Weak vestibular response in persistent developmental stuttering and implications for own voice identification

Max Gattie[1], Elena Lieven[2], Karolina Kluk[1]
University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[1] Manchester Centre for Audiology & Deafness (ManCAD)
[2] LuCiD (the ESRC International Centre for Language and Communicative Development)

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 coincidence detection between cochlear and vestibular streams identifies own voice in mammals (H1) and that the coincidence detection differs in people who stutter (H2). Vestibular-evoked myogenic potential (VEMP), an indirect measure of vestibular function, was assessed in 15 people who stutter, with controls paired on age and sex. VEMP amplitude was 8.5 dB smaller in people who stutter than paired controls (p = 0.035, 95% CI [-0.9, -16.1], t = -2.1, d = -0.8, conditional R^2 = 0.88), suggesting an approximate halving in how they perceptually experience the vestibular component of own voice. H1 and H2 are supported in this initial test of both hypotheses. Discussion covers own voice identification, persistent developmental stuttering, 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, Roelofs & Meyer, 1999; Tourville & Guenther, 2011).

bioRxiv version, 24 November 2020
corresponding author: max.gattie@manchester.ac.uk

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. 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 speech is created out of ambient sounds.

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 study. A hypothesis is formulated for own voice identification. The hypothesis is then developed using the example of persistent developmental stuttering (PDS), a condition characterised by involuntary prolongations and repetitions during speech. The hypotheses are:

H1: Own voice is identified through coincidence detection between vestibular and cochlear afferent streams
H2: Own voice is misidentified in people who stutter

H1 adverts to 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. In section 3.5 it will be proposed that robust a priori identification of own voice, according to H1, is not only essential to solving the cocktail party problem, but moreover underpins auditory stream segregation in general.

Background on the inner ear, and on persistent developmental stuttering, is appropriate before continuing.

Inner ear

The inner ear of mammals is an osseous labyrinth lined with sensory epithelium. Chambers include:

- 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).
- The vestibular system, comprising:
    - Semicircular canals, in which mechanoreceptors are deflected by changes in angular velocity,
    - Gravitoinertial otoliths, in which mechanoreceptors are deflected by changes in linear velocity, and arranged such that resting state deflection corresponds to head orientation (Goldberg, 2012).

However, the traditional discrimination implied above, of cochlear and vestibular chambers with respective hearing and equilibrial functions, is inaccurate (Tait, 1932). It is inaccurate because mammalian vestibular mechanoreceptors are not only deflected by changes in the body's momentum, but are also deflected by vibration and sound. Research since the 1990s shows the vestibular system in mammals responds to 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) (Welgampola, Rosengren, Halmagyi & Colebatch, 2003). When electrophysiological responses of vestibular origin in humans 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). The BC stimulus created by speaking will deflect vestibular mechanoreceptors in humans (Todd, Rosegren & Colebatch, 2008; Curthoys, 2017; Curthoys et al., 2019).

Persistent Developmental Stuttering

Persistent developmental stuttering (PDS) manifests as prolongations or repetitions of speech sounds, characteristically accompanied by increased tension in muscles of the face and articulatory system. The behavioural symptoms are reflected by differences in neurological activity and morphology by comparison with ordinarily fluent speakers.

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, Civier, Ballard & Sowman, 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 (FA) in the callosal body, and in dorsal white matter tracts connecting grey matter regions considered important for auditory and motor function (Neef, Anwander & Friederici, 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, Knösche, & Turner, 2013). FA 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, Baron & Fox, 2014; Belyk, Kraft & Brown, 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.

2. Results

2.1 Statistical Model

Peak to trough (p1-n1) amplitudes and latencies of cervical vestibular-evoked myogenic potentials (VEMPs) were compared between 15 PWS and controls paired on age and sex. VEMP measurement has several confounders (see section 4), shown for response amplitude in figure 1.

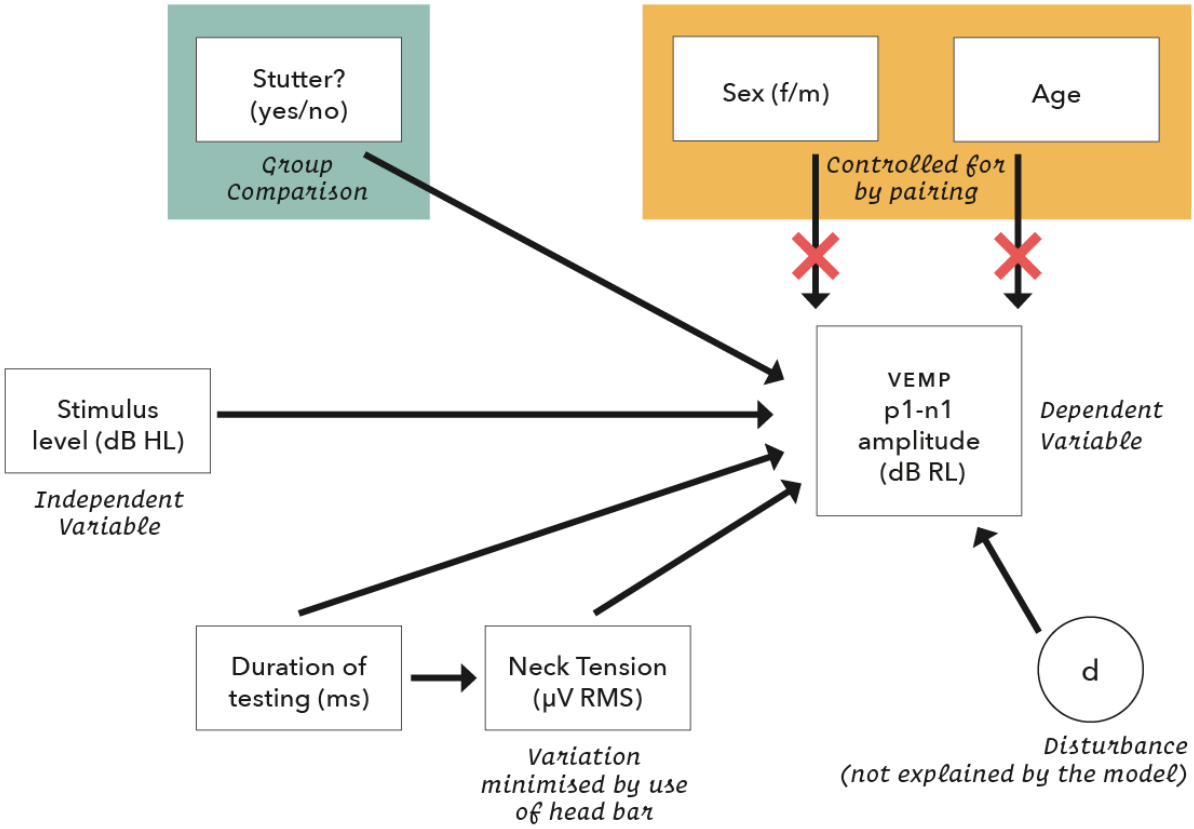

Figure 1: Initial statistical model for VEMP response amplitude. Neck tension was a root mean square of the electromyographic p1-n1 amplitude based on each presentation sequence of 300 repetitions. Age was calculated in days at the time of testing. The dB RL (response level) units used for vestibular response are a log transformation of the VEMP peak to trough amplitude, such that zero dB RL corresponds to vestibular threshold. Possible disturbances include neck size, pulse rate and crossed response. See section 4 for more analysis details.

Initial analysis with 48 control participants (Gattie, Lieven & Kluk, in preparation) eliminated neck tension and duration of testing as confounders for amplitude or latency measures. It also showed that VEMP latency is independent of stimulus level. This simplifies the latency model, because the only remaining predictor is whether or not a participant stutters.

For VEMP amplitude, linear mixed-effects regression modelling (Winter, 2019) will follow the form:

$$VEMP\ amplitude = \beta_{0_j} + \beta_1 \times stutter + \mathcal{E}$$

Where vestibular amplitude is conditioned on whether or not participants stutter, with β_0 as intercept (varies with participant, j) and β_1 as a fixed slope of increase in vestibular response with stimulus level. Varying slope models were also appraised (see Appendix B).

Statistical analysis was conducted with the lme4 package (Bates & Maechler, 2020) in R (R Development Core Team, 2020). Effect size (Cohen's d) was calculated from mixed model t statistics with the EMAtools package for R, version 0.1.3 (R Foundation). Conditional R^2 was calculated according to Nakagawa, Johnson & Schielzeth (2017) using the MuMIn package, version 1.43.17 (R Foundation).

2.2 VEMP amplitude

Figure 2 indicates an approximately normal distribution for data, along with box plots showing participant pairing. Application of the random

intercepts model described in section 2.1 gives the statistically significant result that PWS have a VEMP amplitude 8.5 dB smaller than paired controls for the range of stimulus levels tested (p = 0.035, 95% CI [-0.9, -16.1], Chi-Squared (1) = 4.44, d = -0.8, conditional R^2 = 0.88).

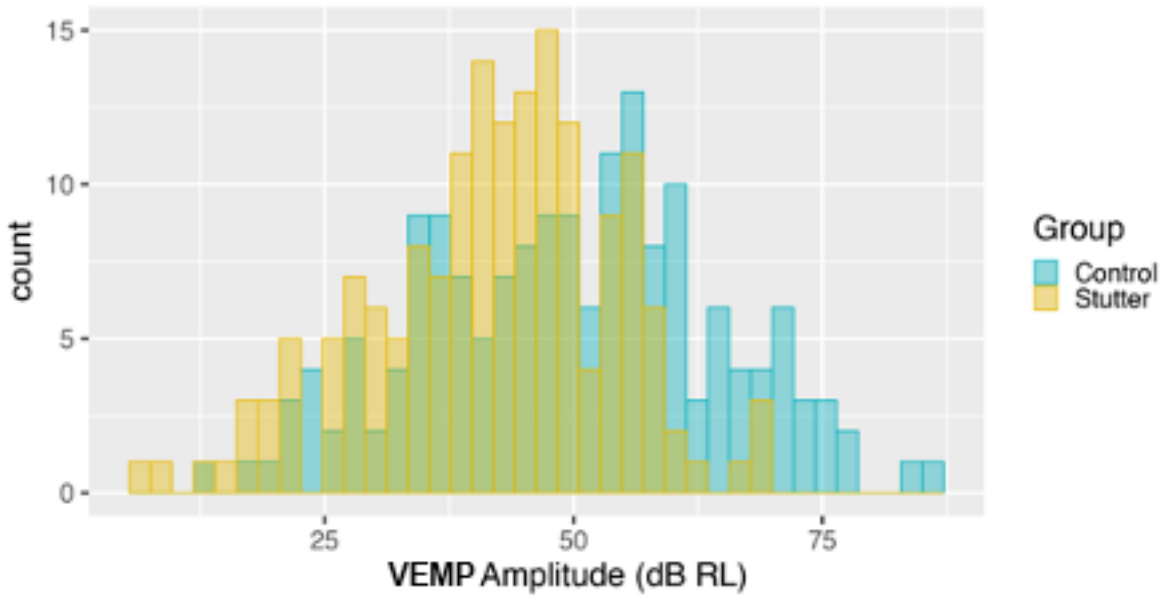

Figure 2: Histogram of VEMP peak to trough (p1-n1) amplitudes for PWS and paired controls. The histogram contains repeated measurements for two groups of 15 participants per group at different stimulus levels. As such, it suggests shape of distribution and direction of group difference, but is not appropriate for statistical comparison (statistical comparison is by linear mixed modelling).

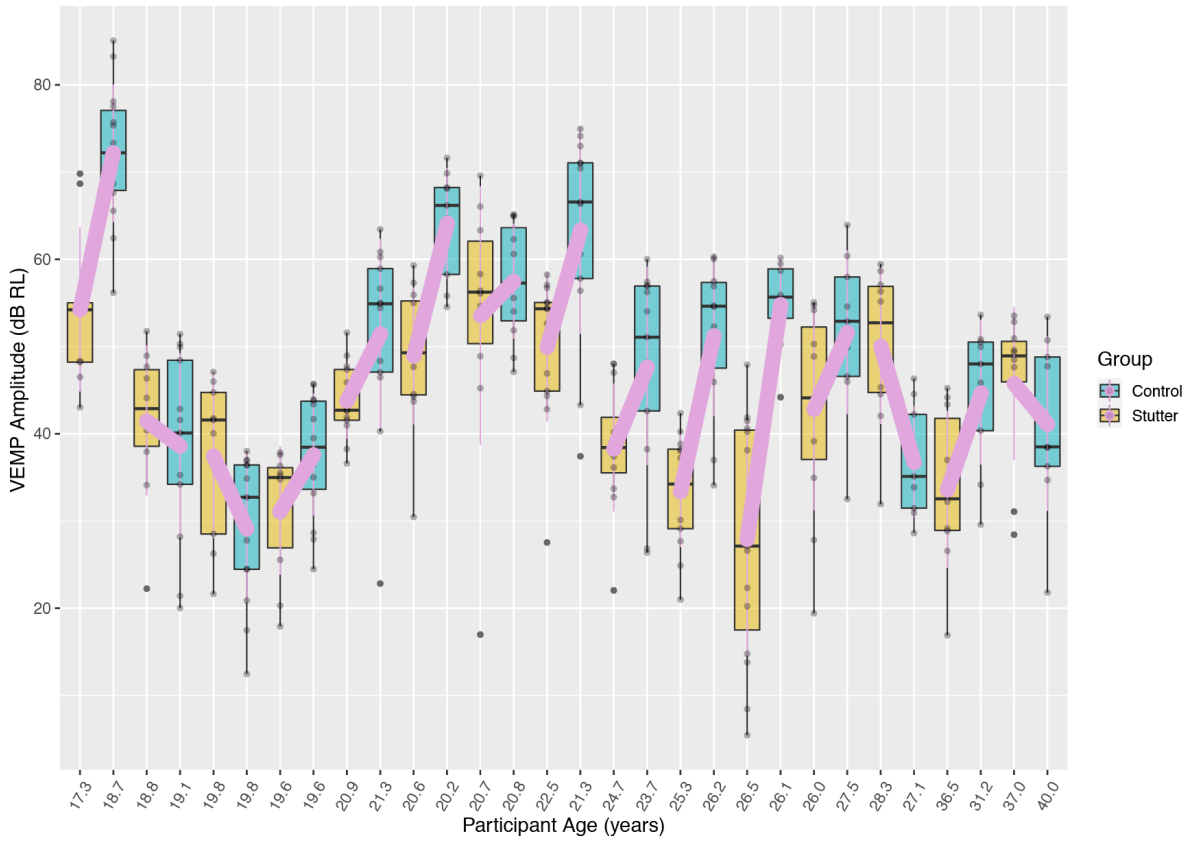

Figure 2 (supplementary): Box plots showing distribution of VEMP amplitudes per participant (across all stimulus levels tested), and participant age. Lines link the mean VEMP amplitudes of participants who stutter with those of their paired controls.

In linear mixed-effects regression modelling, there is a trade-off between greater possibility of type I error when all participants have the same slope but can have different intercepts, versus lower statistical power when both slope and intercept can vary per participant (Barr, Levy, Scheepers & Tily, 2013; Matuschek, Kliegl, Vasishth, Baayen & Bates, 2017). Analysis of a wider range of models, including random slopes, is detailed in Appendix B, with data files and R code available for download. A convergence warning with varying slopes can be removed by removing outlying data. All fixed and varying slope models evaluated have a difference between PWS and control groups of between -7.9 and -8.7 dB RL in VEMP amplitude, with p values between 0.021 and 0.049. Slopes per participant are shown in figure 3. Given that the slopes in figure 3 are approximately parallel, a fixed slope model was chosen. The final model for VEMP amplitude is detailed in figure 4.

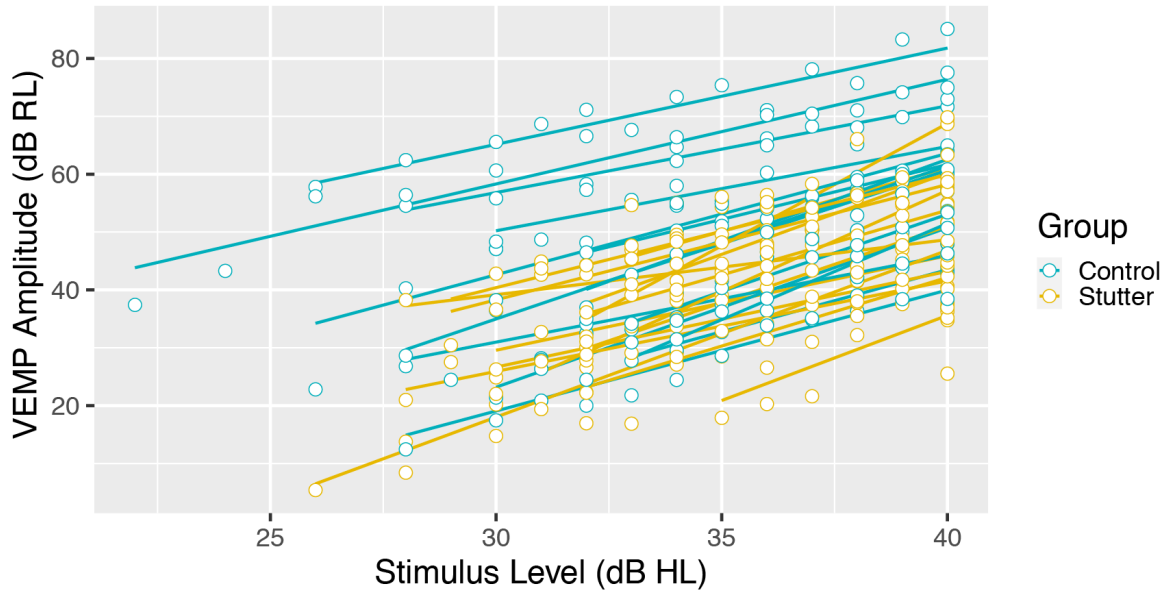

Figure 3: Per participant slopes of stimulus level (dB HL) versus VEMP amplitude (dB RL). A fixed slope, varying intercept model is supported if the least squares fit lines shown in this diagram are considered approximately parallel. See discussion in Appendix B.

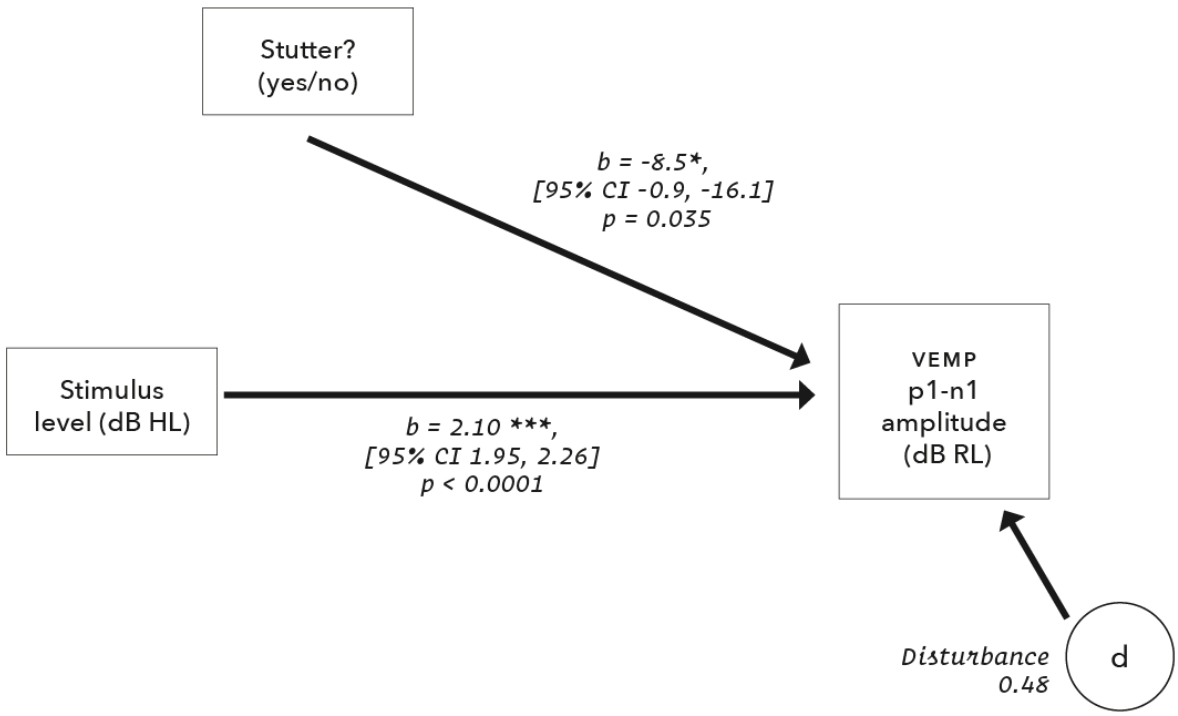

Figure 4: Final model for VEMP amplitude. The disturbance represents influences other than those measured in the model, and is the square root of $(1 - R^2)$, where the conditional R^2 is calculated according to Nakagawa, Johnson & Schielzeth (2017) using the MuMIn package (version 1.43.17). For the control group, using the calibrations and data transformations in this report, the y-axis intercept is - 24.9 dB RL (95% CI [-32.6, -17.2]). The suggestion is of VEMP thresholds at 11.8 dB HL for the control group, and 20.3 dB HL for the PDS group. However, VEMP thresholds estimated in this way (extrapolation to 0 dB RL) assume a linear

relationship between stimulus level and VEMP amplitude over a wider stimulus range than was tested in this report. Such estimates should not be directly compared to VEMP thresholds estimated by clinical search procedures (e.g. as per BSA, 2012).

2.3 VEMP latency

No statistically significant group differences were found for peak to trough VEMP latency. Figure 8 shows latencies collected across all participants and all stimulus levels, including repeat measurements.

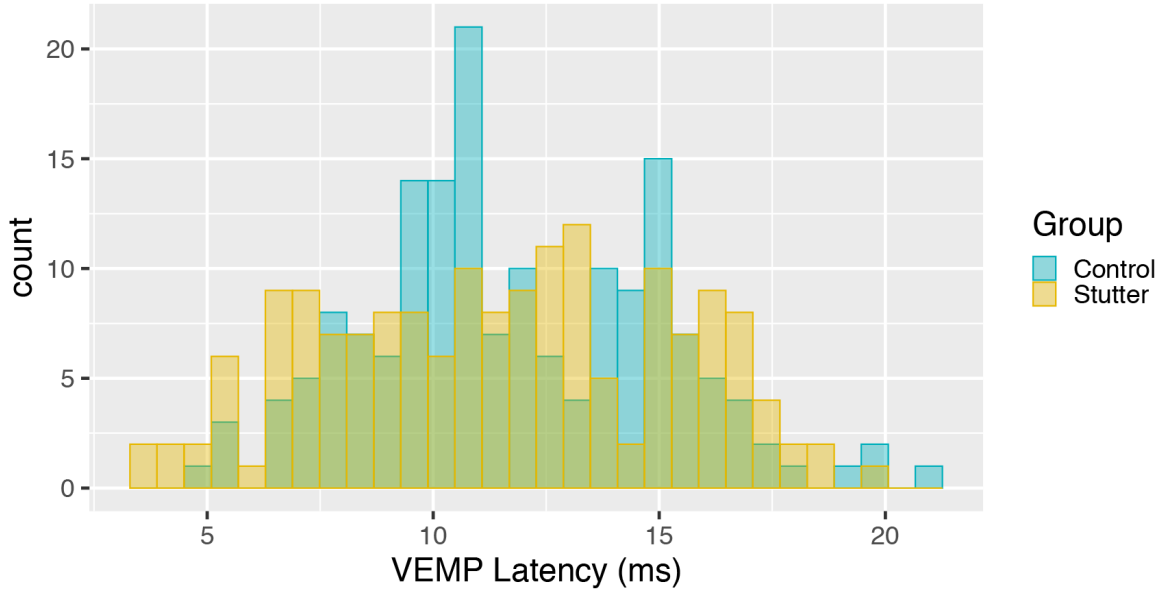

Figure 8: Histogram of peak to trough (p1-n1) latencies for PWS and paired controls. The histogram contains repeated measurements for two groups of 15 participants per group. As such, it suggests shape of distribution and direction of group difference, but is not appropriate for statistical comparison (statistical comparison is by linear mixed modelling).

Data appear normally distributed, with no indication of a group difference. Variation across participants with stimulus level is shown in figure 9. There is no statistically significant interaction. Pearson's correlation coefficient between VEMP latency and stimulus level is $r(165) = 0.13$, $p = 0.10$, 95% CI [-0.03, 0.27] for PWS data; and $r(165) = 0.06$, $p = 0.46$, 95% CI [-0.10, 0.21] for control data.

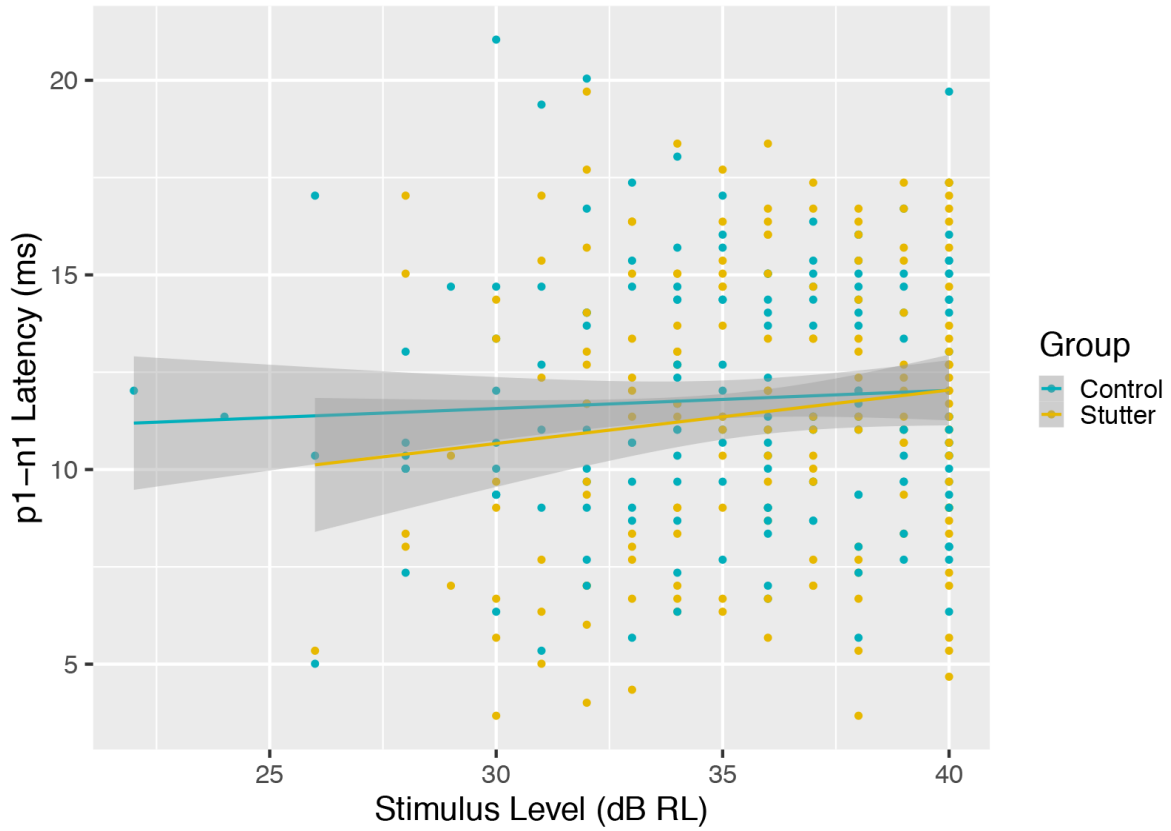

Figure 9: Variation of VEMP peak to trough (p1-n1) latency with stimulus level.

Group comparisons were evaluated through linear mixed model analysis, with p-values generated by likelihood ratio comparisons between the following models:

```
xnull: logp1n1 ~ 1  + (1|PiD)
xmdl:  logp1n1 ~ 1 + isPWS + (1|PiD)
```

There is no statistically significant difference between groups (chi squared (1) 0.07, $p = 0.8$).

3. Discussion

3.1 Summary of results

This pre-registered study assessed two novel hypotheses: that own voice is identified through coincidence detection between vestibular and cochlear streams (H1), and that own voice is misidentified in persistent developmental stuttering (H2). The hypotheses were both developed as best explanation arguments (Lipton, 2004). Through abduction it follows that the combination (H1 + H2) is also a best explanation argument – H1 and H2 are mutually reinforcing.

The finding of a VEMP amplitude 8.5 dB smaller in PWS than in paired controls (p = 0.035, 95% CI [-0.9, -16.1], t = -2.1, d = -0.8) supports H1 and H2. With the units in this report, a 10 dB increase corresponds to perceptual doubling. Thus, the implication is that for PWS the perceptual experience of the vestibular component of own voice is approximately half that of paired controls. This is despite normal hearing as assessed by pure tone audiometry for both PWS and paired controls (see appendix A).

Further analysis details reinforce the finding:

- Of the 15 controls, the 7 controls aged younger than 21 years were selected from a normative sample of 48 undergraduate students (Gattie, Lieven & Kluk, in preparation). VEMP response in these controls is representative of the normative sample of 48, rather than a normative sample of 7 as would have been the case if controls aged younger than 21 years had been sampled randomly from the general population.
- Two of the PWS differed from the remaining 13 PWS. One had both cluttering and stuttering, the other had stuttering with a possible psychogenic rather than developmental origin. In comparison with other PWS, VEMP responses for these PWS were unusually large relative to paired controls. If the two

participants had been excluded, on the basis that they were not typical cases of PDS, analysis of the remaining 13 PWS would have shown a VEMP response 11.2 dB smaller than in their paired controls ($p = 0.007$, 95% CI [-3.6, -18.9]).

- Pilot data for the study were reanalysed using the scripts developed for the main report. Comparison of 5 PWS and matched controls gives a result similar to the main study, with VEMP response 10.1 dB smaller in PWS than in matched controls ($p = 0.044$, 95% CI [-1.3, -18.9]).

Appendix A provides details of participants, and appendices C, D and E describe the pilot study. There is a less than 1 in 20 chance of the finding of a smaller VEMP response in PWS than paired controls being due to sampling variation. The null hypothesis of no group difference should be rejected.

Investigations of the brainstem and inner ear in PDS are substantially fewer than those of the cerebrum. Published literature shows results which are often slender, contradictory or difficult to interpret (e.g. Rousey, Goetzinger & Dirks, 1959; Stromstra, 1972; Hannley & Dorman, 1982; Howell, Marchbanks & El-Yann, 1986; Kramer, Green & Guitar, 1987) with consistent findings compared to controls being that PWS have a lower uncomfortable loudness level (MacCulloch & Eaton, 1971; Brown, Sambrooks & MacCulloch, 1975) and show greater differences in auditory brainstem response when stimuli resemble speech (Tahaei, Ashayeri, Pourbakht & Kamali, 2014, Crivellaro Goncalves, Furquim de Andrade & Gentile Matas, 2015; Mozaffarilegha, Namazi, Tahaei & Jafari, 2019) than when stimuli are clicks (Stager, 1990; Suchodoletz & Wolfram, 1996). Previous research on the vestibular system in PDS indicates that evoked nystagmus during speech is more pronounced in PWS than in controls (Langová, Morávek, Široký & Sváb, 1975). Together with the current result, the suggestion is of divergence in central vestibular function, and/or the nature of conduction along the VIII cranial nerve, between PWS and controls.

The remainder of this section proceeds on the basis that the finding of a smaller VEMP response in PWS than in controls is genuine, supporting H1 and H2. It will outline a biophysical basis for H1. Discussion will cover application to speech-motor research, and PDS in particular, and will describe implications from H1 for auditory scene analysis and theories of mental content.

3.2 Biophysical basis for H1

Own voice is identified in H1 with a biophysically grounded mechanism. The biophysical grounding is that there are two sets of mechanoreceptors for own voice. These are illustrated in figure 10. Sound and vibrational energy from vocalisation which deflects sterocilia in cochlear hair cells corresponds to own voice mixed with ambient environmental sounds. Vibrational energy from vocalisation which deflects stereocilia in vestibular hair cells corresponds to own voice in isolation. Comparison of nerve impulses arising from cochlear and vestibular systems therefore provides a principled distinction between self and environment.

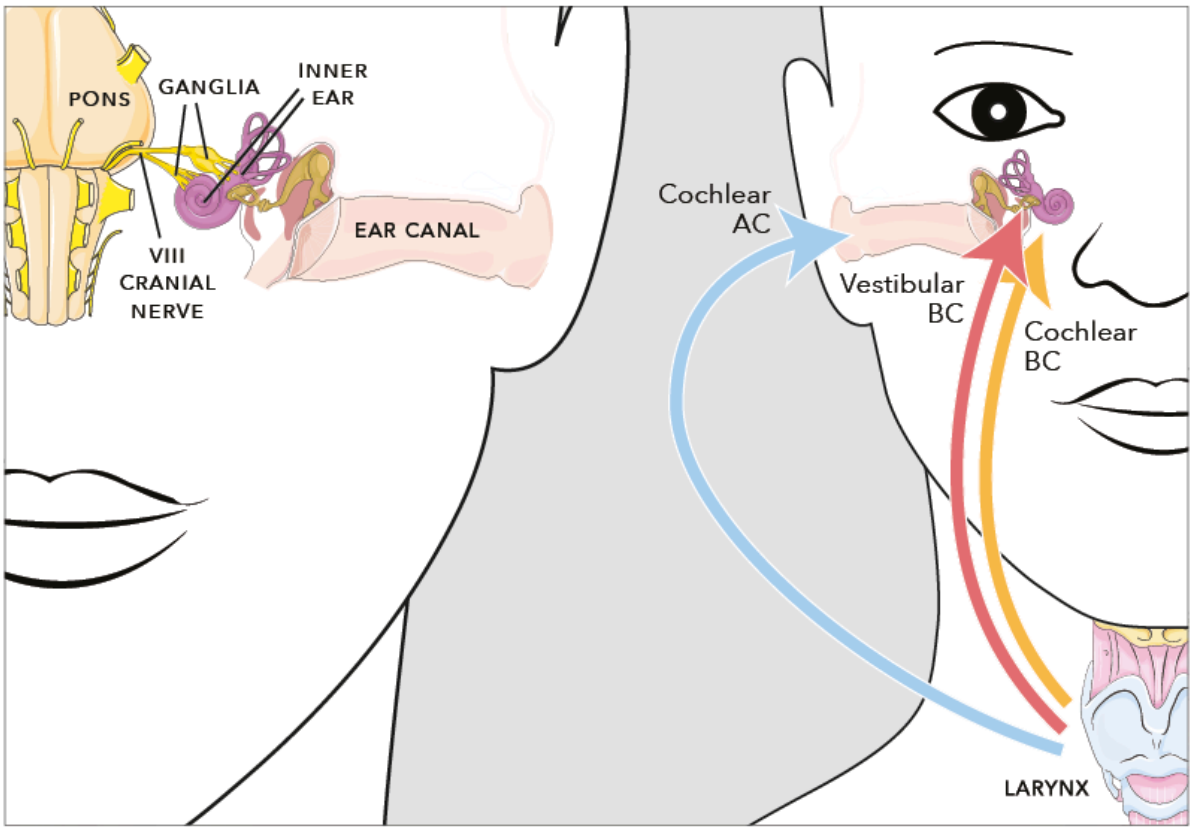

Figure 10: Parameters affecting sound and vibration detection in the human ear.
Left hand side: Anatomical parameters. Dendrons of bipolar ganglion cells terminate on sensory epithelial hair cells in the inner ear. Axons from the ganglia project or branch through the VIII cranial nerve to nuclei of the 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 nuclei. The activity described persists throughout the lifespan.
Right hand side: Acoustic and vibrational parameters. During vocalisation, sound and vibration energy originates at the larynx (and occasionally via tongue clicks, e.g. as in Khoisan languages). 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. AC and BC stimulation above hearing threshold (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. Similar activity in vestibular hair cells follows only from a considerably higher stimulus level. Welgampola, Rosengren, Halmagyi & Colebatch (2003) found vestibular thresholds (reported here in HL equivalents) 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.

© Adapted by Max Gattie from illustrations by Servier Medical Art, https://smart.servier.com. Creative Commons 3.0 licence.

Estimating arrival times for own voice stimuli at the inner ear requires consideration of propagation routes (figure 11). Air-conducted (AC) sound can be direct (dAC) or reflected (rAC), whereas reflection is not an important factor in propagation of body-conducted (BC) vibration. Calculations in figure 11 estimate 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 2 compares the propagation timings. Binaural coincidence detection across cochlear and vestibular mechanoreceptors, based on dAC sound and BC vibration, would identify own voice according to H1.

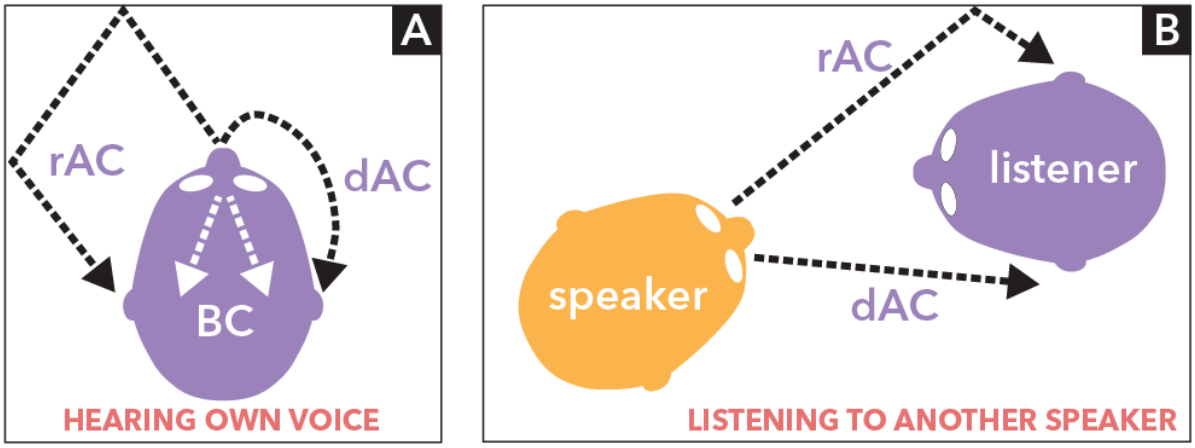

Figure 11: Sound and vibration routes to the ear. Air-conducted sound is split between reflected (rAC) and direct (dAC) routes (Cabrera, Sato, Martens & Lee, 2009; Traer & McDermott, 2016). These are shown in a simplified version. The rAC consists of many environmental reflections with comb filtering (frequencies attenuated or reinforced due to phase differences) as sound energy reaches the ear (Yadav, Cabrera & Martens, 2012; Arend, Stade & Pörschmann, 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).

(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, Perez, Adelman, Sohmer & Kishon-Rabin, 2018). Propagation routes will be 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 is 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, Newman & Rosenzweig, 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, Wilson & Leah, 1996). For animals with a pinna, filtering effects of the pinna reduce localisation inaccuracy for sources within the cone of confusion (Musican & Butler, 1984).

© Creative Commons 4.0 licence.

HEARING OWN VOICE: Routes to each ear

	Sensory organ	Arrival time after vocalisation	Interaural arrival time and intensity	Comment
BC BC	Vestibular system Cochlea	~ 0.5 ms ~ 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.
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, Östli, Håkansson & Stenfelt, 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 2: 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 H1 can be found in the cochlear nucleus and superior olivary complex. See review of brainstem neurons in Golding & Oertel (2012) and review of vestibular inputs to the cochlear nucleus in Newlands et al. (2003) or Smith (2012). 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 these nuclei which correspond to the coincidence detection hypothesised in H1 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 (defined following Bregman, 1990) which identifies own voice. The existence of such an own voice auditory stream, identified through coincidence detection, is the central claim of H1.

3.3 Application to speech-motor research

The own voice auditory stream described by H1 would provide a target for the proposed efference copy of the speech plan in predictive feedback control models (e.g. Hickok & Poeppel, 2007; Roelofs & Meyer, 1999; Tourville & Guenther, 2011). There is no prior research offering a basis by which an own voice auditory stream is specifically distinguished from ambient stimuli (Shamma & Micheyl, 2010; Remez & Thomas, 2013; Bronkhorst, 2015). If applied to speech-motor models, the mechanism described in H1 could improve explanatory power.

A corollary of this proposal is that tests of speech-motor activity should use physiologically valid own voice stimuli, as described by H1 (i.e. with stimuli as in described in figure 11 and table 2). Creation of such tests carries practical difficulty. For example, an ideal test of speech-motor activity would compare the brain response to identical sound and vibrational stimuli in the following conditions:

1. Brain activity generates sound and vibrational stimuli following coordinated nerve impulses to articulatory muscles, and brain activity is altered following audition of the stimuli (i.e. the standard articulatory process).
2. Brain activity is altered solely in response to the stimuli.

The experimental arrangement in (2) is impossible even in animal models because articulation cannot be elicited without disrupting resting

state brain activity. Accordingly, much testing of brain activity during vocalisation has been based around the following comparison:

3. The standard articulatory process (defined as per 1, above), with simultaneous recording of brain activity and the sound and/or vibrations created during vocalisation.
4. Brain activity without vocalisation while the sound and/or vibration recorded in (3) is played back.

Protocols using such comparison must choose a methodology for recording and playback of sound and/or vibration. Possibilities are shown in a Latin square in figure 12. 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, Östli, Håkansson & Stenfelt, 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). This 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). 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 hypothesised in H1.

Figure 12: 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 (i.e. the act of speaking, as per section 3.3), including speech-motor activity along with dAC, rAC and BC stimulation. "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 vocalisation. 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 2 (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 vocalisation.

Many functional imaging studies have compared vocalisation to AC playback of own voice recordings (e.g. with human participants: Numminen, Curio, Neuloh, Jousmüki & Hari, 1998; Numminen & Curio,

1999; Curio, Neuloh, Numminen, Jousmäki, & Hari, 2000; Ford et al., 2001; Houde, Nagarajan, Sekihara, & Merzenich, 2002; Ford & Mathalon, 2004; Ventura, Nagarajan, & Houde, 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, Houde & Rong, 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, Ramanarayanan, Nagarajan & Houde, 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 12), 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. H1 is consistent with this alternative explanation. H1 adds the proposal 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 streams. Whereas in the AC playback condition, operation of the ascending auditory pathway is along the lines described for 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 H1 – but alternatives would be to propose a different method by which own voice is identified, or to stipulate that an auditory target map for own speech is innately specified.

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, Silbersweig, & Frith, 1996; Hirano, Kojima, et al., 1997; Fu et al., 2006; Parkinson et al., 2012; Toyomura et al., 2007; Zarate & Zatorre, 2008; Tourville, Reilly & Guenther, 2008; Zheng, Munhall, & Johnsrude, 2009; Zarate, Wood, & Zatorre, 2010). In auditory perturbation studies, vocalisation is recorded, is optionally digitally manipulated, and is played back with a short delay whilst vocalisation 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. These delays are orders of magnitude longer than the submillisecond timings in table 2. Such 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 12 is a simplification. Stimuli can be further subdivided into those above and below vestibular threshold. Todd, Paillard, Kluk, Whittle & Colebatch (2014a, 2014b) have 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 study will appraise brain activity following combinations of BC and AC stimuli which are respectively above and below vestibular threshold.

3.4 Implications for stuttering research

A difference in the nature of own speech audition between controls and people who stutter (PWS) has been proposed to underlie persistent developmental stuttering (PDS) since at least the introduction of feedback control speech-motor models in the 1950s (see review in Yates, 1963; Bloodstein, 1995). Two proposals for brainstem activity are germane. In the disrupted rhythm hypothesis (Howell, Powell & Khan 1983; Howell, 2004) sensory inputs additional to own speech audition will be maximally fluency-enhancing when they coordinate with ongoing speech. And Doslak, Healey & Riese (1986) observe that stuttering is similar to nystagmus, a repeated involuntary eye movement.

Applying the combination H1 + H2 to both of these proposals adds detail. The detail is that in ordinarily fluent speakers, it is vestibular input to the own voice auditory stream which enhances fluency according to the disrupted rhythm hypothesis. Whereas a smaller vestibular input to the own voice auditory stream for PWS, consistent with the finding of smaller VEMP response for PWS in the current study, underlies the dysfluency in PDS. According to H1 + H2, the observed similarity between stuttering and nystagmus is due to a common cause from vestibular inputs to motor systems: the speech-motor system for stuttering, and the oculo-motor system for nystagmus.

These examples demonstrate how H1 + H2 can create compatibility between otherwise disparate accounts of stuttering. In a similar way, the combination H1 + H2 has the potential to improve understanding of cortical activity in PWS. Recall the hypothesised moderation of temporal cortex activity by speech-motor activity which was discussed in section 3.3. 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, Guenther, Gracco, Ghash & Wallace, 2004; Brown, Ingham, R. J., Ingham, J. C., Laird & Fox, 2005; Hickok, Houde & Rong, 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 3.3, 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 H1 + H2 when applied to a predictive feedback control model.

A complete description of PDS will need to integrate speech-motor and psycholinguistic accounts, as well as predict how, why and when observable stuttering behaviours should manifest, and explain brain activity not discussed here – notably, in the cerebellum and basal ganglia (Belyk, Kraft & Brown, 2015; Alm, 2004). Such a description is beyond the scope of the current article, but a follow-up in preparation will address the theoretical detail.

3.5 Wider implications of H1

The description of H1 in this article is specific to mammals. However, H1 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 animal which produces sound or vibration, and has two or more sets of sensory receptors capable of

detecting sound or vibration. See species surveys in Suthers, Tecumseh Fitch, Fay & Popper (2016) and Pollack, Mason, Popper & Fay (2019).

The prospect of such a wide taxonomic application for H1 suggests a provenance early in evolution. This in turn prompts reconsideration of the role of the inner ear. A follow-up article (in preparation) will contain detailed discussion. Two implications of H1 are outlined here.

1. 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 would identify the auditory streams which comprise auditory scenes:

- Primitives: 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).
- Schemas: Specific processes identifying certain types of sound, e.g. conspecific animal vocalisations or phonemes in human speech (Bey & McAdams, 2002; Billig, Davis, Deeks, Monstrey & Carlyon, 2013; Woods & McDermott, 2018)

The coincidence detection hypothesised in H1 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, Denham & Nelken, 2009; Szabó, Denham & Winkler. 2016; Snyder & Elhilali, 2017; Chakrabarty & Elhilali, 2019). Whichever modelling approach is taken, H1 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. Models will have greater neurological plausibility if the coincidence detection in (ii) occurs early in auditory scene analysis (see section 3.2).
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.

The own voice identification schema of (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 11 and table 2) 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, Cirstea, Pardhan & Moore, 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 own voice identification as described by H1. The vocalisations of neonates have more than sufficient energy to deflect both cochlear and vestibular mechanoreceptors, meaning that auditory learning based on H1 would begin at birth.

2. Distinction between self and environment

H1 provides a principled distinction between self (identification of own voice) and environment (reflection of own voice from surroundings). Such a distinction has importance for cognitive science and philosophy of mind (Wilson & Foglia, 2017). For example, in a representational theory of mind the distinction between self and environment is integral to content determination (Pitt, 2020).

The basis of the self-environment distinction in H1 is the presence of two sets of mechanoreceptors in the ear. One set of mechanoreceptors detects own voice in isolation, the other detects own voice mixed with ambient sound, including reflection of own voice. This is dissimilar to other modalities. For example, the visual analogy to H1 would be identification of one's own hand. However, photoreceptors do not collect sufficient information to identify one's own hand from light waves incident on the retina. Such identification would be possible following multisensory integration, but this is also the case for audition (e.g. as in the combination of audition with proprioception during vocalisation).

Perhaps audition is the only modality within which self and environment can be distinguished. If so, multisensory integrations including audition could underlie self-environment distinction for modalities other than audition. Evolution of any such dependency would have to create phenotypes sufficiently robust to account for self-environment distinction

when hearing ability is absent. See discussions of heritability and innateness in Griffiths (2020), Godfrey-Smith & Sterelny (2016) and Downes & Matthews (2020).

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

3.6 Conclusion

This article developed a hypothesis, H1, that own voice is identified through coincidence detection between cochlear and vestibular auditory streams. A biophysical basis was outlined, involving deflection of cochlear and vestibular mechanoreceptors and firing rates of neurons in the cochlear nucleus and superior olivary complex. Direct investigation of these ideas might best be carried out using animal models.

An alternative is to investigate brain activity in groups which could be expected to show differences from controls in own voice identification. Examples of such groups include those experiencing auditory and/or vestibular neuropathy (Kaga, 2016), auditory hallucination (e.g. in schizophrenia – McLachlan, Phillips, Rossell & Wilson, 2013; Matthews et al., 2013; Weintraub et al., 2012; Waters & Fernyhough, 2019), and the subject of hypothesis H2 – persistent developmental stuttering (PDS). The finding of this study that vestibular response is smaller in people who stutter than in paired controls supports H1 and H2, through inference to the best explanation.

H1 was described in the context of predictive feedback control speech-motor models, with the suggestion that its application could improve

explanatory power of such models. An example of such application was given using PDS, in which misidentification of the own voice auditory stream as proposed in H1 and H2 prompts reinvestigation of hypotheses around speech-motor modification of temporal cortex activity in PDS. These hypotheses were at one time compelling, but their investigation has been discontinued due to what this article proposes is a mistaken lack of empirical support.

It is strongly suggested that rather than using AC stimuli, as is common currently, investigations of own voice should instead use stimuli with physiologically valid propagation routes and inner ear arrival times. As described in table 2 and figure 11, these will consist of an AC/BC combination which is binaurally symmetric, with coincident arrival at both inner ears. The exact nature of stimulus used (e.g. whether the spectral composition is pure tones, chirps or CV syllables, and to what degree spectral composition and stimulus level differ between AC and BC delivery) will depend on the precise research question.

Work in preparation will add detail to the theoretical outlines provided here, and will investigate the brain response to stimuli with physiologically valid propagation routes and inner ear arrival times in people who do and do not stutter.

Acknowledgements

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. Participant expenses were partly covered by the ESRC, and partly covered by a Terence Barry Grant Award from the Stammer Trust, Oxford, UK.

4 Methods

4.1 Background

Vestibular response to sound can be appraised from the cervical vestibular evoked myogenic potential (VEMP), a measure of inhibition of sternocleidomastoid (SCM) muscle spindles (Colebatch & Rothwell, 2004). The inhibition corresponds to activity in a three neuron arc underlying the vestibulo-cervical reflex (Colebatch, Rosengren & Welgampola, 2016; Corneil & Camp, 2018). The VEMP is a large response having a characteristic peak (p1) and trough (n1). VEMPs have been developed clinically as part of a test battery appraising the vestibular periphery and brainstem (Rosengren, Colebatch, Young, Govender & Welgampola, 2019).

Since there is no published study of VEMP in persistent developmental stuttering (PDS), a pilot study (see appendices C, D, E) was conducted based on clinical cervical VEMP protocols (BSA, 2012), with 8 people who stutter (PWS) and 8 matched controls. Although statistically significant differences were not found, power analysis based on the largest group difference (1.5 SD in means for p1-n1 latency) suggested a sample size of 15 PWS, 15 controls for 95% power. The current study was pre-registered (Gattie, Lieven, Howell & Kluk, 2018) on this basis.

4.2 Study design

It was anticipated that VEMP peak to trough (p1-n1) amplitude will:

- Increase with stimulus level (Todd, Rosengren & Colebatch, 2008; Dennis, Govender & Colebatch, 2016). A linear mixed models analysis takes advantage of this relationship, with between group comparisons based on VEMP growth rate parameters.
- Increase with SCM tension (Ochi, Ohashi, & Nishino, 2001), which must be greater than at resting state in order to record a cervical VEMP. Pre-stimulus SCM tension was measured and

analysed as a covariant. Variation was limited by asking participants to maintain a biofeedback target by pushing against a padded head bar.

- Decrease with age (Nguyen, Welgampola & Carey, 2010; Ochi & Ohashi, 2016), controlled for by matching participants on age (see appendix A).
- Exhibit a crossed response: cervical VEMPs are predominantly ipsilateral, but may sometimes have a contralateral component (Colebatch, Halmagyi & Skuse, 1994; Li, Houlden & Tomlinson, 1999; Colebatch & Rothwell, 2004; Ashford et al., 2016). Use of a binaural stimulus limited variation due to any between participant difference in the extent of contralateral activity, because ipsilateral and contralateral components of the VEMP from each ear will be present at both SCM muscles. The arrangement is imperfect, because the mastoid placement for the bone conductor introduces an asymmetry, with approximately 3–5 dB intracranial attenuation for the 500 Hz tone burst used (Stenfelt, 2012). However, this asymmetry in BC stimulation is consistent per participant. One interpretation of the group difference found in the current study is that just one of the PDS or control groups exhibits a crossed response. VEMP recordings using monaural AC stimuli which were collected for this project, but have not yet been analysed, will enable appraisal of the crossed response in PDS versus controls.

Two potential confounders could not be controlled for:

- SCM size and subcutaneous fat are likely to influence VEMP amplitude (Chang, Tsung-Lin Yang, Wang & Young, 2007; Bartuzi, Tokarski & Roman-Liu, 2010). The effect was not appraised, although it was minimised both by the normalisation procedure (see section 4.6) and the use of amplitude growth parameters for between group comparisons.
- Blood has electromagnetic properties (Beving, Eriksson, Davey &

Kell. 1994; Abdalla, 2011) meaning electromagnetic field variations due to blood flow will add noise to electromyographic recordings. The active electrode placement for cervical VEMPs, directly above the carotid artery, suggests that measurement of cervical VEMPs will be particularly affected by blood flow. This is mitigated by the large size of the cervical VEMP response. Stimuli were delivered at a rate of 5.1 per second, whilst resting state pulse rates are approximately one per second. As a result, variations in the electromyographic recording due to carotid artery blood flow will largely cancel out over the approximately one minute recording time, such that noise added due to blood flow is minimal.

Figure 1 (in section 2.1) summarises the model investigated.

4.3 Participants

Stuttering was assessed using the SSI-4 (Riley, 2009). Control participants were all assessed as not stuttering (SSI-4 score lower than 10) whilst PWS had SSI-4 scores between 18 and 39 (described according to the SSI-4 as ranging from “mild” to “very severe”). All participants included in the study had hearing within the normal range as assessed by otoscopy, tympanometry and pure tone audiometry. Participants were paired on sex, and to within 0.05 years (SD 1.05) on age in aggregate. Appendix A has full details of screening and pairing.

Prior to any testing, all participants gave written informed consent according to the Declaration of Helsinki. The University of Manchester Ethics Committee approved the study.

4.4 Electromyography

VEMPs were recorded on an Eclipse EP25 system (Interacoustics AS, Assens, Denmark). Disposable non-metallic silver chloride electrodes were used (type M0835, Biosense Medical, Essex, UK). Skin was prepared with NuPrep® (Weaver and Company, CO, USA) prior to

electrode attachment using Ten20® conductive paste (Weaver and Company, CO, USA). Electrode impedances were maintained below 3 k′Ω. An active electrode was placed over the sternocleidomastoid muscle on the right hand side, with reference and ground electrodes on the upper sternum and nasion, respectively.

The stimulus was a 500 Hz sinusoidal carrier with rectangular windowing generated by the Eclipse. The rise/fall time of zero, and plateau time of 1 ms, gives characteristics intermediate between a tone burst and a click (Lauli & Burkard, 2015). The stimulus was delivered at a rate of 5.1 per second through a B81 bone conductor (Radioear, MN, USA), positioned on the mastoid bone behind the right ear. The bone conductor was calibrated with a Model 4930 artificial mastoid and 2250 Investigator (Brüel and Kjaer, Naerum, Denmark), and an Agilent 54621A 2-Channel Oscilloscope (Keysight, CA, USA). Calibrations were based on the artificial mastoid having a reference equivalent threshold force level re 1μN of 40.2 dB for 500 Hz. Interacoustics provide a correction factor of 69.5 dB for peSPL to nHL conversion of a 2-1-2 500 Hz tone burst. This correction factor was applied to the 0-1-0 500 Hz tone burst, with bone conduction levels accordingly reported in dB HL. Thus, stimulus levels in this report are calculated as they apply to the cochlea, rather than the vestibular system. Maximum stimulus level was set at 40 dB HL, since sine waves with amplitude above 40 dB HL displayed clipped on the oscilloscope.

The electromyographic (EMG) signal was amplified and band-pass filtered prior to sampling on the Eclipse system, using the Interacoustics research licence. Low pass was a digital FIR filter of 102nd order at 1500 Hz, and high pass was an analogue Butterworth filter of 1st order at 6 dB / octave. Sample rate was 3 kHz.

4.5 Procedure

Participants were seated such that their forehead rested against a padded bar, using apparatus specially constructed for this experiment

(figure 13). Participants were instructed to push their heads against the padded bar such that they would maintain an EMG biofeedback target as close as possible to 50 μV RMS throughout testing. If the background EMG was lower than 50 μV RMS, the stimulus would stop playing and participants were instructed to push harder. Participants would rarely push harder than they needed to. The importance of maintaining a constant background EMG and push strategy was relayed to participants, and the experimenter monitored background EMG throughout.

Figure 13: Custom head bar. Participants were instructed to push against a padded bar using the forehead, such that SCM tension was maintained as close as possible to 50 µV RMS throughout testing. Biofeedback in the Eclipse clinical software enabled participants to monitor SCM tension.

Eclipse recordings followed the Interacoustics recommended procedure for VEMPs, with rejection of epochs having peak or trough amplitudes with magnitude larger than ± 800 μV. A software feature compensated for rejected epochs such that the averaged response to exactly 300 epochs was recorded for every stimulus level tested. Such averages of 300 epochs will be referred to henceforth as “sequences”. The initial sequence was recorded with a stimulus level of 40 dB HL, with further

sequences recorded with stimulus level descending in 2 dB steps until 34 dB HL or until the averaged VEMP trace in the sequence was comparable to background noise, whichever came soonest. Comparison of the sequence to background noise was made by the experimenter using the EP25 clinical software. A second series of recordings was initiated at 39 dB HL, with stimulus level descending in 2 dB steps until 35 dB HL or until the averaged VEMP trace in the sequence was comparable to background noise, whichever came soonest. The collection procedure was explained to participants, who could watch their averaged VEMP trace being calculated in real time by the EP25 software on a computer screen. If the participant was willing (e.g. if they had no time constraints) and if participants had shown a response at 34 dB HL, further sequences were recorded at stimulus levels below 34 dB HL. Sessions ended with repeat recording of a sequence using the maximum 40 dB HL stimulus.

4.6 Data Processing

Raw data was processed using custom scripts in MATLAB 2019a (The MathWorks, Inc., Natick, MA). Response amplitudes were transformed into a dimensionless ratio by normalising per participant as follows:

1. For each participant, a pre-stimulus interval of 18 ms was extracted from a mean of the EMG waveforms from the first six sequences of 300 presentations recorded (i.e. it was a pre-stimulus mean of the first 1800 presentations recorded).
2. The pre-stimulus mean per participant from (1) was converted into a root mean square (RMS) background EMG tension per participant.
3. All waveforms for a participant were normalised by dividing them by the pre-stimulus RMS which had been calculated for that participant in step (2).

This normalisation procedure is in principle not necessary, since background EMG tension is already tightly controlled at a target of 50 μV

per participant using the head bar and biofeedback (see section 4.1). The rationalisation for normalisation is to go even further than the tight control of the head bar and biofeedback procedure, by correcting for any small per participant variation in background EMG tension. Such variation may arise, for example, due to subtle differences in push strategy between participants.

Normalisation uses the maximum pre-stimulus data available for every participant (1800 presentations), minimising the presence of random noise per participant in the pre-stimulus RMS background EMG tension. This procedure is preferable to, for example, per sequence normalisation based on pre-stimulus RMS for each sequence of 300 presentations. Per sequence normalisation would introduce noise to data because random fluctuation in pre-stimulus RMS per sequence (i.e. random in addition to any actual change in SCM tension) would affect VEMP amplitudes randomly on a per sequence basis, thereby affecting within participant comparisons. Between participant comparisons use linear mixed-effects regression modelling (section 2), which depends on an accurate within participant measure of VEMP amplitude growth rate with stimulus level. As such, preserving within participant comparisons as accurately as possible – so, identically to the raw data with the normalisation procedures used in this study – is optimal for linear mixed-effects regression modelling.

Peaks and troughs were identified using the “findpeaks” algorithm in the MATLAB Signal Processing Toolbox (waveforms were inverted to find troughs). Initially peaks and troughs were identified for the first 40 dB HL sequence per participant, as follows:

1. Troughs were identified for the entire trace.
2. The most prominent trough (prominence as defined in the findpeaks algorithm) between 15–37 ms was identified as n1.
3. Peaks were identified for the entire trace.
4. Peaks earlier than 5 ms, and later than n1, were discarded.

5. Remaining peaks were ranked as follows:
    a. The three most prominent peaks were awarded 5, 4 and 3 points in order of prominence.
    b. The three most prominent peaks were weighted based on their prominence compared to the most prominent peak: 3 points awarded for greater than or equal to two thirds; 2 points for greater than one third and less than two thirds; and 1 point otherwise.
    c. The five peaks having the smallest time difference from n1 were awarded points from 5 to 1 in a hierarchy with higher points for smaller time difference.
    d. Points from (a), (b) and (c) were summed. The peak with the greatest number of points was identified as p1. Ties were decided in favour of the peak with smaller time difference from n1.

Peaks and troughs for other stimulus levels were identified in a similar manner, except that the trough from the initial 40 dB HL presentation was used as an anchor for trough detection for lower stimulus levels on a per participant basis. Peaks and troughs were rejected (the script returned an empty result) if the p1-n1 amplitude was less than 1.65 times the pre-stimulus RMS for the sequence of 300 repetitions being evaluated.

The script was checked through visual inspection of waveforms for the entire data set collected. This was an iterative procedure, with the script run several times using adjustments to some of the parameters described. Visual inspection showed that the final script identified peaks and troughs with a high degree of fidelity. Identification by the script was final – no data points were removed or adjusted manually.

Data were transformed to a response level (RL) scale by taking the log of p1-n1 amplitude as follows:

$$p1n1amp\ (dB\ RL) = 20\ \times \log_{10}\left(\frac{p1n1amp\ (\mu V)}{prestimulus\ RMS\ (\mu V)}\right) - 20$$

Zero dB RL denotes VEMP threshold (although, not the same as a clinical threshold: see note in figure 4). The transformation is similar to that for the dB HL units widely used for sound levels.

Appendix A: Participant details for Main Study

Participants were recruited differently for the two experimental groups.

Participants who stutter (PWS) were recruited from the Manchester Stammering Support Group. PWS self-referred to the group, either following a suggestion to do so from British National Health Service centres where they had received therapy for stuttering, or following discovery of the group via internet search. All PWS identified as stuttering upon presentation at support group meetings, and were only invited to participate after the lead author had observed them displaying overt stuttering behaviour. Unimpaired hearing was confirmed via self-report before the invitation to participate. PWS were compensated for participating in the two hour session in line with 2019 UK Living Wage Foundation recommendations.

Ordinarily fluent control speakers were recruited from two groups. Participants under 21 years old were psychology undergraduates at the University of Manchester, who participated in return for course credit. Working from a normative sample of 48 undergraduate students (recruited for Gattie, Lieven & Kluk, in preparation) a sample of 7 control participants was selected such that VEMP amplitudes would be representative of the normative sample of 48. This was achieved by choosing the 7 controls so that their average VEMP amplitudes across all stimulus levels tested were comparable (mean 48.8 dB RL, SD 14.8, median 51.3) to those of the 40 participants not chosen from the

normative sample (mean 48.7 dB RL, SD 14.2, median 50.0) and, given that 6 of the 7 undergraduate controls were male, those of the 24 male students in the normative sample of 48 (mean 50.7 dB RL, SD 17.8, median 51.0). Participants older than 21 years were postgraduates at the University of Manchester, and participated without compensation.

Control participants were told that the experiment was about PDS, and asked if they stuttered. No control participants identified as stuttering (if they had done, availability of the support group would have been explained to them). In all cases, unimpaired hearing was confirmed via self-report before the invitation to participate.

For PWS and control groups, the following tests were conducted prior to electrophysiological data collection:

- Otoscopic examination of the tympanic membrane.
- Middle ear impedance at 226 Hz, using an Interacoustics Titan IMP440 (Interacoustics, Denmark).
- Pure tone audiometry over air- and bone-conduction (AC/BC) using a GSI-3B Arrow™ audiometer (Grason-Stadler Inc., MN, USA) with TDH49 headphones (Telephonics Corp., Farmingdale, NY, USA) or a B71 bone conductor (Radioear, MN, USA). Participants were evaluated for thresholds below 20 dB nHL at 1000 Hz, 500 Hz, 250 Hz and 125 Hz (AC), and 1000 Hz, 500 Hz and 250 Hz (BC).

Testing followed British Society of Audiology recommendations. Three participants were excluded because deposits of cerumen were sufficiently substantial as to occlude view of the tympanic membrane. For one of these participants, tympanometry could not be performed successfully. All other participants presented with normal hearing according to the tests described. One participant was excluded due to a growth on the mastoid bone which prevented use of the bone conductor. Replacements were recruited for excluded participants.

Stuttering was appraised quantitatively using the SSI-4 (Riley, 2009). The SSI-4 is a norm-referenced stuttering measure based around count of stuttered syllables. Criteria were that controls would be excluded with an SSI-4 score above 10, and PWS would be excluded with an SSI-4 score below 10. SSI-4 categorisations for controls and PWS are shown in table 3.

Group	SSI-4 score and rating					
	Do not stutter (<10)	Very mild (10 – 17)	Mild (18 – 24)	Moderate (25 – 31)	Severe (32 – 36)	Very severe (37 – 46)
Controls	15	–	–	–	–	–
PWS	–	–	6	4	3	2

Table 3: SSI-4 scores for ordinarily fluent controls and participants who stutter.

A self-assessment questionnaire (attached to this appendix) was used to assess whether anything other than stuttering might affect the results of the VEMP test. Any positive answers to the questionnaire were followed up at interview, along with additional probe questions (see interview sheet attached). The questionnaire and interview were not intended to provide data for quantitative analysis. The intention was partly as an additional screen (given that time constraints meant that all factors of potential interest could not be assessed via test batteries) and primarily to gather hypothesis-generating data. Assessment of the auditory brainstem response to clicks has suggested PWS may separate into one group within, and another without, the response range of matched controls (Stager, 1990). If VEMP responses had suggested a subgrouping of PWS similar to that for the PWS auditory brainstem response to clicks, data collected via questionnaire and interview might have helped to design a follow-up study which could assess heterogeneity in PWS.

The results of the study (figure 2 box plot) showed only one case where a control had consistently smaller VEMP amplitude than a paired PWS.

Whereas in nine cases, VEMP amplitudes of PWS were consistently smaller than in paired controls. In the five remaining cases there was a partial overlap which, in the linear mixed models analysis (section 2.2), was part of a statistically significant group difference between PWS and paired controls for VEMP amplitude. There was therefore little suggestion that PWS should be sorted into subgroups based on VEMP response. As such, data collected from the questionnaire and follow-up interview will not be reported in full. Inferences from such data would be statistically invalid based on the small sample size in this study.

What can be reported, based on the questionnaire and interview, is that no participants presented with non-speech conditions which could have affected test results. Some individual reports from PWS included in the study are of interest:

- One PWS had a spinal condition which might have affected ability to complete the head bar push in the VEMP procedure. The participant had no difficulty completing the procedure, and VEMP recordings were similar to other PWS (i.e. VEMP amplitudes smaller than the paired control).
- Another PWS presented with cluttering in addition to stuttering. The SSI-4 indicated a positive diagnosis for stuttering, as did self-report of the PWS and the assessment of the lead author. The SSI-4 score was the lowest of any PWS tested (19, on a scale from 0 to 46 where scores greater than 10 are classified as stuttering), and the VEMP amplitude was on average slightly higher than the paired control. This is perhaps not a typical case of PDS. If this participant and the paired control had been excluded from analysis (i.e. on the basis that it is not a typical case of PDS), statistical analysis for the remaining 14 PWS with paired controls would have shown stuttering with an effect of -9.4 dB and $p = 0.020$ (compared to -8.5 dB and $p = 0.035$ using 15 PWS with paired controls).
- Another PWS presented with a history consistent with

psychogenic stuttering (van Borsel, 2014). Acquired stuttering is rarer than PDS, typically reported through case study, and occurs in psychogenic and neurogenic subtypes (Baumgartner & Duffy, 1997). Age of onset was 10 years for this participant, which is relatively late for childhood stuttering (Bloodstein, 1995). Although acquired stuttering is typically considered an adult condition, there is no principled basis by which children should be considered unsusceptible to acquired stuttering (Theys, van Wieringen, Tuyls & De Nil, 2009). Thus, it is not clear that this participant had PDS. Despite this, the participant was included in the PDS group. The reason for this was that differential diagnosis is not straightforward (Baumgartner & Duffy, 1997; Ward, 2010), and was further hindered by the 18 year gap between stuttering onset and participation in the current study. The SSI-4 test indicated mild stuttering, and onset at 10 years old suggests PDS, so PDS is a fair categorisation. However, it is notable that this PWS was unusual in having a clearly larger VEMP amplitude than the paired control. If this PWS and paired control had been excluded from analysis (i.e. on the basis of not being a case of PDS), statistical analysis for the remaining 14 PWS with paired controls would have shown that stuttering has an effect of -10.1 dB and $p = 0.015$ (compared to -8.5 dB and $p = 0.035$ using 15 PWS with paired controls).

To summarise: if 2 of the 15 PWS (with paired controls) had been excluded on the basis that their stuttering was atypical of PDS, statistical analysis for the remaining 13 PWS with paired controls would have shown stuttering with an effect of -11.2 dB (95% CI -3.6, -18.9 $p = 0.007$), compared to -8.5 dB (95% CI -0.9 to -16.1, $p = 0.035$) with 30 PWS and paired controls.

Full details of pairing are shown in table 4. For cervical VEMPs, the p1-n1 amplitude in response to BC tone bursts falls linearly with age, at a rate of approximately 0.14% per year (Colebatch, Govender &

Rosengren, 2013). PWS and controls were paired such that mean age difference was -0.1 years (SD 1.8). The mean mismatch in p1-n1 amplitude due to age difference can thus be estimated as -0.2%. Age effects were accordingly considered to cancel in between group comparisons, and age was not a factor in statistical analysis.

MAIN STUDY

Participant ID	Sex	SSI-4	Group	Age (yrs)	difference (yrs)	p1-n1 amp mismatch %
FZNN6543	Male	34	PWS	17.3		
CAMC0468	Male	8	CONT	18.8	1.4	2.0%
JOBA2769	Male	36	PWS	18.8		
JASP1911	Male	9	CONT	19.0	0.3	0.4%
NAHT5214	Male	19	PWS	19.8		
MAIT7329	Male	5	CONT	20.0	0.2	0.2%
BAIB6567	Male	25	PWS	19.6		
JOPS4312	Male	4	CONT	19.6	0.0	0.0%
IS4C4545	Male	38	PWS	20.9		
SCHW7222	Male	2	CONT	21.3	0.3	0.5%
SOO19192	Female	24	PWS	20.6		
JI7634	Female	4	CONT	20.2	-0.4	-0.6%
SAFA1218	Male	26	PWS	20.7		
NI6743	Male	2	CONT	20.8	0.1	0.1%
MILG4836	Male	39	PWS	22.5		
ISAB1291	Male	2	CONT	21.3	-1.2	-1.7%
HARR1598	Male	27	PWS	24.7		
LWIG0921	Male	0	CONT	23.7	-1.0	-1.4%
NIYT6794	Male	22	PWS	25.3		
HE8692	Male	6	CONT	26.2	0.9	1.3%
PA5679	Male	21	PWS	26.5		
RO8893	Male	0	CONT	26.1	-0.4	-0.6%
STEL5152	Female	27	PWS	26.0		
JE8779	Female	6	CONT	27.5	1.5	2.1%
GRTT6591	Male	23	PWS	28.3		
ADAD2376	Male	3	CONT	27.1	-1.3	-1.8%
SQRF5874	Male	35	PWS	36.5		
IB54347	Male	7	CONT	31.2	-5.3	-7.6%
JMJM5454	Male	21	PWS	37.0		
PEDN5219	Male	1	CONT	40.0	3.0	4.3%
	mean (of positive/negative values)			24.2	-0.1 (SD 1.8)	-0.2%
	mean (of absolute values)				1.2 (SD 1.4)	1.6%

Table 4: PWS shown adjacently to paired controls, with estimated error in p1-n1 amplitude due to age mismatch

STUTTERING STUDY
QUESTIONNAIRE

MANCHESTER 1824
The University of Manchester

PARTICIPANT NUMBER

It will help us if you can answer these questions as accurately as possible.

Sometimes you may be in doubt as to whether to reply yes or no. In these cases, please answer yes, and we will discuss your answer in a follow-up interview.

You do not have to answer any questions you do not want to answer. If you don't want to answer a question, leave it blank.

1. Do you tend to favour your right hand, your left hand, or neither hand?

   Right hand
   Left hand
   No hand preference

2. Have you noticed anything unusual about your hearing?

   Yes
   No

   If yes, please describe what's unusual:

3. Do you stutter, either now or in the past?

   Yes
   No

   If yes, please describe your history of stuttering:

4. Do you make any unusual or unexpected movements (e.g. tics)?

   Yes
   No

   If yes, please describe the movements:

5. How would you describe your sense of rhythm and coordination in comparison to your friends' sense of rhythm and coordination?

   Excellent
   Above average
   Average
   A little below average
   I have a rhythm of my own

   Please describe your sense of rhythm and coordination, if it is unusual:

PTO

Form 1: Self-assessment questionnaire, page 1.

6. Have you ever been diagnosed with any of the following disorders?

Articulation
Phonology
Voice
Cluttering
Stuttering
Dysphagia
Receptive language
Expressive language
Specific language impairment
Learning
Literacy
Attention deficit
Central auditory processing
Neuropsychological
Behavioural
Sensory integration
Acquired physical conditions
(e.g. diabetes or asthma)
Autistic spectrum
Epilepsy or other neurological disorder
Mental retardation
Hearing impairment
Congenital physical conditions
(e.g. cleft lip or palate)
Traumatic brain injury
Hormonal imbalance
Emotional disorder
Tourette's syndrome
Cerebral palsy

If you replied yes to any of the above, please give details:

7. Do you have any other medical history you think we should know about?

Yes
No

If yes, please describe the history:

8. Which of the following best describes your formal education?

Level 1 (e.g. GCSE grades D–G)
Level 2 (e.g. GCSE grades A*–C)
Level 3 (e.g. A level or AS level)
Level 4 (e.g. HNC)
Level 5 (e.g. HND)
Level 6 (e.g. Bachelor's undergraduate degree)
Level 7 (e.g. Master's degree)
Level 8 (e.g PhD)

Please return this form to:

Max Gattie
School of Psychological Sciences
Ellen Wilkinson Building
University of Manchester
Oxford Road, Manchester
M13 9PL

Tel: 0161 275 8567
email: max.gattie@postgrad.manchester.ac.uk

Form 2: Self-assessment questionnaire, page 2.

Participant Details

Unique identifier:

Name:

Date of birth:

Contact details:

Mother's education:

Languages:

Music experience:

Hearing history:

Recent loud noises:

Stuttering history:

Other details:

Head size:

- Maximum cross-sectional diameter
- Sagittal nasion to inion
- Coronal top section (from where pinna joins skull)

TURN YOUR PHONE OFF!

Form 3: Interview prompt sheet.

Appendix B: Detailed statistical analysis of VEMP amplitude

Visual inspection of figure 3 shows broadly similar slopes. However, perhaps some PWS slopes are atypical? The mean slope for PWS is 2.3 (SD 0.5), compared to 2.1 (SD 0.3) for controls. A two-sided t-test shows this difference is not statistically significant (t (21) = 1.1, p = 0.30) but a larger sample might indicate, for example, that the slope is constant in controls but differs in PWS. If so, the slope difference may convey physiological detail (e.g. activity in vestibular mechanoreceptors or VIII cranial nerve scales differently between PWS and controls as stimulus level increases) and a varying intercept and slope model would be appropriate.

Linear mixed models with fixed and varying slopes are summarised in table 5, in which p-values were generated by likelihood ratio comparisons to a nested model in which the "isPWS" predictor (true if a participant stutters) was absent.

Linear Mixed Model	Include Outliers	Comments	**PWS-CONT diff** (std err)	**p value** [(df) χ^2]
Fixed slope logp1n1 ~ dB + isPWS + (1|PiD)	Yes		**-8.47** (3.87)	**0.035*** [(1) 4.44]
	No		**-8.14** (3.90)	**0.043*** [(1) 4.08]
Varying slope logp1n1 ~ dB + isPWS + (1 + dB|PiD)	Yes	Convergence warning	**-7.90** (3.72)	**0.049*** [(1) 3.87]
	No		**-8.74** (3.61)	**0.021*** [(1) 5.30]

Table 5: Linear mixed model comparison. All models have varying intercepts. Model description is as per the lme4 syntax: logp1n1 is the VEMP amplitude in dB RL; dB is the stimulus level in dB HL; isPWS is a binary indicator of PWS (true) versus control (false); PiD is the participant ID. Data and R code are available for download with this appendix.

The simplest varying slope model creates a convergence error. The error can be overcome by removing 12 outlying data points (from 334 data points total; outliers are data outside the interquartile range, and

are visible mostly towards the bottom of the figure 2 box plot). When outliers are removed, there is no convergence warning, and PWS have a VEMP amplitude significantly smaller than paired controls by 8.7 dB (Chi-Squared (1) = 5.30, p = 0.021). Whichever model is chosen, the p value is below 0.05.

Appendix C: Results from Pilot Study

In the pilot for the current study, analysis of VEMP thresholds used a clinical threshold estimation procedure (BSA, 2012) and established no statistically significant group difference between PWS and controls. Here, pilot data is reassessed using the same model and analysis scripts described for the main study (section 2 and Appendix B).

Limitations of pilot study procedures, in comparison to procedures described in the main study, are that:

- Presentation was a descending clinical threshold search (BSA, 2012), rather than the counterbalanced even and odd descending sequences described in section 4.5
- An earlier version of the EP25 software was used. This identified the Radioear B81 as the earlier model B71 bone conductor. The B81 is capable of an output between 11–16 dB higher than the B71 at 500 Hz (Jansson, Håkansson, Johannsen & Tengstrand, 2015) but the earlier version EP25 software would not drive the Radioear B81 at such levels.
    - Stimulus levels for the pilot study should be considered as dial settings, since calibrations were manipulated to obtain the highest possible output from the B81.
    - The maximum stimulus level used in the pilot study is lower than that for the main study.
- SCM tension was generated via head raise from supine, and will be more variable than when using the head bar and biofeedback.

Because the head bar was not used to control SCM tension, a model fit was chosen in which the pre-stimulus root mean square of EMG is used as an additional predictor corresponding to background SCM tension. This is in line with the model described in figure 1 (section 2.1).

Figure 14 suggests a normal distribution for VEMP amplitudes (given small sample size), and shows box plots with participant matching. Application of the random intercepts model described in section 2.1 gives the statistically significant result that PWS have a VEMP amplitude 10.1 dB smaller than matched controls for the range of stimulus levels tested (p = 0.044, 95% CI [-1.3, -18.1], Chi-Squared (1) = 4.05). See appendix E for discussion of alternative linear mixed models.

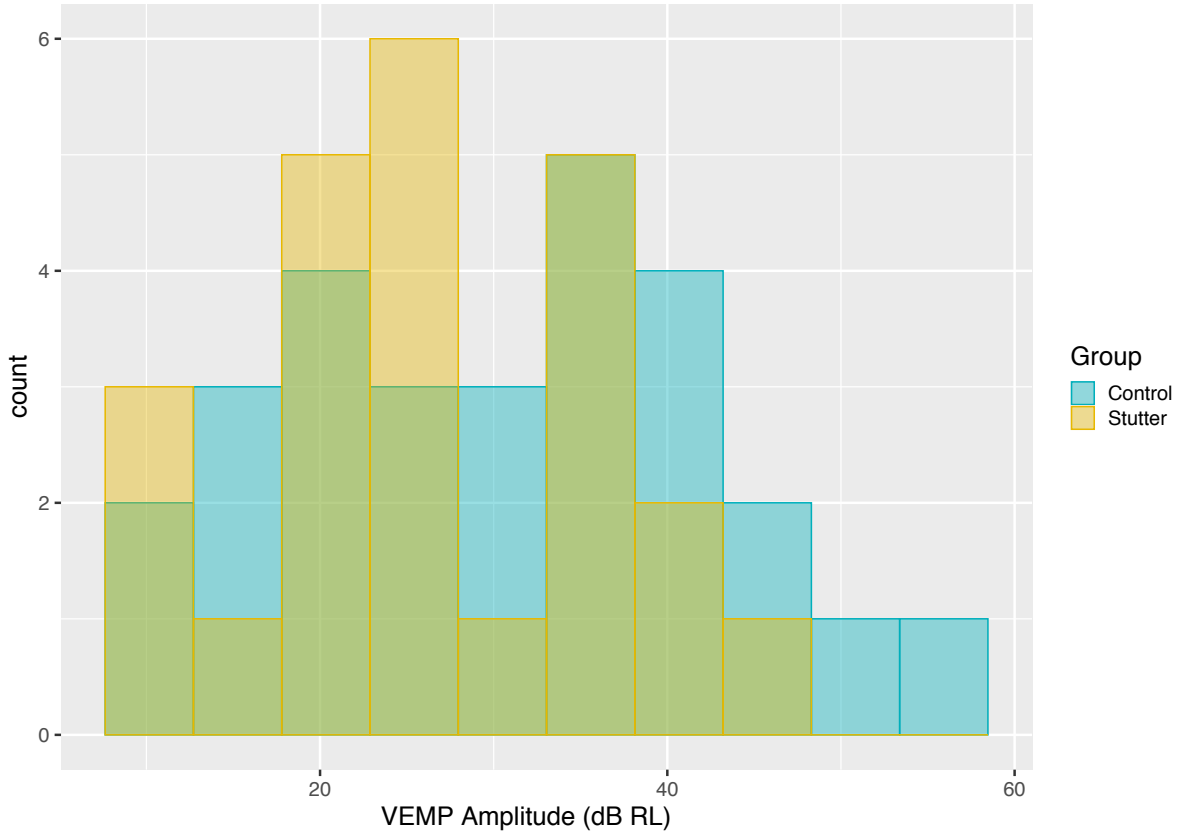

Figure 14: Histogram of VEMP peak to trough (p1-n1) amplitudes for PWS and paired controls for the pilot study (compare with figure 2). The histogram contains repeated measurements for two groups of 5 participants per group at different stimulus levels. As such, it suggests shape of distribution, but is not appropriate for statistical comparison (statistical comparison is by linear mixed modelling).

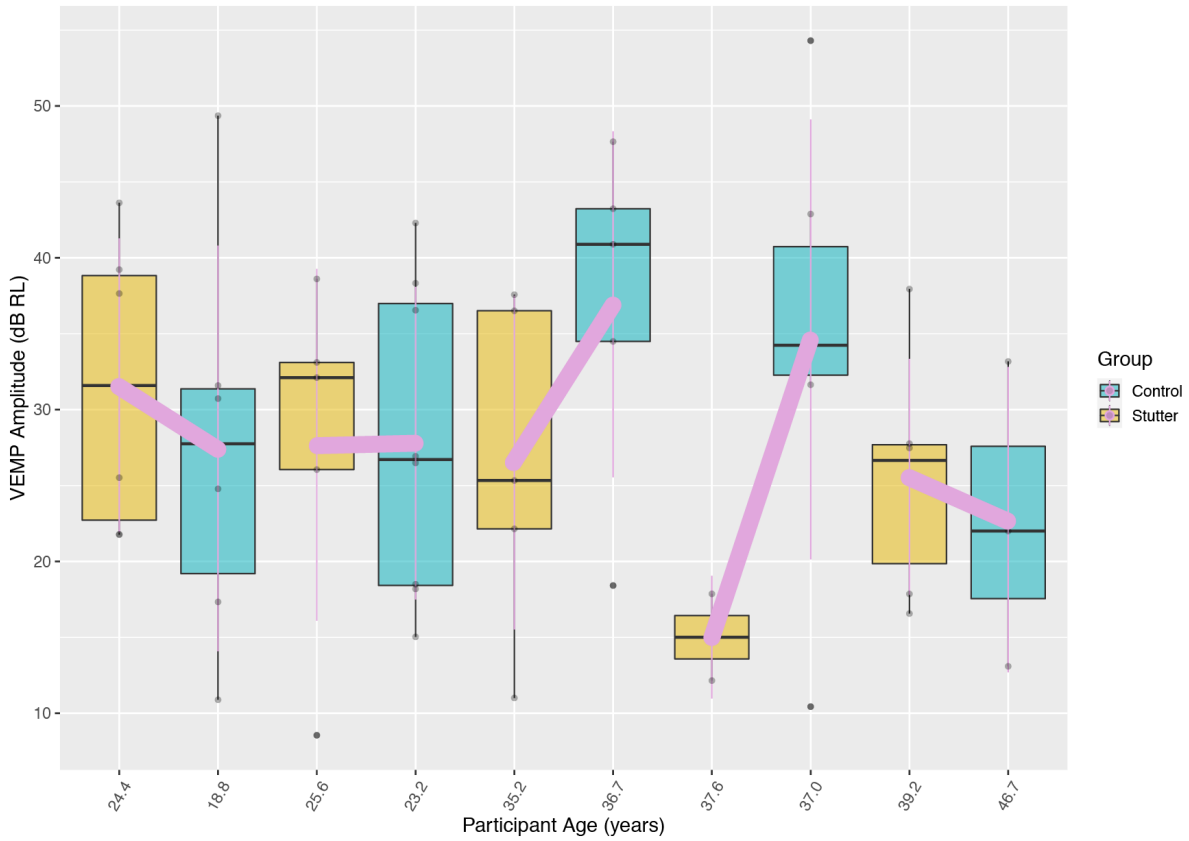

Figure 14 (supplementary): Box plots showing distribution of VEMP amplitudes per participant (across all participants in the pilot study), and participant age. Compare with Figure 2. Lines link the mean VEMP amplitudes of participants who stutter (PWS) with those of their paired controls. Pairing is less exact than for participants detailed in the main study (Appendix A), but age and sex matching is exact in aggregate (more detail in Appendix D).

Individual participant slopes are shown in figure 15. Due to limitations of the pilot study (e.g. absence of the head bar), it is unsurprising that the slopes are more variable than for data in the main study.

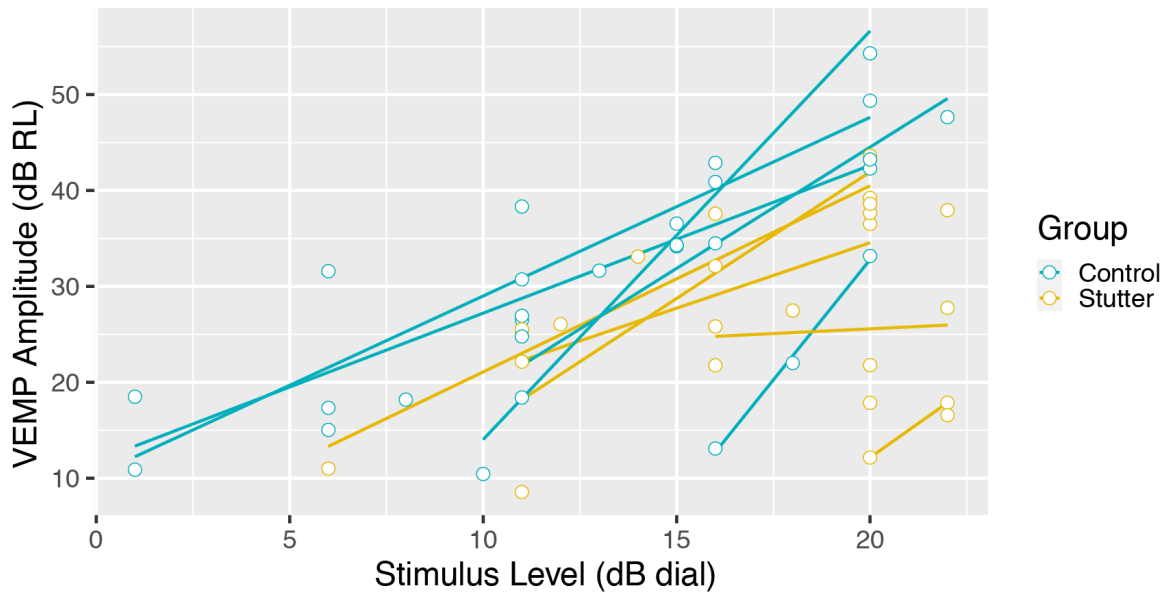

Figure 15: Per participant slopes of stimulus level (dB dial) versus VEMP amplitude (dB RL), showing participants who stutter (PWS) and paired controls. Compare with figure 3. Due to limitations of the pilot study slopes are more variable than the main study. Due to calibration differences stimulus levels are not comparable between main pilot studies.

The final model for the pilot study is shown in figure 16.

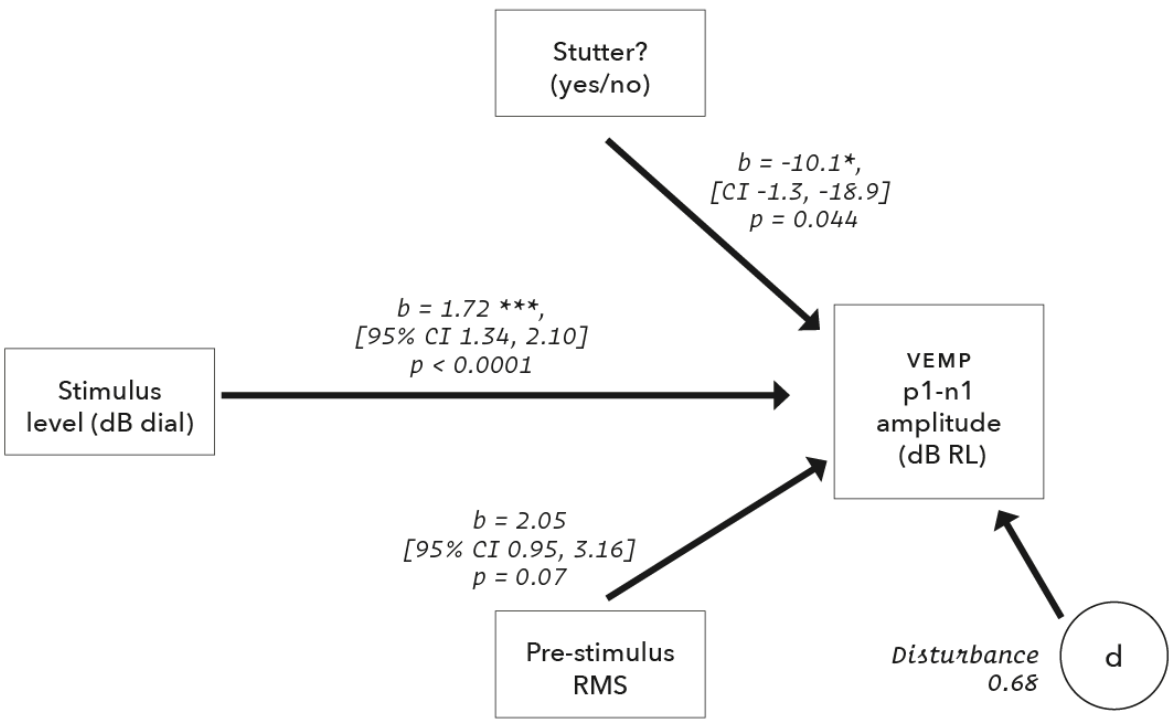

Figure 16: Final model for VEMP amplitude in the pilot study. Compare with figure 4.

No statistically significant group differences were found for peak to trough VEMP latency with pilot study data. Figure 17 shows latencies collected across all participants and all stimulus levels, including repeat measurements.

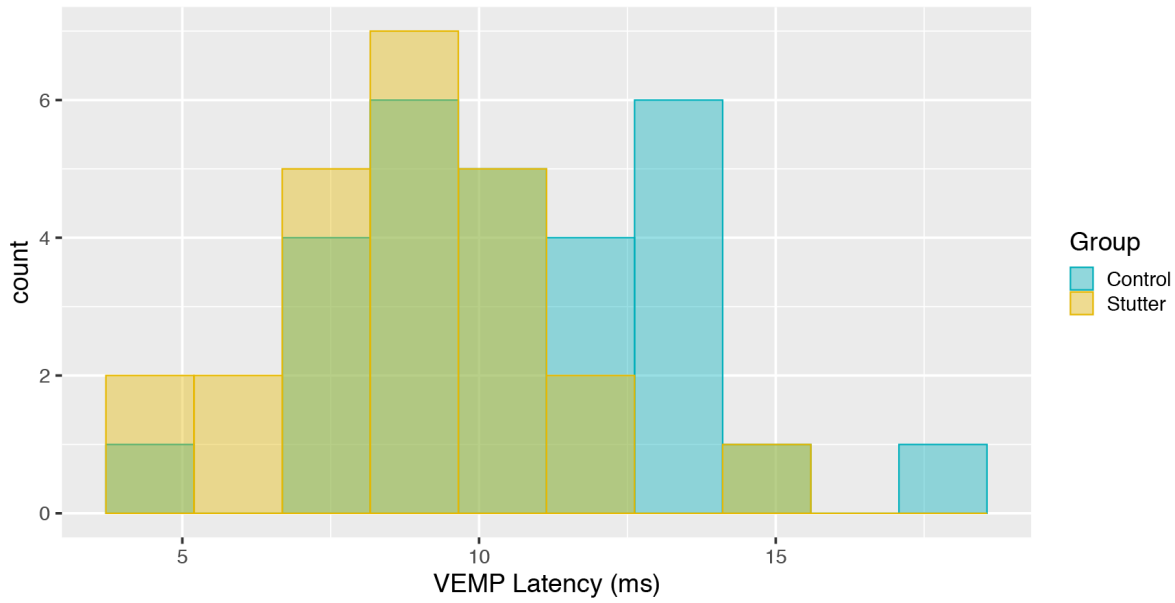

Figure 17: Histogram of VEMP peak to trough (p1-n1) latencies for the pilot study. The histogram contains repeated measurements for two groups of 5 participants per group. As such, it suggests shape of distribution and direction of group difference, but is not appropriate for statistical comparison (statistical comparison is by linear mixed modelling).

Data appear normally distributed, with no indication of a group difference. Variation across participants with stimulus level is shown in figure 18. There is no statistically significant interaction (Pearson's correlation coefficient between VEMP latency and stimulus level is $r(22) = -0.35$, $p = 0.09$, 95% CI [-0.66, 0.06] for PWS data; and $r(26) = -0.30$, $p = 0.12$, 95% CI [-0.61, 0.08] for control data).

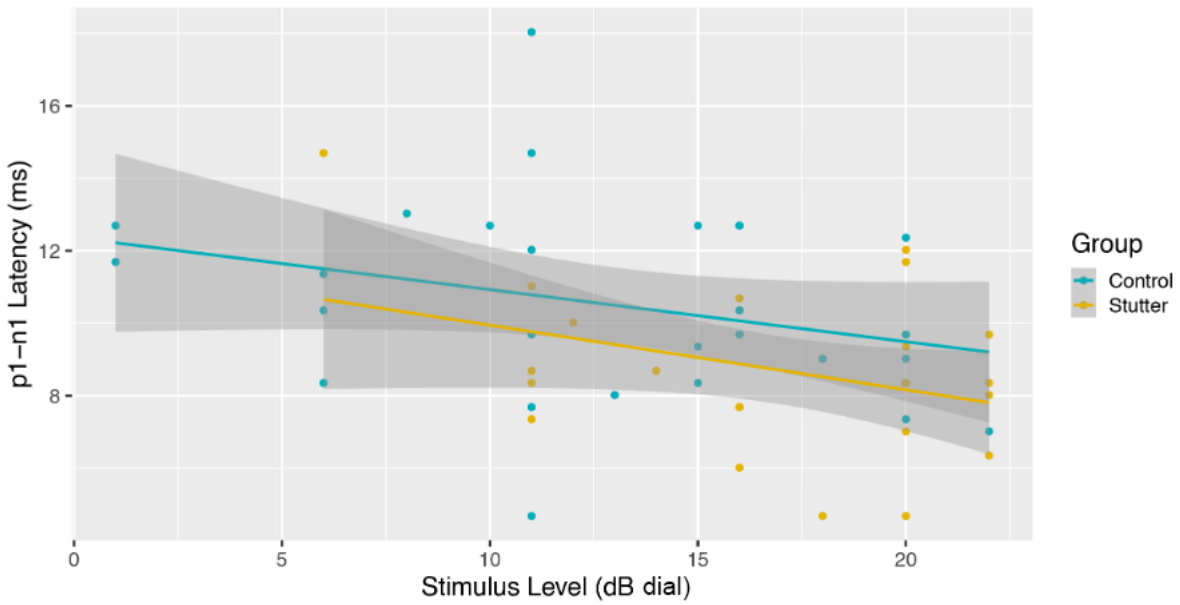

Figure 18: Variation of peak to trough latency with stimulus level for the pilot study.

Group comparison was carried out through linear mixed model analysis, with p-values generated by likelihood ratio comparisons between the following models:

```
xnull: logp1n1 ~ 1  + (1|PiD)
xmdl:  logp1n1 ~ 1 + isPWS + (1|PiD)
```

There is no statistically significant difference between groups (chi squared (1) 2.6, $p = 0.10$).

Appendix D: Participant details for Pilot Study

Participant recruitment for the pilot study was similar to that described in Appendix A, except that pairing on age and sex is only precise in aggregate (table 6). SSI-4 scores are shown in table 7.

PILOT STUDY						
Participant ID	Sex	SSI-4	Group	Age (yrs)	difference (yrs)	p1-n1 amp mismatch %
IDAM2209	Female	33	PWS	24.4		
REET0304	Male	6	CONT	18.8	-5.6	-0.78
SKOT9050	Male	19	PWS	25.6		
FIZZ2001	Male	6	CONT	23.2	-2.4	-0.34
AERO7835	Male	25	PWS	35.2		
PHON8215	Male	3	CONT	36.7	1.5	0.21
JAGG5728	Male	18	PWS	37.6		
TDRR3645	Female	5	CONT	37.0	-0.6	-0.09
TVOY8855	Male	28	PWS	39.2		
BEAG1273	Male	2	CONT	46.7	7.5	1.05
			mean (of positive/negative values)	32.4	-0.1 (SD 4.9)	0.01 %
			mean (of absolute values)		3.5 (SD 2.9)	0.69 %

Table 6: Pilot study, PWS shown adjacently to paired controls, with estimated error in VEMP amplitude due to pairing mismatch

Group	SSI-4 score and rating					
	Do not stutter (<10)	Very mild (10 – 17)	Mild (18 – 24)	Moderate (25 – 31)	Severe (32 – 36)	Very severe (37 – 46)
Controls	5	–	–	–	–	–
PWS	–	–	4	1	–	–

Table 7: Pilot study, SSI-4 scores for ordinarily fluent controls and participants who stutter.

Data collection for the pilot study featured 8 PWS and 8 controls, but three pairs were removed following inability to record a VEMP from one of the pair. Participants in these pairs had ages between 39 and 57 years. Inability to record VEMPs from these participants is likely to be due to VEMP amplitudes declining with age (Colebatch, Govender &

Rosengren, 2013) and the weaker stimuli used in the pilot versus the main study (see limitations described in Appendix C. VEMP p1-n1 amplitudes are proportional to stimulus level; Todd, Rosengren & Colebatch, 2008; Dennis, Govender & Colebatch, 2016).

Two PWS (ages 24.4 and 25.6 in figure 14) from the pilot study were retested, with different controls, for the main study. In the pilot study, VEMP amplitudes for the two PWS were similar to those of matched controls. When the two PWS were retested for the main study (ages 26.0 and 26.5 in figure 2) VEMP amplitudes were clearly lower than more accurately paired controls. Variation due background EMG tension is accounted for in the linear mixed model of the pilot study (figure 16), but is not accounted for the box plot of figure 14. Thus, comparison of box plots between the main study and pilot study will mislead. Statistical analysis using linear mixed models (figures 4 and 16) should be referred to instead.

Appendix E: Pilot study – Detailed statistical analysis of p1-n1 amplitude

Table 8 shows additional linear mixed models for the pilot data. The head bar was not available in the pilot study, so neck tension was generated via head raise from supine. Neck tension was expected to be more variable than in the main study (see discussion in section 4.2). To account for the variability, models were tested with pre-stimulus background RMS tension (which corresponds to neck tension) as a covariate.

Model	Comments	**PWS-CONT diff** (std err)	**p-value** [(df) χ^2]
Fixed slope logp1n1 ~ dB + isPWS + (1|PiD)		**-10.3** (5.0)	**0.060** [(1) 3.52]

Random slope logp1n1 ~ dB + isPWS + (1 + dB|PiD)	Singular fit	**-6.7** (2.9)	**0.040*** [(1) 4.2]
Fixed slope with RMS logp1n1 ~ dB + RMS + isPWS + (1|PiD)	Effect of RMS is 2.1 (std err 1.1)	**-10.1** (4.5)	**0.044*** [(1) 4.05]
Random slope with RMS logp1n1 ~ dB + RMS + isPWS + (1 + dB|PiD)	Convergence Error; effect of RMS is 2.2 (std err 1.1)	**-6.8** (2.7)	**0.035*** [(1) 4.43]

Table 8: Linear mixed model comparison for data from the pilot study. Variables as are described in table 5 for the main study, except for the addition of "RMS", which corresponds to neck tension. "RMS" is the normalised pre-stimulus background RMS EMG waveform for each sequence of 300 stimulus presentations (see section 4.6 for details). Data and R code are available for download with this appendix.

As anticipated, neck tension affected VEMP amplitude (b = 2.05, 95% [CI 0.95, 3.16], p = 0.07). The effect is large in comparison to data collected with a normative group of 48 (Gattie, Lieven & Kluk, in preparation) who used the head bar and had a pre-stimulus neck tension effect of b = 0.32 (95% CI [-0.07, 0.72], p = 0.11). For this reason, the model chosen has neck tension as an additional predictor. For consistency with the main study, a fixed slope model was chosen.

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

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.

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

Alexander, G.E., Newman, J.D. and Symmes, D., Convergence of

prefrontal and acoustic inputs upon neurons in the superior temporal gyrus of the awake squirrel monkey. Brain Res., 116 (1976) 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

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

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

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

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.

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

Bee, M. A., & Micheyl, C. (2008). The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? *Journal of Comparative Psychology*, *122*(3), 235–251. http://doi.org/10.1037/0735-7036.122.3.235

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

Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception & Psychophysics*, *64*(5), 844–854. http://doi.org/10.3758/bf03194750

Beving, H., Eriksson, L. E., Davey, C. L., & Kell, D. B. (1994). Dielectric properties of human blood and erythrocytes at radio frequencies (0.2-10 MHz); dependence on cell volume fraction and medium composition. *European Biophysics Journal : EBJ*, *23*(3), 207–215. http://doi.org/10.1007/BF01007612

Billig, A. J., Davis, M. H., Deeks, J. M., Monstrey, J., & Carlyon, R. P. (2013). Lexical influences on auditory streaming. *Current Biology : CB*, *23*(16), 1585–1589. http://doi.org/10.1016/j.cub.2013.06.042

Bloodstein, O. (1995) A Handbook on Stuttering (5th ed), Singular.

Bregman, A. S. (1990) Auditory scene analysis. MIT Press: Cambridge, MA

British Society of Audiology (2012), Recommended Procedure: Cervical and Ocular Vestibular Evoked Myogenic Potentials

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Atten Percept Psychophys*, *77*(5), 1465–1487. http://doi.org/10.3758/s13414-015-0882-9

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 105–

117. http://doi.org/10.1002/hbm.20140
Brown, T., Sambrooks, J. E., & MacCulloch, M. J. (1975). Auditory thresholds and the effect of reduced auditory feedback of suttering. *Acta Psychiatrica Scandinavica*, *51*(5), 297–311. http://doi.org/10.1111/j.1600-0447.1975.tb00009.x
Budde, K. S., Barron, D. S., & Fox, P. T. (2014). Stuttering, induced fluency, and natural fluency: a hierarchical series of activation likelihood estimation meta-analyses. Brain and Language, 139, 99–107. http://doi.org/10.1016/j.bandl.2014.10.002
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*.
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
Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*(10), 465–471. http://doi.org/10.1016/j.tics.2004.08.008
Chakrabarty, D., & Elhilali, M. (2019). A Gestalt inference model for auditory scene segregation. *PLoS Computational Biology*, *15*(1), e1006711. http://doi.org/10.1371/journal.pcbi.1006711
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
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., 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

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

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

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

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

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

Darwin, C. J. (2007). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences*. http://doi.org/10.1098/rstb.2007.2156

Denham, S. L., & Winkler, I. (2015). Auditory perceptual organization. In J. Wagemans (Ed.), *Oxford Handbook of Perceptual Organization* (pp. 601–620).

Doslak, M. J., Healey, E. C., & Riese, K. (1986). Eye movements of stutterers. *Investigative Ophthalmology & Visual Science*, *27*(9), 1410–1414.

Downes, S.M. and Matthews, L. (2020) “Heritability”, *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Zalta, E.N. (ed.), https://plato.stanford.edu/archives/spr2020/entries/heredity/

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

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

Fairbanks G. (1954) Systematic Research In Experimental Phonetics: 1. A Theory Of The Speech Mechanism As A Servosystem. Journal of Speech and Hearing Disorders; 19(2):133–139. https://doi.org/10.1044/jshd.1902.133 PMID: 13212816

Fenoy, A. J., Severson, M. A., Volkov, I. O., Brugge, J. F., & Howard, M. A. (2006). Hearing suppression induced by electrical stimulation of human auditory cortex. *Brain Research*, *1118*(1), 75–83. http://doi.org/10.1016/j.brainres.2006.08.013

Florentine M., Popper A., Fay R. (eds) Loudness. Springer Handbook of Auditory Research, vol 37. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-6712-1_5

Ford, J. M., & Mathalon, D. H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research*, *38*(1), 37–46. http://doi.org/10.1016/s0022-3956(03)00095-5

Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., . . . McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. Cerebral Cortex, 16(7), 969–977. https://doi.org/10.1093/cercor/bhj039

Goldberg, J.M. (ed) (2012), The vestibular system: a sixth sense. (2012). Oxford University Press Inc

Golding, N. L., & Oertel, D. (2012). Synaptic integration in dendrites: exceptional need for speed. *The Journal of Physiology*, *590*(22), 5563–5569. http://doi.org/10.1113/jphysiol.2012.229328

Grant, W., & Curthoys, I. (2017). Otoliths - Accelerometer and seismometer; Implications in Vestibular Evoked Myogenic Potential

(VEMP). *Hearing Research*, *353*, 26–35. 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

Godfrey-Smith, P. and Sterelny, K. (2016) "Biological Information", *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), Zalta, E.N. (ed.), https://plato.stanford.edu/archives/sum2016/entries/information-biological/

Griffiths, P., "The Distinction Between Innate and Acquired Characteristics", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Zalta, E.N. (ed.), <https://plato.stanford.edu/archives/spr2020/entries/innate-acquired/>.

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.

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

Haykin, S., & Chen, Z. (2005). The cocktail party problem. *Neural Computation*, *17*(9), 1875–1902. http://doi.org/10.1162/0899766054322964

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

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., 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.

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

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.

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

Jeffress, L.A. (1948): A place theory of sound localization. J Comp Physiol Psychol, 41:35-39.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, *24*(10), 1995–2004. http://doi.org/10.1177/0956797613482467

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

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

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

Kramer, M. B., Green, D., & Guitar, B. (1987). A comparison of stutterers and nonstutterers on masking level differences and synthetic sentence identification tasks. *Journal of Communication Disorders*, *20*(5), 379–390.

Langová, J., Morávek, M., Široký & Sváb, L. (1975) "Effect of speaking on evoked vestibular nystagmus in stutterers" *Folia Phoniatrica*, 27(4), 287–291.

Laukli, E., & Burkard, R. (2015). Calibration/Standardization of Short-Duration Stimuli. *Seminars in Hearing*, *36*(1), 3–10. http://doi.org/10.1055/s-0034-1396923

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38– discussion 38–75.

Lewicki, M. S., Olshausen, B. A., Surlykke, A., & Moss, C. F. (2014). Scene analysis in the natural environment. *Frontiers in Psychology*, *5*. http://doi.org/10.3389/fpsyg.2014.00199

Liotti, M., Ingham, J. C., Takai, O., Paskos, D. K., Perez, R., & Ingham, R. J. (2010). Spatiotemporal dynamics of speech sound perception in chronic developmental stuttering. *Brain and Language*, *115*(2), 141–147. http://doi.org/10.1016/j.bandl.2010.07.007

Lipton, P. (2004). *Inference to the Best Explanation*, (2nd ed.), London: Routledge.

MacCulloch, M. J., & Eaton, R. (1971). A note on reduced auditory pain threshold in 44 stuttering children. *The British Journal of Disorders of Communication*, *6*(2), 148–153. http://doi.org/10.3109/13682827109011541

Maffei, C., Sarubbo, S., & Jovicich, J. (2019). A Missing Connection: A

Review of the Macrostructural Anatomy and Tractography of the Acoustic Radiation. *Frontiers in Neuroanatomy*, *13*. http://doi.org/10.3389/fnana.2019.00027

Manley, G. A., Gummer, A., Popper, A.N. & Fay, R.R. (eds) (2017). Understanding the Cochlea. Springer Science & Business Media.

Matthews, N., Todd, J., Mannion, D. J., Finnigan, S., Catts, S., & Michie, P. T. (2013). Schizophrenia Research. *Schizophrenia Research*, *146*(1-3), 344–348. http://doi.org/10.1016/j.schres.2013.02.013

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. http://doi.org/10.1016/j.jml.2017.01.001

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., Ghash 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. CICSD 31:105–122.

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. Brain: A Journal of Neurology, 119(Pt. 3), 907–917.

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

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

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

Müller-Preuss, P. & Ploog, D. (1981). Inhibition of auditory cortical neurons during phonation. *Brain Research*, *215*(1-2), 61–76. http://doi.org/10.1016/0006-8993(81)90491-1

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

Młynarski, W., & McDermott, J. H. (2019). Ecological origins of perceptual grouping principles in the auditory system. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(50), 25355–25364. http://doi.org/10.1073/pnas.1903887116

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.

Numminen,J.,Curio,G.,Neuloh,G.,Jousmüki,V & Hari, R., Speaking modifies utterance-related activity of the human auditory cortex. NeuroImage, 7 (1998) S48.

Numminen, J. & Curio, G. (1999) Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. Neurosci Lett. 1999; 272:29–32. [PubMed: 10507535]

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

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

O'Callaghan, C. (2015). Speech perception. in M. Matthen (ed.), *Oxford Handbook of Philosophy of Perception*, Oxford: Oxford University

Press, pp. 475–494.
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
Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. NeuroImage, 61(1), 314–322. https://doi.org/10.1016/j.neuroimage.2012.02.068
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
Parrell, B., & Houde, J. (2019). Modeling the Role of Sensory Feedback in Speech Motor Control and Learning. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 2963–2985. http://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127
Pitt, David (2020) "Mental Representation", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Zalta, E.N. (ed.), https://plato.stanford.edu/archives/spr2020/entries/mental-representation/
Pollack, G.S., Mason, A.C., Popper, A.N. & Fay, R.R. (eds) (2019) Insect Hearing. Springer Handbook of Auditory Research, vol 55. Springer, New York, NY http://doi.org/10.1016/S0079-6123(08)00673-0
Pörschmann, C. (2000). “Influences of bone conduction and air conduction on the sound of one’s own voice,” Acta Acust. 86, 1038–1045.
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

Riley, G. D. (2009). SSI-4: Stuttering severity instrument - Fourth edition; examiners' manual (4th ed.). Austin, TX: Pro-Ed.

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.

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

Sato, M., & Shiller, D. M. (2018). Auditory prediction during speaking and listening. *Brain and Language*, *187*, 92–103. http://doi.org/10.1016/j.bandl.2018.01.008

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

Shamma S. A., Micheyl C. (2010). Behind the scenes of auditory perception. Curr. Opin. Neurobiol. 20, 361–36610.1016/j.conb.2010.03.009

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

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin*, *133*(5), 780–799. http://doi.org/10.1037/0033-2909.133.5.780

Snyder, J. S., & Elhilali, M. (2017). Recent advances in exploring the neural underpinnings of auditory scene perception. *Annals of the New York Academy of Sciences*, *1396*(1), 39–55. http://doi.org/10.1111/nyas.13317

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

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266. http://doi.org/10.1038/nrn2331

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.

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

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

Szabó, B. T., Denham, S. L., & Winkler, I. (2016). Computational Models of Auditory Scene Analysis: A Review. *Frontiers in Neuroscience*, *10*, 524. http://doi.org/10.3389/fnins.2016.00524

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

Tait, J. (1932). Is All Hearing Cochlear? *Annals of Otology, Rhinology & Laryngology*, *41*(3), 681–704. http://doi.org/10.1177/000348943204100302

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.

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

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. http://doi.org/10.1080/01690960903498424

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. NeuroImage, 39(3), 1429–1443. https://doi.org/10.1016/j.neuroimage.2007.09.054

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

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 Gulick, R. (2018) "Consciousness", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Zalta, E.N. (ed.), https://plato.stanford.edu/archives/spr2018/entries/consciousness/

von Békésy, G. (1949). “The structure of the middle ear and the hearing

of one's own voice by bone conduction," J. Acoust. Soc. Am. **21**, 217–232.

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

Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). "The precedence effect in sound localization," The American Journal of Psychology, 62, 315–336.

Ward, D. (2010). Sudden onset stuttering in an adult: Neurogenic and psychogenic perspectives. *Journal of Neurolinguistics*.

Warren, R. (1973). Quantification of Loudness. *The American Journal of Psychology, 86*(4), 807-825. doi:10.2307/1422087

Waters, F. & Fernyhough, C. (2019). Auditory Hallucinations: Does a Continuum of Severity Entail Continuity in Mechanism? *Schizophrenia Bulletin* 45(4): 717-719.

Weintraub, D. M., Ramage, E. M., Sutton, G., Ringdahl, E., Boren, A., Pasinski, A. C., et al. (2012). Auditory stream segregation impairments in schizophrenia. *Psychophysiology*, *49*(10), 1372–1383. http://doi.org/10.1111/j.1469-8986.2012.01457.x

Welgampola, M. S., Rosengren, S. M., Halmagyi, G. M., & Colebatch, J. G. (2003). Vestibular activation by bone conducted sound. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(6), 771–778. http://doi.org/10.1136/jnnp.74.6.771

Wilson, R.A. and Foglia, L. "Embodied Cognition", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Zalta. E.N. (ed.), https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. http://doi.org/10.1016/j.tics.2009.09.003

Winter (2019) Statistics for Linguists: An Introduction Using R, Routledge. ISBN 9781138056091

Woods, K. J. P., & McDermott, J. H. (2018). Schema learning for the cocktail party problem. *Proceedings of the National Academy of*

Sciences of the United States of America, *115*(14), E3313–E3322. http://doi.org/10.1073/pnas.1801614115

Yadav, M., Cabrera, D., & Martens, W. L. (2012). A system for simulating room acoustical environments for one's own voice. *Applied Acoustics*, *73*(4), 409–414. http://doi.org/10.1016/j.apacoust.2011.10.001

Yates, A. J. (1963). Recent Empirical and Theoretical Approaches to the Experimental Manipulation of Speech in Normal Subjects and in Stammerers. *Behaviour Research and Therapy*, *1*(2-4), 95–119. http://doi.org/10.1016/0005-7967(63)90013-5

Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. Neuropsychologia, 48(2), 607–618. https://doi.org/10.1016/j.neuropsychologia.2009.10.025

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. NeuroImage, 40(4), 1871–1887. https://doi.org/10.1016/j.neuroimage.2008.01.026

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2009). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. Journal of Cognitive Neuroscience, 22(8), 1770–1781. https://doi.org/10.1162/jocn.2009.21324

Zwergal, A., Strupp, M., Brandt, T., & Büttner-Ennever, J. A. (2009). Parallel ascending vestibular pathways: anatomical localization and functional specialization. *Annals of the New York Academy of Sciences*, *1164*(1), 51–59. http://doi.org/10.1111/j.1749-6632.2009.04461.x