S., Brunner, P., Holdgraf, C., Heinze, H. J., Crone, N. E., Rieger, J., Schalk, G., Knight, R. T., & Pasley, B. N. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers of Neuroengineering, 7,* article 14.

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemporary Issues in Communication Science and Disorders, 31,* 105–122.

Maxfield, N. D., Olsen, W. L., Kleinman, D., Frisch, S. A., Ferreira, V. S., & Lister, J. J. (2016). Attention demands of language production in adults who stutter.*Clinical Neurophysiology, 127,* 1942–1960.

Maxfield, N. D., Pizon-Moore, A. A., Frisch, S. A., & Constantine, J. L. (2012). Exploring semantic and phonological picture-word priming in adults who stutter using event-related potentials. *Clinical Neurophysiology, 123,* 1131–1146.

Maxfield, N. D., Huffman, J. L., Frisch, S. A., & Hinckley, J. J. (2010). Neural correlates of semantic activation spreading on the path to picture naming in adults who stutter. *Clinical Neurophysiology, 121,* 1447–1463.

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain, 119,* 907–917.
McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., Frackowiak, R. S., & Frith, C. D. (1996). The neural correlates of inner speech and auditory verbal imagery in schizophrenia: relationship to auditory verbal hallucinations, *The British Journal of Psychiatry, 169,* 148–159.
Meringer, R. & Mayer, C. (1895). Versprechen und Verlesen. Eine Psychologisch-Linguistische Studie. New edition by A. Cutler & D. Fay (1978). Amsterdam Studies in the theory and history of linguistic science (Vol. 2). Amsterdam: John Benjamins.
Metzger, F. L., Auer, T., Helms, G., Paulus, W., Frahm, J., Sommer, M., & Neef, N. E. (2018). Shifted dynamic interactions between subcortical nuclei and inferior frontal gyri during response preparation in persistent developmental stuttering. *Brain Structure and Function, 223,* 165–183.
Miller, A. E. (1985). Cessation of stuttering with progressive multiple sclerosis. *Neurology, 35,* 1341–1343.
Mohan, R. & Weber, C. (2015). Neural systems mediating processing of sound units of language distinguish recovery versus persistence in stuttering. *Journal of Neurodivelopmental Disorders, 7,* 28.
Näätänen, R., Paavilainen, P., Rinne, T., & Albo, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118,* 2544–2590.
Natke, U., Sandrieser, P., van Ark, M., Pietrowsky, R., & Kalveram, K. T. (2004). Linguistic stress, within-word position, and grammatical class in relation to early childhood stuttering. *Journal of Fluency Disorders, 29,* 109–122.
Natke, U., Grosser, J. & Kalveram. K.-T. (2001). Fluency, fundamental frequence, and speech rate under frequency-shifted auditory feedback in stuttering and non-stuttering persons. *Journal of Fluency Disorders, 26,* 227–241.
Neef, N. E., Sommer, M., Neef, A., Paulus, W., Gudenberg, A. W. v., Jung, C., & Wüstenberg, T. (2012). Reduced speech perceptual acuity for stop consonants in individuals who stutter. *Journal of Speech, Language, and Hearing Research, 55,* 276–289.
Neumann, K., Euler, H. A., Gudenberg, A. W. v., Giraud, A.-L., Laufermann, H., Gall, V., & Preibisch, C. (2003). The nature and treatment of stuttering as revealed by fMRI. A within- and between-group comparison. *Journal of Fluency Disorders, 28,* 381–410.
Nooteboom, S. & Quene, H. (2015). Word onsets and speech errors. Explaining relative frequencies of segmental substitutions. *Journal of Memory and Language, 78,* 33– 6.

Ntourou, K., Anderson, J. D., & Wagovich, S. A. (2018). Executive function and childhood stuttering: Parent ratings and evidence from a behavioral task. *Journal of Fluency Disorders, 56, 18–32.*

Packman, A. & Attanasio, J. S. (2004). *Theoretical issues in stuttering.* Hove: Psychology Press (Taylor & Francis).

Palmer, E. D., Rosen, H. J., Ojemann, J. G., Buckner, R. L., Kelley, W. M., & Petersen, S. E. (2001). An event-related fMRI study of overt and covert word stem completion. *NeuroImage, 14,* 182–193.

Poeppel, D, & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language, Cognition, and Neuroscience, 26,* 935–951.

Postma, A. & Kolk, H. (1993). The Covert Repair Hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech & Hearing Research, 36,* 472–487.

Postma, A. & Kolk, H. (1992). Error monitoring in people who stutter: evidence against auditory feedback defect theories. *Journal of Speech and Hearing Research, 35,* 1024–1032.

Preibisch, C., Neumann, K., Raab, P., Euler, H. A., von Gudenberg, A. W., Laufermann, H., & Giraud, A.-L. (2003). Evidence for compensation for stuttering by the right frontal operculum. *NeuroImage, 20,* 1356–1364.

Prestes, R., Neves, A. A., Santos, R. B., Marangoni, A. T., Schiefer, A. M., & Gil, D. (2017). Temporal processing and long-latency auditory evoked potential in stutterers. *Brazilian Journal of Otorhinolaryngology, 83,* 142–146.

Price, J. C., Wise, R. J. S., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K. et al. (1996). Hearing and saying. The functional neuro-anatomy of word processing. *Brain, 119,* 919–931.

Ramtekkar, U. P., Reiersen, A. M., Todorov, A. A., & Todd, R. D. (2010). Sex and age differences in Attention-Deficit/Hyperactivity Disorder symptoms and diagnoses: Implications for DSM-V and ICD-11. *Journal of the American Academy of Child & Adolescent Psychiatry, 49,* 217–228.

Richels, C., Buhr, A., Conture, E., & Ntourou, K. (2010). Utterance complexity and stuttering on function words in preschool-age children who stutter. *Journal of Fluency Disorders, 35,* 314–331.

Rothmond, D. A., Weickert, C. S., & Webster, M. J. (2012). Developmental changes in human dopamine neurotransmission: cortical receptors and terminators. *BMC Neuroscience, 13,* 18.

Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in

speech perception. *NeuroImage, 39,* 1444–1456.

Salmelin, R., Schnitzler, A., Schmitz F, & Freund, H. J. (2000). Single word reading in developmental stutterers and fluent speakers. *Brain, 123,* 1184–1202.

Salmelin, R., Schnitzler, A., Schmitz F., Jäncke, L., Witte, O. W., & Freund, H. J. (1998). Functional organization of the auditory cortex is different in stutterers and fluent speakers. *NeuroReport, 9,* 2225–2229.

Saltuklaroglu, T., Harkrider, A. W., Thornton, D., Jenson, D., & Kittilstved, T. (2017). EEG Mu (μ) rhythm spectra and oscillatory activity differentiate stuttering from non-stuttering adults. *NeuroImage, 153*, 232–245.

Scheerer, N. E., Tumber, A. K., & Jones, J. A. (2016). Attentional demands modulate sensorimotor learning induced by persistent exposure to changes in auditory feedback. *Journal of Neurophysiology, 115,* 826–832.

Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white matter architecture. *Nature Neuroscience, 12,* 1370–1371.

Schwartz, M. F. (1974). The core of the stuttering block. *Journal of Speech and Hearing Disorders, 39*, 169–177.

Segawa, J. A., Tourville, A., Beal, D. S., & Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *Journal of Cognitive Neuroscience, 27,* 819–831.

Seidler, R. D., Kwak, Y., Fling, B. W., & Bernard, J. A. (2013). Neurocognitive mechanisms of error-based motor learning. *Advances in Experimental Medicine and Biology*, *782*, 10.1007/978–1–4614–5465–6_3.

Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E. Jr., Murray, R. M., & McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping, 16,* 219 – 227.

Simonyan, K. & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *Neuroscientist, 17*, 197–208.

Sheeban, J. G. (1974). Stuttering behavior: A phonetic analysis. *Journal of Communication Disorders, 7,* 193–212.

Stager, S. V., Jeffries, K. J., & Braun, A. R. (2003). Common features of fluency-evoking conditions studied in stuttering subjects and controls: an H215O PET study. *Journal of Fluency Disorders, 28,* 319–336.

Sternberg, S. Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech.

Phonetica, 45, 175–197.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: additional evidence. In R. A. Cole (ed.) *Perception and production of fluent speech. Fourteenth Annual Carnegie Symposium of Cognition (pp. 507–534),* Hillsdale, NJ: Lawrence Erlbaum.

Stromsta, C. (1959). Experimental blockage of phonation by distorted sideton. *Journal of Speech and Hearing Research, 2,* 286–301.

Tahaei, A. A., Ashayeri, H., Pourbakht, A., & Kamali, M. (2014). Speech evoked auditory brainstem response in stuttering.*Scientifica (Cairo), a*rticle ID 328646.

Tani, T. & Sakai, Y. (2010). Stuttering after right cerebellar infarction: a case study. *Journal of Fluency Disorders, 35,* 141–145.

Taylor, I. K. (1966). The properties of stuttered words. *Journal of Verbal Learning and Verbal Behavior, 5,* 112–118.

Tian, X. & Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience, 6,* article 314.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forwardmodels. *Frontiers in Psychology, 1,* article 166.

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex, 77,* 1–12.

Tourville, J. A., Cai, S., & Guenther, F. H. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Journal of the Acoustic Society of America, 133,* article 3564.

Toyomura, A., Fujii, T., & Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. *NeuroImage, 57,* 1507–1516.

Tumber, A. K., Scheerer, N. E., & Jones, J. A. (2014). Attentional demands influence vocal compensations to pitch errors heard in auditory feedback. *PLoS ONE, 9,* e109968.

Unger, J. P., Glück, C. W., & Cholewa, J. (2012). Immediate effects of AAF devices on the characteristics of stuttering: a clinical analysis. *Journal of Fluency Disorders, 37,* 122–134.

Usler, E. & Weber-Fox, C. (2015). Neurodevelopment for syntactic processing distinguishes childhood stuttering recovery versus persistence. *Journal of Neurodevelopmental Disorders, 7,* 4.

Van Riper, C. G. (1973). *The Treatment of Stuttering.* Englewood Cliffs, N.J.: Prentice Hall.

Van Riper, C. (1982). *The Nature of Stuttering (2nd ed.*). Englewood Cliffs,

NJ: Prentice-Hall.

Van Riper, C. & Hull, C. J. (1955). The qualitative measurement of the effect of certain situations on stuttering. In W. Johnson & R. R. Leutenegger (Eds.) *Stuttering in Children and Adults*. Minneapolis: University of Minnesota Press.

Vasic, N. & Wijnen, F. (2005). Stuttering as a monitoring deficit. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 226–247). Hove: Psychology Press.

Watkins, K. (2011). Developmental disorders of speech and language: From genes to brain structure and function. [Review]. *Progress in Brain Research, 189,* 225–238.

Watkins, R. V., Yairi, E., & Ambrose, N. G. (1999). Early childhood ctuttering III: initial status of expressive language abilities. *Journal of Speech, Language, & Hearing Research, 42,* 1125–1135.

Webster, R. L. & Lubker, B. B. (1968). Interrelationsships among fluency producing variables in stuttered speech. *Journal of Speech and Hearing Research, 11*, 754–768.

Wymbs, N. F., Ingham, R. J., Ingham, J. C., Paolini, K. E., & Grafton, S. T. (2013). Individual differences in neural regions functionally related to real and imagined stuttering. *Brain and Language, 12,* 153–164.

Xuan, Y., Meng, C., Yang, Y., Zhu, C., Wang, L., Yan, Q., Lin, C., & Yu, C. (2012). Resting-state brain activity in adult males who stutter. *PLoS One, 7*, e30570.

Yairi, E. (1996). Applications of disfluencies in measurements of stuttering. *Journal of Speech and Hearing Research, 39,* 402–404.

Yang, Y., Jia, F., Siok, W. T., & Tan, L. H. (2016). Altered functional connectivity in persistent developmental stuttering. *Scientific Reports, 6,* article 19128.

Zheng, Z. Z., Vicente-Grabovetsky, A., MacDonald, E. N., Munhall, K. G., Cusack, R., & Johnsrude, I. S. (2013). Multi-voxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech. *The Journal of Neuroscience, 33,* 4339–4348.

Zheng, Z. Z., Munhall, K. G., Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience, 22,* 1770–1781.