The role of the ventral MGB in speech in noise comprehension

Paul Glad Mihai[1,2], Nadja Tschentscher[3], Katharina von Kriegstein[1]

[1]Chair of Cognitive and Clinical Neuroscience, Faculty of Psychology, Technische Universität Dresden, Germany

[2]Max Planck Institute for Cognitive and Brain Sciences, Leipzig, Germany

[3]Biologische Psychologie, Ludwig-Maximilians-Universität München, Germany

Abstract

Comprehending speech in noise is a difficult daily activity, yet humans master it with the help of a complex neuronal decoding system spread over the cortical hierarchy. Sensory thalami are central sensory pathway stations on this hierarchical ladder. Recent studies have shown that the left ventral auditory thalamus (ventral medial geniculate body: vMGB) response is modulated by speech recognition tasks when associated with speech recognition abilities. In the context of predictive coding, the vMGB should be more involved when comprehending speech in noise. Here we tested the specific hypothesis that this behaviourally relevant modulation is enhanced when listening in noisy conditions. We found a top-down modulation for the left vMGB when attending a speech comprehension vs. a speaker identity task. This top-down modulation is further enhanced when listening conditions were suboptimally noisy. These results imply that modulation of thalamic driving input to the auditory cortex facilitates speech recognition in noisy conditions.

Author contributions: PGM: collected data, analyzed data, interpreted results, wrote manuscript, corrected manuscript. NT: conceptualized experiment, programmed experiment, corrected manuscript. KvK: conceptualized experiment, interpreted data, wrote manuscript, corrected manuscript.

Introduction

Honking horns and roaring engines, the hammering from a construction site, the mix of music and speech at a restaurant or pub, the chit-chat of many children in a classroom are just some examples of background noises which constantly accompany us. Nevertheless, humans have a remarkable ability to hear and understand speech of a conversation partner even under these difficult listening conditions (Cherry 1953).

Understanding speech in noise is a complex task that involves both sensory and cognitive processes (Bronkhorst 2015; Best et al. 2007; Shinn-Cunningham and Best 2008; Brokx and Nooteboom 1982; Moore, Peters, and Glasberg 1985; Bregman and McAdams 1994; Darwin and Hukin 2000; Parikh and Loizou 2005; Sayles and Winter 2008). Difficulties in understanding speech in noise can occur in age-related hearing impairment, as well as in developmental disorders like autism spectrum disorder or developmental dyslexia (Alcantara et al. 2004; Ziegler et al. 2009; Chandrasekaran et al. 2009; Wong et al. 2009; Schelinski and von Kriegstein 2019; Schoof and Rosen 2016). In contrast, early musical training is associated with better abilities in extracting speech from a noisy background (Strait et al. 2012; Parbery-Clark et al. 2009).

Sound passing from the cochlea to the auditory cortex is intricately processed such that key features are extracted and irrelevancies suppressed (Anderson and Kraus 2010). From Wernicke's research (Wernicke 1874) all the way up to current findings, neuroscientific models of speech recognition have mostly focused on cerebral cortex mechanisms (Hickok and Small 2015; Friederici and Gierhan 2013). Recent findings, however, suggest that a full understanding of speech recognition mechanisms is impossible without understanding cerebral cortex feedback mechanisms to the subcortical sensory pathways (von Kriegstein, Patterson, and Griffiths 2008; Diaz et al. 2012; Díaz, Blank, and von Kriegstein 2018; Mihai et al. 2019; Chandrasekaran et al. 2009; Chandrasekaran, Kraus, and Wong 2011), in

particular, to the sensory thalami (von Kriegstein,et al. 2008; Diaz et al. 2012; Díaz et al. 2018; Mihai et al. 2019).

The classic view, that the sensory thalamus is a passive relay station (Squire et al. 2012) has been by-and-large abandoned over the last two decades. It has become accepted as a structure that is modulated by cognitive demands (Saalmann and Kastner 2015; Saalmann and Kastner 2011; Haynes, Deichmann, and Rees 2005; Antunes and Malmierca 2011; OConnor et al. 2002; McAlonan, Cavanaugh, and Wurtz 2008; Diaz et al. 2012; Mihai et al. 2019; von Kriegstein, Patterson, and Griffiths 2008; Díaz, Blank, and von Kriegstein 2018). For speech processing, previous studies have shown a task-dependent modulation in the MGB for auditory speech recognition (von Kriegstein, Patterson, and Griffiths 2008; Diaz et al. 2012), and in the visual sensory thalamus (lateral geniculate body: LGN) for visual speech recognition (Díaz, Blank, and von Kriegstein 2018). A task-dependent modulation for speech recognition means that sensory thalamus responses were higher for speech tasks (that emphasized recognition of fast-varying speech properties) in contrast to control tasks (that required recognition of relatively constant properties of the speech signal, such as the speaker identity or the sound intensity level). The studies also found that the performance level in auditory speech recognition was positively correlated with the task-dependent modulation in the MGB of the left hemisphere (von Kriegstein, Patterson, and Griffiths 2008; Mihai et al. 2019).

Sensory thalami in itself are complexly organized structures. For example, the MGB consists of three divisions (Winer 1984). Only the ventral MGB (vMGB) can be considered first-order sensory thalamus (Malmierca, Anderson, and Antunes 2015; Winer et al. 2005), as the vMGB receives driving inputs from sources that relay information from the sensory periphery and projects this information to the cerebral cortex (Sherman and Guillery 1998). The ventral MGB also receives modulatory input from the cerebral cortex (Sherman and Guillery 1998). Technical advances have only recently provided such high spatial resolution that allows research on the function of these structures in humans in vivo (Denison et al. 2014; Moerel et al. 2015, Mihai et al 2019). Using these techniques, we previously showed

that the behaviourally relevant task-dependent modulation in the left MGB for speech is located in the vMGB (Mihai et al. 2019), i.e., the first-order sensory thalamus.

The Bayesian brain hypothesis assumes that the brain represents information probabilistically and uses an internal generative model and predictive coding for most effective processing of sensory input (Knill and Pouget 2004; Friston and Kiebel 2009; Friston 2005; Kiebel, Daunizeau, and Friston 2008). Such type of processing has the potential to explain why the human brain is robust to sensory uncertainty, e.g., the ability to recognize speech despite noise in the speech signal (Knill and Pouget 2004; Srinivasan et al. 1982). Although predictive coding is most often discussed in the context of cerebral cortex organization (Hesselmann et al. 2010; Shipp et al. 2013), it might also be a governing principle of the interactions between cerebral cortex and subcortical sensory pathway structures (von Kriegstein et al., 2008; Huang and Rao 2011; Seth and Friston 2016; Bastos et al 2012; Mumford 1992; Adams et al. 2013). In accordance with this suggestion, studies in animals found that feedback from cerebral cortex areas changes the processing in the sensory pathway, i.e., the sensory thalamus and brainstem nuclei (Krupa, Ghazanfar, and Nicolelis 1999; Sillito, Cudeiro, and Jones 2006; Wang et al. 2018). One possible explanation for the task-dependent modulation of the sensory thalamus for speech in humans is that cerebral cortex areas provide predictions about the incoming sensory input to the sensory thalamus , and that this is particularly relevant for fast-varying and predictable stimuli such as speech (von Kriegstein, Patterson, and Griffiths 2008; Diaz et al. 2012). Such a view entails the specific hypothesis that the task-dependent modulation of subcortical sensory processing is especially important when there is high sensory uncertainty (Diaz et al. 2012; Yu and Dayan 2005; Feldman and Friston 2010; Van de Cruys 2014), as for example in people with perceptual difficulties (Diaz et al. 2012; Van de Cruys et al. 2014) or when the incoming signal is disturbed (Yu and Dayan 2005; Friston and Kiebel 2009; Feldman and Friston 2010; Gordon et al. 2017).

An ecologically valid way to increase uncertainty about the speech input is the presentation of speech in background noise (Chandrasekaran and Kraus 2010). A predictive mechanism

at the processing stages of the sensory pathways could explain why most of us can understand speech in noise with relative ease. We therefore tested, whether the task-dependent modulation of the left vMGB for speech is higher when the speech stimuli are heard in a noisy as opposed to a clear background. To do this, we used ultra-high field fMRI at 7 T and a design that has been shown to elicit task-dependent modulation of the vMGB in previous studies (von Kriegstein et al. 2008; Diaz et al. 2012; Mihai et al. 2019). We complemented the design by a noise factor, i.e., the speech stimuli were presented with and without background noise. The design was therefore a 2 × 2 factorial design with the factors task (speech task, speaker task) and noise (noise, clear). To test our hypothesis, we performed a task × noise interaction analysis. We expected that the task-dependent top-down modulation of the left vMGB increases with decreasing signal-to-noise ratios (i.e., increasing uncertainty about the speech sounds).

Results

Participants listened to blocks of auditory syllables (e.g., /ada/, spoken by three different synthesized speakers) and performed either a speech or a speaker task (Figure 1). In the speech task, participants reported via button press whether the current syllable was different from the previous one (1-back task). In the speaker task, participants reported via button press whether the current speaker was different from the previous one. Stimuli were presented embedded in speech-shaped noise (Noise condition) or without background noise (Clear condition).

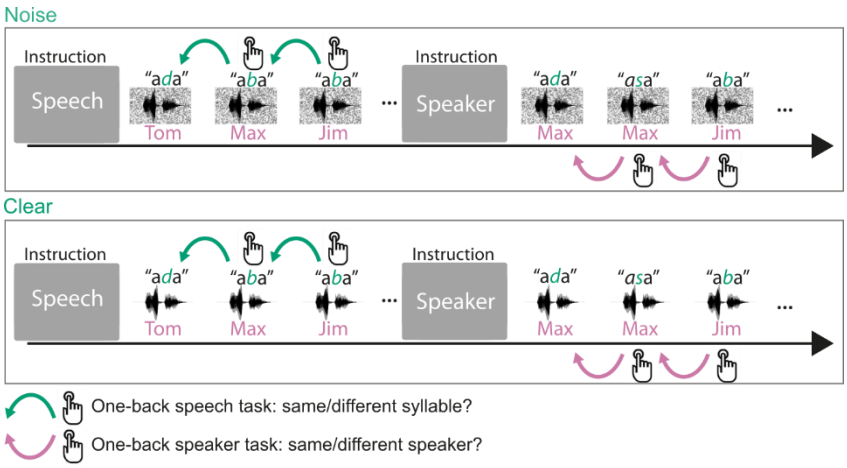

Figure 1. Design and trial structure of the experiment. In the speech task, listeners performed a one-back syllable task. They pressed a button whenever there was a change in syllable in contrast to the immediately preceding one, independent of speaker change. The speaker task used exactly the same stimulus material and trial structure. The task was to press a button when there was a change in speaker identity in contrast to the immediately preceding one, independent of syllable change. An initial task instruction screen informed participants about which task to perform. Participants heard stimuli either with concomitant speech-shaped noise (Noise condition) or without background noise (Clear condition).

Behavioural results

Participants performed well above chance level in all four conditions (> 82% correct; Table 1; Figure 2A).

Table 1. Proportion of hits for each of the four conditions in the experiment. HDP: highest posterior density interval.

	Speech task/ Noise	**Speaker task/ Noise**	**Speech task/ Clear**	**Speaker task/ Clear**
% Mean [95% HPD]	0.82 [0.62, 0.95]	0.87 [0.74, 0.96]	0.92 [0.83, 0.98]	0.90 [0.81, 0.97]

Performing the tasks with background noise was more difficult than the condition without background noise for both the speech and the speaker task (Figure 2B, for details on statistics see figure legend). The rate of hits in the Speech task was the same as in the

Speaker task (Figure 2C). There was a detectable interaction between task and noise (Figure 2D/E), but simple main effects (i.e., Speech task/Noise vs. Speaker task/Noise (Figure 2F) and Speech task/Clear vs. Speaker task/Clear (Figure 2G)) were undetectable.

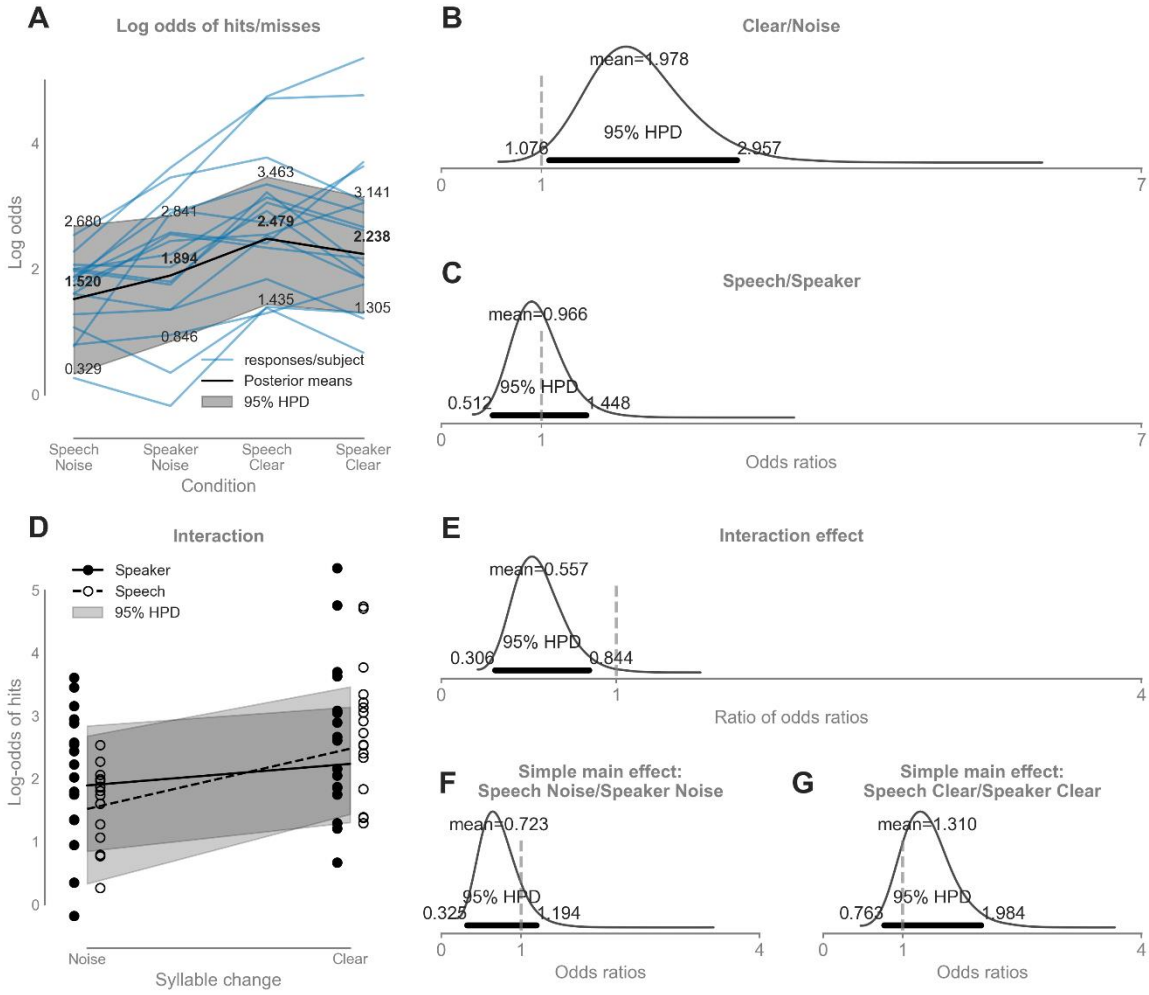

Figure 2. Behavioural results. We performed a binomial logistic regression to compute the rate of hits and misses in each condition, because behavioural data were binomially distributed. For this reason, results are reported in log odds and odds ratios. An odds ratio of 1 means that a difference between the two conditions is lacking (50/50 chance, or a 1:1 ratio; Chen 2003). A detectable difference between conditions is present, if the highest posterior density interval (HPD) of the posterior probability distributions of the odds ratios excludes 1 (McElreath 2016; Bunce and McElreath 2017). A. Mean proportion of correct button presses for each condition (from left to right: Speech Noise, Speaker Noise, Speech Clear, Speaker Clear). The blue lines indicate responses for individual participants, the black line and bold font numbers denote the

posterior mean per condition, and the grey area and normal font numbers demarcate the 95% HPD. The rate of hits compared to misses is plotted on a log scale to allow a linear representation. B. The difference in the rate of hits between the Clear and Noise condition was on average twice as high (1.978 [1.076, 2.957]) and the HPD excluded 1 indicating a clear difference between conditions. C. The rate of hits in the Speech compared to Speaker condition had a mean of ~1 indicating no difference between these conditions. D. Visualization of the interaction effect as a comparison of slopes with 95% HDP. E. Posterior distribution of interaction effect with mean and 95% HDP of 0.557 [0.306, 0.844]. The HPD excluded 1 indicating a difference between odds ratios. F. The rate of hits in the Speech Noise condition was on average ~1/3 lower than the rate of hits in the Speaker Noise condition; however, the HPD strongly overlaps 1 indicating that there was no clear difference between conditions. G. The rate of hits in the Speech Clear condition was on average ~1/3 higher than the rate of hits in the Speaker Clear condition; however, the HPD strongly overlaps 1 indicating that there was no clear difference between conditions.

fMRI Results

We extracted parameter estimates for each condition and participant after first level estimation from an independently localized left vMGB (Mihai et al. 2019). Using Bayesian multi-level linear regression on the extracted parameter estimates, we sampled posterior distributions for each parameter in the linear model (for details refer to Materials and Methods). Posterior distributions update the prior information specified in the model with the observed data (McElreath, 2012). Posterior distributions, in comparison to point estimates, have the advantage of quantifying uncertainty about each parameter. We summarized each posterior distribution using the mean as a point estimate (posterior mean) together with a 95% highest posterior density interval (HPD). The HPD is the probability that the mean lies within the interval (McElreath, 2016; Gelman et al. 2013), e.g., we are 95% sure the mean lies within the specified interval bounds. Differences between conditions were converted to effect sizes (Hedges g* (Hedges and Olkin 1985)). Hedges g*,

like Cohen's d (Cohen 1988), is a population parameter that computes the difference in means between two variables normalized by the pooled standard deviation with the benefit of correcting for small sample sizes. Based on Cohen (1988), we interpreted effect sizes on a spectrum ranging from small ($g^* \approx 0.2$), to medium ($g^* \approx 0.5$), to large ($g^* \approx 0.8$), and beyond. If the HPD did not overlap zero we considered this to be a robust effect (McElreath 2016; Bunce and McElreath 2017). However, we caution readers that if the HPD includes zero it does not mean that the effect is missing (Amrhein et al., 2019). Instead, we choose to quantify and interpret the magnitude (by the point estimate) and its uncertainty (by the HPD) provided by the data and our assumptions (Anderson 2019).

The task-dependent modulation of left vMGB increases for recognizing speech in noise

In accordance with our hypothesis there was increased BOLD response for the task × noise interaction [(Speech task/Noise > Speaker task/Noise) > (Speech task/Clear > Speaker task/Clear)] in the left vMGB. The interaction effect had a mean large effect size ranging from a small effect to a very large effect (g^*=2.549 [0.211, 5.066]; Figure 3B, 3C, and 3D). The 95% HDP of the interaction effect did not include 0. Simple main effect analyses showed that the direction of the interaction was as expected. The Speech task/Noise condition yielded higher responses in contrast to the Speaker task/Noise condition, ranging from a medium to a very large effect (g^* = 1.104 [0.407, 1.798]; Figure 3E). Conversely, there was no vMGB response difference between the Speech task and Speaker task in the clear condition as the HPD overlapped 0 (g^* = 0.243 [-0.366, 0.854]; Figure 3F), ranging from a negative medium effect to a positive large effect.

The results show that the task-dependent modulation of the left vMGB for speech is increased when participants recognize speech with background noise. This cannot be explained by differences in stimulus input as the same stimulus material was used for the speech and the speaker task.

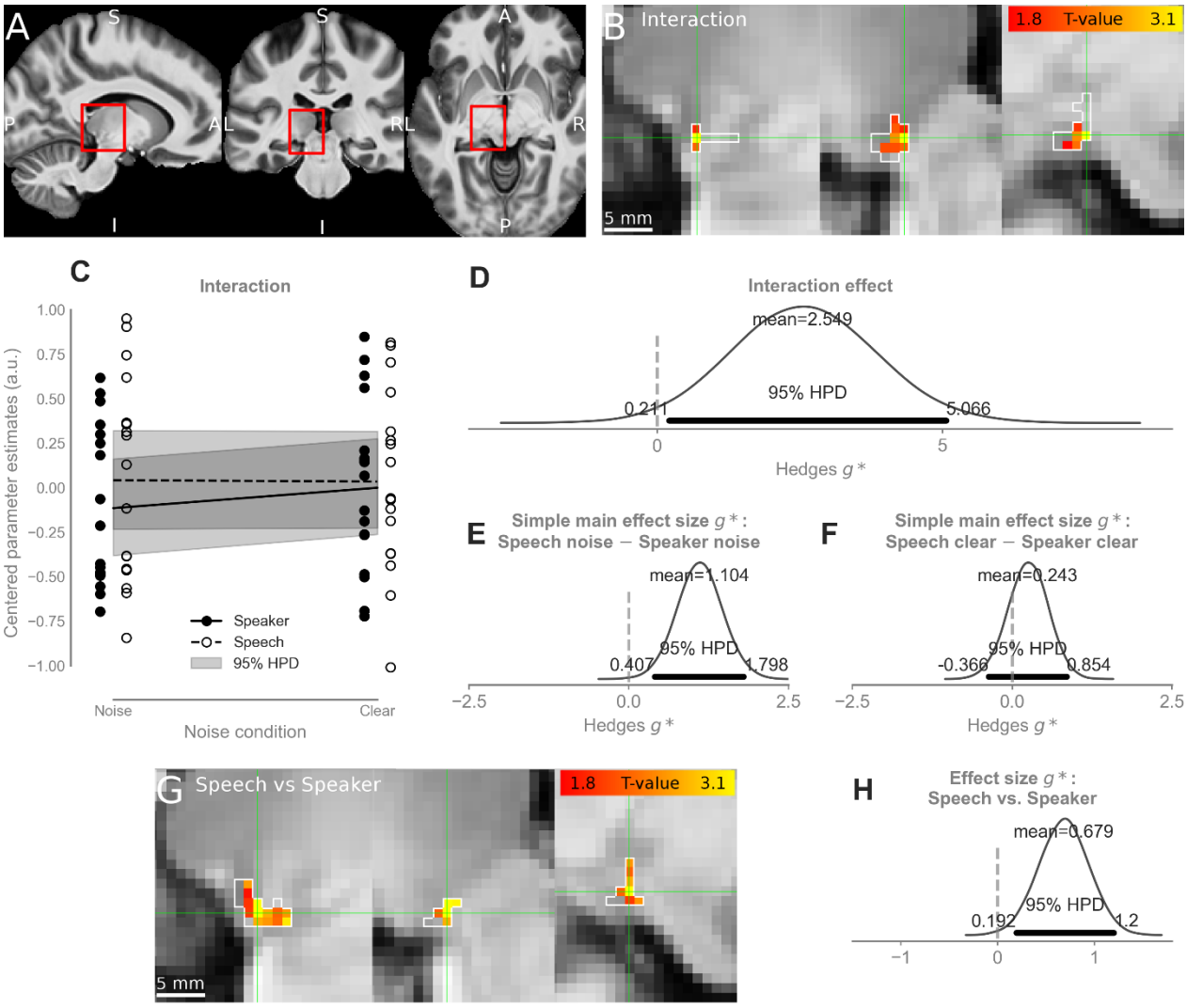

Figure 3. fMRI results. A. The template image in MNI space. Red rectangles denote the approximate location of the left MGB and encompass the zoomed in views in B and G. Letters indicate anatomical terms of location: A, anterior; P, posterior; S, superior; I, inferior; L, left; R, right. Figures A, B, and G share the same orientation across columns; i.e., from left to right: sagittal, coronal, and axial. . B. Interaction (yellow-red colour code): (Speech noise - Speaker noise) - (Speech clear - Speaker clear). White outline shows the boundary of the vMGB mask. C. Centered parameter estimates within the vMGB mask. Open circles denote parameter estimates for the Speech condition; filled circles denote parameter estimates in the Speaker condition. Dashed black line: relationship between noise condition (Noise, Clear) and parameter estimates in the Speech task. Solid black line: relationship between noise condition (Noise, Clear) and parameter estimates in the Speaker task. Shaded grey area shows 95% HPD. D-F Bayesian Analysis of the parameter estimates. D. Interaction: The effect size for the interaction effect was very large (2.549 [0.211, 5.066]) and the HPD excludes zero (indicated by the dashed vertical line). E. Simple main effect: Speech Noise vs. Speaker Noise. The mean effect size was large (1.104 [0.407, 1.798]). The HPD excludes zero. F. Simple main effect: Speech Clear vs. Speaker Clear. The mean effect size was small (0.243 [-0.366, 0.854]). The HPD contained zero. G. The main effect of task (yellow-red colour code): Speech - Speaker.

White outline shows the boundary of the vMGB mask. H. Bayesian Analysis of the parameter estimates for the main effect of task. There was a large effect size between the Speech and Speaker task of 0.679 [0.192, 1.200]. The HPD excluded zero.

Replication of previous findings

In addition to addressing the main hypothesis of the present paper, the data also allowed testing for replication of previous findings. Consistent with previous reports (kvk et al., 2008; Diaz et al., 2012; Mihai et al., 2019) there was a large positive main effect for the Speech vs. Speaker task ranging from a small to a very large effect (g* 0.679 [0.192, 1.200]; Figure 3G & 3H).

In contrast, there was no correlation between the task-dependent modulation (i.e., parameter estimates for the contrast of Speech vs. Speaker) and the correct proportion of hits in the Speech task; the effect size was very small (mean Pearson's r = 0.134 [-0.321, 0.603]; Figure S1) and the HPD included 0. A correlation between task-dependent modulation and speech task performance has been reported in three previous experiments (Exp 1 and 2 of von Kriegstein et al., 2008; Mihai et al., 2019), but was also not present in Diaz et al. 2012.

Discussion

In this study, we showed that the task-dependent modulation of the left hemispheric sensory thalamus (vMGB) for speech is particularly strong when recognizing speech in noisy listening conditions in contrast to conditions where the speech signal is clear. This finding confirms our a-priori hypothesis and is in agreement with explaining task-dependent sensory thalamus modulation within a Bayesian brain framework. Additionally to confirming our main hypothesis, we replicated findings from previous experiments (von Kriegstein, Patterson, and Griffiths 2008; Díaz et al. 2012) that showed task-dependent modulation in the MGB for speech, and furthermore, localized the task-dependent modulation to the vMGB (Mihai et al., 2019).

Studies investigating speech in noise recognition have found a cerebral cortex network that includes areas pertaining to linguistic, attentional, working memory, motor planning, and cognitive processing (Salvi et al. 2002; Scott et al. 2004; Wong et al. 2008; Bishop and Miller 2009). These results suggest that the speech network recruits additional cerebral cortex regions beyond the core speech in order to decipher what is being said in challenging listening conditions. In addition, the quality of speech sound representation as measured with auditory brainstem response recordings (ABR) during passive listening to speech sounds explains inter-individual variability in speech-in-noise recognition abilities (Chandrasekaran et al. 2009;Schoof and Rosen 2016; Song et al. 2011; Selinger et al. 2016). It has been suggested (Chandrasekaran et al. 2009) that the quality of representation might result from cortico-fugal modulation of brainstem nuclei that aids speech comprehension in challenging environments. The source of the high quality representation of speech sound in ABR is to-date, however, unclear. The ABR has a very low spatial resolution and sums responses from the auditory nerve up to the medial geniculate body (Hashimoto et al., 1981), so that it is impossible at the present time point to pinpoint one of the brainstem nuclei. One major candidate is the inferior colliculus (Chandrasekaran et al., 2014). Previous fMRI studies showed a task-dependent modulation of the IC for speech that correlated with speech recognition abilities in two experiments (kvk et al., 2008). Such correlation was, however, not replicated in following experiments (Diaz et al., 2012; Mihai et al., 2019) and was also not found in the present study. Thus, it remains an open question how the present findings of increased task-dependent modulation of the left vMGB for speech in noise recognition relate to potential mechanisms for speech in noise recognition in brainstem nuclei.

Previous experiments and theories emphasize the importance of an anatomical cortical and subcortical hierarchy that is organized according to the timescale of complex stimuli in the natural environment (Giraud et al. 2000; Kiebel, Daunizeau, and Friston 2008; Wang et al. 2008). The theories assume that levels closer to the sensory input encode faster dynamics of the stimulus than levels further away from the sensory input. Correspondingly, the primary sensory thalamus (vMGB, as well as the visual first-order thalamus: lateral

geniculate nucleus [LGN]) is tuned to high frequencies of temporal modulations compared with their associated primary sensory cortical areas (Giraud et al. 2000; Wang et al. 2008; Foster et al. 1985; Hicks, Lee, and Vidyasagar 1983). For humans, the optimized encoding of relatively fast dynamics; e.g., at the phoneme level, is critical for speech recognition and communication (Shannon et al. 1995; Tallal and Piercy 1975; Tallal et al. 1996; Goswami et al. 2011). Formant transitions, voice onset times, or stops, which constitute important speech components, are on time scales of 100 ms or less (Hayward 2000). Additionally, the sound envelope described by relatively fast temporal modulations (1-10 Hz in quiet environments, 10-50 Hz in noisy environments) is important for speech recognition (Shannon et al. 1995; Elliott and Theunissen 2009). Bayesian approaches to brain function propose that the brain uses internal dynamic models to predict the trajectory of the sensory input (Knill and Pouget 2004; Friston and Kiebel 2009; Friston 2005; Kiebel, Daunizeau, and Friston 2008). Thus slower dynamics of the internal dynamic model could be encoded by auditory cortical areas (Giraud et al. 2000; Wang et al. 2008) and provide predictions about the faster dynamics of the input arriving at lower levels of the temporal-anatomic hierarchy (Kiebel, Daunizeau, and Friston 2008; von Kriegstein, Patterson, and Griffiths 2008). In this context, dynamic predictions modulate the response properties of the first-order sensory thalamus to optimize the early stages of speech recognition. This view is consistent with findings in animals, where cortico-thalamic projections outnumber thalamocortical projections (Ojima and Rouiller 2011) and alter the response properties of thalamic neurons (Andolina et al. 2007; Cudeiro and Sillito 2006; Ergenzinger et al. 1998; Krupa, Ghazanfar, and Nicolelis 1999; Sillito, Cudeiro, and Jones 2006; Wang et al. 2018; Sillito et al. 1994; Ghazanfar and Nicolelis 2001). In speech processing such a mechanism might be especially useful as the signal includes both rapid dynamics and is predictable (e.g., due to co-articulation or learned statistical regularities in words) (Saffran 2003). Furthermore, speech needs to be computed online often under suboptimal listening conditions. A Bayesian brain mechanism has the potential to explain why the human brain is robust to (sensory) uncertainty. Uncertainty in this case refers to the limiting reliability of sensory information about the world (Knill and Pouget2004), which may be a combination of, for example, the density of hair cells in the cochlea that limit frequency resolution, the neural noise induced at different processing stages, or background environmental noise

that surrounds the stimulus of interest. The cerebral cortex auditory system acts upon lower levels in the hierarchy to build up accurate predictions within an internal generative model about fast sensory dynamics. In turn, this enhanced subcortical function provides improved signal quality to the auditory cortex. This mechanism results in more efficient processing when taxing conditions such as background noise confront the perceptual system.

We speculate that the task-dependent vMGB modulation might be a result of feedback from cerebral cortex areas. This feedback seems to be enhanced, when stimuli of interest are embedded in background noise. The task-dependent feedback may emanate directly from auditory primary or association cortices, or indirectly via other structures such as the reticular nucleus with its inhibitory connections to the MGB (Rouiller and de Ribaupierre 1985). Feedback cortico-thalamic projections from layer 6 in A1 to the vMGB, but also from association cortices such as the planum temporale (Tschentscher et al. 2019), may modulate information ascending through the lemniscal pathway, rather than convey information to the ventral division (Lee 2013; Llano and Sherman 2008). In theory, the modulation could also be relayed via brainstem nuclei such as the inferior colliculus, although this might not be the most parsimonious explanation.

Developmental disorders like autism spectrum disorder or developmental dyslexia are accompanied by difficulties in understanding speech in noise (Alcantara 2004, Ziegler 2009,Chandrasekaran 2009, Wong 2009, Schelinski 2019, Schoof 2016). In the case of developmental dyslexia, previous studies have found that developmental dyslexics do not have the same amount of task-dependent modulation of the left MGB for speech as controls (Díaz et al. 2012) and also do not display the same context-sensitivity of brainstem responses to speech sounds as typical readers (Chandrasekaran et al. 2009). In addition, diffusion-weigted imaging studies have found reduced structural connections between the sensory thalami (LGN and MGB) and cerebral cortex areas of the left hemisphere (V5/MT; motion-sensitive Planum temporale; Müller-Axt et al., 201x; Tschentscher et al., 2019). Together with the findings of the present study, these deficient mechanisms might account for the difficulties in understanding speech in suboptimal listening conditions in developmental dyslexia. Consider distinguishing speech sounds like “dad” and “had” in a

busy marketplace. For neurotypicals, cortical predictions may modulate vMGB responses to the subtle but predictable spectrotemporal cues that enable the explicit recognition of speech sounds. This modulation would enhance speech recognition. For dyslexics, however, this modulation may be impaired and may explain their difficulty with speech perception in noise (Díaz et al. 2012; Ziegler et al. 2009; Boets et al., 2007; Frey et al., 2018).

Based on previous findings (Mihai et al. 2019; von Kriegstein, Patterson, and Griffiths 2008), we expected a correlation in the vMGB between the Speech vs. Speaker contrast and the proportion of hits in the Speech task. This expectation was not confirmed.. We attribute this lack of correlation to the fact that ~11% of the data had ceiling or near to ceiling behavioural responses; a correlation in such a case is difficult to calculate. Many of the behavioural values were huddled towards the ceiling when plotted against BOLD responses (Figure S1). This was not the case in previous studies.

In a previous study (Mihai et al. 2019) it was surprising that there was no main effect of Speech vs. Speaker. We attributed this to the natural, unmanipulated stimulus material used. The former study used six different natural voices and no noise condition, while the stimuli in the current study used three synthesized voices and a noise condition. Vocal tract length and f0 are relatively stable acoustic cues that do not vary greatly over time in contrast to the highly dynamic cues (e.g., formant transitions, voice onset times, stops; Kent et al., 1992) that are most important for signaling phonemes and are used for speech recognition. However, dynamic cues, such as pitch periodicity, segmental timings, and prosody also aid speaker identification (Benesty et al., 2007). In the previous experiment, which included natural voices, participants might have also used fast changing cues for speaker identity recognition, particularly because the task was difficult. Since dynamic cues are essential for speech recognition, using dynamic cues in a speaker task would render the two tasks less different. The current experiment used three different voices synthesized from one source voice. The variability between these voices was, thus, lower, and three voices are more readily learned than six voices. This simplification resulted in better performance and less ambiguity when recognizing syllables. In sum, these differences

together with the behavioural ceiling effects of the current study may have led to the diverging results outlined here.

In conclusion, the results presented here suggest that the left vMGB is crucially involved in decoding speech as opposed to identifying the speaker and that this decoding is accentuated when there is background noise. This enhancement may be due to top-down processes that act upon subcortical structures, such as the auditory thalamus, to better predict incoming signals.

Materials and Methods

Participants

The Ethics committee of the Medical Faculty, University of Leipzig, Germany approved the study. We recruited 17 participants (mean age 27.7, SD 2.5, 10 female; 15 of these participated in a previous study (Mihai et al., 2019)) from the database of the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. The participants were right-handed (as assessed by the Edinburgh Handedness Inventory (Oldfield 1971)), and native German speakers. Participants provided written informed consent. None of the participants reported a history of psychiatric or neurological disorders, hearing difficulties, or current use of psychoactive medications. Normal hearing abilities were confirmed with pure tone audiometry (250 Hz to 8000 Hz) with a threshold equal to and below 25 dB (Madsen Micromate 304, GN Otometrics, Denmark). To exclude possible undiagnosed developmental dyslexics we tested the participant's reading speed and reading comprehension using the German LGVT: 6-12 test (Schneider, Ennemoser, and Schlagmüller 2007). The cutoff for both reading scores was set to those levels mentioned in the test instructions as the "lower average and above" performance range (i.e., 26% - 100% of the calculated population distribution). None of the participants performed below the cut off performance (mean 67.8%, SD 21.2%, lowest mean score: 36%). In addition, we used the rapid automatized naming (RAN) test of letters, numbers, and objects (Denckla and Rudel 1974; Denckla and Rudel 1976) to assess skills of phonological access. The time

required to name letters and numbers predicts reading ability and is longer in developmental dyslexics compared with typical readers, whereas the time to name objects is not a reliable predictor of reading ability in adults (Semrud-Clikeman et al. 2000). Participants scored well within the range of control participants for letters (mean 17.25, SD 2.52 s), numbers (mean 16.79, SD 2.63 s), and objects (mean 29.65, SD 4.47 s), based on results from a previous study (Diaz et al. 2012). Furthermore, none of the participants exhibited a clinically relevant number of traits associated with autism spectrum disorder as assessed by the Autism Spectrum Quotient (AQ; mean: 15.9, SD 4.1; cutoff: 32-50; (Baron-Cohen et al. 2001)). We tested AQ as autism can be associated with difficulties in speech-in-noise perception (Alcantara et al. 2004; Groen et al. 2009) and has overlapping symptoms with dyslexia (White et al. 2006). Participants received monetary compensation for participating in the study.

Stimuli

The stimuli consisted of 237 vowel-consonant-vowel (VCV) syllables with an average duration of 784 ms, SD 67 ms. These were spoken by one male voice (age 29 years), recorded with a video camera (Canon Legria HFS10, Canon, Japan) and a Røde NTG-1 microphone (Røde Microphones, Silverwater, NSW, Australia) connected to a pre-amplifier (TubeMP Project Series, Applied Research and Technology, Rochester, NY, USA) in a sound-attenuated room. The sampling rate was 48 kHz at 16 bit. Auditory stimuli were cut and flanked by Hamming windows of 15 ms at the beginning and end, converted to mono, and root-mean-square equalized using Python 3.6 (Python Software Foundation, www.python.org). There were three versions of the same 79 auditory files synthesized with TANDEM-STRAIGHT (Banno et al. 2007): 79 with a vocal tract length (VTL) of 17 cm and glottal pulse rate (GPR) of 100 Hz (deep male voice), 79 with VTL of 16 cm and GPR of 150 Hz (voice of a young adult), and 79 with VTL of 14 cm and GPR of 300 Hz (voice of a male child). The different GPR and VTL settings resulted in the percept of three different speakers. We spent a considerable amount of time piloting to find out which combination of GPR and VTL result in a balanced experimental design between the Speech and Speaker task.

Participants were familiarized with the three speakers' voices to ensure that they could perform the speaker-identity task of the main experiment. The familiarization phase took place 30 minutes prior to the fMRI experiment. The speaker familiarization consisted of a presentation of the speakers and a test phase. In the presentation phase, the speakers were presented in six blocks, each containing nine pseudo-randomly chosen stimuli. Participants heard a random choice of the same stimuli used in the experiment. Each block contained one speaker-identity only. Participants were alerted to the onset of a new speaker identity block by presentation of white words on a black screen indicating speaker 1, speaker 2, or speaker 3. Participants listened to the voices with the instruction to memorize the speaker's voice. In the following test phase participants were presented with four blocks of nine trials that each contained syllable pairs spoken by the three speakers. The syllable pairs could be from the same or a different speaker. We asked participants to indicate whether the speakers of the two syllables were the same by pressing keypad buttons "1" for yes and "2" for no. Participants received visual feedback for correct (green flashing German word for correct: "Richtig") and incorrect (red flashing German word for incorrect: "Falsch") answers. The speaker familiarization training consisted of three runs. If participants scored below 80% on the last run, they performed an additional run until they scored above 80%. All participants reached or exceeded the 80% cutoff.

Stimuli were also embedded in background noise (experimental design explained below). The background noise consisted of normally distributed random (white) noise filtered with a speech-shaped envelope. We calculated the envelope from the sum off all VCV stimuli presented in the experiment. We used speech-shaped noise as it has a stronger masking effect than stationary random non-speech noise (Carhartt et al. 1975). Before each run, the noise was computed and added to the stimuli included in the run with a signal-to-noise ratio of 2 dB. Calculations were performed in Matlab 8.6 (The Mathworks Inc., Natick, MA, USA) on Ubuntu Linux 16.04 (Canonical Ltd., London, UK).

Procedure

We conceived the experiment as a 2 × 2 factorial design with the factors task (Speech, Speaker) and background noise (Clear, Noise). Participants listened to blocks of auditory VCV syllables and were asked to perform two types of tasks: a Speech task and a Speaker task. In the Speech task, participants reported via button press whether the current syllable was different from the previous one (1-back task). In the Speaker task, participants reported via button press whether the current speaker was different from the previous one. Task instructions were presented for two seconds prior to each block and consisted of white written words on a black background (German words “Silbe” for syllable, and “Person” for person). After the instruction, the block of syllables started (Figure 1). Each block contained twelve stimuli. Each stimulus presentation was followed by 400 ms of silence. Within one block both syllables and speakers changed at least twice, with a theoretical maximum of nine changes. The theoretical maximum was derived from random sampling of seven instances from three possible change types: no change, speech change, speaker change, and change of speech and speaker. The average length of a block was 7.90 seconds, SD 0.26 seconds. Additionally to the factor task, the experiment included the factor background noise. Blocks had either syllables with background noise (Noise condition) or without background noise (Clear condition).

The experiment was divided into four runs. The first three had a duration of 12:56 min and included 40 blocks: 10 for each of the four conditions (speech noise, speaker noise, speech clear, speaker clear). A fourth run had a duration of 6:32 min and included 20 blocks (5 for each of the four conditions). For two participants only the first three runs were recorded due to time constraints. Participants could rest for one minute between runs.

Participants took part in an experiment familiarization task 30 minutes prior to the fMRI experiment. The experiment familiarization consisted of three 2:50 min run, similar to the runs presented during fMRI measurements. Stimuli were randomly chosen from the same stimulus material used in the experiment. If participants scored below 80% across tasks, they repeated the experiment familiarization training until the score was achieved.

The experiments were programmed in the Matlab Psychophysics Toolbox (Psychtoolbox-3, www.psychtoolbox.com (Brainard 1997)) running on Matlab 8.6 (The Mathworks Inc., Natick, MA, USA) on Ubuntu Linux 16.04 (Canonical Ltd., London, UK). Sound was delivered through a MrConfon amplifier and headphones (MrConfon GmbH, Magdeburg, Germany).

Data Acquisition and Processing

MRI data were acquired using a Siemens Magnetom 7 T scanner (Siemens AG, Erlangen, Germany) with an 8-channel head coil. We convened on the 8-channel coil, due to its spaciousness which allowed the use of higher quality headphones. Functional MRI data were acquired using echo planar imaging (EPI) sequences. We used a field of view (FoV) of 132 mm and partial coverage with 30 slices. This volume was oriented in parallel to the superior temporal gyrus such that the slices encompassed the inferior colliculi (IC), the MGB and the superior temporal gyrus.

The EPI sequences had the following acquisition parameters: TR = 1600 ms, TE = 19 ms, flip angle 65°, GRAPPA (Griswold et al. 2002) with acceleration factor 2, 33% phase oversampling, matrix size 88, FoV 132 mm x 132 mm, phase partial Fourier 6/8, voxel size $(1.5\ \text{mm})^3$, interleaved acquisition, anterior to posterior phase-encode direction. The first three runs consisted of 485 volumes (12:56 min), and the fourth run consisted of 245 volumes (6:32 min). During functional MRI data acquisition, we also acquired physiological values (heart rate, and respiration rate) using a BIOPAC MP150 system (BIOPAC Systems Inc., Goleta, CA, USA).

Structural images were recorded using an MP2RAGE (Marques et al. 2010) T1 protocol: 700 µm isotropic resolution, TE = 2.45ms, TR = 5000 ms, TI1 = 900 ms, TI2 = 2750 ms, flip angle 1 = 5°, flip angle 2 = 3°, FoV 224 mm × 224 mm, GRAPPA acceleration factor 2, duration 10:57 min.

Behavioural Data Analysis

Button presses (hits/misses) were binomially distributed, and were thus modeled using a binomial logistic regression which predicts the probability of correct button presses based on four independent variables (speech task, clear condition; speech task, noise condition; speaker task, clear condition; speaker, task noise condition) in a Bayesian framework (McElreath 2016).

To pool over participants and runs we modelled the correlation between intercepts and slopes. For the model implementation and data analysis, we used PyMC3 3.5 (Salvatier, Wiecki, and Fonnesbeck 2016), a probabilistic programming package for Python 3.6, sampling with a No-U-Turn Sampler (Hoffman and Gelman 2014) with four parallel chains. Per chain, we had 5,000 samples with 5,000 as warm-up. For the effects of interest (Speech vs. Speaker, Noise vs. Clear, Speech Noise vs. Speaker Noise, Speech Clear vs. Speaker Clear), we calculated means from the posterior distributions and 95% highest posterior density intervals (HDP). If the posterior probability distribution does not strongly overlap zero, i.e., the HPD excludes zero, then there was a detectable difference between conditions (McElreath 2016; Bunce and McElreath 2017).

The predictors included in the behavioural data model were: task (x_S:1 = speech, 0 = speaker), and background noise (x_N: 1 = noise, 0 = clear). We also included the two-way interaction of Task and noise condition. Because data were collected across participants and runs we included random effects for both of these in the logistic model. Furthermore, since ~11% of the data exhibited ceiling effects (i.e., some participants scored at the highest possible level) which would result in underestimated means and standard deviations (Uttl 2005), we treated these data as right-censored and modeled them using a Potential class (Lauritzen et al. 1990; Jordan 1998) as implemented in PyMC3. This method integrates out the censored values using the log of the complementary normal cumulative distribution function (McElreath 2016; Gelman et al. 2013). In essence, we sampled twice, once for the observed values without the censored data points, and once for the censored values only.

The model is described below.

$$L_{i,j} \sim Binomial(1, p_{i,j})$$

$$p_{i,j} = \begin{cases} p^*_{i,j}, & \text{for } p^*_{i,j} < c \\ c, & \text{for } p^*_{i,j} \geq c \end{cases}$$

$$logit(p^*_{i,j}) = A_{i,j} + B_{S,i,j}x_S + B_{N,i,j}x_N + B_{SN,i,j}x_S x_N, \text{for } i = 1,\ldots,I; j = 1,\ldots,J$$

$$A_{i,j} = \alpha + \alpha_{participant[i]} + \alpha_{run[j]}$$

$$B_{S,i,j} = \beta_S + \beta_{S,participant[i]} + \beta_{S,run[j]}$$

$$B_{N,i,j} = \beta_N + \beta_{N,participant[i]} + \beta_{N,run[j]}$$

$$B_{SN,i,j} = \beta_{SN} + \beta_{SN,participant[i]} + \beta_{SN,run[j]}$$

$$\begin{bmatrix} \alpha_{participant} \\ \beta_{S,participant} \\ \beta_{N,participant} \\ \beta_{SN,participant} \end{bmatrix} \sim MVNormal\left(\begin{bmatrix} \alpha \\ \beta_S \\ \beta_N \\ \beta_{SN} \end{bmatrix}, S_{participant}\right)$$

$$\begin{bmatrix} \alpha_{run} \\ \beta_{S,run} \\ \beta_{N,run} \\ \beta_{SN,run} \end{bmatrix} \sim MVNormal\left(\begin{bmatrix} \alpha \\ \beta_S \\ \beta_N \\ \beta_{SN} \end{bmatrix}, S_{run}\right)$$

$$S_{subject} = \begin{bmatrix} \sigma_\alpha & 0 & 0 & 0 \\ 0 & \sigma_{\beta_S} & 0 & 0 \\ 0 & 0 & \sigma_{\beta_N} & 0 \\ 0 & 0 & 0 & \sigma_{\beta_{SN}} \end{bmatrix} R_{subject} \begin{bmatrix} \sigma_\alpha & 0 & 0 & 0 \\ 0 & \sigma_{\beta_S} & 0 & 0 \\ 0 & 0 & \sigma_{\beta_N} & 0 \\ 0 & 0 & 0 & \sigma_{\beta_{SN}} \end{bmatrix}$$

$$S_{run} = \begin{bmatrix} \sigma_\alpha & 0 & 0 & 0 \\ 0 & \sigma_{\beta_S} & 0 & 0 \\ 0 & 0 & \sigma_{\beta_N} & 0 \\ 0 & 0 & 0 & \sigma_{\beta_{SN}} \end{bmatrix} R_{run} \begin{bmatrix} \sigma_\alpha & 0 & 0 & 0 \\ 0 & \sigma_{\beta_S} & 0 & 0 \\ 0 & 0 & \sigma_{\beta_N} & 0 \\ 0 & 0 & 0 & \sigma_{\beta_{SN}} \end{bmatrix}$$

$$\alpha \sim Normal(0,5)$$

$$\beta_S \sim Normal(0,5)$$

$$\beta_N \sim Normal(0,5)$$

$$\beta_{SN} \sim Normal(0,5)$$

$$(\sigma_{participant}, \sigma_{run}) \sim HalfCauchy(1)$$

$$\sigma_{corr,participant} \sim HalfCauchy(1)$$

$$\sigma_{corr,run} \sim HalfCauchy(1)$$

$$R_{participant} \sim LKJcorr(4, \sigma_{corr,participant})$$

$$R_{run} \sim LKJcorr(4, \sigma_{corr,run})$$

I represents the participants and *J* the runs. The model is compartmentalized into sub-models for the intercepts and slopes. $A_{i,j}$ is the sub-model for the intercept for observations i, j. Similarly, $B_{S,i,j}$, $B_{N,i,j}$, and $B_{SN,i,j}$ are the sub-models for the speech-speaker slope, clear-noise slope and the interaction slope, resepectively; $S_{subject}/S_{run}$ are the covariance matrices for participant/run and $R_{subject}/R_{run}$ are the priors for the correlation matrices modeled as LKJ probability densities (Lewandowski, Kurowicka, and Joe 2009). Weakly informative priors for the intercept (α) and additional coefficients (e.g., β_S), random effects for participant and run ($\beta_{S,subject}$, $\beta_{S,run}$), and multivariate priors for participants and runs identify the model by constraining the position of $p_{i,j}$ to reasonable values. Here we used normal distributions as priors. Furthermore, $p_{i,j}$ is defined as the ramp function equal to the proportion of hits when these are known and below the ceiling (c), and set to the ceiling if they are equal to or greater than the ceiling *c.*

Functional MRI Data Analysis

Preprocessing of fMRI data

To deal with the noise surrounding the head in MP2RAGE images, these were first segmented using SPM's new segment function (SPM 12, version 12.6906, Wellcome Trust Centre for Human Neuroimaging, UCL, UK, http://www.fil.ion.ucl.ac.uk/spm) running on Matlab 8.6 (The Mathworks Inc., Natick, MA, USA). The resulting grey and white matter segmentations were summed and binarized to remove voxels that contain air, scalp, skull and cerebrospinal fluid from structural images using the ImCalc function of SPM.

We used the template image created for a previous study (Mihai et al. 2019) using structural MP2RAGE images from 28 participants. We chose this template since 15 participants in the current study are included in this image, and the vMGB mask (described below) is in the same space as the template image. The choice of this common template reduces warping artifacts, which would be introduced with a different template, as both the mask and the functional data would need to be warped to a common space. The template was created with ANTs (Avants et al. 2008), which was then registered to the MNI space using the same software package and the MNI152 template provided by FSL 5.0.8 (Smith et al. 2004). All MP2RAGE images were preprocessed with Freesurfer (Fischl et al. 2002; Han and Fischl 2007; Fischl et al. 2004) using the recon-all command to obtain boundaries between grey and white matter, which were later used in the functional to structural registration step.

Preprocessing and statistical analyses pipelines were coded in nipype 1.1.2 (Gorgolewski et al. 2011). Head motion and susceptibility distortion by movement interaction of functional runs were corrected using the Realign and Unwarp method (Andersson et al. 2001) in SPM 12. This step also makes use of a voxel displacement map (VDM), which addresses the problem of geometric distortions in EPI caused by magnetic field inhomogeneity. The VDM was calculated using field map recordings, which provided the absolute value and the phase difference image files, using the FieldMap Toolbox (Jezzard and Balaban 1995) of SPM 12. Outlier runs were detected using ArtifactDetect (composite threshold of translation and rotation: 1; intensity Z-threshold: 3; global threshold: 8;

https://www.nitrc.org/projects/artifact_detect/). Coregistration matrices for realigned functional runs per participant were computed based on each participant's structural image using Freesurfer's BBregister function (register mean EPI image to T1). We used a whole-brain EPI volume as an intermediate file in the coregistration step to avoid registration problems due to the limited FoV of the functional runs. Warping using coregistration matrices (after conversion to ITK coordinate system) and resampling to 1 mm isovoxel was performed using ANTs. Before model creation, we smoothed the data in SPM12 using a 1 mm kernel at full-width half-maximum.

Physiological data

Physiological data (heart rate and respiration rate) were processed by the PhysIO Toolbox (Kasper et al. 2017) to obtain Fourier expansions of each, in order to enter these into the design matrix (see statistical analyses sections below). Since heart beats and respiration result in undesired cortical and subcortical artifacts, regressing these out increases the specificity of fMRI activation to the task of interest (Kasper et al. 2017). These artifacts occur in abundance around the thalamus.

Statistical Analysis of fMRI data

Models were set up in SPM 12 using the native space data for each participant. The design matrix included the following nuisance regressors: three cardiac, four respiratory, and a cardiac × respiratory interaction regressor. We additionally entered a variable number of outlier regressors from the ArtifactDetect step, depending on how many outliers were found in each run. We modelled five conditions of interest: speech clear, speech noise, speaker clear, speaker noise, and task instruction. Onset times and durations were used to create boxcar functions, which were convolved with the hemodynamic response function (HRF) provided by SPM 12. Parameter estimates were computed for each condition at the first level using restricted maximum likelihood (REML) as implemented in SPM 12.

After estimation, contrasts were registered to the MNI structural template using a two-step registration in ANTs. First, a quick registration was performed on the whole head using rigid, affine and diffeomorphic transformations (using Symmetric Normalization, SyN), and the mutual information similarity metric. Second, the high-quality registration was confined to a rectangular prism mask encompassing the left and right MGB, and IC only. This step used affine and SyN transformations, and mean squares and neighbourhood cross-correlation similarity measures, respectively. We performed the registration to MNI space by linearly interpolating the contrast images using the composite transforms from the high-quality registration.

Using the left ventral MGB mask from our previous experiment (Mihai et al. 2019), we extracted parameter estimates for each of the four conditions per participant, averaged over all voxels within the mask. These data were further analyzed in a Bayesian framework (McElreath 2016). The model was implemented in PyMC3 with a No-U-Turn Sampler with four parallel chains. Per chain, we had 5000 samples with 5000 as warm-up. For the distributions of interest, we calculated means and 95% HDP. If posterior probability distributions do not strongly overlap zero, then there was a detectable effect. Differences between conditions were computed using Hedges g* (Hedges and Olkin 1985), which provides an unbiased effect size similar to Cohen's *d* (Cohen 1988), corrected for low sample sizes.

The predictors included in the model were: task (x_S: 1 = Speech, 0 = Speaker), and background noise (x_N: 1 = Noise, 0 = Clear). We also included the two-way interaction of task and noise condition. Because data were collected across participants it is reasonable to include random effects. To pool over participants we modelled the correlation between intercepts and slopes over participants. The interaction model is described below.

$$L_i \sim T(\mu_i, \nu, \lambda)$$

$$\mu_i = A_i + B_{S,i}x_S + B_{N,i}x_N + B_{SN,i}x_S x_N, \text{for } i = 1, \ldots, I$$

$$A_i = \alpha + \alpha_{participant[i]}$$

$$B_{S,i} = \beta_S + \beta_{S,participant[i]}$$

$$B_{N,i} = \beta_N + \beta_{N,participant[i]}$$

$$B_{SN,i} = \beta_{SN} + \beta_{SN,participant[i]}$$

$$\begin{bmatrix} \alpha_{participant} \\ \beta_{S,participant} \\ \beta_{N,participant} \\ \beta_{SN,participant} \end{bmatrix} \sim MVNormal\left(\begin{bmatrix} \alpha \\ \beta_S \\ \beta_N \\ \beta_{SN} \end{bmatrix}, S\right)$$

$$S = \begin{bmatrix} \sigma_\alpha & 0 & 0 & 0 \\ 0 & \sigma_{\beta_S} & 0 & 0 \\ 0 & 0 & \sigma_{\beta_N} & 0 \\ 0 & 0 & 0 & \sigma_{\beta_{SN}} \end{bmatrix} R \begin{bmatrix} \sigma_\alpha & 0 & 0 & 0 \\ 0 & \sigma_{\beta_S} & 0 & 0 \\ 0 & 0 & \sigma_{\beta_N} & 0 \\ 0 & 0 & 0 & \sigma_{\beta_{SN}} \end{bmatrix}$$

$$\alpha \sim T(0,1,3)$$

$$\beta_S \sim T(0,1,3)$$

$$\beta_N \sim T(0,1,3)$$

$$\beta_{SN} \sim T(0,1,3)$$

$$(\sigma_{participant}) \sim HalfCauchy(1)$$

$$\sigma_{corr} \sim HalfCauchy(1)$$

$$R \sim LKJcorr(4, \sigma_{corr})$$

$$\nu \sim Exponential(1/29) + 1$$

$$\sigma \sim HalfCauchy(2)$$

$$\lambda = \sigma^{-2}$$

I represents the participants. The model is compartmentalized into sub-models for the intercepts and slopes. A_i is the sub-model for the intercept for observations i.

Similarly, $B_{S,i}$, $B_{N,i}$, and $B_{SN,i}$ are the sub-models for the speech-speaker slope, clear-noise slope and the interaction slope, respectively; S is the covariance matrix and R is the prior for the correlation matrix modelled as a LKJ probability density (Lewandowski, Kurowicka, and Joe 2009). Weakly informative priors for the intercept (α) and additional coefficients (e.g., β_S), random effects for participant and run ($\beta_{S,subject}$, $\beta_{S,run}$), and multivariate priors for participants identify the model by constraining the position of μ_i to reasonable values. Here we used Student's-T distributions as priors.

Brain-behaviour correlation

We were interested in replicating the brain-behaviour correlation found in three previous experiments (experiment 1 & 2 in von Kriegstein et al. 2008; Mihai et al. 2019). We thus tested whether participants who are better speech recognizers also show increased BOLD response. To answer this question, we performed a correlation calculation between the BOLD response in the left vMGB for the Speech vs. Speaker task together with the proportion of hits in the Speech task. We calculated Pearson's correlation using a multivariate normal distribution with a covariance matrix that automatically includes the correlation term ρ and sampled 2000 samples from the posterior using a No-U-Turn sampler with 2000 samples as warm-up. Mean ρ and 95% HDP were calculated to infer the strength and presence of the correlation. If the HPD included zero, we considered this a lack of correlation.

Acknowledgements

PLOS journals publicly acknowledge the indispensible efforts of our editors and reviewers on an annual basis. To ensure equitable recognition and avoid any appearance of partiality, do not include editors or peer reviewers—named or unnamed—in the Acknowledgments.

Do not include funding sources in the Acknowledgments or anywhere else in the manuscript file. Funding information should only be entered in the financial disclosure section of the submission system.

Supplementary Material

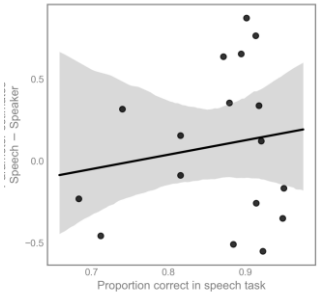

Figure S1. Correlation of Proportion correct in the speech task and parameter estimates from the difference between the Speech and Speaker task. Most data points are close to the ceiling on the right of the behavioural score.

References

Anderson, Samira, and Nina Kraus. 2010. "Sensory-Cognitive Interaction in the Neural Encoding of Speech in Noise: A Review". *Journal of the American Academy of Audiology* 21 (9). American Academy of Audiology: 575–85. doi:10.3766/jaaa.21.9.3.

Hawley, Monica L., Ruth Y. Litovsky, and John F. Culling. 2004. "The Benefit of Binaural Hearing in a Cocktail Party: Effect of Location and Type of Interferer". *The Journal of the Acoustical Society of America* 115 (2). Acoustical Society of America (ASA): 833–43. doi:10.1121/1.1639908.

Bronkhorst, Adelbert W. 2015. "The Cocktail-Party Problem Revisited: Early Processing and Selection of Multi-Talker Speech". *Attention Perception, & Psychophysics* 77 (5). Springer Nature: 1465–87. doi:10.3758/s13414-015-0882-9.

Best, Virginia, Frederick J. Gallun, Simon Carlile, and Barbara G. Shinn-Cunningham. 2007. "Binaural Interference and Auditory Grouping". *The Journal of the Acoustical Society of America* 121 (2). Acoustical Society of America (ASA): 1070–76. doi:10.1121/1.2407738.

Shinn-Cunningham, Barbara G., and Virginia Best. 2008. "Selective Attention in Normal and Impaired Hearing". *Trends in Amplification* 12 (4). SAGE Publications: 283–99. doi:10.1177/1084713808325306.

Brokx, JPL, and SG Nooteboom. 1982. "Intonation and the Perceptual Separation of Simultaneous Voices". *Journal of Phonetics* 10 (1): 23–36.

Moore, Brian C. J., Robert W. Peters, and Brian R. Glasberg. 1985. "Thresholds for the Detection of Inharmonicity in Complex Tones". *The Journal of the Acoustical Society of America* 77 (5). Acoustical Society of America (ASA): 1861–67. doi:10.1121/1.391937.

Bregman, Albert S., and Stephen McAdams. 1994. "Auditory Scene Analysis: The Perceptual Organization of Sound". *The Journal of the Acoustical Society of America* 95 (2). Acoustical Society of America (ASA): 1177–78. doi:10.1121/1.408434.

Darwin, C. J., and R. W. Hukin. 2000. "Effectiveness of Spatial Cues Prosody, and Talker Characteristics in Selective Attention". *The Journal of the Acoustical Society of America* 107 (2). Acoustical Society of America (ASA): 970–77. doi:10.1121/1.428278.

Parikh, Gaurang, and Philipos C. Loizou. 2005. "The Influence of Noise on Vowel and Consonant Cues". *The Journal of the Acoustical Society of America* 118 (6). Acoustical Society of America (ASA): 3874–88. doi:10.1121/1.2118407.

Sayles, M., and I. M. Winter. 2008. "Ambiguous Pitch and the Temporal Representation of Inharmonic Iterated Rippled Noise in the Ventral Cochlear Nucleus". *Journal of Neuroscience* 28 (46). Society for Neuroscience: 11925–38. doi:10.1523/jneurosci.3137-08.2008.

Heinrich, Antje, Bruce A. Schneider, and Fergus I. M. Craik. 2008. "Investigating the Influence of Continuous Babble on Auditory Short-Term Memory Performance". *Quarterly Journal of Experimental Psychology* 61 (5). SAGE Publications: 735–51. doi:10.1080/17470210701402372.

Alcantara, Jose I., Emma J.L. Weisblatt, Brian C.J. Moore, and Patrick F. Bolton. 2004. "Speech-in-Noise Perception in High-Functioning Individuals with Autism or Aspergers Syndrome". *Journal of Child Psychology and Psychiatry* 45 (6). Wiley: 1107–14. doi:10.1111/j.1469-7610.2004.t01-1-00303.x.

Ziegler, Johannes C., Catherine Pech-Georgel, Florence George, and Christian Lorenzi. 2009. "Speech-Perception-in-Noise Deficits in Dyslexia". *Developmental Science* 12 (5). Wiley: 732–45. doi:10.1111/j.1467-7687.2009.00817.x.

Chandrasekaran, Bharath, Jane Hornickel, Erika Skoe, Trent Nicol, and Nina Kraus. 2009. "Context-Dependent Encoding in the Human Auditory Brainstem Relates to Hearing Speech in Noise: Implications for Developmental Dyslexia". *Neuron* 64 (3). Elsevier BV: 311–19. doi:10.1016/j.neuron.2009.10.006.

Wong, Patrick C.M., James Xumin Jin, Geshri M. Gunasekera, Rebekah Abel, Edward R. Lee, and Sumitrajit Dhar. 2009. "Aging and Cortical Mechanisms of Speech Perception in Noise". *Neuropsychologia* 47 (3). Elsevier BV: 693–703. doi:10.1016/j.neuropsychologia.2008.11.032.

Schelinski, Stefanie, and Katharina von Kriegstein. 2019. "Speech-in-Noise Recognition and the Relation to Vocal Pitch Perception in Adults with Autism Spectrum Disorder and Typical Development". *In Prep*.

Schoof, Tim, and Stuart Rosen. 2016. "The Role of Age-Related Declines in Subcortical Auditory Processing in Speech Perception in Noise". *Journal of the Association for Research in Otolaryngology* 17 (5). Springer Nature: 441–60. doi:10.1007/s10162-016-0564-x.

Strait, Dana L., Alexandra Parbery-Clark, Emily Hittner, and Nina Kraus. 2012. "Musical Training during Early Childhood Enhances the Neural Encoding of Speech in Noise". *Brain and Language* 123 (3). Elsevier BV: 191–201. doi:10.1016/j.bandl.2012.09.001.

Parbery-Clark, Alexandra, Erika Skoe, Carrie Lam, and Nina Kraus. 2009. "Musician Enhancement for Speech-In-Noise". *Ear and Hearing* 30 (6). Ovid Technologies (Wolters Kluwer Health): 653–61. doi:10.1097/aud.0b013e3181b412e9.

Zeng, Fan-Gang, and Thomas Lu. 2010. "Faculty of 1000 Evaluation for Musician Enhancement for Speech-in-Noise.". Faculty of 1000 Ltd. doi:10.3410/f.2540956.2189054.

Wernicke, Carl. 1874. *Der Aphasische Symptomencomplex: Eine Psychologische Studie Auf Anatomischer Basis*. Cohn.

Hickok, Gregory, and Steven L. Small. 2015. *Neurobiology of Language*. Academic Press.

Friederici, Angela D, and Sarah ME Gierhan. 2013. "The Language Network". *Current Opinion in Neurobiology*, Macrocircuits, 23 (2): 250–54. doi:10.1016/j.conb.2012.10.002.

von Kriegstein, Katharina, Roy D. Patterson, and T.D. Griffiths. 2008. "Task-Dependent Modulation of Medial Geniculate Body Is Behaviorally Relevant for Speech Recognition". *Current Biology* 18 (23). Elsevier BV: 1855–59. doi:10.1016/j.cub.2008.10.052.

Diaz, B., F. Hintz, S. J. Kiebel, and K. von Kriegstein. 2012. "Dysfunction of the Auditory Thalamus in Developmental Dyslexia". *Proceedings of the National Academy of Sciences* 109 (34). Proceedings of the National Academy of Sciences: 13841–46. doi:10.1073/pnas.1119828109.

Díaz, Begoña, Helen Blank, and Katharina von Kriegstein. 2018. "Task-Dependent Modulation of the Visual Sensory Thalamus Assists Visual-Speech Recognition". *NeuroImage* 178 (September): 721–34. doi:10.1016/j.neuroimage.2018.05.032.

Mihai, Paul Glad, Michelle Moerel, Federico de Martino, Robert Trampel, Stefan Kiebel, and Katharina von Kriegstein. 2019. "Modulation of Tonotopic Ventral MGB Is Behaviorally Relevant for Speech Recognition.: Supplemental Information", January. Cold Spring Harbor Laboratory. doi:10.1101/510164.

Chandrasekaran, Bharath, Nina Kraus, and Patrick C. M. Wong. 2011. "Human Inferior Colliculus Activity Relates to Individual Differences in Spoken Language Learning". *Journal of Neurophysiology* 107 (5): 1325–36. doi:10.1152/jn.00923.2011.

Squire, Larry, Darwin Berg, Floyd E Bloom, Sascha Du Lac, Anirvan Ghosh, and Nicholas C Spitzer. 2012. *Fundamental Neuroscience*. Academic Press.

Saalmann, Yuri B., and Sabine Kastner. 2015. "The Cognitive Thalamus". *Frontiers in Systems Neuroscience* 9 (March). Frontiers Media SA. doi:10.3389/fnsys.2015.00039.

Saalmann, Yuri B., and Sabine Kastner. 2011. "Cognitive and Perceptual Functions of the Visual Thalamus". *Neuron* 71 (2). Elsevier BV: 209–23. doi:10.1016/j.neuron.2011.06.027.

Haynes, John-Dylan, Ralf Deichmann, and Geraint Rees. 2005. “Eye-Specific Effects of Binocular Rivalry in the Human Lateral Geniculate Nucleus”. *Nature* 438 (7067). Springer Nature: 496–99. doi:10.1038/nature04169.

Antunes, F. M., and M. S. Malmierca. 2011. “Effect of Auditory Cortex Deactivation on Stimulus-Specific Adaptation in the Medial Geniculate Body”. *Journal of Neuroscience* 31 (47). Society for Neuroscience: 17306–16. doi:10.1523/jneurosci.1915-11.2011.

OConnor, Daniel H., Miki M. Fukui, Mark A. Pinsk, and Sabine Kastner. 2002. “Attention Modulates Responses in the Human Lateral Geniculate Nucleus”. *Nature Neuroscience* 5 (11). Springer Nature: 1203–9. doi:10.1038/nn957.

McAlonan, Kerry, James Cavanaugh, and Robert H. Wurtz. 2008. “Guarding the Gateway to Cortex with Attention in Visual Thalamus”. *Nature* 456 (7220). Springer Nature: 391–94. doi:10.1038/nature07382.

Díaz, Begoña, Helen Blank, and Katharina von Kriegstein. 2018. “Task-Dependent Modulation of the Visual Sensory Thalamus Assists Visual-Speech Recognition”. *NeuroImage* 178 (September). Elsevier BV: 721–34. doi:10.1016/j.neuroimage.2018.05.032.

Malmierca, Manuel S., Lucy A. Anderson, and Flora M. Antunes. 2015. “The Cortical Modulation of Stimulus-Specific Adaptation in the Auditory Midbrain and Thalamus: a Potential Neuronal Correlate for Predictive Coding”. *Frontiers in Systems Neuroscience* 9 (March). Frontiers Media SA. doi:10.3389/fnsys.2015.00019.

Winer, Jeffery A., Lee M. Miller, Charles C. Lee, and Christoph E. Schreiner. 2005. “Auditory Thalamocortical Transformation: Structure and Function”. *Trends in Neurosciences* 28 (5). Elsevier BV: 255–63. doi:10.1016/j.tins.2005.03.009.

Sherman, S. M., and R. W. Guillery. 1998. “On the Actions That One Nerve Cell Can Have on Another: Distinguishing Drivers from Modulators”. *Proceedings of the National Academy of Sciences* 95 (12). Proceedings of the National Academy of Sciences: 7121–26. doi:10.1073/pnas.95.12.7121.

Calford, M. B. 1983. "The Parcellation of the Medial Geniculate Body of the Cat Defined by the Auditory Response Properties of Single Units". *Journal of Neuroscience* 3 (11): 2350–64. doi:10.1523/JNEUROSCI.03-11-02350.1983.

Rodrigues-Dagaeff, C., G. Simm, Y. De Ribaupierre, A. Villa, F. De Ribaupierre, and E.M. Rouiller. 1989. "Functional Organization of the Ventral Division of the Medial Geniculate Body of the Cat: Evidence for a Rostro-Caudal Gradient of Response Properties and Cortical Projections". *Hearing Research* 39 (1-2). Elsevier BV: 103–25. doi:10.1016/0378-5955(89)90085-3.

Anderson, L.A., M.N. Wallace, and A.R. Palmer. 2007. "Identification of Subdivisions in the Medial Geniculate Body of the Guinea Pig". *Hearing Research* 228 (1-2). Elsevier BV: 156–67. doi:10.1016/j.heares.2007.02.005.

Vasquez-Lopez, Sebastian A, Yves Weissenberger, Michael Lohse, Peter Keating, Andrew J King, and Johannes C Dahmen. 2017. "Thalamic Input to Auditory Cortex Is Locally Heterogeneous but Globally Tonotopic". *ELife* 6 (September). eLife Sciences Publications Ltd. doi:10.7554/elife.25141.

Mothe, Lisa A. De La, Suzanne Blumell, Yoshinao Kajikawa, and Troy A. Hackett. 2006. "Thalamic Connections of the Auditory Cortex in Marmoset Monkeys: Core and Medial Belt Regions". *The Journal of Comparative Neurology* 496 (1). Wiley: 72–96. doi:10.1002/cne.20924.

Anderson, L. A., G. B. Christianson, and J. F. Linden. 2009. "Stimulus-Specific Adaptation Occurs in the Auditory Thalamus". *Journal of Neuroscience* 29 (22). Society for Neuroscience: 7359–63. doi:10.1523/jneurosci.0793-09.2009.

Calford, MB. 1983. "The Parcellation of the Medial Geniculate Body of the Cat Defined by the Auditory Response Properties of Single Units". *The Journal of Neuroscience* 3 (11). Society for Neuroscience: 2350–64. doi:10.1523/jneurosci.03-11-02350.1983.

Cruikshank, S.J, H.P Killackey, and R Metherate. 2001. "Parvalbumin and Calbindin Are Differentially Distributed within Primary and Secondary Subregions of the Mouse Auditory

Forebrain". *Neuroscience* 105 (3). Elsevier BV: 553–69. doi:10.1016/s0306-4522(01)00226-3.

Gonzalez-Lima, F., and A. Cada. 1994. "Cytochrome Oxidase Activity in the Auditory System of the Mouse: A Qualitative and Quantitative Histochemical Study". *Neuroscience* 63 (2). Elsevier BV: 559–78. doi:10.1016/0306-4522(94)90550-9.

Hackett, T.A., I. Stepniewska, and J.H. Kaas. 1998. "Thalamocortical Connections of the Parabelt Auditory Cortex in Macaque Monkeys". *The Journal of Comparative Neurology* 400 (2). Wiley: 271–86. doi:10.1002/(sici)1096-9861(19981019)400:2<271::aid-cne8>3.0.co;2-6.

Morest, D Kent. 1964. "The Neuronal Architecture of the Medial Geniculate Body of the Cat". *Journal of Anatomy* 98 (Pt 4). Wiley-Blackwell: 611.

Winer, Jeffery A, Jack B Kelly, and David T Larue. 1999. "Neural Architecture of the Rat Medial Geniculate Body". *Hearing Research* 130 (1-2). Elsevier BV: 19–41. doi:10.1016/s0378-5955(98)00216-0.

Anderson, Lucy A., and Jennifer F. Linden. 2011. "Physiological Differences between Histologically Defined Subdivisions in the Mouse Auditory Thalamus". *Hearing Research*, Central Auditory Pathways – A Tribute to Jeffery A. Winer, 274 (1): 48–60. doi:10.1016/j.heares.2010.12.016.

Rodrigues-Dagaeff, C., G. Simm, Y. De Ribaupierre, A. Villa, F. De Ribaupierre, and E. M. Rouiller. 1989. "Functional Organization of the Ventral Division of the Medial Geniculate Body of the Cat: Evidence for a Rostro-Caudal Gradient of Response Properties and Cortical Projections". *Hearing Research* 39 (1): 103–25. doi:10.1016/0378-5955(89)90085-3.

Bartlett, Edward L., and Xiaoqin Wang. 2011. "Correlation of Neural Response Properties with Auditory Thalamus Subdivisions in the Awake Marmoset". *Journal of Neurophysiology* 105 (6). American Physiological Society: 2647–67. doi:10.1152/jn.00238.2010.

Ohga, Shinpei, Hiroaki Tsukano, Masao Horie, Hiroki Terashima, Nana Nishio, Yamato Kubota, Kuniyuki Takahashi, Ryuichi Hishida, Hirohide Takebayashi, and Katsuei Shibuki. 2018. "Direct Relay Pathways from Lemniscal Auditory Thalamus to Secondary Auditory

Field in Mice". *Cerebral Cortex* 28 (12). Oxford University Press (OUP): 4424–39. doi:10.1093/cercor/bhy234.

Knill, David C., and Alexandre Pouget. 2004. "The Bayesian Brain: the Role of Uncertainty in Neural Coding and Computation". *Trends in Neurosciences* 27 (12). Elsevier BV: 712–19. doi:10.1016/j.tins.2004.10.007.

Friston, Karl, and Stefan Kiebel. 2009. "Predictive Coding under the Free-Energy Principle". *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1521). The Royal Society: 1211–21. doi:10.1098/rstb.2008.0300.

Friston, Karl. 2005. "A Theory of Cortical Responses". *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1456). The Royal Society: 815–36. doi:10.1098/rstb.2005.1622.

Kiebel, Stefan J., Jean Daunizeau, and Karl J. Friston. 2008. "A Hierarchy of Time-Scales and the Brain". Edited by Olaf Sporns. *PLoS Computational Biology* 4 (11). Public Library of Science (PLoS): e1000209. doi:10.1371/journal.pcbi.1000209.

Krupa, D. J., A. A. Ghazanfar, and M. A. L. Nicolelis. 1999. "Immediate Thalamic Sensory Plasticity Depends on Corticothalamic Feedback". *Proceedings of the National Academy of Sciences* 96 (14). Proceedings of the National Academy of Sciences: 8200–8205. doi:10.1073/pnas.96.14.8200.

Sillito, Adam M., Javier Cudeiro, and Helen E. Jones. 2006. "Always Returning: Feedback and Sensory Processing in Visual Cortex and Thalamus". *Trends in Neurosciences* 29 (6). Elsevier BV: 307–16. doi:10.1016/j.tins.2006.05.001.

Wang, Wei, Ian M. Andolina, Yiliang Lu, Helen E. Jones, and Adam M. Sillito. 2016. "Focal Gain Control of Thalamic Visual Receptive Fields by Layer 6 Corticothalamic Feedback". *Cerebral Cortex*, December. Oxford University Press (OUP). doi:10.1093/cercor/bhw376.

Chandrasekaran, Bharath, and Nina Kraus. 2010. "Music Noise-Exclusion, and Learning". *Music Perception* 27 (4). University of California Press: 297–306. doi:10.1525/mp.2010.27.4.297.

Moerel, Michelle, Federico De Martino, Kâmil Uğurbil, Essa Yacoub, and Elia Formisano. 2015. "Processing of Frequency and Location in Human Subcortical Auditory Structures". *Scientific Reports* 5 (1). Springer Nature. doi:10.1038/srep17048.

Salvi, R.J., A.H. Lockwood, R.D. Frisina, M.L. Coad, D.S. Wack, and D.R. Frisina. 2002. "PET Imaging of the Normal Human Auditory System: Responses to Speech in Quiet and in Background Noise". *Hearing Research* 170 (1-2). Elsevier BV: 96–106. doi:10.1016/s0378-5955(02)00386-6.

Scott, Sophie K., Stuart Rosen, Lindsay Wickham, and Richard J. S. Wise. 2004. "A Positron Emission Tomography Study of the Neural Basis of Informational and Energetic Masking Effects in Speech Perception". *The Journal of the Acoustical Society of America* 115 (2). Acoustical Society of America (ASA): 813–21. doi:10.1121/1.1639336.

Sigalovsky, Irina S., and Jennifer R. Melcher. 2006. "Effects of Sound Level on FMRI Activation in Human Brainstem Thalamic and Cortical Centers". *Hearing Research* 215 (1-2). Elsevier BV: 67–76. doi:10.1016/j.heares.2006.03.002.

Wong, Patrick C. M., Ajith K. Uppunda, Todd B. Parrish, and Sumitrajit Dhar. 2008. "Cortical Mechanisms of Speech Perception in Noise". *Journal of Speech Language, and Hearing Research* 51 (4). American Speech Language Hearing Association: 1026–41. doi:10.1044/1092-4388(2008/075).

Bishop, Christopher W., and Lee M. Miller. 2009. "A Multisensory Cortical Network for Understanding Speech in Noise". *Journal of Cognitive Neuroscience* 21 (9). MIT Press - Journals: 1790–1804. doi:10.1162/jocn.2009.21118.

Smith, S.B., J. Krizman, C. Liu, T. White-Schwoch, T. Nicol, and N. Kraus. 2019. "Investigating Peripheral Sources of Speech-in-Noise Variability in Listeners with Normal Audiograms". *Hearing Research* 371 (January). Elsevier BV: 66–74. doi:10.1016/j.heares.2018.11.008.

Song, Judy H., Erika Skoe, Karen Banai, and Nina Kraus. 2011. "Perception of Speech in Noise: Neural Correlates". *Journal of Cognitive Neuroscience* 23 (9). MIT Press - Journals: 2268–79. doi:10.1162/jocn.2010.21556.

Selinger, Lenka, Katarzyna Zarnowiec, Marc Via, Immaculada C. Clemente, and Carles Escera. 2016. "Involvement of the Serotonin Transporter Gene in Accurate Subcortical Speech Encoding". *The Journal of Neuroscience* 36 (42). Society for Neuroscience: 10782–90. doi:10.1523/jneurosci.1595-16.2016.

Gelman, Andrew, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian Data Analysis*. Chapman and Hall/CRC.

Chen, J. J. 2003. "COMMUNICATING COMPLEX INFORMATION: THE INTERPRETATION OF STATISTICAL INTERACTION IN MULTIPLE LOGISTIC REGRESSION ANALYSIS". *American Journal of Public Health* 93 (9). American Public Health Association: 1376–a–1377. doi:10.2105/ajph.93.9.1376-a.

McElreath, Richard. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Vol. 122. CRC Press.

Bunce, John Andrew, and Richard McElreath. 2017. "Interethnic Interaction Strategic Bargaining Power, and the Dynamics of Cultural Norms". *Human Nature* 28 (4). Springer Nature: 434–56. doi:10.1007/s12110-017-9297-8.

Watanabe, Sumio. 2010. "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory". *Journal of Machine Learning Research* 11 (Dec): 3571–94.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2016. "Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC". *Statistics and Computing* 27 (5). Springer Nature: 1413–32. doi:10.1007/s11222-016-9696-4.

Hedges, Larry V., and Ingram Olkin. 1985. "Multivariate Models for Effect Sizes". In *Statistical Methods for Meta-Analysis*, 205–22. Elsevier. doi:10.1016/b978-0-08-057065-5.50015-4.

Díaz, Begoña, Florian Hintz, Stefan J. Kiebel, and Katharina von Kriegstein. 2012. "Dysfunction of the Auditory Thalamus in Developmental Dyslexia". *Proceedings of the National Academy of Sciences* 109 (34): 13841–46. doi:10.1073/pnas.1119828109.

Luo, F., Q. Wang, A. Kashani, and J. Yan. 2008. "Corticofugal Modulation of Initial Sound Processing in the Brain". *Journal of Neuroscience* 28 (45). Society for Neuroscience: 11615–21. doi:10.1523/jneurosci.3972-08.2008.

Rouiller, E. M., and F. de Ribaupierre. 1985. "Origin of Afferents to Physiologically Defined Regions of the Medial Geniculate Body of the Cat: Ventral and Dorsal Divisions". *Hearing Research* 19 (2): 97–114. doi:10.1016/0378-5955(85)90114-5.

Winer, Jeffery A. 1984. "The Human Medial Geniculate Body". *Hearing Research* 15 (3): 225–47. doi:10.1016/0378-5955(84)90031-5.

Bartlett, Edward L., Srivatsun Sadagopan, and Xiaoqin Wang. 2011. "Fine Frequency Tuning in Monkey Auditory Cortex and Thalamus". *Journal of Neurophysiology* 106 (2): 849–59. doi:10.1152/jn.00559.2010.

Bordi, Fabio, and Joseph E. LeDoux. 1994. "Response Properties of Single Units in Areas of Rat Auditory Thalamus That Project to the Amygdala". *Experimental Brain Research* 98 (2). Springer Nature America Inc: 275–86. doi:10.1007/bf00228415.

Klostermann, Fabian. 2013. "Functional Roles of the Thalamus for Language Capacities". *Frontiers in Systems Neuroscience* 7. doi:10.3389/fnsys.2013.00032.

Giraud, Anne-Lise, Christian Lorenzi, John Ashburner, Jocelyne Wable, Ingrid Johnsrude, Richard Frackowiak, and Andreas Kleinschmidt. 2000. "Representation of the Temporal Envelope of Sounds in the Human Brain". *Journal of Neurophysiology* 84 (3): 1588–98. http://jn.physiology.org/content/84/3/1588.

Wang, X., T. Lu, D. Bendor, and E. Bartlett. 2008. "Neural Coding of Temporal Information in Auditory Thalamus and Cortex". *Neuroscience*, From Cochlea to Cortex: Recent Advances in Auditory Neuroscience, 154 (1): 294–303. doi:10.1016/j.neuroscience.2008.03.065.

Foster, K. H., J. P. Gaska, M. Nagler, and D. A. Pollen. 1985. "Spatial and Temporal Frequency Selectivity of Neurones in Visual Cortical Areas V1 and V2 of the Macaque Monkey.". *The Journal of Physiology* 365 (1): 331–63. doi:10.1113/jphysiol.1985.sp015776.

Hicks, T. P., B. B. Lee, and T. R. Vidyasagar. 1983. "The Responses of Cells in Macaque Lateral Geniculate Nucleus to Sinusoidal Gratings.". *The Journal of Physiology* 337 (1): 183–200. doi:10.1113/jphysiol.1983.sp014619.

Shannon, Robert V., Fan-Gang Zeng, Vivek Kamath, John Wygonski, and Michael Ekelid. 1995. "Speech Recognition with Primarily Temporal Cues". *Science* 270 (5234): 303–4. doi:10.1126/science.270.5234.303.

Tallal, Paula, and Malcolm Piercy. 1975. "Developmental Aphasia: The Perception of Brief Vowels and Extended Stop Consonants". *Neuropsychologia* 13 (1): 69–74. doi:10.1016/0028-3932(75)90049-4.

Tallal, Paula, Steve L. Miller, Gail Bedi, Gary Byma, Xiaoqin Wang, Srikantan S. Nagarajan, Christoph Schreiner, William M. Jenkins, and Michael M. Merzenich. 1996. "Language Comprehension in Language-Learning Impaired Children Improved with Acoustically Modified Speech". *Science* 271 (5245): 81–84. doi:10.1126/science.271.5245.81.

Hayward, Katrina. 2000. *Experimental Phonetics*. Routledge. doi:10.4324/9781315842059.

Elliott, Taffeta M., and Frédéric E. Theunissen. 2009. "The Modulation Transfer Function for Speech Intelligibility". Edited by Karl J. Friston. *PLoS Computational Biology* 5 (3). Public Library of Science (PLoS): e1000302. doi:10.1371/journal.pcbi.1000302.

Ojima, Hisayuki, and Eric M. Rouiller. 2011. "Auditory Cortical Projections to the Medial Geniculate Body". In *The Auditory Cortex*, edited by Jeffery A. Winer and Christoph E. Schreiner, 171–88. Boston, MA: Springer US. doi:10.1007/978-1-4419-0074-6_8.

Andolina, Ian M., Helen E. Jones, Wei Wang, and Adam M. Sillito. 2007. "Corticothalamic Feedback Enhances Stimulus Response Precision in the Visual System". *Proceedings of the National Academy of Sciences* 104 (5): 1685–90. doi:10.1073/pnas.0609318104.

Cudeiro, Javier, and Adam M. Sillito. 2006. "Looking Back: Corticothalamic Feedback and Early Visual Processing". *Trends in Neurosciences*, Neural substrates of cognition, 29 (6): 298–306. doi:10.1016/j.tins.2006.05.002.

Ergenzinger, E. R., M. M. Glasier, J. O. Hahm, and T. P. Pons. 1998. "Cortically Induced Thalamic Plasticity in the Primate Somatosensory System". *Nature Neuroscience* 1 (3): 226–29. doi:10.1038/673.

Wang, Wei, Ian M. Andolina, Yiliang Lu, Helen E. Jones, and Adam M. Sillito. 2018. "Focal Gain Control of Thalamic Visual Receptive Fields by Layer 6 Corticothalamic Feedback". *Cerebral Cortex* 28 (1): 267–80. doi:10.1093/cercor/bhw376.

Sillito, Adam M., Helen E. Jones, George L. Gerstein, and David C. West. 1994. "Feature-Linked Synchronization of Thalamic Relay Cell Firing Induced by Feedback from the Visual Cortex". *Nature* 369 (6480): 479–82. doi:10.1038/369479a0.

Ghazanfar, Asif A., and Miguel A. L. Nicolelis. 2001. "Feature Article: The Structure and Function of Dynamic Cortical and Thalamic Receptive Fields". *Cerebral Cortex* 11 (3): 183–93. doi:10.1093/cercor/11.3.183.

Saffran, Jenny R. 2003. "Statistical Language Learning: Mechanisms and Constraints". *Current Directions in Psychological Science* 12 (4): 110–14. doi:10.1111/1467-8721.01243.

Tschentscher, Nadja, Anja Ruisinger, Helen Blank, Begoña Díaz, and Katharina von Kriegstein. 2019. "Reduced Structural Connectivity between Left Auditory Thalamus and the Motion-Sensitive Planum Temporale in Developmental Dyslexia". *The Journal of Neuroscience*, January. Society for Neuroscience, 1435–18. doi:10.1523/jneurosci.1435-18.2018.

Lee, Charles C. 2013. "Thalamic and Cortical Pathways Supporting Auditory Processing". *Brain and Language* 126 (1): 22–28. doi:10.1016/j.bandl.2012.05.004.

Llano, Daniel A., and S. Murray Sherman. 2008. "Evidence for Nonreciprocal Organization of the Mouse Auditory Thalamocortical-Corticothalamic Projection Systems". *The Journal of Comparative Neurology* 507 (2). Wiley: 1209–27. doi:10.1002/cne.21602.

Oldfield, R. C. 1971. "Edinburgh Handedness Inventory". American Psychological Association (APA). doi:10.1037/t23111-000.

Schneider, Wolfgang, Matthias Ennemoser, and Marco Schlagmüller. 2007. *Lesegeschwindigkeits- Und Verständnistest Für Die Klassen 6-12 (LGVT 6-12)*. Hogrefe Verlag.

Denckla, Martha Bridge, and Rita Rudel. 1974. "Rapid Automatized Naming of Pictured Objects Colors, Letters and Numbers by Normal Children". *Cortex* 10 (2). Elsevier BV: 186–202. doi:10.1016/s0010-9452(74)80009-2.

Denckla, Martha Bridge, and Rita G. Rudel. 1976. "Rapid 'Automatized' Naming (R.A.N.): Dyslexia Differentiated from Other Learning Disabilities". *Neuropsychologia* 14 (4). Elsevier BV: 471–79. doi:10.1016/0028-3932(76)90075-0.

Semrud-Clikeman, Margaret, Kathryn Guy, Julie D. Griffin, and George W. Hynd. 2000. "Rapid Naming Deficits in Children and Adolescents with Reading Disabilities and Attention Deficit Hyperactivity Disorder". *Brain and Language* 74 (1). Elsevier BV: 70–83. doi:10.1006/brln.2000.2337.

Baron-Cohen, Simon, Sally Wheelwright, Richard Skinner, Joanne Martin, and Emma Clubley. 2001. "The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians". *Journal of Autism and Developmental Disorders* 31 (1). Springer Nature: 5–17. doi:10.1023/a:1005653411471.

Groen, Wouter B., Linda van Orsouw, Niels ter Huurne, Sophie Swinkels, Rutger-Jan van der Gaag, Jan K. Buitelaar, and Marcel P. Zwiers. 2009. "Intact Spectral but Abnormal Temporal Processing of Auditory Stimuli in Autism". *Journal of Autism and Developmental Disorders* 39 (5): 742–50. doi:10.1007/s10803-008-0682-3.

White, Sarah, Uta Frith, Elizabeth Milne, Stuart Rosen, John Swettenham, and Franck Ramus. 2006. "A Double Dissociation between Sensorimotor Impairments and Reading Disability: A Comparison of Autistic and Dyslexic Children". *Cognitive Neuropsychology* 23 (5): 748–61. doi:10.1080/02643290500438607.

Banno, Hideki, Hiroaki Hata, Masanori Morise, Toru Takahashi, Toshio Irino, and Hideki Kawahara. 2007. "Implementation of Realtime STRAIGHT Speech Manipulation System:

Report on Its First Implementation". *Acoustical Science and Technology* 28 (3). Acoustical Society of Japan: 140–46. doi:10.1250/ast.28.140.

Brainard, David H. 1997. "The Psychophysics Toolbox". *Spatial Vision* 10 (4). Brill: 433–36. doi:10.1163/156856897x00357.

Griswold, Mark A., Peter M. Jakob, Robin M. Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. 2002. "Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA)". *Magnetic Resonance in Medicine* 47 (6). Wiley: 1202–10. doi:10.1002/mrm.10171.

Marques, José P., Tobias Kober, Gunnar Krueger, Wietske van der Zwaag, Pierre-François Van de Moortele, and Rolf Gruetter. 2010. "MP2RAGE a Self Bias-Field Corrected Sequence for Improved Segmentation and T1-Mapping at High Field". *NeuroImage* 49 (2). Elsevier BV: 1271–81. doi:10.1016/j.neuroimage.2009.10.002.

Salvatier, John, Thomas V. Wiecki, and Christopher Fonnesbeck. 2016. "Probabilistic Programming in Python Using PyMC3". *PeerJ Computer Science* 2 (April): e55. doi:10.7717/peerj-cs.55.

Hoffman, Matthew D, and Andrew Gelman. 2014. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo.". *Journal of Machine Learning Research* 15 (1): 1593–1623.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2015. "Pareto Smoothed Importance Sampling". *ArXiv Preprint ArXiv:1507.02646*.

Grove, Dan, Y. Sakamoto, M. Ishiguro, and G. Kitagawa. 1988. "Akaike Information Criterion Statistics.". *The Statistician* 37 (4/5). JSTOR: 477. doi:10.2307/2348776.

Uttl, B. 2005. "Measurement of Individual Differences: Lessons from Memory Assessment in Research and Clinical Practice.". *Psychol Sci* 16: 460–67.

Lauritzen, S. L., A. P. Dawid, B. N. Larsen, and H.-G. Leimer. 1990. "Independence Properties of Directed Markov Fields". *Networks* 20 (5). Wiley: 491–505. doi:10.1002/net.3230200503.

Jordan, Michael Irwin. 1998. *Learning in Graphical Models*. Vol. 89. Springer Science & Business Media.

Lewandowski, Daniel, Dorota Kurowicka, and Harry Joe. 2009. "Generating Random Correlation Matrices Based on Vines and Extended Onion Method". *Journal of Multivariate Analysis* 100 (9). Elsevier BV: 1989–2001. doi:10.1016/j.jmva.2009.04.008.

Avants, B. B., C. L. Epstein, M. Grossman, and J. C. Gee. 2008. "Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain". *Medical Image Analysis*, Special Issue on The Third International Workshop on Biomedical Image Registration – WBIR 2006, 12 (1): 26–41. doi:10.1016/j.media.2007.06.004.

Smith, Stephen M., Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy E.J. Behrens, Heidi Johansen-Berg, Peter R. Bannister, et al. 2004. "Advances in Functional and Structural MR Image Analysis and Implementation as FSL". *NeuroImage* 23 (January). Elsevier BV: S208–S219. doi:10.1016/j.neuroimage.2004.07.051.

Fischl, Bruce, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, et al. 2002. "Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain". *Neuron* 33 (3). Elsevier: 341–55.

Han, Xiao, and Bruce Fischl. 2007. "Atlas Renormalization for Improved Brain MR Image Segmentation across Scanner Platforms". *IEEE Transactions on Medical Imaging* 26 (4). IEEE: 479–86.

Fischl, Bruce, David H Salat, André JW Van Der Kouwe, Nikos Makris, Florent Ségonne, Brian T Quinn, and Anders M Dale. 2004. "Sequence-Independent Segmentation of Magnetic Resonance Images". *Neuroimage* 23. Elsevier: S69–S84.

Gorgolewski, Krzysztof, Christopher D. Burns, Cindee Madison, Dav Clark, Yaroslav O. Halchenko, Michael L. Waskom, and Satrajit S. Ghosh. 2011. "Nipype: A Flexible Lightweight and Extensible Neuroimaging Data Processing Framework in Python". *Frontiers in Neuroinformatics* 5. Frontiers Media SA. doi:10.3389/fninf.2011.00013.

Andersson, Jesper L.R., Chloe Hutton, John Ashburner, Robert Turner, and Karl Friston. 2001. "Modeling Geometric Deformations in EPI Time Series". *NeuroImage* 13 (5). Elsevier BV: 903–19. doi:10.1006/nimg.2001.0746.

Jezzard, Peter, and Robert S. Balaban. 1995. "Correction for Geometric Distortion in Echo Planar Images from B0 Field Variations". *Magnetic Resonance in Medicine* 34 (1). Wiley: 65–73. doi:10.1002/mrm.1910340111.

Kasper, Lars, Steffen Bollmann, Andreea O. Diaconescu, Chloe Hutton, Jakob Heinzle, Sandra Iglesias, Tobias U. Hauser, et al. 2017. "The PhysIO Toolbox for Modeling Physiological Noise in FMRI Data". *Journal of Neuroscience Methods* 276 (January): 56–72. doi:10.1016/j.jneumeth.2016.10.019.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Laurence Erlbaum Associates.