Speech perception under the tent: A domain-general predictive role for the cerebellum

Jeremy I Skipper[1] and Daniel R Lamett[1, 2]

[1]Experimental Psychology, University College London

[2]Department of Psychology, Acadia University

Jeremy I Skipper
Experimental Psychology
University College London
26 Bedford Way
London WC1H OAP
United Kingdom
E-mail: jeremy.skipper@ucl.ac.uk

Daniel R Lametti
Department of Psychology
Acadia University
15 University Ave
Wolfville NS B4P 2R6
Canada
E-mail: danielrlametti@gmail.com

Keywords: cerebellum, speech perception, speech production, domain-general, prediction, neuroimaging meta-analysis

Abstract

The role of the cerebellum in speech perception remains a mystery. Given its uniform architecture, we tested the hypothesis that it implements a domain-general predictive mechanism whose role in speech processing is determined by connectivity. We used standard and coactivation neuroimaging meta-analyses of speech perception studies *without* motor responses, with (n = 72) or without (n = 92) cerebellum activity. We compared these to speech production studies (n = 175). Results show multiple, distinct regions of perception- and production-related activity throughout the cerebellum with some overlap. Each task had distinct patterns of cortico-cerebellar connectivity. Perception regions/connections were associated with task-related terms mined from thousands of neuroimaging studies that were neither speech nor domain-specific but were prediction related. Finally, when the cerebellum was active, there was less cortical activity compared to when it was inactive, a marker of predictive processing. Results suggest that the cerebellum implements a domain-general mechanism related to prediction during passive speech perception.

Introduction

The cerebellum is a remarkable structure, having about 80% of the neurons in the brain but only around 10% of its mass[1]. Compared to other primates, it is significantly larger in humans relative to the size of the neocortex. This expansion may reflect the need for complex motor programs associated with tool use and speech, hallmarks of human evolution[2,3]. Indeed, the lateral cerebellum scales up with mammals that learn vocally, including elephants, humans, seals, dolphins and whales[4]. Consistent with more expansion in humans and its role in vocal learning, the cerebellum plays an important role in speech production[5]. This is also consistent with the view since the early 1800s that the cerebellum is primarily a motor structure, considered the organ of sexualilty by Gall, with Rolando and Flourens providing the first evidence for its more general role in motor function[6,7]. Over the last 50 years, it has also become apparent that the cerebellum plays *some* role in 'nonmotor' language domains. Activity in the cerebellum is observed during lower-level auditory functions, like speech timing and phonology and higher-level tasks involving semantics, grammar and comprehension[8–10]. However, what the cerebellum contributes to these tasks remains, to quote one 'consensus' paper, 'an ongoing enigma'[11]. Here we address questions about the role of the cerebellum during speech perception.

Domain-generality/Prediction

Though the evidence has remained elusive[12], some theories claim that the function of the cerebellum is domain-general. Thus, computations performed by the cerebellum during vocal learning and speech production should be analogously applied during speech perception. This idea is captured by the Universal Cerebellar Transform (UCT) theory[13,14] , which maintains that, because the cerebellum has a relatively homogenous, 'invariant' architecture, with repeating corticonuclear microcomplexes[15], it performs a 'consistent' computation. Any differences in this computation would be determined by variations in cerebellar location and corresponding cortico-cerebellar connectivity.

There have been a number of proposals as to the nature of the domain-general computation in the cerebellum. One suggestion from the motor control literature is that it implements internal models specifically and prediction more generally[16–20]. Internal models are neural representations of an organism's interactions with the world. These can be used to predict the sensory consequences of movements and maintain movement accuracy. When differences between predicted and actual sensory states occur, internal models are updated such that, on subsequent movements, the discrepancy between predicted and actual sensory feedback is reduced[21–23]. In evidence that the cerebellum plays a key role in this process, participants with cerebellar disruption, due to stroke or brain stimulation, show slow or altered learning in a wide-range of tasks including movement adaptation in response to

visual alterations[24–26] and following physical perturbations of movement[27–29].

Similarly, the cerebellum seems to play a key role in the predictive processing required for the maintenance of accurate speech production. Patients with cerebellar damage frequently present with a range of speech production deficits[5]. They also exhibit impaired feedforward control of speech movements (i.e., an impaired ability to update internal models)[30]. In healthy participants, altering the cerebellum with noninvasive brain stimulation alters sensorimotor adaptation during speech production[31], a result also observed during sensorimotor adaptation associated with limb movements[32,33].

Speech perception

Thus, according to UCT-like theories, the predictive role of the cerebellum in motor control should extend to nonmotor domains. Does the cerebellum play a domain-general and predictive role in speech perception[34]? Despite much theorising[5,9,11,34–39], that question is difficult to answer because the associated neurobiological research is sparse. *Pubmed* (queried May 2020) lists *six* articles with 'cerebellum' or 'cerebellar' and 'speech perception' in the title, half of which are review articles (included in the references in the prior sentence)[1]. And yet, there are thousands of neuroimaging studies involving speech perception that report cerebellar activity (as we later show).

Domain-generality

Explicit evidence for domain-generality of the cerebellum for speech perception does not exist (because speech perception studies by definition only study speech). Nonetheless, a number of task battery and meta-analyses studies more or less address this topic. In two studies, participants did task batteries, some of which were language related[40,41]. The conclusions from one of these is that, despite overlap between language and social cognition tasks, the cerebellum represents cognitive functions in a domain-specific manner because of the spatially modular appearance of these different functions[40]. Though the other task-battery study does not explicitly make this claim, it also shows specific functions mapped to specific cerebellar regions, suggesting specificity[12,41]. Five meta-analyses support the idea that the cerebellum plays specific roles in audition, speech perception and language comprehension[42–46]. Two of these use a large number of studies to profile clusters of cerebellar ROIs in terms of associated behavioural domains[45,46]. Though some of these were associated with speech and language, they were also associated with a range of other domains and subdomains. Though domain-generality is not discussed, the breadth of tasks related to specific regions of cerebellar activity

[1] The following search was used: https://pubmed.ncbi.nlm.nih.gov/?term=%28%22speech+perception%22%5BTitle%5D%29+AND+%28%28cerebellum%5BTitle%5D%29+OR+%28%22cerebellar%22%5BTitle%5D%29%29

seems at odds with a domain-specific account suggested by task battery studies.

Prediction

Prediction likely plays an important role in speech perception. Specifically, because of differences in vocal-tract lengths, accents and speaking contexts, the acoustics of identical phonemes can vary considerably. To help solve this 'lack of invariance' problem, the brain seems to use prior visual and linguistic information to predict forthcoming auditory information, constraining the interpretation of acoustic signals[47–52]. This predictive process involves a wide array of 'motor' regions also involved in speech production, though most of this work has only examined cortical motor regions[47].

The majority of cerebellar studies pertaining to prediction have focused on 'higher level' linguistic prediction (e.g., involving word meaning)[39,53–58]. Only a small number of studies have specifically investigated the role of the cerebellum in speech perception, regardless of its role in prediction. These tend to show that the cerebellum contributes timing signals to the unfolding process of speech perception. Patients with cerebellar damage show impairments in the perception of speech sounds distinguished by temporal cues like voice onset time (e.g., the initial consonant in 'pear' vs 'bear')[59]. Neuroimaging and brain stimulation work supports this link between cerebellar function and temporal processing during speech perception[60–62]. Timing is a clear component of prediction and has been theoretically and empirically linked to predictive processing in speech perception from converging lesion, ERP and fMRI results[34,38,63–66].

Tasks

Existing cerebellar speech perception studies tend to involve tasks that lack ecological validity and involve motor responses, limiting claims about domain-generality and prediction. In terms of validity, the tasks used in most studies are not particularly representative of natural speech perception. For example, in one task battery study, 'language' cerebellar regions are defined by activity associated with listening to and answering questions about Aesop's Fables *subtracted* from activity associated with reading math problems and selecting the correct answer from two alternatives[40]. In another task battery study, language processing seems to be defined using tasks like verbal working memory with letters, verb generation versus reading and/or a two alternative forced choice semantic task following sequential reading of five words. Similarly, the tasks used in the five previously mentioned meta-analyses mostly involved single word generation, repetition, reading or making semantic decisions[42–44].

Perhaps more problematic, most tasks used in these studies required participants to either read, leading to subvocal speech production, or make a meta-linguistic judgement as indicated by a button response.

Though studies often include task subtractions meant to control for motor engagement, it cannot be reasonably demonstrated that this was achieved[67,68]. Similarly, studies having motor responses on discarded trials could still cause motor activation associated with participants expectations that they will need to make responses. This is an important oversight given the historical perspective that the cerebellum is predominantly a motor structure.

Hypotheses

Theory and a small amount of empirical work suggests that the cerebellum might play a domain-general predictive role in speech perception similar to the role played in speech production. There are, however, no studies that directly address both domain-generality and predictive processing. Task-battery and meta-analyses studies suggest mixed results about domain-generality, while studies of prediction during speech perception can likely be counted on one hand and do not address domain-generality. Furthermore, there are few studies of speech perception and the cerebellum that examine natural speech perception i.e., speech perception in the complete absence of movement. To begin to address these gaps in the literature, we performed cerebellar meta-analyses, coactivation based meta-analyses and text-based functional profile analyses. Critically, we used a large number of studies that involve only 'passive' sound and speech perception without an overt motor response on any trial (N = 72). We compared these to speech production studies (N = 175; Figure 1 presents a schematic overview of this work).

Our first hypothesis was that the cerebellum plays a domain-general role in speech perception—that is, a role that is not inherently speech specific. Rather, any speech specificity partly derives from connectivity patterns that give the cerebellum a modular topological appearance. To test this, we examined regions of activity in the cerebellum related to speech perception and production. As these two processes diverge in terms of task decomposition, we anticipated non overlapping regions. However, as perception and production are both sensorimotor processes that share subprocesses, we also expected them to overlap throughout the cerebellum[69]. Using coactivation meta-analysis, we predicted that networks originating from speech perception, production and overlapping regions would have different cortical connectivity. To test for domain-generality, we analysed task-related terms mined from the abstracts of thousands of neuroimaging studies to see which of these predicted activity in cerebellar regions associated with speech perception, production or their overlap. We expected that these regions would also be associated with a wide range of other tasks that are not speech or domain-specific.

Our second hypothesis was that the domain-general role played by the cerebellum and its connections

during speech perception is related to prediction. To assess this, we tested a fundamental tenet of predictive models, that prediction results in metabolic savings because the brain has to do less processing when predictions are accurate and, correspondingly, more processing for unexpected acoustic information[50,57]. We compared speech perception-related whole-brain activity when the cerebellum was active to when it was inactive. We hypothesized that, if the cerebellum is involved in prediction during natural speech perception, there should be a greater amount of activity throughout the brain when the cerebellum is *not* active during this task.

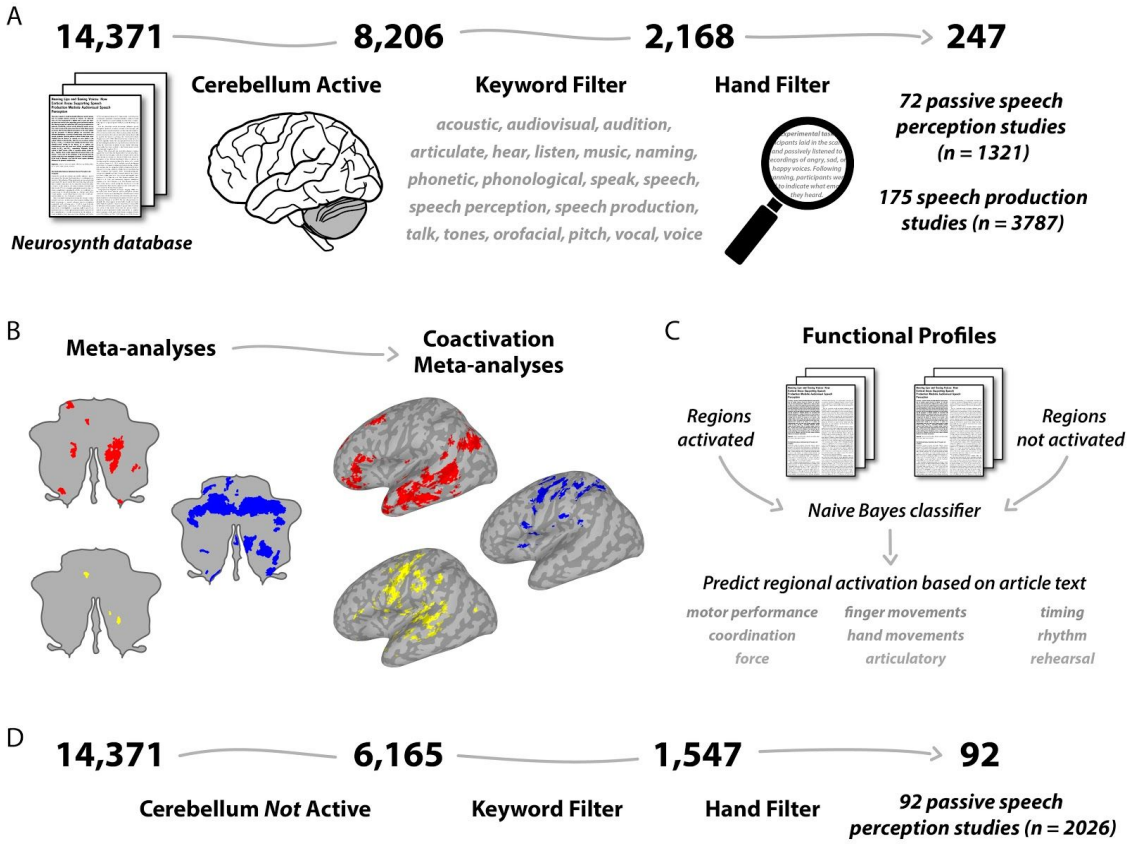

Figure 1. General procedure used for meta-analyses. A) All articles in the Neurosynth database reporting activity in a cerebellar mask were found. These were filtered by speech and language-related keywords and by hand to identify studies involving natural (i.e., 'passive') sound/speech perception and speech production. B) Meta-analyses were run using the identified studies to locate regions of activity in the cerebellum associated with speech perception (red), speech production (blue) and both tasks (yellow). Coactivation meta-analysis was then used to determine which other sub/cortical regions were functionally connected to these sets of regions. C) A classifier was trained to predict articles reporting cerebellar activity in the identified sets of speech perception, production and overlap regions based on keywords in the articles. D) In a second round of article selection, all articles in the Neurosynth database involving natural speech perception that did not report cerebellar activity were found.

Methods

Article selection

Figure 1A outlines the article selection steps. First, we created a maximum probability mask of the cerebellum using a probabilistic cerebellar atlas[70]. The latter was created by averaging the cerebellar lobule masks from 20 participants, aligned to the MNI152 template by nonlinear registration[70]. We then found all of the published articles that had activity somewhere in this mask and that appeared in the Neurosynth database (version 0.7, released July, 2018; https://github.com/neurosynth/neurosynth-data)[71]. This version contains 507,891 activation peaks or centres of mass from 14,371 studies with over 3,200 term-based features. The intersection of the cerebellum mask and database resulted in 8,206 articles (57% of the Neurosynth database).

Next, we searched Pubmed (October, 2018) for all articles that matched a set of 20 search terms (acoustic, audiovisual, audition, articulate, hear, listen, music, naming, phonetic, phonological, speak, speech, speech perception, speech production, talk, tones, orofacial, pitch, vocal, voice) and their variants (e.g., articulate, articulators, articulation, articulatory) and a set of eight Medical Subject Headings (MeSH) terms, a controlled vocabulary used by Pubmed for indexing life science articles (auditory perception, language, verbal behaviour, hearing, hearing tests, speech, speech acoustics, speech production measurements). This search returned 1,002,940 articles. We then found the intersection of these articles and the 8,206 articles in the Neurosynth database reporting cerebellar activity. This resulted in 2,168 articles with cerebellum activity that potentially involved speech, language and/or articulation (15% of the Neurosynth database).

We then went through the abstract and methods of these 2,168 articles by hand to determine if they included a 1) natural speech perception task (i.e., passive speech/sound/music perception that simply involved listening without another explicit task) or 2) a speech production task (i.e., speaking overtly/covertly or moving the articulators). We required that a number of criteria be met for studies to be included. In particular, studies that focused on reading, used patient populations, tested participants younger than 18 or focused on resting state analyses were excluded. Critically, perception studies that involved *any* motor response no matter how minor (e.g., a button press on 5% of trials to maintain alertness) were *not* included. As it remains unclear what the basic 'units' of analysis are in speech (if there are any)[47,72–74], perception studies that involved passive sound and music perception (e.g., listening to tones) were included. Of the original 2,168 articles, 72 (3.32% or 0.50% of the full Neurosynth database; N = 1321 participants) were natural speech/sound perception studies (64% speech, 12.5% sounds, 12.5% instrumental music and 11% tones) and 175 (8.07% or 1.22%; N = 3787 participants) involved speech production or articulation. See Tables S1 and S2 for a complete list of

speech perception and production articles used in meta-analyses.

Cerebellar meta-analyses

We used Neurosynth to conduct meta-analysis on our sample of natural speech perception and production studies that activate the cerebellum. Neurosynth is a database and tool for performing term-based meta-analysis[71]. As designed, it uses a form of kernel density analysis to compare activations reported in studies that frequently use selected psychological terms (e.g., 'language', 'working memory') to activations reported in studies in the rest of the database that do not use these terms. Instead of using Neurosynth to perform a term-based meta-analysis, we simply provided it with the articles found to involve natural speech perception or speech production. Neurosynth compared activations reported in the provided studies to activations reported in the rest of the Neurosynth database. The resulting cerebellum activity maps reflect activations that occur more consistently in our two samples as compared to other studies We examined baseline contrasts and overlaps. For baseline contrasts we used an FDR corrected threshold of $q < .01$ across the whole brain. We examined speech perception and production overlaps at the same individual FDR corrected thresholds. For added protection, we also required that cluster sizes be greater than 10 voxels. Results are displayed on a cerebellar flatmap using version 3.4 of the SUIT Matlab toolbox[75].

Cerebellar coactivation meta-analysis

We next did a meta-analytic coactivation analysis from the regions unique to speech perception, production and their overlap across all 14,371 neuroimaging studies in the Neurosynth database (Figure 1B). This analysis assumes that if a cerebellar region frequently coactivates with other brain regions across many studies and statistical contrasts then that region can be considered to be part of a network with the coactive brain regions. The principle here was to perform a formal contrast between studies that activate each of the three sets of regions as compared to studies that tend to activate the other sets of regions. The resulting statistical maps identify voxels throughout the brain that have a greater probability of coactivating with the identified regions. A two-way chi-square ($\chi 2$) test was used to calculate p-values for each voxel between the sets of studies. The resulting images were again thresholded using an FDR of $q < .01$. We again required that cluster sizes be greater than 10 voxels. This analysis and functional profile analyses (discussed in the next section) were based on de la Vega et al.[76,77] (https://github.com/adelavega/neurosynth-mfc/ and https://github.com/adelavega/neurosynth-lfc).

To provide descriptive functional labels of the resulting coactivation patterns, we calculated the Pearson correlation of each vectorized coactivation map with meta-analyses available in the

Neurosynth database. This resulted in *r-values* that reflect the spatial similarity between each coactivation map and other large-scale meta-analysis (see Table 1).

Cerebellar functional profile analyses

To test for domain-specificity or generality, we next generated functional profiles of the activity patterns in each of the speech perception, production and overlap regions (Figure 1C). This was done by determining which of the terms in the Neurosynth database (which are mined from the text of the 14,371 abstracts) best predicted activity in each of the three sets of cerebellum regions. Specifically, this analysis determines whether a classifier could predict if a study activated specific perception, production or overlap regions in the cerebellum given the terms mentioned in the study's title/abstract.

A naive Bayes classifier was trained to discriminate three sets of high frequency terms associated with activation in each set of regions versus a set of studies that did not produce activation in those regions. Fourfold cross-validation was used for testing and the mean score was calculated across all folds as a summary measure of performance. Models were scored using accuracy, or the fraction of samples correctly predicted. The log odds-ratio (LOR), the probability that a term is present in active versus inactive studies, from the naive Bayes models from each set of regions was used to generate the functional profiles. The LOR indicates whether a term is predictive of activation in a given cerebellar set of regions. We output the terms with LORs that predicted activation in each of the sets of cerebellar regions at an uncorrected statistical threshold of $p < .05$. To conduct functional profile analyses, we went through the terms by hand and labelled each as either anatomical, fMRI or task related. The anatomical label included any term related to brain anatomy (e.g., 'cerebellum'); the fMRI label included any terms related to the fMRI signal (e.g., 'bold signal'), stimuli (e.g., 'video') and methods (e.g., 'contrasted'); and the task label was given to any task-related terms (e.g., 'finger tapping') and their associated functions (e.g., 'speech production').

We then created four further groups of terms. First, to validate the term based approach, we labelled terms as confirmatory if they were specifically related to the cerebellum, natural/passive speech perception or speech production. Second, to determine whether our perception, production and overlap regions were actually speech specific, we labelled each term as to whether it was remotely speech related or whether it had no obvious relationship to speech. Third, to more generally examine the domain-specificity of regions, we labelled each term with the four gross psychological domains: perceptual, motor, cognitive and social/emotional. Anything that did not fit into these categories was labelled as nonspecific. Finally, to provide some insight as to what general functional role the cerebellum may play in speech processing, we created a general category for terms associated with

task demands ('expertise') and mechanisms ('prediction').

No cerebellum meta-analysis

To test the prediction hypothesis, we did a second round of article selection (Figure 1D). Specifically, we repeated the article selections steps outlined above for cerebellum articles but in the N = 6,165 articles in the Neurosynth database that do not report activation in the cerebellar mask. The intersection of these articles and the 1,002,940 articles from our original Pubmed search resulted in a sample of 1547 articles about speech, language and/or articulation. We went through the abstract and methods of these articles by hand to find those involving natural speech perception (i.e., speech/sound/music perception in the complete absence of movement). Studies that explicitly stated that they did not scan the cerebellum were eliminated. This search resulted in 92 (5.94% or 0.64% of the whole Neurosynth database; N = 2026 participants) natural speech/sound perception studies *without* cerebellum activation (64% speech, 18.5% tones, 11% sounds and 6.5% instrumental music). See Tables S1 and S3 for a complete list of speech perception articles with and without cerebellar activity used in this analysis. We did a meta-analysis of speech without a cerebellar activity, using an FDR corrected threshold of $q < .01$ across the whole brain and a minimum cluster size of 10 voxels. We examined how this meta-analysis differed from speech perception when the cerebellum was active by overlapping the speech perception with and without cerebellum activity meta-analyses. We also examined how these compared to speech production (regardless of whether the cerebellum was active or not) across the whole brain.

Results

The aim of the study was to test the hypothesis that the cerebellum has a domain-general organisation whose primary function is related to prediction, with cortico-cerebellar connectivity determining what this computation is applied during speech perception. All articles in the Neurosynth database reporting cerebellar activity were identified (N = 8,206) and from this sample studies involving natural (i.e., completely motor free 'passive') speech perception (N = 72) and speech production (N=175) were found and used in cerebellar meta-analysis (Figure 1A).

Cerebellar meta-analyses

Figure 2A shows significant cerebellar activation during natural speech perception (red) and speech production (blue). Regions activated by both perception and production are in yellow. Speech activity is distributed through much of the cerebellum in a manner that does not correspond to cerebellar lobules. Activity patterns associated with speech perception and production were nearby to each other and showed abrupt transitions. This pattern was not an artifact of statistically contrasting speech

perception and production. Figure 2B demonstrates that this arrangement, i.e., nearby perception and production regions, with small regions of overlap, remains when using less conservative corrections for multiple comparison. (See Table S4 for the volumes and MNI centre of mass coordinates for these cerebellar meta-analyses.)

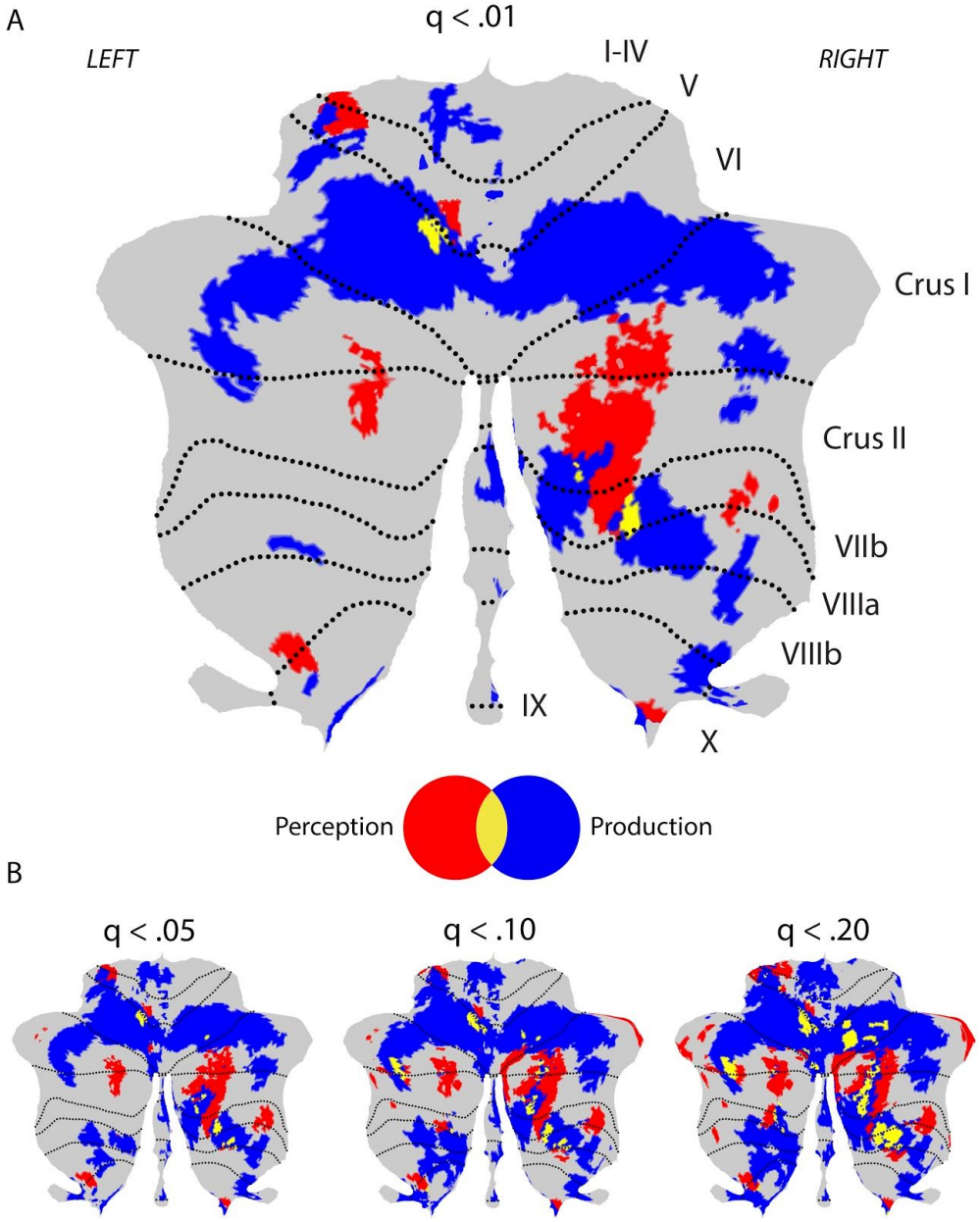

Figure 2. A) Flatmap of the cerebellum with regions of cerebellar activity associated with natural speech perception (red), speech production (blue) and both tasks (yellow). Activity patterns are significant at q < .01 false discovery rate (FDR) corrected for multiple comparisons. B) Flatmaps at increasingly less conservative FDR corrected statistical thresholds. All results are presented with a minimum cluster size of 10 voxels.

Cerebellar coactivation meta-analysis

We identified brain regions that significantly coactivate with perception, production and regions of perception-production overlap in the cerebellum across thousands of studies. Cortically, perceptual cerebellar regions had greater coactivation predominantly with the bilateral middle and anterior temporal cortex, angular gyrus and inferior frontal gyrus (Figure 3, red, 'Perception Network'). Perceptual coactivation also included the caudate bilaterally and the left thalamus. Production regions coactivate more with the precentral and postcentral sulcus and gyrus, the insula, as well as superior parietal regions (Figure 3, blue, 'Production Network'). Medially, production coactivation regions also included the superior frontal gyrus and, subcortically, the putamen and thalamus bilaterally. Finally, regions in the cerebellum activated by both perception and production coactivate with the central sulcus, inferior parietal cortex, the transverse temporal gyrus and sulcus and nearby superior temporal regions (Figure 3, yellow, 'Overlap Network'). Medially, overlap coactivation regions also included the superior frontal gyrus and, subcortically, the caudate on the left and putamen and thalamus bilaterally. These subcortical regions were in different locations than clusters in the same structures associated with perceptual and production coactivation. See Table S4 for the volumes and MNI centre of mass coordinates for regions in this coactivation meta-analysis.

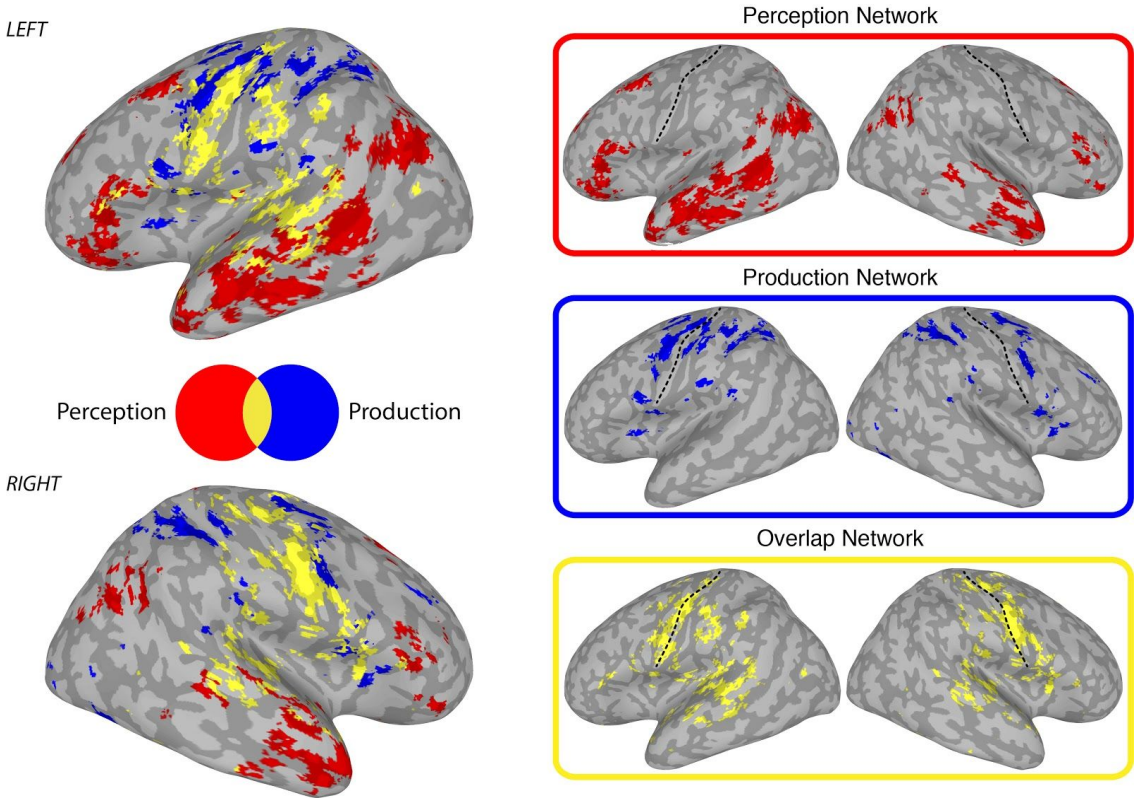

Figure 3. Coactivation networks. Regions in red depict a 'Perception Network' that was coactive with regions in the cerebellum associated with natural/passive speech perception. Regions in blue depict a

'Production Network' that was coactive with regions in the cerebellum associated with speech production. Regions in yellow depict an 'Overlap Network' that was coactive with cerebellar regions associated with both speech perception and speech production. Activity patterns are significant at $q < .01$ false discovery rate (FDR) corrected for multiple comparisons and presented with a minimum cluster size of 10 voxels. The images on the left are a composite of the three networks on the right. The dashed line divides the central sulcus in half.

The perception, production and overlap networks are similar to prior meta-analyses associated with language and semantic processing, motor planning/sequencing and sensorimotor control, respectively. However, they were also similar to meta-analyses associated with tasks that seemingly have little or nothing to do with speech. For instance, the 'Perception Network' was correlated with 'Theory of Mind' meta-analyses and the 'Overlap Network' with finger tapping (Table 1). This suggests that, although cerebellar regions were identified using studies involving only sounds and speech (without any motor response), the cerebellar networks originating from those regions are not necessarily specific to speech or language.

Table 1. The perception, production and overlap networks in Figure 3 were individually correlated with whole brain term based meta-analyses. The following table contains the top 20 correlations, showing that these networks are neither speech or domain-specific. Given the number of voxels in each correlation, these are roughly Bonferroni corrected at a *p-value* of .01/(20*3) < 0.0001.

Perception		Production		Overlap	
Meta-Analysis	**Correlation**	**Meta-Analysis**	**Correlation**	**Meta-Analysis**	**Correlation**
vi	0.41	finger movements	0.42	cerebellar	0.52
cerebellar	0.38	motor	0.41	cerebellum	0.52
lobules	0.35	execution	0.41	vi	0.51
cerebellum	0.34	premotor	0.40	lobules	0.46
mind	0.27	finger	0.40	vermis	0.34
theory mind	0.26	tapping	0.39	nuclei	0.24
mind tom	0.26	premotor cortex	0.38	coordination	0.23
tom	0.25	cerebellum	0.38	motor	0.22
mental states	0.23	movements	0.38	production	0.21
vermis	0.22	supplementary	0.35	speech production	0.20
theory	0.22	finger tapping	0.35	hemispheres	0.19
mentalizing	0.20	motor imagery	0.34	force	0.19
autobiographical	0.19	movement	0.34	cortex cerebellum	0.19
comprehension	0.19	supplementary motor	0.34	finger	0.19
nuclei	0.18	hand	0.33	motor control	0.19
sentence	0.15	sequential	0.31	movement	0.18
sentences	0.15	cerebellar	0.31	finger tapping	0.18

thinking	0.14	imagery	0.29	tapping	0.18
semantic	0.13	dorsal premotor	0.29	loop	0.17
person	0.13	handed	0.28	sensorimotor	0.17

Cerebellar functional profile analyses

Functional profiles of perception, production and overlap regions in the cerebellum were probed to assess domain-specificity or generality. We found all terms in the Neurosynth database that predicted activity in each set of regions at an uncorrected statistical threshold of $p < .05$. A 'confirmatory' qualitative analysis revealed that all three sets of perception, production and overlap regions were associated with many cerebellar anatomical and speech related terms. Speech perception regions were uniquely associated with 'listened', 'listening', 'passive' and 'passively'. They were also associated with speech and language terms and corresponding anatomical regions (e.g., 'comprehension', 'semantic', 'temporal'). The speech production and overlap regions (as compared to the speech perception regions) were uniquely associated with 'active', 'overt' and 'speech production'. Speech production regions were associated with production terms and corresponding motor regions (e.g., 'articulatory', 'motor cortex'). Activity in overlap regions was associated with terms that tended to be more sensorimotor in nature (e.g., 'sensorimotor cortex'), but were not as high-level as terms that predicted activity in speech perception regions (e.g., 'phonological'). Taken together, these results support the validity of the term based approach.

We next filtered out all anatomical and fMRI terms, leaving 168 task related terms. Figure 4 shows the top ten task-related terms (ranked by their standardized log odds ratio) that predicted activity in cerebellar regions associated with speech perception, production and the overlap of the two. Despite the qualitative differences noted above that conform to task differences, a wide range of tasks were associated with activity in each set of regions and there were also similarities in terms between the sets (e.g., 'motor' was the top term associated with each). To quantify this, we calculated the percentage of speech and non speech terms and the percentage of terms in the perceptual, cognitive, social/emotional and motor domains associated with the speech perception, production and overlap regions (Table 2). Only 23.81% of terms that predicted activity in the three sets of regions were speech specific, even though the regions were identified using a large number of studies that involved only sounds and speech . Speech related terms were roughly equally distributed across perception (5.36%), production (10.71%) and overlap regions (7.74%; $\chi^2 = 3.05$, $p > .05$).

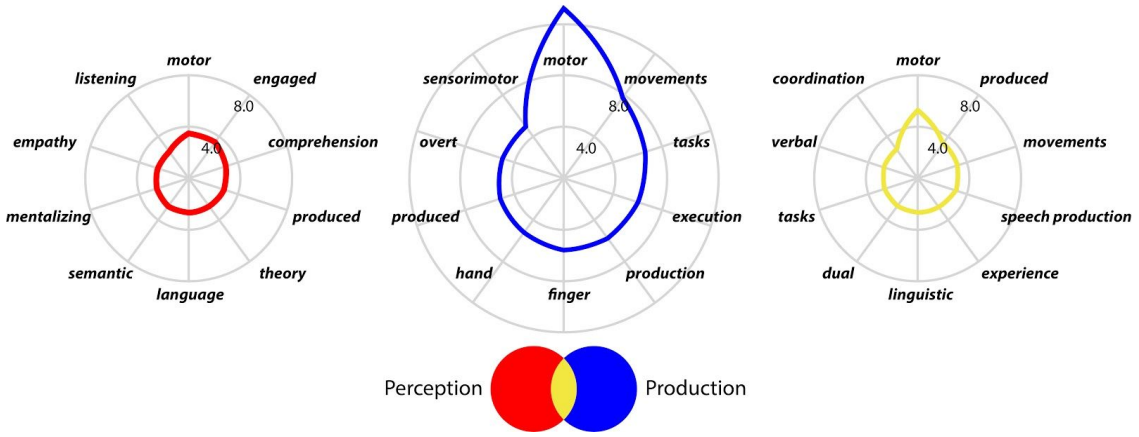

Figure 4. Top ten task related terms from articles that predicted activity in cerebellar regions associated with natural speech perception (red), speech production (blue) and regions activated by both tasks (yellow). The axes display the log odd ratio (in z-scores) for each term.

Table 2. All 168 task related terms associated with each set of regions in the cerebellum were categorised as being speech related or not or into four gross psychological domains, counted and presented as percentages showing that the three sets of cerebellar regions are not speech or domain-specific.

		Domains				
Regions	**Speech**	**Perceptual**	**Cognitive**	**Emotional**	**Motor**	**Unspecified**
Perception	5.36%	2.98%	4.17%	5.95%	2.98%	3.57%
Production	10.71%	8.33%	8.33%	0.00%	27.38%	16.67%
Overlap	7.74%	1.19%	5.95%	0.00%	8.33%	4.17%
Sum	23.81%	12.50%	18.45%	5.95%	38.69%	24.40%

Including all terms (speech and not speech), we examined whether the terms predictive of cerebellar activity in speech perception, production and overlap regions were equally distributed across our four psychological domains: perceptual, cognitive, social/emotional and motor. The distribution of terms was uniform in the case of the speech perception regions ($\chi^2 = 2.48$, $p > .05$), but not uniform in the case of the speech production ($\chi^2 = 61.57$, $p < .001$) and overlap regions ($\chi^2 = 20.15$, $p < .01$). In explanation, more than half of the terms associated with the speech production regions were from the motor category (though *not* speech/articulatory specific) and no terms from the social/emotional category were associated with either the speech production or overlap regions. About a quarter of all terms (24.40%) did not fall into a psychological domain and were classified as nonspecific. These terms were associated with each set of regions, but the distribution was not uniform ($\chi^2 = 22.56$, $p < .001$); four times as many of these terms were associated with production regions in the cerebellum.

Taken together, the results suggest that speech perception, production and overlap regions in the cerebellum are associated with a range of perceptual, cognitive and motor tasks well beyond the domains of speech and language.

Finally, about 31.55% of the terms could be labelled as either associated with demands or mechanisms. About 71.70% of these came from terms that could not be labelled with the perceptual, cognitive, social/emotional or motor categories. Demand related terms were associated with increasing task difficulty and/or expertise (e.g., 'faster'). Mechanism related terms included 'predictions' (speech perception and production regions), 'sequence' and 'sequential' (speech production related regions) and 'coordination' and 'timing' (speech production and overlap regions). These task-independent terms are consistent with prior accounts that link cerebellar functioning to predictive processing. See Table S5 for all terms and categorisations.

No cerebellum meta-analysis

Functional profiles demonstrate that regions in the cerebellum associated with speech perception, speech production and their overlap are also associated with a wide range of tasks well outside of the domain of speech, language and vocal motor control. This result supports a domain-general view of cerebellar processing. A domain-general process often attributed to the cerebellum is prediction or expectancy/timing signals that are an aspect of prediction. A hallmark of predictive processing is metabolic savings (i.e., decreases in activity when events are predicted). To test whether there is a decrease in activity when the cerebellum is active, we identified studies in the Neurosynth database involving natural speech perception that did not report activity in the cerebellum (N=92; Figure 1D). These studies were used in a second meta-analysis that compared whole-brain speech perception related activity when the cerebellum is active versus not active.

As shown in Figure 5, during speech perception, there are striking differences in brain activity as a function of whether the cerebellum is active or not. Specifically, when the cerebellum is active (in red and yellow), cortical activity related to speech perception is primarily located in the superior temporal plane, posterior inferior frontal gyrus and posterior aspect of the superior frontal gyrus. When the cerebellum is not active during speech perception (blue and yellow), there is 1.68 times more brain activity overall that is distributed over a much larger area of the brain. This activity encompasses the same regions as when the cerebellum is active and, additionally, more posterior aspects of the superior and middle temporal gyrus and sulcus, inferior parietal lobule, postcentral gyrus and sulcus, precentral gyrus and sulcus, the inferior frontal gyrus and thalamus. These additional regions are encompassed by a meta-analysis of speech production collapsed over whether or not the cerebellum is active (shown in

white outline). This result is consistent with the idea that the cerebellum plays a role in prediction during passive speech perception.

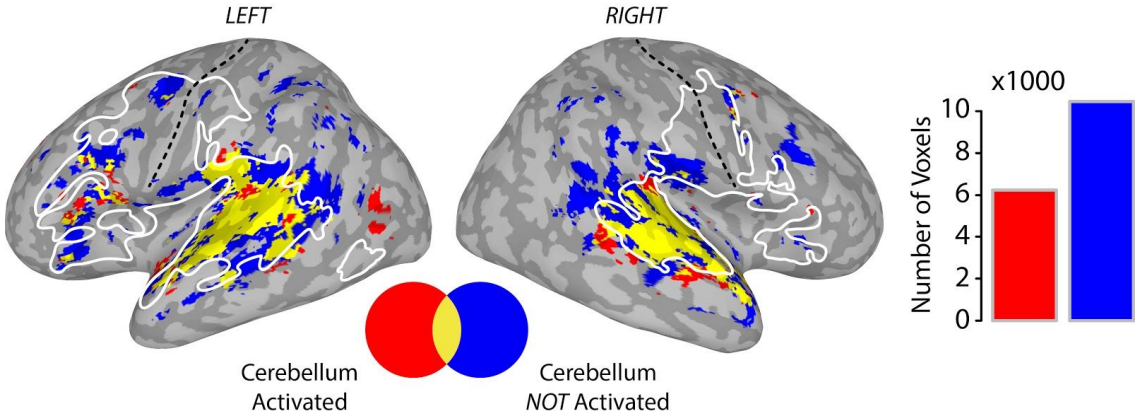

Figure 5. Extra-cerebellar activity patterns associated with natural speech perception when the cerebellum is active (red), not active (blue) and their overlap (yellow) displayed on the lateral surface of the left and right hemispheres. The white outline shows regions active during speech production. Activity patterns are significant at q < .01 FDR corrected and presented with a minimum cluster size of 10 voxels. The bar plot depicts the number of significantly activated voxels for the speech perception when the cerebellum is active (red) or not (blue).

Discussion

We tested the hypothesis that the cerebellum implements a domain-general, predictive mechanism that is deployed during speech perception as a function of connectivity. We identified studies from a large neuroimaging database reporting cerebellar activity during naturalistic (i.e., 'passive') speech perception without any motor response and compared these to speech production studies. We also found studies involving natural speech perception that did not report cerebellar activity. We used these in neuroimaging meta-analyses and coactivation meta-analyses and term-based cerebellar functional profile analyses (Figure 1; Tables S1-S3). We observed multiple regions of activity throughout the cerebellum related to both speech perception and production that were largely separate, but with some overlap (Figure 2; Table S4). These regions had unique patterns of functional connectivity (Figure 3, red, blue and yellow; Table S4). Across thousands of studies, the functional profiles of these networks (Table 1) and their seed cerebellar regions (Table 2) were not speech or domain-specific (see also Table S5). Regions of the cerebellum activated by speech perception studies were also associated with mechanistic terms like 'timing' and 'prediction'. Finally, when the cerebellum was inactive, there was more cortical activity then when it was active (Figure 5). Here, we review these results in relation to the literature on cerebellar topology and use this to discuss how the results support a domain-general, predictive account of the cerebellum in speech perception.

Topology

Modularity

Consistent with other studies using a range of tasks[12,40,41,78–80], the pattern of activity observed for speech traversed lobular boundaries (Figure 2). Though activity patterns do not conform to any obvious anatomy, they do seem to be functionally modular, with sharp boundaries between speech perception and production regions. This is consistent with studies suggesting different functions map to distinct cerebellar regions with abrupt transitions between them[40,41,78,81,82]. The sharp boundaries in our results are even more striking given that activity patterns were *not* the result of a direct contrast between speech perception and production.

From a functional modularity perspective, speech perception and production activity in different regions likely corresponds to the different subtasks that these functions can be decomposed into. Overlapping regions are likely associated with subtasks similar to both. Indeed, after task decomposition, subtasks engage different cerebellar regions. For example, working-memory can be broken down into an 'articulatory loop' and 'phonological store' and these subtasks consistently activate different cerebellar regions[42,83,84]. Conversely, language and working memory[79,85] and social mentalizing[86] may overlap in the cerebellum because they share common subtasks.

Zones

Consistent with the appearance of functional modularity, it is claimed that there is a higher level of cerebellar organisation into a nonmotor (cognitive, emotional and/or social) zone and two motor zones, corresponding to different connectivity patterns. Specifically, past work has observed a nonmotor zone in posterior lateral regions, especially crus I and II, and two motor zones, one in the anterior lobe and the other around lobule VIII[80,87,88]. These have even been further subdivided into triple nonmotor and double motor zones[40,78,80].

If one focuses only on the largest regions of activity in our data, there seems to be a medial speech perception zone (Crus I/III; HVIIb; Figure 2, red) and two speech production zones, one in the anterior and one in the posterior lobe (dorsal and ventral to the medial perception zone in Figure 2, blue). Speech perception also overlaps with speech production in two somatotopically organised motor zones at the same location of lip and tongue representations (e.g., compare Figure 2 to Figure 7B in Boillat et al. 2020)[40,80,89,90].

However, nonmotor vs two motor zones oversimplifies the observed pattern of activity[12]. Our results suggest that there are up to 10 distinct zones for speech perception and more for speech production both distributed throughout both nonmotor and motor zones. This complexity might again be attributed

to the fact that speech is a sensorimotor task that can be decomposed into many subtasks that do not neatly conform to 'cognitive' and 'motor' categories[47]. The latter cerebellar 'sandwich'[91] view (of a cognitive between motor zones) may derive from attempting to map multiple gross functions onto the cerebellum using winner-takes-all-like strategies. Our data maps a single 'function' and shows that large swathes of the cerebellum are involved in speech perception, arguing for finer task decomposition to understand individual regions of activity.

Domain-generality

Based on the uniformity of cerebellar structure, theories like the Universal Cerebellar Transform (UCT)[13,14], claim that cerebellar functions are domain-general, performing a similar computation throughout. Functional specialization in these models is determined by variations in cerebellar location and corresponding cortico-cerebellar connectivity. The fact that our data appears functionally modular while not conforming to any obvious anatomical boundaries could be consistent with either domain-specific or general accounts.

However, domain-generality was supported by the functional profile analyses. Specifically, speech perception regions in the cerebellum were not speech specific. They were also equally associated with perceptual, motor, cognitive and emotional terms generally (see Tables 1 and 2) and (in alphabetical order) attention, audition, finger and hand movements, language, memory, pain, speech, theory of mind and vision terms more specifically (see Table S1). Our speech perception studies were all sound and speech related, with no movements associated with them. It would, therefore, be hard to explain the results with a domain-specific theory as it is unlikely that every one of the nonspeech terms associated with speech perception regions have a linguistic explanation.

Connectivity

If the cerebellum does perform a domain-general role in speech perception, language comprehension and any other domain, that role must be determined by cortico-cerebellar connectivity. As we reviewed, it has been observed that there are abrupt functional divisions in the cerebellum and these have been shown to conform to structural and functional connectivity, perhaps determining the functional specialisation of those regions[13,14]. Furthermore, this connectivity is said to conform to the division of the cerebellum into (multiple) nonmotor zones and (two) motor zones[80]. Indeed, the speech perception, production and overlap sets of regions formed surprisingly distinct networks with other subcortical and cortical regions as determined by coactivation meta-analysis. In functional terms, speech perception regions tended to form networks with "higher-level" language regions, speech production regions formed networks with premotor and other "higher-level" motor/speech production

regions and overlap regions formed networks with primary auditory and primary motor regions. That is, overlap regions were distinctly sensorimotor[47,92]. However, these connectivity patterns were also domain-general (Table 1), suggesting, again, that cerebellar regions are not speech specific. They may become so in speech-only circumstances but we could not determine this as we used *all* studies to do the coactivation analysis.

Prediction

Prediction is an increasingly accepted mechanistic account of how the brain works generally[93,94] and the cerebellum more specifically, especially in the domain of motor control[19,95]. Domain-general theories suggest that the computation the cerebellum performs in any one domain should be similar to the computation it performs in others. The difference between these would be the location in the cerebellum and the attendant cortico-cerebellar connectivity. Indeed, it has been argued that the predictive role of the cerebellum plays in motor control is reused in speech perception and higher level linguistic domains like semantics, suggesting domain-generality[57,65].

Consistent with this account, our functional profiles of speech perception, production and overlap regions were all associated with the terms ‘coordination’, ‘timing’ and ‘prediction’. These regions were each associated with unique cerebellar locations and associated connectivity profiles, consistent with a domain-general account. This suggests the possibility that these regions and associated networks are predicting at different levels of analysis, perhaps corresponding to the superior/inferior motor zone and medial cognitive zones. Indeed, sensory prediction related processes that mediate sensorimotor adaptation have been demonstrated in the superior motor zone where we show speech perception/production overlap[62]. Furthermore, the distribution of activity in the medial portion of the cerebellum completely overlaps the peaks of activity in five fMRI studies of linguistic predictability[34,53,54,57,96,97]. Consistent with this, overlap networks were more associated with sensorimotor regions whereas the perception-only networks were more associated with regions mediating higher-level linguistic processes in prior studies. Indeed, perception-only regions were uniquely associated with the term ‘comprehension’.

We also generated more direct evidence for the predictive account. That is, we tested a key tenant of predictive models, that they result in less activity when predictions are accurate[50,57]. If the cerebellum plays a predictive role, we hypothesized that cortical activity should be reduced when the cerebellum was active in contrast to when it is not. Indeed, we found that when the cerebellum was active there was a nearly two-fold reduction of cortical activity during speech perception compared to when the cerebellum was not active. Furthermore, much of this reduction was in the aforementioned

sensorimotor and higher level linguistic networks associated with the speech perception/production overlap and speech perception only networks, respectively. This suggests the possibility that cortico-cerebellar and cortico-cortical predictions trade off with each other. A hallmark of cerebellar function is expertise wereas a hallmark of the neocortex is flexibility. Perhaps the cerebellum is involved in predictions associated with perceiving more well learned speech whereas the cortex more flexibly applies predictions in new contexts. Consistent with this, the cerebellum seems to play a specific role to 'automatic speech', i.e., overlearned material[36,98].

Limitations

First, our sample of natural speech perception articles includes (as a minority) studies in which participants passively listened to instrumental music, sounds and tones. We made the decision to include these studies because it is far from clear when exactly sound perception changes to speech perception. Are tones less speech-like than a 'ba'? Inclusion of these stimuli at worst reduced our chances of finding cerebellar activity (e.g., because tones are likely less well learned than 'ba's).

Second, our findings reflect the published literature. We cannot control for the quality of the included data, e.g., whether appropriate high level contrasts were used. Furthermore, results may reflect theoretical biases. For instance, predictive coding is a trending topic and there is a known predictive role of the cerebellum in motor control. Thus, there may be a bias to discuss cerebellar activity in the context of a predictive framework. However, we included a large number of studies, almost none of which had any specific interest in the cerebellum (e.g., none of the speech perception articles had 'cerebell'* in the title). They simply happened to report cerebellar activity during a task that met our criteria, likely reducing the impact of bias.

Conclusions

What role does the cerebellum play in speech perception? Our results are consistent with the perspective that the cerebellum plays a domain-general and predictive role in all functions, including speech. Furthermore, the type of prediction (e.g., motor or semantic) must be determined by task (and subtask) specific cortico-cerebellar connectivity.

Acknowledgments

This work was supported by grants from the British Academy, Corpus Christi College, Oxford, the Acadia University research fund and the Natural Sciences and Engineering Research Council (NSERC) of Canada to DRL. It was also partially supported by EPSRC EP/M026965/1 to JIS. JIS would like to thank a Banana. The authors would also like to thank S. Bobbitt, L. Wyatt and N. Guest

for help coding the manuscripts that went into the meta-analysis.

Supplementary Tables

Table S1

Passive speech perception studies with cerebellar activity.

#	Authors	Title	n	Task
1	Mutschler et al. (2007)	A rapid sound-action association effect in human insular cortex.	10	passively listened to piano melodies after either actively or passively learning the melodies
2	Milivojevic et al. (2016)	Coding of Event Nodes and Narrative Context in the Hippocampus.	19	passively watched an audiovisual movie
3	Nastase et al. (2015)	Connectivity in the human brain dissociates entropy and complexity of auditory inputs.	20	passively listened to tones of different levels of entropy and complexity
4	Noohi et al. (2017)	Functional Brain Activation in Response to a Clinical Vestibular Test Correlates with Balance.	16	passively listened to tones and received skull taps
5	Li et al. (2017)	Intranasal oxytocin, but not vasopressin, augments neural responses to toddlers in human fathers.	31	passively viewed pictures of infants and listened to infant cries
6	Li et al. (2015)	Listening to music in a risk-reward context: The roles of the temporoparietal junction and the orbitofrontal/insular cortices in reward-anticipation, reward-gain, and reward-loss.	23	passively listened to music
7	Oh et al. (2018)	Multisensory vestibular, vestibular-auditory, and auditory network effects revealed by parametric sound pressure stimulation.	20	passively listened to tone bursts
8	Wallmark et al. (2018)	Neurophysiological Effects of Trait Empathy in Music Listening.	35	passively listened to different instruments (experiment 1) and songs (experiment 2)
9	Farahani et al. (2017)	Spatiotemporal reconstruction of auditory steady-state responses to acoustic amplitude modulations: Potential sources beyond the auditory pathway.	19	passively listened to modulated white noise while watching a silent movie
10	Li et al. (2017)	The spatiotemporal pattern of pure tone processing: A single-trial EEG-fMRI study.	33	passively listened to tones
11	Lopopolo et al. (2017)	Using stochastic language models (SLM) to map lexical, syntactic, and phonological information processing in the brain.	24	passively listened to stories
12	Downar et al. (2002)	A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities.	10	passively experienced visual, auditory, and tactile stimulation
13	Ogawa et al. (2013)	Audio-visual perception of 3D cinematography: an fMRI study using condition-based and computation-based analyses.	16	passively watched audiovisual movie with alternating 3D and 2D sound and vision
14	Foxe et al. (2002)	Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study.	12	passively listened to scratching sounds and experienced

				somatosensory stimulation on fingers
15	Wilson et al. (2008)	Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension.	24	passively watched an audiovisual movie of an actor describing a story
16	Klein et al. (2006)	Bilingual brain organization: a functional magnetic resonance adaptation study.	16	listened to a list of identical words with the final word either being identical or different and either in the same or different language
17	Hesling et al. (2005)	Cerebral mechanisms of prosodic sensory integration using low-frequency bands of connected speech.	12	passively listened to either filtered or unfiltered speech
18	Wiethoff et al. (2008)	Cerebral processing of emotional prosody--influence of acoustic parameters and arousal.	24	passively listened to words spoken with different emotional prosody
19	Jeong et al. (2011)	Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation.	15	passively listened to happy or sad music while viewing happy or sad faces
20	Vander Wyk et al. (2010)	Cortical integration of audio-visual speech and non-speech stimuli.	19	passively listened to either tones or speech while viewing either ellipses or circles
21	McNealy et al. (2006)	Cracking the language code: neural mechanisms underlying speech parsing.	27	passively listened to strings of syllables linked together
22	Hertrich et al. (2011)	Cross-modal interactions during perception of audiovisual speech and nonspeech signals: an fMRI study.	20	passively listened to speech or nonspeech (tones)
23	Engel et al. (2009)	Different categories of living and non-living sound-sources activate distinct cortical networks.	20	passively listened to sounds of different categories and silently determined which category each one belonged to
24	Leech et al. (2011)	Distributed processing and cortical specialization for speech and environmental sounds in human temporal cortex.	7	passively listened to everyday environmental sounds and speech sounds
25	Halai et al. (2015)	Dual-echo fMRI can detect activations in inferior temporal lobe during intelligible speech comprehension.	20	passively listened to normal speech, rotated speech, and rotated vocoded speech
26	Chapin et al. (2010)	Dynamic emotional and neural responses to music depend on performance expression and listener experience.	21	passively listened to music
27	Chow et al. (2014)	Embodied comprehension of stories: interactions between language regions and modality-specific neural systems.	24	passively listened to stories
28	Obleser et al. (2010)	Expectancy constraints in degraded speech modulate the language comprehension network.	16	passively listened to noise-vocoded sentences
29	Woods et al. (2014)	Expert athletes activate somatosensory and motor planning regions of the brain when passively listening to familiar sports sounds.	57	passively listened to sports sounds and non-sports sounds

30	Eldar et al. (2007)	Feeling the real world: limbic response to music depends on related content.	14	passively watched film clips while listening to music of different emotional valences
31	Dick et al. (2014)	Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech.	14	passively watched audiovisual clip of speaker using gestures
32	Smirnov et al. (2014)	Fronto-parietal network supports context-dependent speech comprehension.	20	passively listened to narratives
33	Specht et al. (2003)	Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task.	12	passively listened to tones, sounds of instruments and animals, or words with one or two syllables
34	Holle et al. (2010)	Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions.	16	passively watched audiovisual clips of an actor speaking while making gestures that matched what was being said
35	Mueller et al. (2011)	Investigating brain response to music: a comparison of different fMRI acquisition schemes.	20	passively listened to manipulated and non-manipulated music clips
36	Herdener et al. (2014)	Jazz drummers recruit language-specific areas for the processing of rhythmic structure.	22	passively listened to a rhythmic sequence
37	Alluri et a. (2012)	Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm.	11	passively listened to music
38	Boldt et al. (2013)	Listening to an audio drama activates two processing networks, one for all sounds, another exclusively for speech.	13	passively watched an audiovisual movie
39	Bengtsson et al. (2009)	Listening to rhythms activates motor and premotor cortices.	17	passively listened to rhythms
40	Skipper et al. (2005)	Listening to talking faces: motor cortical activation during speech perception.	9	passively listened to stories, watched audiovisual movies of stories, or watched silent video clips
41	Lee et al. (2011)	Long-term music training tunes how the brain temporally binds signals from multiple senses.	37	passively watched audiovisual movie of a speaker uttering short sentences or a man playing piano
42	Raposo et al. (2009)	Modulation of motor and premotor cortices by actions, action words and action sentences.	22	passively listened to action words and noise bursts
43	Lambert et al. (2002)	Neural substrates of animal mental imagery: calcarine sulcus and dorsal pathway involvement--an fMRI study.	6	passively listened to animal names
44	Nagels et al. (2013)	Neural substrates of figurative language during natural speech perception: an fMRI study.	16	passively listened to stories
45	Strobel et al. (2008)	Novelty and target processing during an auditory novelty oddball: a simultaneous event-related potential and functional magnetic resonance imaging study.	14	passively listened to sine tones and environmental sounds

46	Holloway et al. (2015)	Orthographic dependency in the neural correlates of reading: evidence from audiovisual integration in English readers.	18	passively listened to letter names, number names, and phonemes
47	Chow et al. (2015)	Personal experience with narrated events modulates functional connectivity within visual and motor systems during story comprehension.	24	passively listened to stories
48	Park et al. (2013)	Personality traits modulate neural responses to emotions expressed in music.	12	passively listened to musical excerpts that conveyed happiness, sadness, or fear
49	Willems et al. (2016)	Prediction During Natural Language Comprehension.	24	passively listened to excerpts from novels
50	Campbell et al. (2007)	Primary and secondary neural networks of auditory prepulse inhibition: a functional magnetic resonance imaging study of sensorimotor gating of the human acoustic startle response.	16	passively listened to white noise (startle stimulus)
51	Lewald et al. (2008)	Processing of sound location in human cortex.	11	passively listened to white noise bursts
52	Giraud et al. (2000)	Representation of the temporal envelope of sounds in the human brain.	5	passively listened to different frequencies of sound
53	Laufer et al. (2008)	Sensory and cognitive mechanisms of change detection in the context of speech.	20	passively listened to nonwords and words
54	Londei et al. (2010)	Sensory-motor brain network connectivity for speech comprehension.	8	passively listened to words, pseudowords, or reversed words
55	Deschamps et al. (2014)	Sequencing at the syllabic and supra-syllabic levels during speech perception: an fMRI study.	15	passively listened to syllables
56	Straube et al. (2010)	Social cues, mentalizing and the neural processing of speech accompanied by gestures.	18	passively watched audio visual movie of speaker performing gestures
57	Nardo et al. (2014)	Spatial orienting in complex audiovisual environments.	26	passively watched audio visual movie portraying everyday situations
58	Yakunina et la. (2013)	Spatiotemporal Segregation of Neural Response to Auditory Stimulation: An fMRI Study Using Independent Component Analysis and Frequency-Domain Analysis.	19	passively listened to continuous broadband noise
59	Rohl et al. (2011)	Spectral loudness summation takes place in the primary auditory cortex.	22	passively listened to pink noise of different bandwidths
60	McGettigan et al. (2012)	Speech comprehension aided by multiple modalities: behavioural and neural interactions.	12	passively watched audiovisual movies of a speaker with varied auditory clarity, visual clarity, and linguistic predictability
61	Paulesu et al. (2009)	Supercalifragilisticexpialidocious: how the brain learns words never heard before.	6	passively listened to and learned a list of nonwords
62	Hugdahl et al. (2003)	The effects of attention on speech perception: an fMRI study.	13	passively listened to single vowels, three-phoneme pseudowords and three- and four-phoneme real nouns and words

63	Lockwood et al. (1999)	The functional anatomy of the normal human auditory system: responses to 0.5 and 4.0 kHz tones at varied intensities.	12	passively listened to sine-wave tones presented to right ear
64	Anders et al. (2008)	The human amygdala is sensitive to the valence of pictures and sounds irrespective of arousal: an fMRI study.	24	passively listened to non-linguistic auditory stimuli
65	Abutalebi et al. (2007)	The neural cost of the auditory perception of language switches: an event-related functional magnetic resonance imaging study in bilinguals.	12	passively listened to stories that switched between L1 and L2
66	Mayer et al. (2009)	The neural networks underlying auditory sensory gating.	21	passively listened to tones
67	Evans et al. (2014)	The pathways for intelligible speech: multivariate and univariate perspectives.	12	passively listened to natural speech, noise-vocoded speech, spectrally rotated speech, and spectrally rotated-noise-vocoded speech
68	Lyons et al. (2010)	The role of personal experience in the neural processing of action-related language.	21	passively listened to sentences
69	Bekinschtein et al. (2011)	Why clowns taste funny: the relationship between humor and semantic ambiguity.	12	passively listened to ambiguous sentences, unambiguous sentences, ambiguous jokes, low-ambiguity jokes, and signal-correlated noise
70	Osaka et al. (2004)	A word expressing affective pain activates the anterior cingulate cortex in the human brain: an fMRI study.	20	passively listened to onomatopoeic words highly suggestive of pain
71	Lepping et al. (2016)	Neural Processing of Emotional Musical and Non-musical Stimuli in Depression.	22	passively listened to music and pure tones
72	Raz et al. (2017)	Robust inter-subject audiovisual decoding in functional magnetic resonance imaging using high-dimensional regression.	23	passively viewed films or passively listened to music

Table S2

Speech production studies with cerebellar activity.

#	Authors	Title	n	Task
1	Van de Putte et al. (2018)	Neural overlap of L1 and L2 semantic representations across visual and auditory modalities: a decoding approach.	22	3 tasks: picture naming, word reading, and word listening
2	Sumiya et al. (2017)	Brain networks of social action-outcome contingency: The role of the ventral striatum in integrating signals from the sensory cortex and medial prefrontal cortex.	38	read punchlines to jokes aloud
3	Lurito et al. (2000)	Comparison of rhyming and word generation with FMRI.	5	subvocally articulated of words on a screen
4	Beaty et al. (2017)	Creative constraints: Brain activity and network dynamics underlying semantic interference during idea production.	24	word generation and cued recall tasks (spoken aloud)
5	Reilly et al. (2016)	Does the sound of a barking dog activate its corresponding visual form? An fMRI investigation of modality-specific semantic access.	18	named objects based on sounds, images, or point-light motion
6	Takashima et al. (2016)	Emergence of representations through repeated training on pronouncing novel letter combinations leads to efficient reading.	25	pronounced pseudowords written in a foreign script
7	Carey et al. (2017)	Functional brain outcomes of L2 speech learning emerge during sensorimotor transformation.	28	produced words containing native/non-native vowels
8	Pattamadilok et al. (2015)	How Early Does the Brain Distinguish between Regular Words, Irregular Words, and Pseudowords during the Reading Process? Evidence from Neurochronometric TMS.	92	read aloud while undergoing TMS
9	Danelli et al. (2017)	How many deficits in the same dyslexic brains? A behavioural and fMRI assessment of comorbidity in adult dyslexics.	23	spoonerism, digit naming, and picture naming tasks (aloud)
10	Ghazi-Saidi et al. (2015)	How native-like can you possibly get: fMRI evidence for processing accent.	12	overt picture naming
11	Flanagin et al. (2017)	Human Exploration of Enclosed Spaces through Echolocation.	11	produced calls and listened to the echoes
12	Quinn et al. (2017)	Learning and retrieving holistic and componential visual-verbal associations in reading and object naming.	20	read words and named objects aloud
13	Weiss et al. (2015)	Many ways to read your vowels--Neural processing of diacritics and vowel letters in Hebrew.	18	read words and pseudowords aloud
14	Markiewicz et al. (2016)	Mapping the cortical representation of speech sounds in a syllable repetition task.	14	listened to syllables and repeated them out loud
15	Hung et al. (2015)	Neural correlates for naming disadvantage of the dominant language in bilingual word production.	21	named pictures while switching between Chinese and English
16	Hung et al. (2015)	Neural correlates of merging number words.	40	read number words aloud
17	Geranmayeh et al. (2014)	Overlapping networks engaged during spoken language production and its cognitive control.	24	named pictures and counted aloud
18	Venezia et al. (2016)	Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech.	20	covertly rehearsed sequences of syllables

19	Gurd et al. (2002)	Posterior parietal cortex is implicated in continuous switching between verbal fluency tasks: an fMRI study with clinical implications.	11	covert word generation within a semantic category
20	Kurland et al. (2018)	Practice effects in healthy older adults: Implications for treatment-induced neuroplasticity in Aphasia.	18	object naming from line drawings
21	Berkovich-Ohana et al. (2015)	Repetitive speech elicits widespread deactivation in the human cortex: the "Mantra" effect?	23	covert word repetition - received auditory cues
22	Purcell et al. (2017)	Shared orthographic neuronal representations for spelling and reading.	16	covertly read or spelled (using keyboard) words presented auditorily
23	Provost et al. (2016)	The effect of aging on the brain network for exception word reading.	32	read pseudowords aloud
24	Reverberi et al. (2018)	The neural basis of free language choice in bilingual speakers: Disentangling language choice and language execution.	15	named objects based on line drawings while randomly switching between English and German
25	Belyk et al. (2016)	The Neural Basis of Vocal Pitch Imitation in Humans.	13	listened to and repeated melodies
26	Bilodeau-Mercure et al. (2015)	The neurobiology of speech perception decline in aging.	22	listened to and repeated syllables at different levels of intelligibility
27	Thothathiri et al. (2015)	Ventral and dorsal streams for choosing word order during sentence production.	18	viewed a scene and provided verbal descriptions of the scene using a pre-determined verb
28	Tian et al. (2016)	Mental imagery of speech implicates two mechanisms of perceptual reactivation.	18	overtly generated a syllable, imagined generating the syllable, or passively listened to the syllable
29	González-García et al. (2016)	Neural activity related to discrimination and vocal production of consonant and dissonant musical intervals.	11	passively listened to or hummed back musical intervals
30	McCrea (2007)	A functional magnetic resonance imaging study of the body schema using full human line-drawing figures in an on-line verbal naming and localization task of single body part words.	9	body part naming
31	Moore et al. (1999)	A functional neuroimaging study of the variables that generate category-specific object processing differences.	8	object naming based on line drawings (silently mouthed the words)
32	Brown et al. (2008)	A larynx area in the human motor cortex.	16	vocalization tasks: phonation (singing on a specific vowel), performing glottal stops, lip protrusion, and vertical tongue movement
33	Zubicaray et al. (2014)	A perfusion fMRI investigation of thematic and categorical context effects in the spoken production of object names.	28	picture naming presented in either homogenous or heterogenous contexts

34	Thürling et al. (2011)	Activation of the dentate nucleus in a verb generation task: A 7T MRI study.	17	reading aloud, silent reading, silent verb generation, and automatized inner speech
35	Sato et al. (2014)	Adaptive coding of orofacial and speech actions in motor and somatosensory spaces with and without overt motor behavior.	14	silent orofacial movements, audible orofacial movements, and syllable production
36	Pichon et al. (2013)	Affective and sensorimotor components of emotional prosody generation.	20	cued to read aloud with different emotional prosody
37	Shuster et al. (2005)	An fMRI investigation of covertly and overtly produced mono- and multisyllabic words.	10	produced monosyllabic and multisyllabic words covertly or overtly
38	Yoon et al. (2006)	An fMRI study of Chinese character reading and picture naming by native Korean speakers.	14	picture naming
39	Garn et al. (2009)	An fMRI study of sex differences in brain activation during object naming.	26	picture naming based online drawings: living, non-living, or scrambled images
40	Piai et al. (2013)	Attention for speaking: domain-general control from the anterior cingulate cortex in spoken word production.	23	picture-word interference task, Stroop task (overt naming)
41	Matchin et al. (2014)	Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI.	50	watched videos of speaker producing /pa/, ta/ and /ka/; had to report the sound they heard while silently articulating "/pa/…/ba/"
42	Raboyeau et al. (2010)	Brain activation and lexical learning: the impact of learning phase and word type.	10	overt object naming in French and Spanish
43	Hervais-Adel man et al. (2015)	Brain functional plasticity associated with the emergence of expertise in extreme language control.	35	listened to French and English sentences and were asked to either listen, shadow (repeat), or interpret
44	Hwang et al. (2009)	Category-specific activations during word generation reflect experiential sensorimotor modalities.	13	covert or overt word generation
45	Frings et al. (2006)	Cerebellar involvement in verb generation: an fMRI study.	16	word reading or verb generation, overt or covert
46	de Zubicaray et al. (2000)	Cerebral regions associated with verbal response initiation, suppression and strategy use.	8	overt category naming, word length judgement
47	Chen et al. (2005)	Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks.	17	subvocally read letters
48	Rauschecker et al. (2008)	Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition.	14	covertly repeated pseudowords following auditory presentation
49	Snyder et al. (2011)	Choosing our words: retrieval and selection processes recruit shared neural substrates in left ventrolateral prefrontal cortex.	18	verb generation after being presented with a noun
50	Soros et al. (2006)	Clustered functional MRI of overt speech production.	9	repeated sublexical speech sounds and

				performed oral movements without vocalization
51	Chang et al. (2009)	Common neural substrates support speech and non-speech vocal tract gestures.	34	produced a variety of articulatory gestures (glottal attacks, lip closures, bilabial bursts, tongue thrusts)
52	Halari et al. (2006)	Comparable fMRI activity with differential behavioural performance on mental rotation and overt verbal fluency tasks in healthy men and women.	19	overt verbal fluency task
53	Peck et al. (2004)	Comparison of baseline conditions to investigate syntactic production using functional magnetic resonance imaging.	10	silent sentence generation based on line drawings, passive object viewing, and object naming
54	Kemeny et al. (2005)	Comparison of continuous overt speech fMRI using BOLD and arterial spin labeling.	6	overt continuous vowel/consonant production
55	Humphreys et al. (2014)	Competitive mechanisms in sentence processing: common and distinct production and reading comprehension networks linked to the prefrontal cortex.	17	covertly read and completed sentence fragments
56	Cai et al. (2013)	Complementary hemispheric specialization for language production and visuospatial attention.	31	silent word generation
57	Abel et al. (2009)	Correct and erroneous picture naming responses in healthy subjects.	34	overt picture naming
58	Benedek et al. (2014)	Creating metaphors: the neural basis of figurative language production.	28	overt metaphor generation
59	Trapp et al. (2014)	Different neural capacity limitations for articulatory and non-articulatory maintenance of verbal information.	14	subvocally rehearsed letters (articulatory and non-articulatory conditions)
60	Nash et al. (2009)	Differential activation of the human trigeminal nuclear complex by noxious and non-noxious orofacial stimulation.	30	orofacial stimulation - participants reported and described their experience of pain while in the scanner
61	Liu et al. (2006)	Dissociation in the neural basis underlying Chinese tone and vowel production.	10	overt word naming
62	Glikmann-Johnston et al. (2015)	Distinct functional connectivity of the hippocampus during semantic and phonemic fluency.	18	covert word generation (semantic and phonemic blocks)
63	Taylor et al. (2014)	Distinct neural specializations for learning to read words and name objects.	19	overt and covert object name retrieval
64	Peeva et al. (2010)	Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network.	18	read pseudowords aloud
65	Palomar-García et al. (2015)	Do bilinguals show neural differences with monolinguals when processing their native language?	44	overt picture naming based online drawings
66	Yoo et al. (2012)	Dual routes for verbal repetition: articulation-based and acoustic-phonetic codes for pseudoword and word repetition, respectively.	22	listened to words with ambiguous vowel sounds and overtly

				repeated what they heard (overt)
67	Wildgruber et al. (1999)	Dynamic pattern of brain activation during sequencing of word strings evaluated by fMRI.	18	covertly recited the months of the year in forwards and backwards order
68	Grabowski et al. (2003)	Effects of gender on blood flow correlates of naming concrete entities.	62	overt object naming
69	Basho et al. (2007)	Effects of generation mode in fMRI adaptations of semantic fluency: paced production and overt speech.	12	overt or covert word generation based on a given category
70	Bavelier et al. (2008)	Encoding, rehearsal, and recall in signers and speakers: shared network but differential engagement.	20	speech and sign (ASL) production
71	Peck et al. (2009)	Event-related functional MRI investigation of vocal pitch variation.	12	produced "uh" at three different pitches
72	Zarate et al. (2008)	Experience-dependent neural substrates involved in vocal pitch regulation during singing.	24	singing pitches aloud with unaltered or pitch-shifted auditory feedback
73	Josse et al. (2008)	Explaining function with anatomy: language lateralization and corpus callosum size.	74	overt picture naming
74	Wattendorf et al. (2013)	Exploration of the neural correlates of ticklish laughter by functional magnetic resonance imaging.	18	tickled inside scanner and were either asked to inhibit laughter or were encouraged to make vocalizations
75	Wilson et al. (2011)	Finding your voice: a singing lesson from functional imaging.	26	covert singing and lexical retrieval
76	Golfinopoulos et al. (2011)	fMRI investigation of unexpected somatosensory feedback perturbation during speech.	13	read pseudowords aloud while wearing a device that disrupted speech movements
77	Hervais-Adelman et al. (2015)	fMRI of Simultaneous Interpretation Reveals the Neural Basis of Extreme Language Control.	43	covertly repeated sentences and translated sentences into native language
78	Grande et al. (2012)	From a concept to a word in a syntactically complete sentence: an fMRI study on spontaneous language production in an overt picture description task.	22	described scenes depicted by line drawings in detail (overt)
79	Green et al. (2015)	Frontopolar activity and connectivity support dynamic conscious augmentation of creative state.	63	presented with noun and had to generate a related verb and were cued to think creatively
80	Tremblay et al. (2013)	Functional and structural aging of the speech sensorimotor neural system: functional magnetic resonance imaging evidence.	39	watched audiovisual movie of speaker producing words and repeated what they heard
81	Abrahams et al. (2003)	Functional magnetic resonance imaging of verbal fluency and confrontation naming using compressed image acquisition to permit overt responses.	18	overt word generation based on an auditorily presented letter
82	Grabski et al. (2012)	Functional MRI assessment of orofacial articulators: neural correlates of lip, jaw, larynx, and tongue movements.	13	overt vowel production and three orofacial motor tasks

83	Schulz et al. (2005)	Functional neuroanatomy of human vocalization: an H215O PET study.	20	voiced and whispered narrative production
84	Zheng et al. (2010)	Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production.	21	producing whispered speech while either hearing this speech through headphones or while hearing masking noise
85	Vitali et al. (2005)	Generating animal and tool names: an fMRI study of effective connectivity.	10	covertly generated words belonging to the "animal" and "tool" categories
86	Yu et al. (2013)	Gum chewing inhibits the sensory processing and the propagation of stress-related information in a brain network.	24	chewed gum while listening to noise stimuli
87	Loucks et al. (2007)	Human brain activation during phonation and exhalation: common volitional control for two upper airway functions.	12	repeated of syllables while exhaling
88	Chang et al. (2014)	Implicit target substitution and sequencing for lexical tone production in Chinese: an FMRI study.	21	listened to tones and produced syllables
89	Hartwigsen et al. (2013)	Increased facilitatory connectivity from the pre-SMA to the left dorsal premotor cortex during pseudoword repetition.	17	repeated auditorily or visually presented pseudowords (overt)
90	McGettigan et al. (2015)	Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity.	21	smiled and wrinkled noses on cue
91	Collina et al. (2014)	Inside the syntactic box: the neural correlates of the functional and positional level in covert sentence production.	20	presented with three words and either had to covertly read them or covertly produce a sentence containing them
92	Seghier et al. (2008)	Inter-subject variability in the use of two different neuronal networks for reading aloud familiar words.	43	read object names aloud
93	Crescentini et al. (2010)	Item retrieval and competition in noun and verb generation: an FMRI study.	14	overt verb and noun generation
94	Hu et al. (2013)	Language aptitude for pronunciation in advanced second language (L2) learners: behavioural predictors and neural substrates.	8	listened to a sentence in L1 or L2 and repeated the sentence
95	Kell et al. (2011)	Lateralization of speech production starts in sensory cortices--a possible sensory origin of cerebral left dominance for speech.	42	covertly or overtly read sentences as they appeared on a screen / made orofacial movements (two experiments)
96	Heim et al. (2009)	Left cytoarchitectonic area 44 supports selection in the mental lexicon during language production.	23	picture naming
97	Guo et al. (2011)	Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals.	24	overtly named pictures in L1 and L2
98	Marsolais et al. (2015)	Marginal neurofunctional changes in high-performing older adults in a verbal fluency task.	28	verbal fluency: word appears on screen and participants said aloud as many words as they could belonging to that category

99	Vogel et al. (2013)	Matching is not naming: a direct comparison of lexical manipulations in explicit and implicit reading tasks.	22	read words aloud as they were presented on screen
100	van Turennout et al. (2003)	Modulation of neural activity during object naming: effects of time and practice.	10	covert object naming based on line drawings
101	Wong et al. (2011)	Motor control of jaw movements: An fMRI study of parafunctional clench and grind behavior.	19	performed orofacial movements (clench and grind on cue)
102	Vingerhoets et al. (2003)	Multilingualism: an fMRI study.	12	covertly generated words starting with a designated letter; covert picture naming; silent reading
103	Zheng et al. (2013)	Multivoxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech.	20	produced monosyllabic words with altered auditory feedback
104	Okada et al. (2000)	Naming of animals and tools: a functional magnetic resonance imaging study of categorical differences in the human brain areas commonly used for naming visually presented objects.	12	silent object naming based on drawings
105	Troiani et al. (2008)	Narrative speech production: an fMRI study using continuous arterial spin labeling.	15	described pictures aloud
106	Berken et al. (2015)	Neural activation in speech production and reading aloud in native and non-native languages.	47	read sentences aloud in L1 and L2
107	Garbin et al. (2011)	Neural bases of language switching in high and early proficient bilinguals.	19	picture naming while switching between L1 and L2
108	Lei et al. (2014)	Neural basis of language switching in the brain: fMRI evidence from Korean-Chinese early bilinguals.	8	covertly generated properties of visually presented words
109	de Diego et al. (2006)	Neural circuits subserving the retrieval of stems and grammatical features in regular and irregular verbs.	12	presented with verbs and had to either covertly produce the verb's present tense or read the verb
110	Consonni et al. (2013)	Neural convergence for language comprehension and grammatical class production in highly proficient bilinguals is independent of age of acquisition.	24	listened to sentences and overtly generated the object or action word in the sentence
111	AbdulSabur et al. (2014)	Neural correlates and network connectivity underlying narrative production and comprehension: a combined fMRI and PET study.	18	retold a story or repeated a nursery rhyme (overt)
112	Park et al. (2011)	Neural correlates in the processing of phoneme-level complexity in vowel production.	14	listened to speech sounds and performed orofacial movements without vocalization
113	Wende et al. (2012)	Neural correlates of continuous causal word generation.	18	generated words that were causally related to a cue word, were associated with cue, or rhymed with cue
114	Kircher et al. (2011)	Neural correlates of rhyming vs. lexical and semantic fluency.	15	overtly generated words starting with a

				designated letter, belonging to a designated category, or rhyming with a pseudoword
115	Christoffels et al. (2007)	Neural correlates of verbal feedback processing: an fMRI study employing overt speech.	14	covert and overt picture naming with or without pink noise masking
116	Wilson et al. (2009)	Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables.	12	overt picture naming
117	Tourville et al. (2008)	Neural mechanisms underlying auditory feedback control of speech.	11	produced monosyllabic words with altered feedback - vowel formants
118	Guenther et al. (2006)	Neural modeling and imaging of the cortical interactions underlying syllable production.	10	read vowels aloud when presented on screen
119	van de Ven et al. (2009)	Neural network of speech monitoring overlaps with overt speech production and comprehension networks: a sequential spatial and temporal ICA study.	12	overt picture naming, overt picture naming with added pink noise, covert picture naming, listening to previously recorded own speech, or listening to pink noise
120	Boukrina et al. (2013)	Neural networks underlying contributions from semantics in reading aloud.	20	read words aloud as they were presented on screen
121	Nagels et al. (2012)	Neural processing of overt word generation in healthy individuals: the effect of age and word knowledge.	56	overt verbal fluency task based on a given category
122	Wilson et al. (2006)	Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception.	12	said "ba ba ba..." on cue
123	Meinzer et al. (2009)	Neural signatures of semantic and phonemic fluency in young and old adults.	32	generated words of a designated category or that started with a designated letter
124	Graves et al. (2010)	Neural systems for reading aloud: a multiparametric approach.	20	read words aloud as they were presented on screen
125	Birn et al. (2010)	Neural systems supporting lexical search guided by letter and semantic category cues: a self-paced overt response fMRI study of verbal fluency.	14	generated words of a designated category or that started with a designated letter
126	Tremblay et al. (2011)	On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception.	21	watched video of speaker and repeated words aloud
127	Kleber et al. (2007)	Overt and imagined singing of an Italian aria.	16	overt singing and mental rehearsal of melodies
128	Herman et al. (2013)	Parsing the phonological loop: activation timing in the dorsal speech stream determines accuracy in speech reproduction.	17	repetition of syllables after hearing them produced by a speaker
129	Liljeström et al. (2008)	Perceiving and naming actions and objects.	15	silently named actions or objects from line drawings

130	Seki et al. (2004)	Phonemic manipulation in Japanese: an fMRI study.	19	listened to words and repeated them while changing the second vowel sound
131	Peschke et al. (2012)	Phonological manipulation between speech perception and production activates a parieto-frontal circuit.	23	heard pseudowords and either had to repeat them normally or phonologically manipulate them
132	Foki et al. (2008)	Probing overtly spoken language at sentential level: a comprehensive high-field BOLD-fMRI protocol reflecting everyday language demands.	23	overtly read sentences from a screen
133	Tremblay et al. (2013)	Regional heterogeneity in the processing and the production of speech in the human planum temporale.	10	whispered words as they were presented; in some trials masking was used
134	Thompson-Schill et al. (1997)	Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation.	5	covertly generated verbs related to a presented noun, classified line drawings of common objects, or compared a target word to probes and judged them for similarity (button press)
135	Hirshorn et al. (2006)	Role of the left inferior frontal gyrus in covert word retrieval: neural correlates of switching during verbal fluency.	10	covertly generated words belonging to a designated category
136	Zou et al. (2012)	Second language experience modulates functional brain network for the native language production in bimodal bilinguals.	27	overt picture naming based online drawings
137	Hocking et al. (2009)	Semantic context and visual feature effects in object naming: an fMRI study using arterial spin labeling.	18	overt object naming based on black and white photos
138	Lowell et al. (2008)	Sensory stimulation activates both motor and sensory components of the swallowing system.	14	overt and covert swallowing
139	Gauthier et al. (2009)	Sex and performance level effects on brain activation during a verbal fluency task: a functional magnetic resonance imaging study.	44	silently produced as many words as possible starting with a designated letter
140	Hamberger et al. (2013)	Shared space, separate processes: Neural activation patterns for auditory description and visual object naming in healthy adults.	13	covertly named orally described target words and pictured objects
141	Grabski et al. (2012)	Somatosensory-motor adaptation of orofacial actions in posterior parietal and ventral premotor cortices.	11	orofacial movements: lip protrusion, tongue retraction, jaw lowering
142	Callan et al. (2006)	Song and speech: brain regions involved with perception and covert production.	16	covertly produced speech and song
143	Videsott et al. (2010)	Speaking in multiple languages: neural correlates of language proficiency in multilingual word production.	20	named pictures in L1 and L2 (overt)
144	Spence et al. (2008)	Speaking of secrets and lies: the contribution of ventrolateral prefrontal cortex to vocal deception.	17	listened to sentences and responded either 'yes' or 'no', either lying or telling the truth
145	Liu et al. (2010)	Speaking words in two languages with one brain: neural overlap and dissociation.	24	overtly named objects in L2 and L2

146	Heim et al. (2008)	Specialisation in Broca's region for semantic, phonological, and syntactic fluency?	28	overt word generation: semantic, syntactic, phonological, and free
147	Meltzer et al. (2009)	Strategies for longitudinal neuroimaging studies of overt language production.	11	overt picture naming of real objects and pseudo-objects
148	Grogan et al. (2009)	Structural correlates of semantic and phonemic fluency ability in first and second languages.	59	named as many fruits or animals as possible in L1 and L2
149	Ocklenburg et al. (2013)	Structural white matter asymmetries in relation to functional asymmetries during speech perception and production.	29	silent verb generation based on a designated letter
150	de Zubicaray et al. (2007)	Support for an auto-associative model of spoken cued recall: evidence from fMRI.	16	learned word pairs and were presented with one of the words inside the scanner; had to recall its pair or respond with the word "pass"
151	Burgund et al. (2003)	Sustained and transient activity during an object-naming task: a mixed blocked and event-related fMRI study.	18	named objects aloud
152	Golestani et al. (2006)	Syntax production in bilinguals.	12	presented with words from L1 or L2 and had to read the word or covertly produce a sentence based on the word
153	McGettigan et al. (2013)	T'ain't what you say, it's the way that you say it--left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations.	23	read sentences aloud in different accents
154	Spalek et al. (2008)	Task-dependent semantic interference in language production: an fMRI study.	21	named colours or objects based on pictures
155	Kemeny et al. (2006)	Temporal dissociation of early lexical access and articulation using a delayed naming task--an FMRI study.	19	viewed pictures and overtly named the object or action shown in the photo
156	Erb et al. (2013)	The brain dynamics of rapid perceptual adaptation to adverse listening conditions.	30	listened to a vocoded sentence and repeated back as much as they could understand
157	Kleber et al. (2010)	The brain of opera singers: experience-dependent changes in functional activation.	49	sang an aria in scanner
158	Brendel et al. (2010)	The contribution of mesiofrontal cortex to the preparation and execution of repetitive syllable productions: an fMRI study.	18	heard clicking and had to produce syllables along with the clicks
159	Heim et al. (2009)	The determiner congruency effect in language production investigated with functional MRI.	16	object naming based on pictures
160	Meister et al. (2009)	The dorsal premotor cortex orchestrates concurrent speech and finger tapping movements.	14	read words aloud or repeated syllables
161	Manan et al. (2015)	The effects of aging on the brain activation pattern during a speech perception task: an fMRI study.	54	listened to words and repeated them aloud
162	Dresel et al. (2005)	The functional neuroanatomy of coordinated orofacial movements: sparse sampling fMRI of whistling.	15	whistled at a constant pitch

163	Saccuman et al. (2006)	The impact of semantic reference on word class: an fMRI study of action and object naming.	13	viewed pictures and overtly named the object or action shown in the photo
164	Hocking et al. (2008)	The influence of colour and sound on neuronal activation during visual object naming.	15	covertly named objects based on pictures
165	Riecker et al. (2008)	The influence of syllable onset complexity and syllable frequency on speech motor control.	9	read pseudowords aloud as they were presented on screen
166	Menenti et al. (2012)	The neuronal infrastructure of speaking.	24	presented pictures and a designated verb and had to describe the pictures using the verb
167	Perani et al. (2003)	The role of age of acquisition and language usage in early, high-proficient bilinguals: an fMRI study during verbal fluency.	11	covertly generated words that started with a designated letter
168	de Zubicaray et al. (2001)	The semantic interference effect in the picture-word paradigm: an event-related fMRI study employing overt responses.	8	presented with pictures and distractor words, had to name the picture while ignoring the word
169	Brown et al. (2009)	The somatotopy of speech: phonation and articulation in the human motor cortex.	16	six oral tasks: speech, monotone phonation, lip protrusion, and vertical tongue movement
170	Ellis et al. (2006)	Traces of vocabulary acquisition in the brain: Evidence from covert object naming.	14	covert picture naming based on line drawings
171	Li et al. (2013)	Two faces, two languages: an fMRI study of bilingual picture naming.	15	overt object naming based on pictures
172	Saur et al. (2008)	Ventral and dorsal pathways for language.	33	listened to and repeated words and pseudowords
173	Zarate et al. (2010)	Vocal accuracy and neural plasticity following micromelody-discrimination training.	20	presented with a target note or melody and had to sing it back
174	Huang et al. (2010)	Study-specific EPI template improves group analysis in functional MRI of young and older adults.	16	performed a word generation task
175	Häberling et al. (2016)	Cerebral asymmetry for language: Comparing production with comprehension.	94	covert speech production and word listening

Table S3

Passive speech perception studies with cerebellar not active.

#	Authors	Title	n	Task
1	Brunetti et al. (2007)	A frontoparietal network for spatial attention reorienting in the auditory domain: a human fMRI/MEG study of functional and temporal dynamics.	10	passively listened to audio stimulus of knife tapping on glass and had to localize sound
2	Czisch et al. (2009)	Acoustic Oddball during NREM Sleep: A Combined EEG/ fMRI Study	18	listened to tones; passive listening in one condition and asleep in the other
3	Cliff et al. (2013)	Aging effects on functional auditory and visual processing using fMRI with variable sensory loading.	21	passively listened to male voice reading list of nouns at 3 different rates (also had visual condition)
4	Yeshurun et al. (2017)	Amplification of local changes along the timescale processing hierarchy	54	listened to two similar stories with a few word changes
5	Büchel et al. (1999)	Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging.	11	listened to neutral auditory tones paired with an aversive sound
6	McGettigan et al. (2012)	An application of univariate and multivariate approaches in FMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes.	20	listened to simple sentences with different amplitudes and spectral modulations
7	Sikka et al. (2015)	An fMRI comparison of neural activity associated with recognition of familiar melodies in younger and older adults	40	listened to familiar tunes, unfamiliar tunes, signal correlated noise, or silence
8	Falkenberg et al. (2011)	Attention and cognitive control networks assessed in a dichotic listening fMRI study.	40	dichotic listening task
9	Dommes et al. (2009)	Auditory cortex stimulation by low-frequency tones-an fMRI study.	17	heard 3 different tones in block design
10	Javad et al. (2013)	Auditory tracts identified with combined fMRI and diffusion tractography.	14	listened to different hierarchical pitches
11	Egorova et al. (2015)	Brain basis of communicative actions in language.	20	watched videos of dialogue between two people about objects on table (either named object or requested the object)
12	Joanisse et al. (2006)	Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short-interval habituation trial paradigm.	10	listened to 4 speech syllables; conditions changed based on relationship of fourth syllable to third (degree of deviance)
13	Hasson et al. (2007)	Brain networks subserving the extraction of sentence information and its encoding to memory.	23	listened to pairs of short auditory narratives - each pair was the same story but differed by one word
14	Herdener et al. (2009)	Brain responses to auditory and visual stimulus offset: shared representations of temporal edges.	17	listened to pure sine tones with different pulse

				repetition speeds depending on condition; also exposed to visual stimuli of red light
15	Chan et al. (2009)	Can Cantonese rhymes be used in the assessment of hemispheric dominance for language?	22	listened to Cantonese rhymes (no semantic meaning) and Cantonese words
16	Harris et al. (2014)	Cerebral activations related to audition-driven performance imagery in professional musicians.	24	professional musicians and non-musicians listened to keyboard music; either imagined they were playing themselves or mentally critiqued it
17	Takeichi et al. (2009)	Comprehension of degraded speech sounds with m-sequence modulation: an fMRI study.	23	listened to parts of news broadcast; partially reversed and partially modulated speech compared to normal speech
18	Lewis et al. (2010)	Cortical networks representing object categories and high-level attributes of familiar real-world action sounds.	14	listened to 3 types of action words (human, mechanical, environmental) and had to decide silently whether a human was involved in production of action sound
19	Neuner et al. (2014)	Cortical response variation with different sound pressure levels: a combined event-related potentials and FMRI study.	16	listened to tones with varying sound pressure levels and amplitudes
20	Narain et al. (2003)	Defining a left-lateralized response specific to intelligible speech using fMRI.	11	listened to speech: normal, autotuned, spectrally rotated normal speech, and spectrally rotated autotuned speech
21	Ho et al. (2017)	Depression alters maternal extended amygdala response and functional connectivity during distress signals in attachment relationship.	29	depressed mothers and healthy mothers passively listened to different baby crying conditions: generic baby cry, imagined their own baby cry, imagined themselves crying, and control noises
22	Patel et al. (2006)	Determining hierarchical functional networks from auditory stimuli fMRI.	11	listened to different sentences spoken by different speakers
23	Herbec et al. (2015)	Differences in fMRI intersubject correlation while viewing unedited and edited videos of dance performance.	16	Watched audiovisual movies of ballet dancers filmed from different viewpoints
24	Lehmann et al. (2006)	Differential patterns of multisensory interactions in core and belt areas of human auditory cortex.	8	listened to sine wave in the presence or absence of a red light flicker

25	Friederici et al. (2010)	Disentangling syntax and intelligibility in auditory language comprehension.	17	listened to German sentences with transitive verbs; sentences were either grammatically correct or incorrect
26	Lewis et al. (2005)	Distinct cortical pathways for processing tool versus animal sounds.	20	listened to tool and animal related sounds and silently categorized the sounds
27	Meyer et al. (2005)	Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex.	12	listened to laughter, sentential speech and non vocal sounds
28	Salmi et al. (2016)	Distributed neural signatures of natural audiovisual speech and music in the human auditory cortex.	16	listened to songs; either piano pieces, vocal melodies, or spoken words
29	Zevin et al. (2010)	Domain general change detection accounts for "dishabituation" effects in temporal-parietal regions in functional magnetic resonance imaging studies of speech perception.	16	listened to syllables with varying intonation, duration and amplitude
30	Raglio et al. (2016)	Effects of active music therapy on the normal brain: fMRI based evidence.	12	listened to excerpts from music therapy sessions
31	Striem-Amit et al. (2011)	Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding FMRI.	10	listened to tone/chirp followed by silence
32	Morrison et al. (2003)	FMRI investigation of cross-cultural music comprehension.	12	subjects (musically trained and not) listened to ensemble music and a solo melody, as well as Cantonese and English speech
33	Dimitrijevic et al. (2008)	Frequency changes in a continuous tone: auditory cortical potentials.	21	listened to continuous tones that varied in frequency
34	Dick et al. (2012)	Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech.	17	exposed to either specific or non specific language, either with or without a visual gesture
35	Rogalsky et al. (2011)	Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging.	20	listened to sentences, scrambled sentences and novel piano melodies
36	Obleser et al. (2007)	Functional integration across brain regions improves speech perception under adverse listening conditions.	16	listened to spoken sentences (half were low predictability and half were high predictability)
37	Dehaene-La mber et al. (2006)	Functional segregation of cortical language areas by sentence repetition.	10	listened to a story with repeated sentences
38	Osaka et al. (2009)	Gaze-related mimic word activates the frontal eye field and related network in the human brain: an fMRI study.	20	heard gaze related words (ex. open eyes) and nonsense syllables (not gaze related)
39	Ahrens et al. (2014)	Gender differences in the temporal voice areas.	27	listened to vocal and non-vocal sounds
40	Evans et al.	Getting the Cocktail Party Started: Masking Effects in	20	listened to stories while

	(2015)	Speech Perception.		facing interference from sounds that differed in similarity to speech
41	Jäncke et al. (2004)	Hearing syllables by seeing visual stimuli.	10	listened to consonant-vowel syllables, then presented with visual stimuli and imagined syllables from the first condition
42	Jamison et al. (2005)	Hemispheric specialization for processing auditory nonspeech stimuli.	14	listened to both temporally and spectrally alternating tones
43	Brunetti et al. (2005)	Human brain activation during passive listening to sounds from different locations: an fMRI and MEG study.	11	heard stimuli from 5 different directions
44	Griffiths et al. (2000)	Human brain areas involved in the analysis of auditory movement.	4	listened to sounds containing changes in interaural phase and intensity
45	Kim et al. (2016)	Identifying Core Affect in Individuals from fMRI Responses to Dynamic Naturalistic Audiovisual Stimuli.	11	viewed audiovisual clips
46	Eisner et al. (2010)	Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations.	20	participants with normal hearing were presented with speech stimuli simulating a cochlear implant
47	Van Atteveldt et al. (2004)	Integration of letters and speech sounds in the human brain.	16	listened to the speech sounds of singular letters while looking at the corresponding letter before, after, and during the presentation of the sound
48	Koeda et al. (2015)	Interaction effect between handedness and CNTNAP2 polymorphism (rs7794745 genotype) on voice-specific frontotemporal activity in healthy individuals: an fMRI study.	108	passively listened to reverse sentences, non vocal sounds, and regular sentences
49	Cong et al. (2013)	Key issues in decomposing fMRI during naturalistic and continuous music experience with independent component analysis.	11	participants with formal music training listened passively to tango music
50	Handjaras et al. (2017)	Modality-independent encoding of individual concepts in the left parietal cortex.	20	two groups of participants (sighted and congenitally blind) listened to or viewed 30 different concrete nouns, that came from 6 different categories
51	Richardson et al. (2011)	Multiple routes from occipital to temporal cortices during reading.	17	passively listened to or viewed sentences (visual and auditory condition)
52	Obleser et al. (2006)	Multiple stages of auditory speech perception reflected in event-related FMRI.	13	listened to 4 consonants (d, t, g, k) with different speakers and in different contexts
53	Pereira et al.	Music and emotions in the brain: familiarity matters.	27	listened to songs of

	(2011)			varying familiarity
54	Tettamanti et al. (2008)	Negation in the brain: modulating action representations.	18	listened to words categorized by concreteness and polarity
55	Virtue et al. (2006)	Neural activity of inferences during story comprehension.	17	listened to short stories describing inference events
56	Heitmann et al. (2014)	Neural correlates of anticipation and processing of performance feedback in social anxiety.	34	listened to feedback on speech given prior to scanning
57	Barrós-Loscertales et al. (2013)	Neural correlates of audiovisual speech processing in a second language.	42	audiovisual speech stimuli in L1 and L2
58	Whitney et al. (2009)	Neural correlates of narrative shifts during auditory story comprehension.	16	listened to short story in scanner
59	Habermeyer et al. (2009)	Neural correlates of pre-attentive processing of pattern deviance in professional musicians.	16	listened to patterns of tones in scanner
60	Andermann et al. (2011)	Neuromagnetic representation of musical register information in human auditory cortex.	34	listened to French horn tones
61	Warson et al. (2013)	People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus.	40	watched audiovisual movie of actor performing speech and non-speech gestures
62	Warren et al. (2002)	Perception of sound-source motion by the human brain.	20	sound motion processing
63	Hall et al. (2008)	Pitch processing sites in the human auditory brain.	16	heard tones of varying pitches
64	Bremmer et al. (2001)	Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys.	8	experienced visual, auditory & somatosensory stimuli; auditory stimulus consisted of sinusoidal tones
65	Mottonen et al. (2002)	Processing of changes in visual speech in the human auditory cortex.	10	watched audiovisual movie of a speaker and listened to syllables which were either congruent or incongruent
66	Buchsbaum et al. (2005)	Reading, hearing, and the planum temporale.	17	listened to and read monosyllabic pseudowords
67	Barker et al. (2013)	Representations of pitch and slow modulation in auditory cortex.	14	listened to temporally modulated tones and tones which varied in pitch
68	Humphries et al. (2005)	Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing	12	listened to sentences that varied in prosodic information and structure
69	Miyamoto et al. (2007)	Saccular stimulation of the human cortex: a functional magnetic resonance imaging study.	12	listened to clicks of varying intensities
70	Bartsch et al. (2006)	Scanning for the scanner: FMRI of audition by read-out omissions from echo-planar imaging.	60	listened to the sound generated by acoustic planar imaging
71	Regev et al.	Selective and invariant neural responses to spoken and	38	either listened to a story,

	(2013)	written narratives.		read a story, or watched audiovisual movie of a story
72	Ghio et al. (2008)	Semantic domain-specific functional integration for action-related vs. abstract concepts.	18	listened to action related sentences
73	Joanisse et al. (2014)	Sensitivity of human auditory cortex to rapid frequency modulation revealed by multivariate representational similarity analysis.	16	listened to bursts of IRN (iterative rippled noise) with varying degrees of time delay between each iteration
74	Kansaky et al. (2000)	Sex differences in lateralization revealed in the posterior language areas.	47	listened to narrative of short essay played in reverse or cut into small parts and played at random
75	Man et al. (2012)	Sight and sound converge to form modality-invariant representations in temporoparietal cortex.	9	listened to the audio of a previously presented audiovisual movie
76	Beilock et al. (2008)	Sports experience changes the neural processing of action language.	29	hockey players and non-hockey players listened to hockey action sentences and everyday action sentences
77	Brennan et al. (2010)	Syntactic structure building in the anterior temporal lobe during natural story listening.	9	listened to Alice in Wonderland
78	Scharpf et al. (2010)	The brain's relevance detection network operates independently of stimulus modality.	24	presented with images and sounds depicting persons and faces as well as non human objects and animals
79	Naci et al. (2013)	The brain's silent messenger: using selective attention to decode human thought for brain-based communication.	16	listened to questions and potential answer words
80	Garcia et al. (2010)	The effect of stimulus context on pitch representations in the human auditory cortex.	15	listened to two different types of pitch-evoking stimuli (harmonic tone vs. complex pitch)
81	Van Atteveldt et al. (2006)	The effect of temporal asynchrony on the multisensory integration of letters and speech sounds.	8	listened to speech sounds (individual letters/phonemes) while viewing corresponding singular letters on a screen
82	Opitz et al. (1999)	The functional neuroanatomy of novelty processing: integrating ERP and fMRI results.	16	listened to a collection of pure sine tones with novel sounds included in the sequence
83	Pernet et al. (2015)	The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices.	218	listened to blocks of either vocal or non-vocal sounds
84	Molholm et al. (2004)	The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators.	20	watched a movie (without sound), while ignoring several tones differing in frequency and duration
85	Patterson et	The processing of temporal pitch and melody information	9	listened to sequences of

	al. (2002)	in auditory cortex.		notes with either random noises, random noises with a fixed pitch, diatonic melodies, or random note melodies
86	Sammler et al. (2010)	The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study.	12	listened to groups of short, unfamiliar songs with different lyrics and tunes
87	Adank et al. (2011)	The role of planum temporale in processing accent variation in spoken language comprehension.	20	listened to recordings of sentences either spoken in standard Dutch or with a novel Dutch accent; the speaker of the sentences also varied
88	Deschamps et al. (2016)	The Structural Correlates of Statistical Information Processing during Speech Perception.	20	listened to sequences of either speech or non-speech sounds that differed in their predictability
89	Jäncke et al. (1999)	The time course of the BOLD response in the human auditory cortex to acoustic stimuli of different duration.	5	listened to digitally sampled tones that varied in duration
90	Willems et al. (2006)	When language meets action: the neural integration of gesture and speech.	16	watched video clips of a woman speaking sentences, each accompanied by co-speech gestures that corresponded with a verb
91	Schirmer et al. (2008)	When vocal processing gets emotional: on the role of social orientation in relevance detection by the human amygdala.	19	presented with neutral or emotional syllable sequences
92	Domahs et al. (2012)	Where the mass counts: common cortical activation for different kinds of nonsingularity.	19	listened to a narrative with different noun types

Table S4

Cerebellar volume size in voxels and MNI x/y/z centre of mass (CM) coordinates for speech perception (red), production (blue) and overlap (yellow) regions for cerebellar meta-analyses (Figure 2) and coactivation meta-analysis (Figure 3). RL is right left; AP is anterior/posterior and IS is inferior/superior. Only clusters with a 10 or more voxels are shown for brevity.

Regions	Meta-Analysis	Volume	CM RL	CM AP	CM IS
Overlap	Coactivation Meta-Analysis	1895	-49.8	6.0	30.8

Overlap	Coactivation Meta-Analysis	1269	49.2	10.7	29.4
Overlap	Coactivation Meta-Analysis	862	52.8	27.5	5.5
Overlap	Coactivation Meta-Analysis	674	-20.7	9.5	2.0
Overlap	Coactivation Meta-Analysis	671	21.9	7.1	2.7
Overlap	Coactivation Meta-Analysis	479	15.7	69.0	-3.8
Overlap	Coactivation Meta-Analysis	422	-53.1	24.4	6.4
Overlap	Coactivation Meta-Analysis	391	3.5	-1.7	52.5
Overlap	Coactivation Meta-Analysis	190	-16.9	58.1	-3.9
Overlap	Coactivation Meta-Analysis	176	35.0	21.7	53.0
Overlap	Coactivation Meta-Analysis	92	15.7	49.9	3.1
Overlap	Coactivation Meta-Analysis	78	55.2	-15.5	11.9
Overlap	Coactivation Meta-Analysis	76	49.4	-2.2	46.9
Overlap	Coactivation Meta-Analysis	61	-44.4	29.7	31.4
Overlap	Coactivation Meta-Analysis	60	-28.0	30.7	7.3
Overlap	Coactivation Meta-Analysis	57	-11.5	42.5	56.0
Overlap	Coactivation Meta-Analysis	51	-51.6	-14.9	-16.5
Overlap	Coactivation Meta-Analysis	50	36.8	12.9	16.8
Overlap	Coactivation Meta-Analysis	49	7.8	89.5	30.7
Overlap	Coactivation Meta-Analysis	44	52.0	-13.6	-15.1
Overlap	Coactivation Meta-Analysis	39	22.3	-32.6	-8.0

Overlap	Coactivation Meta-Analysis	39	-20.9	-4.5	22.8
Overlap	Coactivation Meta-Analysis	38	-39.5	-11.8	-24.0
Overlap	Coactivation Meta-Analysis	37	-39.2	19.5	5.1
Overlap	Coactivation Meta-Analysis	33	14.1	53.3	54.8
Overlap	Coactivation Meta-Analysis	32	-25.3	60.4	8.8
Overlap	Coactivation Meta-Analysis	31	12.9	-11.9	-2.7
Overlap	Coactivation Meta-Analysis	30	25.4	32.1	3.9
Overlap	Coactivation Meta-Analysis	29	5.9	54.3	36.0
Overlap	Coactivation Meta-Analysis	28	-8.2	85.7	25.6
Overlap	Coactivation Meta-Analysis	28	21.6	26.1	73.7
Overlap	Coactivation Meta-Analysis	27	13.6	27.6	-36.9
Overlap	Coactivation Meta-Analysis	26	14.5	40.7	-36.5
Overlap	Coactivation Meta-Analysis	26	-65.4	9.8	-0.5
Overlap	Coactivation Meta-Analysis	25	59.4	-1.6	-17.8
Overlap	Coactivation Meta-Analysis	25	35.8	29.1	-13.4
Overlap	Coactivation Meta-Analysis	25	-56.0	39.8	9.8
Overlap	Coactivation Meta-Analysis	22	-59.0	8.2	-11.8
Overlap	Coactivation Meta-Analysis	22	51.7	-23.8	-1.3
Overlap	Coactivation Meta-Analysis	22	-11.7	95.0	17.4
Overlap	Coactivation Meta-Analysis	22	-21.7	9.7	55.9

Overlap	Coactivation Meta-Analysis	21	17.9	100.5	-7.6
Overlap	Coactivation Meta-Analysis	21	43.0	46.7	1.3
Overlap	Coactivation Meta-Analysis	21	-25.1	-28.5	5.6
Overlap	Coactivation Meta-Analysis	20	10.8	38.1	26.6
Overlap	Coactivation Meta-Analysis	20	14.2	10.1	44.8
Overlap	Coactivation Meta-Analysis	19	-43.3	17.3	-30.9
Overlap	Coactivation Meta-Analysis	19	-49.2	21.3	-20.1
Overlap	Coactivation Meta-Analysis	19	48.4	62.2	6.4
Overlap	Coactivation Meta-Analysis	19	-38.1	36.9	64.0
Overlap	Coactivation Meta-Analysis	18	-48.2	55.4	-7.5
Overlap	Coactivation Meta-Analysis	18	21.9	-2.7	30.3
Overlap	Coactivation Meta-Analysis	17	-1.1	67.9	28.7
Overlap	Coactivation Meta-Analysis	17	20.1	25.9	57.3
Overlap	Coactivation Meta-Analysis	15	-2.7	18.5	-41.0
Overlap	Coactivation Meta-Analysis	15	13.9	-19.1	3.8
Overlap	Coactivation Meta-Analysis	15	-0.1	8.2	5.2
Overlap	Coactivation Meta-Analysis	15	-30.6	55.5	6.7
Overlap	Coactivation Meta-Analysis	14	-22.9	78.1	39.7
Overlap	Coactivation Meta-Analysis	14	-27.3	-30.0	54.7
Overlap	Coactivation Meta-Analysis	13	-61.4	61.3	2.6

Overlap	Coactivation Meta-Analysis	12	-11.0	70.2	-6.5
Overlap	Coactivation Meta-Analysis	12	22.3	24.5	5.1
Overlap	Coactivation Meta-Analysis	11	-57.7	-5.7	-16.7
Overlap	Coactivation Meta-Analysis	10	-28.4	40.5	56.4
Overlap	Meta-Analysis	16	12.8	62.8	-13.4
Overlap	Meta-Analysis	10	-23.2	69.2	-50.4
Perception	Coactivation Meta-Analysis	3518	54.0	6.8	-10.0
Perception	Coactivation Meta-Analysis	1380	-52.4	-1.5	-23.0
Perception	Coactivation Meta-Analysis	1117	5.5	-48.7	38.6
Perception	Coactivation Meta-Analysis	875	46.9	63.1	28.9
Perception	Coactivation Meta-Analysis	516	-0.6	52.8	33.3
Perception	Coactivation Meta-Analysis	465	-50.5	61.2	27.7
Perception	Coactivation Meta-Analysis	428	38.4	-13.8	46.5
Perception	Coactivation Meta-Analysis	308	-53.3	-28.6	-1.3
Perception	Coactivation Meta-Analysis	255	1.2	-48.3	-14.4
Perception	Coactivation Meta-Analysis	238	6.5	24.5	-38.7
Perception	Coactivation Meta-Analysis	217	-23.4	-29.1	47.5
Perception	Coactivation Meta-Analysis	169	16.8	-8.5	11.5
Perception	Coactivation Meta-Analysis	147	27.8	32.5	-16.6
Perception	Coactivation Meta-Analysis	94	-47.4	29.9	-2.5

Perception	Coactivation Meta-Analysis	90	24.0	-42.7	23.5
Perception	Coactivation Meta-Analysis	73	27.3	-15.2	-32.7
Perception	Coactivation Meta-Analysis	71	-16.0	41.5	75.4
Perception	Coactivation Meta-Analysis	61	-23.1	32.5	-16.2
Perception	Coactivation Meta-Analysis	59	-50.7	-38.9	12.3
Perception	Coactivation Meta-Analysis	58	13.3	62.9	-5.4
Perception	Coactivation Meta-Analysis	48	23.6	99.2	-11.7
Perception	Coactivation Meta-Analysis	45	17.9	12.5	-31.7
Perception	Coactivation Meta-Analysis	43	-29.3	6.8	-42.8
Perception	Coactivation Meta-Analysis	43	-15.4	-13.4	15.2
Perception	Coactivation Meta-Analysis	40	-10.4	58.0	1.0
Perception	Coactivation Meta-Analysis	36	-41.8	74.5	26.5
Perception	Coactivation Meta-Analysis	35	-6.3	21.1	66.0
Perception	Coactivation Meta-Analysis	29	7.4	-32.6	-21.9
Perception	Coactivation Meta-Analysis	29	15.7	-40.9	-6.1
Perception	Coactivation Meta-Analysis	24	35.7	12.2	19.5
Perception	Coactivation Meta-Analysis	24	24.3	-39.2	47.7
Perception	Coactivation Meta-Analysis	22	-2.7	-33.6	-5.3
Perception	Coactivation Meta-Analysis	19	34.5	-16.8	-43.1
Perception	Coactivation Meta-Analysis	18	27.9	-34.7	-13.6

Perception	Coactivation Meta-Analysis	18	-1.0	33.9	33.4
Perception	Coactivation Meta-Analysis	17	22.7	10.0	32.1
Perception	Coactivation Meta-Analysis	17	-7.8	-18.7	65.2
Perception	Coactivation Meta-Analysis	16	13.2	40.1	-38.8
Perception	Coactivation Meta-Analysis	16	-33.0	30.9	4.7
Perception	Coactivation Meta-Analysis	15	-6.5	39.7	-47.7
Perception	Coactivation Meta-Analysis	14	26.6	3.5	-25.7
Perception	Coactivation Meta-Analysis	14	13.1	46.6	5.1
Perception	Coactivation Meta-Analysis	13	-65.9	46.5	-7.7
Perception	Coactivation Meta-Analysis	13	7.2	14.0	8.6
Perception	Coactivation Meta-Analysis	12	31.7	1.3	-33.8
Perception	Coactivation Meta-Analysis	12	-0.9	41.0	54.0
Perception	Coactivation Meta-Analysis	11	-3.0	31.3	-54.7
Perception	Coactivation Meta-Analysis	10	18.8	23.2	-23.0
Perception	Coactivation Meta-Analysis	10	4.1	44.9	9.0
Perception	Coactivation Meta-Analysis	10	-42.8	70.3	40.2
Perception	Coactivation Meta-Analysis	10	9.8	30.4	70.2
Perception	Meta-Analysis	200	-20.3	78.9	-41.7
Perception	Meta-Analysis	46	-26.3	75.2	-32.7
Perception	Meta-Analysis	43	-42.3	54.6	-56.1

Perception	Meta-Analysis	29	12.2	40.0	-55.9
Perception	Meta-Analysis	28	11.2	62.3	-9.3
Perception	Meta-Analysis	23	20.6	77.3	-35.7
Perception	Meta-Analysis	14	31.3	64.4	-59.7
Perception	Meta-Analysis	13	20.9	33.2	-26.6
Production	Coactivation Meta-Analysis	2116	42.3	18.5	44.7
Production	Coactivation Meta-Analysis	1143	0.3	-3.3	53.5
Production	Coactivation Meta-Analysis	679	-38.7	43.2	48.8
Production	Coactivation Meta-Analysis	522	-34.1	6.2	55.1
Production	Coactivation Meta-Analysis	402	16.4	11.3	5.4
Production	Coactivation Meta-Analysis	275	-54.7	-5.7	35.2
Production	Coactivation Meta-Analysis	242	-36.6	63.4	-15.8
Production	Coactivation Meta-Analysis	222	-13.4	16.2	7.0
Production	Coactivation Meta-Analysis	135	-49.0	-9.1	3.8
Production	Coactivation Meta-Analysis	123	-36.0	-20.2	4.2
Production	Coactivation Meta-Analysis	99	34.0	62.1	-16.2
Production	Coactivation Meta-Analysis	93	-59.8	27.2	21.1
Production	Coactivation Meta-Analysis	77	-20.3	64.1	55.7
Production	Coactivation Meta-Analysis	70	-22.1	-3.1	6.2
Production	Coactivation Meta-Analysis	70	57.1	17.8	18.7

Production	Coactivation Meta-Analysis	70	-36.3	-37.6	26.9
Production	Coactivation Meta-Analysis	66	32.3	-18.0	2.7
Production	Coactivation Meta-Analysis	44	-31.9	86.9	0.2
Production	Coactivation Meta-Analysis	37	48.2	37.5	23.9
Production	Coactivation Meta-Analysis	23	-8.4	75.5	-10.3
Production	Coactivation Meta-Analysis	22	-16.9	1.6	-2.1
Production	Coactivation Meta-Analysis	13	-43.8	64.8	4.0
Production	Coactivation Meta-Analysis	12	14.1	67.5	59.4
Production	Meta-Analysis	1929	-9.7	62.9	-22.4
Production	Meta-Analysis	184	-26.2	63.2	-53.7
Production	Meta-Analysis	118	40.7	64.6	-31.5
Production	Meta-Analysis	82	-12.2	77.7	-44.5
Production	Meta-Analysis	48	-14.0	39.5	-46.8
Production	Meta-Analysis	25	-5.8	58.9	-31.0
Production	Meta-Analysis	21	7.8	45.0	-17.6
Production	Meta-Analysis	20	-25.2	47.0	-46.3
Production	Meta-Analysis	16	-11.5	67.9	-38.8
Production	Meta-Analysis	15	10.8	53.9	-32.9
Production	Meta-Analysis	13	-0.5	57.4	-0.6
Production	Meta-Analysis	12	29.0	31.3	-27.0

Production	Meta-Analysis	10	1.0	45.4	-7.4

Table S5

All terms associated with each set of regions in the cerebellum were categorised as being speech related or not, into five gross domains of psychological inquiry and other categorisations. These are summarised in Table 2.

Regions	Terms	LOR-Z	P-Value	FDR P-Value	Speech/Not Speech	Task/Anatomical/fMR	Domain	Specific Domain	Other
Overlap	dual	2.71	0.01	0.09	Not Speech	Task	Cognitive	Attention	Demands
Overlap	engage	2.22	0.03	0.21	Not Speech	Task	Unspecific	Unspecific	Demands
Overlap	motor	5.17	0.00	0.00	Not Speech	Task	Motor	Unspecific	Unspecific
Overlap	engaged	2.21	0.03	0.21	Not Speech	Task	Unspecific	Unspecific	Demands
Overlap	engages	2.23	0.03	0.20	Not Speech	Task	Unspecific	Unspecific	Demands
Overlap	episodic	2.21	0.03	0.21	Not Speech	Task	Cognitive	Memory	Demands
Overlap	simultaneous	2.36	0.02	0.17	Not Speech	Task	Unspecific	Unspecific	Demands
Overlap	coordination	2.71	0.01	0.09	Not Speech	Task	Unspecific	Unspecific	Mechanism
Overlap	tasks	2.71	0.01	0.09	Not Speech	Task	Unspecific	Unspecific	Demands
Overlap	produced	3.27	0.00	0.02	Not Speech	Task	Motor	Unspecific	Unspecific
Overlap	movements	3.20	0.00	0.03	Not Speech	Task	Motor	Unspecific	Unspecific
Overlap	speech production	3.09	0.00	0.04	Speech	Task	Motor	Speech	Confirmatory
Overlap	production	2.99	0.00	0.05	Not Speech	Task	Motor	Unspecific	Unspecific
Overlap	experience	2.82	0.00	0.07	Not Speech	Task	Perceptual	Unspecific	Unspecific
Overlap	linguistic	2.75	0.01	0.08	Speech	Task	Cognitive	Language	Unspecific

Overlap	verbal	2.71	0.01	0.09	Speech	Task	Motor	Language	Unspecific
Overlap	words	2.71	0.01	0.09	Speech	Task	Cognitive	Language	Unspecific
Overlap	reading	2.68	0.01	0.09	Speech	Task	Cognitive	Language	Unspecific
Overlap	semantic	2.61	0.01	0.11	Speech	Task	Cognitive	Language	Unspecific
Overlap	hand	2.52	0.01	0.13	Not Speech	Task	Motor	Hand	Unspecific
Overlap	speech	2.50	0.01	0.13	Speech	Task	Motor	Speech	Confirmatory
Overlap	phonological	2.50	0.01	0.13	Speech	Task	Cognitive	Speech	Unspecific
Overlap	sensorimotor	2.48	0.01	0.14	Not Speech	Task	Motor	Unspecific	Unspecific
Overlap	motor control	2.48	0.01	0.14	Not Speech	Task	Motor	Unspecific	Unspecific
Overlap	read	2.43	0.01	0.15	Speech	Task	Cognitive	Language	Unspecific
Overlap	lexical	2.36	0.02	0.17	Speech	Task	Cognitive	Language	Unspecific
Overlap	generation	2.29	0.02	0.19	Not Speech	Task	Motor	Unspecific	Unspecific
Overlap	overt	2.27	0.02	0.19	Not Speech	Task	Motor	Unspecific	Confirmatory
Overlap	auditory	2.13	0.03	0.24	Speech	Task	Perceptual	Auditory	Confirmatory
Overlap	word	2.08	0.04	0.26	Speech	Task	Cognitive	Language	Unspecific
Overlap	timing	2.41	0.02	0.15	Not Speech	Task	Unspecific	Unspecific	Mechanism
Overlap	generated	1.97	0.05	0.30	Not Speech	Task	Motor	Unspecific	Unspecific
Overlap	spoken	1.96	0.05	0.31	Speech	Task	Motor	Speech	Unspecific
Perception	motor	3.46	0.00	0.01	Not Speech	Task	Motor	Unspecific	Unspecific
Perception	difficulty	2.49	0.01	0.13	Not Speech	Task	Unspecific	Unspecific	Demands

Perception	engaged	3.44	0.00	0.01	Not Speech	Task	Unspecific	Unspecific	Demands
Perception	episodic	2.42	0.02	0.15	Not Speech	Task	Cognitive	Memory	Demands
Perception	comprehension	3.01	0.00	0.05	Speech	Task	Cognitive	Language	Unspecific
Perception	produced	2.85	0.00	0.06	Not Speech	Task	Motor	Unspecific	Unspecific
Perception	theory	2.73	0.01	0.09	Not Speech	Task	Emotional	ToM	Unspecific
Perception	successful	2.31	0.02	0.18	Not Speech	Task	Unspecific	Unspecific	Demands
Perception	language	2.68	0.01	0.09	Speech	Task	Cognitive	Language	Unspecific
Perception	semantic	2.68	0.01	0.09	Speech	Task	Cognitive	Language	Unspecific
Perception	mentalizing	2.60	0.01	0.11	Not Speech	Task	Emotional	ToM	Unspecific
Perception	mental states	2.56	0.01	0.12	Not Speech	Task	Emotional	ToM	Unspecific
Perception	empathy	2.53	0.01	0.13	Not Speech	Task	Emotional	ToM	Unspecific
Perception	listening	2.52	0.01	0.13	Speech	Task	Perceptual	Auditory	Confirmatory
Perception	self	2.44	0.01	0.14	Not Speech	Task	Emotional	ToM	Unspecific
Perception	motor control	2.43	0.02	0.15	Not Speech	Task	Motor	Unspecific	Unspecific
Perception	affective	2.42	0.02	0.15	Not Speech	Task	Emotional	Unspecific	Unspecific
Perception	linguistic	2.34	0.02	0.17	Speech	Task	Cognitive	Language	Unspecific
Perception	mental	2.31	0.02	0.18	Not Speech	Task	Emotional	ToM	Unspecific
Perception	predictions	2.06	0.04	0.27	Not Speech	Task	Unspecific	Unspecific	Mechanism
Perception	reading	2.28	0.02	0.19	Speech	Task	Cognitive	Language	Unspecific
Perception	listened	2.24	0.03	0.20	Speech	Task	Perceptual	Auditory	Confirmatory

Perception	theory mind	2.23	0.03	0.20	Not Speech	Task	Emotional	ToM	Unspecific
Perception	sentence	2.23	0.03	0.20	Speech	Task	Cognitive	Language	Unspecific
Perception	experience	2.21	0.03	0.21	Not Speech	Task	Perceptual	Unspecific	Unspecific
Perception	passively	2.17	0.03	0.22	Not Speech	Task	Unspecific	Unspecific	Confirmatory
Perception	speech	2.04	0.04	0.28	Speech	Task	Motor	Speech	Confirmatory
Perception	perspective	2.01	0.04	0.29	Not Speech	Task	Perceptual	Visual	Unspecific
Perception	production	1.99	0.05	0.30	Not Speech	Task	Motor	Unspecific	Unspecific
Perception	mind	1.98	0.05	0.30	Not Speech	Task	Emotional	ToM	Unspecific
Perception	passive	1.97	0.05	0.30	Not Speech	Task	Unspecific	Unspecific	Confirmatory
Perception	experiences	1.97	0.05	0.30	Not Speech	Task	Perceptual	Unspecific	Unspecific
Perception	pain	1.96	0.05	0.31	Not Speech	Task	Emotional	Pain	Unspecific
Production	motor	13.21	0.00	0.00	Not Speech	Task	Motor	Unspecific	Unspecific
Production	movements	7.77	0.00	0.00	Not Speech	Task	Motor	Unspecific	Unspecific
Production	execution	6.04	0.00	0.00	Not Speech	Task	Motor	Unspecific	Unspecific
Production	production	5.78	0.00	0.00	Not Speech	Task	Motor	Unspecific	Unspecific
Production	finger	5.58	0.00	0.00	Not Speech	Task	Motor	Hand	Unspecific
Production	hand	5.24	0.00	0.00	Not Speech	Task	Motor	Hand	Unspecific
Production	produced	5.18	0.00	0.00	Not Speech	Task	Motor	Unspecific	Unspecific
Production	overt	4.99	0.00	0.00	Not Speech	Task	Motor	Unspecific	Confirmatory
Production	sensorimotor	4.98	0.00	0.00	Not Speech	Task	Motor	Unspecific	Unspecific

Production	movement	4.91	0.00	0.00	Not Speech	Task	Motor	Unspecific	Unspecific
Production	motor control	4.59	0.00	0.00	Not Speech	Task	Motor	Unspecific	Unspecific
Production	observation	4.37	0.00	0.00	Not Speech	Task	Perceptual	Visual	Unspecific
Production	coordination	4.72	0.00	0.00	Not Speech	Task	Unspecific	Unspecific	Mechanism
Production	finger movements	4.13	0.00	0.00	Not Speech	Task	Motor	Hand	Unspecific
Production	tapping	3.95	0.00	0.00	Not Speech	Task	Motor	Hand	Unspecific
Production	speech production	3.87	0.00	0.00	Speech	Task	Motor	Speech	Confirmatory
Production	imagery	3.81	0.00	0.00	Not Speech	Task	Perceptual	Visual	Unspecific
Production	rehearsal	3.57	0.00	0.01	Speech	Task	Motor	Unspecific	Unspecific
Production	verbal	3.47	0.00	0.01	Speech	Task	Motor	Language	Unspecific
Production	reading	3.46	0.00	0.01	Speech	Task	Cognitive	Language	Unspecific
Production	phonological	3.45	0.00	0.01	Speech	Task	Cognitive	Speech	Unspecific
Production	learning	3.39	0.00	0.02	Not Speech	Task	Cognitive	Learning	Unspecific
Production	motor imagery	3.38	0.00	0.02	Not Speech	Task	Perceptual	Unspecific	Unspecific
Production	finger tapping	3.37	0.00	0.02	Not Speech	Task	Motor	Hand	Unspecific
Production	motor task	3.26	0.00	0.02	Not Speech	Task	Motor	Unspecific	Unspecific
Production	generation	3.18	0.00	0.03	Not Speech	Task	Motor	Unspecific	Unspecific
Production	pronounced	3.03	0.00	0.04	Speech	Task	Motor	Speech	Unspecific
Production	articulatory	2.92	0.00	0.06	Speech	Task	Motor	Mouth	Unspecific

Production	actions	2.91	0.00	0.06	Not Speech	Task	Motor	Unspecific	Unspecific
Production	imagined	2.91	0.00	0.06	Not Speech	Task	Perceptual	Visual	Unspecific
Production	output	2.90	0.00	0.06	Not Speech	Task	Motor	Unspecific	Unspecific
Production	covert	2.85	0.00	0.06	Not Speech	Task	Cognitive	Speech	Unspecific
Production	visually	2.84	0.00	0.07	Not Speech	Task	Perceptual	Visual	Unspecific
Production	attend	2.31	0.02	0.18	Not Speech	Task	Cognitive	Attention	Demands
Production	behavioral responses	2.33	0.02	0.17	Not Speech	Task	Unspecific	Unspecific	Demands
Production	button	1.97	0.05	0.30	Not Speech	Task	Motor	Hand	Demands
Production	demands	2.06	0.04	0.27	Not Speech	Task	Unspecific	Unspecific	Demands
Production	visual	2.67	0.01	0.09	Not Speech	Task	Perceptual	Visual	Unspecific
Production	visuomotor	2.67	0.01	0.09	Not Speech	Task	Motor	Unspecific	Unspecific
Production	difficulty	2.63	0.01	0.10	Not Speech	Task	Unspecific	Unspecific	Demands
Production	dual	2.37	0.02	0.16	Not Speech	Task	Cognitive	Attention	Demands
Production	engaged	3.80	0.00	0.00	Not Speech	Task	Unspecific	Unspecific	Demands
Production	musical	2.59	0.01	0.11	Speech	Task	Perceptual	Unspecific	Unspecific
Production	imitation	2.58	0.01	0.11	Not Speech	Task	Motor	Unspecific	Unspecific
Production	hands	2.58	0.01	0.11	Not Speech	Task	Motor	Hand	Unspecific
Production	engages	2.15	0.03	0.23	Not Speech	Task	Unspecific	Unspecific	Demands
Production	episodic	2.24	0.03	0.20	Not Speech	Task	Cognitive	Memory	Demands
Production	expertise	2.57	0.01	0.12	Not Speech	Task	Unspecific	Unspecific	Demands
Production	hand movements	2.53	0.01	0.12	Not Speech	Task	Motor	Hand	Unspecific

Production	fusiform face	2.52	0.01	0.13	Not Speech	Task	Perceptual	Visual	Unspecific
Production	mirror	2.51	0.01	0.13	Not Speech	Task	Motor	Unspecific	Unspecific
Production	faster	1.96	0.05	0.31	Not Speech	Task	Unspecific	Unspecific	Demands
Production	fluency	2.15	0.03	0.23	Speech	Task	Unspecific	Unspecific	Demands
Production	force	3.55	0.00	0.01	Not Speech	Task	Motor	Unspecific	Demands
Production	interference	2.29	0.02	0.19	Not Speech	Task	Unspecific	Unspecific	Demands
Production	load	2.30	0.02	0.18	Not Speech	Task	Unspecific	Unspecific	Demands
Production	sensory	2.41	0.02	0.15	Not Speech	Task	Perceptual	Unspecific	Unspecific
Production	executed	2.41	0.02	0.15	Not Speech	Task	Motor	Unspecific	Unspecific
Production	generated	2.38	0.02	0.16	Not Speech	Task	Motor	Unspecific	Unspecific
Production	motor response	2.04	0.04	0.28	Not Speech	Task	Motor	Unspecific	Demands
Production	naming	2.38	0.02	0.16	Speech	Task	Motor	Speech	Unspecific
Production	predictions	2.02	0.04	0.29	Not Speech	Task	Unspecific	Unspecific	Mechanism
Production	motor responses	1.97	0.05	0.30	Not Speech	Task	Motor	Unspecific	Demands
Production	action observation	2.35	0.02	0.17	Not Speech	Task	Perceptual	Unspecific	Unspecific
Production	vocal	2.34	0.02	0.17	Speech	Task	Motor	Mouth	Unspecific
Production	read	2.31	0.02	0.18	Speech	Task	Cognitive	Language	Unspecific
Production	sequence	3.69	0.00	0.01	Not Speech	Task	Unspecific	Unspecific	Mechanism
Production	paced	3.43	0.00	0.01	Not Speech	Task	Unspecific	Unspecific	Demands
Production	sequential	2.47	0.01	0.14	Not Speech	Task	Unspecific	Unspecific	Mechanism

Production	performance	4.16	0.00	0.00	Not Speech	Task	Unspecific	Unspecific	Demands
Production	repetition	2.28	0.02	0.19	Not Speech	Task	Motor	Unspecific	Unspecific
Production	speech	2.28	0.02	0.19	Speech	Task	Motor	Speech	Confirmatory
Production	index finger	2.28	0.02	0.19	Not Speech	Task	Motor	Hand	Unspecific
Production	press	2.62	0.01	0.11	Not Speech	Task	Motor	Hand	Demands
Production	response times	2.41	0.02	0.15	Not Speech	Task	Unspecific	Unspecific	Demands
Production	serial	2.35	0.02	0.17	Not Speech	Task	Unspecific	Unspecific	Demands
Production	simultaneously	2.68	0.01	0.09	Not Speech	Task	Unspecific	Unspecific	Demands
Production	initiation	2.19	0.03	0.22	Not Speech	Task	Motor	Unspecific	Unspecific
Production	face ffa	2.19	0.03	0.22	Not Speech	Task	Perceptual	Visual	Unspecific
Production	oral	2.17	0.03	0.22	Speech	Task	Motor	Mouth	Unspecific
Production	object	2.17	0.03	0.22	Not Speech	Task	Perceptual	Visual	Unspecific
Production	cognitive functions	2.17	0.03	0.22	Not Speech	Task	Cognitive	Unspecific	Unspecific
Production	semantic	2.16	0.03	0.23	Speech	Task	Cognitive	Language	Unspecific
Production	motor function	2.15	0.03	0.23	Not Speech	Task	Motor	Unspecific	Unspecific
Production	skill	2.71	0.01	0.09	Not Speech	Task	Unspecific	Unspecific	Demands
Production	skills	2.56	0.01	0.12	Not Speech	Task	Unspecific	Unspecific	Demands
Production	sustained	2.04	0.04	0.28	Not Speech	Task	Cognitive	Attention	Demands
Production	preparation	2.10	0.04	0.25	Not Speech	Task	Motor	Unspecific	Unspecific

Production	ffa	2.09	0.04	0.26	Not Speech	Task	Perceptual	Visual	Unspecific
Production	action	2.07	0.04	0.27	Not Speech	Task	Motor	Unspecific	Unspecific
Production	task	5.96	0.00	0.00	Not Speech	Task	Unspecific	Unspecific	Demands
Production	task conditions	2.15	0.03	0.23	Not Speech	Task	Unspecific	Unspecific	Demands
Production	task difficulty	2.04	0.04	0.28	Not Speech	Task	Unspecific	Unspecific	Demands
Production	task performance	2.04	0.04	0.28	Not Speech	Task	Unspecific	Unspecific	Demands
Production	tasks	6.64	0.00	0.00	Not Speech	Task	Unspecific	Unspecific	Demands
Production	rhythm	2.03	0.04	0.28	Speech	Task	Motor	Unspecific	Unspecific
Production	timing	3.97	0.00	0.00	Not Speech	Task	Unspecific	Unspecific	Mechanism
Production	internal	2.00	0.05	0.29	Not Speech	Task	Unspecific	Unspecific	Unspecific
Production	pseudowords	1.99	0.05	0.30	Speech	Task	Perceptual	Speech	Unspecific
Production	verbal working	3.18	0.00	0.03	Speech	Task	Cognitive	Language	Demands
Production	working	3.15	0.00	0.03	Not Speech	Task	Cognitive	Memory	Demands
Production	working memory	3.06	0.00	0.04	Not Speech	Task	Cognitive	Memory	Demands

References

1. Herculano-Houzel, S., Catania, K., Manger, P. R. & Kaas, J. H. Mammalian Brains Are Made of These: A Dataset of the Numbers and Densities of Neuronal and Nonneuronal Cells in the Brain of Glires, Primates, Scandentia, Eulipotyphlans, Afrotherians and Artiodactyls, and Their Relationship with Body Mass. *Brain Behav. Evol.* **86**, 145–163 (2015).
2. Barton, R. A. & Venditti, C. Rapid Evolution of the Cerebellum in Humans and Other Great Apes. *Curr. Biol.* **27**, 1249–1250 (2017).
3. MacLeod, C. E., Zilles, K., Schleicher, A., Rilling, J. K. & Gibson, K. R. Expansion of the neocerebellum in Hominoidea. *J. Hum. Evol.* **44**, 401–429 (2003).
4. Smaers, J. B., Turner, A. H., Gómez-Robles, A. & Sherwood, C. C. A cerebellar substrate for cognition evolved multiple times independently in mammals. *Elife* **7**, (2018).
5. Ackermann, H., Mathiak, K. & Riecker, A. The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data. *Cerebellum* **6**, 202–213 (2007).
6. Glickstein, M., Strata, P. & Voogd, J. Cerebellum: history. *Neuroscience* **162**, 549–559 (2009).
7. Macklis, R. M. & Macklis, J. D. Historical and phrenologic reflections on the nonmotor functions of the cerebellum: love under the tent? *Neurology* **42**, 928–932 (1992).
8. Ackermann, H. & Brendel, B. Chapter 7 - Cerebellar Contributions to Speech and Language. in *Neurobiology of Language* (eds. Hickok, G. & Small, S. L.) 73–84 (Academic Press, 2016).
9. Mariën, P. & Manto, M. *The Linguistic Cerebellum*. (Academic Press, 2015).
10. Mariën, P. & Borgatti, R. Chapter 11 - Language and the cerebellum. in *Handbook of Clinical Neurology* (eds. Manto, M. & Huisman, T. A. G. M.) vol. 154 181–202 (Elsevier, 2018).
11. Mariën, P. *et al.* Consensus paper: Language and the cerebellum: an ongoing enigma. *Cerebellum* **13**, 386–410 (2014).
12. Diedrichsen, J., King, M., Hernandez-Castillo, C., Sereno, M. & Ivry, R. B. Universal Transform

or Multiple Functionality? Understanding the Contribution of the Human Cerebellum across Task Domains. *Neuron* **102**, 918–928 (2019).

13. Schmahmann, J. D. The cerebellum and cognition. *Neurosci. Lett.* **688**, 62–75 (2019).
14. Schmahmann, J. D., Guell, X., Stoodley, C. J. & Halko, M. A. The Theory and Neuroscience of Cerebellar Cognition. *Annu. Rev. Neurosci.* **42**, 337–364 (2019).
15. Eccles, J. C., Ito, M. & Szentágothai, J. Architectural Design of the Cerebellar Cortex. in *The Cerebellum as a Neuronal Machine* (eds. Eccles, J. C., Ito, M. & Szentágothai, J.) 195–204 (Springer Berlin Heidelberg, 1967).
16. Taylor, J. A. & Ivry, R. B. Cerebellar and prefrontal cortex contributions to adaptation, strategies, and reinforcement learning. *Prog. Brain Res.* **210**, 217–253 (2014).
17. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. *Trends Cogn. Sci.* **2**, 338–347 (1998).
18. Ito, M. Control of mental activities by internal models in the cerebellum. *Nat. Rev. Neurosci.* **9**, 304–313 (2008).
19. Popa, L. S. & Ebner, T. J. Cerebellum, Predictions and Errors. *Front. Cell. Neurosci.* **12**, 524 (2018).
20. Siman-Tov, T. *et al.* Is there a prediction network? Meta-analytic evidence for a cortical-subcortical network likely subserving prediction. *Neurosci. Biobehav. Rev.* **105**, 262–275 (2019).
21. Lametti, D. R., Smith, H. J., Watkins, K. E. & Shiller, D. M. Robust Sensorimotor Learning During Variable Sentence Level Speech. *Curr. Biol.* (2018).
22. Wolpert, D. M., Diedrichsen, J. & Flanagan, J. R. Principles of sensorimotor learning. *Nat. Rev. Neurosci.* **12**, 739–751 (2011).
23. Shadmehr, R., Smith, M. A. & Krakauer, J. W. Error correction, sensory prediction, and adaptation in motor control. *Annu. Rev. Neurosci.* **33**, 89–108 (2010).

24. Morton, S. M. & Bastian, A. J. Prism adaptation during walking generalizes to reaching and requires the cerebellum. *J. Neurophysiol.* **92**, 2497–2509 (2004).

25. Martin, T. A., Keating, J. G., Goodkin, H. P., Bastian, A. J. & Thach, W. T. Throwing while looking through prisms. I. Focal olivocerebellar lesions impair adaptation. *Brain* **119 (Pt 4)**, 1183–1198 (1996).

26. Baizer, J. S. & Glickstein, M. Proceedings: Role of cerebellum in prism adaptation. *J. Physiol.* **236**, 34P–35P (1974).

27. Rabe, K. *et al.* Adaptation to visuomotor rotation and force field perturbation is correlated to different brain areas in patients with cerebellar degeneration. *J. Neurophysiol.* **101**, 1961–1971 (2009).

28. Smith, M. A. & Shadmehr, R. Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *J. Neurophysiol.* **93**, 2809–2821 (2005).

29. Gibo, T. L., Criscimagna-Hemminger, S. E., Okamura, A. M. & Bastian, A. J. Cerebellar motor learning: are environment dynamics more important than error size? *J. Neurophysiol.* **110**, 322–333 (2013).

30. Parrell, B., Agnew, Z., Nagarajan, S., Houde, J. & Ivry, R. B. Impaired Feedforward Control and Enhanced Feedback Control of Speech in Patients with Cerebellar Degeneration. *J. Neurosci.* **37**, 9249–9258 (2017).

31. Lametti, D. R., Smith, H. J., Freidin, P. & Watkins, K. E. Cortico-cerebellar Networks Drive Sensorimotor Learning in Speech. *J. Cogn. Neurosci.* 1–12 (2017).

32. Jayaram, G., Galea, J. M., Bastian, A. J. & Celnik, P. Human locomotor adaptive learning is proportional to depression of cerebellar excitability. *Cereb. Cortex* **21**, 1901–1909 (2011).

33. Galea, J. M., Vazquez, A., Pasricha, N., de Xivry, J.-J. O. & Celnik, P. Dissociating the roles of the cerebellum and motor cortex during adaptive learning: the motor cortex retains what the cerebellum learns. *Cereb. Cortex* **21**, 1761–1770 (2011).

34. Moberget, T. & Ivry, R. B. Cerebellar contributions to motor control and language comprehension: searching for common computational principles. *Ann. N. Y. Acad. Sci.* **1369**, 154–171 (2016).

35. Hertrich, I., Mathiak, K. & Ackermann, H. The role of the cerebellum in speech perception and language comprehension. in *The linguistic cerebellum* 33–50 (Elsevier, 2016).

36. Ackermann, H. Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends Neurosci.* **31**, 265–272 (2008).

37. Callan, D. E., Kawato, M., Parsons, L. & Turner, R. Speech and song: the role of the cerebellum. *Cerebellum* **6**, 321–327 (2007).

38. Schwartze, M. & Kotz, S. A. Contributions of cerebellar event-based temporal processing and preparatory function to speech perception. *Brain Lang.* **161**, 28–32 (2016).

39. Argyropoulos, G. P. D. The cerebellum, internal models and prediction in 'non-motor' aspects of language: A critical review. *Brain Lang.* **161**, 4–17 (2016).

40. Guell, X., Gabrieli, J. D. E. & Schmahmann, J. D. Triple representation of language, working memory, social and emotion processing in the cerebellum: convergent evidence from task and seed-based resting-state fMRI analyses in a single large cohort. *Neuroimage* **172**, 437–449 (2018).

41. King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B. & Diedrichsen, J. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nat. Neurosci.* **22**, 1371–1378 (2019).

42. E, K.-H., Chen, S.-H. A., Ho, M.-H. R. & Desmond, J. E. A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. *Hum. Brain Mapp.* **35**, 593–615 (2014).

43. Stoodley, C. J. & Schmahmann, J. D. Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage* **44**, 489–501 (2009).

44. Petacchi, A., Laird, A. R., Fox, P. T. & Bower, J. M. Cerebellum and auditory function: an ALE meta-analysis of functional neuroimaging studies. *Hum. Brain Mapp.* **25**, 118–128 (2005).

45. Riedel, M. C. *et al.* Meta-analytic connectivity and behavioral parcellation of the human cerebellum. *Neuroimage* **117**, 327–342 (2015).

46. Balsters, J. H., Laird, A. R., Fox, P. T. & Eickhoff, S. B. Bridging the gap between functional and anatomical features of cortico-cerebellar circuits using meta-analytic connectivity modeling. *Hum. Brain Mapp.* **35**, 3152–3169 (2014).

47. Skipper, J. I., Devlin, J. T. & Lametti, D. R. The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain Lang.* **164**, 77–105 (2017).

48. Skipper, J. I., van Wassenhove, V., Nusbaum, H. C. & Small, S. L. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* **17**, 2387–2399 (2007).

49. Skipper, J. I., Nusbaum, H. C. & Small, S. L. Listening to talking faces: motor cortical activation during speech perception. *Neuroimage* **25**, 76–89 (2005).

50. Skipper, J. I. Echoes of the spoken past: how auditory cortex hears context during speech perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130297 (2014).

51. Skipper, J. I. The NOLB model: a model of the natural organization of language and the brain. in *Cognitive Neuroscience of Natural Language Use* (eds. Willems, R. M. & Willems, R. M.) 101–134 (Cambridge University Press, 2015).

52. Skipper, J. I., Nusbaum, H. C. & Small, S. L. Lending a helping hand to hearing: Another motor theory of speech perception. *Action to language via the mirror neuron system* 250–285 (2006).

53. Lesage, E., Hansen, P. C. & Miall, R. C. Right Lateral Cerebellum Represents Linguistic Predictability. *J. Neurosci.* **37**, 6231–6241 (2017).

54. D'Mello, A. M., Turkeltaub, P. E. & Stoodley, C. J. Cerebellar tDCS Modulates Neural Circuits

during Semantic Prediction: A Combined tDCS-fMRI Study. *J. Neurosci.* **37**, 1604–1613 (2017).

55. Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S. & Miall, R. C. Cerebellar rTMS disrupts predictive language processing. *Curr. Biol.* **22**, R794–5 (2012).
56. Pleger, B. & Timmann, D. The role of the human cerebellum in linguistic prediction, word generation and verbal working memory: evidence from brain imaging, non-invasive cerebellar stimulation and lesion studies. *Neuropsychologia* **115**, 204–210 (2018).
57. Moberget, T., Gullesen, E. H., Andersson, S., Ivry, R. B. & Endestad, T. Generalized role for the cerebellum in encoding internal models: evidence from semantic processing. *J. Neurosci.* **34**, 2871–2878 (2014).
58. Sheu, Y.-S., Liang, Y. & Desmond, J. E. Disruption of Cerebellar Prediction in Verbal Working Memory. *Front. Hum. Neurosci.* **13**, 61 (2019).
59. Ackermann, H., Gräber, S., Hertrich, I. & Daum, I. Categorical speech perception in cerebellar disorders. *Brain Lang.* **60**, 323–331 (1997).
60. Mathiak, K., Hertrich, I., Grodd, W. & Ackermann, H. Cerebellum and speech perception: a functional magnetic resonance imaging study. *J. Cogn. Neurosci.* **14**, 902–912 (2002).
61. Lametti, D. R., Oostwoud Wijdenes, L., Bonaiuto, J., Bestmann, S. & Rothwell, J. C. Cerebellar tDCS Dissociates the Timing of Perceptual Decisions from Perceptual Change in Speech. *J. Neurophysiol.* jn.00433.2016 (2016).
62. Guediche, S., Holt, L. L., Laurent, P., Lim, S.-J. & Fiez, J. A. Evidence for Cerebellar Contributions to Adaptive Plasticity in Speech Perception. *Cereb. Cortex* **25**, 1867–1877 (2015).
63. Kotz, S. A., Stockert, A. & Schwartze, M. Cerebellum, temporal predictability and the updating of a mental model. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130403 (2014).
64. Knolle, F., Schröger, E., Baess, P. & Kotz, S. A. The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *J. Cogn. Neurosci.* **24**, 698–706 (2012).
65. Schwartze, M., Tavano, A., Schröger, E. & Kotz, S. A. Temporal aspects of prediction in

audition: Cortical and subcortical neural mechanisms. *Int. J. Psychophysiol.* **83**, 200–207 (2012).

66. Schwartze, M., Keller, P. E. & Kotz, S. A. Spontaneous, synchronized, and corrective timing behavior in cerebellar lesion patients. *Behav. Brain Res.* **312**, 285–293 (2016).
67. Poeppel, D. A critical review of PET studies of phonological processing. *Brain Lang.* **55**, 317–51; discussion 352–85 (1996).
68. Friston, K. J. *et al.* The trouble with cognitive subtraction. *Neuroimage* **4**, 97–104 (1996).
69. Barton, R. A. Embodied cognitive evolution and the cerebellum. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2097–2107 (2012).
70. Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E. & Ramnani, N. A probabilistic MR atlas of the human cerebellum. *Neuroimage* **46**, 39–46 (2009).
71. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).
72. Goldinger, S. D. & Azuma, T. Puzzle-solving science: the quixotic quest for units in speech perception. *J. Phon.* **31**, 305–320 (2003).
73. Bybee, J. & McClelland, J. L. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* **22**, (2005).
74. Lotto, A. J. & Holt, L. The illusion of the phoneme. (2000).
75. Diedrichsen, J. & Zotow, E. Surface-Based Display of Volume-Averaged Cerebellar Imaging Data. *PLoS One* **10**, e0133402 (2015).
76. Vega, A. de la *et al.* Large-Scale Meta-Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *The Journal of Neuroscience* vol. 36 6553–6562 (2016).
77. De La Vega, A., Yarkoni, T. & Wager, T. D. Large-scale meta-analysis suggests low regional modularity in lateral frontal cortex. *Cerebral* (2018).
78. Guell, X., Schmahmann, J. D., Gabrieli, J. D. E. & Ghosh, S. S. Functional gradients of the cerebellum. *Elife* **7**, (2018).

79. Stoodley, C. J., Valera, E. M. & Schmahmann, J. D. Functional topography of the cerebellum for motor and cognitive tasks: an fMRI study. *Neuroimage* **59**, 1560–1570 (2012).

80. Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C. & Yeo, B. T. T. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 2322–2345 (2011).

81. Marek, S. *et al.* Spatial and Temporal Organization of the Individual Human Cerebellum. *Neuron* **100**, 977–993.e7 (2018).

82. Imamizu, H., Kuroda, T., Miyauchi, S., Yoshioka, T. & Kawato, M. Modular organization of internal models of tools in the human cerebellum. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5461–5466 (2003).

83. Chen, S. H. A. & Desmond, J. E. Temporal dynamics of cerebro-cerebellar network recruitment during a cognitive task. *Neuropsychologia* **43**, 1227–1237 (2005).

84. Chen, S. H. A. & Desmond, J. E. Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. *Neuroimage* **24**, 332–338 (2005).

85. Ashida, R., Cerminara, N. L., Edwards, R. J., Apps, R. & Brooks, J. C. W. Sensorimotor, language, and working memory representation within the human cerebellum. *Hum. Brain Mapp.* **40**, 4732–4747 (2019).

86. Van Overwalle, F., Baetens, K., Mariën, P. & Vandekerckhove, M. Social cognition and the cerebellum: a meta-analysis of over 350 fMRI studies. *Neuroimage* **86**, 554–572 (2014).

87. Stoodley, C. J. & Schmahmann, J. D. Chapter 4 - Functional topography of the human cerebellum. in *Handbook of Clinical Neurology* (eds. Manto, M. & Huisman, T. A. G. M.) vol. 154 59–70 (Elsevier, 2018).

88. Gellersen, H. M., Guo, C. C. & O'Callaghan, C. Cerebellar atrophy in neurodegeneration—a meta-analysis. *J. Neurol. Neurosurg. Psychiatry* (2017).

89. Boillat, Y., Bazin, P.-L. & van der Zwaag, W. Whole-body somatotopic maps in the cerebellum

revealed with 7T fMRI. *Neuroimage* 116624 (2020).

90. Grodd, W., Hülsmann, E., Lotze, M., Wildgruber, D. & Erb, M. Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Hum. Brain Mapp.* **13**, 55–73 (2001).
91. Hurley, S. Perception And Action: Alternative Views. *Synthese* **129**, 3–40 (2001).
92. Skipper, J. I. & Hasson, U. A Core Speech Circuit Between Primary Motor, Somatosensory, And Auditory Cortex: Evidence From Connectivity And Genetic Descriptions. *bioRxiv* 139550 (2017) doi:10.1101/139550.
93. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
94. Keller, G. B. & Mrsic-Flogel, T. D. Predictive Processing: A Canonical Cortical Computation. *Neuron* **100**, 424–435 (2018).
95. Moberget, T. & Ivry, R. B. Prediction, Psychosis, and the Cerebellum. *Biol Psychiatry Cogn Neurosci Neuroimaging* **4**, 820–831 (2019).
96. Bonhage, C. E., Mueller, J. L., Friederici, A. D. & Fiebach, C. J. Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex* **68**, 33–47 (2015).
97. Tourville, J. A., Reilly, K. J. & Guenther, F. H. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* **39**, 1429–1443 (2008).
98. Ackermann, H., Wildgruber, D., Daum, I. & Grodd, W. Does the cerebellum contribute to cognitive aspects of speech production? A functional magnetic resonance imaging (fMRI) study in humans. *Neurosci. Lett.* **247**, 187–190 (1998).