Weak Vestibular Response in Persistent Developmental Stuttering: Implications for Own Voice Identification

Max Gattie[1], Elena Lieven[2], Karolina Kluk[1]*

[1] Manchester Centre for Audiology & Deafness (ManCAD), University of Manchester, Oxford Road, Manchester, M13 9PL, UK

[2] LuCiD (the ESRC International Centre for Language and Communicative Development), University of Manchester, Oxford Road, Manchester, M13 9PL, UK

*** Correspondence:**
Corresponding Author
max.gattie@manchester.ac.uk

Keywords: Stuttering[1], vestibular[2], efference copy[3], speech-motor control[4], own voice identification[5], speech perception[6], auditory scene analysis[7], speech-induced suppression[8].

Abstract

Speech-motor and psycholinguistic models employ feedback control from an auditory stream corresponding to own voice. Such models underspecify how own voice is identified. It is proposed that own voice is identified through coincidence detection between the neural firing rates arising from deflection of cochlear and vestibular mechanoreceptors by the sound and vibration generated during vocalisation. The coincidence detection is proposed to differ in people who stutter. In an update to the approach-avoidance conflict model of Sheehan (1953, 1975) instances of stuttering are proposed to coincide with uncertainty over an ongoing speech act. Discussion covers speech-induced suppression, auditory scene analysis, and theories of mental content.

1. Introduction

Speech-motor and psycholinguistic models describe a feedforward system in which articulatory muscles receive coordinated nerve impulses with sufficient detail to generate speech sounds (e.g. Hickok & Poeppel, 2007; Levelt et al., 1999; Tourville & Guenther, 2011). Typically they employ feedback control as a check for error (Helmholtz, 1886; von Holst & Mittelstädt, 1950; Fairbanks, 1954). Predictive feedback control avoids instability due to timing delay by checking for sensory error against a forward model of the speech-motor plan (see review in Parrell & Houde, 2019). Errors checked for might include articulatory malfunction, or mismatch between spoken and intended message – the nature of the error checked for will vary, depending on the nature of the model.

Such models underspecify how an auditory stream corresponding to own voice is identified (i.e. an auditory stream defined as per Bregman, 1990). A typical requirement is that a mental representation of expected auditory consequences is referred to, or is already identical with, an auditory target map (O'Callaghan, 2015). The question arises of how such reference is managed in the opposite direction – how an auditory target map for own voice is created from ambient sound and vibration.

Greater understanding of own voice identification could improve speech-motor and psycholinguistic models. For example, previously overlooked activity in the auditory brainstem and periphery may explain otherwise intractable difficulties in understanding the cerebral and cerebellar activity accompanying speech and language. Such an approach is taken in the current article. A hypothesis is formulated for own voice identification. The hypothesis is then developed to provide an account of stuttering, a DSM-V diagnosis characterised by involuntary prolongations and repetitions during speech.

The article will proceed as follows. Section 2 will describe the hypothesis of own voice identification. Section 3 will build on the hypothesis of section 2 to present a novel account of stuttering, REMATCH (Reflexivity and Communicative Mismatch). Section 4 will provide discussion of themes arising from sections 2 and 3. In this way, the article will extend from a biophysical account of own voice identification, to a psychosocial account of interpersonal communication. It will progress from audiology, to speech-motor theory, to psycholinguistics and social psychology.

Hypothesis formulation follows inference to the best explanation (Lipton, 2004). Best explanation arguments are mutually supportive. In other words, if one has a best explanation argument of T, and one has a best explanation argument of D, it follows that one has a best explanation argument of (T + D). This pertains even if D is partially reliant on T. This system (sometimes referred to as abduction) differs from, for example, multiplicative combination of probabilities in which the combined probability is lower than either of its constituents. Refuting a best explanation argument requires presentation of a better explanation. The discussion in section 4 will summarise the scope of the best explanation argument. To aid that discussion, hypotheses will be presented following the Methodology of Scientific Research Programmes described by Lakatos (1970). This refers to a “hard core” of (generally unfalsifiable) hypotheses, along with a “protective belt” of testable auxiliary hypotheses. Distinction will also be made between the two kinds of causal explanation described by Botterill (2010). Process explanations are of how something happens, whereas contrastive explanations are of why something happens. These two kinds of explanation interact as understanding of causation is acquired and enhanced.

2. Hypothesis of Own Voice Identification

2.1 Explanatory target

Own voice identification is a specific instance of the cocktail party problem (Bee & Micheyl, 2008), an outstanding issue in auditory scene analysis in which there is no principled basis for discrimination in a multi-talker scenario. It is an example of an ill-posed problem (Hadamard 1902, 1923; Poggio & Koch, 1985), sometimes referred to as an inverse problem, in which there is no mathematically unique solution.

2.2 Candidate explanations

There is no prior research offering a basis by which an own voice auditory stream is specifically distinguished from ambient sound and vibration (Shamma & Micheyl, 2010; Remez & Thomas, 2013; Bronkhorst, 2015). The most closely related literature emphasises the importance of body conducted vibration during own speech (von Békésy, 1949; Maurer & Landis, 1990; Pörschmann, 2000; Sohmer & Freeman, 2001; Shuster & Durrant, 2003; Reinfeldt et al., 2010; Meekings et al., 2015) or else describes

self talk and private speech through a Vygotskian developmental perspective (e.g. Fernyhough & Russell, 1997; Atencio & Montero, 2009; Lupyan & Swingley, 2012).

There is also a large body of work about the role of own voice in speech monitoring systems (e.g. Postma, 2000; Buschbaum, 2001; Ozdemir et al., 2007; Huettig & Hartsuiker, 2010; Nozari et al., 2011; Lind et al., 2014; Acheson & Hagoort, 2014; Kröger et al., 2016) or sensory-motor integration (e.g. Jürgens, 2002; Kaplan et al., 2008; Rosa et al., 2008; Zheng et al., 2010; Hickok et al., 2011; Behroozmand et al., 2015; Houde et al., 2015). This literature takes as a starting point that own voice has already been identified as an ascending auditory stream. It therefore does not address the current explanatory target. Literature concerning sensory-motor integration, and in particular the hypothesis of speech-induced suppression, will be discussed in section 2.4.1.

2.3 A Novel Hypothesis of Own Voice Identification

2.3.1 Introduction

The nature of the speech auditory brainstem response (BinKhamis et al., 2019) suggests that neural activity corresponding to identification of own voice could occur in the auditory brainstem. The auditory brainstem is innervated through the VIII cranial nerve, from bipolar ganglion cells which interface with mechanoreceptors of the inner ear. Neural activity corresponding to own voice could occur at the earliest within the bipolar ganglion cells of the ear itself.

Inner ear structure is common across mammals, consisting of an osseous labyrinth lined with sensory epithelium, and with several chambers. One of the chambers is the cochlea, a coiled structure containing mechanoreceptors which are deflected by ambient sound frequencies ranging from 20 Hz – 20,000 Hz in humans (Manley & Gummer, 2017). Other chambers comprise the vestibular system. These chambers include semicircular canals, in which mechanoreceptors are deflected by changes in angular velocity. There are also gravitoinertial otoliths, arranged such that mechanoreceptors are deflected by changes in linear velocity, and with resting state deflection corresponding to head orientation (Goldberg, 2012).

The traditional discrimination just described, of cochlear and vestibular chambers into hearing and equilibrial functions, is misleading (Tait, 1932). As for other vertebrates, mammalian otolithic receptors are deflected by vibration as well as by changes in body velocity or orientation relative to a fixed gravitational field. The vestibular system in mammals responds to vibrational frequencies up to 1,000 Hz, and may phase lock to higher frequencies (Curthoys et al., 2019).

Vestibular sensitivity is considerably greater to vibrations conducted through the body (BC) than to sound waves in air (AC). Electrophysiological studies show that when human responses of vestibular origin are referenced to a 60 dBA sound level typical of conversational speech, AC thresholds are 10 dB above baseline and BC thresholds 25 dB below baseline (McNerney & Burkard, 2011; Welgampola, Rosengren, Halmagyi & Colebatch, 2003). The act of speaking will deflect vestibular mechanoreceptors in humans (Todd, Rosegren & Colebatch, 2008; Curthoys, 2017; Curthoys et al., 2019).

2.3.2 Concurrency Hypothesis

The core hypothesis is that own voice is identified as an auditory stream through coincidence detection between vestibular and cochlear afferents. This will henceforth be referred to as the Concurrency Hypothesis.

The Concurrency Hypothesis describes a biologically grounded mechanism. The biological grounding is that there are two sets of mechanoreceptors for own voice. Figure 1 gives an overview of relevant details. Sound and vibrational energy deflecting sterocilia in cochlear hair cells corresponds to own voice

mixed with ambient environmental sounds. Concurrently, vibrational energy deflecting stereocilia in vestibular hair cells corresponds to own voice in isolation. Comparison of nerve impulses arising from cochlear and vestibular mechanoreceptors therefore provides a principled distinction between self and environment.

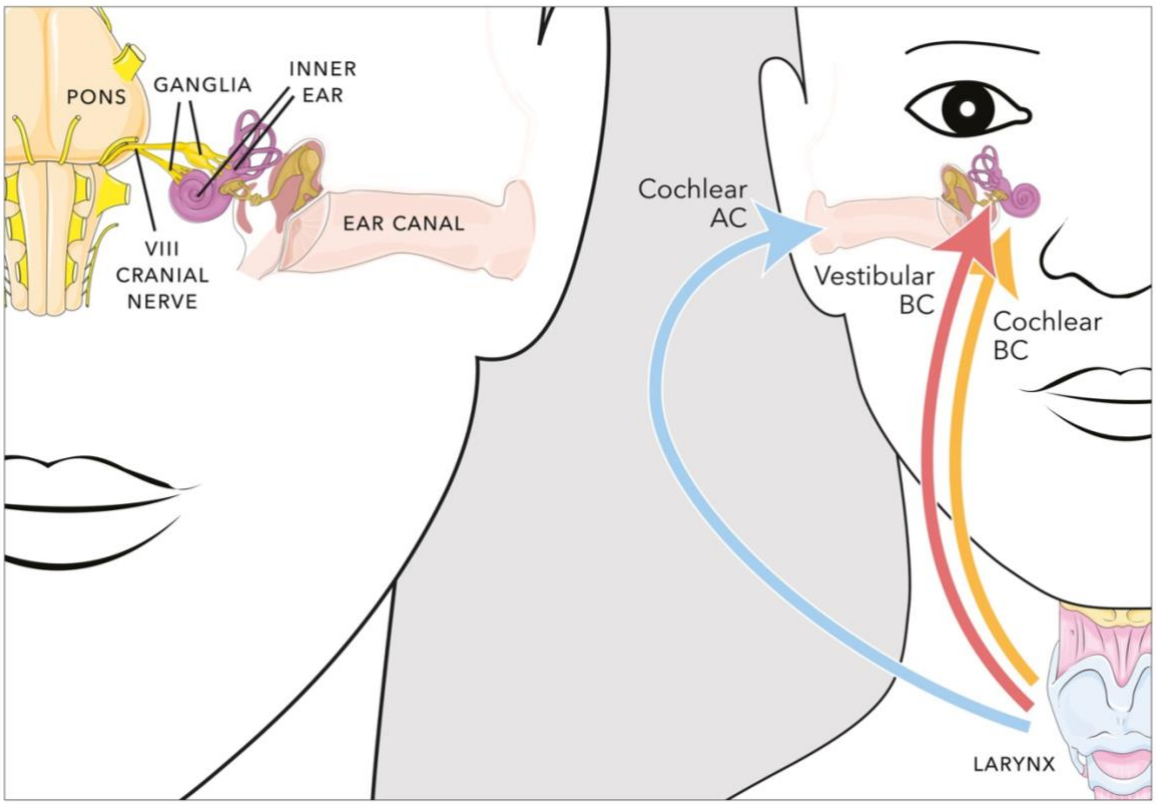

Figure 1: Parameters affecting sound and vibration detection in the human ear. In vivo measurements are difficult, and estimates here are derived from primary sources where possible. For general background on sound source perception, see Yost et al. (2008); for hair cells see Eatock et al. (2006); for voice production see Titze (1994); and for propagation of sound and vibration see Fahy & Thompson (2015).

Left hand side: Anatomical parameters. Dendrons of bipolar ganglion cells terminate on sensory epithelial hair cells in the inner ear. Axons from the ganglia project or branch through the VIII cranial nerve to nuclei of the pons and medulla, and (for some axons from vestibular ganglia) the cerebellum. Sensory hair cells fire continuously, with changes in firing rate following deflections due to sound, vibration and movement. Changes in firing rate will in turn modify long-term potentiation of brainstem and cerebellar nerve cells.

Right hand side: Acoustic and vibrational parameters. During vocalisation, sound and vibration energy originates predominantly at the larynx (and occasionally higher in the vocal tract; Titze, 1994). Energy propagates via two routes to each ear: air conduction (AC) through air surrounding the head, or body conduction (BC) through the neck and head. The inner ear includes cochlear and vestibular sensory hair cells. Sounds are perceived when AC and BC stimulation above hearing threshold (by definition zero dB HL or higher) deflects stereocilia in cochlear hair cells, opening mechanically gated ion channels which set off a chain of activity culminating in release of neurotransmitters, which in turn will raise potentials in dendrites of ganglion cells belonging to the VIII cranial nerve. Deflection of stereocilia in vestibular hair cells requires a considerably higher stimulus level than that for sterocilia in cochlear hair cells. Welgampola et al. (2003) established electrophysiological vestibular thresholds (VEMPs) at sound

levels, as defined at the cochlea, of 31 dB HL for BC stimulation, and 87 dB HL for AC stimulation. Even after adjusting for temporal integration with the brief duration stimuli used in electrophysiological testing, AC vestibular thresholds are 10 dB above, and BC vestibular thresholds 25 dB below, the 60 dBA sound level typical of conversational speech (McNerney & Burkard, 2011). Thus, own voice is either not detected or is very weakly detected via an AC vestibular route. Whereas, unless using alaryngeal speech such as whispering, own voice will consistently be detected by a BC vestibular route. This BC vestibular audition of own voice will persist even if AC and BC cochlear audition of own voice is masked.

Estimating arrival times for own voice stimuli at the inner ear requires consideration of propagation routes (figure 2). Air-conducted (AC) sound can be direct (dAC) or reflected (rAC), whereas body-conducted (BC) vibration can be considered as direct only. Table 1 estimates arrival time at the inner ear at approximately 0.5 ms after vocalisation for both dAC sound and BC vibration. At 60 dBA stimulus levels (typical of vocalisation) BC vibration deflects both cochlear and vestibular mechanoreceptors (McNerney & Burkard, 2011; Welgampola, Rosengren, Halmagyi & Colebatch, 2003). Table 1 compares the propagation timings. Binaural coincidence detection across cochlear and vestibular mechanoreceptors, based on dAC sound and BC vibration, would identify own voice.

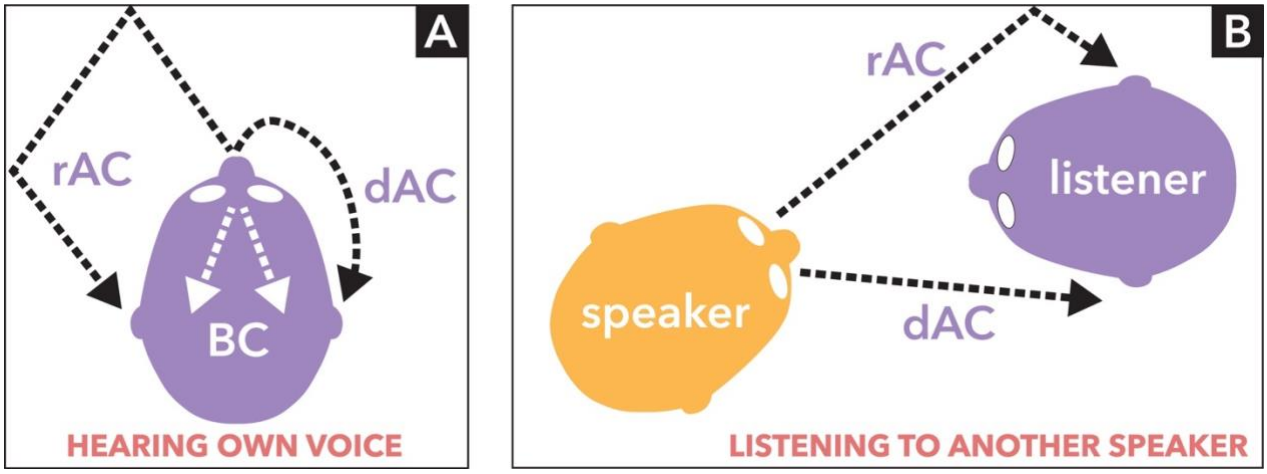

Figure 2: Sound and vibration routes to the ear. Propagation routes are difficult to measure in vivo, and estimates here are derived from primary sources where possible. For general background on sound source perception, see Yost et al. (2008); and for propagation of sound and vibration see Fahy & Thompson (2015).

Air-conducted sound is split between reflected (rAC) and direct (dAC) routes (Cabrera et al., 2009; Traer & McDermott, 2016). These are shown in a simplified version. The rAC consists of many environmental reflections with comb filtering (frequencies attenuated or reinforced due to phase differences) as sound energy reaches the ear (Yadav et al., 2012; Arend et al., 2017). The many possible routes for rAC reflect the relationship between body and environment. If reflections of reflections are present (e.g. standing waves inside a room) rAC becomes reverberation. The dAC route is transmitted directly through the air around the speaker's head. This route includes body reflection, such as that from the shoulders. There is just one form of dAC, which will tend to be stable over the short-term (unless it is windy) and medium-term (unless the head rotates relative to the torso). Conditions in which dAC is unstable tend to also be ones in which conversation is difficult.

(A) Transmission time estimates are based on human head dimensions, and will vary according to skull size and individual physiology. When hearing own voice, dAC sound is transmitted at 340 m/s and so will reach the ear in about 0.5 ms. Body conduction (BC) is through bone or soft tissue (Sohmer, 2017; Chordekar et al., 2018). Propagation routes are complex and frequency dependent, will differ between individuals, and have a nature not fully determined in vivo. However, the complexity of propagation routes will be stable in adults, changing only gradually with head composition and body profile across the lifespan. A propagation rate of 300 m/s is likely in humans (Hotehama & Nakagawa, 2012). If so, BC transmission time can be estimated as similar to the 0.5 ms for dAC. A distance of 1.5 cm between cochlear and vestibular hair cells (Ekdale, 2013) gives propagation time for vibration across the inner ear as 0.05 ms. This becomes an upper limit for arrival time difference from a laryngeal source, meaning BC arrival time is coincident to less than 0.05 ms for vestibular and cochlear mechanoreceptors. Routes to the ear for rAC will typically take 2–20 ms (depending on environmental parameters), and will be considerably less stable than for dAC or BC given that the environment, and the position of the head relative to surroundings, can be expected to change continuously.

(B) When listening to another speaker, dAC sound energy travelling a direct route between interlocutors is heard first. Energy travelling the longer, indirect route of rAC trails dAC slightly (e.g. by 5–10 ms, depending on environment). Thus, changes in firing rates of inner ear hair cells due to a typical 200 ms CV speech syllable travelling dAC and rAC routes will be spread over a further 2–50 ms or more, depending on proximity of interlocuters and environmental reflections. This overlaps with the time window for the Haas, or precedence, effect – a psychoacoustic phenomenon in which sounds separated by less than about 50 ms are perceptually integrated, with longer delays perceived as echo (Haas, 1951; Wallach et al., 1949). Overwhelmingly, dAC and rAC will have different presentations at each ear, along with comb filtering interactions, such that source localisation is via stereo combination following the duplex theory of Rayleigh (1907). There is in principle a confound for sound sources occupying the "cone of confusion" (a set of points equidistant from each ear) in symmetrical environments or those, like an anechoic chamber, with minimal rAC. In practice such a situation is so unlikely to be sustained that it would not normally have developmental impact (but see Cody et al., 1996). For animals with a pinna, filtering effects of the pinna reduce localisation inaccuracy for sources within the cone of confusion (Musican & Butler, 1984).

HEARING OWN VOICE: Routes to each ear

Route	Sensory organ	Arrival time after vocalisation	Interaural arrival time and intensity	Comment
BC	Vestibular system	~ 0.5 ms	Identical (assumes body symmetry)	Insensitive to environmental variation. BC attenuation and filtering are consistent in the short- and medium-term, with only small and gradual long-term changes which follow head composition and body profile across the lifespan.
BC	Cochlea	~ 0.5 ms		

HEARING OWN VOICE: Routes to each ear				
Route	Sensory organ	Arrival time after vocalisation	Interaural arrival time and intensity	Comment
dAC	Cochlea	~ 0.5 ms	Near identical (can vary with head orientation and air turbulence)	During vocalisation, dAC and BC contributions are approximately equal at the cochlea (von Békésy, 1949; Pörschmann, 2000; Reinfeldt et al., 2010). Some spectral variation (e.g. nasal phonemes more prominent over BC) due to filtering differences between air and body. Recordings of own speech (capturing predominantly dAC) are often found by the speaker to differ from what is heard while speaking (a mixture of BC and dAC, with some rAC).
rAC	Cochlea	typically 2–50 ms	Different	Arrives within 2–50 ms (or longer, depending on environment) of dAC sound and BC vibration. Less sonic/vibrational energy than the dAC/BC mixture. Delay relative to dAC/BC creates comb filtering. Delays of rAC above ~50 ms are experienced psychoacoustically as an echo; delays of rAC below ~50 ms are psychoacoustically fused with dAC/BC as in the Haas or precedence effect.

Table 1: Sound and vibrational energy is transmitted to each ear through body conduction (BC) and direct and reflected air conduction (dAC/rAC), and can deflect two sets of mechanoreceptors in each inner ear. At stimulus levels typical of own voice, vestibular mechanoreceptors are only deflected by BC vibration.

Groups of neurons having response properties supporting coincidence detection on the millisecond timescales required for the hypothesised own voice identification mechanism can be found in the cochlear nucleus and superior olivary complex. Review of brainstem neurons can be found in Golding & Oertel (2012) and review of vestibular inputs to the cochlear nucleus in Newlands et al. (2003) or Smith (2012). Figure 3 shows a sagittal view of brain areas innervated by the inner ear, and figure 4 shows cortical areas with connectivity to the vestibular system alongside areas important for speech and language. The cochlear nucleus and superior olivary complex comprise initial stages in a subcortical chain referred to as the ascending auditory pathway (Irvine, 1992). Changes in firing rates within brainstem neurons which correspond to the hypothesised coincidence detection could in turn be expected to change activity at higher stages of the ascending auditory pathway, including inputs to the cortex. Such activity could be interpreted as an auditory stream which identifies own voice.

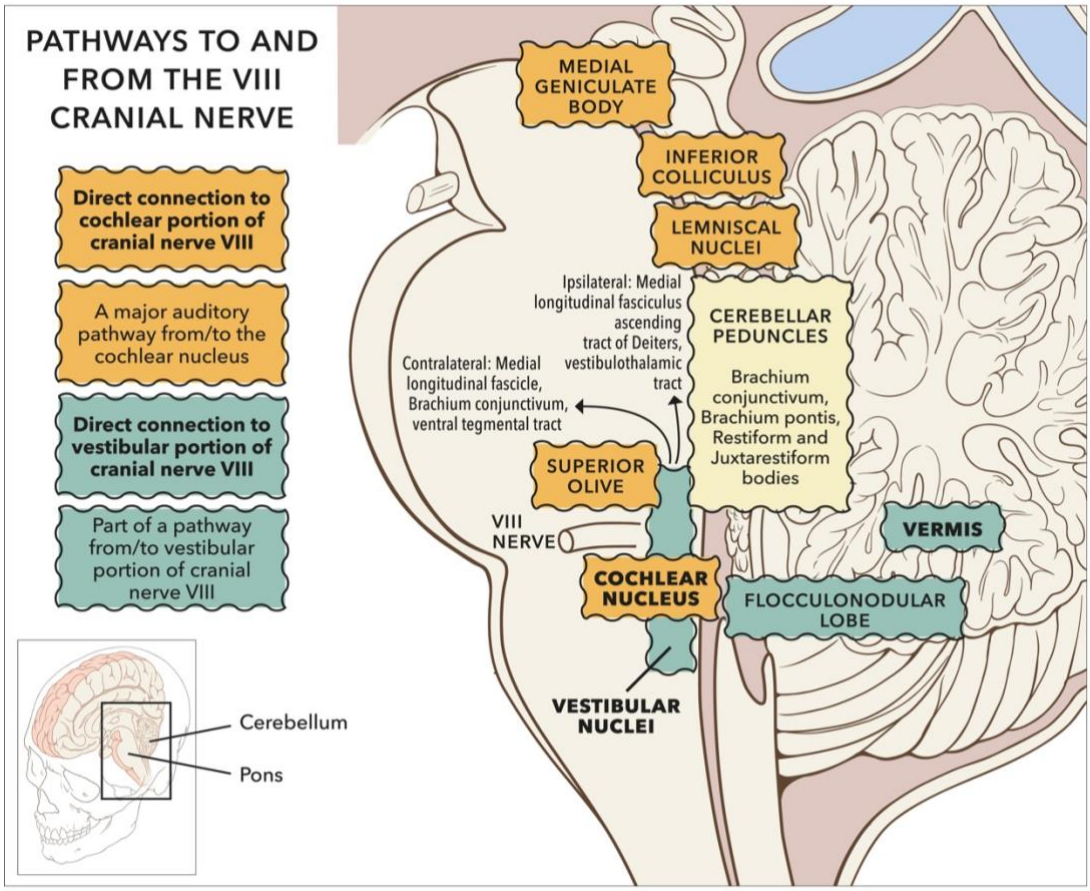

Figure 3: Sagittal view of subcortical pathways to and from the VIII cranial nerve. Whilst the auditory pathway ascending from the cochlear nucleus is relatively well established (Irvine, 1992), pathways to and from vestibular nuclei remain under investigation (Pierrot-Deseilligny & Tilikete, 2008; Zwergal et al., 2009). Investigation is largely using animal models. Projections to vestibular cortex via the thalamus have been established in humans through clinical observation and lesion studies (Conrad et al., 2014; Hitier et al., 2014; Wijesinghe et al., 2015). Vestibular nuclei also project down the spine (not shown).

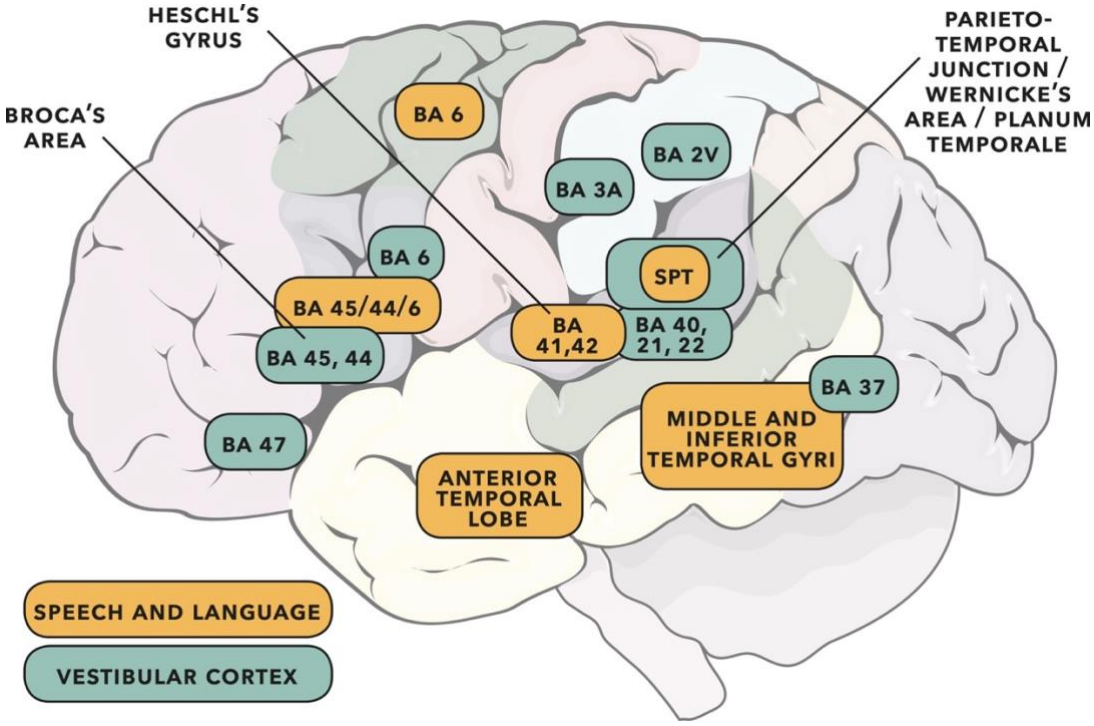

Figure 4: Cortical areas important for speech and language (adapted from the dual-stream model of Hickok & Poeppel, 2007) shown with vestibular cortical areas identified in cats, monkeys and humans (adapted from Ventre-Dominey, 2014; see also Frank & Greenlee, 2018). Cortical activity following vestibular input has wide interpretation (e.g. see reviews of cognition in Hitier et al., 2014, and auditory/rhythm/timing in Todd & Lee, 2015). Some of the vestibular areas identified will be predominantly related to gravitoinertial function (see discussion in Ferrè & Haggard, 2020). Numbers are Brodmann areas – see primary literature for more exact location detail. Spt is the Sylvian parieto-temporal region proposed by Hickok & Poeppel (2007) as a sensorimotor integration area. Vestibular sites in humans have been identified as such when direct electrical stimulation of the cortex gives rise to gravitoinertial illusion. When vestibular sites are identified within BA 21 (lateral temporal lobe) or BA 22 (Wernicke's area), auditory illusion is found to accompany gravitoinertial illusion (Kahane et al., 2003; Fenoy et al., 2006).

2.3.2.1 Explanatory Power

As a process explanation, the Concurrency Hypothesis provides a detailed account of how own voice is identified. The proposed involvement of particular types of brainstem neurons (e.g. octopus cells in the cochlear nucleus, or bipolar principal cells of the medial superior olive) generates testable auxiliary hypotheses (see discussion in sections 2.4 and 3.4). Whereas the existence of an own voice auditory stream, which is identified through coincidence detection between vestibular and cochlear afferents, is the core hypothesis.

There is also a contrastive explanation of why own voice is identified in the way described by the Concurrency Hypothesis. The contrastive explanation addresses evolutionary and philosophical considerations. The Concurrency Hypothesis as described so far is specific to mammals. However, the Concurrency Hypothesis could be extended to all terrestrial and amphibious vertebrates if the basilar papilla is considered in place of the cochlea; to fish if the lagena is considered; and in principle to any

281 animal which produces sound and vibration, and has two or more sets of sensory receptors capable of
282 detecting sound and vibration. See species surveys in Suthers, Tecumseh Fitch, Fay & Popper (2016) and
283 Pollack, Mason, Popper & Fay (2019).

284 The prospect of such a wide taxonomic application for the Concurrency Hypothesis suggests a
285 provenance early in evolution. This in turn prompts reconsideration of the role of the inner ear. The
286 Concurrency Hypothesis provides a principled distinction between self (identification of own voice) and
287 environment (reflection of own voice from surroundings). Such a distinction has importance for cognitive
288 science and philosophy of mind (Wilson & Foglia, 2017). For example, in a representational theory of
289 mind the distinction between self and environment is integral to content determination (Pitt, 2020).

290 The basis for the self-environment distinction in the Concurrency Hypothesis is the presence of two sets
291 of mechanoreceptors in the ear. One set of mechanoreceptors detects own voice in isolation, the other
292 detects own voice mixed with ambient sound, including reflection of own voice. This is dissimilar to other
293 modalities. For example, the visual analogy would be identification of one's own hand. However,
294 photoreceptors do not collect sufficient information to identify one's own hand from light waves incident
295 on the retina. Such identification would be possible following multisensory integration, but this is also
296 the case for audition (e.g. as in the combination of audition with proprioception during vocalisation).

297 As such, audition might be the only modality within which self and environment can be distinguished. If
298 so, multisensory integrations including audition could underlie self-environment distinction for
299 modalities other than audition. Evolution of any such dependency would have to create phenotypes
300 sufficiently robust to account for self-environment distinction when hearing ability is absent. Further
301 consideration of such matters is beyond the scope of this article, but would follow discussions of
302 heritability and innateness such as those in Griffiths (2020), Godfrey-Smith & Sterelny (2016) or Downes
303 & Matthews (2020).

304 Self-environment distinction is also important in our understanding of consciousness (Van Gulick, 2018).
305 For example, our experience of qualia depends on introspection from what we presume to be a shared
306 environment. Our intentionality towards objects other than ourselves rests likewise. From considerations
307 such as these, provision of a principled basis for distinction between self and environment would be a
308 comparably important function of the inner ear as its hearing function.

309 ## 2.4 Discussion

310 This section describes a general application of the Concurrency Hypothesis to speech-motor research and
311 auditory scene analysis. Section 3 will build on the discussion in this section to describe a specific
312 application of the Concurrency Hypothesis to explanation of stuttering.

313 ### 2.4.1 Application to speech-motor research

314 An own voice auditory stream would provide a target for the proposed efference copy of the speech plan
315 in predictive feedback control models (e.g. Hickok & Poeppel, 2007; Roelofs & Meyer, 1999; Tourville &
316 Guenther, 2011). If applied to speech-motor models, the Concurrency Hypothesis has potential to
317 improve explanatory power.

318 A corollary of this proposal is that if the Concurrency Hypothesis is to be tested, speech-motor research
319 should use physiologically valid own voice stimuli. Physiologically valid own voice stimuli are those
320 containing concurrent AC sound and BC vibration, with relative composition and timing as described in
321 figure 1 and table 1.

Creation of such stimuli carries practical difficulty. For example, an ideal test of speech-motor activity would compare brain activity during identical sound and vibrational stimuli in two conditions. The first condition is the standard articulatory process: brain activity generates sound and vibration following coordinated nerve impulses to articulatory muscles, whilst at the same time brain activity is altered following deflection of inner ear mechanoreceptors by the sound and vibration produced during articulation. The second condition should be identical to the first, but without the activity in articulatory muscles being created by brain activity. Instead, the measured brain activity would be solely in response to the sound and vibration produced by articulatory muscles. Unfortunately, the experimental arrangement in the second condition is difficult or impossible even in animal models. The articulatory muscles could in principle be made to produce sound and vibrational stimuli similar to that during vocalisation, for example through electrical stimulus to the articulatory muscles. However, the process of doing so would either be highly traumatic to the host animal, or the animal would have to be sedated. Whatever experimental arrangement is chosen, resting state brain activity in the second condition would differ from that of the first condition (the standard articulatory process) to the extent that comparison of brain activity between the two conditions would be overwhelmingly difficult to interpret.

Accordingly, much testing of brain activity during articulation, or vocalisation, has been based around a simpler comparison. The first condition is the standard articulatory process (i.e. as previously defined), with simultaneous recording of brain activity (e.g. by electrophysiology) and the sound and/or vibrations created during articulation (e.g. using a microphone). The second condition comprises a recording of brain activity without articulation, whilst the sound and/or vibration recorded in the first condition is played back. This comparison would seem to overcome the difficulty with having articulatory muscles create the sound and vibration in the second condition. However, there is a disanalogy in that the sound and vibration in the second condition are not identical to the sound and vibration in the first condition. This disanalogy has potential to invalidate the intended comparison.

Thus, protocols intended to compare brain activity during articulation and the playback of a recording of vocalisation must choose a methodology for recording and playback of the sound and/or vibration. Possibilities are shown in a Latin square in figure 5. Of these, speech-motor investigation has overwhelmingly compared the own voice condition with dAC playback of a dAC recording. Often, participants are invited to adjust sound pressure levels of dAC playback so as to perceptually match the loudness of the AC/BC combination heard during vocalisation. Doing so does not create a stimulus comparable to the stimulus present during vocalisation. Own voice is perceived through an approximately equal combination of air- and body-conducted stimuli (von Békésy, 1949; Pörschmann, 2000; Reinfeldt et al., 2010). Perceptual doubling of the loudness of the AC stimulus, to compensate for the absence of BC stimulus, will for most participants correspond to no more than a 10 dB increase in sound pressure level (Stevens, 1972; Warren, 1973; Florentine, Popper & Fay, 2010). Such an increase will barely bring the AC stimulus to vestibular threshold, which for AC is 10 dB above the 60 dBA level typical of conversational speech. The AC vestibular threshold is moreover 35 dB above the BC vestibular threshold (McNerney & Burkard, 2011; Welgampola, Rosengren, Halmagyi & Colebatch, 2003).

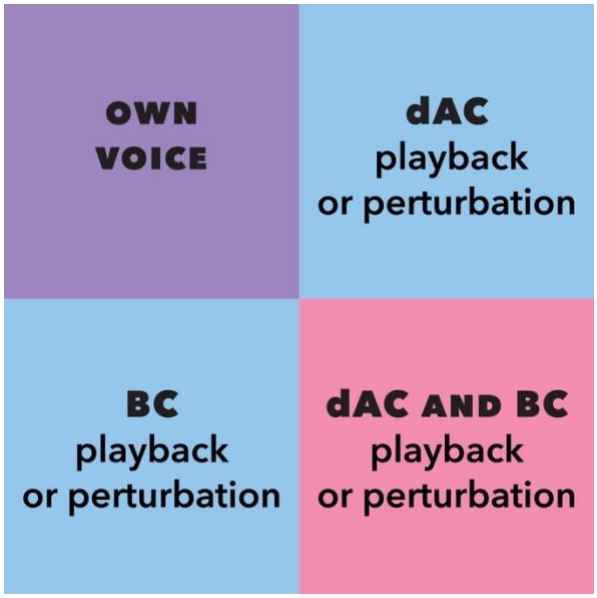

Figure 5: Latin square showing sound and vibrational stimuli which could be used in brain studies of own voice audition. The "own voice" condition is the standard articulatory process. It includes speech-motor brain activity which results in articulation generating dAC and rAC sound, and BC vibration; and at the same time includes the brain activity following deflection of inner ear mechanoreceptors by the dAC and rAC sound, and BC vibration, produced during articulation. "Playback" refers to playback of recordings of sound or vibration made during the standard articulatory process. Playback conditions do not contain speech-motor activity, unless digitally processed playback with a short delay (usually 10 ms or more) is presented concurrently with ongoing articulation. Such short latency digital manipulation is referred to as perturbation, and may also include manipulations to recordings (e.g. frequency shifts or changes to the nature of formants). A limitation for any type of playback is that the sound and vibrational stimuli present in the own voice condition cannot be recreated exactly using the earphones and bone vibrators available in laboratories. Combined air- and body-conducted (AC/BC) playback according to the timings provided in table 1 (i.e. AC and BC playback with binaural arrival at the inner ear coincident to ~ 0.1 ms) offers the closest approximation to the sound and vibrational stimuli present in the own voice condition. Not shown in the diagram is that BC stimulus can be subdivided into levels above and below vestibular threshold. BC stimulus should be above vestibular threshold, and AC stimulus below vestibular threshold, to mimic stimuli present during articulation.

It follows that even after a sound pressure level increase to perceptually match the loudness of own voice, stimulation due to AC playback will either deflect vestibular mechanoreceptors very weakly in comparison to the BC stimulation present during vocalisation, or stimulation due to AC playback will not deflect vestibular mechanoreceptors at all. Firing rates of the vestibular ganglion will be altered barely or not at all from resting state. Action potentials along the VIII cranial nerve will predominantly be altered according to deflection of cochlear mechanoreceptors by AC playback, and an auditory stream corresponding to own voice will not be identified through coincidence detection between cochlear and vestibular streams as per the Concurrency Hypothesis.

Many functional imaging studies have compared vocalisation to AC playback of own voice recordings (e.g. with human participants: Numminen et al., 1998; Numminen & Curio, 1999; Curio et al., 2000; Ford et al., 2001; Houde et al., 2002; Ford & Mathalon, 2004; Ventura et al., 2009; Greenlee et al., 2011; Sato & Shiller, 2018; or using animal models: Müller-Preuss & Ploog, 1981; Eliades & Wang, 2017; Eliades & Tsunada, 2018). A consistent finding in such experiments is that parts of temporal cortex which respond

to sound have reduced activity in the vocalisation condition compared to the playback condition. This has been interpreted as speech-motor activity modulating the temporal cortex (Hickok et al., 2011; Parrell & Houde, 2019). The interpretation is consistent with theoretical models in which attenuating auditory feedback increases accuracy of state estimates of the speech-motor system (Parrell et al., 2019).

Whilst an attractive explanation, motor induced suppression of temporal cortex is not strongly supported by studies comparing vocalisation and AC playback conditions. The reason for this is that vocalisation and playback stimuli differ (as per figure 5), meaning that the observed reduction in temporal cortex activity cannot conclusively be attributed to speech-motor activity modulating temporal cortex. An alternative explanation is that the observed reduction in temporal cortex activity is due to the difference in stimuli between vocalisation and AC playback conditions. The Concurrency Hypothesis is consistent with this alternative explanation. The Concurrency Hypothesis adds the detail that in the vocalisation condition, firing rates of neurons in the ascending auditory pathway will uniquely identify own voice through coincidence detection of cochlear and vestibular afferents. Whereas in the AC playback condition, the ascending auditory pathway functions as it would with any ambient AC stimulus (i.e. as per Irvine 1992; Bregman, 1990).

It is possible that both explanations are correct: that an own voice auditory stream modifies temporal cortex activity, and that articulation modifies temporal cortex activity independently of audition. Exploring these possibilities offers the opportunity to increase explanatory power of speech-motor models, and to make testable predictions. In doing so it is not necessary to use the Concurrency Hypothesis. However, alternatives would be to propose a different method by which own voice is identified as an ascending auditory stream (i.e. a solution to the ill-posed problem of sound source discrimination in auditory scene analysis), or else to stipulate that an auditory target map for own speech is innately specified (e.g. as per Liberman & Mattingly, 1985).

Studies using playback of own voice recordings could be reinterpreted in light of these considerations, and extended to include BC stimuli. Auditory perturbation studies could be similarly reinterpreted (e.g. McGuire et al., 1996; Hirano et al., 1997; Fu et al., 2006; Parkinson et al., 2012; Toyomura et al., 2007; Zarate & Zatorre, 2008; Tourville et al., 2008; Zheng et al., 2009; Zarate et al., 2010). In auditory perturbation studies, vocalisation is recorded, is optionally digitally manipulated, and is played back with a short delay whilst articulation is ongoing. Examples of manipulation include frequency shift or alteration of formants. Recording and playback use AC sound. Digital processing (e.g. with fast Fourier transform) introduces delays which are typically 10 ms or more. Such delays are at least an order of magnitude larger than the sub-millisecond timings in table 1. Thus, auditory perturbation studies assess the effect of keeping the BC vibrational stimulus of vocalisation unchanged, whilst adding a delayed AC stimulus having similar spectral characteristics to the ongoing vocalisation. Effectively they manipulate rAC and (if using insert earphones) attenuate dAC. The protocol could be extended to form part of a larger range of investigation in which BC, and combined AC/BC, manipulations are also evaluated.

The Latin square in figure 5 is a simplification. Stimuli can be further subdivided into those above and below vestibular threshold. Todd et al. (2014a, 2014b) compared cortical response to stimuli above and below vestibular threshold. Electroencephalography showed morphological change in and around the N1 wave upon crossing vestibular threshold, with source analysis indicating origin in cingulate or temporal cortex. The N1 wave (or its M100 equivalent in magnetoencephalography) is the component found to have reduced amplitude when brain activity during vocalisation is compared to brain activity during AC playback of vocalisation. Thus, the suggestion is that in studies comparing vocalisation and playback conditions, the observed brain activity will differ depending on whether playback stimuli are above or

below vestibular threshold. A physiologically valid own voice stimulus will combine BC stimulus above vestibular threshold with AC stimulus below vestibular threshold. Follow-up work to the current article will appraise brain activity following combinations of BC and AC stimuli which are respectively above and below vestibular threshold.

2.4.2 Application to Auditory Scene Analysis

Bregman (1990) proposed that auditory scenes are generated from the neural firing patterns elicited when sound waves are coincident on the biomechanical structure of the middle and inner ears. Auditory scenes would contain detail consistent with our perceptual experience. Two processes are proposed to identify the auditory streams which comprise auditory scenes. Firstly, primitives, which are general purpose segregation and grouping processes based on those developed by the Gestalt school (e.g. common onset, harmonicity, spectral composition, co-variation in amplitude; Carlyon, 2004; Darwin, 2007; Ciocca, 2008; Denham & Winkler, 2015; Młynarski & McDermott, 2019). Secondly, schemas, which are specific processes identifying certain types of sound (e.g. conspecific animal vocalisations or phonemes in human speech; Bey & McAdams, 2002; Billig et al., 2013; Woods & McDermott, 2018).

The Concurrency Hypothesis could be the basis of a schema identifying own voice. Modelling of auditory scene analysis is an active research area (Cooke & Ellis, 2001; Haykin & Chen, 2005; Snyder & Alain, 2007; Winkler et al., 2009; Szabó, Denham & Winkler. 2016; Snyder & Elhilali, 2017; Chakrabarty & Elhilali, 2019). Whichever modelling approach is taken, the Concurrency Hypothesis would be applied through the following principles:

i. Primitive processes are proposed to act on neural firing patterns elicited by deflection of vestibular mechanoreceptors as well as by deflection of cochlear mechanoreceptors.
ii. Whenever firing patterns of vestibular and cochlear origin have similar attributes as identified by primitives, the firing patterns are likely to correspond to own voice.
iii. Activity in the auditory brainstem (BinKhamis et al., 2019) is consistent with substantial processing of speech sounds. As such, models will have greater neurological plausibility if the coincidence detection in (ii) occurs very early in the ascending auditory pathway – for example, in the cochlear nucleus or the superior olivary complex.
iv. Computational modelling of coincidence detection (e.g. through vestibular input to octopus cells in the cochlear nucleus) may require primitives and schemas to be entwined.

An own voice identification schema based on (i – iv) could underpin further schemas. Possibilities are:

v. Vocalisation of conspecifics is likely to be occurring when primitives identify similar neural firing patterns (e.g. spectral composition typical of formants) to those present during own voice coincidence detection, but when vocalisation is not being produced and neural firing patterns arise from cochlear mechanoreceptors only.
vi. If stored in short-term memory, an own voice auditory stream could be compared via primitives to the rAC reflections of own voice (see figure 2 and table 1) to create a schema identifying reflection and reverberation.
vii. Multisensory integration (Stein & Stanford, 2008) of reflections and reverberations from (vi) with head and body position could support a schema for echolocation (see review of human echolocation in Kolarik et al., 2014).
viii. Sound source learning based on (vii), in combination with the generalised vocalisation schema of (v), could support a schema distinguishing sources in multi-speaker scenarios.

ix. Adaptation of the schema in (viii) for sounds other than vocalisation could reinforce learning of sound source location using primitives.

These ideas need development into computational models. The underlying point is that many or all of the schemas required by auditory scene analysis could be based on the Concurrency Hypothesis. The high energy vocalisations of neonates (e.g. crying or wailing) have more than sufficient energy to deflect both cochlear and vestibular mechanoreceptors, meaning that auditory learning based on the Concurrency Hypothesis would begin at birth (and quite possibly, would have a precursor based on the mother's voice in utero).

3. Hypotheses of Stuttering

3.1 Explanatory targets

Explanatory targets for stuttering are extensive. Table 2 shows process explananda (how stuttering happens), whilst table 3 shows contrastive explananda (why stuttering happens). These lists are not intended as exhaustive, but are rather presented as minimal criteria which any hypothesis of stuttering should address.

Priority will be given to addressing process explananda. This is not to downplay the importance of contrastive explananda for stuttering research. However, a comprehensive discussion of contrastive explananda for stuttering (e.g. why there is a sex difference; the role of heredity; whether a particular brain study reflects causation, consequences or correlates of stuttering) encompasses issues wider than those within stuttering research, and is accordingly outside the scope of this article. The current aim is of adequacy for process explanans, with contrastive explanans added as part of ongoing research.

Explananda	Examples	References
Core stuttering behaviours	Prolongation or repetition of speech sounds, including silent blocks to airflow.	Van Riper (1982 ,ch 6); Bloodstein (1995, ch 1–2); Ward (2006, ch 1, 7, 9)
Accessory and interiorised stuttering behaviours	Accessory stuttering includes excess tension or tremor in articulatory muscles; perseveration; changes in breathing; use of fixed posture; postponement of words or substitution of synonyms; and movement of non-articulatory muscles, including limb movements, especially in attempts to time or disguise movement of articulatory muscles. Interiorised stuttering adds word, phoneme or situation fears; situation avoidance; frustration, hostility and guilt. See Iverach et al. (2017) for discussion of overlap between stuttering and social anxiety.	Van Riper (1982 ch 6, 7, 11); Bloodstein (1995, ch 1–2); Ward (2006, ch 1, 7, 9) Iverach et al. (2017)
Linguistic and/or situational	Stuttering increases with propositionality of content. Stuttering is reduced when speaking alone or to animals. Stuttering increases when talking to authority figures, or when the audience appears distracted. Stuttering is mostly word initial, and almost never on the last sound of a word or syntactic structure. Stuttering tends to occur at the beginning of a sentence or grammatical clause. Accented syllables are more likely to be stuttered. Adults tend	Richels et al. (2010); Buhr & Zebrowski (2009); Bloodstein (2002, 2006); Ward (2006, ch 5); Karniol (1995); Bloodstein (1995, ch 7); Van Riper (1982 ch 8); Langová & Sváb (1973); Sheehan et al. (1967);

	to stutter on longer words, on words starting with consonants, and on words with low transition probability (high information load). Adults tend to stutter on content rather than function words, and vice versa for children.	Gould & Sheehan (1967); Eisenson & Horowitz (1945)
Anticipation, Consistency and Adjacency	People who stutter can predict when stuttering will occur with high accuracy. In successive oral readings of the same material, stuttering tends to occur on the same syllables. If previously stuttered words are blotted out, stuttering on subsequent readings tends to be on words adjacent to those previously stuttered.	Garcia-Barrera & Davidow (2015); Brocklehurst et al. (2013); Jackson et al. (2015); Bloodstein (1995, ch 7);
Adaptation	Stuttering is reduced in successive oral readings of the same material (occurs simultaneously with the consistency effect).	Brocklehurst et al. (2013); Max & Baldwin (2010); Bloodstein (1995, ch 8); Wingate (1986 a,b)
Operant conditioning	Stuttering is reduced in response-contingent stimulation experiments (e.g. when using electric shock or time out during stuttered moments).	Ingham (1984, ch 9); Nittrouer & Cheney (1984); Bloodstein (1995, ch 8)
Alteration to audition during speech	Many changes to audition during speech can reduce stuttering. Effective changes include delay; frequency shift; masking; and unison speaking with, or shadowing of, a second speaker.	Bloodstein (1995, ch 2,8); Van Riper (1982 ch 15); Ingham (1984, ch 10); Ward (2006, ch 3); Howell et al. (1987); Yates (1963)
Alteration to stress patterns within vocalisation	Speaking in time with a metronome reduces stuttering, as does singing.	Van Riper (1982 ch 15, 17); Bloodstein (1995, ch 2,8); Wingate (1969)
Therapy effectiveness	Interventions for stuttering have an overall positive effect, although some stuttering usually remains post-intervention. Methodical comparison of interventions is difficult. No intervention is clearly preferred.	Johnson et al. (2015); Baxter et al. (2015); Herder et al. (2006); Ward (2006 ch 15)

Table 2: Process explananda for stuttering

Explananda	Examples	References
Age of onset	Childhood stuttering has a median age of onset between 3–4 years, with a positively skewed distribution and upper limit around 9–12 years. Childhood cases are mostly developmental, but some may be acquired (neurogenic or psychogenic). See Ward (2006, ch 7), Yairi (2007) or Seery et al. (2007) for review of subtypes in children and Van Borsel (2014) for review of acquired stuttering. Adult onset is rare. Adult onset may be acquired, and/or re-emergent childhood stuttering (Van Riper, 1982, p64).	Yairi & Ambrose (2013); Bloodstein (1995 ch 3,6); Craig et al. (2002); Månsson (2000); Andrews & Harris (1964)
Heredity	Averaging across seven twin studies gives a heritability estimate for PDS at 70% (95% CI 59–81%, studies compiled in Frigerio-Domingues & Drayna, 2017). Prospective genetic variations for stuttering have been identified through linkage analysis. Knock-	Benito-Aragón et al. (2020) Frigerio-Domingues & Drayna (2017); Kraft & Yairi (2012); Ward (2006, ch 7);

	in mice carrying one of these variations show stuttering in their ultrasonic vocalisations (Han et al. 2019).	Bloodstein (1995, ch 3)
Incidence and prevalence	Incidence (fraction of the population who have ever stuttered) is between 5–8%. Lifespan prevalence (fraction of the entire population who stutter) is 0.72%, but prevalence can be much higher (e.g. 2–4%) with cohorts aged younger than 12 years, and especially those between 2–6 years.	Yairi & Ambrose (2013); Bloodstein (1995 ch 3,6); Craig et al. (2002); Månsson (2000); Andrews & Harris (1964)
Sex differences	At the typical onset age of 3-4 years old, about 1.5 times more boys than girls stutter. By adulthood, 3 or 4 times more men than women will stutter. Thus, boys are more likely than girls to start stuttering. And girls are more likely than boys to stop stuttering. Drayna et al. (1999) find PWS are more likely to be male in cases with no family history of stuttering (see also Ambrose et al., 1997).	Yairi & Ambrose (2013); Bloodstein (1995 ch 3,6); Craig et al. (2002); Månsson (2000); Andrews & Harris (1964)
Childhood stuttering is frequently transient	Incidence and prevalence data show many children who stutter (60–80%) will stop stuttering, with or without intervention. Systematic review of 35 studies (Sugathan et al., 2020) indicates speech features as predictive of stuttering continuing into adulthood (more stuttering-like dysfluencies including dysrhythmic phonation and monosyllabic word repetition; higher articulatory rate; lower score in phonology tests). Meta-analysis of 11 studies (Singer et al., 2020) adds predictors of: male sex; greater age at onset; heredity; and lower scores in tests of language skills.	Yairi & Ambrose (2013); Sugathan et al. (2020); Singer et al. (2020)
Co-occurring diagnoses	Blood et al. (2003), in a survey of 1184 speech and language pathologists (SLPs), found 37% of 2628 children who stuttered had no co-occurring diagnosable condition. The remaining 63% had an average of 2.2 co-occurring diagnosable conditions. Of these, 33.5% were articulatory, 25.6% concerned expressive or receptive semantics, and 34.5% were non-speech-language (including learning, literacy, attention deficit and central auditory processing). An earlier survey of SLPs (Arndt & Healey, 2001) found 44% of 467 children who stuttered had a co-occurring phonological and/or language diagnosis. There is overlap between stuttering and dyslexia (Elsherif et al., 2021). Generalised household surveys also show a high co-occurrence for stuttering with other developmental categories in self or parent report. Retrospective self-report from adults with latent class analysis (Ajdacic-Gross et al., 2018; see also 2010), established association with atopic disease (e.g. allergy or asthma); psychosocial adversity in childhood; or neurodevelopmental or early anxiety disorder. However, such co-occurences were only present in one of two subgroups.	Elsherif et al. (2021); Briley & Ellis (2018); Ajdacic-Gross et al. (2018); Ajdacic-Gross et al. (2010); Boulet et al. (2009); Blood et al. (2003); Arndt & Healey (2001); Bloodstein (1995, ch 4–6)
Subtle differences from controls	Subtle differences can be found between adults or children who do and do not stutter for tasks involving general motor control. Sometimes differences in the integration of motor control with timing systems, and/or sensory or proprioceptive input, are implied. Examples include movement initiation latency and movement duration, and	Choo et al. (2020); Ofoe et al. (2018); Ntourou et al. (2011); Max (2004); Bloodstein (1995); Rosenfield & Jerger (1984)

	involve use of effectors such as fingers and hands as well as use of the orofacial system (Max, 2004). Subtle differences are also found in many tests of central auditory function (review in Rosenfield & Jerger, 1984). Meta-analysis of language tests in children who stutter (receptive and expressive vocabulary, mean length of utterance, syntactic complexity, homogeneity analysis) shows subtle rather than clinically significant differences from controls (Ntourou et al., 2010). Similarly subtle differences between children who do and do not stutter are reported in tests related to attention and executive function (Ofoe et al., 2018; Choo et al., 2020).	
Neurotransmitters	Positron Electron Tomography and drug trials indicate that neurotransmitters influence stuttering behaviour. Dopamine levels appear particularly important.	Maguire et al. (2020); Metzger et al. (2017); Wu et al. (1997)
Neuroimaging	Differences from controls in areas important for auditory, language and speech-motor function. See discussion in section 3.4.1.	Etchell et al. (2017); Neef et al. (2015); Belyk et al. (2015); Budde et al. (2014)

Table 3: Contrastive explananda for stuttering

3.2 Candidate explanations

Bloodstein (1995) categorises hypotheses of the moment of stuttering into three groups: repressed needs, anticipatory struggle, and breakdown. Research and theoretical development over the last 30 years has overwhelmingly focussed on breakdown hypotheses. As such, repressed needs hypotheses and anticipatory struggle hypotheses will be reviewed only in brief, whilst breakdown hypotheses will be described in greater detail.

3.2.1 Repressed Needs Hypotheses

Originating in the psychoanalytic schools of the 1920s and 1930s, repressed needs hypotheses describe stuttering as a neurotic symptom rooted in unconscious needs. Such hypotheses are outside the mainstream of contemporary stuttering research (Martin, 2016).

3.2.2 Anticipatory Struggle Hypotheses

In anticipatory struggle hypotheses, stuttering is preceded by the speaker's prediction that speech will be difficult to execute. The prediction of difficulty leads to increased muscular tension. The increased muscular tension in turn impairs the coordination usually present during speech, and causes the speech attempt to be stuttered.

Anticipatory struggle hypotheses have seen little development in the last 50 years. For a historical survey, see Bloodstein (1995, ch 2), and for a contemporary perspective see Brocklehurst et al. (2013).

3.2.3 Breakdown Hypotheses

In breakdown hypotheses, stuttering is a behavioural manifestation of vulnerability in speaking ability. The vulnerability is generally proposed to occur in either the language encoding or the speech-motor

system. Breakdown of the vulnerable system is typically attributed to emotional or psychosocial stress (Bloodstein, 1995, p60).

3.2.3.1 Language encoding breakdown

Language encoding breakdown has been described in what Levelt (1989, 1999) refers to as the Formulator. This is a hypothesised stage of speech production between thought and expression, in which lexical and syntactic selection, along with morphological, phonological and phonetic encoding, precedes creation of a motor plan. Levelt's model is shown in figure 6. The Formulator can be described using a spreading activation network (e.g. Dell, 1986; Dell & O'Seaghdha, 1991). In network models, a metrical frame is created for a planned utterance. Phonological segment nodes will then compete for selection, with the nodes filling the frame being those which have the highest activation level at the moment when speech-motor planning commences.

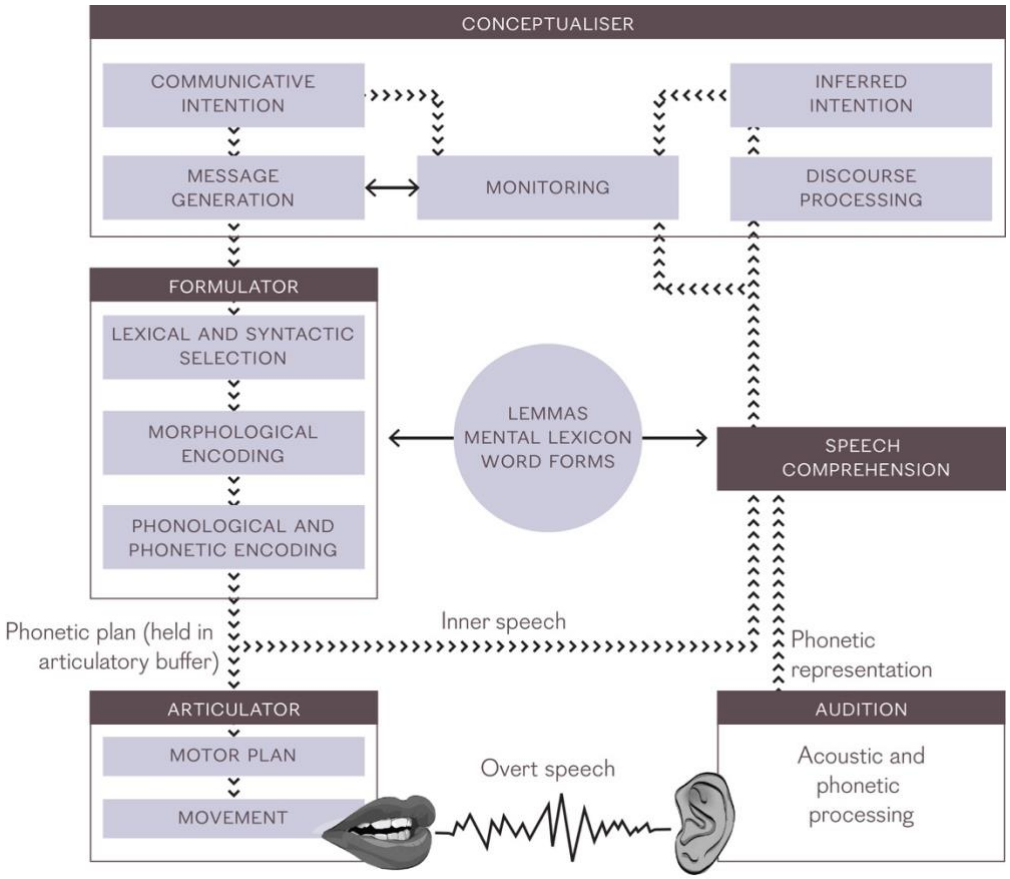

Figure 6: Speech production model of Levelt (1989; see also Levelt et al., 1999). Notable features are an inner and an outer loop, with the parsimony of a shared mental lexicon. Stages include Audition, Speech Comprehension, Conceptualisation, Formulation and Articulation. The Concurrency Hypothesis concerns activity in Audition, and thus addresses a special case of auditory scene analysis (Bregman, 1990). Discussion of Speech Comprehension can be found in Norris et al. (2000), Galantucci et al. (2006) or Poeppel et al. (2008), among others. There is no widely agreed model of the Conceptualiser; any effort to produce one touches on long-standing issues in Cognitive Science, Philosophy of Psychology and Philosophy of Mind (several other hypotheses of the Levelt model, and hypotheses of its constituents, do likewise). Indefrey & Levelt (2004) present a meta-analysis, based on neuroimaging literature, of the time course for processes within the Formulator; see also section 3.2.3.1 for discussion of Dell's (1986) spreading activation network model of the Formulator. Articulation is described by speech-motor control models such as DIVA (Guenther et al., 2006) or FACTS (Parrell et al., 2019).

The Covert Repair Hypothesis (Postma & Kolk, 1993) postulates slower than usual activity in the Formulator for PWS. As a result, a speech plan may be created whilst nodes are still competing for selection. If inappropriate nodes are selected, two possibilities pertain. If the inappropriate nodes are detected prior to articulation (e.g. via an internal monitoring loop), they are repaired covertly. This repair manifests as a silent pause – the speaker wishes to continue, but cannot do so at that moment. Alternatively, if inappropriate nodes are detected during articulation, the speaker will stop and retrace. Phonemes uttered prior to retrace are audible as stuttering for however many reformulations are necessary to correct the speech plan. A variant on this theme is offered by the Vicious Circle Hypothesis (Vasić & Wijnen, 2005; Bernstein Ratner & Wijnen, 2007), which proposes that it is over-vigilance in repair, rather than slower than usual formulation, which causes stuttering.

An alternative breakdown mechanism is described by Howell (2004, 2008). In the EXPLAN hypothesis, breakdown occurs when the rate of speech planning has fallen below that of execution. The available speech plan is repeatedly executed until a continuation of the speech plan is available. EXPLAN entails aspects of both psycholinguistic and speech-motor breakdown. The Variable Release Threshold hypothesis (Brocklehurst et al., 2013) modifies EXPLAN such that the release threshold for a phoneme will vary according to a modified version of Bloodstein's (1975) account of anticipatory struggle.

3.2.3.2 Speech-motor breakdown

Speech-motor breakdown is typically investigated through comparison of people who stutter in fluent versus stuttered speech (state comparison) or people who do and do not stutter during fluent speech (trait comparison). Outcome measurement is via neuroimaging, electromyography of articulatory muscles, or a hybrid design (e.g. studies employing transcranial magnetic stimulation). Differences are reliably and repeatedly established in both trait and state comparisons, and are present even below the threshold for behavioural observation of stuttered speech (Etchell et al., 2017; Neef et al., 2015; Belyk et al., 2015; Budde et al., 2014). Brain areas frequently identified include premotor cortex and the temporo-parietal junction (including white matter connecting those areas), the cerebellum, and the basal ganglia. Stuttering can be emulated neurocomputationally by modelling the brain activity observed in neuroimaging of stuttering (Civier et al., 2013), with over-reliance on auditory feedback a contributing factor (Max et al., 2004; Civier et al., 2010). Arenas (2017) proposes an extension to speech-motor breakdown in which fluctuations in the vigilance of the monitoring system account for the contextual variability of stuttering.

3.3 A Novel Account of Stuttering: REMATCH (Reflexivity and Communicative Mismatch)

3.3.1 Introduction

This section introduces a novel account of stuttering with two core hypotheses: Reflexivity, and Communicative Mismatch. The combination is referred to as REMATCH.

The first core hypothesis in REMATCH concerns a quale referred to as "reflexivity". It proposes that PWS have a subjective experience during speaking in which their own speech has increased salience in comparison to the way that people who do not stutter experience their own speech while speaking. The second describes communicative mismatch, in which a breakdown in communicative choreography between speaker and listener engenders observable stuttering behaviour. The reflexivity proposal develops the Concurrency Hypothesis described in section 2. It is a distal cause of stuttering relative to communicative mismatch.

This section will proceed as follows. The sequence of events leading to a moment of stuttering, consistent with the two core hypotheses, will be described. The core hypotheses will then be applied to the explananda in tables 2 and 3.

3.3.2 Increased Reflexivity

Consider that the subjective experience of seeing the colour red may differ between individuals, even if those individuals can mutually agree that the referenced colour is red (Tye, 2018). Similarly, different speakers may have differing subjective experiences of hearing their own voice during vocalisation. The proposal is that the subjective experience of hearing own voice during vocalisation differs in a principled and consistent manner between people who do and do not stutter.

This subjective experience, or quale, of own voice during vocalisation will henceforth be referred to as "reflexivity". It is related to self-awareness (Gallagher & Zahavi, 2021; Smith, 2020). The exact proposal is that reflexivity is increased for PWS relative to controls. What is meant by increased reflexivity is that the phenomenal experience of own voice is more intense for PWS than for ordinarily fluent speakers. It is as if PWS were speaking through a magical megaphone, which broadcasts only inside the body, and whose effect is to increase salience of the message being delivered rather than volume of the utterance.

Empirical investigation of qualia is achievable through psychophysics, albeit with well-identified difficulties (Fodor, 1987). The proposal of reflexivity as a quale builds on the Concurrency Hypothesis described in section 2, and in particular it follows from the issues around evolution, cognitive science and philosophy of mind discussed in section 2.3.2.1. The hypothesis of Communicative Mismatch, to be introduced in section 3.3.1.2, proposes that a difference in subjective experience of the reflexivity quale between people who do and do not stutter is causative of stuttering behaviour.

A difference in reflexivity between people who do and do not stutter could be expected to coincide with a difference in the auditory feedback whose presence is integral to many types of psycholinguistic and speech-motor models. Alterations to auditory feedback are well-established as reducing stuttering for PWS (Yates, 1963; Howell et al., 1987; Kalinowski et al., 1993; Stuart et al., 2004; Foundas et al., 2013), and hyperfunctional monitoring in stuttering has been proposed from psycholinguistic (Bernstein Ratner, 1997; Bernstein Ratner & Wijnen, 2007) and speech-motor (Arenas, 2017) perspectives. REMATCH is independent of any particular speech-motor or psycholinguistic model. For example, speech may be entirely under feed forward control, or else speech may be best described by paradigms which do away with mental representation entirely (e.g. certain types of dynamical system, or those of extended cognition). For readers who prefer to think in terms of feedback control, the idea would be that an entire person (including the history, personality, hopes, dreams, and so forth) is included in the feedback loop for own voice audition. See Mysak (1969, ch 7) for a systems control account of stuttering along these lines.

3.3.3 Communicative Mismatch

When PWS describe moments of stuttering, the role of the audience and situation are among themes identified (Tichenor & Yaruss, 2018). In a review of linguistic factors, Karniol (1995) suggests that the involvement of motor process in stuttering is a symptom rather than a cause. Pierre (2015), extending beyond linguistics to discuss societal convention more broadly, describes stuttered speech as marginalised relative to dominant choreographies of bodily and inter-bodily communicative practices.

To address perspectives such as these, stuttering will be considered not just as an interruption of speech, but moreover as an interruption of a speech act (i.e. speech act as per Austin, 1955). It will be based around the approach-avoidance conflict hypothesis of stuttering developed by Sheehan (1953, 1958, 1970, 1975). Approach-avoidance conflict was originally formulated as a Gestalt field theory by Lewin (1935). Conflict would follow incompatible goals – for example, accept a substantial pay rise (approach gradient), but only with unpaid weekend work at the employer's discretion (avoidance gradient). Sheehan (1958) proposed that stuttering is a double approach-avoidance conflict, in which "[The person who stutters] can speak, thus achieving his aim of communication, but at the cost of the shame and guilt he has learned to attach to his stuttering. Or he can remain silent, abandon communication, and suffer the frustration and guilt that such a retreat carries with it."

Sheehan was inspired by the work of Miller (1944) who trained rats in a runway first with a food goal, then with electric shock. When Miller presented the previously trained rats with a combination of a food goal and electric shock, the rats would display motor control vacillations similar to those observed in stuttering. In an earlier proposal along similar lines, Wyneken (1868; translated in Van Riper, 1982 p281) describes the will to speak during stuttering as "partially paralysed by doubt … and one which is directly opposed to the will proper". Wyneken goes on to liken stuttering to "…when somebody, for example, wants to venture a jump, but in the very moment in which he leaps doubts that he will succeed. Often he can no longer stop the leap, but also does not jump with sufficient assurance, and so does not reach his goal."

The approach-avoidance proposal is updated in several ways. Firstly, the core hypothesis of increased reflexivity for PWS corresponds to own speech having increased salience when interpreted through the auditory system. Secondly, it is proposed that the unconscious interpretation of own speech operates with the high degree of automaticity proposed for unconscious processes in dual process theory (e.g. as per Evans, 2007; Kahneman, 2011). The double approach-avoidance conflict proposed by Sheehan is thus fragmented between unconscious and conscious processes. The final proposal is that stuttering occurs at times when there is uncertainty about the message being delivered. The uncertainty might, for example, relate to message content (e.g. whether the message being conveyed is accurate) or to message appropriateness (e.g. whether the message should be delivered to a particular audience, or at a particular time). The uncertainty could also be learned (e.g. from previous experience with stuttering – this would account for the difficulty many who stutter have in saying their own name).

Putting all of these components together, the overall proposal is that whenever the speaker unconsciously interprets own speech with uncertainty, nerve signals are created which block the ongoing speech act. At the same time, the speaker notices difficulty and consciously generates nerve signals intended to continue the speech act. Articulatory muscles respond to both conscious and unconscious processes, and so simultaneously receive innervation which is consistent with completion and cessation of an utterance. The resultant activity is behaviourally observable as stuttering. Figure 7 summarises the activity diagrammatically.

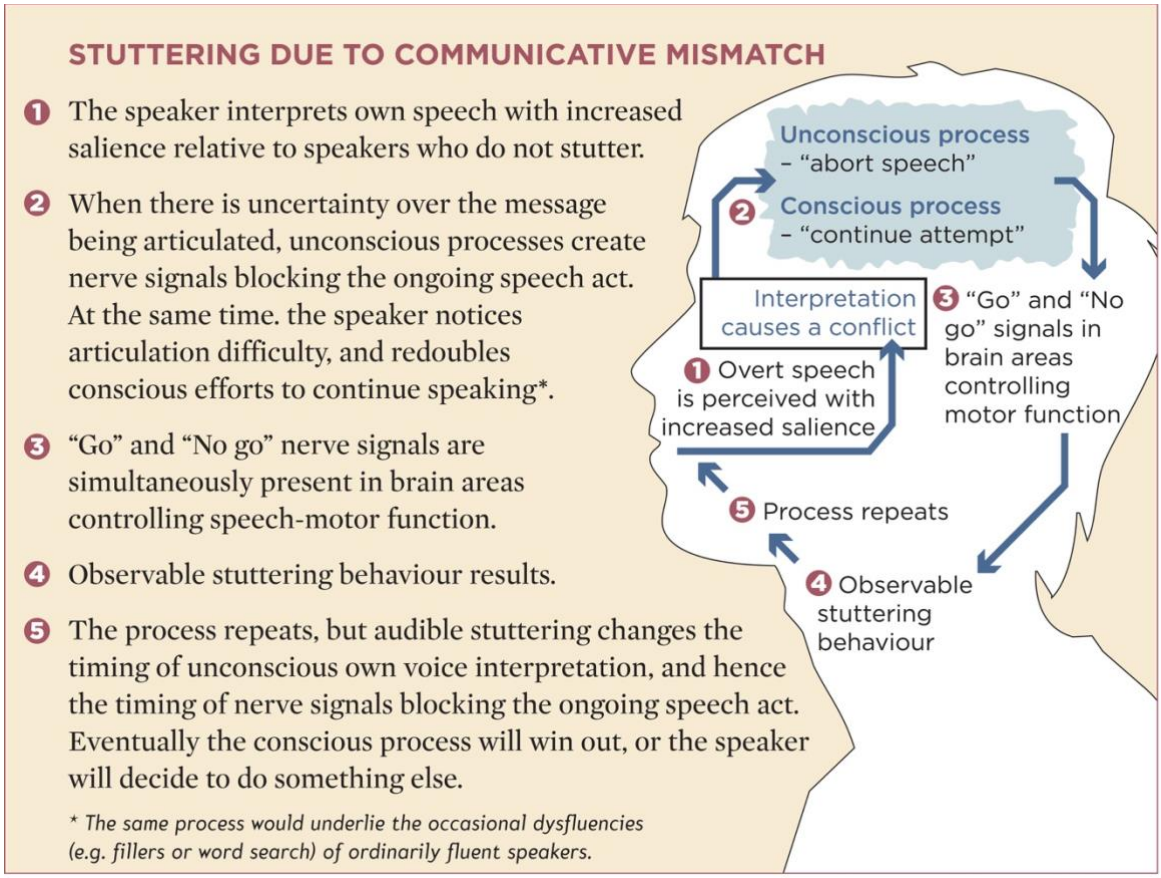

Figure 7: Stuttering due to communicative mismatch. This is based on the approach-avoidance conflict model of stuttering developed by Sheehan (1953, 1958, 1970, 1975), but updated to reflect contemporary understanding of unconscious processes.

From a psycholinguistic perspective, REMATCH places stuttering at the semantic-pragmatic interface. In Levelt's model (figure 6) stuttering would occur within the Conceptualiser. This differs from the psycholinguistic models in section 3.2.3.1, which place stuttering in the Formulator (or, for EXPLAN, between the the Formulator and Articulator). Section 3.4.1.2 will discuss a way to reconcile such psycholinguistic models within REMATCH.

3.4 Explanatory Power

3.4.1 Biological considerations

3.4.1.1 Neurological substrate

Systematic review of grey matter structural neuroimaging in adults who stutter (AWS) shows increased volume in the right superior temporal cortex and right precentral cortex compared to control groups of ordinarily fluent speakers (Etchell et al., 2017; see review for finer granularity and additional areas). These are homologues of areas in the left hemisphere thought to be important for speech and language (Hickok & Poeppel, 2007). Activation likelihood estimation (ALE) meta-analysis of diffusor tensor imaging shows AWS have reduced fractional anisotropy in the callosal body, and in dorsal white matter tracts connecting grey matter regions considered important for auditory and motor function (Neef et al., 2015; see review for finer granularity and additional areas). Possible interpretations of reduced FA include demyelination, larger axon diameter, lower packing density or increased axonal membrane permeability

(Jones et al., 2013). Fractional anisotropy does not detail direction of information flow between grey matter areas.

ALE meta-analyses of functional neuroimaging in AWS during speech tasks show overactivation in areas corresponding to motor activity and underactivation in areas corresponding to auditory activity (Budde et al., 2014; Belyk et al., 2015; see state/trait comparisons in these meta-analyses for additional areas and finer granularity). Cerebellar vermis in AWS is underactive compared to controls, but overactive during stuttering. Neuroimaging of children who stutter shows differences from controls in several of the areas identified for adults who stutter (see, for example, Garnett et al., 2018; Kronfeld-Duenias et al., 2018; Koenraads et al. 2020), suggesting that the neurodevelopmental trajectory for stuttering diverges from that of ordinarily fluent speakers close to stuttering onset.

The core hypotheses of reflexivity and communicative mismatch will be traced through the brain areas just described. The route followed will describe a chronological sequence from audition of own voice, through cortical activity consistent with an ongoing speech act, to the creation of observable stuttering via speech-motor activity. References are as per the review articles already cited (Etchell et al., 2017, Neef et al., 2015, Belyk et al., 2015 and Budde et al., 2014) with further references introduced as necessary.

According to the Concurrency Hypothesis (section 2), own voice will be identified through coincidence detection between cochlear and vestibular afferents. Only two studies have assessed the vestibular system in PWS. Langová et al. (1975) found that horizontal nystagmus evoked during speech is more pronounced in PWS than in controls using rotary chair testing. Gattie et al. (submitted) found the vestibular-evoked myogenic potential, an indirect functional test of the vestibular brainstem and periphery, is smaller in PWS than in controls. The suggestion is of divergence in central vestibular function, and/or the nature of conduction along the VIII cranial nerve, between PWS and controls. Interpreted according to the Concurrency Hypothesis (section 2.3.2), a smaller vestibular input to the auditory brainstem would correspond to a lower likelihood for coincidence detection in cells whose excitation depends upon summation of synaptic input from multiple fibres (e.g. octopus cells in the cochlear nucleus, or principal cells of the medial superior olivary complex; Golding & Oertel, 2012). It follows that the ascending auditory stream at later stages of the ascending auditory pathway, or in temporal cortex, will be more weakly identified as an own voice stream in PWS than in controls. Inputs to cerebellar vermis will also be reduced during vocalisation for PWS (i.e. as per figure 3).

The sum of activity so far (smaller vestibular input to afferent streams of neural activity through the cerebellum and auditory brainstem) would more weakly identify own voice in PWS than in ordinarily fluent speakers. This occurs because the coincidence detection proposed by the Concurrency Hypothesis will be weaker with a smaller vestibular input. The weaker identification of own voice would in turn correspond to the increased reflexivity hypothesised for PWS. It is almost as if own voice is interpreted as for the voice of another speaker. From a systems control perspective (e.g. as per Jones et al., 2016), this would be referred to as inadequate sensory gating of the own voice auditory stream.

In the cerebrum, afferent own voice streams mediated via auditory brainstem and cerebellum could alter function in two brain areas which have repeatedly been identified as important in stuttering research. One of these is the cortico-basal ganglia-thalamo-cortical loop (Milardi et al., 2019), which is reviewed specifically in relation to stuttering by Chang & Guenther (this issue). The other area is temporal cortex, and in particular the temporo-parietal junction. Recall in this regard the discussion of section 2.4.1, that

an own voice auditory stream would provide a target for the proposed efference copy of the speech plan in predictive feedback control models. Several authors have proposed that a difference in such moderation, or in auditory-motor mapping, between PWS and controls underlies the observed stuttering behaviour (Max et al., 2004; Brown et al., 2005; Hickok et al., 2011; Cai et al., 2012). The proposals have received little support in direct tests (e.g. Beal et al., 2010; Liotti et al., 2010). However, tests have used vocalisation versus AC playback protocols. As described in section 2.4.1, vocalisation versus AC playback protocols do not use a physiologically valid own voice stimulus in the playback condition, and as such do not evaluate speech-induced suppression. Accordingly, the proposal that moderation of temporal cortex by speech-motor activity differs between PWS and controls remains live. It is one of the possibilities for the hypotheses of concurrency and reflexivity when applied to stuttering via a predictive feedback control model. Investigation of the temporo-parietal junction is of particular interest, because it has repeatedly been identified as important for self-other distinction (Steinbels, 2016), and contains an area in the Sylvian fissure hypothesised as important for language control (Hickok, 2017). The role of the cerebellum may also be crucial. The cerebellum is repeatedly found to have involvement in speech perception (meta-analysis in Skipper & Lametti, 2021). This includes high level tasks involving semantics, grammar, and comprehension (Ackermann & Brendel, 2016; Mariën & Manto, 2015).

The hypothesis of communicative mismatch is based on approach-avoidance conflict (Sheehan 1953; 1958; 1970; 1975), which contemporary research (review in Aupperle & Paulus, 2010; Barker et al., 2019) places in the insula, amygdala, prefrontal cortex and the basal ganglia. All of these areas have been identified as showing a difference between PWS and controls in neuroimaging research (Yang et al., 2017; Toyomura et al., 2018; Budde et al, 2014; Etchell et al., 2017; see Garcia-Barrera & Davidow, 2015, for discussion of connection between prefrontal and anterior cingulate cortex in error monitoring). The basal ganglia in particular are crucial to the hypothesis of communicative mismatch. This is because conflict between selection and inhibition of competing actions, sometimes termed as “Go” and “No Go” (Mink, 1996; Bahuguna et al., 2015; Dunovan et al., 2015; Mink, 2018), could create involuntary muscular activity similar to that observed in stuttering. Thus, following the sequence described in this section from audition to articulation, basal ganglia activity would be the most proximal cause of observable stuttering behaviour (see Arenas, 2017, for a proposal emphasising functional importance of the subthalamic nucleus). Frontal and parietal cortex associated with speech-motor control would show state and trait differences in stuttering due their involvement in basal ganglia pathways (Albin et al., 1989; DeLong et al, 1990; Calabresi et al., 2014), and also due to white matter connection to temporal cortex important for auditory-motor integration (e.g. the efference copy proposed in speech-motor models), including commissural connection to homologues. The basal ganglia and cerebellum are interconnected via the thalamus (Hoshi et al., 2005; Bostan et al., 2010, Pelzer et al., 2017; Caligiore et al., 2017; Cacciola et al., 2017; Bostan & Strick, 2018) and are both proposed to have involvement in language processing and vocal learning (Booth et al., 2007; Pidoux et al., 2018). This underscores the prospective importance of the vestibular-cerebellar pathway (figure 3) for own voice identification in stuttering, and of cerebellar input to the cortico-basal ganglia-thalamo-cortical loop in stuttering (Chang & Guenther, this issue). Dopamine levels in the basal ganglia affect action selection (Mink, 1996; Reynolds et al., 2001; Haber, 2014; Schultz, 2016), with differences in the dopamine system between PWS and controls found using positron electron tomography (Wu et al., 1997) and through pharmaceutical intervention (Maguire et al., this issue). The basal ganglia have been repeatedly identified as important in stuttering research (Alm, 2004; Metzger et al., 2017; Chang & Guenther, this issue).

3.4.1.2 Stuttering subtypes

The neurological substrate described in the previous section encompasses almost the entire brain. This raises the possibility that preconditions for stuttering may require a difference from ordinarily fluent

788 speakers in the function of not just one brain area, but several (ie. as per Ludlow & Loucks, 2003). Such a
789 view underlies multifactorial models of stuttering (e.g. Smith & Kelly, 1997; Starkweather, 2002; Walden
790 et al., 2012; Smith & Weber, 2017), which examine the interplay between genetic, organismic and
791 environmental contributing factors.

792 One way to develop such models would be through subtyping stuttering. If there are discrete groupings
793 of factors which contribute to stuttering, separation into such groupings prior to data analysis could
794 enable more granular investigation and facilitate hypothesis formulation. Unfortunately, subtyping
795 stuttering is difficult (see review in Yairi, 2007; Seery et al., 2007), largely due to the challenges of
796 longitudinal data collection.

797 Subtyping is proposed here, based on the four track system of Van Riper (1973; 1982 p94–108).
798 Symptomatology is identical to that of Van Riper, but the account is extended with the proposal that
799 causation differs between tracks. Track I is proposed to correspond to stuttering developing as an
800 isolated diagnosis. Track II corresponds to stuttering co-developing with at least one other diagnosis.
801 Tracks III and IV are trauma-based, and may be psychogenic or neurogenic.

802 Track I would have a genetic basis and be based around increased reflexivity. The genetic basis may
803 affect several brain areas (i.e. as per Ludlow & Loucks, 2003). For example, genetic investigation of
804 stuttering has suggested that the nature of white matter may be integral to stuttering behaviour –
805 lyosomal pathways or glial cells are implicated (Han et al., 2019; Benito-Aragón et al. 2020). PWS have
806 reduced fractional anisotropy in dorsal white matter tracts which connect cortical regions having speech-
807 motor and auditory function (Neef et al., 2015; Etchell et al., 2017). It would appear that genetic
808 variations in stuttering might be connected to the structure of these dorsal white matter tracts. If so, it is
809 not clear why the white matter structural variation should be focal to just these dorsal tracts (Watkins &
810 Büchel, 2010; Drayna, 2010). One possibility is that genetic variation affects several white matter tracts.
811 It may, for example, also manifest as reduced fractional anisotropy in the vestibular portion of the VIII
812 cranial nerve. If so, the variation would be consistent with the finding of a weaker vestibular response in
813 PWS by Gattie et al. (submitted), and would support interpretation according to the concurrency and
814 reflexivity hypotheses presented in the current article.

815 This suggestion around genetics is just one example of a long-term investigative target for stuttering
816 research. Many other possibilities pertain – not only variations within the neurological substrate
817 described for stuttering in section 3.4.1.1, but moreover the interplay between genetic, organismic and
818 environmental factors described in multifactorial models of stuttering. Investigation of which factors are
819 necessary and/or sufficient for behavioural stuttering to manifest is a topic for ongoing research.

820 From this perspective, track II stuttering is a particular version of track I in which one of the variations
821 contributing to the co-occurring diagnosis also contributes to stuttering behaviour. The existence of track
822 II stuttering would explain why stuttering co-occurs with other diagnoses at a rate higher than chance.
823 Depending on the nature of co-occurring diagnoses (and accepting that co-occurring diagnoses will
824 sometimes remain pending) existence of track II stuttering could also explain why testing groups of PWS
825 sometimes results in a subtle difference from controls in abilities such as executive function, language,
826 reaction time and general motor control. The explanation would be that a co-occurring diagnosis, rather
827 than stuttering, is causative of the test result. Such a distinction has been established in studies which
828 split the PWS group into those with and without a co-occurring diagnosis (e.g. Cullinan & Springer, 1980;
829 Liebetrau & Daly, 1981; McKnight & Cullinan, 1987; Kobayashi & Hayasaka, 2003)

REMATCH places stuttering at the semantic-pragmatic interface, but is compatible with the psycholinguistic models introduced in section 3.2.3.1. Application of the psycholinguistic models will help to distinguish tracks I–IV. Stuttering in tracks I and II could be described by a combination of EXPLAN (Howell, 2004; 2008) and the Variable Release Threshold hypothesis (Brocklehurst et al., 2013). It is suggested here that the account is updated such that it is REMATCH, rather than the compatible but less detailed anticipatory struggle hypothesis of Bloodstein (1975), which provides detail of the release threshold.

Tracks III and IV stuttering are proposed as having a trauma-based origin which could be psychogenic or neurogenic. When the trauma results in heightened self-awareness (Gallagher & Zahavi, 2021; Smith, 2020) increased attention to self-monitoring would follow. From the perspective of REMATCH the increased attention to self-monitoring would equate to increased reflexivity through a route other than the auditory system. Essentially, there is increased self-doubt about any speech act. In some cases the trauma could follow a profound emotional event (e.g. a bereavement or a family break-up), but it could also follow a more subtle series of events (Starkweather, 2002, lists possible environmental influencers on stuttering). Such stuttering could be described by the Vicious Circle Hypothesis (Vasić & Wijnen, 2005), in which monitoring of phonological error becomes hyper-vigilant. This type of stuttering could alternatively be explained from the perspective of EXPLAN and the Variable Release Threshold hypothesis. It would correspond to a release threshold which varies similarly to that of an ordinarily fluent speaker, but which is continuously subject to a multiplier greater than unity.

Other instances of tracks III and IV could be primarily caused by neural insult (e.g. transient ischemic attack, traumatic brain injury or neurodegenerative disease). If the effect of the neural insult is to alter function of a brain area important for phonological formulation, this type of stuttering could be described by the Covert Repair Hypothesis. However, neurogenic stuttering will be the most difficult to model. If the diagnosis is of a progressive neurological condition, stuttering may be transient prior to being masked by a wider range of symptoms involving language, speech-motor or executive function. In neurogenic stuttering with no other symptoms, behaviour may differ from tracks I–II due to the alteration in brain function having a random structural cause (neural insult) rather than proceeding through a genetic, developmental or psychological route.

Tracks I-IV may show overlap. For example, a child may have a genetic disposition to stuttering (track I) and experience environmental conditions creating psychosocial pressure (track III). This notion underlies the Demands and Capacities model (Adams, 1990; Starkweather & Gottwald, 1990; Starkweather, 2002), which is frequently interpreted as a genetic predisposition to stuttering becoming concrete following environmental influence. However, the predisposition need not be genetic; combinations of any of tracks I–IV, and/or single track etiologies, could just as well result in stuttering behaviour.

Developmental stuttering could involve any of tracks I–IV. Whereas absence of a plausible genetic or developmental contributory mechanism to stuttering in adulthood seems to limit adult onset stuttering to tracks III and IV. Thus, the rarity of adult onset stuttering (Ward, 2006, ch 16) is consistent with Van Riper's (1973; 1982) finding that between 80–90% of his 300-strong caseload were tracks I or II. An exception would be adult onset where there is a history of childhood stuttering (Van Riper, 1982 p66). In such cases, reappearance may have psychogenic or neurogenic influence. For example, Shahed & Jankovic (2001) describe 12 persons who had stuttered in childhood but not as adults, and for whom stuttering reappeared following a diagnosis of Parkinson's disease.

3.4.2 Other explananda

The following sections address the explananda in tables 2 and 3, to which the reader might simultaneously refer.

3.4.2.1 Linguistic and Situational

Variation in phonological formulation between stuttering subtypes was described in section 3.4.2. However, the main linguistic hypothesis within REMATCH is that own speech is interpreted in PWS with increased salience. Recall now the explananda in table 2. Unconscious processes are proposed to block an ongoing utterance whenever there is uncertainty about a speech act. Uncertainty is proposed to increase with propositionality, and hence stuttering will correlate with propositionality. Without an audience, a speech act cannot be performed. This explains why, unless PWS project an audience, stuttering will not occur when alone (Langová & Sváb, 1973). With authority listeners, even ordinarily fluent speakers experience increased salience when executing a speech act. For PWS, increased salience due to the authority listener combines with increased salience due to reflexivity, increasing the propensity for stuttering according to REMATCH.

These proposals are consistent with the observation of Sheehan (1958) that speech breakdown in stuttering coincides with the requirement “to say something important to someone important”. The proposals could be tested by following theoretical frameworks for pragmatics and social convention (e.g. Grice. 1957, 1989; Rescorla, 2019). For example, the exact loci of stuttered instances could be a project in experimental pragmatics (Noveck & Sperber, 2004; Meibauer & Steinbach, 2011; Noveck, 2018). Such a project might initially appear circular (stuttered phonemes are predefined as those with high propositionality). However, corpora of stuttered speech provide rich data, and can therefore be investigated following themes in pragmatics (e.g. Gricean implicatures, epistemic vigilance) using statistical techniques such as latent class or principal components analysis. Such an approach could also appraise changes in language use with development (e.g. within people who stutter there is a tendency for children to stutter on function words, and adults to stutter on content words). See also Eisenson & Horowitz (1945), Sheehan et al. (1967), Gould & Sheehan (1967) and MacKay (1969) for examples of work which could fit within a research programme for experimental pragmatics in stuttering.

3.4.2.2 Anticipation, Consistency and Adjacency

Speakers can unconsciously scan ahead. This applies to spontaneous speech or when reading aloud. If message content scanned ahead is interpreted by the speaker according to REMATCH, the person who stutters will be able to predict when speech difficulty is imminent. In oral readings, uncertainty around any particular word is unchanged on repeated readings, because the underlying message has not changed. Therefore, stuttering has the same loci on repeat readings. When words are blotted out, the reader unconsciously anticipates what the word would have been (or infers intended meaning from the words remaining) leading to stuttering on the word that is unconsciously predicted to convey intended meaning to a listener. This will usually be an adjacent word to the word previously stuttered.

3.4.2.3 Adaptation

Propositionality is reduced on repeat readings, since the listener is already aware of the message being delivered, and the speaker is aware of the message as well. Reduced propositionality in turn reduces salience. According to REMATCH, reduced salience will reduce the tendency for the speaker to unconsciously block an ongoing speech act. Essentially, reduced propositionality acts as a counter for the increased reflexivity proposed in PWS.

3.4.2.4 Operant Conditioning

Speech-motor or psycholinguistic breakdown accounts of stuttering appeal to emotional or psychosocial stress to explain situational variation. A problem for such accounts is that they predict that stress should be very high in laboratory conditions with response-contingent stimulation (e.g. electric shock or time out upon stuttering), and therefore stuttering should increase. However, the converse is found: stuttering decreases with response-contingent stimulation.

The finding can be explained by REMATCH as response-contingent stimulation forcing an attentional shift in the speaker. The attentional shift is towards an increased conscious control of speech. This shift diminishes the influence of unconscious processing of speech, which according to REMATCH (figure 7) will reduce the amount of stuttering. Increasing conscious control is sufficiently effortful that unless an operant speaking technique such as fluency shaping has been learnt, PWS will not increase conscious control volitionally (see Constantino et al., 2020, for extended discussion). However, the continuous presence of response-contingent stimulation in laboratory conditions makes increased conscious control of speech unavoidable for the speaker.

3.4.2.5 Alterations to audition during speech

Alterations to audition during speech will affect own voice identification according to the Concurrency Hypothesis (section 2). Alterations which are effective are proposed to reduce reflexivity, and thereby to reduce stuttering according to REMATCH. The exact detail of audition changes effective for reducing stuttering is a topic for ongoing research. Timings in table 1 show a starting point. The effectiveness of long delays (e.g. 50 ms or more) may have more to do with phoneme or syllable recognition, or word recognition, than with own voice identification. If so, effectiveness of particular delay lengths will be variable, because the duration of word-initial phonemes is variable. The prediction from the Concurrency Hypothesis is that alterations on the time scale of a millisecond or less will be most effective. Such rapid alterations have not been tested other than by Howell et al. (1987), who showed that frequency shifts with a delay on the order of one millisecond were more effective at reducing stuttering than delays of 50 ms. Alterations most effective for reducing stuttering may depend on individual physiologies. If so, there is a prospect for tailoring the delay to individuals depending on EEG measurements. Such a project would be a part of, or be informed by, the investigation of own voice identification outlined in section 2.

In alterations to audition involving a second speaker (shadowing or unison speaking) there is an additional benefit in that propositionality is also reduced (the second speaker is already aware of the message being delivered, and is encouraging delivery of that message). For this reason, unison speaking is the most effective way of reducing stuttering.

3.4.2.6 Therapy effectiveness

REMATCH identifies the proximal cause of core stuttering behaviour as simultaneous “Go” and “No Go” signals in brain areas coordinating articulatory muscles, as described in section 3.4.1.1. Accessory and interiorised stuttering behaviours are explained in this regard as attempts by the speaker to resolve core stuttering behaviours (i.e. explanation as per Van Riper, 1982, ch 6–7).

Early stages of many stuttering therapies (e.g. the motivation, identification and desensitisation stages described by Van Riper, 1973) include psychological therapy, helping speakers to unlearn accessory and interiorised stuttering behaviours which have become engrained through habit. These early stages of stuttering therapy increase approach and decrease avoidance behaviours. For example, desensitisation therapy reduces emotionality attached to speaking situations. It is proposed that reduced emotionality will decrease the tendency to unconsciously block an ongoing speech act, and increase willingness to speak. The effect would be to reduce reflexivity, and thereby decrease stuttering according to REMATCH.

Speech work in therapies (e.g. the variation and adaptation stages in Van Riper, 1973) deliberately introduces prolongation to the beginning of syllables. Prolongation acts similarly to an alteration to audition during speech, and thereby reduces stuttering as described in section 3.4.2.5. An alternative strategy having the same effect would be to deliberately introduce repetition (Johnson. 1961). However, deliberate repetition is seldom used, perhaps because it is more noticeable than prolongation.

A major distinction between therapies is whether prolongation is on every syllable (fluency shaping) or only on syllables where stuttering is anticipated (block modification). See discussion in Ingham (1984, p328) or Gregory (1979). Prolongation on every syllable entails a continued attentional shift whilst talking. According to REMATCH, continued attentional shift will reduce stuttering (see section 3.4.2.4). Thus, fluency shaping has two methods reducing the amount of stuttering: syllable initial prolongation, and attentional shift. This would explain why fluency shaping programmes are often more effective than block modification programmes in reducing the amount of stuttering. However, fluency shaping programmes are effortful for the speaker (Constantino, 2020) and for this reason many PWS will prefer a block modification approach.

3.5 Discussion of the REMATCH hypothesis

The REMATCH hypothesis draws together breakdown and anticipatory struggle hypotheses of stuttering. In this sense, it is similar to, and compatible with, the Variable Release Threshold hypothesis of Brocklehurst et al. (2013). REMATCH goes into additional detail by specifying that the type of anticipatory struggle is an updated version of the approach-avoidance conflict proposed by Sheehan (1953; 1958; 1970; 1975). This update situates REMATCH in what Levelt (1989, 1999) refers to as the Conceptualiser. Thus, REMATCH is fundamentally different from (although compatible with) hypotheses which explain stuttering as breakdown in what Levelt refers to as the Formulator and/or Articulator. From this perspective, a major contribution of REMATCH is to provide a framework through which psycholinguistic and situational variation in stuttering can be investigated.

The updated approach-avoidance conflict in REMATCH is explained through a view of the unconscious proposed to be similar to that in dual process theory (e.g. as per Evans, 2007; Kahneman, 2011), and containing a high degree of automaticity. This provides a basis for investigation using cognitive science methodologies (e.g. as per the experimental pragmatics of Noveck & Sperber, 2004). The unconscious process in REMATCH could just as well have been explained as an update of repressed needs hypotheses of stuttering, in which the view of the unconscious is no longer necessarily that of psychoanalytic theory. This is possible because REMATCH contains a description of the moment of stuttering (figure 7) which can be compared to and informed by first person accounts. Thus, REMATCH promotes integration of qualitative and quantitive work in stuttering, and can furthermore provide a link to phenomenological accounts of stuttering (e.g. Ellis, 2020; Isaacs, 2020). Such integrations could inform psychological therapies for stuttering. They could also help to promote a social model of stuttering (Campbell et al., 2019), even within a world where neuroscientific research will remain within a medical model. Efforts in this direction are important if stuttering research is to be relevant to people who stutter.

4. General discussion

This article has described hypotheses of own voice identification and stuttering. The account has been highly detailed and with very broad scope because, as described in section 1, all of the hypotheses are proposed together as a best explanation argument. As such, it is necessary to show that the combined explanation has a high degree of explanatory power and parsimony.

The crux of this article is the Concurrency Hypothesis that own voice is identified through coincidence detection between the neural firing rates arising from deflection of cochlear and vestibular mechanoreceptors by the sound and vibration generated during vocalisation. Section 2 describes how the Concurrency Hypothesis provides a principled basis for self-environment distinction, with importance for considerations in cognitive science and philosophy of mind. The Concurrency Hypothesis was also applied to speech-motor research, in which it highlighted limitations in empirical support for the proposal that speech-motor activity modulates activity in temporal cortex. Finally, the Concurrency Hypothesis was applied to auditory scene analysis, in which it is proposed to provide the basis for a system of discrimination in multi-talker scenarios.

In section 3, the Concurrency Hypothesis was developed into an explanation of stuttering. The initial step was to propose a quale, reflexivity. This refers to the phenomenology of hearing one's own voice, and is proposed to differ between people who do and do not stutter. The account was then developed into an update of the approach-avoidance conflict model of stuttering (Sheehan 1953; 1958; 1970; 1975), referred to as REMATCH. This explains the moment of stuttering as a communicative mismatch. The speaker experiences own voice with increased salience, but this creates a mismatch whenever there is uncertainty about the ongoing message. In such cases, unconscious processes reinterpreting the message create nerve signals blocking the ongoing speech act, at the same time the speaker is consciously trying to continue. The resultant conflict is behaviourally observable as stuttering.

The Concurrency Hypothesis and REMATCH are core hypotheses. Many auxiliary hypotheses were introduced, mainly within the account of stuttering. These include the neurological substrate for stuttering, a proposal for subtyping stuttering, and a variety of process and contrastive explanations of data from stuttering research. These auxiliary hypotheses are likely to change with time, and are provided here as a snapshot so that the scope of the intended explanation of stuttering is apparent.

The Concurrency Hypothesis could be applied groups other than people who stutter, and who are expected to show differences from controls in own voice identification. Some examples of such groups include those experiencing auditory and/or vestibular neuropathy (Kaga, 2016) and those experiencing auditory hallucination (e.g. in schizophrenia – McLachlan, Phillips, Rossell & Wilson, 2013; Matthews et al., 2013; Weintraub et al., 2012; Waters & Fernyhough, 2019).

Testable predictions generated by the hypotheses in this article are described in sections 2 and 3. One of these predictions is that people who stutter should show a difference from controls in tests of the vestibular system. This was appraised by Gattie et al (submitted) with the finding that vestibular response is weaker in people who stutter than in paired controls. The result is consistent with the only prior research on the vestibular system in people who stutter (Langová et al., 1975) and supports the hypotheses presented in this article.

5. Conclusion

The major recommendation from this article is that researchers should use physiologically valid stimuli when investigating own voice in speech and language research. Using stimulation over air conduction only, even with a sound pressure level increase to perceptually match the loudness experienced during vocalisation, does not generate physiologically valid stimuli. Instead, stimuli should consist of a combination of air conducted sound and body conducted vibration which is binaurally symmetric, and has coincident arrival at both inner ears.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Conceptualization: ideas; formulation of the overarching research goals and aims
MG

Methodology: development or design of methodology or creation of models
MG

Software: programming; software development; designing computer programs; implementation of computer code or algorithms; testing code components
MG

Validation: verification of the replication and reproducibility of results, experiments, or other research outputs
MG

Formal analysis: application of statistical, mathematical, computational, or other techniques to analyse or synthesise data
MG

Investigation: conducting the research and investigation process, specifically performing the experiments or data collection
MG

Resources: provision of study materials, materials, instrumentation, computing resources, or analysis tools
MG

Data curation: annotation, scrubbing, or maintenance of research data (including software code, where it is necessary for interpreting the data itself)
MG

Writing—original draft: preparation, creation and/or presentation of the published work, specifically writing the initial draft
MG

Writing—review & editing: critical review, commentary, or revision
MG, KK, EL

Visualization: preparation, creation and/or presentation of the published work, specifically data presentation or visualisation
MG

Supervision: oversight and leadership responsibilities, including mentorship
KK, EL

Project administration: coordination of the research activity planning and execution
MG, KK, EL

Funding acquisition: acquisition of financial support
MG, KK

Funding

This work was supported by a UK Economic and Social Research Council (ESRC) Collaborative Award in Science and Engineering (CASE) to Max Gattie. The CASE partner was Interacoustics A/S, Middelfart, Denmark. Karolina Kluk was supported by NIHR Manchester Biomedical Research Centre.

Acknowledgments

Thanks to Paul Brocklehurst for comments on this article. Any shortcomings in the finished article are solely the responsibility of the named authors.

References

Abdalla, S. (2011). Low Frequency Dielectric Properties of Human Blood. *IEEE Transactions on NanoBioscience*, *10*(2), 113–120. http://doi.org/10.1109/TNB.2011.2159734

Acheson, D. J., & Hagoort, P. (2014). Twisting tongues to test for conflict-monitoring in speech production. *Frontiers in Human Neuroscience*, *8*, 206. http://doi.org/10.3389/fnhum.2014.00206

Ackermann, H., & Brendel, B. (2016). Cerebellar contributionsto speech and language. In G. Hickok & S. L. Small (Eds.), Neurobiology of language(pp. 73–84). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00007-9

Adams, J. C. (1997). Projections from octopus cells of the posteroventral cochlear nucleus to the ventral nucleus of the lateral lemniscus in cat and human. *Aud. Neurosci.* 3, 335–350.

Adams, M. R. (1990). The Demands and Capacities Model I: theoretical elaborations. *Journal of Fluency Disorders*, **15**, 136–141.

Agmon-Snir, H., Carr, C. E., & Rinzel, J. (1998). The role of dendrites in auditory coincidence detection. *Nature*, *393*(6682), 268–272. http://doi.org/10.1038/30505

Ajdacic-Gross, V., Vetter, S., Müller, M., Kawohl, W., Frey, F., Lupi, G., et al. (2010). Risk factors for stuttering: a secondary analysis of a large data base. *European Archives of Psychiatry and Clinical Neuroscience*, *260*(4), 279–286. http://doi.org/10.1007/s00406-009-0075-4

Ajdacic-Gross, V., Bechtiger, L., Rodgers, S., Müller, M., Kawohl, W., Känel, von, R., et al. (2018). Subtypes of stuttering determined by latent class analysis in two Swiss epidemiological surveys. *Plos One*, *13*(8), e0198450. http://doi.org/10.1371/journal.pone.0198450

Alexander, G.E., Newman, J.D. and Symmes, D. (1976), Convergence of prefrontal and acoustic inputs upon neurons in the superior temporal gyrus of the awake squirrel monkey. Brain Res., 116 334–338.

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. *Journal of Communication Disorders*, *37*(4), 325–369. http://doi.org/10.1016/j.jcomdis.2004.03.001

Albin, R.L., Young, A.B. & Penney, J.B. (1989). The functional anatomy of basal ganglia disorders. Trends Neurosci. 12, 366–375.

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 : JSLHR*, *40*(3), 567–580. http://doi.org/10.1044/jslhr.4003.567

Andrews G. & Harris M. (1964) The Syndrome of Stuttering. London, UK: Spastics Society Medical Education and Information Unit in association with Heinemann Medical.

Andrews, G., Craig, A., Feyer, A.-M., Hoddinott, S., Howie, P., & Neilson, M. (1983). Stuttering: A review of research findings and theories circa 1982. Journal of Speech and Hearing Disorders, 48, 226-246.

Arenas, R. M. (2017). Conceptualizing and investigating the contextual variability of stuttering: The speech and monitoring interaction (SAMI) framework. *Speech, Language and Hearing*, *20*(1), 15–28. http://doi.org/10.1080/2050571X.2016.1221877

Arend, J. M., Stade, P., & Pörschmann, C. (2017). Binaural reproduction of self-generated sound in virtual acoustic environments (pp. 015007–15). Presented at the 173rd Meeting of Acoustical Society of America and 8th Forum Acusticum, Acoustical Society of America. http://doi.org/10.1121/2.0000574

Arndt, J., & Healey, E. C. (2001). Concomitant disorders in school-age children who stutter. Language, Speech, and Hear- ing Services in Schools, 32, 68–78.

Ashida, G., & Carr, C. E. (2011). Sound localization: Jeffress and beyond. *Current Opinion in Neurobiology*, *21*(5), 745–751. http://doi.org/10.1016/j.conb.2011.05.008

Atencio, D. J., & Montero, I. (2009). Private speech and motivation: The role of language in a sociocultural account of motivational processes. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, exec- utive functioning, and the development of verbal self-regulation* (pp. 201–223). Cambridge, UK: Cambridge University Press. http://dx.doi .org/10.1017/CBO9780511581533.017

Aupperle, R. L., & Paulus, M. P. (2010). Neural systems underlying approach and avoidance in anxiety disorders. *Dialogues in Clinical Neuroscience*, *12*(4), 517–531. http://doi.org/10.31887/DCNS.2010.12.4/raupperle

Bahuguna, J., Aertsen, A., & Kumar, A. (2015). Existence and control of Go/No-Go decision transition threshold in the striatum. *PLoS computational biology*, *11*(4), e1004233. https://doi.org/10.1371/journal.pcbi.1004233

Bajo, V. M., & King, A. J. (2012). Cortical modulation of auditory processing in the midbrain. *Frontiers in Neural Circuits*, *6*, 114. http://doi.org/10.3389/fncir.2012.00114

Barker, M., Solinski, H. J., Hashimoto, H., Tagoe, T., Pilati, N., & Hamann, M. (2012). Acoustic overexposure increases the expression of VGLUT-2 mediated projections from the lateral vestibular nucleus to the dorsal cochlear nucleus. Plos One, 7(5), e35955. http://doi.org/10.1371/journal.pone.0035955

Barker, T. V., Buzzell, G. A., & Fox, N. A. (2019). Approach, avoidance, and the detection of conflict in the development of behavioral inhibition. *New Ideas in Psychology*, *53*, 2–12. http://doi.org/10.1016/j.newideapsych.2018.07.001

Barber, V.B. (1939) Studies in the psychology of stuttering: XV. Chorus reading as a distraction in stuttering. Journal of Speech Disorders, 4, 371–383.

Barber, V.B. (1940) Studies in the psychology of stuttering: Rhythm as a distraction in stuttering. Journal of Speech Disorders, 5, 29–42.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68, 255–278. http://dx.doi.org/ 10.1016/j.jml.2012.11.001.

Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes.

Baumgartner, J., & Duffy, J. R. (1997). Psychogenic stuttering in adults with and without neurologic disease. *Journal of Medical Speech-Language Pathology*, *5*(2), 75–95.

Baxter, S., Johnson, M., Blank, L., Cantrell, A., Brumfitt, S., Enderby, P., & Goyder, E. (2015). The state of the art in non-pharmacological interventions for developmental stuttering. Part 1: a systematic review of effectiveness. *International Journal of Language & Communication Disorders*, *50*(5), 676–718. http://doi.org/10.1111/1460-6984.12171

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*(4), 1645–1653. http://doi.org/10.1016/j.neuroimage.2010.04.277

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., & Greenlee, J. D. W. (2015). Sensory-motor networks involved in speech production and motor control: an fMRI study. *NeuroImage*, *109*(C), 418–428. http://doi.org/10.1016/j.neuroimage.2015.01.040

Beiderbeck, B., Myoga, M. H., Müller, N. I. C., Callan, A. R., Friauf, E., Grothe, B., & Pecka, M. (2018). Precisely timed inhibition facilitates action potential firing for spatial coding in the auditory brainstem. *Nature Communications*, *9*(1), 1771–13. http://doi.org/10.1038/s41467-018-04210-y

1173 von Békésy, G. (1949). "The structure of the middle ear and the hearing of one's own voice by bone
1174 conduction," J. Acoust. Soc. Am. 21, 217–232.
1175 Benito-Aragón, C., Gonzalez-Sarmiento, R., Liddell, T., Diez, I., d'Oleire Uquillas, F., Ortiz-Terán, L., et al.
1176 (2020). Neurofilament-lysosomal genetic intersections in the cortical network of stuttering. *Progress*
1177 *in Neurobiology*, *184*, 101718. http://doi.org/10.1016/j.pneurobio.2019.101718
1178 Beving, H., Eriksson, L. E., Davey, C. L., & Kell, D. B. (1994). Dielectric properties of human blood and
1179 erythrocytes at radio frequencies (0.2-10 MHz); dependence on cell volume fraction and medium
1180 composition. *European Biophysics Journal : EBJ*, *23*(3), 207–215.
1181 BinKhamis, G., A. Léger, S. L. Bell, G. Prendergast, M. O'Driscoll, and K. Kluk. 2019. "Speech Auditory
1182 Brainstem Responses: Effects of Background, Stimulus Duration, Consonant–Vowel, and Number of
1183 Epochs." *Ear and Hearing* 40(3): 659–670. doi:10.1097/AUD.0000000000000648.
1184 Blood, G. W., Ridenour, V. J., Jr, & Qualls, C. D. (2003). Co-occurring disorders in children who stutter.
1185 *Journal of Communication Disorders*, *36*, 427–448.
1186 Bloodstein, O. (1995) A Handbook on Stuttering (5th ed), Singular.
1187 Bloodstein, O. (2002). Early stuttering as a type of language difficulty. *Journal of Fluency Disorders*, *27*(2),
1188 163–6– quiz 166–7. http://doi.org/10.1016/s0094-730x(02)00111-0
1189 Bloodstein, O. (2006). Some empirical observations about early stuttering: a possible link to language
1190 development. *Journal of Communication Disorders*, *39*(3), 185–191.
1191 http://doi.org/10.1016/j.jcomdis.2005.11.007
1192 Boulet, S. L., Boyle, C. A., & Schieve, L. A. (2009). Health care use and health and functional impact of
1193 developmental disabil- ities among US children, 1997–2005. Archives of Pediatrics & Adolescent
1194 Medicine, 163, 19–26.
1195 Booth JR, Wood L, Lu D, Houk JC, Bitan T (2007). The role of the basal ganglia and cerebellum in language
1196 processing. Brain Res.;1133:136–44.
1197 Bregman, A. S. (1990) Auditory scene analysis. MIT Press: Cambridge, MA
1198 Briley, P.M. & Ellis, C. Jr (2018). The Coexistence of Disabling Conditions in Children Who Stutter:
1199 Evidence From the National Health Interview Survey. J Speech Lang Hear Res.;61(12):2895-2905. doi:
1200 10.1044/2018_JSLHR-S-17-0378.
1201 British Society of Audiology (2012), Recommended Procedure: Cervical and Ocular Vestibular Evoked
1202 Myogenic Potentials
1203 Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech
1204 production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*(1),
1205 105–117. http://doi.org/10.1002/hbm.20140
1206 Brown, T., Sambrooks, J. E., & MacCulloch, M. J. (1975). Auditory thresholds and the effect of reduced
1207 auditory feedback of suttering. *Acta Psychiatrica Scandinavica*, *51*(5), 297–311.
1208 http://doi.org/10.1111/j.1600-0447.1975.tb00009.x
1209 Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in
1210 phonological processing for speech perception and production. *Cognitive Science*.
1211 Budde, K. S., Barron, D. S., & Fox, P. T. (2014). Stuttering, induced fluency, and natural fluency: a
1212 hierarchical series of activation likelihood estimation meta-analyses. Brain and Language, 139, 99–
1213 107. http://doi.org/10.1016/j.bandl.2014.10.002
1214 Brocklehurst, P. H., Lickley, R. J., & Corley, M. (2013). Revisiting Bloodstein's Anticipatory Struggle
1215 Hypothesis from a psycholinguistic perspective: A Variable Release Threshold hypothesis of stuttering.
1216 *Journal of Communication Disorders*, *46*(3), 217–237. http://doi.org/10.1016/j.jcomdis.2013.04.002
1217 Bernstein Ratner, N. (1997). Stuttering: a psycholinguistic perspective. In: Curlee R, Siegel GM,
1218 editors. *Nature and treatment of stuttering: New directions.* 2. Boston, MA: Allyn & Bacon. pp. 99–
1219 127.
1220 Bernstein Ratner, N., & Wijnen, F. (2007). The Vicious Cycle: Linguistic Encoding, Self- Monitoring and
1221 Stuttering. In Research, Treatment and Self-help in Fluency Disorders – New Horizons: Proceedings of
1222 the Fifth World Congress on Fluency Disorders (pp. 84–90). The International Fluency Association.
1223 ISBN-13 : 978-0955570018.
1224 Bostan A.C., Dum R.P., Strick P.L. (2010). The basal ganglia communicate with the cerebellum. Proc Natl
1225 Acad Sci U S A. 2010 May 4;107(18):8452-6. doi: 10.1073/pnas.1000496107. PMID: 20404184; PMCID:
1226 PMC2889518.

Bostan, A. C., & Strick, P. L. (2018). The basal ganglia and the cerebellum: nodes in an integrated network. *Nature reviews. Neuroscience*, *19*(6), 338–350. https://doi.org/10.1038/s41583-018-0002-7

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

Cabrera, D., Sato, H., Martens, W. L., & Lee, D. (2009). Binaural Measurement and Simulation of the Room Acoustical Response From a Person's Mouth to Their Ears. *Acoustics Australia*, *37*.

Cacciola, A., Calamuneri, A., Milardi, D., Mormina, E., Chillemi, G., Marino, S., Naro, A., Rizzo, G., Anastasi, G., & Quartarone, A. (2017). A Connectomic Analysis of the Human Basal Ganglia Network. *Frontiers in neuroanatomy*, *11*, 85. https://doi.org/10.3389/fnana.2017.00085

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: evidence for abnormal auditory-motor transformation. *Plos One*, *7*(7), e41830. http://doi.org/10.1371/journal.pone.0041830

Calabresi, P., Picconi, B., Tozzi, A., Ghiglieri, V., & Di Filippo, M. (2014). Direct and indirect pathways of basal ganglia: a critical reappraisal. *Nature Publishing Group*, *17*(8), 1022–1030. http://doi.org/10.1038/nn.3743

Caligiore, D., Pezzulo, G., Baldassarre, G., Bostan, A. C., Strick, P. L., Doya, K., Helmich, R. C., Dirkx, M., Houk, J., Jörntell, H., Lago-Rodriguez, A., Galea, J. M., Miall, R. C., Popa, T., Kishore, A., Verschure, P. F., Zucca, R., & Herreros, I. (2017). Consensus Paper: Towards a Systems-Level View of Cerebellar Function: the Interplay Between Cerebellum, Basal Ganglia, and Cortex. *Cerebellum (London, England)*, *16*(1), 203–229. https://doi.org/10.1007/s12311-016-0763-3

Campbell, P., Constantino, C. & Simpson, S. (eds) (2019). Stammering Pride and Prejudice: Difference not Deficit. J & R Press.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, *110*(7), 2653–2658. http://doi.org/10.1073/pnas.1216827110

Chang S.-E., Guenther F. H. (2020). Involvement of the cortico-basal ganglia-thalamocortical loop in developmental stuttering. Front. Psychol. 10:3088. 10.3389/fpsyg.2019.03088

Cherry, C., Sayers, B. M. & Marland, P.M. (1956). Experiments upon the total inhibition of stammering by external control, and some clinical results. *Journal of Psychosomatic Research, 1,* 233–246. https://doi.org/10.1016/0022-3999(56)90001-0

Chordekar, S., Perez, R., Adelman, C., Sohmer, H., & Kishon-Rabin, L. (2018). Does hearing in response to soft-tissue stimulation involve skull vibrations? A within-subject comparison between skull vibration magnitudes and hearing thresholds. *Hearing Research*, *364*, 59–67. http://doi.org/10.1016/j.heares.2018.03.030

Ciocca, V. (2008). The auditory organization of complex sounds. *Frontiers in Bioscience*, *13*(1), 148–169. http://doi.org/10.2741/2666

Civier, O., Bullock, D., Max, L., & Guenther, F. H. (2013). Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. *Brain and Language*, *126*(3), 263–278. http://doi.org/10.1016/j.bandl.2013.05.016

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of Fluency Disorders*, *35*(3), 246–279. http://doi.org/10.1016/j.jfludis.2010.05.002

Choo, A.L., Smith, S.A. & Li, H. (2020). Associations between stuttering, comorbid conditions and executive function in children: a population-based study. BMC Psychol 31;8(1):113. doi: 10.1186/s40359-020-00481-7.

Cody, A. R., Wilson, W., & Leah, J. (1996). Acoustically activated c-fos expression in auditory nuclei of the anaesthetised guinea pig. *Brain Research*, *728*(1), 72–78.

Colebatch, J. G., Govender, S., & Rosengren, S. M. (2013). Two distinct patterns of VEMP changes with age. Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology, 124(10), 2066–2068. http://doi.org/10.1016/j.clinph.2013.04.337

Constantino, C. D., Eichorn, N., Buder, E. H., Beck, J. G., & Manning, W. H. (2020). The Speaker's Experience of Stuttering: Measuring Spontaneity. *Journal of Speech, Language, and Hearing Research*, *63*(4), 983–1001. http://doi.org/10.1044/2019_JSLHR-19-00068

Cooke, M., & Ellis, D. P. W. (2001). The auditory organization of speech and other sources in listeners and computational models. *Speech Communication*, *35*(3-4), 141–177. http://doi.org/10.1016/S0167-6393(00)00078-9

Craig, A., Hancock, K., Tran, Y., Craig, M., & Peters, K. (2002). Epidemiology of stuttering in the community across the entire life span. Journal of Speech, Language, and Hearing Research, 45, 1097–1105.

Crivellaro Goncalves, I., Furquim de Andrade, C. R., & Gentile Matas, C. (2015). Auditory Processing of Speech and Non-Speech Stimuli in Children who Stutter: Electrophysiological Evidences. *Brain Disorders & Therapy*, *04*(05), 1–5. http://doi.org/10.4172/2168-975X.1000199

Cullinan. W.L. and Springer, M.T. (1980) Voice initiation and termination times in stuttering and nonstuttering children., Journal of Speech and Hearing Research, 2, 344–60.

Curthoys, I. S. (2017). The new vestibular stimuli: sound and vibration-anatomical, physiological and clinical evidence. *Experimental Brain Research*, 235(4), 957–972. http://doi.org/10.1007/s00221-017-4874-y

Curthoys, I. S., & Grant, J. W. (2015). How does high-frequency sound or vibration activate vestibular receptors? *Experimental Brain Research*, *233*(3), 691–699. http://doi.org/10.1007/s00221-014- 4192-6

Curthoys, I. S., Grant, J. W., Burgess, A. M., Pastras, C. J., Brown, D. J., & Manzari, L. (2018). Otolithic Receptor Mechanisms for Vestibular- Evoked Myogenic Potentials: A Review. *Frontiers in Neurology*, *9*, 366. http://doi.org/10.3389/fneur.2018.00366

Curthoys, I. S., Grant, J. W., Pastras, C. J., Brown, D. J., Burgess, A. M., Brichta, A. M., & Lim, R. (2019). A review of mechanical and synaptic processes in otolith transduction of sound and vibration for clinical VEMP testing. *Journal of Neurophysiology*. http://doi.org/10.1152/jn.00031.2019

Curthoys, I. S., Kim, J., McPhedran, S. K., & Camp, A. J. (2006). Bone conducted vibration selectively activates irregular primary otolithic vestibular neurons in the guinea pig. *Experimental Brain Research*, *175*(2), 256–267. http://doi.org/10.1007/s00221-006-0544-1

DeLong, M.R. Primate models of movement disorders of basal ganglia origin. (1990).Trends Neurosci. 13, 281–285.

Denhardt (1890). Das Stottern: eine Psychose. Leipsig.

Drayna D, Kilshaw J, Kelly J. (1999). The sex ratio in familial persistent stuttering. Am. J. Hum. Genet.65:1473–75

Dunovan, K., Lynch, B., Molesworth, T., & Verstynen, T. (2015). Competing basal ganglia pathways determine the difference between stopping and deciding not to go. *eLife*, *4* (September 2015), e08723. http://doi.org/10.7554/eLife.08723

Eisenson, J and Horowitz, E (1945) "The influence of propositionality on stuttering" *Journal of Speech Disorders*, 10, 193-197.

Ekdale E. G. (2013). Comparative Anatomy of the Bony Labyrinth (Inner Ear) of Placental Mammals. *PloS one*, *8*(6), e66624. https://doi.org/10.1371/journal.pone.0066624

Eliades, S. J., & Wang, X. (2017). Contributions of sensory tuning to auditory-vocal interactions in marmoset auditory cortex. *Hearing Research*, *348*, 98–111. http://doi.org/10.1016/j.heares.2017.03.001

Eliades, S. J., & Tsunada, J. (2018). Auditory cortical activity drives feedback-dependent vocal control in marmosets. *Nature Communications*, *9*(1), 2540–13. http://doi.org/10.1038/s41467-018-04961-8

Ellis, J. (2020). The Clearing: Music, dysfluency, blackness and time. *Journal of Interdisciplinary Voice Studies*, Volume 5, number 2, 215–233.

Elsherif M.M., Wheeldon L.R., Frisson S. (2021). Do dyslexia and stuttering share a processing deficit? J Fluency Disord. 67:105827. doi: 10.1016/j.jfludis.2020.105827.

Etchell, A. C., Civier, O., Ballard, K. J., & Sowman, P. F. (2017). A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016. *Journal of Fluency Disorders*. http://doi.org/10.1016/j.jfludis.2017.03.007

1332 Evans J. St. (2007). "Dual-processing accounts of reasoning, judgment, and social cognition". Annual
1333 Review of Psychology. 59: 268–269. doi:10.1146/annurev.psych.59.103006.093629. PMID 18154502.
1334 Fairbanks G. (1954) Systematic Research In Experimental Phonetics: 1. A Theory Of The Speech
1335 Mechanism As A Servosystem. Journal of Speech and Hearing Disorders; 19(2):133–139. https://doi.
1336 org/10.1044/jshd.1902.133 PMID: 13212816
1337 Fenoy, A. J., Severson, M. A., Volkov, I. O., Brugge, J. F., & Howard, M. A. (2006). Hearing suppression
1338 induced by electrical stimulation of human auditory cortex. *Brain Research*, *1118*(1), 75–83.
1339 http://doi.org/10.1016/j.brainres.2006.08.013
1340 Fernyhough, C., & Russell, J. (1997). Distinguishing Ones Own Voice from Those of Others A Function for
1341 Private Speech. *International Journal of Behavioral …*.
1342 Florentine M., Popper A., Fay R. (eds) Loudness. Springer Handbook of Auditory Research, vol 37.
1343 Springer, New York, NY. https://doi.org/10.1007/978-1-4419-6712-1_5
1344 Fodor, Jerry A., (1987) "Meaning and the world order" from Fodor, Jerry, A., *Psychosemantics: the*
1345 *problem of meaning in the philosophy of mind* pp.97-127, Cambridge, Massachusetts: The MIT Press
1346 Ford, J. M. & Mathalon, D. H. (2004). Electrophysiological evidence of corollary discharge dysfunction in
1347 schizophrenia during talking and thinking. *Journal of Psychiatric Research*, *38*(1), 37–46.
1348 http://doi.org/10.1016/s0022-3956(03)00095-5
1349 Foundas, A. L., Mock, J. R., Corey, D. M., Golob, E. J., & Conture, E. G. (2013). The SpeechEasy device in
1350 stuttering and nonstuttering adults: fluency effects while speaking and reading. *Brain and Language*,
1351 *126*(2), 141–150. http://doi.org/10.1016/j.bandl.2013.04.004
1352 Frigerio-Domingues, C., & Drayna, D. (2017). Genetic contributions to stuttering: the current evidence.
1353 *Molecular Genetics & Genomic Medicine*, *5*(2), 95–102. http://doi.org/10.1002/mgg3.276
1354 Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams,
1355 S. C. R., Amaro, E., Andrew, C. M., . . . McGuire, P. K. (2006). An fMRI study of verbal self-monitoring:
1356 Neural correlates of auditory verbal feedback. Cerebral Cortex, 16(7), 969–977.
1357 https://doi.org/10.1093/cercor/bhj039
1358 Garcia-Barrera, M. A., & Davidow, J. H. (2015). Anticipation in stuttering: A theoretical model of the
1359 nature of stutter prediction. *Journal of Fluency Disorders*, *44*, 1–15.
1360 http://doi.org/10.1016/j.jfludis.2015.03.002
1361 Gattie, M.C.D., Howell, P., Kluk, K., & Lieven, E. (2019, June 30). Vestibular Evoked Myogenic Potentials in
1362 Persistent Developmental Stuttering. Retrieved from osf.io/9mukg
1363 Gattie, M.C.D., Lieven, E. & Kluk, K. (submitted, 2020) Weak vestibular response in persistent
1364 developmental stuttering
1365 Galantucci B., Fowler C.A., Turvey M.T. (2006). The motor theory of speech perception reviewed.
1366 Psychonomic Bulletin & Review;13(3):361–377.
1367 Gallagher, S. & Zahavi, D. (2021). "Phenomenological Approaches to Self-Consciousness", *The Stanford*
1368 *Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL =
1369 <https://plato.stanford.edu/archives/spr2021/entries/self-consciousness-phenomenological/>.
1370 Goldberg, J.M. (ed) (2012), The vestibular system: a sixth sense. (2012). Oxford University Press Inc
1371 Golding, N. L., & Oertel, D. (2012). Synaptic integration in dendrites: exceptional need for speed. *The*
1372 *Journal of Physiology*, *590*(22), 5563–5569. http://doi.org/10.1113/jphysiol.2012.229328
1373
1374 Gould, E. & Sheehan, J (1967). "Effect of silence on stuttering" Journal of Abnormal Psychology, 72, 5,
1375 441–445.
1376 Govender, S., Todd, N. P. M., & Colebatch, J. G. (2020). Mapping the vestibular cerebellar evoked
1377 potential (VsCEP) following air- and bone-conducted vestibular stimulation. *Experimental Brain*
1378 *Research*, *238*(3), 601–620. http://doi.org/10.1007/s00221-020-05733-x
1379 Grant, W., & Curthoys, I. (2017). Otoliths - Accelerometer and seismometer; Implications in Vestibular
1380 Evoked Myogenic Potential (VEMP). *Hearing Research*, *353*, 26–35.
1381 http://doi.org/10.1016/j.heares.2017.07.012

Greenlee, J. D. W., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., et al. (2011). Human Auditory Cortical Activation during Self-Vocalization. *Plos One*, *6*(3), e14744. http://doi.org/10.1371/journal.pone.0014744

Gregory, H.H. (ed.) (1979). Controversies about Stuttering Therapy. Baltimore: University Park Press.

Grice, P. (1957). 'Meaning', *The Philosophical Review*, 66: 377–88.

Grice, P. (1989). *Studies in the Way of Words* Cambridge MA: Harvard University Press.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*, 280–301.

Haas, H. (1951). "Uber den Einfluss eines Einfachechos auf die Horsamkeit von Sprache," Acustica, 1, 49–58.

Haber S. N. (2014). The place of dopamine in the cortico-basal ganglia circuit. *Neuroscience*, *282*, 248–257. https://doi.org/10.1016/j.neuroscience.2014.10.008

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. Princeton University Bulletin. pp. 49–52.

Hadamard, J. (1923). Lectures on Cauchy's Problem in Linear Partial Differential Equations. Yale University Press, New Haven CT.

Han, T.-U., Root, J., Reyes, L. D., Huchinson, E. B., Hoffmann, Du, J., Lee, W. S., et al. (2019). Human GNPTAB stuttering mutations engineered into mice cause vocalization deficits and astrocyte pathology in the corpus callosum. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(35), 17515–17524. http://doi.org/10.1073/pnas.1901480116

Hannley, M., & Dorman, M. F. (1982). Some observations on auditory function and stuttering. *Journal of Fluency Disorders*, *7*(1), 93–108. http://doi.org/10.1016/S0094-730X(82)80003-X

Helmholtz, H. (1886). *Handbuch der Physiologischen Optik.* Leipzig: Voss.

Herder, C., Howard, C., Nye, C., Vanryckeghem, M., (2006). Effectiveness of behavioural stuttering treatment: a systematic review and meta-analysis. *Contemporary Issues in Communication Science and Disorders*, *33*, 61–73.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. http://doi.org/doi:10.1038/nrn2113

Hickok, G. (2017). A cortical circuit for voluntary laryngeal control: Implications for the evolution language. *Psychon Bull Rev* **24,** 56–63. https://doi.org/10.3758/s13423-016-1100-z

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407–422. http://doi.org/10.1016/j.neuron.2011.01.019

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., . . . Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. NeuroReport, 8(9–10), 2379–2382.

von Holst, E., and Mittelstädt, H. (1950). Das reafferenzprinzip. Naturwissenschaften 37, 464–476. doi: 10.1007/bf00622503

Hoshi E., Tremblay L., Féger J., Carras P.L., Strick P.L. (2005). The cerebellum communicates with the basal ganglia. Nat Neurosci;8(11):1491–1493. doi: 10.1038/nn1544.

Hotehama, T., & Nakagawa, S. (2012). Propagation Velocity of Bone-Conducted Ultrasound in the Human Head. *Japanese Journal of Applied Physics*, *51*(7S), 07GF21. http://doi.org/10.1143/JJAP.51.07GF21

Houde, J. F., & Nagarajan, S. S. (2016). Speech motor control from a modern control theory perspective. *Neurobiology of Language*. http://doi.org/10.1016/B978-0-12-407794-2.00019-5

Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, *33*, 174–181. http://doi.org/10.1016/j.conb.2015.04.006

Howell, P., El-Yaniv, N., & Powell, D. J. (1987). Factors affecting fluency in stutterers when speaking under altered auditory feedback. In H. Peters & W. Hulstijn (Eds.), (pp. 361–369). Speech Motor Dynamics in Stuttering.

Howell, P. (2004). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Research : an on-Line Journal Published by the British Stammering Association*, *1*(1), 31–46.

Howell, P., Marchbanks, R. J., & El-Yann, N. (1986). Middle ear muscle activity during vocalization in normal speakers and stutterers. *Acta Oto-Laryngologica*, *102*(5-6), 396–402. http://doi.org/10.3109/00016488609119423

Howell, P., Powell, D. J., & Khan, I. (1983). Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks. *Journal of Experimental Psychology*, 1–13.

Huettig, F., & Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes*, *25*(3), 347–374. http://doi.org/10.1080/01690960903046926

Irvine D.R.F. (1992) Physiology of the Auditory Brainstem. In: Popper A.N., Fay R.R. (eds) The Mammalian Auditory Pathway: Neurophysiology. Springer Handbook of Auditory Research, vol 2. Springer, New York, NY

Isaacs, D. (2020). 'I Don't Have Time For This': Stuttering and the Politics of University Time. *Scandinavian Journal of Disability Research*, 22(1), 58–67. DOI: http://doi.org/10.16993/sjdr.601

Iverach, L., Rapee, R. M., Wong, Q. J. J., & Lowe, R. (2017). Maintenance of Social Anxiety in Stuttering: A Cognitive-Behavioral Model. *American Journal of Speech-Language Pathology*, *26*(2), 540–17. http://doi.org/10.1044/2016_AJSLP-16-0033

Jackson, E., Yaruss, J. S., Quesal, R. W., Terranova, V., & Whalen, D. (2015). Responses of adults who stutter to the anticipation of stuttering. Journal of Fluency Disorders, 45, 38–51.

Jones, D. K., Knösche, T. R., & Turner, R. (2013). White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. NeuroImage, 73, 239–254. doi:10.1016/j. neuroimage.2012.06.081

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, *26*(2), 235–258.

Johnson, M., Baxter, S., Blank, L., Cantrell, A., Brumfitt, S., Enderby, P., & Goyder, E. (2015). The state of the art in non-pharmacological interventions for developmental stuttering. Part 2: qualitative evidence synthesis of views and experiences. *International Journal of Language & Communication Disorders*, *51*(1), 3–17. http://doi.org/10.1111/1460-6984.12182

Johnson, W. (1961). Stuttering and what you can do about it. Danville, IL: The Interstate Printers & Publishers, Inc

Jones, L. A., Hills, P. J., Dick, K. M., Jones, S. P., & Bright, P. (2016). Cognitive mechanisms associated with auditory sensory gating. Brain and cognition, 102, 33–45. https://doi.org/10.1016/j.bandc.2015.12.005

Kaga, K. (2016). Auditory nerve disease and auditory neuropathy spectrum disorders. *Auris, Nasus, Larynx*, *43*(1), 10–20. http://doi.org/10.1016/j.anl.2015.06.008

Kahneman, D. (2011). Thinking, fast and slow (1st ed.). New York: Farrar, Straus and Giroux. ISBN 9780374275631.

Kalinowski, J., Armson, J., Roland-Mieszkowski, M., Stuart, A. and Gracco, V. L. (1993). Effects of alterations in auditory feedback and speech rate on stuttering frequency. Language and Speech, 36, 1–16.

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: an fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective Neuroscience*, *3*(3), 218–223. http://doi.org/10.1093/scan/nsn014

Karniol, R. (1995) Stuttering, language, and cognition: A review and a model of stuttering as suprasegmental sentence plan alignment (SPA). *Psychological Bulletin*, 117, 104-124.

Kato T, Shiraishi K, Eura Y, Shibata K, Sakata T, Morizono T, Soda T. (1998) A 'neural' response with 3ms latency evoked by loud sound in profoundly deaf patients. Audiol Neurootol;3:253–64.

Kern (1932) Der Einflusz des Hiirens auf das Stottern. Arch. Psychiat. 97, 429449.

Kobayashi, H., & Hayasaka, K. (2003). Manual Diadochokinetic Movements in Children with Stuttering and Phonologicial Disorders. *Japanese Journal of Special Education*, *40*, 649–662.

Kolarik, A. J., Cirstea, S., Pardhan, S., & Moore, B. C. J. (2014). A summary of research investigating echolocation abilities of blind and sighted humans. *Hearing Research*, *310*, 60–68. http://doi.org/10.1016/j.heares.2014.01.010

Kraft, S. J., & Yairi, E. (2012). Genetic Bases of Stuttering: The State of the Art, 2011. *Folia Phoniatrica Et Logopaedica*, *64*(1), 33–46. http://doi.org/10.1159/000331073

1487 Kramer, M. B., Green, D., & Guitar, B. (1987). A comparison of stutterers and nonstutterers on masking
1488 level differences and synthetic sentence identification tasks. *Journal of Communication Disorders*,
1489 *20*(5), 379–390.
1490 Kröger, B. J., Crawford, E., Bekolay, T., & Eliasmith, C. (2016). Modeling Interactions between Speech
1491 Production and Perception: Speech Error Detection at Semantic and Phonological Levels and the Inner
1492 Speech Loop. *Frontiers in Computational Neuroscience*, *10*, 51.
1493 http://doi.org/10.3389/fncom.2016.00051
1494 Lakatos, I.M. (1970) "Falsification and the Methodology of Scientific Research Programmes", in Lakatos &
1495 Musgrave (eds) *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press.
1496 Langová, J & Sváb, L (1973) "Reduction of stuttering under experimental social isolation: The role of
1497 adaptation effect" *Folia Phoniatrica*, 25, 17–22.
1498 Langová, J., Morávek, M., Široký & Sváb, L. (1975) "Effect of speaking on evoked vestibular nystagmus in
1499 stutterers" *Folia Phoniatrica*, 27(4), 287–291.
1500 Laukli, E., & Burkard, R. (2015). Calibration/Standardization of Short-Duration Stimuli. *Seminars in*
1501 *Hearing*, *36*(1), 3–10. http://doi.org/10.1055/s-0034-1396923
1502 Levelt, W.J. (1989) *Speaking: From intention to articulation* Cambridge, MA: MIT Press.
1503 Levelt, W.J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The*
1504 *Behavioral and Brain Sciences*, *22*(1), 1–38– discussion 38–75.
1505 http://doi.org/10.1017/s0140525x99001776
1506 Liebetrau, R. M., & Daly, D. A. (1981). Auditory processing and perceptual abilities of "organic" and
1507 'functional' stutterers. Journal of Fluency Disorders.
1508 Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognitive and
1509 Behavioral Neurology, 21(1), 1–36.
1510 Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Auditory feedback of one's own
1511 voice is used for high-level semantic monitoring: the 'self-comprehension' hypothesis. *Frontiers in*
1512 *Human Neuroscience*, *8*, 166. http://doi.org/10.3389/fnhum.2014.00166
1513 Liotti, M., Ingham, J. C., Takai, O., Paskos, D. K., Perez, R., & Ingham, R. J. (2010). Spatiotemporal
1514 dynamics of speech sound perception in chronic developmental stuttering. *Brain and Language*,
1515 *115*(2), 141–147. http://doi.org/10.1016/j.bandl.2010.07.007
1516 Lipton, P. (2004). *Inference to the Best Explanation*, (2nd ed.), London: Routledge.
1517 Ludlow, C. L., & Loucks, T. (2003). Stuttering: a dynamic motor control disorder. *Journal of Fluency*
1518 *Disorders*, *28*(4), 273–95– quiz 295. http://doi.org/10.1016/j.jfludis.2003.07.001
1519 Lupyan, G., & Swingley, D. (2012). Self-directed speech affects visual search performance. *Quarterly*
1520 *Journal of Experimental Psychology (2006)*, *65*(6), 1068–1085.
1521 http://doi.org/10.1080/17470218.2011.647039
1522 MacCulloch, M. J., & Eaton, R. (1971). A note on reduced auditory pain threshold in 44 stuttering
1523 children. *The British Journal of Disorders of Communication*, *6*(2), 148–153.
1524 http://doi.org/10.3109/13682827109011541
1525 MacKay D. G. (1969). Effects of ambiguity on stuttering: towards a theory of speech production at the
1526 semantic level. *Kybernetik*, *5*(5), 195–208. https://doi.org/10.1007/BF00288943
1527 Maffei, C., Sarubbo, S., & Jovicich, J. (2019). A Missing Connection: A Review of the Macrostructural
1528 Anatomy and Tractography of the Acoustic Radiation. *Frontiers in Neuroanatomy*, *13*.
1529 http://doi.org/10.3389/fnana.2019.00027
1530 Manley, G. A., Gummer, A., Popper, A.N. & Fay, R.R. (eds) (2017). Understanding the Cochlea. Springer
1531 Science & Business Media.
1532 Maraist, J. A. and Hutton, C. (1957) Effects of auditory masking upon the speech of stutterers. J. Speech
1533 Dis. 22, 385-389.
1534 Mason S, Garnham C & Hudson B. (1996) Electric response audiometry in young children before cochlear
1535 implantation: a short latency component. *Ear Hear*. 17:537–43.
1536 Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in
1537 linear mixed models. *Journal of Memory and Language*, *94*, 305–315.
1538 http://doi.org/10.1016/j.jml.2017.01.001
1539 Maurer, D., & Landis, T. (1990). Role of Bone Conduction in the Self-Perception of Speech. *Folia*
1540 *Phoniatrica Et Logopaedica*, *42*(5), 226–229. http://doi.org/10.1159/000266070

Maguire, G. A., Nguyen, D. L., Simonson, K. C., & Kurz, T. L. (2020). The Pharmacologic Treatment of Stuttering and Its Neuropharmacologic Basis. *Frontiers in Neuroscience*, *14*, 158. http://doi.org/10.3389/fnins.2020.00158

Månsson H. (2000) Identification of stuttering in pre- school children: a multifactorial, longitudinal study in development. J Fluency Disord;25(3):241

Martin, D. (2016). Stuttering from the Anus. *Canadian Journal of Disability Studies*, *5*(3), 114–134. https://doi.org/10.15353/cjds.v5i3.299

Mariën, P., & Manto, M. (2015).The linguistic cerebellum. Academic Press.

Matthews, N., Todd, J., Mannion, D. J., Finnigan, S., Catts, S., & Michie, P. T. (2013). Impaired processing of binaural temporal cues to auditory scene analysis in schizophrenia. *Schizophrenia Research*, *146*(1-3), 344–348. http://doi.org/10.1016/j.schres.2013.02.013

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.

Max, L. (2004). Stuttering and internal models for sensorimotor control: A theoretical perspective to generate testable hypotheses. In B. Maassen, R. Kent, H. F. M. Peters, P. van Lieshout, & W. Hulstijn (Eds.), Speech motor control in normal and disordered speech (pp. 357–387). Oxford, UK: Oxford University Press.

Max, L., & Baldwin, C. J. (2010). The role of motor learning in stuttering adaptation: repeated versus novel utterances in a practice-retention paradigm. *Journal of Fluency Disorders*, *35*(1), 33–43. http://doi.org/10.1016/j.jfludis.2009.12.003

McLachlan, N. M., Phillips, D. S., Rossell, S. L., & Wilson, S. J. (2013). Auditory processing and hallucinations in schizophrenia. *Schizophrenia Research*, *150*(2-3), 380–385. http://doi.org/10.1016/j.schres.2013.08.039

McKnight, R. C., & Cullinan, W. L. (1987). Subgroups of stuttering children: Speech and voice reaction times, segmental durations, and naming latencies. Journal of Fluency Disorders, 12(3), 217–233. doi:10.1016/0094-730X(87)90028-3

McNerney, K. M., & Burkard, R. F. (2011). The vestibular evoked myogenic potential (VEMP): air- versus bone-conducted stimuli. *Ear and Hearing*, *32*(6), e6–e15. http://doi.org/10.1097/AUD.0b013e3182280299

Meekings, S., Boebinger, D., Evans, S., Lima, C. F., Chen, S., Ostarek, M., & Scott, S. K. (2015). Do We Know What We're Saying? The Roles of Attention and Sensory Information During Speech Production. *Psychological Science*, *26*(12), 1975–1977. http://doi.org/10.1177/0956797614563766

Meibauer, J. & Steinbach, M. (2011). Experimental Pragmatics/semantics. Amsterdam: John Benjamins Pub. Co. ISBN 9789027255587

Metzger, F. L., Auer, T., Helms, G., Paulus, W., Frahm, J., Sommer, M., & Neef, N. E. (2017). Shifted dynamic interactions between subcortical nuclei and inferior frontal gyri during response preparation in persistent developmental stuttering. *Brain Structure & Function*, *223*(1), 165–182. http://doi.org/10.1007/s00429-017-1476-1

Milardi, D., Quartarone, A., Bramanti, A., Anastasi, G., Bertino, S., Basile, G. A., Buonasera, P., Pilone, G., Celeste, G., Rizzo, G., Bruschetta, D., & Cacciola, A. (2019). The Cortico-Basal Ganglia-Cerebellar Network: Past, Present and Future Perspectives. *Frontiers in systems neuroscience*, *13*, 61. https://doi.org/10.3389/fnsys.2019.00061

Miller N.E. (1944) Experimental Studies of Conflict. In Hunt, J. McV (ed) “Personality and the Behavior Disorders”. New York: Ronald Press.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. Progress in Neurobiology, 50, 381–425.

Mink, J. W. (2018). Basal ganglia mechanisms in action selection, plasticity, and dystonia. *European Journal of Paediatric Neurology : EJPN : Official Journal of the European Paediatric Neurology Society*, *22*(2), 225–229. http://doi.org/10.1016/j.ejpn.2018.01.005

Mozaffarilegha, M., Namazi, H., Tahaei, A. A., & Jafari, S. (2019). Complexity-Based Analysis of the Difference Between Normal Subjects and Subjects with Stuttering in Speech Evoked Auditory

Brainstem Response. *Journal of Medical and Biological Engineering*, *39*(4), 490–497. http://doi.org/10.1007/s40846-018-0430-x

Murofushi, T., Iwasaki, S., Takai, Y., & Takegoshi, H. (2005). Sound-evoked neurogenic responses with short latency of vestibular origin. *Clinical Neurophysiology*, *116*(2), 401–405. http://doi.org/10.1016/j.clinph.2004.09.005

Musicant A. D. & Butler R. A. (1984). The influence of pinnae-based spectral cues on sound localization. The Journal of the Acoustical Society of America, 75(4): 1195-1200

Mysak, E.D. (1969). Speech Pathology and Feedback Theory. Charles C Thomas, Illinois.

Nakagawa, S., Johnson, P.C.D. & Schielzeth, H. (2017). The Coefficient of Determination R2 and Intra-Class Correlation Coefficient from Generalized Linear Mixed-Effects Models Revisited and Expanded. *Journal of the Royal Society Interface* 14 (134): 20170213. https://doi.org/10.1098/rsif.2017.0213.

Neef, N. E., Anwander, A., & Friederici, A. D. (2015). The Neurobiological Grounding of Persistent Stuttering: from Structure to Function. *Current Neurology and Neuroscience Reports*, *15*(9), 63. http://doi.org/10.1007/s11910-015-0579-4

Neelly, J. M. (1961) A study of the speech behaviour of stutterers and nonstutterers under normal and delayed auditory feedback. / . *speech hear. Disord. monogr. Suppl*, No. 7.

Newlands, S. D., Vrabec, J. T., Purcell, I. M., Stewart, C. M., Zimmerman, B. E., & Perachio, A. A. (2003). Central projections of the saccular and utricular nerves in macaques. *The Journal of Comparative Neurology*, *466*(1), 31–47. http://doi.org/10.1002/cne.10876

Nong, D. X., Ura, M., Owa, T., & Noda, Y. (2000). An acoustically evoked short latency negative response in profound hearing loss patients. *Acta Oto-Laryngologica*, *120*(8), 960–966. http://doi.org/10.1080/00016480050218708

Nong, D. X., Ura, M., Kyuna, A., Owa, T., & Noda, Y. (2002). Saccular origin of acoustically evoked short latency negative response. *Otology & Neurotology*, *23*(6), 953–957. http://doi.org/10.1097/00129492-200211000-00024

Norris, D., McQueen, J. M., & Cutler. (2000). Merging information in speech recognition: Feedback is never necessary. *The Behavioral and Brain Sciences*, *23*(3), 299–325. http://doi.org/10.1017/S0140525X00003241

Noveck, I.A., and Sperber, D. (2004). Experimental Pragmatics. Basingstoke, Hampshire: Palgrave Macmillan. ISBN 9781403903501

Noveck, I.A. (2018). Experimental Pragmatics: The Making of a Cognitive Science. Cambridge: CUP. ISBN 9781107084902

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*(1), 1–33. http://doi.org/10.1016/j.cogpsych.2011.05.001

Ntourou, K. , Conture, E. G. , & Lipsey, M. W. (2011). Language abilities of children who stutter: A meta-analytical review. American Journal of Speech-Language Pathology, 20(3), 163–179. https://doi.org/10.1044/1058-0360(2011/09-0102)

Nittrouer, S., & Cheney, C. (1984). Operant techniques used in stuttering therapy: A review. *Journal of Fluency Disorders*, *9*(3), 169–190. http://doi.org/10.1016/0094-730X(84)90011-1

Oliver, D.L., Cant, N.B., Fay, R.R. & Popper, A.N. (eds) (2018) Vertebrate Sound Production and Acoustic Communication. Springer Handbook of Auditory Research, vol 65. Springer, New York, NY

Ofoe L.C., Anderson J.D., Ntourou K. (2018).Short-term memory, inhibition, and attention in developmental stuttering: a meta-analysis. J Speech Lang Hear Res;61(7):1626–1648. doi: 10.1044/2018_JSLHR-S-17-0372.

Ozdemir, R., Roelofs, A., & Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognitive and Behavioral Neurology*, *105*(2), 457–465. http://doi.org/10.1016/j.cognition.2006.10.006

Papathanasiou E, Zamba-Papanicolaou E, Pantziaris M, Kyriakides T, Papacostas S, Myrianthopoulou P, et al. (2003) Click evoked neurogenic vestibular potentials (NVESTEPs): a method of assessing the function of the vestibular system. *Electromyogr Clin Neurophysiol.* 43:399–408.

Papathanasiou ES, Theocharidou EK, Papacostas SS. (2006) Parallel auditory vestibular evoked neurogenic and myogenic potential results in a case of peripheral vestibular dysfunction, showing that the former originates from the vestibular system. *Electromyogr Clin Neurophysiol.* 46:105–11.

Papathanasiou ES, Zamba-Papanicolaou E, Pantziaris M, Kleopas K, Kyriakides T, Papacostas S, et al. (2004) Neurogenic vestibular evoked potentials using a tone pip auditory stimulus. *Electromyogr Clin Neurophysiol.* 44:167–73.

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLoS Computational Biology*, *15*(9), e1007321. http://doi.org/10.1371/journal.pcbi.1007321

Pelzer, E. A., Melzer, C., Timmermann, L., von Cramon, D. Y., & Tittgemeyer, M. (2017). Basal ganglia and cerebellar interconnectivity within the human thalamus. *Brain structure & function*, *222*(1), 381–392. https://doi.org/10.1007/s00429-016-1223-z

Pidoux, L., Le Blanc, P., Levenes, C., & Leblois, A. (2018). A subcortical circuit linking the cerebellum to the basal ganglia engaged in vocal learning. *eLife*, *7*, e32167. https://doi.org/10.7554/eLife.32167

Pierre, J. S. (2015). Distending Straight-Masculine Time: A Phenomenology of the Disabled Speaking Body. *Hypatia*, *30*(1), 49–65. http://doi.org/10.1111/hypa.12128

Poggio, T., & Koch, C. (1985). Ill-posed problems in early vision: From computational theory to analogue networks. *Proceedings of the Royal Society of London - Biological Sciences*, *226*(1244), 303–323. http://doi.org/10.1098/rspb.1985.0097

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech Perception at the Interface of Neurobiology and Linguistics. *Philosophical Transactions: Biological Sciences*, *363*(1493), 1071–1086.

Pörschmann, C. (2000). Influences of bone conduction and air conduction on the sound of one's own voice. *Acta Acustica United with Acustica*, *86*(6), 1038–1045.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognitive and Behavioral Neurology*, *77*(2), 97–132. http://doi.org/10.1016/S0010-0277(00)00090-1

Rayleigh (J.W. Strutt, 3rd Baron Rayleigh) (1907) On our perception of sound direction. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 13(74) XII: 214-232.

Reinfeldt, S., Östli, P., Håkansson, B., & Stenfelt, S. (2010). Hearing one's own voice during phoneme vocalization—Transmission by air and bone conduction. *The Journal of the Acoustical Society of America*, *128*(2), 751–762. http://doi.org/10.1121/1.3458855

Reynolds, J. N., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. Nature, 413, 67–70.

Rescorla, Michael (2019) "Convention", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2019/entries/convention/.

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.

Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition*, *68*(2), 204–214. http://doi.org/10.1016/j.bandc.2008.04.007

Rosenfield D.B., Jerger J. (1984). Stuttering and auditory function. In: Curlee RF, Perkins WH, editors. Nature and treatment of stuttering: New directions. San Diego: College-Hill Press. pp. 73–87.

Rosengren, S. M., & Colebatch, J. G. (2018). The Contributions of Vestibular Evoked Myogenic Potentials and Acoustic Vestibular Stimulation to Our Understanding of the Vestibular System. *Frontiers in Neurology*, *9*, 481. http://doi.org/10.3389/fneur.2018.00481

Rousey, C. L., Goetzinger, C. P., & Dirks, D. (1959). Sound localization ability of normal, stuttering, neurotic, and hemiplegic subjects. *A.M.a. Archives of General Psychiatry*, *1*, 640–645.

Ryugo, D.K., Fay, R.R. & Popper, A.N. (eds) (2011) Auditory and Vestibular Efferents. Springer Handbook of Auditory Research, vol 38. Springer, New York, NY

Sato, Y., Mori, K., Koizumi, T., Minagawa-Kawai, Y., Tanaka, A., Ozawa, E., et al. (2011). Functional lateralization of speech processing in adults and children who stutter. *Frontiers in Psychology*, *2*, 70. http://doi.org/10.3389/fpsyg.2011.00070

Schlindwein, P., Mueller, M., Bauermann, T., Brandt, T., Stoeter, P., & Dieterich, M. (2008). Cortical representation of saccular vestibular stimulation: VEMPs in fMRI. *NeuroImage*, *39*(1), 19–31. http://doi.org/10.1016/j.neuroimage.2007.08.016

Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. Nat Rev Neurosci. 2016 March ; 17(3): 183–195. doi:10.1038/nrn.2015.26.

Seery C.H., Watkins R.V., Mangelsdorf S.C., Shigeto A. (2007). Subtyping stuttering II: contributions from language and temperament. J Fluency Disord.;32(3):197–217. 10.1016/j.jfludis.2007.07.001 .

Shahed, J., & Jankovic, J. (2001). Re-emergence of childhood stammering in Parkinson's disease: A hypothesis. *Movement Disorders, 16*, 114–118.

Shane, M. L. S. (1955) Effect on stuttering of alteration in auditory feedback. In Stuttering in children and adults. University of Minnesota Press, Minneapolis.

Sheehan, J.G. (1953) "Theory and treatment of stuttering as an approach-avoidance conflict", *Journal of Psychology*, 36, 27–49.

Sheehan, J.G., (1958) Conflict theory of stuttering. In Eisenson, J. (ed.), Stuttering: A Symposium. New York: Harper and Row.

Sheehan, J.G. (1970). *Stuttering: Research and therapy.* New York: Harper & Row.

Sheehan, J.G. (1975). Conflict theory and avoidance reduction therapy. In J. Eisenson (Ed.), *Stuttering: A second symposium* (pp. 97-198). New York: Harper & Row.

Sheehan, J., Hadley, R. & Gould, E., (1967). "Impact of authority on stuttering" Journal of Abnormal Psychology, 72, 290–293.

Shuster, L. I., & Durrant, J. D. (2003). Toward a better understanding of the perception of self-produced speech. *Journal of Communication Disorders*.

Singer, C. M., Hessling, A., Kelly, E. M., Singer, L., & Jones, R. M. (2020). Clinical Characteristics Associated With Stuttering Persistence: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *63*(9), 2995–3018. http://doi.org/10.1044/2020_JSLHR-20-00096

Skipper, J.I. & Lamettie, D.R. (2021) Speech Perception under the Tent: A Domain-general Predictive Role for the Cerebellum. *Journal of Cognitive Neuroscience* 1–18. https://doi.org/10.1162/jocn_a_01729

Smith A., & Kelly E. (1997). Stuttering: A dynamic, multifactorial model. In Curlee R. & Siegel G. (Eds.), Nature and treatment of stuttering: New directions (pp. 204–217). Boston, MA: Allyn & Bacon.

Smith, A., & Weber, C. (2017). How Stuttering Develops: The Multifactorial Dynamic Pathways Theory. *Journal of speech, language, and hearing research : JSLHR*, *60*(9), 2483–2505. https://doi.org/10.1044/2017_JSLHR-S-16-0343

Smith, J., "Self-Consciousness" (2020). *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/self-consciousness/>.

Smith, P. F. (2012). Interactions between the vestibular nucleus and the dorsal cochlear nucleus: implications for tinnitus. *Hearing Research*, *292*(1-2), 80–82. http://doi.org/10.1016/j.heares.2012.08.006

Sohmer, H., & Freeman, S. (2001). The latency of auditory nerve brainstem evoked responses to air-and bone-conducted stimuli. *Hearing Research*.

Sohmer, H. (2017). Soft Tissue Conduction: Review, Mechanisms, and Implications. *Trends in Hearing*, *21*, 2331216517734087. http://doi.org/10.1177/2331216517734087

Stager, S. V. (1990). Heterogeneity in stuttering. *Journal of Fluency Disorders*, *15*(1), 9–19. http://doi.org/10.1016/0094-730X(90)90028-Q

Steinbeis, N. (2016). The role of self–other distinction in understanding others' mental and emotional states: neurocognitive mechanisms in children and adults. Phil. Trans. R. Soc. B, 371:20150074 http://doi.org/10.1098/rstb.2015.0074

Starkweather, C. W. & Gottwald, S. R., (1990). The demands and capacities model: II. Clinical applications. *Journal of Fluency Disorders*, **15**, 143–157.

Starkweather, C. W. (2002). The epigenesis of stuttering. *Journal of Fluency Disorders*, *27*(4), 269–87–quiz 287–8. http://doi.org/10.1016/s0094-730x(02)00144-4

Stevens, S.S. (1972) Perceived level of noise by mark VII and decibels (E). J Acoust Soc Am 51:575–601.

Stromsta, C. (1972). Interaural phase disparity of stutterers and nonstutterers. *Journal of Speech and Hearing Research*, *15*(4), 771–780.

Stuart, A., Kalinowski, J., Rastatter, M., Saltuklaroglu, T., & Dayalu, V. (2004). Investigations of the impact of altered auditory feedback in-the-ear devices on the speech of people who stutter: initial fitting and 4-month follow-up. *International Journal of Language & Communication Disorders*, *39*(1), 93–113. http://doi.org/10.1080/13682820310001616976

Suchodoletz, W., & Wolfram, I. (1996). Brainstem auditory evoked potentials (BAEP) in children with developmental language disorders. *Klinische Pädiatrie*, *208*(05), 290–293. http://doi.org/10.1055/s-2008-1046485

Sugathan, N., & Maruthy, S. (2020). Predictive factors for persistence and recovery of stuttering in children: A systematic review. *International Journal of Speech-Language Pathology*, 1–13. http://doi.org/10.1080/17549507.2020.1812718

Suthers, R.A., Tecumseh Fitch, W., Fay, R.R. & Popper, A.N. (eds) (2016) Vertebrate Sound Production and Acoustic Communication. Springer Handbook of Auditory Research, vol 53. Springer, New York, NY

Tahaei, A. A., Ashayeri, H., Pourbakht, A., & Kamali, M. (2014). Speech evoked auditory brainstem response in stuttering. *Scientifica*, *2014*, 328646. http://doi.org/10.1155/2014/328646

Theys, C., van Wieringen, A., Tuyls, L., & De Nil, L. (2009). Acquired stuttering in a 16-year-old boy. *Journal of Neurolinguistics*, *22*(5), 427–435. http://doi.org/10.1016/j.jneuroling.2009.02.001

Tichenor, S., & Yaruss, J. S. (2018). A Phenomenological Analysis of the Experience of Stuttering. *American Journal of Speech-Language Pathology*, *27*(3S), 1180–1194. http://doi.org/10.1044/2018_AJSLP-ODC11-17-0192

Tichenor, S., & Yaruss, J. S. (2019). Stuttering as Defined by Adults Who Stutter. *Journal of Speech, Language and Hearing Research*, *vol 62*, 4356–4369. http://doi.org/10.1044/2018_AJSLP-ODC11-17-0192

Todd, N. P. M., Paillard, A. C., Kluk, K., Whittle, E., & Colebatch, J. G. (2014a). Vestibular receptors contribute to cortical auditory evoked potentials. *Hearing Research*, *309*, 63–74. http://doi.org/10.1016/j.heares.2013.11.008

Todd, N. P. M., Paillard, A. C., Kluk, K., Whittle, E., & Colebatch, J. G. (2014b). Source analysis of short and long latency vestibular-evoked potentials (VsEPs) produced by left vs. right ear air-conducted 500 Hz tone pips. *Hearing Research*, *312*(C), 91–102. http://doi.org/10.1016/j.heares.2014.03.006

Todd, N. P. M., Rosengren, S. M., & Colebatch, J. G. (2003). A short latency vestibular evoked potential (VsEP) produced by bone-conducted acoustic stimulation. *Journal of the Acoustical Society of America*, *114*(6 Pt 1), 3264–3272. http://doi.org/10.1121/1.1628249

Todd, N.P.M., Rosengren, S.M., Colebatch, J.G., 2008. A source analysis of short- latency vestibular evoked potentials produced by air- and bone-conducted sound. *Neurophysiol. Clin.* 119 (8), 1881e1894.

Todd, N. P., & Lee, C. S. (2015). The sensory-motor theory of rhythm and beat induction 20 years on: a new synthesis and future perspectives. *Frontiers in Human Neuroscience*, *9*. http://doi.org/10.3389/fnhum.2015.00444

Toyomura A., Fujii T., Yokosawa K., & Kuriki S. (2018). Speech disfluency-dependent amygdala activity in adults who stutter: Neuroimaging of interpersonal communication in MRI scanner environment. *Neuroscience*, 374, 144–154.

Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., & Kuriki, S. (2007). Neural correlates of auditory feedback control in human. Neuroscience, 146(2), 499–503. https://doi.org/10.1016/j.neuroscience.2007.02.023

Traer, J., & McDermott, J. H. (2016). Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences*, *113*(48), E7856–E7865. http://doi.org/10.1073/pnas.1612524113

Tye, M. (2018). "Qualia", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/qualia/>.

Van Borsel, J. (2014). Acquired stuttering: A note on terminology. *Journal of Neurolinguistics*, *27*(1), 41–49. http://doi.org/10.1016/j.jneuroling.2013.09.003

Van Riper, C. (1971), *The Nature of Stuttering*, 1st ed, Englewood Cliffs: Prentice-Hall.

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

1803 Van Riper, C. (1982), *The Nature of Stuttering*, 2nd ed, Englewood Cliffs: Prentice-Hall.
1804 Walden, T. A., Frankel, C. B., Buhr, A. P., Johnson, K. N., Conture, E. G., & Karrass, J. M. (2012). Dual
1805 diathesis-stressor model of emotional and linguistic contributions to developmental stuttering.
1806 *Journal of abnormal child psychology*, *40*(4), 633–644. https://doi.org/10.1007/s10802-011-9581-8
1807 Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). "The precedence effect in sound localization,"
1808 The American Journal of Psychology, 62, 315–336.
1809 Ward, D. (2010). Sudden onset stuttering in an adult: Neurogenic and psychogenic perspectives. *Journal*
1810 *of Neurolinguistics*.
1811 Warren, R. (1973). Quantification of Loudness. *The American Journal of Psychology, 86*(4), 807-825.
1812 doi:10.2307/1422087
1813 Waters, F. & Fernyhough, C. (2019). Auditory Hallucinations: Does a Continuum of Severity Entail
1814 Continuity in Mechanism? *Schizophrenia Bulletin* 45(4): 717-719.
1815 Weintraub, D. M., Ramage, E. M., Sutton, G., Ringdahl, E., Boren, A., Pasinski, A. C., et al. (2012).
1816 Auditory stream segregation impairments in schizophrenia. Psychophysiology, 49(10), 1372– 1383.
1817 http://doi.org/10.1111/j.1469-8986.2012.01457.x
1818 Welgampola, M. S., Rosengren, S. M., Halmagyi, G. M., & Colebatch, J. G. (2003). Vestibular activation by
1819 bone conducted sound. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(6), 771–778.
1820 http://doi.org/10.1136/jnnp.74.6.771
1821 Wingate, M. E. (1986a). Adaptation, consistency and beyond. I. Limitations and contradictions. Journal of
1822 Fluency Disorders, 11(1), 1–36 http://dx.doi.org/10.1016/0094-730X(86)90002-1.
1823 Wingate, M. E. (1986b). Adaptation, consistency and beyond. II. An integral account. Journal of Fluency
1824 Disorders, 11(1), 37–53 http://dx.doi.org/10.1016/0094-730X(86)90003-3.
1825 Wingate, M.E. (1969). Sound and pattern in “artificial” fluency. Journal of Speech and Hearing Research
1826 12: 677–86.
1827 Wyneken, C. (1868). Über das Stottern und dessen Heilung. Zeitschrift für rationelle Medizin, 31, 1–29.
1828 Wu, J. C., Maguire, G., Riley, G., Lee, A., Keator, D., & Tang, C. (1997). Increased dopamine activity
1829 associated with stuttering. *Clinical Neuroscience and Neuropsychology*, *8*, 767–770.
1830 Yadav, M., Cabrera, D., & Martens, W. L. (2012). A system for simulating room acoustical environments
1831 for one’s own voice. *Applied Acoustics*, *73*(4), 409–414.
1832 http://doi.org/10.1016/j.apacoust.2011.10.001
1833 Yates, A. J. (1963). Recent Empirical and Theoretical Approaches to the Experimental Manipulation of
1834 Speech in Normal Subjects and in Stammerers. *Behaviour Research and Therapy*, *1*(2-4), 95–119.
1835 http://doi.org/10.1016/0005-7967(63)90013-5
1836 Yairi E. (2007) Subtyping Stuttering I: A Review. Journal of Fluency Disorders. 32(3):165-96. doi:
1837 10.1016/j.jfludis.2007.04.001
1838 Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. Journal of Fluency
1839 Disorders, 38(2), 66–87. http://doi.org/10.1016/j.jfludis.2012.11.002
1840 Yang, Y., Jia, F., Siok, W. T., & Tan, L. H. (2017). The role of anxiety in stuttering: Evidence from functional
1841 connectivity. *Neuroscience*, *346*, 216–225. http://doi.org/10.1016/j.neuroscience.2016.11.033
1842 Yates, A. J. (1963). Recent Empirical and Theoretical Approaches to the Experimental Manipulation of
1843 Speech in Normal Subjects and in Stammerers. *Behaviour Research and Therapy*, *1*(2-4), 95–119.
1844 http://doi.org/10.1016/0005-7967(63)90013-5
1845 Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary
1846 vocal pitch regulation in experienced singers. Neuropsychologia, 48(2), 607–618. https://
1847 doi.org/10.1016/j.neuropsychologia.2009.10.025
1848 Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch
1849 regulation during singing. NeuroImage, 40(4), 1871–1887. https://doi.org/10.1016/
1850 j.neuroimage.2008.01.026
1851 Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2009). Functional overlap between regions involved in
1852 speech perception and in monitoring one’s own voice during speech production. Journal of Cognitive
1853 Neuroscience, 22(8), 1770–1781. https:// doi.org/10.1162/jocn.2009.21324
1854 Zwergal, A., Strupp, M., Brandt, T., & Büttner-Ennever, J. A. (2009). Parallel ascending vestibular
1855 pathways: anatomical localization and functional specialization. *Annals of the New York Academy of*
1856 *Sciences*, *1164*(1), 51–59. http://doi.org/10.1111/j.1749-6632.2009.04461.x

1857