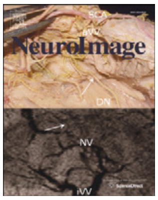
NeuroImage 64 (2013) 616–629

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Approaching expert results using a hierarchical cerebellum parcellation protocol for multiple inexpert human raters

John A. Bogovic [a,*], Bruno Jedynak [b], Rachel Rigg [e], Annie Du [f], Bennett A. Landman [d], Jerry L. Prince [a,c,e], Sarah H. Ying [c,f,g]

[a] Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA
[b] Department of Applied Mathematics & Statistics, Johns Hopkins University, Baltimore, MD, USA
[c] Department of Radiology and Radiological Science, Johns Hopkins Medical Institutions, Baltimore, MD, USA
[d] Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA
[e] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
[f] Department of Neurology, Johns Hopkins Medical Institutions, Baltimore, MD, USA
[g] Department of Ophthalmology, Johns Hopkins Medical Institutions, Baltimore, MD, USA

ARTICLE INFO

Article history:
Accepted 27 August 2012
Available online 4 September 2012

Keywords:
Human cerebellum
Manual labeling
Delineation
Parcellation
STAPLE
STAPLER
Label fusion

ABSTRACT

Volumetric measurements obtained from image parcellation have been instrumental in uncovering structure–function relationships. However, anatomical study of the cerebellum is a challenging task. Because of its complex structure, expert human raters have been necessary for reliable and accurate segmentation and parcellation. Such delineations are time-consuming and prohibitively expensive for large studies. Therefore, we present a three-part cerebellar parcellation system that utilizes multiple inexpert human raters that can efficiently and expediently produce results nearly on par with those of experts. This system includes a hierarchical delineation protocol, a rapid verification and evaluation process, and statistical fusion of the inexpert rater parcellations. The quality of the raters' and fused parcellations was established by examining their Dice similarity coefficient, region of interest (ROI) volumes, and the intraclass correlation coefficient of region volume. The intra-rater ICC was found to be 0.93 at the finest level of parcellation.

Introduction

The cerebellum has a remarkably complex structure that coordinates numerous vital functions of the human body. It is involved in tasks such as eye-movement (McCormick and Thompson, 1984; Ritchie, 1976), speech (Silveri et al., 1994), balance, fine motor control, motor learning, and cognition (Leiner et al., 1986; Schmahmann, 1991). Like the cerebral cortex, the human cerebellum exhibits functional localization. This is reflected in part by its anatomic structure. There are two macroscopic levels of organization: medial–lateral and rostral–caudal. The medial–lateral anatomical divisions of the cerebellum echo the differences in connectivity between the medially located spinocerebellum, and the lateral cerebrocerebellum, or "neo-cerebellum." The spinocerebellum consists of the wormlike "vermis" and the more lateral paravermis, or "intermediate zone." As its name suggests, these regions receive afferents primarily from the spinal cord. The evolutionarily ancient flocculonodular lobe or "vestibulocerebellum" is intimately associated with the vestibular system, and therefore highly influences spatial orientation and balance. In the rostral–caudal direction, transverse fissures create divisions in the cerebellum called lobules.

The study of cerebellar substructures has been confounded by inconsistencies in nomenclature. This has been remedied in the study of humans by the general acceptance of the standard introduced by Schmahmann et al. (1999, 2000). In their work, the cerebellar lobules are numbered from I to X. These lobules stem from white matter branches rooted in the central mass of cerebellar white matter, called the corpus medullare (CM). Under this convention, lobule I is located most rostrally, with lobule numbering increasing caudally. In this work, we adopt the numbering standard of Schmahmann, but will refer to super-groupings as follows: anterior (I–V), superior posterior, or middle (VI and Crus I and II of VIIA, and VIIB), inferior posterior (VIII, IX), and caudal (VIII, IX, and X). Three of these conventions and super-groupings are illustrated in Fig. 1.

Region-specific changes in the cerebellum have been correlated with a number of diseases and functional deficits. For example, regionally selective degeneration of the vermis and anterior lobe has been observed over the course of aging (Andersen et al., 2003; Raz et al., 1998). A decrease in size of the inferior posterior vermis has been observed in boys with attention-deficit and hyperactivity disorder (ADHD) relative to normals (Berquin et al., 1998; Mostofsky et al., 1998). Several studies have shown changes in volumes of the vermis (Nopoulos et al., 1999; Okugawa et al., 2003) and vermian white matter (Levitt et al., 1999) in patients with schizophrenia. Evidence suggests that the vermis and

---

* Corresponding author.
E-mail address: bogovic@jhu.edu (J.A. Bogovic).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.08.075

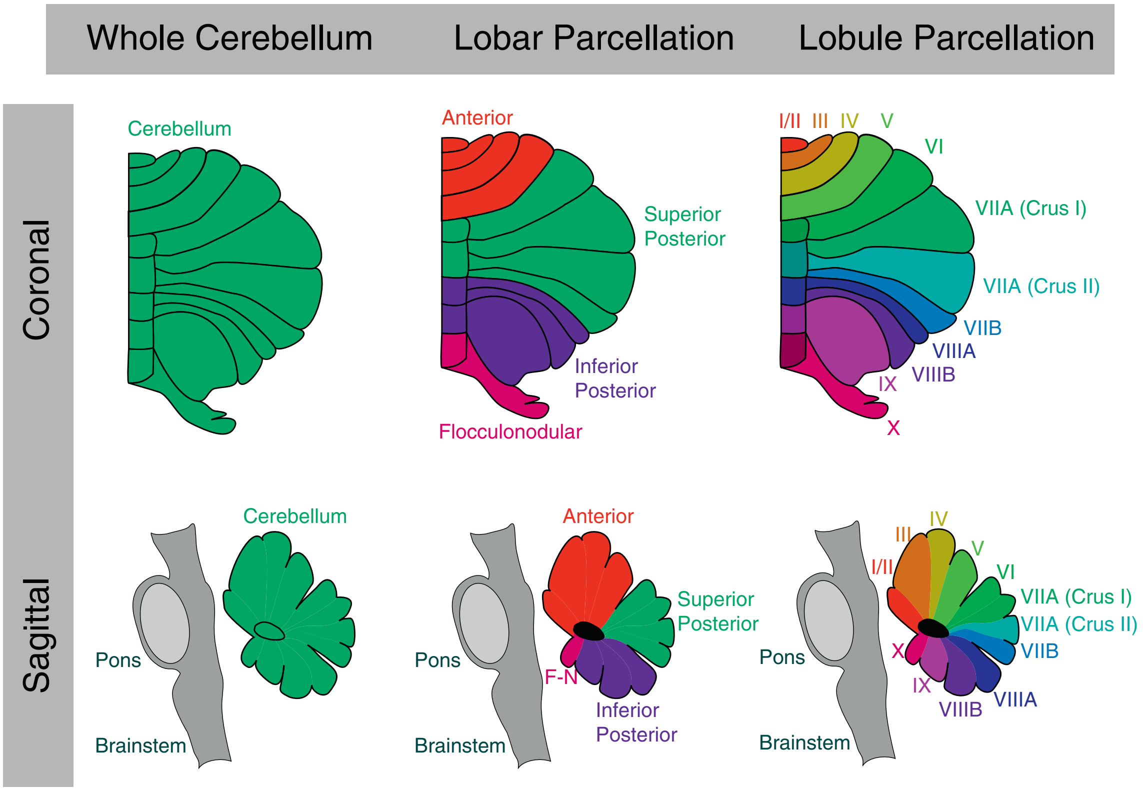


Fig. 1. Illustration of cerebellar anatomy and lobule grouping.

flocculus are targeted by chronic alcoholism (Baker et al., 1999; Cavanagh et al., 1997). Superior posterior lobe volumes have been shown to decrease in patients with Alzheimer's disease relative to controls (Thomann et al., 2008). Several types of cerebellar ataxia have also demonstrated region-specific atrophy within the cerebellum (Brenneis et al., 2003; Jung et al., 2011; Ying et al., 2006).

Clearly, the success of such studies depends on accurate and precise measurements for the structures of interest. Segmentations produced by human experts remain the gold standard despite the progress achieved in automated segmentation algorithms. However, the training of experts is a very long process, typically requiring thousands of hours. The high quality measurements produced by such raters are therefore time-consuming and expensive. Here we briefly review examples of three approaches that have been used to limit the amount of human expertise required: 1) limit the scope of the study, 2) employ automated or semi-automated image analysis methods, or 3) employ inexpert human raters.

In order to make best use of the experts' time, research hypotheses may be tested using a small cohort of subjects, on a small number of subregions, or using a coarse parcellation. For example, one may limit study to the cerebellar vermis. Raz et al. (1998) measured the cross-sectional area of the cerebellar vermis grouping lobules I–V, VI–VII, and VIII–X. A coarse parcellation considered by Levitt et al. (1999), among others, consists of hemispheric white and gray matter, vermian white matter, and gray matter split into three subgroups consisting of lobules I–V, VI–VII, and VIII–X.

A few methods have been introduced that produce a full parcellation of the cerebellar lobules from magnetic resonance (MR) images. A semi-automatic surface-based method presented by Makris et al. (2003, 2005) has the advantage of parcellating the cerebellum into medial-to-lateral subdivisions. Pierson et al. (2002) developed manual and semi-automated methods for delineating the corpus medullare, anterior lobe, superior posterior lobe, and inferior posterior lobe. These manual and semi-automated methods have produced good results, but rater training and delineation time are not discussed. Significant knowledge of cerebellar anatomy is required to follow these protocols, and their use will yield adequate results only when used by expert human raters. As a consequence, they tend also to be time-consuming and expensive.

To our knowledge, there is currently only one publicly available automated method for cerebellar parcellation. Diedrichsen (2006), Diedrichsen et al. (2009) describe an automatic method based on nonlinear registration of a cerebellum label template (SUIT). This methodology was employed in (Donchin et al., 2012) in the case of focal lesions and cerebellar atrophy and showed lobule-specific changes using voxel-based morphometry. However, an explicit evaluation of segmentation performance was not performed. For our data, the SUIT template produced a mean Dice similarity[1] (standard deviation) of 0.55 (0.16) across 15 subjects and 24 labels (see Appendix A for details). Since this method uses a template constructed from control subjects, it is not surprising that results are poor for patients with severe cerebellar degenerative disease (especially those including diverse forms of lobule-specific atrophy). Our data, which include nine subjects with significant degeneration, demonstrate this phenomenon. In any case, testing this approach (or any alternative) would require numerous manual parcellations for validation. The present paper describes a cost-effective approach to produce such manual parcellations on both normal and ataxia subjects.

Recently, the "multi-atlas" segmentation framework has been shown to achieve excellent results for many anatomical segmentation tasks (Heckemann et al., 2006; Isgum et al., 2009). We explored the performance of such a method (see Appendix A for details) for cerebellar parcellation and found that it produced unsatisfactory results: mean Dice similarity (standard deviation) across 15 subjects and 28 labels of

---

[1] The Dice similarity coefficient (DSC) (Dice, 1945) was computed with the "gold standard" expert rater delineation. For a given label l, the DSC of a parcellation with the expert parcellation is given by: $DSC(l)=2(|S_{Rl}\cap S_{El}|)/(|S_{El}|+|S_{Rl}|)$, where S_{Rl} and S_{El} denote the sets of voxels assigned to label l by the rater and expert, respectively.