Brain structural imaging of receptive speech and beyond: a review of current methods

Damien Marie & Narly Golestani

To cite this article: Damien Marie & Narly Golestani (2017) Brain structural imaging of receptive speech and beyond: a review of current methods, Language, Cognition and Neuroscience, 32:7, 870-890, DOI: 10.1080/23273798.2016.1250926

To link to this article: http://dx.doi.org/10.1080/23273798.2016.1250926

Published online: 09 Nov 2016.
ABSTRACT
This review provides an overview of brain structural imaging methods for examining neural correlates of performance, skill and learning in the domain of language and beyond. We first provide a historical overview of structural imaging, followed by an overview of brain structural measures/variables, assumptions regarding their neurophysiological basis, a description of the challenges and solutions for using structural imaging for studying receptive speech, a review of key empirical contributions in the field, and finally, some future directions for this research approach. Studies having examined multiple structural measures and modalities in parallel are still sparse. However, recent methodological advances which allow the non-invasive estimation of regional distribution of histological properties such as myelination, along with the increasing resolution of in vivo structural imaging are opening new and exciting avenues which will enable a better understanding of the brain structural underpinnings of language, and of the physiological relevance of these structural features.

1. Historical background
The study of brain structure is complementary to brain functional investigations in that function ultimately arises from particular neuroanatomical substrates. It is now widely accepted that experience-dependent changes in structure can take place not only early in life but also in the adult brain (Boyke, Driemeyer, Gaser, Buchel, & May, 2008; Draganski et al., 2004; Opendak & Gould, 2015; Pfefferbaum, Sullivan, & Carmelli, 2004; Richardson & Price, 2009), often in the very regions known to functionally subserve the task or domain of processing at hand (Golestani, 2014; Zatorre, Fields, & Johansen-Berg, 2012).

Historically, the evaluation of brain structure is intimately related to research on speech, starting with neuropsychological characterisation of Broca’s (1861) and Wernicke’s (1874) aphasias. Fundamental first works on brain anatomy relevant to the language system includes the early cytoarchitectonic studies serving to parcellate the auditory subfields (von Economo & Horn, 1930; von Economo & Koskinas, 1925). Other early works were post-mortem myelin staining of the acoustic radiation (fibre tract which project from the thalamus to primary auditory cortex), and characterisation of macrostructural variation of Heschl’s gyrus (HG), which includes primary auditory cortex, in relation to the former profession, age and gender of individuals (Pfeifer, 1920). Building on these advances, Geschwind and Levitsky demonstrated a striking leftward asymmetry in the length of the planum temporale (PT), delineated manually. This asymmetry was proposed to underlie left-hemispheric specialisation for language (Geschwind & Levitsky, 1968). Further work showed that the PT, this time cytoarchitectonically delineated (i.e. area Tpt), was also leftward asymmetric in volume, and further, that cytoarchitectural and macrostructural asymmetries were correlated (Galaburda, Sanides, & Geschwind, 1978). Interestingly, similar macrostructural asymmetries were indirectly shown in non-human primates, suggesting the early evolutionary presence of this auditory cortex structural feature (LeMay, 1976). Early postmortem work in humans also showed higher cortical folding, or gyriﬁcation, in prefrontal and temporal cortices compared to non-human primates (Zilles, Armstrong, Schleicher, & Kretschmann, 1988). Since the advent of in vivo neuroimaging in the 1970–1980s, structural magnetic resonance imaging (MRI) has become the method of choice for the non-invasive assessment of brain structure (Symms, Jager, Schmierer, & Yousry, 2004). The first applications relevant to language included conﬁrmation of the leftward asymmetry of the PT (Larsen, Ødegaard, Grude, & Høien, 1989; Shapleske, Rossell, Woodruff, & David, 1999; Steinmetz, Rademacher, et al., 1989), and ﬁndings that this asymmetry...
is altered in dyslexia (Larsen, Høien, Lundberg, & Ædegaard, 1990). Another early use of structural MRI was its registration to functional data for localisation of the latter, and the development of a standard, Talairach space for group-level data analysis (Evans, Marrett, et al., 1992; Fox, Perlmutter, & Raichle, 1985; Seitz et al., 1990; Talairach et al., 1967). The Talairach space being limited in terms of its generalisability (since it was derived from one hemisphere of a 60-year old female), more common now is the use of population-averaged templates and the associated standard spaces (e.g. Montreal Neurological Institute space: Evans, Collins, et al., 1993). Early works relevant to language having used a common reference space for group-level structural analyses included findings of leftward asymmetry in the posterior Sylvian fissure (Steinmetz, Fürst, & Freund, 1989), development of macrostructural probabilistic maps of HG (Penhune, Zatorre, MacDonald, & Evans, 1996) and of brain regions involved in linguistic processing (e.g. the pars opercularis of the inferior frontal gyrus: Tomaiuolo et al., 1999) and in cognitive and language control (e.g. the cingulate and paracingulate gyri: Paus, Otaky, et al., 1996; Paus, Tomaiuolo, et al., 1996), and development of microstructural probabilistic maps (e.g. primary auditory areas and acoustic radiations: Rademacher, Burgel, & Zilles, 2002; Rademacher et al., 2001).

The above developments together with ultra-high resolution MRI (i.e. 7 Tesla field strength) and new and ever improving structural MRI data analysis methods have allowed this domain of in vivo research to vastly expand. In the last years, a number of papers have reviewed brain structural imaging studies in the context of learning and development in different domains of processing (Draganski & May, 2008; Lovden, Backman, Lindenberger, Schaefer, & Schmiedek, 2010; May, 2011; Mills & Tamnes, 2014; Tardif et al., 2016; Voelcker-Rehage & Niemann, 2013). This review will outline structural MRI studies in the domain of language that have focused on receptive speech (i.e. speech perception and comprehension), at low (i.e. phonetic) to higher (i.e. sentence level) levels of the language processing hierarchy. We will outline available methods, challenges and solutions in using structural imaging for studying receptive speech, key empirical contributions and promising future developments.

2. Overview of the method

2.1. Range of possible behavioural and brain structural measures

Structural MRI data can be preprocessed and analysed using different software packages, ranging from manual or semi-automated (e.g. labelling in MRicro suite, http://www.mccauslandcenter.sc.edu/micro/ or semi-automatic segmentation in ITK-SNAP, http://www.itksnap.org) to fully automated ones (e.g. voxel-based morphometry (VBM), http://www.fil.ion.ucl.ac.uk/spm/, Ashburner, 2009, or FSL, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/, Smith et al., 2004; BrainVoyager, http://www.brainvoyager.com, Goebel, 2012; FreeSurfer, http://freesurfer.net, Fischl, 2012; see Nitzken et al., 2014 for a more detailed overview). Some of these are more suitable for exploratory, whole-brain analyses while others are more suitable for region-of-interest (ROI) analyses, where specific a priori predictions are made. An overview of these, however, is beyond the scope of this review. The selection of the data analysis approach(es) ought to be determined by the questions posed by the experiment and answerable according to the study design. Ideally, one should aim to show convergence between the results of different analysis approaches, and across data sets where those might be available. Also, results of one analysis can lead to new/specific hypotheses about the origin of the underlying structural difference, which can, in turn, be tested using different analysis tools (see, e.g. Golestani, Molk, Dehaene, LeBihan, & Pallier, 2007; Golestani, Paus, & Zatorre, 2002; Golestani, Price, & Scott, 2011). Note that except in the context of certain experimental designs such as well-controlled longitudinal studies, it can only be speculated as to whether the measured brain structural differences have arisen from differences in previous experience and learning, or whether they already existed before the training began, and thus possibly play a causal role in modulating performance.

2.1.1. Behavioural measures/variables

Structural imaging studies on language have typically used the approach of correlating performance on a particular task (i.e. a continuous measures of performance on a task, see, e.g. Golestani et al., 2002) with brain structure, or of comparing brain structure across groups that differ along some measure or characteristic (e.g. good or poor performance on a learning task, see Golestani et al., 2007; Wong et al., 2008), comparison of expert to non-expert group (see, e.g. Golestani et al., 2011). In longitudinal designs, brain structure is compared before and after some intervention or learning (e.g. Stein et al., 2012) in the same group of participants, sometimes in comparison to a control group who has also been scanned at the same intervals (e.g. Hervais-Adelman, Moser-Mercer, Murray, & Golestani, submitted for publication).
2.1.2. Brain structural measures/variables

The brain structural measures that one can obtain following the analysis of structural MRI data include measures of (1) grey or white matter probability, or concentration (also called “density”, see below), following VBM (e.g. Ashburner & Friston, 2000; Golestani et al., 2007), (2) grey or white matter volume following either VBM where modulation is applied during the pre-processing steps (Mechelli, Price, Friston, & Ashburner, 2005; Radua, Canales-Rodríguez, Pommol-Clotet, & Salvador, 2014), or following deformation-based morphometry (DBM) (e.g. Ashburner et al., 1998; Golestani et al., 2002) or manual labelling of ROIs (e.g. Marie et al., 2015), (3) cortical surface area (e.g. Burgaleta, Baus, Díaz, & Sebastián-Gallés, 2014), (4) cortical thickness (e.g. Burgaleta et al., 2014; Hutton, De Vita, Ashburner, Deichmann, & Turner, 2008), (5) gyriﬁcation (e.g. Golestani et al., 2011; Marie et al., 2015; Schae et al., 2008), (6) position (e.g. Golestani et al., 2002), (7) length (e.g. Bonte et al., 2013), (8) depth (e.g. Bonte et al., 2013) or (9) shape (e.g. Burgaleta, Sanjuán, Ventura-Campos, Sebastián-Galles, & Ávila, 2016; Le Goualher et al., 1999; Nitzken et al., 2014) of particular structures (e.g. gyri, sulci or subcortical nuclei). Several automated surface-based morphometry (SBM) packages exist that allow not only to automatically parcellate the brain into atlas-derived regions of interest, but also to extract some of these brain structural measures (e.g. volume, surface area, thickness, sulcal information and gyriﬁcation, Fischl, 2012; Greve et al., 2013; Mangin et al., 2004; Reuter, Schmansky, Rosas, & Fischl, 2012; Riviere et al., 2002; Schae et al., 2012; Van Essen et al., 2001). These and other packages are also adapted for analysing subcortical structures (Fischl et al., 2002; Liem et al., 2015; Patenaude, Smith, Kennedy, & Jenkinson, 2011). Hemispheric symmetries can also be computed for any of the above measures (e.g. volume, surface area, gyriﬁcation, etc.). Asymmetry is typically quantiﬁed using a lateralisation index (LI), which is calculated as a ratio (e.g. LI = [(X^R − X^L)/(X^R + X^L)/2], where X indicates the brain measure at hand, and where R and L indicate the right and left hemispheres, respectively; cf. Penhune et al., 1996; Golestani et al., 2011). Whole-brain voxel- or vertex-wise asymmetry maps can also be generated (see Kurth, Gaser, & Luders, 2015; Greve et al., 2013 for respective descriptions). The above measures (i.e. volume, surface area, cortical thickness, etc.) can be used for analyses of structural covariance, which can be performed voxel-wise, vertex-wise or based on predefined ROIs. Analyses of covariance allow to assess how variation in the volume (or thickness, or other) of a particular structure covaries with that of other regions across the brain, across individuals or between groups (Mechelli, Friston, Frackowiak, & Price, 2005).

The above brain structural measures are mostly derived from T1-weighted structural MRI scans. More recent advances have led to the development of quantitative MRI (qMRI) sequences (previously known as “quantitative multiparameter mapping”, or MPM), which allow to acquire a range of quantitative, complementary MRI parameters in parallel (Weiskopf, Mohammadi, Lutti, & Callaghan, 2015). As is described in more detail below (see Section 2.2), qMRI allows to obtain various measures (e.g. R1, MT, and R2* maps) which provide estimates of brain myelination in vivo.

Structural brain imaging data can also be analysed using multivariate and machine learning approaches that are relatively more data driven and computationally demanding than more traditional, univariate ones (Chung, 2012; Lemm, Blankertz, Dickhaus, & Muller, 2011). They rely on several assumptions about the data model and its statistical properties, and therefore their interpretation may not be as straightforward as more conventional methods with respect to elucidation of the underlying anatomy and its localisation. The measures obtained using machine learning approaches such as principal component analysis (PCA) or independent component analysis (ICA) on brain structural data can take the form of abstract, latent variables that explain patterns or group differences in the data. For example, a multivariate alternative to VBM, called source-based morphometry, involves performing an ICA analysis on the data after preprocessing as would be done for VBM. This approach allows to identify naturally grouping, maximally independent sources of variance in the data, on which further statistical analyses are performed. The identified “source networks”, or groups of spatially distinct regions showing common covariance patterns among subjects, provide information about localisation of the structural measure at hand (e.g. grey matter), and its variation between individuals (see Xu, Groth, Porcell, Schretlen, & Calhoun, 2009, for a full description of the method). Other multivariate methods include multi-voxel pattern analysis (MVPA), also known as pattern classification or analysis, of brain structural data (e.g. Hoeft et al., 2011), which allows to determine distributed structural patterns that can distinguish two groups, or that can predict disease or recovery. Depending on the choice of approach and its implementation, the resulting patterns can be distributed in the brain and therefore difficult to interpret with respect to neuroanatomy. In research on language, the MVPA approach has so far mainly been applied to brain functional data (Evans et al., 2014; Hervais-Adelman, Moser-Mercer, & Golestani, 2015; Raizada,
Tsao, Liu, & Kuhl, 2010; Raizada, Tsao, Liu, Holloway, et al., 2010) although some studies have examined brain structural differences in the context of reading skill (He et al., 2013) and reading deficit (Hoeft et al., 2011). There also exist related methods such as canonical correlational analysis, where sets of continuous features are related to one another in a multivariate manner (Hackmack et al., 2012). These multivariate classification approaches may allow to reliably predict which group a new, unseen structural MRI may belong to, or they may allow to predict long-term outcome for example in dyslexia (e.g. Hoeft et al., 2011). Also, the extracted explanatory dimensions (e.g. in the cases of PCA or ICA) may be abstract, and may thus need to be “back-projected” into “brain space” in order to be interpretable with respect to the underlying anatomy. Finally, results of these multivariate methods need to be validated using cross-validation schemes (see He et al., 2013; Hoeft et al., 2011). Bootstrapping approaches are also useful for assessing the stability of results (Nakagawa, 2004).

Anatomical atlases of the whole brain and probabilistic maps of specific structures of interest have been developed, both for use during the preprocessing steps for VBM and other analyses, and also for the purpose of ROI-based analyses (Auzias, Coulon, & Brovelli, 2016; Destrieux, Fischl, Dale, & Halgren, 2010; Evans, Janke, Collins, & Baillet, 2012; Paus, Tomaiuolo, et al., 1996; Penhune et al., 1996; Rademacher et al., 2001; Tomaiuolo et al., 1999; Tzourio-Mazoyer et al., 2002). In conjunction with probabilistic maps of regional differences in cytoarchitecture (Amunts, Schleicher, Ditterich, & Zilles, 2003; Bludau et al., 2014; Lorenz et al., 2015; Morosan et al., 2001) these atlases can be used to localise structural (or functional) results. Other developments include registration of sulcus-based models of cortical organisation onto flat representations (Auzias et al., 2013), and corresponding parcellation models (Auzias et al., 2016) that can be used for anatomical localisation. Additionally, recent interest has been growing for the use of “sulcal pits” (points of maximum depth within folds) to compare brains or to register them (Auzias, Brun, Deruelle, & Coulon, 2015).

### 2.2. Neurophysiological basis

Structural MRI provides information about brain structure at the macroscopic scale, and this imaging method has also been used in non-human primates to demonstrate rapid brain structural change as a function of learning (Quallo et al., 2009). Animal studies, using microscopy mainly in mice and rats, and in vitro studies, have also directly explored the neurophysiological underpinnings of experience-dependent plasticity (Buchs & Muller, 1996; Holtmaat, Wilbrecht, Knott, Welker, & Svoboda, 2006; Segal & Andersen, 2000). As reviewed by Thomas and Baker (2013), these animal studies are critical for explaining spatial and temporal profiles of structural plasticity and how such changes are reflected in the MRI signal. While grey matter changes detected by MRI in the human brain are likely not related to neurogenesis, which might be minimal, sparse or even absent after the normal developmental period (Bhardwaj et al., 2006; Ernst & Frisen, 2015), or which may only occur in restricted brain regions (Bowers & Jessberger, 2016; Marques et al., 2016), numerous other potential candidate neurophysiological mechanisms have been proposed, as reviewed by several papers (Draganski & May, 2008; Tardif et al., 2016; Zatorre et al., 2012). For grey matter, these include dendritic and axonal arborisation and synaptogenesis, neuronal size, neuropil volume, changes to glia, microglia, astrocytes, cell swelling and angiogenesis (i.e. vascular changes). Many of these factors can lead to differences in the computational capacities of the cortex (e.g. number of neurons and their organisation). Differences in grey matter volume observed using structural MRI could also arise from more cortical folding or from thicker cortex, or could even partly be an artifact of poor grey/white matter demarcation (Mechelli, Price, et al., 2005). In white matter, approximately 50% of macromolecules constitute the myelin sheaths, which envelop axons, axonal membranes and neurofibrils, and which constrain water diffusion. Consequently, structural features other than myelin might give rise to anisotropy (Beaulieu, 2002; Tardif et al., 2016). Despite this, myelination is the candidate mechanism that has most often been put forth as underlying white matter differences or change. Myelination differences can arise from myelin formation and oligodendroglial proliferation, differences in fibre density or organisation, differences in axon number (e.g. due to pruning or axonal branching) and differences in axonal diameter or trajectory. Angiogenesis, and increases in glial cell size, number and swelling have also been proposed (Draganski & May, 2008; Tardif et al., 2016; Zatorre et al., 2012). Discussion of the neurophysiological basis underlying diffusion tensor imaging (DTI) differences is beyond the scope of this review; readers are referred to several existing papers that discuss this topic (Beaulieu, 2002; Dubois, Hertz-Pannier, Dehaene-Lambertz, Cointepas, & Le Bihan, 2006).

One study has directly analysed the microscopic underpinnings of macroscopic grey matter probability differences obtained using VBM in humans by performing histological analyses on brain tissue that was resected for clinical purposes in patients with temporal...
lobe epilepsy (Eriksson et al., 2009). This study showed that macroscopic measures of grey matter probability did not correlate with quantitative histopathological measurements of neuronal density. There was also no relationship with a composite measure of neuronal components such as neuropil, neuronal size, dendritic or axonal arborisation, but these non-findings do not exclude the possible contribution of other neuronal markers. These results confirm that the term “density”, often used to describe VBM results, is a misnomer, and that VBM results do not reflect cell packing density as measured cytoarchitectonically, but rather that they reflect the concentration, or probability of grey matter presence at a particular location in the brain (Ashburner & Friston, 2000; Eriksson et al., 2009). In contrast, recent work using an in vivo index of histological cortical thickness by imaging of gamma-aminobutyric acid (GABA) receptors (tracer labelled positron emission tomography) together with surface-based measures of cortical thickness (using MRI) has shown that in temporal and occipital regions, there is a negative relationship between neuronal density and surface-based cortical thickness. In fronto-parietal regions however, neuronal density is relatively constant, and independent of surface cortical thickness, but regional differences in GABA receptor concentrations may have affected these results (la Fougere et al., 2011).

While cortical volume is the product of cortical thickness and surface area, cortical volume could be more closely related to cortical surface area than to thickness. Also, grey matter volume has been shown to be more strongly genetically and environmentally correlated with surface area than with thickness (Winkler et al., 2010). Nonetheless, similar underlying neurophysiological underpinnings have been proposed for cortical thickness and for grey matter volume (e.g. dendritic and axonal arborisation and synaptogenesis, neuronal size, neuropil volume, etc., see, for example, Zatorre et al., 2012). For cortical surface area, different explanations have been offered. For example, the radial-unit hypothesis postulates that cortical surface area is principally determined by the number and spacing of radial cortical columns, rather than by the number and size of cells within a column, packing density or number of connections. These latter measures are thought to instead underlie cortical thickness and grey matter volume (Eickhoff et al., 2005; Rakic, 1995). Thus, variation in cortical surface area could relate to differences in the number and spacing of radial cortical columns (Rakic, 1988, 2009).

Gyrification and sulcus establishment has theoretically been related to morphogenetic (Rogers et al., 2010; Toro et al., 2008) and to mechanical factors (e.g. axonal tension, grey matter tangential expansion mechanisms, establishment of connectivity; Tallinen, Chung, Biggins, & Mahadevan, 2014; Van Essen, 1997) as well as to in utero regional changes in tissue growth (Scheinost et al., 2015), but its neurobiological origins remain poorly understood (Razavi, Zhang, Liu, & Wang, 2015). Also, cytoarchitectonic probability atlases and classical postmortem studies show that at least in certain brain regions, sulcal boundaries tend to demarcate different cytoarchitectonic zones, although the mapping remains quite variable (Welker, 1990; Zilles & Amunts, 2010). Importantly, differences in gyrification can arise from factors including differential growth of cortical layers, which are composed of different cell types, and also, cortical folding can impose constraints on the shape of cells, especially in the outer cortical layers (White, Su, Schmidt, Kao, & Sapiro, 2010; Zilles, Palomer-Gallagher, & Amunts, 2013).

As can be seen from the above, our incomplete understanding of the relationship between tissue microstructure and morphological measures extracted from T1-weighted structural MRI data limits the inferences that can be made regarding underlying, tissue level differences. In fact, it has recently been demonstrated that microstructural differences in brain tissue may lead to the spurious detection of differences in morphological measures of grey matter volume and of cortical thickness (Lorio et al., 2016). Indeed, while the image intensity in T1-weighted images is predominantly driven by T1, other MRI parameters also affect the image contrast. Lorio and colleagues acquired different quantitative maps (e.g. 1/T1, R1 and other maps) that contribute to determining signal intensity in T1-weighted images. They then used these to create synthetic T1-weighted images, and showed that morphometric measurements derived from VBM and SBM vary depending on the MRI parameters that are considered. Specifically, because the different MRI parameters are correlates of different microstructural features of brain tissue, depending on which features one wants to examine, one should carefully select which MRI parameters to use for the extraction of morphometric measurements.

As described in Section 2.1.2, qMRI techniques that overcome the limitations of standard anatomical data are emerging. qMRI data provide quantitative markers of tissue microstructure, allowing more direct insight into the mechanisms underlying brain plasticity (Callaghan et al., 2014). For instance, the close correspondence between regional distributions of the R1 parameter measured by qMRI and cytoarchitectonically defined probability maps of primary sensory brain regions supports the validity of R1 as an index of brain myelination, highlighting recent advances towards “in vivo histology” (for review, see Lutti, Dick, Sereno, & Weiskopf, 2014).
Furthermore, the parallel quantification of complementary MRI parameters by qMRI allows the simultaneous assessment of multiple histological properties of brain tissue (Weiskopf et al., 2015). In particular, the quantification of the R2* parameter provides an index of iron concentration, which may be used for example in the assessment of iron deposition in the basal ganglia with ageing (Lorio et al., 2014). Other approaches also exist for modelling myelination patterns, for example using ultra-high resolution (i.e. 7 Tesla) structural MRI, which have also been shown to display good correspondence with ex vivo histology (Dinse et al., 2015; Eickhoff et al., 2005).

3. Challenges and solutions for studying spoken language

A number of challenges are present when using structural imaging to understand differences or changes in brain structure across individuals, between groups or over time. Some of these issues are not specific to studies on language, whereas others are uniquely present in studies on language as compared to ones in other cognitive domains. Many of these issues, especially ones related to the methods used for data analysis and to subject selection, have recently been described in detail in the context of structural imaging studies on bilingualism; readers are referred to this paper for more information on these topics (García-Pentón, Fernández García, Costello, Duñabeitia, & Carreiras, 2015).

3.1. Issues relevant to all structural imaging studies

3.1.1. MRI scanner

Neuroimaging research is increasingly performed in the context of consortia and of large databases (e.g. Ottet et al., 2013). In some cases, the data are acquired using different sequences and/or scanners, at different imaging sites and at different times of the day. These sequence and scanner differences, however, add noise and variability to the data, and can lead to systematic differences in results (Mechelli, Price, et al., 2005). Time of day can also impact morphometric measurements, as pointed out by a recent study (Trefler et al., 2016). In the case that data from different scanners and/or sequences have to be compared, one should add scanner as a covariate at the data analysis stage (Pardoe, Pell, Abbott, Berg, & Jackson, 2008; Segall et al., 2009; Stonnington et al., 2008; Takao, Hayashi, & Ohtomo, 2014). Recent developments in qMRI, however, allow for improved data comparability across scanners, and for higher sensitivity in multi-centre studies (Weiskopf et al., 2013). This is due to the quantitative nature of this type of data, to the fact that many methodological biases (e.g. sensitivity to field inhomogeneities) are eliminated (Tardif et al., 2016).

Other scanner-related issues include the fact that although ultra-high field MRI offers greater resolution than does more conventional (i.e. 3 Tesla) structural MRI, the former also results in larger distortions, especially in inferior and frontal parts of the cortex, due to proximity to air-filled cavities (Duchin, Abosch, Yacoub, Sapiro, & Harel, 2012). These distortions can be partly but not completely corrected (May, 2011). Also, signal hyperintensity in the anterior temporal pole can result in poor segmentation using SBM (Dale, Fischl, & Sereno, 1999), however, this can be manually corrected (see, e.g. Marie, Maingault, Crivello, Mazoyer, & Tzourio-Mazoyer, 2016).

3.1.2. Motion

Functional brain imaging data can be corrected for movement, and until recently this was not possible for structural scans. Now there are methods available to correct for motion in structural images retrospectively, even for sub-voxel head motion (Gallichan, Marques, & Gruetter, 2016). This can also be done prospectively, by monitoring fine movements in real time during scanning and correcting for it via online adjustments to the scanning protocol (Ooi, Krueger, Thomas, Swaminathan, & Brown, 2009; Todd, Josephs, Callaghan, Lutti, & Weiskopf, 2015).

3.1.3. Issues related to image quality and data analysis

Issues related to image quality that are relevant to all structural imaging studies include factors such as intensity inhomogeneities, random noise and spatial resolution (Thomas & Baker, 2013). For instance, intensity inhomogeneities, although they can partly be corrected for, affect the quality of grey and white matter segmentations. This is especially critical for longitudinal studies since systematic differences in intensity inhomogeneity between scans obtained over time can be confounded with the effect of interest (training or other), that also happens over time (Lewis & Fox, 2004). Some of these issues can be improved by using multiple structural scans and averaging them for each individual, to improve the signal-to-noise ratio for structural analyses, or by using qMRI, which is less sensitive to image distortions. Quality assessment can help to identify potential problems with the data (Ducharme et al., 2016). There are also methodological developments at the data analysis stage that can lead to improvements in the quality of structural MRI analyses, in particular when the data are of
poor quality, or have been acquired some years ago. These include the improved nonparametric nonuniform intensity normalisation (N4ITK; Tustison et al., 2010), and the spatially adaptive non-local means denoising filter (Manjon, Coupe, Marti-Bonmati, Collins, & Robles, 2010). Other issues related to the data itself concern the difficulty in spatial normalisation of atypical brains, which can for example lead to residual macroscopic differences after spatial normalisation and/or to sub-optimal segmentation in patients (Mechelli, Price, et al., 2005).

There are many issues that can arise at the data analysis stage, a detailed overview of which is beyond the scope of this review. Some issues relate for example to the use of different operating systems (Glator et al., 2015). Others related to the robustness of data analysis approaches (e.g. standard parametric tests, see Thomas & Baker, 2013). During VBM analyses, choices regarding the VBM analysis pipeline (i.e. FSL or SPM, Thomas & Baker, 2013), approach (i.e. optimised versus non-optimised VBM), the implementation of modulation versus not during the preprocessing stage, and the choice of the smoothing kernel size can modulate results and account for some of the inconsistencies in the literature (Mechelli, Price, et al., 2005). Modulation refers to the scaling of the warped tissue map by the Jacobian determinant of the deformation to compensate for the normalisation, and to provide an absolute value of tissue volume instead of relative tissue concentration (Ashburner, 2009; Mechelli, Price, et al., 2005). Although modulation has yielded neurologically interesting results that were not detected in non-modulated analyses (Good et al., 2001b; Keller, Wilke, Wieshmann, Sluming, & Roberts, 2004), a recent study has shown that, contrary to the generalised idea that modulation should be mandatory, modulation is not always better. This study tested the effect of modulation within standard (i.e. low-dimensional spatial normalisation on the whole brain, before segmentation), optimised (i.e. low-dimensional spatial normalisation on tissue probability maps) and advanced VBM (high-dimensional spatial normalisation based on differomorphic transformation algorithms) procedures, and showed that especially for optimised and advanced VBM, modulation may potentially worsen VBM outcome (Radua et al., 2014). Specifically, by multiplying unmodulated data by the Jacobian determinants, modulation can introduce multiplicative noise and therefore a decrease in statistical power, and this is relatively more the case when normalisation is stronger (e.g. with advanced VBM algorithms) (Radua et al., 2014). This is consistent with previous suggestions that standard VBM may be more sensitive to subtle brain variations than optimised VBM, due to the finer normalisation imposed by tissue priors in the latter, which can “wash away”, or mask existing differences/effects (Mechelli, Price, et al., 2005).

Other issues to consider at the data analysis stage in VBM analyses relate to the choice of template. Studies have shown differences in grey/white matter probability when using different templates, suggesting template effects, and the use of population-specific templates has been recommended (Salmond et al., 2002; Senjem, Gunter, Shiung, Petersen, & Jack, 2005; Shen, Sterr, & Szameit, 2005; Shen, Szameit, & Sterr, 2007). The choice of template is also crucial in the context of longitudinal designs, where the normalisation should be performed on a template constituted of images acquired at the first time point in order to allow the detection of changes arising between the first and second time points (Boyke et al., 2008; Draganski et al., 2006; Driemeyer, Boyke, Gaser, Buchel, & May, 2008; Holzel et al., 2011; Ilg et al., 2008; May et al., 2007; Thomas et al., 2009). Relative to VBM, the template effect may be less important for SBM as demonstrated by a recent study (Marie et al., 2016). However as noted in this study, FreeSurfer’s segmentation should be checked because of the presence of inaccuracies in 10% of the cases, generally at the temporal pole. Furthermore, many SBM pipelines work on the two hemispheres separately, and results are therefore hemisphere-specific, although it is still possible to compute vertex-wise asymmetry maps (Greve et al., 2013). Overall, for VBM (and in some cases also for SBM), the choice of template, modulation, optimised procedures or not, kernel size, etc. ought to be dictated by the experimental question(s) at hand (cf. Mechelli, Price, et al., 2005).

Because manual labelling is subject to errors in accuracy and to poor reproducibility (Mills & Tamnes, 2014), it is important to establish high inter-rater reliability when tracing a region of interest, and that the labelling be performed blind to group and to hemisphere. If this is not possible, ratings ought to be performed twice by the same person, and high reliability ought to be established across ratings (see, e.g. Golestani et al., 2007). Demonstration of high inter- or intra-rater reliability allows to ensure that the approach is consistent and robust. This is especially critical for regions with high anatomical variability.

One more issue related to the analysis of structural data is that due to the distributed nature of higher-level cognitive (i.e. linguistic) processing in the brain, structural differences or change often occur in a network of regions, and not in one region in isolation. One limitation of many structural imaging studies is that in order to demonstrate statistical significance, one or a small number of ROIs are selected, and this results in evaluation of focal change rather than of
changes that may have occurred in a network. Similarly, if a multiple comparisons correction such as a Bonferroni correction is applied when a larger number of ROIs are examined, statistical significance of each result diminishes, resulting in non-findings. This can constitute a type II error (a miss) in cases where a structural effect actually exists. In other words, especially when changes are expected to occur in multiple regions, Bonferroni adjustments may be too strict and may result in a lack of power for the detection of existing results (Nakagawa, 2004; Perneger, 1998). Some solutions to this problem include the use of more powerful multiple comparisons corrections methods (e.g. the Holms method; Aickin & Gensler, 1996), or the use of graph theory for the assessment of patterns of change across a network of brain regions. While for structural imaging graph theory has so far mostly been applied to diffusion weighted imaging data (Hagmann et al., 2008; Ottet et al., 2013), it can also be applied to examine structural covariance differences between groups or over time (e.g. Scheinost et al., 2015).

3.1.4. Experimental design
The design of studies aiming to reveal the structural correlates of different behavioural measures including receptive speech is critical in order to optimise the chance of detecting possibly subtle existing structural differences. For instance, inappropriate subject selection can lead to false negative or positive results. The interpretation and validity of findings depends on the careful selection and matching of the participants with respect to age, sex, handedness, educational level, language experience (in the case of studies on speech), etc. so as to ensure that group differences are not driven by confounds of non-interest. Educational level and socio-economic status (SES) are often taken into account in brain structural imaging studies including ones on language, in particular because it has been shown that language is one of the cognitive domains most affected by SES, and also because SES itself has been shown to be related to brain structural differences in brain regions involved in reading, in semantic processing and in memory (Jednoróg et al., 2012). In the case of studies on speech processing however, few studies control for musical training even though it could impact performance on some behavioural tasks that are relevant to speech, such as the processing of rapidly changing acoustic cues (Elmer, Hänggi, Meyer, & Jäncke, 2013) or such as speech-in-noise perception (Francois, Grau-Sanchez, Duarte, & Rodriguez-Fornells, 2015; Slater et al., 2015).

When it is not possible to fully match confounding variables, statistically covarying for the confounding parameter (e.g. age) may be suitable (Pell et al., 2008). However, if that parameter varies too widely or if it correlates with a different parameter of interest, then covarying out that variable may either not be sufficient, or it may wash out an effect of interest. For example, the measure of interest might be the number of years of experience with a particular (linguistic) task, but if this measure correlates highly with age, then the brain structural features that correlate with the amount of experience will no longer be detectable when age is included as a covariate. Also, due to non-linear relationships between brain structure and factors such as head size, age and gender, covarying out these factors is not optimal, and having well-matched groups is preferred (see, e.g. Barnes et al., 2010). It is also possible that brain structural correlates of performance on domain-specific tasks including receptive speech may vary over time, due to the evolution of developmental, maturational and ageing-related factors across the lifespan.

Finally, careful selection of behavioural measures is required in order to yield valid results that can be interpreted meaningfully and that can be generalisable. The measures that are being correlated with brain structure (e.g. phonetic learning skill) have to be stable over time within individuals. In other words, behavioural measures that fluctuate widely across testing sessions (e.g. due to variation in attention, concentration, or other factors, or due to lack of test–retest stability) cannot be used for correlations with brain structure since they would not yield statistically robust correlations with brain anatomy (Golestani, 2014).

3.1.5. Issues related to the interpretation of structural results
A challenge in studies on brain structural differences related to training and skill, and in studies on brain plasticity, is the interpretation of the direction of results, which is not always consistent across studies. Most structural studies report increases in measures of regional brain size or volume as a function of training or skill (Golestani, 2014), but this is not always the case, as has been discussed in recent papers on the structural correlates of bilingualism (García-Pentón et al., 2015) and of training (Thomas & Baker, 2013). This is also an issue with functional imaging studies on plasticity, and has also been discussed in published reviews on specific domains of processing (e.g. phonetic processing, see Golestani, 2015) and training (e.g. motor learning, see Dayan & Cohen, 2011), and on functional plasticity more generally (Kelly & Garavan, 2005). One possibility for reconciliation regarding the direction of training effects in brain structural studies is that local brain volume (or thickness, or other) may increase at early
3.2. Issues specific to studies on language

3.2.1. Anatomical variability of language areas, and definition of regional boundaries

A challenge associated with the structural imaging of receptive speech is the variability in the morphology of brain regions that are important for speech perception. Primary sulci and gyri are established early in development and are less variable between individuals than are secondary and tertiary structures (White et al., 2010). Despite this, even primary and secondary regions display considerable variability between individuals. HG, which includes primary auditory cortex, is a good example. HG differs in size and number of convolutions between hemispheres and individuals (Marie et al., 2015; Penhune et al., 1996). Given that according to conventional anatomical definitions, the supplementary convolutions of HG are considered as belonging to the PT, anatomical variability of the PT is also non-negligible. Furthermore, given that gross morphology (i.e. the location of gyri and of sulci) does not directly correspond to different cytoarchitectonic territories (e.g. for HG and primary auditory cortex, see Morosan et al, 2001; Rademacher et al., 2001), it is unknown to what extent gross anatomy reflects microstructure and related functional differences.

Another issue is that the definitions of the borders of specific ROIs are arbitrary, and thus structural differences may be present between groups when adopting one but not another definition. Several studies have attempted to circumvent this issue by employing multiple, different rules to define the borders of particular regions in the very same data (e.g. Alarelli et al., 2014; Marie et al., 2015), and then determining under which rule(s) group differences exist. Another way to work around this issue is to use relatively more data-driven approaches such as VBM to locate group differences. Other, more sophisticated computational morphometry methods have also been developed in an attempt to automate the parcellation even of such a small and variable ROI such as the auditory cortex, using for example landmark-based approaches (Engel, Toennies, & Brechmann, 2011). Yet others have used MR sequences that are sensitive to myelination (De Martino et al., 2015; Dick et al., 2012; Sigalovsky, Fischl, & Melcher, 2006; Wasserthal, Brechmann, Stadler, Fischl, & Engel, 2014), or pattern classification approaches (Schonwiesner, Dechent, Voit, Petkov, & Krumholz, 2015) to locate the primary auditory cortex in vivo. Automated methods can be more accurate and easy to develop for the segmentation and labelling of “closed” structures such as subcortical nuclei than for structures such as HG or the PT that lie on an open surface (in this case the superior temporal plane), since in the former case, signal intensity can be used to assist automatisation of the procedure (see, e.g. Coupe et al., 2011; Igual et al., 2011).

4. Key empirical contributions

Over the last 15–20 years, the number of studies having examined brain structural correlates of individual
differences in aspects of global (Sun & Hevner, 2014; Thompson et al., 2001; Toro et al., 2008) and local brain anatomy (Chiarello, Vazquez, Felton, & McDowell, 2016; Marie et al., 2015, 2016; Penhune et al., 1996; Rademacher et al., 2001; Tomaiuolo et al., 1999) has greatly increased. A medline search on brain structural MRI studies in relation to behaviour in healthy individuals shows exponential growth in the number of publications on this topic over the last 20 years. Several recent papers have reviewed the literature on the structural correlates of individual differences in behaviour and cognition (Kanai & Rees, 2011), of plasticity (Johansen-Berg, 2007) and on the possible microscopic, physiological bases of such differences (Tardif et al., 2016; Zatorre et al., 2012), attesting to the growing interest in this topic.

Studies have examined brain structure in relation to characteristics such as handedness or language lateralisatation (Chiarello, Vazquez, Felton, & Leonard, 2013; Dorsaint-Pierre et al., 2006; Good et al., 2001a; Greve et al., 2013; Leroy et al., 2015; Marie et al., 2015; McDowell, Felton, Vazquez, & Chiarello, 2015; Tzourio-Mazoyer et al., 2015) and sex (Davatzikos & Resnick, 1998; Good et al., 2001a; Leonard, Towler, Welcome, & Chiarello, 2009; Leonard et al., 2008; Welcome et al., 2009), which are known to modulate language processing and its neural bases (Chiarello, Welcome, & Leonard, 2012; Chiarello, Welcome, Halderman, & Leonard, 2009; Chiarello, Welcome, Halderman, Towler, et al., 2009). Also relevant to language processing, studies have explored brain structure in the auditory cortex per se (Marie et al., 2016; Penhune et al., 1996; Rademacher et al., 2001) and its relation to auditory (Foster & Zatorre, 2010; Sutherland et al., 2012; Warrier et al., 2009) and linguistic processing (Ressel et al., 2012; Seither-Preisler, Parnnutt, & Schneider, 2014), as well as to genetics (Cai et al., 2014).

In the domain of language, reviews have been published that provide overviews of brain structural imaging studies of language and of language-related processing in healthy individuals (Golestani, 2014; Richardson & Price, 2009; Zatorre, Belin, & Penhune, 2002). Reviews have also been published on the brain structural (and also functional) correlates of second language learning and bilingualism (Costa & Sebastián-Gallés, 2014; García-Pentén et al., 2015; Li, Legault, & Litcowsky, 2014; Stein, Winkler, Kaiser, & Dierks, 2014; Wong, Yin, & O’Brien, 2016). In this section, we will provide a brief overview of the key empirical contributions on brain structural differences and plasticity in relation to auditorily presented receptive speech (i.e. speech processing via the auditory and not via the written or other modalities).

The earliest studies on brain structural correlates of receptive speech addressed the question of brain structural differences in relation to individual differences in foreign speech sound learning. Apart from a few studies having examined brain structure in musicians (Schlaug, Jancke, Huang, & Steinmetz, 1995; Zatorre, Perry, Beckett, Westbury, & Evans, 1998), these were the first studies that attempted to relate individual differences in brain structure to individual differences in performance on higher-level, cognitive measures. In the first such study, 59 people were trained to identify the difficult, non-native dental-velar phonetic contrast. Using VBM followed by analyses on the location and thickness of sulci and then by DBM, it was shown that faster phonetic learners have more white matter in the parietal cortex compared to slower learners, especially in the left hemisphere (Golestani et al., 2002). This white matter difference could arise from greater myelination in the former group, which would allow more efficient neural processing in the parietal cortex and in connected brain regions. The inferior parietal cortex, together with auditory and frontal regions, is part of the dorsal stream known to be important for phonological processing and working memory (Aboitiz, 2012; Hickok & Poeppel, 2007; Paulesu, Frith, & Frackowiak, 1993; Rodriguez-Fornells, Unillera, Menestres-Misse, & de Diego-Balaguer, 2009). In a follow-up study, a different group of participants was trained to hear the same difficult phonetic contrast, and four complementary structural brain imaging analysis methods were used to look for differences between faster and slower learners: optimised VBM, DBM, manual labelling and analyses on sulci. Here, the original parietal cortex finding was partly replicated, with a greater left > right asymmetry in parietal lobe volumes in faster learners. There were also additional findings: both VBM and manual labelling revealed a larger left HG in faster learners, and this difference was relatively more driven by white than by grey matter differences. Also, faster learners were more likely to have multiple transverse gyri (sometimes called HG, although by definition the duplications belong to the PT) in the left hemisphere compared to slower learners (Golestani et al., 2007). A re-analysis of the data from the original study (i.e. Golestani et al., 2002), this time using optimised VBM, showed that the group difference in the left HG also existed in that original dataset. Last, the follow-up study also identified a positional difference in the location of the right insula and HG, with a relatively more superior position of these structures in the slower learners (Golestani et al., 2007). These findings suggest that more white matter in the left auditory cortex, and possibly more myelination and therefore greater auditory processing efficiency, may partly predict individual differences in the learning of speech sounds that rely on the processing of rapidly
changing acoustic information (i.e. stop consonants). The findings also suggest that faster and slower phonetic learners show global differences in the position of components of a right hemispheric language network, possibly reflecting individual differences in the functional anatomy and lateralisation of language processing. The finding of a larger left HG in faster phonetic learners converges with the results of a study that examined linguistic pitch learning, where individuals were trained to associate monosyllabic pseudowords which varied in terms of their pitch contour, with pictures. This study showed that people who were better at learning this task had larger volumes of left HG compared to the less successful learners (Wong et al., 2008). These auditory cortex findings in relation to phonetic and to linguistic pitch learning are also convergent with findings of larger HG volumes in children with better frequency modulation detection thresholds, an ability that is known to correlate with phonological awareness and literacy skills (Sutherland et al., 2012).

Another structural imaging study having examined the structural correlates of phonetic processing did so by examining brain structure in phonetics experts. VBM, SBM and manual labelling were used to examine brain structure in phonetics experts compared to non-expert, control participants. It was found that the auditory cortex (i.e. HG and the second or third transverse gyri, when they were present) was bigger bilaterally in phoneticians compared to controls. Also, as in the study on faster versus slower learners described above, the phoneticians were more likely than the controls to have multiple transverse gyri in the left hemisphere. The phoneticians also had larger grey matter volumes in the left pars opercularis (which corresponds to Brodmann’s area 44 of Broca’s area) compared to the slower learners, and within the phoneticians group, the surface area of this very region correlated positively with the number of years of transcription training experience across individuals (Golestani et al., 2011). The left pars opercularis is known to be involved in phonetic analysis and segmentation (Gough, Nobre, & Devlin, 2005; Nixon, Lazarova, Hodinott-Hill, Gough, & Passingham, 2004; Paulesu et al., 1993; Zatorre, Meyer, Gjedde, & Evans, 1996), and as mentioned above, together with the inferior parietal cortex, is part of the dorsal audio-motor integration stream (Hickok & Poeppel, 2007; Paulesu et al., 1993; Rodriguez-Fornells et al., 2009). These findings, consistent with the results of a DTI study in largely the same sample (Vandermosten, Price, & Golestani, 2016), suggest that phonetic transcription training results in experience-dependent structural plasticity in the left inferior frontal gyrus. In addition, the group differences in the auditory cortex, and in particular the gyrification difference, may predate the phonetic training, and may represent a brain structural phenotype predicting domain-specific aptitude for fine auditory processing (Golestani et al., 2011). This interpretation is in line with heritability studies having shown that HG morphology is more highly heritable than is that of Broca’s area (Peper, Brouwer, Boomsma, Kahn, & Poll, 2007; see also Zatorre, 2013 for a review on the contributions of nature versus nurture in speech and music processing).

Moving up from lower-level (i.e. phonetic) speech processing to the processing of phonetic information in a lexical context, two studies examined brain structure, seemingly in the same large group of Spanish-Catalan bilingual participants (N = 126), in relation to the ability to perceive a difficult vowel contrast from their second (but early learnt) language. Participants were tested both at pre-lexical (i.e. phonetic) and at lexical levels (i.e. within the context of words), and were assigned to two groups according to their ability to perceive this contrast. The first, VBM study revealed higher white matter volumes bilaterally in the insulo-fronto-opercular region in poor perceivers (Sebastián-Gallés et al., 2012). This result converges with those of a previous functional magnetic resonance imaging (fMRI) study having shown that people who are slower at learning to hear a foreign phonetic contrast recruit this same region in the left hemisphere more than do faster learners (Golestani & Zatorre, 2004). In the second study in the Spanish-Catalan bilinguals, group differences in cortical thickness and surface area were examined. It was found that poorer perceivers have thicker cortex as compared to better perceivers in regions including the left temporal cortex, angular gyrus and precuneus, consistent with the functional roles of these regions in phonologically based word recognition (Burgaleta et al., 2014).

To our knowledge, not many studies have examined brain structural correlates of spoken or receptive language processing at higher levels of language processing. Two such studies have examined brain structural correlates of degraded speech comprehension, at the word (Harris, Dubno, Keren, Ahlstrom, & Eckert, 2009) and sentence levels (Wong, Ettlinger, Sheppard, Gunasekera, & Dhar, 2010). In the first such study, fMRI and structural MRI were used to examine functional and structural correlates of the ability to comprehend words embedded in parametrically varying levels of noise. It was found that the volume of the left HG/superior temporal gyrus (STG) was positively correlated with the ability to recognise words in noise in both older and younger adults. Further, causal path modelling of the performance and of functional and structural imaging data suggested that in the older adults, age-related changes in left HG/STG morphology contribute to poorer performance on the speech in noise task.
Directly highlighted the functional relevance of fine anatomical investigations. Helbling and colleagues (2015) combined myelin mapping with magnetoencephalography, and showed that variation in myelination across the cortex predicts variation in electrophysiological response, and is thus of functional relevance (Helbling et al., 2015). This underscores the importance of better understanding the relationship between function and structure, in particular with respect to myelination, and opens very exciting avenues for the future of in vivo histology. Also, receptor mapping in language (and other) brain regions will continue to open new avenues for linking brain function and structure (Amunts et al., 2010).

Other promising avenues for future research are on the relative contributions of predisposition versus experience-dependent plasticity to brain structure–behaviour relationships. For example, in the auditory domain this has been done using molecular genetics (Cai et al., 2014; Guadalupe et al., 2015), and twin studies have also explored the contribution of predisposition to brain structure (Thompson et al., 2001). Conversely, a few longitudinal investigations of brain structural plasticity have been done in the domain of language-related training and expertise (Hervais-Adelman, submitted for publication), but more are needed. These longitudinal investigations are crucial in helping to disentangle whether individual differences in brain structure predate training, or whether they arise from experience-dependent brain plasticity. Also, work on brain structural development and its relationship to language is crucial (Dehaene-Lambertz & Spelke, 2015; Ortiz-Mantilla, Choe, Flax, Grant, & Benasich, 2010; Pujol et al., 2006), both for understanding brain development per se, and for addressing the nature-nurture question. Future work could jointly examine genetics and training, to see to what extent predisposition may facilitate domain-specific learning and associated structural plasticity.

Data analysis methods that will likely continue to lead to significant advances on the topic of brain structural correlates of language processing, in particular for the detection of finer/more subtle differences, include the use of “advanced VBM”, which relies on the use of finer normalisation algorithms such as Advanced Normalisation ToolS (ANTS, Avants, Epstein, Grossman, & Gee, 2008) and Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) (Ashburner, 2007). Also, despite the existence of the different automated labelling approaches and related atlases previously described in this review, novel approaches that are targeted at specific brain structures such as the hippocampus (Tangaro et al., 2015) and subcortical structures (Babalola et al., 2009) exist. These remain to be

From this overview it can be seen that structural imaging studies on receptive speech processing show structural correlates in largely the very regions known to be functionally involved in the respective tasks. As such, these studies are complementary to functional imaging studies, and are informative regarding how brain functional change gives rise to brain structural change.

5. Future directions

Brain structural imaging studies are complementary to functional ones in that they inform us about brain features that systematically vary in relation to a behavioural measure, or as a function of training and experience. Over the last years, methods such as quantitative structural imaging have been developed which contribute to our understanding of the brain and how it relates to behaviour beyond that made by other techniques, and in a manner that is more biologically meaningful. Also ultra-high resolution MRI, at 7 Tesla and now even at 9.4 Tesla in humans, allows for the acquisition of laminar resolution images. It can be used to examine brain structure per se, in relation to individual differences in language (or other) measures and to learning, or it can be used to highly improve localisation of brain function (Waehnert et al., 2014, 2016). To our knowledge, such studies have not to date been done on receptive speech per se. However, relevant to speech is a recent study on auditory processing that used high resolution quantitative R1 mapping (to index cortical myelination) together with tonotopic mapping of brain function in order to locate the primary auditory cortex (Dick et al., 2012). Similarly, laminar resolution myelin maps acquired at ultra-high resolution have enabled cortical depth-dependent assessment of myelin contrast within grey matter to locate the primary auditory cortex (De Martino et al., 2015). Importantly, a recent study has directly highlighted the functional relevance of fine structure predate training, or whether they arise from experience-dependent brain plasticity. Also, work on brain structural development and its relationship to language is crucial (Dehaene-Lambertz & Spelke, 2015; Ortiz-Mantilla, Choe, Flax, Grant, & Benasich, 2010; Pujol et al., 2006), both for understanding brain development per se, and for addressing the nature-nurture question. Future work could jointly examine genetics and training, to see to what extent predisposition may facilitate domain-specific learning and associated structural plasticity.

Data analysis methods that will likely continue to lead to significant advances on the topic of brain structural correlates of language processing, in particular for the detection of finer/more subtle differences, include the use of “advanced VBM”, which relies on the use of finer normalisation algorithms such as Advanced Normalisation ToolS (ANTS, Avants, Epstein, Grossman, & Gee, 2008) and Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) (Ashburner, 2007). Also, despite the existence of the different automated labelling approaches and related atlases previously described in this review, novel approaches that are targeted at specific brain structures such as the hippocampus (Tangaro et al., 2015) and subcortical structures (Babalola et al., 2009) exist. These remain to be
adapted however for language-related brain regions such as the auditory cortex that have very high inter-subject anatomical variability, and for the characterisation not just of volume but of the shape of specific brain structures. Related is the development of a tool called BrainPrint, which allows to capture shape information from an ensemble of cortical and subcortical structures, and to use these to predict which group people belong to (e.g. age group, sex, etc.), or to assess genetic influences on brain structure (Wachinger et al., 2015). Analysis of structural covariance (e.g. on measures of cortical thickness, see Lerch et al., 2006) looking at brain regions involved in receptive speech, or in relation to measures of language skill and learning are sparse (Mechelli, Friston, et al., 2005). Also, studies that examine both grey and white matter in the same individuals, and that combine different measures of brain structure and also different brain imaging modalities (e.g. structural MRI and DTI) in order to show convergence, are lacking. The recent studies having shown the functional relevance of fine and even microscopic (i.e. in vivo histology) indices of brain structure are very important catalysts for further significant developments in our understanding of the neural bases of language, and will likely multiply, thereby providing us with a better understanding of the physiological relevance of brain structural differences.

Acknowledgements

We wish to thank Dimitri Van De Ville, Antoine Lutti, Bogdan Draganski and two anonymous reviewers for their helpful comments on the methodological section of this review.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Swiss National Science Foundation grants awarded to N.G. [grant number PP00P3_133701], [grant number PP00P3_163756].

ORCID

Damien Marie http://orcid.org/0000-0001-8199-8260

References


